Using microscopic and integral measurements to infer nuclear data parameters

Daan Houben

November 27, 2025

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 including 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. A case study is performed 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.

2 Background and mathematical motivation

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:

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

Where:

- $\mu_{GP}(\boldsymbol{\theta}^*)$ is the predicted mean response.

- $\Sigma_{GP}(\boldsymbol{\theta}^*)$ 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.

3 Markov Chain Monte Carlo (MCMC) Framework

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

$$P(\boldsymbol{\theta} \mid D) \propto P(D \mid \boldsymbol{\theta})P(\boldsymbol{\theta}), \tag{2}$$

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

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

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

Taking the natural logarithm yields the log-likelihood:

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

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

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:

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

4.1 Integral Experiments (Scalar Data)

Integral experiments provide a single measured value d_{int} with an associated experimental uncertainty σ_{int}. The Gaussian Process predicts a scalar mean μ_{GP} and variance σ_{GP}^2. To account for both experimental error and emulator uncertainty, we sum the variances:

$$\chi^2_{int} = \frac{(d_{int} - \mu_{GP}(\boldsymbol{\theta}))^2}{\sigma_{int}^2 + \sigma_{GP}^2(\boldsymbol{\theta})} \tag{7}$$

The contribution to the log-likelihood is:

$$ll_{int} = -0.5 \cdot \chi^2_{int} \tag{8}$$

4.2 Microscopic Experiments (Vector Data with High Correlation)

Microscopic experiments yield a vector of N data points $D_{micro} = [d_1, ..., d_N]$ (e.g., a spectrum or time series). However, exploratory analysis revealed extreme multicollinearity in the data (correlation coefficients $\rho > 0.99$), rendering the covariance matrix singular and ill-conditioned.

Treating these N points as independent would falsely inflate the degrees of freedom, leading to artificially narrow posterior distributions. To address this, we adopt a **dimensionality reduction approach**, treating the highly correlated vector as a single effective data point representing the average behavior.

4.2.1 Methodology: The "Super-Point" Approximation

1. **Data Reduction:** We calculate the scalar mean of the experimental vector:

$$\bar{D} = \frac{1}{N} \sum_{i=1}^{N} d_i \tag{9}$$

2. **Model Reduction:** We similarly average the GP prediction vector:

$$\bar{M}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^{N} \mu_{GP,i}(\boldsymbol{\theta}) \tag{10}$$

3. **Uncertainty Quantification:** Due to the near-perfect correlation, the uncertainty of the mean does *not* decrease by $1/\sqrt{N}$. Instead, the relative uncertainty of the "super-point" is constrained by the representative experimental uncertainty σ_{exp}:

$$\sigma_{eff} \approx \sigma_{exp} \tag{11}$$

The effective Chi-squared is calculated as:

$$\chi^2_{micro} = \left(\frac{\bar{D} - \bar{M}(\boldsymbol{\theta})}{\sigma_{eff}} \right)^2 \tag{12}$$

This formulation prevents the "double-counting" of information inherent in strongly correlated datasets while preserving the global signal-to-noise ratio. The log-likelihood contribution is:

$$ll_{micro} = -0.5 \cdot \chi^2_{micro} \tag{13}$$

5 Total Log-Likelihood

Assuming independence between the integral and microscopic experiments, the total log-likelihood used to drive the MCMC sampling is the weighted sum of the individual components:

$$\ln(\mathcal{L}_{total}) = \sum w_{int} \cdot ll_{int} + \sum w_{micro} \cdot ll_{micro} \tag{14}$$

This methodology ensures a statistically robust parameter estimation that honors the physical constraints of the data, the uncertainty of the surrogate model, and the effective information content of correlated experiments.

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.