Appendix A: Data Processing

Table of Contents

List of Tables

A.1 Calculation of Average Daily Values for open sources

To prepare the 2012 NAEMS datasets of 30-minute values for use in calculating daily averages, EPA adjusted the NH_3 values per Table A-1. Adjustments factors were not provided in literature for the bLS NH_3 values for the WA5A and TX5A sites. In cases where 30-minute emissions flux values were available for both the bLS model and VRPM, we used the average of the bLS and VRPM values. Table A-2 presents an example calculation for two days at site IN5A (one day with both bLS and VRPM data, and one day with data from only measurement methodology).

Table A-1. Open Source Data Adjustment Factor

Site	Adjustment	Reference
IA3A	bLS/0.63	Grant and Boehm (2018)
IN4A	RPM*1.12	Grant, Boehm, and Heber (2016)
IN5A	bLS*1.19	Grant and Boehm 2020
NC3A	RPM*1.02	Grant, Boehm, and Heber (2016)
NC4A	RPM*1.57	Grant, Boehm, and Heber (2016)
OK3A	RPM*0.95	Grant, Boehm, and Heber (2016)
OK4A	RPM*1.08	Grant, Boehm, and Heber (2016)
WA5A	None	Not applicable
WI5A	bLS*1.13	Grant and Boehm 2020
TX5A	None	Not applicable

Table A-2. Example Calculation of Average NH_3 Emissions for IN5A

Date and Time	2012 NAEMS VRPM Flux Values (g/s)	2012 NAEMS bLS Flux Values (g/s)	Adjusted bLS (g/s)	Average of VRPM and Adjusted bLS 30-min Values (g/s)
9/30/08 12:15 AM	0.450	0.630	0.750	0.600
9/30/08 12:45 AM	0.410	0.491	0.584	0.497
9/30/08 1:15 AM	0.420	0.474	0.564	0.492
9/30/08 4:15 AM	0.130	0.014	0.017	0.073
9/30/08 4:45 AM		0.064	0.076	0.076
9/30/08 6:15 AM	0.080	0.027	0.032	0.056
9/30/08 6:45 AM		0.040	0.047	0.047
9/30/08 7:15 AM		0.002	0.002	0.002
9/30/08 12:15 PM		0.037	0.044	0.044
9/30/08 12:45 PM		0.038	0.045	0.045
9/30/08 1:15 PM	0.110	0.090	0.107	0.108
9/30/08 1:45 PM	0.160	0.125	0.149	0.154
9/30/08 2:15 PM	0.130	0.139	0.165	0.148
9/30/08 2:45 PM	0.200	0.213	0.253	0.227

Table A-2. Example Calculation of Average NH_3 Emissions for IN5A

Date and Time	2012 NAEMS VRPM Flux Values (g/s)	2012 NAEMS bLS Flux Values (g/s)	Adjusted bLS (g/s)	Average of VRPM and Adjusted bLS 30-min Values (g/s)
9/30/08 3:15 PM	0.140	0.166	0.198	0.169
9/30/08 4:15 PM	0.190	0.179	0.213	0.202
9/30/08 4:45 PM	0.190	0.169	0.201	0.196
9/30/08 5:15 PM	0.250	0.239	0.284	0.267
9/30/08 5:45 PM	0.240	0.236	0.281	0.260
9/30/08 6:45 PM	0.270	0.199	0.237	0.253
9/30/08 7:15 PM	0.350	0.341	0.406	0.378
9/30/08 7:45 PM		0.308	0.367	0.367
9/30/08 8:15 PM	0.370	0.560	0.666	0.518
9/30/08 9:15 PM		0.478	0.569	0.569
9/30/08 9:45 PM		0.553	0.658	0.658
9/30/08 10:15 PM	0.340	0.380	0.452	0.396
9/30/08 10:45 PM	0.370	0.479	0.570	0.470
9/30/08 11:45 PM	0.190	0.153	0.182	0.186
2/1/09 12:45 AM		0.115	0.137	0.137
2/1/09 1:15 AM		0.111	0.132	0.132
2/1/09 1:45 AM		0.071	0.085	0.085
2/1/09 2:15 AM		0.121	0.144	0.144
2/1/09 3:15 AM		0.113	0.134	0.134
2/1/09 3:45 AM		0.127	0.151	0.151
2/1/09 4:15 AM		0.111	0.132	0.132
2/1/09 4:45 AM		0.094	0.111	0.111
2/1/09 5:15 AM		0.137	0.163	0.163
2/1/09 5:45 AM		0.129	0.154	0.154
2/1/09 6:15 AM		0.154	0.183	0.183
2/1/09 6:45 AM		0.109	0.130	0.130
2/1/09 7:15 AM		0.098	0.117	0.117
2/1/09 7:45 AM		-0.006	-0.007	-0.007
2/1/09 8:15 AM		0.073	0.087	0.087
2/1/09 9:15 AM		0.085	0.101	0.101
2/1/09 9:45 AM		0.076	0.091	0.091
2/1/09 10:15 AM		0.051	0.060	0.060
2/1/09 10:45 AM		0.028	0.033	0.033
2/1/09 11:15 AM		0.013	0.015	0.015
2/1/09 11:45 AM		0.017	0.020	0.020
2/1/09 4:15 PM		0.032	0.038	0.038
2/1/09 5:15 PM		0.013	0.016	0.016
2/1/09 5:45 PM		0.004	0.005	0.005
2/1/09 6:45 PM		-0.003	-0.003	-0.003

Table A-2. Example Calculation of Average NH_3 Emissions for IN5A

Date and Time	2012 NAEMS VRPM Flux Values (g/s)	2012 NAEMS bLS Flux Values (g/s)	Adjusted bLS (g/s)	Average of VRPM and Adjusted bLS 30-min Values (g/s)
2/1/09 7:15 PM		0.007	0.009	0.009
2/1/09 7:45 PM		-0.020	-0.024	-0.024
2/1/09 8:15 PM		-0.004	-0.005	-0.005
2/1/09 9:15 PM		-0.012	-0.014	-0.014
2/1/09 10:15 PM		-0.002	-0.003	-0.003
2/1/09 10:45 PM		-0.011	-0.012	-0.012
2/1/09 11:15 PM		0.003	0.004	0.004
2/1/09 11:45 PM		0.020	0.024	0.024

Appendix B: Estimating Annual Uncertainty Association with the Application of the Daily Emission Estimating Methodology Statistical Models

Table of Contents

List of Tables

List of Figures

B.1 Introduction

Annual emissions for an individual source on a farm (e.g., barn or lagoon) are estimated by the summing the daily emissions over the year, where the daily emissions are estimated by a statistical model and appropriate daily input variables. To determine the random error or uncertainty associated with applying this methodology over a year, a non-parametric approach was originally developed. This approach has been previously described in the "Emission Estimation Methods for Animal Feeding Operations (Process Overview)" report and briefly in section 7 of each animal sector report. In response to USDA comments and also in a desire to provide a clearer description of the approach, the method for estimating annual uncertainty was reviewed by EPA. As a result of the review, a new simpler approach for determining annual uncertainty has been developed, which gives almost identical results to the original approach. This document provides a more extensive description of the original non-parametric approach and a description of the new approach, which is based on parametric principles (referred to hereafter as the parametric approach). Both non-parametric and parametric approach descriptions use the swine barn emissions-estimating methodologies (EEMs) as examples. This document also provides analysis and discussion of the two different approaches as well as some conclusions.

B.2 Approaches for Determining Annual Uncertainty

B.2.1 Original Non-Parametric Approach

The non-parametric approach starts by first determining the standard deviation of the model predicted daily emission residuals in units of mass per time. This value in conjunction with the average residual value (arbitrary for determining random error) is used to generate a normal distribution. For each day of the year, residuals from the normal distribution are randomly selected 10,000 times using the Monte Carlo simulation and are added to the predicted daily emission from the statistical model (e.g., for swine grow-finish shallow pit NH3 EEM: $\ln(NH_3) = 1.1422 + 0.0091$*Ambient Temperature + 0.0085*Live Animal Weight). The daily emission prediction is generated using a meteorological profile while holding the other predictor variables constant (e.g., live animal weight (LAW)). Daily emission predictions are then summed to produce 10,000 estimates of the annual single source total emissions. The statistics of the 10,000 annual emission estimates are then calculated, including annual mean, annual median, and the 2.5 and 97.5 percentiles. Multiple sets of 10,000 simulations (referred to as emission intervals) are run by varying one of the predictor variables. As an example of this, Table B-1 shows the Monte Carlo simulation statistics for the swine grow-finish shallow pit NH_3 model (daily residual standard deviation value (S_r) of 2.0439 kg day^{-1}) for 45 emission intervals that use different LAW input values that range from 0 to 132,000 kg for the example year of 2009 at NC3B. In Table B-1, it can be observed that the annual mean and median are very similar, which is related to the properties of a normal distribution. Furthermore, it can be observed that the % uncertainty for emission intervals decreases as emissions increase, whereas the range (difference between the 2.5 and 97.5 percentile) stays fairly constant. For each emission interval, the percent uncertainty was calculated according to Equation 1.

$$\%\,Uncertainty = 0.5 \times \left(\frac{Range}{Median\ Annual\ Emission}\right) \times 100 \qquad \text{Equation 1}$$

The % uncertainty for each emission interval was then plotted against mean annual emission with the relationship between the two variables defined as y=k/x, where k is a constant and is unknown. The k-value is then determined through regression (see Figure B-1 for an example). Once the k-value has been determined, the % annual uncertainty can be determined using the following equation:

$$\% \, Annual \; Uncertainty = \frac{k}{Annual \; Emissions} \qquad \text{Equation 2}$$

From Equation 2, the annual emission uncertainty (U_{an}) can be determined using the annual emission value:

$$U_{an} = Annual \; Emission \; \times \left(\frac{\% \, Annual \; Uncertainty}{100}\right) \qquad \text{Equation 3}$$

Applying Equations 1 and 2 to a swine grow-finish shallow pit operation that has NH_3 emissions of 1500 kg yr^{-1} would yield % annual uncertainty and U_{an} values of 5.1% and 76.28 kg, respectively, with a k-value of 7,628.4.

Analysis of S_r and U_{an} for EEMs showed a relationship between these two terms with U_{an} approximately 37.5 times greater than S_r. As an example of this, S_r and U_{an} values and their corresponding ratio for the swine barn EEMs are shown in Table B-2.

B.2.2 New Parametric Approach for Determining Annual Uncertainty

The parametric approach characterizes the random error or uncertainty in the statistical model prediction using the gaussian error of propagation. Accordingly, the annual standard deviation (S_{an}) for n days can be determined as:

$$S_{an} = \sqrt{(S_{r1})^2 + (S_{r2})^2 + \ldots (S_{rn})^2} \qquad \text{Equation 4}$$

where S_r is the same daily residual standard deviation used in the Monte Carlo simulations. If S_r is the same value for each day (i.e., $S_{r1} = S_{r2} = S_{rn}$), Equation 4 simplifies to:

$$S_{an} = S_r n^{0.5} \qquad \text{Equation 5}$$

Thus, for the swine grow-finish shallow pit NH_3 EEM with a S_r value of 2.0439 kg day^{-1} and a n value of 365, S_{an} is determined as follows:

$$S_{an} = 2.0439 \times (365^{0.5}) = 39.05 \text{ kg} \qquad \text{Equation 6}$$

In the previously described non-parametric approach, 95% of residual distribution was considered (i.e., the range was the difference between the 97.5 and 2.5 percentiles). If we make this approach analogous to the non-parametric approach by applying similarly it to 95% of the distribution or 1.96 standard deviations, the annual uncertainty (U_{an}) can be approximated to:

$$U_{an} \approx 1.96 \, S_{an} \qquad \text{Equation 7}$$

Thus, if applying the S_{an} value from Equation 6 for the swine grow-finish shallow pit NH_3 EEM:

$$U_{an} \approx 1.96 \, S_{an} \approx 1.96 \times 39.05 \approx 76.54 \text{ kg} \qquad \text{Equation 8}$$

In addition, combining Equation 7 and Equation 5 with an n-value of 365 representing the number of days in the annual uncertainty calculation yields:

$$U_{an} \approx 1.96\, S_{an} \approx 1.96\, S_r n^{0.5} \approx 1.96\, S_r * (365)^{0.5} \approx 37.45\, S_r \qquad \text{Equation 9}$$

The parametric approach was used to determine U_{an} for other EEMs. As an example, the results for the swine barns EEMs are shown in Table B-3.

B.3 Analysis and Discussion

Using swine barn EEMs as an example, the non-parametric and parametric approaches for determining U_{an} produce extremely similar values, with the % difference being less than 0.5% (see Tables B-2 and B-3). The slight difference is due to the nature of the Monte-Carlo simulations used in the non-parametric method, which has intrinsic variation. The observed ratio between U_{an} and S_r values in the non-parametric approach for EEMs (Table B-2) is extremely similar to the relationship shown in Equation 9. As highlighted when describing the non-parametric approach, creating a normal distribution based on residual standard deviation produces similar mean and median values and thus their application in the non-parametric and parametric approaches produces similar estimates of annual uncertainty.

The parametric approach was further tested by combining Equations 5 and 7 so that:

$$U_{an} \approx 1.96\, S_r n^{0.5} \qquad \text{Equation 10}$$

or

$$\frac{U_{an}}{S_r} \approx 1.96 n^{0.5} \qquad \text{Equation 11}$$

Equation 11 was tested by determining U_{an} for different number of simulated days (n) using the Monte Carlo method. For a range of different n-values (1-365) and a constant S_r value, a 1,000 simulations were run using the same Monte Carlo method as used previously for the emission intervals. The $\frac{U_{an}}{S_r}$ values were plotted against n with a power function line of best fit fitted to the data (Figure B-2). It can be observed that the values for the equation determined by the power function line of best fit match the constants in Equation 11. This analysis supports the validity of the parametric approach for determining annual uncertainty.

B.4 Conclusions

The original non-parametric and new parametric approach yield almost identical estimates of annual uncertainty, however, the advantage of the parametric approach over the non-parametric approach is that it is easier to apply and understand. In addition, the parametric approach can easily be adapted to estimate uncertainty for a different number of days if needed (see Equation 10). Accordingly, EPA will adopt the parametric approach for estimating annual uncertainty associated with EEM statistical models. It should be noted that both approaches assume there is no relationship between the magnitude of the residual and the magnitude of emissions, and thus the approach when applied will provide an estimate of annual uncertainty that is independent of a barn's emission magnitude. As a result, the annual uncertainty for a farm will be dependent on the daily residual standard deviation of the corresponding EEM model and the number of individual sources (e.g., barns and lagoons) on a farm (see Equation 13) in "Emission Estimation Methods for Animal Feeding Operations (Process Overview)" report). Furthermore, it should be noted that both approaches do not consider systematic error (e.g., bias) in the EEM statistical models.

Table B-1. Monte Carlo simulation statistics and calculations for the swine grow-finish shallow pit NH_3 model for 45 emission intervals with different LAW input values. Each emission interval or record/line is from 10,000 simulations.

LAW (Mg)	Mean (kg)	Median (kg)	2.5 % (kg)	97.5 % (kf)	Range (kg)	% Uncertainty
0	599.05	598.93	522.95	675.84	152.89	12.76
3	634.57	634.49	559.72	711.85	152.12	11.99
6	670.70	670.50	594.05	746.47	152.42	11.37
9	706.64	706.78	629.14	783.93	154.79	10.95
12	745.00	744.93	671.12	820.96	149.84	10.06
15	783.68	783.46	707.13	858.69	151.55	9.67
18	823.53	823.55	748.11	899.95	151.84	9.22
21	864.07	863.88	789.25	940.53	151.28	8.76
24	906.15	906.41	830.27	982.96	152.69	8.42
27	950.09	949.99	874.40	1025.91	151.51	7.97
30	992.62	992.78	915.89	1067.82	151.93	7.65
33	1038.46	1038.72	961.45	1114.74	153.29	7.38
36	1084.53	1084.53	1008.09	1160.31	152.23	7.02
39	1132.48	1132.68	1056.09	1209.66	153.57	6.78
42	1181.08	1180.68	1104.17	1258.65	154.49	6.54
45	1231.46	1231.18	1155.79	1307.29	151.51	6.15
48	1282.21	1282.76	1205.13	1359.16	154.03	6.00
51	1335.10	1335.57	1258.19	1410.30	152.12	5.69
54	1389.23	1388.97	1311.69	1466.46	154.76	5.57
57	1444.51	1444.70	1368.57	1521.28	152.71	5.29
60	1502.36	1502.59	1426.85	1577.39	150.54	5.01
63	1560.19	1560.19	1482.81	1635.78	152.97	4.90
66	1620.61	1620.34	1545.01	1697.99	152.98	4.72
69	1682.46	1682.47	1606.43	1758.14	151.71	4.51
72	1744.62	1744.49	1667.06	1821.06	154.00	4.41
75	1809.97	1810.44	1732.07	1884.86	152.79	4.22
78	1875.81	1876.48	1798.65	1951.74	153.09	4.08
81	1943.96	1943.73	1867.14	2019.72	152.58	3.92
84	2013.79	2013.67	1939.27	2091.08	151.81	3.77
87	2085.73	2085.65	2008.39	2163.04	154.65	3.71
90	2159.63	2160.21	2083.89	2234.94	151.06	3.50
93	2234.82	2234.86	2157.13	2311.34	154.21	3.45
96	2311.99	2312.15	2235.66	2389.71	154.05	3.33
99	2391.27	2390.90	2315.83	2468.78	152.95	3.20
102	2473.55	2473.60	2396.41	2549.72	153.31	3.10
105	2557.14	2557.61	2479.78	2632.16	152.39	2.98

LAW (Mg)	Mean (kg)	Median (kg)	2.5 % (kg)	97.5 % (kf)	Range (kg)	% Uncertainty
108	2642.76	2642.65	2566.85	2720.30	153.45	2.90
111	2730.36	2730.57	2654.02	2806.76	152.74	2.80
114	2820.53	2820.25	2742.48	2897.84	155.36	2.75
117	2912.99	2913.01	2836.55	2989.45	152.90	2.62
120	3008.60	3008.65	2933.02	3085.04	152.02	2.53
123	3105.88	3105.76	3029.06	3182.53	153.46	2.47
126	3206.97	3206.59	3130.43	3284.60	154.18	2.40
129	3309.12	3309.54	3232.44	3385.15	152.71	2.31
132	3413.73	3414.15	3337.60	3488.32	150.72	2.21

Table B-2. Daily residual standard deviation (S_r) and corresponding annual uncertainty (U_{an}) using the non-parametric approach for swine barn EEMs.

Swine Barn Source	Pollutant	S_r[a] (g day^{-1} or kg day^{-1})	U_{an}[b] (g or kg)	U_{an}-S_r Ratio
Farrowing Room	H_2S	111.4	4176.3	37.5
Farrowing Room	NH_3	0.2	7.5	37.4
Farrowing Room	PM_{10}	18.4	691.7	37.6
Farrowing Room	$PM_{2.5}$	2.6	98.1	37.6
Farrowing Room	TSP	53.8	2023.5	37.6
Gestation Barn	H_2S	3292.6	123570.9	37.5
Gestation Barn	NH_3	10.3	385.0	37.5
Gestation Barn	PM_{10}	182.2	6834.1	37.5
Gestation Barn	$PM_{2.5}$	23.9	898.3	37.6
Gestation Barn	TSP	351.9	13204.9	37.5
Gestation Barn-Deep pit	H_2S	2876.3	107955.6	37.5
Gestation Barn-Deep pit	NH_3	9.1	342.1	37.5
Gestation Barn-Shallow Pit	H_2S	2876.3	107637.5	37.4
Gestation Barn-Shallow Pit	NH_3	9.1	342.0	37.5
Grow-Finish Barn	H_2S	423.5	15887.0	37.5
Grow-Finish Barn	NH_3	2.1	79.2	37.6
Grow-Finish Barn	PM_{10}	103.2	3860.5	37.4
Grow-Finish Barn	$PM_{2.5}$	8.3	311.8	37.4
Grow-Finish Barn	TSP	303.5	11386.1	37.5
Grow-Finish Barn- Deep Pit	H_2S	405.0	15203.6	37.5
Grow-Finish Barn- Deep Pit	NH_3	2.0	76.8	37.6
Grow-Finish Barn- Shallow Pit	H_2S	405.0	15153.9	37.4
Grow-Finish Barn- Shallow Pit	NH_3	2.0	76.5	37.4

[a] Units are kg day^{-1} for NH_3 and g day^{-1} for all other pollutants.

[b] Units are kg for NH_3 and g for all other pollutants.

Table B-3. Daily residual standard deviation (S_r) and corresponding annual uncertainty (U_{an}) using the parametric approach for swine barn EEMs.

Swine Barn Source	Pollutant	S_r^a (g day^{-1} or kg day^{-1})	U_{an}^b (g or kg)
Swine Farrowing Room	H_2S	111.4	4171.4
Swine Farrowing Room	NH_3	0.2	7.5
Swine Farrowing Room	PM_{10}	18.4	689.1
Swine Farrowing Room	$PM_{2.5}$	2.6	97.7
Swine Farrowing Room	TSP	53.8	2016.4
Gestation Barn	H_2S	3292.6	123292.0
Gestation Barn	NH_3	10.3	384.5
Gestation Barn	PM_{10}	182.2	6824.5
Gestation Barn	$PM_{2.5}$	23.9	894.5
Gestation Barn	TSP	351.9	13176.0
Gestation Barn-Deep pit	H_2S	2876.3	107706.0
Gestation Barn-Deep pit	NH_3	9.1	341.7
Gestation Barn-Shallow Pit	H_2S	2876.3	107706.0
Gestation Barn-Shallow Pit	NH_3	9.1	341.7
Grow-Finish Barn	H_2S	423.5	15857.5
Grow-Finish Barn	NH_3	2.1	79.0
Grow-Finish Barn	PM_{10}	103.2	3866.1
Grow-Finish Barn	$PM_{2.5}$	8.3	312.1
Grow-Finish Barn	TSP	303.5	11363.0
Grow-Finish Barn, Deep Pit	H_2S	405.0	15164.2
Grow-Finish Barn, Deep Pit	NH_3	2.0	76.5
Grow-Finish Barn, Shallow Pit	H_2S	405.0	15164.2
Grow-Finish Barn, Shallow Pit	NH_3	2.0	76.5

[a] Units are kg day^{-1} for NH_3 and g day^{-1} for all other pollutants.

[b] Units are kg for NH_3 and g for all other pollutants.

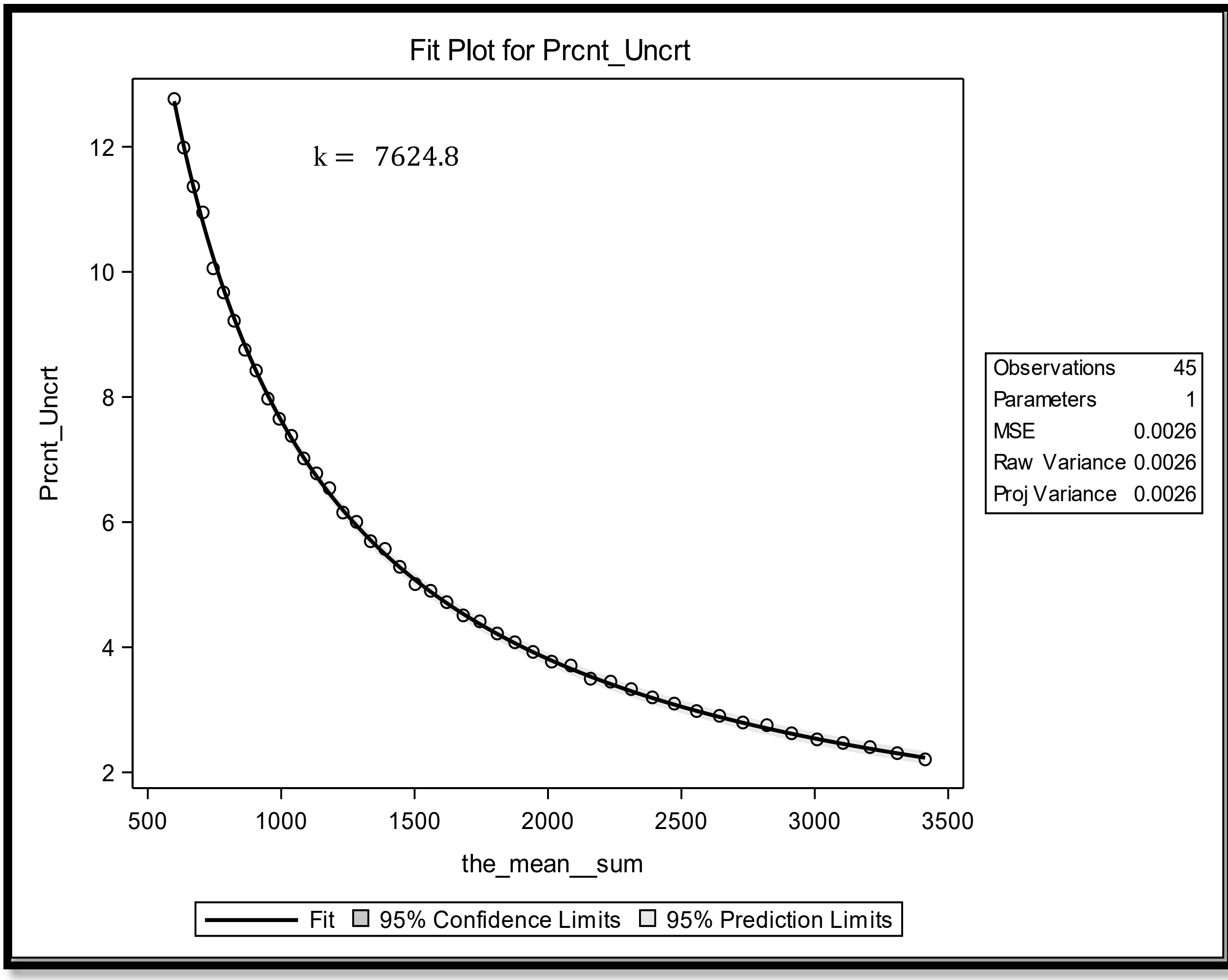

Figure B-1. y=k/x relationship (k=7624.8) between mean annual emissions and % uncertainty for the swine grow-finish shallow pit NH_3 EEM.

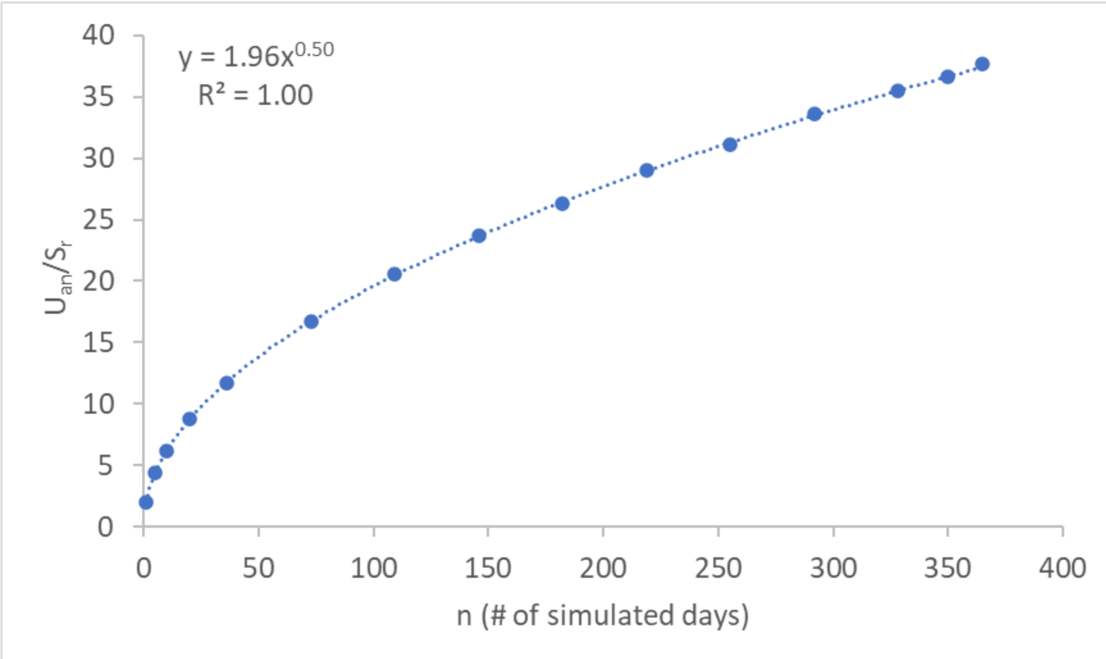

Figure B-2. Relationship (power function line of best fit) between the ratio of annual uncertainty to residual standard deviation ($\frac{U_{an}}{S_r}$) and the number of simulated days using a constant S_r value.