



**TECHNISCHE
UNIVERSITÄT
WIEN**
Vienna University of Technology

DISSERTATION

Large Scale Bayesian Nuclear Data Evaluation with Consistent Model Defects

Ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften
unter der Leitung von

Ao. Univ. Prof. Helmut Leeb

Inst. Nr.: E141

Atominstitut

eingereicht an der Technischen Universität Wien
Fakultät für Physik

von

Dipl.-Ing. Georg Schnabel

Matrikelnummer: e0525028

Weißes Kreuz-Gasse 57, A-2340 Mödling

Wien, Juni 2015

Kurzfassung

Das Ziel von Kerndatenevaluationen ist die zuverlässige Bestimmung von Wirkungsquerschnitten und verwandten Observablen der Atomkerne. Die Bestimmung dieser Observablen geschieht mittels einer Evaluationsmethode, in der die Information aus Experimenten mit den Ergebnissen aus Modellrechnungen kombiniert wird. Die auf diese Weise erhaltenen Schätzwerte mitsamt den zugehörigen Unsicherheiten und Korrelationen werden zu Datensätzen zusammengefasst, welche für die Entwicklung von zukünftigen nuklearen Einrichtungen, wie Fusionsreaktoren zur Energiegewinnung und Beschleuniger getriebenen Systeme zur Verbrennung nuklearen Mülls, benötigt werden. Die Effizienz und Sicherheit solcher Einrichtungen wird in hohem Maße von der Qualität der Datensätze und folglich auch von der Zuverlässigkeit der eingesetzten Evaluationsmethoden abhängen.

In dieser Arbeit wurde ein Großteil der existierenden Evaluationsmethoden in zwei Szenarien studiert. Dabei hat sich herausgestellt, dass in nahezu allen Methoden eine wesentliche Annahme fehlt. Üblicherweise basieren Nuklearmodelle auf Näherungen, weshalb die Vorhersagen dieser Modelle schon alleine aus diesem Grund von zuverlässigen experimentellen Daten abweichen können. Dennoch bleibt diese Unzulänglichkeit der Modelle in den Evaluationsmethoden unberücksichtigt. Dieser Umstand kann, wie in dieser Arbeit demonstriert wurde, zu evaluierten Schätzwerten und Unsicherheiten führen, die nicht mit den experimentellen Daten verträglich sind.

Deshalb wurde im Rahmen dieser Arbeit eine Erweiterung Bayesscher Evaluationsmethoden vorgeschlagen um einen eventuellen Modellfehler zu berücksichtigen. Diese Erweiterung basiert auf der Modellierung des Modellfehlers durch einen Gauß-Prozess. In dieser neuen Formulierung werden die Summenregeln zwischen Reaktionskanälen erhalten und weiters ist eine explizite Abschätzung des Modellfehlers möglich. Die letztgenannten Besonderheiten der neuen Formulierung sind bisher in keiner anderen Evaluationsmethode vorhanden.

Außerdem wurden zwei Verbesserungen für existierende Evaluationsmethoden erarbeitet. Die eine Verbesserung betrifft Methoden, die Ergebnisse mittels Monte-Carlo-Simulation erzielen. Es wurde ein Metropolis-Hastings-Algorithmus mit einer speziellen Vorschlagsverteilung entwickelt, der sich in den studierten Szenarien als effizienter als bestehende Monte-Carlo-Simulations-Techniken erwiesen hat.

Die zweite Verbesserung betrifft Evaluationsmethoden, die das Nuklearmodell auf bestimmte Weise vereinfachen. Bisher war die Anwendung dieser Methoden auf zehntausende Observablen beschränkt. Eine neue Formulierung ermöglicht die Auswertung von zehn Millionen und mehr Observablen und kann problemlos um die Berücksichtigung von Modellfehlern erweitert werden.

Der neue Ansatz ist bestens für die Implementierung als Datenbankanwendung geeignet. Die Realisierung einer solchen Anwendung und ein öffentlicher Zugang versprechen die beschleunigte Erzeugung von zuverlässigen Datensätzen. Weiters erlaubt die um Modellfehler erweiterte Formulierung die automatische Erkennung von Unzulänglichkeiten des Modells und der experimentellen Daten. Folglich würde auch die Entwicklung von Nuklearmodellen mit hoher Vorhersagekraft von einer solchen Datenbankanwendung profitieren.

Abstract

The aim of nuclear data evaluation is the reliable determination of cross sections and related quantities of the atomic nuclei. To this end, evaluation methods are applied which combine the information of experiments with the results of model calculations. The evaluated observables with their associated uncertainties and correlations are assembled into data sets, which are required for the development of novel nuclear facilities, such as fusion reactors for energy supply, and accelerator driven systems for nuclear waste incineration. The efficiency and safety of such future facilities is dependent on the quality of these data sets and thus also on the reliability of the applied evaluation methods.

This work investigated the performance of the majority of available evaluation methods in two scenarios. The study indicated the importance of an essential component in these methods, which is the frequently ignored deficiency of nuclear models. Usually, nuclear models are based on approximations and thus their predictions may deviate from reliable experimental data. As demonstrated in this thesis, the neglect of this possibility in evaluation methods can lead to estimates of observables which are inconsistent with experimental data.

Due to this finding, an extension of Bayesian evaluation methods is proposed to take into account the deficiency of the nuclear models. The deficiency is modeled as a random function in terms of a Gaussian process and combined with the model prediction. This novel formulation conserves sum rules and allows to explicitly estimate the magnitude of model deficiency. Both features are missing in available evaluation methods so far.

Furthermore, two improvements of existing methods have been developed in the course of this thesis. The first improvement concerns methods relying on Monte Carlo sampling. A Metropolis-Hastings scheme with a specific proposal distribution is suggested, which proved to be more efficient in the studied scenarios than the Monte Carlo sampling schemes of available evaluation methods.

The second improvement concerns Bayesian evaluation methods based on a certain simplification of the nuclear model. These methods were restricted to the consistent evaluation of tens of thousands of observables. In this thesis, a new evaluation scheme has been developed, which is mathematically equivalent to existing methods, but allows the consistent evaluation of dozens of millions of observables.

The new scheme is suited for the implementation as a database application. The realization of such an application with public access can help to accelerate the production of reliable nuclear data sets. Furthermore, in combination with the novel treatment of model deficiencies, problems of the model and the experimental data can be tracked down without user interaction. This feature can foster the development of nuclear models with high predictive power.

Acknowledgments

First of all, I want to thank my doctoral adviser Helmut Leeb who employed me from spring 2013 and gave me the opportunity to work in the exciting field of nuclear data evaluation. In the initial phase of my thesis, I benefited from joint discussions, and later on, from his encouragement to pursue my own ideas. He granted me a lot of freedom in my working style, but also confronted me with tough deadlines. I think this management style was perfectly suited for me.

I also want to express my gratitude to Arjan Koning who promptly answered all my emails with questions about his TALYS code. Furthermore, without the existence of the TALYS code and its clear and comprehensive documentation, this thesis would not have been possible. Therefore—and I am sure that I speak not only for myself—I thank him again for his great efforts developing such a versatile tool for nuclear data evaluation.

The progress in the subject of my thesis profited considerably from the work on several projects funded by different institutions. I acknowledge the funding from *EURATOM* for the CHANDA project 6050203, from the *European Union* for the ENSAR project 262010, from the *Kommission für die Koordination der Kernfusionsforschung in Österreich* for the project IPN-2013-7, and the reception of the F4E-FPA-168 specific grant 1 from *Fusion For Energy*. Although the results obtained within these projects helped in the development of this thesis, the views and conclusions of this thesis are not necessarily shared by these institutions.

Furthermore, the fast progress of this thesis benefited significantly from the expressive power of the *R programming language for statistical computing* (R Development Core Team, 2008). Especially, the *ggplot2* package for visualization (Wickham, 2009) and the *data.table* package (Dowle et al., 2014) proved to be very helpful. I also acknowledge the usage of the Vienna Scientific cluster for the achievement of certain computational results.

Finally, I want to express my gratitude to a few persons who did not directly contribute to my thesis, but accompanied me on the way.

It was a convenient experience to find in my PhD colleague Thomas Srdinko a person who shares my fascination for computer programming and technology. The nice coffee breaks with talks about non-thesis related topics were a welcome distraction from the often demanding work tasks. And of course, I really appreciated sitting outside in the park, enjoying the sun, and smoking a cigar like a boss on my birthday.

I was grateful to find a refuge at my family's place. In times when the workload forced me to restrict my diet to goulash from the tin and black coffee, the

occasional proper meals with my family were a welcome change. I really enjoyed these sporadic gatherings with my family.

Last but not least, I want to thank my friend Olha. Despite her own tight schedule during the last year, she lent me an ear when I needed it. Furthermore, she listened to me when I enthused about the most recent machine learning algorithms or cool statistical methods and how awesome they are. Thereby, she motivated me to steadily increase my knowledge and to stay creative in tackling the research questions of my thesis.

Contents

1	Introduction	1
I	Large Number of Observables	5
2	Bayesian statistics	6
2.1	Foundations of statistics	7
2.1.1	Countably many propositions	7
2.1.2	Uncountably many propositions	11
2.2	Multivariate normal distribution	14
2.3	Principle of maximum entropy	22
3	Nuclear data evaluation	27
3.1	Accounting for experimental data	29
3.2	Bayesian update of parameters	33
3.3	Bayesian update of observables	38
3.4	Mapping model data to experiment	43
3.4.1	Mapping of angle-integrated cross sections	45
3.4.2	Mapping spectra	48
3.5	Uncertainty about the incident energy	53
3.6	Conservation of features	56
4	Existing methods	61
4.1	Backward-Forward Monte Carlo	62
4.2	Bayesian Monte Carlo	64
4.3	Unified Monte Carlo-G	67
4.4	Unified Monte Carlo-B	69
4.5	Full Bayesian Evaluation Technique	70
4.6	Summary of methods	72
5	Large number of observables	75

5.1	Ensemble representation	76
5.2	Calculating the weights	80
5.3	Interpretation as Gaussian process	84
5.4	Database approach	86
5.5	Comparison to Monte Carlo methods	91
II	Consistent Model Defects	95
6	Consequences of model defects	96
6.1	Deficient linear model	97
6.2	Deficient non-linear model	109
6.3	Analysis of the EMPIRE-Kalman method	115
6.4	Analysis of the FBET/EMPIRE-MC method	120
6.5	Summary and conclusions	131
7	Consequences in Monte Carlo methods	134
7.1	Analysis of the UMC-G method	134
7.2	Refined Monte Carlo scheme	142
7.3	Analysis of the UMC-B method	143
7.4	Analysis of the BMC method	154
7.5	Analysis of the BFMC method	164
7.6	Summary and conclusions	173
8	Existing approaches	175
8.1	chi-square rescaling	176
8.2	Symmetric Monte Carlo	178
8.3	Ornstein-Uhlenbeck process	182
8.4	Scaling procedure	184
8.5	Conclusions	188
9	Consistent treatment of model defects	190
9.1	Model defects as Gaussian processes	190
9.2	Integration into the Bayesian update	197
9.3	Linearization of the Bayesian update	199
9.4	Marginal likelihood maximization	202
9.5	Showcase: deficient non-linear model	206
9.6	Sum rule conserving model defects	215
9.7	Assignment of model defects	221
9.8	Showcase: conservation of sum rules	224

10	Summary and outlook	238
A	Nuclear models	242
A.1	Neutron and proton optical potential	242
B	Statistics	246
B.1	Conjugate prior with respect to a likelihood	246
B.2	chi-square statistics	247
B.3	Woodbury identity	248
	Bibliography	249

Introduction

The aim of nuclear data evaluation is the generation of reliable and consistent estimates of reaction cross sections and related properties of the atomic nuclei. These estimates represent our best knowledge of these quantities (cross sections, decay rates, etc.) and are provided together with associated covariance matrices in the form of so-called evaluated nuclear data files to the user community. These evaluated data files are an important prerequisite for a wide range of applications such as dosimetry, medical diagnosis and therapy, and environmental research. Furthermore, they are the basis for the development of novel nuclear technology, such as nuclear fusion and nuclear waste incineration. The availability of uncertainty information in form of covariance matrices in recently established nuclear data files allows for extended simulations of novel facilities in order to optimize important design parameters with regard to efficiency and safety.

Originally, nuclear data files have been set up for neutron induced reactions in order to provide a basis for the design and construction of nuclear reactors. Nowadays, there is a more extended demand from the user community and evaluated data files for γ -, proton-, and deuteron-induced reactions have been set up. These data files are based on two sources of information: experimental data and results of model calculations. Latter are important to provide reliable predictions of observables in energy regions where experimental data are scarce or not accessible. For example, experimental information on neutron induced cross section data is limited for incident energies beyond 20 MeV, but required for the development of novel applications. Unfortunately, the nuclear many-body problem is of high complexity and hampers calculations based on first principles. Hence nuclear models have to be used which describe essential facets of the problem, but rely on parameters whose values are to some extent uncertain. Consequently, evaluation methods are applied to combine the information of the experimental data with the

results of model calculations to generate estimates of cross sections and associated covariance matrices at all relevant incident energies.

Several evaluation methods exist such as the FBET (Neudecker, 2012), BFMC (Bauge, Hilaire, and Dossantos-Uzarralde, 2007), BMC (Koning, 2015), GANDR (Muir et al., 2007), KALMAN (Kawano and Shibata, 1997), UMC-G (Smith, 2004), and UMC-B (Capote, Smith, et al., 2012). The majority of these methods has a clear interpretation within Bayesian statistics. Bayesian statistics is a well-founded framework which dates back to the work of Bayes (1763) and allows a consistent combination of prior knowledge with observations. In Bayesian statistics knowledge (or uncertainty for that matter) is modeled in terms of probability distributions. The so-called prior distribution reflects the state of knowledge before taking into account the observations. The likelihood is also a probability distribution and represents the information given by the observations. The multiplication of prior and likelihood yields besides a normalization constant the posterior probability distribution, which expresses a refined state of knowledge. The described calculation rule is known as the Bayesian update formula.

The reason for the increasing application of Bayesian statistics for problems of inference not only in nuclear data evaluation may be found in its theoretical properties. Bayesian inference coincides with logical reasoning in the case of certainty about the involved propositions. In the case of uncertainty about the propositions, Bayesian statistics complies with principles of common sense reasoning. Even the opposite direction has been proved by Cox (1946). Coherence with logical reasoning and principles of common sense reason leads necessarily to the Bayesian update formula.

Even though the majority of methods is clearly interpretable within Bayesian statistics, they are distinct from each other by two aspects. The first aspect concerns whether the original nuclear model is replaced by a simplified model. Using a simplified model leads to closed-form expressions for the calculation of the estimates and associated covariance matrices. If the exact model is employed, the latter information has to be obtained via Monte Carlo sampling. The second aspect is the choice of the probability distribution of the likelihood. Most methods use a multivariate normal distribution suggested by the principle of maximum entropy (Jaynes, 1957a). However, especially Monte Carlo methods often make other choices. So far, no consensus has been reached about the proper way of doing nuclear data evaluation. This situation is not satisfactory, because different evaluation methods may produce different estimates and uncertainty assessments in the same evaluation scenario. Usually, the validity of evaluations and hence also evaluation methods is tested by comparing results of simulations based on the evaluated data with measurements from integral benchmark experiments (e.g.

Markovskij et al. (2003)).

In this thesis another approach was followed to test the evaluation methods. The above mentioned methods were applied in a situation with a simple linear model and hypothetical experimental data. The experimental data were constructed in such a way that the model was not able to describe them properly. The evaluation methods relying on a multivariate normal likelihood produced unreasonable small uncertainties. Only evaluated uncertainties of some Monte Carlo methods were higher due to an alternative choice of the likelihood. However, it was clearly demonstrated that the evaluated covariance matrices of all methods only permit variations of the cross sections within the possibilities of the model. This property of the methods is undesirable in the case of a nuclear model which cannot mimic reliable experimental data.

In addition, the methods were studied in an evaluation of the neutron-induced total cross section of ^{181}Ta . The optical model in the parametrization of Koning and Delaroche (2003) was employed as nuclear model. The results showed the importance to consider model defects in an evaluation. Otherwise, uncertainties may become unreasonable small and the evaluated cross sections may be inconsistent with experimental data included in the evaluation. Furthermore, the modification of the likelihood in the BMC and BFMC method yielded uncertainties which seemed to be too large when taking into account the strong experimental evidence for a certain cross section curve. These findings indicate that the modification of the likelihood is not the proper way to take into account the deficiency of the nuclear model.

Due to this insight, a new formulation of the Bayesian update formula has been developed in this thesis to account for model defects. The prediction of the nuclear model is combined with a so-called model error function. The latter function describes the shape of the model error. Because the shape of this function is unknown before taking into account the experimental data, it is modeled as a random function governed by a Gaussian process. Gaussian process regression is a statistical method for non-parametric regression and has already found wide application in other fields such as geostatistics, economy, meteorology, and machine learning. Harnessing the respective methodology makes it possible to explicitly estimate the form of the model error, which is done for the first time in nuclear data evaluation. This information can be useful to refine the nuclear models. In contrast to existing approaches which account for model defects, the new approach preserves the sum rules of cross sections.

Another achievement of this thesis is the reformulation of the Bayesian update formula as implemented by the FBET (Neudecker, 2012) and EMPIRE-MC method (Herman et al., 2007). The evaluation according to the old scheme was

limited to tens of thousands of observables. In contrast to that, dozens of millions of observables can be simultaneously updated using the new formulation. The new formulation is well suited to be implemented as a database application: users can send a request with experimental data and an associated covariance matrix to the server, and the desired evaluated cross sections are computed and sent back.

Finally, the slow convergence of the Monte Carlo schemes introduced by the BMC, UMC-G and UMC-B in the evaluation scenario of the neutron-induced total cross section of ^{181}Ta necessitated a new Monte Carlo scheme. In this thesis a simple but effective Metropolis-Hastings algorithm is suggested which resolved the issues of the original Monte Carlo schemes.

This thesis is structured into two parts. Part I introduces Bayesian statistics and outlines existing methods to perform nuclear data evaluation. The last chapter of this part details the reformulated Bayesian update scheme to evaluate a large number of observables.

Part II deals with the consistent treatment of model defects. First, the consequences of model defects are studied in a scenario with a linear model and in an evaluation of the neutron-induced total cross section of ^{181}Ta . Then, existing methods to take into account model defects are discussed. Finally, the new approach to deal with model defects is presented and applied in an evaluation of the neutron-induced total cross section and the (n,2n) cross section of ^{181}Ta .

Part I

Large Number of Observables

Bayesian statistics

Methods of nuclear data evaluation are in general based on statistics. Using statistics, information about the properties of the nuclei is expressed in terms of probability distributions. Attempts to define the term probability have led to different results and were accompanied by philosophical discussion. Two definitions of probability are of particular importance for nuclear data evaluation.

In the frequentist interpretation, probability is defined as the relative frequency of events in a well-defined experiment when the number of trials approaches infinity. This definition is suited for measurements of nuclear quantities in experiments. Apart from constraints given by available measurement time and money, experiments can be repeated arbitrary many times. Counting rates of detectors are clearly defined events.

In the Bayesian interpretation, probability represents a degree of belief about the truth of a proposition. This definition enables to apply statistics to a broader range of problems than possible with the frequentist interpretation. For instance, knowledge about the range of a parameter in a nuclear model can be cast into a probability distribution. Values for that parameter associated with higher probability are believed to be more likely than those with lower probability. The frequentist definition, in contrast, would not allow such a usage of probabilities because of the missing link to a well-defined experiment.

This difference in interpretation splits the field of statistics in two schools of thought, Bayesian statistics and frequentist statistics. Both schools of thought work with the same set of basic formulas to calculate probabilities. They differ in how these formulas are applied for statistical inference.

In nuclear data evaluation, we often face the problem of insufficient data. Designing modern nuclear facilities requires knowledge of reaction cross sections at incident energies above 20 MeV, yet experimental data is often only available be-

low that energy. The remedy is to take into account results of model calculations. These results are dependent on the values chosen for the parameters of the nuclear models. There is uncertainty about the true values of these parameters, which has to be accounted for in an evaluation. Consequently, contemporary evaluation methods are based on Bayesian statistics.

The three fundamental concepts in Bayesian statistics are the prior, the likelihood, and the posterior. Assume that we want to infer the value of a cross section. The prior represents our belief about the true value of the cross section before looking at the data. The likelihood expresses how likely observing a certain outcome in an experiment is taking into account the prior knowledge. After performing an experiment, the combination of the prior and the likelihood yields the posterior. The posterior represents an improved knowledge about the value of the cross section based on both sources of information, the prior knowledge and the experimental data.

The specification of a probability distribution for the prior is the first step of an evaluation. Usually, prior knowledge about a cross section is limited to a best guess and an idea about the accuracy of this guess. This information is not sufficient to unambiguously define a probability distribution. In order to reduce subjectivity in the assignment, principles such as the principle of maximum entropy (Jaynes, 1957a,b) and the principle of invariant transformation groups (Jaynes, 1968) are applied.

The next section outlines the foundations of probability theory and introduces the Bayesian update formula, which is fundamental for this thesis. Thereafter, we elaborate on the multivariate normal distribution, which is without exception used in this thesis to specify the probability distributions for the Bayesian update formula.

2.1 Foundations of statistics

Statistical inference operates on probabilities. We follow the Bayesian interpretation and define probability as degree of belief and introduce the basic concepts of statistics in a Bayesian diction. First, we discuss the basic concepts and relations of statistics for countably many propositions. In a second step, we extend the concepts to the case of uncountably many propositions.

2.1.1 Countably many propositions

A set of countably many proposition is given if there is an enumeration scheme that does not leave out any of the propositions. Therefore, every finite set of proposi-

tions is countable. An example for a countable set of infinitely many propositions is

$$A_n \equiv n \text{ persons wait in the queue.} \quad (2.1)$$

A probability to each of the propositions A_n can be assigned reflecting the belief about its truth. These assignments must conform to the probability axioms set by Kolmogorov, e.g. (Kolmogorov, 2000). We introduce them in a Bayesian diction.

Probability axioms

1. The probability of a proposition A is a non-negative number,

$$P(A_i) \geq 0. \quad (2.2)$$

2. The probability that any of the mutually exclusive propositions A_1, A_2, \dots is true is given by

$$P(A_1 \vee A_2 \vee \dots) = P(A_1) + P(A_2) + \dots \quad (2.3)$$

3. For a set of mutually exclusive propositions A_i of which one has to be true, it holds that

$$P\left(\bigvee_i A_i\right) = \sum_i P(A_i) = 1. \quad (2.4)$$

Two mutually exclusive propositions can never be true at the same time. For instance, the propositions in Equation 2.1 are mutually exclusive. First and third axiom imply that a probability is a number in the closed interval $[0, 1]$. Propositions associated with probability one are considered to be certainly true. Conversely, propositions associated with probability zero are considered to be certainly false.

An important concept of statistics is *conditional probability*. For two propositions A and B , the conditional probability is defined as

$$P(A|B) := \frac{P(A \wedge B)}{P(B)}. \quad (2.5)$$

It denotes the probability for proposition A to be true under the condition that B is known to be true. From the definition of conditional probability and the probability axioms two important formulas to calculate probabilities follow.

Rules to calculate probabilities

$$\text{sumrule} \quad P(A \vee B) = P(A) + P(B) - P(A \wedge B) \quad (2.6)$$

$$\text{productrule} \quad P(A \wedge B) = P(A|B)P(B) \quad (2.7)$$

The probability axioms of Kolmogorov are not the only possible approach to probability theory. The approach of Cox (2001) is of particular significance for the Bayesian school of thought. His derivation promotes the laws of probabilities as an extension of Aristotelian logic where the truth of propositions is uncertain.

Inference in Bayesian statistics is performed by the application of the Bayesian update formula. The Bayesian update formula is the cornerstone of Bayesian statistics.

Bayesian update formula

Let A and B be propositions. $P(A)$ and $P(B) > 0$ are the associated probabilities. The Bayesian update formula is given by

$$P(A|B) = \frac{P(B|A)}{P(B)}P(A), \quad (2.8)$$

where $P(A)$ is called the prior probability distribution, $P(B|A)$ the likelihood, $P(B)$ the evidence, and $P(A|B)$ the posterior probability distribution.

The Bayesian update formula follows directly from the product rule. Using the fact that $A \wedge B$ and $B \wedge A$ denote the same proposition, we can apply Equation 2.7 once for $P(A \wedge B)$ and another time for $P(B \wedge A)$. Equating the right hand sides leads to Equation 2.8.

We give an example of Bayesian inference, which demonstrates the interpretation of Bayesian statistics as an extension of Aristotelian logic. Suppose we want to know whether it rained. We can try to obtain an answer by examining the ground. If the ground is dry, we can conclude that it did not rain. If the ground is wet, we have some indication that it indeed rained. However, a wet ground is not a proof for rain because perhaps it was just artificially irrigated. We want to obtain an answer whether it rained by performing Bayesian inference. To this end, we must assign probabilities to the propositions. Table 2.1 shows our choices for this example.

Assume that we observed wet ground. To get a refined estimate whether it rained, we use the Bayesian update formula, Equation 2.8, to include the additional information obtained by our observation,

$$P(\text{rain} | \text{wet ground}) = \frac{P(\text{wet ground} | \text{rain})}{P(\text{wet ground})}P(\text{rain}). \quad (2.9)$$

The probability $P(\text{wet ground})$ is not listed in the table. To keep expressions succinct, we denote the converse of hypothesis A by $\neg A$. It can be computed from the listed probabilities by applying the product rule and the sum rule,

$$P(\text{wet ground}) = P(\text{rain})P(\text{wet ground} | \text{rain}) + P(\neg\text{rain})P(\text{wet ground} | \neg\text{rain}). \quad (2.10)$$

Given the numbers in the table, we have $P(\text{wet ground}) = 0.6$. Inserting the values for all required probabilities in Equation 2.9 yields $P(\text{rain} | \text{wet ground}) = 0.8\dot{3}$. The observation of wet ground leads us to refine our assessment. Now we regard the proposition that it rained more likely than before. However, we are still not certain about whether it rained, because rain is not the only possible cause for a wet ground.

Let us consider the opposite case. If we observe dry ground, the Bayesian update formula takes the form

$$P(\text{rain} | \neg\text{wet ground}) = \frac{P(\text{wet ground} | \text{rain})}{P(\neg\text{wet ground})}P(\text{rain}) \quad (2.11)$$

We get $P(\text{rain} | \neg\text{wet ground}) = 0$. The observation of dry ground allows us to conclude with certainty that it did not rain. This inference is equivalent to the syllogism

From dry ground it follows that it did not rain.

The ground is dry.

Hence, it did not rain.

The first statement is equivalent to the proposition ‘rain causes wet ground’. The inference made with Bayesian statistics coincides with classical logic. This finding demonstrates an important consistency feature of Bayesian statistics. In the limiting case of certainty about propositions, Bayesian inference reduces to logical reasoning.

Prior	Likelihood
$P(\text{rain}) = 0.5$	$P(\neg\text{wet ground} \text{rain}) = 0$
$P(\neg\text{rain}) = 0.5$	$P(\text{wet ground} \text{rain}) = 1$
	$P(\neg\text{wet ground} \neg\text{rain}) = 0.8$
	$P(\text{wet ground} \neg\text{rain}) = 0.2$

Table 2.1: Inferring whether it rained by examining the ground. Assignment of probabilities to the propositions for Bayesian inference. The equal prior probabilities reflect our indifference a priori whether it rained. The first two probabilities under likelihood correspond to our assumption that rain always causes a wet ground. The remaining two probabilities express that a wet ground is indication for rain but not a proof.

2.1.2 Uncountably many propositions

A set of uncountably many propositions is given if there exists no scheme to exhaustively enumerate over the propositions. Uncountably sets contain always an infinite number of elements. An example for a set of uncountably many propositions is

$$A(t) \equiv \text{The temperature in this room is } t^\circ \text{ Celsius.} \quad (2.12)$$

The temperature t is a real number, which makes it impossible to enumerate over all propositions without leaving out any. In general, the assignment of probabilities to propositions is not possible anymore. An assignment of non-zero probabilities to an uncountable subset of the propositions would violate Equation 2.4. The notion of probability has to be extended.

Probability density function

A function $\rho : \mathbb{R} \rightarrow [0, \infty)$ is called a probability density function (pdf) if it is associated to a bijective mapping from real numbers to mutual exclusive propositions $A(t)$, exactly one of them true, such that

$$P(t_1 \leq t \leq t_2) = \int_{t_1}^{t_2} \rho(t) dt, \quad (2.13)$$

where $P(t_1 \leq t \leq t_2)$ is the probability for the true proposition being located in the set of propositions corresponding to the numbers in the interval $[t_1, t_2]$. Taking $t_1 = -\infty$ and $t_2 = +\infty$ yields 1, because then the true proposition is in the covered interval for sure.

In contrast to probabilities, probability density functions are allowed to yield values greater than one. Only integrals of them must not be greater than one. This thesis deals with the estimation of cross sections, which are continuous quantities. Thus, propositions of interest are of the form

$$\begin{aligned} A(x_1, x_2, \dots) &= A_1(x_1) \wedge A_2(x_2) \wedge \dots \quad \text{with} & (2.14) \\ A_1(x_1) &= \text{the total cross section is } x_1 \text{ mBarn,} \\ A_2(x_2) &= \text{the elastic cross section is } x_2 \text{ mBarn,} \\ &\vdots \end{aligned}$$

Two composite propositions $A(x_1, x_2, \dots)$, $A(x'_1, x'_2, \dots)$ are distinct if $x_i \neq x'_i$ for some i . Furthermore, we assume that two distinct propositions are mutually exclusive. Maybe we are not certain which choice of the x_i corresponds to the true proposition, but we know that exactly one of the propositions has to be true.

We restrict further discussion to a composite proposition $A(x, y)$ with two arguments. However, the extension to more variables is straight-forward. The probability density function for a compound proposition $A(x, y)$ can be defined analogously to the univariate case and we write $\rho(x, y)$. The probability that a composite proposition in the set defined by $x_1 < x < x_2, y_1 < y < y_2$ is true is given by

$$P(x_1 < x < x_2 \wedge y_1 < y < y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} \rho(x, y) dx dy. \quad (2.15)$$

Suppose we are only interested in the probability density distribution of x irrespective of the value of the other quantity y . We have to integrate over y to obtain the *marginal probability density*,

$$\rho_X(x) := \int_{-\infty}^{+\infty} \rho(x, y) dy. \quad (2.16)$$

Using the definition of the marginal density distribution, we can define the *conditional probability density*,

$$\rho(x|y) := \frac{\rho(x, y)}{\rho_Y(y)}, \quad (2.17)$$

which denotes the probability density for x under the condition that we know the true value of y . Finally, having defined the marginal pdf and the conditional pdf, we can introduce the Bayesian update formula for the case of uncountably many propositions.

Bayesian update formula for uncountably many propositions

Let $\rho(x, y)$ be a bivariate probability density distribution. Let $\rho_X(x)$ and $\rho_Y(y)$ be the associated marginal probability distributions. The Bayesian update formula is given by

$$\rho(x|y) = \frac{\rho(y|x)}{\rho_Y(y)} \rho_X(x). \quad (2.18)$$

We call $\rho(x|y)$ the posterior pdf, $\rho(y|x)$ the likelihood, $\rho_Y(y)$ the evidence, and $\rho_X(x)$ the prior pdf.

We conclude this section with an example. Suppose we want to infer the current room temperature. We model our prior belief about the true temperature t as a standard normal distribution,

$$\rho_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right). \quad (2.19)$$

This probability density distribution encodes our belief that the most likely temperature is 0° Celsius and we do not expect the true temperature to deviate more

than a few degree from this value. More details of the normal distribution follow in the next section.

We want to refine our prior belief by measuring the temperature with a thermometer. We know that the measured temperature t' usually differs from the true temperature t due to measurement error. We assume that the deviation also follows a normal distribution,

$$\rho(t'|t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(t-t')^2\right). \quad (2.20)$$

The application of the Bayesian update formula, Equation 2.18, additionally requires the calculation of the evidence $\rho_{T'}(t')$. The evidence is the marginal distribution of $\rho(t, t')$ where the true temperature t has been integrated out, see Equation 2.16. We remark that also the prior pdf $\rho_T(t)$ can be regarded as the marginal distribution where the measurement t' has been integrated out. The definition of conditional probability density, Equation 2.17, enables us to calculate the evidence,

$$\rho_{T'}(t') = \int_{-\infty}^{\infty} \rho(t, t') dt = \int_{-\infty}^{\infty} \rho(t'|t) \rho_T(t) dt = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}t'^2\right). \quad (2.21)$$

Inserting prior pdf (Equation 2.19), likelihood (Equation 2.20), and evidence Equation 2.21) into the Bayesian update formula (Equation 2.18) yields the posterior pdf,

$$\rho(t|t') = \frac{\rho(t'|t)}{\rho_{T'}(t')} \rho_T(t) = \frac{1}{\sqrt{\pi}} \exp\left(-(t-t'/2)^2\right). \quad (2.22)$$

The posterior pdf is a function of the true temperature t . The other variable t' has to be regarded as a constant which is determined by measurement. For instance, if $t' = 2$ has been measured, we use $\rho(t' = 2|t)$ for the likelihood in the Bayesian update formula. The relations of prior, likelihood and posterior are illustrated in Figure 2.1. The posterior pdf is a normal distribution, which is due to the fact that both likelihood and prior are normal distributions. In general, if the posterior pdf lies in the same class of distributions as the prior pdf, one speaks of a *conjugate prior* with respect to the likelihood (see section B.1). The concept of conjugate priors was introduced by Raiffa and Schlaifer (1961). Working with conjugate priors is convenient, because the Bayesian update formula reduces to calculating the characterizing parameters of the distribution. In the case of the (univariate) normal distribution, these parameters are the mean and the variance. More about these quantities in the next section.

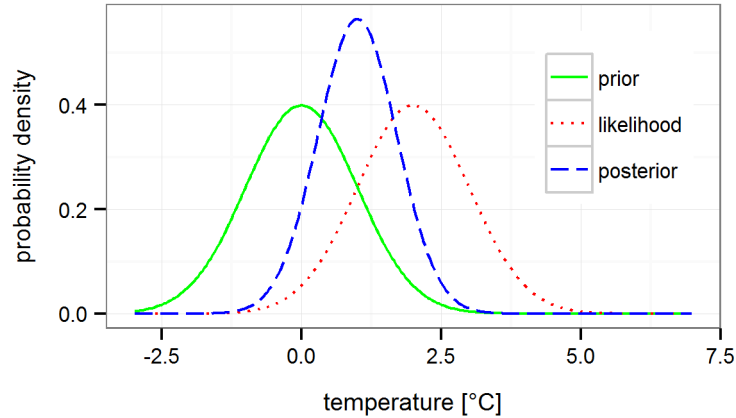


Figure 2.1: Bayesian inference of the room temperature. The most likely temperature according to the prior is 0° Celsius. The measured temperature is 2° Celsius which leads to the shown likelihood. Combining prior and likelihood with the Bayesian update formula yields the posterior. The tighter shape of the posterior compared to the prior corresponds to a reduced uncertainty about the true value of the temperature.

2.2 Multivariate normal distribution

The multivariate normal distribution takes a special role in this thesis. Priors and likelihoods for cross sections, spectra, and angle-differential cross sections will be without exception specified by multivariate normal distributions. Because of this choice, the resulting posteriors will also follow multivariate normal distributions. This feature allows for the efficient computation of the Bayesian update formula. Theoretical justification for using multivariate normal distributions comes from the principle of maximum entropy (Jaynes, 1957a), which is discussed in the next section.

A multivariate normal distribution is a probability distribution for a set of random variables X_i , which we combine to a vector \vec{X} . It is characterized by the mean values of the X_i and the covariances between them. For the following considerations we distinguish between the random variable X_i and a realization x_i thereof. The former refers to the random variable with all its properties, such as its probability distribution. The latter is just a number we could draw from the random variable, for instance in terms of a measurement in an experiment.

The mean value and the covariance can be calculated for arbitrary probability distributions $\rho(\vec{x})$. Especially, we obtain the mean value of the random variable X_i in terms of the expectation operator,

$$\mu_i = \text{E}[X_i] = \int x_i \rho(\vec{x}) dx_1 \dots dx_n. \quad (2.23)$$

Integration has to be performed over the set of feasible realizations \vec{x} . The covari-

ance between two random variables X_i and X_j can be computed by applying the covariance operator,

$$\text{Cov}[X_i, X_j] = \int (x_i - \text{E}[X_i]) (x_j - \text{E}[X_j]) \rho(\vec{x}) dx_1 \dots dx_n. \quad (2.24)$$

Obviously, the covariance operator is symmetric with respect to its arguments, $\text{Cov}[X_i, X_j] = \text{Cov}[X_j, X_i]$. The covariance of a variable with itself is an important quantity and referred to as variance. We introduce the variance operator,

$$\text{Var}[X_i] = \text{Cov}[X_i, X_i] = \int (x_i - \text{E}[X_i])^2 \rho(\vec{x}) dx_1 \dots dx_n. \quad (2.25)$$

These quantities represent first and second order moments of the probability distribution and yield important information about the random variable. For instance, if some X_i represents a nuclear model parameter and we assume that X_i is governed by a normal distribution, then parameter values in the interval $\text{E}[X_i] \pm \sqrt{\text{Var}[X_i]}$ are believed to be about two times more likely than values outside this interval. In more technical terms, the probability mass within this interval is approximately 0.683 and outside that interval 0.317. The smaller the variance of a random variable, the smaller the uncertainty about its true value.

The mean value of a random variable $\text{E}[X_i]$ has the important property that it leads to minimal variance. Replacing $\text{E}[X_i]$ in Equation 2.25 with any other value would increase the result. Due to this property, the mean value is a reasonable choice as a best guess for a random variable governed by a symmetrical and unimodal probability distribution.

The covariance between two random variables $\text{Cov}[X_i, X_j]$ is an unnormalized measure of their mutual dependence. To get a normalized measure, we introduce the correlation operator,

$$\text{Cor}[X_i, X_j] = \frac{\text{Cov}[X_i, X_j]}{\sqrt{\text{Var}[X_i] \text{Var}[X_j]}}. \quad (2.26)$$

The correlation can take values in the closed interval $[-1, 1]$. A positive correlation indicates that a higher value x_i is more likely accompanied by a higher value x_j . Conversely, a negative correlations indicates that a higher value x_i is accompanied by a lower value x_j . Furthermore, the closer the correlation is to the boundaries of the interval, the more the values of the random variables are dependent on each other. In the extreme case of a perfect positive or negative correlation, $\text{Cor}[X_i, X_j] = \pm 1$, knowledge of the value x_i is sufficient to infer the value x_j . In general, knowing x_i leads to a relative reduction of the variance of x_j by $\text{Cor}[X_i, X_j]^2$.

Because the expectation operator is a linear functional, it satisfies

$$\text{E}[\alpha + \beta X_i + \gamma X_j] = \alpha + \beta \text{E}[X_i] + \gamma \text{E}[X_j], \quad (2.27)$$

for arbitrary scalars α, β, γ . Similarly, we have for the covariance operator

$$\text{Cov}[\alpha + \beta X_i + \gamma X_j, X_k] = \beta \text{Cov}[X_i, X_k] + \gamma \text{Cov}[X_j, X_k]. \quad (2.28)$$

Additive constants leave the result unaffected. Due to the symmetry of the covariance operator, the same identity holds for the second argument. For the variance operator as a special case of the covariance operator, we get

$$\text{Var}[\alpha + \beta X_i + \gamma X_j] = \beta^2 \text{Var}[X_i] + \gamma^2 \text{Var}[X_j] + 2\beta\gamma \text{Cov}[X_i, X_j]. \quad (2.29)$$

These linear relationships are of central importance for the developments presented in this thesis. First, they lead in a straightforward way to the conservation of sum rules and the relation between angle-integrated cross sections and the associated angle-differential cross sections. Furthermore, the novel treatment of model defects introduced in chapter 6 also relies on them.

After this concise elaboration on the expectation, variance and covariance operator, we introduce the multivariate normal distribution. A collection of random variables X_i follows a multivariate normal distribution if and only if all possible linear combination of them follow a univariate normal distribution,

$$\sum_{i=1}^N c_i X_i \sim \mathcal{N}(\mu, \sigma^2), \quad (2.30)$$

where the constants c_i are arbitrary real numbers. The probability density function of a univariate normal distribution with mean μ and standard deviation σ for a random variable X is

$$\rho(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right). \quad (2.31)$$

The application of the expectation operator and the covariance operator yields

$$\text{E}[X] = \mu \quad \text{and} \quad \text{Var}[X] = \sigma^2. \quad (2.32)$$

Mean value and variance are sufficient to unambiguously determine the shape of the normal distribution. The standard deviation σ and the variance σ^2 give the same information. The specific task of the statistical study considered provides usually criteria which one is more convenient to work with. In the case of the normal distribution, the standard deviation is directly related to probabilities. The probability mass in the interval $[\mu - \sigma, \mu + \sigma]$ is 0.683 and in the interval $[\mu - 2\sigma, \mu + 2\sigma]$ 0.954. However, calculations relying on the identities Equation 2.27, Equation 2.28 or Equation 2.29 are better carried out at the level of variances.

Another—and for our purposes more suitable—definition is to directly specify the probability density distribution of the multivariate normal distribution. To this end, we introduce for the set of random variables $\{X_i\}_{i=1:N}$ the abbreviations

$$\mu_i = \text{E}[X_i] \quad \text{and} \quad v_{ij} = \text{Cov}[X_i, X_j]. \quad (2.33)$$

The mean values μ_i can be combined into the mean vector $\vec{\mu}$ and the covariances v_{ij} in the covariance matrix \mathbf{V} . Then, the pdf of the multivariate normal distribution is given by

$$\rho(\vec{x}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{V}|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{V}^{-1}(\vec{x} - \vec{\mu})\right), \quad (2.34)$$

where d is the dimension of \vec{x} . This probability density distribution is ubiquitous in this thesis. Priors, likelihoods and posteriors without exception are modeled as multivariate normal distributions. The elements of $\vec{\mu}$ and \vec{x} represent angle-integrated cross sections at different incident energies, angle-differential cross sections at different incident energies and angles, and spectra at different incident energies and emission energies.

However, definition Equation 2.34 is only applicable if \mathbf{V} is positive definite. For some multivariate normal distributions \mathbf{V} is only positive semi-definite. For instance, consider the two random variables X_1 and X_2 which are related by $X_2 = cX_1$. In this case, the covariance matrix takes the form

$$\mathbf{V} = \begin{pmatrix} v_{11} & c v_{11} \\ c v_{11} & c^2 v_{11} \end{pmatrix}. \quad (2.35)$$

The second row is a multiple of the first row. Thus, the covariance matrix has not full rank and its inverse is not defined. Consequently, the probability density distribution in Equation 2.34 is also not defined. Nevertheless, if X_1 follows a univariate normal distribution, the definition relying on linear combinations, Equation 2.30, still applies.

To explicitly obtain the probability density function of a multivariate normal distribution whose covariance matrix \mathbf{V} is only positive semi-definite, one can perform an eigen decomposition. We assume that the eigenvalues λ_i are sorted in decreasing order. With \vec{e}_i we refer to the eigenvector associated with the eigenvalue λ_i . One takes the d leading eigenvectors associated to non-zero eigenvalues and assembles them into a matrix,

$$\mathbf{P} = (\vec{e}_1, \vec{e}_2, \dots, \vec{e}_d) \quad \text{where } \lambda_i > 0, \forall i \leq d. \quad (2.36)$$

Instead of working with the original random vector \vec{X} we work with the random vector \vec{Y} . The relation between these two vectors is

$$\vec{X} = \mathbf{P}\vec{Y}. \quad (2.37)$$

The dimension of \vec{Y} is given by the number of non-zero eigenvalues d of the covariance matrix \mathbf{V} . Finally, we need to transform the mean vector $\vec{\mu}$ and the covariance matrix \mathbf{V} ,

$$\vec{\tau} = \mathbf{P}^T \vec{\mu} \quad \text{and} \quad \mathbf{W} = \mathbf{P}^T \mathbf{V} \mathbf{P}. \quad (2.38)$$

Now, because the inverse of \mathbf{W} exists, the probability density function can be stated,

$$\rho(\vec{x}) = \frac{1}{\sqrt{2\pi|\mathbf{W}|}} \exp\left(-\frac{1}{2}(\vec{y} - \vec{\tau})^T \mathbf{W}^{-1}(\vec{y} - \vec{\tau})\right). \quad (2.39)$$

In order to demonstrate the procedure we consider the example of Equation 2.35. This covariance matrix resulted from the relation $X_2 = cX_1$ between the random variables. Its eigenvalues and associated eigenvectors are

$$\lambda_1 = v_{11}(1 + c^2), \quad \vec{e}_1 = \frac{1}{\sqrt{1 + c^2}} \begin{pmatrix} 1 \\ c \end{pmatrix} \quad \text{and} \quad \lambda_2 = 0, \quad \vec{e}_2 = \frac{1}{\sqrt{1 + c^2}} \begin{pmatrix} c \\ -1 \end{pmatrix}. \quad (2.40)$$

The normalized eigenvectors are orthogonal, $\vec{e}_1 \perp \vec{e}_2$. The pairwise orthogonality of eigenvectors is a general property of covariance matrices due to their symmetry, $V_{ij} = V_{ji}$. Even though we consider two random variables, there is only one degree of freedom because of the linear dependence of the rows of the covariance matrix. Thus, the new random variable \vec{Y} contains only one element Y_1 . By applying the projection operator $\mathbf{P} = \vec{e}_1$, we get a parameter equation for the straight line that corresponds to feasible combinations (X_1, X_2) ,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \frac{1}{\sqrt{1 + c^2}} \begin{pmatrix} 1 \\ c \end{pmatrix} Y_1. \quad (2.41)$$

The transformed covariance matrix \mathbf{W} contains only one element W_{11} , which represents the variance of Y_1 . We get $W_{11} = \mathbf{P}^T \mathbf{V} \mathbf{P} = v_{11}(1 + c^2)$, which coincides with the eigenvalue λ_1 . This finding is of general validity: If we take some normalized vector \vec{k} pointing into an arbitrary direction, and ‘sandwich’ the covariance matrix, $\vec{k}^T \mathbf{V} \vec{k}$, we get the variance along the direction of \vec{k} . If the vector \vec{k} is an eigenvector \vec{e}_i of the covariance matrix, $\vec{e}_i^T \mathbf{V} \vec{e}_i$ yields the respective eigenvalue λ_i . Therefore, the eigenvalues of the covariance matrix are the variances along the principal axis of the covariance matrix. The first principal axis is associated with the direction of largest variance. The second principal axis is associated with the largest variance under the condition that it is orthogonal to the first principal axis. In general, the n^{th} principal axis is defined by largest variance under the condition that it is orthogonal to all previous $(n - 1)$ principal axes. Multivariate normal distributions for less or equal three random variables can also be visualized by plotting their confidence ellipses. As an example, Figure 2.2 shows a bivariate normal distribution.

So far, we have introduced the definition of the multivariate normal distribution and the important notions of the expectation, variance, covariance and correlation. Next, we discuss the marginal probability density and the conditional probability density of the multivariate normal distribution. The general definition of these

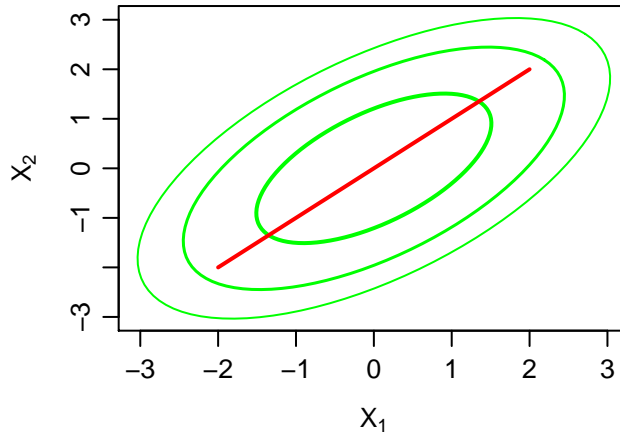


Figure 2.2: The green ellipses are a visualization of a bivariate normal distribution with mean vector $\vec{0}$, variances $\text{Var}[X_1] = \text{Var}[X_2] = 1$ and covariance $\text{Cov}[X_1, X_2] = 0.6$. The probability to obtain a combination of realizations (x_1, x_2) enclosed by one of the green ellipses is 0.68 for the inner ellipse, 0.95 for the middle ellipse, and 0.99 for the outer ellipse. The principal axes of the ellipses are associated with the eigenvectors of the covariance matrix. The feasible combinations of realizations (x_1, x_2) span an area. However, if the random variables are perfectly correlated due to linear dependence, the ellipses degenerate to a straight (red) line. In this case, the probability to obtain a combination not lying on the line is zero. The mathematical signature of linear dependent random variables is a non-invertible covariance matrix.

concepts was given in Equation 2.16 and Equation 2.17. Assume that we have a set of random variables $\{X_i\}_{i=1:n}$. We can partition them into the two disjoint subsets $\{X_i\}_{i=1:k}$ and $\{X_i\}_{i=(k+1):n}$. We combine the random variables of the first subset to the random vector \vec{Y} and those of the second vector into the random vector \vec{Z} . Let the expectation of \vec{Y} be denoted by \vec{a} and the expectation of \vec{Z} be denoted by \vec{b} . The matrix \mathbf{U} contains the covariances for \vec{Y} and the matrix \mathbf{V} the covariances for \vec{Z} . The covariance matrix relating \vec{Y} and \vec{Z} is denoted by \mathbf{C} . Using this notation, the mean vector and covariance matrix for the complete set of random variables governed by a multivariate normal distribution can be written as

$$\vec{\mu} = \begin{pmatrix} \vec{a} \\ \vec{b} \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} \mathbf{U} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{V} \end{pmatrix}. \quad (2.42)$$

We use this partitioned form of the mean vector and the covariance matrix as a basis to discuss the marginal distribution and the conditional distribution.

The marginal distribution of a subset of the random variables is also a multivariate normal distribution. It can be obtained by removing elements in the mean vector and the covariance matrix that are connected to random variables not element of the subset. For instance, the marginal distribution of the random variables in \vec{Y} is obtained by only keeping \vec{a} in the mean vector and \mathbf{U} in the

covariance matrix,

$$\rho_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int \rho(x_1, \dots, x_n) dx_{k+1} \dots dx_n \equiv \mathcal{N}(\vec{a}, \mathbf{U}). \quad (2.43)$$

Being able to obtain *marginal probability densities* by removing elements of the mean vector and the covariance matrix represents a huge advantage. In general, the evaluation of the multidimensional integral for the marginal probability densities for general distributions is very difficult if not infeasible.

If we know the realizations x_i of a subset of the random variables, we are perhaps able to learn something about the unobserved random variables due to correlations. For instance, the weather of two cities situated nearby is not independent. Knowing that the sun is shining in one city could increase the probability that the sun is also shining in the other city. Suppose we know the realizations \vec{z} of the random vector \vec{Z} . The probability distribution for the random vector \vec{Y} conditioned on \vec{z} is also a multivariate normal distribution. Its mean vector $\vec{a}_{\vec{z}}$ and covariance matrix $\mathbf{U}_{\vec{z}}$ is given by (e.g. Mises and Geiringer (1966))

$$\vec{a}_{\vec{z}} = \vec{a} + \mathbf{C}\mathbf{V}^{-1}(\vec{z} - \vec{b}) \quad (2.44)$$

$$\mathbf{U}_{\vec{z}} = \mathbf{U} - \mathbf{C}\mathbf{V}^{-1}\mathbf{C}^T \quad (2.45)$$

The formula for the conditional distribution of the multivariate normal distribution is one way to derive the Bayesian update formula specialized to the case of multivariate prior and likelihood.

Another important property of the multivariate normal distribution is its invariance under linear transformations of the random variable. Suppose we have a random variable \vec{X} governed by $\mathcal{N}(\vec{\mu}, \Sigma)$ and perform an affine transformation

$$\vec{Y} = \mathbf{S}\vec{X} + \vec{b}. \quad (2.46)$$

The transformed variable \vec{Y} is multivariate normal with mean vector and covariance matrix

$$\vec{\mu}' = \mathbf{S}\vec{\mu} + \vec{b} \quad \text{and} \quad \Sigma' = \mathbf{S}\Sigma\mathbf{S}^T. \quad (2.47)$$

In particular, the expression for the transformed covariance matrix is frequently denoted as *sandwich formula*. The invariance of the multivariate normal distribution under linear transformations is particularly useful for nuclear data evaluation. For instance, apriori knowledge of nuclear model parameters is usually specified in terms of a multivariate normal distribution. The linearization of the nuclear model leads to a linear relation between observables and nuclear model parameters, $\vec{\sigma} = \mathbf{S}\vec{p} + \vec{\sigma}_0$. Hence, one obtains a multivariate normal distribution for the observables which can be easily computed from the multivariate normal distribution of the model parameters.

Finally, both the product and the convolution of the probability density functions of two multivariate normal distributions yield a multivariate normal distribution. However, in the case of the product, the resulting multivariate normal distribution is not normalized anymore.

In general, the convolution is related to the sum of random vectors. Suppose we want to know the probability density function of the sum of independent multivariate normal random vectors,

$$\vec{Z} = \vec{X}_1 + \vec{X}_2 + \cdots + \vec{X}_n. \quad (2.48)$$

We take it as granted that \vec{Z} will also follow a multivariate normal distribution. Because the X_i are independent, their joint probability density function is given by

$$\rho(\vec{x}_1, \dots, \vec{x}_n) = \rho_1(\vec{x}_1) \rho_2(\vec{x}_2) \cdots \rho_n(\vec{x}_n). \quad (2.49)$$

To get the probability density for one particular realization \vec{z} , we have to integrate over all possible combinations of the \vec{x}_i that sum up to \vec{z} . We end up with the convolution

$$\rho(\vec{z}) = \int \rho_1 \left(\vec{z} - \sum_{i=2}^n \vec{x}_i \right) \rho_2(\vec{x}_2) \cdots \rho_n(\vec{x}_n) dx_2 \cdots dx_n. \quad (2.50)$$

We can make use of the properties of the expectation operator and covariance operator (see Equation 2.27 and Equation 2.28) to obtain the distribution parameters. The mean vector and the covariance matrix associated with \vec{Z} are

$$\vec{\mu} = \text{E}[\vec{Z}] = \text{E}[\vec{X}_1] + \text{E}[\vec{X}_2] + \cdots + \text{E}[\vec{X}_n], \quad (2.51)$$

$$\Sigma = \text{Var}[\vec{Z}] = \text{Var}[\vec{X}_1] + \text{Var}[\vec{X}_2] + \cdots + \text{Var}[\vec{X}_n]. \quad (2.52)$$

Here, we extended the applicability of the variance operator to random vectors. The variance operator applied to a matrix shall give the covariance matrix for the elements of the random vector. To summarize, we have for the random vector \vec{Z}

$$\vec{Z} \sim \mathcal{N} \left(\sum_i \text{E}[X_i], \sum_i \text{Var}[X_i] \right). \quad (2.53)$$

Not only the convolution of multivariate normal distributions yields a multivariate normal distribution, but also their product. Assume we have the two probability density function $\rho_1(\vec{x}) = \mathcal{N}(\vec{\mu}_1, \mathbf{A}_1)$ and $\rho_2(\vec{x}) = \mathcal{N}(\vec{\mu}_2, \mathbf{A}_2)$. Their product is given by

$$\rho_1(\vec{x}) \rho_2(\vec{x}) = Z^{-1} \frac{1}{\sqrt{2\pi|\mathbf{C}|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{z})^T \mathbf{C}^{-1} (\vec{x} - \vec{z}) \right\}, \quad (2.54)$$

with

$$\vec{z} = \mathbf{C}(A_1^{-1}\vec{\mu}_1 + A_2^{-1}\vec{\mu}_2) \quad \text{and} \quad \mathbf{C} = (\mathbf{A}_1^{-1} + \mathbf{A}_2^{-1})^{-1}. \quad (2.55)$$

The probability distribution in Equation 2.54 is not normalized anymore due to the occurrence of the extra factor Z^{-1} , which is specified by

$$Z^{-1} = \frac{1}{\sqrt{(2\pi)^d |\mathbf{A}_1 + \mathbf{A}_2|}} \exp \left\{ -\frac{1}{2} (\vec{\mu}_1 - \vec{\mu}_2)^T (\mathbf{A}_1 + \mathbf{A}_2)^{-1} (\vec{\mu}_1 - \vec{\mu}_2) \right\}. \quad (2.56)$$

Concerning the multivariate normal distribution, the outlined properties and relations give sufficient background information to discuss two developments in this thesis: the extension of the applicability of the linearized Bayesian update formula to a large number of observables, and the consistent treatment of model defects within the framework of Bayesian statistics. The next section discusses the principle of maximum entropy which gives theoretical support for the use of multivariate normal distributions instead of other distributions.

2.3 Principle of maximum entropy

The application of Bayesian statistics requires the specification of a prior probability distribution. However, existing prior knowledge is usually not sufficient to narrow down possible choices of the prior probability distribution to just one. This arbitrariness is not desirable, because Bayesian inference starting from different prior distributions leads to different results. Scientific methods for the acquisition of knowledge should be objective. The result of a scientific method should not be dependent on the person who applies it or the person's beliefs. This arbitrariness in the prior specification has been a common criticism of Bayesian statistics. Frequentist statistics does not suffer from this problem, but is too restrictive for the application in nuclear data evaluation. To alleviate the prior assignment problem, principles are introduced which remove the ambiguity. Ideally, these principles are of general nature and considered reasonable by everyone. The application of these principles restores objectivity in Bayesian inference. In this section, we first discuss the concept of information entropy and afterward the principle of maximum entropy introduced by Jaynes (1957a). The principle of maximum entropy gives theoretical support for the use of the multivariate normal distribution if only the means, variances and covariances of the random variables are known.

The concept of information entropy was introduced in a seminal papers of Claude Elwood Shannon (Shannon, 1948a,b). Given a set of possible events $\{E_i\}_{i=1:n}$ and for each event E_i a probability of its occurrence p_i , the measure of information entropy $H(p_1, p_2, \dots, p_n)$ quantifies the uncertainty about the outcome. Shannon demanded the information entropy to obey three properties, from

which he derived its functional form. We cite these properties here from Shannon (1948b, p. 10):

Characterization of the information entropy H

1. H should be continuous in the p_i .
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n . With equally likely events there is more choice, or uncertainty, when there are more possible events.
3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H . [...]

These general properties constrain the functional form of the information entropy up to a constant factor K . The information entropy is given by (Shannon, 1948b, p. 11)

$$H(p_1, \dots, p_n) = -K \sum_{i=1}^n p_i \log p_i. \quad (2.57)$$

The arbitrary factor K is due to the fact that the base of the logarithm can be chosen freely. It is a remarkable fact that the above stated rather general properties lead to an unambiguous functional form. Considering this, the information entropy appears to be a fundamental quantity to measure information.

The information entropy attains its largest value if all events have equal probability, $p_i = \frac{1}{n}$, which is $H = \log n$. The smallest value is attained if the probability for a single event is one and for all other events zero. Defining $0 \log 0 = 0$ due to $\lim_{p \rightarrow 0^+} p \log p = 0$, we obtain $H = 0$. The results of these limiting cases are in line with our intuitive understanding of the term uncertainty. If asked to predict the next event and all events are equally likely, we can only blindly guess. We are maximally uncertain about the outcome. However, if probabilities of some events are higher than others, we would choose some of the events with higher probability. The uncertainty is reduced. If only a single event has probability one and all others zero, it means we are absolutely certain about the outcome. Therefore, the information entropy takes its lowest possible value, which is zero.

The definition of information entropy in Equation 2.57 is for a discrete set of events only. Its generalization to an uncountable set of events—also given by Shannon—is called *differential entropy*. It is defined as

$$\mathcal{H}[X] = - \int p(x) \log p(x) dx, \quad (2.58)$$

where $p(x)$ is the probability distribution of the continuous random variable X . The differential entropy does not possess all the features of the discrete version. First, it can be negative, and second, it is not invariant under a transformation of the variable. However, the difference $\mathcal{H}_1(X) - \mathcal{H}_2(X)$ is still a useful measure to compare uncertainties. Two ways to restore the properties of the discrete version were suggested. The first way suggested by Jaynes (1957a) is the introduction of an invariant measure. The second way is the *Kullback-Leibler divergence* (Kullback and Leibler, 1951). The resulting definitions differ only by the sign. Here, we state the Kullback-Leibler divergence,

$$\mathcal{D}(p \parallel q) = \int p(x) \log \frac{p(x)}{q(x)} dx. \quad (2.59)$$

The function $q(x)$ has to be a normalized probability distribution and is called the *reference measure* (or reference probability distribution). The properties of the discrete definition are restored. Especially important for the remainder of the section, the positivity,

$$\mathcal{D}(p \parallel q) \geq 0. \quad (2.60)$$

Having laid out the definition of the information entropy and the differential entropy, we state the principle of maximum entropy.

Principle of maximum information entropy

From all possible probability distributions that are consistent to given constraints, one should pick the probability distribution that possesses maximum information entropy. Choosing any other probability distribution means to implicitly assume further knowledge one does not have.

The use of the multivariate normal distribution for the prior distribution in this thesis is justified by the principle of maximum entropy. Inserting its definition Equation 2.34 into Equation 2.58 yields the differential entropy of a multivariate normal random vector $\vec{X} = (X_1, X_2, \dots, X_d)^T$,

$$\mathcal{H}(\vec{X}) = \frac{1}{2} \ln ((2\pi e)^d |\mathbf{V}|). \quad (2.61)$$

For convenience, we stated the result in terms of the natural logarithm. The information entropy depends only on the determinant of the covariance matrix \mathbf{V} and its dimension d .

If only the means $E[X_i]$ and the covariances $\text{Cov}[X_i, X_j]$ of the random variables X_i are known, the distribution with maximum entropy is the multivariate normal distribution. Because of the importance of this result for this thesis, we give a proof here. Without loss of generality we can make a translation of the random

vector \vec{X} so that its mean vector is the zero vector. Further, we combine the covariances to the covariance matrix \mathbf{V} . As a first step of the proof, we expand the Kullback-Leibler divergence,

$$\mathcal{D}(p \parallel q) = \int p(\vec{x}) \log p(\vec{x}) d\vec{x} - \int p(\vec{x}) \log q(\vec{x}) d\vec{x}. \quad (2.62)$$

The first term on the right side is the negative differential entropy, see Equation 2.58. Assuming a zero-mean multivariate normal distribution with covariance matrix \mathbf{V} as reference distribution $q(\vec{x})$ and using Equation 2.60, we can write

$$\mathcal{H}_p(\vec{X}) \leq - \int p(\vec{x}) \left\{ -\frac{1}{2} \ln((2\pi)^d |\mathbf{V}|) - \frac{1}{2} \vec{x}^T \mathbf{V}^{-1} \vec{x} \right\} d\vec{x}. \quad (2.63)$$

Due to the required normalization of $p(x)$, we are able to evaluate the integral of the first term,

$$\mathcal{H}_p(\vec{X}) \leq \frac{1}{2} \left\{ \ln((2\pi)^d |\mathbf{V}|) + \mathbb{E}_p[\vec{x}^T \mathbf{V}^{-1} \vec{x}] \right\}. \quad (2.64)$$

The integral of the second term was rewritten in terms of the expectation operator. Using the fact that the trace of a scalar is the scalar itself, the cyclicity of the trace, and the commutative property of trace and expectation operator, we obtain for the second term

$$\mathbb{E}_p[\vec{x}^T \mathbf{V}^{-1} \vec{x}] = \text{Tr}[\mathbb{E}_p[\vec{x} \vec{x}^T] \mathbf{V}^{-1}] = \text{Tr}[\mathbf{V} \mathbf{V}^{-1}] = d. \quad (2.65)$$

The result is the dimension d of the covariance matrix. Inserting this result into Equation 2.64 gives

$$\mathcal{H}_p(\vec{X}) \leq \frac{1}{2} \left\{ \ln((2\pi)^d |\mathbf{V}|) + d \right\} = \mathcal{H}_q(\vec{X}). \quad (2.66)$$

Therefore, any other distribution $p(x)$ with the same mean vector and covariance matrix has lower information entropy than the multivariate normal distribution. The proof is completed.

We conclude this section by pointing out the geometric interpretation of the information entropy of the multivariate normal distribution. Figure 2.2 shows confidence regions of the bivariate normal distribution. Points associated with the same probability density form an ellipse. For a trivariate normal distribution, points with the same probability density form an ellipsoid. For even more variables, one obtains hyperellipsoids. In general, the volume of the d -dimensional hyperellipsoid spanned by the standard deviations δ_i along the principal axes of the covariance matrix \mathbf{V} is given by the expression

$$\mathcal{V}_d = \pi \prod_{i=1}^d \delta_i. \quad (2.67)$$

To construct a confidence region that includes the true vector with probability p , we additionally need the χ^2 -distribution (see section B.2). If $\chi_d^2(p)$ denotes the quantile function for probability p of the χ^2 -distribution with d degrees of freedom, the volume of the respective confidence region is $\mathcal{V}_d(p) = [\chi_d^2(p)]^{\frac{d}{2}} \mathcal{V}_d$.

We can compare the hyperellipsoid given in Equation 2.67 to the information entropy of the multivariate normal distribution given in Equation 2.61. For this purpose, we identically rewrite the information entropy,

$$\mathcal{H}(\vec{X}) = \ln \left((2\pi e)^{\frac{d}{2}} |\mathbf{V}|^{\frac{1}{2}} \right) = \ln \left(\prod_{i=1}^d \delta_i \sqrt{2\pi e} \right). \quad (2.68)$$

Here, we replaced $|\mathbf{V}|^{\frac{1}{2}}$ by the product of the square roots of the eigenvalues of the covariance matrix, $\sqrt{\lambda_i} = \delta_i$. Comparing Equation 2.67 and Equation 2.68, we recognize the following relation between the information entropy and the hyperellipsoid spanned by the standard deviations along the principal axes of the covariance matrix. If we inflate the latter mentioned hyperellipsoid by stretching each axis by a factor $\sqrt{2\pi e}$, divide the result by π , and take the logarithm, we obtain the information entropy.

Nuclear data evaluation

The aim of nuclear data evaluation is to produce best and consistent estimates and associated uncertainties of observables, especially reaction related ones for atomic nuclei throughout the nuclear charts (Chadwick et al., 2006). For instance, these observables may be angle-integrated cross sections, spectra, and angle-differential cross sections of various reaction channels. Knowledge about these observables is required for the design of novel nuclear facilities, and to assess their efficiency and safety. Two information sources are available to determine these observables: experimental data and nuclear models. Experimental data usually do not cover the whole range of incident energies that is of interest. Therefore, the predictions of nuclear models are required to fill the gaps. Bayesian statistics offers a well-founded framework to combine the experimental data with the results of model calculations to obtain estimates and uncertainties at all relevant incident energies.

We stated the general form of the Bayesian update formula in Equation 2.18. In order to make the distinction between prior distribution and likelihood apparent from the function symbol, we use from now on π for the prior pdf and ℓ for the likelihood function. Because the posterior can be used as the new prior distribution in another evaluation, we also use the function symbol π for the posterior pdf. Throughout this thesis, model parameters are denoted by \vec{p} and vectors of cross sections by $\vec{\sigma}$. Using the adjusted notation, the Bayesian update formula takes the form

$$\pi(\vec{p} | \vec{\sigma}, \mathcal{M}) = \frac{\ell(\vec{\sigma} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M})}{\int \ell(\vec{\sigma} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M}) d\vec{p}}. \quad (3.1)$$

The denominator represents the evidence. The given form is obtained by using $\pi(\vec{\sigma}, \vec{p} | \mathcal{M}) = \ell(\vec{\sigma} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M})$, which follows from the definition of conditional probability density in Equation 2.17, and then marginalizing over the model parameters, see Equation 2.16. The prior pdf $\pi(\vec{p} | \mathcal{M})$ expresses our prior knowledge

about the values of the model parameters before looking at the data. Parameter sets for which the prior pdf takes higher values are believed to be more likely than those with lower values. The likelihood $\ell(\vec{\sigma} | \vec{p}, \mathcal{M})$ gives the probability density to measure a certain realization of the observables in the experiment under the assumption that the true parameter vector is given by \vec{p} . The result of prior pdf times likelihood divided by the evidence yields the posterior pdf. The posterior pdf represents a refined knowledge about the model parameters by taking into account the experimental data.

All occurring probability density distributions are conditioned on \mathcal{M} , which indicates the systematics of the model. Therefore, results of the Bayesian inference are conditioned on the fact that the model is a perfect description of reality. In section 6.1, the consequences are worked out if this assumption does not hold. An extended form of the Bayesian update formula is presented in chapter 9 which accounts for possible deficiencies of the model. The fact that results of the Bayesian inference are restricted to the possibilities of the model are evident in Equation 3.1, because the refined probability distribution refers to the model parameters. However, in section 3.3 the Bayesian update formula is formulated at the level of cross sections, which camouflages the conditioning on the model systematics.

The evaluation of the likelihood $\ell(\vec{\sigma} | \vec{p}, \mathcal{M})$ requires nuclear model with the parameter vector \vec{p} . Throughout this thesis we exclusively use the nuclear model code TALYS (Koning, Hilaire, and Duijvestijn, 2008). TALYS implements a variety of nuclear models, such as the optical model for elastic scattering and direct reactions, and the exciton model for pre-equilibrium reactions. Depending on the incident energy of the projectile and the mass of the target nucleus, computation times may range from minutes to hours. In general, due to the computation time of the model code, the computation of the exact posterior pdf is difficult. One remedy is to linearize the relation between parameters and observables. This approach is detailed in section 3.2. Another approach is to construct a surrogate model that can be evaluated much faster than the original model. This approach is detailed in section 3.3. In both approaches, the prior pdf is specified as a multivariate normal distribution, either at the level of parameters or at the level of cross sections. Yet another approach to deal with the complexity of the nuclear model is to apply Monte Carlo procedures, such as the Bayesian Monte Carlo method (Koning, 2015).

Once the observables of the experiment are predicted by a model calculation, the likelihood takes the same form in each approach. The next section elaborates on the construction of the likelihood function based on the available data.

3.1 Accounting for experimental data

The concept of the likelihood is used in both Bayesian statistics and frequentist statistics. The likelihood gives the probability to obtain a certain outcome, such as a measured value of a cross section, under the assumption that a certain set of model parameters \vec{p} is the true one. The underlying statistical model has the form,

$$\vec{\sigma} = \mathcal{M}(\vec{p}) + \vec{\varepsilon}, \quad (3.2)$$

where $\vec{\sigma}$ denotes the experimental measurement, $\mathcal{M}(\vec{p})$ the model prediction based on the model parameters \vec{p} , and $\vec{\varepsilon}$ the measurement error. Consequently, in the absence of measurement errors, the model prediction of observables using the true model parameter vector \vec{p}_{true} should perfectly match with corresponding measurements.

The specification of the probability density distribution for the likelihood is dependent on the assumptions made for the measurement error $\vec{\varepsilon}$. In some cases, the physical process determines the appropriate distribution. For instance, the number of decay events in a radioactive material occurring in a given time span is best described by a Poisson distribution. In other cases, it might be not clear which distribution to choose. If the choice is not clear, there are several reasons to use a (multivariate) normal distribution. First, many distributions such as the binomial or the Poisson distribution, have as limiting distribution the normal distribution. Second, if the measurement error is a superposition of many independent contributions, their sum follows a normal distribution according to the central limit theorem (e.g. Vaart (1998)). Third, if only mean values and variances are known, the principle of maximum entropy leads to a normal distribution. In addition, if also correlations between the observables are given, a multivariate normal distribution should be used. Because of these arguments, we assume pdf of the measurement error to be given by a multivariate normal distribution. This assumption leads to the likelihood

$$\ell(\vec{\sigma} | \vec{p}, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{B}|}} \exp \left\{ -\frac{1}{2} (\vec{\sigma} - \mathcal{M}(\vec{p}))^T \mathbf{B}^{-1} (\vec{\sigma} - \mathcal{M}(\vec{p})) \right\}. \quad (3.3)$$

The matrix \mathbf{B} is the covariance matrix of the experiment, and d is the dimension of the vector $\vec{\sigma}$ containing the measurement results. The nuclear model $\mathcal{M}(\vec{p})$ yields the model predictions of the observables given in $\vec{\sigma}$ and determines the center of the distribution.

We want to make plausible which considerations help to determine the variances and the covariances contained in the covariance matrix \mathbf{B} . We restrict the further discussion to a single reaction channel. Suppose that the vector of measurement

results is given by

$$\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)^T, \quad (3.4)$$

where an element σ_i represents the cross section of the i^{th} -experimental data point measured at incident energy E_i . Generally, measurements are subject of two types of uncertainties: statistical uncertainties and systematic errors. The total uncertainties associated with the measurement is the sum of these two contributions,

$$\vec{\varepsilon}_{\text{tot}} = \vec{\varepsilon}_{\text{stat}} + \vec{\varepsilon}_{\text{sys}}. \quad (3.5)$$

For both contributions, a covariance matrix has to be specified. The sum of them yields the experimental covariance matrix, $\mathbf{B} = \mathbf{B}_{\text{stat}} + \mathbf{B}_{\text{sys}}$. We first consider the construction of the covariance matrix \mathbf{B}_{stat} for the statistical uncertainties. Each diagonal element $B_{\text{stat},ii}$ represents the variance of the measured cross section σ_i associated with the i^{th} experimental data point due to the statistical uncertainty. Off-diagonal elements are zero. The magnitudes of the variances are determined by the measurement process.

As an example, we briefly outline a schematic measurement process and its statistical treatment. A cross section at a specific incident energy may be measured by leading an incident beam through the target material and measure the reduction of intensity of the outgoing flux in forward direction. Let N_{inc} denote the average number of incident particles entering the target material each second. Further, let N_{out} denote the average number of incident particles just passing through the target material each second without being scattered. The respective cross section is given by

$$\sigma = c \frac{N_{\text{inc}} - N_{\text{out}}}{N_{\text{inc}}}. \quad (3.6)$$

The factor c subsumes other quantities that determine the cross section, such as the density of the target material. They are not relevant for the point to be made. The scattering process itself is not deterministic. In the short run, the number of incident particles each second deviates from the average N_{inc} . Furthermore, whether a certain incident particle hits upon a nucleus in the target material is a matter of chance. We are dealing with a random process. We assume that we know the value of N_{inc} . The average number of incident particles just passing through each second, N_{out} , is what we have to infer to obtain the cross section. The number of counts $n(T)$ of incident particles passing through without interaction in a time period T follows a Poisson distribution. The probability to measure k counts in the time span T is given by

$$\Pr[n(T) = k] = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{with } \lambda = TN_{\text{out}}. \quad (3.7)$$

Expectation and variance of the Poisson process are

$$E[n(T)] = \lambda \quad \text{and} \quad \text{Var}[n(T)] = \lambda. \quad (3.8)$$

The distribution parameter λ indicates the number of expected counts when measuring for the duration T . Therefore, an estimate of N_{out} and the associated variance δ^2 based on the measured number of counts k is

$$\hat{N}_{\text{out}} = \delta^2 = \frac{k}{T}. \quad (3.9)$$

When extending the measurement time, the parameter λ increases as well. For λ large enough (larger than twenty), the Poisson distribution approaches the normal distribution. Therefore, the assumption of a multivariate normal distribution for the likelihood seems justified. Equation 3.6 is a linear relation between N_{out} and the cross section σ . Therefore, because of Equation 2.29, their variances are related by

$$\text{Var}[\sigma] = \frac{c^2}{N_{\text{inc}}^2} \text{Var}[N_{\text{out}}]. \quad (3.10)$$

Given the number of counts k_i and the measurement time T_i at the incident energy E_i of each experimental data point, the variance of N_{out} of each experimental data point can be estimated with Equation 3.9. Using Equation 3.10, the diagonal elements of the covariance matrix \mathbf{B}_{stat} for the cross sections can be determined. Off-diagonal elements are zero. The zero off-diagonals mean that the multivariate normal distribution can be factorized into a product of univariate normal distributions—one univariate normal distribution for each experimental data point. This error component is called statistical error and can easily be interpreted as a consequence of the inherent randomness of the physical process. The only possibility to reduce this error component is by increasing measurement time or using incident beams with a higher flux density. However, it is impossible to completely eliminate the statistical error.

In contrast to statistical errors, the systematic errors may be completely eliminated, at least in theory. They are the consequence of imperfect knowledge about the experimental setup. The worst case is to be unaware of them. Then, what is actually measured deviates from what should be measured. Large measurement times may reduce the statistical error to an extent that the true value is not contained within the calculated confidence interval anymore due to the systematic error. The best case is to be aware of the systematic error and to be able to correct the measurement results accordingly. The case between these two extreme positions is to be aware of the systematic error but to be unable to precisely quantify its magnitude. We give a schematic example for the latter situation.

The relation between the cross section and the counting rate N_{out} was sketched in Equation 3.6. An estimate of N_{out} was given in Equation 3.9. However, not every particle hitting the detector is detected. The proportion of detected particles to all particles hitting the detector is referred to as *detector efficiency* η . The relation between the detected counts and the true number of counts is

$$n_{\text{true}}(T) = \frac{n_{\text{det}}(T)}{\eta}. \quad (3.11)$$

Therefore, the cross section σ_i at incident energy E_i computed from the counts without taking into account the detector efficiency has to be corrected. The corrected cross section is given by $\sigma_{\text{cor},i} = \sigma_i/\eta$. If the detector efficiency is not perfectly known, we have to account for its uncertainty in \mathbf{B}_{sys} . Assuming independence of the detector efficiency from the incident energy and the count rate¹, we may model the uncertainty about the detector efficiency η as a normal distribution,

$$\eta \sim \mathcal{N}(\eta_0, \tau^2). \quad (3.12)$$

The distribution parameter η_0 represents a best estimate of the efficiency and the standard deviation τ reflects the trust in this estimate.

The detector efficiency affects the measurement result at each incident energy E_i in the same way. This fact is reflected in non-zero off-diagonal elements $B_{\text{sys},ij}$ in the covariance matrix \mathbf{B}_{sys} for the systematic error. The covariances are

$$\text{Cov}[\sigma_{\text{cor},i}, \sigma_{\text{cor},j}] = \text{Cov}\left[\frac{\sigma_i}{\eta}, \frac{\sigma_j}{\eta}\right] \approx \text{Cov}\left[\frac{\eta - \eta_0}{\eta_0^2}, \frac{\eta - \eta_0}{\eta_0^2}\right] = \frac{1}{\eta_0^4} \text{Var}[\eta] = \frac{\tau^2}{\eta_0^4}. \quad (3.13)$$

This result is based on a first order Taylor expansion of σ_i/η with respect to η . Due to Equation 2.28 the constant zero-order term σ_i/η_0 does not show up in the covariance. Noteworthy, all elements of \mathbf{B}_{sys} take the same value, which is τ^2/η_0^4 . Thus, the correlation between all experimental points are one if only taking into account the covariance matrix \mathbf{B}_{sys} . However, the correlations in the experimental covariance matrix $\mathbf{B} = \mathbf{B}_{\text{sys}} + \mathbf{B}_{\text{stat}}$ are smaller due to the uncorrelated statistical uncertainty modeled with \mathbf{B}_{stat} . In general, the larger the statistical error compared to the systematic error, the lower the correlations between the experimental data points.

This section dealt with the specification of the likelihood in the Bayesian update formula. Once the predictions of the model are obtained, the computation of the likelihood can be quickly performed. The evaluation of the posterior probability density function, however, is difficult. In order to calculate the posterior pdf

¹This assumption is made for the sake of simplicity. Usually, the efficiency will depend on the energy and sometimes also on the count rate.

for several parameter vectors \vec{p} , the model has to be evaluated for each model parameter vector of the statistical ensemble. The next section presents one way to deal with this situation, which is the linearization of the model.

3.2 Bayesian update of parameters

The Bayesian update formula for nuclear data evaluation was given in Equation 3.1. In section 3.1 we argued the choice of a multivariate normal distribution for the likelihood. The two missing pieces to apply the Bayesian update formula are the evidence and the prior distribution. The evidence represents a normalization constant for the posterior distribution and the calculation of its value is not required, as will be seen in this section. If our prior knowledge about the model parameters is given only by best estimates and associated variances and covariances, we should choose—according to the principle of maximum entropy—a multivariate normal distribution. Therefore, we introduce the multivariate normal distribution as the prior distribution for the model parameters,

$$\pi(\vec{p} | \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{A}_0|}} \exp \left\{ -\frac{1}{2} (\vec{p} - \vec{p}_0)^T \mathbf{A}_0^{-1} (\vec{p} - \vec{p}_0) \right\}. \quad (3.14)$$

The vector \vec{p}_0 represents the best apriori estimate of the model parameters, the covariance matrix \mathbf{A}_0 contains associated variances and covariances, and d is the number of model parameters contained in \vec{p}_0 .

Even though, both likelihood and prior are given as multivariate normal distributions, the posterior distribution is still difficult to evaluate. This is due to the fact that the evaluation of the likelihood $\ell(\sigma | \vec{p}, \mathcal{M})$ involves the nuclear model. Frequently, nuclear model calculations require solutions of differential equations. Thus, a relationship between physical observables and the model parameters does not exist in general. Furthermore, depending on the requested incident energy, model calculations may be very time demanding.

One possible solution is to linearize the nuclear model, which means to use a first order Taylor expansion instead of the true model,

$$\mathcal{M}_{\text{lin}}(\vec{p}) = \vec{\sigma}_{\text{ref}} + \mathbf{S}(\vec{p} - \vec{p}_{\text{ref}}) \quad \text{with} \quad \mathbf{S} = \left. \frac{\partial \mathcal{M}(\vec{p})}{\partial \vec{p}} \right|_{\vec{p}=\vec{p}_{\text{ref}}}. \quad (3.15)$$

The vector $\vec{\sigma}_{\text{ref}}$ is the result of the exact model prediction $\mathcal{M}(\vec{p}_{\text{ref}})$ at the expansion point \vec{p}_{ref} . The sensitivity matrix \mathbf{S} represents the Jacobian matrix of the nuclear model at the expansion point. Noteworthy, using the center vector \vec{p}_0 of the prior distribution as the expansion point is a reasonable choice, but not a necessity.

Replacing the original model by the linearized model in the likelihood, Equation 3.3, yields

$$\ell(\vec{\sigma} | \vec{p}, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{B}|}} \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p}) \right\} \quad (3.16)$$

with

$$\vec{\sigma}_{\text{eff}} = \vec{\sigma} - \vec{\sigma}_{\text{ref}} + \mathbf{S}\vec{p}_{\text{ref}}. \quad (3.17)$$

The matrix \mathbf{B} is the covariance matrix of the experiment, the vector $\vec{\sigma}$ contains the measured observables, and n is its dimension.

We have to calculate the product of likelihood and prior to get the posterior. The following consideration is helpful to achieve this goal. Technically, the likelihood gives the probability density to obtain a certain measurement $\vec{\sigma}$ if the true values of the observables are given by the model prediction. However, the exponent is symmetric concerning the measurement and the model prediction. Swapping $\vec{\sigma}_{\text{eff}}$ and $\mathbf{S}\vec{p}$ in Equation 3.16 does not alter the result. Therefore, we may formally interpret this equation as a multivariate normal distribution for the model parameter vector \vec{p} . We accept the broken normalization for the moment.

How to find the center vector \vec{p}_ℓ and the covariance matrix \mathbf{B}_ℓ of this multivariate normal distribution? We use ℓ as subscript to indicate that \vec{p}_ℓ and \mathbf{B}_ℓ are the distribution parameters for the model parameter vector associated with the likelihood. We can expand the relevant part of the exponent of a multivariate normal distribution with center vector \vec{p}_ℓ and covariance matrix \mathbf{B}_ℓ ,

$$\begin{aligned} (\vec{p} - \vec{p}_\ell)^T \mathbf{B}_\ell^{-1} (\vec{p} - \vec{p}_\ell) &= \vec{p}^T \mathbf{B}_\ell^{-1} \vec{p} + \vec{p}^T \mathbf{B}_\ell^{-1} \vec{p}_\ell \\ &\quad + \vec{p}_\ell^T \mathbf{B}_\ell^{-1} \vec{p} + \vec{p}_\ell^T \mathbf{B}_\ell^{-1} \vec{p}_\ell. \end{aligned} \quad (3.18)$$

Carrying out the same expansion for the exponent in Equation 3.16 yields

$$\begin{aligned} (\mathbf{S}\vec{p} - \vec{\sigma}_{\text{eff}})^T \mathbf{B}^{-1} (\mathbf{S}\vec{p} - \vec{\sigma}_{\text{eff}}) &= \vec{p}^T \mathbf{S}^T \mathbf{B}^{-1} \mathbf{S} \vec{p} + \vec{p}^T \mathbf{S}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{eff}} \\ &\quad + \vec{\sigma}_{\text{eff}}^T \mathbf{B}^{-1} \mathbf{S} \vec{p} + \vec{\sigma}_{\text{eff}}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{eff}}. \end{aligned} \quad (3.19)$$

Comparing the first terms on the r.h.s. of Equation 3.18 and Equation 3.19, we identify

$$\mathbf{B}_\ell^{-1} = \mathbf{S}^T \mathbf{B}^{-1} \mathbf{S}. \quad (3.20)$$

Comparing the second terms on the r.h.s. of Equation 3.18 and Equation 3.19, we identify

$$\mathbf{B}_\ell^{-1} \vec{p}_\ell = \mathbf{S}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{eff}} \Rightarrow \vec{p}_\ell = \mathbf{B}_\ell \mathbf{S}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{eff}}. \quad (3.21)$$

We obtained the center vector \vec{p}_ℓ and the covariance matrix \mathbf{B}_ℓ for the parameter vector \vec{p} in the likelihood. Inserting found expressions again into Equation 3.18, we realize that the 4th term on the r.h.s. differs from the corresponding term

in Equation 3.19. This constant deviation in the exponent represents merely a rescaling of the multivariate normal distribution, but does not alter its functional form.

Because both the prior given in Equation 3.14 and the likelihood given in Equation 3.16 are multivariate normal distributions, also the posterior is a multivariate normal distribution. Using Equation 2.54, we obtain the center vector \vec{p}_1 and the covariance matrix \mathbf{A}_1 of the posterior,

$$\mathbf{A}_1 = (\mathbf{A}_0^{-1} + \mathbf{B}_\ell^{-1})^{-1} = (\mathbf{A}_0^{-1} + \mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1}, \quad (3.22)$$

$$\vec{p}_1 = \mathbf{A}_1 (\mathbf{A}_0^{-1} \vec{p}_0 + \mathbf{B}_\ell^{-1} \vec{p}_\ell) = \mathbf{A}_1 (\mathbf{A}_0^{-1} \vec{p}_0 + \mathbf{S}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{eff}}). \quad (3.23)$$

The application of the Woodbury matrix identity (Woodbury, 1950) (see section B.3) allows to rewrite these formulas to

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0 \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} \mathbf{S} \mathbf{A}_0, \quad (3.24)$$

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}_0 \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} (\vec{\sigma}_{\text{eff}} - \mathbf{S} \vec{p}_0). \quad (3.25)$$

In general, the Bayesian update formulas given by Equation 3.24 and Equation 3.25 appear to be numerically preferable to those in Equation 3.22 and Equation 3.23 because only one inversion on the grid of the experimental data is required.

SAMMY (Larson, 1998), a popular computer code for making R-matrix fits, uses both versions of the update formulas. In the diction of the SAMMY manual, the first block here corresponds to the (M+W) inversion scheme and the second block to the (N+V) inversion scheme. SAMMY was originally published in 1984, when computer power was very limited. Under the assumption that the covariance matrices for both the experimental data and the model parameters are diagonal, the update formulas Equation 3.22 and Equation 3.23 are beneficial because the inversion of both covariance matrices is trivial and the final more involved inversion is performed on the low dimensional space of model parameters. However, using a diagonal experimental covariance matrix means to neglect systematic errors of the experiment; or to only keep their diagonal contribution and thereby throwing away information.

The linearization of the nuclear model lead to a multivariate normal posterior distribution. Because we know the functional form of a properly normalized multivariate normal distribution, the computation of the evidence is not needed. In general, the purpose of the evidence is to take care of the proper normalization of the posterior distribution.

Two possible issues have to be considered when performing the Bayesian update procedure with the linearized model. For describing the first issue, assume that the experimental data suggest a parameter vector \vec{p} far away from the parameter

vector \vec{p}_0 at which the nuclear model had been linearized. The updated parameter vector \vec{p}_1 may substantially deviate from the result using the exact nuclear model. This situation is shown in the left diagram of Figure 3.1. A possible solution of this problem is an iterative evaluation of the Bayesian update formula. In each iteration step using the updated parameter vector \vec{p}_1 from the previous iteration as expansion point \vec{p}_{ref} . The iteration may continue until convergence is achieved. The second issue concerns evaluated confidence intervals that significantly differ from those resulting from using the exact nuclear model. This issue is illustrated in the right part of Figure 3.1.

In general, however, the Bayesian update formula with the linearized model can be expected to produce reasonable results if two criteria are met: The expansion point of the linearized model is close to the maximum of the posterior distribution, and the experimental data have small uncertainties, thereby narrowing down the uncertainties of the parameters to the linear domain.

One remaining question is the choice of proper correlations between model parameters in the prior covariance matrix. If mathematical relations between model parameters are known, these relations determine the correlations to use. If we

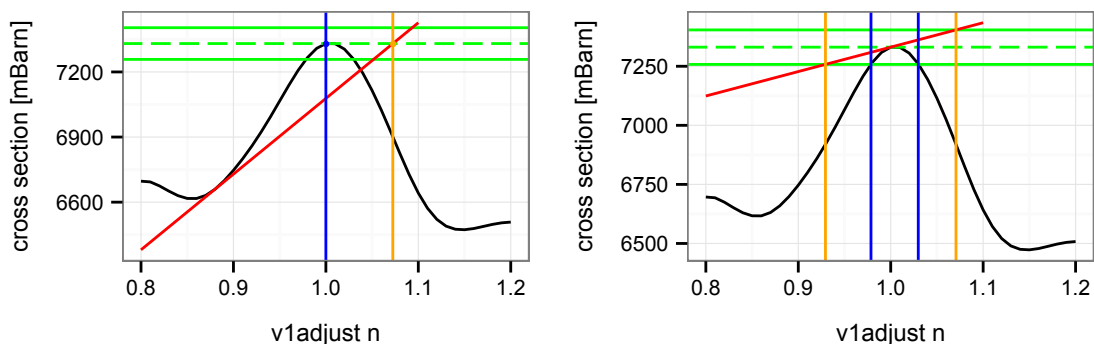


Figure 3.1: Both diagrams show the total cross section of ^{181}Ta at 2 MeV as a function of the optical model parameter v_1 . The parameter v_1 is given relative to the default value of TALYS. The dashed green line indicates a hypothetical experimental measurement and the green solid lines indicate the corresponding experimental confidence interval. Each diagram illustrates a possible issue of the Bayesian update when working with a linearized model. Left diagram: the model is linearized around $v_1 = 0.88$. The red line shows the systematics of the linearized model. After the update with experimental data, the new best estimate is $v_1 \approx 1.07$, marked by the orange line. However, the new best estimate obtained by using the exact model would be $v_1 = 1$, marked by the blue line. Right diagram: the model is linearized around the optimal value $v_1 = 1$. The red line corresponds to the linearized model. The updated confidence interval for v_1 , indicated by the orange lines, is larger than that obtained by using the exact model, indicated by the blue lines.

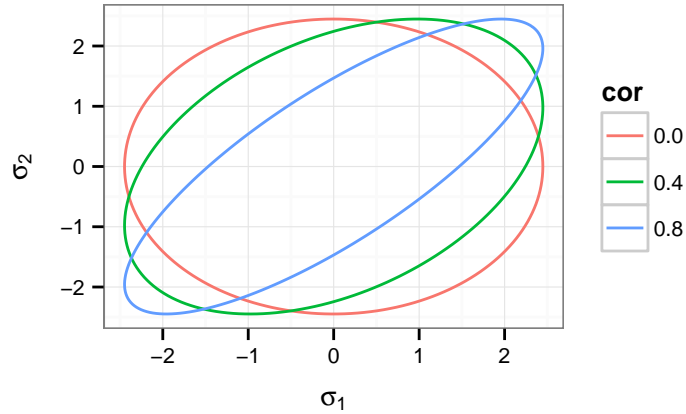


Figure 3.2: 95% confidence ellipses of bivariate normal distributions with $\text{Var}[\sigma_1] = \text{Var}[\sigma_2] = 1$. The depicted bivariate normal distributions differ by the assumed correlation between the two variables. With increasing correlation, the confidence ellipses become thinner and more tilted. Also, the area surrounded by the ellipse gets smaller. This is the geometric expression of the fact that correlations lower the determinant of the covariance matrix.

are not aware of such relations or they simply do not exist, the principle of maximum entropy demands to treat the model parameters as uncorrelated. To justify this statement, the differential entropy of the multivariate normal distribution is needed. Recalling its definition from Equation 2.61, we realize that its value is only dependent on the determinant and the dimension of the covariance matrix. A special case of Hadamard's inequality (Hadamard, 1893) for positive semi-definite matrices, such as covariance matrices, is

$$|\mathbf{V}| \leq \prod_{i=1}^d V_{ii}. \quad (3.26)$$

This inequality states that the product of the diagonal elements V_{ii} is always greater than or equal the determinant of the matrix. If the model parameters are uncorrelated, off-diagonal elements of the covariance matrix \mathbf{V} are zero. In this case, the determinant is exactly given by the product of the diagonal elements, which yields the largest possible determinant according to Hadamard's inequality. The introduction of off-diagonal elements, thereby correlating parameters, can only lower the value of the determinant. Therefore, the maximal differential entropy is attained if parameters are uncorrelated.

As pointed out at the end of section 2.3, the determinant of the covariance matrix is also related to the confidence hyperellipsoids of the multivariate normal distribution. The change of its shape with increasing correlations for a bivariate normal distribution is displayed in Figure 3.2.

3.3 Bayesian update of observables

Linearizing the model around a reference set of model parameters has the benefit that it leads to analytic matrix formulas for the Bayesian update, as has been discussed in section 3.2. In contrast, using the exact model requires the application of Monte Carlo procedures, which are computationally involved. However, the linearization of the model has two drawbacks:

1. If the reference point for the linearization is far away from the optimal parameter vector according to model prior and experimental data, the Bayesian update using the linearized model may yield a bad approximation of the best parameter vector. The magnitude of this effect depends on the non-linearity of the model.
2. If experimental evidence does not narrow down parameter uncertainties to a sufficiently linear domain of the model, evaluated uncertainties may be misleading.

Both issues were illustrated in Figure 3.1 of the last section. In this section, we present another way of linearizing the model, which follows the maxim *to be on the safe side*.

The probability density functions chosen for the likelihood and the prior have been discussed in section 3.1 and section 3.2. For reference, we repeat the functional form of the model prior here. The model prior is given by

$$\pi(\vec{p} | \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{A}_0|}} \exp \left\{ -\frac{1}{2} (\vec{p} - \vec{p}_0)^T \mathbf{A}_0^{-1} (\vec{p} - \vec{p}_0) \right\}, \quad (3.27)$$

with the prior best estimate \vec{p}_0 and the associated covariance matrix \mathbf{A}_0 . The number of model parameters contained in \vec{p}_0 is denoted by d .

In order to make the Bayesian update formula analytically tractable, the original nuclear model may be replaced by a simple surrogate model. This surrogate model is constructed in the following way. Parameter vectors \vec{p}_i are sampled from the prior distribution given in Equation 3.27. The nuclear model is applied for each sampled parameter vector to get the model prediction, $\vec{\sigma}_i = \mathcal{M}(\vec{p}_i)$. Then, a multivariate normal distribution is constructed based on the model predictions $\vec{\sigma}_i$. This requires the specification of a center vector and a covariance matrix. There are two possibilities to specify the center vector. The first possibility is to evaluate the nuclear model for the center vector \vec{p}_0 of the prior distribution. Then, the center vector at the level of physical observables is given by $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$. The second possibility is to get the center vector $\vec{\sigma}_0$ by computing the arithmetic

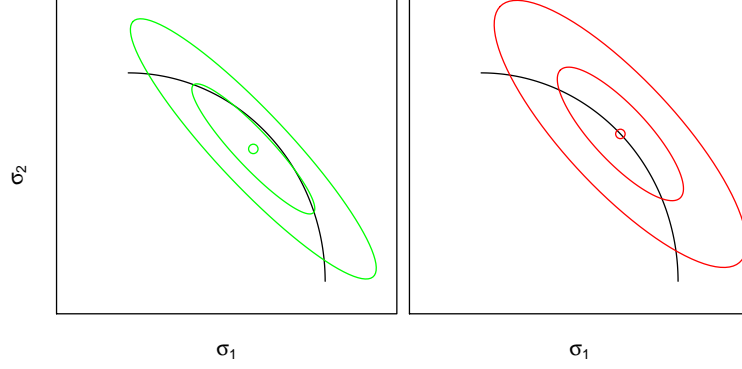


Figure 3.3: The dependence of the covariance matrix on the chosen center point $\vec{\sigma}_0$. The black curve indicates predictions of cross sections σ_1 and σ_2 by a hypothetical model. In the left diagram, the center is given by the arithmetic mean. This choice leads to a covariance matrix which is visualized by the green confidence ellipses. In the right diagram, the center is given by $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$. The confidence ellipses of the associated covariance matrix are drawn in red. Due to the minimum variance property of the arithmetic mean, the left confidence ellipses enclose a smaller area. However, the center point is not located on the model curve.

mean of the drawn samples $\{\vec{p}_i\}_{i=1:m}$,

$$\vec{\sigma}_0 = \frac{1}{m} \sum_{i=1}^m \vec{\sigma}_i. \quad (3.28)$$

Having determined the center vector $\vec{\sigma}_0$, the covariance matrix can be computed by

$$\tilde{\mathbf{A}}_0 = \frac{1}{n-1} \sum_{i=1}^m (\vec{\sigma}_i - \vec{\sigma}_0)(\vec{\sigma}_i - \vec{\sigma}_0)^T. \quad (3.29)$$

The computation of the covariance matrix is dependent on the center vector $\vec{\sigma}_0$. The advantage of using the arithmetic mean for $\vec{\sigma}_0$ is that this choice leads to minimal variance. The disadvantage may be that the vector $\vec{\sigma}_0$ represents a collection of values for the observables, which cannot be reached by any choice of model parameters. The vector $\vec{\sigma}_0$ would be outside the possibilities of the model. These arguments are illustrated in Figure 3.3. In this thesis, we consider the use of $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$ as the center vector to be the better option due to the latter reason.

The result of the outlined procedure is a statistical surrogate model $\mathcal{M}_{\text{sur}}(\vec{\sigma}_{\text{mod}})$, which replaces the original nuclear model in the Bayesian update procedure. Using the surrogate model, the model prior takes the form

$$\pi(\vec{\sigma}_{\text{mod}} | \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^n |\tilde{\mathbf{A}}_0|}} \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{mod}} - \vec{\sigma}_0)^T \tilde{\mathbf{A}}_0^{-1} (\vec{\sigma}_{\text{mod}} - \vec{\sigma}_0) \right\}, \quad (3.30)$$

Noteworthy, the probability distribution of the surrogate model is specified at the level of physical observables, such as cross sections. The model parameters

of the surrogate model are the physical observables. Hence, the connection to model parameters is lost. The new evaluation scheme to tackle a large number of observables developed in this thesis recovers this broken link. It is presented in the chapter 5.

At this point it is important to remark that the modified prior distribution refers to physical observables specified at the energies of the model mesh. For instance, the model calculation may yield the cross sections at incident energies from 1 to 20 MeV in steps of 1 MeV. In contrast, the vector of experimental measurements may contain the cross sections at incident energies in-between these mesh points. Therefore, to be more precise, we refer with $\vec{\sigma}_{\text{exp}}$ to the observables measured in the experiment and with $\vec{\sigma}_{\text{mod}}$ to the observables given at the incident energies of the model mesh. We assume that the interpolation of the model predictions to the incident energies of the experiment can be expressed as a linear transformation,

$$\vec{\sigma}_{\text{int}} = \mathcal{M}_{\text{sur}}(\vec{\sigma}_{\text{mod}}) = \mathbf{S}\vec{\sigma}_{\text{mod}}, \quad (3.31)$$

where \mathbf{S} is the transformation matrix. The construction of this matrix is discussed in section 3.4. Because of using the surrogate model with the mapping rule in Equation 3.31, the likelihood introduced in Equation 3.3 is modified to

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{\sigma}_{\text{mod}}, \mathcal{M}_{\text{sur}}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{B}|}} \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\sigma}_{\text{mod}})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\sigma}_{\text{mod}}) \right\}, \quad (3.32)$$

where \mathbf{B} is the experimental covariance matrix, and n is the number of experimental data points. The structure of the prior in Equation 3.30 and the likelihood in Equation 3.32 is identical to the forms in Equation 3.14 and Equation 3.16 presented in section 3.2. Thus, we can reuse the Bayesian update formulas given in Equation 3.24 and Equation 3.25 with variable names substituted appropriately,

$$\tilde{\mathbf{A}}_1 = \tilde{\mathbf{A}}_0 - \tilde{\mathbf{A}}_0 \mathbf{S}^T \left(\mathbf{S} \tilde{\mathbf{A}}_0 \mathbf{S}^T + \mathbf{B} \right)^{-1} \mathbf{S} \tilde{\mathbf{A}}_0, \quad (3.33)$$

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \tilde{\mathbf{A}}_0 \mathbf{S}^T \left(\mathbf{S} \tilde{\mathbf{A}}_0 \mathbf{S}^T + \mathbf{B} \right)^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\sigma}_0). \quad (3.34)$$

Applying these update formulas yields a modified center $\vec{\sigma}_1$ and covariance matrix $\tilde{\mathbf{A}}_1$ for the multivariate normal distribution of the surrogate model parameters $\vec{\sigma}_{\text{mod}}$.

It is important to consider the role of the surrogate model in more detail. Its essential benefit is the convenient numerical treatment due to the simplification of the original model. The surrogate model approximates relationships of observables, such as cross sections at different incident energies, by linear relationships. The error thereby made due to non-linearities is taken into account as additional uncertainty. The situation is shown for a schematic example in Figure 3.4. The

fact that non-linear features of the original model are lost in the surrogate model leads to lower predictive power compared to the original model. For instance, a precise measurement at one incident energy may suffice to predict accurately the cross section at another energy using the original model. The surrogate model, in contrast, is limited to accounting for linear trends. Therefore, the uncertainties in the predictions are always higher compared to the original model. Simply said, the uncertainties due to non-linearities are just added to the uncertainties of the prediction.

The loosening of the model systematics can be either an advantage or a disadvantage. If the original model is a very good description of reality, it is not desirable to throw away some of its features. However, if the original model is not that good but still useful, one might favor using the surrogate model. General trends of the models are still captured, but details are thrown away. This opens up the possibility that observables resulting from a Bayesian update reasonable resemble included experimental data, even though the original model would not be able to describe the data.

To conclude this section, Figure 3.5 and Figure 3.6 give impressions of the non-linearities of the models used in TALYS to predict the neutron-induced total cross section of ^{181}Ta at low incident energies. Only the three most sensitive optical model parameters r_v , a_v and v_1 are taken into account.

The next section elaborates on the specification of the transformation matrix to map the predictions from the model mesh to the experimental observables.

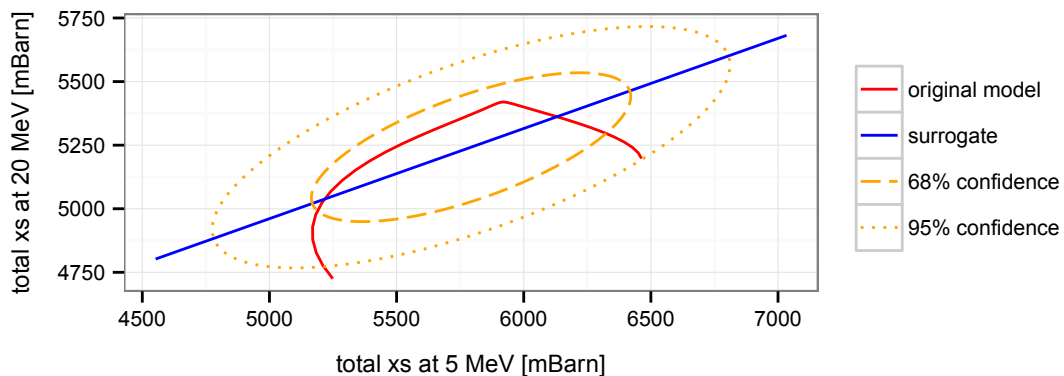


Figure 3.4: Visualization of the surrogate model. The original model (red) is replaced by a linear surrogate model (blue). Non-linear features of the original model are captured as additional uncertainties in the surrogate model.

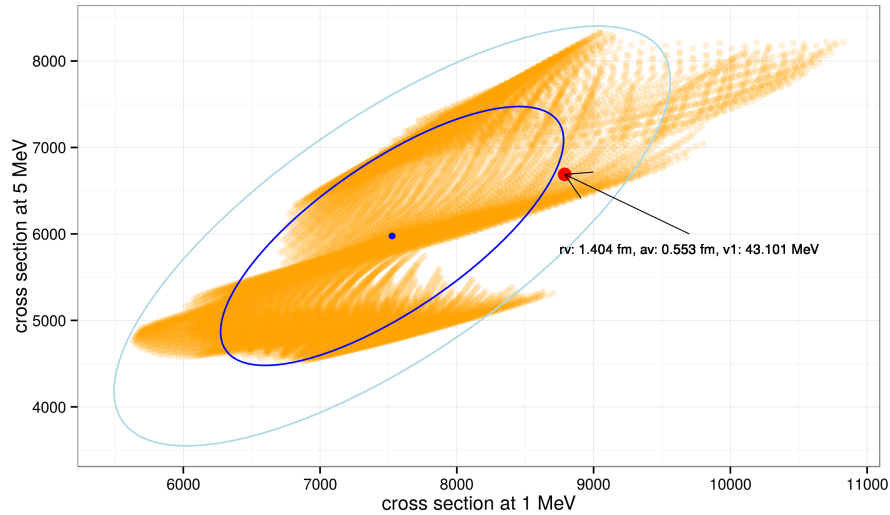


Figure 3.5: The cloud of orange points visualizes the accessible combinations of values (in millibarn) for the neutron-induced total cross section of ^{181}Ta at 1 MeV and 5 MeV that can be reached by varying the optical model parameters r_v , a_v , and v_1 . The interval of variation for each model parameter relative to the TALYS default value is given by $[0.8, 1.2]$. The surrogate model approximates this cloud of model possibilities by an ellipse. White regions not accessible by the original model are possible results if using the surrogate model. The systematics of the original model are loosened.

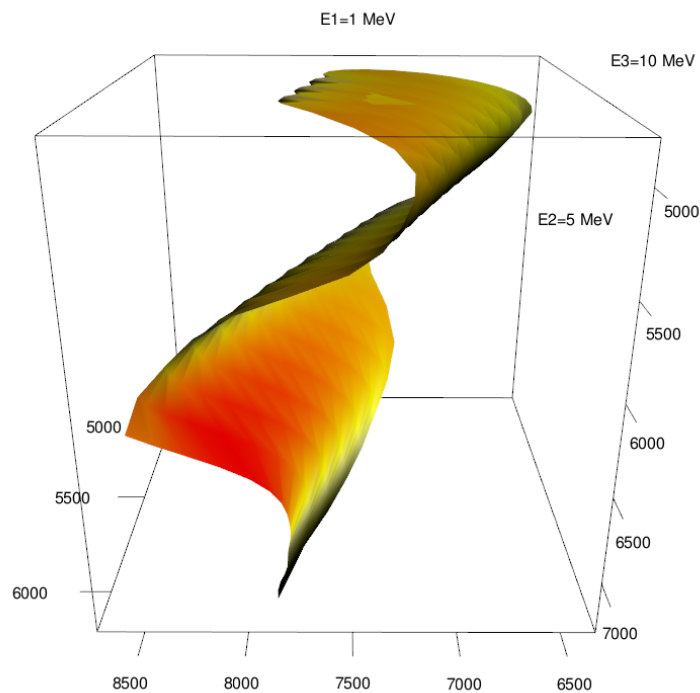


Figure 3.6: Visualization of the non-linear model systematics. The axes denote the neutron-induced total cross section (in millibarn) of ^{181}Ta at 1, 5, and 10 MeV. The depicted manifold visualizes possible combinations of the cross sections when varying optical model parameters r_v and v_1 up to 20% around TALYS default values. The model parameter a_v is set to the TALYS default.

3.4 Mapping model data to experiment

Model calculations require the specification of energies and angles at which the values of observables should be computed. Two approaches are available to map the calculated observables at the mesh points to the energies and angles of the experimental data.

The first approach uses mesh points that coincide with the energies and angles of the experimental data. If only few experimental data points are involved, this approach can considerably speed up the search for the best set of model parameters because usually the time needed for a model calculation depends on the chosen mesh size. However, the comparison of calculated results with experimental data that have not been taken into account for the model calculation is not possible.

The second approach makes use of a sufficiently dense grid that spans the complete energy and angular range of the experimental data. An interpolation scheme is applied to obtain observables at energies and angles in-between the mesh points. This approach is beneficial if plenty of experimental data are available lying nearby in terms of energy or angle. Computing the observables at only a few mesh points significantly reduces the computation time compared to calculations at all energies of the included experimental data. In general, cross sections at high incident energies (above few MeV) are smooth functions of the energy and therefore interpolation between mesh points yields sufficiently accurate results.

We follow the second approach because of its greater flexibility. Especially, it allows predictions at energies where no experimental data are available, one motivation of using Bayesian inference in nuclear data evaluation.

The Bayesian inference applied in this thesis is based on the multivariate normal distribution which exhibits the important property that a linear transformation of the random vector leads again to a multivariate normal distributions. More precisely, if the pdf of \vec{X} is a multivariate normal distribution with mean vector $\vec{\mu}$ and covariance matrix Σ , then the pdf of the transformed variable $\vec{Y} = \mathbf{S}\vec{X} + \vec{b}$ is also a multivariate normal distribution with mean vector $\vec{\mu}' = \mathbf{S}\vec{\mu} + \vec{b}$ and covariance matrix $\Sigma' = \mathbf{S}\Sigma\mathbf{S}^T$. Due to this fact interpolation schemes that are given as linear transformations of the model vector are numerically beneficial because a closed-form solution of the Bayesian update formula is preserved.

Consider a vector of cross sections $\vec{\sigma}$ resulting from a model calculation, where each of its components σ_i is the cross section at a certain incident energy E_i . Assume that the cross sections of the experiment have been measured at different energies E'_j . An interpolation scheme based on a linear transformation to map the cross sections from the energies of the model mesh to the energies of the

experiment takes the form

$$\sigma(E'_j) = \sum_i \alpha(E'_j, E_i) \sigma(E_i). \quad (3.35)$$

The corresponding transformation sensitivity matrix \mathbf{S} is given by

$$\mathbf{S} = \begin{pmatrix} \alpha(E'_1, E_1) & \dots & \alpha(E'_1, E_n) \\ \vdots & \ddots & \vdots \\ \alpha(E'_m, E_1) & \dots & \alpha(E'_m, E_n) \end{pmatrix}. \quad (3.36)$$

Hence, the mapping of the vector of model cross sections is performed by $\vec{\sigma}' = \mathbf{S}\vec{\sigma}$. Noteworthy, such a linear transformation does not imply linear interpolation in-between the energies. As an example, consider three cross sections $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)^T$ at energies E_1, E_2, E_3 from a model calculation. These cross sections can be perfectly fitted with a polynomial of degree two,

$$f(E) = c_1 E^2 + c_2 E + c_3. \quad (3.37)$$

The coefficients $\vec{c} = (c_1, c_2, c_3)$ are the solution of the set of linear equations

$$\vec{\sigma} = \mathbf{X}\vec{c} \quad \text{with} \quad \mathbf{X} = \begin{pmatrix} E_1^2 & E_1 & 1 \\ E_2^2 & E_2 & 1 \\ E_3^2 & E_3 & 1 \end{pmatrix}, \quad (3.38)$$

leading to the formal solution $\vec{c} = \mathbf{X}^{-1}\vec{\sigma}$. Knowledge of the coefficients \vec{c} allows to interpolate the model cross sections at the experimental energies E'_i ,

$$f(E'_i) = (E_i'^2, E'_i, 1) \vec{c} = (E_i'^2, E'_i, 1) \mathbf{X}^{-1} \vec{\sigma}. \quad (3.39)$$

From these considerations, we can identify the linear transformation matrix $\mathbf{S} = (E_i'^2, E'_i, 1) \mathbf{X}^{-1}$ and thus a polynomial of degree two is a linear transformation. The matrix \mathbf{S} is fully specified by the model energies and the experimental energies. Once it is calculated, the model prediction at the experimental energies is $\vec{\sigma}' = \mathbf{S}\vec{\sigma}$ where $\vec{\sigma}$ contains the predictions at the model energies. More generally, all kinds of polynomials such as splines and trigonometric polynomials can be expressed as linear transformations.

Throughout this thesis, we choose a simple linear interpolation for integral cross sections and bilinear interpolation for spectra and angle-differential cross sections. Linear and bilinear interpolation lead to very sparse transformation matrices. Maximally four elements per row in the transformation matrix are distinct from zero. This property can be exploited to speed up matrix multiplications required in the Bayesian update procedure.

3.4.1 Mapping of angle-integrated cross sections

Mapping in the case of cross sections (perfect monochromatic incident beam) In this subsection we describe the mapping of angle angle-integrated cross section data from the model mesh to the incident energies of the experimental data. For simplicity we restrict ourselves to cross sections of one reaction channel. Let $\vec{E} = (E_1, \dots, E_n)^T$ be the mesh of model energies and $\vec{E}' = (E'_1, \dots, E'_m)^T$ the energies with available experimental data. The model energies E_j are assumed to be in ascending order, $E_j < E_{j+1}$. The model cross sections at \vec{E} are denoted by $\vec{\sigma} = (\sigma_1, \dots, \sigma_n)^T$. The interpolated model cross section σ'_i for one particular experimental energy E'_i obtained through linear interpolation is

$$\begin{aligned} \sigma'_i &= \sigma_j + \frac{\sigma_{j+1} - \sigma_j}{E_{j+1} - E_j} (E'_i - E_j) \\ &= \left(\frac{E_{j+1} - E'_i}{E_{j+1} - E_j} \right) \sigma_j + \left(\frac{E'_i - E_j}{E_{j+1} - E_j} \right) \sigma_{j+1}, \end{aligned} \quad (3.40)$$

where E_j is the closest energy to E'_i from below and E_{j+1} is the closest energy to E'_i from above. Thus, the elements of the transformation matrix \mathbf{S} are given by

$$S_{ij} := \begin{cases} \frac{E_{j+1} - E'_i}{E_{j+1} - E_j} & \text{if } E_j \text{ is closest neighbor to } E'_i \text{ from below} \\ \frac{E'_i - E_{j-1}}{E_j - E_{j-1}} & \text{if } E_j \text{ is closest neighbor to } E'_i \text{ from above} \\ 0 & \text{otherwise} \end{cases} \quad (3.41)$$

Each row of the transformation matrix \mathbf{S} describes the mapping from model cross sections to one particular experimental energy and contains at most two non-vanishing elements. Hence only $2m$ multiplications required in order to map the model cross sections to the experimental mesh. In contrast, spline interpolation would require $m \times n$ multiplications where n is the number of model mesh points.

Mapping in the case of cross sections (non-monochromatic incident beam, Gaussian folding) However, in a real experiment this simple mapping rule does not always provide an adequate description. Incident particle beams are usually not perfectly monochromatic. Therefore, the angle-integrated cross section measured in such an experiment is the superposition of cross sections in a certain interval of incident energies. As long as cross sections are sufficiently smooth functions of the incident energy, effects of non-monochromaticity can be neglected and one can use the linear interpolation scheme described above. However, if the cross section fluctuates significantly with varying energy, the interpolation scheme has to be adapted.

In the following, we derive the adapted interpolation scheme for the case that the energies in the incident particle beam are distributed according to a normal

distribution. Introducing the probability density function of the standard normal distribution,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad (3.42)$$

with vanishing mean value $\vec{\mu} = E[x] = 0$ and standard deviation $\delta = 1$, we define the cumulative distribution function

$$\Phi(x) = \int_{-\infty}^x \phi(x') dx', \quad (3.43)$$

which satisfies $\Phi(\infty) = 1$. Using the pdf of Equation 3.42, we can express normal distributions with arbitrary means μ and standard deviations δ ,

$$\phi(x; \mu, \delta) = \frac{1}{\delta} \phi\left(\frac{x - \mu}{\delta}\right), \quad (3.44)$$

$$\Phi(x; \mu, \delta) = \Phi\left(\frac{x - \mu}{\delta}\right). \quad (3.45)$$

In our present considerations, mean and standard deviation define center and spread of the energies in the incident particle beam. In order to get the proper mapping from the model cross sections to the observables of the experiment, we have to evaluate the energy-weighted integral

$$\sigma(E) = \int_{E_{\min}}^{E_{\max}} \sigma(E') \phi(E'; \mu, \delta) dE'. \quad (3.46)$$

Here, the integration boundaries E_{\min} and E_{\max} refer to the smallest and biggest energy of the model mesh. The model cross section $\sigma(E')$ can be obtained at arbitrary energies E' by linear interpolation, Equation 3.40. Thus, we can evaluate the integral piecewise, where the boundaries of each segment are defined by consecutive energies E_j and E_{j+1} of the model mesh. Within one segment, the integral takes the form

$$\int_{E_j}^{E_{j+1}} \left[\left(\frac{E_{j+1} - E'}{E_{j+1} - E_j} \right) \sigma_j + \left(\frac{E' - E_j}{E_{j+1} - E_j} \right) \sigma_{j+1} \right] \phi(E'; \mu, \delta) dE' \quad (3.47)$$

This integral is of the form $\int (aE' + b)\phi(E'; \mu, \sigma) dE'$ and can be solved analytically in terms of the cumulative distribution function. For the sake of compact expressions, we introduce the functions

$$g(E_j, E_{j+1}; \mu, \sigma) = \phi(E_{j+1}; \mu, \delta) - \phi(E_j; \mu, \delta), \quad (3.48)$$

$$G(E_j, E_{j+1}; \mu, \sigma) = \Phi(E_{j+1}; \mu, \delta) - \Phi(E_j; \mu, \delta). \quad (3.49)$$

Thus the solution of the integral Equation 3.47 can be written

$$\begin{aligned} \frac{\sigma_j}{E_{j+1} - E_j} \left[(E_{j+1} - E_j - \mu) G(E_j, E_{j+1}; \mu, \delta) + \delta g(E_j, E_{j+1}; \mu, \delta) \right] + \\ \frac{\sigma_{j+1}}{E_{j+1} - E_j} \left[\mu G(E_j, E_{j+1}; \mu, \delta) - \delta g(E_j, E_{j+1}; \mu, \delta) \right]. \end{aligned} \quad (3.50)$$

In order to assemble the solutions of all piecewise integrations, we use the abbreviations

$$c_j(\mu, \delta) := \frac{1}{E_{j+1} - E_j} \left[(E_{j+1} - E_j - \mu) G(E_j, E_{j+1}; \mu, \delta) - \delta g(E_j, E_{j+1}; \mu, \delta) \right], \quad (3.51)$$

$$\tilde{c}_j(\mu, \delta) := \frac{1}{E_j - E_{j-1}} \left[\mu G(E_{j-1}, E_j; \mu, \delta) - \delta g(E_{j-1}, E_j; \mu, \delta) \right], \quad (3.52)$$

to specify the coefficients

$$a_j(\mu, \delta) := \begin{cases} c_1(\mu, \delta) & \text{for } j = 1 \\ c_j(\mu, \delta) + \tilde{c}_j(\mu, \delta) & \text{for } 2 \leq j \leq n-1 \\ \tilde{c}_j(\mu, \delta) & \text{for } j = n \end{cases} . \quad (3.53)$$

Finally, the folded cross section from the model that can be compared with the experimental observable can be written as

$$\sigma(\mu, \delta) = \sum_{j=1}^n a_j(\mu, \delta) \sigma_j . \quad (3.54)$$

As a reminder, the energy mesh of the model consists of n energies E_j and σ_j are the associated values of the cross sections obtained by the model calculation. The mean μ corresponds to the energy of the incident beam at which the experiments are performed while the standard deviation δ characterizes the energy spread of the incident particle beam. Suppose we have m experimental points (μ_i, δ_i) . Then, the corresponding transformation matrix \mathbf{S} is given by

$$\mathbf{S} = \begin{pmatrix} a_1(\mu_1, \delta_1) & \cdots & a_n(\mu_1, \delta_1) \\ \vdots & \ddots & \vdots \\ a_1(\mu_m, \delta_m) & \cdots & a_n(\mu_m, \delta_m) \end{pmatrix} . \quad (3.55)$$

Note that the folding of model cross sections leads to a dense transformation matrix. Analyzing the expression in more detail reveals that this is only partially true. In fact, the contributions of model cross sections at energies of the mesh a few δ_i away from μ_i are almost vanishing and can be neglected in good approximation.

The variances and covariances of the mapped model cross sections play an important role in Bayesian inference with multivariate normal distributions. The variance of a mapped model cross section can be expressed in terms of the variances of the model cross sections at the mesh points,

$$\text{Var}[\sigma(\mu, \delta)] = \sum_{j=1}^n [a_j(\mu, \delta)]^2 \text{Var}[\sigma_j] + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} a_i(\mu, \delta) a_j(\mu, \delta) \text{Cov}[\sigma_i, \sigma_j] . \quad (3.56)$$

The covariance between two mapped model cross sections $\sigma(\mu_1, \delta_1)$ and $\sigma(\mu_2, \delta_2)$ is given by

$$\text{Cov}[\sigma(\mu_1, \delta_1), \sigma(\mu_2, \delta_2)] = \sum_{i=1}^n \sum_{j=1}^n a_i(\mu_1, \delta_1) a_j(\mu_2, \delta_2) \text{Cov}[\sigma_i, \sigma_j]. \quad (3.57)$$

Variances and covariances of mapped model cross sections are given as linear combinations of variances and covariances of the model cross sections at the mesh points. This is due to the fact that the folding integral Equation 3.46 represents a linear functional.

It must be remarked that the inclusion of the energy spread of the incident beam is conceptually different from taking into account an error in the determination of the incident energy. Accounting for the energy spread means to model a known property of the experimental setup. In contrast, accounting for the uncertainty of the incident energy means to account for the imperfect knowledge of the actual state of the incident beam. In other words, we may model the energy spread of the incident beam with a normal distributions $\mathcal{N}(\mu, \sigma)$, but we are still uncertain about the exact values of μ and σ . The proper treatment of the uncertainty of the incident energy is the subject of section 3.5.

3.4.2 Mapping spectra

The treatment of angle-differential cross sections as well as spectra of emitted particles in the update procedure is more involved because of their dependence on two variables. Hence a different procedure for the mapping of model data to the energies of the experiment must be applied. In the following we restrict the explanation to the treatment of spectra.

The collision of an incident beam of energy E with some target nuclei will lead to a variety of reactions which are compatible with energy conservation. Depending on the actual reaction, particles with different energies E' will be emitted. Thus we can a spectrum of the emitted particles described by the cross section $s(E, E')$.

The mathematical relation between a cross section $\sigma(E)$ at a certain incident energy E and the associated spectrum $s(E, E')$ with E' being the energy of the scattered particle is

$$\sigma(E) = \int_0^{\infty} s(E, E') dE'. \quad (3.58)$$

The upper boundary for integration is due to the conservation of energy. Knowledge of the spectrum $s(E, E')$ at every emitted energy E' contains the information needed to calculate the associated reaction cross section $\sigma(E)$.

In order to apply Bayesian inference, model predictions of the spectrum at incident energies E_i and emitted energies E'_j of the model mesh have to be mapped

to the incident and emitted energies of the experimental data. Two cases have to be addressed. If the emitted energy is about the incident energy, the spectrum is given by a number of spikes associated with the resolved excitation energies of the residual target nucleus. If the emitted energy is sufficiently smaller than the incident energy, the spectrum can be regarded as a continuous function of the emitted energy due to the limited resolution of detectors.

We first consider the case of a continuous spectrum. We use the abbreviation $s_{ij} \equiv s(E_i, E_j)$ to denote the values of the spectrum calculated in a model calculation at the mesh points (E_i, E_j) . Let \tilde{E} and \tilde{E}' be the incident and emitted energy of one experimental data point, respectively. In order to determine the corresponding value of the model, bilinear interpolation is used in this thesis. The formula for bilinear interpolation is

$$s(\tilde{E}, \tilde{E}') = \frac{1}{(E_2 - E_1)(E'_2 - E'_1)} \left[(E_2 - \tilde{E})(E'_2 - \tilde{E}') s_{11} + (E_2 - \tilde{E})(\tilde{E}' - E'_1) s_{12} + (\tilde{E} - E_1)(E'_2 - \tilde{E}') s_{21} + (\tilde{E} - E_1)(\tilde{E}' - E'_1) s_{22} \right]. \quad (3.59)$$

The quantities E_1 and E'_1 are the closest energies to \tilde{E} and \tilde{E}' from below within the set of all energies of the model mesh. Similarly, the quantities E_2 and E'_2 are the closest energies to \tilde{E} and \tilde{E}' from above. Bilinear interpolation is a linear transformation from the vector of computed values of the spectrum to the incident energies and emitted energies of the experimental data.

Suppose that the model mesh consists of n incident energies E_i and k emitted energies E'_j . We assume the incident energies and the emitted energies to be ordered, $E_i < E_{i+1}$ and $E'_j < E'_{j+1}$. A reasonable structure of the model vector \vec{s} containing the spectrum values could be $\vec{s} = (s_{11}, \dots, s_{1k}, \dots, s_{i1}, \dots, s_{ik}, \dots, s_{n1}, \dots, s_{nk})^T$. However, other schemes to assign values of the spectrum to positions in the model vector are equally possible. In order to formulate the transformation matrix, we introduce the following quantities. The quantity $V_i^<$ denotes the incident energy E_l of the model mesh that is closest from below to the incident energy \tilde{E}_i of the i^{th} experimental data point. The quantity $V_i^>$ denotes the incident energy E_{l+1} of the model mesh that is closest from above to the incident energy \tilde{E}_i of the i^{th} experimental data point. The quantities $W_i^<$ and $W_i^>$ refer to the emitted energy and are analogously defined. Additionally, we abbreviate $d_i = [(V_i^> - V_i^<)(W_i^> - W_i^<)]^{-1}$. With $E_{\#j}$ we refer to the incident energy associated with the j^{th} element of the model vector. Analogously, with $E'_{\#j}$ we refer to the emitted energy associated with the j^{th} element of the model vector. For instance, using the structure of the model vector \vec{s} outlined above, we have $E_{\#2} \equiv E_1$ and $E'_{\#2} \equiv E'_2$. As a reminder,

\tilde{E}_i and \tilde{E}'_i refer to incident energy and emitted energy of the i^{th} experimental data point, respectively.

Using the introduced notation, the elements S_{ij} of the transformation matrix \mathbf{S} for the spectra are

$$S_{ij} := \begin{cases} d_i (V_i^> - \tilde{E}_i)(W_i^> - \tilde{E}'_i) & \text{if } V_i^< = E_{\#j} \text{ and } W_i^< = E'_{\#j} \\ d_i (V_i^> - \tilde{E}_i)(\tilde{E}'_i - W_i^<) & \text{if } V_i^< = E_{\#j} \text{ and } W_i^> = E'_{\#j} \\ d_i (\tilde{E}_i - V_i^<)(W_i^> - \tilde{E}'_i) & \text{if } V_i^> = E_{\#j} \text{ and } W_i^< = E'_{\#j} \\ d_i (\tilde{E}_i - V_i^<)(\tilde{E}'_i - W_i^<) & \text{if } V_i^> = E_{\#j} \text{ and } W_i^> = E'_{\#j} \\ 0 & \text{otherwise} \end{cases} \quad (3.60)$$

The i^{th} row describes the mapping of spectrum values of the model vectors to the incident energy and emitted energy of the i^{th} experimental data point. The mapped values of the spectrum \vec{s}' are given by $\mathbf{S}\vec{s}$. The transformation matrix \mathbf{S} is very sparse, each row contains maximally four non-zero elements. Exploiting this sparsity is a very important step in dealing with spectral data. For instance, mapping to 1000 experimental data points from model data given on a mesh of 100 incident energies and 100 emitted energies leads to a transformation matrix of dimension $10^3 \times 10^4$. A matrix of this dimension amounts to 80 MBytes if stored in double precision, which is easily manageable on a modern computer. However, one goal of this thesis is to consistently evaluate dozens or even hundreds of spectra of different reaction channels together with cross sections and angle-differential cross sections. Under this point of view, the storage requirement of 80 MBytes for one reaction channel is already too much. However, exploiting the sparsity of \mathbf{S} greatly reduces the storage requirement. In the example, the reduced storage requirement would amount to merely 32 KBytes.

The transformation matrix Equation 3.60 is derived under the assumption of a monochromatic beam of incident particles and a detector with perfect energy resolution. However, experiments to measure spectra have not only a certain spread of energies in the incident beam but also a limited resolution in measuring the energy of the emitted particles. If both the spread δ of energies in the incident beam and the energy resolution δ' for detecting emitted particles can be described by a normal distribution, the following formula maps the spectrum from the model mesh to the incident energy μ and emitted energy μ' of the experimental data

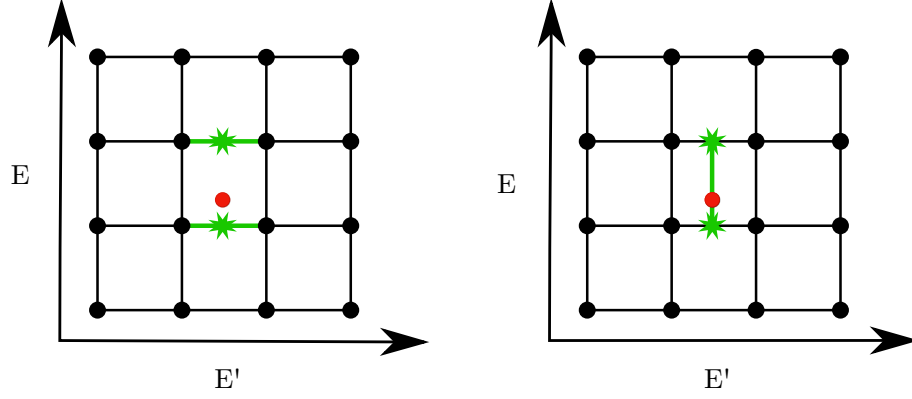


Figure 3.7: Bilinear interpolation as a sequence of two linear interpolations. First, linear interpolation is done along the emitted energy E' in the two rows corresponding to the incident energies of the model grid that are nearest to the incident energy of the experimental data point. Second, linear interpolation is done along the incident energy E between the obtained values (green stars) from the first step to get the model value at the incident energy of the experimental data point (red circle).

point,

$$\tilde{s}(\mu, \mu') = \int_{E_{\text{inc}}^{\text{min}}}^{E_{\text{inc}}^{\text{max}}} \int_{E_{\text{out}}^{\text{min}}}^{E_{\text{out}}^{\text{max}}} s(E, E') \phi(E; \mu, \delta) \phi(E'; \mu', \delta') dE dE'. \quad (3.61)$$

The limits of the integrals refer to the smallest and largest incident and emitted energies of the model mesh. The definition of the two normal distributions denoted with function symbol ϕ is given in Equation 3.42 and Equation 3.44. The value of $s(E, E')$ can be obtained by bilinear interpolation from the calculated spectrum at the model mesh points, Equation 3.59. Bilinear interpolation can be carried out by sequentially performing linear interpolations. Figure 3.7 illustrates this approach. First we perform linear interpolation with respect to E , which allows us to use Equation 3.54 for the folding of an angle-integrated cross section,

$$\tilde{s}(\mu, \mu') = \int_{E_{\text{out}}^{\text{min}}}^{E_{\text{out}}^{\text{max}}} \sum_{j=1}^n a_j(\mu, \delta) s(E_j, E') \phi(E'; \mu', \delta') dE'. \quad (3.62)$$

The coefficients $a_j(\mu, \delta)$ are defined as in Equation 3.53. The sum is over all n incident energies E_j of the model mesh.

Applying linear interpolation with respect to the emitted energy E' leads in a further step to

$$\tilde{s}(\mu, \mu') = \sum_{j=1}^n \sum_{k=1}^m a_j(\mu, \delta) a'_k(\mu', \delta') s(E_j, E'_k). \quad (3.63)$$

The coefficients $a'_k(\mu', \delta')$ are calculated in the same way as $a_k(\mu, \delta)$, but with the emitted energy E'_k inserted. The sum is carried out over all m emitted energies of the model mesh.

Suppose there exist several experimental values of the spectrum. Let $\mu_i, \delta_i, \mu'_j, \delta'_j$ denote the measurement conditions for the i^{th} experimental data point. Then, the elements S_{ij} of the transformation matrix are

$$S_{ij} := a_{p_j}(\mu_i, \delta_i) a_{q_j}(\mu_i, \delta_i), \quad (3.64)$$

where p_j is the index of the incident energy in the sorted mesh of incident energies that corresponds to the incident energy associated with the j^{th} element of the model vector $\vec{\sigma}$. The q_j are analogously defined for the emitted energies.

The discussion so far was for the case of a continuous spectrum. However, if the emitted energy is just slightly smaller than the incident energy, only few low lying excitation levels of the target nucleus are involved in the scattering process. The energy difference between these levels is reflected in the energy differences of the emitted particles. Therefore, if the levels are sufficiently spaced, the detector can resolve the energy differences and the spiky structure of the spectrum becomes observable. With $\sigma_{\text{xc},i}$ being the cross section for scattering at the i^{th} level of the target nucleus, we can write the spectrum as

$$s(E, E') = \sum_i \sigma_{\text{xc},i} \delta(E_{\text{xc},i} - E - E'). \quad (3.65)$$

In order to compare the results of model calculations with the experiment, we have to fold the true spectrum with the normal distributions $\phi(E; \mu, \delta)$ and $\phi(E'; \mu', \delta')$ characterizing the energy spread of the incident beam and the detector resolution,

$$\tilde{s}(\mu, \mu') = \int_0^\infty \int_0^\infty s(E, E') \phi(E; \mu, \delta) \phi(E'; \mu', \delta') dE dE'. \quad (3.66)$$

Due to the Dirac delta-distribution, the integral with respect to E' can be immediately evaluated,

$$\tilde{s}(\mu, \mu') = \int_0^\infty \sum_i \sigma_{\text{xc},i} \phi(E; \mu, \delta) \phi(E - E_{\text{xc},i}; \mu', \delta') dE. \quad (3.67)$$

The remaining integral has a closed-form solution,

$$\tilde{s}(\mu, \mu') = \sum_i \sigma_{\text{xc},i} \phi\left(E_{\text{xc},i}; (\mu - \mu'), \sqrt{\delta^2 + \delta'^2}\right) G\left(0, \infty; \tilde{\mu}, \tilde{\delta}\right), \quad (3.68)$$

with the function $G(\cdot)$ defined in Equation 3.49 and the abbreviations

$$\tilde{\mu} = \frac{(E_{\text{xc},i} + \mu') \delta^2 + \mu \delta'^2}{\delta^2 + \delta'^2} \quad \text{and} \quad \tilde{\delta} = \frac{\delta \delta'}{\sqrt{\delta^2 + \delta'^2}}. \quad (3.69)$$

If we assume $\tilde{\mu} \gg \tilde{\delta}$, then $G(0, \infty; \tilde{\mu}; \tilde{\delta}) \approx 1$ and therefore Equation 3.70 reduces to

$$\tilde{s}(\mu, \mu') = \sum_i \sigma_{\text{xc},i} \phi\left((\mu - \mu'); E_{\text{xc},i}, \sqrt{\delta^2 + \delta'^2}\right). \quad (3.70)$$

We swapped the arguments $(\mu - \mu')$ and $E_{xc,i}$, which is possible because of the symmetry of the normal distribution, see Equation 3.42. Therefore, the result is a sum of normal distributions centered at the $E_{xc,i}$ and scaled by the value of the respective cross section $\sigma_{xc,i}$. Finally, the complete spectrum can be obtained by adding up the solution for the folded continuous spectrum, Equation 3.63, and the solution for the folded discrete spectrum, Equation 3.70.

In general, the energy spread of the incident beam and the resolution of the detector may be neglected in the domain where the spectrum can be regarded as continuous. However, for emitted energies close to the incident energy, the discrete nature of the spectrum becomes noticeable. In this case, it is of utmost importance to carefully assess the details of the experiment and fix spread δ and resolution δ' properly. In addition, the assumption of the normal distribution can be questioned. For instance, the geometry of the experimental setup may suggest using a rectangular kernel instead of a normal kernel. A non-adequate specification of the folding kernel can lead to severe consequences for the comparability of model results and experimental data. Even more severe, estimates or associated uncertainties obtained through Bayesian inference may lose their validity.

3.5 Uncertainty about the incident energy

Specifics of the experimental setup, such as a finite energy spread in the incident beam, can be taken into account in the Bayesian update by constructing the sensitivity matrix \mathbf{S} appropriately. Details on its construction have been given in the previous two sections. Error sources, such as the imperfect knowledge about the detector efficiency, can be taken into account by a proper construction of the experimental covariance matrix \mathbf{B} . A general discussion on its construction has been given in section 3.1. Yet another possible error source requires special treatment. The consequences of having a wrong idea about the incident energy cannot be assessed by merely studying the experimental design. To assess the impact of an inaccurately determined incident energy requires knowledge of the model systematics. This section deals with the proper Bayesian treatment of an uncertainty in the incident energy.

Starting point is an extended version of the Bayesian update formula. We introduce a new random variable \vec{E} , which contains the incident energies associated with the experimental data points. Taking into account this new variable, the Bayesian update formula takes the form

$$\pi(\vec{p}, \vec{E} | \vec{\sigma}, \mathcal{M}) \propto \ell(\vec{\sigma} | \vec{p}, \vec{E}, \mathcal{M}) \pi(\vec{p}, \vec{E} | \mathcal{M}) \quad (3.71)$$

with the factorized prior

$$\pi(\vec{p}, \vec{E} | \mathcal{M}) = \pi(\vec{p} | \mathcal{M}) \pi(\vec{E}). \quad (3.72)$$

The posterior pdf refers to both the vector of model parameters \vec{p} and the vector of incident energies \vec{E} associated with the vector of measurements $\vec{\sigma}$. The factorized form of the prior indicates that model parameters and incident energies of the experiment are assumed uncorrelated before taking into account the data. However, in the resulting posterior pdf, correlations between these two variables will exist in general. The specification of the prior for the model parameter has been discussed in section 3.2. In the following we want to keep the numerical benefits of the Bayesian update based on the linearization of the model. Therefore, we assume again a multivariate normal distribution for the prior on the incident energies,

$$\pi(\vec{E}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K}|}} \exp \left\{ -\frac{1}{2} (\vec{E} - \vec{E}_0)^T \mathbf{K}^{-1} (\vec{E} - \vec{E}_0) \right\}. \quad (3.73)$$

The best guess of the incident energies associated with the experimental data points is denoted by \vec{E}_0 . The associated uncertainties are expressed in terms of the covariance matrix \mathbf{K} . The number of experimental data points is denoted by n .

In order to use the analytical Bayesian update formulas, we have to linearize the nuclear model $\mathcal{M}(\vec{p}; \vec{E})$ with respect to the model parameters and the incident energies,

$$\mathcal{M}_{\text{lin}}(\vec{p}; \vec{E}) = \vec{\sigma}_0 + \mathbf{S}(\vec{p} - \vec{p}_0) + \mathbf{T}(\vec{E} - \vec{E}_0), \quad (3.74)$$

with $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0; \vec{E}_0)$ and

$$\mathbf{S} = \left. \frac{\partial \mathcal{M}(\vec{p}; \vec{E})}{\partial \vec{p}} \right|_{\vec{p}=\vec{p}_0} \quad \text{and} \quad \mathbf{T} = \left. \frac{\partial \mathcal{M}(\vec{p}; \vec{E})}{\partial \vec{E}} \right|_{\vec{E}=\vec{E}_0}. \quad (3.75)$$

The Jacobian matrices \mathbf{S} and \mathbf{T} can be determined numerically. We used the center vectors \vec{p}_0 and \vec{E}_0 of the prior distributions as expansion points. Inserting this linearized model in the likelihood, Equation 3.3, yields

$$\ell(\vec{\sigma} | \vec{p}, \vec{E}, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{B}|}} \times \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p} - \mathbf{T}\vec{E})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p} - \mathbf{T}\vec{E}) \right\}, \quad (3.76)$$

where

$$\vec{\sigma}_{\text{eff}} = \vec{\sigma} - \vec{\sigma}_0 + \mathbf{S}\vec{p}_0 + \mathbf{T}\vec{E}_0. \quad (3.77)$$

The vector $\vec{\sigma}$ contains the measured values of the observables. The terms independent of \vec{p} and \vec{E} were merged with the experimental vector to simplify further treatment. Thus we introduce an effective measurement vector $\vec{\sigma}_{\text{eff}}$.

In order to derive the distribution parameters of the posterior pdf, we note that we are dealing with the statistical model

$$\vec{\sigma} = \mathcal{M}_{\text{lin}}(\vec{p}; \vec{E}) + \vec{\varepsilon} = \vec{\sigma}_0 + \mathbf{S}(\vec{p} - \vec{p}_0) + \mathbf{T}(\vec{E} - \vec{E}_0) + \vec{\varepsilon}. \quad (3.78)$$

The vector $\vec{\varepsilon}$ denotes the measurement error. We modeled the uncertainty about the true values of the variables \vec{p} , \vec{E} , and $\vec{\varepsilon}$ as multivariate normal distributions with covariance matrices

$$\text{Var}[\vec{p}] = \mathbf{A}, \text{Var}[\vec{E}] = \mathbf{K}, \text{ and } \text{Var}[\vec{\varepsilon}] = \mathbf{B}. \quad (3.79)$$

Due to the invariance of the multivariate normal distribution under linear transformations, the quantities $\mathbf{S}\vec{p}$ and $\mathbf{T}\vec{E}$ are multivariate normal distributions with covariance matrices

$$\text{Var}[\mathbf{S}\vec{p}] = \mathbf{SAS}^T \text{ and } \text{Var}[\mathbf{T}\vec{E}] = \mathbf{TKT}^T. \quad (3.80)$$

Using the folding theorem Equation 2.48, we know that the sum in Equation 3.78 is again a multivariate normal distribution. We are able to compute the covariance matrices between $\vec{\sigma}$ and the terms in the sum. Recalling that \vec{p} , \vec{E} and $\vec{\varepsilon}$ are pairwise independent, the application of Equation 2.28 yields

$$\text{Cov}[\vec{\sigma}, \mathbf{S}\vec{p}] = \mathbf{SA} \text{ and } \text{Cov}[\vec{\sigma}, \mathbf{T}\vec{E}] = \mathbf{TK}. \quad (3.81)$$

The covariance matrix of $\vec{\sigma}$ is given by

$$\text{Var}[\vec{\sigma}] = \mathbf{SAS}^T + \mathbf{TKT}^T + \mathbf{B}. \quad (3.82)$$

Using the results in Equation 3.81 and Equation 3.82, we can explicitly write the center vector and the covariance matrix of the joint multivariate normal distribution of the random vector $(\vec{\sigma}, \vec{p}, \vec{E})^T$,

$$\vec{\mu} = \begin{pmatrix} \vec{\sigma}_0 \\ \vec{p}_0 \\ \vec{E}_0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} (\mathbf{SAS}^T + \mathbf{TKT}^T + \mathbf{B}) & \mathbf{SA} & \mathbf{TK} \\ \mathbf{AS}^T & \mathbf{A} & \mathbf{0} \\ \mathbf{KT}^T & \mathbf{0} & \mathbf{K} \end{pmatrix}. \quad (3.83)$$

These formulas yield the updated center vector \vec{p}_1 and updated covariance matrix \mathbf{A}_1 of the posterior multivariate normal distribution, where the incident energy is marginalized out,

$$\pi(\vec{p} | \vec{\sigma}, \mathcal{M}) = \int \pi(\vec{p}, \vec{E} | \vec{\sigma}, \mathcal{M}) d\vec{E} \sim \mathcal{N}(\vec{p}_1, \mathbf{A}_1). \quad (3.84)$$

If we obtained from an experiment a measurement vector $\vec{\sigma}$, also $\vec{\sigma}_{\text{eff}}$ is determined due to Equation 3.77. We can condition the multivariate normal distribution of $(\vec{\sigma}, \vec{p}, \vec{E})^T$ upon $\vec{\sigma}$. Using Equation 2.44 and Equation 2.45, we obtain the characterizing quantities of the posterior distribution,

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}\mathbf{S}^T(\mathbf{S}\mathbf{A}\mathbf{S}^T + \mathbf{T}\mathbf{K}\mathbf{T}^T + \mathbf{B})^{-1}(\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p}_0 - \mathbf{T}\vec{E}_0), \quad (3.85)$$

$$\mathbf{A}_1 = \mathbf{A} - \mathbf{A}\mathbf{S}^T(\mathbf{S}\mathbf{A}\mathbf{S}^T + \mathbf{T}\mathbf{K}\mathbf{T}^T + \mathbf{B})^{-1}\mathbf{S}\mathbf{A}. \quad (3.86)$$

As a side note, we can summarize $\tilde{\mathbf{B}} = \mathbf{T}\mathbf{K}\mathbf{T}^T + \mathbf{B}$ and may refer to $\tilde{\mathbf{B}}$ as the experimental covariance matrix. If we denote by σ_i the prediction of the model $\mathcal{M}(\vec{p}; \vec{E}_i)$ at incident energy E_i , the covariance matrix $\tilde{\mathbf{B}}$ can be written as

$$\tilde{B}_{ij} = \frac{\partial \sigma_i}{\partial E_m} \text{Cov}[E_m, E_n] \frac{\partial \sigma_j}{\partial E_n} + B_{ij}. \quad (3.87)$$

This formula just performs linear error propagation, e.g. Smith (1991). Thus, standard linear error propagation can be interpreted in the Bayesian framework as expressing uncertainty about the observables as multivariate normal distributions and linearizing the relationships.

Finally, we can also compute the posterior distribution parameters of the incident energy. In the same way Equation 3.85 and Equation 3.86 were derived, we can get a refined assessment of the incident energy,

$$\vec{E}_1 = \vec{E}_0 + \mathbf{K}\mathbf{T}^T(\mathbf{S}\mathbf{A}\mathbf{S}^T + \mathbf{T}\mathbf{K}\mathbf{T}^T + \mathbf{B})^{-1}(\vec{\sigma}_{\text{eff}} - \mathbf{S}\vec{p}_0 - \mathbf{T}\vec{E}_0), \quad (3.88)$$

$$\mathbf{K}_1 = \mathbf{K} - \mathbf{K}\mathbf{T}^T(\mathbf{S}\mathbf{A}\mathbf{S}^T + \mathbf{T}\mathbf{K}\mathbf{T}^T + \mathbf{B})^{-1}\mathbf{T}\mathbf{K}. \quad (3.89)$$

We showed in this section the feasibility of a sound Bayesian treatment of uncertainties with regard to the incident energies. The linearization of the model with respect to incident energies has led to closed-form expressions for the Bayesian update. Even though the formulas were derived on the basis of a linearized model, analogous update formulas can be obtained for the surrogate approach.

3.6 Conservation of features

In the previous sections we have introduced simplified models which allowed us to derive closed form expressions for the Bayesian inference. Especially the linearized nuclear model approach (section 3.2) and the surrogate approach section 3.3 are considered. Due to the simplification not all properties of the original model will be preserved. However, we can show that two important features are preserved. One feature is the conservation of sum rules. The other feature is the conservation of the relation between angle-differential cross section (or spectra) and associated angle-integrated cross section.

We start the discussion with the conservation of sum rules. Suppose we have the total cross section σ_{tot} , the elastic cross section σ_{el} , and the non-elastic cross section $\sigma_{\text{non-el}}$ at the same incident energy. It must always hold that

$$\sigma_{\text{tot}} = \sigma_{\text{el}} + \sigma_{\text{non-el}}. \quad (3.90)$$

The result of a nuclear model calculation obeys this relation. We combine these three cross sections to a vector $\vec{\sigma} = (\sigma_{\text{tot}}, \sigma_{\text{el}}, \sigma_{\text{non-el}})^T$. To check whether the linearized nuclear model,

$$\mathcal{M}_{\text{lin}}(\vec{p}) = \vec{\sigma}_0 + \mathbf{S}(\vec{p} - \vec{p}_0), \quad (3.91)$$

also obeys the sum rule, we have to study the sensitivity matrix \mathbf{S} because the sum rule is evidently preserved in $\vec{\sigma}_0$ due to $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$. The sensitivity matrix is the Jacobian matrix of the nuclear model at the expansion point \vec{p}_0 . Thus, we can write

$$\mathbf{S} = \begin{pmatrix} \frac{\partial \mathcal{M}}{\partial p_1} & \frac{\partial \mathcal{M}}{\partial p_2} & \cdots & \frac{\partial \mathcal{M}}{\partial p_d} \end{pmatrix}. \quad (3.92)$$

Each partial derivative $\partial \mathcal{M} / \partial p_i$ represents a column in the sensitivity matrix. The numerical evaluation of these column vectors is carried out by taking the discrete derivative,

$$\frac{\partial \mathcal{M}}{\partial p_i} = \frac{\mathcal{M}(\vec{p} + \vec{\epsilon}_i) - \mathcal{M}(\vec{p})}{\epsilon}. \quad (3.93)$$

In the vector $\vec{\epsilon}_i$ the only non-zero component is the i^{th} element associated with the model parameter p_i . Its value is given by a small numerical constant ϵ . For numerical derivatives carried out in this thesis, we choose ϵ to be between 1% and 1‰ of the model parameter value. In order to show that the linearized model preserves sum rules, we have to show that the difference vector in the numerator of Equation 3.93 preserves sum rules. Explicitly writing the numerator yields

$$\mathcal{M}(\vec{p} + \vec{\epsilon}_i) - \mathcal{M}(\vec{p}) = \begin{pmatrix} \sigma'_{\text{tot}} \\ \sigma'_{\text{el}} \\ \sigma'_{\text{non-el}} \end{pmatrix} - \begin{pmatrix} \sigma_{\text{tot}} \\ \sigma_{\text{el}} \\ \sigma_{\text{non-el}} \end{pmatrix}. \quad (3.94)$$

The primed cross sections are the result of $\mathcal{M}(\vec{p} + \vec{\epsilon}_i)$ and those without primes are the result of $\mathcal{M}(\vec{p})$. Each of the two vectors is consistent with the sum rule, because they are the result of an exact model calculation. Therefore, we have

$$\begin{aligned} \sigma'_{\text{tot}} - \sigma_{\text{tot}} &= (\sigma'_{\text{el}} + \sigma'_{\text{non-el}}) - (\sigma_{\text{el}} + \sigma_{\text{non-el}}) \\ &= (\sigma'_{\text{el}} - \sigma_{\text{el}}) + (\sigma'_{\text{non-el}} - \sigma_{\text{non-el}}). \end{aligned} \quad (3.95)$$

The expressions in the brackets of the second row are the values of the elastic and non-elastic component of the difference vector. Therefore each column in the

sensitivity matrix obeys sum rules. Using the representation of the sensitivity matrix in Equation 3.92, we can rewrite Equation 3.91 to

$$\mathcal{M}_{\text{lin}}(\vec{p}) = \vec{\sigma}_0 + \sum_{i=1}^d \frac{\partial \mathcal{M}}{\partial p_i} (p_i - p_{0,i}). \quad (3.96)$$

Because all occurring vectors obey the sum rule, their linear combination does as well.

Having shown the conservation of sum rules for the linearized model, we continue the discussion for the surrogate model. The surrogate model introduced in section 3.3 is given by

$$\mathcal{M}_{\text{sur}}(\vec{\sigma}_{\text{mod}}) = \mathbf{S} \vec{\sigma}_{\text{mod}}. \quad (3.97)$$

The application of the transformation matrix \mathbf{S} interpolates the cross sections $\vec{\sigma}_{\text{mod}}$ given at the incident energies \vec{E} of the model mesh to the incident energies \vec{E}' of the experimental data. For the current consideration we assume the model vector to be partitioned as $\vec{\sigma}_{\text{mod}} = (\vec{\sigma}_{\text{tot}}, \vec{\sigma}_{\text{el}}, \vec{\sigma}_{\text{non-el}})^T$. Each component $\vec{\sigma}_{\text{tot}}$, $\vec{\sigma}_{\text{el}}$, $\vec{\sigma}_{\text{non-el}}$ contains the cross sections on the same mesh of incident energies \vec{E} . The mapping to the incident energies of the experiment \vec{E}' is done via the transformation matrices \mathbf{S}_{tot} , \mathbf{S}_{el} , and $\mathbf{S}_{\text{non-el}}$. The formula for the interpolation with matrices written element wise is

$$\begin{pmatrix} \vec{\sigma}'_{\text{tot}} \\ \vec{\sigma}'_{\text{el}} \\ \vec{\sigma}'_{\text{non-el}} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{\text{tot}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\text{el}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{\text{non-el}} \end{pmatrix} \begin{pmatrix} \vec{\sigma}_{\text{tot}} \\ \vec{\sigma}_{\text{el}} \\ \vec{\sigma}_{\text{non-el}} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{\text{tot}} \vec{\sigma}_{\text{tot}} \\ \mathbf{S}_{\text{el}} \vec{\sigma}_{\text{el}} \\ \mathbf{S}_{\text{non-el}} \vec{\sigma}_{\text{non-el}} \end{pmatrix}. \quad (3.98)$$

The cross section vectors with primes on the left hand side contain the interpolated values. In section 3.4 we showed that the transformation matrices are unambiguously determined by the incident energies of the model mesh \vec{E} and the incident energies of the measurements \vec{E}' . Because \vec{E} and \vec{E}' are the same for each component, we have

$$\mathbf{S}^* = \mathbf{S}_{\text{tot}} = \mathbf{S}_{\text{el}} = \mathbf{S}_{\text{non-el}}. \quad (3.99)$$

In order to check the conservation of sum rules for the interpolated values, we evaluate

$$\mathbf{S}^* \vec{\sigma}'_{\text{tot}} = \mathbf{S}^* \vec{\sigma}_{\text{el}} + \mathbf{S}^* \vec{\sigma}_{\text{non-el}} = \mathbf{S}^* \overbrace{(\vec{\sigma}_{\text{el}} + \vec{\sigma}_{\text{non-el}})}^{\vec{\sigma}_{\text{tot}}}. \quad (3.100)$$

Thus, under the condition that sum rules are preserved by the predicted cross sections on the model mesh, they are also preserved by the interpolated cross sections.

The question remains whether the surrogate model satisfies sum rules at the mesh points. The surrogate model is constructed by computing a center vector $\vec{\sigma}_0$

and a covariance matrix $\tilde{\mathbf{A}}_0$ from the results of model calculations σ_i . For both approaches to construct the center vector, either as a direct result of a model calculation or as the arithmetic mean of model results, the center vector obeys sum rules. The reason in the latter approach is that a linear combination of vectors preserves sum rules if each vector obeys them. The prior distribution for the surrogate parameter vector $\vec{\sigma}_{\text{mod}}$ determines which combinations of cross sections are feasible. Sum rules are linear relationships between cross section. Linear relationships can be perfectly reflected in a covariance matrix. This means that a cross section vector violating a sum rule is associated with probability zero. Therefore, the surrogate model preserves sum rules.

Finally, we discuss the conservation of the relationships between differential cross sections and integrated cross sections. As an example, the spectrum $s(E, E')$ at incident energy E and emitted energy E' is related to the integrated cross section $\sigma(E)$ by

$$\sigma(E) = \int_0^E s(E, E') dE'. \quad (3.101)$$

If the spectrum is given on a sufficiently dense grid, we can approximate the integral as a sum

$$\sigma(E) \approx \sum_{i=1}^n \alpha_i s(E, E'_i). \quad (3.102)$$

The coefficients α_i depend on the interpolation scheme for the spectrum. Equation 3.102 represents a linear relationship between the values of the spectrum and the integral cross sections and can be regarded as a special case of a sum rule. Hence, also this consistency property of cross sections and spectra (or angle-differential cross section) is preserved by both the linearized model and the surrogate model. A possible violation occurs in the magnitude due to the approximation of the integral by a sum, which usually can be neglected.

Finally we note that the surrogate model is capable of mimicking more features of the nuclear model than those discussed so far. The smoothness of cross sections is such an additional feature. In general, it is difficult to assess which features are preserved. The term feature itself is difficult to define. However, one qualitative measure for the amount of preserved features is the number of non-vanishing eigenvalues of the prior covariance matrix. The number of non-vanishing eigenvalues reveals how many axes in the space of cross section vectors are associated with non-vanishing variation. For instance, consider a vector containing the cross sections at hundred different incident of the same reaction channel. Vectors characterized by a distinct choice of cross sections span a vector space. Theoretically the nuclear model could permit variation along hundred orthogonal axes in this vector space. Yet, the number of axes with non-vanishing variation is typically

significantly lower. For the linearized model discussed in section 3.2, the number of axes with non-vanishing variation is given by the number of model parameters. In contrast, the surrogate model discussed in section 3.3 is more flexible due to the fact that non-linear features are taken into account as additional uncertainties. For instance, Figure 3.4 in section 3.3 demonstrated that a non-linear nuclear model with one model parameter would be approximated by a surrogate model allowing variation along two axes. Figure 3.8 shows the rapid decline of eigenvalues in the prior covariance matrix for the neutron-induced total cross section of ^{181}Ta . Thus, also in the surrogate model the effective dimension is much lower than the maximal possible.

In this section we showed that sum rules, and the relationship between differential cross sections and integrated cross sections are preserved by the Bayesian update formula based on a simplified model. These consistency properties are an essential aspect of the method because evaluated nuclear data files are only considered to be reliable and applicable if they satisfy these consistency properties. In chapter 9 we will make use of the fact that the multivariate normal distribution can perfectly capture sum rules to define a consistent model defect.

The next chapter gives an overview over existing methods for nuclear data evaluation. Either they are based on one of the approaches detailed in section 3.2 and section 3.3 or on a Monte Carlo procedure.

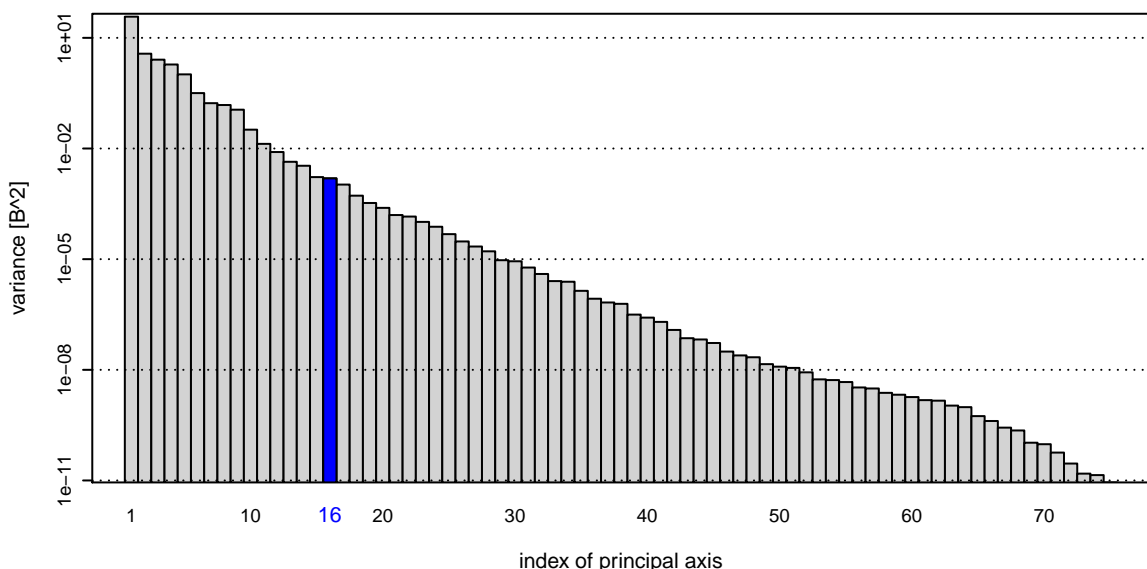


Figure 3.8: Eigenvalues of the prior covariance for the neutron-induced total cross section of ^{181}Ta constructed from a sample of model predictions with varied optical model parameters. Sixteen parameters of the neutron optical potential were varied according to a uniform distribution.

Existing methods

Most existing methods for nuclear data evaluation are based on Bayesian statistics. Strong arguments for the use of Bayesian statistics are given in the textbook ‘Probability Theory: The Logic of Science’ by Jaynes (2003). One argument is that Bayesian statistics reduces to logical reasoning if there is certainty about the truth of propositions. In the face of uncertainty, conclusions resulting from Bayesian inference are still in line with common sense reasoning. The notion of common sense is formalized in the book. The important message is that an inference procedure, which is not equivalent to Bayesian statistics, must necessarily be in disagreement with classical logic or in disagreement with cogent principles of common sense reasoning. These features of Bayesian inference are in complete agreement with the goals and requirements of nuclear data evaluation. Hence sharing Jaynes viewpoint leads to the conclusion that proper nuclear data evaluation should be performed within the framework of Bayesian statistics. Therefore, the only non-Bayesian evaluation method (BFMC) of Bauge, Hilaire, and Dossantos-Uzarralde (2007) presented in this chapter will also be interpreted within the framework of Bayesian statistics.

The outlined methods are distinct from each other in two aspects. The first aspect concerns the underlying statistical model, which means the choice of probability distribution for the likelihood and the prior distribution. The second aspect concerns if and how approximations are made to obtain results. Following, we present various methods and discuss their advantages and disadvantages.

4.1 Backward-Forward Monte Carlo

The Backward-Forward Monte Carlo (BFMC) method of Bauge, Hilaire, and Dossantos-Uzarralde (2007) relies on two steps. In the first step—the backward step—model parameter vectors \vec{p}_i are drawn from a uniform distribution. Then, for each parameter vector \vec{p}_i a model prediction vector $\vec{\sigma}_i = \mathcal{M}(\vec{p}_i)$ is obtained. The cross sections in a model prediction vector $\vec{\sigma}_i$ are computed at the incident energies associated with the cross sections in the measurement vector $\vec{\sigma}_{\text{exp}}$. Thereafter, the generalized χ^2 -value of each model prediction is calculated using the measurement vector,

$$\chi_i^2 = (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i)^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i). \quad (4.1)$$

The covariance matrix of the experimental data is denoted by \mathbf{B} . As a side remark, this expression is proportional to the exponent of the likelihood in Equation 3.3. Thus, finding a parameter vector \vec{p}^* whose prediction $\vec{\sigma}^*$ leads to a minimum of the χ^2 -value is equivalent to maximizing the likelihood given by a multivariate normal distribution.

In the BMFC method, the χ^2 -values are used to construct a weight for each sampled vector $\vec{\sigma}_i$. The weights are given by the expression

$$\omega_i = C \exp \left[- \left(\frac{\chi_i^2}{\chi_{\min}^2} \right)^2 \right], \quad (4.2)$$

where C is a normalization constant and χ_{\min}^2 is the smallest χ^2 -value which occurred for the samples. Based on these weights, a multivariate normal distribution for the parameter vector is constructed. Its center vector is computed by

$$\vec{p}' = \frac{\sum_{i=1}^n \omega_i \vec{p}_i}{\sum_{i=1}^n \omega_i}, \quad (4.3)$$

where n is the number of samples. The associated covariance matrix is given by

$$\mathbf{A}' = \frac{\sum_{i=1}^n \omega_i (\vec{p}_i - \vec{p}') (\vec{p}_i - \vec{p}')^T}{\sum_{i=1}^n \omega_i}. \quad (4.4)$$

With the calculation of the distribution parameters \vec{p}' and \mathbf{A}' the backward step is completed.

In the second step—the forward step—the uncertainties about the cross sections are computed based on the uncertainties about the model parameters. To this end, model parameter vectors \vec{q}_i are sampled from the multivariate normal distribution $\mathcal{N}(\vec{p}', \mathbf{A}')$. Then, for each parameter vector \vec{q}_i the model predictions $\vec{\tau}_i = \mathcal{M}(\vec{q}_i)$ are calculated. A center vector (or best estimate) is then

$$\vec{\tau}' = \frac{1}{m} \sum_{i=1}^m \vec{\tau}_i, \quad (4.5)$$

with the associated covariance matrix

$$\mathbf{C} = \frac{1}{m-1} \sum_{i=1}^m (\vec{\sigma}_i - \vec{\sigma}') (\vec{\sigma}_i - \vec{\sigma}')^T. \quad (4.6)$$

The variable m denotes the number of drawn samples. Noteworthy, m and n do not have to be equal. Whether to use m or $m-1$ in the denominator can be discussed. The former choice leads to a biased estimator for finite sample sizes whereas the latter choice to an unbiased one. However, for sample sizes large enough, the difference between these two estimators may be of no significance.

An advantage of the BFMC method is that the original nuclear model is used. No approximations are introduced. A disadvantage is that the method is not based on Bayesian statistics. Nevertheless, we can interpret the method within the framework of Bayesian statistics and point out the statistical assumptions made. We repeat the Bayesian update formula here,

$$\pi(\vec{p} | \vec{\sigma}_{\text{exp}}, \mathcal{M}) \propto \ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M}). \quad (4.7)$$

We want to identify the prior distribution and the likelihood which are implicitly assumed in the BFMC method. Because parameter vectors are sampled uniformly in the backward step, the prior distribution is given by a uniform distribution. Scaling the obtained samples with the weights in Equation 4.2 means to assume the same expression for the likelihood. If we rearrange Equation 4.2, we get for the likelihood

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) \propto \left[\exp \left\{ - \left(\frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}))^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p})) \right)^2 \right\} \right]^{\left(\frac{2}{\chi_{\min}^2} \right)^2}. \quad (4.8)$$

The sample vector $\vec{\sigma}_i$ had been replaced by the prediction $\mathcal{M}(\vec{p})$ for an arbitrary parameter vector \vec{p} . This form of the likelihood is difficult to justify. In section 3.1 we considered several motivations for the use of a multivariate normal distribution for the likelihood if there are no arguments supporting another choice. The likelihood in Equation 4.8 is distinct from the multivariate normal distribution by two modifications. These modifications are opposite in their effect. One modification is the squaring of the exponent which makes the distribution more peaked compared to the multivariate normal distribution. The other modification, which is taking the $(\chi_{\min}^2/2)^2$ -th root of the exponential function, flattens the probability distribution.

The original paper (Bauge, Hilaire, and Dossantos-Uzarralde, 2007) clearly articulates the assumption of a perfect model. We cite the relevant sentence: ‘The basis of the BFMC method relies on the assumption that the only source of un-

certainty of the calculated cross section is the imperfect knowledge of the model parameters.’

In chapter 6 we study the consequences for the evaluated cross sections and associated uncertainties if the model is not a perfect description of reality and the evaluation method does not take into account model defects. One consequence are unreasonable small evaluated uncertainties. The analysis of the BFMC in section 7.5 supports the assumption that the likelihood of the BFMC method is in most situations much flatter than a multivariate normal distribution. If this assumption holds, model predictions significantly deviating from the cross sections of the experiment are associated with higher probability than in the case of a multivariate normal distribution. As a consequence, evaluated uncertainties are also bigger. Therefore, one could interpret the non-normal likelihood of the BFMC method as a measure to take into account model defects. Such a treatment, however, is not satisfactory. By mixing a model defect treatment into the likelihood in such a way, the fundamental assumptions made for the model defect are not clear. The consistent treatment of model defects in chapter 9 is well separated from the likelihood, statistically sound, and the basic assumptions are clear.

4.2 Bayesian Monte Carlo

The Bayesian Monte Carlo procedure (BMC) of Koning (2015) deliberately relies on Bayesian statistics. The prior distribution for the model parameters $\pi(\vec{p} | \mathcal{M})$ is given by a uniform distribution. In order to understand the choice of the likelihood $\ell(\vec{\sigma}_{\text{exp}} | \mathcal{M})$, we have to analyze the details of the evaluation method. The BMC method is a Monte Carlo procedure that makes use of importance sampling with a certain kind of self-learning capability. We are interested in the basic assumptions of the statistical model. Thus, we want to know the choice made for the prior and the likelihood. Because the self-learning capability distracts from understanding the likelihood, it will not be discussed here.

Before the self-learning comes into effect, the parameter vectors \vec{p}_i are drawn from the uniform prior distribution. Each parameter vector \vec{p}_i is used to predict the observables measured in the experiments. We denote a specific observable measured in an experiment by $\sigma_{\text{exp},cmk}$. The index c identifies the reaction channel, the index m the experimental data set, and the index k the measurement point within the data set. The corresponding model prediction associated with the parameter vector \vec{p}_i is denoted by $\sigma_{cmk,i}$. The standard error associated with the experimental point $\sigma_{\text{exp},cmk}$ is denoted by $\delta_{\text{exp},cmk}$. The BMC method proceeds by calculating a χ^2 -like quantity for each model prediction based on available

experimental data,

$$\tilde{\chi}_i^2 = \frac{1}{C} \sum_{c=1}^C \sum_{m=1}^{M_c} \frac{1}{N_m} \sum_{k=1}^{N_m} \left(\frac{\sigma_{cmk,i} - \sigma_{\text{exp},cmk}}{\delta_{\text{exp},cmk}} \right)^2, \quad (4.9)$$

with the total number of channels C , the number of experimental datasets M_c for a channel c , and the number of data points N_m in the experimental dataset m . This quantity is used to compute a weight for each model prediction i ,

$$\tilde{\omega}_i \propto \exp(-\tilde{\chi}_i^2). \quad (4.10)$$

We omitted a scaling factor given in the original paper which is of no relevance to determine the likelihood. Its only purpose is to rescale the weights so that they are representable by the standard floating point data type used in computer programs.

Knowledge of the weights $\tilde{\omega}_i$ allows to calculate summary statistics of the posterior distribution. For instance, the evaluated mean vector of cross sections is given by

$$\vec{\sigma}' = \frac{\sum_{i=1}^n \omega_i \vec{\sigma}_i}{\sum_{i=1}^n \omega_i}. \quad (4.11)$$

Similarly, the associated covariance matrix can be determined,

$$\mathbf{A}' = \frac{\sum_{i=1}^n \omega_i (\vec{\sigma}_i - \vec{\sigma}') (\vec{\sigma}_i - \vec{\sigma}')^T}{\sum_{i=1}^n \omega_i}. \quad (4.12)$$

In the following, we discuss the implications of using the χ^2 -like quantity of Equation 4.9. The standard definition of the χ^2 statistics is

$$\chi_i^2 = \sum_{c=1}^C \sum_{m=1}^{M_c} \sum_{k=1}^{N_m} \left(\frac{\sigma_{cmk,i} - \sigma_{\text{exp},cmk}}{\delta_{\text{exp},cmk}} \right)^2. \quad (4.13)$$

If we would use this definition and calculate

$$\omega_i = \exp\left(-\frac{1}{2}\chi_i^2\right), \quad (4.14)$$

we would obtain weights that imply the use of a multivariate normal distribution for the likelihood. To recognize the correspondence, we can combine the model predictions $\sigma_{cmk,i}$ for each parameter vector \vec{p}_i to a vector $\vec{\sigma}_i$, and the experimental data $\sigma_{\text{exp},cmk}$ to the vector $\vec{\sigma}_{\text{exp}}$. Further, we construct a diagonal covariance matrix \mathbf{B} whose elements are $\delta_{\text{exp},cmk}^2$. Now we can verify that

$$-\frac{1}{2}(\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i)^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i) = -\frac{1}{2} \sum_{c=1}^C \sum_{m=1}^{M_c} \sum_{k=1}^{N_m} \left(\frac{\sigma_{cmk,i} - \sigma_{\text{exp},cmk}}{\delta_{\text{exp},cmk}} \right)^2. \quad (4.15)$$

Therefore, using the standard definition of the χ^2 -statistics in Equation 4.13 and the expression for the weights in Equation 4.14, we would indeed weight the samples according to a posterior that is defined by a uniform prior distribution and a multivariate normal likelihood with zero-correlations between the individual experimental data points. The multivariate normal distribution with a diagonal covariance matrix is the proper choice according to the principle of maximum entropy if the correlations are not known. The justification was given at the end of section 3.2.

The previous consideration has shown that the likelihood in the BMC method does not agree with the principle of maximum entropy. In order to get more insight, we analyze Equation 4.9 to understand the assumptions of the implied likelihood. We start with the auxiliary quantity

$$\tilde{\chi}_{cm}^2(\vec{p}) = \frac{1}{N_m} \sum_{k=1}^{N_m} \left(\frac{\mathcal{M}(\vec{p}) - \sigma_{\text{exp},cmk}}{\delta_{\text{exp},cmk}} \right)^2, \quad (4.16)$$

where we substituted $\vec{\sigma}_{cmk,i}$ by the model prediction for an arbitrary parameter \vec{p} . Dividing by N_m means to summarize the information of all points of an experimental dataset in one point. Technically, this measure significantly lowers the $\tilde{\chi}_i^2$ -value compared to the standard definition in Equation 4.13, and hence flattens the likelihood.

Using Equation 4.16, the likelihood of the BMC method has the form

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) \propto \left[\exp \left(-\frac{1}{2} \sum_{c=1}^C \sum_{m=1}^{M_c} \tilde{\chi}_{cm}^2(\vec{p}) \right) \right]^{\frac{2}{C}}. \quad (4.17)$$

Similar to the BFMC method, the $C/2$ -th root is taken of the result of the exponential function, which is another measure to make the likelihood significantly flatter compared to the multivariate normal distribution.

One aim of the BMC method is to make comprehensive use of all available experimental data. According to the principle of maximum entropy, the proper form of the likelihood is a multivariate normal distribution. The proper form combined with a huge amount of data leads to vanishing evaluated uncertainties. Because models are often not perfect descriptions of reality, the evaluated cross sections will not be consistent with experimental data. These findings will be discussed in more detail in chapter 6. The flattened likelihood of the BMC method can be seen as a way to take into account model defects. However, the same objections as in the case of the BFMC method apply, i.e. the treatment of model defects should be statistically sound and based on clear assumptions.

An advantage of the BMC method is that no approximations are introduced concerning the evaluation of the nuclear model. The exact systematics of the

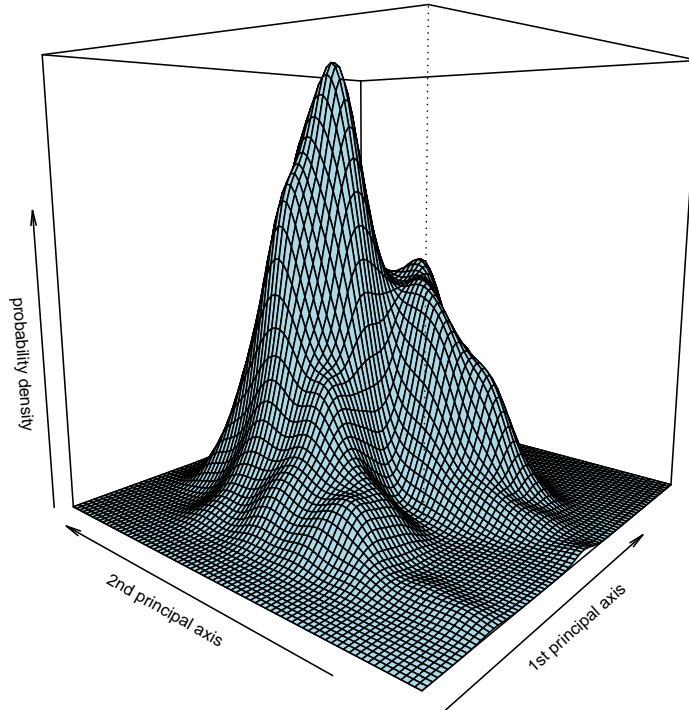


Figure 4.1: Projection of the probability distribution of the neutron-induced total cross sections of ^{181}Ta onto the two directions of highest variance. This distribution at the level of cross sections corresponds to a uniform distribution at the level of the optical model parameters.

model are preserved. Furthermore, assuming a uniform prior distribution for the model parameters is associated with a non-trivial distribution at the level of cross sections. A Bayesian approach based on a surrogate model approximates this non-trivial distribution by a multivariate normal distribution. The BMC method, in contrast, perfectly takes this non-trivial distribution into account. A projection of this distribution for the neutron-induced total cross section of ^{181}Ta is shown in Figure 4.1.

4.3 Unified Monte Carlo-G

Two versions of the Uniform Monte Carlo (UMC) method exist (Capote, Smith, et al., 2012; Smith, 2004). Both versions are based on Bayesian statistics. This section is focused on the original version of Smith (2004), which is termed UMC-G. It can be regarded as a hybrid approach relying on both linearization and Monte Carlo sampling. The nuclear model is replaced by a surrogate model. The details have been discussed in section 3.3. We briefly recapitulate the construction of the prior distribution for the cross sections. First, model parameter vectors \vec{p}_i are

drawn from a uniform distribution.¹ Next, for each parameter vector \vec{p}_i the model prediction $\vec{\sigma}_i = \mathcal{M}(\vec{p}_i)$ is computed. The obtained cross section vectors are used to calculate a center vector $\vec{\sigma}'$ and a covariance matrix \mathbf{A}' . These quantities define a multivariate normal distribution $\mathcal{N}(\vec{\sigma}', \mathbf{A}')$ on the space of cross section vectors, which is used as prior distribution.

Combining this prior distribution with the likelihood, summary statistics such as the mean vector or the covariance matrix are extracted from the posterior distribution by means of a Monte Carlo procedure. The suitability of two sampling algorithms has been studied by Capote and Smith (2008), a brute force algorithm and a Metropolis-Hastings algorithm. Here we only outline the original brute force approach. The results of an in-depth analysis of both approaches can be found in section 7.1. Assume that a cross section vector is given by $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)^T$. An ensemble of m cross section vectors $\vec{\sigma}_i$ is produced. Each cross section vector $\vec{\sigma}_i$ is generated by drawing a value from a uniform distribution for each of its element $\sigma_{i,j}$. The support of the uniform distribution, $\sigma_{\min,j} < \sigma_{i,j} < \sigma_{\max,j}$, has to be sufficiently large to cover the regions of significant posterior probability density. Based on these samples, summary statistics $f(\vec{\sigma})$ can be estimated by

$$E[f(\vec{\sigma})] = \frac{\sum_{i=1}^m f(\vec{\sigma}_i) \pi(\vec{\sigma}_i | \vec{\sigma}_{\text{exp}}, \mathcal{M}_{\text{sur}})}{\sum_{i=1}^m \pi(\vec{\sigma}_i | \vec{\sigma}_{\text{exp}}, \mathcal{M}_{\text{sur}})}, \quad (4.18)$$

where the posterior distribution is given by

$$\pi(\vec{\sigma} | \vec{\sigma}_{\text{exp}}, \mathcal{M}_{\text{sur}}) \propto \ell(\vec{\sigma}_{\text{exp}} | \vec{\sigma}, \mathcal{M}_{\text{sur}}) \pi(\vec{\sigma} | \mathcal{M}_{\text{sur}}). \quad (4.19)$$

For instance, an estimate for the mean vector can be obtained by

$$E[\vec{\sigma}] = \frac{\sum_{i=1}^m \vec{\sigma}_i \pi(\vec{\sigma}_i | \vec{\sigma}_{\text{exp}}, \mathcal{M}_{\text{sur}})}{\sum_{i=1}^m \pi(\vec{\sigma}_i | \vec{\sigma}_{\text{exp}}, \mathcal{M}_{\text{sur}})}. \quad (4.20)$$

As discussed above, the prior distribution is approximated by a multivariate normal prior distribution, $\pi(\vec{\sigma} | \mathcal{M}_{\text{sur}}) \sim \mathcal{N}(\vec{\sigma}', \mathbf{A}')$. In this algorithm, every probability distribution is conditioned upon the surrogate model \mathcal{M}_{sur} . The experimental measurements are contained in $\vec{\sigma}_{\text{exp}}$.

The UMC-G approach is beneficial if the link between the cross sections of the surrogate model $\vec{\sigma}$ and the quantities of the experiment $\vec{\sigma}_{\text{exp}}$ cannot be properly described by a linear transformation

$$\vec{\sigma}_{\text{exp}} = \mathbf{S}\vec{\sigma}. \quad (4.21)$$

The Bayesian methods detailed in section 3.2 and section 3.3 are based on the assumption that such a linear relationship is adequate. However, for instance, if the

¹A multivariate normal distribution would also be an appropriate choice.

ratio of two cross sections was measured in an experiment with large uncertainties, the linear relationship could be inappropriate. An advantage of the BMC-G method is that non-linear relationships between model cross sections and experimental observables can be treated exactly. Another advantage is the possibility to draw samples at high speed due to the use of the simplified surrogate model. The latter advantage, however, represents also a possible disadvantage because non-linear features of the nuclear model are not preserved in the surrogate model (see section 3.3).

Finally, the UMC-G method makes no reference to the use of a specific likelihood. Its features are the use of a surrogate model and a Monte Carlo sampling scheme to obtain summary statistics of the posterior distribution.

4.4 Unified Monte Carlo-B

The other version of the Uniform Monte Carlo method (UMC-B) of Capote, Smith, et al. (2012) succeeds the original UMC-G method. It relies also on Monte Carlo sampling but does not replace the nuclear model by a surrogate model. In this sense, the method is superior, because it preserves the non-linear features of the nuclear model.

In order to introduce the method, we remind that a summary statistics $f(\vec{\sigma})$ of the posterior distribution is given by

$$\mathbb{E}[f(\vec{p})] = \int f(\vec{p}) \pi(\vec{p} | \vec{\sigma}_{\text{exp}}, \mathcal{M}) d\vec{p}. \quad (4.22)$$

The integration has to be performed over the volume of feasible parameter vectors. Summary statistics at the level of cross sections can be computed by using

$$\tilde{f}(\vec{\sigma}) = f(\mathcal{M}(\vec{p})) . \quad (4.23)$$

Summary statistics given as functions of cross sections are of main interest for applications. Simple examples are the expectation vector $\tilde{f}(\vec{\sigma}) = \vec{\sigma}$ or the covariance matrix $\tilde{f}(\vec{\sigma}) = \vec{\sigma}\vec{\sigma}^T - \mathbb{E}[\vec{\sigma}]\mathbb{E}[\vec{\sigma}^T]$. A more complicated example is the expectation value of an observable in an integral measurement. In this case, $f(\vec{\sigma})$ would be defined by nuclear models, material properties, and the geometry of the experiment.

In order to explain the UMC-B method, we have to expand the posterior distribution as a product of prior and likelihood in Equation 4.22,

$$\mathbb{E}[f(\vec{p})] = \int f(\vec{p}) \ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M}) d\vec{p}. \quad (4.24)$$

Because this integral has to be carried out in the high-dimensional space of model parameter vectors, deterministic integration approaches are at troublesome.

Therefore, the UMC-B method relies on Monte Carlo integration. In the UMC-B method, parameter vectors \vec{p}_i are drawn from the prior distribution $\pi(\vec{p} | \mathcal{M})$. Under this circumstance, an approximation of the integral in Equation 4.24 is

$$E[f(\vec{p})] \approx \frac{\sum_{i=1}^n f(\vec{p}_i) \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}{\sum_{i=1}^n \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}. \quad (4.25)$$

Within the Monte Carlo approach, the likelihood does not necessarily have to be a multivariate normal distribution. However, the multivariate normal distribution is in many cases the proper choice (section 3.1). Its assumption would lead to

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M}) \propto \exp\left(-\frac{1}{2}(\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}_i))^T \mathbf{B}^{-1}(\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}_i))\right), \quad (4.26)$$

where \mathbf{B} is the experimental covariance matrix. The determination of the normalization factor of the likelihood is not necessary, because it cancels out in Equation 4.25.

Both the BMC and the UMC-B method rely on Bayesian statistics and use the exact nuclear model without any approximation. Both methods deliberately. The likelihood in the BMC is much flatter than the multivariate normal distribution, which in many situations would be the proper choice. The unconventional choice of likelihood camouflages the problem of model defects. In contrast, the paper of (Capote, Smith, et al., 2012) on the UMC-B methods makes reference to the principle of maximum entropy and specifies probability distributions accordingly. However, the paper on UMC-B only demonstrates the method in a simple toy example with three data points. The application of the UMC-B method in a situation with plenty of high quality experimental data can yield undesirable results. Evaluated cross sections may disagree with experimental data and uncertainties may be unreasonable low due to the use of the nuclear model without any approximation. This issues will be discussed in detail in chapter 6.

4.5 Full Bayesian Evaluation Technique

The Full Bayesian Evaluation Technique (FBET) of Neudecker (2012) is based on Bayesian statistics and implements the surrogate approach detailed in section 3.3. The EMPIRE-MC method (Herman et al., 2007) also implements the surrogate approach. We briefly recapitulate the procedure.

Model parameter vectors \vec{p}_i are drawn from a uniform or multivariate normal distribution. The associated model predictions $\vec{\sigma}_i = \mathcal{M}(\vec{p}_i)$ are computed. Then, a center vector $\vec{\sigma}'$ and a covariance matrix \mathbf{A}' are estimated from the model predictions. These two quantities define a multivariate normal distribution, which is used as the prior distribution. As a side note, the UMC-G method constructs the

prior in the same way. A linear mapping, such as spline interpolation or linear interpolation, is assumed to transfer the cross sections from the model mesh to the incident energies of the experimental data. The further assumption of a multivariate normal distribution for the likelihood leads to analytic update formulas for the center vector and the covariance matrix.

The FBET differs from the EMPIRE-MC method by an extended surrogate model. Instead of the covariance matrix \mathbf{A}' estimated from the model samples, the FBET uses

$$\mathbf{A}'_{\text{FBET}} = \mathbf{A}' + \mathbf{A}_{\text{def}}. \quad (4.27)$$

The matrix \mathbf{A}_{def} accounts for the fact that models can be deficient. Different algorithms for its construction were proposed, see Leeb, Neudecker, and Srdinko (2008). These algorithms will be discussed in detail in chapter 8.

Finally, we mention SAMMY (Larson, 1998) and the EMPIRE-Kalman method (Kawano and Shibata, 1997) which are based on the linearization of the model. Details were given in section 3.2.

We make a brief comparison of the methods based on the linearization of the model (SAMMY and EMPIRE-KALMAN) and the methods based on the surrogate model (EMPIRE-MC and FBET). Both approaches use closed-form update formulas and hence are numerically less involved than the Monte Carlo procedures outlined in the previous sections. The advantage of linearizing the method is that the link to the model parameters is maintained. The advantage of the surrogate approach is the greater flexibility of the simplified model, which leads to better agreement with the included experimental data and more conservative evaluated uncertainties. Therefore, the surrogate approach is less sensitive to the problem of unreasonably low uncertainties. We can interpret the relaxation of the original model as a measure to account for model defects. However, the non-linearity of the model and its accuracy are independent qualities of the model. Therefore, the linearization of the model cannot be regarded as the proper solution to take into account the deficiencies of the model.

Another advantage of the linearized model is the smaller number of model parameters. For instance, the number of optical model parameters for the individual particles is in the order of dozens to hundreds. In contrast, the surrogate approach operates at the level of cross sections, which implies cross sections at dozens to hundred incident energies for dozens to hundreds of reaction channels. Due to the large number of ‘parameters’, the actual application of this approach is limited to the evaluation of angle-integrated cross sections. However, the refined update scheme presented in chapter 5 removes this limitation.

4.6 Summary of methods

The presented methods except one rely on Bayesian statistics. The prior distribution for the model parameters is either the uniform distribution or the multivariate normal distribution. As long as the prior does not restrict the parameter range too much, the evaluated cross sections and uncertainties will be determined by the likelihood—within the possibilities of the model. Under these conditions, the choice of the prior distribution is of minor importance with regard to the result. The likelihood is characterized by the estimates, uncertainties and correlations of the experimental data. It is an important conclusion of the given comparison of methods that the main purpose of the prior is not the specification of best estimates of parameters and associated uncertainties but the conditioning of the evaluated cross sections upon the model systematics. Of course, if only few experimental data points with moderate to high uncertainties are involved, the prior probability distribution for the model parameters takes care of regularization. Values of model parameters should remain within reasonable ranges. Concerning the prior, there is somewhat agreement between the methods.

The first aspect that sets methods apart are the involved approximations. The model can be taken exactly into account. Information about the posterior has then to be obtained through Monte Carlo sampling. If not treating the model exactly, two prevalent approximations exist. First, the linearization of the nuclear model at a specific set of model parameters. And second, sampling from the original model and fitting a multivariate normal distribution based on the ensemble of model predictions. The latter approach can be seen as trying to mimic the exact distribution in cross section space that corresponds to the uniform/multivariate normal distribution in parameter space by a multivariate normal distribution. This approach can be justified by the principle of maximum entropy.

If we trust the model to be a perfect description of reality, using the exact nuclear model is the best approach. All features of the nuclear model are preserved. However, the required Monte Carlo sampling is computationally involved. Methods based on an approximation of the model produce results much faster, but discard non-linear features of the original model. Methods using the linearization of the model are superior, if the inclusion of experimental data narrows uncertainties down to a linear domain of the model. Methods based on the surrogate approach are beneficial if the experimental data does not narrow down the possibilities to the linear domain of the model. Then, non-linear features are at least taken into account as uncertainties. In contrast, the linearized model restricts the possibilities to a tangent hyperplane of the non-linear model manifold in cross section space whose dimension is given by the number of model parameters. Therefore,

the linearized model is more rigid than the surrogate model.

Disagreement between the methods concerns the choice of the probability distribution of the likelihood. The general observation is that deterministic methods always use a multivariate normal distribution for the likelihood. In section 3.1 we elaborated why the multivariate normal distribution is usually the proper choice. Among the Monte Carlo procedures, the UMC-B and UMC-G method also use a multivariate normal likelihood, whereas the BMC and BFMC method make other choices. The UMC-B method can be considered as the most exact method among the presented methods. It is the only method that fulfills two criteria: the nuclear model without any approximation is used and the likelihood is chosen to be a multivariate normal distribution. However, to the knowledge of the author the UMC-B method has only been demonstrated on a toy example with three data points, whereas other Monte Carlo procedures such as BFMC and BMC have already been applied in nuclear data evaluation.

The information of the last paragraph can help to understand why the issue of model defects has not received the attention which it deserves. The surrogate model used in UMC-G, EMPIRE-MC, and FBET is more flexible than the original model. The greater flexibility of the surrogate model alleviates the issue of model deficiencies. However, in some situations unreasonable results still occur, such as negative cross sections at reaction thresholds. Such results are the manifestation of a severe mismatch between data and the model systematics, which means either the model is not perfect or the experimental data are inconsistent.

The Monte Carlo methods BFMC and BMC take the nuclear model exactly into account and have also been applied on real world experimental data. In this chapter it has been shown that the employed likelihood is much flatter than compared to the multivariate normal distribution. Due to this choice, evaluated uncertainties are bigger than if using the multivariate normal distribution. In this thesis it will be demonstrated in chapter 9 that the proper choice of the likelihood leads to unreasonable low uncertainties due to the fact that often nuclear models are not perfect descriptions of reality. The issue becomes most apparent in the situation when taking the nuclear model exactly into account and plenty of high quality experimental data are available. The pioneering ideas of Pigni and Leeb (2003) and Leeb, Neudecker, and Srdinko (2008) regarding model defects are extended to a method whose foundation is fully justified within the framework of Bayesian statistics.

To conclude this chapter, Figure 4.2 gives a graphical overview over evaluation methods and their similarities and differences.

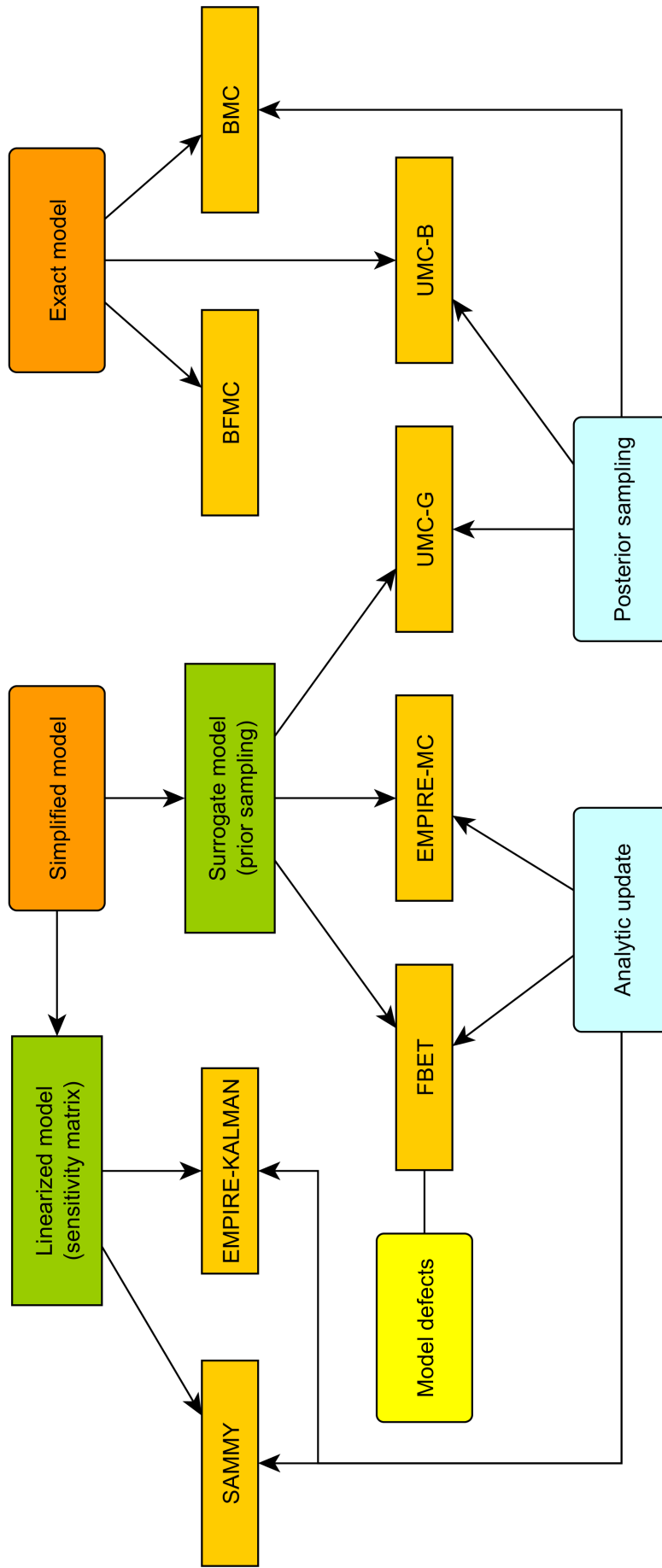


Figure 4.2: Overview over existing methods. The prior can be constructed based on either the exact nuclear model or a simplified model. The simplification of the model can be either the linearization of the nuclear model or the use of a statistical surrogate model at the level of observables. Center vector and covariance matrix of the resulting posterior distribution are obtained either by analytical update formulas or by a Monte Carlo procedure (posterior sampling).

Large number of observables

Nowadays, the significantly increased computer power and the vast amount of available storage space offers exciting possibilities to deal with data of unprecedented complexity. Sophisticated algorithms in internet search engines help to accurately locate the desired information among several hundred of millions of websites in a fraction of a second. Machine learning algorithms scan through databases to generate or test hypotheses, classify information, and detect noteworthy anomalies. Examples range from spam filters of email programs (which often rely on Bayesian statistics) over recommendation systems of internet retailers to algorithms employed by credit institutes to detect fraud based on transaction behavior. Thus it can be expected that many algorithms developed in the field of machine learning dealing with big data will sooner or later find their way into the field of nuclear data evaluation. In this chapter a revised Bayesian update scheme is presented that can be interpreted as a technique to extract the best prediction with regard to experimental information from a database of model calculations. The method can be easily parallelized and the database can be distributed among several computers. This development can be seen as a step into the above mentioned direction. We term the method Large Scale Bayesian Evaluation Technique (LSBET).

Efforts have already been undertaken to harness today's computer power and inexpensive storage space in nuclear data evaluation. Two examples are the Total Monte Carlo (TMC) method (Koning and Rochman, 2008; Rochman, Koning, Cruz, et al., 2010; Rochman, Koning, and Marck, 2009) and the GANDR evaluation tool (Muir et al., 2007). In the TMC method, uncertainties of model parameters are propagated to obtain uncertainties of integral observables, without any approximation of the model or linearization of dependencies. This approach has become feasible only with the increased computer power. The aim of

the GANDR tool is to assist in deciding which measurements could supply the most significant information for the design of new nuclear facilities. Its evaluation methodology relies on the generalized least squares approach (GLSQ) whose formulas are equivalent to those discussed in section 3.3. The outstanding feature is the high dimension (order 90000×90000) of the prior covariance matrix which is completely and consistently updated according to experimental data. This matrix needs about 30 gigabytes of storage space.

The new evaluation method presented in this chapter, which is a main result of this thesis, extends the Bayesian approach discussed in section 3.3 to a large number of observables. Millions or even dozens of millions of observables stemming from a model calculations can be updated with experimental data. The storage requirement of the complete prior covariance matrix would be in the magnitude of hundreds of terabytes or even petabytes. The new approach circumvents its explicit calculation, yet the update formulas are equivalent to the original approach.

Finally, we have to remark here, that the GANDR approach and the new approach to be presented have to be seen as complementary due to different demands on the methods. The basic assumption of the new evaluation scheme is that observables are updated with experimental data in one step, whereas several update steps are possible in the GANDR approach. Furthermore, the assumption of the new approach is that the prior is constructed on the basis of predictions from a nuclear model code. The number of observables that can be updated with the new approach exceeds the possible number of observables in the GANDR approach by order of magnitudes.

5.1 Ensemble representation

The new evaluation scheme is based on the Bayesian update formula of the surrogate approach which has been discussed in section 3.3. The relevant update formulas for the evaluated center vector $\vec{\sigma}_1$ and the covariance matrix \mathbf{A}_1 are

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \mathbf{A}_0 \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S} \vec{\sigma}_0), \quad (5.1)$$

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0 \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} \mathbf{S} \mathbf{A}_0, \quad (5.2)$$

where \mathbf{A}_0 is the prior covariance matrix, $\vec{\sigma}_0$ is the prior best estimate, and \mathbf{S} is the sensitivity matrix transferring the predictions from the model mesh to the incident energies, emitted energies, and angles of the experiment. The measured quantities are contained in $\vec{\sigma}_{\text{exp}}$, and \mathbf{B} is the associated experimental covariance matrix.

The prior best estimate and prior covariance matrix are calculated from an ensemble of model prediction vectors $\vec{\tau}_i$,

$$\vec{\sigma}_0 = \frac{1}{n} \sum_{i=1}^n \vec{\tau}_i \quad \text{and} \quad \mathbf{A}_0 = \frac{1}{n} \sum_{i=1}^n (\vec{\tau}_i - \vec{\sigma}_0) (\vec{\tau}_i - \vec{\sigma}_0)^T. \quad (5.3)$$

We are interested in the consistent evaluation of all reaction observables which can be reliably determined with a nuclear model code, such as TALYS (Koning, Hilaire, and Duijvestijn, 2008), for a set of nuclei of interest. For instance, observables are integrated cross sections, angle-differential cross sections and spectra. An evaluation up to 200 MeV requires a mesh of incident energies with about hundred points. Also the mesh size for the angles and emitted energies has to be sufficiently dense. TALYS using the default settings outputs angle-differential cross sections in steps of two degrees, which leads to a mesh with 90 points. Finally, we assume that the same mesh size is used for emitted energies. At higher energies, hundreds of reactions might be possible. Under these assumptions, the dimension of the prior best guess $\vec{\sigma}_0$ would be in the magnitude of one million. The corresponding covariance matrix \mathbf{A}_0 would be of dimension $10^6 \times 10^6$, which corresponds to a storage requirement of 80 terabytes. Hence, the Bayesian update via Equation 5.1 and Equation 5.2 is not feasible anymore.

An important insight of the considerations given in the previous chapters is the fact that all the required information to perform the update is already contained in the model predictions $\vec{\tau}_i$. If we assume the generation of 10^3 model prediction vectors, each containing 10^6 elements, we obtain 10^9 elements of significant information. Estimating a covariance matrix from these samples inflates the number of elements to 10^{12} without adding any new information.

The idea of the new approach (Schnabel and Leeb, 2015) is to work directly with an ensemble of samples and to circumvent the generation of the prior covariance matrix. The same idea has been successfully applied in other contexts. Examples are the L-BFGS method (Byrd et al., 1995) for optimization and the ensemble Kalman filter (Evensen, 1994).

In order to explain the procedure, we introduce the shifted sample vectors

$$\vec{u}_i = \vec{\tau}_i - \vec{\sigma}_0. \quad (5.4)$$

With this abbreviation the expression of the prior covariance matrix is

$$\mathbf{A}_0 = \frac{1}{n} \sum_{i=1}^n \vec{u}_i \vec{u}_i^T. \quad (5.5)$$

Explicitly inserting this expression into Equation 5.1, we get

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \frac{1}{n} \sum_{i=1}^n \vec{u}_i \vec{u}_i^T \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S} \vec{\sigma}_0). \quad (5.6)$$

The quantity

$$\omega_i = \frac{1}{n} \vec{u}_i^T \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S} \vec{\sigma}_0). \quad (5.7)$$

is just a number and can be considered as a weight. Thus we can cast the update formula Equation 5.6 in the form

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \sum_{i=1}^n \omega_i \vec{u}_i. \quad (5.8)$$

Consequently, the Bayesian update formula is just a linear combination of the shifted sample vectors and therefore also a linear combination of the model predictions $\vec{\tau}_i$. Model features such as the conservation of sum rules or energy balances are linear constraints on the observables, which are fulfilled within every $\vec{\tau}_i$. Because Equation 5.8 is a linear combination of the vectors $\vec{\tau}_i$, also the updated vector will be consistent with regard to the linear constraints. Furthermore, the resulting vector of observables is always located on the hyperplane spanned by the model predictions $\vec{\tau}_i$.

Due to the limited number of model parameters varied for the generation of samples, the effective dimension is usually much smaller than the number of observables. Therefore, the required number of samples has only to be sufficiently larger than the effective dimension of the problem and not larger than the number of observables. The effective dimension is typically in the order of the number of the varied model parameters. For instance, in the case of the optical model the typical number of varied model parameters is about fifty. Experience showed that sufficient convergence is achieved if the number of samples is in the order of thousand. Because of the rather low effective dimension in general no regularization is required to estimate the mean vector and the covariance matrix. However, to be on the safe side, one can check if the mean vector and the standard deviations of its elements are sufficiently converged.

Considering the high number of observables, it is infeasible (or at least unreasonable due to time constraints) to compute the complete posterior covariance matrix \mathbf{A}_1 of the observables. Nevertheless, simulations of nuclear facilities require at least blocks of the updated covariance matrix. For instance, one block could be associated with the covariance matrix of an individual reaction channel. Using the definition of the prior covariance matrix in Equation 5.3 and the shifted sample vectors defined in Equation 5.4, we can rewrite the formula for the posterior covariance matrix, Equation 5.2,

$$\mathbf{A}_1 = \frac{1}{n} \sum_{i=1}^n \vec{u}_i \vec{u}_i^T - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \vec{u}_i \vec{u}_i^T \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} \mathbf{S} \vec{u}_j \vec{u}_j^T. \quad (5.9)$$

We identify again a scalar weight term

$$\omega_{ij} = \frac{1}{n^2} \vec{u}_i^T \mathbf{S}^T (\mathbf{S} \mathbf{A}_0 \mathbf{S}^T + \mathbf{B})^{-1} \mathbf{S} \vec{u}_j \quad (5.10)$$

which allows us to express Equation 5.9 in the compact form

$$\mathbf{A}_1 = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\delta_{ij}}{n} + \omega_{ij} \right) \vec{u}_i \vec{u}_j^T, \quad (5.11)$$

with $\delta_{ij} = 1$ if $i = j$ and zero otherwise. Therefore, both the mean vector and the covariance matrix can be updated by computing linear combinations of the shifted sample vectors or their dyadic products.

To evaluate a block of the posterior covariance matrix, one keeps only the elements of interest in the sample vectors. For instance, to obtain the covariance block given by row indices $1 \dots 49$ and column indices $50 \dots 99$, one uses the reduced sample vectors $\vec{v}_i = (u_{i,1}, \dots, u_{i,49})^T$ and $\vec{w}_j = (u_{j,50}, \dots, u_{j,99})^T$. The respective block of the posterior covariance matrix can then be computed by

$$\mathbf{A}_{(1:49),(50:99)} = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\delta_{ij}}{n} + \omega_{ij} \right) \vec{v}_i \vec{w}_j^T. \quad (5.12)$$

One may doubt the benefit of rewriting the Bayesian update formulas in this way, because the computation of the weights still requires to evaluate the product $\mathbf{S} \mathbf{A}$. For instance, consider 10^3 experimental data points and 10^6 observables to update. The sensitivity matrix would be of dimension $10^3 \times 10^6$ and the prior covariance matrix of dimension $10^6 \times 10^6$. The time complexity for the required multiplication $\mathbf{S} \mathbf{A}$ would be $10^3 \times 10^6 \times 10^6 = 10^{15}$. Taking into account that a contemporary desktop computer is capable of about 10^{10} instructions per second, the multiplication would take 10^5 seconds, which is approximately one month. In addition, the previous consideration neglects the fact that a covariance matrix of 80 terabytes does not fit into working memory and has to be loaded piecewise. Access times and throughputs of hard disks are magnitudes smaller than those of the working memory. The next section shows that harnessing the sparsity of the sensitivity matrix and the ensemble representation of the covariance matrix allows for the fast computation of the weights.

The important statement of this section is that—once the weights are available—both updating the complete mean vector for millions of observables and computing blocks of the posterior covariance matrix are feasible. Updating the mean vector of ten million observables may take less than half an hour on a modern desktop computer.

5.2 Calculating the weights

The direct calculation of the weights is infeasible for a large number of observables because of the required determination of the involved matrix product $\mathbf{S}\mathbf{A}_0\mathbf{S}^T$. However, the computation of the weights is well manageable if making use of the ensemble representation of the covariance matrix and exploiting the sparsity of the sensitivity matrix \mathbf{S} .

Starting point of the procedure are the expressions for the weights Equation 5.7 and Equation 5.10. We use the representation Equation 5.5 of \mathbf{A}_0 to express $\mathbf{S}\mathbf{A}_0\mathbf{S}^T$,

$$\omega_i = \frac{1}{n} (\mathbf{S}\vec{u}_i)^T \left(\frac{1}{n} \sum_{k=1}^n (\mathbf{S}\vec{u}_k) (\mathbf{S}\vec{u}_k)^T + \mathbf{B} \right)^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\sigma}_0), \quad (5.13)$$

$$\omega_{ij} = \frac{1}{n^2} (\mathbf{S}\vec{u}_i)^T \left(\frac{1}{n} \sum_{k=1}^n (\mathbf{S}\vec{u}_k) (\mathbf{S}\vec{u}_k)^T + \mathbf{B} \right)^{-1} (\mathbf{S}\vec{u}_j). \quad (5.14)$$

Each vector $\vec{r}_i = \mathbf{S}\vec{u}_i$ contains the model prediction associated with the i^{th} calculation interpolated to the energies and angles of the experimental data. The number of available experimental data points is usually in the order of thousands, whereas the number of values of observables predicted by the nuclear model code may exceed several millions. If we assume that the vectors \vec{r}_i are available, the inversion is the most time consuming operation. The inversion is performed on the mesh of the experimental data, which means that the dimension of the matrix is in the magnitude 1000×1000 . The inversion of such a matrix takes not more than half a minute on contemporary desktop computers.

Another important aspect is that the expression

$$\mathbf{X} = \left(\frac{1}{n} \sum_{k=1}^n (\mathbf{S}\vec{u}_k) (\mathbf{S}\vec{u}_k)^T + \mathbf{B} \right)^{-1} \quad (5.15)$$

has to be evaluated only once and the result can then be reused for the computation of every weight. Also the expression $\vec{r}_0 = (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\sigma}_0)$ can be reused. Using these abbreviations, the formulas for the weights become particularly succinct,

$$\omega_i = \frac{\vec{r}_i^T \mathbf{X} \vec{r}_0}{n} \quad \text{and} \quad \omega_{ij} = \frac{\vec{r}_i^T \mathbf{X} \vec{r}_j}{n^2}. \quad (5.16)$$

We note that once $\vec{a} = \mathbf{X}\vec{r}_0$ is computed, the weights for updating the mean vector are given by $\omega_i = \vec{r}_i^T \vec{a}$. The dimension of the vectors in this scalar product equals the number of experimental data points. If M is the number of experimental data points and N the number of model predictions, the time complexity to evaluate all weights ω_i is $M \times N$. In a common scenario with about 10^3 experimental

data points and 10^3 model predictions, the computation time can be neglected. The most expensive operation that has to be repeated for every weight ω_{ij} is the computation of the matrix product $\mathbf{X}\vec{r}_j$. For 10^3 experimental data points, the time complexity is $10^3 \times 10^3 = 10^6$. Furthermore, assuming 10^3 samples, we get a time complexity of $10^6 \times 10^3 = 10^9$ to compute all weights. Recalling that a modern desktop computer can perform about 10^{10} instructions per second, a conservative guess of the computation time is a few seconds.

The discussion so far elaborated on the calculation of the weights once the vectors $\vec{r}_i = \mathbf{S}\vec{u}_i$ are available. The remainder of this section deals with the calculation of the \vec{r}_i . In section 3.4 the construction of the sensitivity matrix for integrated cross sections using linear interpolation and spectra using bilinear interpolation was explained. To recapitulate the mapping scheme, consider an experimental data vector $\vec{\sigma}_{\text{exp}}$ with observations (integrated and differential cross sections) at certain energies and angles. Further consider a vector of model predictions $\vec{\sigma}_{\text{mod}}$ at different energies and angles. Using linear Equation 3.40 and bilinear interpolation Equation 3.59, an interpolation from the model mesh to the experimental mesh always has the form

$$\sigma_i^{\text{int}} = \alpha_{i1}\sigma_{\beta_{i1}}^{\text{mod}} + \alpha_{i2}\sigma_{\beta_{i2}}^{\text{mod}} + \alpha_{i3}\sigma_{\beta_{i3}}^{\text{mod}} + \alpha_{i4}\sigma_{\beta_{i4}}^{\text{mod}}. \quad (5.17)$$

The observables in the vector $\vec{\sigma}^{\text{int}}$ are specified at the same energies and angles as the observables in $\vec{\sigma}_{\text{exp}}$. Each interpolated data point is given as the weighted sum of up to four elements of the model prediction vector $\vec{\sigma}_{\text{mod}}$. For linear interpolation only two elements of $\vec{\sigma}_{\text{mod}}$ are involved, hence one may assign $\alpha_{i3} = \alpha_{i4} = 0$ and $\beta_{i3} = \beta_{i4} = 1$. The constants α_{ik} can be arranged into a matrix $\boldsymbol{\alpha}$. Also the β_{ik} can be arranged into a matrix $\boldsymbol{\beta}$. The i^{th} row of these matrices contains the information to compute the interpolation σ_i^{int} associated with the i^{th} element of $\vec{\sigma}_{\text{exp}}$. The dimension of both matrices is $M \times 4$ where M is the number of observations in $\vec{\sigma}_{\text{exp}}$. The interpolation is computationally inexpensive when the matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are given. Only $M \times 4$ multiplications have to be performed to obtain the complete interpolated vector.

In order to determine $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we introduce the index table \mathcal{I} associated with a cross section vector $\vec{\sigma}$. The index table contains the reaction channel, incident energy, and outgoing energy or angle of each observable included in $\vec{\sigma}$. A complete characterization of observables requires both the index table and the cross section

vector. As a schematic example, consider the pair

$$\vec{\sigma} = \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_{d-1} \\ \sigma_d \end{pmatrix}, \quad \mathcal{I} = \begin{array}{|c|c|c|c|} \hline \mathbf{Pos} & \mathbf{Reac} & \mathbf{Einc} & \mathbf{Eout} \\ \hline 1 & \text{CS/TOT} & 1.5 & 0 \\ 2 & \text{CS/TOT} & 2.0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ d-1 & \text{SP/PART/n} & 8.0 & 5.5 \\ d & \text{SP/PART/n} & 8.0 & 6.5 \\ \hline \end{array}. \quad (5.18)$$

In this example, the index table tells us that the first element of the cross section vector is the total cross section at 1.5 MeV incident energy. The reaction channel of an element in the cross section vector is identified by a reaction string. The reaction string also gives the information whether the associated element in the cross section vector is an integrated cross section, a spectrum value or an angle-differential cross section. The example shows that the reaction string of integrated cross sections starts with ‘CS’. Because no emission energy has to be specified then, the entry in the column ‘Eout’ is set to zero. The last column was labeled ‘Eout’ for brevity. If the reaction string indicates an angle-differential cross section, the value in the last column represents the angle.

In the following we consider an experimental vector $\vec{\sigma}_{\text{exp}}$ with index table \mathcal{I}_{exp} and a model vector $\vec{\sigma}_{\text{mod}}$ with index table \mathcal{I}_{mod} . The construction of the matrices α and β can be performed solely on the basis of \mathcal{I}_{mod} and \mathcal{I}_{exp} . For a fast construction of α and β , the index table \mathcal{I}_{mod} is sorted primarily according to ‘Reac’, secondarily according to ‘Einc’ and tertiarily according to ‘Eout’. For two arbitrary rows i and j of \mathcal{I}_{mod} , it can be decided whether they are related by

$$(\text{Reac}, \text{Einc}, \text{Eout})_i < (\text{Reac}, \text{Einc}, \text{Eout})_j \quad (5.19)$$

or not. Due to the fact that the rows in \mathcal{I}_{mod} are sorted, a binary search (also known as bisection method) can be applied to quickly obtain the relevant rows. The binary search is defined by the recursive function

$$\text{bs}(x, r_1, r_2, \mathcal{I}) := \begin{cases} \text{bs}(x, r_1, \lfloor \frac{r_1+r_2}{2} \rfloor, \mathcal{I}) & \text{if } x < \text{row}(\mathcal{I}, \lfloor \frac{r_1+r_2}{2} \rfloor) \\ \text{bs}(x, \lfloor \frac{r_1+r_2}{2} \rfloor, r_2, \mathcal{I}) & \text{if } x > \text{row}(\mathcal{I}, \lfloor \frac{r_1+r_2}{2} \rfloor) \\ r_1 & \text{otherwise} \end{cases} \quad (5.20)$$

where x denotes a triple (Reac,Einc,Eout) to which the information specified in \mathcal{I} should be mapped. The row numbers r_1 and r_2 define the boundaries of the range in which the search has to be performed. Finally, the notation $\text{row}(\mathcal{I}, i)$ represents the content of the i^{th} row of \mathcal{I} . The function $\text{bs}(\cdot)$ returns the number of the row in \mathcal{I} whose content is closest from below to x out of all rows. The binary search is schematically depicted in Figure 5.1.

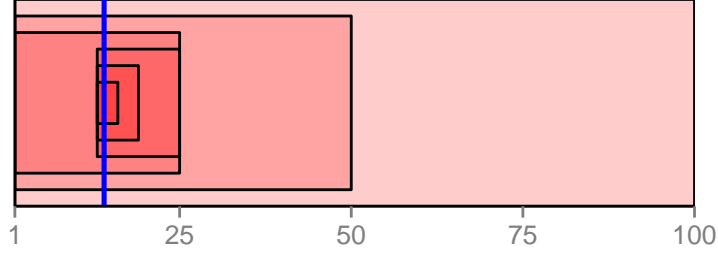


Figure 5.1: Illustration of binary search. Because rows are sorted, a particular row can be found by recursively halving the search interval.

The matrices α and β are constructed row-wise. For every row i in \mathcal{I}_{exp} , the relevant rows in \mathcal{I}_{mod} are determined. The obtained row positions are then stored in the i^{th} row of β . Based on the information in the relevant rows of \mathcal{I}_{mod} , the associated coefficients in Equation 5.17 are computed and stored in the i^{th} row of the matrix β . Following we describe the process to determine one row in α and β for an integrated cross section. The algorithm is analogous for a spectrum value or angle-differential cross section. Assume that the i^{th} row of \mathcal{I}_{exp} contains $Reac = CS/TOT$, $E = Einc = 5.0$, and $Eout = 0.0$. To construct a linear mapping from a model prediction vector to this observation of the experiment, two rows in \mathcal{I}_{mod} have to be located. Both rows must contain $Reac = CS/TOT$. One row must contain the incident energy E_1 which is closest from below to the incident energy $E = 5.0$ of the experimental data point. The other row must contain the incident energy E_2 which is closest from above to the incident energy of the experimental data point. Because \mathcal{I}_{mod} is sorted, these two rows are adjacent. The location of this pair is determined by a binary search. Using the binary search function defined in Equation 5.20, the relevant rows in \mathcal{I}_{mod} can be determined by $k = \text{bs}(\text{row}(\mathcal{I}_{\text{exp}}, i), 1, n, \mathcal{I}_{\text{mod}})$, where n is the total number of rows in \mathcal{I}_{mod} . Let the incident energy in the k^{th} and $(k+1)^{\text{th}}$ row be E_k and E_{k+1} respectively. Now, the i^{th} row of β is

$$\beta_i = (k; k + 1; 1; 1). \quad (5.21)$$

The coefficients for linear interpolation are written in the i^{th} row of α ,

$$\alpha_i = \left(\frac{E_{k+1} - E}{E_{k+1} - E_k}; \frac{E - E_k}{E_{k+1} - E_k}; 0; 0 \right). \quad (5.22)$$

The last two coefficients are set to zero, because only two elements of the model prediction vector are involved in linear interpolation.

The binary search is very efficient. For instance, if \mathcal{I}_{mod} contains 10^6 rows, only twenty comparisons are needed to locate the rows in \mathcal{I}_{mod} that are relevant to map a model prediction vector to one row in \mathcal{I}_{exp} . In general, the time complexity to

compute α and β is proportional to $M \times \log(N)$ where M is the number of rows in \mathcal{I}_{exp} and N is the number of rows in \mathcal{I}_{mod} . Therefore, the mapping can be quickly performed even for $N \simeq 10^7$.

This section showed that the calculation of the weights is computationally expensive. In fact, the bottleneck is the storage device. For instance, storing 10^3 model prediction vectors with 10^7 observables in double precision requires 80 gigabytes. The working memory of typical contemporary desktop computers is too small to keep all model prediction vectors at once. Thus, model predictions have to be read piecewise from the hard disk. Hence, the time demand is primarily determined by the access speed of the storage device. Nevertheless, the whole procedure is well-manageable.

5.3 Interpretation as Gaussian process

Gaussian process regression is a powerful method for non-parametric regression and well interpretable within the framework of Bayesian statistics. In this section we show that the described procedure to update a large number of observables can be interpreted as a Gaussian process regression. This interpretation will significantly simplify the treatment of model defects in chapter 9. The notation we introduce in this section will also be helpful to explain the database approach introduced in the next section. A more formal introduction of Gaussian processes will be given in section 9.1.

A Gaussian process can be regarded as the generalization of a multivariate normal distribution. A multivariate normal distribution is a probability distribution for a finite number of random variables. A Gaussian process is a probability distribution for an infinite number of random variables. A collection of infinitely many random variables can be regarded as a random function $f(x)$, where the value of the function for any permissible choice of x is a random variable. The defining property of a Gaussian process is that for every possible finite set $\{x_1, x_2, \dots, x_n\}$ of values of the function argument, the associated random variables $f(x_1), f(x_2), \dots, f(x_n)$ are governed by a multivariate normal distribution.

The scheme for updating a large number of observables is based on the surrogate approach discussed in section 3.3. In this approach the prior is constructed by fitting a multivariate normal distribution to an ensemble of model prediction vectors. In order to recognize the link to Gaussian processes, we restrict ourselves to model prediction vectors which contain only the values of the total cross section at different incident energies E_i . Therefore, also the fitted multivariate normal distribution refers only to the total cross section. The total cross section is

a function of the incident energy. Prior estimates of total cross sections in-between the incident energies of the model mesh are obtained by linear interpolation. This means that the prior mean vector $\vec{\sigma}$ with associated energies \vec{E} is augmented to a continuous function $\sigma(E)$ via linear interpolation. Linear interpolation represents a linear transformation of the random variables (the cross section values) defined at the energies of the model mesh. As pointed out on section 2.2, a linear transformation of a multivariate normal random vectors yields another multivariate normal random vector. Therefore, the assumption of a multivariate normal distribution for the model vector together with linear interpolation defines a Gaussian process. The same argument applies for bilinear interpolation of spectra and angle-differential cross sections.

Having established the link to Gaussian processes, one might refrain from considering the model prediction *vector* as object of primary interest and instead considers a model prediction *function*. The scheme to update a large number of observables conforms well to this interpretation. In section 5.2, we discussed efficient interpolation schemes to determine from the vector of model predictions the corresponding values of the model on the mesh of an experiment vector $\vec{\sigma}_{\text{exp}}$. Instead of building up the complete sensitivity matrix \mathbf{S} and perform the matrix multiplication $\mathbf{S}\vec{\sigma}_{\text{mod}}$, only the operations required for linear and bilinear interpolation are carried out. We emphasize this fact by introducing the mapping operator

$$\hat{S}(\mathcal{I}_1, \mathcal{I}_2, \vec{v}) \equiv \mathbf{S}\vec{v} \quad (5.23)$$

whose action is equivalent to the evaluation of the matrix product $\mathbf{S}\vec{v}$. The index table \mathcal{I}_2 specifies the meaning of the values in \vec{v} whereas the index table \mathcal{I}_1 specifies the mesh to which values should be interpolated. An example for the structure of an index table was given in Equation 5.18.

Using the mapping operator and Equation 5.3, we can compute the prior estimates of observables at arbitrary energies and/or angles

$$\sigma_0(\mathcal{I}) = \hat{S}(\mathcal{I}, \mathcal{I}_{\text{mod}}, \vec{\sigma}_0). \quad (5.24)$$

The vector $\vec{\sigma}_0$ contains the prior estimates of the observables on the model mesh. The index table \mathcal{I}_{mod} describes the reaction type, energies, and angles associated with values in $\vec{\sigma}_0$. The index table \mathcal{I} specifies the observables at arbitrary energies or angles within the boundaries of the mesh. The mapping operator takes care of the proper interpolation from the discretized energy mesh to the specified energies in \mathcal{I} . The discrete nature of the energy and angle meshes is hidden in the internals of $\sigma_0(\mathcal{I})$. From the outside, the object $\sigma_0(\mathcal{I})$ appears as a vector-valued function with several arguments, given that we keep the number of rows in \mathcal{I} constant and

only allow the adjustment of the energies and angles. Therefore, we can regard $\sigma_0(\mathcal{I})$ as the *prior mean function*.

Similarly, prior covariances between observables at arbitrary energies and angles are given by

$$A_0(\mathcal{I}_1, \mathcal{I}_2) = \frac{1}{n} \sum_{i=1}^n \hat{S}(\mathcal{I}_1, \mathcal{I}_{\text{mod}}, \vec{u}_i) \hat{S}(\mathcal{I}_2, \mathcal{I}_{\text{mod}}, \vec{u}_i)^T, \quad (5.25)$$

where \vec{u}_i are the shifted sample vectors defined in Equation 5.4, and n is the number of sample vectors. Again, from the outside $A_0(\mathcal{I}_1, \mathcal{I}_2)$ appears as a vector-valued function if the number of rows in \mathcal{I}_1 and \mathcal{I}_2 are kept fixed, and we can regard it as the prior covariance function.

The updated quantities correspond also to a Gaussian process. Using the mapping operator and Equation 5.8, the posterior mean function has the form

$$\sigma_1(\mathcal{I}) = \hat{S}(\mathcal{I}, \mathcal{I}_{\text{mod}}, \vec{\sigma}_0) + \sum_{i=1}^n \omega_i \hat{S}(\mathcal{I}, \mathcal{I}_{\text{mod}}, \vec{u}_i), \quad (5.26)$$

with the weights ω_i defined in Equation 5.7. Based on Equation 5.11, the posterior covariance function is given by

$$A_1(\mathcal{I}_1, \mathcal{I}_2) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\delta_{ij}}{n} + \omega_{ij} \right) \hat{S}(\mathcal{I}_1, \mathcal{I}_{\text{mod}}, \vec{u}_i) \hat{S}(\mathcal{I}_2, \mathcal{I}_{\text{mod}}, \vec{u}_j)^T \quad (5.27)$$

with the weights ω_{ij} defined in Equation 5.10. The efficient computation of the weights has been discussed in section 5.2.

This section demonstrated that the surrogate approach, and hence also the scheme to update a large number of observables, can be interpreted as Gaussian process regression. Both the prior and the posterior can be interpreted as Gaussian processes. This interpretation will be tremendously helpful in chapter 9 to define consistent model defects. Once the technical details such as the implementation of the mapping operator are solved, the Bayesian update becomes very simple. The next section elaborates on these technical details.

5.4 Database approach

The basis to perform the Bayesian update for a large number of observables are the model prediction vectors $\vec{\tau}_i$. Each of them contains the observables predicted by a nuclear model using specific values of the model parameters. The first important design consideration is the storage format of the ensemble of model prediction vectors $\vec{\tau}_i$ with the associated index tables $\mathcal{I}_{\text{mod},i}$. We discuss two options.

A straight-forward approach is to generate a model prediction matrix

$$\boldsymbol{\tau} = (\vec{\tau}_1, \vec{\tau}_2, \dots, \vec{\tau}_n), \quad (5.28)$$

where each column is given by a particular model prediction vector $\vec{\tau}_i$. To construct such a matrix, every $\vec{\tau}_i$ must contain the same number of elements. In addition, it is reasonable to demand that all $\vec{\tau}_i$ possess the same structure. Therefore, all index tables $\mathcal{I}_{\text{mod},i}$ are identical and we just write \mathcal{I}_{mod} . The matrix $\boldsymbol{\tau}$ can be stored in a single file and the index table \mathcal{I}_{mod} in another file. For vectors containing 10^7 variables, the memory requirement of the index table \mathcal{I}_{mod} is about 300 megabytes and fits completely into working memory. The binary search, see Equation 5.20, allows a quick determination of the relevant rows in $\boldsymbol{\tau}$ for the Bayesian update. For instance, to calculate the weights based on 10^3 experimental cross sections involves up to 2000 rows of $\boldsymbol{\tau}$. If stored suitably, individual rows of $\boldsymbol{\tau}$ can be read at once. Conventional hard disks have a throughput about hundred megabytes per second. Assuming 10^3 model prediction vectors equivalent to 10^3 columns, reading one row takes about 0.1 milliseconds. However, the access time to move the read head to the start of the row is in the magnitude of milliseconds. Therefore, the time to retrieve the values of 2000 rows is in the magnitude of dozens of seconds. Due to the regular structure of a matrix, this storage scheme is computationally efficient. Splitting the matrix $\boldsymbol{\tau}$ into blocks and distributing these blocks across several hard disks can help to speed up the computation. Relevant information from the model prediction vectors can then be retrieved in parallel. The disadvantage of this storage scheme is its inflexibility. Either adding new model prediction vectors to $\boldsymbol{\tau}$ or augmenting the existing vectors with a new observable is an expensive operation. The situation is illustrated in Figure 5.2.

The lack of flexibility is a severe drawback of this storage format. For instance, assume that observables were collected from model calculations and stored into such a matrix. While gathering experimental data, one finds data for a reaction channel that has not been included in the model prediction matrix $\boldsymbol{\tau}$. Extracting the corresponding observables from the model calculations, and adding them to $\boldsymbol{\tau}$ takes time. Eventually, hundreds of gigabytes have to be written again to hard disk. Further, the number of provided observables of a model code may vary with the choice of model parameters. At which incident energy reaction channels open up may depend on model parameters. It then becomes difficult to define a common mesh of observables, which is necessary to store the observables of the model calculations as a matrix.

Therefore, another storage format based on the original output of the model code is proposed which avoids the drawbacks discussed above. The idea of the other approach is to work with the original output of the nuclear model code. We

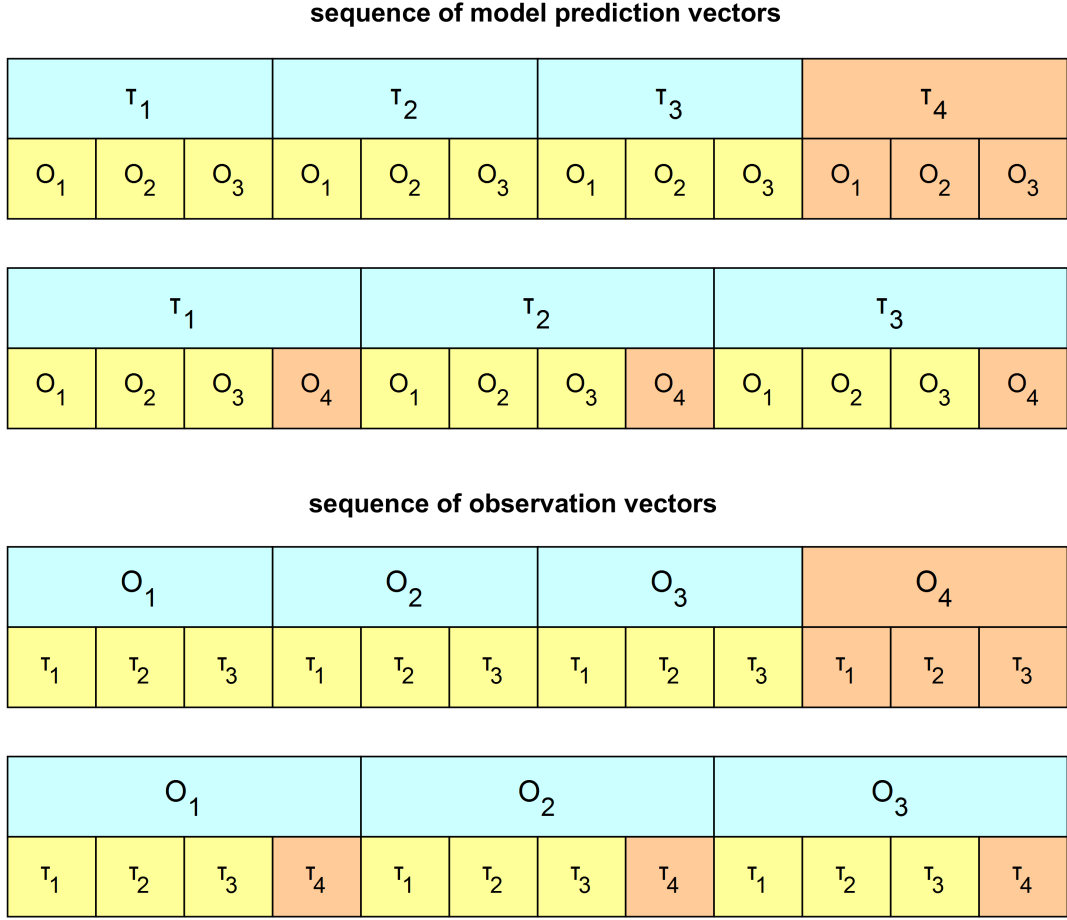


Figure 5.2: Ordering schemes of the matrix containing the model prediction vectors. If each block contains the observables O_i associated with one model prediction $\vec{\tau}_j$, the insertion of a new model prediction at the end of the chain is fast, whereas amending an observation to every prediction slow. In the latter case, the whole structure has to be copied. Runtime complexities of these two insertion operations are switched, if each block contains the same observable O_i from all the model prediction vectors $\vec{\tau}_j$.

assume that the output of each model calculation based on model parameter set \vec{p}_i is located in a directory D_i on the hard disk. The output files in one directory D_i contain all the information required to build up the model prediction vector $\vec{\tau}_i$ and also the index table $\mathcal{I}_{\text{mod},i}$. Based on the mapping operator in Equation 5.23 we introduce another operator \hat{T} ,

$$\hat{T}(\mathcal{I}, D_i) \equiv \hat{S}(\mathcal{I}, \mathcal{I}_{\text{mod},i}, \vec{\tau}_i). \quad (5.29)$$

The operator \hat{T} extracts the relevant information from directory D_i and applies an interpolation scheme to produce the predictions of the observables specified in \mathcal{I} . The operator \hat{T} does not require $\mathcal{I}_{\text{mod},i}$ or $\vec{\tau}_i$ as arguments, because this information is available in the files contained in D_i . For convenience, we further

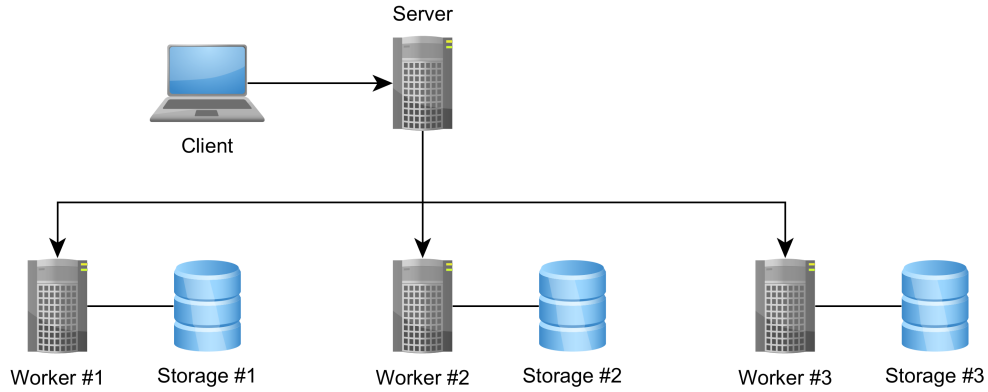


Figure 5.3: The update scheme for a large number of observables as database application. This diagram shows the associated computer infrastructure.

define

$$\hat{T}(\mathcal{I}, D_i - D_0) := \hat{T}(\mathcal{I}, D_i) - \hat{T}(\mathcal{I}, D_0), \quad (5.30)$$

where D_0 is the directory that contains the model results based on the prior best parameter set \vec{p}_0 .

In the diction of programming, the operator \hat{T} is a function. The form of Bayesian update is unaffected by the operations within this function. Due to this fact, the approach becomes very flexible. If linear interpolation is not desirable, one can rewrite \hat{T} to implement some other interpolation scheme. If a new version of the nuclear model code is made available which predicts a new type of observable, it suffices to extend the function to read from the appropriate files.

The update scheme for a large number of observables can be implemented as a database application, which can work with a database distributed across several computers. A client requests specific evaluated observables based on an experimental data set from the server. The server distributes the computation tasks among several worker computers, each connected to a storage device containing a subset of the model calculation directories. Partial results are sent back to the server and combined there. The final result is transmitted back to the client. The layout of the described infrastructure is depicted in Figure 5.3.

In the following, we describe all steps involved starting from the client request and ending with sending back the evaluated observables. Let C_i denote the i^{th} worker computer. Further let \mathcal{D}_i denote the set of directories located on the storage device connected to C_i . The directory D_0 shall contain the output files of the calculation made with the prior best parameter set and is stored on the server.

1. The client requests an evaluation of the observables specified in $\mathcal{I}_{\text{eval}}$ based on experimental data $\vec{\sigma}_{\text{exp}}$, \mathcal{I}_{exp} and associated covariance matrix \mathbf{B} . This

information is sent to the server.

2. The server sends to each worker computer C_i the request to compute $\hat{T}(\mathcal{I}_{\text{exp}}, D)$ and $\hat{T}(\mathcal{I}_{\text{eval}}, D)$ for all $D \in \mathcal{D}_i$. The resulting vectors $\vec{\tau}_k$ and $\vec{\sigma}_k$, respectively, are sent back to the server.
3. The server computes the shifted prediction vectors $\vec{r}_k = \vec{\tau}_k - \hat{T}(\mathcal{I}_{\text{exp}}, D_0)$, and based on them the expression in Equation 5.15

$$\mathbf{X} = \left(\frac{1}{n} \sum_{k=1}^n \vec{r}_k \vec{r}_k^T + \mathbf{B} \right)^{-1}. \quad (5.31)$$

4. The server computes the expression $\vec{a} = \mathbf{X} (\vec{\sigma}_{\text{exp}} - \hat{T}(\mathcal{I}_{\text{exp}}, D_0))$. Then, the weights to update the mean vector and the covariance matrix are calculated,

$$\omega_i = \frac{\vec{r}_i^T \vec{a}}{n} \quad \text{and} \quad \omega_{ij} = \frac{\vec{r}_i^T \mathbf{X} \vec{r}_j}{n^2}. \quad (5.32)$$

5. The server computes the shifted prediction vectors $\vec{u}_k = \vec{\sigma}_k - \hat{T}(\mathcal{I}_{\text{eval}}, D_0)$. These vectors are used to evaluate the requested observables,

$$\vec{\sigma}_{\text{eval}} = \hat{T}(\mathcal{I}_{\text{eval}}, D_0) + \sum_{k=1}^n \vec{u}_k \omega_k, \quad (5.33)$$

and the associated covariance block

$$\mathbf{A}_{\text{eval}} = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\delta_{ij}}{n} + \omega_{ij} \right) \vec{u}_i \vec{u}_j^T. \quad (5.34)$$

6. The evaluated observable vector $\vec{\sigma}_{\text{eval}}$ and the associated covariance matrix \mathbf{A}_{eval} are sent back to the client.

Parallelization is not restricted to the retrieval of information of the directories which contain the model predictions and interpolation of this information (step 2). Calculating the sums in step three and five, and the calculation of the weights in step four can also be parallelized. However, the speed gain achieved by these measures is much smaller than achieved by the parallelization in step two. As has been pointed out in section 5.2, the time limiting factor of the procedure is the time needed to retrieve the information from the storage devices. Compared to this time, the evaluation of the sums is negligible.

If such a server/worker infrastructure is not available, the whole procedure can be carried out on a single computer. The directories may be still distributed across several hard disks. The use of solid state disks whose random access time is about two magnitudes lower than those of conventional hard disks makes possible to further reduce the overall time of the procedure.

5.5 Comparison to Monte Carlo methods

The update scheme for a large number of observables (LSBET) bears a strong resemblance to the Monte Carlo methods, such as BMC and UMC-G, discussed in chapter 4. For a comparison of the methods, we insert the shifted sample vector defined in Equation 5.4 into Equation 5.3,

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \sum_{i=1}^n \omega_i (\vec{\tau}_i - \vec{\sigma}_0), \quad (5.35)$$

where $\vec{\sigma}_0$ is the prior best estimate and $\vec{\tau}_i$ are the model prediction vectors. The model prediction vectors are obtained by drawing model parameter vectors \vec{p}_i from the prior distribution and then computing the associated model predictions, $\vec{\tau}_i = \mathcal{M}(\vec{p}_i)$.

If we use the arithmetic mean to construct the prior best estimate $\vec{\sigma}_0$,

$$\vec{\sigma}_0 = \frac{1}{n} \sum_{i=1}^n \vec{\tau}_i, \quad (5.36)$$

the formula to update the mean vector, Equation 5.35, can be written as

$$\vec{\sigma}_1 = \frac{1}{n} \sum_{i=1}^n \vec{\tau}_i + \sum_{i=1}^n \vec{\tau}_i \omega_i - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \vec{\tau}_i \omega_j = \quad (5.37)$$

$$= \sum_{i=1}^n \vec{\tau}_i \left[\omega_i + \frac{1}{n} \left(1 - \sum_{j=1}^n \omega_j \right) \right] = \sum_{i=1}^n \vec{\tau}_i \tilde{\omega}_i. \quad (5.38)$$

Thus, the updated mean vector is a weighted sum of the model prediction vectors obtained by sampling from the prior distribution of the model parameters. Also in Monte Carlo methods, such as UMC-B, an estimate of the mean vector is given by a weighted sum of the model prediction vectors drawn from the prior distribution. The formula to estimate statistics from model prediction vectors in Monte Carlo procedures is given in Equation 4.25. Especially, the mean value is given by

$$\vec{\sigma}_1^{\text{MC}} = \sum_{i=1}^n \vec{\tau}_i \left(\frac{\omega_i^{\text{MC}}}{\sum_{i=1}^n \omega_i^{\text{MC}}} \right) = \sum_{i=1}^n \vec{\tau}_i \tilde{\omega}_i^{\text{MC}} \quad (5.39)$$

In the update scheme for a large number of observables, the likelihood is a multivariate normal distribution. Using the same likelihood in a Monte Carlo procedure, the weights ω_i^{MC} become

$$\omega_i^{\text{MC}} \propto \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathbf{S} \vec{\tau}_i)^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S} \vec{\tau}_i) \right\}, \quad (5.40)$$

with the vector of measurements $\vec{\sigma}_{\text{exp}}$ and the associated covariance matrix \mathbf{B} . The sensitivity matrix \mathbf{S} maps the model predictions in $\vec{\tau}_i$ to the mesh of $\vec{\sigma}_{\text{exp}}$.

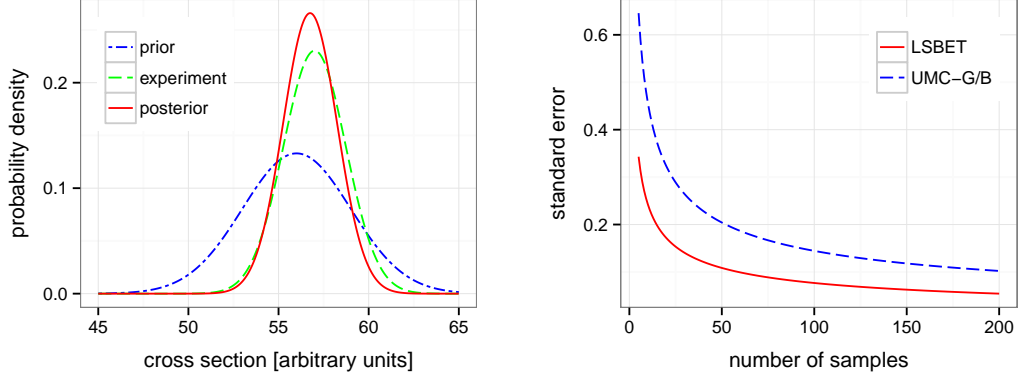


Figure 5.4: Distributions of the first schematic scenario are shown in the left diagram. A comparison of UMC-G/B and LSBET in terms of efficiency is shown in the right diagram.

This expression for the weights is very different from Equation 5.13. In order to understand the consequence of this difference, we discuss two schematic scenarios.

In the first scenario, we consider the update of a single cross section σ . We assume the prior to be a multivariate normal distribution, $\pi(\sigma) \sim \mathcal{N}(\mu_0 = 56; A_0 = 9)$. Further, we assume that an experiment yields the estimate $\sigma_{\text{exp}} = 57$ and variance $B = 3$. Because prior and experiment refer to the same cross section, the sensitivity matrix can be discarded from the update formula. Then the Bayesian update yields

$$\sigma_1 = \sigma_0 + A_0 S(A_0 + B)^{-1} (\sigma_{\text{exp}} - \sigma_0) = 56 + \frac{9}{9 + 3} (57 - 56) = 56.75, \quad (5.41)$$

$$A_1 = A_0 - A_0 (A_0 + B)^{-1} A_0 = 9 - \frac{81}{9 + 3} = 2.25. \quad (5.42)$$

Therefore, the posterior is the multivariate normal distribution $\mathcal{N}(\sigma_1 = 56.75, A_1 = 2.25)$. Under the assumption of a (multivariate) normal prior and (multivariate) normal likelihood, the computed parameters of the posterior distribution represent the exact solution. The distributions are shown in the left diagram of Figure 5.4.

In both the LSBET and a Monte Carlo procedure, such as the UMC-G or UMC-B method, samples $\vec{\tau}_i$ are drawn from the prior distribution. A refined estimate of the cross section vector is then given by a weighted sum of the sample vectors. We are interested in the efficiency of the LSBET and Monte Carlo methods. Drawing a certain number N of samples, how large is the expected difference to the exact solution? Due to the sampling from the prior, the difference fluctuates randomly. The standard error represents the expected difference between the estimate of the mean vector and the exact solution; consequently 68% of the drawn cases lie within this interval. We computed the standard error of LSBET and UMC-G/B for the schematic scenario as a function of the number of sample vectors N . The result is shown in the right diagram of Figure 5.4. The standard errors of LSBET and

UMC-G/B are

$$\epsilon(N; \text{LSBET}) = \frac{0.767}{\sqrt{N}} \quad \text{and} \quad \epsilon(N; \text{UMC}) = \frac{1.44}{\sqrt{N}}. \quad (5.43)$$

Roughly 3.5 times more samples are required for the UMC-G/B method to achieve the same accuracy as the LSBET. If the experimental data is very unlikely according to the prior distribution, the spread can be even larger. In general, the LSBET is more efficient if the prior distribution is a multivariate normal distribution.

In the second schematic scenario, we investigate the behavior if the prior distribution is different from a multivariate normal distribution. We consider the update of two cross section $\vec{\sigma} = (\sigma_1, \sigma_2)^T$. The prior should be constructed based on the non-linear model

$$\mathcal{M}(\alpha) = \begin{pmatrix} 50 + 10 \sin(\alpha) \\ 50 + 10 \cos(\alpha) \end{pmatrix}. \quad (5.44)$$

We assume a uniform distribution in the interval $0 \leq \alpha \leq \pi/2$ as the prior distribution of the model parameter. Because the LSBET is equivalent to the surrogate approach discussed in section 3.3, drawing a sufficiently large number of samples leads to the multivariate normal prior specified by

$$\vec{\sigma}_0 = \begin{pmatrix} 56.3524 \\ 56.3524 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_0 = \begin{pmatrix} 9.7444 & -8.9323 \\ -8.9323 & 9.7444 \end{pmatrix} \quad (5.45)$$

Let us consider a measurement of these two cross sections to be characterized by

$$\vec{\sigma}_{\text{exp}} = \begin{pmatrix} 57 \\ 54 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 3 & 0 \\ 0 & 0.05 \end{pmatrix} \quad (5.46)$$

The bivariate prior distribution used in LSBET and the experiment distribution are illustrated in the left diagram of Figure 5.5. Performing the Bayesian update with the LSBET leads to the posterior distribution with

$$\vec{\sigma}_{\text{LSBET}} = \begin{pmatrix} 57.9772 \\ 54.0269 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{LSBET}} = \begin{pmatrix} 1.0428 & -0.0297 \\ -0.0297 & 0.0493 \end{pmatrix} \quad (5.47)$$

In contrast, performing the Bayesian update with the UMC-B method leads to the posterior distribution with

$$\vec{\sigma}_{\text{UMC-B}} = \begin{pmatrix} 59.1540 \\ 54.0181 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{UMC-B}} = \begin{pmatrix} 0.0097 & -0.0221 \\ -0.0221 & 0.0502 \end{pmatrix} \quad (5.48)$$

We note that in the case of a non-linear model and a sufficient large sample size, the results of the UMC-G method and the LSBET are virtually identical. This is due to the fact that both methods approximate the non-linear model by the

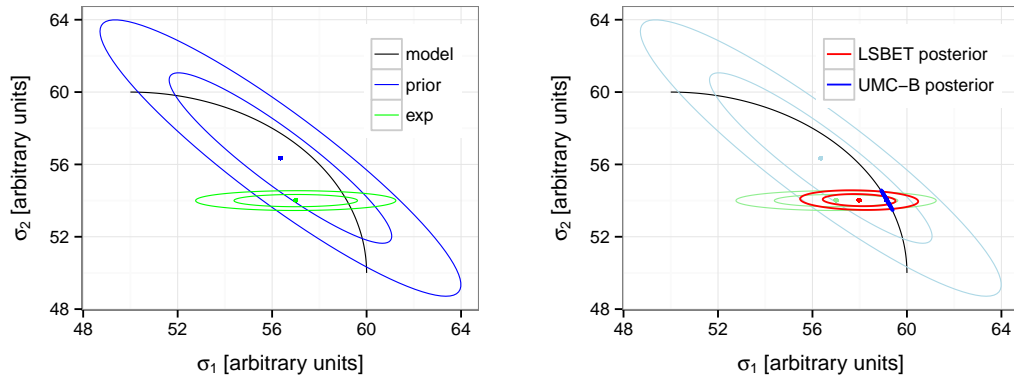


Figure 5.5: Distributions of the second toy scenario are depicted in the left diagram. The evaluated cross sections of LSBET and UMC-B and associated 68%- and 95% confidence ellipses are shown in the right diagram.

multivariate normal distribution given in Equation 5.45. In the UMC-B method the exact model is used and thus the posterior preserves all model systematics. The evaluated correlation between the two cross sections is $\rho_{\text{LSBET}} = \text{Cor}[\sigma_1, \sigma_2] \approx -0.1312$ if using the LSBET, and $\rho_{\text{UMC-B}} \approx -0.9989$ if using the UMC-B method. The high evaluated correlation of the UMC-B method indicates that both cross sections are virtually linear dependent.

The evaluated cross sections and associated uncertainty ellipses of both methods are depicted in the right diagram of Figure 5.5.

If we believe the model to accurately describe reality, the result of the LSBET method does not yield the proper solution. The UMC-B method, on the other hand, comes arbitrary close to the proper estimate when the number of samples gets larger and larger. The magnitude of the bias introduced by approximating the model in the LSBET depends on the non-linearity of the model. However, we have also to note that the proper solution on the model manifold is not excluded by the evaluated covariance matrix. Updating with another experiment which is concentrated on a point on the model curve would remove the bias in the LSBET.

We conclude by summarizing the two important insights of this section. If the prior uncertainties of the model parameters are small enough to constrain the model to a linear domain, and furthermore using a multivariate normal distribution for the model parameters, the LSBET is more efficient. A smaller number of samples is needed to get accurate results. If the uncertainties do not restrict the model to a linear domain, evaluated quantities of the LSBET are biased compared to the exact solution. In general, evaluated covariance matrices of LSBET are less stiff than of UMC-B. Evaluated quantities do not have to lie on the manifold of model possibilities.

Part II

Consistent Model Defects

Consequences of model defects

If a model prediction does not agree with experimental data, then either the experimental data are inconsistent or the model does not account for all relevant physical aspects. Because nuclear models are an essential element in an evaluation procedure, results are significantly affected by the quality of the employed nuclear models.

In real evaluation scenarios, experimental data are often scarce and thus evaluators have to take into account model predictions. In addition, the lack of data makes it difficult to assess the quality of the nuclear models. Indications for the reliability of evaluated data are usually obtained by a comparison of simulations based on these data with measurements from integral benchmark experiments (e.g. Markovskij et al. (2003)).

In this and the next chapter we follow another approach to study the impact of the quality of the nuclear model on the results produced by different evaluation methods. We investigate the performance of the methods in two evaluation scenarios with a large amount of experimental data available.

In the first scenario we assume a simple linear model which has the advantage that the simplification of the nuclear model (i.e. linearization of the model or the surrogate approach) does not affect the result. Hence evaluation methods based on the simplification of the model, such as FBET, and Monte Carlo methods, such as UMC-B, collapse to one common approach. Differences in the results are then only due to different choices of the likelihood.

In the second scenario we use the various methods to perform an evaluation of the neutron-induced total cross section of ^{181}Ta . In contrast to the first scenario, the employed optical model establishes a non-linear relationship between the parameters of the optical potential and the values of cross sections at different

incident energies. Therefore, results can be expected to be dependent on whether the nuclear model is replaced by an approximation or accounted for exactly.

In this chapter we analyze the evaluation methods EMPIRE-Kalman, EMPIRE-MC, FBET, UMC-G, UMC-B, BMC and BFMC in the scenario with the linear model. Furthermore, we introduce the details of the evaluation scenario dealing with the neutron-induced total cross section of ^{181}Ta and apply the EMPIRE-Kalman and FBET/EMPIRE-MC method, which use an approximation of the nuclear model. Monte Carlo methods which use the exact nuclear model will be discussed separately in chapter 7.

6.1 Deficient linear model

In order to study the consequences of model defects, we consider a very simple hypothetical model \mathcal{M} to predict the cross sections of a single reaction channel. The model assumes that the cross sections σ at different incident energies E are located on a straight line. The functional form of the model is

$$\mathcal{M}(E) = kE + d. \quad (6.1)$$

The slope k and the intercept d of the straight line are the model parameters. Of course, this simple model is not suitable to predict any real cross section curve. Nevertheless, this model like any actual model represents a function that predicts cross sections once the model parameters are specified. Furthermore, because the model is linear, Monte Carlo methods, such as UMC-G, UMC-B, and linearized methods, such as FBET and EMPIRE-MC, yield the same solution. The following results were obtained by the application of the update formulas of section 3.2.

In actual nuclear data evaluation, we never know the true values of the observables, e.g. cross sections, but only results of measurements. Contrary in the current schematic scenario, we can assume that a certain cross section curve represents the truth. In the present toy example, this true cross section curve should be given by

$$\sigma_{\text{true}}(E) = (E - 15)^2 + 10. \quad (6.2)$$

The true cross sections reside on a parabola whereas the model assumes that cross sections are located on a straight line. Hence, the model is deficient and cannot provide a proper description of the true values.

Measurements are always afflicted by measurement error. Therefore, measured values are only approximations to the true values. Let us assume that independent measurements at 20 energies were made, each with a standard error (or uncertainty) of $\delta = 2$ mBarn. Thus the associated covariance matrix \mathbf{B} is diagonal, and each diagonal element is given by $\delta^2 = 4$. In order to simulate the

Einc	xs	Einc	xs	Einc	xs	Einc	xs
10.0000	35.7150	12.6316	11.8115	15.2632	9.7647	17.8947	15.1050
10.5263	27.9861	13.1579	16.2711	15.7895	8.9097	18.4211	20.4166
11.0526	27.0942	13.6842	10.9031	16.3158	11.3483	18.9474	23.6427
11.5789	21.8605	14.2105	12.6094	16.8421	13.8749	19.4737	28.2947
12.1053	15.8224	14.7368	8.3059	17.3684	12.7900	20.0000	36.3257

Table 6.1: Values of the hypothetical measurements rounded to four decimal places. Each value is associated with the same standard error (=uncertainty) $\delta = 2$. The experimental data points are assumed to be uncorrelated.

hypothetical measurements, we added Gaussian noise $\mathcal{N}(0, \delta^2)$ to the true cross sections given by Equation 6.2. In this way obtained ‘experimental values’ are listed in Table 6.1. In this simple example, the sensitivity matrix \mathbf{S} of dimension 20×2 consists of the elements $S_{i1} = E_i$ and $S_{i2} = 1$, where E_i is the incident energy of the i^{th} experimental data point.

In order to apply Bayesian statistics, a prior has to be specified for the model parameters. We assume a multivariate normal distribution for the model parameters $\vec{p} = (k, d)$ with

$$\vec{p}_0 = \begin{pmatrix} k_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 15 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_0 = \begin{pmatrix} 2 & 0 \\ 0 & 400 \end{pmatrix}. \quad (6.3)$$

The true cross section curve, the generated experimental values, and the model prior are displayed in Figure 6.1.

Performing the Bayesian update yields the multivariate normal posterior distribution with

$$\vec{p}_1 = \begin{pmatrix} k_1 \\ d_1 \end{pmatrix} = \begin{pmatrix} 0.0328 \\ 19.3382 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = \begin{pmatrix} 0.0212 & -0.3182 \\ -0.3182 & 4.9707 \end{pmatrix}. \quad (6.4)$$

The corresponding evaluated cross section curve is displayed in Figure 6.2. We remark that assuming perfectly correlated experimental errors does not make any difference for the further discussion (evaluated uncertainties of cross section estimates would be around 1.4 mBarn). The basic assumption that the model perfectly describes reality is reflected in the evaluated cross section curve. The model is a straight line and consequently also the evaluated cross section curve is a straight line. No amount of experimental data suggesting a different shape can change this behavior. Furthermore, the evidence given by the experimental data leads to rather small uncertainties of the evaluated cross section. From the point of Bayesian statistics, the obtained mean values and the associated uncertainties are completely correct.

However, they do not reflect the obvious energy dependence of the available set of reliable experimental data. Especially, the given error band appears too

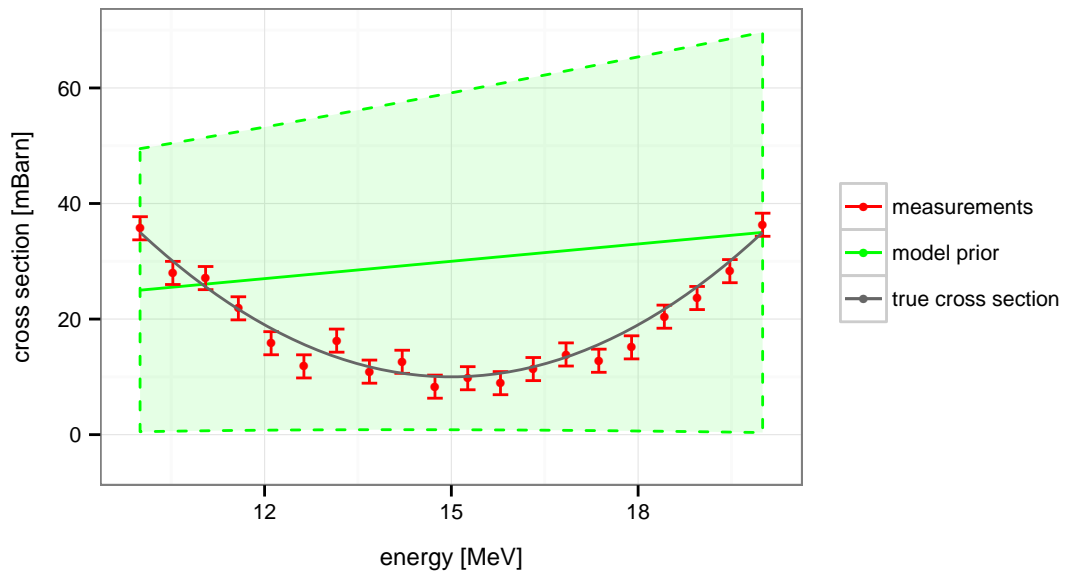


Figure 6.1: True cross section curve, the generated values of the measurements and the model prior in the hypothetical scenario. The green line denotes the prior best estimate and the pale green area indicates the associated 1-sigma (68%) uncertainty band. The vertical extent of the 95% uncertainty would be roughly two times larger.

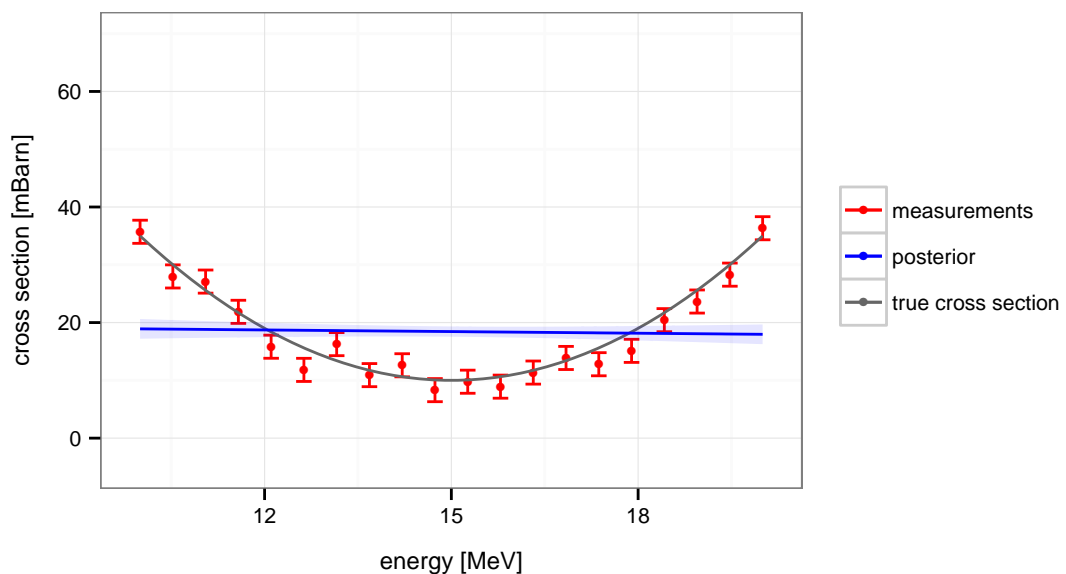


Figure 6.2: True cross section curve, the generated values of the measurements and the evaluated cross section in the hypothetical scenario. The blue line denotes the posterior estimate and the pale area indicates the associated 2-sigma (95%) uncertainty band.

narrow and does not correspond to well founded expectations. Consequently, such a result would not be acceptable for nuclear data evaluation. The origin of this unsatisfactory posterior is the incorrect assumption that the model provides a perfect description of the ‘true’ value of the observables.

This simple toy example reveals a fundamental shortcoming of most contemporary evaluation methods, i.e. the assumption that the involved nuclear models yield a perfect description of the observables. With regard to the complexity of the nuclear many-body problem the validity of this assumption is highly questionable. The given example clearly shows the significant impact of ignored model defects on the final evaluation. The ignorance of model defects usually results in biased estimates of the mean values and unrealistic (frequently underestimated) error bands. Hence, the reliability of the evaluation must be questioned.

As discussed in chapter 4, the BFMC and BMC method are based on unconventional choices for the likelihood. These unconventional choices lead to evaluated uncertainties differing from those produced by the methods mentioned in the last paragraph. In the following, we investigate these differences also on the basis of the toy model.

The likelihood of the BMC method was given in Equation 4.17. The value of the likelihood depends explicitly on the number of experimental datasets. This property may be criticized, because the result of an evaluation should be independent of how the data is split. Only the actual information in the data should determine the result. If we assume that each experimental data point represents an individual dataset, the result of the BMC method coincides with the result Equation 6.4 obtained by the FBET and the other approaches. The evaluated cross section curve is illustrated in Figure 6.2. An alternative visualization of the update procedure in terms of the model parameters k and d is presented in Figure 6.4.

We have the other extreme case of the BMC method if all experimental data points belong to the same dataset. Rewriting the exponent in Equation 4.17 in matrix notation, we get for the toy scenario (a single reaction channel and only one dataset)

$$\ell_{\text{BMC}}(\vec{p} | \vec{\sigma}_{\text{exp}}, \mathcal{M}) \propto \exp \left(-\frac{1}{2} (\mathcal{M}(\vec{p}) - \vec{\sigma}_{\text{exp}})^T (N\mathbf{B})^{-1} (\mathcal{M}(\vec{p}) - \vec{\sigma}_{\text{exp}}) \right), \quad (6.5)$$

where N is the number of experimental data points contained in the vector $\vec{\sigma}_{\text{exp}}$. The covariance matrix in this likelihood $\tilde{\mathbf{B}} = N\mathbf{B}$ is scaled by a factor N compared to the conventional choice. Performing the update with the BMC method yields

$$\vec{p}_{\text{BMC}} = \begin{pmatrix} k_1 \\ d_1 \end{pmatrix} = \begin{pmatrix} 0.1115 \\ 16.7523 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{BMC}} = \begin{pmatrix} 0.2977 & -4.4207 \\ -4.4207 & 69.6138 \end{pmatrix}. \quad (6.6)$$

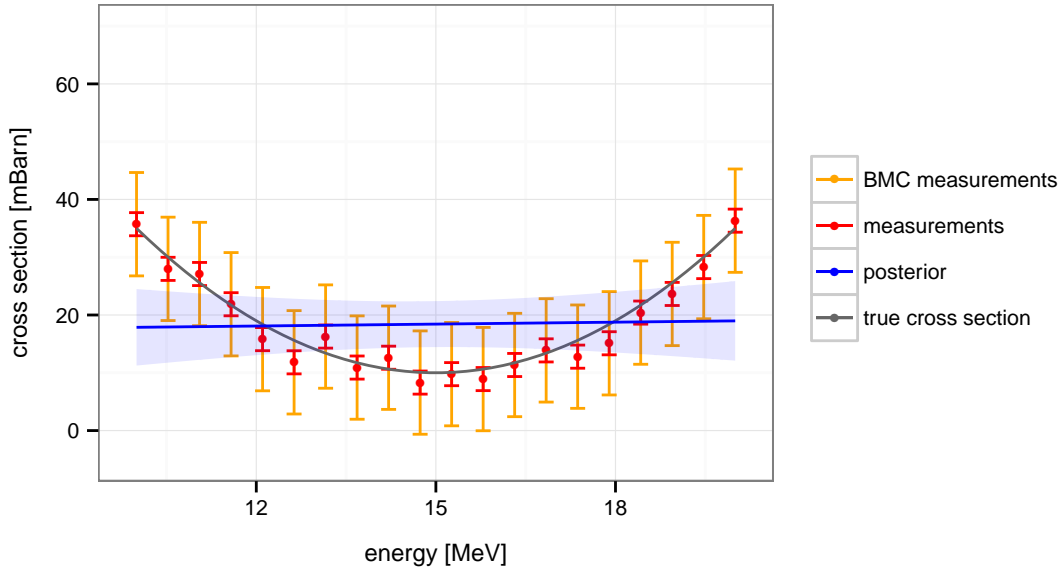


Figure 6.3: The evaluated cross section curve and the associated 95% confidence band resulting from updating with the BMC method. The BMC method inflates the uncertainties of the experimental data points. This measure leads to larger evaluated uncertainties, compare to Figure 6.2.

Evaluated uncertainties are about a factor 3.7 larger than those associated with Equation 6.4. The evaluated cross sections are shown in Figure 6.3. The update in terms of model parameters is visualized in Figure 6.5. Even though, the uncertainties are larger, the evaluation is still not consistent to experimental data.

Hence, a modified likelihood does not represent a viable possibility to account for model defects. If the model and the experimental data are reliable, evaluated uncertainties are larger than required. On the other hand, in case of severe model deficiency, the evaluated mean values and uncertainties may still not reflect the information contained in the experimental data.

There is yet another reason why the modification of the likelihood cannot be considered as an appropriate measure to treat model defects. Not only the evaluated cross section curve is restricted to the possibilities of the model but also evaluated covariance matrices at the level of cross sections inherently include the systematics of the model. In the case of model deficiency, however, these dependencies are not given and therefore inappropriate. Figure 6.6 shows cross section curves drawn from the BMC posterior. Clearly, admissible deviations from the posterior best estimate are confined to straight lines. Every evaluation method not taking into account model defects possesses the same feature. Only methods, such as FBET and EMPIRE-MC, that replace the original model by a surrogate model allow freedom beyond the possibilities of the model. The degree of freedom depends on the non-linearity of the nuclear model. The implied assumption that

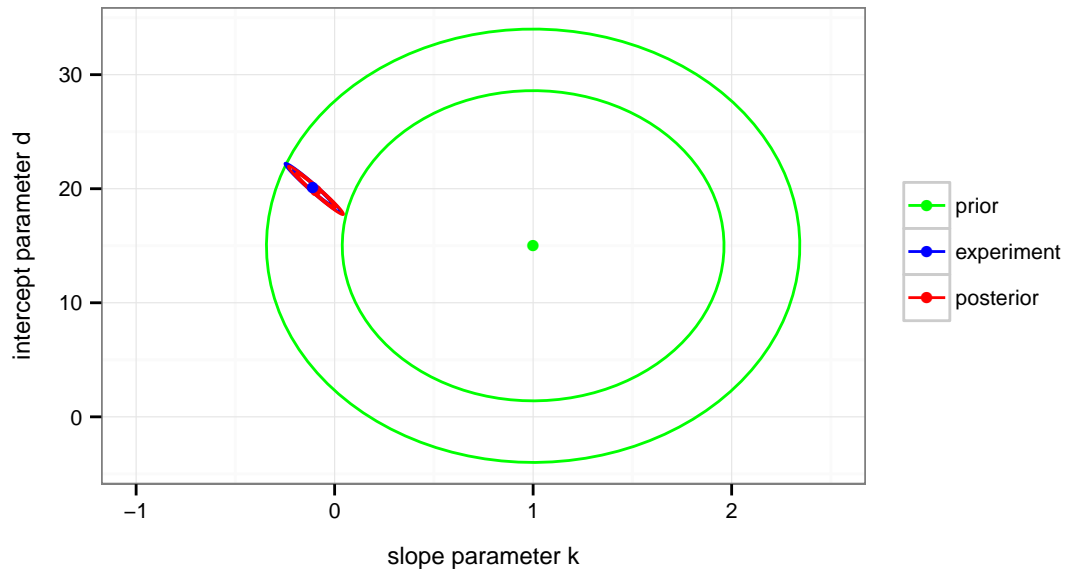


Figure 6.4: The update procedure at the level of model parameters. Inner ellipses denote the 68% confidence area and outer ellipses the 95% confidence area of the probability distributions introduced in the schematic scenario. The displayed posterior distribution results from the application of methods such as FBET, EMPIRE-MC, EMPIRE-Kalman, UMC-B, and UMC-G. The BMC method also produces this result if each experimental data point represents an individual dataset. Because of the rather accurate estimate associated with the experiment compared to the prior, the posterior roughly coincides with the probability distribution of the experiment.

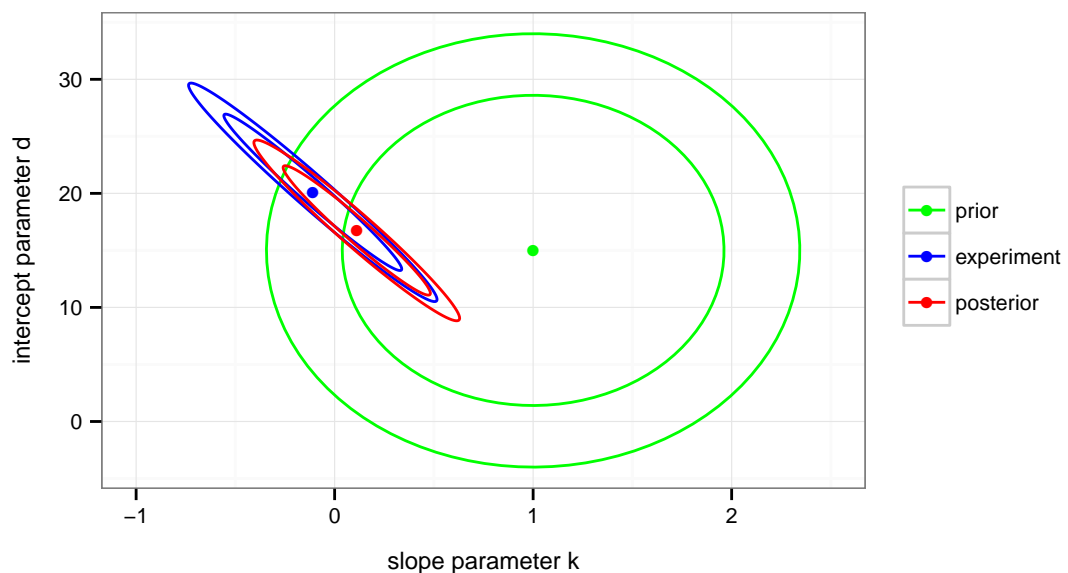


Figure 6.5: The update procedure at the level of model parameters. Inner ellipses denote the 68% confidence area and outer ellipses the 95% confidence area of the probability distributions introduced in the schematic scenario. The depicted posterior distribution results from the application of the BMC method if all experimental points are considered as belonging to one dataset. The likelihood (experiment) is inflated compared to Figure 6.4.

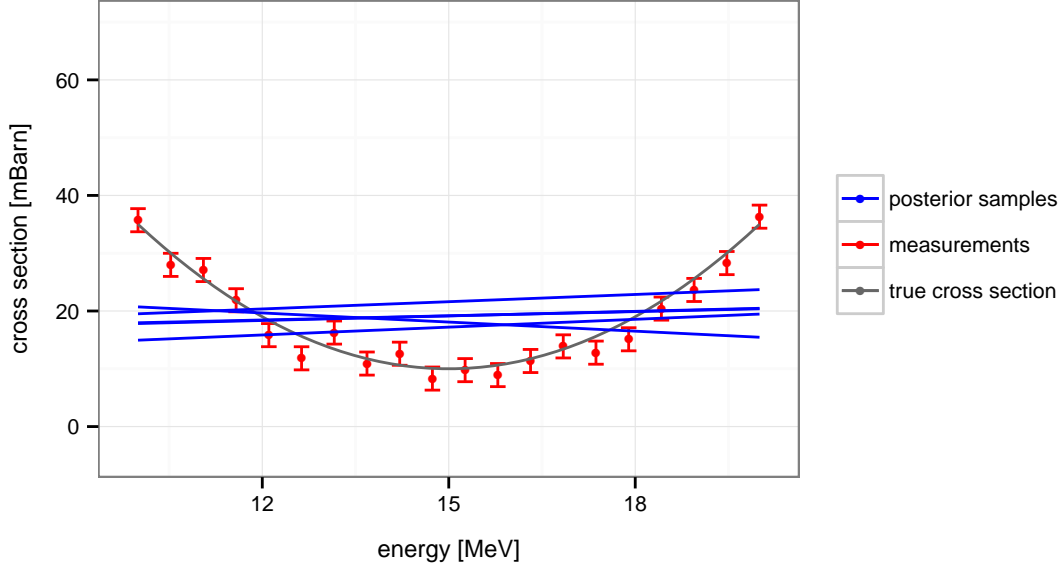


Figure 6.6: Cross section curves drawn from the posterior distribution obtained by the BMC method.

there is a relationship between the non-linearity of the model and the extent of the model defect is difficult to justify. Furthermore, in view of the large amount of data produced by a nuclear model code and the non-negligible runtime of these codes, the degree of non-linearity of the model is difficult to assess. Thus, assumptions for the model defect remain unknown.

The issues addressed so far also apply for the BFMC method. For the sake of completeness, we have also applied the BFMC method to the toy scenario. The likelihood of the BFMC method was given in Equation 4.8. We repeat it here specialized to the case of the linear model $\mathcal{M}(\vec{p}) = \mathbf{S}\vec{p}$,

$$\ell_{\text{BFMC}}(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) = C \left[\exp \left\{ - \left(\frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{p})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{p}) \right)^2 \right\} \right]^{\left(\frac{2}{\chi_{\min}^2} \right)^2} \quad (6.7)$$

In contrast to the likelihood in Equation 6.5 of the BMC method, the exponent of the BFMC likelihood is squared. This likelihood significantly differs from a multivariate normal distribution. For the linear model $\mathcal{M}(\vec{p}) = \mathbf{S}\vec{p}$ of the toy scenario, the model parameters k and d occur up to the fourth power in the exponent of the likelihood. Thus, one cannot expect to easily (if at all) obtain analytic update formulas. The derivation of such formulas would require the solution of a system of equations where the variables occur up to the third power. Therefore, even for the simple linear model of the toy scenario, we have to resort to numeric methods.

We start with the determination of the χ_{\min}^2 -value. In the BFMC method, the χ_{\min}^2 -value is given by the lowest χ_i^2 -value of the sampled model prediction vectors

$\vec{\sigma}_i$. The formula to compute the χ_i^2 -values is

$$\chi_i^2 = (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i)^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_i) , \quad (6.8)$$

where $\vec{\sigma}_{\text{exp}}$ contains the measured cross sections and \mathbf{B} is the associated covariance matrix. In the toy scenario, we can compute the theoretically lowest possible χ_i^2 -value that could appear in a sample. The χ_{min}^2 -value is associated with the parameter vector \vec{p}_ℓ for which the exponential function in Equation 6.7 attains its largest value. Because squaring a non-negative function preserves the location of the maximum, the same parameter vector \vec{p}_ℓ is associated to the maximum of the multivariate normal distribution

$$\tilde{C} \exp \left\{ -\frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{p})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{p}) \right\} . \quad (6.9)$$

We derived the parameter vector \vec{p}_ℓ that leads to a maximum likelihood in section 3.2. Adapting the solution in Equation 3.21 to the current situation,

$$\vec{p}_\ell = (\mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1} \mathbf{S}^T \mathbf{B}^{-1} \vec{\sigma}_{\text{exp}} , \quad (6.10)$$

we obtain $\vec{p}_\ell = (-0.1100, 20.0919)^T$ and hence $\chi_{\text{min}}^2 = 362.792$. The χ_{min}^2 -value quantifies the adequacy of the model taking into account the deviation from experimental data. For a good model the quantity χ_{min}^2/N with N being the number of experimental data points should be about one. As expected, the value $362.792/20 \approx 18$ indicates the deficiency of the linear model in the toy scenario. The exponent $(2/\chi_{\text{min}}^2)^2$ in Equation 6.7 flattens the likelihood according to the magnitude of the model deficiency. In this sense, the BFMC method is the only Monte Carlo procedure that accounts for model defects. Nevertheless, evaluated estimates and covariance matrices are still determined by the restriction of the (maybe very deficient) model, and therefore biased mean values and questionable uncertainties might be obtained.

With the knowledge of the χ_{min}^2 -value, we can normalize the likelihood in Equation 6.7 to unity,

$$\int_{\mathbb{R}^N} \ell_{\text{BFMC}}(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) d\vec{p} = 1 . \quad (6.11)$$

We evaluated this integral by Monte Carlo integration and obtained the normalization constant $C = 0.0955 \pm 0.0004$. Knowledge of this normalization constant is not required in the BFMC method. However, here we need C for a comparison of evaluated confidence areas to those of other methods.

Using again Monte Carlo integration, we get the mean parameter vector \vec{p}_ℓ and the associated covariance matrix \mathbf{B}_ℓ ,

$$\vec{p}_\ell = \begin{pmatrix} k_\ell \\ d_\ell \end{pmatrix} = \begin{pmatrix} -0.1216 \\ 20.3300 \end{pmatrix} \quad \text{and} \quad \mathbf{B}_\ell = \begin{pmatrix} 1.2535 & -18.8492 \\ -18.8492 & 295.0867 \end{pmatrix} . \quad (6.12)$$

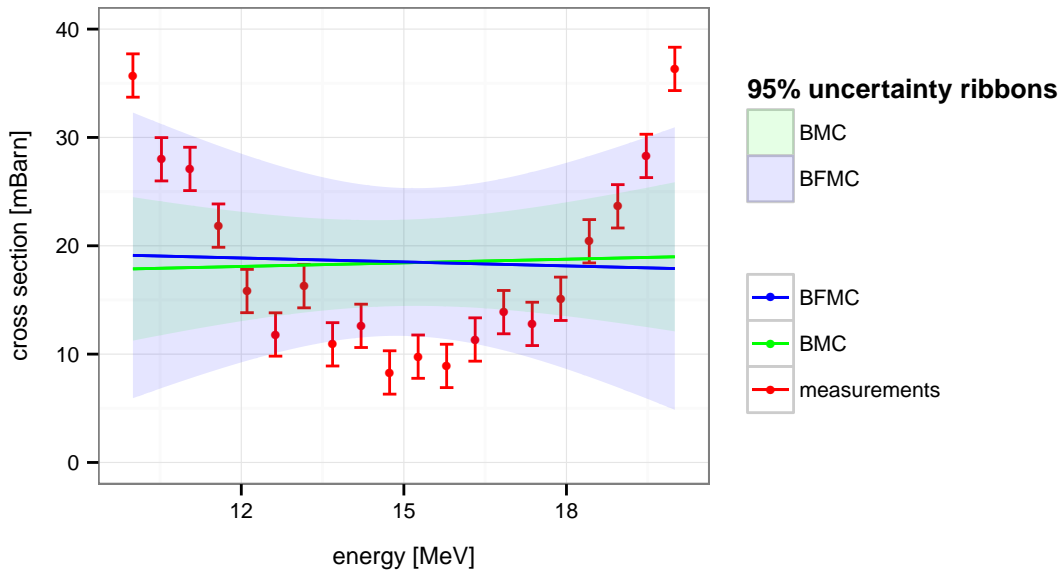


Figure 6.7: Comparison of the evaluated cross section curves and associated 95% uncertainty bands obtained by the BMC and BFMC method.

The mean vector \vec{p}_ℓ is slightly different from the values given under Equation 6.10 because of the statistical error of the Monte Carlo integration. The statistical fluctuations are by a factor hundred smaller than the evaluated uncertainties $\Delta k = \sqrt{1.2634} \approx 1$ and $\Delta d = \sqrt{295.0867} \approx 17$, and hence of no significance. If we assume that the prior is sufficiently uninformative, the posterior is essentially determined by the likelihood. Consequently, the evaluated uncertainties of k and d are about two times larger than in the BMC method. In general, the uncertainties generated by the BFMC method strongly depend on the actual χ^2_{\min} -value. Figure 6.7 compares the evaluated cross section curves resulting from the BMC and the BFMC method. Due to the rescaling of the likelihood, the BFMC method encloses most experimental data points in the 95% uncertainty band. The possibilities are still restricted to a straight line, and realizations of cross section curves drawn from the posterior would be exclusively given by straight lines, similar to those in Figure 6.6 but more scattered.

Furthermore, we want to discuss the functional form of the BFMC likelihood, Equation 6.7. This likelihood is not a multivariate normal distribution; nevertheless the isolines of the probability density are ellipsoids of the same shape as those of the multivariate normal distribution determined by the same covariance matrix \mathbf{B} . Equal shape means that the ratio of the major axis and the minor axis of the ellipsoids is the same for both the BFMC likelihood and the corresponding multivariate normal likelihood.

In the forward step of the BFMC method, the probability distribution of the likelihood is approximated by a multivariate normal distribution whose first and

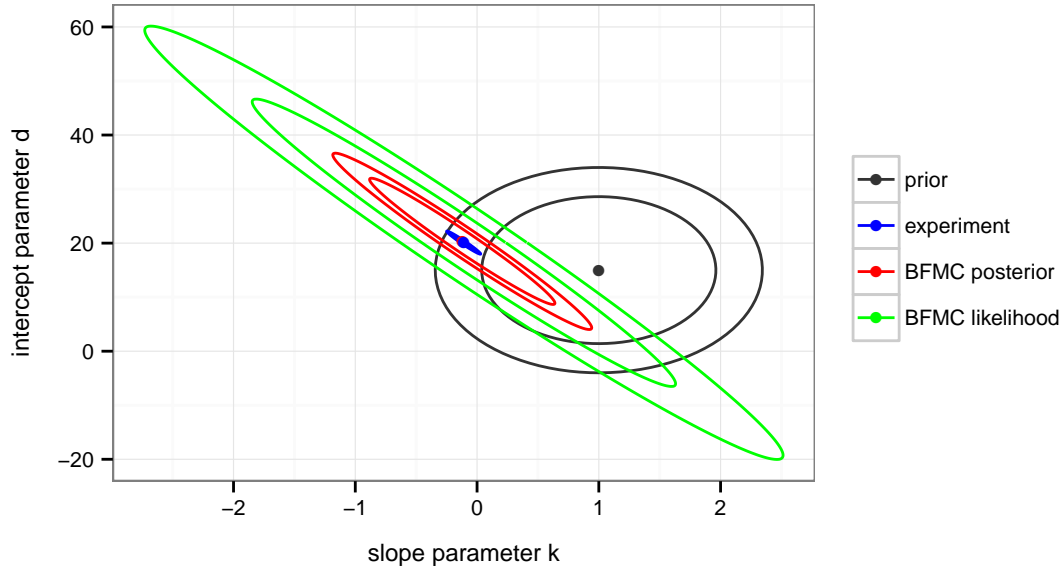


Figure 6.8: Visualization of the BFMC likelihood and the BFMC posterior distribution. The ellipsoids indicate 68% and 95% uncertainty areas. The BFMC posterior is obtained by fitting a multivariate normal distribution to samples weighted according to the BFMC likelihood. The prior distribution employed by other methods, such as FBET, UMC-G, BMC, etc. is shown for comparison.

second moments match those of the exact likelihood. The first moments represent the mean vector and the second moments the covariance matrix. The weights obtained for the drawn parameter vectors in the backward step are used in the forward step to construct a multivariate normal distribution for the parameters, hence we refer to the latter distribution as the BFMC posterior. The probability distributions are visualized in Figure 6.8. The fact that both isolines of the BFMC likelihood enclose the respective isolines of the BFMC posterior might irritate. This phenomenon can be attributed to the specific functional form of the BFMC likelihood. Figure 6.9 shows the decline of probability density as function of the distance from the center point along the major axis of the uncertainty ellipsoid for both the BFMC likelihood and the BFMC posterior visualized in Figure 6.8. The BFMC likelihood is flatter around the center than the BFMC posterior with the same mean vector and covariance matrix. Therefore, confidence intervals associated with the same level of trust are wider.

In this section we showed at the example of a simple linear model that a deficient model may lead to biased evaluated cross section curves whose meaning are questionable. The results are summarized in Table 6.2. In the example, the slope k and the intercept d exhibit a strong negative correlation for all considered methods. Propagating the uncertainties Δk and Δd from the model parameters to the cross sections, we obtain for the correlation between two cross sections σ_i

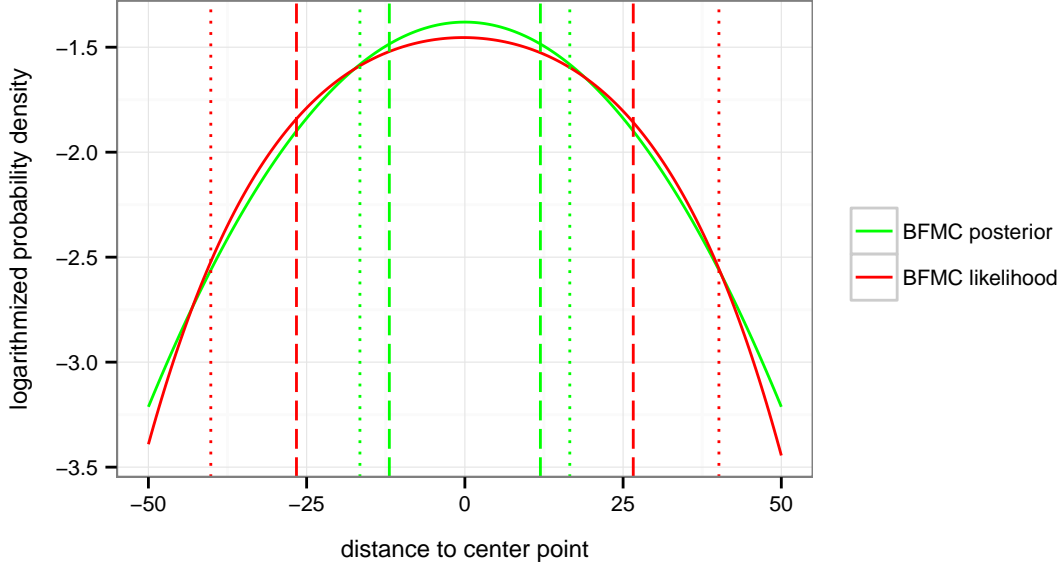


Figure 6.9: Logarithmized probability density of the BFMC likelihood and the BFMC posterior along the major axis of the green ellipsoids shown in Figure 6.8. The dashed lines indicate the intersection points with the 68% uncertainty isolines and the dotted lines the intersection points with the 95% uncertainty isolines.

at energy E_i and σ_j at energy E_j

$$\text{Cor}[\sigma_i, \sigma_j] = \frac{E_i E_j \Delta k^2 + \Delta d^2 + \rho \Delta k \Delta d (E_i + E_j)}{E_i E_j \Delta k^2 + \Delta d^2 + \Delta k \Delta d (E_i + E_j)}, \quad (6.13)$$

where ρ is the evaluated correlation between slope and intercept, i.e. $\rho = \text{Cor}[k, d]$. All involved quantities are positive except the correlation coefficient ρ . Perfect positive correlation between cross sections is achieved for $\rho = 1$. The evaluated negative correlation $\rho \approx -1$ of all methods might be interpreted as the attempt to reduce the correlations between cross sections at different incident to a minimum, in order to agree with the assumption of uncorrelated experimental data points. The latter assumption is reflected by the diagonal experimental covariance matrix \mathbf{B} . The correlations of the evaluated cross sections are displayed in Figure 6.10.

Method	k	d	ρ	Δk	Δd	$\Delta \sigma$
FBET	0.0328	19.3382	-0.9805	0.1456	2.2300	0.6021
BMC	0.1115	16.7523	-0.9711	0.5456	8.3435	2.5073
BFMC	-0.1216	20.3300	-0.9801	1.1200	17.1781	4.6093

Table 6.2: Comparison of the results obtained by the methods in the toy scenario. Given are standard errors (=uncertainties) Δk , Δd and the correlation ρ between the slope k and the intercept d . The quantity $\Delta \sigma$ denotes the average standard error of the cross sections. The result of FBET is identical to that of EMPIRE-Kalman, EMPIRE-MC, UMC-B, and UMC-G for large sample sizes.

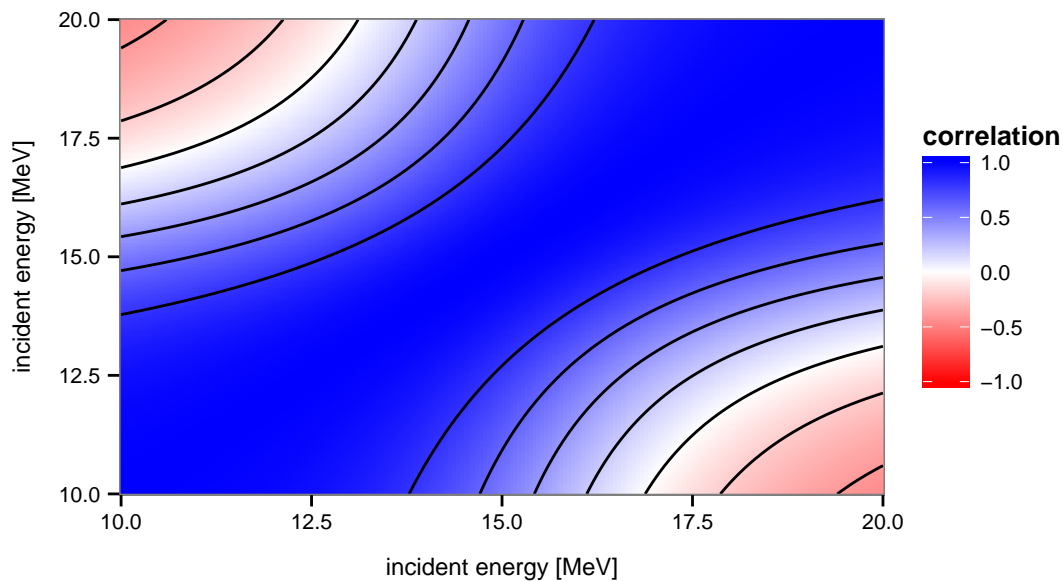


Figure 6.10: Evaluated correlations between the cross sections resulting from the application of the FBET in the toy scenario.

As pointed out above, among the presented methods only the BFMC method takes into account model defects. However, the modification of the likelihood function with a quantity characterizing the quality of the model does not provide a satisfactory treatment of model defects. Even though evaluated uncertainty bands are widened, the posterior still allows variations only within the possibilities of the model. Not a single cross section curve sampled from the posterior would reflect the visible trend in the experimental data. Consequently, a proper treatment of model defects in evaluation methods is required. Such a treatment should allow evaluated cross section curves to be outside the possibilities of the model whenever experimental data indicates deficiencies of the model. Returning to the toy model, a reasonable evaluation accounting for model deficiencies might lead to a result as illustrated in Figure 6.11. The points made about model defects in the toy scenario are of general validity.

The considerations within the simple toy model clearly indicate that the proper treatment of the model deficiencies requires an extension of the Bayesian update procedure. In chapter 9 we will show that an extension of the prior with Gaussian processes provides a proper framework. This extension for model defects can be applied to almost all available Bayesian evaluation techniques because UMC-B, UMC-G, FBET, EMPIRE-Kalman, and EMPIRE-MC differ only in the applied method for the evaluation of the Bayesian update formula, either by a Monte Carlo procedure or the approximation of the nuclear model.

The next section demonstrates the consequences of model defects for a serious case in nuclear data evaluation.

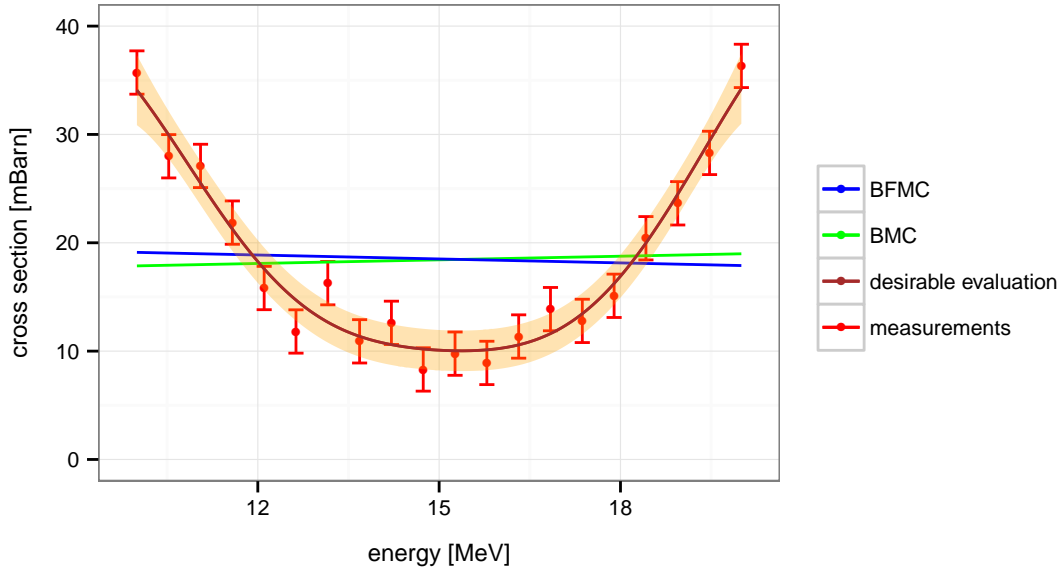


Figure 6.11: Illustration of an evaluated cross section curve that can be considered reasonable. It agrees with experimental data and the 95% uncertainty band has a reasonable width.

6.2 Deficient non-linear model

In this section we consider a realistic application, i.e. the evaluation of the neutron-induced total cross section of ^{181}Ta . The study of this example is of particular interest with regard to nuclear data evaluation. It is well known that the available optical model potential provides a good description of this cross section. The agreement between experimental data and model predictions may give rise to the assumption that the issue of model defects is of no relevance here. A careful analysis of the various evaluation methods outlined in this section refutes this assumption. In the following we give the details of the evaluation scenario.

The global parametrization of the neutron-nucleus optical potential of Koning and Delaroche (2003) depends on more than twenty parameters (see section A.1). Optimizing only the three most significant parameters yields a very accurate fit to the experimental data. The possibility to restrict the variation to three model parameters is of great value for studying the methods. The low dimensional space of model parameters can be exhaustively explored, which in consequence allows to perform any evaluation method—Monte Carlo or not—in short time. The generation of Monte Carlo histories consisting of hundred thousand model calculations takes less than a minute. Finally, a wealth of high-quality experimental data is available, which makes possible to benchmark the evaluation methods.

E [MeV]	σ [mbarn]	ϵ_{EXFOR}	ϵ_{norm}	ϵ_{stat}	ϵ_{tot}
5.2925	5800.8700	16.1500	29.0043	48.4500	56.4682
7.8956	5085.2900	12.9200	25.4264	38.7600	46.3556
11.7788	5227.6700	15.8500	26.1384	47.5500	54.2606
17.7486	5417.0700	20.7300	27.0853	62.1900	67.8322
26.4779	4769.3300	16.6900	23.8467	50.0700	55.4587
39.5005	4055.2800	14.0500	20.2764	42.1500	46.7734
59.5202	4327.4100	14.9800	21.6370	44.9400	49.8775
88.7940	4436.7500	16.1800	22.1837	48.5400	53.3690
132.4650	3574.9900	14.4600	17.8749	43.3800	46.9184
199.6020	2724.0600	13.4300	13.6203	40.2900	42.5300

Table 6.3: Excerpt from the Finlay et al. (1993) dataset containing measurements of the neutron-induced total cross section σ of ^{181}Ta at different incident energies. The column ϵ_{EXFOR} lists the statistical errors given in the EXFOR database. The columns ϵ_{norm} , ϵ_{stat} , and ϵ_{tot} list the normalization error (0.5%), statistical error, and total error that we use in the evaluation scenario. The unit of all given error contributions is millibarn.

We used the experimental data of Finlay et al. (1993) either as input or as benchmark of the results. These data are included in the EXFOR database (National Nuclear Data Center, 2015b; Otuka et al., 2014) where a normalization error below 0.5% and a statistical error about 0.4% is given. We assumed the normalization error to be 0.5% and rescaled the statistical error by a factor of three to be on the safe side. These assumptions lead to an overall error about 1.2% for the data points. We restricted the evaluation to incident energies between 5 and 200 MeV. The dataset of Finlay contains 364 points in this range. An excerpt from the dataset is shown in Table 6.3.

To determine the most sensitive parameters, we numerically calculated the quantities $\partial\sigma_i/\partial p_j$ for a dense mesh of incident energies E_i and the 24 neutron optical model parameters p_j . We used a difference quotient to approximate the derivative

$$\frac{\partial\sigma_i}{\partial p_j} \approx \frac{\sigma_i(p_j + h_j) - \sigma_i(p_j)}{h_j}. \quad (6.14)$$

The constant h_j was chosen to be 1% of the default value of the model parameter p_j . Figure 6.12 illustrates the sensitivity of the cross sections at different incident energies with regard to the change of a particular model parameter (variation 10%). Hence, the figure visualizes the quantities $J_{ij} = 0.1p_j(\partial\sigma_i/\partial p_j)$. In order to make all evaluation methods tractable, we restricted possible variations to the model parameters r_v , a_v and v_1 . Sensitivities with respect to these parameters are visualized in the bottom three rows of Figure 6.12. All other model parameters were left at their default values given by the global parametrization of Koning and Delaroche (2003). First we prove whether the variation of the three selected

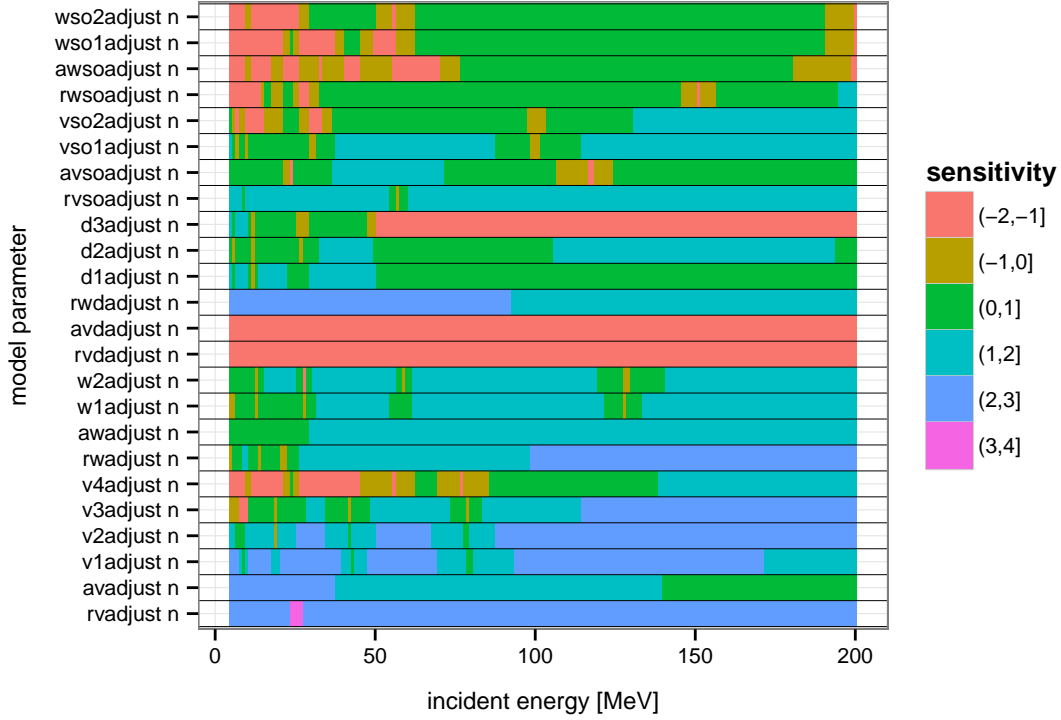


Figure 6.12: Sensitivity of the neutron-induced total cross section of ^{181}Ta at different incident energies to variations of the neutron optical model parameters. The colors of the cells indicate the change of the cross section value when the model parameter is changed by 10% from its default value. The change is measured in powers of 10, e.g. blue color indicates that the change of cross section is between 100 and 1000 mBarn. The parameters are identified via their keywords used in TALYS input files. Details are given in the TALYS manual (Koning, Hilaire, and Goriely, 2013).

parameters is sufficient to obtain a good fit to the data.

The restriction to three model parameters makes it possible to exhaustively explore the parameter space. We defined for the parameter a_v the mesh $A := \{0.85d_1, 0.86d_1, \dots, 1.14d_1, 1.15d_1\}$ where d_1 represents the default value of a_v . The mesh for r_v is given by $R := \{0.85d_2, 0.86d_2, \dots, 1.14d_2, 1.15d_2\}$ with d_2 the default value of r_v . And finally, the mesh for v_1 is $V := \{0.85d_3, 0.86d_3, \dots, 1.14d_3, 1.15d_3\}$ with d_3 the default value of v_1 . The compound mesh for all three parameters is given by the cartesian product $A \times R \times V$. We performed a TALYS calculation for each combination of parameter values occurring in this mesh. These $31 \times 31 \times 31 = 29791$ calculations were carried out on the Vienna Scientific Cluster. The mesh of incident energies E_i used in all calculations comprised 235 points from 0.1 MeV to 210 MeV and was generated with the formula

$$E_i = 0.1 \exp\left(\frac{(i-1)(\ln 210 - \ln 0.1)}{235}\right), i = 1..235. \quad (6.15)$$

In the following we denote $\tilde{a}_v = a_v/d_1$, $\tilde{r}_v = r_v/d_2$, $\tilde{v}_1 = v_1/d_3$ as the model

parameters.

Based on these model predictions on the 4-dimensional mesh $(E, \tilde{a}_v, \tilde{r}_v, \tilde{v}_1)$ the use of linear interpolation with respect to each parameter and the incident energies allows us to compute cross sections at arbitrary incident energies for arbitrary combinations of parameter values $\tilde{a}_v, \tilde{r}_v, \tilde{v}_1$ within the boundaries of the meshes. The results of the model calculations together with the interpolation scheme define a surrogate model $M_{\text{sur}}(\vec{p}, \vec{E})$. Due to the dense mesh of model parameters and incident energies, the predictions of the surrogate model reflect to a great extent the features of the original model. Moreover, calculations of the surrogate model can be performed very fast (predictions for thousands of model parameter sets in a few seconds) whereas calculations of the original model are rather involved.

To verify the adequacy of the model, we determined the most appropriate parameter vector $\vec{p} = (\tilde{a}_v, \tilde{r}_v, \tilde{v}_1)^T$ of the surrogate model by minimization of the generalized $\chi^2(\vec{p})$ -value

$$\chi^2(\vec{p}) = (\vec{\sigma}_{\text{exp}} - \mathcal{M}_{\text{sur}}(\vec{p}))^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}_{\text{sur}}(\vec{p})). \quad (6.16)$$

The vector $\vec{\sigma}_{\text{exp}}$ contains the measurements and \mathbf{B} is the associated experimental covariance matrix. We employed the L-BFGS algorithm (Byrd et al., 1995) to find the parameter vector associated with minimal χ^2 . The L-BFGS algorithm allows to specify ranges in which the parameter values have to reside. This feature of the algorithm is a necessity in our case, because cross sections can only be interpolated within the boundaries of the mesh $A \times R \times V$ of model parameters and the mesh of incident energies.

We performed several optimization runs, using either the experimental covariance matrix \mathbf{B} containing all correlations or a stripped version $\mathbf{B}_{\text{strip}}$ thereof in which only diagonal elements are non-zero. Independent of the parameter set from which the optimization started, we obtained with the full covariance matrix \mathbf{B} the most appropriate parameter set,

$$\begin{pmatrix} \tilde{a}_v \\ \tilde{r}_v \\ \tilde{v}_1 \end{pmatrix} = \begin{pmatrix} 1.0611 \\ 0.9762 \\ 1.0167 \end{pmatrix} \quad \text{and} \quad \chi^2 = 348.6544. \quad (6.17)$$

When we used the stripped covariance matrix $\mathbf{B}_{\text{strip}}$ we got

$$\begin{pmatrix} \tilde{a}_v \\ \tilde{r}_v \\ \tilde{v}_1 \end{pmatrix} = \begin{pmatrix} 1.0618 \\ 0.9772 \\ 1.0161 \end{pmatrix} \quad \text{and} \quad \chi^2 = 279.2549. \quad (6.18)$$

In both cases, the estimates of the optimal parameters are approximately the same. The quantity χ^2/N with $N = 364$ being the number of experimental data

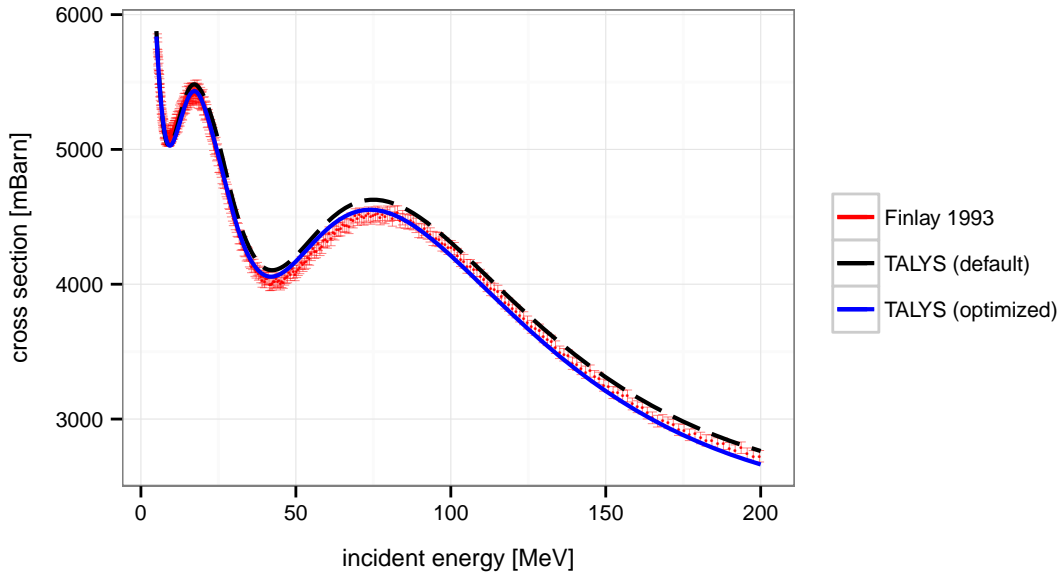


Figure 6.13: Neutron-induced total cross section of ^{181}Ta . The experimental dataset of Finlay is used in the evaluation scenario. Error bars denote the 95% confidence interval. The shown TALYS predictions are based on the default parameter set and an optimized parameter set.

points is in both cases around one. Therefore, the model can be regarded as adequate. For comparison, the default parameter vector $(1, 1, 1)^T$ is associated with $\chi^2 = 696.3525$. Noteworthy, the parameter set which minimizes the χ^2 value maximizes the corresponding likelihood (see Equation 3.3). Furthermore, if the prior distribution in the Bayesian update is uninformative, the parameter set with minimal χ^2 is also associated with the highest possible posterior probability density. Experimental data and the TALYS predictions based on the default parameter set and the optimized parameter set are shown in Figure 6.13.

The calculations so far show the suitability of the nuclear model and that the adjustment of the three most significant model parameters is sufficient to fit the model to the data. In the following we introduce a benchmark procedure for the assessment of the evaluation methods. We are interested in the following characteristics of the evaluation methods:

1. **Descriptive power.** How well does a method describe the experimental data included in the evaluation process?
2. **Predictive power.** How well does a method predict experimental data which had not been included in the evaluation process?

Clearly, both features are associated with the quality of the employed nuclear model. A good model will describe included data well and make accurate predic-

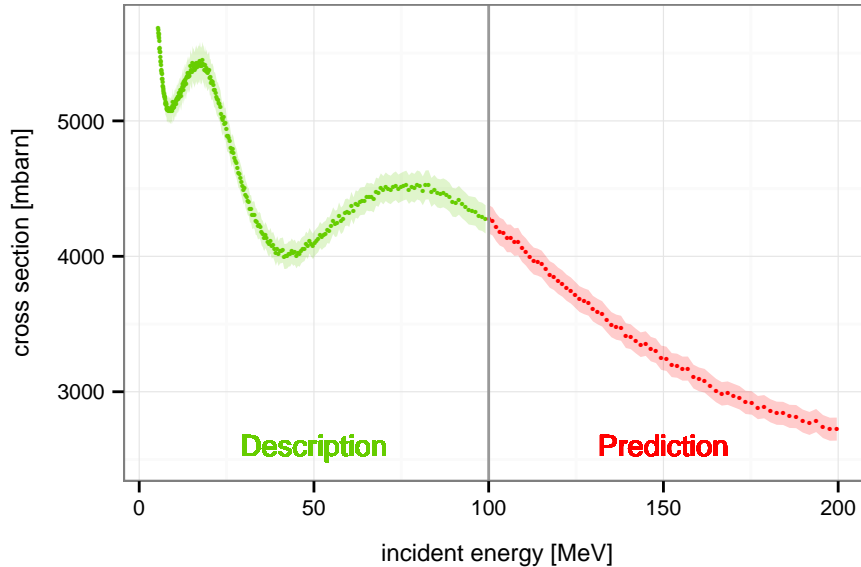


Figure 6.14: Benchmark procedure for the evaluation methods. The experimental data are split into two sets. The set \mathcal{A} on the left half is used in the evaluation process whereas the set \mathcal{B} on the right half is only used to benchmark the evaluated cross sections.

tions. However, the nuclear model has to be seen separate from the evaluation method. An evaluation method is (or at least should be) a statistical procedure to produce reliable estimates. Reliable estimates are not necessarily accurate estimates. Whether the estimates are accurate depends on the quality of the employed nuclear model and is outside the scope of the evaluation method. However, to generate reliable estimates is a key requirement of an evaluation method. Reliable estimates are accompanied by a reasonable assessment of their uncertainty. The use of nuclear models of high quality should lead to sharp predictions of the observables. On the other hand, the use of poor nuclear models should lead to high uncertainties (large standard errors) of the evaluated estimates.

We have already verified that the optical model is able to rather accurately predict the total cross section. In order to study the reliability of the various evaluation methods, we split the experimental data into two sets \mathcal{A} and \mathcal{B} . The set \mathcal{A} contains the data from 5 to 100 MeV and is used in the evaluation process. The set \mathcal{B} contains the data from 100 to 200 MeV and is used to benchmark the reliability of evaluated cross sections. We denote the complete set of experimental data as \mathcal{C} . These sets are displayed in Figure 6.14.

In order to apply a Bayesian evaluation method, a prior distribution for the model parameters is required. For the sake of simplicity, we assume a uniform distribution in the interval $[0.85, 1.15]$ for each of the parameters \tilde{r}_v , \tilde{a}_v and \tilde{v}_1 . For evaluation methods which require a multivariate normal distribution as prior

distribution, we use the standard deviations $\Delta\tilde{r}_v = 0.15$, $\Delta\tilde{a}_v = 0.15$ and $\Delta\tilde{v}_1 = 0.15$. The parameters are always assumed to be uncorrelated. In all considerations we use the center vector is assumed to be $\vec{p}_0 = (1, 1, 1)^T$. The chosen boundaries of the uniform distribution and the chosen standard deviations of the multivariate normal distribution express a rather vague knowledge of the model parameters. The optimal parameters obtained by a fit to the complete dataset in Equation 6.17 are consistent to these choices. In the following analysis of the evaluation methods, the term uncertainty denotes the standard deviation ($\approx 68\%$ confidence interval of the normal distribution).

6.3 Analysis of the EMPIRE-Kalman method

We start with the analysis of the EMPIRE-Kalman method which implements the approach detailed in section 3.2. In this approach the KALMAN code (Kawano and Shibata, 1997) is coupled with the EMPIRE code (Herman et al., 2007). Especially, the update step is performed with Equation 3.25. In order to perform the update, the sensitivity matrix \mathbf{S} is computed at the prior center vector $\vec{p}_0 = (1, 1, 1)$. The setup of the prior covariance matrix and the experimental covariance matrix \mathbf{B} was explained in section 6.2. Updating with the complete set of experimental data \mathcal{C} yields the evaluated parameter set \vec{p}_1 and the associated covariance matrix \mathbf{A}_1 ,

$$\begin{pmatrix} \tilde{a}_v \\ \tilde{r}_v \\ \tilde{v}_1 \end{pmatrix} = \begin{pmatrix} 1.0974 \\ 0.9909 \\ 1.0001 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-6} \begin{pmatrix} 56.2529 & 5.4510 & -8.2628 \\ 5.4510 & 5.6739 & -6.9469 \\ -8.2628 & -6.9469 & 10.6774 \end{pmatrix}. \quad (6.19)$$

The evaluated uncertainties of the parameters are in the order of 10^{-3} . The uncertainties at the level of cross sections are about 0.4%. The evaluated cross sections are shown in Figure 6.15.

The evaluated parameter set differs significantly from the optimal solution in Equation 6.17. Furthermore, the 95% confidence band is inconsistent with the experimental data. These deviations may indicate the following: Either the prior has a significant influence on the inference or the linear approximation of the model at the prior center vector is improper. The latter would be the case if non-linear features of the optical model calculation between the prior center vector and the optimal solution cannot be neglected.

In order to decide which of these two possible causes is true, we extend the EMPIRE-Kalman method to an iterative procedure. In each step, the nuclear model is replaced by a linear approximation with the current parameter vector as expansion point. The Bayesian update formula is then solved analytically which

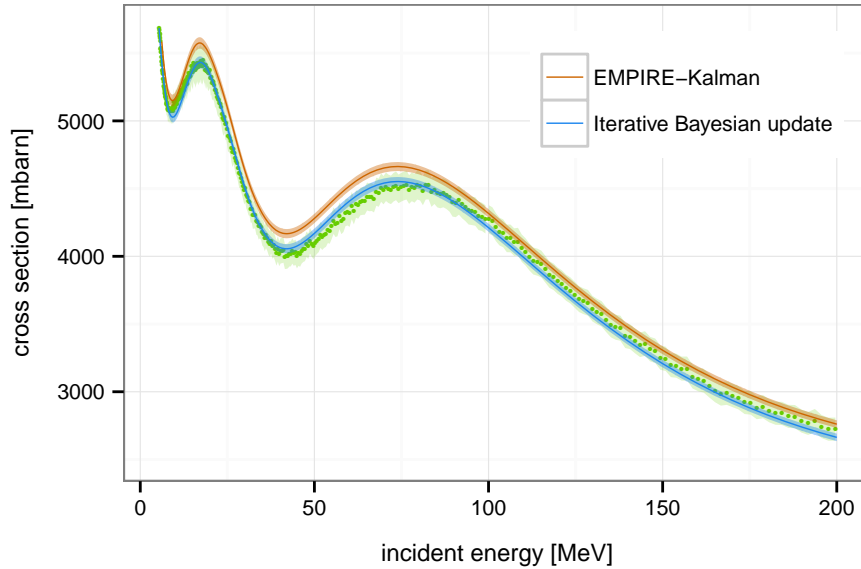


Figure 6.15: Evaluated cross sections obtained by the EMPIRE-Kalman method and the iterative approach. The complete experimental data set \mathcal{C} was included in the evaluation. The shown are the 95% confidence bands of the evaluated cross sections and the experimental data.

yields an estimate of the best posterior parameter vector. The iteration continues until convergence is achieved. This approach is also implemented in the code SAMMY (Larson, 1998) to generate R-matrix fits.

Depending on the initial expansion point \vec{p}_{ref} for the linear approximation, the iterative approach converged to two slightly different solutions. The better solution (in terms of χ^2) is given by

$$\vec{p}_{\text{b}} = \begin{pmatrix} 1.0610 \\ 0.9762 \\ 1.0167 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{b}} = 10^{-6} \begin{pmatrix} 75.4753 & 8.8800 & -13.4959 \\ 8.8800 & 5.0312 & -7.1248 \\ -13.4959 & -7.1248 & 11.7834 \end{pmatrix}. \quad (6.20)$$

The evaluated parameters are almost identical to the best solution in Equation 6.17. Therefore, the prior is uninformative enough and the result of the inference is essentially determined by the information of the highly peaked likelihood. Further, non-linear features of the optical model calculation play an important role and we can expect differences between Monte Carlo methods and methods based on a simplified model.

The cross section prediction of the slightly worse solution is shifted approximately 1% upward compared to the better solution. The evaluated quantities are given by

$$\vec{p}_{\text{w}} = \begin{pmatrix} 1.0701 \\ 0.9810 \\ 1.0104 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{w}} = 10^{-6} \begin{pmatrix} 67.7976 & 8.9601 & -11.9581 \\ 8.9601 & 5.4067 & -6.8238 \\ -11.9581 & -6.8238 & 10.3352 \end{pmatrix}. \quad (6.21)$$

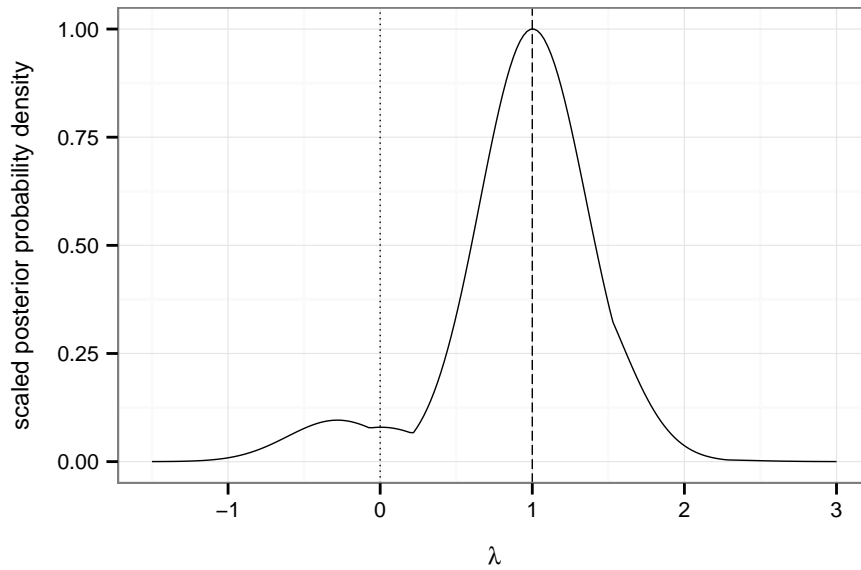


Figure 6.16: Posterior probability density along the straight line defined by the solutions \vec{p}_b and \vec{p}_w . Dotted and dashed line mark the λ value associated with \vec{p}_w and \vec{p}_b , respectively.

The covariance matrices associated with the two solutions obtained by the iterative approach are comparable. The uncertainties of the parameters are of order 10^{-3} and the associated uncertainties of the cross sections are about 0.4%. These uncertainties are comparable to those obtained by the EMPIRE-Kalman method.

We observed a limitation of the accuracy of 10^{-3} in the case of convergence to the slightly worse solution \vec{p}_b . As soon as this accuracy was reached, further iteration steps caused the parameter vector to cyclically jump in the magnitude of this accuracy. It seems as convergence of the iterative linearized Bayesian update approach is not mathematically guaranteed.

The existence of two solutions is associated with a multimodal posterior distribution. We studied its shape by scanning the parameter space along a straight line defined by the two parameter vectors \vec{p}_b and \vec{p}_w . The functional form of this line is

$$\vec{p}(\lambda) = \lambda \vec{p}_b + (1 - \lambda) \vec{p}_w. \quad (6.22)$$

The posterior probability density associated with this line is visualized in Figure 6.16. Noteworthy, along this line there appears another local maximum of the posterior probability density associated with $\lambda \approx -0.3$. The solution \vec{p}_w does not coincide with this local maximum. Even choosing the parameter vector associated with $\lambda = -0.3$ as the initial one, the iterative approach still converges to \vec{p}_w . Therefore, the iterative linearized Bayesian update procedure does not necessarily yield points in parameter space which are associated with local maxima of the

posterior probability density distribution. The reason for this can be attributed again to the linear approximation of the nuclear model.

Because the evaluated cross sections obtained by the EMPIRE-Kalman method are not consistent with experimental data, we restrict the following discussion to the iterative linearized Bayesian update procedure. Several restarts using different initial parameter vectors warrant that the best solution in terms of the χ^2 value is found. Nevertheless, the essential arguments in the further discussion apply equally to the EMPIRE-Kalman method.

Albeit the evaluated cross sections obtained by the iterative approach conform well with the experimental data, the associated uncertainty of 0.4% with regard to the location of the experimental data points may be regarded as too low. If this is the case, it means at least one of the prior assumptions that entered the evaluation is problematic. Several assumptions might be questioned. Following we list some of these assumptions, revise them, and check the change in the evaluated observables:

1. The statistical error in the experimental data points about 1% is underestimated. If we assume a statistical error of 5%, we obtain an evaluated uncertainty of 0.7% for the cross sections.
2. The normalization error in the experimental data points about 0.5% is underestimated. If we assume a normalization error of 5%, we obtain an evaluated uncertainty of 0.6% for the cross sections.
3. Both statistical error and normalization error are underestimated. If we assume a statistical error of 5% and a normalization error of 5%, we obtain an evaluated uncertainty of 2.6% for the cross sections.
4. The uncertainty about the energy dependent detector efficiency is not taken into account. We add to the experimental matrix \mathbf{B} the contribution

$$B_{\text{eff},ij} = 0.05^2 \sigma_i \sigma_j \left(1 - \frac{|E_i - E_j|}{10} \right)_+^{2.5}, \quad (6.23)$$

where E_i is the incident energy associated with the cross section σ_i . The notation $(x)_+$ means that the bracket should evaluate to zero whenever x is negative. This contribution expresses an additional uncertainty for each data point of 5%. The correlations in B_{eff} vanish if the difference between E_i and E_j is larger than 10 MeV. We checked the positive definiteness of this matrix by a Cholesky decomposition. Now the application of the Bayesian update leads to an evaluated uncertainty of 1.3%. If we take instead of 5% only 1% uncertainty due to detector efficiency for each cross section, the evaluation would yield an uncertainty of 0.5%.

5. The assumption that the model is a perfect description of reality does not hold. Maybe because of experiences made in past evaluations, we assume that the model prediction—even with the best possible parameter set—can deviate by 1% from reliably measured experimental data. This type of prior assumption cannot be expressed by a certain choice of the experimental covariance matrix \mathbf{B} .

It may be the case that the rough construction of the experimental covariance matrix \mathbf{B} only based on a global energy independent normalization error and a statistical error is inappropriate. However, most of the adjustments of (1)-(4) may be considered as to warily, yet they still yield evaluated uncertainties ranging from 0.5% to 0.7%. If these are considered too low with regard to the locations of the experimental data points, it has to be due to (5). In chapter 9 we elaborate how a certain distrust in the model can be mathematically expressed to be amenable for an inclusion in the Bayesian update procedure.

We proceed by performing the benchmark explained in section 6.2. The inclusion of only the experimental data set \mathcal{A} in the iterative linearized Bayesian update procedure yields

$$\vec{p}_{\mathcal{A}} = \begin{pmatrix} 1.1032 \\ 0.9557 \\ 1.0389 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\mathcal{A}} = 10^{-6} \begin{pmatrix} 104.4001 & 0.6011 & -6.2205 \\ 0.6011 & 8.0949 & -11.9070 \\ -6.2205 & -11.9070 & 19.2916 \end{pmatrix}. \quad (6.24)$$

Uncertainties of evaluated cross sections are again about 0.4%. For comparison, we also used the EMPIRE-Kalman method to update with \mathcal{A} . The evaluated cross sections are shown in Figure 6.17 and the evaluated correlations in Figure 6.19.

The evaluated cross sections obtained by the iterative approach are inconsistent with the benchmark data set \mathcal{B} which had not been included in the evaluation procedure. In this energy domain, the 95% confidence band of the evaluated cross sections hardly overlaps with the 95% confidence band of the experiment. Therefore, according to the uncertainties of the experimental data, the evaluated cross sections are very unlikely to be true. And conversely, according to the evaluated uncertainties, the experimental data are very unlikely to be true. If we assume that we constructed the experimental covariance matrix \mathbf{B} coherently to our belief about their validity, this undesirable evaluation result must be due to the inappropriate assumption of a perfect model.

Even in the energy domain where data had been included, the evaluated cross sections slightly below 100 MeV are inconsistent with experimental data. The mismatch between data is even more severe for the EMPIRE-Kalman method. The result that the EMPIRE-Kalman evaluation matches well experimental data not included in the update process has to be seen as a coincidence.

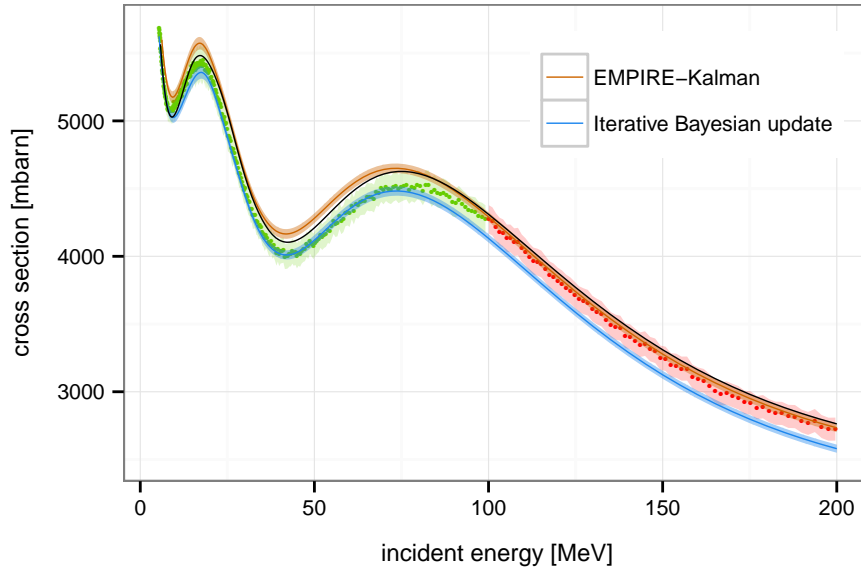


Figure 6.17: Evaluated cross sections obtained by the EMPIRE-Kalman method and the iterative approach. Only data set \mathcal{A} (green points) was included in the evaluation. Shown are the 95% confidence bands of the evaluated cross sections and the experimental data. The black line denotes the cross sections associated with the default parameter set $(1, 1, 1)^T$.

We arrive at the conclusion that accounting for model defects is a necessity in evaluation procedures. Otherwise evaluation results may be inconsistent with experimental data and obtained uncertainties are undesirably low. The use of high quality models do not justify to neglect the treatment of model defects. Continuous technological development will lead to an increasing amount of precise experimental data revealing new features not contained in present models.

6.4 Analysis of the FBET/EMPIRE-MC method

The Full Bayesian Evaluation Technique (FBET) and the EMPIRE-Monte Carlo (EMPIRE-MC) method are based on the surrogate approach detailed in section 3.3. They have also been briefly discussed in section 4.5.

The basic idea is to fit a multivariate normal distribution to an ensemble of model predictions and then to use this multivariate normal distribution as the prior. The development of the FBET is associated with the development of algorithms to construct so-called model defect covariance matrices. If no model defect covariance matrix is used in the evaluation, the FBET and the EMPIRE-MC method are identical. Because we want to study the consequences of model defects, we do not include a model defect covariance matrix in the Bayesian update procedure here. The construction of model defect covariance matrices will be

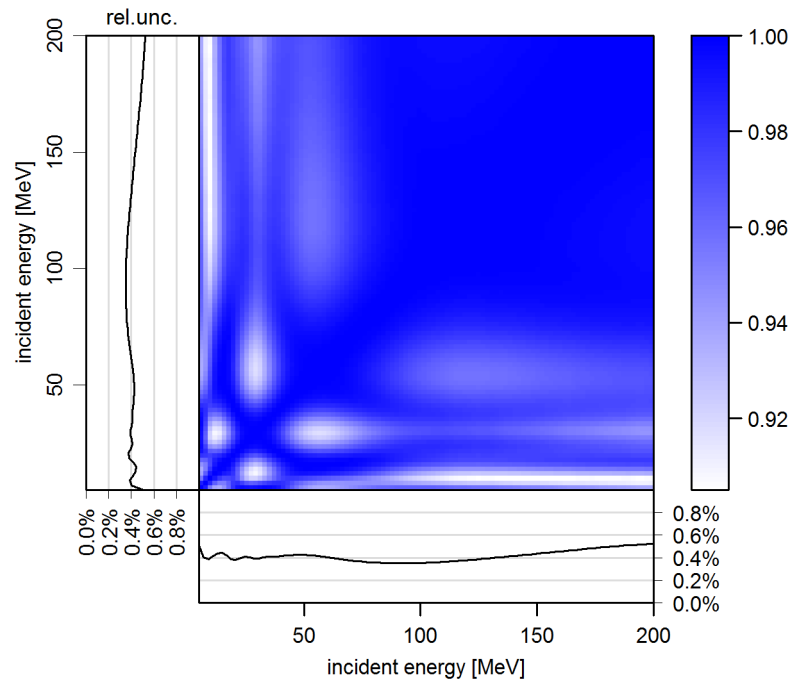


Figure 6.18: Evaluated uncertainties and correlations for the neutron induced total cross section of ^{181}Ta . The iterative linearized Bayesian update was used and the update was performed with the complete experimental data set \mathcal{C} .

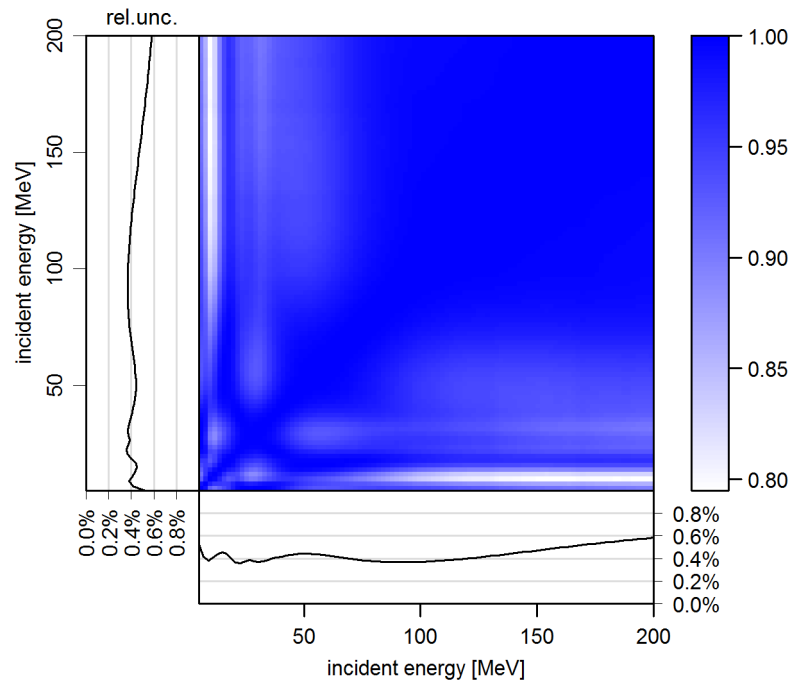


Figure 6.19: Evaluated uncertainties and correlations for the neutron induced total cross section of ^{181}Ta . The iterative linearized Bayesian update was used and the update was performed only with the experimental data set \mathcal{A} .

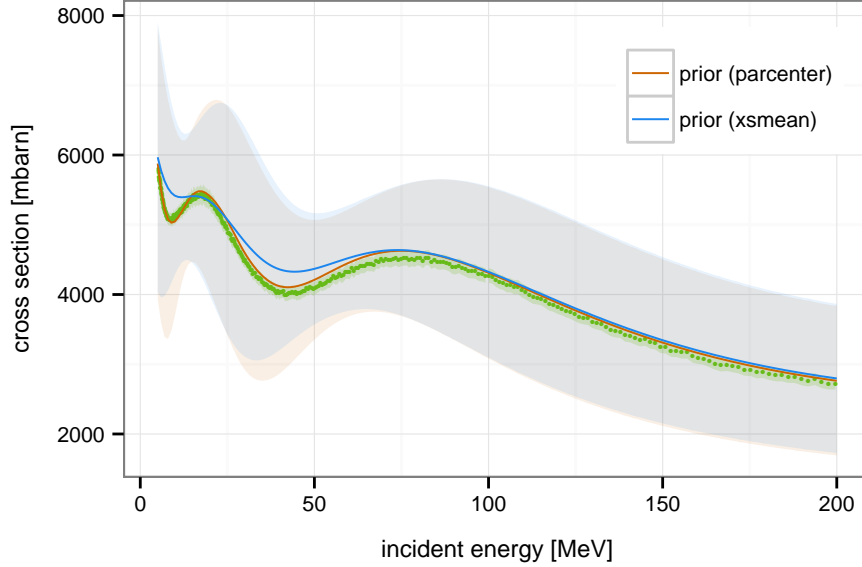


Figure 6.20: Prior cross section curves and associated 95% confidence bands. The cross section curve labeled ‘xsmean’ represents the arithmetic mean of the model prediction vectors. The other curve labeled ‘parcenter’ represents the cross section curve associated with the prior best parameter vector $\vec{p}_0 = (1, 1, 1)^T$.

discussed separately in chapter 9.

As a first step, we draw 10000 parameter vectors $\vec{p}_i = (\tilde{a}_v, \tilde{r}_v, \tilde{v}_1)^T$ from a uniform distribution. The interval of the uniform distribution for each parameter $\tilde{a}_v, \tilde{r}_v, \tilde{v}_1$ is given by $[0.85, 1.15]$. From the associated model prediction vectors $\vec{\tau}_i = \mathcal{M}(\vec{p}_i)$ we calculate the covariance matrix

$$\mathbf{A}_0 = \frac{1}{10000} \sum_{i=1}^{10000} (\vec{\tau}_i - \vec{\sigma}_0)(\vec{\tau}_i - \vec{\sigma}_0)^T. \quad (6.25)$$

The center vector $\vec{\sigma}_0$ is given by the arithmetic mean of the $\vec{\tau}_i$ or the model prediction using the center parameter vector $\vec{p}_0 = (1, 1, 1)^T$. The prior center cross section curve and associated 95% uncertainty band for both choices is shown in Figure 6.20. The difference of these prior cross section curves, especially at lower energies, is due to the non-linearity of the model. The center vector and covariance matrix unambiguously determine the multivariate normal distribution which serves as prior knowledge.

Replacing the non-trivial prior distribution at the level of cross sections by a multivariate normal distribution leads to a less restrictive prior distribution. Cross section curves excluded by the original model are possible solutions when using the surrogate model. Because we allowed variation of only three model parameters ($\tilde{a}_v, \tilde{r}_v, \tilde{v}_1$), admissible combinations of values in a vector $\vec{\sigma}$ containing cross sections at different incident energies are very limited. If for a parameter vector \vec{p} the model yields the cross section vector $\vec{\sigma} = \mathcal{M}(\vec{p})$, admissible cross section vectors in the

vicinity are restricted to a three-dimensional tangent hyperplane. As discussed in section 3.3, non-linearities loosen this restriction in the surrogate model. We want to study the increased flexibility of the surrogate model prior compared to the original prior.

The inspection of the correlation matrix associated with \mathbf{A}_0 is one way to learn about the restrictions. Figure 6.21 and Figure 6.22 show correlations and uncertainties associated with the prior covariance matrix for the two possible choices of the prior center vector $\vec{\sigma}_0$. The correlation structure resulting from these choices is slightly different and the correlations are stronger when using the arithmetic mean. The latter observation is associated with the minimum variance property of the arithmetic mean and therefore of general validity.

The uncertainties ranging from 8% to 20% are large enough to enclose the experimental data in the 68% confidence band. Cross sections at different incident energies are highly positively correlated. This finding indicates that many features of the original model are also preserved in the surrogate model.

We know that the original model possesses three degrees of freedom. In order to determine the degrees of freedom of the surrogate model, we perform an eigendecomposition of the prior covariance matrix. The obtained eigenvalues λ_i represent the variances along the directions given by the associated eigenvectors \vec{e}_i . Because covariance matrices are symmetric, the eigenvectors define an orthogonal basis in cross section space. Thus, it is possible to express any cross section vector $\vec{\sigma}$ whose elements are arbitrary real numbers as a linear combination of the eigenvectors,

$$\vec{\sigma} = \sum_{i=1}^d \beta_i \vec{e}_i. \quad (6.26)$$

The dimension of $\vec{\sigma}$ is denoted by d , which is equal to the number of incident energies in the mesh. For the following, it is particularly convenient to write a cross section as

$$\vec{\sigma} = \vec{\sigma}_0 + \sum_{i=1}^d \beta_i \vec{e}_i, \quad (6.27)$$

where $\vec{\sigma}_0$ is the model prediction based on the best prior parameter vector $\vec{p}_0 = (1, 1, 1)^T$. In Equation 6.27 we expressed a cross section in the eigenbasis of the covariance matrix. In contrast to the original cross sections σ_i contained in $\vec{\sigma}$, the new coordinates β_i are a priori uncorrelated. Therefore, the prior variance of a cross section σ_k at some incident energy E_k can be expressed as

$$\text{Var}[\sigma_k] = \sum_{i=1}^d \text{Var}[\beta_i] e_{ik}, \quad (6.28)$$

where e_{ik} is the k^{th} component associated with the energy E_k of the eigenvector \vec{e}_i . The variances of the β_i are given by the eigenvalues λ_i of the covariance matrix.

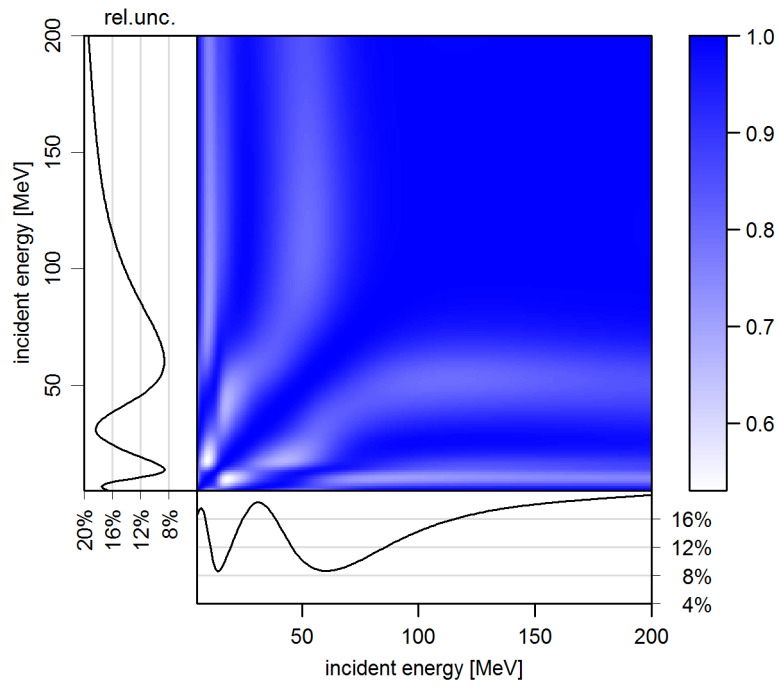


Figure 6.21: Uncertainties and correlations associated with the prior covariance matrix if using $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$.

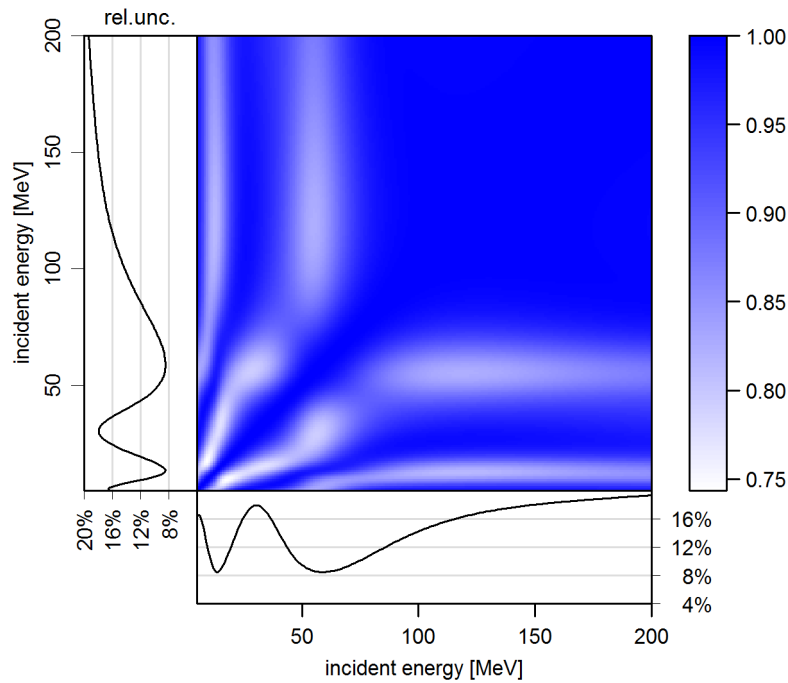


Figure 6.22: Uncertainties and correlations associated with the prior covariance matrix if using the arithmetic mean of the cross section vectors $\vec{\tau}_i$.

If $\text{Var}[\beta_i]$ is close to zero, it means the direction associated with \vec{e}_i is frozen out. Evidence provided by experimental data that the prior cross section vector $\vec{\sigma}_0$ should be changed in the direction of \vec{e}_i will consequently be ignored. Hence, this direction is not relevant in the Bayesian update procedure. In order to determine the relevant eigenvectors, we introduce the vector $\vec{q}_i = \sqrt{\text{Var}[\beta_i]} \vec{e}_i$. The vector \vec{q}_i is a measure for the significance of the direction \vec{e}_i in the Bayesian update procedure. The vectors \vec{q}_i corresponding to the first fifteen eigenvectors associated with largest variance are illustrated in Figure 6.23. Here we used $\vec{\sigma}_0 = \mathcal{M}(\vec{p})$ as the center vector to construct the prior covariance matrix.

Evidently, the number of relevant eigenvectors is about ten. The inclusion of further eigenvectors does not significantly change the result of the Bayesian update procedure. Corresponding changes in the cross sections most likely will not exceed a few millibarn. Therefore, we have about ten degrees of freedom in the surrogate model. Even though this number is significantly higher than for the original model, the surrogate model is still rather restrictive. Effectively, evaluated cross section curves are a linear combination of about ten eigenvectors.

Performing the Bayesian update with the FBET or EMPIRE-MC method and including the complete dataset \mathcal{C} yields the evaluated cross sections illustrated in Figure 6.24. The evaluated correlations are shown in Figure 6.25. For these results we used $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$ as the prior center vector. Using the arithmetic mean leads to the same results and the cross section curves, uncertainties and correlations are indistinguishable.

The evaluated cross sections generated by the iterative linearized Bayesian update illustrated in Figure 6.15 matched the experimental data very well. The evaluated cross sections generated by the FBET/EMPIRE-MC method follow the experimental data perfectly. This is due to the fact that the linear approximation of the model at a certain parameter vector is more restrictive than the surrogate model. We can verify how many degrees of freedom (number of eigenvectors) of the surrogate model are involved in the evaluated cross section curve. To this end, we project the difference vector $\vec{\sigma}_{\text{eval}} - \vec{\sigma}_0$ onto the eigenvectors of the prior covariance matrix. We obtain the evaluated β_i coefficients of Equation 6.27

$$\beta_{\text{eval},i} = \vec{e}_i \cdot (\vec{\sigma}_{\text{eval}} - \vec{\sigma}_0). \quad (6.29)$$

This equation holds when the orthogonal eigenvectors \vec{e}_i of the prior covariance matrix are normalized. In order to determine the eigenvectors of significant influence on the result, we compute for each eigenvector the quantity

$$\delta_{\text{abs},i} = \max_{1 \leq k \leq d} \{ |e_{ik}(\sigma_{\text{eval},k} - \sigma_{0,k})| \}, \quad (6.30)$$

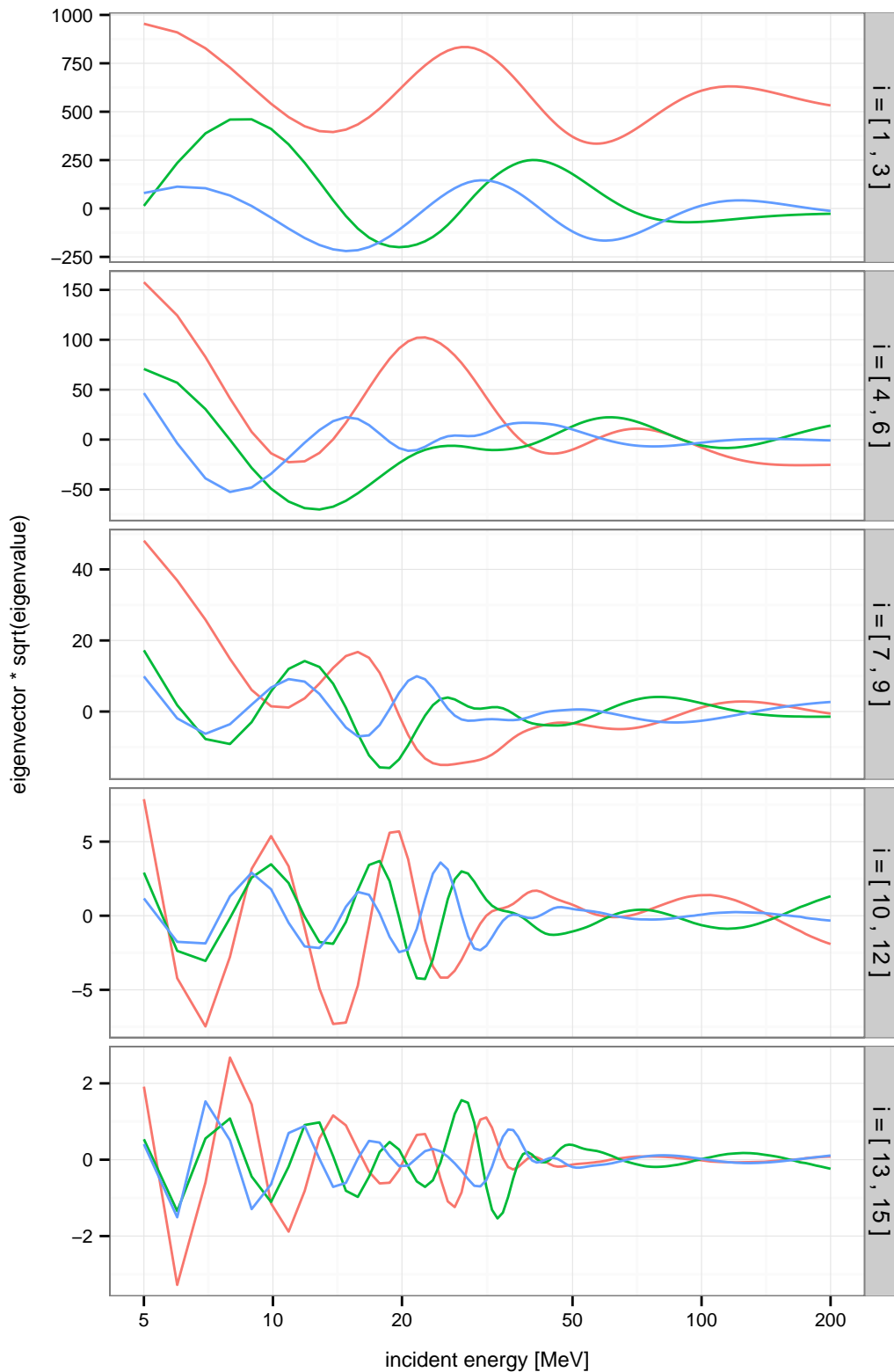


Figure 6.23: Eigenvectors of the prior covariance matrix. Each eigenvector \vec{e}_i is scaled by the square root of the associated eigenvalue λ_i . The annotation $i = [x, x + 2]$ of each plot indicates that the eigenvectors \vec{e}_x (red), \vec{e}_{x+1} (green), and \vec{e}_{x+2} (blue) are shown. Eigenvectors are sorted decreasingly with respect to the associated eigenvalues.

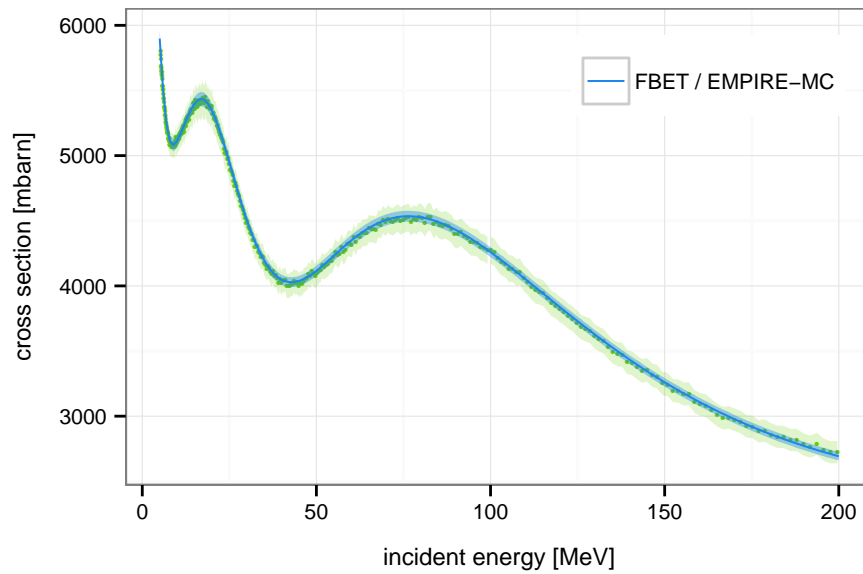


Figure 6.24: Evaluated cross sections and associated 95% confidence band obtained by the FBET/EMPIRE-MC method when updating with the complete experimental data set \mathcal{C} and using $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$.

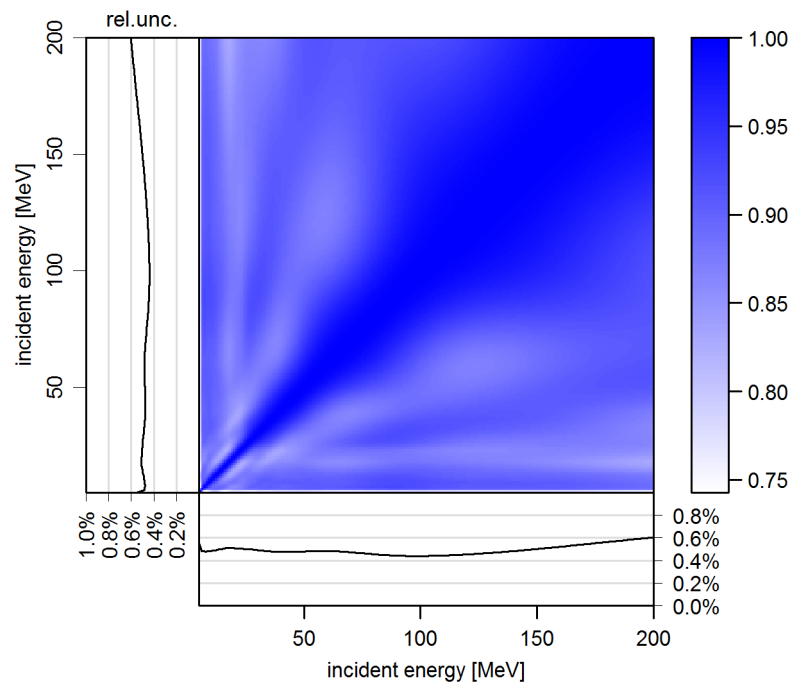


Figure 6.25: Uncertainties and correlations obtained by the FBET/EMPIRE-MC method when updating with the complete experimental data set \mathcal{C} and using $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$.

i	$\beta_{\text{eval},i}$	$\delta_{\text{abs},i}$	$\delta_{\text{rel},i}$	i	$\beta_{\text{eval},i}$	$\delta_{\text{abs},i}$	$\delta_{\text{rel},i}$
1	8.25e+02	9.66e+01	1.95e-02	11	-9.92e+00	2.70e+00	5.14e-04
2	-4.22e+01	1.18e+01	2.35e-03	12	9.89e-01	3.33e-01	6.54e-05
3	-2.68e+02	5.22e+01	9.64e-03	13	2.82e+00	1.46e+00	2.65e-04
4	-6.93e+00	2.36e+00	4.02e-04	14	1.28e+00	3.73e-01	8.51e-05
5	2.44e+02	6.90e+01	1.31e-02	15	6.73e-01	2.68e-01	5.12e-05
6	1.21e+02	4.71e+01	9.27e-03	16	7.22e-02	2.18e-02	4.08e-06
7	-3.13e+02	1.52e+02	2.59e-02	17	-3.53e-03	1.06e-03	1.95e-07
8	-5.83e+01	1.96e+01	3.33e-03	18	7.45e-02	2.55e-02	5.04e-06
9	-8.70e+01	2.28e+01	4.15e-03	19	-3.96e-02	1.22e-02	2.83e-06
10	-3.60e+01	1.05e+01	1.93e-03	20	-5.69e-03	1.72e-03	3.96e-07

Table 6.4: Significance of the directions given by the eigenvectors \vec{e}_i of the prior covariance matrix for the evaluated cross section curve. The coefficient $\beta_{\text{eval},i}$ denotes the difference between the prior cross section vector and the evaluated cross section vector projected onto the direction of the eigenvector \vec{e}_i . The measures $\delta_{\text{abs},i}$ and $\delta_{\text{rel},i}$ defined in Equation 6.30 and Equation 6.31 indicate how much the shift along the direction \vec{e}_i contributes to the evaluated cross section vector.

which picks for a given direction (eigenvector) \vec{e}_i the biggest absolute difference between the prior cross section $\sigma_{0,k}$ and the evaluated cross section $\sigma_{\text{eval},k}$ occurring at some energy E_k . Furthermore, we define

$$\delta_{\text{rel},i} = \max_{1 \leq k \leq d} \left\{ \left| \frac{e_{ik}(\sigma_{\text{eval},k} - \sigma_{0,k})}{\sigma_{0,k}} \right| \right\} \quad (6.31)$$

which yields for the eigenvector \vec{e}_i the biggest relative difference between the prior cross section $\sigma_{0,k}$ and the evaluated cross section $\sigma_{\text{eval},k}$ occurring at some energy E_k . The values of these quantities for the twenty eigenvectors associated with largest variance are stated in Table 6.4. The relative contribution of eigenvectors \vec{e}_i with $i > 10$ is below one tenth of a percent and absolutely measured below five millibarn. This finding supports the conclusion of our prior analysis that about ten eigenvectors are significant in the Bayesian update procedure.

Next, we perform the benchmark procedure specified in section 6.2. We include only the dataset \mathcal{A} in the Bayesian update. The dataset \mathcal{A} contains measured cross sections associated with incident energies between 5 and 100 MeV. The evaluated cross section curve is shown in Figure 6.26 and the associated correlations are shown in Figure 6.27. In contrast to the iterative linearized Bayesian update procedure (see Figure 6.17), the evaluation performed with the FBET/EMPIRE-MC method conforms well with the experimental data. It follows perfectly the data which had been included in the update and resembles well the data used to benchmark the evaluation. The reason will become fully apparent in chapter 9 where we introduce the new evaluation methodology to consistently take into account model defects. Nevertheless, we give a qualitative explanation here.

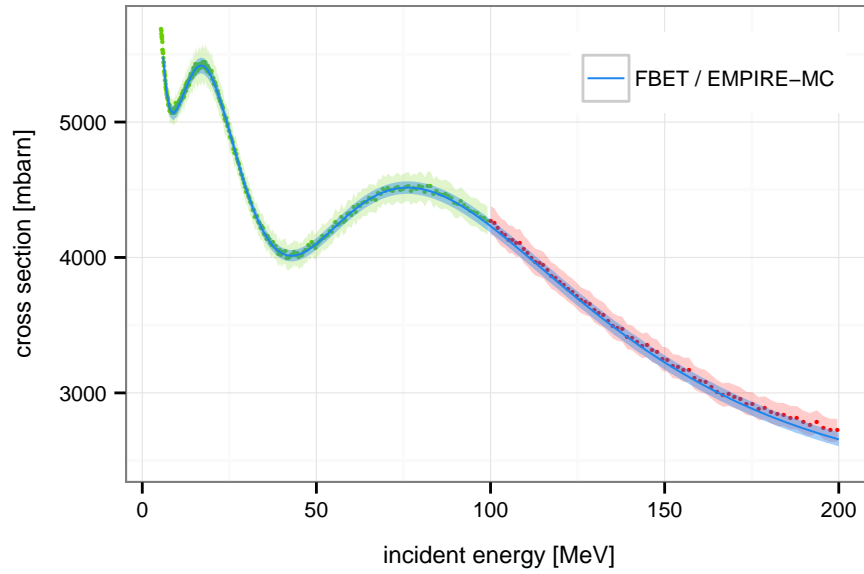


Figure 6.26: Evaluated cross sections and associated 95% confidence band obtained by the FBET/EMPIRE-MC method when only updating with dataset \mathcal{A} and using $\sigma_0 = \mathcal{M}(\vec{p}_0)$

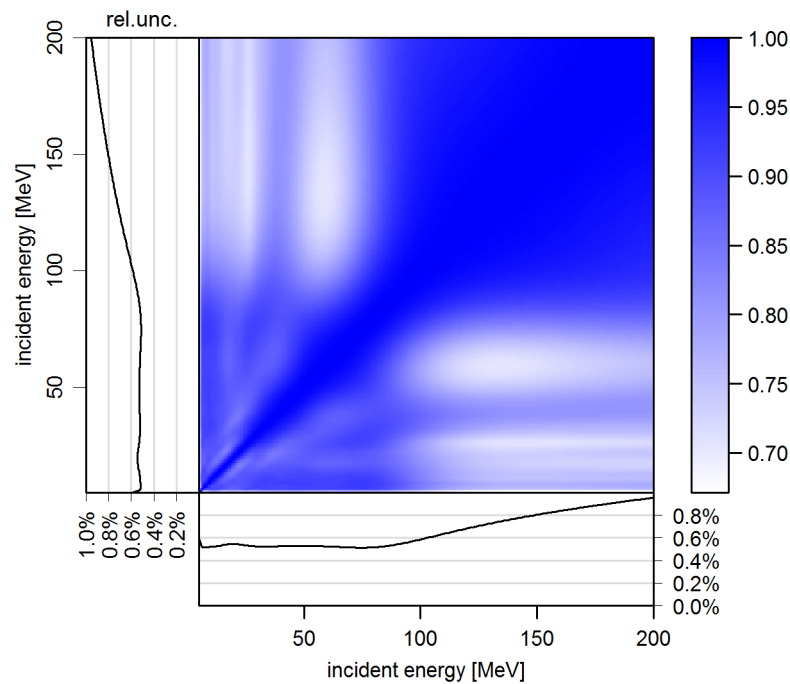


Figure 6.27: Evaluated correlations and uncertainties obtained by the FBET/EMPIRE-MC method when only updating with dataset \mathcal{A} and using $\sigma_0 = \mathcal{M}(\vec{p}_0)$.

The original nuclear model is able to predict the cross section rather well. A fit of the nuclear model to the complete experimental dataset \mathcal{C} yields a description that does not deviate more than 2% from the measurements (see Figure 6.13). However, the model is slightly deficient at low energies. It cannot simultaneously mimic the experimental data in the first valley at approximately 7 MeV and the rest of the data. Due to plenty of experimental data in this valley, the Bayesian update tries to align the model to these data at the cost of a worsened alignment in the domain of higher incident energies. Because the iterative linearized Bayesian update procedure (and also the EMPIRE-Kalman) method is restricted to the possibilities of the model, the evaluation becomes inconsistent. In contrast to that, the surrogate model cannot preserve non-linear relationships. These non-linear features are transformed into additional uncertainties in the surrogate model. The discrepancy between $\bar{\sigma}_0 = \mathcal{M}(\vec{p}_0)$ and the arithmetic mean of the cross section samples at low energies reveals the non-linearity of the optical model in this energy domain (see Figure 6.20). Consequently, the model defect of the original nuclear model disappears in the surrogate model. Furthermore, due to the non-linearity also the connection between cross sections at low incident energies to those of higher incident energies is weakened. This fact is reflected in white stripes appearing near the left edge and the bottom edge of Figure 6.27

At higher incident energies the relationships between cross section are in good approximation linear and the predictive power of the original model is preserved. Including data from 50 to 100 MeV is sufficient to predict the data at incident energies above 100 MeV. To underpin the current argument, we performed the Bayesian update with both the FBET/EMPIRE-MC method and the iterative linearized Bayesian update procedure including only dataset \mathcal{B} . The obtained results are illustrated in Figure 6.28. Evidently, the non-linearity of the original model leads to greater flexibility of the surrogate model at low energies. Hence, also the predictive power of the surrogate model is weakened and the uncertainty band of the FBET/EMPIRE-MC method at low energies is larger than those of the iterative linearized Bayesian update procedure.

This section showed that the FBET/EMPIRE-MC method produces reliable estimates and uncertainties of the neutron-induced total cross section of ^{181}Ta —despite the small deficiency of the nuclear model. However, the conclusion that methods such as FBET or EMPIRE-MC based on the surrogate model are superior to Monte Carlo methods or methods relying on the linearization of the model cannot be made for three reasons. First, model defects do not necessarily appear in energy domains where the nuclear model is highly non-linear. If the model defect would appear in an energy domain where the original model is rather linear, predictions made by the surrogate approach can also be inconsistent

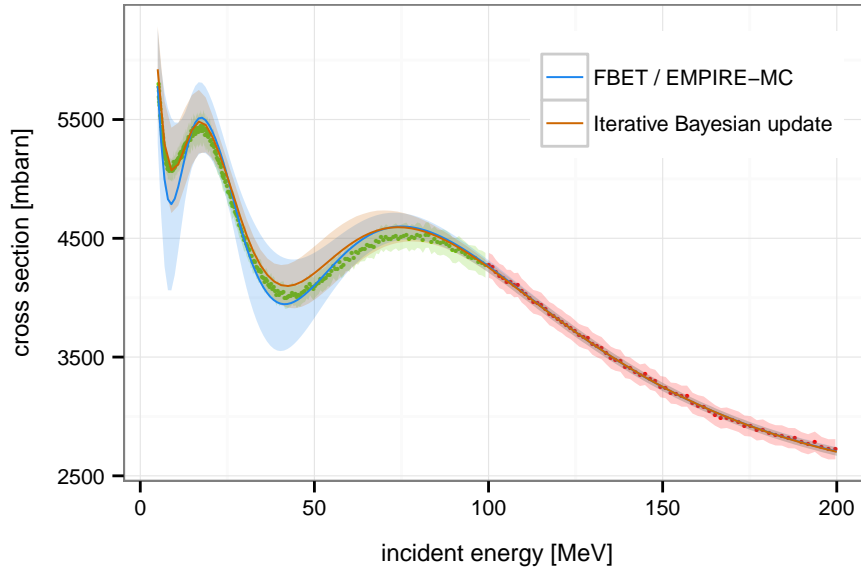


Figure 6.28: Comparison of the evaluated cross sections and associated 95% confidence bands obtained by FBET/EMPIRE-MC and the iterative linearized Bayesian update when only updating with dataset \mathcal{B} (red points).

with experimental data. This fact had been demonstrated in section 6.1. Second, the surrogate model has less predictive power than the original model due to its greater flexibility. Information of the experimental data is not as efficiently used as it could be. Methods such as the iterative linearized Bayesian update procedure or Monte Carlo procedures such as UMC-B have higher predictive power. And third, it remains unclear in which energy domains the model is adequate and in which it is deficient.

6.5 Summary and conclusions

In this chapter we first studied the methods EMPIRE-Kalman, EMPIRE-MC, FBET, UMC-G, UMC-B, BMC and BFMC in an evaluation scenario with a linear model. Due to the application of a linear model, the methods EMPIRE-Kalman, EMPIRE-MC, FBET, UMC-B and UMC-G collapsed to one common approach. Hence, in the following, references to FBET will mean all of them. Only the BMC and BFMC methods have to be considered as different approaches due to an alternative choice of the likelihood.

The experimental data were constructed in a way to be not representable by the linear model. The evaluation on the basis of this deficient model lead to unreasonable low uncertainties obtained by the FBET which assumes a multivariate normal distribution for the likelihood. The evaluated 95% uncertainty band did not enclose the majority of the experimental data points. Furthermore, the al-

lowed variations permitted by the posterior covariance matrix were restricted to the representation possibilities of the nuclear model. Especially in the studied schematic scenario, the experimental data clearly indicated the true cross section curve to be a parabola whereas the predictions of the model were given by straight lines.

In contrast to that, the evaluated uncertainties of the BMC and BFMC were larger compared to those of the FBET, which can be attributed to the alternative choice of the likelihood. However, in the case of the BMC method a non-negligible number of experimental data points were still located outside the 95% confidence band. Furthermore, it may be regarded as an undesirable feature that the likelihood of the BMC method explicitly depends on the number of datasets and data points within these datasets. In which structure the observations are presented should not affect the result of an evaluation. From a conceptual point of view, only the information about the observations, such as the measured values and associated uncertainties, should have an impact on the result.

Also the BFMC method uses a likelihood different from a multivariate normal distribution. The employed likelihood is determined by the smallest χ^2 value within the obtained ensemble of model predictions. Therefore, among all studied methods only the BFMC method takes into account model defects. Consequently, the evaluated uncertainties were larger compared to those of all other methods and consistent with the experimental data. However, as in the case of all other studied methods, the posterior covariance matrix only allowed deviations within the possibilities of the nuclear model.

In addition, we detailed the specifics of the second scenario dealing with the evaluation of the neutron-induced total cross section of ^{181}Ta in this chapter and performed an evaluation with the EMPIRE-Kalman and the FBET (or equivalently EMPIRE-MC) method. The important difference to the first scenario is the non-linear relationship between model parameters and the cross sections at different incident energies established by the optical model. For the evaluation we split the complete dataset \mathcal{C} into the subsets \mathcal{A} and \mathcal{B} containing the data below and above 100 MeV, respectively.

The EMPIRE-Kalman method, which relies on the linearization of the model, did not produce reliable results. Therefore, we extended the method to an iterative procedure as implemented in the code SAMMY. The update with the complete dataset \mathcal{C} led to a rather good reproduction of experimental data. However, considering the evaluated cross section curve with regard to the experimental data, the evaluated uncertainties about 0.4% may be considered too low. The update only with dataset \mathcal{A} led to an inconsistency between the evaluation and the experimental data in dataset \mathcal{B} . The uncertainties of the experimental data excluded

the evaluated cross section curve, and conversely, the evaluated cross section curve excluded the experimental data.

In contrast to that, the evaluation performed with the FBET/EMPIRE-MC approach was consistent with experimental data, independent from the dataset used for the update. We argued that the nuclear model is not able to mimic the data at low and high incident energies equally well due to a model defect at low energies. However, the employed surrogate model is associated with more degrees of freedom than the original model. Therefore, the model deficiency at low incident energies (as a ‘feature’ of the original model) is not preserved in the surrogate model.

We conclude that evaluation methods must account for model defects. Otherwise, evaluated cross section curves might be inconsistent with included experimental data. The modification of the likelihood as in the BMC and BFMC is not the proper solution to deal with model defects, because the posterior covariance matrix still reflects the systematics of the nuclear model. However, in the case of model deficiency, the systematics of the model is inadequate. Also the linearization of the model cannot be regarded as proper measure to deal with model defects. In general, the non-linearity of the nuclear model and a possible deficiency have to be regarded as independent aspects.

Consequences in Monte Carlo methods

We continue the discussion of the evaluation methods at the example of the scenario introduced in section 6.2. We briefly recapitulate the important aspects: The methods are applied in an evaluation of the neutron-induced total cross section of ^{181}Ta . We use the optical model in the global parametrization of Koning and Delaroche (2003) with possible variations restricted to the parameters a_v, r_v, v_1 . For convenience, in the course of this chapter we work with the rescaled model parameters $\tilde{a}_v = a_v/d_1, \tilde{r}_v = r_v/d_2, \tilde{v}_1 = v_1/d_3$ with d_1, d_2, d_3 being the default parameters in the global parameterization of Koning and Delaroche (2003). The experimental data from Finlay et al. (1993) are used for the Bayesian update. The full dataset \mathcal{C} is split in two subsets \mathcal{A} and \mathcal{B} where the former contains the data below 100 MeV and the latter above 100 MeV.

In this section we focus on Monte Carlo methods, hence we will also investigate the efficiency of the applied sampling schemes. All studied methods except the UMC-G method use the original nuclear model without any approximation.

7.1 Analysis of the UMC-G method

The Unified Monte Carlo-G (UMC-G) method described in section 4.3 is based on the surrogate approach. Therefore, the prior distribution is constructed in the same way as in the FBET and the EMPIRE-MC method, which were analyzed in section 6.4. In the evaluation scenario introduced in section 6.2, the likelihood is assumed to be a multivariate normal distribution. Due to this choice, the FBET/EMPIRE-MC method and the UMC-G method evaluate the same multivariate

normal posterior distribution. The only difference between the methods is the determination of the posterior distribution. The FBET/EMPIRE-MC method uses closed-form expressions to compute the mean vector and the covariance matrix of the posterior distribution whereas the UMC-G method uses Monte Carlo sampling to obtain these quantities. The UMC-G method converges for a sufficient large sample size to the solution of the FBET/EMPIRE-MC method. Hence the analysis carried out in section 6.4 also applies to the UMC-G method. Albeit nothing new can be learned about model defects at the example of the UMC-G method, we still want to study its performance in our evaluation scenario. Insights on the efficiency and convergence will be helpful in the discussion of other Monte Carlo procedures.

The accurate approximation of the nuclear model described in section 6.2 enables us to assess the performance of the sampling schemes proposed by Capote and Smith (2008) at the example of the neutron-induced total cross section of ^{181}Ta . The main difference compared to study presented in the paper is the number of experimental data points which are included in the evaluation. The evaluation scenario studied in the paper includes seven experimental data points whereas we are going to include up to 364 points. The wealth of experimental data in our evaluation scenario is associated with a likelihood that is extremely peaked compared to the likelihood considered in the paper.

In the following discussion, we denote with $\vec{\sigma}_0$ and \mathbf{A} the prior center vector and the prior covariance matrix, respectively. These quantities unambiguously specify the shape of the multivariate normal prior distribution. Because the surrogate model approach is applied, this distribution refers to cross section vectors and not model parameters. We denote the vector of measurements by $\vec{\sigma}_{\text{exp}}$ and the associated covariance matrix by \mathbf{B} . In this section we include the complete experimental data set \mathcal{C} in the likelihood.

An essential prerequisite to apply the UMC-G method is the determination of the prior center vector and the prior covariance matrix. For the construction of the prior, we sampled 10000 parameter vectors $\vec{p}_i = (\tilde{a}_{v,i}, \tilde{r}_{v,i}, \tilde{v}_{1,i})$ from a uniform distribution. The range of the uniform distribution for each model parameter is given by [0.85, 1.15]. The model prediction vectors $\vec{\tau}_i = \mathcal{M}(\vec{p}_i)$ associated with the sampled parameter vectors \vec{p}_i are used to construct the covariance matrix,

$$\mathbf{A} = \frac{1}{10000} \sum_{i=1}^{10000} (\vec{\tau}_i - \vec{\sigma}_0)(\vec{\tau}_i - \vec{\sigma}_0)^T, \quad (7.1)$$

where the prior center vector is given by $\vec{\sigma}_0 = \mathcal{M}(\vec{p}_0)$. With the knowledge of the prior distribution and the likelihood, the posterior distribution can be determined by means of Monte Carlo sampling. Capote and Smith (2008) suggested two

sampling schemes: a brute-force scheme and the Metropolis-Hastings algorithm. In the following we benchmark both schemes.

In the *brute-force scheme*, cross section vectors $\vec{\tau}_i$ are drawn from a uniform distribution. The boundaries of the uniform distribution for the elements $\tau_{i,k}$ of the vector $\vec{\tau}_i$ are determined by

$$\sigma_{0,k} - \psi\sqrt{A_{kk}} \leq \tau_{i,k} \leq \sigma_{0,k} + \psi\sqrt{A_{kk}}. \quad (7.2)$$

The boundaries depend on the elements $\sigma_{0,k}$ of the prior center vector, the diagonal elements A_{kk} of the prior covariance matrix and the coefficient ψ . The chosen value of ψ can be expected to have great impact on the efficiency of the UMC-G method.

In the next step, the associated posterior probability density is calculated for each sampled cross section vector $\vec{\tau}_i$,

$$\begin{aligned} \pi(\vec{\tau}_i | \vec{\sigma}_{\text{exp}}) \propto \exp\left(-\frac{1}{2}(\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\tau}_i)^T \mathbf{B}^{-1}(\vec{\sigma}_{\text{exp}} - \mathbf{S}\vec{\tau}_i)\right) \times \\ \exp\left(-\frac{1}{2}(\vec{\tau}_i - \vec{\sigma}_0)^T \mathbf{A}^{-1}(\vec{\tau}_i - \vec{\sigma}_0)\right). \end{aligned} \quad (7.3)$$

The sensitivity matrix \mathbf{S} maps the cross sections from the model mesh to the incident energies of the experimental data. Finally, the posterior center vector can be estimated by

$$\vec{\sigma}_1 = \frac{\sum_{i=1}^n \vec{\tau}_i \pi(\vec{\tau}_i | \vec{\sigma}_{\text{exp}})}{\sum_{i=1}^n \pi(\vec{\tau}_i | \vec{\sigma}_{\text{exp}})}, \quad (7.4)$$

and the posterior covariance matrix is given by

$$\mathbf{A}_1 = \frac{\sum_{i=1}^n (\vec{\tau}_i - \vec{\sigma}_1)(\vec{\tau}_i - \vec{\sigma}_1)^T \pi(\vec{\tau}_i | \vec{\sigma}_{\text{exp}})}{\sum_{i=1}^n \pi(\vec{\tau}_i | \vec{\sigma}_{\text{exp}})}. \quad (7.5)$$

Even though, the brute-force scheme is straight forward, we were unable to apply it successfully. The first obstacle was the reduced rank of the prior covariance matrix \mathbf{A} . The variation of only three model parameters lead to very strong correlations between cross sections at different incident energies. We used a mesh of 200 incident energies for the model calculations. Because of the limited precision of floating point numbers, an eigendecomposition of \mathbf{A} revealed about 100 from the 200 eigenvalues to be almost zero and 56 of them even slightly negative. The smallest eigenvalue was approximately -10^{-7} . The rank deficiency of \mathbf{A} made it impossible to evaluate the posterior probability density with Equation 7.3. Therefore, we applied the procedure explained on page 17 in section 2.2 to deal with covariance matrices that are only positive-semidefinite. Hence we replaced $\vec{\sigma}$, $\vec{\sigma}_0$ and \mathbf{A} in the second exponential function of Equation 7.3 by

$$\vec{\lambda} = \mathbf{P}^T \vec{\sigma}, \quad \vec{\mu} = \mathbf{P}^T \vec{\sigma}_0, \quad \text{and} \quad \mathbf{W} = \mathbf{P}^T \mathbf{A} \mathbf{P}. \quad (7.6)$$

ψ	$\log \tilde{\omega}_2$	$\log \tilde{\omega}_3$	$\log \tilde{\omega}_4$	$\log \tilde{\omega}_5$
1.000	-9.90e+07	-1.09e+08	-1.29e+08	-1.42e+08
0.500	-1.23e+07	-2.01e+07	-2.11e+07	-2.26e+07
0.100	-3.75e+05	-3.90e+05	-8.93e+05	-9.28e+05
0.050	-8.64e+04	-1.03e+05	-1.30e+05	-2.60e+05
0.010	-1.36e+04	-1.76e+04	-1.80e+04	-2.02e+04
0.005	-1.64e+01	-1.86e+03	-2.01e+03	-2.16e+03
0.001	-1.93e+01	-4.77e+01	-5.79e+01	-7.57e+01

Table 7.1: The largest weights occurring in the ensemble of one million cross section vectors for each choice of ψ . The weights are given relative to the largest weight ω_1 occurring in the ensemble, $\tilde{\omega}_i = \omega_i/\omega_1$. Scaled weights are stated in terms of the logarithm to base 10.

The columns of the projection matrix $\mathbf{P} = (\vec{e}_1, \vec{e}_2, \dots)$ are the eigenvectors of the prior covariance matrix which are associated with positive eigenvalues.

We tried different choices of ψ in Equation 7.2 to set the limits of the uniform distribution and generated for each choice $n = 10^6$ cross section vectors. We did not observe convergence for any choice of ψ . The largest weights encountered in the ensemble of cross section vectors for each choice of ψ are summarized in Table 7.1. The largest weight in an ensemble is orders of magnitudes larger than all other weights. Therefore, actually only one term in Equation 7.4 and Equation 7.5 determines the result. Effectively we have drawn just a single cross section vector.

In a real evaluation, increasing the size of the ensemble is not an option. Considering the run times of nuclear model codes, the generation of an ensemble of 10^6 model predictions can already exceed the capabilities. Moreover, we tentatively tried an ensemble size of 10^7 , but the rapid decline of weights persisted.

The essential problem is directly related to the fact that the sampling distribution is determined by the characteristics of the prior. However, in the present evaluation scenario the posterior distribution resembles more the distribution of the sharply peaked likelihood.

The second scheme for the generation of samples is the *Metropolis-Hastings (MH) algorithm* (Hastings, 1970; Metropolis et al., 1953). We explain the MH algorithm in the diction of the current evaluation scenario. However, the MH algorithm can be applied in any scenario where samples should be drawn from a certain probability distribution and other methods, such as inverse transform sampling, do not work.

In order to draw an ensemble of cross section vectors from a probability distribution $p(\vec{\sigma})$, the MH algorithm constructs a chain of samples stepwise. In each step, based on the current cross section vector $\vec{\sigma}_i$ a new cross section vector $\vec{\sigma}'$ is proposed by drawing a sample from the *proposal distribution* $q(\vec{\sigma}' | \vec{\sigma}_i)$. We restrict

the following explanation to proposal distributions which are symmetric in their arguments, $q(\vec{\sigma} | \vec{\tau}_i) = q(\vec{\tau}_i | \vec{\sigma})$. The acceptance of a proposed cross section vector is a matter of chance. A number γ in the interval $[0, 1]$ is drawn from a uniform distribution. If the condition

$$\frac{p(\vec{\tau}')}{p(\vec{\tau}_i)} \geq \gamma \quad (7.7)$$

holds, the proposal is accepted and the next cross section vector in the chain is $\vec{\tau}_{i+1} = \vec{\tau}'$. Otherwise, the current cross section vector $\vec{\tau}_i$ is repeated, $\vec{\tau}_{i+1} = \vec{\tau}_i$. The thereby obtained chain of cross section vectors represents a sample drawn from the probability distribution $p(\vec{\sigma})$.¹

In the paper of Capote and Smith (2008), the proposals of cross section vectors $\vec{\tau}_i$ are generated by

$$\tau'_k = \vec{\tau}_{i,k} + (2\eta - 1)\delta\sqrt{A_{kk}}, \quad (7.8)$$

where η is a number in the interval $[0, 1]$ drawn from a uniform distribution and δ is a constant which specifies the step length. Albeit its similarity to Equation 7.2 there is an essential difference because the center of the uniform distribution in the former expression is given by the current cross section vector $\vec{\tau}_i$ in the chain whereas in the latter equation it is frozen to the center vector of the prior distribution.

The initial cross section vector may lie in a domain in cross sections space associated with small posterior probability. It may require a certain number of steps until the chain enters into the region of significant probability density. This initial stage is called burn-in phase. In order to avoid distortions of the summary statistics, a certain number of cross section vectors drawn at the start of the Monte Carlo chain is usually discarded.

The other crucial aspect is the value of δ which determines the step length. A smaller step length leads to a higher acceptance rate, but also to slower mixing. For small step sizes, subsequent cross section vectors in the chain are located nearby and it takes more steps to explore the probability distribution. Therefore, the optimal choice of δ represents a compromise between acceptance rate and speed of exploration.

In the current evaluation scenario, we know the closed-form solution for the posterior mean vector. We can choose the cross section vector associated with highest posterior probability density as initial cross section vector. Then, no cross section vectors at the start of the chain have to be discarded due to the burn-in phase. This makes it possible to investigate the impact of the choice of δ separately from the burn-in phase. The following table summarizes acceptance rates estimated from 10^4 samples:

¹Technically, the proposal distribution has to meet some criteria to ensure that the cross section vectors in the chain represent an ensemble drawn from the probability distribution $p(\vec{\sigma})$

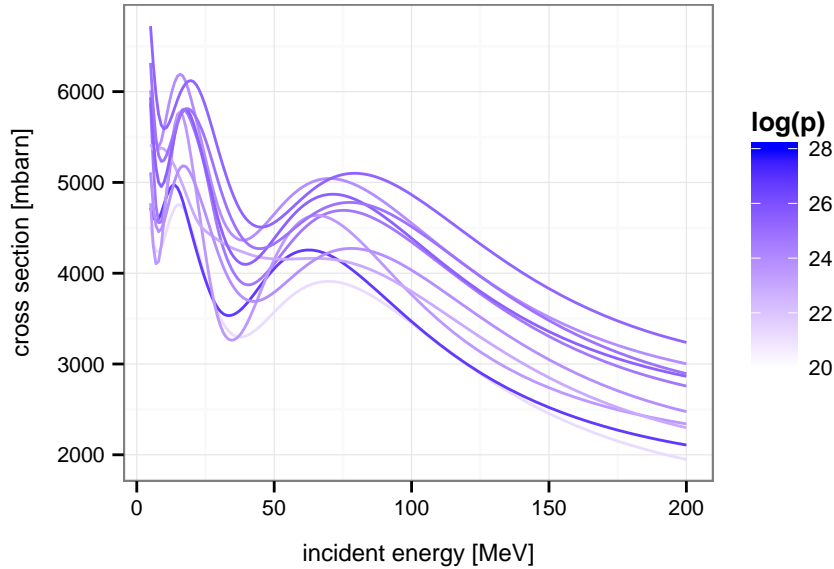


Figure 7.1: Cross section curves drawn from the prior probability distribution of the surrogate model. The curves are colored according to the associated logarithmized prior probability density.

δ	1.0e-07	1.0e-06	5.0e-06	1.0e-05	1.5e-05	2.0e-05	3.0e-05
acc.rate	99.88%	94.99%	71.74%	47.80%	28.85%	16.13%	2.43%

Evidently, the acceptance rate is very sensitive on the chosen value of δ . An acceptance rate between 20% and 40% is considered a good compromise between speed of exploration and the acceptance of samples, see e.g. Brooks (2011). The feasible interval of δ is very narrow. This finding suggests that the proposal distribution is not well aligned to the posterior distribution in the present evaluation scenario. The following consideration supports this assumption.

The prior covariance matrix reflects the features of the nuclear model. For example, the smoothness of the cross section curve is such a feature of the optical model which can be verified by drawing cross section vectors from the prior distribution, see Figure 7.1.² If an experimental data point included in the evaluation corrects the prior cross section at the respective energy upward, also cross sections at nearby incident energies will be pulled upward due to the smoothness constraint reflected in the prior covariance matrix. Hence nearby cross sections are strongly positively correlated. The proposal distribution in Equation 7.8 does not account for this fact, but samples the cross section at every incident energy independently from cross sections at other incident energies. The posterior distributions for two cross sections at nearby incident energies and the proposal distribution are visu-

²When the number of incident energies in the model mesh goes to infinity, predicted cross section vectors become cross section curves. Conversely, a cross section vector with a finite number of elements can be regarded as a discretized version of a cross section curve.

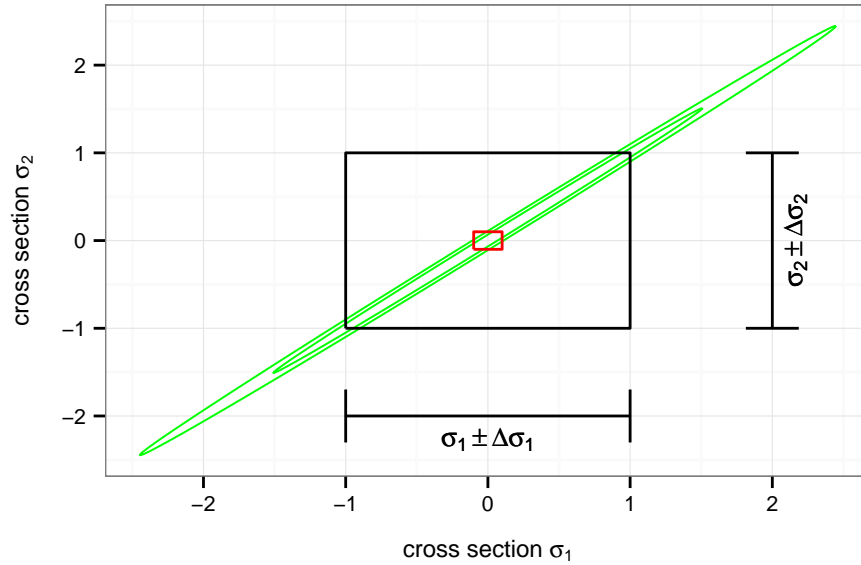


Figure 7.2: Schematic illustration. The green ellipses represent the 68% and 95% confidence areas for two cross sections σ_1 and σ_2 according to the posterior distribution. Proposals are selected within a box whose proportion is determined by the prior standard deviations $\Delta\sigma_1$ and $\Delta\sigma_2$ of the cross sections. Rescaling the larger box by a small constant δ leads to a box size associated with a sensible acceptance rate.

alized in Figure 7.2. It is apparent from the graph that δ has to be a very small value to shrink down the box of the uniform distribution sufficiently. Only then the acceptance rate falls within the desirable range. However, the desirable acceptance rate leads to a bad mixing of the chain. The traversal of the posterior probability distribution along the direction of highest variance requires plenty of steps.

The question of the efficiency of the MH algorithm to explore the cross section space in domains of significant probability density as well as the number of steps associated with the burn-in phase are related. In order to study the number of iterations associated with the burn-in phase, we used the prior center vector $\vec{p}_0 = (\tilde{a}_v, \tilde{r}_v, \tilde{v}_1)^T = (1, 1, 1)^T$ as initial vector for the MH algorithm and investigated the evolution of the probability density associated with the sampled cross section vectors. A representative result is illustrated in Figure 7.3. The chain does not walk up the probability distribution, but instead walks down slightly and remains in a certain range of the posterior probability density. This result is doubtful because we know from the previous sections that the prior center vector is rather unlikely under the posterior distribution. Hence the probability density associated with the cross section vectors should in average increase with an increasing number of iterations.

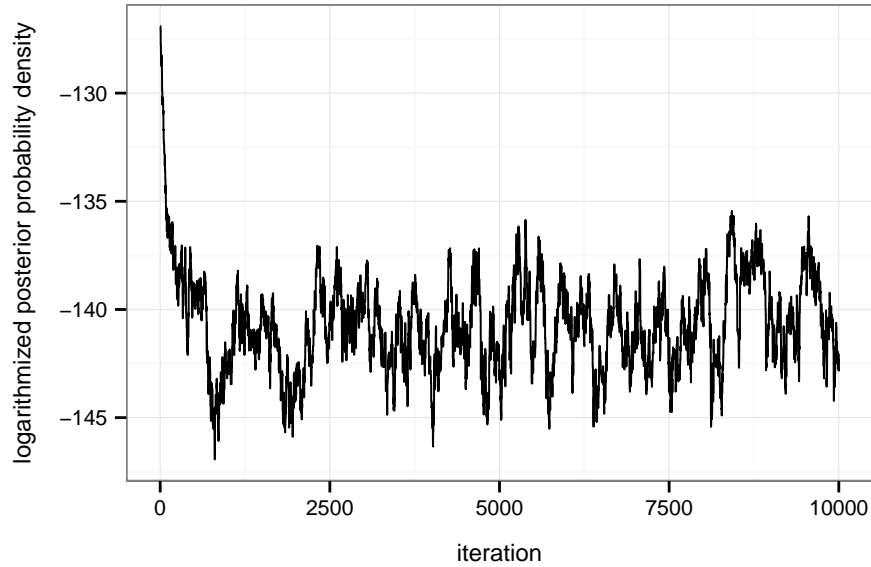


Figure 7.3: Evolution of the posterior probability density associated with the cross sections vectors in the Monte Carlo chain. The proposal function in Equation 7.8 is employed with a step length $\delta = 10^{-5}$. The depicted logarithmized posterior probability density is not normalized.

We argued at the example of Figure 7.2 that the proposal distribution is not well aligned to the posterior distribution in our evaluation scenario. Furthermore, the prior covariance matrix is degenerated due to the variation of only three model parameters. Thus, feasible cross sections are located in a linear subspace of cross section space, yet the proposal distribution gives non-vanishing probabilities for a jump along all directions of the cross section space. In terms of the schematic illustration in Figure 7.2, the posterior distribution is a straight line. Ideally, no finite size of the box indicating the boundaries of the uniform distribution leads to proposals of cross section vectors associated with non-vanishing probability. Practically, we obtained a certain acceptance rate because we projected the proposed cross section vectors onto the subspace of feasible cross section vectors.

To verify the argument, we set the off-diagonal elements of the prior covariance matrix to zero and applied again the MH-algorithm with the proposal function in Equation 7.8 and a step length $\delta = 10^{-2}$. A representative result is shown in Figure 7.4. Now the Monte Carlo chain possesses the expected behavior and climbs up the posterior probability density distribution until cross section vectors fall into a significant domain.

We conclude that the sampling schemes studied by Capote and Smith (2008) are not adequate in scenarios with a large number of experimental data and hence sharply peaked likelihoods. In addition, due to smoothness constraints reflected in the prior distribution, these sampling schemes become less efficient with an

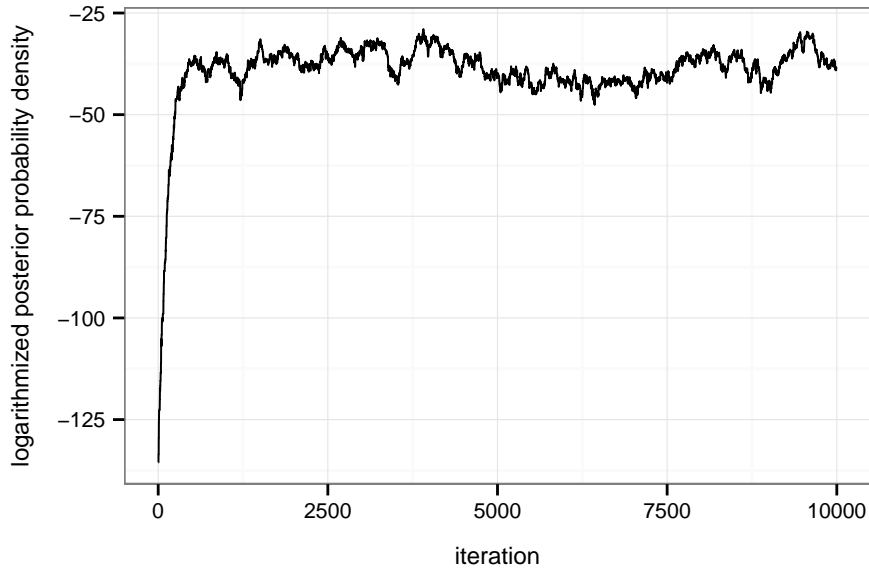


Figure 7.4: Evolution of the posterior probability density associated with the cross sections vectors in the Monte Carlo chain. The proposal function in Equation 7.8 is employed with a step length $\delta = 10^{-2}$. In addition, off-diagonal elements of the prior covariance matrix A_0 were set to zero. The depicted logarithmized posterior probability density is not normalized.

increasing density of incident energies in the model mesh.

7.2 Refined Monte Carlo scheme

The previous section showed that the two Monte Carlo schemes for the UMC-G method suggested by Capote and Smith (2008) are not applicable in our evaluation scenario. Also the paper on the UMC-B approach (Capote, Smith, et al., 2012) leaves the question of an appropriate sampling scheme unanswered. Therefore, an efficient Monte Carlo scheme is needed to make the UMC-G/B approach feasible in cases where the likelihood is sharply peaked.

In this section, we introduce a simple and effective proposal distribution for the Metropolis-Hastings (MH) algorithm. In the further discussion the posterior distribution of the parameter vector \vec{p} is denoted by $\pi_1(\vec{p})$. The basic idea is to linearize the nuclear model at the prior center vector. Then we can use the closed-form expression of Equation 3.24 to compute the posterior covariance matrix \mathbf{A}_1 . We use this covariance matrix to specify the proposal distribution,

$$q(\vec{p}' | \vec{p}) \sim \mathcal{N}(\vec{p}, \delta \mathbf{A}_1), \quad (7.9)$$

where δ determines the step length. In order to decide whether a jump from \vec{p} to the proposed parameter vector \vec{p}' is accepted, a random number η in the interval

$[0, 1]$ is drawn from a uniform distribution. A move is accepted if

$$\frac{\pi_1(\vec{p}')}{\pi_1(\vec{p})} \geq \eta. \quad (7.10)$$

In contrast to the two Monte Carlo schemes presented in the previous chapter, the proposal distribution in Equation 7.9 is adapted to an approximation of the posterior covariance matrix. Consequently, the choice of the step length is far less dependent on the specifics of the evaluation scenario, such as the number of included experimental data points.

We briefly repeat the analysis of the UMC-G method using the MH algorithm and the refined proposal distribution in Equation 7.9. In the UMC-G method the original model is already replaced by a surrogate model. We can directly use the update formula in Equation 3.33. The following table states acceptance rates for different choices of the step length δ :

δ	4e-02	6e-02	8e-02	1e-01	2e-01	3e-01	4e-01
acc.rate	84.43%	77.76%	70.45%	62.77%	34.38%	14.75%	5.23%

The evolution of the posterior probability density associated with the cross section vectors for $\delta = 0.2$ is illustrated in Figure 7.5. After a few hundred iterations the chain enters a domain of significant posterior probability density. We discarded the first thousand cross section vectors and calculated an estimate of the posterior center vector according to Equation 7.4. The maximal relative deviation from the analytic solution is approximately 0.1%.

We will use the MH algorithm with the proposal distribution in Equation 7.9 to obtain results with the UMC-B method. Also the efficiency and accuracy of the BMC and BFMC method will be compared to this MH algorithm.

7.3 Analysis of the UMC-B method

The Unified Monte Carlo-B (UMC-B) method (Capote, Smith, et al., 2012) can be regarded as the most exact evaluation method among all the methods analyzed in this chapter. The combination of two features distinguishes this method from other approaches: the nuclear model without any approximation is used and the likelihood is specified according to the principle of maximum entropy. However, to our knowledge the UMC-B method has not yet been applied in a realistic evaluation scenario. The method was introduced in (Capote, Smith, et al., 2012) and only a demonstration in a schematic example with three experimental data points was given. In this section we apply the UMC-B method to the evaluation of the neutron-induced total cross section of ^{181}Ta with a rich set of available experimental data.

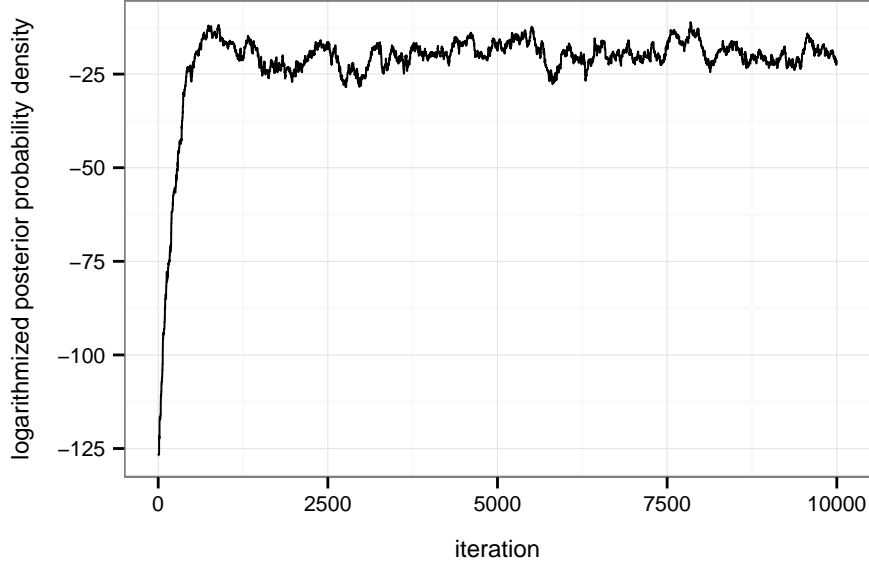


Figure 7.5: Evolution of the posterior probability density associated with the cross sections vectors in the Monte Carlo chain. The prior is specified by the surrogate model. The proposal distribution in Equation 7.9 with a step length $\delta = 0.2$ is used.

The UMC-B method makes use of Bayesian statistics leading to the posterior distribution

$$\pi(\vec{p} | \vec{\sigma}_{\text{exp}}, \mathcal{M}) \propto \ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) \pi(\vec{p} | \mathcal{M}), \quad (7.11)$$

where the vector \vec{p} contains the values of the model parameters \tilde{a}_v , \tilde{r}_v , and \tilde{v}_1 . In our evaluation scenario, the likelihood is a multivariate normal distribution. Therefore it is given by

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \mathcal{M}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{B}|}} \exp \left\{ -\frac{1}{2} (\vec{\sigma} - \mathcal{M}(\vec{p}))^T \mathbf{B}^{-1} (\vec{\sigma} - \mathcal{M}(\vec{p})) \right\}, \quad (7.12)$$

where $\vec{\sigma}_{\text{exp}}$ denotes the vector of cross section measurements and \mathbf{B} is the associated experimental covariance matrix. In contrast to the UMC-G method, the nuclear model \mathcal{M} is not replaced by an approximation in the UMC-B method.³

The approach applied by Capote, Smith, et al. (2012) draws an ensemble of parameter vectors \vec{p}_i from the prior distribution $\pi(\vec{p} | \mathcal{M})$ and determines summary statistics of the posterior distribution in terms of a weighted sum of the associated model predictions $\vec{\tau}_i = \mathcal{M}(\vec{p}_i)$. Especially, the posterior mean vector $\vec{\sigma}_1$ can be estimated by

$$\vec{\sigma}_1 = \frac{\sum_{i=1}^n \vec{\tau}_i \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}{\sum_{i=1}^n \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}. \quad (7.13)$$

³This essential difference between the UMC-G and the UMC-B method remains without remark in the original paper.

Further, the posterior covariance matrix is given by

$$\mathbf{A}_1 = \frac{\sum_{i=1}^n (\vec{\tau}_i - \vec{\sigma}_1)(\vec{\tau}_i - \vec{\sigma}_1)^T \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}{\sum_{i=1}^n \ell(\vec{\sigma}_{\text{exp}} | \vec{p}_i, \mathcal{M})}. \quad (7.14)$$

In the description of the schematic example in (Capote, Smith, et al., 2012), the choice made for the prior distribution $\pi(\vec{p} | \mathcal{M})$ is not explicitly mentioned. However, there are indications that a uniform distribution was used. For reasons discussed in section 7.1, this sampling scheme is not feasible in our evaluation scenario. Due to the highly peaked likelihood, model parameter vectors sampled from the prior distribution are associated with almost vanishing weights.

A simple consideration allows us to estimate the number of samples required to draw one sample with a significant value of the likelihood when sampling from the prior. The iterative Bayesian update yielded for each of the evaluated model parameters \tilde{a}_v , \tilde{r}_v , and \tilde{v}_1 an uncertainty about 10^{-3} (see Equation 6.20). In the current evaluation scenario, the prior distribution is a uniform distribution for each parameter in the range $[0.85, 1.15]$. Therefore, the probability that a parameter vector drawn from the prior distribution has significant weight is approximately given by

$$\frac{(10^{-3})^3}{0.3^3} = \frac{10^{-9}}{27 \cdot 10^{-3}} \approx 3 \cdot 10^{-6}. \quad (7.15)$$

Hence, only one sample out of $3 \cdot 10^6$ samples can be expected to be associated with a significant value of the likelihood. Clearly, this sampling scheme is inapplicable.

Instead of sampling from the prior distribution as done by Capote, Smith, et al. (2012), we use the Metropolis-Hasting algorithm to directly draw model parameter vectors from the posterior distribution $\pi(\vec{p} | \vec{\sigma}_{\text{exp}})$. In order to construct the proposal distribution introduced in section 7.1, we compute an approximation of the posterior covariance matrix. To this end, we linearize the nuclear model at the prior center vector and apply the analytical update formula in Equation 3.22. We already performed such a calculation for the EMPIRE-Kalman method. The resulting posterior covariance matrix for the model parameters is given in Equation 6.19.

We included the complete experimental data set \mathcal{C} in the specification of the likelihood and created a Monte Carlo chain with $n = 10^5$ parameter vectors $\vec{\gamma}_i$ using the step length $\delta = 0.8$ in the proposal distribution yields an acceptance rate of 34%. Figure 7.6 shows the evolution of the posterior probability density. Comparing this posterior probability density with those illustrated in Figure 7.5, we notice that the present probability density is much lower. This is due the fact that the UMC-B method relies on the exact model and the UMC-G method on a more flexible surrogate model, which can better adapt to the data. Furthermore,

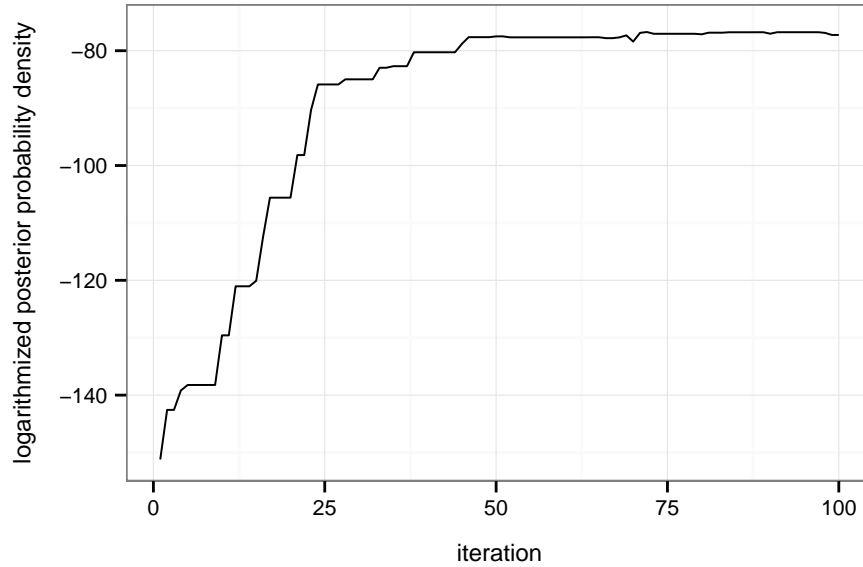


Figure 7.6: Evolution of the posterior probability density associated with the cross sections vectors in the Monte Carlo chain. The prior is specified by the a uniform distribution for the model parameters and the exact model is employed. The proposal distribution in Equation 7.9 with a step length $\delta = 0.2$ is used. The depicted logarithmized posterior probability density is not normalized.

the burn-in phase is rather short. After about fifty iterations, there is no visible trend anymore.

Because the obtained parameter vectors represent an ensemble drawn from the exact posterior distribution, the mean vector \vec{p}_1 and the covariance matrix \mathbf{A}_1 of the posterior distribution are determined by the standard formulas

$$\vec{p}_1 = \frac{1}{n} \sum_{i=1}^n \vec{\gamma}_i \quad \text{and} \quad \mathbf{A}_1 = \frac{1}{n} \sum_{i=1}^n (\vec{\gamma}_i - \vec{p}_1) (\vec{\gamma}_i - \vec{p}_1)^T. \quad (7.16)$$

Using these formulas, we obtain

$$\begin{pmatrix} \tilde{a}_v \\ \tilde{r}_v \\ \tilde{v}_1 \end{pmatrix} = \begin{pmatrix} 1.0614 \\ 0.9766 \\ 1.0162 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-6} \begin{pmatrix} 90.0749 & 13.1820 & -19.2156 \\ 13.1820 & 6.8861 & -9.3364 \\ -19.2156 & -9.3364 & 14.5449 \end{pmatrix}. \quad (7.17)$$

In order to assess the accuracy of these estimates, we created several Monte Carlo chains each consisting of 10^5 samples. The fluctuations observed in the mean vector were in the order of 10^{-4} . The fluctuations in the diagonal of the covariance matrix were about 10^{-6} or less.

The obtained mean vector is approximately the same as in the iterative linearized Bayesian update procedure (see Equation 6.20). This finding reflects the sharply peaked likelihood which reduces the uncertainty in the model parameters

to a linear domain of the model. However, the difference between the respective covariance matrices is statistically significant. The diagonal elements of the covariance matrix produced by the iterative linearized Bayesian update are roughly 20% smaller. This deviation can be attributed to a local maximum of the posterior probability distribution. The marginal density distributions for pairs of model parameters are shown in Figure 7.10, Figure 7.12, and Figure 7.14.

The linearization of the model in the iterative Bayesian update procedure makes this method blind for local maxima. If local maxima exist, uncertainties are then probably underestimated. In contrast to that, a Monte Carlo method based on the Metropolis-Hastings algorithm is able to discover these local maxima. In the present situation, a linearized method might be still regarded more efficient because the sharply peaked likelihood restricts the nuclear model approximately to a linear domain. The confidence regions enclosing the global maximum are of elliptical shape which is a signature of linearity. Only in confidence areas associated with high certainty enclosing a larger domain in the space of parameter vectors, the non-linearity of the model becomes noticeable.

In order to study the reliability of the predictions obtained by the UMC-B method, we made an additional evaluation only including the dataset \mathcal{A} in the likelihood. Based on a Monte Carlo chain with 10^5 parameter vectors and using a step length $\delta = 0.8$, we obtained

$$\begin{pmatrix} \tilde{a}_v \\ \tilde{r}_v \\ \tilde{v}_1 \end{pmatrix} = \begin{pmatrix} 1.1037 \\ 0.9569 \\ 1.0372 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-6} \begin{pmatrix} 112.3212 & 0.5221 & -6.9062 \\ 0.5221 & 11.7173 & -16.6597 \\ -6.9062 & -16.6597 & 25.6224 \end{pmatrix}. \quad (7.18)$$

The evaluated cross section curve is visualized in Figure 7.7. The associated uncertainties and correlations are shown in Figure 7.8. Due to the small deficiency of the model at low incident energies, the model prediction is inconsistent with the dataset \mathcal{B} , which had not been included in the likelihood.

The comparison of the evaluated correlations with those resulting from an update with \mathcal{C} (see Figure 7.9) yields at first glance a surprising result: The uncertainties of the evaluated cross sections are larger when updating with the complete dataset \mathcal{C} than when only updating with \mathcal{A} . This observation may seem doubtful because new information should never increase the uncertainty. Despite this observation, the uncertainty is indeed reduced, but not along directions associated with the diagonal of the covariance matrix for the cross sections. The reduced uncertainty is reflected in the higher correlations of the covariance matrix resulting from an update with \mathcal{C} . The reduced uncertainty can also be verified at the evaluated parameter covariance matrix, in which the diagonal elements are indeed

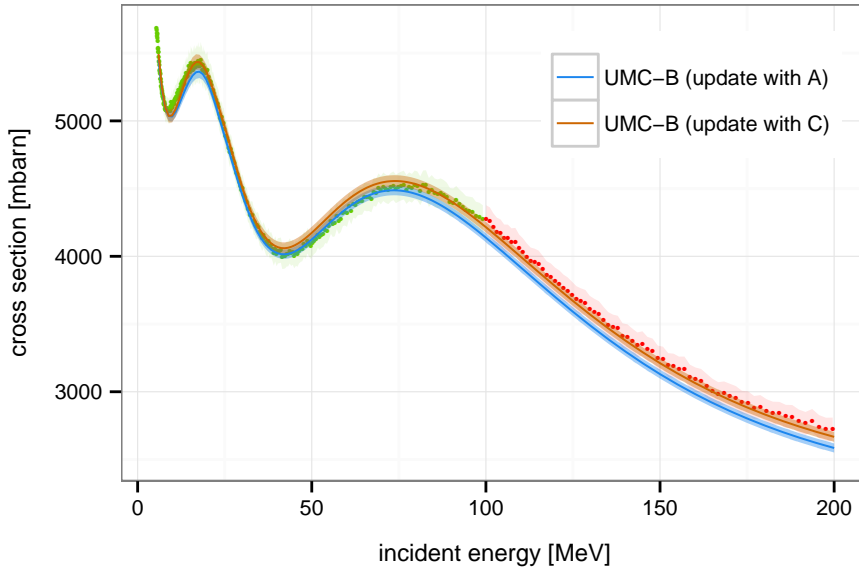


Figure 7.7: Evaluated cross sections obtained by the UMC-B method. In one evaluation the update was performed with dataset \mathcal{A} (red) whereas in the other evaluation the update was performed with dataset \mathcal{C} . The shown confidence bands indicate the 95% trust level.

smaller than those obtained with the full data set \mathcal{C} .

The comparison of the results obtained by the UMC-B method and the iterative linearized Bayesian update Equation 6.24 shows that evaluated mean vectors are in very good agreement. As in the case of only updating with \mathcal{C} , the differences in the covariance matrices are statistically significant. The increased uncertainty in the case of UMC-B can be attributed to the non-linearity of the model. Figure 7.11, Figure 7.13, and Figure 7.15 show marginal density plots of the posterior distribution of the model parameters.

Posterior distributions that substantially differ from the multivariate normal distribution can only be accurately treated with Monte Carlo methods. In the presence of significant non-linearities, methods based on either the linearization of the model (see section 3.2) or the surrogate approach (see section 3.3) produce different results. Whether evaluated uncertainties of methods linearizing the model are lower or higher compared to those obtained by Monte Carlo methods depends on the specific evaluation scenario. In contrast to that, evaluated uncertainties produced by methods based on the surrogate approach are always higher and yield the most cautious predictions. However, it is also not desirable to be overly cautious in evaluations because results could lead to the wrong conclusion that some experiment has to be carried out to get more certainty. Another consequence could be that nuclear facilities are not as efficiently constructed as they could be.

Consequently, it is desirable to keep the complete (non-linear) systematics of the

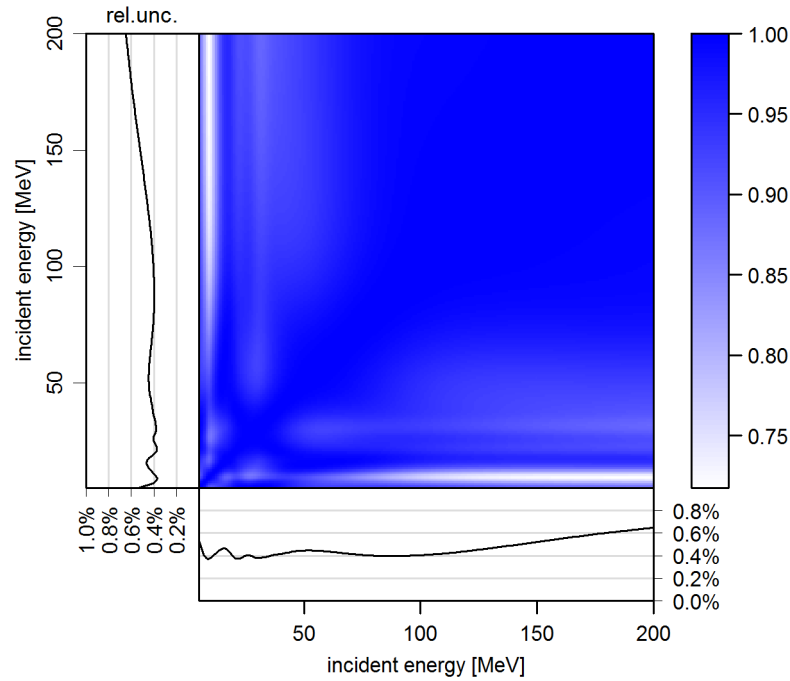


Figure 7.8: Evaluated uncertainties and correlations obtained by the UMC-B method when including the complete dataset \mathcal{A} . These estimates of the uncertainty and the correlations are based on 10^5 parameter vectors drawn from the posterior distribution.

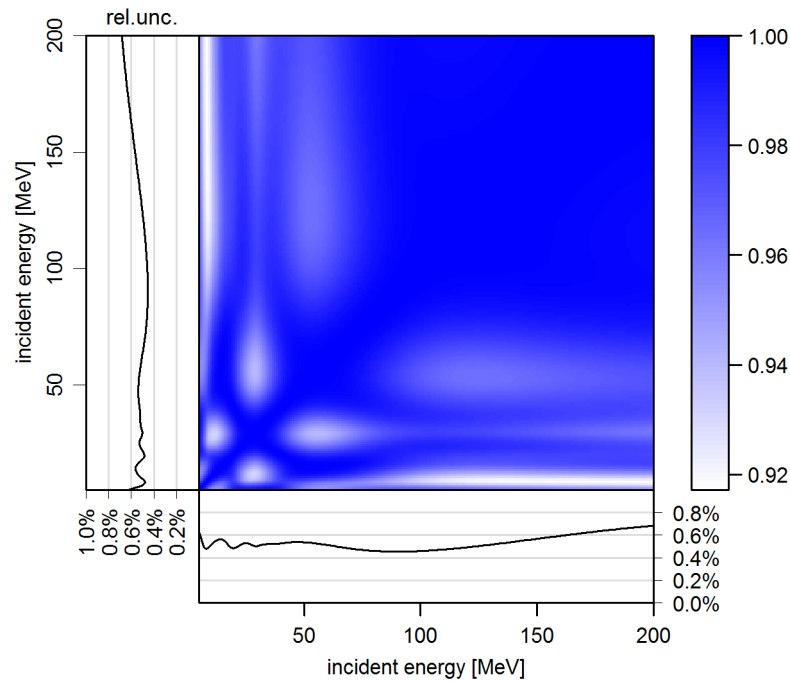


Figure 7.9: Evaluated uncertainties and correlations obtained by the UMC-B method when including the complete dataset \mathcal{C} . These estimates of the uncertainty and the correlations are based on 10^5 parameter vectors drawn from the posterior distribution.

nuclear model to get accurate predictions and tight confidence intervals. However, in the evaluation with the UMC-B method including only the data set \mathcal{A} the predicted cross section curve (illustrated in Figure 7.7) was misleading. Therefore, it is of utmost importance to account for the expected reliability of the nuclear model in the prior distribution. The mathematical modeling of this trust is the subject of chapter 9.

We have extensively analyzed the UMC-B method in this section for three reasons. First, the exact nuclear model is taken into account for the inference and consequently the results of the Bayesian inference are not distorted or diluted by approximations of the nuclear model. Second, the likelihood is a multivariate normal distribution which is a choice based on sound principles (see section 3.1). Third, using the Metropolis-Hastings algorithm provides a theoretical guarantee that we indeed sample from the posterior distribution. Thus, every part of the update procedure stands on solid ground. If a Monte Carlo method does not converge to the result of the UMC-B method, it necessarily disagrees with one of the three mentioned aspects.

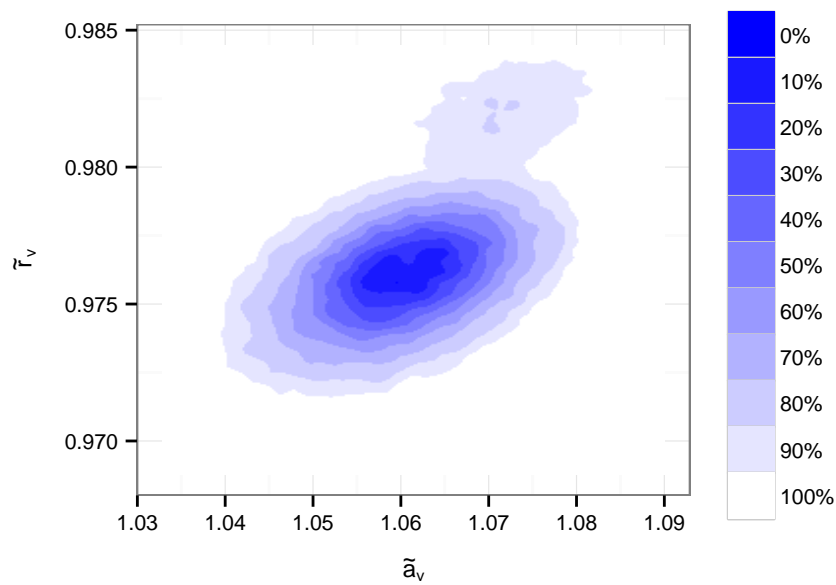


Figure 7.10: Marginal density distribution of the posterior where \tilde{v}_1 has been integrated out. The complete dataset \mathcal{C} entered the likelihood. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

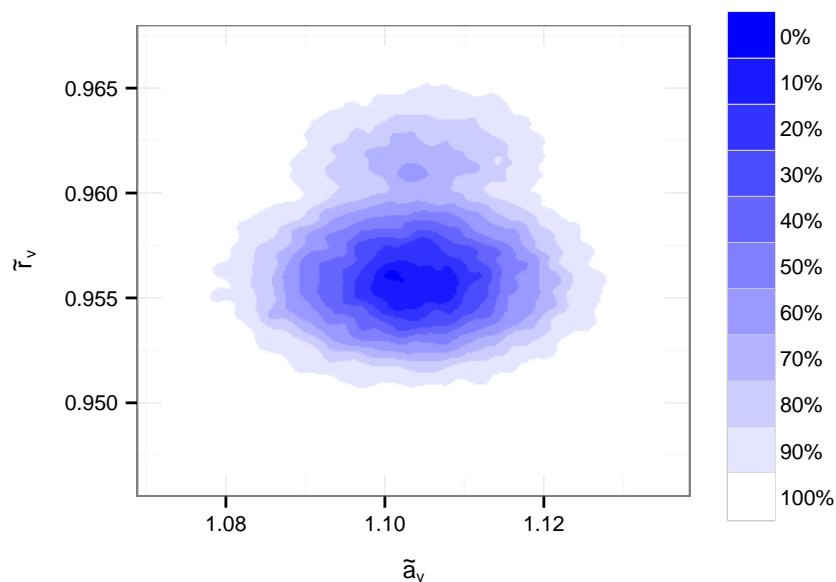


Figure 7.11: Marginal density distribution of the posterior where \tilde{v}_1 has been integrated out. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

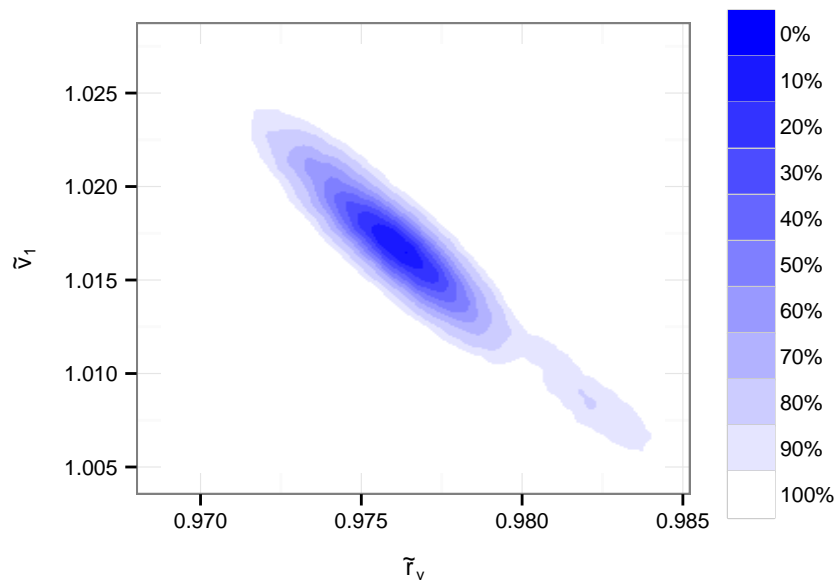


Figure 7.12: Marginal density distribution of the posterior where \tilde{a}_v has been integrated out. The complete dataset \mathcal{C} entered the likelihood. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

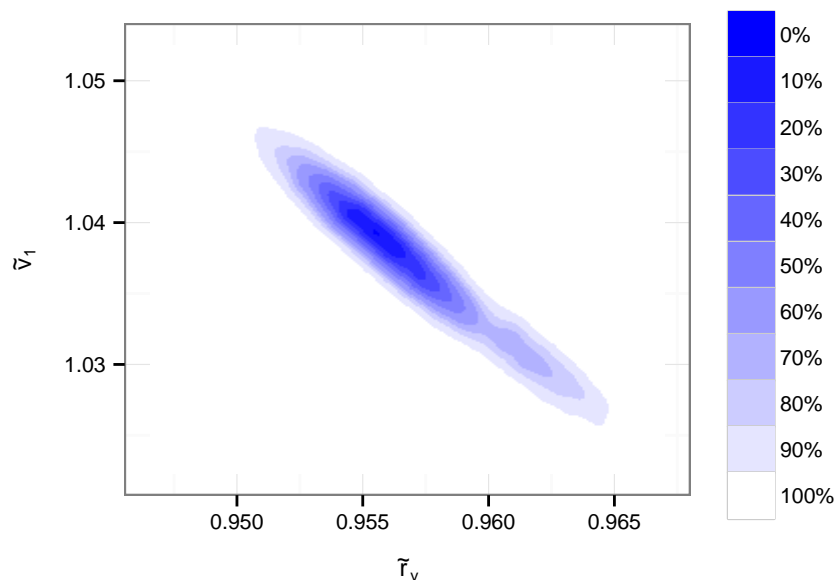


Figure 7.13: Marginal density distribution of the posterior where \tilde{a}_v has been integrated out. Only the dataset \mathcal{A} entered the likelihood. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

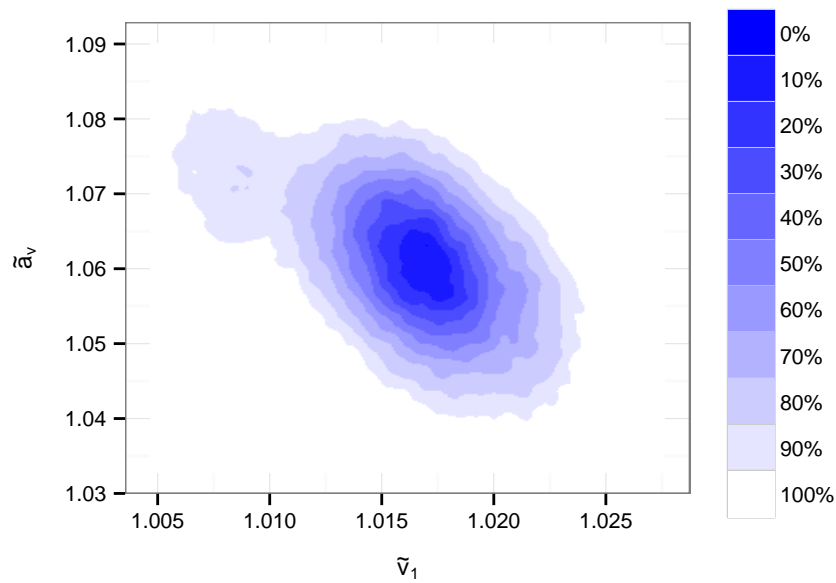


Figure 7.14: Marginal density distribution of the posterior where \tilde{r}_v has been integrated out. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

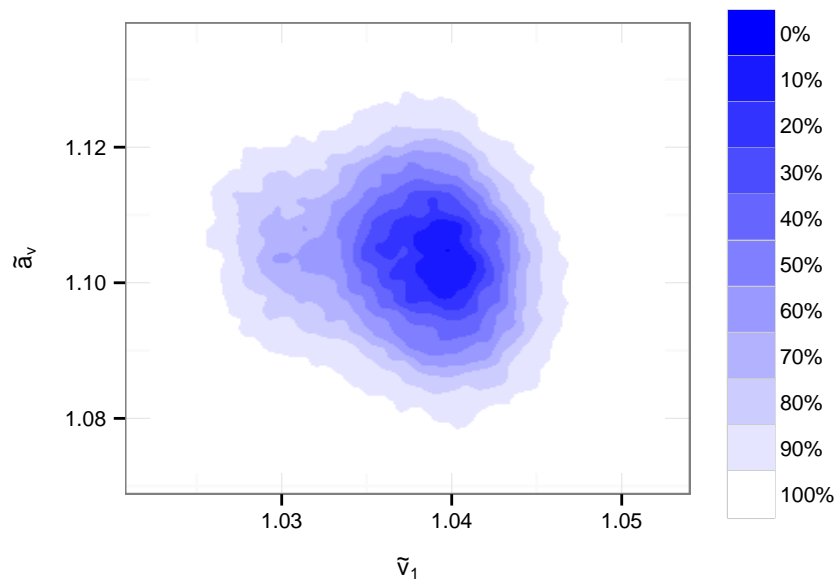


Figure 7.15: Marginal density distribution of the posterior where \tilde{r}_v has been integrated out. Only the dataset \mathcal{A} entered the likelihood. The optical model without any approximation is used to map the model parameters to cross sections. The contours denote confidence areas.

7.4 Analysis of the BMC method

The Bayesian Monte Carlo (BMC) method (Koning, 2015) uses Monte Carlo sampling to obtain summary statistics of the posterior distribution. The specification of the likelihood (see Equation 4.17) is different from that of the UMC-B method. Furthermore, the BMC method implements a certain kind of self-adaptive importance sampling scheme.

Before we elaborate on the specific details of the self-adaptive sampling scheme, we outline the basics of importance sampling. Assume that we want to compute a summary statistics $f(\vec{p})$ of the probability distribution $\pi(\vec{p})$. One way to achieve this goal is to evaluate the integral

$$\mathbb{E}_\pi[f(\vec{p})] = \int_{\mathcal{S}} f(\vec{p})\pi(\vec{p}) \, d\vec{p}. \quad (7.19)$$

The subscript π indicates that the expectation is with respect to the probability distribution $\pi(\vec{p})$. The integration has to be performed over the support \mathcal{S} of $\pi(\vec{p})$. If the integral is not tractable by means of deterministic quadrature schemes, an alternative is the sampling of \vec{p}_i according to the distribution $\pi(\vec{p})$ and the estimation of the integral by

$$\mathbb{E}_\pi[f(\vec{p})] \approx \frac{1}{n} \sum_{i=1}^n f(\vec{p}_i). \quad (7.20)$$

In some cases, directly sampling from $\pi(\vec{p})$ is also not possible. Then, one can sample instead from another so-called *instrumental distribution* $\phi(\vec{p})$. The idea behind this modified evaluation of the integral is best seen rewriting Equation 7.19 as

$$\mathbb{E}_\pi[f(\vec{p})] = \int \frac{f(\vec{p})\pi(\vec{p})}{\phi(\vec{p})} \phi(\vec{p}) \, d\vec{p}. \quad (7.21)$$

Interpreting the fraction in the integrand as a summary statistics of $\phi(\vec{p})$, an estimate of the integral is given by

$$\mathbb{E}_\phi[f(\vec{p})] \approx \frac{1}{n} \sum_{i=1}^n \frac{f(\vec{p}_i)\pi(\vec{p}_i)}{\phi(\vec{p}_i)}. \quad (7.22)$$

The number of samples to obtain a certain accuracy can be significantly reduced if the distribution $\phi(\vec{p})$ resembles closely the original distribution $\pi(\vec{p})$. If one of the probability distributions is not normalized, the following integral yields $c \neq 1$,

$$\int \pi(\vec{p}) \, d\vec{p} = \int \frac{\pi(\vec{p})}{\phi(\vec{p})} \phi(\vec{p}) \, d\vec{p} = c. \quad (7.23)$$

In this case, we must estimate c as well,

$$\mathbb{E}_\phi[1] \approx \frac{1}{n} \sum_{i=1}^n \frac{\pi(\vec{p}_i)}{\phi(\vec{p}_i)} \approx c \quad (7.24)$$

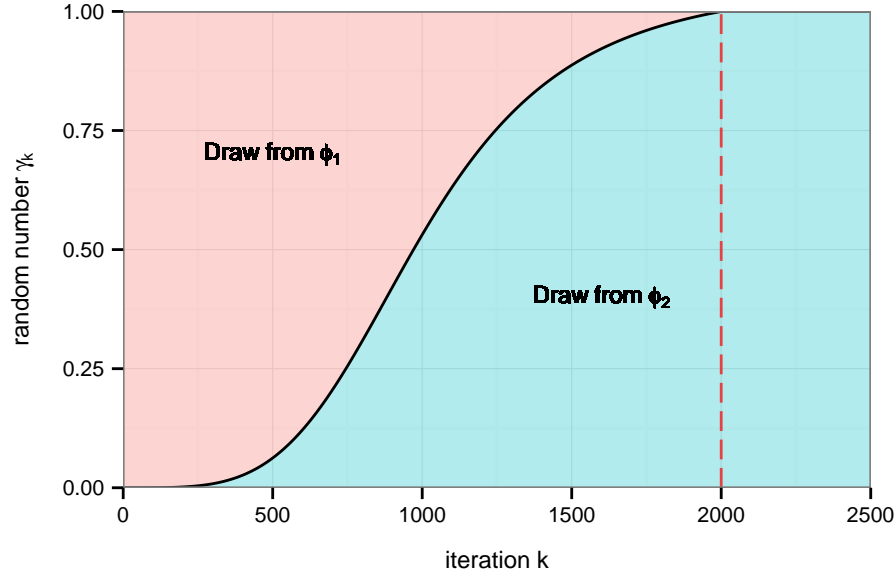


Figure 7.16: Illustration of the function $S(k)$ which determines the probabilities to choose between $\phi_1(\vec{p})$ and $\phi_2(\vec{p})$ in an iteration k .

and modify Equation 7.22 accordingly,

$$\mathbb{E}_\phi[f(\vec{p})] \approx \frac{\frac{1}{n} \sum_{i=1}^n \frac{f(\vec{p}_i) \pi(\vec{p}_i)}{\phi(\vec{p}_i)}}{\frac{1}{n} \sum_{i=1}^n \frac{\pi(\vec{p}_i)}{\phi(\vec{p}_i)}} = \frac{\sum_{i=1}^n f(\vec{p}_i) \omega_i}{\sum_{i=1}^n \omega_i}, \quad (7.25)$$

where $\omega_i = \pi(\vec{p}_i) / \phi(\vec{p}_i)$.

The objective of Monte Carlo methods relying on Bayesian statistics is the computation of summary statistics, such as the mean vector, of the posterior distribution $\pi(\vec{p} | \vec{\sigma}_{\text{exp}})$. The self-adaptive sampling scheme of the BMC method uses two different instrumental distributions $\phi_1(\vec{p})$, $\phi_2(\vec{p})$. In each iteration k a random number γ_k is drawn from a uniform distribution in the range $[0, 1]$. The value γ_k is then compared to the value of the function

$$S(k) = \begin{cases} \frac{17k^4}{16k^4 + N_{\text{burn}}^4} & \text{for } k < N_{\text{burn}} \\ 1 & \text{for } k \geq N_{\text{burn}} \end{cases}. \quad (7.26)$$

If $\gamma_k > S(k)$ holds, the parameter vector in the k^{th} iteration is drawn from $\phi_1(\vec{p})$, otherwise from $\phi_2(\vec{p})$. The function $S(k)$ is illustrated in Figure 7.16.

(Koning, 2015) choose a uniform distribution $\phi_1(\vec{p})$ for the model parameters. Its support should be large enough to remove as much as possible the influence of the prior on the result. In our evaluation scenario we assume

$$\phi_1(\vec{p}) := \begin{cases} \left(\frac{3}{10}\right)^3 & \text{for } \vec{p} \in [0.85, 1.15]^3 \\ 0 & \text{otherwise} \end{cases} \quad (7.27)$$

The specification of the instrumental distribution $\phi_2(\vec{p})$ in the k^{th} iteration depends on the parameter vectors $\vec{p}_1, \vec{p}_2, \dots, \vec{p}_{k-1}$ drawn so far, thus a better notation

is $\phi_2^k(\vec{p})$. Let the parameter vectors $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_m$ be the subset of the parameter vectors drawn before the k^{th} iteration that were sampled from the prior distribution $\phi_1(\vec{p})$. We denote the associated weights by $\tilde{\omega}_i = \pi(\vec{q}_i | \vec{\sigma}_{\text{exp}}) / \phi_1(\vec{q}_i)$. Now the instrumental distribution $\phi_2^k(\vec{p})$ can be defined as

$$\phi_2^k(\vec{p}) : \begin{cases} \frac{\tilde{\omega}_i}{\sum_{j=1}^m \tilde{\omega}_j} & \text{if a } \vec{q}_i \text{ exists so that } \vec{p} = \vec{q}_i \\ 0 & \text{otherwise} \end{cases} \quad (7.28)$$

In the paper of (Koning, 2015) not many details are given on the construction of $\phi_2^k(\vec{p})$. Defining $\phi_2^k(\vec{p})$ based only on the samples drawn from $\phi_1^k(\vec{p})$ seems reasonable to us, yet this choice might differ from the choice in the paper.

Thus, the BMC sampling scheme is completely specified and we can apply it to the neutron-induced total cross section of ^{181}Ta . As shown in section 6.1, the value of the likelihood depends on the number of involved channels C and the number of points N_m in each experimental data set m . For convenience, we repeat the functional form of the likelihood here,

$$\ell(\vec{\sigma}_{\text{exp}} | \vec{p}) \propto \exp \left[-\frac{1}{C} \sum_{c=1}^C \sum_{m=1}^{M_c} \frac{1}{N_m} \sum_{k=1}^{N_m} \left(\frac{\sigma_{\text{exp},cmk} - \mathcal{M}(\vec{p})}{\delta_{\text{exp},cmk}} \right)^2 \right]. \quad (7.29)$$

The subscript cmk identifies a cross section by its channel c , to which dataset m it belongs and the position k within this dataset. The experimental measurement values are denoted by $\sigma_{\text{exp},cmk}$ and the associated uncertainties by $\delta_{\text{exp},cmk}$. The function $\mathcal{M}(\vec{p})$ gives the appropriate model prediction for the experimental data point $\sigma_{\text{exp},cmk}$. This likelihood does not include correlations between the data points.

In our evaluation scenario we consider only one channel, hence $C = 1$. We included the complete dataset \mathcal{C} in the likelihood and assumed that all points belong to one dataset. According to Equation 6.5, this corresponds to a rescaling of the experimental covariance matrix \mathbf{B} by the number of data points (364). Because correlations between the data points are ignored, the matrix \mathbf{B} contains non-zero values only in the diagonal.

First, we generated 10^4 parameter vectors and used $N_{\text{burn}} = 10^4$. The sampling process is visualized in Figure 7.17. Inserting $\vec{p}_1 = f(\vec{p}) = \vec{p}$ and $f(\vec{p}) = (\vec{p} - \vec{p}_1)(\vec{p} - \vec{p}_1)^T$ into Equation 7.25, we obtained the evaluated mean vector and covariance matrix

$$\vec{p}_{\text{BMC}} = \begin{pmatrix} 1.0234 \\ 0.9833 \\ 1.0172 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{BMC}} = 10^{-4} \begin{pmatrix} 60.3391 & -13.1244 & 15.8424 \\ -13.1244 & 3.7115 & -5.0835 \\ 15.8424 & -5.0835 & 15.9559 \end{pmatrix}. \quad (7.30)$$

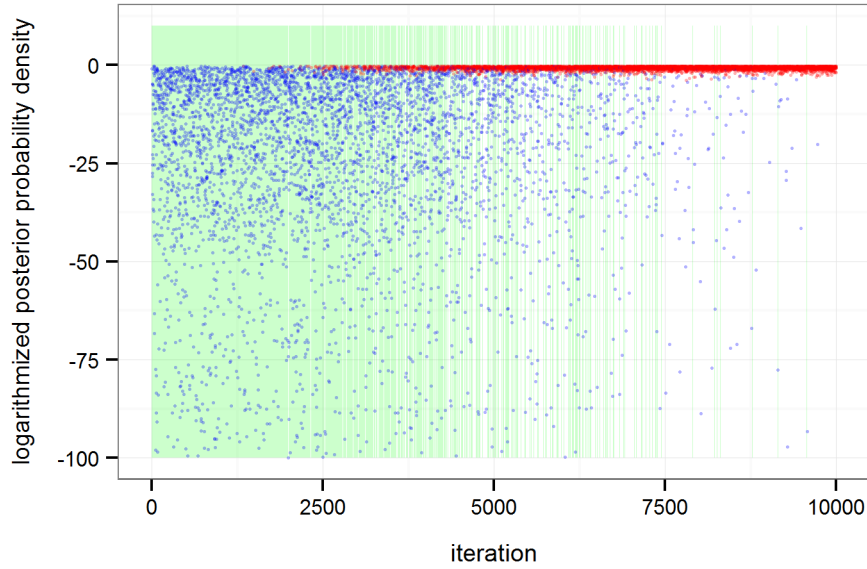


Figure 7.17: Bayesian Monte Carlo self-adapting scheme. Shown are the posterior probability densities (unnormalized and logarithmic) associated with the sampled parameter vectors. Points in blue were sampled from $\phi_1(\vec{p})$ and points in red from $\phi_2(\vec{p})$.

The evaluation at the level of cross sections is shown in Figure 7.18 and the associated uncertainties and correlations are illustrated in Figure 7.19. The evaluated uncertainties are larger than those obtained by the EMPIRE-Kalman, FBET and UMC-B method, which can be attributed to the alternative specification of the likelihood.

As a reference evaluation, we also applied the UMC-B method using the same choice of the likelihood. The evaluated mean vector and covariance matrix obtained by the UMC-B method based on 10^5 samples from the posterior distribution, and using a step length $\delta = 0.8$, are given by

$$\vec{p}_{\text{UMC-B}} = \begin{pmatrix} 1.0226 \\ 0.9839 \\ 1.0151 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{UMC-B}} = 10^{-4} \begin{pmatrix} 56.3889 & -11.5769 & 13.2041 \\ -11.5769 & 3.3926 & -4.7509 \\ 13.2041 & -4.7509 & 16.2219 \end{pmatrix}. \quad (7.31)$$

The mean vectors estimated with the BMC method and the UMC-B method are approximately equal. The occurring differences are well below the evaluated uncertainties of these parameters. Hence, the two evaluation methods are consistent with each other. At the level of cross sections, evaluated cross section curves and associated confidence bands are hardly distinguishable by eye.

Next, we investigate the change of evaluated quantities with an increasing number of sampled parameter vectors. After iteration $k = N_{\text{burn}}$, samples are drawn exclusively from the instrumental distribution $\phi_2^k(\vec{p})$. This means that new samples

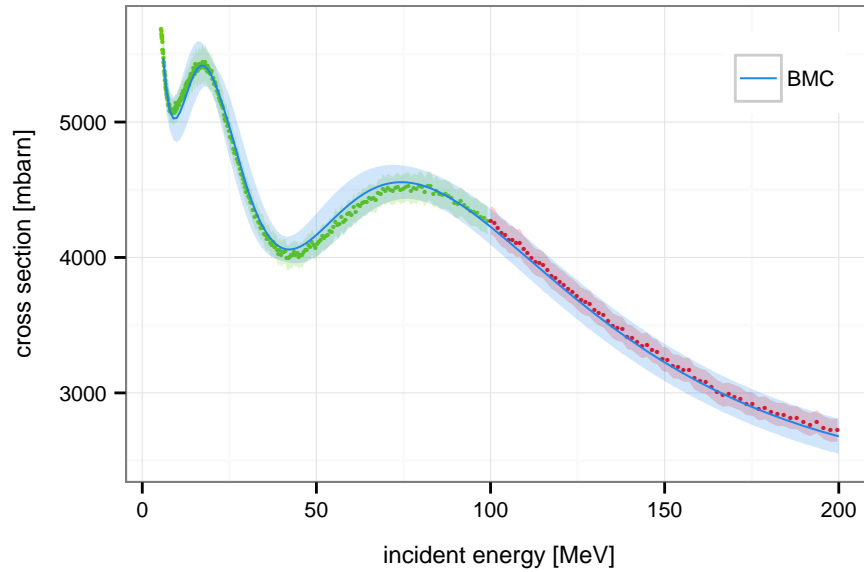


Figure 7.18: Evaluated cross section curve obtained by the BMC method when updating with the complete dataset \mathcal{C} . 10^4 parameter vectors were sampled and N_{burn} was set to the same value.

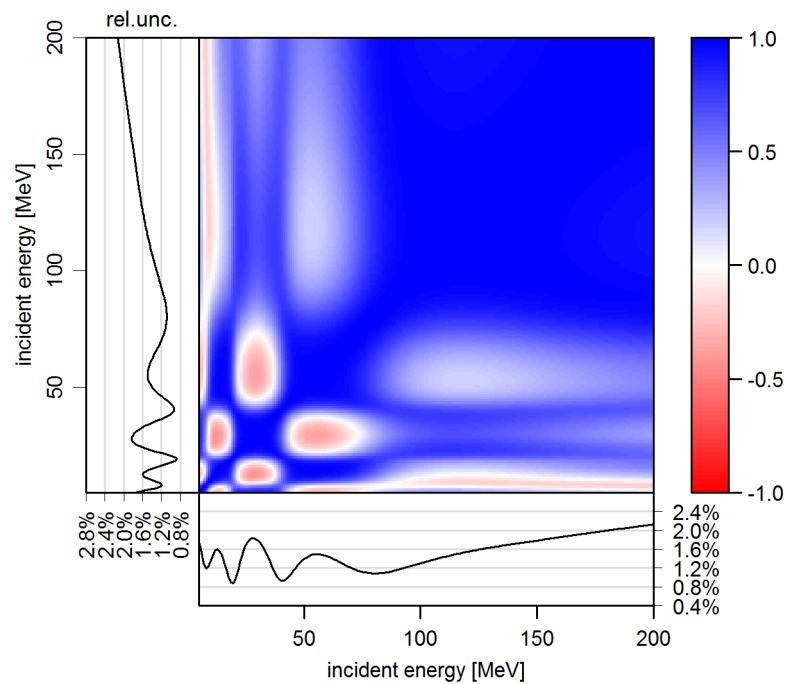


Figure 7.19: Evaluated uncertainties and correlations obtained by the BMC method when updating with the complete dataset \mathcal{C} . 10^4 parameter vectors were sampled and N_{burn} was set to the same value.

are generated by randomly picking from old samples. The probability to select a certain sample is proportional to its weight. We reused the samples which yielded the result in Equation 7.30 and generated on their basis using $\phi_2^k(\vec{p})$ another $9 \cdot 10^4$ samples. We obtained the estimates

$$\vec{p}_{\text{BMC}} = \begin{pmatrix} 1.0230 \\ 0.9834 \\ 1.0171 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{BMC}} = 10^{-4} \begin{pmatrix} 60.2108 & -13.0818 & 15.6801 \\ -13.0818 & 3.7009 & -5.0432 \\ 15.6801 & -5.0432 & 15.8447 \end{pmatrix}. \quad (7.32)$$

They slightly differ from the results in Equation 7.30 where we stopped the iteration after $N_{\text{burn}} = 10^4$ samples.

In the following we are interested in estimates of summary statistics $f(\vec{p})$ in the limit of infinitely many iterations. After the initial burn-in phase, samples are exclusively drawn from $\phi_2^k(\vec{p})$. Therefore $\phi_2^k(\vec{p}) = \phi_2^n$ holds for all $k, n > N_{\text{burn}}$ and we can drop the superscript k for the following consideration. We rewrite the probability density function of $\phi_2(\vec{p})$ given in Equation 7.28 as

$$\phi_2^k(\vec{p}) = C \sum_{i=1}^m \tilde{\omega}_i \delta(\vec{p} - \vec{q}_i), \quad (7.33)$$

where $\delta(\cdot)$ denotes the Dirac-delta functional and \vec{q}_i are the samples drawn from $\phi_1(\vec{p})$ before a number of N_{burn} iterations is reached. The factor C denotes the normalization constant and is given by $C^{-1} = \sum_{j=1}^m \tilde{\omega}_j$.

One possibility to compute a summary statistics $f(\vec{p})$ is to sample from $\phi_2(\vec{p})$ and then estimate its value using Equation 7.20. However, because we know $\phi_2(\vec{p})$ explicitly, we can also directly solve the integral in Equation 7.19,

$$\int f(\vec{p}) C \sum_{i=1}^m \tilde{\omega}_i \delta(\vec{p} - \vec{q}_i) d\vec{p} = C \sum_{i=1}^m f(\vec{p}_i) \tilde{\omega}_i. \quad (7.34)$$

This result shows that recycling samples does not offer an advantage, because the result in the limit of infinitely many iterations is analytically available. This result is also important with respect to the Total Monte Carlo (TMC) framework.

In the TMC framework, uncertainties in model parameters are propagated to uncertainties of observables, which are relevant for applications, such as the criticality k_{eff} of reactors. Dependencies between quantities are exactly accounted for—no perturbation theory or sensitivity analysis is applied. Model parameters are generated according to a probability distribution, which reflects the uncertainty in the value of the parameter. For each set of sampled model parameters, the whole application chain is executed. Albeit the computations performed in this application chain are complex, in terms of statistics it just represents the evaluation of a summary statistics $f(\vec{p})$. For this reason, in the remainder of this

section we always create exactly a number of N_{burn} samples and then calculate the mean vector and the covariance matrix analytically.

Next, we perform an evaluation with the BMC only including dataset \mathcal{A} in the likelihood. The estimates for the parameter mean vector and the covariance matrix are given by

$$\vec{p}_{\text{BMC}} = \begin{pmatrix} 1.0369098 \\ 0.9793851 \\ 1.0171299 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_{\text{BMC}} = 10^{-4} \begin{pmatrix} 59.9923 & -13.6613 & 13.3971 \\ -13.6613 & 4.1351 & -4.8254 \\ 13.3971 & -4.8254 & 14.2780 \end{pmatrix}. \quad (7.35)$$

The evaluated cross section curve is shown in Figure 7.20 and associated uncertainties and correlations are shown in Figure 7.21. The uncertainties at high incident energies are considerably larger than those obtained in updates with the complete dataset \mathcal{C} (see Equation 7.31). However, the evaluated parameter vectors are approximately equal. Further, the oscillation of the uncertainty at low incident energies clearly shows the influence of the model systematics, because the uncertainties of the experiment reflected in the covariance matrix \mathbf{B} are rather uniform. In contrast to the result of the UMC-B method in Equation 7.18, the evaluation is consistent with the experimental dataset \mathcal{B} which had not been included in the update process. This result can be attributed to the modified specification of the likelihood.

Multiplying the covariance matrix \mathbf{B} by the number of experimental data points (dataset \mathcal{A} contains 294 points) corresponds to the assumption of larger experimental uncertainties. Thus, the experimental data does not give enough evidence for a cross section curve that cannot be described by the model. Hence the model deficiency does not affect the result of the evaluation in an undesirable way.

However, rescaling the covariance matrix means to throw away information from the measurements. If a sound assessment of an experiment yields an uncertainty of 1% in the cross section values, there is no cogent reason to enlarge this uncertainty dependent on the number of data points. For instance, in Figure 7.20 the evaluated uncertainty at 50 MeV might be larger than necessary. Another point of criticism may be that also around 50 MeV the evaluated curve deviates noticeably from the experiment which was included in the evaluation. If the experimental data are reliable, it is a desirable feature of the evaluated cross section curve to mimic well the experimental data points.

As the last practical investigation, we want to use a multivariate normal likelihood and apply the self-adaptive Monte Carlo scheme of the BMC method. According to Equation 7.29, the BMC likelihood coincides with a multivariate normal

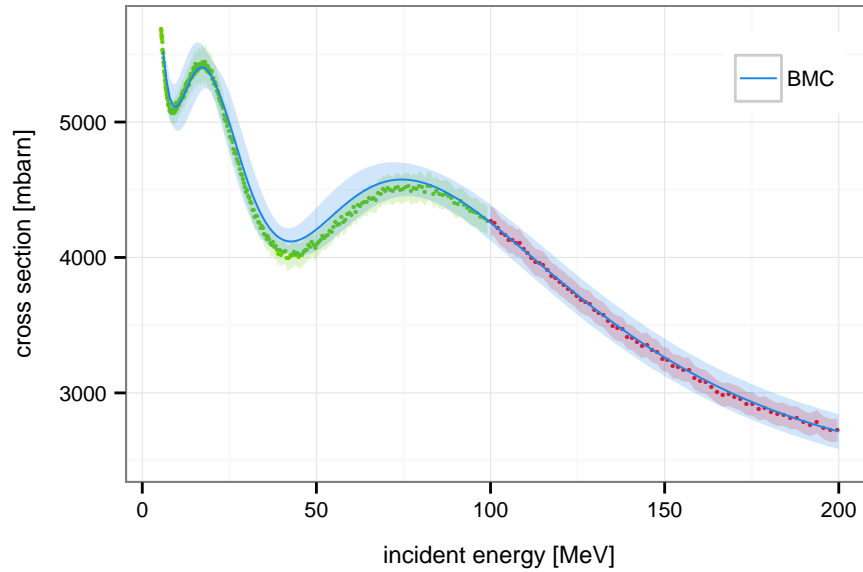


Figure 7.20: Evaluated cross section curve obtained by the BMC method when updating only with the dataset \mathcal{A} . 10^4 parameter vectors were sampled and N_{burn} was set to the same value.

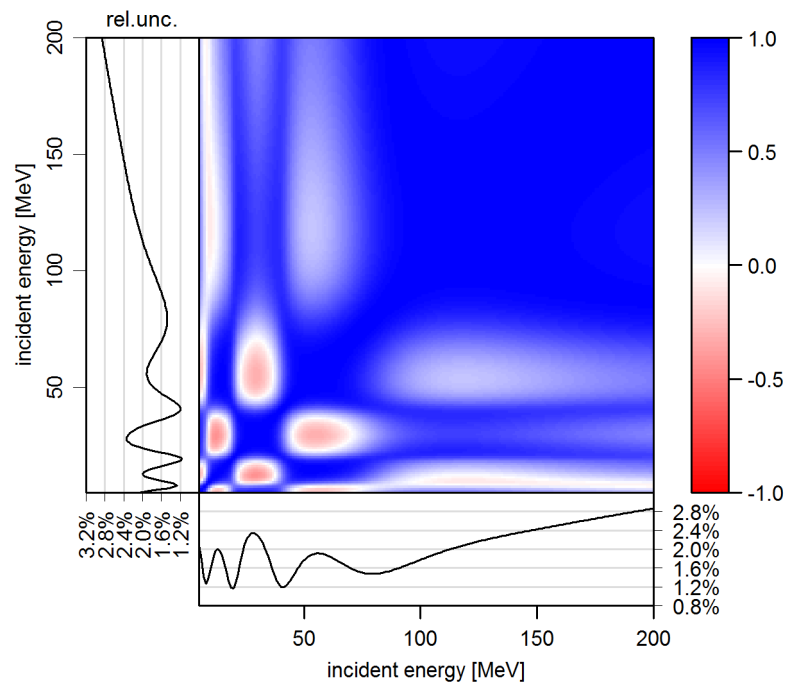


Figure 7.21: Evaluated uncertainties and correlations obtained by the BMC method when updating only with the dataset \mathcal{A} . 10^4 parameter vectors were sampled and N_{burn} was set to the same value.

distribution⁴ if only one channel is included in the update and each experimental data point is associated with a separate dataset. We used the full dataset \mathcal{C} and the form of the likelihood given in Equation 7.12 for the evaluation and performed 10^4 iterations with $N_{\text{burn}} = 10^4$. We found the same issue as for the UMC-G and UMC-B method relying on the originally proposed brute-force scheme (see section 7.1): The spread of the obtained weights $\tilde{\omega}_i$ was very large and only a few parameter vectors \vec{q}_k had significant weight. The smallest weight was about 10^{12244} times smaller than the largest weight. For the sake of illustration, we state the seven largest weights among the 10^4 weights we obtained:

$$1.00\text{e}+00 \quad 2.43\text{e}-01 \quad 9.94\text{e}-02 \quad 1.98\text{e}-03 \quad 8.42\text{e}-04 \quad 1.42\text{e}-04 \quad 3.37\text{e}-09$$

The weights are divided by the largest weight for comparability. Consequently, only a few weights among the 10^4 weights are non-negligible. In the extreme case it could happen that only one weight associated with some parameter vector \vec{q}_l is several orders of magnitude larger than all other weights. Then, the instrumental distribution $\phi_2^k(\vec{p})$ (see Equation 7.33) would be very accurately given by

$$\phi_2^k(\vec{p}) = C \sum_{i=1}^m \tilde{\omega}_i \delta(\vec{p} - \vec{q}_i) \approx \delta(\vec{p} - \vec{q}_l), \quad (7.36)$$

and consequently only the parameter vector \vec{q}_l is drawn from $\phi_2^k(\vec{p})$. In the present result only the parameter vectors associated with the six largest weights have a realistic chance to get recycled in 10^4 iterations. On the basis of just a few parameter vectors no reliable estimates of summary statistics are possible.

Finally, we want to discuss a theoretical issue that might affect estimates of summary statistics unfavorably. As discussed at the beginning of this section, the weights resulting from importance sampling are given by $\omega(\vec{p}) = \pi(\vec{p})/\phi(\vec{p})$ where $\pi(\vec{p})$ is the distribution of interest and $\phi(\vec{p})$ is the instrumental distribution from which the samples are drawn. The explanation of the self-adaptive importance sampling scheme of the BMC method ((Koning, 2015)) suggests that $\phi(x)$ for the calculation is either given by $\phi_1(\vec{p})$ or $\phi_2(\vec{p})$. However, the random process to select between these distributions itself has to be considered. Therefore, the proper instrumental distribution is given by

$$\phi^k(\vec{p}) = (1 - S(k))\phi_1(\vec{p}) + S(k)\phi_2^k(\vec{p}). \quad (7.37)$$

The probability to draw samples from $\phi_1(\vec{p})$ declines to zero as the number of iterations approaches N_{burn} . This means that the probability to draw a vector

⁴Strictly speaking, also the BMC likelihood is a multivariate normal distribution, but the covariance matrix is rescaled dependent on the number of channels involved and the number of measurement points in the datasets.

\vec{p} which was not already sampled declines to zero. In consequence, the weights for such parameter vectors diverge. After N_{burn} iterations, there is zero probability to explore new parameter vectors. This is in contradiction to an essential requirement of importance sampling. Every parameter vector \vec{p} associated with a non-zero probability density under $\pi(\vec{p})$ must have a chance to be sampled from the instrumental distribution $\phi(\vec{p})$. To see the reason for this requirement numerically, we may assign a small probability ε to draw from $\phi_1(\vec{p})$ after N_{burn} iterations. The instrumental distribution after N_{burn} steps then takes the form

$$\phi(\vec{p}) = \varepsilon\phi_1(\vec{p}) + (1 - \varepsilon)\phi_2(\vec{p}). \quad (7.38)$$

The expectation value of a summary statistics $f(\vec{p})$ under the distribution $\pi(\vec{p})$ is given in Equation 7.22. In the following we abbreviate $\mathcal{I} = \mathbb{E}_\pi[f(\vec{p})]$. An estimate I_n of \mathcal{I} based on a finite number n of samples is afflicted with a statistical error. If we use importance sampling and some instrumental distribution $\phi(\vec{p})$, the variance of the estimate is given by

$$\text{Var}_\phi[I_n] = \frac{1}{n} \text{Var}_\phi \left[\frac{f(\vec{p})\pi(\vec{p})}{\phi(\vec{p})} \right] = \frac{1}{n} \int_{\mathcal{S}} \left(\frac{f(\vec{p})\pi(\vec{p})}{\phi(\vec{p})} - \mathcal{I} \right)^2 \phi(\vec{p}) \, d\vec{p}. \quad (7.39)$$

We assume that $f(\vec{p})\pi(\vec{p})$ is bounded in \mathcal{S} and further that there exists a measurable subset of \mathcal{S} in which $f(\vec{p})\pi(\vec{p}) \neq 0$. Due to the definition in Equation 7.21, $f(\vec{p})\pi(\vec{p}) \neq 0$ implies $f(\vec{p})\pi(\vec{p}) > 0$ in some region of \mathcal{S} . In general, these assumptions hold if $f(\vec{p})$ is a reasonable summary statistics. Now we return to the specific instrumental distribution in Equation 7.38. Because $\phi_2(\vec{p})$ is defined only on a discrete subset of \mathcal{S} , there exists a subset \mathcal{A} such that $\phi_2(\vec{p}) = 0$ and $f(\vec{p})\pi(\vec{p}) > 0$ for all $\vec{p} \in \mathcal{A}$. Let \mathcal{B} be the complementary set to \mathcal{A} . We can split the integral in Equation 7.39 into the sum

$$\int_{\mathcal{A}} \left(\frac{f(\vec{p})\pi(\vec{p})}{\varepsilon\phi_1(\vec{p})} - \mathcal{I} \right)^2 \varepsilon\phi_1(\vec{p}) \, d\vec{p} + \int_{\mathcal{B}} \left(\frac{f(\vec{p})\pi(\vec{p})}{\phi(\vec{p})} - \mathcal{I} \right)^2 \phi(\vec{p}) \, d\vec{p} \quad (7.40)$$

Now we can expand the first integral as

$$\begin{aligned} \int_{\mathcal{A}} \left(\frac{(f(\vec{p})\pi(\vec{p}))^2}{\varepsilon\phi_1(\vec{p})} - 2f(\vec{p})\pi(\vec{p})\mathcal{I} + \varepsilon\mathcal{I}^2\phi_1(\vec{p}) \right) d\vec{p} = \\ \frac{1}{\varepsilon} \int_{\mathcal{A}} \frac{(f(\vec{p})\pi(\vec{p}))^2}{\phi_1(\vec{p})} d\vec{p} - \int_{\mathcal{A}} (2f(\vec{p})\pi(\vec{p})\mathcal{I} + \varepsilon\mathcal{I}^2\phi_1(\vec{p})) d\vec{p} \end{aligned} \quad (7.41)$$

Therefore, $\text{Var}_\phi[I_n] \rightarrow \infty$ when $\varepsilon \rightarrow 0$. The variance of the estimate I_n gets arbitrarily large for ε declining towards zero. This argument proves that assigning non-zero probabilities only to parameter vectors drawn in an earlier stage of the sampling process is illicit in importance sampling.

7.5 Analysis of the BFMC method

The Backward-Forward Monte Carlo method (Bauge and Dossantos-Uzarralde, 2011; Bauge, Hilaire, and Dossantos-Uzarralde, 2007) is not intentionally designed as Bayesian method, though its underlying assumptions can be interpreted within the framework of Bayesian statistics (see section 4.1).

In the BFMC method, parameter vectors \vec{p}_i are drawn from a uniform distribution. For each obtained parameter vector \vec{p}_i the χ^2 -value is computed by

$$\chi_i^2 = (\vec{\sigma}_{\text{exp}} - \mathcal{M}\vec{p}_i)^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}\vec{p}_i), \quad (7.42)$$

where the vector $\vec{\sigma}_{\text{exp}}$ contains the measurements and \mathbf{B} is the associated experimental covariance matrix. We denote the smallest χ^2 -value occurring in the set of drawn parameter vectors \vec{p}_i as χ_{min}^2 . As the next step, a weight for each parameter is calculated by

$$\omega_i = \exp\left(-\left(\frac{\chi_i^2}{\chi_{\text{min}}^2}\right)^2\right). \quad (7.43)$$

The occurrence of χ_{min}^2 in the exponential function distinguishes this method from all other methods. The analysis of the UMC-G method using the brute-force sampling scheme in section 7.1 showed that due to the sharply peaked likelihood no significant weights are generated if drawing from the uniform prior distribution. The same would hold for the BFMC method if the likelihood would not be adjusted by the smallest χ^2 -value.

We generated 10^5 parameter vectors by drawing from a uniform distribution in the interval $[0.85, 1.15]^3$. The complete dataset \mathcal{C} was used to calculate the χ^2 -values. The following table shows the decline of the weights as calculated in the BFMC method and those used in the UMC-B method, $\omega_{\text{UMC-B},i} = \exp(-0.5\chi_i^2)$. Starting from the parameter vector associated with the largest weight, every fifth weight is stated:

i	1	6	11	16	21	26	31
ω_{BFMC}	1.00e+00	9.69e-01	9.52e-01	9.48e-01	9.42e-01	9.31e-01	9.25e-01
$\omega_{\text{UMC-B}}$	1.00e+00	6.17e-02	1.43e-02	9.94e-03	5.52e-03	2.11e-03	1.20e-03

All displayed weights are divided by the largest weight. The UMC-B weights decline rapidly and only dozens of the 10^5 generated weights significantly contribute to the estimation of summary statistics. For the generated ensemble of parameter vectors, the largest weight was about 10^{12830} times larger than the smallest one. Concerning the BFMC weights, the ratio of the largest to the smallest weight was with about 10^{12376} comparable in size. However, the BFMC weights decline much slower than the UMC-B weights and hence many of them contribute in the

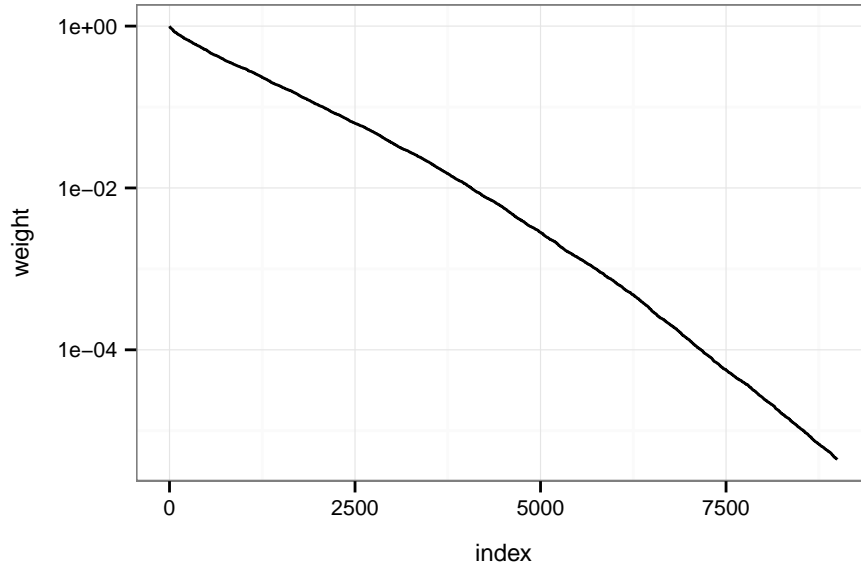


Figure 7.22: The decline of the weights as computed in the BFMC method. The largest 9000 weights out of 10^5 weights associated with the parameter vectors are shown in decreasing order.

estimation of summary statistics. Figure 7.22 illustrates the decline of the BFMC weights in our evaluations scenario when using the complete dataset \mathcal{C} . About 5000 from the 10^5 weights are significant in the estimation of summary statistics.

On the basis of the computed weights ω_i associated with the drawn parameter vectors \vec{p}_i , the BFMC method uses the following formulas to calculate the mean vector and the covariance matrix for the parameter vectors:

$$\vec{p}' = \frac{\sum_{i=1}^n \omega_i \vec{p}_i}{\sum_{i=1}^n \omega_i} \quad \text{and} \quad \mathbf{A}' = \frac{\sum_{i=1}^n \omega_i (\vec{p}_i - \vec{p}') (\vec{p}_i - \vec{p}')^T}{\sum_{i=1}^n \omega_i}. \quad (7.44)$$

We obtained on the basis of 10^5 parameter vectors and including the complete dataset \mathcal{C} the result

$$\vec{p}' = \begin{pmatrix} 1.0475 \\ 0.9772 \\ 1.0184 \end{pmatrix} \quad \text{and} \quad \mathbf{A}' = 10^{-4} \begin{pmatrix} 38.9246 & 5.8039 & -8.4423 \\ 5.8039 & 3.7445 & -4.9616 \\ -8.4423 & -4.9616 & 7.6248 \end{pmatrix}. \quad (7.45)$$

Several runs of the BFMC method, in each drawing 10^5 parameter vectors from a uniform distribution, indicated the stability of this solution. Differences of the values in the mean vectors were about 10^{-3} and in the elements of the covariance matrix about 10^{-5} .

Before we discuss the forward step of the BFMC procedure, we want to mention the following observation. Even though, the mean vector and the covariance matrix for the parameters were approximately the same in each run of the BFMC

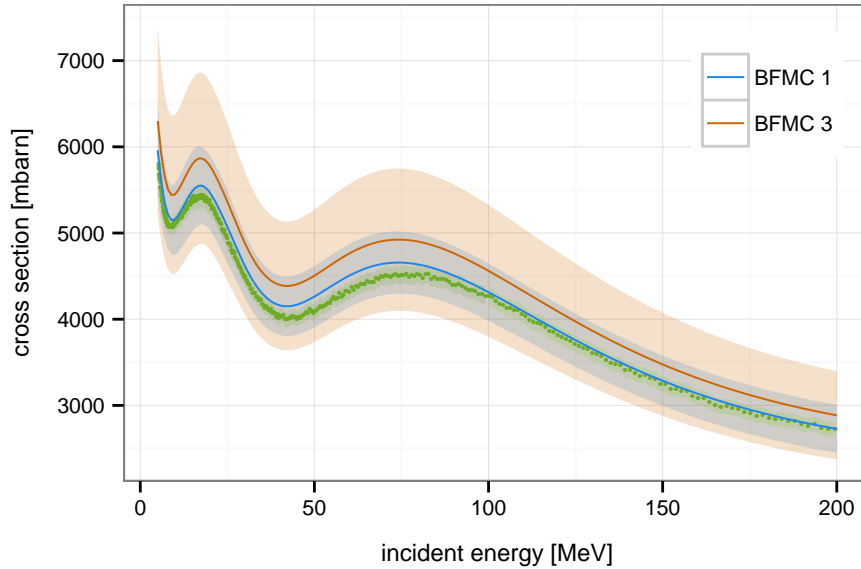


Figure 7.23: Differences between the cross section curves of different runs with the BFMC method. The cross section curve and the uncertainties were directly estimated using the weights from the backward step.

method, we obtained big differences when we estimated the cross section curves by

$$E[\sigma(E)] \approx \frac{\sum_{i=1}^n \omega_i \mathcal{M}(\vec{p}_i; E)}{\sum_{i=1}^n \omega_i}. \quad (7.46)$$

The results stemming from different runs with the BFMC method are shown in Figure 7.23. We will discuss this observation together with the cross section estimates resulting from the forward step, which is the last step of the procedure.

In the forward step, parameter vectors \vec{p}_i are drawn from the multivariate normal distribution defined by the mean vector and the covariance matrix given in Equation 7.45. Associated cross section vectors $\vec{\sigma}_i = \mathcal{M}(\vec{p}_i)$ are computed relying on the nuclear model \mathcal{M} . The mean vector and the covariance matrix for the cross sections are now estimated by

$$\vec{\sigma}_{\text{fwd}} = \frac{1}{n} \sum_{i=1}^n \vec{\sigma}_i \quad \text{and} \quad \mathbf{A}_{\text{fwd}} = \frac{1}{n} \sum_{i=1}^n (\vec{\sigma}_i - \vec{\sigma}_{\text{fwd}})(\vec{\sigma}_i - \vec{\sigma}_{\text{fwd}})^T. \quad (7.47)$$

Noteworthy, the big differences of the cross section curves between several runs when using the weights of the backward step do not occur in $\vec{\sigma}_{\text{fwd}}$ anymore. Figure 7.24 allows the comparison of the cross section curves resulting from the forward step of the same two runs which are shown in Figure 7.23.

In order to understand the significant difference when using the weights of the backward step and why these differences vanish after the forward step, we estimated the probability density in parameter space based on the parameter vectors

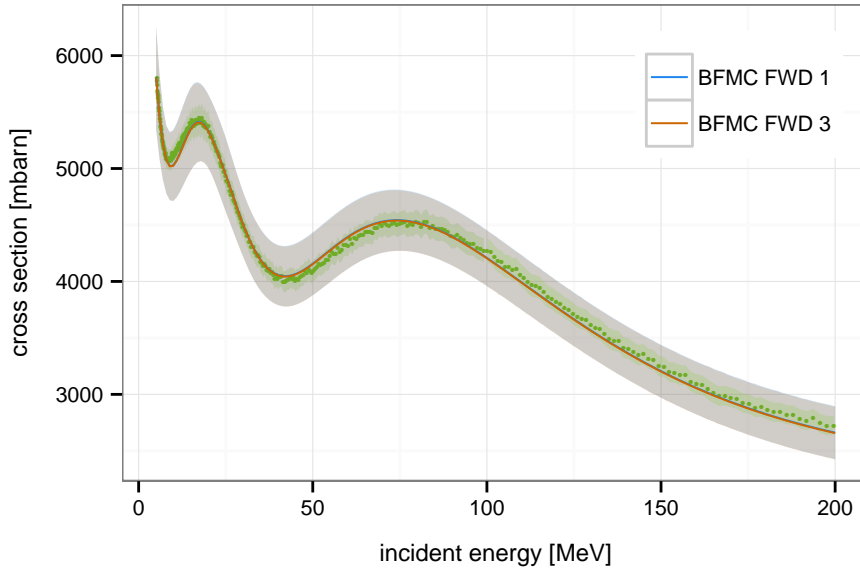


Figure 7.24: Cross section curves and 95% confidence band resulting from two runs of the BFMC method after the forward step. The same two runs lead to the cross section curves illustrated in Figure 7.23.

drawn in the backward step and their associated weights by kernel density estimation. Two projections of the probability density in parameter space onto the $\tilde{v}_1 - \tilde{a}_v$ plane are shown in Figure 7.25 and Figure 7.26. Due to the large amount of precise experimental data points, parameter vectors associated with large χ^2 -values are concentrated in a small region of parameter space. The consideration in section 7.3 (page 145) showed that the probability for a parameter vector drawn from the uniform prior distribution to fall into this region is negligible. The distance of drawn parameter vectors to this region is a matter of chance. Due to this fact associated χ^2 -values fluctuate significantly between different runs. This fluctuation is reflected in the plots of the marginal probability density distributions. Therefore, the values within the parameter vectors associated with largest weights of different runs can differ significantly, which explains the large differences of cross sections estimates based on the weights of the backward step occurring in different runs.

However, in the forward step this highly fluctuating estimate of the probability distribution in parameter space is replaced by a multivariate normal distribution. Fluctuations are not preserved by the multivariate normal distribution. Only the gross structure of the probability density distribution is preserved, which is approximately the same in each run. We arrive at the conclusion that the forward step in the BFMC method is a necessity to guarantee a certain reproducibility of evaluations.

To bring the discussion back to more practical matters, Figure 7.27 shows eval-

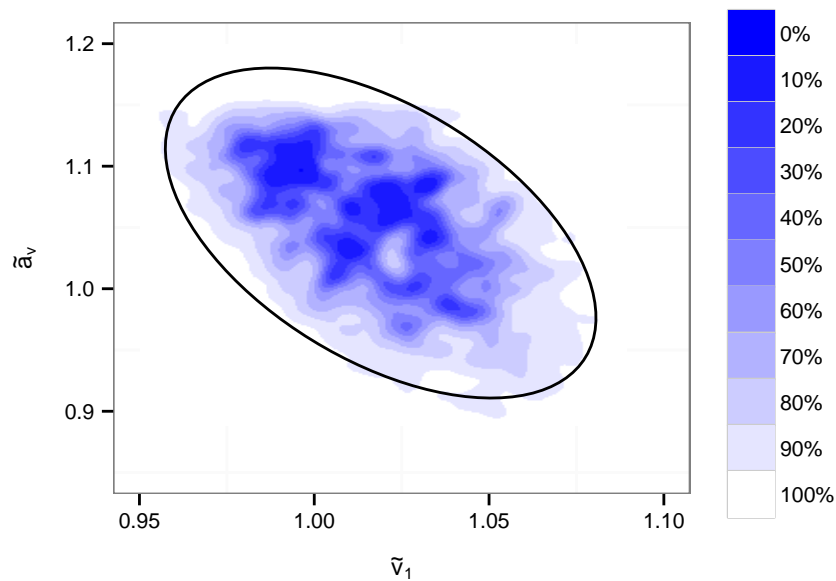


Figure 7.25: Marginal probability density distribution of \tilde{v}_1 and \tilde{a}_v estimated on the basis of the parameter vectors and their associated weights generated in the backward step of BFMC run 1. The ellipse indicates the 90% confidence interval of the multivariate normal distribution estimated from the drawn parameter vectors and their weights.

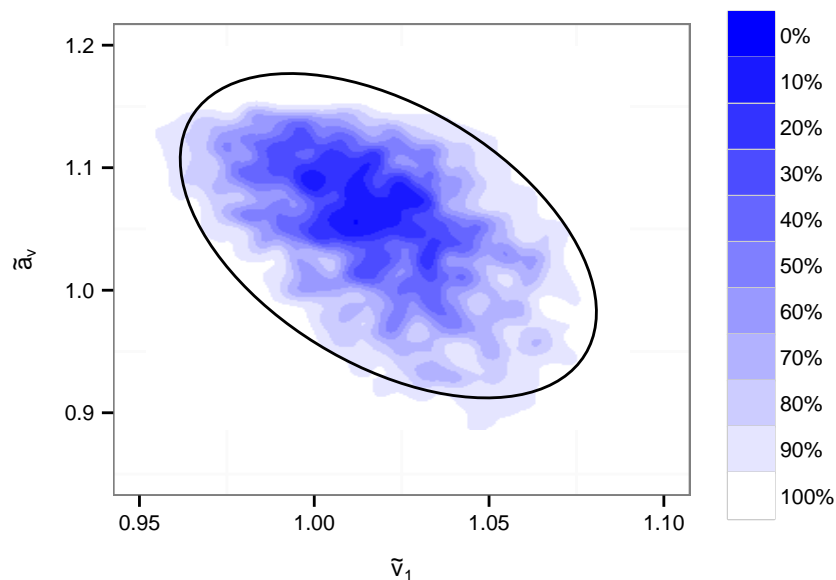


Figure 7.26: Marginal probability density distribution of \tilde{v}_1 and \tilde{a}_v estimated on the basis of parameter vectors and their associated weights generated in the backward step of BFMC run 3. The ellipse indicates the 90% confidence interval of the multivariate normal distribution estimated from the drawn parameter vectors and their weights.

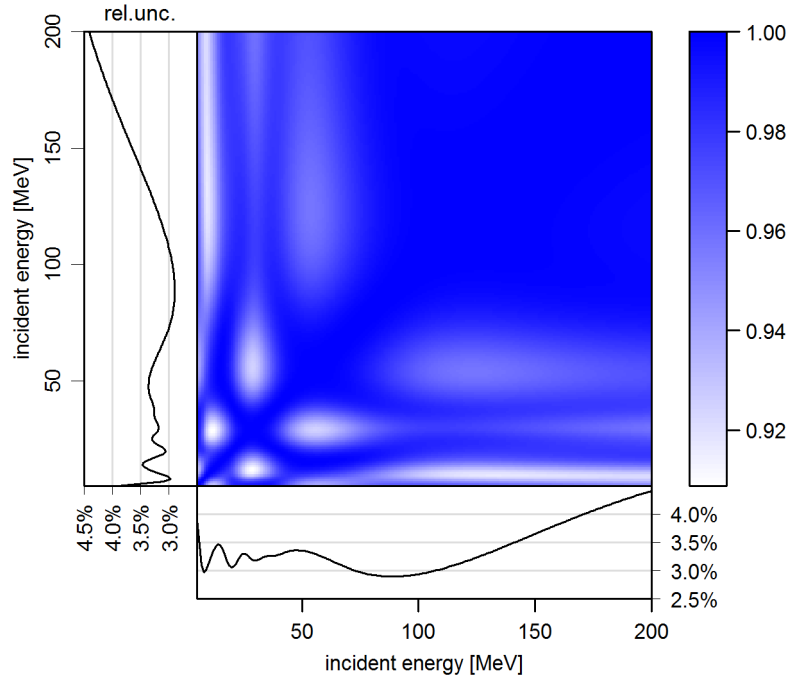


Figure 7.27: Evaluated uncertainties and correlations obtained by the BFMC method when including the complete dataset \mathcal{C} in the procedure.

uated uncertainties and correlations for ^{181}Ta obtained by the BFMC method including the complete dataset \mathcal{C} . The uncertainties ranging from slightly below 3% to 4.5% are the largest among all analyzed methods. This is due the fact that the BFMC method adjusts the likelihood on the basis of the the minimal χ^2 -value encountered in the ensemble of parameter vectors.

Despite the large uncertainties of the cross sections at different energies, they are strongly positively correlated. This result is opposed to that of the BMC method which produced lower uncertainties and weaker correlations (see Figure 7.19).

We generated cross section curves from the BFMC posterior and the BMC posterior to get an understanding of the different uncertainty-correlation structure of the methods. They are shown in Figure 7.28 and Figure 7.29. In the BFMC method, cross section curves with high probability density are distinct from each other by a global shift. In contrast to that, cross section curves with high posterior probability density in the BMC procedure often cross each other at lower incident energies and are only roughly parallel above 80 MeV. These differences can be attributed to the different choices of the likelihoods made in these methods in combination with the normalization uncertainty of the experimental data.

The BMC method uses a multivariate normal likelihood, but discards correlations between the experimental data points. Additionally, the uncertainties of the experimental data points are inflated by a factor which depends on the number

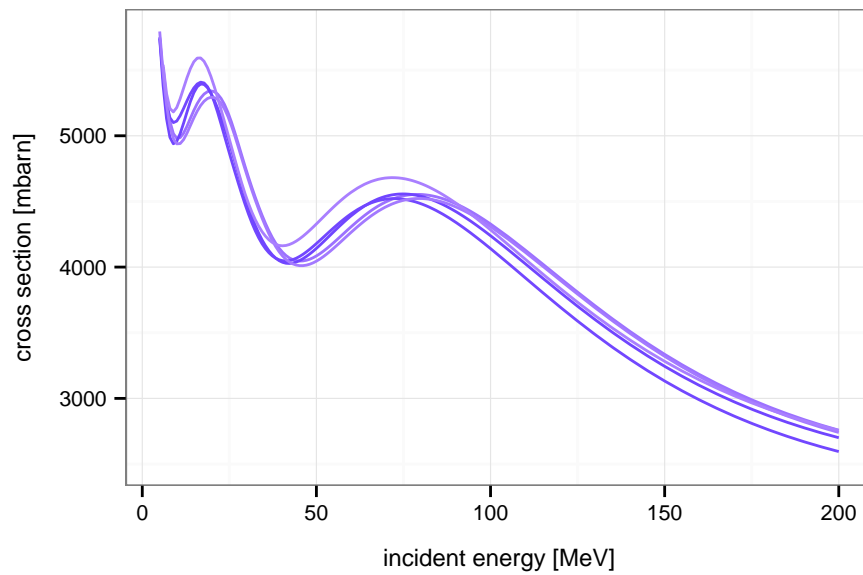


Figure 7.28: Samples drawn from the posterior distribution of the BMC method. The multivariate normal likelihood involved uses a modified version \mathbf{B}' of the experimental covariance matrix \mathbf{B} : the original matrix \mathbf{B} is multiplied by the number of experimental data points and off-diagonal elements are set to zero.

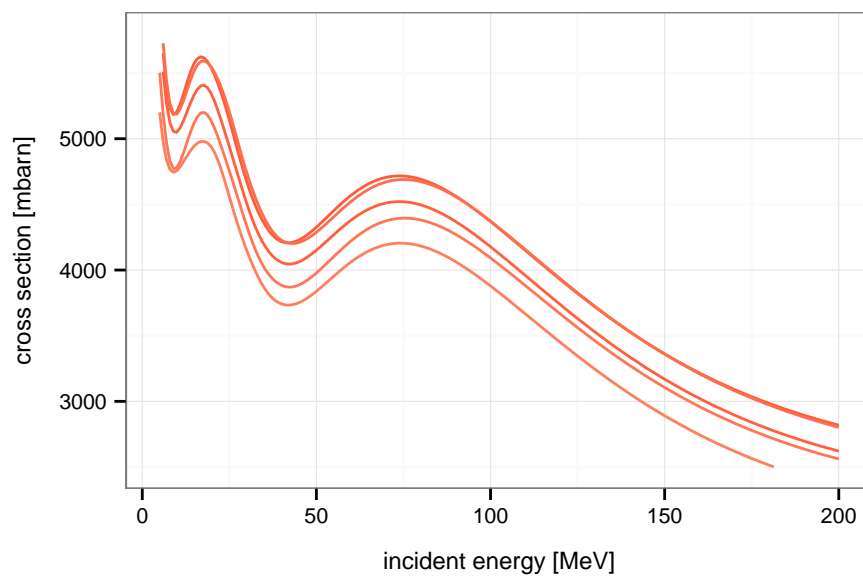


Figure 7.29: Samples drawn from the posterior distribution of the BFMC method. The original experimental covariance matrix \mathbf{B} enters the likelihood but the likelihood is flattened based on the minimal χ^2 -value.

of experimental data points. In contrast to that, the BFMC method retains the correlations between experimental data points, but does not use a multivariate normal distribution as likelihood. It uses a functional form for the likelihood that is flattened compared to the multivariate normal distribution dependent on the minimal χ^2 -value encountered for the ensemble of drawn parameter vectors.

Especially, the result of the BFMC evaluation demonstrates the importance of a consistent treatment of model defects in the evaluation method. The 95% confidence band in Figure 7.24 is consistent with the experimental data. However, the observation that the most likely cross section curves according to the BFMC posterior are different from each other merely by a global shift exclusively reflects the systematics of the model. Furthermore, the uncertainty of about 4% of the evaluated cross sections seems to be overestimated.

We complete the investigation of the BFMC method by performing an evaluation only with dataset \mathcal{A} . The evaluated parameter mean vector and the associated covariance matrix based on 10^5 parameter vectors are given by

$$\vec{p}' = \begin{pmatrix} 1.0432 \\ 0.9761 \\ 1.0200 \end{pmatrix} \quad \text{and} \quad \mathbf{A}' = 10^{-4} \begin{pmatrix} 41.0839 & 6.2775 & -9.0406 \\ 6.2775 & 3.8941 & -5.1925 \\ -9.0406 & -5.1925 & 7.9738 \end{pmatrix}. \quad (7.48)$$

The evaluated cross section curve is shown in Figure 7.30. Evaluated uncertainties and correlations are visualized in Figure 7.31. This result is approximately equal to the solution resulting from the inclusion of the complete dataset \mathcal{C} (see Equation 7.45).

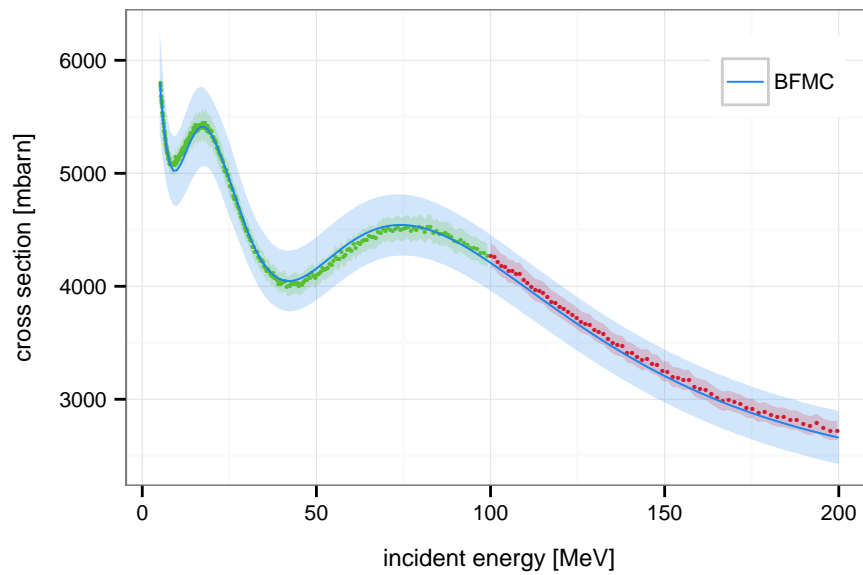


Figure 7.30: Evaluated cross section curve obtained by the BFMC method when only including dataset \mathcal{A} . The shown band indicates the 95% confidence level.

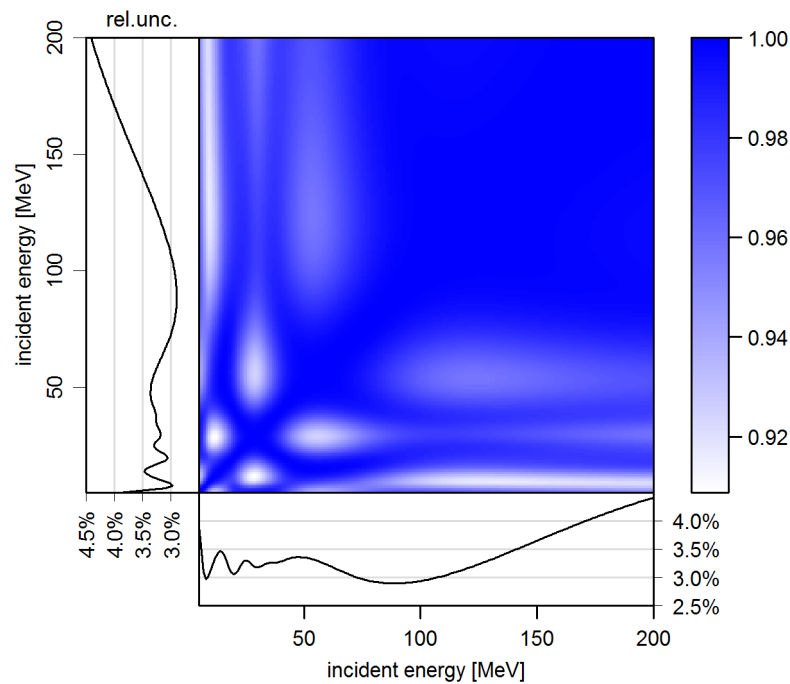


Figure 7.31: Evaluated uncertainties and correlations obtained by the BFMC method when only including dataset \mathcal{A} .

7.6 Summary and conclusions

In this chapter we studied the Monte Carlo methods UMC-G, UMC-B, BMC and BFMC in an evaluation of the neutron-induced total cross section of ^{181}Ta . The former two methods use a multivariate normal distribution for the likelihood whereas the latter two make other choices. All of these methods except the UMC-G approach use the exact nuclear model without any approximation for the Bayesian inference.

The UMC-G method is based on the surrogate approach outlined in section 3.3. Hence, in principle the analysis of the FBET/EMPIRE-MC method in section 6.4 also applies for the UMC-G method. However, contrary to the FBET the UMC-G method obtains results by means of Monte Carlo sampling. We studied the brute-force sampling scheme and the Metropolis-Hastings algorithm proposed by Capote and Smith (2008). Due to several reasons the proposed sampling schemes were not applicable in our evaluation scenario with plenty of experimental data available. The primary reason was the neglect of correlations between cross sections which led to the generation of cross section vectors with almost vanishing weights.

Therefore, we suggested a new sampling scheme based on the Metropolis-Hastings algorithm with a refined proposal distribution. The idea is to construct the proposal distribution on the basis of the posterior covariance matrix obtained by a Bayesian update using a simplified model. The refined proposal distribution resolved the issues of the original sampling schemes suggested by Capote and Smith (2008).

We applied the refined Metropolis-Hastings algorithm to perform the evaluation with the UMC-B method. The inclusion of the complete dataset \mathcal{C} yielded an evaluated cross section curve which followed well the trend of the experimental data. However, only considering the dataset \mathcal{A} containing data points below 100 MeV led to an evaluation which was inconsistent with the dataset \mathcal{B} . The results were almost equal to those obtained by the iterative procedure analyzed in section 6.3. Only in confidence regions in parameter space associated with a large trust level, the non-linearity of the nuclear model became apparent in the form of a local maximum of the probability density. The possibility to discover such local maximums represents a clear benefit of Monte Carlo methods over methods that linearize the model.

The BMC method relies on a modified likelihood which explicitly depends on the number of datasets and data points therein. As discussed in the previous section, this dependence is from a conceptual point of view not satisfactory. Under the assumption that each experimental data point belongs to the same dataset, the BMC likelihood can be interpreted as a multivariate normal distribution where

the experimental covariance matrix is scaled by the number of experimental data points. Using this choice of likelihood, the evaluation with the BMC method was consistent with the experimental data. However, two concerns were raised with respect to the suggested adaptive sampling strategy. First, the recycling of samples does not offer an advantage, because the asymptotic result of this sampling process is analytically available; and second, exclusively recycling samples drawn in an earlier stage of the sampling process violates a basic requirement of importance sampling: If a point is associated with non-vanishing probability density under the target distribution, this point must also be associated with non-vanishing probability density under the instrumental distribution.

Finally, the evaluation obtained by the BFMC method was also consistent with experimental data, independent from whether the complete dataset or only a subset was included in the update. However, the evaluated uncertainties seemed to be exaggerated with regard to the strong evidence given by the experimental data. Evaluated uncertainties were about three times larger at each point than the respective experimental uncertainties. Furthermore, we recognized that the forward step is essential to guarantee the reproducibility of results among several runs.

To summarize, from the three Monte Carlo methods UMC-B, BMC and BFMC, which use the exact nuclear model, only the latter two produced an evaluation consistent with experimental data. Also the latter two rely on a likelihood different from a multivariate normal distribution, which might advocate the modification of the likelihood as a proper measure to deal with model defects. However, the evaluated uncertainties, especially in the case of the BFMC method, did not conform to the experimental uncertainties. Moreover, accounting for the nuclear model without any approximation restricts cross section curves associated with non-zero probability density under the posterior distribution to the possibilities of the model. Therefore, the modification of the likelihood cannot be regarded as a satisfactory solution to account for model defects.

Existing approaches

The analysis of the previous two chapters showed the necessity of the consistent treatment of model deficiencies in an evaluation method. Otherwise, evaluations might be inconsistent with reliable experimental data, even if the employed nuclear model gives accurate predictions.

Few approaches exist to deal with model defects. Monte Carlo methods such as the BMC and BFMC modify the functional form of the likelihood. The adjustment of the likelihood means to modify the assumptions about the uncertainties of the experimental data. This measure implies the assumption that the model is perfect and its disagreement with experimental data is due to an inadequate uncertainty assessment of the experimental data. However, it is generally known that nuclear models can only describe some facets of the nuclear many-body problem due to the required approximations in nuclear structure and reaction calculations. Thus the concept of a perfect model is not a realistic option.

In this section we revisit the existing approaches to deal with model defects. First, we discuss χ^2 -rescaling. Conclusions will be representative for all methods that adjust the likelihood to deal with model defects. Afterward, we study the Symmetric Monte Carlo procedure (Rochman, Koning, Bauge, et al., 2014), which accounts for model defects in a pragmatic way. Finally, we discuss the use of an Ornstein-Uhlenbeck-process¹ (Pigni and Leeb, 2003) and the scaling procedure (Leeb, Neudecker, and Srdinko, 2008), which are well embedded in the framework of Bayesian statistics.

¹No name was originally given to this approach.

8.1 chi-square rescaling

The χ^2 -value is a measure how well a model prediction fits the experimental data. For a satisfactory description the quotient χ^2/N with N being the number of experimental data points (degrees of freedom) should be about one. If the χ^2 -value is significantly larger than N , either the model cannot describe the data or the uncertainties of the experimental data are underestimated. In such a case, the uncertainty band of the evaluated cross section curve is inconsistent with the experimental data. A common approach to address this issue is to rescale the evaluated covariance matrix according to the χ^2 -value. This can only be justified if the uncertainties of the experimental data are underestimated. The rescaling of the evaluated covariance matrix corresponds to a rescaling of the experimental covariance matrix.

In the following we check whether the rescaling of the covariance matrix is a proper procedure to account for model defects. For the sake of simplicity, we assume a linear model $\mathcal{M}(\vec{p}) = \mathbf{S}\vec{p}$. In section 3.2, we derived the formula to update the covariance matrix \mathbf{A}_0 of the model parameters based on the experimental covariance matrix \mathbf{B} ,

$$\mathbf{A}_1 = (\mathbf{A}_0^{-1} + \mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1}. \quad (8.1)$$

In general, we can expect \mathbf{A}_0 and $\mathbf{S}^T \mathbf{B} \mathbf{S}$ to be positive definite. We can use the matrix identity

$$(\mathbf{U} + \mathbf{V})^{-1} = \mathbf{U}^{-1} - \mathbf{U}^{-1} \mathbf{V} (\mathbf{U} + \mathbf{V})^{-1} \quad (8.2)$$

to construct the expansion

$$(\mathbf{U} + \mathbf{V})^{-1} = \mathbf{U}^{-1} - \mathbf{U}^{-1} \mathbf{V} \mathbf{U}^{-1} + \mathbf{U}^{-1} \mathbf{V} \mathbf{U}^{-1} \mathbf{V} - \dots \quad (8.3)$$

Expanding Equation 8.1 in this way yields

$$(\mathbf{A}_0^{-1} + \mathbf{S}^T \mathbf{B} \mathbf{S})^{-1} = (\mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1} - (\mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1} \mathbf{A}_0^{-1} (\mathbf{S}^T \mathbf{B}^{-1} \mathbf{S})^{-1} + \dots \quad (8.4)$$

The covariance matrix \mathbf{A}_0 can be decomposed into $\mathbf{P} \mathbf{D} \mathbf{P}^T$, where \mathbf{P} is an orthogonal matrix, hence $\mathbf{P}^T = \mathbf{P}^{-1}$, and \mathbf{D} is a diagonal matrix which contains the eigenvalues λ_i of \mathbf{A}_0 . Further we can also decompose \mathbf{B} into $\mathbf{Q} \mathbf{E} \mathbf{Q}^T$. The inversions of the decomposed matrices are given by $(\mathbf{A}_0)^{-1} = \mathbf{P} \mathbf{D}^{-1} \mathbf{P}^T$ and $\mathbf{B}^{-1} = \mathbf{Q} \mathbf{E}^{-1} \mathbf{Q}^T$. The inverse of the diagonal matrices are simply $\mathbf{D}^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \dots)$ and analogously $\mathbf{E}^{-1} = \text{diag}(1/\tau_1, 1/\tau_2, \dots)$. The eigenvalues represent the uncertainties along the principal axes of the covariance matrix. If the uncertainties reflected in \mathbf{B} are much smaller than those in \mathbf{A}_0 , the second and higher order terms in Equation 8.4 can be neglected. Consequently, the updated parameter covariance matrix is in good approximation determined by the experimental covariance matrix. Rescaling the evaluated covariance matrix by factor λ can then

be interpreted as rescaling the experimental covariance matrix that entered the update process by the same factor,

$$\lambda \mathbf{A}_1 \approx (\mathbf{S}^T (\lambda \mathbf{B})^{-1} \mathbf{S})^{-1}. \quad (8.5)$$

In the following, we motivate when χ^2 rescaling is a proper measure and afterward explain on the basis of Equation 8.5 why it is not a suitable measure to address model defects.

Generally, measurements are afflicted with statistical and systematic errors. In the following we assume a vector of measurements $\vec{\sigma}_{\text{exp}}$ to be the sum of the true cross section vector $\vec{\sigma}_{\text{true}}$ and the measurement error $\vec{\varepsilon}$,

$$\vec{\sigma}_{\text{exp}} = \vec{\sigma}_{\text{true}} + \vec{\varepsilon}. \quad (8.6)$$

The specifics of the experiment determine the probability distribution of the measurement error. Often, a multivariate normal distribution is the proper choice, $\vec{\varepsilon} \sim \mathcal{N}(\vec{0}, \mathbf{B})$. If we have an accurate estimate of the true cross section vector and the probability distribution of the measurement error reflects the uncertainties of the experiment, it follows that the quantity

$$X = (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_{\text{true}})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_{\text{true}}) \quad (8.7)$$

corresponds to a χ_k^2 -distribution with k degrees of freedom, where k is the number of experimental data points. Mean value and variance of the χ_k^2 -distribution are given by

$$\text{E}[X] = k \quad \text{and} \quad \text{Var}[X] = 2k. \quad (8.8)$$

Therefore, X/k is in the order of one if the required conditions are satisfied. If X/k is significantly larger than one, then either the experimental covariance matrix is misspecified or the estimate of the true cross section vector $\vec{\sigma}_{\text{true}}$ is inappropriate. If the former is the case, we can enforce a χ^2 -value of k by multiplying the experimental covariance matrix with X/k ,

$$\tilde{X} = (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_{\text{true}})^T \left(\frac{X}{k} \mathbf{B} \right)^{-1} (\vec{\sigma}_{\text{exp}} - \vec{\sigma}_{\text{true}}) = k. \quad (8.9)$$

In order to understand problems arising with the χ^2 -dependent rescaling of the experimental covariance matrix, we have to consider the estimation of $\vec{\sigma}_{\text{true}}$. In most available methods, the fundamental assumption is that the true cross section vector can be predicted by a nuclear model if only the model parameters are adjusted properly, $\vec{\sigma}_{\text{true}} \equiv \vec{\sigma}_{\text{mod}} \equiv \mathcal{M}(\vec{p}_{\text{true}})$. In section 6.1 we discussed at the example of a simple linear model that the evaluated covariance matrix reflects the systematics of the employed nuclear model. Rescaling the evaluated covariance matrix

does not change this feature. Therefore, the best posterior estimate will be still in disagreement with experimental data and variations allowed by the evaluated covariance matrix are still restricted to cross section curves which can be predicted by the model and do not resemble the curve indicated by the experimental data.

8.2 Symmetric Monte Carlo

The Symmetric Monte Carlo (SMC) method of Rochman, Koning, Bauge, et al. (2014) is a pragmatic solution to merge the information from model calculations and experimental data. Evaluated covariance matrices of Monte Carlo methods are in general associated with high correlations due to the dependence on relatively few model parameters (compared to the number of observables). This feature is often regarded as undesirable, which expresses a certain distrust in the nuclear model. If we have no doubts about the reliability of the model, we have to accept any degree of evaluated correlation. The SMC method is constructed in a way that evaluated correlations are weakened and evaluated cross sections are closer to the experimental data points, even if the nuclear model is not able to predict them.

The basic idea of the SMC method is to generate not only a sample of model predictions on the basis of a nuclear model with varied model parameters but also a sample of experimental data points. The sampling of model predictions can be done for instance with the BFMC method (see section 4.1).

For the outline of the idea, it suffices to restrict the discussion to the cross sections of different channels at the same incident energy. Let us denote with $\sigma_i^M(a)$ the model prediction made with parameter set \vec{p}_i for the cross section of channel a . The quantity $\sigma^E(c)$ denotes the experimental measurement of the cross section of channel c . All cross sections are at the same incident energy. The SMC method is capable to deal with a large amount of experimental data points and associated correlations. For a basic understanding of the method, one experimental data point is sufficient.

The value of the experimental measurement $\sigma^E(c)$ and the associated uncertainty $\delta^E(c)$ define a normal distribution for the true cross section value, $\sigma_{\text{true}} \sim \mathcal{N}(\sigma^E(c), \delta^E(c))$. In the SMC method, cross section values $\sigma_i^E(c)$ are drawn from this normal distribution. Then, each sampled model prediction $\sigma_i^M(c)$ is combined with one $\sigma_i^E(c)$ to form a combined cross section

$$\sigma_i^{ME}(c) = \omega_i \sigma_i^M(c) + (1 - \omega_i) \sigma_i^E(c). \quad (8.10)$$

Several choices for the weights ω_i are discussed by (Rochman, Koning, Bauge, et al., 2014). For instance, one possible choice is the weight $\omega = \sigma_i^E(c)/\delta^E(c)$

where $\delta^E(c)$ is the uncertainty of the experimental data point. The higher the uncertainty of the experimental data point, the closer the combined cross section is to the model prediction. The combined cross sections can be used in the TMC procedure for the determination of summary statistics of integral observables.

The sampling of experimental data points can be circumvented if we are only interested in a consistent ENDF file with best estimates and associated covariance matrices. For this restricted goal, it suffices to calculate the expression

$$\begin{aligned} \mathbb{E}[\sigma_i^{ME}(c)] &= \mathbb{E}[\omega_i \sigma_i^M(c)] + \mathbb{E}[(1 - \omega_i) \sigma_i^E(c)] = \\ &= \frac{1}{\delta^E(c)} \mathbb{E}[\sigma_i^E(c) \sigma_i^M(c)] + \mathbb{E}[\sigma_i^E(c)] - \frac{1}{\delta^E(c)} \mathbb{E}[\sigma_i^E(c)^2] = \\ &= \frac{1}{\delta^E(c)} \mathbb{E}[\sigma_i^M(c)] \mathbb{E}[\sigma_i^E(c)] + \mathbb{E}[\sigma_i^E(c)] - \frac{1}{\delta^E(c)} \mathbb{E}[\sigma_i^E(c)^2]. \end{aligned} \quad (8.11)$$

Here, we used the linearity of the expectation operator (see Equation 2.27) and the independence of the random variables $\sigma_i^M(c)$ and $\sigma_i^E(c)$. Thus, we were able to substitute $\mathbb{E}[\sigma_i^E(c) \sigma_i^M(c)]$ by $\mathbb{E}[\sigma_i^M(c)] \mathbb{E}[\sigma_i^E(c)]$. The expectations $\mathbb{E}[\sigma_i^E(c)]$ and $\mathbb{E}[\sigma_i^E(c)^2]$ of a normally distributed random variable are known in closed form,

$$\mathbb{E}[\sigma_i^E(c)] = \sigma^E(c) \quad \text{and} \quad \mathbb{E}[\sigma_i^E(c)^2] = (\sigma^E(c))^2 + (\delta^E(c))^2, \quad (8.12)$$

while the expectation of $\sigma_i^M(c)$ has to be estimated from the sample of model predictions,

$$\mathbb{E}[\sigma_i^M(c)] \approx \frac{1}{n} \sum_{i=1}^n \sigma_i^M(c) = \hat{\sigma}^M(c). \quad (8.13)$$

Putting everything together, we obtain the expression for the best estimate

$$\mathbb{E}[\sigma_i^{ME}(c)] \approx \hat{\sigma}^M(c) \frac{\sigma^E(c)}{\delta^E(c)} + \sigma^E(c) - \frac{(\sigma^E(c))^2 + (\delta^E(c))^2}{\delta^E(c)} = \hat{\sigma}^{ME}(c) \quad (8.14)$$

Similar calculations lead to the variance. Sampling in the suggested form becomes only necessary, if the sampling distribution of the experimental data point would be conditionally dependent on the value of the model prediction. Analogous formulas are available for the multivariate case.

Evaluations with the SMC method are not restricted to the representation possibilities of the nuclear model which is a clear benefit of the method. However, to allow deviations from the predictions of the model might lead to inconsistencies such as the violation of sum rules. Hence, for each combined cross section $\sigma_i^{ME}(c)$ the excess

$$\alpha_i(c) = \sigma_i^M(c) - \sigma_i^{ME}(c) \quad (8.15)$$

has to be distributed to other channels in order to preserve sum rules. In order to sketch the formalism of redistribution used in the SMC method, we introduce

the following quantities. The estimate of the expectation $\hat{\sigma}^M(a)$ for an arbitrary channel a can be calculated analogously to Equation 8.13. An estimate of the uncertainty $\delta^M(a)$ of a cross section $\sigma^M(a)$ stemming from the model is given by

$$\delta^M(a) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\sigma_i^M(a) - \hat{\sigma}^M(a))^2}. \quad (8.16)$$

Further, the correlation between two cross sections of different channels can be estimated from the model predictions by

$$\rho(a, b) = \frac{1}{(n-1)} \sum_{i=1}^n \frac{(\sigma_i^M(a) - \hat{\sigma}^M(a))(\sigma_i^M(b) - \hat{\sigma}^M(b))}{\delta^M(a)\delta^M(b)}. \quad (8.17)$$

Using the introduced quantities, the formula for the redistribution of the excess $\alpha_i(c)$ in channel c to another channel b is given by

$$\sigma_i^{ME}(b) = \frac{\rho(c, b) (\sigma_i^M(b))^{1.5}}{\sum_{j \neq c} |\rho(c, j)| (\sigma_i^M(j))^{1.5}} \times \alpha_i(c) + \sigma_i^M(b). \quad (8.18)$$

Although this expression looks reasonable, it does not allow a clear statement concerning the conservation of sum rules. Therefore, we tested the formula for a schematic case. We consider the total $\sigma(\text{tot})$, elastic $\sigma(\text{el})$ and non-elastic $\sigma(\text{nonel})$ cross section which can be combined to a vector $\vec{\sigma} = (\sigma(\text{tot}), \sigma(\text{el}), \sigma(\text{nonel}))^T$. We assume that the estimate of the covariance matrix \mathbf{A} and the associated correlation matrix Φ from the model predictions is given by

$$\mathbf{A} = \begin{pmatrix} 50 & 30 & 20 \\ 30 & 20 & 10 \\ 20 & 10 & 10 \end{pmatrix} \quad \text{and} \quad \Phi \approx \begin{pmatrix} 1.0000 & 0.9487 & 0.8944 \\ 0.9487 & 1.0000 & 0.7071 \\ 0.8944 & 0.7071 & 1.0000 \end{pmatrix}. \quad (8.19)$$

From the correlation matrix we see that $\rho(\text{el}, \text{nonel}) \approx 0.7071$, $\rho(\text{el}, \text{tot}) \approx 0.9487$ and $\rho(\text{nonel}, \text{tot}) \approx 0.8944$. The covariance matrix has not full rank which is due to the constraint $\sigma(\text{el}) + \sigma(\text{nonel}) = \sigma(\text{tot})$. We further assume that we sampled the model prediction vector

$$\vec{\sigma}_1^M = \begin{pmatrix} 100 \\ 70 \\ 30 \end{pmatrix}. \quad (8.20)$$

Finally, we assume that due to a measurement of the elastic cross section and some sampled value from the respective distribution, we obtained the combined cross section $\sigma_1^{ME}(\text{el}) = 50$, which yields the excess $\alpha_1(\text{el}) = 70 - 50 = 20$. Inserting all numbers into Equation 8.18 gives

$$\sigma_1^{ME}(\text{tot}) = \frac{0.9487 \times 100^{1.5}}{0.9487 \times 100^{1.5} + 0.7071 \times 30^{1.5}} \times 20 + 100 = 117.8178 \quad (8.21)$$

and

$$\sigma_1^{ME}(\text{nonel}) = \frac{0.7071 \times 30^{1.5}}{0.9487 \times 100^{1.5} + 0.7071 \times 30^{1.5}} \times 20 + 30 = 32.1822. \quad (8.22)$$

The resulting cross sections violate the sum rule

$$\sigma_1^{ME}(\text{el}) + \sigma_1^{ME}(\text{nonel}) = 50 + 32.1822 \neq 117.8178 = \sigma_1^{ME}(\text{tot}). \quad (8.23)$$

Therefore, from a mathematical point of view, the Equation 8.18 for redistributing the excess does not guarantee the conservation of sum rules.

Due to this finding, we want restrict the further analysis to a single channel. In the course of this thesis, we regard Bayesian inference as the proper tool for nuclear data evaluation. Therefore, in the following we interpret the SMC method within Bayesian statistics and point out the implied assumptions.

In Equation 8.10 we gave the definition of the combined cross section generated from experimental data and model predictions. In the following discussion we only consider a single cross section at some incident energy and regard the weight ω_i as independent from the value of the drawn quantities $\sigma_i^M(c)$ and $\sigma_i^E(c)$ and hence just write ω . We will also drop the channel specification in the quantities, because we consider only a single channel. Under these assumptions, we obtain for the expectation of the combined cross section

$$\sigma_1 = \text{E}[\sigma_i^{ME}] = \omega \text{E}[\sigma_i^M] + (1 - \omega) \text{E}[\sigma_i^E]. \quad (8.24)$$

The quantity σ_1 would enter a nuclear data file as the best estimate. Even though not necessarily true in practice, the model predictions are assumed to be drawn from a normal distribution, $\sigma_i^M \sim \mathcal{N}(\sigma_0^M, \delta_0^M)$. The quantities σ_0^M and δ_0^M can be estimated from the model predictions. As a reminder, we also assume $\sigma_i^E \sim \mathcal{N}(\sigma^E, \delta^E)$. Under these assumptions Equation 8.24 can be given in the compact form

$$\sigma_1 = \omega \sigma_0^M + (1 - \omega) \sigma^E. \quad (8.25)$$

In Bayesian statistics, the assumption of a normal distribution for the model cross section and the experimental cross section leads to the prior $\pi(\sigma^M) \sim \mathcal{N}(\sigma_0^M, \delta_0^M)$ and the likelihood $\ell(\sigma^E | \sigma^M) \sim \mathcal{N}(\sigma^M, \delta^E)$, respectively. Thus, the solution for the mean vector of the posterior distribution is available in closed form,

$$\sigma_1 = \frac{(\delta^E)^2 \sigma_0^M + (\delta_0^M)^2 \sigma^E}{(\delta_0^M)^2 + (\delta^E)^2}. \quad (8.26)$$

Consequently, the SMC method coincides with Bayesian inference if the weight is chosen as

$$\omega = \frac{(\delta^E)^2}{(\delta_0^M)^2 + (\delta^E)^2}. \quad (8.27)$$

Another assumption of the SMC method concerns the priority given to experimental data, i.e. features of the model such as smoothness are considered of minor importance at incident energies where experimental data are available.

is that the features of the model such as smoothness have to be completely dis-trusted at incident energies where an experimental data point is available. Because of this assumption, evaluated cross section curves may look jagged. If such a result is undesirable, it means we have some prior assumptions about the smoothness of the cross section curve which did not enter the inference procedure.

8.3 Ornstein-Uhlenbeck process

Another approach of Pigni and Leeb (2003) aims directly at the specification of the model defect. The approach was developed as an extension to the surrogate approach (see section 3.3). The idea is to use a prior covariance matrix of the form

$$\mathbf{A} = \mathbf{A}_{\text{mod}} + \mathbf{A}_{\text{def}}, \quad (8.28)$$

where the model covariance matrix \mathbf{A}_{mod} is constructed based on a sample of model predictions with varied parameter sets. The model defect covariance matrix \mathbf{A}_{def} expresses information about the deficiency of the model. Deficiency of the model means the deviation of the best model prediction from the (unobservable) true values.

In the following we describe the construction of the model defect covariance matrix. Suppose that σ_k^{exp} are the measured values of cross sections at energies E_k of the same reaction channel. The associated uncertainties are denoted by δ_k^{exp} . Furthermore, we denote by σ_k^{mod} the model predictions for the cross sections measured in the experiment. Using these quantities, we can introduce the relative errors

$$\delta u_k = \frac{|\sigma_k^{\text{mod}} - \sigma_k^{\text{exp}}|}{\sigma_k^{\text{mod}}}, \quad (8.29)$$

and the associated weights

$$\omega_k = \left(\frac{\sigma_k^{\text{mod}}}{\delta_k^{\text{exp}}} \right)^2. \quad (8.30)$$

The weights are larger for cross sections with lower experimental uncertainty. The introduced quantities serve for the definition of

$$(\delta u)^2 = \frac{\sum_{k=1}^n \omega_k (\delta u_k)^2}{\sum_{k=1}^n \omega_k}, \quad (8.31)$$

which represents a global measure for the deviation of the model from experimental data. It can be termed as the weighted mean squared relative error of the model.

The original paper suggests that experimental data for the calculation of $(\delta u)^2$ should not be included in the Bayesian update procedure in order to avoid double-counting of experimental data. We will argue in section 9.4 why it could be still reasonable to use the experimental data which are included in the update for the determination of the model defect whenever no other experimental data are available.

The calculated δu is used to specify covariances between two model errors $\varepsilon(E_1)$ and $\varepsilon(E_2)$ at energies E_1 and E_2 , respectively,

$$\text{Cov}[\varepsilon(E_1), \varepsilon(E_2)] = \sigma^{\text{mod}}(E_1)\sigma^{\text{mod}}(E_2)\mathcal{C}(E_1, E_2). \quad (8.32)$$

The covariances $\mathcal{C}(E_1, E_2)$ on a relative scale are given by

$$\mathcal{C}(E_1, E_2) = (\delta u)^2 \exp\left(\frac{\delta u}{\delta u_0} |\ln E_1 - \ln E_2|\right). \quad (8.33)$$

In the original paper the choice for the constant δu_0 is 0.01. The covariance function in Equation 8.32 serves to define a model covariance matrix \mathbf{A}_{def} for an arbitrary mesh of incident energies.

The addition of \mathbf{A}_{def} to the prior covariance matrix extends the representation possibilities of the original model, which is a proper measure to address the issue of model misspecification. However, covariances between distinct channels are assumed to be zero. Due to this assumption, consistency properties such as the fulfillment of sum rules are not necessarily preserved by the evaluated quantities. The conservation of sum rules has been an important objective in the novel approach to treat model defects which will be presented in chapter 9.

The form of the relative covariance function in Equation 8.33 defines a Gaussian process, in particular an Ornstein-Uhlenbeck (OU) process (Uhlenbeck and Ornstein, 1930). A random process is the generalization of the concept of a random variable to functions. Thus, a random process defines a probability distribution on a function space. In order to get an idea about the implied assumptions by the choice of an OU process, we set $\delta u = 0.1$ and draw a sample (a function) from the probability distribution defined by the OU process. The obtained function is shown in Figure 8.1.

The illustrated model error function looks very jagged. This observation reflects a theoretical property of the OU process: possible functions are continuous but not differentiable. In the fast neutron region, cross section curves are expected to be rather smooth, therefore the non-differentiability of functions from the OU process may be not in line with our prior belief. Other choices of the covariance function are possible which yield smoother curves.

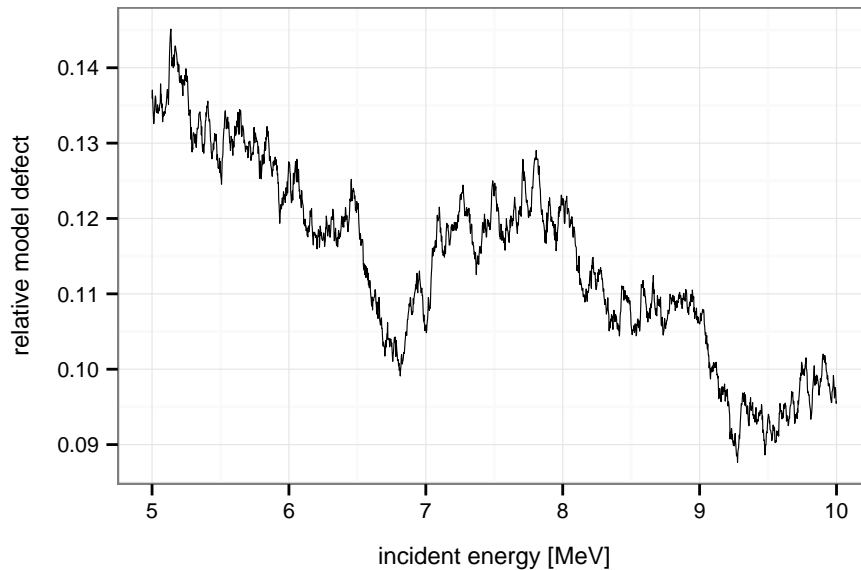


Figure 8.1: A function drawn from an Ornstein-Uhlenbeck process. The parametrization of the process is given in Equation 8.33 with $\delta u_0 = 0.01$ and $\delta u = 0.1$.

8.4 Scaling procedure

The scaling procedure (Leeb, Neudecker, and Srdinko, 2008) aims directly at the specification of the model defect. Analogously to the approach outlined in the previous section, the scaling procedure is an algorithm to construct a model defect covariance matrix \mathbf{A}_{def} , which is added to the prior covariance matrix \mathbf{A}_{mod} of the surrogate approach (see section 3.3).

The fundamental assumption of the scaling procedure is that the model performs comparably well for different isotopes of the same element. If the predictions systematically deviate from experimental data for one isotope, then the same systematic deviation is expected for the isotope of current interest.

The scaling procedure does not provide guidance for the specification of the covariances between cross sections of different reaction channels. They are assumed to be zero. Hence, model defects of different channels are assumed to be independent, and thus we can restrict the further discussion to cross sections of a single reaction channel. In the following we outline the procedure.

The energy region is divided into $m = 1 \dots M$ bins with the energies E_m at the centers of the bins. Further, we consider $n = 1 \dots N$ different isotopes. Let σ_i^{exp} be a measured cross section of isotope n_i at incident energy E_i associated with the energy bin m_i . The cross section σ_i^{mod} is the model prediction for the measured cross section σ_i^{exp} . The index set $\text{bin}(n, m)$ contains all indices i for which $n_i = n$ and $m_i = m$ —in other words for which the cross sections σ_i^{exp} are associated with

isotope n and falling into the energy bin m .

As a first step, the following weights are calculated,

$$\omega_i = \frac{\sigma_i^{\text{mod}}}{\sum_{j \in \text{bin}(n_i, m_i)} \sigma_j^{\text{mod}}} . \quad (8.34)$$

Now a weighted mean value $\langle D_{(n,m)} \rangle$ of the relative deviation of the experimental cross sections from the model predictions can be calculated for every bin(n, m),

$$\langle D_{(n,m)} \rangle = \sum_{i \in \text{bin}(n,m)} \omega_i \left(\frac{\sigma_i^{\text{exp}}}{\sigma_i^{\text{mod}}} \right) . \quad (8.35)$$

Similarly, a weighted non-central second moment $\langle (D_{(n,m)})^2 \rangle$ of the relative deviation of the experimental cross sections from the model predictions for each bin(n, m) is given by

$$\langle (D_{(n,m)})^2 \rangle = \sum_{i \in \text{bin}(n,m)} \omega_i \left(\frac{\sigma_i^{\text{exp}}}{\sigma_i^{\text{mod}}} \right)^2 . \quad (8.36)$$

The weights in Equation 8.34 give emphasis in each bin(n, m) on deviations associated with larger model predictions $\sigma_i^{\text{mod}}, i \in \text{bin}(n, m)$. Please also note that the sum of all weights associated with a single bin(n, m) is one by construction.

In the scaling procedure, the prior cross section curve is rescaled by the following factor

$$D = \frac{1}{N} \sum_{i=1}^N \langle D_{(n)} \rangle \quad \text{with} \quad \langle D_{(n)} \rangle = \sum_{m=1}^M \tilde{\omega}_{(n,m)} \langle D_{(n,m)} \rangle . \quad (8.37)$$

The weights $\tilde{\omega}_{(n,m)}$ should be in correspondence with the weights ω_i . This can be achieved by the assignment

$$\tilde{\omega}_{(n,m)} = \frac{\sum_{i \in \text{bin}(n,m)} \omega_i}{\sum_{m=1}^M \sum_{i \in \text{bin}(n,m)} \omega_i} . \quad (8.38)$$

The elements of the model defect covariance matrix defined on a relative scale are specified as

$$\mathcal{C}(E_m, E_{m'}) = \frac{1}{N} \sum_{k=1}^N \{ [\langle D_{(k,m)} \rangle - D] \times [\langle D_{(k,m')} \rangle - D] + T_{m,m'} \} . \quad (8.39)$$

The term $T_{mm'}$ is defined as

$$T_{m,m'} = \delta_{m,m'} \left(\langle (D_{(n,m)})^2 \rangle - \langle D_{(n,m)} \rangle \right) \quad (8.40)$$

and its purpose is to account for the limited accuracy of the experimental data.

Now the covariance between two model defects $\varepsilon(E_m)$ and $\varepsilon(E_{m'})$ for energy bin m and m' , respectively, is given by

$$\text{Cov}[\varepsilon(E_m), \varepsilon(E_{m'})] = \sigma^{\text{mod}}(E_m) \sigma^{\text{mod}}(E_{m'}) \mathcal{C}(E_m, E_{m'}) . \quad (8.41)$$

where $\sigma^{\text{mod}}(E_m)$ is the model prediction for the isotope of interest at the center energy E_m of the m^{th} energy bin. Using these covariances, we can specify the elements $(\mathbf{A}_{\text{def}})_{ij}$ of the model defect covariance matrix by

$$(\mathbf{A}_{\text{def}})_{ij} = \text{Cov}[\varepsilon(E_i), \varepsilon(E_j)]. \quad (8.42)$$

The whole procedure appears rather involved. In order to understand the structure of the model defect as defined in Equation 8.42, we study an idealized scenario. We consider a single angle integrated cross section channel. As in section 6.1, we assume that the model is given by a straight line,

$$\mathcal{M}(E) = kE + d, \quad (8.43)$$

where k is the slope and d the intercept of the straight line. Given the incident energy E , this model predicts the respective cross section. Suppose that several isotopes of the chemical element of interest exist. Further assume that we are able to measure the cross sections of each isotope with perfect precision, and the measured cross section curves of the isotopes are identical. We assume that the cross section curve of each isotope determined by the measurements is given by

$$\sigma_{\text{true}}(E) = (E - 15)^2 + 10. \quad (8.44)$$

The model parameters which describe this curve between 10 and 20 MeV best (in the sense of χ^2) on the basis of thousand equidistant experimental data points are

$$k = 0 \quad \text{and} \quad d \approx 18.35. \quad (8.45)$$

The estimate of d varies slightly with the number of data points used for χ^2 -fitting. The situation is illustrated in Figure 8.2. Under the assumed conditions, the term $T_{m,m'}$ in Equation 8.39 vanishes.

We can create a dense grid of perfectly precise measurements σ_i^{exp} and assign exactly one measurement to each energy bin. Hence, all weights ω_i are one and we can discard them from the formulas. We computed the model defect covariance matrix for this scenario. The corresponding correlation matrix and uncertainties are illustrated in Figure 8.3. The model defect covariance matrix \mathbf{A}_{def} has rank one. Consequently, only one eigenvector is associated with a non-zero eigenvalue. This eigenvector determines the shape of the model error. The permitted shapes are visualized in Figure 8.4.

Finally, we perform a Bayesian update using the uninformative prior

$$\begin{pmatrix} k_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 15 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_0 = \begin{pmatrix} 15 & 0 \\ 0 & 100 \end{pmatrix}. \quad (8.46)$$

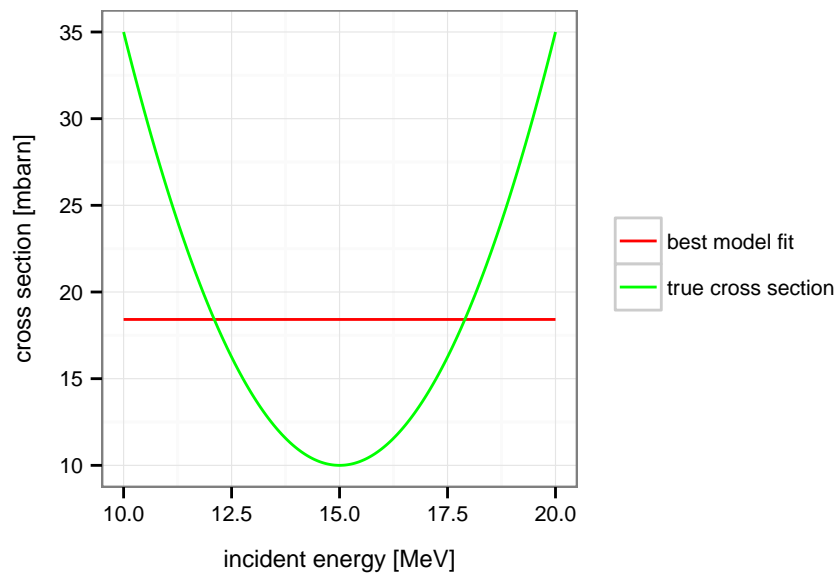


Figure 8.2: Schematic scenario. Perfectly precise measurements are located on the true cross section curve. The best model fit to that curve is a horizontal line. The true cross section curve is identical for all isotopes and so are the best model fits.

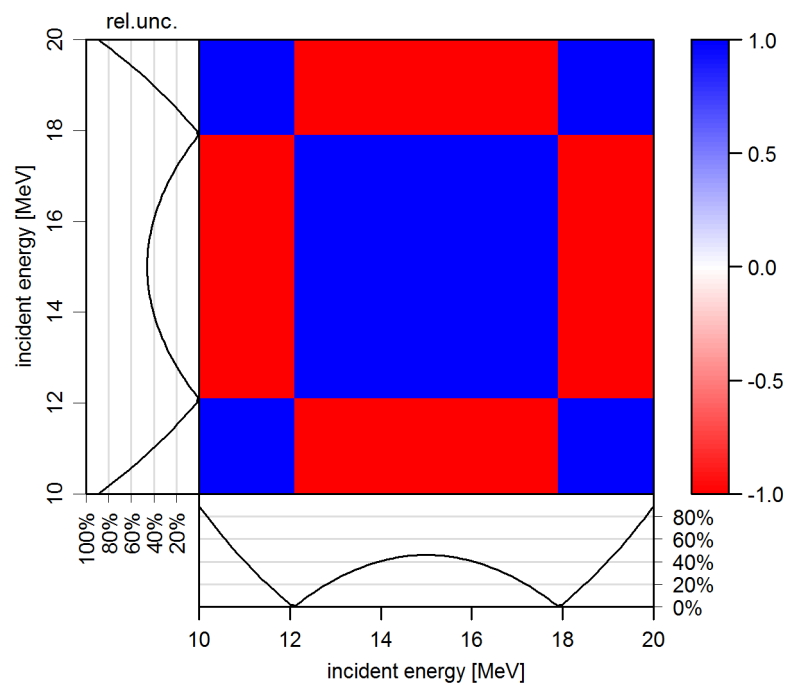


Figure 8.3: The correlations of the model defect covariance matrix in the schematic scenario.

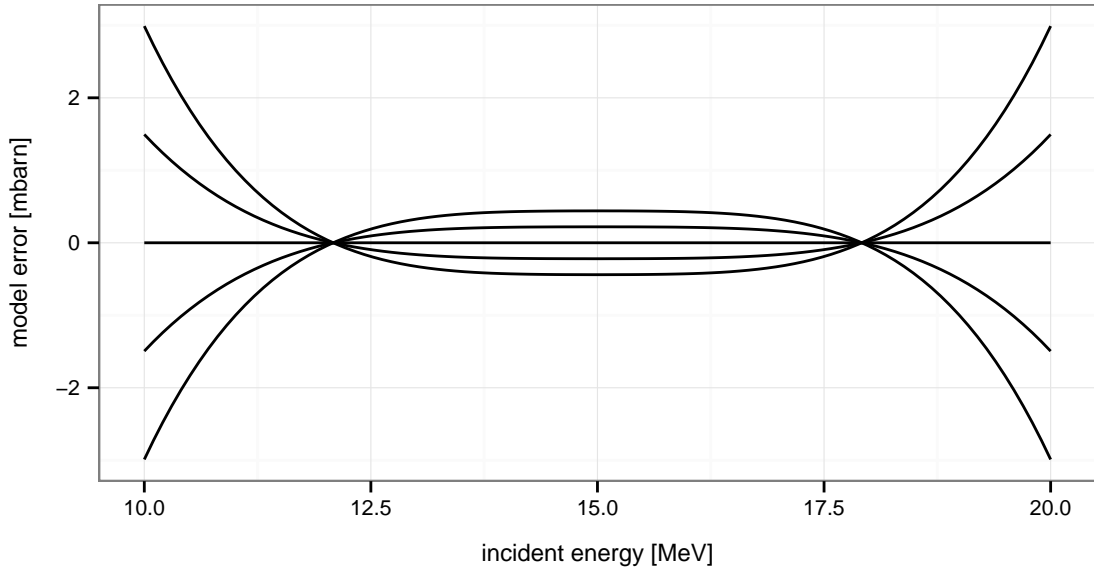


Figure 8.4: Possible shapes of the model error permitted by the model defect covariance matrix.

Using the sensitivity matrix \mathbf{S} characterized by $S_{i1} = E_i$ and $S_{i2} = 1$, we can specify the complete prior covariance matrix \mathbf{A} which includes the model defect covariance matrix,

$$\mathbf{A} = \mathbf{S}\mathbf{A}_0\mathbf{S}^T + \mathbf{A}_{\text{def}}. \quad (8.47)$$

Due to the assumption of perfectly precise measurements, the experimental covariance matrix would be given by the matrix of zeros. But then we would not be able to apply the update formulas Equation 3.33 and Equation 3.34. Hence we added a small value 10^{-6} to the diagonal elements of the experimental covariance matrix to make the formulas usable. The result of the Bayesian update is illustrated in Figure 8.5. The evaluated cross section curve is closer to the true cross section. However, in the schematic scenario with perfect knowledge about the true cross section curve and the assumption that the systematic deviations of the model prediction from the true cross section curve is the same for each isotope, we would have expected a perfect match of the evaluated cross section curve and the true cross section curve. The schematic scenario can be regarded as a consistency test for methods that take into account model defects and base its construction on the available information about neighbor isotopes.

8.5 Conclusions

None of the existing approaches guarantees the conservation of important consistency properties such as the fulfillment of sum rules. None of the existing

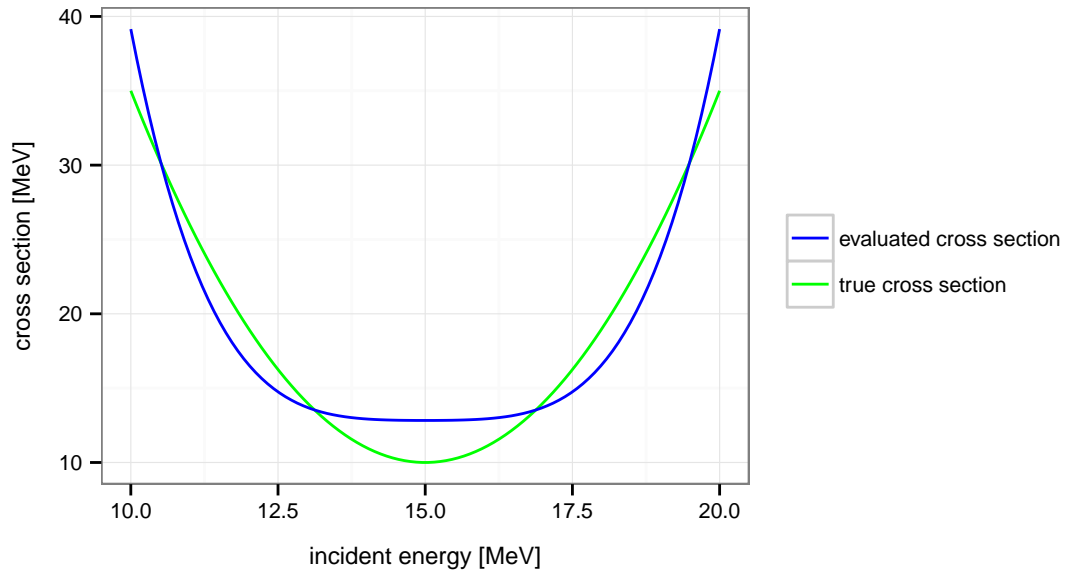


Figure 8.5: Update in the schematic scenario. The model defect covariance matrix \mathbf{A}_{def} has been included in the prior.

approaches takes into account smoothness constraints. Consequently, given a sufficiently dense mesh of incident energies, evaluated cross section curves may look jagged. However, in the fast neutron energy region, cross section curves are expected to be rather smooth. Even though possible for the scaling procedure and the Ornstein-Uhlenbeck process, the evaluated model defect function has never been computed and studied. Knowledge about the form of the model defect can be regarded as valuable information to refine the physical models.

In the next chapter we formulate a solution of these issues by means of Gaussian processes. With the exception of χ^2 -rescaling, all of the presented methods can be interpreted as special forms of Gaussian processes.

Consistent treatment of model defects

In this chapter, we introduce a new way to take into account model defects, which preserves important consistency constraints such as the sum rules of cross sections. Furthermore, we establish the link to Gaussian processes and exploit the respective methodology. The use of Gaussian processes enables us to explicitly estimate the magnitude and the functional form of the model deficiency. Furthermore, we obtain information on how the systematics of the model is inadequate.

The new approach is fully justified within the framework of Bayesian statistics. Therefore, every evaluation method based on Bayesian statistics can be augmented by the new approach to account for model defects. Specific details about how the Bayesian update formula is evaluated—by the linearization of the model or Monte Carlo sampling—are not relevant in this respect.

9.1 Model defects as Gaussian processes

Gaussian process regression is a powerful statistical tool to perform non-parametric regression. Historically, this technique was for the first time described in the master thesis of Krige (1951), which dealt with the estimation of abundances of gold on the basis of known abundances at some places. The technique was later formalized by Matheron (1962). In contrast to regression based on a parametric model, the specification of a function for which the optimal parameters are sought is not required. Instead a probability distribution is defined on a function space. We will show in this section that this approach is well suited for the treatment of model defects of nuclear models.

Usually there is no prior knowledge about the specific functional form of the model defect, otherwise it would be possible to construct a better nuclear model from the beginning. However, there are expectations about the smoothness of an evaluated cross section curve and to what extent reliable experimental data can be described by the model. These expectations provide already enough information to define a probability distribution on a function space.

Nowadays, Gaussian process regression is profitably applied in many fields, such as geostatistics, meteorology, economics, robotics, and machine learning. Frequently, it is used to estimate a function from data in order to bypass the difficulty to specify a parametric model. In contrast to that, the idea in this thesis is to use the parametric nuclear model as a basis and to model its deficiency as a Gaussian process. The idea to use a Gaussian process to describe model deficiency as such is not particularly new. Blight and Ott (1975) described how the model deficiency in form of a Gaussian process can be taken into account in polynomial regression.

The Ornstein-Uhlenbeck process and the scaling procedure discussed in chapter 8 as an extension of the surrogate approach rely on special forms of Gaussian processes. However, the link to Gaussian process regression was not realized. Consequently, the respective statistical methodology was not exploited. As an example, the approach relying on an Ornstein-Uhlenbeck process leads to shapes of model defects that are nowhere differentiable, which may not properly reflect prior assumption on the smoothness of cross section curves. Further, the specification $\delta u_0 = 0.01$ in Equation 8.33 appears rather ad-hoc. Such parameters of the so-called *covariance functions* can be estimated by maximizing the marginal likelihood, which is a well-defined statistical method (see section 9.4).

The approaches so far that directly aim at the specification of the model defect were exclusively combined with linearized versions of the Bayesian update formula. This fact might suggest that model defects cannot be properly taken into account in Monte Carlo procedures. However, the linearization of the nuclear model is independent from the treatment of model defects. Sound Bayesian statistics can be carried out equally well using linearized models (and hence closed form Bayesian update formulas) and exact models (and hence using Monte Carlo sampling). After this brief introduction and important clarifications, we formulate in this section model defects as Gaussian processes.

We start with the definition of a Gaussian process:

Definition of a Gaussian process

Let the function space \mathcal{F} contain all functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Consider a probability distribution $\rho(\cdot)$ on \mathcal{F} that assigns to every function $f \in \mathcal{F}$ the probability density $\rho(f)$. We equally well think of the probability density as a functional $\rho[f(\vec{x})] : \mathcal{F} \rightarrow [0, \infty]$. The probability density distribution $\rho(\cdot)$ induces a Gaussian process if for any finite set of $\{\vec{x}_i\}_{i=1..N}, \forall i : \vec{x}_i \in \mathbb{R}^d$ the associated function values $\{f(\vec{x}_i)\}_{i=1..N}$ follow a multivariate normal distribution. In other words, a Gaussian process is defined over all marginal distributions of finitely many function values.

A general introduction to Gaussian processes is given by Rasmussen and Williams (2006).

Because it is not possible to exhaustively enumerate the mean vectors and covariance matrices for all possible sets $\{f(\vec{x}_i)\}_{i=1..N}$, the mean vectors and covariance matrices have to be defined over a so-called *mean function* $m(\vec{x})$ and *covariance function* $k(\vec{x}, \vec{x}')$. These functions can be expressed in terms of the expectation and covariance operator,

$$m(\vec{x}) = \mathbb{E}[f(\vec{x})], \tag{9.1}$$

$$k(\vec{x}, \vec{x}') = \text{Cov}[f(\vec{x}), f(\vec{x}')]. \tag{9.2}$$

The mean function and the covariance function completely characterize the Gaussian process. In analogy to a random variable which is characterized by a probability distribution over its possible realizations, the probability distribution induced by the mean function and covariance function characterizes a *random function*. We express this by writing

$$f(\vec{x}) \sim \mathcal{GP}(m(\vec{x}), k(\vec{x}, \vec{x}')). \tag{9.3}$$

Realizations of a random function can be generated by drawing from the respective probability density distribution. For this purpose, a mesh of functions arguments $\{\vec{x}_i\}_{i=1..N}$ has to be specified. Then, for every tuple $(\vec{x}_i, \vec{x}_j), i = 1..N, j = 1..N$, the value $k(\vec{x}_i, \vec{x}_j)$ of the covariance function is evaluated. Further, the mean function is evaluated for every $\vec{x}_i, i = 1..N$. On the basis of these results, a mean vector \vec{x}_0 and a covariance matrix \mathbf{A}_0 are constructed,

$$x_{0,i} = m(\vec{x}_i) \text{ and } A_{0,ij} = k(\vec{x}_i, \vec{x}_j). \tag{9.4}$$

Now one can draw from the multivariate normal distribution $\mathcal{N}(\vec{x}_0, \mathbf{A}_0)$ to obtain the values $f(\vec{x}_i)$ of a particular realization of the random function. Under

the condition that the functions associated with non-vanishing probability density are sufficiently smooth, interpolating the values between the mesh points yields a good approximation to the realization of the random function. The generation of realizations of random functions is useful to learn which functions shapes are considered very likely. We can then judge if the assumed Gaussian process adequately reflects our prior beliefs, for instance concerning the smoothness of the function. However, we have to remark that the Bayesian inference procedure does not require sampling from the Gaussian process.

For the following discussion, we assume that the mean function is zero everywhere, $m(\vec{x}) \equiv 0$. With regard to the specification of a model defect, this assumption seems to be reasonable. Without prior knowledge about the failure of the nuclear model, one might consider the model prediction still as the best estimate if no experimental data are available to give evidence against this assumption.

We restrict ourself to one-dimensional arguments x for the moment. The extension to vector-valued arguments will become relevant for the treatment of model defects in several channels (see section 9.6). To give some practical intuition about Gaussian processes, we introduce the *squared exponential* covariance function, which is a very common choice in the field of machine learning,

$$k(E, E') = \delta^2 \exp\left(-\frac{1}{2\lambda}(E - E')^2\right). \quad (9.5)$$

It depends on two *hyperparameters* δ and λ . The term hyperparameter should remind that these parameters do not characterize a certain function but characterize the probability distribution on a function space. In order to visualize the impact of the values of the hyperparameters on the shapes of the most likely functions, we generated several realizations of the random function for two choices of hyperparameter sets (δ, λ) . Figure 9.1 shows samples for the choice $\delta = 0.3$ and $\lambda = 20$ and Figure 9.2 for the choice $\delta = 0.1$ and $\lambda = 5$. We labeled the axes according to the meaning in nuclear data evaluation. From Figure 9.1 and Figure 9.2 the meaning of the hyperparameters is well conceivable. The parameter λ determines the length-scale of changes with incident energy of the most likely error functions. The expected number of upcrossings $E[N]$ of the zero-line in a unit interval for a stationary covariance function $k(x - x')$ is given by (Adler, 2010, Theorem 4.1.1)

$$E[N] = \frac{1}{2\pi} \sqrt{\frac{-k''(0)}{k(0)}}. \quad (9.6)$$

This quantity takes the value $(2\pi\lambda)^{-1}$ for the squared exponential. In some cases, this theoretical result makes allows a direct specification of the length-scale. For instance, a statistical (physical) model may be only capable to describe the average

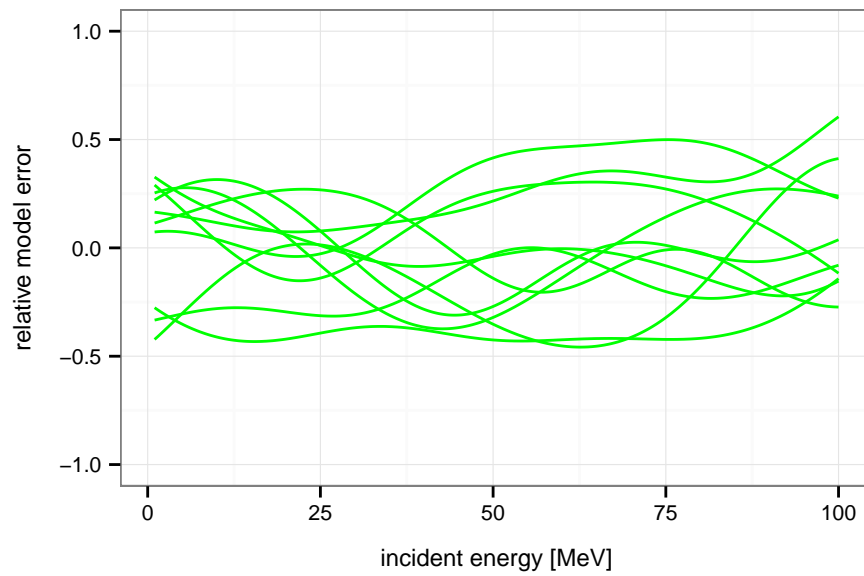


Figure 9.1: Realizations of the Gaussian process defined by the squared exponential covariance function with $\delta = 0.3$ and $\lambda = 20$.

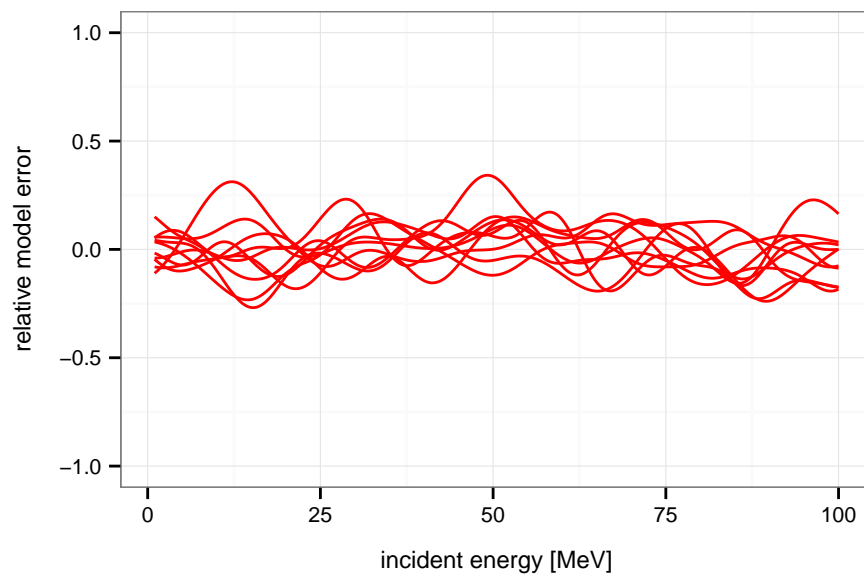


Figure 9.2: Realizations of the Gaussian process defined by the squared exponential covariance function with $\delta = 0.1$ and $\lambda = 5$.

cross section but not the fluctuations due to resonances. If an expectation about the typical resonance width Γ exists, one can set $\lambda = \Gamma/\pi$. If we expect the fluctuations of the model defect to decrease with increasing incident energy, we are free to take the logarithm of the incident energies in the exponent,

$$k(E, E') = \delta^2 \exp\left(-\frac{1}{2\lambda}(\log E - \log E')^2\right). \quad (9.7)$$

The other hyperparameter δ specifies the expected magnitude of the model error. If there is an expectation about the typical χ^2 -value of model fits to experimental data, the quantity $\chi^2/N - 1$ with N the number of data points can serve as a good guess for δ^2 . However, there is also the possibility to set hyperparameters automatically by maximizing the marginal likelihood.

In the following sections, we will usually make use of the squared exponential covariance function. However, other choices are possible. For instance, the covariance function of the Ornstein-Uhlenbeck process in Equation 8.33 is another option. Both the Ornstein-Uhlenbeck and the squared exponential covariance function are within the Matern class given by (Stein, 1999)

$$k_{\text{Matern}}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r\sqrt{2\nu}}{\lambda}\right)^\nu K_\nu\left(\frac{r\sqrt{2\nu}}{\lambda}\right), \quad (9.8)$$

where $r = |E - E'|$ and ν and λ are positive hyperparameters. The function K_ν is a modified Bessel function (Abramowitz and Stegun, 1965, sec. 9.6). The choice $\nu = 1/2$ leads to the Ornstein-Uhlenbeck process and the squared exponential covariance function is realized for $\nu \rightarrow \infty$. The hyperparameter ν determines the roughness of the functions. Higher values of ν are associated with smoother functions.

An important property of covariance functions is their additivity in order to construct new covariance functions. This property is very useful to tailor the covariance function to the specific situation. For instance, the optical model is known to give a very accurate prediction of the total cross section at higher incident energies, but cannot properly describe the resonance range. In this case, we can define a Gaussian process to capture resonances at low incident energies,

$$k_{\text{res}}(E, E') = \theta(E_c - E)\theta(E_c - E')\delta_1^2 \exp\left(-\frac{1}{2\lambda}(\log E - \log E')^2\right), \quad (9.9)$$

where $\theta(E_c - E) = 1$ for $E < E_c$ and zero otherwise. Another Gaussian process k_{glob} can be defined for the complete range of incident energies to capture global model deficiencies. The combined Gaussian process is then given by $k(E, E') = k_{\text{res}}(E, E') + k_{\text{glob}}(E, E')$. This property of Gaussian processes is also

the key to construct model defects that preserve the sum rules of cross sections (see section 9.6).

Independent of the choice of the covariance function, the model error can be defined either on a relative scale or an absolute scale. If we directly use a covariance function such as Equation 9.7 to specify the model error, we would use an absolute scale. However, this choice is problematic due to the dynamic range of cross section values. A cross section may rise from zero at its threshold energy to several thousand millibarn. A stationary covariance function on an absolute scale would either allow negative cross sections at energies associated with very low cross section values or assign unreasonable low uncertainty to large cross sections. Consequently, it is reasonable to define the covariance function $k(E - E')$ on a relative scale. This can be achieved by

$$\tilde{k}(E - E') = \sigma(E)\sigma(E')k(E, E'), \quad (9.10)$$

where $\sigma(E)$ is the prior best estimate of the cross section stemming from a model calculation. The covariance function $k(E - E')$ on a relative scale is rescaled to obtain the covariance function $\tilde{k}(E - E')$ on an absolute scale.

Finally, we point out the relationships between existing approaches for model defects with Gaussian processes. The method based on the Ornstein-Uhlenbeck process obviously uses a Gaussian process. The scaling procedure discussed in section 8.4 constructs a model defect covariance matrix \mathbf{A}_{def} in a specific way. In the Bayesian inference procedure \mathbf{A}_{def} is transformed to the incident energies of the experimental data by $\mathbf{S}\mathbf{A}_{\text{def}}\mathbf{S}^T$, where \mathbf{S} is the sensitivity matrix. As shown in section 3.4, elements S_{ij} of the sensitivity matrix are fully determined by the incident energies E_i of the experimental data points and the incident energies E_j of the model mesh. Because the sensitivity matrix can be constructed for experiments with arbitrary incident energies E within the boundaries of the model mesh, we may equally write $S(E, E_j)$. Covariance elements associated with the model defect covariance matrix between arbitrary energies are then given by

$$A_{\text{def}}(E, E') = \sum_{i=1}^M \sum_{j=1}^M S(E, E_i)A_{\text{def},ij}S(E', E_j), \quad (9.11)$$

which represents a covariance function. At the end of section 8.2 we showed that the Symmetric Monte Carlo procedure coincides with Bayesian inference for a certain choice of the weights (see Equation 8.27) and the assumption of a normal distribution for the cross sections predicted by the nuclear model. Therefore, we can interpret the SMC procedure in the following way. The original nuclear model is replaced by a Gaussian process with mean function $m(E) = \text{E}[\sigma^{\text{mod}}(E)]$ where $\sigma^{\text{mod}}(E)$ is the arithmetic mean of the drawn model predictions, and covariance function $k(E, E') = \delta(E - E')\text{Var}[\sigma^{\text{mod}}(E)]$.

9.2 Integration into the Bayesian update

The previous section was focussed on the formulation of the model defect in terms of a Gaussian process. This knowledge on model defects represents apriori knowledge which has to be combined with the prior knowledge about the values of the model parameters in the Bayesian inference procedure. Before we combine these two pieces of knowledge, we want to recapitulate Bayesian inference as commonly applied technique in nuclear data evaluation.

In nuclear data evaluation one usually assumes that the experimental cross section vector $\vec{\sigma}_{\text{exp}}$ is given as the sum of the model prediction $\vec{\sigma}_{\text{mod}} = \mathcal{M}(\vec{p})$ and the measurement error $\vec{\varepsilon}_{\text{exp}}$ of the experiment,

$$\vec{\sigma}_{\text{exp}} = \vec{\sigma}_{\text{mod}} + \vec{\varepsilon}_{\text{exp}}. \quad (9.12)$$

The additional assumption that the measurement error is governed by a multivariate normal distribution $\vec{\varepsilon} \sim \mathcal{N}(\vec{0}, \mathbf{B})$ leads to the likelihood $\ell(\vec{\sigma}_{\text{exp}} | \vec{p}) = \mathcal{N}(\mathcal{M}(\vec{p}), \mathbf{B})$. If we further assume that also the uncertainty about the values of the model parameters is specified as a multivariate normal distribution, $\pi(\vec{p}) \sim \mathcal{N}(\vec{p}_0, \mathbf{A}_0)$, we obtain the Bayesian update formula

$$\begin{aligned} \log \pi(\vec{p} | \vec{\sigma}_{\text{exp}}) = \log C - \frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}))^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p})) \\ - \frac{1}{2} (\vec{p} - \vec{p}_0)^T \mathbf{A}_0^{-1} (\vec{p} - \vec{p}_0). \end{aligned} \quad (9.13)$$

where the normalization constant C is given by

$$C^{-1} = \int \ell(\vec{\sigma}_{\text{exp}} | \vec{p}) \pi(\vec{p}) \, d\vec{p}. \quad (9.14)$$

Moments of the posterior distribution in Equation 9.13 can be evaluated via Monte Carlo sampling techniques, such as importance sampling or the Metropolis-Hastings algorithm. If the model is linearized (section 3.2) or replaced by a surrogate model (section 3.3), the posterior is also multivariate normal and closed form expressions are available to calculate its mean vector and covariance matrix. Independent of how the moments of the posterior are obtained, most evaluation techniques are based on the assumption of a perfect nuclear model which yields a perfect description of reality. Thus, the errors are composed of parameter uncertainties and experimental uncertainties. The analysis in chapter 6 showed that even small model defects may lead to undesirable evaluated cross section curves.

The inclusion of model defects in an evaluation is straight-forward. The statistical concept of nuclear data evaluation has to be extended with a model deficiency term,

$$\vec{\sigma}_{\text{exp}} = \vec{\sigma}_{\text{mod}} + \vec{\varepsilon}_{\text{mod}} + \vec{\varepsilon}_{\text{exp}}. \quad (9.15)$$

The model deficiency term $\vec{\varepsilon}_{\text{mod}}$ is specified by a Gaussian process, $\mathcal{GP}(m(x), k(x, x'))$. It should be remarked that in the novel formulation, the true cross section vector is the sum of the model prediction and the model deficiency, $\vec{\sigma}_{\text{true}} = \vec{\sigma}_{\text{mod}} + \vec{\varepsilon}_{\text{mod}}$. This is the essential conceptual difference to the standard Bayesian inference procedure used in nuclear data evaluation so far. Assuming a multivariate normal distribution for the measurement error, $\vec{\varepsilon}_{\text{exp}} \sim \mathcal{N}(\vec{0}, \mathbf{B})$, the likelihood is given by $\ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \vec{\varepsilon}_{\text{mod}}) = \mathcal{N}(\mathcal{M}(\vec{p}) + \vec{\varepsilon}_{\text{mod}}, \mathbf{B})$. Whenever $\mathcal{M}(\vec{p})$ appears, it denotes the model prediction vector for some parameter set \vec{p} that fits in the context. The model defect vector $\vec{\varepsilon}_{\text{mod}}$ contains the model defects at the incident energies of the measurement vector $\vec{\sigma}_{\text{exp}}$. We also have to introduce the vector $\vec{\varepsilon}_{\text{pred}}$ which contains the model defects at incident energies of interest. Both vectors are combined to $\vec{\varepsilon}_{\text{comb}}^T = (\vec{\varepsilon}_{\text{pred}}^T, \vec{\varepsilon}_{\text{mod}}^T)$. With the choice of a multivariate normal prior $\pi(\vec{p}) = \mathcal{N}(\vec{p}_0, \mathbf{A}_0)$ and the assumption of apriori independence of model parameters and the model deficiency, $\pi(\vec{p} | \vec{\varepsilon}_{\text{mod}}) = \pi(\vec{p})$, we obtain the posterior distribution

$$\begin{aligned} \log \pi(\vec{p}, \vec{\varepsilon}_{\text{comb}} | \vec{\sigma}_{\text{exp}}) &= \log \tilde{C} + \log \ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \vec{\varepsilon}_{\text{mod}}) + \log \pi(\vec{p}) + \log \pi(\vec{\varepsilon}_{\text{comb}}) \\ &= \log C - \frac{1}{2} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}) - \vec{\varepsilon}_{\text{mod}})^T \mathbf{B}^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}(\vec{p}) - \vec{\varepsilon}_{\text{mod}}) \\ &\quad - \frac{1}{2} (\vec{p} - \vec{p}_0)^T \mathbf{A}_0^{-1} (\vec{p} - \vec{p}_0) \\ &\quad - \frac{1}{2} (\vec{\varepsilon}_{\text{comb}})^T \mathbf{K}_0^{-1} (\vec{\varepsilon}_{\text{comb}}). \end{aligned} \tag{9.16}$$

In the case of a uniform prior distribution for the the model parameters, the term $\log \pi(\vec{p})$ can be discarded from the equation. Elements of \mathbf{K}_0 are determined by the covariance function of the Gaussian process, $K_{0,ij} = k(E_i, E_j)$, where the E_i are the incident energies associated with the elements of $\vec{\varepsilon}_{\text{comb}}$. We remind that the prior mean vector of the model defect is the zero vector and therefore does not show up in Equation 9.16.

Important summary statistics can be computed from this posterior. The posterior expectation of the true cross section $\vec{\sigma}_{\text{true}}(\vec{p}, \vec{\varepsilon}_{\text{pred}}) = \mathcal{M}(\vec{p}) + \vec{\varepsilon}_{\text{pred}}$ is given by

$$\vec{\sigma}_{\text{true}}^{\text{best}} = E[\vec{\sigma}_{\text{true}}(\vec{p}, \vec{\varepsilon}_{\text{pred}})] = \int \vec{\sigma}_{\text{true}}(\vec{p}, \vec{\varepsilon}_{\text{pred}}) \pi(\vec{p}, \vec{\varepsilon}_{\text{comb}} | \vec{\sigma}_{\text{exp}}) d\vec{p} d\vec{\varepsilon}_{\text{comb}}. \tag{9.17}$$

The cross sections in $\vec{\sigma}_{\text{true}}^{\text{best}}$ enter nuclear data files as best estimates. Another important summary statistics is the posterior covariance matrix

$$\begin{aligned} \mathbf{A}_{\text{true}}^{\text{best}} = \text{Var}[\vec{\sigma}_{\text{true}}] &= \int (\vec{\sigma}_{\text{true}}(\vec{p}, \vec{\varepsilon}_{\text{pred}}) - \vec{\sigma}_{\text{true}}^{\text{best}}) (\vec{\sigma}_{\text{true}}(\vec{p}, \vec{\varepsilon}_{\text{pred}}) - \vec{\sigma}_{\text{true}}^{\text{best}})^T \times \\ &\quad \pi(\vec{p}, \vec{\varepsilon}_{\text{comb}} | \vec{\sigma}_{\text{exp}}) d\vec{p} d\vec{\varepsilon}_{\text{comb}}, \end{aligned} \tag{9.18}$$

which enters nuclear data files to characterize the uncertainty of the best estimate $\vec{\sigma}_{\text{true}}^{\text{best}}$. A new feature of the approach is the possibility to explicitly estimate the form of the model deficiency,

$$\vec{\varepsilon}_{\text{pred}}^{\text{best}} = \text{E}[\vec{\varepsilon}_{\text{pred}}] = \int \vec{\varepsilon}_{\text{pred}} \pi(\vec{p}, \vec{\varepsilon}_{\text{comb}} | \vec{\sigma}_{\text{exp}}) d\vec{p} d\vec{\varepsilon}_{\text{comb}}. \quad (9.19)$$

The associated covariance matrix can be calculated analogously to Equation 9.18. A quantity that may be considered of fundamental interest is the posterior covariance matrix between \vec{p} and $\vec{\varepsilon}_{\text{pred}}$, which gives information about the systematics the model is not able to capture

$$\mathbf{C} = \text{Cov}[\vec{\varepsilon}_{\text{pred}}, \vec{p}] = \int (\vec{\varepsilon}_{\text{pred}} - \vec{\varepsilon}_{\text{pred}}^{\text{best}}) (\vec{p} - \vec{p}^{\text{best}})^T \times \pi(\vec{p}, \vec{\varepsilon}_{\text{comb}} | \vec{\sigma}_{\text{exp}}) d\vec{p} d\vec{\varepsilon}_{\text{comb}}, \quad (9.20)$$

where $\vec{p}^{\text{best}} = \text{E}[\vec{p}]$ is calculated analogously to Equation 9.19.

These formulas are difficult to evaluate due to the high dimensionality of the parameter space and the time needed to evaluate the nuclear model for one parameter set. One possibility to address this difficulty is to use Monte Carlo integration; another possibility is to work with a simplified nuclear model. In this thesis we follow the latter approach.

9.3 Linearization of the Bayesian update

In chapter 3 we described two approaches to simplify the model in order to obtain closed form Bayesian update formulas. In the following we restrict ourselves to the linearized model, while the surrogate approach (section 3.3) can be regarded as a special case of the linearization of the model (section 3.2). The linearized version of the model \mathcal{M} has the form

$$\mathcal{M}_{\text{lin}}(\vec{p}) = \vec{\sigma}_{\text{ref}} + \mathbf{S}(\vec{p} - \vec{p}_{\text{ref}}) = \vec{\Delta} + \mathbf{S}\vec{p}, \quad (9.21)$$

where $\vec{\sigma}_{\text{ref}}$ is the model prediction at the expansion point \vec{p}_{ref} , the vector \vec{p} contains the model parameters, and the sensitivity matrix \mathbf{S} is the Jacobian matrix $\partial\vec{\sigma}/\partial\vec{p}$ of the nuclear model. In the case of the surrogate approach, we identify $\vec{p} = \vec{\sigma}_{\text{mod}}$, $\vec{p}_{\text{ref}} = \vec{0}$, $\vec{\sigma}_{\text{ref}} = \vec{0}$.

For the sake of clarity we have to introduce some additional notation. In the following discussion three meshes of incident energies will be relevant: 1) the mesh \mathcal{I}_{exp} associated with the measurement vector $\vec{\sigma}_{\text{exp}}$, 2) the mesh $\mathcal{I}_{\text{pred}}$ which contains the incident energies for which we want to obtain estimates of cross sections, and 3) the combined mesh $\mathcal{I}_{\text{comb}}$ which contains both the incident energies of \mathcal{I}_{exp} and

$\mathcal{I}_{\text{pred}}$. The sensitivity matrices for the mapping of a model parameter vector to cross section values at the incident energies of these meshes are denoted by \mathbf{S}_{exp} , \mathbf{S}_{pred} , and \mathbf{S}_{comb} , respectively. The model defect vector $\vec{\varepsilon}_{\text{mod}}$ is defined on the mesh $\mathcal{I}_{\text{comb}}$. The matrices \mathbf{T}_{exp} and \mathbf{T}_{pred} select from $\vec{\varepsilon}_{\text{mod}}$ the entries associated with \mathcal{I}_{exp} and $\mathcal{I}_{\text{pred}}$, respectively. The quantity $\vec{\Delta}$ in Equation 9.21 is defined on the mesh $\mathcal{I}_{\text{comb}}$.

First, we construct the statistical model. Inserting the linearized model into Equation 9.15 yields

$$\vec{\sigma}_{\text{exp}} = \mathbf{T}_{\text{exp}}\vec{\Delta} + \mathbf{S}_{\text{exp}}\vec{p} + \mathbf{T}_{\text{exp}}\vec{\varepsilon}_{\text{mod}} + \vec{\varepsilon}_{\text{exp}}. \quad (9.22)$$

We make the following choices for the prior probability distributions:

$$\vec{p} \sim \mathcal{N}(\vec{p}_0, \mathbf{A}_0), \vec{\varepsilon}_{\text{mod}} \sim \mathcal{N}(\vec{0}, \mathbf{K}_0), \text{ and } \vec{\varepsilon}_{\text{exp}} \sim \mathcal{N}(\vec{0}, \mathbf{B}). \quad (9.23)$$

The covariance matrix \mathbf{K}_0 is constructed with the covariance function of the Gaussian process. Further we assume the mutual independence of \vec{p} , $\vec{\varepsilon}_{\text{mod}}$ and $\vec{\varepsilon}_{\text{exp}}$ a priori. Thus, the statistical model is completely defined.

For the derivation of the Bayesian update formulas, we need all covariances between $\vec{\sigma}_{\text{exp}}$ on the left hand side of Equation 9.22 and the variables on the right hand side:

$$\text{Cov}[\vec{\sigma}_{\text{exp}}, \vec{p}] = \mathbf{S}_{\text{exp}}\mathbf{A}_0, \quad (9.24)$$

$$\text{Cov}[\vec{\sigma}_{\text{exp}}, \vec{\varepsilon}_{\text{mod}}] = \mathbf{T}_{\text{exp}}\mathbf{K}_0. \quad (9.25)$$

In addition, we need the prior covariance matrix of $\vec{\sigma}_{\text{exp}}$,

$$\text{Var}[\vec{\sigma}_{\text{exp}}] = \mathbf{S}_{\text{exp}}\mathbf{A}_0\mathbf{S}_{\text{exp}}^T + \mathbf{T}_{\text{exp}}\mathbf{K}_0\mathbf{T}_{\text{exp}}^T. \quad (9.26)$$

Therefore, the prior joint multivariate normal distribution of the random vector $\vec{x}^T = (\vec{\sigma}_{\text{exp}}^T, \vec{p}^T, \vec{\varepsilon}_{\text{mod}}^T)$ is characterized by the mean vector

$$\vec{x}_0 = \begin{pmatrix} \mathbf{T}_{\text{exp}}\vec{\Delta} + \mathbf{S}_{\text{exp}}\vec{p}_0 \\ \vec{p}_0 \\ \vec{0} \end{pmatrix}, \quad (9.27)$$

and the covariance matrix

$$\Sigma_0 = \begin{pmatrix} (\mathbf{S}_{\text{exp}}\mathbf{A}_0\mathbf{S}_{\text{exp}}^T + \mathbf{T}_{\text{exp}}\mathbf{K}_0\mathbf{T}_{\text{exp}}^T + \mathbf{B}) & \mathbf{S}_{\text{exp}}\mathbf{A}_0 & \mathbf{T}_{\text{exp}}\mathbf{K}_0 \\ \mathbf{A}_0\mathbf{S}_{\text{exp}}^T & \mathbf{A}_0 & \mathbf{0} \\ \mathbf{K}_0\mathbf{T}_{\text{exp}}^T & \mathbf{0} & \mathbf{K}_0 \end{pmatrix}. \quad (9.28)$$

Before taking into account experimental data, the variable $\vec{\sigma}_{\text{exp}}$ has to be considered as a random variable. If we perform a measurement, we obtain a realization

$\tilde{\sigma}_{\text{exp}}$ of the random variable $\vec{\sigma}_{\text{exp}}$. We can then derive the Bayesian update formulas by conditioning the random variables \vec{p} and $\vec{\varepsilon}_{\text{mod}}$ on the known first component of the joint multivariate normal distribution. For convenience, we introduce the abbreviations

$$\vec{\alpha} = (\tilde{\sigma}_{\text{exp}} - \mathbf{T}_{\text{exp}}\vec{\Delta} - \mathbf{S}_{\text{exp}}\vec{p}_0), \quad (9.29)$$

$$\mathbf{X} = (\mathbf{S}_{\text{exp}}\mathbf{A}_0\mathbf{S}_{\text{exp}}^T + \mathbf{T}_{\text{exp}}\mathbf{K}_0\mathbf{T}_{\text{exp}}^T + \mathbf{B})^{-1}. \quad (9.30)$$

The application of Equation 2.44 yields the posterior mean vectors

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}_0\mathbf{S}_{\text{exp}}^T\mathbf{X}\vec{\alpha}, \quad (9.31)$$

$$\vec{\varepsilon}_1 = \mathbf{K}_0\mathbf{T}_{\text{exp}}^T\mathbf{X}\vec{\alpha}. \quad (9.32)$$

These posterior mean vectors can be transferred to the mesh $\mathcal{I}_{\text{pred}}$ which contains the incident energies of interest,

$$\vec{\sigma}_{\text{pred}} = \mathbf{T}_{\text{pred}}\vec{\Delta} + \mathbf{S}_{\text{pred}}\vec{p}_1 \quad (9.33)$$

$$\vec{\varepsilon}_{\text{pred}} = \mathbf{T}_{\text{pred}}\vec{\varepsilon}_{\text{mod}} \quad (9.34)$$

$$\vec{\sigma}_{\text{true}} = \vec{\sigma}_{\text{pred}} + \vec{\varepsilon}_{\text{pred}}. \quad (9.35)$$

The cross sections in $\vec{\sigma}_{\text{true}}$ would enter nuclear data files as best estimates. A novelty of this approach is that we can also estimate the magnitude of model deficiency at every incident energy of interest. This feature is definitely an improvement compared to assessing the model quality by its χ^2 -value. As a side remark, high values of the model defect at some incident energy do not necessarily indicate the deficiency of the nuclear model. It could also indicate that inconsistent experimental data were included in the update procedure.

The derivation of the posterior covariance matrices can be performed with Equation 2.45 and yields

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0\mathbf{S}_{\text{exp}}^T\mathbf{X}\mathbf{S}_{\text{exp}}\mathbf{A}_0, \quad (9.36)$$

$$\mathbf{K}_1 = \mathbf{K}_0 - \mathbf{K}_0\mathbf{T}_{\text{exp}}^T\mathbf{X}\mathbf{T}_{\text{exp}}\mathbf{K}_0. \quad (9.37)$$

For the derivation of the posterior covariance matrix \mathbf{A}_{true} associated with $\vec{\sigma}_{\text{true}}$, we consider the random variable

$$\vec{u} = \mathbf{S}_{\text{comb}}\vec{p} + \vec{\varepsilon}_{\text{mod}}. \quad (9.38)$$

Its prior covariance matrix is given by

$$\mathbf{U}_0 = \text{Var}[\vec{u}] = \mathbf{S}_{\text{comb}}\mathbf{A}_0\mathbf{S}_{\text{comb}}^T + \mathbf{K}_0, \quad (9.39)$$

and the prior covariance matrix between $\vec{\sigma}_{\text{exp}}$ and \vec{u} by

$$\mathbf{C}_0 = \text{Cov}[\vec{u}, \vec{\sigma}_{\text{exp}}] = \mathbf{S}_{\text{comb}} \mathbf{A}_0 \mathbf{S}_{\text{exp}}^T + \mathbf{K}_0 \mathbf{T}_{\text{exp}}^T. \quad (9.40)$$

Using again Equation 2.45, we obtain

$$\mathbf{U}_1 = \mathbf{U}_0 - \mathbf{C}_0 \mathbf{X} \mathbf{C}_0^T. \quad (9.41)$$

All evaluated covariance matrices stated so far are defined on the mesh $\mathcal{I}_{\text{comb}}$. We can transfer them to the mesh $\mathcal{I}_{\text{pred}}$,

$$\mathbf{A}_{\text{pred}} = \mathbf{T}_{\text{pred}} \mathbf{A}_1 \mathbf{T}_{\text{pred}}^T, \quad (9.42)$$

$$\mathbf{K}_{\text{pred}} = \mathbf{T}_{\text{pred}} \mathbf{K}_1 \mathbf{T}_{\text{pred}}^T, \quad (9.43)$$

$$\mathbf{A}_{\text{true}} = \mathbf{T}_{\text{pred}} \tilde{\mathbf{A}}_1 \mathbf{T}_{\text{pred}}^T. \quad (9.44)$$

Finally, the covariances between $\vec{\sigma}_{\text{pred}}$ and $\vec{\varepsilon}_{\text{pred}}$ are of particular interest for the development or refinement of nuclear models, because they reflect the systematics which the nuclear model is unable to describe. Using the linearity of the variance operator (see Equation 2.29),

$$\text{Var}[\vec{\sigma}_{\text{true}}] = \text{Var}[\vec{\sigma}_{\text{pred}}] + \text{Var}[\vec{\varepsilon}_{\text{pred}}] + 2\text{Cov}[\vec{\sigma}_{\text{pred}}, \vec{\varepsilon}_{\text{pred}}], \quad (9.45)$$

we obtain

$$\mathbf{D}_{\text{pred}} = \text{Cov}[\vec{\sigma}_{\text{pred}}, \vec{\varepsilon}_{\text{pred}}] = \frac{1}{2} (\mathbf{A}_{\text{true}} - \mathbf{A}_{\text{pred}} - \mathbf{K}_{\text{pred}}). \quad (9.46)$$

9.4 Marginal likelihood maximization

The prior covariance matrix \mathbf{K}_0 of the model defect vector $\vec{\varepsilon}_{\text{mod}}$ is constructed with the covariance function $k(E, E')$ of a Gaussian process. The choice of the covariance function and specific values for its hyperparameters induce a probability distribution on a function space. For instance, the squared exponential covariance function introduced in Equation 9.5 depends on the amplitude δ and the length-scale λ . Although we gave some intuition in section 9.1 which considerations may help to specify the hyperparameters, the specification of the hyperparameters remains in many cases a difficult task. The maximization of the marginal likelihood is a statistical technique to automatically adjust the hyperparameters on the basis of experimental data.

The marginal likelihood, also called the evidence, represents the probability distribution for the measurement $\vec{\sigma}_{\text{exp}}$. Under the assumption of the statistical model, the marginal likelihood gives the information how likely it is to make a certain observation. In our case of a prior probability distribution for the model

parameters and another one for the model defect, the marginal likelihood is given by

$$\pi(\vec{\sigma}_{\text{exp}} | \delta, \lambda) = \int \ell(\vec{\sigma}_{\text{exp}} | \vec{p}, \vec{\varepsilon}_{\text{exp}}) \pi(\vec{p}) \pi(\vec{\varepsilon}_{\text{exp}} | \delta, \lambda) d\vec{p} d\vec{\varepsilon}_{\text{exp}}. \quad (9.47)$$

We made explicit reference to the dependence of the distribution on the hyperparameters by writing $\pi(\vec{\varepsilon}_{\text{exp}} | \delta, \lambda)$. Suppose that the result of a measurement is the vector $\vec{\sigma}_{\text{exp}}$. This measurement is associated with a certain probability density. The approach suggested in this thesis is to choose values for the hyperparameters δ and λ which maximize the marginal likelihood,

$$(\delta, \lambda) = \underset{\delta, \lambda \in \mathbb{R}^+}{\operatorname{argmax}} \rho(\vec{\sigma}_{\text{exp}} | \delta, \lambda). \quad (9.48)$$

This approach clearly violates the Bayesian principle that the prior should not be specified based on the data included in the update process. Hence, we leave the school of classical Bayesian statistics. However, in the following we argue why the consideration of experimental data to set the hyperparameters of the model defect distribution is reasonable. Generally, Bayesian methods which take into account the experimental data for adjusting the prior hyperparameters are known as Empirical Bayes methods, e.g. Casella (1985).

The principal objection against using the observations that enter the Bayesian inference for the specification of the prior is the double-counting of data. For instance, consider ten noisy observations x_1, x_2, \dots, x_{10} . Their mean value \hat{x} as a best guess for the true value is associated with the variance

$$\operatorname{Var}[\hat{x}] = \frac{1}{10^2} \sum_{i=1}^{10} \operatorname{Var}[x_i] = \frac{1}{10} \operatorname{Var}[x]. \quad (9.49)$$

If we just duplicate each observation x_i , denote the duplicate by \tilde{x}_i , and calculate the variance again, without taking into account the perfect correlations in the dataset, we obtain

$$\operatorname{Var}[\hat{x}] = \frac{1}{20^2} \sum_{i=1}^{10} (\operatorname{Var}[x_i] + \operatorname{Var}[\tilde{x}_i]) = \frac{1}{20} \operatorname{Var}[x]. \quad (9.50)$$

The variance is incorrectly reduced, suggesting that we have a more accurate guess for the true value than is the case. If we would perform a proper calculation and account for the perfect correlations, the result would be

$$\operatorname{Var}[\hat{x}] = \frac{1}{20^2} \sum_{i=1}^{10} (\operatorname{Var}[x_i] + \operatorname{Var}[\tilde{x}_i] + 2 \overbrace{\operatorname{Cov}[x_i, \tilde{x}_i]}^{=1}) = \frac{1}{10} \operatorname{Var}[x], \quad (9.51)$$

and hence nothing is changed. These considerations clearly show that including the data several times in the Bayesian inference procedure renders the reliability of the evaluated uncertainties void.

However, estimating the hyperparameters of the model defect is conceptually different. Compared to the Bayesian update procedure that does not take into account model defects, the maximization of the marginal likelihood always leads to larger evaluated uncertainties, and hence to more reliable estimates.

We want to explain the philosophy of maximizing the marginal likelihood with an illustrative example. Suppose that we ask a weatherman for a prediction of the weather of today and tomorrow. We give him the information about today's weather, such as the temperature and the cloudiness, which is available to us. Suppose that we give him the information that the sun is shining today and the temperature is 25°C. If we ask him then for the prediction of today's weather, and he answers us that the sky is completely clouded and the temperature is 15°C, we very likely would react in two ways. First, we would correct his prediction of today's weather to match better our observation, and second, we would have an increased distrust in his prediction of tomorrow's weather.

The situation in nuclear data evaluation is analogous. The prior distribution of the model parameters is updated according to experimental data. If the evaluated cross section curve is inconsistent with the experimental data, we would have an increased distrust in the reliability of the nuclear model. Consequently, we would increase the uncertainties at incident energies without experimental data and adjust the evaluated cross section curve to better match the available observations. Adjusting the hyperparameters of the probability distribution for the model defect function represents exactly this strategy.

After the discussion of this conceptual aspect, we want to derive the specific form of the marginal likelihood. The evaluation of the integral in Equation 9.47 is an involved task in general. Frequently, Monte Carlo methods are applied for its evaluation. However, in the present case due to the linearization of the nuclear model and the assumption of multivariate normal distribution for all variables, the integral can be solved analytically. Our starting point for a simple derivation of the solution is the statistical model

$$\vec{\sigma}_{\text{exp}} = \mathbf{T}_{\text{exp}}\vec{\Delta} + \mathbf{S}_{\text{exp}}\vec{p} + \mathbf{T}_{\text{exp}}\vec{\varepsilon}_{\text{mod}} + \vec{\varepsilon}_{\text{exp}}. \quad (9.52)$$

with the assignments

$$\vec{p} \sim \mathcal{N}(\vec{p}_0, \mathbf{A}_0), \vec{\varepsilon}_{\text{mod}} \sim \mathcal{N}(\vec{0}, \mathbf{K}_0), \text{ and } \vec{\varepsilon}_{\text{exp}} \sim \mathcal{N}(\vec{0}, \mathbf{B}). \quad (9.53)$$

The notation was explained at the beginning of section 9.3. Because all random variables on the right hand side of Equation 9.52 are multivariate normal, also the variable $\vec{\sigma}_{\text{exp}}$ follows a multivariate normal distribution. Therefore, it suffices to calculate its mean vector and covariance matrix. The mean vector is given by

$$\vec{m} = \mathbf{E}[\vec{\sigma}_{\text{exp}}] = \mathbf{T}_{\text{exp}}\vec{\Delta} + \mathbf{S}_{\text{exp}}\vec{p}_0, \quad (9.54)$$

and for the covariance matrix we obtain

$$\mathbf{H}_{(\delta,\lambda)} = \text{Var}[\vec{\sigma}_{\text{exp}}] = \mathbf{S}_{\text{exp}}\mathbf{A}_0\mathbf{S}_{\text{exp}}^T + \mathbf{T}_{\text{exp}}\mathbf{K}_0(\delta, \lambda)\mathbf{T}_{\text{exp}}^T + \mathbf{B}. \quad (9.55)$$

The probability density distribution of the marginal likelihood is thus $\pi(\vec{\sigma}_{\text{exp}}) = \mathcal{N}(\vec{m}, \mathbf{H})$ and its logarithm takes the form

$$\log \pi(\vec{\sigma}_{\text{exp}} | \delta, \lambda) = -\frac{1}{2} \log \det \mathbf{H}_{(\delta,\lambda)} - \frac{1}{2} (\vec{\sigma}_{\text{exp}} - \vec{m})^T \mathbf{H}_{(\delta,\lambda)}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{m}) - \frac{n}{2} \log(2\pi), \quad (9.56)$$

where n is the dimension of $\vec{\sigma}_{\text{exp}}$. We are interested to find the hyperparameters that maximize the marginal likelihood. For instance, increasing the hyperparameter δ leads to larger variances in the diagonal of \mathbf{K}_0 and hence the second term in Equation 9.56 becomes larger (less negative). Simultaneously, the determinant of $\mathbf{H}_{(\delta,\lambda)}$ gets larger and thereby the first term becomes smaller. The first term is besides an additive constant the negative information entropy (see Equation 2.61). Therefore, maximizing the marginal likelihood denotes a trade-off between entropy minimization and the maximization of the probability density to observe $\vec{\sigma}_{\text{exp}}$.

The search for the parameters that maximize Equation 9.56 can be performed with a gradient-based optimization algorithm such as L-BFGS-B (Byrd et al., 1995). The L-BFGS-B algorithm allows the restriction of parameter ranges in the optimization procedure. This feature is convenient because the hyperparameters δ and λ must be non-negative. Further, the restriction to certain intervals can be regarded as a measure to take into account prior knowledge.

The gradient of $\log \pi(\vec{\sigma}_{\text{exp}} | \delta, \lambda)$ with respect to the hyperparameters is analytically available. Using the matrix identities (e.g. Petersen et al. (2006))

$$\frac{\partial}{\partial \phi} \log \det \mathbf{H} = \text{Tr}(\mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \phi}) \quad \text{and} \quad \frac{\partial}{\partial \phi} \mathbf{H}^{-1} = -\mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \phi} \mathbf{H}^{-1} \quad (9.57)$$

we obtain

$$\begin{aligned} \frac{\partial}{\partial \phi} \log \pi(\vec{\sigma}_{\text{exp}} | \delta, \lambda) = & -\frac{1}{2} \text{Tr}(\mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \phi}) \\ & + \frac{1}{2} (\vec{\sigma}_{\text{exp}} - \vec{m})^T \mathbf{H}^{-1} \frac{\partial \mathbf{H}}{\partial \phi} \mathbf{H}^{-1} (\vec{\sigma}_{\text{exp}} - \vec{m}). \end{aligned} \quad (9.58)$$

In the first term only the diagonal elements of the matrix-matrix product have to be calculated for the evaluation of the trace. In the second term, the quantity $\vec{\alpha} = \mathbf{H}^{-1}(\vec{\sigma}_{\text{exp}} - \vec{m})$ has to be computed only once. The remaining calculation of $\vec{\alpha}^T (\partial \mathbf{H} / \partial \phi) \vec{\alpha}$ can be performed fast.

Finally, we remark that the maximization of the marginal likelihood is not necessarily in contradiction to classical Bayesian statistics. If experimental data of neighbor isotopes are available, the hyperparameters can be adjusted on the

basis of these data. The most conservative values of the hyperparameters which occurred for some isotope can be chosen for the current evaluation. The most conservative values are those which lead to maximal information entropy of the constructed model defect covariance matrix. In the case of the squared exponential covariance function (see Equation 9.5) and the Ornstein-Uhlenbeck process (see Equation 8.32), the information entropy increases with increasing amplitude δ and decreasing length-scale λ .

9.5 Showcase: deficient non-linear model

The analysis in chapter 6 showed that not accounting for model defects can have severe consequences for the reliability of an evaluation. The evaluated cross section curve may disagree with experimental data included in the Bayesian inference and might be also inconsistent with experimental data which had not been included. In this section we perform an evaluation of the neutron-induced total cross section of ^{181}Ta with the extended Bayesian update which accounts for model defects. The evaluation scenario was detailed in section 6.2. We briefly recapitulate the important aspects.

The complete experimental dataset \mathcal{C} from Finlay et al. (1993) in the energy range from 5 to 200 MeV is split into two datasets. Dataset \mathcal{A} contains the data points below 100 MeV and dataset \mathcal{B} the data points above 100 MeV. The experimental data are illustrated in Figure 6.14. The evaluation is performed with one of these datasets and the resulting cross section curve is compared to the datasets which had not been included in the update.

The optical model is employed and the variation of model parameters is restricted to a_v , r_v and v_1 . Instead of these parameters we use the dimensionless parameters $\tilde{a}_v = a_v/a_v^{\text{Def}}$, $\tilde{r}_v = r_v/r_v^{\text{Def}}$ and $\tilde{v}_1 = v_1/v_1^{\text{Def}}$, where the superscript Def denotes the choice according to the global parameterization of Koning and Delaroche (2003).

We will use the iterative linearized Bayesian update procedure outlined in section 6.3. Without the inclusion of model defects, the evaluated cross section curve was inconsistent with dataset \mathcal{B} if only the dataset \mathcal{A} was included in the Bayesian update (see Equation 6.24 and Figure 6.17). In this section we will show how the inclusion of model defects restores consistency. The inclusion of model defects requires the following modification of the iterative linearized Bayesian update approach.

Iterative linearized Bayesian update with model defects

1. Set the reference parameter vector \vec{p}_{ref} to the best prior estimate \vec{p}_0 .
2. Linearize the model at \vec{p}_{ref} (see Equation 9.21).
3. Find the hyperparameters δ and λ that maximize the marginal likelihood given in Equation 9.56.
4. Construct the covariance matrix \mathbf{K}_0 of the model defect with the covariance function using the values of the hyperparameters δ and λ obtained in the previous step.
5. Perform the Bayesian update with Equation 9.31 to obtain the posterior parameter vector \vec{p}_1 .
6. If \vec{p}_1 did not converge yet, choose \vec{p}_1 as reference parameter vector \vec{p}_{ref} and continue with step two.

In the following evaluation, the probability distribution of the model defect function is given by the Gaussian process $\varepsilon(E) \sim \mathcal{GP}(m(E), k(E, E'))$ with $m(E) \equiv 0$ and the squared exponential covariance function

$$k(E, E') = \sigma(E)\sigma(E')\delta^2 \exp\left(-\frac{1}{2\lambda^2}(\log E - \log E')^2\right). \quad (9.59)$$

with the best prior estimate of the cross section $\sigma(E)$, the magnitude of the model error δ and the length-scale λ . The prior distribution of the nuclear model parameters is given by $\vec{p} \sim \mathcal{N}(\vec{p}_0, \mathbf{A}_0)$ with

$$\vec{p}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_0 = \begin{pmatrix} 0.15^2 & 0 & 0 \\ 0 & 0.15^2 & 0 \\ 0 & 0 & 0.15^2 \end{pmatrix}. \quad (9.60)$$

The inclusion of the complete experimental dataset \mathcal{C} in the iterative Bayesian update yields

$$\vec{p}_1 = \begin{pmatrix} 1.0411 \\ 0.9832 \\ 1.0144 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-4} \begin{pmatrix} 15.0534 & -2.2677 & 2.2481 \\ -2.2677 & 0.6182 & -0.7722 \\ 2.2481 & -0.7722 & 1.8753 \end{pmatrix}. \quad (9.61)$$

The evaluated uncertainties of the model parameters are about 10^{-2} . For comparison, the evaluated uncertainties of the iterative Bayesian update without model defects were in the order of 10^{-3} (see Equation 6.20). The values of the hyperparameters found by maximizing the marginal likelihood are $\delta = 0.0117$ and

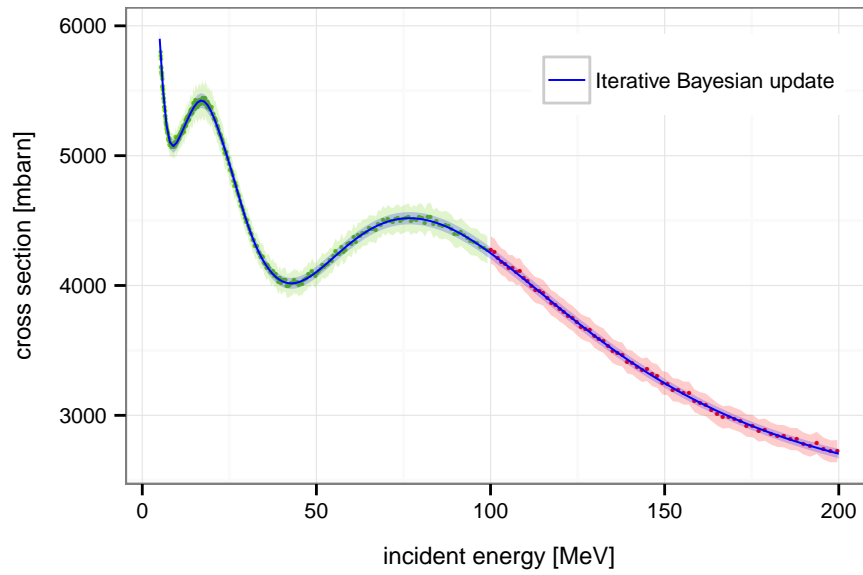


Figure 9.3: Evaluated cross section curve obtained by the iterative linearized Bayesian update procedure with model defects. The complete experimental dataset \mathcal{C} had been included in the update. Shown are also the 95% confidence bands.

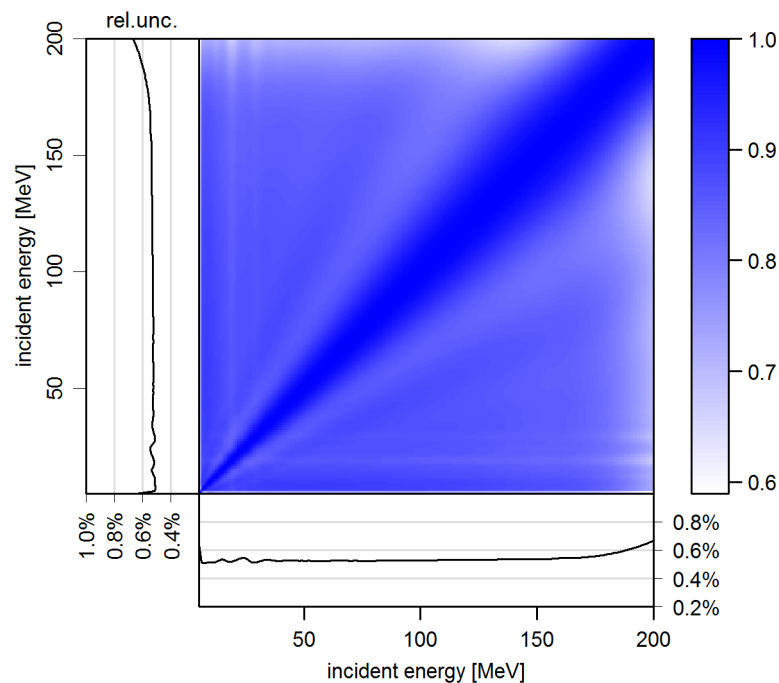


Figure 9.4: Evaluated correlations and uncertainties obtained by the iterative Bayesian update procedure with model defects. The complete experimental dataset \mathcal{C} had been included in the update.

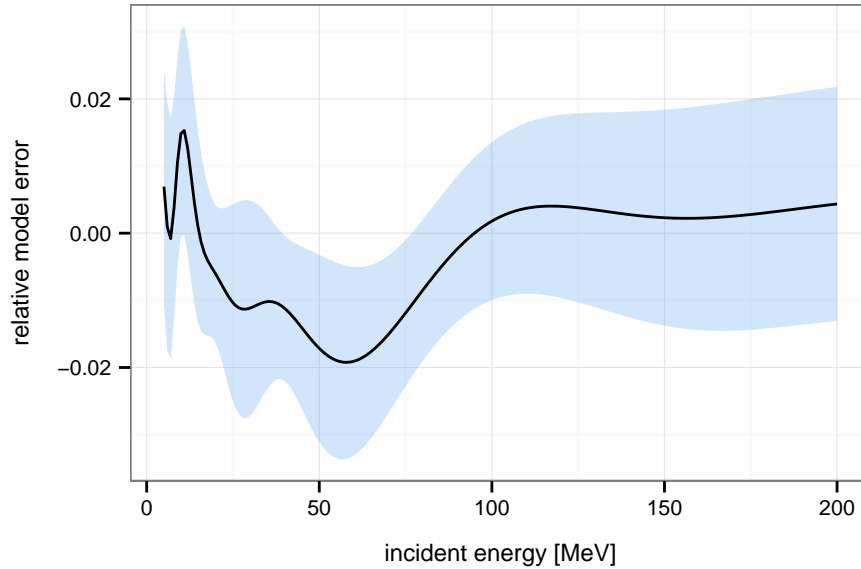


Figure 9.5: Evaluated model defect on a relative scale and the associated 95% confidence band obtained by the iterative linearized Bayesian update procedure. The complete dataset \mathcal{C} was included in the update.

$\lambda = 0.3051$. Hence, the optical model is able to fit the data with an accuracy of about 1% globally. Therefore, we can consider the optical model as a rather accurate and useful model. The evaluated cross sections computed with Equation 9.35 are shown in Figure 9.3. The evaluated cross section curve is smooth and matches the data perfectly. In contrast to that, the evaluated cross section curve obtained by the iterative Bayesian update procedure without model defects was good but deviated at some energies slightly from the experimental data (see Figure 6.15).

The evaluated correlations and uncertainties are shown in Figure 9.4. The inclusion of the model defect in the prior knowledge automatically weakens the evaluated correlations (compare for instance with Figure 6.15). Even though, the linearization of the nuclear model leads typically to stiff correlation matrices, the inclusion of the model defect causes a cancellation of model systematics. It is automatically detected that the systematics of the model is inappropriate at the level of accuracy given by the experimental data. The evaluated correlations appear uniform and reflect the normalization error of the experimental uncertainty. However, due to the finite length-scale of the covariance function, the correlations of cross sections at different incident energies significantly decrease with an increasing difference of the incident energies. For comparison, the cross sections obtained by the FBET/EMPIRE-MC method based on the surrogate approach without model defects are stronger correlated (see Figure 6.25).

The evaluated model defect is shown in Figure 9.5. We recognize again that the optical model is an accurate and useful model. The hypothesis of a perfect

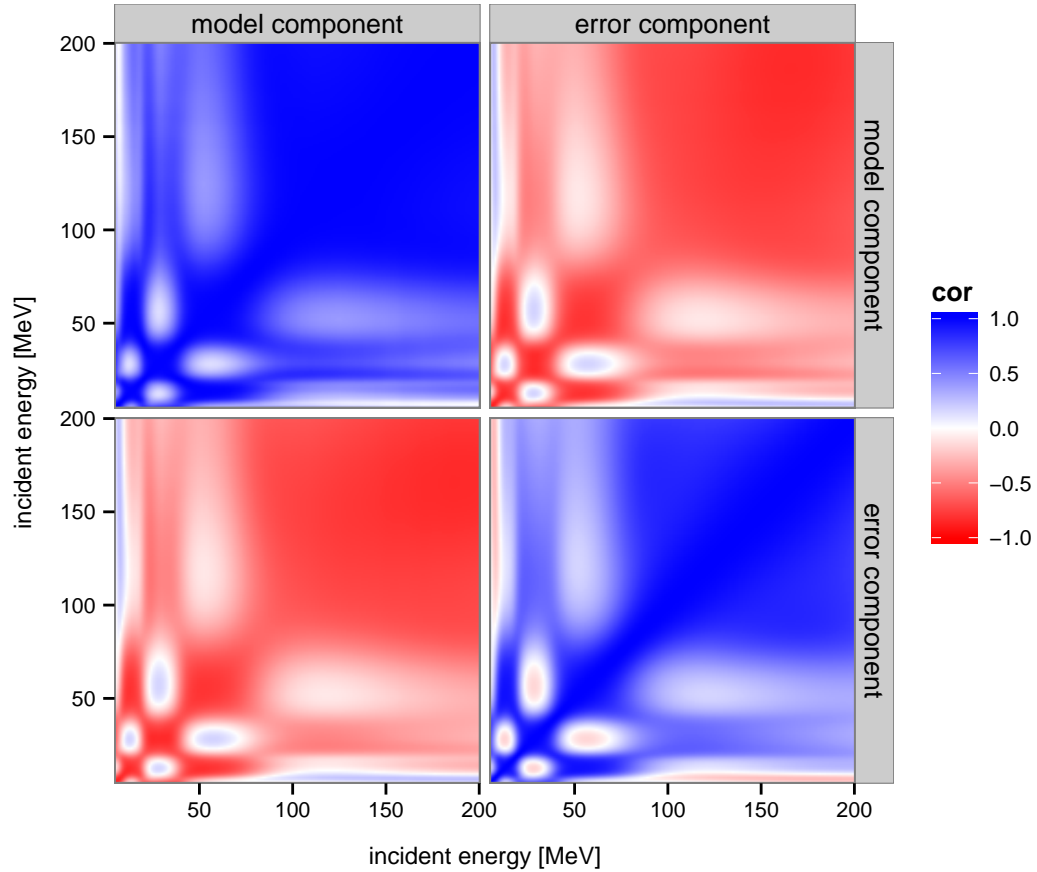


Figure 9.6: Evaluated correlations of the model prediction and the model defect. Also the correlations between the model prediction and the model defect are shown.

model cannot be rejected at most incident energies, because the 95% confidence band encloses the zero-line. However, at around 10 MeV and 50 MeV the model is clearly deficient. The deficiency in these regions is estimated to be slightly below 2%. These deficiencies were also the reason, why evaluated cross section curves of methods using the exact model such as the UMC-B method became inconsistent with experimental data. The evaluated correlations of the model prediction and the model defect are displayed in Figure 9.6. The correlations of the model prediction and those of the model defect are of similar structure. Model prediction and model defect are negatively correlated, and thus the model defect weakens the correlations of the model prediction.

Next, we perform the iterative linearized Bayesian update procedure with model defects including only dataset \mathcal{A} . We obtain the evaluated parameter mean vector

and covariance matrix,

$$\vec{p}_1 = \begin{pmatrix} 1.0737 \\ 0.9735 \\ 1.0242 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-4} \begin{pmatrix} 21.7842 & -4.3239 & 4.6015 \\ -4.3239 & 1.2633 & -1.5280 \\ 4.6015 & -1.5280 & 2.8174 \end{pmatrix}. \quad (9.62)$$

The best hyperparameters $\delta = 0.012$ and $\lambda = 0.299$ are approximately equal to those obtained by updating with the complete dataset \mathcal{C} . The evaluated cross section curve is illustrated in Figure 9.7. The cross section curve perfectly agrees with included experimental data and gives a reasonable prediction in the range of dataset \mathcal{B} . In contrast to that, the evaluation with the iterative Bayesian update procedure without model defects was inconsistent with dataset \mathcal{B} (see Figure 6.17). The evaluated uncertainty and correlations are shown in Figure 9.9. The evaluated uncertainty is about 0.5% in the energy range from 5 to 100 MeV from which data were included in the update. In the energy range above 100 MeV where data had not been included, the uncertainty gradually increases to about 1.8%. In contrast to the evaluation without model defects (see Figure 6.18) the cross sections above 100 MeV are only weakly correlated to those below 100 MeV. The estimate of the model defect is shown in Figure 9.9. Finally, in Figure 9.7 the model component of the evaluated cross section curve is shown which deviates slightly from the experimental data due to the model deficiency.

As a last benchmark, we perform the update with dataset \mathcal{B} . The evaluated parameter mean vector and covariance matrix are given by

$$\vec{p}_1 = \begin{pmatrix} 1.0757 \\ 0.9818 \\ 1.0134 \end{pmatrix} \quad \text{and} \quad \mathbf{A}_1 = 10^{-4} \begin{pmatrix} 151.7837 & -12.2696 & -6.6565 \\ -12.2696 & 1.1863 & -0.2601 \\ -6.6565 & -0.2601 & 5.2059 \end{pmatrix}. \quad (9.63)$$

The evaluated cross section curve is shown in Figure 9.11. The maximization of the marginal likelihood yielded the hyperparameters $\delta = 0.0001$ and $\lambda = 9.237$. The model defect is virtually switched off because of the small value of δ . This is due to the fact that the model is deficient at low incident energies. Because only the dataset \mathcal{B} associated with energies above 100 MeV was included, the model deficiency could not be inferred from the data.

The results of this section indicate that the problems of available evaluation methods studied in chapter 6 are resolved:

1. The evaluated cross section curve mimics well the experimental included in the update process and associated uncertainties reflect the experimental uncertainties.

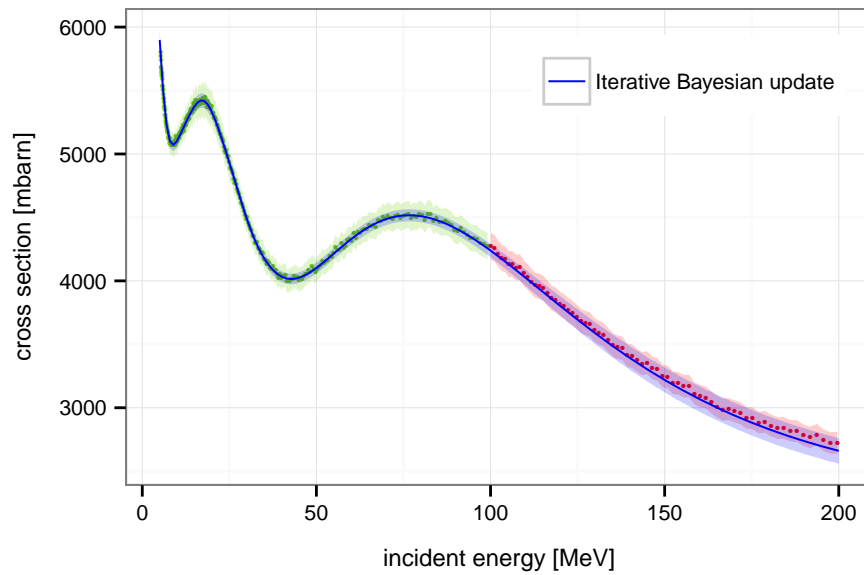


Figure 9.7: Evaluated cross section curve and 95% confidence band obtained by the iterative linearized Bayesian update procedure with model defects. The complete experimental dataset \mathcal{A} was included in the update.

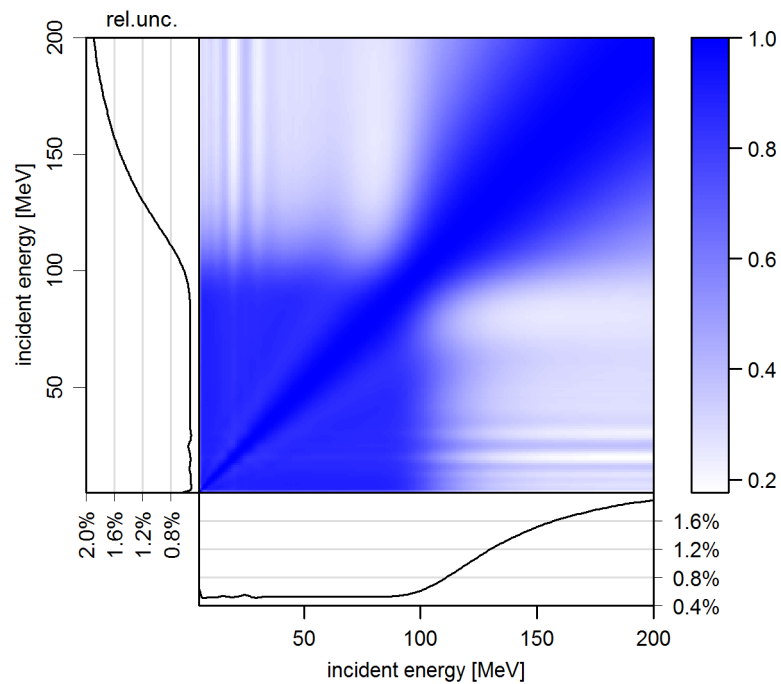


Figure 9.8: Evaluated correlations and uncertainties obtained by the iterative Bayesian update procedure with model defects. The complete experimental dataset \mathcal{A} was included in the update.

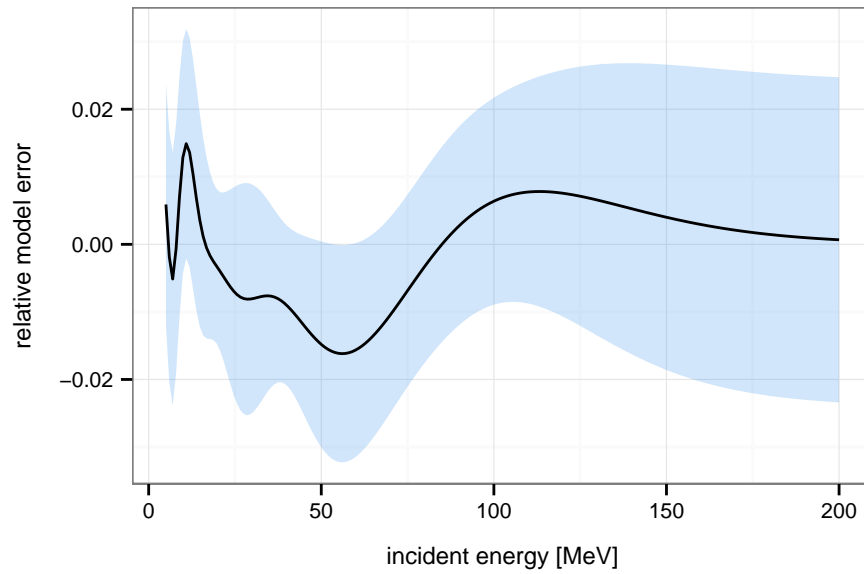


Figure 9.9: Evaluated model defect and associated 95% confidence band on a relative scale obtained by the iterative linearized Bayesian update procedure. Only dataset \mathcal{A} has been included in the update.

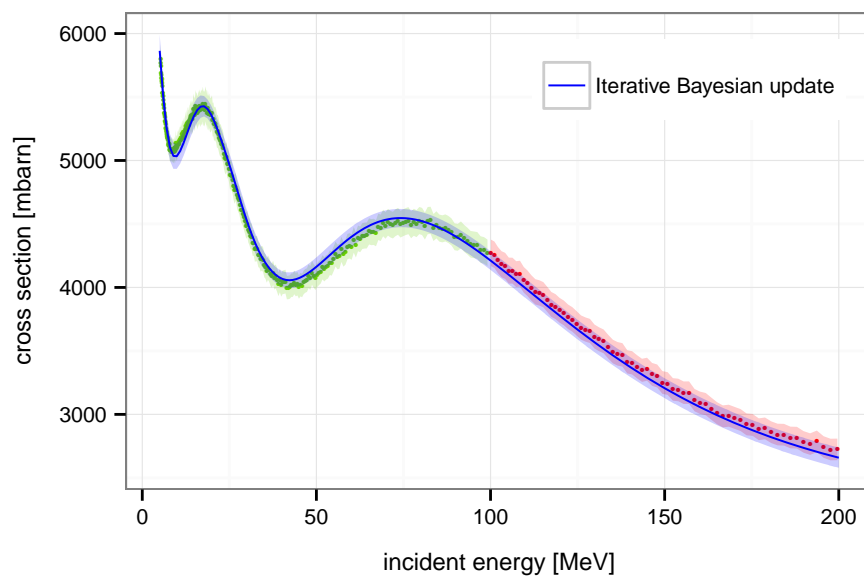


Figure 9.10: Evaluated model component and associated 95% confidence band obtained by the iterative linearized Bayesian update procedure with model defects. Only dataset \mathcal{A} has been included in the update.

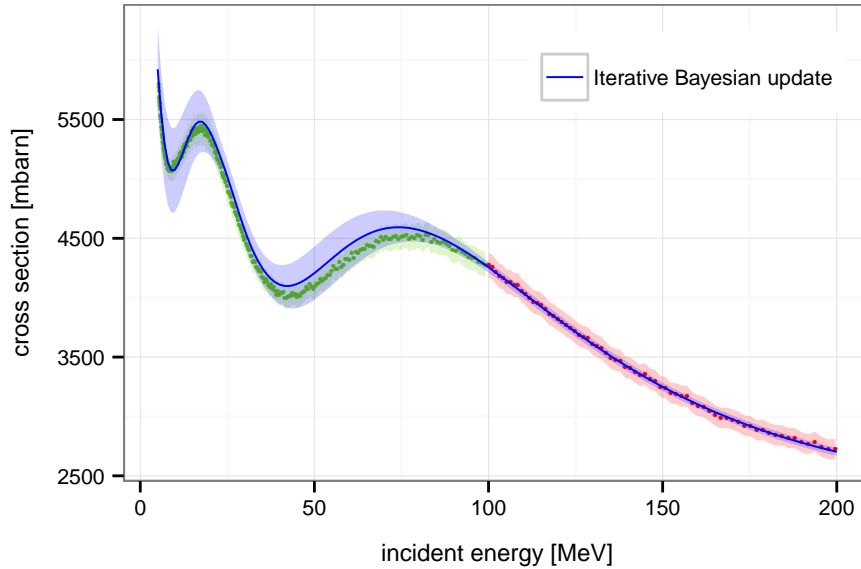


Figure 9.11: Evaluated cross section curve and associated 95% confidence band obtained by the iterative linearized Bayesian update procedure with model defects. Only dataset \mathcal{B} was included in the update.

2. The evaluated cross section curve is consistent with experimental data which were not included in the evaluation procedure and associated uncertainties are increased due to the deficiency of the nuclear model.
3. The evaluated cross section curve is a smooth function. It follows the trend of the experimental data but does not follow every local fluctuation.
4. The evaluated uncertainties are generated directly from the evaluation procedure. No ad-hoc rescaling with the χ^2 -value due to too small uncertainties or a mismatch between experimental data and evaluated cross section curve is necessary.

The new approach to deal with model defects combines the advantages of no-model fit techniques, such as smoothing splines, and model fits without their disadvantages. No-model fits yield an excellent reproduction of the included experimental data, yet they have no predictive power beyond the energy range of included data. In contrast to that, a model fit can predict cross sections in energy domains without data, but due to imperfections of the model might deviate significantly from included experimental data.

Within this special evaluation scenario, the new approach seems to be the jack of all trades. However, also the new approach is based on assumptions and if these are violated, we still can obtain unsatisfactory evaluations. The assumptions are reflected in the choice of the covariance function in Equation 9.59:

1. The relative model defect is expected to be roughly the same at every incident energy.
2. The expected fluctuations of the model defect decrease with increasing incident energy
3. The model defect is a smooth function with derivatives of any order.

For instance, the violation of assumption (1) leads to overly cautious values of δ and λ obtained by the maximization of the marginal likelihood in order to be consistent with the data in the energy domain of most severe model deficiency. However, other choices of the covariance function are possible which imply different assumptions.

9.6 Sum rule conserving model defects

The introduction of a model defect function in the Bayesian update allows for deviations of the evaluated cross section curves from the systematics of the model. On one hand, this extended flexibility helps to overcome the deficiencies of the model, but on the other hand consistency constraints such as the sum rules of cross sections may be violated. For instance, at the same incident energy E the total cross section is always given as the sum of elastic and non-elastic cross section,

$$\sigma_{n,\text{tot}}(E) = \sigma_{n,\text{el}}(E) + \sigma_{n,\text{non-el}}(E). \quad (9.64)$$

As another example, the neutron production cross section is given by

$$\sigma_{n\text{-prod}}(E) = \sigma_{n,n}(E) + 2\sigma_{n,2n}(E) + 2\sigma_{n,2np}(E) + \dots \quad (9.65)$$

In general, some cross sections are defined as linear combinations of other cross sections,

$$\sigma(E) = \sum_i \alpha_i \sigma_i(E), \quad (9.66)$$

and evaluated cross sections must obey these relations. From now on we refer to relationships between cross sections of the form in Equation 9.66 as sum rules. In this section we extend the formulation of model defects to ensure their fulfillment.

The essential key for this extension is the fact that a multivariate normal distribution is able to perfectly capture linear relationships between observables. Consequently, also the Bayesian update preserves these relationships. Due to the importance of this feature for the construction of consistent model defects, we give a proof here.

Considering the relation in Equation 9.66, the variance of $\sigma(E)$ is given by

$$\text{Var}[\sigma(E)] = \sum_{i,j} \alpha_i \alpha_j \text{Cov}[\sigma_i(E), \sigma_j(E)]. \quad (9.67)$$

Covariances between $\sigma(E)$ and the cross sections on the r.h.s of Equation 9.66 are given by

$$\text{Cov}[\sigma(E), \sigma_i(E)] = \sum_j \alpha_j \text{Cov}[\sigma(E), \sigma_j(E)]. \quad (9.68)$$

Using the abbreviation $c_{ij} = \text{Cov}[\sigma_i(E), \sigma_j(E)]$, the covariance matrix \mathbf{A} which contains the variances and covariances of the cross section σ and the cross sections σ_i can be written as

$$\mathbf{A} = \begin{pmatrix} \sum_{ij} \alpha_i \alpha_j c_{ij} & \sum_j \alpha_j c_{1j} & \sum_j \alpha_j c_{2j} & \dots \\ \sum_j \alpha_j c_{1j} & c_{11} & c_{12} & \dots \\ \sum_j \alpha_j c_{2j} & c_{12} & c_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (9.69)$$

The first row is a linear combination of the other rows, hence the rank of this covariance matrix is reduced by one. We can immediately identify the (not normalized) eigenvector

$$\vec{e}_0 = (1, -\alpha_1, -\alpha_2, \dots)^T, \quad (9.70)$$

which is associated with the eigenvalue $\lambda_0 = 0$. Because a covariance matrix is symmetric, its eigenvectors are pairwise orthogonal, $\vec{e}_i^T \vec{e}_j = 0, i \neq j$. Due to the definition of orthogonality, the eigenvectors orthogonal to \vec{e}_0 have to obey the relation

$$(x_0, x_1, x_2, \dots) \vec{e}_0 = x_0 - \alpha_1 x_1 - \alpha_2 x_2 - \dots = 0, \quad (9.71)$$

which is exactly the sum rule that caused the reduced rank of the covariance matrix. Therefore, the components of all eigenvectors except \vec{e}_0 are in agreement with the sum rule. We can write the covariance matrix in its spectral decomposition

$$\mathbf{A} = \sum_i \lambda_i \vec{e}_i \vec{e}_i^T, \quad (9.72)$$

and insert it into the linearized Bayesian update formula (see e.g. Equation 3.34)

$$\vec{\sigma}_1 = \vec{\sigma}_0 + \mathbf{A}_0 \vec{u} = \vec{\sigma}_0 + \sum_i \lambda_i \vec{e}_i (\vec{e}_i^T \vec{u}). \quad (9.73)$$

The only contribution in the sum associated with \vec{e}_0 that could lead to a violation of the sum rule vanishes because of the zero eigenvalue. Therefore, sum rules are perfectly preserved in the linearized Bayesian update formula.

For the construction of a consistent (=sum rule preserving) model defect, we have to construct covariance matrix elements between channels in such a way

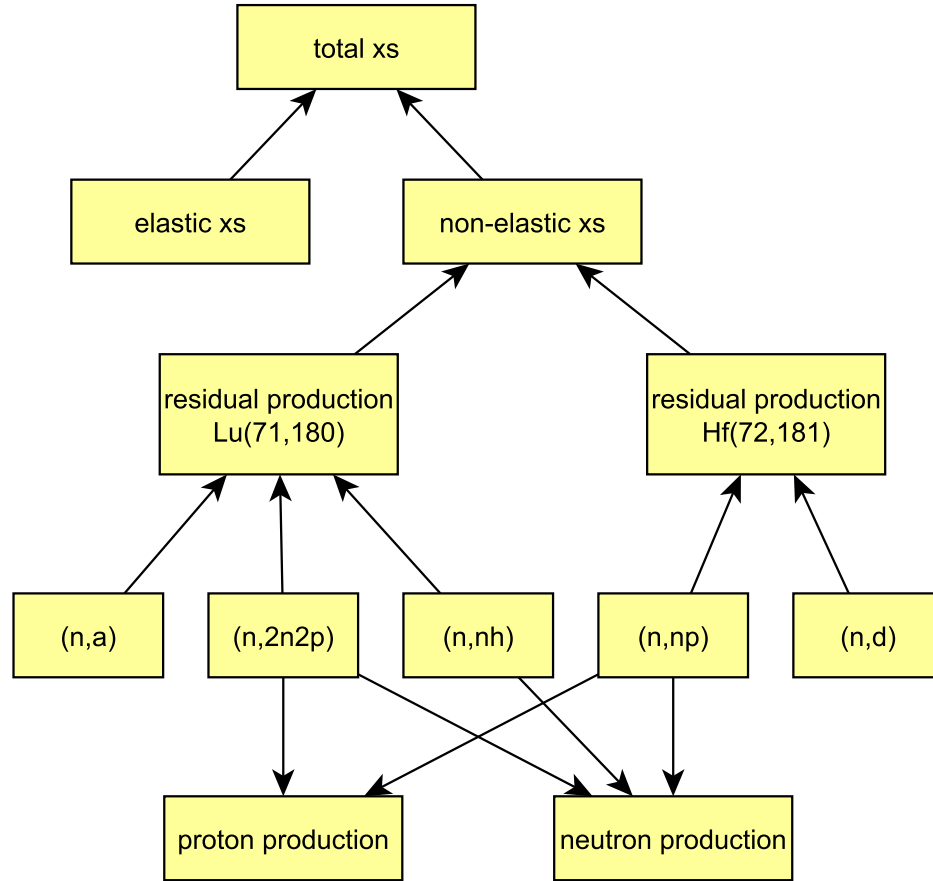


Figure 9.12: Some relationships of neutron-induced cross sections of ^{181}Ta visualized as a graph. (a=alpha, n=neutron, p=proton, h=Helium-3, d=deuteron).

that Equation 9.67 and Equation 9.68 are satisfied whenever a sum rule of the form Equation 9.66 exists. In practice, there are many cross sections of interest which are defined as linear combinations of other cross sections such as residual production cross sections. Figure 9.12 illustrates some of these relationships for neutron-induced cross sections of ^{181}Ta .

We restrict the further discussion to cross sections of different channels at the same incident energy. The generalization to different energies will be discussed afterward. The following discussion requires an extended notation. Especially, we denote by $\sigma(c)$ the cross section associated with reaction channel c . If a cross section $\sigma(c)$ is defined as a linear combination of other cross sections $\sigma(c')$, we denote by $\alpha(c|c')$ the multiplication constant associated with $\sigma(c')$ in the linear combination. For instance, the total cross section can be written as

$$\begin{aligned} \sigma(\text{tot}) &= \alpha(\text{tot}|\text{el})\sigma(\text{el}) + \alpha(\text{tot}|\text{nonel})\sigma(\text{nonel}), \\ &\text{with } \alpha(\text{tot}|\text{el}) = \alpha(\text{tot}|\text{nonel}) = 1. \end{aligned} \quad (9.74)$$

Another possibility is to express the total cross section as sum of *elementary cross*

sections

$$\begin{aligned} \sigma(\text{tot}) = & \alpha(\text{tot}|\text{el})\sigma(\text{el}) + \alpha(\text{tot}|\text{n,inel})\sigma(\text{n,inel}) + \\ & \alpha(\text{tot}|\text{n,2n})\sigma(\text{n,2n}) + \alpha(\text{tot}|\text{n,a})\sigma(\text{n,a}) + \dots \end{aligned} \quad (9.75)$$

In principle, we could have omitted the multipliers $\alpha(.,.)$ because they are one in this example, but we want to keep the discussion general and in other cases such as production cross sections they are not necessarily one. With elementary cross sections we denote cross sections that are the building blocks of other cross sections. These elementary cross sections fulfill two criteria:

1. An elementary cross section cannot be expressed as a linear combination of other elementary cross sections.
2. All non-elementary cross sections can be expressed as linear combinations of elementary cross sections, $\sigma = \sum_i \alpha_i \sigma_i$, with non-negative multipliers α_i .

Figure 9.13 schematically illustrates elementary and non-elementary cross sections. If we start from an elementary cross section, following a path along the directions of the arrows and encounter in this order the cross sections $\sigma(c_1), \sigma(c_2), \sigma(c_3)$, their multipliers satisfy the relation

$$\alpha(c_3|c_1) = \alpha(c_3|c_2)\alpha(c_2|c_1). \quad (9.76)$$

In order to construct consistent model defects, we make the following fundamental assumption:

Apriori independence of model defects

Model defects of elementary cross sections associated with different elementary reaction channels are assumed to be independent of each other. This choice is in agreement with the principle of maximum entropy. If there is no knowledge about the form of the dependence, assuming independence leads to a maximum of the information entropy.

Independence between reaction channels means zero-correlations between them. A covariance matrix with given diagonal elements has maximal information entropy if all off-diagonal covariance elements are zero (see Equation 3.26). We remark that the assumption of a-priori independence enters the prior distribution, but an update with experimental data will in general introduce correlations between model defects of different channels.

For the following discussion we introduce the index set \mathcal{E} which contains all elementary cross sections. Further, we introduce the concept of *common elementary*

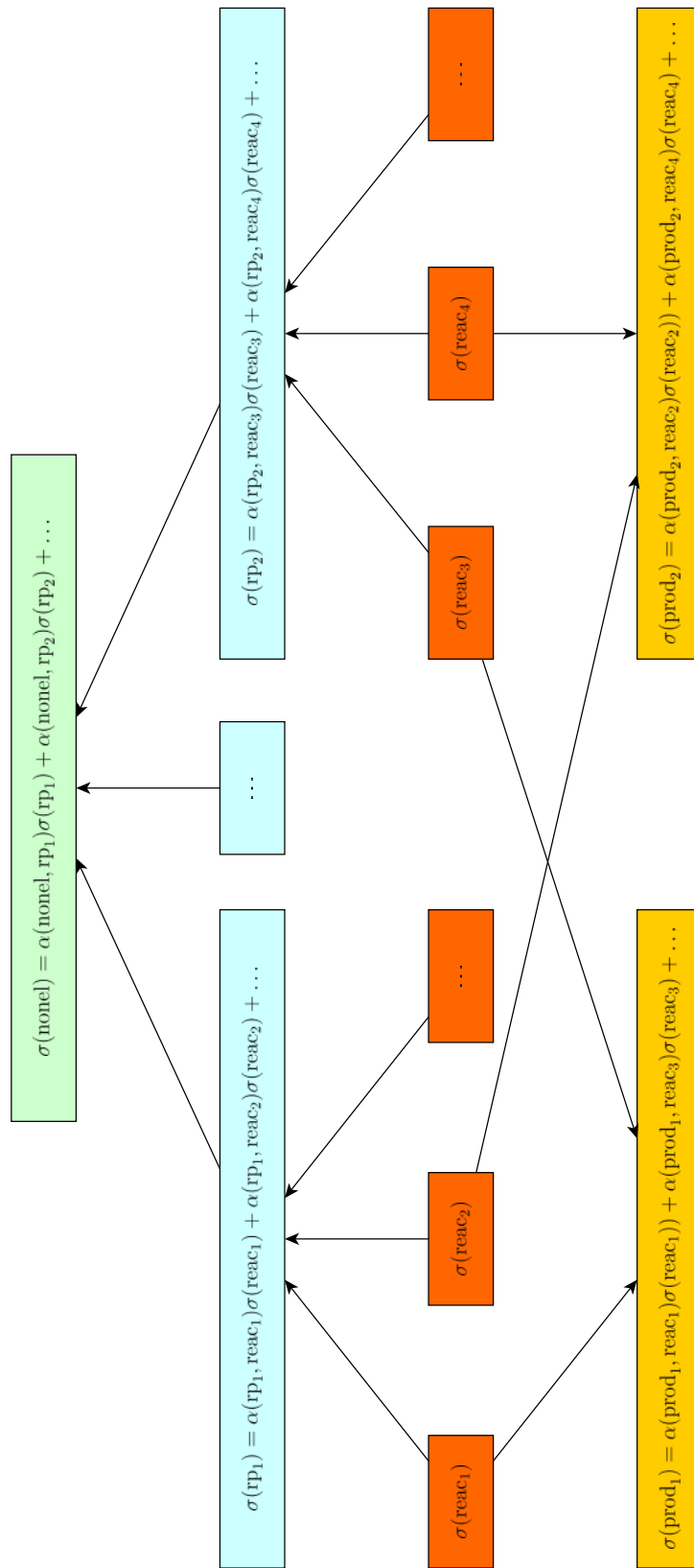


Figure 9.13: Dependencies between cross sections. The elementary cross sections $\sigma(\text{reac}_1), \sigma(\text{reac}_2), \dots$ are the building blocks for other non-elementary cross sections.

cross section. Suppose that we have two cross sections given as linear combinations of elementary cross sections,

$$\sigma(c_1) = \sum_{c' \in \mathcal{E}} \alpha(c_1|c')\sigma(c') \quad \text{and} \quad \sigma(c_2) = \sum_{c' \in \mathcal{E}} \alpha(c_2|c')\sigma(c'). \quad (9.77)$$

We call an elementary cross section which occurs both in the linear combination of $\sigma(c_1)$ and $\sigma(c_2)$ a common elementary cross section. We denote the set of all common elementary cross sections of $\sigma(c_1)$ and $\sigma(c_2)$ as $\text{CEC}(c_1, c_2)$. Noteworthy, $\text{CEC}(c, c') = \{c'\}$ if $\sigma(c')$ is an elementary cross section and contributes in the linear combination of $\sigma(c)$; and $\text{CEC}(c_1, c_2) = \emptyset$ if both $\sigma(c_1)$ and $\sigma(c_2)$ are elementary cross sections and not identical. Further, we denote by $\varepsilon(c)$ the model error associated with the cross section $\sigma(c)$.

The assumption of a priori independence between model defects associated with elementary cross sections greatly simplifies the task to construct consistent model defects. The constraint

$$\sigma(c) = \sum_{c' \in \mathcal{E}} \alpha(c|c')\sigma(c'). \quad (9.78)$$

induces the following constraint on the variance,

$$\text{Var}[\sigma(c)] = \sum_{c'_1 \in \mathcal{E}} \sum_{c'_2 \in \mathcal{E}} \alpha(c|c'_1)\alpha(c|c'_2)\text{Cov}[\sigma(c'_1), \sigma(c'_2)], \quad (9.79)$$

which in case of the model defects due to the independence assumption, $\text{Cov}[\varepsilon(c'_1), \varepsilon(c'_2)] = 0, \forall c'_1, c'_2 \in \mathcal{E}$, reduces to

$$\text{Var}[\varepsilon(c)] = \sum_{c' \in \mathcal{E}} \alpha(c|c')^2 \text{Var}[\varepsilon(c')]. \quad (9.80)$$

Similarly, the induced constraint on the covariance between a non-elementary cross section $\sigma(c)$ and an elementary cross section $\sigma(c')$,

$$\text{Cov}[\sigma(c), \sigma(c')] = \sum_{c'' \in \mathcal{E}} \alpha(c|c'')\text{Cov}[c'', c'], \quad (9.81)$$

simplifies for the model defect to

$$\text{Cov}[\varepsilon(c), \varepsilon(c')] = \alpha(c|c')\text{Var}[\varepsilon(c')]. \quad (9.82)$$

Covariances between two model defects associated with arbitrary cross sections $\sigma(c_1)$ and $\sigma(c_2)$ are given by

$$\text{Cov}[\varepsilon(c_1), \varepsilon(c_2)] = \sum_{c' \in \text{CEC}(c_1, c_2)} \alpha(c_1|c')\alpha(c_2|c')\text{Var}[\varepsilon(c')]. \quad (9.83)$$

We note that Equation 9.80 and Equation 9.82 are special cases of Equation 9.83. If we define $\sum_{c' \in \text{CEC}(c_1, c_2)} \dots = 0$ for $\text{CEC}(c_1, c_2) = \emptyset$, and $\alpha(c|c) = 1$, then Equation 9.83 can be used to calculate variances and covariances of and between model defects associated with arbitrary—elementary or non-elementary—cross sections.

We emphasize the important result that variances and covariances of model defects associated with arbitrary cross sections are completely determined by the variances of the model defects associated with elementary cross sections and the multipliers in the definitions of the derived cross sections. If the model defect covariance matrix satisfies Equation 9.83, and additionally model defects associated with elementary cross sections are uncorrelated, the fulfillment of sum rules is guaranteed.

9.7 Assignment of model defects

We extend the discussion of the previous section to cross sections in different channels and at different incident energies. Let $\sigma(c, E)$ denote the cross section of reaction channel c at incident energy E . The corresponding model defect is denoted by $\varepsilon(c, E)$. Model defects of all elementary cross section channels $c \in \mathcal{E}$ can be specified as Gaussian processes,

$$\varepsilon(c, E) \sim \mathcal{GP}(m_c(E), k_c(E, E')), \quad (9.84)$$

where $m_c(E) \equiv 0$. The same form of the covariance function can be taken for each channel or different choices can be made for different channels. The variance of a model error associated with an elementary cross section $\sigma(c, E)$ is given by $k_c(E, E)$. Therefore, according to Equation 9.83, covariances between different channels have to be specified as

$$k_{c_1, c_2}(E_1, E_2) = \sum_{c' \in \text{CEC}(c_1, c_2)} \alpha(c_1|c')\alpha(c_2|c')k_{c'}(E_1, E_2). \quad (9.85)$$

If $c_1 \neq c_2$ and $c_1, c_2 \in \mathcal{E}$ and thus $\text{CEC}(c_1, c_2) = \emptyset$, then $k_{c_1, c_2}(E_1, E_2) = 0$. A model defect covariance matrix \mathbf{K}_0 constructed with $k_{c_1, c_2}(E_1, E_2)$ is guaranteed to preserve sum rules.

Next, we discuss the assignment of model defects to cross section channels. Assume that we take the squared exponential covariance function which relies on two hyperparameters δ and λ . At high incident energies, hundred of elementary cross sections are open and the assignment of a covariance function to each of them means to introduce hundreds of hyperparameters δ_c, λ_c for each $c \in \mathcal{E}$. However, without experimental data and insufficient prior knowledge, there is no way to specify the values of the hyperparameters. In addition, it is often desirable to specify a model defect for a non-elementary cross section such as the total cross section which can be accurately measured. For these reasons, we introduce a scheme which allows to specify model defects for arbitrary elementary or non-elementary cross section channels. The model defect is distributed to elementary channels in order to enforce sum rules.

Depending on the incident energy a cross section may vary over several orders of magnitude. Therefore we define the covariance function $\tilde{k}_c(E, E')$ of every channel c on a relative scale,

$$k_c(E, E') = \sigma_M(c, E)\sigma_M(c, E')\tilde{k}_c(E, E'), \quad (9.86)$$

where $\sigma_M(c, E)$ is the prior best model prediction. For example, the relative covariance function $\tilde{k}_c(E, E')$ could be given as the squared exponential covariance function

$$\tilde{k}_c(E, E') = \delta_c \exp\left(-\frac{1}{2\lambda_c^2}(\log E - \log E')^2\right) \quad (9.87)$$

with hyperparameters δ_c, λ_c . If a relative covariance function $\tilde{k}_c(E, E')$ is defined for a non-elementary cross section channel, we construct the covariance matrix on an absolute scale as

$$k_c(E, E') = \sum_{c' \in \text{CEC}(c, c)} \alpha(c|c')^2 \sigma_M(c', E)\sigma_M(c', E')\tilde{k}_c(E, E'). \quad (9.88)$$

In order to preserve sum rules, we make for all elementary channels c' which contribute to c the choice

$$k_{c'}(E, E') = \sigma_M(c', E)\sigma_M(c', E')\tilde{k}_c(E, E') \text{ for all } c' \in \text{CEC}(c, c). \quad (9.89)$$

Hence, we assume the same functional form of the relative covariance function of the non-elementary channel also for the elementary channels. Only the hyperparameters associated with $\tilde{k}_c(E, E')$ are introduced and all elementary channels which contribute to c are associated with the same values of the hyperparameters. The choices in Equation 9.88 and Equation 9.89 lead to the preservation of sum rules.

An additional issue which has to be addressed is the question how to resolve contradictory assignments. For instance, the specification of a relative covariance function $\tilde{k}_{\text{tot}}(E, E')$ for the total cross section \tilde{k} leads via Equation 9.89 to the assignment of the same relative covariance function to all elementary cross sections. If in a second step a model defect is assigned to the non-elastic cross section channel, some assignments of covariance functions to elementary cross sections have to be changed accordingly. In order to accomplish this task, an assignment table can be used. As an example, it could be given by

elementary channel	reac ₁	reac ₂	reac ₃	reac ₄	...
assignment	(n,tot)	(n,tot)	-	(n,nonel)	...

This table indicates which relative covariance function is used for a particular elementary reaction channel. With $\mathcal{A}(c')$ we denote the assignment to the elementary

channel c' , for instance $\mathcal{A}(\text{reac}_1) = (\text{n,tot})$ in the current example. It is also possible that an elementary channel is not associated with a model defect (indicated by the dash).

If a new model defect is specified for a particular reaction channel c , the following rules determine the adjustments of the assignment table.

Model defect assignment rules

Suppose that a model defect is introduced for a channel c . The index set $\mathcal{E}(c) = \text{CEC}(c, c)$ contains all elementary reaction channels that contribute in the linear combination of $\sigma(c)$. The current assignment of a particular elementary reaction channel c' is denoted by $\mathcal{A}(c')$. The assignments $\mathcal{A}(c')$ of all elementary channels c' are altered according to the following rules:

1. If $\mathcal{E}(c) \subset \mathcal{E}(\mathcal{A}(c'))$ then change the assignment to $\mathcal{A}(c') = c$.
2. If $\mathcal{E}(\mathcal{A}(c')) \subset \mathcal{E}(c)$ then leave the assignment $\mathcal{A}(c')$ unaffected.
3. If neither the condition of (1) nor (2) is satisfied, but $\mathcal{E}(c) \cap \mathcal{E}(\mathcal{A}(c')) \neq \emptyset$ leave the assignment $\mathcal{A}(c')$ unaffected.

The idea behind rule (1) is that if a non-elementary cross section $\sigma(c)$ is defined as a linear combination of elementary cross sections which are a subset of the elementary cross sections in the linear combination of $\sigma(c')$, then $\sigma(c)$ is more directly related to these elementary cross sections than $\sigma(c')$. Therefore, it seems reasonable to chose the relative covariance function associated with $\sigma(c)$ also for these elementary cross sections. Rule (2) represents the opposite case and is stated just for the sake of completeness. Even though rule (3) is also superfluous in the current form, we stated it to emphasize the fact that some cross sections are not in a parent-child relation. For instance, the neutron production cross section and the residual production cross section are not in a parent-child relation but have common elementary cross sections. Using a first come-first serve principle for the condition in (3) is a pragmatic solution. However, the existence of a better decision rule whether to leave the assignment or to change it is very likely.

On the basis of the assignment table, Equation 9.85 and Equation 9.89, covariances associated with the model defect between arbitrary cross sections can be calculated by

$$k_{c_1, c_2}(E_1, E_2) = \sum_{c' \in \text{CEC}(c_1, c_2)} \alpha(c_1|c')\alpha(c_2|c')\sigma_M(c', E)\sigma_M(c', E')\tilde{k}_{\mathcal{A}(c')}(E, E'). \tag{9.90}$$

9.8 Showcase: conservation of sum rules

In this section we demonstrate the effect of sum conserving model defects at the example of an evaluation of the $(n,2n)$ cross section of ^{181}Ta . The $(n,2n)$ scattering process leaves the residual nucleus either in the ground state or an isomer state. The sum of these contributions has to equal the $(n,2n)$ total cross section and hence this cross section is suited to verify whether sum rules are preserved by the model defects.

We used the surrogate approach described in section 3.3, especially the update formulas in Equation 3.33 and Equation 3.34. A sample of thousand model predictions was generated by randomly varying parameters of the optical model and the Fermi gas model. The model parameters associated with the optical model in the parametrization of Koning and Delaroche (2003) were drawn from a uniform distribution within a 20% window around the default values. Table 9.1 lists the TALYS adjustment factors of these model parameters. The level density parameters of the Fermi gas model were drawn from a uniform distribution within the intervals listed in Table 9.2. The allowed intervals of variation of the optical model parameters roughly matches the choice of Leeb, Schnabel, et al. (2015) and the boundaries of the intervals for the level density parameters are identical.

The experimental data for the $(n,2n)$ -total, $(n,2n)$ -ground and $(n,2n)$ -isomer cross section included in the evaluation are listed in Table 9.3.

First, we performed an evaluation without model defects. The result is shown in Figure 9.14. The $(n,2n)$ -ground cross section is adequately described whereas the evaluated cross section curve of the $(n,2n)$ -total channel is above and those of the $(n,2n)$ -isomer channel is below the experimental data. The experimental data of these three channels are inconsistent because they violate the sum rule. The Bayesian update procedure preserves sum rules and consequently the evaluated cross section curves in the individual channels represent a compromise. Plenty of data are available in the $(n,2n)$ -ground channel, hence the evaluation within this channel is rather fixed. For the other two channels, the number of experimental data points is much smaller. The evaluation is shifted upwards from the experimental data in the $(n,2n)$ -total channel in order to get closer to the data in the $(n,2n)$ -isomer channel. Noteworthy, the evaluated cross sections of all channels between 7.5 and 8 MeV near the threshold energy are slightly negative (about -10 millibarn) which is due to model deficiency.

The evaluated correlations of and between the cross section channels are shown in Figure 9.15. Correlations of the $(n,2n)$ -total and $(n,2n)$ -isomer cross section look almost identical. Furthermore, the same structure can be found for the correlations between these channels. Therefore, a shift of the cross section curve in one of these

rvadjust	avadjust	v1adjust	v2adjust	v3adjust	v4adjust
rwadjust	awadjust	w1adjust	w2adjust	rvadjust	avadjust
rwdadjust	d1adjust	d2adjust	d3adjust	rvsoadjust	avsoadjust
vso1adjust	vso2adjust	rwsadjust	awsadjust	wso1adjust	wso2adjust

Table 9.1: TALYS (Koning, Hilaire, and Duijvestijn, 2007) keywords to adjust the default parameters of the optical model in the global parametrization of Koning and Delaroche (2003). Each particle (neutron, proton, ...) is associated with its own set of these adjustment factors. Thousand parameter sets were created for the construction of the prior mean vector and the prior covariance matrix. Adjustment factors of the neutron, proton and alpha-particle were drawn from a uniform distribution in the interval $[0.8, 1.2]$.

keyword	alphald	betald	Pshiftconstant	gammashell
lower limit	0.011402	0.126247	-3.000000	0.331541
upper limit	0.030059	0.332833	3.000000	0.615719

Table 9.2: Ranges of the level density parameters of the Fermi Gas model in which values were drawn from a uniform distribution.

Reference	EXFOR #	points
(n,2n) total cross section		
Veeseer, Arthur, and Young (1977)	10445009	10
Frehaut (1980)	20416016	14
Takahashi (1992)	22136026	1
(n,2n)-ground level		
Peiguo (1985)	30733002	1
Han-Lin, Wen Rong, and Guo (1985)	30724003	33
Kasugai (1992)	22351005	6
Filatenkov (1999)	41240103	8
Filatenkov (2001)	41424056	4
Zhu et al. (2011)	32696009	6
(n,2n)-isomer level		
Ikeda (1988)	22089106	8

Table 9.3: Experimental data included in the evaluation.

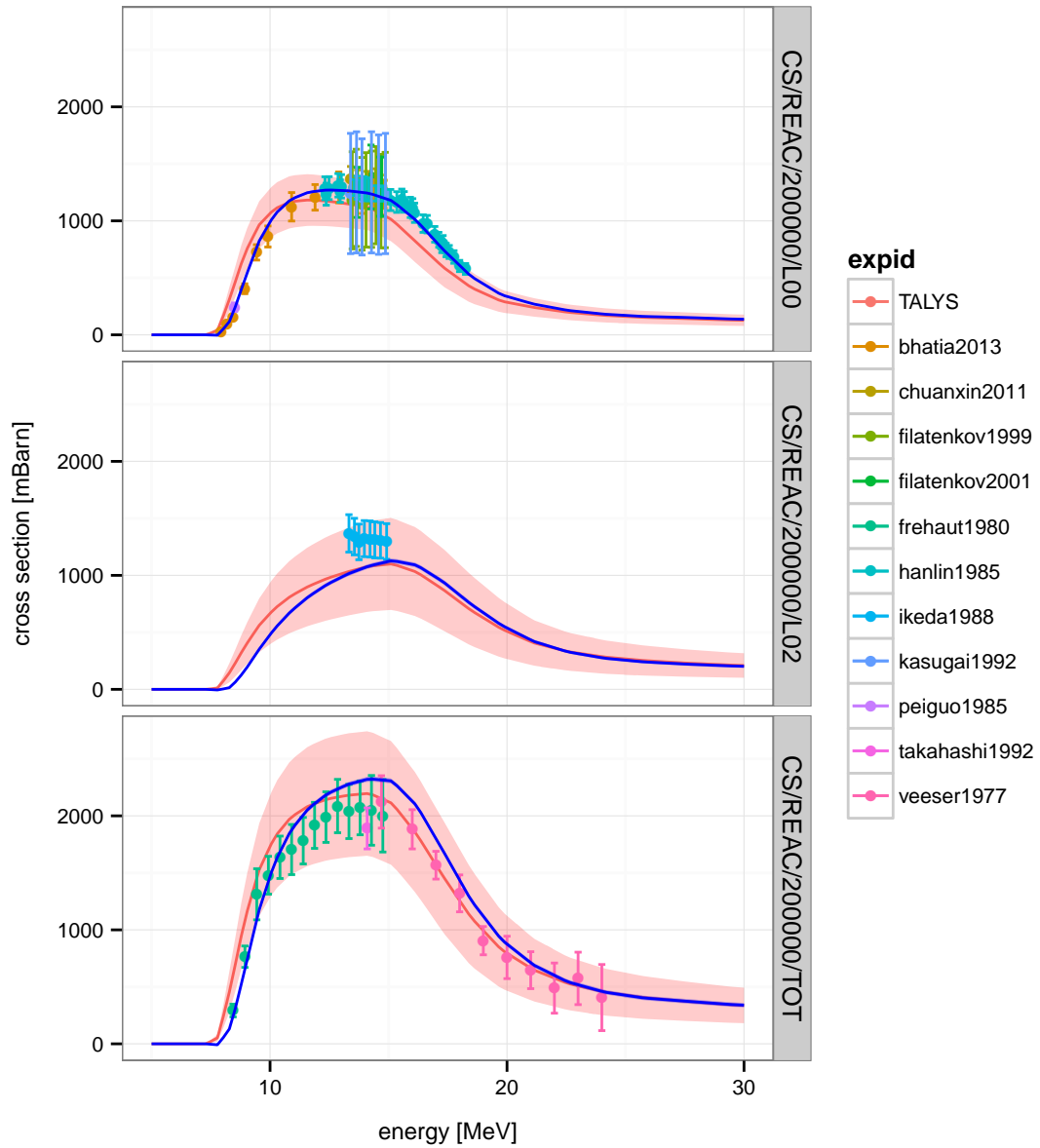


Figure 9.14: Evaluated cross sections (blue lines) obtained by the surrogate approach without model defects. Illustrated are from top to bottom the (n,2n)-ground, (n,2n)-isomer, and (n,2n)-total cross section of ^{181}Ta . Error bars of the experiment indicate the 95% confidence interval and the error band indicates the 68% confidence interval of the prior cross section curve.

channels leads to approximately the same shift in the other channel.

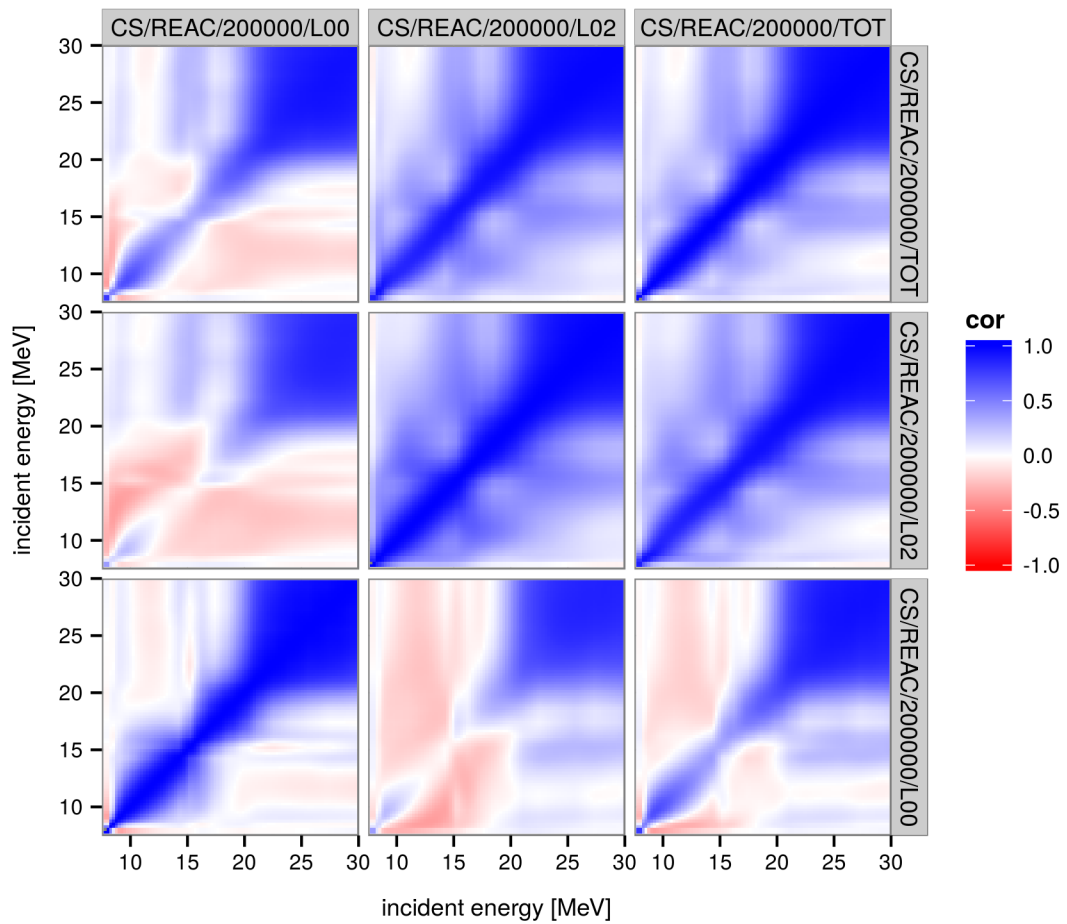


Figure 9.15: Evaluated correlations obtained by the surrogate approach without model defects. The blocks in the diagonal show from bottom-left to top-right the correlation matrices of the (n,2n)-ground, (n,2n)-isomer, and (n,2n)-total cross section channel. Off-diagonal blocks visualize the correlations between these reaction channels.

In the following we perform the evaluation with model defects, hence the formulas of section 9.3 were applied. We assume two Gaussian processes for the model defect in the (n,2n)-ground and (n,2n)-isomer channel (abbreviated as *gnd* and *iso*, respectively),

$$\varepsilon(\text{gnd}, E) \sim \mathcal{GP}(m(E), k_{\text{gnd}}(E, E')) \quad \text{and} \quad \varepsilon(\text{iso}, E) \sim \mathcal{GP}(m(E), k_{\text{iso}}(E, E')). \quad (9.91)$$

The mean function for both model defects is given by $m(E) \equiv 0$. We choose for both channels a squared exponential covariance function,

$$\tilde{k}_c = \delta_c^2 \exp\left(-\frac{1}{2\lambda_c}(E - E')^2\right). \quad (9.92)$$

Therefore, four hyperparameters $\delta_{\text{gnd}}, \lambda_{\text{gnd}}, \delta_{\text{iso}}, \lambda_{\text{iso}}$ are involved in the evaluation. The covariances of and between all channels are determined by Equation 9.90. The maximization of the marginal likelihood (see Equation 9.56) yields

δ_{gnd}	λ_{gnd}	δ_{iso}	λ_{iso}
0.2612	3.0018	0.1695	0.5718

The evaluated cross sections are shown in Figure 9.16. Due to the specific construction of the model defects, the sum rules are preserved. The sum of the (n,2n)-ground and (n,2n)-isomer cross section deviates not more than 3×10^{-4} millibarn from the (n,2n)-total cross section. The small deviation can be attributed to numerical imprecisions in the calculation.

The evaluated cross section curve, especially in the (n,2n)-isomer and (n,2n)-total cross section channel may not be in line with our expectations about the proper shape. However, we have to consider the assumptions which entered the evaluation:

1. Estimates and associated uncertainties of the experimental data are reliable.
2. The model defect is a smooth function with derivatives of any order.
3. Indifference about the length-scale of the model defect.

According to these prior assumptions, the evaluated cross section curves makes sense, because:

1. All experimental data included in the evaluation are perfectly described, except at the incident energies where they violate the sum rule.
2. Abrupt changes of the evaluated cross section curve do not occur due to the assumed smoothness of the model defect, hence also the oscillations.

3. The length-scale λ_{iso} is adjusted to a small value in order to allow the rapid increase of the model defect in the (n,2n)-total cross section channel. Due to the small length-scale, the evaluated (n,2n)-total cross section curve is able to quickly deviate from the experimental data in order to represent a good compromise with the experimental data in the (n,2n)-isomer channel.

We redo the evaluation, but this time forbid length-scales smaller than 3 MeV. The L-BFGS-B algorithm (Byrd et al., 1995) used for the maximization of the marginal likelihood can deal with this type of restriction. The result of the evaluation is shown in Figure 9.17. The obtained hyperparameters are given by

δ_{gnd}	λ_{gnd}	δ_{iso}	λ_{iso}
0.2503	5.9159	0.4721	3.0000

Evaluated cross section curves do not exhibit the oscillations anymore and may be considered more in line with the expectation about the proper shape. Compared to the evaluation without model defects, the resulting cross section curves from the evaluation with model defects better resemble the experimental data. In addition, also the negative evaluated cross sections at the threshold energy do (almost) not appear anymore. Without model defects, the most negative cross section was about 10 millibarn. In contrast to that, the most negative cross section occurring when taking into account model defects is approximately -0.1 millibarn—a huge improvement.

The evaluated model defects are illustrated in Figure 9.18. The graph clearly indicates the deficiency of the model in the (n,2n)-ground channel near the threshold energy. Exactly this deficiency caused the negative cross section in the evaluation which did not account for model defects. Further, we recognize the power of Gaussian processes in predicting the shape of unknown functions. Even though, the employed stationary covariance function implies the prior assumption that the relative model error is in the same magnitude at every incident energy, the evaluated model error can have significant different values at different incident energies. In the current example, the obtained value $\delta_{\text{iso}} = 0.4721$ reflects the prior assumption that the model deficiency is in the order of 50%, yet the evaluated model error at the threshold energy in the (n,2n)-isomer channel is greater than 300%.

Further, the assumption of a uniform model error seems not to hold. If the model deficiency in several evaluations is always very pronounced at the threshold energy and almost zero at higher incident energies, one may consider to refine the assumptions about the model defect. It is possible to construct a covariance function which implies the assumption of a high and quickly varying model defect

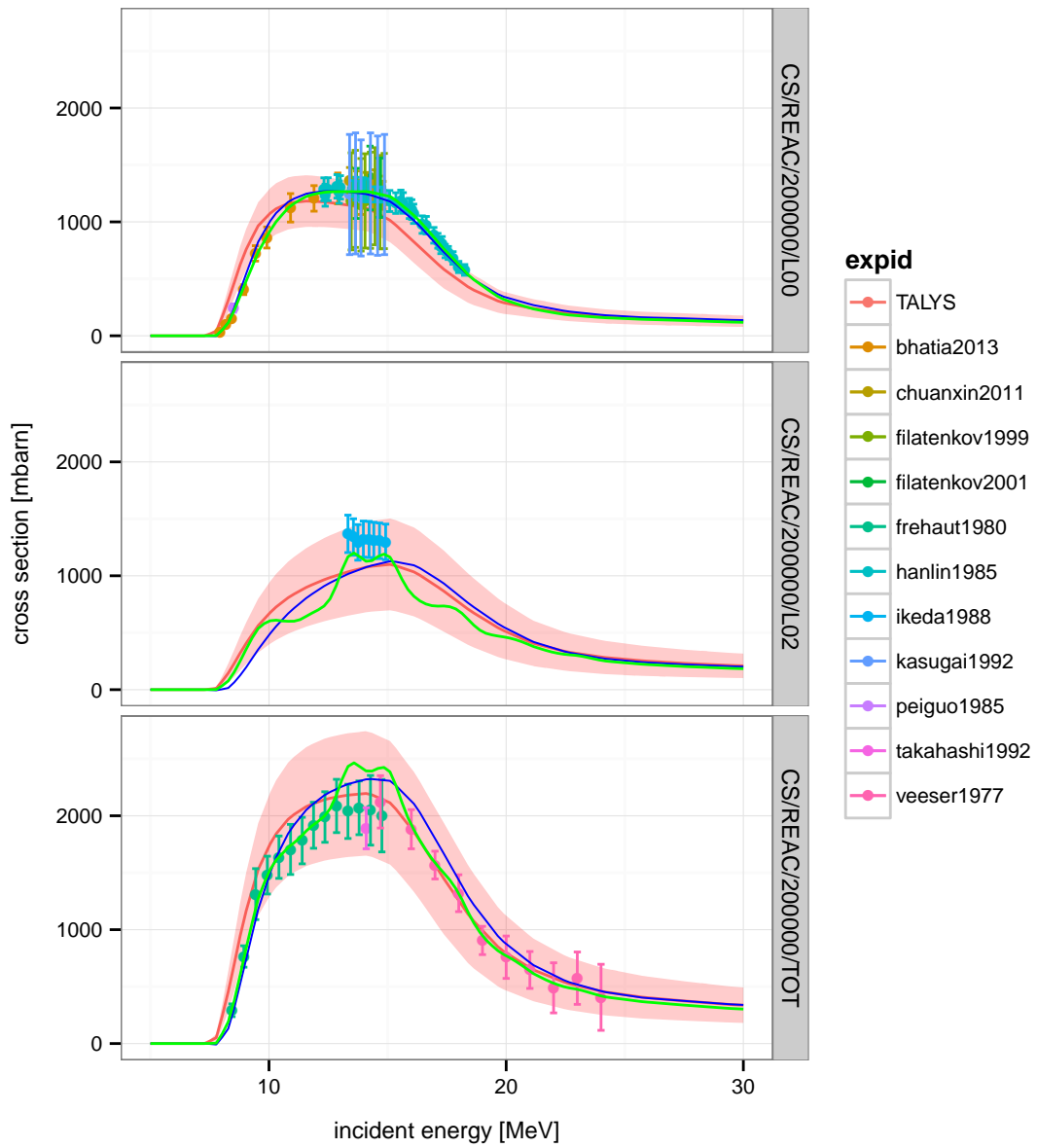


Figure 9.16: Evaluated cross section curves (green lines) obtained by the surrogate approach with model defects. The graphs display from top to bottom the (n,2n)-ground, (n,2n)-isomer, and (n,2n)-total cross section. The squared exponential covariance function was employed for the model defects.

at the threshold energy and a slower varying model defect component at higher incident energies.

The evaluated correlations are shown in Figure 9.19. These correlations are for the quantity which we denoted the ‘true cross section’, $\vec{\sigma}_{\text{true}} = \vec{\sigma}_{\text{pred}} + \vec{\varepsilon}_{\text{mod}}$. The structure visible in these graphs is different from the correlations resulting from an evaluation without model defects (see Figure 9.15). Correlations within channels and also between different channels are rather dimmed. Only cross sections at nearby incident energies are strongly correlated due to the finite length-scale of the covariance function of the model defect. However, the correlations between the cross sections of the (n,2n)-isomer and the (n,2n)-total channel at nearby energies are very pronounced. This finding reflects the sum rule constraint between the cross section channels: The evaluated cross section curve in the (n,2n)-ground channel is fixed due to plenty of experimental data; and due to the sum rule between the cross section channels only one degree of freedom remains for the (n,2n)-isomer and (n,2n)-total channel. If the cross section curve in the (n,2n)-isomer channel is shifted upwards, the same shift has to occur in the (n,2n)-total channel.

The evaluated correlations of the model defect are shown in Figure 9.21. Even though the model defect of elementary channels are assumed to be a priori uncorrelated, the evaluated model defects of the (n,2n)-ground and (n,2n)-isomer channel are slightly correlated. Further, two domains of the model defect can be distinguished in the (n,2n)-ground channel: The model defects below about 16 MeV are strongly correlated with each other, and the model defect above 16 MeV are rather strongly correlated. However, the correlation of the model defects between these domains is very weak.

The evaluated correlations between the model prediction and the model defect are illustrated in Figure 9.21. All correlations are negative, which is the expected result. The model prediction is adapted to the experimental data as good as possible and the model defect helps to overcome the remaining deviation. If the model prediction slightly changes in some direction, the model defect has to change in the other direction in order to still yield a good description of the data.

Finally, the evaluated correlations and uncertainties of the (n,2n)-total cross section are displayed in Figure 9.22. The smallest uncertainty is about 1% where experimental data are available. In the high energy tail the uncertainty rises to about 30%. Near the threshold energy the evaluated uncertainty is even larger than 100%, hence would need to be cut before entering a nuclear data file. For comparison, also the evaluated correlations of the (n,2n)-total cross section obtained by the surrogate approach without model defects are shown in Figure 9.23.

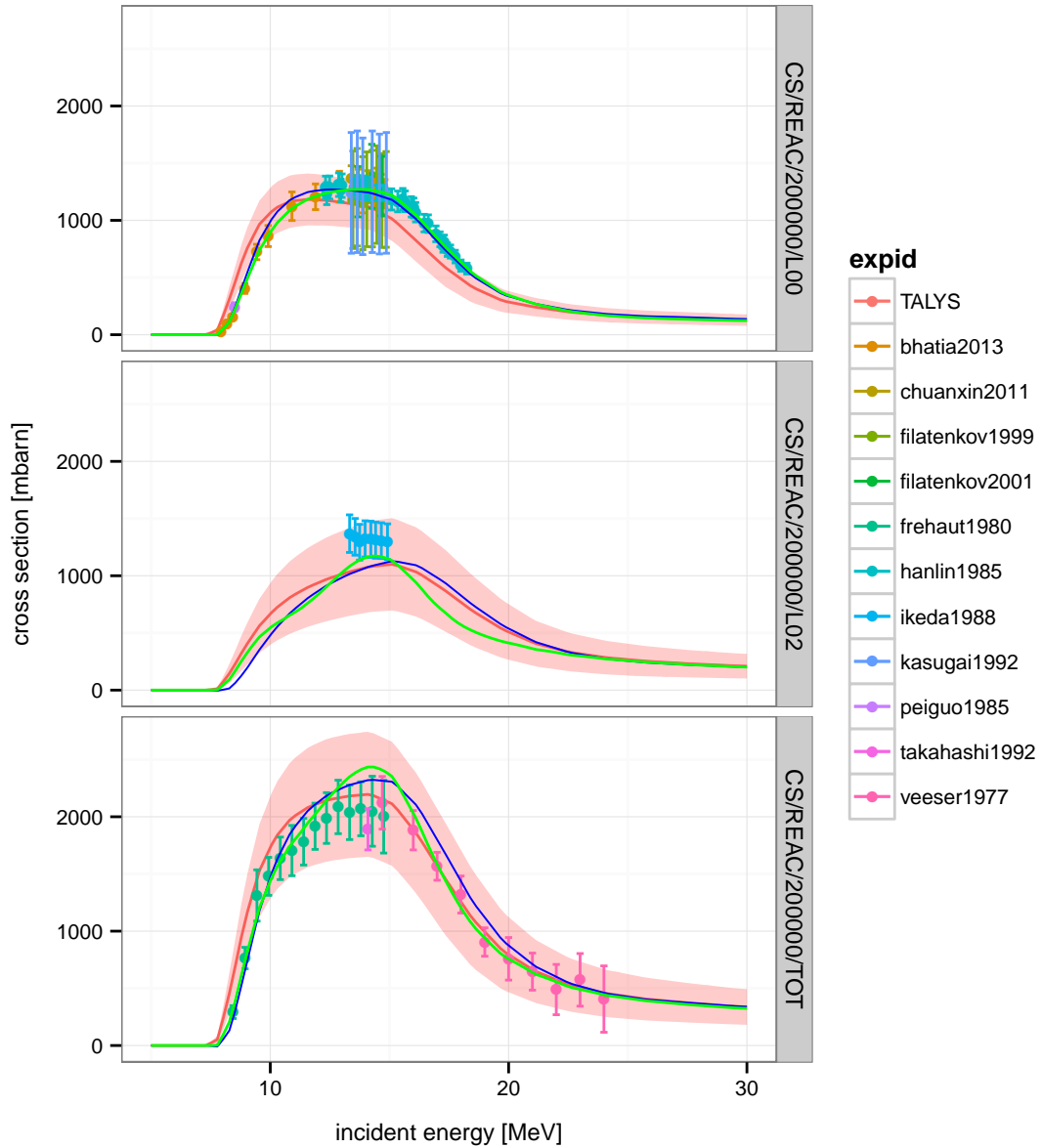


Figure 9.17: Evaluated cross section curves (green lines) obtained by the surrogate approach with model defects. The graphs display from top to bottom the (n,2n)-ground, (n,2n)-isomer, and (n,2n)-total cross section. The squared exponential covariance function was employed for the model defects. Length-scales of the model defects were restricted to values greater than 3 MeV in the maximization of the marginal likelihood.

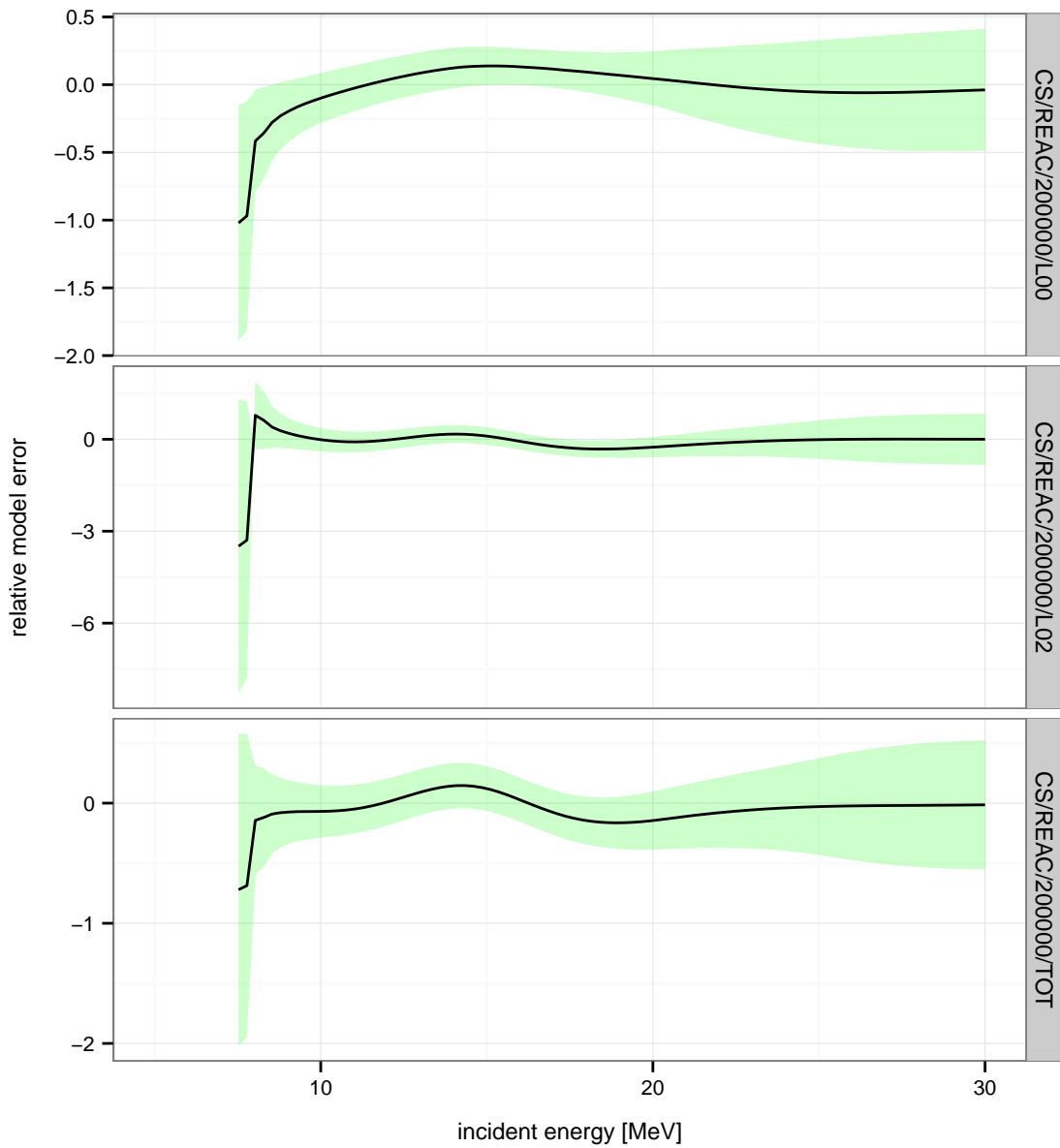


Figure 9.18: Evaluated relative model errors with associated 95% confidence bands. Shown are from top to bottom the (n,2n)-ground, (n,2n)-isomer, and (n,2n)-total cross section channel. If included experimental data are reliable, then the model is clearly deficient near the threshold energy in the (n,2n)-ground channel. Everywhere else the nuclear model seems to be reliable.

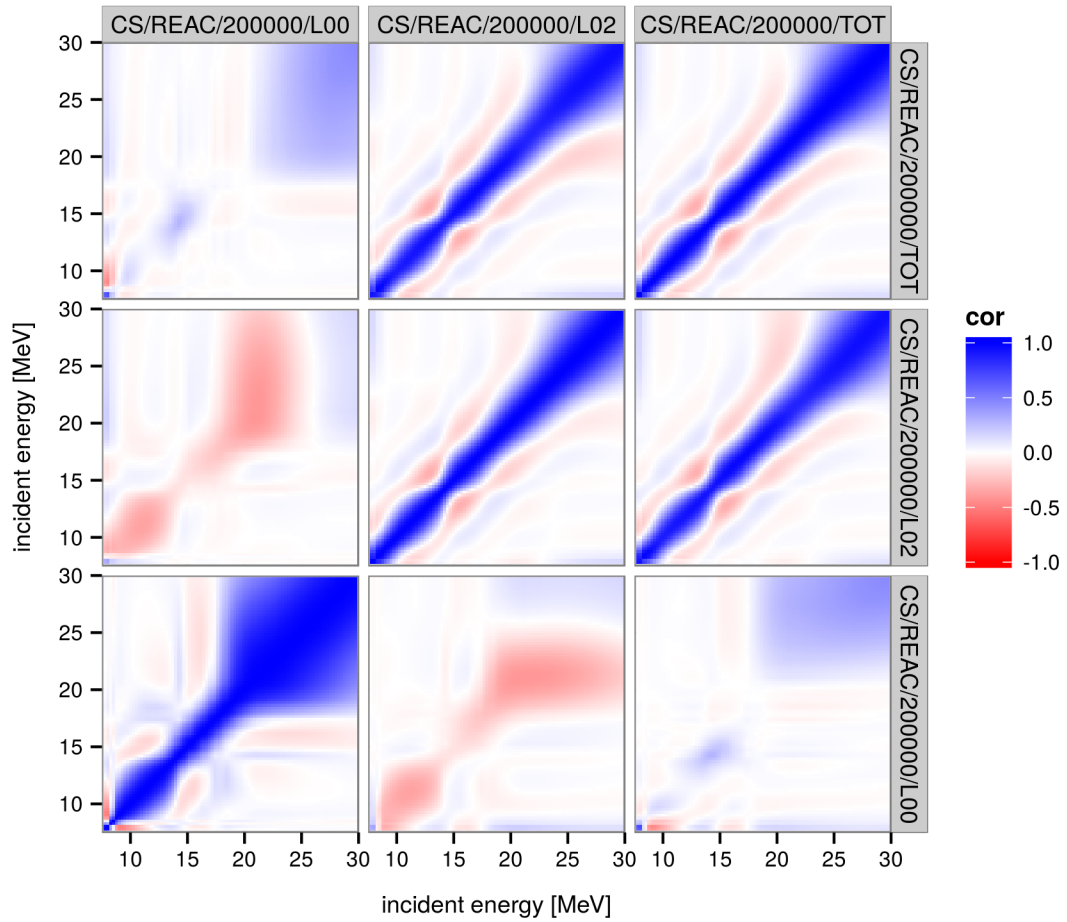


Figure 9.19: Evaluated correlations of cross sections within and between the reaction channels obtained by the surrogate approach with model defects. The channels in the diagonal are from bottom-left to the top-right the (n,2n)-ground, (n,2n)-isomer and (n,2n)-total cross section channel.

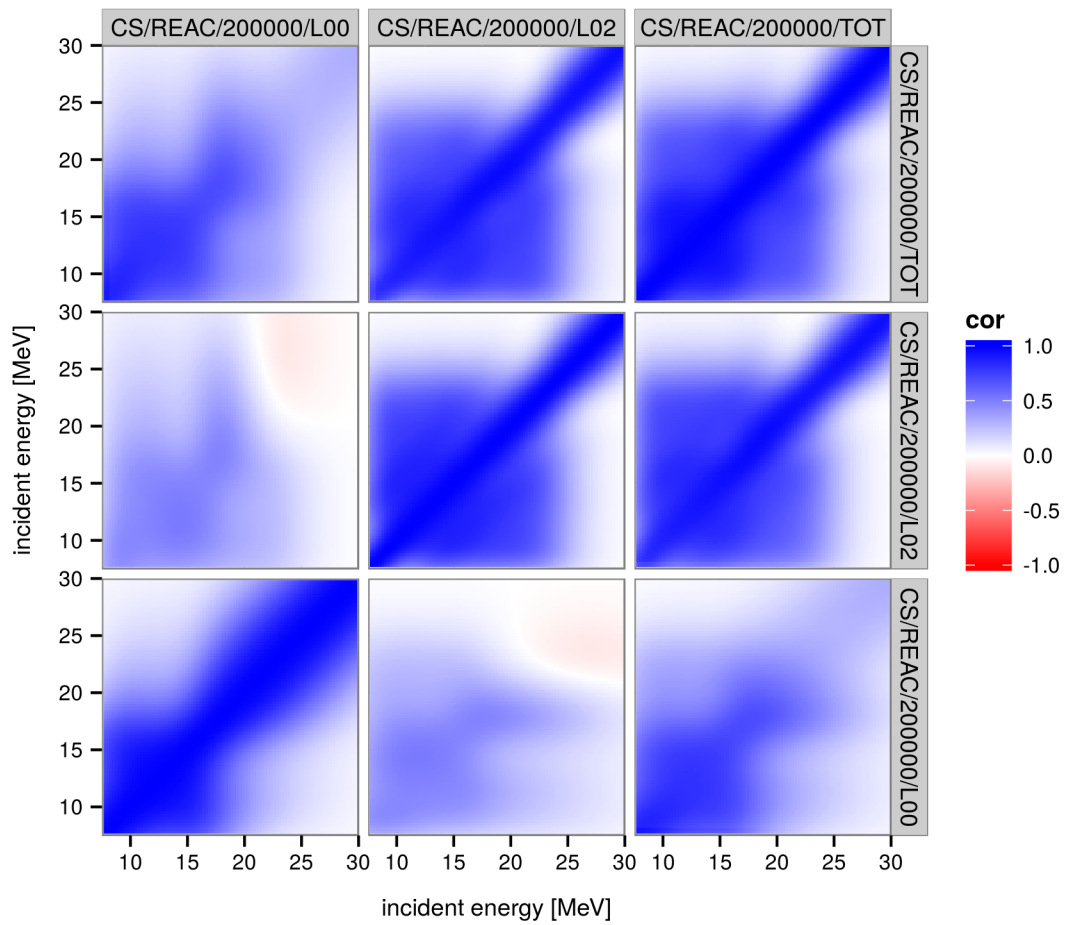


Figure 9.20: Evaluated correlations of the model defect within and between the reaction channels obtained by the surrogate approach. The channels in the diagonal are from bottom-left to the top-right the (n,2n)-ground, (n,2n)-isomer and (n,2n)-total cross section channel.

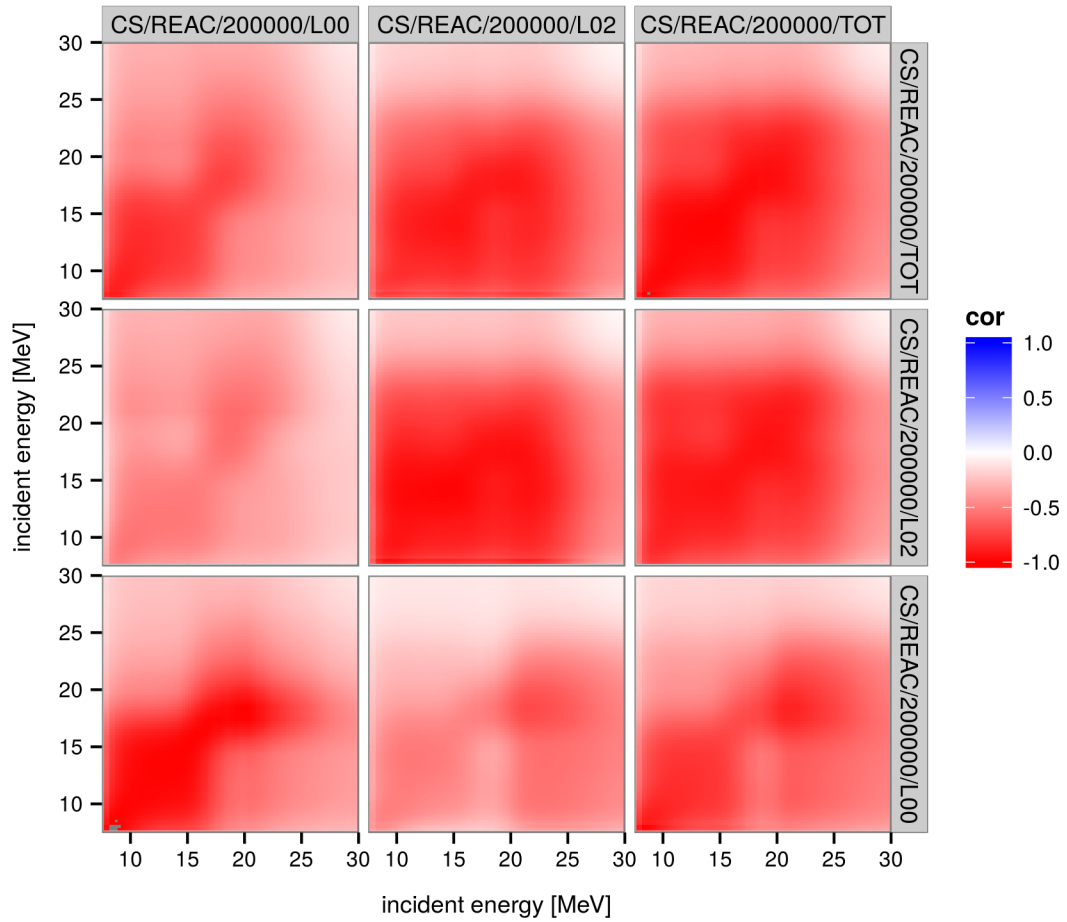


Figure 9.21: Evaluated correlations between the model defect and the model prediction within and between the reaction channels obtained by the surrogate approach. The channels in the diagonal are from bottom-left to the top-right the (n,2n)-ground, (n,2n)-isomer and (n,2n)-total cross section channel.

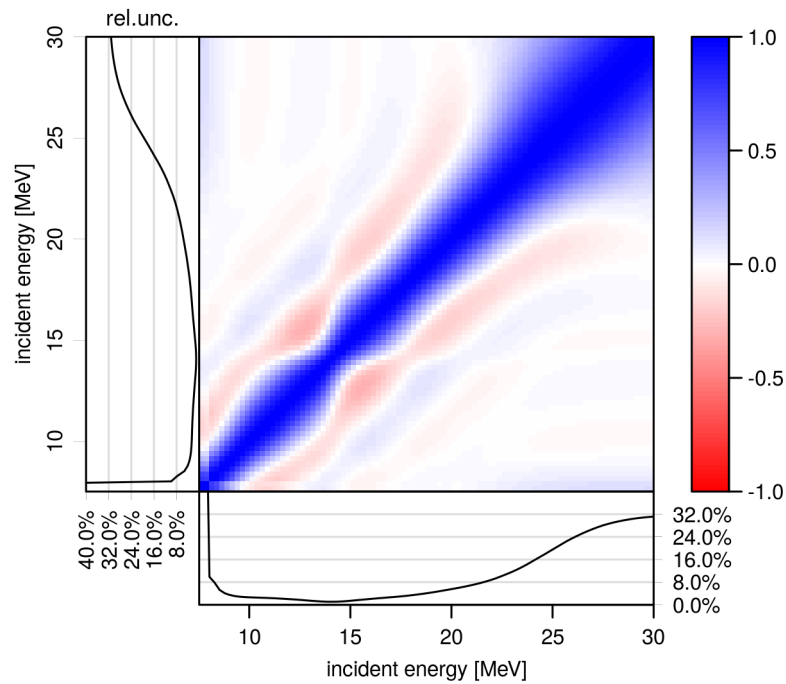


Figure 9.22: Evaluated correlations of the evaluated (n,2n)-total cross section and associated uncertainties obtained by the surrogate approach with model defects.

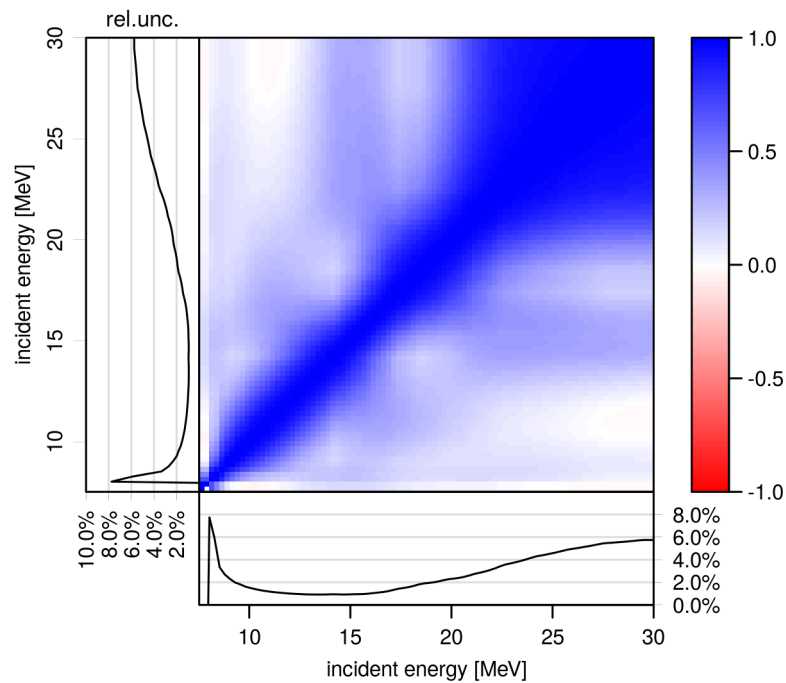


Figure 9.23: Evaluated correlations of the evaluated (n,2n)-total cross section and associated uncertainties obtained by the surrogate approach without model defects.

Summary and outlook

The existing evaluation methods (FBET, EMPIRE-MC, EMPIRE-Kalman, UMC-G, UMC-B, BMC, BFMC) were thoroughly studied in two evaluation scenarios. In the first scenario, the evaluation was performed with a linear model and hypothetical experimental data. The experimental data were specifically constructed in order to be not representable by the linear model. Due to the choice of a linear model, all methods except the BFMC and BMC method collapsed to one common approach.

The evaluated estimates and uncertainties of all studied methods except BFMC were inconsistent with the included experimental data. The possibility for the experimental data points to represent the true cross sections was excluded by the evaluated uncertainties. Only the uncertainties obtained by the BFMC method were in agreement with the experimental data due to an alternative choice of the likelihood.

In the second scenario, an evaluation of the neutron-induced total cross section of ^{181}Ta was performed. The optical model in the parametrization of Koning and Delaroche (2003) was employed and the experimental data of Finlay et al. (1993) were included in the evaluation. The important difference to the first scenario is the non-linear relationship between model parameters and cross sections established by the optical model. It has been demonstrated that the optical model is capable to describe the experimental data rather well.

From the evaluation methods which use a multivariate normal likelihood, only the FBET, EMPIRE-MC and UMC-G method based on the surrogate approach produced evaluations consistent with the experimental data. The BFMC and BMC method using a likelihood different from a multivariate normal distribution were also consistent, but their evaluated uncertainties were significantly larger than suggested by the experimental data. In addition, the evaluated covariance matrix

restricted possible variations of the cross section curve to possibilities within the model. An undesirable feature if the model is deficient. The evaluations obtained by all other studied methods were inconsistent with the experimental data.

The investigation of the evaluation methods in the two evaluation scenarios leads us to the following conclusions:

1. The form of the likelihood reflects the assumptions about the uncertainties in the experiment. Hence, from a conceptual point of view it is inappropriate to modify its form in order to obtain evaluated uncertainties of reasonable magnitude.
2. Using a simplified model (surrogate approach or linearization of the model) cannot be regarded as a sufficient measure to deal with the deficiencies of the nuclear model. The deficiency of a model cannot be expected to be dependent on the degree of non-linearity of the model. Yet, regarding the surrogate approach as a proper measure to address model defects would imply this assumption.
3. Proper Bayesian inference can be performed on the basis of evaluation methods using a simplified model and Monte Carlo methods using the exact nuclear model. The choice among these two classes of methods is a choice on how the Bayesian update formula should be evaluated. However, from a conceptual point of view it is more important to consider the assumptions which enter the Bayesian inference procedure.
4. Assumptions about the quality, predictive power or efficiency of the nuclear model have to enter the prior distribution. If the nuclear model is not considered to be capable to perfectly predict the ‘true’ cross section curve, an essential assumption is missing in most contemporary evaluation methods.

On the basis of these conclusions, the Bayesian update formula was extended to account for the deficiency of the nuclear model. The assumption of contemporary evaluation methods that the true cross section curve can be predicted by the nuclear model was replaced by the assumption that the true cross section curve is the sum of the model prediction and a model error function. Because the shape of the latter function is unknown, it has to be regarded as a random function which was modeled as a Gaussian process. Therefore, no specific functional form needs to be specified for the evaluation. Instead, a probability function on possible functions is constructed based on the expected smoothness of the evaluated cross section curve and the expected deviations of the model prediction from reliable experimental data. The maximization of the marginal likelihood was proposed

to find the so-called hyperparameters of the Gaussian process, which determine smoothness and magnitude of the model error function. Moreover, the prior on the model error functions of different channels was constructed in a way to preserve the sum rules of cross sections. Exploiting the statistical methodology of Gaussian process regression, it was shown how the functional form of the model defect can be explicitly estimated. Both the possibility to explicitly estimate the shape of the model error and a Bayesian formulation accounting for model defects which preserves sum rules is unprecedented in nuclear data evaluation.

The extended Bayesian update formula was tested in the evaluation of the neutron-induced total cross section of ^{181}Ta . The update formula was evaluated with the iterative linearized Bayesian update approach¹. The evaluated cross section curve and uncertainties perfectly matched included experimental data. In addition, the evaluated uncertainties were consistent with the experimental data which had not been included in the evaluation. Therefore, the open questions of available methods have been resolved. A second evaluation of the (n,2n) cross section of ^{181}Ta was performed in order to demonstrate the conservation of sum rules.

The squared exponential covariance function was chosen in the evaluations to demonstrate the features of the extended Bayesian approach. This choice implies certain assumptions on the model error function such as the existence of derivatives of any order and the prior expectation that the relative model error is independent of the incident energy. Without further prior knowledge, both assumptions can be regarded reasonable. However, in some situations prior knowledge may exist which suggests different assumptions. In such cases, the covariance functions can be tailored according to the specific prior knowledge. It is even possible to use other processes than Gaussian processes to model the prior knowledge about the model error. However, the important conclusion of this thesis is that the model error must be defined as a random function, which is added to the model prediction in order to form the true cross section.

How the extended Bayesian formula is evaluated, either by the simplification of the nuclear model or Monte Carlo methods, is then a completely different question. If Monte Carlo methods are feasible, they are the preferred choice to evaluate the Bayesian update formula. The exact model is taken into account and thus a clear separation between the model prediction and the model error is possible. If the model is replaced by a surrogate model, it remains an open question whether the model is accurate or its deficiency (as a model feature) was not captured by the surrogate model.

¹This approach is for instance implemented in the code SAMMY.

The existing Monte Carlo schemes suggested for the BMC, UMC-B and UMC-G method proved to be not efficient in the evaluation scenarios. Their sampling distributions are determined by the prior distribution, yet often the posterior distribution is to a large extent determined by the likelihood. Additionally, the proposed sampling distributions for the UMC-G method do not account for the prior correlations between cross sections, which has a severe negative impact on convergence.

Therefore, as another achievement of this thesis, a more efficient sampling scheme has been proposed. The idea is to apply the Metropolis-Hastings algorithm and to use an approximation of the posterior covariance matrix to define a multivariate normal distribution as proposal distribution. The approximation is constructed by using the Bayesian update procedure with a simplified model. In the evaluation of the neutron-induced total cross section of ^{181}Ta with plenty of experimental data available, the refined sampling scheme has been demonstrated to be more efficient than the other sampling schemes. However, very likely still better sampling schemes exist to further increase the speed of convergence.

Finally, the Bayesian update formulas based on the surrogate approach as implemented by the FBET and the EMPIRE-MC method have been reformulated in order to avoid the explicit calculation of the prior covariance matrix. In the conventional update scheme it is not feasible to evaluate more than tens of thousands of observables. In contrast to that, the new scheme allows the evaluation of dozens of millions of observables. Furthermore, model defects can be included in the new scheme in a straight-forward way.

The new scheme is suited to be implemented as a database application: A client sends experiment data and an associated covariance matrix to the server and requests specific evaluated observables (cross sections, angle-differential cross sections, etc.). The server retrieves the desired observables from a sample of model calculations with varied model parameters and weights them according to received experimental data. Simple vector and matrix computations yield then the posterior means and covariances of the demanded observables, which are sent back to the client. Such a database application accessible over the Internet could prove useful to accelerate the production of reliable nuclear data. Due to the inclusion of model defects, deficiencies of the nuclear models or inconsistent experimental data can be automatically tracked down without user interaction. Therefore, also the development of more accurate nuclear models can profit from such a database.

Nuclear models

The nuclear models code TALYS (Koning, Hilaire, and Duijvestijn, 2007) was used to obtain angle-integrated cross sections. Especially, the parameters of the optical potential were varied in order to generate the prior distributions. In the following we outline the optical potential as implemented in TALYS. We closely follow the explanation of the TALYS manual (Koning, Hilaire, and Goriely, 2013).

A.1 Neutron and proton optical potential

The optical potential is a phenomenological approach to describe the interaction between a nucleon and a nucleus. It explicitly describes shape elastic scattering while non-elastic channels are only accounted for globally. The numerical solution of the Schrödinger equation with the optical potential yields several quantities of interest such as the angle-integrated and angle-differential cross sections of the shape elastic component. TALYS uses the optical potential in the global parametrization of Koning and Delaroche (2003).

The optical potential is of the form

$$\begin{aligned} \mathcal{U}(r, E) = & -\mathcal{V}_v(r, E) - i\mathcal{W}_v(r, E) - i\mathcal{W}_d(r, E) \\ & + \mathcal{V}_{so}(r, E)\ell \cdot \sigma + i\mathcal{W}_{so}(r, E)\ell \cdot \sigma + \mathcal{V}_C(r) \end{aligned} \quad (\text{A.1})$$

with the real and imaginary volume-central components $\mathcal{V}_v(r, E)$ and $\mathcal{W}_v(r, E)$, the imaginary surface central component $\mathcal{W}_d(r, E)$, the real and imaginary spin orbit component $\mathcal{V}_{so}(r, E)$ and $\mathcal{W}_{so}(r, E)$ and the real Coloumb component $\mathcal{V}_C(r)$. All components are separated in well depths V_v, W_v, W_d, V_{so} , and W_{so} , which depend on the laboratory energy of the incident particle, and energy-independent radial

parts $f = f(r)$,

$$\begin{aligned}\mathcal{V}_v(r, E) &= \mathcal{V}_v(E)f(r, R_v, a_v), \\ \mathcal{W}_v(r, E) &= \mathcal{W}_v(E)f(r, R_v, a_v), \\ \mathcal{W}_d(r, E) &= -4a_d\mathcal{W}_d(E)\frac{d}{dr}f(r, R_d, a_d), \\ \mathcal{V}_{so}(r, E) &= V_{so}(E)\left(\frac{\hbar}{m_\pi c}\right)^2\frac{1}{r}\frac{d}{dr}f(r, R_{so}, a_{so}), \\ \mathcal{W}_{so}(r, E) &= \mathcal{W}_{so}(E)\left(\frac{\hbar}{m_\pi c}\right)^2\frac{1}{r}\frac{d}{dr}f(r, R_{so}, a_{so}).\end{aligned}$$

The form factor $f(r, R_i, a_i)$ is given by a Wood-Saxon shape,

$$f(r, R_i, a_i) = \left[1 + \exp\left(\frac{r - R_i}{a_i}\right)\right]^{-1} \quad (\text{A.2})$$

where the radius is determined by $R_i = r_i A^{1/3}$, with the atomic mass number A and the diffuseness parameter a_i . If the particle is charged, the Coloumb term V_C is given by the potential of a uniformly charged sphere,

$$\mathcal{V}_C(r) = \begin{cases} \frac{Zze^2}{2R_C}\left(3 - \frac{r^2}{R_C^2}\right) & \text{for } r \leq R_C \\ \frac{Zze^2}{r} & \text{for } r > R_C \end{cases}, \quad (\text{A.3})$$

with the charge of the target and the the projectile Z and z , respectively. The Coloumb radius is given by $R_C = r_C A^{1/3}$. The energy-dependent potential depths for neutrons and protons take the form

$$\begin{aligned}\mathcal{V}_v(E) &= v_1[1 - v_2(E - E_f) + v_3(E - E_f)^2 - v_4(E - E_f)^3], \\ \mathcal{W}_v(E) &= w_1\frac{(E - E_f)^2}{(E - E_f)^2 + w_2^2}, \\ \mathcal{W}_d(E) &= d_1\frac{(E - E_f)^2}{(E - E_f)^2 + d_3^2}\exp[-d_2(E - E_f)], \\ \mathcal{V}_{so}(E) &= v_{so,1}\exp[-v_{so,2}(E - E_f)], \\ \mathcal{W}_{so}(E) &= w_{so,1}\frac{(E - E_f)^2}{(E - E_f)^2 + w_{so,2}^2}.\end{aligned} \quad (\text{A.4})$$

The Fermi energy for neutrons is given by

$$E_f^n = -\frac{1}{2}[S_n(Z, N) + S_n(Z, N + 1)], \quad (\text{A.5})$$

where S_n denotes the neutron separation energy. The Fermi energy for the proton is determined by

$$E_f^p = -\frac{1}{2}[S_n(Z, N) + S_n(Z + 1, N)], \quad (\text{A.6})$$

with the proton separation energy S_p available in the nuclear structure database (National Nuclear Data Center, 2015a). For some isotopes TALYS provides optimized values for the parameters v_1, v_2, \dots . If such an optimized parameter set is not available, the global parametrization of Koning and Delaroche (2003) is used, which is valid for mass numbers in the range $24 \leq A \leq 209$. The optical potential parameters for neutrons are given by

$$\begin{aligned}
r_v^n &= 1.3039 - 0.405A^{-1/3}, \\
a_v^n &= 0.6778 - 1.487 \times 10^{-4}A, \\
r_d^n &= 1.3424 - 0.01585A^{1/3}, \\
a_d^n &= 0.5446 - 1.656 \times 10^{-4}A, \\
r_{so}^n &= 1.1854 - 0.647A^{-1/3}, \\
a_{so}^n &= 0.59, \\
v_1^n &= 59.30 - 21 \frac{N - Z}{A} - 0.024A, \\
v_2^n &= 0.007228 - 1.48 \times 10^{-6}A, \\
v_3^n &= 1.994 \times 10^{-5} - 2 \times 10^{-8}A, \\
w_1^n &= 12.195 + 0.0167A, \\
w_2^n &= 73.55 + 0.0795A, \\
d_1^n &= 16 - 16 \frac{N - Z}{A}, \\
d_2^n &= 0.018 + \frac{0.003802}{1 + \exp[(A - 156)/8]}, \\
d_3^n &= 11.5, \\
v_{so,1}^n &= 5.922 + 0.0030A, \\
v_{so,2}^n &= 0.0040, \\
w_{so,1}^n &= -3.1, \\
w_{so,2}^n &= 160, \\
E_f^n &= -11.2814 + 0.02646A.
\end{aligned} \tag{A.7}$$

In the case of the proton, the following optical potential parameters are differently parametrized,

$$\begin{aligned}
a_d^p &= 0.5187 - 5.205 \times 10^{-4} A, \\
v_1^p &= 59.30 + 21 \frac{N - Z}{A} - 0.024A, \\
v_2^p &= 0.007067 + 4.23 \times 10^{-6} A, \\
v_3^p &= 1.729 \times 10^{-5} + 1.136 \times 10^{-8} A, \\
w_1^p &= 14.667 + 0.009629A, \\
d_1^p &= 16 + 16 \frac{N - Z}{A}, \\
E_f^p &= -8.4075 + 0.01378A, \\
r_C &= 1.198 + 0.697A^{-2/3} + 12.994A^{-5/3}, \\
\bar{V}_C &= \frac{1.73ZA^{-1/3}}{r_C}.
\end{aligned} \tag{A.8}$$

In TALYS the default values can be modified with keywords such as **rvadjust**, **avadjust**, etc. To which nucleon the modification applies is specified by appending the particle type to the keyword. For instance, **rvadjust n 1.05** scales the default value of r_v^n by 1.05.

Statistics

B.1 Conjugate prior with respect to a likelihood

In Bayesian statistics, the posterior distribution is proportional to the product of prior distribution and likelihood. If the functional form of the latter product is of the same form as the prior distribution, the Bayesian update formula can be conveniently evaluated. Then, it suffices to calculate the distribution parameters which completely characterize the functional form of the posterior distribution. Because this property depends not only on the form of the prior distribution but also on the likelihood, one speaks of a conjugate prior *with respect to the likelihood*.

In the following we give a formal definition taken from Felsenstein (2006).

Definition

A class of prior density functions $\Pi = \{\pi_\psi | \psi \in \Psi\}$ for the parameters θ in the parameter space Θ with a set of hyperparameters Ψ is conjugated to a family of probability density functions $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$ if there exists for every sample $x \in \mathcal{X}$ of the sample space \mathcal{X} and every $\pi_\psi \in \Pi$ a $\psi_x \in \Psi$ such that

$$\theta | X = x \sim \pi_{\psi_x} \in \Pi, \quad (\text{B.1})$$

for $X | \theta \sim P_\theta$ and $\theta \sim \pi_\psi$.

The notation $x \sim P$ indicates that the random variable x is distributed according to the probability density distribution P . As an example, consider the estimation of the mean value m of a random variable x distributed according to a normal distribution with known variance v . Hence, for a given choice of m we have $x \sim \ell(x|m) = \mathcal{N}(m, v)$. The set of all admissible values of m represents the parameter space Θ . The prior distribution $\pi(m)$ for the mean value m could

be given as a normal distribution with center μ and standard deviation δ . It reflects the uncertainty about the unknown value of m . The quantities μ and δ are called hyperparameters because they determine the probability distribution for the parameter m which has to be inferred. A concrete measurement \tilde{x} is associated with the likelihood $\ell(\tilde{x}|m)$ which is reduced to a function of m . The posterior $\pi(m|\tilde{x}) \propto \ell(\tilde{x}|m)\pi(m)$ is again a normal distribution for the mean value m characterized by new hyperparameters μ' and δ' .

B.2 chi-square statistics

The χ^2 -distribution plays a prominent role in statistics. It appears in several statistical tests, such as the χ^2 -test of independence, and is used to assess the quality of a model fit to data. In this thesis, the χ^2 -distribution is frequently used to determine confidence regions associated with the multivariate normal distribution.

Let the random variables Z_1, \dots, Z_k be distributed according to a standard normal distribution, $Z_i \sim \mathcal{N}(0, 1)$, then the sum of their squares is distributed according to a χ^2 -distribution with k degrees of freedom,

$$Z_1 + Z_2 + \dots + Z_k = X \sim \chi_k^2. \quad (\text{B.2})$$

The probability density distribution $f_k(x)$ of the χ^2 -distribution with k degrees of freedom is given by

$$f_k(x) = \begin{cases} \frac{x^{k/2-1} \exp(-x/2)}{2^{k/2} \Gamma(k/2)} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \quad (\text{B.3})$$

The associated cumulative distribution function $F_k(x) = \int_{-\infty}^{\infty} f_k(x) dx$ is often given in tabulated form. However, if k is an even number, it takes the simple form

$$F_k(x) = 1 - \exp\left(-\frac{x}{2}\right) \sum_{i=0}^{k/2-1} \frac{1}{\Gamma(i+1)} \left(\frac{x}{2}\right)^i. \quad (\text{B.4})$$

Using the χ^2 -distribution, confidence regions of a multivariate normal distribution with mean vector $\vec{\mu}$ and covariance matrix Σ can be determined. The confidence region within the true vector is located with probability p is given by

$$(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) \leq \chi_k^2(p), \quad (\text{B.5})$$

where $\chi_k^2(p)$ denotes the quantile function for probability p of the χ^2 -distribution.

B.3 Woodbury identity

The Woodbury matrix identity (Woodbury, 1950) is used in section 3.2 to obtain a particular form of the Bayesian update formula and is given by

$$(\mathbf{A} + \mathbf{UCV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \quad (\text{B.6})$$

with \mathbf{A} , \mathbf{U} , \mathbf{C} , and \mathbf{V} being matrices of appropriate dimension.

The validity of the formula can be verified by multiplying both sides by $(\mathbf{A} + \mathbf{UCV})$ from the left. Evidently, the left side equals the identity matrix, and therefore it suffices to evaluate the right hand side of Equation B.6,

$$\begin{aligned} & (\mathbf{A} + \mathbf{UCV})[\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}] \\ &= \mathbb{1} + \mathbf{UCVA}^{-1} - (\mathbf{U} + \mathbf{UCVA}^{-1}\mathbf{U})(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \\ &= \mathbb{1} + \mathbf{UCVA}^{-1} - \mathbf{UC}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1} \\ &= \mathbb{1} + \mathbf{UCVA}^{-1} - \mathbf{UCVA}^{-1} = \mathbb{1}. \end{aligned} \quad (\text{B.7})$$

Bibliography

- Abramowitz, M. and I. A. Stegun (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. 55. Courier Corporation (cit. on p. 195).
- Adler, R. J. (2010). *The Geometry of Random Fields*. Society for Industrial and Applied Mathematics. ISBN: 978-0-89871-693-1 978-0-89871-898-0 (cit. on p. 193).
- Bauge, E. and P. Dossantos-Uzarralde (2011). “Evaluation of the Covariance Matrix of ^{239}Pu Neutronic Cross Sections in the Continuum Using the Backward-Forward Monte-Carlo Method”. In: *Journal of the Korean Physical Society* 59.23, p. 1218. ISSN: 0374-4884. DOI: 10.3938/jkps.59.1218 (cit. on p. 164).
- Bauge, E., S. Hilaire, and P. Dossantos-Uzarralde (2007). “Evaluation of the Covariance Matrix of Neutronic Cross Sections with the Backward-Forward Monte Carlo Method”. In: Int. Conf. Nuclear Data for Science and Technology. EDP Sciences. DOI: 10.1051/ndata:07339 (cit. on pp. 2, 61–63, 164).
- Bayes, T. (1763). “An Essay towards Solving a Problem in the Doctrine of Chances.” In: *Philosophical Transactions (1683-1775)* 53, pp. 370–418. ISSN: 0260-7085. JSTOR: 105741 (cit. on p. 2).
- Blight, B. J. N. and L. Ott (1975). “A Bayesian Approach to Model Inadequacy for Polynomial Regression”. In: *Biometrika* 62.1, p. 79. ISSN: 00063444. DOI: 10.2307/2334489. JSTOR: 2334489?origin=crossref (cit. on p. 191).
- Brooks, S., ed. (2011). *Handbook for Markov chain Monte Carlo*. Boca Raton: Taylor & Francis. 592 pp. ISBN: 978-1-4200-7941-8 (cit. on p. 139).
- Byrd, R. H. et al. (1995). “A Limited Memory Algorithm for Bound Constrained Optimization”. In: *SIAM Journal on Scientific Computing* 16.5, pp. 1190–1208. ISSN: 1064-8275, 1095-7197. DOI: 10.1137/0916069 (cit. on pp. 77, 112, 205, 229).
- Capote, R. and D. L. Smith (2008). “An Investigation of the Performance of the Unified Monte Carlo Method of Neutron Cross Section Data Evaluation”. In: *Nuclear Data Sheets* 109.12, pp. 2768–2773. ISSN: 00903752. DOI: 10.1016/j.nds.2008.11.007 (cit. on pp. 68, 135, 138, 141, 142, 173).
- Capote, R., D. L. Smith, et al. (2012). “A New Formulation of the Unified Monte Carlo Approach (UMC-B) and Cross-Section Evaluation for the Dosimetry

- Reaction $^{55}\text{Mn}(n,g)^{56}\text{Mn}$ ". In: *Journal of ASTM International* 9.3, p. 104115. ISSN: 1546962X. DOI: 10.1520/JAI104115 (cit. on pp. 2, 67, 69, 70, 142–145).
- Casella, G. (1985). "An Introduction to Empirical Bayes Data Analysis". In: *The American Statistician* 39.2, p. 83. ISSN: 00031305. DOI: 10.2307/2682801. JSTOR: 2682801?origin=crossref (cit. on p. 203).
- Chadwick, M. B. et al. (2006). "ENDF/B-VII.0: Next Generation Evaluated Nuclear Data Library for Nuclear Science and Technology". In: *Nuclear Data Sheets* 107.12, pp. 2931–3060. ISSN: 00903752. DOI: 10.1016/j.nds.2006.11.001 (cit. on p. 27).
- Cox, R. T. (1946). "Probability, Frequency and Reasonable Expectation". In: *American Journal of Physics* 14.1, p. 1. ISSN: 00029505. DOI: 10.1119/1.1990764 (cit. on p. 2).
- (2001). *Algebra of Probable Inference*. Baltimore Md: Johns Hopkins Press: Johns Hopkins University Press. 127 pp. ISBN: 978-0-8018-6982-2 (cit. on p. 9).
- Dowle, M. et al. (2014). *data.table: Extension of data.frame*. R package version 1.9.4. URL: <https://github.com/Rdatatable/data.table/wiki> (visited on 07/01/2015) (cit. on p. v).
- Evensen, G. (1994). "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model using Monte Carlo Methods to Forecast Error Statistics". In: *Journal of Geophysical Research* 99 (C5), p. 10143. ISSN: 0148-0227. DOI: 10.1029/94JC00572 (cit. on p. 77).
- Felsenstein, K. (2006). "Bayes Statistik. Lecture Notes of a Master Course in Statistics for Mathematicians" (cit. on p. 246).
- Filatenkov, A. A. (1999). *Leningrad Reports*. 252. Khlopin Radium Institute (cit. on p. 225).
- (2001). *Leningrad Reports*. 258. Khlopin Radium Institute (cit. on p. 225).
- Finlay, R. et al. (1993). "Neutron Total Cross Sections at Intermediate Energies". In: *Physical Review C* 47.1, pp. 237–247. DOI: 10.1103/PhysRevC.47.237 (cit. on pp. 110, 134, 206, 238).
- Frehaut, J. (1980). "Status of (n,2n) Cross Section Measurements at Bruyeres-Le-Chatel". In: *Symposium on neutron cross-sections from 10 to 50 MeV*. Vol. 1. New York: Brookhaven National Laboratory, p. 399 (cit. on p. 225).
- Hadamard, J. (1893). "Etude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann". In: *Journal de Mathématiques Pures et Appliquées*, pp. 171–216. ISSN: 0021-7874 (cit. on p. 37).
- Han-Lin, L., Z. Wen Rong, and F. P. Guo (1985). "Measurement of the Neutron Cross Sections for the Reactions $^{169}\text{(n, 2n)}^{168}\text{Tm}$, $^{169}\text{Tm}(n,3n)^{167}\text{Tm}$, and $^{181}\text{Ta}(n,2n)^{180m}\text{Ta}$ ". In: *Nuclear Science and Engineering* 90.3, pp. 304–310. ISSN: 00295639. DOI: 10.13182/NSE85-3 (cit. on p. 225).

- Hastings, W. K. (1970). “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57, p. 97 (cit. on p. 137).
- Herman, M. et al. (2007). “EMPIRE: Nuclear Reaction Model Code System for Data Evaluation”. In: *Nuclear Data Sheets* 108.12, pp. 2655–2715. ISSN: 00903752. DOI: 10.1016/j.nds.2007.11.003 (cit. on pp. 3, 70, 115).
- Ikeda, Y. (1988). *JAERI Reports*. 1312 (cit. on p. 225).
- Jaynes, E. T. (1957a). “Information Theory and Statistical Mechanics”. In: *Physical Review* 106.4, pp. 620–630. ISSN: 0031-899X. DOI: 10.1103/PhysRev.106.620 (cit. on pp. 2, 7, 14, 22, 24).
- (1957b). “Information Theory and Statistical Mechanics. II”. In: *Physical Review* 108.2, pp. 171–190. ISSN: 0031-899X. DOI: 10.1103/PhysRev.108.171 (cit. on p. 7).
- (1968). “Prior Probabilities”. In: *IEEE Transactions on Systems Science and Cybernetics* 4, pp. 227–241 (cit. on p. 7).
- (2003). *Probability Theory: The Logic of Science*. Ed. by G. Larry Bretthorst. Cambridge, UK ; New York, NY: Cambridge University Press. 753 pp. ISBN: 978-0-521-59271-0 (cit. on p. 61).
- Kasugai, Y. (1992). *JAERI-M reports*. 93-046, p. 277 (cit. on p. 225).
- Kawano, T. and K. Shibata (1997). *Covariance Evaluation System* (cit. on pp. 2, 71, 115).
- Kolmogorov, A. N. (2000). *Foundations of the Theory of Probability*. 2., Engl. ed., reprint. Providence, RI: Amer. Mathem. Soc. 84 pp. ISBN: 0-8218-2648-4 0-8218-2648-4 (cit. on p. 8).
- Koning, A. J. (2015). “Bayesian Monte Carlo Method for Nuclear Data Evaluation”. In: *Nuclear Data Sheets* 123, pp. 207–213. ISSN: 00903752. DOI: 10.1016/j.nds.2014.12.036 (cit. on pp. 2, 28, 64, 154–156, 162).
- Koning, A. J. and J. P. Delaroche (2003). “Local and Global Nucleon Optical Models from 1 keV to 200 MeV”. In: *Nuclear Physics A* 713.3–4, pp. 231–310. ISSN: 0375-9474. DOI: 10.1016/S0375-9474(02)01321-0 (cit. on pp. 3, 109, 110, 134, 206, 224, 225, 238, 242, 244).
- Koning, A. J., S. Hilaire, and M. C. Duijvestijn (2007). “TALYS-1.0”. In: EDP Sciences. DOI: 10.1051/ndata:07767 (cit. on pp. 225, 242).
- Koning, A. J., S. Hilaire, and S. Goriely (2013). *TALYS-1.6 - A Nuclear Reaction Program*. URL: <http://www.talys.eu> (visited on 07/01/2015) (cit. on pp. 111, 242).
- Koning, A. J. and D. Rochman (2008). “Towards Sustainable Nuclear Energy: Putting Nuclear Physics to Work”. In: *Annals of Nuclear Energy* 35.11, pp. 2024–2030. ISSN: 03064549. DOI: 10.1016/j.anucene.2008.06.004 (cit. on p. 75).

- Koning, A.J., S. Hilaire, and M.C. Duijvestijn (2008). “TALYS-1.0”. In: *Proceedings of International Conference on Nuclear Data for Science and Technology*. Nice, France: EDP Sciences, pp. 211–214 (cit. on pp. 28, 77).
- Krige, D. G. (1951). “A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand”. In: *Journal of the Chemical, Metallurgical and Mining Society* 52, pp. 119–139 (cit. on p. 190).
- Kullback, S. and R. A. Leibler (1951). “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729694 (cit. on p. 24).
- Larson, N. M. (1998). *Updated user’s guide for SAMMY: Multilevel R-Matrix Fits to Neutron Data Using Bayes’ Equation, Revision 4*. ORNL/TM-9179/R4. Oak Ridge: Oak Ridge National Laboratory (cit. on pp. 35, 71, 116).
- Leeb, H., D. Neudecker, and Th. Srdinko (2008). “Consistent Procedure for Nuclear Data Evaluation Based on Modeling”. In: *Nuclear Data Sheets* 109.12, pp. 2762–2767. ISSN: 00903752. DOI: 10.1016/j.nds.2008.11.006 (cit. on pp. 71, 73, 175, 184).
- Leeb, H., G. Schnabel, et al. (2015). “Bayesian Evaluation Including Covariance Matrices of Neutron-induced Reaction Cross Sections of ^{181}Ta ”. In: *Nuclear Data Sheets* 123, pp. 153–158. ISSN: 00903752. DOI: 10.1016/j.nds.2014.12.027 (cit. on p. 224).
- Markovskij, D. V. et al. (2003). “Integral Benchmark Experiments with 14-MeV Neutrons for Testing the Nuclear Data of Vanadium”. In: *Fusion Engineering and Design*. 22nd Symposium on Fusion Technology 69.1–4, pp. 419–423. ISSN: 0920-3796. DOI: 10.1016/S0920-3796(03)00084-X (cit. on pp. 3, 96).
- Matheron, G. (1962). *Traité de géostatistique appliquée*. Vol. 14. Paris: Editions Technip (cit. on p. 190).
- Metropolis, N. et al. (1953). “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6, p. 1087. ISSN: 00219606. DOI: 10.1063/1.1699114 (cit. on p. 137).
- Mises, R. von and H. Geiringer (1966). “Mathematical Theory of Probability and Statistics.” In: *Journal of the Royal Statistical Society. Series A (General)* 129.2, p. 289. ISSN: 00359238. DOI: 10.2307/2343628. JSTOR: 10.2307/2343628?origin=crossref (cit. on p. 20).
- Muir, D. W. et al. (2007). “The Global Assessment of Nuclear Data, GANDR”. In: EDP Sciences. DOI: 10.1051/ndata:07635 (cit. on pp. 2, 75).
- National Nuclear Data Center (2015a). *Evaluated Nuclear Structure Data File Retrieval*. URL: <http://www.nndc.bnl.gov/ensdf/> (visited on 07/01/2015) (cit. on p. 244).

- (2015b). *Experimental Nuclear Reaction Data (EXFOR)*. URL: <https://www-nds.iaea.org/exfor/exfor.htm> (visited on 07/01/2015) (cit. on p. 110).
- Neudecker, D. (2012). “The Full Bayesian Evaluation Technique and its Application to Isotopes of Structural Materials”. PhD thesis. Austria: Technische Universität Wien. 188 pp. (cit. on pp. 2, 3, 70).
- Otuka, N. et al. (2014). “Towards a More Complete and Accurate Experimental Nuclear Reaction Data Library (EXFOR): International Collaboration Between Nuclear Reaction Data Centres (NRDC)”. In: *Nuclear Data Sheets* 120, pp. 272–276. ISSN: 00903752. DOI: [10.1016/j.nds.2014.07.065](https://doi.org/10.1016/j.nds.2014.07.065) (cit. on p. 110).
- Peiguo, F. (1985). In: *Chinese Journal of Nuclear Physics* 7.3, p. 242 (cit. on p. 225).
- Petersen, K. B. et al. (2006). *The Matrix Cookbook* (cit. on p. 205).
- Pigni, M. T. and H. Leeb (2003). “Uncertainty Estimates of Evaluated ^{56}Fe Cross Sections Based on Extensive Modelling at Energies Beyond 20 MeV”. In: *Proc. Int. Workshop on Nuclear Data for the Transmutation of Nuclear Waste. GSI-Darmstadt, Germany* (cit. on pp. 73, 175, 182).
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing (cit. on p. v).
- Raiffa, H. and R. Schlaifer (1961). *Applied Statistical Decision Theory*. Harvard University Graduate School of Business Administration (Division of Research); Bailey & Swinfen (cit. on p. 13).
- Rasmussen, Carl Edward and Christopher K. I Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, Mass.: MIT Press. ISBN: 0-262-18253-X 978-0-262-18253-9 (cit. on p. 192).
- Rochman, D., A. J. Koning, E. Bauge, et al. (2014). “From Flatness to Steepness: Updating TALYS Covariances with Experimental Information”. In: *Annals of Nuclear Energy* 73, pp. 7–16. ISSN: 03064549. DOI: [10.1016/j.anucene.2014.06.016](https://doi.org/10.1016/j.anucene.2014.06.016) (cit. on pp. 175, 178).
- Rochman, D., A. J. Koning, D. F. da Cruz, et al. (2010). “On the Evaluation of ^{23}Na Neutron-induced Reactions and Validations”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 612.2, pp. 374–387. ISSN: 0168-9002. DOI: [10.1016/j.nima.2009.10.147](https://doi.org/10.1016/j.nima.2009.10.147) (cit. on p. 75).
- Rochman, D., A. J. Koning, and S. C. van der Marck (2009). “Uncertainties for Criticality-Safety Benchmarks and Distributions”. In: *Annals of Nuclear Energy* 36.6, pp. 810–831. ISSN: 0306-4549. DOI: [10.1016/j.anucene.2009.01.018](https://doi.org/10.1016/j.anucene.2009.01.018) (cit. on p. 75).

- Schnabel, G. and H. Leeb (2015). “A New Module for Large Scale Bayesian Evaluation in the Fast Neutron Energy Region”. In: *Nuclear Data Sheets*. Special Issue on International Workshop on Nuclear Data Covariances April 28 - May 1, 2014, Santa Fe, New Mexico, USA <http://t2.lanl.gov/cw2014> Special Issue on International Workshop on Nuclear Data Covariances 123, pp. 196–200. ISSN: 0090-3752. DOI: 10.1016/j.nds.2014.12.034 (cit. on p. 77).
- Shannon, C. E. (1948a). “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3, pp. 379–423. ISSN: 00058580. DOI: 10.1002/j.1538-7305.1948.tb01338.x (cit. on p. 22).
- (1948b). “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.4, pp. 623–656. ISSN: 00058580. DOI: 10.1002/j.1538-7305.1948.tb00917.x (cit. on pp. 22, 23).
- Smith, D. L. (1991). *Probability, Statistics, and Data Uncertainties in Nuclear Science and Technology*. Neutron physics and nuclear data in science and technology v. 4. LaGrange Park, Ill: American Nuclear Society. 269 pp. ISBN: 0-89448-036-7 (cit. on p. 56).
- Smith, D. L. (2004). *Covariance Matrices for Nuclear Cross Sections Derived from Nuclear Model Calculations*. ANL/NDM-159. Argonne National Lab., IL (USA) (cit. on pp. 2, 67).
- Stein, M. L. (1999). *Interpolation of Spatial Data Some Theory for Kriging*. New York, NY: Springer New York. ISBN: 978-1-4612-7166-6 1-4612-7166-5 978-1-4612-1494-6 1-4612-1494-7 (cit. on p. 195).
- Takahashi, A (1992). *OKTAVIAN report*. 92. Osaka University, p. 1 (cit. on p. 225).
- Uhlenbeck, G. E. and L. S. Ornstein (1930). “On the Theory of the Brownian Motion”. In: *Physical Review* 36.5, pp. 823–841. ISSN: 0031-899X. DOI: 10.1103/PhysRev.36.823 (cit. on p. 183).
- Vaart, A. W. van der (1998). *Asymptotic Statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press. 443 pp. ISBN: 0-521-49603-9 (cit. on p. 29).
- Veeser, L. R., E. D. Arthur, and P. G. Young (1977). “Cross Sections for (n, 2 n) and (n, 3 n) Reactions above 14 MeV”. In: *Physical Review C* 16.5, p. 1792 (cit. on p. 225).
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer New York. ISBN: 978-0-387-98140-6 (cit. on p. v).
- Woodbury, M. A. (1950). *Inverting Modified Matrices*. Statistical Research Group, Memo. Rep. no. 42. Princeton University, Princeton, N. J. 4 pp. (cit. on pp. 35, 248).

- Zhu, C. et al. (2011). “Measurements of (n, 2n) Reaction Cross Sections at 14 MeV for Several Nuclei”. In: *Nuclear Science and Engineering* 169.2, pp. 188–197. ISSN: 00295639. DOI: 10.13182/NSE10-35 (cit. on p. 225).