Supplementary Material

# Sensitivity Analysis of Parameters

Before executing MCMC, a sensitivity analysis was conducted within the exploration range of each parameter. Parameters that were manually fitted to the spectrum of Mixture No. 11 were taken as the true values. From these, only one parameter was varied within the MCMC exploration range to generate a spectrum, and the difference from the observed data was calculated.

As a result, as shown in Figure 6‑1, it was found that the monoisotopic mass exhibits a steep sensitivity characteristic. This is due to the peak width of the detector response being at most 0.05 Th at m/z: 1,972, which is extremely small (0.0025%) relative to the mass space analyzable by Q-TOF, ranging from 10 to 40,000 Th. It was also confirmed that the monoisotopic mass exhibits multi-modality within the exploration space.

The parameters of charge state influence the macro distribution shape, but do not affect the intervals between the comb-like peaks of mass-to-charge ratio, resulting in a broad sensitivity characteristic as shown in Figure 6‑5 and Figure 6‑6. Likewise, the isotopic parameters influence the micro peak width, but do not significantly change the mass, resulting in a broad sensitivity characteristic as seen in Figure 6‑3 and Figure 6‑4.



(b)

(a)

Figure ‑ Sensitivity characteristic of monoisotopic mass (a) Overall view, (b) Enlarged view



Figure ‑　 Sensitivity characteristic of ion counts



Figure ‑　 Sensitivity characteristic of representative number of atoms



Figure ‑ Sensitivity characteristic of representative isotopic abundance



Figure ‑　 Sensitivity characteristic of representative number of functional groups



Figure ‑　 Sensitivity characteristic of representative charge rate

# Parameter Exploration Using the Markov Chain Monte Carlo Method

From the posterior probability distribution $P\left(\overbar{S}\_{obs}\left(t\right)\right)$, we sample the parameter $θ$ to select the one that maximizes the posterior probability. We employ the No-U-Turn Sampler (NUTS) for sampling, a recent and popular variant of the Markov Chain Monte Carlo (MCMC) method. NUTS is a type of MCMC, especially a derivative of the Hamiltonian Monte Carlo method (HMC) (Betancourt 2017; Neal. 2011). HMC uses concepts from physics to efficiently sample from high-dimensional probability distributions. However, choosing an appropriate number of leapfrog(Birdsall and Langdon 2004) steps (the number of steps the parameter moves during simulation) in HMC can be challenging. If there are too few steps, the sampler cannot effectively move across the exploration space; and too many, it will U-turn back toward its starting point.

NUTS dynamically selects an appropriate number of steps to explore the Hamiltonian's energy surface based on the principle of stopping the step before the sampler begins a U-turn. This addresses the problem of adjusting the HMC leapfrog steps, enabling efficient sampling from high-dimensional probability distributions. The max tree depth, equivalent to the maximum number of search steps in a single iteration, was set to 10. The Hamiltonian is defined as:

$$H(q, p) = U(q) + K(p)$$

where $q$ is the variable (position) we want to sample from the target probability distribution, $p$ is an auxiliary variable (momentum), $U(q)$ represents potential energy, and $K(p) $represents kinetic energy.

First, we randomly initialize the momentum $p$ for the current position $q$. Using the leapfrog method, we compute a new $(q, p)$ pair. In this process, the momentum p is initially updated by half a step:

$$p(t+\frac{ε}{2}) = p(t)-\frac{ε}{2}∙\frac{∂U}{∂q(t)}$$

where $ε$ is the step size, and $t$ represents the current time step. Next, the position $q$ is updated for the one step:

$$q\left(t+ε\right)=q\left(t\right)+ ε∙\frac{∂K}{∂p(t+\frac{ε}{2})}$$

Finally, the momentum $p$ is updated by another half step:

$$p(t+ε)=p(t+\frac{ε}{2})-\frac{ε}{2}∙\frac{∂U}{∂q(t+ε)}$$

If $p∙p^{'}\leq 0$, it is determined that the sampler has made a U-turn, and the exploration is terminated. After all steps are completed, an acceptance/rejection step is performed using the Metropolis method(Wasserstein, Kalos, and Whitlock 1989; Tierney 1994). During this step, the difference in Hamiltonian energy is computed to get $ΔH=H(q', p')-H(q, p)$. Here, $(q, p)$ is the current sample and $(q', p')$ is the proposed new sample. If the Hamiltonian energy of the new sample is lower or equal to the current one $\left(ΔH\leq 0\right)$, the new sample is accepted. Conversely, if the Hamiltonian energy of the new sample is higher $\left(ΔH>0\right)$, it's accepted with probability $min(1, exp(-ΔH))$.

For this study, the step size $ε$ of the sampling algorithm was tuned to achieve an acceptance rate of 0.5. If the acceptance rate is too high, only steps near the current parameter value might be accepted, possibly preventing full exploration of the parameter space. If the rate is too low, many proposed steps will be rejected, increasing the time taken for sampling.

The number of samples in this study was set to 1,000. Although the initial state of MCMC is chosen randomly, this state often lies in a domain different from the target probability distribution. Reaching closer to the target distribution requires a certain number of steps (iterations). However, samples generated in this initial phase often do not reflect the posterior probability distribution correctly. Therefore, we discard samples from this initial phase. This process is called "Burn-in," and was set to 1,000 samples in this study. Before the 1,000 step sampling for the parameter search, this burn-in sampling was performed.

There's also autocorrelation between samples produced by MCMC, implying that consecutive samples depend on each other. This autocorrelation can impact statistical inference. To reduce the correlation between acquired samples, we sampled every other step. Furthermore, MCMC sampling is dependent on its initial state, which, induces in high-dimensional spaces, risks getting trapped in local optima. By sampling from multiple initial values, we can explore the parameter space more broadly and reduce this risk. In this study, we started from four different initial values.

The domain definitions for each parameter are as per Table 2‑1. When the number of constituents is $k=k’(>2)$, from among the $k'$ constituents, the prior distribution of parameters for constituents 1 to $k'-1$ is determined using a narrowed prior distribution centered around the optimal parameters estimated in the model for $k = k'-1$. Based on this new prior distribution, we estimate the optimal parameters and acquire the maximum posterior probability.

$m\_{j}^{'}|\_{k=k^{'}}$ represents the monoisotopic mass when the number of constituents is $k’$. When $j<k'$, the search range is limited to $\pm Δm$ from the value obtained at $k=k’-1$. For $j=k'$, the entire pre-set search range is explored. The same applies to $I\_{j}|\_{k=k^{'}}$、$n\_{j}|\_{k=k^{'}}$、$u\_{j}|\_{k=k^{'}}$、$l\_{j}|\_{k=k^{'}}$、$v\_{j}|\_{k=k^{'}}$. As the number of constituents increases, the area that a single constituent occupies in the observed data spectrum becomes smaller. Therefore, the lower limit of $I\_{j}|\_{k=k^{'}}$ is divided by $k’$.

Table ‑ The domain of the parameters

|  |  |  |
| --- | --- | --- |
| Parameter | Range | Constant |
| $$m\_{j}^{'}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[m\_{min}, m\_{max}\right]\left(j=k'\right)\\\left[\left(m\_{j}^{'}|\_{k=k^{'}-1}-∆m\right), \left(m\_{j}^{'}|\_{k=k^{'}-1}+∆m\right)\right] \left(j<k'\right)\end{array}\right.$$ | $$m\_{min}=6300.0$$$$m\_{max}=6400.0$$$$∆m=4.0$$ |
| $$I\_{j}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[I\_{min}, I\_{max}\right]\left(j=k'\right)\\\left[\frac{I\_{j}|\_{k=k^{'}-1}}{3k'}, I\_{max}\right]\left(j<k'\right)\end{array}\right.$$ | $$I\_{min}=10000$$$$I\_{max}=300000$$ |
| $$n\_{j}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[n\_{min}, n\_{max}\right]\left(j=k'\right)\\\left[\left(n\_{j}|\_{k=k^{'}-1}\*\left(1-∆n\right)\right), \left(n\_{j}|\_{k=k^{'}-1}\*\left(1+∆n\right)\right)\right]\left(j<k'\right)\end{array}\right.$$ | $$n\_{min}=\frac{m\_{j}^{'}|\_{k=k^{'}}}{12.0}$$$$n\_{max}=\frac{m\_{j}^{'}|\_{k=k^{'}}}{8.0}$$$$∆n=0.05$$ |
| $$u\_{j}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[u\_{min}, u\_{max}\right]\left(j=k'\right)\\\left[\left(u\_{j}|\_{k=k^{'}-1}-∆u\right), \left(u\_{j}|\_{k=k^{'}-1}+∆u\right)\right]\left(j<k'\right)\end{array}\right.$$ | $$u\_{min}=0.001$$$$u\_{max}=0.011$$$$∆u=0.001$$ |
| $$l\_{j}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[l\_{min}, l\_{max}\right]\left(j=k'\right)\\l\_{j}|\_{k=k^{'}-1}\left(j<k'\right)\end{array}\right.$$ | $$l\_{min}=\frac{m\_{j}^{'}|\_{k=k^{'}}}{40.0}$$$$l\_{max}=\frac{m\_{j}^{'}|\_{k=k^{'}}}{20.0}$$ |
| $$v\_{j}|\_{k=k^{'}}$$ | $$\left\{\begin{array}{c}\left[v\_{min}, v\_{max}\right]\left(j=k'\right)\\v\_{j}|\_{k=k^{'}-1}\left(j<k'\right)\end{array}\right.$$ | $$v\_{min}=0.01, $$$$v\_{max}=0.05$$ |

# Parameter Optimization by Stochastic Variational Inference

After executing MCMC, the parameters of the maximum posterior probability obtained are inherited as initial values, and optimization of the parameters is performed using Stochastic Variational Inference (SVI(Wingate and Weber 2013; Ranganath, Gerrish, and Blei 22--25 Apr 2014; Kingma and Welling 2013)). In SVI, instead of directly computing the posterior probability distribution, an easier-to-compute distribution is used to approximate the posterior probability distribution. A good approximation of the posterior probability distribution is obtained by minimizing the KL divergence between this approximate distribution and the true posterior probability distribution. However, since KL divergence cannot be directly calculated, the optimal variational function is determined by maximizing the ELBO (Evidence Lower Bound) (Blei, Kucukelbir, and McAuliffe 2017). For this study, only the value of the maximum a posteriori (MAP) was required, so a delta function is used to approximate the posterior probability distribution in the model of each constituent number.

To maximize the ELBO, Adam(Kingma and Ba 2014) (Adaptive Moment Estimation), a type of Stochastic Gradient Descent (SGD), is used. Adam is widely used in machine learning. By individually adjusting the learning rate α for each parameter, Adam allows parameters with steeper gradients to receive smaller updates, while parameters with gentler gradients receive larger updates, automatically scaling the problem. Additionally, Adam reduces the oscillations that were a challenge with SGD by considering both the first moment $v\_{t}$ and the second moment $s\_{t}$ of past gradients.

The parameter θ is updated as follows:

$$v\_{t}=β\_{1}v\_{t-1}+\left(1-β\_{1}\right)G$$

$$s\_{t}=β\_{2}s\_{t-1}+\left(1-β\_{2}\right)G^{2}$$

$$θ\_{t+1}=θ\_{t}-α\frac{v\_{t}}{\sqrt{s\_{t}+ε}}$$

$$β\_{1}=0.9, β\_{2}=0.999, ε=10^{-8}, G:gradient of loss function$$

The initial value of the learning rate α is set to 0.0005, and the parameter update is performed 20,000 times.

# References

Betancourt, Michael. 2017. “A Conceptual Introduction to Hamiltonian Monte Carlo.” *ArXiv [Stat.ME]*. arXiv. http://arxiv.org/abs/1701.02434.

Birdsall, C. K., and A. B. Langdon. 2004. *Plasma Physics via Computer Simulation*. CRC Press.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. “Variational Inference: A Review for Statisticians.” *Journal of the American Statistical Association* 112 (518): 859–77.

Kingma, Diederik P., and Jimmy Ba. 2014. “Adam: A Method for Stochastic Optimization.” *ArXiv [Cs.LG]*. arXiv. http://arxiv.org/abs/1412.6980.

Kingma, Diederik P., and Max Welling. 2013. “Auto-Encoding Variational Bayes.” *ArXiv [Stat.ML]*. arXiv. http://arxiv.org/abs/1312.6114v11.

Neal., R. 2011. *Handbook of Markov Chain Monte Carlo, Chapter 5: MCMC Using Hamiltonian Dynamics*. CRC Press.

Ranganath, Rajesh, Sean Gerrish, and David Blei. 22--25 Apr 2014. “Black Box Variational Inference.” In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, edited by Samuel Kaski and Jukka Corander, 33:814–22. Proceedings of Machine Learning Research. Reykjavik, Iceland: PMLR.

Tierney, Luke. 1994. “Markov Chains for Exploring Posterior Distributions.” *The Annals of Statistics* 22 (4): 1701–28.

Wasserstein, Ronald L., Malvin H. Kalos, and Paula A. Whitlock. 1989. “Monte Carlo Methods, Volume 1: Basics.” *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences* 31 (2): 269.

Wingate, David, and Theophane Weber. 2013. “Automated Variational Inference in Probabilistic Programming.” *ArXiv E-Prints*, January, arXiv:1301.1299.