ATTACHMENT 1

QUANTITATIVE POISSON REGRESSION MODELING

Conceptual Model

In this approach, the concentration values for LA in ABS air observed at any particular location are assumed to depend on a wide range of random variables, resulting in a distribution of values. The distribution of values at a location is bounded by zero (concentrations cannot be less than zero), and are generally observed to be right skewed. For mathematical convenience, this type of data set is assumed to be reasonably well characterized by a gamma distribution[1]:

$$C \sim \text{gamma}(k, \Theta)$$

where:

C = Asbestos concentration (f/cc)
k = shape parameter
Θ = scale parameter

The mean and variance of a gamma distribution are as follows::

$$\text{Mean} = k \cdot \Theta$$
$$\text{Variance} = k \cdot \Theta^2$$

The mean of multiple ABS air samples collected at location “j” is assumed to be linearly dependent on the level of LA in the soil at that location:

$$\text{Mean}_j = a + b \cdot x_j$$

In all cases, a and b are unknown constants and x_j is the value of the soil metric at location "j". The value of the shape parameter is assumed to be constant between locations. This is analogous to assuming a constant coefficient of variation (stdev / mean = constant)

The goal of the analysis is to fit the model to the available ABS data in order to derive estimates of the model parameters (a, b, and k).

[1] A gamma distribution is similar to a lognormal distribution, but is not quite so "tail-heavy". EPA ProUCL guidance generally prefers the gamma distribution to the lognormal distribution for most right-skewed environmental datasets.

Poisson Variability in Air Values

Given a random sample "i" of air drawn from C ~ gamma(k,Θ), the average number of LA fibers that will be observed in that sample during TEM analysis depends on true concentration (c_i, f/cc) in that sample and on the volume of air (cc) that is analyzed:

$$\lambda_i = c_i \cdot v_i$$

The volume of air analyzed (v_i) is not an intrinsic parameter, but is determined both by the volume of air passed through the filter during sample collection and by the fraction of the filter that is analyzed:

$$v_i = V_i \cdot 1000 \cdot \frac{GO_i \cdot Ago}{EFA}$$

where:

EFA = Effective filter area (mm^2)
GO_i = Number of grid openings evaluated during the analysis
Ago = Area of one grid opening (mm^2)
V_i = Volume of air drawn through the filter (L)
1000 = cc per L

The actual number of fibers y_i observed during the analysis of a sample with expected average counts of λ_i is a Poisson random variable:

$$y_i \sim Poisson(\lambda_i) = \frac{\lambda^y}{y!}\exp(-\lambda) \qquad y = 0, 1, 2, \ldots$$

Data

The data used for fitting includes the full set of 3 ABS samples from 75 locations for each of two rounds (total = 450 samples). For each sample, the data are:

$y_{i,j}$ = number of PCME LA structures observed in sample i from location j
$v_{i,j}$ = volume of air analyzed for sample i from location j
x_j = soil metric for location j

Maximum Likelihood Fitting

In the Maximum Likelihood Estimation (MLE) approach, the likelihood of observing the data is evaluated as a function of the model parameters (a, b, and k), and the parameter set that yields the highest likelihood is identified as the best estimate of the true values. For the Poisson-gamma distribution, the log-likelihood (LL) of a given observation is given by:

$$LL = z1 + z2 + z3$$

where:

$$z1 = \ln[\Gamma(y+k)] - \ln[\Gamma(k)] - \ln[\Gamma(y+1)]$$
$$z2 = y \cdot [\ln(y) - \ln(k+\lambda]$$
$$z3 = -k \cdot [\ln(k+\lambda) - \ln(k)]$$

and:

Γ = Gamma funtion
y = number of PCME LA structures observed
λ = expected number of structures = $(a + b \cdot x) \cdot v$

The parameter values are found by maximizing the sum of the LL values across all samples.

Results

The results of this model fitting effort are provided in Figures A1-1 to A1-6 (one figure per data set). The MLE parameters for each data set are summarized below:

Data Set	Soil Metric	ABS Round	Parameter Value		
			a	b	k
1	Visible detection frequency	1 + 2	8.88E-03	2.76E-01	1.53E-01
2		1	1.35E-02	6.33E-01	1.96E-01
3		2	4.07E-03	3.34E-02	1.57E-01
4	Visible score	1 + 2	9.33E-03	2.12E-01	1.47E-01
5		1	1.40E-02	5.18E-01	1.87E-01
6		2	4.16E-03	1.85E-02	1.57E-01

The MLE parameters across both ABS rounds are utilized in the screening level risk characterization (rows are shaded in grey).

Because simple visual inspection of the data reveal that the relationship between LA levels in soil and in ABS is not strong, a second model fitting effort was performed in which the value of b was constrained to be zero. This is analogous to assuming that the concentration of PCME LA in ABS air does not depend on the level of LA in soil. The fits of the two models (with and without an assumption of soil dependency) were compared using Akaike's Information Criterion (AIC):

$$AIC = -2P - 2L$$

where:

P = Number of parameters in the model
L = Log-likelihood

The results, expressed as the difference between the two AIC values, are shown below:

Data Set	Δ AIC
1	83
2	64
3	19
4	66
5	54
6	18

In general, differences in AIC on the order of 2 or more indicate that there is a significant difference in model fit between the two models being evaluated. As seen, in all cases the difference in AIC is much higher than this, indicating that the assumption that levels of PCME LA in ABS air are a function of LA levels in soil is appropriate.

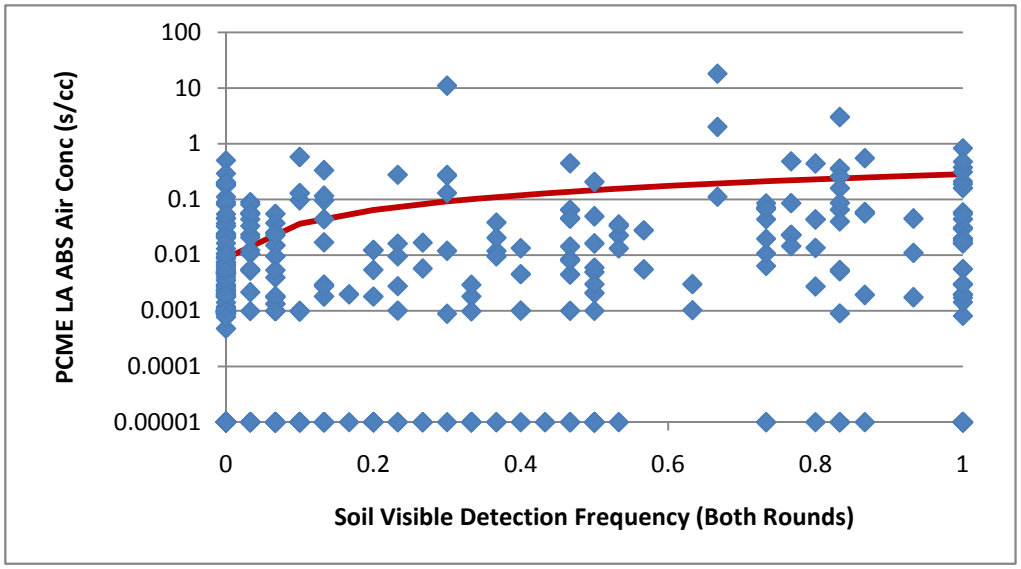
REFERENCES

Handbook of Poisson Distribution
Frank A. Haight
John Wiley & Sons
1967

Regression Analysis of Count Data
A. Colin Cameron & Pravin K. Trivedi
Cambridge University Press
1998

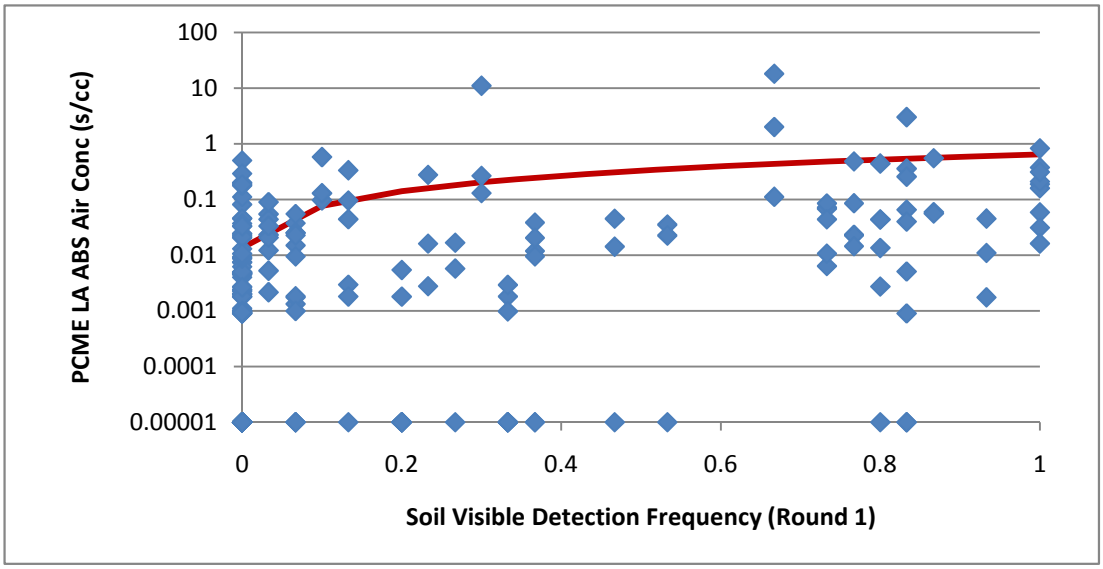
Quantitative Microbial Risk Assessment
Charles Haas, Joan B. Rose, Charles P. Gerba.
John Wiley & Sons, Inc.
1999

FIGURE A1-1. SOIL VISIBLE DETECTION FREQUENCY (BOTH ROUNDS) vs. PCME LA ABS AIR CONCENTRATION



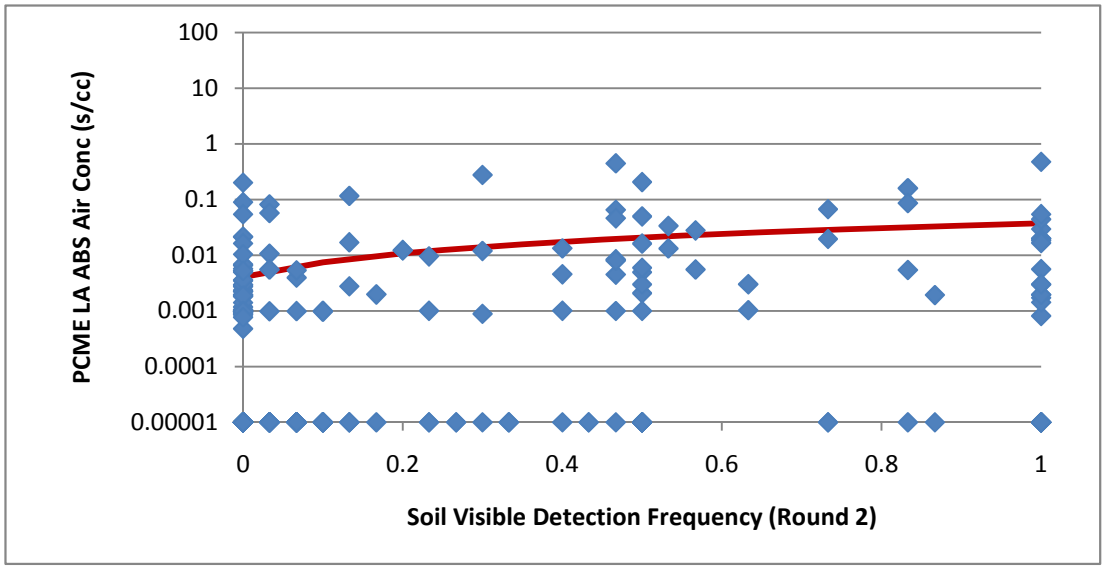
Non-detects are displayed at 0.00001 s/cc.

FIGURE A1-2. SOIL VISIBLE DETECTION FREQUENCY (ROUND 1) vs. PCME LA ABS AIR CONCENTRATION



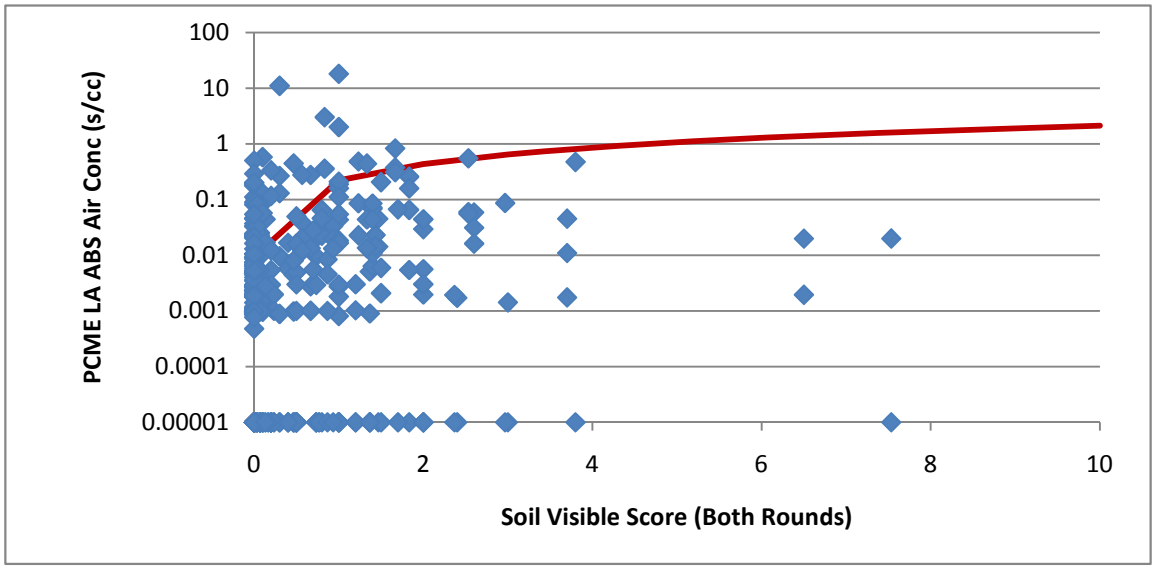
Non-detects are displayed at 0.00001 s/cc.

FIGURE A1-3. SOIL VISIBLE DETECTION FREQUENCY (ROUND 2) vs. PCME LA ABS AIR CONCENTRATION



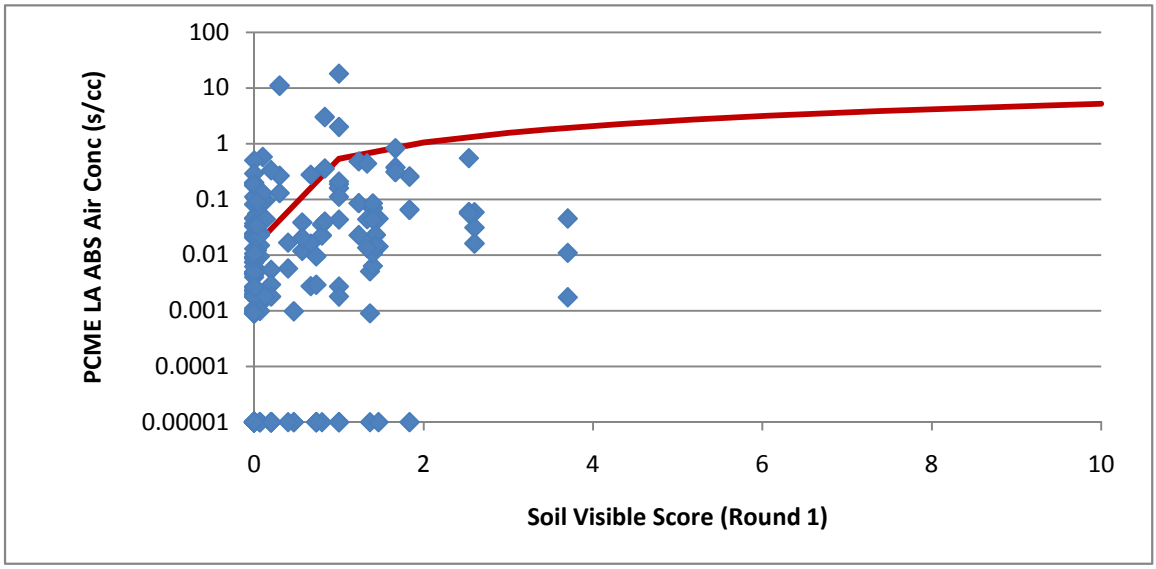
Non-detects are displayed at 0.00001 s/cc.

FIGURE A1-4. SOIL VISIBLE SCORE (BOTH ROUNDS) vs. PCME LA ABS AIR CONCENTRATION



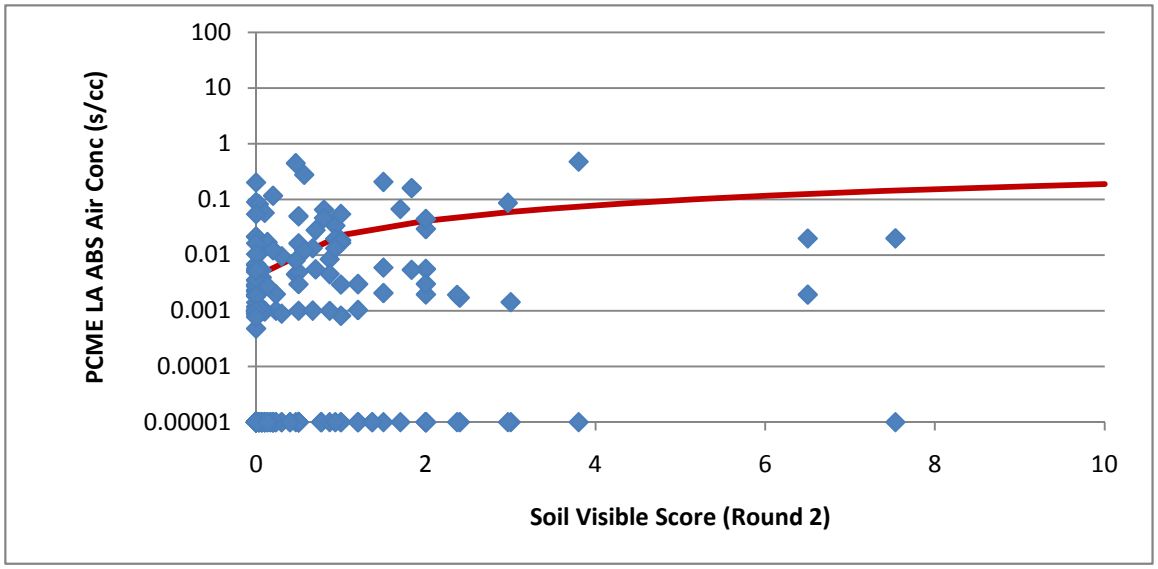
Non-detects are displayed at 0.00001 s/cc.

FIGURE A1-5. SOIL VISIBLE SCORE (ROUND 1) vs. PCME LA ABS AIR CONCENTRATION



Non-detects are displayed at 0.00001 s/cc.

FIGURE A1-6. SOIL VISIBLE SCORE (ROUND 2) vs. PCME LA ABS AIR CONCENTRATION



Non-detects are displayed at 0.00001 s/cc.