



MEMORANDUM

SUBJECT: A Methodology For Incorporating Short-Term Variable Background Concentrations In Risk Assessments

FROM: John Langstaff *John Langstaff*

TO: PM NAAQS Review Docket (OAR-2001-0017)

DATE: December 17, 2004

One of the analyses we perform as part of our PM risk assessment is a calculation of health risks assuming that the non-anthropogenic background concentrations fluctuate daily. Assigning daily background concentrations to observed concentrations for our risk assessments has been problematic. If we randomly assign background concentrations according to their distribution, sometimes the background value will be higher than the observed value. Easy fixes to this unrealistically distort the distribution of concentrations.

This memorandum describes a methodology for specifying the short-term background concentrations in a more realistic way. The key idea of this method is to use the available distributional information for the observed and background concentrations to estimate their joint distribution, which yields the distribution of the background concentrations conditioned on the level of the observed concentrations. Then to assign a background value to an observed concentration, we randomly select a value from the conditional distribution corresponding to the observed value. We also describe an application of this method to daily PM_{2.5} concentrations in an urban area.

Description of Method

For PM we can argue heuristically that background and non-background concentrations are not highly correlated since they derive from independent processes, although meteorology can induce some correlation. However, we can specify some degree of correlation for an analysis of sensitivity to this factor.

For ease of presentation we will refer to the non-background concentrations as anthropogenic concentrations in the remainder of this memorandum.

Denote by A, B, C the random variables and by f_A, f_B, f_C the probability distribution functions for the anthropogenic, background, and observed concentrations, respectively. Denote by μ_A, μ_B, μ_C the means and by $\sigma_A, \sigma_B, \sigma_C$ the standard deviations of A, B, C , and let ρ be the Pearson's correlation between A and B .

We have $C = A + B$ (1)

therefore $\mu_C = \mu_A + \mu_B$ (2)

and $\sigma_C^2 = \sigma_A^2 + \sigma_B^2 + 2\rho\sigma_A\sigma_B$ (3)

Let $f_{B|C=c_0}$ denote the distribution of B conditioned on $C=c_0$. In other words, suppose one looks at the all of the values of background concentrations at those times when the observed concentration is close to c_0 . $f_{B|C=c_0}$ is the probability distribution of those values. If background were a constant b_0 , then $f_{B|C=c_0}$ would just be b_0 (with probability 1), and there wouldn't be a distribution of different daily values.

We model f_A and f_B by 3-parameter gamma distributions of the form

$$f(x) = \frac{\left(\frac{x-\tau}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x-\tau}{\beta}\right)}{\beta \Gamma(\alpha)}, \quad x \geq \tau; \quad \alpha, \beta > 0, \quad (4)$$

where Γ is the gamma function, α is the shape parameter, β is the scale parameter, and τ is the location (threshold) parameter.

We find better fits to $PM_{2.5}$ data when the location parameter τ is taken to be zero (instead of estimated from the data) and so we do that here. We estimate the mean and standard deviation μ_C and σ_C of the observed concentrations by the sample mean and standard deviation of the observed concentrations. We assume that we have in hand estimates of the mean and standard deviation of the short-term background concentrations, μ_B and σ_B , and that we also have an estimate of the correlation ρ between the anthropogenic and background concentrations.

We use a Monte Carlo approach to estimate the joint distribution of B and C , in the following way. First, we generate a large sample from the joint distribution of f_A and f_B : $\{a_i\}$ and $\{b_i\}$ (with correlation ρ), which is fully specified by equations 1-4. Then we set $c_i = a_i + b_i$, generating the joint sample $\{a_i, b_i, c_i\}$. We can then estimate the conditional distribution $f_{B|C=c_0}$ nonparametrically directly from this.

To assign a short-term background value to an observed concentration x for the risk assessment, we randomly sample a value from the distribution $f_{B|C=x}$.

Application to Ambient Daily-Average $PM_{2.5}$ Concentrations in St. Louis

For the health risk assessment in support of the PM NAAQS (Abt, 2005), we have a time series C of 365 monitor-composite daily averages of $PM_{2.5}$ concentrations in St. Louis for the year 2003 for which we want to generate a series of background concentrations B . The sample

mean μ_C is 14.02 and the sample standard deviation σ_C is $7.09 \mu\text{g}/\text{m}^3$. Fitting a gamma distribution to these data using maximum likelihood estimation (MLE) produces the parameter estimates $\alpha_C = 4.11$ and $\beta_C = 3.41$. Figure 1 shows the fitted gamma distribution function overlaid on a histogram of the observed concentrations. We are assuming that the mean μ_B and standard deviation σ_B of the daily background concentrations of $\text{PM}_{2.5}$ are 3.5 and $2.5 \mu\text{g}/\text{m}^3$ respectively. This mean value is provided by the PM CD (EPA, 2004) and referenced in the draft PM Staff Paper (EPA, 2005). The basis for the standard deviation is given in Langstaff (2005).

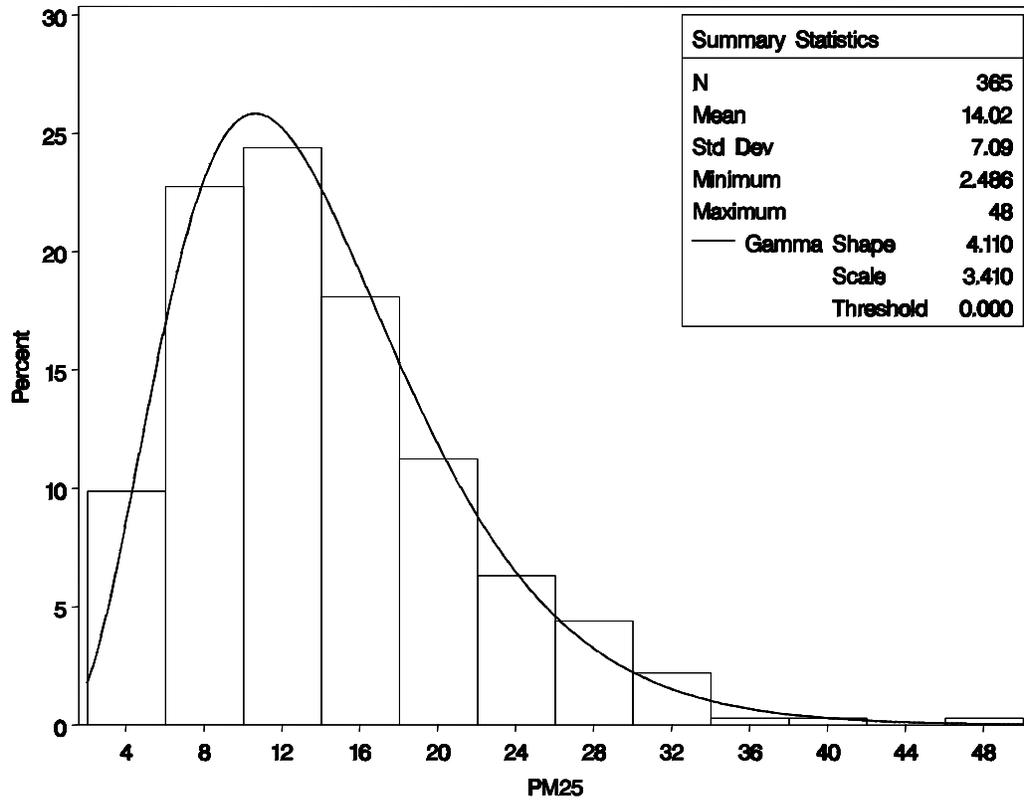


Figure 1. The distribution of the observed $\text{PM}_{2.5}$ concentrations in St. Louis, with fitted gamma pdf ($\mu\text{g}/\text{m}^3$)

We estimate the parameters of the gamma distribution of background concentrations by the method of moments:

$$\alpha_B = (\mu_B / \sigma_B)^2 = 1.96 \quad (5)$$

$$\beta_B = \mu_B / \alpha_B = 1.79 \quad (6)$$

To estimate the gamma parameters for A, we set

$$\mu_A = \mu_C - \mu_B = 14.02 - 3.5 = 10.52 \quad (7)$$

and find values of α_A and β_A such that

$$\beta_A = \mu_A / \alpha_A \quad (8)$$

and $C = A + B$ has a gamma(α_C, β_C) distribution.

We do this by generating sets of series A and B (with correlation $\rho = 0$) for different values of α_A ; for each of these form $C=A+B$; then calculate MLE estimates of α_C and β_C ; and then choose α_A such that α_C and β_C are the same as the gamma parameters 4.11 and 3.41 for the observed $PM_{2.5}$ data. (See Figures 2 and 3.)

Since A, B, and C are not distributed exactly as gamma, we estimate α_A in this way instead of analytically. Then we set $\beta_A = \mu_A / \alpha_A$. This results in $\alpha_A = 2.59$ and $\beta_A = 4.06$.

At this point we have parametric forms of the distributions of A and B, and we generate a large number of independent variates $\{a_i\}$ and $\{b_i\}$ from these distributions.

To assign a short-term background value to an observed concentration x for the risk assessment, we randomly sample a value from the distribution $f_B|C=x$. We do this by subdividing the range of the observed concentrations into nonoverlapping bins G_1, G_2 , etc. We used bin sizes of 0.2 from 0 to 20, 0.5 from 20 to 30, and 1.0 above 30 ($\mu\text{g}/\text{m}^3$). Then for each observed x , we identify the bin G_x that it falls in, and randomly sample one of the b_i from the set $\{a_i, b_i, c_i\}$ where c_i falls in G_x . This yields a series $\{b_j^*, j=1 \text{ to } 365\}$, which we refer to as a "realization." Note that in order for this method to be stable, there needs to be a sufficient number of values in each of the bins G_i . We generated three million values for each of A, B, and C, which resulted in each bin having at least 440 values. Sensitivity analyses have shown this to be sufficient to characterize the joint distribution of A and B.

Although $\{b_i\}$ has mean 3.5 and standard deviation 2.5, any realization $\{b_j^*\}$ sampled from $\{b_i\}$ is likely to have mean and standard deviation that differ from these values (as a result of sampling variability). Figure 4 illustrates this variability for 240 realizations of B with distributions of the means, standard deviations, minima and maxima of these realizations. The variability of the corresponding realizations of A and of the correlations of A and B are illustrated in Figures 5 and 6. We require the mean of the selected realization $\{b_j^*\}$ to be 3.5, the standard deviation to be close to 2.5, and the correlation between $\{a_j^*\}$ and $\{b_j^*\}$ to be near zero, where $a_j^* = x_j - b_j^*$, and x_j is the j^{th} observed concentration. We generate 240 realizations, narrow these down to realizations with mean and standard deviation within one percent of the observed $PM_{2.5}$ statistics and correlation with magnitude less than 0.1 (Table 1), and from these select the background series to be the realization with mean closest to 3.5.

The distributions of the selected background and anthropogenic concentration series are shown in Figures 7 and 8.

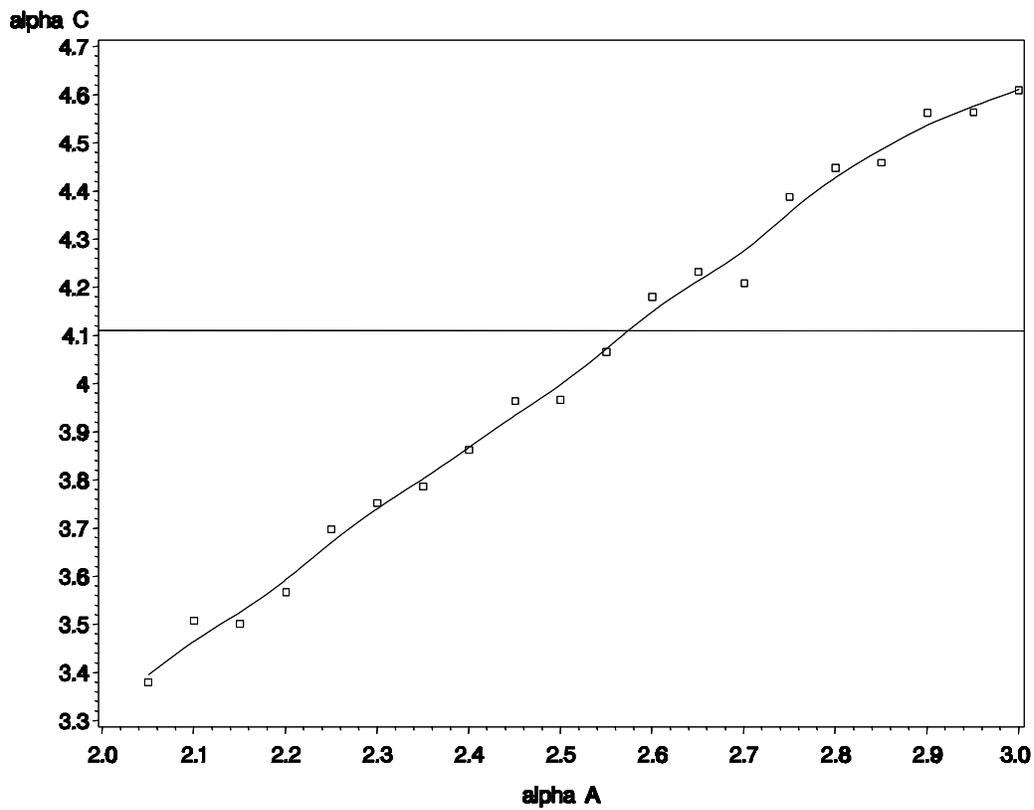


Figure 2. MLE estimates of α_C as a function of α_A

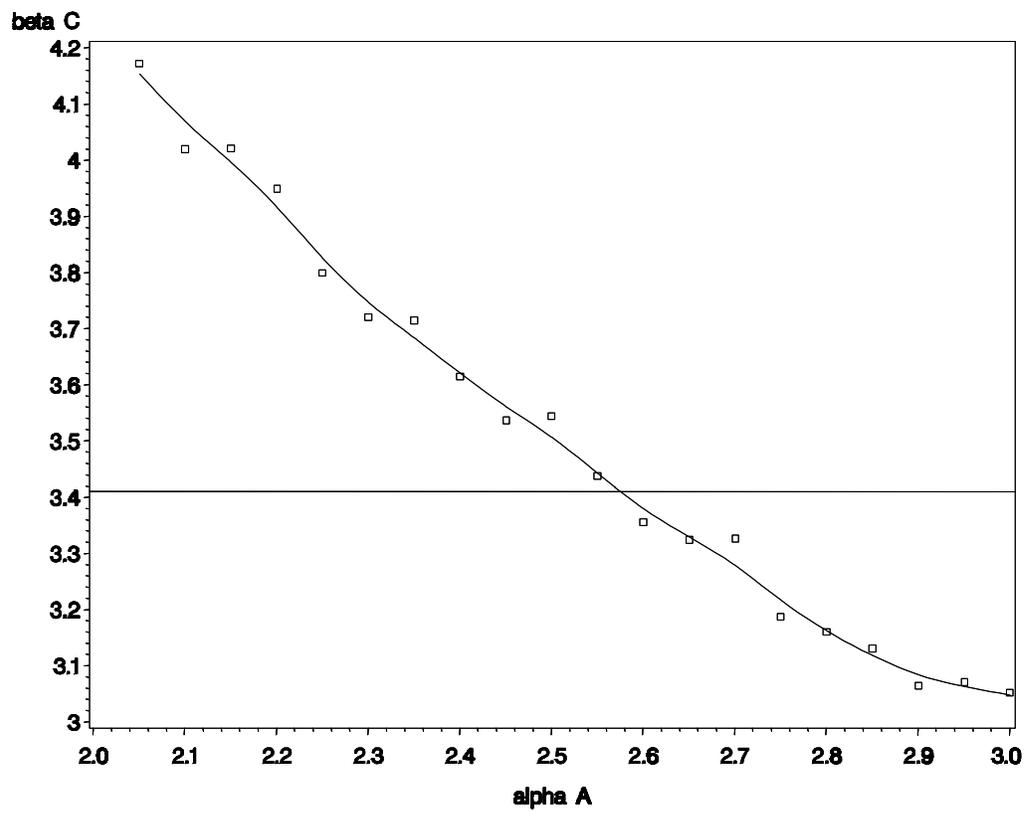


Figure 3. MLE estimates of β_C as a function of α_A

Table 1. Realizations with mean and standard deviation within one percent of the observed PM_{2.5} statistics and correlation with magnitude less than 0.1. Realization index (j), means and standard deviations (µg/m³) of the realization with the percent differences between the realizations and the observed data, and the correlations between A and B are tabulated.

j	Mean of B	Percent diff	St Dev of B	Percent diff	Correlation of A and B
153	3.474	.75%	2.517	(.66%)	-0.087
219	3.477	.66%	2.522	(.88%)	0.012
216	3.482	.53%	2.520	(.81%)	0.069
149	3.483	.49%	2.499	.02%	0.061
107	3.484	.46%	2.477	.93%	-0.029
160	3.486	.39%	2.516	(.64%)	0.010
147	3.499	.03%	2.505	(.19%)	0.046
109	3.499	.02%	2.484	.64%	-0.029
8	3.500	.00%	2.521	(.84%)	0.012
39	3.512	(.35%)	2.475	.99%	0.010
90	3.532	(.91%)	2.522	(.88%)	0.008

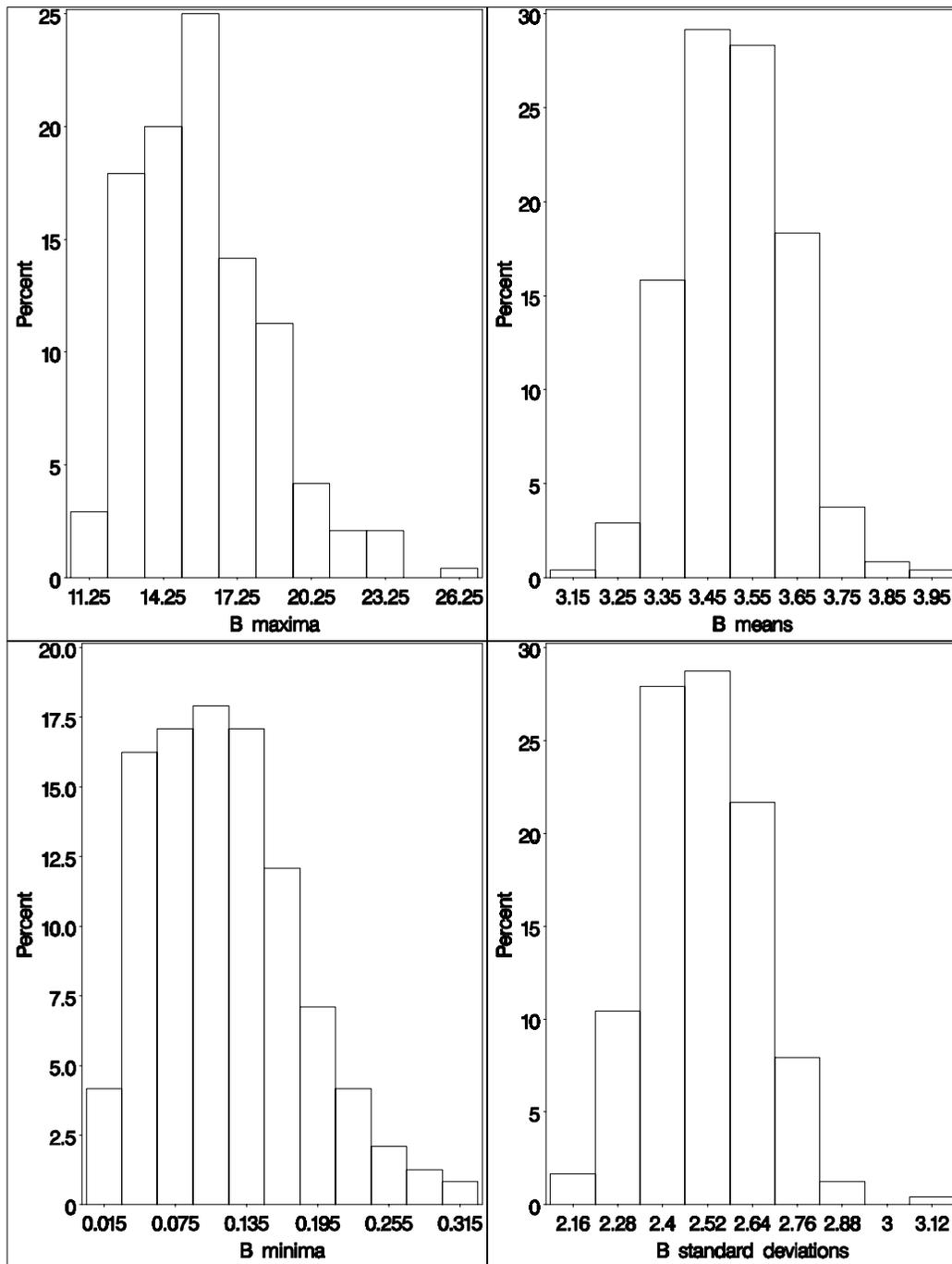


Figure 4. Variability of statistics of 240 realizations of B for St. Louis ($\mu\text{g}/\text{m}^3$)

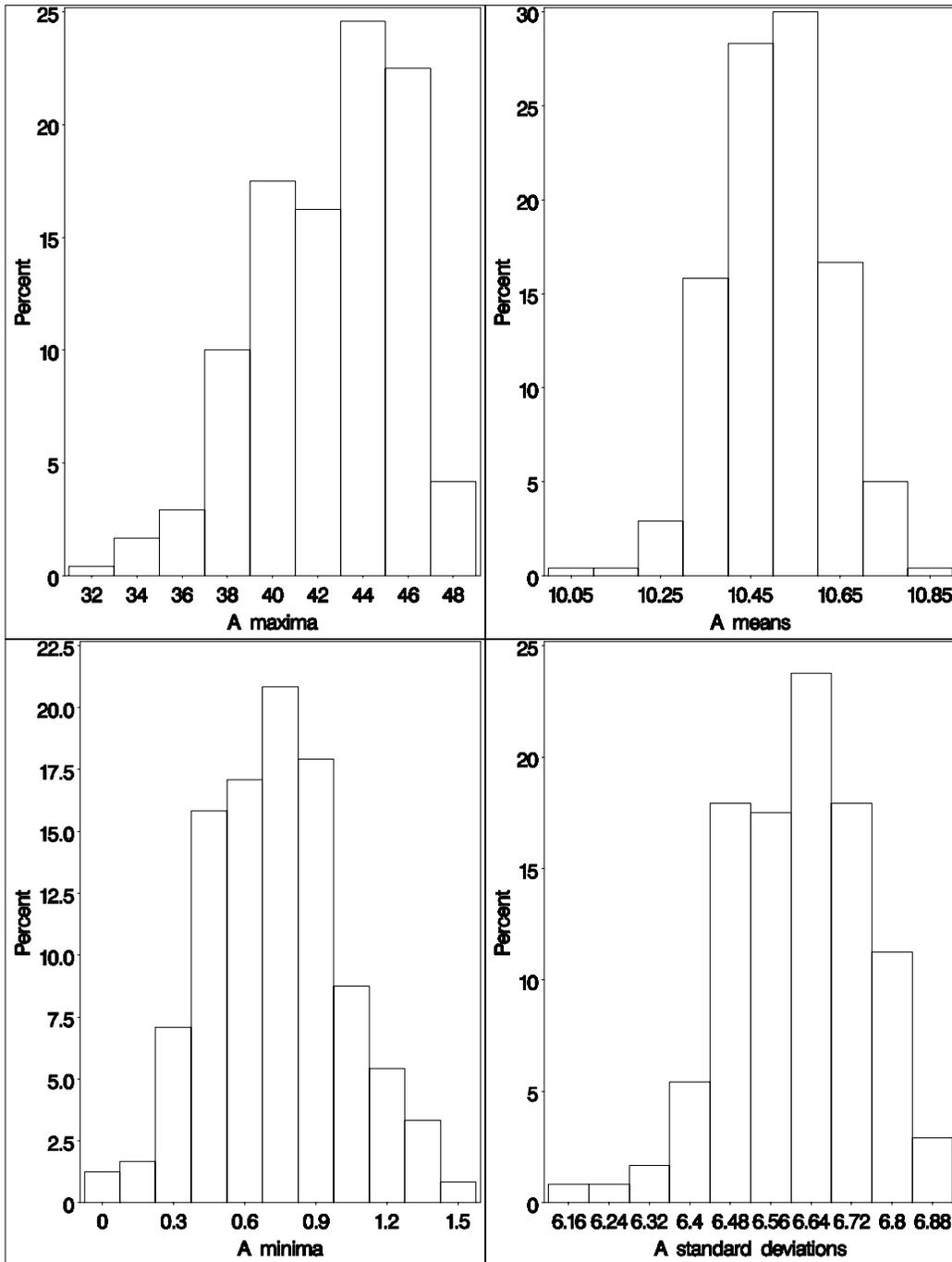


Figure 5. Variability of statistics of 240 realizations of A for St. Louis ($\mu\text{g}/\text{m}^3$)

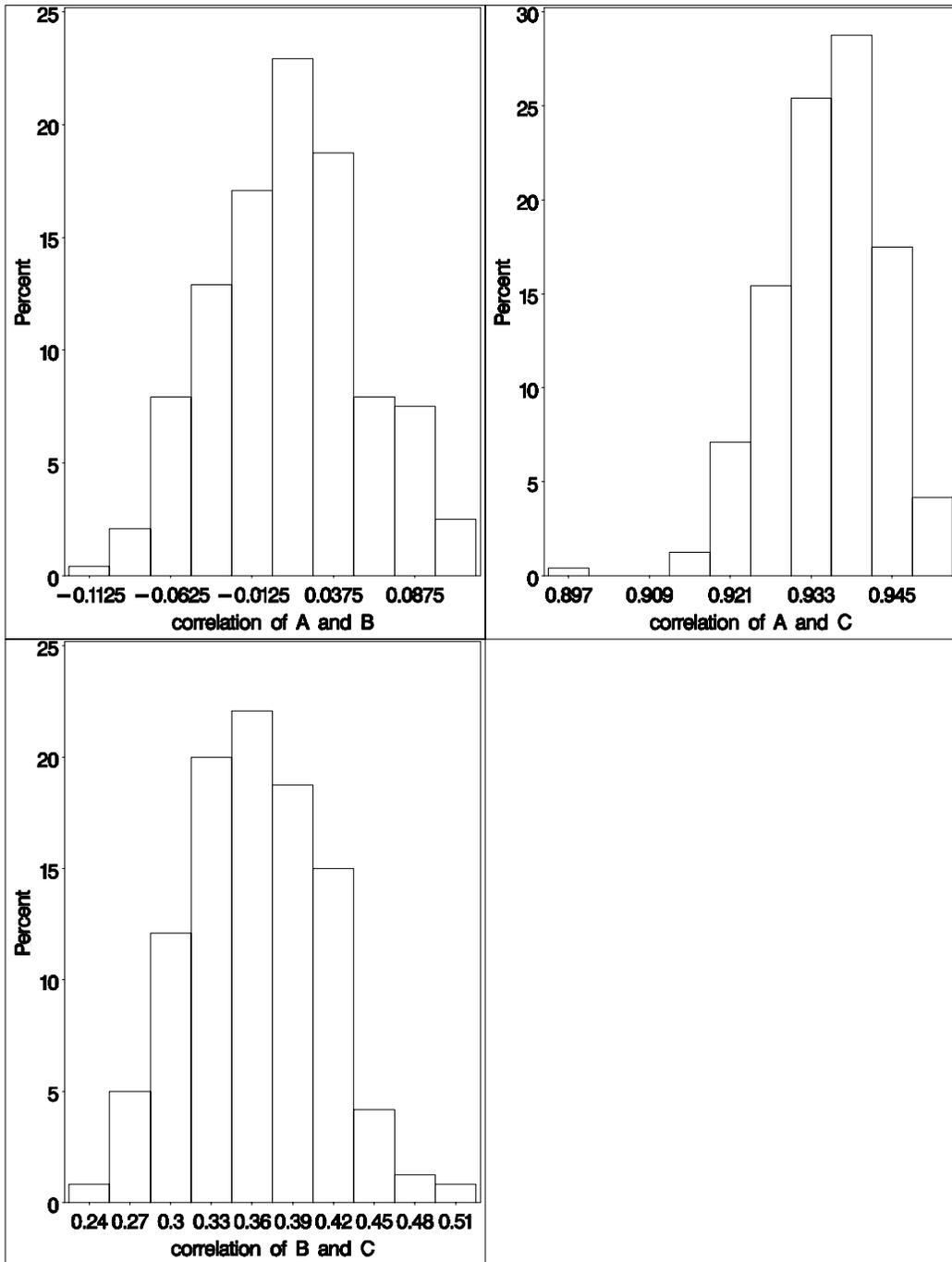


Figure 6. Variability of correlations of 240 realizations of A and B for St. Louis

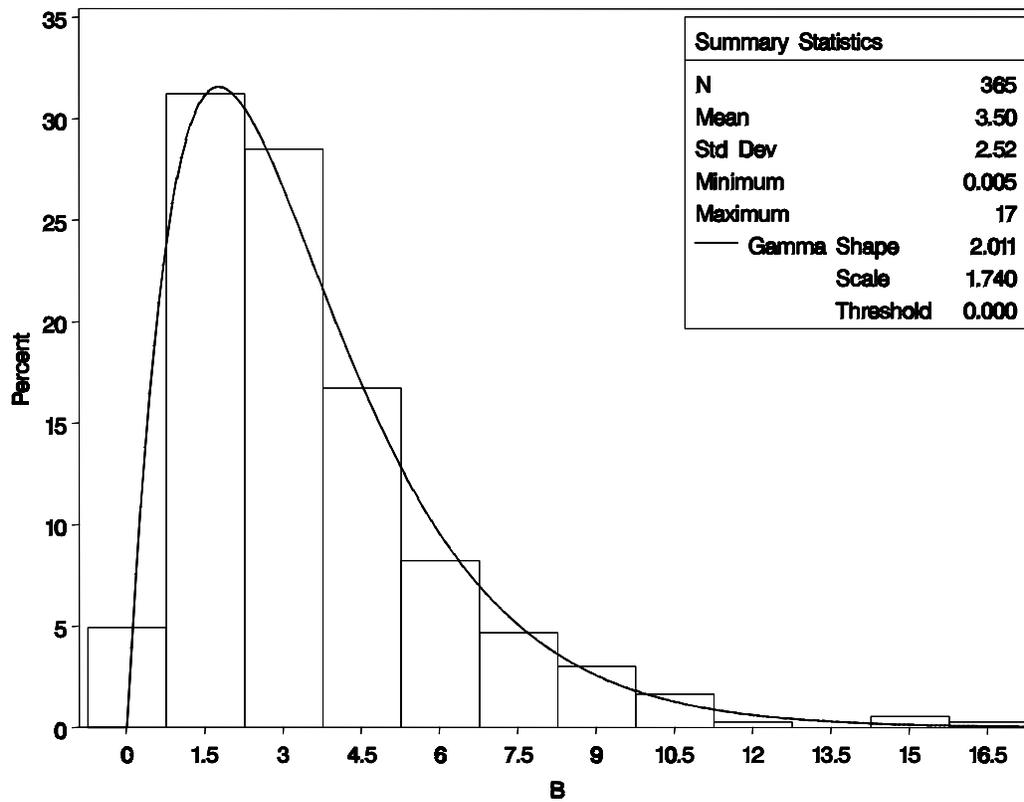


Figure 7. The distribution of the PM_{2.5} background concentration series for St. Louis ($\mu\text{g}/\text{m}^3$)

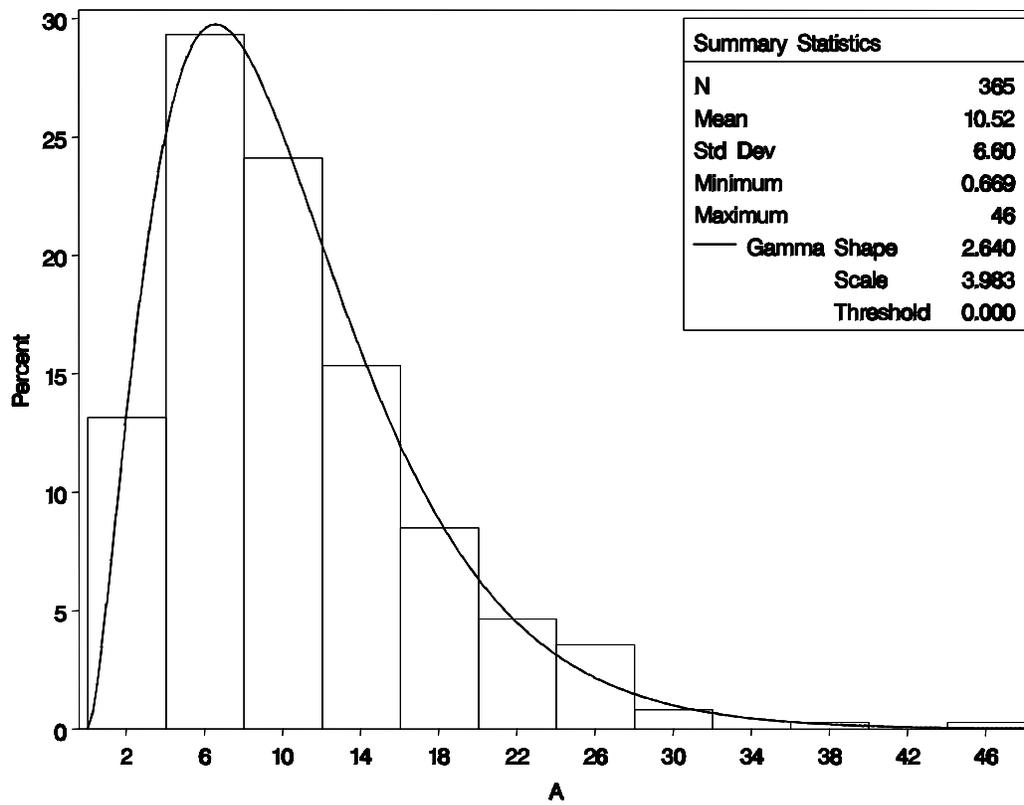


Figure 8. The distribution of the PM_{2.5} anthropogenic concentration series for St. Louis ($\mu\text{g}/\text{m}^3$)

We also generate a series of background concentrations with the assumption that the background and anthropogenic concentrations are positively correlated following the same steps described above, except ρ is taken to be non-zero. Figure 9 shows the distribution of background values assuming a correlation of 0.4 between the background and anthropogenic concentrations.

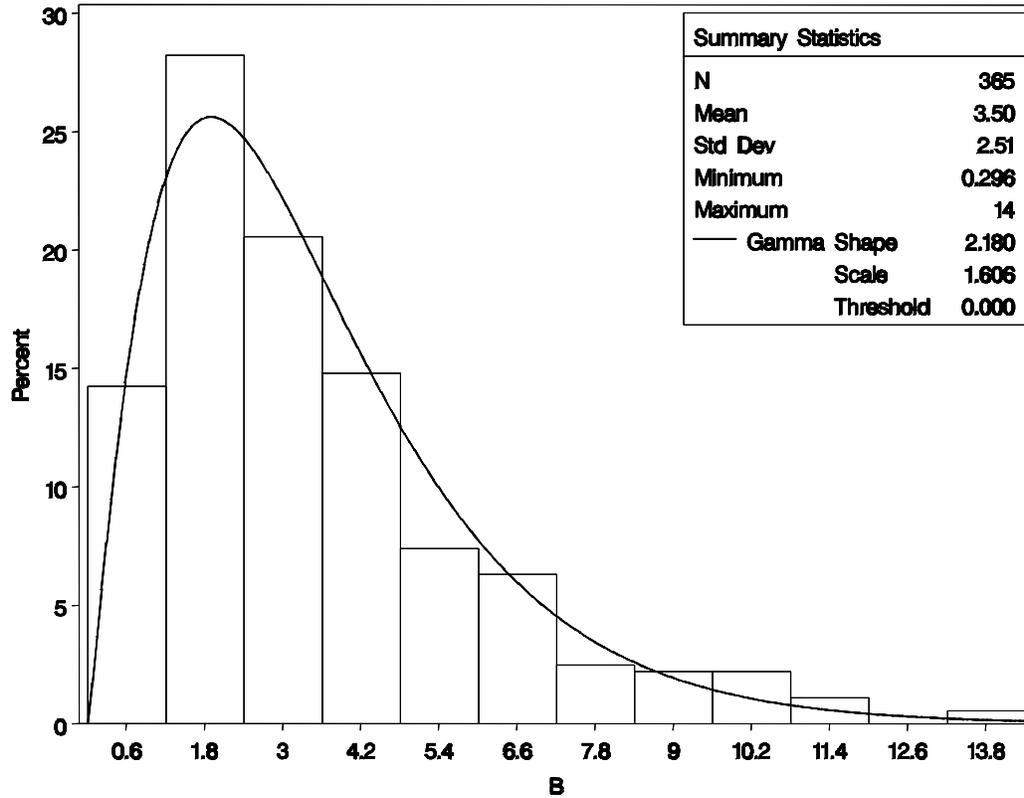


Figure 9. The distribution of the PM_{2.5} background concentration series for St. Louis, assuming a correlation of 0.4 between the background and anthropogenic concentrations ($\mu\text{g}/\text{m}^3$)

Application to Ambient Daily-Average PM_{2.5} Concentrations in Detroit

We have a time series C of 357 PM_{2.5} concentrations in Detroit for the year 2003 for which we want to generate a series of background concentrations B. The sample mean μ_C is 15.73 and the sample standard deviation σ_C is 9.29. Fitting a gamma distribution to these data yields the parameter estimates $\alpha_C = 3.183$ and $\beta_C = 4.943$. Figure 10 shows this fitted gamma distribution function overlaid on a histogram of the observed concentrations. Following the procedure described above, we generate a background series corresponding to the observed concentrations with mean 3.5 and standard deviation 2.5 $\mu\text{g}/\text{m}^3$. Figure 11 illustrates the distribution of this series.

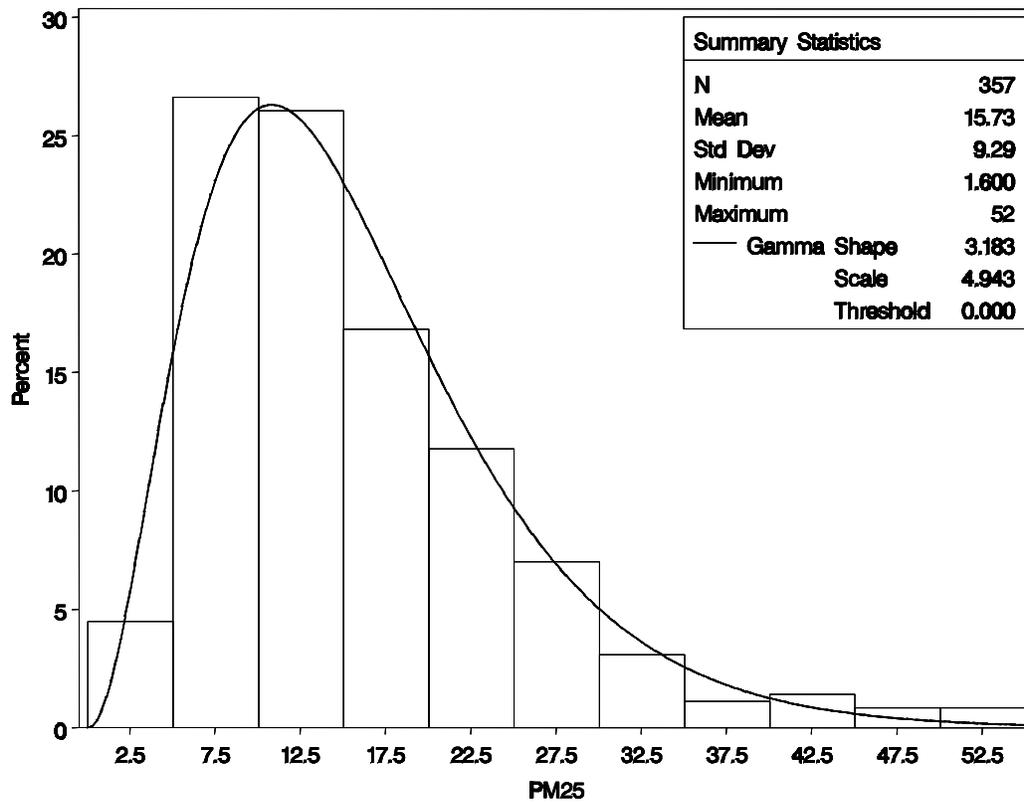


Figure 10. The distribution of the observed PM_{2.5} concentration series for Detroit ($\mu\text{g}/\text{m}^3$)

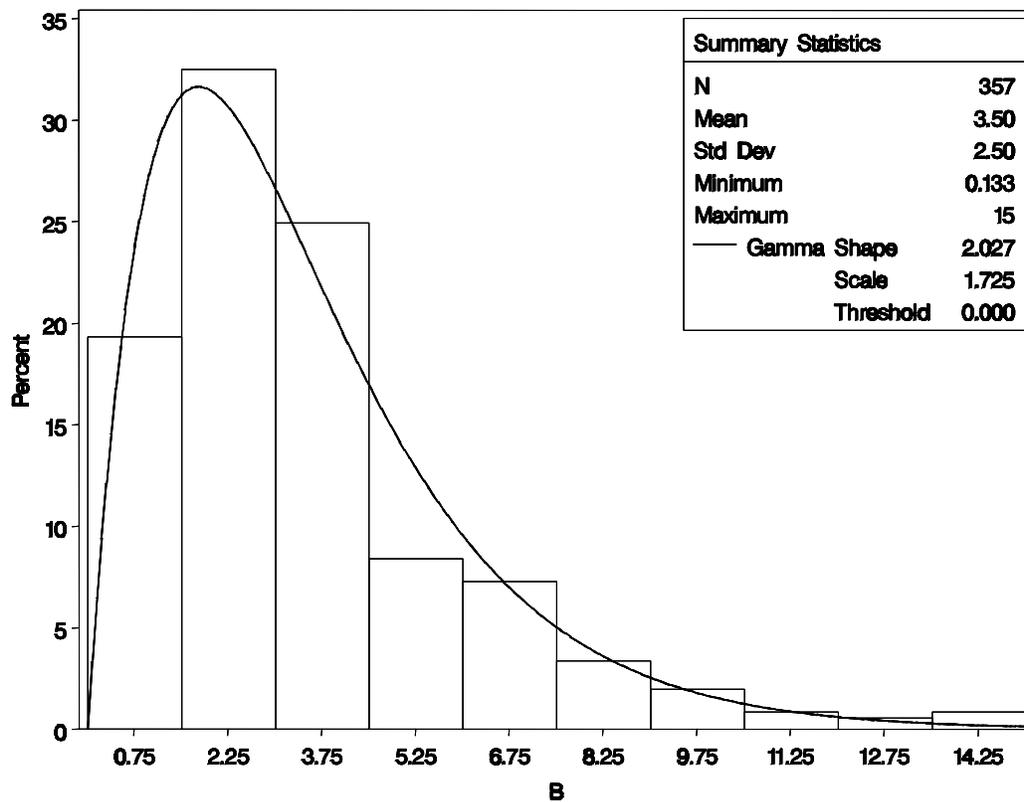


Figure 11. The distribution of the PM_{2.5} background concentration series for Detroit ($\mu\text{g}/\text{m}^3$)

References

Abt Associates, Inc. (2005). Particulate Matter Health Risk Assessment for Selected Urban Areas. Draft Report.

EPA (2004). Air Quality Criteria for Particulate Matter. EPA National Center for Environmental Assessment, Research Triangle Park, NC 27711. EPA/600/P-99/002aD.

EPA (2005). Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information, Second Draft PM Staff Paper. EPA Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711.

Langstaff, J. (2005). OAQPS Staff Memorandum to PM NAAQS Review Docket (OAR-2001-0017). Subject: Estimation of Policy-Relevant Background Concentrations of Particulate Matter.