ON ISOTOPIC RATIO SIGNATURES FOR NUCLEAR ARCHAEOLOGY

A Bayesian Approach for Verifying Fissile Material Declarations

Von der Fakultät für Mathematik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

BENJAMIN JUNG, M.SC.

aus

Troy, USA

Berichter: Prof. Dr. rer. nat. Malte Göttsche
Univ.-Prof. Dr. rer. nat. Achim Stahl

Tag der mündlichen Prüfung: 31.07.2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek verfügbar.

ABSTRACT

Peaceful, cooperative, and sustainable nuclear disarmament requires means for technical verification, such as those comprising "nuclear archaeology". This work continues development of a novel nuclear archaeology technique called Bayesian Reprocessing Waste Analysis Method (BRAM), which proposes the use of isotopic signatures in samples of reprocessing waste to reconstruct reactor operating histories for the purpose of verifying fissile material baseline declarations.

First, a new, genetic algorithm was developed to optimize the choice of isotopic ratios that are used to inform the reconstruction process. This proposed set of isotopic ratios is chosen to be able to discriminate between three different reactor types, namely PHWR, PWR, and Magnox; and reconstruct previously neglected inference parameters, that is, initial enrichment and average power density.

With two systematic, simulation-based test series this ratio set was shown to successfully reconstruct four reactor operating parameters covering a large variety of possible true parameter combinations while focusing on the low burnup regime ($0.1\,\mathrm{MW\,d\,kg^{-1}}$ to $9\,\mathrm{MW\,d\,kg^{-1}}$), which is more relevant for military plutonium production. The first test series showed the potential to reconstruct parameters when waste from two reactor cycles from different reactors has been mixed in a sample. The second test series showed the possibility to discriminate between three reactor types, if the origin of the sample is unknown, while simultaneously inferring the operating history parameters.

Building on these simulation-based results, steps towards validating this Bayesian framework are taken with experimental data from the SFCOMPO database. Here, parameters such as average burnup, power density, initial enrichment are reconstructed successfully for a PWR reactor and burnup, cooling time, and power density for a Magnox reactor type, demonstrating the working principle of the method; while highlighting the need for accurate uncertainty assessments.

Finally, a simulation-based case study of reconstructing the operating history of the Democratic People's Republic of Korea's 5 MWe reactor demonstrates the potential utility of the Bayesian framework, showing that existing assumptions about the reactor's history could be corroborated or disproven and that there is potential to ascertain more information about the early reactor operations. All in all, this work shows promising steps forward in the development of the BRAM as a tool for reconstructing reactor operating histories and verifying nuclear disarmament.

ZUSAMMENFASSUNG

Friedliche, kooperative und nachhaltige nukleare Abrüstung erfordert technische Verifikationsmethoden, wie die der „nuklearen Archäologie“. In dieser Arbeit wird die „Bayesian Reprocessing Waste Analysis Method“ (BRAM) weiterentwickelt. Diese Methode sieht vor, anhand von Isotopensignaturen in Proben von Wiederaufbereitungsabfällen den Betriebsverlauf von Reaktoren zu rekonstruieren, um Dekalarationen von Spaltmaterial zu verifizieren.

Zunächst wurde ein neuer Algorithmus entwickelt, um die Auswahl der Isotopenverhältnisse zu optimieren, die als Grundlage für die Rekonstruktion dienen. Dieser vorgeschlagene Satz von Isotopenverhältnissen wurde so gewählt, dass er eine Unterscheidung zwischen drei verschiedenen Reaktortypen—PWR, PHWR und Magnox—ermöglicht und bisher vernachlässigte Inferenzparameter, nämlich den Anreicherungsgrad und die Leistungsdichte, rekonstruiert.

Mit zwei systematischen, simulationsbasierten Testreihen wurde gezeigt, dass diese Isotopenverhältnisse vier Reaktorbetriebsparameter erfolgreich rekonstruieren können. Dabei wurde eine Bandbreite möglicher Parameterkombinationen abgedeckt, wobei der Schwerpunkt auf niedrigen Abbrandwerten ($0.1\,\mathrm{MW\,d\,kg^{-1}}$ bis $9\,\mathrm{MW\,d\,kg^{-1}}$) lag. Die erste Testreihe zeigte, dass eine Rekonstruktion von Parametern möglich ist, wenn Abfälle aus zwei Reaktorzyklen verschiedener Reaktoren in einer Probe vermischt wurden. Die zweite Testreihe zeigte, dass es möglich ist, bei unbekannter Herkunft der Probe zwischen drei Reaktortypen zu unterscheiden und gleichzeitig Rückschlüsse auf die Betriebsparameter zu ziehen.

Aufbauend auf diesen Ergebnissen wurden erste Schritte zur Validierung der Methode anhand experimenteller Daten aus der SFCOMPO-Datenbank unternommen. Dabei ließen sich Parameter wie der Abbrand, die Leistungsdichte und der Anreicherungsgrad für einen Druckwasserreaktor sowie der Abbrand, die Abkühlzeit und die Leistungsdichte für einen Magnox-Reaktor erfolgreich rekonstruieren, was die Funktionsweise der Methode demonstriert und gleichzeitig die Notwendigkeit genauer Unsicherheitsanalysen hervorhebt.

Schließlich verdeutlicht eine simulationsbasierte Fallstudie anhand des nordkoreanischen $5\,\mathrm{MWe}$-Reaktors den potenziellen Nutzen des Bayes’schen Ansatzes: Sie zeigt, dass bestehende Annahmen über die Geschichte des Reaktors bestätigt oder widerlegt werden könnten und dass die Möglichkeit besteht, weitere Informationen über den frühen Reaktorbetrieb zu gewinnen. Insgesamt zeigt diese Arbeit vielversprechende Fortschritte bei der Entwicklung der BRAM als Methode zur Rekonstruktion der Betriebsgeschichte von Reaktoren und zur Überprüfung der nuklearen Abrüstung.

PUBLICATIONS

The research methods and results discussed in this thesis have contributed to the following journal articles and conference proceedings. The algorithms and frameworks developed as part of this research have been released in open-source software repositories.

JOURNAL ARTICLES

- B. Jung, A. Figueroa, and M. Göttsche, "Towards inferring reactor operations from high-level waste," Nuclear Engineering and Technology (2024) DOI: 10.1016/j.net.2024.02.031.
- B. Jung, J. Bosse, and M. Göttsche, "Reconstructing North Korea's plutonium production history with Bayesian inference-based reprocessing waste analysis," Nuclear Engineering and Technology **58**, 103849 (2026) DOI: 10.1016/j.net.2025.103849.

CONFERENCE PAPERS

- M. Göttsche, K. Ermert, B. Jung, and L. Rademacher, "Verifying initial declarations using nuclear archaeology methods," in IAEA Symposium on International Safeguards (Oct. 31–, Nov. 4, 2022).
- B. Jung and M. Göttsche, "Inferring reactor parameters from nuclear waste with Bayesian inference," in Proceedings of the INMM 65th Annual Meeting (July 21–25, 2024).

SOFTWARE RELEASES

- B. Jung, *Jung-benjamin/gp-surfer: v0.5.3*, version v0.5.3, Zenodo, June 2, 2025, DOI: 10.5281/zenodo.15576629.
- B. Jung, *Jung-benjamin/rasor: v0.3.1*, version v0.3.1, Zenodo, Apr. 22, 2026, DOI: 10.5281/zenodo.19696049.
- B. Jung, *Jung-benjamin/bram-inferno: v0.8.1*, version v0.8.1, Zenodo, June 2, 2025, DOI: 10.5281/zenodo.15577482.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to all the people who have supported, taught, and guided me throughout this work.

I am tremendously grateful to my supervisor, Prof. Malte Göttsche, for the opportunity to work on this research project. I am grateful for his support and guidance. He provided many opportunities to engage with new topics and meet fascinating people.

I thank Prof. Achim Stahl for sharing his knowledge. I have learned from him since he taught my very first lecture on experimental physics over ten years ago.

I am deeply grateful to my colleagues of the former Nuclear Verification and Disarmament group, all of whom contributed to a stimulating work environment and entertaining after-work exploits:

to Max Schalz for all the moments of shared joy, excitement and frustration; for the hours together in the office, debugging code and discussing new ideas; for making great coffee.

to Ana de la Varga for all the advice, the humor, and the sunny outlook on life; to Sophie Kretzschmar for the discussions on science, politics, and philosophy and for reading earlier drafts.

to Antonio Figueroa, Lukas Rademacher and Yan-Jie Schnellbach for enlightening conversations about nuclear archaeology, reactor simulations and Bayesian inference.

to Katja Ermert and Johannes Bosse who were eager students and helped me grow as a teacher.

I thank Stefaan van Winckel for sharing his knowledge on radiochemistry and isotope ratio measurements.

My thanks also go out to Fabian Unruh, Jakob Böttcher, Christopher Fichtlscherer and Hanno Jacobs for reading the earlier, messier drafts of this thesis.

I am beyond grateful to my parents for their unwavering support without which I would not be where I am today. They taught me patience, nurtured my curiosity, and were always there when I needed

them. Special thanks go to my sister, who has been a lifelong companion and greatest of friends; who inspires and motivates me.

Thank you Cristina, for your support and love. I am grateful for every second together, for all the exciting adventures, all the moments of peaceful relaxation, and all the time in between.

I gratefully acknowledge the computing time provided to me at the NHR Centers NHR4CES at RWTH Aachen University (project number p0020230) and at TU Darmstadt (project number p0022838). This is funded by the Federal Ministry of Research, Technology and Space, and the state governments participating on the basis of the resolutions of the GWK for national high performance computing at universities (https://www.nhr-verein.de/unsere-partner).

CONTENTS

I TOWARDS VERIFIABLE NUCLEAR DISARMAMENT 1

1 INTRODUCTION 3
- 1.1 The destructive power of chain reactions . . . 3
- 1.2 Verifying fissile material declarations for nuclear disarmament . . . 4
- 1.3 Nuclear archaeology . . . 5
- 1.4 Developing a nuclear archaeology method for analyzing reprocessing waste . . . 6

II A BAYESIAN FRAMEWORK FOR RECONSTRUCTING REACTOR OPERATING PARAMETERS 9

2 MODELLING THE COMPOSITION OF HIGH-LEVEL NUCLEAR WASTE 11
- 2.1 On the physics of plutonium production . . . 11
  - 2.1.1 Nuclear Reactor Physics . . . 11
  - 2.1.2 Nuclear Reprocessing . . . 14
- 2.2 Modelling nuclear reactor physics . . . 15
  - 2.2.1 Depletion Calculations . . . 15
  - 2.2.2 Neutron Transport . . . 17
  - 2.2.3 Modelling the time evolution of nuclear fuel . . . 21
  - 2.2.4 Modeling Reactor Geometry and Approximations 22
  - 2.2.5 Nuclear Data Libraries . . . 23
  - 2.2.6 Using OpenMC . . . 23

3 THE BAYESIAN REPROCESSING WASTE ANALYSIS METHOD 25
- 3.1 Model Inversion with Bayesian Inference . . . 25
  - 3.1.1 Bayesian inference . . . 25
  - 3.1.2 Model inversion . . . 26
  - 3.1.3 Comparing models . . . 27
- 3.2 Markov Chain Monte Carlo Sampling . . . 28
  - 3.2.1 Convergence of Markov chain Monte Carlo . . . 30
- 3.3 Inferring Reactor Operating Parameters from Nuclear Waste . . . 31
- 3.4 Surrogate modelling with Gaussian Processes . . . 34
  - 3.4.1 Gaussian Processes . . . 35
  - 3.4.2 Hyperparameter Optimization . . . 36
  - 3.4.3 Training Gaussian Process Surrogate Models in Practice . . . 36
- 3.5 Software for the Bayesian reprocessing waste analysis method . . . 37

4 SELECTING SUITABLE ISOTOPIC RATIOS 39
- 4.1 Selecting Isotopic Ratio Candidates . . . 40

4.1.1 Converting spent nuclear fuel (SNF) to highly active raffinate (HAR) . . . 40
4.1.2 Detection threshold . . . 43
4.1.3 Excluding volatile nuclides and decay products 43
4.1.4 Selecting candidate ratios . . . 44
4.2 Optimizing a Subset of the Ratio Candidates . . . 46
4.2.1 Objective functions . . . 46
4.2.2 A combinatorial approach to comparing ratio lists 49
4.2.3 A genetic algorithm for comparing ratio lists . . 49
4.2.4 On the number of selected ratios . . . 51
4.3 Comparing the suitability of ratio lists . . . 52
4.3.1 Test criteria . . . 52
4.4 Selecting a new ratio list . . . 54
4.4.1 Tested algorithm configurations . . . 54
4.4.2 Ratio selection results . . . 55
4.4.3 Discussion of selected ratios . . . 58

III TESTING, VALIDATION AND APPLICATIONS 61

5 SIMULATION-BASED TESTING 63
5.1 Reactor models . . . 63
5.1.1 Surrogate models . . . 66
5.2 The extended "two-batch" scenario . . . 67
5.2.1 A likelihood function for mixed-waste samples 68
5.2.2 Mixture model testing . . . 68
5.2.3 Discussion of the two-batch scenario . . . 71
5.3 Discriminating reactor types . . . 74
5.3.1 Bayesian model classification . . . 75
5.3.2 Classifier testing . . . 75
5.3.3 Discussing reactor type classification . . . 78

6 VALIDATING WITH SFCOMPO DATA 83
6.1 Validation with PWR data . . . 84
6.1.1 Samples from the Obrigheim reactor . . . 84
6.1.2 An Obrigheim inference model . . . 84
6.1.3 Obrigheim validation results . . . 86
6.2 Validation with Magnox data . . . 87
6.2.1 Magnox measurement data . . . 87
6.2.2 A Magnox inference model . . . 88
6.2.3 Exploring model calibration . . . 89
6.2.4 Validation results: Magnox . . . 91
6.3 Discussion of validation results . . . 93

7 POTENTIAL APPLICATIONS IN THE DPRK 95
7.1 A brief history of plutonium production in North Korea 96
7.1.1 Modelling the 5 MWe reactor . . . 96
7.2 Reconstructing the 5 MWe reactor's operating history . 97
7.3 The *1994 discharge* scenario with varying priors . . . 99
7.4 Discriminating *pre-1994* and *1994* discharge scenarios . 102

7.4.1 A posterior plutonium estimate 102
7.5 Discussing application potential 103
8 CONCLUSIONS 105
8.1 Summary 105
8.2 Implications 106
8.3 Outlook 107

IV APPENDIX 109
A COMPUTATIONALLY EFFICIENT APPROXIMATION OF MARGINAL POSTERIOR DISTRIBUTIONS 111
B PERFORMANCE TEST STATISTICS 113
B.1 Test statistics of the "two-batch" mixture scenario . . . 114
B.2 Test statistics of the classifier scenario 129

BIBLIOGRAPHY 131

LIST OF FIGURES

Figure 1 Sequential reactor physics simulation approac 21
Figure 2 Calibration data for nuclide pre-selection . . . 41
Figure 3 Ratio selection success statistics 56
Figure 4 Ratio selection: information gain 57
Figure 5 Reactor model schematics 64
Figure 6 A posterior of a mixture model. 70
Figure 7 Convergence ratio of two-batch mixture 71
Figure 8 Success rate in two-batch mixtures. 72
Figure 9 KL-divergence in two-batch mixtures. 73
Figure 10 Confusion matrix 76
Figure 11 Posterior of classification model. 77
Figure 12 Reactor classification convergence statistics . . 78
Figure 13 Reactor classification convergence by parameter 80
Figure 14 Reactor classification performance statistics . . 81
Figure 15 Obrigheim reactor model 86
Figure 16 Posteriors for PWR SFCOMPO sample 87
Figure 17 Bradwell-1 reactor model 89
Figure 18 Magnox model calibration 91
Figure 19 Posteriors for Magnox SFCOMPO samples . . 92
Figure 20 Timeline of the 5 MWe reactor operation. . . . 96
Figure 21 DPRK scenario posterior parameter 101
Figure 22 DPRK scenario posterior plutonium estimates 103
Figure 23 Mixture scenario convergence statistics vs BU. 114
Figure 24 Mixture scenario convergence statistics vs CT. 115
Figure 25 Mixture scenario convergence statistics vs IE. . 116
Figure 26 Mixture scenario convergence statistics vs PD. 117
Figure 27 Mixture scenario convergence statistics vs α. . 118
Figure 28 Mixture scenario success statistics vs BU. . . . 119
Figure 29 Mixture scenario success statistics vs CT. . . . 120
Figure 30 Mixture scenario success statistics vs IE. 121
Figure 31 Mixture scenario success statistics vs PD. . . . 122
Figure 32 Mixture scenario success statistics vs α. 123
Figure 33 Mixture scenario information gain statistics vs BU. 124
Figure 34 Mixture scenario information gain statistics vs CT. 125
Figure 35 Mixture scenario information gain statistics vs IE. 126
Figure 36 Mixture scenario information gain statistics vs PD. 127

Figure 37 Mixture scenario information gain statistics vs α. 128
Figure 38 Classifier success statistics scatterplot 129
Figure 39 Classifier information gain statistics scatterplot 130

LIST OF TABLES

Table 1 Pre-selected candidate nuclides 45
Table 2 Optimized ratio list 58
Table 3 Reactor design parameters 66
Table 4 Performance test priors 67
Table 5 Ratios of PWR nuclide data (SFCOMPO). . . . 85
Table 6 Inference parameters for PWR validation . . . 85
Table 7 Ratios of Magnox nuclide data (SFCOMPO). . 88
Table 8 Inference parameters for Magnox validation . 88
Table 9 Design parameters of the 5 MWe reactor 97
Table 10 Two potential operating histories of the 5 MWe reactor in Yongbyon 98
Table 11 Isotopic ratios in the DPRK scenario 99
Table 12 Prior variations for 5 MWe inference test cases 100
Table 13 DPRK scenario model weights 102

ALGORITHMS

Algorithm 1 Objective function for ratio selection. 48
Algorithm 2 Combinatorial ratio selection algorithm 49
Algorithm 3 A basic genetic algorithm 50
Algorithm 4 Genetic ratio selection algorithm 51

Algorithm 5 Approximating marginal distributions 112

ACRONYMS

ABC approximate Bayesian computation

BRAM Bayesian Reprocessing Waste Analysis Method
BU Burnup
CANDU CANada Deuterium Uranium
CT Cooling Time
DPRK Democratic People's Republic of Korea
elpd Expected Log Pointwise Predictive Density
ELWR Experimental Light-Water Reactor
ENDF Evaluated Nuclear Data File
GIRM Graphite Isotope Ratio Method
GP Gaussian process
HAR highly active raffinate
HDI highest (posterior) density interval
HEU highly-enriched uranium
HLW high-level waste
HMC Hamiltonian Monte Carlo
IAEA International Atomic Energy Agency
ICP-MS inductively coupled plasma mass spectrometry
IE Initial Enrichment
KL Kullback-Leibler
LOO-CV Leave-One-Out Cross-Validation
Magnox Magnesium Non-OXidising
MCMC Markov Chain Monte Carlo
NPT Treaty on the Non-Proliferation of Nuclear Weapons
NUTS No-U-Turn Sampler
PD Power Density
PHWR Pressurized Heavy-Water Reactor
PSIS Pareto smoothed importance sampling
PUREX Plutonium Uranium Reduction EXtraction
PWR Pressurized Water Reactor
SFCOMPO Spent Fuel COMPOsition
SNF spent nuclear fuel
TIMS thermal ionization mass spectrometry

Part I

TOWARDS VERIFIABLE NUCLEAR DISARMAMENT

1 INTRODUCTION

1.1 THE DESTRUCTIVE POWER OF CHAIN REACTIONS

A "blinding light that flashed as if a thousand magnesium bulbs had been turned on all at once" are the words that Kikue Shiota uses to describe her recollection of the nuclear bomb that destroyed Hiroshima. She survived the blast trapped under the rubble of her family's house, more than one kilometer away from the center of the explosion [1].

Dropping a similarly sized bomb on Aachen—where I am sitting as I write this—would destroy buildings and ignite fires inside a 1.7 km radius around its famous cathedral, instantly killing about 30 000 people and injuring twice as many more[1]. Inside the blast zone, people would be left to fend for themselves, as the severe destruction of infrastructure, including one or perhaps two of Aachen's three hospitals, would hinder access for emergency response personnel, who would not be able to approach the center of the explosion without severe risks to their own lives. The injured and dying would be left without assistance or means of communication, cut off from the rest of world [3].

[1] *Casualties were estimated with* NUKEMAP *[2], an online tool for simulating and visualizing the destructive effects of nuclear weapons.*

The largest nuclear device currently deployed in a nuclear arsenal has an estimated yield of 5 Mt. Dropped on Berlin, one such device would kill approximately 1.4 million people; dropped on Paris, it would kill up to 3 million people, and in New Delhi 5 million human lives could be extinguished in little more than an instant.

There are nine nuclear weapons possessing states, with an estimated 12 241 warheads between them, of which 2100 were kept in a state of operational alert [4]. The lion's share of these weapons is possessed by the United States and Russia. A major nuclear war fought between these two nuclear behemoths would cause hundreds of millions of immediate deaths and potentially leave billions of people starving. But even a limited, "small-scale" nuclear war between India and Pakistan, with only 100 "Hiroshima-sized" nuclear explosions is likely to cause widespread fires, sending tonnes of ash and soot into the upper atmosphere, lowering global temperatures by several degrees for many years. This effect is called nuclear winter; scientists estimate it would lead to collapse of trade, a significant portion of the world population would starve [5]. Nuclear conflict would not be restricted to the nations that fight it, but cause people all across the globe to suffer and die.

In the minds of many people, the world is safer today because the prospective horrors of nuclear war and the notion of "mutually

assured destruction" prevent major global conflicts between superpowers. This belief follows the rationale that no rational actor would begin a conflict that risks nuclear war. What if those in control of the "Big Red Button" are not inclined to behave rationally?

1.2 VERIFYING FISSILE MATERIAL DECLARATIONS FOR NUCLEAR DISARMAMENT

Their immense destructive power and the conviction that their possession can deter a rival from aggressive military action, gives nuclear weapons an important role in some state's national security strategies. Due to the sometimes fragile relationships between rival states, nuclear non-proliferation, nuclear arms control and, especially, nuclear disarmament require mechanisms to facilitate cooperation and build confidence; they require means to verify that parties to an agreement are each upholding their end of the bargain [6].

If weapons stockpiles are close to or at zero, just one or a few devices could drastically alter geopolitical power distribution. Transparent verification becomes necessary to provide assurance that there are no undisclosed nuclear weapons [7]. Since the knowledge of how to build the weapons is impossible to eliminate, the materials from which they are made, that is highly-enriched uranium (HEU) and weapons-grade plutonium, need to be controlled. To be confident that a state does not secretly possess nuclear weapons, it would be important to be confident that all of a state's fissile material can be accounted for.

Therefore, the parties to the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) rely on the International Atomic Energy Agency (IAEA) to implement nuclear safeguards and verify that nuclear material in non-nuclear weapons states is not being used to build nuclear weapons [8]. Member states enter into an agreement with the IAEA that allows access to nuclear facilities and observe nuclear activities [9]. Each state regularly provides the agency with an account of its fissile material stockpiles, that the IAEA then needs to verify as a part of its responsibilities. Given the general success of the NPT regime, the use of fissile material declarations has also been advocated for in other nuclear arms control and disarmament contexts [6, 10]. The very first declaration, before the commencement of verification activities, is often called a *fissile material baseline declaration*.

In the NPT regime, the IAEA then needs to verify that the initial report (or baseline declaration) is correct and complete; verifying completeness is challenging if states have preexisting nuclear programs, and more so, if states have previously possessed nuclear weapons. Such was the case in South Africa, where the agency effectively spent years reconstructing the history of the fuel cycle to deduce that it was

highly unlikely that South Africa could have other, hidden stockpiles of material or nuclear weapons [11].

1.3 NUCLEAR ARCHAEOLOGY

The prospective application of reconstructing a state's historic production and usage of fissile material to facilitate nuclear disarmament and arms control is the driving motivation behind research in nuclear archaeology [12, 13]. This field of research is dedicated to developing technical methods to support related verification regimes. The majority of nuclear archaeology research focuses on using measurements of the isotopic composition of materials at fissile material production facilities as evidence for reconstructing how the facility was operated [14], although some recent studies have investigated methods to systematically analyze historical documentation of nuclear facilities to detect potential inconsistencies [15, 16].

The key facilities of interest are enrichment plants, nuclear reactors, and reprocessing facilities, where HEU and plutonium are produced.

To verify operational histories of enrichment plants, several different types of evidence have been investigated. The composition of the waste product of enrichment plants[2] [13, 17], the uranium deposition in separation membranes from diffusion plants [18] or in corrosion layers in the connecting pipes [19], and material defects caused by α-particles in the gaskets of gas centrifuges [20] have all shown promise as potential indicators for enrichment operations.

[2] *Waste products from uranium enrichment are called "tails".*

To reconstruct the past plutonium production of a nuclear reactor, the Graphite Isotope Ratio Method (GIRM) [21] was developed. This method relies on the fact that the changes to the isotopic composition of impurities, such as boron or titanium, in the permanent structures of a reactor core are intrinsically linked to the amount of plutonium that is produced in the fuel, by virtue of the neutron flux that permeates all materials in the reactor core and drives nuclear transmutation. Although GIRM was first developed for reactors with a solid graphite moderator, it has since been studied for use in heavy-water moderated reactors [22] and light-water reactors [23]. Research has also indicated that it could be possible to differentiate between different production modes, e. g., plutonium vs. tritium production [24]. This method is perhaps the most studied technique in the nuclear archaeology toolkit; yet open questions remain regarding its capabilities, limitations, and associated uncertainty assessments [25] and development efforts continue, applying novel computational inference techniques and searching for new, suitable isotopic signatures [26].

More generally, individual nuclear archaeology methods are still under development, requiring methodological refinement, systematic uncertainty assessments, and more experimental testing. Recently, an "integrated approach" has been advocated for, seeking to exploit syn-

ergies between the different nuclear archaeology methods [27], leading to the development of holistic reconstruction approaches that model the entire nuclear fuel cycle as one system to reconstruct HEU and plutonium production [28].

This quest for synergistic sources of evidence has also led to the development of a novel concept aiming to reconstruct the operating history of nuclear reactors from measurements of the composition of waste from reprocessing activities, where plutonium and uranium are separated from the irradiated nuclear fuel. Reconstructing the irradiation history of discharged nuclear fuel is a typical monitoring task performed by operators of nuclear power plants and for safeguards activities; targeting samples of reprocessing waste adds complexity to this task by the information about the reactor model that are necessary for accurate computational modelling and inference.

1.4 DEVELOPING A NUCLEAR ARCHAEOLOGY METHOD FOR ANALYZING REPROCESSING WASTE

A Bayesian inference framework was developed for this new concept by Antonio Figueroa [29] in his doctoral thesis. He showed that it is, in principle, possible to reconstruct two parameters, namely the *average burnup* of the fuel and the *cooling time* of a batch of reprocessed nuclear fuel from the composition of the waste product. He also demonstrated an inference model that can be used to reconstruct these parameters when the waste from two different batches of reprocessed fuel has been mixed in a sample. Groundbreaking and foundational, his research showed the potential of this idea and opened new research questions that warrant investigation in order to establish this concept as a new tool in the nuclear archaeology toolkit.

In this thesis, I will explore answers to some of these questions, intending to advance this method, which I call Bayesian Reprocessing Waste Analysis Method (BRAM), towards application readiness. The first question concerns the complexity of the inference scenario, seeking to reconstruct additional reactor operating parameters and study different reactor model types. This avenue of investigation includes the scenario that the reactor model may be unknown and needs to be inferred as well. The second question concerns the isotopic signatures in the reprocessing waste, aiming to develop sophisticated techniques to choose optimal signatures for various use cases. Third is the question of experimental validation, which I aim to advance using the limited amount of experimental data that is publicly available.

Thus, in the following chapters, I will first explain the theoretical and computational foundations of modelling the composition of irradiated nuclear reactor fuel and then describe the statistical inference framework that is needed to reconstruct the reactor operating parameters.

I will go on to discuss the isotopic signatures in the reprocessing waste, focusing on a new genetic algorithm for optimizing the choice of observed isotopic evidence. Then, I will present two simulation-based studies that systematically explore the performance of the framework with increased model complexity. The first focuses on mixture scenarios that include different reactor types and more reactor parameters, that is *initial enrichment* and *power density* in addition to *average burnup* and *cooling time*. The second explores the possibility of dealing with unknown reactor types, combining model classification with parameter reconstruction.

Afterward, I will show two validation case studies with experimental data from the Spent Fuel COMPOsition (SFCOMPO) database, before discussing a hypothetical application scenario to demonstrate the potential utility for verifying nuclear disarmament. Finally, I will discuss the implications of this research and outline avenues of investigation for future researchers.

Part II

A BAYESIAN FRAMEWORK FOR RECONSTRUCTING REACTOR OPERATING PARAMETERS

2

MODELLING THE COMPOSITION OF HIGH-LEVEL NUCLEAR WASTE

The goal of nuclear archaeology is to estimate the amount of fissile material that was produced nuclear facilities by reconstructing their operating histories. In the context of this thesis, the targeted production facilities are nuclear reactors, where plutonium is created as a product of nuclear reactions that transmute the uranium in the fuel. The intended source of evidence are the isotopic signatures in samples of waste from reprocessing activities, where the plutonium is separated from the irradiated nuclear fuel. To reconstruct information about past reactor operations, it is crucial to understand and be able to model the physical processes that relate the composition of the nuclear waste to the operating conditions in the reactor.

This chapter first explains fundamental concepts of nuclear physics and reactor engineering that are relevant to the production of plutonium. Then, it describes the mathematical models and computational approaches for simulating the interactions between neutrons and matter in a nuclear reactor and for predicting the changes of the isotopic composition of its fuel.

2.1 ON THE PHYSICS OF PLUTONIUM PRODUCTION

Plutonium is an element that, unlike uranium, does not occur naturally, except in trace amounts. Instead, it is primarily created from uranium by way of a neutron capture and two subsequent decay processes:[3]

$$^{238}_{92}\mathrm{U} + ^{1}_{0}\mathrm{n} \rightarrow ^{239}_{92}\mathrm{U} \xrightarrow{\beta^-} ^{239}_{93}\mathrm{Np} \xrightarrow{\beta^-} ^{239}_{94}\mathrm{Pu}. \tag{1}$$

To produce plutonium in significant amounts, large amounts of $^{238}\mathrm{U}$ must be exposed to a strong neutron flux. Such conditions are commonly created in a nuclear reactor, where a self-sustaining fission chain reaction generates neutrons that, in turn, interact with the surrounding matter, transmuting some of the uranium fuel. Thus, plutonium is created in the fuel of a nuclear reactor during its operation.

[3] *Other isotopes of plutonium, e. g.,* $^{240}\mathrm{Pu}$, $^{241}\mathrm{Pu}$, *or* $^{238}\mathrm{Pu}$, *can be produced by further neutron capture and decay reactions.*

2.1.1 *Nuclear Reactor Physics*

A nuclear reactor is a device that can be used to control a fission chain reaction. To sustain a stable fission chain reaction, each fission event must produce enough neutrons to induce on average one new fission event, a condition that is referred to as *criticality*. The criticality

is determined by the *effective multiplication factor* k_{eff}. There are many ways to define k_{eff}. The most readily interpretable definition is:

$$k_{eff} = \frac{\text{neutron production rate}}{\text{neutron absorption rate + neutron leakage rate}}. \quad (2)$$

Obviously, $k_{eff} = 1$ is the condition for criticality. If $k_{eff} < 1$, the system is *subcritical* and no chain reaction can be sustained. If $k_{eff} > 1$, the system is *supercritical* and the chain reaction grows exponentially, typically with harmful consequences.[4]

[4] *Precisely this effect is exploited in the design of nuclear weapons.*

The production and absorption rates are primarily determined by two factors: the *cross-sections* for interactions of neutrons with the nuclides in the system and the *energy spectrum* of the neutron flux. Typically, the interaction cross-sections are larger for slow neutrons, but many nuclei do not undergo fission unless a threshold energy is exceeded by the incident neutron. Although the fraction of fission reactions tends to increase for fast neutrons, slow neutrons produce more overall fissions due to the higher total cross-section. To practically fulfill the criticality condition, a system requires a sufficient amount of a fissile isotope, which is an isotope that can undergo fission even if the absorbed neutron is slow.

The most common fissile isotope is ^{235}U, which is present in natural uranium at a concentration of about 0.711 %-wt. Through a process called *enrichment*, the concentration of ^{235}U in uranium can be increased, which is useful for increasing the macroscopic cross-section for fission in nuclear fuel and can compensate for higher neutron loss rates due to other design choices.[5]

[5] *And which is necessary for some reactor designs to work*

There are many different concepts and designs for nuclear reactors. A fundamental characteristic of a reactor is the shape of its *neutron energy spectrum*. The neutrons produced in fission reactions are considered to be *fast* neutrons ($E \propto 1\,MeV$). Their energy spectrum peaks around 1 MeV [30]. Reactors that use these neutrons are typically designated as *fast neutron reactors*.

On the other hand, *thermal reactors* exploit the higher fission cross-sections of ^{235}U at thermal energies ($E \propto 0.025\,eV$). In such designs, fast fission neutrons are slowed down through scattering reactions that occur in a so called *moderator*, that surrounds and separates fuel elements. The moderator is typically a material with a low atomic mass, such as light-water, heavy-water[6], or graphite. Depending on the moderator material, thermal reactors use fuel with 3–5 % ^{235}U or even natural uranium.

[6] *Heavy-water is water primarily composed of D_2O.*

The enrichment of the fuel also affects the fission rate that can be achieved in the reactor, which determines the rate of energy release in the fuel, that is, the reactor's thermal power output. A single fission event releases on average 200 MeV of energy, while a modern nuclear reactor can produce up to $1000\,MW_t$ of thermal power. While the the thermal power output primarily depends on the size of (i. e., the amount of fuel in) the core, the power density , given in W/gHM,

measures the amount of thermal power produced per unit mass of heavy metal (i. e., uranium) in the fuel and is a useful indicator of the fission rate in the reactor.

Over the course of being irradiated, the fission reactions occur in the fuel and energy is released. Consequently, the number of fissile isotopes decreases[7] and its capacity to fission diminishes. This effect is quantified by the *burnup*. It is calculated by integrating the thermal power density p over time:

$$\mathrm{BU} = \int_{t_0}^{t_1} p(t)\,\mathrm{d}t. \tag{3}$$

Consequently, the unit of burnup[8] is MWd/kgHM. It describes the total amount energy that has been released in the fuel and is used to determine how much energy can still be released. The maximum attainable burnup depends on the reactor design and the enrichment level of the fuel and is important for the fuel economy of a reactor. Once nuclear fuel has been irradiated to a burnup where it can no longer sustain a fission chain reaction, it is discharged from the reactor core. Such fuel is called spent nuclear fuel (SNF).

Since plutonium is created as a consequence of neutron capture reactions, the amount of plutonium in the fuel increases steadily[9] with the received irradiation, and thus with the burnup. Initially, only ^{239}Pu is produced, but as the burnup increases, the amount of ^{240}Pu also increases. In plutonium intended for use in nuclear weapons, the concentration of ^{239}Pu should be high, while the concentration of ^{240}Pu should be low. A common definition for *weapons-grade* plutonium is plutonium containing at least 93 % ^{239}Pu [31], although it is generally accepted knowledge that nuclear explosive devices can be constructed with plutonium with much higher concentrations of ^{240}Pu [32]. Typically, the limit of weapons-grade plutonium is exceeded at very low ($\lesssim 1$ MWd/kgHM) burnup levels. Therefore, plutonium production reactors discharge fuel before it reaches its maximum burnup to optimize the plutonium isotope vector for military use.

Irradiating nuclear fuel also produces a plethora of other elements and isotopes besides plutonium. They are created by a variety of nuclear reactions, such as *induced fission* or *radiative neutron capture*, and by *nuclear decay*. The nuclides created by these processes can be loosely grouped into three respective categories: *fission products*, *actinides*[10] and other *activation products*, and *decay products*. Many of these products are highly radioactive, harmful to human health, and produce significant amounts of thermal energy even without being irradiated by neutrons. Therefore, SNF needs to be cooled (usually in a pool of water) for some time before it can be repurposed or disposed of.

[7] *At least the amount of ^{235}U decreases. ^{239}Pu grows at first an can also fission.*

[8] *To be precise, this quantity is called the* specific burnup. *The burnup is technically defined as just the integral of the thermal power over time. However, the specific burnup is much more frequently used and often simply referred to as the burnup.*

[9] *If fuel receives a sufficient amount of irradiation, the composition of the fuel can change such that the loss rate of ^{239}Pu (being a fissile nuclide) exceeds its production rate and the amount of ^{239}Pu in the fuel diminishes again.*

[10] *Although fission and decay products can also be transmuted by radiative neutron capture, the trans-uranic actinides, especially, are produced by this reaction.*

2.1.2 *Nuclear Reprocessing*

After SNF, or *irradiated* nuclear fuel, is removed from the reactor, it is typically stored in a cooling pool for some time. Then, it can be *reprocessed* to extract the uranium and the plutonium. A number of chemical processes can be used to separate uranium and plutonium from the other elements contained in irradiated nuclear fuel. The most common, and de facto standard, reprocessing technique is the Plutonium Uranium Reduction EXtraction (PUREX) process [33, 34].

In concept, the PUREX process consists of two steps. First, the uranium and plutonium are separated from the other elements. Then, the plutonium is separated from the uranium.

First, spent fuel rods are mechanically cut into short pieces, which are then heated to remove any radioactive gases, such as tritium and krypton. Then the pieces of fuel are dissolved in concentrated nitric acid (HNO_3). During this step, the uranium and plutonium become highly oxidized.

In the next step, an organic solvent—30 % tributyl phosphate in kerosene or dodecane—is added to the solution to transfer the uranium and plutonium into an organic phase, while the remaining actinides and fission products remain in the nitric phase. Then, the organic solution and the aqueous nitric solution are separated from each other. The former contains almost all of the uranium and plutonium and only trace amounts of other elements, and the latter contains all the other material and barely any uranium or plutonium.

Next, a solution of a chemical-reducing agent, e. g., a ferrous compound, is added to solution of uranium and plutonium. This agent reduces the plutonium to a lower (3^+) oxidation state, but leaves the uranium in its higher (6^+) oxidation state. Thus, the plutonium is no longer soluble in the tributyl phosphate and can be removed via the aqueous solution of nitric acid.

Finally, the uranium is transferred from the organic solution back into an aqueous solution. Some additional processing steps can be used to further purify the uranium and plutonium solutions.[11]

Thus, the reprocessing ends with three products: uranium nitrate, plutonium nitrate, and a solution of nitric acid containing all the other elements is a highly active raffinate (HAR), which is a waste product of the process. It is often called PUREX raffinate, or more generally, high-level waste (HLW) or *reprocessing waste*.

[11] *If the process is performed efficiently, more than 99.9 % of the uranium and plutonium can be recovered [35].*

2.2 MODELLING NUCLEAR REACTOR PHYSICS

The growth rate of ^{239}Pu can be approximated very roughly by its rate of production through neutron capture of ^{238}U minus its rate of decay:

$$\frac{dN_{^{239}Pu}}{dt} = \phi \cdot \sigma_{^{238}U,\gamma} N_{^{238}U} - \lambda_{^{239}Pu} N_{^{239}Pu}. \quad (4)$$

Here, the production rate is approximated as the product of the neutron flux ϕ, the cross-section for radiative neutron capture $\sigma_{^{238}U,\gamma}$ by ^{238}U, and the number of ^{238}U atoms.

A simple approximation of the neutron flux ϕ in a nuclear reactor can be obtained by relating the thermal power P to the cross-section for fissions $\sigma_{^{235}U,fis} \cdot N_{^{235}U}$ and the average energy released per fission E_{fis}:

$$\phi = \frac{P}{E_{fis} \cdot \sigma_{^{235}U,fis} \cdot N_{^{235}U}} \quad (5)$$

To obtain precise predictions of the behavior of a nuclear reactor, a vast number of coupled physical processes must be modeled. Sophisticated modelling techniques include multi-physics approaches to couple the neutronics[12] and the thermo-hydraulic behavior of the reactor core.

In this work, the primary interest is the isotopic composition of reprocessing waste, which is a derivate of the SNF. To predict the isotopic composition of nuclear fuel, the physical processes that transmute its components need to be modelled. A well established approach is to link Monte Carlo neutron transport simulations with depletion calculations in a coupled sequence to approximate a solution to the governing differential equations.

I used the open-source Monte Carlo neutron and photon transport simulation code `OpenMC` [36]. This software also has depletion capabilities [37] and, thus, serves for modelling the time evolution of the nuclear fuel composition. Some of the models developed during earlier stages of this research use the depletion software `ONIX` [38]. [13] The description of modeling approaches in the following sections focuses on the implementation in `OpenMC`. Other, similar, software tools may use slightly different approaches, although the general concepts remain valid regardless.

[12] Neutronics *refers to the modeling of the movement and interactions of neutrons in matter.*

[13] *Since `ONIX` is not actively maintained, I transitioned to using the more recently developed depletion capabilities of `OpenMC`.*

2.2.1 *Depletion Calculations*

The Bateman equations describe the time evolution of the isotopic composition of nuclear fuel [39]. Neglecting any spatial movement of the nuclides, the number density N of a nuclide i at a specific position $\vec{r}$ in the reactor core is:

$$\begin{aligned}\frac{dN_i(t)}{dt} &= \sum_{j\neq i}\left(\int_0^\infty a_{j\to i}(E,t)\,\sigma_{a_j}(E,t)\,\phi(E,t)\,dE\right)N_j(t)\\ &\quad+\sum_{k\neq i} b_{k\to i}\lambda_k N_k(t)\\ &\quad-\left(\int_0^\infty \sigma_{a_i}(E,t)\,\phi(E,t)\,dE+\lambda_i\right)N_i(t).\end{aligned} \tag{6}$$

The first two terms on the right-hand side describe the production of nuclide i, either by the decay of another nuclide k or by the interaction with a neutron of a nuclide j. The third term describes the destruction of nuclide i and comprising nuclear decay and neutron absorption.

In this equation, $\phi(E,t)$ is the scalar neutron flux energy spectrum, describing the number of neutrons moving through a unit area per unit of time at the position $\vec{r}$ and as a function of neutron energy. The energy-dependent cross-section for any neutron-absorbing reaction of either nuclide i or nuclide j is denoted by $\sigma_{a_{i/j}}(E,t)$. Generally, physical constants such as microscopic cross-sections are independent of time. However, in Equation (6) the time-dependency of these constants reflects their dependency on the temperature of the material, which is not a variable of the equation, but can nevertheless change with time.

The most common reactions are fission (z,fis) and radiative capture (z,γ), but other reactions can also occur, such as (z,α), which produces a helium nucleus. In the case of fission reactions, the branching ratio $a_{j\to i}$ describes the energy dependent probability that a fission neutron will produce a nuclide i. The nuclear decay constant is denoted with $\lambda_{k/i}$, with the branching ratio $b_{k\to i}$ describing the probability that a decay of nuclide k will produce a nuclide i.

While not all nuclides can be produced by all other nuclides, it is clear from Equation (6) that the Bateman equations form a system of coupled differential equations. The added challenge to solving this system of equations is the fact that the neutron flux ϕ depends on the number densities of the nuclides, because the nuclear reactions absorb and produce neutrons. This dependency will be addressed in the next section.

Assuming that the neutron flux is known and fixed in time, the Bateman equations can be written in matrix form as:

$$\frac{d\vec{N}(t)}{dt} = \mathbf{A}(t)\,\vec{N}(t). \tag{7}$$

$\mathbf{A}$ is called the *burnup* or *transition* matrix; each of its elements $a_{ij}(t)$ corresponds to the terms in Equation (6) that create or destroy a nuclide. Accordingly, the solution is:

$$\vec{N}(t) = \exp\left(\int_{t_0}^{t}\mathbf{A}(t')\,dt'\right)\vec{N}_{t_0}. \tag{8}$$

To obtain a practical solution, it is useful to discretize time into steps where the $\mathbf{A}$ is approximately constant, which allows the solution to be written as:

$$\vec{N}(t_n) = \exp\left(\int_{t_{n-1}}^{t_n} A_{t_{n-1}} \, dt'\right) \vec{N}(t_{n-1}) = \exp\left(A_{t_{n-1}} \Delta t\right) \vec{N}(t_{n-1}) \tag{9}$$

Then, computing the solution comes down to computing the matrix exponential of $\mathbf{A}\Delta t$, which can be achieved with any of a number of approaches [40]. A particular challenge to many of these methods is the *stiffness* of the matrix. The transition matrix is considered stiff, because its elements vary across several orders of magnitude. For example, the half-lives of the various nuclides can range from fractions of milliseconds to millions of years. `OpenMC` uses the *Chebyshev Rational Approximation Method* [41], which is the current state-of-the-art in reactor modeling.

The discretization introduced in Equation (9) is especially convenient, as it implies that the assumption that the neutron flux is known and fixed only needs to hold for one timestep. The following section explains how the neutronics parameters for one such timestep can be calculated.

2.2.2 *Neutron Transport*

The transmutation of nuclei in the reactor depends on the distribution of the neutrons that permeate it. The neutron population in a physical system is given by the *angular neutron density* $N(\vec{r}, \vec{\Omega}, E, t)$, which specifies the number of neutrons at a point in space $\vec{r}$, with energy E, traveling in the direction $\vec{\Omega}$, at a given time t. N is a density on the 6-dimensional phase space $(\vec{r}, \vec{\Omega}, E)$ with the units $\mathrm{cm}^{-3}\,\mathrm{MeV}^{-1}$.

To model the interactions of neutrons with atomic nuclei, the *angular neutron flux* $\phi(\vec{r}, \vec{\Omega}, E, t)$ is a slightly more useful quantity. It is defined as the product of the neutron velocity $v = \sqrt{2E/m}$ and the angular density N:

$$\phi\left(\vec{r}, \vec{\Omega}, E, t\right) = vN\left(\vec{r}, \vec{\Omega}, E, t\right) \tag{10}$$

and has the units $\mathrm{cm}^{-2}\,\mathrm{MeV}^{-1}\,\mathrm{s}^{-1}$. The angular neutron flux can be interpreted as the number of neutrons that travel through a point in the phase space in a given direction *per unit of time.* This definition differs slightly from the neutron flux in Equation (6) in that it still distinguishes the directions of travel explicitly.

2.2.2.1 *Neutron Transport Equation*

The propagation of neutrons through a physical system is described by the *neutron transport equation*, which is sometimes called the *linear Boltzmann equation*. It includes the movement of neutrons through

space, the scattering of neutrons off of atomic nuclei, the absorption of neutrons by nuclei, and the expulsion of neutrons by fission events.

Like the Bateman equations, the neutron transport can be written as a conservation equation, meaning that the rate of change is equal to the rate of gain minus the rate of loss. The rate of loss is given by the collision rate and the net leakage rate. The rate of gain comprises the "in-scattering" rate, the fission rate, and the source rate. Here, it is useful to distinguish between the *prompt fission neutrons* and the *delayed fission neutrons*. The former describes neutrons that are produced immediately during a fission event, while the latter describes neutrons that are produced by the decay of the radioactive fission products.

The following terms contribute to neutron transport equation, which is given further below in Equation (11). The notation $\phi(\cdot) = \phi(\vec{r}, \vec{\Omega}, E, t)$ and $\phi(\cdot') = \phi(\vec{r}, \vec{\Omega}', E', t)$ is used to distinguish between outgoing and incoming neutrons, respectively.

THE NET LEAKAGE RATE is given by the projection of the gradient of the neutron flux into the direction of travel: $\vec{\Omega} \cdot \nabla\phi(\cdot)$.

THE COLLISION TERM is product of the total interaction cross-section and the neutron flux: $\Sigma_t \phi(\cdot)$.

THE "IN-SCATTERING" RATE is obtained by integrating the product of scattering cross-section $\Sigma_s(\vec{\Omega}' \cdot \vec{\Omega}, E' \to E)$ and neutron flux over all incoming directions of travel Ω' and incoming energies E'.

THE PROMPT FISSION RATE is governed by the fission cross-section $\Sigma_f(E')$,[14] the average number of neutrons released per fission ν, and the fraction of delayed neutrons $\beta(E')$. $\chi_p(E)$ represents the energy spectrum of prompt fission neutrons.

[14] *The average energy of prompt fission neutrons is* 1.98 MeV *[33].*

THE DELAYED FISSION RATE depends on the amount of unstable fission products, which are described as *precursor densities* $C_j(\vec{r}, t)$.[15] It is common practice to divide the precursor nuclides into six groups, each with its own decay constant λ_j and neutron energy spectrum $\chi_j(E)$.

[15] *The half-lives of the six precursor groups range from* 0.23 s *to* 55.72 s *[33].*

THE SOURCE TERM $Q(\vec{r}, E, t)$ describes any other internal or external sources of neutrons.

$$\begin{aligned}
&\frac{1}{\nu}\frac{\partial\phi}{\partial t}\left(\vec{r},\vec{\Omega},E,t\right)+\vec{\Omega}\cdot\nabla\phi\left(\cdot\right)+\Sigma_t\left(E\right)\phi\left(\cdot\right)\\
&\quad=\int_0^\infty\int_{4\pi}\Sigma_s\left(\vec{\Omega}'\cdot\vec{\Omega},E'\to E\right)\phi\left(\cdot'\right)d\Omega'dE'\\
&\qquad+\frac{\chi_p\left(E\right)}{4\pi}\int_0^\infty\int_{4\pi}\left[1-\beta\left(E'\right)\right]\nu\Sigma_f\left(E'\right)\phi\left(\cdot'\right)d\Omega'dE'\\
&\qquad\quad+\frac{1}{4\pi}\sum_{j=1}^{6}\chi_j\left(E\right)\lambda_j C_j\left(\vec{r},t\right)+\frac{1}{4\pi}Q\left(\vec{r},E,t\right)
\end{aligned}\tag{11}$$

To account for the delayed fission neutrons, the neutron transport equation is coupled to the equations that describe the change of the precursor densities C_j (see Equation (12)). Here, the loss term is simply the decay rate $\lambda_j C_j$ and the production term is given by the fission rate $\Sigma_f\phi$ multiplied by the number of delayed neutrons $\beta_j\nu$ per fission.

$$\frac{\partial C_j}{\partial t}\left(\vec{r},t\right)+\lambda_j C_j=\int_0^\infty\int_{4\pi}\beta_j\left(E'\right)\nu\Sigma_f\left(E'\right)\phi\left(\cdot\right)d\Omega'dE'\tag{12}$$

In notion, the angular neutron flux ϕ is obtained by solving Equations (11) and (12) with some boundary conditions $\phi(\vec{r}\in\partial V,\cdot)=\phi^b(\cdot)$ and some initial conditions $\phi(\cdot,t{=}0)=\phi^i(\cdot)$ and $C_j(\vec{r},t{=}0)=C_j^i(\vec{r})$. However, most neutron transport problems of practical interest are too complex to be solved analytically and require either approximations or numerical solutions.

2.2.2.2 *The Steady-State Simplification*

For practical modeling purposes, it is useful to first consider the steady-state neutron transport equation. Neglecting external sources, the steady-state neutron equation can be rewritten as an *eigenproblem*, with the *eigenvalue* k_{eff} and the *eigenvector* ϕ (see Hébert [42]). The solution with the maximum possible value of k_{eff} is the solution that has physical meaning with a flux distribution $\phi(\vec{r},E)$ that is almost positive everywhere on the domain. Importantly, this flux distribution can be arbitrarily normalized. A physically meaningful normalization can be derived from the thermal power of the reactor being modeled.

2.2.2.3 *Monte Carlo Neutron Transport*

Current state-of-the art neutron transport codes use the Monte Carlo method to approximate solutions to the neutron transport equation. Trajectory parameters, such as the distance traveled before a collision or the scattering angle after a collision, are randomly sampled from the respective cumulative distribution functions. Reaction types are randomly selected according to their relative probabilities. A specific

number of *neutron histories*, i. e., set of events that a neutron undergoes from its creation to its destruction, are simulated and used to calculate relevant neutronics parameters, such as the reaction rates R_i of each type of reaction with each nuclide.

In the neutron transport code `OpenMC`, a transport simulation proceeds according to the following steps [43] to simulate a single neutron history.

1. The particles properties are initialized depending on a source distribution.
2. The energy-dependent cross-sections of the material located at the particles coordinates are determined.
3. The distance to the next interaction is sampled as $d = -\frac{\xi}{\Sigma_t}$, with the total material cross-section Σ_t and a pseudo-randomly sampled number $\xi \sim \mathcal{U}[0, 1)$.
4. If the distance to the next interaction is larger than the distance to the nearest boundary with a different material, the particle is moved and step 2 is repeated.
5. Otherwise, the nuclide with which the particle interacts is sampled with probability $P(i) = \frac{\Sigma_{t,i}}{\Sigma_t}$. Here, $\Sigma_{t,i}$ are the total cross-sections of the individual nuclides in the material.
6. Then the reaction type is randomly sampled based on the microscopic cross-sections. The probability of a reaction x is $P(x) = \frac{\sigma_x}{\sigma_t}$.
7. If the sampled reaction is a scattering reaction, the outgoing energy and angle are sampled from the corresponding distributions. If the sampled reaction is absorption, the particle is removed. If the reaction is fission, the particle is removed and fission sites are created and stored in the fission bank. The fission bank is used to determine source sites for the next generation of particles.

These steps are repeated for a fixed number of N particles (i. e., neutrons) to obtain an estimate of physical quantities of interest.[16] The process is repeated for a fixed number of *batches* or *generations* to determine a statistical estimate for the given quantities.

Such a calculation inherently neglects any time dependency of the equation and is restricted to solving the steady-state transport equation. Although fixed sources of neutrons can be modeled with the above approach, the interesting problem for modeling nuclear reactors is the eigenproblem described in Section 2.2.2.2.[17] In this case, the source distribution is not known precisely at the start of the calculation. Modeling the system for multiple successive generations, allows the method to converge to the true neutron source distribution, since

[16] *This estimate is called* tally*, as it is estimated by counting certain events in the simulation, such as neutron-nucelus collisions.*

[17] *Simulations that model a system with fissile isotopes and no fixed source distribution are, therefore, commonly called* k-eigenvalue *calculations.*

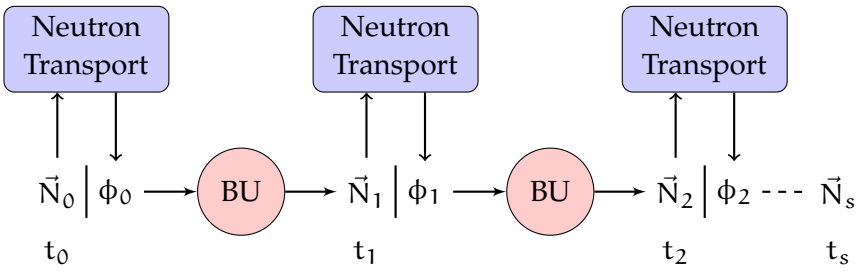


Figure 1: The sequential approach to Monte Carlo reactor simulations: Starting with a given material composition $\vec{N}_0$, the neutron transport module calculates the neutron flux. Then the burnup module (BU) calculates the changes in the material composition. This process is repeated for the given number of steps. (Graphic reproduced from [44].)

the source sites of each generation is derived from the fission sites of the previous generation. Therefore, it is recommended to ignore the first couple of generations, as their source distribution has not yet converged[18].

Given a method to solve the steady-stage neutron flux, the next step is to introduce time-dependency into the model.

[18] *There are methods to determine how many batches should be skipped [43], but I will not discuss them here.*

2.2.3 *Modelling the time evolution of nuclear fuel*

The depletion equation (6) and the neutron transport equation (11) are coupled by way of the reaction rates R_i of the neutrons with the various nuclides in the material. To overcome this problem, modern simulation approaches solve the two equations in a sequence of steps where the results of the previous neutron transport simulation are used to solve the next depletion calculation and vice versa. This approach is illustrated in Figure 1. During the neutron transport simulation, the material composition is unchanging and for the depletion calculation, the neutron flux is assumed constant. By choosing a suitable partitioning of the time steps, good approximations of the fuel depletion can be obtained.

There are several different methods, called *coupling schemes*, to link the neutron transport simulation and the depletion calculation. The most straight-forward coupling scheme is simply the *Predictor* method, which is the one depicted in Figure 1. The neutron flux is simulated with the material composition at t_0. Then, this neutron flux is used to calculate the fuel depletion over a time interval Δt to determine the nuclide vector at $t_1 = t_0 + \Delta t$. Then, this nuclide vector is used to compute the neutronics parameters for the next time step.

Other coupling schemes use more than one neutron transport simulation per depletion time step to determine the neutronics parameters. For example, the *Predictor-Corrector* concept combines beginning-of-step and end-of-step neutronics parameters to re-calculate the fuel

depletion. The new result then serves as the beginning-of-step nuclide vector for the next time step in the sequence. Other, more sophisticated methods exist as well and can be reviewed in [37].

2.2.4 *Modeling Reactor Geometry and Approximations*

In the previous sections, the spatial inhomogeneity of the physical system that is a nuclear reactor core has been implied. To model the physical processes accurately, the spatial distribution of the different materials in the reactor needs to be modeled. In `OpenMC` [19], the modeled system is created by defining geometric surfaces in a three-dimensional Euclidian space. Each surface divides space into two *half-spaces*. Arbitrarily complex shapes can be defined as unions and/or intersections of such half-spaces. These geometric shapes divide the space into cells, each of which is defined with an associated material that specifies physical quantities, such as the nuclide composition, density, and temperature.

[19] *Other codes follow similar principles, although some of the details may vary.*

To keep track of the neutrons as they traverse the system, each surface is defined with a boundary type that specifies what happens if a neutron hits the surface. Common boundary types are *transmissive*, *reflective*, and *vacuum*. The first allows neutrons to pass into the adjacent geometric region, while the last removes the neutron from the system. The reflective boundary reflects the neutrons back into the region with a new direction of travel, which depends on the shape of the surface.

Another useful boundary type is the *periodic* boundary. A surface with a periodic boundary has an associated partner. When a neutron hits the surface, it is translocated to the partner surface as if it had entered the region through that surface. This surface type is especially useful to model rotationally symmetric systems.

In general, it is common practice to exploit symmetries in the physical system to reduce the size of a model and the associated computational cost of the simulation. For example, it is often useful to model only a quarter of a reactor core and make use of reflective or periodic boundaries at the appropriate surfaces.

Sometimes, it is also sufficient to neglect the axial inhomogeneity of the system and model it as a two-dimensional plane. This approach neglects boundary effects at the top and bottom of the reactor core, but can reduce computational cost significantly.

2.2.4.1 *Infinite lattice approximation*

Another frequently used approach to simplify a reactor model is the *infinite lattice approximation*. This approach models a small sub-region of the reactor core, whose structure is repeated regularly throughout the core with a lattice-like pattern. Such a region is called a *unit cell*.

Examples are a single fuel channel, one fuel assembly, or a small $N \times N$ lattice of fuel channels.

The unit cell is repeated infinitely in each direction perpendicular to the fuel rods through the application of periodic or reflective boundary conditions. The axial direction of such a model can be finite, bounded by materials that represent the bottom and top of the reactor core. To further simplify such a model, the axial extent of the model can be reduced, and the axial boundary conditions can also be made reflective. This is a 2D infinite lattice model. It approximately reflects the conditions near the center of the reactor core, where neutron losses through leakage are minimal.

2.2.5 *Nuclear Data Libraries*

Modeling the cornucopia of nuclear reactions requires an extensive amount of data to describe the occurence-probability and products of each reaction [45]. The backbone of the required data are the energy dependent cross-sections for neutron nucleus interactions. These are carefully curated in evaluated nuclear data libraries that combine experimental data with theoretical models to provide cross-sections for around 400 atomic nuclei at incident neutron energies from 10^{-11} to 20 MeV. The most important data are the absorption, fission, and scattering cross-sections. The primary absorption reaction is the radiative neutron capture (n, γ), but other reactions, such as (n, p) and (n, α), are also considered. Some nuclides require special thermal scattering data that reflect their binding states in certain molecules, for example, H_2O.

Other relevant data are the neutron yield, the fission product yield, and the radioactive decay data. The neutron yield data describes the number of prompt and delayed neutrons produced by a fission reaction. The fission product yield describes the likelihood of producing given fission products, and the decay data comprises decay constants and branching ratios.

After a complex evaluation process, such libraries are validated against benchmark experiments. The largest evaluated neutron data libraries are developed and maintained by the major nuclear energy-producing countries or regions: ENDF/B (USA), JEFF (Europe), JENDL (Japan), BROND (Russia), and CENDL China.

The nuclear data used in this work is provided by the developers of `OpenMC` in a ready-to-use format. The source of the data is the Evaluated Nuclear Data File (ENDF)/B-VII.1 [46] library.

2.2.6 *Using OpenMC*

To run a neutron transport simulation with `OpenMC`, the user needs to specify which materials are in the modeled system, how these mate-

rials occupy space, and which settings to use for simulating particles. Materials are defined by their nuclide composition and density and temperature. They also need to be associated with a nuclear cross-section data library. The geometry is defined as cells that occupy regions in space and are filled with materials (see Section 2.2.4). The settings can be used to modify many parameters of the simulation algorithm, most importantly the number of particles per batch and the number of active and inactive batches that are modeled.

While the backbone of `OpenMC` is written in `C++`, there is a practical and powerful `Python API` that can be used to generate the `XML`-structured input files.

The depletion module is implemented fully in `Python` and comprises algorithms to compute the matrix exponential (see Section 2.2.1) and for coupling the neutron transport and fuel depletion calculations (see Section 2.2.3). Important, user-specifiable variables are the time steps and the thermal power output of the system. Additionally, any depletion specific nuclear data, such as decay constants, need to me specified.

3

THE BAYESIAN REPROCESSING WASTE ANALYSIS METHOD

Nuclear archaeology is often conceived of as an inverse problem. A given configuration of variables, i. e., parameters pertaining to the operation of a nuclear reactor, generates, through the underlying physical processes, some observable quantities, i. e., the composition of the nuclear waste. The goal is to infer the configuration of variables from observations of these quantities.

The previous chapter introduced a computational approach to modelling the physical processes in a reactor and predict the isotopic composition of spent nuclear fuel (SNF). In other words, it presented a "forward" model. This chapter presents the theoretical foundation of the Bayesian Reprocessing Waste Analysis Method (BRAM), a framework for inverting such a model to solve the inverse problem that is the topic of this thesis. First, I introduce the fundamental concepts of Bayesian model inversion. Then, I delineate some Markov Chain Monte Carlo (MCMC) sampling algorithms that are useful for solving Bayesian inference problems. Third, I explain how the reactor operating parameters are fused with the Bayesian framework, and finally, I discuss Gaussian-process based surrogate modelling, a necessary ingredient to make the whole method computationally feasible.

3.1 MODEL INVERSION WITH BAYESIAN INFERENCE

Given a model that predicts some observable data, the *inverse problem* consists of adjusting the model's input parameters to bring its predictions into agreement with the observations. Bayesian model inversion conceptualizes this inverse problem as a statistical inference problem, that is, a *Bayesian* inference problem, in which the model inputs are treated as random variables and the information about these variables is encoded in probability distributions. The solution to the inverse problem is the posterior probability distribution of the input parameters conditional on the measurements of the observable quantities.

Unless otherwise noted, the following explanations are based on Gelman et al. [47] and Wagner et al. [48].

3.1.1 *Bayesian inference*

Given some non-observable parameters of interest $\theta \in \mathcal{D}_\theta$, statistical inference aims to draw conclusions about θ using information from

observations y of some observable quantities $Y \in \mathcal{D}_Y$ that are related to θ, either directly or indirectly. *Bayesian inference* approaches this problem through the lens of Bayes' theorem of conditional probability, offering the *posterior probability distribution* $P(\theta|y)$ as a solution:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}. \tag{13}$$

The *prior probability distribution* $P(\theta)$ describes the information available about θ before the data y is observed, and the *likelihood function* $P(y|\theta)$ establishes the connection between the data y and the parameters θ. It can be understood as the "likelihood" of observing the data y given values of θ, hence its name. The likelihood function can be expressed as:

$$\mathcal{L}(\theta, y) := P(y|\theta) = \prod_{i=1}^{N} P(y_i|\theta), \tag{14}$$

given the common assumption that y is a set of N independent measurements $\{y_1, \ldots, y_N\}$ of N realizations of Y. For a fixed y, $P(y)$ is independent of θ and is, therefore, commonly treated as a normalization constant, which leads to the *unnormalized posterior distribution*:

$$P(\theta|y) \propto \mathcal{L}(\theta, y) \cdot P(\theta). \tag{15}$$

Although there are examples of analytical solutions to computing the posterior distribution, the posterior is often analytically intractable, especially for complex models with many parameters [47]. In these cases, numerical algorithms are necessary to approximate the posterior distribution appropriately.

One, widely used, approach is to use Monte Carlo-based algorithms to generate samples from the unnormalized posterior distribution (see Section 3.2) and compute the normalization from the generated samples.

3.1.2 *Model inversion*

Applying the Bayesian inference framework to the model inversion problem requires two ingredients: prior distributions $P(\theta)$ of the model input parameters and a likelihood function $\mathcal{L}$. The prior distributions are defined according to the available information, such as physical constraints, expert knowledge, or previous inference iterations.

Given a model $M(\theta)$ that predicts values of the observable quantity Y, the likelihood function describes the difference between the model predictions and the observations y_i of the observable Y. With an additive discrepancy model, the difference ϵ, defined as:

$$\epsilon = y_i - M(\theta), \tag{16}$$

is normally distributed $\epsilon \sim \mathcal{N}(0, \Sigma)$. The covariance Σ typically describes both the *measurement error* and the *modeling error*. With this discrepancy model, the conditional probability of observing y_i, given θ and Σ, can be written as:

$$P(y_i|\theta) = \mathcal{N}(y_i|M(\theta), \Sigma), \tag{17}$$

which can be substituted into Equation (14) to obtain the likelihood function. Thus, the posterior distribution of the parameters θ of a model $M(\theta)$ that predicts the observables Y, conditioned on measurements y_i of Y, is given by:

$$P(\theta|y) = \frac{P(\theta) \cdot \prod_{i=1}^{N} \mathcal{N}(y_i|M(\theta), \Sigma)}{P(y)}, \tag{18}$$

where the prior $P(\theta)$ is chosen by the user and $P(y)$ is treated as a normalization constant to be calculated from the numerically generated samples of the unnormalized posterior density.

3.1.3 *Comparing models*

It is not uncommon to have several options to model the likelihood, and sometimes the purpose of an analysis can even be to determine which model is more suitable to describe the observations. In a Bayesian analysis with several competing models, the Expected Log Pointwise Predictive Density (elpd) can be computed for each model and used to rank them according to their performance to choose the best model. It is also possible to compute weights for each model and use a weighted average of all models as the result of the analysis.

Vehtari et al. [49], and Yao et al. [50] offer some recommendations in this regard, which were followed in this work, using the implementation provided in the `ArviZ` library [51] for `Python`. In the following, the specific methods used in this work are outlined briefly and curious readers are encouraged to consult the above publications for more detailed explanations.

The posterior predictive accuracy describes how well a conditioned Bayesian model predicts new, "unseen" datapoints. A measure of this accuracy can be defined as the Expected Log Pointwise Predictive Density (elpd):

$$\text{elpd} = \sum_{i=1}^{n} \int P_t(\tilde{y}_i) \log\left(P(\tilde{y}_i|y)\right) d\tilde{y}_i, \tag{19}$$

with the "true" distribution of the data $P_t(\tilde{y}_i)$ and the posterior predictive density $P(\tilde{y}_i|y)$ of the conditioned model. Since the true distribution P_t is generally not known, the elpd must be approximated. The recommended approach [49] to approximate this quantity is to use Leave-One-Out Cross-Validation (LOO-CV), approximated by Pareto

smoothed importance sampling (PSIS). This combination of methods offers a computationally efficient estimate of the elpd, which, when computed for each model, can be used to rank them.

The *stacked* posterior predictive density distribution $\hat{P}(\tilde{y}|y)$ [50] is:

$$\hat{P}(\tilde{y}|y) = \sum_{k=1}^{n} \hat{\omega}_k P(\tilde{y}|y, M_k), \tag{20}$$

with the weights $\hat{\omega}_k$ for n Bayesian models M_k. These weights are computed with a PSIS-LOO-CV estimate of the respective posterior densities. The stacked posterior predictive distribution can be used to obtain "model-averaged" predictions of new datapoints.

3.2 MARKOV CHAIN MONTE CARLO SAMPLING

The Bayesian inference problems in this thesis, and many Bayesian inference problems in general, do not offer analytical solutions, but require numerical methods for finding the posterior distributions. A powerful way of approximating the posterior distribution is to use an algorithm to generate random samples from it. The samples are then used to characterize the distribution. Such an approach resembles a typical Monte Carlo simulation.

The expectation of a function $f(\theta)$ with respect to the posterior distribution $P(\theta|y)$ can be approximated by the sample mean of the function evaluated at the samples θ_i:

$$\mathbb{E}\left(f\left(\theta\right)|y\right) = \int f\left(\theta\right) P\left(\theta|y\right) d\theta \approx \frac{1}{N} \sum_{i=1}^{N} f(\theta_i). \tag{21}$$

In this manner, summary statistics can be computed. For example, the expectation value $\mathbb{E}(\theta)$ is obtained by setting $f(\theta) = \theta$. Thus, is approximated by the arithmetic mean of the samples θ_i.

The challenge lies in generating the samples θ_i. Drawing samples directly from $P(\theta|y)$[20] is often impossible or computationally infeasible. Instead, Markov Chain Monte Carlo (MCMC) methods are useful to generate samples iteratively from a distribution, called *proposal distribution*, that is updated with each iteration and converges to the target distribution. The proposal distribution for each new sample is defined so that it depends only on the previous sample. Thus, the samples construe a *Markov chain*, which is a sequence of random variables for which the probability distribution of the next variable depends only on the current variable.

[20] *Note that I write* $P(\theta|y)$*, although the common procedure is to generate samples from the* unnormalized *posterior and compute the normalization afterward.*

There are many variants of MCMC algorithms. In this thesis, the MCMC algorithm of choice is the No-U-Turn Sampler (NUTS) [52]. The NUTS is one variant of the Hamiltonian Monte Carlo (HMC) class of algorithms, which in turn are a version of the very general *Metropolis-Hastings* algorithm. The following explanations of these algorithms are adapted from the book *Bayesian Data Analysis* [47].

The *Metropolis-Hastings* algorithm generates samples θ_t ($t = 1, 2, \ldots$) by iteratively repeating the following three steps:

1. Sample a proposal θ^* from the proposal distribution $J\left(\theta^*|\theta^{t-1}\right)$.[21]
2. Compute the acceptance probability $r = \frac{P(\theta^*|y)J\left(\theta^*|\theta^{t-1}\right)}{P(\theta^{t-1}|y)J(\theta^{t-1}|\theta^*)}$.
3. Set $\theta_t = \begin{cases} \theta^* & \text{with probability } \min(1, r) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$

[21] *The plain* Metropolis *algorithm is a special case of the Metropolis-Hastings algorithm which sets the condition of a symmetric proposal distribution. In this case, the acceptance probability simplifies to* $r = \frac{P(\theta^*|y)}{P(\theta^{t-1}|y)}$.

Among the many adaptations and extensions of the Metropolis-Hastings algorithm, the HMC algorithm is a powerful variant that is particularly useful for high-dimensional problems, as it does not exhibit the "random-walk" behavior that is typical for the Metropolis-Hastings algorithm. HMC suppresses the random walk behavior by introducing an additional variable ϕ_j for each variable θ_j. The target distribution becomes a joint distribution $P(\theta, \phi|y) = P((\theta|y)(\phi)|y)$. In analogy to Hamiltonian mechanics, θ and ϕ are interpreted as position and momentum. Typically, the momentum distribution $P(\phi) = \mathcal{N}(0, M)$ is modelled as a multivariate normal distribution with zero mean. [22]

[22] *The covariance matrix* M *is called the "mass matrix", in reference to the physical analogy.*

Like in the Metropolis-Hastings algorithm, θ and ϕ are updated iteratively and accepted or rejected according the acceptance ratio. However, the procedure for proposing a new sample in each iteration is more complicated, involving a series of so-called leapfrog steps where θ is updated according to its momentum and the gradient of the log-posterior. Each iteration of an HMC algorithm proceeds as follows:

(1) Sample $\phi \sim \mathcal{N}(0, M)$. Update θ and ϕ in a sequence of L "leapfrog" steps that each proceed as follows:

(a) Make a half-step of ϕ with the gradient of the log-density of θ:

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon\nabla_\theta \log P(\theta|y).$$

(b) Update θ with the momentum vector:

$$\theta \leftarrow \theta + \epsilon M^{-1}\phi.$$

(c) Make another half-step to update ϕ:

$$\phi \leftarrow \phi + \frac{1}{2}\epsilon\nabla_\theta \log P(\theta|y).$$

(2) Calculate the acceptance probability:

$$r = \frac{P(\theta^*|y)(\phi^*)}{P(\theta^{t-1}|y)P(\phi^{t-1})}.$$

Here, θ^{t-1} and ϕ^{t-1} are the values of θ and ϕ at the beginning of the leapfrog sequence, and θ^* and ϕ^* are the proposed values.

(3) Accept or reject the new sample:

$$\theta_t = \begin{cases} \theta^* & \text{with probability } r \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

This procedure can be tuned in three places: the mass matrix M, the step size ϵ, and the number of leapfrog steps L. Each of these parameters influences the performance of the algorithm and how efficiently it converges to the target distribution.

The samples used in this work, NUTS, is a variant of HMC with procedures for adaptively setting the step size ϵ and the mass matrix M. It also includes criteria for adapting the number of leapfrog steps L with each iteration. It is common practice to start with a tuning phase during which these parameters are varied to improve the performance. The samples generated during this phase are discarded, and the algorithm continues with fixed parameters to generate the sequence of samples from the target distribution. Consequently, this algorithm requires very little manual tuning by the user and still offers very good performance.

3.2.1 *Convergence of Markov chain Monte Carlo*

To use the samples generated by an MCMC algorithm to approximate quantities of interest, it is necessary to ensure that the samples are representative of the target distribution. In the colloquial sense, this trait is called *convergence*. If convergence is reached, the within-sequence correlation of the Markov chain samples does not impair any estimates of the target distribution quantities.

It is common practice to generate several independent chains of samples and compare them to assess convergence. To assess the convergence of the MCMC algorithm, two qualities can be observed: *mixing* and *stationarity*. The former implies that sequences of samples computed from different starting points cover the same space. The latter means that different segments of the same sequence are equivalent, e.g., the first half of the sequence should traverse the same space as the second half.

Qualitatively, mixing and stationarity can be assessed by examining their trace plots, that is a graph that traces the value of the samples as a function of their place in the Markov-chain. Visual inspection would reveal, for instance, whether the first half of one chain differs from the second half, or whether two independent chains cover different parts of the sampling space. However, a visual inspection of graphs does not provide a strict criterion by which to determine convergence.

As a quantitative measure of convergence, this work uses the *potential scale reduction factor* $\widehat{R}$. To be more precise, it uses the rank-normalized split-$\widehat{R}$ described by Vehtari et al. [53] and follows the authors' recommendation of using $\widehat{R} \leqslant 1.01$ as a threshold for convergence. A derivation of this quantity can be found in the above-mentioned article [53]. To compute $\widehat{R}$ for Bayesian inference samples, this work uses the `ArviZ` [51] library for `Python`, which provides an implementation.

3.3 INFERRING REACTOR OPERATING PARAMETERS FROM NUCLEAR WASTE

To build a Bayesian model for inferring reactor operating parameters from nuclear reprocessing waste, the components of the posterior distribution (Equation (18)) need to be defined.

THE OBSERVABLE DATA Y are the nuclides comprising the nuclear waste, or more specifically, isotopic ratios of their concentrations. Nuclear waste typically contains a large number of different nuclides, most of which are created by the various nuclear reactions that occur during the irradiation of the fuel. Categories of nuclides are major and minor actinides[23], fission products, decay products, activation products, and products of other nuclear reactions. In this work, Y is a set of isotopic ratios, comprising some, but not all, of the nuclides that are expected to be present in the waste. The choice of which ratios to include in this set is discussed in Chapter 4. These observables need to be measured in samples of the waste.

[23] *In nuclear engineering, isotopes of thorium, uranium, and plutonium are referred to as* major *actinides. Other actinides are* minor *actinides.*

Measurements of the isotopic ratios are obtained by assaying samples of the spent nuclear fuel (SNF) or the reprocessing waste, which can be performed with a variety of analytical techniques. These techniques can be loosely grouped into two categories: radiometric techniques that measure the radiation produced by the decay of radioactive isotopes, and mass spectrometry techniques that separate ionized atoms with an electrical or magnetic field based on their mass. Generally, characterizing SNF and, in particular, reprocessing waste is a challenging task [54], first because the material is highly radioactive and second because the matrix of nuclides is highly complex, and their relative concentrations vary by orders of magnitude. Some related uncertainty sources and limitations are will be discussed in Section 4.1.

The amount of each nuclide in spent fuel depends on the reactor design (e. g., core layout, moderator material), the conditions in the reactor during operation, and the time since the fuel was irradiated. Some of these operating conditions can be known or estimated for modelling, and the others are the targets for the reconstruction.

THE INFERENCE PARAMETERS θ are parameters that describe the irradiation history of the SNF and the time since it was irradiated. In this work, I generally consider the four parameters Burnup (BU), Cooling Time (CT), Initial Enrichment (IE), and Power Density (PD):

BURNUP represents the *average burnup* of the fuel when it is discharged from the reactor. A definition of the physical quantity burnup is given in Section 2.1. It correlates with the SNF composition via the fission products and products of other reactions that increase with the amount of irradiation the fuel receives.

COOLING TIME is the time since the fuel was last irradiated. It is typically expressed in years or days and often corresponds closely to the time since the fuel was discharged from the reactor core. CT correlates with the SNF composition via the decay of the radioactive isotopes in the SNF.

INITIAL ENRICHMENT is the percentage of the isotope ^{235}U in the uranium of unirradiated fuel, defined as $x_{U-235} = N_{U-235}/N_U$. N refers to the amount of a respective isotope or element. It is usually specified as a percentage of mass (%-wt). This convention is heeded in this thesis, unless noted otherwise. IE correlates with the SNF directly via the ^{235}U content and indirectly via its impact on the power density and reaction rates.

POWER DENSITY refers to the time-averaged power density in the fuel during irradiation. The different definitions of power density are explained in Section 2.1. The power density reflects the fission rate in the fuel, which is indicative of the total reaction rate. Thus, it impacts the relationship between the production and destruction rates of some of the shorter-lived fission products; depending on cross-sections and decay constants, the production and destruction certain nuclides can reach equilibrium states that depend on the power density, which determines their contribution to the SNF composition.

Often, I refer to these inference parameters as the "operating parameters" or the "reactor parameters". These parameters are relevant for the application purposes of BRAM for various reasons. BU and IE are directly related to fissile material production and consumption. The former is linked to the amount of plutonium that is created during reactor operation, and the latter reflects how much ^{235}U is consumed by a reactor, and not used in other parts of the fuel cycle. While CT is not directly linked to fissile material balances, it can be useful to know the time since fuel was discharged to reconstruct a timeline of reactor operations. It is also necessary to model cooling time as a variable parameter, because many of the observable isotopic ratios undergo

radioactive decay. Likewise, PD is treated as a variable parameter because it influences the concentrations of some unstable isotopes in the waste.

PRIOR DISTRIBUTIONS $P(\theta)$ need to be defined for each of the abovementioned inference parameters. Generally, it is best to use a probability distribution that describes the existing state of knowledge as accurately as possible. However, since the prior distribution is equivalent to the likelihood in its impact on the posterior, the prior should be selected with great care, and, in some contexts, it may be better to err on the side of caution and use a prior that is more "naive" than strictly necessary to avoid introducing biases.

In this work, I use uniform distributions as priors. These priors are somewhat uninformative, in the sense that, although they are bounded, they do not prefer or penalize any particular regions of the parameter space within those bounds. The bounds of the uniform distributions are chosen to reflect the physical limits of the parameters or (usually in the case of CT) the context of the problem.

For example, an upper limit on Burnup (BU) can be determined by assessing when the fuel becomes unable to sustain a fission chain reaction and would impede the reactor operation. Limits on Initial Enrichment (IE) can be determined from the reactor type and design. PWRs typically require higher IE (2–5 %) than, for example, PHWRs or Magnox reactors, which can operate with natural uranium. Limits on Power Density (PD) are perhaps more complicated to determine as they require knowledge in nuclear engineering to assess the reactor design. A physical limit is perhaps given by the melting temperature of the fuel[24], although the maximal power determined during the approval procedure of a reactor is a more useful limit.

[24] *This limit should never be reached by the reactor, as it can cause catastrophic failure...*

The Cooling Time (CT) is perhaps the most challenging of the four parameters when it comes to choosing limits for a prior distribution, as it requires an assessment of the time period during which the last irradiation cycle ended before the sampled material was discharged from the reactor. The measurement date of the sample, and the date of sample collection can be precisely known, which makes it possible to determine a lower limit on CT. However, the earliest possible (or likely) discharge date is more difficult to determine and strongly depends on the context of the scenario. A very naive prior might use the construction date of the reactor to define an upper limit, although other sources of information may allow more precise knowledge of the time of discharge.

THE LIKELIHOOD FUNCTION $P(y|\theta)$ is the centerpiece of the Bayesian framework, which describes the relationship between the observable data and the variables of interest. Given a set of isotopic ratios Y, the observed quantity is the vector of isotopic ratio measure-

ments $\vec{y} = (y_r, r \in Y)$. The model $\vec{M}(\theta)$ predicts values of $\vec{y}$ given the inference parameters $\theta = (BU, CT, IE, PD)$ as input variables. Here, the vector notation $\vec{M}$ is used to refer to the whole set of ratio, distinguishing from a model M_r of an individual ratio. With the predictive model $\vec{M}$ for these isotopic ratios and the discrepancy model in Equation (17), the likelihood function is:

$$P(y|\theta) = \mathcal{N}\left(\vec{y} - \vec{M}(\theta), \Sigma\right) = \prod_{r \in Y} \mathcal{N}(y_r - M_r(\theta), \sigma_r). \tag{22}$$

Here, the multivariate normal distribution factorizes under the assumption that the isotopic ratios are independent, i. e., that the covariance matrix Σ contains non-zero elements only on its diagonal. This approximation, while not strictly correct, is used to simplify the computation of the likelihood.

The model M could be based on any of a variety of models, algorithms, or approximations; such as the Monte Carlo-based neutron transport and fuel depletion codes described Section 2.2.2.3. However, using a model based on `OpenMC` or a similar code is impractical at best and impossible at worst when it comes to generating samples from the Bayesian posterior distribution with MCMC algorithms. A single reactor simulation can take hours or even days to compute, and MCMC samplers require at least one model evaluation per generated sample. Drawing thousands of samples becomes computationally expensive and, in many cases, unfeasible, especially since the potential to parallelize most MCMC algorithms is limited due to their sequential nature.

Although it may be possible to mitigate some of the computational cost with smart model simplifications and approximations, it is more convenient to use surrogate modelling techniques, especially for the simulation-based development studies in this thesis.

3.4 SURROGATE MODELLING WITH GAUSSIAN PROCESSES

Surrogate modelling replaces a computationally expensive model with a cheaper model that is fitted to data generated with the expensive model. In this work, surrogate models are used to replace neutron transport and fuel depletion calculations in the likelihood. There are a number of machine learning methods that could be used to create surrogate models of the reactor models to predict the isotopic composition of the nuclear waste. The method of choice are Gaussian processes (GPs), which come with two benefits: their gradient can be computed analytically, and they require less training data compared to other techniques. The former is necessary to be able to use gradient-based MCMC samplers (e. g., NUTS), and the latter reduces the computational cost of generating training data. Previous work has demonstrated that GPs are well suited to model the isotopic composition of nuclear fuel [55].

3.4.1 *Gaussian Processes*

The following, simplified explanation is based on "Gaussian Processes for Machine Learning" by Rasmussen and Williams [56]. The authors define a Gaussian process (GP) as a "collection of random variables, any finite number of which have a joint Gaussian distribution". Given a function $f(\vec{x})$, a finite set of function values f^* with associated input values X^* is treated as normally distributed with a mean value m and covariance $K(X^*, X^*)$:

$$f^* \sim \mathcal{N}(m, K(X^*, X^*)). \tag{23}$$

K is the covariance matrix containing the elementwise kernel evaluations between the input points X^*. I follow the common practice and set $m = 0$.

An example of a commonly used kernel function is the *squared exponential* kernel (in one dimension):

$$k(x, x') = \exp\left(-\frac{1}{2l^2}(x - x')^2\right). \tag{24}$$

With this kernel, the covariance depends on the distance between two input points and a parameter l, called *lengthscale*. The lengthscale determines how quickly the function changes across the input space, that is, input values that are farther apart than l are less strongly correlated.

To serve as a surrogate model, a GP needs a method to predict function values f^* at new input points X^* given a set of observations f at known input points X, a. k. a. training data. Here, the newly predicted function values are called *test outputs* and the observed function values *training outputs*. The joint distribution of the training outputs and the test outputs is a multivariate Gaussian distribution (with $m = 0$):

$$\begin{bmatrix} f \\ f^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X, X) & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{bmatrix}\right). \tag{25}$$

The distribution of the test outputs conditional on the training outputs is then given by:

$$\begin{aligned} f^*|X^*, X, f &\sim \mathcal{N}(\mu(f, X, X^*), \Sigma(X, X^*)), \text{ with} \\ \mu(f, X, X^*) &= K(X^*, K)K^{-1}(X, X)f \text{ and} \\ \Sigma(f, X, X^*) &= K(X^*, X^*) - K(X^*, X)K^{-1}(X, X)K(X, X^*). \end{aligned} \tag{26}$$

After μ and Σ are computed, samples are drawn from the predictive distribution. However, to use the GP as a surrogate model in the typical sense, the expectation value $\mu(f, X, X^*)$ serves as the prediction of the function value at the new input points X^*.

3.4.2 *Hyperparameter Optimization*

Typically, the covariance functions contain some free parameters, such as the lengthscale l in Equation (24), which are called *hyperparameters* and are denoted with η. Varying the hyperparameters evidently changes the value of $\mu(f, X, X^*)$ in Equation (26). To improve the predictive performance, the hyperparameters can be optimized. The log marginal likelihood with respect to the hyperparameters can be defined as:

$$\log p(f|X, \eta) = -\frac{1}{2} f^T K^{-1} f - \frac{1}{2} \log |K| - \frac{n}{2} \log(2\pi), \tag{27}$$

with n being the number of training data points, i. e., the number of points on which the GP is conditioned and K^{-1} the inverse of the kernel matrix. Gradient-based optimization algorithms are used find the hyperparameters that maximize the log marginal likelihood.

Since it is by no means given that the optimization converges to a global maximum, it can be useful to repeat the optimization with different starting values and use the best result out of all repetitions. To ensure that the hyperparameters are sufficiently optimized, a test dataset can be used to validate the model's predictive performance, which is common practice in machine learning applications. The validation procedure also serves to avoid "overtraining", which occurs when a model is optimized to precisely predict the values in the training data but produces inaccurate predictions on input values that are not in the training data.

3.4.3 *Training Gaussian Process Surrogate Models in Practice*

In this thesis, the purpose of the surrogate models is to predict concentrations of isotopes in reprocessing waste given reactor operating parameters as input variables. There are three fundamental steps to creating such a surrogate model:

1. Generate a set of input data, i. e., reactor parameters BU, CT, IE, PD
2. Simulate output values, i. e., the SNF composition, for each input with Monte Carlo reactor physics simulations.
3. Optimize the hyperparameters of the GP model, using the input values with their associated simulated output.

The training inputs are sampled randomly from the targeted input parameter space, which is determined by the range of parameters covered by the prior distributions of the inference parameters. The surrogate model should offer accurate predictions in any region of the parameter space that the sampling algorithm of the inference routine

can potentially enter. For this purpose, the Sobol sequence is used to generate the input data, which ensures a more even coverage of the parameter space compared to quasi-random sampling-based input values [57].

The quality of the surrogate models is assessed with the *coefficient of determination* R^2, which is defined as:

$$R^2 = 1 - \frac{\sum_i \left(y_i^* - f_i^*\right)^2}{\sum_i \left(y_i^* - \bar{y}^*\right)^2}. \tag{28}$$

Here, f_i^* are the predictions obtained on a set of test inputs with the optimized GP hyperparameters, y_i^* are the corresponding test output values, and $\bar{y}^*$ is the mean of the y_i^*.

An iterative procedure is used to repeat the hyperparameter optimization with new starting values if the optimization does not converge to a satisfactory value of R^2. If the condition is met, an additional R^2 is computed for a third, independent validation dataset. If it meets the condition $R^2 > 0.95$, the hyperparameters are chosen. In practice, a training dataset with 200–500 points was found to be sufficient to obtain surrogate models for all necessary isotopes. An individual model is trained for each isotope or isotopic ratio, that is, each model has a single output value.

3.5 SOFTWARE FOR THE BAYESIAN REPROCESSING WASTE ANALYSIS METHOD

The inference framework is implemented in `Python` and has been released as an open-source library under the name `bram-inferno` [58]. It makes extensive use of the `PyMC` library [59]. Processing and analysis of the output of the sampling algorithm is performed in `ArviZ` [51], a `Python` library for analyzing Bayesian models.

For surrogate modelling purposes, I implemented procedures for Gaussian process regression that I have released as the `Python` package `gp-surfer` [60]. The Sobol sequence algorithm provided by the `SciPy` [61] library was used in this package and throughout this work.

4

SELECTING SUITABLE ISOTOPIC RATIOS

The Bayesian Reprocessing Waste Analysis Method (BRAM) conditions posterior distributions of reactor operating parameters on observations of isotopic ratios in samples of reprocessing waste. A large number of different isotopes are created in nuclear fuel during and after reactor operation that end up in the reprocessing waste. To be useful, isotopic ratios need to be measurable with the current state-of-the-art methods, such as mass spectrometry[25].

As the number of measurable isotopic ratios can be large, it may be desirable or necessary to use only a subset of the available ratios. A practical reason for this is to ensure that the MCMC algorithms (see Section 3.2.1) can converge to the posterior distribution with a reasonable computational effort. Using large ratio sets ($\gtrsim 50$) provides a lot of potential information for the posterior, but it can lead to increased computational cost and convergence problems during sampling. Another reason to limit the number of ratios can be the economic cost of assaying the sample. Generally, the high-precision measurement techniques required for a precise and accurate assay are expensive. If many samples from different sources need to be analyzed during an investigation, it may be preferable to measure only a small set of ratios that is sufficiently sensitive to the inference parameters to limit the cost. Thus, the objective of this exercise is to select a suitable subset of the measurable isotopic ratios and ensure that something can be "learned" through the inference; that the ratios encode information about the inference parameters.

Fundamentally, the relation between the reactor operating parameters and the measurable nuclide quantities and ratios are governed by the physical processes that determine their creation and destruction and removal in and from the nuclear fuel, during and after its residence in the reactor core, storage facilities and reprocessing facilities. These physical processes, while explainable and to a certain degree predictable with theoretical and computational models, constitute a vast and complex linked net of processes, whose outcome is difficult to model and predict with precision. Specific indicator nuclides and ratios are frequently used to characterize SNF in safeguards and radioactive waste management [62]; indicators have also been studied in nuclear forensics contexts [63]. These nuclide ratios do not serve our purpose, mainly because there are very few *intra-element* isotopic ratios among them.

In his previous research, Figueroa [29] developed and algorithm to select a set of isotopic ratios optimized for reconstructing Bur-

[25] *Inductively coupled plasma mass spectrometry (ICP-MS) and thermal ionization mass spectrometry (TIMS) can be used to precisely characterize spent nuclear fuel and reprocessing waste [54].*

nup (BU) and Cooling Time (CT) of waste from a Pressurized Water Reactor (PWR). For the present thesis, a new set of isotopic ratios was selected that is better suited to serve as evidence for reconstructing more parameters and different reactor types.

This chapter first discusses considerations for selecting isotopic ratio candidates that are expected to be present and measurable in reprocessing waste. Then, it presents two algorithms for selecting a subset of these isotopic ratios by optimizing an objective functions that quantifies the expected utility of the ratios for conditioning the posterior. Finally, this chapter presents a set of ratios that performs better than the previously used set and will be used to further develop BRAM in the following chapters.

4.1 SELECTING ISOTOPIC RATIO CANDIDATES

Monte Carlo reactor physics simulations and depletion calculations track the time evolution of a vast number of nuclides. This includes nuclides that are created in quantities that are, while numerically computable, physically insignificant, down to sub-integer atomic concentrations. Such nuclides need to be filtered to obtain a plausible set of useful isotopic ratios based on simulated data.

Furthermore, the simulation output contains nuclide number densities as expected in spent nuclear fuel. SNF is diluted during reprocessing (see Section 2.1) and consequently the isotopic concentrations are different in the highly active raffinate (HAR) (reprocessing waste) than in SNF, which affects their detectability. Some elements are also affected by the chemical processes and do not pass through the reprocessing procedure without losses. In the following, I outline the procedure I used to select a set of isotopic ratio candidates that are likely to be measurable and can serve as a basis for further optimization. There are three principal steps to this pre-selection procedure:

1. Convert simulated SNF number densities to approximated HAR number densities.
2. Filter nuclides below a *detectability threshold*.
3. Remove nuclides that are otherwise compromised by the reprocessing steps.

4.1.1 *Converting SNF to HAR*

Reprocessing involves a series of chemical processes that ultimately convert SNF to HAR. Although the general reprocessing procedure has been discussed in academic and non-academic literature, the precise process parameters required for accurate calculation of concentrations are not publicly available. Instead, I approximated a calibra-

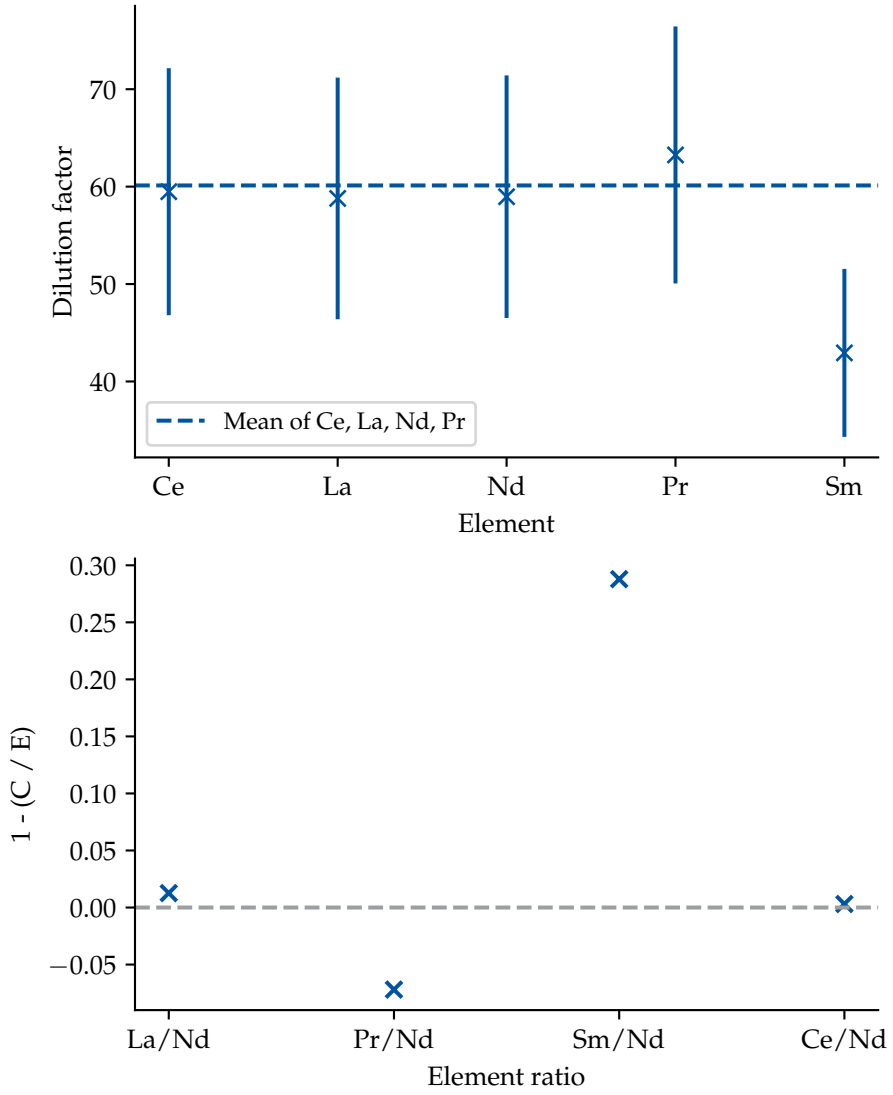


Figure 2: Dilution factor to convert simulated SNF nuclide densities to HAR nuclide densities. The upper graph shows the computed dilution factor for each of the available Lanthanides. The errorbars in the upper graph indicate the variance due to varying burnup input values for the computed values. The experimental data in Comte et al. [64] is specified without measurement uncertainties. The lower graph shows the deviation between computed element mass density ratios C and experimental mass density ratios E. Here, the errorbars have been omitted from the lower graph to improve legibility.

tion using single set of published experimental data [64], with the advice of Stefaan van Winckel of the Joint Research Center in Karlsruhe, who is a knowledgeable expert in the field of radiochemistry. Since reprocessing serves to separate only uranium and plutonium, the concentrations of the fission products and other nuclides remains approximately unchanged relative to each other. They are all diluted approximately equally during reprocessing.

The experimental data contains concentration measurements for only a few useful elements, namely the lanthanides cerium, lanthanum, praseodymium, neodymium, and samarium. I define the dilution factor DF_l for each lanthanide

$$\mathrm{DF}_l = \rho_{C,l}/\rho_{E,l} \tag{29}$$

as the ratio of its simulated mass density ρ_C in the SNF and its experimental mass density ρ_E in the HAR. Since the corresponding fuel burnup for the measurement data is not given in the paper, I used the average ρ_C of a simulated dataset with sampled burnup values typical of spent PWR fuel, ranging from $20\,\mathrm{MW\,d\,kg^{-1}}$ to $45\,\mathrm{MW\,d\,kg^{-1}}$.

The top graph in Figure 2 shows the dilution factor for each of the five Lanthanides. The dilution factor of samarium deviates significantly from the other elements. This deviation becomes more apparent in the lower graph of Figure 2, which shows the deviation between computed and experimental ratios of the lanthanide concentrations. Given the assumption that these elements are largely unaffected by the reprocessing steps, the mass density ratios between the elements of this group are expected to be the same before and after reprocessing. Again, the ratio of Sm/Nd deviates significantly more than the other elemental ratios. The deviation between computed and experimental Sm/Nd is approximately 30 %, whereas the other ratios deviate by less than 10 %. The cause for this deviation is unclear; it could stem from simulation inaccuracy or measurement uncertainty.[26]

[26] *Note that Comte et al. [64] do not specify measurement errors.*

Given the unexplainable deviation, samarium is excluded when computing the dilution factor. Averaging DF of the other elements yields:

$$\widehat{\mathrm{DF}} = 60,$$

which is applied to all nuclides in the dataset to convert SNF densities to HAR densities.

It is clear that DF is a crude approximation. This approximation is intentionally simplistic, because the available information about the experimental data is limited. A more sophisticated estimate would require more information about the reprocessing steps and the operating conditions in the reactor. It is noteworthy that DF was computed with simulated and experimental data for a PWR. The reprocessing steps are different depending on the type of fuel. Thus, while it seems plausible to assume that the approximated dilution factor also holds

for SNF from a Pressurized Heavy-Water Reactor (PHWR), since they also use UO_2 fuel, it is perhaps less reasonable to expect the same dilution factor for Magnesium Non-OXidising (Magnox) fuel, which is uranium metal. Nonetheless, for lack of a better approximation, I use the same dilution factor for all three reactor types.

4.1.2 *Detection threshold*

In this work, I assume that a sample of reprocessing waste will be analyzed with the latest state-of-the-art methods. Nevertheless, or perhaps as a consequence, it is difficult to define a general detection threshold. Van Winckel et al. [54] provides an insightful overview of the various techniques for assaying SNF. The same techniques can be applied to HAR.

However, exact detection limits vary from element to element and also depend on the composition of the material that is being assayed. Generally, 10 parts-per-trillion is a reasonable approximation as detection limit for sophisticated assay techniques; for some nuclides it will be lower and for some it will be higher. Detection limits can be lowered slightly [65] by using chemical processes to separate elements and increase their concentration and to reduce uncertainty inducing effects, such as isobaric interference.[27] By increasing the concentration of elements of interest, the detection limit may be improved by one order of magnitude. Therefore, the threshold for selecting candidate nuclides was set to 1 part-per-trillion. This is an optimistic threshold, aiming to include many as nuclides. Nonetheless, it is important to emphasize that the detection threshold and the measurement uncertainty depend on the composition of the material. It may be the case that not all nuclides in the dataset can be measured at such low concentrations.

To convert the approximated nuclide number densities in the HAR to concentrations, I use a typical value for HAR density: $1.3\,\mathrm{g\,cm^{-3}}$ [65].

[27] *Isobaric nuclides cannot be separated by mass-spectrometry because their mass differences are too small. In most applications they are considered to have the same mass.*

4.1.3 *Excluding volatile nuclides and decay products*

Other nuclides and elements must be excluded because their losses during reprocessing cannot be predicted accurately. The most obvious of these are the noble gases, because they are chemically inert and escape at an unknown rate during reprocessing. Decay products of the radioactive isotopes of noble gases are also excluded as the unpredictable escape of the parent isotopes makes the entire decay chain unpredictable. The element iodine has also been excluded since it is volatile and can be expected to escape in significant quantities during reprocessing.

A similar effect needs to be considered for the decay products of U and Pu isotopes. Since U and Pu are removed at an unknown point in time, the production rate of their decay progeny is abruptly changed, and they cannot be computed reliably.

It may be prudent to exclude other nuclides as well, depending on the reprocessing infrastructure and the timescale of the various steps. Such exclusions would be case-specific and were not considered here.

4.1.4 *Selecting candidate ratios*

The pre-selection procedure is performed with a dataset of simulated nuclide densities that correspond to a set of uniformly generated input parameter values that cover the expected input parameter space. A nuclide needs to pass the pre-selection procedure in each of these samples to enter the set of candidates to ensure that the final ratio set is sensitive on the entire parameter space.

This process is repeated on three datasets, obtained with three reactor models: Magnox, PHWR, and PWR. The final set of isotopic ratio candidates is the intersection of the ratio sets of each reactor type. It contains 127 isotopes, belonging to 27 different elements, which are listed in Table 1.

These nuclides can be combined into 270 *intra-element* isotopic ratios. Although ratios of isotopes of different elements are used to characterize SNF [66], this work only uses ratios of isotopes of the same element, i. e., intra-element isotopic ratios. This restriction is necessary due to the assumption that the sample will have undergone chemical separation processes, such as the PUREX processes or procedures used to enhance the sample purity for assaying. The separation efficiency of such processes varies for different elements [63]. Thus, the chemical separation processes would potentially add biases to ratios between different elements. Using intra-element ratios mitigates this problem, although it does not avoid it entirely, because radionuclides and their decay products may be separated with varying efficiency. This effect is neglected in the present, simulation-based research; its effect on the ratio uncertainty needs to be determined in an experimental setting.

Table 1: Nuclides that meet the pre-selection criteria. Intra-element ratios thereof serve as the candidate pool for selecting a smaller, optimized set of ratios.

Ba	^{134}Ba, ^{136}Ba, ^{137}Ba, ^{138}Ba
Cd	^{110}Cd, ^{111}Cd, ^{112}Cd, ^{113}Cd, ^{113m}Cd, ^{114}Cd, ^{116}Cd
Ce	^{140}Ce, ^{142}Ce
Cs	^{133}Cs, ^{134}Cs, ^{137}Cs
Dy	^{160}Dy, ^{161}Dy, ^{162}Dy, ^{163}Dy, ^{164}Dy
Er	^{166}Er, ^{167}Er, ^{168}Er
Eu	^{151}Eu, ^{152}Eu, ^{153}Eu, ^{154}Eu, ^{155}Eu
Ga	^{69}Ga, ^{71}Ga
Gd	^{152}Gd, ^{154}Gd, ^{155}Gd, ^{156}Gd, ^{157}Gd, ^{158}Gd, ^{160}Gd
Ge	^{72}Ge, ^{73}Ge, ^{74}Ge, ^{76}Ge
In	^{113}In, ^{115}In
La	^{138}La, ^{139}La
Mo	^{100}Mo, ^{95}Mo, ^{96}Mo, ^{97}Mo, ^{98}Mo
Nb	^{93}Nb, ^{93m}Nb†
Nd	^{142}Nd, ^{143}Nd, ^{144}Nd, ^{145}Nd, ^{146}Nd, ^{148}Nd, ^{150}Nd
Pd	^{104}Pd, ^{105}Pd, ^{106}Pd, ^{107}Pd, ^{108}Pd, ^{110}Pd
Pu	^{238}Pu, ^{239}Pu, ^{240}Pu, ^{241}Pu, ^{242}Pu
Ru	^{100}Ru, ^{101}Ru, ^{102}Ru, ^{104}Ru, ^{99}Ru
Sb	^{121}Sb, ^{123}Sb, ^{125}Sb
Se	^{77}Se, ^{78}Se, ^{79}Se, ^{80}Se, ^{82}Se
Sm	^{147}Sm, ^{148}Sm, ^{149}Sm, ^{150}Sm, ^{151}Sm, ^{152}Sm, ^{154}Sm
Sn	^{115}Sn, ^{116}Sn, ^{117}Sn, ^{118}Sn, ^{119}Sn, ^{120}Sn, ^{121m}Sn†, ^{122}Sn, ^{124}Sn, ^{126}Sn
Sr	^{86}Sr, ^{88}Sr, ^{90}Sr
Te	^{122}Te, ^{124}Te, ^{125m}Te†, ^{126}Te, ^{128}Te, ^{130}Te
U	^{234}U, ^{235}U, ^{236}U, ^{238}U
Y	^{89}Y, ^{90}Y
Zr	^{90}Zr, ^{91}Zr, ^{92}Zr, ^{93}Zr, ^{94}Zr, ^{96}Zr

† Here, m indicates the first metastable excited states.

4.2 OPTIMIZING A SUBSET OF THE RATIO CANDIDATES

To select an "optimized" subset of isotopic ratios from the set of candidates, this thesis frames the inference process as a model with the isotopic ratio set as input and posterior probability distributions as outputs. To optimize the input (ratio set), an algorithm requires a quantitative measure of the "goodness" of the posteriors to compare different ratio sets and select those that perform "better". Here, this measure is called *objective function*. This section first discusses the objective function, then describes two different approaches to optimizing it.

4.2.1 *Objective functions*

The objective function needs to meet several requirements in order to serve its purpose as a measure to be optimized for selecting a useful set of isotopic ratios. First, it should assess to what degree potential posterior distributions will agree with the corresponding ground truth, i. e., true parameter values. Second, it should evaluate this agreement for the whole parameter space of potential true values to ensure the general suitability of the ratio set. Third, it should be relatively efficient to compute so that it can be used in iterative algorithms that require many evaluations.

The objective function (and its variations) that I describe here uses the work of Figueroa [29] as a foundation. He developed the concept for approximating marginal posterior distributions to compute summary statistics that measure the suitability of a set of ratios. I introduced new concepts to his algorithm to significantly improve computational efficiency and introduced alternative summary statistics that better capture the agreement between posterior and ground truth.

Summarizing marginal posterior distributions

Estimators of the mean $\hat{\mu}$ and the standard deviation $\hat{\sigma}$ are used as summary statistics for approximated marginals of the Bayesian posterior distributions.[28] They are computed using a method called approximate Bayesian computation (ABC), which proceeds as follows:

1. Sample input value from the prior: $\theta_s \propto p(\theta)$.
2. Compute a ratio estimate with the surrogate model: $y\prime = M(\theta_S)$.
3. Compute weights given the likelihood conditional on a test point: $w_l = \mathcal{N}(y\prime|y_{test}, \sigma_{test})$.
4. Repeat for N samples.

[28] *The quantities used to summarize and compare the posterior distributions here are different to those that I use later to evaluate the performance of the inference framework (see Chapter 5). The principal reason is the need for computational speed.*

The summary statistics $\hat{\mu}$ and $\hat{\sigma}$[29] are computed with the samples θ_s and the likelihood-weights w_l.

This method is conceptually similar to the Markov Chain Monte Carlo (MCMC) approach. They key difference is the input sampling procedure: in ABC the input samples are generated to uniformly cover the input space, whereas in MCMC each new sample is generated given a proposal distribution that depends on the previous sample and then needs to fulfill an acceptance criterion (see Section 3.2). The former is typically faster to compute, while the latter typically yields more accurate posterior distributions.

This application of ABC was first implemented by Figueroa [29] with a grid-based sampling approach for models with a two-dimensional prior input space. For the present work, I implemented several changes to the algorithm to reduce the computational cost and enable higher dimensional input spaces. These changes are outlined in Appendix A.

[29] *$\hat{\sigma}$ is a summary statistic for the marginal posterior distribution. σ_{test} is a parameter of the likelihood that accounts for measurement and model uncertainty.*

Fitness functions

The *fitness function*, which is distinct from the objective function, is a quantity that summarizes the marginal posterior distributions in a manner that reflects their "goodness". They consist of two parts: a score computed for each marginal distribution derived from their respective summary statistics $\hat{\mu}_i$ and $\hat{\sigma}_i$, and a function that combines the scores into a single value, either the sum or the maximum of the individual scores. For this research, the following scores were implemented and tested:

$$
\begin{aligned}
f_1 &= \max\left(\left\{\frac{\hat{\sigma}_i}{\hat{\mu}_i}\right\}\right)\cdot 100 \\
f_2 &= \max\left(\left\{\frac{|\hat{\mu}_i - x_{test,i}|}{x_{test,i}}\right\}\right)\cdot 100 \\
f_3 &= \max\left(\left\{\frac{|\mu_i - x_{test,i}|}{x_{test,i}} + \frac{\hat{\sigma}_i}{\hat{\mu}_i}\right\}\right)\cdot 100 \\
f_4 &= \sum_i\left(\frac{\hat{\sigma}_i}{\hat{\mu}_i}\right)\cdot 100 \\
f_5 &= \sum_i\left(\frac{|\hat{\mu}_i - x_{test,i}|}{x_{test,i}}\right)\cdot 100 \\
f_6 &= \sum_i\left(\frac{|\mu_i - x_{test,i}|}{x_{test,i}} + \frac{\hat{\sigma}_i}{\hat{\mu}_i}\right)\cdot 100 \\
f_7 &= \max\left(\left\{\frac{2\cdot|\hat{\mu}_i - x_{test,i}|}{x_{test,i} + \hat{\mu}_i}\right\}\right) \\
f_8 &= \sum_i\left(\frac{2\cdot|\hat{\mu}_i - x_{test,i}|}{x_{test,i} + \hat{\mu}_i}\right)
\end{aligned}
\tag{30}
$$

Here, x_{test} is ground truth of the approximated posterior of the respective parameter, and the indices i iterate over the parameters, i. e., BU, CT, IE, and PD. Each fitness function is conditional on a given set of isotopic ratios Y.

Fitness function f_1 is the maximum of the relative standard deviations of the marginals. This function was used by Figueroa [29]. Using the relative standard deviation has the potential drawback that, while

it selects for "narrow" posteriors, it does not ensure that the true parameter values are included in the posterior probability space.

The other functions in Equation (30) were introduced in this work to capture the deviation between the posterior mean and the true parameter value, to ensure that the posteriors correctly infer the true values.

Parameter space coverage

The fitness functions in Equation (30) provide a quantitative measure to compare sets of isotopic ratios given a single test point. For practical applications, a set of ratios needs to perform well on the entire, anticipated, parameter space. Therefore, the objective function computes the fitness values for many test points and averages the results to obtain a single, comparable quantity. Each test point x_{test} is a realization the vector of inference parameters. The set of test points TP should cover the anticipated parameter space evenly. To this end, the Sobol sequence [57] is utilized to construct quasi-random test point samples.

A separate set of test points is constructed for each reactor model that is to be considered in the later inference, and the average of the reactor model objective functions is computed. This yields the objective function O_k:

$$O_k(Y) = \frac{1}{M} \sum_{m}^{M} \frac{1}{N} \sum_{n}^{N} f_{k,m}\left(x_{test,n,m}, Y\right) \tag{31}$$

for a given set of ratios Y. Here, M is the number of reactor models, $f_{k,m}$ is the fitness function f_k (Equation (30)) for a specific reactor m, and N is the number of test points x_{test}.

This objective function, implemented according to Algorithm 1, now provides a single quantitative measure for the "goodness" of a set of isotopic ratios, that, by way of numerical optimization (minimization), can be used to compare different sets of ratios.

Algorithm 1 The parameter space-averaging objective function

```
function ObjectiveFunction(test-points, models, Y)
    scores ← [ ]
    for 𝓜 in models do
        for x_test in test-points do
            μ, σ ← SummarizeMarginals(x_test, Y, 𝓜)
            fitness← FitnessFunction(μ, σ, x_test)
            scores.append(fitness)
        end for
    end for
    return mean(scores)
end function
```

4.2.2 *A combinatorial approach to comparing ratio lists*

A straightforward approach to finding the optimal subset of isotopic ratios is to compare all possible subsets. I call this approach the *combinatorial* algorithm. Since it is computationally infeasible to test all possible combinations of isotopic ratios of all possible list lengths, Figueroa [29] chose to compare only pairs of isotopic ratios, and iterate over all possible pairs. The objective function is then computed for each pair and a set of N pairs with the highest objective function values is selected. This approach is depicted in Algorithm 2.

Algorithm 2 Combinatorial approach to comparing ratio lists. The objective function is described in Algorithm 1.

```
1:  procedure CombinatorialRatioSelection(test-points, models)
2:      Ratios ← RatioCandidates
3:      Pairs ← AllPairs(Ratios)
4:      ObjectiveScores ← [ ]
5:      for pair in Pairs do
6:          score ← ObjectiveFunction(test-points, models, pair)
7:          ObjectiveValues.append(score)
8:      end for
9:      return TopN(Pairs, ObjectiveValues)
10: end procedure
```

It is to be noted that the approach chosen in this work deviates from the work of Figueroa [29] by averaging the fitness function values over the test points. In Figueroa's work, the best pairs are chosen individually for each test point and combined into a single set afterward, with the goal to include at least some ratios that perform well on each test point and thus ensure good performance all across the parameter space.

A potential weakness of the combinatorial approach is that it only compares pairs of isotopic ratios directly, and assumes that the union of the "best" pairs of ratios still performs "well". To improve the ratio selection procedure, an alternative algorithm was developed to allow direct comparison of larger sets of isotopic ratios.

4.2.3 *A genetic algorithm for comparing ratio lists*

This new approach to ratio selection uses a genetic algorithm to explore the space of possible ratio sets. Genetic algorithms are heuristic search approaches that seek to mimic the process of natural selection in evolution. A basic genetic algorithm is depicted in Algorithm 3. For a more detailed introduction to genetic algorithms, see Kramer [67].

The core elements of a genetic algorithm are a mapping between a *genotype* and a *phenotype*, a set of operations that mutate the genotype, and a fitness (or objective) function that compares the phenotypes. The algorithm starts with a, usually randomly selected, initial population of genotypes, which is iteratively updated. Each iteration of the algorithm is called a generation. In each generation, the genotypes of the population are mutated, mapped to the corresponding phenotype, and the fitness function of the phenotype is computed. Based on the fitness values, a new "parent" population is selected for the next generation. The algorithm stops if a stopping condition is met, which can be a fixed number of generation, or another convergence criterion.

Algorithm 3 A basic genetic algorithm

```
1: procedure GeneticAlgorithm
2:     Initialize population
3:     repeat
4:         Mutate members of population
5:         Map genotype to phenotype
6:         Evaluate fitness of population
7:         Select members for reproduction
8:     until stopping condition
9: end procedure
```

In the context of selecting an optimized subset of isotopic ratios, the genotype of an individual in the population is a set of isotopic ratios, which is drawn from the set of candidate ratios. The corresponding phenotype of a genotype are the marginal distributions of the Bayesian posteriors, which are approximated as described in Section 4.2.1. The "genetic fitness" is evaluated by one of the variants (Equation (30)) of the objective functions in Equation (31).

First, the fittest members of the population are chosen for reproduction. From the fittest members, the next generation is generated by applying mutations to their genotypes. In this context, a *gene* refers to a single isotopic ratio. To explore the space of possible genotypes, the genetic ratio-selection algorithm uses the following mutations to modify the genotypes of the fittest members and thus generate a new population:

ELITISM The fittest members of the population are copied to the next generation.

CROSS-OVER A new genotype is generated by randomly choosing genes from the fittest members of the population.

GENE SWAP A single gene is replaced by a randomly chosen gene from the *gene pool* (set of all candidate ratios).

NOVELTY SEARCH Randomly sample a new genotype unrelated to the previous generation.

The *novelty search* mutation reduces the risk of the algorithm getting stuck in a local optimum of the fitness function and promotes a wider coverage of the space of possible ratio sets.

It is possible to let the algorithm vary the number of ratios in a single genotype, both by allowing *deletion* and *addition* as mutations that remove or add a random ratio to the *elites* and by allowing the number ratios in the randomly sampled ratio sets to vary randomly as well. In light of the stated aim to optimize the set of ratios with a constrained set size, these mutations were not used in this work.

Algorithm 4 The Genetic algorithm for selecting isotopic ratios. The objective function is outlined in Algorithm 1.

```
procedure GeneticRatioSelection(test-points, models)
    Initialize population
    for _ in range(N) do Fitness ← [ ]
        for member in population do
            Score ← ObjectiveFunction(test-points, models, member)
            Fitness.append(Score)
        end for
        Elites ← fittest members of population
        New population ← Elites
        New population ← New population ∪ CrossOver(Elites)
        New population ← New population ∪ GeneSwap(Elites)
        New population ← New population ∪ NoveltySearch
    end for
end procedure
```

My implementation of the genetic algorithm uses a fixed number of iterations, instead of a stopping criterion. The algorithm is implemented in `Python` and uses the `NumPy` [68] and `SciPy` [61] libraries for numerical computations. It has been published in open-source repositories [69].

4.2.4 *On the number of selected ratios*

The number of ratios in the selected set can be variable or fixed in both selection algorithms. In the combinatorial algorithm, the number of chosen pairs can be set, leaving open how many unique ratios end up in the final set. Alternatively, the desired size of the set can be specified by adding the "next-best" ratio pair to the set until it has the desired size. With the genetic algorithm, mutations that add or remove ratios to the set can be activated or deactivated and the size of freshly sampled ratio sets can be randomized within user-specified minimum and maximum limits.

Preliminary tests showed that the size of the ratio set has a significant impact on the results of the inference, perhaps more so than the choice of algorithm or objective function. In the following, the size of the ratio set was fixed beforehand. Thus, the performance evaluation compares ratio sets of the same size, leading differences back to the different choices of ratios (by the respective algorithms) rather than the number of ratios. Furthermore, the set size needs to be fixed to ensure that the more complex inference models developed in later chapters can also converge with reasonable computational resources.

4.3 COMPARING THE SUITABILITY OF RATIO LISTS

Both the combinatorial and the genetic algorithm optimize the ratio set using approximated summary statistics that can be computed efficiently. To gain confidence that the ratio set will actually lead to informative posterior distributions, I use simulated test scenarios to evaluate how well the MCMC sampling converges, whether the posterior probability density agrees with the true parameter values, and how much information is learned.

This test framework comprises three steps: generating a dataset of associated reactor parameter and isotopic ratio vectors, running the inference for each item in the dataset, and evaluating performance statistics that measure the above criteria.

The test dataset is a mapping of test input parameters $x_{test} = (BU, CT, IE, PD)$ to isotopic ratio vectors $\vec{y}_{obs}$. A total of N_{tp} test inputs, which are henceforth called *ground truth*, are generated with the quasi-random Sobol sequence to cover the input parameter space as evenly as possible. For each test point, isotopic ratio values are computed with the GP surrogate models. The ratio values are then perturbed with Gaussian noise σ_{sam} to simulate measurement uncertainty. For each test point, N_{sam} differently perturbed samples are generated, leading to a total of $N = N_{tp} \times N_{sam}$ test cases in a test scenario.

Once the posterior distributions have been inferred for each test case, the following test criteria are evaluated to assess the *performance* of the ratio set.

4.3.1 *Test criteria*

Convergence

The first step to analyzing the quality of a ratio set is to assess the convergence of the MCMC runs. This serves two purposes: first, it excludes runs that have not converged from the other test metrics, and second, it identifies whether a set of ratios leads to frequent convergence issues, which is undesirable. Here, I use the $\widehat{R}$ statistic, as

described in Section 3.2.1, with a threshold value of 1.01, which is recommended by Vehtari et al. [53]. Only test cases in which all marginal posterior distributions yield an $\widehat{R}$ statistic below 1.01 are considered to have converged. The remaining test cases are excluded from the other test criteria.[30]

[30] *The individual samples corresponding to one test point are evaluated separately.*

Posterior–ground truth agreement

The second step is to ensure that the posterior distributions include the ground truth values with a high-enough probability. If the prior and the likelihood are correctly defined, the *true* posterior distribution covers the ground truth values. Since MCMC only obtains a finite number of samples from the posterior distribution, there is no guarantee that the ground truth values are covered by it. Indeed, due to the finite sample size, it is not possible to define a range within which the true value is guaranteed to lie. The Bayesian solution to this problem is the highest (posterior) density interval (HDI). The HDI is the smallest Bayesian credible interval, which is any interval that the value of the inferred parameter will fall into with a probability of γ. A common notation is also γ-HDI. In this work, unless otherwise noted, I use the 95%-HDI, which, per the above definition, is the smallest interval that contains the true value with a probability of 95%.

The test criterion for this test series is that the ground truth value of a parameter is included in the 95%-HDI of its posterior distribution. I call this the *success* criterion S, which has a binary `true` or `false` value for each parameter in a test run:

$$S = \begin{cases} 1 & x_{true} \in \text{HDI} \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

Since I am simulating the observed data with a known approach, an instinctive assumption may be that the ground truth should be included in the 95%-HDI in 95% of the test cases. This is not the case. If the observable quantities, i. e., the isotopic ratios, were completely uncorrelated with the parameters, the posterior would reflect the prior, which is uniformly distributed. The HDI would then be arbitrarily placed to cover 95% of the prior space and thus easily exclude a ground truth value near the edge of the input space, regardless of how often I repeat the inference with simulated observables. On the other hand, if I were to use informative prior distributions, the coverage probability could be much higher than 95%.

Information gain

The success criterion is the most important metric, as it is imperative that the inference results are reliable. However, if the success criterion is met, it is also useful to consider how much information is gained

with the inference with respect to the prior information. To this end, I use the Kullback-Leibler (KL) divergence [70] to measure the difference in information between the prior $p(\theta)$ and the marginal posterior distributions $p(\theta|y_{obs})$:

$$D_{KL}\left(p\left(\theta|y_{obs}\right), p\left(\theta\right)\right) = \int \log\left[\frac{p\left(\theta|y_{obs}\right)}{p\left(\theta\right)}\right] p\left(\theta|y_{obs}\right) d\theta. \qquad (33)$$

With this definition,[31] a larger D_{KL} implies a greater difference between the prior and the posterior and, thus, signifies that more information has been gained. In this test framework, the information gain is subordinate to the success criterion, as it is only relevant if the inference is successful.

[31] *In some sense, S and D_{KL} correspond to the concepts of* accuracy *and* precision *in the context of measurement.*

4.4 SELECTING A NEW RATIO LIST

The ratio selection procedure is executed with four different configurations. The testing framework described above is used to compare the performance of the resulting four ratio list and select the most useful. This list will be used in Chapter 5 to investigate more complex inference scenarios.

4.4.1 *Tested algorithm configurations*

In the following, algorithm configurations were chosen based on preliminary testing. The two best combinations *combinatorial* algorithm and fitness function, and the two best combinations the *genetic algorithm* and fitness function were chosen, yielding a total of four algorithm variants that are tested in dept:

COMBINATORIAL-f_1 refers to the combinatorial algorithm with fitness function f_1. This is the performance measure used by Figueroa [29] in his thesis.

COMBINATORIAL-f_4 refers to the combinatorial algorithm with fitness function f_4. Preliminary testing showed this to produce superior results to f_1.

GENETIC-f_2 refers to the genetic algorithm with fitness function f_2.

GENETIC-f_8 refers to the genetic algorithm with fitness function f_8.

The length of the ratio list was fixed to 30 in all four variants. This choice was made to preemptively mitigate convergence issues, as experience has shown that longer ratio lists frequently cause problems for the MCMC sampler with complex models like the ones used in Chapter 5. Furthermore, limiting the ratio list length to 30 permits a fairer comparison to Figueroa's ratio list, since the length of the ratio list significanly impacts the performance.

All four variants used a *model-averaging* objective function that considers three reactor models (Magnox, PHWR, and PWR) to produce good performance in inference models that simultaneously consider these three reactor types.

4.4.2 *Ratio selection results*

The test framework was executed for each of these ratio lists with $N_{tp} = 40$ and $N_{sam} = 20$; the perturbation was samples with a relative standard deviation $\sigma_{sam} = 10\,\%$. 15 % relative standard deviation was used in the likelihood to account for additional modelling error. The procedure was repeated for the three different reactor types, i. e., three sets of test points and synthetic measurements were used.

Figure 3a shows the number of converged test runs N_{conv} for each test scenario. The boxplots indicate the distribution of N_{conv} across the 20 samples, that is, each boxplot summarizes 20 points. Here, the two genetic-algorithm-derived ratio lists have much higher rates of convergence than the combinatoric-algorithm-derived ratios. The latter perform especially poorly in the Magnox test scenario, although this observation defies explanation as of now.

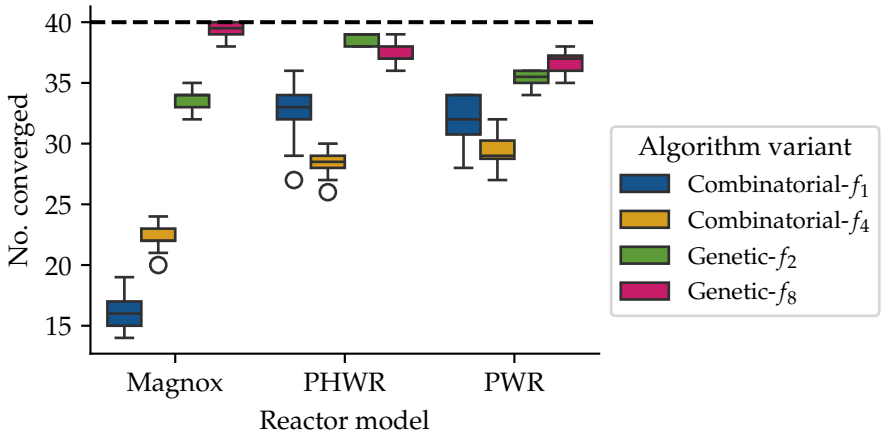
The distribution of the normalized *success* rate are shown in Figure 3b. Success is evaluated independently for each of the four inference parameters according to Equation (32) and the number of successes is normalized to the number of converged runs. Here, there is less variation than in the convergence rate; the normalized success rate is above 0.95 on average for all parameters, reactor types and ratio lists. Thus, convergence is a strong indicator for successful parameter reconstruction. At the same time it is the more useful criterion for determining the suitability of a ratio list.

The expected information gain is measured by the KL-divergence (see Section 4.3.1). Figure 4 shows the KL-divergence averaged over the test points and distributed over the perturbed samples. A larger KL-divergence implies a greater difference between the prior and the posterior, which in this case means the posterior is "narrower" and contains more information.

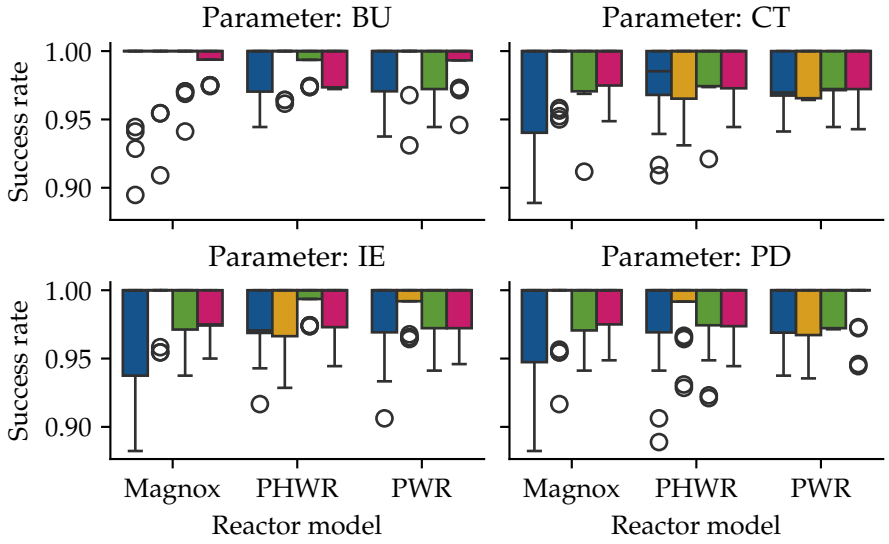
A clear difference can be observed between the parameters BU and IE compared to PD and CT. Whereas the two genetic-algorithm-derived ratio lists tend towards a larger KL-divergence on BU and IE, they produce much lower KL-divergence distributions than the combinatoric-algorithm-derived lists on CT and PD.

Given the above results, I chose the `genetic-f8` ratio list as the best ratio list as it demonstrates the best convergence and consequently leads to more successful reconstructions.[32] In the context of verifying fissile material stockpile declarations, the parameters BU and IE are more important that CT and PD. The former are directly correlated to fissile material production (BU) or consumption (IE), whereas the

[32] *The total success rate of this ratio list is highest compared to the other ratio lists due to its higher convergence rate.*



(a) Convergence rate of inference test runs with four algorithm-derived ratio lists. The y-axis indicates the number of test points from the test space with converged posterior distributions. The horizontal dashed line shows the number of test points N_{tp} in the test dataset, which implies the maximum number of converged posteriors.



(b) Success rate of inference test runs with four algorithm-derived ratio lists. The y-axes show the number of test points that were successfully reconstructed normalized to the number of converged test points.

Figure 3: Performance statistics of inference tests of four algorithm-derived ratio lists with $N_{sam} = 20$ and $N_{tp} = 40$. The boxplots show the distribution of the statistics over the samples repetitions. The x-axes separate the results by reactor type and the color indicates the algorithm configuration that selected the ratio list.

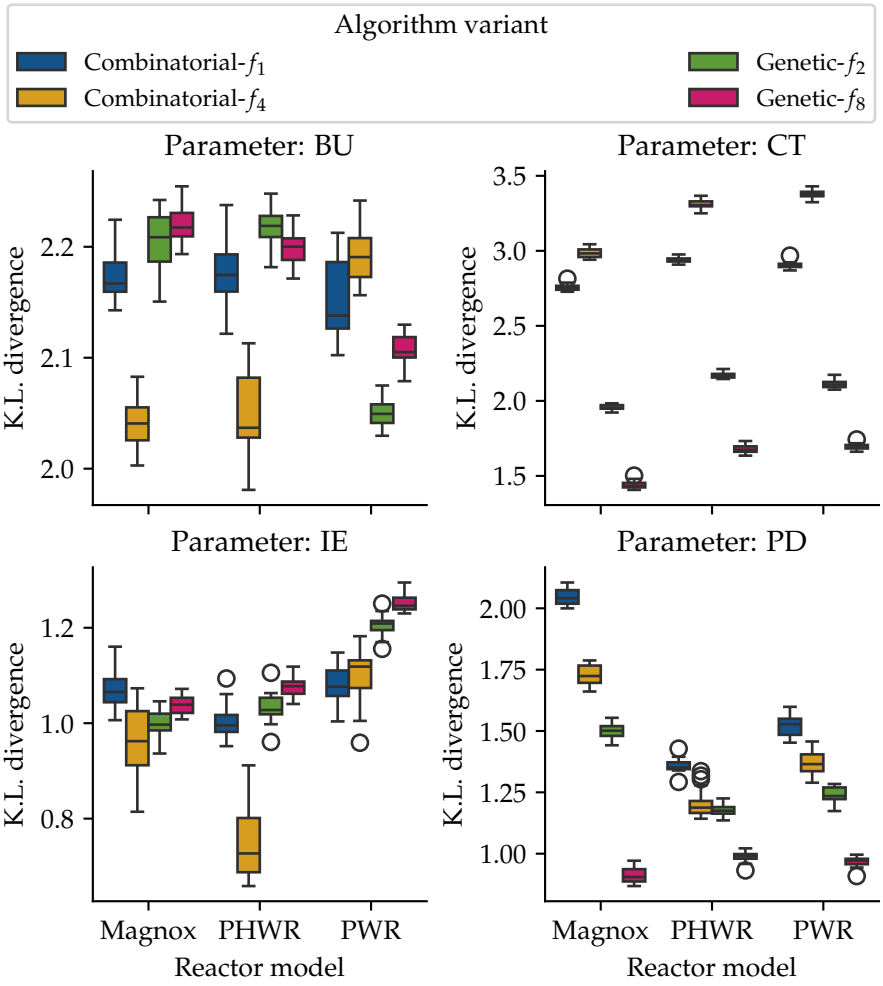


Figure 4: Information gain in inference test runs with four algorithm-derived ratio lists. The y-axes show the value of the *Kullback-Leibler divergence* between the prior and the posterior distribution. The four graphs show the results for the four inference parameters. The x-axes separate the results by the reactor model and the colors refer to the ratio selection algorithm.

Table 2: The list of ratios selected with the genetic algorithm using fitness function f_8.

Samarium	Cadmium	Uranium	Gadolinium
$^{147}Sm/^{149}Sm$	$^{111}Cd/^{113}Cd$	$^{234}U/^{236}U$	$^{155}Gd/^{157}Gd$
$^{147}Sm/^{154}Sm$	$^{112}Cd/^{113}Cd$	$^{234}U/^{238}U$	$^{155}Gd/^{160}Gd$
$^{149}Sm/^{151}Sm$	$^{113}Cd/^{114}Cd$	$^{235}U/^{236}U$	
$^{149}Sm/^{152}Sm$	$^{113}Cd/^{116}Cd$	$^{235}U/^{238}U$	
$^{149}Sm/^{154}Sm$	$^{113m}Cd/^{114}Cd$	$^{236}U/^{238}U$	
$^{150}Sm/^{151}Sm$	$^{113m}Cd/^{116}Cd$		
$^{151}Sm/^{154}Sm$			
Neodymium	**Tin**	**Barium**	**Molybdenum**
$^{143}Nd/^{144}Nd$	$^{115}Sn/^{119}Sn$	$^{136}Ba/^{138}Ba$	$^{100}Mo/^{96}Mo$
$^{144}Nd/^{145}Nd$	$^{121m}Sn/^{126}Sn$		
Lead	**Tellurium**	**Yttrium**	**Zirconium**
$^{105}Pd/^{110}Pd$	$^{126}Te/^{130}Te$	$^{89}Y/^{90}Y$	$^{90}Zr/^{96}Zr$

latter need to be modelled to account for their influence on the observable ratios. Therefore, the lower information gain of the `genetic-f8` ratio list for those parameters is tolerable. This set of ratios is listed in Table 2.

4.4.3 *Discussion of selected ratios*

The genetic algorithm is a probabilistic process that is not guaranteed to find a global optimum; repeating the selection algorithm with different starting conditions could lead to different ratio lists. The ratios in Table 2 are the best ratios of the present thesis and will be studied further in the following chapter. This set of ratios warrants some discussion.

Though an optimistic detection threshold was chosen in Section 4.1, none of the nuclides in the chosen ratio list in Table 2 occur in concentrations close to the detection limit in any of the three simulated dataset. Their respective concentrations are, on average, greater than 1×10^{-10} g per gram of HAR. Only a few nuclides have individual datapoints as outliers close to the approximated detection threshold (1×10^{-12}). Thus, it is likely that the chosen ratios will be present in the HAR in detectable quantities.

Nonetheless, this does not guarantee that they can be measured with high precision. In the selection algorithm's objective function, each ratio was given the same relative uncertainty of 10 %. For many ratios, the uncertainty will be much lower, and for some it will likely

be larger. Though expert analysis for each individual nuclide would be required, the algorithm can, in principle, take varying individual uncertainty into account. However, such an analysis was beyond the scope of this thesis.

In the published literature on SNF characterization, uranium, plutonium, and neodymium are used frequently to determine BU [62, 71, 72]. Most of the relevant neodymium and uranium isotopes are included in the selected ratio list, though plutonium isotopes are notably absent. Furthermore, radionuclides commonly used to characterize cooling time, for example, ^{134}Cs, ^{137}Cs, or ^{154}Eu [62, 73] are also absent. Some of these, e. g., ^{144}Ce, are excluded due to the condition that they must exceed the threshold concentration of one part-per-trillion in the HAR after 10 000 d of cooling. Others simply do not pass the genetic algorithm.

The selected list does contain isotopes of samarium and barium that have been studied for use in nuclear forensics [74] to determine the origin of and characterize separated plutonium, which contains trace amounts thereof. The selected ratio list also contains several nuclides that are not common in the literature. Most of these are stable or very long-lived nuclides in the decay chains of fission products and are strong indicators for burnup and are somewhat sensitive to variations of the initial enrichment. Only ^{90}Y and ^{113m}Cd decay at rates that are relevant in the given parameter space. These characteristics reflect the observation in the previous section that the ratios are more sensitive to BU and IE than CT and PD.

The selected ratio list should be taken as a theory-based proposition. The simulation-based test cases in Chapter 5 will demonstrate its potential utility on a theoretical level; experiments will need to determine if it can be used in practice. Nonetheless, the principle of the selection algorithm will be useful even if some of the nuclides considered here need to be excluded from its list of candidates.

Part III

TESTING, VALIDATION AND APPLICATIONS

5

SIMULATION-BASED TESTING

Compared to the Graphite Isotope Ratio Method (GIRM), the Bayesian Reprocessing Waste Analysis Method (BRAM) faces two additional challenges due to using reprocessing waste as the basis for reconstructing reactor parameters.

First, a reprocessing plant may process many *batches*[33] of spent fuel, and the waste may be stored and mixed in the same container. The nuclide signature of a sample of such a mixture would differ to that of individual batches and require more complex modelling.

[33] *A "batch" of fuel refers to fuel that was irradiated under the same or similar conditions and shares a set of average operating parameters.*

In his thesis, Figueroa [29] addresses this challenge with his "two-batch scenarios" and introduces a likelihood function to reconstruct the Burnup (BU) and Cooling Time (CT) of two batches of waste from a Pressurized Water Reactor (PWR). This chapter builds on his analysis by adding the Initial Enrichment (IE) and the average Power Density (PD) as two additional inference parameters and studying test cases that include two additional reactor types: a Magnox reactor and a Pressurized Heavy-Water Reactor (PHWR). A preliminary version of this research has been published in conference proceedings [75].

The second challenge is the fact that the type of reactor that produced the waste that is sampled may not be known with certainty and may need to be reconstructed as well. I address this question with a second test scenario in which the source reactor type is reconstructed simultaneously with the operating parameters. This work builds on and extends concepts I published in an earlier article [76].

This chapter first describes the three reactor types and their implementation as computational models. Then, two different inference scenarios are discussed. The first extends the previous research on "two-batch mixtures", and the second studies reactor type reconstruction. Each section first explains the theoretical model before presenting the results of systematic, simulation-based performance tests. Both performance test sets are performed with two sets of isotopic ratios: the set used in Figueroa's thesis [29] and the set of ratios derived earlier in this thesis with the new genetic algorithm (see Table 2 in Chapter 4). Comparing the results obtained with these two ratio lists will be instructive regarding the value of the ratio selection algorithm.

5.1 REACTOR MODELS

The inference scenarios in this chapter consider three reactor types. The first is the Magnox type reactor. This type of nuclear reactor is cooled with CO_2 gas, moderated with solid blocks of graphite, and

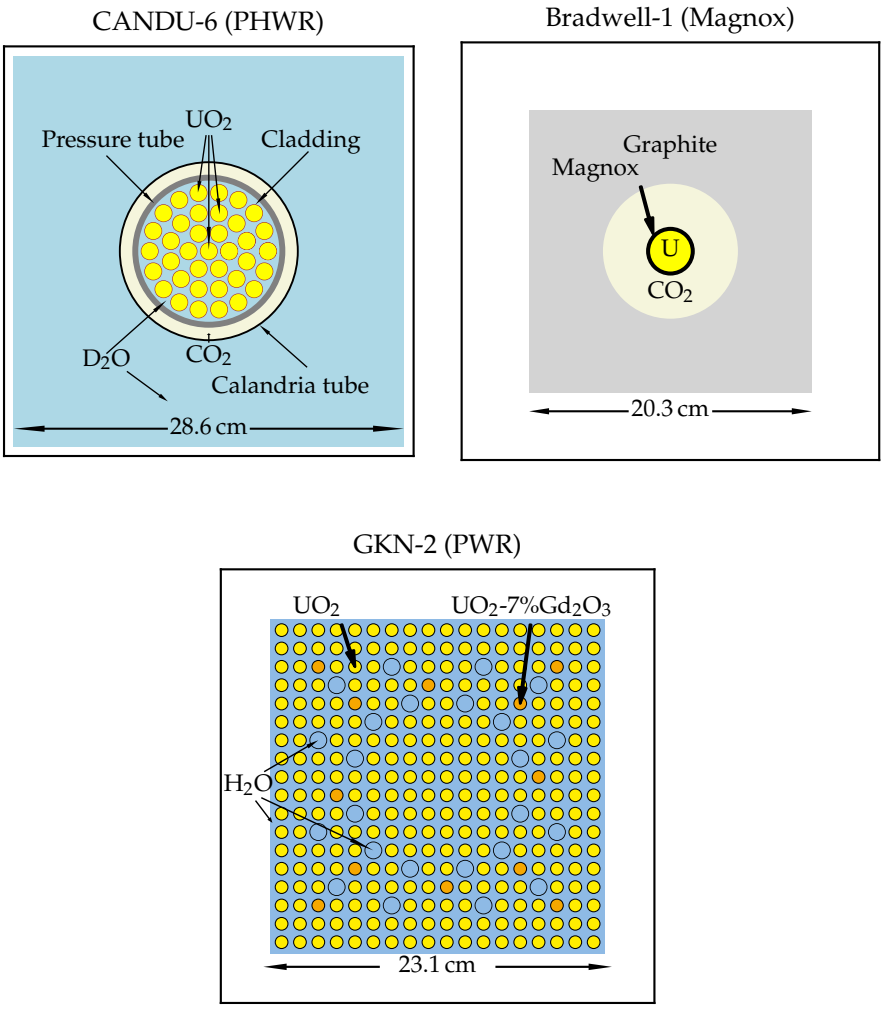


Figure 5: Unit cells of CANDU-6 (PHWR), Bradwell-1 (Magnox) and GKN-II (PWR) reactor models, as implemented in `OpenMC`. Reflective boundary conditions are applied at the borders of the colored areas; he black borders serve to visually compare the size of the unit cells.

fuelled with natural uranium metal. Its name stands for Magnesium Non-OXidising, which refers to the magnesium oxide that was initially used as cladding for the fuel. This type of reactor was used during the early years of the United Kingdom's nuclear programme for commercial electricity production and to produce plutonium for military purposes. The only remaining operational Magnox reactor is the 5 MWe reactor in North Korea, which is of particular interest from a nuclear archaeology perspective and will be discussed in more detail in Chapter 7.

The second type of reactor is the Pressurized Heavy-Water Reactor (PHWR). This reactor model uses D_2O as coolant[34] and as moderator material. It is usually fuelled with UO_2, either with natural or slightly enriched ^{235}U content. Historically, PHWR type reactors, such as those at the Savannah River plant [77], have been used to produce plutonium and tritium for nuclear weapons programs. CANada Deuterium Uranium (CANDU) is a line of reactors designed for commercial power production operated mostly in Canada, but also in other parts of the world, such as India and China. This work models a CANDU-6 reactor, whose design is well documented in publicly available literature [78].

[34] *Some PHWR designs also use H_2O as coolant, but these are no longer common.*

The third type of reactor is the Pressurized Water Reactor (PWR), which is the most common type reactor and usually used for commercial power production. It uses H_2O as its coolant and moderator and, therefore, requires fuel with enriched ^{235}U content. The enrichment level varies from 2 % to 5 % across different PWR designs. Though PWRs have historically not been used for military plutonium production, I include them in this analysis for two reasons. First, PWRs are in widespread use so it is not unlikely that nuclear archaeological or nuclear forensic investigations encounter samples of their waste, either in mixtures or separately. Second, the current nuclear weapons states built dedicated plutonium production reactors for their military programs; in the future, rogue states may proliferate with the reactor technology they already have, which in many cases would be a PWR.

For the following test scenarios, I modelled one reactor of each abovementioned type: a generic CANDU-6, a Magnox reactor from the United Kingdom called *Bradwell-1*, and a PWR from Germany called *Gemeinschaftskernkraftwerck Neckarwestheim 2* (GKN-II). Each reactor was modelled as an infinite lattice of a unit cell of the respective core, with reflective boundary conditions. I implemented each model in `OpenMC` [36], using cross-section, fission yield and, nuclear decay data from the ENDF/B-vii.1 library [46]. A schematic of each unit cell is shown in Figure 5 and corresponding material parameters are listed in Table 3. The table lists nominal values for enrichment, although it is varied during the process of generating training data for surrogate modelling purposes.

Table 3: Material parameters of the modelled reactor types [78–80]. Cladding and insulator materials, though implemented in the models, are omitted here for brevity's sake.

	Magnox	PHWR	PWR
		Fuel	
^{235}U content [%-wt U]†	0.711	0.711	3.789
Material	U metal	UO_2	UO_2
Density [g cm^{-3}]	18.5	10.6	10.4
Temperature [K]	700	960	1027
Fuel diameter [cm]	2.9	1.215	0.805
Pins per channel / assembly	1	37	288
		Burnable absorber	
Material	–	-	UO_2-7%Gd_2O_3
^{235}U content [%-wt U]†	–	–	2.6
Density	–	-	10.13
Temperature [K]	–	–	1027
Pin diameter [cm]	–	–	0.805
Pins per assembly	–	–	12
		Moderator	
Material	Graphite	DO_2	H_2O‡
Density [g cm^{-3}]	1.6	1.0085	0.7005
Temperature [K]	530	343	575
		Coolant	
Material	CO_2	D_2O	H_2O
Density [g cm^{-3}]	8.57×10^{-3}	0.8143	0.7005
Temperature [K]	558	563	575

† Nominal enrichment values. In the `OpenMC` models, fuel enrichment was variable according to Table 4

‡ There is no distinction between coolant and moderator in a PWR.

5.1.1 *Surrogate models*

The reactor models described above were used to generate three sets of training data for Gaussian process-based surrogate models, which were trained according to the procedure described in Section 3.4. Input samples were generated to uniformly cover the four-dimensional parameter space of (BU,CT,PD,IE) bounded by the limits in Table 4.

These bounds were chosen to reflect plausible ranges for each parameter, although two limits deviate from this principle.

Table 4: Upper and lower bounds of the uniform prior distributions in the performance test scenarios.

	Magnox	PHWR	PWR
Burnup [MW d kg^{-1}]	0.09–9	0.09–9	0.09–9
Power density [W g^{-1}]	0.1–4	10–40	10–60
Cooling time [day]	1000–10,000	1000–10,000	1000–10,000
Initial enrichment [% ^{235}U]	0.711–1.5	0.711-1.5	2–5
Mixing ratio† (α)	0.5–1.5	0.5–1.5	0.5–1.5

† Only relevant for the *Mixture* scenario.

First, the burnup of the PWR model is bounded at 9 MW d kg^{-1}, mirroring the Magnox and PHWR bounds, which correspond roughly to the highest burnup that can be achieved with natural uranium fuel. For PWRs operated in commercial settings, this bound is low, as they typically attain burnup values between 20 MW d kg^{-1} to 50 MW d kg^{-1}; although the fuel could potentially be discharged at lower values. Considering military applications, low burnup ($\leqslant$ 5 MW d) is more interesting, because the lower ^{240}Pu content is more favorable for building nuclear weapons. Considering such a wide range of values (from low, military-application relevant BU to high, civilian-power production levels) thins the density of training data points at the low end of the range and makes the surrogate models less precise. Therefore, I restricted the burnup range for the PWR to the same bounds as the other two reactors.

Second, the enrichment of the Magnox reactor is varied up to 1.5 %. Magnox reactors did not typically operate with enriched fuel, although there is some indication that the United Kingdom used a Magnox reactor to produce tritium for its nuclear program, which would likely require enriched fuel. A similar scenario has been discussed for the North Korean 5 MWe reactor. Therefore, the enrichment of the Magnox model is varied as well.

5.2 THE EXTENDED "TWO-BATCH" SCENARIO

The two-batch mixture scenario engages with the challenge that reprocessing waste samples can consist of batches of waste with different irradiation histories. These batches can originate from different operating cycles of the same reactor or from different reactors. While this scenario remains limited to mixtures of two batches, the complexity of the inference task is increased compared to previous research:

first, by including two more reactor types, i. e., Magnox and PHWR; and second, by adding two inference parameters, that is, Power Density (PD), and Initial Enrichment (IE).

5.2.1 *A likelihood function for mixed-waste samples*

The likelihood function discussed in Equation (22) requires a model to predict isotopic ratio values. A generalized model M_r for the isotopic ratios in a mixed waste sample can be defined as:

$$M_r(\theta) = \frac{\sum_b^B \alpha_b M_{i,b}(\theta_b)}{\sum_b^B \alpha_b M_{j,b}(\theta_b)}, \tag{34}$$

with the batch index b, the number of batches B, the nuclide indices i and j, the mixing factor α_b, and the inference parameters θ_b. The models M_b of the individual batches can be of the same reactor type or of different reactor types, although, for the following study, the number of batches is limited to two and the correct reactor types are fixed and known.

Each batch has four individual inference parameters:

$$\theta_b = (BU_b, CT_b, PD_b, IE_b)$$

and is weighted by a mixing factor α_b that describes the relative amounts of material from each batch that were mixed. Here, both mixing factors are variable inference parameters.[35]

[35] *In contrast, Figueroa [29] defines $\alpha_1 = 1$ as fixed and only lets α_2 vary.*

Thus, there are ten inference parameters in total: $\theta = (\alpha_1, \alpha_2, \theta_1, \theta_2)$. Their prior distributions are modelled as uniform distributions between the bounds listed in Table 4. These bounds are more restrictive than those used by Figueroa [29]. Most notably, α_b runs from 0.5 to 1.5, restricting the maximal mixing ratio to 1:3. In Figueroa's thesis, α_b varied from 0.1 to 10, which corresponds to maximum mixing ratio of 1:10. In either case, the mixing factor priors are selected somewhat arbitrarily, with the modest purpose to introduce the concept of this parameter, not to describe it as realistically as possible.

With the surrogate models, likelihood function, and prior distribution in hand, the next step is to apply the inference framework to some test cases and evaluate how well it performs.

5.2.2 *Mixture model testing*

To systematically test the performance of this inference model, I use the test framework outlined in Section 4.3. One test dataset is created for each of the 6 possible pairwise combinations of the three reactor types; each dataset contains $N_{tp} = 100$ different ground truth inputs, with $N_{sam} = 20$ simulated measurements each. The nuclide concentrations of the simulated measurements are first computed with the

respective individual surrogate model and then added together with sampled weights that correspond to the mixing factor. Then, the ratio values are computed and perturbed with the Gaussian noise to add synthetic measurement error. The perturbation uses a 10 % relative standard deviation. In contrast, the inference likelihood uses 15 % relative standard deviation to include model uncertainty in addition to the measurement error. This value does not represent a methodical uncertainty analysis, rather, it is a rough approximation. The results in Chapter 6 provide some evidence that it is a reasonable assumption.

The inference model is applied to each of the test cases. Each run comprises four MCMC sampling chains, with 1000 samples each. Before drawing the samples, 2000 sampling steps are used to tune the hyperparameters of the NUTS algorithm. After sampling, the performance criteria are computed (Section 4.3.1). First, convergence is assessed, then, success is determined, and finally, the information gain is estimated.

The graphs in Figure 6 exemplify the result of one test case. It shows the posterior distributions of the inference parameters with both ratio lists for a PHWR-PWR mixture sample. This example is considered a *successful* reconstruction, as the true values are inside the 95 %-HDI of the respective posterior distributions of all ten parameters. However, some parameters are reconstructed more precisely than others, i. e., parameters are reconstructed with varying information gain. For example, BU and IE are reconstructed precisely as clearly separated, peaked distributions, that cover only parts of the prior space. PD and α are reconstructed less precisely, especially with the genetic-algorithm-derived ratio list of this thesis. The CT posteriors computed with Figueroa's ratio list are also sharper than those obtained with the genetic algorithm ratio list. These observations reflect a general trend that is observable in the statistics discussed in the following.

The aggregated results of the convergence assessment are shown in Figure 7, which shows two remarkable trends. First, the convergence rate is significantly higher in test cases with two different reactor types. Second, the convergence rate is higher when the new, genetic-algorithm derived ratio list is used. Overall, the variance of the convergence rate over the sampled repetitions is small. The lower convergence rate seems correlated to test cases with low ground truth BU, which is shown in more detail in Figure 23 in Appendix B.1.

The aggregated results for success are displayed in Figure 8, normalized to the convergence rate. Generally, the success rate is high ($\approx$ 90 %), although there is a slight distinction between test scenarios with the same reactor type and scenarios with different reactor types. Scenarios with different reactor types have a slightly higher success rate for all parameters, irrespective of which reactor types. Contrary

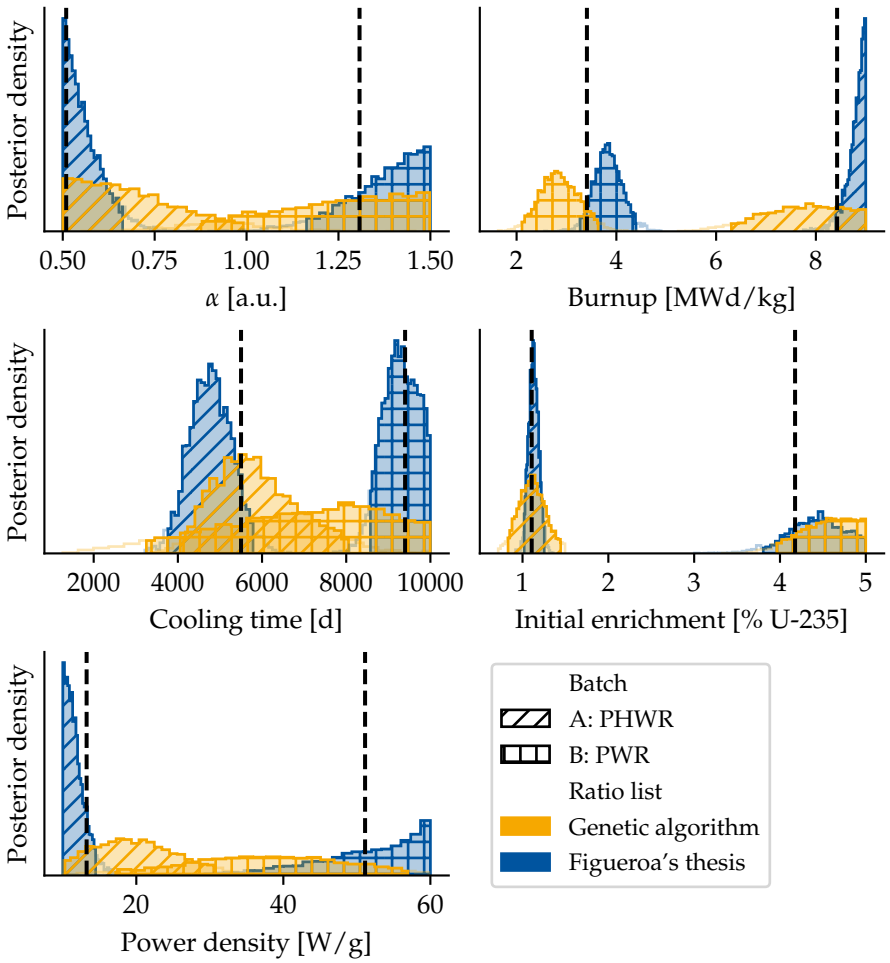


Figure 6: Posterior distributions of a "two-batch" scenario test case with a PHWR-PWR mixture. The hatches indicate the reactor type that each parameter is associated with, and the color indicates the ratio list used to infer the posterior. The shaded are under each graph indicates the 95 %-HDI, and the true parameter values are represented by the vertical dashed lines. This graph shows a successful reconstruction, because the 95 %-HDI intervals include the respective ground truth values.

to the convergence rate, the normalized success rate statistics show no significant difference between the two ratio lists. It should be noted that the total number of successfully reconstructed test cases is much lower with Figueroa's list, because the convergence rate is much lower.

The information gain is displayed in Figure 9, which shows the KL-divergence averaged over the test point set, distributed over the samples, for each parameter, reactor type combination and ratio list. Generally, BU is reconstructed with the most information gain and α with almost no information gain. There is a slight trend towards higher information gain with Figueroa's ratio list, than with the genetic-algorithm-derived ratio list. This effect is most visible for the param-

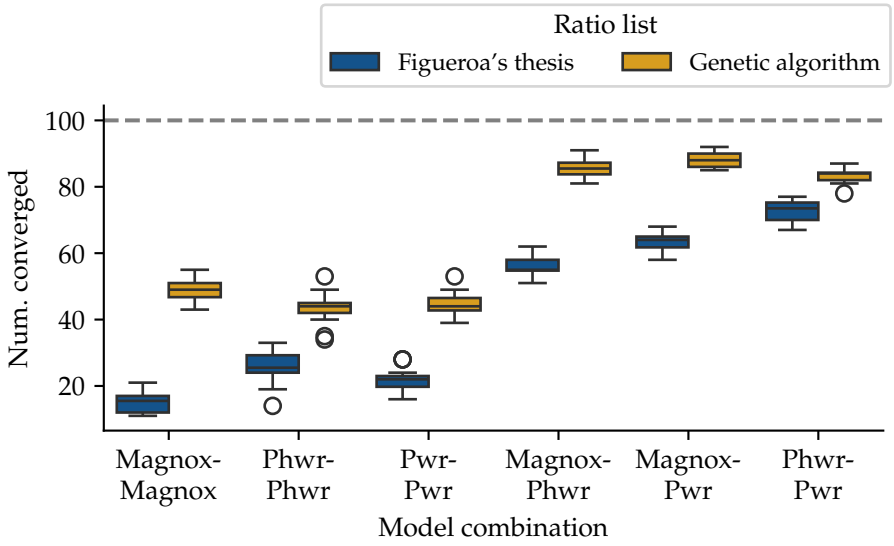


Figure 7: Convergence rate of the mixture inference model by reactor type combination. The y-axis shows the number of test points with converged posterior samples (max. 100). The boxplots indicate the distribution over the perturbed samples. The color of the boxplot indicates the ratio list.

eter CT in all scenarios and also for PD in some scenarios. For BU and IE the difference between ratio lists is not significant.

5.2.3 *Discussion of the two-batch scenario*

The above results lead to a number of insights but also to some question. First, it is clear that this scenario tends to encounter convergence issues. In a specific application scenario, such issues could be mitigated by case-specific investigation of their cause. Convergence could be improved with adaptations to the surrogate model, the sampling algorithm, the ratio list, or even the prior distributions Such adaptations were not done here, as they require case-by-case analysis; such detailed analysis was not feasible in the context of this systematic parameter space scan.

Among the converged samples, the rate of successfully reconstructed parameters was generally high ($\approx 90\,\%$), which confirms that the new ratio list of this thesis, though optimized for the "one-batch" scenario, performs well in this, more complex, scenario. The improved convergence rate obtained with the genetic-algorithm derived ratio list compared to Figueroa's ratio list demonstrates the usefulness of the new ratio selection algorithm.

Each test point was tested with multiple Gaussian-noise perturbed synthetic measurements. The variance within the results of each test point was generally low, indicating that measurement uncertainty can be captured by the likelihood. However, there are instances where

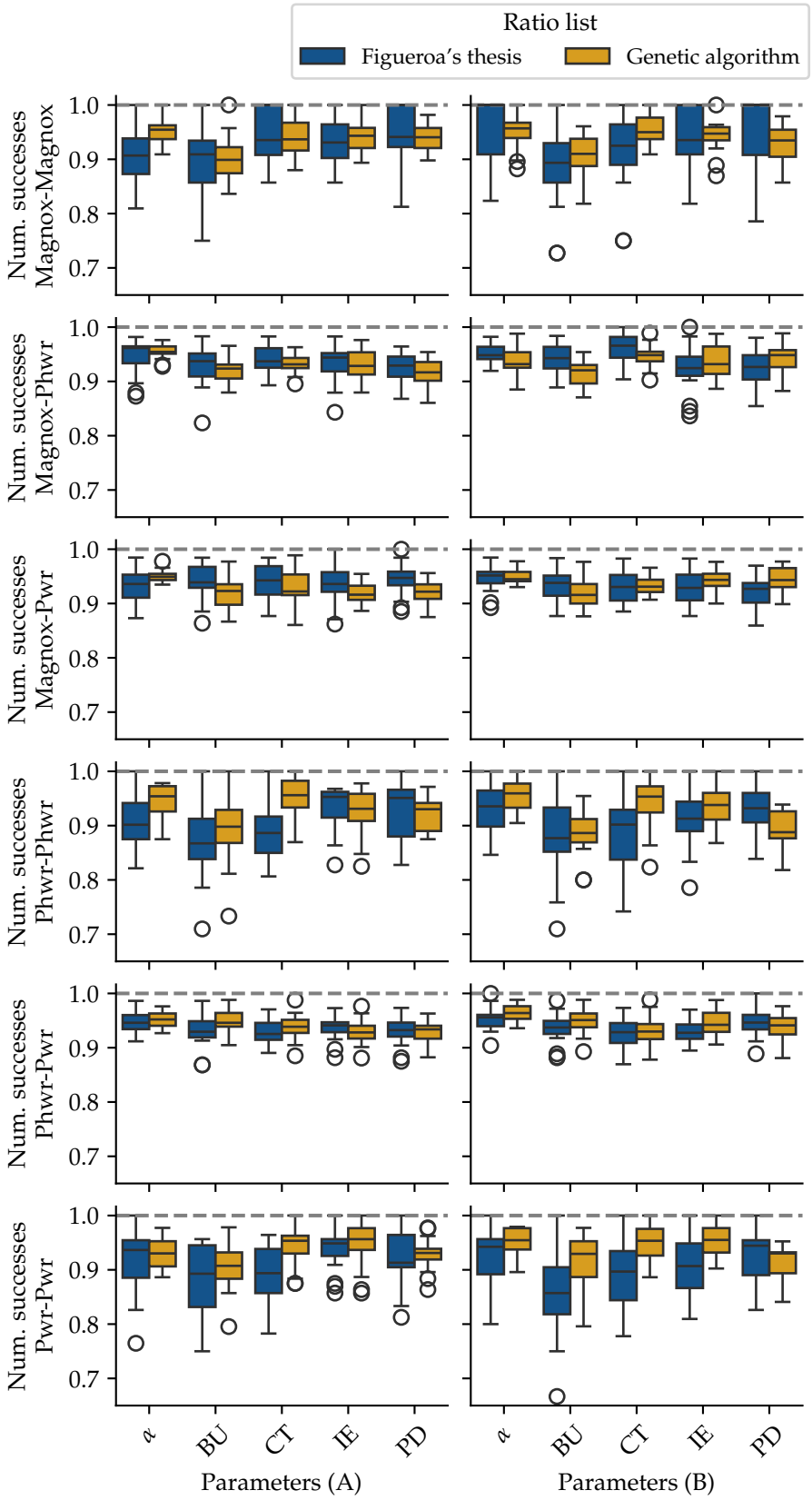


Figure 8: Distribution of successfully reconstructed samples for each pairwise combination of reactor types. The success rate is normalized to the number of converged test points. The dashed horizontal line indicates the maximum, i. e., a success rate of 1 implies all converged posteriors are successful. The boxplots show the distribution of the success rate over the 20 perturbed samples.

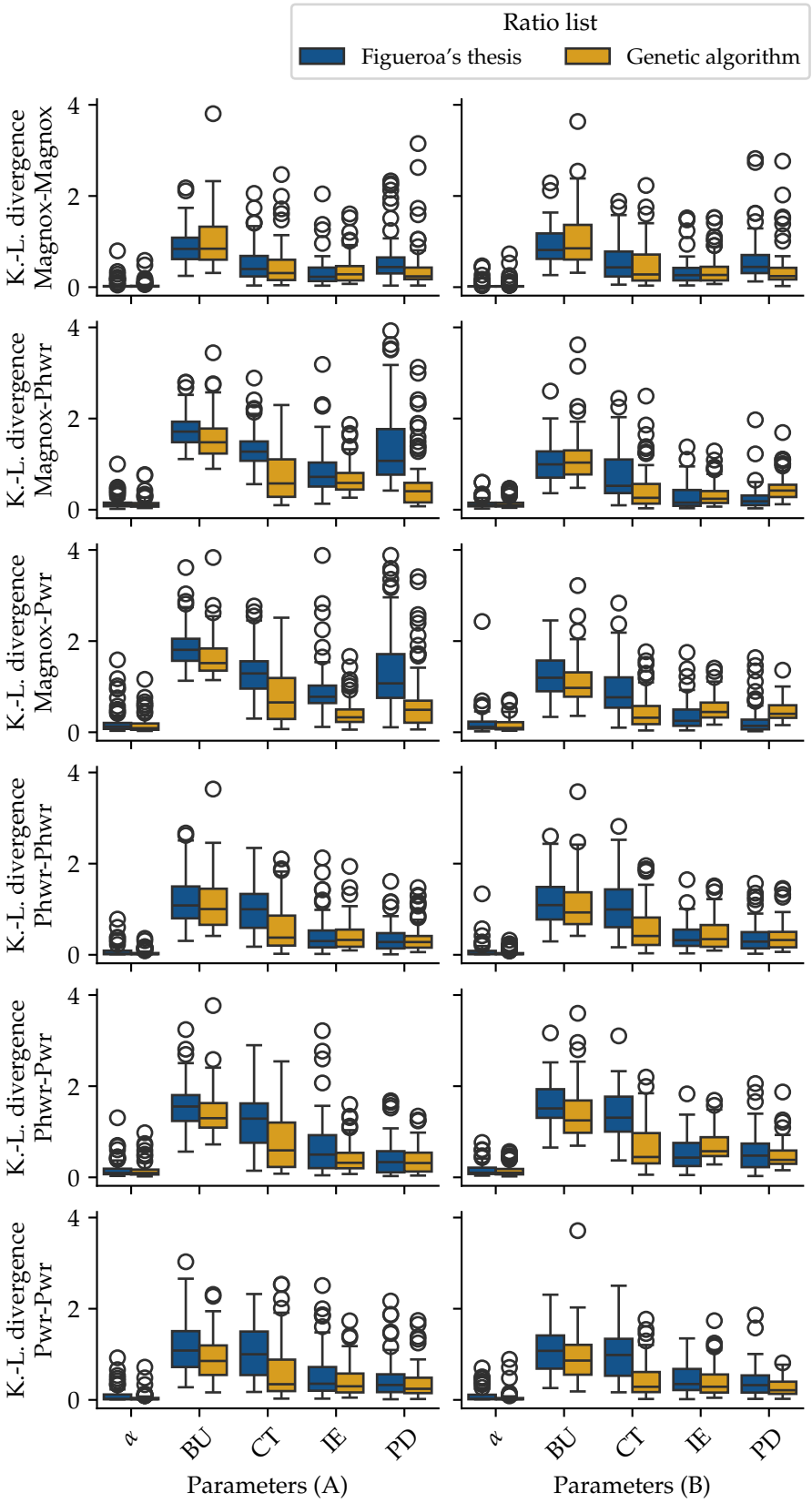


Figure 9: Distribution of Kullback-Leibler divergence between the posterior and the prior in each pairwise combination of reactor types. The Kullback-Leibler divergence is a measure of the information gain; larger values indicate more information gain. The boxplots show its distribution over the 20 perturbed samples.

the measurement error causes problems, as shown by the fact that for some test points there are samples that converge and others that do not. This effect appears despite the fact that the likelihood overestimated the relative error by 5 %.

There is a clear trend towards better overall performance in test scenarios with two different reactor types compared to test scenarios with the same reactor type. A likely explanation is the fact that each of the combinations of different reactor has at least one disjunct prior parameter space, which may help to separate the components of a mixture. The scenarios with the same reactor type are symmetrical with respect to the inference parameter spaces of each batch, so that a specific combination of parameter values has the same likelihood value irrespective of which batch index it is associated with. This problem may be mitigated by adding a step to the analysis to separate the batches in the posteriors. For example, a Gaussian mixture model have shown some promise in this regard in a Master's thesis that I supervised [81].

5.3 DISCRIMINATING REACTOR TYPES

In the "two-batch" scenario, the reactor types are assumed to be known prior to the analysis. This assumption does not usually hold for potential application scenarios, unless constraining information about a state's nuclear fuel cycle is available beforehand. Therefore, this scenario explores whether it is possible to use BRAM to determine the reactor of origin of a waste sample, while simultaneously reconstructing the operating history.

To develop this concept, I return to a "one batch" scenario; the supposed sample originates from a single batch of waste from one reactor, and the task is to reconstruct its irradiation history. It seems unlikely to find reprocessing waste that has not been mixed in the real world. The more likely "real-world" counterpart to this scenario may be nuclear forensics, where the origin of radioactive material needs to be determined in the context of criminal investigations. In nuclear forensics, the reactor type can sometimes be determined from morphological properties of the material [82] if the samples are in a solid state.

Thus, the scenario presented here is a sort of nuclear archaeology–forensics hybrid, postulating a sample of liquid HAR that cannot be attributed to its source from its morphological properties. The goal is not to present a realistic application scenario, but to extend and test the capabilities of the inference framework.

5.3.1 *Bayesian model classification*

In a sense, simultaneously inferring the reactor type and operating parameters is a combination of the typical "classification" and "regression" tasks in machine learning applications. Previously [76], I developed a model for this task based on hybrid *Gibbs-NUTS* MCMC sampling. The Gibbs sampler is used to sample a categorical variable that represents the reactor type and the NUTS sampler uses the corresponding reactor model to draw samples from the reactor parameters. The results were promising, however, the hybrid sampling can be limited by computational cost.

Here, I use a different approach. Posterior distributions are first sampled independently with "one-batch" likelihoods for each of the three reactor types. Then Leave-One-Out Cross-Validation (LOO-CV) (see Section 3.1) is used to compare, weight, and rank the models with a Bayesian model averaging method called *stacking* [50]. I used the implementation of these algorithms provided by the `arviz` [51] library.

The model with the highest weight is taken as the "predicted" model,[36] which can be evaluated as correct, incorrect, or inconclusive. The prediction is inconclusive if there is more than one model with equal weights that are the highest weight.

[36] *or* label, *to use typical classifier terminology.*

To reconstruct the inference parameters, a weighted sampling procedure is used to create a single weight-averaged posterior distribution for each parameter, containing samples from the three separate posteriors according to the weights assigned to their model types.

The priors, like in the previous section, are modelled as uniform distributions with the bounds listed previously in Table 4. There is no explicit prior on the reactor type, as no prior information on the weights is passed to the LOO-CV algorithm. Implicitly, the reactor type prior can be viewed as a categorical distribution over three variables with equal weights equal to 1/3.

5.3.2 *Classifier testing*

The test cases are set up analogously to Sections 4.3.1 and 5.2.2. There are 200 test points per "true" reactor type, with 20 samples per test point to simulate measurement uncertainties. Again, the likelihood uncertainty (15 %) is chosen to overestimate the sampling uncertainty (10 %), and the test cases are run with the both the genetic-algorithm-derived and Figueroa's ratio list.

In this scenario, the primary goal of the classification model is to correctly predict the reactor type label among three options: Magnox, PHWR, and PWR. Figure 10 shows two confusion matrices, one for each ratio list. Each element of the confusion matrix shows how many of

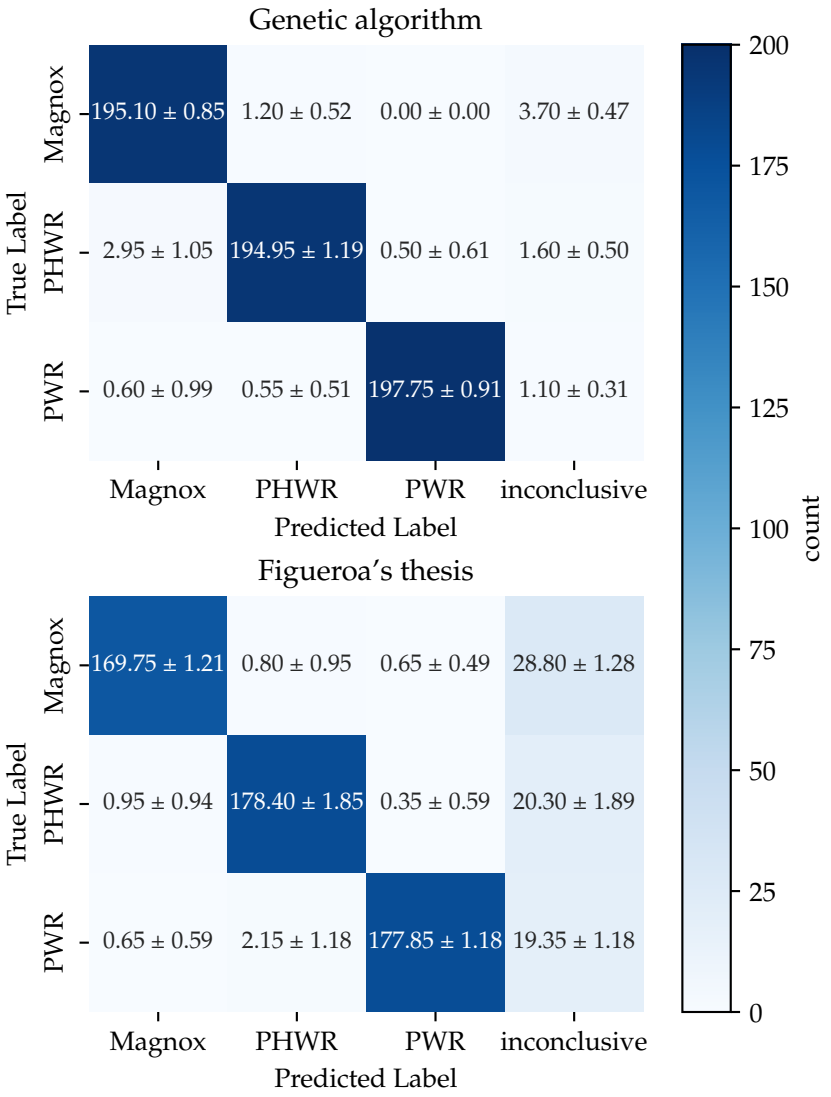


Figure 10: Confusion matrix of the classifier test scenario for both ratio lists.

the 200 true labels were identified as which label by the inference model, with values in each element averaged over the 20 samples.

Both confusion matrices are large diagonal elements, indicating a high rate of successful reactor type classification. The new, genetic-algorithm-derived ratio list performs slightly better, classifying around 195 test cases correctly for each reactor type. The old ratio list identifies between 170 and 180 labels correctly. Notably, the lower rate of successful classifications is not due to a higher rate of misclassification, but rather a higher number of inconclusive outcomes.

This outcome is acceptable, as it implies that an analyst would be prompted to investigate further if the classification is uncertain, rather than obtaining and incorrect result.

The secondary purpose of this model is to reconstruct the reactor parameters BU, CT, IE and PD. Posteriors for one test sample are shown

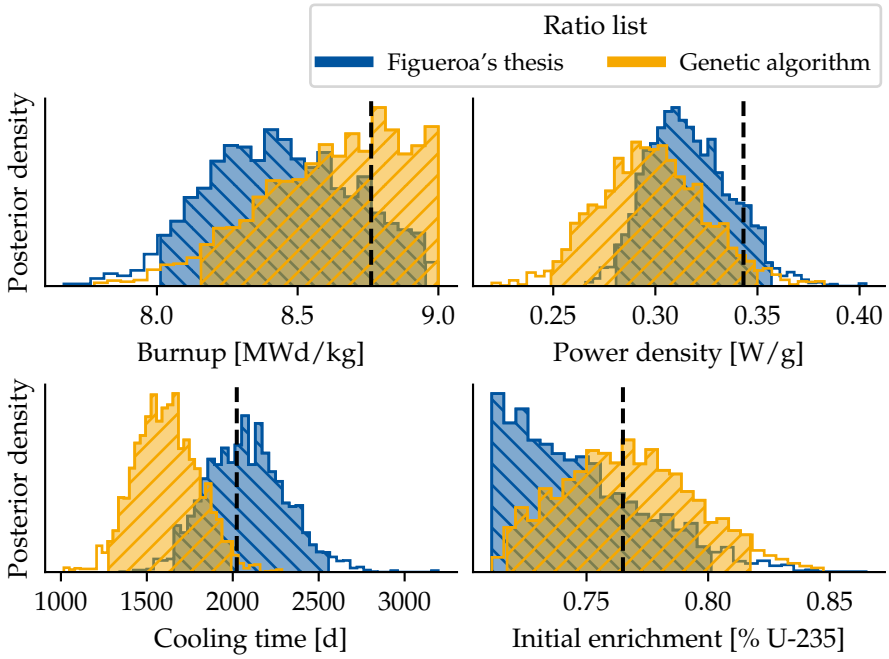


Figure 11: Example of posterior distributions of the classification model applied to a simulated sample of a Magnox reactor. The reactor type was identified correctly with both ratio lists. The histograms represent the posteriors probability density of the reactor parameters. The true parameter values are indicated by the dashed vertical lines and the shaded areas indicate the 95 %-HDI.

in Figure 11. The example is a correctly-labelled Magnox test case. BU and IE are reconstructed well with the new, genetic-algorithm-derived ratio list; PD and CT not as much, with ground truth values scraping the edges of the posterior probability density. Conversely, Figueroa's ratio list reconstructs CT well and its PD posterior density is shifted closer to the ground truth. Instead, the BU and IE posterior densities are shifted further away from the respective ground truth, although they are inferred successfully according to the 95 %-HDI-criterion.

Convergence statistics for the test scenarios are shown in Figure 12, which displays the number of converged posteriors for each true reactor type and both ratio lists. Generally, the rate of convergence is high ($\approx$ 95 %) with the genetic-algorithm-based ratio list and slightly lower with Figueroa's list. The latter shows remarkably lower convergence rate for test cases with the Magnox ($\approx$ 60 %) reactor model, compared to the PWR and PHWR models (both $\approx$ 80 %).

The same convergence statistics are displayed again in Figure 13, this time as a function of the ground truth parameter values. To make broad trends visible, the test points are grouped into bins and the convergence rates are averages for each parameter bin. While the convergence rate is roughly constant over the CT, IE, and PD ranges, it decreases towards lower BU values for all three reactor types. This

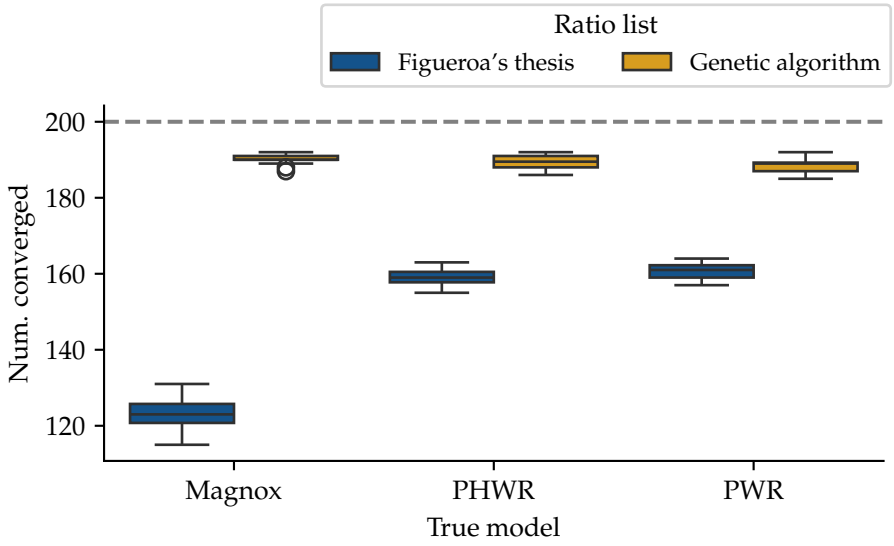


Figure 12: Convergence statistics of the weight-averaged posterior distributions of the reactor type classifier model. The y-axis shows the number of test points (out of 200) with converged posterior distributions. The boxplots indicate the distribution over the perturbed samples of each test point.

trend is stronger, with Figueroa's ratio list than with the new ratio list.

Figure 14a shows the aggregated success rates normalized to the convergence rates. The normalized success rate is high ($\approx 90\,\%$) for all three reactor types and all four inference parameters and, there are no obvious differences between the ratio lists. The normalized success rate does not show the same decrease towards low BU values as the convergence rate. The corresponding graphs are shown in Figure 38 in Appendix B.2. These results indicate that the reconstruction performs well if convergence is achieved and that the convergence criterion is useful for assessing the accuracy of the posterior distributions.

The information gain statistics are shown in Figure 14b, with small observable differences between the ratio list. As observed previously, Figueroa's ratio list produced slightly better information gain, i. e., posteriors with lower variance. This effect is most visible in for the parameters CT and PD. Overall, the information gain tends to be larger in this scenario than in the "two-batch" scenario, which is consistent with simpler inference problem.

5.3.3 *Discussing reactor type classification*

The results of this test scenario show that the Bayesian framework can, in principle, be used to discriminate between different reactor types with LOO-CV-based model comparison techniques. Simultaneous pa-

rameter inference works well, indicating that BRAM can be used to extract useful information from a sample even if the reactor of origin is unknown.[37]

Furthermore, the results underscore the utility of the new ratio selection algorithm, as the new ratio list consistently outperforms the previous one. Generally, low burnup ($\lesssim 3\,\mathrm{MW\,d\,kg^{-1}}$) values cause convergence problems for the framework, even though the scenario already limits itself to a burnup range below $9\,\mathrm{MW\,d\,kg^{-1}}$. This effect is likely due to inaccurate surrogate model predications at low burnup values. Gaussian process surrogates generally perform worse at the edge of their training parameter space. At low burnup, many of the nuclides are only present at very low concentrations; for some of these nuclides, the surrogate models predictions have a very large variance at low burnup, which can even lead to negative predictions[38]. This behavior conflicts with the gradient calculated by the sampler and thus lead to convergence issues. Using different kernels for the GP-models or adding additional training data for low burnup could mitigate this issue. The ratio selection algorithm could also be adjusted to place greater emphasis on that region of the parameter space and thus avoid problematic isotopic ratios.

Comparing the results obtained with the LOO-CV-based classification approach to the results obtained with the hybrid Gibbs-NUTS sampling approach [76] is difficult, due to differences in the test scenario setup and the ratio list. In the paper, I did not use the Gaussian-noise perturbed samples of ratios. Instead, only one run with the predicted ratio vector[39] was used, neglecting uncertainty modelling. Furthermore, a reduced version of Figueroa's ratio list was used, missing $^{94}\mathrm{Nb}/^{93\mathrm{m}}\mathrm{Nb}$, $^{93}\mathrm{Nb}/^{93\mathrm{m}}\mathrm{Nb}$, $^{132}\mathrm{Ba}/^{137\mathrm{m}}\mathrm{Ba}$, $^{213}\mathrm{Bi}/^{211}\mathrm{Bi}$, $^{209}\mathrm{Pb}/^{214}\mathrm{Pb}$, and $^{211}\mathrm{Po}/^{215}\mathrm{Po}$. This reduction was necessary due to poor surrogate model performance for the Magnox reactor type; an issue that has since been addressed by increasing the amount of training data.

The results presented in the paper appear better, with slightly higher classification success and parameter inference success and information gain. Whether these differences are due to the different classification approach or due to the uncertainty sampling added in this study cannot be determined due to the caveats mentioned above.

[37] *Besides the applications discussed in this thesis, this model may also be useful in nuclear forensics.*

[38] *Negative predictions are cast to zero.*

[39] *that would correspond to the mean of the Gaussian sampling distribution.*

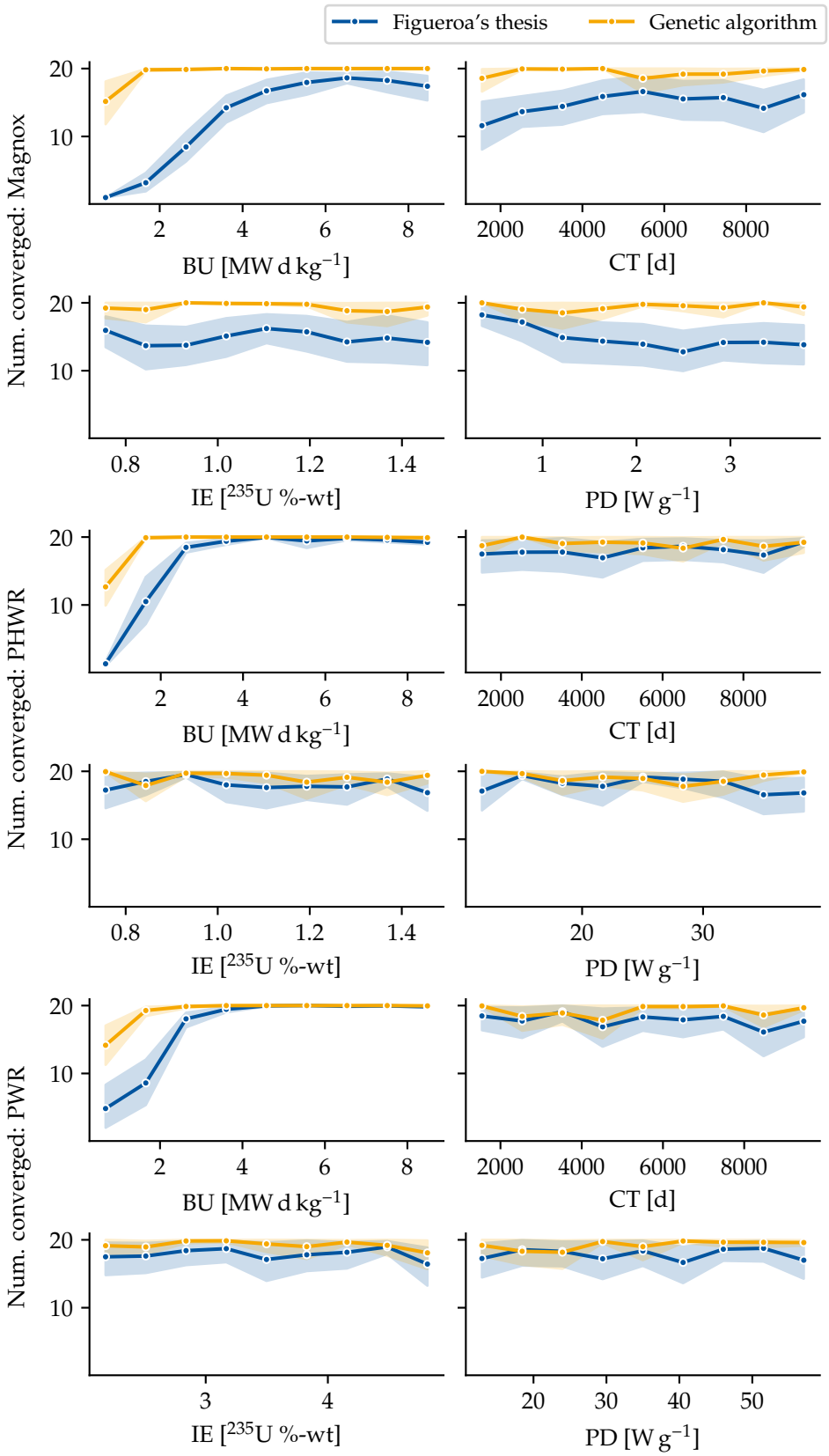
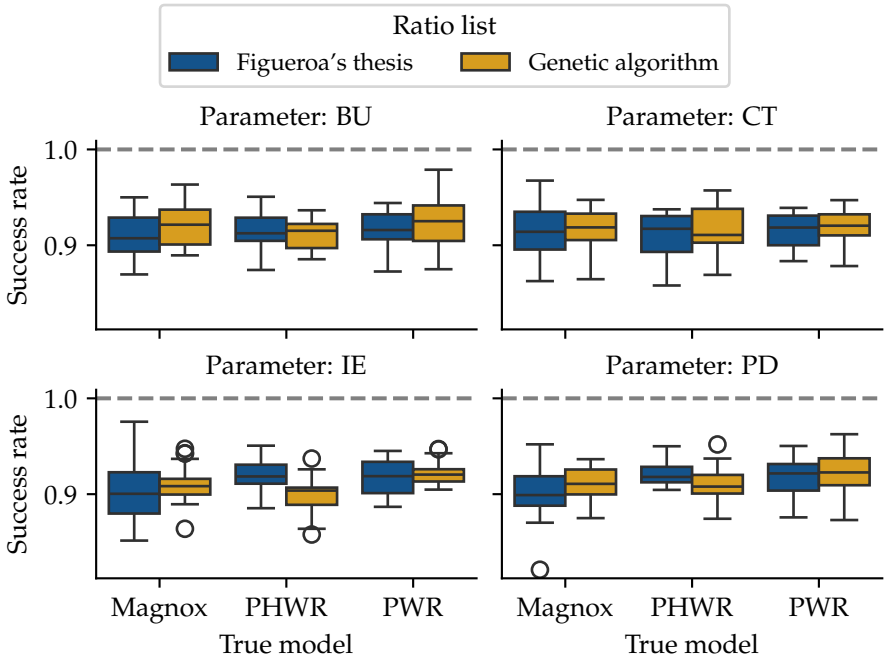
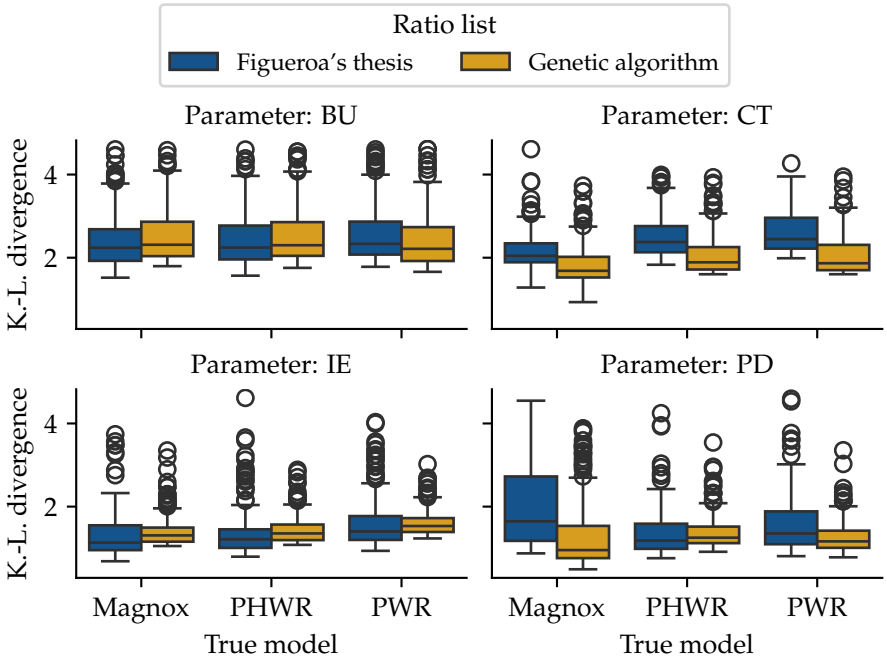


Figure 13: Convergence statistics dependent on ground truth parameter values.



(a) Summary statistic for successfully reconstructed parameters in the classifier test scenario. The success rate is normalized to the convergence rate. Misclassified test cases are included.



(b) Summary statistic of the information gain in the classifier test scenario, measured by the Kullback-Leibler-divergence between the posterior and prior distributions.

Figure 14: Reactor classification performance statistics

6

VALIDATING WITH SFCOMPO DATA

Previous research focused on developing the Bayesian Reprocessing Waste Analysis Method (BRAM) has relied almost exclusively on simulations for testing. That is, synthetic datasets of reactor parameter and isotopic ratio vector pairings were created with simulation tools, which are the same as (or similar to) those employed in the inference model. The performance of BRAM was evaluated by applying it to these datasets. This raises the valid question whether modelling uncertainties and biases due to the necessary model approximations and simplifications will reduce the method's accuracy or render it useless when confronted with experimental data from reprocessing waste samples.

Thus, while the simulation-based results are promising, validation with experimental data is desirable to demonstrate the method's reliability and is necessary to transition the framework from a theoretical concept to a practical tool. However, validation proves challenging because reprocessing waste is a dangerous substance due to its high radioactivity and its chemical nature. The necessary protective measures only further increase the already high cost of the measurement techniques[40] required for a precise analysis of its isotopic composition. Data from routine fuel monitoring activities and reporting obligations of reactor operators could be used instead, but these are rarely made transparently accessible to the broader public or academic research community.

[40] *E.g., mass spectrometry*

The exception that breaks the rule is the SFCOMPO database [83], which contains experimental isotopic composition measurements for spent nuclear fuel (SNF) from a variety of different reactor types, including Magnox reactors, PWRs, and PHWRs. The measurement data is accompanied by design information for the fuel assemblies and reactor core, as well as information on the operating conditions, which is relevant for modelling purposes.

The database contains data for SNF, which is not necessarily representative of highly active raffinate (HAR) from reprocessing. Also, the nuclides that have been assayed vary from sample to sample, and it is not possible to use those isotopic ratios that were selected in Chapter 4.

Nonetheless, for lack of other experimental data, I use data from SFCOMPO to advance the validation of BRAM. This chapter describes two validation test cases with respective datasets. The first uses experimental data from SNF assemblies of a PWR. The second test case uses data from Magnox reactor. The available dataset from PHWRs only con-

tained measurements of uranium and plutonium nuclides, which is not suitable for the purposes of this analysis. Therefore, no validation was performed on PHWR samples.

6.1 VALIDATION WITH PWR DATA

6.1.1 *Samples from the Obrigheim reactor*

Among the many database entries for different PWR reactors there is a set of ten samples from the Obrigheim reactor that are of particular interest to this validation endeavor. The samples, numbered 86–95, correspond to five different fuel assemblies that were irradiated concurrently. The assemblies were placed at different locations within the reactor core during different reactor cycles; each one has a unique power density history but a similar final burnup to the others. Hence, I consider these assemblies to be representative of an *average* assembly.

According to Radulescu et al. [80], each assembly was cut in half, and each half was dissolved completely before being assayed. If the dissolved assemblies were to be mixed in an imagined waste storage tank, any of the individual differences would disappear when the solution homogenizes and any samples taken afterward would reflect an average nuclide composition. Thus, by averaging the measurement data of these samples, I create a proxy for a reprocessing waste sample with a single, *average* irradiation history, which is to be reconstructed. The inverse of the squared measurement uncertainty was used to compute weighted averages of the measurement data and correspondingly propagated uncertainty estimates.

The sample dataset contains isotopic concentration data for nuclides of uranium, plutonium, americium and curium; atomic ratio data for nuclides of neodymium, krypton, and xenon; and the activity ratio of two caesium isotopes. Krypton and xenon are noble gases that are presumed to be unreliable evidence due to their propensity to volatilize and escape during reprocessing. They are excluded from this analysis to ensure that the results are not affected—positively or negatively—by ratios that would not be useable in the intended application. Although both of the americium measurements and one of the curium measurements are given with an uncertainty estimate of 100 %, they are included in the analysis because averaging the samples reduces this uncertainty. The nuclides that were used are listed in Table 5.

6.1.2 *An Obrigheim inference model*

The goal, then, is to reconstruct three parameters of this sample: Burnup (BU), Power Density (PD), Initial Enrichment (IE). Unfortunately

Table 5: Intra-element ratios for inferring reactor parameters of the Obrigheim SNF samples from the SFCOMPO databse [83].

$^{143}Nd/^{148}Nd$	$^{238}Pu/^{239}Pu$	$^{236}U/^{238}U$	$^{134}Cs/^{137}Cs$
$^{146}Nd/^{145}Nd$	$^{240}Pu/^{241}Pu$	$^{235}U/^{238}U$	$^{241}Am/^{243}Am$
$^{144}Nd/^{148}Nd$	$^{242}Pu/^{239}Pu$		$^{242}Cm/^{244}Cm$

for this analysis, the data for these samples has been back-propagated to the time of discharge, which makes the effective Cooling Time (CT) zero. Including CT as an inference parameter with ground truth zero is likely to be problematic, because it corresponds to the edge of the parameter space. GP-surrogate models tend to perform worse at the edges of their parameter space, and the posterior distributions become compressed towards the edge of the prior . Therefore, CT is omitted from the parameter space in this test case.

Table 6: Prior and ground truth of the inference parameters in the PWR validation case with the averaged samples of the Obrigheim reactor.

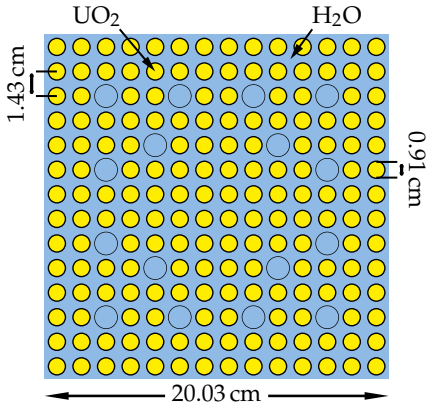
	Prior	Ground truth[†]
Burnup [MW d kg^{-1}]	20–40	28.3± 1.3
Power density [W g^{-1}]	10–50	21.4± 0.9
Initial enrichment [% ^{235}U]	2–5	3.13[‡]

[†] Mean and standard deviation of the sample values.
[‡] There was no variance between enrichment levels of the different samples.

The ground truth values and the chosen parameter space are given in Table 6. The final BU estimates for each sample are averaged to obtain the ground truth of BU. The ground truth of PD is the average of the time-averaged power density sequences that are provided in the database.

To model the likelihood of the sample, predictive computational models are required for each of the nuclides in the dataset. As outlined in Section 3.4, I use GP-based surrogate models, sampling 500 training data input values uniformly within the prior limits in Table 6 and CT implicitly set to zero for each datapoint. The training data outputs were computed with an infinite lattice Monte Carlo model of an Obrigheim assembly. Figure 15 displays the design and material parameters of the model. The schematic in Figure 15a shows one fuel assembly. In the actual implementation, the symmetry of the assembly can be exploited to reduce the model size. Only a quarter of the assembly needs to be modelled, with reflective boundary conditions at the symmetry axes.

The standard deviation $\sigma_{\mathcal{L},R}$ of each ratio R in the likelihood function $\mathcal{L}$, was modelled as a relative error, with $\sigma_{\mathcal{L},R} = \hat{y}_R \sigma_{R,rel}$ and



(a) Layout and geometry parameters of one fuel assembly (unit cell).

	ρ [g cm^{-3}]	T [K]
UO_2	10.05	800
Cladding	6.56	605.2
H_2O	0.7299	571

(b) Material parameters of the simulation model

Figure 15: Parameters of the simulation model of the Obrigheim reactor. The yellow circles represent the fuel elements in the assembly, which are surrounded by the H_2O moderator. The empty grid points are guide tubes, that can be used to insert control rods (to shut down the reactor) or measurement instrumentation. Unless noted otherwise, parameter values were taken from the SFCOMPO database.

$\sigma^2_{R,rel} = \sigma^2_{i,rel} + \sigma^2_{j,rel}$. The relative error $\sigma_{i,rel}$ of each nuclide was taken as $\sigma_{i,rel} = \sigma^2_{SFCOMPO,i} + \sigma^2_{model}$. Precise uncertainty estimates of modelling error are challenging, as it depends on the uncertainty of the nuclide data and simulation parameters such as the timestep sequence and model simplifications such as 2D- or infinite-lattice approximations. I chose $\sigma_{model} = 15\,\%$ as a moderately conservative heuristic for all nuclides, using the experiment-to-calculation deviations reported for the `Scale` software [80] to guide judgment. $\hat{y}_R$ and $\hat{\sigma}_{SFCOMPO,i}$ are the weighted average and weighted propagated uncertainty of the measurement data of the individual samples. Given the reported measurement uncertainties, $\sigma^2_{i,rel}$ is between 15 % to 17 % for each of the nuclides, except for ^{241}Am (35 %), ^{243}Am (36 %), and ^{242}Cm (36 %).

6.1.3 *Obrigheim validation results*

The posterior distributions obtained for the averaged Obrigheim sample are displayed in Figure 16. While the BU and IE posteriors are shaped like Gaussian distributions, the PD posterior is asymmetrical, with a higher probability density towards the lower bound of the prior range where the distribution is cut off. The peaks the BU and IE distributions are close to the respective ground truth value, and the 95 %-HDIs of all three parameters cover the ground truth.

Overall, this constitutes a very good agreement between the posterior probability densities and the ground truth values, which reflects

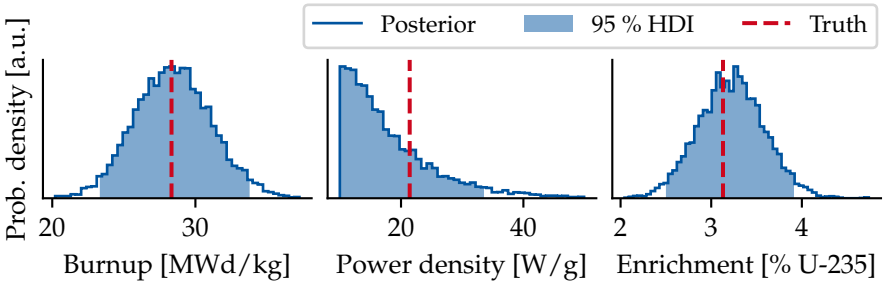


Figure 16: Inference results for the Obrigheim reactor parameters. The histograms represent binned samples of the posterior distribution normalized to be a probability density. The shaded areas indicate the 95 %-HDI of each posterior. The vertical, dashed lines represent the ground truth of each parameter.

positively on BRAM from a validation perspective. A reasonable concern regarding potential applications of the framework is whether modelling uncertainties due to necessary model approximations and simplifications will render it useless when confronted with real data from reprocessing waste samples.

This result supports the rationale that a simple reactor model can be used to reconstruct average parameters that describe the irradiation history of a whole batch of reprocessed spent nuclear fuel and that the spatial inhomogeneities of its composition would "average out" when the waste is mixed during or after reprocessing.

6.2 VALIDATION WITH MAGNOX DATA

6.2.1 *Magnox measurement data*

The SFCOMPO database contains four samples from Magnox reactors: three from the reactor *Hunterston A-1* (samples A–C) and one from the reactor *Bradwell-1* (sample D). Each sample corresponds to a slice of a fuel element that was irradiated in the reactor; the location of the fuel element within the reactor and any potential fuel movement or shuffling sequences or refuelling procedures are not specified.

The thermal power history, estimated final burnup, and the irradiation start and end dates of each sample are specified in the database. Although samples A–C stem from the same reactor and were irradiated for a similar duration, their final burnup and discharge dates vary significantly, casting doubt on whether they can be treated as representing an average sample. Therefore, each sample serves as evidence in a separate inference scenario with three inference parameters: BU, CT, and PD. The parameter IE is not treated as an inference parameter here because does not vary between the samples.

For each sample, there is a dataset with isotopic concentrations of various isotopes of uranium, plutonium, neodymium, curium, cae-

Table 7: Intra-element ratios for inferring reactor parameters of the Magnox SNF samples from the SFCOMPO databse [83].

$^{142}Nd/^{143}Nd$	$^{238}Pu/^{239}Pu$	$^{236}U/^{238}U$	$^{241}Am/^{243}Am$
$^{144}Nd/^{145}Nd$	$^{240}Pu/^{241}Pu$	$^{235}U/^{238}U$	$^{242}Cm/^{244}Cm$
$^{146}Nd/^{148}Nd$	$^{242}Pu/^{239}Pu$		$^{134}Cs/^{137}Cs$
$^{150}Nd/^{143}Nd$			

† Data for americium was not available for samples A and C.
‡ Data for ^{234}U was not available for samples C and D.

sium, and americium[41]. As evidence for the inference, I used all possible intra-element isotopic ratios, which are listed in Table 7.

[41] *The datasets also contain radioactive activity data for ^{95}Nb, ^{95}Zr, ^{144}Pr, and ^{144}Ce; however, no other isotopes of the same elements was available. Therefore, no intra-element ratios could be computed.*

6.2.2 *A Magnox inference model*

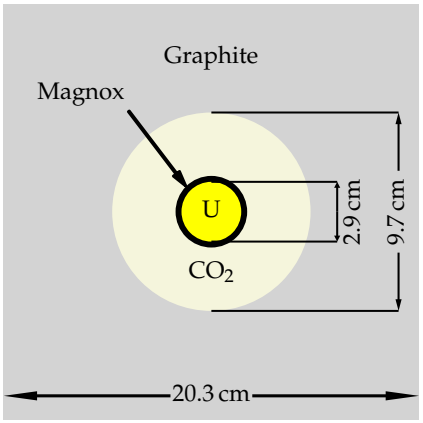
Table 8: Prior and ground truth of the inference parameters of each sample in the Magnox validation case.

		Samples			
	Prior	A	B	C	D
Burnup [MW d kg^{-1}]	0.09–9	3.27	6.58	6.9	8.93
Power density [W g^{-1}]	0.1–4.5	1.53	3.03	3.57	2.26
Cooling time [day]	50–1000	236	246	438	370

Given the operating history, the true values of each parameter can be calculated. The true cooling time is the time between the discharge date and the measurement date. The true power density is approximated by calculating a time-average of the given power density history. The true burnup is given directly in the database. Thus, I obtained measurement-based dataset with four ground truth to isotopic ratio vector pairings to validate the inference framework with a Magnox type reactor model. These ground truth values are listed in Table 8.

The table also shows the bounds of the parameter space of this scenario, which are used as limits for the uniform prior distributions during inference. The BU bounds reflect the range of values that can be achieved with natural uranium fuel in a Magnox reactor. The PD bounds cover the typical range for Magnox reactors [79], and the CT bounds were chosen as a moderately wide interval around the ground truth.

The standard deviation $\sigma_{\mathcal{L},R}$ of each ratio R in the likelihood function $\mathcal{L}$, was modelled as a relative error, with $\sigma_{\mathcal{L},R} = y_R \sigma_{R,rel}$ and $\sigma^2_{R,rel} = \sigma^2_{i,rel} + \sigma^2_{j,rel}$. The relative error $\sigma_{i,rel}$ of each nuclide was



(a) Layout and geometry parameters of the fuel channel (unit cell)

	Fuel vector [%-wt]
^{234}U	5.4079×10^{-3}
^{235}U	0.710 97
^{238}U	99.2836

	ρ [g cm^{-3}]	T [K]
U	18.5	700
Magnox	1.743	559
Graph.	1.6†	530
CO_2‡	8.57×10^{-3}	558

† "Early design" from [84]
‡ Calculated for 9 bar

(b) Material parameters of the simulation model

Figure 17: Parameters of the simulation model of the Bradwell-1 reactor. Unless specified otherwise, parameter values are taken as given in the SFCOMPO database.

taken as $\sigma_{i,rel} = \sigma^2_{SFCOMPO,i} + \sigma^2_{model}$, with $\sigma_{model} = 15\,\%$ like in Section 6.1, and $\sigma_{SFCOMPO,i}$ given by the SFCOMPO database.

The surrogate models for the likelihood are again created according the procedure in Section 3.4. A simulated dataset of 500 training and 100 test data points was generated with Monte Carlo reactor physics simulations for a quasi-randomly generated[42] sequence of input samples.

[42] *using the Sobol sequence.*

The simulation model consisted of an infinite lattice unit cell of the Bradwell reactor. The material and design parameters of the model are shown in Figure 17. Although there are some differences between the designs of the Bradwell and the Hunterston reactors, there was little difference between the relevant simulation output obtained with the two different models during testing. Therefore, I used the Bradwell model-generated dataset to train one set of surrogates, which I used for all four samples.

6.2.3 *Exploring model calibration*

The GP-surrogate models generally perform well at predicting simulated nuclide concentrations. Nevertheless, significant prediction errors can occur when comparing to the measurement data, given that the Monte Carlo simulations also come with prediction errors. These errors can stem from model approximations, e.g., the infinite lattice approximation, or from nuclear data uncertainties. They might also occur because relevant information pertaining to the sample irradi-

ation history is unknown. Since such conditions are likely to be encountered in the anticipated applications[43] of the framework, I use the Magnox data to explore the potential of calibrating the surrogate models with measurement data.

[43] *i. e., nuclear verification*

I test the concept with three nuclides: ^{235}U, ^{238}U, and ^{239}Pu. This choice is motivated by their significant contributions to the processes in a nuclear reactor: ^{235}U is the primary source of fission neutrons and the transmutation of ^{238}U is a major neutron absorption mechanism and the driver of ^{239}Pu production. Thus, their concentrations in the sample are strongly correlated to its burnup.

To calibrate the surrogate models for one of the four samples, the average relative deviation between surrogate model prediction and measured values is used, excluding the datapoint of the sample that will be inferred. The calibration factor $CF_{i,s}$ for nuclide i and sample s is:

$$CF_{i,s} = \frac{1}{N} \sum_{j \neq s}^{N} \frac{M_i \left(BU_j, CT_j, PD_j \right) - n_{i,j}}{n_{i,j}}. \tag{35}$$

Here, $n_{i,j}$ is the nuclide mass density measurement of nuclide i of sample j in the database. M_i is the surrogate model for the nuclide, which is evaluated at the true parameter values of the respective sample. Then, the calibrated surrogate model of nuclide i is:

$$M_{s,i,cal.}\left(BU, CT, PD\right) = M_i\left(BU, CT, PD\right)\left(1 + CF_{i,s}\right), \tag{36}$$

which is used for inference with the data from sample s.

Figure 18 compares the calibrated and the uncalibrated surrogate model predictions. It shows the deviation between the computed predictions C and the experimental value E, both for the calibrated and the uncalibrated surrogate model. Generally, the calibration reduces the prediction error. This effect is most noticeable for ^{239}Pu, where the error is offset by over 5 %. Sample D breaks this trend. Here, the error increases with the calibration. Given that the surrogate model underpredicts for samples A–C, the calibration factor for $M_{235,D,calib.}$ shifts the prediction upwards. Since the model already overpredicts for sample D, the calibration increases the error.

This systematic difference between samples D and A–C has no clear explanation. Reactor model differences are an unlikely, although not impossible, source of the discrepancy, given that the model used to generate the training data used the Bradwell design parameters, which is the reactor from where sample D originates. Due to the scarcity of information about the location of the samples within the reactor, other avenues of investigation could not be pursued, and the discrepancy could not be reconciled.

In the following section, inference results with the calibrated and the uncalibrated model are compared. Given the correlation between the calibrated nuclides and the parameter BU, I expect the effect of the calibration to be most pronounced in the BU posterior.

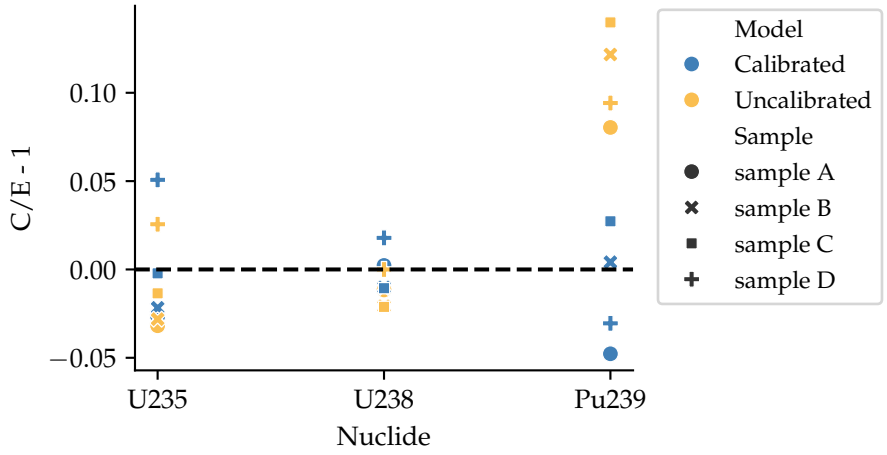


Figure 18: Calibrated and uncalibrated model predictions of ^{235}U, ^{238}U, and ^{239}Pu for the four Magnox samples. The y-axis shows the ratio of the predicted nuclide concentrations (C) to the measured nuclide concentrations (E). Calibrated predictions are shown in blue and uncalibrated predictions are shown in orange.

6.2.4 *Validation results: Magnox*

The inference results for all four samples in this validation dataset are shown in Figure 19. It shows the posterior distributions obtained with both the calibrated and the uncalibrated surrogate models[44].

Overall, the graph indicates good inference model performance. In almost all cases, the 95 %-HDI of the posterior includes the respective true parameter value. With a few exceptions, the posterior distributions resemble Gaussian distributions. One exception is the BU posterior of sample D. Here, the probability mass compressed towards the upper bound of the prior distribution because the true value is close to the bound.

The PD posteriors of samples B and C both have two regions of concentrated probability mass: one around the true value and one at the upper bound of the prior. Notably, the "correct" peak in the posterior of sample C is very slight.

The largest deviation is observed in the CT posterior of sample C. Here, the true value lies above the upper bound of the 95 %-HDI, barely at the edge of the region covered by the samples of the marginal posterior distribution. This deviation may be explained by the effect of the ratios ^{242}Cm/^{244}Cm and ^{134}Cs/^{137}Cs on the likelihood function. They are the two ratios most sensitive to CT in the given parameter range due to the respective half-lives of ^{242}Cm (162.8 d) and ^{134}Cs (754.2 d); the other nuclides in the sample are stable or have half-lives greater than 5000 days.

At the true parameter values of sample C, both ^{242}Cm and ^{134}Cs are predicted with a significant deviation from the measured values by the respective surrogate model. ^{242}Cm is underpredicted by ap-

[44] *Nuclides other than ^{235}U, ^{238}U, and ^{239}Pu are included in the calibrated inference, their surrogates are the same as in the uncalibrated inference.*

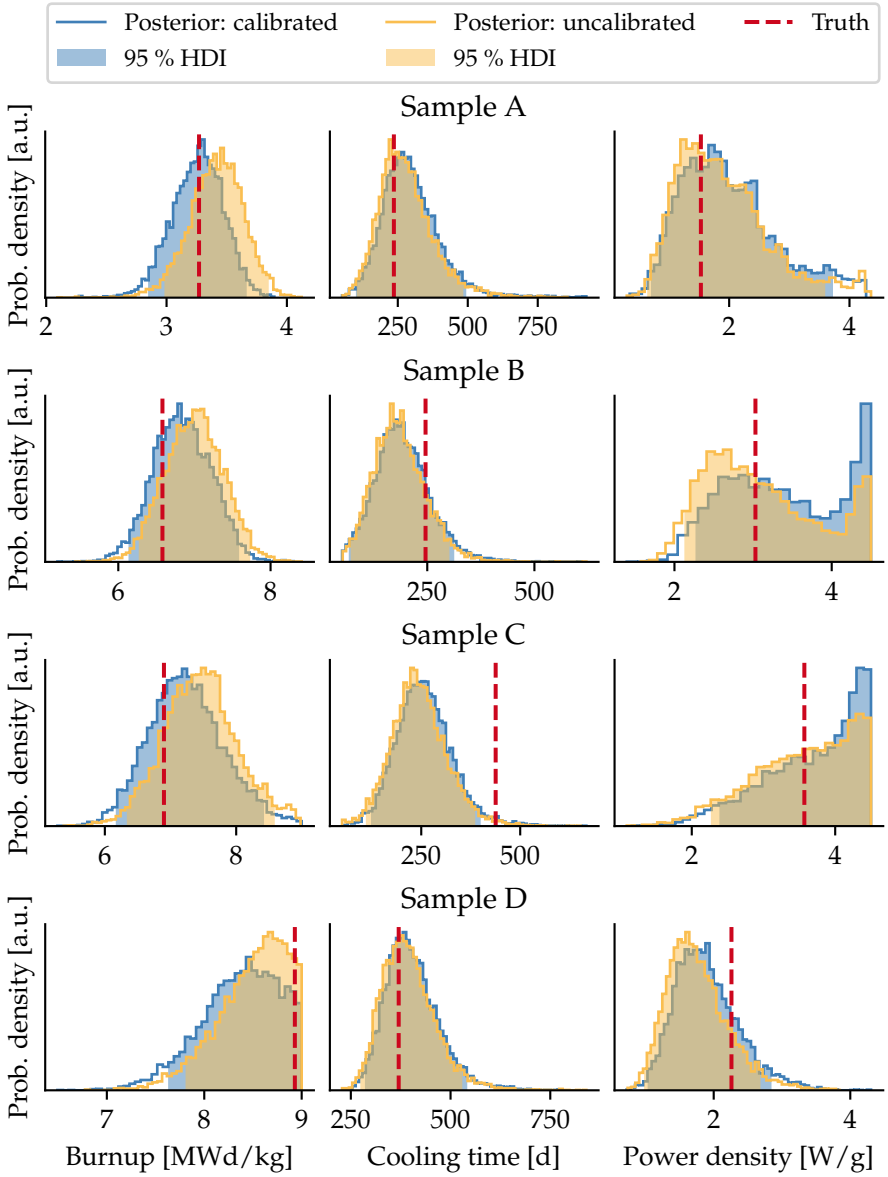


Figure 19: Inference results for reactor parameters inferred from experimental data from the SFCOMPO database. The histograms represent binned samples of the posterior distribution normalized to be a probability density. The shaded areas underneath each histogram indicate the 95 %-HDI. The vertical, dashed lines represent the respective ground truth parameter values.

proximately 40 % and ^{134}Cs is underpredicted by approximately 20 %. Both nuclides are given with less than 3 % measurement error in the database; thus, their likelihood uncertainty is dominated by the 15 % modelling error that was used for all four samples. Given that both nuclides are biased to lower nuclide density values, the maximum of the joint likelihood is shifted to lower CTs than it should be. At the true CT, the nuclide predictions are too low compared to the measurement; at lower CTs they are closer to the true value.

Figure 19 also shows that, for samples A–C, there is slight shift of the calibrated posterior density towards lower BU values, that is, towards the true value, compared to the uncalibrated posterior. For sample D, the opposite effect is observable, with the posterior density flattening and allocating higher probability to BU values that are further away from the true value. This pattern is similar to the pattern observed in Figure 18, where the model calibration for sample D actually increased the deviation between computed and experimental ^{235}U concentration.

6.3 DISCUSSION OF VALIDATION RESULTS

The results of both validation scenarios are a promising start towards validating BRAM. The PWR study looks almost like a "textbook example", with single-modal posterior distributions that peak close to the ground truth values.

The posterior distributions obtained with the Magnox data do not peak as close to the ground truth, but their respective 95 %-HDIs include the true values. Since the Magnox samples come from various, partially unknown, locations within the reactor core, they reflect the "local" isotopic composition, rather than a core average (like the PWR sample). Such samples are affected by inhomogeneities in the neutron flux distributions and as such it is to be expected that they are more challenging to model and reconstruct with a simple infinite lattice model.

With this test case, the potential to use experimental data to "calibrate" the likelihood model was successfully demonstrated. Although the calibration was simplistic, it led to improved posterior distributions for BU, which is arguably the most important parameter. In future validation efforts, the calibration approach should be explored further.

Taken together, the two scenarios indicate that it is possible to reconstruct average reactor operating parameters with simple models. If the likelihood uncertainty is modelled accurately, the inference framework produces promising results even with samples affected by local inhomogeneities. The results obtained with Magnox sample C underline this need to accurately calibrate the model and assess un-

certainties. In this case, the ^{242}Cm is not modelled well enough to reconstruct the cooling time correctly.

There are some obvious limitations to using the SFCOMPO data for validation. First, the database contains measurements of nuclide densities in SNF, not in HAR. While the intra-element isotopic ratios are mostly unaffected by the chemical reprocessing steps, it would be preferable in future work to validate with samples from real reprocessing waste. Second, the number of nuclides that is available in the SFCOMPO data is limited. The two cases discussed in this chapter use nuclides that are commonly used to characterize SNF [62]. The two radionuclides ^{242}Cm and ^{134}Cs are intuitively useful for reconstructing the cooling time due to their comparatively short half lives (162.8 d and 2.065 y, respectively [85]); the other nuclides are longer lived or stable and are useful burnup indicators. The results here show that they can also be used to infer the initial enrichment of the fuel and the average power density in the reactor. Enrichment is primarily inferred from the uranium isotopes themselves, though the changes in relative abundance between ^{238}U and ^{235}U also affect the abundance of respective fission and transmutation products. The informational content on power density is due to its changing of nuclear reaction rates relative to nuclear decay rates, that change the respective production and destruction rates of each radionuclide.

Of the nuclides used in this chapter, only the uranium and some of the neodymium isotopes are among the ratios listed in Table 2. Thus, it is not possible to validate the framework with those ratios that have been selected for improved reconstruction performance in Chapter 4.

7

POTENTIAL APPLICATIONS IN THE DPRK

The Democratic People's Republic of Korea (DPRK) is a former party to the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) that, since announcing its withdrawal from the treaty in 2003 [86], has conducted a series of nuclear test explosions [86] and is now a nuclear weapons-possessing state. There is a strong interest in the international community to convince North Korea to disarm its nuclear arsenal and return to the NPT regime. As North Korea has previously not been entirely forthright and cooperative with the IAEA inspectors and current diplomatic relations are burdened with mistrust, it seems likely that future steps towards disarmament will be heavily scrutinized by the international community and credibility would be bolstered by the transparent use of technical verification tools, such as those comprising the domain of nuclear archaeology. In fact, the use of nuclear archaeology techniques has been advocated [27, 87] and studied [88, 89] in the past.

Therefore, the DPRK is an interesting context under which to further explore the Bayesian Reprocessing Waste Analysis Method (BRAM). The comparatively[45] small scale of its nuclear fuel cycle permits simulated scenarios that expand on the previous developmental studies in Chapter 5 in a moderated but meaningful manner by increasing the complexity of the inference model. At the same time, these scenarios remain sufficiently close to reality to explore the potential utility and relevance of the framework in a realistic application.

[45] *compared to other nuclear-weapons(-possessing) states*

The research described in this chapter has been published in a peer-reviewed article [90], co-authored by Johannes Bosse, who conducted his master's thesis [81] research under my supervision. The description of the methods and results, as well as the figures have been adapted.

The following sections outline the history of the North Korean nuclear program, describe the relevant adjustments to the inference model, and discuss simulation-based inference test cases. Two applications are investigated: first, burnup and cooling times are reconstructed for a "five-batch" mixture of waste, then, two competing assessments of the North Korean 5 MWe reactor's history are modelled and compared.

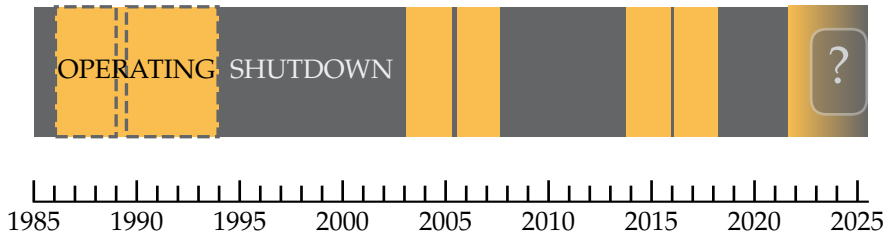


Figure 20: The estimated operating timeline of then DPRK's 5 MWe reactor. Each orange section indicates an operating window after which one full core of spent fuel was discharged. The dashed boxes show the potential alternate history where fuel was discharged from the reactor prior to 1994.

7.1 A BRIEF HISTORY OF PLUTONIUM PRODUCTION IN NORTH KOREA

The DPRK's main pathway for plutonium production has been the so-called 5 MWe[1] reactor, a graphite-moderated reactor of the Magnox design. The reactor is fueled with approximately 50 tonnes of natural uranium, can produce approximately 6 kg of plutonium per year, and has been operating intermittently since 1986 [91]. The presumed operating history of the 5 MWe reactor is shown in Figure 20. The commonly assumed timeline, indicated by the orange blocks, has the reactor starting its 6th operating cycle in July 2021 [86]. It is uncertain if this operating cycle has ended, although here have been some indications that the fuel was discharged in 2024 [92].

Several sources also report suspicions that fuel was discharged from the reactor during a shut-down in 1998, before the end of its first "official" operating cycle [89, 93]. Although it is unclear whether a full refuelling could have been carried out during this shut-down, this work assumes a possible alternate operating history with an early extra operating cycle, adopting the assessment of de Troullioud de Lanversin and Kütt [89]. The alternate history is indicated by the dashed, gray boxes in Figure 20.

7.1.1 *Modelling the 5 MWe reactor*

The 5 MWe reactor is a graphite-moderated Magnox-type reactor with a nominal power of 25 MWt. It uses natural uranium metal as fuel. Many of its design parameters are known from observations by IAEA inspectors [93], or Siegfried Hecker and his team[2] [94]; or can be

1 It is allegedly capable of producing 5 MWe of electric power, which North Korea insists is its purpose.

2 Siegfried Hecker is an American nuclear scientist, former director of the Los Alamos National Laboratory, and research professor at Stanford University. Between 2004

gauged from the design of British Magnox reactors [79], on which the 5 MWe reactor is based. A number of studies have modelled the 5 MWe reactor to estimate its plutonium production [89, 95], explore potential applications of GIRM [88, 89], or estimate its potential to produce tritium [96]. These studies are consistent in most of the design parameters, although there are some differences, e. g., in the assumed densities of the fuel or the amount of Al in the fuel. Table 9 shows the design parameters of the 5 MWe reactor used in this work.

Like the Magnox reactors in Chapters 5 and 6, this reactor was modelled as an infinite-lattice model, applying periodic boundary conditions to a unit-cell model of one fuel channel. The simulation code `ONIX` [38] was used with nuclear data from the ENDF/b-v.iii library to generate training data for surrogate modelling (see Section 3.4). The training data has two input dimensions, BU and CT, that map to nuclide densities in the spent nuclear fuel (SNF).

Table 9: Design parameters of the 5 MWe reactor. The values are based on Kang [88], de Troullioud de Lanversin and Kütt [89], Albright and O'Neill [93], and Park and Hong [95].

Fuel channel		Full core	
Fuel comp.	nat. U 0.5 %at Al	Moderator mat.	Graphite
Fuel diameter	2.9 cm	Moderator density	1.74
Fuel length	52 cm	Num. fuel channels	$\approx$ 800
Cladding comp.	Mg 1 %at Al	Tot. fuel mass	$\approx$ 50 t
Cladding thick.	0.5 cm	Thermal power	20–25 MWt
Channel diam.	6.5 cm	Fuel density	18.45 g cm^{-3}
Lattice pitch	20 cm	Fuel temperature	873 K

7.2 RECONSTRUCTING THE 5 MWE REACTOR'S OPERATING HISTORY

To explore the potential of using BRAM to verify the 5 MWe's operating history, the first step is to investigate the *1994 discharge* scenario, attempting to reconstruct BU and CT of the five commonly presumed operating cycles. Later, I assess the potential to discriminate between the two possible operating histories. In this work, the last operating cycle is ignored because public sources are vague as to when it may have ended and whether the core was actually discharged.

and 2010, he and several colleagues were invited to several visits of the Yongbyon Nuclear Complex by the North Korean government. A significant portion of the publicly available information about the North Korean nuclear program originates from these visits.

Table 10: Two potential operating histories of the 5 MWe reactor in Yongbyon.

Operating window	Burnup [MWd/kg]	
	Pre-1994 discharge	1994 discharge
1986–1989	0.185	-
1989–1994	0.5–0.7	-
1986–1994	-	0.6–0.7
Jan. 2003–Apr. 2005	0.23–0.33	0.23–0.33
Jun. 2005–Jul. 2007	0.22–0.31	0.22–0.31
Aug. 2013–Oct. 2015	0.22–0.31	0.22–0.31
Jan. 2016–Mar. 2018	0.23–0.32	0.23–0.32

The DPRK operates a single reprocessing plant, which prompts the assumption that all of its reprocessing waste is mixed together. To model the likelihood for the isotopic ratio in such a mixture, a simplified version of the mixture-model (Equation (34)) from Section 5.2.1 is used:

$$M_r(\theta) = \frac{\sum_b^B N_i(\theta_b)}{\sum_b^B N_j(\theta_b)}. \tag{37}$$

[46] *Each batch corresponds to one operating cycle.*

The number of batches[46] B is fixed to five or six and each batch is modelled with a separate pair of inference parameters $\theta_b = (BU_b, CT_b)$. The mixing factors are neglected as inference parameters. This simplification primarily serves to reduce the model complexity and can be justified with the assumption that each cycle represents one full core discharge that is reprocessed with the same PUREX process parameters. Therefore, the volume of waste added to the mixture by each cycle is approximately equal, eliminating the need to model the mixing factor. The likelihood uncertainty was set to 5 % for each nuclide.

The isotopic ratios for this scenario were selected with the combinatorial ratio selection algorithm (Algorithm 2)[3]. The model in Equation (37) has more parameters than the models tested in Chapter 5, incurring a larger computational cost. During preliminary testing with longer ratio lists, the MCMC sampling was slow, requiring many model evaluations to generate a new sample from the posterior, and ultimately failing to finish sampling in a reasonable time. As a precautionary measure to preclude convergence issues, the length of the selected ratio list was limited to ten. They are shown in Table 11.

Prior distributions on the average burnup and the cooling time of each batch are defined as uniform distributions, based on the assumed operating timeline in Figure 20. Using the burnup limits esti-

3 The combinatorial algorithm was used because this study predates the development of the new, genetic algorithm.

Table 11: List of isotopic ratios selected for the DPRK scenarios.

Ba-134/Ba-136	Ba-134/Ba-137	Cd-110/Cd-113
Eu-151/Eu-155	Eu-153/Eu-155	Gd-154/Gd-155
Gd-154/Gd-157	Sm-148/Sm-149	

mated by de Troullioud de Lanversin and Kütt [89], two scenarios for the operating history are defined, analogously to their analysis. The first scenario, called *1994 discharge* scenario, assumes that the first full core discharge occurred in 1994, and the second, *pre-1994 discharge* scenario assumes an additional full core discharge in 1989. Thus, the hypothetical waste contains five batches in the former scenario and six batches in the latter. Table 10 shows limits of the uniform prior distributions on the BU parameter. The cooling time (CT) is defined as the time between the respective shut-down date of each cycle and a hypothetical measurement date on Jan. 1st 2020. Lower and upper limits are given by a one-year interval centered on the cycle end dates in Table 10.

In the following section, two additional variations of these prior distributions are investigated to explore the effect of prior assumptions on the inference results.

7.3 THE *1994 discharge* SCENARIO WITH VARYING PRIORS

This section compares results obtained with the "five-batch" inference model for the *1994 discharge* scenario and three different sets of prior distributions.[47] The limits of these variations are specified in Table 12. The first test case uses the prior limits described in the previous section. Hereafter, these priors are called *realistic* priors, because they are based on the current, publicly available knowledge on the reactor's operating history. They reflect an assessment by hypothetical inspectors relying on open-source intelligence analysis to inform their reconstruction.

[47] *The results in this section were produced together with my student, Johannes Bosse.*

The second case extends the prior intervals to assess the sensitivity of the selected isotopic ratios with respect to the inference parameters. These priors are called *wide* priors. The same, larger BU interval is used for all five batches and each respective CT interval is set to cover multiple years. These intervals are still physically reasonable in the sense that the burnup is well below the limit where the reactor loses criticality.

The third case is called *wrong priors* and is intended to represent a scenario where inspectors have incorrect information about the reactor. The prior limits on BU are shifted, such that the lower limit is greater than the true value, corresponding to potential errors in

Table 12: Limits of the uniform prior distributions for three variations of the *1994 discharge* scenario.

Realistic Priors		Wide Priors		Wrong Priors	
BU [MWd/kg]	CT [d]	BU [MWd/kg]	CT [d]	BU [MWd/kg]	CT [d]
0.23-0.32	417-783	0-0.8	0-1050	0.32-0.42	417-783
0.22-0.31	1317-1683	0-0.8	1100-3000	0.31-0.42	1317-1683
0.22-0.31	4287-4653	0-0.8	3050-4800	0.31-0.41	4287-4653
0.23-0.33	5098-5462	0-0.8	4850-7000	0.33-0.43	5098-5462
0.6-0.7	9088-9452	0-0.8	7050-10000	0.7-0.8	9088-94

estimates of the capacity factor or power. The cooling time intervals remain unchanged compared to the realistic case.

The three test cases use the same ground truth and simulated measurement data, corresponding to surrogate model predictions with the center of each *realistic* prior interval as input. The posterior distributions of these three test cases are shown in Figure 21. In the *realistic* test case, the prior distributions appear roughly uniformly distributed, except for the BU posterior of the first cycle (batch E), which has a significantly higher true BU value than the other batches.

In the *wide* test case, there is some variation in the CT posteriors, and, more relevantly, the BU posteriors show distinct peaks compared to their priors. Comparing this case to the realistic case, it appears that the realistic priors are so narrow that the isotopic ratios are not sensitive to variations in BU on such a small scale.

The *wrong* test case results in BU posteriors that are concentrated at the lower bound of the prior. In a sense, they "point" towards the true parameter value. Here, the ratio sensitivity that can be observed in the *wide* test case is instructive, implying that the flat posteriors in the realistic case can be interpreted as confirmation that the ground truth lies near the center or at least within the prior interval, because otherwise, posteriors resembling the *wrong* case would be observed.

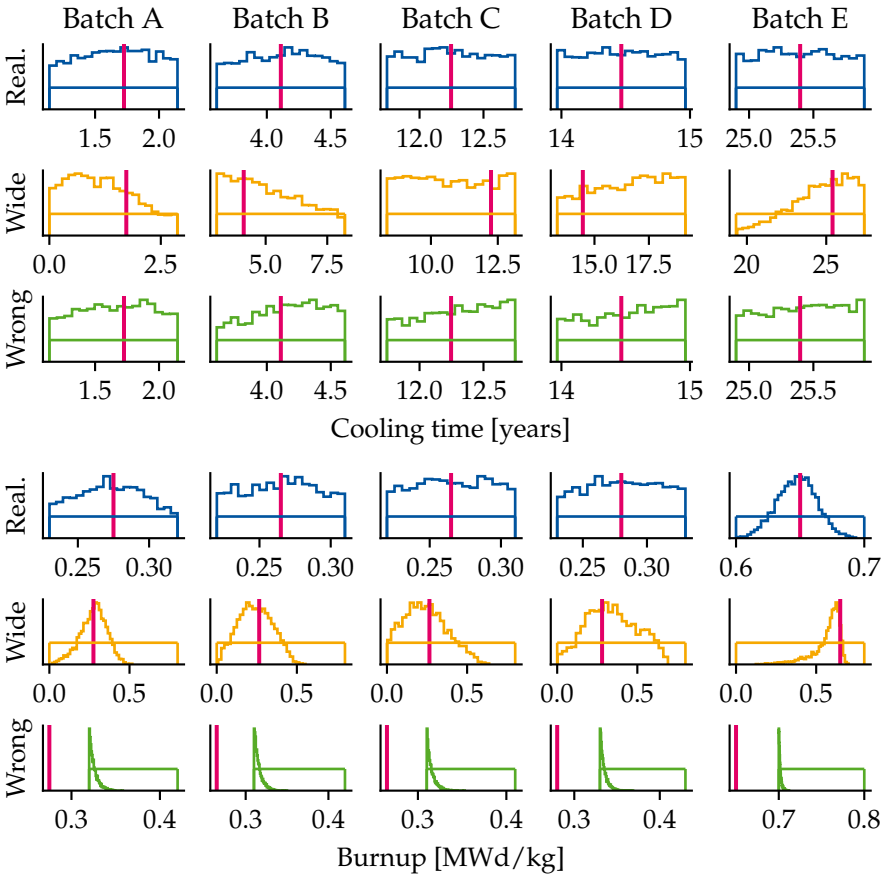


Figure 21: Prior and posterior distributions of the inference parameters of the three scenarios. The upper graphs show posteriors for cooling time, and the lower graphs show posteriors for burnup. The vertical line indicates the "true" parameter value as used to simulate the isotopic ratio observations and the flat horizontal lines indicate the prior distribution.

7.4 DISCRIMINATING *pre-1994* AND *1994* DISCHARGE SCENARIOS

Having seen that the *1994 discharge* scenario can be verified with the Bayesian framework, the next question is whether it is possible to distinguish between the *1994 discharge* scenario and the suspected alternative *pre-1994 discharge scenario*. To this end, the LOO-CV-based model comparison framework is tested, which was used in Section 5.3 to discriminate between reactor types. A "five-batch" and a "six-batch" variant of the likelihood function (see Equation (37)) is run with priors that correspond to the realistic limits of the respective scenario in Table 10. Ranking the models with LOO-CV indicates which model is preferred. The assigned weights can be interpreted as a degree of confidence in this preference.

Two test cases are studied, one corresponding to a true *pre-1994 discharge* and the other to the *1994 discharge* scenario. The LOO-CV algorithm compares the two models, respectively, for each test case. One weight is calculated for each ratio in the likelihood; mean and standard deviation of the weights are shown in Table 13. In each true scenario, the respective "correct" inference model is clearly preferred according to the weights, that is, the 1994-model is preferred in the "true" *1994 discharge* scenario and the opposite is true for the true *pre-1994 discharge* scenario.

Each test case contains one isotopic ratio with a weight that indicates the incorrect model. In the *1994 discharge* scenario, $^{134}\mathrm{Ba}/^{136}\mathrm{Ba}$ prefers the incorrect model; in the *pre-1994 discharge* scenario, it is $^{148}\mathrm{Sm}/^{149}\mathrm{Sm}$.

7.4.1 *A posterior plutonium estimate*

In the imagined application scenarios, the goal is to verify a declaration of plutonium stocks by reconstructing the production history. To estimate the amount of plutonium that was produced, samples are drawn from the posterior distributions of the BU parameters and insert them into the forward model. By weighting the samples from each model with the corresponding LOO-CV weights, a *model averaged*

Table 13: Averaged weights for the two model comparison scenarios. Weights are averaged over each ratio in the likelihood. The columns show the ground truth scenario and the indicate are the inference model.

	1994 scenario	pre-1994 scenario
1994 model	0.88 ± 0.35	0.12 ± 0.35
pre-1994 model	0.13 ± 0.35	0.88 ± 0.35

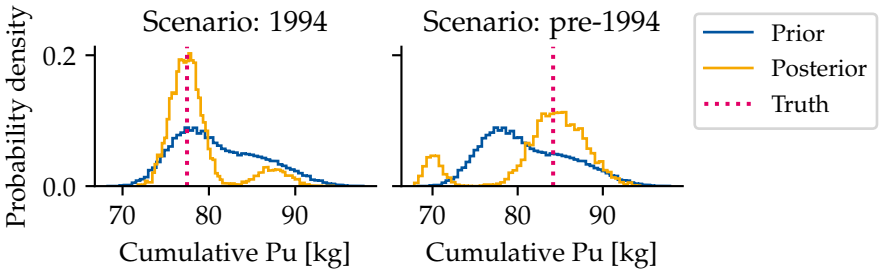


Figure 22: Weight-averaged posterior predictive distribution of the plutonium production estimate of two possible "true" scenarios. The dashed vertical line represents the true (simulated) amount of plutonium, the blue histogram indicates the prior predictive distribution, and the orange histogram shows the posterior distribution.

posterior plutonium estimate is obtained. This distribution is shown in Figure 22. For comparison, prior plutonium estimates are drawn, using equal weights for both models.

The approach described above is inspired by *posterior predictive sampling*, although it would not be quite accurate to call it such. In posterior predictive sampling, samples from the posterior are used to predict new observations of the observables, which are the isotopic ratios in Table 11. Here, the models compute the plutonium density.

7.5 DISCUSSING APPLICATION POTENTIAL

This chapter presented promising results that underscore the potential utility of BRAM to support future fissile material verification efforts in the DPRK. While realistic prior distributions are already quite precise, the posterior distributions can indicate if their limits were incorrect and point towards inconsistencies in a hypothetical declaration. Conversely, flat posteriors can be taken as confirmation that a declaration is accurate, as the method is sensitive to the inference parameters.

Using the LOO-CV-based model comparison method, there is potential to discriminate between competing assessments of the reactor operating history during the early stages of the DPRK's nuclear program. A posterior plutonium estimate, derived with weighted models, can give a more accurate estimate of the produced quantities.

The simulated test cases in this chapter used an optimistic uncertainty model, compared to the experimental validation cases in Chapter 6. Using larger uncertainty estimates would reduce sensitivity to the inference parameters. However, the results in Chapter 6 indicate that experimental data can be used to calibrate the inference model and improve results. Thus, it may be possible to use data from com-

mercial Magnox reactors to calibrate inference models in preparation for their application to the 5 MWe reactor.

The analysis here used only ten isotopic ratios to condition the posterior distributions. Using more ratios could lead to greater sensitivity, if the greater computational cost is acceptable or the efficiency of the sampling algorithm can be improved.

8 CONCLUSIONS

8.1 SUMMARY

The goal of this work has been to advance the development of the novel Bayesian Reprocessing Waste Analysis Method (BRAM) as a tool with the potential to support future disarmament endeavors by enabling transparent verification of fissile material declarations. To this end, some of the open questions have been investigated, primarily on the basis of computerized simulations, but also with publicly available experimental data.

I investigated the suitability of different isotopic ratios to perform well as evidence for reconstructing parameters pertinent to the operating history of nuclear reactors. For this purpose, a new *genetic* algorithm was developed to optimize the choice of ratios, given constraints on the number of rations that can be measured. Using some basic approximations regarding the expected detectability of various nuclides in reprocessing waste, a new set of 30 isotopic ratios was selected from a much larger set of options. This set of ratios was optimized to produce accurate parameter inference for an average fuel burnup below $9\,\mathrm{MW\,d\,kg^{-1}}$ in PWR, PHWR, and Magnox type reactors; while also reconstructing the cooling time, initial enrichment, and average power density.

The suitability of this set of ratios was demonstrated in two systematic, simulation-based, test series; each exploring a different, not-yet-studied use case. The first demonstrated that it is possible to reconstruct not only *burnup* and *cooling time*, but also *initial enrichment* and *power density* of samples containing a mixture of waste from two different, but a priori known, reactor types. The second demonstrated that, assuming waste has not been mixed, the reactor type can also be reconstructed with a high accuracy using Bayesian model comparison methods, while operating history parameters are simultaneously inferred. The test series covered a large range of possible parameter values and demonstrated an overall high success rate, although low ($\lesssim 3\,\mathrm{MW\,d\,kg^{-1}}$) can lead to convergence issues while sampling the posteriors. In both test scenarios, the new selection of isotopic ratios performed better than those used in previous studies.

To initiate efforts to experimentally validate this novel nuclear archaeology technique, two test cases were created with experimental data from the SFCOMPO database. The first test case used measurement data from several fuel assemblies of the PWR "Kernkraftwerk Obrigheim". By averaging these samples, a proxy for measurement

of reprocessing waste from one reactor core with one set of average operating history parameters was created. The average burnup, the initial enrichment, and the average power density could be reconstructed with high accuracy, with posterior distributions matching the known true values given by the database.

The second test case used four experimental samples from British Magnox reactors. Their operating parameters were reconstructed individually, since they did not share a common irradiation history. Burnup and power density were reconstructed successfully for all four samples, although the difference between the true parameter values and the density peaks of the posterior probability distributions were larger than for the PWR test case. The cooling time was reconstructed successfully for three samples, for the fourth sample the cooling time was underestimated, most likely due to modelling errors for the nuclide ^{242}Cm.

Both test cases support the conclusion that BRAM can perform its intended task, although one has to allow for and handle modelling uncertainty; here 15 % modelling uncertainty led to mostly successful parameter reconstruction. With the Magnox test case, the possibility to calibrate the inference model with other measurements for the same reactor type was successfully demonstrated; simple calibration factors for the nuclides ^{235}U, ^{238}U, and ^{239}Pu led to slight improvements in the reconstructed burnup posteriors.

Finally, a scenario for potential applications to verify nuclear disarmament in the Democratic People's Republic of Korea (DPRK) was studied. Although the potential to gain more precise knowledge of the plutonium production is limited due to already very precise prior information, the isotopic ratios used in the simulation are sensitive on the expected scale and, thus, the method could be used to confirm the existing estimates, or indicate if current assumptions are incorrect. Combining the analysis of mixed waste samples and the model comparison method shows potential for discriminating between competing hypotheses for the early operating history of the reactor.

8.2 IMPLICATIONS

Generally, the simulation-based test scenarios highlight the potential utility, while indicating some limitations and opening further questions. Dealing with samples containing mixed reprocessing waste stemming from several unknown reactors seems like the ultimate challenge. The results indicate that treating the sample as a mixture comprised of "batches" of waste, each with individual parameters, can work in simple scenarios, when reactor types and the number of batches are known. Here, the mixing ratio between the components was not prioritized, and previous research showed that large mixing ratios can negatively impact the parameter reconstruction. Regard-

ing potential applications, it is unclear whether reprocessing waste is frequently disposed of in the same container; perhaps the number of batches to be expected in one sample is limited. However, certain knowledge of the contributing reactors is unlikely.

The potential to reconstruct the reactor type is promising in that regard. It also points to other possible applications, for example in the field of nuclear forensics, where radioactive material needs to be attributed to a source. While this is often successfully attempted on the basis of structural properties of the material, the Bayesian Reprocessing Waste Analysis Method (BRAM) could offer additional confirmation or provide insights in challenging cases.

Modelling errors, stemming from nuclear data uncertainties or necessary model simplifications, play a crucial role in determining the success of the reconstruction. The efforts to validate the methodology with experimental data show that these uncertainties can be accounted for and even reduced by calibrating the inference model with data from similar reactors. Nonetheless, these uncertainties determine the resulting uncertainty on the reconstructed parameters. Given that validation was undertaken with nuclides that are frequently studied in the context of reactor operations, the question remains whether the success with the new set of ratios, which draws from a wider set of candidates with likely less accurate nuclear data, can be recreated experimentally.

8.3 OUTLOOK

Looking forward, the research in this thesis opens new questions and indicates promising new avenues or investigation. First, the results of the experimental validation, alongside the demonstrated impact of modelling uncertainties, motivates further experimental research and validation. In this context, more sophisticated model calibration techniques could be explored and developed with a stronger foundation of data. This line of research could also assess the potential to "realize" the improvements promised by the new set of isotopic ratios, by investigating potential measurement and modelling uncertainties of the "less-studied" nuclides.

The Bayesian model comparison method can be explored in different scenarios. It may be useful to distinguish whether a reactor has been used to produce tritium or plutonium. Such a distinction between production modes may be possible by simply reconstructing the initial enrichment, although it seems worth investigating whether the presence of lithium targets in the core has other effects on the isotopic signature of the spent nuclear fuel and reprocessing waste.

Moving towards scenarios where the number of batches in the mixture, and their reactor of origin, are unknown, the LOO-CV-based model comparison may again be a useful tool.

Future research endeavors might coordinate themselves as an effort to prepare the Bayesian Reprocessing Waste Analysis Method for potential verification activities in the DPRK, anticipating the state's cooperative intent to return to compliance with its obligations under its safeguards agreement and the NPT. Such efforts could seek to refine the inference model for mixtures of reprocessing waste from the 5 MWe, taking potentially new reactor cycles and hypothesized alternative production modes [96] into account. The targeted set of isotopic ratios could be further refined, including uncertainty assessments and calibration for individual ratios. Finally, the possibility to reconstruct samples including waste from the newly operational Experimental Light-Water Reactor (ELWR) [97] should be investigated.

Overall, the potential utility of the Bayesian Reprocessing Waste Analysis Method (BRAM) has been demonstrated in this thesis, while motivating further research to develop and validate its practical relevance.

Part IV

APPENDIX

A

COMPUTATIONALLY EFFICIENT APPROXIMATION OF MARGINAL POSTERIOR DISTRIBUTIONS

The approximate Bayesian computation (ABC) algorithm implemented by Figueroa [29], while foundational in the context of this work, is computationally efficient *only* under the caveat that it scales very poorly with the number of dimensions D of the inference parameter space. This drawback is owed to the fact that the likelihood function for one set of ratios is evaluated on a regular grid of size N^D of the input space. This exponential scaling causes two computational bottlenecks: the size of the input value array starts to exceed the available memory even on high-performance compute clusters and the number of likelihood function calls increases the compute time. The former could be addressed by storing input values and loading them into memory sequentially; however, this approach clashes with the latter bottleneck, as it further increases the necessary time. The compute-time can be reduced by running the independent steps of the computation in parallel, which, in turn, increases the required memory. Thus, increasing D forces a reduction of grid steps. For example, to run ABC with the four input dimensions in this thesis ($D = 4$), N could not be above approximately 200. In contrast, Figueroa [29] used $N = 1000$ for his study with two input dimensions.

For this work, I developed a modified version of the algorithm that scales better with the number of dimensions. [48] Instead of evaluating the likelihood function on a regular grid, I use the Sobol sequence [57, 98] to generate a quasi-random sequence of M points in the input space. This sequence generates an input matrix of dimension $M \times D$, which consequently only requires M evaluations of the likelihood function. The procedure to compute the marginal approximations is outlined in Algorithm 5. After generating the input samples with the Sobol sequence and computing the likelihood value of each sample, a binning structure is created for each dimension. For each input dimension D, each sample is assigned to a bin based on its value in the respective input dimension (see Algorithm 5). To obtain the marginal distribution in one dimension, the likelihood values corresponding to test points in the same bin are summed. Finally, by numerical integration, the marginal distributions are normalized.

[48] *It should be mentioned that Figueroa has also released Sobol-sequence-based implementation of this algorithm GitHub. Though we had discussed the possibility of this adaptation together, I did not become aware of his version until after I had developed my own.*

Generally, this approach computes worse approximations than the grid-based approach. However, the grid-based approach is not feasible in higher dimensions, which makes this approach useful.

I also implemented features to improve the computational performance of the ratio selection algorithm. A *lookup table*-based approach

Algorithm 5 Approximating marginal distributions

```
1: procedure MARGINALS(LIM, D, N, M)        ▷ LIM is a 2 × D array
2:     samples ← Sobol(LIM, M)
3:     likelihood ← pdf(samples)
4:     bins ← linspace(LIM, N + 1)
5:     idx ← []
6:     for i in range(D) do                                  ▷ ...
7:         idx ← digitize(samples[:, i], bins[i])
8:     end for
9:
10:    marginals ← empty(D × N)
11:    for i in range(D) do                                  ▷ ...
12:        for j in range(N) do
13:            marginals[i,j] ← sum(likelihood[idx[i][j]])
14:        end for
15:    end for
16:    return normalize(marginals)                           ▷ ...
17: end procedure
```

was added to reduce the amount of redundant surrogate model evaluations. This approach removed the need for interpolation-based surrogate models and enabled the direct use of the GP-based surrogate models that are also used by the inference framework.

B

PERFORMANCE TEST STATISTICS

This appendix shows the performance test statistics of the analysis in Chapter 5 plotted against the ground truth values of the inference parameters. For the "two-batch"-mixture scenario, *success* and *information-gain* of a parameter are displayed only as a function of its own ground truth value, i. e., success rate of BU is plotted against the true BU value, and so forth.

The success statistics are the normalized success rate, i. e., number of successfully reconstructed samples divided by the number of converged samples.

In all figures, the statistics are binned and averaged to make the trends visible and shaded areas indicate the 75 % percentile intervals of the statistics in each bin.

B.1 TEST STATISTICS OF THE "TWO-BATCH" MIXTURE SCENARIO

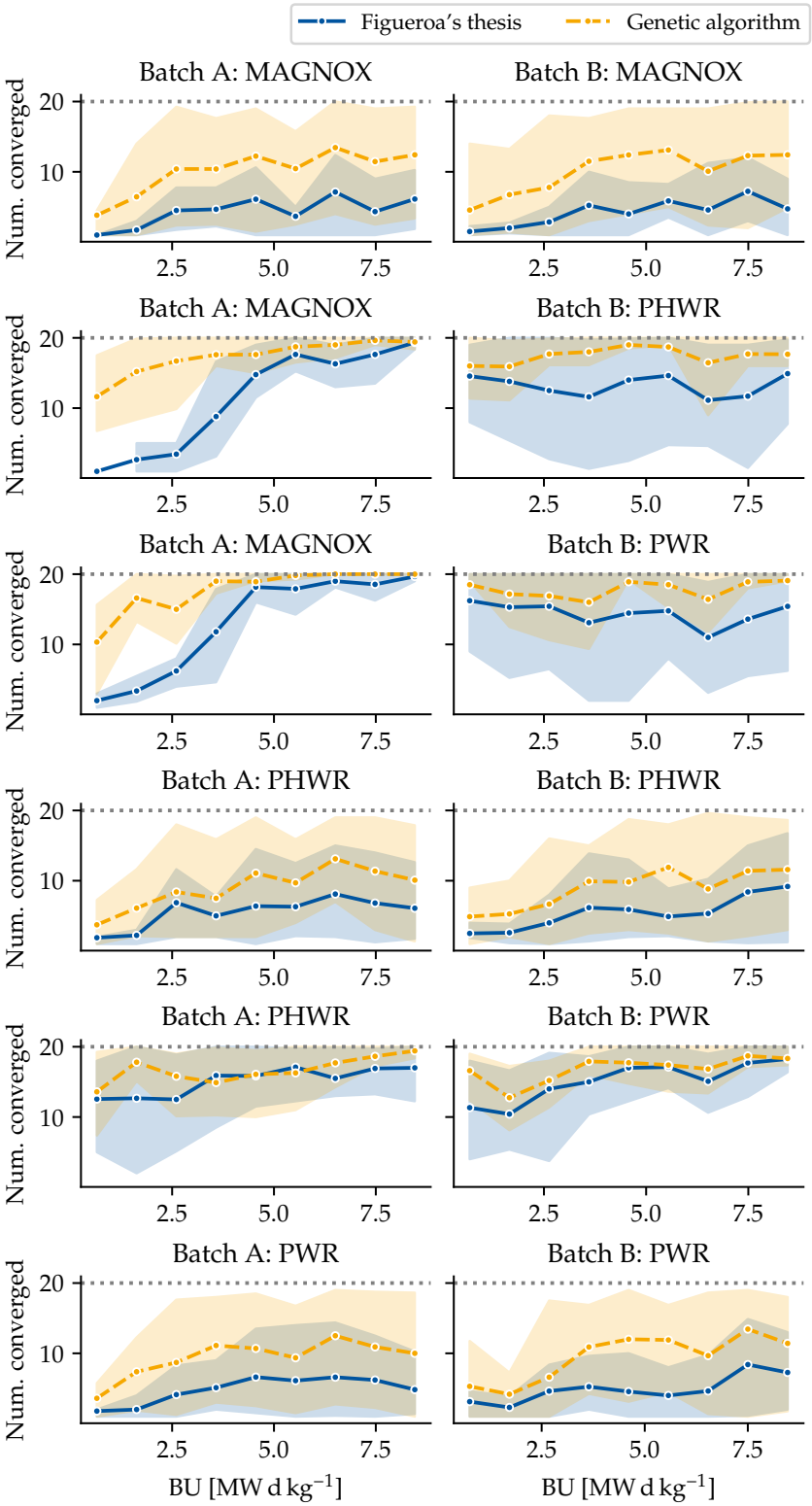


Figure 23: Mixture scenario convergence statistics vs BU.

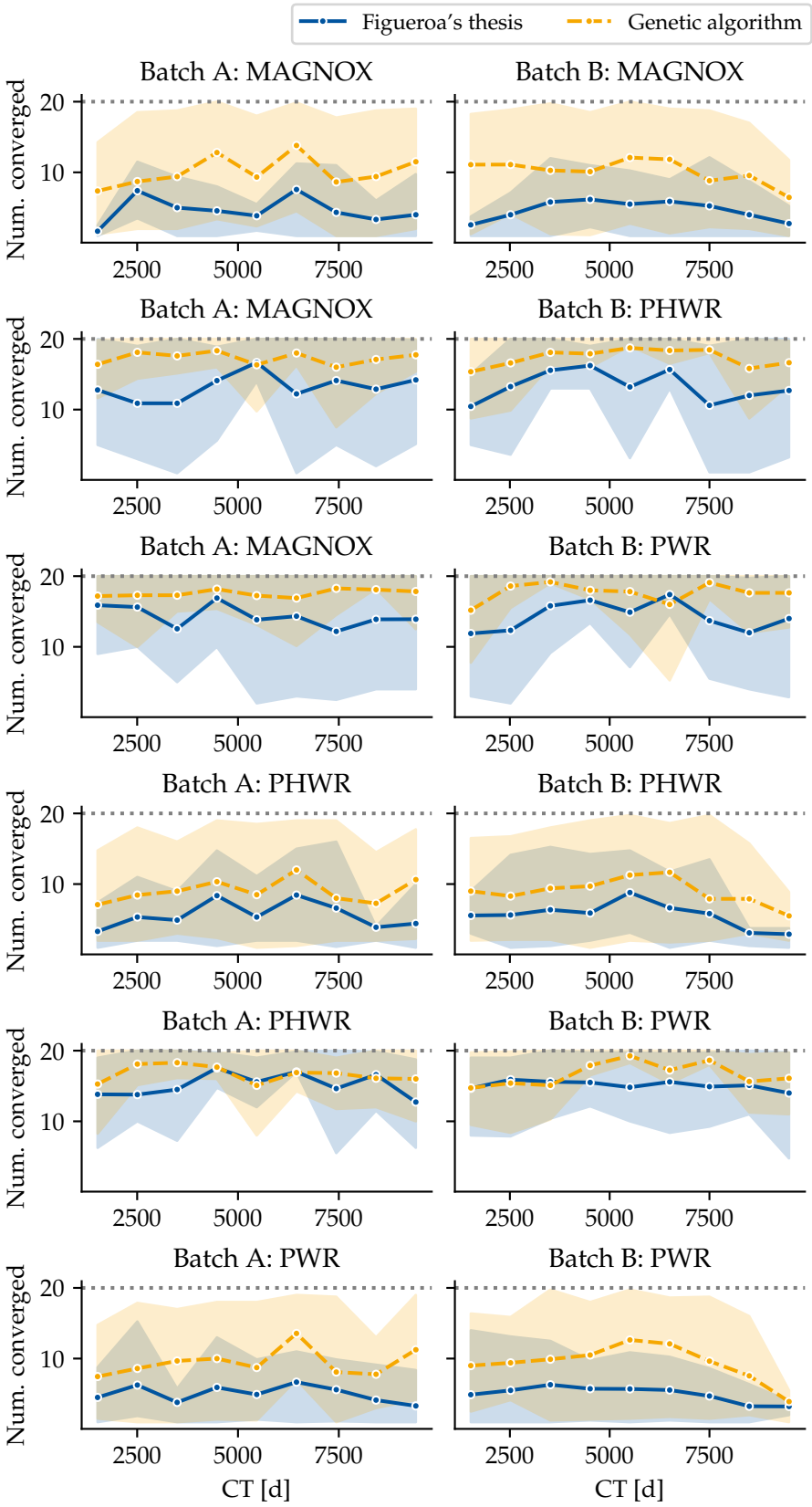


Figure 24: Mixture scenario convergence statistics vs CT.

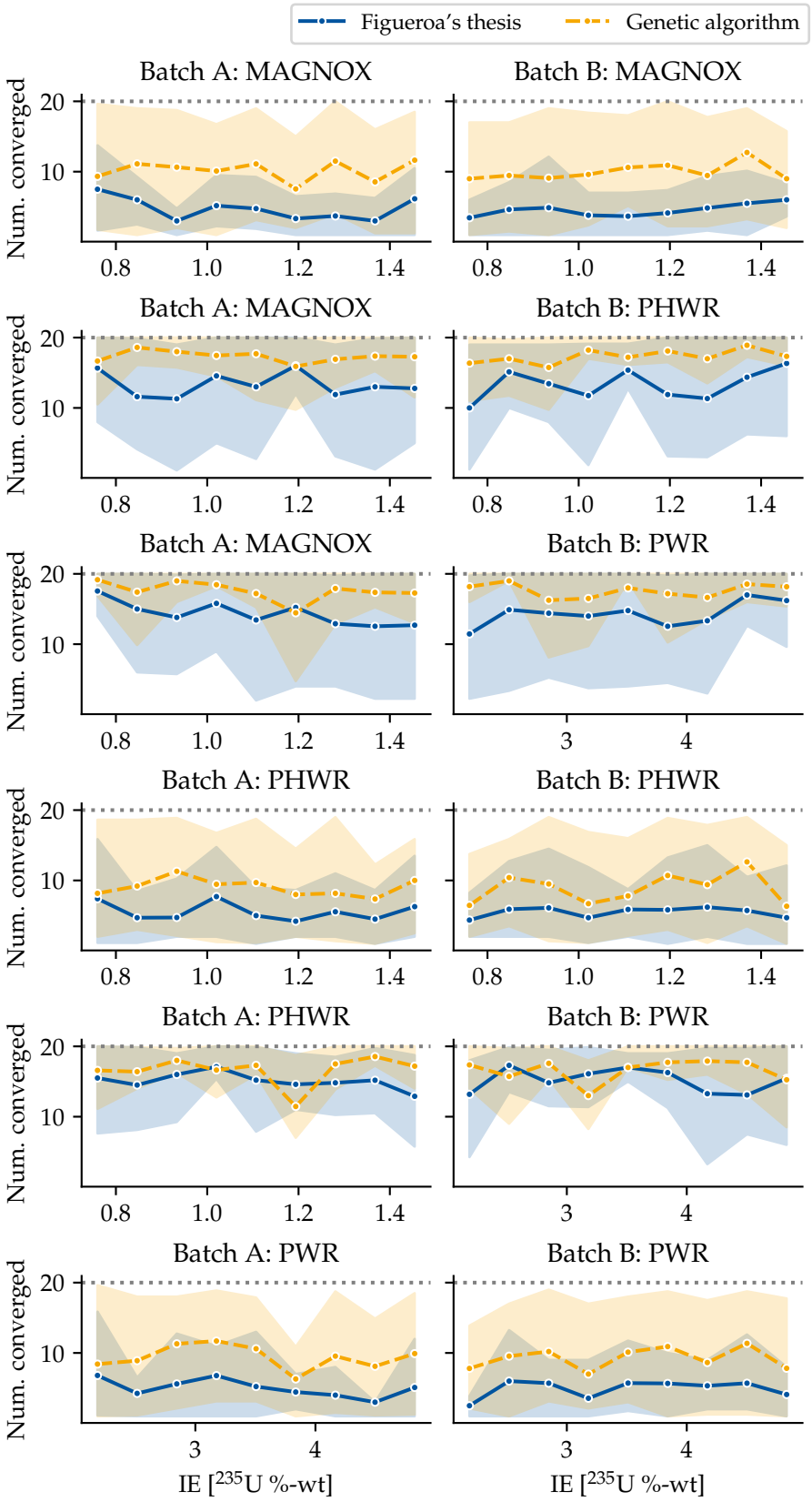


Figure 25: Mixture scenario convergence statistics vs IE.

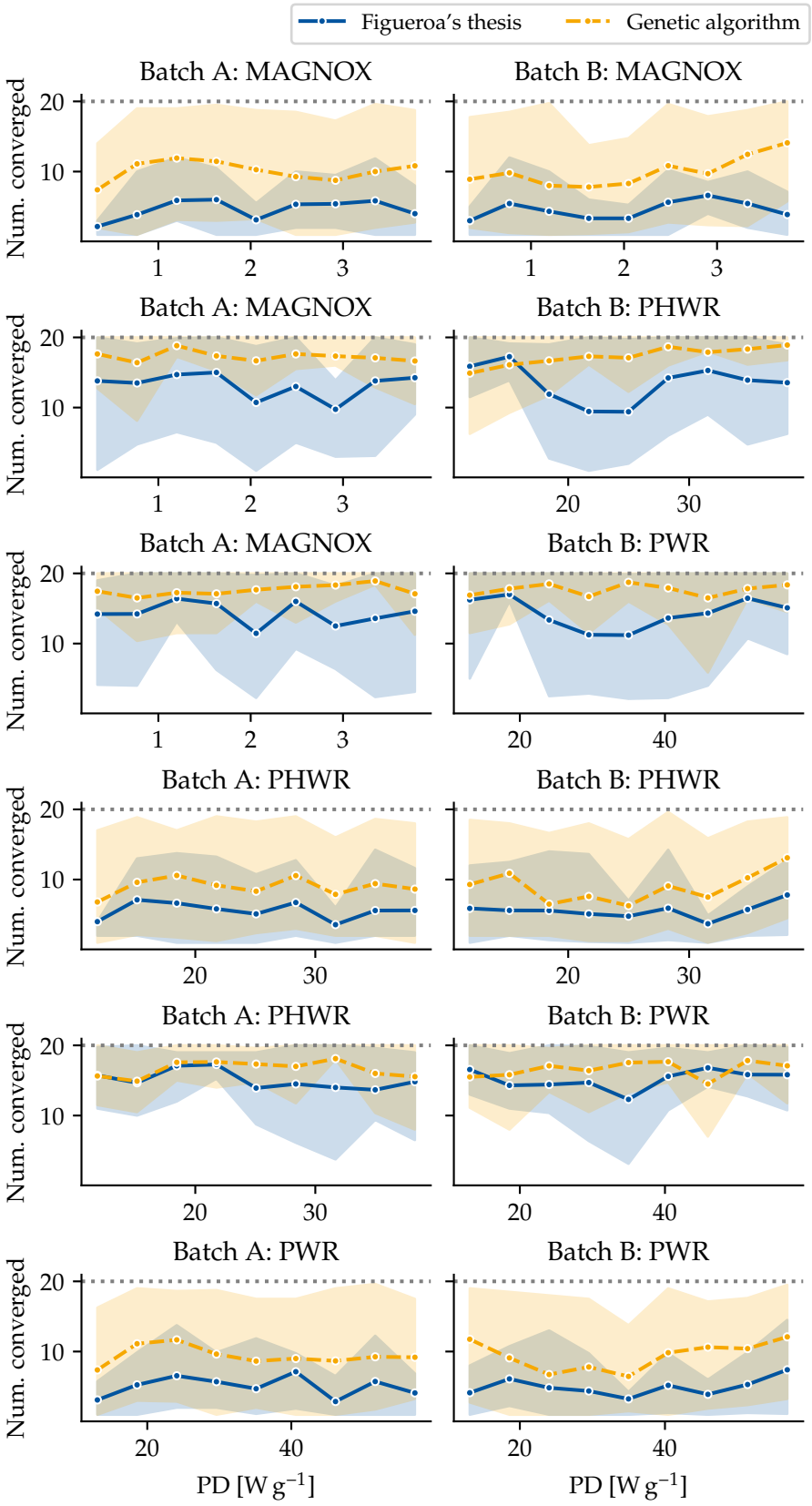


Figure 26: Mixture scenario convergence statistics vs PD.

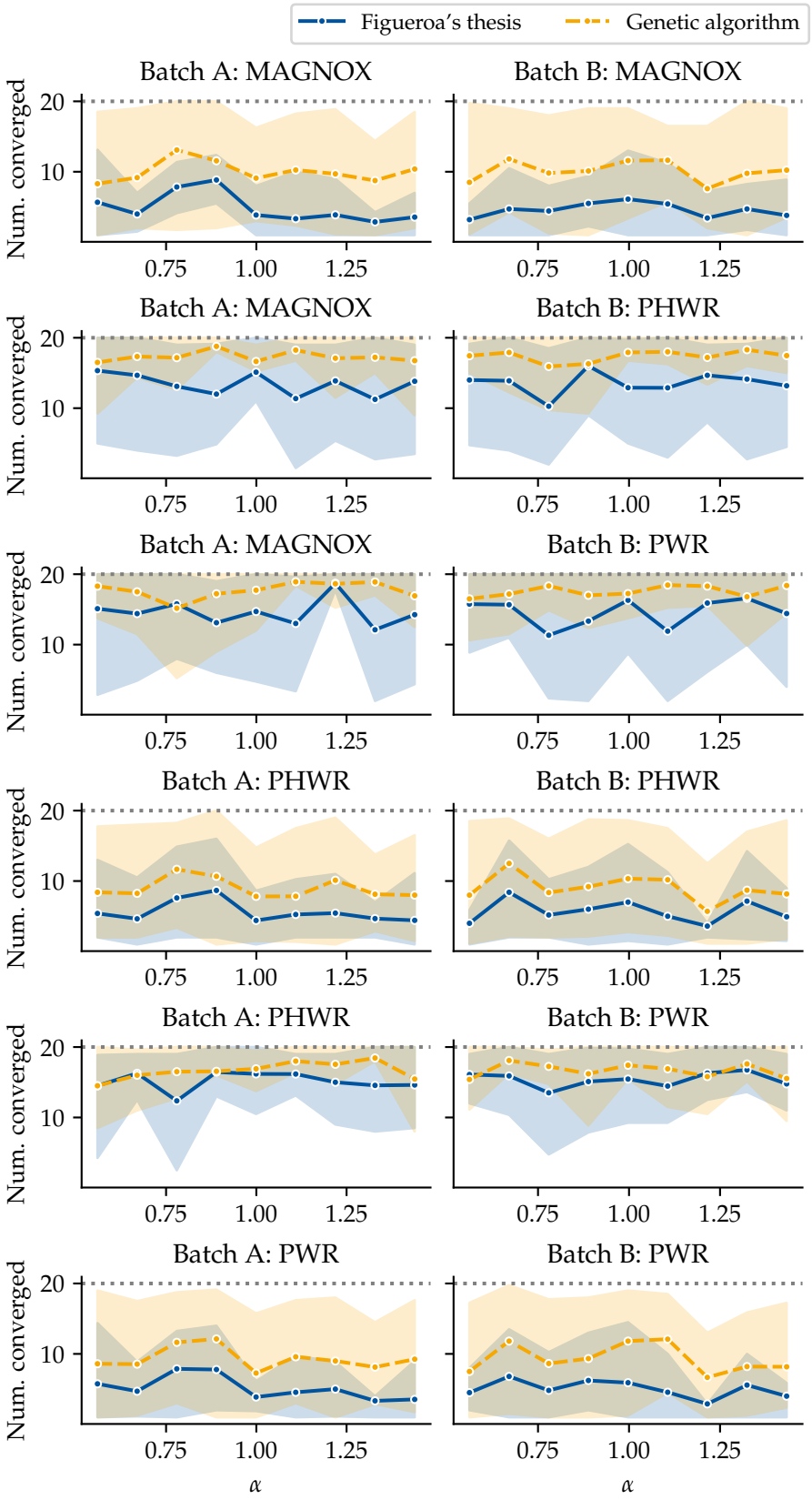


Figure 27: Mixture scenario convergence statistics vs α.

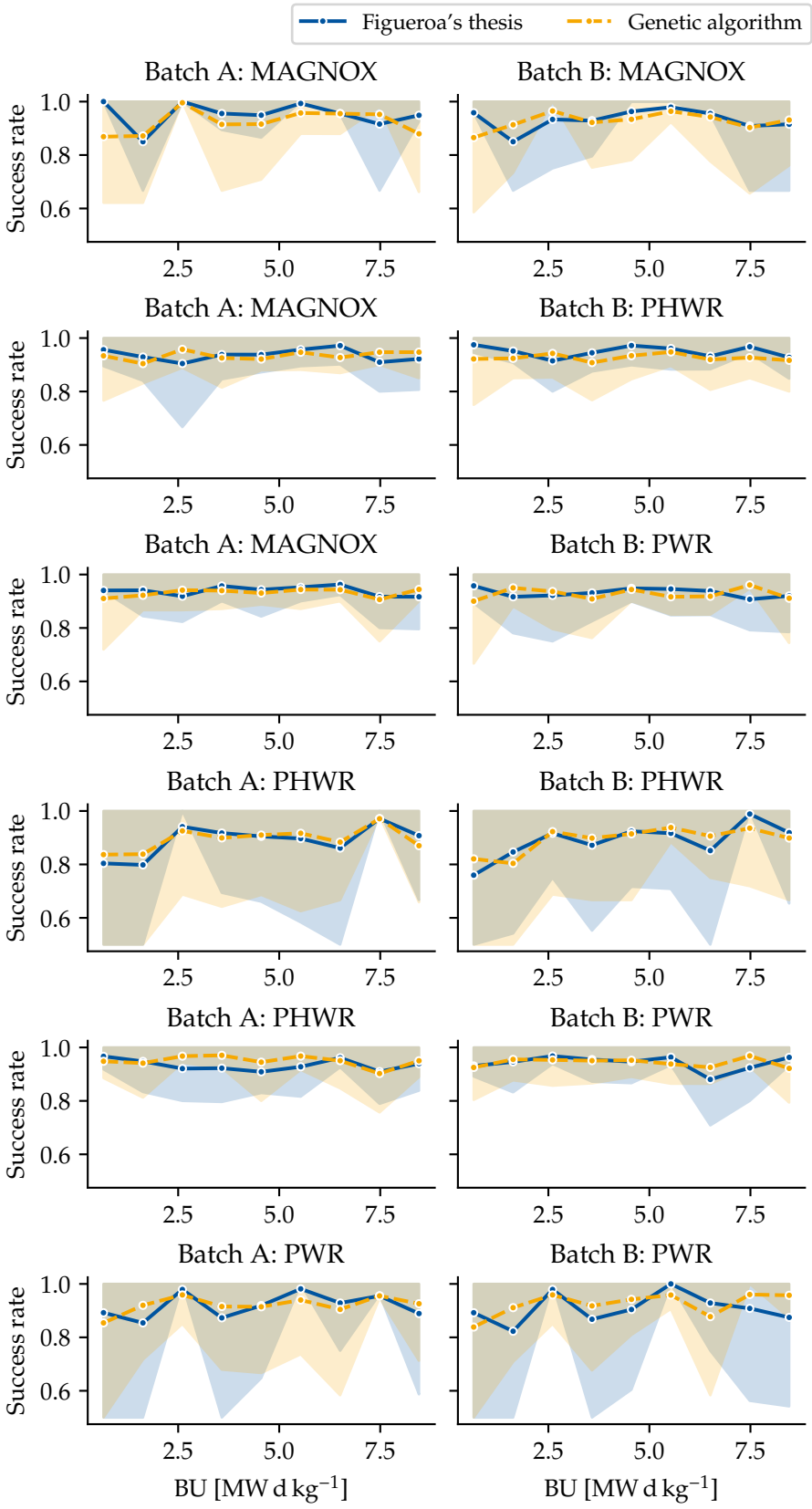


Figure 28: Mixture scenario success statistics vs BU.

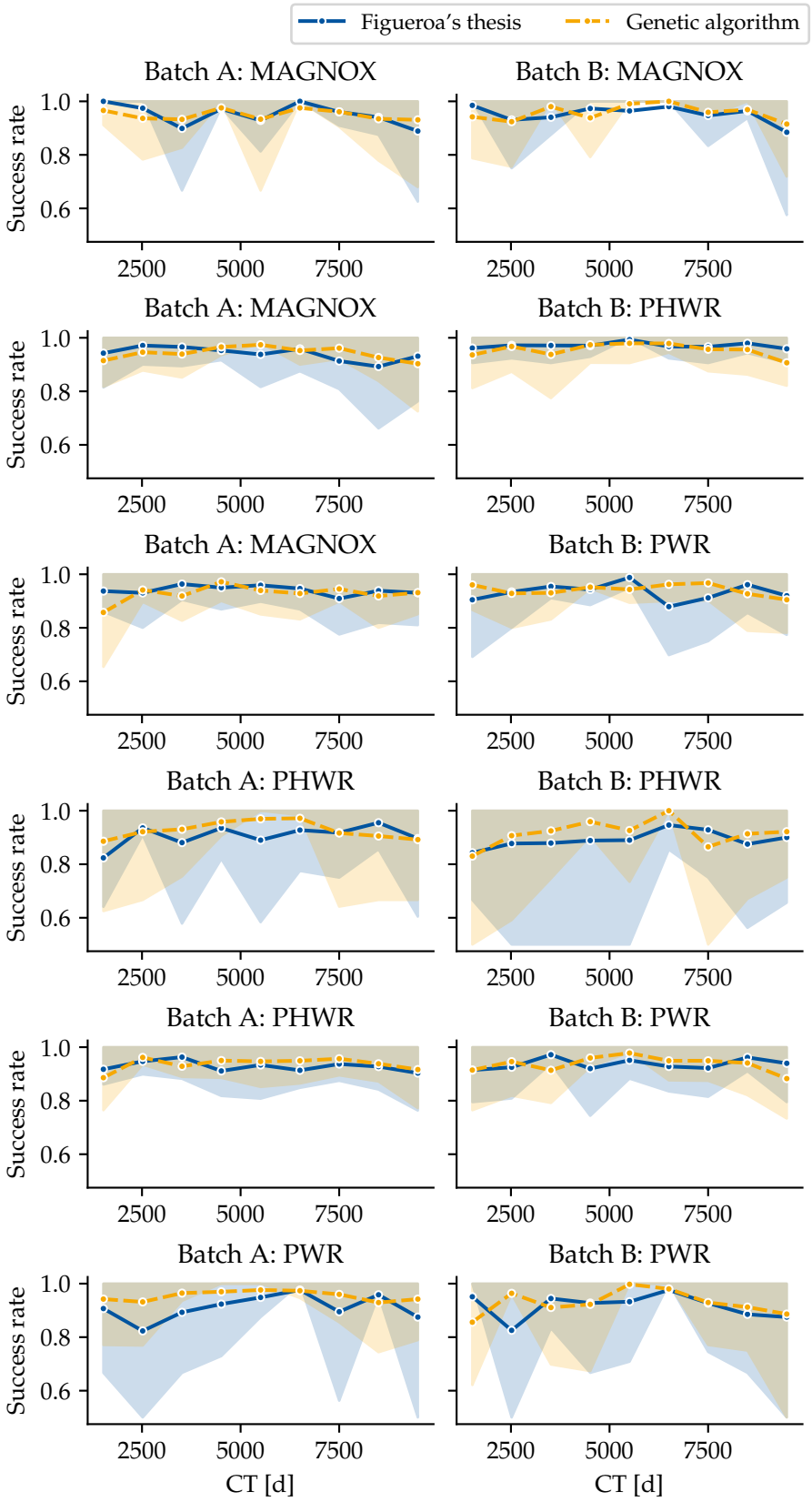


Figure 29: Mixture scenario success statistics vs CT.

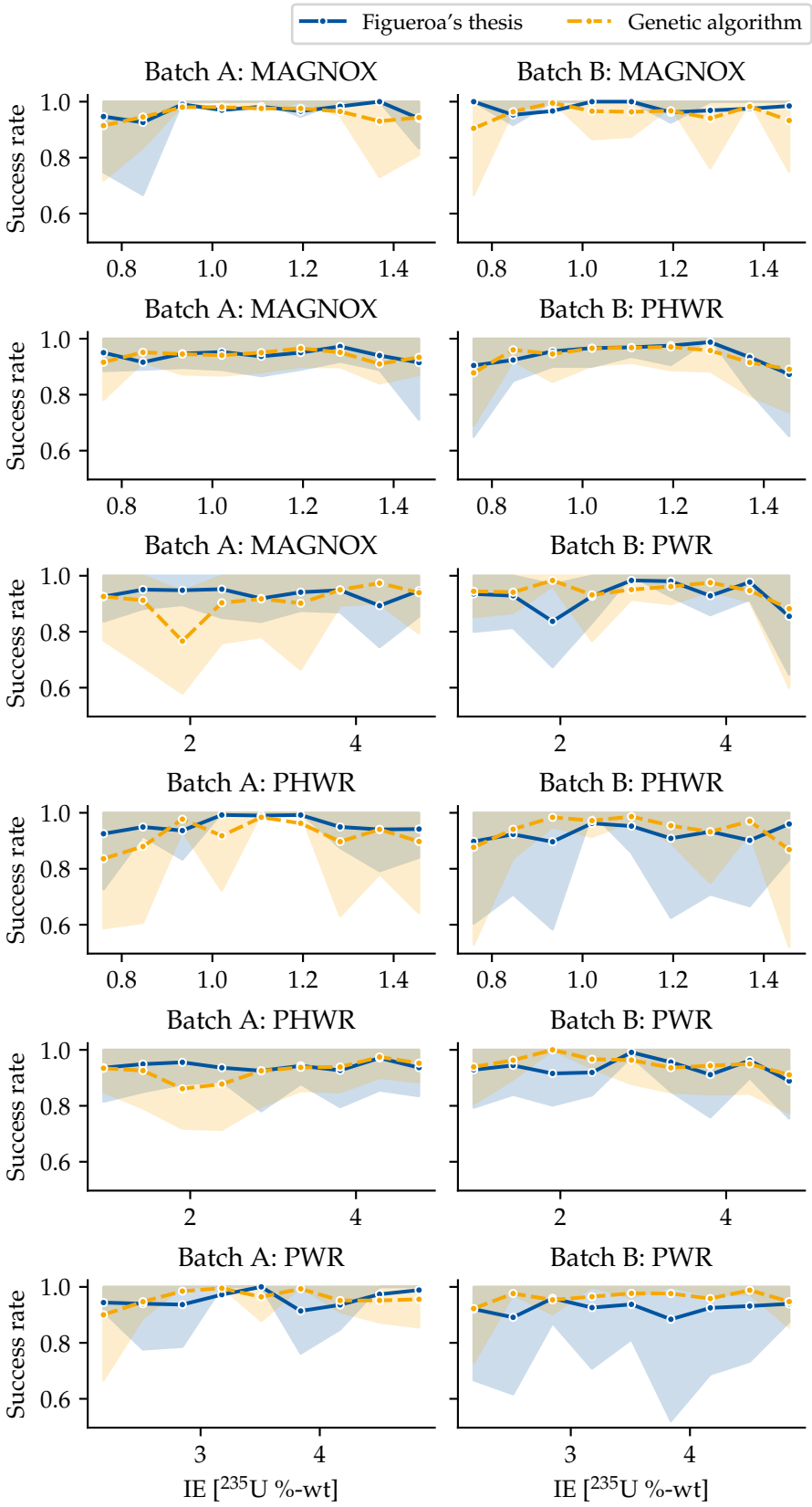


Figure 30: Mixture scenario success statistics vs IE.

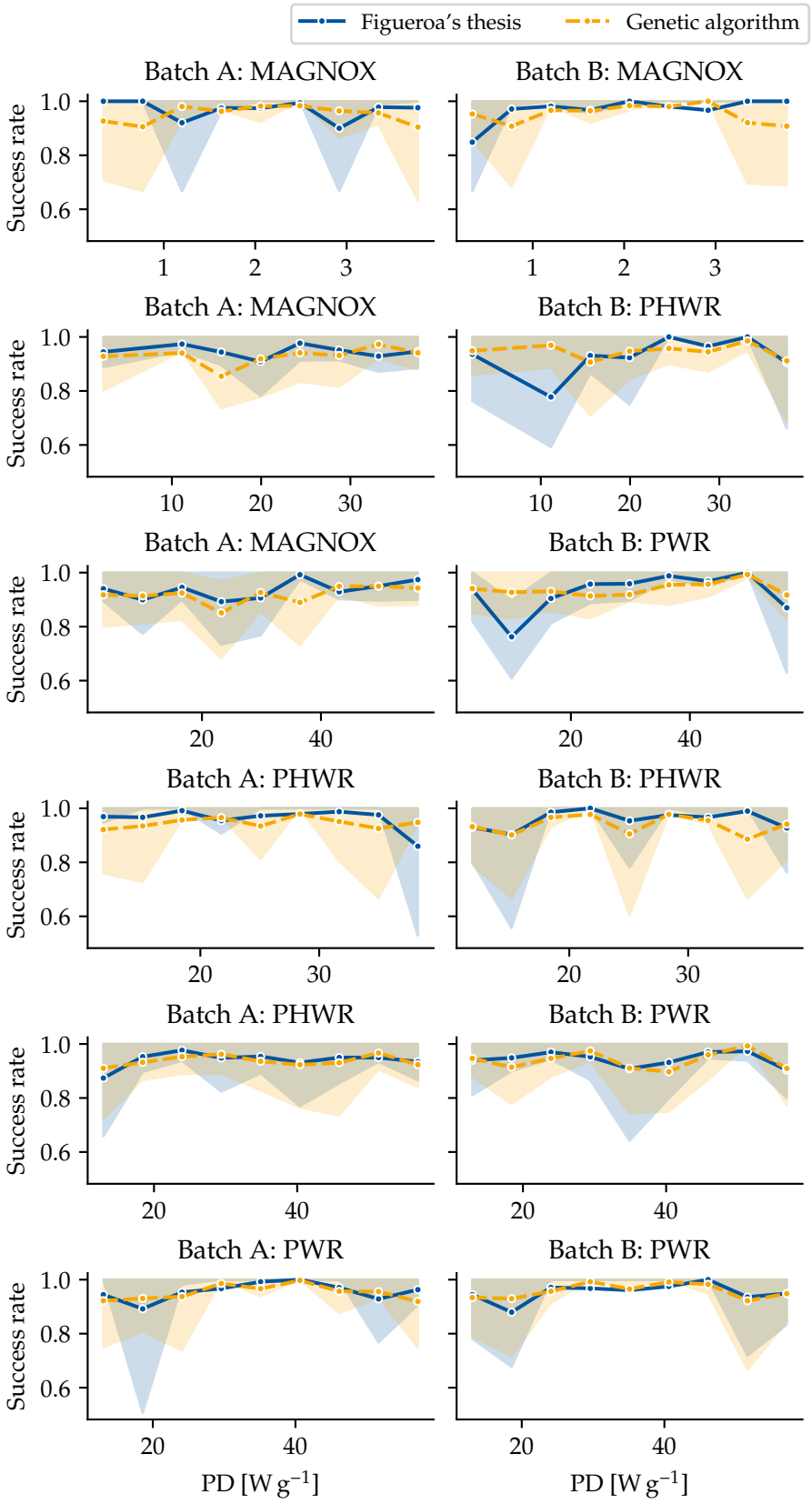


Figure 31: Mixture scenario success statistics vs PD.

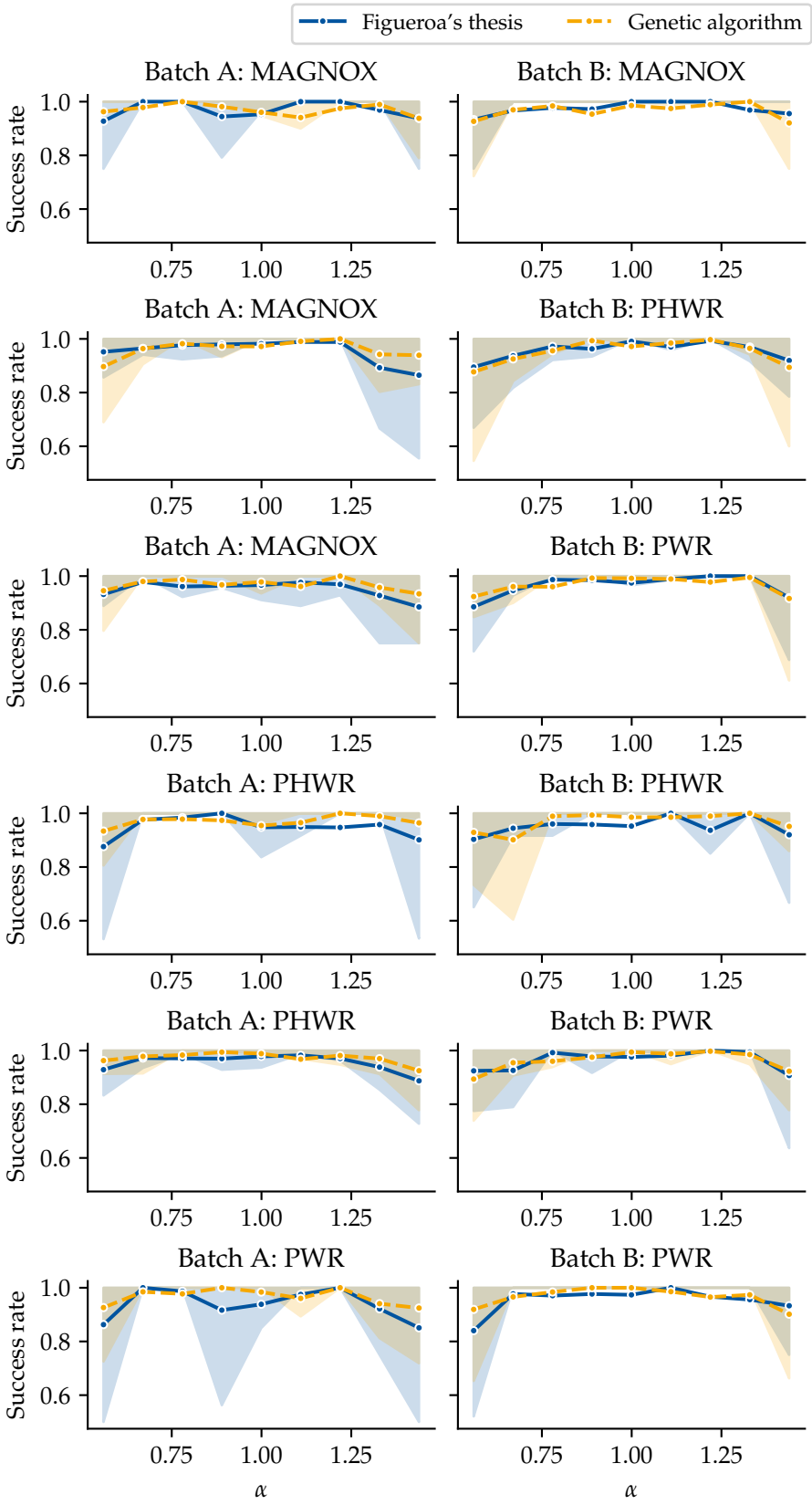


Figure 32: Mixture scenario success statistics vs α.

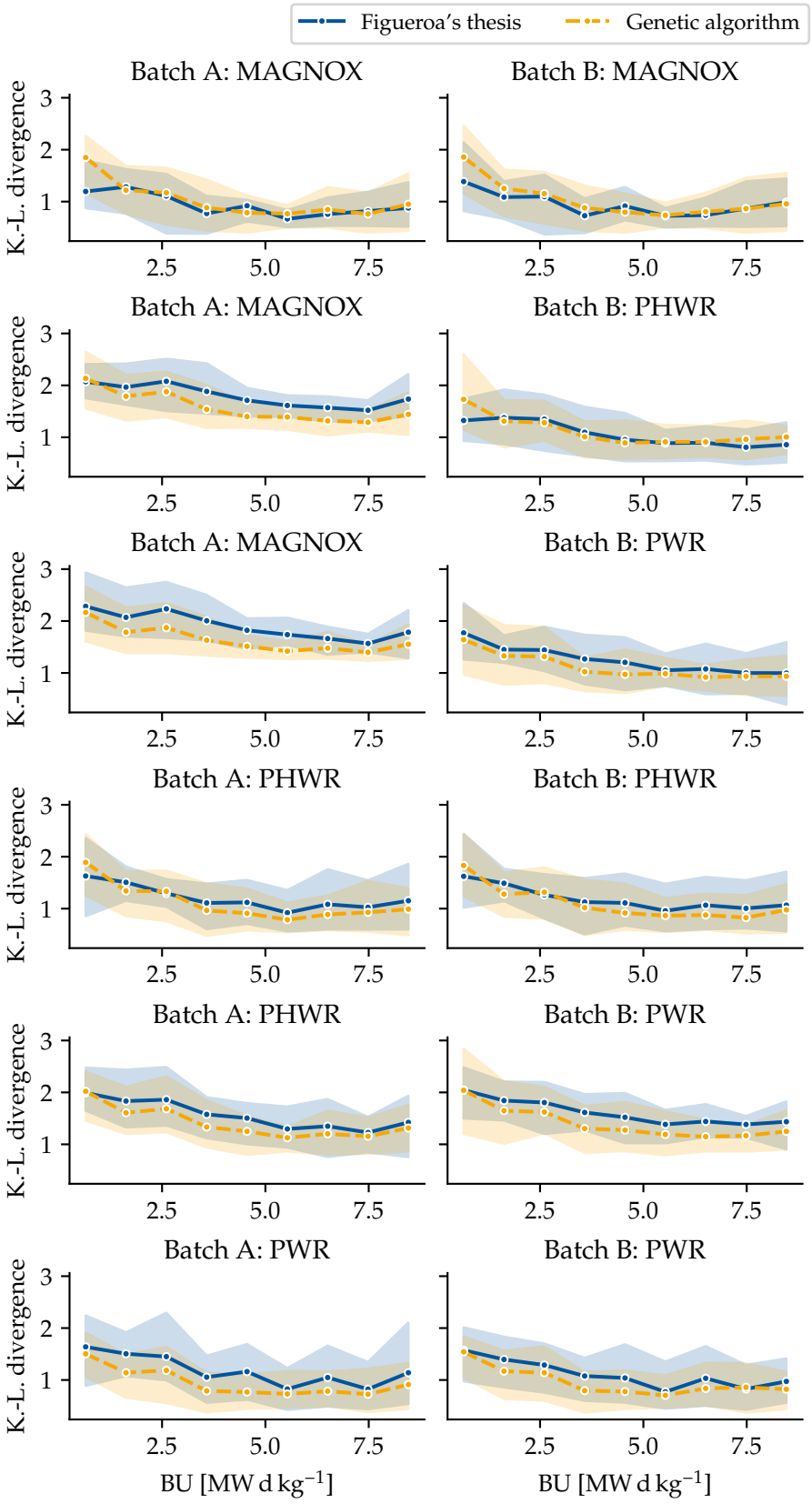


Figure 33: Mixture scenario information gain statistics vs BU.

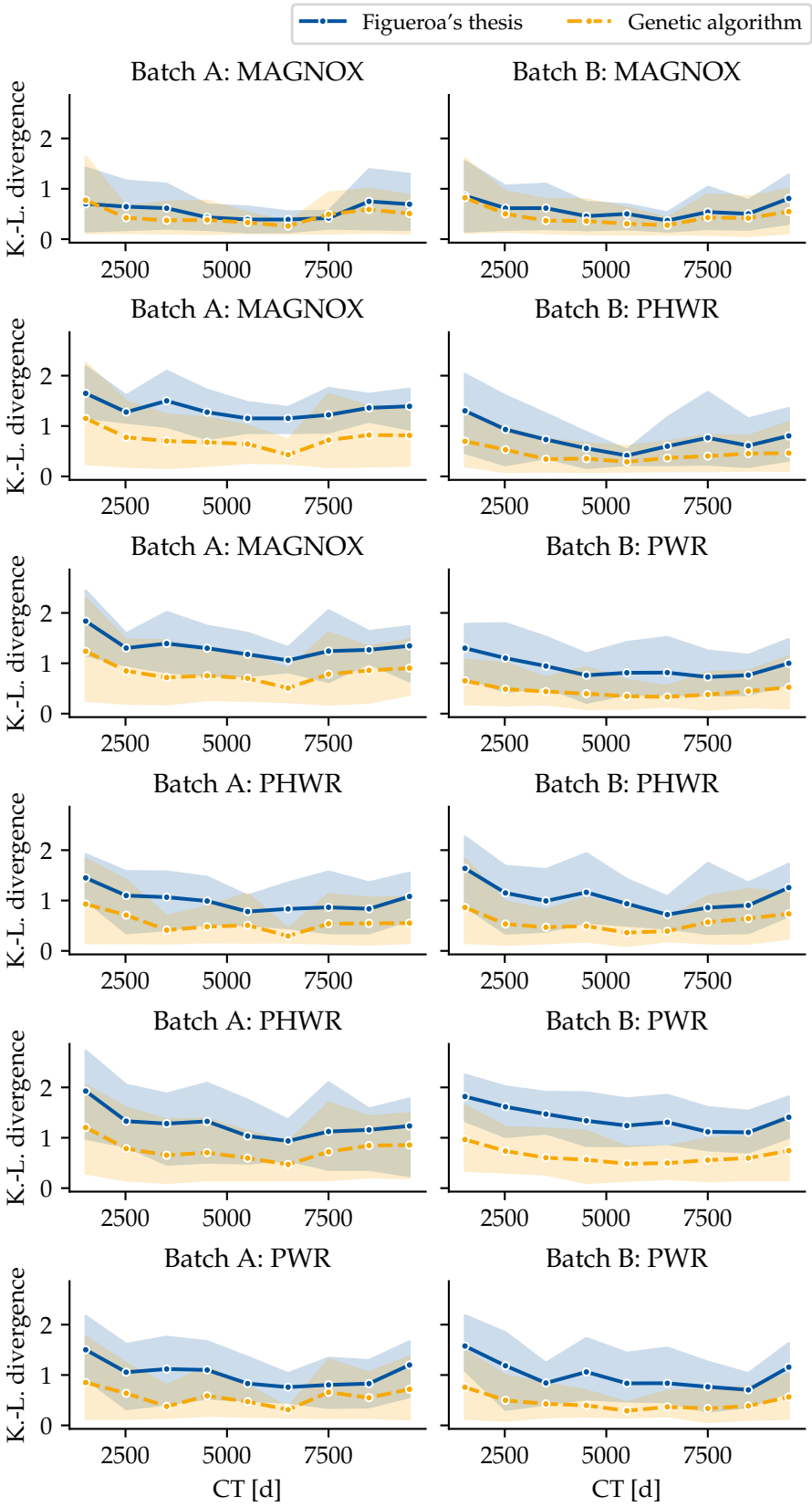


Figure 34: Mixture scenario information gain statistics vs CT.

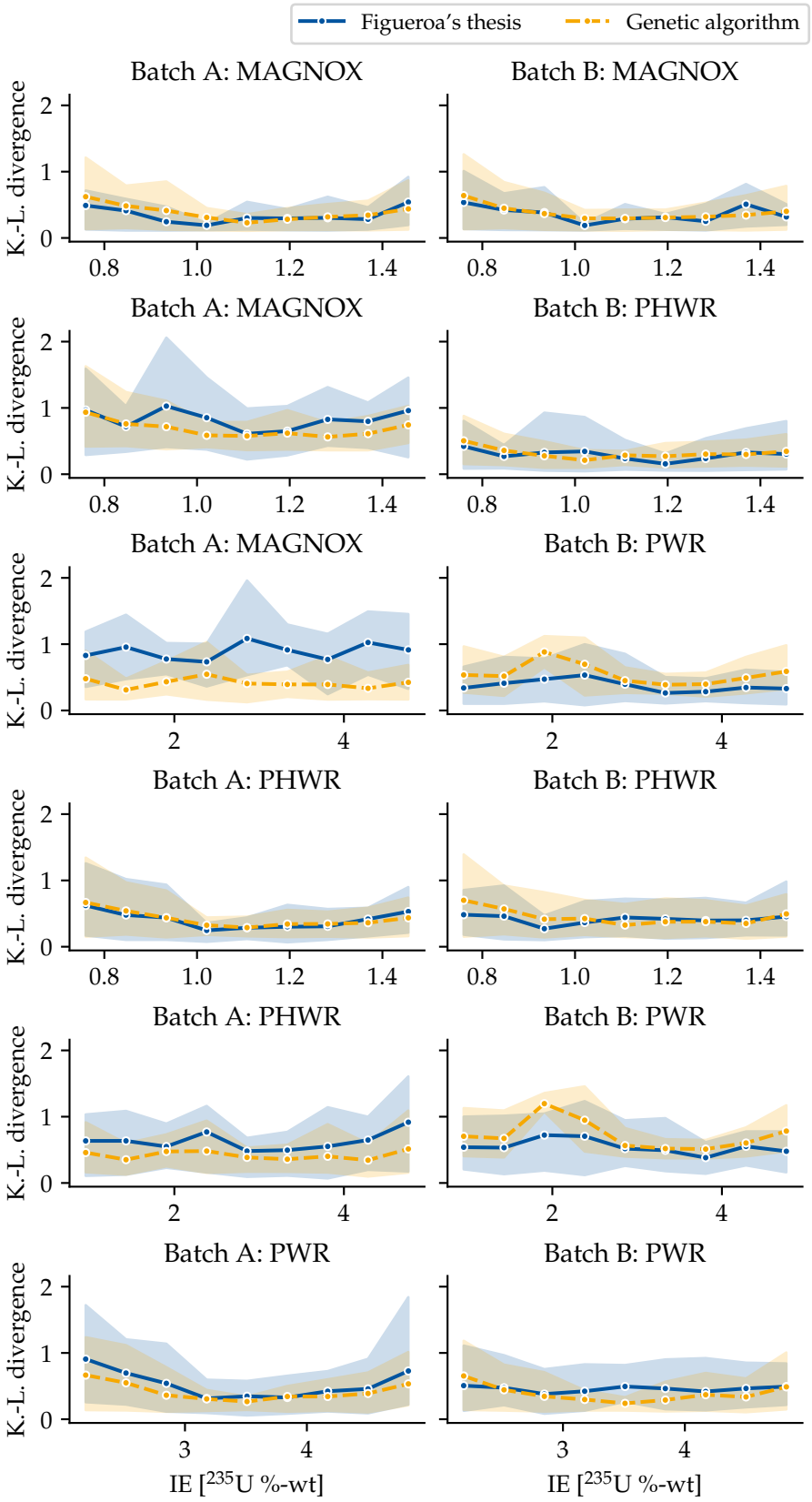


Figure 35: Mixture scenario information gain statistics vs IE.

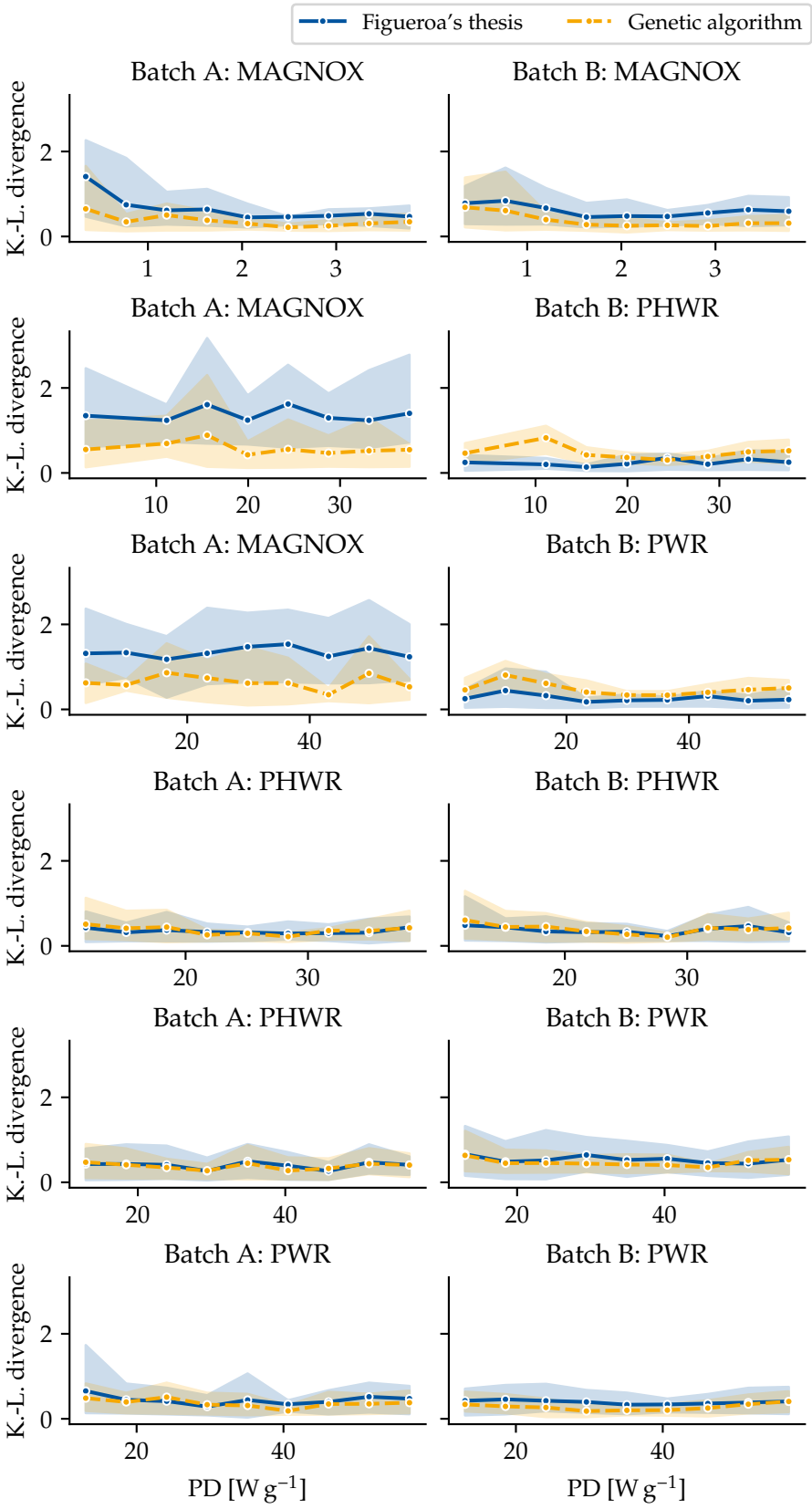


Figure 36: Mixture scenario information gain statistics vs PD.

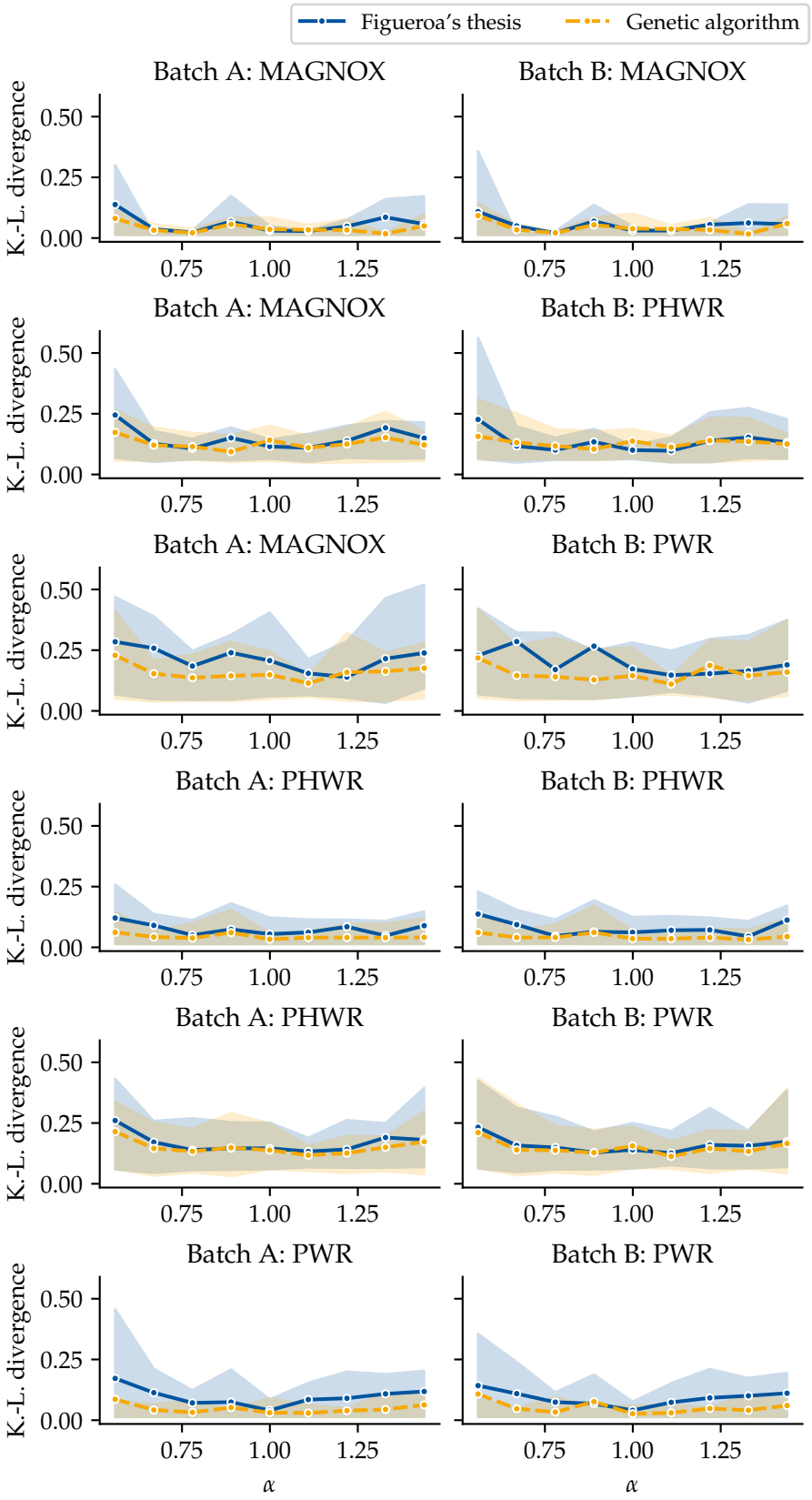


Figure 37: Mixture scenario information gain statistics vs α.

B.2 TEST STATISTICS OF THE CLASSIFIER SCENARIO

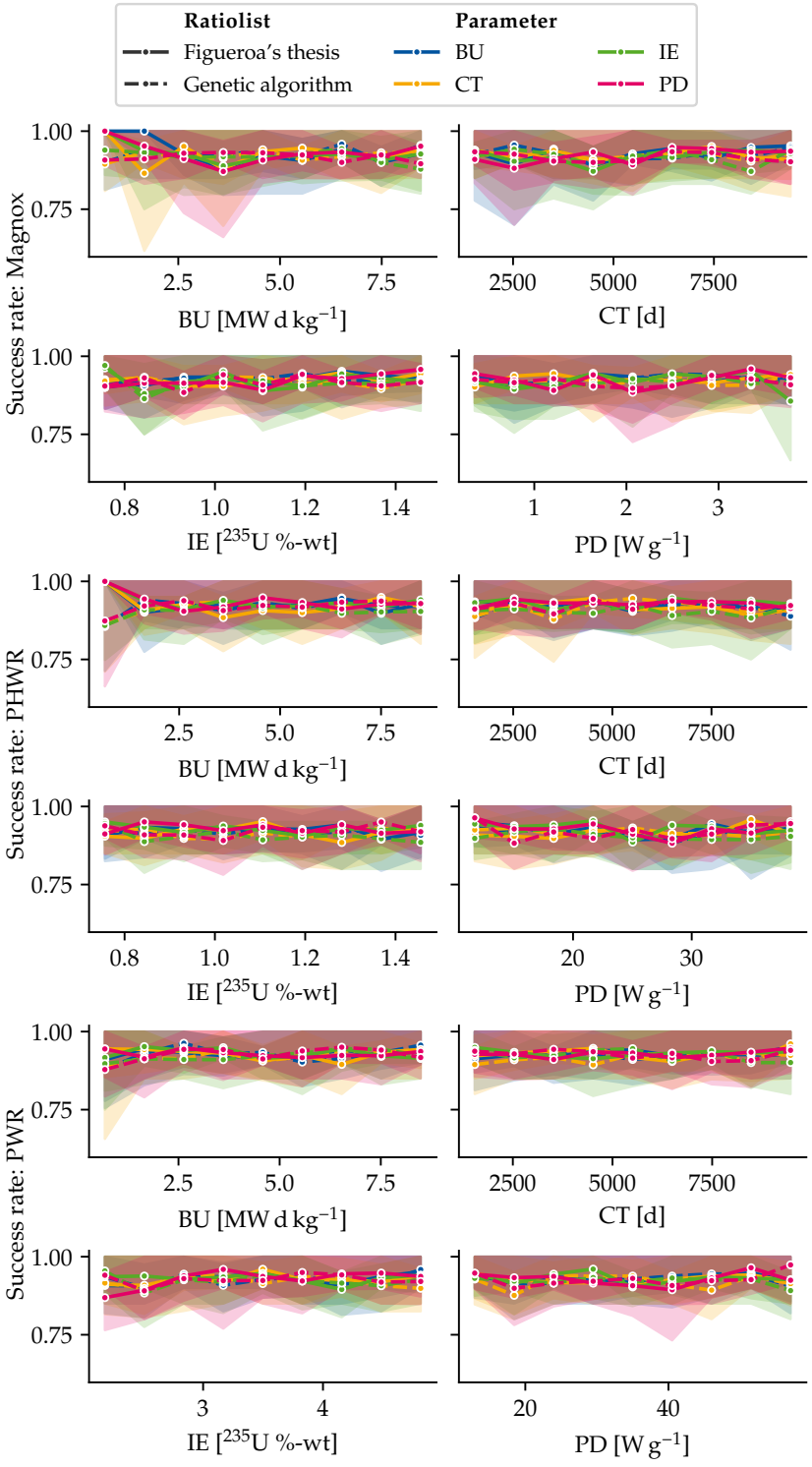


Figure 38: Success statistics of the classifier scenario dependent on the ground truth parameter values.

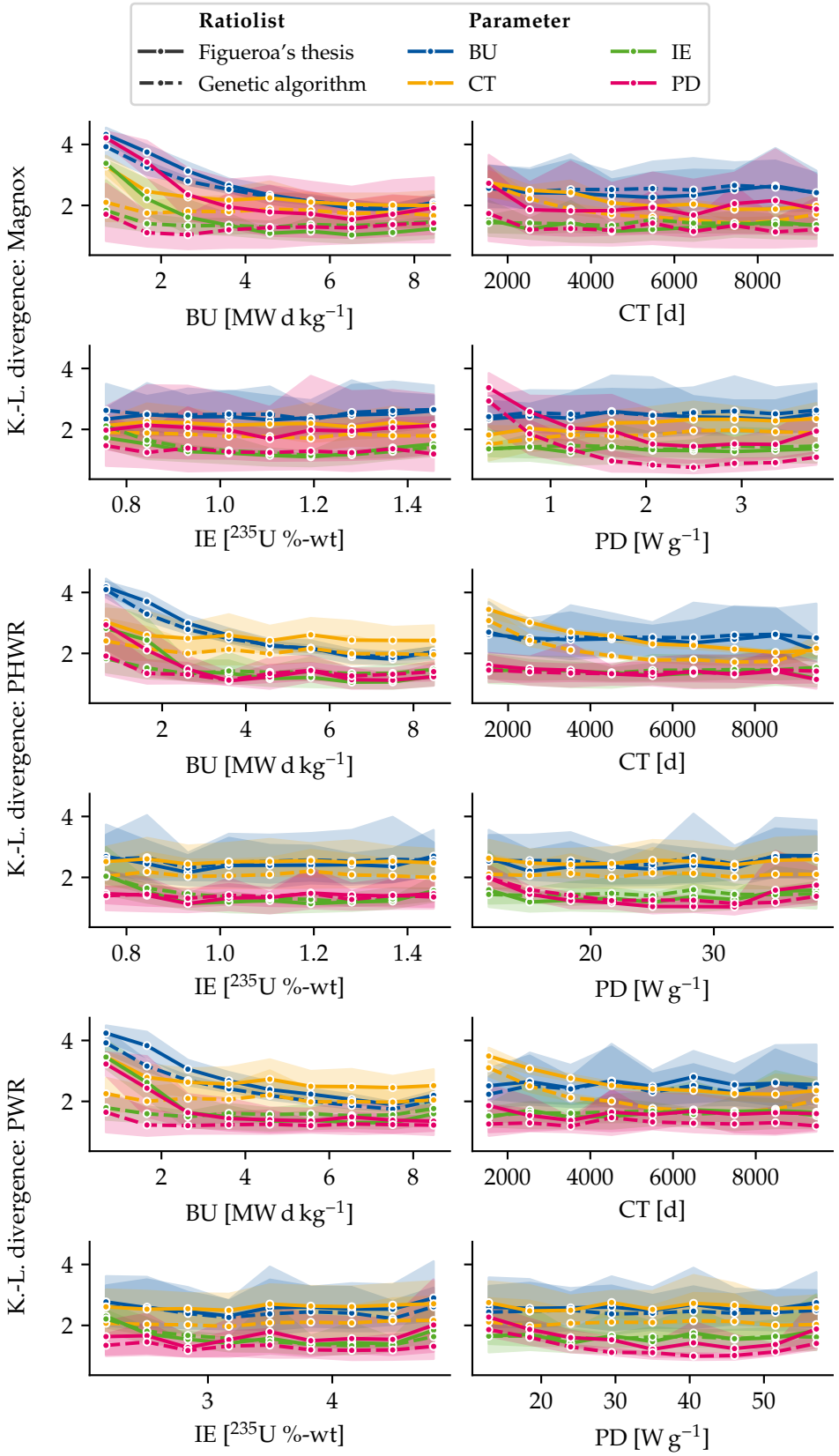


Figure 39: Information gain statistics of the classifier test scenario in dependence of the ground truth parameter values.

BIBLIOGRAPHY

[1] M. Solly, "Nine Harrowing Eyewitness Accounts of the Bombings of Hiroshima and Nagasaki," Smithsonian Magazine (2020).

[2] *NUKEMAP by Alex Wellerstein*, NUKEMAP by Alex Wellerstein, https://nuclearsecrecy.com/nukemap/ (visited on 01/23/2026).

[3] *NukeEXPO - Online report*, Norwegian People's Aid, https://www.npaid.org/news/nukeexpo-online-report (visited on 07/02/2024).

[4] H. M. Kristensen and M. Korda, "World nuclear forces," in *SIPRI Yearbook 2025*, SIPRI Yearbook (SIPRI, 2025).

[5] L. Xia et al., "Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection," Nature Food **3**, 586–596 (2022) DOI: 10.1038/s43016-022-00573-0.

[6] M. Göttsche and A. Glaser, eds., *Toward Nuclear Disarmament: Building up Transparency and Verification* (Federal Foreign Office, 2021).

[7] A. Muti, G. Christopher, and N. Stott, "The Role of Verification in Supporting Irreversible Nuclear Disarmament," Journal for Peace and Nuclear Disarmament **6**, 292–302 (2023) DOI: 10.1080/25751654.2023.2291892.

[8] I. Susenau, "The NPT and IAEA safeguards," IAEA Bulletin **62** (2021).

[9] *INFCIRC/153(Corrected) - The Structure and Content of Agreements Between the Agency and States Required in Connection with the Treaty on the Non-Proliferation of Nuclear Weapons*, IAEA, 1972.

[10] P. Podvig, A. S. Diakov, A. Glaser, and M. Göttsche, *FM(C)T meeting series - Verifiable declarations of fissile material stocks: challenges and solutions* (UNIDIR, Geneva, 2017).

[11] O. Heinonen, "Verifying the dismantlement of South Africa's nuclear weapons program," in *Nuclear weapons materials gone missing: what does history teach?* (2014), pp. 163–184.

[12] F. von Hippel, "Warhead and fissile-material declarations," in *Reversing the arms race: How to achieve and verify deep reductions in the nuclear arsenals*, Science & Global Security Monograph Series (Gordon and Breach, New York, 1990), pp. 61–81.

[13] S. Fetter, "Nuclear archaeology: Verifying declarations of fissile-material production," Science & Global Security **3**, 237–259 (1993) DOI: 10.1080/08929889308426386.

[14] T. W. Wood et al., "The Future of Nuclear Archaeology: Reducing Legacy Risks of Weapons Fissile Material," Science & Global Security **22**, 4–26 (2014) DOI: 10.1080/08929882.2014.874217.

[15] O. Reistad, A. Glaser, R. D. Frank, and S. H. Kaald, "Document-Based Nuclear Archaeology," Science & Global Security **30**, 95–107 (2022) DOI: 10.1080/08929882.2022.2111880.

[16] S. Kretzschmar, L. Rademacher, and M. Göttsche, "Reconstructing nuclear histories with archives – hands-on lessons learned," in Proceedings of the INMM 65th Annual Meeting (July 21–25, 2024).

[17] M. Sharp, "Applications and limitations of nuclear archaeology in uranium enrichment plants," Science & Global Security **21**, 70–92 (2013) DOI: 10.1080/08929882.2013.755028.

[18] S. Philippe and A. Glaser, "Nuclear archaeology for gaseous diffusion enrichment plants," Science & Global Security **22**, 27–49 (2014) DOI: 10.1080/08929882.2014.871881.

[19] D. Willingham, B. E. Naes, K. A. Burns, and B. D. Reid, "Secondary ion mass spectrometry signatures for verifying declarations of fissile-material production," Applied Radiation and Isotopes **97**, 125–129 (2015) DOI: 10.1016/j.apradiso.2014.12.015.

[20] R. C. Connick et al., "Measuring very low radiation doses in PTFE for nuclear forensic enrichment reconstruction," Radiation Physics and Chemistry **226**, 112256 (2025) DOI: 10.1016/j.radphyschem.2024.112256.

[21] C. J. Gesh, *A Graphite Isotope Ratio Method: A Primer on Estimating Plutonium Production in Graphite Moderated Reactors*, PNNL-14568 (Pacific Northwest National Laboratory, Feb. 2004), DOI: 10.2172/15010619.

[22] A. Gasner and A. Glaser, "Nuclear Archaeology for Heavy-Water-Moderated Plutonium Production Reactors," Science & Global Security **19**, 223–233 (2011) DOI: 10.1080/08929882.2011.616124.

[23] D. C. Gerlach et al., *Final Report on Isotope Ratio Techniques for Light Water Reactors*, PNNL-18573 (July 2009), DOI: 10.2172/986731.

[24] J. de Troullioud de Lanversin, M. Göttsche, and A. Glaser, "Nuclear Archaeology to Distinguish Plutonium and Tritium Production Modes in Heavy-Water Reactors," Science & Global Security **26**, 70–90 (2018) DOI: 10.1080/08929882.2018.1518693.

[25] B. Jung and M. Göttsche, "Assessing Uncertainty in Plutonium Production Estimates Based on the Isotope Ratio Method," Science & Global Security **30**, 3–21 (2022) DOI: 10.1080/08929882.2022.2060599.

[26] L. Rademacher and M. Göttsche, "Forensic Measurements for Nuclear Archaeology – Trawsfynydd revisited," in Proceedings of the INMM 65th Annual Meeting (July 21–25, 2024).

[27] M. Göttsche, "Nuclear archaeology to assess fissile-material inventories for nuclear safety and security, transparency, and disarmament," The Nonproliferation Review **30**, 131–150 (2024) DOI: 10.1080/10736700.2024.2314371.

[28] M. Schalz, L. Bormann, and M. Göttsche, "Bicyclus: A statistical fuel cycle simulation framework for nuclear verification," Annals of Nuclear Energy **196**, 110221 (2024) DOI: 10.1016/j.anucene.2023.110221.

[29] A. Figueroa, "A novel method for reconstructing a nuclear reactor's operation history based on reprocessing waste measurements," PhD thesis (RWTH Aachen University, Aachen, 2024).

[30] W. M. Stacey, *Nuclear reactor physics*, 2nd ed., completely rev. and enlarged (Wiley-VCH, Weinheim, 2007), 706 pp.

[31] USDOE, *Plutonium: The first 50 years. United States plutonium production, acquisition, and utilization from 1944 through 1994*, DOE/DP–0137, 219368 (Washington DC (United States), Feb. 1, 1996), DOI: 10.2172/219368.

[32] J. C. Mark, "Explosive properties of reactor-grade plutonium," Science & Global Security **4**, 111–128 (1993) DOI: 10.1080/08929889308426394.

[33] J. R. Lamarsh and A. J. Baratta, *Introduction to nuclear engineering*, 3rd ed, Addison-Wesley Series in Nuclear Science and Engineering (Prentice Hall, Upper Saddle River, N.J, 2001), 783 pp.

[34] E. R. Irish and W. H. Reas, *The PUREX process - A solvent extraction reprocessing method for irradiated uranium*, HW-49483 A, 4341712 (Apr. 1, 1957), DOI: 10.2172/4341712.

[35] E. R. Irish, *Description of Purex Plant Process*, HW–60116, 4248051 (May 19, 1959), DOI: 10.2172/4248051.

[36] P. K. Romano et al., "OpenMC: A state-of-the-art Monte Carlo code for research and development," Annals of Nuclear Energy **82**, 90–97 (2015) DOI: 10.1016/j.anucene.2014.07.048.

[37] P. K. Romano, C. J. Josey, A. E. Johnson, and J. Liang, "Depletion capabilities in the OpenMC Monte Carlo particle transport code," Annals of Nuclear Energy **152**, 107989 (2021) DOI: 10.1016/j.anucene.2020.107989.

[38] J. de Troullioud de Lanversin, M. Kütt, and A. Glaser, "ONIX: An open-source depletion code," Annals of Nuclear Energy **151**, 107903 (2021) DOI: 10.1016/j.anucene.2020.107903.

[39] D. G. Cacuci, ed., *Handbook of nuclear engineering*, Springer Reference (Springer, New York ; London, 2010), 5 pp.

[40] C. Moler and C. Van Loan, "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later," SIAM Review **45**, 3–49 (2003) DOI: 10.1137/S00361445024180.

[41] M. Pusa, "Rational Approximations to the Matrix Exponential in Burnup Calculations," Nuclear Science and Engineering **169**, 155–167 (2011) DOI: 10.13182/NSE10-81.

[42] A. Hébert, "Multigroup Neutron Transport and Diffusion Computations," in *Handbook of Nuclear Engineering*, edited by D. G. Cacuci (Springer US, Boston, MA, 2010), pp. 751–911, DOI: 10.1007/978-0-387-98149-9_8.

[43] OpenMC Contributors, *Theory and Methodology — OpenMC Documentation*, https://docs.openmc.org/en/stable/methods/index.html (visited on 05/04/2025).

[44] B. Jung, "Reconstructing Plutonium Production with the Isotope Ratio Method: Uncertainty Assessments," M.Sc. (RWTH Aachen University, Aachen, Oct. 14, 2020).

[45] P. Obložinský, M. Herman, and S. F. Mughabghab, "Evaluated nuclear data," in *Handbook of Nuclear Engineering*, edited by D. G. Cacuci (Springer US, Boston, MA, 2010), pp. 83–187, DOI: 10.1007/978-0-387-98149-9_2.

[46] D. A. Brown, "Status of the ENDF/B Nuclear Data Library," Nuclear Data Sheets **118**, 98–103 (2014) DOI: 10.1016/j.nds.2014.04.010.

[47] A. Gelman et al., *Bayesian Data Analysis, Third Edition*, 3rd ed., Texts in Statistical Science (2013).

[48] P.-R. Wagner, S. Marelli, and B. Sudret, "Bayesian model inversion using stochastic spectral embedding," Journal of Computational Physics **436**, 110141 (2021) DOI: 10.1016/j.jcp.2021.110141.

[49] A. Vehtari, A. Gelman, and J. Gabry, "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC," Statistics and Computing **27**, 1413–1432 (2017) DOI: 10.1007/s11222-016-9696-4.

[50] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)," Bayesian Analysis **13**, 917–1007 (2018) DOI: 10.1214/17-BA1091.

[51] R. Kumar, C. Carroll, A. Hartikainen, and O. Martin, "ArviZ a unified library for exploratory analysis of Bayesian models in Python," Journal of Open Source Software **4**, 1143 (2019) DOI: 10.21105/joss.01143.

[52] M. D. Hoffman and A. Gelman, "The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo," Journal of Machine Learning Research **15**, 1593–1623 (2014).

[53] A. Vehtari et al., "Rank-Normalization, Folding, and Localization: An Improved R^ for Assessing Convergence of MCMC (with Discussion)," Bayesian Analysis **16**, 667–718 (2021) DOI: 10.1214/20-BA1221.

[54] S. Van Winckel, R. Alvarez-Sarandes, D. Serrano Purroy, and L. Aldave de las Heras, "Assay data of spent nuclear fuel: the lab-work behind the numbers," Frontiers in Energy Research **11** (2023) DOI: 10.3389/fenrg.2023.1168460.

[55] A. Figueroa and M. Göttsche, "Gaussian processes for surrogate modeling of discharged fuel nuclide compositions," Annals of Nuclear Energy **156**, 108085 (2021) DOI: 10.1016/j.anucene.2020.108085.

[56] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*, Adaptive Computation and Machine Learning (MIT Press, Cambridge, Mass, 2006), 248 pp.

[57] I. M. Sobol', "On the distribution of points in a cube and the approximate evaluation of integrals," USSR Computational Mathematics and Mathematical Physics **7**, 86–112 (1967) DOI: 10.1016/0041-5553(67)90144-9.

[58] B. Jung, *Jung-benjamin/bram-inferno: v0.8.1*, version v0.8.1, Zenodo, June 2, 2025, DOI: 10.5281/zenodo.15577482.

[59] PyMC-Devs, *PyMC*, version v5.23.0, Zenodo, May 28, 2025, DOI: 10.5281/zenodo.15532438.

[60] B. Jung, *Jung-benjamin/gp-surfer: v0.5.3*, version v0.5.3, Zenodo, June 2, 2025, DOI: 10.5281/zenodo.15576629.

[61] P. Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," Nature Methods **17**, 261–272 (2020) DOI: 10.1038/s41592-019-0686-2.

[62] Nuclear Energy Agency, *Spent nuclear fuel assay data for isotopic validation - state-of-the-art report* (Paris, 2011).

[63] S. S. Chirayath, J. M. Osborn, and T. M. Coles, "Trace Fission Product Ratios for Nuclear Forensics Attribution of Weapons-Grade Plutonium from Fast and Thermal Reactors," Science & Global Security **23**, 48–67 (2015) DOI: 10.1080/08929882.2015.996079.

[64] J. Comte et al., "Determination of selenium-79 in solutions of fission products after pre-treatment by ion exchange chromatography and ETV-ICP-MS," Journal of Analytical Atomic Spectrometry **18**, 702 (2003) DOI: 10.1039/b209253k.

[65] S. Van Winckel, *Personal communication*, Aug. 15, 2025.

[66] A. Favalli et al., "Determining initial enrichment, burnup, and cooling time of pressurized-water-reactor spent fuel assemblies by analyzing passive gamma spectra measured at the Clab interim-fuel storage facility in Sweden," Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **820**, 102–111 (2016) DOI: 10.1016/j.nima.2016.02.072.

[67] O. Kramer, *Genetic Algorithm Essentials*, Vol. 679, Studies in Computational Intelligence (Springer International Publishing, Cham, 2017), DOI: 10.1007/978-3-319-52156-5.

[68] C. R. Harris et al., "Array programming with NumPy," Nature **585**, 357–362 (2020) DOI: 10.1038/s41586-020-2649-2.

[69] B. Jung, *Jung-benjamin/rasor: v0.3.1*, version v0.3.1, Zenodo, Apr. 22, 2026, DOI: 10.5281/zenodo.19696049.

[70] S. Kullback and R. A. Leibler, "On Information and Sufficiency," The Annals of Mathematical Statistics **22**, 79–86 (1951) DOI: 10.1214/aoms/1177729694.

[71] B. D. Roach et al., "The need for speed – Burnup determination of spent nuclear fuel," Talanta Open **6**, 100152 (2022) DOI: 10.1016/j.talo.2022.100152.

[72] N. N. Wanna et al., "Quantification of uranium, plutonium, neodymium and gadolinium for the characterization of spent nuclear fuel using isotope dilution HPIC-SF-ICP-MS," Talanta **221**, 121592 (2021) DOI: 10.1016/j.talanta.2020.121592.

[73] V. Mishra, Zs. Elter, E. Branger, and S. Grape, "Statistical analysis of fuel cycle data from Swedish Pressurized Water Reactors and the impact of simplifying assumptions on simulated nuclide inventories," Progress in Nuclear Energy **155**, 104483 (2023) DOI: 10.1016/j.pnucene.2022.104483.

[74] J. M. Osborn et al., "Experimental validation of a nuclear forensics methodology for source reactor-type discrimination of chemically separated plutonium," Nuclear Engineering and Technology **51**, 384–393 (2019) DOI: 10.1016/j.net.2018.11.003.

[75] B. Jung and M. Göttsche, "Inferring reactor parameters from nuclear waste with Bayesian inference," in Proceedings of the INMM 65th Annual Meeting (July 21–25, 2024).

[76] B. Jung, A. Figueroa, and M. Göttsche, "Towards inferring reactor operations from high-level waste," Nuclear Engineering and Technology (2024) DOI: 10.1016/j.net.2024.02.031.

[77] W. H. Carlton, A. A. Simpkins, and G. T. Jannik, *Radionuclides in the Savannah River Site Environment*, WSRC-MS-99-00667 (1999).

[78] R. Chaplin, "Genealogy of CANDU reactors," in *The essential CANDU* (2016).

[79] S. E. Jensen and E. Nonbol, *Description of the Magnox Type of Gas Cooled Reactor (MAGNOX)*, NKS-2 (Riso National Laboratory, Nov. 1998).

[80] G. Radulescu, I. C. Gauld, and G. Ilas, *SCALE 5.1 Predictions of PWR Spent Nuclear Fuel Isotopic Compositions*, ORNL/TM-2010/44, 983556 (Mar. 1, 2010), DOI: 10.2172/983556.

[81] J. Bosse, "Applying Bayesian inference-based nuclear archaeology to a simulated case study of DPRK," M.Sc. (RWTH Aachen University, Aachen, June 2024).

[82] Z. Varga et al., "Trends and perspectives in Nuclear Forensic Science," TrAC Trends in Analytical Chemistry **146**, 116503 (2022) DOI: 10.1016/j.trac.2021.116503.

[83] F. Michel-Sendis et al., "SFCOMPO-2.0: An OECD NEA database of spent nuclear fuel isotopic assays, reactor design specifications, and operating data," Annals of Nuclear Energy **110**, 779–788 (2017) DOI: 10.1016/j.anucene.2017.07.022.

[84] I. I. Al-Qasir et al., "Neutron thermalization in nuclear graphite: A modern story of a classic moderator," Annals of Nuclear Energy **161**, 108437 (2021) DOI: 10.1016/j.anucene.2021.108437.

[85] *Evaluated Nuclear Structure Data File (ENSDF)* (National Nuclear Data Center; Brookhaven National Laboratory (BNL), Upton, NY (United States), 2022), DOI: 10.18139/NNDC.ENSDF/1845010.

[86] M. Flory, *Application of Safeguards in the Democratic People's Republic of Korea*, GOV2022/40-GC(66)/16 (International Atomic Energy Agency, Sept. 7, 2022).

[87] A. Glaser and Z. Mian, "Denuclearizing North Korea: A verified, phased approach," Science **361**, 981–983 (2018) DOI: 10.1126/science.aau4817.

[88] J. Kang, "Using the Graphite Isotope Ratio Method to Verify the DPRK's Plutonium-Production Declaration," Science & Global Security **19**, 121–129 (2011) DOI: 10.1080/08929882.2011.586309.

[89] J. de Troullioud de Lanversin and M. Kütt, "Verifying North Korea's Plutonium Production with Nuclear Archaeology," Science & Global Security **29**, 145–166 (2021) DOI: 10.1080/08929882.2021.1999054.

[90] B. Jung, J. Bosse, and M. Göttsche, "Reconstructing North Korea's plutonium production history with Bayesian inference-based reprocessing waste analysis," Nuclear Engineering and Technology **58**, 103849 (2026) DOI: 10.1016/j.net.2025.103849.

[91] S. S. Hecker, C. Braun, and C. Lawrence, "North Korea's Stockpiles of Fissile Material," Korea Observer **47**, 721–749 (2016).

[92] P. Makowsky, J. Liu, M. Williams, and I. Ragnone, "Yongbyon Nuclear Scientific Research Center: Construction nearing completion at Radiochemical Laboratory," 38 North (2025).

[93] D. Albright and K. O'Neill, *Solving the North Korean Nuclear Puzzle* (The Institute for Science and International Security, Washington DC (United States), 2000).

[94] S. S. Hecker, *North Korea's Yongbyon Nuclear Complex: A Report by Siegfried S. Hecker* (Center for International Security and Cooperation, Stanford University, Nov. 20, 2010).

[95] G. H. Park and S. G. Hong, "An estimation of weapon-grade plutonium production from 5 MWe YongByon reactor through MCNP6 core depletion analysis," Progress in Nuclear Energy **130**, 103533 (2020) DOI: 10.1016/j.pnucene.2020.103533.

[96] S. Yang et al., "Estimating North Korea's nuclear capabilities: Insights from a study on tritium production in a 5MWe graphite-moderated reactor," Nuclear Engineering and Technology **56**, 2666–2675 (2024) DOI: 10.1016/j.net.2024.02.026.

[97] C. Gustavsson et al., "Modeling Plutonium Production in the Experimental Light Water Reactor of North Korea," Science & Global Security, 1–21 (2025) DOI: 10.1080/08929882.2025.2583613.

[98] S. Joe and F. Y. Kuo, "Constructing Sobol Sequences with Better Two-Dimensional Projections," SIAM Journal on Scientific Computing **30**, 2635–2654 (2008) DOI: 10.1137/070709359.