United States Environmental Protection Agency

Overview of Statistical Methodology Used to Develop EEMs for Swine and Dairy Manure Storage

3/15/2012

United States Environmental Protection Agency

Outline of Presentation

- Overview of Lagoon/Basin Monitoring
- Swine and Dairy Lagoon/Basin Emissions-Estimating Methodologies (EEMs) Development
- Charge questions to the Science Advisory Board

United States Environmental Protection Agency

Sites Selected for Monitoring

- A total of 9 liquid open source sites were monitored
- Swine sites
  - 6 sites located in the southeast, midwest and west
  - Monitored both sow and finishing farms
  - 5 lagoons and 1 basin were monitored
- Dairy sites
  - 3 sites located in the midwest and northwest
  - Monitored 2 lagoons, 1 basin
- Emissions measurement:
  - Two sites were monitored continuously for approximately 1 year
    - 1 dairy and 1 swine
  - Remaining sites were to be measured for up to 21 days each season over 2 years

United States Environmental Protection Agency

EEMs Submitted to SAB for Review

- The EPA developed three alternative NH_3 EEMs for SAB's evaluation that use:
  - Ambient meteorological data, and
  - Paired combinations of static, farm-based variables (i.e., animal type, surface area, and farm size)

- The EPA will revisit the EEMs pending SAB's review and feedback.

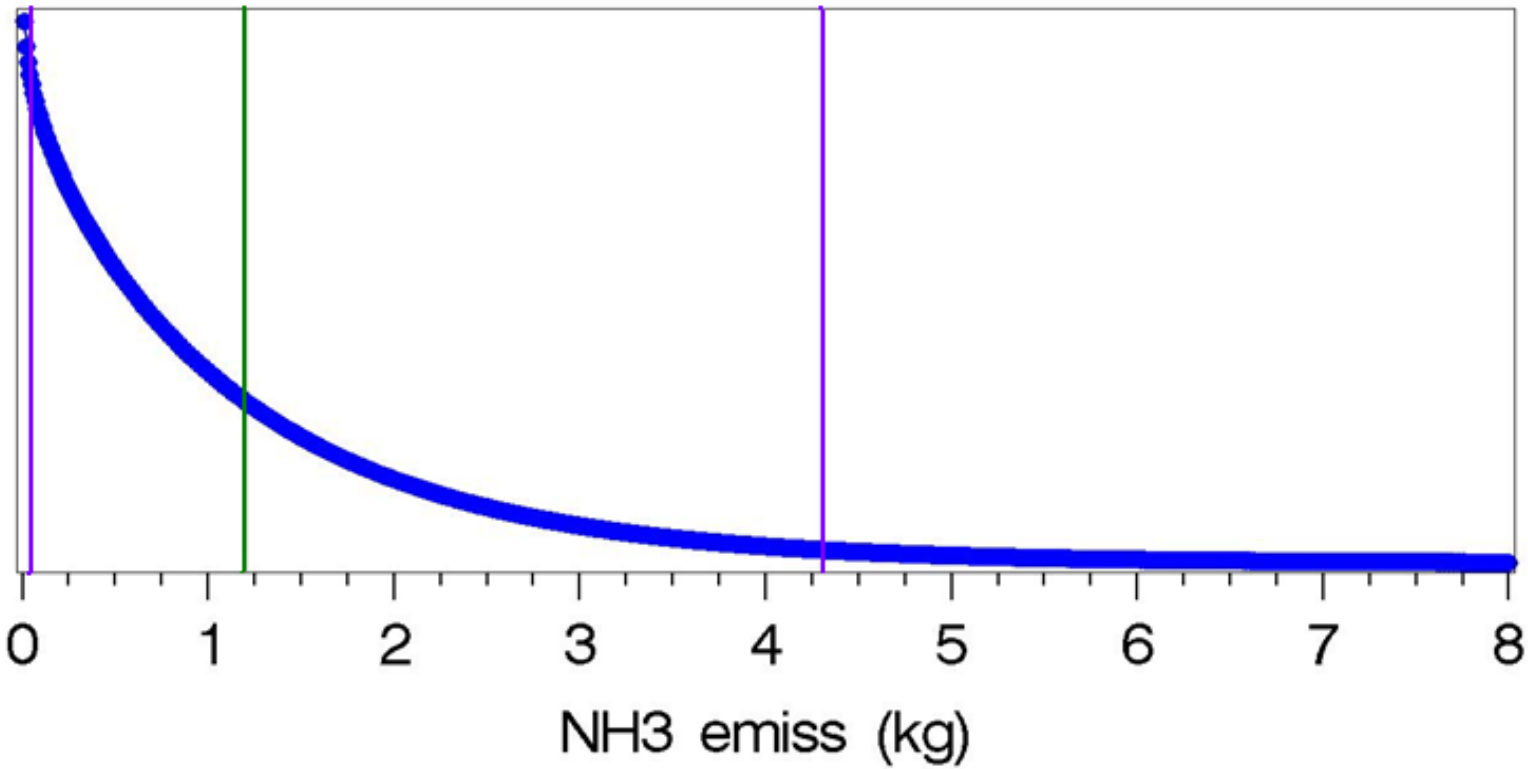
United States
Environmental Protection
Agency

EEM DEVELOPMENT

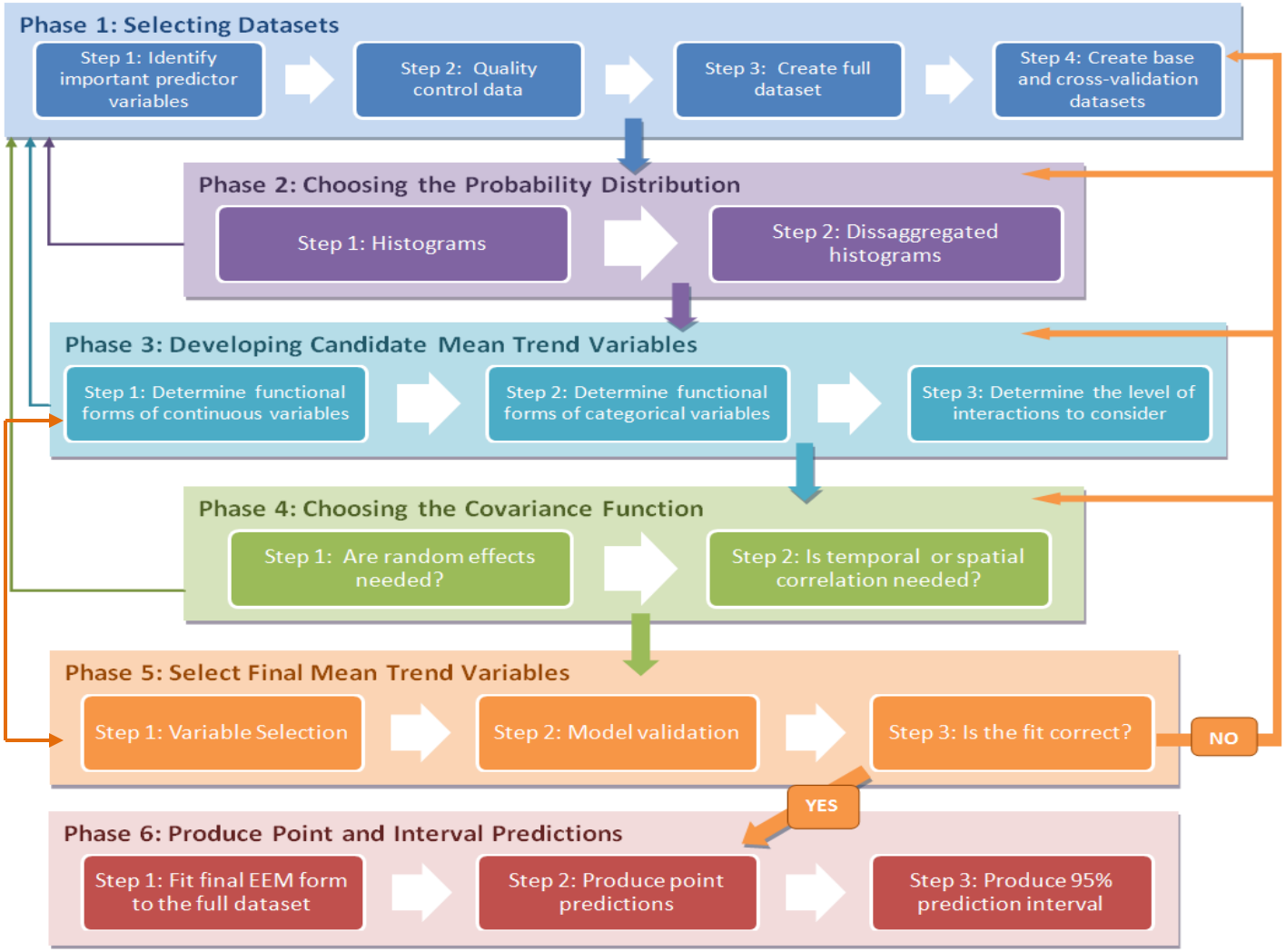
Example Half-hour NH_3 Emissions

Point prediction: 1.2 kg

95% prediction interval: (0.036, 4.3) kg



Overview of EEM Development Approach



United States Environmental Protection Agency

NAEMS Data for Dairy and Swine Open Sources

Category	Description	Frequency
Meteorology	Ambient temperature (ºC)	Continuous
	Relative humidity (%)	Continuous
	Atmospheric pressure (kPa)	Continuous
	Dew point temperature (ºC)	Continuous
	Solar radiation (W/m^2)	Continuous
	Surface wetness (Ω)	Continuous
	Wind speed (m/s)	Continuous
Lagoon liquid	Total Kjeldahl Nitrogen (TKN) content (wet weight %)	Periodic
	Solids content (wet weight %)	Periodic
	NH_3 content (wet weight %)	Periodic
	pH (pH)	Continuous
	Oxidation/reduction potential (mV)	Continuous
	Temperature (ºC)	Continuous

United States Environmental Protection Agency

NAEMS Data for Dairy and Swine Open Sources (cont.)

Category	Description	Frequency
Farm Characteristics	NAEMS site ID	Static
	Animal type (Swine or dairy)	Static
	Farm animal capacity (head)	Static
	Average animal weight (kg)	Static
	Average animal weight (piglet) (kg)	Static
	Number of confinement structures on the farm (structures)	Static
	Farm manure management system	Static
	Solids separation (Y/N)	Static
	Odor control (Y/N)	Static
	Farm age (years)	Static
	Animal Inventory (head)	Periodic

United States Environmental Protection Agency

NAEMS Data for Dairy and Swine Open Sources (concl.)

Category	Description	Frequency
Lagoon Characteristics	Impoundment type (lagoon or basin)	Static
	Lagoon configuration (e.g., single stage, multiple stage)	Static
	Lagoon volumetric loading rate (lb VS/d-1,000 ft^3)	Static
	Lagoon surface loading rate (lb VS/d-ac)	Static
	Lagoon volume (ft^3)	Static
	Lagoon surface area (1,000 ft^2)	Static
	Lagoon liquid depth (ft)	Static
	Lagoon sludge depth (ft)	Static
	Number of manure inlets to the lagoon (inlets)	Static
	Manure discharge schedule (days)	Static
	Lagoon pump-out frequency (days)	Static
	Lagoon agitation prior to pump-out (Y/N)	Periodic
	Manure discharge to lagoon event	Periodic
	Natural lagoon cover (%)	Periodic

Initial NH_3 Data Sub-setting

NAEMS Data Submitted to EPA:
12,854 30-min NH_3 emissions observations
13 Time-varying predictor variables
25 Static predictor variables

Data Completeness and Usability Assessment

Preliminary Dataset:
10,783 30-min NH_3 emissions observations
5 Time-varying predictor variables
8 Static predictor variables

United States Environmental Protection Agency

NH_3 Data

- Missing NH_3 data
  - Hours missing within days
  - Whole days missing
- Course of action
  - Did not aggregate to daily
  - Used half-hour values

United States Environmental Protection Agency

Lagoons Liquid Data

- Observed missing data for liquid measurements
  - pH, oxidation reduction potential, temperature
  - Would reduce NH_3 data available for EEM development
- Course of action
  - Omit from analysis
  - Used static farm-based predictors as surrogates

United States Environmental Protection Agency

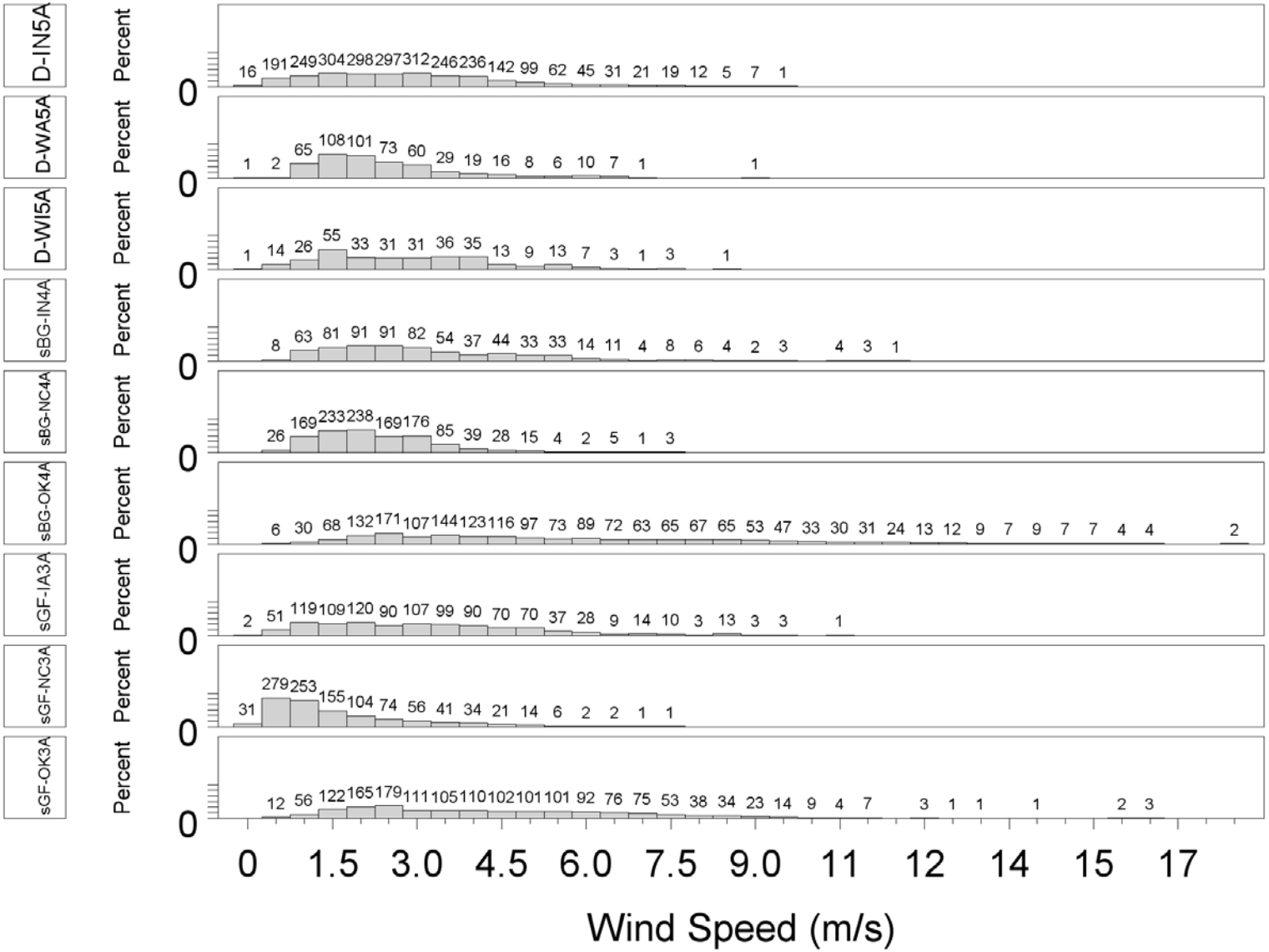
Candidate Predictors

Category	Description	Predictor Variable	Units
Ambient	Temperature	*ta*	^{o}C
	Relative humidity	*ha*	%
	Wind speed	*ws*	m/s
	Hour of the day	*hour*	hour
	Julian day (day of the year)	*jday*	day
Farm and lagoon	Animal type (Dairy or Swine indicator)	*animal*	NA
	Capacity of farm (number of animals)	*capacity*	head
	Average adult animal weight	*adultwt*	lb
	Number of confinement structures	*barns*	barns
	Manure management system	*mms*	NA
	Surface area	*sa*	1000 ft^2
	Number of manure inlets into lagoon	*inlets*	inlets
	Whether an odor control agent was used on a given day	*odorctrl*	NA

NA = not applicable

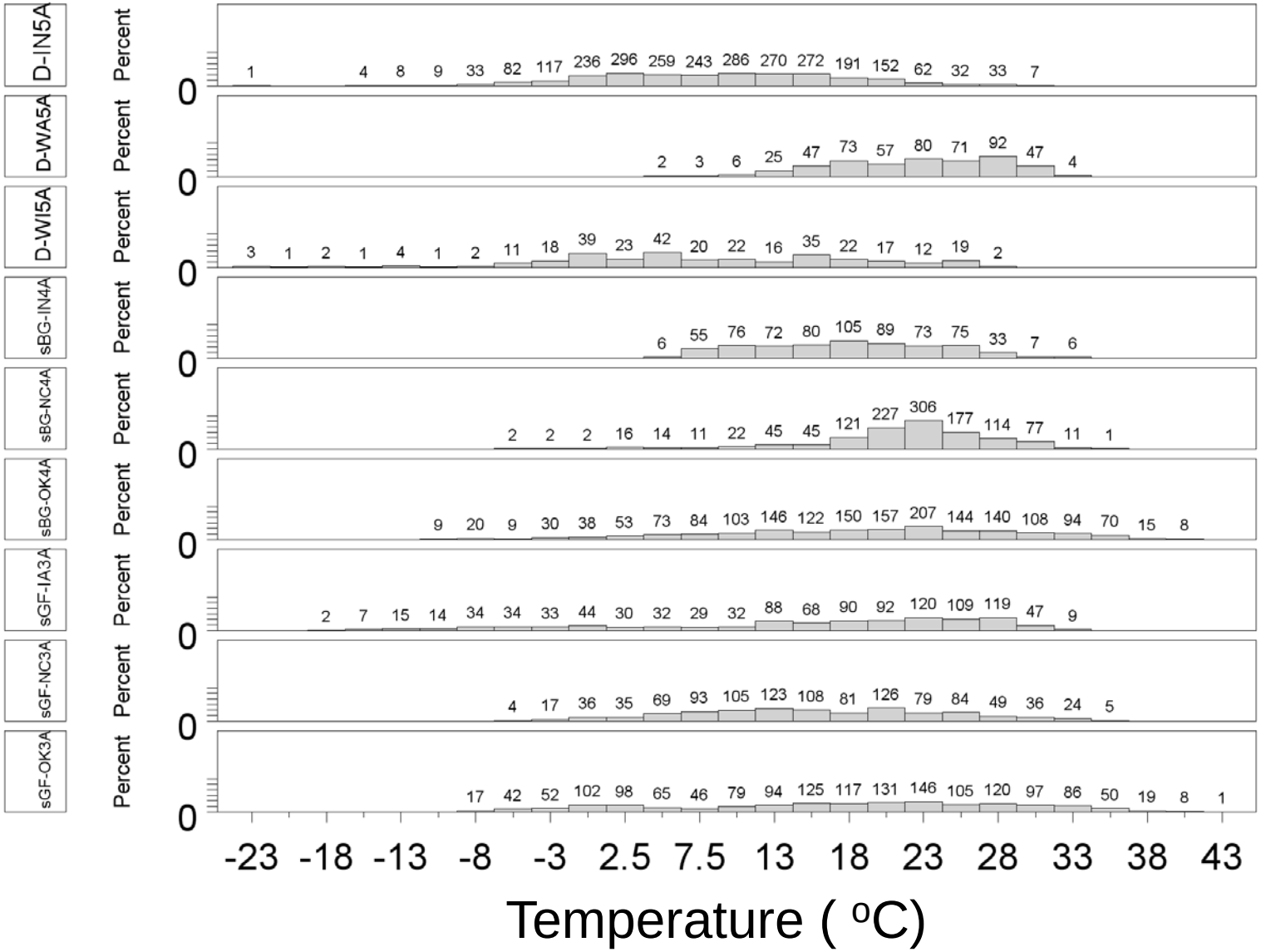
United States Environmental Protection Agency

Data Limitation: Wind Speed



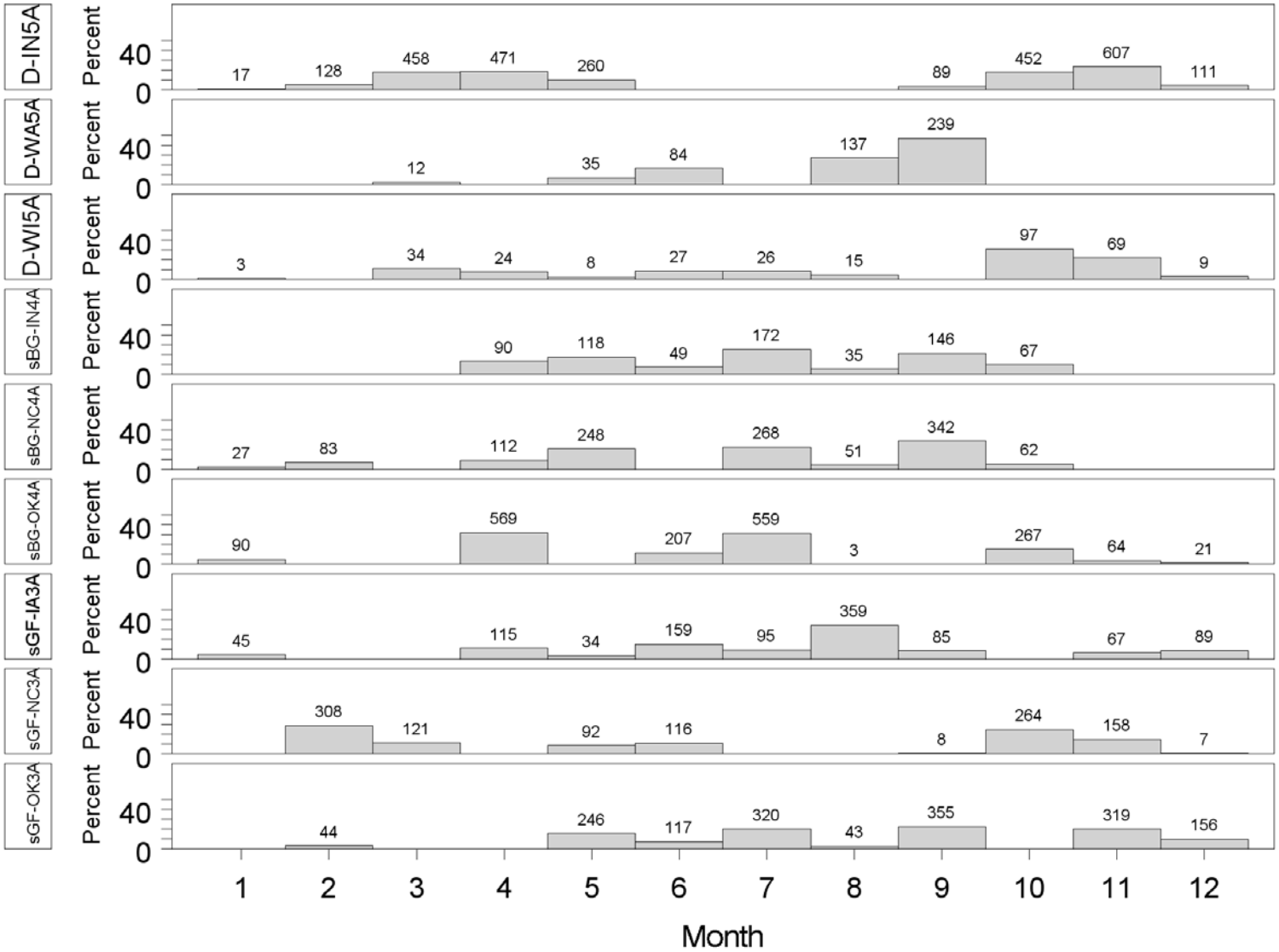
United States Environmental Protection Agency

Data Limitations: Temperature



United States Environmental Protection Agency

Data Limitations: Season



United States Environmental Protection Agency

Data Limitations: Summary

Dairy

- Limited high winds speed data
- Limited high temperature data
- Limited summer data

Swine

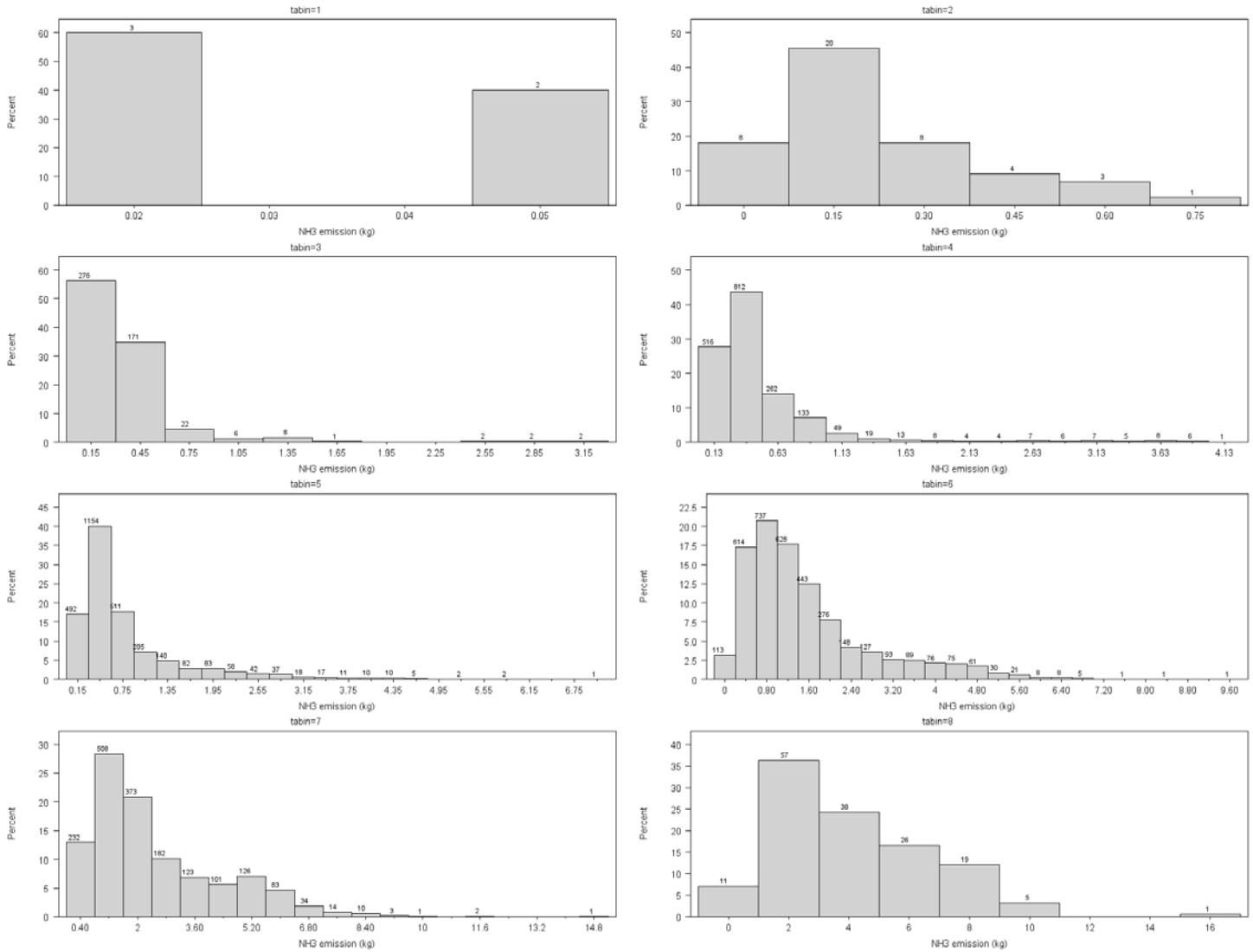
- Winter months underrepresented

Decision

- Combined swine and dairy data to learn about full range of meteorological conditions
- This does not imply that emissions from both animal types are the same

United States Environmental Protection Agency

Rationale for Gamma Distribution



United States Environmental Protection Agency

Static Farm-based Variables

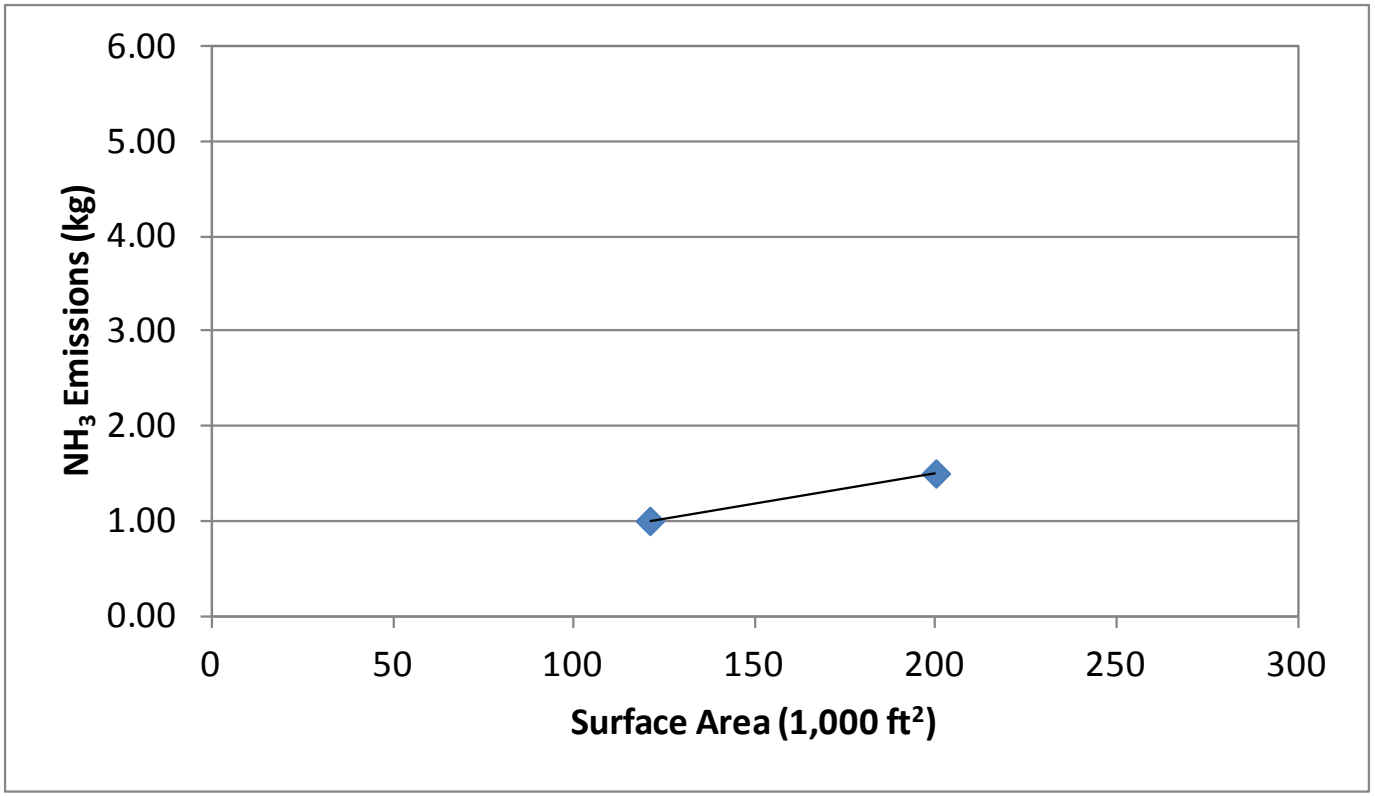
Category	Description	Predictor Variable	Units
Farm and lagoon	Animal type (Dairy or Swine indicator)	*animal*	NA
	Capacity of farm (number of animals)	*capacity*	head
	Average adult animal weight	*adultwt*	lb
	Number of confinement structures	*barns*	barns
	Manure management system	*mms*	NA
	Surface area	*sa*	1000 ft^2
	Number of manure inlets into lagoon	*inlets*	inlets
	Whether an odor control agent was used on a given day	*odorctrl*	NA

NA = not applicable

United States Environmental Protection Agency

Degrees of Freedom Challenge

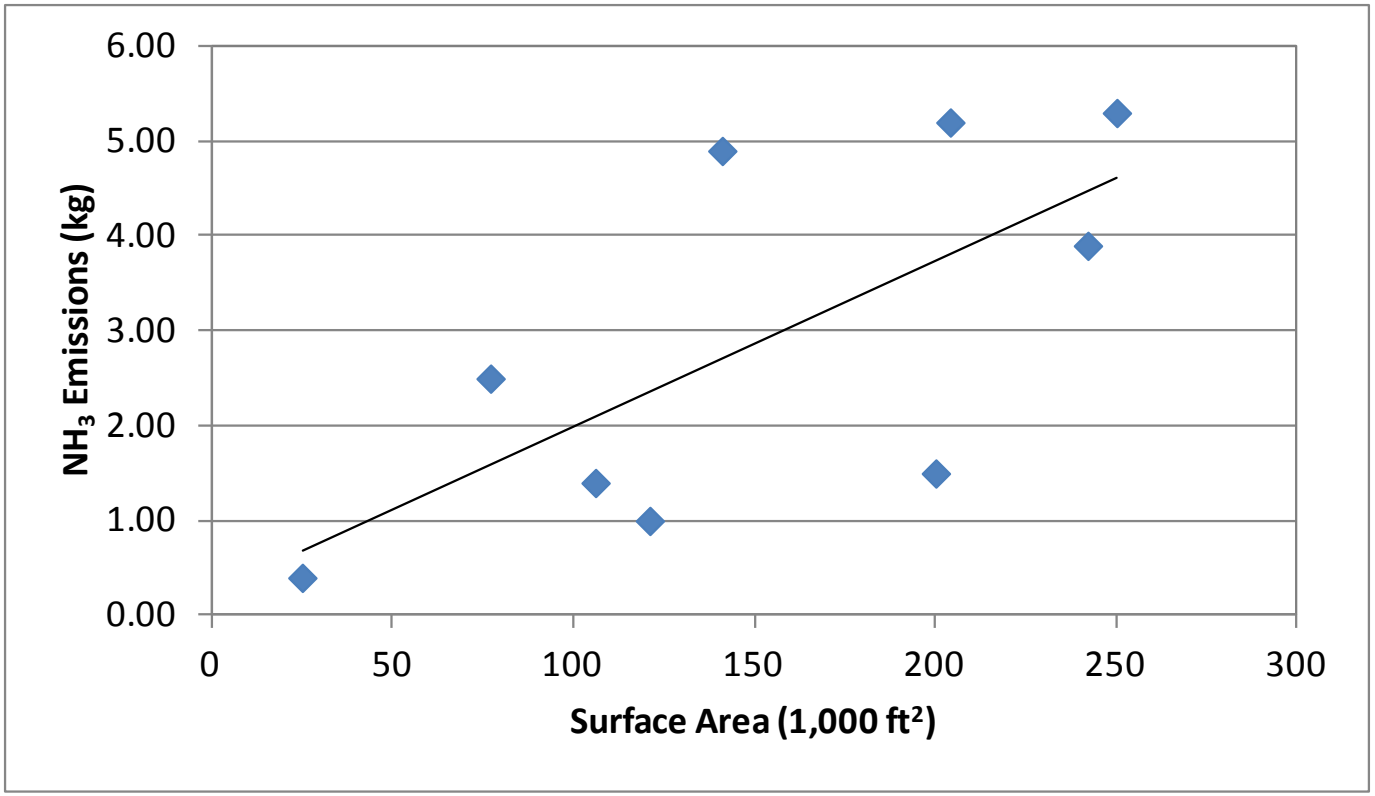
Number of predictors vs. number of sites



United States Environmental Protection Agency

Degrees of Freedom Challenge

Number of predictors vs. number of sites



United States Environmental Protection Agency

Static Farm-based Variables

Description	Predictor Variable	Units	Keep?
Animal type (Dairy or Swine indicator)	*animal*	NA	Y
Capacity of farm (number of animals)	*capacity*	head	Y
Average adult animal weight	*adultwt*	lb	Y
Number of confinement structures	*barns*	barns	N
Manure management system	*mms*	NA	N
Surface area	*sa*	1000 ft^2	Y
Number of manure inlets into lagoon	*inlets*	inlets	N
Whether an odor control agent was used on a given day	*odorctrl*	NA	N

NA = not applicable

United States Environmental Protection Agency

Static Farm-based Variables Considered

Description	Predictor Variable	Units	Keep?
Animal type (Dairy or Swine indicator)	*animal*	NA	Y
Capacity of farm (number of animals) * Average adult animal mass	*size*	head	Y
Surface area	*sa*	1000 ft^2	Y

NA = not applicable

United States Environmental Protection Agency

Additional Sub-setting

NAEMS Data Submitted to EPA:
12,854 30-min NH_3 emissions observations
13 Time-varying predictor variables
25 Static predictor variables

Data Completeness and Usability Assessment

Preliminary Dataset:
10,783 30-min NH_3 emissions observations
5 Time-varying predictor variables
8 Static predictor variables

Learnability Assessment

Full Dataset:
10,783 30-min NH_3 emissions observations
5 Time-varying predictor variables
3 Static predictor variables

Base Dataset:
8,592 30-min NH_3 emissions observations (~80%)
5 Time-varying predictor variables
3 Static predictor variables

Cross-validation Dataset:
2,191 30-min NH_3 emissions observations (~20%)
5 Time-varying predictor variables
3 Static predictor variables

United States Environmental Protection Agency

Features Considered

- Serial Correlation
  - Difficulties diagnosing
  - Difficulties fitting
- Random Effect
  - Would use one degree of freedom
- Link Function
  - Compared identity, reciprocal and log
  - Log was most appropriate

United States Environmental Protection Agency

Final NH_3 EEM Fit Statistics

Fit Statistic	Animal/ surface area	Animal/ size	Surface area/ size
Negative two log likelihood (-2LL)	3,811	3,676	3,577
Bayesian information criterion (BIC)	3,815	3,684	3,586
Percent in Prediction Interval (% in PI)	99	99	99
Prediction Interval width (kg)	4.6	4.5	4.6
Root Mean Squared Error (RMSE) (kg)	0.73	0.83	0.80
R^2	0.74	0.66	0.68

Animal/Surface Area EEM Examples

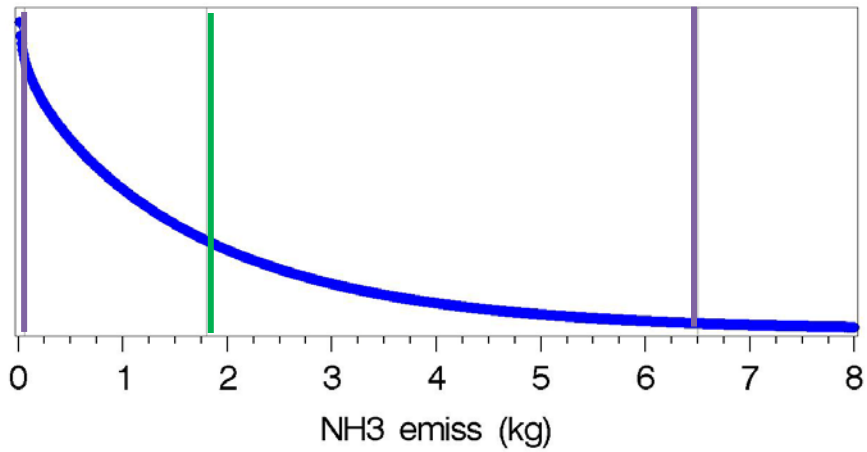
Same predictor values for a half-hour period

Date	August 30
Hour	6 p.m.
Temperature	29° C (80° F)
Humidity	40%
Wind speed	4.1 m/s (9 mi/hr)
Surface area	11,240 m^2 (121,000 ft^2)

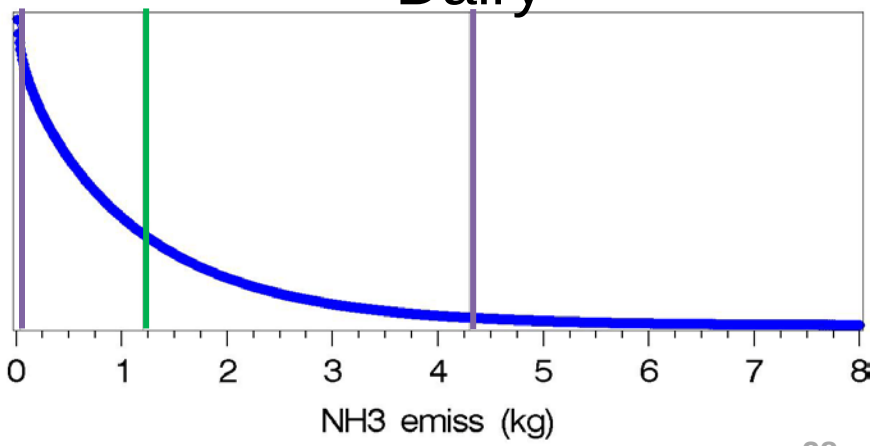
Example EEM Results

Animal Type	Swine	Dairy
Point Prediction (kg)	1.8	1.2
Lower Bound (kg)	0.055	0.036
Upper Bound (kg)	6.5	4.3

Swine



Dairy



United States Environmental Protection Agency

Summary

- Limited coverage of meteorological conditions → Combine swine and dairy data
- Limited lagoon liquid data →
  - Omit from analysis
  - Use farm-based variables as surrogates
- Degrees of freedom challenge →

  Use three most meaning full static variables

  Compare three EEMs

United States Environmental Protection Agency

SAB CHARGE QUESTIONS

United States Environmental Protection Agency

Statistical Methodology

- Statistical Methodology used to develop draft EEMs:
  - To ensure that the dataset for swine and dairy manure storage units represented all seasonal meteorological conditions, the EPA combined the swine and dairy data so that the EEM could learn about effects of all met conditions from both animals simultaneously.

- Question 2: Please comment on the EPA's decision to combine the swine and dairy data to learn about full range of meteorological conditions.

United States Environmental Protection Agency

Statistical Methodology (concl.)

- Question 3: Please comment on the use of static predictor variables as surrogates for data on the liquid conditions in manure storage units.

- Question 4: Does the SAB recommend that the EPA consider alternative approaches for developing the draft EEMs that balance the competing needs for a large dataset to reflect seasonal conditions and considering of liquid conditions?

United States Environmental Protection Agency

Negative and Zero Data

- EPA used the following general approach regarding inclusion of negative and zero emissions values in our data set
  - The EPA evaluated whether the negative or zero values represent the variability in emissions measurements due to the means of obtaining the measurements
  - The EPA included zero values because these values potentially represent instances where the emissions from the source are zero
  - The EPA reviewed the data to determine if data quality measures were properly performed according to the Quality Assurance Project Plan. If the data did not follow data quality measures, the EPA contacted the Science Advisor to determine if the corrected data could be submitted.

- Questions 5 & 6: Please comment on the EPA's approach for handling negative or zero emission measurements and provide any alternative approaches.