This document describes our likelihood-based approach to multiply impute censored data of the form $\mathbf{Y}_{T\times P}$, where T is the number of observations and P is the number of variables. In $\mathbf{Y}$, there are some data which are censored For each observation t, we assumed $\mathbf{y_t} \sim \text{MVN}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$.

First $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are estimated using a Markov Chain Monte Carlo approach to sample from the posterior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, since the censored data make using standard maximum likelihood estimators difficult. We used conjugate priors $\boldsymbol{\theta} \sim \text{MVN}(\mathbf{0}, 10^5\mathbf{I})$ and $\boldsymbol{\Sigma} \sim \text{inv-Wishart}(P+1, \mathbf{I})$, where $\mathbf{I}$ is the $P \times P$ identity matrix. We directly sampled from the posterior distributions of $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, and the censored constituent concentrations using Gibbs sampling. Letting $\mathbf{Y} = \log(\mathbf{X})$, the full conditionals for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ are

$$(\boldsymbol{\theta} \mid \boldsymbol{\Sigma}, \mathbf{Y}) \sim \text{MVN}\Big(\left(10^{-5}\mathbf{I} + n\boldsymbol{\Sigma}^{-1}\right)^{-1} \left(n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}\right), \;\; \left(10^{-5}\mathbf{I} + n\boldsymbol{\Sigma}^{-1}\right)^{-1} \Big) \tag{1}$$

$$(\boldsymbol{\Sigma} \mid \boldsymbol{\theta}, \mathbf{Y}) \sim \text{inv-Wishart}\left(n + P + 1, \; \mathbf{I} + \sum_{t=1}^{n} (\mathbf{y}_t - \boldsymbol{\theta})(\mathbf{y}_t - \boldsymbol{\theta})^T \right) \tag{2}$$

where $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, ..., \bar{y}_P)^T$. For each observation t, let y_{tp} be the data for a censored variable p and $\mathbf{y}_{tq}$ be the data for the remaining q variables. The distribution of y_{tp} conditional on $\mathbf{y}_{tq}$ is truncated normal,

$$(y_{tp} \mid \mathbf{y}_{tq}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) \sim \; \text{trunc-N}\left(\theta_p + \boldsymbol{\Sigma}_{pq}\boldsymbol{\Sigma}_q^{-1}(\mathbf{y}_{tq} - \boldsymbol{\theta}_q), \; \Sigma_p - \boldsymbol{\Sigma}_{pq}\boldsymbol{\Sigma}_q^{-1}\boldsymbol{\Sigma}_{pq}^T \right) \tag{3}$$

where y_{tp} is censored, $\boldsymbol{\Sigma}_{pq}$ is the covariance between variable p and the remaining variables q, and θ_p, $\boldsymbol{\theta}_q$, Σ_p, and $\boldsymbol{\Sigma}_q$ refer to the subsets of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ corresponding to constituents p and q.

To impute the censored data, the function draws N samples from the joint distribution of $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, and the censored data by iteratively sampling from the three distributions (equations (1), (2), and (3)) and updating the values for $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$, and each censored y_{tp}. Let $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}$ be the posterior means of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. Each censored observation y_{tp} is imputed using a random draw from the truncated normal in equation 3, conditioning on observed variables on day t and replacing $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}$.