SPOILAGE IDENTIFICATION OF BEEF USING AN ELECTRONIC NOSE SYSTEM

S. Balasubramanian, S. Panigrahi, C. M. Logue, M. Marchello, C. Doetkott, H. Gu, J. Sherwood, L. Nolan

ABSTRACT. *A commercially available Cyranose−320 ™ conducting polymer−based electronic nose system was used to analyze the volatile organic compounds emanating from fresh beef strip loins (M. Longisimmus lumborum) stored at 4˚C and 10˚C. Two statistical techniques, i.e., linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), were used to develop classification models from the collected sensor signals. The performances of the developed models were validated by two different methods: leave−1−out cross−validation, and bootstrapping. The developed models classified meat samples based on the microbial population into "unspoiled" (microbial counts <6.0 $\log_{10}$ cfu/g) and "spoiled" (microbial counts ≥ 6.0 $\log_{10}$ cfu/g). Overall, quadratic discriminant−based classification models performed better than linear discriminant analysis based models. For the meat samples stored at 10˚C, the highest classification accuracies obtained by the LDA method with leave−1−out and bootstrapping validations were 87.10% and 85.87%, respectively. On the other hand, classification by QDA and subsequent validation by leave−1−out and bootstrapping provided highest accuracies of 87.5% and 97.38%, respectively. For samples stored at 4˚C, the LDA method provided highest classification accuracies of 79.17% and 85.64% using leave−1−out and bootstrapping validation, respectively. When the QDA method was used, the highest classification accuracies obtained for the samples stored at 4˚C were 87.50% and 98.48%, respectively, with leave−1−out and bootstrapping validations.*

Keywords. *Bootstrap, Electronic nose, Food safety, Intelligent sensors, Meat, Statistical analysis.*

The food and agriculture industry is the largest in the country, representing 16% of our gross national product. Measures to ensure a safe and high−quality food supply are a critical priority for our nation (NSTC, 1996). Food safety problems can be broken down into three major categories: microbial contamination, chemical hazards, and natural toxins. The detection of microbial contamination in foods relies chiefly on culture−based methods that do not examine all products or all surfaces of every product and that also have built−in delays during detection. Non−destructive and real−time means of detecting microbial contamination could greatly reduce the risk of foodborne illness.

The use of electronic noses for sensing the quality of foodstuffs as a means of non−destructive sensing is becoming widespread, and moreover, electronic noses are fast and reliable. Applications of electronic noses for determining food quality have been reported by several researchers (Blixt and Borch, 1999; Schaller et al., 1998; Jonsson et al., 1997; Natale et al., 1997). They have reported the use of various sensors for sensing the quality of meat, grains, coffee, mushrooms, beer, cheese, fish, beverages, blueberries, and even sugars. Boothe and Arnold (2002) employed an electronic nose to analyze the volatile compounds emitted from poultry meat samples. Their study revealed that the electronic nose was able to detect changes in the volatile compounds associated with chicken meat based on the storage time and temperature.

Conducting polymers have been also used as detectors in electronic nose systems (Stella et al., 2000). Conducting organic polymer sensors (CP) exhibit a change of resistance when any gas is adsorbed by the sensor (Schaller et al., 1998). This change of resistance is sensed and delivered as the output. These sensors are very sensitive to polar compounds (Schaller et al., 1998) and can be used at ambient temperatures (Annor−Frempong et al., 1998). However, they also have some disadvantages. For example, because of their low operating temperatures (<50˚C), these sensors are very sensitive to moisture. Electronic nose systems based on conducting polymers have been used for different applications, including monitoring of wine fermentation (Pinheiro et al., 2002), differentiating boar taint (Annor−Frempong et al., 1998), classification of olive oils (Stella et al., 2000),

Article was submitted for review in November 2003; approved for publication by the Food & Process Engineering Institute Division of ASAE in April 2004. A portion of the data used in this study was used for presentation of a poster (92B−28) at the 2003 IFT Annual Meeting.

The mention of a product or a company name does not imply any endorsement, recommendation, exclusion or any other type of implication by any of the authors or their affiliated entities.

The authors are **Sundar Balasubramanian, ASAE Student Member,** Graduate Student, and **Suranjan Panigrahi, ASAE Member Engineer,** Associate Professor, Department of Agricultural and Biosystems Engineering, **Catherine M. Logue,** Assistant Professor, Department of Veterinary and Microbiological Sciences, **Martin Marchello,** Professor, Department of Animal and Range Sciences, **Curt Doetkott,** Statistical Consultant, Information Technology Services, **Huanzhong Gu,** Research Specialist, Department of Agricultural and Biosystems Engineering, and **Julie Sherwood,** Research Specialist, Department of Veterinary and Microbiological Sciences, North Dakota State University, Fargo, North Dakota; and **Lisa Nolan,** Professor, Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa. **Corresponding author:** Suranjan Panigrahi, 1221 Albrecht Blvd. P.O. Box: 5626, Fargo, ND 58105; phone: 701−231−7270; fax: 701−231−1008; e−mail: s.panigrahi@ndsu.nodak.edu.

  ISSN 0001−2351

classification of mite infestation in wheat (Ridgway et al., 1999), and differentiation of spoilage fungi (Keshri et al., 1998).

Appropriate data analysis and pattern recognition techniques should be applied to construct a reliable algorithm for interpreting the acquired signal or smell patterns for classification or prediction purposes. The smell patterns obtained from the electronic nose sensors can be analyzed using various statistical and neural network tools. Pattern recognition techniques like principal component analysis (PCA), partial least squares (PLS), functional discriminant analysis (FDA), cluster analysis, and fuzzy logic or artificial neural networks (ANN) have been used for data analysis in electronic nose applications (Haugen and Kvaal, 1998). Siegmund and Pfannhauser (1999) performed a PCA analysis of the discriminant factors to classify cooked chicken meat samples based on the responses obtained from an electronic nose. Pinheiro et al. (2002) emphasized autoscaling the data prior to principal component analysis, as autoscaling could prevent bias due to high sensor responses dominating the analysis.

If the sample size is small, then building a reliable prediction or classification model is often difficult. Similarly, even with adequate samples, it is critical to build a reliable and robust model that can perform satisfactorily in real−world conditions. Several techniques such as leave−*k*−out (leave−1−out) and bootstrap methods have been used for this purpose. Panigrahi et al. (1998) described a bootstrap procedure employed for building a three−way classification model to determine the quality of edible beans. A single bootstrap sample is created by randomly drawing *n* observations with replacement from the original sample set. However, this bootstrap approach, which is appropriate for small sample sizes and had lower variability than the cross−validation approach, was found to be more biased (Efron, 1983). Efron and Tibshirani (1993) proposed a simple method to correct this bias and that could improve the quality of the bootstrap error rate estimates.

Meat including beef is a staple food in the U.S. Thus, it is critical to ensure the supply of high−quality, safe meat products. The development and evaluation of intelligent quality sensors to provide critical quality information about food and agricultural products are justified. Although different types of sensors have been reported for determining the quality of food products, we hypothesize that the volatile organic compounds (gases) generated by the activities of spoilage organisms on meat could be used as indicators of bacterial presence and, subsequently, of meat quality. Thus, the objective of this article is to evaluate the capability of a commercially available electronic nose system for identification of meat (beef) spoilage. Emphasis is given to the use of selected statistical techniques for developing classification models.

Materials and Methods

Beef Samples

For each experiment, beef strip loin (*M. Longisimmus lumborum*) samples from different animals (obtained from the Animal and Range Sciences Department, North Dakota State University, Fargo, N.D.) were taken after 24 h of chilling. From each beef carcass, a 100 g piece of strip loin was taken. This 100 g meat sample was divided into two equal samples. Each sample was packed separately using a polystyrene base tray and was covered with a commercial food−grade stretch wrap polymer (Filmco Meat Stretch Wrap, Filmco Industries, Aurora, Ohio). This packaging simulated the packaging of meat in retail grocery stores. One of these two samples was used for the electronic nose analysis, and the other was used for analyzing the microbial flora present in it. We assumed that the microbial data from this meat (50 g) sample would be representative of the microbial data in the other 50 g of sample meant for headspace analysis. The packaged meat samples were stored at 10˚C (50˚F) and 4˚C (37˚F). The 10˚C storage condition was chosen to expedite the spoilage of meat, and the 4˚C storage represented the ideal storage temperature of meat in grocery stores.

Three experiments were carried out in different months. In the first experiment, there were two replications for each sample collected from each carcass. For the other two experiments, there was one replication for each sample collected from each carcass. The experiments were conducted in October 2002, March 2003, and May 2003, respectively. The number of samples for each experiment and the sampling interval for microbial analysis and meat headspace analysis were carried out as shown in table 1.

Microbiological Analysis

The microbiological analysis of the meat samples was carried out as required. Aerobic plate counts of the meat were carried out in the Department of Veterinary and Microbiological Sciences at North Dakota State University using standard protocols (U.S. FDA, 1998). Briefly, a small portion of the meat sample weighing approximately 10 g was aseptically removed and diluted to a 1:10 ratio in maximum recovery diluent (MRD, CM733, Oxoid, Inc., Ogdensberg, N.Y.). The meat sample was homogenized in a stomacher (IUL Masticator, Torrent de Lestadella, Barcelona, Spain) for 2 min and serially diluted in 9 mL of MRD as necessary. Plate counts were obtained by plating out samples on the surface of plate count agar (PCA, Oxoid, Inc.) with incubation of the plates at 25˚C for 3 days. Counts were obtained by enumerating typical colonies present and calculated as $\log_{10}$ cfu/g of the meat sample.

Table 1. Electronic nose sampling protocol employed for the stored beef strip loins.

Storage Temperature	Days of Sampling	October, 2002		March, 2003		May, 2003	
		Replications	Samples	Replications	Samples	Replications	Samples
10°C (50°F)	0[a], 1, 2, 3, 4,5,6,7	6	48	4	32	3	24
4°C (37°F)	3,6,9,12,15	6	30	4	20	3	15
Total samples for one experiment	13 days total		78		52		39

[a] Denotes that on day 0 the readings from the sensor were used for both 10°C and 4°C data.

Electronic Nose Setup

A commercially available Cyranose−320™ electronic nose (Cyranose−320™, Cyrano Sciences, Pasadena, Cal.) was used to obtain the smell patterns from the headspace of the beef sample packs. This electronic nose contains an array of 32 conducting polymer sensors. Each sensor has a certain degree of affinity towards specific chemical or volatile compounds. When the sensor is exposed to a chemical, the chemical is adsorbed by the sensing element; subsequently, a change in resistance is experienced by the sensor that is proportional to the amount of chemical absorbed by the conducting polymer surface. This change in resistance over a specific time interval constitutes the signal or the response of the electronic nose. This signal is stored as an output file, which can be exported to MS Excel (Microsoft Corporation, Seattle, Wash.). Out of the 32 sensors, four (sensors 5, 6, 23, and 32) were sensitive to polar compounds (water vapor) present in the headspace due to the respiration of meat. Therefore, the responses from these four polar−sensitive sensors were not included in the data sets used for this study.

Sampling Procedure

Responses of the electronic nose system (Cyranose−320™) to the headspace of meat samples were acquired starting from day 0 until day 7 (a total of 8 days) for the samples stored at 10˚C. From the meat samples stored at 4˚C, similar responses of the electronic nose system were acquired starting from day 3 until day 15 at an interval of 3 days (a total of 5 days). It should be noted that the sampling was carried out at storage intervals appropriate to the storage temperature of meat. In the case of meat held at 10˚C and above, spoilage can occur at a rapid rate; hence, typical sampling intervals are usually daily in order to capture the spoilage counts more accurately. In meat samples stored at temperatures less than 4˚C, spoilage occurs at a much slower rate (Boothe and Arnold, 2002) because the low storage temperature slows down the spoilage characteristics. Due to this reason, we chose a 3−day sampling interval and examined spoilage over a longer period of time. The sampling interval used in acquiring the sensor responses and conducting the microbial analysis is presented in table 1.

Table 2 shows the salient data acquisition parameters. The meat samples were discarded after each day's analysis. Atmospheric air was used as the reference. Prior to purging the sensor array with atmospheric air, the air was conditioned by passing it through desiccants, e.g., drierite (anhydrous calcium sulfate) (fig. 1). A small syringe (open at both ends) was attached to the opening of the purge inlet of the electronic nose system. The syringe was filled with about 3 g of drierite and dry cotton balls. Atmospheric air used for purging the sensors passed through the syringe filled with drierite and cotton. Drierite helped to remove any moisture and unnecessary odor from the atmospheric air. This processed air was then used to purge the sensors for 10 s prior to sampling the headspace gas. The sensors were maintained at a temperature of 42˚C during data acquisition. The electronic nose was programmed such that any data−processing algorithm available in the electronic nose system did not process the acquired data. Thus, the data acquired were the raw responses of the sensor. The setup for data acquisition using the Cyranose−320™ electronic nose is schematically represented in figure 1.

Table 2. Cyranose−320™ sampling cycle for beef headspace analysis.

Operation	Time (s)
Baseline correction (laboratory air intake)	10
Sample draw−in	120
Laboratory air purge	5
Sample purge	30
Manual purge	240
Total run time for one sample	405

Data Pre−Processing and Feature Extraction

The overall data collection and processing techniques used for development of the classification models are shown in figure 2. The commercial electronic nose system has a provision for using several statistical techniques for classification of sensor responses. It should be noted that the sensor response for the meat samples (at day 0) stored at 10˚C was also used as the day 0 reading for the samples stored at 4˚C.

Figure 3 shows a typical raw signal/response of a sensor (detector) of the electronic nose. The acquired raw sensor signals were first pre−processed using binomial smoothing, averaging, and normalization techniques. After the binomial smoothing was carried out, the averaging technique further smoothed the data using a 2−1 data reduction technique. These processing operations were carried out using appropriate functionalities available in GRAMS/32 software (Thermo Galactic, Woburn, Mass.).

The smoothed sensor data were then normalized using the following equation:

$$R_n = \frac{R_i - R_{\min}}{R_{\min}} \quad (1)$$

where

R_n = normalized sensor response at a given instant i
R_i = sensor response at a given instant i
$R_{\min}$ = minimum sensor response obtained between time intervals of 0 and 30 s, where 0 s represents the start of purging with ambient air (fig. 3).

A separate program was written to automatically determine $R_{\min}$ and the corresponding time value of $T_{\min}$. Figure 4 shows a typical plot of sensor signals after they were smoothed and normalized. Subsequently, the area under the curve for the sensor responses between $T_{\min}$ (time at which

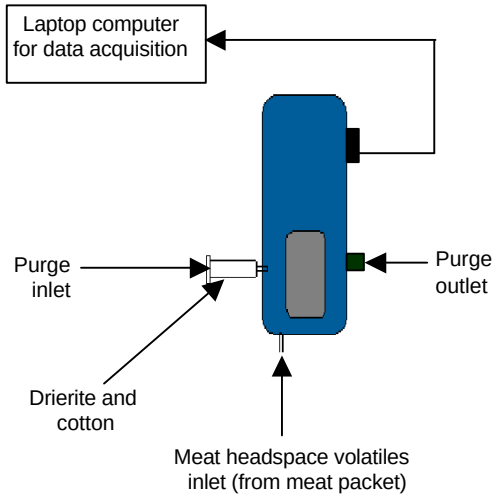


Figure 1. Setup for Cyranose 320™ sensor for headspace volatile sampling.

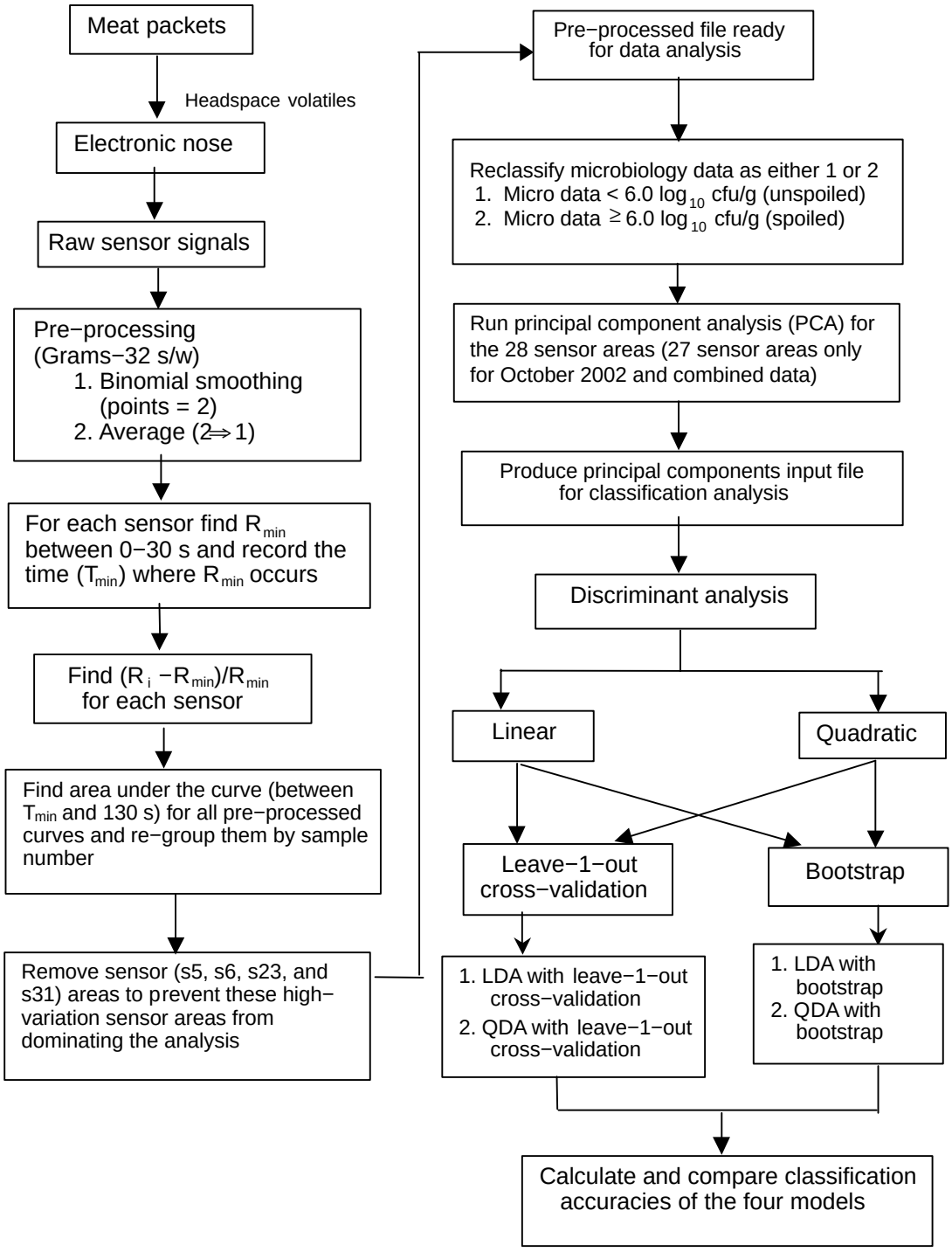


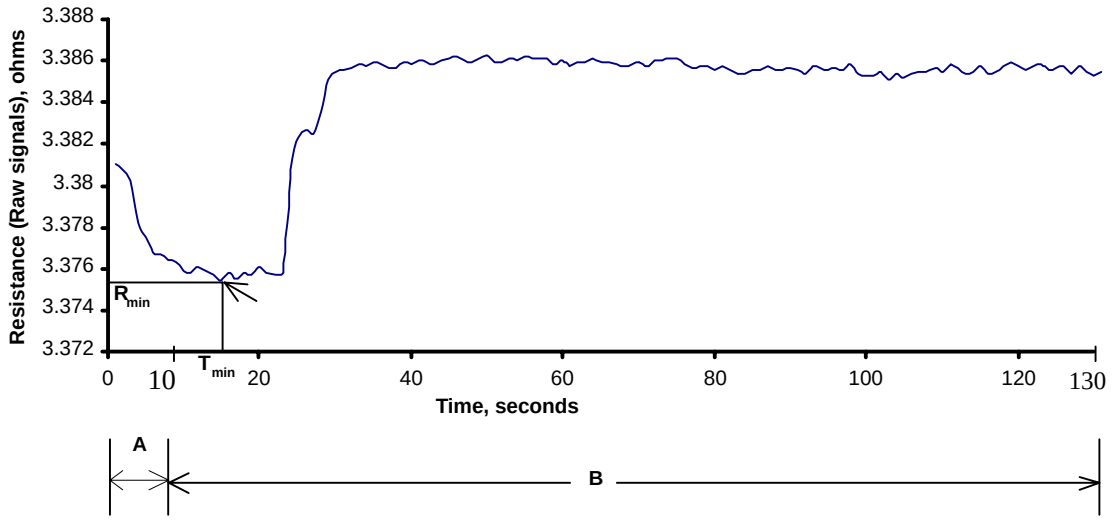
Figure 2. Data collection, processing, and statistical analysis methods used for developing the classification models.

the minimum sensor response, R_{min}, occurs) and 130 s (the time at which data collection from the meat headspace was stopped) was extracted, and this file was stored for future processing.

Twenty−eight areas (features) corresponding to 28 sensors were extracted for each meat sample. For October 2002 and combined datasets only, 27 area features were used, corresponding to 27 sensors, as one of the sensors was not functioning properly during the October 2002 experiment. As the numbers of observations (days) were small (i.e., eight for 10˚C and six for 4˚C), further reductions in the dimensionalities of our data set were needed. Principal component analysis (PCA) is a suitable method that can reduce the dimensionalities while eliminating redundancies in the data set. Therefore, PCA was applied, and the principal components were extracted so that they accounted for 99% or more of the variation. These extracted data sets were further used for development and evaluation of the classification models.

Model Development for Classification

Two types of discriminant analysis models (linear and quadratic) were used to classify the stored beef samples into



A = 10 s (sensor purge for baseline correction)
B = 120 s (Headspace Sample collection from meat package)
R_{min}= Minimum Sensor Response

Figure 3. A typical raw signal/response of a sensor of an electronic nose.

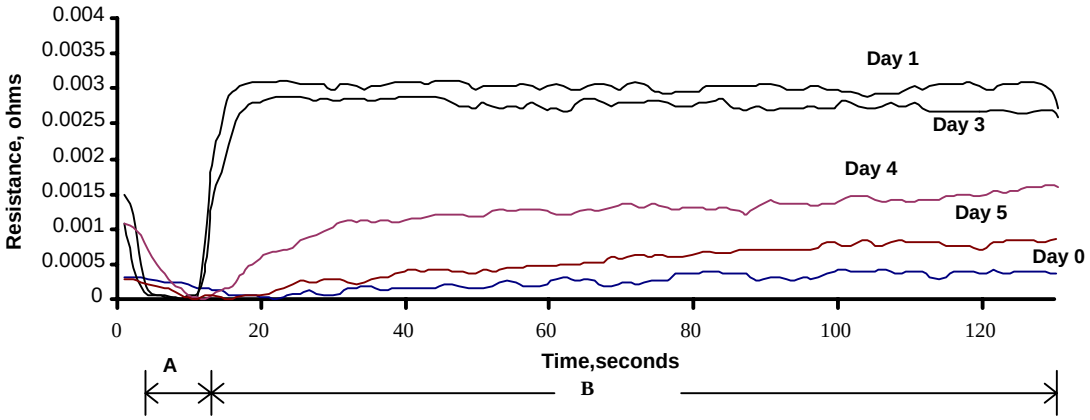
two groups based on their microbial counts. Results of the aerobic plate counts ranged from <1.7 $\log_{10}$ cfu/g to about 9.63 $\log_{10}$ cfu/g. Samples with microbial counts $\geq$ 6.0 $\log_{10}$ cfu/g were classified as "spoiled," and samples with microbial counts <6.0 $\log_{10}$ cfu/g were classified as "unspoiled." The PROC DISCRIM procedure in SAS (version 8.2, SAS Institute, Inc., Cary, N.C.) was used for developing linear discriminant and quadratic discriminant analysis models. PROC DISCRIM calculates the generalized squared distance (Rao, 1973). Each observation (response variable) is then assigned a probability of belonging to a given group based on the generalized squared distance from the group mean. The response variable can thus be grouped into one of the two groups. If the classification is based on the pooled covariance matrix, then the resulting discriminant function is linear (Rao, 1973). If the classification criterion is based on the individual within−group covariance matrices, it results in a quadratic discriminant function.

As the number of observations was small for each experiment, separation of the data into training and testing data sets further reduced the number of data in the "test" data set. Therefore, to increase the robustness of our classification models, we used two different validation techniques: leave−1−out and bootstrap analysis.

In the leave−1−out method, for a given data set with *n* observations, one observation is randomly removed. The model is developed using the rest (*n* − 1) of the observations, which is called the training or calibration data set. The single observation is now used as a validation/test data set. This process of data separation is continued to create *n* training and testing data sets. For each pair of training and testing data set, the model is developed and validated. The performances of the model (classification accuracies) on *n* number of validation (test) data sets were determined, and the average classification accuracy was noted.

The Bootstrap Approach

Bootstrap analysis is an emerging and relatively novel concept to validate a given model in a rigorous manner. Since our data sets, in most cases, consisted of less than 30 observa−



A= Sensor Purge cycle
B=Volatile Sampling cycle

Figure 4. Typical sensor response curves (normalized and smoothed) obtained for beef stored for different days under 10˚C (50˚F).

Table 3. Calculation of the refined estimate of true total error due to bootstrapping for 10˚C (50˚F) meat samples analyzed by quadratic discriminant analysis.

Data Set	Err (B, O*)	Err (B, B*)	Err (O, O*)	Bias/Optimism	Refined Estimator	Accuracy
October, 2002	0.1989	0.1656	0.2444	0.0333	0.2777	0.7223
March, 2003	0.0316	0.0000	0.0000	0.0316	0.0316	0.9684
May, 2003	0.0305	0.0043	0.0000	0.0262	0.0262	0.9738
March−May 2003	0.1489	0.0460	0.2000	0.1029	0.3029	0.6971
Combined (October, March, May)	0.2395	0.2073	0.2400	0.0322	0.2722	0.7278

tions, it was obvious that the error would be very high. In addition, the process of splitting this data set into training and testing sets in the ratio 3:1 or 2:1 would make the training set even smaller. This would introduce a higher percentage of error and a greater degree of uncertainty. To minimize this error and create a better sampling procedure for validating the model created with a small data set, it was decided to use bootstrapping. The approach described by Panigrahi et al. (1998) was used. For our study, 1000 bootstrap samples, each with 50 observations, were created. A single bootstrap sample was created by randomly drawing *n* observations with replacements from the original sample set. The classification developed on each bootstrap sample was validated against the original sample (data set), and the associated estimates of error of prediction were determined. For *N* bootstrap samples, the average estimates of error of prediction were further calculated. As this error of prediction is reported to be biased (Efron, 1983), the following method, as proposed by Effron and Tibshirani (1993), was used to determine the refined bootstrap estimator (error):

- First, use the original sample as both the training (O) and validation (O*) set. Let the error rate for this test be Err (O, O*).
- Next, compute the error rate for both training (B) and validation (B*). Let this error rate be denoted by Err (B, B*).
- Finally, use the bootstrap sample as the training set and the original sample as the validation set. Compute the error. Let this error be Err (B, O*).
- The bias or optimism of the simple bootstrap analysis is now defined as the difference between Err (B, O*) and Err (B, B*) averaged over the *N* bootstrap samples. The refined bootstrap estimator is now given by the optimism added to Err (O, O*).

The results obtained by the LDA− and QDA−based bootstrap technique were corrected for the bias using the process described above, and the refined (or "true") bootstrap estimator/error rate was calculated for the classification models developed. The mean total classification accuracy achieved for each group was also calculated from these refined estimators. Table 3 shows a typical refined estimate of true total classification error for the different data sets computed by the above method for bootstrap analysis. These results were compared with the classification results obtained from discriminant analysis (both LDA and QDA) based on leave−1−out cross−validation.

For this study, a total of five data sets were used for the development and evaluation of the classification models (table 4). Data sets 4 and 5 were created by combining the primary data sets (1, 2, and 3). This combination was done to increase the number of data sets, and to create data sets with additional variability that might be specific to meat/experimental conditions on a given experiment date.

Results and Discussions

Microbial Analysis

Figure 5 shows a typical growth curve of spoilage bacteria ($\log_{10}$ cfu/g) obtained for the beef samples during the experimental period at 10˚C and 4˚C. From the figure, it is evident that the meat samples stored at 10˚C undergo spoilage at a faster rate than the meat samples stored at 4˚C.

LDA and QDA with Leave-1-Out Cross-Validation for Samples Stored at 10˚C

The classification results by LDA and QDA obtained for the meat samples stored at 10˚C (50˚F) and cross−validated by the leave−1−out method are shown in table 5. The highest overall classification accuracy obtained by LDA was about 87.1%. For QDA, the highest overall classification accuracy obtained was 87.5%. The overall classification accuracies achieved by LDA and QDA were very similar (with a maximum of 4% difference in overall accuracies). For individual groups, the highest accuracies obtained by LDA were 92.3% and 100% for the unspoiled and spoiled samples, respectively. On the other hand, quadratic discriminant analysis provided the highest accuracies of 100% and 88.9%, respectively, for the unspoiled and spoiled meat samples. For one data set (March 2003), 0% classification accuracy was provided by QDA for the spoiled samples. The small number of spoiled samples (5) as compared to the number of unspoiled samples (26) could be one of the reasons for this low accuracy. Overall, the unequal distribution of the number of unspoiled (<6.0 $\log_{10}$ cfu/g) and spoiled samples ($\geq$ 6.0 $\log_{10}$ cfu/g) could be one of the reasons for the observed variations among the classification accuracies.

LDA and QDA with Bootstrapping for Samples Stored At 10˚C

Table 6 summarizes the classification results by linear and quadratic discriminant analysis validated by bootstrapping. It can be seen that the highest overall accuracy achieved by LDA was 85% and was provided by two data sets (March 2003 and May 2003). QDA similarly provided overall accuracies >96% for these two data sets. However, when

Table 4. Data sets used for building the classification models.

Period Data Collected	Storage Temperature (°C)	No. of Observations	Data Set
October, 2002	10	45	1
	4	32	
March, 2003	10	31	2
	4	24	
May, 2003	10	24	3
	4	18	
March−May, 2003	10	55	4
	4	42	
Combined (October, March, May)	10	100	5
	4	74	

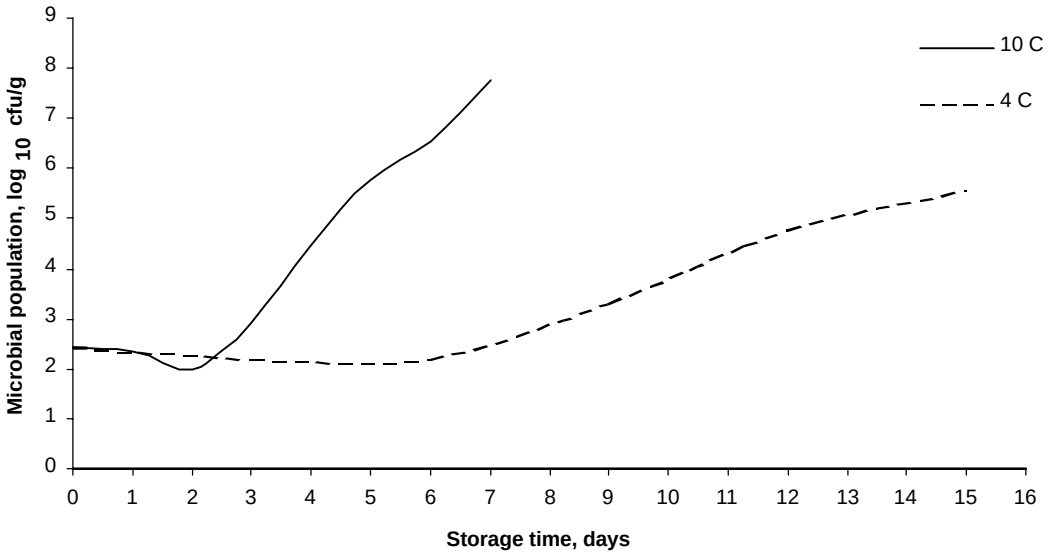


Figure 5. A typical growth curve of spoilage bacteria (log_{10} cfu/g) in beef samples stored at 10˚C and 4˚C, respectively.

these two data sets were combined, both methods showed an accuracy of approximately 70% for the combined data set (March−May 2003). This type of variation in classification accuracy has been observed by Ridgway et al. (1999) while combining data sets. In their study, using a conducting polymer−based electronic nose to classify mite infestation in wheat, they obtained a classification accuracy of around 80% when they considered the weeks of storage separately. However, when they combined the weeks of storage, the classification accuracy went down. Ridgway et al. (1999) reasoned that this drop in accuracy might be due to major changes in the sample (wheat) composition between the first and second week of storage. Thus, for this study, the variation in the experimental conditions pertaining to the temperature, relative humidity, storage conditions, handling of the meat packets, and meat sample conditions (related to the feed, sex, age, and breed of the animal) could contribute to this variation in accuracies.

Additional analyses of within−group accuracies indicated the highest accuracy for the unspoiled meat samples by LDA to be 89.08% (table 6). The maximum accuracy for the spoiled meat samples was 10% higher than that obtained for the unspoiled ones. In fact, three data sets (October 2002, May 2003, and combined) yielded above 90% accuracies for the spoiled samples when analyzed by LDA and bootstrapping. This result is similar to the results obtained when the same three data sets were analyzed by LDA with leave−1−out cross−validation. QDA with bootstrapping provided the highest classification accuracy achieved (100%) for the unspoiled samples and 93.78% for the spoiled samples.

Table 5. Classification accuracies obtained by leave−1−out cross−validation for meat samples stored at 10˚C (50˚F).

				Cross−Validation Accuracy (%)					
	No. of Samples			Linear Discriminant Analysis			Quadratic Discriminant Analysis		
Data Set	Total	Unspoiled[a]	Spoiled[b]	Unspoiled	Spoiled	Total	Unspoiled	Spoiled	Total
October, 2002	45	33	12	66.67	91.67	73.33	69.70	83.33	73.33
March, 2003	31	26	5	92.31	60.00	87.10	100.00	0.00	83.87
May, 2003	24	14	10	71.43	100.00	83.34	100.00	70.00	87.50
March−May, 2003	55	40	15	65.00	66.67	65.45	75.00	46.67	67.30
Combined (October, March, May)	100	73	27	68.49	92.59	75.00	69.86	88.89	75.00

[a] Unspoiled = <6.0 log_{10} cfu/g.
[b] Spoiled = ≥6.0 log_{10} cfu/g.

Table 6. Classification accuracies (based on refined estimate of true total error) obtained by bootstrap based validation for meat samples stored at 10˚C (50˚F).

				Cross−Validation Accuracy (%)					
	No. of Samples			Linear Discriminant Analysis			Quadratic Discriminant Analysis		
Data Set	Total	Unspoiled	Spoiled	Unspoiled	Spoiled	Total	Unspoiled	Spoiled	Total
October, 2002	45	33	12	64.59	91.04	71.70	68.80	81.84	72.23
March, 2003	31	26	5	89.08	69.61	85.87	100.00	80.46	96.84
May, 2003	24	14	10	75.77	99.56	85.67	99.92	93.78	97.38
March−May, 2003	55	40	15	68.52	74.19	70.02	73.84	58.92	69.71
Combined (October, March, May)	100	73	27	66.53	93.00	73.51	70.35	78.65	72.78

Table 7. Classification accuracies obtained by leave−1−out cross−validation for meat samples stored at 4˚C (37˚F).

				Cross−Validation Accuracy (%)			Cross−Validation Accuracy (%)		
	No. of Samples			Linear Discriminant Analysis			Quadratic Discriminant Analysis		
Data Set	Total	Unspoiled[a]	Spoiled[b]	Unspoiled	Spoiled	Total	Unspoiled	Spoiled	Total
October, 2002	32	25	7	60.00	28.57	53.13	84.00	0.00	65.63
March, 2003	24	21	3	76.19	100.00	79.17	100.00	0.00	87.50
May, 2003	18	12	6	33.33	16.67	27.78	66.67	16.67	50.00
March−May, 2003	42	33	9	60.61	55.56	59.50	84.85	33.33	73.80
Combined (October, March, May)	74	58	16	62.07	25.00	54.05	62.07	43.75	58.10

[a] Unspoiled = <6.0 $\log_{10}$ cfu/mL.
[b] Spoiled = ≥6.0 $\log_{10}$ cfu/mL.

LDA and QDA with Leave-1-Out Cross-Validation for Samples Stored at 4˚C

The highest overall classification accuracy achieved by LDA and leave−1−out cross−validation was 79.17% (table 7). The highest overall classification accuracy provided by QDA was around 8% higher than that provided by LDA. The differences in overall classification accuracies between LDA and QDA varied from 4% to 22%. However, the overall classification accuracies obtained for the samples stored at 4˚C were lower than those obtained for the 10˚C samples. The smaller samples sizes for the 4˚C storage condition (as compared to 10˚C) could be one of the reasons. In addition, the sampling intervals were 1 day for the meat stored at 10˚C, as compared to 3 days for the samples stored at 4˚C. This increased sampling interval for the 4˚C samples might have contributed to this variation. This observation is similar to findings that high inconsistencies existed in the microflora and the resultant volatile production in processed poultry (Senter et al., 2000).

The Cyranose 320™ was used for classifying microwave−cooked turkey samples by Mueller et al. (2002), who used all 32 sensors to obtain a sample recognition confidence of 100%. Our experiments did not use all 32 sensors of the Cyranose−320™ as we did not take into consideration the sensors related to humidity information. Our study also differs from that of Mueller et al. (2002) in that they dealt with warmed−over flavor where humidity might not be an issue. Moreover, they used neural networks for data classification. We anticipate that the classification accuracy for our study could be increased by using neural network techniques.

Additional investigation of within−group accuracies revealed 76.19% accuracy by LDA for the unspoiled samples. The spoiled samples were classified with a maximum accuracy of 100%. For QDA, the highest accuracy achieved for the unspoiled samples among all experiments was 100%. On the other hand, the spoiled samples were classified with a maximum accuracy of 43.75%.

LDA and QDA with Bootstrapping for Samples Stored at 4˚C

The results of the classification accuracies of meat samples (stored at 4˚C) using LDA and QDA with bootstrapping are presented in table 8. The maximum accuracy obtained by LDA and bootstrapping was 85.64%, which was provided by the March 2003 data set. A maximum accuracy of 98.48% was obtained by QDA and bootstrapping with the March 2003 data set. QDA with bootstrapping performed better than LDA in classifying the meat samples. This indicated that a non−linear classifier like QDA performed better than a linear classifier like LDA. The comparison of overall classification accuracies between LDA and QDA indicated that the accuracies could vary between 10% and 25% with bootstrapping. Boothe and Arnold (2002) reported that the data obtained from poultry samples stored at 4˚C using a conducting polymer electronic nose clustered more closely than the data obtained from samples stored at 13˚C. This variation was attributed to the greater diversity in volatile compounds generated when the poultry samples were stored at a higher temperature (13˚C). We observed similar trend in our analysis.

The maximum classification accuracies obtained for the unspoiled and spoiled samples by LDA were 83.63% and 99.83%, respectively. Using QDA, the maximum classification accuracies obtained for the unspoiled and spoiled samples were 100% and 87.88%, respectively.

Summary and Conclusions

A Cyranose−320™ electronic nose system was used to classify beef stored at 10˚C and 4˚C into two classes (“spoiled” and “unspoiled”). The electronic nose system was used to acquire the raw signals only. The acquired signals were pre−processed to reduce noise and were then processed to build statistical−based classification models. The statistical techniques used to build the models were linear and

Table 8. Classification accuracies (based on refined estimate of true total error) obtained by bootstrap based validation for meat samples stored at 4˚C (37˚F).

				Cross−Validation Accuracy (%)					
	No. of Samples			Linear Discriminant Analysis			Quadratic Discriminant Analysis		
Data Set	Total	Unspoiled	Spoiled	Unspoiled	Spoiled	Total	Unspoiled	Spoiled	Total
October, 2002	32	25	7	62.56	59.32	61.80	86.55	77.36	84.51
March, 2003	24	21	3	83.63	99.83	85.64	100.00	87.88	98.48
May, 2003	18	12	6	44.71	56.38	48.36	73.06	73.58	73.25
March−May, 2003	42	33	9	57.18	62.77	57.21	84.11	52.26	77.26
Combined (October, March, May)	74	58	16	61.87	12.78	50.72	61.40	58.28	60.45

quadratic discriminant analysis. The performances of the developed models were validated using leave−1−out cross−validation and bootstrapping.

The highest overall classification accuracy obtained for the meat samples stored at 10˚C by LDA and cross−validated by the leave−1−out method was about 87.10%. QDA also provided similar accuracies (87.50%) when cross−validated by the same method. With bootstrapping, overall classification accuracy obtained by QDA increased by 10% (97.38%). The highest overall accuracy obtained with LDA and bootstrapping was 85.87%.

For the samples stored at 4˚C, LDA and leave−1−out cross−validation provided the highest overall accuracy of 79.17%. The highest accuracy obtained with QDA by the same method of cross−validation was 8% higher. With bootstrapping, the highest classification accuracies obtained were 85.64% and 98.48%, respectively, by LDA and QDA methods. The overall classification accuracies obtained by QDA were higher (from 4% to 25%) than those obtained by LDA for the meat samples stored at 4˚C.

This study indicated the capability of an electronic nose system to classify stored beef into two classes ("spoiled" and "unspoiled"). Our ongoing work focuses on the use of neural networks for developing classification models and will be presented in future reports. Future work will also include the use of different extracted features and validation of the classification models on larger data sets.

ACKNOWLEDGEMENTS

The authors would like to express their profound gratitude to the USDA Cooperative State Research, Education, and Extension Service (USDA−CSREES) for their financial support for this research. Cyranose−320™ is the trademark of Cyrano Sciences, Inc., Pasadena, California.

REFERENCES

Annor−Frempong, I. E., G. R. Nute, J. D. Wood, F. W. Whittington, and A. West. 1998. The measurement of the responses of different odor intensities of 'boar taint' using a sensory panel and an electronic nose. *Meat Science* 50(2): 139−151.

Blixt, Y., and E. Borch. 1999. Using an electronic nose for determining the spoilage of vacuum packaged beef. *International J. Food Microbiology* 46(2): 123−134.

Boothe, D. D. H., and J. W. Arnold. 2002. Electronic nose analysis of volatile compounds from poultry meat samples, fresh and after refrigerated storage. *J. Sci. Food and Agric.* 82(3): 315−322.

Efron, B. 1983. Estimating the error rate of a prediction rule: Some improvements on cross−validation. *J. American Statistical Assoc.* 92: 548−560.

Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York, N.Y.: Chapman and Hall.

Haugen, J. E., and K. Kvaal. 1998. Electronic nose and artificial neural network. *Meat Science* 49(3): 273−286.

Jonsson, A., F. Winquist, J. Schnurer, H. Sundgren, and I. Lundstrom. 1997. Electronic nose for microbial quality classification of grains. *International J. Food Microbiology* 35(2): 187−193.

Keshri, G., N. Magan, and P. Voysey. 1998. Use of an electronic nose for the early detection and differentiation between spoilage fungi. *Letters in Applied Microbiology* 27(5): 261−264.

Mueller, K. M., S. J. Formaian, H. Carnahan, J. Forman, N. Shilton, K. Mallikarjunan, D. Vaughan, and D. V. Deventer. 2002. Detection and discrimination of warmed−over flavor in pre−cooked turkey meat using electronic nose systems. ASAE Paper No. 026095. St. Joseph, Mich.: ASAE.

Natale, C. D., A. Macagnano, F. Davide, A. D'Amico, R. Paolesse, T. Boschi, M. Faccio, and G. Ferri. 1997. An electronic nose for food analysis. *Sensors and Actuators B* 44(1−3): 521−526.

NSTC. 1996. Meeting the challenge: A research agenda for America's health, safety, and food. Washington, D.C.: National Science and Technology Council, Committee on Health, Safety, and Food.

Panigrahi, S., C. Doetkott, and R. Marsh. 1998. Advanced information technologies for objective quality sensing of edible beans. In *Proc. Sensoral 98*: *International Workshop on Sensing Quality of Agricultural Products*, 16−20. V. Bellon−Maurel, ed. Montpellier, France.

Pinheiro, C., C. M. Rodrigues, T. Schafer, and J. G. Crespo. 2002. Monitoring the aroma production during wine−must fermentation with an electronic nose. *Biotech. and Bioeng.* 77(6): 632−640.

Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*. 2nd ed. New York, N.Y.: John Wiley and Sons.

Ridgway, C., J. Chambers, E. Portero−Larragueta, and O. Prosser. 1999. Detection of mite infestation in wheat by electronic nose with transient flow sampling. *J. Sci. of Food and Agric.* 79(15): 2067−2074.

Schaller, E., J. O. Bosset, and F. Escher. 1998. 'Electronic noses' and their application to food. *Lebensm.−Wiss. und Technol.* 31(4): 305−316.

Senter, S. D., J. W. Arnold, and V. Chew. 2000. APC values and volatile compounds formed in commercially processed raw chicken parts during storage at 4˚C and 13˚C and under simulated temperature abuse conditions. *J. Sci. Food and Agric.* 80(10): 1559−1564.

Siegmund, B., and W. Pfannhauser. 1999. Changes of the volatile fraction of cooked chicken meat during chill storing: Results obtained by the electronic nose in comparison to GC−MS and GC olfactometry. *Z. Lebensm Unters Forsch A.* 208(5−6): 336−341.

Stella, R., J. N. Barisci, G. Serra, G. G. Wallace, and D. D. Rossi. 2000. Characterization of olive oil by an electronic nose based on conducting polymer sensors. *Sensors and Actuators B* 63(1−2): 1−9.

U.S. FDA. 1998. *Bacteriological Analytical Manual*. 8th ed. Gaithersburg, Md.: AOAC International.