

Bayesian Optimisation using microscopic and integral measurements to infer nuclear data parameters

Daan Houben, Mathieu Hursin

December 5, 2025

Abstract

Test 123

1 Introduction

Nuclear data is considered as the major source of uncertainty in some reactor observables, most notably effective multiplication factor (k_{eff}). The nuclear data available in evaluated nuclear data libraries, such as JEFF-4.0, are a result of a complex fitting procedure including theoretical models, microscopic experiments, expert judgement and sometimes integral experiments. Although integral experiments are often not directly included in the fitting procedure due to the computational cost, they form an important step towards validating the nuclear data vector. In this work, a Bayesian Optimization framework is proposed which enables consolidating microscopic energy dependent measurements with integral experiments to estimate nuclear data parameters.

Due to the computational cost of evaluating the performance of nuclear data in high-fidelity simulation software, most notably in integral experiments, Gaussian Processes (GPs) are trained. Subsequently, the GPs are used in a Markov Chain Monte Carlo (MCMC) algorithm in order to obtain a posterior estimate of the parameter(s) to infer. Two case studies are presented, first on the Γ_γ width of ^{53}Cr in the keV energy range using data recently published by P. Maroto et al. (2025) [1], complemented with integral experiments sensitive to ^{53}Cr . Secondly, on a hypothetical problem in which the integral experiments are more sensitive to the input parameters.

2 Background and mathematical motivation

2.1 Bayes theorem

2.2 Surrogate modelling

To enable efficient sampling of the parameter space, we replace the computationally expensive high-fidelity models $f(\boldsymbol{\theta})$ with Gaussian Process (GP) emulators. A GP defines a distribution over functions, specified by a mean function $m(\boldsymbol{\theta})$ and a covariance kernel $k(\boldsymbol{\theta}, \boldsymbol{\theta}')$.

For an unknown parameter set $\boldsymbol{\theta}^*$, the GP predicts the model output as a normal distribution:

$$P(f(\boldsymbol{\theta}^*) \mid \mathcal{D}_{\text{train}}) \sim \mathcal{N}(\mu_{\text{GP}}(\boldsymbol{\theta}^*), \Sigma_{\text{GP}}(\boldsymbol{\theta}^*)) \quad (1)$$

Where:

- $\mu_{\text{GP}}(\boldsymbol{\theta}^*)$ is the predicted mean response.
- $\Sigma_{\text{GP}}(\boldsymbol{\theta}^*)$ captures the epistemic uncertainty of the emulator itself.

In this work, the GP is trained on a set of input-output pairs generated from the high-fidelity models describing the microscopic energy dependent and integral experiments. For microscopic energy dependent experiments, the SAMMY resonance fitting tool is used to calculate a goodness-of-fit in terms of χ^2 . For the integral experiments, the continuous energy Monte Carlo transport code SERPENT-2 is used to obtain the calculated response. For any candidate $\boldsymbol{\theta}$ proposed during MCMC, the GP prediction $\mu_{GP}(\boldsymbol{\theta})$ is used as the model response.

2.3 Markov Chain Monte Carlo (MCMC) Framework

We seek the posterior distribution $P(\boldsymbol{\theta} | D)$ of the parameters given the experimental data D . According to Bayes' theorem

$$P(\boldsymbol{\theta} | D) \propto P(D | \boldsymbol{\theta})P(\boldsymbol{\theta}), \quad (2)$$

where $P(\boldsymbol{\theta} | D)$ is the posterior distribution, $P(D | \boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$ is the Likelihood (the goodness-of-fit) and $P(\boldsymbol{\theta})$ is the prior (initial beliefs about parameter bounds). $P(D | \boldsymbol{\theta})$ is evaluated using the product of all experimental combinations.

We employ the Emcee sampler algorithm to find the posterior distribution. The unnormalized posterior distribution is calculated using Eq. 2. The unnormalized posterior is transformed to the log-space to prevent numerical underflow.

2.4 Likelihood Formulation

The definition of the log-likelihood $\ln(\mathcal{L})$ depends on the nature of the experimental data. We distinguish between integral experiments (scalar values) and microscopic experiments (highly correlated data points). Under the assumption of Gaussian experimental errors, the probability density function (PDF) for an observations d given prediction $M(\boldsymbol{\theta})$ is given by

$$P(D|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(D - M(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1}(D - M(\boldsymbol{\theta}))\right) \quad (3)$$

Taking the natural logarithm yields the log-likelihood:

$$\ln(\mathcal{L}) = -\frac{1}{2} \underbrace{(D - M(\boldsymbol{\theta}))^T \boldsymbol{\Sigma}^{-1}(D - M(\boldsymbol{\theta}))}_{\chi^2} - \underbrace{\frac{1}{2} \ln((2\pi)^N |\boldsymbol{\Sigma}|)}_{\text{Constant}} \quad (4)$$

$$\ln(\mathcal{L}) = -\frac{1}{2}\chi^2 + C \quad (5)$$

In MCMC, we calculate the *difference* between log-likelihoods. Since the covariance matrix $\boldsymbol{\Sigma}$ (and thus the constant C) depends only on the data and not the parameters $\boldsymbol{\theta}$, it cancels out. Therefore, we evaluate the log-likelihood simply as:

$$\ln(\mathcal{L}) \propto -\frac{1}{2}\chi^2 \quad (6)$$

References

- [1] P. Pérez-Maroto, C. Guerrero, A. Casanovas, B. Fernández, E. Mendoza, V. Alcayne, J. Lerendegui-Marco, C. Domingo-Pardo, J. M. Quesada, R. Capote, and t. n. Collaboration, “50Cr and 53Cr neutron capture cross sections measurement at the n_tof facility at CERN,” 2025. Version Number: 1.