

**Preliminary Summary Points from U.S. Environmental Protection Agency Science
Advisory Board Animal Feeding Operations Emissions Review Panel**

**For Purposes of Discussion at March 16, 2012 Animal Feeding Operation Emission Review
Panel Open Public Meeting**

Summary of Comments on Section 9 – Charge Question 1

Development of Decaking and Full Litter Clean-out Period EEMs

3/16/12 draft

Section 9:

Charge Question 1: Please comment on the statistical approach used by EPA for developing the draft EEMs for broiler confinement houses.... In addition, please comment on using this approach for developing draft EEMs for egg-layers, swine and dairy confinement houses.

DRAFT Panel Response:

The SAB supports EPA’s use of emission factors to estimate emissions from decaking and full litter clean-out operations at broiler operations, but the SAB recommends that EPA express emission factors in more appropriate units than those proposed in the February 2012 draft methodology.

- EPA proposes to estimate emissions by using an emission factors in terms of “g pollutant/kg-bird-day” multiplied by a calculation of litter buildup. These emission factors increase with increasing numbers of flocks (days) on given litter. An “infinitely increasing” emission factor **does not** match available data in the literature.
- Other research has shown that the nitrogen content of litter reaches a balance after 3 to 5 flocks.
 - o During the first four of 18 flocks, Coufal et al. (2006) reported little correlation between nitrogen volatilization and average temperature ($R^2 = 0.27$)
 - o After 3-5 flocks, nitrogen volatilization is highly correlated to average temperature ($R^2 = 0.88$ in Coufal et al., 2006).
 - o Experience shows that after 3-4 flocks, most of the litter material is effectively composted and has reached capacity to retain any more nitrogen per unit mass, thus the mechanism regulating ammonia emissions changes at this point, and the pool of available ammonia remains relatively consistent.

- **Therefore, a more appropriate method of estimating NH₃ emissions (for litter on which \geq 3-4 flocks have been raised) would be in terms of “mass of pollutant per unit mass of litter removed”** (e.g. lb NH₃/ton litter removed). For litter on which < 3-4 flocks have been raised, an emission factor for ammonia in terms of “g pollutant/kg-bird-d” may be appropriate (per EPA’s approach).
- To our knowledge, no similar information exists for constituents of interest to EPA aside from ammonia. However, it is expected that similar generation mechanisms would govern emissions of other gaseous constituents, so a similar emission factor units would be appropriate. PM emissions would be expected to correlate with the mass of litter handled and litter age.
- EPA proposed emission factors calculated based on three different estimates of birds weight (i.e. cumulative weight, total shipped weight, or maximum weight). Litter production (for \geq 3-4 flocks on litter) or total mass of birds raised on litter (for < 3-4 flocks) should be based on cumulative mass of birds raised on litter, as:
 1. This is the unit most pertinent to quantifying litter production,
 2. Bird inventory numbers are regularly collected by producers, and
 3. Bird mass is well modeled and can easily be predicted using growth curves.
 4. The market weight of birds is variable across the industry and is not fully represented in the data collected.

**Preliminary Summary Points from U.S. Environmental Protection Agency
Science Advisory Board Animal Feeding Operations Emissions Review Panel**

For Purposes of Discussion at

March 16, 2012 Animal Feeding Operation Emission Review Panel Open Public Meeting

Additional Data Needs:

3/15/12 draft

Broiler report:

1) Paul Sampson: Mean variance relationships are not in document, nor is a comparison of diagnostic variance vs. means. Were other assessments conducted other than just normality?

Dr. Nail: She can give residual plots.

2) Rick Kohn: How much data was eliminated due to incompleteness?

Dr. Nail: She can get get back to panel on this.

3) Deanne Meyer: Is there a summary table on detection limits for analytical methods?

Larry Elmore: Yes, it's in QAPP. Will identify where.

4) Dave Allen: Measurements were made of cleaned-out litter; where was sample taken in this litter?

Dr. Huber: Would need to check SOP for where and how floor and blowdown samples were taken.

Peter Bloomfield: Table 7-13 on page 7-43: please clarify what is being shown, and what is shown in table and what should occur when eliminating a variable.

Dr. Nail: She can provide clarification.

5) Dave Allen: Panel should consider providing advice on how EPA could use more data than was used in the report. Panel could suggest criteria for looking at additional data and how to use it. Any particular criteria to specify?

Brock Faulkner: It would be helpful to know why certain percentages of data were disqualified in the first place. With that information, could specify criteria.

Larry Elmore: NAEMS data used by EPA to generate EEMS is presented on EPA's website at: <http://www.epa.gov/airquality/agmonitoring/data.html>. This data has already been QA/QC'd and includes qualifications, and followed QAPP per the consent agreement.

Dave Allen: Let's further consider whether raw, unqualified data should also be presented on EPA website.

7) Viney Aneja: Why did EPA check negative measurements?

Robin Dunkins: Will check and get back on this.

Lagoon report:

8) Dave Allen: Were all negative values for Lagoon Report thrown out, or were they screened and selectively used? How did EPA handle negative values? For broilers saw why; what drove negative values in this set of data? Drift?

Dr. Nail: Zero or negative values from this dataset had different reasons for omission than in the Broiler data set.

Robin Dunkins and Dr. Heber: Not sure; will get back on this. In some situations, concentrations are low.

9) Deanne Meyer: EPA should provide information on how outliers were handled, and provide a better description on what the negative value is.

10) Deanne Meyer: Since EPA assumed that lagoon volume capacity for static assumed same volume, EPA should describe lagoon shape (e.g., typical sides; straight vs curved; sloping shape; range of design shapes, etc.).

Danny Greene: Will provide information on slope of lagoons.

11) Ronaldo Maghirang: Did parameter measurements have calculated uncertainty values (e.g., 10%? 20%?)

Danny Greene: Will provide information on that.

12) Robert Hagevoort: Lagoon Report, Section 3:

a) Table 3-1, page 3-3. What does 'not received' data mean. If data exists, should be reviewed. Clarify caption D regarding nitrogen. Conduct a nitrogen mass balance on this data.

b) Table 3.2: H₂S for IN5 data: clarify why average emissions are zero, and max daily as 42.9.

c) Clarify how much manure is captured in lagoons. Did all manure end up in lagoons?

d) Clarify whether solids separation systems are in place. Settling ponds? Solids separation?

13) Deanne Meyer: Lagoon Report, Section 3:

a) Discuss cow excretion data, what % of structures contribute to the volatile concentrations, and % of residence time manure is on hard surface.

b) Need table of when lagoon data was collected on each facility.

14) April Leytem: Provide scientific basis for model selection. Both models may work equally well. If another model works better, discuss why not use it. Provide comparison of the two models, through a validation effort. A reference has shown that RPM model has shown 30% variation, and BLM model has shown 2% variation; consider this reference.

**Preliminary Summary Points from U.S. Environmental Protection Agency Science
Advisory Board Animal Feeding Operations Emissions Review Panel**

**For Purposes of Discussion at March 16, 2012 Animal Feeding Operation Emission Review
Panel Open Public Meeting**

Charge Question 1 and Section 7 of Broiler Report:

3/16/12 draft

Please comment on the statistical approach used by the EPA for developing the draft EEMs for broiler confinement houses and swine and dairy lagoons/basins. In addition please comment on the approach for developing draft EEMs for egg-layers, swine and dairy confinement houses.

Current Approach

We understand that the EPA needs a method to routinely estimate air emissions on animal feeding operations.

The data used for the statistical approach cannot be assumed to represent other farms in the US and therefore the regression models are wholly inadequate to routinely estimate emissions on farms outside the dataset.

Alternative Model Structure

The EPA needs to use a process-based modeling approach to make predictions of air emissions on farms.

The EPA should develop process-based models of different levels of complexity as needed. For example, a very simple model with only a few variables (e.g. number of animals, stage of growth, surface area, production rate) could be used to roughly estimate emissions on different farms. More complex models including additional variables (e.g. manure composition, feed composition, weather or climate parameters) could be used when more accurate estimates are desired.

The alternative structure will be simpler, more easily developed, and more robust than the statistical approach.

Existing data from other studies can also be used.

Model Evaluation

Model-development efforts should be iterative with model evaluation. With each improvement or addition of complexity to a model, the model should be evaluated against its predecessor.

Models should be evaluated using techniques of sensitivity analysis, boundary value analysis, and accuracy (model prediction error).

Regression analysis of predicted vs. observed should be replaced with regression of residuals. R^2 , slope and intercept should be replaced with model prediction error, mean bias, and slope bias.

The collected NAEMS data should be used for model evaluation of the process-based model. This data may also be used to improve parameter estimates of the process-based model.

**Preliminary Summary Points from U.S. Environmental Protection Agency Science
Advisory Board Animal Feeding Operations Emissions Review Panel**

**For Purposes of Discussion at March 16, 2012 Animal Feeding Operation Emission Review
Panel Open Public Meeting**

Broiler VOC EEM:

3/16/12 draft

Charge Question 7: Please comment on the approach EPA used to develop the draft broiler VOC EEM.

The Panel identified significant limitations with the broiler VOC data, and concluded that the broiler VOC data do not support generating a broiler VOC EEM at this time. However, there are valuable components of the VOC data that should be used as appropriate. If VOC data are used to develop an EEM, the methods used to collect the VOC data need to be more extensively documented.

EPA is required to provide an EEM for daily and annual VOC emissions; however, there is provision in the Consent Agreement that if the SAB decides that the available data are not adequate to support development of the EEM, the EPA can delay that development until adequate data are available.

The limitations of the broiler VOC data include:

- VOC data is difficult to gather, and data is not available to support a sound prediction of VOC emissions.
- Canisters were used to sample VOCs, and canisters only assess certain type of VOCs; would need to use other techniques to gather other VOCs that canisters cannot collect.
- Two separate instruments were used to collect VOC data, using two different methods, with two different levels of data completeness.

CA VOC Data:

- The procedure used to collect VOC data in California has not produced useful data for empirical model development and should not be used in the EEM.
 - Used THM analyzer; did not use FID. Tried to use a photobooster analyzer, but resulted in bad data.
 - Did not do measure total VOC.

KY VOC data:

Based on EPA's presentation on KY VOC data, that data appears generally valid and usable.

Concerns include:

- KY VOC data is a very limited data set, and thus concerned about applying it across USA as representative of VOC emissions.
- Unknown recovery rate from the canister. In the lab, was not able to get all compounds out of electropolished canister onto sorbent tube.
- Seven quarters of sampling were conducted. At each sampling event, four samplers were sent to a facility, with two placed at outlets at each barn. Concern is inlet samples were not taken.

Valuable components of the VOC data that should be used as appropriate include.

- Speciation data coupled with reactivity information provides some general information on broiler emissions.

If VOC data are used to develop an EEM, the Panel needs more information on how VOCs were measured, particularly at the Kentucky site.

- Need more information on what drives VOC emissions, particularly at broiler facilities.
- Reactivity of the VOCs included in the measure of non methane hydrocarbons is important and should be available for use with any EEM that is developed and applied to agricultural operations.
- Some indication of the relative magnitude of the VOC emissions from broiler facilities to background levels and other sources would be helpful information.
- VOC speciation data are important and should be described further. The individual compounds found should be reported for both the Kentucky and California sites.
 - Identify what compounds were measured for both the Kentucky and California sites and which compounds might be extractable from GC/MS.
 - The most important compounds should be identified. For example, list the top twenty VOCs that should be collected, and describe why these are the top twenty.
 - Discuss QA/QC procedures that might assess speciation.
 - Show total and speciated VOC data.
 - Identify what fraction of the total was speciated in the results.
 - Compare data to other available AFO data, and indicate whether the VOC numbers are really significant.
 - Note whether there are reactive VOCs. Reactive VOCs provide less confidence in the overall data.
 - Identify what fraction of total was speciated in the results.

**Preliminary Summary Points from U.S. Environmental Protection Agency Science
Advisory Board Animal Feeding Operations Emissions Review Panel**

**For Purposes of Discussion at March 16, 2012 Animal Feeding Operation Emission Review
Panel Open Public Meeting**

Negative or Zero Emission Measurements:

3/16/12 draft

Question 5: Please comment on the EPA's approach for handling negative or zero emission measurements.

Question 6: In the interest of maximizing the number of available data values for development of the draft H₂S EEMs for swine and dairy lagoons/basins, does SAB recommend any alternative approaches for handling negative and zero data other than the approach used by the agency.

Summary: There are two types of negative values – raw concentration data and calculated concentration differences. Calculation of concentration differences (outlet minus inlet) can create negative emission rate estimates; negative concentrations may or may not lead to negative emission rate estimates. For both the raw concentration data and the calculated concentration differences, there was a diversity of Panel opinion on whether EPA should use negative and zero values in calculating EEMs. Panel members raised the following points without reaching agreement on whether to keep or discard zero or negative data:

Keep all zeros and all negative values.

- **Calculated data:** Zero and negative values should be included because:
 - If the calculated value is 0, it should be used.
 - Background values, used to create the calculated values (measured – background), were measured either intermittently (twice a day for gas), or continuously without correction for lag time in the barn (PM data). This could lead to a bias either up or down, potentially creating negative data values.
 - If an event occurred outside the barn (i.e., other barn cleanout, manure movement, etc.), or meteorological conditions created the exhaust air to come back into the barn, these events may create a spike or change in measured values that effect the calculated values.
 - If negative values are excluded due to calculated error, then there is a bias toward those values that were overestimated on the positive side (these values were not taken out of the data set). The average value should be zero. Standard deviation is half positive, half negative. The true estimated value would be closer if all estimated values were included.
 - Negative emissions also had high QA, and there is reason why they washed through.
 - If the calculated value is negative, the raw data can be consulted to discover if it is a calculated effect or other.

- Relatively small number of data points were negative. Negative data does not appear regularly on a daily basis. Seem to be small offsets.
 - Generally can estimate uncertainty with those negative values.
 - If you build a model to make predictions, and the model excluded negative values, the model will be biased when predicting at low emission rates. Any bias in the model weakens the case for use of the end product.
 - There is already a lot of uncertainty in the measurements, which speaks to inclusion of negative values that qualify.
 - There is no statistical problem with inclusion of negative values into the model.
 - If negative rate was due to negative measured rates, go to one half of minimum detection limit (MDL), and include them.
- **Raw data.**
 - If the raw data is 0 after instrument calibration adjustment, it should be used.
 - If the instrument produces a negative concentration value that is due to a “below detection” or “minimum detection limit (MDL)” reading, but within instrument limits, the number should be used. Suggestions on the use of negative values from instrument:
 - Convert negative value, that is within the instrument error to 0 and use.
 - Use the negative value produced if it is within instrument error. Often times values fall below the standard curve as part of the variation of equipment, error, etc.
 - Use EPA procedure of using half of the MDL when observed value is below limit of detection.
 - Outliers should be treated per the QA/QC process first outside of this process, then assessed for negative and zero effects based on above criteria.
 - If raw data is deemed negative after adjustment due to calibration, the value should be included in the data set. If not, there is a bias to those data that are positive due to the same process.

Use some zeros and negative values.

- Data should be qualified on an individual basis.
- Don't make blanket statement. Consider including negative values in some cases. Go element by element. There also may not be a situation/statement that applies for one compound under all circumstances (e.g., due to differences in background concentrations depending on location).
- Raw negative data should be qualified, and considered.

Exclude zeros and negative values.

- If the measured concentration value is below lowest detection limit for the instrument *and* out of instrument error, limits, or uncertainty, then the value should be removed from data set. Qualify data individually.
- If there is instrument drift, ok to not include. Majority of negative values for H₂S EEMs for swine and dairy lagoons/basins were with instrument drift.
- Observed data should not be negative and *any* negative data is not good data and should be considered for elimination.
- If negative value is outlier, then OK to exclude.

Information needs:

- Consider whether raw, unqualified data should also be presented on EPA website.
- Clarify why EPA checked negative measurements.
- Clarify whether all negative values for Lagoon Report were thrown out, or screened and selectively used.
- Clarify how EPA handled negative values for lagoon document. Discuss what drove negative values in this set of data. Drift?
- Prepare an outlier analysis. Get information on how outliers were handled, and provide a better description on what the negative is.

Alternative approaches.

If model is not right shape, can do linear regressions. The magnitude of error is percentage error. If percent is true value, this is a multiplicative error. That error may be a small component of overall regression model. If small number of zeros, that is significant. If have negative value where error ends is different if negative error when hit zero.

There are general methods available to fit non linear models with negative values. There is no statistical problem with fitting such values, and No methodological reasons for excluding negative values. It may be important to fit negative values for these emissions. If truncate model at zero, you may be turning a linear into a non linear model.

**Preliminary Summary Points from U.S. Environmental Protection Agency Science
Advisory Board Animal Feeding Operations Emissions Review Panel**

**For Purposes of Discussion at March 16, 2012 Animal Feeding Operation Emission Review
Panel Open Public Meeting**

Charge Question 1 and Section 7 of Broiler Report:

3/16/12 draft

Discussion of response to Charge Question 1: *Please comment on the statistical approach used by the EPA for developing the draft EEMs for broiler confinement houses and swine and dairy lagoons/basins. In addition please comment on the approach for developing draft EEMs for egg-layers, swine and dairy confinement houses.*

Development and Structure of the broiler Emissions Estimating Models (Section 7)

1. The panel recommends the EPA develop a modeling approach that is more consistent with the sampling design structure and data limitations. Model development needs to consider effects of location, house within location and flocks within house in model inference and prediction. Model uncertainty needs to recognize the limitations in using a small number of locations. The panel is concerned that any model developed from information on two sites is not applicable to all sites in the US.
2. The panel further recommends the EPA carefully consider the process for developing the statistical model, paying attention to the mean and variance components of the model. In particular, the approach for evaluating random effects requires attention. The panel also expressed concern about using a polynomial model for estimating the relationship between animal mass and concentrations.
3. The panel recommends the EPA consider other approaches to the crossvalidation method used to evaluate the model. K-fold crossvalidation methods are preferable to simple data splitting. Splitting of data based on factors related to model usage (such as flock, house and location) should be considered as a way to evaluate model predictive ability.
4. The panel recommends that residual analyses have more importance in the report and modeling process. It is preferable to plot residuals to look for oddities, lack of fit, serial correlation and lack of support for the probability model rather than histograms of the data. The mean and variance specifications should be assessed in an extensive analysis of residuals. The covariance structure, especially the possible contemporaneous correlation among residuals for different houses at a single site, should also be assessed using the same residuals.

General comments

1. The modeling approach generally ignores the sampling/design structure of the data and implications. It is not clear from the model development process if the overall goal is prediction or inference. The sampling design determines the ability to make statements about the collection of potential samples. In this design, there are locations, sites/houses within locations and flocks within houses. These factors, which might be called design factors are mostly ignored. They are however rather important when it comes to making inferences about what factors and interactions are important as they affect the variance estimates and degrees of freedom for testing. While it would be useful to add factors associated with year and season, I suspect that the imbalance in the data will cause limitations when the model is applied to new sites.

The EPA attempts to remove problems due to inadequate sample design by combining the information from separate sets into a single data set. While this may be required to develop a model, the inference may be limited to the locations and houses that are available. From a broad inference perspective the model involves $N=3$ sites. This is a rather small sample for developing models for use in other locations. It is also difficult to estimate variance components with this number of sites.

2. The process of developing the statistical model for predicting each pollutant should begin with finding appropriate specifications of:

- the mean, as a function of the predictor variables;
- the variance, as a function of the mean and/or the predictor variables.

The distributional form of the observations, identification of which is the first step in the process described in the Draft, is generally accepted as less important than the mean and variance specifications. The panel expressed concern about the both deterministic and stochastic components of the model. Specifically

- Nonlinear models: Polynomial regression, such as the use of cubic functions to represent nonlinear dependence in average mass of animals, leads to poor predictions near the extremes of the experimental conditions, and can lead to disastrous extrapolations only just beyond those extremes. It would be useful to plot the model to see the values that might occur for maximum bird mass. Use of a nonlinear model may be a possibility here although there are potential problems with these models as well. The restriction on the range of mass should be reported if the cubic model is used. Plots of individual flocks suggest that a different models might be appropriate for different flocks or that a random effect due to flock is needed. Alternative strategies to polynomials for nonlinear relationships: perhaps low degree of freedom splines that are linear at the boundaries? If polynomials are to be used, the panel recommends us of orthogonal polynomials. With

these one can arguably consider eliminating some interaction terms rather than keeping all three polynomial terms in any interaction considered.

- Correlation structure: It is not clear that the very high temporal correlation structure has been adequately modeled. Usual time series tools (ACF and PACF) should be considered to assess the adequacy of the AR(1) model. The defense of the current model seems to be based entirely on the coverage of predictive intervals. While this is important, this does not guarantee a good model (overall coverage near 95% does not necessarily mean that coverage conditional on other factors is also 95%). The extremely high autocorrelation suggests that perhaps there are some other temporal trend features that could/should be identified.
- Random effects: The analysis approach must consider random effects for flocks. It is possible that other factors (such as buildup) may account for most of the flock effects, but it is still necessary to consider a flock random effect to account for what must otherwise be considered dependent observations (beyond the temporal dependence). Although house and location are also considered as potentially random, there are too few levels of the house and site factors to analyze them as random effects. They should be modeled and tested as fixed effects. We would hope that the house and site factors would act like additive blocking effects in addition to other predictors, but it could be necessary to consider interaction effects permitting other predictors to have different coefficients at different sites.

2. Crossvalidation is a useful tool for model selection and for evaluating predictive ability. Its value is constrained by the method for selecting the test set for model evaluation. By selecting a random sample of observations, the results concerning predictive ability are limited. It is not clear if the method will give a good measure of the predictive ability for a site in Florida, or another state or another location within Kentucky. It should be possible to estimate prediction error for different flocks, for different houses and for different locations by running exercises using these factors to select holdout samples. The crossvalidation exercise could help identify the limitations to the model and to obtain a better estimate of the prediction error at new locations or new flocks.

The exercise described in the Draft as “cross-validation” is not what most statisticians understand by that description. Five-fold cross-validation would involve a similar division of the data set into fifths, but each would be held out in turn, and predicted using a model fitted to the other four fifths. The exercise described in the Draft is also not a true validation, because the performance of candidate models in predicting the hold-out data was used in the model selection process. In a true validation, the test data would be held out of the entire model selection and estimation process. Model and analysis must incorporate the factors in the experimental design. That is, the “house”, “site” and “flock” factors must be part of any analysis.

3. The panel recommends that residual analyses be part of the report. Histograms are used to indicate that the data are skewed however, these plots are rather limited, as the authors point out. It is preferable to plot residuals; to look for oddities, lack of fit, serial correlation and lack of normality. The mean and variance specifications should be assessed in an extensive analysis of residuals. The covariance structure, especially the possible of contemporaneous correlation among residuals for different houses at a single site, should also be assessed using the same residuals. Table 7-9 is definitely not a good way to assess mean-variance relationship as the constant range of NH₃ values in the rows of the table constrain the SDs to be similar.

4. The variable selection approach in the model building is likely suboptimal with respect to the goal of accurate prediction. We would recommend a modern text focusing on prediction, such as “The Elements of Statistical Learning” by Hastie, Tibshirani and Friedman. Because the primary aim is prediction there is no reason to base variable selection on backward elimination with a conservative $p < .001$ criterion. The apparent significance of individual predictors is not a primary concern, especially in the context of (somewhat) correlated predictors. While the final choice of model was not completely automatic according to the backward elimination algorithm, there seems no reason not to consider the results of an all subsets regression procedure rather than backward elimination (although this would only be possible without all the interaction effects) using a BIC criterion. Uncertainty in the “best” model could be assessed with cross-validation (see below).

Miscellaneous:

Page 7-29 Bottom of second paragraph - centering does not produce data where 50% are below zero and 50% above unless you are centering by the median.

Page 7-31 is there any justification for using a change in R² for adding interaction terms?

Page 7-37 (2nd paragraph) Can three sites really be representative of all sites? Consider rewriting the sentence.

Figure legends could contain more information. For example, Table 7-16, mention this is standardized data.

Diets can have a large impact but are not included.

Season is not in the model. The data cover somewhat different time periods. It may however be difficult to include season in a simple manner. The California sites are sampled from Sept 2007-Oct 2009 while Kentucky sites were sampled from February 2006 to March 2007. There is

confounding between location and time of sampling (year of sampling). This may affect inference for seasonality.

Table 7-2 parameters picked based on chemistry and knowledge - additional variables should have been in the model.

The use of the regression of predicted versus observed is potentially difficult as one may obtain an R^2 of 1.0 when there is a biased model.

Check calculations of LL and BIC - these seem to be based on REML rather than ML

It is generally not clear how different results will be when they are applied to new sites.

A relevant variable that should be include is the nitrogen inputs.

Some of the variables exhibit measurement error (number of birds and average bird weight are estimates). Consider accounting for the error because if it is not accounted for the relationship between the concentration and predictors is attenuated.

In Table 7.8 it appears that the variance component for house is significant (or this is a typo).

It might be useful to consider a method such as quantile regression for estimating the percentiles of the distribution rather than the average values.

There is clear lack of constant variance. As is common in data of this kind, the variance and the mean are correlated, so that, e.g., the variance in the response increases as the mean response also increases. A simple solution to this problem is to transform the response variable since transformation sometimes disentangles the mean and the variance. One possibility is to just use a square root of a log transformation (assuming that we report the zeros as censored).

Table 7-2 refers to selected variables, which were based on knowledge of chemistry (p. 7-14). Air flow, temperature, and time variables are related to chemistry. However, feed rate and composition, water management, and manure composition (moisture and N) also relate to the chemistry.

Table 7-3. Note that more than half of data is missing in fall, 79% missing in California.

Centering and scaling the predictor variables (usually termed “standardization”) has no effect on collinearity, except between a predictor and the constant term.

When a model estimated using “base” data is evaluated by comparing its predictions for “hold-out” data, the rmse (root mean squared prediction error) is the most important summary. The R^2 in the regression of the hold-out data on the predictions is less relevant; testing that the regression has a slope of 1 and an intercept of 0 gives some information about possible differences between base and hold-out data.

In Section 7.4.3: “a small p-value indicates that the estimated value of the parameter is not significantly different from zero ” is the opposite of the correct interpretation; a small p-value indicates that the estimated value of the parameter *is* significantly different from zero.

It is not clear how important some factors are since tests are not reported. It does appear that animal mass seems to be most important.

All relevant variables were not used. More consideration to mass balance and process-based models is needed.

The panel thought that there might be evidence for variance heterogeneity and requests that this be given additional attention.

There is a lack of data/modeling of correlation between houses on a farm which is necessary in order to put confidence limits on a total farm emission estimate. Future data collection should include information on multiple houses. The report shows concern at the end of section 7.4 (p. 7-37) for making predictions at sites not included in the NAEMS. Unfortunately, if there are any significant differences between the sites available, whether in variance and auto-correlation parameters or other fixed effect parameters, any such predictions cannot be justified. It is a limitation of the study design that collected data on only 3 sites in 2 states. One cannot ignore heterogeneity across sites and rely on predictions assuming no heterogeneity, as suggested at the end of section 7.4.

Consider a joint test of significance of the slope and intercept in the model that compares predictions and actual values (p 7-42). . Also a plot of these values is warranted.

Negative values

About negative/zero values: I am not sure whether I completely understand what is going on here, but here is what I think I understand.

- There is no such thing as a negative emission. If a negative value is recorded, this **MUST** be due to measurement error.
- If we know the minimum detection level of the instrument that made the measurement, then any measurement below that detection limit is **CENSORED** (it

something between 0 and the MDL) and should be treated as such in the analysis. This would include any value below the MDL, including negative values, zeros, etc.

- The simple solution is to use half of the MDL in place of the measurement, as Wendy (I think) suggested.
- The not-so-simple but statistically more correct approach is to toss the standard regression model and fit a model that allows for the presence of censoring. SAS can handle that, for example, but it requires a bit of an overhaul of the statistical models in Section 7.
- If a calculated value such as measurement – background is negative, we will cannot report a negative value, it seems to me. I would probably again just report that the emission is below background and either leave it at that or again treat it as censored, although this is a bit more difficult because the background is also an estimated value and using it as a fixed censoring threshold is difficult.

Section 8

- Plots in section 8 really suggest analysis on a log scale would be appropriate, although this is not certain without diagnostics. Log scale not strongly suggested for VOC.