Winter 12-4-2009

Development of the Mask Scentometer, a Comparison of Ambient Odor Assessment Methods, and their Application in Ground Truthing Atmospheric Dispersion Models

Christopher G. Henry
University of Nebraska at Lincoln, chenry1@unl.edu

DEVELOPMENT OF THE MASK SCENTOMETER, A COMPARISON OF AMBIENT ODOR ASSESSMENT METHODS, AND THEIR APPLICATION IN GROUND TRUTHING ATMOSPHERIC DISPERSION MODELS

by

Christopher G. Henry

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Engineering
(Agricultural and Biological Systems Engineering)

Under the Supervision of Professor Dennis D. Schulte

Lincoln, Nebraska

December, 2009

DEVELOPMENT OF THE MASK SCENTOMETER, A COMPARISON OF AMBIENT ODOR ASSESSMENT METHODS, AND THEIR APPLICATION IN GROUND TRUTHING ATMOSPHERIC DISPERSION MODELS

Christopher Garrett Henry, Ph. D.

University of Nebraska, 2009

Adviser: Dennis D. Schulte

This dissertation is organized as four stand-alone papers. Paper No. 1 describes the development of the Mask Scentometer and reports dilution ratios measured during use by twelve different people. Dilution ratios at the Mask Scentometer's five dilution-to-threshold (D/T) settings were found to be 0.35, 1, 2, 4.5 and 18. In Paper No.'s 2 and 4, ambient odor assessment methods were compared in both controlled laboratory conditions and in the field. Laboratory analysis of ambient air samples using dynamic triangular forced-choice olfactometry (DTFCO) did not correlate well with any of the ambient odor assessment methods. Average intensity-predicted D/T was roughly five times higher than D/T measured using a Nasal Ranger®, and D/T obtained using a Nasal Ranger® was roughly two to five times higher than corresponding D/T from a Mask Scentometer. In Paper 3, odor intensity ratings and Mask Scentometer readings were used to calibrate the AERMOD dispersion model for predicting odor concentrations downwind of area sources. Dispersion of odor from a swine waste treatment lagoon and two cattle feedlots was modeled with AERMOD and the predictions were compared to the observations using a statistical approach to develop scaling factors. These were found to be 12 for odor intensity and 0.15 for the Mask Scentometer (although a scaling factor between 0.5 and 0.7 is also justified). Random effects and autocorrelation were found to be significant in ambient odor assessment data. In Paper 4, the dispersion of odors from a swine production building complex was studied. CALPUFF and AERMOD were used to predict short-time-step (one minute) odor concentrations. Source emission measurements and meteorological data were collected to coincide with ambient odor measurements obtained using the Nasal Ranger®, Mask Scentometer, field odor intensity ratings, and DTFCO. In general, odor concentrations predicted using CALPUFF were found to be about twice those predicted by AERMOD. Model predictions agreed best with the readings from the Nasal Ranger® and Mask Scentometer; and both of these ambient odor assessment methods are suited for ground truthing AERMOD and CALPUFF, although some model scaling factor adjustment is needed.

ACKNOWLEDGEMENTS

I dedicate this dissertation to my wife Heather and son Connor, I am thankful for their support and steadfast love during this pursuit. Many evenings and weekends were lost between us during the writing of this Ph. D. The love and support of my parents, Bob and Janis have been unconditional and without question, for the weekends they came to watch Connor so that I could "work on the Ph.D."

I wish to thank my advisor, Dennis Schulte, for the "brutal" demands I placed on him during this endeavor. Many others at UNL contributed to my success, Rick Koelsch, my former supervisor, allowed and encouraged me to improve myself. Rick was also instrumental in acquiring the original Nebraska Department of Agriculture grant which was the foundation of Paper No. 3. I want to express appreciation to Rick Stowell and David Billesbach for their review of the first draft of my dissertation. Special thanks to Anne Parkhurst for the many meetings in her office going over SAS and R code for the statistics presented in this work. I cannot thank Jason Gross enough, for his "holding up the fort" of our small AFO Extension program at times when I needed to do "odor work," especially during my Fulbright.

Larry Jacobson and Steve Hoff were instrumental in this work, and I am most grateful for their efforts and support. Without Steve's ambition and dedication, the USDA NRI project would not have been so successful. Much of the data presented in this dissertation is the result of Steve's hard work and resolve during the experiments.

To Ron Sheffield, for his work on the Mask Scentometer, and for sharing his original prototype, from which I further improved and developed for one-handed operation.

I wish to also thank Robin Ormerod, Kristin Zeise, Geordie Galvin, and Peter D'Abreaton, all of PAE Holmes Ltd, of whom I am deeply grateful. They were instrumental in completing the CALPUFF modeling for Paper No. 4. Through no fault of their own, we found modeling errors in the first runs, and it took another two to "get it right." But they came through for me, although I am sure that they considered it the "project that never ends." Their insights into modeling odors (or odours in Australia), was extremely valuable, and I wish I would have had more time in the Brisbane office with them while I was on my Fulbright.

GRANT INFORMATION

This work was supported by the following funding agencies:

Nebraska Department of Agriculture

United States Department of Agriculture, Cooperative State Research, Education and Extension Service, National Research Initiative

Australian-American Fulbright Commission and the Institute of International Education

TABLE OF CONTENTS

LIST OF TABLES

Paper No. 1

Paper No. 2

LIST OF TABLES, CONTINUED

Paper No. 3

LIST OF TABLES, CONTINUED

Paper No. 4

LIST OF FIGURES

Paper No. 1

Paper No. 2

LIST OF FIGURES, CONTINUED

Paper No. 3

LIST OF FIGURES, CONTINUED

INTRODUCTION AND ORGANIZATION OF DISSERTATION

Odors from livestock and poultry production facilities can generate nuisance complaints and have become an issue that limits the viability and growth of the agricultural sector in the United States. Largely because of odor concerns, producers looking to build new facilities often face arduous public hearings, conservative separation distances and setback requirements, and the prospects of lawsuits being filed against them. Producers are encouraged to screen sites and plan facilities for minimal odor risk to neighbors. Unfortunately, quantifying and predicting odor risk, especially from diffuse area sources is not an easy task. Uncertainty in differentiating between perceived and potential odor risk to neighbors is a major impediment to finding common ground toward minimizing odor problems and reducing barriers for growth in the livestock sector. Two challenges in overcoming this impediment involve developing trustworthy tools to predict odor conditions – and resulting odor risk – and accurately measuring existing odor conditions in areas around livestock operations.

Atmospheric dispersion models and olfactometry offer great promise to address these challenges. Odor and gas dispersion from livestock facilities is a complicated process that depends on many factors, such as the production system, stocking density, season, localized weather patterns, terrain, and receptor locations relative to the production areas. Atmospheric dispersion models harness the power of computers to simulate the atmospheric processes that control the dispersion of airborne contaminants and predict the fate and transport of those contaminants in the atmosphere. Olfactometry encompasses many methods and procedures involving the science of smell, for laboratory and ambient applications. This dissertation relates measurements from existing ambient odor assessment methods to the output of two popular atmospheric dispersion models

(AERMOD and CALPUFF). It was thought that this feat had yet to be accomplished in any published work.

This dissertation differs from the traditional thesis in that chapters are presented as stand-alone manuscripts. Some of the manuscripts include additional tables, figures, original data, and discussion that will not be submitted for publication, but are presented at the end of each respective paper to provide a more complete research record.

Paper No. 1, entitled the "Mask Scentometer for Assessing Ambient Odors," discusses scentometer development and reports results of a study on the operating parameters of the Mask Scentometer. Previously, it was assumed that the Mask Scentometer presented the user with the same dilution levels as the original Barneby and Sutcliff Box Scentometer, from which it is derived. Pressure transducers and the orifice relationship were used to measure the inhalation airflow rates and resulting dilution ratios through the Mask Scentometer during use by twelve panelists. The objective of this study was to determine the dilution-to-threshold (D/T) settings of the Mask Scentometer. The Mask Scentometer is a common thread in this dissertation, and the D/T settings established for the Mask Scentometer in this initial work are embodied in Papers 2-4.

Paper No. 2, entitled "Comparison of Ambient Odor Assessment Techniques in a Controlled Environment," compares ambient odor assessment methods in a laboratory setting where the ventilation air was controlled. In this work, which was part of a United Stated Department of Agriculture National Research Initiative project, odor assessors were presented different swine odor levels in a room within the Iowa State University Air Dispersion Laboratory. Some assessors took D/T readings using field olfactometers (either the Nasal Ranger® or a Mask Scentometer), while others rated odor intensity using an odor intensity reference scale (OIRS) and air samples were collected for analysis in an olfactometry laboratory. The objective of this paper was to compare and relate ambient odor assessment methods.

Paper No. 3, entitled "Ground Truthing AERMOD for Area Source Livestock Odor Dispersion using Odor Intensity and the Mask Scentometer," reports scaling factors that were developed for modeling odor transport from waste treatment lagoons and cattle feedlots. The scaling factors will further development of the Nebraska Odor Footprint Tool (NOFT) and related simple tools, which assist livestock producers and the general public in assessing odor risk. This pioneering work with large area sources employed both D/T readings from Mask Scentometers and odor intensity ratings (OIRS) to ground-truth a dispersion model. To our knowledge this was the first time that more than one ambient odor assessment method was used in the field in this manner. The objectives of this paper were to determine scaling factors for large area sources and to compare two ambient odor assessment methods when used to calibrate a dispersion model.

Paper No. 4 is entitled "Ground Truthing CALPUFF and AERMOD for Odor Dispersion from Swine Barns using Ambient Odor Assessment Techniques." This work was conducted on a four-building, swine-finishing site that was participating in the Aerial Pollutant Emissions from Animal Concentrated Buildings (APECAB) Study and took advantage of instrumentation and logistics that were put in place for that study. Intensive odor sampling of the barns' emissions was done and samples were analyzed in the olfactometry labs at both Iowa State University and the University of Minnesota, likely making it the most intensively sampled livestock odor dispersion project in the United States. Atmospheric dispersion models used micro-meteorological data collected on the site to predict odor concentrations at downwind locations where odor assessors rated odor intensity and used the Nasal Ranger® and the Mask Scentometer to take D/T readings. The objective of this paper was to statistically analyze the relationships between the ambient odor assessment techniques and dispersion model predictions (CALPUFF and AERMOD) for building sources.

The following research questions are posed and answered during the course of this dissertation.

1. What are the correct dilution-to-threshold ratios for the Mask Scentometer?
2. What are the relationships between the commonly used ambient odor assessment techniques?
3. What scaling factors are appropriate for modeling air dispersion from area sources?
4. How reliable are atmospheric dispersion models in predicting ambient odor and what effect does the ambient odor assessment method have in this evaluation?

Paper No. 1

MASK SCENTOMETER FOR ASSESSING AMBIENT ODORS

C. G. Henry, G. E. Meyer, D. D. Schulte, R. R. Stowell,

A. M. Parkhurst, and R. E. Sheffield

ABSTRACT. This paper summarizes the development and operation of a Mask Scentometer and reports air dilution ratios measured during its use, which were used to establish the device's dilution-to-threshold settings. The Mask Scentometer is a facial respirator that has been modified to operate conceptually like the Barneby and Sutcliffe Box Scentometer. The Mask Scentometer is comprised of a ¼-face respirator with two modified spin-on cartridges, one per side, which facilitate mixing ambient air with filtered air for presentation to an odor assessor at user-selected dilution ratios. The 'clean side' cartridge includes an activated carbon filter canister with two half-inch orifices for metering clean air into the mask chamber. The 'ambient side' cartridge includes an adjustable dial with five different orifices for metering ambient air directly into the mask [unfiltered] at a range of flow rates. Prior to this study, the dilution ratios of air presented to an assessor using the Mask Scentometer were assumed to be the same as the dilution-to-threshold settings of the Barneby and Sutcliffe Box Scentometer: 170, 31, 15, 7 and 2 volumes of clean air per volume of ambient air. In a controlled laboratory environment, airflow rates were measured through both cartridges of a Mask Scentometer using a pressure transducer while twelve different assessors used the device. The flow-weighted average dilution ratios produced within the Mask Scentometer were 18, 4.5, 2, 1, and 0.35.

Investigators using the Mask Scentometer to measure ambient odor concentrations are advised to use these dilution-to-threshold values.

INTRODUCTION

Field olfactometry can be defined as a field technique or device used to determine the dilution to threshold (D/T) of an odor – that is, the odor concentration - in the ambient atmosphere. The use of one or more of these techniques or devices may also be referred to as ‘nasal ranging’, ‘field sniffing’, ‘field odor assessing’, and ‘field odor surveying.’

Laboratory-based approaches, like dynamic triangular forced-choice olfactometry (DTFCO), typically are not very effective at assessing the low odor concentrations encountered in the ambient air, such as those downwind of a livestock facility (CEN, 2003). The cost to analyze air samples using an olfactometry laboratory can be very expensive ($100-$300 or more per sample). By using field olfactometry, one can obtain more data over a longer period of time for less expense, making these approaches attractive for calibrating and verifying models, as well as for general assessments of odor (Sheffield and Ndegwa, 2008). For example, using a single laboratory olfactometer, with DTFCO one is generally able to analyze at most a dozen ambient air samples (collected in 10-50 L containers). With field olfactometry, an individual can quickly and easily take many more readings (several per minute), which enables assessors to better capture fluctuating ambient conditions and collect data over a much longer period of time.

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 initially developed by Sheffield (2004) and improved by Henry (2004). The Box Scentometer was first developed in the 1950's as a result of grants provided by the US Public Health Service (Huey et al., 1959 and 1960). The original Box Scentometer was manufactured and sold by the Barnebey-Chaney Company, which later became the Barnebey and Sutcliffe Corporation (Barneby and Sutcliffe Corporation, 1959; Huey et al., 1960; Barneby and Sutcliffe Corporation, 1995), now owned by the Calgon Carbon Corporation. The Box Scentometer works by diluting incoming ambient air with cleaned air (air that has been passed through an activated carbon filter) to obtain discrete dilution ratios. Two Box Scentometer models were developed: a 1959 model and the 1959-A model. The 1959-A model included a second filter (top and bottom) to double the filtering capacity.

Huey et al. (1960) defined the dilution ratio (or dilution factor) of a scentometer as:

Dilution ratio = Volume of carbon-filtered air / Volume of odorous air

On a dynamic basis, the dilution ratio is a ratio of airflow rates. A 'dilution-to-threshold' or D/T reading is the dilution ratio that corresponds with reaching the assessor's odor detection threshold.

The disadvantages of the Box Scentometer include lack of control of inhalation rates by assessors, discomfort of the glass inhalation tubes, odor fatigue caused by poor nasal sealing to the glass ports, and inability to prevent physical fatigue during measurements and olfactory fatigue between measurements. The poor sealing to the glass tubes likely contributes to measurement error, resulting in higher D/T than actually exist.

The second-year US Public Health Service report (Huey et al., 1959) describes how the Box Scentometer D/T settings were originally verified. A frictionless piston created with a soap bubble in a cylinder was used to measure the flow from each of the orifices. Air was withdrawn from the device (by a human nose for the 1959 model and using an air pump for model 1959-A) and the volume swept by the piston was assumed to be a measure of the airflow rate through the orifice(s) during the flow period. As shown in Table 1, the dilution ratios for the four D/T settings are not exactly equal to the theoretical dilution ratios that are supplied for the instrument. Modifications made to the 1959 model (an additional carbon chamber and reconfigured orifices) made the dilution ratios measured with the 1959-A model perform closer to the theoretical ratios. One shortcoming in the verification is that air flow was only measured under one flow rate for a very short period of time.

Table 1. Comparison of air dilution ratios with the Barneby and Sutcliffe Box Scentometer (Huey et. al., 1959)

	Model 1959	Model 1959-A	
Theoretical dilution ratio	Measured dilution ratio	Measured dilution ratio	Published D/T setting
2	1.8 (range 1.8-1.9)	1	2
8	9.2 (range 8-11)	6.2	7
32	27 (range 18-36)	31	31
128	77 (range 64-98)	169	170

The Nasal Ranger® provides a better seal, improved control of airflow rates, and more comfort for panelists during inhalation than the Box Scentometer. Physical fatigue with the hand-held device still occurs during measurements, though, and olfactory fatigue can be a problem if a mask is not worn between measurements. The cost of the device may also be an issue.

The Mask Scentometer (Henry, 2004; Sheffield et al., 2004) is functionally similar to a Box Scentometer, but with a more convenient configuration. It is comprised of a ¼-face respirator mask with two cartridges, one on each side. One cartridge has a charcoal filter with a plenum that has two ¼-inch (0.63 cm) holes in it. As ambient air is drawn through this cartridge, the air is scrubbed of odors before entering the mask chamber. The other cartridge has a dial mechanism with holes of various sizes that allow an assessor to adjust the ratio of ambient air into the mask chamber where dilution with carbon filtered air occurs. Advantages of the Mask Scentometer are that it minimizes discomfort and odor fatigue, allows for one-handed operation, and can be operated over long periods of time without operational fatigue. It also costs less to construct than a Nasal Ranger®.

Thus, there are significant potential advantages to using field olfactometry and, specifically, to using a Mask Scentometer to assess ambient odor. The actual dilution characteristics of a Mask Scentometer had previously not been measured, though. The orifice diameters in the Mask Scentometer are the same as those in the Box Scentometer, which led to the presumption that air was presented to the assessor at dilution ratios of 170, 31, 15, 7 and 2. The hypothesis that these presumed dilution ratios are accurate was tested in this project.

PURPOSE OF WORK

The dilution ratios delivered by the Mask Scentometer were thought to be equivalent to those of the original Barneby and Sutcliffe Box Scentometer. The purpose of this project was to test this hypothesis. This paper reports the measured dilution ratios of the Mask Scentometer while incorporating the influence of assessor sniffing and breathing effects on the flow rates, and thus dilutions, achieved during normal use. Since a Mask

Scentometer is worn continuously during an odor assessment period, the user must breathe through the unit even when not taking a reading (sniffing). There are no airflow control devices, thus the Mask Scentometer provides a 'working' dilution-to-threshold determination that are affected by a user's entire breathing pattern during an assessment.

DESCRIPTION OF DEVICE

The Mask Scentometer, shown in Figure 1, is comprised of a butyl rubber ¼-face respirator mask (manufactured by North™ Model 7700-30, P/N 7700-1114) with two cartridges. One cartridge, referred to as the 'clean side' or 'clean air delivery cartridge', has two orifices (Figure 2) and a canister of activated carbon medium (organic vapors and acid gases, P/N N7500-3) for cleaning one air stream. The other cartridge has five holes (orifices) of varying sizes. This 'ambient air delivery cartridge' (Figure 3) has a fixed Plexiglas disk with a single half-inch-diameter hole in it. A Plexiglas dial with a series of orifices (1/2", 1/4", 3/16", 1/8" and 1/16" diameters) is mounted flush with the disk and can be rotated about a plastic bolt to introduce different amounts of ambient air to the mask chamber and present a range of air dilution ratios to the user. The dial and cartridge with the series of orifices has a thin rubber gasket between them to prevent air leaks. During operation, users 'sniff' while cycling through the five ambient air delivery ports, proceeding from the smallest to the largest orifice (largest to smallest dilution ratio). Users adjust the dial every 5-8 seconds, which generally allows a D/T measurement to be taken every 30 seconds (time needed to cycle though all five ports).

Figure 1. Mask Scentometer as worn during use.

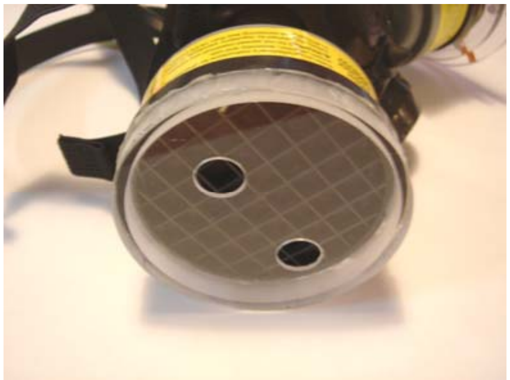

Figure 2. Clean air delivery side of the Mask Scentometer with filter cartridge. Air is drawn through the two half-inch-diameter orifices and activated-carbon medium. A half-inch plenum separates the intake orifices from the carbon filter medium. The device is operated with both orifices open.

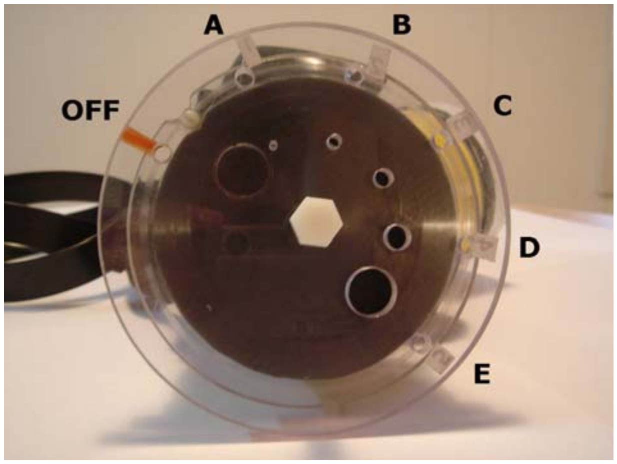

Figure 3. Ambient air delivery side of the Mask Scentometer. The adjustable dial has five functional D/T settings (A-E), each corresponding to a different orifice size, ambient air delivery rate, and dilution ratio. The device is shown in the "OFF" position, so only air enters from the filtered 'clean' side is presented to the user .

MATERIALS AND METHODS

To assess the actual dilution ratios achieved during use of the Mask Scentometer, the flow rates through each side of the mask, clean and ambient, must be known. The orifices on the Mask Scentometer were assumed to be sharp-edged holes of negligible thickness, so the orifice equation was used (Munson et al., 1994):

$$Q = CA\sqrt{\frac{2(P_1 - P_2)}{\rho}}$$

Where,

Q = Flow rate in cm^3/s

C = Orifice coefficient and correction factor, found experimentally to be 1.57 for the clean side and 1.08 for the ambient side (the same factor could be applied to all orifice sizes on the ambient side cartridge)

A = Orifice area, cm^2

$P_1 - P_2$ = Pressure difference across the orifice, reported in Pascals (kg-m/s^2-m^2, or g-cm/s^2-cm^2). In this application, P_1 is the pressure inside the cartridge and P_2 is atmospheric pressure outside the cartridge (measured at the pressure transducer housing)

ρ = air density, 0.001184 g/cm^3 at 25 degrees Celsius

Airflow rates (Q) through the cartridge orifices - which determine the dilution ratio of air within the Mask Scentometer - were measured using pressure transducers. A Mask Scentometer was instrumented by installing small ports just inside the cartridges near the orifices. This, allowed 1/8-inch ID vinyl hoses to be connected to differential pressure transducers. Differential pressure transducers rated for 0-2.5 inches of water (Omega Engineering, PX-160) were installed on each cartridge. These were interfaced to a datalogger. The pressure transducer specifications were ±1.0% linearity, a full scale repeatability of ±0.25%, and a 1 ms response time (Omega Engineering, ND). A Labjack™ U12 datalogger (www.labjack.com) capable of making 50 measurements per second was used to read the data from the pressure transducers. A program was developed in Labview (National Instruments Labview 8.5, www.ni.com), to collect data from both of the transducers. Labview’s Virtual Instrument (VI) platform was programmed to show, graph, calculate and save the flow rates of the Mask Scentometer as well as the corresponding dilution ratios.

Verification of the data acquisition system was done by comparing its response to known flow rates measured by mechanical ball rotameters (flow meters). Flow rates calculated from measured pressure differentials across the orifices were verified against the actual flow rates delivered by a vacuum pump with a regulating valve and a rotameter. For airflow rates between 1-10 L/m, two air sampling pumps (with integrated rotameters) were used in parallel. For airflow rates between 10-50 L/min, a large rotary vane vacuum pump (Busch RC 0021 15 CFM) and rotameter were used. Flow rates through each of the five orifices in the ambient air delivery cartridge and through the two clean air orifices were measured over the expected range of flow rates. Flow was maintained for at least five minutes after a change to ensure that steady-state flow and pressure conditions prevailed.

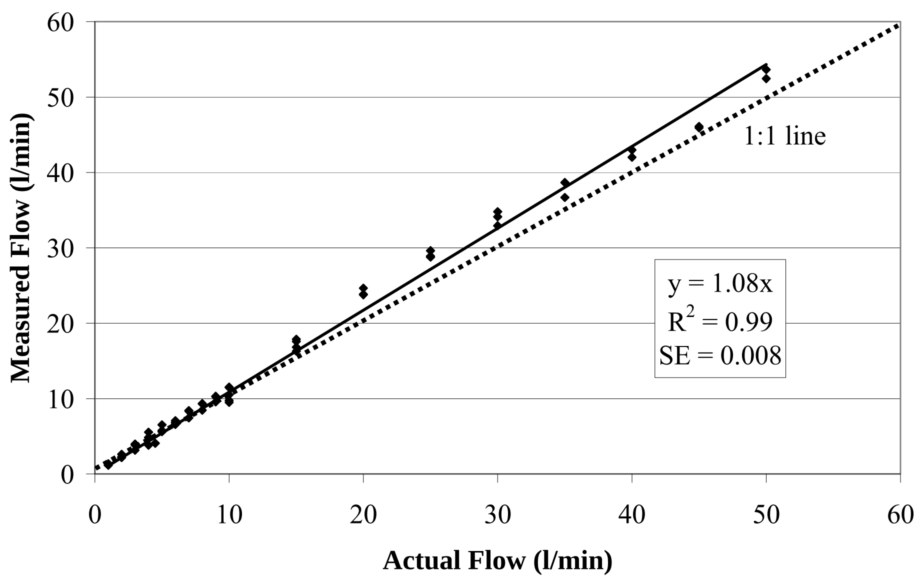

Figure 4. Airflow rates through the ambient air delivery cartridge as measured using pressure transducers compared to actual flow rates as measured directly using a rotameter (all orifices).

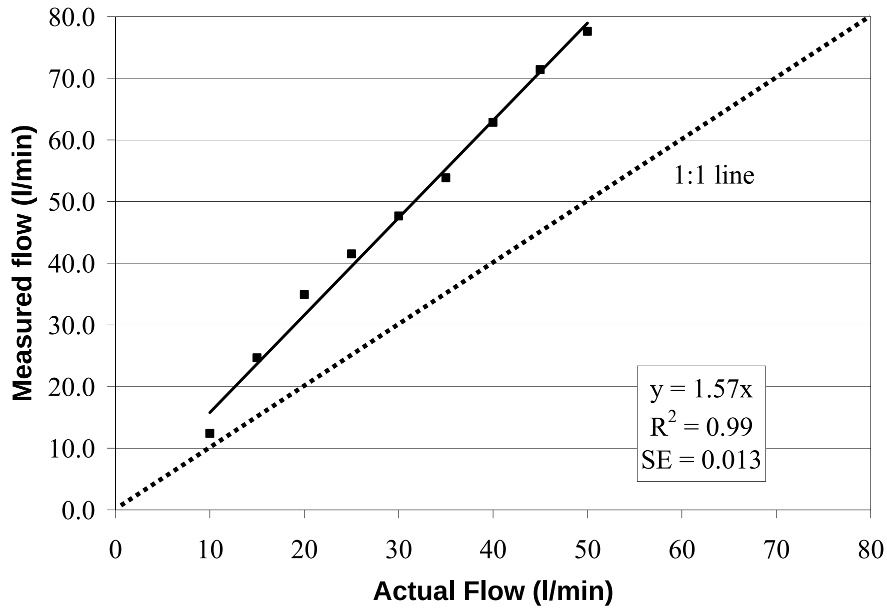

Figure 5. Airflow rates through the clean air delivery cartridge as measured using pressure transducers compared to actual flow rates measured directly using a rotameter.

Comparison of measured and actual flow rates yielded strong linear relationships (zero-intercept $R_o^2 = 0.99$) for both the clean and ambient air delivery sides. The slopes shown in Figure 4 and Figure 5 become the correction factors or orifice coefficients (C) for each cartridge. The correction factor is larger for the clean air delivery side than for the ambient side. This is thought to be due to the pressure loss across the carbon filter medium. The lowest pressure difference that could be measured was 0.025 cm of water. Figure 4 represents flows for all orifices, so the correction factor developed could be the same for all settings. Each setting was tested from it's lowest measurable flow rate to it's highest flowrate or 50 L/min. Preliminary testing had indicated that maximum flowrates for a person using the Mask Scentometer would be around 30-40 L/min, so factors were done at 1 L/min increments to 10 L/min, then 5 L/min increments until 50 L/min. Flow ranges were 1-4 L/min, 1-10 L/min, 4-15 L/min, 10-50 L/min, and 15-50 L/min for settings A, B, C, D, and E ambient cartridge settings, respectively. For the clean air delivery cartridge, airflow below 10 L/min could not be measured (pressure difference was less than 0.025 cm of water (0.01 inches of water). For pressure differences less than 0.025 cm of water, the flow was assumed to be zero for both sides. As a result, there were some limitations.

EXPERIMENT

Twelve trained assessors used the instrumented Mask Scentometer in a laboratory fume hood where a series of odor concentrations using the standard odorant, n-butanol in liquid form, were presented in a pan (Figure 6). The n-butanol concentrations were used to provide a target odor for the panelists to assess during the expreiment. Ten presentations of n-butanol at corresponding intensity levels of 0, 1, 2, 3 and 4, in duplicate, were presented randomly to each assessor. A static-scale method based on ASTM Standard E 544-99, *Standard Practices for Referencing Suprathreshold Odor Intensity* was used. The 0-4 static scale was based on the geometric progression of 3 starting with 25 ppm n-butanol in air for 1; 75 ppm n-butanol in air for 2; 225 ppm n-butanol in air for 3; and 675 ppm n-butanol in air for 4. These are the expected headspace concentrations as defined by the standard.

Figure 6. Experimental setup for flow measurement through the Mask Scentometer. Assessors were presented with n-butanol solutions to assess the odor dilution to threshold while airflow was measured.

The assessors were asked to evaluate the level of odor in the hood using the Mask Scentometer. Flow rates and mask settings were monitored with a Data AcQuisition System (DAQ). Hood flow was constant during the experiment and assessors were asked to perform their assessments so that all assessors were about the same distance from the pan (where the n-butanol concentrations were varied). A metal partition was used in the ventilation hood to allow the n-butanol concentration to build up and remain stable so the assessors could detect the pan concentration. Since the measured dilution ratios were dynamically changing with flow during respiration, flow-weighted dilution ratios were calculated for each setting over the entire dataset of assessments for each assessor. The following equations outline how flow-weighted average dilution ratios were calculated:

For each instantaneous measurement (i),

$$DR_i = \frac{Q_{cleanair}}{Q_{ambientair}}$$

$$Q_i = Q_{cleanair} + Q_{ambientair}$$

For all flows, where $Q_{cleanair}$ is the clean air flow rate, $Q_{ambientair}$ is the ambient air flow rate, Q_i is the instantaneous total flow rate, and DR_i is the instantaneous dilution ratio when at mask setting A, B, C, D, or E, and for the total number (N) of (*i*)'s for that setting,

$$DR_{setting} = \frac{\sum_{i=1}^{N} DR_i Q_i}{\sum_{i=1}^{N} Q_i}$$

Thus, the results in Table 2 below are the weighted averages composited from every instance (instantaneous measurement, *i*) that the assessor used a particular setting over the entire experiment. The total number of instantaneous measurements varied and was dependent on the number of breaths used by the assessor during each device setting and the number of times that setting was used during the ten odor level sessions. More data was collected for the first setting, A than for E because assessors always started with the A setting first and did not go through all of the settings (A-E) if odor was detected before reaching setting E. At least four of the ten assessments required the person to use all of the settings (A through E). Each assessment resulted in three to six breaths, and each breath resulted in 30 to 40 instantaneous flow and D/T measurements. So for setting E, which had the least number of measurements, a minimum of 500 instantaneous measurements represented the D/T for that assessor (sniffer) while setting A had about 2,500 instantaneous measurements.

RESULTS AND DISCUSSION

For each assessor and their device setting readings, the mean dilution ratio for each assessor and setting was weighted for flow and averaged to compensate for the varying nature of breathing (sniffing) during the assessment. The results of the experiment are shown in Table 2 and Table 3.

Table 2. Experimentally determined Mask Scentometer dilution ratios.

	Mask dial setting (ambient orifice diameter)				
Assessor	A (1/16")	B (1/8")	C (3/16")	D (1/4")	E (1/2")
a	20.65	5.04	2.14	1.21	0.36
b	16.54	4.13	1.80	0.99	0.36
c	17.35	4.21	1.72	0.96	0.34
d	16.47	4.05	1.74	0.98	0.35
e	15.07	3.72	1.60	0.87	0.44
f	21.53	5.32	2.33	1.27	0.33
g	17.80	4.48	1.92	1.09	0.37
h	16.29	4.04	1.75	0.99	0.38
i	16.64	4.21	1.81	0.99	0.35
j	17.39	4.60	2.07	1.21	0.43
k	17.27	4.29	1.85	1.01	0.34
l	17.90	4.35	1.86	1.04	0.33
Average	17.58	4.37	1.88	1.05	0.36
Standard Deviation	1.82	0.44	0.20	0.12	0.04
Max.	21.53	5.32	2.33	1.27	0.44
Min	15.07	3.72	1.60	0.87	0.33

In Table 3, assumed dilution ratios, ambient orifice diameters, actual flow-weighted average dilution ratios, standard deviations, and recommended D/T settings are shown.

Table 3. Dilution ratio results and D/T settings for each Mask Scentometer dial setting.

Dial setting	Dial orifice diameter	Assumed D/T setting	Measured* (actual) dilution ratio		Recommended D/T setting**
			Average	Standard deviation	
OFF	None	0	NA	NA	0
A	1/16”	170	17.58	1.82	18
B	1/8”	31	4.36	0.44	4.5
C	3/16”	15	1.88	0.20	2
D	1/4”	7	1.05	0.12	1
E	1/2”	2	0.36	0.04	0.35

** Measured dilution ratios are flow-weighted averages.*
*** Recommended setting is not significantly different (α=0.05) from the mean measured dilution ratios of the twelve assessors.*

During data collection it was noticed that some individuals inhaled deep and long during their assessment while others inhaled using shorter (but still deep) breaths, possibly contributing to variation in flow rates. It was also observed during the experiment that female assessors tended to sniff deeper and shorter than male assessors. A long and deep inhalation resulted in a slightly lower dilution ratio while a short but deep inhalation resulted in a slightly higher dilution ratio. A t-test was used to determine if the dilution results for each dial setting (A-E) were the same between male (n=6) and female (n=6) assessors. The differences were not statistically significant (p values were 0.10, 0.31, 0.39, 0.83, and 0.13 for settings A, B, C, D and E, respectively, α=0.05), so the variations in the results were likely not biased by gender. Additionally, a check to see if there was a significant difference among assessors was used (The null hypothesis was that the flow-weighted D/T’s for individual assessors were not different from one another). The final device D/T’s were selected by testing the hypothesis, of whether the six position means were significantly different from the nearest half D/T. This was done to make the final device settings more user friendly. For example, for setting C, a D/T of 2 is more user-friendly than 1.8. The twelve assessors' results for each setting were not significantly different (α=0.05) from the values of 18, 4.5, 2, 1, and 0.35 using a general linear model of variance analysis (ANOVA) to test these proposed device settings.

MASK SCENTOMETER RESULT REPEATABILITY

A shortcoming of the results presented is that only one set of Mask Scentometer cartridges were used (each sniffer used individual masks that were fitted and assigned to them). To ensure that the device-related results were repeatable and consistent between Mask Scentometers, a series of tests was conducted on eight clean air delivery cartridges. For this test, the airflow through each cartridge was tested at 15, 25, 35 and 45 L/min. The data were then analyzed using ANOVA with the mean flow rate from a set of flows collected over the operating range to test for variation between cartridges. Duncan's Multiple Range Test (MRT), Least Significant Difference (t-test LSD), Bonferroni, and Scheffe's multiple comparison tests were used to detect differences. The analyses were performed using SAS (2008). Only LSD results are shown in Table 4; however, the results were the same for Scheffe's, Duncan's MRT, and Bonferroni tests. From this analysis, it was concluded that the replication of the device is reliable. The evaluation was not conducted for the ambient air delivery cartridge due to the expense associated with sacrificing ambient air cartridges for this test.

Table 4. Measured flow rates through eight clean air delivery cartridges.

Actual flow (L/min)	Cartridge							
	1	2	3	4	5	6	7	8
15	15.5	16.2	15.5	16.4	16.4	15.4	16.5	15.5
25	25.3	26.4	25.4	27.0	26.9	21.3	27.3	25.1
35	33.0	34.3	32.9	34.9	34.4	32.0	34.7	34.6
45	41.9	43.8	42.4	45.0	44.7	42.2	45.7	42.8
Mean*	28.93[a]	31.18[a]	29.05[a]	30.83[a]	30.60[a]	27.73[a]	31.05[a]	29.50[a]

**Within the row, letters with identical superscripts means are not significantly different at alpha level of 0.05*

General Care, Preparation and Operation of Mask Scentometer

Based on extensive experience using the Mask Scentometer at the Universities of Idaho and Nebraska, the following general guidelines for its use are recommended. It should be baked at 120 degrees Fahrenheit for five hours after every eight hours of use. This minimizes odors from previous respiration, volatile compounds from the butyl rubber, and glue that has not dried from fabrication. The mask should be cleaned with an alcohol wipe after every use and kept in a plastic bag to prevent contamination. On occasion the integral check valves, comprised of thin rubber disks - two located at the inlets and one located at the exhaust port of the mask - may stick and prevent airflow. They should be removed and cleaned periodically. The valves can be torn easily, so must be handled with care, gripping closer to the center where they are attached when removing.

Users must not consume coffee, spicy foods, or caffeinated drinks on days that they are using this device. To use this device accurately, users cannot be smokers. The mask should be donned before reaching the area where measurements are to be made and should be in the “OFF” position. The user should cycle the dial through all five positions, taking a normal sniff at all settings while noting any background odors. This will help the user recognize the difference between an ambient odor and background odor from the mask.

The device should be worn for at least two minutes before making an assessment, to allow the user's olfactometry nerves to become acclimated to the background odor of the mask, and to minimize any error from trying to distinguish between the target odor and the background odor of the device.

Users should have a respirator fit test before using the device and be capable of wearing the device for an extended period of time and be comfortable breathing normally with a respirator. The user should pressure test the mask to ensure it is fitting properly, and that air cannot leak past the seals. This can be accomplished by placing the palm of one's hand on the clean side cartridge with the mask in the "OFF" position. The wearer should not be able to inhale (a vacuum will be created). This indicates a good seal with the mask and face. If a leak is detected, tighten the straps on the mask or tighten the screw on the dial. If the leak cannot be stopped, then a different size mask may be needed. North™ Model 7700-30, P/N 7700-1114 masks come in small, medium, and large sizes.

Wearing the mask can make the user very sensitive to other odors, especially at low concentrations of the target odor. The user must be able to "recognize" the odor that is being assessed to correctly assess the D/T. When at the location where a measurement is to be taken, the user should briefly remove the mask, and take a sniff to identify and locate the odor that is to be assessed. Failure to do this could result in the measurement of another ambient odor not associated with the source being investigated.

Users need to train themselves to differentiate between breathing through the Mask Scentometer and "sniffing." Sniffing is described as the action of drawing in air so as to pass it over the olfactory nerves in the sinus cavity. Simply dawning the Mask Scentometer without regard for breathing technique may not produce a reliable result. This is an important distinction, as the device does not allow the user to physically separate normal respiration from how one breathes when searching for a scent. The user should differentiate between background odors and the target odor, other non-target odors may be noticed in the mask during an assessment, but users should train themselves to be consistent at detecting the target odor. If the target odor has not been located at the onset, then move to a location where the odor is detected with the mask briefly removed.

Once the receptor location has been established, a stopwatch should be started to coordinate the measurements. The user should then move the dial to the first position A, sniff two to four times through the nose taking care to ensure that air is being passed over the olfactometry nerves. If the target odor is not detected, then move the dial to position B and repeat the process. Continue through the positions until odor is detected (record setting, i.e. "A") or not detected (record as non-detect). After 30 seconds has passed, begin the process again. When a mask position has been reached and an odor can be detected but not recognized, then move on to the next position until it can be determined that the odor is the target odor. When the odor can be recognized, record this mask setting and return the mask to the "OFF" position. Do not go any further than the first position that the odor was recognized. For example, if an odor can be detected at B, but cannot be recognized as the target odor, then the user should move on to position C, if the odor can be recognized as the target odor, such as the odor from a pig barn, then the user is to record "C" and return the dial to the "OFF" position. The user would not assess positions D and E. This procedure will minimize olfactory fatigue.

Odor measurements should be taken for a period of 15 minutes at 30-second intervals. An experienced user can cycle through all five settings in a period of 30 seconds. Since some measurements reflect odor peaks or puffs from the source, the 15-minute period allows for a short averaging time and reduces errors associated with a single point in time measurement. This method results in 30 individual measurements that can be averaged over the time period.

SUMMARY AND CONCLUSIONS

The Mask Scentometer, similar in concept to the Barneby and Sutcliffe Box Scentometer, is equipped with a series of different-sized orifices that facilitate the dilution and mixing of ambient air with air that has been cleaned with an activated carbon filter to achieve set dilution ratios. The device is used in field olfactometry to determine dilutions to threshold, with its main advantage over other ambient odor measurement techniques being that it is worn tightly on the face, which minimizes the likelihood of sensory fatigue by the user. The assumption prior to completing this work and the hypothesis tested was that the dilution-to-threshold settings were the same as with the Box Scentometer: 170, 31, 15, 7, and 2. This work measured actual flow rates through both the ambient and clean air delivery cartridges during use and found the dilution ratios to be 18, 4.5, 2, 1 and 0.35, respectively, at each of the five D/T settings. The dilution ratios were not the same for the Mask Scentometer and the Box Scentometer. Thus, these measured dilution ratios should be used in any work where the Mask Scentometer is being used to assess ambient odors. This work also brings into question the actual dilutions to threshold of the Barneby and Sutcliffe Box Scentometer during operation since the original work by Huey et al. (1959) did not measure the dilution ratios during operation.

REFERENCES

ASTM. 1999. E544-99. Standard Practices for Referencing Suprathreshold Odor Intensity. *Annual Book of ASTM Standards.* Philadelphia, PA. American Society of Testing and Materials.

Barnebey and Sutcliffe Corporation. 1995. Technical Bulletin No. 105, Odor Control with Activated Carbon.

Barnebey and Sutcliffe Corporation. 1959. Scentometer: An Instrument for Field Odor Measurement, U.S. Public Health Designation - Scentometer Model 1959-A, Barnebey & Sutcliffe Designation – Model SCC (Modified to extend range).

Committee for European Normalization (CEN). 2003. CEN/TC264/WG2/N222/e Air Quality – Determination of Odour Concentration by Dynamic Olfactometry, Brussels, Belgium. European Committee for Standardization.

Henry, C. G. 2004. Instruction Manual for Mask Scentometer. University of Nebraska-Lincoln. Lincoln, Nebraska.

Huey, N. A, L. C. Broering, and C. W. Gruber. 1959. Odor Measurement Techniques II, December, Second Year Final Report. US Public Health Service, Community Air Pollution Demonstration, Project Grant A-59-541. Bureau of Air Pollution Control and Heating Inspection, Department of Safety, Cincinnati, OH.

Huey, N. A., Broedering, L. C., Jutze, G. A., and Gruber, C. W. 1960. Objective odor pollution control investigations. Presented at the 83rd Annual Meeting of APCP. May 22-26.

Munson, B. R., D. R Young, T. H. Okiishi. 1994. *Fundamentals of Fluid Mechanics.* Second Edition. John Wiley and Sons, New York, NY. ISBN 0-471-57958-0.

Omega Engineering. No date. Low-Pressure Transducer for Vacuum/Differential/Gage for measurement of clean gasses. PX-Series datasheet. Omega Engineering, Stamford, CT.

Sheffield, R. E. and P. Ndegwa. January 2008. Sampling Agricultural Odors. PNW 595. University of ID.

Statistical Analysis Software. 2008. Version 9.1. SAS Institute Inc. Cary, NC.

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

Verein Deutscher Ingenieure (VDI). 1993. VDI 3940, Determination of Odorants in Ambient Air by Field Inspections, Beuth Verlag, Düsseldorf, Germany.

Paper No. 2

COMPARISON OF AMBIENT ODOR ASSESSMENT TECHNIQUES IN A CONTROLLED ENVIRONMENT

C. G. Henry, D. D. Schulte, S. J. Hoff, L. D. Jacobson, and A. M. Parkhurst

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 $(D/T)_G$.

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.

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). Intensity ratings based on an Odor Intensity Reference Scale (OIRS), may be used as predictors of odor concentration.

The Box Scentometer was first developed as a result of grants provided by the US Public Health Service in the 1950's (Huey et al., 1959 and 1960). The original Box Scentometer was manufactured and sold by the Barnebey-Chaney Company, which later became the Barnebey Sutcliffe Corporation (Barnebey and Sutcliffe Corporation, 1959; Huey et al., 1960; Barnebey and Sutcliffe Corporation, 1995), now owned by the Calgon Carbon Corporation. The Box Scentometer works by diluting incoming ambient air with cleaned air (air that has been passed through an activated carbon filter) to obtain discrete dilution ratios. The disadvantages of the instrument include a lack of control of inhalation rates by assessors, the discomfort of glass inhalation tubes, odor fatigue caused by poor nasal sealing to the glass ports, and the inability to prevent physical fatigue during measurements and olfactory fatigue between measurements.

The Mask Scentometer (Henry, 2004) is functionally similar to a Box Scentometer. It is comprised of a ¼-face respirator mask with two cartridges, one on each side. One cartridge has a charcoal filter with a plenum that has two ¼-inch (0.63 cm) holes in it. As ambient air is drawn through this cartridge, the air is cleaned before entering the mask chamber. The other cartridge has a dial mechanism with holes that allow an assessor to adjust the flow rate of ambient air into the mask chamber where dilution with clean air occurs. Although Sheffield et al., (2004) and Henry (2004) used the same orifice diameters for ambient and clean air that had been established by the Barnebey-Chaney Box Scentometer, the D/T settings had never been verified during the development of the Mask Scentometer. In the study reported in Paper No. 1, the actual dilution ratios of the

Mask Scentometer were measured during use and found to be 0.35, 1, 2, 4.5 and 18. These D/T settings were subsequently used in the analysis of the results from the Mask Scentometer for this research. The advantage of the Mask Scentometer is that it minimizes odor fatigue, allows for one-handed operation, and can be easily operated over long periods of time without fatigue. In addition, the Mask Scentometer overcomes many of the disadvantages of the Barnebey and Sutcliffe Box Scentometer, specifically the poor sealing and discomfort of the glass ports and risk of odor fatigue. Assessors 'sniff' through the series of five mask settings, advancing the dial every 5-8 seconds, which allows a measurement to be taken every 30 seconds (shortest time to rotate through all five settings).

The Nasal Ranger® is another relatively new device. Developed by St. Croix Sensory, Inc., of St. Croix, Minnesota, it is designed to combine the portability and relatively low cost ($1,500) of a scentometer with the dilution control of more expensive laboratory olfactometers (McGinley and McGinley, 2003). The Nasal Ranger® dilutes air in a similar fashion to the other scentometers, but incorporates a flow meter and feedback indicator to help ensure the user is being presented with the indicated dilution. An orifice selector dial on the device contains six ambient air inlets for D/T settings of 2, 4, 7, 15, 30 and 60, although customized D/T settings are available. Between each selectable D/T setting there are blank positions, in which all incoming air is passed through the two replaceable carbon filter cartridges (presenting the user with only odor-free air). Flow control is maintained by a static pressure sensor and LED's, which provide feedback to the person using the device to inform them if they should increase or decrease their air intake rate. The static pressure target range is equivalent to 16-20 L/min sniffing rate, enabling the unit to be attached to a pump for direct calibration of the dilution ratios. The sniffing mask on the Nasal Ranger® is equipped with an inlet and outlet, both of which have check valves, which allows an assessor to keep the unit over the nose during an entire odor assessment (McGinley and McGinley, 2003). However, unless the instrument is held to the nose during the entire assessment, which results in physical fatigue, it does not protect against olfactory fatigue.

When assessors rate odor intensity using an odor intensity reference scale (OIRS), they use a reference scale to assess target odors (McGinley, 2002). The ASTM Standard E 544-99, *Standard Practices for Referencing Suprathreshold Odor Intensity*, describes dynamic and static scale methods of assessing target odors using a standard odorant and the Dutch Standard (VDI 1993), *Determination of Odorants in Ambient Air by Field Inspections* describes a system that is not based on a reference odorant but uses descriptive levels. The specified standard odorant in the ASTM standard is 1-butanol (*n*-butanol). Where an olfactometer is available to present diluted n-butanol concentrations, the dynamic method is relevant; however, in the absence of an olfactometer, the static scale can be used. A simple numerical scale is used, for example, 1-12, 1-10, or 1-5 that correspond to the intensity levels. The standard specifies the static scale method as dilutions in air ranging from 10 to 10,000 ppm n-butanol in air with a geometric progression of 2. However, two other scales are commonly used, one is a 10-point static scale starting at 12 ppm with a geometric progression of 2, and the other is a 5-point static scale starting at 25 ppm with a geometric progression of 3. The latter method was used to determine odor intensity in this experiment because it is easy for assessors to commit the scale to memory for use in the field.

When measuring odor intensity in the field, odor assessors use a charcoal-filtered respirator mask to maintain osmotic sensitivity between readings and expose their noses for only a few seconds to rate the odor intensity, replacing the mask and repeating the procedure every 10-15 seconds, typically for a period of 10-15 minutes.

The first method of laboratory-based olfactometry used a syringe dilution technique called *Standard Test Method for Measurement of Odor in Atmospheres*, published as ASTM D1391. Later in 1979, came ASTM E679-79, *Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits*, which was updated in 1991 and again in 1999. The latter method uses an olfactometer to dynamically dilute odorous air with clean zero-odor air and present it to a panelist, along with two blind presentations. The panelist sniffs three presentations (two blanks and one sample) and must select the one that is different from the other two. Their selection can be designated as a guess, detection or recognition, but they are forced to choose one declaration. This approach is called a triangular forced-choice method. This method presents the diluted odor samples to each panelist in an ascending concentration series, starting with the lowest concentration level (highest dilution), and steps up the concentrations in a series (e.g. two or three times higher) until the panelist reports detection.

Dynamic triangular forced-choice olfactometry (DTFCO) is considered the gold standard of odor measurement. DTFCO involves collecting odor samples in special bags or containers (common bag materials are Tedlar™ and Melenix) which are problematic in and of themselves because of bag material background odors (Koziel et al., 2004; Keener et al., 2002; and Qu and Feddes, 2006). DTFCO also has problems with lower detection limits (LDL) that are commonly much higher than odor levels experienced in the ambient atmosphere (1-30 D/T). Parker et al. (2003) reported most olfactometer LDL to be around 50 D/T. Parker did report that the LDL could be reduced to around 15 if baked Tedlar™ bags were used for feedlot odors to reduce the background odor from new Tedlar™ bags which have a manufacture background odor that is estimated between 20 to 60 OU_E/m^3 (European Odor Unit). Qu and Feddes (2006) found that measured concentrations of SKC 10L Tedlar™ bags ranged between 8 to 71 OU_E/m^3 and recommended that odor levels in samples should be equal to or higher than 608 OU_E/m^3 for new Tedlar™ bags and 544 OU_E/m^3 for self-made Tedlar™ bags. Because of the odor interactions with sample bags and high LDL of olfactometers, some believe DTFCO to be inappropriate for ambient odor assessment. The European Standard for olfactometry does not recommend DTFCO as an appropriate method for ambient odor assessment (CEN, 2003).

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. Pearson correlation coefficients of -0.01 to 0.5 were found and were rarely significant between the field and laboratory odor intensity methods suggesting a sample storage issue with Tedlar bags. Moderate correlations were found between field ammonia and hydrogen sulfide sampling methods and the olfactometry and odor intensity methods used. They noted that assessing different types of odor sources, which included dairy, beef, food processing, and wastewater was a primary factor in the lack of consistent performance of the methods.

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 measured hydrogen sulfide and ammonia using a Jerome meter and Dragger diffusion tubes, respectively. They found a moderate correlation between D/T and H_2S/Total Reduced Sulphur (TRS) which appeared to increase slightly with receptor distance from the source. They found the Pearson correlation coefficients between the Nasal Ranger® D/T and H_2S/TRS (n=520 for each odor source) to be 0.523, 0.625 and 0.664 for 0 meters, 50 meters, and 200 meters, respectively. Regression of all of the data yielded an R^2 of 0.21 (y = 0.0025x - 0.042) between H_2S/TRS and D/T.

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. The relationships they found between the field olfactometer D/T and the room hydrogen sulfide concentrations (measured with a Jerome Meter) are shown in Table 1.

Table 1. Relationship between field olfactometer D/T and H_2S level (McGinley and McGinley, 2003)

2 D/T	4 D/T	7 D/T	15 D/T	30 D/T	60 D/T
2-4 ppm H_2S	4-5 ppm H_2S	4-11 ppm H_2S	11-17 ppm H_2S	17-28 ppm H_2S	28-40+ ppm H_2S

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:

- <u>Dynamic triangular forced-choice olfactometry (DTFCO)</u> DTFCO was used to analyze air samples collected 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 Paper No. 1.
- 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. An 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. Figure 1 shows study participants assessing odor in the test room on day 2.

Figure 1. Odor assessors record their observations within the ISU Air Dispersion Laboratory test room.

The following relationship (Sheffield et al., 2004) was used to obtain geometric average dilutions to threshold $(D/T)_G$ for the field olfactometers (Mask Scentometer and Nasal Ranger®):

$$(D/T)_{G,n} = 10^{\frac{\log D/T_n + \log D/T_{(n+1)}}{2}}$$

Where D/T_n is the current setting step number (unit D/T) and DT_{n+1} is the next setting step number. For example, for the Nasal Ranger® setting #3 (unit D/T = 4) and setting #4 (unit D/T = 7), the $(D/T)_{G,3}$ = 5.3. This was done to normalize the peaks and keep extremely high or low values from skewing the results. One of the issues with data from field olfactometers is how to deal with non-detects, where no odor level is measured. It is not possible to take the geometric average of results that include zeros. Taking the geometric average of the device settings is preferred, and reasonable since the geometric average of settings 3 and 4 for example is 5.3 which is between 4 and 7, so when the device reports a D/T of 4, it is more likely that it is somewhere between a 4 and a 7. The geometric D/T settings are shown in Table 2. For the non-detect level, a value was assumed that was about two-thirds between zero and the first D/T setting on the device (0.2 for the Mask Scentometer and 1.4 for the Nasal Ranger®). The $(D/T)_G$ have the effect of increasing the overall result of an assessment compared to taking an average of the D/T settings, and a drawback that when no odor is present, an odor level is reported (although small, i.e. not less than 0.35 or 1.4). In this situation, it is assumed that an odor was present during the assessments, so the effect of the non-detects is small on the overall results. Also, no attempt was made to adjust the highest settings of the devices (18 DT for the Mask Scentometer and 60 for the Nasal Ranger® since they are the limits of the instruments (their $(D/T)_G$ could be between 18/60 and infinity).

Table 2. 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 for each method were compared to panel D/T obtained via DTFCO. To analyze the data, a lack-of-fit test was used first to determine if the data were linear. Next the data were evaluated for interactions between assessors. If there was significant interaction, suspect assessors were removed until there was no interaction. Next, simple linear regression (with zero intercept) was performed between methods to determine the slope of the relationship and correlation.

Since a simple linear model was desired between methods and bias between assessors was a concern, the lack-of-fit test was used to screen the dataset for bias. This test evaluates the relationship between systematic error (bias) and pure error (random in nature). SAS (Statistical Analysis Software, 2008) version 9.1 was used for the lack-of-fit test. The procedure is to calculate the mean response for each method (y) at each (x) odor concentration (DTFCO); in this case, the session mean for an assessor was compared to the D/T as determined by DTFCO sample analysis for that same session. Then, the difference was calculated between each session DTFCO D/T and each of the associated raw responses (each assessor's session mean was a response), and the sum of the squares of each difference was calculated. The lack-of-fit test determines if the residual means deviate from zero enough to suggest bias and, consequently, if the linear model is inadequate.

Next, consistency between days - sessions across assessors - was examined. Tukey's test for non-additivity (interaction) was performed in SAS using code developed by Parkhurst (2008). In this analysis, since the experiment was conducted over the course of two days (days were blocked) across 13 sessions (treatments), the interaction was tested for any effect due to days or sessions for each method (experimental unit).

Once interactions were removed, the 13 session means for each method were recomputed using the averages of all remaining assessors. Simple linear regression analysis (with zero intercept) was performed on the session means to determine relationships between laboratory-based olfactometry (DTFCO) and the Nasal Ranger®, Mask Scentometer, field intensity ratings, and intensity-predicted D/T using the statistical package R (Development Core Team, 2008).

The data was then analyzed using the general linear model analysis of variance (ANOVA) using the session means for each assessment method to test the variation between methods. The least-significant-difference multiple comparison test (t-test LSD) was used to detect D/T differences between odor assessment methods (DTFCO, intensity-predicted ratings, Nasal Ranger®, and the Mask Scentometer using SAS (2008).

RESULTS AND DISCUSSION

Developing statistical relationships between ambient odor assessment methods was complicated by the fact that a different number of odor assessors used each method. Additionally, for each method a different number of observations were made by each assessor during the 10-minute sessions: the intensity assessors took readings every 15 seconds yielding 40 observations per session; Mask Scentometer users took readings every 30 seconds yielding 20 observations; Nasal Ranger® users took a reading twice during each session for two observations; and one air sample was taken for laboratory olfactometry (DTFCO). The means for each method and session were calculated and the regression procedure (Proc Reg) was used. 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.

As can be seen in Table 3, none of the methods produced measurements that were biased on a statistically significant basis ($P > 0.05$) with respect to DTFCO data; that is, all of the methods shared a linear relationship with DTFCO. It should be noted, that the Mask Scentometer is just nearly linear, with $P=0.0549$, while the other methods have values that clearly indicate they are linear. Sessions 4 and 6 resulted in very high Mask Scentometer assessments of 5.9 and 7.1 the highest two sessions for the device, yet these sessions yielded some of the lowest DTFCO assessments of 99.7 and 63.3 D/T, which skewed the results.

Table 3. Results of lack-of-fit test, showing lack of bias between assessment methods and DTFCO.

Method	P value	n
Mask Scentometer	0.0549	58
Nasal Ranger®	0.7585	53
Intensity-predicted D/T	0.9973	58
Intensity (0-5)	0.6884	58

Tukey's test revealed some interactions between days and assessors. Thus, assessor data were removed from the sessions where there were significant interactions, until the overall interactions were no longer significant. This procedure was performed separately for D/T (Table 4) and geometric D/T or $(D/T)_G$ (Table 5) for the Nasal Ranger® and the Mask Scentometer data. Average-intensity-predicted D/T are directly based on intensity ratings so the test for interaction and lack-of-fit was only necessary for intensity.

Table 4. Tukey's Test results for D/T interactions between days and assessors.

Day	Mask Scentometer	Nasal Ranger®	Field intensity	Intensity-predicted D/T
1 (Sessions 1-6)	P = 0.0063 for 3 assessors P = 0.165 with one assessor removed	P = 0.0647 for 3 assessors None removed	P = 0.2119 for 5 assessors, None removed	P = 0.0847 for 5 assessors None removed)
2 (Sessions 7-13)	P = 0.1259 for 6 assessors None removed	P = 0.0012 for 5 assessors P = 0.049 with one assessor removed	P = 0.2287 for 4 assessors, None removed	P = 0.0080 for 4 assessors P = 0.0576 with one assessor removed

Table 5. Tukey's Test results for $(D/T)_G$ interactions between days and assessors.

Day	Mask Scentometer	Nasal Ranger®
1 (Sessions 1-6)	P = 0.9592 for 3 assessors None removed	P = 0.3694 for 3 assessors None removed
2 (Sessions 7-13)	P = 0.5417 for 5 assessors None removed	P = 0.0179 for 5 assessors P = 0.2494 with one 1removed

The Pearson's product-moment correlation coefficient and Spearman's Rank Correlation (ρ) are used to indicate the strength and direction of the linear relationship between two random variables. The Pearson's product-moment correlation coefficient is a parametric test and is defined as the sum of the products of the standard scores (number of standard deviations above and below) of the two measures divided by the degrees of freedom. The Spearman's Rank Correlation (ρ) is a non-parametric test and is a special case of the Pearson Product-moment coefficient, in which the two sets of data are ranked before calculating the coefficient. The raw scores are converted to ranks and the differences between the ranks of each observation on the two variables are calculated.

The correlation coefficients lie between -1 and 1, with 1 indicating a strong linear relationship (-1 indicates a strong inverse relationship, i.e. negative slope) and a 0 indicating no linear relationship. Both statistics were calculated, however, non-parametric tests are considered to be more robust because they do not rely on the assumption that the data comes from a probability distribution (a normal distribution of odor levels was measured by the methods, i.e. range of possible values and the probability that the measurement was in that range). It is unknown whether this assumption was met in this experiment. Also, since the intensity rating can be considered an ordinate scale, the non-parametric test (Spearman's ρ) is more applicable. However, both methods lead to the same conclusions for this analysis.

Table 6. 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.*

From the results shown in Table 6, 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 $(D/T)_G$).

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 (x_i) 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 7, 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 (see example shown in Appendix). 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 7. Note that the R_o^2 are the same for each of the linear models. From Table 7 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 $(D/T)_G$. Meanwhile, since the slope of Nasal Ranger® $(D/T)_G$ as a function of intensity-predicted D/T was 0.18, intensity-predicted D/T were nearly five times higher than Nasal Ranger® $(D/T)_G$ and over 25 times higher than Mask Scentometer $(D/T)_G$. Going one step further, with a slope of 0.42 between intensity-predicted D/T and DTFCO D/T, olfactometry laboratory D/T were about 2.5 times higher than intensity-predicted D/T and 50 to 100 times higher than Mask Scentometer $(D/T)_G$ and D/T values, respectively.

Table 7. 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.*

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 method used to calculate R_o^2 is described in detail in the Appendix. The relationship between Intensity-predicted D/T and Average-intensity-predicted D/T is closest to a 1:1 slope at 1.79 (Table 7) 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 7), 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 8) 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 8. 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. While the D/T settings for the Mask Scentometer from Paper No. 1 were used, this analysis had also been done with the previously assumed D/T values of 170, 31, 15, 7, and 2. This scale setting resulted in better agreement (slope of 0.40 and a R_o^2 of 0.67) with the Nasal Ranger® D/T. Slopes and R_o^2's were also better for comparison with intensity ratings, resulting in a slope of 0.16 and R_o^2 of 0.76 for intensity and a slope of 11.1 and a R_o^2 of 0.67 for Intensity-predicted D/T, in all cases an improvement in the results. Therefore, the range limitation 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.

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 9 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.

Table 9. 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.*

An intensity rating between 1.5 and 2 has been discussed as being a threshold at which odor annoyance occurs (Zhu et al., 2000; Jacobsen et al., 2000; Stowell et al., 2007; and Halverson et al., 2007). In this study (see Table 9), 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.

Another interesting observation is the difference between intensity ratings and intensity-based D/T predictions. While the methods use the same raw data, they do not yield the same results. An intensity of two is equivalent to an intensity-predicted D/T of 72, and this study found that intensity-predicted D/T's were about twice that of average-intensity-predicted D/T. Averaging the odor intensity numbers dampens the high's and low's experienced during an assessment, which in this study led to results that were more comparable to other odor assessment methods than did averaging the set of individual predicted D/T values.

SUMMARY AND CONCLUSIONS

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

ASTM. 1979. D1391. Standard Test Method for Measurement of Odor in Atmospheres. *Annual Book of ASTM Standards.* American Society of Testing and Materials. Philadelphia, PA, USA.

ASTM. 1979. E679-79. 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.

ASTM. 1991. E679-91. 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.

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.

ASTM. 1999. E544-99. Standard practices for Referencing Suprathreshold Odor Intensity. *Annual Book of ASTM Standards.* American Society of Testing and Materials. Philadelphia, PA.

Barnebey and Sutcliffe Corporation. 1995. Technical Bulletin No. 105, Odor Control with Activated Carbon.

Barnebey and Sutcliffe Corporation. 1959. Scentometer: An Instrument for Field Odor Measurement, U.S. Public Health Designation-Scentometer Model 1959-A Barnebey & Sutcliffe Designation – Model SCC (Modified to extend range).

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.

Halverson, M. J., D. D. Jones, and D. D. Schulte. 2007. A knowledge-based model to assess odor annoyance. Presented at the International Symposium on Air Quality and Waste Management for Agriculture. 16-19 September 2007. Broomfield, CO. ASABE Publication number: 701P0907cd.

Henry, C. G. 2004. Instruction Manual for Mask Scentometer. University of Nebraska-Lincoln. Lincoln, NE.

Huey, N. A, L. C. Broering, and C. W. Gruber. 1959. Odor Measurement Techniques II, December 1959, Second Year Final Report. US Public Health Service, Community Air Pollution Demonstration, Project Grant A-59-541. Bureau of Air Pollution Control and Heating Inspection, Department of Safety, Cincinnati, OH.

Huey, N. A., Broedering, L. C., Jutze, G. A., and C. W. Gruber. 1960. Objective odor pollution control investigations. Presented at the 83rd Annual Meeting of APCP, May 22-26, 1960.

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.

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 *Proc. Int. Symp. on Gaseous and Odor Emissions from Animal Production Facilities*, 320-330. Horsens, Denmark.

Keener, K. M., J. Zhang, R. W. Bottcher, and R. D. Munilla. 2002. Quantification of odorants from Tedlar™ bags. ASAE Paper No. 024165, Chicago, IL, ASAE.

Koziel, J. A., J. P. Spinhime, J. D. Lloyd, D. B. Parker, D. W. Wright, and R. W. Kuhrt. 2004. Evaluation of sample recovery of maladorous gasses from air sampling bags, SPME, and sampling canisters. ASAE Paper No. 044129. Ottawa, ON, Canada. ASAE, St. Joseph, MI.

McGinley, C. M. 2002. Standardized odor measurement practices for air quality testing. Presented at the Air and Waste Management Association Symposium on Air Quality Measurement Methods and Technology. November 13-15, 2002. San Francisco, CA.

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.

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.

Nicolai, R. E., C. J. Clanton, and H. Guo. 2000. Modeling the relationship between threshold and intensity of swine odors. *Proc. Second International Conference on Air Pollution from Agricultural Operations.* St. Joseph, Mich.: ASAE.

Parker, D. B., M. B. Rhoades, J. Koziel, and J. Spinhirne. 2003. Background odors in Tedlar® bags used for CAFO odor sampling. ASAE Paper No. 034144. Las Vegas, NV. ASAE, St. Joseph, MI.

Parkhurst, A. M. 2009. Personal Communication.

Parkhurst, A. M. 2008. SAS code macro for Tukey's Test for Interaction. Department of Statistics, University of Nebraska, Lincoln, NE.

Stowell, R. R, C. G. Henry, R. K. Koelsch, and D. D Schulte. 2007. Association of odor measures with annoyance: An odor-monitoring field study. Presented at the International Symposium on Air Quality and Waste Management for Agriculture. 16-19 September. Broomfield, CO. ASABE publication number: 701P0907cd.

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

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.

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

Qu, G., and J. Feddes. 2006. Estimation of measurable error caused by background odor in sampling bags. ASAE Paper No. 064141. ASABE, St. Joseph, MI.

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.

Statistical Analysis Software. 2008. Version 9.1. SAS Institute Inc. Cary, NC.

Verein Deutscher Ingenieure (VDI) 1993. VDI 3881: Olfactometry odour threshold determination. Berlin, Germany.

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.

APPENDIX

Additional information not submitted to the journal is included in this appendix. The method to calculate the coefficient of determination (R_o^2) throughout this dissertation is provided followed by Table 10 which reports the session means used in this paper. Figures 1 though 16 show the plotted regression results. Following the plots is an example of how the standard error of the estimate was used to compare model slopes.

Coefficient of Determination Calculation

The coefficient of determination, R^2 is the proportion of variability about Y bar explained by regression. It is the most frequently used goodness-of-fit technique. This no-intercept model, was used to calculate the coefficient:

$$R_o^2 = \frac{\sum_{i=1}^{n} \hat{Y}_i^2}{\sum_{i=1}^{n} Y_i^2}$$

It indicates the proportion of variability around the origin accounted for by regression. Occasionally, R_o^2 is larger than R^2 even though the residual mean square for the intercept model is smaller than that for the no-intercept model. This comes about because R_o^2 is computed using the uncorrected sums of squares (Parkhurst, 2009).

Session Means for Odor Assessment Methods

Table 10. Session means for odor assessment methods.

Session	DTFCO lab D/T	Intensity rating (0-5)	Intensity-predicted D/T	Average intensity-predicted D/T	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$	Nasal Ranger® D/T	Nasal Ranger® $(D/T)_G$
1	32.0	0.6	18.6	16.1	0.8	0.5	15.7	22.6
2	76.3	2.2	136.5	91.6	3.2	3.4	29.0	32.8
3	87.7	1.7	89.3	50.1	2.4	4.6	17.3	24.6
4	99.7	1.0	33.2	24.3	5.9	2.8	8.8	12.4
5	136.3	1.9	148.5	67	1.3	7.4	13.7	19.4
6	63.3	0.7	21.5	17.6	7.1	1.3	4.3	6.1
7	27.7	0.8	7.8	20.7	0.7	2.4	5.9	8.2
8	59.7	2.7	290.4	148.8	2.2	6.9	27.1	35.3
9	144.0	2.0	142.1	70.7	1.8	6.3	31.4	35.0
10	197.0	1.6	66.0	46.4	1.2	4.7	15.5	22.0
11	331.0	0.8	13.6	19.8	0.5	2.2	9.0	12.6
12	208.7	1.9	98.1	67	2.0	6.5	22.0	28.0
13	283.3	1.7	91.4	54.6	1.8	4.8	10.9	15.4

PLOTTED REGRESSION RESULTS

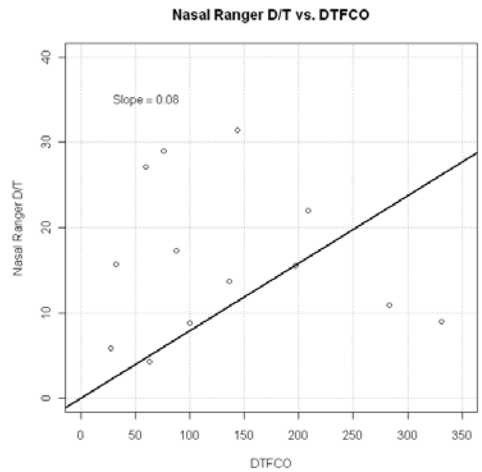

Figure 2. Nasal Ranger® D/T versus DTFCO

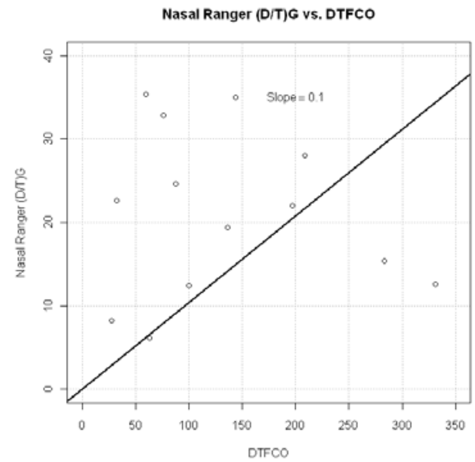

Figure 3. Nasal Ranger® $(D/T)_G$ versus DTFCO

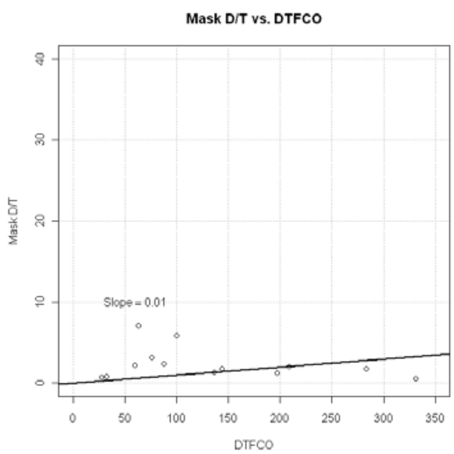

Figure 4. Mask Scentometer D/T versus DTFCO

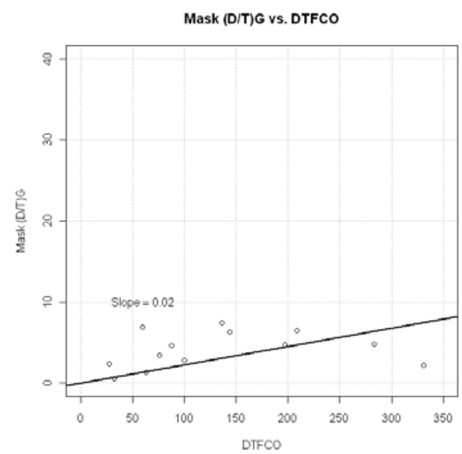

Figure 5. Mask Scentometer $(D/T)_G$ versus DTFCO

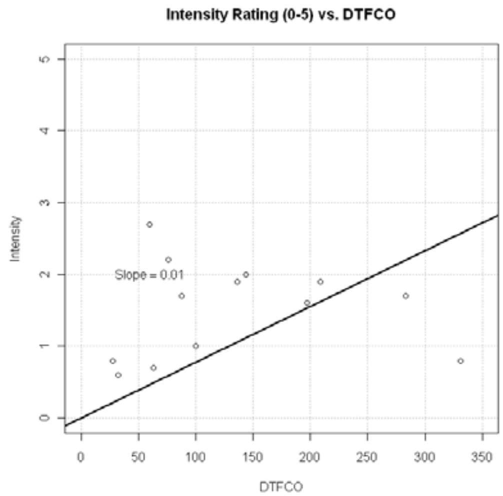

Figure 6. Intensity Rating versus DTFCO

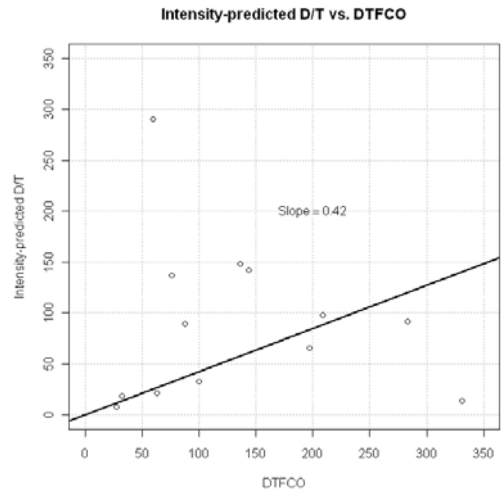

Figure 7. Intensity-predicted D/T versus DTFCO

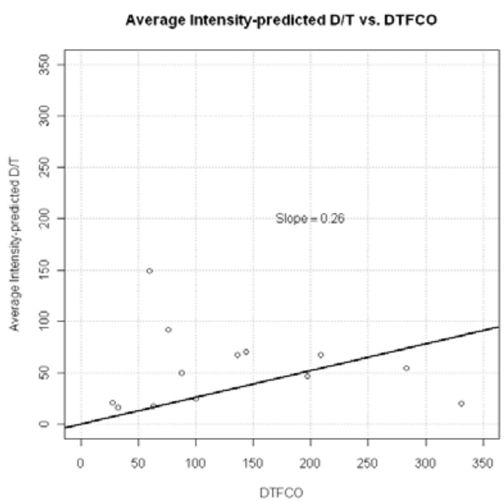

Figure 8. Average intensity-predicted D/T versus DTFCO

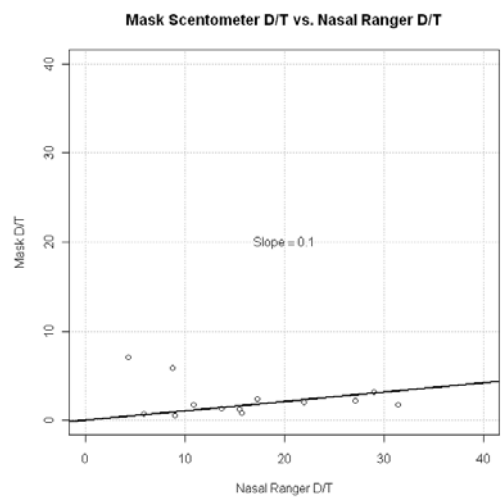

Figure 9. Mask Scentometer D/T versus Nasal Ranger® D/T

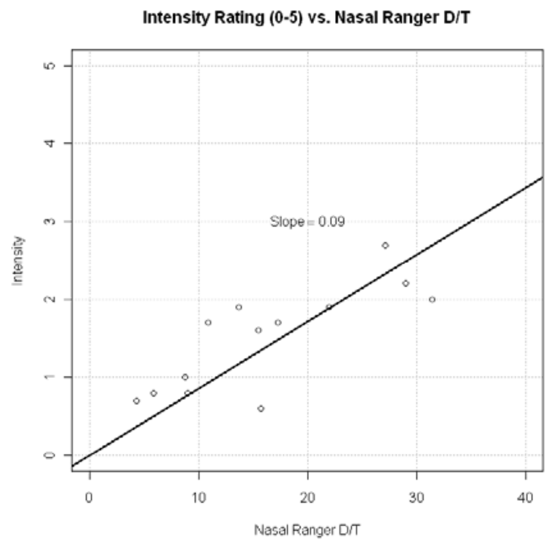

Figure 10. Intensity Rating (0-5) versus Nasal Ranger® D/T

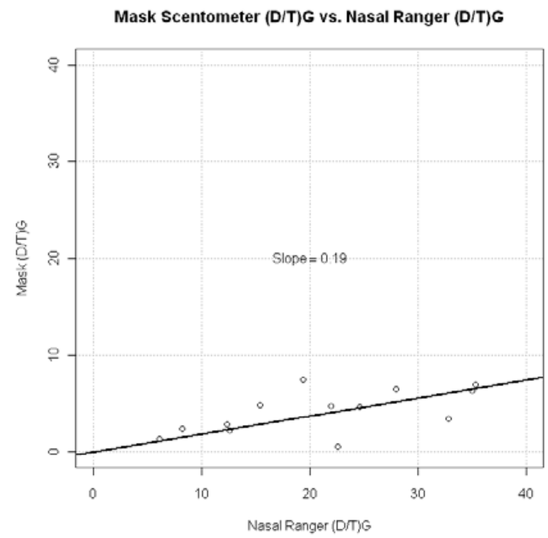

Figure 11. Mask Scentometer $(D/T)_G$ versus Nasal Ranger® $(D/T)_G$

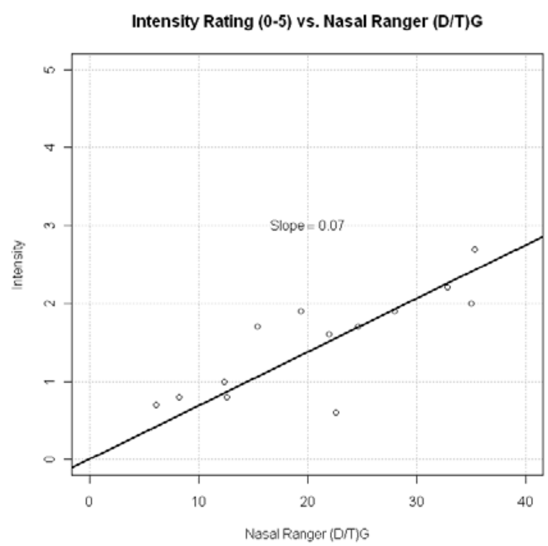

Figure 12. Intensity Rating (0-5) versus Nasal Ranger® $(D/T)_G$

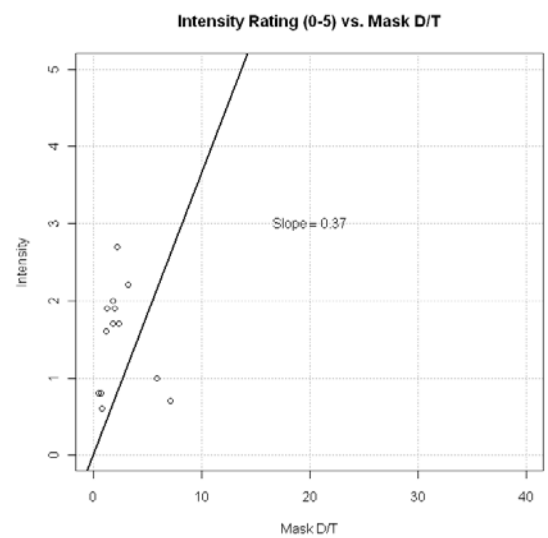

Figure 13. Intensity Rating versus Mask Scentometer D/T

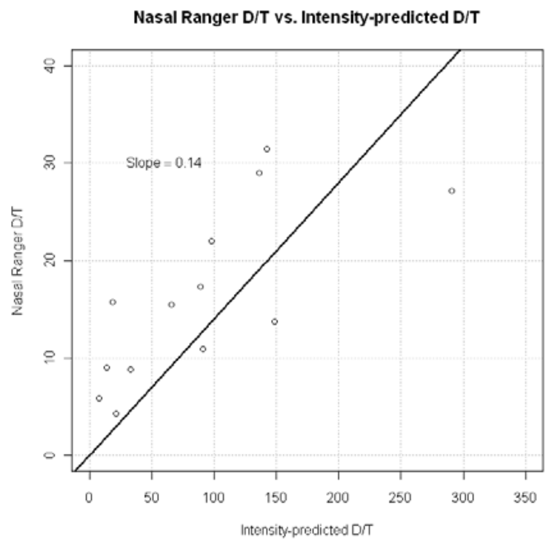

Figure 14. Nasal Ranger® D/T versus Intensity-predicted D/T

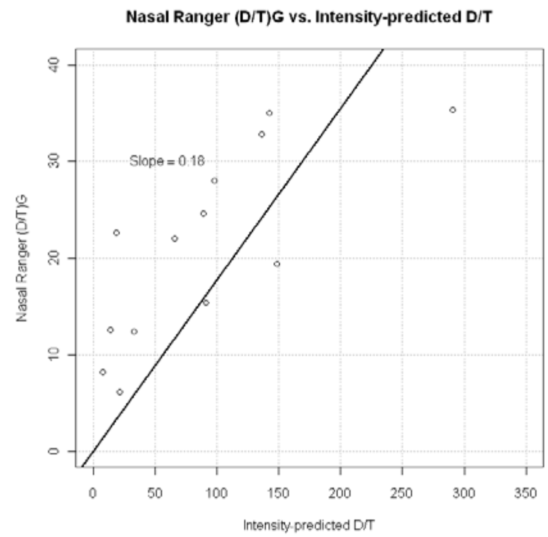

Figure 15. Nasal Ranger® $(D/T)_G$ versus Intensity-predicted D/T

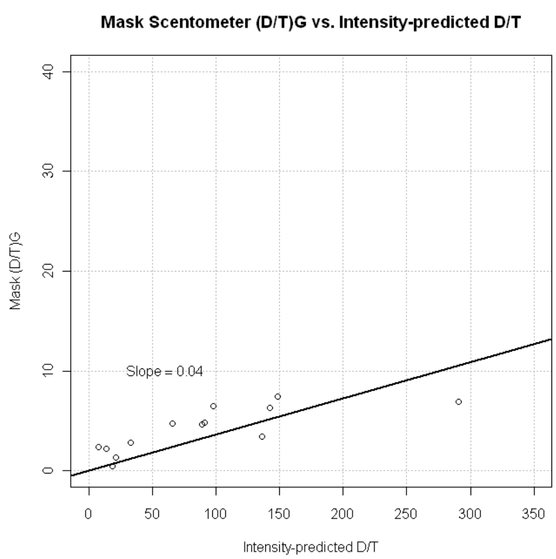

Figure 16. Mask Scentometer $(D/T)_G$ versus Intensity-predicted D/T

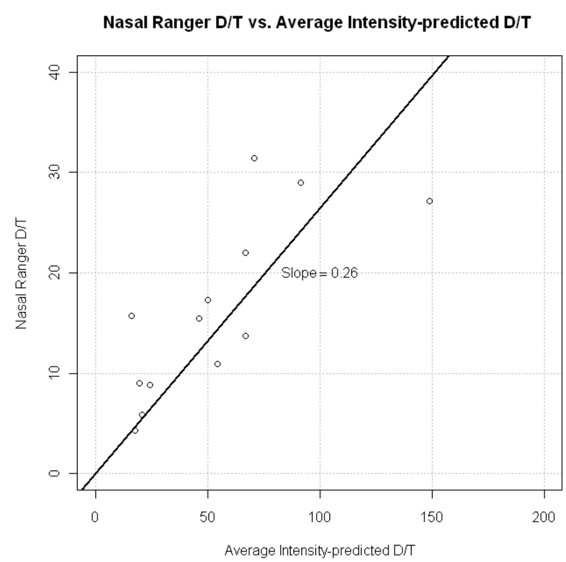

Figure 17. Nasal Ranger® D/T versus Average Intensity-predicted D/T

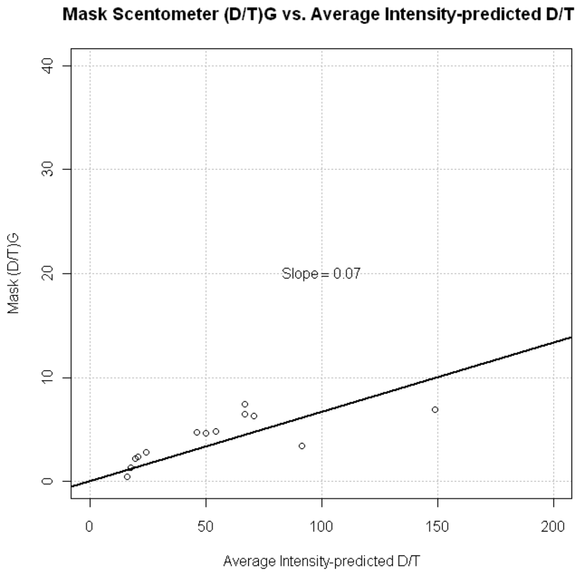

Figure 18. Mask Scentometer $(D/T)_G$ versus Average Intensity-predicted D/T

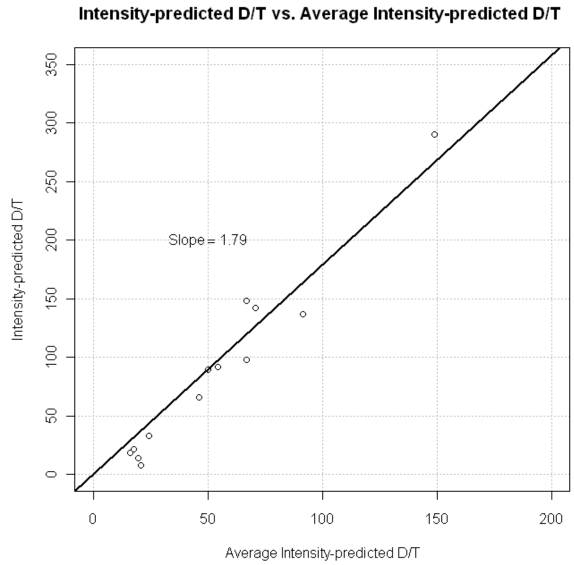

Figure 19. Intensity-predicted D/T versus Average Intensity-predicted D/T

Model Selection Example

The standard error of the estimate is a measure of error of prediction and was used to select the best model from a set of two data sets. That is, the lower the standard error the higher the precision, and thus the more preferable model. This example shows the regression results between Average intensity-predicted D/T and DTFCO. Figure 20 shows Average intensity-predicted D/T as the independent variable (predictor) and DTFCO as the dependent variable (response). The slope for this model is 1.65. Figure 21 shows DTFCO as the independent variable (predictor) and Average intensity-predicted D/T as the dependent variable (response). The slope for this model is 0.26. So for Figure 20 one can predict DTFCO as a function of Average intensity-predicted D/T and for Figure 21 one can predict Average intensity-predicted D/T as a function of DTFCO. The standard error was calculated for each model (slope of 1.65 and slope of 0.26) and an error of 0.5 was found for the model slope of 1.65 (Figure 20) and 0.09 for the model slope of 0.26 (Figure 21). When comparing two models, the model with the lowest error (less variance) is the better model. Therefore the model with the slope of 0.26, which is the model with DTFCO as the independent variable (predictor) and Average intensity-predicted D/T, is more precise and has less variance than the other model. For this model, for a DTFCO of 1 D/T an assessment using Average intensity-predicted D/T would be 0.26 D/T. To relate an Average intensity-predicted D/T to a DTFCO one would use the inverse slope (1/0.26=3.8). To relate a an assessment of 1 D/T made using Average intensity-predicted D/T to a DTFCO we would multiply the assessment times 3.8, resulting in an equivalent DTFCO D/T of 3.8.

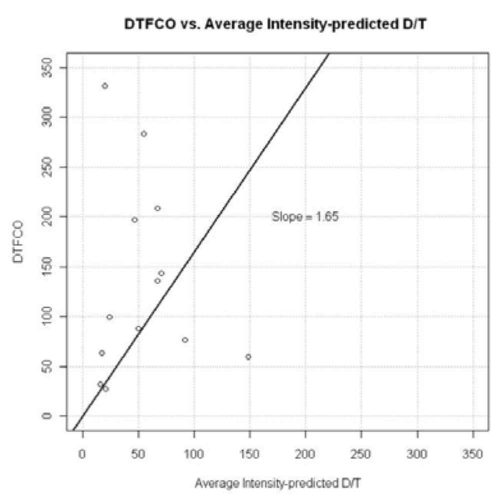

Figure 20. Model Selection Example: DTFCO versus Average Intensity-predicted D/T

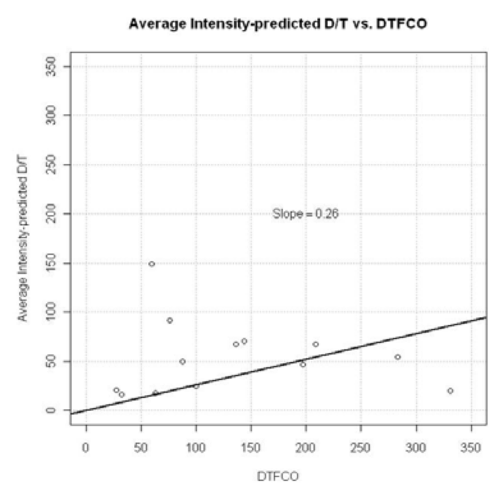

Figure 21. Model Selection Example: Average Intensity-predicted D/T versus DTFCO

Paper No. 3

GROUND TRUTHING AERMOD FOR AREA SOURCE LIVESTOCK ODOR DISPERSION USING ODOR INTENSITY AND THE MASK SCENTOMETER

C. G. Henry, D. D. Schulte, A. M. Parkhurst, R. R. Stowell,

D. P. Billesbach and N. Ebrahim

ABSTRACT. Ambient odors from feedlots and lagoons located in transects within 500 meters of these area sources were assessed by Mask Scentometer and an Odor Intensity Reference Scale (OIRS). Odor dispersion was modeled with AERMOD, using high quality meteorological data, and the predictions were compared to ambient observations using a statistical approach to develop scaling factors. Random effects and autocorrelation were found to be significant in both odor assessment techniques. Scaling factors (slopes) for ground truthing AERMOD with the Mask Scentometer and OIRS methods that accounted for random effects and autocorrelation were within the ranges of 0.1–0.2 and 3.6-29.1, respectively. Using all of the observed data with a mixed effects model (and autocorrelation), as opposed to mean results, provided a better agreement with AERMOD predictions. The Mask Scentometer was more consistent than OIRS when comparing results to AERMOD predictions and is recommended for use in ground-truthing models.

INTRODUCTION

The assessment and prediction of the impact of odors on humans is an area of intense research for the development of simple planning tools in the livestock industry. Assessment techniques, mitigation strategies, and tools to estimate odor risk are especially needed as policy develops in the US to deal with the balance of livestock production and quality of life of rural communities. Ambient odor measurement has historically been a relatively low-technology approach. Air quality models have been used for decades in Europe, South Africa, and Australia to model the impact of odors, for regulatory purposes. Much research has been conducted to calibrate regulatory air quality models with single point monitoring devices and ambient air quality data measured by instruments. These air quality models are now being used in the US to predict odor dispersion and impact, however, one of the significant shortcomings is that, unlike other air quality contaminates, odors are very difficult to measure at ambient levels. No direct machine method has been developed to measure odors without the use of the human olfactory sense. In order to use air quality models for odors, they must be calibrated in a similar fashion as when they were developed for air quality (i.e. single component) contaminates. Thus, consistent and reliable methods for ambient odor assessment are needed that can be used to ground truth or calibrate these models. Such techniques are important in the development of simple odor risk assessment tools for use by the general public.

The research question posed in this paper is, can atmospheric dispersion models be used to reliably predict ambient odor measured by the Mask Scentometer and the Odor Intensity Reference Scale (OIRS) techniques? The objective of this research is to develop relationships between odor levels measured by field methods and predictions from dispersion modeling. Specifically, compared are AERMOD predictions from an area source with measured ambient odor levels based upon an Odor Intensity Referencing System (OIRS) referred to as field intensity or the intensity rating method (using n-butanol as a reference) and the Mask Scentometer. This data was previously reported by Ebrihim (2006) and Henry et al. (2006) and resulted in very high scaling factors. This paper is an effort to account for the variation in the dataset. The outcomes of this work can be used to add source odor emission numbers (OEN) to such simple tools by applying these scaling factors determined in this study to area emission source values for lagoons and open feedlots.

BACKGROUND

ATMOSPHERIC DISPERSION MODELS USED FOR ODOR PREDICTION

Use of atmospheric dispersion models to predict odors downwind of livestock facilities began in the early 1980's (Janni, 1982; Carney and Dodd, 1989; Ormerod, 1991). Currently there are several models being used in the United States to evaluate odor impact. Researchers at the University of Minnesota have used INPUFF-2 (Integrated PUFF), a US EPA Gaussian puff model (Peterson and Lavdas, 1986) to model odors in the development of OFFSET (Odor From Feedlots Separation Estimation Tool) (Jacobsen et al., 2003). The University of Nebraska-Lincoln is using a Gaussian dispersion model, AERMOD (AMS/EPA Regulatory Model) which was developed through a joint effort between the American Meteorological Society and the US EPA. It was selected as a replacement for ISC3 (Industrial Source Complex) in the development of the Odor Footprint Tool (OFT) (Koppolu et al., 2004, Schulte et al., 2004). Lastly, Iowa State University has developed CAM (Community Assessment Model) for predicting odor dispersions in a community (Hoff and Bundy, 2003).

There appears to only be a few dispersion models developed specifically with odor in mind: An Australian model based on Gaussian dispersion called STINK, a German Lagrangian particle model called AUSTAL2000G (Argusoft, 2000), ODODIS (ODOur DISpersion software), which is an un-validated odor model based on the theory of Högström (1972) as described by De Melo Lisboa et al.(2005), and a fluctuating plume dispersion model based on the theory of Gifford (1959) and developed by Mussio et al., (2001). STINK was developed by Smith (1993, 1994) specifically for modeling the dispersion of livestock odors from feedlot pads. AUSTAL2000G a model that accounts for the specific odor perception function of the human nose by incorporating the peak-to-mean approach into the model's predictions. ODODIS shows good agreement with the Prairie Grass database from 1959. The fluctuating plume model facilitates the prediction of odor-impact frequencies in a community. The latter showed good agreement, based on a limited study, with the maximum odour levels reported and was not sensitive to atmospheric stability class or distances between source and receptors (De Melo Lisboa et al., 2005).

Several other models have been adapted and augmented for odor concentration prediction. The Air Pollution Model (TAPM) developed by the Commonwealth Scientific and Industrial Research Organization (CSIRO), and AUSPLUME (Lorimer, 1986, an adaptation of the Industrial Source Complex models). AUSPLUME was developed specifically as a regulatory tool for use primarily in odor impact assessment by the Victoria EPA. Both are examples of existing dispersion models that were adapted from another model and altered to predict odor impact. The Atmospheric Studies Group at TRC Companies Inc., the developers of the CALPUFF modeling system, has been modified by others to be used directly for odor impact assessment. CALPUFF is a multi-layer, multi-species, non-steady-state puff dispersion model that simulates the temporal and spatial variability of meteorological conditions on pollutant transport, transformation, and removal (Scire et al., 2000).

Using models to predict odors is subject to limitations. For example, some models are only able to use basic meteorology for dispersion prediction. INPUFF-2 and AUSPLUME use atmospheric stability classes to select dispersion coefficients. More sophisticated models such as CALPUFF and AERMOD use Monin-Obukhov similarity theory to estimate atmospheric stability, thus providing better algorithms to estimate dispersion parameters. Another primary limitation is the inability of all models to account for short term peak prediction of odors. Conventional dispersion models predict concentrations on an hourly basis and there are a variety of methods to estimate the short-time interval concentrations from hourly observations (e.g. peak-to-mean ratio's).

Xing et al. (2006) compared of ISC3, ASUPLUME, CALPUFF, and INPUFF-2 using ambient odor assessors trained to report an 8 scale 1-butanol OIRS downwind of a series of swine buildings with earthen storages in Canada. They compared two different odor concentrations to intensity prediction equations, one from Zhang et al. (2005) and another from the University of Alberta. They found that these prediction equations were very important in model performance and had mixed results in agreement ranging from 13% to 76% depending on the model, experiment, and prediction equation used. They found scaling factors for the models to be in the range of 1.2 to 7.9, and concluded that none of the models studied were obviously better than another. Finally a few researchers have investigated Computational Fluid Dynamics (CFD) for odor dispersion, and found good agreement between model predictions and measured intensity (OIRS) assessments (Bjerg et al., 2004; Li and Guo, 2006).

AMBIENT ODOR MEASUREMENT METHODS

The Mask Scentometer (Henry, 2004) is functionally similar to a box scentometer. It is comprised of a ¼-face respirator mask with two cartridges. The first cartridge contains a charcoal filter with a plenum and two 1/2-inch holes for drawing in and scrubbing air into the mask. The second cartridge is comprised of a dial mechanism that is used to select one of 5 orifices that adjusts the ratio of ambient air to clean air presented to the user. Paper No. 1 reported the dilution ratios of the Mask Scentometer as 0.35, 1, 2, 4.5, and 18. The advantage of the Mask Scentometer is that it minimizes odor fatigue, allows for one-handed operation, and can be easily operated over long periods of time without fatigue. It also overcomes many of the disadvantages of the Barneby and Sutcliffe scentometer, specifically the poor sealing and discomfort of the glass ports and risk of odor fatigue before use. Users "sniff" through the series of five ports adjusting the dial every 5-8 seconds, allowing a measurement to be taken every 30 seconds.

"Odor Intensity Reference Scales" (OIRS) use a standard odorant as a reference to assess a target odor that is unrelated in character (McGinley, 2002). The ASTM Standard E 544-99 "Standard Practices for Referencing Suprathreshold Odor Intensity" describes a dynamic and static scale method. The standard specifies the static scale method of 12 levels from 10 to 10,000 ppm n-butanol with a geometric progression of two. In this work, a 0-5 scale based on the geometric progression of three, using 25 ppm to represent one, 75 ppm for two, 225 ppm for three, 675 ppm for four, and 2025 ppm for five was used. This is the same technique used by Jacobson et al. (2000); Jacobson et al. (2003); Nicolai et al. (2000); Zhu et al. (2000), however several other methods are known to exist (Misselbrook et al., 1993, Chen et al., 1999; Zhang et al., 2003; Zhang et al., 2005), all based on relationships originally established by Fechner (1860), Stevens (1957), or Beidler (1954).

In the field, the intensity method involves using a respirator mask to maintain osmotic sensitivity and exposing one's nose for a few seconds to take an odor assessment, replacing the mask and repeating the procedure every 10-15 seconds, typically for a period of 10-15 minutes.

PREVIOUS WORK SCALING MODEL RESULTS FOR ODORS

Traditionally, agreement between predicted and observed odor measurements has been achieved using a peak-to-mean methodology. The first references in the literature of peak-to-mean ratios were by Högström (1972) and Smith (1973), where in general, (C_{peak}/C_{mean}) the ratio of the peak measured concentration and the mean predicted concentration is equal to the response time of the human nose (1-5 seconds) divided by the modeling time, typically 60 minutes (t_{peak}/t_{mean}) raised to a power "n". The peak-to-mean ratios are essentially correction (or scaling) factors applied to scale dispersion model results with observations. They were devised to account for the short term fluctuations of odor in the atmosphere. Most dispersion models predict concentrations based on standard micrometeorological observations that are readily available; the shortest time step available is generally one hour. However, peak values as defined by Wilson (1996) as "the concentration that is exceeded m times in a statistically independent ensemble of n repeats of an event" are generally much higher than the average concentration. Several methodologies have been developed and much of the work and controversy centers around the exponent of the power term, which ranges between 0.167 and 0.65 (Pope and Diosey, 2000; Katestone Scientific, 1995; Mejer and Krause, 1986; Mahin, 1998; Best, 2000; Schaugerger et al., 2000 and 2001, Duffee et al., 1991) with the common range being between 0.2 and 0.4.

Curran et al. (2002) in Ireland, used US EPA Industrial Source Complex 3 model (ISC3) and the UK's ADSM3 (Atmospheric Dispersion Modeling System) to demonstrate that setback distances for swine facilities are more appropriately established using models rather than absolute values. In 2007, Curran et al. compared ISC3 and CALPUFF for odors. They measured the downwind odor intensity from a 514 head sow operation. They used olfactometry and ventilation rates to measure odor emission rates (treated as point sources), but performed their field assessments using an OIRS that included a 1-butanol scale and VDI 3940 as a guideline. They found the average predicted versus measured peak-to-mean concentration ratio on the sampling days to vary from 1.4 to 9.4. They found that over 80% of the model predictions were larger than their field observations and concluded that both models gave conservative estimates of downwind odor concentration. Wang et al. (2007) used CALPUFF and ISCST3 for comparing flux hood emission rates to model back-calculated emission rates from feedlot surfaces. She found that CALPUFF could fairly well predict downwind odor concentrations, but ISCST3 tended to under-predict downwind concentrations. They also found that modeled emission rates were higher than flux hood measurements.

Diosey et al. (2002) compared three dispersion model predictions from a simulated waste water treatment plant to measured hydrogen sulfide concentrations and found that AERMOD predictions were one and two orders of magnitude less than those of ISCST3 and CALPUFF, respectively. Modi (2006) compared AERMOD predictions against Nasal Ranger® observations, for four swine barns in Iowa, and found overall scaling factors of 1.66, with the model under predicting the Nasal Ranger measurements. For the same emissions and Nasal Ranger® data, Henry et al. (2007) found a scaling factor of 0.99 for CALPUFF.

The University of Minnesota used sampled emissions from 280 animal buildings and manure storages and used ambient assessors stationed in transects to capture plumes using a 0-5 OIRS based on 1-butanol to ground-truth INPUFF-2, driven by on-site micrometeorological data, for the development of a simple tool they called OFFSET. They successfully developed a tool that provides a setback based on a user selected odor-annoyance-free level of risk and facility types (Guo et al., 2006; Jacobson et al., 2000; Jacobson et al., 2005). The foundation for this work was laid by Zhu et al. (2000a) and Guo et al. (2001), who found that the model predictions of INPUFF-2 were always lower than the observations, and derived scaling factors of 10 for manure storages and 35 for building sources.

The work described here is an extension of the work described by the Master's thesis of Ebrihim (2006) and Henry et al. (2006), where preliminary scaling factors for AERMOD were reported. Ebrihim (2006) inferred scaling factors to peak to mean ratios using measured and predicted results. Ebrihim also developed scaling factors from regression models (what would be considered a cell means fixed effects model in this work). Intensity D/T (also referred to as Intensity-predicted D/T) scaling factors of 1837 for lagoons (or 56 if one lagoon was removed) and 58 for feedlots were reported by Henry et al., (2006). For the Mask Scentometer, scaling factors of 1.8 and 32 were necessary for lagoons and feedlots, respectively. In related work, Koppolu (2002) reported that scaling factors in the range for 0.2 to 3900 were needed to adjust AERMOD predictions to short-term measurement based on the emission and receptor dataset on which OFFSET is based (Jacobson et al., 2000). When Koppolu (2002) used the lagoon data from this study, she found Intensity-predicted D/T scaling factors of 0.2 and 7.0. As can be seen, a method to account for the large variation in scaling factors is needed.

Table 1. Scaling factors reported in previous studied using this dataset.

Study	Lagoon Intensity-predicted D/T	Feedlots Intensity-predicted D/T	Lagoon Mask Scentometer*	Feedlot Mask Scentometer*
Koppolu (2002)	0.2-7.0			
Henry et al., (2006) (scaled by each sniffer and predicted average observation, then averaged all sniffer scaling factors)	1837 or 56	58	1.8	32
Ebrihim (2006) Peak to Mean Method	5.9	4.6	3.5	8.2
Ebrihim (2006) Regression Method (same as cell means fixed effects model in this paper)	15.9	15.3	1.6	4.4

** Factors were developed using D/T values for Mask Scentometer of 2, 7, 15, 31, and 170.*

MATERIALS

In order to compare dispersion model predictions with ambient odor assessment techniques, four trials were set up to place assessors downwind of area odor sources. Area odor sources were selected because open air lagoons and feedlots have not been studied as intensively as ventilated building sources (represented as point sources) and are important odor sources in the livestock industry.

The next generation regulatory dispersion model AERMOD (USEPA, 2004) was selected as the model to predict concentrations downwind of the experiments. This work is an extension of the Master's thesis of Ebrahim (2007). Area source emission rates were measured using a stainless steel wind tunnel that was constructed according to plans from Schmidt and Bicudo (2002). The tunnel was originally designed by Jiang et al. (1995), and consisted of an inlet PVC stack, blower, expansion chamber, air filter, pressure gauge, tunnel body, mixing chamber, outlet PVC "T" and two gas sampling ports. Wind tunnels are portable, open-bottomed enclosures that are placed over the emitting surface. Ambient or filtered air is drawn or blown through the tunnel to mix with and transport gases away from the emitting surface. The air flow through the device is intended to simulate the convective mixing that is responsible for transport processes present during ambient conditions.

Odor emissions were estimated by collecting a bag sample from the outlet port of the wind tunnel and having that sample analyzed by an olfactometry lab according to procedures described by Duyson et al. (2003) and Byler et al. (2004). Swine odors were analyzed by the University of Minnesota odor lab and feedlot odors were analyzed by the West Texas A&M University odor laboratory. Actual emission rates vary with wind speed and atmospheric stability effects. Odor concentrations (reported as dilution to threshold) measured using wind tunnels were adjusted to a standard 1m/s and 1 meter high standardized emission rate, using a procedure outlined in Smith and Watts (1994).

Assessments were made at a swine manure treatment lagoon (all production stages) and two cattle feedlots during the fall of 2003 and spring of 2004. Sites were selected that were topographically level surrounding the source, were isolated from other odor sources, and had few or no obstructions surrounding the site within 500 meters.

LAGOON

The anaerobic lagoon selected for this study treated manure from 1,250 swine finishers, 800 sows and gilts, and 800 nursery pigs. This lagoon was typical of many in Southeast Nebraska and was uniquely situated 1/2 mile to the east of the production barns. This allowed the researchers to isolate the lagoon odors from other odors produced by the production facility or neighboring sources. The lagoon was situated just south of an east-west gravel road, and was surrounded by cropland with little topographical relief surrounding the facility in all directions. The area of the lagoon was 12,022 square meters, and the range of emissions for the first (lagoon 1) and second experiment (lagoon 2) were 2.7-4.7 OU/m^2s and 7-9.8 OU/m^2s, respectively. The flat terrain comprised of rural cropland (harvested cornstalks) allowed “sniffers” to be located in lines perpendicular to and within the influence of the odor plume on a series of transects to the north of the facility on a fall day when the prevailing wind was from the southwest. Transects for lagoon 1 were established at 111, 153, and 198 meters downwind from the lagoon. Eight sniffers were used that day, which was warm and sunny, with no cloud cover and a very gentle and shifting breeze in November of 2003 (0.9 to 2.7 m/s, B and C atmospheric stability class).

On the day of the spring assessments (lagoon 2), the wind was from the northeast. Transects were established downwind at 58, 73, 103, and 134 meters. Atmospheric conditions during the day in February of 2004 were sunny with 3.3 to 6.5 m/s wind speeds (B and C atmospheric stability class) on a day.

FEEDLOTS

The two feedlots selected were a 1,000 head (feedlot 1) and 4,200-head facility (feedlot 2) located in east central and central Nebraska, respectfully. Terrain surrounding both facilities had very little topographical relief and was characteristic of rural farmland (harvested cornstalks). The width of the source was several times greater than the width of the sniffer transect, the area of feedlot 1 and 2 were 51,214 sq meters and 184,845 sq meters, respectively. Transects were established downwind at 106, 308 and 505 meters for feedlot 1 and 150, 265, 390, and 504 meters for feedlot 2. Odor monitoring was conducted on a cold, windy, and overcast day in November of 2003 at feedlot 1 (5.3 m/s to 8.6 m/s, atmospheric stability class D), once in the morning (Feedlot 1 AM) and again in the afternoon (Feedlot 1 PM). For Feedlot 2 odor monitoring was done mid-day on a warm, sunny and windy day in February of 2004 for (atmospheric stability class B, wind speeds between 3.2 and 4.9 m/s). Odor emission rates for the morning experiments for Feedlot 1 were 13.6 to 15.1 OU/(m^2s) and for the afternoon were between 11.2 and 12.3 OU/(m^2s). The unit odor emission rate for Feedlot 2 were between 2.4 and 3.2 OU/(m^2s).

DATA COLLECTION AND MODELING METHODOLOGY

Individuals referred to as field sniffers, refrained from consuming caffeinated drinks, eating spicy foods, and wearing perfume or cologne on days when participating in the study. They first attended a daylong training seminar that instructed them in the use of the Mask Scentometer and how to assess odor intensities. Mask Scentometer readings or “Mask D/T’s” were taken by turning a dial on the mask which selected one of a series of

5 orifices that controlled the dilution ratio of ambient air (odorous air) to air that was cleaned with a carbon filter (clean air) (Henry, 2004). When the sniffer first reached the point at which the odor was recognized, the D/T setting from the dial was recorded. This D/T is considered to be fundamentally the same as an Odor Unit (OU), which can be used directly for comparison with model predicted concentration.

Next, sniffers recorded odor intensities. Sniffers were trained to correlate livestock odors with a reference odorant, n-butanol. Sniffers used a five point OIRS static scale starting at 25 ppm with a geometric progression of three. Sniffers removed their masks briefly to take these measurements.

A micrometeorological station was set up at the facility the day before a sniffing event. It was located in the plume where the sniffers were expected to work the next day. The station was instrumented to record net solar radiation, temperature, relative humidity, wind speed, wind direction, down-welling short wave radiation, and barometric pressure at a height of 3.8 meters above the surface. In addition, a sonic anemometer was installed that allowed estimation of sensible heat flux (H), friction velocity (u*), and Monin-Obukhov length (L) from raw 10 Hz wind speed data. Micrometeorological data were averaged every minute and the results were used to drive the dispersion model. AERMOD and most other dispersion models only accept 1-hour time step meteorological data, so the 1-minute micrometeorological data were used as the input to the model and the hourly model outputs were assumed to be 1-minute predictions without any model adjustment. The authors recognize that atmospheric conditions calculated on a 1-minute time step will not be representative of the transport processes over the entire experimental footprint. Rather, they will be localized around the instrument tower. This will introduce variability into our analysis. However, given the very close nature of the source emissions and sniffers, the authors decided that this was preferable to assuming one atmospheric condition during each assessment. Other researchers have used a similar approach (Li and Guo, 2006; Zhu et al., 2000a; Zhu et al., 2005; Xing et al., 2006). Previous work at the University of Nebraska by Koppolu et al., (2002) compared AERMOD and STINK in small-scale experiments where Volatile Fatty Acids (VFA) dispersion from small children's swimming pools was measured downwind with thermal desorption tubes and SPME fibers. Koppolu evaluated meteorological averaging times of 1, 5, 15, and 30 minutes and found, so long as extreme variations in the wind speed and direction did not occur, shorter periods had better agreement with model predictions. These experiments however, were conducted over much shorter receptor distances where the short averaging times would have minimal effects. She found that AERMOD and STINK both gave comparable predictions of VFA.

The AERMET meteorological preprocessor parameters, surface roughness (range 0.05 to 0.20), Bowen's ratio (range (0.5 to 0.7), and albedo (range 0.18 to 0.20) were input for the conditions of each experiment and used to generate surface and profile data for input to AERMOD. Receptor coordinates, elevations, and heights above ground, were processed in AERMAP, a modeling system preprocessor for source, receptor and terrain inputs, using terrain data supplied by the USGS (Ebrahim, 2007).

On the day of the experiment, sniffers calibrated their olfactory senses with n-butanol at a location away from the plume but near the site. They also calibrated themselves against each other, by subjecting themselves to varying levels of the target odor (near the outer fringes of the plume) and agreeing among themselves that they were all reporting intensity and scentometer readings consistently. They donned their masks before entering the plume. First, Mask Scentometer measurements were taken every 30 seconds, adjusting the dial on the Mask Scentometer until odor was detected. If no odor was detected, a 0 was recorded. Mask Scentometer measurements were taken for 15 minutes. Next, field intensity measurements were taken. The sniffers were instructed to become familiar with the target odor and make their assessment of the target odor, this was done to minimize the influence of background odors that at low concentrations could produce a false positive. Sniffers collected intensity data every 15 seconds for 15 minutes, and dilution to threshold (DT) from a Mask Scentometer every 30 seconds for 15 minutes. Each sniffer was stationed at a location in the transect (Figure 1) for a 30 minute period before moving to another location. Sniffers were randomly located in each transect, and each sniffer self-ensured that they were never in the same transect position on the same day. The mask allowed sniffers to collect two odor measurements sequentially. After sniffers were notified to start taking measurements, they began by taking Mask Scentometer assessments. After they finished, they waited for the odor intensity start signal, and then collected odor intensity measurements. When they finished with these they would move on to the next transect.

Data were manually recorded on pre-printed data sheets. Stopwatches were used that could be set to chime at a set interval. A lead sniffer was identified and synchronized his watch with the weather station clock so measurements would correspond to modeled data. The lead sniffer would then start all of the other sniffers at the same time.

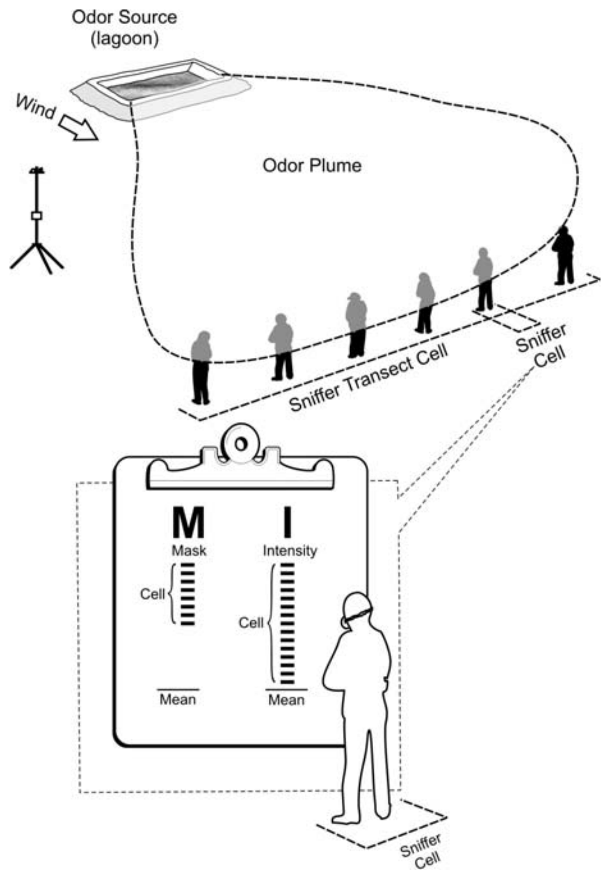

Figure 1. Graphical representation of experiments showing sniffer-transect cells and sniffer cells.

DATA ANALYSIS METHODOLOGY

A comparison of the ambient odor assessments was done using the R statistical package (R Development Team, 2008) to evaluate fixed and random effects that could be described by a mixed effects model. The goal of this procedure was to remove variability from the data and produce a scaling factor (slope) between model predictions and ambient odor assessment observations. Additionally, the purpose of this methodology was to find the best fitting model with a slope (scaling factor) that best represented all of the data for each experiment. To analyze the data, four statistical models were used, of increasing in complexity. These were:

1. Cell means fixed effects model (using the average assessment value for each sniffer in each 15 minute session).
2. Fixed effects model (uses all of the observations from each 15 minute session),
3. Mixed effects model (accounts for random effects),
4. Mixed effects model with autocorrelation effects (accounts for random effects and autocorrelation).

The approach used the following two steps until the best model with either separate transect slopes (multiple slopes model) or a single common slope was found for each experiment:

1. Each of the four models were developed in two ways, first the slope for each sniffer-transect (7-10 sniffers per transect) was developed, then the data was pooled (all transects (sniffer-transect) for the experiment) to estimate a common slope for the dataset. An ANOVA was used to test whether the common slope was representative of the transects for the experiment. .
2. Test the less complex model against the more complex model (using ANOVA and goodness of fit tests). If the more complex model is a better fit, then test against the next more complex model. If the less complex model is a better fit, then that model's slope is the scaling factor.

Autocorrelation was used to find repeating patterns, such as the presence of periodic signals which could possibly be buried under noise and to reduce the sensitivity of slope estimates to data outliers. Since autocorrelation violates the assumptions of model errors being independent, it becomes a parameter in the model, and is the amount of the previous error term that needs to be added to the current measurement.

The simplest model tested was the cell means fixed effects model. The observations for each of the sniffers were averaged and a single value for each transect (sniffer-transect cell) was used for this model. There were 60 observations for odor intensity assessments and 30 observations for the Mask Scentometer (see Figure 1). Each of these is defined as a sniffer cell. Each transect, comprised of six to ten sniffers, produced a set of sniffer-transect cells. Two cell means models were developed, a common slope model which used all of the data from the transects together and then models for each of the transects. The null hypothesis, Ho: Transect slopes are not different from the common slope of all of the sniffer-transects combined were tested against Ha: at least one slope is different between the models slopes. If there was a significant difference (reject Ho, accept alternative hypothesis, then we concluded that the slopes were not common, meaning that there was too much variation between sniffer-transects to find a common slope. The cell means model had the following form:

$$y_{ij} = \beta_i x_{ij} + \varepsilon_{ij}$$ Assumptions: $\varepsilon_{ij} \sim N(0,\sigma^2)$ Identically & Independently Distributed

Where *i*, represents the transect, *j* represents the sniffer cell data, .y_{ij} is the dependent variable, the average of the sniffer observations, x_{ij} is the independent variable, the average of the AERMOD predictions, β_i is the slope of the line, which is interpreted to be the model scaling factor, and ε_{ij} is the error term which indicates the deviation of a particular sniffer-transect cell or observation from the line. For the multiple slopes models the slope (β) is calculate for each of the transects ($\beta_{Transect\ 1}$, $\beta_{Transect\ 2}$, etc.) and for the common slope model, all of the averaged observation and prediction data from the transects is pooled and used to calculate the slope (β_{common}).

Next we used all of the data (all observations from sniffer) in the sniffer cells to determine if the transect slopes are common using a fixed effect model with all of the sniffer observations. The fixed effect model has the following form (through origin).

$$y_{ijk} = \beta_i x_{ijk} + \varepsilon_{ijk}$$

Assumptions: $\varepsilon_{ij} \sim N(0,\sigma^2)$ Identically & Independently Distributed

Where *i* represents the transect, *j* denotes the sniffer cell and *k* denotes the observation in that sample (within the sniffer cell). The dependent variable (y_{ijk}), is the response or sniffer observation, x_{ijk} is the independent variable , the AERMOD prediction. The slope of the line for the common slope is β and the slope for the ith transect is β_i for the multiple slope model, and ε_{ijk} is the error term which indicates the deviation of a particular sniffer-transect cell or observation from the line.

Next, random effects were added to the model, referred to as the mixed effects model, which is comprised of both fixed and random effects. This model uses all of the data in the observations in the sniffer cells, like the previous model and the variation in multiple observations is removed from the error term and used to estimate the random effects sniffers have on the slope. The same procedure is followed for this model, first determining transect slopes and common slopes and comparing them to see if they are common. . The mixed effects model has the following form (through origin):

$$y_{ijk} = x_{ijk}\left(\beta_i + b_{ij}\right) + \varepsilon_{ijk}$$

Assumptions: $b_{ij} \sim N(0,\sigma^2)$ and $\varepsilon_{ijk} \sim N(0,\sigma^2)$ Identically & Independently Distributed

For the multiple slopes models the slope (β_i) is calculate for each of the transects and for the common slope model, all of the data from the transects is pooled and used to calculate the slope (essentially β_{common}). The additional parameter, b_{ij} denotes the random effect variable that is added to the equation from the transect-sniffer cell. The random effect due to the slope of the ij^{th} transect-sniffer cell (b_{ij}) is assumed to be normally distributed with a mean of zero and a variance of σ^2_{cell}. The residual error for the ijk^{th} observation, (ε_{ijk}) is assumed to be normally distributed with a mean of zero and a variance of σ^2. This allows for the possibility that predictions from the fitted equation have two sources of error, the usual residual error associated with the measurement process within the cell and the error associated with the sampling process for the transect-sniffer cells.

The last model adds autocorrelation to the mixed effects model. We first check for common slopes between the individual transects and then determine if the effect of autocorrelation is significantly different (and improved) from the mixed effects model without autocorrelation. Again the estimates of the individual transect slopes are tested against the estimate of the common slope. The mixed effects model with autocorrelation has the following form (through origin).

$$y_{ijk} = x_{ijk}\left(\beta_i + b_{ij}\right) + \phi y_{i-1,j-1} + \varepsilon_{ijk}$$

Assumptions: b_{ij}~N(0,σ^2) and ε_{ijk}~N(0,σ^2) Identically & Independently Distributed

The additional parameter to the mixed effects model with autocorrelation, φ is the autocorrelation effect of the previous observation ($y_{i-1,j-1}$) on the next (y_{ij}).

The following eight results are the possible outcomes of this procedure:

- A cell means fixed effects model with either (i) a common slope or (ii) separate estimate of transect slopes;
- A fixed effects model with either (iii) a common slope or (iv) separate estimate of transect slopes;
- A mixed effects model accounting for random effects only with either (v) a common slope or (vi) separate estimate of transect slopes; or
- A mixed effects model accounting for random and autocorrelation effects with either (vii) a common slope or (viii) separate estimates of transect slopes.

INTENSITY-PREDICTED D/T RESULTS

The results for the field sniffers were analyzed using the intensity data converted to intensity-predicted D/T using the relationships established from Jacobsen et al. (2000) and Nicolai et al.,(2000) for beef and swine odors. The data was analyzed for each experiment separately.

$$D/T_{beef\,(feedlots)} = 9.429 \cdot e^{1.085I}$$

$$D/T_{swine(lagoon)} = 8.367 \cdot e^{1.078I}$$

Where I is the intensity of the odor as determined by a trained assessor and D/T is referred to as Intensity-predicted D/T for the remainder of this paper.

LAGOON 1 INTENSITY

Results for lagoon 1 Intensity-predicted D/T are shown in Table 2. For the cell means model, the slopes of the individual transects (111m, 153m, 198m) were found to be 3.2, 1.0, and 2.8, respectively with a common slope of 2.8 (p=0.62). When a fixed effects model using all of the sniffer observations in the cells was used to estimate the slopes, the individual transect slopes were found to be 1.3, 0 and 0.6. Although the common slope was estimated 1.0, it was not-representative of the transect slopes ($p<0.0001$). When random effects were included in the model, there was a significant improvement ($p=0.0001$). For this mixed effects model, the individual transect slopes were 14.7, 0 and 0.7 with a common slope of 6.8 (standard error for this slope is 3.8, $p=0.16$). There was no improvement when autocorrelation was incorporated in the mixed model ($p=0.22$). Thus, a mixed model with a common slope of 6.8 without autocorrelation best represents the Intensity-predicted D/T scaling factor for this lagoon.

Table 2. Slope Results for Lagoon 1 Intensity-predicted D/T

Model	111 m	153 m	198 m	Common Slope
Cell Means Fixed Effects Model Slope	3.2	1.0	2.8	Yes 2.8
Fixed Effects Model Slope (all data)	1.3	0	0.6	No (1.0)
Mixed Effects Model Slope	14.7	0	0.7	Yes 6.8*
Mixed Effects Model and Autocorrelation Slope	N/A	N/A	N/A	N/A

**Conclusion: Mixed effects model with common slope of 6.8 without autocorrelation.*

LAGOON 2 INTENSITY

The results of this experiment, shown in Table 3, were previously reported in Ebrahim (2007) and are summarized here for further comparison. Ebrahim (2007) reported individual transect slopes of 26.5, 6.7, 37.4 and 70.6 and a common slope of 30.8 ($p=0.31$) using a cell means model. He found the random effects to be significant ($p=<0.001$) and autocorrelation to be significant ($p=<0.0001$), however the common slope of 13.8 (standard error of 2.4) did not represent all of the transects adequately. Using a mixed model and autocorrelation resulted in transect slopes of 8.7, 3.6, 17.7 and 29.1 with standard errors of 3.4, 3.2, 3.6, 3.5 respectively. Therefore the mixed effects model produced the best result with an autocorrelation parameter of 0.6 and that separate slope estimates for transects were necessary to represent scaling factors for Intensity-predicted D/T for lagoon 2.

Table 3. Slope Results for Lagoon 2 Intensity-predicted D/T

Transect	58 m	73 m	103 m	134 m	Common Slope
Cell Means Fixed Effects Model Slope	26.5	6.7	37.4	70.6	Yes 30.8
Fixed Model Slope (all data)	13.9	4.5	26.2	44.1	No (19.7)
Mixed Effects Model Slope	12.6	4.4	25.3	46.9	Yes 22.1
Mixed Effects Model Slope and Autocorrelation	*8.7	*3.6	*17.7	*29.1	No (13.8)

** Conclusion: Mixed effects model with separate transect slope estimates (8.7, 3.6, 17.7, 29.1) and autocorrelation (0.63).*

FEEDLOT 1 AM INTENSITY

Results for Feedlot 1 AM Intensity are shown in Table 4. Slopes of 15.1, 16.7 and 22.3 and a common slope of 15.7 were found using a fixed effects cell means model. When all of the observations in the sniffer-transect cells was used to construct a fixed effects model, transect slopes of 10.9, 8.3 and 20.9 were found but the common slope of 11.2 was not representative of the three transect slopes ($p<0.0001$). When random effects ($p=0.001$) and autocorrelation ($p=0.0001$) were included in the model there was significant improvement. However, no common slope for the mixed effects model was found ($p=0.0259$) using an alpha of 0.05. If strict accordance to an alpha of 0.05 is observed, then the mixed effects model with a transect (distance) effect is warranted of 8.6, 6 and 20 and an autocorrelation parameter of 0.3. The standard errors for these transect slopes are 2.9, 6.0, and 20.0, respectively. However, using a very significant alpha level of 0.01 ($p=0.0259$) would allow a common slope of 10.7 (standard error of

2.2) to be used and the same autocorrelation parameter. By strict adherence to an alpha of 0.05 for common slope definition, a mixed effects model with an autocorrelation effect of 0.3 and separate transect slopes (8.6, 6, 20) are necessary. However, the slopes are nearly common, so a mixed effects model with autocorrelation parameter of 0.3 and a common slope of 10.7 is considered to be sufficient to characterize the Intensity-predicted D/T scaling factor for this experiment.

Table 4. Slope Results for Feedlot 1 AM Intensity-predicted D/T

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	15.1	16.7	22.3	Yes 15.7
Fixed Model Slope (all data)	10.9	8.3	20.9	No (11.2)
Mixed Effects Model Slope	10.5	8.1	20.1	Yes 12.7
Mixed effects Model Slope and Autocorrelation	*8.6	*6.0	*20.0	No/Maybe (**10.7)

**Conclusion: Mixed effects model with common slope (10.7) and autocorrelation (0.3).*

FEEDLOT 1 PM INTENSITY

Results for Feedlot 1 PM are shown in Table 5. Transect slopes of 11.1, 13.9 and 7.9 were found using the cell means fixed effects model. A common slope of 11.2 was found to represent the three transects. Using all of the observations in the cell decreased the slopes to 7.3, 8.3 and 4.3, but the common slope of 7.3 was not adequate to represent the transect slopes. When the random effects ($p=0.0001$) were included in the model there was significant improvement ($p=0.0001$). For the mixed effect model transect slopes were 6.8, 7.7 and 4.9 with a common slope of 6.7 ($p=0.61$). When autocorrelation was included in the model it resulted in significant improvement ($p=0.001$, parameter of 0.3). The transect slopes for this model (6.5, 7.1, and 4.2) were represented adequately by the common slope of 6.3. The standard error for this model was 1.06. The mixed effects model with a common slope of 6.3 and an autocorrelation parameter of 0.3 characterize the Intensity-predicted D/T scaling factor for this experiment.

Table 5. Slope Results for Feedlot 1 PM Intensity-predicted D/T

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	11.1	13.9	7.9	Yes 11.2
Fixed Effects Model Slope (all data)	7.3	8.3	4.3	No (7.3)
Mixed Effects Model Slope	6.8	7.7	4.9	Yes 6.7
Mixed Effects Model Slope and Autocorrelation	6.5	7.1	4.2	Yes *6.3

**Conclusion: Common Slope Mixed effects model (6.3) with autocorrelation (0.3)*

FEEDLOT 2 INTENSITY

The results of Feedlot 2 are shown in Table 6. For the cell means model the slopes of the individual transects (150, 265, 390, 504 meters) were found to be 141.8, 396.8, 35.2, and 61.8, respectively with a common slope of 48.1 (p=0.095). When a fixed effects model using all of the observations from the sniffer cells was used to estimate slopes the individual slopes of 21.8, 18.0, 17.1 and 10.5, respectively with a common slope of 15.6 (p=0.81). Unlike the other experiments, the random effects were not significant (p=0.52), and resulted in the same slope (15.7). When autocorrelation was included in the model, the effect was significant (p=0.0001, parameter=0.3). The fixed effect model with a common transect slope estimate of 15.6 and an autocorrelation parameter of 0.3 characterized the Intensity-predicted D/T scaling factor for this experiment.

Table 6. Slope Results for Feedlot 2 Intensity-predicted D/T

Transect	150 m	265 m	390 m	504 m	Common Slope
Cell Means Fixed Effects Model Slope	141.8	396.8	35.2	61.8	Yes 48.1
Fixed Effect Model Slope (all data)	21.8	18.0	17.1	10.5	*Yes 15.6
Mixed Effects Model Slope	N/A	N/A	N/A	N/A	15.7
Mixed Effects Model Slope and Autocorrelation	N/A	N/A	N/A	N/A	15.7 (autocorrelation is significant)

** Conclusion: Common slope fixed effect model (15.6) with autocorrelation (0.3)*

INTENSITY-PREDICTED D/T RESULTS DISCUSSION

The results of the experiments using intensity are shown in Table 7. Except for lagoon 1, the autocorrelation effect was always significant, and except for feedlot 2, random effects were significant in the data. It is not clear why this is the case, but it is assumed for the lagoon 1 experiment, that the low odor concentrations experienced by the sniffers compared to the other experiments is the culprit. What is interesting is that much of the variation was removed from the data sets by using all of the data in the sniffer-transect cells and adding the additional random effect and autocorrelation parameters in the model. For example, using only the means for Feedlot 1 resulted in very large slopes for the transects (141, 396, 35, and 61) but by using all of the data in each of the sniffer cells, these extreme slopes were reduced to a common slope of 15.7. Likely some “puffs” of odor or sniffer errors resulted in outliers that skewed the results when averaged. There is a reduction of slopes in all cases from a mean cell fixed effects model, to a fixed effects model using all of the observations in the cells, and a further reduction in slope when the random effects are accounted for, and again when autocorrelation is included in the model.

Unfortunately for lagoon 2 and feedlot 1 AM common transect slopes could not be found. If a uniform slope had to be established for area sources, the average of all the experiments is 12, with a range between 3.6 and 29.1 (the geometric mean of these two values is 15.9). There is no trend in the data that would allow us to assess whether large or small distances were a factor (feedlot 1 AM and PM were at the same distance), nor if there was an odor type effect (beef versus swine). Also, the effect of autocorrelation was found in four of the five experiments. Random effects and autocorrelation were generally present in these datasets. Furthermore, if the cell means fixed effects model common slopes are averaged for Intensity-predicted D/T (all fixed effects models resulted in common slopes), 2.8, 30.8, 15.7, 11.2, and 48.1, their average is a slope of 22. We can surmise that the difference between 22 and 12 is due to random effects and autocorrelation attributable to the odor assessors and the nature of odor dispersion.

Table 7. Summary of Intensity-predicted D/T Slope Results

Experiment	Slope	Random Effects	Autocorrelation	SE
Lagoon 1	6.8	Yes	No	3.8
Lagoon 2	8.7, 3.6, 17.7, & 29.1	Yes	Yes (0.6)	3.4, 3.2, 3.6, 3.5
Feedlot 1 AM	8.6, 6, & 20	Yes	Yes (0.6)	2.8, 3.2, 3.8
Feedlot 1 PM	6.3	Yes	Yes (0.3)	1.1
Feedlot 2	15.6	No	Yes (0.3)	3.4
Range	3.6-29.1		(0.3-0.6)	1.1-3.8
Average	12		0.45	

MASK SCENTOMETER RESULTS

The following relationship, first used by Sheffield et al. (2004), was used to geometrically average the Dilutions to Threshold $(D/T)_G$ readings for Field Olfactometers:

$$(D/T)_{G,n} = 10^{\frac{\log D/T_n + \log D/T_{(n+1)}}{2}}$$

Where n is the setting of the device, i.e. 3rd setting is 7 D/T.

This procedure normalizes the peaks and keeps extremely high or low values from skewing the results. One of the issues with data from field olfactometers is how to deal with non-detects, where no odor level is measured because it is not possible to take the geometric average of results that include zeros. Rather, taking the geometric average of the adjacent device settings is preferred and reasonable since the geometric average of settings 2 and 3 for example is 0.6 which is between 0.35 and 1, so when the device reports a 0.35 (setting 2) it is more likely that it is somewhere between a 0.35 and a 1. The geometric D/T's used are shown in Table 8. The $(D/T)_G$ have the effect of increasing the overall mean result of an assessment period compared to taking the average of the unit's D/T's, and the drawback that when no odor is present, an odor level is reported (although small, i.e. not less than 0.2 D/T). In this situation, it is assumed that an odor was present during the assessments, so the effect of the non-detects is small on the overall results. Two caveats exist with the $(D/T)_G$ approach, first that professional judgment was used to approximate the geometric mean D/T between non-detect of 0 and 0.35, 0.2 D/T was used. Second, no attempt was made to adjust the last setting (6) of 18 D/T (D/T could be between 18 and infinity). The results were analyzed using the Mask

Scentometer data from the field sniffers first using the unit D/T and geometric D/T shown in Table 8.

Table 8. Geometric Dilutions to Threshold $(D/T)_G$ used for Mask Scentometer

Mask Scentometer		
Unit D/T	Setting, n	Geometric D/T
18	6	18
4.5	5	9
2	4	3
1	3	1.4
0.35	2	0.6
0 / No detect	1	0.2

Lagoon 1 Mask Scentometer Results

The data were analyzed for D/T and $(D/T)_G$; the results are shown in Table 9. For the D/T data, the cell means fixed effect model results in transect slopes of 0, 0.04 and 0.02, but the common slope of 0.01 was not representative of the transects. When all of the cell data in a fixed effects model was used, the transect slopes were found to be 0, 0.01 and 0, but the common slope of 0 was not representative of these individual transect slopes. These slopes were not found to be statistically different from zero. For D/T it was found that including the random effects did not significantly improve the model ($p=0.99$). However, autocorrelation did improve the model significantly ($p=0.0001$) for the mixed effects model, resulting in a parameter of 0.3. Because the slopes are so near to zero, we concluded that the results are not reliable. There was very little odor experienced by the sniffers, and very little predicted by the model. The odor that was present appears to have been below the detection threshold of the Mask Scentometer.

Table 9. Summary of Lagoon 1 Mask Scentometer D/T Slope Results

Transect	111 m	153 m	198 m	Common Slope
Cell Means Fixed Effects Model Slope	0	0.04	0.02	No (0.01)
Fixed Effect Model Slope (all data)	0	0.01	0	No (0)
Mixed Effects Model Slope	0	0	0	0
Mixed Effects Model Slope and Autocorrelation	0	0	0	0

Conclusion: no reliable result, slope is 0.

The analysis was repeated using $(D/T)_G$ instead of D/T, and the results are shown in Table 10. When the $(D/T)_G$ are used in the analysis, the fixed effects model using the cell means produced uncommon slopes (p=0.00001) of 0.1, 0.1, and 0.2. When all of the observations from the sniffer cells were used, individual transect slopes of 0 were found. While random effects were found to be significant (p=0.001), and autocorrelation was found to be significant, the resulting slopes were not different from zero, so for $(D/T)_G$, (like for D/T) we concluded that no conversion factor (slope=0) exists from this dataset.

Table 10. Summary of Lagoon 1 Mask Scentometer $(D/T)_G$ Slope Results

Transect	111 m	153 m	198 m	Common Slope
Cell Means Fixed Effects Model Slope	0.1	0.1	0.2	No (0.1)
Fixed Effect Model Slope (all data)	0	0	0	No (0)
Mixed Effects Model Slope	0	0	0	No (0)
Mixed Effects Model Slope and Autocorrelation	0	0	0	Yes (0)

Conclusion: no reliable result, slope is 0.

LAGOON 2 MASK SCENTOMETER RESULTS

Unlike the lagoon 1 experiment, the lagoon 2 experiment was conducted on a day that was more conducive to odor emission and transport and the sniffers were considerably closer to the source. When D/T was analyzed using the cell means fixed effects model, the transect slopes (53, 73, 103, and 134 m) were 0.3, 0.1, 0.5, and 1.6 with a common slope of 0.3. However the common slope (0.3) was not representative of the individual transect slope estimates (p=0.0001). When all of the sniffer cell observations were used in the fixed effects model, individual transect slopes of 0.5, 0.3, 0.9 and 1.3 were found with a common slope of 0.5; however, the commons slope was again not representative of the individual transect slopes (p=0.005). When random effects were included in the model, there was significant improvement (p=0.001) but the common slope was still not representative.. Finally, when autocorrelation was included in the model, there was significant improvement (p=0.0001) with a parameter of 0.8. The individual transect slope estimates were found to have a common slope of 0.1 (p=0.71) and represented those transect slopes. The standard error for the slope of the mixed effects model slope (with autocorrelation) is 0.02. The mixed effects model with a common transect slope estimate of 0.1 and an autocorrelation parameter of 0.8 characterizes the scaling factor for this experiment.

Table 11. Summary of Lagoon 2 Mask Scentometer D/T Slope Results

Transect	53 m	73 m	103 m	134 m	Common Slope
Cell Means Fixed Effects Model Slope	0.3	0.1	0.5	1.6	No (0.3)
Fixed Effects Model Slope (all data)	0.2	0.1	0.4	0.6	No (0.2)
Mixed Effects Model Slope	0.2	0.1	0.5	0.8	No (0.4)
Mixed Effects Model Slope and Autocorrelation	0.1	0.1	0.2	0.1	Yes *0.1

Conclusion: Common transect slope of 0.1 from mixed effects model with autocorrelation (0.8)

When the data was reanalyzed using $(D/T)_G$, similar results were found as can be seen in Table 12. The cell means fixed effect model estimated individual transect slopes of 0.5, 0.2, 0.8 and 2.7, but the common slope of 0.5 was not representative of the individual transect slopes (p=0.006). When all of the data were used, the individual transect slope estimates were 0.4, 0.2, 0.7, and 1.0 but the common slope of 0.5 was again not representative of the individual slopes. When random effects were included in the model, there was significant improvement (p=0.001). Autocorrelation also improved the results (p=0.001) for the mixed effects model. The autocorrelation parameter was found to be 0.7. No significant difference between transect slopes were found, so a common slope of 0.2 is representative of the individual transect slopes. The standard error for the mixed effect model is 0.05. The mixed effects model with a common transect slope estimate of 0.2 and an autocorrelation parameter of 0.7 characterizes the scaling factor for this experiment.

Table 12. Summary of Lagoon 2 Mask Scentometer $(D/T)_G$ Slope Results

Transect	53 m	73 m	103 m	134 m	Common Slope
Cell Means Fixed Effects Model Slope	0.5	0.2	0.8	2.7	No (0.5)
Fixed Effect Model Slope (all data)	0.4	0.2	0.7	1.0	No (0.4)
Mixed Effects Model Slope	0.4	0.2	0.7	1.0	No (0.6)
Mixed Effects Model Slope and Autocorrelation	0.1	0.1	0.4	0.2	Yes *0.2

** Conclusion: Mixed effects model with autocorrelation with a common slope of 0.2 and autocorrelation (0.7)*

FEEDLOT 1 AM MASK SCENTOMETER RESULTS

Table 13 shows the results for Feedlot 1 AM. For the individual transects (106, 308, and 505 meters) slope estimates using the cell means fixed effects model were found to be 0.6, 0.7, and 0.7 with a common slope of 0.6 (p=0.98) for D/T. A common slope of 0.5 was also found using the sniffer cell observations. When random effects were included in the model, there was further significant improvement (p=0.0001). Additionally, when autocorrelation was included in the mixed model, there was significant improvement (p=0.0001) again and the parameter was found to be 0.9. A mixed model with a common slope of 0.2 (the standard error was 0.05) including the effect of autocorrelation best represents this experiment.

Table 13. Summary of Feedlot 1 AM Mask Scentometer D/T Slope Results

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	0.6	0.7	0.7	Yes 0.6
Fixed Effect Model Slope (all data)	0.5	0.5	0.7	Yes 0.5
Mixed Effects Model Slope	0.5	0.5	0.6	Yes 0.6
Mixed Effects Model Slope and Autocorrelation	0.2	0.2	0.4	Yes *0.2

** Conclusion: Mixed effects model with autocorrelation with common slope of 0.2 and autocorrelation (0.9).*

Similar results (Table 14) was found for $(D/T)_G$, using the cell means fixed effects model resulted in slopes of 0.7, 0.8, and 0.9 and a common slope of 0.8 (p=0.92) represented the transect slopes. A similar result was found when all of the sniffer cell observations were used. When random effects were included in the model, there was significant improvement (p=0.001). Autocorrelation also significantly improved the model (p=0.0001, parameter =0.9). For this model the common slope of 0.1 is representative of the individual transect slope estimates of 0.1, 0.2, and 0.4 (p=0.49). The mixed effects model with a common slope of 0.1 (the standard error was 0.15) and the autocorrelation parameter of 0.9 best represents this experiment.

Table 14. Summary of Feedlot 1 AM Mask Scentometer $(D/T)_G$ Slope Results

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	0.7	0.8	0.9	Yes 0.8
Fixed Effect Model Slope (all data)	0.6	0.6	0.8	Yes 0.6
Mixed Effects Model Slope	0.6	0.7	0.8	Yes 0.7
Mixed Effects Model Slope and Autocorrelation	0.1	0.2	0.4	Yes *0.1

**Conclusion: Mixed Effects model with autocorrelation with a common slope of 0.1 and autocorrelation (0.9).*

FEEDLOT 1 PM MASK SCENTOMETER RESULTS

The results for Feedlot 1 PM D/T are shown in Table 15. For the cell means fixed effects model, the slopes for the individual transects (106 m, 308 m, and 505 m) were found to be 0.2, 0.4 and 0 with a common slope of 0.2 and was found to be representative of the individual transects (p=0.19). When the sniffer cell observations were used (fixed effects model), the transect slopes of 0.1, 0.3 and 0 represented the transect slopes, but they were not common to a slope of 0.1. When random effects were included in the model, there was significant improvement (p=0.001) and the inclusion of the autocorrelation effect further improved the model (p=0.0001, parameter=0.5). The transect slopes of this mixed effects model were common (0.1) to two of the transect slopes of 0.1and 0.1 (p=0.13). The mixed effects model with a common slope of 0.1 (standard error of 0.02) and autocorrelation parameter of 0.5 best represent the scaling factors for this experiment.

Table 15. Summary of Feedlot 1 PM Mask Scentometer D/T Slope Results

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	0.2	0.4	0	Yes 0.2
Fixed Effect Model Slope (all data)	0.1	0.3	0	No (0.1)
Mixed Effects Model Slope	0	0.3	0	Yes 0.1
Mixed Effects Model Slope and Autocorrelation	0	0.1	0.1	Yes *0.1

Conclusion: Mixed effects Model with autocorrelation with a common slope of 0.1 and autocorrelation of (0.5).

Similar results were found for $(D/T)_{G.}$ as shown in Table 16. The individual transect slope estimates for the cell means fixed effects model were found to be 0.2, 0.4, and 0.1 with a common slope of 0.2 that did represent the individual transects (p=0.19). When all of the observations were used in the fixed effects model the individual transect slopes were found to be 0.1, 0.3 and 0.1 with a common slope of 0.1, but it did not represent the individual transect slopes (p<0.0001). When random effects were included in the model there was significant improvement (p<0.0001). When autocorrelation was included in the model, there was further improvement (parameter=0.05, p<0.0001). For the mixed model with autocorrelation, two of the individual transect slopes were found to be 0.04 and 0.2 were representative of the common slope of 0.1 (p=0.136, standard error 0.02. The mixed effects model with a common slope of 0.1 (p=0.136) and an autocorrelation parameter of 0.5 best represent the scaling factor for this experiment.

Table 16. Summary of Feedlot 1 PM Mask Scentometer $(D/T)_G$ Slope Results

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	0.2	0.4	0.1	Yes 0.2
Fixed Effect Model Slope (all data)	0.1	0.3	0.1	No 0.1
Mixed Effects Model Slope	0.1	0.3	0.1	Yes 0.1
Mixed Effects Model Slope and Autocorrelation	0.04	0.2	0	Yes *0.1

**Conclusion: Mixed effects model with autocorrelation with a common slope of 0.1 with autocorrelation (0.5).*

FEEDLOT 2 MASK SCENTOMETER RESULTS

The results for Feedlot 2 Mask Scentometer D/T are shown in Table 17. For the cell means fixed effects model, the slopes for the individual transect (150m, 265m, 390m, and 504m) were found to be 0.8, 10.7, 0.7, and 0.6. A common slope of 0.8 was found to represent the individual transect slopes. When all of the sniffer observations were used, (fixed effects model), the transect slopes were found to be 0.1, 2.0, 0.3, and 0.1, and the common slope of 0.3 was found to be representative of the individual slopes. When random effects and autocorrelation are included in the model, there was significant improvement (both p=0.0001). For the mixed effects model with autocorrelation, the individual transect slopes of 0.3, 0.6, 0.1, and 0.04 with a common slope of 0.2 was found to be representative of the individual transect slopes (p=0.66, standard error 0.1). It should be noted that the 265 meter transect slopes are considerably larger than the other three transects, and the reason for this is unknown. The discrepancy between the model predictions and measured Mask Scentometer readings are much greater for this transect. The mixed effects model with autocorrelation with a common slope of 0.2 and the autocorrelation parameter of 0.6 best represents this experiment.

Table 17. Summary of Feedlot 2 Mask Scentometer D/T Slope Results

Transect	150 m	265 m	390 m	504 m	Common Slope
Cell Means Fixed Effects Model Slope	0.8	10.7	0.7	0.6	Yes 0.8
Fixed Effect Model Slope (all data)	0.1	2.0	0.3	0.1	Yes 0.3
Mixed Effects Model Slope	0.5	2.0	0.3	0.1	Yes 0.6
Mixed Effects Model Slope and Autocorrelation	0.3	0.6	0.1	0.04	Yes *0.2

**Conclusion: Mixed effects model with autocorrelation with a common slope of 0.2 and autocorrelation (0.6)*

When D/T_G is used, a similar result (Table 18) is found, the cell means fixed effect model produces a common slope of 1.4, from transect slopes of 1.4, 16.5, 1.6, and 1.2. Using all of the observations from a sniffer cell yields transect slopes of 0.7, 3.4, 0.6, and 0.3 with a common slope of 0.5 that does represent the individual slopes (p=0.057). However, when random effects (p=0.0001) and autocorrelation (p=0.0001) are included in the model there was a significant improvement. For the mixed effects model, individual transect slopes of 0.9, 3.6, 0.6, and 0.3 are common to a single slope of 0.6 (p=0.12). The mixed effects model with an autocorrelation effect produced individual transect slopes of 0.2, 0.1, 0.1 and 0 a common slope (p=0.65) of 0.10, an autocorrelation parameter of 0.8, and a standard error of 0.1. The mixed effects model with a common slope of 0.1 and autocorrelation parameter of 0.8 best represent this experiment.

Table 18. Summary of Feedlot 2 Mask Scentometer D/T_G Slope Results

Transect	150 m	265 m	390 m	504 m	Common Slope
Cell Means Fixed Effects Model Slope	1.4	16.5	1.6	1.2	Yes 1.4
Fixed Effect Model Slope (all data)	0.7	3.4	0.6	0.3	Yes 0.5
Mixed Effects Model Slope	0.9	3.6	0.6	0.3	Yes 0.6
Mixed Effects Model Slope and Autocorrelation	0.2	0.1	0.1	0	Yes *0.1

**Conclusion: Mixed effects model with autocorrelation with a common slope of 0.1 and autocorrelation (0.8)*

MASK SCENTOMETER RESULTS DISCUSSION

The Mask Scentometer results for the experiments with D/T and for $(D/T)_G$ are shown in Table 19 and Table 20, respectively. Except for the lagoon 1 experiment, random effects and autocorrelation were consistently present across the experiments. Unlike the intensity method, a common slope was found for each experiment. The range of slopes for D/T and $(D/T)_G$ of the Mask Scentometer were nearly the same, 0.1-0.2. These slopes lie in a much tighter range than those from Intensity-predicted D/T measurements. It does not appear to matter whether D/T or $(D/T)_G$ was used in this type of analysis. It appears that AERMOD under predicts the Mask Scentometer. If AERMOD predicted an odor concentration of 6 Odor Units, it would be roughly equivalent to a 1 D/T assessment by a Mask Scentometer if random effects and autocorrelation are accounted for. The wind tunnel is a likely source of large error. Slightly higher source emission rates could very well bring the slopes closer to 1:1. Therefore, what is important from a consistency perspective, is not that the slope deviates from 1:1, but rather that the slopes found with the Mask Scentometer are clustered in a tight group near a slope of 0.15.

In our analysis, the mixed effects model with autocorrelation is analyzing whether or not a sniffer's response is auto-correlated to the response just before, then looks ahead to the next response to see if the effect exists again with the one just before it. The autocorrelation parameter can range from -1 to 1, and a value of 0 means that there is no autocorrelation effect. When the autocorrelation parameter is positive, it suggests the next response in a cell is trending higher, and a negative parameter suggests the next response is trending lower. Similar to intensity, the autocorrelation effect was found in the Mask Scentometer dataset, even the lagoon 1 data. The range of autocorrelation parameters for the Mask Scentometer was 0.5 to 0.9 for both D/T and $(D/T)_G$. Intensity autocorrelation parameters ranged from 0.3 to 0.6, indicating a slightly higher autocorrelation effect from the Mask Scentometer.

This gives us much greater confidence in the Mask Scentometer method over the Intensity-predicted D/T method (and possibly other similar OIRS methods), when used to ground truth models. The lower standard errors in the final Mask Scentometer models (range 0.02-0.15) compared to the Intensity-predicted D/T data (1.0-3.8), provide more evidence that the Mask Scentometer method is more robust. However, there is generally more of an autocorrelation effect with the Mask Scentometer, than the Intensity-predicted D/T method. We were able to account for the random effects consistently with the Mask Scentometer and always found a common transect slope for the experiments, which was not the case with the Intensity-predicted D/T method, suggesting that the Mask Scentometer method is more reliable.

Table 19. Summary of Slope Results for All Mask Scentometer D/T Experiments

Mask D/T				
	Random Effects	Slope	Autocorrelation	SE
Lagoon 1	No Reliable Result			
Lagoon 2	Yes	0.1	0.8	0.02
Feedlot 1AM	Yes	0.2	0.9	0.05
Feedlot 1 PM	Yes	0.1	0.5	0.02
Feedlot 2	Yes	0.2	0.6	0.1
Range		0.1-0.2	0.5-0.9	0.02-0.1

Table 20. Summary of Slope Results for All Mask Scentometer $(D/T)_G$ Experiments

Mask $(D/T)_G$				
	Random Effects	Slope	Autocorrelation	SE
Lagoon 1	No Reliable Result			
Lagoon 2	Yes	0.2	0.7	0.11
Feedlot 1AM	Yes	0.1	0.9	0.15
Feedlot 1 PM	Yes	0.1	0.5	0.02
Feedlot 2	Yes	0.1	0.8	0.1
Range		0.1-0.2	0.5-0.9	0.02-0.15

The low slopes of the Mask Scentometer could be explained by several shortcomings of the experiment. First, the lagoon and feedlot emission rate samples used to calculate down wind odor intensities were very low and close to or below the Lower Detection Limit (LDL) of the olfactometry labs. Odor samples were between 8.5 to 34 OU D/T compared to much higher concentrations experienced from buildings and manure storages >100 D/T. Ebrahim (2007) reported emission rates between 1 and 19 OU/(m^2-s) for all of these same experiments. There are also some uncertainties with the odor emission sampling device, the wind tunnel, especially at such low odor concentrations, these could have overestimated the emission rate used in the modeling. Additionally, the wind speed and stability conversion developed by Smith and Watts (1994) has never been validated in the literature. So there is some uncertainty as to the exact emission rate from the area sources, more so than would be present in the current literature. With that said, it is clear that when we compare the ranges of slopes from the experiments, after the random effects and autocorrelation are accounted for, we can produce a reliable slope from the Mask Scentometer for area odor sources.

The results for all slopes from the models are shown in Table 21. While we found random effects and autocorrelation to be significant and while we did not always find a common slope with the cell means fixed effects model, the fixed effects model, and the mixed effects model, their slopes are closer to 1:1 than the mixed effects model with autocorrelation. For the cell means fixed effects model using $(D/T)_G$, the range of slopes is 0.2 to 1.4 (average slope of 0.5 for D/T and 0.7 for $(D/T)_G$). This gives the slope range and average closest to a 1:1. For those using the Mask Scentometer to ground-truth dispersion models, this approach provides the scaling factor nearest to one (a scaling factor of one being an ideal scaling factor).

Table 21. Model Slopes for Mask D/T and Mask D/T_G

Mask D/T/ Mask D/T_G Common Transect Slopes				
	Cell Means Fixed Effects Model slope	Fixed Effects Model slope (all data)	Mixed effects Model slope	Mixed Effects Model slope and Autocorrelation
Lagoon 1	No reliable result			
Lagoon 2	0.3 No / 0.5 No	0.2 No / 0.4 No	0.4 No / 0.6 No	0.1 Yes / 0.2 Yes
Feedlot 1AM	0.6 Yes / 0.8 Yes	0.5 yes / 0.6 yes	0.6 Yes / 0.7 Yes	0.2 Yes / 0.1 Yes
Feedlot 1 PM	0.2 Yes / 0.2 Yes	0.1 No / 0.1 Yes	0.1 Yes / 0.1 Yes	0.1 Yes / 0.1 Yes
Feedlot 2	0.8 Yes / 1.4 Yes	0.3 Yes / 0.5 Yes	0.6 Yes / 0.6 Yes	0.2 Yes / 0.1 Yes
Range for D/T_G	0.2-1.4	0.1-0.6	0.1-0.7	0.1-0.2
Average	0.5/0.7	0.3/0.4	0.4/0.5	0.2/0.1

DISCUSSION

When we combine the results from the two methods and all experiments, some trends appear. When intensity was converted to D/T using the relationship developed by Jacobson et al., (2000), random effects were significant for four of the five experiments. Autocorrelation was significant for four of the five experiments with the range of parameters being between a positive 0.3 to 0.6.

Slopes or correction factors of 3.6 to 29.1 were found with standard errors for the mixed effects models to be in the range of 1.0 to 3.8. From this it is difficult to select a single slope that could be used as a universal scaling factor for modeling. Other similar studies that used an OIRS with dispersion modeling (Zhu et al.,2000a and Guo et al., 2001) reported scaling factors of 10 for manure storages and 35 for buildings (point sources), which is of the same order of magnitude as our scaling factors for AERMOD (INPUFF-2 was used in their work). Xing et al., (2006) found scaling factors of 1.2 to 7.9 for ISCST3, AUSPLUME, CLAPUFF and INPUFF-2 using a similar OIRS. We conclude that Intensity-predicted D/T scaling factors are dependent on the model and intensity conversion equation. From our data, the mean of the slopes was 12, meaning that we need to factor model predictions by 12 to match intensity observations. One caveat, is that while this appears to agree with other studies, these studies did not account for random and autocorrelation effects in their analysis.

The Mask Scentometer was analyzed two different ways, first using the reported D/T's from the instrument and using a geometric scale (D/T_G). For D/T, random effects were significant in four of the five experiments. Autocorrelation was significant in all of the experiments; the range of parameters was 0.5-0.9. Slopes or correction factors ranged from 0.1 to 0.2 with standard errors of the mixed effects model slopes being in the range of 0.02 to 0.15. Autocorrelation was found in all of the experiments (range 0.7 to 0.9), except lagoon 1, where likely the odor was below the detection limit of the Mask Scentometer. This analysis gives more support to the scaling factor of 0.15, if we select a value in the middle of the range.

It appears that the statistical techniques used to analyze the data were able to account for a large portion of the variability in the dataset. In a practical sense, the random effects of odor measurement and autocorrelation of the human subjects were accounted for in the model. This effectively “removed” them from the slope estimate (scaling factor) which results in a true slope, unaffected by those effects. Using the mixed effects model with autocorrelation resulted in a very consistent scaling factor (slope) for the Mask Scentometer.

The effect of autocorrelation was present in the dataset except for one Intensity-predicted D/T experiment. The range of autocorrelation parameters for the Mask Scentometer were positive and ranged from 0.5 to 0.9 for D/T and $(D/T)_G$. The autocorrelation effect was more pronounced in the Mask dataset, than the Intensity dataset (range 0.3 to 0.6) suggesting that the sniffers observations from the Intensity OIRS method were more independent. There are three likely causes of the autocorrelation and random effects. First there is a physical effect and the random nature of odor transport that the models cannot accurately predict. Second there is a physiological effect due to the sensitivity and repetitive activity associated with the sniffers olfactometry nerves. The repeated firing of these nerves may explain the random and autocorrelation effect. Finally, the third suspect to these effects is the psychology of the human psyche that they recall their last assessment and that likely plays a role in their next observation.

For all of the experiments the slope was smaller when using all of the sniffer observations than the model slopes where cell means were used, and generally there was a significant difference between them. This is an important finding, as much of the previous work has averaged the observations of a sniffer session. Information is lost when only the means are used. Other researchers should exercise caution when "averaging" all of the assessments from an observer, as this practice significantly impacts the results. This is definitive for Intensity-predicted D/T measurements; however, for the Mask Scentometer, the smallest AERMOD model scaling factors were found using the cell means fixed effects model, 0.5 for D/T and 0.7 for $(D/T)_G$. The inverse of this slope is 2 for D/T and 1.4 for $(D/T)_G$, which would be the scaling factors that need to be applied to the Mask Scentometer to scale them up to AERMOD predictions. While random and autocorrelation effects are present and assuming that source emissions were not overestimated it may be more valuable when the Mask Scentometer is used for model ground-truthing to use the cell means fixed effects model and ignore the significance of the random effects and autocorrelation. In a practical sense, this is the most straight forward means to analyze data from this method, if one is seeking a 1:1 relationship and the goal is to match model results with in field observations. Based on this, we need to either factor down one minute averaged AERMOD predictions down by 0.5 for $(D/T)_G$ or 0.7 for D/T, or factor up the average assessment from a Mask Scentometer by 1.4 for $(D/T)_G$ or 2 for D/T to match model predictions (note that this is the opposite of intensity scaling).

CONCLUSIONS

The dispersion of odors from large area sources, lagoons and feedlot pads were studied using a Mask Scentometer and a 0-5 point OIRS referenced to 1-butanol. Field assessors were located in transect lines downwind of the odor source, in the plume where on-site micrometeorological data were obtained to drive the AERMOD dispersion model. The observed results from field assessors were compared to model predictions using statistical approaches that accounted for random and autocorrelation effects. The conclusions are:

1. The scaling factor (slope) relating Intensity-predicted D/T to AERMOD predictions was found to be between 3.6 and 29.1. We recommend using 12 for low emission rate odor area sources such as lagoons and feedlot surfaces.
2. The scaling factor (slope) relating the Mask Scentometer to AERMOD results was found to be between 0.1-0.2 for both D/T and $(D/T)_G$. A scaling factor of 0.15 is suggested instead of the values 1.6 and 8.2, (for lagoons and feedlots, respectively) as reported by Ebrahim (2006). The geometric scale (D/T_G) does not provide a significant advantage over the D/T scale for the Mask Scentometer.
3. A caveat to the scaling factors for Intensity-predicted D/T and the Mask Scentometer is that there is uncertainty surrounding the emission estimation used in this study. The fact that such low odor emissions were present, and our ability to measure low odor emission rates and to adjust them for wind speed and stability play a large role in the scaling factors. However, the narrow range of scaling factors and their closer proximity to 1:1 for the Mask Scentometer than for the Intensity-predicted D/T method, suggests that the Mask Scentometer is more reliable for ambient odor assessment methods.

4. Autocorrelation parameters were found to be positive and lie between 0.3 to 0.6 for Intensity-predicted D/T and between 0.5 to 0.9 for the Mask Scentometer for D/T and $(D/T)_G$. Intensity-predicted D/T demonstrated slightly less autocorrelation than the Mask Scentometer. The effect was pronounced in both methods, which suggests that in both methods, a previous observation affects the next one. The autocorrelation effect is significant in field odor assessments.
5. Information is lost when only the means from an assessment are used. The fixed effects model using all of the sniffer's observations significantly improved the results over that of the cell means fixed effects model.
6. Random effects were significant in our dataset, for nearly all the experiments. Using a mixed effects model was effective in accounting for these effects. In all cases the mixed effects model accounted for the random variation in the sniffer observations.
7. Using the mixed effects model accounting for autocorrelation to ground-truth dispersion models using an OIRS by scaling the AERMOD model predictions by 12 (slope) and accounting for random and autocorrelation effects is recommended. Conversely, the Intensity-predicted D/T cell means model resulted in an average slope of 22 (12 when effects were accounted for). The difference between these slopes can be explained by the random effects and autocorrelation effects attributable to those that assess odors.
8. For the Mask Scentometer the cell means model produced an AERMOD scaling factor of 0.5 for D/T and 0.7 for $(D/T)_G$, (meaning Mask Scentometer assessments should be factored by 2 for D/T and 1.4 for $(D/T)_G$ to match model predictions). While it requires a smaller scaling factor, this approach has the disadvantage of

ignoring the random effects and autocorrelation. Conversely, the difference between the slopes of the cell means fixed effects model (0.5/0.7) and mixed effects model with autocorrelation (0.15) can be explained by the random effects and autocorrelation effects attributable to those that assess odors. However, the cell means fixed effects scaling factors of 0.5 for D/T and 0.7 $(D/T)_G$ have value for simplicity and comparing scaling factors found in other studies.

ACKNOWLEDGEMENTS

The authors would like to thank the Nebraska Department of Agriculture for their financial support for this work. The lead author would also like to thank the faculty and staff at the University of Nebraska for their support.

REFERENCES

ArguSoft-Germany, Austria and Switzerland, 2000: http://www.austal2000g.de/austal2000g.htm

ASTM. 2000. D 6589-00: Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance. Annual Book of ASTM Standards. Philadelphia, PA: American Society of Testing and Materials.

ASTM. 1999. E 544-99. Standard practices for referencing suprathreshold odor intensity. American Society for Testing Materials, Philadelphia, PA.

Beidler, L.M. 1954. A Theory of taste stimulation. *J Gen Physiol*. 38(2):133-139.

Best P.R. 2000. Averaging Time Corrections for Estimating Extreme Air Quality Statistics, 15th International Clean Air Conference, Sydney, Australia.

Bjerg B., P. Kai, S. Morsing, and H. Takai. 2004. CFD Analysis to Predict Close Range Spreading of Ventilation Air from Livestock Buildings. Agricultural Engineering

International: the *CIGR Journal of Scientific Research and Development Manuscript* BC 03 014. Vol. VI.

Byler, J., D. D. Schulte and R. K. Koelsch. 2004. Odor, H2S and NH-3 Emissions from Phototrophic and Non-Phototropic Anaerobic Swine Lagoons. Written for presentation at the 2004 ASAE/CSAE Annual International Meeting. Fairmont Chateau Laurier, The Westin, Government Centre, Ottawa, Ontario, Canada.

Carney, P. G. and V. A. Dodd. 1989. A comparison between predicted and measured values for the dispersion of malodours from slurry. *J agric. Eng. Res*. 44(1):67-76.

Chen, Y., D.S. Bundy, S.J. Hoff. 1999. Development of the relationship between odor intensity and threshold dilution ratio for swine units. *J. of the Air and Waste Management Association.* 49:1082-1088.

Curran, T. P., B.A. Sheridan, E.T. Hayes and V. A. Dodd. 2002. Comparison of two dispersion models for odour nuisance prediction from pig units. Paper No. 02-4014. ASAE/CIGR Annual International Meeting, Chicago, July 28-31, ASAE, St. Joseph, MI.

Curran, T.P., V. A. Dodd, W. L. Magette. 2007. Evaluation of ISC3 and CALPUFF Atmospheric Dispersion Models for Odor Nuisance Prediction. Paper number 074181. Annual International Meeting, Minneapolis, MN, 17-20 June. ASAE, St. Joseph, MI.

Disoey, P.G., M.E. Hess and L. Farrell. 2002. What smells here? *Water Env. & Tech*: pp 36-39.

De Melo Lisboa, H., J.M. Guillot, J. L. Fanlo, P. L. Cloirec. 2005. Dispersion of odorours gases in the atmosphere-Part I: Modeling approaches to the phenomenon. *Science of the Total Environment*.

Duffee, O'Brien and Ostojic. 1991. Odor Modeling-Why and How. Recent Developments and Current Practices in Odor Regulations, Controls and Technology, Air and Waste Management Association.

Duyson, R., Erickson, G., Schulte, D. & Stowell, R. (2003). Ammonia, Hydrogen Sulfide, and Odor Emissions from a Beef Cattle Feedlot. Written for presentation at the 2003 ASAE International Meeting, ASAE., Las Vegas, NV.

Ebrahim, N. 2007. Process Based Modeling of Odor Dispersion from Area Source Livestock Facilities. Unpublished MS Thesis. Biological Systems Engineering, University of Nebraska, Lincoln, NE.

Fechner, G.T. 1860. Elemente der psychophsik. Leipzig: Breitkopf and Hartel. English translation of Vol 1 by Helmut E. Adler. 1966. Elements of psychophysics. (D. H. Howes and E.G. Boring, Eds). Holt, Rinehart and Winston, New York.

Gifford, F. A. 1959. Statistical properties of a fluctuating plume dispersion model. *Advances in Geophysics* 6: 177-137.

Guo, H, L.D. Jacobson, D.R. Schmidt, and R.E. Nicolai. 2001. Calibrating INPUFF-2 model by resident panelists for long-distance odor dispersion from animal production sites. *Applied Eng in Agric* 17(6): 859-868.

Guo, H., L.D. Jacobson, D.R. Schmidt, R.E. Nicolai, J. Zhu, K.A. Janni. 2005. Development of the OFFSET model for determination of odor-annoyance-free setback distances from animal production sites: Part II. Model development and evaluations. *Transactions of the ASAE* 48(6).

Guo, H., Z. Yu, C. Lague. 2006. Livestock Odour Dispersion Modeling: A Review. Written for presentation at the CSBE/SCGAB 2006 Annual Conference Edmonton, Alberta, July 16-19, 2006. Paper no. 06-170. The Canadian Society for Bioengineering.

Glostelow, P., Longhurst, P., Parsons, S. A., Stuetz, R. M.. 1993. Sampling for measurement of odours. IWA Publishing, London, UK.

Hartung, E. and T. Jungbluth. 1997. Determination of the odour plume boundaries from animal houses. In Proc. 5th Int. Symposium on Livestock Environment, Bloomington, Minn., 1, 163-179. St Joseph, Mich., ASAE.

Henry, C. G. 2004. Instruction Manual for Mask Scentometer. University of Nebraska-Lincoln, Lincoln, NE.

Henry C. G., D.D. Schulte, R.K. Koelsch, R.R. Stowell, D.P. Billesbach, N. Ebrihim, A.M. Parkhurst, and D. B. Parker. 2006. Comparing Field Odor Assessment Methods with an Atmospheric Dispersion Model for Calibrating Setback Estimation Tools for Livestock. Proceedings in the Workshop on Agricultural Air Quality: State of the Science. Bolger Conference Center, Potamac, MD. June 5-8, 2006.

Henry, C. G., P. C. D'Abreton, R. J. Ormerod, G. G. Galvin, S. J. Hoff, L. D. Jacobsen, D.D. Schulte, & D. L. Billesbach. 2007. Downwind Odor Predictions from Four Swine Finishing Barns Using CALPUFF. Presented at the International

Symposium on Air Quality and Waste Management for Agriculture. 16-19 September 2007. Publication number 701P0907cd, ASABE, St. Joseph, MI

Hoff, S. J. and D. S. Bundy 2003. Modeling odor dispersion from multiple sources to multiple receptors. In: Proceedings of the International Symposium on Gaseous and Odour Emissions from Animal Production Facilities. Norsens, Denmark. pp 331-339.

Högström, U. 1972. A method for predicting odour frequencies from a point source. *Atmos Environment*, 6: 103-121.

Huey, N.A., Broedering, L.C., Jutze, G.A., and Gruber, C. W. 1960. Objective Odor Pollution Control Investigations. Presented at the 83rd Annual Meeting of APCP, May 22-26, 1960.

Jacobson, L.D., H. Guo, D.R. SchmiD/T, 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.

Jacobsen, L. D., H. Guo, D. R. SchmiD/T, 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: Proceedings of the International Symposium on Gaseous and Odour Emissions from Animal Production Facilities. Horsens, Denmark, June 1-4, 2003. pp 320-330.

Jacobson, L. D., H. Guo, D. R. SchmiD/T, R. E. Nicolai, J. Zhu, K. A. Janni. 2005. Development of the OFFSET model for determination of odor-annoyance-free setback distances from animal production sites: Part I: Review and experiment. *Transactions of the ASAE* 48(6).

Janni, K. A. 1982. Modeling dispersion of odorous gases from agricultural sources. *Transactions of the ASAE* 25: 1721-1723.

Jiang K., Bliss, P. J. and T. J. Schuls. 1995. The Development of a sampling system for determining odour emission rates from areal surfaces: Part 1. Aerodynamic performance. *J. of Air and Waste Management Association* 45:917-922

Katestone Scientific. 1995. Peak-to-Mean ratios for odour assessment. Report to the Environmental Protection Authority of New South Wales.

Koppolu, L., D. D. Schulte, S. Lin, M.J. Rinkol, D.P. Billesbach and S. B. Verma. 2002. Comparison of AERMOD and STINK for dispersion modeling odorous compounds. Paper no. 02-4015. St. Joseph, MI.: ASAE.

Koppolu, L., D. SchmiD/T, D. D. Schulte, and L. Jacobson 2004. Development of scaling factors (peak-to-mean ratios) through dispersion modeling with AERMOD and field-based odor measurements for livestock facilities. ASAE paper # 044196. ASAE, St. Joseph, MI.

Li, J., D. S. Bundy, S. J. Hoff, and Q. Liu. 1994. Field odor measurement and applications of Gaussian plume model. ASAE Paper No 94-4054. St. Joseph, Mich, ASAE.

Li, J. and Guo H. 2006.Evaluating the Impact of Computation Time Interval on Livestock Odour Dispersion using a Computation Fluid Dynamics Model. Written for presentation at the CSBE/SCGAB 2006 Annual Conference, Edmonton Alberta, July 16-19. Paper number 06-173, The Canadian Society for Bioengineering.

Lin, S. 2002. Modeling and Evaluation the Atmospheric Dispersion of Volatile Fatty Acids. MS Thesis. Lincoln, NE University of Nebraska.

Lorimer, G. D., 1987. The AUSPLUME Gaussian Plume Dispersion Model, Environmental Protection Authority of Victoria, Australia.

McGahan, Eugene; Alan Skerman; Ken Casey and Tim Streeten. 2001. Separation Guidelines for Queensland Piggeries. State of Queensland, Department of Primary Industries and Fisheries, Queeensland.

McGinley, C. M. 2002. Standardized Odor Measurement Practices for Air Quality Testing. Presented at the Air and Waste Management Association Symposium on Air Quality Measurement Methods and Technology. San Francisco, CA.

Mejer, G.J. and K.H. Krause. 1986. Dispersion models fro emissions from agricultural sources. In *Odor Prevention and Control for Organic Sludge and Livestock Farming*. pp 99-112. V.C. Nielsen, J.H. Voorberg and P.L. Hermite, eds. Elsevier Appl. Sci. Publ. New York, N.Y.

Misselbrook, T.H., C.R. Clarkon, B.F. Pain. 1993. Relationship between concentration and intensity of odours for pig slurry and broiler houses. *J. Agric. Eng. Res.* 55:163-169.

Modi, M. 2006. Modeling odor dispersion form a swine facility using AERMOD. Unpublished MS thesis. Department of Biological Systems Engineering. University of Nebraska Lincoln, Lincoln, NE. 86p

Mussio, P., A.W. Gnyp, P. F. Henshaw. 2001. A fluctuating plume dispersion model for the prediction of odour-impact frequencies from continuous stationary sources. *Atmospheric Environment*, 35:2955-2962.

Ormerod, R. J. 1991. Modelling Emission and Dispersion of Feedlot Odour – A Case Study. *In Proc of Workshop on Agricultural Odours*, Toowoomba, February 1991.

Nicolai, R. E., C. J. Clanton, and H. Guo. 2000. Modeling the relationship between threshold and intensity of swine odors. *Second International Conference on Air Pollution from Agricultural Operations.* St. Joseph, Mich.: ASAE.

Parker D. B., M. B. Rhoades, J. Koziel, and J. Spinhirne, 2003. Background Odors in Tedlar® Bags Used for CAFO Odor Sampling. ASAE Paper No. 034144. Las Vegas, NV: ASAE.

Parker D. B., Rhoades M. B., Schuster G. L., Koziel J. A., Perschbacher-Buser (2005) *Odor characterisation at open-lot beef cattle feedyards using triangular forced choice olfactometry*. Transactions of the ASAE 48.

Peterson, B. W., and L. G. Lavas. 1986. Inpuff 2.0 – A multiple source Gaussian puff dispersion algorithm: User Guide. Research Triangle Park, N. C.: Atmospheric Sciences Research Laboratory. Office of Research and Development, USEPA.

Pope, R.J. and P.G. Diosey. 2000. Odor dispersion: Models and Methods.

Qu Guoliang and John Feddes. 2006. Estimation of measurable error caused by background odor in sampling bags. ASAE Paper No. 064141. Portland, OR: ASABE.

Rinkol, M. J. 2000. Modeling the Atmospheric Dispersion of Odors from Agricultural Sources. M.S. Thesis. University of Nebraska. Lincoln, NE.

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.

Schauberger G., M. Piringer, E. Petz. 2000. Diurnal and annual variation of the sensation of distance of odour emitted by livestock buildings calculated by the Austrian odour dispersion model (AODM). *Atmospheric Environment*, 34:18-28.

Schauberge G., M. Piringer. E. Petz. 2001. Separation distance to avoid odour nuisance due to livestock calculated by the Austrian odour dispersion model (AODM). *Agriculture, Ecosystems and Environment* 87:13-28

Schmidt, D. R and J. R. Bicudo. 2002. Using a wind tunnel to determine odor and gas fluxes from manure surfaces. Paper number 024083. ASAE, St. Joseph, MI.

Scire, J.S., Strimaitis, D.G. and Yamartino, R.J. 2000. A user's guide for the CALPUFF dispersion model (version 5), Earth Tech, Inc., Concord, MA.

Sheffield,R. E., M. Thompson, B. Dye and D. Parker. 2004. Evaluation of Field-based Odor Assessment Methods. Water Environment Federation and the Air and Waste Management Association Conference on Odor and Air Quality. March 19-21, 2004. Tacoma, WA.

Smith, M. E. 1973. Recommended Guide for the Prediction of the Dispersion of Airborne Effluents. ASME, New York.

Smith, R. J. 1993. Dispersion of odors from ground level agricultural sources. *J. Agric. Eng. Res.* 54: 187-200

Smith R. J.and P. J Watts. 1994. Determination of odors from ground level agricultural sources. *J. Agric. Eng. Res*. 54: 187-200.

Stevens, S.S. 1957. On the Psychophysical Law. *The Psychological Review.* 64(3):153-181.

Stowell, R.R., L. Koppolu, D. D. Schulte, and R. K. Koelsch 2005. Proceedings of the Seventh International Symposium, 18-20 May 2005 (Beijing, China) Publication Date 18 May 2005.

Unites States Environmental Protection Agency. 2004. User's Guide for the AMS/EPA Regulatory Model-AERMOD. EPA-454/B-03-001, September 2004.

Verein Deutscher Ingenieure (VDI) 1993. VDI 3881: Olfactometry Odour Threshold Determination. Berlin, Germany.

Wilson J. D. 1996. Concentration Fluctuation and Averaging Times in Vapor Clouds. American Institute of Chemical Engineers.

Xing Y., H. Guo, J. Feddes, S. Shewchuck. 2006. Evaluation of Air Dispersion Models Using Swine Odour Plume Measurement Data. CSBE paper No. 06-172.

Zhang, Q. J.J.R. Feddes, I.K. Edeogu, and X. J Zhou. 2003. Correlation between odour intensity assessed by human assessors and odour concentrations measured with olfactometers. *Canadian Biosystems Eng*. 44(6):27-32.

Zhang, Q., X.J. Zhou, H.Q. Guo, Y.X. Li, and N. Cicek. 2005. Odour and greenhouse gas emissions from hog operations. Final Report Submitted to Manitoba Livestock Manure Management Initiative Inc.

Zhou, X. J., Q. Zhang, H. Guo, Y. X. Li. 2005. Evaluation of air dispersion models for livestock odour application. CSAAE/SGIR 2005 meeting, Winnipeg, Manitoba, Paper number 05-026.

Zhu, J., L. D. Jacobson, D. R. Schmidt, and R. Nicolai. 2000. Evaluation of INPUFF-2 model for predicting downwind odours from animal production facilities. *Applied Engineering in Agriculture*. 16 (2): 159-164.

APPENDIX

An example of the analysis demonstrated in this paper is shown. The analysis procedure for Feedlot1 PM for Intensity-predicted D/T is described in further detail. Table 5, from the results section is reprinted below for the reader's reference.

Table 5. Results for Feedlot 1 PM Intensity-predicted D/T

Transect	106 m	308 m	505 m	Common Slope
Cell Means Fixed Effects Model Slope	**11.1**	**13.9**	**7.9**	**Yes 11.2**
Fixed Effects Model Slope (all data)	**7.3**	**8.3**	**4.3**	**No (7.3)**
Mixed Effects Model Slope	**6.8**	**7.7**	**4.9**	**Yes 6.7**
Mixed Effects Model Slope and Autocorrelation	**6.5**	**7.1**	**4.2**	**Yes *6.3**

**Conclusion: Common Slope Mixed effects model (6.3) with autocorrelation (0.3)*

For Feedlot1 PM ten sniffers were used (labeled as A-H and PI) and three transects (labeled as 1-3) with transect 1 being the nearest transect (106 meters) and 3 being the furthest transect (505 meters) from the source. The first step is to develop the cell means fixed effects model and evaluate that model for a singular slope that represents the three transects. The data is plotted in Figure 2 and suggests that a common slope may represent the three transects. The slope for transect 1 is 11.1, 13.9 for 2, and 7.9 for 3. An ANOVA which tested maximum likelihood or referred to as an F-test for the purpose of this Appendix, between the transect slopes and the common slope (11.2) was accepted (p=0.723) indicating that the transects can be represented with a common, singular representative slope.

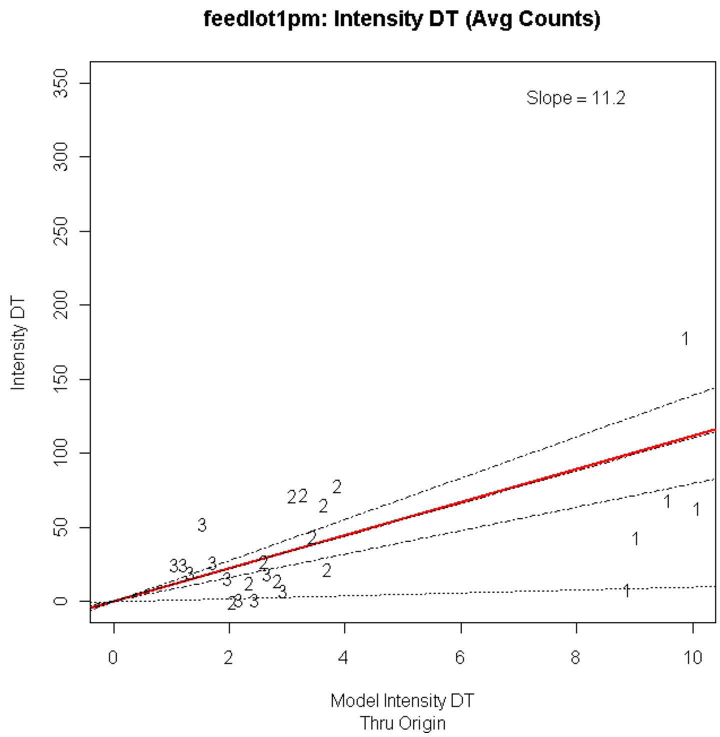

Figure 2. Feedlot 1 PM: Intensity-predicted D/T versus Model Predictions.

The cell means fixed effects model fit to the raw data gives an estimate for the common slope of 11.2. This analysis restricts inferences to this sample of transect-sniffer cells. That is, it is restricted to the intensity-predicted D/T data observed on these cells.

The same process is repeated for the fixed effects models where each of the individual assessments are used develop the model rather than the average of the session. Again the two slope models are tested (ANOVA) to see if the common slope of 7.3 was representative of the transect slopes of 7.3, 8.3, and 4.3. A F-test was significant ($p>0.0001$) meaning that the common slope of 7.3 was not representative of the transects. Additionally the fixed effects (individual transect model) was tested against the cell means fixed effects model and was found to be significant ($p>0.0001$) meaning that the fixed effects model provided a better fit to the data.

Next the mixed effects models were developed for both a common slope and separate transect slopes. The separate slopes were found to be 6.8, 7.7, and 4.9 and the common slope model was 6.7. The F-test was not significant (p=0.31) meaning that the common slope was representative of the individual transect slopes. An ANOVA comparing the mixed effects model and fixed effects model was significant (F-test, p<0.0001) meaning that the mixed effects model gave a better fit to the data than did the fixed effects model. Figure 3 enables comparison of the individual estimates of slopes (lis) with the individual with-in cell "random effects estimate" (lme) and with the estimate for the population average slope (fixed effects coefficient). This plot shows how the within cell estimates represent a compromise between the separate slope estimates and the common fixed effects slope. The mixed-effects model (data indicated with a "+") shrinks the separate slope estimates (data represented by "o") toward the population average.

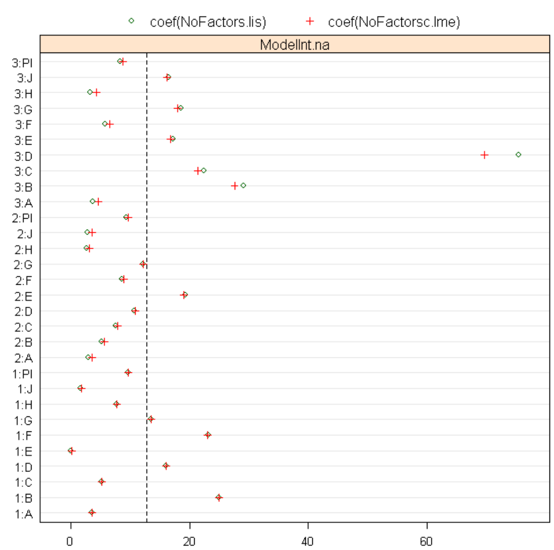

Figure 3. Plot of sniffer averages of the fixed effects model (lis) and the mixed effects model (lme) showing the deviation from the population average slope (experiment mean for all sniffers and all transects).

Finally the effect of autocorrelation is incorporated to the mixed effects model, the separate transect slope estimates are 6.5, 7.1, and 4.2 and the model with a singular common slope is 6.3. Comparing the two slope models with an ANOVA is not significant ($p=0.62$) meaning that the common slope is representative of the individual transect slopes. Testing the effect of autocorrelation is significant ($p>0.001$) meaning that including autocorrelation improved the mixed model.

The pooling of cells in the lme (mixed model) estimation gives a certain amount of robustness to accounting for individual outlier behavior. This is better illustrated by Figure 4. The assumption of independent residuals has been relaxed by including a test for autocorrelation within the counts of each transect-sniffer cell. Taking autocorrelation into account was significant ($\varphi=0.3$, $p=0.0001$) and reduced the estimate of the common slope from 6.7 to 6.3 by reducing the models sensitivity to outliers. Thus the errors in the model can no longer be assumed to be independent. The autocorrelation parameter represented the correlation between the current error and the previous one. It is the fraction of the previous error term that needs to be added to the current error term.

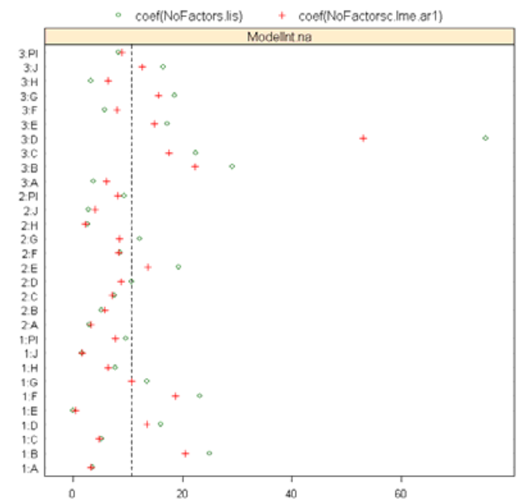

Figure 4. Plot of sniffer averages of the fixed effects model (lis) and the mixed effects model with autocorrelation (lme.ar1) showing the deviation from the population average slope (experiment mean for all sniffers and all transects).

Where model comparisons are significant between the cell means fixed effects model, fixed effects model, mixed effects models, and mixed effects model with autocorrelation, statistical tests, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select the best fitting model. In all cases the better fitting model was the more complex model. This was done as a check to ensure that significance was due to the more complex model.

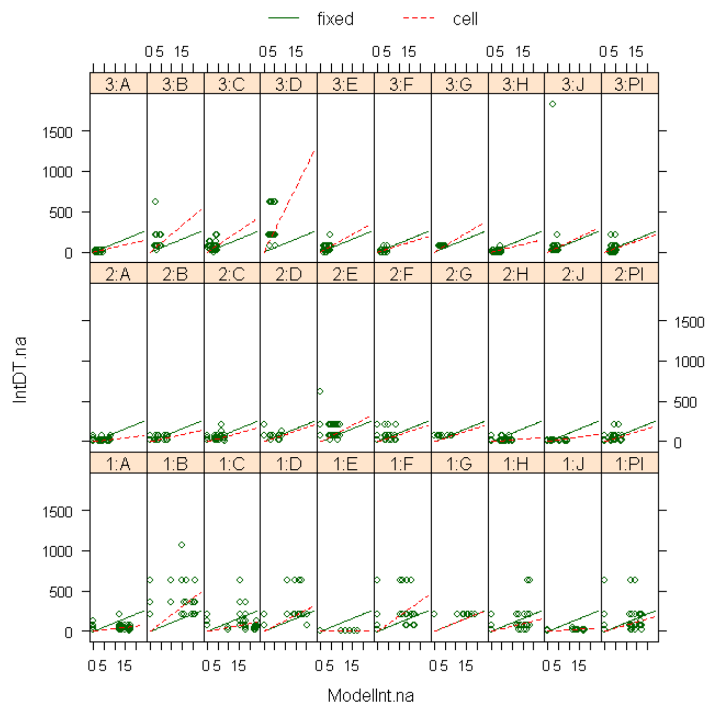

Figure 5. Individual plots by sniffer and transect of data with fixed effects model (solid line) and mixed effects model with autocorrleation (dashed line)

Shown in Figure 5 are the fixed effects model slopes (solid) and the mixed effects model with autocorrelation (dashed). Comparison of these models showed greater sensitivity of the mixed model (dashed slopes, lme.ar1) to extreme observations. Since regression slopes are the same, little predictive power is lost by combining data to estimate a single slope, resulting in one slope (scaling factor) of 6.3 being representative for the experiment (Feedlot1 PM, Intensity-predicted D/T).

Paper No. 4

Ground Truthing CALPUFF and AERMOD for Odor Dispersion from Swine Barns using Ambient Odor Assessment Techniques

C. G. Henry, P. C. D'Abreton, R. J. Ormerod, G. G. Galvin, S. J. Hoff, L. D. Jacobson, D. D. Schulte, and D. P. Billesbach

ABSTRACT. A collaborative research effort by several institutions investigated the dispersion of odors from a swine production facility. Trained human receptors measured downwind odor concentrations from four tunnel-ventilated swine finishing barns near Story City, Iowa, during twenty measurement events conducted between June and November 2004. Odor concentrations were modeled for short time steps using CALPUFF and AERMOD atmospheric dispersion models to compare predicted and measured odor levels. Source emission measurements and extensive micrometeorological data were collected along with ambient odor measurements using the Nasal Ranger® device (St. Croix Sensory, St. Paul MN), Mask Scentometer, odor intensity ratings, and air sample analysis by dynamic triangular forced-choice olfactometry (DTFCO). AERMOD predictions fit the odor measurements slightly better than CALPUFF with predicted concentrations being about half those predicted by CALPUFF. The Mask Scentometer and Nasal Ranger® measurements related best to the dispersion model output, and scaling factors of 3 are suggested for the Nasal Ranger® and 0.5 for the Mask Scentometer. Measurements obtained using the Nasal Ranger®, Mask Scentometer, and odor intensity ratings correlated well to each other, had the strongest

linear relationships, and provided slopes (measured: modeled) closest to 1.0. Odor intensity-predicted D/T did not correlate and relate as well, and this method was deemed less desirable for ambient odor assessment. Collection of ambient air samples for analysis in a DTFCO laboratory displayed poor correlations with other methods and should not be used to assess ambient odors.

INTRODUCTION

Odor issues have become a limiting factor in the viability and growth of livestock and poultry production in the United States. Odor dispersion from livestock facilities is a complicated process that depends on many factors, such as the production system, stocking density, season, localized weather patterns, terrain, and receptor locations relative to the production areas. The National Research Council (NRC, 2003) suggested that one of the two major ways to assess the effects of airborne emissions from animal feeding operations is to replace the current emission factor approach with process-based modeling. Methods and tools are needed to assist in describing the odor risk posed by new and existing facilities on the neighboring community. Such methods and processes would be valuable to the livestock and poultry industry and rural communities when siting new facilities and expanding current production facilities. There would also be benefits in the evaluation and adoption of control and mitigation strategies.

Currently, there are several models being used in the United States to evaluate odor dispersion. Researchers at the University of Minnesota used INPUFF-2 (Bee-Line Software Co, Asheville, NC), a US EPA Gaussian puff model described by Peterson and Lavdas (1986), to predict odor levels in the development of the Odor From Feedlots Setback Estimation Tool or OFFSET (Jacobson et al., 2005 and Guo et al., 2005). The

University of Nebraska-Lincoln is using AERMOD (AMS/EPA Regulatory Model) to develop an odor risk-assessment tool called the Nebraska Odor Footprint Tool or NOFT (Koppolu et al., 2004; Schulte et al., 2004; Stowell et al., 2005; and Niemeir et al., 2008). AERMOD is a Gaussian dispersion model developed in a joint effort by the American Meteorological Society and the US EPA, as the replacement to ISC3 (Industrial Source Complex). Iowa State University has also developed its own model, called CAM (Community Assessment Model) for predicting odor dispersions in a community (Hoff and Bundy, 2003).

The objectives of the work reported in this paper are to 1) develop model-specific scaling factors for CALPUFF and AERMOD associated with measurements made using the following ambient odor assessment techniques: Nasal Ranger®, Mask Scentometer, intensity ratings, and laboratory analysis of collected air samples; and 2) compare the performance of CALPUFF and AERMOD for odor prediction. An underlying goal of this work was to find the model and ambient odor assessment technique combination that gave the best agreement between predicted and observed concentrations.

PREVIOUS WORK

In agriculture, the primary driver behind ground-truthing odor models with field observations has been the development of "simple tools" that can be used to quickly and inexpensively assess the odor risk presented by a proposed facility on neighbors. Such tools are very useful for planning facilities and screening prospective sites, and can be easily used by livestock producers, planning and regulatory officials, and the general public to envision the odor risk of livestock facilities. Historically, such tools were based solely on model results, but more recent work has incorporated observations of trained assessors to calibrate the models by scaling model results.

Traditionally, agreement between predicted and observed odor measurements have been achieved using peak-to-mean scaling. The first references in the literature of peak-to-mean ratios were by Högström (1972) and Smith (1973) where, in general, the ratio of the peak measured concentration to the mean predicted concentration equals the response time of the human nose (1-5 seconds) divided by the modeling time, typically 60 minutes, raised to a power n, as shown in the following equation:

$$(C_{peak}/C_{mean}) = (t_{peak}/t_{mean})^n$$

Peak-to-mean ratios are essentially a correction factor (or scaling factor) used to match dispersion model results with observations. This is done because of the short-term fluctuations observed in ambient atmospheric odor concentrations due to transport processes. Most dispersion models predict concentrations based on historical meteorological observations that are readily available, where the shortest time step is usually one hour. However, peak values, defined by Wilson (1996) as "the concentration that is exceeded m times in a statistically independent ensemble of n repeats of an event," are generally much higher than the average concentration. Several peak to mean adjustment methodologies have been developed and much of the work and uncertainty centers around the power term n, which ranges between 0.167 and 0.65 (Pope and Diosey, 2000, Katestone Scientific, 1995; Mejer and Krause, 1986; Mahin, 1998; Best, 2000; Schaugerger et al., 2000 and 2001, Duffee et al., 1991) with the typical range being between 0.2 and 0.4.

Plume dispersion modeling has undergone significant refinement in recent years. Steady-state Gaussian plume air dispersion models such as Industrial Source Complex version 3 (ISC3), which formed the basis of air dispersion modeling, are now being replaced by a new generation of models, most notably CALPUFF and AERMOD. These new models incorporate many additional algorithms addressing influences that are ignored by the steady-state Gaussian models, but are known to significantly influence plume dispersion.

The US EPA regulatory model, CALPUFF, is a multi-layer, multi-species, non-steady-state puff dispersion model that can simulate the effects of temporal and spatial variability of micrometeorological conditions on pollutant transport, transformation and removal (Scire et al., 2000). The model contains algorithms for near-source effects such as building downwash, partial plume penetration, sub-grid scale interactions, as well as longer-range effects such as pollutant removal, chemical transformation, vertical wind shear, and coastal interaction effects. The model employs dispersion equations based on a Gaussian distribution of pollutants across the puff and takes into account the complex arrangement of emissions from point, area, volume, and line sources.

These models generally depend on high-resolution micrometeorological data for best performance. The averaging times for the odor observations must be matched to the model's temporal resolution. Normally, CALPUFF calculates hourly average concentrations that are based on hourly average micrometeorological data. However, Pacific Air and Environment, Brisbane, Australia, coded a short-time-step version of CALPUFF to deal with situations where model output could be obtained at shorter time steps. Since this work, versions of CALPUFF have been released that permit sub-hourly time-steps (D'Abreton, 2006). The time-step may be reduced to one minute, which provides very good resolution. This short-time-step version of the model is available for the quasi-three-dimensional mode operating with measured turbulence parameters.

In 1991, the American Meteorological Society (AMS) and the US Environmental Protection Agency (US EPA) initiated a formal collaboration with the goal of introducing planetary boundary layer (PBL) concepts into regulatory models. The result of this work is AERMOD (AMS/EPA Regulatory Model), which was intended to replace the Industrial Source Complex Model systems. AERMOD is a steady-state plume model that calculates concentrations at hourly time steps based on temporally averaged micrometeorological inputs. In the stable boundary layer (SBL), AERMOD assumes the concentration distribution to be Gaussian in both the vertical and horizontal planes. In the convective boundary layer (CBL), the horizontal distribution is also assumed to be Gaussian, but the vertical distribution is described with a bi-Gaussian probability density function (pdf). Dispersion coefficients used to calculate concentrations in both the SBL and CBL are the summation of the lateral and vertical concentration distributions (dispersion coefficients) and ambient turbulence and dispersion from plume buoyancy. The improvements of AERMOD over ISC3 are summarized in US EPA (2004) and include; 1) dispersion in both the convective and stable boundary layers; 2) plume rise and buoyancy; 3) plume penetration into elevated inversions; 4) computation of vertical profiles of wind, turbulence, and temperature; 5) treatment of receptors on all types of terrain from the surface to and above the plume height; 6) inclusion of building wake effects; 7) improved characterization of the fundamental boundary layer parameters, and 8) the treatment of plume meander.

Curran et al., (2002) used US EPA's ISC3 and the UK's ADSM3 (Atmospheric Dispersion Modeling System) in Ireland to demonstrate that setback distances for swine facilities are more appropriately established using models rather than absolute values. In 2007, Curran et al. compared ISC3 and CALPUFF for odors. They measured the downwind odor intensity from a 514 head sow operation. They used olfactometry and measured ventilation rates to determine odor emission rates and treated the facilities as point sources. Their field odor assessments consisted of odor intensity ratings using a 1-butanol reference scale and VDI 3940 as a guideline. They found that the average predicted-to-measured mean concentration ratio on the sampling days varied from 1.4 to 9.37. They also found that over 80% of the model predictions were greater than their field observations and concluded that both models gave conservative estimates of downwind odor concentration.

Diosey et al. (2002) compared AERMOD predictions to hydrogen sulfide measurements simulating hydrogen sulfide stack emissions from wastewater treatment plants and found the model predictions to be one and two orders of magnitude less than those of ISCST3 and CALPUFF, respectfully.

A research team from the University of Minnesota measured emissions from 280 animal buildings and manure storages. This team then used trained ambient odor assessors stationed in transects within odor plumes (to measure odor intensity using a 0-5, 1-butanol reference scale) and on-site micrometeorological data to ground-truth INPUFF-2 during the development of a simple tool they called OFFSET. They successfully developed this tool, which provides a separation distance based on a selected level of odor risk and facility information (Guo et al., 2005; Jacobson et al., 2000; and Jacobson et al., 2005). The foundation for this work was laid by Zhu et al., (2000) and Guo et al., (2001), which showed that the model predictions of INPUFF-2 were always lower than the observed odor concentrations, leading to scaling factors of 10 for manure storages and 35 for building sources.

Li and Guo (2006) utilized a Computational Fluid Dynamics (CFD) Model to study the effect of time step on model predictions for a 3000-sow swine farm. They found that the longer the time step (1 hr) the higher the odor concentration prediction and the longer the travel distance, and the shorter the time step (1 minute was the shortest they used) the lower the concentration and shorter the travel distance.

Xing et al. (2006) did a comparison of ISCST3, AUSPLUME, CALPUFF, and INPUFF-2 using ambient odor assessors trained to report odor intensity (on a 0-8, 1-butanol reference scale) from a series of swine buildings with earthen storages in Canada. In their work they compared two different equations to predict odor concentrations from intensity, with one equation from Zhang et al. (2005) and the other from Segura and Feddes (2005). They found that the equation selected was very important in model performance and produced mixed results with agreement ranging from 13% to 76% depending on the model, experiment and equation used. They found scaling factors for the models to be in the range of 1.2 to 7.9, and concluded that none of the models studied were obviously better than another.

In previous analyses of data from the monitoring project described in this paper, Modi (2006) and Schulte et al. (2007) compared AERMOD concentration predictions to Nasal Ranger® observations and found overall scaling factors of 1.66, with the model under-predicting the Nasal Ranger® observations. However, these results were only from a subset of ten experiments. This work includes an additional ten experiments (for a total of 20). D'Abreton et al. (2007) modeled downwind odor concentrations for this experiment using the CALPUFF modeling system and found that the 1-minute time step version of CALPUFF was able to mimic the variable nature of odors and that the Nasal Ranger® observations were within the range of model predictions 44% of the time. Henry et al. (2007) analyzed these results further and found a scaling factor of 0.99, implying that a scaling factor may not be needed when CALPUFF predictions are intended to match Nasal Ranger® observations. This paper is a continuation and extension of the work of D'Abreton et al., (2007); Henry et al., (2007); Modi (2006); Schulte et al., (2007).

METHODOLOGY

Trained human receptors measured odor concentrations downwind of four tunnel-ventilated swine finishing barns near Story City, Iowa, during twenty-six measurement events conducted between June and November 2004. Due to micrometeorological equipment failure, data from six of these events were not used, leaving twenty usable events. For each of the twenty 15-minute measurement events, receptors measured odor levels at four locations in the plume, resulting in 80 observations (n=80) for which model predictions and odor measurements could be compared over the same time period and place.

AMBIENT ODOR ASSESSMENT METHODS

At each of the four downwind measurement locations, ambient odor was assessed using the following techniques:

Nasal Ranger®. Two assessors from Iowa State University trained by St. Croix used the Nasal Ranger® field olfactometer (www.nasalranger.com). Unit setting numbers that corresponded to dilution-to-threshold readings were made twice during each 15-minute assessment period, once at the beginning of the session and the second time at the 7.5-minute mark. The unit settings for the Nasal Ranger® were 0 (no detect), 2, 4, 7, 15, 30, and 60 D/T. The average of each set of four D/T readings made (two assessors x 2 readings) was used in the analysis (n = 80).

Mask Scentometer. During several but not all measurement events, one or two assessors trained by the University of Nebraska used the Mask Scentometer – as described by Sheffield et al. (2004) and Henry (2004). These assessors recorded unit setting numbers that corresponded to dilution-to-threshold readings every 30 seconds during each 15-minute measurement event, for a total of 30 D/T readings per assessor per event. The unit settings used in this work for the Mask Scentometer were 0 (no detect), 0.35, 1, 2, 4.5 and 18 D/T, and the average of the 30 individual D/T readings was used in the analysis. If there was more than one assessor at the receptor location then the arithmetic average of their results was used in the analysis (n = 55).

Odor Intensity Rating Scale (OIRS). Two assessors trained by the University of Minnesota rated odor intensity 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 concentrations in air with a geometric progression of three, with 25 ppm representing I=1 and 2,025 ppm representing I=5. This is the same technique used by Jacobson et al., (2000); Jacobson et al., (2003); Nicolai et al., (2000); Zhu et al., (2000). During each 15-minute event, each assessor rated odor intensity 60 times (every 15 seconds). Average intensities for both assessors during the event were averaged and this intensity value was used in the analysis (n=80).

In the data analyses, comparisons were also made using odor concentration values that were predicted from the average intensities. For clarity, these empirically derived odor concentration values are referred to as ‘Average intensity-predicted D/T’ in this paper. Jacobson et al. (2000) published a relationship between intensity and dilutions to threshold (D/T) as determined from the analysis of odors using a laboratory olfactometer. For swine odors, they used the relationship $D/T_{swine} = 8.367\ e^{1.0781I}$ to obtain D/T values from intensities. Jacobson et al., (2003); Nicolai et al., (2000); Zhu et al., (2000) also used this prediction equation. In the current work, a D/T value was predicted for each measurement event using this equation and the average intensity for the event (n=80).

DTFCO. Ambient air samples were collected in the field for subsequent odor analysis using dynamic triangular forced-choice olfactometry (DTFCO). One air sample was collected in a new unflushed Tedlar 10 L bag during the first four minutes of each 15-minute measurement event. The Iowa State University Odor Lab analyzed odor samples following the ASTM Standard E679-97, “Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits”. The lab was in compliance with the European Standard for olfactometry (CEN, 2003). All samples were analyzed within 24 hours of collection (n=80).

All assessors recorded readings on pre-printed data sheets. Field samples and measurement times were synchronized with the on-site micrometeorological station clock so measurement intervals would correspond to modeled time steps. One challenge with analyzing scentometer data arises from the crisp, nonlinear unit D/T settings. When an odor concentration of 2 D/T is reported by an assessor using a device having unit settings of 2 and 7 D/T, it is very likely that the actual concentration was somewhere between 2 and 7 D/T – so, actual odor concentrations equating to 3 and 6 D/T both would be reported as 2 D/T, which leads to results being skewed downward. To account for this undesired influence, dilution-to-threshold data for the Mask Scentometer and Nasal Ranger® were adjusted using Equation 1 (Sheffield et al., 2004) to obtain geometric average dilution-to-threshold $(D/T)_G$ readings.

Equation 1. $$D/T_{G,n} = 10^{\frac{\log D/T_n + \log D/T_{(n+1)}}{2}}$$

Where n is the device setting number reported by an assessor for a reading, and D/T_n is the D/T specified for that setting (referred to as the 'unit D/T'). For example, for a reported Nasal Ranger® setting of three (unit D/T of 4 and next higher unit D/T of 7), the $(D/T)_G$ is 5.3. The $(D/T)_G$ used in representing field olfactometer readings are shown in Table 1. Another issue with data from field olfactometers is how to deal with non-detects. The fact that odor was not detected at the lowest device setting does not mean no odor existed in the ambient air. Also, it is not possible to take the geometric average of results that include zeros. For non-detect readings, the $(D/T)_G$ was assumed to be about two-thirds between zero and the first unit D/T on the device (0.2 for the Mask Scentometer and 1.4 for the Nasal Ranger®). No attempt was made to adjust the D/T for the last settings of the devices (18 D/T for the Mask Scentometer and 60 D/T for the Nasal Ranger®) since they are the limits of the instruments (their $(D/T)_G$ could be anywhere between 18/60 and infinity). Using $(D/T)_G$ data increases the odor concentration for an assessment compared to using the unit D/T, which has a justifiable basis. The main drawback is that when no odor is actually present, a small odor level is reported (i.e. $(D/T)_G$ cannot be less than 0.35 or 1.4). In this situation, it was assumed that an odor was present during the assessments, so there should be little effect of assigning a $(D/T)_G > 0$ for non-detects on the overall results.

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

Mask Scentometer		Setting	Nasal Ranger®	
Unit D/T	Geometric (D/T)	n	Unit D/T	Geometric (D/T)
na	na	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 / No detect	0.2	1	0 / No detect	1.4

Modeling Methodology

The AERMOD and CALPUFF models require meteorological data, source emission rate data, facility layout and dimensions and receptor location information. AERMOD is designed to accept hourly micrometeorological data. However, data with a 1-minute time step can be input as hourly data to produce 1-minute predictions, without any model adjustment, when corresponding micrometeorological data are available (Modi, 2006). The meteorological data were averaged every minute and this data was used in the modeling. The authors recognize that representing atmospheric conditions on a 1-minute time step will not be representative of transport processes over the entire experimental footprint. Rather, the conditions will be representative of localized areas near the instrument tower. This will introduce variability into the analyses. However, given the relatively close proximity of the source emissions and field assessors, the authors decided that this was preferable to assuming one atmospheric condition during each assessment period. Previous work at the University of Nebraska by Koppolu et al., (2002) compared AERMOD and STINK where dispersion of volatile fatty acids (VFA's) from a one-meter-diameter source was measured downwind with thermal desorption tubes and SPME fibers. They evaluated meteorological averaging times of 1, 5, 15, and 30 minutes and found, so long as extreme variations in the wind speed and direction did not occur, shorter periods had better agreement with model predictions. The experiments, however, were conducted over much shorter receptor distances, where the effects of short averaging time were minimized. They also found that AERMOD and STINK both gave comparable predictions of VFA. Other researchers have used similar approaches (Li and Guo, 2006; Zhu et al., 2000a; Zhou et al., 2005; Xing et al., 2006).

METEROLOGICAL DATA

For this project, extensive meteorological data were collected from two on-site weather stations. A 10-meter tower measured temperature, relative humidity, wind direction and speed (cup anemometer), and net solar radiation over one-minute intervals. The tower was located 5 meters east of the eastern edge of the barns, between barns 2 and 3 (Figure 1). A sonic anemometer on a 3.8-meter tower measured wind speed and direction at 10 Hz. This facilitated calculation of turbulence parameters (σ_u, σ_v and σ_w) and other derived variables such as sensible heat flux, friction velocity (u*), roughness length (z_o), and temperature (T*). In addition, relative humidity (capacitance sensor), temperature (platinum resistance thermometer), short-wave solar radiation (upwelling and down-welling pyronometers), and soil heat flux were measured over one-minute intervals. This second meteorological station was located approximately 80 m (260’) to the northeast of barn 4 (Figure 1).

SWINE FACILITY

The production site was located on flat agricultural terrain and consisted of four swine finishing barns as shown in Figure 1. Each barn housed 950 head of finishing pigs and was 13.7 m wide by 61 meters long, with 19.8 meters between buildings. Internal temperature and indoor air quality were maintained by a combination of ventilation modes:

Pit fans - air was drawn through vents located on the sidewalls down both sides of the room and expelled from two fans that were located on the south side of each barn. The total pit fan airflow rate for each barn was approximately 7,000 m^3/hour.

Tunnel fans - air was primarily drawn through large inlet openings located on the west endwall of each building – opposite the fans – at a total airflow rate of between 98,000 m^3/hour and 15,000 m^3/hour. Each barn had five such fans located on the eastern side.

Receptors were located either in a diamond pattern or a longitudinal transect within the odor plume. The receptor locations were modeled as a “localized grid” of receptors in a 4-meter grid. One-minute model predictions were averaged over the 15-minute assessment periods for comparison with observations.

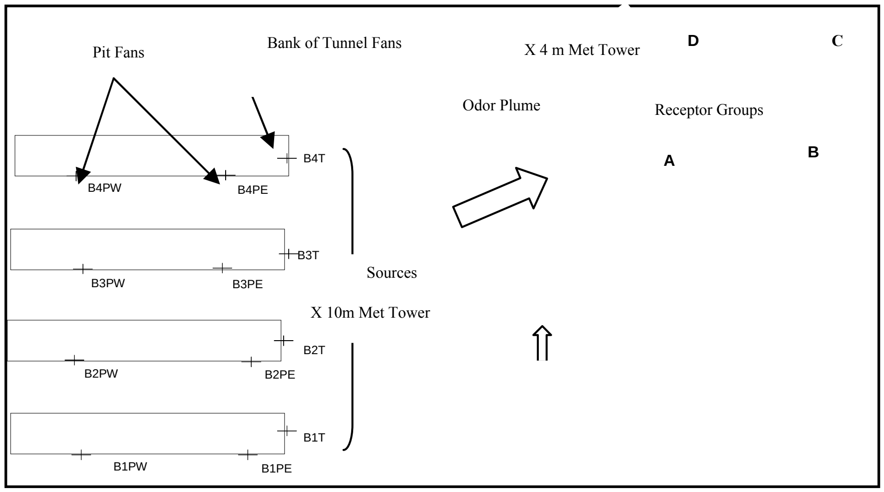

Figure 1. Site layout showing locations of the swine barns, pit fans (e.g. B4PW), tunnel fans (e.g. B1T represents a group of 5 fans in Barn 1), and the meteorological stations. Receptor locations are also shown for measurement events 1-3.

EMISSIONS DATA

Source emission rates were developed at the time of the field assessments (and modeling periods) from well-instrumented buildings on the site. A mobile emission-monitoring laboratory recorded fan data that was used to calculate ventilation rates (USDA, 2001). Barns 2 and 3 were instrumented to send data to the mobile laboratory. Exhaust air was collected twice from the fans during each of the 15-minute measurement periods (once at the beginning of the measurement event, and again at the 7.5-minute mark of the event). The sample bags were transported to the Iowa State University olfactometry lab within 24 hours of the event and analyzed for dilution to threshold using a venturi-type dynamic dilution olfactometer (AC'SCENT® International Olfactometer, St. Croix Sensory, Inc. Stillwater, MN). Odor detection threshold, defined as the concentration that the panelists first detect a difference between the sampled air and two clean air streams, was measured in accordance with ASTM standard E679-91 using trained panelists.

Emissions from the exhaust fans were directed horizontally, with no significant vertical velocity component. Temperature differences between the barns and the 4-meter tower (ambient) existed, with barn temperatures being up to 1°C warmer than corresponding ambient temperatures during the measurement events. Atmospheric temperature differentials, as small as 1-2°C, may have a significant effect on near-field ground-level concentrations (Ormerod et al., 2003). Therefore, any plume rise occurred as a result of thermal buoyancy of the emissions, with no plume rise component due to vertical momentum. To deal with the effects of buoyancy in this study, the horizontally directed fans were modeled as four pseudo-point sources to negate mechanically generated plume rise while maintaining the thermal mass of the plume. To achieve this, vertical efflux velocity was set at the recommended value of 0.01 m/s (e.g. NCDENR, 2003; NDEQ, 2001; NMAQB, 2003).

STATISTICAL ANALYIS METHODOLOGY

The data from the ambient odor assessments and model predictions were analyzed using bias and error analysis. Initial investigation of the data indicated that it was linear in nature. Additionally, linear regression was used to develop scaling factors, between the models and respective ambient odor assessment methods. By using both analysis techniques, the best paired ambient odor assessment technique and dispersion model combination can be established. That is the combination with the lowest error and bias, and the combination with the best fitting scaling factor. The slope for regression, or scaling factor, would be a slope nearest 1.0 and 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. Using this two tiered approach should provide the most reliable technique to use with CALPUFF and AERMOD with the smallest best fitting scaling factor and lowest bias and error.

As a separate analysis, odor methods were compared among themselves using linear regression to develop scaling factors between the different methods. In this analysis, the best fitting slope or scaling factor between methods was sought. The ultimate goal of which was to find the methods that were the most comparable to each other, or equivalent in their assessment of D/T, and if not comparable, how much scaling would be necessary to relate one method to the other (i.e. a Mask Scentometer D/T to a DTFCO lab D/T).

The Fractional Bias Test and Model Bias Test are commonly used to compare model predictions with observations (ASTM, 2000; Wilmott, 1981; and Pielke, 1984). The Fractional Bias Test is symmetrical and bounded by values between 2 and -2. A value of zero indicates no bias. Model bias is another measure of gross error resulting from both bias and scatter, and the ideal value is also zero. The root mean squared error (RMSE) and the normalized mean squared error (NMSE) are typically used for statistically assessing the performance of atmospheric dispersion models (ASTM, 2005). The advantage of NMSE and RMSE is that the normalization allows comparisons between experiments with different average values. In the equations shown in Figure 2, O_i represents the observed values and P_i represents the corresponding predicted values. The subscript i represents paired values and the over bar indicates an average.

$$FB = \overline{FB_i} \quad where \quad FB_i = \frac{2 \cdot |P_i - O_i|}{P_i + O_i} \qquad RMSE = \sqrt{\overline{(P_i - O_i)^2}}$$

$$MB = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i) \qquad NMSE = \frac{\overline{(P_i - O_i)^2}}{\overline{PO}}$$

Figure 2. Equations for Root Mean Square Error (RMSE), Normalized Mean Square Error (NMSE), Fractional Bias (FB), and Model Bias (MB).

The mean results for each receptor location (localized grid) were analyzed in the R statistical package (R Development Core Team, 2008) using the "lm" (linear model) command and regression through the origin.

RESULTS AND DISCUSSION

MODEL BIAS AND ERROR ANALYSIS

Model performance statistics are shown in Table 2 and Table 3 for CALPUFF and AERMOD, respectively, along with general measurement and modeling results. The results shown for the field olfactometers are based upon unit D/T values, but the analyses were also done for $(D/T)_G$ (not shown) and produced the same results. For the Nasal Ranger®, the use of $(D/T)_G$ moved the statistics away from the ideal value; however, for the Mask Scentometer statistics improved slightly. Model Bias, Fractional Bias, NMSE, and RMSE using $(D/T)_G$ for the Nasal Ranger® and Mask Scentometer were 8.4, 0.8, 2.7 and 1.8 for the Nasal Ranger® and -5.8, -0.5, 2.8, and 1.2 for the Mask Scentometer, respectively. From the performance statistics, shown in Table 2 and

Table 3, several trends emerge. First, for CALPUFF, all of the methods have fractional bias within the ideal range. For Model Bias, the method with the smallest Model Bias are Intensity (-4.9), followed by the Mask Scentometer (-5.0), the Nasal Ranger® (5.9), Average intensity-predicted D/T (45.2), and DTFCO (60.4). A similar result was found when comparing the odor assessment methods to AERMOD for bias, that is all methods have fractional bias within the ideal range. For model bias, the method with the smallest bias is the Nasal Ranger® (4.1), Intensity (-6.9), Mask Scentometer (-7.1), Average-intensity-predicted D/T (43.2), and DTFCO (58.4).

Table 2. Model performance statistics for CALPUFF contrasted against ambient odor assessment method.

Statistical Test	Nasal Ranger® D/T	Mask Scentometer D/T	Intensity	Average intensity-predicted D/T	DTFCO Lab D/T	Ideal value is…
Mean observed measure (units vary by method)	12.2	2.3	1.3	51.4	95.2	
Mean modeled D/T	6.4	6.4	6.4	6.4	6.4	
Standard deviation of observed values (units vary by method)	12.1	1.7	1.7	54.4	70.5	
Standard deviation of modeled D/T	7.7	7.7	7.7	7.7	7.7	
Model bias (MB)	6.1	-5.0	-4.9	45.2	60.4	0
Fractional bias (FB)	0.64	-0.9	-1.3	1.6	1.7	-2 to 2
Normalized mean square error (NMSE)	2.6	5.7	9.5	15.1	15.3	0
Root mean square error (RMSE)	1.6	1.0	1.0	7.8	10.7	0

Table 3. Model performance statistics for AERMOD contrasted against ambient odor assessment method.

Statistical Test	Nasal Ranger® D/T	Mask Scentometer D/T	Intensity	Average intensity-predicted D/T	DTFCO Lab D/T	Ideal value is…
Mean observed measure (units vary by method)	12.3	2.3	1.3	51.4	95.2	
Mean modeled D/T	8.2	8.2	8.2	8.2	8.2	
Standard deviation of observed values (units vary by method)	12.1	1.7	1.7	54.4	70.5	
Standard deviation of modeled D/T	9.2	9.2	9.2	9.2	9.2	
Model bias (MB)	4.1	-7.1	-6.9	43.2	58.4	0
Fractional bias (FB)	0.4	-1.1	-1.4	1.4	1.7	-2 to 2
Normalized mean square error	1.8	6.6	11.6	11.2	10.9	0
Root mean square error	1.5	1.2	1.3	7.7	10.3	0

For CALPUFF, the methods with the smallest error (NMSE) in descending order are the Nasal Ranger® (2.6), Mask Scentometer (5.7), intensity (9.5), Average intensity-predicted D/T (15.1), and DTFCO (15.3). For RMSE, the methods with the smallest error are in descending order: intensity (1.0), the Mask Scentometer (1.0), the Nasal Ranger® (1.6), Average intensity-predicted D/T (7.8), and DTFCO (10.7). For AERMOD, the methods with the smallest error for NMSE are, in descending order: the Nasal Ranger® (1.8), Mask Scentometer (6.6), DTFCO (10.9), Average intensity-predicted D/T (11.2), and intensity (11.6). For RMSE, the methods with the smallest error are, in descending order: the Mask Scentometer (1.2), intensity (1.3), the Nasal Ranger® (1.5), Average intensity-predicted D/T (7.7), and DTFCO (10.3). In general, the methods found to consistently have the lowest bias and error were the Nasal Ranger®, intensity, and Mask Scentometer.

REGRESSION RESULTS

Linear regression analysis results (slopes, coefficients of determination, and standard errors) are shown in Table 4. The slopes represent scaling factors needed to relate values obtained from the various models and odor assessment methods to each other. Traditionally in linear regression analysis, one variable is the independent variable or predictor (x) and a relationship can be found for the dependent variable or response (y). One of the underlying assumptions is that the regressors (x_i) 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 4, 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 (see example shown in Appendix, Paper No. 2). The slope with a "*" produced the lowest error and is the more precise relationship. Note that the coefficients of determination (R_o^2) are the same for each of the linear models.

From Table 4 one can relate one method to another and assess the scale difference 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 (0.20), so Nasal Ranger® $(D/T)_G$ readings were about 5 times higher than Mask Scentometer $(D/T)_G$. Meanwhile, the slope of Average intensity-predicted D/T (y) as a function of Nasal Ranger® $(D/T)_G$ (x) was 2.7, however, the Nasal Ranger® $(D/T)_G$ (y) as a function of Average intensity-predicted D/T (x) resulted in a slope of 0.19 and had a lower error (*). Therefore, to relate an assessment made with Average intensity-predicted D/T (y) to an assessment made with a Nasal Ranger® $(D/T)_G$, we would select the stronger relationship, which is 1 Average intensity-predicted D/T = 5.2 Nasal Ranger® $(D/T)_G$ (1/0.19 = 5.2).

Table 4. Regression Results: Slope or Scaling Factor (top), Coefficient of Determination, R_o^2 (middle), and Standard Error (bottom).

(Response) Y► (Predictor) X▼	CALPUFF model D/T	AERMOD model D/T	Nasal Ranger® D/T	Nasal Ranger® $(D/T)_G$	Intensity	Average Intensity-predicted D/T	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$	DTFCO Lab D/T
CALPUFF model D/T		0.80 0.42 0.11	0.99 0.33 0.2	1.19 0.37 0.2	0.09* 0.35 0.01	3.79 0.25 0.07	0.19* 0.47 0.03	0.39* 0.51 0.06	9.06 0.29 1.9
AERMOD model D/T	0.52* 0.42 0.07		0.87 0.38 0.1	1.06 0.44 0.1	0.09* 0.46 0.01	3.32 0.30 0.6	0.14* 0.55 0.02	0.27* 0.52 0.04	5.37 0.39 0.9
Nasal Ranger® D/T	0.33* 0.33 0.05	0.44* 0.38 0.06			0.07* 0.63 0.006	3.30 0.58 0.3	0.12* 0.73 0.01	0.20* 0.49 0.03	4.9 0.43 0.8
Nasal Ranger® $(D/T)_G$	0.31* 0.37 0.05	0.41* 0.44 0.05			0.06* 0.64 0.005	2.73 0.52 0.3	0.11* 0.73 0.01	0.19* 0.55 0.03	4.13 0.45 0.06
Intensity	3.8 0.35 0.60	5.4 0.46 0.7	8.6 0.63 0.8	10.1 0.64 0.9		44.0 0.83 2.3	1.40 0.76 0.1	2.56 0.63 0.3	57.5 0.55 6.9
Average Intensity-predicted D/T	0.07* 0.26 0.10	0.09* 0.30 0.02	0.17* 0.58 0.02	0.19* 0.52 0.02			0.02* 0.64 0.003	0.04* 0.35 0.008	1.28 0.46 0.2
Mask Scentometer D/T	2.5 0.47 0.43	3.87 0.55 0.56	6.1 0.73 0.6	6.6 0.73 0.6	0.54* 0.76 0.05	26.2 0.64 3.1			29.6 0.47 5.9
Mask Scentometer $(D/T)_G$	1.29 0.51 0.21	1.88 0.52 0.29	2.50 0.49 0.4	2.87 0.55 0.4	0.25* 0.63 0.3	9.7 0.35 2.0			15.45 0.47 3.0
DTFCO Lab D/T	0.03* 0.29 0.006	0.07* 0.39 0.01	0.09* 0.43 0.01	0.11* 0.45 0.02	0.009* 0.55 0.001	0.36* 0.46 0.05	0.02* 0.47 0.003	0.03* 0.47 0.006	

** Indicates more precise relationship based on having the lowest standard error.*
To scale a Nasal Ranger D/T to a Mask Scentometer $(D/T)_G$, multiply the Nasal Ranger value times 0.20 (i.e. 1 Nasal Ranger D/T = 0.20 Mask Scentometer $(D/T)_G$).
To scale methods using scaling factors in the light grey boxes, use the inverse slope, for example to relate a Nasal Ranger $(D/T)_G$ to an intensity-predicted D/T, the stronger relationship is 0.19 (rather than 2.7), so multiply the Nasal Ranger value times 5.2 (1/0.19 = 5.2) (i.e. 1 Nasal Ranger $(D/T)_G$ = 5.2 intensity-predicted D/T).

CALPUFF

Shown in Table 4, slopes between CALPUFF and AERMOD results were found to be 0.80 and 0.52, with a slope of 0.52 being the stronger relationship (higher R_o^2). This means that a CALPUFF prediction is about twice that of an AERMOD prediction. The ambient odor assessment method showing the strongest relationship with CALPUFF was the Mask Scentometer (R_o^2 = 0.47 for D/T and R_o^2 = 0.51 for $(D/T)_G$). The next strongest R_o^2 was for the Nasal Ranger® $(D/T)_G$ (R_o^2 = 0.37), Intensity rating (R_o^2 = 0.35) and Nasal Ranger® D/T (R_o^2 = 0.33). Both Average intensity-predicted D/T and DTFCO had the lowest coefficients of determination (R_o^2 = 0.25 and R_o^2 = 0.29). The methods with the slopes nearest one, using the lowest standard error, were Nasal Ranger® $(D/T)_G$ (slope=1/0.37=2.7), the Mask Scentometer $(D/T)_G$ (slope=0.39), Nasal Ranger® (slope=1/0.31=3.2), Mask Scentometer D/T (slope=0.19), DTFCO (slope=9.0), Intensity rating (slope=1/0.07=14), and Average intensity-predicted D/T (slope=14.2). These slopes represent the odor assessment method dependent scaling factors for CALPUFF. Based on this dataset and using the R_o^2 and a slope nearest one representing the best relationship as criteria, both the Nasal Ranger and Mask Scentometer appear to be best matched to CALPUFF predictions. From the entire dataset, the Nasal Ranger has a slightly better scaling factor (2.7-3.2), but the Mask Scentometer's scaling factors of 0.19 and 0.39 (slope) are a better fit (R_o^2=0.47-0.51).

It should be noted that the Mask Scentometer data set is much smaller (only half the number of observations) and that the range of the instrument is limited (max assessment of "18 D/T"). The effect of this limitation was studied further, and CALPUFF predictions that were higher than the theoretical limit of 18 Odor Units (assumed equivalent to a D/T) were removed from the dataset (deleted 22 observations with 23 observations remaining). When these high model predictions were removed, which were primarily the closest locations to the barn where the highest odor concentrations were experienced, the slopes improved from 0.19 to 0.28 for Mask Scentometer D/T and 0.39 to 0.52 for the Mask Scentometer $(D/T)_G$ and a modest reduction in the R_o^2 as shown in

Table 5. With the high model predictions removed, the Mask Scentometer $(D/T)_G$ has the slope closest to one (0.52) and a coefficient of determination (0.42) better than the Nasal Ranger. While likely not the entire reason, the range limitation of the Mask Scentometer was likely a factor in the results.

Table 5. Slope (top), Coefficient of determination R_o^2, and standard error for CALPUFF predictions less than 18 OU.

	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$
CALPUFF	0.28* 0.47 0.05	0.52* 0.42 0.1
AERMOD	0.22* 0.44 0.04	0.54* 0.56 0.08

** Indicates the lowest standard error, where Mask Scentometer measurements are y, and model predictions are the x variables.*

When $(D/T)_G$ is used instead of D/T for the Nasal Ranger® and Mask Scentometer, the slopes for the Nasal Ranger® modestly decreased from 0.33 to 0.31 yet for the Mask Scentometer improved from 0.19 to 0.39 (or 0.28 to 0.52 when high model predictions were removed). In both cases R_o^2 improved slightly. This showed that $(D/T)_G$ provides a better result for the Mask Scentometer, but is not necessarily helpful in improving results for the Nasal Ranger®. In summary, the ambient odor assessment methods showing the best relationship to the CALPUFF model predictions were the Nasal Ranger® and Mask Scentometer. Curran et al., (2007) found the average predicted to measured mean concentration ratio on the sampling days to vary from 1.4 to 9.37 (which would relate to model scaling factors of 0.7 and 0.10), for an OIRS that included a 1-butanol scale and using VDI 3940 as a guideline when comparing ISC3 and CALPUFF for odors. Li and Guo (2006) found scaling factors from 1.2-7.9 using CALPUFF and an OIRS (0-8 scale). Zhu et al., (2001) found scaling factors of 10 for manure storages and 35 for buildings for

INPUFF-2 using the same OIRS (0-5 scale) as this study. In this study, the most comparable method, Average intensity-predicted D/T, produced a scaling factor of 14 for CALPUFF.

AERMOD

The method with the strongest R_o^2 is the Mask Scentometer (0.55 for D/T and 0.52 for $(D/T)_G$). The next best R_o^2 was Intensity Rating (0.46), the Nasal Ranger® $(D/T)_G$ (0.44), DTFCO lab D/T (0.39), Nasal Ranger® D/T (0.38), and Average-intensity-predicted D/T (0.30). The methods with the slopes nearest one, using the more precise relationships(standard error), were Nasal Ranger® D/T (1/0.44=2.3), Nasal Ranger® $(D/T)_G$ (1/0.41=2.4), the Mask Scentometer $(D/T)_G$ (0.27), DTFCO (5.3), Mask Scentometer D/T (0.14), and Intensity rating and Average-intensity-predicted D/T both the same (1/0.09=11.1). These slopes represent the odor assessment method dependent scaling factors for AERMOD. Based on this dataset and using the R_o^2 and a slope nearest one representing the best relationship as criteria, both the Nasal Ranger and Mask Scentometer appear to be best matched to AERMOD predictions because they have the strongest R_o^2 and slopes closest to one. From the entire dataset, the Nasal Ranger has a slightly better scaling factor (2.3-2.4), but the Mask Scentometer's scaling factor (slope) is a better fit (R_o^2=0.52-0.55). Again, like the CALPUFF results, $(D/T)_G$ related better to the AERMOD predictions for the Mask Scentometer, but does not appear to have any substantial effect for the Nasal Ranger®. Clearly, the intensity rating, DTFCO Lab D/T and the Average-intensity-predicted D/T do not relate as well with AERMOD predictions when using R_o^2 and nearness to a 1:1 slope as criteria.

Similar to CALPUFF, when the high AERMOD predictions greater than 18 OU (assumed equivalent to a D/T), were removed from the dataset, the slopes improved as shown in Table 5. The more precise model as determined by the lowest standard error was by using the model prediction as the independent variable and the Mask Scentometer measurement as the dependent variable. A modest reduction in the R_o^2 and a slope improvement from 0.14 to 0.22 (Mask Scentometer D/T) and from 0.27 to 0.54 (Mask Scentometer $(D/T)_G$ using the dataset with the high model predictions removed for the Mask Scentometer D/T and $(D/T)_G$ respectively.

In Paper No. 3 the Mask Scentometer and an OIRS (Intensity D/T) method were used to ground-truth AERMOD using swine lagoons and beef feedlots as the odor sources (area sources). Unfortunately, the analysis performed in Paper No. 3 could not be replicated on this dataset could only be performed on the intensity rating method and Mask Scentometer methods. A better result was found when autocorrelation and random effects were accounted for in with the Intensity-predicted D/T method (each individual assessment was predicted to a D/T, rather than the average intensity rating, which results in a arithmetically different result), and found AERMOD scaling factors of 22 using a cell means model (same as this study) and 12 (accounting for random effects and autocorrelation). In this study we found a scaling factor of 11 for Average–intensity-predicted D/T with no accounting for random effects and autocorrelation. Paper No. 2 found that an Intensity-predicted D/T was about twice an Average intensity-predicted D/T, so the equivalent factor may be closer to 12 or 6. For the Mask Scentometer, the cell means fixed effects model is the same as used in this study and an average slope of 0.5 was found for Mask Scentometer D/T and 0.7 for Mask Scentometer $(D/T)_G$. In this study, when high model predictions were removed that would have obviously been beyond the range of the Mask Scentometer, slopes were found to be 0.22 and 0.54. The results found in this study are analogous to the results found in Paper No. 3.

SUGGESTED CALPUFF AND AERMOD MODEL SCALING FACTORS

Shown in Table 6 are the suggested scaling factors to be applied to CALPUFF and AERMOD odor predictions for livestock building sources found from this study. These scaling factors would be applicable to modeling predictions that were done at one-minute time steps, additional scaling factors may be needed for hourly model predictions.

Table 6. Suggested Scaling Factors for CALPUFF and AERMOD

Model	Nasal Ranger® (either D/T or $(D/T)_G$)	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$	Average-intensity-predicted D/T
CALPUFF	3.0	0.39	0.52	14
AERMOD	2.4	0.22	0.54	11

The peak-to-mean power term, n was also calculated, assuming a one second averaging time for the ambient assessment methods (t_{peak}) and one hour for the modeling time (t_{mean}) with the (C_{peak}/C_{mean}) equal to the scaling factor found in Table 6. These power terms, n are shown in Table 7 and are all within the range of n= 0.167 to 0.65 reported in the literature. The power term for the Mask Scentometer is negative because the model over-predicts the observed of the Mask Scentometer. The most relevant power term that would be comparable past work is Average intensity-predicted D/T which reports a power term much higher than the reported range, 0.64 for CALPUFF and 0.59 for AERMOD. Power terms, n were lower for the Nasal Ranger® and Mask Scentometer than for Average intensity-predicted D/T.

Table 7. Power Terms, n for peak-to-mean ratios

Model	Nasal Ranger® (either D/T or $(D/T)_G$)	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$	Average intensity-predicted D/T
CALPUFF	0.27	-0.23	-0.16	0.64
AERMOD	0.21	-0.37	-0.15	0.59

ODOR METHODS RESULTS

The Spearman's Rank Correlation coefficients (ρ) were used to indicate the strength and direction of the linear relationship between two random variables. The Spearman's Rank Correlation is a special case of the Pearson product-moment coefficient, in which the two sets of data are ranked before calculating the coefficient. In Table 8 the raw scores are converted to ranks and the differences between the ranks of each observation on the two variables were calculated using the "cor.test" command in the R statistical package (R Development team, 2008). The correlation coefficient lies between -1 and 1, with 1 indicating a strong linear relationship (-1 strong inverse relationship, i.e. negative slope) and a 0 indicating no linear relationship. The Spearman's Rank correlation was used because non-parametric tests are considered to be more robust because they do not rely on the assumption that the data comes from a normal distribution of odor levels (i.e. range of possible values and the probability that the measurement was in that range). Also since the intensity rating can be considered an ordinate scale, the non-parametric test (Spearman's ρ) is more applicable and allows for an all methods comparison. The null hypothesis is that no correlation exists between odor measurement methods, the alternative is that correlation (a relationship) exists between odor measurement methods (P values must be less than $\alpha=0.05$ and $\alpha=0.10$ to indicate significant correlation exists) and that the relationship is linear.

Table 8. Spearman's Correlation Coefficient, ρ

	Intensity Rating (0-5)	Average-intensity-predicted D/T	Mask D/T	Mask $(D/T)_G$	DTFCO Lab D/T
Nasal Ranger® D/T	0.50*	0.50*	0.53*	0.32*	0.15
Nasal Ranger® $(D/T)_G$	0.48*	0.48*	0.52*	0.33*	0.12
Intensity Rating (0-5)		1.0*	0.53*	0.26**	0.09
Average-intensity-predicted D/T			0.53*	0.26**	0.09
Mask D/T					0.01
Mask $(D/T)_G$					0.09

** Denotes P<α=0.05, there is a correlation between methods. ** P<α=0.10*

When correlating the odor assessment methods in Table 8 some patterns emerge. First it is apparent that DTFCO Lab D/T does not correlate to any of the other odor measurement methods. The strongest correlations exist between the Mask Scentometer D/T (0.53), Intensity Rating (0.53), Average-intensity-predicted D/T (0.53) and both Nasal Ranger® D/T (0.53) and $(D/T)_G$ (0.52). As expected intensity rating and Average intensity-predicted D/T are perfectly correlated (1.0). The Nasal Ranger® also correlates well with intensity rating (0.48 and 0.50), and Average intensity-rating D/T (0.48 and 0.50). The Mask Scentometer $(D/T)_G$ is not as strongly correlated as Mask Scentometer D/T, but is still significantly correlated to the Nasal Ranger® and to intensity rating and Average intensity-predicted D/T (α= 0.10).

To relate the methods to each other, linear regression was performed as shown previously in Table 4. The method with the strongest fit are intensity rating (0-5) and Average intensity-predicted D/T (R_o^2=0.83). The next best R_o^2 is between intensity and the Mask Scentometer D/T (R_o^2=0.76). Another good R_o^2 is 0.73 between both Nasal

Ranger® D/T and $(D/T)_G$ and Mask Scentometer D/T, although correlation degrades when compared to Mask Scentometer $(D/T)_G$ (R_o^2=0.54 and 0.55). The next best correlations are between intensity and the Nasal Ranger® $(D/T)_G$ (R_o^2=0.64) and the Mask Scentometer D/T (0.76) and $(D/T)_G$ (0.63), and DTFCO (0.55). In general DTFCO Lab D/T has the weakest R_o^2 to any of the other methods.

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 near 1.0. The methods with the slope nearest one are, Mask Scentometer D/T and intensity (1/0.54=1.9), DTFCO lab D/T and Average intensity-predicted D/T ((1/0.36=2.8), Mask Scentometer $(D/T)_G$ and intensity (1/0.25=4), Mask Scentometer $(D/T)_G$ to Nasal Ranger® D/T (0.20) and $(D/T)_G$ (0.19), Nasal Ranger $(D/T)_G$ to Average intensity-predicted D/T (1/0.19=5.2) and Mask Scentometer D/T to Nasal Ranger® D/T (0.12) and $(D/T)_G$ (0.11).

In the CALPUFF and AERMOD analysis, the limitation of the Mask Scentometer was evaluated by removing data points that were greater than the last setting of the device, 18 D/T. Data points were removed when the Nasal Ranger® assessments were greater than 18 for both D/T and $(D/T)_G$ (n=31). This resulted in the slopes nearly doubling, from 0.12 to 0.20 for D/T and 0.20 to 0.48 for $(D/T)_{G.}$ The R_o^2 is still strong at 0.63 and 0.60 for D/T and $(D/T)_G$. We conclude from this analysis that the limited range of the Mask Scentometer is a factor in the results. 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 less data available from the Mask Scentometer readings (n=55) than for the intensity and Nasal Ranger® assessments (n=80), so with more replication, the results may be different.

Table 9. Device Limitation: Regression Results, slope (top), Coefficient of Determination, R_o^2 (middle) and Standard Error (bottom) for Nasal Ranger® assessments less than 18 D/T.

	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$
Nasal Ranger D/T	0.20 0.63 0.03	
Nasal Ranger $(D/T)_G$		0.48 0.60 0.08

Paper No. 2 compared the same methods in a controlled chamber with the same instruments and methods that were used in this study. In that study $(D/T)_G$ provided better results. Table 10 shows the slopes found in this study with the results from the controlled chamber study. As can be seen, many of the slopes are similar. The slopes between Nasal Ranger® $(D/T)_G$ and the Mask Scentometer $(D/T)_G$ are nearly the same. The conversion factor between Intensity rating and Average intensity-predicted D/T is "about 40" since it was found to be 38.2 and 44 between the two studies, but because of the exponential nature of the conversion equation, it is not truly linear. A more reliable comparison is that between the Nasal Ranger and Mask Scentometer, consistently a a slope of 0.10-0.12 for D/T and 0.19 for $(D/T)_G$ (except with the high's removed the slope is 0.48). This comparison provides some confidence that results are comparable and repeatable between a controlled environment and ambient outdoor (uncontrolled) odor assessments. From Table 10 one could assemble a "conversion table" to relate one method's D/T to another, with some relationships being more robust than others.

Table 10. Comparison of slopes between ambient odor assessment methods

X	Y	Controlled Laboratory Comparison Paper No. 2	In-Field Comparison This Study Paper No. 4
Average intensity-predicted D/T	Mask Scentometer D/T	0.03	0.02
Average intensity-predicted D/T	Mask Scentometer $(D/T)_G$	0.07	0.04
Average intensity-predicted D/T	Nasal Ranger® D/T	0.26	0.17
Average intensity-predicted D/T	Nasal Ranger® $(D/T)_G$	0.34	0.19
DTFCO Lab D/T	Average intensity-predicted D/T	0.26	0.36
DTFCO Lab D/T	Intensity	0.007	0.009
DTFCO Lab D/T	Mask Scentometer D/T	0.01	0.02
DTFCO Lab D/T	Mask Scentometer $(D/T)_G$	0.02	0.03
DTFCO Lab D/T	Nasal Ranger® D/T	0.08	0.09
DTFCO Lab D/T	Nasal Ranger® $(D/T)_G$	0.10	0.11
Intensity	Average-intensity-predicted D/T	38.2	44.0
Mask Scentometer D/T	Intensity	0.37	0.54
Mask Scentometer $(D/T)_G$	Intensity	0.34	0.25
Nasal Ranger® D/T	Intensity	0.19	0.07
Nasal Ranger® D/T	Mask Scentometer D/T	0.10	0.12 (0.20 highs removed)
Nasal Ranger® $(D/T)_G$	Intensity	0.07	0.06
Nasal Ranger® $(D/T)_G$	Mask Scentometer $(D/T)_G$	0.19	0.19 (0.48 highs removed)

In summary, many of the slopes are the same or nearly the same for the chamber study in Paper No. 2 as this in-field study. From these two studies, one can now relate with a relative level of confidence (some relationships are stronger and more consistent than others) the relationships between ambient odor methods.

SUMMARY AND CONCLUSIONS

Odor emissions from four swine barns were intensively sampled during a series of twenty 15-minute experiments, where ambient downwind odors were assessed using the Nasal Ranger®, Mask Scentometer, intensity, and DTFCO. Micrometeorological parameters were measured with a cup anemometer and wind vane as well as a high frequency sonic anemometer, temperature and relative humidity at four elevations, short-wave net solar radiation and soil heat flux. Predictions from measured odor emission rates and micrometeorological conditions were generated with CALPUFF and AERMOD and compared to one another and to the ambient odor assessment techniques. The following conclusions resulted:

1. It is clear from this work that results of ambient odor assessments depend on the method and device employed and, while results for one method may correlate and relate well to another method, they are not the same, even if the results are both reported as dilutions to threshold (D/T). When D/T are reported or utilized (i.e. modeling, regulations) they should be reported as a D/T evaluated by the method used (i.e. a 7 D/T was assessed using a Mask Scentometer). In light of this research, it is clear that standards for ambient odor assessment are needed.
2. Statistical bias and error methods were used to evaluate the agreement between model prediction and ambient assessment methods. Although DTFCO Lab D/T and Average intensity-predicted D/T have fractional bias within the ideal range, these methods have the largest model bias and error. The Nasal Ranger®, Intensity rating, and Mask Scentometer have the least model bias and statistical error of the five methods. There was very little difference between models, except that there was slightly better agreement for the Nasal Ranger®, Average intensity-predicted D/T and DTFCO with

AERMOD than CALPUFF. For the Mask Scentometer and Intensity rating, slightly better agreement was observed with CALPUFF than AERMOD.

3. In general, the ambient odor assessment methods showed a slightly better relationship to AERMOD than CALPUFF. CALPUFF predictions were about twice that of AERMOD predictions. Scaling factors for the three ambient odor assessment methods that performed the best (lowest bias and error, regression slope nearest one, and coefficient of determinations) were The Nasal Ranger® and Mask Scentometer. intensity rating was also good to a lesser extent, but is less practical because it is not a comparable measurement to an Odor Unit. The recommended scaling factors for CALPUFF and AERMOD are 3.0 and 2.4 for The Nasal Ranger® D/T and $(D/T)_G$ assessments. For the Mask Scentometer CALPUFF scaling factors were found to be 0.52 for $(D/T)_G$ and 0.39 for D/T and AERMOD scaling factors were found to be 0.54 for $D/T)_G$ and 0.22 for D/T. Scaling factors for Average intensity-predicted D/T were found to be 14 for CALPUFF and 11 for AERMOD. Power terms, n for peak to mean ratios, were found to be within the range of those reported in the literature.
4. The Spearman's Correlation Coefficient (ρ), linear regression, and standard error were used to develop and find the best fitting relationships between the odor methods used in this study between ambient odor methods, and these scaling factors agreed with those found in Paper No. 2. In general an Average intensity-predicted D/T is about a three to five times that of a Nasal Ranger® Assessment and a Mask Scentometer $(D/T)_G$ is about a fifth to a half of a Nasal Ranger® Assessment.
5. The range of the Mask Scentometer appears to be a factor in the results, and is a likely reason for the scaling factors found in this study to be less than 1.0. When high

values were removed from the dispersion model and Nasal Ranger® data sets the scaling factors to the Mask Scentometer improved (moved close to 1.0) when compared to the entire dataset.

6. Geometric average $(D/T)_G$ of the Mask Scentometer and Nasal Ranger® setting was used in this work. This relationship had a pronounced effect on the Mask Scentometer results, but not on the Nasal Ranger Results. This is likely due to the difference in the number of observations taken during the averaging periods. Field olfactometers (specifically the Mask Scentometer, but this may apply also to the Nasal Ranger) that take repeated measures over a period of time, do appear to benefit by having better agreement with dispersion models and other odor methods, from the application of $(D/T)_G$ to their reported scales.
7. Our recommendation is to use the Nasal Ranger® or Mask Scentometer when ground-truthing AERMOD and CALPUFF odor concentration predictions. It provides an overall result that is very near model predictions, when an adequate amount of both ambient and emission data are collected. OIRS and DTFCO had larger scaling factors and weaker relationships to models than the Nasal Ranger® and Mask Scentometer. In general, AERMOD has a slightly better agreement over CALPUFF, but not likely significant, so we cannot recommend one model over the other. While we ran the models at one-minute intervals as opposed to a longer averaging period, we did not evaluate if this was a factor in our results. This work has the most implication to those that model odor dispersion with CALPUFF and AERMOD and use ambient odor techniques to verify model results.

ACKNOWLEDGEMENTS

The project was funded by a USDA CSREES NRI Integrated Programs grant, Verification of Odor Dispersion Modeling for Siting Livestock and Poultry Production Systems, and received additional support from the Agricultural Research Divisions of the respective universities.

REFERENCES

American Society for Testing and Materials (ASTM). 1991. E679-97. Standard practices for determination of odor and taste thresholds by a forced-choice ascending concentration series method of limits. *Annual Book of ASTM Standards.* Philadelphia, PA.: American Society of Testing and Materials.

American Society for Testing and Materials (ASTM). 1999. E 544-99. Standard practices for referencing suprathreshold odor intensity. American Society for Testing Materials, Philadelphia, PA.

American Society for Testing and Materials (ASTM). 2000. D 6589-00:Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance. Annual Book of ASTM Standards. Philadelphia, PA: American Society of Testing and Materials.

Best P.R. 2000. Averaging Time Corrections for Estimating Extreme Air Quality Statistics, 15th International Clean Air Conference, Sydney, Australia.

Committee for European Normalization (CEN) 2003. CEN/TC264/WG2/N222/e Air Quality – Determination of Odour Concentration by Dynamic Olfactometry, Brussels, Belgium. European Committee for Standardization.

Curran, T. P., B.A. Sheridan, E.T. Hayes and V. A. Dodd. 2002. Comparison of tow dispersion models for odour nuisance prediction from pig units. Paper No. 02-4014.

ASAE/CIGR Annual International Meeting, Chicago, July 28-31, ASAE, St. Joseph, MI.

Curran, T.P., V. A. Dodd, and W. L. Magette. 2007. Evaluation of ISC3 and CALPUFF Atmospheric Dispersion Models for Odor Nuisance Prediction. Paper number 074181. Annual Internation Meeting, Minneapolis, MN, 17-20 June. ASAE, St. Joseph, MI.

D'Abreton, P.C. 2006. Personal Communication.

D'Abreton, P. C., C Henry and R Ormerod. 2007. Downwind Odour Predictions from Frou Swine Finshing barns using CALPUFF. Consultancy report for Iowa State University. Prepared by Pacific Air and Environment Pty, Ltd., South Brisbane/Sydney, Australia.

Diosey, P.G., M.E. Hess and L. Farrell. 2002. What smells here? *Water Env. & Tech*: pp 36-39.

R. A. Duffee, M. A. O'Brien and N. Ostojic. 1991. Odor Modeling-Why and How. Recent Developments and Current Practices in Odor Regulations, Controls and Technology, Air and Waste Management Association. 1991.

Guo, H, L.D. Jacobson, D.R. Schmidt, and R.E. Nicolai. 2001. Calibrating INPUFF-2 model by resident panelists for long-distance odor dispersion from animal production sites. Applied Eng in Agric 17(6): 859-868.

Guo, H., L.D. Jacobson, D.R. Schmidt, R.E. Nicolai, J. Zhu and K.A. Janni. 2005. Development of the OFFSET model for determination of odor-annoyance-free setback distances from animal production sites: Part II. Model development and evaluations. *Transactions of the ASAE* 48(6).

Guo, H., Z. Yu and C. Lague. 2006. Livestock Odour Dispersion Modeling: A Review. Written for presentation at the CSBE/SCGAB 2006 Annual Conference Edmonton, Alberta, July 16-19, 2006. Paper no. 06-170. The Canadian society for Bioengineering.

Henry, C. G., P C. D'Abreton, R. J. Ormerod, G. G. Galvin, S. J. Hoff, L. D. Jacobsen, and D.D. Schulte. 2007. Downwind Odor Predictions From Four Swine Finishing Barns Using CALPUFF. In Proceedings of the International Symposium on Air Quality and Waste Management for Agriculture. September 16-19, 2007 Omni Interlock Resort Broomfield, Colorado.

Henry, C.G. 2004. Operation Manual for Mask Scentometer.

Hoff, S. J. and D. S. Bundy 2003a. Modeling odor dispersion from multiple sources to multiple receptors. In: Proceedings of the International Symposium on Gaseous and Odour Emissions from Animal Production Facilities. Norsens, Denmark. June 1-4, 2003. pp 331-339.

Högström, U. 1972. A method for predicting odour frequencies from a point source. Atmospheric Environment, 6: 103-121.

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.

Jacobsen, 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: Proceedings of the International Symposium on Gaseous and Odour Emissions from Animal Production Facilities. Horsens, Denmark, June 1-4, 2003. pp 320-330.

Jacobson, L. D., H. Guo, D. R. Schmidt, R. E. Nicolai, J. Zhu and K. A. Janni. 2005. Development of the OFFSET model for determination of odor-annoyance-free setback distances from animal production sites: Part I: Review and experiment. Transactions of the ASAE 48(6)

Katestone Scientific. 1995. Peak-to-Mean ratios for odour assessment. Report to the Environmental Protection Authority of New South Wales.

Koppolu, L., D. D. Schulte, S. Lin, M.J. Rinkol, D.P. Billesbach and S. B. Verma. 2002. Comparison of AERMOD and STINK for dispersion modeling odorous compounds. Paper no. 02-4015. St. Joseph, MI.: ASAE.

Koppolu, L., D. Schmidt, D. D. Schulte, and L. Jacobson 2004. Development of scaling factors (peak-to-mean ratios) through dispersion modeling with AERMOD and field-based odor measurements for livestock facilities. ASAE paper # 044196. ASAE, St. Joseph, MI.

Li, J. and Guo H. 2006. Evaluating the Impact of Computation Time Interval on Livestock Odour Dispersion using a Computation Fluid Dynamics Model. Written for presentation at the CSBE/SCGAB 2006 Annual Conference, Edmonton Alberta, July 16-19. Paper number 06-173, The Canadian Society for Bioengineering.

Mahin T.D. 1997. Using dispersion modeling odor dilutions to threshold (D/T) odor levels toe meet regulatory requirements for composting facilities. For presentation at the Air and Waste Management Associations 90th annual meeting and exhibition, June 8-13, 1997, Toronto, Ontario, Canada.

Mejer, G.J. and K.H. Krause. 1986. Dispersion models fro emissions from agricultural sources. In *Odor Prevention and Control for Organic Sludge and Livestock Farming*. pp 99-112. V.C. Nielsen, J.H. Voorberg and P.L. Hermite, eds. Elsevier Appl. Sci. Publ. New York, N.Y.

Modi, M. 2006. Modeling odor dispersion form a swine facility using AERMOD. Unpublished MS thesis. Department of Biological Systems Engineering. University of Nebraska Lincoln, Lincoln, NE. 86p

Niemeir, K, R., C. A. Powers, R.R. Stowell, D.D. Schulte. 2008. Validating the Odor Footprint Tool using Field Data. Written for presentation at the 2008 ASABE Annual Meeting, Providence, Rhode Island, June 29-July 2, 2008. Paper number 084577.

NCDENR 2003. Guidelines for Evaluating the Air Quality Impacts of Toxic Pollutants in North Carolina, North Carolina Department of Natural Resources, Division of Air Quality.

Nebraska Department of Environmental Quality 2001. Atmospheric Dispersion Modeling Guidance for Permits, Nebraska Department of Environmental Quality. Lincoln, Nebraska.

Nicolai, R. E., C. J. Clanton, and H. Guo. 2000. Modeling the relationship between threshold and intensity of swine odors. *Second International Conference on Air Pollution from Agricultural Operations.* St. Joseph, Mich.: ASAE.

New Mexico Air Quality Bureau (NMAQB) 2003. Dispersion Modeling Guidelines, New Mexico Air Quality Bureau.

National Research Council. 2003. The Scientific basis for estimating emissions from animal feeding operations. National Research Council. Washington, D. C.

Ormerod, R.J, D'Abreton, P.C and Holmes 2003. Buoyancy effects associated with non-point odour sources: modelling issues and implications, 18th Biennial Clean Air Society of Australia and New Zealand Conference, Newcastle.

Pielke, R. A. 1984. Mesoscale Meterological Modelling, Academic Press, Orlando.

Peterson, W. B. and L. A. Lavas 1986. INPUFF-2.0. A multiple source Gaussian puff dispersion algorithm. Users Guide. EPA/600/8-86/024. August.

Pope, R.J. and P.G. Diosey. 2000. Odor dispersion: Models and Methods.

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.

Scire, J.S., Strimaitis, D.G. and Yamartino, R.J. (2000) A user's guide for the CALPUFF dispersion model (version 5), Earth Tech, Inc., Concord, MA.

Schauberger G., M. Piringer, E. Petz. 2000. Diurnal and annual variation of the sensation of distance of odour emitted by livestock buildings calculated by the Austrian odour dispersion model (AODM). Atmospheric Environment, 34:18-28.

Schauberger G., M. Piringer. E. Petz. 2001. Separation distance to avoid odour nuisance due to livestock calculated by the Austrian odour dispersion model (AODM). Agriculture, Ecosystems and Environment 87:13-28

Schulte, D. D., L. Koppolu, R. K. Koelsch, and D. P. Billesbach. 2004. Odor Footprint Modeling for Confinement Animal Feeding Operations. Water Environment

Federation. Presented at the Annual Wastewater Management Association, Bellevue, WA. April 17th, 2004.

Schulte, D.D., M.R. Modi, C.G. Henry, R.R. Stowell, D.P. Billesbach, S.J. Hoff and L.D. Jacobson. 2007. Modeling Odor Dispersion from a Swine Facility using AERMOD. In Proceedings of the International Symposium on Air Quality and Waste Management for Agriculture. September 16-19, 2007 Omni Interlock Resort Broomfield, Colorado.

Segura, J. C. and J. J. R. Feddes. 2005. Relationship between Odour Intensity and Concentration of n-Butanol. Written for presentation at the CSAE/SCGR 2005 Meeting, Winnipeg, Manitoba, June 26-29, 2005.

Sheffield, R.E., M. Thompson, B. Dye and D. Parker 2004. Evaluation of Field-based Odor Assessment Methods. Water Environment Federation and the Air and Waste Management Association Conference on Odor and Air Quality. March 19-21, 2004. Tacoma, WA.

Smith, M. E. 1973. Recommended Guide for the Prediction of the Dispersion of Airborne Effluents. ASME, New York.

Stowell, R R. L. Koppolu, D.D. Schulte and R. K. Koelsch. 2005. Applications of Using the Odor Footprint Tool. Presented at the VII International Livestock Environment Symposium, Beijing, China, May 18-20, 2005. ASAE Pub #701P0205.

United States Department of Agriculture (USDA). 2001. USDA-IFAFS Project “Aerial Pollutant Emissions from Confinement Animal Buildings (APECAB)”. Six state integrated research project between Minnesota, Indiana, Iowa, Illinois, North Carolina, and Texas. 09/2001 to 09/2004.

United States Environmental Protection Agency (USEPA) 1985. User's Guide for the Building Profile Input Program, United States Environmental Protection Agency, EPA-454/R-93-038

United States Environmental Protection Agency (USEPA) 2003. Appendix W to Part 51 – guideline on air quality models, EPA Pt. 51, App. W.

Willmott, C. J. 1981. On the validation of Models, Phys. Geography, 2, 184-194.

United States Environmental Protection Agency (USEPA) 2004. AERMOD: Description of Model Formulation. EPA-454/R-03-004. September 2004. Office of Air Quality Planning and Standards. Research Triangle Park, North Carolina.

Wilson J. D. 1996. Concentration Fluctuation and Averaging Times in Vapor Clouds. American Institute of Chemical Engineers. ISBN: 0-8169-0679-3. 181 p.

Xing Y., H. Guo, J. Feddes, S. Shewchuck. 2006. Evaluation of Air Dispersion Models Using Swine Odour Plume Measurement Data. CSBE paper No. 06-172.

Zhou, X. J., Q. Zhang, H. Guo, Y. X. Li. 2005. Evaluation of air dispersion models for livestock odour application. CSAAE/SGIR 2005 meeting, Winnipeg, Manitoba, Paper number 05-026.

Zhu, J., L. D. Jacobson, D. R. Schmidt, and R. Nicolai. 2000. Evaluation of INPUFF-2 model for predicting downwind odours from animal production facilities. Applied Engineering in Agriculture. 16(2): 159-164.

APPENDIX

Table 11. Data Used in Analysis, Means of Ambient Odor Methods and CALPUFF and AERMOD Models by Experiment.

Experiment	Receptor	CALPUFF	Nasal Ranger® D/T	AERMOD	DFTCO	Intensity	Mask D/T	Nasal Ranger® $(D/T)_G$	Mask $(D/T)_G$	Average-intensity-predicted D/T
1	A	1.93	16	10	111	1.7	NA	22.6	NA	52.3
1	B	3.76	28	14	158.5	2.1	NA	33.5	NA	80.5
1	C	0.06	1.38	1	173	0	NA	2.1	NA	8.4
1	D	0.51	2.88	4	160	0.8	NA	4.2	NA	19.8
2	A	2.03	16.75	23	77.5	2.2	NA	23.8	NA	89.6
2	B	6.05	14.75	14	46	1.9	1.14	21.0	1.975	64.9
2	C	0.06	1.25	2	36.5	0.1	0.17	1.9	0.505	9.3
2	D	NA	18.5	9	37	2.2	4.63	26.3	8.17	89.6
3	A	2.54	22.25	30	212.5	2.4	5.40	25.4	7.65	111.2
3	B	9.17	4.25	3	246.5	0.8	0.12	5.9	0.4525	19.8
3	C	0.06	12.75	4	249.5	1.6	1.13	17.9	1.82	47.0
3	D	0.75	22.5	8	346	1.7	NA	31.8	NA	52.3
4	A	22.14	37.5	22	52.5	2.1	5.85	46.8	9.9	80.5
4	B	1.61	4.25	4	35	1	0.17	5.9	0.5475	24.6
4	C	1.69	1.13	2	36	0.8	NA	1.7	NA	19.8
4	D	5.92	18.5	6	36	2.2	2.96	26.3	4.84	89.6
5	A	9.48	6.25	20	73	2.1	2.65	9.0	4.64	80.5
5	B	16.9	41.25	31	82.5	1.3	4.68	45.9	8.25	34.0
5	C	0.52	0.5	1	75	0.2	0.07	1.0	0.4	10.4
5	D	NA	14	8	67.5	2	0.78	19.8	1.4825	72.3
6	A	25.56	30	45	241	3	4.03	36.2	6.81	212.4
6	B	0.9	2.13	3	62	0.5	0.54	3.0	0.8975	14.3
6	C	1.63	11	6	202	1.7	1.13	15.7	1.82	52.3
6	D	8.24	16.75	15	53.5	1.1	0.76	23.8	1.2175	27.4
11	A	8.93	7.5	10.8	34	0.8	NA	10.8	NA	19.8
11	B	2.16	0.35	1.1	58.5	0.3	NA	0.8	NA	11.6
11	C	1.48	0.6	14.7	86.5	0.6	NA	1.1	NA	16.0
11	D	1.19	0.3	13.8	121.5	0.2	NA	0.7	NA	10.4
12	A	8.39	25.2	5.4	28	1.4	NA	30.6	NA	37.8
12	B	2.14	0.15	0.7	26	0.9	NA	0.5	NA	22.1
12	C	1.44	5.9	12.7	38	1.3	NA	8.3	NA	34.0
12	D	1.15	1	12	31.5	0.3	NA	1.6	NA	11.6
13	A	17.48	30	22.4	NA	1.2	NA	37.5	NA	30.5
13	B	6.64	1.1	2	NA	1.9	NA	1.7	NA	64.9
13	C	2.54	5.8	0.5	NA	0.3	NA	8.4	NA	11.6
13	D	1.83	0.2	0.2	NA	0	NA	0.6	NA	8.4
14	A	42.71	19.4	22.4	NA	1.3	NA	27.5	NA	34.0
14	B	18.84	0.35	2	NA	0.9	NA	0.8	NA	22.1
14	C	12.43	3.3	0.2	NA	0.5	NA	4.8	NA	14.3
14	D	9.68	0	1.2	NA	0.1	NA	0.4	NA	9.3
15	A	12.72	14.2	6.8	72.5	0.6	NA	20.1	NA	16.0
15	B	10.25	2.1	6.1	122	1.2	NA	3.0	NA	30.5
15	C	3.91	6	3.7	193.5	0.8	NA	8.5	NA	19.8
15	D	3.83	1.3	3.2	193	0.6	NA	2.0	NA	16.0
16	A	14.29	11.6	10.1	54.5	0.6	NA	16.4	NA	16.0
16	B	11.48	2.35	9.7	47.5	1.6	NA	3.3	NA	47.0
16	C	4.07	8.6	6.2	26.5	0.9	NA	12.2	NA	22.1
16	D	4.21	2.9	6.3	44	0.5	NA	3.9	NA	14.3
17	A	4.86	5	15.6	167.5	2	3.60	5.3	6.75	72.3
17	B	2.67	20	5.3	146	1.3	2.80	21.2	7.985	34.0
17	C	2.23	29.84	3.4	97.5	1.2	0.97	26.3	2.68875	30.5
17	D	1.28	6	1.1	51.5	2.8	NA	5.3	NA	171.2
18	A	3.16	2.67	41.2	60.5	2	2.95	2.8	4.35	72.3
18	B	2.77	17.33	25.2	55	1.4	2.80	21.2	7.985	37.8
18	C	1.18	26.17	6.1	59	0.9	1.45	26.3	1.47	22.1
18	D	0.71	9.67	2.1	45	2.3	NA	10.2	NA	99.9
19	A	14.22	50	0.5	NA	3.4	3.88	42.4	1.667083	326.8
19	B	16.48	2.67	0.7	NA	2.4	1.45	2.8	1.47	111.2
19	C	5.86	12.33	0	NA	1.1	1.97	10.2	1.96	27.4
19	D	3.86	8.17	0	NA	1.4	NA	10.2	NA	37.8
20	A	3.99	50	0.5	NA	3.2	7.10	42.4	1.977619	263.5
20	B	5.77	2.17	7	NA	1.8	0.38	2.8	2.12	58.2
20	C	0.56	12.33	0	NA	0.9	0.97	10.2	2.68875	22.1
20	D	1.22	5.67	0	NA	1.2	NA	6.5	NA	30.5
23	A	5.09	12.33	8	88.5	0.9	2.38	10.2	5.1	22.1
23	B	2.81	6.5	5	80	1.4	2.36	6.5	5.44913	37.8

23	C	1.58	25	3	103	1.9	1.34	21.2	5.7	64.9
23	D	1.06	6	2	81	2.2	NA	5.3	NA	89.6
24	A	4.23	20	8	59.5	0.9	2.98	21.2	6.05	22.1
24	B	2.43	9.665	4	52	1.3	3.60	10.2	6.75	34.0
24	C	1.54	30	3	77	2.2	1.30	42.4	7.05	89.6
24	D	0.74	5	2	61	2.2	NA	5.3	NA	89.6
25	A	6.94	9.67	14	NA	1.1	2.65	10.2	7.16	27.4
25	B	3.87	9.67	7	NA	0.8	2.05	10.2	7.61	19.8
25	C	2.74	4.165	5	NA	2.2	2.90	5.3	7.8	89.6
25	D	1.49	40	3	NA	1.4	NA	42.4	NA	37.8
26	A	32.98	7	13	NA	0.7	2.80	10.2	7.985	17.8
26	B	18.14	12.33	6	NA	0.8	1.13	10.2	9.225	19.8
26	C	9.96	5.5	4	NA	2.3	0.90	7.8	11.4	99.9
26	D	3.99	30	2	NA	1.1	NA	42.4	NA	27.4

Closing Comments and Recommendations

Closing Comments

This research has increased the state of knowledge about and our understanding of the relationships between atmospheric dispersion models and ambient odor assessment techniques. It has also shed light on the relationships between different odor assessment methods.

In achieving the first objective of this dissertation, the dilution-to-threshold settings of the Mask Scentometer – previously thought to provide air dilution ratios of 2, 7, 15, 31 and 170 – were measured to actually be 0.35, 1, 2, 4.5 and 18 D/T. What we learned in this phase of the research (Paper No. 1) was that the Mask Scentometer has a much more limited D/T range than previously thought. However, these dilution ratio's have the advantage of providing more resolution at very low odor concentrations (for D/T less than 4.5). The pressure loss across the cartridge on the clean side of the Mask Scentometer is likely greater than with the comparable activated carbon chambers of the Box Scentometer, resulting in the discrepancy in dilution achieved. More comparative work between the Mask and Box Scentometers is warranted. Further development of the Mask Scentometer to extend its range beyond 18 D/T may yield a better instrument for ambient odor assessment.

When the ambient assessment methods were related to each other, both in the environmentally controlled chamber (Paper No. 2) and in the field (Paper No. 4), interesting trends emerged to answer the second objective of this dissertation, what are the relationships between the commonly used ambient odor assessment techniques? Rating odor intensity on a 0-5 reference scale correlates well to the other methods, but it is still a different measure (intensity rating is not the same as concentration) making it difficult in practice to relate intensity values to readings from other methods and to model output. In fact, when intensity was used to predict odor concentration using laboratory-based relationships (that is, Intensity-predicted D/T and verage intensity-predicted D/T), subsequent relationships became weaker. Additionally, laboratory analysis (DTFCO) of odor in ambient air samples, intensity ratings, Average intensity-predicted D/T, and Intensity-predicted D/T were more comparable to each other than to the other (field olfactometry) methods. While lab analysis of samples and intensity ratings may have provided good correlations in prior work, they are likely not as robust as the more direct odor assessment methods, the Nasal Ranger® and Mask Scentometer.

Additionally, some information surfaced regarding how to utilize the equation for predicting D/T from intensity. In practice, one can first use an equation to predict D/T from individual intensity ratings and then average the resulting D/T for analysis and comparison; or, the individual intensity ratings can be averaged and the average intensity can be used to predict a D/T. Paper No. 2 reports the effects of the method used to predict a D/T from a set of assessor intensity ratings. The D/T predicted using an average intensity (Average intensity-predicted D/T) correlated better, had slopes nearer to 1:1, and had stronger coefficients of determination (R_o^2) to readings from the other methods than did averaging of the predicted D/T (Intensity-predicted D/T). In fact, an Average intensity-predicted D/T was roughly half the value of an Intensity-predicted D/T. Average intensity-predicted D/T correlated well with readings from the Nasal Ranger and Mask Scentometer and the session means for the methods were not significantly different. Average intensity-predicted D/T were about three times corresponding Nasal Ranger®

$(D/T)_G$ and about 13 times Mask Scentometer D/T and $(D/T)_G$. In the environmentally controlled ISU chamber, intensity-predicted D/T were roughly 3-4 times higher than D/T obtained using a Nasal Ranger®, and D/T from a Nasal Ranger® were roughly five times higher than a Mask Scentometer.

Many of the slopes (scaling factors) were similar between the two studies. The slopes between the Nasal Ranger® $(D/T)_G$ and Mask Scentometer $(D/T)_G$ were the same, 0.19, in Papers No.'s 2 and 4. DTFCO had slopes that were the furthest from one (largest scaling factors), lowest coefficients of determination for both studies, and was not significantly correlated to the other methods. While laboratory analysis of samples is the 'gold standard' for source odor measurement, DTFCO should not be used to assess ambient odors.

In general the Mask Scentometer, Nasal Ranger®, and intensity ratings had slopes nearest one (smallest scaling factors), significant correlations, and highest coefficients of determination. A reoccurring theme throughout both the lab and field studies was that something is lost in the prediction of D/T from intensity and that the prediction equation is not robust. Correlations degraded and Intensity-predicted D/T did not relate as well to Mask Scentometer and Nasal Ranger® as did the original intensity ratings. This should be a caution to researchers who use intensity ratings to assess ambient odors and predict D/T.

Table 10. Comparison of slopes between ambient odor assessment methods (from Paper No. 4).

X	Y	Controlled Laboratory Comparison Paper No. 2	In-Field Comparison This Study Paper No. 4
Average intensity-predicted D/T	Mask Scentometer D/T	0.03	0.02
Average intensity-predicted D/T	Mask Scentometer $(D/T)_G$	0.07	0.04
Average intensity-predicted D/T	Nasal Ranger® D/T	0.26	0.17
Average intensity-predicted D/T	Nasal Ranger® $(D/T)_G$	0.34	0.19
DTFCO Lab D/T	Average intensity-predicted D/T	0.26	0.36
DTFCO Lab D/T	Intensity	0.007	0.009
DTFCO Lab D/T	Mask Scentometer D/T	0.01	0.02
DTFCO Lab D/T	Mask Scentometer $(D/T)_G$	0.02	0.03
DTFCO Lab D/T	Nasal Ranger® D/T	0.08	0.09
DTFCO Lab D/T	Nasal Ranger® $(D/T)_G$	0.10	0.11
Intensity	Average-intensity-predicted D/T	38.2	44.0
Mask Scentometer D/T	Intensity	0.37	0.54
Mask Scentometer $(D/T)_G$	Intensity	0.34	0.25
Nasal Ranger® D/T	Intensity	0.19	0.07
Nasal Ranger® D/T	Mask Scentometer D/T	0.10	0.12 (0.20 highs removed)
Nasal Ranger® $(D/T)_G$	Intensity	0.07	0.06
Nasal Ranger® $(D/T)_G$	Mask Scentometer $(D/T)_G$	0.19	0.19 (0.48 highs removed)

Some general slope trends emerge from the two studies as can be seen from Table 9 from Paper No. 4. An Average intensity–predicted D/T is about 33-50 times that of a Mask Scentometer assessment and about 4-6 times that of a Nasal Ranger® assessment. A DTFCO Lab D/T is about 150 times that of an intensity rating, about 30 to 100 times more than that of a Mask Scentometer assessment, about 10-13 times that of a Nasal Ranger® assessment, and about 3-4 times that of an Average intensity-predicted D/T. Average intensity-predicted D/T is about 40 times that of an intensity rating. Finally, Nasal Ranger® assessments are about 2-10 times that of assessments from a Mask Scentometer.

In Paper No. 3, random effects and autocorrelation were found to be significant in ambient odor assessment. While it has always seemed intuitive that odor moves as fluctuating filaments drifting in the atmosphere, this work shows a method for accounting for this random nature of ambient odor assessment. Paper No. 3 addresses the third objective of the dissertation, what scaling factors are appropriate for modeling air dispersion from area sources? A scaling factor of about 12 is needed to adjust AERMOD modeled concentrations to match intensity-predicted D/T. Accounting for random effects and autocorrelation brought the slope closer to 1:1. When random effects and autocorrelation were accounted for, the Mask Scentometer method was very consistent across experiments with slopes between AERMOD within a very tight range near 0.15. Slightly more autocorrelation was present in the Mask Scentometer data than the Intensity-predicted D/T data. However, if these effects were neglected – by using the cell means fixed-effects model for $(D/T)_G$ – an AERMOD scaling factor (slope) of 0.5 for D/T and 0.7 for $(D/T)_G$ was produced. In Paper 3, we concluded that using a Mask Scentometers $(D/T)_G$ to ground-truth AERMOD is preferable to using methods based upon intensity ratings (OIRS).

Paper No. 4 answers the final objective of this dissertation, how reliable are atmospheric dispersion models in predicting ambient odor and what effect does the ambient odor assessment method have in this evaluation? Paper No. 4 showed that dispersion models can be used to reliably predict ambient odors. In general, AERMOD has a slight advantage over CALPUFF, but not likely a significant advantage, so it is difficult to recommend one model over the other based upon the data. The overall recommendation is to use the Nasal Ranger® or Mask Scentometer when ground-truthing dispersion models for predicting odor risk. These methods provide results that are very near model predictions (small scaling factors), when an adequate amount of both ambient and emission data are collected. The best correlations were between Mask Scentometer D/T, intensity ratings, and Nasal Ranger® D/T. However, I would tend to dismiss the application of intensity to ground-truth models because it is more difficult to relate intensity ratings to model output. Perhaps in the future, models will be developed to predict odor intensity and thus dispense of the need to predict ambient concentration.

RECOMMENDATIONS

- When using ambient odor assessment methods that take repeated measures in short time spans, be aware that random effects and an autocorrelation effect are likely present in the data and significant. Yet a method exists to account for these effects and those who use ambient odor data can now make informed decisions about how and if to account for these effects.
- When ground truthing dispersion models that predict odor concentration (i.e. output units are odor units, OU or D/T) the best ambient odor assessment methods appears to be the Nasal Ranger® and Mask Scentometer. They provide the closest to a 1:1 relationships to CALPUFF and AERMOD predictions with strong coefficient of determinations, and have lower bias and error than the other methods.
- When using the Mask Scentometer, the actual clean air to ambient air dilution ratios are 0.35, 1, 2, 4.5 and 18. These dilution ratios should be used when

reporting Mask Scentometer results. A shortcoming of the device may be the limited range (max D/T is 18). At very low odor concentrations the device is likely to perform well, however, when a high concentration is experienced, (>18 D/T), then the device is likely not to capture this high concentration which is believed to be the reason for the slope difference between it and the Nasal Ranger®. Development effort is recommended to improve the range of the Mask Scentometer, specifically work to decrease the pressure drop across the clean side cartridge which should have the effect of increasing the dilution ratios.

- For those who use an OIRS (odor intensity ratings) to assess odors for modeling. The conversion equations for OIRS, should be applied to the average intensity ratings of an assessment period and then converted to D/T. This Average intensity-predicted D/T gave a better result than did the average of each predicted D/T of each assessment (Intensity-predicted D/T).
- A geometric relationship, $(D/T)_G$ was applied to the data for the Nasal Ranger® and Mask Scentometer scales. In general, this scale resulted in better results for both devices. Applying $(D/T)_G$ to the unit D/T is more important for the Mask Scentometer because of the repeated measures it takes during a period (i.e. takes 30 assessments as opposed to 2 assessments for the Nasal Ranger®). $(D/T)_G$ (the geometric mean of instrument settings) should be used rather than the unit D/T (instrument settings).
- There is no evidence to suggest that AERMOD performs significantly better than CALPUFF, they both performed well in the studies. The models were provided high-quality, fast-response wind speed and direction data, and were operated at

one-minute time steps. Care should be taken when applying the results to one-hour time-step modeling results where short-time step are used for the ambient odor assessment techniques. To directly apply these results (consider this a disclaimer), either relationships need to be applied to hourly modeled results (peak-to-mean relationships for the method used), or models should be operated in faster (1 minute) time steps to predict odor concentrations.

- Laboratory-based olfactometry (DTFCO) is not a good ambient odor assessment method. It does not provide reliable, consistent data and the resulting D/T do not relate well to data from any of the other methods. There is no compelling reason to use it for ambient odor assessment. Furthermore, the inability to find “solid” correlations and slopes between methods brings to question the practice of using DTFCO to measure emissions if a different odor assessment method is to be used to ground-truth model predictions.
- For future odor work, the same method should be used to measure emissions as downwind assessments. The best approach being either the Nasal Ranger® or Mask Scentometer to measure very near (emissions) and far (receptor) odor concentrations and use a model back-calculation technique to estimate emissions (from the very near assessments). This way, the source emission method is the same as the downwind assessment technique. It is clear from the collective results, that a large source of variation can be attributed to systemic differences in methods.
- Finally it is clear from this work that results of ambient odor assessments depend on the method and device employed and, while results for one method may

correlate and relate well to another method, they are not the same, even if the results are both reported as dilutions to threshold (D/T). When D/T are reported or utilized (i.e. modeling, regulations) they should be reported as a D/T evaluated by the method used (i.e. a 2 D/T was assessed using a Mask Scentometer). In light of this research, it is clear that standards for ambient odor assessment are needed.