Bayesian Statistical Analysis for Mass Spectrometric Data Processing

Ellis McLarty[1,2], Elizabeth LaBone[1], Kyle Samperton[3], Shelby Bowden[3], Alexis Riche[3]

[1]Environment and Legacy Management, SRNL | [2]School of Mathematical and Statistical Sciences, Clemson University | [3]Global Security, SRNL

Introduction

Analytical techniques such as Thermal Ionization Mass Spectrometry (TIMS) are routinely employed at Savannah River National Laboratory (SRNL), other National Laboratories, and in academia to determine the precise isotopic composition of diverse natural and anthropogenic samples (e.g., rocks and nuclear materials). Robust uncertainty estimation is critical to accurate nuclear material characterization, control, and safety, with implications for nuclear nonproliferation, disposition, and forensics.

Quantifying and reporting uncertainty in such analyses, while regularly performed, have a rigorous statistical foundation. The *Guide to the Expression of Uncertainty in Measurement*[3] (GUM) outlines conventional techniques used to assess such uncertainty. As the accessibility and speed of statistical computing increase, there is a need to modernize conventional techniques. For example, Supplement 1[2] to the GUM suggests the use of approximation methods as an updated approach to the GUM.

We argue that a transition from frequentist to Bayesian inference, as employed in Supplement 1 to the GUM, is a worthy consideration in metrology. Specifically, we suggest the use of Markov chain Monte Carlo (MCMC) algorithms to better characterize the measurement uncertainty of traceable isotope ratios.

How do Markov chain Monte Carlo algorithms perform in assessing the measurement uncertainty of traceable isotope ratios?

Methods

To quantify the associated uncertainty in TIMS isotopic analyses, we employ a Metropolis-Hastings algorithm. This MCMC is a common approach in other fields but is not widely applied in mass spectrometry. For the first implementation, we propose a Gaussian likelihood function and uniform prior.

The implementation of our MCMC algorithm is shown to the right. TIMS analyses can be run at different turret levels, between 2 and 21, and these differences are accounted for in the code. There are additional tuning parameters in our work related to turret level such as providing relative noise based on the number of turret slots in use.

Note that a burn in period of 1e5 is used. Additionally, the acceptance level is approximately 30%.

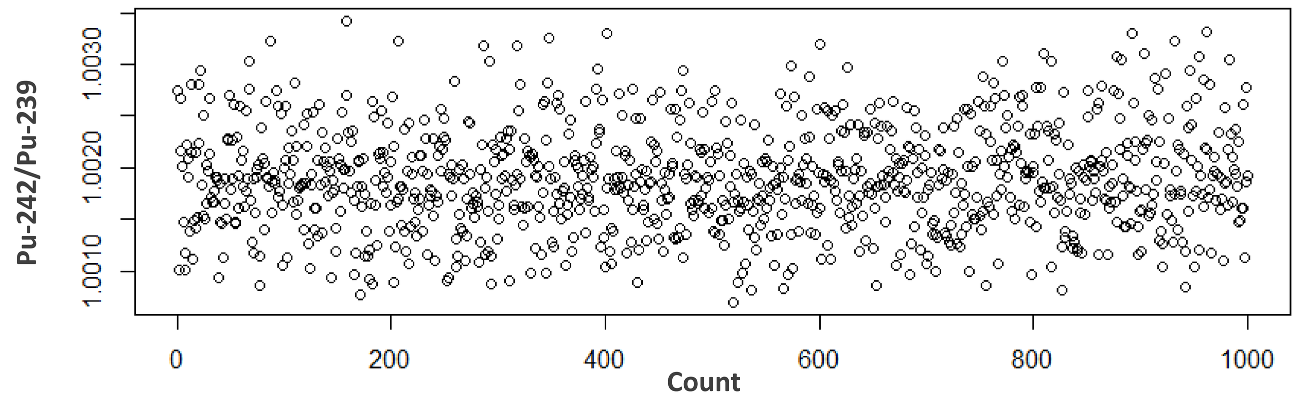
```
# MCMC algorithm:
proposalfunction = function(param){

  return(rnorm(2,
             mean = param,
             sd = c(0.0005,
                    0.000001)))
}


mu.true   <- mean(data) # input true mean value
sigma.true <- sd(data) # input true standard deviation value
turret_num <- 21

chain_list <- c()
acceptance_df <- data.frame()

# loop through iterations of model
for(k in 1:10){

  total_acceptance <- rep(NA, (turret_num-1))

  #loop through number of samples per turret
  for(turret_positions in 2:turret_num){
   print(turret_positions)
  # generate N synthetic analyses sampled from true distribution
  N <- turret_positions #number of Pu standard analyses
  sim <- rnorm(N,
               mu.true,
               sigma.true)
  mu.obs <- mean(sim)
  sigma.obs <- sd(sim)

  startvalue = c(mu.obs,
                  sigma.obs)

  chain = run_metropolis_MCMC(startvalue,
                              1e5,
                              sim)


  burnIn = 1e4
  acceptance = 1-mean(duplicated(chain[-(1:burnIn),]))
  total_acceptance[turret_positions] <- acceptance

 }

}
```

Workflow

Data Collection

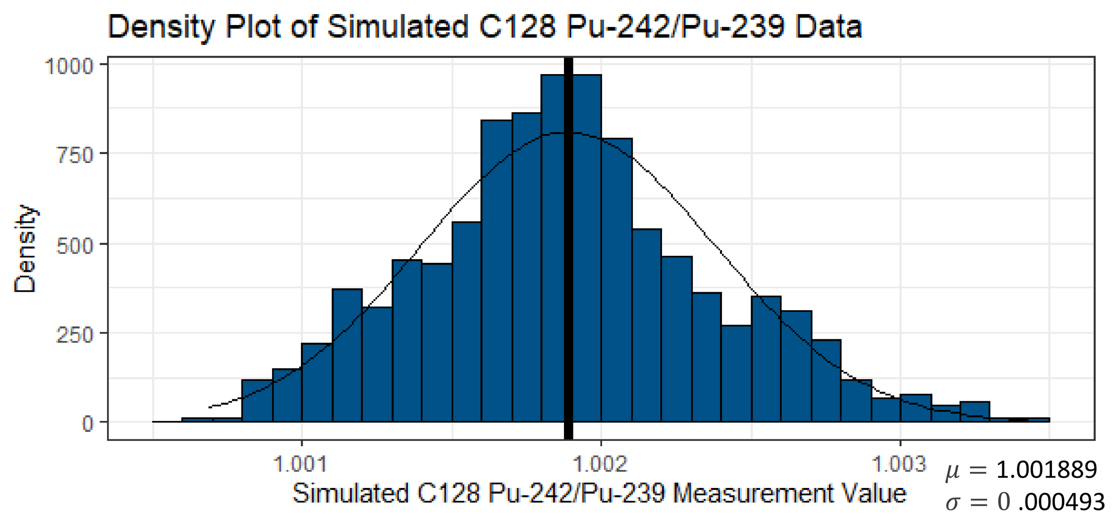
Thermo Scientific TRITON™ PlusTIMS at SRNL

Data Simulation

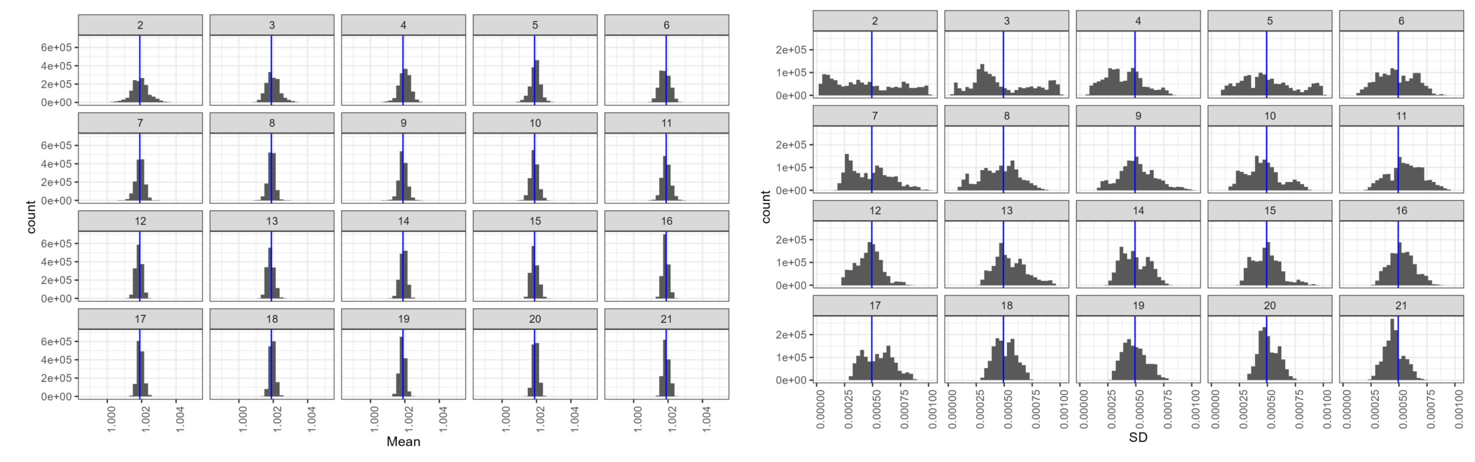


Simulated C128 Pu-242/Pu-239 TIMS Data

Data Distribution



Posterior Means and Standard Deviations

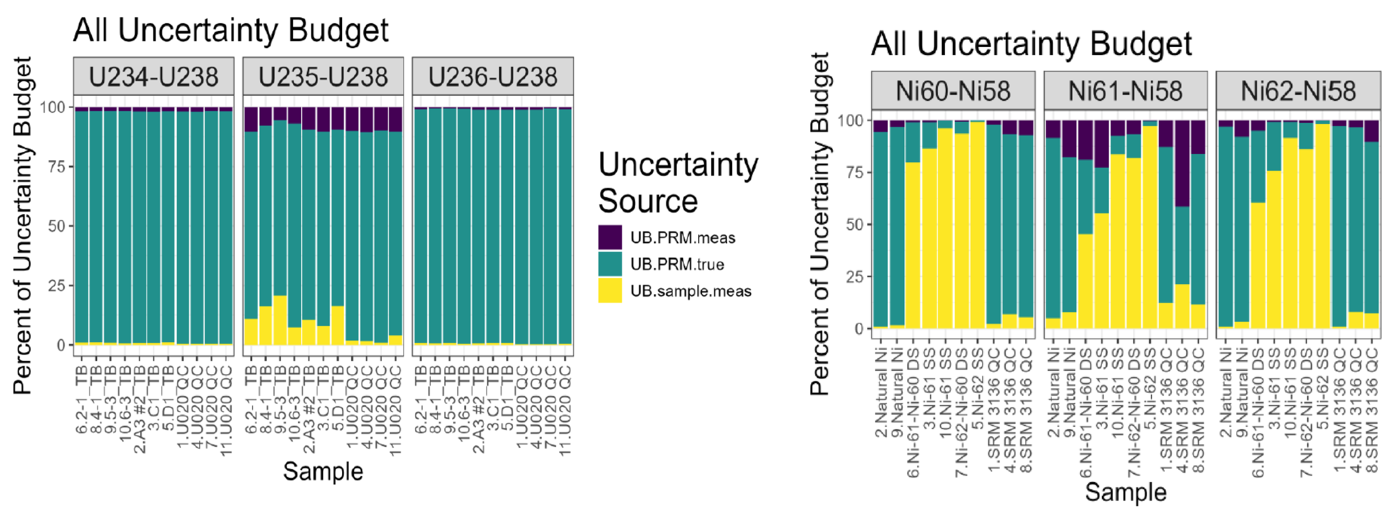


Discussion

Based on initial testing, an MCMC algorithmic framework proves viable in assessing the measurement uncertainty of TIMS data. Algorithm development is ongoing, but early-stage results are promising. Currently, the model assumes a fixed mean and variance for the priors and no overdispersion. Our next steps include increasing the flexibility and strength of prior data, accounting for overdispersion, and building robust data validation techniques.

Furthermore, this research has been developed in conjunction with an R package centered on a graphical user interface (GUI) for analyzing high-precision mass spectrometric data. There is currently a lack of advanced, generalized software for the complete analysis of raw data produced by modern isotope ratio mass spectrometers, including multicollector—inductively coupled plasma—mass spectrometry (MC-ICP-MS) and TIMS. Our work aims to fill this gap with a GUI accessible to those without programming experience, written in the open source and multi-platform programming language R.

This poster outlines the Bayesian approach to uncertainty quantification which will be readily accessible on the GUI. A traditional frequentist approach will be available as well to provide practitioners with the ability to toggle between statistical methods easily. In addition to TIMS plutonium measurements, recent work includes employing a Gibbs sampler to examine the uncertainty quantification of uranium isotopic data run on MC-ICP-MS. We will continue to research additional algorithmic approaches, expand to other elements, and tune the algorithms to reflect the respective measurement system.



Example output of the GUI; visualization of the overall uncertainty budget by uncertainty source for uranium (left) and nickel (right)

References

[1] Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer. 10.1007/978-0-387-92407-6

[2] JCGM. (2008). Evaluation of measurement data — supplement 1 to the "Guide to the expression of uncertainty in measurement" — propagation of distributions using a Monte Carlo method. https://doi.org/10.59161/jcgm101-2008

[3] JCGM. (2008). Guide to the expression of uncertainty in measurements. https://doi.org/10.59161/jcgm100-2008e

SRNL-STI-2025-00212

Acknowledgements

This effort was funded by SRNL LDRD-2024-00020 (*Rapid, rigorous, reproducible data analysis software for high-precision mass spectrometry*).

Savannah River National Laboratory®