Estimation for the unknown true concentration of environmental analytical data: Two-component mix model and Gamma mixed model

Pei-Shan Yen

June 5 2019

1 Introduction

The traditional environmental analysis commonly accepts the original measurements as true concentration without considering the uncertainty, which may not be realistic. However, statistical models can help evaluate the environmental analytical measurements. One specific topic is to reach an accurate estimation of the true concentration of asbestos fibers, and the characteristic of the data is the heteroscedastic measurement errors and between-laboratories variability.

The statistical model is constructed using the nonlinear format instead of the traditional linear format, for accounting for the larger variation of the analyte in a higher concentration level. We estimate the model parameters following the work of Kim and Bhaumik (2018), which adopts the two-component mixed model and gamma mixed model that is fast in execution. After obtaining the model parameters, we further apply this model to estimate the unknown true concentration and verify the variation property of the analytic observations.

2 Methodology

The assumptions of the model for the asbestos fibers are given as the following:

(1) Y_{ijk} is the kth replicate measurement of the jth concentration nested in the ith laboratory, where $i = 1, 2, ..., N$ and $j = 1, 2, ..., n_i$ as well as $k = 1, 2, ..., K_{ij}$.

(2) μ_j is the jth unknown concentration and it can be replaced by $\bar{y}_j$

(3) ϵ_{ijk} follows a normal distribution with a mean of 0 and the variance of σ_ϵ^2

(4) the regression coefficients of the fixed effects is defined as $\beta = (\beta_0, \beta_1)$ where β_0 is the intercept coefficient and β_1 the slope coefficient.

(5) the regression coefficients of the lab-specific random effects is defined as $u = (u_{0i}, u_{1i})$ where u_{0i} is the intercept coefficient and u_{1i} the slope coefficient. It is also assumed that the random effects u follow a bivariate normal distribution with a mean of 0 and the covariance matrix Σ_u with the components σ_0^2, σ_1^2, and σ_{01}.

2.1 Two-component mixed model

The characteristic of the environmental analytical measurements is that the variations increases in a higher concentration level. In other words, the variation remains constant at near-zero concentrations while the variation increases at larger concentrations. Rocke and Lorenzato (1995) proposed the nonlinear two-component model (1) to incorporate the characteristic of the analytics:

$$Y_{jk} = \beta_0 + \beta_1 \mu_j e^{\eta_j} + \epsilon_{jk} \tag{1}$$

In order to deal with the issue of uncertainty, the environmental samples are usually sent to multiple laboratories. To measure this kind of data, Bhaumik and Gibbons (2005) extended the model (1) to the model (2) so that the between-laboratories variability can be included.

$$Y_{ijk} = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j e^{\eta_i jk} + \epsilon_{ijk} \tag{2}$$

The iteratively reweighted maximum marginal likelihood (IWMML) method is used to estimate the model parameters and the variance component. IWMML is a combination of iteratively reweighted least squares (IWLS) method and the marginal maximum likelihood (MML) method. The IWMML procedure is proposed by Gibbons and Bhaumik (2001) to deal with multi-laboratories scenario. After obtaining the model parameters, we proceed to applying the model (2) to predict the true level of concentration. Finally, we perform the simulations to verify the estimation results.

2.1.1 IWMML Procedure

In what follows, steps 1-6 give the procedure for performing IWLS method, and steps 7-10 gives the ones for performing the MML methods. The two methods are alternated until all parameters converge.

1. Use $\bar{y}_j$ to replace the unknown concentration μ_j and define the initial value of β_0, β_1, σ_ϵ^2 and σ_η^2. Here $\beta_0^{(0)}$ and $\beta_1^{(0)}$ are obtained in virtue of the fitting to the linear model $Y_{ijk} = \beta_0 + \beta_1 \mu_j$ using the ordinary least square. $\sigma_\epsilon^{(0)}$ is the standard deviation of the measurements at the lowest concentration, and $\sigma_\eta^{(0)}$ is the standard deviation of the measurements at the highest concentration.

2. Update β_0 and β_1 by fitting the linear regression model $Y_{ijk} = \beta_0 + \beta_1 \mu_j + e$ using weighted linear regression model with weight $w_j = [\sigma_\epsilon^2 + (\beta_1 \mu_j)^2 e^{\sigma_\eta^2}(e^{\sigma_\eta^2} - 1)]^{-1}$.

3. Compute the mean of squared residuals $s^2(\mu_j) = \sum_{i=1}^{N} \frac{(\hat{\beta}_0 + \hat{\beta}_1 \mu_j - y_{ij})^2}{m_j K_{ij}}$ where m_j is the number of laboratories that analyze the measurements at concentration level j.

4. Construct a variance function model $s^2(\mu_j) = \gamma + \delta\mu_j^2 + e$, then fitting the linear model with weight $w_j = \frac{m_j K_{ij}}{s^2(\mu_j)}$ where $\gamma = \sigma_\epsilon^2$ and $\delta = \beta_1^2 e^{\sigma_\eta^2}(e^{\sigma_\eta^2} - 1)$.

5. Update $\sigma_\epsilon^2 = \gamma$ and $\sigma_\eta^2 = ln\left[\frac{1+\sqrt{\frac{1+4\delta}{\beta_1^2}}}{2}\right]$.

6. Iterate step 1-5 until achieving the convergence, and so gives $\hat{\sigma_\epsilon}^2$ and $\hat{\sigma_\eta}^2$.

7. Define the initial value of lab-specific parameters β_{0i} and β_{1i}. This lies in the fitting to the lab-specific linear model by using ordinary least square.

8. Fit the linear model $\boldsymbol{W_i^{\frac{1}{2}} Y_i} = \boldsymbol{W_i^{\frac{1}{2}} M_i \beta_i} + \boldsymbol{W_i^{\frac{1}{2}} \epsilon_i}$, $w_j = \frac{m_j K_{ij}}{[\hat{\sigma_\epsilon}^2 + (\hat{\beta}_{1i}\mu_j)^2 e^{\hat{\sigma_\eta}^2}(e^{\hat{\sigma_\eta}^2} - 1)]}$.

9. Estimate $\beta_0, \beta_1, \Sigma_\mu$, and the lab-specific parameters β_{0i} and β_{1i} through the MML method.

10. Update $\hat{\sigma_\epsilon}^2$ and $\hat{\sigma_\eta}^2$ by fitting the variance function model in step 4.

2.1.2 Point Estimation and Interval Estimation for unknown near-zero concentration

We can define the point estimation for the unknown near-zero concentration:

$$\hat{\mu}_j = \frac{1}{m_j}\sum_{i=1}^{m_j} \frac{y_{ij} - \hat{\beta}_{0i}}{\hat{\beta}_{1i} e^{\frac{\hat{\sigma_\eta^2}}{2}}} \tag{3}$$

where $\hat{\mu}_j$ is asymptotically unbiased. Let $\hat{\mu}_0$ be the average of the near-zero measurements. According to the central limit theorem, $\bar{y}_0$ will follow an approximately normal distribution:

$$\bar{Y_0} = \sum_{i=1}^{m_0} \frac{y_{i0}}{m_0} \sim N\Big(\beta_0 + \beta_1 \mu_0 e^{\frac{\sigma_\eta^2}{2}}, \frac{V(Y_{i0})}{m_0}\Big) \tag{4}$$

Based on the Slusky's theorem, the limiting distribution of $\hat{\mu}_0$ is a normal distribution:

$$\hat{\mu}_0 = \frac{\bar{y}_0 - \hat{\beta}_0}{\hat{\beta}_1 e^{\frac{\hat{\sigma_\eta^2}}{2}}} \sim N\Big(\mu_0, \frac{V(Y_{i0})}{m_0 \beta_1^2 e^{\sigma_\eta^2}}\Big) \tag{5}$$

Therefore, the 100(1-α)% confidence interval at the unknown near-zero concentration is

$$\hat{\mu}_0 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{V}(Y_{i0})}{m_0 \hat{\beta}_1^2 e^{\hat{\sigma}_\eta^2}}} \tag{6}$$

2.1.3 Point Estimation and Interval Estimation for unknown high concentration

The measurements at the high unknown concentration follow an approximately lognormal distribution. Hence, we can define

$$T_{ij} = ln\Big(\frac{Y_{ij} - \beta_{0i}}{\beta_{1i}\mu_j}\Big) \sim N\Big(0, ln\frac{1}{2}\Big\{1 + \Big(1 + \frac{4V(Y_{ij})}{\beta_{1i}^2 \mu_j^2}\Big)\Big\}\Big) \tag{7}$$

The pivot statistics $Z(\mu_j)$ follows an approximately standard normal distribution:

$$Z(\mu_j) = \frac{1}{\sqrt{m_j}}\sum_{i=1}^{m_j} \frac{T_{ij}}{\sqrt{V(Y_{ij})}} \tag{8}$$

We would then have the $100(1-\alpha)\%$ confidence interval for the unknown concentration:

$$Z(\mu_j) = \Big\{\mu_j : -z_{\frac{\alpha}{2}} \le Z(\mu_j) \le z_{\frac{\alpha}{2}}\Big\} \tag{9}$$

2.2 Gamma Mixed Model

In addition to the two-component mixed model, the gamma mixed model with identity link is an alternative approach to model the asbestos data. This is because the variance of a gamma distribution can also be expressed as a quadratic function of the mean. Here we does not use the gamma model $(Y_{ijk}|u = \beta_0 + \beta_1\mu_j)$ because the between-laboratories variability must be considered.

2.2.1 Conditional Distribution

A multiplicative gamma mixed model can be constructed to describe the random variable, namely, Y_{ijk} conditional on the random effect u

$$Y_{ijk}|u = [(\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j]\epsilon_{ijk} \tag{10}$$

where the random variable ϵ_{ijk} is independently and identically distributed as $Gamma(\alpha, \frac{1}{\alpha})$. The expectation and variance of ϵ_{ijk} are:

$$E(\epsilon_{ijk}) = \alpha\Big(\frac{1}{\alpha}\Big) = 1 \quad and \quad Var(\epsilon_{ijk}) = \alpha\Big(\frac{1}{\alpha}\Big)^2 = \frac{1}{\alpha} \tag{11}$$

Then we can obtain the conditional mean of Y_{ijk}:

$$E[Y_{ijk}|u] = E(\epsilon_{ijk})[\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j] = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j \tag{12}$$

as well as the conditional variance of Y_{ijk}:

$$Var[Y_{ijk}|u] = Var(\epsilon_{ijk})[\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j]^2 = \frac{[\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j]^2}{\alpha} \tag{13}$$

Because $Y_{ijjk}|u$ is a linear combination of the random variable ϵ_{ijk}, we can use the moment generating function to prove that $Y_{ijk}|u$ will follow a gamma distribution. Therefore, the model (10) can be expressed as

$$Y_{ijjk}|u \sim Gamma(\alpha, \theta) \tag{14}$$

where the scale parameter θ is $\frac{(\beta_0+u_{0i})+(\beta_1+u_{1i})\mu_j}{\alpha}$. Note that θ follows a normal distribution with mean $\frac{\beta_0+\beta_1\mu_j}{\alpha}$ and variance $\frac{\sigma_0^2+2\sigma_{01}\mu_j+\sigma_1^2\mu_j^2}{\alpha^2}$ because θ is a linear combination of the random effects u which follow a bivariate normal distribution.

2.2.2 Marginal Distribution

Applying (12) leads to the marginal mean of Y_{ijk}:

$$E(Y_{ijk}) = E[E(Y_{ijk}|u)] = (\beta_0 + u_{0i}) + (\beta_1 + u_{1i})\mu_j \tag{15}$$

The marginal variance of Y_{ijk} is the sum of the variance of the conditional expectation of $Y|u$ and the expected conditional variance $Y|u$:

$$Var(Y_{ijk}) = Var[E(Y_{ijk}|u)] + E[Var(Y_{ijk}|u)] \tag{16}$$

where $Var[E(Y_{ijk}|u)]$ is defined as the between-laboratory variance:

$$Var[E(Y_{ijk}|u)] = Var[\alpha\theta] = \alpha^2 Var[\theta] = \sigma_0^2 + 2\sigma_{01}\mu_j + \sigma_1^2\mu_j^2 \tag{17}$$

and $E[Var(Y_{ijk}|u)]$ is the within-laboratory variance:

$$E[Var(Y_{ijk}|u)] = E[\alpha\theta^2] = \alpha\{Var[\theta] + E[\theta]^2\} = \frac{\beta_0^2 + \sigma_0^2 + 2(\beta_0\beta_1 + \sigma_{01})\mu_j + (\beta_1^2 + \sigma_1^2)\mu_j^2}{\alpha} \tag{18}$$

Here the between-laboratory variance and the within-laboratory variance are both quadratic functions of concentration μ_j.

2.2.3 Parameter Estimation

Because (3) is one of the generalized linear mixed models (GLMM), the Marginal Maximum Likelihood (MML) is used to estimate the parameters. However, due to the complexity of the marginal likelihood equation, the MLE of the unknown parameter cannot be solved analytically. Consequently, the approximate estimation of the mixed model in an iterative numerical optimization procedure, for instance, the root-finding algorithm such as the adaptive Gauss-Hermite quadrature (AGQ) method or Laplace Approximation, are widely sought for.

The AGQ method is adopted in this study. The reason is that the large number of samples for environmental monitoring are typically limited, and Laplace Approximation produces less accurate estimates under the small size setting. When dealing with AGQ method, 10-20 quadrature points can produce good performance of estimation without pronounced computational cost. After the maximum likelihood estimation (MLE) of β, u, and σ_u are obtained by maximizing the full marginal likelihood function, we further use the Empirical Bayes (EB) method to estimate the lab-specific random effects u_i.

2.2.4 Point Estimation and Interval Estimation for no replicated measurement ($K = 1$)

In this section we discuss the point estimation and interval estimation of unknown true concentration when there is no repeated measurement ($K = 1$).

The point estimation for the unknown true jth level of concentration is

$$\hat{\mu}_j = \frac{1}{m_j}\sum_{i=1}^{m_j}\frac{y_{ij} - \hat{\beta}_{0i}}{\hat{\beta}_{1i}} \tag{19}$$

where $\hat{\beta}_{0i} = \hat{\beta}_0 + \hat{u}_{0i}$ and $\hat{\beta}_{1i} = \hat{\beta}_1 + \hat{u}_{1i}$ are lab-specific intercept and slope estimates.

The cube root of the gamma mixed model follows approximately normal distribution

$$Y_{ij}^{\frac{1}{3}}|u_i \sim Normal(\lambda_{ij}, \tau_{ij}^2) = Normal\left(\frac{\theta_{ij}^{\frac{1}{3}}\Gamma(\alpha+\frac{1}{3})}{\Gamma(\alpha)}, \frac{\theta_{ij}^{\frac{2}{3}}\Gamma(\alpha+\frac{2}{3})}{\Gamma(\alpha)} - \lambda_{ij}^2\right) \tag{20}$$

The lab-specific confidence interval for the unknown true concentration can then be derive from (20).

However, if we would like to send new environmental samples to brand new laboratories, the global confidence interval needs to be constructed based on the following procedures. Because the joint distribution of $(Y_{ij}^{\frac{1}{3}}, u)$ is an approximately bivariate normal distribution, the marginal distribution $Y_{ij}^{\frac{1}{3}}$ will follow an approximately normal distribution. The mean λ_j^* and variance τ_j^{*2} of the marginal distribution of Y_{ij} can be obtained by using the delta method:

$$\lambda_j^* \approx \frac{\Gamma(\alpha+\frac{1}{3})}{\Gamma(\alpha)}\Big(\frac{\beta_0+\beta_1\mu_j}{\alpha}\Big)^{\frac{1}{3}} \tag{21}$$

$$\tau_j^{*2} \approx \left(\frac{\beta_0+\beta_1\mu_j}{\alpha}\right)^{\frac{2}{3}}\left\{\frac{\Gamma(\alpha+\frac{2}{3})}{\Gamma(\alpha)} - \left(\frac{\Gamma(\alpha+\frac{1}{3})}{\Gamma(\alpha)}\right)^2\right\} + \left\{\frac{\Gamma(\alpha+\frac{1}{3})}{\Gamma(\alpha)}\right\}^2\frac{\sigma_0^2 + 2\sigma_{01}\mu_j + \sigma_1^2\mu_j^2}{9\alpha^{\frac{2}{3}}(\beta_0+\beta_1\mu_j)^{\frac{4}{3}}} \tag{22}$$

The pivot statistics $Z(\mu_j)$ follows a approximately standard normal distribution:

$$Z(\mu_j) = \frac{1}{\sqrt{m_j}}\sum_{i=1}^{m_j}\frac{y_{ij}^{\frac{1}{3}} - \hat{\lambda}_j^*}{\hat{\tau}_j^*} \tag{23}$$

As a result, the $100(1-\alpha)\%$ confidence interval for the unknown jth concentration is

$$Z(\mu_j) = \left\{\mu_j : -z_{\frac{\alpha}{2}} \le Z(\mu_j) \le z_{\frac{\alpha}{2}}\right\} \tag{24}$$

2.2.5 Point Estimation and Interval Estimation for replicated measurement ($K > 1$)

Lastly, we discuss the point estimation and interval estimation of unknown true concentration when the replicated measurement ($K > 1$) occurs.

We can define the point estimation for the unknown true jth level of concentration by letting $\bar{y}_{ij.} = \sum_{k=1}^{K_{ij}} \frac{y_{ijk}}{K_{ij}}$ and

$$\hat{\mu}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} \frac{\bar{y}_{ij.} - \hat{\beta}_{0i}}{\hat{\beta}_{1i}} \tag{25}$$

The confidence interval for the unknown true concentration is constructed on the approximately normal distribution for the cube root of the gamma mixed model. The observations $Y_{ij}^{\frac{1}{3}} | u_i$ follow an approximately normal distribution with mean $\lambda_{ij} = \frac{\theta_{ij}^{*\frac{1}{3}} \Gamma(\alpha_{ij}^* + \frac{1}{3})}{\Gamma(\alpha_{ij}^*)}$ and variance $\tau_{ij}^2 = \frac{\theta_{ij}^{*\frac{2}{3}} \Gamma(\alpha_{ij}^* + \frac{2}{3})}{\Gamma(\alpha_{ij}^*)} - \lambda_{ij}^2$. Here $\alpha_{ij}^* = \alpha K_{ij}$ and $\theta_{ij}^* = \frac{\theta_{ij}}{K_{ij}}$.

The marginal distribution of $Y_{ijk}^{\frac{1}{3}}$ will follow an approximately normal distribution with mean λ_j^* and variance τ_j^{*2}. This can be again obtained by using the delta method:

$$\lambda_j^* \approx \frac{\Gamma(\alpha_{ij}^* + \frac{1}{3})}{\Gamma(\alpha_{ij}^*)} \Big(\frac{\beta_0 + \beta_1 \mu_j}{\alpha_{ij}^*} \Big)^{\frac{1}{3}} \tag{26}$$

$$\tau_j^{*2} \approx \left(\frac{\beta_0 + \beta_1 \mu_j}{\alpha_{ij}^*} \right)^{\frac{2}{3}} \left\{ \frac{\Gamma(\alpha_{ij}^* + \frac{2}{3})}{\Gamma(\alpha_{ij}^*)} - \left(\frac{\Gamma(\alpha_{ij}^* + \frac{1}{3})}{\Gamma(\alpha_{ij}^*)} \right)^2 \right\} + \left\{ \frac{\Gamma(\alpha_{ij}^* + \frac{1}{3})}{\Gamma(\alpha_{ij}^*)} \right\}^2 \frac{\sigma_0^2 + 2\sigma_{01}\mu_j + \sigma_1^2 \mu_j^2}{9 \alpha_{ij}^{*\frac{2}{3}} (\beta_0 + \beta_1 \mu_j)^{\frac{4}{3}}} \tag{27}$$

The pivot statistics $Z(\mu_j)$ follows a approximately standard normal distribution:

$$Z(\mu_j) = \frac{1}{\sqrt{m_j}} \sum_{i=1}^{m_j} \frac{\bar{y}_{ij.}^{\frac{1}{3}} - \hat{\lambda}_j^*}{\hat{\tau}_j^*} \tag{28}$$

We would then have the $100(1-\alpha)\%$ confidence interval for the unknown concentration:

$$Z(\mu_j) = \left\{ \mu_j : -z_{\frac{\alpha}{2}} \leq Z(\mu_j) \leq z_{\frac{\alpha}{2}} \right\} \tag{29}$$

3 Simulation

3.1 Two-component mixed model

In order to examine the two-component mixed model (2), a simulation exercise which consists of 1000 simulated datasets is performed. Each dataset is generated by the model whose parameters are given as the following:

1. fixed effects: $\beta_0 = 0$, $\beta_1 = 1$
2. random effects: $\sigma_0^2 = 0.0020$, $\sigma_{01} = -0.0015$, and $\sigma_1^2 = 0.0250$
3. random errors: $\sigma_\epsilon^2 = 0.0015$, and $\sigma_\eta^2 = 0.0500$

Table 1 demonstrates the average and the root mean square error (RMSE) of the parameter estimation. The estimation of the mean unknown concentration perform well and the RMSE decreases when the number of replication increases. Table 2 shows the result of point estimation as well as the interval estimation. The point estimate has a better performance when the number of laboratories increases. The length of the confidence interval become wider as the concentration level increase. This reflects the heteroscedastic characteristic of the data.

Table 1: Parameter estimation for two component mixed model

Parameters	True Values	N (# of Labs) / K (# of replicates)						
		Estimate	30/10	30/05	30/01	20/10	20/05	20/01
β_0	0.0000	MEAN	0.0039	0.0044	0.0077	0.0034	0.0041	0.0078
		RMSE	0.0076	0.0078	0.0100	0.0092	0.0098	0.0118
β_1	1.0000	MEAN	1.0171	1.0138	0.9932	1.0173	1.0153	0.9933
		RMSE	0.0299	0.0309	0.0357	0.0361	0.0375	0.0442
σ_0^2	0.0020	MEAN	0.0016	0.0015	0.0017	0.0016	0.0015	0.0016
		RMSE	0.0004	0.0004	0.0009	0.0005	0.0006	0.0009
σ_{01}	-0.0015	MEAN	-0.0011	-0.0011	-0.0014	-0.0011	-0.0011	-0.0011
		RMSE	0.0012	0.0013	0.0021	0.0016	0.0017	0.0023
σ_1^2	0.0250	MEAN	0.0249	0.0250	0.0242	0.0246	0.0246	0.0221
		RMSE	0.0068	0.0069	0.0096	0.0083	0.0087	0.0110
σ_ϵ^2	0.0015	MEAN	0.0013	0.0013	0.0010	0.0013	0.0012	0.0011
		RMSE	0.0003	0.0005	0.0011	0.0004	0.0006	0.0012
σ_η^2	0.0500	MEAN	0.0475	0.0468	0.0452	0.0476	0.0470	0.0443
		RMSE	0.0037	0.0049	0.0110	0.0047	0.0061	0.0138

Table 2: Interval estimation for two component mixed model

	$N = 30$ Labs, $K = 1$ (No replicates)				$N = 20$ Labs, $K = 1$ (No replicates)			
μ	$\hat{\mu}$.	95% CI	Length	C.P.	$\hat{\mu}$	95% CI	Length	C.P.
0.0600	0.0584	(0.0401, 0.0767)	0.0366	0.9950	0.0583	(0.0359, 0.0808)	0.0449	0.9940
0.1500	0.1433	(0.1192, 0.1674)	0.0482	0.9810	0.1440	(0.1143, 0.1736)	0.0553	0.9850
0.3000	0.2938	(0.2593, 0.3284)	0.0692	0.9750	0.2940	(0.2516, 0.3363)	0.0847	0.9730
0.4000	0.3950	(0.3622, 0.4254)	0.0633	0.9030	0.3946	(0.3554, 0.4324)	0.0770	0.9060
0.5000	0.4945	(0.4554, 0.5332)	0.0778	0.9300	0.4950	(0.4481, 0.5427)	0.0946	0.9180
0.6000	0.5935	(0.5475, 0.6398)	0.0923	0.9310	0.5964	(0.5413, 0.6538)	0.1126	0.9330
0.8000	0.7950	(0.7352, 0.8574)	0.1222	0.9280	0.7962	(0.7242, 0.8725)	0.1483	0.9360
1.2000	1.1986	(1.1103, 1.2930)	0.1827	0.9390	1.2031	(1.0967, 1.3185)	0.2218	0.9400
2.8000	2.8014	(2.5964, 3.0207)	0.4243	0.9520	2.8018	(2.5568, 3.0697)	0.5129	0.9550
4.0000	4.0063	(3.7135, 4.3197)	0.6062	0.9590	4.0199	(3.6662, 4.4010)	0.7348	0.9320

3.2 Gamma mixed model

A simulation study is performed to verify the gamma mixed model (3). Total of 1000 simulated datasets are generated following the model with parameter values:

1. fixed effects: $\beta_0 = 0$, $\beta_1 = 1$
2. random effects $\sigma_0^2 = 0.0005$, $\sigma_{01} = -0.0002$, and $\sigma_1^2 = 0.0100$
3. the shape parameter of the gamma model $\alpha = 13$

Table 3 demonstrates the average and the root mean square error (RMSE) of the parameter estimations. Table 4 shows that the result of point estimation and interval estimation.

Table 3: Parameter estimation for gamma mixed model

Parameters	True Values	N (# of Labs) / K (# of replicates)						
		Estimate	30/10	30/05	30/01	20/10	20/05	20/01
β_0	0	MEAN				-0.0019		
		RMSE				0.0052		
β_1	1	MEAN				1.0008		
		RMSE				0.0272		
σ_0^2	0.0005	MEAN				0.0004		
		RMSE				0.0002		
σ_{01}	-0.0002	MEAN				-0.0002		
		RMSE				0.0007		
σ_1^2	0.010	MEAN				0.0095		
		RMSE				0.0054		
α	13	MEAN				13.0930		
		RMSE				1.1707		

Table 4: Interval estimation for gamma mixed model

	$N = 30$ Labs, $K = 10$ (replicates)				$N = 30$ Labs, $K = 1$ (No replicates)			
μ	$\hat{\mu}$.	95% CI	Length	C.P.	$\hat{\mu}$	95% CI	Length	C.P.
0.0600	0.0607				0.0629			
0.1500	0.1505				0.1526			
0.3000	0.3002				0.3027			
0.4000	0.4006				0.4027			
0.5000	0.5009				0.3037			
0.6000	0.6010				0.6015			
0.8000	0.8002				0.7996			
1.2000	1.1199				1.2001			
2.8000	2.7990				2.7999			
4.0000	3.9993				4.0078			

4 Discussion

The parameter estimation for the two-component mixed model can be successfully performed by using R statistical software. However, the estimation for the gamma mixed model fails in R due to the complexity of the likelihood function. The following R packages are used, including the packages lme4, GLMMadaptive, R2admb, glmmADMB, and the Zelig, but all return the error messages about the failure of the numerical solution. As a result, the work of the parameter estimation is performed in SAS. The parameter estimation and the point estimation of the unknown true concentration can be done in SAS. In the future, the interval estimation would be continually developed.

5 Reference

Bhaumik, D. K., and R. D. Gibbons. 2005. Confidence regions for random-effects calibration curves with heteroscedastic errors. Technometrics 47:223–31.

Gibbons, R. D.. 2001, Statistical Methods for Detection and Quantification of Environmental Contamination, NY: Wiley.

Gibbons, R. D., and D. K. Bhaumik. 2001. Weighted random-effects regression models with application to interlaboratory calibration. Technometrics 43:192–98.

Gibbons, R.D., Bhaumik, D.K., Aryal, S., 2009. Statistical Methods for Groundwater Monitoring, NY: Wiley.

Hedeker, D., and R. D. Gibbons. 2006. Longitudinal data analysis. Hoboken, NJ: Wiley.

Kim, Y., and Bhaumik, D. K.. 2018. Calibration confidence regions for asbestos fibers with heteroscedasticity and interlaboratory variability. Journal of statistical theory and practice 12:635-656 .

Rocke, D. M., and S. Lorenzato. 1995. A two-component model for measurement error in analytical chemistry. Technometrics 37:176–84.