POLITECNICO DI TORINO

Master's Degree in Nuclear and Energy Engineering
Sustainable Nuclear Energy



Master Thesis

Machine Learning-Based Interpolation Schemes for the Fission Matrix Method

Advisors:
Dr. Abrate
Prof. Dulla
Ing. Vita

Candidate
Marco Betti - 331325

Academic Year 2025 - 2026

Abstract

The accurate and efficient simulation of neutron transport phenomena in nuclear reactor cores is a computationally demanding task, particularly when high-fidelity models are employed. Among the different approaches used to describe neutron behaviour, the fission matrix method represents a possible solution to have reliable results in a computationally efficient way.

This work investigates the development and optimisation of a machine learning-based surrogate modelling framework for the generation of Fission Matrices, with application to the VTB HPMR reactor. The reference dataset is obtained through high-fidelity simulations performed with the FIM4NTRA code, which provides physically consistent Fission Matrix solutions used as training and validation data.

The proposed surrogate model is based on the Proper Orthogonal Decomposition (POD) combined with Kriging regression. The POD is employed to reduce the dimensionality of the Fission Matrix representation by extracting the most relevant spatial modes, while the Kriging model is used to map the input parameter space to the reduced latent space. The reconstructed Fission Matrices are obtained through the inverse projection of the predicted POD coefficients.

A central aspect of this work is a parametric study aimed at evaluating the influence of the active learning strategy on the overall performance of the surrogate. In particular, the effect of a centrality-based penalization mitigating the spontaneous over-sampling of the domain boundaries is systematically analysed, together with a comparison between single-model and ensemble-based uncertainty metrics as candidate selection criteria. The impact of these choices on the convergence rate and the stability of the surrogate accuracy is assessed in order to identify the best configuration able to reduce the number of high-fidelity simulations required to achieve a given level of accuracy.

The results demonstrate that the proposed POD-Kriging framework is capable of accurately reproducing reference Fission Matrices while significantly reducing the computational cost compared to direct high-fidelity simulations. The study paves the way for more complex applications involving a larger number of parameters, where keeping the database size as small as possible becomes essential to maintain the attractiveness of the proposed methodology.

Contents

1 Introduction **1**
- 1.1 Background and motivation . . . 1
- 1.2 Objective of the thesis . . . 2
- 1.3 Original contributions . . . 2
- 1.4 Limitations . . . 3
- 1.5 Thesis Outline . . . 3

2 The Fission Matrix Method **5**
- 2.1 Introduction . . . 5
- 2.2 Theoretical Derivation of the Method . . . 6
  - 2.2.1 Integral formulation of the transport problem . . . 6
  - 2.2.2 Adjoint formulation . . . 7
  - 2.2.3 Discretization: the forward fission matrix . . . 8
  - 2.2.4 Discretization: the adjoint fission matrix . . . 8
  - 2.2.5 Relation between the direct and adjoint fission matrix . . . 9
- 2.3 Applications of the Fission Matrix . . . 9
  - 2.3.1 Higher eigenmodes of the discretized problem . . . 9
  - 2.3.2 Higher-Mode Fluxes and Reaction Rates . . . 10
  - 2.3.3 From a Single Matrix to a Reactor Model: The RAPID Approach . . . 10
- 2.4 FIM4NTRA Code . . . 11
- 2.5 Summary . . . 11

3 Machine Learning Theoretical Background **13**
- 3.1 Motivation . . . 13
- 3.2 Proper Orthogonal Decomposition . . . 13
  - 3.2.1 Construction of the snapshot matrix . . . 14
  - 3.2.2 POD as an eigenproblem . . . 14
  - 3.2.3 The Snapshot Method and its SVD formulation . . . 15
  - 3.2.4 Mode Truncation and Dimensionality Reduction . . . 16
- 3.3 Kriging: Gaussian Process Regression . . . 16
  - 3.3.1 Definition and probabilistic setting . . . 16
  - 3.3.2 Joint distribution and predictive equations . . . 17
  - 3.3.3 Kernel choice and hyperparameter estimation . . . 18
  - 3.3.4 Application to the POD-reduced fission matrix surrogate . . . 19

4 Methodology and computational framework **21**
4.1 Reactor and geometry . . . 21
4.2 Parameter space definition . . . 23
4.3 Surrogate Calibration . . . 24
4.3.1 Mode truncation . . . 24
4.3.2 Kernel and hyperparameter configuration . . . 26
4.3.3 Truncation depth within the active learning loop . . . 26
4.3.4 Exclusion of already measured points . . . 27
4.4 Acquisition score . . . 28
4.5 Adaptive learning algorithm . . . 29

5 Parametric study and results **33**
5.1 Centrality Correction . . . 33
5.1.1 Boundary weighting formulation . . . 33
5.1.2 Boundary weight tuning . . . 34
5.1.3 Spatial distribution of selected points . . . 35
5.1.4 Evolution of uncertainty and acquisition score during the loop . . . 35
5.2 Sanity Check Against Random Sampling . . . 38
5.2.1 Verification of the Adaptive Learning sampling . . . 38
5.2.2 Results . . . 38
5.2.3 Interpretation . . . 40
5.3 Power distribution error . . . 40
5.3.1 Fission source error . . . 40
5.3.2 Generalisation to an independent test set . . . 41
5.4 Fission matrix reconstruction versus direct prediction . . . 42
5.5 Computational cost of the surrogate . . . 43

6 Conclusions and Future Work **45**
6.1 Necessity of a boundary correction . . . 46
6.2 Choice of uncertainty source . . . 46
6.3 Mode Truncation Criterion . . . 47
6.4 Limitations of the proposed approach . . . 47
6.5 Future work . . . 48

A Sobol Low-Discrepancy Sequences **49**

Chapter 1

Introduction

1.1 Background and motivation

The modelling and prediction of the neutronics of a nuclear reactor core across its full range of operating conditions is a central task of reactor physics design and safety analysis. Nowadays, the design stage of the core is carried out using deterministic diffusion and transport codes due to their relatively easy implementation and smaller computational cost, but paying the price of a more difficult representation of complex geometries and worse results accuracy due to numerical approximations. The Monte Carlo method, a high-fidelity statistical approach, is on the other hand often considered, yet with some limits (summarized in 2.2), the gold standard [1] for geometry precision and results accuracy, but its heavy computational cost relegates it mostly to the verification stage of a design. The fission matrix (FM) method sits in the middle, representing a bridge between these two classes of methods. The heart of the FM approach consists in the deterministic solution of a linear eigenvalue problem, in which the elements of a matrix operator, the fission matrix, represent the average number of neutrons born in a spatial region of the core due to an average neutron born by fission in another spatial region [2]; the dominant eigenpair of this operator reproduces the system multiplication factor k_{eff} and the spatial shape of the fission source. Once available, the fission matrix is a physically consistent representation of the reactor's neutronic state. Starting from a reduced database of Monte-Carlo-generated fission matrices, describing key parametrized operational stages, any other FM from the parameter space can be quickly generated via interpolation, and then solved via fast deterministic techniques. Here lies the interest in this method: combining high-fidelity and deterministic approaches to balance the pros and cons of both methods [2]. The accuracy of the FM technique depends on the resolution of the database and the the kind of interpolation used. A higher resolution would give a more accurate surrogate model, but since the available data in a Monte Carlo environment is very computationally expensive, the dimension of the dataset has to be built trying to maximize performances at a reduced quantity of data. While most of the literature focuses on non-adaptive sampling strategies to populate the parameter space (e.g. uniform grids, Latin Hypercube, or Sobol sequences) [2] [3], the focus of this work is on interpolation scheme itself: testing machine learning, in particular POD (Proper Orthogonal Decomposition) coupled with Kriging, as an alternative matrix interpolation scheme, combined with

an uncertainty driven adaptive sampling startegy.

This thesis addresses the Virtual Test Bed (VTB) Heat Pipe Microreactor (HPMR), whose operating regime is parametrised by the core temperature $T \in [300, 1000]\,\mathrm{K}$ and the control-drum rotation angle $\theta \in [0°, 180°]$. High-fidelity fission matrices for this core are generated, as training data for the model, with the FIM4NTRA code [4] in place of actual Monte Carlo simulations. This choice (Section 2.4) is motivated by FIM4NTRA's computational cost, much smaller with respect to Monte Carlo, without compromising the study's scope of interpolating a given set of matrices. A single evaluation is not prohibitively expensive by itself, but a design or safety study that requires the fission matrix as a continuous function of (T, θ), rather than at a handful of nominal points, quickly becomes impractical if done by random or uniform sampling of the parameter space, requiring a much higher number of high-fidelity evaluations.

1.2 Objective of the thesis

The objective of this work is to demonstrate that fission matrices representing the physics of the VTB HPMR core can be generated over the full (T, θ) operating ranges using significantly fewer high-fidelity FIM4NTRA evaluations than brute-force sampling would require, by choosing a data-driven surrogate model trained through adaptive sampling over a simple uniform parameter space subdivision. The surrogate combines a linear dimensionality reduction of the fission matrix using Proper Orthogonal Decomposition (POD) with a Kriging (also known as Gaussian Process Regression, GPR [5]) model, which selects iteratively the points in the parameter space to evaluate next on the basis of their uncertainty, adaptively individuating the areas with higher improvement margin.

The key technical matter addressed in this work is the proper tuning of the parameters of this adaptive learning strategy, together with the implementation of corrections to limit its biases and to maximize its prediction accuracy. As shown in Chapter 5 an uncorrected posterior-variance criterion is found to be as relevant as random sampling on this problem, because of a systematic tendency to over-sample the boundary of the parameter domain.

1.3 Original contributions

This work makes the following contributions:

- A practical proof that the standard criterion for POD mode truncation based on the fraction on variance explained is misleading for this application: truncation orders that explain even more than 99.7% of the variance can still leave a considerably larger error on k_{eff} than a marginally deeper truncation (Section 4.3.1), motivating the use of the full available rank of the training data as the operational truncation order rather than a fixed cut-off based on the explained variance.
- A comparison, as candidate sources of predictive uncertainty for the POD-Kriging, between the single-model Gaussian Process posterior variance and the disagreement

among an ensemble of GPs, together with empirical justification, based on the stability of the fitted hyperparameters and the log marginal likelihood of the training data.

- An adaptive sampling strategy for the surrogate construction of fission matrices, in which candidate points are ranked using an acquisition score that combines the GP posterior predictive variance with a centrality-based penalization term (Section 5.1.1) correcting the tendency of raw predictive-variance sampling to oversample the domain boundaries..

- A quantitative description of the boundary-oversampling bias justifying the need for this correction: without it, 48% of the points selected by the surrogate model fall on the domain boundary, against about 17% found under random sampling (Section 5.1.2), yielding a model which is over-trained in reproducing the boundary but with too little information on the centre part of the domain.

- An assessment of the final surrogate against a random sampling strategy, proving that the accuracy gain from the score-based acquisition strategy increases not due to the simple addition of points to the dataset, but to the addition of points with higher importance for the surrogate model (Section 5.2).

1.4 Limitations

The limitations of this work are the following. To reduce the computational burden, the results presented in this work are obtained against the FIM4NTRA code, instead of a Monte Carlo transport code reference such as Serpent 2. The parameter space explored is two-dimensional, and the active learning cycle studied throughout Chapter 5 spans 40 iterations from an initial dataset of 16 points. The baseline acquisition criterion is a procedure based on a score that depends on the posterior variance of the POD-Kriging model: the point with the highest score at every iteration is therefore the "weakest" point in terms of quality of the information, and it thus is selected to be added to the the dataset. Furthermore no statistical noise is propagated onto the POD coefficients during training. The acquisition criterion, based purely on the GP posterior variance, does not differentiate between the uncertainty due to sparse sampling and the aleatoric uncertainty due to MC statistical noise. These limitations are discussed further, and their implications for generalising the proposed strategy beyond the present case study assessed, in Chapter 6.

1.5 Thesis Outline

The thesis content is structured as follows. Chapter 2 introduces the neutron transport background and the fission matrix theoretical derivation, its applications and the specificities of the FIM4NTRA code. Chapter 3 presents the theoretical background for the two machine learning blocks of the surrogate: the Proper Orthogonal Decomposition and Kriging, underlying the important figures of merit and parameters. Chapter 4 describes the specific reactor core geometry, the parameter domain and the adaptive learning workflow. This chapter

contains also the surrogate calibration: the POD truncation, the kernel and hyperparameter configuration configuration, and the choice of uncertainty source. Chapter 5 presents the main results of this work: the centrality correction and its calibration, the sanity check against random sampling, and the validation of the resulting surrogate on an independent test set and against direct prediction of k_{eff} and the fission source. Chapter 6 discusses the results and limitations of these this study and outlines directions for future work.

Chapter 2

The Fission Matrix Method

2.1 Introduction

Continuous-energy Monte Carlo codes are widely regarded as the reference tool for reactor physics analysis [2], since they can represent geometry, energy, and angular variables without the discretization approximations required by deterministic transport methods. An important limit of these codes, however, is the elevated computational price. Criticality calculations are performed through the power method, where batches of neutrons are tracked generation by generation. Since individual batch estimates are affected by statistical noise, the outer-iteration acceleration schemes commonly used in deterministic codes (e.g. Chebyshev acceleration) cannot be applied [2], and full three-dimensional, pin-resolved power distributions with many axial levels are difficult to obtain with an acceptable statistical uncertainty in a reasonable computing time. Two additional difficulties intensify this cost: convergence of the fission source can be even slower in loosely coupled systems (large dominance ratio) [1], and standard Monte Carlo provides no direct route to the higher eigenmodes of the transport operator [2], which are useful tools for perturbation theory, kinetics, and diagnosing source convergence.

The baseline of full-core design and safety analysis is still represented by deterministic diffusion and transport codes, precisely because they are fast enough to be run many times, but their accuracy can be difficult to predict *a priori*, particularly for geometries or reactor types (research reactors, microreactors, nuclear thermal propulsion cores) that are not meant to fully fit the assumptions for diffusion theory. Between these two extremes, a class of hybrid methods has been created with the objective of combining the accuracy of a Monte Carlo solution with a fast, deterministic-like solve for arbitrary system states. The *fission matrix* method belongs to this class, with theoretical basis established for continuous-energy transport by Carney et al. [2]. The general formulation of the method involves the reworking of the fission source eigenvalue problem as a finite linear system whose coefficients are tallied, without approximation, during an ordinary Monte Carlo criticality run.

The idea of using the fission matrix for criticality problems is not a recent concept, dating back to the earliest Monte Carlo literature [6, 7], but it long remained a subject of mostly

academic interest because of the prohibitive storage and computational cost of the full matrix for realistic geometries. However, over the last two decades the method has gained some interest due to both algorithmic advances and to the recognition that a single, precomputed fission matrix can be reused, at negligible additional cost, to answer questions about other operational states that would otherwise each require a dedicated Monte Carlo run.

This chapter is organized as follows. Section 2.2 derives the forward and adjoint fission matrix equations from the continuous-energy transport equation, following Carney et al. [2]. Section 2.3 then turns to the question of what further information can be extracted from the forward and adjoint fission matrices once they are available. Specifically, it is shown how the fission matrix gives direct access to the higher eigenmodes of the transport operator, and how the corresponding higher-mode fluxes and reaction rates can in turn be obtained. Section 2.5 closes the chapter with a brief summary.

2.2 Theoretical Derivation of the Method

The *fission matrix* method addresses two well-known limitations of Monte Carlo criticality codes [2]: the impossibility of applying classical outer-iteration acceleration schemes, made ineffective by the statistical noise inherent to batch-wise tallies [2], and the inability to directly reconstruct higher eigenmodes of the fission source distribution. The core idea is to discretize the spatial domain and represent the neutron transport operator, restricted to the fission source alone, as a matrix whose eigenvalues and eigenvectors approximate the spectrum of the continuous transport equation. The derivation below is a condensed reformulation of the theoretical treatment given by Carney et al. [2].

2.2.1 Integral formulation of the transport problem

The fission matrix uses the Green's function approach to neutron transport. Instead of solving directly for the flux, the response of the system to a point (Dirac-delta) source placed anywhere in phase space is characterized, and the physical flux is then reconstructed as a superposition of these elementary responses.

The k-eigenvalue form of the neutron transport equation is:

$$\mathcal{M}\,\phi(\vec{r}, E, \hat{\Omega}) = \frac{1}{k}\,\frac{\chi(E)}{4\pi}\,S(\vec{r}), \tag{2.1}$$

where $\mathcal{M}$ is the net-loss operator (streaming, total collision, and scattering), $\phi(\vec{r}, E, \hat{\Omega})$ is the angular flux that represents the travelled path length per unit volume, energy and angle, $\chi(E)$ is the fission neutron emission energy spectrum (assumed isotropic and independent of the position and energy of the fissioning neutron), and $S(\vec{r})$ is the fission source, defined as:

$$S(\vec{r}) = \oint d\hat{\Omega}' \int_0^{+\infty} dE' \nu\Sigma_F(\vec{r}, E')\,\phi(\vec{r}, E', \hat{\Omega}'). \tag{2.2}$$

Let $G(\vec{r}_0, E_0, \hat{\Omega}_0 \to \vec{r}, E, \hat{\Omega})$ be the Green's function, defined as the solution of the transport equation with a Dirac-delta source in phase space. By linearity, the angular flux in the

fission source can be written as a superposition of these point-source responses, weighted by the emission spectrum $\chi(E)$.

In a multiplying system, however, $S(\vec{r})$ is itself generated by the flux it contains, through Eq. (2.2). Adapting the Green's function approach to this case means including the fission response back into the source term, yielding the k-eigenvalue form of the Peierls equation [2]. Multiplying by $\nu\Sigma_F(\vec{r}, E)$ and integrating over energy and angle then produces a closed integral equation for the fission source alone:

$$S(\vec{r}) = \frac{1}{k} \int d\vec{r}_0 \; S(\vec{r}_0)\, H(\vec{r}_0 \to \vec{r}), \tag{2.3}$$

where the kernel $H(\vec{r}_0 \to \vec{r})$ is the energy-angle-averaged Green's function, weighted by the incoming emission spectrum and the outgoing fission production rate:

$$H(\vec{r}_0 \to \vec{r}) = \oint d\hat{\Omega} \oint d\hat{\Omega}_0 \int_0^{+\infty} dE \int_0^{+\infty} dE_0 \;\; \nu\Sigma_F(\vec{r}, E)\, \frac{\chi(E_0)}{4\pi}\, G(\vec{r}_0, E_0, \hat{\Omega}_0 \to \vec{r}, E, \hat{\Omega}). \tag{2.4}$$

This purely spatial kernel maps a unit fission source at $\vec{r}_0$ to the fission-neutron production it induces at $\vec{r}$. The kernel hence the continuous counterpart of the fission matrix. The key point is that H can be estimated *without any approximation* from a Monte Carlo simulation (except the statistical uncertainty). The Green's function is exactly what simulated particle transport produces, and the energy-angle integration defining H reduces to a simple spatial tally, binned according to the neutron's initial and final spatial position.

2.2.2 Adjoint formulation

The adjoint formulation follows an entirely analogous path. The adjoint k-eigenvalue transport equation:

$$\mathcal{M}^\dagger \, \phi^\dagger(\vec{r}, E, \hat{\Omega}) = \frac{1}{k}\, \nu\Sigma_F(\vec{r}, E)\, S^\dagger(\vec{r}), \tag{2.5}$$

with adjoint source $S^\dagger(\vec{r}) = \iint dE'\, d\hat{\Omega}' \;\; \frac{\chi(E')}{4\pi}\, \phi^\dagger(\vec{r}, E', \hat{\Omega}')$, admits an adjoint Green's function $G^\dagger$; multiplying by $\chi(E)/4\pi$ and integrating over energy and angle, exactly as before, leads to an integral equation for the adjoint source:

$$S^\dagger(\vec{r}) = \frac{1}{k} \int d\vec{r}_0 \; S^\dagger(\vec{r}_0)\, H^\dagger(\vec{r}_0 \to \vec{r}). \tag{2.6}$$

Using the reciprocity relation between the direct and adjoint Green's functions:

$$G^\dagger(\vec{r}_0, E_0, \hat{\Omega}_0 \to \vec{r}, E, \hat{\Omega}) = G(\vec{r}, E, \hat{\Omega} \to \vec{r}_0, E_0, \hat{\Omega}_0), \tag{2.7}$$

(valid despite the energy asymmetry introduced by neutron slowing-down) one shows that the adjoint kernel equals the direct kernel with its arguments swapped:

$$H^\dagger(\vec{r}_0 \to \vec{r}) = H(\vec{r} \to \vec{r}_0), \tag{2.8}$$

so that the adjoint integral equation becomes

$$S^\dagger(\vec{r}) = \frac{1}{k}\int d\vec{r}_0\; S^\dagger(\vec{r}_0)\, H(\vec{r}\to\vec{r}_0). \tag{2.9}$$

It is worth noting that, while for the one-speed transport problem the kernel H is symmetric (so that the operator is self-adjoint), this symmetry does not generally hold for the multigroup or continuous-energy case. As a consequence, the existence of real eigenvalues and real eigenfunctions for higher modes cannot be proven rigorously in general; it is instead adopted as a working assumption [2] and checked numerically. Denoting by k_n, $S_n(\vec{r})$, and $S_n^\dagger(\vec{r})$, respectively, the eigenvalue, direct eigenfunction, and adjoint eigenfunction of order n (with $n = 0$ the fundamental mode), the mode-resolved versions of Eqs. (2.3) and (2.9) satisfy the orthogonality relation

$$(k_p - k_q)\int d\vec{r}\; S_p(\vec{r})\, S_q^\dagger(\vec{r}) = 0, \tag{2.10}$$

which guarantees biorthogonality between direct and adjoint modes of different order, but *not* orthogonality among the direct modes alone.

2.2.3 Discretization: the forward fission matrix

To make the problem numerically tractable, the spatial domain is divided into N regions. Integrating Eq. (2.3) over the volume V_i of the final region i, and normalizing by the region-integrated source S_j of the starting region j, yields the discrete form:

$$S_i = \frac{1}{k}\sum_{j=1}^{N} F_{i,j}\, S_j, \qquad F_{i,j} = \int_{V_i} d\vec{r}\int_{V_j} d\vec{r}_0\; \frac{S(\vec{r}_0)}{S_j}\, H(\vec{r}_0\to\vec{r}), \tag{2.11}$$

with $S_j = \int_{V_j} S(\vec{r}_0)\, d\vec{r}_0$. The matrix element $F_{i,j}$ has an immediate physical interpretation: it is the expected number of fission neutrons produced in region i per average fission neutron born in region j. In matrix–vector form:

$$\vec{S} = \frac{1}{k}\, F\, \vec{S}, \tag{2.12}$$

where F is the $N \times N$ *fission matrix* and k formally coincides with the effective multiplication factor k_{eff}. The accuracy of this discretization depends on mesh refinement: only for sufficiently small regions does the ratio S_i/V_i correctly approximate the local continuous source $S(\vec{r})$.

2.2.4 Discretization: the adjoint fission matrix

In a perfectly symmetric fashion, integrating Eq. (2.9) over the same regions gives the adjoint fission matrix:

$$S_i^\dagger = \frac{1}{k}\sum_{j=1}^{N} F_{i,j}^\dagger\, S_j^\dagger, \qquad F_{i,j}^\dagger = \int_{V_i} d\vec{r}\int_{V_j} d\vec{r}_0\; \frac{S^\dagger(\vec{r}_0)}{S_j^\dagger}\, H(\vec{r}\to\vec{r}_0), \tag{2.13}$$

or, in matrix form, $\vec{S}^\dagger = \frac{1}{k} F^\dagger \vec{S}^\dagger$. The components of $\vec{S}^\dagger$ represent the importance, with respect to overall neutron multiplication, of fission neutrons born in each region.

2.2.5 Relation between the direct and adjoint fission matrix

Comparing the element $F_{i,j}$ of the direct matrix with the corresponding transposed element $F^\dagger_{j,i}$ of the adjoint matrix, one finds that both integrals involve the same energy-angle kernel (the same Green's function and cross-section data), but are weighted by different spatial weighting functions: $S(\vec{r})/S_j$ in the direct case, $S^\dagger(\vec{r})/S^\dagger_i$ in the adjoint case. If the mesh is fine enough such that these weights are approximately uniform within each region, i.e.

$$\frac{S^\dagger(\vec{r})}{S^\dagger_i/V_i} \approx 1 \quad \forall\, \vec{r} \in V_i, \qquad \frac{S(\vec{r}_0)}{S_j/V_j} \approx 1 \quad \forall\, \vec{r}_0 \in V_j, \tag{2.14}$$

then the two integrals coincide numerically, giving the central result

$$F^\dagger = F^T \tag{2.15}$$

In other words, in the limit of fine spatial discretization the adjoint fission matrix equals the transpose of the direct matrix. The right eigenvectors of F yield the direct source modes, while the left eigenvectors (equivalently, the right eigenvectors of F^T) yield the adjoint modes. Besides being consistent with the reciprocity relation of the continuous problem (Eq. (2.9)), this property is the practical foundation for computing higher eigenmodes with the method, since it removes the need to separately tally F and $F^\dagger$ during the Monte Carlo simulation: accumulating the direct matrix F alone, and taking its transpose, is sufficient to also obtain the adjoint spectrum. This is the point of departure for the applications discussed next. The treatment of the adjoint equation is present for theoretical completeness, even though the computation of higher modes goes beyond the scope of this study.

2.3 Applications of the Fission Matrix

Once the forward and adjoint fission matrices of Sec. 2.2 are available, the k_{eff} and the fundamental-mode fission source can be extracted. However, these are not the only pieces of information that a single tallied fission matrix carries.

2.3.1 Higher eigenmodes of the discretized problem

Section 2.2 showed that, once the spatial domain is divided into N regions, the k-eigenvalue transport problem reduces to the matrix eigenvalue equation $\vec{S} = \frac{1}{k} F \vec{S}$, and that, for a sufficiently fine mesh, the adjoint fission matrix is simply the transpose of the direct one, $F^\dagger = F^T$ (Eq. (2.15)). This last property is what makes higher-mode analysis practical: instead of tallying F and $F^\dagger$ separately, only the direct matrix F needs to be accumulated during the Monte Carlo run, and its full eigenvalue decomposition then yields *both* sets of modes at once:

$$F\,\vec{S}_n = k_n\,\vec{S}_n, \qquad F^T\,\vec{S}_n^{\dagger} = k_n\,\vec{S}_n^{\dagger}, \qquad n = 0, 1, \ldots, N-1, \tag{2.16}$$

ordered by decreasing magnitude of k_n, with $n = 0$ the fundamental mode. The forward eigenvectors $\vec{S}_n$ are the right eigenvectors of F and represent the nth spatial mode of the fission source. The adjoint eigenvectors $\vec{S}_n^{\dagger}$ are the corresponding left eigenvectors and represent the importance $\psi = \phi^{\dagger}$ of a fission neutron born in each region toward exciting that same mode. Because F is in general non-symmetric, forward modes of different order are not mutually orthogonal, but the forward and adjoint sets are always biorthogonal, $(k_p - k_q)\,\vec{S}_q^{\dagger} \cdot \vec{S}_p = 0$, which is what makes them a useful basis for expansions of an arbitrary source distribution.

For matrices of full 3D cores, the eigenvalue problem is not solved by direct diagonalization but by an iterative sparse eigensolver. The implicitly restarted Arnoldi method (IRAM) offers the best combination of speed and low memory footprint, since it requires only repeated matrix–vector products and never needs the full matrix to be held explicitly [8]. The eigenvalue spectrum obtained this way is only meaningful once two convergence conditions are met: the spatial mesh must be fine enough that the discretization implicit in Eqs. (2.11) and (2.15) is accurate (checked by refining the mesh until the low-order eigenvalues stop changing), and enough particle histories must be simulated that the matrix elements, and hence the eigenvalues, lose their imaginary component, since a non-symmetric matrix estimated with statistical noise can otherwise return complex-conjugate eigenvalue pairs even though the underlying physical spectrum is expected to be real for reactor-type problems.

2.3.2 Higher-Mode Fluxes and Reaction Rates

A right eigenvector $\vec{S}_n$ of the fission matrix is a spatial source distribution, not a flux. To obtain the associated angular flux ϕ (and any reaction rate derived from it), a fixed-source Monte Carlo calculation is run using $\vec{S}_n$, normalized by the fission spectrum, as the starting distribution, with fission treated as a pure loss (capture) so that the calculation does not regenerate the fundamental mode. Since higher-mode sources take on both positive and negative values, this requires tagging each starting particle with the sign of its local source weight and tracing positive and negative contributions separately through the tallies. The resulting higher-mode fluxes tend to reflect the fission source shape at fast energies, while at thermal energies they develop peaks in the moderator, in accordance with the physics of neutron thermalization. Relative uncertainties in these fixed-source solutions are generally well controlled where the mode itself is significantly different from zero, degrading mainly near its zero-contour lines where score cancellation is largest [2].

2.3.3 From a Single Matrix to a Reactor Model: The RAPID Approach

It is worth citing the RAPID (Real-time Analysis for Particle transport and In-situ Detection) methodology of Walters et al. [1], due to its importance in the development of the fission matrix method. The method answers a practical problem: for a large, pin-resolved, three-dimensional power reactor even storing this matrix directly is impractical: a full-size

PWR core can involve tens of trillions of matrix elements. The RAPID method addresses this by building an approximate full-core fission matrix out of a comparatively small library of precalculated coefficients, based on three simplifying assumptions.

1. **Geometric similarity**. Away from the reactor boundary, the coefficient coupling two regions depends only on their relative position, not on their absolute location.

2. **Boundary correction**, rather than tallying separate coefficients for every boundary configuration.

3. **Material discontinuity**: due to the different assembly or pin types used in real cores, it combines coefficient libraries computed independently for each material type into a single, mixed-core matrix. To keep the linear system a manageable size, RAPID additionally performs an **axial-slice homogenization** using a weighting fission density obtained by first solving the uncollapsed two-dimensional problem for each slice.

With these approximations, Walters et al. [1] show that RAPID reproduces reference Serpent eigenvalues to within about 10 pcm and pinwise fission-source distributions to within root-mean-square errors of about 1%, and over 2500 times faster than the reference Monte Carlo calculation, not counting the (one-time, reusable) precalculation of the coefficient database.

However this work does not use the RAPID code since here the fission matrices are generated directly by high-fidelity criticality runs, thus simulating the full core.

2.4 FIM4NTRA Code

The object of this work is not to verify the fidelity of the Fission Matrix method against Monte Carlo simulations *per se*, but to test the accuracy of alternative interpolation techniques for the FM method, with a given database of fission matrices. Ideally, the database would be made up of fission matrices from Monte Carlo runs, but due to their heavy computational cost and since this is not relevant to the scope of this study, the database is generated via the FIM4NTRA code.
FIM4NTRA (FIssion Matrix method for NeuTRonics Analysis) is a Python code developed by C. Vita [4] *ad hoc* for the FM method, with an object-oriented architecture. The code, applied to the VTB Heat-Pipe Micro-Reactor (HPMR), showed, for example, to be able to find the criticality configuration of the control drums rotation angle in 4.3 s, yielding a fission matrix with a 4 pcm reactivity difference with respect to its Serpent-generated counterpart. The fission matrices in the training and testing databases used in this study are generated with the FIM4NTRA code, which relies on a further reduced Monte Carlo database.

2.5 Summary

Taken together, the theoretical derivation of Sec. 2.2 and the applications reviewed in Sec. 2.3 explain the formalization of the fission matrix problem starting from the transport equation, and show how to construct a practical tool out of it. The fission matrix, tallied

at essentially no extra cost during an ordinary Monte Carlo criticality calculation, gives direct access to the higher eigenmodes and, through a fixed-source calculation on each such mode, to the corresponding higher-mode fluxes and reaction rates [2]. The fission matrices can be computed via any high-fidelity method, like the FIM4NTRA code, and then used as databases for various interpolation schemes to swiftly solve other system states of interest.

Chapter 3

Machine Learning Theoretical Background

3.1 Motivation

The fission matrix (FM) method, as explained in Chapter 2, solves the neutron transport eigenvalue problem as a linear map between spatial fission-source distributions, with the fundamental eigenpair of the matrix reproducing the system multiplication factor and the fission source shape [2]. The generation of a fission matrix database over a continuous range of operating conditions (here the VTB HPMR core temperature T and the rotation angle of the control drums θ) with dedicated Monte Carlo fixed-source simulations would be computationally prohibitive [1, 2]. The present work adopts the FIM4NTRA code, in place of the actual Monte Carlo runs, for the generation of a database of a limited number of high-fidelity simulations and uses it to train a data-driven surrogate model. The surrogate model evaluates, at negligible cost, the point in the parameter space with the highest uncertainty, and generates the corresponding fission matrix, adding it to the database. This process is repeated for a given number of iterations, while monitoring the trend of the uncertainty. The two techniques for the surrogate model are the *Proper Orthogonal Decomposition* (*POD*) and the *Kriging* regression. The first one is used reduce the dimensionality of the fission matrix, which would be otherwise an object too large to handle, while the second one is a regression model, trained on the reduced coordinates, with a useful built-in predictor of its own uncertainty. This last property is what makes an adaptive learning (AL) strategy possible: the surrogate does not simply predict the fission matrix at a new parameter point, it also reports how much it does not trust that prediction, and this information is used to decide where FIM4NTRA should be run next.

3.2 Proper Orthogonal Decomposition

The Proper Orthogonal Decomposition (POD) is a linear dimensionality-reduction method originally introduced in the fluid-dynamics literature to analyse numerically the turbulent velocity fields [9]. In the field of statistics the same procedure is under the name of Principal Component Analysis (PCA). Independently of the application field, the underlying idea is

the same: given a finite ensemble of high-dimensional observations, the POD constructs an orthonormal basis, ranked by the fraction of variance that each basis vector reproduces, such that a small number of basis vectors is sufficient, depending on the application, to reconstruct the ensemble to a prescribed accuracy.

In the present work, the POD is applied to the fission matrix, a high-dimensional operator that represents the average number of neutrons born in a spatial region of the core due to an average neutron born by fission in another spatial region, for a given reactor configuration. The parameter space explored is two-dimensional, composed of core temperature $T \in [300, 1000]\,\mathrm{K}$ and the control drums rotation angle $\theta \in [0°, 180°]$. Temperature and control drums angle were selected as parameters since they represent the two dominant reactivity feedback and control mechanisms in the core: T activates the Doppler broadening of resonance cross-sections (the intrinsic, passive feedback), while θ regulates the reflector/absorber worth of the control drums (the active control). The scope of the POD in this case is to compress the database of parametrised fission matrices into a set of orthogonal basis described by modes that effectively capture most of the information of the matrix without needing to model every single entry. These parameter-dependent coefficients can then be regressed with the Kriging surrogate, as detailed in Section 3.3.

3.2.1 Construction of the snapshot matrix

Let $\mathbf{y}(T, \theta) \in \mathbb{R}^n$ denote the fission matrix generated with FIM4NTRA for a given parameter pair (T, θ), flattened into a column vector of length n (the number of FM entries). Given a training set of m high-fidelity evaluations $\{(T_i, \theta_i)\}_{i=1}^m$, the corresponding flattened fission matrices are stacked as rows of a *snapshot matrix*:

$$\mathbf{S} = \begin{pmatrix} \mathbf{y}(T_1, \theta_1)^\top \\ \mathbf{y}(T_2, \theta_2)^\top \\ \vdots \\ \mathbf{y}(T_m, \theta_m)^\top \end{pmatrix} \in \mathbb{R}^{m \times n}. \tag{3.1}$$

As is standard practice in POD, the decomposition is performed on the fluctuating part of the field. The mean FM field $\bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{y}(T_i, \theta_i)$ is subtracted from every row of $\mathbf{S}$ to obtain the centred snapshot matrix:

$$\mathbf{U} = \mathbf{S} - \mathbf{1}_m \, \bar{\mathbf{y}}^\top, \tag{3.2}$$

being $\mathbf{1}_m \in \mathbb{R}^m$ a column vector of ones. In the present application the number of FM entries n is several orders of magnitude larger than the number of training points m ($n \gg m$), which has direct consequences for the choice of algorithm discussed in Section 3.2.3.

3.2.2 POD as an eigenproblem

Formally, POD solves for an orthonormal basis $\{\mathbf{\Phi}_r\}_{r=1}^n$ of $\mathbb{R}^n$, ranked by the fraction of variance explained, obtained from the eigendecomposition of the $n \times n$ covariance matrix:

$$\mathbf{C} = \frac{1}{m-1}\,\mathbf{U}^{\top}\mathbf{U} = \mathbf{\Phi}\,\mathbf{\Lambda}\,\mathbf{\Phi}^{\top}, \tag{3.3}$$

where the columns of $\mathbf{\Phi}$ are the *POD modes* and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$, are the associated eigenvalues. This direct formulation, however, requires constructing and diagonalising an $n \times n$ matrix, which is impractical in this application due to the dimensions of the matrix. Section 3.2.3 introduces the reformulation actually used in the implementation.

3.2.3 The Snapshot Method and its SVD formulation

The *Snapshot Method*, introduced by Sirovich [10], is based on the fact that the roles of the spatial n and sample m dimensions in Eq. (3.1) can be interchanged without loss of generality [9]. Instead of the $n \times n$ covariance matrix $\mathbf{C}$, a much smaller $m \times m$ correlation matrix can be formed:

$$\mathbf{C}_s = \frac{1}{m-1}\,\mathbf{U}\,\mathbf{U}^{\top} = \mathbf{A}\,\mathbf{\Lambda}\,\mathbf{A}^{\top}, \tag{3.4}$$

whose eigenvalues coincide exactly with the non-zero eigenvalues of matrix $\mathbf{C}$ in Eq. (3.3). The modes are then extracted by projecting the centred data onto the eigenvectors of $\mathbf{C}_s$, $\mathbf{\Phi} = \mathbf{U}^{\top}\mathbf{A}$, followed by column-wise normalisation to unit norm. Because $\mathbf{C}_s$ is only $m \times m$, this reformulation reduces the eigendecomposition cost from $\mathcal{O}(n^3)$ to $\mathcal{O}(m^3)$ and returns at most $\min(m, n) = m$ non-trivial modes, being the maximum number resolvable from m training snapshots in any case.

In practice, both the direct and the snapshot formulations are particular views of the Singular Value Decomposition (SVD) of the centred snapshot matrix.

$$\frac{\mathbf{U}}{\sqrt{m-1}} = \mathbf{L}\,\mathbf{\Sigma}\,\mathbf{R}^{\top}, \tag{3.5}$$

with $\mathbf{L} \in \mathbb{R}^{m\times m}$ and $\mathbf{R} \in \mathbb{R}^{n\times n}$ orthogonal, and $\mathbf{\Sigma}$ a rectangular diagonal matrix of singular values $\sigma_1 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$. Comparing Eq. (3.5) with Eqs. (3.3) and (3.4) shows that $\mathbf{\Phi} = \mathbf{R}$, $\mathbf{A} = \mathbf{L}$, and $\lambda_r = \sigma_r^2$ for every mode r. This equivalence is the formulation actually implemented in the model, since a direct call to an SVD routine on $\mathbf{U}/\sqrt{m-1}$ is both numerically more stable and more concise than explicitly forming and diagonalising $\mathbf{C}_s$.

Once $\mathbf{\Phi}$ and $\mathbf{A}$ are obtained, the centred data can be expressed exactly as a sum of m contributions

$$\mathbf{U} = \mathbf{A}\,\mathbf{\Phi}^{\top} = \sum_{r=1}^{m} \mathbf{a}_r\,\mathbf{\Phi}_r^{\top}, \tag{3.6}$$

where $\mathbf{a}_r \in \mathbb{R}^m$, the r-th column of $\mathbf{A}$, is the coefficient vector of mode r: the projection of every training fission matrix onto $\mathbf{\Phi}_r$. In this work, $\mathbf{a}_r$ represents a scalar field over the (T, θ) parameter space, and it is this field, evaluated at every point as $a_r(T, \theta)$, that the Kriging surrogate is trained to reproduce (Section 3.3).

3.2.4 Mode Truncation and Dimensionality Reduction

The great contribution of POD is in dimensionality reduction: in the classical approach the field is approximated by retaining only the $r^* \leq m$ modes associated with the largest singular values:

$$\mathbf{y}(T,\theta) \approx \bar{\mathbf{y}} + \sum_{r=1}^{r^*} a_r(T,\theta)\, \mathbf{\Phi}_r. \tag{3.7}$$

The classical criterion for selecting the truncation order r^* is energetic: modes are added until a prescribed fraction of the total variance is captured, being $\lambda_r / \sum_j \lambda_j$ is the fraction of variance explained by mode r [9]. When the objective of the POD is a qualitative or energetic reconstruction of a field, as in the fluid-dynamics literature, this criterion is the standard practice.

However in the present application the quantity of interest is not the variance explained by the reconstructed fission matrix itself, but the propagated error on k_{eff}. Section 4.3.1 shows empirically that, for the fission matrix problem, these two objectives are actually not equivalent: truncations that explain up to 99.7% of the total variance can still leave a significantly large k_{eff} error. This phenomenon happens because the residual, low-energy modes are not uniformly irrelevant to the physics described by the fission matrix. For this reason, Section 4.3.1 assesses the need to use the full column rank of the training data, $r^* = m - 2$.

3.3 Kriging: Gaussian Process Regression

3.3.1 Definition and probabilistic setting

Each POD coefficient $a_r(\mathbf{x})$ is modelled independently as a Kriging surrogate. Kriging originates in geostatistics, where it was introduced as a method for spatial interpolation of ore-grade estimates [11] and later formalised mathematically [12]. The same formulation is now an established method in the machine-learning literature under the name Gaussian Process Regression [13], and it is a widely used technique for expensive black-box simulators [14, 15].

The Gaussian Process Regression (GPR) models an unknown function that describes a *Gaussian process*, that is a collection of random variables where any finite subset of them follows a joint multivariate Gaussian distribution [13]. Formally, a scalar random field $Z(\mathbf{x})$, $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$, is a Gaussian process if, for any finite collection of input points $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the vector $(Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))$ is jointly Gaussian. The process is characterized by a mean function $m(\mathbf{x}) = \mathbb{E}[Z(\mathbf{x})]$ and a covariance (kernel) function $k(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}(Z(\mathbf{x}), Z(\mathbf{x}'))$.

By prescribing a full probability law, GPR does not limit itself to the prediction of the value of a point, but it produces a complete predictive distribution giving a built-in uncertainty measure of its own estimation [5].

3.3.2 Joint distribution and predictive equations

Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ denote the training inputs and $\mathbf{y} = (y_1, \dots, y_n)^\top$ the corresponding (possibly noisy) observations

$$y_i = Z(\mathbf{x}_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n, \tag{3.8}$$

where σ^2 represents an additive noise term (in the geostatistical literature referred to as the *nugget*), $\mathcal{N}$ is a Gaussian distribution and i.i.d. stands for "independent and identically distributed". Under the Gaussian process assumption, the training observations follow

$$\mathbf{y} \sim \mathcal{N}\big(m(\mathbf{X}),\ k(\mathbf{X}, \mathbf{X}) + \sigma^2 I\big), \tag{3.9}$$

where $m(\mathbf{X}) = [m(\mathbf{x}_1), \dots, m(\mathbf{x}_n)]^\top$ is the mean function evaluated at the training inputs, and $k(\mathbf{X}, \mathbf{X})$ is the $n \times n$ covariance matrix obtained by evaluating the kernel on every pair of training points. The kernel defines how similar two points in the parameter space are, and hence how much the information observed at one has to influence the prediction at the other.

For a new input $\mathbf{x}_*$, Gaussianity implies that the joint distribution of the training points and the unknown value $Z_* = Z(\mathbf{x}_*)$ are themselves Gaussian:

$$\begin{pmatrix} \mathbf{y} \\ Z_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} m(\mathbf{X}) \\ m(\mathbf{x}_*) \end{pmatrix}, \begin{pmatrix} k(\mathbf{X}, \mathbf{X}) + \sigma^2 I & \mathbf{k}_* \\ \mathbf{k}_*^\top & k(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix} \right), \tag{3.10}$$

with $\mathbf{k}_* = k(\mathbf{X}, \mathbf{x}_*)$ being the vector of cross-covariances between the training points and $\mathbf{x}_*$.

The predictive distribution is then obtained by conditioning this joint Gaussian on the observed data, using the standard closed-form result for conditional distributions of multivariate Gaussians [13]:

$$Z_* \mid \mathbf{y} \ \sim\ \mathcal{N}\big(\mu_*,\ \sigma_*^2\big), \tag{3.11}$$

$$\mu_* = m(\mathbf{x}_*) + \mathbf{k}_*^\top \big(k(\mathbf{X}, \mathbf{X}) + \sigma^2 I\big)^{-1} \big(\mathbf{y} - m(\mathbf{X})\big), \tag{3.12}$$

$$\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top \big(k(\mathbf{X}, \mathbf{X}) + \sigma^2 I\big)^{-1} \mathbf{k}_*. \tag{3.13}$$

Equations (3.12) and (3.13) constitute the core of GPR: the predictive mean μ_* is a weighted linear combination of the deviations of the training points from the prior mean, while the predictive variance σ_*^2 depends only on the input locations (through the kernel), and not on the observed values $\mathbf{y}$. This variance is the quantity used by the active learning cycle adopted in this work (Section 4.5): it provides an intrinsic measure to the model of where the surrogate is least confident, and then it is used to guide the selection of the new points to generate. Section 3.3.4 extends this single-mode quantity to an aggregate measure of uncertainty at the level of the reconstructed fission matrix, which is the quantity actually used to drive point selection.

If the point estimate $\hat{Z}_* = \mu_*$ is taken as predictor, it corresponds to the maximum a

posteriori (MAP) value of the predictive distribution in Eq. (3.11); this is the value returned by the predictor in the implementation adopted in this work.

3.3.3 Kernel choice and hyperparameter estimation

The predictive equations above assume the kernel $k(\cdot,\cdot)$ is known. In practice, the kernel is chosen from a parametric family and its hyperparameters $\boldsymbol{\psi}$ are tuned for the specific application. Two choices are involved: the *family* of the kernel, which regulates the smoothness assumed for the guessed function, and the *hyperparameters* within that family (an output-scale and a set of length scales), which are tuned automatically by maximizing the log marginal likelihood of the training data.

For stationary, anisotropic kernels of the kind used in this work, the covariance between two inputs depends only on their scaled distance

$$r_d = \frac{x_d - x'_d}{\ell_d}, \qquad d = 1, \dots, D, \tag{3.14}$$

where $D = 2$ is the input dimension (temperature T and rotation angle θ) and ℓ_d is an independent length scale for each dimension. This anisotropic parametrisation, usually named *automatic relevance determination* (ARD), allows the model to capture the fact that the fission matrix response may vary over different characteristic scales along T and θ, instead of assuming the same scale of variation.

Two kernel families are considered as candidates in this work:

- **Squared-exponential RBF (Radial Basis Function).** Among the many different kinds of RBF functions, the squared-exponential one is the most common choice in GPR,

$$k_{\mathrm{SE}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\psi}) = \sigma_f^2 \exp\left(-\tfrac{1}{2}\sum_{d=1}^{D} r_d^2\right), \tag{3.15}$$

  which assumes that the function is infinitely mean-square differentiable (i.e. maximally smooth). It will be the default option in this work when no other informations suggest differently.

- **Matérn class.** A more flexible family that relaxes the assumption of infinite smoothness of the RBF,

$$k_\nu(\mathbf{x}, \mathbf{x}'; \boldsymbol{\psi}) = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\, r\right)^{\nu} K_\nu\!\left(\sqrt{2\nu}\, r\right), \qquad r = \left(\sum_{d=1}^{D} r_d^2\right)^{1/2}, \tag{3.16}$$

  where K_ν is the modified Bessel function of the second kind and $\nu > 0$ is a smoothness parameter. As ν tends to infinity the Matérn kernel recovers the squared-exponential, and returns functions with only $\lceil \nu \rceil$ mean-square derivatives for finite ν. Common choices are $\nu = 3/2$ and $\nu = 5/2$, corresponding to once and twice mean-square differentiable functions, respectively.

For both families, the hyperparameter vector $\boldsymbol{\psi}$ collects the output-scale σ_f^2 and the D length scales ℓ_d. The output-scale σ_f^2 sets the vertical scale of the guessed function, while each ℓ_d determines the speed of decay of the function correlation along dimension d: a small ℓ_d allows the model to follow rapid local variation but restricts extrapolation to nearby training points, while a large ℓ_d gives a smoother prediction more globally informed. GPR estimates the hyperparameter vector $\boldsymbol{\psi}$ by directly maximizing the log marginal likelihood of the observed data [5, 13]:

$$\log p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\psi}) = -\frac{1}{2}\mathbf{y}^\top K_\psi^{-1}\mathbf{y} - \frac{1}{2}\log|K_\psi| - \frac{n}{2}\log 2\pi, \tag{3.17}$$

where $K_\psi = k(\mathbf{X}, \mathbf{X}; \boldsymbol{\psi}) + \sigma^2 I$. This practice balances model fit (first term) against model complexity (second term), giving an automatic form of regularization that avoids both under- and over-fitting. This process does not require a separate validation step, simplifying the implementation of the model. Section 4.3.2 presents the considerations made to choose both the kernel family and the value of the nugget term σ^2.

3.3.4 Application to the POD-reduced fission matrix surrogate

In this work, GPR is not applied directly to the full order fission matrix $\mathbf{Y} \in \mathbb{R}^{n\times n}$ generated by FIM4NTRA, but to its low-dimensional representation obtained through Proper Orthogonal Decomposition (POD). In this specific application, the fission matrix has a cardinality $n = 90$. Each training snapshot Y_i is flattened and projected onto the r^* POD modes, returning a coefficient vector $\mathbf{a}_i \in \mathbb{R}^{r^*}$. A number of r^* independent GPR models is then trained to map the two-dimensional input $\mathbf{x} = (T, \theta)$ to each POD coefficient

$$a_r(\mathbf{x}) \sim \mathcal{GP}\big(0,\ k_r(\mathbf{x}, \mathbf{x}'; \boldsymbol{\psi}_r)\big), \qquad r = 1, \ldots, r^*, \tag{3.18}$$

each governed by its own kernel hyperparameters $\boldsymbol{\psi}_r$, estimated independently via Eq. (3.17). The prediction for a point $\mathbf{x}_*$ not included in the training set is obtained by evaluating the predictive mean of Eq. (3.12) for every mode, $\hat{\mathbf{a}}_* = (\mu_*^{(1)}, \ldots, \mu_*^{(n_c)})$, and reconstructing the full-order fission matrix through the inverse POD mapping, $\hat{Y}_* = \boldsymbol{\Phi}\hat{\mathbf{a}}_*$, where $\boldsymbol{\Phi}$ denotes the POD basis. It is worth noting that this reconstruction is only as good as the POD basis itself: $\boldsymbol{\Phi}$ is estimated from a finite, and in this context comparatively small, set of training snapshots, and is therefore implicitly assumed to span the relevant solution subspace. Any component of the true fission matrix orthogonal to this subspace cannot be recovered by the surrogate, regardless of how accurately the Kriging model predicts the retained coefficients $\hat{\mathbf{a}}_*$. The accuracy of $\hat{Y}_*$ is thus bounded above by the completeness of $\boldsymbol{\Phi}$, not by the interpolation error alone. Moreover, a non-negativity constraint, $\hat{Y}_* = \max(\hat{Y}_*, 0)$, is then enforced to preserve the physical consistency of the reconstructed fission matrix, since its entries represent the non-negative numbers of neutrons born in a region due to neutrons born by fission in another region 2.1.

Since the surrogate model is made of r^* independent GP models rather than a single one, the posterior variance $\sigma^2_{*,r}(\mathbf{x})$ from Eq. (3.13), describing a single mode is not by itself a valid measure of the uncertainty on the reconstructed fission matrix as a whole. The actual aggregate uncertainty measure used by the adaptive learning algorithm to evaluate

its confidence over the parameter space is instead defined by combining the posterior standard deviations of all retained modes.

$$\sigma(\mathbf{x}) = \left(\sum_{r=1}^{r^*} \sigma_{*,r}^2(\mathbf{x}) \right)^{1/2}, \tag{3.19}$$

This is therefore the quantity evaluated by the point-selection criterion of Section 4.5 and reported throughout Chapter 5. Equation (3.19) reduces to a single term of Eq. (3.13) in the limit case $r^* = 1$, and is otherwise the Euclidean norm of the "per-mode" standard deviations, so that the uncertainty of the fission matrix reconstruction is influenced by any of the retained modes rather than requiring all modes to be simultaneously uncertain.

It is worth making two practical remarks on the specific implementation to this work:

- **Input normalization.** Since T and θ ranges are very different in size ($[300, 1000]$ K versus $[0°, 180°]$), the input space is rescaled to $[0, 1]^2$ before the fitting stage. This ensures that the length-scale hyperparameters ℓ_T, ℓ_θ are directly comparable and that their optimization is coherent across both dimensions.

- **Nugget regularization.** A small noise term σ^2 is added to the diagonal of the covariance matrix. This term has a numerical purpose in addition to being just an observational noise: it guarantees that K_ψ remains invertible even when the training set contains duplicated or near-duplicated input locations. As a consequence, $\sigma(\mathbf{x})$ does not vanish exactly even at already-measured input locations, but settles to a small residual floor (Section 4.3.3).

Instead of the aggregate posterior standard deviation of Eq. (3.19), built from the posterior variance of Eq. (3.13) of each single-mode, Section 6.2 discusses also the possibility of using as predictive uncertainty of the fission matrix the disagreement between an ensemble of independently trained POD+GPR surrogates.

Chapter 4

Methodology and computational framework

4.1 Reactor and geometry

The reactor used in this study is the VTB (Virtual Test Bed) Heat Pipe Micro Reactor (HPMR), a reactor designed by ANL to analyse some modelling challenges within the industry, like the heat pipes for heat removal, the adoption of TRi-structural ISOtropic (TRISO) fuel to manage higher temperatures and the use of rotating control drums in the radial deflectors [16]. The reactor layout is shown in Figure 4.1 and the core parameters are listed in Table 4.1.

Parameter	**Value**
Rated thermal power	2 MW
Fuel type	TRISO
Enrichment	19.95 at% LEU
Fuel form	UCO
Matrix	Hexagonal graphite
Packing fraction	40%
Core length	160 cm
Axial reflectors (upper/lower)	20 cm, beryllium metal
Number of fuel assemblies	30
Radial reflector	1 ring of beryllium
Number of control drums	12

Table 4.1: Main parameters of the HPMR core (VTB).

Additionally to the graphite structure, yttrium-hydride (YH2) pins are used thanks to their excellent moderating capabilities. The clad is structured as follows: an outer stainless steel envelope, a helium gap, the YH2, another helium gap, and then graphite monolith. The heat pipes also have a stainless-steel envelope, cooled by potassium as working fluid. A helium gap separates the heat pipe form the graphite monolith.

The reactor's control system consists of 12 control drums positioned within the radial reflector, capable of achieving cold shutdown at any point during reactor operation. A central shutdown rod is additionally provided for redundancy.

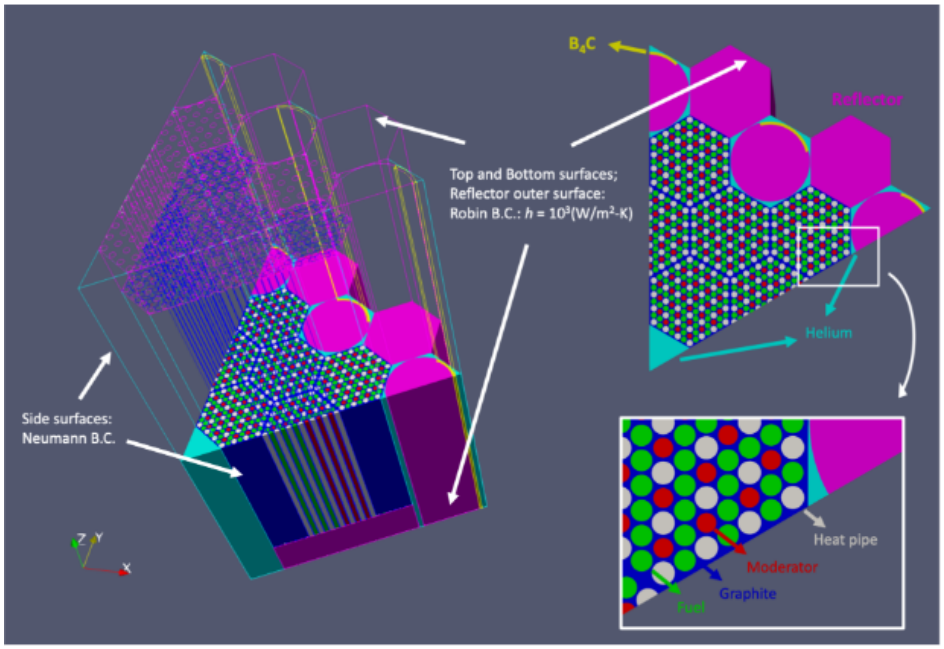


Figure 4.1: Layout of the core for the VTB HPMR model. Only one sixth of the core is shown. Source: [16].

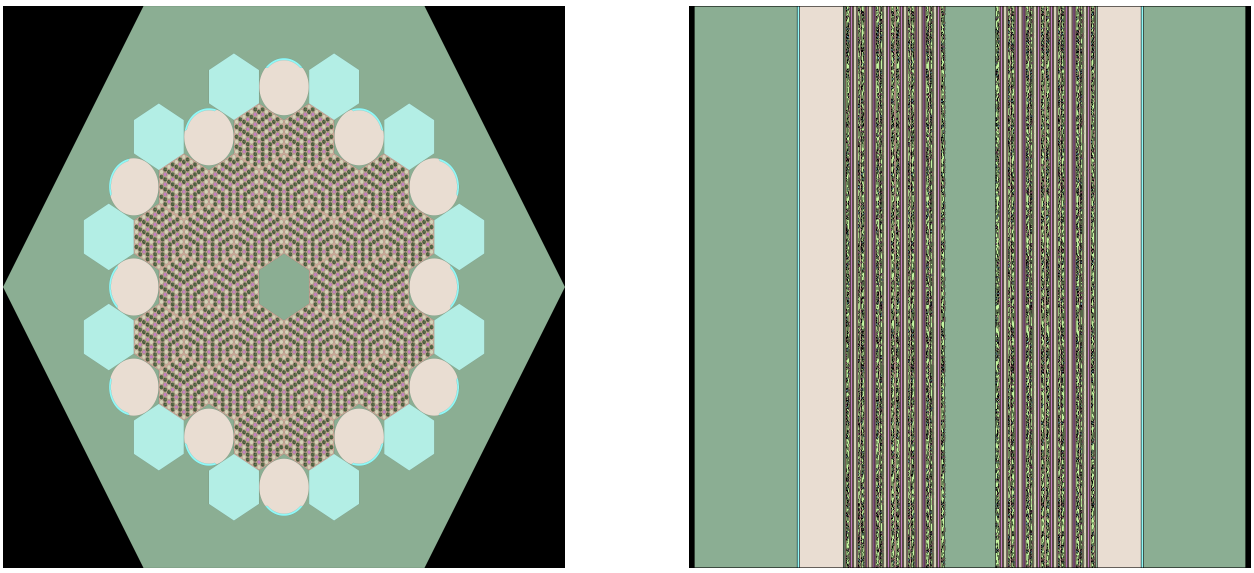


Figure 4.2: Serpent picture of VTB HPMR core, radial (left) and axial(right) cross-section.

In Figure 4.2 a top and side view of the core is shown, with the control drums fully rotated outwards.

4.2 Parameter space definition

The number of variables featuring the parameter spce is set to be 2 because, remaining within FIM4NTRA's capabilities, a higher order space would have needlessly increased the computational burden with respect to the objective of the study, as well as complicated the mapping of the parameter space. The variables adopted to independently qualify the different operational states of the reactor are temperature T, in the range 300 K to 1000 K, and rotation angle of the control drums θ in the range 0° to 180°. The surrogate model is hence trained to reproduce the fission matrix as a function of this two-dimensional parameter space

$$\Omega = [T_{min}, T_{max}] \times [\theta_{min}, \theta_{max}] \subset \mathbb{R}^2. \tag{4.1}$$

These two parameters were selected because they represent the dominant sources of variability in the neutron spectrum and core geometry addressed in this work.

The **initial dataset** for the training dataset X_0 is generated by uniform grid sampling over Ω, obtained via a Cartesian product of 4 linearly spaced points along each dimension, with a total of 16, as shown in Figure 4.3.

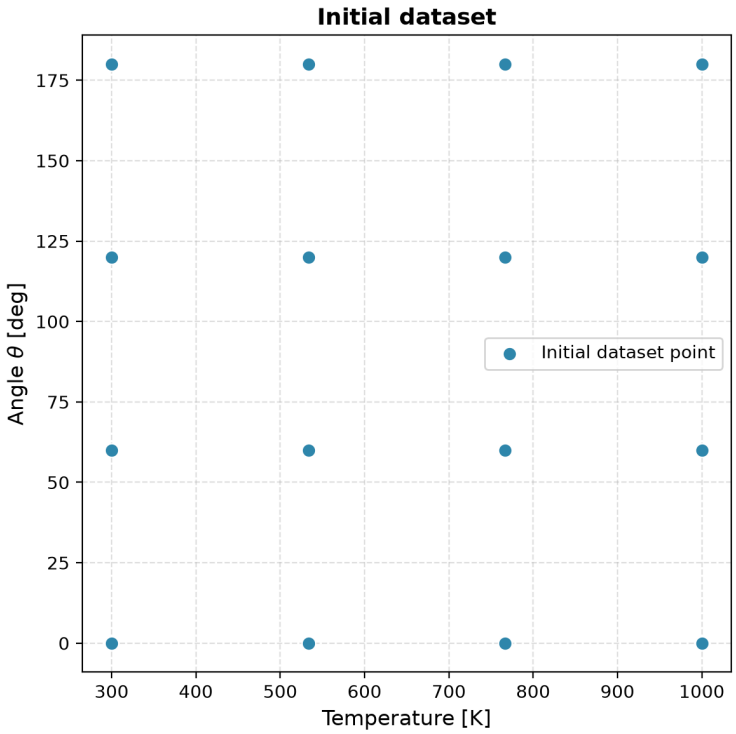


Figure 4.3: Initial dataset X_0 with 16 points.

This choice guaranties a regular, reproducible coverage of the domain at initialization, which is a necessary condition for the following adaptive sampling strategy to correctly identify more uncertain regions without mistaking grid irregularity for actual model uncertainty. Candidate points $\{x_i\}_{i=1}^{N_{\text{CANDIDATES}}}$ proposed at each active learning iteration are drawn from

the same normalized space and filtered against the current training set X to exclude points already sampled, within a numerical tolerance ϵ.

4.3 Surrogate Calibration

The first step of the work, before optimizing the adaptive sampling strategy of Section 4.5, is to make three design choices regarding the surrogate itself: the POD truncation order, the Kriging kernel configuration, and the source of predictive uncertainty. In this section is reported the operative analysis that determined each of these choices and that motivates their use throughout the parametric study of Chapter 5.

4.3.1 Mode truncation

As stated in Section 3.2.4, the classical criterion for selecting the POD truncation order r^* in Eq. (3.7) is energetic: modes are retained until a prescribed fraction of the total variance is captured, or "explained". To test whether this criterion is valid also for the present application, a reference training set of $N = 56$ points was generated (the 16 initial points of Section 4.2 plus 40 points drawn at random from the candidate grid), and the POD spectrum was computed for increasing subsets of size $n \in \{16, 26, 36, 56\}$ for each truncation order $r \in \{1, 2, 3, 5, 8, 10, 15, 20, 30, 40\}$. Then, 3 observations were made.

1. For every subset n Figure 4.4 (a) shows the variance ratio explained by every mode r singularly. The first mode carries the largest piece of information about the variance, while the following modes, smaller, carry less information. These eigenvalues decay rapidly and similarly for each n: the spectrum does not change shape significantly as n increases.

2. Figure 4.4 (b) shows that the residual variance, i.e. the complementary to the explained variance ratio, for every n, decreases as more modes are retained.

3. From Figure 4.4 one could conclude that only few modes are necessary to describe the variance. Figure 4.5 verifies this assertion by quantifying the effect of mode truncation directly on the k_{eff}. Two errors are defined: one due to the *POD only* and one due to the *POD + Kriging*. For the first one, the POD base is computed on the training set, and the high-fidelity ("ground truth") FMs from the testing set are projected onto that base, retaining only the first r modes. The *POD only* error is the difference in pcm between the k_{eff} of the true full-order FMs and the k_{eff} of the FMs reconstructed from the truncated modes. The second error takes into account also the effect of the GPR. The Kriging model is trained on the already truncated coefficients of FMs from the training set, and then model predicts the FMs of new (T, θ) cases in the testing set. The *POD + Kriging* error measures the difference between the k_{eff} of the high-fidelity FMs and the k_{eff} of the FMs reconstructed from the predicted (truncated) coefficients. This error is hence always equal or larger than the *POD only* one. Figure 4.5 thus shows that, differently from what Figure 4.4 would suggest, retaining every mode has a non-negligible impact on the k_{eff}.

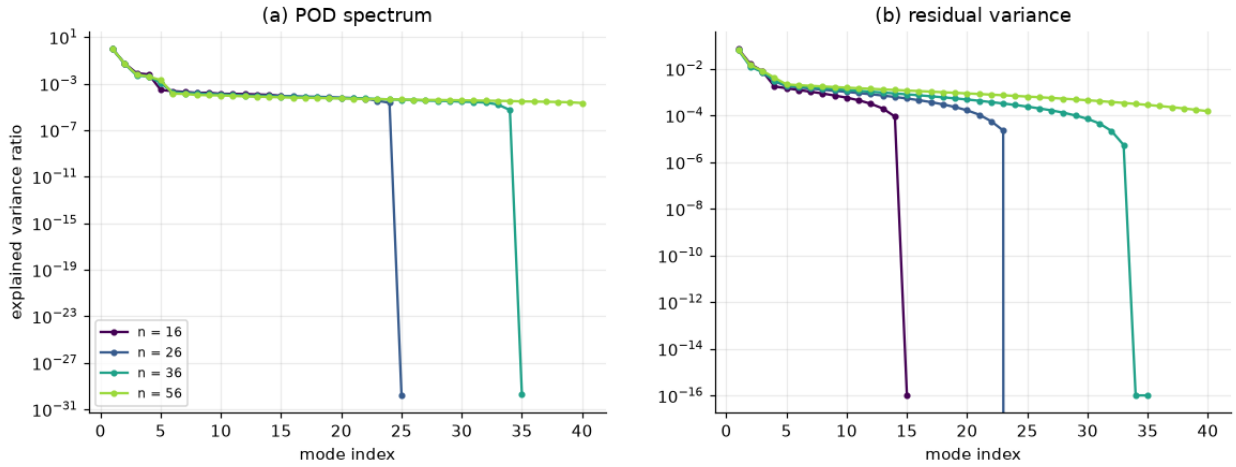


Figure 4.4: POD spectrum.

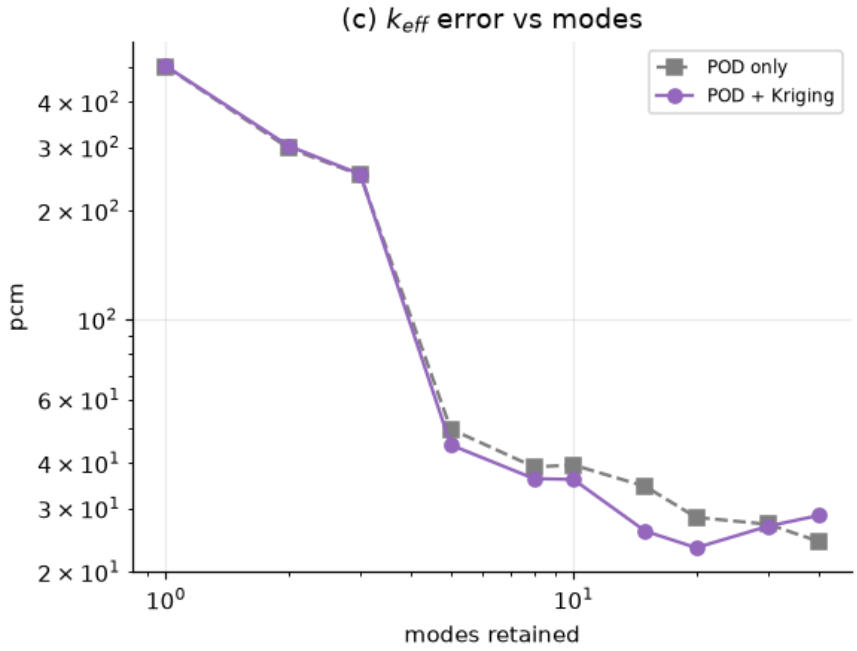


Figure 4.5: *POD only* and *POD + Kriging* errors on the k_{eff}.

The two metrics, in accordance, give a very different result with respect to the energetic criterion. At $r = 5$ modes, 99.79% of the total variance is already captured, yet the resulting k_{eff} error is 45 pcm; increasing the truncation order to $r = 15$ raises the explained variance by only 0.10 percentage points (to 99.89%) while nearly halving the error to 26 pcm. The fraction of variance left unexplained therefore saturates well before the physically relevant error does: the higher-order modes, individually responsible for a negligible share of the total variance, are not correspondingly uninfluential on the neutronic quantity of interest. Consequently, in the final configuration all modes are retained. It should be stressed that, even when all modes are used, the dimensionality reduction is still considerably large, since

$r^* = m - 2$ and $m \ll n$. The computational benefit of the POD reduction is a structural consequence of the problem it is applied to, rather than simply the choice of modes retained.

4.3.2 Kernel and hyperparameter configuration

Having set the truncation order to its maximum admissible value (see Section 4.3.3), six variants of the Kriging kernel configuration of Section 3.3.3, varying also the nugget term σ^2, were compared on the same reference training set $\mathbf{X}_{t1}$ (Table 4.2). The choice between an RBF and a Matérn-5/2 kernel has only a marginal effect on accuracy when left with the nugget at its default order of magnitude ($\sigma^2 = 10^{-6}$), revealing to be statistically indistinguishable on a single training-set realisation (19.7 vs. 19.1 pcm, averaged over three independent realisations of the training set). The results confirm RBF as the marginally better choice, and it is thus retained for consistency with the rest of workflow. Three alternative configurations have instead clearly worse performances:

1. increasing or decreasing the nugget by two orders of magnitude lowers the accuracy by a factor of 4 and 5 (86.9 and 97.3 pcm);

2. the method of rescaling the POD coefficients to unit variance before fitting the GPR models is strongly disadvantageous (3165.6 pcm);

3. training an independent GPR per mode without a shared kernel family also offers no benefit (24.6 vs. 19.7 pcm).

Configuration	k_{eff} error [pcm]
RBF, $\sigma^2 = 10^{-6}$ (adopted)	19.7
Matérn 5/2, $\sigma^2 = 10^{-6}$	19.1
RBF, $\sigma^2 = 10^{-8}$	86.9
RBF, $\sigma^2 = 10^{-4}$	97.3
RBF, standardised POD coefficients	3165.6
RBF, one independent GPR per mode	24.6

Table 4.2: Comparison of Kriging kernel configurations, k_{eff} error on the testing set, full-rank truncation.

4.3.3 Truncation depth within the active learning loop

Section 4.3.1 established that the truncation order matters for accuracy, but left to investigate whether the order used during the active learning loop should be different from the order of the final model used on the testing set. To test this, the adaptive learning strategy, with the same initial database X_0 and repeated iteratively on 40 new points, was carried out for six values of the truncation order inside the loop, $r_{\text{loop}} \in \{10, 15, 23, 30, 40, r^*\}$, and each resulting dataset was then re-evaluated at seven values of the truncation order of the final

model, $r_{\text{final}} \in \{10, 15, 20, 23, 30, 40, k^*\}$. The $r^* = n - 2$ indicates the maximum truncation order admissible by the rank of the training data (Section 4.3.1). The resulting 6×7 matrix of k_{eff} errors (Figure 4.6) shows that the choice of r_{final} has a larger effect than the choice of r_{loop} (6.7 vs. 3.5 pcm of spread across the respective best-case values), and that $r_{\text{final}} = r^*$ is the best choice for every value of r_{loop} tested. Due to these considerations, the truncation order is fixed at its maximum admissible value, $r^* = n - 2$, both throughout the active learning loop and in the final surrogate model used in Chapter 5, rather than being treated as a free hyperparameter to be tuned.

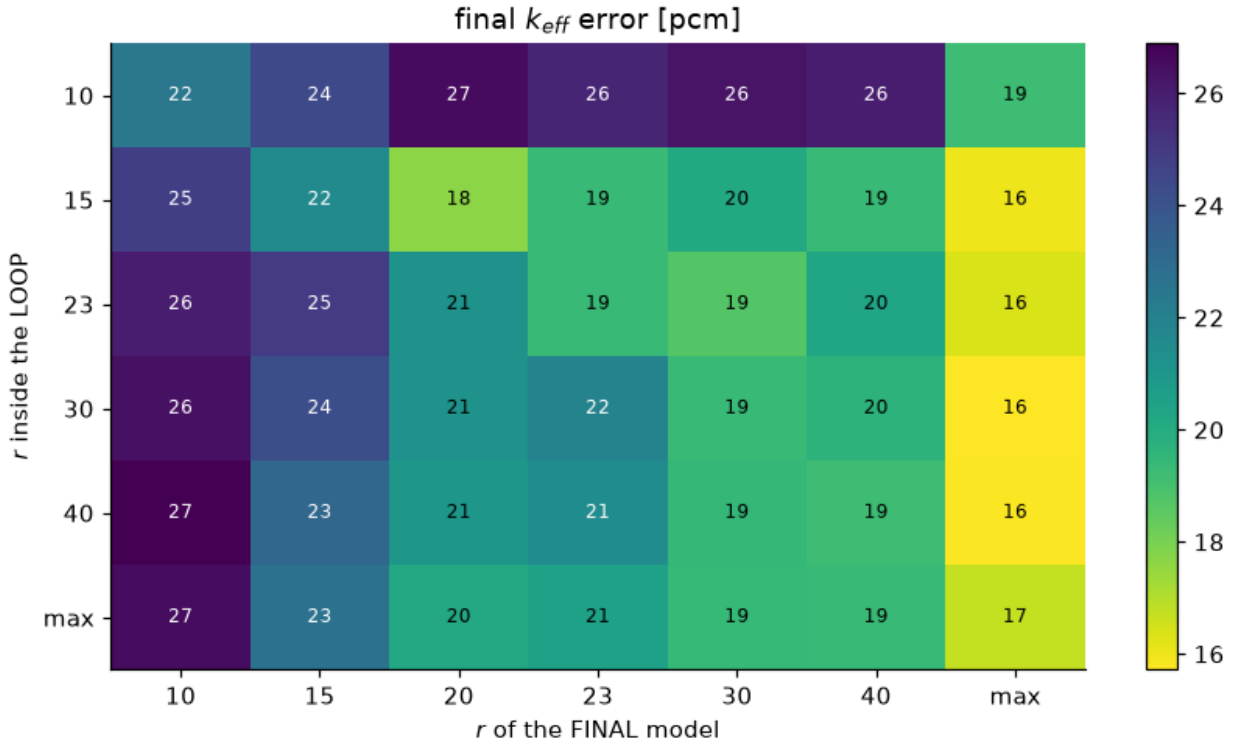


Figure 4.6: Matrix of k_{eff} errors changing the number of modes in the AL loop and in the final configuration.

4.3.4 Exclusion of already measured points

Another independent verification concerns a practical requirement of the point-selection criterion itself: whether candidates that coincide with, or lie arbitrarily close to, already measured points must be explicitly excluded from selection.

To assess this, a purely uncertainty-driven loop was run for 14 iterations starting from the initial dataset $\mathbf{X}_0$. At each iteration the surrogate is refit, the aggregate *posterior* standard deviation $\sigma(\mathbf{x})$ of Eq. (3.19) is evaluated both on the candidate grid and on the already-measured points, and the candidate with maximum σ is added to the training set. Figure 4.7 reports, at every iteration, the maximum posterior standard deviation among the test candidates together with the mean *posterior* standard deviation at the already measured points.

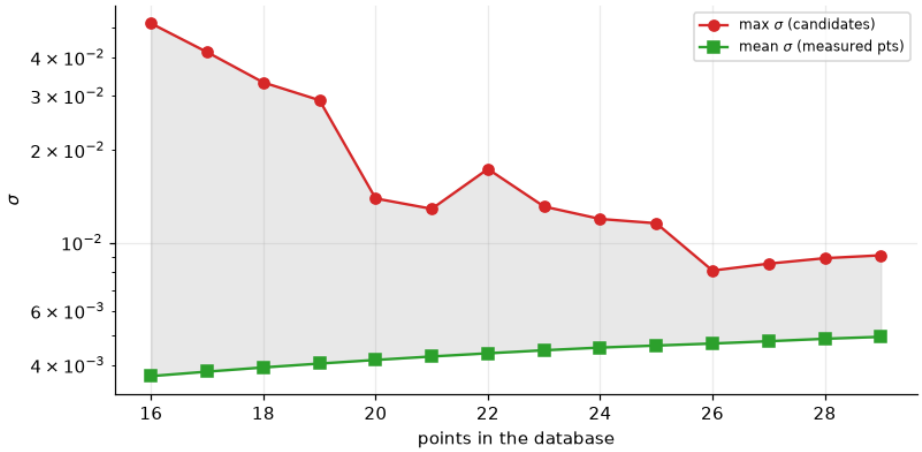


Figure 4.7: Evolution of the posterior standard deviation during a purely uncertainty-driven active learning loop: maximum σ among unmeasured candidates (red) versus mean σ at already-measured points (green).

The mean *posterior* standard deviation at measured points does not decay to zero as the loop progresses: because of the presence of the nugget as regularisation term, σ^2, present in each single-mode *posterior* variance of Eq. (3.13), none of the GPs interpolates exactly even at training locations. This residual propagates through the aggregation of Eq. (3.19), so the combined $\sigma(\mathbf{x})$ also settles to a background noise instead of disappearing at already-measured points. The maximum *posterior* standard deviation among test candidates, by contrast, decreases rapidly as the training set grows. The ratio between the two quantities is a direct measure of how distinguishable a really informative candidate is from the residual floor, and drops from a factor of 14.0× at the first iteration to only 1.8× after 14 iterations. As the domain becomes progressively better sampled, the margin between a truly uninformative, already measured location and a still informative candidate therefore narrows substantially, to the point of becoming barely distinguishable as the iterations go on.

This result motivates the explicit exclusion of already measured points from the candidate pool at every iteration of the active learning loop: without this filter, the selection criterion could increasingly mistake the residual noise floor at an already sampled location for a really uninformative region, particularly once the unexplored fraction of the parameter space becomes small.

4.4 Acquisition score

The point-selection criterion used so far ranks unmeasured candidates by their aggregate *posterior* standard deviation $\sigma(\mathbf{x})$ (Eq. (3.19)). Applied without modification, this criterion shows to be biased due to the location of the selected points within the parameter domain: since the training data extends only up to the boundary of $[T_{\min}, T_{\max}] \times [\theta_{\min}, \theta_{\max}]$, the GP *posterior* variance is systematically higher near the edges of the domain. The boundary candidates have in fact fewer neighbouring points available to give informations about them.

If not corrected, this bias causes the acquisition criterion to over-select boundary points, as quantified in Section 5.1.2.

To balance this effect, the acquisition score for each candidate $\mathbf{x}$ is defined as the product of the aggregate *posterior* standard deviation and a centrality weight:

$$s(\mathbf{x}) = \sigma(\mathbf{x}) \cdot w(\mathbf{x}), \tag{4.2}$$

where the weight is a Gaussian function of the distance from the domain boundary. The definition and tuning of the boundary weight is reported in detail in Chapter 5.

4.5 Adaptive learning algorithm

Together with the surrogate definition of Chapter 3 and the acquisition score of Eq. (4.2) in place, this section describes how the two are combined into the operational adaptive learning loop used throughout Chapter 5. The surrogate is not fitted once and called repeatedly: both the POD basis and the Kriging models are refitted at every iteration since the POD basis depends on the training data and therefore changes as new points are added. The full procedure is summarised in the following steps.

1. Define the parameter space (T, θ) and the candidate grid (22×22).
2. Define the training set $\mathbf{X}_0$ (16 points, Cartesian product) and the test sets $\mathbf{X}_{t1}$, $\mathbf{X}_{t2}$ (100 points each, Sobol sampling).
3. Generate the fission matrices via FIM4NTRA for $\mathbf{X}_0$, $\mathbf{X}_{t1}$ and $\mathbf{X}_{t2}$.
4. **[begin loop, 40 iterations]**
   (a) Reduce the dataset dimensionality via POD.
   (b) Train the Kriging model on the POD coefficients.
   (c) Compute the aggregate posterior uncertainty $\sigma(\mathbf{x})$ (Eq. (3.19)) on the candidate grid, excluding already-measured points.
   (d) Compute the corrected acquisition score $s(\mathbf{x}) = \sigma(\mathbf{x}) \cdot w(\mathbf{x})$.
   (e) Run FIM4NTRA on the candidate with highest $s(\mathbf{x})$, and add the resulting pair $(\mathbf{x}, Y)$ to the training set.

   [end loop]
5. Predict the POD coefficients for a new $\mathbf{x}_*$ via Kriging.
6. Reconstruct the fission matrix via inverse POD projection, $\hat{Y}_* = \mathbf{\Phi}\hat{\mathbf{a}}_*$ (Eq. (3.7)).

7. Assess: compute k_{eff} and the fission source from $\hat{Y}_*$ and compare against the true values on $\mathbf{X}_{t1}$ and $\mathbf{X}_{t2}$.

8. If the error is not acceptable, return to step 5 and recalibrate the truncation order, the kernel configuration, or the boundary weight w_b (Chapter 5).

The general workflow of this study on the adaptive learning model is sketched in Figure 4.8.

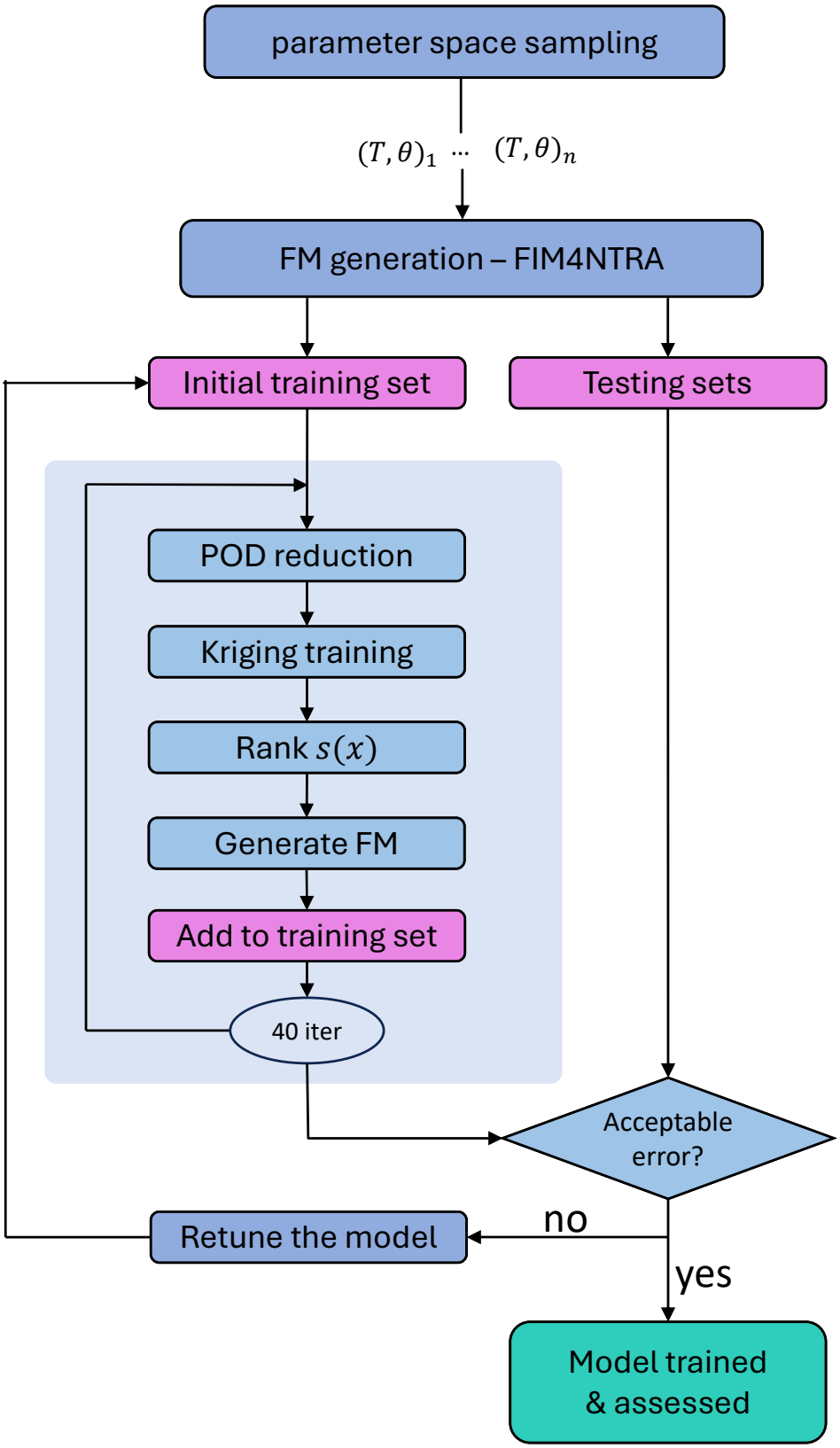


Figure 4.8: Workflow of the metamodel training.

Chapter 5

Parametric study and results

This chapter applies the adaptive learning algorithm of Section 4.5 to the VTB HPMR core and evaluates its performance against the baseline *posterior*-variance criterion. Section 5.1 identifies and quantifies the boundary oversampling bias of the uncorrected acquisition score, introduces the centrality weighting used to correct it, and calibrates its free parameter, the boundary weight w_b. Section 5.2 tests whether the resulting gain in accuracy is to attribute to the adaptive selection rule, by comparing it against random sampling baselines sharing the same initial conditions. Section 5.3 extends the error evaluation to the spatially resolved fission source, and assesses that the optimal configuration generalises to an independent test set.

5.1 Centrality Correction

As introduced qualitatively in Section 4.4, the oversampling at the domain border motivates the definition of the acquisition score $s(\mathbf{x}) = \sigma(\mathbf{x}) \cdot w(\mathbf{x})$ (Eq. (4.2)). This section derives the centrality weight $w(\mathbf{x})$ explicitly, quantifies the boundary bias it is designed to correct, and calibrates its single free parameter.

5.1.1 Boundary weighting formulation

The GP *posterior* standard deviation $\sigma(\mathbf{x})$ used in the raw acquisition criterion is systematically higher near the edges of the domain $[T_{\min}, T_{\max}] \times [\theta_{\min}, \theta_{\max}]$, since boundary candidates have fewer neighbouring points available to inform their prediction.

To balance this effect, the acquisition score for each candidate $\mathbf{x}$ is defined as the product of the aggregate posterior standard deviation and a centrality weight:

$$s(\mathbf{x}) = \sigma(\mathbf{x}) \cdot w(\mathbf{x}), \tag{5.1}$$

where the weight is a Gaussian function of the (normalised) distance $d_b(\mathbf{x})$ to the nearest domain boundary,

$$w(\mathbf{x}) = \exp\left[- \left(\frac{1 - d_b(\mathbf{x})}{s(w_b)} \right)^{2n} \right], \qquad s(w_b) = \big(- \ln w_b \big)^{-1/(2n)} \tag{5.2}$$

$$d_b(\mathbf{x}) = \frac{\min(\tilde{x}_1, 1-\tilde{x}_1, \tilde{x}_2, 1-\tilde{x}_2)}{0.5} \tag{5.3}$$

with $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2) \in [0,1]^2$ the input rescaled to the unit square, $n = 2$ a fixed shape exponent, and $s(w_b)$ a scale factor. The scale factor $s(w_b)$ is not a free parameter of the model but is fixed analytically by requiring $w(\mathbf{x}) = w_b$ exactly at the domain boundary. Since $d_b(\mathbf{x}) = 0$ at the boundary, Eq. (5.2) there reduces to

$$w_b = \exp\left[-\left(\frac{1}{s}\right)^{2n}\right] \iff \left(\frac{1}{s}\right)^{2n} = -\ln w_b \iff s = \big(-\ln w_b\big)^{-1/(2n)}, \tag{5.4}$$

which is the expression for $s(w_b)$ used above. This ensures that, whatever value of w_b is chosen, the resulting weight function always spans exactly from w_b at the edge to 1 at the centre ($d_b = 1$), so that w_b alone controls the strength of the correction without affecting its shape.

The single free parameter $w_b \in (0,1]$ is the *boundary weight*: $w_b = 1$ corresponds to the uncorrected criterion ($w(\mathbf{x}) \equiv 1$ everywhere), while $w_b \to 0$ increasingly eliminates candidates near the boundary regardless of their posterior uncertainty. Already-measured candidates are excluded from the acquisition score (Section 4.3.3), instead of being only down-weighted.

5.1.2 Boundary weight tuning

The boundary weight w_b was chosen over eight values, $w_b \in \{1, 0.8, 0.6, 0.4, 0.2, 0.05, 0.01, 0.005\}$, each tested with a dedicated active learning run of 40 iterations, starting from the same 16 initial points $\mathbf{X}_0$. Table 5.1 reports, for each value, the resulting k_{eff} error, fission-source error, and fraction of selected points landing on the domain boundary, all evaluated on the independent test set $\mathbf{X}_{t1}$.

w_b	k_{eff} error [pcm]	fission source [%]	boundary fraction
1	21.07	0.217	48%
0.8	17.61	0.219	35%
0.6	17.96	0.208	20%
0.4 (optimum)	**17.38**	0.217	0%
0.2	19.28	0.216	0%
0.05	19.91	0.223	0%
0.01	25.08	0.251	0%
0.005	35.96	0.288	0%

Table 5.1: Tuning of the boundary weight w_b: final k_{eff} error, fission source error, and fraction of the 40 selected points on the domain boundary.

The minimum k_{eff} error is reached at $w_b^* = 0.4$, with a value of 17.38 pcm against 21.07 pcm for the uncorrected criterion: an 18% reduction. The error is not monotonic in

the strength of the correction: weakening the penalty from $w_b^* = 0.4$ to $w_b = 1$ progressively reintroduces a boundary overselection, degrading accuracy. The same behaviour appears when strengthening it beyond $w_b^* = 0.4$ to $w_b \to 0$, with the error more than doubling at $w_b = 0.005$. This identifies an interior optimum: the boundary correction must be present and calibrated. An excessive penalization of the boundary candidates deprives the criterion of informative points that are near the domain edge, lowering accuracy just as severely as no correction at all. $w_b^* = 0.4$ is adopted as the operational configuration throughout the rest of this work. The trends of the mean k_{eff} for the whole sweep is shown in Figure 5.1.

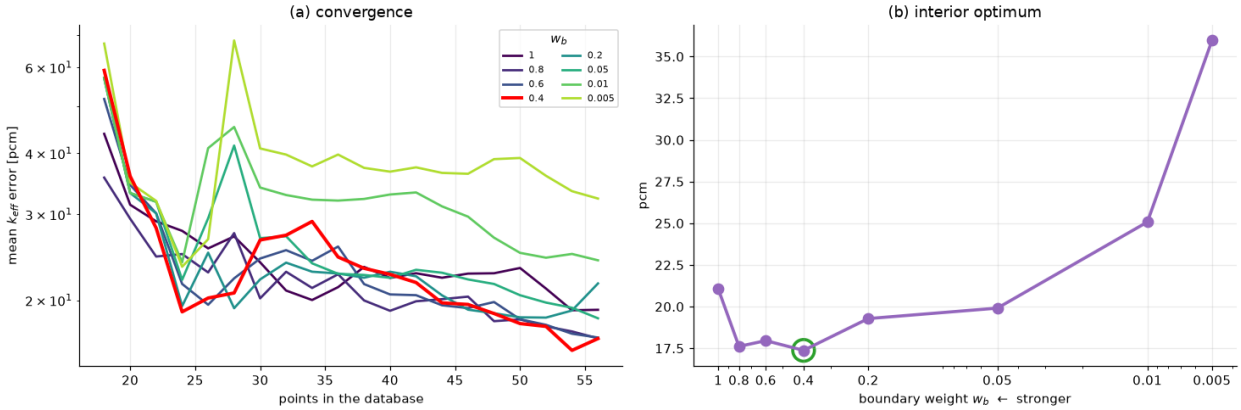


Figure 5.1: Mean k_{eff} error trends for different valeus of the *boundary weight* w_b.

5.1.3 Spatial distribution of selected points

Figure 5.2 compares the spatial distribution of the 40 points selected by the uncorrected criterion ($w_b = 1$) against the corrected criterion at its optimal setting ($w_b^* = 0.4$). Figure 5.3 shows the cumulative fraction of selected points landing on the boundary as the loop progresses. A uniform random sampling of the candidate grid, as reference, is expected to generate 17% of points on the domain boundary (the ratio of boundary to total candidates on the 22×22 grid).

Without correction, 48% of the 40 selected points end up on the domain boundary, nearly three times the 17% ratio expected in random sampling. This bias is already clear after only a handful of iterations: 60% of the first 5, and of the first 10, selected points lie on the boundary. With the optimal correction, the boundary fraction drops to is 0% during the entire run, at every stage of the cumulative count shown in Figure 5.2 (right). This confirms that the systematic preference for boundary candidates identified in Section 5.1.1 is not a transient phenomenon of the early iterations but a persistent bias of the uncorrected criterion based on the posterior variance, which is properly handled by the chosen weighting.

5.1.4 Evolution of uncertainty and acquisition score during the loop

To visualize how the implementation of the boundary weight $w(\mathbf{x})$ in the score system impacts the acquisition as the training set grows, Figure 5.4 shows, at five snapshots of the

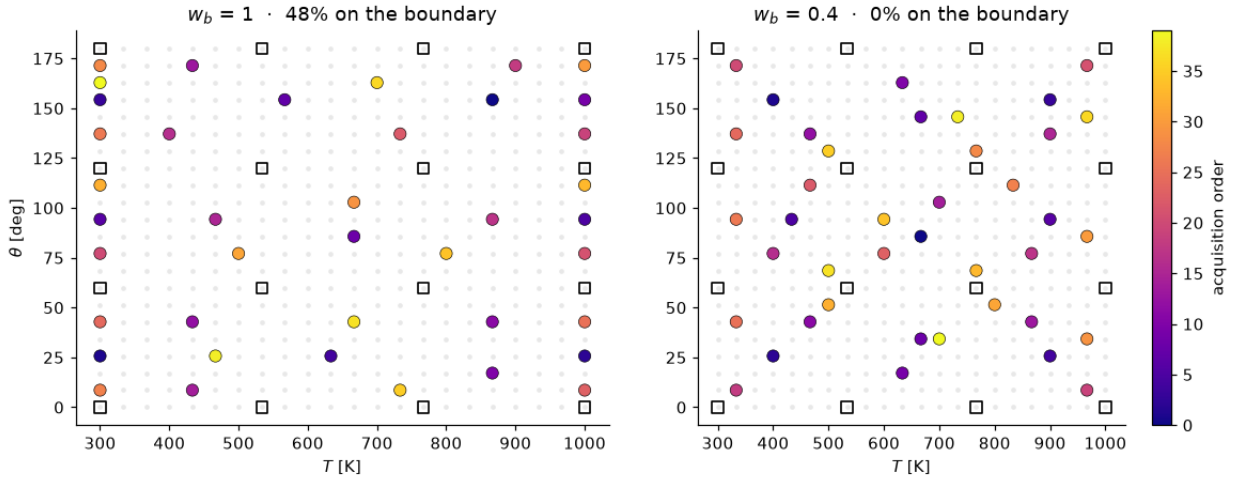


Figure 5.2: Location of the 40 acquired points (the colour indicates the acquisition order) for the uncorrected criterion ($w_b = 1$, left) and the optimal boundary correction ($w_b^* = 0.4$, right), with the initial points $\mathbf{X}_0$ shown as open squares.

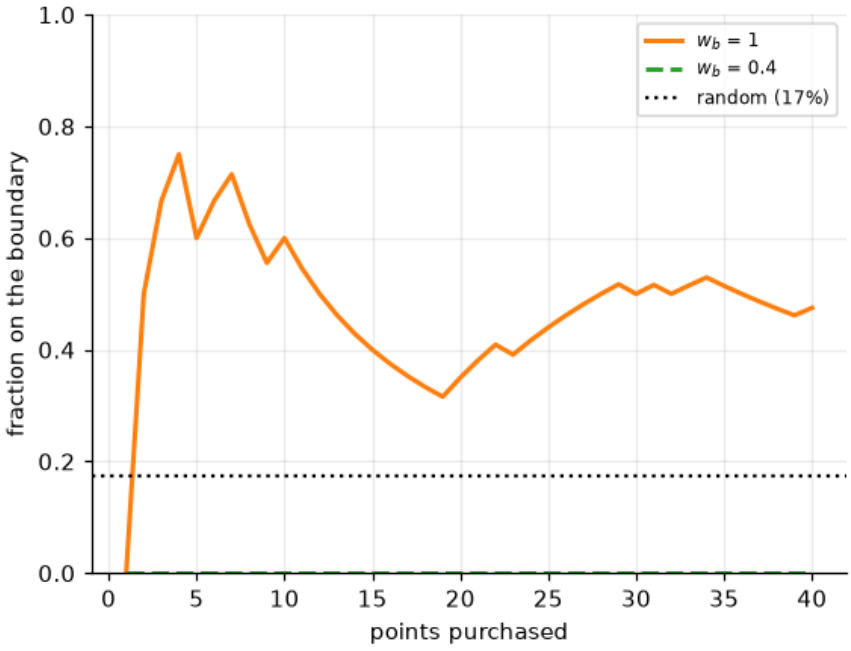


Figure 5.3: Cumulative fraction of selected points on the domain boundary as a function of the number of points acquired, compared to the 17% expected under random sampling.

optimal run ($w_b^* = 0.4$, after 16, 20, 25, 35 and 55 points), the spatial map of $\sigma(\mathbf{x})$ (top row) and of the resulting acquisition score $s(\mathbf{x}) = \sigma(\mathbf{x})\, w(\mathbf{x})$ (bottom row). The top row shows at each snapshot the candidate with the highest posterior variance, hence the one that would have been chosen by the uncorrected algorithm, while the bottom shows the point actually chosen by the corrected version. An arrow highlights the difference in the two locations, giving a qualitative indication on the impact of the correction on the points acquired.

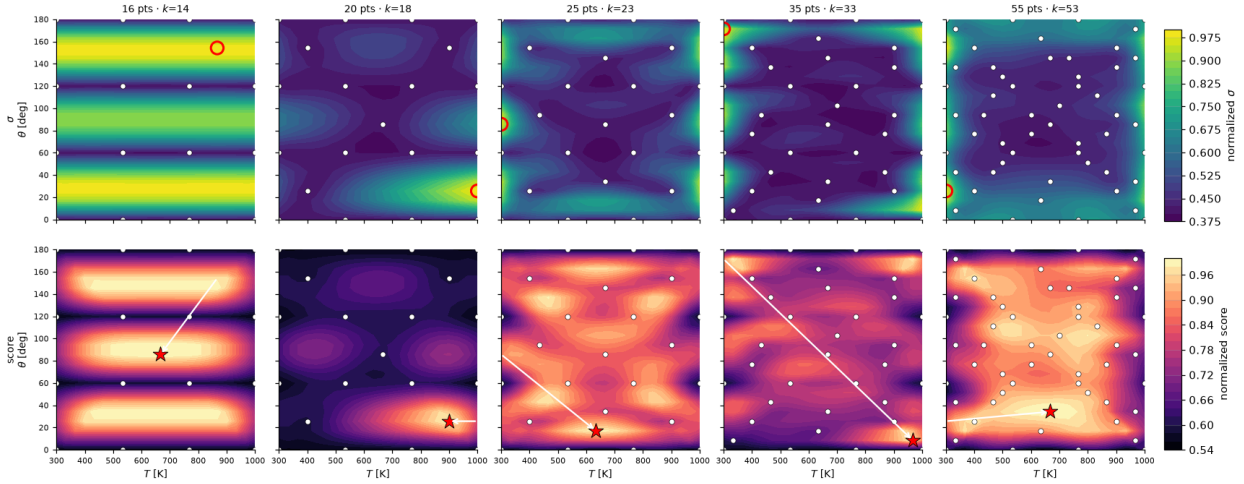


Figure 5.4: Top row: spatial map of the aggregate posterior standard deviation $\sigma(\mathbf{x})$, at five stages of the optimal run ($w_b^* = 0.4$); white markers show the current training points, the cyan circle marks the candidate that would be selected by uncertainty alone. Bottom row: the resulting acquisition score $s(\mathbf{x})$ of Eq. (4.2); the red star marks the point actually selected, with an arrow indicating the shift from the uncertainty-only choice.

Table 5.2 reports, at each snapshot, the locations of the two candidates and the distance between them in normalised coordinates. The centrality weighting consistently moves the acquisition away from the point of maximum uncertainty, differently across the iterations (0.14 to 1.31 in normalised units) depending on how close the point lies to the boundary at that stage of the run.

Points in database	$\arg\max \sigma$ (T, θ)	selected point (T, θ)	displacement
16	(867, 154)	(667, 86)	0.476
20	(1000, 26)	(900, 26)	0.143
25	(300, 86)	(633, 17)	0.610
35	(300, 171)	(967, 9)	1.314
55	(300, 26)	(667, 34)	0.526

Table 5.2: Displacement between the candidate maximising posterior variance $\sigma(\mathbf{x})$ and the candidate actually selected by the full acquisition score $s(\mathbf{x})$, at the five snapshots of Figure 5.4. Distance is computed in normalised input coordinates $\tilde{\mathbf{x}} \in [0, 1]^2$.

Across the run, the maximum value of $\sigma(\mathbf{x})$ over the candidate grid decreases by a factor of 5.6, from 5.16×10^{-2} at 16 points to 9.27×10^{-3} at 55 points, consistent with the surrogate

progressively resolving the parameter space as training data accumulates. Throughout this decrease, the centrality weighting remains active at every iteration, correcting the acquisition mainly on the domain edges. The correction is in fact not applied only in the high-uncertainty stages of the loop, but to the entire duration of the active learning cycle.

5.2 Sanity Check Against Random Sampling

The results of Section 5.1 establish that correcting the acquisition bias on the boundary of the domain increases the model performances with respect to a score only based on the posterior variance. As the loop iterates, the uncertainty is proven to decrease, but it is still not clear whether this improvement is due to the adaptive learning action, or only to the fact that the database size increases. Therefore this section contains the verification of whether the surrogate is just adding new points, or whether it is adding the appropriate new points. Since the acquisition criterion is deterministic, and the improvement of the boundary correction (Section 5.1.2) could in principle be a property of the grid geometry rather than of the adaptive strategy itself, this comparison is the verification that ultimately determines if the proposed method is worth adopting.

5.2.1 Verification of the Adaptive Learning sampling

The sanity check consists in the comparison between the adaptive algorithm sampling and a random sampling case. Twenty independent random baselines were generated, each starting from the same 16 initial points $\mathbf{X}_0$ used throughout this chapter and selecting the same number of additional points ($N_{\text{iter}} = 40$) from the parameter space, picked without replacement using an independent random seed ($1000 + s$ for baseline $s = 0, \ldots, 19$). Each baseline is evaluated on the same independent test set $\mathbf{X}_{t1}$, using the same metrics and the same final averaging (mean of the last 5 evaluations) as the adaptive runs of Section 5.1.2, so that the comparison isolates the effect of the selection rule alone.

5.2.2 Results

Across the 20 random baselines, the final k_{eff} error is 19.45 ± 1.96 pcm (mean $\pm$ standard deviation) and the fission-source error is $0.244 \pm 0.013\%$. Table 5.3 places the errors alongside the adaptive strategy at three points of the boundary weight sweep of Section 5.1.2.

Three observations follow directly from Table 5.3 and Figure 5.5:

- **Without boundary correction, the adaptive strategy is statistically indistinguishable from random sampling.** At $w_b = 1$, the adaptive error (21.07 pcm) falls within the range of the random distribution: 15 of the 20 random baselines achieve a lower error. The posterior variance criterion alone, without correction, provides no measurable advantage over uniform random selection on this problem.

- **With the optimal boundary correction, the adaptive strategy outperforms random sampling by a wide margin.** At $w_b^* = 0.4$, the adaptive error (17.38 pcm) is lower than 18 of the 20 random baselines. Only 2 random seeds achieve a lower

Strategy	k_{eff} error [pcm]	random baselines outperforming
Adaptive, $w_b = 1$ (uncorrected)	21.07	15/20
Adaptive, $w_b^* = 0.4$ (optimal)	**17.38**	2/20
Adaptive, $w_b = 0.005$ (over-corrected)	35.96	20/20
Random sampling (20 seeds)	19.45 ± 1.96	—

Table 5.3: Final k_{eff} error for the adaptive strategy at three boundary weight settings, compared against 20 independent random baselines sharing the same initial points and candidate pool.

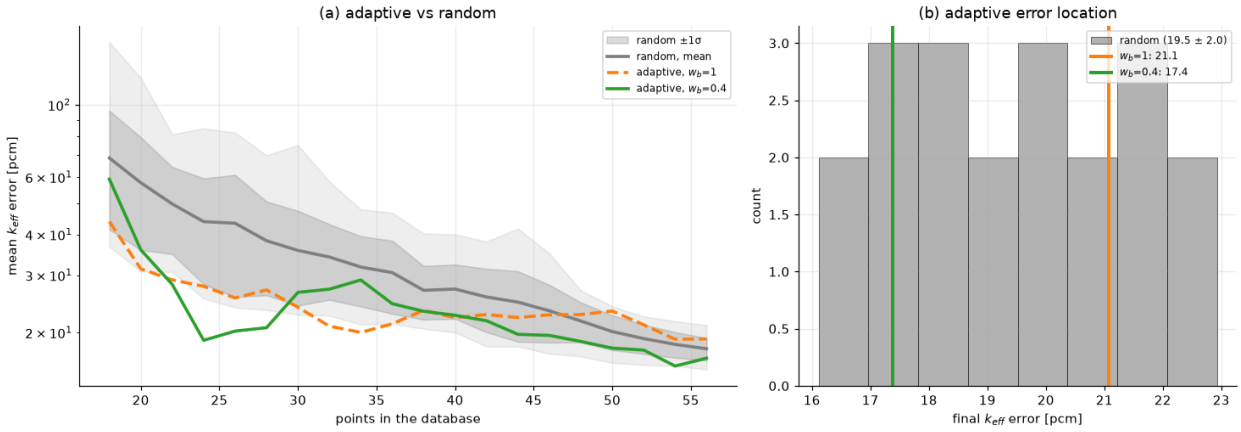


Figure 5.5: (a) Convergence of the mean k_{eff} error as a function of the number of points in the training database: random baselines (grey band, min–max range; darker band, ±1 standard deviation; solid grey line, mean) versus the adaptive strategy without boundary correction (orange dashed) and with the optimal correction (green solid). (b) Distribution of the final k_{eff} error over the 20 random baselines, with the adaptive strategy's uncorrected and optimally corrected results shown as vertical reference lines.

error, and none by a large margin. This means that the adaptive result lies more than one standard deviation below the random mean ($19.45 - 1.96 = 17.49$ pcm), despite having been produced by a single deterministic run rather than the best of many random attempts.

- **Over-correction is worse than no correction at all.** At the other extreme of the sweep, $w_b = 0.005$, every one of the 20 random baselines outperforms the adaptive strategy. This confirms, from the perspective of a random-sampling comparison, the interior optimum already identified in Section 5.1.2: an excessive suppression of the candidates on the boundary actively makes the adaptive strategy worse than guessing.

5.2.3 Interpretation

Taken together, these results indicate that the *posterior* variance uncertainty criterion, by itself, is a necessary but not sufficient tool for an effective active learning strategy on this problem: it must be paired with an explicit, calibrated information on where it is not convenient to sample, without which its practical benefit over random sampling vanishes entirely. This is, to the best of the author's knowledge, the central empirical finding of this work, and the primary justification for treating the centrality correction of Section 5.1 as a structural component of the proposed sampling strategy.

5.3 Power distribution error

The results of Sections 5.1 and 5.2 measure the adaptive strategy accomplishments only with the k_{eff} error on the evaluation set $\mathbf{X}_{t1}$ (Sobol seed 7, see Appendix A) used to drive the boundary weight testing. Two questions remain open: I) whether the increase in accuracy from the centrality correction translates to the *fission source* distribution and II) whether the optimal configuration identified on $\mathbf{X}_{t1}$ generalises to an independent test set. The fission source differs from the k_{eff} by being a local quantity, rather than an integral parameter.

5.3.1 Fission source error

The fission source error is defined as the mean, over the spatial regions of the core, of the relative deviation between the predicted and the reference (FIM4NTRA) normalised fission source, expressed in percent. Table 5.4 reports this metric, together with its 95^{th} percentile and maximum over the spatial regions, for every value of the boundary weight w_b tested in Section 5.1.2.

Unlike the k_{eff} (Table 5.1), the mean fission source error is comparatively uniform across the central portion of the batch ($w_b \in [0.2, 1]$, all within 0.208–0.219%) and reaches its minimum at $w_b = 0.6$ rather than at $w_b^* = 0.4$, the optimum found for k_{eff}. The two metrics are therefore not optimised by exactly the same boundary weight, even though the difference between the two (0.208% at $w_b = 0.6$ versus 0.217% at $w_b^* = 0.4$) is small relative to the range across the sweep as a whole. At the operational configuration $w_b^* = 0.4$, the fission source error is 0.215% (Table 5.5), compared to 0.244% for the random baselines of Section 5.2.

w_b	mean fs error [%]	95$^{\text{th}}$ pct. [%]	max [%]
1	0.217	0.56	1.55
0.8	0.219	0.58	1.99
0.6	0.208	0.56	1.81
0.4	0.217	0.59	1.82
0.2	0.216	0.61	1.69
0.05	0.223	0.61	1.87
0.01	0.251	0.71	2.62
0.005	0.288	0.80	2.97

Table 5.4: Fission source error (mean, 95$^{\text{th}}$ percentile, and maximum over spatial regions) across 8 boundary weights, evaluated on $\mathbf{X}_{t1}$.

None of the 20 random seeds achieves a lower fission source error than the adaptive strategy at its operational setting.

Strategy	fission source error [%]	random baselines outperforming
Adaptive, $w_b^* = 0.4$	0.215	0/20
Random sampling (20 seeds)	0.244	—

Table 5.5: Fission source error at the operational configuration versus random sampling (20 seeds).

As observed in the case of the boundary weight testing on k_{eff} (Section 5.1.2), the over-correction of the acquisition score increases the fission source error monotonically beyond the optimum. At $w_b = 0.005$ the mean error more than doubles to 0.288% and the maximum local error reaches 2.97%, roughly twice its value at the operational setting.

5.3.2 Generalisation to an independent test set

To verify that the boundary weight selected on $\mathbf{X}_{t1}$ is not a result specific to that particular random draw, every trained dataset from the sweep of Section 5.1.2 was additionally evaluated on a second, independent test set $\mathbf{X}_{t2}$ of 100 points (Sobol seed 99, excluding $\mathbf{X}_0$). Table 5.6 reports the resulting k_{eff} error with the corresponding value on $\mathbf{X}_{t1}$.

The optimal boundary weight identified on $\mathbf{X}_{t1}$, $w_b^* = 0.4$, is confirmed also the optimum on the independent set $\mathbf{X}_{t2}$. Moreover the two test sets agree not only on which configuration is best, but on the relative ranking of essentially every value of w_b tested. This indicates that the accuracy gain from centrality correction, and the specific weight at which it is maximised, is a property of the boundary driven selection itself rather than of the particular evaluation set used to identify it in Section 5.1.2.

w_b	k_{eff} error, $\mathbf{X}_{t1}$ [pcm]	k_{eff} error, $\mathbf{X}_{t2}$ [pcm]
1	19.1	18.6
0.8	16.7	16.5
0.6	16.8	16.5
0.4 (optimum on both)	**16.7**	**15.6**
0.2	21.7	20.4
0.05	18.4	17.7
0.01	24.2	24.2
0.005	32.3	32.8

Table 5.6: k_{eff} error across the boundary weight sweep, evaluated on both the test set $\mathbf{X}_{t1}$ (Sobol seed 7) and an independent test set $\mathbf{X}_{t2}$ (Sobol seed 99).

5.4 Fission matrix reconstruction versus direct prediction

This section compares the prediction of the fission matrix field (via POD+Kriging, then deriving k_{eff} and the fission source from it) against the prediction of k_{eff} and the fission source directly with a single GPR, bypassing the fission matrix entirely.

Five independent training sets (16 initial points $\mathbf{X}_0$ + 40 random candidates, independent seeds) were evaluated on $\mathbf{X}_{t1}$ at five training-set sizes, $N \in \{16, 26, 36, 46, 56\}$. Direct prediction uses an independent GPR with the same kernel configuration as Section 4.3.2, trained directly on the scalar $k_{\text{eff}}(T, \theta)$ and on the fission source vector.

The results are reported in Table 5.7 and Figure 5.6.

N	k_{eff}, via FM [pcm]	k_{eff}, direct [pcm]	fs, via FM [%]	fs, direct [%]
16	85.1 ± 0.0	41.9 ± 0.0	0.476 ± 0.000	0.567 ± 0.000
26	46.3 ± 19.1	80.2 ± 42.0	0.443 ± 0.093	0.663 ± 0.089
36	36.7 ± 10.8	71.7 ± 28.8	0.356 ± 0.031	0.657 ± 0.081
46	25.0 ± 4.9	51.1 ± 12.5	0.282 ± 0.039	0.606 ± 0.056
56	19.5 ± 0.7	57.9 ± 32.5	0.236 ± 0.012	0.544 ± 0.101

Table 5.7: k_{eff} and fission source error (mean $\pm$ std. over 5 realisations), via fission matrix vs. direct prediction.

The prediction done using the POD+Kriging on the whole fission matrix is more accurate on the fission source error at every N (e.g. 0.236% vs. 0.544% at $N = 56$). On the k_{eff}, the direct prediction is less accurate and less stable ($\pm$32.5 vs. $\pm$0.7 pcm at $N = 56$), and its error does not decay monotonically with N. Direct prediction has a better performance only at $N = 16$ (41.9 vs. 85.1 pcm). At every larger N, the prediction via FM is better and more consistent.

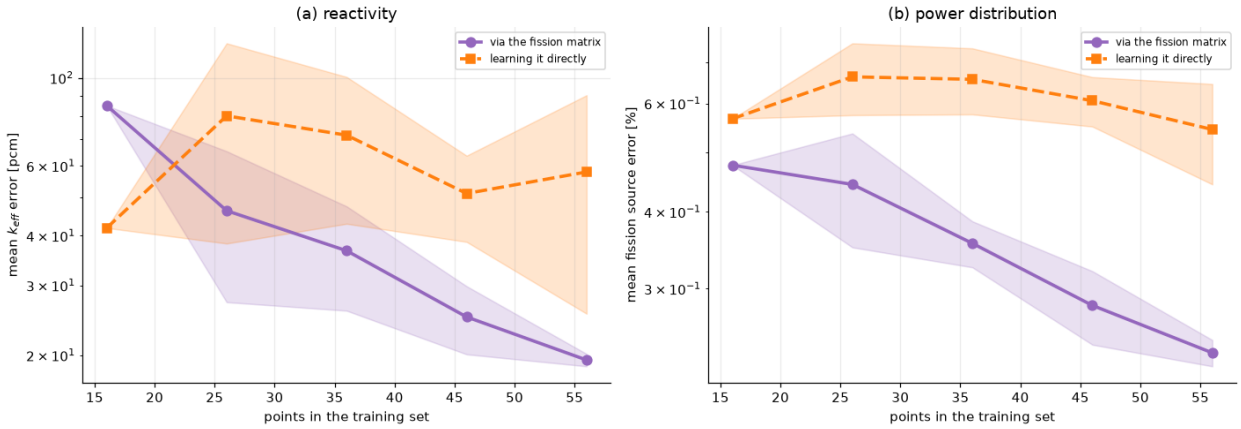


Figure 5.6: Comparison between the performances of the prediction based on the FM reconstruction and the direct prediction on the a) k_{eff} error and b) on the fission source error.

5.5 Computational cost of the surrogate

The computational cost of the surrogate itself is negligible with respect to the reference simulations. Fitting the POD+Kriging model on the full active-learning dataset (56 points) takes around one second (0.93 ± 0.1 s), dominated by the optimization of the Gaussian Process hyperparameters, while a single prediction requires around 2.3 ms and 30 ms for a 100 points-batch (0.30 ms/point). These numbers however are not directly comparable to the interpolation times reported for FIM4NTRA [4], since the latter already operates on a precomputed database and performs an analogous low-cost interpolation.
The actual computational benefit lies in the reduction of Monte Carlo simulations used for the training of the metamodel. Since surrogate evaluation cost is essentially free in comparison to a single reference simulation, the overall computational saving of the method is governed almost entirely by the reduction in the number of Serpent evaluations needed to reach a target accuracy.

Chapter 6

Conclusions and Future Work

The objective of this work is to determine whether physically representative fission matrices could be generated over a chosen two-dimensional parameter space with a machine learning based interpolation from a reduced database of high-fidelity simulations, using the VTB HPMT core. The parameters chosen are temperature and control drums rotation angle, and the high-fidelity code used to train the model is FIM4NTRA. The adaptive learning consists of an iterative cycle of 40 iterations in which at every step a new point of the domain is added to the dataset, choosing the point on the basis of an uncertainty score. The goal of the adaptive learning strategy is to use substantially fewer high-fidelity evaluations than uniform or random sampling, due to the high cost of fission matrices generated by a Monte Carlo run. The clear finding of this work is that the answer to this question depends heavily on the settings of the POD+Kriging model and the definition of the uncertainty score. Chapter 5 finds that a score based only on the posterior variance, applied without correction, is statistically indistinguishable from random sampling on this problem (21.07 pcm, outperformed by 15 of 20 random baselines), since it systematically over-samples the boundary of the parameter domain (48% of selected points, against 17% expected for a random selection). Once the score is integrated with a correction based on the distance of the candidates from the domain boundaries, the same acquisition criterion becomes effective: the fraction on points selected on the boundary drops to 0%, with an error on the k_{eff} of 17.38 pcm, corresponding to an 18% reduction. This criterion allows the surrogate model to outperform 18 out of the 20 random baselines (Section 5.2.2). This important results is shown to be valid also for the distribution of the fission source and for the prediction of the entire fission matrix itself. In addition to this key result, this work presents three additional findings in the model calibration sections (Chapter 4). First, the classical "energetic" criterion based on the fraction of total variance explained, in the case of fission matrix method, is not a suitable metric to assess the right POD truncation order. It in fact reveals that that fraction explained does not translate to an analogous reduction of the error on k_{eff}, that requires a full column rank with no truncations. Second, the disagreement among an ensemble of GPR models as measure of the predictive uncertainty is found to be unsuited for this problem with respect to the single model posterior variance, due to the fact the the log marginal likelihood has a dominant optimum. Third, the practice of deriving of the k_{eff} from the the predicted fission matrix is found be more accurate and stable with respect to a direct prediction of k_{eff} itself.

6.1 Necessity of a boundary correction

The result of Section 5.2 that uncertainty-driven sampling alone does not guarantee a measurable advantage over random acquisition is, at first sight, a negative result for the active learning strategy. A machine learning tool with capabilities not distinguishable form randomness feels useless at best, or even counterproductive. The present findings do not indicate that a posterior-variance is disadvantaged with respect to a random sampling criterion, but rather that it is not unconditional. It shows in fact that it can be entirely absorbed by a structural bias specific to the parameter space. One aspect of this case study could reasonably increase the relevance of this bias with respect to other applications. In fact the domain is bounded on both parameters (T and θ), so every point of the database within one length scale of either boundary contributes to an outward-facing region where the uncertainty is intrinsically more elevated. Using a domain unbounded in at least one direction, or periodic in the rotation angle θ (which the present formulation does not assume), would probably reduce this boundary behaviour. On the other hand, the domain is only two-dimensional, so the boundary is a comparatively small fraction of the domain total volume (17% of the candidate grid, Section 5.1.2). In parameter spaces of higher dimensions the boundary would occupy an even larger share of the volume of a bounded hyper-rectangle [17], suggesting that the bias observed in this work would not simply persist but likely worsen as additional operating parameters are introduced, rather than being a specific characteristic of the two-parameter case studied.

6.2 Choice of uncertainty source

Section 4.3.3 lays out the empirical motivation that justifies the adoption of the single-model posterior variance instead of a GP ensemble disagreement. The reason for this choice is that the log marginal likelihood for this problem is found to have a single clear optimum. Perturbing artificially the fitted length scales away from that optimum even by a factor 2 drops the log-likelihood by 347 points: the training data strongly disfavoured it, indicating that ensemble built by forcing diversity in the length scales would use points that the training set rejects.

Moreover, changing only the optimiser's random seed does not result in meaningfully different models. Five GPs trained on the same dataset, with same kernel, changing only the seed of the optimizer log-likelihood give identical length scales down to the sixth decimal place., with a dispersion between the predictions of the order of 10^{-11} (numerical noise).

Another contribution of Section 4.3.3 is the addition of a constraint: whatever the uncertainty source may be, a filter for duplicate candidates is needed, since as convergence is reached, the new truly informative signal from the points can not be told apart by the residual noise introduced by the nugget.

6.3 Mode Truncation Criterion

The important contribution of Section 4.3.1 is that the explained variance criterion is a misleading guide to the POD truncation order for applications in which the accuracy of the guess of the k_{eff} is central. POD modes are ranked in order of scale, for the largest to the smallest, reflecting also how much of the overall variance they explain. This criterion has no direct knowledge of k_{eff}, an eigenvalue of the reconstructed matrix. A relatively small mode that has a negligible contribution to the field's total variance can still make an important addition to its dominant eigenvalue if it perturbs the operator in a direction to which that eigenvalue is particularly sensitive. This consideration is not tied specifically to fission matrices: it is a general risk in every case where a POD-reduced surrogate is used to reconstruct a field that is not itself the variance (an eigenvalue, a spatial peak, a derived safety margin). This work therefore suggests that the order truncation choices should be taken, in general, against the quantity of ultimate interest, as done throughout this work, rather than against the explained variance ratio alone.

6.4 Limitations of the proposed approach

The limitations of this work, as defined in Section 1.4, are discussed here.

Reference limitation. The reference simulations used throughout this work are obtained with the FIM4NTRA code, which is an interpolated database itself, instead of Monte Carlo runs, for time management reasons. Therefore the accuracy reported in the study is a measure of the ability of the surrogate to reproduce that specific database, and not its absolute accuracy against transport calculations.
However all the results reported can be generalized with confidence to a case where Monte Carlo softwares, like Serpent 2, are used as the high-fidelity simulations. The reason behind this is that the surrogate limits itself to learn from a training dataset and verify against a testing set, hence not caring about the source of the high-fidelity matrices, whose generation can be seen as a "black box".

Dimensionality of the parameter space. As discussed in Section 6.1 the parameter space studied in this work is 2-dimensional, in temperature T and in control drums rotation angle θ. In higher dimensional parameter spaces the boundary bias observed in Section 5.1.2 is expected to grow in relevance, and the numerical tuning of the model (optimal boundary weight $w_b^* = 0.4$ and truncation order $r^* = m - 2$) is case specific and can not be transferred unmodified in differently shaped domains.

Single deterministic run. The AL strategy has been mostly evaluated as a single deterministic run per configuration, with the exception of the random sampled alternatives in Section 5.2. Even if in Section 4.3.3 the usefulness of an ensemble disagreement is verified to be non existent, it is worth mentioning that this study does not quantify the sensitivity to a different initial dataset or a different candidate grid resolution.

Error propagation. No statistical noise is propagated onto the POD coefficients during

training. Each training point yields a single point estimate of the reduced coefficients, with no variance term coming from the underlying Monte Carlo tallies or, as in this case, the FIM4NTRA code outputs. As a consequence, the acquisition criterion, based purely on the GP *posterior* variance, does not differentiate between the uncertainty due to sparse sampling and the aleatoric uncertainty due to MC statistical noise.

6.5 Future work

The results and limitations of this work point at the possible future directions of this work.

Extension to a higher-dimensional parameter space. The present study limits the operational exploration to a bidimensional space of temperature and angle. Section 6.1 hypothesizes that the issue with the oversampling of the domain boundary in a Kriging environment is expectable to worsen when translated to a higher-dimensional parameter space. The additional, for example of the extraction height of a central control rod, or water level in a PWR reactor, would increase the boundary to volume ratio of the domain, reasonably increasing the bias. The truncation order is not expected to be impacted, but the centrality penalty, through the boundary weight w_b would need to be recalibrated and verified.

Training and assessment against a Monte Carlo reference. The fission matrices and the values of k_{eff} guessed by the surrogate model in this work are all compared to those generated by the FIM4NTRA code, that acts as the high-fidelity source. The construction of a training and testing database with Monte Carlo simulations, with a Serpent 2 oracle for example, as well as remake of the sanity check (Section 5.2) procedure, would give a more solid background to the results of this study, opening to the chance of comparing the results directly with the reference literature.

Applications to other core designs. The truncation order, the hyperparameters tuning, the chosen uncertainty, as well as the centrality based correction proposed in this study are properties or the POD+Kriging model and the parameter space dimensions. However, paring the observations on the VTB HPMR core with a similar study on other types of reactor would test the generality claim presented here.

Appendix A

Sobol Low-Discrepancy Sequences

The Sobol sequence is a deterministic, low-discrepancy point set in $[0,1]^s$ that fills the hypercube more uniformly than pseudo-random sampling, reducing clustering and gaps typical of random samples [18].

Let $I^s = [0,1]^s$ and let $P_N = \{x_1, \dots, x_N\} \subset I^s$. For an axis-aligned box $B = \prod_{i=1}^{s} [0, b_i) \subset I^s$, the local discrepancy is

$$D_N(B) = \left| \frac{A(B; P_N)}{N} - \mathrm{vol}(B) \right|, \tag{A.1}$$

where $A(B; P_N)$ counts the points of P_N inside B. The star discrepancy is $D_N^*(P_N) = \sup_B D_N(B)$. A sequence is called low-discrepancy if

$$D_N^*(P_N) = O\left(\frac{(\log N)^s}{N} \right), \tag{A.2}$$

which is asymptotically better than the $O(N^{-1/2})$ rate of independent uniform random points [18].

The Sobol sequence is a (t, s)-sequence in base 2. Each coordinate is built from *direction numbers* $\{v_1, v_2, \dots\}$, generated recursively from a polynomial over GF(2):

$$p(z) = z^d + a_1 z^{d-1} + \cdots + a_{d-1} z + 1, \qquad a_j \in \{0, 1\}, \tag{A.3}$$

via

$$v_i = a_1 v_{i-1} \oplus \cdots \oplus a_{d-1} v_{i-d+1} \oplus v_{i-d} \oplus \left\lfloor v_{i-d} / 2^d \right\rfloor, \tag{A.4}$$

with $\oplus$ the boolean XOR. The n-th point is obtained as $x_n = \bigoplus_k b_k \, v_{k+1}$, where b_k are the bits of the Gray-code representation of n (i.e. binary code where each number differs by 1 bit). Using the Gray code lets each new point be generated from the previous one in $O(1)$ operations. Each of the s dimensions uses its own primitive polynomial and initial direction numbers, so that coordinates remain individually valid 1-D sequences while being uncorrelated across dimensions. Poorly chosen initialisations can otherwise cause visible correlation

between dimensions.

Sobol sequences present the following **properties**.

- (t, m, s)**-net property.** Every elementary interval of volume 2^{t-m} contains exactly 2^t points out of the first 2^m, guaranteeing balanced coverage at every dyadic scale.
- **Extensibility.** The sequence can be extended incrementally without discarding previous points, unlike fixed-size designs such as Latin Hypercube Sampling.
- **Convergence rate.** For bounded-variation functions, quasi-Monte Carlo integration with Sobol points achieves $O((\log N)^s/N)$ error, compared to $O(N^{-1/2})$ for pseudo-random sampling [18].
- **Related variance-reduction use.** A related application of the same low-discrepancy construction is the variance-reducing multiplier technique for global sensitivity indices.

Bibliography

[1] William J. Walters, Nathan J. Roskoff, and Alireza Haghighat. The RAPID fission matrix approach to reactor core criticality calculations. *Nuclear Science and Engineering*, 192(1):21–39, 2018.

[2] Sean E. Carney, Forrest B. Brown, Brian C. Kiedrowski, and William R. Martin. Theory and applications of the fission matrix method for continuous-energy monte carlo. *Annals of Nuclear Energy*, 73:423–431, 2014.

[3] Binhang Zhang, Yanzhao Bi, Changxiao Guan, Xianbao Yuan, Yonghong Zhang, and Yunlong Xiao. Adaptive POD-ROM based on Grassmann manifold: A robust and efficient approach for neutronics-depletion coupling calculations in nuclear reactors. *Progress in Nuclear Energy*, 189:105932, 2025.

[4] Christian Vita, Nicolo' Abrate, Sandra Dulla, William Walters, Valerio Mascolino, Nicolas Stauff, and Fausto Franceschini. FIM4NTRA: A code for neutronics analyses based on the fission matrix method. In *Proceedings of the ANS Winter Meeting*. American Nuclear Society, 2026. DOI to be assigned by ANS.

[5] Marius Marinescu. Explaining and Connecting Kriging with Gaussian Process Regression. *arXiv preprint arXiv:2408.02331*, 2025. Updated October 14, 2025.

[6] K. W. Morton. Criticality calculations by Monte Carlo methods. Technical Report T/R-1903, United Kingdom Atomic Energy Research Establishment, Harwell, 1956.

[7] E. L. Kaplan. Monte Carlo methods for equilibrium solutions in neutron multiplication. Technical Report UCRL-5275-T, Lawrence Radiation Laboratory, 1958.

[8] C. Josey and M. D. Veit. Eigenfunction decomposition of reactor perturbations and transitions using MCNP Monte Carlo. Technical Report LA-UR-13-26449, Los Alamos National Laboratory, 2013.

[9] Julien Weiss. A tutorial on the proper orthogonal decomposition. In *AIAA Aviation 2019 Forum*, Dallas, Texas, United States, June 2019. American Institute of Aeronautics and Astronautics. Accepted manuscript (postprint).

[10] Lawrence Sirovich. Turbulence and the dynamics of coherent structures. Part I: Coherent structures. *Quarterly of Applied Mathematics*, 45(3):561–571, 1987.

[11] Daniel G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.

[12] Georges Matheron. *Principles of geostatistics*, volume 58. 1963.

[13] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[14] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*. Springer, 2nd edition, 2018.

[15] Kai-Tai Fang, Runze Li, and Agus Sudjianto. *Design and Modeling for Computer Experiments*. Chapman and Hall/CRC, 2016.

[16] N Stauff, A Abdelhameed, Y Cao, Y Miao, K Mo, D Nunez, et al. Multiphysics analysis of load following and safety transients for microreactors. Technical Report ANL/NEAMS-22/1, Argonne National Lab.(ANL), Argonne, IL (United States), 2022.

[17] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2nd edition, 2009.

[18] I. M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

[19] Adam Rau and William J. Walters. Nuclear science and engineering. Technical report, 2021.