HAL
open science

Few group cross section modeling by machine learning for nuclear reactor

Esteban Alejandro Szames

To cite this version:

Esteban Alejandro Szames. Few group cross section modeling by machine learning for nuclear reactor. Neural and Evolutionary Computing [cs.NE]. Université Paris-Saclay, 2020. English. NNT : 2020UPASS134 . tel-03117471

HAL Id: tel-03117471

https://theses.hal.science/tel-03117471v1

Submitted on 21 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Few-group cross section modeling by machine learning for nuclear reactors

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 576, particules hadrons énergie et noyau : instrumentation, image, cosmos et simulation (Pheniics)
Spécialité de doctorat: énergie nucléaire
Unité de recherche: Université Paris-Saclay, CEA, Département de Modélisation des Systèmes et Structures, 91191, Gif-sur-Yvette, France.
Référent: : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris, le 3 Juillet 2020, par

Esteban Alejandro Szames

Composition du jury:

Pierre Désesquelles Directeur de recherche, Ecole Doctorale Pheniics, Université Paris-Saclay	Président
Bertrand Bouriquet Ingénieur de recherche, Électricité de France (EDF)	Rapporteur & Examinateur
Paolo Vinai Professeur Associé, Chalmers University of Technology	Rapporteur & Examinateur
Mathilde Mougeot Directrice de recherche, Centre de Mathématiques et de Leurs Applications (ENSIIE)	Examinatrice
Alberto Previti Ingénieur de recherche, Framatome, Department of Computer Codes and Methods Development	Examinateur
Jean-Marc Martinez Directeur de recherche, Université Paris-Saclay	Directeur
Karim Ammar Ingénieur de recherche, CEA, Saclay	Invité
Eduardo Villarino Professeur Associé, INVAP S.E.	Invité
Daniele Tomatis Ingénieur de recherche, CEA, Saclay	Invité

Thèse de doctorat

NNT: 2020UPASS134

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Contents

Abstract 15

1 Introduction 21
1.1 Nuclear Power Plants . . . 21
1.1.1 Nuclear reactors . . . 22
1.2 Interaction of neutrons with matter . . . 24
1.2.1 Nuclear cross sections . . . 24
1.2.2 Neutron transport equation . . . 25
1.2.3 Isotopic depletion . . . 26
1.2.4 Other magnitudes of interest . . . 26
1.3 Core simulations . . . 27
1.3.1 Challenges of core calculations . . . 27
1.3.2 Two-step deterministic calculation scheme . . . 28
1.3.3 Modeling errors of the two-step modeling scheme . . . 30
1.3.4 Lattice calculation implementation . . . 31
1.3.5 Cross section variables . . . 31
1.4 Cross sections representation . . . 32
1.4.1 Modeling requirements . . . 34
1.4.2 Modern trends in core simulations . . . 36
1.5 Modeling challenges and state-of-the-art . . . 37
1.5.1 Final thoughts and discussion . . . 41
1.6 Organization of thesis . . . 42

2 Use case and performance metrics 45
2.1 PWR standard UO_2 fuel assembly . . . 45
2.2 Lattice calculation settings . . . 45
2.2.1 Input space composition . . . 46
2.2.2 Data set discretization . . . 47
2.2.3 Output space composition . . . 48
2.3 Quantities of interest and performance metrics . . . 49
2.3.1 Multiplication factor sensitivity to cross section error . . . 50
2.3.2 Important microscopic cross sections . . . 51
2.4 Analysis of variance . . . 52
2.5 Final thoughts and discussion . . . 55

3 Spline interpolation 57
3.1 Interpolating approximation using univariate basis functions . . . 57
3.2 Global polynomial representation . . . 58
3.2.1 Limitations of global approximations . . . 59
3.3 High order piece-wise polynomial approximations . . . 59
3.3.1 B-splines as basis of a piece-wise polynomial space . . . 61
3.3.2 Interpolant piece-wise approximations with B-splines . . . 62
3.4 Knot vector constructions . . . 62

3.5 Spline cross section approximation . . . 64
3.6 Spline error analysis for $d = 1$ . . . 64
3.6.1 Non-monotonically error behavior high order splines . . . 65
3.6.2 The drive of spline oscillations . . . 67
3.6.3 Theoretical investigation of the spline error . . . 68
3.6.4 Multi-dimensional spline approximation in a Cartesian grid . . . 69
3.7 Spline approximation in higher dimensional space . . . 69
3.8 Final thoughts and discussion . . . 70

4 Kernel methods and active learning 73
4.1 Machine learning and kernel methods . . . 73
4.2 Introduction to Reproducing Kernel Hilbert Spaces (RKHS) . . . 73
4.2.1 Approximation by kernel methods . . . 74
4.2.2 Advantages and drawbacks of the methodology . . . 76
4.3 Local function spaces reproduced by the kernel . . . 77
4.3.1 The Bernoulli kernel . . . 77
4.3.2 The spline kernel . . . 77
4.4 Results . . . 78
4.4.1 Comparison of Spline and Bernoulli kernels . . . 78
4.5 Supervised Learning . . . 79
4.5.1 Active learning for cross section modeling . . . 80
4.5.2 Computational cost of Active Learning . . . 81
4.6 Kernel methods using active learning in a shared support scheme . . . 82
4.7 Cross section clustering in active learning . . . 86
4.7.1 Independent cross section pool active learning . . . 86
4.8 Higher order RKHS approximation . . . 91
4.8.1 Regularization and preconditioning . . . 91
4.8.2 Higher order spline approximation with RKHS and active learning . . . 91
4.9 Conditioning of the Gram matrix . . . 94
4.9.1 Kernel method's coefficient distribution and shrinking . . . 94
4.10 Investigation on evaluation speed . . . 94
4.10.1 Evaluation speed of independent cross section models . . . 95
4.10.2 Evaluation speed of cross section models with common evaluation vector 96
4.10.3 Active learning computing time . . . 97
4.11 Final thoughts and discussion . . . 97

5 Artificial Neural Networks 101
5.1 Machine learning . . . 101
5.2 Artificial neural networks in nuclear engineering . . . 102
5.2.1 Feed-forward fully connected artificial neural networks . . . 103
5.3 Artificial neutral networks sensitivity studies . . . 104
5.3.1 Cross section subset . . . 104
5.3.2 Implementation setup . . . 105
5.3.3 ANN architecture . . . 108
5.3.4 Qualitative analysis on the function space generated by artificial neural networks . . . 111
5.4 Industry data set . . . 112
5.5 Multi-output network . . . 116
5.5.1 Parametric analysis in the number of neurons for a two-layered network . 116
5.5.2 Error analysis for a two-layer, 20 neurons multi-output artificial neural network . . . 116
5.5.3 Weight evolution with training . . . 121
5.6 Performance remarks . . . 121

5.7 Final thoughts and discussion . . . 123

6 Conclusion and Perspectives 127
6.1 Main conclusion . . . 127
6.2 Additional contributions of the thesis . . . 130
6.2.1 Spline interpolation . . . 130
6.2.2 Kernel methods . . . 131
6.2.3 Artificial neural networks . . . 133
6.2.4 History Effects . . . 134
6.2.5 Developed packages . . . 134
6.3 Perspectives . . . 134
6.3.1 Use case . . . 134
6.3.2 Spline interpolation . . . 135
6.3.3 Kernel methods . . . 135
6.3.4 Artificial neural networks . . . 136

A Spline interpolation using B-splines 149
A.1 Piece-wise polynomial representation with B-splines theoretical discussion . . . 149
A.2 Difficult cross section for Spline approximation . . . 150

B Kernel Methods 153
B.1 Other kernels . . . 153
B.2 High order Bernoulli kernel methods . . . 153
B.2.1 Kernel method's coefficient distribution and shrinking . . . 154

C History effect modeling 159
C.1 History effects phenomena . . . 159
C.2 History parameters . . . 161
C.3 Use cases for numerical tests . . . 163
C.4 Implementation . . . 164
C.5 Results . . . 168
C.5.1 Analysis of the microscopic cross sections . . . 168
C.5.2 Analysis of the macroscopic cross sections . . . 172
C.6 Concentrations of important isotopes . . . 176
C.6.1 Analysis of the infinite multiplication factor . . . 177
C.7 Conclusion . . . 179

List of Figures

1.1 Nuclear power plants (NPP) in the world. 22
1.2 Nuclear reactor power generation. 23
1.3 Typical fuel assembly of a PWR reactor. 23
1.4 Nuclear fission cross section of ^{235}U. 25
1.5 The multi-scale structure involved in core calculations. 28
1.6 Two-step calculation scheme. 30
1.7 Nominal calculations where other history parameters may be present defining one or several depletion histories. 32
1.8 Cross section in function of burnup and the fuel temperature for ^{239}Pu, Σ_{res}, ^{238}U. 34
1.9 Most important few-group cross section modeling requirements. 36
1.10 Chronological literature review of few-group cross sections modeling. 37

2.1 Layout of 1/4 UO_2 PWR fuel element used in this thesis. 46
2.2 Density of pool and test points for the use case 48
2.3 Evolution of cross section importance with burnup 52
2.4 Zero mean cross section projections for the isotope ^{235}U using the entire data set ($d = 3$) discriminated by reaction and group 53
2.5 Zero mean cross section projections for Σ_{res} in using entire data set ($d = 3$) discriminated by reaction and group. 54
2.6 Cross sections variation with burnup discriminated by isotope, reaction and group. 55

3.1 Divergent global polynomial approximation in cross section modeling. 60
3.2 Properties B-splines basis. 62
3.3 Error profiles for different spline approximation for $d = 1$ 65
3.4 RE_σ for the oddend spline with the support $\mathcal{X}_\star$ and $d = 1$. 66
3.5 High order spline approximation for difficult cross section 67
3.6 Modified high order spline approximation. 68
3.7 AV_σ for higher order modified splines using a linear function space for the fuel temperature and boron concentration. 70

4.1 Average microscopic cross section error using piece-wise approximations with different active learning strategies in a shared support scheme. 84
4.2 Relative error histogram over the test set $\mathcal{T}_{flat}$ for every cross section using the support of interest using a linear approximation 85
4.3 Loss function values discriminated by cross sections 87
4.4 Errors for linear piece-wise approximations with independent active learning per cross section (no shared support). 89
4.5 Library composition at maximal size for the two for active learning without a shared support condition. 90
4.6 Effect of regularization in microscopic cross section approximation. 92
4.7 Error histogram using different regularization values. 92
4.8 RKHS approximation using a high order spline kernel with preconditioning and a regularization. 93

5.1 Scheme of a single-layer artificial neural network. 104
5.2 Cross section relative error (RE_σ) in training using different learning rates for a single-layer network with 20 neurons. 107
5.3 Non linear activation function for artificial neural networks. 108
5.4 Average relative error in $\Sigma_{res,a,1}$ with training for single-layer networks with a varying number of neurons. 109
5.5 RE_σ for $\sigma_{235,f,2}$ in function of training for different typologies at constant network size. 111
5.6 Shallow network of 8 neurons approximating $\sigma_{235,f,2}$. 112
5.7 Absolute error E_σ for important cross section during training using the industry set and the $\mathcal{T}_{Xe}$ for test. 114
5.8 Error histogram for $N = 13/13$ discriminated by cross section. 114
5.9 Multi-output network performance. 117
5.10 Modeling of $\sigma_{235,f,2}$ with a $n = 20/20$ multi-output network 118
5.11 Modeling of $\Sigma_{res,a,1}$ with a $n = 20/20$ multi-output network 119
5.12 Modeling of $\sigma_{239,f,2}$ and $\sigma_{240,f,1}$ with a multi-output $N = 20/20$ network 120
5.13 Network parameter evolution with training. 122

A.1 Spline approximation for difficult cross section for $d = 1$. Oscillation are reduced with a support of 119 points per cross sections. 150
A.2 Standard and modified higher order spline approximation using the using $\mathcal{T}_{Xe}$ test. 152

B.1 Cross section error with other kernels using and hyper-parameter full grid search. 155
B.2 Cross section average relative error for the Bernoulli kernel of order 1, 2 and 3. . 156
B.3 Conditioning number and corresponding coefficients module with varying regularization. 157

C.1 Ratio of the history parameters P, SH, SI for the different use cases over the nominal case. 165
C.2 Scheme of the history parameterization. 166
C.3 History coefficients of the most important isotopes. 167
C.4 History coefficients of $\sigma_{a,1}$ of ^{240}Pu for the history parameters. 168
C.5 Microscopic cross section $\nu\sigma_{f,2}$ of ^{235}U along the burnup with different use cases. 169
C.6 Burnup-dependent average errors of all the microscopic cross sections for all history parameters. 170
C.7 Burnup-averaged error with all the history parameters and for the given use cases. 171
C.8 Distribution of microscopic cross section error (bins of 0.25%). 172
C.9 $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ error for $\mathcal{B}_{c1}$. 173
C.10 $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ in $\mathcal{B}_{c2}$. 174
C.11 Cumulative error of the macroscopic cross sections in the fast group at 22.5 GWd/t (half of the 2-nd cycle). 175
C.12 Ratio of the corrected and uncorrected microscopic absorption cross sections in the fast group; the errors are burnup-averaged in the second cycle of $\mathcal{B}_{c2}$. 175
C.13 Concentration of important isotopes. 176
C.14 (a) $\sigma_{a,1}$ of ^{240}Pu in the 2-nd cycle of $\mathcal{B}_{c2}$. Concentrations of important isotopes for the reference cases $\mathcal{B}_{c2}$, $\mathcal{B}_N$, $\mathcal{B}_I$, $\mathcal{B}_O$. (b) Net production rate for ^{240}Pu when inserting control rods in $\mathcal{B}_N$,$\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$. 177
C.15 ϵ_{k_∞} with the different correction methods, with the neutron reactivity of all use cases (top left). 178
C.16 Histograms of ϵ_{k_∞} and ϵ_{SI} (uncorrected profile in dotted line); bins of 10 pcm and 0.025% respectively. 180

List of Tables

1.1 Library size with increasing input and output space 35

2.1 Constitution and sampling of the input space. 47
2.2 Variance of some important cross sections 53

3.1 Approximations for $d = 1$ with different spline types differing in the order and knot vector construction strategy. 64
3.2 69

4.1 Approximation's error using the Bernoulli and Spline kernel 79
4.2 Loss functions ($\mathcal{L}$) acting on σ, Σ or k_∞. 81
4.3 All approximations consider a linear approximation space with different grid types. 83
4.4 High order spline kernel with preconditioning, regularization and active learning. 94
4.5 Conditioning number of the system $(K + \lambda I)$ 95
4.6 Evaluation speed for RKHS and B-splines utilizing first and second order spaces 96

5.1 Cross section relative error (RE_σ) using different normalizations for a single-layer network with 20 neurons at 1E5 epochs. 106
5.2 Cross section relative error using different activation functions for a single-layer network with 20 neurons at 1E5 epochs. 107
5.3 Cross section relative error using different loss functions for a single-layer network with 20 neurons at 1E5 epochs. 108
5.4 Average relative error with training for single-layer networks with a varying number of neurons at 1E5 epochs. 8 neurons is the smallest number for which the error are within the target. 109
5.5 Average relative error with training for single-layer networks with a varying number of neurons at 1E5 epochs. 8 neurons is the smallest number for which the error are within the target. 110
5.6 Average relative errors AV_σ using the 80/20 split and the industry set for multilinear interpolation and artificial neural networks. 115
5.7 Training and evaluation times for network that model a single cross section. . . 123

6.1 Library size reduction, accuracy and evaluation speed of the three studied methodologies of the thesis: Splines using B-splines, kernel methods (RKHS) with active learning, and artificial neural networks. 129

A.1 Modified Not-a-Knot splines with different smoothness conditions for the first burnup values. 151
A.2 Modified Splop splines with different smoothness conditions for the first burnup values. 151

B.1 Other kernels analyzed. 154

C.1 Depletion histories considered to reproduce the history effects and to test the methodologies proposed in the literature. 164

C.2 Material composition of the AISI316L control rods. 164
C.3 Description of the history effect parameterization. 165
C.4 Standard deviation, mean and maximum error of $\epsilon_{\sigma,\mathcal{B},\theta}(Bu)$ in %. 172
C.5 Standard deviation (SD), mean and maximum absolute error (MAX) of ϵ_{k_∞} in pcm and of ϵ_{SI} in %. 179

Acronyms

ANN Artificial Neural Network.

BWR Boiling Water Reactor.

CPU Central Processing Unit.

CR Control Rod.

EPR European Pressurized Reactor.

F Set of all approximations.

FA Fuel Assembly.

FP Fission Product.

GET Generalized Equivalence Theory.

GPU Graphic Processing Unit.

HFP Hot Full Power.

HZP Hot Zero Power.

LWR Light Water Reactor.

MH Moderator History.

MOC Method of Characteristics.

MSE Mean Square Error.

MTR Material Testing Reactor.

NPP Nuclear Power Plant.

P Plutonium history parameter.

PWR Pressurized Water Reactor.

RAM Random Access Memory.

RKHS Reproducing Kernel Hilbert Spaces.

SH Spectral History.

SI Spectral Index.

SMR Small Modular Reactor.

TPU Tensor Processing Unit.

UO_2 Uranium oxide ceramic.

Glossary

A Nucleus of atom.

Bu Burnup, power release per unit of initial heavy metal. Unit Giga-Watts-day per tonne of heavy metal GWd/t.

C_B Boron concentration.

D Diffusion coefficient.

E Energy of neutron. Unit in electronvolt eV.

I_σ Cross section importance.

L Number of layers.

L_x Evaluation functional.

T Temperature of the medium.

T_f Fuel temperature averaged over the fuel rod. Unit in degree Celsius °C.

$\$_{K,\vec{t}}$ Linear space of splines of order K and knot vector $\vec{t}$.

$\Pi_{K,\vec{\xi}}$ Linear space of piece-wise polynomials of order K and breaks vector $\vec{\xi}$.

Π_K Linear space of polynomials of order K.

Σ Macroscopic cross section. Unit in $1/cm$.

γ Gamma ray.

$\mathcal{B}$ Depletion calculation.

$\mathcal{G}$ Set of energy groups.

$\mathcal{H}$ Hilbert Space.

$\mathcal{H}_k$ Reproducing kernel Hilbert Space (RKHS).

$\mathcal{I}$ Set of specialized isotopes.

$\mathcal{J}$ Penalty function.

$\mathcal{L}$ Loss function.

$\mathcal{R}$ Set of reaction types.

$\mathcal{S}_\sigma$ Support of a cross section model.

$\mathcal{T}$ Test set.

$\mathcal{U}$ Domain of cross section.

$\mathcal{V}$ Set of homogenized region.

$\mathcal{X}$ Normalized domain of cross section.

$\mathcal{X}_\mathcal{P}$ Sampling of the total pool of data.

$\mathcal{Y}$ Set of cross sections within $\mathcal{I}$, $\mathcal{R}$, $\mathcal{G}$, $\mathcal{V}$.

ν Amount of neutrons emitted per fission.

ϕ Scalar neutron flux.

σ Microscopic cross section usually noted as $\sigma_{i,r,g}$ being i the isotope, r the reaction type and g the energy group. Unit in barns $1b$ =1E-24 cm^2.

$\vec{\iota}$ Instantaneous state-variables.

$\vec{\theta}$ History state-variables.

$\vec{r}$ Vector position in space.

ξ Microscopic reaction rate.

d Dimension of a cross section's domain.

k Kernel function.

k_∞ Infinite multiplication factor. Unit in in pcm =1E-6.

n Number of neurons.

t Time. Unit in seconds s.

v Velocity of neutron.

Synthèse en français: Dans les réacteurs nucléaires, une réaction de fission en chaine est induite par les neutrons et est contrôlée pour générer de l'énergie thermique utilisée pour la production d'électricité. Le type de réacteur le plus commun au monde est le REP, Réacteur à Eau Pressurisée qui constitue actuellement la majorité du parc des réacteurs nucléaires en France. Ces réacteurs sont caractérisés par l'utilisation d'eau pour la réfrigération du cœur et la modération des neutrons, ainsi que par l'utilisation d'uranium enrichi placé dans des crayons de combustibles eux-mêmes disposés dans des assemblages "rectangulaires" de 17 par 17 crayons. L'énergie produite dans le cœur est transportée (via le réfrigérant, qui reste à l'état liquide) vers les générateurs de vapeur. L'énergie électrique est alors produite par un turbine alimentée par la vapeur produite par les générateurs de vapeur.

Pour analyser l'état du réacteur en condition d'opérations normales et accidentelles, l'équation du transport de Boltzmann, qui régule les changements dans la population de neutrons, doit être résolue. Or, il s'agit d'un défi important, car c'est une équation intégro-différentielle en 7 variables devant être résolues pour des milieux complexes de grande dimension comme les cœurs des les réacteurs nucléaires. A cette complexité géométrique vient s'ajouter la complexité due aux phénomènes multi-échelles des énergies des neutrons et aux aspects multi-physique dus aux échanges avec les matériaux combustible et le fluide modérateur (phénomènes de thermo-hydraulique, thermomécanique et neutronique sont fortement interdépendants). Les sections efficaces, qui modélisent l'interaction entre les neutrons et les matières, sont les paramètres clés de l'équation du transport. Leurs valeurs varient en fonction des paramètres thermo-hydrauliques du réacteur. Pour estimer la répartition de la puissance au sein d'un réacteur nucléaire, il est donc nécessaire de modéliser l'interdépendance entre les phénomènes neutroniques et thermo-hydrauliques de manière pratique et en restant modeste sur l'utilisation des moyens numériques.

Une stratégie est possible à travers un schéma de calcul déterministe, qui discrétise l'équation du transport en espace, en énergie et en angle. La résolution des équations du système se fait en deux deux étapes distinctes : le calcul réseau, réalisé sur un motif de petite dimension (typiquement un assemblage), où les données de sections efficaces sont tabulées en fonction des paramètres de fonctionnement du réacteur puis condensées en énergie et homogénéisées en espace. Puis, dans la seconde étape, le calcul cœur, où ces librairies de sections sont utilisées pour la simulation du système entier, avec une discrétisation beaucoup plus grossière qui permet des calculs plus rapides. Si cela est exécuté d'une manière correcte, un réacteur de puissance modélisé ainsi permet des simulations précises avec un coût numérique modéré.

Les librairies des sections efficaces homogénéisées à peu de groupes d'énergies sont le lien entre ces deux étapes de modélisation, mais aussi une des sources principales d'erreurs dues aux approximations. Ces sections évoluent de manière relativement régulière lors des changements d'état du réacteur, avec une dépendance d'ordre bas dans les variables principales (comme la température et le taux de combustion du combustible, la température du modérateur...) et sans bruit numérique. En revanche, de forts changements localisés peuvent être observés dans certaines régions du domaine.

Cette thèse est consacrée à la modélisation des sections efficaces par de nouvelles techniques innovantes et performantes, mais suffisamment matures pour être appliquées aux problèmes industriels. La performance d'un modèle est principalement définie par le

nombre de coefficients qui le caractérisent (c'est-à-dire l'espace mémoire nécessaire pour le stocker), la vitesse d'évaluation, la précision, la robustesse par rapport au bruit numérique, la complexité, etc. Dans les anciens travaux de modélisation, seules des sections efficaces macroscopiques étaient modelées à travers des facteurs de correction construits d'une manière empirique, basée sur "l'avis d'experts". La plupart de ces exemples peuvent être regroupés sous la dénomination d'expansions de Taylor définies par "essai et erreur".

Au début des années 2000, la modélisation est passée à des simulations neutroniques-thermo-hydrauliques tridimensionnelles et entièrement couplées. L'interpolation multilinéaire dans une grille complète a été couramment utilisée pour reconstruire les sections efficaces, en utilisant des bibliothèques de très grande taille, afin de répondre aux exigences de précision. Mais un nombre de points à croissance exponentielle est nécessaire pour interpoler les sections lorsque l'on augmente la taille du modèle. Cela est appelé la "malédiction de la dimensionnalité" (curse of dimensionality). Des approximations d'ordre supérieur ou des projections dans des sous-bibliothèques spécialisées sont utilisées pour essayer de réduire l'occupation mémoire.

Pour diminuer la taille de la libraire de sections efficaces, des méthodes de régression ont été explorées. La principale difficulté de ces méthodologies est le coût des calculs nécessaires pour la résolution de la quadrature intégrale qui contient les coefficients de l'approximation. Jusqu'à un million de points de calcul sont requis pour traiter un problème de faible dimension.

La dépendance des sections efficaces aux variables d'état peut être considérée comme la somme de fonctions de dimension relativement faible, elles présentent donc une "dimensionnalité pratique réduite". En conséquence des grilles "creuses" (sparse , c'est-à-dire non complètes, mais régulières) ont été explorées pour l'approximation des fonctions multidimensionnelles. Ceci entraîne une réduction très significative de la taille de la bibliothèque de sections efficaces, cela même avec un espace d'approximation au premier ordre. Des espaces d'approximation d'ordre élevé augmentent le taux de convergence, mais nécessitent des polynômes globaux et des discrétisations très particulières (nœuds de Tchebychev) pour traiter le phénomène de Runge-Kutta

Les objectifs de cette thèse sont d'analyser la modélisation des sections efficaces en utilisant différentes méthodes : les Splines avec des B-splines, les méthodes à noyaux combinées avec de l'apprentissage actif et des réseaux de neurones artificiels à sortie unique ou multidiomensionelle. Ces modèles sont évalués en vue de la précision requise, d'une vitesse d'évaluation compatible avec les standards industriels et d'une réduction des besoins de stockage de la bibliothèque. Les splines représentent l'état de l'art dans de nombreux codes industriels, et sont une approximation locale d'ordre supérieur. Nous proposons les méthodes à noyau du type RKHS (Reproducing Kernel Hilbert Space) qui permettent d'envisager des supports très généraux. Cette approche combinée à des techniques d'apprentissage actif permet de choisir un support optimal de point pour l'approximation. Cela a pour effet de diminuer la taille mémoire occupée pour l'approximation des sections efficaces. L'approche RKHS permet une grande variété d'espaces de fonctions qui peuvent reproduites avec un coût d'évaluation relativement réduit. Enfin, les réseaux de neurones artificiels reconnus pour leurs capacités d'approximation de fonctions complexes peuvent apprendre des relations directement à partir des données, ce qui permet de traiter la totalité des sections efficaces avec un seul modèle. Ces modèles sont d'autant plus perfor-

mants aujourd'hui car ils tirent profit de la performance des outils et des logiciels existant aujourd'hui.

Dans cette thèse, nous nous sommes limités à l'étude d'un assemblage REP standard de combustible UOX et nous avons analysé la dépendance des sections efficaces en fonction de trois variables d'état : le burnup (traduisant l'usure du combustible), la température du combustible et la concentration de bore. La taille de stockage des bibliothèques est optimisée en cherchant à maximiser la vitesse et la précision de l'évaluation, tout en cherchant à réduire l'erreur de reconstruction des sections efficaces microscopiques, macroscopiques et le facteur de multiplication infini.

La dépendance des sections efficaces peut être divisée entre les variables d'état qui varient de manière instantanée dans le cœur, et les variables dites "historiques" qui varient lentement avec l'exposition du combustible nucléaire au flux neutronique. Même si la plupart des modélisations ont été menées pour les variables instantanées (qui vient du calcul du type "branch" dans le code de calcul réseau) une étude de modélisation de variables historiques a été aussi effectuée pour les variables principales d'un REP. Différentes méthodologies de paramétrisation au premier ordre ont été réalisées, notamment en considérant la concentration de plutonium et des variables spectrales comme le spectral index ou l'histoire spectrale. Des améliorations significatives ont été trouvées. En revanche, pour des cas où la barre de contrôle change de position pendant l'évolution, des erreurs significatives de modélisation ont été décelées.

SPLINE : La plupart des méthodologies développées le sont pour les variables de type instantanées. Les stratégies que nous proposons sont comparées entre elles, et à l'interpolation multilinéaire sur une grille cartésienne qui est la méthode utilisée usuellement dans l'industrie. L'interpolation par splines représente la méthode "state-of-the-art" dans l'industrie. Elle utilise une base de B-splines qui génère des approximations locales d'ordre élevé. Une grille cartésienne est utilisée comme support, ce qui permet l'utilisation de routines performante numériquement. Un traitement formel de la construction d'espaces d'approximation locaux avec les B-splines permet d'analyser les différentes conditions de bords avec les Splines. De faibles améliorations sont obtenues en relaxant les conditions des dérivées sur les bords du domaine, ce qui en aide à la résolution des problèmes d'oscillations. Néanmoins lorsque l'on augmente le nombre de dimensions, le bénéfice d'utiliser un espace d'approximation d'ordre élevé est réduit par l'utilisation d'une grille cartésienne

RKHS : La deuxième méthode appartient au domaine du machine learning, et en particulière les "kernel machines". Ces méthodes à noyaux utilisent le cadre général de l'apprentissage machine et sont capables de proposer, dans un espace vectoriel normalisé, une grande variété de modèles de régression ou de classification. Les méthodes à noyaux peuvent reproduire différents espaces de fonctions en utilisant un support non structuré. Support optimisé avec des techniques d'apprentissage actif. Les approximations sont trouvées grâce à un processus d'optimisation convexe facilité par "l'astuce du noyau" (kernel trick). Le caractère modulaire intrinsèque de la méthode facilite la séparation des phases de modélisation : sélection de l'espace de fonctions, application de routines numériques, et optimisation du support par apprentissage actif. Différentes critères représentés par des fonction de perte (loss function) ont été étudiés, sur les sections efficaces microscopiques, macroscopiques ainsi que sur le facteur multiplicatif avec différentes métriques.

Le meilleur compromis entre précision et performance numérique a été trouvé en privilégiant les sections efficaces qui contribuent le plus à la section macroscopique. L'analyse de l'évolution de la fonction de perte a permis d'observer l'inclusion de points d'un bas intérêt en la plupart de sections efficaces lorsque la condition de support partagé est imposée. En conséquence, une réduction en la taille de la librairie d'un maximum d'un ordre de grandeur a été obtenue en faisant la construction du domaine de manière indépendant pour chaque section efficace.

Nous avons utilisé des RKHS permettant de générer des fonctions dans des espaces de Sobolev (fonctions continues à dérivées bornées en norme L2). Cela permet de reproduire des approximations semblables à celles des splines sans contraintes sur le placement des points supports. Cette propriété permet d'exploiter des grilles non cartésiennes ce qui réduit considérablement le nombre de points nécessaires à l'approximation des sections efficaces. L'approche par RKHS exige la résolution d'un système linéaire dense dans notre cas. Des problèmes de conditionnement numériques sont apparus. Nous avons dû les résoudre par des techniques de régularisation. Pour l'ensemble de ces méthodes, différentes améliorations numériques (notamment le développent de modules en fortran compilé et importé dans Python à travers des packages comme F2PY) ont permis l'obtention de temps de calcul compatibles avec les standards industriels.

RÉSEAUX DE NEURONES : Les réseaux de neurones sont des méthodes d'approximation universelles capables d'approcher de façon arbitraire des fonctions continues sans formuler de relations explicites entre les variables. Ces « model-free mapping systems » sont capables de considérer des espaces d'entrées et de sorties de très haute dimension, sans le besoin d'imposer des lois données par l'utilisateur et sans un surcoût computationnel excessif. Ils utilisent les données en forme de « batch » ou en « streaming » sans changement significatif de leur structure et sont capables d'exploiter les architectures modernes de type GPU. Une fois formés avec des paramètres d'apprentissage adéquats, les réseaux de neurones à sorties multiples (intrinsèquement parallélisables) réduisent au minimum les besoins de stockage tout en offrant une vitesse d'évaluation élevée.

Des analyses de sensibilité ont été conduites pour les différentes options de normalisation, les paramètres d'apprentissage et les fonctions d'activation. Différentes topologies de réseaux de neurones ont été étudiées en trouvant des systèmes qui minimisent le stockage de chaque section efficace (en utilisant une couche cachée à 8 neurones) ou des systèmes à 2couches cachées qui maximisent le taux de convergence pendant l'étape d'apprentissage.

Ces systèmes ont été testés avec différents jeux de données pour reproduire les cas caractéristiques trouvées dans les cas industriels. Une réduction très significative de la taille des librairies de sections efficaces a été ainsi obtenue. Les temps d'entrainement sont acceptables spécialement lorsque l'on considère les réseaux de neurones à double couche qui maximisent le taux de convergence.Néanmoins, et surtout pour les jeux de données plus petits, des phénomènes de sur-apprentissage ont été observés.

Les caractéristiques de ces systèmes d'approximation permettent une extension très facile à un schéma de prédiction à sorties multiples. À travers des analyses de sensibilités sur des réseaux de neurones multicouches, qui modélisent la totalité des sections efficaces dans la librairie, un modèle unique a été obtenu. Cela permet une réduction de la taille de la librairie de 98% et une réduction du le temps d'obtention du modèle proportionnelle à la

quantité totale de sectionnes efficaces modelés. Les temps d'évaluation de ces modèles sont compatibles avec les standards de l'industrie notamment lorsque l'accélération GPU est disponible. L'ensemble des données, des outils, et des scripts développés sont disponibles librement sous licence MIT.

Conclusion : La conclusion centrale de la thèse est que les méthodes à noyaux, qui utilisent des espaces d'approximation locaux et régularisés, sont bien adaptées à la reconstruction des sections efficaces. Avec une sélection des points supports optimisés par apprentissage actif, la méthode est performante, modulaire et relativement facile à implémenter. Les réseaux de neurones sont potentiellement plus performants, notamment en considèrent les réseaux multi sorties. Ils peuvent exploiter les représentations (patterns) communes dans les données et ainsi obtenir une compression significative de la mémoire occupée et une vitesse d'évaluation maximale. En revanche, cette méthode manque de méthodologies standardisées pour l'obtention de ses hyper paramètres et n'est pas encore assez mature pour être industrialisée.

Chapter 1

Introduction

In this chapter a brief introduction to nuclear power plants, their governing equations and the computational tool used for simulating their behaviour are provided. Core calculations using the two-step modeling scheme are presented to identify the challenges posed by modern industry calculations. Few group cross sections, that represent the interaction of neutrons with matter, are the key linking elements between the two phases of this modeling approach. Their characteristics, requirements and implementation constrains are presented in the context of modern core calculations. Different approximation procedures with their advantages and drawbacks are extracted from a thorough literature review guiding the main objectives of this thesis.

1.1 Nuclear Power Plants

Commercial Nuclear Power Plants (NPP) produce electricity by generating thermal energy from nuclear fission providing access to carbon-free, non intermittent energy. There is a total of 450 NPP currently in operation, with an installed power of 396 GW(e) accounting for about 10 % of total electricity production [1]. An additional of 55 units with a total capacity of 57 GW(e) are under construction.

Leading countries in nuclear energy production are the United States, France, Russia, and China. The United States is the world's largest producer of nuclear power accounting for about 30% of the worldwide nuclear generation of electricity. France has 58 NPPs with a nuclear share in electricity production of 71 %, the highest in the world. In China 10 NPP are under construction and nuclear energy share is expected to increase from 4% to 10% by 2030 [1].

An expected growth in the world-wide electricity demand of 2.1% per year pushes for an increase in the nuclear share of up to 5% by the year 2050 [1]; a doubling of the current installed power. Additionally, half of the nuclear fleet is over 30 years old, and even when accounting for lifespan extensions, a significant demand for new reactors is in sight. In Fig. 1.1 current planning of new NPP is shown.

Asides from power generation, other application of nuclear technology includes medical treatment, radioisotopes production, materials research and development, naval propulsion, desalination, space applications, etc.

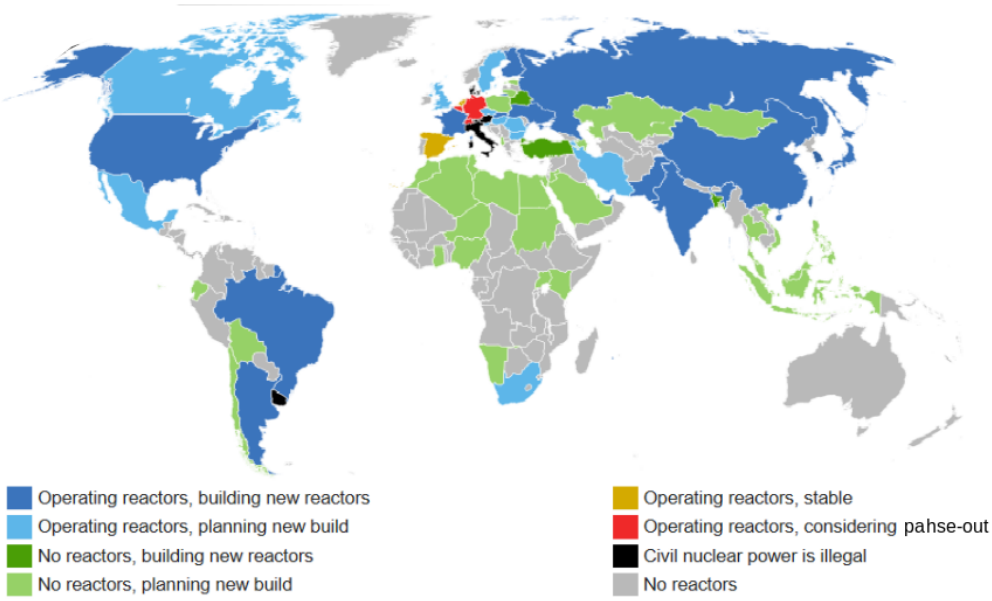


Figure 1.1: Nuclear power plants (NPP) in the world [1]. Many countries count with NPP and several more are projected or underway.

1.1.1 Nuclear reactors

In nuclear reactors, a self-sustained and controlled fission chain reaction generates thermal energy used to produce electricity. Neutrons induce fission reactions in atoms of uranium or plutonium which, when split, release large amounts of energy and high-speed neutrons. They lose energy by colliding with nuclei present in the core in a process called moderation, arriving at energies of a few eV which favors fission reactions. Light Water Reactors (LWR) are a type of nuclear reactor that uses water (H_2O)[1] as both coolant and neutron moderator. With light nucleus such as the Hydrogen neutrons can lose their entire kinetic energy in a single interaction.

The most common LWR is the pressurized water reactor (PWR) which constitutes the majority of the world's NPP. As the name implies, a heavy pressure vessel and piping is used to reach about 155 bars maintaining the water in the primary circuit in a liquid state at about 300 °C. The coolant is pumped at high pressure through the core carrying the heated water to the steam generator. In the secondary circuit, the steam flows to the turbines that spin an electric generator producing electricity. A scheme of this process is shown in Fig. 1.2. To control the rate at which the neutron population changes i.e., regulate the power output, neutron absorbers such as retractile control rods made of boron, cadmium or indium which are potent neutron absorbers, are use at the core. Boron diluted in the primary circuit is also used. The fissile material is in the form of fuel

[1] As opposed to heavy water consisting mainly of D_2O.

pellets composed of a uranium oxide ceramic (UO2) enriched up to 3% in ^{235}U or MOX fuels[2]. These are inserted into Zircaloy tubes of about 1 cm in diameter and 4 meters long that are bundled together forming a Fuel Assembly (FA) typically in a 14×14 or 17×17 arrangement as shown in Fig. 1.3.

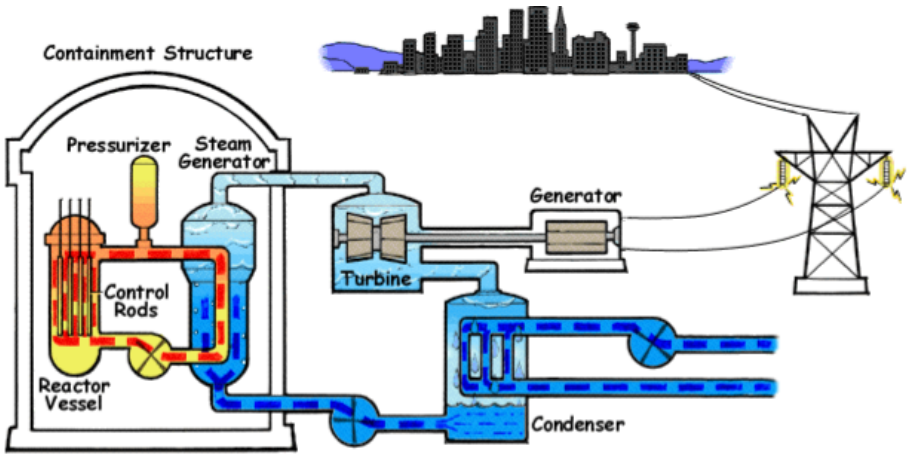


Figure 1.2: Nuclear reactor power generation. In the primary circuit the pressurizer is used to control the pressure and the control rods to control the neutron reaction. Heat produced in the core is transported through the primary circuit towards the steam generator. In the secondary circuit the produced steam is used to spin the generator and produce electricity. Image from [2].

In NPP the mechanical integrity of the fuel assembly must be assured for extreme conditions of very high pressure and temperature under irradiation, where complex processes such as pellet-cladding interaction and fuel swelling take place. The fuel temperature depends on the thermo-hydraulic regime throughout the core. A reflector composed of steel and water usually surrounds the core to reduce the amount of neutron leaving the system and reaching the vessel.

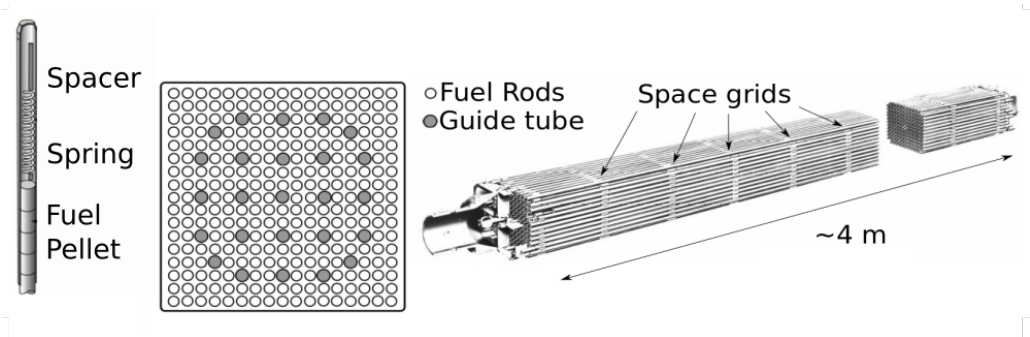


Figure 1.3: Typical fuel assembly of a PWR reactor. The square 17×17 arrangement can be noticed together with the upper nozzle and the space grids to reduce the vibrations caused by the coolant. The cladding is made of Zircaloy, a special alloy transparent to neutrons. The fuel pellets are composed of an uranium oxide ceramic (UO_2) enriched up to 3% in ^{235}U.

[2]Mixed oxide fuel (MOX) is a blend of plutonium, occasionally sourced from reprocessing, with natural or depleted uranium. MOX fuel is an alternative to UO2 which currently predominates in LWR cores.

1.2 Interaction of neutrons with matter

Neutrons interactions with matter are described by *nuclear cross section.* An intuitive definition is as a proportionality factor in irradiation experiments over a target: it's observed that the surfacic reaction rate dR_r (number of nuclear reactions of type r per unit of time and unit surface) is proportional to the number density of nuclei in the target C, the intensity of the beam I and the width of the target ds. Equality is obtained by the introduction of nuclear cross section (σ_r)

$$dR_r = \sigma_r C I ds. \tag{1.1}$$

This relation dictates that σ_r has units of cm^2 usually measured in barns $1b$ =1E-24 cm^2 [3]. Nuclear cross section data is tabulated in the form $\sigma_{r,i}(E,T)$ for the reaction r of isotope i in function of energy E and the temperature T of the medium[3]. These, sometimes called point-wise evaluated cross sections, are obtained from experiments coupled with mathematical models for different reaction types: absorption, fission or scattering.

1.2.1 Nuclear cross sections

Nuclear cross sections result from nuclear interactions between an incident neutron (n) on a target nucleus A. A complete definition from fundamental physics can be found in [5].

A *scattering cross section* is in the form $n + A \rightarrow A + n$ for potential and elastic (a compound nucleus is formed) interactions where the neutron may transfer energy to the nucleus A. In inelastic scattering, the target nucleus may additionally lose energy by emitting gamma photons (γ).

Fission cross sections are in the form $n+A \rightarrow FP_1+FP_2+\nu n$ which is an exothermic process in which the target nucleus is split into two *fission products* (FP) liberating $\sim$ 200 MeV. Among other particles (α, β), an amount of ν neutrons are emitted. Fission is responsible for the energy generated in a reactor and for sustaining the chain reaction. The neutron energy threshold for which this reaction takes place differentiates fissile isotopes (^{235}U, ^{239}Pu) from fissionable isotopes (^{238}U, ^{240}Pu).

In a capture interaction of the form $n + A \rightarrow (A + 1) + \gamma$ the neutron is trapped by the nucleus. This type of reaction causes a loss of neutrons in the system but is also essential to control the reaction chain. The *absorption cross section* is the sum of fission and capture cross sections. In general, nuclear cross section are inversely proportional to the neutron's velocity v, following the $1/v$ law as can be seen in Fig 1.4. An intuitive explanation is that slower neutrons stay a longer time in the vicinity of a nucleus thus increasing the probability of a nuclear interaction. Absorption and fission cross sections present resonances when the incident neutrons has a kinetic energy matching the discrete excitation states of the nucleus. As energy increases, resonances start to overlap reaching the "unresolved resonance region".

[3]"medium" refers to isotope, temperature and the chemical, or crystal structure of matter. The latter becomes relevant for low energy neutrons [4].

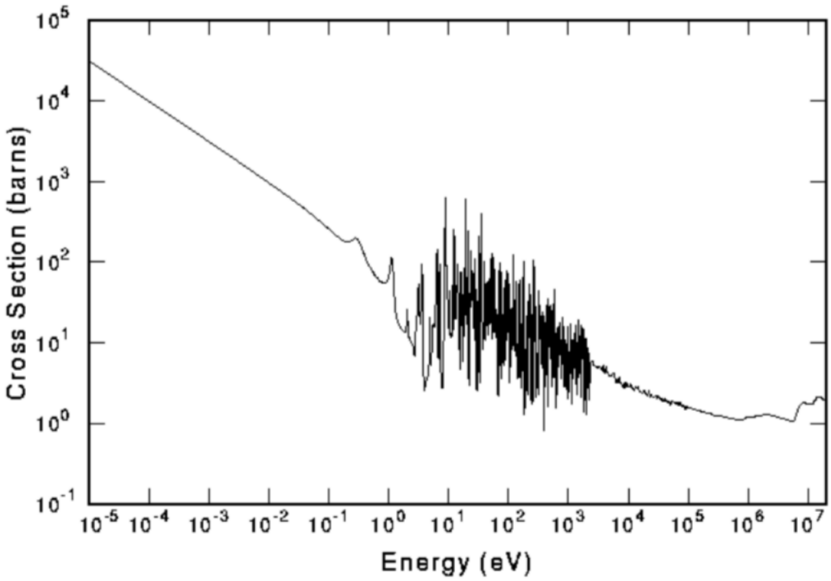


Figure 1.4: Nuclear fission cross section of ^{235}U. Three distinct regions can be observed: of $1/v$ up to 1E-1 eV, the region of resolved resonances whose overlapping results in the unresolved or continuum region after 6E3 eV.

1.2.2 Neutron transport equation

The thermal output in a nuclear power plant is proportional to the neutron population in the core. Nuclear cross sections are used in the *Boltzmann neutron transport* equation to describe the neutron field in a multiplying media in function of space, energy, angle, and time [6]

$$\begin{aligned}\frac{1}{v}\frac{\partial\psi(\vec{r},E,\vec{\Omega},t)}{\partial t} &= -\vec{\Omega}.\vec{\nabla}\psi(\vec{r},E,\vec{\Omega},t) - \sum_k C_k(\vec{r},t)\sigma_k(E)\psi(\vec{r},E,\vec{\Omega},t)\\ &+\sum_k C_k(\vec{r},t)\int_0^\infty dE'\int_{4\pi} d\vec{\Omega}'\sigma_{s,k}(E'\to E,\vec{\Omega}'\to\vec{\Omega})\psi(\vec{r},E',\vec{\Omega}',t)\\ &+\frac{1}{4\pi}\sum_k C_k(\vec{r},t)\int_0^\infty dE'\nu_{p,k}(E')\sigma_{f,k}(E')\chi_{p,k}(E'\to E)\phi(\vec{r},E',t)\\ &+\frac{1}{4\pi}\sum_k \nu_{p,fs,k}\lambda_{fs,k}C_k(\vec{r},t)\chi_{p,fs,k}(E)\\ &+\frac{1}{4\pi}\sum_k \lambda_{d,k}C_k(\vec{r},t)\chi_{d,k}(E) + S_{ext}(\vec{r},E,\vec{\Omega},t).\end{aligned} \tag{1.2}$$

This integro-differentiable equation establishes a neutron balance of 7 terms for the variation of the flux in a phase-space defined by $d\vec{r}d\vec{\Omega}dE$. These are, in order: neutrons traversing the boundaries of the volume near $\vec{r}$, interactions that remove neutrons by absorption or scattering, neutrons entering the phase space due to scattering, fission source of prompt neutrons, fission source due to isotope decay, fission source of delayed neutrons and an external source. More details can be found in [3, 6].

The solution of interest is the (scalar) neutron flux $\phi = \int_{4\pi}\psi d\vec{\Omega}$ used to calculate the

reaction rates that define the power output and the fuel evolution.

1.2.3 Isotopic depletion

The exposure of matter to the neutron flux produces nuclear reactions that modify its composition. This change is called *depletion*, mainly produced by neutron absorption, fission or radioactive decay. For each depleting isotope i, the time-dependent number density $C_i(\vec{r},t)$ is

$$\begin{aligned}\frac{dC_i(\vec{r},t)}{dt} = &\sum_{m\neq i}(\zeta_{i\leftarrow m}(\vec{r},t)+\lambda_{i\leftarrow m})C_m(\vec{r},t)\\ &-(\lambda_i C_i(\vec{r},t)+\zeta_i(\vec{r},t))C_i(\vec{r},t).\end{aligned} \tag{1.3}$$

The change in time of $C_i(\vec{r},t)$ depends on the reaction rates[4] $\zeta(\vec{r},t)$ and the radioactive decay (λ); specifically on the formation of isotope i from other m isotopes due to nuclear interactions and radioactive decay, and its own rate of destruction due to these processes.

This equation governs the concentration of every isotope present in the core that undergoes depletion forming the *Bateman equations*. Some produced isotopes may, in time, decay to another different isotope [5] forming a set of relations called depletion chains. Heavy nuclei present in the fuel or neutron absorbers in the control rods change the most during reactor operation. Different fuel types usually employ different depletion chains which are subject to modeling and simplification. The time scale of this process is slow enough to use a steady-state solution of the neutron transport equation.

1.2.4 Other magnitudes of interest

Macroscopic cross sections describe the probability that the neutron undergoes a specific interaction in a given volume of material accounting for the concentrations (C) of all isotopes (I) therein

$$\Sigma_r(\vec{r},E,t) = \sum_{j=1}^{I}\sigma_j(\vec{r},E)C_j(\vec{r},t). \tag{1.4}$$

In the following section the difficulties posed by Eq. 1.2 and how to deal with them will be presented. One of these steps is defining the multi-group balance equations [6] which can be written as a generalized eigenvalue problem by the introduction of the multiplication factor k_∞[6]

$$\begin{cases}\left(\Sigma_{a,1}+\Sigma_{1\to 2}\right)\phi_1 = \Sigma_{2\to 1}\phi_2 + \left(\nu_1\Sigma_{f,1}\phi_1+\nu_2\Sigma_{f,2}\phi_2\right)/k_\infty,\\ \left(\Sigma_{a,2}+\Sigma_{2\to 1}\right)\phi_2 = \Sigma_{1\to 2}\phi_1.\end{cases} \tag{1.5}$$

[4] the microscopic reaction rate $\zeta_{i\leftarrow m}(\vec{r},t) = \sum_q \int_0^\infty \sigma_{q,k\leftarrow m}(E)\phi(\vec{r},E,T)dE$ is the rate of formation of isotope i from isotopes m for reaction types q. The scalar flux is required for this calculation.

[5] Specifically beta decays give origin to the term λC in Eq. 1.2 which is what actually allows nuclear reactors to be controlled as they have a characteristic time of a few seconds.

[6] For a two-group infinite homogeneous medium. More details can be found in [7].

Thus,

$$k_\infty = \frac{\nu_1 \Sigma_{f,1} \mathrm{SI} + \nu_2 \Sigma_{f,2}}{(\Sigma_{a,1} + \Sigma_{1\to 2})\mathrm{SI} - \Sigma_{2\to 1}}. \tag{1.6}$$

k_∞ defines the rate of change in the neutron population of the core. The complementary SI is defined as

$$\mathrm{SI} = (\Sigma_{a,2} + \Sigma_{2\to 1})/\Sigma_{1\to 2}, \tag{1.7}$$

is the ratio of fast flux to thermal flux which is an important parameter for the reactor state and fuel evolution.

1.3 Core simulations

It is not a trivial task to properly convey the extent of what lays under the term "core simulations". They are used throughout the lifetime of a reactor from design, construction, operation to eventual decommissioning; allowing to simulate the reactor for day-to-day steady-state operation, transients and accidental conditions not leading to core damage (i.e. not leading to any change in the geometry of the core), and severe accidents (where there is core damage) [8]. Core simulations are assessed from a safety, logistic, and operational perspectives, which usually are in contradiction with each other. Typical calculations also include fuel management, power calculation, safety margins, reload operations, core planning, isotope follow-up (over the lifetime of each full assembly), evaluation of every system component, etc.

1.3.1 Challenges of core calculations

Solving the transport equation for assessing the state of the core is a challenging task as NPP are big, heterogeneous, complex, multi-scale systems, typically containing between 150-800 fuel assemblies (PWR) [8] of square lattices forming the reactor core as shown in Fig 1.5. Local perturbations may modify the state of the whole plant while a global perturbation changes the properties of the media in all scales.

The reactor core is also a multi-physics system with a high degree of coupling between the neutron and the thermo-hydraulic phenomena. For example in the thermal Doppler broadening effect, by an increase of the kinetic energy of materials (temperature) the nuclear resonances broaden, increasing the probability of the nuclear reaction. This phenomenon plays a major role in the negative fuel temperature coefficient in LWR. The thermalization processes, in which neutrons lose energy by scattering interactions reaching thermal energies, depends on the temperature and density of water, the moderator.

A direct calculation of neutron flux distribution as expressed in Eq. 1.2 using an ultra-fine discretization in energy and space requires an unattainable computational effort in both processing and memory requirements. Dropping the time dependence, Eq. 1.2 is still a 7 variable integro-differential problem requiring discretizations of mm and eV to model, for example, localized depletion (plutonium skin effect, burnable poison onion effect) and nuclear resonances. The thermo-hydraulic field, tightly coupled with the neutron

phenomena, changes throughout the core in both radial and axial dimensions. Each fuel assembly undergoes a different depletion processes under these conditions. Dedicated computational tools able to cope with this modeling problem have been developed over the last 70 years.

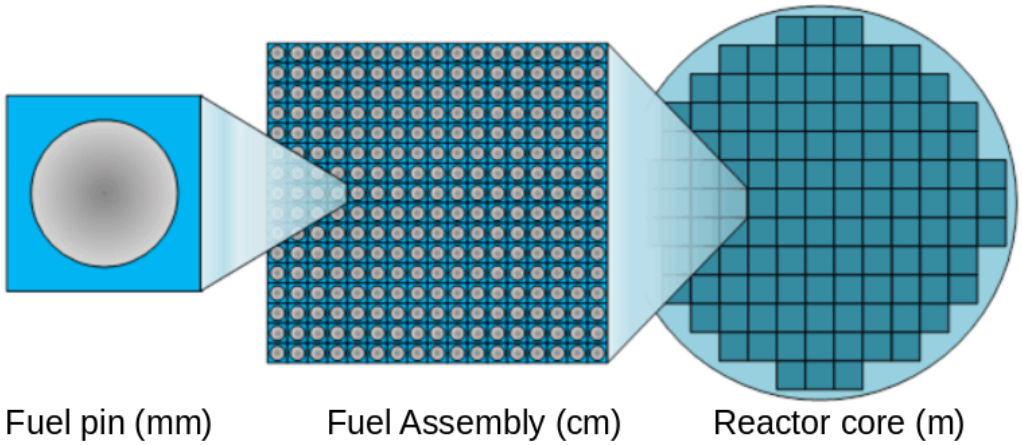


Figure 1.5: The multi-scale structure involved in core calculations. About 200 fuel assemblies typically compose the core of dimensions $\sim 4m \times 4m$. Each fuel assembly is a bundle arranged in a 17×17 lattice. Each fuel pin contains UO_2 pellets. About 50000 fuel rods and 18 million fuel pellets may be present in a PWR core.

One strategy to treat the three-dimensional continuous-energy equation directly, is by *stochastic methods* which perform very few approximations. Many millions of particles histories are simulated based on sequences of random numbers taking into account the neutron interactions, using a fine energy discretization, and an accurate representation of the geometric domain. Though elegant, these methods are slow to converge, especially for big heterogeneous systems such as reactor cores. They are quite suited for certain tasks (biological shielding, dose calculations, reference solutions, very small systems, or with high neutron streaming or otherwise pure particle transport problems) but unpractical for day-to-day operation PWR calculations.

Another strategy is the *deterministic calculation scheme* where numerical analysis techniques are used to estimate the solution of Eq. 1.2 by discretizing the space, energy and angular domains. The core modeling consists of calculations carried out in two distinct steps. If done correctly, NPP are modeled at a relatively modest computational cost, and in a practical way. *Cross section libraries*, analyzed in this thesis, are the key linking element between these *two-step calculations*.

1.3.2 Two-step deterministic calculation scheme

In the deterministic calculation schemes a two-step approach is used were nuclear data are *homogenized in space* and *condensed in energy* by means of *lattice calculations*. In the first step (1) *lattice or transport calculations* are performed with high spatial and energy discretization but only for small representative regions of the core. Resulting homogenized *few group cross sections* feed *core calculation* (2) which are carried out using coarse meshes

but for the entire system. In this way local effects at the fuel assembly level are modeled somewhat independently of global reactor effects, that are treated for the entire core.

(1) Lattice calculations

Corrected nuclear cross sections accounting for depression in the flux due to resonances are obtained from self-shielding calculations[7]. These are used in a highly discretized spatial mesh of the two-dimension fuel lattice to calculate a multi-group scalar flux $\phi(\vec{r}, E)$. It can be obtained using different treatments of the transport equation for single or multiple fuel assemblies as dictated by the core plan. A periodic environment is normally considered by reflective boundary conditions. Cross sections are then condensed in energy to a broad group structure and homogenized in space

$$\sigma_{v,i,r,g} = \frac{\int_v \int_g (\sigma_{i,r}\phi)(\vec{r}, E) dE d\vec{r}}{\int_v \int_g \phi(\vec{r}, E) dE d\vec{r}}. \tag{1.8}$$

A scheme of this process is presented in Figure 1.6. The fuel inventory is calculated at different burnup steps using complete depletion chains, and critical neutron spectra is achieved by a leakage model [9]. The few groups structure and their energy boundaries depend on the specific application. PWR modeling usually uses two groups, simply called the fast and the thermal group with energy cut-off at 0.625 eV while for fast reactors, where fission take place at higher energies, require more groups. Homogenization may be performed over the entire fuel assembly, quarters or pin-by-pin depending on the modeling objectives and the specific reactor. The results of this calculation are the *homogenized few-group cross sections* and possibly the additional equivalence factors, which are stored in a compact *few-group cross section library* for later use in the whole core calculation.

Calculations in detailed environments of multiple assemblies (colorset calculations) are quite common for modeling the reflector. Some important industry codes are APOLLO2 [10], APOLLO3 [11], CASMO [12], DRAGON [13], etc.

(2) Core calculations

Resulting homogenized few-group cross sections are the input of the two group diffusion equations [8], which are the most extended option for three dimensional LWR core calculation in industry [14]. Cross section data, calculated a priori in step (1) and store in *few-groups cross section libraries*, are used to reconstruct cross sections on demand. Modern codes also provide the capability to compute the evolution of the main nuclide concentrations throughout the fuel cycle. These are the most important fissile isotopes, fission products, and burnable absorbers using reduced depletion chains. Some important industry codes counting with these capabilities are APOLLO3 [11], the package ARCADIA [15], SIMULATE4 [16], POLCA-T [17], COCAGNE [18], etc.

[7] For performing the self-shielding calculation an analytical energy spectrum consists of a fission spectrum for the high-energy range, a slowing-down spectrum for the intermediate range and a Maxwellian thermal spectrum for energy lower than 0.625 eV may be used.

[8] The diffusion equations are obtained by using the Flick law between the integrated flux and the currents, i.e. $\vec{J}(\vec{r}, E) = -D(\vec{r}, E)\nabla\phi(\vec{r}, E)$ in the multi-group balance equation.

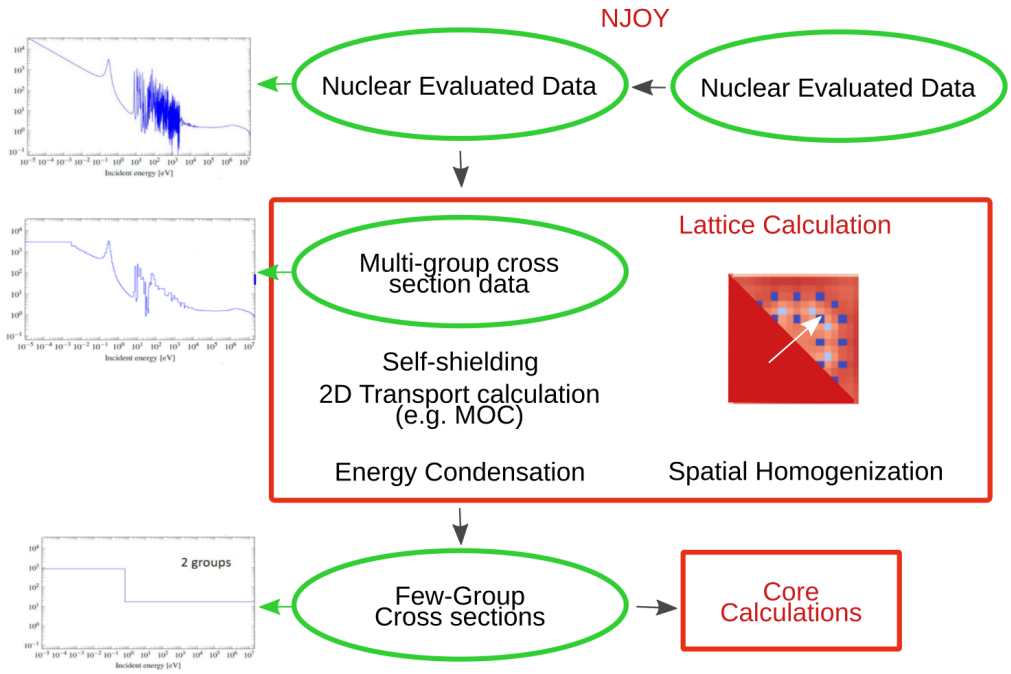


Figure 1.6: Two-step calculation scheme. Evaluated nuclear data are treated by NJOY which generates continuous energy files to be used by the lattice code. The latter performs self-shielding and flux calculations to homogenize and condense nuclear data generating few-group cross section libraries which are the input of core calculations.

1.3.3 Modeling errors of the two-step modeling scheme

Different sources of errors result from this scheme, mainly: [19]

- The significant transport effect (neutron streaming) in very inhomogeneous cores (core calculations using MOX/UO_2 fuels or being axially asymmetrical) are intrinsically difficult to model using the diffusion approximation. Higher order transport solvers such as SP_N may be used instead [20].

- The spatial homogenization and energy collapsing of nuclear data. Reflective boundary conditions are used in a priori cross section preparation, so the collapsing flux used in Eq. 1.8 does not take into account actual reactor conditions which may introduce significant flux gradient that differ from the infinite medium distribution. Classical techniques to deal with these issues belong to the Generalized Equivalence Theory (GET). If proven insufficient, pin-by-pin homogenization may be considered resulting in a more heterogeneous core model [21]. Techniques for embedding the lattice transport [19] in core calculations dropping energy condensation altogether, are being studied. These are generally considered too computationally expensive[9] especially for transient calculations [14].

- The few group cross section representation used at the core level, which are recon-

[9] In a standard lattice calculation the neutron flux is solved for each neutronically distinct reactor component (different fuels and reflector) with a calculation time of $\mathcal{O}$(min) for every calculation point.

structed on the fly from the pre-calculated data libraries in step (1).

The subject of this thesis unfolds within this third item. In what follows only few-group (homogenized) cross sections will be analyzed, simply called "cross sections".

1.3.4 Lattice calculation implementation

In order to represent operational and accidental conditions homogenized data are functionalized by means of reactor "state-parameters", which are physical quantities showing high sensitivity on the neutron reactivity, being reactor-dependent as well. In standard lattice calculations the wide range of possible reactor states are approximated as perturbed conditions with respect to a reference configuration, which is the most probable along fuel exposure at hot full power[10]. So, "branch calculations" are derived for instantaneous changes at operation from a "nominal depletion history", where in principle the only varying state parameter is the burnup in GWd/t[11] (or fuel exposure) as depicted in Fig 1.7. Since the weighting neutron spectra depends on the exposure history, the lattice depletion conditions ought to be the most representative of the average core history for attaining accurate cross section reconstructions.

The burnup discretization depends in part on inner convergence criteria of the lattice code, and sometimes the exact value cannot be easily imposed. Some configuration of state-parameters may result in a non-converging lattice calculation, at least for the settings used in that particular calculation. Even if branch calculation could be calculated freely, many core codes may require a specific construction e.g. a Cartesian scheme for the cross section library equipped with its corresponding "generator vectors" for indexing. In practice the freedom of the user to impose in which points the cross section library is generated may be somewhat limited.

1.3.5 Cross section variables

The difference between depletion and branch calculations gives rise to the so called *history variables* that account for condition changes during isotope depletion. Indeed the burnup is the first history variable, and others may be considered, e.g. moderator density or control rods. This implementation caveat, has significant consequences in how deterministic calculations are set up, and can explicitly appear in cross section dependencies as $\sigma(x_{h,1}, \ldots, x_{h,n}; x_{b,1}, \ldots, x_{b,m})$. We identified n history variables (h) and m as branch (or instantaneous) variables (b) being $n + m = d$, i.e. the total amount of dimensions.

Though conceptually different, in the majority of the thesis no distinction is made

[10] Other schemes have been proposed where many independent burnup calculations are performed with rather few branch calculations. In such scheme the history parameter space covers in a significant extent the classical instantaneous parameter space [22].

[11] This parameter describes the isotope evolution with exposure under nominal conditions and is expressed in units of time-integrated power release per unit of initial heavy metal, i.e. (Giga-Watts-day per tonne of heavy metal)

between the two. Cross section are therefore treated as

$$\sigma(\vec{x}) = \sigma(x_1, \ldots, x_d). \tag{1.9}$$

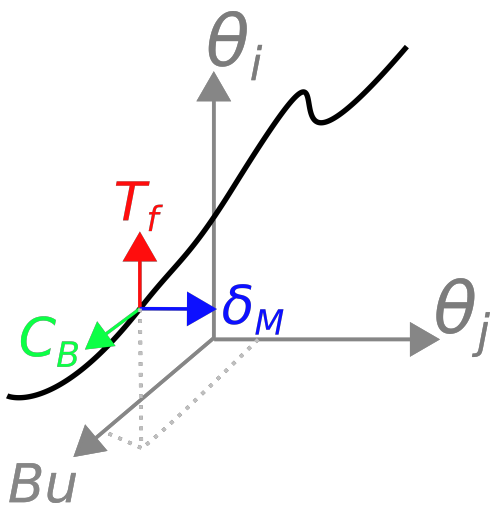


Figure 1.7: Nominal calculations where other history parameters may be present defining one or several depletion histories. For each point, branch calculation in different instantaneous variables may be performed. In this illustration the fuel temperature (T_f), boron concentration (C_B) and moderator density (δ_M) are considered.

Common choices of instantaneous variables for PWR calculations are the temperature or/and density of the coolant, covering all regimes of heat transfer and thermal-hydraulics in the core, an average temperature in the fuel elements to reproduce the Doppler effect, the amount of ^{135}Xe at equilibrium with the power level and the amount of diluted boron in water. Other required variables may be control rods[12], blades, detectors, water gap, proximity to reflector, etc[13].

1.4 Cross sections representation

Provided linear independence of all the parameters in the model cross sections are treated as real valued scalar functions defined over a physical domain $\mathcal{U}^d$. Without loss of generality, we consider the mapping into the unit hyper-cube

$$\mathcal{X} = \{\vec{x} = (x_1, \ldots, x_d) \in [0,1]^d : x_i = \frac{\iota_i - a_i}{b_i - a_i}, \quad \iota_i \in \mathcal{U}_i, \quad 1 \leq i \leq d\}, \tag{1.10}$$

[12]Control rods can sometimes induce additional $\Delta\sigma^{CR}$ terms which can be simpler than the original cross section data. Adequate lattice calculation data, possibly matching the remainder state-parameters, is needed [23].

[13]From an implementation point of view, this history parameters may pose a challenge to lattice calculations parallelization, while in principle calculations in different instantaneous parameters are intrinsically parallelizable.

called *input space*, of dimensionality $d = |\vec{x}|, \vec{x} \in \mathcal{X}$. For obtaining cross section data, this space must be sampled by the lattice code in some specific way. Many times a "full grid" or Cartesian grid resulting from tensor product rule is used

$$\mathcal{X}_{\mathcal{P}} = \vec{\mathcal{X}}_1 \times \ldots \times \vec{\mathcal{X}}_d = \prod_{i=1}^{d} \vec{\mathcal{X}}_i. \tag{1.11}$$

The subscript in $\mathcal{X}_{\mathcal{P}}$ indicates that this is a discrete sampling of the space $\mathcal{X}$ having an amount of point conforming it (cardinality) of $|\mathcal{X}_{\mathcal{P}}|$. This is the standard sampling strategy in many industry applications [24], were d and the variables range ($\mathcal{U}_i$, $1 \leq i \leq d$) depend on the reactor and scope of the simulation. The word *pool* ($\mathcal{P}$) is suggestive, indicating the totality of available data, and as it will be seen later, significant parts of it may prove superfluous when obtained from a Cartesian sampling rule. One of the properties of such domain is being "rectangular" with an hyper-dimensional enveloping box of the data. It's also regular meaning that the points belong to an underling structure. In this particular case corresponding to the generator vectors $\vec{\mathcal{X}}_i$, $1 \leq i \leq d$. In an unstructured support no reduced representation of the data points is possible.

Each lattice calculation point delivers the entire cross section set denominated *output space*

$$\mathcal{Y} = \{\sigma_{irg} : \vec{x} \to W, \vec{x} \in \mathcal{X}, \quad i, r, g \in \mathcal{I} \cdot \mathcal{R} \cdot \mathcal{G}\}^{14}, \tag{1.12}$$

normally $W \subset \mathbb{R}_{\geq 0}$. The total amount of cross sections is the cardinality, $|\mathcal{Y}| = |\mathcal{I} \cdot \mathcal{R} \cdot \mathcal{G}|$ for each specialized isotope $i \in \mathcal{I}$, reaction type $r \in \mathcal{R}$ and group $g \in \mathcal{G}$. The notation $\mathcal{I} \cdot \mathcal{R} \cdot \mathcal{G}$ designates a product different from a tensor rule since, for example, fission products do not have fission cross sections or the amount of scattering cross section can be isotope dependent (for a large $\mathcal{G}$).

In general, cross sections present smooth non-linear profiles with possibly strong variations in localized regions, low order dependence among the variables and low noise [15]. As seen in Fig 1.8. Cross section variance may not be equally distributed in the different dimensions where much of the complex behavior lays, for example, in the burnup variable. This type of physical insight on cross section dependence on state-parameters is a valuable asset which, unfortunately, may not always be available.

For each cross section, a subset of data composed of pairs $\{\vec{x}, \sigma(\vec{x})\}$ extracted from the pool is called the *support*

$$\mathcal{S}_\sigma = \{\sigma(\vec{x}_i), \vec{x}_i \in \mathcal{X}_{\mathcal{S}}\}. \tag{1.13}$$

This is the available information to build a model sometimes called "data sites", "response vectors" or "learning space" with $|\mathcal{S}_\sigma| = |\mathcal{X}_{\mathcal{S}}|$, $\mathcal{X}_{\mathcal{S}} \subset \mathcal{X}_{\mathcal{P}}$. A lattice calculation point

[14] The isotope i is marked with the mass number of the nuclide A, thus unequivocally identify each isotope of each element in this work. In this work homogenization is performed over the entire fuel assembly.

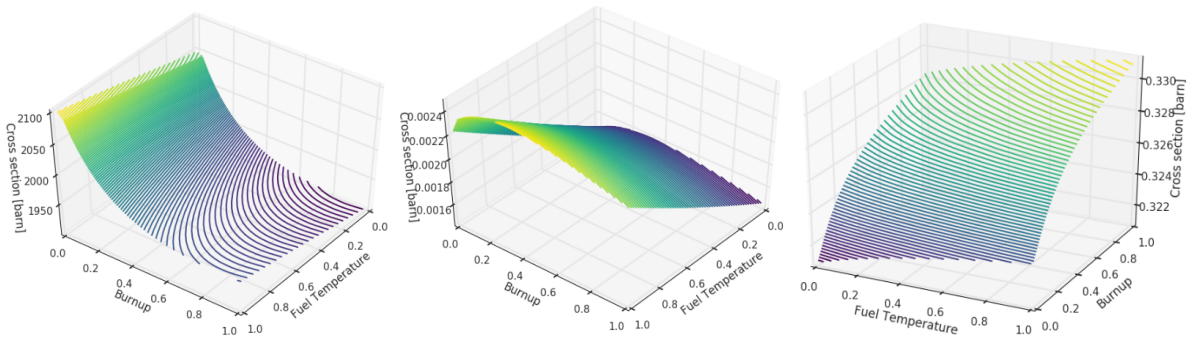
[15] The "low noise" claim is somewhat tricky. Deterministic calculations present convergence errors which, in principle are small in magnitude. Activation/deactivation of different modules, whether be modeling of physical phenomena, numerical solver or acceleration techniques may as well introduce discontinuities or noise. If cross sections were to come from stochastic calculations significant statistical noise may be encountered [25].

delivers all the cross section so, in principle, they "share support". This tends to be the case in many industry settings though it is not mandatory.

Now we are in condition to state **the subject of this thesis: finding the set of approximations** [16]

$$F = \{\hat{\sigma} \simeq \sigma | \mathcal{S}_\sigma, \forall \sigma \in \mathcal{Y}\}. \tag{1.14}$$

Which reads as "finding the set of approximations called F composed of $\hat{\sigma}$ similar to σ knowing the support $\mathcal{S}_\sigma$ for every cross section in $\mathcal{Y}$ ".



(a) (b) (c)

Figure 1.8: Cross section in function of the burnup and the fuel temperature for ^{239}Pu, Σ_{res} and ^{238}U. (a) Thermal fission cross section $\sigma_{239,f,2}$ which becomes an important fissile isotope as burnup increases .(b) Up-scattering cross section $\Sigma_{res,2\to1}$ mainly due to the water in the moderator modeled in the residual cross section. (c) Fast fission cross section $\sigma_{238,f,1}$ which present a very high concentration in UO_2 fuels and contributes significantly to the reactivity even in a thermal reactor.

1.4.1 Modeling requirements

Current cross sections models need not only to deal with ever growing volumes of data [26], but also a larger quantity of cross section reconstructions per core calculation point. In many industry implementations cross sections are modeled by a first order piece-wise polynomial interpolation, here called multi-linear, usually adopting a Cartesian sampling rule of the input space as presented in Eq. [24]. In this simple schema data are simply stored to be quickly interpolated on demand. However, the number of data points may grow exponentially with increasing dimensions which is known as the "Curse of dimensionality" [27]. Highly discretized grids may be specially required for $\mathcal{O}(1)$ approximations such as multi-linear, especially in view of modern microscopic cross section target relative errors laying between 1E-1% and 1E-2% [28, 29]. Under the reasonable assumption that a shared support is used by all cross section, built using a tensor product rule, this requirement raises a storage problem that can be easily understood with an example. Let $\mathcal{R} = 3$ for absorption, fission, and scattering, the latter increasing quadratically with the number of groups. Indeed scattering reactions are in the form $\{\sigma_{g\to g'}, \forall g, g' \in \mathcal{G}\}$. This

[16] Identification of the best state-parameters (type of variables and dimensionality of $\mathcal{X}$), or the definition of the set $\mathcal{Y}$ for that matter, is not treated in this work and relays in "expert's knowledge" for the modeling problem at hand.

data is normally store in full rank matrices at least up to 20 groups of energy taking large percentages of the cross section data. Then, let $size(float) = 32$ byte,

$$size(\mathcal{Y} \times \mathcal{X}) \sim |\mathcal{I}| \times |\mathcal{V}| \times (2|\mathcal{G}| + |\mathcal{G}|^2) \times (N_{Bu} + N_b^{d-1}) \times \frac{32\text{GB}}{1\text{E-9}}. \tag{1.15}$$

For $d = 4$, let $\mathcal{X} = Bu \times T_f \times C_b \times \delta_M$ with discretization $N_{Bu} = 50$ for the burnup and $N_b = 6$ for each branch variable, an acceptable choice for multi-linear interpolation in most cases. Library sizes are presented in Tab. 1.1. An 8 group scheme using 16 homogenization regions or a 20 group scheme with a pin-by-pin homogenization are common in industry [18]. When further increasing the dimensions of the input and output space the library size can reach the TB. And this, for a single depletion history while up to 400 can be required for modeling large cores [19]. In [30] for example, a 1 GB is already considered an excessive library size.

Dimensions (d)	Isotopes ($\|\mathcal{I}\|$)	Groups ($\|\mathcal{G}\|$)	Regions ($\|\mathcal{V}\|$)	Library size
3	10	2	1	1 MB
4	30	2	4	500 MB
4	30	8	16	1 GB
4	10	20	289	300 GB
4	100	20	289	3 TB

Table 1.1: Library size with increasing input (dimension d) and output space (specialized isotopes $|\mathcal{I}|$, energy groups $|\mathcal{G}|$, and homogenization regions $|\mathcal{V}|$). Two common discretization schemes are 16 regions for an 8 group energy structure and pin-by-pin homogenization (289 regions) for a 20 group energy structure. Very large library sizes may be encountered.

For a standard PWR reactor, a gross estimate of the amount of cells (N_{cell}) composing the core model is

$$\begin{aligned} N_{cell} \sim number(\text{Fuel Assembly}) \times number(\text{Axial Discretization}) \times \\ number(\text{Homogenization regions}) \sim 300 \times 50 \times |\mathcal{V}| \end{aligned} \tag{1.16}$$

For 50 specialized isotopes in a 8 group energy scheme $|\mathcal{V}| \sim 16$ are used, which implies a total of 1E9 cross section reconstructions. If 20 energy groups are used in a pin-by-pin discretization ($|\mathcal{V}| \sim 289$) the amount of cross section reconstruction goes to 1E11 for a single core calculation. Between 10 and 30 iterations may be required to converge the thermo-hydraulic module. Compromises are made to define the adequate cardinality of the input and output space depending on the scope of the simulation, yet, a reconstruction speed of at least $\sim$ 1E-5 s/σ is required to perform a core calculation point in a time of $\mathcal{O}(min)$. This speed constraint means that the storage size can be a crippling factor in cross section reconstructions due to the memory accessing speed. Retrieving data at CPU cache, RAM and disk memory is about 10, 100, 1000 clock cycles respectively. The accessing speed problem, induced by the size of the library, undermines the homogenization and equivalence effort of the two-step calculation strategy presented in Section 1.3.

Recapitulating, *accuracy*, *library size*, and *evaluation speed* will be the main criteria for judging the models in this thesis in concordance with other works [31]. Secondary modeling objectives are shown in Fig 1.9. The *lattice calculation time* may be up to 4 min per calculation point which is considerable. In reactor design, cross section data may be

generated several times, and requiring less, rather than more points is preferred. Though the *preparation* of the library itself may be regarded as an "off-line" phase, this may have to be carried out several times for different studies and so a predictable, consistent, and if possible shorter time is preferred. The possibility of including lattice calculation points in an on-line fashion is of interest as well. Finally, all thing being equal, a low *complexity* model, that may be interpretable in physical terms is preferred.

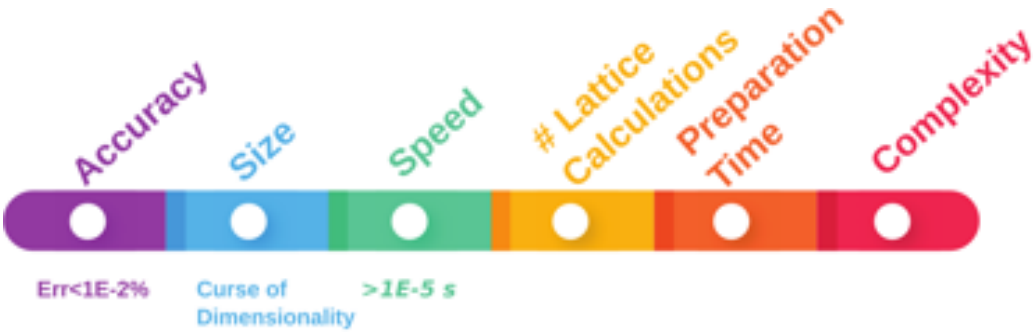


Figure 1.9: Most important few-group cross section modeling requirements. An accuracy in the vicinity of 1E-2% average error and an evaluation speed >1E-5 seconds per cross section (or evaluation time <1E-5 seconds) are typical thresholds found in the literature. The size of the library can suffer from the "Curse of dimensionality" specially for tensor product rules as commonly used in industry. Other criteria for judging a model is the total amount of required lattice calculation, the preparation time of the library itself and the complexity of the model.

1.4.2 Modern trends in core simulations

Modern modeling trends mentioned so far are: increasing in the number of energy groups [20] and specialized isotopes, and the coupling with multi-physics codes [8, 23]. Others are:

- Interested in reactor simulations in a load following scheme which may require to examine current history modeling strategies with changing depletion conditions with special attention to control rod insertion.
- Modeling of UO_2/MOX mix loaded cores, which could result in larger library sizes in order to deal with the boundary condition problems or additional cross section variables.
- An effort to unify codes dedicated to steady-state and transient calculations, as to assure consistency and reduce overall iterations between simulations [23]. Cross section models able to perform well in these two distinct types of simulations could prove more challenging.
- For dealing with innovative reactor concepts, that present strong axial asymmetries, it has been suggested to produce the cross section library using stochastic codes [32]. Such stochastic/deterministic scheme could require advanced regularization techniques in cross section modeling.

1.5 Modeling challenges and state-of-the-art

A chronological literature review of cross sections modeling is presented in Fig. 1.10 where some distinct periods and corresponding strategies can be devised.

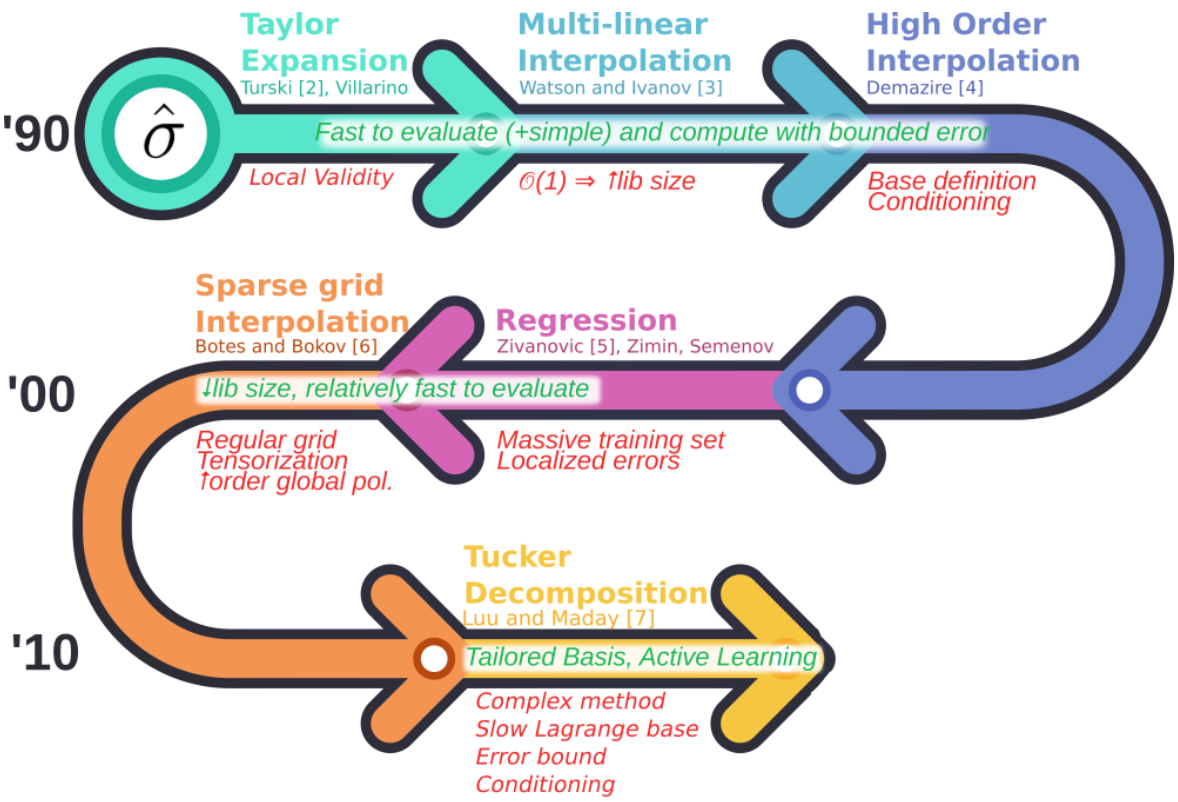


Figure 1.10: Chronological literature review of few-group cross sections modeling. Three periods can be distinguished in a loosely manner. The '00 characterized by Taylor expansions tweaked by the user's expert knowledge relaying in few terms for the expansion. Multi-linear and higher order interpolation in full grids to better cope with multi-physics coupling and improve best estimate calculations. Then, several works on regression and sparse grid interpolation in order to reduce the libraries size. The latest works concentrate in machine learning techniques to let the approximation process learn which is the optimal function space to use and/or in which support points.

In some early works, only macroscopic cross sections where tabulated applying correction factors in the form $\Sigma_0 + \sum_i \Delta\Sigma_i$ which where constructed in a rather heuristic way, based on physical insight and different for every reactor type accounting for both history and instantaneous effects. Discussion was focused in minimizing the amount of variables by, for example, parameterizing with time instead of burnup to account for the boron concentration in an easier way since it was hard-coded by the lattice solver into Σ_0. Only 1-dimensional linear interpolation on exposure was used [33].

Approximations had simple dependencies without cross terms. The POLX-1 code [34] for example, used a quadratic polynomial in the form

$$\Sigma(x_1, \ldots, x_d) = \Sigma_0 + \sum_i^d C_{1,i} x_i + C_{2,i} x_i^2. \tag{1.17}$$

Coefficients where found by ordinary least squares. Higher order polynomials where explicitly avoided due to possible oscillations and, if necessary, the (exposure) domain was sub-divided by hand or pre-treated to ease the modelization (e.g. parameterizing with the square root of the fuel temperature). In [35] a similar form to Eq. 1.17 but up to degree 3 was used considering cross terms of only two variables at the most. These examples and others [36, 37, 38] can be globally encapsulated as *Taylor expansion* where retained terms and truncation was defined by "trial and error" [35]. Models came from "expert knowledge" and were ultimately case dependent. Resulting libraries were small, with models fast to evaluate, as only few coefficients were used. These lend themselves to physical interpretation, as for example a positivity thermal absorption coefficient with fuel temperature due to the Doppler broadening effect or a scattering increase with moderator density, as detailed in [39]. Though such approximations only hold local validity [26], they can be used for small perturbations like usually encountered in open pool MTRs [40, 41] or steady-state calculations [42]. With a proper assessment of the quality of the representation [39], they can still be used like in POLCA7 [43] which counts with several history parameters, spectral corrections and specialized isotopes in a multi-table approach. They have been proposed for pin-by-pin homogenization [44].

By the early '00 point kinetic/1-dimensional thermo-hydraulic modeling shifted to fully coupled three-dimensional neutronic-thermal-hydraulic simulations for transient and accidental analysis [26]; of particular use in PWR having an open lattice with radial cross-flow [45]. Additionally, further core optimization and improvement in best estimate calculations for increase plant availability [46] required an update in the cross section models [45]. In a asymmetric steam line brake accident for example, that requires a 3D modeling of the moderator density, difference between commercial codes (SIMULATE-3 and CORETRAN) were observed, specifically due to the cross section models [47]. Though the same homogenize cross section data was used for both, simplifications in the cross section dependence in boron concentration and moderator density which led to non-negligible differences in power evolution during the accident of up to 20% in the pin power distribution.

Multi-linear interpolation in a full grid has commonly been used to address these kind of issues [46, 29]. User-imposed coefficients are avoided, all inter-dependencies among the variables are accounted for, the reconstruction process is simple and evaluation fast, provided an efficient sorting algorithm. However with an error convergence of only $\mathcal{O}(1)$, usually large library sizes are needed to meet the accuracy requirements [27]. Not only cross section accuracies are of concern, but also the estimation of core feedback coefficient which many times amount to reactivity derivatives. If a Cartesian quadrature is used, an exponential in d amount of computationally expensive lattice calculations are needed, raising the *Curse of Dimensionality* [48] which may hinder evaluation times [17]. For this reason in some implementations a division exists between a master library considering only history variables from which a transient library can be derived to model specific transients by performing branch calculations on the region of interest only [19]. Higher order approximation (splines) [49] and projection into dedicated sub-libraries can mitigate this to some extent. Splines ensure a smooth interpolating approximation that requires fewer points, sometimes called "High-order table lookup method" [45]. Possible oscillation

[17] A naive implementation of multi-linear interpolation on unstructured grids that require complicated searches of the neighbor's points, can prove quite impractical [14].

problems have been reported [46]. Examples of codes using parametrized libraries are CORETRAN while using tables are SIMULATE, CRONOS, DYN3D [14].

An hybrid spline interpolation/polynomial regression scheme was tested in [29] where macroscopic cross section took the form $\Sigma = \Sigma^{base}(Bu) + \Sigma^{branch}_{Bu}(x_2, \ldots, x_d)$. A quintic spline was used only in the burnup and a step-wise algorithm provided a global polynomial for the instantaneous variables. Coefficients were obtained with last square regression. A significance test based on a "bias vs variance" trade-off was used to generate a suitable, though not necessarily optimal model. This selection methodology was tested in a somewhat moderate amount of state variables [22]. A Legendre polynomial base was used to improve matrix conditioning.

In [31] macroscopic cross sections were modeled with global Legendre polynomials of the form

$$\Sigma = \sum_{\vec{k}} \beta_{\vec{k}} \phi_{\vec{k}} \tag{1.18}$$

for a pebble bed modular reactor (using a somewhat unorthodox parameter state selection as required by the benchmark). The main issue of such methodology is the computation cost of solving the integral quadrature for obtaining the coefficients $\beta_{\vec{k}}$ given the basis $\phi_{\vec{k}}$. In [31] a Randomized Quasi-Monte-Carlo Integration (RQMCI) was used to find by quasi-regression, rejecting unimportant ones (only 30 coefficients were retained from the original 15600)[18]. Low-discrepancy Sobol sequences that sample the multidimensional space in a relatively uniform way exhibited the best convergence rates and data scrambling was used to estimate the numerical error of the coefficients $\beta_{\vec{k}}$. However this came at the cost of 1E6 lattice calculation points for a relatively low d. These works show that in practice, cross section dependence on state-variables can be thought of as the sum of relatively low-dimension functions, they exhibit "low practical dimensionality" [27] coherent with previous observations [35]. This insight allowed the enhancement of the methodology by using Sparse Grid Integration [50] which explicitly restrains the order of the approximation and thus the terms of the expression in Eq. 1.18 [27].

Quasi-regression has to deal with both truncation and coefficient's estimation errors. Additionally an enveloping multidimensional square box is required for the domain since basis function need to be defined everywhere. Even if some heuristic schemes of exploration/estimation were suggested for dealing with a larger amount of coefficients, according to [22] actual accuracy needs may require unrealistically big amount of lattice calculations. Possible difficulties in capturing local behavior (e.g. xenon transient) or the retention of very high order terms can also occur [51].

Approximation of smooth functions in a high dimensional domain can profit from a sparse grid discretization [52]. Multi-linear hierarchical interpolation using equidistant nodes was used to approximate cross section in a level-based function representation [53]. The central idea is to restrict the tensor product rule, thus obtaining a sparse grid. Even for a possibly suboptimal first order approximation space, the use of sparse grid in comparison to tensor product rules resulted in an order of magnitude reduction in the library size. The methodology was later utilized [51] with a Lagrange basis where Chebyshev

[18] Coefficient rejection or shrinking a posteriori are techniques for discarding unimportant terms in function approximation. Some sort of significance test is employed to judge which coefficient to reject. With orthonormal basis the coefficient's module is usually used.

nodes were required to deal with the Runge phenomenon[19] [54]. Dimension-wise sparse grid anisotropy allowed to better sample the dimensions that contribute more to the variance [28]. Provided a multidimensional cardinal base, nested rules yielded the inversions of smaller, level-based collocation (inversion) systems at the expense of including all the previous nodes in the basis used to perform the expansion. These methods resulted in global approximations on sparse, though regular and nested, domains attaining the desired accuracy with a relatively small cross section library after shrinking. As with any global expansion method using orthonormal basis, the contribution to the total variance of the approximation is proportional to the coefficient's module. This allows to wage rejection techniques to perform shrinking a posterior and facilitates the interpretation of cross section dependences. Interpolation with a Lagrange base yields superior accuracy than quasi-regression in sparse grid according to [55]. In a posterior study, this methodology was successfully tested on MOX fuel as well [55].

Optimization on the anisotropic sampling can be found in [14] with respect to k_∞, though limited to non-uniform Cartesian grids. In this article, it's noted that integral parameters such as the multiplication factor are not equally sensitive to the different cross sections discussing the possibility of considering other elements for the optimization of the library [14].

Though modern machine learning terminology is not present [29], [31], [27] are examples of a supervised learning process, where expansion bases are predefined without knowing in advanced the retained combinations, which result from significance tests. In [14, 28] supervised searches for optimal sampling schemes are presented, though constrained to regular grids.

Another example of supervised learning for model selection can be found in [56]. A genetic algorithm selects from the model by combining monomial functions that minimize the MSE. Discussion is framed in terms of optimal custom function space but the global polynomial used has relativity high errors of about 1%. Here the concept of optimizing the expansion space itself instead of the support is discussed. This is the case in [57] where each cross section participates in the formation of an optimal function space on which to perform the expansion using the Karhunen–Loève quadrature. Additionally the "Empirical Interpolation Method" defines an un-structured, or scattered support on which to build the library. Yet, this supervised learning implementation is severely intertwined with the expansion space: support points candidates result from a combinatorial procedure based on the error of the eigen-functions. The sampling is restrained to a "Tucker grid" (built to obtain the eigen functions) and subjected to a second selection process for lowering the amount of additional lattice calculations. A posteriori shrinking is then performed to eliminate unimportant terms. The error analysis was somewhat limited and possible conditioning number problems not reported.

[19] Runge phenomenon is the problem of very strong oscillations in polynomial interpolation when using equidistant nodes. This can occur at relatively low order polynomial orders that approximate well behaved smooth functions.

1.5.1 Final thoughts and discussion

As indicated in Fig 1.10 three distinct periods are loosely identified. In the '90, in the context of very low computation resources by today standards, core simulations and cross section models were quite reduced in scope, the latter basically circumscribed to global polynomials with a few terms. In the '00 a demand for best estimates calculations converged to piece-wise approximations in full grids, possibly employing sub-tables to render the library size more manageable. From then onward the design of more challenging reactors, a shift towards the treatment of high dimensionality problems and interest in black-box approximations schemes stimulated the search for automatic model constructions. Many works are dedicated to supervised learning techniques for sampling the cross section domain while others try to produce tailored function spaces to facilitate the approximation.

This cross section modeling results in some highlights guiding our work:

- A full grid support suffers not only from the "Curse of dimensionality", but can include significant amounts of unnecessary data as evidenced by the low errors attained with sparse grids [28], by the variability of retained regression coefficient cross section-wise [22], and the success of a posteriori shrinking in virtually every method counting with an orthonormal bases.
- A posteriori shrinking has proven quite successful, but it also implies that effort in producing lattice calculation data and preparing the library has been discarded.
- A tabulated library not including any type of functionalization doesn't easily allow for an undesirable, but possible request of extrapolation in cross section evaluation.
- In traditional regression techniques the tensor product rule is not required for sampling the approximation's support. However the Curse of dimensionality can still "creep into" the regression model in the form of the amount of high order terms that need to be retained [27] or the user's effort in imposing the right structure.
- Problems with global approximations, either in regression or interpolation, have been reported particularly for capturing local variations and avoiding the Runge phenomenon. Local approaches are therefore of interest.
- Piece-wise approximations with splines are reported in the literature, but without detailing its properties. Using a B-spline basis is only mentioned without providing additional information on the boundary conditions, knot vector, order of the approximation, or typical approximation difficulties of higher order interpolation.
- In general, cross sections exhibit smooth profiles, which is a natural consequence of the underling physical phenomena. However, in practice, numerical lattice code can occasionally introduce noise [27] and approximations should, if possible, be able to deal with this potential issue.
- Benefits of using unstructured support with respect to full grids have been suggested though the majority of supervised learning techniques have only been applied to regular grids [28, 14]. Many times nodes are selected in view of the quadrature

system (Tucker or Chebyshev), but are clearly not optimal from an information point of view. For example in a Chebyshev discretization the first and last burnup regions are equally sampled, which incidentally may not be attainable with the precision required to deal with possible numerical problems in the approximation.

- Works using unstructured supports [57] have shown great potential but without providing comparisons with more traditional function spaces as used in industry.
- In general no comparative analysis is to be found on selection criteria for supervised learning techniques that compare σ, Σ, and k_∞ and their resulting errors. Neither the interplay between cross section sharing support and the supervised learning procedure.
- Custom function spaces [57] are a very interesting feature though, if low order functions can indeed describe well the data [27], the benefit of such tailored basis may be of secondary importance.
- Artificial Neural Networks have been used before in Nuclear Engineering but sparsely in cross section modeling [58]. This area of nuclear engineering has not profited from the rapidly developing machine learning community and their computational frameworks.
- No work is to be found were a single model deals with the entire cross section set. A "multi-output" central model has not been treated yet.

The objectives of this thesis is to analyze few-group cross section modeling using splines with B-splines, Kernel Methods with Pool Active learning and Artificial Neural Networks in single and muli-output models. These models are assessed in view of an adequate accuracy, evaluation speed and library storage reduction. Splines represent the state-of-the-art in many industry codes, and are a high order local approximation. Kernel Methods allow to consider very general supports, subjecto to active learning techniques, where a vast variety of function space can be reproduced at a reduced computational cost. Artificial Neural Networks can learn relations from data allowing to treat the totality of the cross section set with a single, very efficient model. Hopefully this thesis is well aligned with the points previously mentioned for which new answers or research paths will be provided. Each methodology will be introduced in their corresponding chapters and compared to each other. The approximations resulting from these methodologies will be compared to multi-linear interpolation in a Cartesian grid which is the standard in many industry applications [59, 18, 24, 30, 46, 45, 29].

1.6 Organization of thesis

This thesis is organized as follows:

- In chapter 1 a brief introduction to nuclear power plants, nuclear reactors and their governing equations was provided. Different calculation strategies were discussed, emphasizing the deterministic two-step scheme from which few-group cross sections

are obtained, presenting its implementation and requirements. This was followed by a thorough literature review extracting the main modeling tendencies with their advantages and drawbacks. The thesis' main objectives and modeling strategies were outlined.

- In chapter 2 the use case considered throughout the thesis is provided, recalling the most important characteristics and detailing the lattice calculation settings. The constitution of the input and output space is presented showing some grid examples. The cross section data division to construct and test the approximations is shown. The performance metrics are presented for σ, k_∞ and Σ introducing for the latter the notion of importance. An analysis of the cross section's variance is carried out as well.

- In chapter 3 a review of global and local interpolating approximations using univariate expansion basis is carried out showing its benefits and limitations. Piecewise polynomial interpolation is analyzed in detail specifically for Splines utilizing a B-spline basis, which is a classical approximation technique, representing the state-of-the-art in many industry implementations. Some aspects of this type of approximations such as the knot vector construction and possible oscillations in the approximation are examined setting the ground of comparison for other, more innovative methodologies.

- In chapter 4, Kernel methods are used for cross section modeling in the context of ridge regression approximations. A kernel induces a function space that requires no particular structure of the support allowing to wage Active learning techniques to select the support points of the cross section data set. The decision criterion is based on loss functions that use σ, Σ or k_∞ with and without a shared support condition for the cross section set. A sensitivity analysis on regularization, preconditioning and the approximation's order is carried out. The higher order approximations obtained overcome the difficulties found in chapter 3 related to the use of the tensor product rule and the need of regularization.

- In chapter 5 approximation by Artificial Neural Networks is presented discussing the main aspect of the methodology. A parametric study on learning hyper-parameters, network's size, and topology is carried for feed forward fully connected multi-layer networks. Results are compared to multi-linear interpolation in terms of library size reduction, accuracy and evaluation speed. Multi-output models were the entire cross section library is modeled with a single network are examined together with their performance and characteristics.

- In chapter 6 the main conclusions and overall thesis contributions and perspectives are provided.

In the Appendix C a review of history effects modeling strategies is carried out analysing their improvement capabilities and possible difficulties with certain non-nominal exposure conditions. This is a self-contained, independent study.

Chapter 2

Use case and performance metrics

In this chapter the use case considered throughout the thesis is presented, recalling the most important characteristics and detailing the lattice calculation settings. The constitution of the input and output space is provided showing some grid examples and the data division to construct and test the approximations. The performance metrics are presented for σ, k_∞, and Σ, introducing for the latter the notion of importance. The error bounds discussed in the introduction are further studied with the aid of a sensitivity coefficient on the multiplication factor. An analysis of the cross sections variance and scatter projections of the data for important cross sections is carried out as well.

2.1 PWR standard UO_2 fuel assembly

The modeling methodologies in this thesis are analyzed with a specific *use case* provided by the OECD-NEA Burn-up Credit Criticality Benchmark (Phase-IID) which studies the effects on depletion of prolongued control rod insertions [60]. It's a classical PWR fuel assembly as shown in Fig. 2.1 composed of 17×17 UO_2 fuel rods with 4% enrichment w/o ^{235}U and with 25 guide tubes. Material and geometrical specifications are fully available in the benchmark report [60]. The State-parameters used for the depletion calculation are: a constant boron concentration of $C_{B,0} = 456$ *ppm* diluted in the moderator, a fuel temperature of $T_{f,0} = 600$ °C and a moderator density of 0.72 g/cm^3. The concentration of ^{135}Xe is not fixed by the power but set "free" to change with exposure.

2.2 Lattice calculation settings

The lattice code employed was APOLLO 2.8 [10]. Self-shielding calculations were performed using the Livolant-Jeanpierre formalism [61, 3] for all actinide isotopes, principal fission products and the constituents of the cladding, burnable absorber pins and control rods. These are: ^{107}Ag, ^{109}Ag, ^{110}Cd, ^{113}Cd, ^{241}Am, ^{243}Am, ^{nat}Cr, ^{133}Cs, ^{153}Eu, ^{nat}Fe, ^{154}Gd, ^{155}Gd, ^{156}Gd, ^{157}Gd, ^{158}Gd, ^{160}Gd, ^{115}In, ^{95}Mo, ^{143}Nd, ^{145}Nd, ^{nat}Ni, ^{237}Np, ^{238}Pu, ^{239}Pu, ^{240}Pu, ^{241}Pu, ^{242}Pu, ^{103}Rh, ^{101}Ru, ^{147}Sm, ^{149}Sm, ^{150}Sm, ^{151}Sm, ^{152}Sm, ^{99}Tc, ^{234}U,

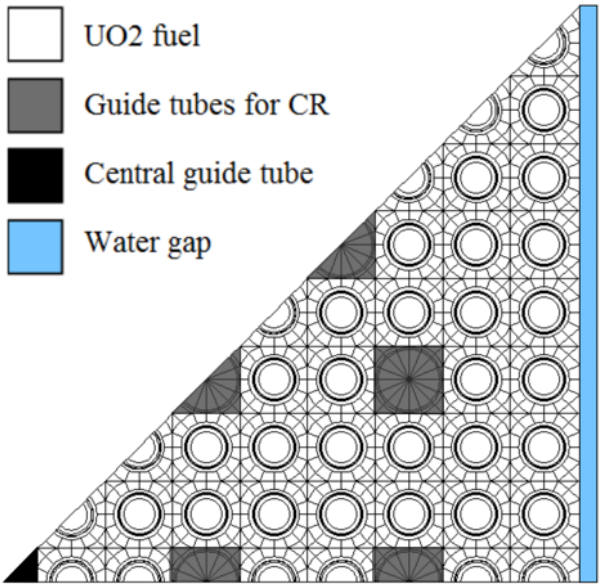


Figure 2.1: Layout of 1/4 UO_2 PWR fuel element for the Credit Criticality Benchmark [60]. The meshing of depleting materials is shown. A reflective boundary condition is used.

^{235}U, ^{236}U, ^{238}U, and ^{nat}Zr. The options for the treatment of the resonance interferences (resonant mixtures model) are chosen for ^{235}U, ^{238}U, ^{239}Pu, and ^{240}Pu.

Flux calculations were carried out using the Method of Characteristics (MOC) to model the transport equation with the P_3 anisotropic scattering. The spatial mesh is presented in the Fig. 2.1. A 281-multi-group cross section library based on JEFF-3.1.1 with a reflective boundary conditions are used imposing the B_1 fundamental mode leakage model to obtain a critical flux [3]. The trajectories along which the MOC solves the balance and transmission equations were defined using parallel trajectories spacing equal to ~ 0.05 cm, with the angular quadrature of product type, where the azimuthal spacing was $\pi/24$, with three polar angles following the Legendre quadrature between 0 and $\pi/2$.

Every pin and ring within depletes independently. The calculation uses the predictor-corrector scheme based on polynomial interpolation of the reaction rates, while solving the Bateman equations with the fourth order Runge-Kutta method. At each depletion step the flux is recalculated. Aditional details can be found in [62] and [63].

2.2.1 Input space composition

In this thesis the state-variables that conform the input space are the burnup (Bu), fuel temperature (T_f), and boron concentration (C_B) with cross sections in the form $\sigma(Bu, T_f, C_B)$. The target exposure of 45 GWd/t represents a plausible utilization option. The boron concentration is in the interval [100,1500] ppm and the fuel temperature, averaged over the fuel rod, in the range [200,2000] °C. In this thesis this three dimensional input space ($d = 3$) is normalized to the unit hyper-cube as presented in Eq. 1.10, hence $\vec{x} \in [0, 1]^3$.

In the Appendix C, for analysing the history effects, another calculation scheme presented therein is used. It basically consists in several independent burnup calculations at different conditions (moderator density, control rod position, etc), but without branch calculations.

2.2.2 Data set discretization

The data set is formed from a Cartesian sampling of the domain using a discretization $Bu \times T_f \times C_B = 170 \times 16 \times 16$ as presented in Table 2.1. This is a sufficiently dense[1] sampling of the input space to extract two disjoints sets, one to build the approximation, and others to perform the error analysis.

Index(i)	Variable	Range	Discretization ($\lvert\mathcal{X}_i\rvert$)	
			Data set	Industry set ($\mathcal{X}_\star$)
1	Bu [GWd/t]	[0,45]	170	35
2	T_f [°C]	[200,2000]	16	6
3	C_B [ppm]	[100,1500]	16	5

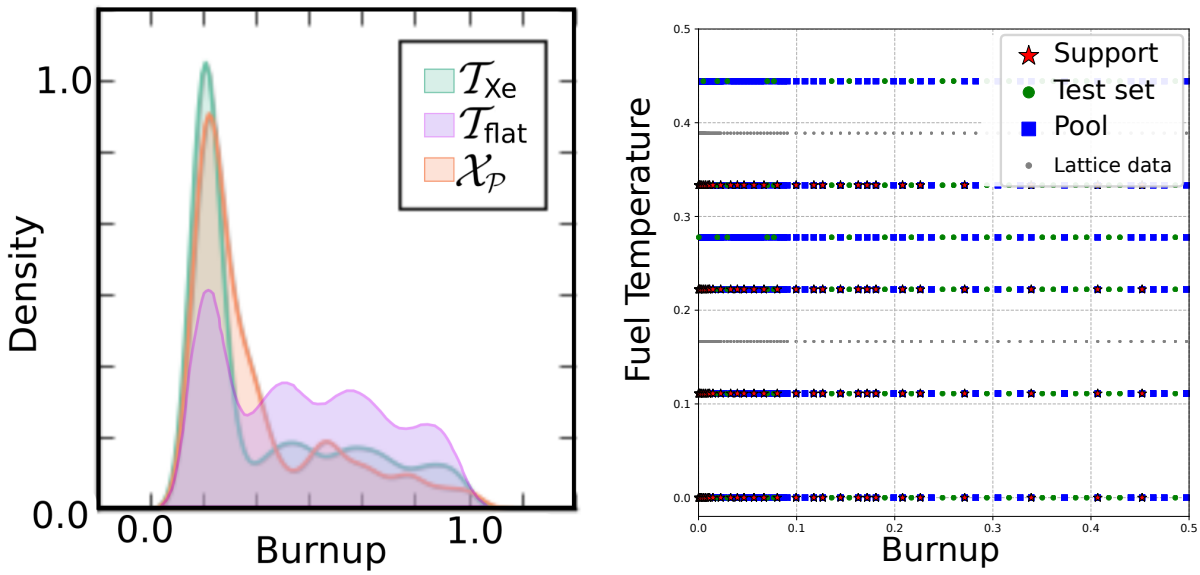
Table 2.1: Constitution and sampling of the input space. The three state-variables considered in this thesis are the burnup, fuel temperature, and boron concentration sampled using a Cartesian grid whose discretization is presented in the Data set column. Additionally an industry set is shown representing a plausible discretization option for this type of fuel assembly.

A special support called the Industry set ($\mathcal{X}_\star \subset \mathcal{X}_\mathcal{P}$), having 1050 points for each cross section, represents a reasonable discretization for an UO_2 fuel assembly in this range of variation and provides a specific support to compare the methodologies. It's common practice to perform lattice calculations using a fine discretization for the burnup, as to assure proper convergence, and then build the cross section library from a subset of the data.

The data set coming from lattice calculations is divided into the pool $\mathcal{X}_\mathcal{P}$, from which different supports are extracted $\mathcal{X}_\mathcal{S} \subset \mathcal{X}_\mathcal{P}$, and the test set $\mathcal{T}$ for evaluating the cross sections errors. The large amount of points allows for such separation, fixing the test set independently of any support, resulting in a stable error evaluation as $\mathcal{X}_\mathcal{S} \subset \mathcal{X}_\mathcal{P} \cap \mathcal{T} = \emptyset$ for any methodologies treated in this work. Care has been taken to assure that if $|\mathcal{X}_\mathcal{S}| < |\mathcal{X}_\star|$ then $\mathcal{X}_\mathcal{S} \subset \mathcal{X}_\star$ as to have a pleasant progression towards the error of the industry data set. Also, the borders of the hyper-cube are always considered for any support, as to exclude extrapolations in the error analysis. Throughout the majority of the thesis the size of the pool is limited to $|\mathcal{X}_\mathcal{P}| = 2500$ (instead of $170 \times 16 \times 16$) which is enough to contain the Industry set and attain the target errors. It has been noticed that additional, independent lattice calculations, especially for different burnup discretizations, can produce incongruent tests with respect to the original data set, difficult to deal with. As very high cross section accuracies are required, these type of inconcistancies can have deleterious affects in the error analysis [64].

[1] This represent about 500MB of data with 43520 points of data for every cross section. In [30] for example, a GB of data is already considered a problematic library size.

Two different test sets are considered. In the first one a relativity homogeneous sampling distribution is obtained, by dropping points in the first burnup values, $|\mathcal{T}_{flat}| = 3000$. This will be the test set mostly used in the thesis. The second one follows the distribution of the original data set thus having a high density in the first burnup values, $|\mathcal{T}_{Xe}| = 5000$. In Fig. 2.2a the density of the pool and test points is shown for the two data sets. In Fig. 2.2b grid points are shown for the burnup and fuel temperature at nominal boron concentration. The Cartesian structure can be appreciated were data has been divided into the pool, which is the total set of available points for defining a support (blue), and the test set $\mathcal{T}_{flat}$ for evaluating the approximations (green). Test points are located only where pool points are also available to facilitate the comparison with multi-linear and to reduce the computational burden of the error analysis. As an example the Industry support presented in Tab. 2.1 is shown in red.



(a) (b)

Figure 2.2: Density of pool and test points projected onto the burnup dimension. $\mathcal{T}_{Xe}$ concentrate more test points in the first burnup values. (b) Example of a 2D grid slice showing the pool and test points. In red the Industry set support which is extracted from the pool which hereditates the regular structure from the original lattice data. In green the test points used for error analysis. In gray unused lattice calculation data.

2.2.3 Output space composition

In regards to the output space, the spatial homogenization of nuclear data is performed over the whole assembly and a two energy group scheme is used (cut-off at 0.625 eV). The set of specialized isotopes is:

$$\mathcal{I} = \{^{109}\text{Ag}, ^{243}\text{Am}, ^{153}\text{Eu}, ^{155}\text{Gd}, ^{95}\text{Mo}, ^{143}\text{Nd}, ^{145}\text{Nd}, ^{237}\text{Np}, ^{238}\text{Pu}, ^{239}\text{Pu}, ^{240}\text{Pu}, ^{241}\text{Pu}, ^{242}\text{Pu}, ^{103}\text{Rh}, ^{101}\text{Ru}, ^{147}\text{Sm}, ^{149}\text{Sm}, ^{150}\text{Sm}, ^{151}\text{Sm}, ^{152}\text{Sm}, ^{99}\text{Tc}, ^{234}\text{U}, ^{235}\text{U}, ^{236}\text{U}, ^{238}\text{U}, ^{133}\text{Cs}, ^{135}\text{Xe}, ^{135}\text{I}\}, \quad (2.1)$$

following the recommended isotopic specialization for criticality safety analysis [65]. The remaining isotopes are lumped together in a residual macroscopic cross section Σ_{res} that

has a constant concentration of 1, see Eq. 2.10. The total number of isotopes treated in this thesis is $|\mathcal{I}|+1 = 30$. This is a significant amount of specialized isotopes, for example in the state-of-the-art ARTEMIS core simulator about 50 are normally used [23]. In this work only cross section participating in the infinite multiplicity factor are analyzed, being $\{\sigma_{a1}, \sigma_{a2}, \sigma_{f1}, \sigma_{f2}, \sigma_{1\to2}, \sigma_{2\to1}\}$[2]. The cardinality of the output space, i.e. the total amount of functions to approximate, is $|\mathcal{Y}| = 144$. The total size of the cross section library using the Industry set is therefore $|\mathcal{Y}| \times |\mathcal{X}_\star| = 144 \times 1050$ =1.5E5 terms.

2.3 Quantities of interest and performance metrics

In this work the accuracy of cross sections approximations (see Eq. 1.14) is assessed by computing the absolute value of the error

$$\Delta\sigma = \text{abs}(\hat{\sigma} - \sigma). \tag{2.2}$$

The approximation $\hat{\sigma}$ is usually evaluated in the test set. A direct analysis of the error distribution can facilitate the detection of "difficult" regions to approximate. A mean cross section error is defined as

$$\text{E}_\sigma = \frac{\sum_{i=1}^{|\mathcal{T}|} \Delta\sigma_i}{|\mathcal{T}|}, \tag{2.3}$$

measured in absolute values $1\text{E}-24$ cm^2=1 *barn*. A mean relative error is

$$\text{RE}_\sigma = \frac{100}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \Delta\sigma_i/\sigma_i, \tag{2.4}$$

in % which is convenient due to the significant variation in cross sections absolute values ($\sigma^{^{135}\text{Xe},a,2} \sim$ 1E6 barn while $\sigma^{238,f,1} \sim$ 1E-1 barn)[3]. RE_σ helps to identified approximations with high error.

The model's overall accuracy can be set forth in different ways. The simplest one is considering the error in both the test set $\mathcal{T}$ for all cross section of the model $\mathcal{Y}$ (i.e. the approximation's errors in F, see Eq. 1.14)

$$\text{AV}_\sigma = \frac{\sum_{i=1}^{|\mathcal{Y}|} \text{RE}_i}{|\mathcal{Y}|}. \tag{2.5}$$

This gives an overall model's error which can be easily compared between different methodologies. As mentioned in Sec. 1.4.1 modern target relative errors are between 1E-2% and 1E-1%. Another useful point of comparisons is multi-linear interpolation in a

[2]In this work the fission cross section is affected by the ν parameter being actually $\nu\sigma_f$. Reactions in the form $n + A \to 2n + (A-1)$ $n + A \to 3n + (A-2)$ are not included in the absorption cross section. Only the scattering cross section corresponding to the P_0 term are analyzed.

[3]A tolerance is considered for a possible case where the error may be small, but the cross section's absolute value so small that the relative becomes disproportional large. The condition to exclude such point is $\Delta\sigma)/\sigma > 5$ and $.98 \leq \hat{\sigma}/\sigma \leq 1.02$. The first part requires the relative error to be bigger than 5%. The second demands for the approximation $\hat{\sigma}$ to be within $\pm2\%$ of σ. Bad approximations won't be excluded of the error analysis but good ones with high relative errors will. In practice this condition never rose.

full grid using the Industry set, representative of industry applications. This type of error however, may average out variations in E_σ which thus analyzed as well.

The overall model error can, for example, be discriminating by groups

$$\begin{aligned} AV_{\sigma,1} &= \sum_{i=1}^{|\mathcal{Y}|_{g=1}} RE_i/|\mathcal{Y}|_{g=1}, \\ AV_{\sigma,2} &= \sum_{i=1}^{|\mathcal{Y}|_{g=2}} RE_i/|\mathcal{Y}|_{g=1}, \end{aligned} \tag{2.6}$$

and so forth for reaction type r or even isotope i. The use of these types of errors will be limited, but they can depict the error distribution within $\mathcal{Y}$.

The maximum relative error (in %) is

$$MAX_\sigma = \max_{x\in\mathcal{T},\sigma\in\mathcal{Y}} 100\Delta\sigma(x)/\sigma(x), \tag{2.7}$$

for all the cross section in $\mathcal{Y}$. Similar error definitions are used for macroscopic cross sections Σ which can be though off as an error on σ weighted by the concentrations. Also the error on infinite multiplication factor k_∞, derived from macroscopic cross sections, are analyzed. L1 norms on relative errors are used in this work following the recomendations found in litterature for assessing the cross section models [14, 29]. Many times cross section library size will be compared at constant error. This viewpoint, implies a "test of significance" of the parameters of the model, quantifying how much information are they able to convey. All the methodologies were tested with a CPU hardware of 24 cores of 2300 MHz and 126 GB of RAM with a GeForce GTX 1080, Python2.7 [66], and Cuda 10.1 [67].

2.3.1 Multiplication factor sensitivity to cross section error

To provide some insight into the rationale behind the relative errors targets of [1E-2%, 1E-1%] a sensitivity coefficient from Eq. 1.6 is considered. The derivative of the multiplication factor with respect to the macroscopic thermal fission cross section is

$$S_{f,2} = \frac{\partial k_\infty}{\partial \nu_2\Sigma_{f,2}} = \frac{1}{(\Sigma_{a,1}+\Sigma_{1\to2})(\Sigma_{a,2}+\Sigma_{2\to1})/\Sigma_{1\to2}-\Sigma_{2\to1}}. \tag{2.8}$$

For the use case treated in this work, $S_{f,2}(0, T_0, C_{B,0}) \sim 9$. Then,

$$|\Delta k_\infty| \sim 9|\Delta\nu_2\Sigma_{f,2}|. \tag{2.9}$$

In a UO_2 fresh fuel, $\nu_2\Sigma_{f,2} = C_{^{235}U}\sigma_{235,f,2}$ with, in this case, $C_{^{235}U} \sim$ 2.51E-4/(barn cm) and $\sigma_{235,f,2} \sim 600$ barn. Then a $\Delta\sigma_{235f,2} = 1\%$, which is about 6 barn, results in an unacceptable $|\Delta k_\infty| \sim 1300$ pcm[4] only due to reconstruction error in the cross sections. If a more acceptable error of $|\Delta k_\infty| = 10$ pcm is demanded, a $|\Delta\sigma_{235,f,2}| \sim 0.05$ barn, i.e., about 1E-2% relative error is required. Though this is possibly an upper bound, Σ is usually composed of many cross sections where error cancellation can occur and the error targets prescribed by the literature seem justified.

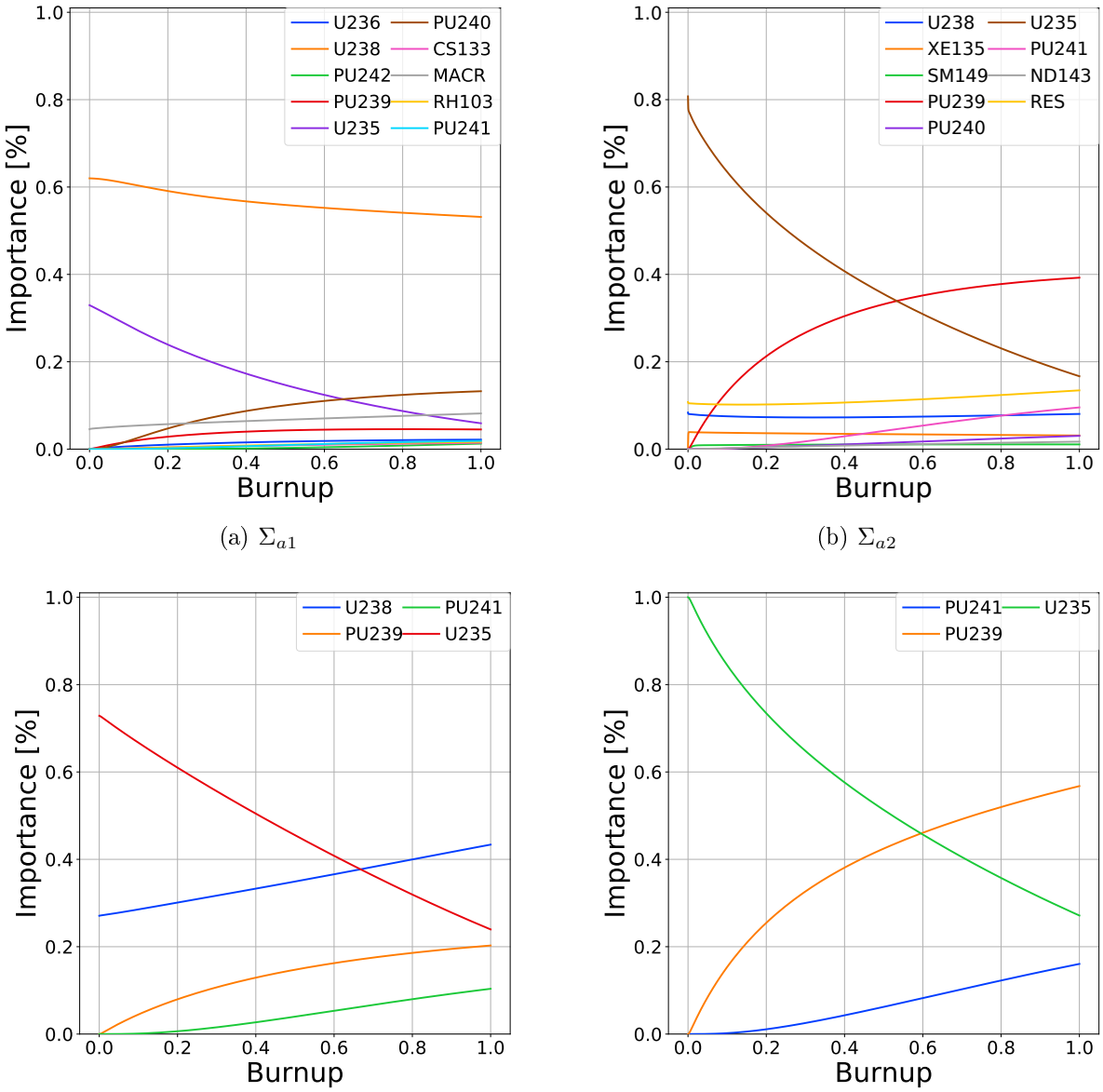
[4]This is about two dollars worth of reactivity, being the β_{eff} of a PWR about 650 pcm. This unit is relative to the prompt critical state, in which the chain reaction is no longer controlled. Such error, is of course, unacceptable.

2.3.2 Important microscopic cross sections

Isotopes participate differently in the macroscopic cross section due to their changing concentration $C_i(Bu)$ with burnup and cross section absolute value. For a given reaction and group, the *Importance* is

$$I_\sigma(Bu) = I_{i,r,g}(Bu) = \sigma_{i,r,g}(Bu, T_{f,0}, C_{b,0})C_i(Bu)/\Sigma_{r,g}(Bu, T_{f,0}, C_{b,0}). \quad i \in \mathcal{I}, \tag{2.10}$$

being $0 \leq I_\sigma \leq 1$. The importance function depends mainly on the burnup showing low sensitivity to instantaneous values so nominal are used. It's shown in Fig 2.3 for isotopes with $\max(I_\sigma) > 0.1$ were the apparition of fission products in the absorption cross section, of plutonium in thermal fission and the effect of spectrum hardening in the fast fission (^{238}U) can be noted. With respect to $\Sigma_{1\to2}$ it's composed up to 99.7% of Σ_{res} and $\Sigma_{2\to1}$ is 90% Σ_{res} and 9% ^{238}U (1% others). This already announces a potential conflict in modeling objectives: not only few specialized isotopes participate significantly in Σ, but in particular the scattering cross sections, that can be the most demanding in library storage, can have negligible fractions of Σ.



(a) Σ_{a1} (b) Σ_{a2}

(c) Σ_{f1} (d) Σ_{f2}

Figure 2.3: Evolution of cross section importance with burnup. At very low values ^{235}U, ^{238}U, and Σ_{res} constitute the totality of the macroscopic cross section. With increasing burnup actinides and fission product start to form. Only cross section with $\max(I_\sigma(Bu)) > 0.1$ are plotted. (a) In fast absorption the Pu element becomes progressively important. (b) A similar process is observed including the neutronic poisons ^{135}Xe and ^{149}Sm. (c) Fissionable isotopes such as ^{238}U participates in the fast fission macroscopic cross section. Being the concentration of ^{238}U practically constant it gains importance as ^{235}U depletes. (d) Only isotopes with a pair mass number are fissile. ^{239}Pu plays a major role in sustaining the chain reaction at high burnup values.

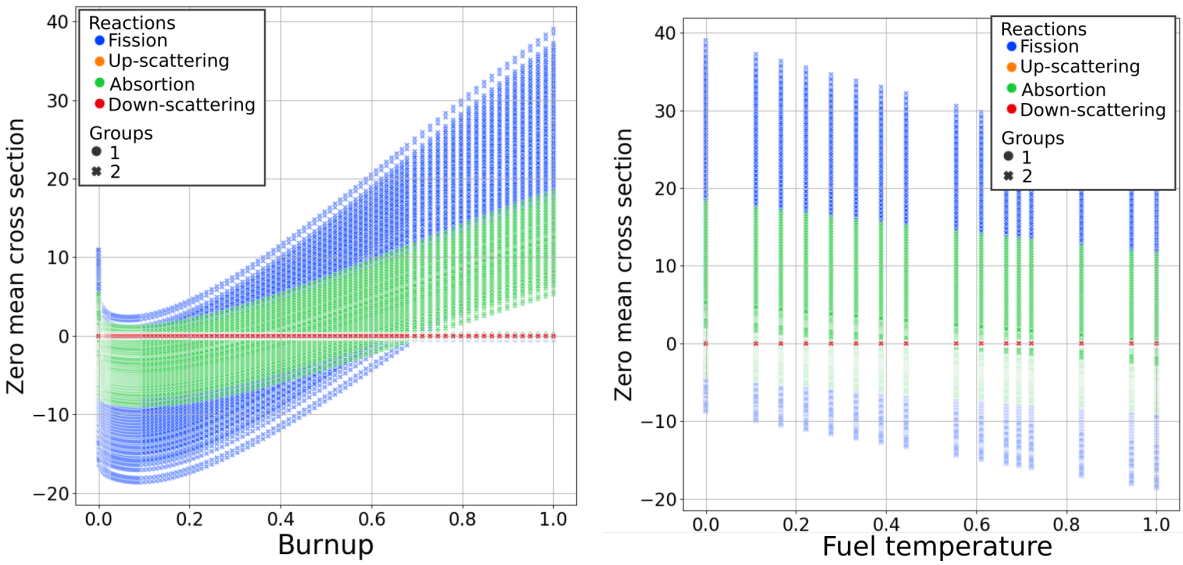
2.4 Analysis of variance

Cross section data projection into the state-parameters is displayed in Fig. 2.4 for the isotope ^{235}U and in Fig. 2.5 for Σ_{res} discriminating by reaction type and energy group.

To ease the comparisons, each cross section is affected by

$$\sigma \leftarrow \sigma - \text{mean}(\sigma), \forall \sigma \in \mathcal{Y}, \tag{2.11}$$

thus having a zero mean. It can be seen that cross section data exhibits a complex dependence with burnup, particularly for thermal absorption and fission. The "effective low order" (yet not-linear) dependence for instantaneous variables can be observed. ^{235}U has a similar dependence on both instantaneous variables while Σ_{res} changes significantly with boron concentration since it includes the moderator on which the boron is diluted. In Fig. 2.5a each line of data for $\Sigma_{a,2}$ includes all the values with changing T_f that has a very mild spectral influence on the residual cross section, stripped away of the majority of fission product and actinides. For every cross section, the cloud of data is proportional to the variance that could be correlated with complexity and thus difficulty of approximation. The discretization discussed in Table 2.1 are apparent here were 16 columns of data corresponding to the Cartesian sampling scheme for the lattice calculations can be seen.



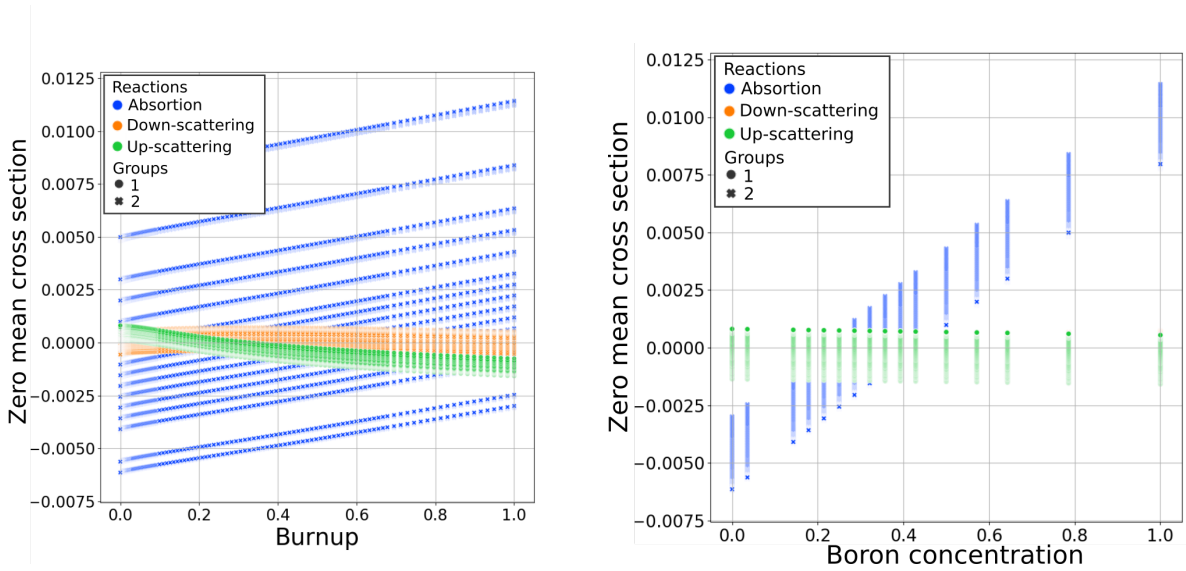
(a) ^{235}U (b) ^{235}U

Figure 2.4: Zero mean cross section projections for the isotope ^{235}U using the entire data set ($d = 3$) discriminated by reaction and group. (a) Significant variation is observed in function of the burnup, specially for fission and absorption cross section. The scattered cloud of mainly thermal cross section data is due to the projection of the fuel temperature and boron concentration into the burnup dimension. In (b) "low order" but not linear dependence of cross section data with the fuel temperature can be observed.

σ	$\sigma_{^{238}U,f,1}$	$\sigma_{235,f,2}$	$\sigma_{235,a,1}$	$\sigma_{235,a,2}$	$\sigma_{239,f,2}$	$\sigma_{^{238}Pu,f,1}$	$\sigma_{^{135}Xe,a,2}$	$\Sigma_{2\to1}$	$\Sigma_{1\to2}$
Var(σ)	7.9E-6	92.29	0.006	20.07	4.5E3	0.25	1.6E9	7.1E-8	3.4E-7
Var(σ)/ mean(σ)	2.4E-5	1.5E-1	5.1E-4	0.07	2.4	1.0E-5	1.1E3	2.0E-5	3.4E-5

Table 2.2: Variance of some important cross sections. It can be noticed that in general thermal cross sections exhibit a higher variance than fast cross sections.

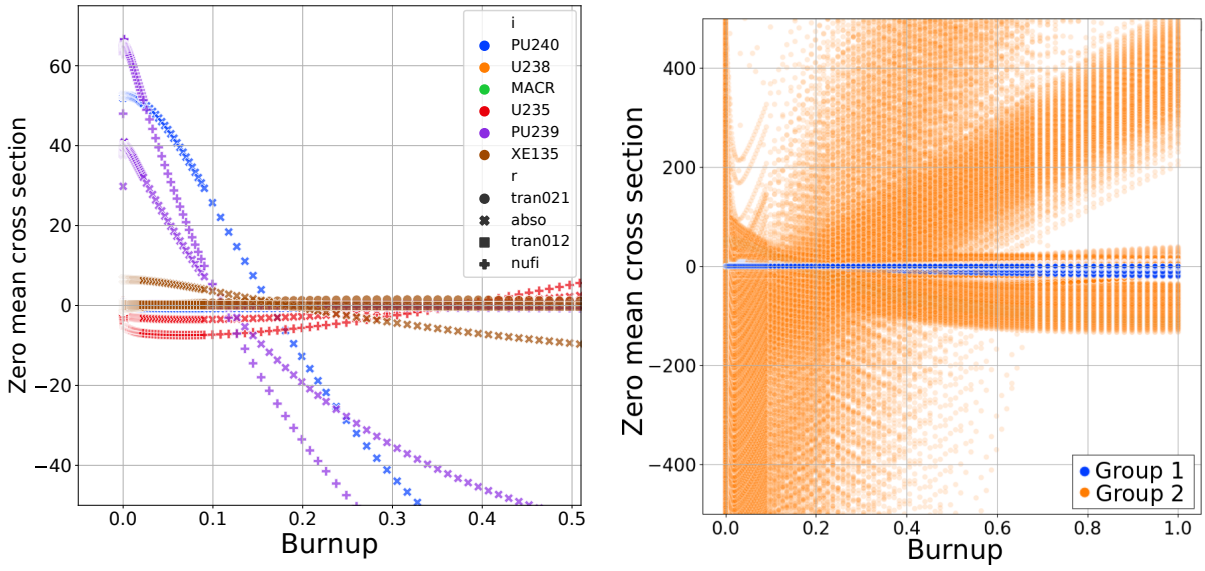
The variance is presented in Table 2.2 for some important cross sections. It can change significantly between isotopes and specially between groups. In Fig. 2.6a several more



(a) Σ_{res} (b) Σ_{res}

Figure 2.5: Zero mean cross section projections for Σ_{res} in using entire data set ($d = 3$) discriminated by reaction and group. (a) Significant variations are observed for the absorption reaction. Each line includes all the data for the fuel temperature which has a mild spectral effect in the moderator's absorption. In (b) the dependence of the moderator's absorption cross section with boron concentration is apparent since this neutron poison is diluted in the water.

isotopes are plotted for $\sigma(Bu, T_{f,0}, C_{B,0})$ to have a clearer figure (i.e. $d = 1$). Significant variations within the first burnup values are noticed, specially for the plutonium element. In [30] it's suggested that these strong variations are due to passing from an initial zero concentration infinite-diluted high cross section values to self-shielded values. Also, as the isotopic content of the fuel changes, the spectrum of the collapsing flux varies as well. Difference between the fast and thermal group are further exhibited in Fig. 2.6b by plotting the entire data set ($d = 3$) discriminated by energy group. Fast cross section have quite smaller values and are thus relatively more sensitive to possible numerical artifacts.



(a) (b)

Figure 2.6: Cross sections variation with burnup discriminated by isotope, reaction and group. (a) Cross section discriminating by reaction and isotope for $\sigma(Bu, T_{f,0}, C_{B,0})$. Absorption and fission cross section tend to present the highest variance. (b) Entire cross section data set discriminated only by energy group. Higher absolute values and variance are observed for thermal cross sections.

2.5 Final thoughts and discussion

This chapter presented the use case to assess the approximation methodologies detailing the calculation settings and the composition of the input and output space. The main conclusions are:

- A standard PWR fuel assembly composed of 17×17 UO_2 fuel rods with 4% enrichment w/o ^{235}U [60] was presented for classical lattice calculation settings using a full grid sampling of the domain.
- The output space considered counts with 29 specialized isotopes from which $\mathcal{Y} = 144$ functions are retained for approximation. This kind of isotope specialization can be found in criticality safety analysis [65] and is not very different to those utilized in state-of-the-art industry applications [23]. The reactions considered are $\{\sigma_{a1}, \sigma_{a2}, \sigma_{f1}, \sigma_{f2}, \sigma_{1\to2}, \sigma_{2\to1}\}$.
- A three dimensional ($d = 3$) input space consisting of the state-parameters burnup, fuel temperature, and boron concentration with plausible utilization ranges is considered.
- Though a relativity low dimensionality input space is used [5], it is sufficiently densely

[5]Difficulties were encountered in cross section data generation. When scheduled, the APOLLO3 code was in development and the APOLLO2 code was used instead. Auxiliary executable programs were provided to read the cross section data files (sbr.exe) that, after debugging, still required about 100 seconds to extract each calculation point for large files. These issues are being process limiting the scope of the considered dimensions to only 3.

sampled to extract two distinct test sets which are independent of the pool used to build the approximations.

- The data, pool, and test set were shown together with some support examples. The Industry set was defined standing as an additional element of comparison between the cross section models.

- The performance metric, i.e. the cross section error definitions were presented emphasizing the need to consider relative errors though discussing possible pitfalls for fast cross sections. Errors discriminated by isotope, reaction, or group can be used to depict the error distribution within the set and reveal a possible error compensation.

- A sensitivity coefficient on the multiplication factor due to cross section error was presented. The target relative errors discussed in chapter 1, laying between [0.1 %, 0.01 %], imply a maximum error of 10 *pcm* while an 1 % error in $\sigma_{235,f,2}$ can lead to a multiplication factor error of 1300 *pcm* which is unacceptably big due to safety concerns. An error of 0.01 % can be really small of only 0.5 barn which is why very high accuracy models are needed.

- A definition of cross section importance was provided, showing that only a few cross sections have a large participation in the macroscopic cross section. This is especially the case for scattering reaction that can require large portions of the library size.

- Cross section variance and projections in the state-parameters were presented showing the variability within $\mathcal{Y}$ for different isotopes, reactions, and groups. In particular a significant part of the function's variance is located in the burnup dimension and specifically in the first burnup values.

- Thermal cross sections exhibit larger variances than fast cross section data, especially for the absorption and fission cross section on the burnup dimension. The other exhibit a "low order", yet non linear, dependencies and some insights of the underling physical phenomena governing this behaviour were presented.

Chapter 3

Spline interpolation

In this chapter a review of global and local interpolating approximations using univariate expansion basis is carried out showing its benefits and limitations. Piece-wise polynomial interpolation is analyzed in detail specifically for splines utilizing a B-spline basis, which is a classical approximation technique representing the state-of-the-art in many industrial implementations. Some aspects of this type of approximations such as the knot vector construction and possible oscillations in the approximation are examined setting the ground of comparison for other, more innovative methodologies.

3.1 Interpolating approximation using univariate basis functions

An approximation $\hat{y} \simeq y$ (y is the function being approximated, the cross section σ of Eq. 1.14) can be built as a linear combination of the tensor product of univariate basis functions $\psi(x_i), x_i \in \vec{x}, 1 \leq i \leq d$

$$y(\vec{x}) \simeq \hat{y}(\vec{x}) = \sum_{j_1=1}^{N_1} \cdots \sum_{j_d=1}^{N_d} \alpha_{j_1,\ldots,j_d} \prod_{i=1}^{d} \psi_{i,j_1,\ldots,j_d}(x_i), \tag{3.1}$$

being d the dimension of the input space. For finding the coefficients $\vec{\alpha}$ which are the parameters of the model, a support coming from the pool $\mathcal{S} \in \mathcal{P}$ is used to impose the interpolation condition

$$\hat{y}(\vec{x}_j) = y(\vec{x}_j), \quad 1 \leq j \leq |\mathcal{S}|. \tag{3.2}$$

In view of Eq. 3.1 the *collocation system* is

$$A\vec{\alpha} = \vec{y}, \tag{3.3}$$

where $A_{k,j} = \prod_{i=1}^{d} \psi_{i,j_1,\ldots,j_d}(x_{i,k})$ with $1 \leq k \leq |\mathcal{S}|$ and the basis function with the index $1 \leq j_i \leq N_i$ for $1 \leq i \leq d$. Using a tensor product rule for the support the total amount of basis functions, equal to the amount of coefficients, is $\prod_{i=1}^{d} N_i = \prod_{i=1}^{d} |\mathcal{X}_i|$; thus accounting for all possible cross-terms between the variables, effectively treating y as a "black-box". A vast amount of "classical" representations methodologies can be found under this scheme. However, "The major limitation of the methodology described is the requirement for the function to be defined everywhere in the rectangular problem domain. As a result, it is not applicable in the cases where there are unphysical domains of state parameters" [31].

3.2 Global polynomial representation

Lets start with a 1 dimensional case for the use case considered in this work corresponding to cross sections in the form $\sigma(Bu, T_{f,0}, C_{B,0})$. Polynomials can be evaluated, differentiated and integrated easily using basic arithmetic operations. A *global polynomial* is defined as

$$\hat{y}(x) = p(x) = \sum_{j=1}^{N} \alpha_j \psi_j(x). \tag{3.4}$$

They are a very well understood function space and the building blocks of more advanced representation methods. The polynomials p of order N and degree $N-1$ form the linear space $p \in \Pi_{<N}$ from which a unique polynomail that interpolates the support $\mathcal{X}_{\mathcal{S}}$ can be extracted [68].

Possibly the simplest base example is the monomial base composed of $\psi_j(x) = x^{j-1}$ for $1 \leq j \leq N$. This basis, used in [56], leads to impractical full rank matrices to invert in Eq. 3.3. A more convenient basis in this regard is the Lagrange base $\psi_j(x) = \prod_{k=1,k\neq j}^{N} (x-x_k)(x_j-x_k)$, $1 \leq j \leq N$ defining a trivial collocation system since $\psi_j(x_k) = \delta_{jk}$. This property was exploited in [28] to facilitate the independent addition of new points to the system. Unfortunately this base also entails a relativity expensive evaluation effort[1].

The polynomial interpolation package used in this work [69] utilizes the Newton base defined by $\psi_j(x) = \prod_{k=1}^{j-1}(x - x_k), \quad 2 \leq j \leq N$, $(\psi_1(x) = 1)$. The divided differences algorithm can be used to find $\vec{\alpha}$ and the Horner's scheme to evaluate $\hat{y}$ in a particullary efficent way[2].

[1]Rejection techniques can be used due to the orthonormal character of the Lagrange base. This technique is exploited in many works but attention must be drawn to the points actually defining the base as the productory in Lagrange base is not changed by shrinking and low evaluation speed of $\hat{y}$ could be encountered. The basis evaluation cost is $(2N-2)A + (N-2)M + (N-1)D$ operations where A, M, D stand for addition, multiplication and division respectively.

[2]The divided differences algorithm has a computational cost of $N(2N-1)A + N(N-1)D/2$ operations [68] and the Horner's scheme profiting from a nested expression of $\hat{y}$ a cost of $(N-1)(2A+M)$ operations [70].

3.2.1 Limitations of global approximations

Global polynomials are efficient solutions for "small" problems but struggle when increasing the size of the support, especially for non-smooth functions. A function y badly behaved anywhere can badly condition $\hat{y}$ everywhere [68]. A bound on the global polynomial approximation is $||p|| \leq \lambda ||y||$ with the uniform norm defined by

$$||y|| = \max_{x \in \mathcal{X}}(\text{abs}(y(x))). \tag{3.5}$$

If, for example, the support is sampled using an uniform distribution as done in [53], the Lebesgue constant is $||\lambda|| \sim 2^N/(e\ln(N))$ scaling quite badly with N. A Chebyshev distribution for sampling the support mitigates this to some extent [68], with a Lebesgue constant of $||\lambda|| = ||\lambda^c|| \sim (2/\pi)\ln(N) + 4$ which can be convenient, even at the expense suboptimal burnup grid as usued in [28, 42]. Nonetheless the order of the polynomial approximation is not the only caveat. A bound to the approximation error is

$$||y - \hat{y}|| \leq (1 + ||\lambda^c||)\text{dist}(y, \Pi_{<N}),$$

with the term distance to the global polynomial function space bounded by

$$\text{dist}(y, \Pi_{<N}) = \min_{p \in \Pi_{<N}} ||y - \hat{y}||.$$

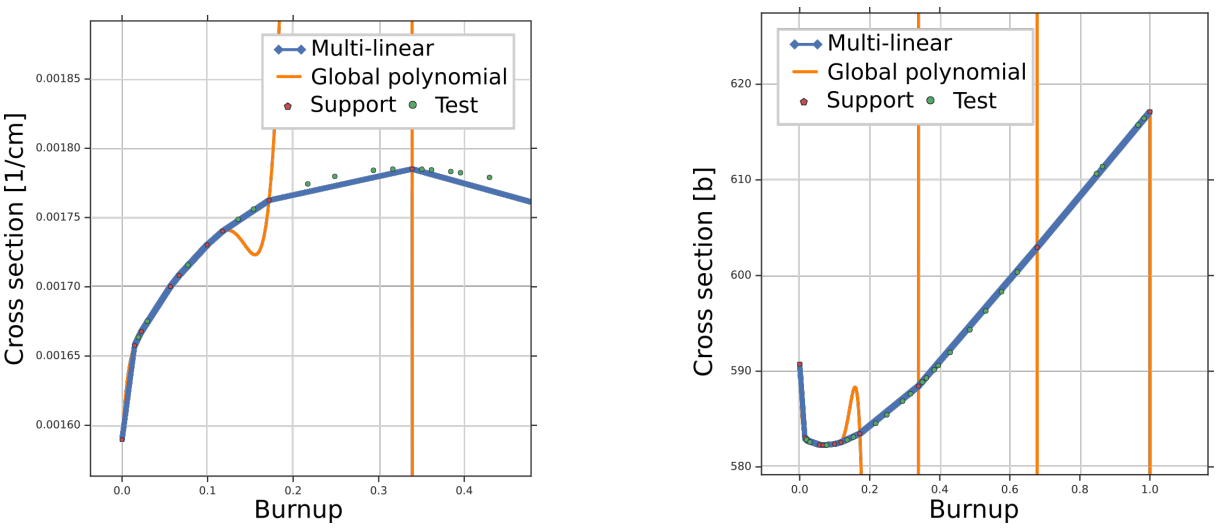
Considering an approximation using Chebyshev nodes results in

$$\text{dist}(y, \Pi_{<N}) \leq 2\left(\frac{1}{4}\right)^N \frac{||y^{(N)}||}{N!}$$

meaning that for non-smooth functions the error may still grow at the rate of $||y^{(N)}||$. These two limitations related to the size of the support ($|\mathcal{X}_\mathcal{S}|$) and the variations of the function being approximated ($||y^{(N)}||$) may render global polynomials inadequate for cross sections modeling. A mere $|\mathcal{X}_\mathcal{S}| \sim 10$ can be sufficient to triggering un-bounded oscillations as shown in Fig. 3.1 (Runge phenomenon [68]). In fact, high values of $||y^{(N)}||$ near the border of the domain, and possible large $|\mathcal{X}_\mathcal{S}|$ with non-uniform distributions, are the norm rather than the exception.

3.3 High order piece-wise polynomial approximations

Some of the problems analyzed in the previous section can be overcome by subdividing the domain and approximating the resulting pieces with lower order polynomials. These *piece-wise polynomials* noted as *pp*, are composed of l polynomials of order $K \ll N$ and joined together at the breaks ($\vec{\xi} \in \mathcal{X}$) that form the strictly increasing sequence $|\vec{\xi}| = l + 1$. *pp* are *local* approximations as only data belonging to the sub-domains defined at the breaks are used in each piece.



(a) (b)

Figure 3.1: Divergent global polynomial approximation in cross section modeling using only support 11 points marked in red that defines the breaks ($\vec{\xi}$) of the piece-wise approximation. A local linear approximation on the other hand, is bounded by the support. (a) $\Sigma_{res,1\to 2}$ exhibit unbound oscillation due to the Runge phenomenon. (b) A divergent error is observed for $\sigma_{235,f,2}$. The first point corresponding to a zero concentration of ^{135}Xe badly conditions the system locally affecting the global approximation everywhere in the domain.

These functions define the space $pp \in \Pi_{<K,\vec{\xi}}$ of dimension Kl that can be thought of as the direct sum of l pieces of $p \in \Pi_{<N}$. Approximations need not to be smooth or continuous at the breaks, though such level of generality is not usually desired in many application including cross section modeling. A classical example of piece-wise approximation is linear interpolation of order $K = 2$ (and degree 1). As only continuity is imposed, the approximation is local though not smooth. Points are connected by straight lines and the first derivative may not exist at the breaks as can be observed in Fig. 3.1. Higher order *pp* approximation allow to impose additional smoothness conditions ideally requiring less points to describe a smooth curve. A common term to refer to these type of curves, specially when showcasing a "clean" and "smooth" profile without wiggles is *splines* in relation to the tools used in naval engineering at the beginning of the XXth century.

Smooth conditions at the breaks are formally defined with the aid of the continuity vector $\vec{\nu}$ that expresses the amount of derivatives that exist at the breaks. This is a piece-wise polynomial sub-space $\Pi_{<K,\vec{\xi},\vec{\nu}} \subset \Pi_{<K,\vec{\xi}}$[3]. Its construction and evaluation is achieved in a stable and efficient way by using B-splines where "B" stands for basis.

[3]The continuity vector imposes the existence of the derivative up to $(\nu_j - 1)$ imposing that no "jump" occures at the breaks: $jump(\hat{y}(\xi_j)^{(k-1)}) = 0$, for $1 \leq k \leq \nu_j$ and $2 \leq j \leq l$. Finding a proper base for $\Pi_{<K,\vec{\xi},\vec{\nu}}$ results in solving a smaller collocation system than in $\Pi_{<K,\vec{\xi}}$ as in the latter the additional (homogeneous) continuity conditions must be included. In fact, for different supports the solutions in the same functional space $\Pi_{<K,\vec{\xi}}$ require solving the very same continuity conditions repeatedly (see page 80 of [68]).

3.3.1 B-splines as basis of a piece-wise polynomial space

Let $K, N \in \mathbb{Z}_{>1}$ and the non-decreasing *knot vector* $\vec{t}$ with $t_j \in \mathbb{R}, 1 \leq j \leq N+K$. Then, a B-spline ($B_{j,K}$) is a piece-wise polynomial of order K defined by the recurrence relation

$$\begin{gathered} B_{j,K} = w_{j,K} B_{j,K-1} + (1 - w_{j+1,K}) B_{j+1,K-1}, \quad w_{j,K} = \frac{x - t_j}{t_{j+K-1} - t_j}, \\ B_{j,1} = \begin{cases} 1, & t_j \leq x \leq t_{j+1}, \\ 0, & \text{otherwise.} \end{cases} \end{gathered} \tag{3.6}$$

Some of these basis functions are shown in Fig 3.2a. The smoothness or regularity of each basis term increases with the order K. A spline is a linear combination of B-splines that can be used to approximate y

$$y(x) \simeq \hat{y}(x) = s(x) = \sum_{j=1}^{N} \alpha_j B_{j,K}(x), \tag{3.7}$$

with $\alpha_j \in \mathbb{R}$ forming the linear space $s \in \$_{K,\vec{t}}$. The connection between the spline space $\$_{K,\vec{t}}$ and $\Pi_{<K,\vec{\xi},\vec{\nu}}$ is provided by the Curry and Schoenberg theorem presented in the Appendix A.1. B-splines are characterized by:

- *Positivity*: $B_{j,K} > 0$ if $x \in [t_j, t_{j+K}]$.
- *Locality*: $B_{j,K} = 0$ if $x \notin [t_j, t_{j+K}]$.
- *Partition of unity*: $\sum_{j=1}^{N} B_{j,K} = 1$ if $x \in I_{K,\vec{t}} = [t_K, t_{N+1}]$.

These properties make of B-splines an extremely convenient basis for representing functions. As shown in Fig. 3.2b they define positive banded matrices for the collocation system in Eq. 3.3 for which very efficient resolution routines are available [71]. Being the basis local and having partition of unity the approximation is bounded by the coefficients (Convex hull property)

$$min(\alpha_{j+1-K}, \ldots, \alpha_j) \leq s(x) \leq max(\alpha_{j+1-K}, \ldots, \alpha_j), \tag{3.8}$$

for $t_j \leq x \leq t_{j+1}$. For this reason the coefficients are sometimes referred to as "control points", and a total of only K participates in the evaluation of $s(x)$[4]. Due to the bounded nature of B-splines and the very well conditioned collocation matrices the coefficient's module have values that reflect the support's data. Depending on the field the word "splines" may refer to parametric curves consisting of a set of spline functions in the form ($s_1(\theta)$,$s_2(\theta)$,...).

[4]Just as it would be the case in a piece-wise representation $\Pi_{K,\vec{\xi},\vec{\nu}}$ that requires the evaluation of a polynomial of order K.

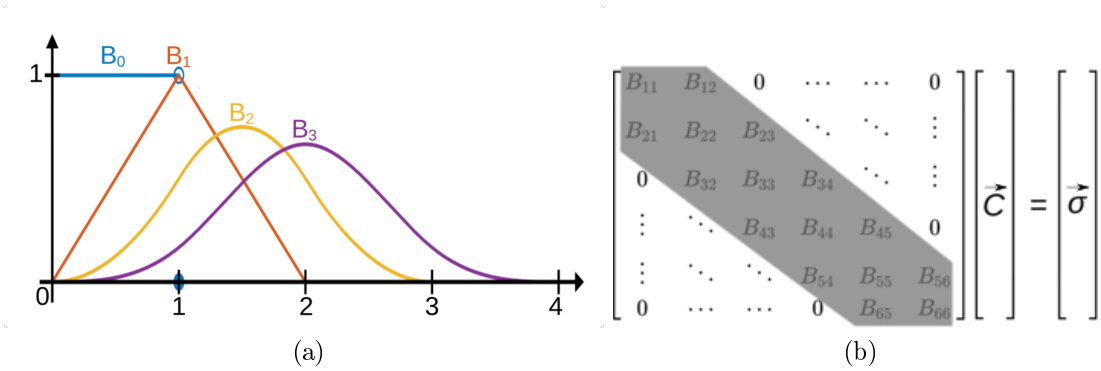


Figure 3.2: Properties B-splines basis. (a) The order of the spline defines its regularity. This local basis has non-zero values in a small region of the domain defined by the order K. (b) The properties of locality, positivity, and partition of unity positive banded matrices in the collocation system for which very efficient routines are available.

3.3.2 Interpolant piece-wise approximations with B-splines

An approximation $s \in \$_{K,\vec{t}}$ is a solution of Eq. 3.2 for a support defined over $\mathcal{S}$ by the fulfilment of Theorem 1 which establishes two condition on $\vec{t}$. First, the maximum multiplicity r, the repetitions of an element in the knot vector, for any interior knot must be $r < K - 1$, only the first and last knots are allowed to have $r = K$. The multiplicity defines the order up to which the derivative exists at the breaks. A total amount of interior knots $|I_{K,\vec{t}}| = |\vec{t}| - 2K = N - K$ is defined from $|\mathcal{S}| - 2 = N - 2$ unused support points. The interplay between these two, the data for interpolation, and the location of the breaks with their corresponding multiplicity define the quality of the spline approximation.

Theorem 1 (Schoenberg and Whitney theorem [68]). Let $\mathcal{X}_{\mathcal{S}}$ be strictly increasing sequence and $0 < t_j = \cdots = t_{j+r} = x_j < 1$ implies

$$r < K - 1 \tag{3.9}$$

then the system in Eq. 3.2 is invertible if and only if $B_{j,K}(x_j) > 0$ which happens if

$$t_j < x_j < t_{j+K} \text{ for } 1 \leq j \leq N. \tag{3.10}$$

$x_j = t_j$ is allowed for $j = 1$ and $j = N$ regardless of condition 3.9 and 3.10.

3.4 Knot vector constructions

Smooth approximations increase cross section accuracy and are particualrry relevant in core feedback coefficients calculations. Several recepies for computing valid knot vectors from a given support are presented in [68].

Linear approximation

For $K = 2$, $r = 0$ meaning that every interior knot can appear only once thus imposing a single condition at the breaks: continuity. $I_{2,\vec{t}} = N$ and therefore

$$t_{j+2} = x_{j+1} \text{ for } 1 < j < N, x \in \mathcal{X}_{\mathcal{S}}, \tag{3.11}$$

thus $\vec{t} = (0, 0, x_2, \ldots, x_{N-1}, 1, 1)$. A solution to Eq. 3.3 is trivial, $\alpha_j = y(x_j)$ for $1 \leq j \leq N$, and due to the convex hull property the local interpolant approximation is bounded by the support. Let y have two continuous derivatives ($y^{(2)}$), an error bound is

$$||y - \hat{y}|| \leq \frac{1}{8}||\Delta\mathcal{X}||^2||y^{(2)}||, \tag{3.12}$$

which is of $\mathcal{O}(N^{-2})$. A support increase leading to a reduction of the interval in the support $||\Delta\mathcal{X}||$, improves the accuracy without changing the complexity of $\hat{y}$ which are straight lines. It does increase the total amount of points required by the approximation[5].

Not-a-knot spline

For $K > 2$ interior knots of multiplicity equal to 1 correspond to a smooth approximation at the breaks being a solution of Eq. 3.2 whose existence is assure by the Schoenberg and Whitney theorem. The amount of interior knots $t_{K+1}, \ldots, t_N$ is smaller than the amount of support points. One strategy for computing $\vec{t}$ is, for an even order $K = 2m$ with $m \in \mathbb{Z}_{\geq 1}$ the assignation

$$t_{K+j} = x_{m+j}, \quad 1 \leq j \leq N - K. \tag{3.13}$$

It can be noticed that indeed m data sites do not appear in the knot vector. For $K = 4$ this is the cubic spline with the "not-a-knot" end condition.

Oddend spline

If K is odd an assignation of the form 3.13 can be consider by dropping an additional data sites at either extreme of the interior knots. Dropping the last knot is considered.

Splop spline

A so called optimal or "crowned" knot vector is obtained by using the routine of the same name found in [68][6]. Another knot vector, similar to splop is given by

$$t_{K+j} = (x_{j+1} + \cdots + x_{j+K-1})/(K-1), \quad 1 \leq j \leq n - K. \tag{3.14}$$

which is called here Pre-Splop.

[5] In the particular case of a uniform mesh the classical error bound for linear interpolation is obtained $||y - \hat{y}|| \leq \frac{1}{8}\left(\frac{b-a}{n-1}\right)^2 ||y^{(2)}||$ with in this case $b = 1$ and $a = 0$.

[6] The motivation is to minimize the term $(1 + ||\hat{y}||)$ of the error bound presented detailed later in Eq 3.17.

Approxima- -tion type	MAX_σ	AV_σ	MAX_Σ	AV_Σ	MAX_{k_∞}	AV_{k_∞}	Library Size	
	[%]				[*pcm*]		Coef.	Δ[%]
Multi-linear	7E-1	1.25E-2	8E-2	1.1E-2	2.0	4.7	5E3	-
Not-a-knot	5E-1	2E-3	2E-2	6E-4	2.0	0.5	2.2E3	56
Splop	5E-1	2E-3	2E-2	8E-4	2.0	0.6	2.2E3	56
Pre-plop	5E-1	2E-3	2E-2	8E-4	2.0	0.5	2.2E3	56
Odd-end	4E-1	3.5E-3	2E-2	1.8E-3	4.1	1.1	2.9E3	42

Table 3.1: Approximations for $d = 1$ with different spline types differing in the order and knot vector construction strategy. Maximum and average errors are calculated at the support corresponding to the industry set. The library size and corresponding reductions are calculated at the constant average error ($AV_{\sigma,\star}$) of multi-linear.

3.5 Spline cross section approximation

In this section the different model's errors for piece-wise approximation using linear and higher order splines are presented. In this work we consider five types of spline approximation: Linear ($K = 2$), Not-a-knot ($K = 4$), Odd-end ($K = 3$), Splop ($K = 3$), and Pre-plop ($K = 3$).

3.6 Spline error analysis for $d = 1$

Average and maximum cross section errors (see Eq. 2.5 and Eq. 2.7) are presented in Fig. 3.3a and Fig. 3.3b in function of the library size (total amount of coefficients). For the linear spline (order $K = 2$) the mean error monotonously diminishes when increasing the support. Higher order interpolation tends to perform better than linear interpolation: a smaller support is required at a given error (e.g. at $\mathrm{AV}_{\sigma,*}$ or 1E-2%), or equivalently a better accuracy is obtained for a given library size. Different knot vectors offer the same representation quality with the exception of oddend that has a somewhat higher error[7]. Plots with the same overall error tendencies are obtained for k_∞ and thus are not shown. Errors for the industry set, which was first introduced in Section 2.2.2, are presented in Table 3.1.

Even for this rather simple 1-dimensional case, higher order splines maximum or average errors do not decrease monotonously. If the order of the spline is further increased this problem gets exacerbated as shown in Fig 3.3d. With a low support, all cross sections approximations have a similar poor accuracy which can be observed in both Fig. 3.3a and Fig. 3.3a. By comparing them it can be noted that the error profile of linear approximations are similar in σ and Σ. This is not the case for high order spline approximations, when increasing the support, since $\mathrm{AV}_\Sigma < \mathrm{AV}_\sigma$. In order to find the roots of this behavior it's necessary to further analyze the cross section errors.

[7]The knot vector construction for $k = 3$ dropping a data site at the beginning instead of at the end performed poorly having even divergent errors. A higher order for the odd-end knot vector resulted in a divergent approximation. It's therefore not presented nor discussed further.

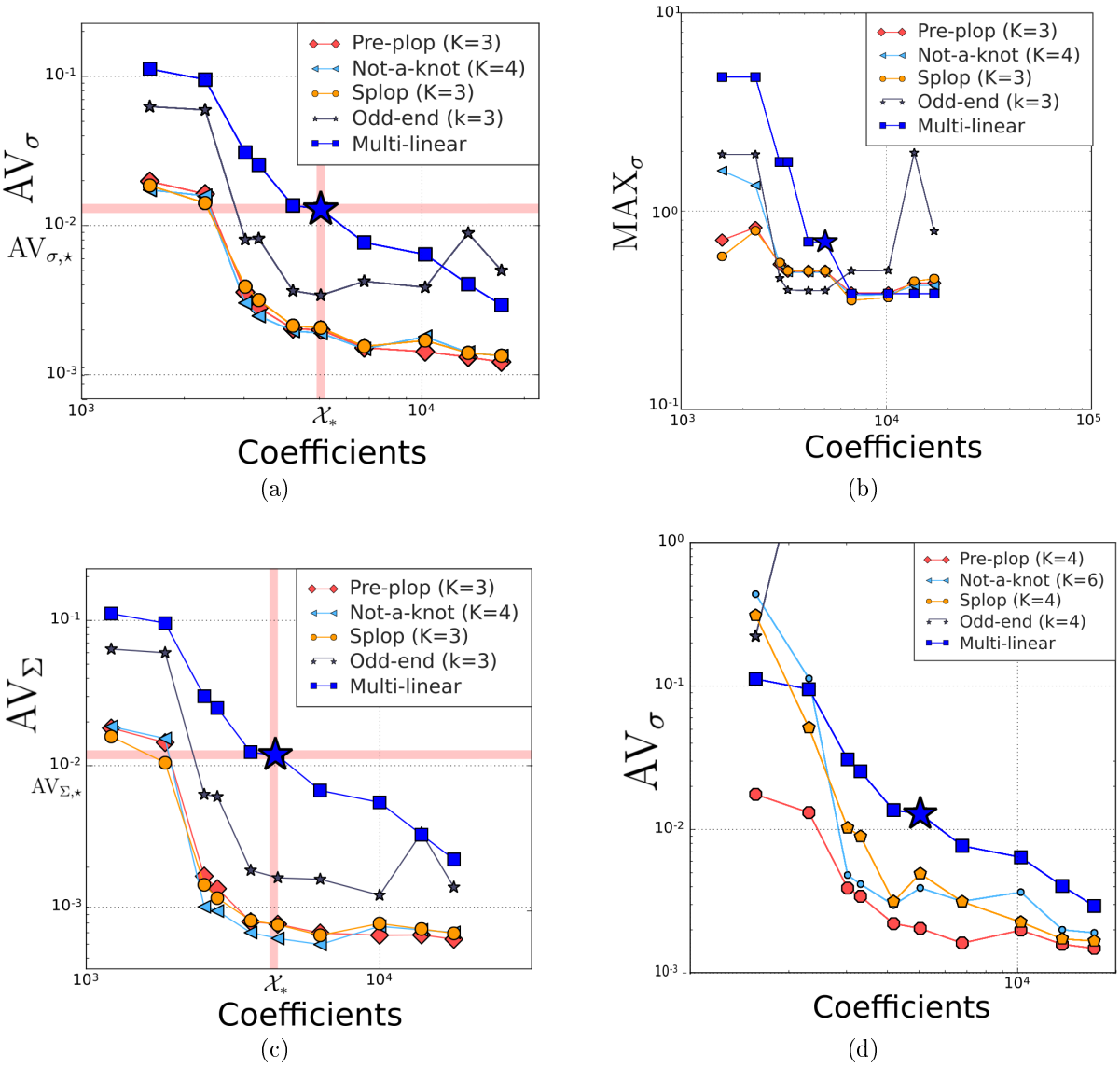


Figure 3.3: Error profiles for different spline approximation for $d = 1$. (a) Average error in σ. Multi-linear error diminishes monotonously and the Industry set error is withing the target boundaries. Higher order splines exhibit a higher accuracy. (b) Maximum error in σ. Higher order splines can present larger errors than multi-linear. (c) Average error in Σ. Multi-linear error diminishes monotonously and the Industry set error is withing the target boundaries. Higher order splines perform better in this case than with σ. (d) Higher order splines with additional smoothness conditions. The performance deteriorates. Straight line helps to visualize the trend.

3.6.1 Non-monotonically error behavior high order splines

To further explore the non-monotonously decreasing error with support in higher order splines, the error RE_σ for the Industry set is presented in Fig. 3.4. This is an information dense figure, but helps to pinpoint the sources of the error. In the horizontal axis reaction type per group are located and the specialized isotopes in the vertical. Some cross sections present a much higher error that the average as for example, $\sigma_{240,a,1}$ that participates significantly in Σ (see Fig. 2.3a).

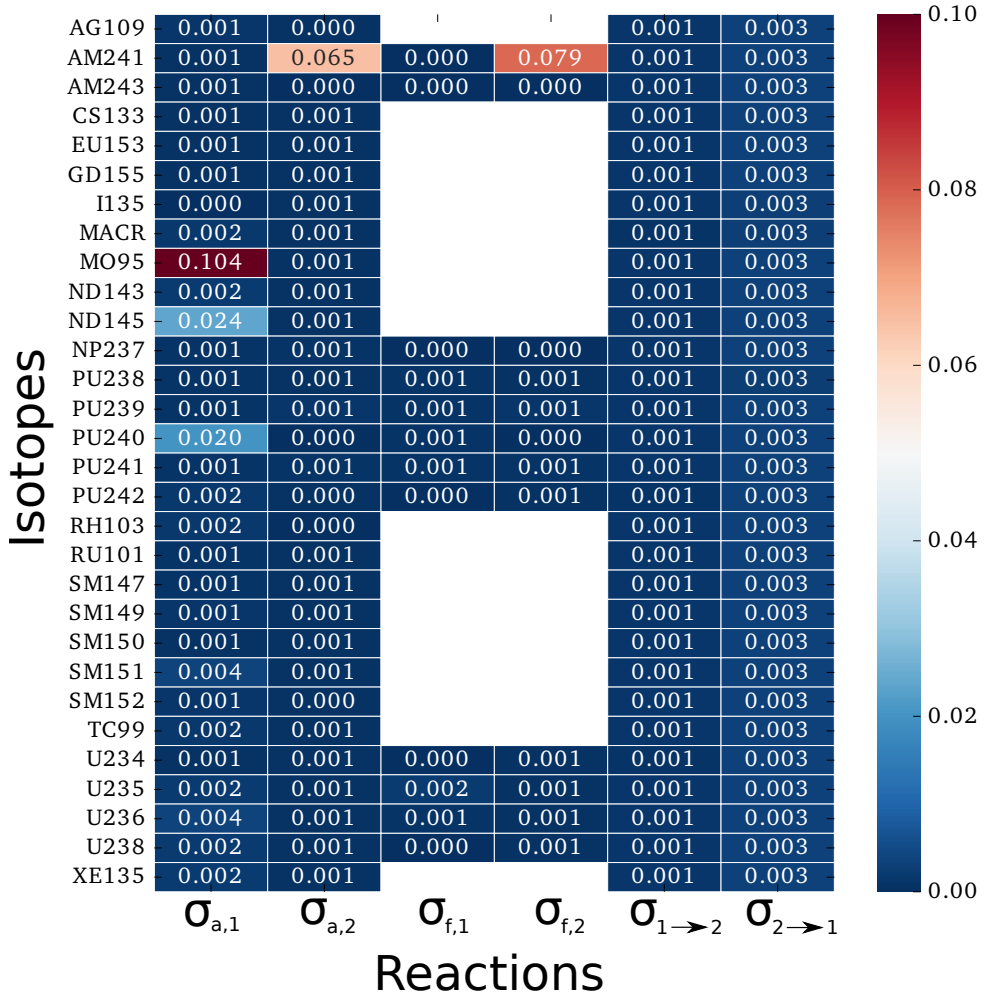


Figure 3.4: RE_σ for the oddend spline with the support $\mathcal{X}_\star$ and $d = 1$. Though the majority of the cross sections are well predicted having a relative error value smaller than 0.05%, others present large average relative errors.

Some of the cross sections with the highest error are plotted in Fig 3.5. It can be seen that abrupt changes in the cross section values for the very first burnup points triggers spline overshoots and oscillations. This problem does not affect the linear interpolation, bounded by the support. Near $\mathcal{S}_\star$ the support is large enough to include points in this difficult regions, inducing the oscillations, but sufficient to limit them. Since this happens for low burnup values, where many isotopes have near zero concentrations, $\mathrm{AV}_\Sigma < \mathrm{AV}_\sigma$ as noticed in Fig. 3.3a.

It could be argue that these errors are small and maybe globally inconsequential. Indeed they are on the order of ~ 1 *barn* and, at least for a 1-dimensional case, a library reduction is obtained with higher order splines. To this, it's important to note that in view of isotopes concentrations, actual cross section values, and spectral conditions it's actually hard to judge the aggregated effect of such nonphysical modeling errors in for example, the integral or feedback parameters of core calculations. The onset of oscillations leading to non-trivial dependence with the support in the region of interest are unexpected and hard to predict. Though small in absolute values, these overshoots are of the same order of magnitude than the spectral effect ^{135}Xe concentration change. In the Appendix A.1 it's shown that further increase in the support does help, though performance varies depending

on the test set used. This type of oscillation problems have been mentioned before [46], but not investigated in detail for cross section approximations. From a modeling point a view, the need of regularization is evident which will be treated in the next chapter.

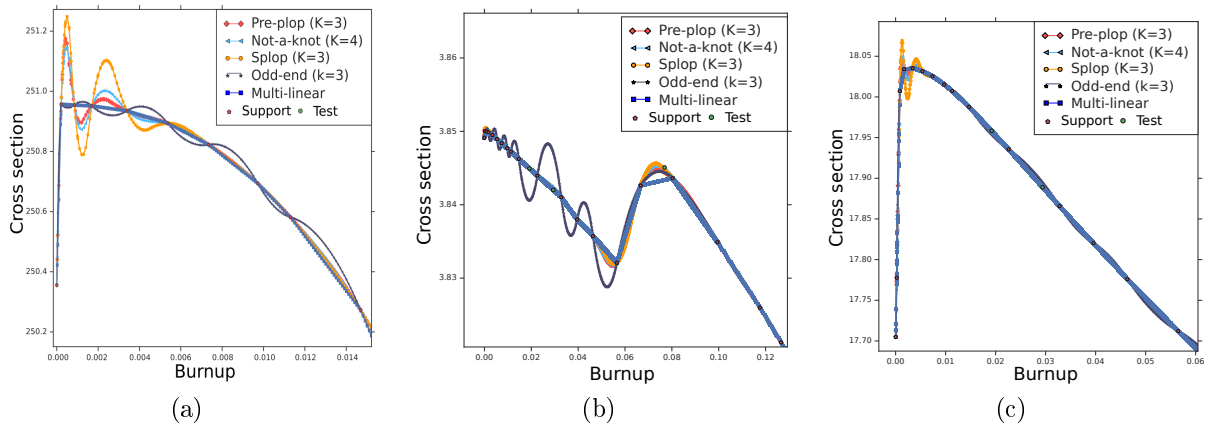


Figure 3.5: High order spline approximation for difficult cross section. Nonphysical oscillations of the same order of magnitude that the spectral effect of the Xe transient can be observed. The cross sections presented are $\sigma_{240,a,1}$, $\sigma_{95,a,1}$, $\sigma_{241,f,2}$ in (a), (b), and (c) respectively.

3.6.2 The drive of spline oscillations

Another example is for $\sigma_{240,a,1}$ presented in Fig 3.6. The variation of cross section data could be possibly due to the activation of the self-shielding phenomenon combined with the xenon build-up and even numerical noise which can be more relevant for fast cross section. Indeed this behavior happens, more often than not, for cross section with zero initial concentration. The oscillations are driven by the smooth condition imposed at the knots. By taking a closer look, we observe that the fourth knot actually falls within the "jump"

$$\vec{t}_{pre-splop} = (0.0, 0.0, 0.0, 5.29\text{E}-4, 1.27\text{E}-3, 2.54\text{E}-3, 4.38\text{E}-3 \ldots \\ \ldots, 1, 1, 1\}. \tag{3.15}$$

Therefore by considering a double knot in this specific position only continuity is imposed and the spline's wiggles disappear. This "relaxed", hand crafted knot vector

$$\vec{t}_{pre-splop} = (0.0, 0.0, 0.0, 5.29\text{E}-4, 1.27\text{E}-3, 1.27\text{E}-3, 4.38\text{E}-3 \ldots \\ \ldots, 1, 1, 1\}, \tag{3.16}$$

generates an approximation smooth everywhere except at the "jump" where only continuity is imposed. Inspired by this example, several strategies to modified the knot vector by introducing multiplicities and thus relaxing the smooth condition have been tested and are presented in the Appendix A.2. Unfortunately the overall error improvement remains marginal.

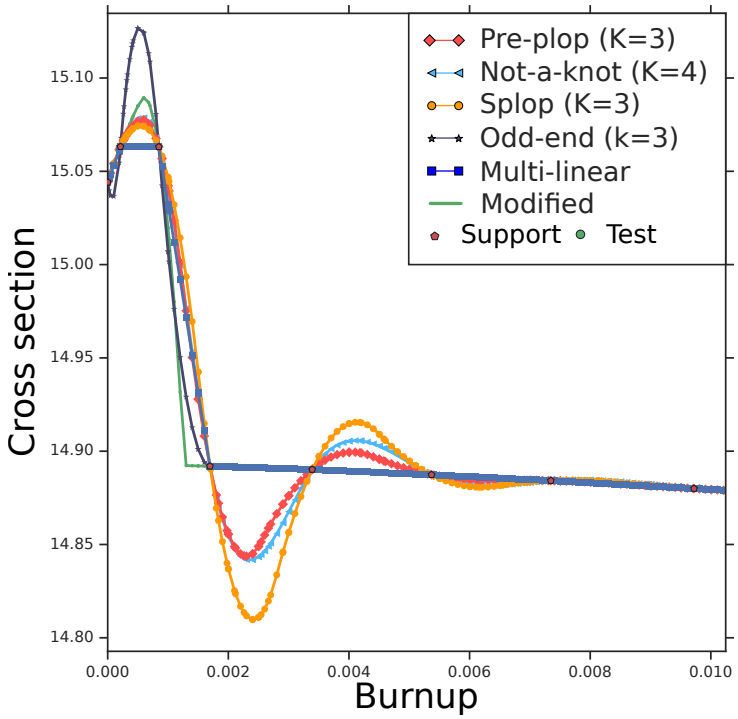


Figure 3.6: A high order modified spline approximation based on the pre-splop knot vector construction considers a double multiplicity knot at the data "jump" tempering the oscillations. This "relaxed" knot vector generates a smooth approximation everywhere except at this point in which only continuity is imposed.

3.6.3 Theoretical investigation of the spline error

A spline approximation error bound is [68]

$$||y - \hat{y}|| \leq (1 + ||\hat{y}||)\text{dist}(y, \$_{K,\vec{t}}). \tag{3.17}$$

Let y have K continuous derivatives, then the distance of the function to the spline function space is bounded by

$$\text{dist}(y, \$_{K,\vec{t}}) \leq \text{const}_K \max(\Delta t)^K ||y^{(K)}||, \tag{3.18}$$

which means that the best possible (though indeterminate) spline approximation $s \in \$_{K,\vec{t}}$ to a smooth function can at best go to zero as fast as the K-th power of the mesh size. Though this type of bound can profit from the approximation's order and the discretization of $\mathcal{X}$, significant variation in the data, specially for higher orders, can compromise the error bound. The choice of the data sites not only impacts the bound in Eq. 3.17 by defining the information available to $\hat{y}$ but also in view of the following bound

$$||\hat{y}|| \geq \text{cte}_K \min_{x_i \in \mathcal{X}_S} \frac{\{t_{j+K-1} - t_j : (t_j \ldots t_{j+K-1}) \bigcap (x_i \ldots x_{i+1}) \neq \emptyset\}}{\Delta x_i}. \tag{3.19}$$

The numerator grows as the intersection between the knot vector and grid in which the support is defined augments. The denominator diminishes with increasing support

Approxima- -tion type	MAX_σ	AV_σ	MAX_Σ	AV_Σ	MAX_{k_∞}	AV_{k_∞}	Library Size	
	[%]				[pcm]		Coef.	Δ[%]
Multi-linear	1.9	2.1E-2	3E-1	1.8E-2	90	15	1.5E5	-
Not-a-knot (Mod.)	1.9	1.4E-2	2E-1	1.0E-2	80	12	1.1E5	26

Table 3.2: The high order approximation considered a modified not-a-knot spline for the burnup dimension. Maximum and average errors are calculated at the support corresponding to the industry set. The library size and corresponding reductions are calculated at the constant average error ($AV_{\Sigma,\star}$) of multi-linear.

cardinality, having smaller minimum intervals. Therefore irregularity in the sampling of the support can both increase the numerator and decreace the denominator augmenting the error bound. Indeed with increasing $||\hat{y}||$ "the approximation process becomes less local. This means that the error at one part of the domain may very well depend on the function y anywhere. [...] If this is detected it makes no sense to keep adapting the knot vector to the irregularities of y"[68]. This explains the difficulties observed with splines exacerbated by the order K, the variation in σ and irregularities in the sampling of $\mathcal{X}$ as empirically observed. Indeed the optimal order K is not evident and any $K > 4$ showed very poor accuracy.

3.6.4 Multi-dimensional spline approximation in a Cartesian grid

The advantage of a Cartesian support is that the multi-dimensional collocation system can be expressed in the form of

$$(A_1 \otimes \cdots \otimes A_d)\vec{\alpha} = \vec{\sigma}, \tag{3.20}$$

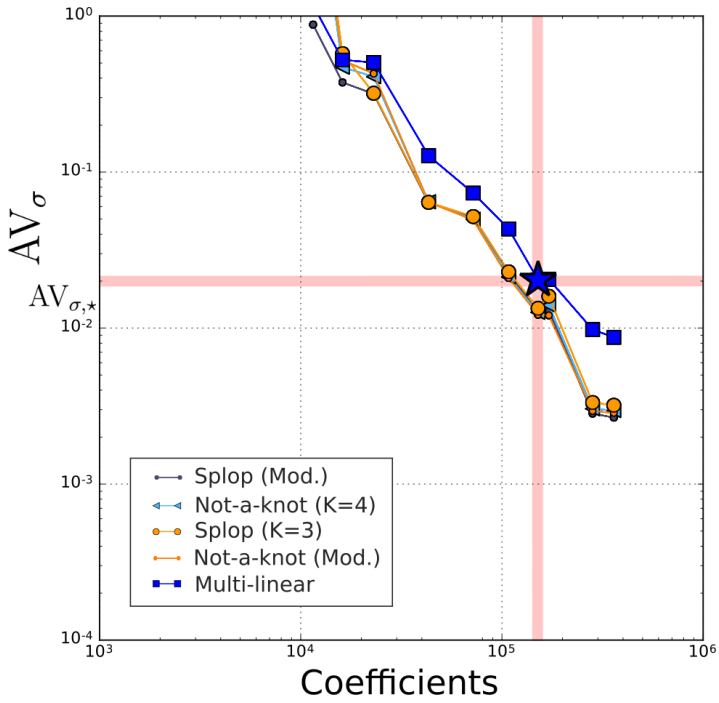
with A_i a 1-dimensional collocation system (see Eq. 3.3) which can be solved by a repeted application of their corresponding and possibly optimized solvers [71]. This is exploited in the the package used in this thesis [69]. It uses very efficient solvers to deal with the collocation system obtained from [72], and that have been refracted for increased performance.

3.7 Spline approximation in higher dimensional space

In Figure 3.7 a high order spline space is shown in function of the support size for the three dimensional use case ($d = 3$). Multi-linear approximation's errors diminishes monotonically, bounded by $\sigma(\vec{x})$, though not reaching zero, being the *useful* information in the pool limited, and $\mathcal{T}_{flat}$ fixed.

A higher order spline is able to offer some marginal library size reduction which is detailed in Table 3.2. However the gains remain very small. This is made worst by the

fact that a modified spline knot vector is actually being used and combine with multi-linear in the other instantaneous variables for the $\mathcal{T}_{flat}$ test case that has a lower density of points in the first burnup values. All other combinations of function spaces with or without modified knot vector are treated in the Appendix A.2 where performance deteriorates even more, specially for $\mathcal{T}_{Xe}$. The tensor product rule and the lack of regularization greatly hinders the storage reduction capabilities of these approximations. This is why research using full grids were abandoned from the '00 onward, though they still are the industrial standard.



(a) σ

Figure 3.7: AV_σ for higher order modified splines using a linear function space for the fuel temperature and boron concentration. Though some accuracy gains are observed these tend to be rather small. The test set is on $\mathcal{T}_{flat}$ straight line helps to visualize the trend.

3.8 Final thoughts and discussion

In this chapter global and local polynomial approximations were studied, the latter using high order splines with B-spline basis trying different orders and knot vectors constructions. We observed that:

- Polynomial approximations dispose of an extensive and accessible literature armed with a rich set of computational tools and libraries. The monomial, Lagrange and Newton basis, appearing in the literature review of chapter 1, were presented discussing some of their characteristics.
- Unfortunately, for cross section approximation global polynomials interpolation proved inadequate, at least without using special discretizations in $\mathcal{X}_\mathcal{S}$. Divergent

errors where obtained with just a few support points. Setting free the ^{135}Xe concentration probably augments significantly the modeling challenge as reported in [64].

- For 1-dimensional cross section modeling, spline approximation resulted in an accuracy increase. Though this is not an exhaustive study on knot vector construction, several recepies as suggested in [68] were tested in accordance to questions raised in [46]. At least for this use case these have a minor impact on the spline approximation power.

- A formal treatment of spline interpolation and its link to a B-spline basis were presented introducing the two main Theorems that govern this function space (Theo. 1 and Theo. 2, presented in Appendix A.1). Specifically giving the condition for which a $s \in \$_{K,\vec{t}} = \Pi_{K,\vec{\xi},\vec{\nu}} \subset \Pi_{K,\vec{\xi}}$ fulfills the interpolation condition $s(x_i) = \sigma(x_i) \forall x_i \in \mathcal{S}_\sigma$. This degree of formalism has not been provided before in cross section modeling even when using splines and facilitates the manipulation of the smooth conditions at the breaks.

- B-splines are indeed a convenient base with a positive banded collocation matrix and low variation in the coefficient's module of the expansion terms. Only K expansion terms are required in evaluation due to the local support property regaining a polynomial piece-wise evaluation cost.

- A tensor product rule for the support allows fo the iterative application of 1-dimensional solvers to find the coefficients of the multi-dimensional collocation system. These algorithms are readily available in efficient FORTRAN routines [72] and nicely packed into the Python package developed recently [69].

- Splines using B-splines are bounded by the coefficients (convex hull property), equal to cross section values for multi-linear interpolation ($K = 2$). On the other hand high order splines can present oscillations and large enough relative errors that compromise the accuracy in the region of interest.

- A theoretical investigation showed that though the approximations are local, the error bound increases with the order (K), the irregularity of either the support or knot vector, and the derivative of the data as empirically observed.

- Oscillation problems have been mentioned before [46]. These were unexpected and hard to predict, leading to non-trivial dependence of the error with the support in the region of interest. Equipped with the complete B-spline formalism, attempt were made to modify the knot vector by increasing the multiplicity of certain knots for the Splop and not-a-knot vector constructions. Unfortunately only a mild improvement was obtained not leading to a major library size reduction .

- Different combinations of functional spaces for any of the two test sets resulted in moderate library size reductions for $d = 3$. This is mainly attributed to the underling tensor product rule usually used in industrial implementations and mandatory in [69]. This observation is in concordance with other works [30] where higher order piece-wise models were dropped altogether in favor of linear ones, which are simpler to code and to use, without penalizing significantly the accuracy.

- In [23] a tensor product of 1 dimensional splines (or polynomials) is used with the additional property of reducing, possibly a posteriori, certain terms of the representation. Though this technique is of interest, no details of the procedure are provided and the need of regularization remains unattended.

- Multi-linear interpolation presented a monotonically diminishing error with the support size for σ, Σ and k_∞. Due to the bounding of the approximation with σ, important properties such as cross sections positivity is assured. Multi-linear interpolation can only have a second order error in relations like $\Sigma_{t,1} = \Sigma_{a,1} + \Sigma_{1\to1} + \Sigma_{1\to2}$.

In view of these results and the industrial applications multi-linear interpolation using a Cartesian grid is retained as the default methods for comparing other, more advanced methodologies. The next step is the use of local smooth approximation including regularization techniques even at the expense of non-interpolating approximations and without a tensor grid construction. This will be studied in the next chapter.

Chapter 4

Kernel methods and active learning

In this chapter, "Kernel methods" are used for cross section modeling in the context of ridge regression approximations. A kernel induces a function space that requires no particular structure of the support allowing to wage Active learning techniques to select the support points of the cross section data set. The decision criterion is based on loss functions that use σ, Σ, or k_∞ with and without a shared support condition for the cross section set. A sensitivity analysis on regularization, preconditioning and the approximation's order is carried out. The higher order approximations obtained overcome the difficulties found in chapter 3, related to the use of the tensor product rule and the need of regularization.

4.1 Machine learning and kernel methods

Reproducing Kernel Hilbert spaces (RKHS) has been largely used in the machine learning community for at least 20 years [73] reaching its apogee in the '90. They are a very general framework used to pose in a vector space, equipped with a norm, a vast quantity of machine learning problems. RKHS belongs to a broader family of methodologies called "Kernel Methods" that include, Kriging and Gaussian processes [74] usually employed for regression, and Support Vector Machines (SVM) for classification problems. Their underling feature is that strongly non linear problems, become separable and are linearly solved in a higher dimensional space without incurring in the significant costs of explicitly operating in this augmented space. In this thesis the terms RKHS and kernel methods are used interchangeably.

4.2 Introduction to Reproducing Kernel Hilbert Spaces (RKHS)

Kernel functions $k(x, z)$ associate objects of a rather general domain $\mathcal{X}$ to $\mathbb{R}$ as

$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}. \tag{4.1}$$

They are positively definite

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0, \tag{4.2}$$

given $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathbb{R}$. Also called *similarity function* they are symmetric, $k(x, z) = k(z, x)$.

For each $x \in \mathcal{X}$ a function space $\mathcal{H}$ composed of $k(x, \cdot) \in \mathcal{H}$ can be constructed. The dot notation $(\cdot)$ indicates a variable and in this case x could be though of as a parameter. Being the kernel symmetric these positions are, of course, interchangeable. The Hilbert space $\mathcal{H}$ is a *Reproducing Kernel Hilbert Space* (RKHS) noted as $\mathcal{H}_k$ if

$$\begin{aligned} (1)\ \forall x \in \mathcal{X}, k(x, \cdot) \in \mathcal{H}_k, \\ (2)\ \psi(x) = \langle \psi(\cdot), k(\cdot, x)\rangle_{\mathcal{H}_k} \forall x \in \mathcal{X}, \forall \psi(\cdot) \in \mathcal{H}_k. \end{aligned} \tag{4.3}$$

Let the function $\psi(\cdot) = k(z, \cdot), z \in \mathcal{X}$ and $\psi(x) = k(z, x), x \in \mathcal{X}$ then, by (2)[1],

$$k(z, x) = \langle k(\cdot, z), k(\cdot, x)\rangle_{\mathcal{H}_k}, \tag{4.4}$$

which is known as the *kernel trick* [77]. The Theorem of Moore-Aronszajn establishes that to every positively definite kernel an RKHS corresponds and vice-versa. We hence equate a positively definite kernel to a reproducing kernel nominated just as *kernel*. For any $\psi \in \mathcal{H}_k$, an inner product induces a norm $||\psi||_{\mathcal{H}_k} = \sqrt{\langle \psi, \psi, \rangle_{\mathcal{H}_k}}$ which can be obtained by kernel evaluation without explicitly formulating $\psi(\cdot)$ due to the kernel trick. Its utility will be apparent in the next section when dealing with a ridge regression problem. Other properties of RKHS are:

- The kernel associated to an RKHS is unique.
- Let k_1, k_2 be valid kernels, then they can be added $k = \alpha k_1 + \beta k_2, \alpha, \beta \in \mathbb{R}$ and multiply $k = k_1 \times k_2$ allowing the construction of more complex, compound kernels.

4.2.1 Approximation by kernel methods

Provided a support $\mathcal{S}$ (see Eq. 1.13) composed by a number of observations $N = |\mathcal{S}|$, a general regression problem using a function space belonging to an RKHS is

$$\arg\min_{\hat{y} \in \mathcal{H}_k} \sum_{i=1}^{N} \mathcal{L}(y_i, \hat{y}(\vec{x}_i)) + \lambda \mathcal{J}(f), \text{ with } \lambda \in \mathbb{R}. \tag{4.5}$$

For the use case considered in this work $\vec{x} \in [0, 1]^3$. $\mathcal{L}$ is a measure of discrepancy with respect to the available data and $\mathcal{J}$ a penalty function acting on the approximation. The

[1]The provided definition of an RKHS and its linking with reproducing kernels can be considered somewhat dogmatic. A more sophisticated elaboration demands only that a RKHS have bounded evaluation functional: $|L_x f| = |f(x)| \leq M||f||_{\mathcal{H}}, M > 0 \in \mathbb{R}, \forall f \in \mathcal{H}, \forall x \in \mathcal{X}$ [75]. This allows for $\mathcal{H}_k$ to use the *Riesz representation theorem:* each linear functional $\phi \in \mathcal{H}_k$, can be expressed in the form $\phi(h) = \langle h, f\rangle_{\mathcal{H}_k}, \quad f, h \in \mathcal{H}_k$ where f is uniquely determined by the functional ϕ. The theorem provides the connection between the inner products within $\mathcal{H}_k$ that is *reproduced* by the evaluation of elements in $\mathcal{H}_k$. The details of such presentation are beyond the scopes of this work and can be found elsewere [76].

hyper-parameter λ defines what is commonly called the "variance versus bias" trade-off, i.e. the fidelity of the approximation to the data in opposition to the bias ("smoothness" or "regularity") of the approximation $\hat{y}(\vec{x})$. Excessive variance can lead to over-fitting, were very irregular approximations follow the support data quite well, and excessive bias to the opposite situation called under-fitting. The approximation problem in Eq. 4.5 can be generalized in terms of linear operators using the average square loss for the evaluation functional and a square penalty functional leading to a *ridge regression* approximation problem

$$\arg\min_{\hat{y}\in\mathcal{H}_k} \sum_{i=1}^{N} (y_i - \hat{y}(\vec{x}_i))^2 + \lambda ||\hat{y}||^2_{\mathcal{H}_k}, \text{ with } \lambda \in \mathbb{R}. \tag{4.6}$$

By the *Representer Theorem* [73] the optimal approximation can be expressed in terms of the kernel (reproducing property)

$$\hat{y}(\vec{x}) = \sum_{i=1}^{N} \alpha_i k(\vec{x}, \vec{x}_i), \tag{4.7}$$

with $\vec{x}_i \in \mathcal{X}_{\mathcal{S}}$. Then from Eq. 4.7, $||\hat{y}||^2_{\mathcal{H}_k} = \vec{\alpha}^T K \vec{\alpha}$ being the positive definite and symmetric Gram matrix $K_{i,j} = \langle k(\vec{x}_i, \cdot), k(\vec{x}_j, \cdot)\rangle_{\mathcal{H}_k} = k(\vec{x}_i, \vec{x}_j)$ of size $N \times N$. The ridge regression problem can then be written as

$$\arg\min_{\vec{\alpha}\in\mathbb{R}^N} \frac{1}{N}(\vec{y} - K\vec{\alpha})^T(\vec{y} - K\vec{\alpha}) + \lambda \vec{\alpha}^T K \vec{\alpha}, \tag{4.8}$$

and by considering the derivative with respect to $\vec{\alpha}$

$$0 = \frac{2}{N}(\vec{y} - K\vec{\alpha})(-K) + 2\lambda K\vec{\alpha}. \tag{4.9}$$

K^{-1} exist for distinct sampling of $x_i \in \mathcal{X}$. The solution is found by solving the linear system

$$(K + N\lambda I)\vec{\alpha} = \vec{y}, \tag{4.10}$$

which is invertible if the regression coefficient $\lambda \geq 0$ and for $\lambda = 0$ the solution interpolates the support points[2]. The terms α_i can be interpreted as a priori weights of the observations $\vec{x}_i$ from the functions $k(\vec{x}_i, \cdot)$.

In Eq. 4.10, solving a linear system leads to a non-linear solution belonging to the function space reproduced by the kernel. Through the kernel trick $\vec{\alpha}$ is obtained by kernel evaluation without explicitly finding the mapping or feature space[3]. The kernel *reproduces* the function space and the inner product allows to even treat an infinite amount of dimensions with exact computations.

[2]In some methodologies such as Kriging, maximum likelihood techniques are used to estimate the coefficient directly from the data, specially in the presence of significant noise or numerical error [76].

[3]We recognize $\psi(\cdot) = k(x, \cdot) \in \mathcal{H}_k$. For example, the polynomial kernel $k(\vec{x}, \vec{z}) = ((\vec{x} \cdot \vec{z}) + 1)^d$ reproduces a function space of polynomials of degree d with a mapping $\psi : \mathcal{X} \to \mathcal{H}_k$. Let $d = 2$, $x = (x_1, x_2) \in \mathcal{X}$, $\psi(\cdot) = (x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, 1)$ that using the kernel trick $\langle\psi(\vec{x}), \psi(\vec{z})\rangle = ((\vec{x} \cdot \vec{z}) + 1)^2 = k(\vec{x}, \vec{z})$. The computation gains are evident, specially for very high values of d.

4.2.2 Advantages and drawbacks of the methodology

One of the interests of Kernel Methods, is that the representation of data, whether be cross sections, text, DNA molecule, engineering data, or other, is embedded in the feature space generated by the kernel and thus separated from the resolution and optimization methods. This is the source of many advantages:

- With no restriction in the structure of the support ($\mathcal{X}_\mathcal{S}$), Active Learning optimization techniques can be used to choose the most convenient observations. This feature is aligned with modern works on the application of machine learning techniques [57, 14].
- By the Representer Theorem the ridge regression problem (Eq. 4.10) is transformed into a convex optimization problem, independent of the function space.
- The approximations are expressed in terms of the kernel, not requiring to explicitly compute the mapping from $\psi : \mathcal{X} \to \mathcal{H}_k$ (kernel trick). An exact evaluation of the terms in the Gram matrix is provided by the kernel avoiding numerical estimations.
- A large variety of kernel functions are available reproducing a quite diverse collection of function spaces. Further customization is possible with hyper-parameters [74, 78].
- The regularization term λ can be used to easily handle variance versus bias trade-off defining the smoothness or regularity of the approximation. An automatic estimation methodology called "generalized cross validation" is proposed in [76].
- When modeling in this framework the function space to represent the data, the numerical solvers, and the possible active learning procedures are decoupled.

As with any methodology, drawbacks can be identified:

- A linear system given from a full rank Gram matrices of size $N \times N$ needs to be solved, possibly using the Cholesky or Singular Value Decomposition methods.
- These may present large conditioning numbers, especially for big support sizes. Indeed this kind of methodologies are not well suited for a support sizes with $N >$5E3, worsening with increased dimensionality [79]. Pre-conditioning is usually required.
- The evaluation cost of an approximation in the form of Eq. 4.7 increases linearly with the support's size N, since all observations are recalled to perform an evaluation.
- Kernel selection or even kernel design is required. The hyper-parameters can have a strong influence in the approximation quality and sensitivity studies are usually performed.
- Kernels are not orthonormal functions and possible high coefficients values may be encountered in relation to the bad conditioning number of the Gram matrix. Rejection techniques cannot therefore be easily applied.

4.3 Local function spaces reproduced by the kernel

Piece-wise polynomials are local approximation well suited for cross section approximation that exhibit "low practical dimensionality" [27]. However, difficulties were encountered in chapter 3 with respect to the need of regularization and the use of a tensor product structure in the support. This limiting set up is encountered in many state-of-the-art industry applications using, for example, multi-linear interpolation in a Cartesian grid.

The type of Spline spaces used in the previous chapter belongs to a Sobolev space

$$W_m([0,1]) = \{f \in [0,1] \to \mathbb{R}/f^{(1)}, f^{(2)}, \ldots, f^{(m-1)} \subset C^1, f^{(m)} \in L^2([0,1])\}. \tag{4.11}$$

with m the order of the functions f populating the space. For $m > 1$ these functions are smooth, adequate to model smooth data. For reproducing this function space in RKHS two kernel types are considered: the spline kernel and the Bernoulli kernel.

4.3.1 The Bernoulli kernel

A $W_m([0,1])$ is reproduced by the Bernoulli kernel

$$k_m(x,z) = \sum_{l=0}^{m} \frac{1}{(l!)^2} B_l(x) B_l(z) + \frac{(-1)^{m+1}}{(2m)!} B_{2m}(abs(x-z)), \tag{4.12}$$

Though the kernel may seem complicated, we notice that in actuality for $m = 1$, $k_1(x,z) = 1 + B_1(x)B_1(z) + \frac{1}{2}B_2(abs(x-z))$ with

$$\begin{aligned} &B_0(x) = 1, \quad B_1(x) = x - \frac{1}{2}, \quad B_2(x) = x^2 - x + 1/6, \\ &B_4(x) = x(x(x(x(-2)+1))) - 1/30, \text{ (for } m = 2) \ . \end{aligned} \tag{4.13}$$

In particular the norm of a $f \in W_m[0,1]$ is

$$||f||^2 = \sum_{l=0}^{m-1} \int_0^1 (f^{(l)}(x))^2 dx + \int_0^1 (f^{(m)}(x))^2 dx. \tag{4.14}$$

4.3.2 The spline kernel

For an m differentiable function space the spline kernel is

$$k_{m,0}(x,z) = \sum_{i=0}^{m-1} \frac{x^i z^i}{i!i!} + k_{m,1}(x,z), \quad k_{m,1}(x,z) = \int_0^1 \frac{(x-u)_+^{m-1}(z-u)_+^{m-1}}{(m-1)!(m-1!)} du. \tag{4.15}$$

By kernel composition $k_m(x,z) = k_{m,0}(x,z) + k_{m,1}(x,z)$. Then, for $m = 1$ the spline kernel is

$$k_1(x,z) = 1 + \max(x,z). \tag{4.16}$$

reproducing a linear space. For order $m = 2$

$$k_2(x, z) = 1 + uv + u^2(v - u/3)/2, \tag{4.17}$$

with $u = x, v = z$ if $x < z$ or $u = z, v = x$ if $x > z$. In particular the norm of a $f \in W_m[0, 1]$ is

$$||f||^2 = \sum_{l=0}^{m-1} (f^{(l)}(0))^2 + \int_0^1 (f^{(m)}(x))^2 dx. \tag{4.18}$$

The spline kernel reproduces an approximation where the moment of order 2 ($\int_0^1 (\hat{y}^{(m)}(u))^2 du$) can be interpreted as an energy minimization. According to [76] the functional space reproduced by these two kernels types is of special interest since is one of "the best non-parametric curve smoothing method".

4.4 Results

In the following sections results are presented using the Bernoulli and the Spline kernels simply reproducing a high order approximation space for $m \geq 2$. For treating a multidimensional case the tensorization

$$k(\vec{x}, \vec{x}_i) = \prod_{j=1}^{d} k_j(x_j, x_{j,i}), \quad d = 3, \tag{4.19}$$

has been considered. Though kernel composition allows to select different k_j for $1 \leq j \leq d$, the same kernel is used in every dimension for simplicity . The approximation's errors are computed on the test set $\mathcal{T}_{flat}$ which stands as a relatively homogeneous sampling of the input space (see Section 2.2.2) not belonging to the support $\mathcal{S}_\sigma$. It's important to remark that though local approximation are used, an evaluation in the form of Eq. 4.7 requires the computation of all the observations.

4.4.1 Comparison of Spline and Bernoulli kernels

In Tab. 4.1 the relative error average AV_σ, is shown for the spline and Bernoulli kernels (see Eq. 4.15, 4.12) using the Industry set for the support. The linear spline kernel reproduces the multi-linear approximation achieving the same bounded interpolation error, without requiring preconditioning or regularization ($\lambda = 0$). This facilitates a systematic analysis on the different active learning strategies and, though maybe sub-optimal for smooth data, it's a function space of particular interest since is often used in industry. In concordance with the studies of chapter 3, higher order approximation requires regularization and possibly preconditioning to be consistently superior to linear schemes. These local approximations perform well in comparison to other kernels that use hyper parameters which are examined in the Appendix B.1.

Approximation name	Order (m)	AV_σ	AV_Σ	AV_{k_∞}
Multi-linear (B-splines)	1	0.021	0.016	14
Bernoulli kernel	1	0.051	0.032	21
Bernoulli kernel	2	0.059	0.02	23
Spline kernel	1	0.021	0.016	14
Spline kernel	2	0.061	0.022	35

Table 4.1: Approximation's error using the Bernoulli and Spline kernels for the full grid support $\mathcal{S}_\star$. The spline kernel reproduces the multi-linear error (in green) while this is not the case for the Bernoulli kernel (in red). High order approximation exhibit larger errors possibly requiring regularization. The same errors trends are present for all supports.

4.5 Supervised Learning

The literature review carried out in chapter 1 showed that tensor product rules compromise library reduction techniques by including vast amounts of possibly unnecessary information. This is precisely the type of grid used in industry settings for multi-linear interpolation. Kernel Methods on the other hand, impose no condition on the support's structure, thus enabling the use of *supervised learning*[4] procedures from the field of machine learning. Pool active learning, referred to from now on as active learning, is performed by computing the extrema of a *loss function* $\mathcal{L}$, which lays at the heart of the active learning process defining the optimization metrics. The term "pool" indicates that a possibly big and fix amount of information is available. The learning procedure must be aligned with the modeling objectives, and can be conceptualize as [80]:

- Assure an *information* dense set by selecting points with high incertitude/error, that maximize the model's variation or investigate regions of high disagreement between models (query by committee).
- Assure a *representative* coverage of the domain for a given metric (homogeneous, based on the test density, etc) or correlate the sampling to the variance density explicitly modulating the support's anisotropy. Care must be given to the treatment of outliers that could claim a lot of attention.
- Assure a *diverse* sampling of the space under some metric, for example by exploring every possible clusters of data, independently of the local error or variance within them.

A clear cut distinction between these criteria is not always straightforward, depending on the algorithm and the approximation at hand. This thesis mainly focuses on the first criterion: maximize prediction capabilities while minimizing the number of support points.

[4]Supervised learning refers to the learning agent being subject to a "teacher that supervises" the process by correcting the learner's predictions in view of the correct answers. In other forms of machine learning such as unsupervised learning the "correct" answers may not exist beforehand.

4.5.1 Active learning for cross section modeling

Cross section approximations are the agent subject to active learning as presented in pseudo code in Algo. 1. The model starts with an initial small support $\mathcal{S}_0 \subset \mathcal{S}_\mathcal{P}$, of 5 points chosen at random, and loss function values are computed within the loop to find new optimal points, $\vec{x}_\dagger \in \mathcal{X}_\dagger \subset \mathcal{X}_\mathcal{P}$. The algorithm iterates until the break condition is reached, in this case a user defined budget b, for the library size. The result is an optimal support $\mathcal{S}_\dagger = \{\sigma(\vec{x}_i), \vec{x}_i \in \mathcal{X}_\dagger\}$ with $\mathcal{X}_\dagger$ sheared between the cross sections.

Data: $\mathcal{S}_\mathcal{P}$, $\mathcal{S}_0$, b, $\mathcal{L}$
1 $\mathcal{X}_\dagger = \mathcal{X}_0$
2 **while** $|\mathcal{X}_\dagger| \leq b$ **do**
3 $\quad \hat{y} = \hat{y}(\vec{x}|\mathcal{S}_\dagger), \forall y \in \mathcal{Y}$
4 $\quad \vec{x}_\dagger = \underset{\vec{x}\in\mathcal{X}_\mathcal{P}\setminus\mathcal{X}_\dagger}{\arg\max}\ \mathcal{L}(\hat{y}(\vec{x}), y(\vec{x}))\ \mathcal{X}_\dagger = \mathcal{X}_\dagger \cup \vec{x}_\dagger$
5 **end**
Result: An optimal sampling by active learning $\mathcal{X}_\dagger$ from a discrepancy function $\mathcal{L}$ of size b.

Algorithm 1: Active learning pseudo code for cross sections sharing support. The variable y can be σ, Σ, or k_∞. An initial small support, of size $|\mathcal{S}_0| = 5$, is increased by adding points $\vec{x}_\dagger \in \mathcal{X}_\mathcal{P}$ that define an optimal sampling $\mathcal{X}_\dagger$, with respect to a loss function $\mathcal{L}$. It recieves a vector (of errors) from which the corresponding $x_\dagger$ is selected. In each step the approximations $\hat{y}$ are computed as required by the loss function for the sampling $\mathcal{X}_\dagger$ defined in the previous iteration until the break condition defined by budget b is reached.

In other works an optimization using the infinite multiplication factor in an interative way, one dimension at the time [14] or microscopic cross sections [31, 55] (essentialy by hand) has been explored but restricted to a Cartesian and sparse grid respectively. In [57] an unstructured support is explored but limited to the grid employed to numerically solve a quadrature problem (Tucker grid).

In this thesis a broader range of loss functions listed in Tab. 4.2, in increasing order of complexity, are examined. First, a simple random selection of the support's points (RAND) analyzes the effect of using unstructured supports as opposed to full grid, but without active learning. With U only the uranium thermal fission cross section is considered. For RXS, RXSI, and XSI all cross sections are used, weighting with the importance (see Eq. 2.10) in the latter two. XSI is the only loss function that uses absolutes errors for the cross section set. The loss functions M and MF consider the absolute error of the set of macroscopic cross sections and the multiplication factor respectively. They require computing the approximation for every cross section.

Loss function ($\mathcal{L}$)		Description
Name	Function	
RAND	rand	Random error.
U	$\lvert\hat{\sigma}_{235,f,2} - \sigma_{235,f,2}\rvert$	Absolute errors of a single cross section.
RXS	$\arg\max_{\sigma\in\mathcal{Y}} \Delta\sigma/\sigma$	Relative errors for the entire cross sections set.
RXSI	$\arg\max_{\sigma\in\mathcal{Y}} I_\sigma\Delta\sigma/\sigma$	Relative errors for the entire cross sections set weighted with the importance.
RIXS	$\arg\max_{\sigma\in\mathcal{Y} \text{ if } max(I_\sigma)>0.2} I_\sigma\Delta\sigma/\sigma$	Relative error for a subset of cross section with high importance.
XSI	$\arg\max_{\sigma\in\mathcal{Y}} I_\sigma\Delta\sigma$	Absolute errors for the entire cross sections set.
M	$\arg\max_{\Sigma\in\mathcal{Y}_\Sigma} \Delta\Sigma$	Absolute error for the entire macroscopic cross section set reconstructed from microscopic cross sections
MF	$\lvert\hat{k}_\infty - k_\infty\rvert$	Absolute error for the infinite multiplication factor reconstructed from macroscopic cross sections.

Table 4.2: Loss functions ($\mathcal{L}$) acting on σ, Σ or k_∞ expressed as y. When a set of approximation is considered, that with the highest error computed on the unused points of the pool is selected (this is the approximation participating in the active learning for that point). The loss functions RXS, RXSI, XSI, M and MF require the computation of the entire microscopic cross section set $|\mathcal{Y}| = 144$. The acronyms stand as R for relative, XS for including all cross section and I for importance.

4.5.2 Computational cost of Active Learning

The Active learning process presented in Algo. 1 can be computationally expensive. The size of the support $N = |\mathcal{X}_\mathcal{S}|$, that defines the Gram matrix of $size(K) = N \times N$, and the amount of coefficient in Eq. 4.7 increases as $|\mathcal{X}_0| \to b < |\mathcal{X}_\mathcal{P}|$. For each cross section and iteration (W) of Algo. 1, the step 3 requires the computation of K and $\vec{\alpha}$[5] while the step 4 of $\hat{y}(\vec{x}_i), \forall\vec{x}_i \in \mathcal{X}_\mathcal{P}\backslash\mathcal{X}_\dagger$. Evaluations are performed on $M = |\mathcal{X}_\mathcal{P}| - |\mathcal{X}_\dagger|$ points candidates from the pool, not used by the support. The amount of kernel evaluation (Eq. 4.1) for each cross section and each iteration is

$$d(N^2 + NM), \tag{4.20}$$

were the tensorization has been accounted for. In order to provide a simple estimation of the active learning cost, expressed in total kernel evaluation being $\mathcal{X}_0 = 5$ and $|\mathcal{X}_\mathcal{P}| = 2500$, let $N = M = W = 1250$ (i.e. half the pool). Then, for a single cross section, the total amount of kernel evaluations is $\sim$ 1.1E11, which is significant.

The type of loss function considered directly impacts the computational cost. RXS, RXSI, XSI, M, and MF require the computation of the entire cross section set. RIXS only uses a subset of important cross sections (of $\max(I_\sigma) > 0.2$) reducing the overall active learning cost. A random selection of support points has no significant associated cost. Actual active learning times are presented in Sec. 4.10. The cost of computing $\vec{\alpha}$ has been neglected in this analysis.

[5]Possibly thorough a Cholesky or singular value decomposition.

4.6 Kernel methods using active learning in a shared support scheme

The error AV_σ in function of the library size is presented in Fig. 4.1a for the loss functions of Tab. 4.2. The same linear piece-wise function space is used, indicated in blue the use of a Cartesian grid and specifically with a star the Industry set. Errors, bounded by the support, diminish monotonically and converge as the support increases. The different grid's anisotropies induced by $\mathcal{L}$, are compared for a library size of 2.2E4 coefficients where they are the most diverse[6]. Grid density plots projected onto the burnup are shown in Fig 4.1b, 4.1c, 4.1d. For this relativity low support the full grid is still missing some points between 0.5 and 0.75.

A random selection (RAND) presented in Fig. 4.1b already produces significant improvements as expected: collinearity undermines the informativeness of an approximation's support [29, 22]. In fact, for every support size, this accuracy gain rivalizes with the use of higher order splines in chapter 3 even if it's only following the discretization of the pool. In a way this isolates the benefit of just using an unstructured support reducing the library size by half without added computational cost. All active learning strategies further improve the accuracy.

The lowest AV_σ error is obtained with RXS since each new point $\vec{x}_\dagger$ is chosen by comparing relative errors throughout the entire cross section set. The grid distribution is shown Fig. 4.1b where a high density is observed at the first burnup values. Error histogram are shown in Fig. 4.2 characterized by centred means, almost a normally distributed shape, small standard deviations and no error trails for RXS which is not the case for multi-linear with a small support. Similar histograms are noticed for the other active learning strategies.

Absolute errors, even if weighted by the importance (XSI), result in a single cross section participating in the active learning: $\sigma_{135,a,2}$ due to its significantly larger absolute values. Consequently, the computation of step 3 and step 4 in Algo. 1 is carried for every cross section but systematically unused $|\mathcal{Y}| - 1$ times. The support's anisotropy and resulting error profile are however, rather similar to the other strategies as shown in Fig. 4.1d.

In Tab. 4.3 the error for σ, Σ, and k_∞ is presented for a library size of 2.2E4 coefficients. Library size reductions are provided at constant average macroscopic cross section error, $\mathrm{AV}_{\Sigma,\star}$. Considering the infinite multiplication factor (M) does have a marginal benefit in macroscopic cross sections maximum error though comparable results are obtained with RXSI or RIXS. These loss functions seem to allocate more points towards the middle of the burnup as shown in Fig. 4.1c, away of regions with very low concentrations like the low burnup values for many isotopes. MF achieves the lowest errors in k_∞ but not in σ or Σ.

The magnitude in library size reduction with respect to a Cartesian grid is coherent

[6]Since It's sufficient to manifest the effect of different active learning strategies without imposing additional point due to a budget similar to the size of the pool. Indeed as the unstructured grid $\mathcal{X}_\dagger \to \mathcal{X}_\mathcal{P}$ the grid density tends towards the pool presented in chapter 2 (see Fig. 2.6).

Grid type	MAX_σ	AV_σ	MAX_Σ	AV_Σ	MAX_{k_∞}	AV_{k_∞}	Library Size	
	[%]				[*pcm*]		Coef.	Δ[%]
Cartesian Grid	13.488	0.503	2.500	0.285	498	149	1.5E^5	-
RAND	13.457	0.084	2.256	0.047	161	13	4.3E^4	21
U	1.465	0.025	0.318	0.012	56	7	1.8E^4	88
XSI*	1.627	0.019	0.318	0.010	38	5	1.9E^4	87
RXS*	1.901	0.014	0.345	0.009	31	5	1.3E^4	91
RXSI*	1.933	0.018	0.319	0.010	26	4	1.4E^4	90
RIXS	1.671	0.022	0.318	0.010	37	6	1.1E^4	92
M*	1.946	0.016	0.318	0.009	21	3	1.1E^4	92
MF*	1.890	0.036	0.443	0.019	21	3	2.6E^4	83

Table 4.3: All approximations consider a linear approximation space with different grid types. The Cartesian grid is defined by the user. RAND is an unstructured grid chosen randomly. The others are subject to active learning using different loss functions. Maximum and average errors are calculated at the support corresponding to a library size of 2.2×10^4 coefficients. The library size and corresponding reductions are calculated at the constant average error ($AV_{\Sigma,\star}$) of multi-linear. * approximations are computed for the entire cross section set $\mathcal{Y} = 144$.

with other works [53] that use the same linear function space. On the other hand this methodology, RKHS subject to active learning, is not constrained to a regular structure and requires no particular discretization of the pool. For this case similar gains are observed when performing the optimization using very diverse objects (σ, Σ and k_∞) suggesting a certain robustness of the methodology. The use of integral parameters, as done elsewere [14], does not provide significant benefits.

Though active learning could be considered an "off-line" task, performed only once during cross section preparation, is computationally intensive as discussed in Section 4.5.2. The loss functions XSI, RXS, RXSI, M, and MF require the computation of the whole cross section set unlike RIXS or U, that attain similar error profiles at smaller computational cost. RIXS is selected as the "best" strategy maximizing the *information* of the model at a relativity moderated computational cost.

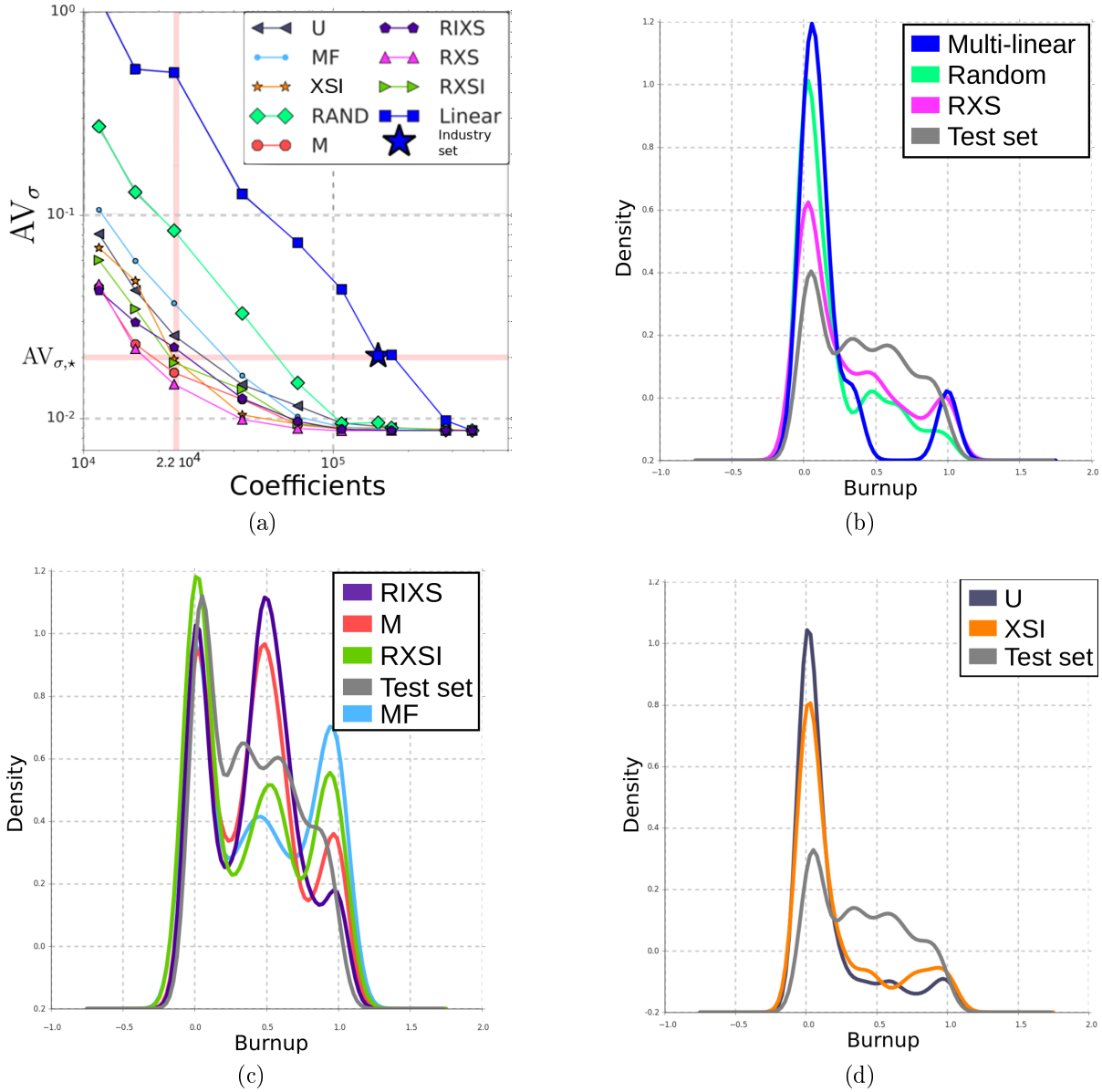
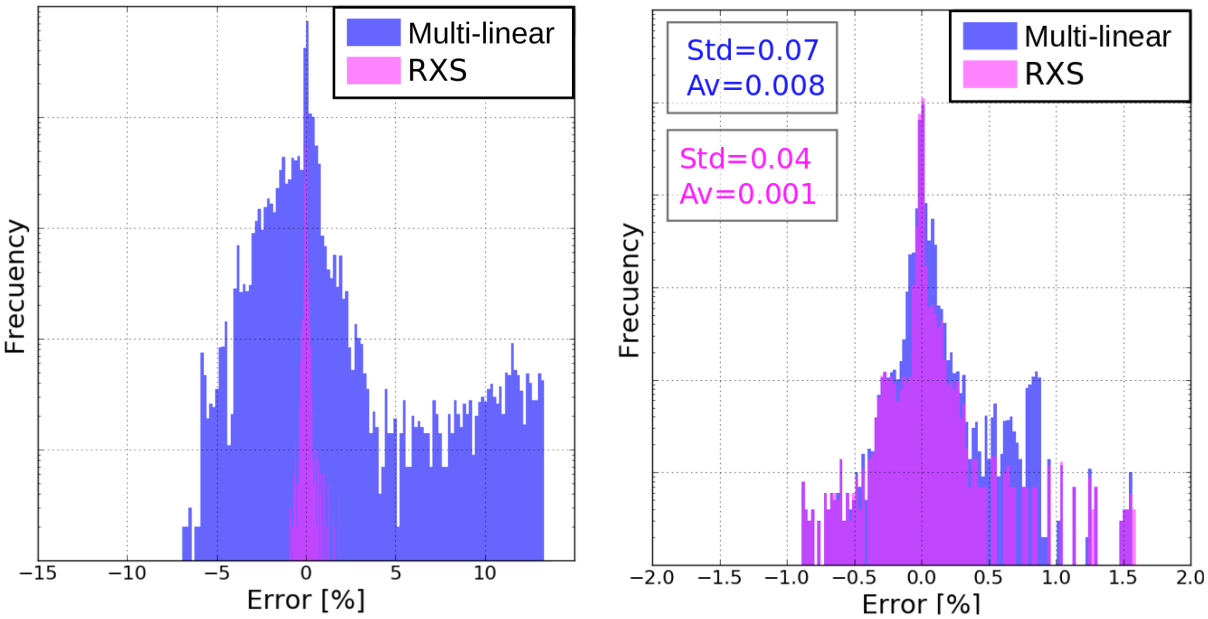


Figure 4.1: Average microscopic cross section error using linear piece-wise approximations with different grids in a shared support scheme. In blue a Cartesian grid is used while in the others different active learning strategies are used. (a) AV_σ error noting two lines of comparison, grid distribution at relativity low support (2.2E4 coefficients or 150 coefficients per cross section) and line of constant error ($AV_{\sigma,\star}$). Grid distribution projected in the burnup are presented in (b), (c), and (d). The test set $\mathcal{T}_{flat}$ is also presented facilitating comparisons between the plots. (b) A random selection of points follows the original data distribution of the pool. (c) When considering the isotope's concentration the middle burnup region is sampled more. (d) When optimizing with a single cross section, the optimal support has a similar profile than with RXS.



(a) Library size of 2.2×10^4

(b) Library size of $|\mathcal{X}_\star|$

Figure 4.2: Relative error histogram over $\mathcal{T}_{flat}$ for every cross section. A Cartesian grid is compared to the RXS active learning strategy which exhibit smaller errors using the same linear approximation space. (a) RXS is characterized by centered means, almost normally distributed shape, small standard deviations and no error trails which is not the case for multi-linear with a total library size of 2.2E4 coefficients (b). When the Industry set is used, multi-linear's error improves though outperformed by RXS.

4.7 Cross section clustering in active learning

Actual loss function values are plotted for RXS and RXSI in Fig.4.3a discriminating per cross section. A *clustering effect* can be noticed where the cross sections participating (independently) in the selection of each support point, end up grouped together in a few blocks mainly dominated by $\sigma_{239,f,2}$, $\sigma_{235,f,1}$, $\Sigma_{res,2\to1}$, and $\Sigma_{res,a,1}$.

The use of the importance further solidifies this trend: for RXS a total of 50 cross sections appear (many marked in gray) while for RXSI only 10, since $I_\sigma \sim 0$ for a large amount of cross sections. $\mathcal{L}$ values are bigger for RXS than for RXSI (since $I_\sigma \leq 1$) but the overall profile and error stagnation ($|\mathcal{X}_\mathcal{S}| > 1000$) remains the same, suggesting that the useful information is being extracted at a similar rate from $\mathcal{X}_\mathcal{P}$. For this reason the importance was successfully used to pre-select a small cross section subset (RIXS) obtaining similar grid distributions at a lower computational cost. The profiles present breaks in the derivative (at about 200 and 1000) that can be understood as a change in the "relevance" of the points being added to the model and could even be used as a stop criterion for Algo. 1.

To further exhibit this, partial loss function values

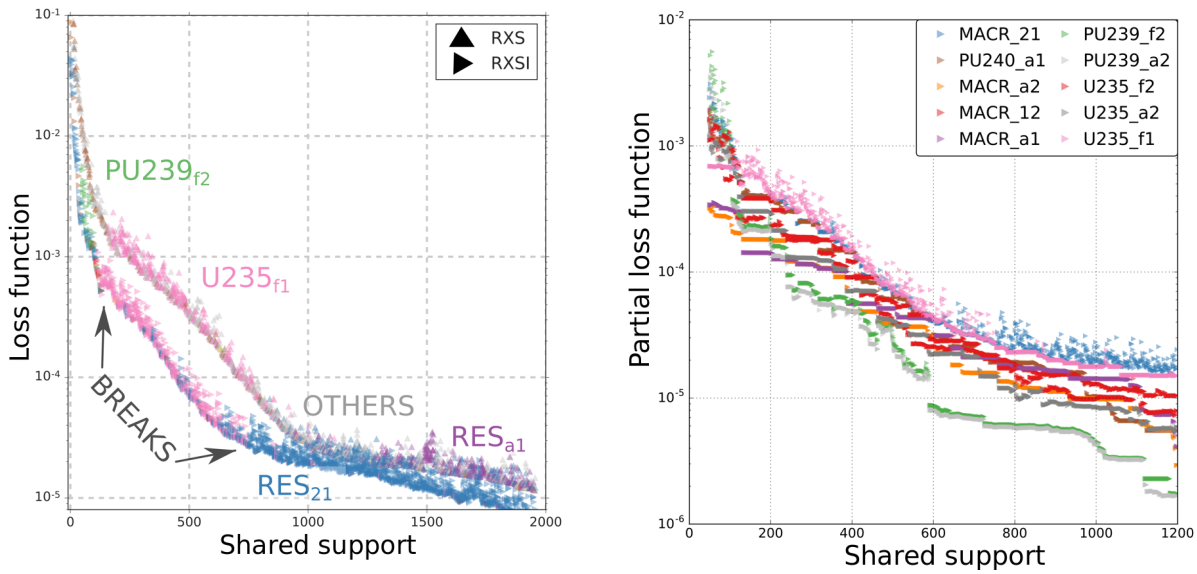
$$\mathcal{L}_\sigma = 100|\hat{\sigma}/\sigma - 1|_\infty, \tag{4.21}$$

are presented in Fig. 4.3b for a few cross section. $\mathcal{L}_{\sigma_{235,f,1}}$ and $\mathcal{L}_{\Sigma_{res,2\to1}}$ exhibit jumps due to the ongoing active learning process [7]. However, for the remaining cross section not in the cluster, significant plateaus in their partial loss function profiles are observed, induced by the shared support condition that forces the model to incorporate a common $\vec{x}_\dagger$ for every cross section. For example, $\mathcal{L}_{\sigma_{239,f,2}}$ clearly stagnates, especially after 600 points. The clustering effect is a consequence of using a share support scheme. At least with respect to the infinite norm used in this work, the optimization effort is subordinated to a small subset of cross sections, forcing the model to incorporate large amounts of unnecessary data.

4.7.1 Independent cross section pool active learning

In the cross section modeling literature it was observed that cross section complexity varies significantly within $\mathcal{Y}$. Partial loss function values suggest that the model could benefit from dropping the shared support condition and performing an active learning process independently per cross sections, especially since only the approximation's coefficients need to be stored. This is presented in Algo. 2 using, again, a linear function space. For the error tolerance in the break condition two methods are considered. In "No shared support"(NoS) a constant tolerance $\delta_\sigma = \delta$ is used for every cross section. In "Importance no shared support" (ImpNoS) the tolerance is weighted with the importance for each cross section, $\delta_\sigma = mean(I_\sigma)\delta$. The constant tolerance values used for both methods is δ =0.1, 0.01, 0.001.

[7] Though difficult to appreciate in the picture the very first cross section used for selecting the points is $\sigma_{239,a,1}$ that in relative terms can manifest big maximums errors as analyzed in chapter 3.



(a) (b)

Figure 4.3: (a) Loss function values for RXS and RXSI discriminated by cross section. Only a few subset of cross sections dominate the process. Loss function profiles quantify the "relevance" of the support points being added to the model. (b) Partial loss function values ($\mathcal{L}_\sigma$) for RXSI showing significant plateaus (with respect to the infinite norm) for several isotopes. These cross section do not participate in the active learning, but due to the share support condition are forced to incorporate additional support points.

AV_Σ in function of the library size is presented in Fig. 4.4a. Errors for Σ and k_∞ are presented in Tab. 4.4. With NoS a reduction in the library size (or gain in accuracy) is observed with respect to both a Cartesian grid and RIXS. For ImpNoS at $\mathrm{AV}_{\Sigma,\star}$ the library size is reduced another order of magnitude arriving at 1% of multi-linear in a Cartesian grid. The multiplication factor is also quite well predicted. However only cross section with significant I_σ have a small error tolerance and thus high AV_σ is noticed. Though integral parameters such as k_∞ or Σ have important physical meaning, the latter being actually used by the core solver, the need of keeping track of the original functions in multi-objective optimization is evident in this case.

The constitution of these two types of cross section libraries are further studied in Fig. 4.5b and Fig. 4.5a for a size of 1E4 coefficients. A Square plot is used, where the amount of terms relative to the library size defines the area of each cross section. The same colorbar indicates their absolute amount. In the ImpNoS cross section library only $\sigma_{235,a1}$, $\sigma_{235,f1}$, $\Sigma_{res,1\to2}$, and $\Sigma_{res,2\to1}$ surpass 300 coefficients, while this is the case for the majority of cross sections in NoS. The active learning only focuses in cross sections relevant to Σ. The scattering cross sections are strongly affected by this since many of them have $I_\sigma = 0$.

In Fig. 4.5c relative cross section errors for the ImpNoS library are presented discriminated by isotope, reaction, and group. Many unimportant cross section indeed exhibit big errors. Scattering cross sections are only well approximated for Σ_{res} and ^{238}U and to some extent ^{235}U. A reaction may be well approximated in one energy group but not in the other as it's the case for $\sigma_{135,a,2}$ and $\sigma_{238,f,1}$. For NoS all cross sections are equally

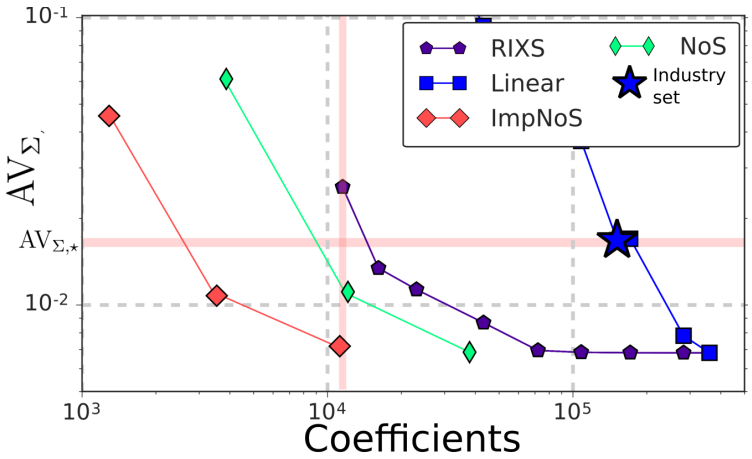
Data: $\mathcal{S}_{\mathcal{P}}$, $\mathcal{S}_0$, δ_σ
for $y \in \mathcal{Y}$ **do**
  $\mathcal{X}_{\sigma,\dagger} = \mathcal{X}_0$
  while $max(err) \leq \delta_\sigma$ **do**
    $\hat{\sigma} = \hat{\sigma}(\vec{x}|\mathcal{S}_{\sigma,\dagger})$
    $err = \hat{\sigma}(\vec{x})/\sigma(\vec{x}) - 1$
    $\vec{x}_{\sigma,\dagger} = \underset{\vec{x}\in\mathcal{X}_{\mathcal{P}}\setminus\mathcal{X}_{\sigma,\dagger}}{\arg\max}\ err$
    $\mathcal{X}_{\sigma,\dagger} = \mathcal{X}_{\sigma,\dagger} \cup \vec{x}_{\sigma,\dagger}$
  end
end
Result: An optimal set $\mathcal{X}_{\sigma,\dagger}$.

Algorithm 2: Pseudo code for Pool Active Learning without cross section sharing the support. The support is optimized independently for every cross section resulting in a varying number of coefficient when the break condition is reached. This is defined as an error tolerance that can depend on the cross section. Relative errors with the infinite norm are used.

well approximated and thus the plot not shown.

In [29] it was reported that thermal cross section required the double of regression terms than fast cross sections. This coincides with the observations in chapter 2 that they, in general, have larger variances. On the other hand with relative errors, required to make meaningful comparisons among cross section, fast cross section could pose an additional challenge. Without the shared support condition for a fix tolerance (NoS) using a linear function space, it was observed a rate of fast/thermal coefficients of 0.56, 0.73, 1.05 for $\delta_\sigma = 0.1, 0.01, 0.001$. It could be argued that when decreasing the relative error tolerance, the modeling challenge driven by the variance is progressively dominated by the bigger relative challenge and even possible numerical noise or outliers in fast cross sections.

When using an expansion method for cross section approximations that only requires to store the coefficients, dropping the share support condition could be convenient as the required amount of support points vary within the cross section set. It could even be argue that in an active learning context the amount of required observation is proportional to the cross section complexity, possibly characterized by the derivative being thus $|\mathcal{S}_\sigma| \propto \sigma^{(2)}$ which further specializes the library without a shared support condition. The actual applicability of weighting with the importance is probably limited. In this particular case, unimportant scattering cross section were shut-down from the library, which may, or may not be acceptable depending on the modeling objectives. Conceptually the loss function is the right place to define the modeling objectives of the cross section library and where any physical insight could be enforced.

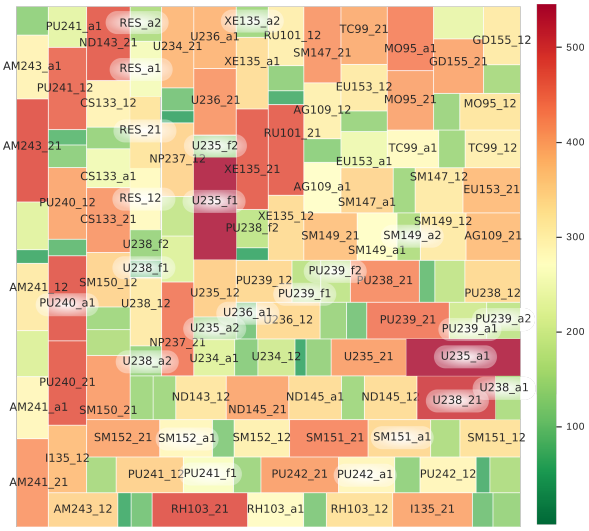


(a)

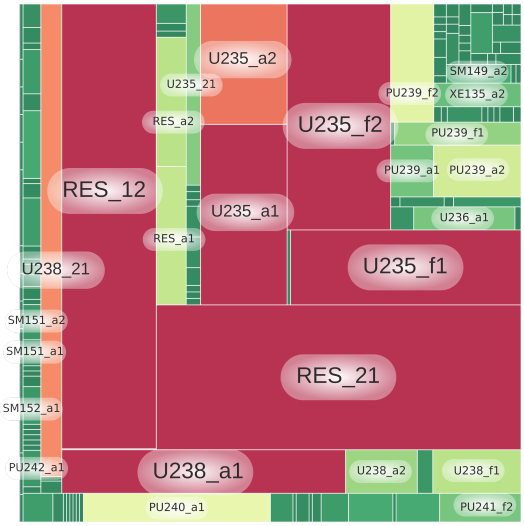
Approximation	MAX_σ	AV_σ	MAX_Σ	AV_Σ	MAX_{k_∞}	AV_{k_∞}	Library Size	
	[%]				[*pcm*]		Coef.	Δ[%]
Multi-linear	40	1.1	4.5	1	1200	400	$1.5E^5$	-
NoS	1.8	1.4E-2	3E-1	0.011	45	9	$9.0E^3$	94
ImpNoS	48.847	2.569	3E-1	0.007	15	3	$2.2E^3$	99
RIXS	2.9	6E-2	4E-1	2.8E-2	80	18	$1.1E^4$	92

(b)

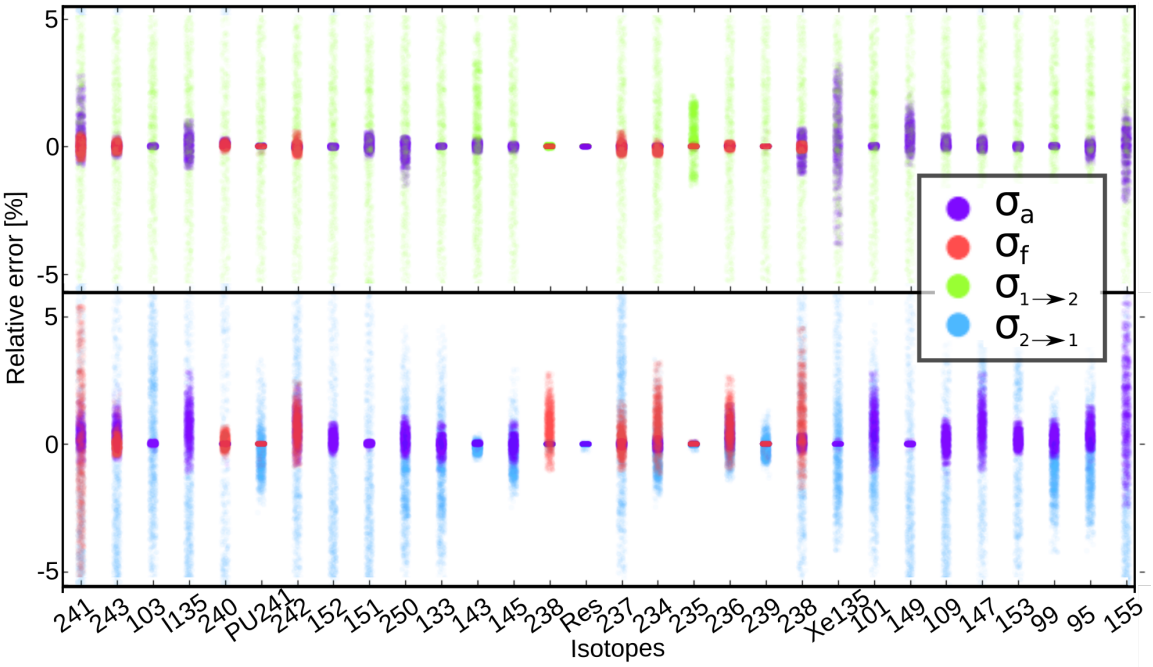
Figure 4.4: Errors for linear piece-wise approximations with independent active learning per cross section (no shared support). The three points of the curve correspond to a δ of 0.1, 0.01, and 0.001. In ImpNoS this value is affected by the importance. (a) AV_Σ error without a shared support condition. At constant error $AV_{\Sigma,\star}$ almost two order of magnitude in library reduction are obtained with ImpNoS where error tolerance is modulated by the cross section importance. NoS also reduced the library size with respect to a Cartesian grid and to RIXS. (b) Errors for active learning models without shared support for σ, Σ, k_∞. Errors at constant support $\sim$ 1.1E4 coefficients. Library size at constant error $AV_{\Sigma,\star}$. High microscopic cross section error can be seen for ImpNoS.



(a) NoS (Library size of 3.7E4)



(b) ImpNoS (Library size of 1.1E4)



(c) ImpNoS (Library size of 1.1E4)

Figure 4.5: Library composition at maximal size for the two for active learning without a shared support condition. The area marks the amount of coefficients relative to the library size. The same colorbar is used in both figures. (a) In NoS the amount of coefficients vary in order to attain the same error tolerance for each cross section. (b) In ImpNoS the importance is considered in the active learning process and thus the cross section relevant to *Sigma* compose the majority of the library. (c) σ relative errors of ImpNoS at 1.1E4 discriminated by reaction, group, and isotope. A strong correlation with the number of coefficients composing the library is noted.

4.8 Higher order RKHS approximation

Local smooth approximations are of interest due to the underlying physical phenomena producing cross section data. RKHS can reproduce this function space without the limitations previously encountered in chapter 3, since a regularization coefficient λ naturally appears in the kernel ridge regression problem (Eq. 4.6).

4.8.1 Regularization and preconditioning

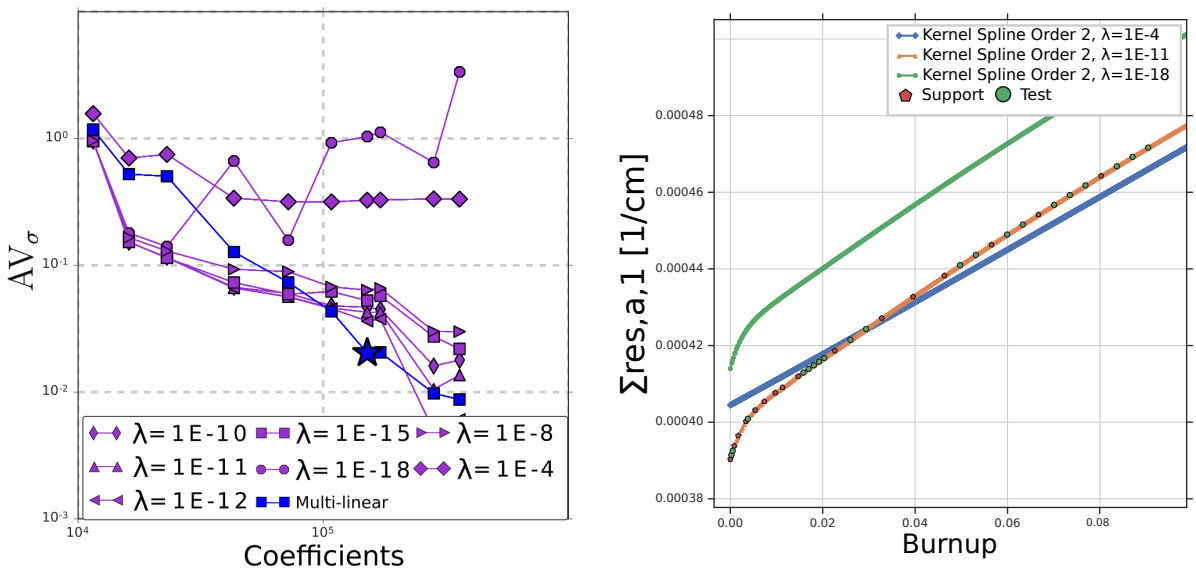
A sensitivity study was carried out for the regularization term λ of Eq. 4.10 that defines the "variance versus bias" trade off. In Fig. 4.6a the results are shown for the spline kernel presented in Eq. 4.17. High errors are found for λ ∼1E-4 (high bias) and λ ∼1E-18 (high variance) being the best compromise λ ∼1E-11. Average k_∞ errors are ∼ 200 *pcm* and maximum errors can reach 1000 *pcm* with inadequate regularization. In Fig. 4.6b an example for the cross section in function of the burnup $\Sigma_{res,a,1}(Bu, T_{f,0}, C_{B,0})$ is provided where high bias does not allow the approximation to follow the data while high variance generate an offset, possible due to numerical errors associated to the bad condition number of the Gram matrix.

Error histograms for the absorption cross section, discriminated by isotope, are presented in Fig. 4.7 for λ=1E-4 (high bias), λ=1E-11 (optimal) and λ=1E-18 (high variance). Only the 10 cross section with the highest errors are shown. Errors with high bias do exhibit a zero mean but have large standard deviations. For λ=1E-11 normal distribution are seen with a relatively reduced amount of outliers. For λ=1E-18 many error have significant offset errors and large standard deviations.

Utilizing square root for the variables marginally improves the conditioning number of the matrix K, either first or second order splines as shown in Tab. 4.5. Therefore a λ=1E-11 and the square root for preconditioning are the two candidates to generate adequate high order smooth cross section approximations, able to cope with strong variation in the first burnup regions.

4.8.2 Higher order spline approximation with RKHS and active learning

An active learning procedure using the RIXS loss function for a high order spline kernel with preconditioning (square root) and regularization (λ =1E-11) is presented in Fig. 4.8a. Though preconditioning or regularization alone are not satisfactory, the combination of the two result in a monotonically decreasing error similar to linear approximations. However, being a second order differentiable space, a gain in accuracy of up to one order of magnitude , is systematically obtained. In Fig. 4.8b the maximum relative error of the entire cross section set is presented for $\mathcal{T}_{Xe}$ where a very fast convergence is obtained, showing that the approximation is, in practice, bounded by the support. Two cross section examples are shown in Fig. 4.8c and Fig. 4.8d where this regularized spline approximation is able to follow the data quite well. Errors for Σ and k_∞ are presented in Table 4.4 where



(a) (b)

Figure 4.6: Effect of regularization in microscopic cross section approximation using a high order spline kernel. (a) AV_σ for different regularization coefficients λ. The lowest error is achieved with λ=1E-11 (b) Example of cross section approximation $\Sigma_{res,a,1}(Bu, T_{f,0}, C_{B,0})$ using the industry set support $\mathcal{X}_\star$. An approximation with a large λ is unable to follow the data, while a low λ can cause numerical errors that generate offsets in the approximation.

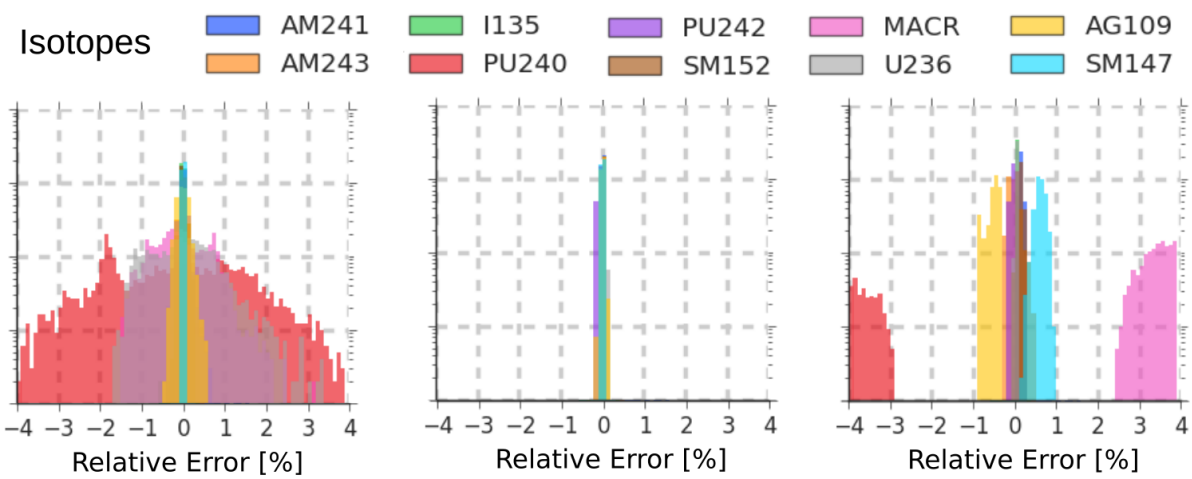
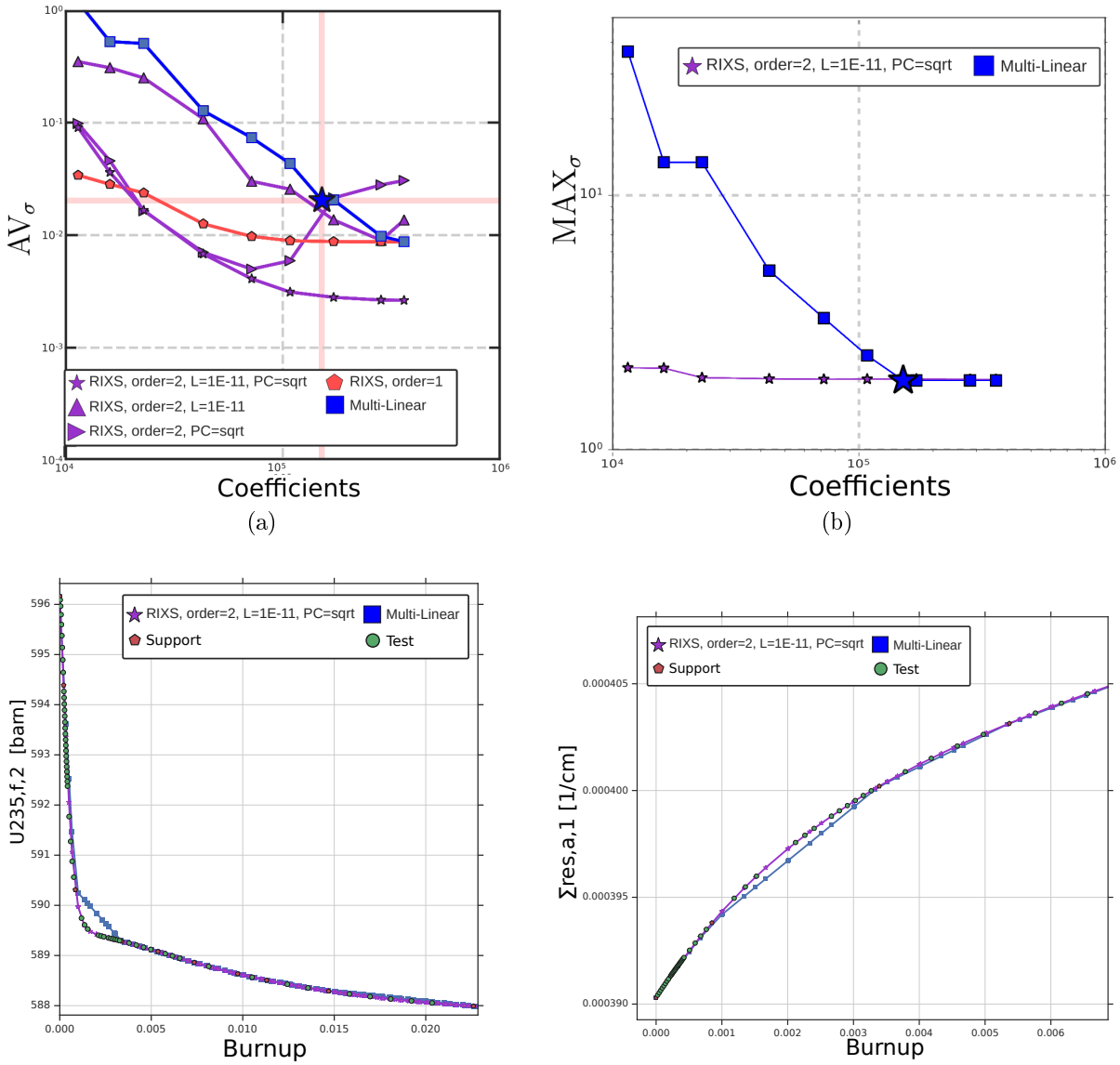


Figure 4.7: Error histogram of fast absorption cross section discriminated by isotope using different regularization values for a high order spline kernel in a full grid. From left to right: a high bias approximation (λ =1E-4) has centered means but very large standard deviations since the approximation fails to follow the data. Adequate compromise (λ =1E-11) with centered means and acceptable errors. High variance approximation (λ =1E-18) where possible numerical errors generate significant offsets.

the maximum error of k_∞ passes from 500 *pcm* with multi-linear in a Cartesain grid to 15 *pcm*. This is the type of approximation searched for: a local, smooth, that (practicality) interpolates the data, and can cope with strong local variation being in practice, bounded by the arbitrarily chosen support.

Similar results are presented in the Appendix B.2 for the Bernoulli kernel where third order kernel are explored as well, not leading to further improvement.



(a) (b) (c) (d)

Figure 4.8: RKHS approximation using a high order spline kernel with preconditioning (square root) and a regularization (λ=1E-11) subject to active learning with RIXS loss functions. (a) Average microscopic cross section error in $\mathcal{T}_{flat}$ monotonously diminishes with increasing support resulting in an order of magnitude gain in library size or accuracy. (b) Maximum error for $\mathcal{T}_{Xe}$ is virtually flat with the support increase suggesting an error bounded by the support. (c) and (d) the approximation method is able to follow strong variation of cross sections without oscillations.

Approximation	MAX_σ	AV_σ	MAX_Σ	AV_Σ	MAX_{k_∞}	AV_{k_∞}	Library Size	
	[%]				[*pcm*]		Coef.	Δ[%]
Multi-linear	1.9	2.1E-2	3E-1	1.8E-2	90	15	$1.5\mathrm{E}^5$	-
RKHS ($m = 2$)	1.9	3E-3	1.5E-1	1.8E-3	8	0.6	$2.1\mathrm{E}^4$	86
RKHS ($m = 1$)	1.9	9E-3	3E-1	6.5E-3	18	2	$1.1\mathrm{E}^4$	92

Table 4.4: High order spline kernel with preconditioning (square root), regularization (λ=1E-11) using the RIXS loss function for active learning. Average and maximum errors for σ, Σ, k_∞ at constant support $\mathcal{X}_\star$. Library size at constant error $\mathrm{AV}_{\Sigma,\star}$.

4.9 Conditioning of the Gram matrix

In RKHS the potentially badly conditioned system of Eq. 4.10 needs to be solved to find the approximation's coefficients. It's known that for a support size larger than 3000-5000 points, inversion problems can rise [79]. In this work the inversion is performed with the Moore-Pernose pseudo inverse routine that uses a Singular Value Decomposition with a default tolerance of λ =1E-15 [81].

The conditioning number increases with the order of the spline kernel as shown in Tab. 4.5 and with the support size as shown in Appendix B.2.1. The conditioning number strongly depends on the size of the support. For the range of interest using a first order spline kernel it ranges from 1E6 to 1E10. When increasing the order of the approximation very high conditioning number leading to numerical errors in the coefficients obtention are observed without adequate regularization. Active Learning procedure does not affect the conditioning number suggesting that a high density of point near each other plays a secondary role. Preconditioning provides a marginal benefit.

4.9.1 Kernel method's coefficient distribution and shrinking

In The Appendix B.2.1 the coefficient module are plotted showing that they are indirectly proportional to λ thus arriving a possible numerical error for very low regularization. Rejection techniques discarding the coefficients with the lowest module are not normally contemplated in RKHS and the expansion basis does not form an orthonormal system. When discarding coefficients of low absolute value ($\sim$ 0.1% of the total) a steep error increase was observed, specially for the second degree spline kernel. This high coefficients values are possibly a result of ill condition systems and is a weakness of the method.

4.10 Investigation on evaluation speed

Another potential drawbacks of RKHS is related to the evaluation speed since all the observations in the support participate in Eq. 4.7. The cases of interest analyzed in this work range $|\mathcal{S}| \sim$ 100-1000. Active learning has a major influence in performance by keeping this number low.

Spline kernel order (m)	Support	Regularization (λ)	Pre-conditioning	Conditioning number
1	$\lvert\mathcal{S}\rvert$=2.2E4	1E-15 (defualt)	No	1E6
1	$\mathcal{S}_\star$	1E-15 (defualt)	No	1E10
1	$\mathcal{S}_\star$	1E-15 (defualt)	sqrt(Bu)	1E8
2	$\mathcal{S}_\star$	1E-18	No	1E20
2	$\mathcal{S}_\star$	1E-11	No	1E14
2	$\mathcal{S}_\star$	1E-15 (defualt)	No	1E20
2	$\mathcal{S}_\star$	1E-15 (defualt)	sqrt(Bu)	1E19
2	$\mathcal{S}_\star$	1E-15 (defualt)	sqrt(all)	1E18
2	$\mathcal{S}_\star$	1E-4	No	1E7
2	$\mathcal{S}_\star$	1E-11	sqrt(all)	1E14
2	$\mathcal{S}_\star$	1E-15 (defualt)	sqrt(all)	1E20

Table 4.5: Conditioning number of the system $(K + \lambda I)$ required to find the coefficients $\vec{\alpha}$ for the RKHS approximations. Preconditioning marginally improves the conditioning number while active learning doesn't change it. A significant worsening is observed when increasing the order of the approximation (m). The default limit of Moore-Pernose inversion routine is 1E-15.

The evaluation speeds depends on the amount of support point and the cost of computing the kernel. In this work a Python2.7-F2PY-Fortran[8] implementation was developed in an Object-Oriented approach. In what follows time estimates are provided which are of course machine dependent though they convey an order of magnitude of evaluation speeds or at least an upper bound.

4.10.1 Evaluation speed of independent cross section models

A first numerical experiment is simply considering, for a single cross section, the evaluation time increase with the support size $N = |\mathcal{S}|$. A total of 1E5 points coming from a sequence of random number in $[0, 1]^d$ are serially evaluated in a Python loop. A linear increase in evaluation time $\hat{t}$ in function of the support size N is observed

$$\hat{t}_N = \frac{\partial \hat{t}}{\partial N} N + \hat{t}_0. \tag{4.22}$$

The slope $\frac{\partial \hat{t}}{\partial N}$ and the intercept $\hat{t}_0$ is presented in Tab. 4.6 for RKHS approximations and also Splines using B-splines. A $\hat{t}_0$ ~5E-5 s for both B-splines and RKHS is observed which could be condition by the Python loop itself. Still, RKHS evaluation times do increase linearly with the support size and proportionally to the order of the kernel. Specific evaluation times are presented for the support sizes obtained with active learning ($N \sim 150$) and the industry set ($N \sim 1000$). Due to the Fortran routines used by

[8]The object-oriented classes developed to test these methodologies allows to import compiled modules for performing the most demanding computations since an interpreted lenguage such as Python is intrinsically slower. All kernels operations, tensorization, inner product, and matrix calculation routines were written in Fortran 90 [82] and imported to Python using the F2PY package [83]. The script is fully available in [84].

Approximation		Derivative	Intercept	$\hat{t}$ for [s]	
Type	Order	$(\frac{\partial \hat{t}}{\partial N})$	$(\hat{t}_0)$	$\lvert\mathcal{S}\rvert$ =1E2	$\lvert\mathcal{S}\rvert$ =1E3
B-spline (Multi-linear)	1	4.1E-10	6.9E-5	6.9E-5	6.9E-5
B-spline	2	1.3E-9	6.8E-5	6.8E-5	6.8E-5
RKHS	1 (Spline)	1.2E-7	2.6E-5	4.8E-5	1.4E-4
RKHS	2 (Spline)	1.3E-7	2.6E-5	4.8E-5	1.6E-4
RKHS	1 (Bernoulli)	5.9E-7	2.5E-5	1.2E-4	6.2E-4
RKHS	2 (Bernoulli)	7.7E-7	2.1E-5	1.4E-4	8.0E-4
RKHS (accel.)	1 (Spline)	2.3E-9	1.0E-5	1.0E-5	1.2E-5
RKHS (accel.)	2 (Spline)	2.4E-9	1.0E-5	1.0E-5	1.2E-5
RKHS (accel.)	1 (Bernoulli)	5.7E-9	1.0E-5	1.1E-5	1.7E-5
RKHS (accel.)	2 (Bernoulli)	6.9E-9	9.9R-6	1.1E-5	1.8E-5

Table 4.6: Evaluation speed in seconds per cross section evaluation point for RKHS and B-splines approximations. First and second order approximations are used for B-spline and RKHS with a spline and Bernoulli kernel. The derivative marks the evaluation time increase with the support. Accelerated RKHS using a common evaluation vector are almost independent of the support in the considered range of up to 2500 coefficients. Two support sizes of interest are ∼150 coefficient (satisfactory error with active learning) and ∼1500 coefficients (Industry set support) for which evaluation times are shown.

the approximation class in Python, for a small enough support of $N \sim 150$ the RKHS implementation is actually faster than B-splines from the package [69]. This result does moderately deteriorates with increasing the support size. Preconditioning, regularization and different grid distributions due to active learning have negligible effects on the final evaluation speed.

For a given support, the actual calculation time of a cell in the core could be compose of: compute the basis vector ($compute(\vec{B})$), find the coefficients ($find(\vec{\alpha})$), perform the inner product ($\langle \vec{B}, \vec{\alpha} \rangle$). This operations is performed $|\mathcal{Y}|$ times since the entire cross section set is needed. Indeed B-splines scale very well with the support size since they are local basis with a few terms different than zero and so, $compute(\vec{B})$ is independent of the support. However searching the coefficient vector in disk memory in a large library size ($find(\vec{\alpha})$) could severely hinder evaluation times. On the same token performing the operation $\langle \vec{B}, \vec{\alpha} \rangle$ cache memory is practically independent of the vector size itself. In this scenario the library reductions obtained with active learning could significantly surpass the convenience of B-spline basis.

4.10.2 Evaluation speed of cross section models with common evaluation vector

In a shared support scheme using the same function space for every cross section, the same operation $compute(\vec{B})$ is performed $|\mathcal{Y}|$ times. An acceleration technique could be to store this evaluation vector in the RAM memory. This conforms the "accelerated" methods in Tab. 4.6 for RKHS. All resulting evaluation times are smaller than with B-

splines. Though they don't share the evaluated basis for the cross sections[9] it's composed of only a few terms. Especially in this scenario it's observed that all method have relatively similar evaluation speed within the targets, yet noticing that order 1 is faster than order 2 and that the Spline kernel is faster than the Bernoulli kernel.

This type of optimization is at the expense of RAM memory which may be limited. If a different function space was used for each cross section, as it's the case in [57] or without the share support condition, this type of acceleration would not be possible.

4.10.3 Active learning computing time

In view of these evaluation times, let the amount of average kernel evaluation for a single cross section be ∼1.1E10 as presented in Section 4.5.2. If a single non tensorized kernel evaluation is ∼ 6E-8 s, then loss functions considering the 144 cross sections have an active learning time of $6.6\text{E-}8 \times 1.1\text{E}10 \times 144 \frac{h}{3600} = 31$ h. Considering a much larger pool of data for this study would have raise difficulties[10].

4.11 Final thoughts and discussion

RKHS employed a scattered support which is an effective tool to deal with the curse of dimensionality, further enhanced by dropping the share support condition[11]. Satisfactory higher order approximation schemes compatible with active learning improved the approximations accuracy. This type of methodical analysis has not been addressed in the literature before, especially comparing to multi-linear interpolation in a Cartesian grid as used in industry applications. The main results are:

RKHS in cross section modeling

- Kernel methods can utilize unstructured supports allowing to wage machine learning optimization techniques. This is a quite synergic combination: active learning has a complete freedom to select the observations, whose quantity minimization is crucial to improve the conditioning number of the Gram matrix and the evaluation speed.
- This is a realistic supervised learning scenario in a "one-trough" two-step calculation scheme, where rich sets of cross section data are available, but an iterative process between lattice calculations and library preparation is discouraged.
- Kernels are "observation based" similarity function that don't require to be defined over an hyper-dimensional rectangle encompassing the input space, which is a

[9] This option or the possibility to provide evaluation vector was not available in the B-spline package [69].

[10] To perform these sensitivity studies an elemental Message passing interface was implemented in Python [85]. In this way each active learning process could utilize a separate CPU.

[11] Library size reduction were analyzed at constant $\text{AV}_{\Sigma,\star}$ instead of $\text{AV}_{\sigma,\star}$ to facilitate the comparison with no shared support using the importance

problematic limitation of traditional expansion methods [27].

- Though this work was centered in kernels reproducing local approximations, a large variety of other functional spaces can be reproduced with RKHS. Kernel design could lead to very specialized approximation spaces in a similar vain of [57]. Other kernel functions, provided by [78], did not excel as "out of the box" tools, further solidifying the interest in piece-wise local function spaces for cross section approximation.

- The linear spline kernel reproduced muli-linear interpolation as commonly used in industry, but without limitations with respect to the support structure or regularization. Profiting from the linear error bounds it was used for assessing different loss function in Pool active learning.

Active learning with a first order spline kernel with a shared support

- Loss functions acting on σ, Σ, and k_∞ were assess in terms of their performance: accuracy gains, library size reduction, and computational cost. All active learning strategies improved the cross section library showing the robustness of the methodology.

- A random selection (RAND) of the support already produced a significant improvement as expected: collinearity undermines the informativeness of an approximation's support [29, 22]. The gains in accuracy rivalizes with higher order splines utilized in chapter 3 reducing the library size by half.

- The loss function that considers the relative error of all the cross sections (RXS) exhibited the lowest AV_σ performing well in AV_Σ and AV_{k_∞}. Relative errors are required to make meaningful comparisons among cross sections.

- Absolute values, even if weighted with the importance, result in the participation of only one cross section $\sigma_{135,a,2}$. The computation effort for all the others is thus wasted. Nonetheless the grid distribution obtained is rather similar to the corresponding to U, which produces near optimal results.

- The loss functions U, RXS, XSI exhibit a significant density of point near the first burnup values, where many cross section show strong variations. On the other hand when considering the concentration (RXSI, M, MF) more points are allocated towards the middle burnup region. These method also need the evaluation of the entire cross section set without a tangible benefit. For this use case optimizing for integral parameter as done in [14] yielded no additional gains.

- RXSI uses the importance to select a subset of cross sections resulting in the best compromise between maximizing the information for the cross section's set and the computational cost of the active learning procedure.

Active learning with a first order spline kernel without a shared support

- In a shared support scheme large amount of unnecessary data can be included in the model as manifested by the clustering of cross section participating in the active

learning process. Long plateaus of partial loss function values were observed for many cross section.

- By dropping the shared support condition, an independent active learning per cross section (NoS) was carried out further improving the cross section library. This type of gains could be considered as an *a priori* optimization equated to *a posteriori* rejection techniques, but noting that in the former, no calculation effort is wasted in discarding terms.

- By additionally adjusting the error tolerance by the importance (ImpNoS) a significant library size reduction is obtained. Large error in many microscopic cross sections are noted however, specially for scattering cross sections. Since it's not possible to know in advance which microscopic cross section will be required in a given simulation, this type of optimization could prove unsatisfactory.

- Though integral parameters such as k_∞ or Σ have important physical meaning, the benefit of keeping track of the original functions in multi-objective optimization is showcased by this example (ImpNoS).

- Without the shared support condition for active learning using a fix tolerance (NoS) and relative error, it was observed an increase in the ratio of fast to thermal coefficients when decreasing the error tolerance. It's possible that the modeling difficulty shifts from the cross section's variance to numerical artifacts in some fast cross sections.

High order RKHS approximation with spline kernels

- With preconditioning and regularization the high order spline kernel combined with active learning presented a monotonically decreasing error with the support size, achieving an order of magnitude improvement with respect to multi-linear.

- The maximum error of this approximation in the $\mathcal{T}_{Xe}$ was virtually flat being thus, in practice, a local second order approximation bounded by the support. This type of smooth approximations are of particular interest for core feedback coefficient [64].

- Thus, the main limitations discussed in chapter 3 are effectively dealt with by this methodology. Similar results were obtained with the Bernoulli kernel at a slightly higher computational cost, though the first order does not exactly reproduces multi-linear interpolation. An order 3 approximations did not improve the results further.

Performance and evaluation speed of RKHS approximations

- Though active learning can be considered an "off line" task, a computation time of up to 1 day is considerable. On the other hand this methodology assures an optimal support since all available data is evaluated in each step. In other works [14] the optimization is a one dimensional iterative process that reaches a local minimum in $\mathcal{L}$. If only one cross section is considered near optimal results are obtained in 15 minutes.

- Loss function profiles with increasing support allow to quantify the amount of useful information in the data set and its extraction rate providing a direct mesure of the error (in the pool) similar to other methodologies [53, 29].

- Especially without a shared support, it could be argued that the amount of required observations in RKHS with active learning is proportional, not to the dimensionality of the input space but to the cross section's complexity e.g. expressed in terms of the second derivative $|\mathcal{X}_\dagger| \sim \sigma^{(2)}$. Moreover if an increase in the quantity of cross sections is correlated to a mean decrease in their complexity, then the aggregated set of required observations (i.e. the library size) could remain bounded. This suggests a rather scalable technique with both the input and output space increase.

- For this use case no more than 100 observations per cross section were required with active learning and up to 2500 were sucesfully tested. In more challenging use cases the amount of observations usually lays between 1000 and 1600 [55] which falls well beneath the limit of ∼5E3 under which RKHS is effective [79].

- The evaluation cost of RKHS is linear with the support size and the slope proportional the kernel's evaluation cost. In the range of 100-1000 terms RKHS approximation had acceptable evaluation speeds being actually faster than splines using B-splines for 100 terms.

- In actual core calculation the entire cross section set is required. Sharing the evaluation vector among the cross section at the expense of RAM memory resulted in an evaluation speed under 1E-5 seg for any support and kernel. The benefits of such acceleration would increase with the output space cardinality.

Though this is a rough, machine-dependent estimation, it showed that Kernel Methods can be competitive performance-wise for the functional space of interest in cross section modeling.

Chapter 5

Artificial Neural Networks

In this chapter, artificial neural networks are used for cross section modeling. They adapt to data trough a supervised learning process, called training, without requiring the formulation of explicit relations among the variables, being naturally able to cope with high dimensionality in both the input and output spaces. Firstly, a parametric study on learning hyper-parameters, network's size, and topology is carried out selecting models of interest to test with the industry set. Results are compared to multi-linear interpolation in terms of library size reduction, accuracy and evaluation speed. Multi-output models where the entire cross section library is modeled with a single network are explored as well.

5.1 Machine learning

A modeling methodology that is currently leading a machine learning revolution is Deep Learning [86] using Artificial Neural Networks (ANN) which are computing systems composed of processing elements, called neurons, that connected to each other, emulate their biological counterparts. State-of-the-art implementations in artificial intelligence, computer vision, natural language processing have achieve outstanding performances [86]. This is mainly driven by [87] data availability (e.g. internet of things), hardware power increase at a reduced cost (GPU [88], TPU) and breakthroughs in learning algorithms (back-propagation and its implementations) coupled with free license and mature frameworks (TensorFlow [89], PyTorch [90]). For regression problems they are sometimes referred to as black-box universal approximators since a reasonable, well behaved function, can theoretically be approximated up to an arbitrary degree with an adequate artificial neural network [91]. Some of the main characteristics are:

- Artificial neural networks are able to approximate non-linear function without direct user intervention being quite robust with noise, and imposing no prior task-specific rules which are derived from training, a supervised learning task. Indeed the functional form requires no explicit formulation among the variables prior to the training process, in contrast to classical techniques.

- These learning systems are naturally able to cope with high dimensionality in both the input and output spaces; sometimes defined as "A computational mechanism able to acquire, represent, and compute mapping from one multi-variate space of information to another, given a set of data representing that mapping" [92].

- Though the definition of an adequate feature representation (i.e. choosing the input space variables) is not dealt with in this thesis, a general purpose representation system such as multi-layered networks can identified the features automatically (auto-encoders) [86]. This is one of the main reason of their success in many fields of computer science [87].

- The training process, which can naturally be performed on-line (mini-batch setup), can lead to neurons having a small subset of connections resulting in local approximations.

- Consolidated frameworks to utilize and train artificial neural networks are readily available [89, 90] facilitating the use of GPU acceleration. Easy-to-use high level packages are also available (Keras [93]). After the training process, that might be platform dependent if it uses GPU acceleration, only the network's parameters require storage.

Some of the shortcomings are:

- Possibly long training times depending on the quantity of data and the size of the network. In principle, convergence is not guarantee [92]. Due to their stochastic nature of the optimization algorithms different sessions may result in different models.

- The amount of architectural and training hyper-parameters to be defined is quite large, depending as well in the chosen training algorithm. Time consuming parametric studies are usually required and dwelling into the details of the training algorithm and how they condition the emerging network model can be challenging.

- It's hard to either impose or extract physical insights onto the model as well as providing an interpretation on the network's structure. Providing error bounds can be difficult.

5.2 Artificial neural networks in nuclear engineering

They have been applied to a broad range of problems such as core parameter prediction and control, reactor kinetics [94], fuel management optimization [95], heat transfer modeling [96, 97], and inventory estimation [98]. In general, rather classical machine learning settings have been used: feed-forward fully connected multi-layer networks of only a few

layers[1]. In the specific field of cross section modeling their use has been quite limited. Good performances were obtained in [58] but using simple architectures with only one state-parameter, the burnup.

5.2.1 Feed-forward fully connected artificial neural networks

In a feed-forward network the information travels forward through the layers as the name implies. They are sometimes called *Shallow* if composed of a single layer or *Deep* or multi-layer for more. Neurons between layers are *fully connected* by weights and non-linear *activation functions* $f(x), x \in \mathbb{R}$ complete the output of each processing unit (neurons) enabling the network to learn complex patterns. A neuron processes a linear combination of inputs, weights and biases in the form $f(\vec{w} \cdot \vec{z} + b)$ with $b \in \mathbb{R}$ and $\vec{w}, \vec{z} \in \mathbb{R}^{N^{l-1}}$ being N^{l-1} the number of neurons of the previous layer. The aggregated output of all the activation function in a layer is

$$f^l : \mathbb{R}^{N^{l-1}} \to \mathbb{R}^{N^l}, \quad 2 \leq l \leq L. \tag{5.1}$$

For L layers the output of an network is

$$\begin{aligned} \hat{y}(\vec{x}) = f^L \circ f^{L-1} \circ \cdots \circ f^1(\vec{x}), \\ \hat{y} : \vec{x} \in \mathcal{X} \to \mathbb{R}^O, \end{aligned} \tag{5.2}$$

being O the cardinality of the output space in the last layer. This is presented in Fig. 5.1 for a shallow network of 4 neurons $\hat{y} = \sum_{j=1}^{4} w_{2j1} f(\sum_{i=1}^{3} w_{1ji} x_i + b_{1j}) + b_{21}$ and a scalar output ($O = 1$).

The weights, that represents the strength of the connections, and the biases are determined by a supervised learning process called training that consists on the minimization of a *loss function*, $\mathcal{L}(y, \hat{y})$. A training set provides the samples used in a forward pass to evaluate the network and define the gradient of the loss function with respect to the parameters of the network, $\nabla_{\vec{w},b} \mathcal{L}(y, \hat{y})$. The gradients are obtained by a *back-propagation* algorithm which is a computationally convenient implementation of the chain rule [100]. For actually optimizing the parameters, stochastic gradient descent algorithms are used. Conceptually the operation is

$$(\vec{w}, b) \leftarrow (\vec{w}, b) - \alpha \nabla_{\vec{w},b} \mathcal{L} \tag{5.3}$$

where $(\vec{w}, b)$ are the parameters of the network. They are updated following the slope of the cost function downwards being $\alpha > 0$ the *learning rate*. It's called stochastic because the small set of randomly chosen examples gives an estimations on the gradient used to examine the parameter-space $(\vec{w}, b)$. One of the most popular choices for gradient descent optimization is the Adam (Adaptive moment estimation) [101] algorithm that includes an adaptive momentum term to accelerate convergence and avoid oscillations. It's considered

[1]Some exceptions are: in [95] a cascade-network in which each hidden layer receives an input from all previous layers is reported to accelerate the convergence rate. Extreme machine learning have been reported to have particularly low training times as the training uses ordinary least square instead of back-propagation [94]. Nucleide classification in radioactivity monitoring have recently utilized convoluted neural networks minimizing the feature selection process [99]. In another very recent work [96], properly deep neural network has been used to model the heat tranfer process in PWR.

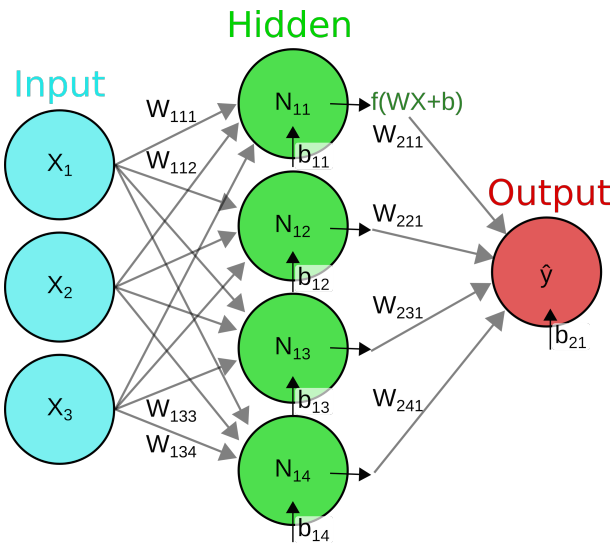


Figure 5.1: Scheme of a single-layer artificial neural network. The three type of layers can be seen in cyan the input (In), in green the hidden layer and in red the output (Out). The network is fully connected as weights link all the elements from layer to layer. This shallow network counts with 4 neurons and has an output in the form of $\hat{y} = \sum_{j=1}^{4} w_{2j1} f(\sum_{i=1}^{3} w_{1ji} x_i + b_{1j}) + b_{21}$ with the weights and biases learned by training.

one of the best optimizes currently available also having intuitive hyper-parameters, capability of using non-stationary loss functions and automatic learning rate annealing and low memory requirements[2]. It's used in this thesis following literature recommendations [96] and for testing "out of the box" tools, readily available with PyTorch. Is customary to divide the training data into *batches* and a iteration on the whole training set completes an *epoch*.

5.3 Artificial neutral networks sensitivity studies

Even for a classical feed forward networks the quantity of problem-dependent parameters to consider is rather large [92]. The design itself requires to define the amount of hidden layers, neurons per layer, activation function, parameter's initialization and possible regularization layers. Pre-processing of data is mandatory. Training hyper-parameters must be chosen as well: learning rate, number of batches, loss function with possible regularization, etc. The design of an optimal network remains a difficult and multistage iterative process due to the high amount of parameters, the wide scale in which they can vary and the strong inter-dependence among them. This defines sensitivity studies to perform.

5.3.1 Cross section subset

Only the cross sections $\sigma_{235,f,2}$ and $\Sigma_{res,a,1}$ are considered in order to reduce the computational burden of the sensitivity studies. These two important cross section are different in

[2]Default gradient decay parameters of the PyTorch package are used: $\beta_1 = 0.9$ and $\beta_2 = 0.99$.

kind, being one a microscopic cross section and the other one a residual one. They are also of different energy groups and reaction type. A separated network is used to model each. A single layer network of neurons $N = 20$ with the activation function $f(x) = tanh(x)$ is the default model following literature recommendations [98].

5.3.2 Implementation setup

The PyTorch package was used in this work [90] and the developed script is available in GitHub under MIT license [102]. In order to follow typical training setups, and in view of the high performance of these packages, data was divided in a randomly sampled *Train/Test* split of 80%/20% [3]. Train data are further divided in 5 randomly sampled batches [4] and thus 5 network adaptations occur per epoch. This helps the optimizer to exit possible local minima adding also a regularization effect to the network. The parametric studies are performed by analyzing the RE_σ error in both the test and training set up to 1E5 epochs. Comparing these two errors shows the generalization capability of the model.

Data normalization

Data normalization has a strong effect in the gradient's absolutes values and thus in the overall training process. The average relative error RE_σ is presented in Table 5.1 for different normalizations of $\mathcal{X}$ and $\mathcal{Y}$. Without any cross section normalization, the optimizer is unable to converge with errors of up to 100%. Normalizing by the $\max(\sigma(\vec{x}))$ bounds cross section values to 1. $\mathcal{X}$ is already contained in the unit hyper-cube and errors are marginally decrease when considering $\sqrt{\mathrm{Bu}}$ in order to standardize the statistical distribution of burnup data. Normalizing by the variance exhibits the best results, slightly further improved by using $log(\sigma)$.

Specially for deep neural networks the initialization of the parameters can severely condition the training. If the weights are too small the forward pass suffers from diminishing values that may nullify the final output preventing the computation of the gradients. For exessive initial weight values the variance grow between layer may compromise the learning algorithm. However for these small networks different initializations of the weights using Random, Normal, Uniform, or Xavier-Uniform distributions (see [90]) were analyzed with negligible impact in the RE_σ and $\mathcal{L}$ profiles.

[3] This is the recommended scheme if the original data set is dense enough to properly analyze the generalization error [100, 96]. Due to time constraints and in view of the results in previous chapter only microscopic cross section are analysed. The computational cost of other arrangements such as k-fold cross validation was considered excessive.

[4] If the gradient utilized to update the network's weight is calculated for each data point the methodology is called stochastic gradient decent. If the entire training set is used it's sometimes called batch gradient decent. If the training data is divided in subset as done in this thesis, the learning method is called mini-batch gradient descent.

Normalization		RE_σ	
Input space	Output space	$\sigma_{235f,2}$	$\Sigma_{res,a,1}$
-	$\sigma/\max(\sigma)$	3.0E-1	2.1E-1
$x_1 \leftarrow \sqrt{Bu}$	$\sigma/\max(\sigma)$	1.5E-1	1.5E-1
$x_1 \leftarrow \sqrt{Bu}$	$\sigma - \text{mean}(\sigma) \leftarrow \sigma/\max(\sigma)$	1.0E-1	1.5E-1
$x_1 \leftarrow \sqrt{Bu}$	$(\sigma - \text{mean}(\sigma))/\text{std}(\sigma)$	2.0E-2	5.0E-2
$x_i \leftarrow x_i - \text{mean}(x_i)$, $1 \leq i \leq d, x_1 \leftarrow \sqrt{Bu}$	$(\sigma - \text{mean}(\sigma))/\text{std}(\sigma) \leftarrow log(\sigma)$	2.0E-2	3.0E-2

Table 5.1: Cross section relative error (RE_σ) using different normalizations for a single-layer network with 20 neurons at 1E5 epochs. Utilizing the variance of σ combine with the *log* exhibits the lowest error. Without any normalization the network in unable to learn.

Learning rate

The learning rate (α) modulates the actualization of the weights influencing the convergence rate and the optimizer's capability to escape possible local minima. If it is too high the algorithm may not converge and if it is too low it may do so too slowly. This behavior is confirmed. In Fig. 5.5 for an α =1E-1 errors fluctuated up to 2 orders of magnitude around a RE_σ of 5E-1 while for an α =1E-5 it converges smoothly but very slowly and possibly getting trapped in a local minimum. A value $\alpha = 0.001$ is chosen as a good compromise which systematically provides the lowest error ([1E3,1E5] epochs) though with a somewhat slow initial convergence (up to 1E3 epochs). Different batches of 1, 5, 100, and 1000 number of training points per batch were tested. For this use case, the totality of the training data was able to fit into GPU memory which is a typical training limitation[5]. A single batch exhibits relatively slow convergence and slightly higher errors, whilst with 1000 batches stronger error fluctuations with increase computational cost was observed. A value of 5 is chosen as good compromise. For every learning rate a complete overlap between test and training error is observed. In the 80%/20% split, abundant training data avoids over-fitting while at the same time, the smooth nature of cross section data, makes the training quite representative of the test data.

[5]Limitation that may condition the model itself. The batch size can have a regularization effect which can affect what the optimal training hyper-parameters and network's topology (e.g. use of regularization layers). It could also limit the fraction of the problem that can be modeled with a single network, thus changing the scope and therefore characteristics of the network. Hardware limitation ought to be considered in the iteration loop of approximation with artificial neural networks.

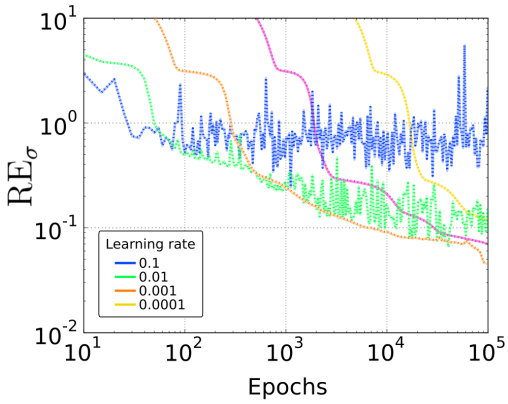


Figure 5.2: $\Sigma_{res,a,1}$ relative error (RE_σ) in training using different learning rates for a single-layer network with 20 neurons. The most stable and fastest convergence (after 1E3 epochs) is observed for a learning rate $\alpha = 0.001$. The test error overlaps with the training error.

Activation functions

The activation function influences the convergence rate during training and the evaluation cost of the network. The Sigmoid function has been historically popular due to its biological interpretation. However, it can compromise the training of deep networks due to a the possible saturation problem in very deep networks. The Tanh activation function is preferred since the smooth output is contained in [-1,1] with a zero mean. The Relu activation function does not suffer from saturation, and is very fast to evaluate though not smooth. However, specially for very small networks, they may generate discontinuities in the derivatives and "dead neurons" as $\text{Relu}(x) = 0, \quad \forall x < 0$ caused by a bad initialization or learning rate. Some of these issues are addressed by the ELU activation function used in deep learning [96] though it's more expensive to evaluate. In Table 5.2 the relative error RE_σ at 1E5 epochs is presented for these activation functions. Relu and Tanh are presented in Fig. 5.3. With the exception of HardS all exhibit a similar performances. Tanh is retained as recommended by the literature.

$f(x)$		RE_σ	
		$\sigma_{235,f,2}$	$\Sigma_{res,a,1}$
HardS	(see [90])	3E-1	3E0
Relu	$max(0,x)$	2E-2	1E-1
Elu	$max(0,x)+min(0,e^x-1)$	2E-2	6E-2
Sigmoid	$(1+e^{-x})^{-1}$	2E-2	6E-2
Tanh	$(e^x-e^{-x})(e^x+e^{-x})^{-1}$	2E-2	3E-2

Table 5.2: Cross section relative error (RE_σ) using different activation functions for a single-layer network with 20 neurons at 1E5 epochs. With the exception of HardS all activation function have a similar performance.

[5]The HardS activation function is defined as $HardS = \begin{cases} x, & \text{if } x > 0.5, \\ x, & \text{if } x < -0.5, \\ 0, & \text{otherwise.} \end{cases}$

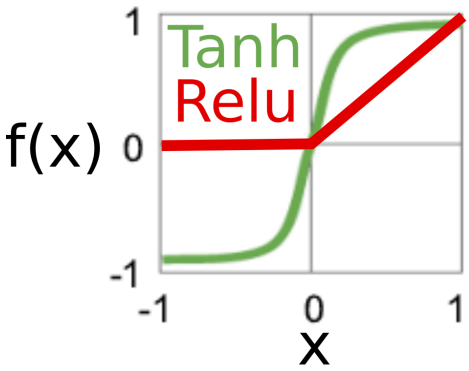


Figure 5.3: Non linear activation function for artificial neural networks. The Tanh activation function produces a smooth output contained in [-1,1] with a zero mean. Relu are very fast to evaluate though not smooth.

Loss functions

The loss functions considered are presented in Table 5.3. With respect to the RE_σ, a metric based on the absolute value of the error, an L_1 loss function showed the bests results, as expected [95].

$\mathcal{L}(\hat{\sigma}, \sigma)$	RE_σ	
	$\sigma_{235,f,2}$	$\Sigma_{res,a,1}$
mean($\lvert\hat{\sigma} - \sigma\rvert$)	2E-2	3E-2
SmoothL1Loss (see [90])	2E-2	5E-2
mean($(\hat{\sigma} - \sigma)^2$)	2E-2	4E-2
max($\lvert\hat{\sigma} - \sigma\rvert$)	1E-1	1E-1

Table 5.3: Cross section relative errors (RE_σ) using different loss functions for a single-layer network with 20 neurons at 1E5 epochs. With the exception of using the maximum error all loss function present a similar error.

5.3.3 ANN architecture

The *degrees of freedom* defined by network's topology shapes its approximation power, provided the proper training hyper-parameters. Though some rules of thump exist, parametric studies are required since the optimal structure ultimately depends on the training set size, the algorithm, and the function's characteristics such as noise, complexity, and regularity [100]. Indeed " It's impossible to predict the optimal architecture for a given problem prior to modeling" [97].

Varying number of neurons for a single layer network

Fewer neurons imply a simpler parameter space to explore, faster training times, and a resulting smaller library size which is faster to process and evaluate. RE_σ errors for

N	RE_σ	
	$\sigma_{235,f,2}$	$\Sigma_{res,a,1}$
2	2E-1	5E-1
5	3E-2	2E-1
8	2E-2	4E-2
15	2E-2	3E-2
20	2E-2	3E-2
30	2E-2	3E-2

Table 5.4: Average relative error with training for single-layer networks with a varying number of neurons at 1E5 epochs. 8 neurons is the smallest number for which the error are within the target.

shallow network with varying number of neurons are presented in Table 5.4 and Fig. 5.4. Error profiles during training are quite similar for $N \geq 8$ at least up to 1E5 epochs, and the number of neurons $N = 8$ is the minimal size to reach the target accuracy[6]. By close inspection of Fig. 5.4 it would seem that errors stabilize inversely to the number of neurons of the network.

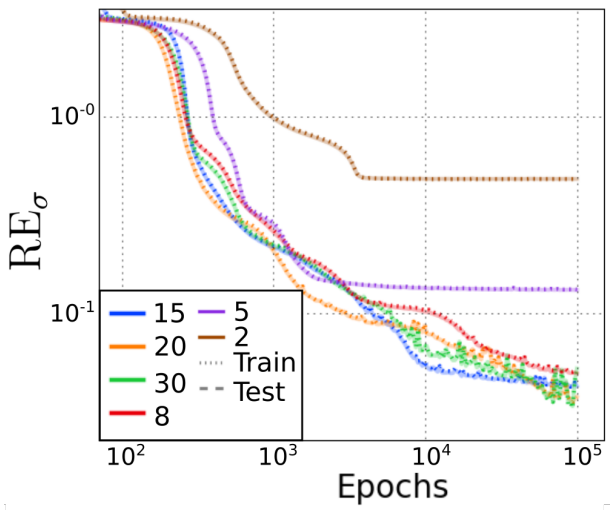


Figure 5.4: Average relative error in $\Sigma_{res,a,1}$ with training for single-layer networks with a varying number of neurons. It can be observed that with very few neurons the error stagnates as the model is not equipped with sufficient degrees of freedom to further adapt. Rapid variations are noted at with 30 neurons after 1E4 epochs. Due to the abundance of smooth cross section data in the 80%/20% split train errors overlaps with test errors.

Varying number of layers at constant library size

For a fully connected network with N neurons, L layers, I inputs, and O outputs the amount of network parameters is

$$|ANN| = (L-1)N^2 + (I+L+O)N + O. \tag{5.4}$$

[6]It has been observed that when using a rich data set the approximation power of even very small artificial neural network can significantly surpass traditional interpolation techniques [103].

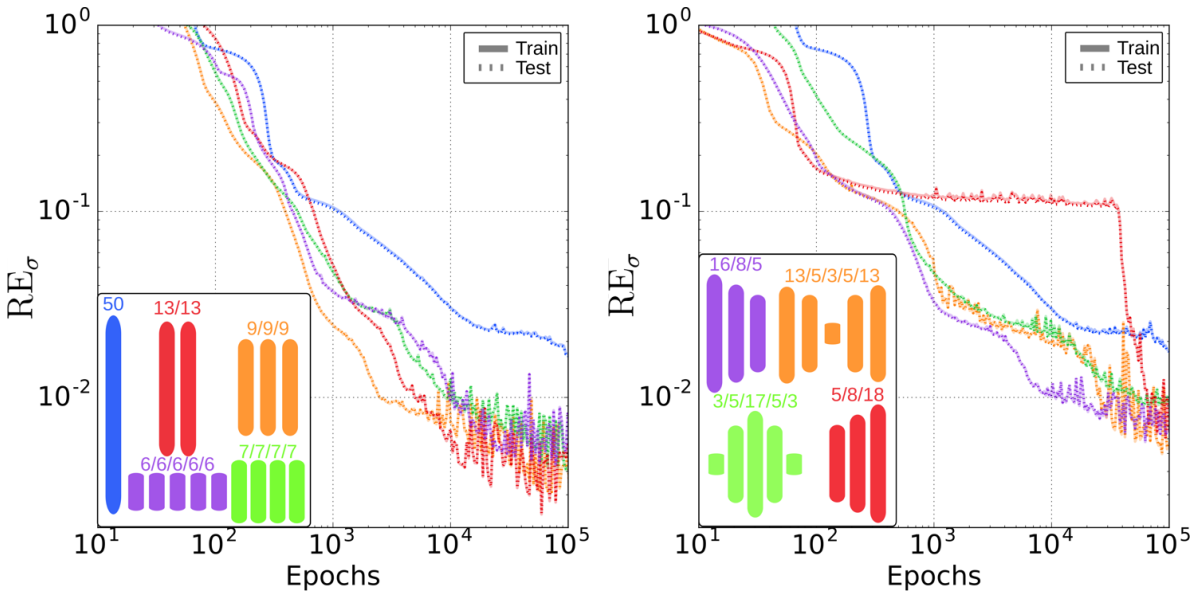
N	RE_σ	
	$\sigma_{235,f,2}$	$\Sigma_{res,a,1}$
5	1E-2	1E-1
10	4E-3	4E-2
13	1E-3	4E-2
15	2E-3	4E-2
20	4E-2	6E-2

Table 5.5: Average relative error with training for single-layer networks with a varying number of neurons at 1E5 epochs. 8 neurons is the smallest number for which the error are within the target.

For the models considered in these sections $I = 3$ and $O = 1$. If augmenting the number of neurons of a shallow network doesn't provide the desired accuracy, it's recommended to consider a two-layered topology [100]. Therefore, a parametric study on the network's layers, at constant library size by virtue of Eq. 5.4, was carried out. "Rectangular" designs of layers L =1, 2, 3, 4, 5 and neurons N =50, 13, 9, 7, 6 respectively are presented in Fig. 5.5a where each layer is separated by "/" in the scheme. The RE_σ errors for the $\sigma_{235,f,2}$ decreases monotonously up to 1E3 epochs. Then, fluctuations are noticed, possibly as the optimizer explores the parameter space to escape a possible local minimum. More than a single layer provides smaller errors and faster convergence, specially for 2 layers which coincides with other works that approximate similar real value continuous functions [95, 58]. Test and training errors overlap since cross sections tend to be smooth, noise-free functions, and training data very abundant in this 80%/20% split (with respect to cross section variance). In Fig. 5.5b other types of topologies commonly found in literature [104] are tested, but they do not offer any advantage under these settings. Similar results are obtained for other cross sections which are thus not shown. In [105] it's argued that a 2 layer network is a system able to produce "bumps" of arbitrary high and position and their number depends on the amount of neurons. Other works provide a similar interpretation by stating than two layers are required to properly model discontinuities in the data or strong variations [92]. More than two layers can induce an overly complicated parameter space that slows down error convergence [100] which is why "for some problems not only architectures with lower number of neurons suffice but they are able to produce better results than architectures with higher number of neurons" [97].

Varying the number of neurons for a 2-layer network

The RE_σ is presented at 1E5 epochs for two-layered networks with varying number of neurons. In [92] it's suggested that the maximum number of neurons in the hidden layer (of a two layered network) should be the double of the amount of inputs, 6 in this case. This is not confirmed being the optimal size 13 neurons.



(a) (b)

Figure 5.5: RE$_\sigma$ for $\sigma_{235,f,2}$ in function of training for different typologies at constant network size. (a) Constant number of neurons per layer. All networks with more than one layer perform better than the shallow network. Fluctuation in the error with training are noticed for more than one layer. The best results are obtained with a 2-layer network for which the lowest errors are attained. Increasing or decreasing the amount of layer at constant amount of parameters slows down error progression. (b) Common architectures with varying neurons per layer at constant network size. All of them perform better than a single-layer network, particularly the N =16/8/5 network, though do not outperform the N =13/13 network.

5.3.4 Qualitative analysis on the function space generated by artificial neural networks

The network's activation function influences the training process but also the quality of the approximation. Especially for regression problems, some works relate the activation function to a sort of function space used by the network [106]. For the minimal network size of 8 neurons, cross section approximation is presented for $\sigma_{235,f,2}$ in Fig. 5.6. The limitations of the Relu activation function for cross section modeling become apparent in Fig. 5.6a; it shares the multi-linear drawback of breaks in the derivative (in either of the two dimensions, see Fig 5.6c), but without being bounded by the support. With a higher number of neurons this phenomenon is attenuated but still present. On the other hand, the Tanh activation function results in a quite satisfactory approximation. It follows the training data reasonable well though, as observed in the zoom section of Fig. 5.6b, the cross section's rapid change is not followed exactly due to the stochastic nature of the learning process that introduces some regularization. The 2-dimensional plot is presented in Fig. 5.6d which is smooth and indistinguishable from the one corresponding to the network of 20 neurons.

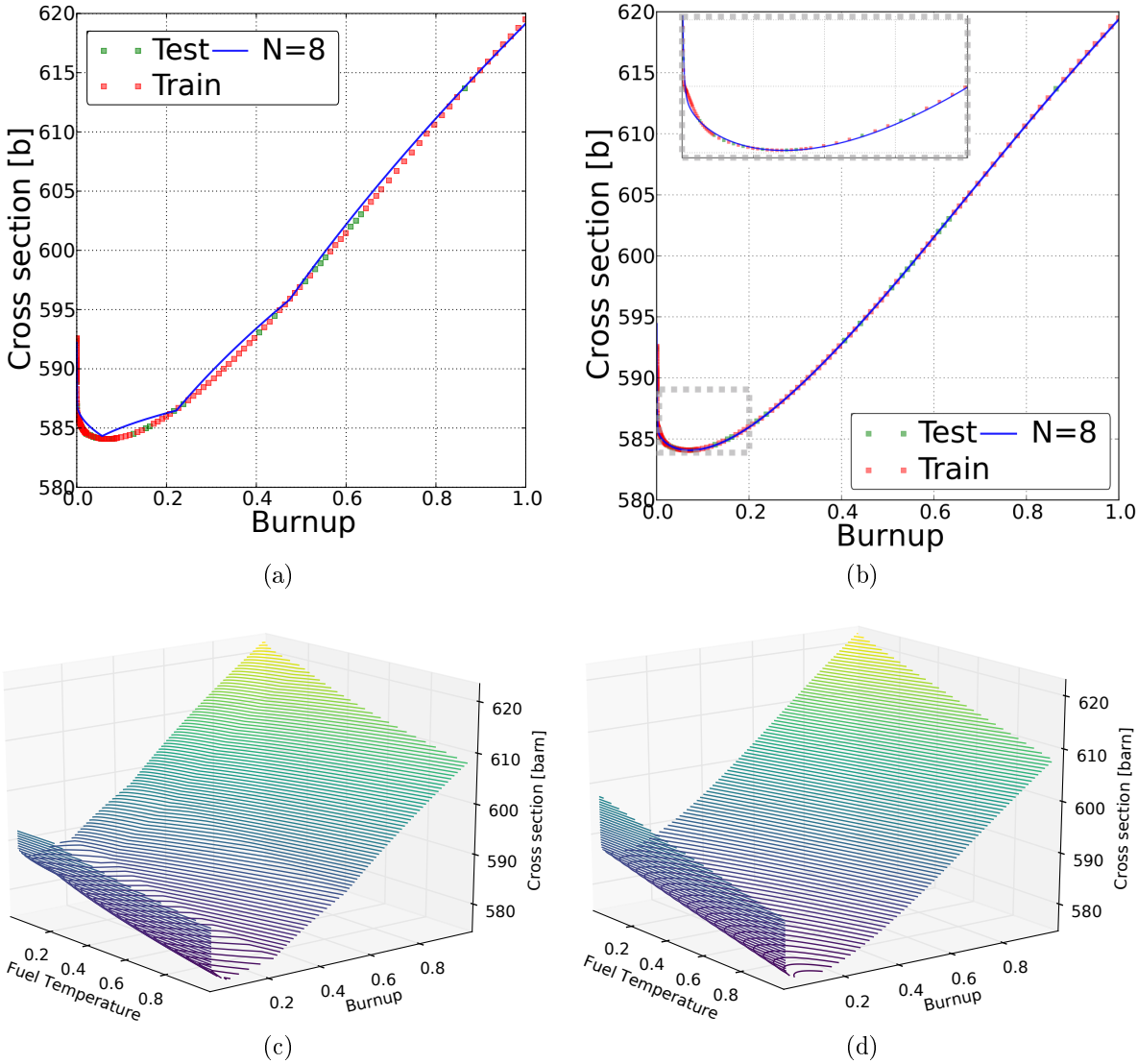


Figure 5.6: Shallow network of 8 neurons approximating $\sigma_{235,f,2}$. For the line (Bu, T_{f0}, C_{B0}): (a) Relu activation function. Breaks in the derivative of the approximations can be observed. (b) Tanh activation function which produces a smooth approximation that follows the strong variation in the data though introducing some regularization. For the surface (Bu, T_f, C_{B0}): (c) Relu activation function, resulting in breaks in the derivative in the two dimensions. (d) Tanh activation function which produces a smooth approximation.

5.4 Industry data set

The sensitivity studies allowed to define the training hyper-parameters and to identify two networks of interest: one that maximizes library size reduction ($N = 8$) and one that converges fast ($N = 13/13$) both reaching the target accuracy. In actual core calculations, the amount of lattice data may be closer to the industry set than to the denser training set considered so far, which could potentially hamper the supervised learning process. It is therefore of interest to assess the network's approximation quality using this restricted data set. With regard to its structure, for this kind of engineering applications the

regularity of a Cartesian rule is considered convenient [92]. This shows that machine learning implementations are fairly empirical approaches. The test error is computed on the $\mathcal{T}_{Xe}$ as to compare with the methodologies of the preceding chapters. The E_σ error evolution with training is showed in Fig. 5.7a and 5.7b for the networks $N = 8$ and $N = 13/13$. Important cross sections are shown.

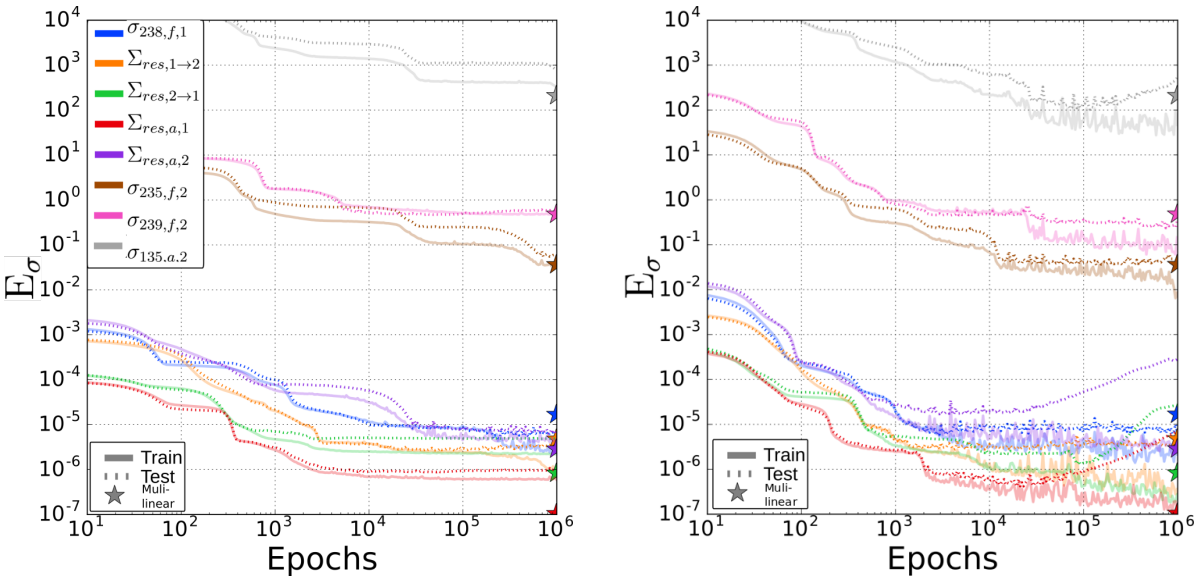
Multi-linear errors, marked with a star, are reached trough training though after a relatively high amount of epochs. The network $N = 13/13$ starts to overfit this smaller data set after 1E5 epochs and test errors stagnate or even increases for some cross sections. Training errors on the contrary keep diminishing even reaching ML for all cross section as the set is being memorized by the network. In [107] the recommended minimum ratio of training data to wheights is 3 to 1 which is attained for both networks in either the 80%/20% split (162 points per weight) or the industry set (5 points per weight). For typical lattice calculation data, a relatively small 2-layered network can over-fit and an "early stopping" criterion is needed.

In Table 5.6 the AV_σ and the library size reduction, defined as the ratio of parameters in the models

$$100(1 - |\mathrm{ANN}|/|\mathcal{X}_\star|), \tag{5.5}$$

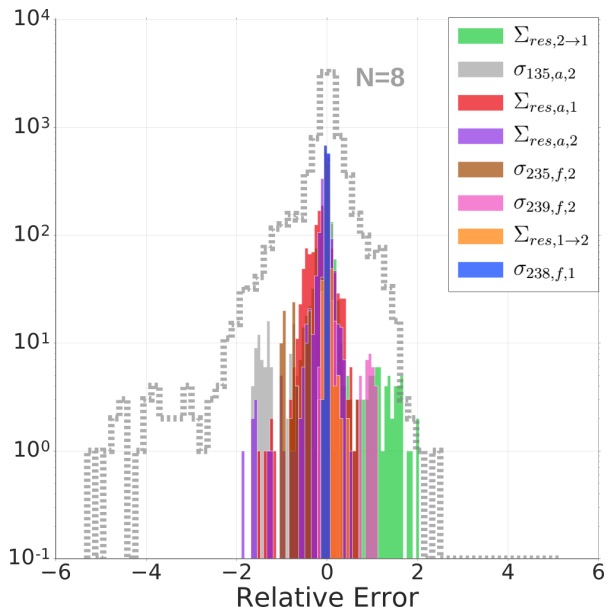
is presented also for the split 80%/20%. Errors are reported at 1E6 epochs for $N = 8$ and 2E4 epochs for $N = 13/13$ (early stop criterion[7]). Network's errors are lower than 1E-1% and a significant library size reduction of up to 96% is obtained. Actually, this upper limit is reached with only 6E3 epochs for $N = 8$ and 7E2 epochs for $N = 13/13$. An error histogram discriminated by cross section is presented in Fig. 5.8a at 7E2 epochs for $N = 13/13$. It shows centered means, and low overall errors without tails for this netwoks that converges faster than $N = 8$ whose profile it's also presented. The stochastic nature of the learning process requires a large amount of epochs to reach the target accuracy.

[7]Defined as a break condition in the training process when the test error starts to increment while the training error keeps diminishing.



(a) (b)

Figure 5.7: Absolute error E_σ for important cross section during training using the industry set and the $\mathcal{T}_{Xe}$ for test. In general training errors tend to attain multi-linear errors which are marked with a star. (a) A single layer network of $N = 8$ which maximizes the library size reduction. The upper limit of 1E-1% is reached with only 6E3 epochs. (b) A two-layered network of $N = 13/13$ where over-fitting starts to occur and the test error departs from training error. The optimal model is obtained at 2E4 epochs. The upper limit of 1E-1% is reached with only 7E2 epochs.



(a)

Figure 5.8: Error histogram using the industry set for $N = 13/13$ at 7E2 epochs discriminated by cross section characterized by centered means, low standard deviation without tails. In gray the corresponding error histogram for the network $N = 8$.

Approximation	AV_σ	80/20	industry
Multi-linear	Test	1.2E-3	2.1E-2
	Size	36096	1050
ANN ($N = 8$, $O = 1$)	Train	4.7E-2	3.9E-2
	Test	4.4E-2	7.3E-2
	$\|ANN\|$	41	
Library reduction		99%	96%
ANN ($N = 13/13$, $O = 1$)	Train	9.0E-3	2.1E-2
	Test	8.5E-3	6.0E-2
	$\|ANN\|$	222	
Library reduction		99%	79%
ANN ($N = 20/20$, $O = 144$)	Train	2.5E-2	1.9E-2
	Test	2.5E-2	3.0E-2
	$\|ANN\|$	3.1E3	
Library reduction		99%	98%

Table 5.6: Average relative errors AV_σ using the 80/20 split and the industry set for multi-linear interpolation and artificial neural networks. The errors are tested on $\mathcal{X}_{Xe}$ and reported at 1E6 epochs for $N = 8$ and 2E4 epochs for $N = 13/13$ (early stop criterion). The error for the artificial neural network are in the region of interest, achieving significant library size reductions though ultimately unable to reach the corresponding multi-linear accuracy. Nonetheless both networks are able to profit from the richer data set achieving a higher accuracy with the 80%/20% that with the industry set. The multi-output network having 3.1E3 coefficients (equivalent to 3.1E3/144=21 coefficients per cross section) achieves the highest compression.

5.5 Multi-output network

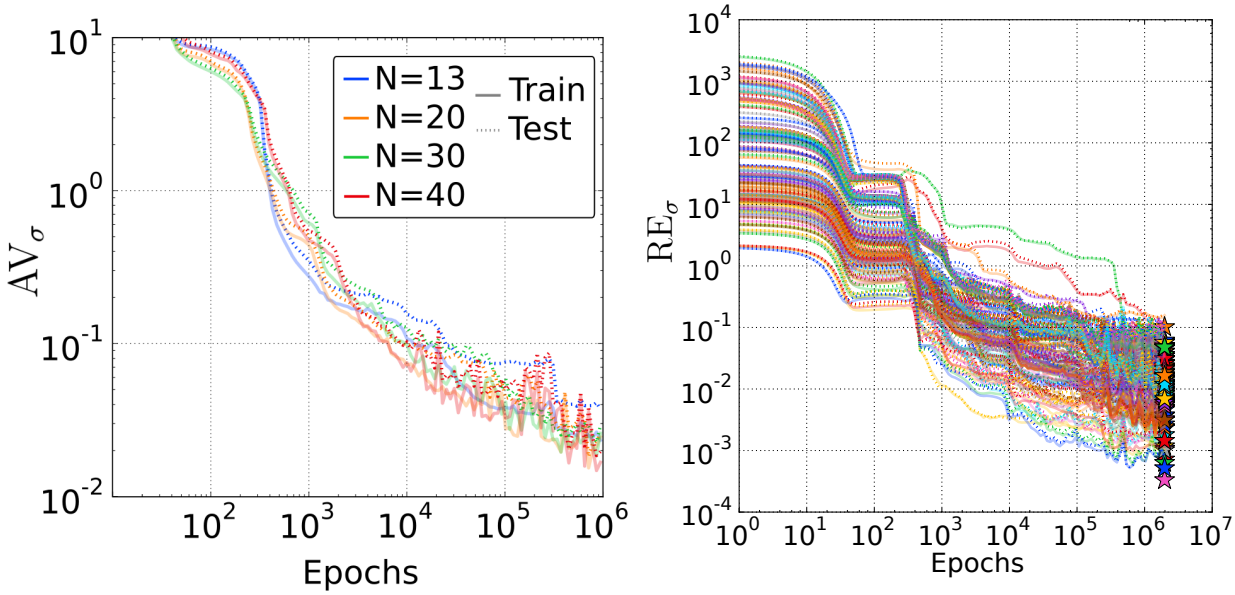
Artificial neural networks can be easily extended from single output ($O = 1$, as treated so far) to "multi-output" systems ($O > 1$) without significant change in the training process, storage requirements, or parallelizable capabilities. In this sense they truly are mapping devices between multi-dimensional spaces which is the ultimate cross section modeling problem. An implicit idea, in a common central model, is its capability to share learned structures from the data across the output vector. This means that a capture pattern is stored only once, and used as required for the different cross section. Such behaviour is explicitly observed in (very) deep neural networks in Computer vision problems. Though the analysis and identification of emergent features in the network [108] is beyond the scope of this thesis, it's worth recalling this concept, as any library compression does relays in exploiting common elements across the cross section set.

5.5.1 Parametric analysis in the number of neurons for a two-layered network

In view of previous results, a parametric study on the topology of a two-layer network is presented in Fig. 5.9a where *a single network models the totality of the cross section set* i.e., a function $f : x \in \mathcal{X}^{d=3} \rightarrow \mathbb{R}^{|\mathcal{Y}|=144}$. The AV_σ error on the train and test set is shown during training. All the networks reach the targets error of 2E-2 (similar to multi-linear). A slightly better convergence is observed for a number of neurons $N \geq$20. No further improvement is observed with additional neurons, hence a multi-output network of $N = 20/20$ is considered optimal. From Eq. 5.4 the network's parameters are $|ANN|$ =3.1E3 which stands as a maximal library reduction of 98% as presented in Table 5.6. In view of these results no further exploration in subdividing the cross section set was considered. A multi-output network of $N = 20/20$, has 3.1E3 parameters which is equivalent to 22 parameters per cross section. A multi-output network of $N = 13/13$ has 2.1E3 coefficients which is equivalent to 14 parameters per cross section. This is equivalent to a single-output shallow network of 2 neurons, quite unable to fit the data, since 8 were needed to meet this target (having 41 parameters). This evidences that learned patterns in the unique model are shared among the cross section. Provided an adequate network's structure and training, as long as each cross section's error, and in consequence the average model's error, meets the targets the *storage size is independent of both input and output space cardinality.* This is the ultimate feature of a library reduction modeling technique.

5.5.2 Error analysis for a two-layer, 20 neurons multi-output artificial neural network

In multi-output networks, an implicit metric is being defined among the cross section since the layers are fully connected. Moreover, a secondary source of regularization is added, mainly the minimization of a much larger error set, spammed throughout the entire domain, which could be detrimental to the capture of the details in the cross section's profiles for regions of rapid variation. In Fig. 5.10 and Fig. 5.11a the RE_σ with



(a) (b)

Figure 5.9: (a)Topological study for a multi-output two-layered network. The error AV$_\sigma$ is presented in function of the training where all the networks reach the target errors. The optimal amount of neurons is 20. (b) Every cross section error in function of training for a multi-output 2 layered network of 20 neuron. In general the multi-linear error marked with a star is reached and an overlap is observed between the error on the train and test sets. Up to 1E3 epochs, the error evolution is rather similar for all the cross section of the set and differential error progression is noticed afterwards.

training and cross section plots are presented for $\sigma_{235,f,2}$ and $\Sigma_{res,a,1}$ respectively. Though resulting from a single multi-output network model, they are rather indistinguishable from those obtained using a dedicated single model per cross section. No perturbation or abnormality is observed. By close inspection in the zoom region of Fig. 5.10b a somewhat paradoxical observation can be made: this multi-output network using the industry set is able to follow better the cross section profile than a single shallow network using the much larger 80%/20% split (see Fig. 5.6b). From this observation a very simple rule of thumb could be extracted for cross section modeling: without some separation between test and training error (i.e. the onset of over-fitting) small features on the data profile could be being regularized out of the approximation, a consequence of the stochastic learning process with very large data set.

In Fig. 5.12a its shown that the approximation is quite able to follow a difficult set of data for the important cross section $\sigma_{239,f,2}$. This is further exhibited in Fig. 5.12c where, again, a smooth approximation is obtained. In Fig. 5.12b however, some oscillations are noticed for the $\sigma_{240,f,1}$ cross section.

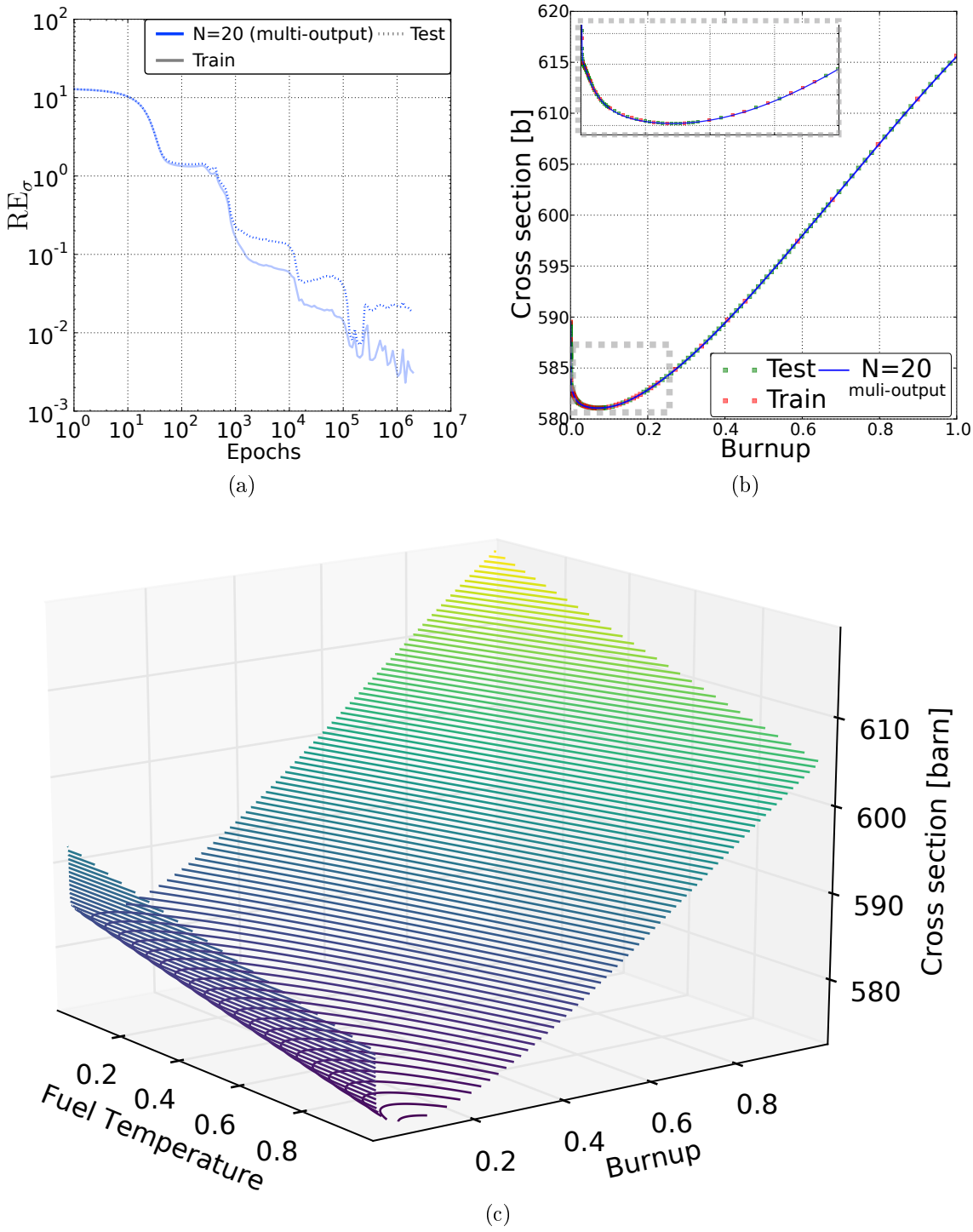


Figure 5.10: Modeling of $\sigma_{235,f,2}$ with a multi-output $N = 20/20$ network using the industry set. (a) The onset of over-fitting is observed at 1E5 epochs. The network reached the error targets. (b) Cross section at $\sigma_{235,f,2}(Bu, T_f, C_{B0})$. The smaller industry data set facilitates following the profile change even if the entire cross section set is being modeled. (c) A smooth approximation without abnormalities is observed. In practice indistinguishable from a single dedicated model.

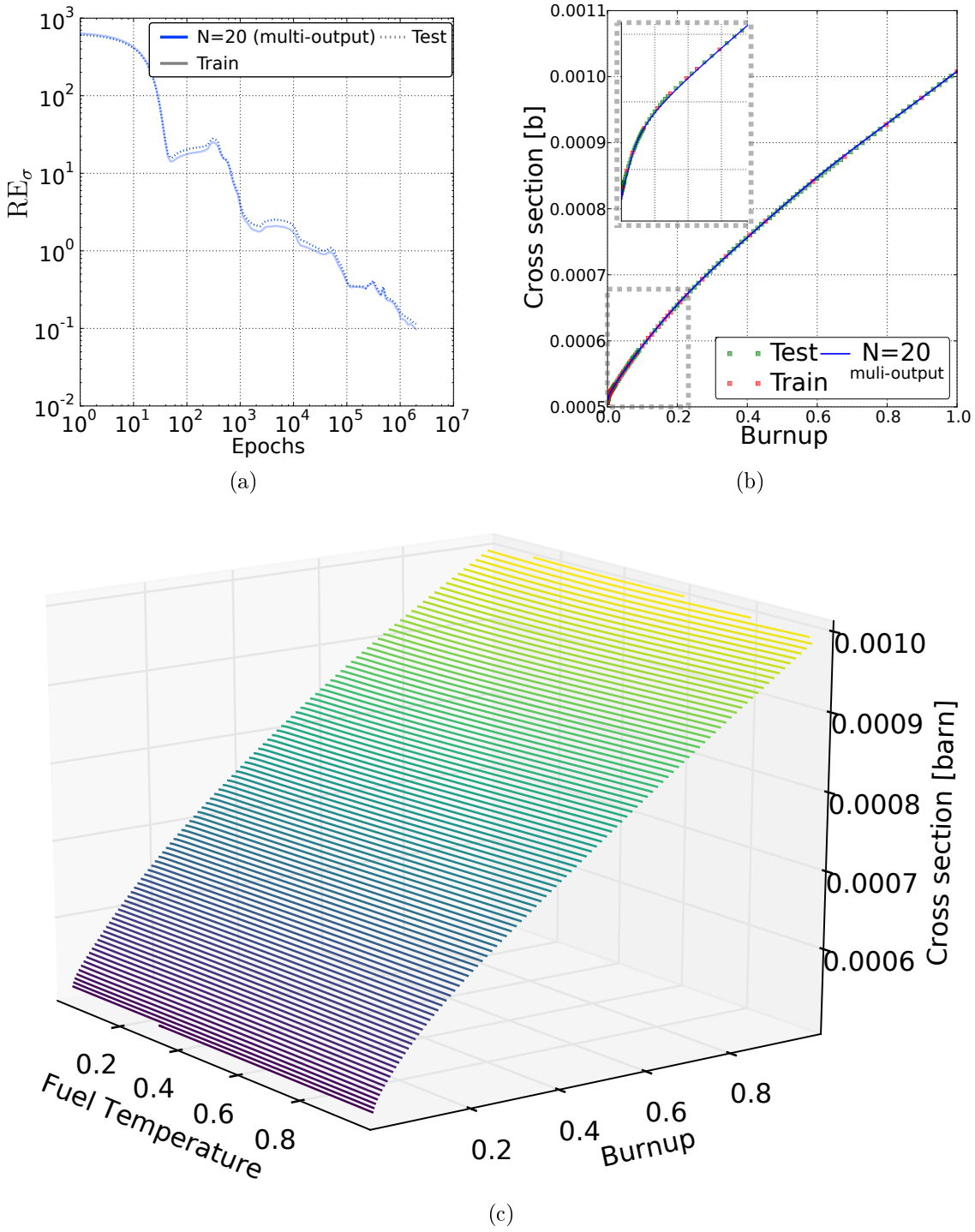
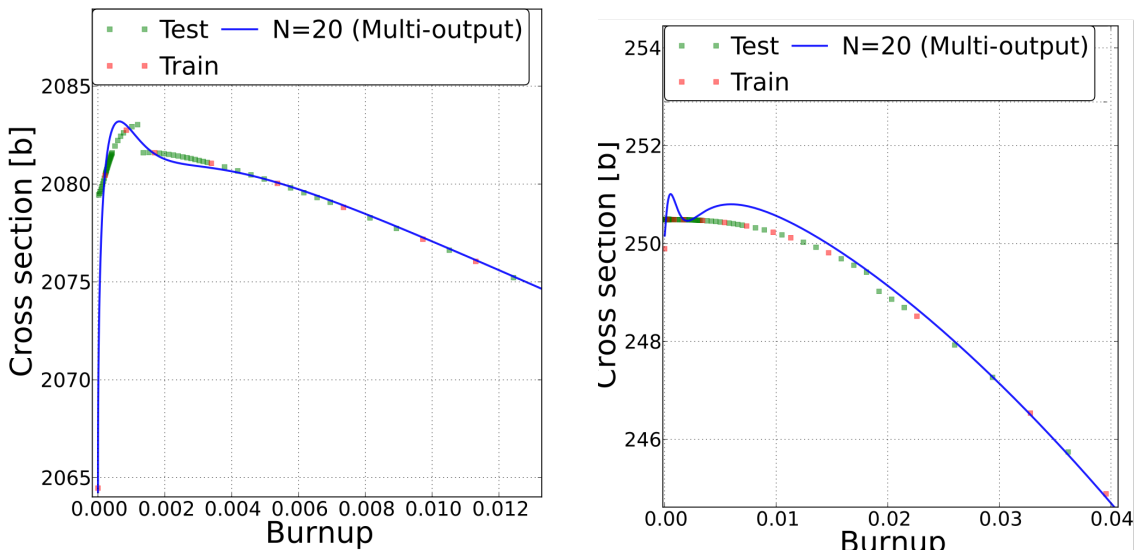
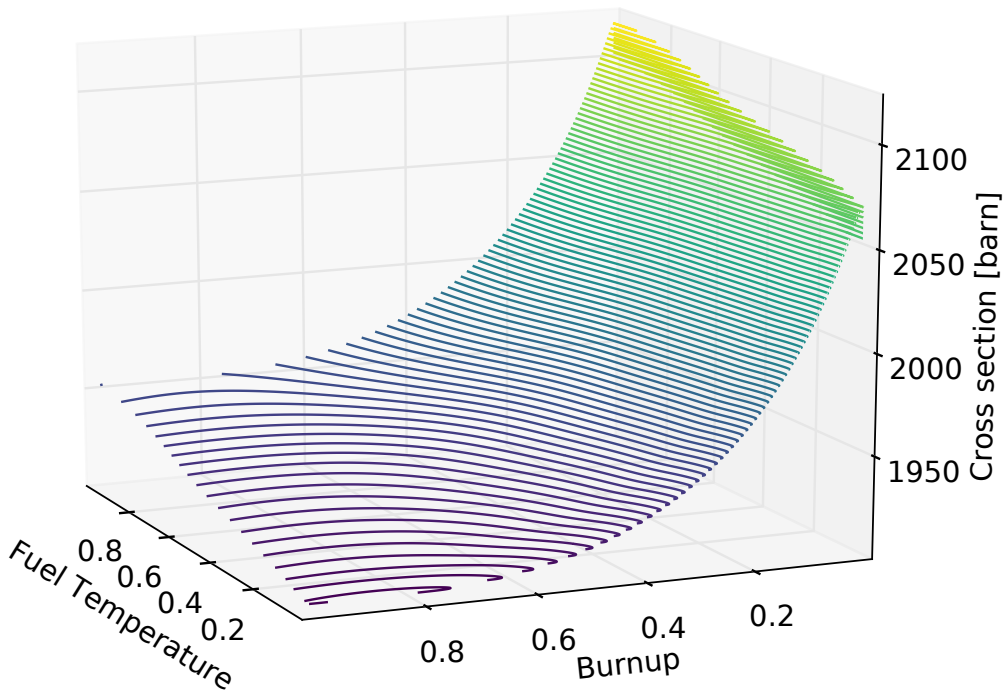


Figure 5.11: Modeling of $\Sigma_{res,a,1}$ with a multi-output $N = 20/20$ network using the industry set. (a) No over-fitting is observed. (b) Cross section at $\sigma_{235,f,2}(Bu, T_f, C_{B0})$. The smaller industry data set facilitates following the profile change even if the entire cross section set is being modeled. (c) A smooth approximation without abnormalities is observed. In practice indistinguishable from a single dedicated model.



(a) (b)



(c)

Figure 5.12: Modeling of $\sigma_{239,f,2}$ and $\sigma_{240,f,1}$ with a multi-output $N = 20/20$ network using the industry set. (a) Cross section at $\sigma_{239,f,2}(Bu, T_f, C_{B0})$ is able to almost interpolate training data. (b) Cross section at $\sigma_{240,f,1}(Bu, T_f, C_{B0})$ presents an overshot. (c) A smooth approximation without abnormalities is observed.

5.5.3 Weight evolution with training

The parameters of the network determine its behaviour [109], capability of exploiting common patters in the data, and possibly the error bounds [110]. Though a detailed analysis of these properties, which belong to an ongoing area of research, are beyond the scope of this thesis, some observations can be made in view of their evolution with training. The network's weights (that connect the neurons) and baiases evolution during training are presented in Fig. 5.13 discriminated by layer. The networks used in this thesis are shown (using tanh activation function). Small initialization values give way, with training, to a single large weight and bias in a shallow network ($O = 1$) as shown in Fig. 5.13a. A similar plot is obtained for the network of 8 neurons. This suggest a similar "network utilization" which is coherent with the approximation's plot shown previously[8]. This few but large parameters, remained bounded in these training settings, and are located in the hidden layers.

It could be argue that the multi-output network exhibits a higher degree of "parameters utilization". Incidently they depart from the bulk of zero centred values at about 1E5 epochs, where differential error evolution was clearly observed in the cross section (see Fig. 5.9b). Inner layers could define a logical structure within the network while only the last layer, modulated by the activation function, accounts for the output values of the network that conform the approximation. The last layer's parameters (Out) are usually bounded by 1, as the normalized cross section values.

5.6 Performance remarks

Training and evaluation times are presented in Table 5.7 with a single network adaptation per epoch (only one batch, equal to the entier training set). The training time ranges from 1E-3 s/epoch to 1E-1 s/epoch depending on the availability of GPU acceleration. In the parametric studies, up to 1E6 epochs with 5 batches per epoch were used, which is a non-negligible calculation time of about 5×(1E-3 s)×1E6=4 h/σ similar to reported elswere for this kind of networks and accuracy [58]. On the other hand, 7E2 epochs requiring only a few seconds already produced satisfactory results.

An evaluation vector composed of random numbers in the hyper-cube of size 2.4E5 was used to estimate the evaluation time expressed per evaluation point and cross section[9]. Unlike the Kernel methods or B-spline implementation, the PyTorch module is optimized for vectorial calculations. Evaluation speed was quite compatible with industry requirements especially if GPU acceleration is available.

[8] Indeed several experiments with up to 1E7 epochs with single layer networks of 100, 200, and 500 neurons systematically resulted in the same parameter's profile: a few parameters different from zero and similar (training) error than, for example, a two layer network of $N = 13/13$. No training settings/network topology was able to have interpolating like training errors. In [111] it is argue that an excessively complex parameter space can be counter-productive due to the increase probability of the training algorithm getting trapped in local minima.

[9] Representing an entire core calculation with the amount of cells equal to 300×50×16.

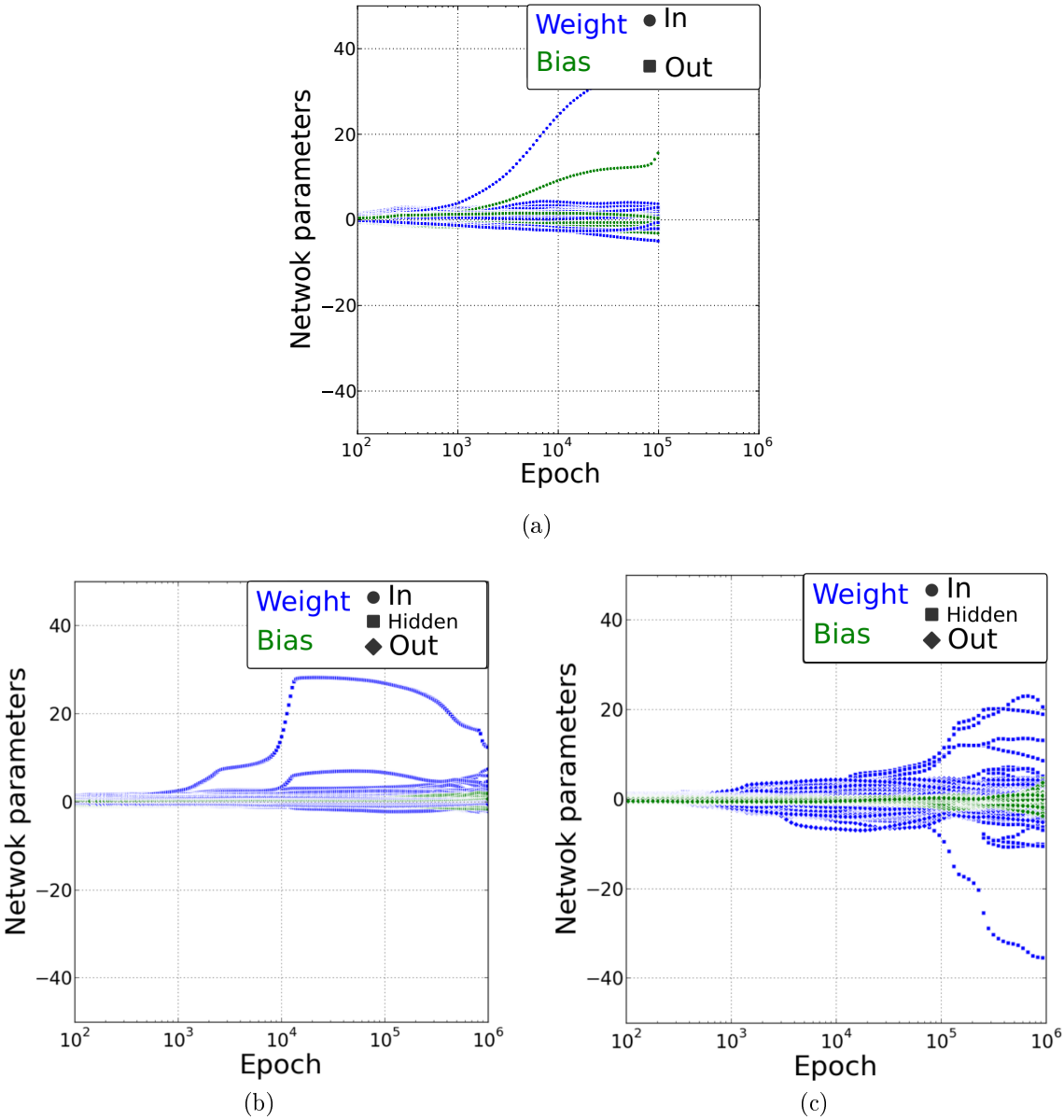


Figure 5.13: Network parameter evolution with training. Tanh activation function is used. (a) 8 neurons shallow network. (b) Two layered 13/13 network. (c) Two layered $n = 20/20$ multi-output network.

The training and evaluation times of a network is practically independent of the dimensionality of the output. Hence, multi-output times are very similar to those of a single output, especially if GPU acceleration is available, though for treating the entire cross section set. Thus a reduction is obtained in a factor proportional to the amount of cross section modeled by the network. Possible complications with bigger multi-output networks could arise from GPU memory limitations. To address this the training size batches (dependent on the input space dimensionality) must be changed depending on the quantity of cross sections being modeled by the network. Both aspects can influence the optimal networks topology. Precisely this kind of inter-dependencies are a major shortcoming of artificial neural networks, since they are generally resolved in a quite heuristic way.

N	GPU			CPU		
	Training [s/epoch]		Evaluation	Training [s/epoch]		Evaluation
	80/20	Industry	[s/point]	80/20	Industry	[s/point]
8	2E-3	2E-3	1E-9	7E-3	1E-3	2E-8
50	2E-3	2E-3	2E-9	2E-2	2E-3	3E-7
200	3E-3	2E-3	4E-9	1E-1	5E-3	1E-6
13/13	3E-3	3E-3	8E-10	2E-2	3E-3	1E-7

Table 5.7: Training and evaluation times for artificial neural networks. Training times using with the 80%/20% split (36096) points and the industry set (1050) pints is presented per epoch with a single batch. They exhibit a relatively low dependence on the number of neurons for both CPU and GPU calculations. Vectorialized evaluation times per point and cross section are quite compatible with industry's speed requirements. Multi-output network training and evaluation times are practically the same since they are independent of the dimensionality of the output.

5.7 Final thoughts and discussion

In this chapter artificial neural network were used to model cross section, discussing their main characteristics and shortcomings:

Artificial neural networks

- Multi-variate function approximation is often a generalization of uni-variate representation methodologies, as for example the tensorization used in the preceding chapters. Artificial neural networks are a fundamentally different technique, described as "universal approximators" or "model-free mapping systems" that can inherently cope with a high dimensionality of both input and output spaces without user imposed rules among the variables nor added computational cost.

- They are an intrinsically incremental learning method quite able to cope with an on-line data steam, not limited to an "off-line" preparation phase as common in a one-through calculation scheme.

- The training process is an error driven supervised learning task for models that are able to generalize quite well provided a dense enough data set. Indeed since cross sections tend to be smooth, noiseless functions, with the 80%/20% split a complete overlapping between test and training errors was observed for all the networks. This allows for an error-controlled library preparation as in other works [31].

Sensitivity analysis

- Data normalization had a strong effect in the training process. Without it, the optimizer was unable to converge with resulting approximations not fitting the data and errors of 100%. Normalizing the burnup variable with the square root, and the cross section value by the variance with a zero mean, exhibited the best results.

- The learning rate (α) modulates the actualization of the weights influencing the convergence rate and the optimizer's possibility of escaping local minima. The value $\alpha = 0.001$ was found optimal. It's worth mentioning that big learning rate values ($\alpha = 0.1$) did not result in divergent errors.
- With the exception of HardS all activation function exhibit a similar performance for these small networks and Tanh was selected following literature recommendations. Qualitative analysis on cross section approximation showed that the Relu activation function produces nonphysical breaks in the derivative, not present with smooth activation functions.
- For the use case considered, a single-layer network of only $N = 8$ having 41 parameters is the smallest model that attains the target accuracy.
- Under the condition of constant library size, a two layered network provided the faster convergence in accordance with [100, 95, 58]. Other topologies commonly found in literature, provided no further advantage in these settings.
- The parametric studies resulted in optimizations that did not necessarily comply with rules of thumb found in literature.
- The stochastic nature of the training process has several consequences. Variability in the final errors of different training session of ±2E-2 was observed. Denser data sets provide a stronger regularization effect, increased by the use of randomized batches. When trying to minimize the amount of parameters, over-fitting is not a major concern and weight decay or regularization layer were not analyzed. Due to cross section's regularity, training error generalizes quite well.

Industry set

Two networks of interest were identify, of maximal library size reduction ($N = 8$) and of fast error convergence ($N = 13/13$) that were tested using the industry set.

- Using the industry set for training, multi-linear errors are generally attained by the training error though after a relatively high amount of epochs. A significant library size reduction of up to 96% is obtained.
- Over-fitting was observed in the $N = 13/13$ network using this smaller data set after 1E5 epochs (test errors increasing with training). The amount of parameters of such relativity small network are sufficient to over-fit data sets of this size as could be found in industry. Training errors on the other hand reach those of multi-linear for all cross section.
- Similar results are obtained for the 80%/20% split, actually reaching an AV_σ under 1E-2 though never attaining the multi-linear error.
- It could be argue that artificial neural networks trade storage memory for computation time, which is a desirable feature in an off-line preparation phase. Indeed the $N = 13/13$ network was able to profit from the 80%/20% split achieving a lower error than with the industry set.

- An empirical observation for cross section modeling under these settings is that in order to reproduce the precise shape of a given cross section variation some separation between training and test error has to occur (i.e. the onset of overfitting).

Multi-output

Unlike the other methodologies, artificial neural network can be naturally and easily extended to multi-output models that approximate the entire cross section set.

- A single network was able to model the entire cross section set reaching an AV_σ ~3E-2, similar to multi-linear for the industry set. A maximal library size reduction of 98% was obtained having a total of 3.1E3 parameters.
- This would be equivalent to 20 per cross section, that for single-output shallow networks, represents about 3 neurons which are insufficient to produce satisfactory approximations. This shows that learnt common patterns are being shared among the cross sections, enhancing the library reduction capabilities of the method.
- A parametric study on a two-layer multi-output network resulted in 20 neurons as the optimal value. In view of the satisfactory results no further segregation of cross sections based on difficulty, reaction type, group, etc was considered.
- Qualitative analysis of important cross sections prove indistinguishable from dedicated (single output) models. No perturbation or abnormal behaviour was observed. Additionally, the training points of the (smaller) industry set were approximated better (than a single-output network using the 80%/20% split), even if optimizing the entire cross section set.
- Multi-output network's parameters exhibited a larger rate of change with training, especially after 1E5 epochs where differential error progression was observed in the cross section set. This suggest that the degrees of freedom are better utilized than with a single-output network.

Performance

Modern frameworks for artificial neural networks can profit from GPU acceleration:

- The evaluation speed was quite compatible with industry standards, especially if GPU acceleration is available. CPU evaluation times were inversely proportional to the amount of neurons in a measurable amount.
- A similar dependency was observed for the training. Some of the studies in this work considered up to 1E6 epochs using 5 batches per epoch, with thus 4 h worth of training per cross section, which is considerable. Very high accuracies prove challenging for artifical neural networks, requiring long training times as reported elsewhere [98].

- However, acceptable errors were obtained with only 6E3 epochs for $N = 8$ and 7E2 epochs for $N = 13/13$ which corresponds to a training time of a few seconds. For the networks of interest, the computational cost was independent of the network's topology if GPU acceleration is available.
- The training and evaluation time of a multi-output network was independent of the output's cardinality, being thus effectively reduced in a factor proportional that amount. The totality of the training data was able to fit into GPU memory.

Provided an adequate topology and training parameters, multi-output network have indeed behaved as "model-free" mapping devices. To the author knowledge, no other, more extensive studies of cross section modeling using artificial neural networks exists to date.

Chapter 6

Conclusion and Perspectives

This thesis was devoted to few-group cross section approximations methodologies in view of modern core calculations requirements, characterized by an increase in the amount of cross sections to model (output space) and the dimensions of their domain (input space). In general, these are smooth, low order, and virtually noise-free functions of varying complexity across the state-variables and the cross section set. The main modeling objectives were to reduce the storage requirements of the library, in order to increase the computational performance of core calculations, at a target accuracy ($\mathrm{AV}_\sigma \sim$ 1E-2%) and evaluation speed (>1E-5 s)[1]. Secondary objectives were to lower the amount of lattice calculation required, library preparation time, and model's complexity with a good conditioning if an expansion basis is used.

The current industry standard is multi-linear or second order interpolation using Cartesian grids. Though straightforward, simple, and efficient under the right circumstances, it does suffers from the "Curse of dimensionality" where storage requirements may grow exponentially rendering the library inefficient and cumbersome to work with. Classical methodologies can enforce large amounts of unnecessary data into the model as evidenced by the success of lighter sampling rules (sparse grids), a posteriori rejection, regression, or even Taylor expansion methods. The multi-variate modeling aspect of such method is usually restrained to an extension of uni-variate approximations by a tensor product rule, for each cross section, independently.

6.1 Main conclusion

Table 6.1 summarizes the main results of the thesis showing the reduction in the library size[2], the accuracy, the evaluation speed and the underling library structure for the three studied methodologies: **Spline interpolation with a B-spline basis, kernel methods**

[1] As an industrial reference for comparison the commercial code APOLLO3 [11] provides an evaluation speed of 1E-7 s per evaluation point per cross section.

[2] The library reduction is measured at constant $\mathrm{AV}_{\sigma,\star}$ error in percentage as $100(1-|Coef|/|\mathcal{Y}|\times|\mathcal{X}_\star|)$ which is: the total amount of parameter for the entire cross section set ($|Coef|$) versus the total amount of terms in the multi-linear interpolation using the industry set ($|\mathcal{Y}\times\mathcal{X}_\star|$)

(RKHS) with pool active learning, and artificial neural networks.

Spline approximation using B-splines with a Cartesian grid is an industrial state-of-the-art methodology that sets the ground of comparison for other, more innovative methods. It uses a piece-wise smooth approximation space exploiting the computationally efficient, and freely available B-splines routines that produce well conditioned bases, resilient to numerical noise. The evaluation time, proportional to the regularity of the approximation (m), meets the modeling objectives of this proven, yet classical methodology. The overall improvement with respect to multi-linear is about 26%, rather small due to the underling library structure and the lack of regularization.

Kernel methods (RKHS) allow to reproduce several function spaces, including high order splines, but introducing regularization and unstructured supports, which are selected by active learning. The kernel trick avoids explicit computations in this feature space, facilitating the obtention of the approximations, and making the methodology moderately resilient to an input dimensionality increase. Supervise learning is able to reflect the function's complexity across the state-parameters ($|\mathcal{X}_\dagger| \sim \sigma^{(2)}$) and even the cross section set (by dropping the share support) achieving a library reduction of up to 85%. A very high accuracy was obtained with a second order regularized spline space which also uses an optimized support. The evaluation speed is similar to splines and systematically above the modeling objectives if the evaluation vector is shared among the cross sections. High conditioning number of the kernel matrix and large coefficient's module in the approximations were observed. Though with just a few cross section the active learning procedure was able to select a near optimal support, long active learning times of up to a day were encountered for the most demanding loss functions.

Artificial neural networks are fundamentally different "model-free" mapping devices able to cope, in principle, with arbitrary dimensionality increases of both input and output spaces. These intrinsically parallelizable models, able to profit from GPU acceleration, drop entirely the notion of support, being able to model a single cross section or the complete set. A maximal library size reduction, evaluation speed, and good accuracy, are obtained for a model independent of the cardinality of both input and output space, provided a suitable network's topology under an adequate and sufficient training. Indeed long preparation times may be encountered (especially for single-output networks for the entire set), and a large quantity of problem dependent, interrelated, hyper-parameter must be defined. These are not hardware independent, and the resulting models can be hard to interpret.

These methods used a pool setup, where all the data is already present and without iterations with the lattice code, which is a natural consequence of the underlying "one-through" two-step calculation scheme. They could be extended to an "on-line" mode, especially artificial neural networks. In regards to the approximation robustness expressed as the module of the coefficients: B-splines systematically offered a very well conditioned matrix and bounded coefficients value, though for regularity $m > 1$ over-shots or oscillation were found for some support configurations. Kernel method conditioning number and coefficients module was systematically large. Nonetheless the spline space reproduced, if linear, was naturally bounded by the support and for $m > 1$ regularization was successfully applied. Though effort has been made to achieve acceptable computational performance (by developing relativity efficient compiled libraries) further optimization

Methodology			Size at $AV_{\sigma,\star}$ (Δ Coef. %)	Accuracy at $\lvert\mathcal{X}_\star\rvert$ (%)	Evaluation speed	
Name	Detail	Library structure			[s/σ]	$\propto$
B-splines	Multi-linear	$\lvert\mathcal{Y}\rvert \times \lvert\mathcal{X}_\mathcal{S}\rvert$ ($\lvert\mathcal{X}_\mathcal{S}\rvert \sim N^d$)	1.5E5(0.5 GB)	2.1E-2	1E-5	m
	Splines		1.1E5 (27)	1.5E-2		
Kernel Methods	Shared support	$\lvert\mathcal{Y}\rvert \times \lvert\mathcal{X}_\dagger\rvert$ ($\lvert\mathcal{X}_\dagger\rvert \sim \max(\sigma^{(2)})$)	2.1E4 (86)	3E-3 ($m=2$)	[1E-5,1E-3]	$\lvert\mathcal{X}_\dagger\rvert$
	~~Shared support~~	$\sum_i^{\lvert\mathcal{Y}\rvert} \lvert\mathcal{X}_{\dagger,i}\rvert$ ($\lvert\mathcal{X}_{\dagger,i}\rvert \sim \sigma_i^{(2)}$)	1E4(93[a])	9E-3 ($m=1$)		
Networks	per σ	$\lvert\mathcal{Y}\rvert \times \lvert ANN\rvert(\lvert ANN\rvert \sim \mathcal{N}^2)$	5.9E3 (96)	8.5E-3[b]	[1E-7,1E-9]	$\mathcal{N}$
	Multi-output	$\lvert ANN\rvert \sim \mathcal{N}^2$	3.1E3 (98%)	2.E-2	1E-9	

Table 6.1: Library size reduction, accuracy and evaluation speed of the three studied methodologies of the thesis: Splines using B-splines, kernel methods (RKHS) with active learning, and artificial neural networks. Maximal reduction of the library size and highest evaluation speed is obtained with artificial neural networks, especially with "multi-output" networks. Very good performances are obtained with kernel methods (RIXS loss function) without the shared support condition. The evaluation speed of B-splines, kernel methods and networks is proportional to the order m, the size of the support $\lvert\mathcal{X}_\mathcal{S}\rvert$ and the amount of neurons $\mathcal{N}$ respectively. When the evaluation vector is shared among cross sections kernel method evaluation speed meets the target for the entire range support sizes and kernel types, being faster than B-splines. In regards to the library structure, splines using B-splines use a Cartesian grid for each cross section. Kernel method utilize an optimal support proportional to the cross section's complexity, which can be independently optimized without a shared support condition. Artificial neural networks drop the notion of support altogether, and when considering a multi-output network the entire cross section set is treated with a model depending solely on the amount of neurons.

[a] Value for NoS strategy. Without share support condition, if the error tolerances are modulated by the cross section importance (ImpNoS), 3E3 (98%) coefficients are required at $AV_{\Sigma\star}$ but with very high AV_σ error.

[b] AV_σ error in the $\mathcal{T}_{Xe}$ (test error) for the two layered network $\mathcal{N} = 13/13$ and the 80%/20% split, for which a lower error is achieved than when training with the industry set.

work would probably result in improved evaluation speed.

In view of these results *kernel methods utilizing a regularized piece-wise function space with an optimized support* provide an efficient and relatively straightforward method to improve cross section modeling in industry applications[3] [112]. *Artificial neural networks offer the highest performance potential* [113], especially in view of multi-output networks, capable of exploiting common patterns in the data achieving maximal compression and evaluation speed, but lacking standardization of the modeling phase, or some sort of guidelines, probably required in industrial applications. *High order approximation improve the accuracy,* but its benefits in terms of library reduction are limited, especially if using the intrinsically penalizing Cartesian grids, or possibly even regular grid. To the author's knowledge, no other, more extensive studies of cross section modeling using kernel methods or artificial neural networks exist to date. In the following sections, additional **contribution of the thesis** and observation are presented, followed by future perspectives.

6.2 Additional contributions of the thesis

6.2.1 Spline interpolation

Piece-wise interpolation in a Cartesian grid stands as a classical industry technique for cross section modeling which was explored in the third chapter.

- Polynomial approximations dispose of an extensive and accessible literature armed with a rich set of computational tools and libraries. Unfortunately for cross section approximation, global polynomials interpolation proved inadequate, at least without special discretizations in the sampling of the support. Divergent errors where obtained with just a few points.
- Multi-linear interpolation presented a monotonically diminishing error with the support size for σ, Σ, and k_∞. In virtue of the error bound with the support, important properties such as cross sections positivity and other relations are preserved.
- **A formal development of spline interpolation using B-splines, not treated before in the cross section literature,** was presented introducing their main theorems. B-splines are indeed a convenient base, with a positive banded collocation matrix, requiring a small quantity of evaluation terms proportional to the regularity of the approximation.
- Overshots and oscillations were detected in some cross sections, leading to non-trivial dependence of the error with the support. These issues have been mentioned before [46], but not analyzed in the literature. Though this is not an exhaustive study on knot vector construction, **equipped with the complete B-spline formalism**

[3] The intrinsic modular character of the method is well suited for actual industry implementations: the function space selection, the resolution routines of the resulting linear systems and the active learning phase, further determined by the lose function, are basically decoupled phases.

several recepies, as suggested in [68], were tested and attempt made to improve the knot vector. Unfortunately, at least for this use case, knot vector manipulations had only a minor impact on the spline representation quality.

- Higher order splines outperformed multi-linear interpolation to a small extent, in accordance with other works [30]. **This is mainly attributed to the lack of regularization in the interpolating solutions, and the underling tensor product rule**.

6.2.2 Kernel methods

In this thesis several kernels (RKHS) were explored for the cross section modeling of a standard PWR fuel assembly, performing the corresponding parametric studies, and assessing different active learning strategies for optimizing the support. Results were presented in chapter four comparing to multi-linear interpolation.

- **A formal presentation of Kernel method with active learning was provided for cross section modeling** noticing the synergic combination: active learning has a complete freedom to select the observations, whose quantity minimization is crucial for performant RKHS.
- **These innovative, "observation based" kernels functions, impose no conditions on the data pool thus overcoming the limitations of classical expansion methods** which are circumscribed to an hyper-dimensional rectangle in the input space [27].
- Several kernels were explored with their corresponding hyper-parameters resulting in the selection of **the regularized spline function space. This matches the insight obtained from the thorough literature review: efficient cross sections representation may benefit more from flexibility in the selection of the support points than the use of exotic function spaces.** Muli-linear interpolation was reproduced, retaining its error bounds, but for an arbitrary grid. This facilitated an analysis of active learning strategies **highlighting the potential gains for industry settings.**
- A random selection of support points, that reduces the collinearity in the data, **substantially improved the models even with a linear function space, showcasing the interest of unstructured supports.**
- **In this thesis loss functions acting on σ, Σ, and k_∞ were confronted** in terms of accuracy gains, library size reduction, and computational cost. **Comparative studies with such a varied set of strategies demonstrated the robustness of the methodology.**
- The best compromise between library reduction and the computational cost of active learning was achieved when using the importance to pre-select a subset of the cross sections (RXSI). **Loss function profiles allowed to quantify the amount**

of useful information in the data set and its extraction rate providing a direct measure of the error. **The need of relative errors to make meaningful comparisons among cross sections** was discussed in detail, indicating as well possible difficulties in the fast group.

- **In this thesis the effect of the shared support condition was examined by the clustering of cross section participating in the active learning process and plateaus in the partial loss function values**. A shared support penalizes the active learning by the most complex cross section, forcing the incorporation of large volumes of unnecessary data.

- For a given pool of data, **the modeling objectives in library preparation can be imposed through the loss function in the active learning phase**. Without a shared support, when modulating the error tolerance with the importance (ImpNoS), a good accuracy in Σ was obtained with a minimal library size though high error in σ, the original function being approximated were encountered. This example **suggests caution in the use of integral parameters without overseeing the original cross section approximations.** A detailed characterization of the σ error discriminated by isotope, reaction, and group showed the link between the loss function, the library structure, and the corresponding errors.

- **A second order spline space with energy minimization was combined with active learning resulting in a monotonically decreasing average error with the (optimized) support size**, achieving an order of magnitude improvement with respect to multi-linear and a bounded maximum error. **Third order approximation spaces provided no further gains**.

- **Particular attention was devoted to the evaluation cost of RKHS** which is linear with the support size and proportional to the kernel's evaluation cost. For this use case, 100 observations selected by active learning were sufficient to achieve the target accuracy, and up to 2500 were successfully tested. In this range, RKHS approximations had acceptable evaluation speeds being actually faster than B-splines up to 100 terms. **By sharing the evaluation vector across the entire cross section set required in core calculations, an evaluation time under 1E-5 s was systematically obtained.**

Kernel methods may prove quite scalable with both the input and output space increase. The relative low dimensionality of the input space considered in this work, may even penalize the library size reduction with respect to a Cartesian grid, being thus a lower bound. As long as cross section exhibit a low order dependence among the variables, the amount of observation required should grow with the cross section's complexity, here marked with the second derivative: $|\mathcal{X}_\dagger| \sim \sigma^{(2)}(x_1, \ldots, x_d) \ll \prod_{i=1}^{d} |\mathcal{X}_i|$. With respect to an output space cardinality increase, if correlated to a mean decrease in cross section's complexity, requiring thus fewer observation in each, the aggregated set of observations (i.e. the library size) could remain bounded. Attention must be drown to the shared support case were, in principle, $|\mathcal{X}_\dagger| \sim \max_{i \in \mathcal{Y}}(\sigma_i^{(2)})$. On the other hand a shared support facilitated sharing the evaluation vector, which is thus the same among the cross sections. The benefit of this consideration would increase with the output space cardinality.

6.2.3 Artificial neural networks

In this thesis artificial neural networks were introduced for modeling the cross section of a standard PWR fuel assembly performing topological and hyper-parametric studies using the PyTorch package. Results were presented in chapter five comparing to multi-linear interpolation .

- A dense enough data set using a 80%/20% split **resulted in models that generalize quite well and a complete overlapping between test and training errors was observed** allowing for an error-controlled library preparation.
- **Several data normalization strategies were analysed for both the input and output space:** the square root for the burnup and for cross sections the variance with a zero mean exhibited the best results. **Additional parametric studies included the learning rate, activation functions, and the network's parameter initialization.** A qualitative analysis showed a **detrimental effect of non smooth activation function (Relu)** in cross section approximation
- **Topological parametric studies were carried out obtaining a single layered network of 8 neurons as the smallest model that attains the target accuracy, and a two layered network of 13 neurons providing the faster convergence** in accordance with other works [100, 95, 58]. Other topologies commonly found in literature, provided no further advantage in these settings.
- Satisfactory results were obtained for **networks using the more restrictive Industry set,** though over-fitting was observed.
- Artificial neural networks can be naturally extend to multi-output models. Qualitative analysis showed that approximations from multi-output model do not differ from dedicated (single output) models.
- In a 13/13 multi-output network the equivalent amount of parameters per cross section corresponds to 2 neurons in a shallow single-output network, which is unable to fit the data. **This suggests that some commonalities of the data set are exploited by a single central model.**
- After a topological parametric study **the entire cross section set was modeled with a single network** attaining maximal library size reduction of 98%, highest evaluation speed, and an accuracy comparable to multi-linear.
- A non negligible training time of up to 4 h per cross section was observed without GPU acceleration. On the other hand, acceptable errors were obtained with a training time of a few seconds. **The training and evaluation time of a multi-output network was independent of the output's cardinality**, effectively reducing training time in a factor equal to the amount of cross section modeled. All evaluation times were quite compatible with industry standards, especially if GPU acceleration is available.

Artificial neural networks use intrinsically incremental learning methods, quite able to cope with an on-line data steam, not limited to an "off-line" preparation phase. Provided

an adequate topology and training parameters, multi-output network have indeed behaved as "model-free" mapping devices. To the author knowledge, no other, more extensive studies of cross section modeling using artificial neural networks exists to date.

6.2.4 History Effects

A review of history effects modeling was carried out analyzing different history variables for a first order Taylor expansion on σ, Σ, and k_∞. **The plutonium isotope exhibit the best performance reducing the average relative error from 0.8% to 0.1% for most cases**. However for certain control rod insertion scheme during burnup, a non monotonic behaviour of the plutonium concentration lead to an error increase for this parametrization methods suggesting the need of additional history variables.

6.2.5 Developed packages

The developed packages for the cross section modeling are available under a MIT license for kernel methods [84] and artificial neural networks [102]. The dataset used is also public [114].

6.3 Perspectives

During the development of the thesis several lines of research were identified for follow up works in cross section modeling.

6.3.1 Use case

Several specialized isotopes were considered in this thesis but for a relatively small input space, of only 3 state-variables. Assessing the scalability capabilities of the developed methodologies and their possible pitfalls would be of interest.

- About 4 state-variables are usually used in steady-state calculation and 5 or more for transient safety studies. The most challenging input space considered in the entire literature was of 15 variables [22]. A MOX fuel [14] in a transient variable range seems to be a quite demanding use case to consider, and of particular interest since they are a motivation of updates and improvements in core calculations methods [20].

- A robust test could include PWR and MOX fuels, with a 2, 8, or 20 groups of energy for the steady-state or transient state-parameters range. Additionally with

or without modeling the Xe buildup. Two fuels type that have received less attention are for MTR and especially for PHWR[4].

- In regards to augmenting the output space, a subject of debate detected in the literature review is cross section's complexity with energy groups: if a somewhat unlucky partitioning isolates the difficult resonance regions, an increase in the complexity of the most difficult cross sections could take place. Additionally for an arbitrary energy group partitioning insights from the 2-group energy scheme are not necessarily valid. On this regard an increase in difficulty is reported in [25], mentioned in [28], and also in [64]. On the other hand in [14] it is reported that the modeling problem becomes easier. The same was reported for MOX fuel in [55] and for VHTR in [30].

- Few group cross section data is usually considered relatively noise-free. This is no longer the case if Monte-Carlo tools are used [25], and testing the regularization capabilities of the developed methods could be of interest.

6.3.2 Spline interpolation

- Optimal knot placement and multiplicity control by machine learning techniques could lead to an improvement in the approximating power of splines, retaining their convenient properties. Preliminary knot vector manipulations mildly improved the representation quality.

- Boundary condition analysis in spline approximations has been raised in [46] and only partially treated in this work, and in a very pragmatical way.

- In view of the results of kernel methods, incorporating in industry application regression approximations using splines that minimize the energy could be considered (thin splines).

6.3.3 Kernel methods

- Some of the studied kernels present a very low evaluation cost. Nonetheless the kernel representation itself can be explicitly optimized for improved evaluation speed.

- Possibly a major point of improvement for industry application is the the conditioning number of the Gram matrix for finding the coefficients and their resulting absolute values, to reduce sensitivity to numerical noise.

- To further compare with some works found in the cross section literature, reproducing an interpolating polynomial space could be used to assess different grid types (e.g. regular to unstructured grids).

- The relationship between the spline spaces of kernels and B-splines could be formalizing further, i.e. the exact link between $\mathcal{H}_k$ and $\$_{K,\vec{t}}$.

[4] Of course due to the technology being less widespread and thee simpler state-parameter dependency (low EOC burnup, very high termalization with constant moderator density and low spectral interaction among fuels, no boron concentration or void fraction and thus low amount of history effects, etc.)

- Kernel design could lead to very specialized approximation spaces in a similar vain of the tailored basis functions in [57]. It's even conceivable a kernel composition which explicitly incorporates known physical phenomena, such as the xenon or plutonium build up, the depletion of fissile material or burnable poisons.

- Though kernel methods may be an effective strategy [79], they can only deal with a moderate amount of samples $\leq$5E3 which, in principle, could increase with the amount of dimensions as well as the evaluation cost of the kernel itself. This should be subject to further testing.

- The considered active learning was quite straightforward, based solely in the approximation's error. Other, possibly more sophisticated ways of assessing the observations is in view of the "coherence" of the new mapping functions with respect to the observation set. This kind of methodologies, sometimes used in on-line RKHS, could be of interest [115].

- Least Absolute Shrinkage and Selection Operator (LASSO [116]) regression could be of interest. It considers an L1 regularization allowing for the complete removal of unwanted coefficients in the preparation process (not a posteriori).

6.3.4 Artificial neural networks

- A complete parametric study in multi-output architecture on a bigger use case investigating the error bounds would be the next logical step. Further examination on the interest (if any) of common rules of thumb is of use[5].

- A larger data set could pose a challenge to multi-output artificial neural networks. In such a case a "smart" regrouping (unsupervised learning process) of cross section to be subject to to a common model could result in improved performance.

- If multi-output artificial neural networks are indeed being able to exploit common patterns in cross section data, and especially in view of possible GPU memory limitation in training, two promising use cases could be: pin-by-pin discretizations and the scattering reactions of a 20 group energy scheme with several specialized isotopes. The latter take a very large percentage of the library size, they may be less demanding in terms of accuracy (due to their low importance), tend to be easer to model, and the set could have a similar functionality with state-parameter which can be exploited with a central model. Indeed how to regroup the cross section in a convenient way for multi-output modeling is in itself an unsupervised classification learning problem.

- A fascinating area of research, well suited for relativity complex multi-out networks, is the study of its inner structure (why it works). In [109] by studying the response vector to specific input patterns the coefficients of a pseudo Taylor expansion for the networks are obtained, facilitating the examinations of its properties.

[5] In [117] it's recommended that the number of neuron in a feed forward network be about 2/3 of the input variables and the totality of the outputs; the successful preliminary studies on multi-output modeling considered far fewer neurons.

- For the relatively small networks used in this work, second order optimization algorithm that uses the Hessian matrix (Newton method) could provide an accuracy increase without significant additional cost [100]. A Newton-saddle free optimization algorithm able to converge faster and to exit local minima is successfully tested with small networks in [111]. Additional learning algorithms using feed forward artifical neural networks in engineering modeling problems can be found in [98].

- Dynamic graphs are supported by machine learning framework in Python allowing to explore adaptive networks of changing topology during training. Target errors could potentially be reached with the minimal degrees of freedom required and thus maximizing library size reduction and avoiding over-fitting.

- If required, special artificial neural networks regression techniques such as dropout [118] could be analysed.

- Though all activation functions had a similar performance for cross section modeling, networks using spline activation function subject to training have been reported to achieve high accuracies [119].

- Special interpolant networks exist [120] though requiring $|\mathcal{X}|$ neurons. Such special systems cannot be easily extended to several output systems. A somewhat ad hoc, but simple, experiment to improve the accuracy, was to consider two networks in tandem, one that fits the original function and the second one than fits the residual error. It's the author's opinion that other type of networks or learning algorithms should be studied first.

- Artificial neural networks could be particularly well suited for dealing with a large amount history-variables, not requiring them to be define in any particular structure with regards to instantaneous state-variables. Also, artificial neural networks are a convenient tool for capturing hidden relations such as, for example, the unexpected dependencies of the fast diffusion coefficient with ^{235}U concentration [22].

The nuclear engineer community as a whole could not only benefit but to contribute as well to the growing machine learning community, characterized by the free access to powerful frame-works and novel modeling approaches.

Bibliography

[1] *Energy, Electricity and Nuclear Power Estimates for the Period up to 2050.* No. 1 in Reference Data Series, Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2019.

[2] U.S. NRC, "The pressurized water reactor," 2017.

[3] A. Hébert, *Applied reactor physics.* Presses inter Polytechnique, 2009.

[4] E. Villarino and I. Mochi, "Thermal hydraulic models for neutronic and thermaly hydraulic feedback in citvap code," 09 2014.

[5] W. Stacey, "Nuclear reactor physics," *John Wiley and Sons Inc.*, 2001.

[6] *La neutronique.* No. 1 in Monographies DEN/CEA, Paris: Groupe Moniteur, 2013.

[7] D. J., *Nuclear Reactor Analysis.* 1976.

[8] C. Demazière, "Development of computational methods and their applications for the analysis of nuclear power plants," *International Journal of Nuclear Energy Science and Technology*, vol. 4, 01 2009.

[9] R. Stamm'ler and M. Abbate, *Methods of Steady-state Reactor Physics in Nuclear Design.* Academic Press, 1983.

[10] R. Sanchez *et al.*, "APOLLO2 Year 2010," *Nucl. Eng. and Technology*, vol. 42, pp. 474–499, 2010.

[11] D. Schneider *et al.*, "Apollo3 ® : Cea/den deterministic multi-purpose code for reactor physics analysis," 05 2016.

[12] J. Rhodes, K. Smith, and D. Lee, "CASMO-5 development and applications," vol. 2006, 01 2006.

[13] R. R. G. Marleau, A. Hébert, ""A User Guide for DRAGON 3.05D"," 2006.

[14] S. Sánchez-Cervera, N. García-Herranz, J. Herrero, and O. Cabellos, "Optimization of multidimensional cross-section tables for few-group core calculations," *Annals of Nuclear Energy*, vol. 69, pp. 226 – 237, 2014.

[15] S. Kuch *et al.*, "Transient validation of areva's new arcadia™ code system," p. V02BT09A019, 07 2014.

[16] S.-O. Lindahl *et al.*, "SIMULATE-4 developments," 01 2008.

[17] D. Panayotov, "Validation of westinghouse integrated code polca-t against oecd neacrp-l-335 rod ejection benchmark," vol. 3, 09 2008.

[18] A. Calloo *et al.*, "COCAGNE: EDF new neutronic core code for ANDROMEDE calculation chain," 04 2017.

[19] B. Ivanov, "Methodology for embedded transport core calculation," *PhD Thesis,*, 01 2007.

[20] T. Downar, C. H Lee, and G. Jiang, "An assessment of advanced nodal methods for mox fuel analysis in light water reactors," 11 2018.

[21] M. TATSUMI and A. YAMAMOTO, "Advanced pwr core calculation based on multi-group nodal-transport method in three-dimensional pin-by-pin geometry," *Journal of Nuclear Science and Technology*, vol. 40, no. 6, pp. 376–387, 2003.

[22] J. Dufek, "Building the nodal nuclear data dependences in a many-dimensional state-variable space," *Annals of Nuclear Energy*, vol. 38, no. 7, pp. 1569 – 1577, 2011.

[23] N. Martin, M. Riedmann, and J. Bigot, "Latest developments in the artemis tm core simulator for bwr steady-state and transient methodologies," 04 2017.

[24] U. e. a. Grundmann, "Dyn3d version 3.2 - code for calculation of transients in light water reactors (lwr) with hexagonal or quadratic fuel elements - description of models and methods -," *Wissenschaftlich-Technische Berichte*, 06 2019.

[25] J. Hou, F. Heidet, P. Gorman, and E. Greenspan, "On multi-group cross sections for breed-and-burn reactors," 11 2014.

[26] C. Demaziere and M. Stalek, "Feasibility study of coupling the CASMO-4/TABLES-3/SIMULATE-3 code system to TRACE/PARCS," 2004.

[27] R. Zivanovic and P. M. Bokov, "Cross-section parameterization of the pebble bed modular reactor using the dimension-wise expansion model," 2010.

[28] D. Botes and P. M. Bokov, "Polynomial interpolation of few-group neutron cross sections on sparse grids," *Annals of Nuclear Energy*, vol. 64, pp. 156 – 168, 2014.

[29] V. G. Zimin and A. A. Semenov, "Building neutron cross-section dependencies for few-group reactor calculations using stepwise regression," *Annals of Nuclear Energy*, vol. 32, no. 1, pp. 119 – 136, 2005.

[30] L. Cal *et al.*, "Status of reactor physics activities on cross section generation and functionalization for the prismatic very high temperature reactor and development of spatially heterogeneous codes," *ANL*, vol. 68, 2006.

[31] P. M. Bokov, "Automated few-group cross-section parameterization based on quasi-regression.," 2009.

[32] A. Rais, D. Siefman, G. Girardin, M. Hursin, and A. Pautz, "Methods and models for the coupled neutronics and thermal-hydraulics analysis of the crocus reactor at efpl," *Science and Technology of Nuclear Installations*, vol. 2015, 05 2015.

[33] S. E. Aumeier, J. C. Lee, D. M. Cribley, and W. R. Martin, "Cross-section parameterization using irradiation time and exposure for global depletion analysis," *Nuclear Technology*, vol. 108, no. 3, pp. 299–319, 1994.

[34] H. Ras, "Polx-1: A code for generating polynomial fits to neutron cross section data as functions of reactor core conditions," 1991.

[35] R. Turski, E. Morris, T. Taiwo, and J. Cahalan, "Macroscopic cross section generation and application for coupled spatial kinetics and thermal hydraulics analysis with sas-dif3dk," 08 1997.

[36] S. E. Aumeier, J. C. Lee, D. M. Cribley, and W. R. Martin, "Cross-section parameterization using irradiation time and exposure for global depletion analysis," *Nuclear Technology*, vol. 108, no. 3, pp. 299–319, 1994.

[37] I. ABB Combustion Engineering Nuclear Power, *The Advanced PHEONIX and POLCA Codes for Nuclear Design of Boiling Water Reactors*. 1999.

[38] W. Zhao and L. Xudong, "Treatment program of fuel assembly cross section parameter," 1995.

[39] J. C. B. Fiel, F. C. da Silva, A. S. Martinez, and L. C. Leal, "Parameterized representation of macroscopic cross section for pwr reactor," *Annals of Nuclear Energy*, vol. 75, pp. 736 – 741, 2015.

[40] I. M. Eduardo Villarino, "Mecánica computacional vol xxxiii," 09 2014.

[41] H. Sarabia and E. Villarino, "InterpolaciÓn de secciones eficacespara el cÁlculo de reactores de mÚltiple dependencias," 09 2014.

[42] J. C. B. Fiel and T. F. Belo, "Parameterized representation of macroscopic cross section in burn-up cycles," *Annals of Nuclear Energy*, vol. 94, pp. 472 – 477, 2016.

[43] R. Sarwar, "Validation of POLCA7 Cross section Model," *Chalmers university of technology*, 2013.

[44] T. Fujita, T. Endo, and A. Yamamoto, "A macroscopic cross-section model for bwr pin-by-pin core analysis," *Journal of Nuclear Science and Technology*, vol. 51, no. 3, pp. 282–304, 2014.

[45] K. Ivanov and M. Avramova, "Challenges in coupled thermal–hydraulics and neutronics simulations for lwr safety analysis," *Annals of Nuclear Energy*, vol. 34, pp. 501–513, 06 2007.

[46] J. K. Watson and K. N. Ivanov, "Improved cross-section modeling methodology for coupled three-dimensional transient simulations," *Annals of Nuclear Energy*, vol. 29, no. 8, pp. 937 – 966, 2002.

[47] H. Ferroukhi, O. Zerkak, and R. Chawla, "Cross-section modelling effects on pressurised water reactor main steam line break analyses," *Annals of Nuclear Energy*, vol. 36, no. 8, pp. 1184 – 1200, 2009.

[48] R. E. Bellman, *Adaptive control processes: a guided tour*. 1961.

[49] J. G. Herriot and C. H. Reinsch, "Algorithm 600: Translation of algorithm 507. procedures for quintic natural spline interpolation," *ACM Trans. Math. Softw.*, vol. 9, pp. 258–259, 1983.

[50] S.A.Smolyak, "Quadrature and interpolation formulas for tensor products of certain classes of functions," *Dokl. Akad. Nauk SSSR*, vol. 148, no. 5, pp. 1042–1045, 1963.

[51] P. Bokov, D. Botes, and V. Zimin, "Pseudospectral chebyshev representation of few-group cross sections on sparse grids," 05 2012.

[52] V. Barthelmann, E. Novak, and K. Ritter, "High dimensional polynomial interpolation on sparse grids," *Adv. Comput. Math.*, vol. 12, pp. 273–288, 03 2000.

[53] D. Botes and P. M. Bokov, "Hierarchical, multilinear representation of few-group cross sections on sparse grids," 2011.

[54] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2001.

[55] D. Botes, S. Chifamba, and P. M. Bokov, "A study of the performance of a sparse grid cross section representation methodology as applied to mox fuel," *Annals of Nuclear Energy*, vol. 99, pp. 444 – 454, 2017.

[56] S. Yum1, H. C. Shin, M. Park1, J. Choe, P. Zhang, and D. Lee1, "Application of gmdh to cross section functionalization," International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, (Jeju, Korea), 2017.

[57] T. H. Luu, M. Guillo, Y. Maday, and P. Guérin, "Use cases of tucker decomposition method for reconstruction of neutron macroscopic cross-sections," *Annals of Nuclear Energy*, vol. 109, pp. 284 – 297, 2017.

[58] B. Leniau and et al, "A neural network approach for burn-up calculation and its application to the dynamic fuel cycle code class," *Annals of Nuclear Energy*, vol. 81, pp. 125 – 133, 2015.

[59] T. H. Luu, *Amélioration du modèle de sections efficaces dans le code de cœur COCAGNE de la chaîne de calculs d'EDF*. PhD thesis, 2017. Thèse de doctorat dirigée par Guillo, Matthieu et Maday, Yvon Mathématiques Appliquées Paris 6 2017.

[60] A. Barreau, "Burn-up Credit Criticality Benchmark," Tech. Rep. Phase, IID, OECD-NEA, 2006.

[61] M. Coste-Delclaux, "Modélisation du phénomene d'autoprotection dans le code de transport multigroupe APOLLO2," *CEA-R-6114*, 2006. (in French).

[62] D. Tomatis, A. Galia, S. Pastoris, and I. Zmijarevic, "Quantification of history effects in pwr modelling," 2017.

[63] E. Szames and D. Tomatis, "A review of history parameters in pwr core analysis," *Nuclear Engineering and Design*, vol. 347, pp. 158 – 174, 2019.

[64] D. Botes, *Few group cross section representation based on Sparse Grid Methods.* PhD thesis, 04 2013.

[65] G. Radulescu, I. Gauld, G. Ilas, and J. Wagner, "An approach for validating actinide and fission product burnup credit criticality safety analyses - isotopic composition predictions," 09 2011.

[66] G. Van Rossum and F. L. Drake Jr, *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[67] S. Cook, *CUDA Programming: A Developer's Guide to Parallel Computing with GPUs.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1st ed., 2012.

[68] C. de Boor, "A practical guide to splines," 2001.

[69] "D. tomatis, "pppack"." `https://github.com/ndarmage/pppack`, 2017.

[70] R. L. Burden and J. D. Faires, "Numerical analysis," 2005.

[71] C. de Boor, "Efficient computer manipulation of tensor product," 1979.

[72] C. de Boor, "Piecewise polynomial package," 2019.

[73] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* Cambridge, MA, USA: MIT Press, 2001.

[74] C. E. Rasmussen, *Gaussian Processes in Machine Learning*, pp. 63–71. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[75] N. Akhiezer and I. Glazman, *Theory of Linear Operators in Hilber Space.* New York: Frederick Ungar Publishing, 1966.

[76] G. Wahba, *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics, 1990.

[77] C. Saunders, A. Gammerman, and V. Vovk, "Ridge regression learning algorithm in dual variables.," pp. 515–521, 01 1998.

[78] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[79] J. An and A. Owen, "Quasi-regression," *Journal of Complexity*, vol. 17, no. 4, pp. 588 – 607, 2001.

[80] D. Wu, "Pool-based sequential active learning for regression," *IEEE transactions on neural networks and learning systems*, 2018.

[81] G. STRANG, "Linear algebra and its applications (second edition)," pp. 409 – 414, Academic Press, second edition ed., 1980.

[82] L. P. Meissner, *FORTRAN 90.* USA: PWS Publishing Co., 1st ed., 1995.

[83] P. Peterson, "F2py: a tool for connecting fortran and python programs," *International Journal of Computational Science and Engineering*, vol. 4, no. 4, pp. 296–305, 2009.

[84] "E. szames, kernel methods and pool active learning." `https://github.com/BetanSz/Kernel-Methods-and-PAL.git`, 2020.

[85] L. D. Dalcin *et al.*, "Parallel distributed computing using python," *Advances in Water Resources*, vol. 34, no. 9, pp. 1124 – 1139, 2011. New Computational Methods and Software Tools.

[86] L. Yann, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 2, pp. 436 – 444, 2015.

[87] W. Liu *et al.*, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11 – 26, 2017.

[88] D. Kirk, "Nvidia cuda software and gpu parallel computing architecture," in *Proceedings of the 6th International Symposium on Memory Management*, ISMM '07, (New York, NY, USA), p. 103–104, Association for Computing Machinery, 2007.

[89] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[90] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," *NIPS*, 2017.

[91] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, no. 2, pp. 251 – 257, 1991.

[92] M. Rafiq *et al.*, "Neural network design for engineering applications," *Computers and Structures*, vol. 79, no. 17, pp. 1541 – 1552, 2001.

[93] C. Francois *et al.*, "Keras," 2015.

[94] P. Picca and R. Furfaro, "Application of extreme learning machines to inverse neutron kinetics," *Annals of Nuclear Energy*, vol. 100, pp. 1 – 8, 2017.

[95] A. Hedayat *et al.*, "Estimation of research reactor core parameters using cascade feed forward artificial neural networks," *Progress in Nuclear Energy*, vol. 51, no. 6, pp. 709 – 718, 2009.

[96] Y. Liu, N. Dinh, Y. Sato, and B. Niceno, "Data-driven modeling for boiling heat transfer: Using deep neural networks and high-fidelity simulation results," *Applied Thermal Engineering*, vol. 144, pp. 305 – 320, 2018.

[97] D. Wijayasekara *et al.*, "Optimal artificial neural network architecture selection for performance prediction of compact heat exchanger with the ebalm-otr technique," *Nuclear Engineering and Design*, vol. 241, no. 7, pp. 2549 – 2557, 2011.

[98] S. Mirvakili *et al.*, "Developing a computational tool for predicting physical parameters of a typical vver-1000 core based on artificial neural network," *Annals of Nuclear Energy*, vol. 50, pp. 82 – 93, 2012.

[99] D. Liang *et al.*, "Rapid nuclide identification algorithm based on convolutional neural network," *Annals of Nuclear Energy*, vol. 133, pp. 483 – 490, 2019.

[100] D. Scozil *et al.*, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43 – 62, 1997.

[101] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv e-prints*, p. arXiv:1412.6980, Dec 2014.

[102] E. Szames, "E. szames, FFANN for regression." `https://github.com/BetanSz/FFANN-for-regression`, 2019.

[103] R. Raturi, "Large data analysis via interpolation of functions: Interpolating polynomials vs artificial neural networks," 2018.

[104] M. G. Fernandez *et al.*, "Nuclear energy system's behavior and decision making using machine learning," *Nuclear Engineering and Design*, vol. 324, pp. 27 – 34, 2017.

[105] A. Lapedes and R. Farber, "How neural nets work," in *Proceedings of the 1987 International Conference on Neural Information Processing Systems*, NIPS'87, (Cambridge, MA, USA), pp. 442–456, MIT Press, 1987.

[106] X. Cheng, B. Khomtchouk, N. Matloff, and P. Mohanty, "Polynomial regression as an alternative to neural nets," 06 2018.

[107] G. M. Maggiora, D. W. Elrod, and R. G. Trenary, "Computational neural networks as model-free mapping devices," *Journal of Chemical Information and Computer Sciences*, vol. 32, no. 6, pp. 732–741, 1992.

[108] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1 – 15, 2018.

[109] C. Cui and T. Fearn, "Modern practical convolutional neural networks for multivariate regression: Applications to nir calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 182, 07 2018.

[110] A. Virmaux and K. Scaman, "Lipschitz regularity of deep neural networks: analysis and efficient estimation," in *Advances in Neural Information Processing Systems 31* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), pp. 3835–3844, Curran Associates, Inc., 2018.

[111] R. Pascanu, Y. N. Dauphin, S. Ganguli, and Y. Bengio, "On the saddle point problem for non-convex optimization," 2014.

[112] E. Szames, K. Ammar, D. Tomatis, , and J. Martinez, "Few-group cross sections library by active learning with spline kernels," *Proceding of the conference: Physics of reactor conference, Cambridge, United Kingdom*, 06 2020.

[113] E. Szames, K. Ammar, D. Tomatis, , and J. Martinez, "Few-group cross sections modeling with artificial neural networks," *Proceding of the conference: Physics of reactor conference, Cambridge, United Kingdom*, 06 2020.

[114] E. Szames, "E. szames, few-group cross section data set." `https://github.com/BetanSz/Few-group-XS-data-set.git`, 2020.

[115] P. Honeine, *Contributions to signal processing with kernel-based machine learning.* Habilitation à diriger des recherches, Habilitation à Diriger des Recherches, de l'Ecole Doctorale de l'Université de Technologie de Compiègne, Dec. 2013. 164 pages.

[116] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[117] S. Thamatampalli, G. Vinod, R. Saraf, A. Ghosh, and H. Kushwaha, "Application of artificial neural networks to nuclear power plant transient diagnosis," *Reliability Engineering and System Safety*, vol. 92, pp. 1468–1472, 10 2007.

[118] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 06 2014.

[119] A. S. Douzette, "B-splines in machine learning," 2017.

[120] B. Llanas and F. Sainz, "Constructive approximate interpolation by neural networks," *Journal of Computational and Applied Mathematics*, vol. 188, no. 2, pp. 283 – 308, 2006.

[121] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122, 2013.

[122] M. Erwin, M. Larry, and Z. Baocheng, "Reactor physics methods development at Westinghouse," in *Proc. Int. Conf. Nuclear Energy for New Europe 2007*, 2007.

[123] T. Hino, J. Miwa, T. Mitsuyasu, Y. Ishii, M. Ohtsuka, K. Moriya, K. Shirvan, V. Seker, A. Hall, T. Downar, P. M. Gorman, M. Fratoni, and E. Greenspan, "Core design and analysis of axially heterogeneous boiling water reactor for burning transuranium elements," *Nuclear Science and Engineering*, vol. 187, no. 3, pp. 213–239, 2017.

[124] A. Grossetete, "Le pilotage de l'EPR: mode T," *Revue générale nucléaire*, no. 3, pp. 37–41, 2007. (in French).

[125] M. Onoue, T. Kawanishi, W. R. Carlson, and T. Morita, "Application of MSHIM core control strategy for Westinghouse AP1000 nuclear power plant," *GENES4/ANP2003*, pp. 15–18, 2003.

[126] L. van der Merwe and C. J. Hah, "Reactivity balance for a soluble boron-free small modular reactor," *Nuclear Engineering and Technology*, 2018.

[127] J. Choe, H. C. Shin, and D. Lee, "Optimal Control Rod for Boron-Free Small Modular PWR," in *Proceedings of The 20th Pacific Basin Nuclear Conference*, (Singapore), pp. 809–815, Springer Singapore, 2017.

[128] H. Kim, C. Chung, M. Yahya, and Y. Kim, "Core reactivity control for a soluble boron free small modular reactor," *American Nuclear Society*, vol. 116, 2017.

[129] A. Dall'Osso, "Modeling local history effects in the nodal core ARTEMIS," *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering*, 2017.

[130] R. Mosteller, "Impact of moderator history on physics parameters in pressurized water reactors," *Nuclear Science and Engineering*, vol. 98, no. 2, pp. 149–153, 1988.

[131] P. Guimarães and E. Müller, "Parameterization of two-group nodal cross section data for POLCA-T BWR transient applications," in *International Conference on Mathematics, Computational Methods, & Reactor Physics*, 2009.

[132] T. Bahadir, S. Lindahl, and S. Palmtag, "SIMULATE-4 multigroup nodal code with microscopic depletion model," *Mathematics and Computation, Supercomputing, Reactor Physics and Nuclear and Biological Applications*, 2005.

[133] K. R. Rempe, K. S. Smith, and A. F. Henry, "SIMULATE-3 pin power reconstruction: Methodology and benchmarking," *Nuclear Science and Engineering*, vol. 103, no. 4, pp. 334–342, 1989.

[134] C. Lee, Y. Kim, J. Song, and C. Park, "Incorporation of a new spectral history correction method into local power reconstruction for nodal methods," *Nuclear science and engineering*, vol. 124, no. 1, pp. 160–166, 1996.

[135] T. Iwamoto and M. Yamamoto, "Advanced nodal methods of the few-group BWR core simulator NEREUS," *Journal of nuclear science and technology*, vol. 36, no. 11, pp. 996–1008, 1999.

[136] D. Baturin and S. Vygovskii, "Taking account of the spectral history of fuel burnup during the preparation of the neutron-physical constants for VVER-1000 fuel assemblies," *Atomic Energy*, vol. 90, no. 4, pp. 267–272, 2001.

[137] J. M. A. Beltrán, "Fundamentals of 3-d neutron kinetics and current status," in *Proceedings of THICKET-2008: Seminar on the Transfer of Knowledge Gained through CSNI Activities in the Field of Thermal-hydraulics*, (Francia), Nuclear Energy Agency of OECD, 2008.

[138] Y. T. Toshikazu Ida, "Two group micro-depletion correction model for alpha/phoenix-p/anc code system," 2000.

[139] B. Zhang *et al.*, "A study on generic two-group cross-section representation methodology," 2002.

[140] M. Larry, Z. Baocheng, and S. D. Jung, "Pwr core modeling using the nexus once-through cross-section model," 2006.

[141] I. Bilodid and S. Mittag, "Use of the local Pu-239 concentration as an indicator of burnup spectral history in DYN3D," *Annals of Nuclear Energy*, vol. 37, no. 9, pp. 1208–1213, 2010.

[142] Y. Bilodid, "Spectral History Modeling in the Reactor Dynamics Code DYN3D," *Technischen Universitat Dresden*, 2014.

[143] Y. Bilodid, E. Fridman, and S. Kliem, "Microscopic depletion with the correction of microscopic cross section nodeal difussion code dyn3d," *PHYSOR*, 2016.

[144] Girieud, "SCIENCE, Version 2: The Most Recent Capabilities of the Framatome 3D Nuclear Code Package," *FRAMATOME Nuclear Operations*, 2000.

[145] K. Rosman and P. Taylor, "Isotopic composition of the elements 1997," *Pure and Applied Chemistry*, vol. 70, no. 1, pp. 217–235, 1998.

[146] Y. Li, S. Gao, H. Wu, L. Cao, and W. Shen, "Pwr few-group constants parameterization analysis," *Progress in Nuclear Energy*, vol. 88, pp. 104–117, 04 2016.

Appendix A

Spline interpolation using B-splines

A.1 Piece-wise polynomial representation with B-splines theoretical discussion

In order to assure that $\$_{K,t} = \Pi_{<K,\xi,\nu}$ the continuity conditions in $\vec{\nu}$ needs to be expressed in the space $\$_{K,\vec{t}}$ by an adequate construction of the knot vector $\vec{t}$ as defined by the Curry and Schoenberg theorem 2. The theorem enables the construction of a basis of $\Pi_{K,\vec{\xi},\vec{\nu}}$ by providing the recipe for an appropriate $\vec{t}$. The set of knots $\{t_j : t_j \in (a,b), \quad K+1 \leq j \leq n\}$ are called interior knots.

Theorem 2 (Curry and Schoenberg theorem)**.** For a given strictly increasing sequence $\vec{\xi} = (\xi_j)_1^{l+1}$ and a non-negative integer sequence $\vec{\nu} = (\nu_j)_2^l$ with $\nu_j \leq K$, $\forall j$, set

$$n = K + \sum_{j=2}^{l}(K - \nu_j) = Kl - \sum_{j=2}^{l}\nu_j = \dim(\Pi_{<K,\vec{\xi},\vec{\nu}}), \tag{A.1}$$

and let $\vec{t} = (t_j)_1^{n+K}$ be the non-decreasing sequence obtained from $\vec{\xi}$ by the following two requirements:

- for $2 \leq j \leq l$ the number ξ_j occurs exactly $K - \nu_j$ times in $\vec{t}$.
- $t_1 \leq t_2 \leq \cdots \leq t_K \leq \xi_1$ and $\xi_{l+1} \leq t_{n+1} \leq \cdots \leq t_{n+K}$.

then, the sequence $B_{1,K}, \ldots, B_{n,K}$ of B-splines of order K and knot sequence $\vec{t}$ is a basis for $\Pi_{<K,\vec{\xi},\vec{\nu}}$ considered as functions on $I_{K,\vec{t}} = [t_K, t_{n+1}]$, i.e.

$$\$_{K,\vec{t}} = \Pi_{<K,\vec{\xi},\vec{\nu}} \text{ on } I_{K,\vec{t}}. \tag{A.2}$$

The choice at the endpoints of the knot vector is open, but it is customary to consider

$$\begin{aligned} \xi_1 = a = t_1 = t_2 = \cdots = t_K < t_{K+1} \leq \ldots \\ \cdots \leq t_n < t_{n+1} = t_{n+2} = \cdots = t_{K+n} = b = \xi_{l+1}, \end{aligned} \tag{A.3}$$

which translates into $\nu_1 = \nu_{l+1} = 0$ so no continuity condition is imposed in the endpoints $\xi_1 = a$ and $\xi_{l+1} = b$.

A.2 Difficult cross section for Spline approximation

In Fig. A.1 the cross section that exhibited the highest error in Fig. 3.4 are presented where it can be seen that for the majority of the approximations the oscillations have been controlled as the support increases.

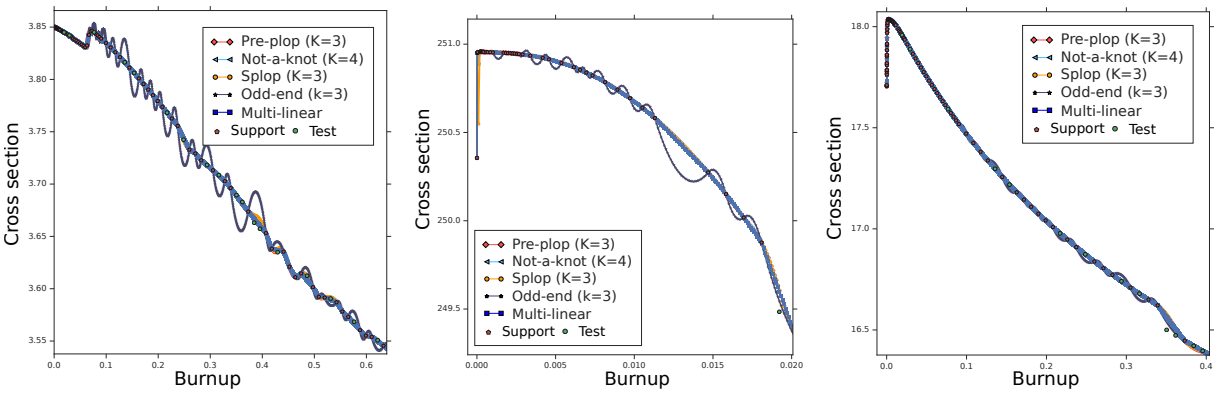


Figure A.1: Spline approximation for difficult cross section for $d = 1$. Oscillation are reduced with a support of 119 points per cross sections.

However in Fig. A.2a and Fig. A.2b the error on $\mathcal{X}_{Xe}$ is presented for $d = 1$ and $d = 3$ using the standard knot vector constructions. It can be noted that the solution greatly deteriorates and no significant library reduction can be obtained with higher order splines. An effort was made for trying to define modified knot vectors, relaxing the smoothness conditions by increasing the multiplicity of certain knots, as to reduce oscillations and increase the accuracy. The proposed knot vector are presented in Table A.1 and Table A.2. The results are presented in Fig. A.2c and Fig. A.2d for $\mathcal{X}_{Xe}$ in $d = 1$. Though some improvement are noticed the library reductions remain small. Many strategies not presented here resulted in divergent approximations, especially for the odd-end knot vector.

Name	Modification
M1	Relaxes the smooth condition at the first inner k
M3	Every three inner knots the third gains multiplicity $K-2$.
M4	The first 1/3 of the knot vector is composed of knots of multiplictiy $K-2$
M6	The third knot gains multiplicity $K-1$ at the expense of its right and left neighbor.
M8	The third knot gains multiplicity $K-1$ at the expense of its right and left neighbor
M9	The second knot gains multiplicity $K-1$ at the expense of its right and left neighbor
M10	The knots of the first 1/3 of the vector gain multiplicty $K-1$ at the expense of its right and left neighbors.

Table A.1: Modified Not-a-Knot splines with different smoothness conditions for the first burnup values.

Name	Modification
M1	The first inner knot (position $K+1$) gains multiplicity $K-1$ at the expense of the knot in the right.
M3	The second inner knot (position $K+2$) gains multiplicity $K-1$ at the expense of the knot in the right.
M4	The third inner knot (position $K+3$) gains multiplicity $K-1$ at the expense of the knot in the right.
M6	The first and third inner knot gains multiplicity $K-1$ at the expense of the knots in the right.
M10	Knots located in the first third of knot vector gain multiplicty $K-1$ at the expense of its right neighbor.

Table A.2: Modified Splop splines with different smoothness conditions for the first burnup values.

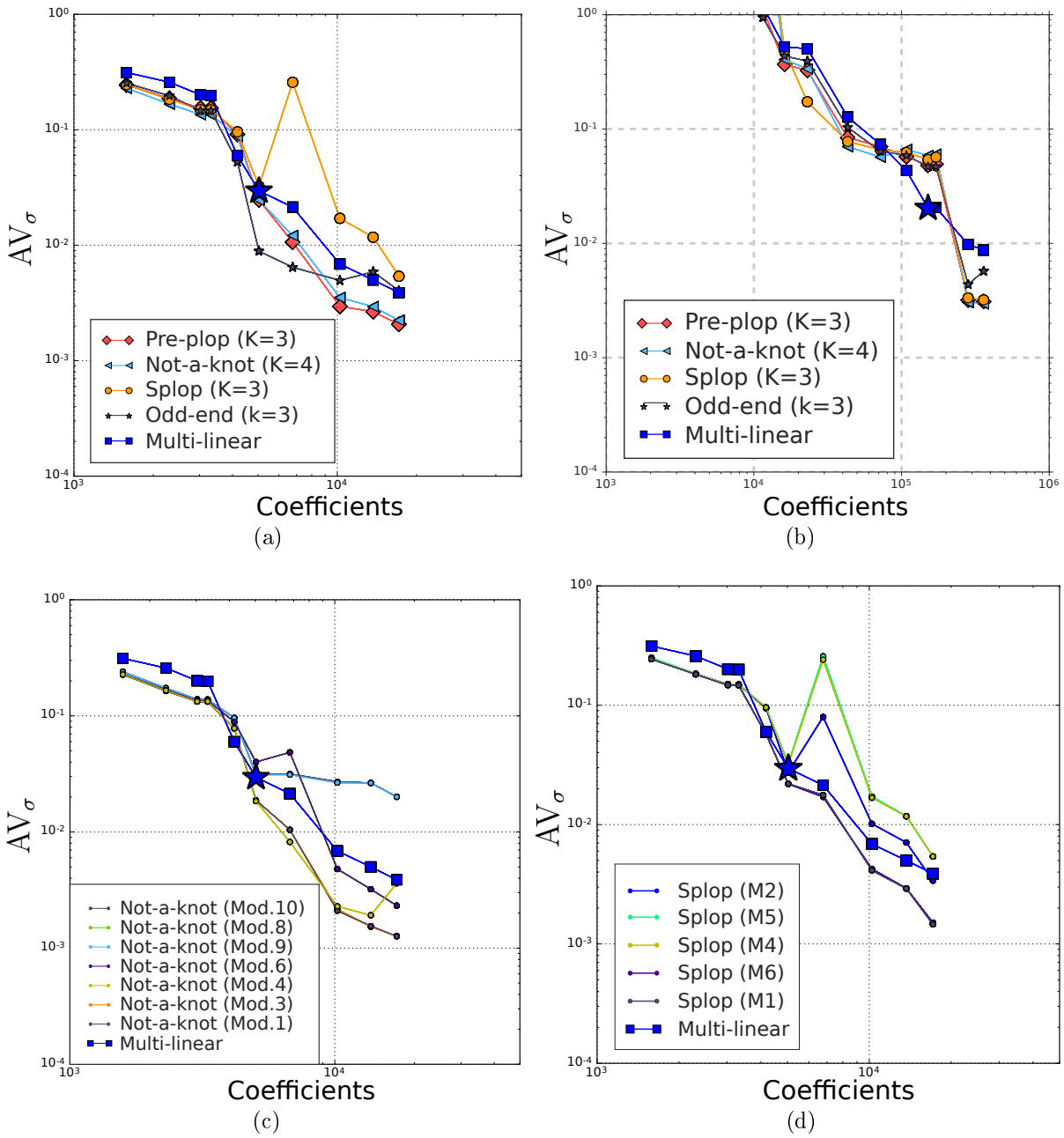


Figure A.2: Standard and modified higher order spline approximation for using the using $\mathcal{T}_{Xe}$ test.(a) Standard knot vector constructions exhibit a poor performance for $d = 1$. (b) and (c) Modified not-a-knot and Splop knot vector respectively. The modified Splop vector M1 and M6 improve significantely the approximation accuracy. The not-a-knot modified vector M1 and M3 also improve the accuracy for $d = 1$. (d) FRE for a) Standard knot vector constructions using the complete test set. (b) Modified not-a-knot knto vector for the complete test set. (c) Modified splop knto vector for the complete test set.

Appendix B

Kernel Methods

B.1 Other kernels

In practice the kernel is defined directly implicitly generating a mapping from $\mathcal{X}$ to $\mathcal{H}$. A vas quantity of "standard" kernel choices are available reproducing different function spaces. Many of them are directly included in the scikit-learn package [121]. An extensive and detailed analysis on the characteristic of each kernel, their hyper-parameters and the resulting approximating quality is beyond the scope of this work. However for every kernel shown in Table B.1 a full grid support was used with an hyper parameter search for every cross section [1]. The resulting errors are presented in Fig. B.1. Though a finer analysis on the range of interest for the hyper-parameters and maybe kernel composition would probably have surpass mulit-linear interpolation, its interesting to note that this is not the case for a blind use of tools "out of the box". For each support an independent full grid hyper parameter search is carried out for every cross section. Yet the only promising kernel in these settings is the Chi2 kernel surpassing multi-liner for some supports.

s

B.2 High order Bernoulli kernel methods

The first order Bernoulli kernel was not used in the main part of the analysis since its order 1 kernel did not reproduce a multi-linear space, and it actually had a higher error. However, a similar analysis on pre-conditioning and regularization was conducted which resulted in similar optimal values for this kernel. Cross section errors on $\mathcal{T}_{flat}$ are presented in Fig. B.2 using the Bernoulli kernel of order 1,2 and 3 and the RIXS loss function. Similar results to those of the spline kernel are obtained for the order 1 and 2. An order 3 kernel does not produce further improvement.

[1] Using the function GridSearchCV of the scikit-learn package using a 5 fold cross-validation and an $r2$ score. A regularization coefficient $\lambda = [1e-2, 1e-3, 1e-4, 0]$ following the package recommendations for these kernel.

Name	Kernel	Hyper-parameter
InnerProduct	$x^T y$	No
Sigmoid	$tanh(\gamma\langle x, y\rangle + c_0)$	γ, c_0
Laplacian	$e^{-\gamma\|\|x-y\|\|_1}$	γ
Polynomial	$(\gamma x^T y + c)^d$	γ, c_0, d
Radial Basis Functions	$e^{-\gamma(\|\|x-y\|\|^2)}$	γ
Matern	see [121]	a, b
Chi2	$e^{(-\gamma\sum_i[(x_i-y_i)^2/(x_i+y_i)])}$	γ

Table B.1: Other kernels analyzed. Many use hyper-parameters which are, with their corresponding possible values $\gamma = [0.1, 10, 100]$, $c_0 = [0.1, 1, 10, 100]$, $d = [2, 3, 4]$, $a = [0.1, 1, 10]$ and $b = [0.5, 1.5, 2.5]$. An independent full grid search is carried out for every cross section.

B.2.1 Kernel method's coefficient distribution and shrinking

In Fig. B.3b the coefficient are plotted ordered by the module for second order RKHS approximation with different regularization coefficients λ. The absolute value and to a certain extent the "slope" increases indirectly proportional to λ. In Fig. B.3a the conditioning number in function of the library size is provided for an RKHS approximation using the Bernoullie kernel for orders 1,2 and 3. It can be observed that preconditioning helps but regularization is needed for the second and third order.

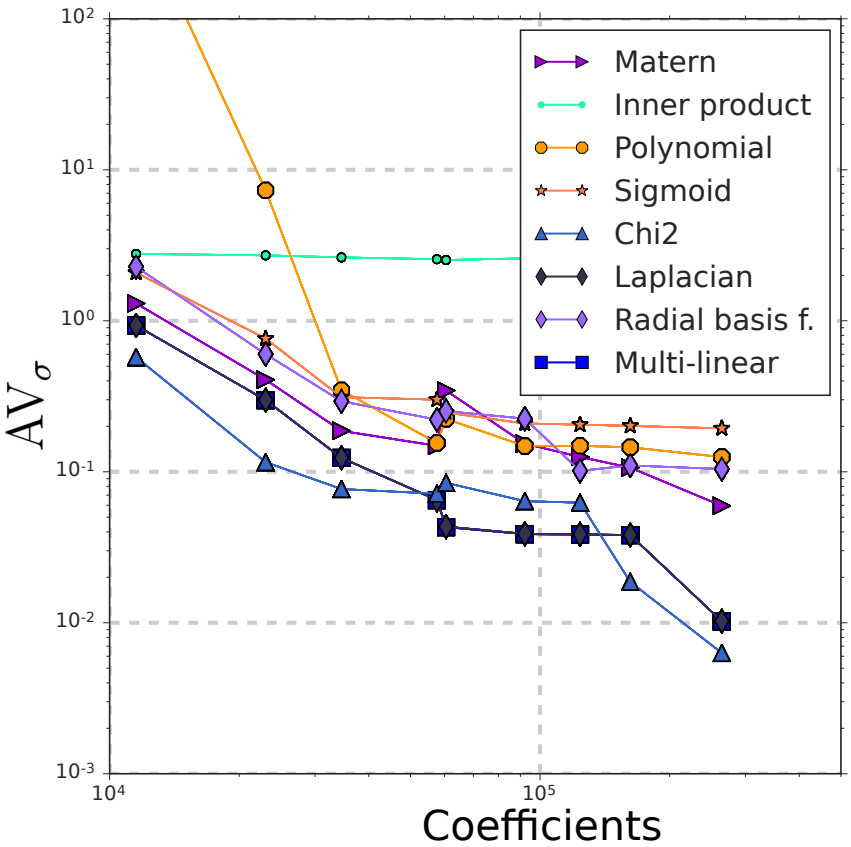


Figure B.1: Cross section error with other kernels using and hyper-parameter full grid search. Only the Chi2 exhibits a better accuracy than multi-linear for some supports.

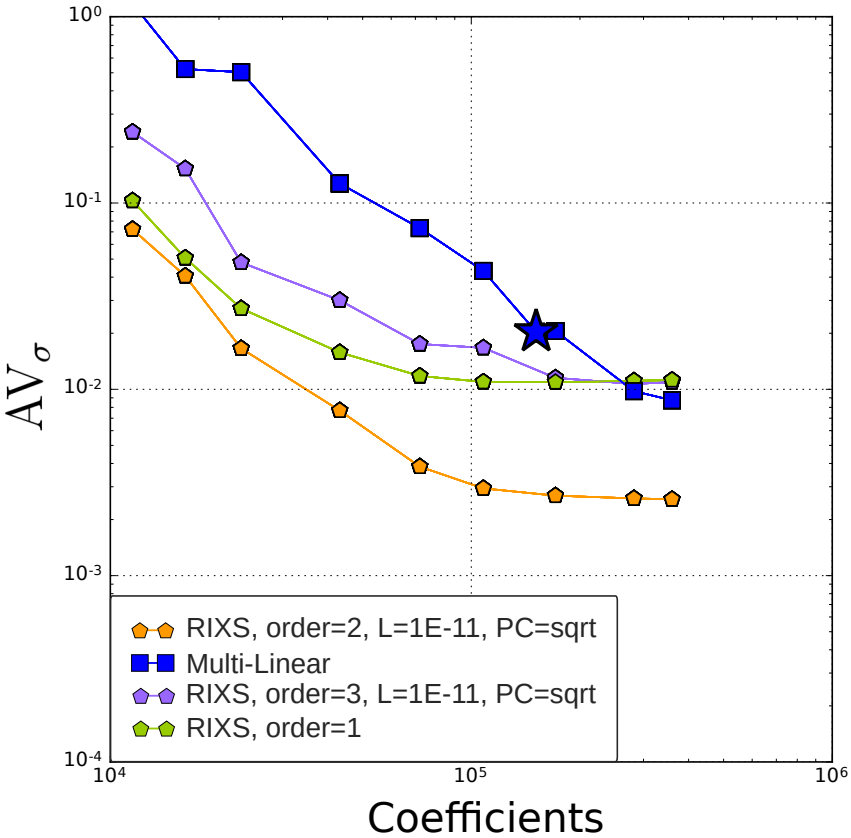
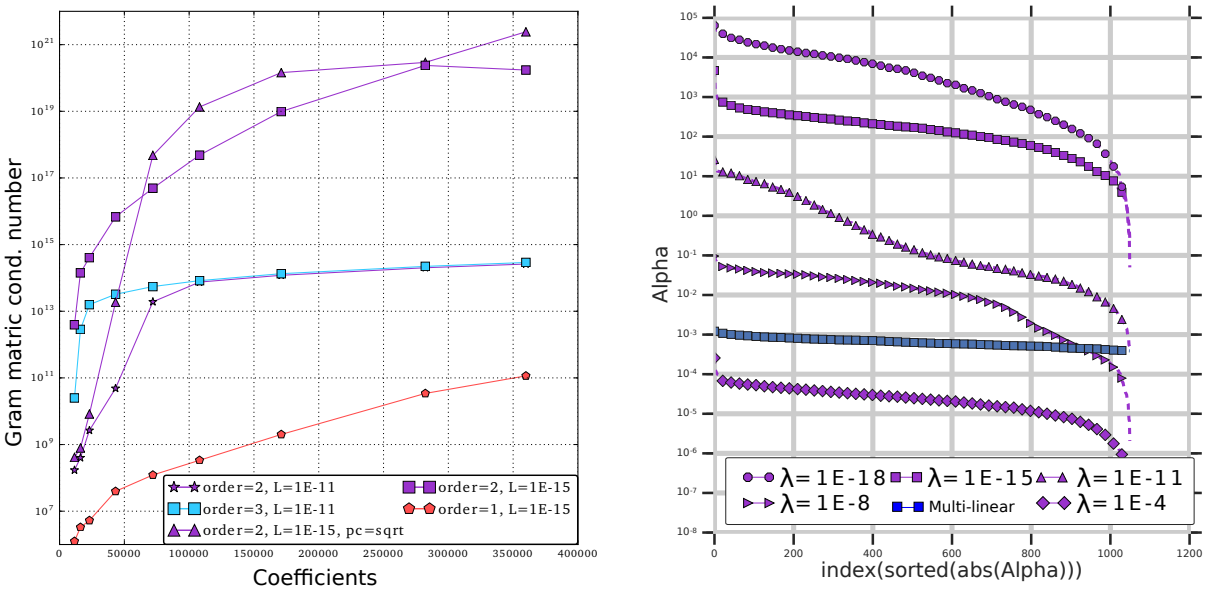


Figure B.2: Cross section average relative error for the Bernoulli kernel of order 1, 2 and 3. An order 3 in unable to provide further improvement in accuracy.



(a) (b)

Figure B.3: (a)Matrix conditioning number for the spline kernel of order 1, 2 and 3 with different regularization terms in function of the total library size for $\mathcal{Y} = 144$. Very large conditioning number can be encountered. (b) Order by absolute value coefficient distribution ($\vec{\alpha}$) for the second order spline kernel for the Industry set ($\mathcal{S}_\star$). Large variation from maximum to minimum values can be observed, inversely proportional to the regularization coefficient (λ).

Appendix C

History effect modeling

A review of the different techniques reported in literature to account for the history effects in core calculations for PWR is presented [63]. This is a somewhat separate work that required a different calculation scheme and is thus presented in its totality as a separate Annex. With respect to the nomenclature developed in Section 1.3.5 its important to note that $\sigma(\vec{x}) = \sigma(x_1, \ldots, x_d) = \sigma(x_{h,1}, \ldots, x_{h,n}; x_{b,1}, \ldots, x_{b,m}) = \sigma(Bu, \vec{\iota}, \vec{\theta})$ being the instantaneous state-variables $\vec{\iota} = x_{b,1}, \ldots, x_{b,m}$ and the history variables $\vec{\theta} = x_{h,1}, \ldots, x_{h,n}$. Another nomenclature related point to make is that in Section 2.2.1 a single depletion history was used for provide branch calculation on instantaneous state-variables (fuel temperature and boron concentration) while in this annex the process is different. Calculations in different state-variables throughout the fuel evolution (marked with a $\mathcal{B}$) are considered but without branch calculations. The nominal fuel evolution is $\mathcal{B}_N$ and other, independent fuel evolution with control rods ($\mathcal{B}_{c1}$, $\mathcal{B}_{c2}$), moderator density ($\mathcal{B}_I$, $\mathcal{B}_O$) or fuel temperature ($\mathcal{B}_f$) are carried out.

C.1 History effects phenomena

The few-group cross section data depends on the weighting neutron spectrum used at condensation and homogenization. In turn, this spectrum follows from the isotopic content resulting from previous depletion conditions. Hence, the inter-dependence of the spectrum and of the local concentrations at these given conditions defines a depletion history.

Whenever temporary deviations from the nominal conditions are noticed along the real exposure observed in the nodes of the core calculations, local spectra gets farer from those assumed a-priori at cross section preparation and the predicted homogenized cross sections may not be sufficiently accurate. Exposure under different spectral conditions yields inevitably different nuclide contents and homogenizing spectra, thus causing a "history effect" or spectrum-induced error on the homogenized cross sections from lattice calculations.

This phenomenon is of particular importance in BWR[1] reactor analysis due to the severe change of the void fraction along the core's height and permanently controlled fuel assemblies in the bottom region. For instance, in the NEXUS package developed at Westinghouse three different depletion histories are usually considered [122] while some advanced BWR designs show up to four [123].

On the contrary, a single nominal history is usually chosen with PWRs, where branch calculations are motivated by relatively short insertion periods of control elements during power operation. Effects due to history phenomena are usually considered as second order terms in PWR analysis, and this is certainly true for short deviations in time from nominal conditions.

Nonetheless, the need of load following capabilities has called for more controlled core patterns in PWRs, especially in France where the electrical power production is largely coming from nuclear power plants. In addition, it is possible to notice as a general trend in the design of new PWR units, like the EPR from AREVA [124] or the AP1000 from Westinghouse [125], enhanced control by gray banks permanently inserted or with prolonged insertion, with the goal of reducing the operational costs of the chemical shim. Moreover, innovative boron-free Small Modular Reactors envision control rods as a central feature for reactivity control at normal operation [126, 127, 128], motivating further the reason of this work.

The presence of control elements in the assemblies hardens the neutron spectrum because of increased absorption in the thermal range and decreased efficiency in the neutron slowing down rate. The strength of the control bank in the neutron absorption is indicated as gray or black. The first means only stainless steel in the rod cluster, while rods made of boron carbide (B_4C) are also used for the second. Safety issues could arise such as an unexpected positive reactivity excess after the extraction of a bank inserted for a long period of time, because of a mistaken condensation by a softer spectrum and for a higher amount of fissile material in the reactor with respect to the computed amount. Unfortunately, this kind of situations are rather frequent in real calculations, and they can be source of major error in a few extreme cases as shown by Tomatis et al. [62].

A similar process takes place with the enthalpy increase of the coolant along the fuel element's channels resulting in different moderating conditions for neutrons. Although higher temperature promotes the scattering with hydrogen bound in water, the reduction of the moderator density decreases the neutron slowing-down rate. Spectrum hardening is expected towards the core outlet, and enhanced thermalization at inlet.

Different zones in the core, exposed to different power levels, can then undergo different spectral histories. This requires the introduction of the history parameters, in addition to the customary instantaneous ones.

[1] A Boiling Water Reactor (BWR) is the second most common type of electricity generating NPP. With a bigger vessel, designed for a lower pressure than the PWR, the water in the primary circuit is turned directly into steam driving an integrated steam turbine. A significant axial asymmetry in the moderator density results from this design which is somewhat offset by using control rods inserted from below the core in normal reactor operation.

C.2 History parameters

In standard cross section preparation the isotopic changes along fuel exposure are represented by the burnup (Bu), which can be considered as a first history parameter. The most probable physical conditions of exposure in the reactor define the nominal depletion history ($\mathcal{B}_N$) and the state parameters are introduced to model a departure from such conditions.

Provided linear independence of all the parameters in the model, the general functionalization of the cross section is $\sigma(Bu, \vec{\iota}, \vec{\theta})$, where $\vec{\iota} = (\iota_1, \ldots, \iota_d)$ represents d instantaneous state parameters and $\vec{\theta} = (\theta_1, \ldots, \theta_h)$ features the set of h history parameters. The dependence on the first parameters is reproduced by branch calculations, while new depletion histories are used for the second ones. These last histories are called "off-nominal" in this work. A Cartesian grid is then constructed with all these calculation points provided that branch calculations are performed on all off-nominal calculations.

Of course, the selection of the supporting depletion calculations is driven by feasibility reasons according to the available computational resources. For instance, the size of the data libraries for core calculations is estimated in the order of several gigabytes, and even if computing a lattice calculation point is in the order of a few tenths of seconds, the library production time may become promptly prohibitive for the industrial work-frame. In the previous generation of core calculations few particularized isotopes ($I << 10$) were used and history parameters were a needful requirement for realistic simulations. Denser depletion chains in core calculations have mitigated the history effects, moving the attention on the pin power distributions. Indeed non-conservative errors in the power form factors are reported in [23] and the use of history parameters was proposed in [129] to reduce them.

In an early work [130] the moderator history was used for computing empirical corrective terms stored in look-up tables. It is defined as a burnup-averaged quantity of the instantaneous moderator density δ:

$$\mathrm{MH} = \frac{1}{Bu} \int_0^{Bu} \delta(Bu')dBu'. \tag{C.1}$$

Later, in the code POLCA [131], cross sections calculated at hot full power (HFP) condition were corrected, with additive terms presenting a quadratic dependence with the MH. Nowadays, many industrial code simulators take into account history parameters, like SIMULATE-4 [132] or ARCADIA-BWR [23]. However, when several history parameters are considered, a large number of additional depletion histories is required. This occurs with the code NECP [44] that accounts for the control rod and fuel temperature history for a total of 30 additional depletion calculations.

According to the available literature, a single history parameter was considered in PWR modeling. A common assumption asserts that changes in the cross sections induced by spectral variations can be considered as independent of its causing phenomenon [133]. The spectral history SH [134, 135] is defined as:

$$\mathrm{SH} = \frac{1}{Bu}\int_0^{Bu} \frac{\mathrm{SI}(Bu')}{\mathrm{SI}_N(Bu')} dBu', \tag{C.2}$$

with the ratio between the actual spectral index (SI) and the one from the nominal calculation here indicated by the subscript N. The spectral index in the two-group model is the ratio of the fast flux ϕ_1 over the thermal flux ϕ_2 (cut-off energy at 0.625 eV):

$$\mathrm{SI} = \phi_1/\phi_2. \tag{C.3}$$

Although other definitions of the SI are noticed in literature [136, 137], they are not considered in this work in virtue of the simpler and standard definition above. For example, the code NEREUS [135] is reported to add a quadratic correction of the kind $\sum_{i=0}^{2} a_i(Bu)(\mathrm{SH}-1)^i$ to the macroscopic cross sections.

Despite the spectral index is here considered as an instantaneous parameter, the code PHOENIX/ANC used the SI evaluated at HFP and hot zero power (HZP) for correcting microscopic cross sections with a second order polynomial, whose coefficients came by ordinary least square regression. This correction was applied only to the fission and to the absorption thermal cross sections of ^{235}U and ^{239}Pu [138]. The model was later enhanced to correct also the fast group data [139], as reported in the code package NEXUS by Westinghouse dedicated to LWR simulation [140].

In [141] relative differences in the concentration of the ^{239}Pu are considered, here simply noted with the chemical symbol and the mass number, as a reliable indicator of the spectral history in the fuel. Changes in the homogenized cross sections affected by the history were found to be proportional to the square root of ^{239}Pu so a new history parameter was proposed with the following linearization [142]:

$$\mathrm{P} = \sqrt{^{239}\mathrm{Pu}}. \tag{C.4}$$

In the same work, it is explicitly stated as underling hypothesis that correlations in the change of the isotopic content causing the history effect can be modeled with a unique history parameter. This method was implemented in the code DYN3D [143].

The SCIENCE code package by Framatome uses the ratio between the concentrations of all plutonium isotopes (Pu) and ^{238}U to recover the history [144], i.e.

$$\mathrm{PU} = \mathrm{Pu}/^{238}\mathrm{U}. \tag{C.5}$$

A new irradiation history with a different moderator density provides the additional off-nominal calculation for the parameterization. This interesting feature is retained in this work for the implementation of the different methodologies under study.

Several authors suggest a unique linear correction term [133, 130, 141, 43] for the history effect. In general, this can be formalized through first order Taylor expansions around the history parameter evaluated on the nominal depletion history, i.e. $\sigma_N =$

$\sigma(Bu, \vec{\iota}, \theta_N(Bu))$. For every isotope, reaction, energy group and homogenization region it is:

$$\sigma(Bu, \vec{\iota}, \theta) \simeq \hat{\sigma} = \sigma_N \left[1 + \frac{\theta_N}{\sigma_N} \frac{\partial \sigma}{\partial \theta} \bigg|_{\theta=\theta_N} \left(\frac{\theta}{\theta_N} - 1 \right) \right]. \tag{C.6}$$

The history parameter θ results from the actual node's state in the core calculation, while the value of θ_N from the nominal depletion history must also be determined online within the core calculations. The derivative is approximated by finite differences requiring the additional "off-nominal" calculations at different depletion conditions. Although Equation C.6 holds for any $\vec{\iota}$, the derivative term is assumed as weakly dependent on $\vec{\iota}$ itself, so it is here evaluated at a given $\vec{\hat{\iota}}$ as:

$$\frac{\partial \sigma}{\partial \theta} \bigg|_{\theta=\theta_N} \simeq \frac{\partial \sigma}{\partial \theta} \bigg|_{\theta=\theta_N, \vec{\iota}=\vec{\hat{\iota}}} \simeq \frac{\sigma_{off}(Bu, \vec{\hat{\iota}}, \theta_{off}(Bu)) - \sigma_N(Bu, \vec{\hat{\iota}}, \theta_N(Bu))}{\theta_{off} - \theta_N}. \tag{C.7}$$

This allows a reduction in the number of lattice calculations and also in the storage requirements. Considering $\vec{\hat{\iota}}$ at the nominal state permits the evaluation at the most probable conditions of operations but requires an additional branch calculation on the supporting off-nominal history. This can be avoided by evaluating the derivative at the off-nominal condition. Of course, the same $\vec{\hat{\iota}}$ is used in Equation C.7 for the cross sections in order to capture only the target history effect. Lastly, σ_N is customarily reconstructed in the core calculation by interpolation on $\vec{\iota}$ and Bu.

C.3 Use cases for numerical tests

The irradiation histories calculated by the lattice code are denoted by the character $\mathcal{B}$ in Table C.1, with the subscripts indicating the different depletion conditions. These histories are reproduce by the curves $(Bu, \iota_{\mathcal{B}}(Bu), \theta_{\mathcal{B}}(Bu))$. Indeed, they allow to reproduce the behavior of the core's node delivering the actual θ values for Equation C.6 and the isotopes concentrations for Equation 2.10. $\mathcal{B}_N$, $\mathcal{B}_I$ and $\mathcal{B}_O$ are respectively the irradiation histories at nominal, inlet and outlet conditions. Branch calculations are executed only on these depletion histories for computing the derivatives and to implement the methodologies. $\mathcal{B}_f$ presents a case with high temperature in the fuel, a challenging condition for spectral history modeling [136].

Gray control rods made of stainless steel AISI316L with a density of 7.8481 g/cm^3 and suited for reactor control by prolonged insertion, are used in the first cycle for $\mathcal{B}_{c1}$ and in the second cycle for $\mathcal{B}_{c2}$. Their specification is presented in table C.2. They are characterized by a inner radius of 0.18840 cm and outer radius of 0.43130 cm.

An overview of the departure of the history parameters from the nominal case is presented in Figure C.1. The plotted relative values allow easily to determine the term

	Description	T_f [°C]	δ_M [g/cm^3]	CR-IP [GWd/t]
$\mathcal{B}_N$	Nominal condition	600	0.72	Out
$\mathcal{B}_I$	Inlet condition	600	0.75	Out
$\mathcal{B}_O$	Outlet condition	600	0.66	Out
$\mathcal{B}_f$	High fuel temperature	1200	0.72	Out
$\mathcal{B}_{c1}$	CR in the 1st cycle	600	0.72	In [0,15]
$\mathcal{B}_{c2}$	CR in the 2nd cycle	600	0.72	In [15,30]

Table C.1: Depletion histories considered to reproduce the history effects and to test the methodologies proposed in the literature. The insertion period of the control rods (CR-IP) is given in burnup units. T_f is the average fuel temperature in the pellets and δ_M is the density of the cooling water.

Isotope	Concentration [1/barn/cm]
^{nat}Fe	5.5366×E^{-02}
^{nat}Cr	1.5452×E^{-02}
^{nat}Ni	9.6629×E^{-03}
^{nat}Mo	1.2317×E^{-03}
^{55}Mn	1.7206×E^{-03}
^{nat}Si	1.6827×E^{-03}
^{nat}C	3.1505×E^{-04}

Table C.2: Material composition of the AISI316L control rods. The superscript "nat" denotes chemical elements with the natural isotopic abundances, see [145].

$(\theta/\theta_N - 1)$ of Equation C.6 and to show the variation among the history parameters for the different depletion histories. All off-nominal cases provide reference cross section values at distinct depletion conditions for benchmarking the physical accuracy.

The spectrum hardening, noticed at the outlet condition with a 15% increment of the SI, promotes the production of ^{239}Pu by an increasing capture rate of ^{238}U due to epithermal resonances. All cases show $\theta_{\mathcal{B}} > \theta_N$ but the inlet $\mathcal{B}_I$, which is the only depletion curve that exhibits a decrease in the Pu concentration caused by a softer spectrum. The monotonic behavior of the SI within the fuel cycles results from the hardening of the neutron spectrum with exposure since the fissile material in the fuel is depleted at a constant power level. When the movement of the control rod changes the configuration of the assembly, the SI shows discontinuity because of the high reactivity worth of the isotopes in the control rods. Instead the SH always responds in a continuous way due to its integral definition. After the control rod is withdrawn, the history parameters tend towards their nominal values.

C.4 Implementation

As defined in Equation C.6, off-nominal calculations are needed to use a new parameterization on θ and for the evaluation of the derivative in Equation C.7. The same

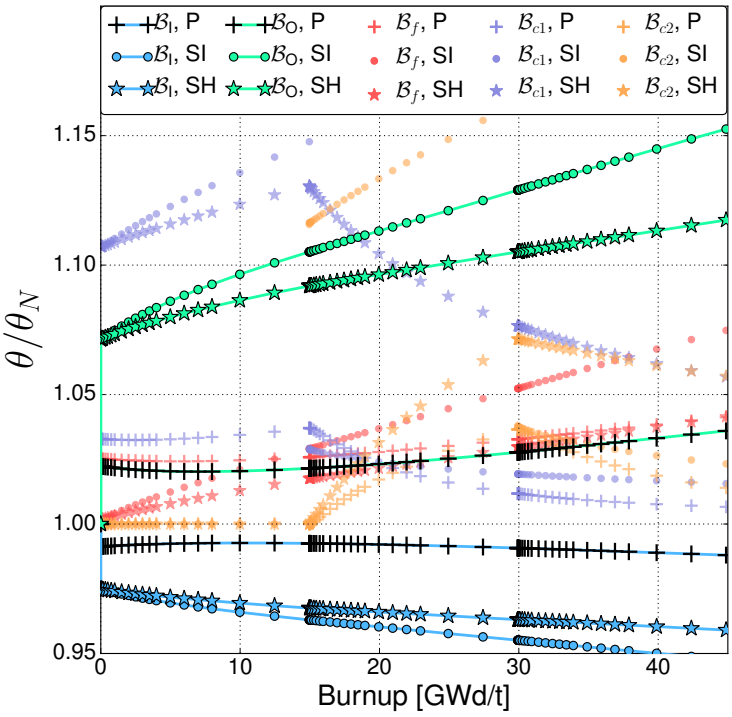


Figure C.1: Ratio of the history parameters P, SH, SI for the different use cases over the nominal case.

Equation C.6 is specifically introduced in this work to have a common basis to compare the performances of the different methodologies under reviewing and to offer a unique implementation. The possibilities to implement the methods considered in this work are presented in Table C.3 and tested with the use cases defined in Table C.1.

Method	History parameter (θ)	Off-nominal history (θ_{off})	Instantaneous parameter ($\vec{\iota}$)
P1	(C.4)	$\mathcal{B}_I$	$\mathcal{B}_I$
P2			$\mathcal{B}_N$
P3		$\mathcal{B}_O$	$\mathcal{B}_O$
P4			$\mathcal{B}_N$
PU	(C.5)		$\mathcal{B}_O$
SI	(C.3)		
SH	(C.2)		
MH	(C.1)		

Table C.3: Description of the history effect parameterization; Numbers in the second column refer to the equations used to define the given history parameter, whereas the depletion histories whose parameters belong to are indicated in the others.

Since the most of the cases show history effects similar to those caused by spectrum hardening at the core outlet, we choose $\mathcal{B}_O$ as the off-nominal depletion history to implement the different methodologies from Section C.2. All the combinations of θ_{off} and $\vec{\iota}$ are considered only with the parameter P. As well, in the PU method the sum is performed over all available plutonium isotopes from $\mathcal{I}$, see Equation 2.1. The methods P1, P2, P3, P4 are here indicated by P*, and when also considering PU they are referred to as "the Pu methods", while those based on SI and SH are called "the spectral methods".

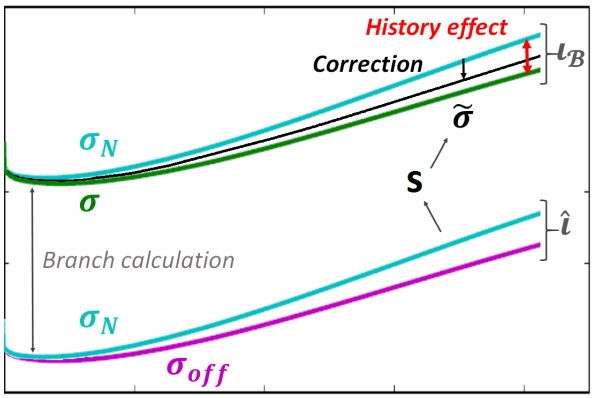


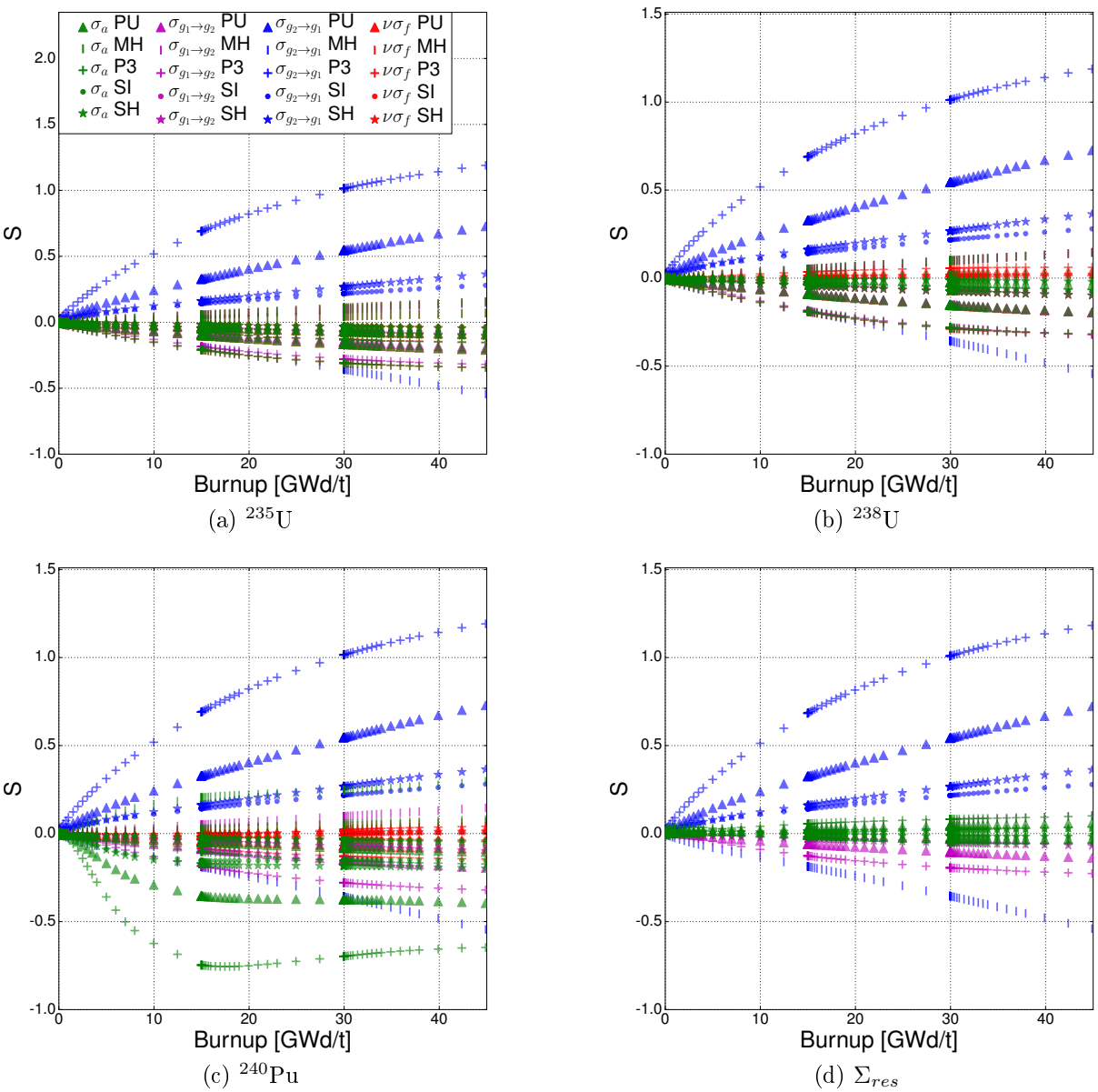
Figure C.2: Scheme of the history parameterization.

When performing statistical analysis only points where the history effect has taken place are considered so the first cycle of $\mathcal{B}_{c2}$ is excluded, and points with 0 exposure for all the cases. For uncorrected, P1, P2 and P4 all the use cases are considered. For P3, PU, SH, SI all cases except $\mathcal{B}_O$ are considered. For MH only $\mathcal{B}_I$ is considered.

Figure C.2 illustrates briefly how the corrections are computed. The coefficients $S = (\theta_N/\sigma_N)(\Delta\sigma/\Delta\theta)$ are determined by σ_N and σ_{off} evaluated at $\vec{\hat{\iota}}$ by means of branch calculations for different values of the fuel assembly burnup. They are then stored on data libraries for subsequent core calculations, where the cross section σ_N corrected by S will provide $\hat{\sigma}$, i.e. an approximation of the target value σ (in green). These coefficients become part of the process of the cross section preparation and they must be stored in the few-group cross section library.

The history coefficients of a few important isotopes are presented in Figure C.3 as an example. The importance of an isotope is here estimated simply as the fraction of macroscopic cross sections as $(C_i\sigma_{i,r,g})/\Sigma_{r,g}$ where C_i is the nuclide concentration of the isotope i. The isotopes shown in Figure C.3 are those that demonstrate higher importance for the most of the whole exposure length. Although other definitions focusing on the isotope importance for the neutron reactivity are available from perturbation theory, this choice allows for fast classifications when analyzing the results. The absolute values of the coefficients start at zero, since no history effect has taken place yet, to then smoothly increase with the burnup. In general if $\theta_{off} > \theta_N$, negative values can be seen as a harder spectrum lowers the overall chance of neutron capture in the thermal energy range, i.e. $\sigma_O < \sigma_N$ and contrarily $\sigma_I > \sigma_N$. If $\theta_{off} < \theta_N$, like with the method MH, the coefficients are usually positive. Neutron up-scattering increases with a harder spectrum whereas down-scattering decreases.

The weak dependence of the coefficients on $\vec{\hat{\iota}}$ is shown in Figure C.4 for the coefficient of $\sigma_{a,1}$ of the isotope ^{240}Pu, in fact negligible differences can be observed between P1 and P2 and between P3 and P4. This allows to avoid additional branch calculations on the off-nominal depletion calculation (here $\mathcal{B}_I$ or $\mathcal{B}_O$) to homogenize new cross sections at the nominal condition $\vec{\hat{\iota}}_N$. The observations about the ^{240}Pu are a general trend for all other isotopes. On the other hand, a significant difference of up to 25% can be noticed by changing θ_{off} from inlet to outlet, like from P1 to P3 for instance. This indicates possible non linearity and relevance of the higher order terms in the Taylor expansion,



(a) ^{235}U (b) ^{238}U

(c) ^{240}Pu (d) Σ_{res}

Figure C.3: History coefficients of the most important isotopes. σ_a and $\nu\sigma_f$ identify the values of both the fast and the thermal groups. The legend in the plot of ^{235}U is common to the other figures.

here disregarded to comply with the methods from the literature. Smaller absolute values for the SI and for the SH coefficients are observed with respect to the Pu methods.

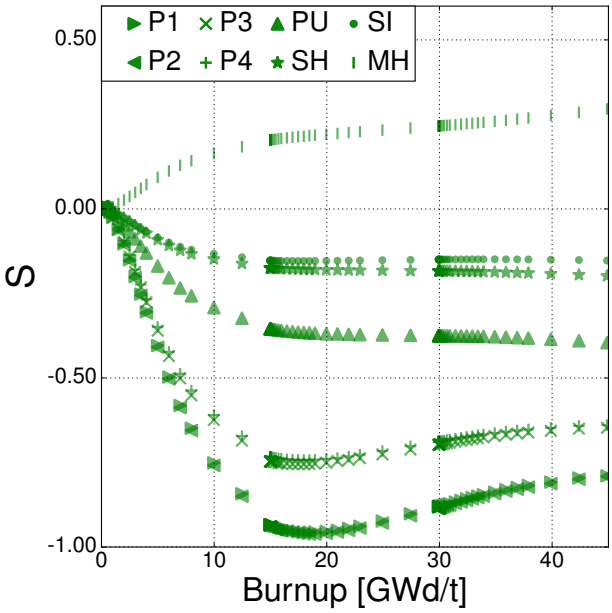


Figure C.4: History coefficients of $\sigma_{a,1}$ of ^{240}Pu for the history parameters.

C.5 Results

The performance of the methods to predict accurate microscopic and macroscopic cross sections (σ and Σ), still homogenized in the fuel assembly quarter and condensed in the two group energy mesh, is studied in this section. The multiplication factor and the fundamental flux calculated by the associated eigenvalue problem in the infinite homogeneous medium are also analyzed.

C.5.1 Analysis of the microscopic cross sections

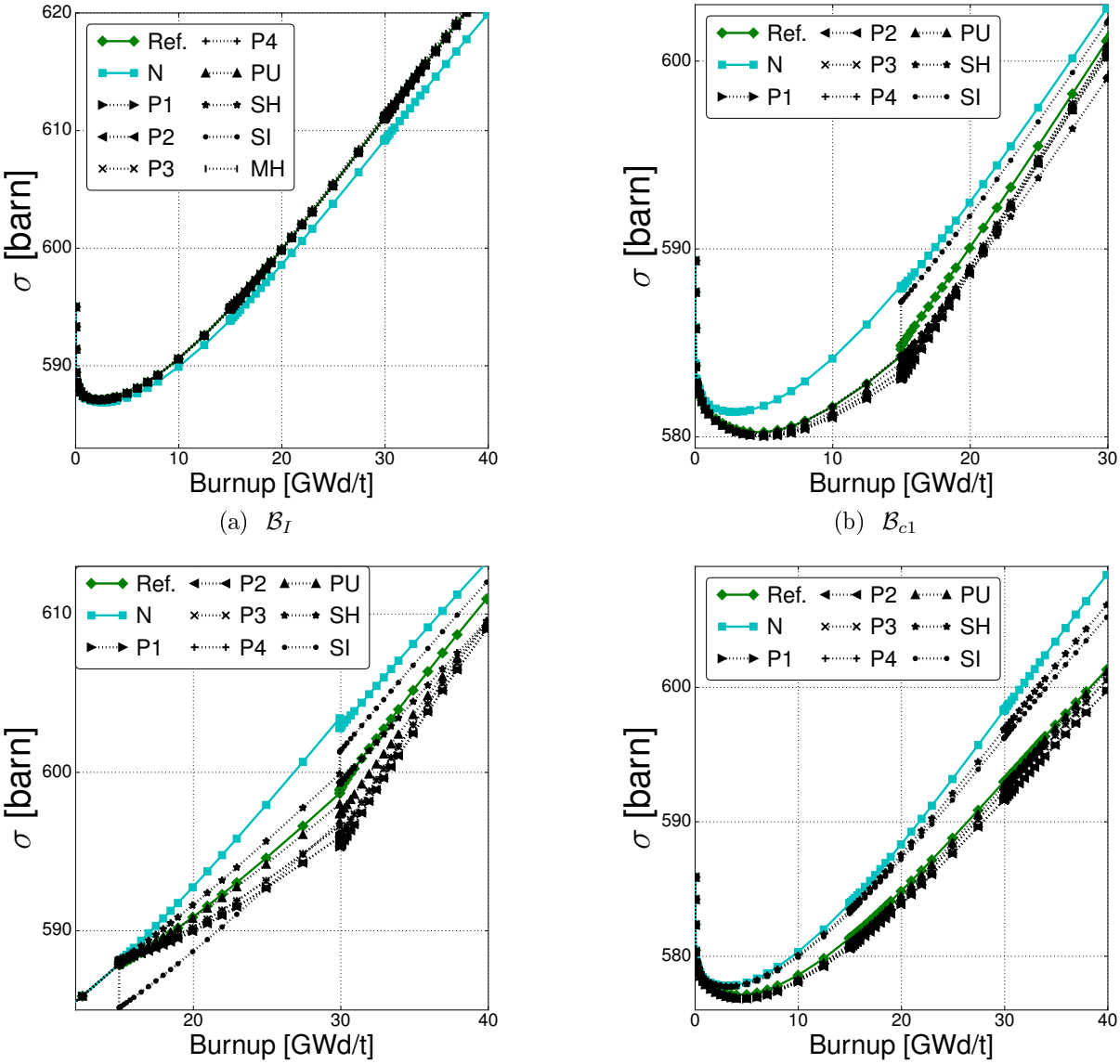
The relative error of the microscopic cross sections for each isotope, reaction and energy group is defined as:

$$\epsilon_{\sigma,\mathcal{B},\theta}(Bu) = \hat{\sigma}/\sigma_{\mathcal{B}} - 1. \tag{C.8}$$

$\hat{\sigma}$ comes from Equation C.6 while $\sigma_{\mathcal{B}}$ is calculated on the depletion history $\mathcal{B}$. Although both cross sections are evaluated over the curve $(Bu, \vec{\iota}_{\mathcal{B}}(Bu), \theta_{\mathcal{B}}(Bu))$ the isotopic inventories of the corresponding calculations reproducing them by homogenization differ. Branch calculations are performed at $\vec{\iota}_{\mathcal{B}}(Bu)$ thus avoiding the need of any interpolation to determine σ_N and the introduction of other source of error. Relative errors permit to account for every σ regardless of the variations in their absolute values.

Illustrative examples of the history effects are presented in Figure C.5 for the thermal fission production cross section of ^{235}U, $\nu\sigma_{f,2}$. For instance in $\mathcal{B}_{c1}$, a maximum error of 8 barn, about 1.35%, is achieved at the end of the first cycle for the uncorrected cross section (σ_N in light blue). This error is often reduced when considering the corrections of the history parameters, that is approaching the green curves of the reference target values. After the withdrawal of the control rod at 15 GWd/t in $\mathcal{B}_{c1}$ and at 30 GWd/t in $\mathcal{B}_{c2}$, the SI exhibits a prompt response underestimating the intended correction, while the

SH shows a delayed response typical of its integral character. The Pu method provides better correction here with this fissile isotope.



(a) $\mathcal{B}_I$ (b) $\mathcal{B}_{c1}$ (c) $\mathcal{B}_{c2}$ (d) $\mathcal{B}_f$

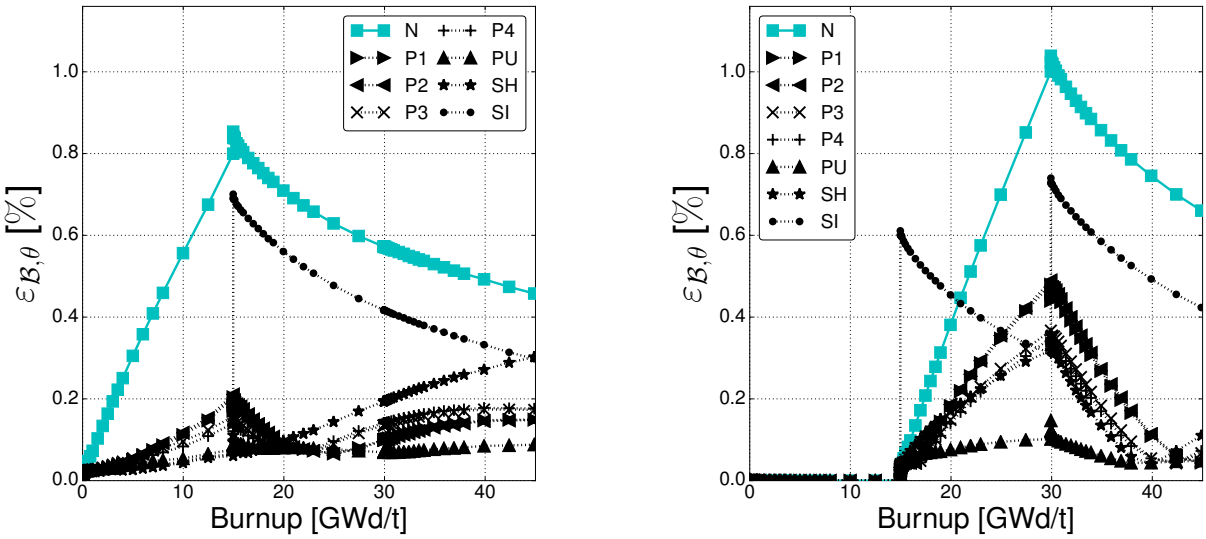
Figure C.5: Microscopic cross section $\nu\sigma_{f,2}$ of ^{235}U along the burnup with different use cases.

A simple arithmetic average of the absolute error is proposed in the following to assess a global overview of the behavior of the error:

$$\varepsilon_{\mathcal{B},\theta}(Bu) = \frac{\sum_{i=1}^{I}\sum_{r}\sum_{g}|\epsilon_{\sigma_{i,r,g}}|}{M}, \tag{C.9}$$

where I is inferred from Equation 2.1, and the sum on the reactions and energy groups is taken on the elements of the set $\{\nu\sigma_{f,1}, \nu\sigma_{f,2}, \sigma_{a,1}, \sigma_{a,2}, \sigma_{1\to2}, \sigma_{2\to1}\}$. M is the total number of terms in the sum. These averages provide a global trend, without any quantification of the error on the neutron reactivity, treated further on. They are available for all cases $\mathcal{B}$ and history parameters θ. The in-scattering cross sections are disregarded in this study,

because they can be removed from the neutron balance equation. Figures C.6 show the averages ε in the two cases $\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$.



(a) $\mathcal{B}_{c1}$ (b) $\mathcal{B}_{c2}$

Figure C.6: Burnup-dependent average errors of all the microscopic cross sections for all history parameters.

The uncorrected error shows a monotonic behavior in the different cycles, with a negative trend after rod withdrawal. Maximum values around 1% are detected at the end of the insertion periods. General improvement is observed with the history parameters but for the SI, which overshoots its correction. As already mentioned above, the presence of strong neutron absorbers in the control rods has a considerable influence on the SI, and consequently also on the difference ($\mathrm{SI}/\mathrm{SI}_N - 1$) in Equation C.6. In the first cycle of $\mathcal{B}_{c1}$, the module of its history coefficient varies smoothly with the burnup, see Figure C.3, and so it does accordingly the average error on all the microscopic cross sections. After the control rods are withdrawn, $\mathrm{SI}/\mathrm{SI}_N$ gets closer to 1 since the isotopic contents, produced under different depletion conditions, have a negligible impact on the neutron spectrum in the assembly and $\varepsilon_{\mathcal{B}_{c1},\mathrm{SI}}$ follows the profile of $\varepsilon_{\mathcal{B}_{c1},\mathrm{N}}$. At the beginning of the second cycle in $\mathcal{B}_{c2}$ the severe jump of the SI, together with non-zero history coefficients, leads to un-physical $\hat{\sigma}$, even if no history effect has occurred yet. $\mathcal{B}_{c2}$ seems slightly more challenging than $\mathcal{B}_{c1}$ for the higher errors noticed. Besides, hidden compensation effects could be quite relevant by considering also the greater coefficients S and ratios θ/θ_N computed in the second cycle range. The method PU shows the best results by reducing the error below 0.1%, outperforming the other P* methods.

Burnup-averaged errors offer quick comparisons to compare the performances of the methods in the use cases. They are defined hereafter within the generic interval (Bu_1, Bu_2) as:

$$\bar{\varepsilon}_{\mathcal{B},\theta} = \frac{1}{\Delta Bu} \int_{Bu_1}^{Bu_2} |\varepsilon_{\mathcal{B},\theta}(Bu)| dBu', \tag{C.10}$$

and presented in Figure C.7 on the complete exposure interval ([0,45] GWd/t) for $\mathcal{B}_I$, $\mathcal{B}_O$, $\mathcal{B}_f$, and separated per cycle for the other cases.

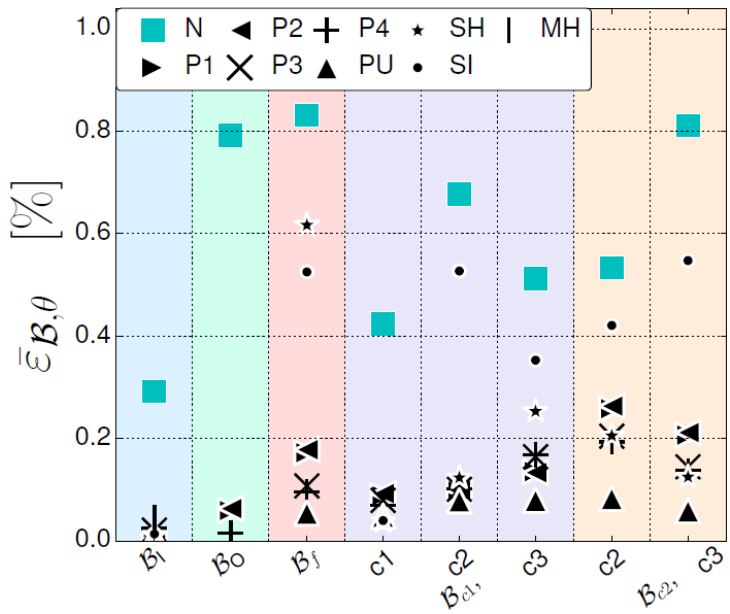


Figure C.7: Burnup-averaged error with all the history parameters and for the given use cases.

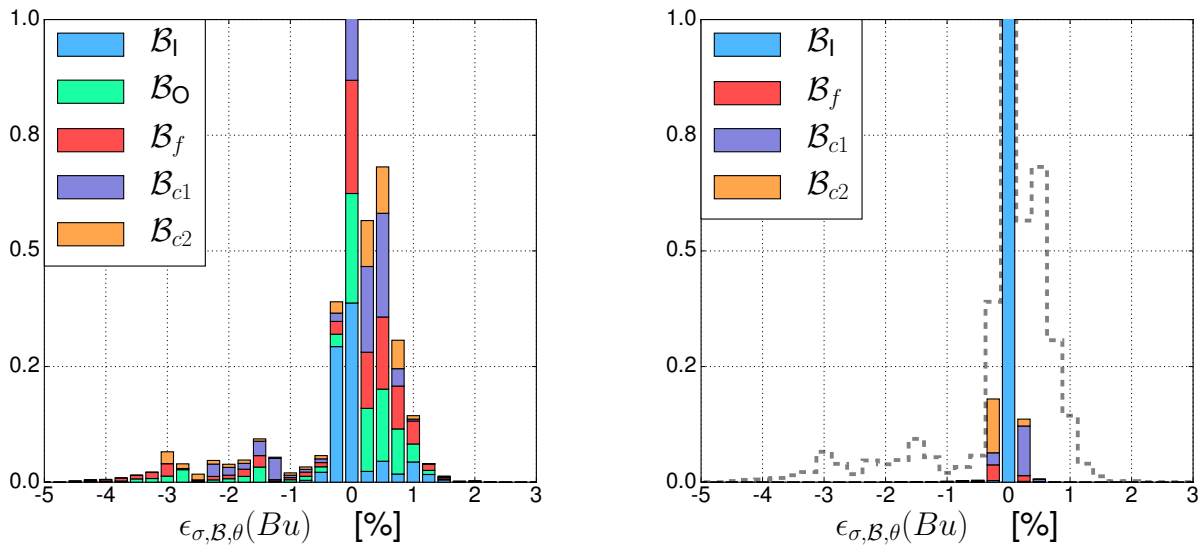
A reduction of one order of magnitude in the error is observed with all methods when testing against $\mathcal{B}_I$ and $\mathcal{B}_O$. This fact is expected indeed, because these cases are used as off-nominal calculations to compute the history coefficients. Of course, when a case is used for the implementation of the correction method, then it is automatically neglected for the tests.

Since inlet conditions are closer to $\mathcal{B}_N$, the uncorrected average error is smaller. No significant change with different off-nominal calculations or history parameters can be seen for both $\mathcal{B}_I$ and $\mathcal{B}_O$. The best performance is achieved with these cases as expected, since they were used to compute the history coefficients.

For $\mathcal{B}_f$ a reduction in the error can also be seen, though as noted in literature [136], the spectral methods exhibit poorer performances by reducing the error from 0.8% to about 0.6%. A possible explanation could be in view of the depletion condition, for a broadening of the resonances caused by the Doppler effect, that has a direct consequence on the ^{239}Pu build-up with exposure. As the spectral characteristics of the system are dominated by the moderator, that remains unaltered, spectral-derived history parameter may be inadequate for this case just as it was after the control rod removal, presented in Figure C.6a. For example in Figure C.1, it can be observed that the parameter P in $\mathcal{B}_f$ is higher than in $\mathcal{B}_O$, whilst the contrary is true for the SI indicating a possible failure of this methodology in reproducing the off-nominal condition. This applied to the SH too, which is derived from the SI.

About $\mathcal{B}_I$, $\mathcal{B}_O$ and $\mathcal{B}_f$ the depletion conditions are constant and the cross section error increases linearly with burnup. This means that $\bar{\varepsilon}_{\mathcal{B},\theta}$ is quite representative of the slope of $\varepsilon_{\mathcal{B},\theta}(Bu)$. As the spectral conditions of the depletion histories are closer to $\mathcal{B}_O$ than to $\mathcal{B}_I$, a marginal gain can indeed be seen with P3 or P4 in comparison to P1 or P2. On the other hand, considering different $\vec{\iota}$ (P1 versus P2 or P3 versus P4) does not improve further the error reduction.

Finally Equation C.8 is analyzed with the standard deviation (SD), the mean and the maximum error of all the cross sections by considering all the reference cases detailed in Section C.4. In Figure C.8, the error distribution for the microscopic cross sections is presented for σ_N and $\hat{\sigma}$ with PU and in Table C.4 for all the methods.



(a) Error distribution for σ_N; peak value of 2.1 (outside chart).

(b) Error distribution for $\hat{\sigma}$, with θ =PU. In dotted line the profile for σ_N; peak value of 7.4 (outside chart).

Figure C.8: Distribution of microscopic cross section error (bins of 0.25%).

	MEAN($\epsilon_{\sigma,\mathcal{B},\theta}(Bu)$)	SD	MAX($\lvert\epsilon_{\sigma,\mathcal{B},\theta}(Bu)\rvert$)
N	−0.06	0.8934	4.41
P1	<0.01	0.1984	2.98
P2	<0.01	0.2022	3.00
P3	<0.01	0.1635	2.31
P4	<0.01	0.1389	2.27
PU	−0.01	0.0878	1.79
SH	−0.03	0.4688	2.96
SI	−0.04	0.5755	2.40
MH	−0.01	0.0653	0.45

Table C.4: Standard deviation, mean and maximum error of $\epsilon_{\sigma,\mathcal{B},\theta}(Bu)$ in %.

A reduction of 90% in the SD is achieved with the PU method followed by P4, P3 with 83% and then P1 and P2 with 77%. The SH presents a reduction of 47% followed by 34%. A similar trend is found for the maximum error. Uncorrected cross sections do not present any bias, as well as the ones corrected with the history parameters.

C.5.2 Analysis of the macroscopic cross sections

The input of the nodal equations are macroscopic cross section that define with the flux the reaction rates dictating the state of the core and the fuel evolution. A similar error

definition to Equation C.8 is here considered:

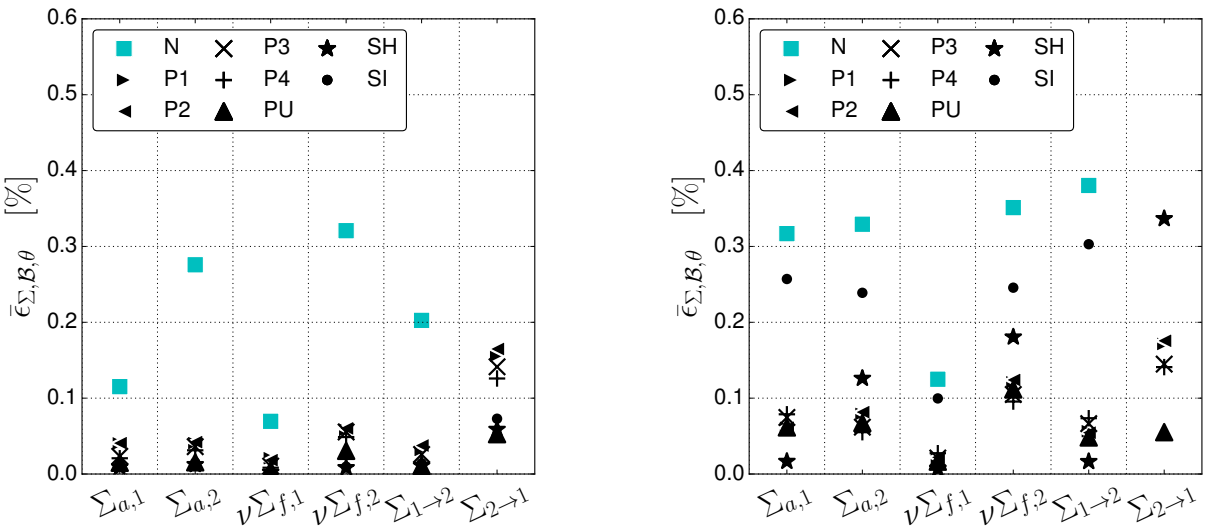
$$\epsilon_{\Sigma,\mathcal{B},\theta}(Bu) = \hat{\Sigma}/\Sigma_{\mathcal{B}} - 1. \tag{C.11}$$

Macroscopic cross sections $\hat{\Sigma}$ and $\Sigma_{\mathcal{B}}$ are computed respectively from $\hat{\sigma}$ and $\sigma_{\mathcal{B}}$ by Equation 2.10 using the reference concentrations from the case $\mathcal{B}$ for both of them. Consequently $\epsilon_{\Sigma,\mathcal{B},\theta}$ relates exclusively to spectrum induced errors, unlike other works where, due to the cross section representation model, nominal concentrations were used for $\hat{\Sigma}$, emphasizing more the history effect [146].

In Figure C.9, the burnup-averaged of the relative error in Equation C.11:

$$\bar{\epsilon}_{\Sigma,\mathcal{B},\theta} = \frac{1}{\Delta Bu}\int_{Bu_1}^{Bu_2} |\epsilon_{\Sigma,\mathcal{B},\theta}(Bu)| dBu', \tag{C.12}$$

is presented for $\mathcal{B}_{c1}$ (with integration per cycle).



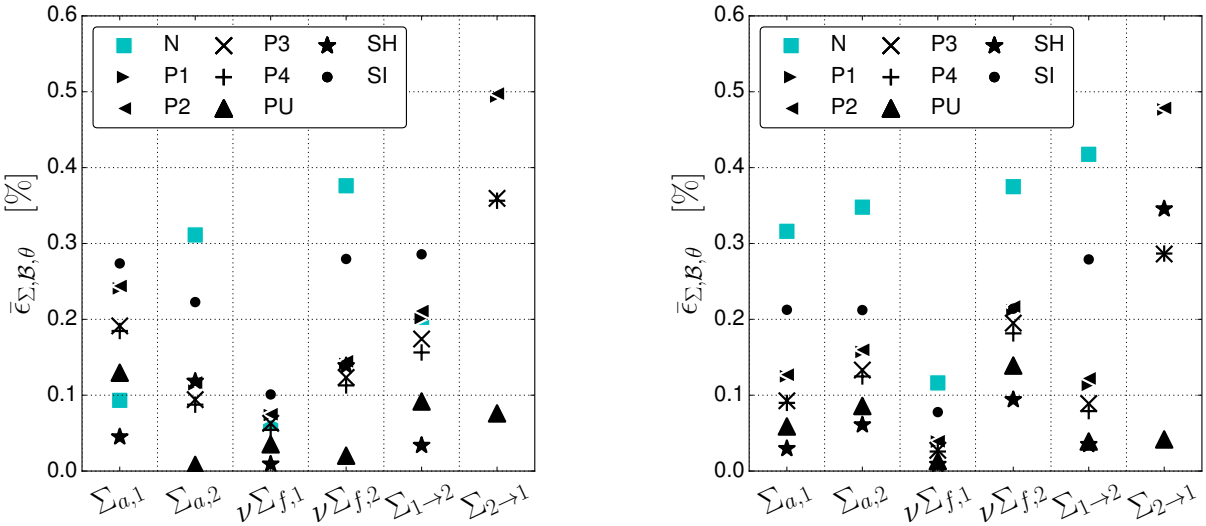
(a) First cycle. (b) Second cycle.

Figure C.9: $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ error for $\mathcal{B}_{c1}$.

A reduction in the error can be seen for every cross section, in particular for the up-scattering $\Sigma_{2\to1}$, where the uncorrected error of 1.21% (outside chart) is reduced to about 0.1%. The error of the absorption and of the fission cross section is roughly 0.05%, regardless of the size of the uncorrected error in the thermal group in the first cycle. In the second cycle instead, the history parameterization continues to reduce the error albeit to a lesser extent. As previously explained, the SI can not improve the results after the control rod withdrawal. In the third cycle of $\mathcal{B}_{c1}$ and $\mathcal{B}_f$, $\mathcal{B}_I$ and $\mathcal{B}_O$ similar reductions in $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ are observed.

The error $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ is presented in Figure C.10 for the second and for third cycle of $\mathcal{B}_{c2}$. Bad results are observed for the fast cross sections $\Sigma_{a,1}$, $\nu\Sigma_{f,1}$ and $\Sigma_{1\to2}$. This is addressed in Section C.5.2. The PU methods significantly outperforms the P* methods especially in these cases, as already noted in Figure C.5.1. The instantaneous behavior of the SI

explained for Figure C.6 compromises the correction capabilities in both the first and second cycle. Here again a marginal gain can be seen for P3 and P4 in comparison to P1 and P2.



(a) Second cycle (b) Third cycle

Figure C.10: $\bar{\epsilon}_{\Sigma,\mathcal{B},\theta}$ in $\mathcal{B}_{c2}$.

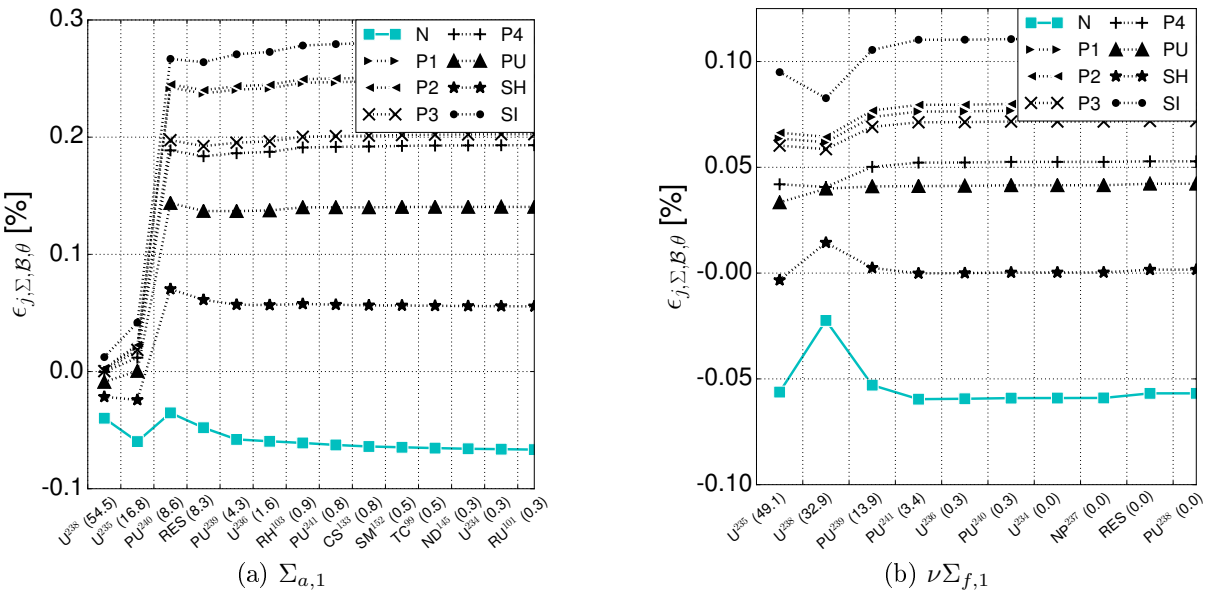
Investigation of the poor performances in $\mathcal{B}_{c2}$

The methodologies examined so far could not provide satisfactory corrections with the case $\mathcal{B}_{c2}$, which presents the all-rods-in configuration of the assembly all along the second cycle. Varying depletion conditions along the fuel irradiation were noted as very challenging for the modeling of the history effects [146]. In this section, we investigate the causes of the bad performances noticed with the tested parameterizations.

A cumulative error introduced hereafter is used to spot the major contribution to the missed corrections:

$$\epsilon_{j,\Sigma,\mathcal{B},\theta} = \frac{\sum_{i=1}^{I_j}(\sigma_{i,\mathcal{B}} - \hat{\sigma}_i)C_i}{\Sigma_{\mathcal{B}}}, \tag{C.13}$$

with the concentration C_i of the isotope i, $1 \leq i \leq I_j$ and $1 \leq j \leq I$. I_j corresponds to the set of j isotopes $\mathcal{I}_j \subseteq \mathcal{I}_I$, being for instance $\mathcal{I}_1 = \{^{238}\mathrm{U}\}$, $\mathcal{I}_2 = \{^{238}\mathrm{U}, ^{235}\mathrm{U}\}$, ..., until finding the full error on the macroscopic cross section $\Sigma_{\mathcal{B}}$ with $\mathcal{I}_I$. This criterion is affected by the compensation of errors with different sign coming from successive isotopes. Hence the j-th contribution is to be compared with the $(j-1)$-th to discriminate the major contributor. The type of cross section with the highest error is the fast neutron absorption, and specifically the one of ^{240}Pu overshoots the expected correction, see Figure C.11. This behavior is observed with all history parameters, as illustrated in Figure C.12, where the errors of this cross section averaged within the second cycle are divided by the original errors obtained by the uncorrected σ_N. Values smaller than 1 indicate an improvement with the applied correction, and bad prediction otherwise. Both figures sort the values by isotope according to their importance on the macroscopic cross sections (values indicated in parentheses).



(a) $\Sigma_{a,1}$ (b) $\nu\Sigma_{f,1}$

Figure C.11: Cumulative error of the macroscopic cross sections in the fast group at 22.5 GWd/t (half of the 2-nd cycle).

About the fission production cross section, the error is mainly determined by ^{235}U. We also noticed wrong predictions of the up-scattering cross section due to the contribution of the residual mixture, that is $\Sigma_{2\to1} \approx \Sigma_{2\to1,res}$. Thermal cross sections were correctly reproduced instead. The methods SH and PU provide the better results. In particular, the slow response in burnup of the SH limits the entity of the unwanted correction.

The control rod insertion in already irradiated fuel induces history effects that become similar to those noticed in $\mathcal{B}_I$ at the beginning of the cycle, see Figure C.14. $\mathcal{B}_I$ has certainly a more thermal neutron spectrum. The same trend is observed on the concentration and on $\sigma_{a,1}$ of ^{240}Pu. The corrections estimated by all methods rely on a supposed spectrum hardening in time, that does not occur at the assumed rate. The ^{240}Pu history coefficient is negative (see Figure C.4) and $\theta_{c2} > \theta_N$ (see Figure C.1), thus yielding a correction in the wrong direction for $\sigma_{a,1}$.

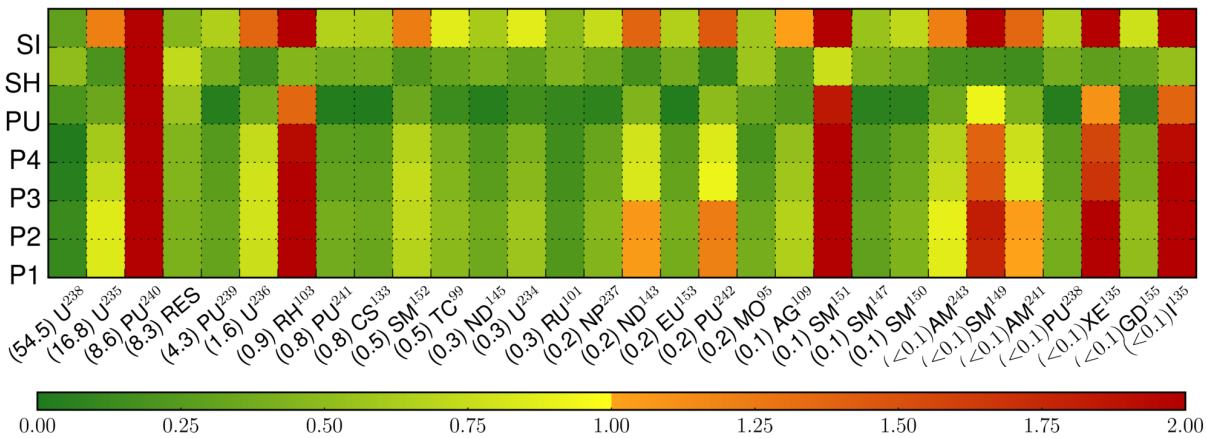
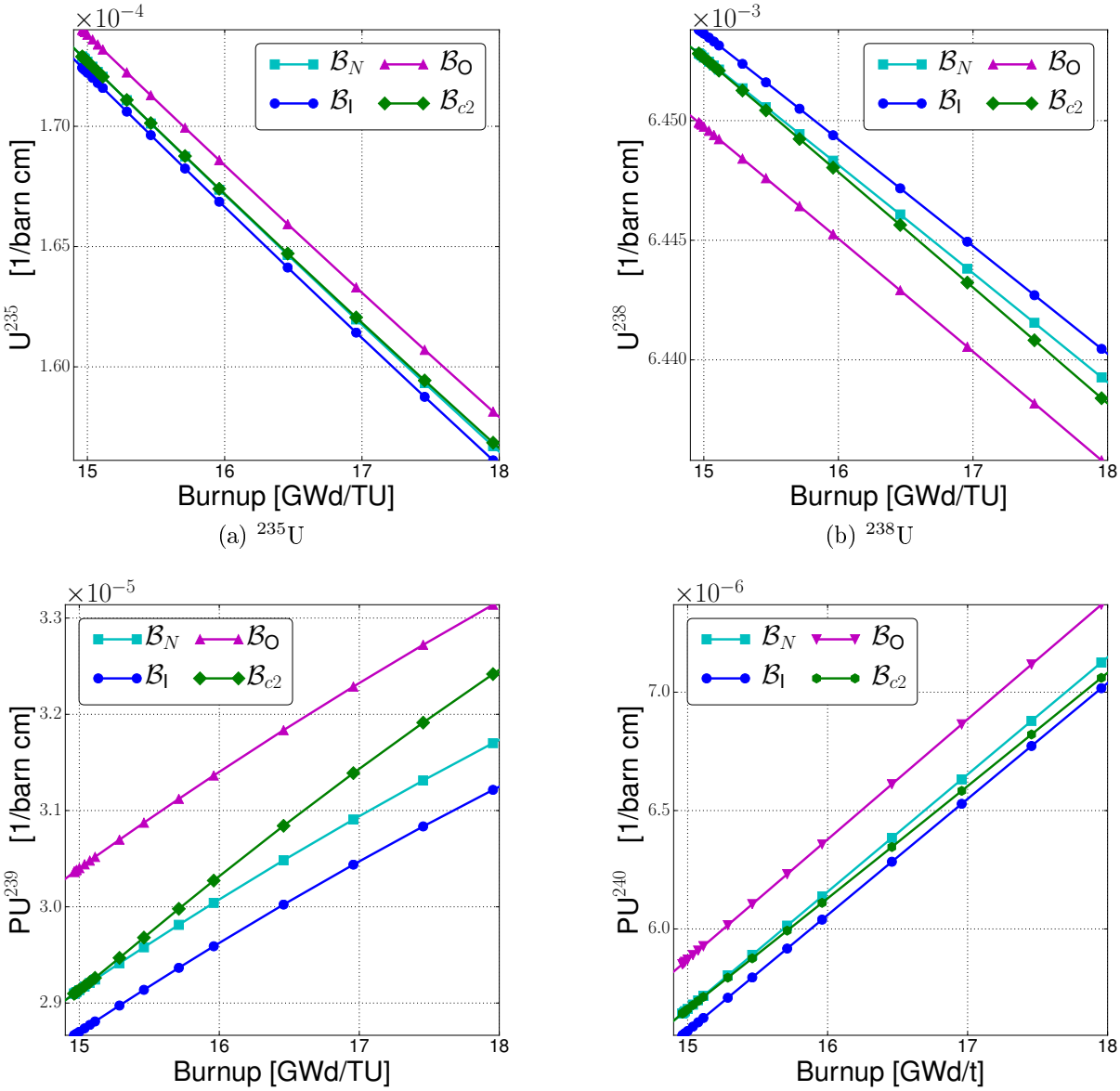


Figure C.12: Ratio of the corrected and uncorrected microscopic absorption cross sections in the fast group; the errors are burnup-averaged in the second cycle of $\mathcal{B}_{c2}$.

C.6 Concentrations of important isotopes

In Fig. C.13 the concentration of some important isotopes are presented for the reference case $\mathcal{B}_{c2}$ together with $\mathcal{B}_N$, $\mathcal{B}_I$ and $\mathcal{B}_O$. A non monotonic behavior can be seen in the concentration of ^{240}Pu that up to 18 GWd/t actually diminishes towards inlet concentration (with a hardening spectrum), only to start increasing after 24 GWd/t.



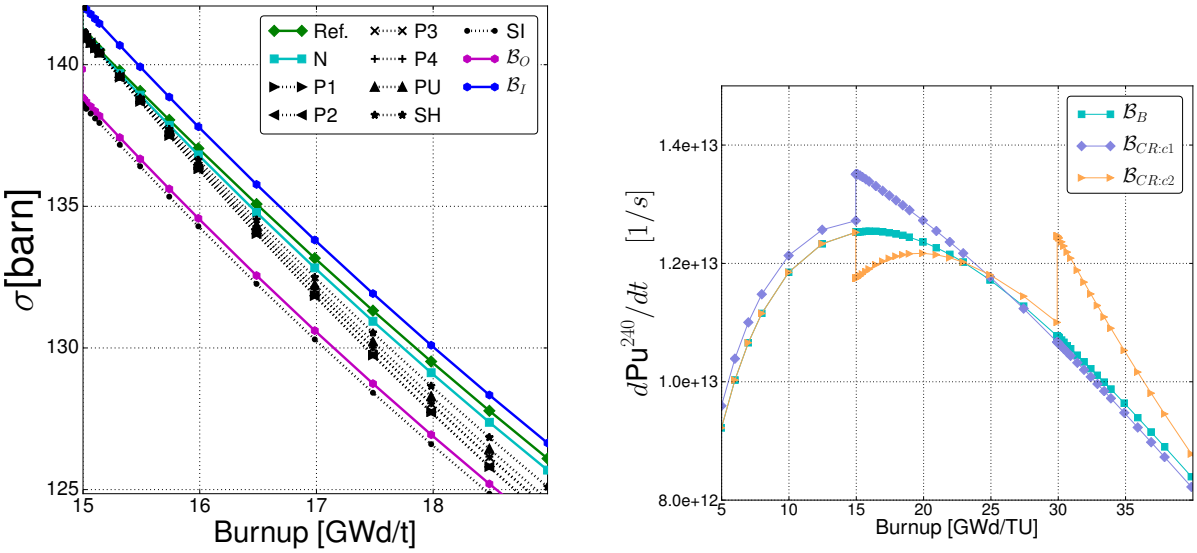
(a) ^{235}U (b) ^{238}U (c) ^{239}Pu (d) ^{240}Pu

Figure C.13: Concentration of important isotopes.

The concentration of ^{240}Pu results from the balance equation

$$\frac{d^{240}\mathrm{Pu}}{dt} = P - D, \tag{C.14}$$

with the production rate $P = (\sum_{g=1}^{2} \sigma_{^{239}\mathrm{Pu},g,c}\phi_g)C_{^{239}\mathrm{Pu}}$ and the destruction rate $D =$



(a) (b) Concentration of ^{240}Pu

Figure C.14: (a) $\sigma_{a,1}$ of ^{240}Pu in the 2-nd cycle of $\mathcal{B}_{c2}$. Concentrations of important isotopes for the reference cases $\mathcal{B}_{c2}$, $\mathcal{B}_N$, $\mathcal{B}_I$, $\mathcal{B}_O$. (b) Net production rate for ^{240}Pu when inserting control rods in $\mathcal{B}_N$,$\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$.

$(\sum_{g=1}^{2} \sigma_{^{240}\mathrm{Pu},g,a}\phi_g)\ C_{^{240}\mathrm{Pu}}$. These are plotted in Figure C.14b for the cases $\mathcal{B}_N$, $\mathcal{B}_{c1}$, $\mathcal{B}_{c2}$ were indeed a reduction in the overall production rate can be seen when inserting the control rods for $\mathcal{B}_{c2}$. This is contrary to what is observed for $\mathcal{B}_{c1}$ in the first cycle. For both $\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$ the production term P starts to increase when the spectrum hardens due to the increase in the ^{239}Pu concentration. However, the destruction term D instantaneously grows as well when the control rod is inserted in the second cycle of $\mathcal{B}_{c2}$. This is not the case for $\mathcal{B}_{c1}$ as no ^{240}Pu is yet present at BOC for the UO_2 fuel.

A single history parameter may not be enough to reproduce the history of $\mathcal{B}_{c2}$. More independent variables may be necessary for the intended corrections. Also, in [142] it's suggested that, a big module of the coefficient, like for ^{240}Pu, is a marker of high sensitivity of the cross section to the changes in the concentration of the isotope, and this could demand for the use of non-linear functions.

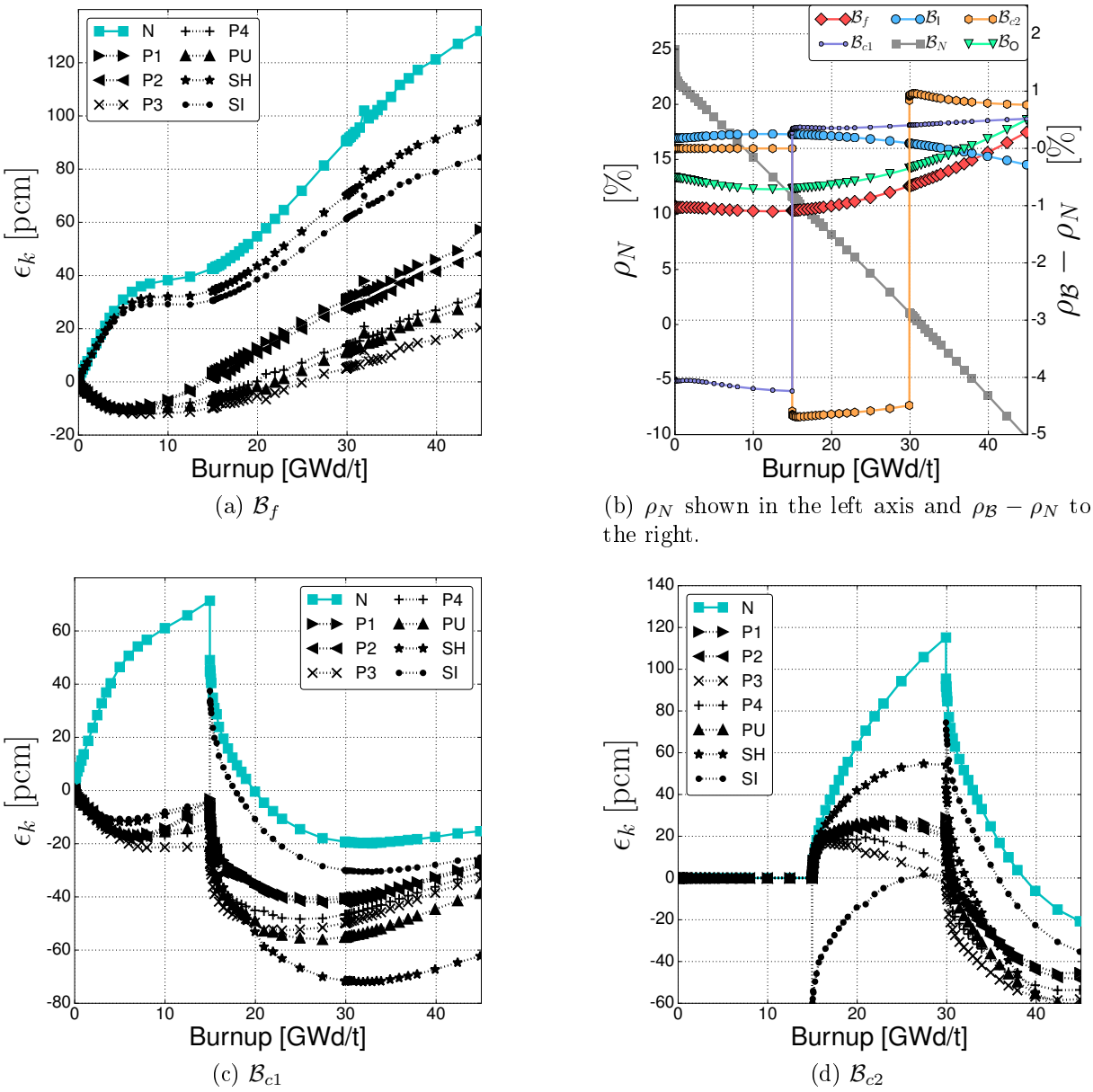
C.6.1 Analysis of the infinite multiplication factor

Hereafter, we examine the performances of the different history models by comparing the eigen-pairs of the eigenvalue problems obtained with the corrected cross sections and without any leakage model. The reference values are produced by the cross sections calculated by the lattice code on the curves $\mathcal{B}$. The same nuclide concentrations coming from these curves are used to build the macroscopic cross sections with the corrected microscopic cross sections. The relative errors of the two integral parameters are computed as:

$$\epsilon_{\mathrm{SI},\mathcal{B},\theta}(Bu) = \hat{\mathrm{SI}}/\mathrm{SI}_{\mathcal{B}} - 1, \tag{C.15a}$$

$$\epsilon_{k_\infty,\mathcal{B},\theta}(Bu) = \hat{k}/k_{\infty,\mathcal{B}} - 1. \tag{C.15b}$$

The Figures C.15 plots the evolution of ϵ_{k_∞} in $\mathcal{B}_f$, $\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$, together with the neutron reactivity along the burnup. About the cases $\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$, the corrections achieve a good reduction of ϵ_{k_∞} in the insertion periods. But they can not reproduce correctly the assembly reactivity after the insertion, suggesting a change in the correction methodology at control rod withdrawal by using the SI. The trend observed in $\mathcal{B}_I$ and $\mathcal{B}_O$ is very similar to $\mathcal{B}_f$. Again, the spectral methods are less effective with the applied corrections.



(a) $\mathcal{B}_f$

(b) ρ_N shown in the left axis and $\rho_{\mathcal{B}} - \rho_N$ to the right.

(c) $\mathcal{B}_{c1}$

(d) $\mathcal{B}_{c2}$

Figure C.15: ϵ_{k_∞} with the different correction methods, with the neutron reactivity of all use cases (top left).

We remind that the possible improvements observed in the macroscopic cross sections may bring higher error in the fundamental eigen-pair at the end because of systematic error compensation. However, the differences in the original uncorrected reactivities are

rather small compared to those arising in the core calculations, due to the uncertainty related to the physical core state and to the nuclide concentrations.

The standard deviation, the mean and the maximum absolute error of ϵ_{k_∞} are presented in Table C.5 for all the history parameters. The statistics take into account the values from all the available cases from Section C.4. The Pu methods manage on average to get better results. The method MH is only tested against $\mathcal{B}_I$, for lack of cases with a different moderator density during depletion; at least in this test case, it performs reasonably well. The distribution of these errors are also shown in the Figures C.16, where only the method PU is reported, being the most successful. The components from the separate cases appear with different colors in the columns of the histograms.

	ϵ_k			ϵ_{SI}		
	MEAN	SD	MAX	MEAN	SD	MAX
N	27	39	132	-0.017	0.062	0.217
P1	2	20	57	-0.014	0.048	0.138
P2	-0	20	48	-0.010	0.048	0.151
P3	-10	21	59	-0.027	0.052	0.129
P4	-5	19	54	-0.028	0.049	0.134
PU	-7	23	61	-0.028	0.048	0.134
SH	6	40	98	-0.025	0.050	0.153
SI	10	31	84	-0.026	0.053	0.141
MH	8	6	20	0.002	0.001	0.006

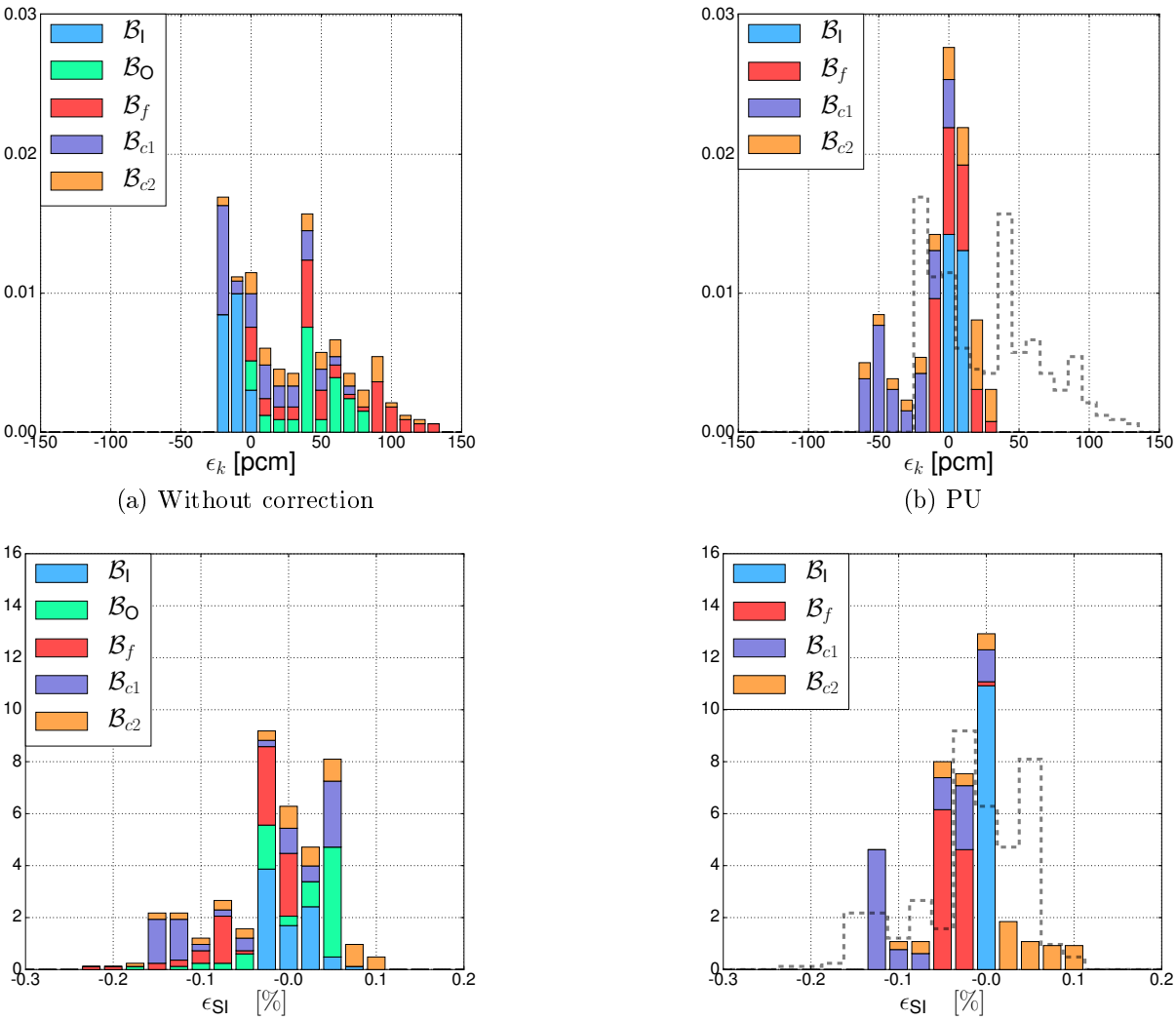
Table C.5: Standard deviation (SD), mean and maximum absolute error (MAX) of ϵ_{k_∞} in pcm and of ϵ_{SI} in %.

The reduction in the standard deviation is strongly driven by the reference cases $\mathcal{B}_I$ and $\mathcal{B}_f$. The missed corrections after the rods withdrawal in $\mathcal{B}_{c1}$ and $\mathcal{B}_{c2}$ are the main cause for the high tail in Figure C.6.1, with errors of about -50 pcm in $\hat{k}_\infty$.

C.7 Conclusion

This work discussed a few relevant history effects in PWR analysis caused by the approximations of the standard cross section preparation. These effects arise in core calculations whenever the fuel assemblies are burning for prolonged periods in time at different exposure conditions other than the ones used by the lattice code in burnup calculations. In fact, the traditional cross section preparation is generally reproducing the behavior of the fuel at normal operation on base power load, allowing only short variations in time of the core state. Recent design of PWR units and boron-free SMRs largely employs protracted mechanical shim to control and operate the reactor, thus incurring in possible issues with the modeling of the history effects.

Indeed, this constitutes a well-known problem since the beginning of the development of light water reactor technology, thus motivating the literature review which is at the base of this article. This review suggests possible resolutions by increasing the number of



(a) Without correction (b) PU

(c) Without correction (d) PU

Figure C.16: Histograms of ϵ_{k_∞} and ϵ_{SI} (uncorrected profile in dotted line); bins of 10 pcm and 0.025% respectively.

specialized isotopes in the reduced depletion chains, and the introduction of various models to correct the homogenized cross sections directly online during the core calculation. These methodologies are here reviewed and tested with a set of depletion histories of topical interest on a typical UO_2 17×17 fuel assembly, whose data specifications follow from the "Burn-up Credit Criticality Benchmark".

Although some methods came originally with empirical formulae, we propose here a common expression justified by a first order Taylor expansion which offer also a common background for their implementation and discussion. A comparison of the Taylor coefficients for the first derivative approximated at the core inlet and outlet seems indicating possible non-linear behaviors, likely recoverable by higher order terms in the expansion. On the other hand, a weak dependence on the instantaneous parameters of the history coefficients is noted, thence allowing to avoid additional branch calculations on the off-nominal burnup calculations. This interesting outcome limits the implementation effort in the existing schemes of cross section preparation based on lookup tables.

We used a representative list of isotopes in our reduced depletion chain to cover the majority of core applications with the current standards in industry. The list is fine enough to yield small errors in the neutron reactivity and in the spectral index of the fundamental flux calculated with the two group cross sections, which are spatially homogenized in the fuel assembly quarter. The reference nuclide concentrations were used in all tests in order to restrain the source of error to the only microscopic cross sections. Furthermore, all cross sections requested at the given instantaneous state parameters were produced by lattice calculations, without making any approximation by the common data interpolation. The errors noticed in our tests must then be considered as a lower bound for the real error in core calculations, due to the additional differences arising in the nuclide inventory along exposure.

Of course, the use of history parameters and of the number of isotopes in the simplified depletion chains must comply with the computational constraints in time of reactor calculations. More state parameters or specialized nuclides imply a longer cross section preparation by lattice calculations and larger reactor data-libraries, as well as more expensive data interpolation for the multi-physics core calculations.

About the microscopic cross sections, a general reduction of the error is observed with all the history parameters, but in the use case $\mathcal{B}_{c2}$. In this case all methods failed with the fast absorption cross section of ^{240}Pu. This outcome is not noticed in $\mathcal{B}_{c1}$ suggesting that a dedicated off-nominal calculation should be used to reproduce the physics of plutonium during protracted insertion periods in spent fuel with relevant amount of ^{240}Pu.

About the history parameterization, the Pu methods performed better than the spectral methods, with the PU method showing the best results in all test cases. In particular, the spectral methods did not provide any benefit with the case $\mathcal{B}_f$. The SH avoids however the sudden changes in value characterizing the SI and canceling the potential corrections. In conclusion our recommendation is to use the parameter PU.

Titre: Amélioration du modèle reconstruction des sections efficaces dans un code de calcul de neutronique à l'echelle cœur.

Mots clés: Calcul cœur, Sections efficaces homogénéisées à peu de groupes, Apprentissage machine, Méthodes à noyaux (RKHS), Réseau de neurones, Apprentissage actif, B-splines, Effets d'historique

Résumé: Pour estimer la répartition de la puissance au sein d'un réacteur nucléaire, il est nécessaire de coupler des modélisations neutroniques et thermohydrauliques. De telles simulations doivent disposer des valeurs sections efficaces homogénéisées à peu de groupes d'énergies qui décrivent les interactions entre les neutrons et la matière. Cette thèse est consacrée à la modélisation des sections efficaces par des techniques académiques innovantes basées sur l'apprentissage machine. Les premières méthodes utilisent les modèles à noyaux du type RKHS (Reproducing Kernel Hilbert Space) et les secondes par réseaux de neurones.

La performance d'un modèle est principalement définie par le nombre de coefficients qui le caractérisent (c'est-à-dire l'espace mémoire nécessaire pour le stocker), la vitesse d'évaluation, la précision, la robustesse au bruit numérique, la complexité, etc. Dans cette thèse, un assemblage standard de combustible UOX REP est analysé avec trois variables d'état : le burnup, la température du combustible et la concentration en bore. La taille de stockage des bibliothèques est optimisée en cherchant à maximiser la vitesse et la précision de l'évaluation, tout en cherchant à réduire l'erreur de reconstruction des sections efficaces microscopiques, macroscopiques et du facteur de multiplication infini. Trois techniques d'approximation sont étudiées.

Les méthodes de noyaux, qui utilisent le cadre général d'apprentissage machine, sont capables de proposer, dans un espace vectoriel normalisé, une grande variété de modèles de régression ou de classification. Les méthodes à noyaux peuvent reproduire différents espaces de fonctions en utilisant un support non structuré, qui est optimisé avec des techniques d'apprentissage actif. Les approximations sont trouvées grâce à un processus d'optimisation convexe facilité par "l'astuce du noyau". Le caractère modulaire intrinsèque de la méthode facilite la séparation des phases de modélisation : sélection de l'espace de fonctions, application de routines numériques, et optimisation du support par apprentissage actif.

Les réseaux de neurones sont des méthodes d'approximation universelles capables d'approcher de façon arbitraire des fonctions continues sans formuler de relations explicites entre les variables. Une fois formés avec des paramètres d'apprentissage adéquats, les réseaux à sorties multiples (intrinsèquement parallélisables) réduisent au minimum les besoins de stockage tout en offrant une vitesse d'évaluation élevée.

Les stratégies que nous proposons sont comparées entre elles et à l'interpolation multilinéaire sur une grille cartésienne qui est la méthode utilisée usuellement dans l'industrie. L'ensemble des données, des outils, et des scripts développés sont disponibles librement sous licence MIT.

Université Paris-Saclay, Espace Technologique/Immeuble Discovery, Route de l'Orme aux Merisiers RD 128/91190 Saint-Aubin, France

Title: Few-group cross section modeling by machine learning for nuclear reactors

Keywords: Core Calculations, neutron transport, Homogenized Few-Group Cross Sections, Machine Learning, Kernel Methods, Artificial Neural Networks, Pool Active Learning, B-splines, History Effects

Abstract: Modern nuclear reactors utilize core calculations that implement a thermo-hydraulic feedback requiring accurate homogenized few-group cross sections. They describe the interactions of neutrons with matter, and are endowed with the properties of smoothness and regularity, steaming from their underling physical phenomena. This thesis is devoted to the modeling of these functions by industry state-of-the-art and innovative machine learning techniques. Mathematically, the subject can be defined as the analysis of convenient mapping techniques from one multi-dimensional space to another, conceptualize as the aggregated sum of these functions, whose quantity and domain depends on the simulations objectives. Convenient is intended in terms of computational performance, such as the model's size, evaluation speed, accuracy, robustness to numerical noise, complexity, etc; always with respect to the engineering modeling objectives that specify the multi-dimensional spaces of interest. In this thesis, a standard UO_2 PWR fuel assembly is analyzed for three state-variables, burnup, fuel temperature, and boron concentration. Library storage requirements are optimized meeting the evaluation speed and accuracy targets in view of microscopic, macroscopic cross sections and the infinite multiplication factor. Three approximation techniques are studied:

The state-of-the-art spline interpolation using computationally convenient B-spline basis, that generate high order local approximations. A full grid is used as usually done in the industry.

Kernel methods, that are a very general machine learning framework able to pose in a normed vector space, a large variety of regression or classification problems. Kernel functions can reproduce different function spaces using an unstructured support, which is optimized with pool active learning techniques. The approximations are found through a convex optimization process simplified by the kernel trick. The intrinsic modular character of the method facilitates segregating the modeling phases: function space selection, application of numerical routines and support optimization through active learning.

Artificial neural networks which are "model free" universal approximators able to approach continuous functions to an arbitrary degree without formulating explicit relations among the variables. With adequate training settings, intrinsically parallelizable multi-output networks minimize storage requirements offering the highest evaluation speed.

These strategies are compared to each other and to multi-linear interpolation in a Cartesian grid, the industry standard in core calculations. The data set, the developed tools, and scripts are freely available under a MIT license.

Université Paris-Saclay, Espace Technologique/Immeuble Discovery, Route de l'Orme aux Merisiers RD 128/91190 Saint-Aubin, France