Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.OA-2025-0025

Vol. **18**, No. 4, pp. 1066-1099
November 2025

Machine Learning-Based Bias Correction for the GEIM Model

Taidong Niu[1†], Mingzhe Lyu[1,2†], Helin Gong[3], Han Zhang[1], Zhang Chen[4] and Zhiyong Wang[1,*]

[1] *School of Mathematical Sciences, University of Electronic Science and Technology of China, Sichuan 611731, China*
[2] *Shanghai Zhangjiang Institute of Mathematics, Shanghai 201203, China*
[3] *Paris Elite Institute of Technology, Shanghai Jiao Tong University, Shanghai 200240, China*
[4] *Nuclear Power Institute of China, Chengdu 610041, China*

Received 28 February 2025; Accepted (in revised version) 2 June 2025

Abstract. This study focuses on real-time reconstruction of the spatial distribution of nuclear power using limited measurement observations. While physical models, such as the generalized empirical interpolation method (GEIM), can reconstruct the spatial field, they often cause bias if the model used to construct it is biased. The parametrized background data weak (PBDW) method attempts to mitigate this model bias, but its effectiveness is limited. To improve model bias correction, this paper proposes leveraging machine learning techniques-specifically, support vector regression, K-nearest neighbors, and decision trees to enhance the GEIM method. These techniques predict model bias distributions across the entire field based on observed model bias at measurement points. The results demonstrate that Gaussian process based correction performs comparably to PBDW, both offering superior accuracy and robustness against noise, while other machine learning methods exhibit instability under varying parameter settings.

AMS subject classifications: 65M32

Key words: Model order reduction, model bias correction, Gaussian processes, generalized empirical interpolation method, physical field reconstruction.

1. Introduction

Accurate modeling and optimization are fundamental to improving reactor performance and ensuring the safety and regulatory compliance of nuclear facilities [11, 12, 45]. Nuclear reactors are inherently complex systems, where various physical

*Corresponding author. *Email address:* `zhywang@uestc.edu.cn` (Z. Wang)

†Contributed equally

phenomena-such as neutron transport, fluid dynamics, and heat transfer-interact in highly nonlinear and coupled ways. These complexities make it difficult to construct high-fidelity models that are accurate across a broad range of operating conditions. While full-order models offer high accuracy, they are often computationally intensive, rendering them impractical for real-time applications in safety-critical scenarios [38, 39]. The challenge is further compounded by the high dimensionality and large number of variables involved.

To improve computational efficiency while preserving predictive accuracy, significant research has focused on developing methods for reconstructing reactor fields from sparse measurements, enabling faster and more adaptive predictions in practice [44]. Such methods integrate physical modeling with real-world data to produce real-time approximations of system behavior, thereby improving control and predictive capabilities.

Recent progress in data assimilation and model order reduction (MOR) offers new opportunities for efficient field reconstruction. Among MOR techniques, proper orthogonal decomposition (POD) is widely adopted for extracting dominant mode shapes from experimental data or high-fidelity simulations [5, 31]. POD enables the representation of complex high-dimensional systems using a small number of basis functions, significantly reducing computational cost.

Another prominent method is the generalized empirical interpolation method, which extends the classical EIM by jointly selecting basis functions and observation sensors to minimize interpolation error [27]. GEIM has shown strong performance in Banach spaces and enhanced stability over standard interpolation schemes [2, 32]. Despite these advantages, reduced-order models (ROMs) are still subject to model bias, which can arise from several sources:

(1) Simplified or imperfect physical modeling assumptions [24].

(2) Narrow sampling of parameter spaces [14].

(3) Insufficient basis function richness.

(4) Observation noise or numerical artifacts [18, 37].

These biases hinder the ability of ROM to generalize and perform reliably, especially under changing operational conditions.

To address model bias, the parametrized background data weak method combines reduced modeling with data assimilation, using update spaces to adjust for discrepancies [29]. However, the performance of PBDW is fundamentally limited by the quality of the chosen basis and the dimensionality of the reduced spaces, often resulting in suboptimal bias correction.

In recent years, machine learning (ML) approaches have shown great potential in physical field reconstruction tasks [13, 20, 36]. Techniques such as Gaussian processes (GP), support vector regression (SVR), K-nearest neighbors (KNN), and decision trees (DT) are capable of learning complex mappings from data, making them promising