TU WIEN

**TECHNISCHE
UNIVERSITÄT
WIEN**
Vienna University of Technology

DIPLOMARBEIT

Extension of the Full Bayesian Evaluation Technique to Differential Cross Sections

Ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Diplom-Ingenieurs unter der Leitung von

Ao. Univ. Prof. Helmut Leeb
Inst. Nr.: 141
Atominstitut

eingereicht an der Technischen Universität Wien
Fakultät für Physik

von
Georg Schnabel
Matrikelnummer: e0525028
Weißes Kreuz-Gasse 57, A-2340 Mödling

Wien, März 2012

Kurzfassung

In Kerndatenevaluationen werden experimentelle Daten und physikalische Modelle zu Datensätzen für nukleare Anwendungen wie zum Beispiel die Verbrennung von nuklearem Müll, Dosimetrie oder medizischen Anwendungen kombiniert. Die von Evaluationsmethoden erstellten Datensätze sollen die in den Evaluationsprozess einbezogenen experimentellen Daten wiedergeben und in Energiebereichen, in denen diese fehlen, vernünftige Vorhersagen treffen.

Die vollständige Bayesianische Evaluations-Methode [26] bewältigt diese Anforderungen unter Verwendung von Bayesscher Statistik. Die evaluierten Daten bestehen aus integralen Wirkungsquerschnitten und Kovarianzmatrizen für einfallende Neutronen oberhalb des Resonanzbereichs bis zu 150 MeV. Der eingesetzte Prior ist durch eine multidimensionale Normalverteilung gegeben. Die ihn definierenden statistischen Kenngrößen, ein Mittelwert-Vektor und eine Kovarianzmatrix, werden durch ein Stichprobenverfahren ermittelt, das sich auf Ergebnisse des nuklearen Modell-Programms TALYS [23] stützt.

In dieser Arbeit wurde die vollständige Bayesianische Evaluations-Methode erweitert, um auch die Einbeziehung differentieller Wirkungsquerschnittsdaten in Evaluationen zu ermöglichen. Auf diese Weise soll das vorhandene experimentelle Datenmaterial besser ausgeschöpft werden. Die Eigenschaften der erweiterten vollständigen Bayesianischen Evaluations-Methode wurden allgemein studiert und die Anwendbarkeit am Beispiel von Tantal-181 für den totalen und den differentiellen, elastischen Wirkungsquerschnitt geprüft.

Es konnte gezeigt werden, dass die Annahme einer Normalverteilung die Systematik der in TALYS implentierten Modelle auch für den differentiellen, elastischen Wirkungsquerschnitt adequat beschreibt. Als Nebenprodukt wurde eine Methode entdeckt, um das Resultat von Evaluationen weniger abhängig von experimentellen Daten zu machen, die als unphysikalische Ausreißer zu betrachten sind. Die beispielhafte Anwendung der vollständigen Bayesianischen Evaluations-Methode an Tantal-181 zeigte den Einfluss der Prior-Erwartungshaltung auf das Ergebnis. Die starken Korrelationen im Prior, in Kombination mit der großen Menge an experimentellen Datenpunkten, führten zu evaluierten Fehlern in der Größenordnung des systematischen Fehler der Experimente. Weiters war ein Großteil der experimentellen Daten nicht im resultierenden 1-σ Konfidenzintervall enthalten. Dieser Effekt kann im Falle des differentiellen, elastischen Wirkungsquerschnitts vermutlich zu einem Teil auf die Prior-Erwartungshaltung bezüglich des Verhältnisses zwischen integralem elastischen- und Reaktionswirkungsquerschnitt zurück geführt werden, die im Widerspruch zu den experimentellen Daten steht. Der allgemeine Grund für das Auftreten dieses Effekts wurde tiefgehend analysiert. Wesentliche Aspekte, die weiterer Untersuchung bedürfen, wurden aufgezeigt. Das sind im Besonderen die Gültigkeit der verwendeten physikalischen Modelle zur Beschreibung der differentiellen Wirkungsquerschnittskurven im Bereich großer Winkel und die Verfeinerung der Heuristiken zur Festlegung der Grenzen für die Parameter des optischen Potentials.

Abstract

In Nuclear Data evaluations, experimental data are combined with model knowledge to form suitable data sets required for the development of nuclear applications, e.g. nuclear waste incineration, dosimetry and medical applications. The data sets produced by evaluation methods should be in accordance with the included experimental data and give reasonable predictions at energies where experimental data is missing.

The Full Bayesian Evaluation Technique (FBET) [26] tackles these demands by means of Bayesian statistics. It provides angle-integrated neutron cross sections and covariance matrices from the unresolved energy range up to 150 MeV. The prior is assumed to be of Gaussian shape. Its defining quantities, a vector of means and a covariance matrix, are obtained by a sampling procedure based on the nuclear models code TALYS [23].

The goal of this thesis was to extend the Full Bayesian Evaluation Technique to allow the inclusion of experimental angle differential cross section data in order to make more efficient use of the information available. General features of the extended method were studied and its applicability was tested for tantalum-181.

It could be shown that the prior sufficiently incorporates the systematics of the physical model also for the angle differential elastic CS channel. As a side product, a procedure to adapt the prior to reduce the sensitivity of the FBET to experimental outliers was discovered. The evaluation of tantalum-181 revealed the prior expectation to have a noticeable influence on the evaluated data sets. Due to strong model correlations and a large number of included experimental data points, the uncertainties of the evaluated cross sections dropped to the overall normalization error of the experiments and the bulk of the experimental data was not contained within the 1-σ confidence intervals. In the case of the differential elastic CS, the occurrence of this effect may be partially attributed to the expectation of the prior about the ratio of the integral elastic CS to the reaction CS, which is in disagreement with the included experimental data. The reason for the occurrence of this effect in general was discussed in detail and the key aspects which require further investigation were determined, i.e. the validity of physical models for differential cross sections in the large angle domain and the refinement of the heuristics applied for the specification of the optical potential parameter boundaries used for the construction of the prior.

Acknowledgements

I want to thank my supervisor Prof. Helmut Leeb for his great support throughout the development process of this thesis. He was always available for valuable discussion which enriched my view on the topic and frequently led to new ideas. In the final stage, I profited a lot from his advice on how to precisely formulate statements.

I am very grateful to Veronika Doblhoff. She dedicated her time to proofread my thesis. Her corrections and advice helped me to considerable improve its quality. I am ever thankful for her support.

Contents

1 Introduction

The discovery of radioactivity by Henri Becquerel and further work by Pierre and Marie Currie at the end of the 19$^{\text{th}}$ century can be regarded as the starting point of nuclear physics. Since then, research in this field has lead to many important technical applications such as nuclear reactors for energy supply or imaging methods in medicine. Albeit the emphasis is on the development of the renewable energy technology at present, the efforts to develop novel nuclear technologies are also required to satisfy environmental demands and the energy needs of the future. Aspects to be improved are safety measures, efficiency and the reduction of radioactive waste. Additionally, there is the optimistic hope to have attained enough knowledge, desirably in a few decades, to construct fusion reactors capable to satisfy a good fraction of tomorrow's energy needs.

Advancement in these issues requires profound knowledge of the structural properties of the nuclei. Especially, accurate estimates for different types of cross sections are needed as they allow for reactor simulations on the computer, which saves costs required for the development of new nuclear technologies.

Not only for this reason, nuclear data libraries such as JEFF-3.1.1 [2] in Europe or ENDF/B-VII.1 [5] in the U.S. were set up to concentrate all the pieces of knowledge obtained at various facilities. The data included in these libraries are the result of evaluation processes in which physical models are combined with experimental data. Thus, the systematic knowledge distilled from experiments is taken into account in order to make the best use of the information available. At the moment, however, important data needed for the design of new reactor types, as e.g. the (n,n) cross sections, are still unsatisfactorily known for the energy range from 20 MeV to 200 MeV. Therefore, evaluations are required, which give reasonable extrapolated estimates and associated uncertainties in terms of covariance matrices for this energy range. For this purpose, several evaluation techniques exist and different techniques are used at different facilities; until now no standard procedure for uncertainty estimation has been established. An introduction to the various approaches is given in [3].

One particular method developed by H. Leeb at al. [24] is the Full Bayesian Evaluation Technique (FBET) which provides by means of Bayesian inference estimates of cross sections based on both model and experimental data. For obtaining model data, the nuclear models code implemented in TALYS [23] is used. Since every model is only an approximation to reality, the method also accounts for systematic deviations of the model from the experimental data expressed in terms of covariance matrices, which are obtained by a well defined procedure.

Currently, only integral cross sections are taken into account. This work is the attempt to extend the FBET to include also angle differential cross section data. Thus, the information contained in the angular differential data become accessible to the evaluation

and it is expected that considering differential cross section data also improves the estimates for integral cross sections. Studying whether this is true is one issue of this thesis. The general issue of this thesis is to study the properties of the extended FBET and its applicability to nuclear data evaluation.

In the current implementation of the FBET, the systematics of the physical model implemented in TALYS are captured in terms of covariance matrices. These are able to describe only linear relationships between cross sections at different energies and angles. Therefore, the applicability of covariances as uncertainty measures depends on whether the underlying physical model can be approximated by linear relationships. This issue was examined for the total cross section and the angle differential elastic cross section. Another criterion for the applicability of the FBET is the validity of the underlying physical models. Thus, the effects of an inadequate model on an evaluation were also studied.

In order to test the extended FBET, an evaluation of tantalum-181 total neutron cross section and the angle differential elastic cross section with data from Smith [36] and Finlay [9] was performed. Since both Smith and Finlay gathered a huge amount of experimental data points, a comparison of the experimental data and the evaluated data was possible over a great energy range in the case of the total cross section and over a great angular range in the case of the angle differential elastic cross section.

In the past, the FBET had been criticised for producing unphysical kinks or oscillations in the evaluated cross section curves when outliers are present in the experimental data. Therefore, it was also studied how the method could be adapted to make it less sensitive to outliers.

The structure of this thesis is as follows: Chapter 2 outlines basic statistical definitions and introduces the framework of Bayesian inference. Furthermore, the principle of maximum entropy is presented, which is used to assign probabilities to possible values for the cross sections before experimental data is entered into the evaluation process. Chapter 3 summarizes the Full Bayesian Evaluation Technique as implemented until now, thereby showing the actual application of Bayesian statistics. A few remarks and a brief discussion of conceptual details of the method end this chapter.

Chapter 4 and 5 are the main parts of this thesis. Implementation details of the above mentioned extensions are discussed and the performance and validity of the extended method is tested for tantalum-181.

2 Bayesian statistics as methodical framework

The task to create estimates and associated uncertainty boundaries using multiple information sources – in this work given by model data and various experiments – each of which comprises some uncertainty about the true value can be tackled by methods for statistical inference. By the choice which sources are used and how they enter the inference process, one inevitably follows the methodology of one of two statistical schools of thought, Bayesianism or Frequentism. Both schools work with the same set of equations for operating on probabilities and for both of them, a probability is a real number between zero and one. However, they differ in their interpretation of the term 'probability'. Frequentists define probability as the limit of the relative frequency of an event when the number of trials goes to infinity. Thus, Frequentists regard probabilities as something measurable and restrict their assignment to events which could occur within a well-defined experiment that could be repeated arbitrarily many times. Bayesians think of probability as a measure for plausibility. So, according to their view, a probability is not a physical measureable quantity, but rather reflects the current state of knowledge.

The method used in this work incorporates experimental results and model data given by the nuclear model code TALYS [23] to produce improved estimates for cross section values. Results for cross sections obtained by model calculations are dependent on optical potential parameters characterizing the nucleus. For these parameters, reasonable boundaries have to be imposed and probabilities for intermediate values have to be assigned. Although these specifications are made by considering experimental results and applying reasonable principles, they are not absolutely compelling. Other choices may be equally justifiable. Hence, these probabilities have to be interpreted in the Bayesian sense as degree of belief or plausibility and consequently the method of this work utilizes Bayesian inference.

The next section briefly outlines the fundamental equations of probability theory which are the cornerstone of the Bayesian inference formula. This formula will be elaborated on thereafter.

2.1 Basics of probability theory

Several approaches exist to deduce the rules for dealing with probabilities. Two of them are the Kolmogorov axioms [21] and Cox's theorem [6]. Although the first approach relies on set and measure theory and regards events as elements in some set whereas the second considers propositions and is based on consistency postulates, they are – accord-

ing to E.T. Jaynes – equivalent [20]. The resulting rules are stated here in a Bayesian diction without derivation.

Laws of probability theory

Let A and B be propositions that could either be true or false. $P(X)$ is the plausibility for some proposition X; it can assume values in the closed interval $[0, 1]$; $P(X) = 1$ represents absolute certainty about the truth of proposition X and $P(X) = 0$ absolute certainty about its falseness. $P(A \,|\, B)$ expresses the plausibility of A under the assumption that B is true. $A \vee B$ stands for the composite proposition 'A or B' and $A \wedge B$ for 'A and B'. The following equations are available to compute plausibilities:

$$\textbf{sum rule} \qquad P(A \vee B) = P(A) + P(B) - P(A \wedge B) \qquad (2.1)$$

$$\textbf{product rule} \qquad P(A \wedge B) = P(A \,|\, B)P(B) \qquad (2.2)$$

The term plausibility was used here rather than probability to make clear that probabilities have to be understood in the Bayesian sense. Nevertheless, having mentioned that point, the more common term probability will be used exclusively from now on.

Noteworthy, these equations can be derived from Cox's theorem which itself was derived by demanding consistency and compatibility to common sense reasoning. Consistency means that whenever there are several ways to calculate a certain probability, all ways should lead to the same result. Compatibility to common sense means that if all statements are certain – each having a probability of 1 assigned to it – the equations should implement Aristotelian logic perfectly. Hence, these rules entail constraints on the assignment of probabilities to statements. Also, knowing about the incorporated consistency features and the possibility to model vague knowledge about the validity of propositions, these equations recommend themselves as foundation for a method of inference using incomplete information – Bayesian inference. The respective formula which will be discussed in more detail in the next section is contained in the Bayes theorem.

Theorem 2.1.1 Bayes equation: *A and B are propositions, $P(A)$ and $P(B) > 0$ their assigned probabilities, then their mutual conditional probabilities $P(A \,|\, B)$ and $P(B \,|\, A)$ relate to each other by*

$$P(A \,|\, B) = \frac{P(B \,|\, A)P(A)}{P(B)}. \qquad (2.3)$$

Proof: Exchange propositions A and B in the product rule (2.2). Because $P(A \wedge B) = P(B \wedge A)$, one can equate the right hand sides of original and altered equation.

Since this work deals with the estimation of different types of cross section which are continuous quantities, the above notion of probability has to be adapted. One can handle these infinitely many propositions by parameterising them, e.g.

$$A(t) \equiv \text{the total cross section of Tantal-181 is 't' millibarn}$$

t being a real number. To abbreviate the notation, instead of passing a proposition $A(t)$ as argument to the probability function P, only the real valued parameter t is passed to

a corresponding function ρ yielding the same probability

$$P(A(t)) = \rho(t). \tag{2.4}$$

Furthermore, the case of infinitely many statements requires $\rho(t)$ to be a probability density function instead of a probability function. Here, only a loose definition is given suited for the needs of this work:

Definition 2.1.1 *A function $\rho : \mathbb{R} \to [0, \infty)$ 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*

$$P(t_1 \le t \le t_2) = \int_{t_1}^{t_2} \rho(t) \, dt, \tag{2.5}$$

where $P(t_1 \le t \le t_2)$ 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.

As only one value is possible for a certain cross section, propositions stating different values to be the true one are necessarily mutually exclusive. Thus, the given definition is completely sufficient for the estimation task to be tackled. Because propositions in this work exclusively contain statements about numerical values of cross sections, it is henceforth assumed that propositions exclusively make statements about the value of some numerical quantity and that the real number indexing a proposition is exactly the value stated in that proposition for that quantity. Thus, the word proposition will be dropped in favour of quantity.

Even though the value of $\rho(t)$ for a given t cannot be interpreted as probability anymore, it still can be regarded as a measure for 'likeliness', meaning that if $\rho(t_1)$ is twice as high as $\rho(t_2)$ then the probability for the true value of t being located within a small interval centered around t_1 is approximately twice as high as the probability of it being located within an interval ofthe same length centered around t_2.

If there are two quantities to be considered, the density just gets another argument $\rho(s, t)$ where s is the value of some quantity q_1 and t of some quantity q_2. $\rho(s, t)$ expresses the likeliness that the true value of q_1 is s and that of q_2 is t simultaneously. Therefore it has the analogous meaning as $P(A \wedge B)$ for probabilities. The extension of (2.5) to the case of multiple quantities is straight-forward. The probability for s in $[s_1, s_2]$ and t in $[t_1, t_2]$ can be calculated through

$$P(s_1 \le s \le s_2 \wedge t_1 \le t \le t_2) = \int_{s_1}^{s_2} \int_{t_1}^{t_2} \rho(s, t) \, ds \, dt. \tag{2.6}$$

If only the probability for one quantity lying in a certain interval is of interest, independently of the other quantity, the following procedure – called marginalization – is applied:

$$P(t_1 \le t \le t_2) = \int_{t_1}^{t_2} \int_{-\infty}^{+\infty} \rho(s, t) \, ds \, dt \equiv \int_{t_1}^{t_2} \rho_s(t) \, dt, \tag{2.7}$$
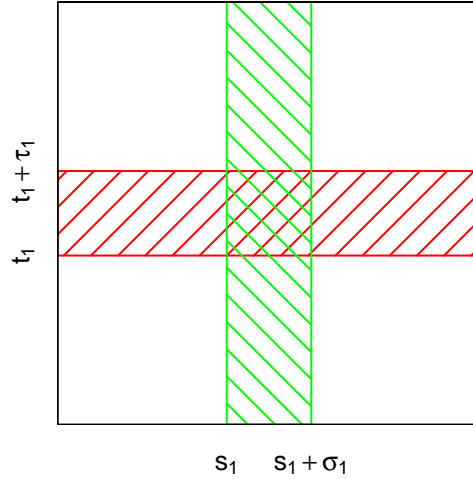
Figure 2.1: Venn diagram: A (green), B (red) and $A \wedge B$ (green & red)

where $\rho_s(t)$ is called the 'marginal probability density'. Restating Bayes Theorem (2.1.1) in terms of probability densities requires the introduction of the concept of 'conditional probability density'. For this purpose assume the following propositions

$$A \equiv s_1 < s < s_1 + \sigma_1 \text{ and t arbitrary} \tag{2.8}$$

$$B \equiv t_1 < t < t_1 + \tau_1 \text{ and s arbitrary} \tag{2.9}$$

with s_1, t_1 arbitrary real numbers and σ, τ_1 positive real numbers. Figure 2.1 visualizes the situation. Each point in the plane represents an elementary hypothesis, expressing that the considered quantities take the values s and t respectively. Composite propositions A (green) and B (red) are sets of elementary propositions. The area shared by both of them represents $A \wedge B$. The briefly introduced concept of conditional probability, written as $P(A \,|\, B)$, now gets an intuitive meaning. If proposition B is known to be true, then proposition A could only be realized by elementary propositions in the area covered by both A and B. $P(A \,|\, B)$ could be greater or less than $P(A)$ depending on the amount of elementary propositions from A being also contained in B and which probability is assigned to them in relation to the probability of proposition B, so

$$P(A \,|\, B) = \frac{P(A \wedge B)}{P(B)}, \tag{2.10}$$

which is the restated product rule (2.2). Inserting the integral expressions for $P(A \wedge B)$ (2.6) and $P(B)$ (2.7), one obtains

$$P(A|B) = \frac{\int_{s_1}^{s_1+\sigma_1} \int_{t_1}^{t_1+\tau_1} \rho(s,t) \, \mathrm{d}s \, \mathrm{d}t}{\int_{t_1}^{t_1+\tau_1} \int_{-\infty}^{+\infty} \rho(s,t) \, \mathrm{d}s \, \mathrm{d}t}. \tag{2.11}$$

6

When probabilities can be calculated by integrating over probability densities, then conversely the derivation of probabilities in respect to the integral upper boundary leads to probability densities. This allows obtaining the conditional probability density by taking the limit $P(A \mid B)/\sigma_1$, $\sigma_1 \to 0 \wedge \tau_1 \to 0$. Assuming σ_1 and τ_1 to be sufficient small, the first mean value theorem for integration yields

$$\frac{1}{\sigma_1} P(A \mid B) \approx \frac{1}{\sigma_1} \frac{\sigma_1 \tau_1 \rho(s_1, t_1)}{\tau_1 \int_{-\infty}^{+\infty} \rho(s, t_1) \, \mathrm{d}s} = \frac{\rho(s_1, t_1)}{\int_{-\infty}^{+\infty} \rho(s, t_1) \, \mathrm{d}s}. \tag{2.12}$$

Taking the limit $\sigma_1 \to 0, \tau_1 \to 0$ turns the approximation into an exact equation. Remembering the definition of a marginalized probability density (2.7), the conditional probability density $\rho(s \mid t)$ finally can be expressed as

$$\rho(s \mid t) = \frac{\rho(s, t)}{\rho_s(t)}, \tag{2.13}$$

which can also be regarded as the restated product rule for the continuous quantity case. Being aware that the same limit procedure applied to the conditional probability $P(B \mid A)$ gives

$$\rho(t \mid s) = \frac{\rho(s, t)}{\rho_t(s)}, \tag{2.14}$$

one can make $\rho(s, t)$ explicit in both equations and then equate them, which results in the Bayes theorem for the continuous case.

Theorem 2.1.2 Bayes equation for continuous quantities: *Let s and t be the values of two quantities, $\rho(s, t)$ their associated probability density, $\rho_s(t)$ the probability density for a certain value t without considering values of s (marginal probability density), and $\rho(s \mid t)$ the probability density for s under the assumption that the other quantity takes value t (conditional probability density), then the following equation relates the conditional probability densities*

$$\rho(t \mid s) = \frac{\rho(s \mid t)\rho_s(t)}{\rho_t(s)}. \tag{2.15}$$

Having established the Bayes equation for both the 'finite number of propositions' and the 'continuous quantity' case, the next section elaborates on how it can be used for an incremental inference procedure.

2.2 Bayesian inference

The Full Bayesian Evaluation Technique relies on Bayesian inference specialized on the needs of nuclear data evaluation. In order to ease comprehension, this section describes the basic working of Bayesian inference along with its common notions. Implementation details of the Full Bayesian Evaluation Technique for nuclear data evaluation are the topic of the next chapter.

Apart from strict logical deduction of propositions from given axioms, the validity of an inference is not absolutely compelling. This is a consequence of the fact that propositions from which useful conclusions can be drawn exhibit a rather general character, making it impossible to check their validity directly. For instance, one might believe the proposition 'Up until now no human has ever reached the age of 130 years' is almost certainly true. Hence, when asked about the life span of a particular deceased person, it would allow the conclusion 'not longer than 130 years' and it should be regarded as almost certainly true as well. Maybe the high probability of the proposition about the life span is based on the fact that one has never seen or heard about a person contradicting it. However, there is no final proof for its validity and perhaps newly acquired information may change that opinion to a lower probability tomorrow. In addition, the credibility of the information source could also be subject to an opinion. If one is very sure about the upper limit for the human life span and a bit in doubt about the credibility of the information source, one might doubt it even more.

This example for inferences on the basis of vague knowledge can be abstractly stated as follows: There are some preconceptions about a matter in form of propositions (hypotheses), each associated with a certain degree of belief (probability). Then on the basis of newly acquired information (data), the probabilities of the hypothesis should be adapted in accordance to obtained data.

This is exactly what the method of Bayesian inference accomplishes in a quantitative manner. It also assures that probabilities assigned to propositions always remain consistent to the principles of logic. For example, if there are two propositions contradicting each other and one of them is known to be certainly true, then every change in their probabilities due to newly acquired data must obey the constraint that they must add up to one.

The Bayesian inference formula (2.1.1) for a finite number of propositions has already been introduced in the last section. Using a modified notation for propositions to account for their meaning in the inference procedure leads to

$$P(H \,|\, D) = \frac{P(D \,|\, H)}{P(D)} P(H),\tag{2.16}$$

where H represents a hypothesis for which the probability should be updated due to the data D. $P(H)$ is the 'prior probability' for hypothesis H to be true. The term on the left hand side, $P(H \,|\, D)$, is the 'posterior probability' stating the new probability for H after inclusion of the data D in the Bayesian inference. The term $P(D \,|\, H)$ is called 'likelihood' and $P(D)$ 'evidence'. The ratio likelihood to evidence can be regarded as a scaling factor indicating how many times more likely the hypothesis got compared to before considering the new data.

The probability assignment for $P(H)$ and $P(D)$ is undetermined without further information. The assignment must be made by considering a-priori knowledge and experience of the experimenter. Therefore, different people may make different choices depending on their experience and their view on the topic. This represents the subjectivists view of Bayesian probabilities. On the other hand there is the objectivists approach, in which

prior probabilities have to be constructed by some principle which can be regarded rationally. The prior probability (density) assignment to nuclear potential parameters in this work is achieved by the maximum entropy principle described in the next section.

Because of the dependence of $P(H)$, $P(D)$ and the other terms on a-priori knowledge I, they are frequently expressed as conditional probabilities, e.g. $P(H) \rightarrow P(H \mid I)$. For simplicity's sake this is not done in this work, as it does not change the calculations, but is solely a reminder. Having stated it here explicitly, there is no necessity to be reminded of it during every formula.

Two examples – one for the finite number of propositions and one for the continuous quantity case – should demonstrate the features of Bayesian inference. First, assume following two hypotheses

$$H_V \equiv \text{guinea pigs eat exclusively vegetables}$$
$$H_M \equiv \text{guinea pigs eat exclusively meat}$$

Evidently they are mutual exclusive. Furthermore, assume they are also exhaustive, meaning that one of them must be true. These two conditions will also be fulfilled later on when tackling differential cross sections. Every other possibility such as 'They never eat' or 'They eat both of it' is assigned an a-priori probability of zero and no amount of data can change that. These assignments are part of the a-priori knowledge.

Now assume that one is rather sure about guinea pigs being vegetarians, so prior probabilities might be $P(H_V) = 0.8$ and $P(H_M) = 0.2$. Data are given by several guinea pig observations where they have eaten either vegetables V or meat M. The conditional probabilities necessary for calculating the likelihoods are $P(V \mid H_V) = 1, P(M \mid H_V) = 0, P(V \mid H_M) = 0, P(M \mid H_M) = 1$. By relying on the sum rule, the product rule and logical conversion of propositions therein one can calculate $P(D)$ by

$$P(D) = P(D \mid H_V)P(H_V) + P(D \mid H_M)P(H_M). \tag{2.17}$$

If a guinea pig eating a vegetable was observed, the application of the Bayesian update formula for both hypotheses would yield

$$P(H_V \mid V) = \frac{P(V \mid H_V)}{P(V \mid H_V)P(H_V) + P(V \mid H_M)P(H_M)}P(H_V) = 1 \tag{2.18}$$

$$P(H_M \mid V) = \frac{P(V \mid H_M)}{P(V \mid H_V)P(H_V) + P(V \mid H_M)P(H_M)}P(H_M) = 0 \tag{2.19}$$

In accordance with the stated hypotheses, one observation would be enough to decide between them with absolute certainty. However, if the hypothesis of mixed eating had been included, probability changes would have been smaller and a probability of one could have never been reached by H_V or H_M, because no amount of e.g. observations of vegetable eating guinea pigs could exclude the mixed eating hypothesis, as one single contrary observation would suffice to make it true.

The situation is slightly modified if propositions related to continuous quantities are considered. Let one integral reaction cross section serve as an example. Each distinct value for x shall represent a hypothesis expressing x to be the true value. To every distinct x, a prior probability density has to be assigned. This can be done for all values of x at once by setting up the prior probability density function (pdf). In the Full Bayesian Evaluation Technique a multi-dimensional Gauss distribution is chosen. A one-dimensional Gaussian pdf is assumed for the current example:

$$\rho(x \mid \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}) = \frac{1}{\sigma_{\mathrm{mod}}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{\mathrm{mod}}^2}(x - \mu_{\mathrm{mod}})^2\right). \tag{2.20}$$

The notation $\rho(x \mid \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}})$ was used in order to clarify that the shape of the prior pdf is parameterised by the so called 'hyperparameters' μ_{mod} and σ_{mod} standing for the mean and standard deviation respectively. To set up the prior properly, definite values must be selected for the hyperparameters. Appropriate choices can be made guided by former experimental results. How this is done in the Full Bayesian Evaluation Technique is sketched in the next chapter.

The term in the numerator, previously introduced as likelihood, yields the probability density with which an experimental value is obtained under the assumption that a specific value for x is the true one. Experimental data is given in terms of an estimate for the true value μ_{exp} and its assigned standard error σ_{exp}, indicating how accurate the estimate is. It is nearly always assumed that this estimate obeys a Gauss distribution, thus

$$\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} \mid x, \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}) = \frac{1}{\sigma_{\mathrm{exp}}\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{\mathrm{exp}}^2}(x - \mu_{\mathrm{exp}})^2\right). \tag{2.21}$$

A point worth mentioning is the symmetry in the density $\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} \mid x) = \rho(x \mid \mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}})$. The last missing piece is the denominator, the evidence, $\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} \mid \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}})$. Because the generic hypotheses for different cross section values are mutually exclusive and exhaustive, this term may be calculated – in analogy to (2.17) – by

$$\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} \mid \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}) = \int_{-\infty}^{+\infty} \rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} \mid x)\, \rho(x \mid \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}})\, \mathrm{d}x \tag{2.22}$$

With the choice $s \equiv \mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}}$, $t \equiv x$ and using the product rule $\rho(s,t) = \rho(s|t)\rho(t)$, one sees that this is really the marginal probability density $\rho_t(s)$ as defined in (2.7).

Therefore, the general Bayesian update formula reads

$$\rho(x | \mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}}, \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}) = \frac{\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} | x, \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}})}{\rho(\mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}} | \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}})} \rho(x | \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}). \tag{2.23}$$

With all expressions plugged in and all constants outside the exponential function (independent of x) subsumed in N, this becomes

$$\rho(x | \mu_{\mathrm{exp}}, \sigma_{\mathrm{exp}}, \mu_{\mathrm{mod}}, \sigma_{\mathrm{mod}}) = N \exp\left(-\frac{1}{2\sigma_{\mathrm{exp}}^2}(x - \mu_{\mathrm{exp}})^2 - \frac{1}{2\sigma_{\mathrm{mod}}^2}(x - \mu_{\mathrm{mod}})^2\right). \tag{2.24}$$

Since a density integrated over the complete real range must yield one, the normalization factor N can be determined afterwards. Consequently, there is no need to compute the evidence (2.22) explicitly. The product of two Gauss distributions with arbitrary means and standard deviations yields a Gauss distribution again, which can be seen by quadratic extension. It is said that the prior is conjugated to the posterior with respect to the likelihood. Thus, knowing the posterior to be Gaussian, it suffices to calculate its mean and variance (standard deviation squared) from the mean and variance of the likelihood and prior to specify it uniquely:

$$\mu_{\mathrm{new}} = \frac{\mu_{\mathrm{exp}}\,\sigma_{\mathrm{mod}}^2 + \mu_{\mathrm{mod}}\,\sigma_{\mathrm{exp}}^2}{\sigma_{\mathrm{exp}}^2 + \sigma_{\mathrm{mod}}^2} \quad \text{and} \quad \sigma_{\mathrm{new}}^2 = \frac{\sigma_{\mathrm{exp}}^2 \sigma_{\mathrm{mod}}^2}{\sigma_{\mathrm{exp}}^2 + \sigma_{\mathrm{mod}}^2}. \tag{2.25}$$

The posterior mean is the weighted average of likelihood and prior mean, where each of them is weighted by the variance of the other. The new variance resembles the 'reduced mass' in mechanics. If σ_{exp} and σ_{mod} differ in the order of magnitude, σ_{new} is very accurately given by the variance having the smaller value. Moreover, σ_{new} is always smaller than σ_{mod} and σ_{exp}.

Updating with the data from several experiments can be performed sequentially, where in each update step the posterior distribution parameters μ_{new} and σ_{new} of the previous update step are taken as new hyperparameters for the prior distribution.

Remarkably, formula (2.24) is symmetric with respect to the prior information (later called model data) and the experimental data. Hence, one may regard the experimental data as prior information and the model data as likelihood; or one may lay down the
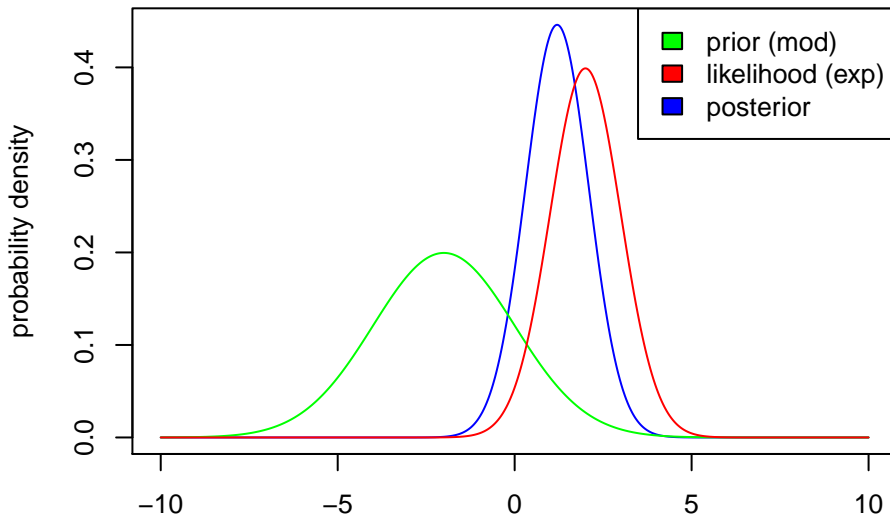

Figure 2.2: The Bayesian update formula updates the prior probability density (green) by inclusion of experimental data (red) to the posterior probability density (blue); prior: $\mu_{\mathrm{mod}} = -2, \sigma_{\mathrm{mod}} = 2$, likelihood: $\mu_{\mathrm{exp}} = 2, \sigma_{\mathrm{exp}} = 1$, posterior: $\mu_{\mathrm{new}} = 1.2, \sigma_{\mathrm{new}} \approx 0.89$

11

notions of prior and likelihood completely and think of several information sources being on equal footing. Each of them predicts a certain value to be the true one and gives an estimate of the accuracy in terms of a standard deviation. The new prediction included in the posterior is then a compromise of all these sources. Fig. 2.2 gives an impression of how the Bayesian update procedure changes the probability density function.

The last example has already shown the essential steps in the Full Bayesian Update Technique. The main difference is that in the latter case the prior is not only dependent on one cross section x but on several types of them. This means that the one-dimensional Gauss distribution is extended to a multi-dimensional one. A physical model places constraints on variations of these cross sections – if one cross section takes a certain value, the possible range of values of other cross section types is reduced. This coupling between different cross section types is achieved by the specification of a covariance matrix which is a parameter (or rather: a set of parameters) in the multi-dimensional Gauss distribution. A covariance matrix is the generalization of the variance for more than one dimension.

2.3 Principle of maximum entropy

Since the method of this work implements Bayesian inference, it relies on the specification of a prior pdf. In order to unambiguously specify it, reasonable principles are necessary.

One of these principles introduced by E.T. Jaynes is the principle of maximum entropy [19]. It states that one should choose prior probabilities in such a way that the information entropy is maximised under consideration of eventually imposed constraints, e.g. the mean being a known quantity. The information entropy is defined as [34]

$$H = -\sum_{i=1}^{n} p_i \log(p_i), \tag{2.26}$$

with n being the number of propositions and p_i the probabilities assigned to them. In the following, it is shown why pdfs obtained by maximising the information entropy can be regarded as plausible choices.

If no constraints are imposed on the shape of a pdf, equal probabilities should be assigned to the propositions since there is no reason to favour any proposition over any other. However, constraints which are imposed on the shape of a pdf very often forbid to assign equal probabilities. In this case, a pdf should be chosen which follows the uniform distribution as close as possible. In order to quantify the 'closeness' or distance to the uniform distribution a measure is required. In the following description of a betting game – taken from Jaynes [20, p. 351-355] and adapted – the information entropy will naturally appear as such and its maximisation is associated with the most promising bet.

Assume that there are m propositions to which probabilities have to be assigned. Imagine two persons Mr. A and Mrs. B playing the following game. Mr. A has an amount of N probability quanta available to distribute them among the propositions. Because

he has no idea of how to do that, he takes one probability quantum at a time and assigns it randomly, according to a uniform distribution, to some proposition i. Maybe he tosses a symmetrical and unbiased dice with m sides for that purpose. He does so for every probability quantum at his avail. When he has worked off all of them, each quantum is associated with a proposition. Thereafter, he calculates the corresponding probability for each proposition i by

$$p_i = \frac{n_i}{N},\tag{2.27}$$

where n_i is the amount of quanta assigned to proposition i. Then, he shows his result to Mrs. B. Mrs. B has more information than him. She knows about some constraints that should be regarded by the probability distribution. If his proposition is consistent with the constraints, she accepts it, otherwise rejects it. In this manner Mr. A and Mrs. B continue for several games.

After a while they stop playing and analyse the outcomes of the games where the probability distribution was accepted. Assuming they played sufficiently many games, what relative frequency for the realization of a probability p_i for a proposition i would they find?

To resolve this question, one can assume for a moment that there have not been any constraints and thus Mrs. B would have accepted every outcome as valid. In this case, the number of possibilities to arrive at the same assignment of probability quanta is given by the multinomial coefficient

$$W = \frac{N!}{n_1!\, n_2!\, n_3! \cdots n_m!}.\tag{2.28}$$

For instance, if there are $m = 2$ propositions and $N = 2$ probability quanta for distribution, obtaining the outcome $n_1 = 1$ and $n_2 = 1$ can be realized by Mr. A first assigning a quantum to proposition 1 and then the other quantum to proposition 2 or first assigning a quantum to proposition 2 and then the other to proposition 1 – the same outcome reached in two different ways, thus $W = 2$. Since every way is equally likely, the probability for a certain outcome is proportional to the number of ways it can be realized – hence proportional to W. To scale the multiplicity W to represent the probability in the unconstrained case, one has to multiply it by m^{-N}.

The important observation is that imposing constraints on the probability distribution does not change the number of ways allowed outcomes can be realized, their multiplicity remains the same. Only the scaling factor to transform multiplicities in probabilities takes a different value.

Assume that another person Mr. C should bet on the outcome of the next game which is going to be accepted by Mrs. B. As Mr. C is also aware of the constraints, he reasonably bets on the outcome which is associated with the highest multiplicity W of all accepted outcomes. Consequently, he takes values for n_i allowed by the constraints which maximize (2.28). Because maximization of every monotonically increasing function of W

in place of directly maximising W leads to the same result, one could use the function

$$
\begin{aligned}
\frac{1}{N} \log(W) &= \frac{1}{N} \log \left(\frac{N!}{n_1! \, n_2! \, n_3! \cdots n_m!} \right) \\
&= \frac{1}{N} \log \left(\frac{N!}{(Np_1)! \, (Np_2)! \, (Np_3)! \cdots (Np_m)!} \right) \\
&= \frac{1}{N} \left(\log(N!) - \sum_{i=1}^{m} \log((Np_i)!) \right).
\end{aligned}
\tag{2.29}
$$

Because probabilities are 'smooth' quantities in the interval $[0,1]$ and (2.27) produces rather rough jumps for small N, one could look at the limiting case $N \to \infty$ In this limiting case, Stirling's approximation formula

$$
\log(N!) = N \log(N) - N + \sqrt{2\pi N} + \frac{1}{12N} + O\left(\frac{1}{N^2} \right).
\tag{2.30}
$$

reduced to the leading term can be applied to Eq. (2.29):

$$
\begin{aligned}
\lim_{N \to \infty} \log(W) &= \frac{1}{N} \left(N \log(N) - \sum_{i=1}^{m} Np_i \log((Np_i)) \right) \\
&= \log(N) - \sum_{i=1}^{m} p_i \log((Np_i)) \\
&= \log(N) - \log(N) \sum_{i=1}^{m} p_i - \sum_{i=1}^{m} p_i \log(p_i) \\
&= - \sum_{i=1}^{m} p_i \log(p_i).
\end{aligned}
\tag{2.31}
$$

This is exactly the information entropy (2.26) and the probabilities which maximise it are associated with the most probable outcome of the described betting game.

The whole derivation is a very convincing argument for using the principle of maximum entropy to select prior probabilities, because a probability assignment is regarded as the result of a random process where no proposition is favoured over any other. If no constraints are imposed on the pdf, a uniform distribution is obtained, in accordance with the principle of indifference.

The entropy formula (2.26) only works out for a finite set of propositions. The 'continuous quantity' version of it is given by [20, p. 375]

$$
H^c = - \int \mathrm{d}x \, \rho(x) \log \left[\frac{\rho(x)}{m(x)} \right],
\tag{2.32}
$$

with $\rho(x)$ being a probability density function and $m(x)$ an invariant measure. The introduction of $m(x)$ is necessary as a variable transformation might not change anything

about the problem, but the principle of maximum entropy would lead to another pdf. For instance, if there is an experiment in which some length is measured and one has to set up a prior expressing complete ignorance about the outcome, the prior pdf should be the same on every length scale. The invariant measure $m(x)$ ascertains the invariance of the prior pdf under variable transformations. However, if the scale is known, e.g. the standard deviation takes a certain value at a definite scale and only the mean is unknown, the invariant measure would only ascertain invariance in respect to translation. Transformation to another length scale would then indeed change the functional form of the prior pdf.

Very often one has testable information about a pdf such as knowledge of the arithmetic mean or variance to which the pdf should be constrained. If there are m different constraints of the form

$$\int \mathrm{d}x\, \rho(x) f_k(x) = F_k \tag{2.33}$$

and, additionally, a constraint due to normalization

$$\int \mathrm{d}x\, \rho(x) = 1, \tag{2.34}$$

then maximisation of H^c (2.32) results in the pdf

$$\rho(x) = \frac{1}{Z(\lambda_1, \lambda_2, \ldots, \lambda_m)} m(x) \exp\left(\lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_m f_m(x)\right). \tag{2.35}$$

Z is called partition function because it has an analogous meaning as the partition function used in statistical mechanics. It assures the normalization of the pdf and therefore is given by

$$Z(\lambda_1, \lambda_2, \ldots, \lambda_m) = \int \mathrm{d}x\, m(x) \exp\left(\lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_m f_m(x)\right). \tag{2.36}$$

The Lagrange multipliers can be determined by solving the transcendental set of equations:

$$F_k = -\frac{\partial}{\partial \lambda_k} \log Z(\lambda_1, \lambda_2, \ldots, \lambda_m). \tag{2.37}$$

If there are no constraints of the form (2.33), the maximum entropy distribution is given by

$$\rho(x) = \left(\int m(x)\,\mathrm{d}x\right)^{-1} m(x). \tag{2.38}$$

Hence, it is only determined by the invariant measure. If the quantity x is known to be restricted to the interval $[a, b]$ on a definite scale, the maximum entropy pdf reads

$$\rho(x) = \frac{1}{b-a}. \tag{2.39}$$

As closing example, assume that the mean μ and the variance σ^2 of a distribution are known. These quantities are defined by

$$\mu = \int x\rho(x)\mathrm{d}x \qquad and \qquad \sigma^2 = \int (x-\mu)^2\rho(x)\mathrm{d}x. \qquad (2.40)$$

These two formulas can be combined to obtain a single constraint for the pdf:

$$\int \mathrm{d}x\rho(x)\,(x-\mu)^2 = \sigma^2. \qquad (2.41)$$

If it is known to which scale these distribution parameters refer, the invariant measure is given by $m(x) = 1$. Inserting $f_1 = (x-\mu)^2$ in Eq. (2.35) leads to a pdf of Gaussian shape:

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{1}{2\sigma^2}(x-\mu)^2\right). \qquad (2.42)$$

The multidimensional normal distribution is the pdf chosen in this work. The reason for the restriction to the leading two moments is given in the next chapter. The next chapter recapitulates the (detailed) implementation of Bayesian inference in the Full Bayesian Evaluation Technique.

3 The Full Bayesian Evaluation Technique recapitulated

The Full Bayesian Evaluation Technique [26] implements Bayesian inference as discussed in the previous chapter to combine experimental data with model data to generate improved estimates of integral cross sections. Thus, the specification of a prior pdf is necessary, which is done with the aid of the nuclear reaction code TALYS [23], consideration of physical boundary conditions, the principle of maximum entropy [19] and some heuristics. These topics are discussed in more detail in the following section 'Prior Knowledge'. Since for both the prior pdf – relying on model data – and the likelihood – relying on experimental data – a Gauss distribution is assumed and the model data is mapped to the experimental grid by a linear transformation, the posterior pdf is also of Gaussian shape. This allows to calculate the two characteristic quantities, a vector of mean values and a covariance matrix (the analogon to the variance in the one dimensional case) to uniquely specify the posterior pdf. The derivation of these formulas is the topic of the section 'Linearized Bayesian Update Procedure'.

Up to now FBET is only capable of dealing with integral cross section data. Details about the extension to differential cross sections, the subject of this work, are treated in chapter 4.

Before elaborating on technical details of FBET let introduce the notion of an integral cross section and make plausible why accurate knowledge about this quantity is helpful. Integral cross sections are defined by

$$\sigma = \frac{I_S}{I_I}, \tag{3.1}$$

where I_I is the current density of incident particles and I_S is the current density of scattered particles. If a multitude of target nuclei is within the scope of the incident particle stream, the factor $1/N$ has to be applied on the right hand side of Eq. (3.1).

The range of interaction depends on the projectile as well as the structure of the target nucleus. For instance, only for charged projectiles the scattering process is influenced by the Coloumb interaction. Therefore, cross sections are frequently categorized according to the type of projectile. Furthermore, an interaction may occur in several ways. An incident particle might be absorbed by the target nucleus and integrated into its structure; or the structure of the target nucleus is preserved and the incident particle only experiences a change in its moving direction. Hence, cross sections are further categorized as shape-elastic and reaction cross section. In case of absorption, different processes can take place. At the end of such a process, the nucleus may re-emit another particle, e.g. an α-particle and thereby turn to a more stable structure. Consequently, reaction cross

sections are again further divided by the type of (if there are any) outgoing particle(s). Commonly, all these categorizations are abbreviated in the following formal notation

$$\text{target nucleus (projectile, ejectile) final nucleus} \quad \text{e.g.} \quad {}^{181}_{73}\text{Ta}\,(n,\alpha)\,{}^{178}_{71}\text{Lu} \qquad (3.2)$$

If the specific target nucleus is clear from context, its specification is dropped. Additionally, if the final nucleus is already determined by the specification of target nucleus, projectile and ejectile it may also be dropped. For some cross section types various final nuclei occur and they are not measured because of technical limitations. This is the case for the total cross section (n,tot) where every interaction is taken into account.

All these different kinds of cross sections reveal particular aspects of interactions in play. As technological progress continues more and better experiments yield more and more accurate data. This knowledge, in turn, is used to design physical models for predicting outcomes of scattering processes. At the moment, there are several models based either on microscopic descriptions of the nucleus or phenomenological considerations. None of them is able to describe all scattering processes for all nuclei at all incident energies correctly – each model has its domain of validity. However, in the unresolved energy range reaching from about some MeV to around $200\,\text{MeV}$, the optical potential approach [17], for instance, describes the total cross section (n,tot) rather well.

Thus, confidence in a certain model may rise to a degree so that one starts using it for predicting cross sections in energy ranges without experimental data. Furthermore, models are used in energy ranges where data is available to correct estimates stemming from experiments to better fulfill the systematics of the model. Incorporating physical models enjoying some trust helps to narrow down uncertainties in areas with and even without experimental data. Having the predictions of the model with associated uncertainty information for energy ranges without experimental data could render some experimental setups expendable. Thus, costs are saved and experiments are only performed when they are absolutely necessary.

Basically, FBET is not a nuclear model but a statistical meta-model which calculates cross section estimates by relying on model data and experimental data. In the procedure described here, the nuclear code TALYS [23], which gives reasonable predictions in the unresolved energy range, is utilized to produce the model data. However, this is not a necessity. Data stemming from every other model could be used equally well as long as some criteria are fulfilled. These will become evident during the course of introducing the technical details of FBET.

3.1 Prior knowledge

The prior probability density function (prior pdf) in the Full Bayesian Evaluation Technique is specified with the aid of a physical model.

Such a model is based on concepts of the nucleus structure and relevant interactions between nucleons. As soon as values are assigned to the parameters required by the model, the latter can be used to predict outcomes of scattering experiments. Therefore,

one can regard a model as a function

$$\sigma_c(E) \equiv \sigma_c(E\,; r_1, r_2, \cdots), \tag{3.3}$$

where one passes the incident energy E as argument and gets the value of a certain cross section associated to a channel c, e.g. (n, tot), as result. The model parameters r_i characterize the structural properties of the nucleus and are thus – ideally – fixed quantities. Effectively, these parameters are obtained by fitting the model to experimental data, so experimental uncertainties are transferred to uncertainties in the model parameters. The following section describes how exactly this type of uncertainty enters the prior.

Cross section uncertainties due to parameter uncertainties

TALYS [23] is a phenomenological model that implements the optical potential approach [17]. It is able to describe cross section from above the resonance range to about 200 MeV reasonably well. Above that range channels associated with π-particle production open up which cannot be described properly by a nucleon-nucleon potential. Therefore, the evaluation is limited to the energy range below 200 MeV.

Together with the also incorporated global parameterization by Koning and Delaroche [22], TALYS steadily computes cross sections between Mg and Bi by relying only on the specification of a few parameters. These parameters are determined by fitting the global optical potential to total CS data obtained for various nuclei. Physical considerations allow to state intervals for them in which the true values have to lie with high probability. Their determination is explained in detail in the Ph.D. thesis of Neudecker [26].

Once parameter boundaries are established for the model parameters, a probability density function for allowed model parameters has to be specified. Minimizing the information entropy (2.32) and applying the principle of indifference yields an uniform distribution.

The prior in FBET is formulated in terms of cross sections and not parameter values, so the uniform pdf $\rho(r_1, \ldots, r_n | E, c)$ in parameter space has to be mapped to the associated pdf in cross section space $\tilde{\rho}(x|E, c)$. The incident energy E and channel c are regarded as fixed quantities which specify the scattering process in view. Once they are set up, the probability for obtaining a certain value x for the selected cross section is completely determined by the uncertainty of model parameters r_i. However, instead of only regarding a single cross section, the joint pdf for several cross sections at several energies is considered

$$\tilde{\rho}(\sigma_{c_1}(E_1), \, \ldots \,, \sigma_{c_1}(E_N), \, \ldots \,, \sigma_{c_M}(E_1), \, \ldots \,, \sigma_{c_M}(E_N)), \tag{3.4}$$

where $\sigma_{c_i}(E_j)$ is the value for the cross section associated to channel c_i at incident energy E_j. If each of the M cross sections in the joint pdf is stated for the same N energies, then all these values can be thought of as constituting a vector containing $Z = M \times N$ elements. In the following let σ_i be the i-th entry of this vector. Assuming that the number of cross section values Z is equal to the number of model parameters L and

that distinctive realizations of model parameters $\vec{P}_i$ are mapped to distinctive vectors of cross section values $\vec{\sigma}_i$, the mapping from parameter space to cross section space is a bijective function $\vec{\sigma}(\vec{P})$. In this case, the pdf for the cross section space is given by

$$\tilde{\rho}(\vec{\sigma}) = \rho(\vec{P}) \det \left(\frac{\partial \vec{\sigma}}{\partial \vec{P}} \right)^{-1}, \tag{3.5}$$

where $\partial \vec{\sigma}/\partial \vec{P}$ is the Jacobian matrix for the mapping $\vec{\sigma}(\vec{P})$. If $\vec{\sigma}$ is of greater dimension than $\vec{P}$, the mapping $\vec{P} \to \vec{\sigma}$ may still be assumed to be injective, but the determinant is not applicable anymore because the Jacobian matrix is rectangular. Thus, formula (3.5) has to be adapted. To this end, one has to realize that the above determinant is a scaling factor stating how many times bigger the infinitesimal volume $d\tilde{V}$ in cross section space is compared to the corresponding volume dV in parameter space. Because corresponding volumes in both spaces have to be equally probable, a blow up of $d\tilde{V}$ compared to dV has to be compensated by a reduction of the pdf at that point,

$$\tilde{\rho}(\vec{\sigma}) = \rho(\vec{P}) \frac{dV}{d\tilde{V}}. \tag{3.6}$$

Since the dimension of $\vec{P}$ is lower than of $\vec{\sigma}$, the pdf for the cross section space is only defined for points reachable by the mapping $\vec{\sigma}(\vec{P})$. In a sufficiently small environment around a point $\vec{\sigma}_0$ reachable by some $\vec{P}_0$, the Jacobian matrix can be used to construct a local approximation to the function $\vec{\sigma}(\vec{P})$:

$$\vec{\sigma}(\vec{P}) = \vec{\sigma}(\vec{P}_0) + \frac{\partial \vec{\sigma}}{\partial \vec{P}}(\vec{P} - \vec{P}_0). \tag{3.7}$$

Thus, the physical model can locally be regarded as linear mapping from $\mathbb{R}^L$ to $\mathbb{R}^Z$ with the image restricted to an L-dimensional subspace. The columns of the Jacobian matrix define a basis for this subspace. Infinitesimal volume elements $d\tilde{V}_r$ in the cross section subspace and dV in the parameter space are related through the square root of the Gram determinant by

$$d\tilde{V}_r = dV \sqrt{\det \left(\left(\frac{\partial \vec{\sigma}}{\partial \vec{P}} \right)^T \frac{\partial \vec{\sigma}}{\partial \vec{P}} \right)}. \tag{3.8}$$

Using equations (3.6) and (3.8) yields the following expression for the pdf in cross section space:

$$\tilde{\rho}(\vec{\sigma}_c) = \int_{V(P)} \rho(\vec{P}) \det \left((\partial \vec{\sigma}/\partial \vec{P})^T (\partial \vec{\sigma}/\partial \vec{P}) \right)^{-1/2} \delta \left(\vec{\sigma}_c - \vec{\sigma}(\vec{P}) \right) \mathrm{d}^Z P. \tag{3.9}$$

Integration is thereby performed over the whole range of allowed parameter vectors. The argument was renamed to $\vec{\sigma}_c$ to distinguish it from the function symbol $\vec{\sigma}$ establishing the mapping.

If this exact expression would be chosen for the prior pdf, then – independent of what experimental data constitutes the likelihood – the posterior pdf would always be

restricted to cross section vectors reachable by the model. In the case of an inadequate model, no amount of experimental data could compensate for that.

In practice, it is not possible to use the exact expression, as one TALYS model calculation takes, depending on chosen incident energies, from several minutes to over an hour. Even then, only the mapping of one parameter vector to the corresponding cross section vector is determined. Therefore, the construction of the prior pdf for cross section space cannot be done exactly and needs to take some assumptions into account. The procedure chosen in FBET is to sample parameter vectors according to a uniform distribution and calculate the corresponding cross section vectors. Because the resulting point density in cross section space is far too low to estimate the real pdf non-parameterically, it is assumed to be of Gaussian shape. The specification of a multidimensional Gauss distribution requires the calculation of the arithmetic mean

$$\overline{\sigma} = \frac{1}{l_{\max}} \sum_{l=1}^{l_{\max}} \vec{\sigma}(\vec{P}_l), \tag{3.10}$$

with $l_{\max}$ being the number of sample points in parameter space and $\vec{P}_l$ a certain parameter vector obtained during sampling. Furthermore, the covariance matrix A^{PU} is needed:

$$A^{\mathrm{PU}} = \frac{1}{l_{\max}} \sum_{l=1}^{l_{\max}} \left(\vec{\sigma}(\vec{P}_l) - \overline{\sigma} \right) \left(\vec{\sigma}(\vec{P}_l) - \overline{\sigma} \right)^T. \tag{3.11}$$

One element of the covariance matrix is given by

$$A_{ij}^{\mathrm{PU}} = \mathrm{cov}(\sigma_i, \sigma_j) = \frac{1}{l_{\max}} \sum_{l=1}^{l_{\max}} (\sigma_i(\vec{P}_l) - \overline{\sigma}_i)(\sigma_j(\vec{P}_l) - \overline{\sigma}_j). \tag{3.12}$$

which is a scale dependent measure for the linear dependence between $\vec{\sigma}_i$ and $\vec{\sigma}_j$. A measure which is invariant under linear transformations of the respective quantities is the correlation coefficient

$$\mathrm{corr}(\sigma_i, \sigma_j) = \frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}}. \tag{3.13}$$

It takes a value in the interval $(-1, 1)$: $\mathrm{corr}(\sigma_i, \sigma_j) = 1$ for a perfect positive linear dependence ($\sigma_i = \alpha \, \sigma_j, \alpha > 0$) and $\mathrm{corr}(\sigma_i, \sigma_j) = -1$ for a perfect negative one ($\alpha < 0$).

The Gaussian associated to that mean vector and covariance matrix is

$$\tilde{\rho}(\vec{\sigma}) = \frac{1}{(2\pi)^{Z/2}\sqrt{\det(A^{\mathrm{PU}})}} \exp\left(\frac{1}{2}(\vec{\sigma} - \overline{\sigma})^T (A^{\mathrm{PU}})^{-1} (\vec{\sigma} - \overline{\sigma}) \right). \tag{3.14}$$

This is also the distribution which maximizes the information entropy if only arithmetic means $\overline{\sigma}_i$ and covariances $\mathrm{cov}(\sigma_i, \sigma_j)$ are given. The consideration of higher moments would alter the distribution of maximal entropy which might approximate the real distribution more accurately. However, the enormously increased computational costs do not seem to justify this step.
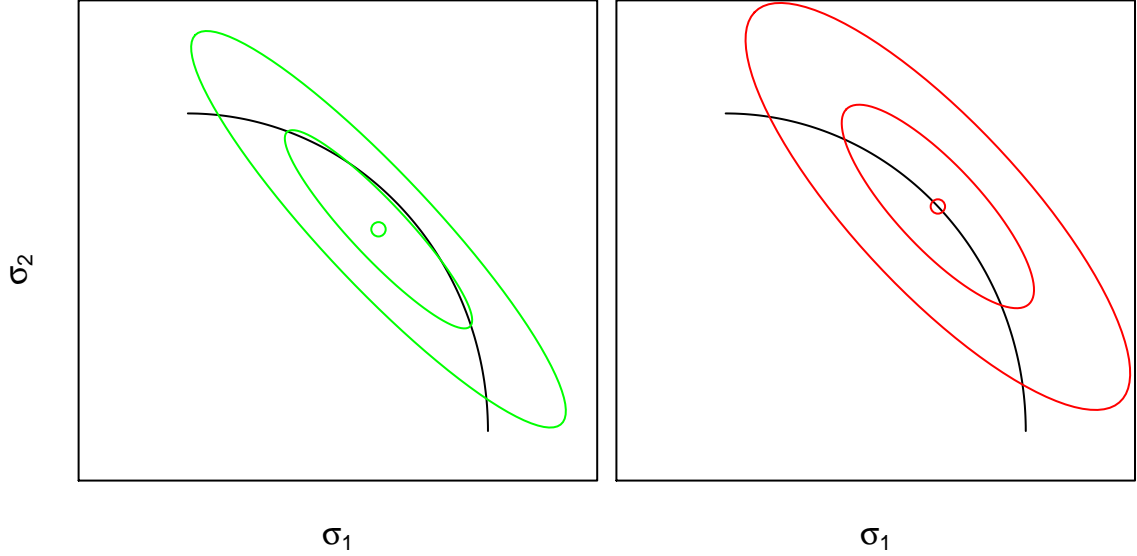
21

Figure 3.1: Influence of the choice of the center point on the shape of the Gauss distribution. Green: centered at the mean of sampled cross section vectors $\overline{\sigma}$; red: centered at mapped mean parameter vector $\vec{\sigma}\left(\overline{P}\right)$

Due to the non-linear nature of the mapping from the parameter space to the cross section space, the mean vector in the cross section space differs from the mapping of the mean parameter vector, so $\vec{\sigma}\left(\overline{P}\right) \neq \overline{\sigma}$. The situation is visualized in Fig. 3.1. Assume that the model relies only on one parameter s and maps to two cross sections σ_1, σ_2. The reachable points in the cross section space are represented by the black line, where each point on this line might be considered equally probable. Taking the arithmetic mean $\overline{\sigma}$ as center point leads to the green Gauss distribution; taking the mapped mean parameter vector $\vec{\sigma}\left(\overline{P}\right)$ leads to the red one. As can be seen, the red distribution better approximates the slope around the point associated with the mean parameter vector, but at the cost of a higher uncertainty. In the original work of Neudecker [26] $\vec{\sigma}(\overline{P})$ was chosen as center point. This work will, instead, use $\overline{\sigma}$ which can be justified by several reasons. Firstly, the cross section vectors were obtained by relying on a uniform distribution in parameter space. Because every parameter set is considered equally likely, the parameter vector $\overline{P}$ is not in any way special. The Gauss distribution centered at the arithmetic mean is – according to the principle of maximum entropy – the more honest description, in better accordance with the uniform distribution in parameter space. Secondly, the increased uncertainty when using $\vec{\sigma}\left(\overline{P}\right)$ increases the probability for points not predicted by the model. Thirdly, the trend of the green distribution follows the model more rigorously regarding all points whereas the red one is only valid for points around $\vec{\sigma}\left(\overline{P}\right)$.

Even if the most probable cross section point is then not the best guess for the true point, the true point is still considered very likely. In the above figure, which clearly shows a highly non-linear model, the true point is within the 70% confidence interval.

Furthermore, the purpose of the model is not to be accurate from the beginning, but to adapt experimental data to better fulfill the systematics of the model.

Another important aspect is the singularity of A^{PU}. If the parameter space is less dimensional than the cross section space, model points are restricted to a subspace. In Fig 3.1, the model is restricted to a line. Non-linearity of the model together with approximating the true pdf with a Gauss distribution results in the assignment of non-vanishing probabilities to points not reachable by the model. Nonetheless, sometimes A^{PU} does not have full rank and thus the required inversion in formula (3.14) is not possible. A covariance matrix is singular if there are directions along which the variance vanishes. Therefore, one can perform an eigenvalue decomposition to detect these directions. This yields the covariance matrix relative to its eigenbasis

$$A_{\mathrm{eigen}}^{\mathrm{PU}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \qquad \lambda_i \geq 0, \tag{3.15}$$

where λ_i are the eigenvalues. They state the variance along the principal axes. In order to obtain a matrix of full rank, one has to remove columns and rows of this matrix containing a zero eigenvalue. The resulting matrix $\tilde{A}_{\mathrm{eigen}}^{\mathrm{PU}}$ is associated with an eigenbasis which is only a basis for a linear subspace in the cross section space. Thus, vectors that had to be multiplied with the original covariance matrix have to be projected into this subspace before multiplication. Let $\vec{e}_i$ be the normalized eigenvector associated to an eigenvalue λ_i, then e.g.

$$\vec{x}^T \left(A^{\mathrm{PU}} \right)^{-1} \vec{x} \to \vec{x}^T \left[\vec{e}_1, \cdots, \vec{e}_i, \cdots, \vec{e}_n \right] \left(\tilde{A}_{\mathrm{eigen}}^{\mathrm{PU}} \right)^{-1} \left[\vec{e}_1, \cdots, \vec{e}_i, \cdots, \vec{e}_n \right]^T \vec{x}. \tag{3.16}$$

Only those $\vec{e}_i$ are included which correspond to non-zero eigenvalues λ_i. Through this procedure no information is lost, because the model would not allow any deviation in directions associated with zero eigenvalues. The last entity relies on the fact that a matrix A owns the same eigenbasis as A^{-1}.

Cross section uncertainties due to model deficiencies

Although the approximation of the pdf for the cross section space with a Gauss distribution leads to non-vanishing probability densities for points not lying on the model manifold, thereby allowing experimental data to draw the predicted value away from it, a model inadequate under some circumstance may still constrain the space of possible cross section vectors too much. Therefore, the systematic error of the model was taken into account in the FBET expressed in terms of a covariance matrix A^{MD}. Its determination for a certain isotope relied on other isotopes of the same element The complete prior covariance matrix is given by $A_0 = A^{\mathrm{PU}} + A^{\mathrm{MD}}$. The procedure to obtain model

deficiency matrices A^{MD} is formulated for integral cross sections. The extension of this procedure to differential cross sections is not within the scope of this work. Nevertheless, as will be seen in the next chapter, it is possible to incorporate the uncertainty information for integral cross sections included in A^{MD} consistently into the extended FBET. Details of the 'scaling procedure' to calculate A^{MD} are given in the work of D. Neudecker [26].

3.2 A linearized Bayesian Update Procedure

In section 3.1, the sampling procedure was described, which yielded a set of cross section vectors. Each component σ_i of these vectors represented a cross section value for a certain channel c_i at an incident energy E_i. This sample of cross section vectors was used to estimate a mean vector $\overline{\sigma}_{\mathrm{mod}}$ and a covariance matrix A^{PU} to build up the Gaussian prior

$$\rho(\vec{\sigma}|\overline{\sigma}_{\mathrm{mod}}, A^{\mathrm{PU}}) = -\frac{1}{(2\pi)^{Z/2}\sqrt{\det(A^{\mathrm{PU}})}} \exp\left(-\frac{1}{2}(\vec{\sigma} - \overline{\sigma}_{\mathrm{mod}})^T (A^{\mathrm{PU}})^{-1}(\vec{\sigma} - \overline{\sigma}_{\mathrm{mod}})\right),$$
(3.17)

which states the probability density that $\vec{\sigma}$ represents the real cross section vector under the assumed hyperparameters $\overline{\sigma}_{\mathrm{mod}}$ and A^{PU} derived from model data. Z is given by the dimension (number of rows or columns) of A^{PU}.

The likelihood is also given by a Gauss distribution specified by the estimate $\overline{\sigma}_{\mathrm{exp}}$ for the true cross section vector and covariance matrix B. In general, the dimensions of $\vec{\sigma}_{\mathrm{exp}}$ and $\vec{\sigma}_{\mathrm{mod}}$ differ. One may only use experimental data giving estimates for channels contained in the prior to compose the likelihood. Still, it is possible that experimental data contains estimates for energies not stated in the prior, e.g. the model vector $\vec{\sigma}_{\mathrm{mod}}$ contains cross section values for channel c at energies E_1, E_2, E_3 whereas the experiment states cross section values for the same channel at energies E'_1, E'_2 lying between the model energies. In this case, the application of an interpolation scheme is required. For this purpose, FBET uses spline interpolation, which can be regarded as linear mapping from the energy grid of the model to the energies of the experiment

$$\begin{pmatrix} \sigma_{c;E'_1} \\ \sigma_{c;E'_2} \\ \vdots \\ \sigma_{c;E'_m} \end{pmatrix} = S \begin{pmatrix} \sigma_{c;E_1} \\ \sigma_{c;E_2} \\ \vdots \\ \sigma_{c;E_n} \end{pmatrix},$$
(3.18)

with S being an $m \times n$ matrix. Basically, every function of the form

$$f(E) = \sum_i \alpha_i \, g_i(E),$$
(3.19)

where the $g_i(E)$ are arbitrary functions defined at all E values of interest, establishes a linear mapping, as long as all α_i are linear functions of the σ_{c,E_i}. Thus, instead of splines, it would be possible to use e.g. Fourier series or Legendre series.

Assume some cross section is given for the energies E_i and it has to be mapped to energies E_i'. The mapping matrix S can be determined by setting $\sigma_{c,E_i} = 0, \forall i \neq j$ and $\sigma_{c,E_j} = 1$. For this choice, one can determine the coefficients α_i in f which leads to the function $f_j(E)$. Now f_j is calculated for all m energies E_i'. This procedure is performed for all $j \in \{1, \cdots, n\}$. The resulting mapping matrix is given by

$$
S = \begin{pmatrix}
f_1(E_1') & f_2(E_1') & \cdots & f_n(E_1') \\
f_1(E_2') & f_2(E_2') & \cdots & f_n(E_2') \\
\vdots & \vdots & \ddots & \vdots \\
f_1(E_m') & f_2(E_m') & \cdots & f_n(E_m')
\end{pmatrix}.
$$ (3.20)

Let S be the matrix which maps a model cross section vector to the associated vector in the experiment space, then the likelihood is given by

$$
\rho(\overline{\sigma}_{\mathrm{exp}}|\vec{\sigma}) = \frac{1}{(2\pi)^{Z/2}\sqrt{\det(B)}} \exp\left(-\frac{1}{2}(S\vec{\sigma} - \overline{\sigma}_{\mathrm{exp}})^T B^{-1}(S\vec{\sigma} - \overline{\sigma}_{\mathrm{exp}})\right).
$$ (3.21)

It states the probability density that an experimental estimate $\overline{\sigma}_{\mathrm{exp}}$ is obtained assuming $\vec{\sigma}$ to be the true cross section vector. The true cross section is given with respect to the model grid. Therefore, the mapping matrix S is applied to transfer it to the experiment space.

The product of prior and likelihood yields – up to a normalization factor N – the posterior pdf

$$
\rho(\vec{\sigma}|\overline{\sigma}_{\mathrm{exp}}, B, \overline{\sigma}_{\mathrm{mod}}, A) =
$$
$$
N \exp\left(-\frac{1}{2}(S\vec{\sigma} - \overline{\sigma}_{\mathrm{exp}})^T B^{-1}(S\vec{\sigma} - \overline{\sigma}_{\mathrm{exp}}) - \frac{1}{2}(\vec{\sigma} - \overline{\sigma}_{\mathrm{mod}})^T A^{-1}(\vec{\sigma} - \overline{\sigma}_{\mathrm{mod}})\right).
$$ (3.22)

The product of two Gauss distributions is itself a Gauss distribution. Assuming this to be true, it should be possible to restate the exponent of Eq. (3.22) in the following form:

$$
-\frac{1}{2}(\vec{\sigma} - \tilde{\sigma})^T \tilde{A}^{-1}(\vec{\sigma} - \tilde{\sigma}) = -\frac{1}{2}\left\{\vec{\sigma}^T \tilde{A}^{-1}\vec{\sigma} - \vec{\sigma}^T \tilde{A}^{-1}\tilde{\sigma} - \tilde{\sigma}^T \tilde{A}^{-1}\vec{\sigma} + \tilde{\sigma}^T \tilde{A}^{-1}\tilde{\sigma}\right\},
$$ (3.23)

with $\tilde{\sigma}$ the new mean and $\tilde{A}$ the new covariance matrix of the Gauss distribution. Rearranging the exponent of Eq. (3.22), one obtains

$$
-\frac{1}{2}\Big\{\vec{\sigma}^T \left(A^{-1} + S^T B^{-1} S\right) \vec{\sigma} - \vec{\sigma}^T \left(A^{-1}\overline{\sigma}_{\mathrm{mod}} + S^T B^{-1}\overline{\sigma}_{\mathrm{exp}}\right)
$$
$$
- \left(\overline{\sigma}_{\mathrm{mod}}^T A^{-1} + \overline{\sigma}_{\mathrm{exp}}^T B^{-1} S\right) \vec{\sigma}
$$
$$
+ \left(\overline{\sigma}_{\mathrm{mod}}^T A^{-1}\overline{\sigma}_{\mathrm{mod}} + \overline{\sigma}_{\mathrm{exp}}^T B^{-1}\overline{\sigma}_{\mathrm{exp}}\right) \Big\}.
$$ (3.24)

The rightmost summand in Eq. (3.23) and (3.24) is not important in the current consideration, because it only changes the normalization factor. If proper normalized pdfs

for likelihood and prior are used, the Bayesian inference formula guarantees the normalization of the posterior pdf. Comparing (3.23) with (3.24), one can immediately identify

$$\tilde{A}^{-1} = A^{-1} + S^T B^{-1} S \tag{3.25}$$

and

$$\tilde{A}^{-1}\tilde{\sigma} = \left(A^{-1}\overline{\sigma}_{\text{mod}} + S^T B^{-1}\overline{\sigma}_{\text{exp}}\right) \tag{3.26}$$

$$\tilde{\sigma} = \left(A^{-1} + S^T B^{-1} S\right)^{-1} \left(A^{-1}\overline{\sigma}_{\text{mod}} + S^T B^{-1}\overline{\sigma}_{\text{exp}}\right). \tag{3.27}$$

In order to obtain the expressions as given in the work of Neudecker [26], one has to utilize the Woodbury matrix identity [37]

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VAU)^{-1}VA^{-1}. \tag{3.28}$$

Applied to Eq. (3.25), one obtains

$$\tilde{A} = A - AS^T \left(SAS^T + B\right)^{-1} SA \tag{3.29}$$

for the covariance matrix. Inserting (3.29) into (3.27) yields

$$\tilde{\sigma} = \overline{\sigma}_{\text{mod}} - AS^T \left(SAS^T + B\right)^{-1} S\overline{\sigma}_{\text{mod}} + \\ AS^T B^{-1}\overline{\sigma}_{\text{exp}} - AS^T \left(SAS^T + B\right)^{-1} SAS^T B^{-1}\overline{\sigma}_{\text{exp}}. \tag{3.30}$$

The second line is equal to

$$AS^T \left(SAS^T + B\right)^{-1} \overline{\sigma}_{\text{exp}}, \tag{3.31}$$

which follows from

$$AS^T B^{-1}\overline{\sigma}_{\text{exp}} - AS^T \left(SAS^T + B\right)^{-1} SAS^T B^{-1}\overline{\sigma}_{\text{exp}} = AS^T \left(SAS^T + B\right)^{-1} \overline{\sigma}_{\text{exp}} \\ B^{-1} - \left(SAS^T + B\right)^{-1} SAS^T B^{-1} = \left(SAS^T + B\right)^{-1} \\ \mathbb{1} - \left(SAS^T + B\right)^{-1} SAS^T = \left(SAS^T + B\right)^{-1} B \\ \left(SAS^T + B\right) - SAS^T = B. \tag{3.32}$$

Hence, replacing the second line of (3.30) by (3.31) and rearranging yields the final expression for the mean value $\tilde{\sigma}$

$$\tilde{\sigma} = \overline{\sigma}_{\text{mod}} + AS^T(SAS^T + B)^{-1}(\overline{\sigma}_{\text{exp}} - S\overline{\sigma}_{\text{mod}}). \tag{3.33}$$

This derivation shows that, as long as a linear mapping is chosen for the transfer of the model cross section vectors to the experiment space, the posterior pdf (3.22) is again a Gauss distribution. In order to obtain the important parameters, i.e. the mean value $\tilde{\sigma}$ and the covariance matrix $\tilde{A}$ which completely specify its shape, it sufficed to identify which terms in the exponent of (3.22) constitute the mean vector and which the

covariance matrix. As Eq. (3.25) and (3.29) are equivalent, connected by the Woodbury matrix identity, it would also be possible to calculate the covariance matrix directly:

$$\tilde{A} = \left(A^{-1} + S^T B^{-1} S\right)^{-1}. \tag{3.34}$$

However, this is not done in practice, as it would require the inversion of the huge model covariance matrix A. Because A is known in the beginning, the Woodbury matrix identity allows to calculate the inverse of (3.34) without the need to compute A^{-1}. Therefore, only the inversion of the much smaller matrix $(SAS^T + B)$ is required.

Another issue to be resolved is whether the update formulas (3.29) and (3.33) are always valid. For instance, the direct formula (3.34) relies on the inversion of A which is not always possible, because A is only guaranteed to be positive semi-definite. The application of the Woodbury matrix identity leads to the allegedly equivalent formula (3.29) requiring the inverse of A^{-1}, which would be just A. As A^{-1} sometimes does not exist, it has to be clarified what is really calculated by Eq. (3.29) in such a case. On close inspection, it turns out that formula (3.29) invisibly implements the projection of A into a subspace as described around formula (3.16). To see this, let T be the transformation matrix for the mapping into this subspace. Its rows are given by the eigenvectors of A associated with non-zero eigenvalues. Furthermore, assume A_{red} to be the covariance matrix A in respect to this subspace, then

$$A = T^T A_{\mathrm{red}} T \qquad \text{and} \qquad A_{\mathrm{red}} = T A T^T. \tag{3.35}$$

Because A is symmetric, the eigenvectors are orthogonal and so is the transformation matrix they build up, which means $T^T = T^{-1}$. Every matrix X of same dimensions as A could be equally projected by calculating $X_{\mathrm{red}} = T X T^T$. Transforming all quantities into the subspace, A_{red}^{-1} is now well-defined, and thus formula (3.34) reads

$$\tilde{A}_{\mathrm{red}} = \left(A_{\mathrm{red}}^{-1} + T S^T B^{-1} S T^T\right)^{-1}. \tag{3.36}$$

Applying the Woodbury matrix identity (3.28) once again yields

$$\tilde{A}_{\mathrm{red}} = A_{\mathrm{red}} - A_{\mathrm{red}} T S^T (S T^T A_{\mathrm{red}} T S^T + B)^{-1} S T^T A_{\mathrm{red}}. \tag{3.37}$$

Up-projecting by multiplying both sides with T^T from the left and T from the right and substituting $T^T A_{\mathrm{red}} T \rightarrow A$, the original expression (3.29) is obtained. The last consideration also shows that experimental uncertainties associated with directions not lying in this subspace are simply trimmed out. This is reasonable, as the model does not allow any deviation in these directions; thus the uncertainty of the experiment along the same direction does not have any influence on the prediction.

Performing several updates with experimental data, the question arises whether the order matters in which data is included. To answer this, assume two experimental data sets $\overline{\sigma}_{\mathrm{exp},1}$, B_1 with the mapping matrix S_1 and $\overline{\sigma}_{\mathrm{exp},2}$, B_2 with S_2. Instead of considering the change of $\tilde{\sigma}$ and $\tilde{A}$ due to updates one can also look at $\tilde{A}^{-1}$ and $\tilde{A}^{-1}\tilde{\sigma}$. The first update with experiment 1 yields, according to equations (3.25) and (3.26),

$$\tilde{A}_1^{-1} = A^{-1} + S_1^T B_1^{-1} S_1 \tag{3.38}$$

$$\tilde{A}_1^{-1}\tilde{\sigma}_1 = A^{-1}\overline{\sigma}_{\mathrm{mod}} + S_1^T B_1^{-1}\overline{\sigma}_{\mathrm{exp},1}. \tag{3.39}$$

Thereafter, using the obtained posterior quantities as prior quantities and updating with experiment 2 results in

$$\tilde{A}_{12}^{-1} = \tilde{A}_1^{-1} + S_2^T B_2^{-1} S_2 \qquad = A^{-1} + S_1^T B_1^{-1} S_1 + S_2^T B_2^{-1} S_2 \qquad (3.40)$$

$$\tilde{A}_{12}^{-1} \tilde{\sigma}_{12} = \tilde{A}_1^{-1} \tilde{\sigma}_1 + S_2^T B_2^{-1} \overline{\sigma}_{\text{exp},2} = A^{-1} \overline{\sigma}_{\text{mod}} + S_1^T B_1^{-1} \overline{\sigma}_{\text{exp},1} + S_2^T B_2^{-1} \overline{\sigma}_{\text{exp},2}. \qquad (3.41)$$

Updating in the opposite order amounts to a change of indices $1 \rightarrow 2, 2 \rightarrow 1$. As both $\tilde{A}_{12}^{-1}$ and $\tilde{A}_{12}^{-1} \overline{\sigma}_{12}$ are symmetric in respect to index swapping, the result remains unchanged. Hence, as long as experiment 1 is independent of experiment 2, i.e. if their common covariance matrix is block diagonal, the order of updating is irrelevant. However, if there are correlations between the experiments, they must both be included during the same update step.

To conclude this section, it shall be explained why this update procedure is referred to as 'linearized'. The first reason is the choice of a Gauss distribution for the prior. This distribution only takes linear relationships between different cross sections in the model cross section vector $\vec{\sigma}$ into account. This is probably always sufficient to model dependencies between different energies, which are not far apart in the same channel. Non-linear features of the physical model may dominate when energy differences increase. If the real probability density function has several maxima, the Gauss distribution would also be inadequate. In extreme cases, the correlation coefficient could turn to zero, although one cross section value is able to exactly predict the other. Another domain where the Gauss approximation could fail is within the resonance energy range. When both the location of peaks as well as their width is sensitively dependent on the model parameters, the assumption of linear relationships between cross sections perhaps cannot explain their dependency.

The second reason is the restriction to linear mappings from the model space to the experiment space. Only in this case, the likelihood and, consequently, also the posterior are of Gaussian shape. As long as the model energy grid is dense compared to changes of the slope of $\sigma(E)$, there is no cogent reason to doubt about the applicability of spline interpolation. Reversely, if the grid were too rough, no interpolation scheme – lacking knowledge of the function trend between the grid points – would be any more plausible. For a non-linear mapping, one would have to minimize (3.22) numerically.

However, as this work relies on spline interpolation, the mapping from model space to experimental space is linear. Thus, the important update formulas for this work are:

$$\tilde{\sigma} = \overline{\sigma}_{\text{mod}} + A S^T (S A S^T + B)^{-1} (\overline{\sigma}_{\text{exp}} - S \overline{\sigma}_{\text{mod}}) \qquad (3.42)$$

$$\tilde{A} = A - A S^T \left(S A S^T + B \right)^{-1} S A. \qquad (3.43)$$

3.3 Comparison of the linearized Bayesian Update Procedure with MLE

Maximum likelihood estimation (MLE) is a common method in conventional statistics (Frequentism) to estimate the parameters of a model. In this section, the conceptual differences between the linearized Bayesian update procedure (LBUP) and MLE will be demonstrated.

MLE is based on the assumption that the model adequately describes reality but, due to measuring inaccuracies, the observed data cannot be fitted exactly by the model. Errors of measurement are assumed to be normally distributed. This can be expressed as

$$\vec{\sigma} = m(\vec{P}_0) + \varepsilon \quad \text{with} \quad \varepsilon \sim \mathcal{N}(\vec{0}, B), \tag{3.44}$$

where $m(\vec{P}_0)$ is a function which maps the correct parameter vector $\vec{P}_0$ to the true cross section vector $\vec{\sigma}_0$. However, due to the noise ε, the observed quantity is $\vec{\sigma}$. The covariance matrix B is estimated from experimental data. The reasoning for obtaining an estimate for the true parameter vector $\hat{P}_0$ is as follows: Suppose the true value $\vec{\sigma}_0$ and hence also $\vec{P}_0$ were known. Then, the pdf for obtaining $\vec{\sigma}$ in a measurement is given by

$$\rho(\vec{\sigma}) = N \exp \left(\left(\vec{\sigma} - m(\vec{P}_0) \right)^T B^{-1} \left(\vec{\sigma} - m(\vec{P}_0) \right) \right) \tag{3.45}$$

The maximum likelihood principle states that the best guess for $\vec{P}_0$ is the parameter vector $\hat{P}_0$ which maximizes the probability density function for obtaining the observed cross section vector $\vec{\sigma}$. Because only model parameters are adjusted, the best estimate $\hat{\sigma}_0$ is always reachable by the model, in other words, included in the image of $m(\vec{P})$.

If the Bayesian update procedure would use the distribution in the cross section space which is in exact correspondence to the uniform distribution in the parameter space, the Bayesian update procedure would also restrict evaluated cross section vectors to elements of the image of $m(\vec{P})$. In this case, the difference between MLE and the Bayesian procedure amounts to different a-priori probability assignments for the vectors in cross section space. The Bayesian procedure favours points where the mapping $\vec{P} \rightarrow \vec{\sigma}$ is contracted, i.e. where the ratio V_c/V_p of corresponding volumes in cross section space (V_c) and in parameter space (V_p) is small. Contrarily, MLE does not favour any cross section vectors a priori. Thus, MLE is equivalent to a Bayesian procedure in which a uniform prior in respect to the cross section space is chosen.

Both methods perfectly regard the dependencies between different cross sections imposed by the model. To which amount their predictions differ depends on the uncertainty of the experimental data compared to the prior uncertainty. If the prior is rather broad and the experiment makes a sharp prediction where the prior assumes a high probability, the location of the posterior is almost completely determined by the experiment. This is visualized in the left box of Fig. 3.2. In this case, the model only imposes its systematics on the posterior, i.e. the best guess is restricted to the model manifold and the Bayesian update yields similar results to MLE. If the prediction of the experiment is not regarded
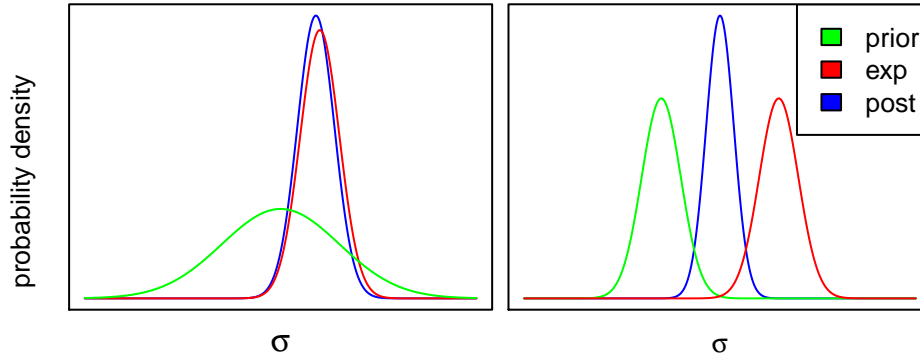
Figure 3.2: left: unspecific prior and sharp prediction by experiment lead to compareable results of MLE and the Bayesian update procedure; right: prior is at odds with experiment about the true value, MLE and Bayes procedure lead to substantially different results

to be very probable by the prior, MLE and the Bayesian update procedure substantially differ. The behaviour of the Bayes procedure in this scenery is qualitatively shown in the right image in Fig. 3.2.

Since the parameter boundaries set by D. Neudecker [26] are fairly cautious, the prior in cross section space turns out to be rather broad for integral cross sections. Hence, similar results as those of MLE can be expected. However, the previous discussion assumed the perfect Bayesian update procedure. In the real linearised Bayesian update procedure of this work, the pdf in cross section space is approximated by a Gauss distribution. By approximating the non-linear features of the model linearly, the perfect systematics of the model are loosened and the prediction could be drawn to points not within the model space. If the model is absolutely accurate, this is a disadvantage, but otherwise could be an advantage.

4 Extension to differential cross sections

The extension of the Full Bayesian Evaluation Technique to differential cross sections as performed in this work is only an extension of parameter space, but not in concept. All structural properties of the method described in the last chapter remain unchanged: the prior is still given by a Gauss distribution and the mapping from model cross section space to experiment space is unaltered conveyed by a linear transformation. In short, the linearised Bayesian Update Procedure (LBUP) as presented in the previous chapter is used.

Merely the prior hyperparameters for the Gauss distribution are adapted. Extra blocks are added in the covariance matrix to account for uncertainty information of the differential cross section data stemming from the physical model. Thus, the inclusion of experimental data measuring this channel becomes feasible in LBUP. The specific modifications in the prior are the topic of the next section.

Thereafter, the details of the mapping of the model data to the energies and angles of the experiment are described. At that point, all methodical and structural properties of the extended FBET are completely defined.

In the remaining two sections, the impact of the included experimental data on the posterior is studied and an alternative update scheme for the posterior cross section vector is described.

4.1 Modification of the prior

Integral reaction cross sections are determined by measuring the fraction of incident particles experiencing a reaction of a specific type. Differential cross sections include additional information on either the angular distribution or the energy distribution of outgoing particles. Both quantities reflect essential information on the structure of the target nucleus.

From the point of view of informatics, these two quantities are not really different, because they can be modelled by the same data structures. Nevertheless, this work exclusively deals with differential cross sections with respect to angle. If the differential cross section shows no dependence on the azimuthal angle, it is related to the integral cross section by

$$\sigma_c(E) = 2\pi \int_{-\pi}^{\pi} \frac{d\sigma_c}{d\theta}(E, \theta) \, \sin(\theta) \, \mathrm{d}\theta, \tag{4.1}$$

31

where c denotes a certain channel, E is the energy of incident particles and θ is the scattering angle.

At a certain incident energy E, the differential cross section is a function of the scattering angle θ. Since its functional form cannot be expressed analytically, its function values have to be calculated numerically for a discrete set of angle values. Therefore, one ends up with a function defined roughly by a table containing pairs of parameters and function values.

The degree of smoothness of the function $d\sigma/d\theta$ determines the mesh of angles required to describe the underlying angular dependence sufficiently well. For instance, sharp oscillatory features require a rather dense grid to ensure that peaks are not completely between two angle points and thereby remain unrecognised.

The amount of grid points necessary to adequately describe a function can eventually be reduced by expanding it into a sum of linearly independent functions. In this work $d\sigma/d\theta(E, \theta)$ is expanded in terms of Legendre polynomials

$$\frac{d\sigma}{d\theta}(E, \theta) = \sum_{l=0}^{\infty} \alpha_l(E) P_l(\cos\theta). \tag{4.2}$$

The coefficients α_l depend on the incident energy but not on the angle. There are two reasons for the choice of Legendre polynomials as expansion functions: At low incident energies, the first few leading terms already suffice to describe the differential cross section very accurately over the complete angular range (as the energy increases, more terms are necessary for the same level of accuracy). The second reason concerns the possibility to use model deficiency covariance matrices (A^{MD}), generated for integral cross sections in the course of Neudeckers Ph.D. thesis [26]. Inserting the expansion (4.2) into formula (4.1) and relying on the orthogonality between the P_l, it turns out that the first coefficient α_0 completely defines the integral cross section:

$$\sigma_c(E) = 4\pi\alpha_0(E) \ . \tag{4.3}$$

Thus, the integral cross section of some channel is nicely separated from features of the angular distribution. As $P_0(\cos\theta) = 1$, the coefficient α_0 represents an overall shift for the complete angular differential cross section curve. This makes it easy to account for monitoring errors in differential cross section experiments. Uncertainty information included in A^{MD} can be incorporated by adding its elements $cov(\sigma_c(E), \sigma_{c'}(E'))$ divided by $(4\pi)^2$ to the elements $cov(\alpha_{c,0}(E_i), \alpha_{c',0}(E_j))$ of the extended A^{PU}, see Eq. (3.12).

The model cross section space of the original FBET was constituted by vectors of the form

$$\vec{\sigma} = (\sigma_{c_1, E_1}, \ \cdots, \ \sigma_{c_1, E_n}, \ \sigma_{c_2, E_1}, \ \cdots, \ \sigma_{c_2, E_n}, \ \cdots, \ \sigma_{c_m, E_1}, \ \cdots, \ \sigma_{c_m, E_n})^T, \tag{4.4}$$

with σ_{c_i, E_j} being the integral cross section associated to channel c_i at incident energy E_j. As each TALYS calculation returns every cross section at the same incident energies specified in the input file, every model cross section vector contains $m \times n$ elements – m channels and for each channel n incident energies.

In the extended FBET, every σ_{c_i, E_j} corresponding to a channel for which differential cross section data should be taken into account is removed from Eq. (4.4). In their place, a certain amount of coefficients of the Legendre series (4.2) is inserted. As was pointed out in Eq. (4.3), the information on the integral cross section is already contained in α_0. Therefore, a model cross section vector of the extended FBET could read

$$\vec{\sigma} = \left(\sigma_{c_1, E_1}, \cdots, \sigma_{c_1, E_n}, \alpha^0_{c_2, E_1}, \cdots, \alpha^L_{c_2, E_1}, \cdots, \alpha^0_{c_2, E_n}, \cdots, \alpha^L_{c_2, E_n} \right)^T . \qquad (4.5)$$

$\alpha^i_{c_j, E_k}$ represents the i-th order Legendre coefficient of the differential cross section for the channel c_j at incident energy E_k. Since TALYS outputs at most 61 coefficients for a differential cross section curve, this was also the choice for the model cross section vector.

To estimate the hyperparameters $\overline{\sigma}_{\mathrm{mod}}$ and A^{PU} the same procedure as described in section (3.1) is applied. In particular, these two quantities are calculated by relying on Eq. (3.10) and (3.11).

4.2 Transfer from model cross section space to experiment cross section space

In Section 3.2 the procedure to construct mapping matrices to transfer an integral cross section defined on the energy grid of the model to the energies of the experiment was outlined. In addition, differential cross section data have to be mapped from the representation in Legendre coefficients to the angles of the experiment. For the following, it is assumed that the model cross section vector contains only differential cross section data for a single channel c

$$\vec{\sigma}^{\mathrm{DA}}_c = \left(\vec{\alpha}(E_1)^T, \vec{\alpha}(E_2)^T, \cdots, \vec{\alpha}(E_m)^T \right)^T \quad \text{with} \quad \vec{\alpha}(E_i)^T = (\alpha_0(E_i), \cdots, \alpha_L(E_i)) . \quad (4.6)$$

The transformation matrix S can be constructed in a two step procedure: In the first step, the matrix S_E is built up, which maps the Legendre coefficients given for incident energies E_i to the energies E'_j of the experiment (see Fig. 4.1). The effect of S_E is the same as splining the α_i based on model energies E_i and subsequently using the obtained spline to calculate $\alpha_i(E'_j)$. In matrix notation, the mapping reads

$$\begin{pmatrix} \vec{\alpha}(E'_1) \\ \vec{\alpha}(E'_2) \\ \vdots \\ \vec{\alpha}(E'_m) \end{pmatrix} = S_E \begin{pmatrix} \vec{\alpha}(E_1) \\ \vec{\alpha}(E_2) \\ \vdots \\ \vec{\alpha}(E_n) \end{pmatrix} . \qquad (4.7)$$

In the second step, the resulting Legendre coefficients $\alpha_i(E'_j)$ are mapped to the angles $\theta_k(E'_j)$ at which differential cross sections were measured in the experiment. For one
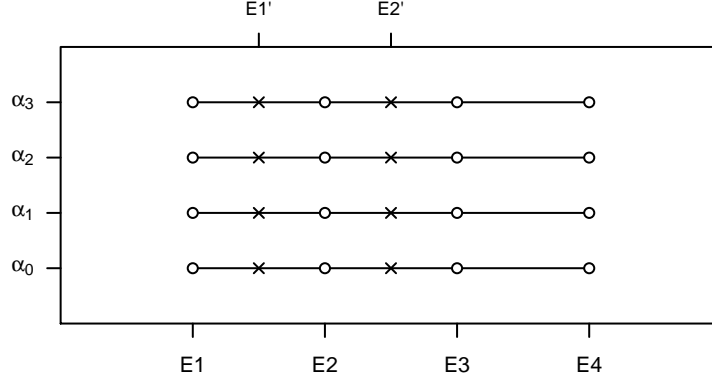
Figure 4.1: Legendre coefficients α_i of every order are splined based on model energy points E_i and subsequently their intermediate values at experimental energies E'_j predicted. The mapping from the set of energies $\{E_i\}$ to energies $\{E'_j\}$ can be expressed as a linear transformation.

particular obtained $\vec{\alpha}(E'_j)$, the mapping is given by

$$
\begin{pmatrix} \theta_1(E'_j) \\ \theta_2(E'_j) \\ \vdots \\ \theta_m(E'_j) \end{pmatrix} = \overbrace{\begin{pmatrix} P_0(\cos\theta_1) & P_1(\cos\theta_1) & \ldots & P_n(\cos\theta_1) \\ P_0(\cos\theta_2) & P_1(\cos\theta_2) & \ldots & P_n(\cos\theta_2) \\ \vdots & \vdots & \ddots & \vdots \\ P_0(\cos\theta_m) & P_1(\cos\theta_m) & \ldots & P_n(\cos\theta_m) \end{pmatrix}}^{S_\theta} \begin{pmatrix} \alpha_0(E'_j) \\ \alpha_1(E'_j) \\ \vdots \\ \alpha_n(E'_j) \end{pmatrix} . \tag{4.8}
$$

In general, in an experiment data at different angles may have been measured at different energies; hence the notation $S_\theta(E'_i)$ is used indicating that the matrix S_θ with angles $\vec{\theta}(E'_i) = (\theta_1(E'_i), \theta_2(E'_i), \cdots)$ provides the mapping at incident energy E'_i. The complete transformation for all $\vec{\alpha}(E'_i)$ to all angles $\vec{\theta}(E'_i)$ is given by

$$
\begin{pmatrix} \vec{\theta}(E'_1) \\ \vec{\theta}(E'_2) \\ \vdots \\ \vec{\theta}(E'_m) \end{pmatrix} = \overbrace{\begin{pmatrix} S_\theta(E'_1) & 0 & \ldots & 0 \\ 0 & S_\theta(E'_2) & \ldots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \ldots & S_\theta(E'_m) \end{pmatrix}}^{S_\theta(\vec{E})} \begin{pmatrix} \vec{\alpha}(E'_1) \\ \vec{\alpha}(E'_2) \\ \vdots \\ \vec{\alpha}(E'_m) \end{pmatrix} . \tag{4.9}
$$

As the diagonal entries $S_\theta(E'_i)$ are rectangular matrices, the '0's are also rectangular zero matrices of appropriate dimension.

The complete mapping matrix S^{DA} performing the transfer of Legendre coefficients to new energies and consecutively transforming them to angles is obtained by

$$
S_c^{\mathrm{DA}} = S_\theta(\vec{E})S_E, \tag{4.10}
$$

where the label c indicates that this matrix maps differential cross section data of a particular channel. Finally, the mapping of a model cross section vector (containing

differential cross section data for multiple channels) to the space of experiments is performed by

$$\begin{pmatrix} \vec{\sigma}^{\mathrm{DA}}_{c_1,\exp} \\ \vec{\sigma}^{\mathrm{DA}}_{c_2,\exp} \\ \vdots \\ \vec{\sigma}^{\mathrm{DA}}_{c_n,\exp} \end{pmatrix} = \begin{pmatrix} S^{\mathrm{DA}}_{c_1} & 0 & \cdots & 0 \\ 0 & S^{\mathrm{DA}}_{c_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & S^{\mathrm{DA}}_{c_n} \end{pmatrix} \begin{pmatrix} \vec{\sigma}^{\mathrm{DA}}_{c_1,\mathrm{mod}} \\ \vec{\sigma}^{\mathrm{DA}}_{c_2,\mathrm{mod}} \\ \vdots \\ \vec{\sigma}^{\mathrm{DA}}_{c_n,\mathrm{mod}} \end{pmatrix}. \tag{4.11}$$

It turns out that mapping matrices S are always block-diagonal, because mapping does not occur between different channels. This is a consequence of the fact that the transformation takes place between different representations of the same cross section data of a specific channel. If necessary, the block diagonal feature can be used to speed up the matrix multiplication conveying the mapping.

4.3 Effect of updating a single channel

The following example demonstrates the effect of the inclusion of experimental data and its covariance via a Bayesian update procedure.

In regard to an experiment, cross section data of the prior can be categorized in two groups: cross section data associated with the channel c of the experiment and the cross section data for other channels. In principle, each of these other channels has to be updated in the same way. Thus, it is sufficient to consider two channels.

Let $\vec{\sigma}_{\exp}$ be the vector composed of M experimental cross section estimates for channel c and B the associated covariance matrix. Model cross section data consisting of N estimates for the same channel are subsumed in $\vec{\sigma}_c$. The vectors $\vec{\sigma}_{\exp}$ and $\vec{\sigma}_{\mathrm{mod}}$ may differ in number of included data points as well as style of representation. For instance, the experiment vector $\vec{\sigma}_{\exp}$ could state differential cross section values at certain angles whereas the model vector states them in terms of Legendre coefficients. Additionally, the prior contains L estimates for another channel $\vec{\sigma}_r$. The necessary quantities for computing the posterior estimate vector (3.33) and covariance matrix (3.29) are

$$\vec{\sigma}_{\mathrm{mod}} = \begin{pmatrix} \vec{\sigma}_r \\ \vec{\sigma}_c \end{pmatrix}, \qquad S = \begin{pmatrix} 0_{M \times L} & S_2 \end{pmatrix}, \qquad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{4.12}$$

The mapping of the model cross section vector to the experimental cross section grid is achieved by

$$\vec{\sigma}^{\mathrm{map}}_{\mathrm{mod}} = S\vec{\sigma}_{\mathrm{mod}} = S_2\vec{\sigma}_c. \tag{4.13}$$

The partial expressions needed for calculating the posterior quantities are

$$\begin{aligned} AS^T &= \begin{pmatrix} A_{12}S_2^T \\ A_{22}S_2^T \end{pmatrix}, & SAS^T &= S_2 A_{22} S_2^T, \\ SA &= \begin{pmatrix} S_2 A_{21} & S_2 A_{22} \end{pmatrix}, & X_c &= (S_2 A_{22} S_2^T + B)^{-1}. \end{aligned} \tag{4.14}$$

Inserting these expressions into Eq. (3.43) and Eq. (3.42) one obtains

$$\tilde{\sigma} = \begin{pmatrix} \tilde{\sigma}_r \\ \tilde{\sigma}_c \end{pmatrix} = \begin{pmatrix} \vec{\sigma}_r + A_{12} S_2^T X_c (\vec{\sigma}_{\exp} - S_2 \vec{\sigma}_c) \\ \vec{\sigma}_c + A_{22} S_2^T X_c (\vec{\sigma}_{\exp} - S_2 \vec{\sigma}_c) \end{pmatrix} \tag{4.15}$$

and

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix} = \begin{pmatrix} \left[A_{11} - A_{12} S_2^T X_c S_2 A_{21} \right] & \left[A_{12} - A_{12} S_2^T X_c S_2 A_{22} \right] \\ \left[A_{21} - A_{22} S_2^T X_c S_2 A_{21} \right] & \left[A_{22} - A_{22} S_2^T X_c S_2 A_{22} \right] \end{pmatrix}. \tag{4.16}$$

These equations for the posterior quantities show that it is possible to update with experimental data for a channel not considered in the prior initially. As long as S_2, X_c together with the samples which were used to produce prior estimate and covariance matrix are available, every other channel not included in the prior in the beginning can be updated later, e.g. when a prediction for it is really needed.

Assume the vectors $\vec{\sigma}_{\exp}$, $\vec{\sigma}_c$ and $\vec{\sigma}_r$ contain only a single data point. Further assume, $\vec{\sigma}_c$ and $\vec{\sigma}_{\exp}$ are given in the same representation, i.e. $S_2 = 1$. For instance, this would be the case if both are integral cross sections for the same incident energy. Under these assumptions, the corresponding quantities in Eq. (4.15) and (4.16) are scalars. The posterior estimate for channel σ_c is given by

$$\tilde{\sigma}_c = \sigma_c + \frac{A_{22}}{A_{22} + B}(\sigma_{\exp} - \sigma_c) = \frac{B\sigma_c + A_{22}\sigma_{\exp}}{A_{22} + B}, \tag{4.17}$$

which is just the weighted mean as already obtained in the example with the one dimensional Gauss distribution, see Eq. (2.25). Prior estimates and covariance elements of other channels have no influence on its prediction. The other model cross section value σ_r for the channel without experimental data is

$$\tilde{\sigma}_r = \sigma_r + \frac{A_{12}}{A_{22} + B}(\sigma_{\exp} - \sigma_c). \tag{4.18}$$

If the experimental covariance B is much smaller than the model covariance A_{22}, then $A_{12}/(A_{22}+B) \approx A_{12}/A_{22}$. In this frequently occurring situation Eq. (4.18) is exactly the formula for linear regression with $\sigma_{\exp}$ resembling the independent predictor variable, σ_c its assumed mean value and σ_r the dependent observable. Furthermore, if the covariance A_{12} vanishes, there is no influence on the channel without experimental data.

The posterior covariance matrix $\tilde{A}$ is exclusively determined by A and B. The model estimate and the experimental estimate have no influence. The variance associated with σ_r is reduced by

$$\Delta A_{11} = -\frac{A_{12}^2}{A_{22} + B} \overset{B \ll A_{22}}{\approx} -\mathrm{corr}^2(\sigma_r, \sigma_c) A_{11}. \tag{4.19}$$

The squared correlation coefficient (Eq. 3.13), yields the percental reduction of A_{11} if the experimental variance is much smaller than the model variance.

4.4 Another update approach

Besides the Bayesian update formula for the mean (3.42) there is the possibility of directly weighting the samples with regard to experimental data. Under the assumption that sampling vectors $\vec{X}_i$ are distributed according to a multinomial Gauss distribution

$$\vec{X}_i \sim \mathcal{N}(\vec{\mu}, \Sigma), \tag{4.20}$$

both update schemes yield the same result in expectation. The quantity $\vec{\mu}$ is the real center point of the Gaussian distribution which is estimated by $\vec{\sigma}_{\text{mod}}$ from the drawn samples $\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_N$ (Eq. (3.10). The true covariance matrix Σ is estimated by the sample covariance matrix A (Eq. 3.11). Assume experimental data $\{\vec{\sigma}_{\text{exp}}, B\}$, the posterior estimate $\tilde{\sigma}$ can be calculated by

$$\tilde{\sigma} = \frac{1}{N} \sum_{i=1}^{N} \vec{X}_i \, C_L \exp\left(-\frac{1}{2}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}_i\right)^T B^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}_i\right)\right), \tag{4.21}$$

with C_L the normalization factor for the likelihood. This yields under the assumption (4.20) in expectation the same result as

$$\tilde{\sigma} = \vec{\sigma}_{\text{mod}} + AS^T(SAS^T + B)^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{\sigma}_{\text{mod}}\right). \tag{4.22}$$

In order to prove this statement, the definition of the expectation value is required,

$$E[X] = \int_{V(X)} X \, \rho(X) \, \mathrm{d}X, \tag{4.23}$$

with X a random variable X and the probability density function $\rho(X)$. The integration is performed over the complete range of allowed values. The expectation operator E is linear

$$E[cX + Y] = cE[X] + E[Y]. \tag{4.24}$$

Using (4.23) and (4.24), the expectation value of (4.21) can be obtained,

$$E\left[\frac{1}{N}\sum_{i=1}^{N} \vec{X}_i \, C_L \exp\left(-\frac{1}{2}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}_i\right)^T B^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}_i\right)\right)\right] \tag{4.25}$$

$$= E\left[\vec{X} C_L \exp\left(-\frac{1}{2}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}\right)^T B^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}\right)\right)\right] \tag{4.26}$$

$$= \int \vec{X} \, C_L \exp\left(-\frac{1}{2}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}\right)^T B^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{X}\right)\right) \tag{4.27}$$

$$C_P \exp\left(-\frac{1}{2}\left(\vec{X} - \vec{\mu}\right)^T \Sigma^{-1}\left(\vec{X} - \vec{\mu}\right)\right) \mathrm{d}\vec{X}$$

$$= \vec{\mu} + \Sigma S^T(S\Sigma S^T + B)^{-1}\left(\vec{\sigma}_{\text{exp}} - S\vec{\mu}\right), \tag{4.28}$$

with the normalization factor C_P for the Gauss distribution of the vectors X_i. It was possible to write the result of the integral (4.27) immediately, because the multiplication

of the two Gauss distributions is of the same form as in section 3.2, where the expression for the mean was proven to be of the form (4.28). Since A and $\vec{\sigma}_{\mathrm{mod}}$ are unbiased estimators, the expectation value of (4.22) is also given by (4.28).

However, the validity of this result requires the assumption (4.20) to hold. If it does not hold, the weighting procedure yields a better estimate for the true mean in expectation as the Bayesian update procedure does. This is due to the fact that the Bayesian update procedure assumes a Gaussian pdf for the prior whereas no assumptions are made for the prior pdf in the weighting approach – it is implicitly determined by the sampling process and the utilized pdf for the parameters of the optical potential.

Unfortunately, tentative application of the weighting approach indicated that the number of samples which can be calculated in a reasonable amount of time is far too low to obtain reasonable results. Experimental data is usually measured with uncertainties in the percent range. Hence, if only a few points in a sample vector deviate from the experimental estimates by a few percent, the assigned weight by the experimental data is already very low. Furthermore, weights of different sample vectors differ orders of magnitude which causes the posterior estimate to be almost exclusively determined by the sample vector with the highest weight. In order to circumvent this problem, it would be necessary to apply importance sampling, i.e. adaptively alter boundaries for the parameters of the optical potential to obtain vectors with significant weights.

5 Application of the extended Full Bayesian Evaluation Technique to tantalum-181

5.1 Parameter uncertainties for tantalum-181

The nuclear model code TALYS 1.2 [23] was used to produce 1000 cross section samples containing total CS, integral elastic CS and differential elastic CS with neutrons as incident particles. Each of these cross sections was computed for the same set of 75 incident energies ranging from 0.3 MeV to 160 MeV. With increasing incident energy, the distance between two energies of the grid was also increased. The following table depicts the utilized step sizes for different energy intervals:

energy [MeV]	0.3	1.0	10	30	50	60	160
increment [MeV]	0.1	0.2	1	2.5	5	10	–

Table 5.1: Two consecutive energies E_1 and E_2 in the first row define an energy interval. The distance between the mesh points in this interval is given by the value in the second row below E_1.

For example, in the energy interval from 0.3 MeV to 1.0 MeV a mesh size of 0.1 MeV was used. From these samples, the mean cross section vector and the covariance matrix were calculated to define the Gaussian prior.

One of the models implemented in TALYS is the optical potential approach [17] with the global parametrisation of Koning and Delaroche [22]. There is an optical potential assigned to every particle type such as neutrons and protons. Because only neutrons were considered as incident particles, it would have sufficed to specify the parameters for the neutron optical potential to describe the total and shape-elastic cross section well in the unresolved energy range. However, as experiments are unable to distinguish between shape- and compound-elastic CS and hence can only measure the total elastic CS, the calculation of the compound-elastic CS is necessary to compare model predictions against experimental results.

The compound-elastic CS is part of the reaction channel and competes with other reactions and therefore it depends on optical potential parameters of other involved

particles. In addition, the parameters of the level density models implemented in TALYS have an impact on these channels.

Nonetheless, an analysis showed that the compound-elastic CS is of minor significance for the integral elastic CS. Above an incident energy of $2\,\mathrm{MeV}$, its fraction of the integral elastic CS is less than 3% (Fig 5.1). Thus, one will not lose important features by varying only the neutron optical potential parameters. This fact was not clear in the beginning, and therefore the neutron optical potential parameters as well as level density parameters for the Fermi gas model were varied.

The parameter boundaries were taken as obtained by D. Neudecker during the course of her Ph.D. thesis [26]. They are given in Table 5.2. Values in between were sampled according to an uniform distribution. A random number generator of Mersenne-Twister [25] type was used in order to avoid the unpleasant feature of linear congruential generators which deliver random numbers restricted to a set of parallel hyperplanes.

Neutron optical potential (global parametrisation)					
r_v	1.0645	1.4000	d_1	10.6937	15.1184
a_v	0.4700	0.8318	d_2	0.0171	0.0192
v_1	40.4318	61.3587	d_3	9.8829	13.1171
v_2	0.0051	0.0088	r_{SO}	0.9098	1.2322
v_3	0.000016	0.000016	a_{SO}	0.4500	0.7300
w_1	13.3422	17.0932	v_{SO1}	5.1359	7.7941
w_2	79.1430	96.7360	v_{SO2}	0.0035	0.0045
r_d	1.0715	1.4340	w_{SO1}	-2.7179	-3.4821
a_d	0.3716	0.6576	w_{SO2}	160.000	160.000
Level density parameters (Fermi gas model)					
α	0.011402	0.030059	γ_1	0.331541	0.615719
β	0.126247	0.332833	γ_2	0.000000	0.000000
Δ	-3.000000	3.000000			

Table 5.2: Lower and upper boundaries of neutron optical potential and level density parameters used in the sampling process. Intermediate values were selected according to an uniform distribution.
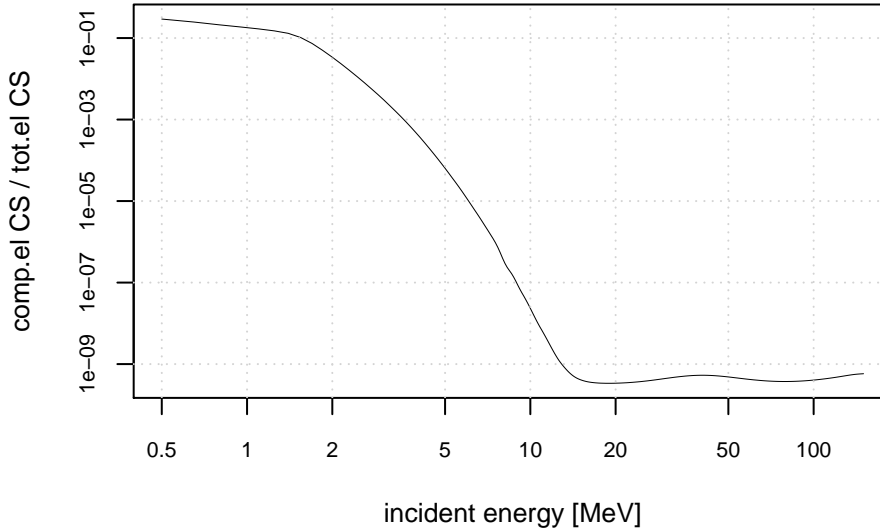
Figure 5.1: Portion of the compound-elastic CS on the integral elastic CS. It amounts to 30% at 0.5 MeV, drops below 10% at 1.6 MeV and is about 3% at 2 MeV

5.2 Prior studies

In principle, the Full Bayesian Evaluation Technique allows the use of very general probability density functions (pdf). However, in actual implementations for pragmatic reasons the prior is assumed to be of Gaussian shape. As was pointed out, the probability distribution in cross section space results from the uniform distribution in parameter space and is given by $G^{-1/2}$, where G is the Gram determinant (see Eq. 3.8).

Thus, the applicability of the actual implementation of the FBET depends on whether $G^{-1/2}$ resembles the essential features of a Gauss distribution. The high dimensionality of the parameter space – 16 parameters of the neutron optical potential are varied – renders the calculation of G for a sufficiently dense grid almost impossible. This is the reason why the mapping from parameter space to cross section space is only determined for a sample of parameter sets. The sample size is far too small to obtain a non-parametric estimate of the resulting probability density distribution. Hence, the calculation is restricted to the lowest two moments of the cross section vectors. Knowing only these two quantities, the principle of maximum entropy leads to a pdf of Gaussian shape.

In the following we study the features of the sample distribution in order to gain knowledge on the true distribution and thus of the validity of the implemented FBET. The more features the real distribution shares with the Gauss distribution, the better the prior implements the systematics of the model.

Another aim is the investigation of the mapping from parameter space to cross section space established by the model. In the vicinity of a parameter vector, the function can be approximated by a linear mapping from $\mathbb{R}^{16}$ (due to 16 optical potential parameters) to a 16-dimensional linear subspace in cross section space, see (Eq. (3.7)). Globally, the

41

parameter vectors are mapped to a manifold in cross section space which may be bent as a whole and may exhibits points of contraction. Points located within a small volume in the vicinity of these points of contractions get mapped to an even smaller volume in cross section space leading to an increased probability density there. Hereby, volumes are defined by the Lebesgue measure[1].

The following analyses were carried out with the statistical programming language R [29]. In particular, the following packages were utilized: 'robustbase' [32] for outlier detection, 'np' [16] for non-parametrically density estimation, 'mclust' [12, 13] for cluster analysis and 'car' [11] for scatterplots.

5.2.1 Analyzing (n,tot)

The total CS was calculated for 75 incident energies. Therefore, every obtained cross section curve can be regarded as a vector in a 75-dimensional vector space.

Outliers. First of all, an outlier test based on the Stahel-Donoho outlyingness [14] was performed. The utilized measure of outlyingness for a vector X_i of the sample is given by

$$\text{out}(X_i) := \sup_{\alpha} \left(\text{out}(X_i, \alpha)\right) = \sup_{\alpha} \left(\frac{|\alpha^T X_i - \text{med}(\alpha^T X)|}{\text{IQR}(\alpha^T X)} \right), \tag{5.1}$$

where α is a normalized vector; $\alpha^T X_i$ represents the projection of X_i onto the direction given by α; $\text{med}(\alpha^T X)$ denotes the median of the projected samples $\{X_j\}, j = 1, \ldots, n$ with n being the number of sample vectors and $\text{IQR}(\alpha^T X)$ denotes the interquartile range of the projected sample vectors. Thus, the outlyingness of a vector X_i is the maximal distance occurring for this vector from the median normalized by the IQR in a particular projection direction.

In practice, the outlyingness can only be approximately computed by projecting the samples onto hundreds or thousands of randomly selected vectors α and take the highest obtained value for $\text{out}(X_i, \alpha)$ as the outlyingness.

A cross section vector was treated as outlier if its outlyingness was greater than two. If samples are obtained according to a Gauss distribution the outlier test is expected to identify about 0.5% of the vectors as outliers which are actually not.

The number of projection directions was increased until convergence in the number of outliers was observed. 66 curves out of 1000 were identified as outliers – by far more than the expectable 0.5%. Some of these are shown in the top row of Fig. 5.2. Nearly all of them possess remarkable features at $3.2\,\text{MeV}$. Either there is a huge reduction of the CS perceivable as a sharp negative peak or local oscillations occur. Further analysis revealed the negative peak to stem from the calculated reaction cross sections (see Fig. 5.3).

This sudden drop – in some cases over several barn – appears unphysical and can probably be attributed to numerical instabilities in the TALYS code. Thus, these outliers were removed from the sample for the subsequent analysis.

---

[1]In one, two and three dimensions this amounts to length, area and volume in the usual sense.

Additionally, the sampled vectors were classified according to their number of maxima. The number of maxima (= number of minima) ranged from two to six and a negligible amount of vectors – all identified as outliers – featured more than four maxima (cf. Fig. 5.4).
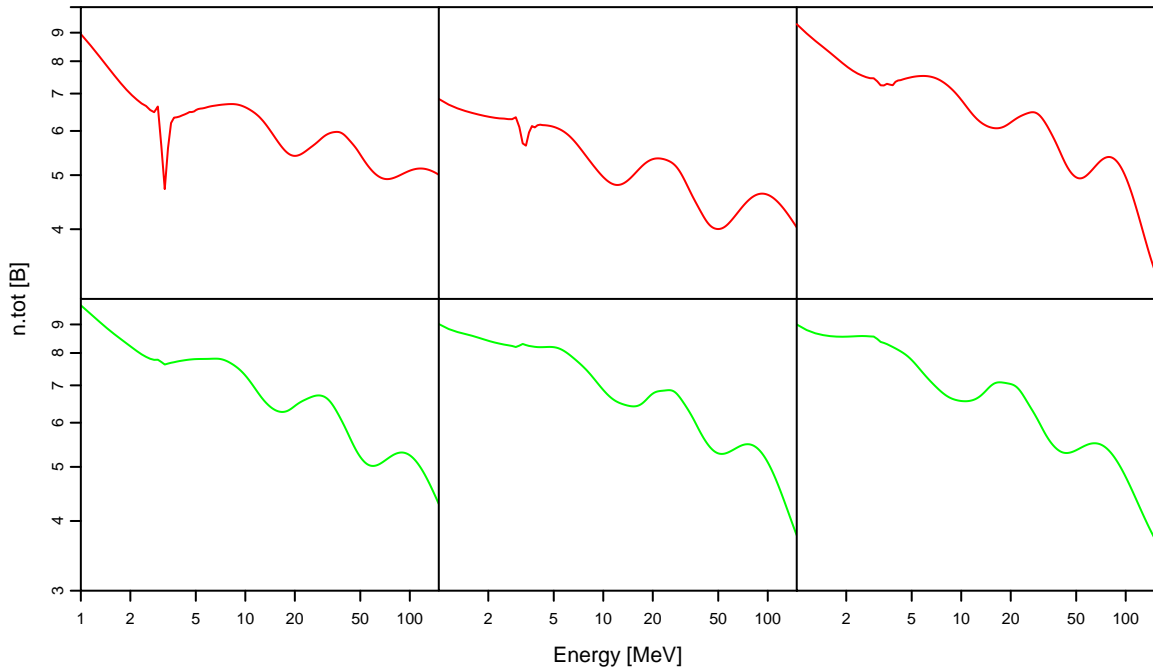
Figure 5.2: The top row displays some of the total cross section curves rated worst by the outlier test. The bottom row shows the worst rated vectors after outliers were removed. Nearly all outliers exhibit peculiar features at 3.2 MeV.

43
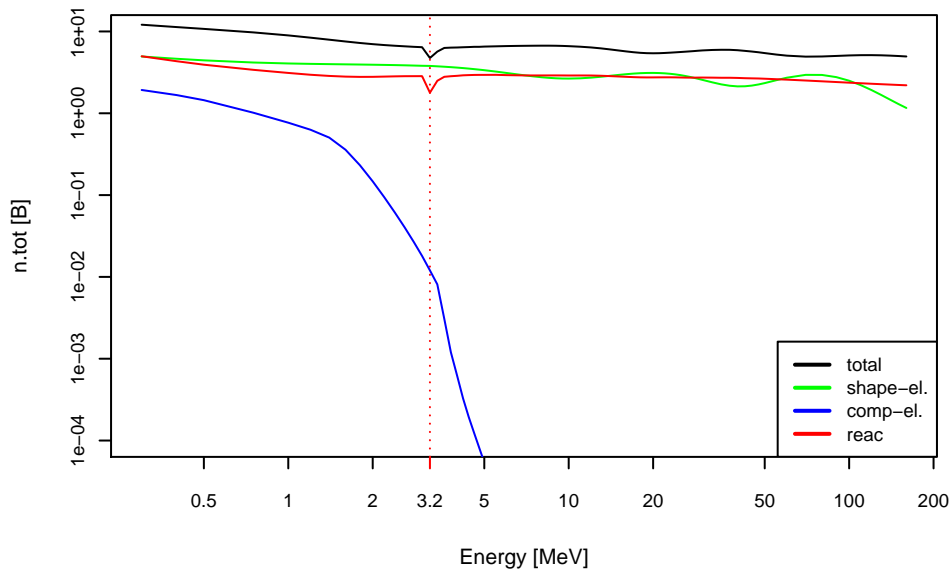
Figure 5.3: Total, shape-elastic, compound-elastic and reaction cross section in comparison for the top-left curve of Fig. 5.2. The negative peak is caused by the reaction channel. Since the compound-elastic CS has a small portion of the total CS, the negative peak must be associated with other reaction channels.
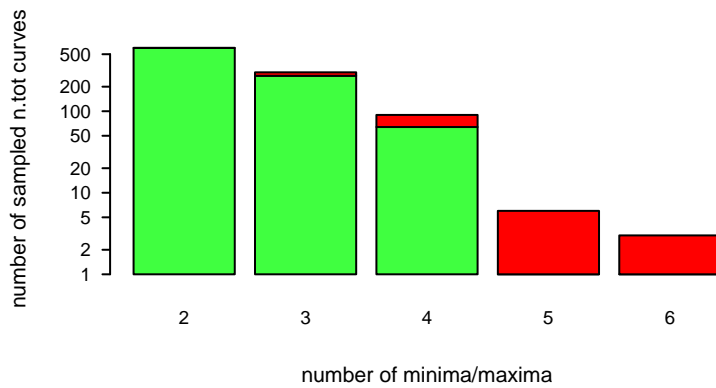
Figure 5.4: Classification of the vectors within the sample according to their number of maxima (=number of minima). The stacked red bars show the number of outliers within each class. Classes associated with higher numbers of maxima contain a higher percentage of outlier curves.

44

Principal component analysis. Since only 16 optical potential parameters were varied and level density parameters are – except perhaps for the outlier vectors – of minor significance for the result, cross section curves must not have more than about 16 degrees of freedom. In the case of a linear mapping from parameter space to cross section space, the curves would be restricted to a linear 16-dimensional subspace. Non-linear features of the mapping may cause the cross section manifold to be bent, thereby reaching more than 16 dimensions (see Fig. 3.1). To study how many dimensions are important, an eigenvalue decomposition of the covariance matrix was performed. This pictorially means to align the coordinate axes parallel to the principal axes of the assumed Gauss distribution. The eigenvalues represent the variances along these axes. Directions associated with the highest variances contain the most significant information about the features of a vector. This decomposition technique, invented in 1901 by K. Pearson, is called principal component analysis [27].

The result of the eigenvalue decomposition is illustrated in Fig. 5.5. The corresponding normalized eigenvectors are visualized in Fig. 5.6. The stated 'mean' values denote the expectation and the stated 'sd' values the uncertainty (standard deviation) the prior assigns to the directions given by the eigenvectors. The prior estimate for the total CS curve (cf. 5.28) can be obtained by $\sigma(E) = m_1 v_1 + \cdots + m_n v_n$ with m_i being the mean value for the direction given by eigenvector v_i. The 1-σ-confidence interval can be calculated by $\delta = (\delta_1^2 + \cdots + \delta_n^2)^{1/2}$ with δ_i being the standard deviation in direction of the eigenvector v_i.
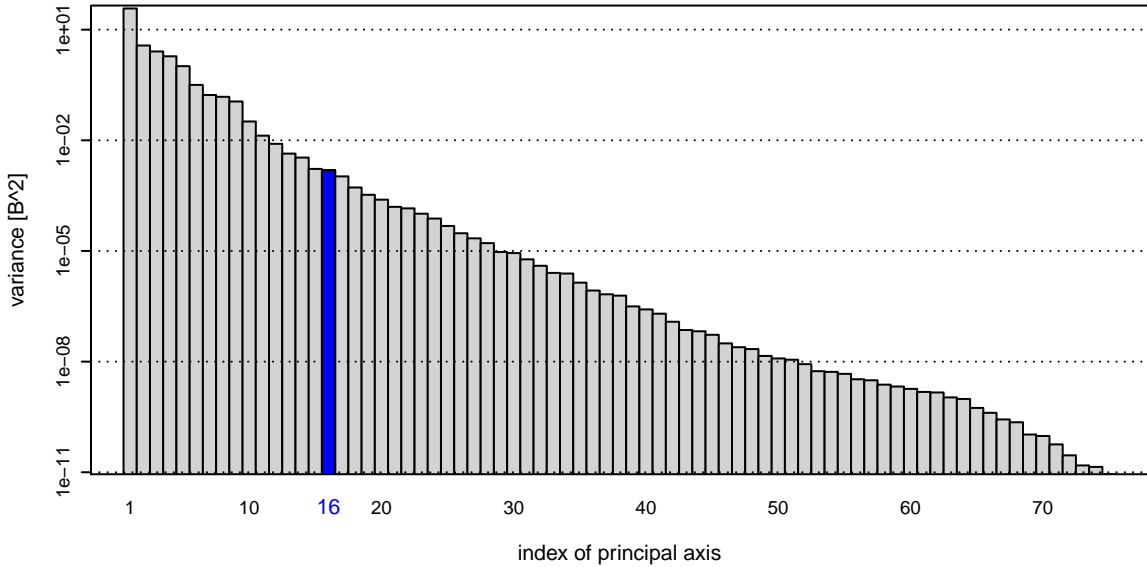
Figure 5.5: The eigenvalues of the covariance matrix for (n, tot) in descending order. They represent the variances along the principal axes of the assumed Gauss distribution. Their rapid decline indicates there are a lot of negligible directions.
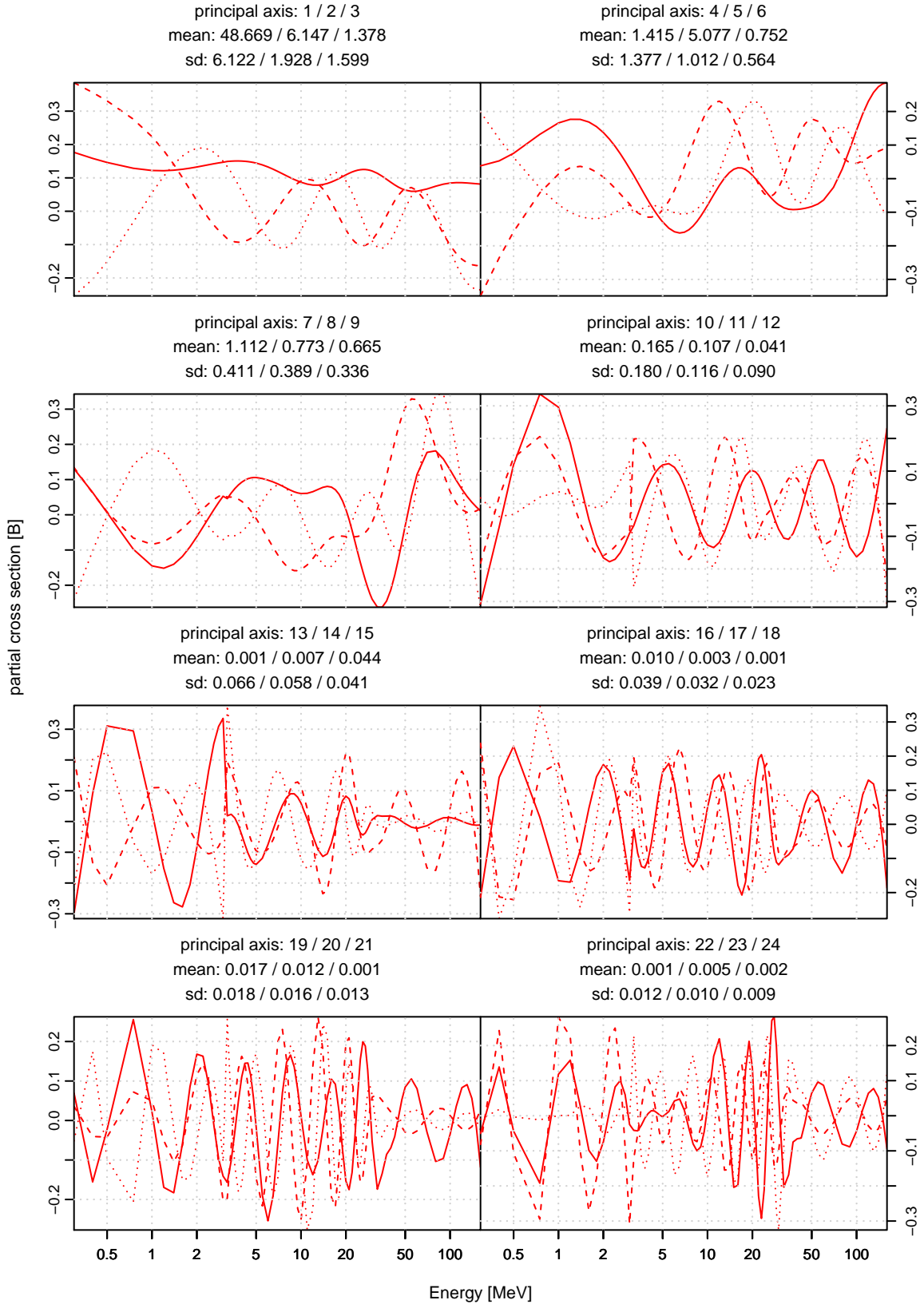
Figure 5.6: Normalized eigenvectors of the prior covariance matrix for the total CS

46

As can be seen, the variances rapidly decrease and the variance associated with the 16th principal axis is already about a factor 10^{-4} smaller than the one of the first principal axis. Hence, a linear combination of only a few of the eigenvectors associated with the biggest eigenvalues should suffice to reproduce the majority of the cross section vectors of the sample.

In order to determine the number of required eigenvectors $\vec{e}_i$, the following procedure was applied: All cross section vectors within the sample were projected onto the principal axes of the covariance matrix. Thereafter, only the coordinates x_i for the axes corresponding to the n biggest eigenvalues were kept, all other were set to zero. Each cross section vector was subsequently approximated by a linear combination of the eigenvectors associated with the coordinates which were kept: $x_1 \vec{e}_1 + x_2 \vec{e}_2 + \cdots + x_n \vec{e}_n$. The resulting vectors were compared to the original ones in terms of several error measures. As error measure for a single curve, either the maximal absolute residue or the median of the absolute residues was taken. The error measure for the whole sample was either given by the maximal error occurring for a vector therein or by the median of the errors for the vectors. These errors were calculated for different numbers of considered coordinates n. The results are shown in Fig. 5.7. Furthermore, the same error measures were used for the relative residues (Fig. 5.8). These figures indicate that the leading 16 eigenvectors suffice to describe all vectors with good accuracy. The median of the single curve errors given by the maximal residue is about 10 millibarn; the maximal percental deviation occurring in the sample lies below 1%. The median of the maximal percental deviation is in the magnitude of one-tenth of a percent (see Fig. 5.8). Thus, the model can be indeed understood as a mapping from $\mathbb{R}^{16}$ to a linear 16-dimensional subspace in cross section space. Moreover, even without further research, the assumption of bijectivity seems very plausible.

In order to gain an even better understanding of the quality of the approximation, the probability density to obtain a certain relative deviation of the approximated vector from the original one was estimated for each incident energy in the grid. In other words, the conditional probability density $\rho(r_{\text{rel}}|E)$ was calculated. The result is shown in Fig. 5.9.
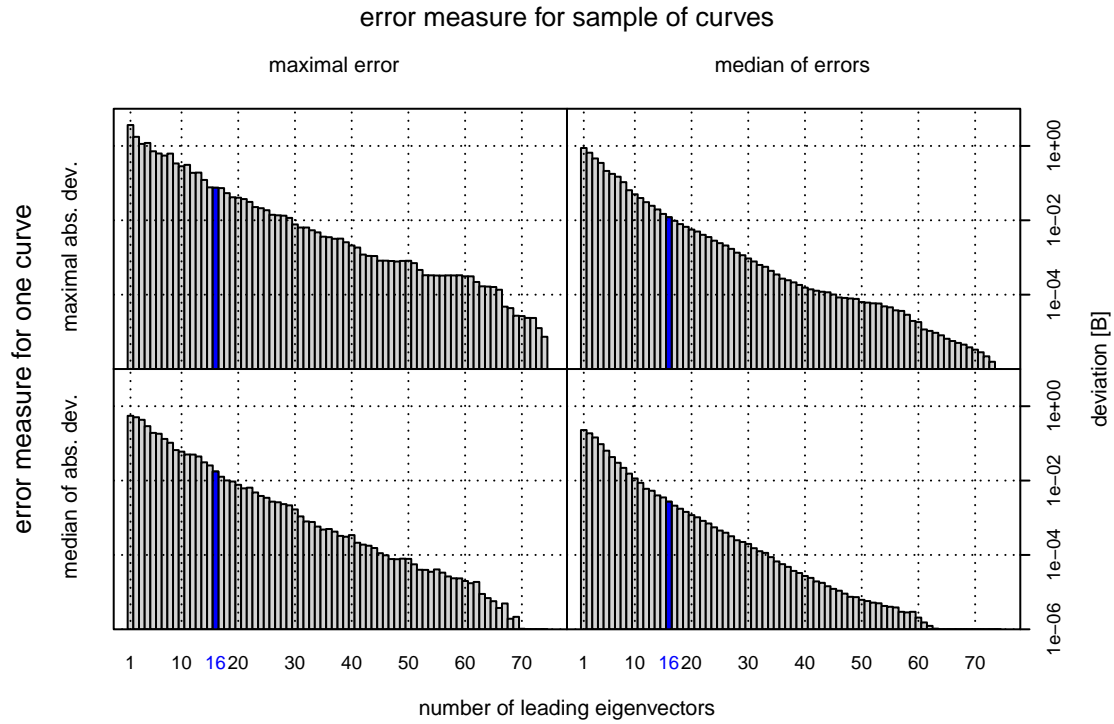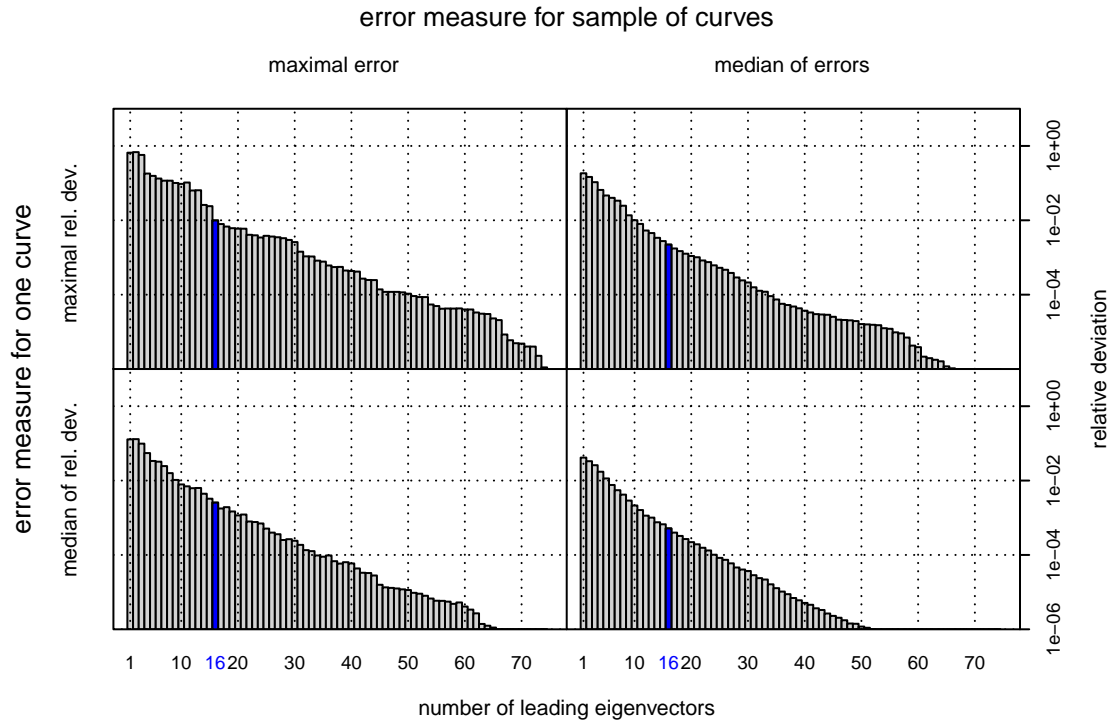
Figure 5.7: The bars represent the error for the whole sample if approximations with a specific number of eigenvectors are used. The error for the whole sample is based on different error measures for a single vector. For the sub-diagrams of the top row, the error measure for a vector is given by the maximal absolute deviation of the approximation from the original vector. For the sub-diagrams of the bottom row, the error measure for a vector is given by the median of the absolute deviations of the approximation from the original vector. The error measure for the whole sample in the left column is determined by the maximal error occurring for a vector within the sample. The error measure for the whole sample in the right column is determined by the median of the errors for the vectors in the sample.

Figure 5.8: The bars represent the error for the whole sample if approximations with a specific number of eigenvectors are used. The error for the whole sample is based on different error measures for a single vector. For the sub-diagrams of the top row, the error measure for a vector is given by the maximal relative deviation of the approximation from the original vector. For the sub-diagrams of the bottom row, the error measure for a vector is given by the median of the relative deviations of the approximation from the original vector. The error measure for the whole sample in the left column is determined by the maximal error occurring for a vector within the sample. The error measure for the whole sample in the right column is determined by the median of the errors for the vectors in the sample.
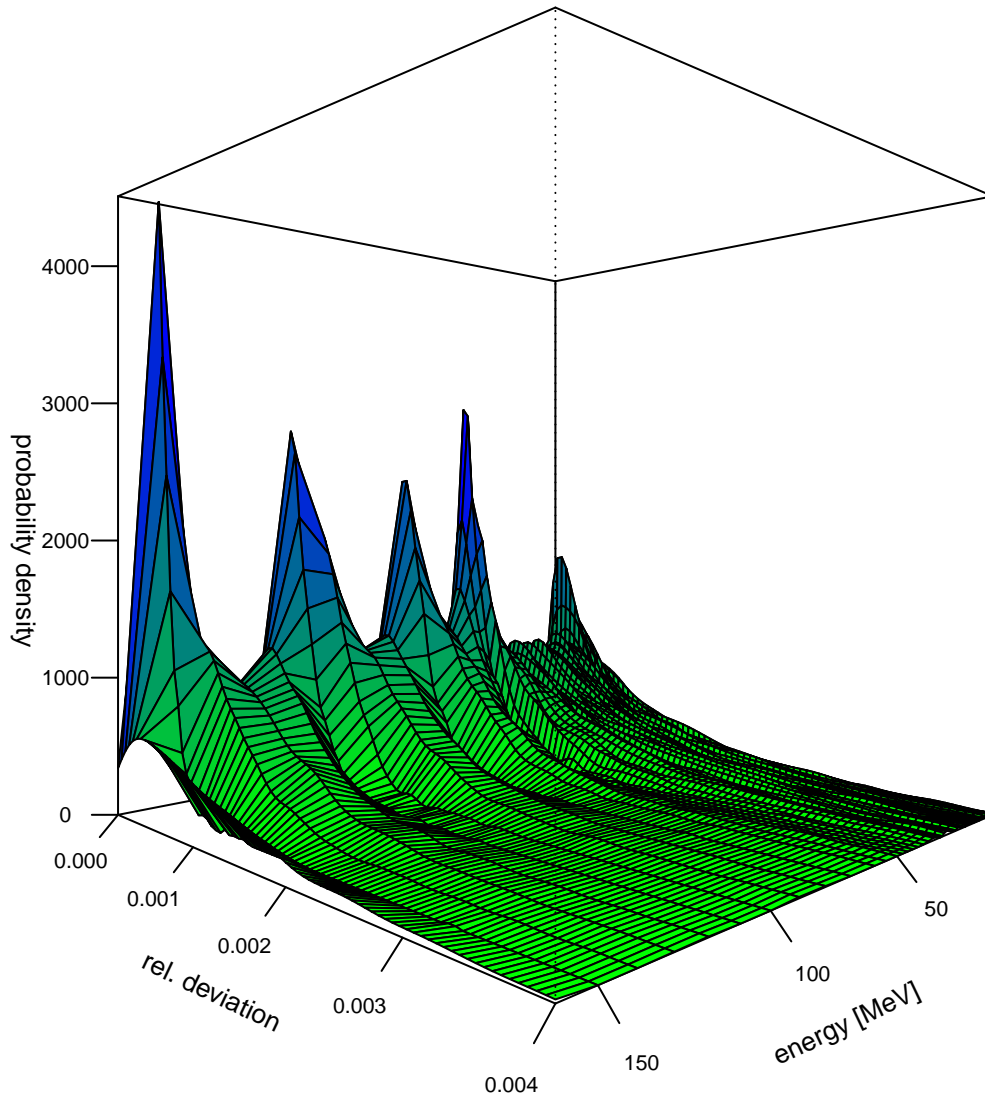
Figure 5.9: The estimated probability density for the relative deviation of the approximated (=dimension reduced) curves from the original vectors at each incident energy. For a randomly selected cross section curve of the sample, the probability for the approximation of this curve to deviate from the original by more than 0.3% at any energy is negligible.

Investigation of the mapping. The previous analysis has indicated the existence of a mapping from $\mathbb{R}^{16}$ to a 16-dimensional linear subspace in cross section space, the question arises whether it is approximately even a linear or quadratic mapping. Thus, the following three relationships were considered:

$$f_1: \qquad \sigma_i = a_i + \sum_{j=1}^{n} b_{ij} P_j \tag{5.2}$$

$$f_2: \qquad \sigma_i = a_i + \sum_{j=1}^{n} b_{ij} P_j + \sum_{j=1}^{n} c_{ij} P_j^2 \tag{5.3}$$

$$f_3: \qquad \sigma_i = a_i + \sum_{j=1}^{n} b_{ij} P_j + \sum_{j=1}^{n} \sum_{k=j}^{n} c_{ijk} P_j P_k \tag{5.4}$$

where σ_i is the i-th component of the cross section vector, P_j the j-th component of the parameter vector, $n = 16$ the number of parameters and a_i, b_{ij}, c_{ijk} the coefficients which have to be determined. The values of the latter were obtained by a linear regression analysis for each σ_i individually. Neither the linear model f_1 nor the quadratic models without mixed terms f_2 and with mixed terms f_3 were able to describe the cross section curves in a satisfying manner. The biggest deviations were found below 2 MeV. Above this energy, the approximations f_1, f_2 and f_3 were adequate for most vectors. As was to be expected, the quadratic approximation with mixed terms f_3 always outperformed f_1 and f_2. One sample vector for which the approximations were less satisfactory is shown in Fig. 5.10.
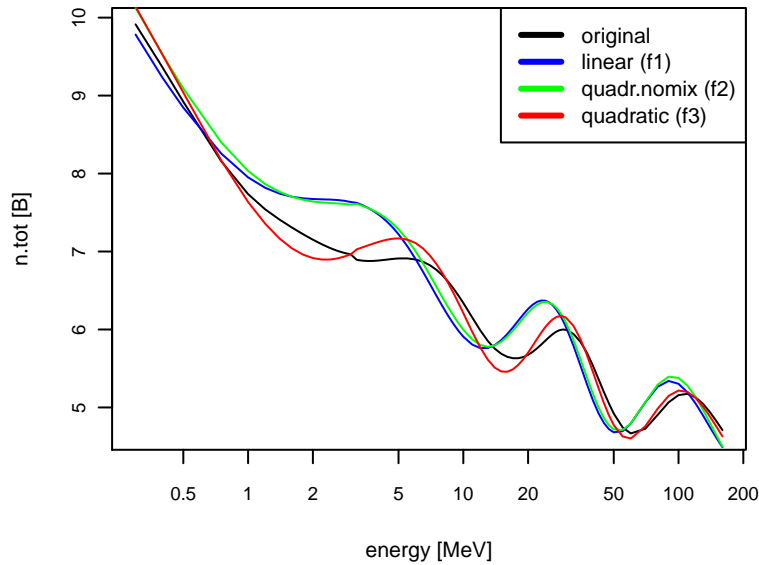
Figure 5.10: A total cross section curve which was not approximated well by f_1, f_2 and f_3. The quadratic relationship f_3 outperforms the linear relationship f_1 and the quadratic relationship without mixed terms f_2. As f_2 is not better than f_1, the mixed terms $P_i P_j$ seem to be important.

Marginal probability density. Although the cross section space has too many dimensions to estimate the pdf in a non-parametric way, the marginal probability densities for two degrees of freedom can be estimated well. This means to project the sample cross section vectors onto a 2-dimensional plane and estimate the probability density for the points thereon. If the distribution in cross section space is a multi-dimensional Gauss distribution, the cross section vectors projected onto arbitrary planes should also follow a Gauss distribution. The first two principal axes of the assumed Gauss distribution are associated with the highest variances. Hence, they contain the highest amount of information about the shape of a particular cross section vector and thus their joint probability density distribution was examined. The estimate for the joint pdf was determined with the aid of the R-package 'np' [16]. It implements kernel-density-estimation with likelihood cross-validation. A diagonal Gaussian kernel was chosen, therefore the
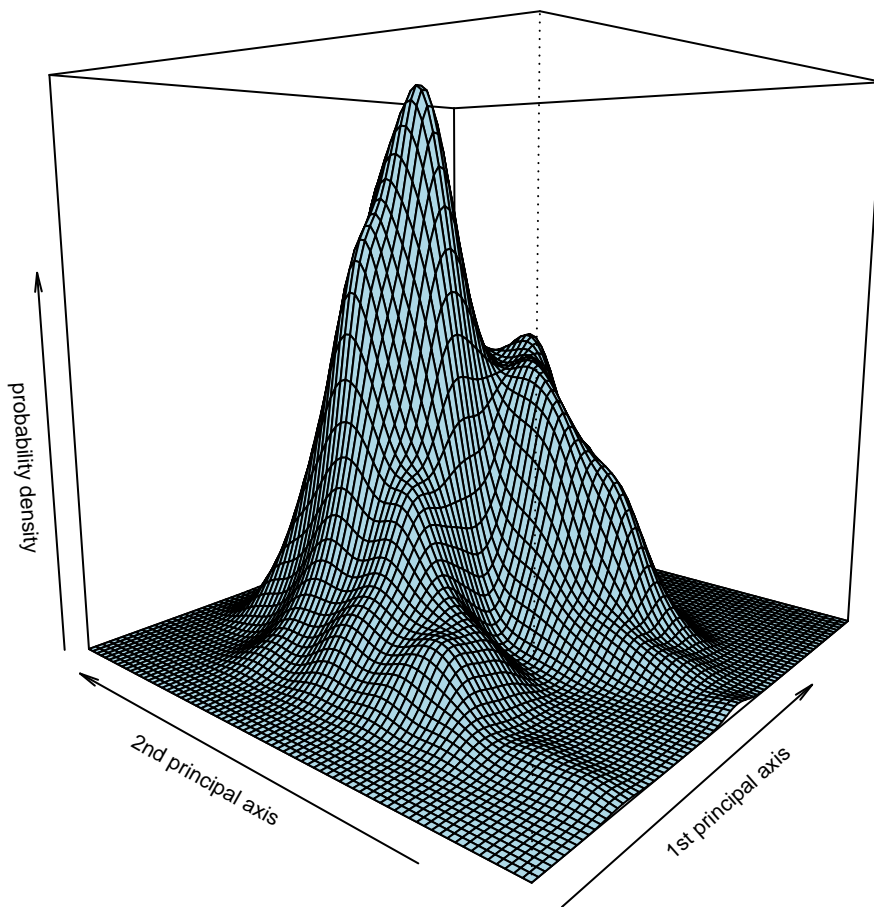
Figure 5.11: The probability density function projected onto the first two principal axes obtained via kernel density estimation.

true pdf was approximated by

$$\hat{f}(x_1, x_2; h_1, h_2) = \frac{1}{2\pi n h_1 h_2} \sum_{i=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{x_1 - x_{i,1}}{h_1}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{x_2 - x_{i,2}}{h_1}\right)^2\right),$$
(5.5)

where the coordinate of the first principal axis is x_1 and of the second principal axis x_2. The quantities $x_{i,1}$ and $x_{i,2}$ are the projections of the i-th sample vector onto the first and second principal axis respectively. The optimal bandwidths h_1 and h_2 for determining the degree of smoothing were obtained by likelihood cross-validation. This method splits the sample vectors repeatedly into two sets, a training set and a validation set, and generates an estimate of the pdf (5.5) by using only vectors from the training set. The log-likelihood according to the obtained pdf is subsequently calculated for the validation set. It is defined by

$$\log L(x_1, \ldots, x_n | \hat{f}) = \sum_{i=1}^{n} \log(\hat{f}(x_{i,1}, x_{i,2})) .$$
(5.6)

Values for the bandwidth parameters h_1 and h_2 are selected which maximize the log-likelihood function. The result is presented in Fig. 5.11.

Obviously, the pdf does not really have a Gaussian shape. The local maximum located on the right hand side of the global maximum is related to a volume where the Gram determinant takes low values. However, as this is only the pdf for a projection of the 75-dimensional cross section space onto a plane, more points of contraction could exist which are not steadily discernible in this projection plane. It would be an analogous problem to try to extract information about the position and structure of furniture in a room by only looking at the shadows on one wall. If the shadows of two objects overlap, it is not possible to discriminate between them. Perhaps the bump in the front is also associated with a point of contraction.

Cluster analysis. In order to detect preferably all points of contraction a clustering algorithm [12] was applied. Firstly, all cross section curves were projected onto the subspace spanned by the leading 15 eigenvectors of the covariance matrix associated with the highest variances. After this reduction of dimension, different Gaussian mixture models were fitted by the expectation-maximisation algorithm [7] as implemented in the R-package 'mclust' [13] to describe the pdf of the reduced cross section vectors as well as possible. A Gaussian mixture model is defined by

$$\rho(x) := \sum_{k=1}^{K} \omega_k \, (2\pi)^{-p/2} \det(\Sigma_k)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right),$$
(5.7)

with the weighting factor ω_k, the mean vector μ_k and the covariance matrix Σ_k for the k-th Gauss distribution, also called the k-th component. The dimension p is given by the number of elements of the occurring vectors x and μ_k. In order to ensure normalization the ω_k have to sum up to one.

The suitability of several Gaussian mixture models which differ in the number of components and number of degrees of freedom for each component was examined. The specific restrictions which are imposed on the components can be explained on the basis of the covariance matrices, parametrised in the following way [13, p. 53]:

$$\Sigma_k = \lambda_k D_k A_k D_k^T, \tag{5.8}$$

with the orthogonal matrix of eigenvectors D_k, the diagonal matrix A_k containing the scaled eigenvalues $\tilde{e}_{k,i}$ of Σ_k and a scalar λ_k. The scalar and the scaled eigenvalues are determined by the conditions $\prod_{i=1}^p \lambda_k \tilde{e}_{k,i} = 1$ and $\lambda_k \tilde{e}_{k,i} = e_{k,i}$ where $e_{k,i}$ denotes the original unscaled eigenvector. D_k determines the orientation, i.e. the alignment of the principal axes, A_k determines the shape and λ_k determines the volume of an ellipsoid resembling the density contours of a component k.

Only the specific restrictions of the models tagged EEV, VEV, VVV in Fig. 5.12 are described since they are of particular interest. For a general description of all probability density models occurring in Fig. 5.12, the reader is relegated to the 'mclust' manual [13].

The model EEV constrains all components to have the same values for λ_k and A_k and allows them to have only different choices for D_k. Thus, the ellipses defined by the isosurfaces of probability density for different components have equal volumes and equal shapes. They are only allowed to differ in their orientation. The model VEV permits individual choices of λ_k and A_k for the components, therefore they are restricted to have equal shape but are allowed to differ in volume and orientation. Finally, the model VVV permits the components to differ in volume, shape and orientation.
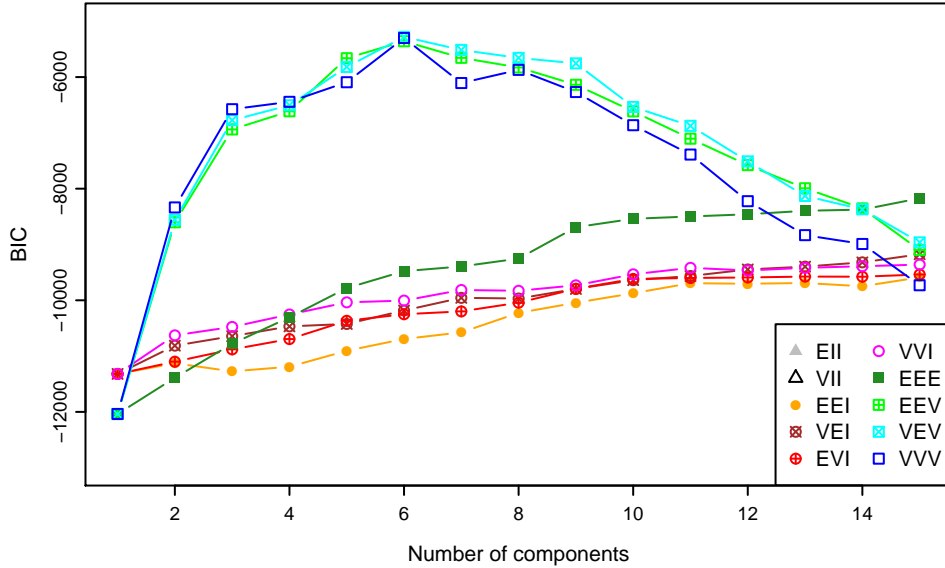
Figure 5.12: BIC values for different models. The models EEV (=equal volume and shape, varying orientation), VEV (=equal shape, varying volume and orientation) and VVV (=varying volume, shape and orientation) obtained the highest values for 6 components.

54

These models were fitted to the dimension reduced cross section sample vectors by the expectation-maximisation algorithm [7]. In order to determine their suitability the Bayesian information criterion (BIC) [33] was used. The BIC value for a model is given by

$$BIC \equiv 2 \log L - N \log n, \tag{5.9}$$

where L is the maximised likelihood of the model, N is the number of free parameters of the model and n is the number of observations, i.e. number of sample vectors. The higher the BIC value the better rated the model. The likelihood is given by

$$L = \prod_{i=1}^{n} \rho(X_i), \tag{5.10}$$

where X_i denotes a sample vector and ρ is the probability density function (Eq. 5.7). The Bayesian information criterion is a measure for goodness of fit which takes into account that an increasing amount of model parameters automatically improves the fit although the model is still inadequate. Therefore, in order to avoid over-fitting and find an adequate model, the Bayesian information criterion incorporates a term which punishes models relying on more free parameters.

Fig. 5.12 shows the BIC values obtained for different models for a varying number of components. The models EEV, VEV and VVV with six components attain a maximal BIC and they are superior to the other considered models. The current implementation of FBET works with the density model VVV and one component. Thus, the switch
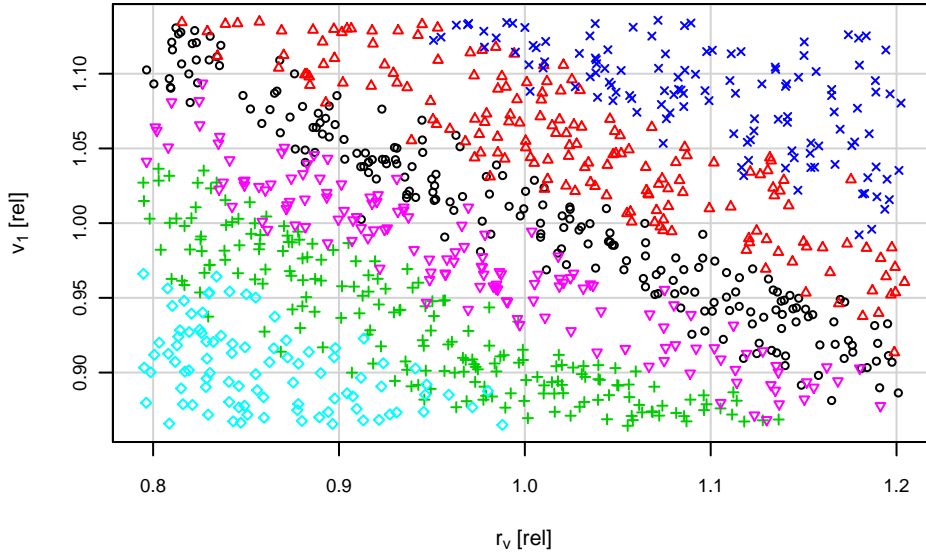
Figure 5.13: The classification of cross section vectors was transferred to the corresponding parameter vectors. The parameter vectors of different classes are nicely separated in the (r_v, v_1)-plane. The values of r_v and v_1 are given relative to the middle values of the allowed parameter ranges.

to Gaussian mixture models with more components would allow to better mimic the unknown true probability density function.

Gaussian mixture models are not only used to model probability density distributions but also for the identification of latent variables, i.e. attributes assigned to sample vectors which were not determined during the sampling process. For instance, one can hardly expect the weight distribution of fish to follow a Gauss distribution because the mean weight of a shark is very different from the mean weight of a goldfish. Nonetheless, the assumption of a Gauss distribution for each 'class' of fish may be fulfilled.

Therefore, one can interpret the result that six components describe the data best as there are six classes of cross section vectors. According to the model VVV with six components which obtained the best BIC, most of the cross section vectors are well separated, i.e. clearly attributable to a component in the Gaussian mixture model. For 90% of the cross section vectors the conditional probability to stem from a certain component is over 99%; for 95% of the vectors it is still more than 96%.

Tracing back the cross section vectors to their respective optical potential parameter vectors and imposing the classification in cross section space onto the parameter vectors, the classes are well separated in respect to the optical potential parameters r_v and v_1 (cf. Fig. 5.13)

Marginal probability density within a class. In the present situation where a uniform distribution is chosen in parameter space, the features of the distribution in cross section space are only defined by the functional relationship from parameter space to cross section space given by the physical model. Consequently, the estimation of the probability density reveals properties of the mapping. More specific, it is connected to values of the Gram determinant which itself can be regarded as a modulus of the derivation generalized to arbitrary dimensional vector spaces.

The preceding analyses demonstrated the mapping to be of a complex structure. Nonetheless, a cluster analysis detected clearly six different clusters with a low probability for misclassification. This means there are distinct domains in parameter space where the mapping strongly contracts volumes compared to the transient regions in between these domains.

The question emerges whether the mapping is of a less complicated structure within these classes. Various density estimations for projection planes indicated that this is not the case. For example, Fig. 5.14 shows the projection plan with highest variance for the class associated with parameter sets in the bottom left of Fig. 5.13.
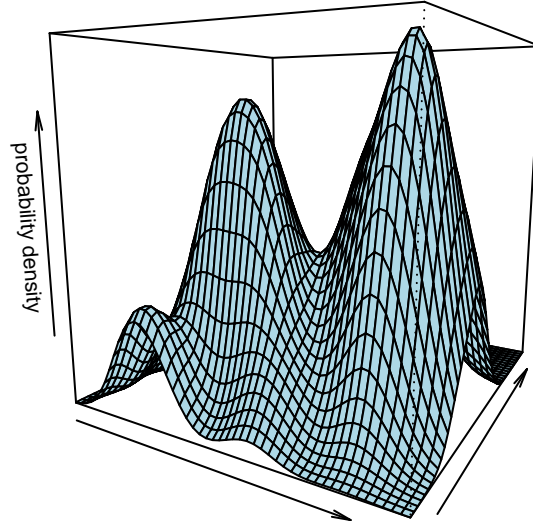
Figure 5.14: Marginal density distribution for the class associated with the parameters on the bottom left of Fig. 5.13 for the projection plane with highest variance.

Intermediate summary. The previous analyses allow two important conclusions. Firstly, considering the total cross section, the physical model is, in good approximation, a mapping to a 16-dimensional subspace in cross section space. The energy grid seems – except for the outliers – sufficiently dense to have smooth transitions from one grid point to another. Thus, a linear combination of 16 basis functions is appropriate to describe each of the 1000 cross section curves within the sample sufficiently well. These are shown in Fig. 5.6.

Secondly, the density estimation and the cluster analysis proved the true probability density to deviate substantially from the Gauss distribution – a sum of six Gauss distributions resembles the real situation more accurately. However, the investigation of marginal density distributions within different classes indicated features which even cannot be aptly described by a mix of Gauss distribution (Fig. 5.14). This does not necessarily mean the model manifold in cross section space is left. In the first place it simply means that the Gauss distribution is not in exact correspondence to the uniform distribution assumed for the optical potential parameters.

These two findings point at ways for future improvements: the use of dimension reduction techniques to extract only the significant information and thereby allowing to handle the information of numerous sources more efficiently; and the incorporation of more sophisticated probability density models, i.e. Gaussian mixture models, which better resemble the true density in order to better follow the systematics of the model and the uniform distribution in parameter space. Nonetheless, in this work the Full Bayesian Evaluation Technique is exactly implemented as described in the previous chapters.

General distribution features. In the following analysis, a more specific point of view is taken. Fig. 5.16 visualizes the probabilities for points in the E-CS diagram to be reached by cross section curves. Increasing intensity of the blue colour denotes areas with increased number of sample vectors in vicinity. The cross section values show a big spread of several Barn in energy ranges where the curves tend to decrease. Additionally, a high spread can be observed for energies above 100 MeV. The very blue areas are caused by the crossing of decreasing curves which have not yet returned to a positive slope and curves which have already performed that transition. Notably, some of the curves seem to fork between 70 MeV and 80 MeV.

The probability density functions for cross section values at some incident energies are shown in Fig. 5.17. These correspond to vertical slices in Fig. 5.16. Even for incident energies associated with very blue colours in Fig. 5.16, denoting a high concentration of cross section curves, the standard deviation is still between 10% and 20%.

The observed fork in Fig. 5.16 at around 70 MeV appears as a small notch in the bottom right diagram of Fig. 5.17. As can be seen, the assumed Gauss distribution, denoted by the green curve, is in spite of the notch still a fairly good approximation for the true probability density function.

Due to the high spread of cross section values over the whole range of incident energies compared to typical experimental uncertainties of some percent, it can be expected that the model merely imposes its systematics on the experimental data. This means if only a single cross section value for a particular energy is included during a Bayesian update step, the resulting prediction for the cross section value at the energy of the experiment is primarily given by the estimate of the experiment. Furthermore, the resulting uncertainty is also mainly determined by the experiment, consider Eq. (2.25). However, the situation changes if experimental points for several energies are included. As was shown, the model – in good approximation – only allows curves which are a linear combination of 16 basis functions. This can be seen by considering the standard deviations listed in Fig. 5.6. The 16th basis function (eigenvector) is associated with a standard deviation of 0.039 barn. Together with the fact that this basis function has a maximal value of 0.2 barn around 0.5 MeV, this leads to the conclusion that the deviation of the cross section curve caused by the coefficient of this base function is not more than 10 millibarn. Since the cross section takes values in the magnitude of 10 barn at 0.5 MeV, the relative standard deviation associated with the 16th basis function is of the magnitude of 0.1%. Thus, if the experimental data points are not representable by the set of basis functions, the predicted curve after the Bayesian update step will be – in good approximation – a linear combination of them again.

Because the regarded cross section space is 75-dimensional, the approximate restriction to 16 dimensions indicates that there are many directions in which no flexibility is permitted. This in turn implies high correlations which show up in the correlation matrix (Fig. 5.15). Notably, all the correlations are positive. Above an incident energy of 100 MeV all cross sections are highly correlated with each other.

In addition, some stripes of looser correlations occur in the graph, correlations attenuate for increasing energy differences and then rise again. They regain their highest values for energy differences between 20 to 30 MeV. Expectedly, cross sections for energies up

to 60 MeV are only weakly correlated with cross sections above 100 MeV, which can be seen by the brighter areas in the top left and bottom right corner of Fig. 5.15.

The high correlations even for energy differences of more than 40 MeV indicate that despite the true probability distribution is not Gaussian, a Gauss distribution is still able to capture a lot of the systematics of the model.
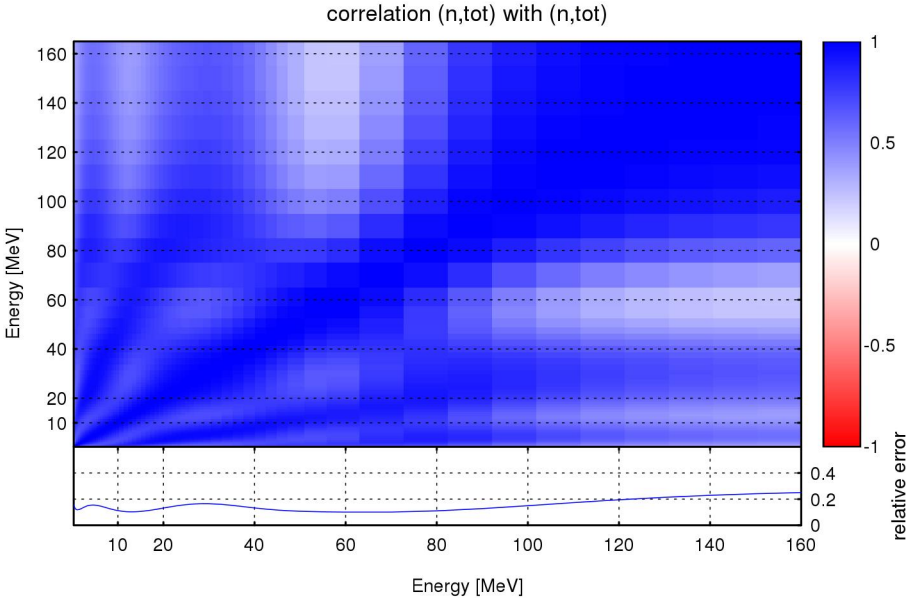

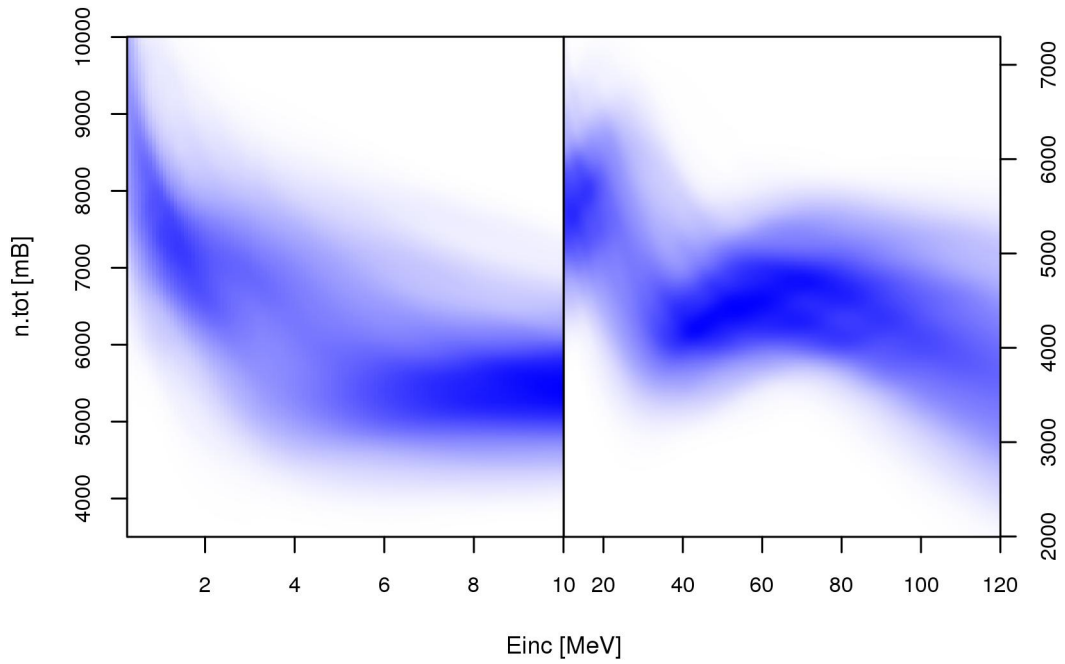Figure 5.15: Correlations between different incident energies within the total cross section channel.

Figure 5.16: Probability density that a model cross section curve for (n,tot) crosses some point in the (E,CS) plane. Bluer points are reached by more cross section curves than whiter ones.
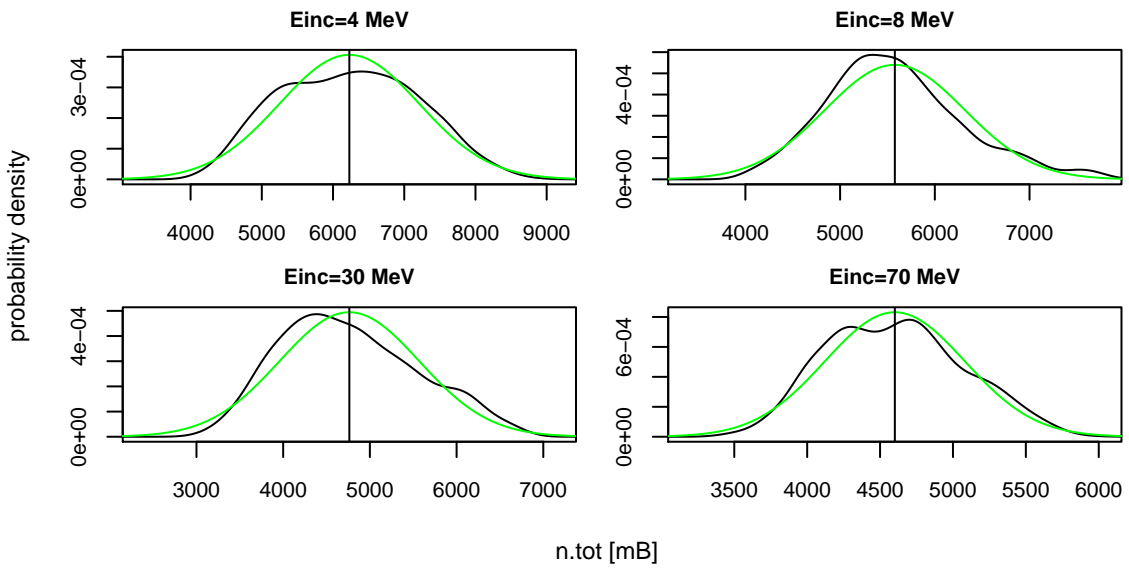
Figure 5.17: These probability density functions correspond to vertical slices in Fig. 5.16

5.2.2 Analyzing (n,el)

The primary goal of this master thesis is the inclusion of differential cross section data into the FBET. Especially, the angle differential elastic cross section should be included in this work. Therefore, differential elastic cross section data are considered. Differential cross sections are expressed for every of the 75 incident energies in terms of Legendre coefficients up to the order of 60. Hence, a sample cross section vector contains $61 \times 75 = 4575$ elements associated with the elastic channel. The corresponding block in the covariance matrix consists of $4575^2 = 20930625$ elements which amounts to approximately 170 MByte. Despite this size, the calculation of the Bayesian update was not a problem. One update step took a few minutes on an Intel CPU with 3 GHz and 4 Gbyte of main memory. Nonetheless, in the future differential data for other channels may be included in the prior and as the number of elements in the covariance matrix scales with the square of elements in the sample cross section vectors, computability and storage needs have to be taken into account. Therefore, the first analysis is going to focus on the issue whether the same dimension reduction technique as for the total cross section channel is applicable to the elastic one.

The second analyses tackles the issue to what extent the assumed Gauss distribution is able to capture the systematics of the model for differential elastic data. Thus, the correlations within the elastic channel between different energies and also between different angles are considered. Additionally, the magnitude of correlations between differential elastic cross sections and the total cross section is investigated.

Dimensionality. A vector with differential angular cross section data can be considered as the discretised image of a function in two variables, i.e. order of Legendre coefficients and the incident energy. The function for the mean sample vector is shown in Fig. 5.18. Assuming that the functions represented by the other sample vectors are of a similar structure, one may expect that there exists a suitable representation in which much less than $75 \times 61 = 4575$ (number of energies and number of Legendre coefficients per energy) values are needed for their characterisation. Furthermore, these vectors are still the result of a mapping from the 20-dimensional parameter space[2] and consequently they are restricted to a 20-dimensional manifold in this 4575-dimensional cross section space.

The result of the principal component analysis [27] is presented in Fig. 5.19. The variances decrease quickly but not as quickly as for the total cross section. A value of around 10^{-5} already reached at the 30th principal axis in the case of the total cross section is not reached before the 90th principal axis for the differential elastic one, compare to Fig. 5.5. Consequently, a higher number of eigenvectors has to be taken into account to provide a satisfying description of the differential CS data.

In order to determine the required number of eigenvectors, the same procedure as for the total cross section was applied: For every sample vector all coordinates associated with the differential CS were determined by projection to the eigenbasis of the covari-

---

[2]16 varied parameters for the optical potential and 4 varied parameters for the Fermi level density model. It is not certain whether the latter can be neglected for the differential elastic CS.
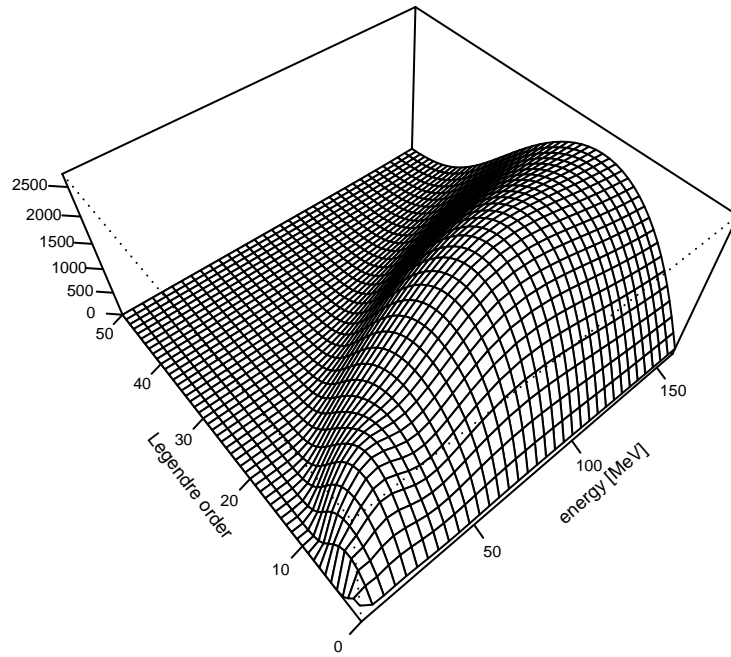
Figure 5.18: The shape of the function for differential elastic cross sections given by the mean sample vector. Angular data are expressed in terms of Legendre coefficients. The function was interpolated for non-integral values of the Legendre order for the sake of visualisation.
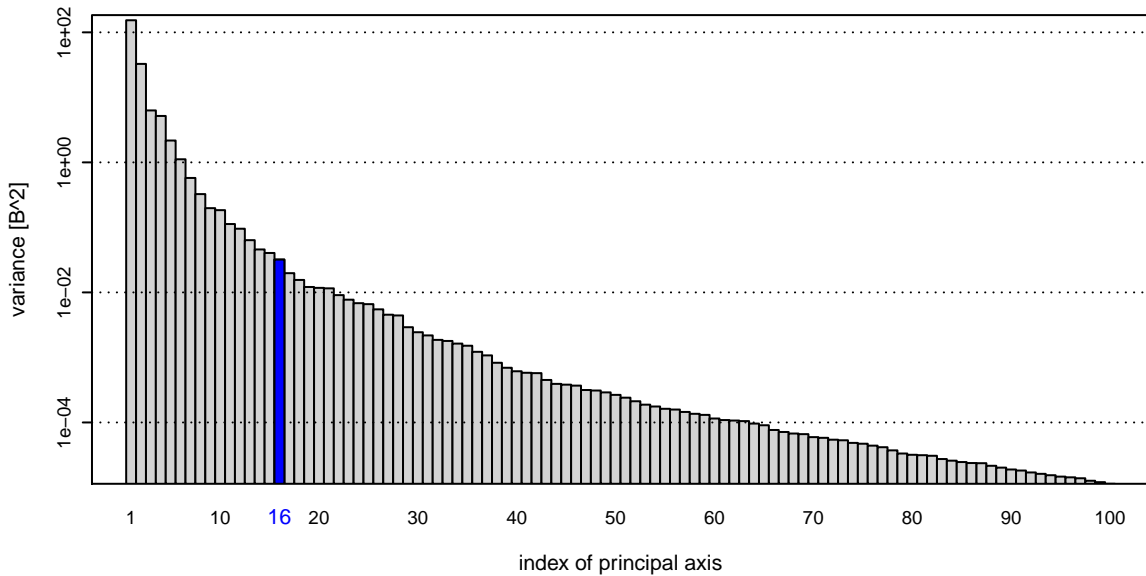

Figure 5.19: Variances along the principal axes of the Gauss distribution associated with the covariance matrix for the differential elastic cross section vectors. As in the case of (n, tot) variances rapidly diminish.

ance matrix. Subsequently, all coordinates except for the leading n were set to zero. Thereafter, all curves were back-projected onto the original basis and the maximal absolute deviation of an approximated curve from the original curve was determined. This procedure was performed for different choices of n. The result is shown in Fig. 5.20.

There are some aspects concerning this diagram which require discussion. The maximal absolute error is not a monotonically decreasing function of the number of utilized eigenvectors. At first glance this seems strange, since the addition of vectors in the linear combination should always reduce the error. However, if the coefficients associated to the eigenvectors are determined by projection – which was done here – only the error given by the euclidean distance is to be guaranteed to decrease monotonically. It would be another situation if the values of the coefficients were found by directly minimizing the maximal absolute error. Then additional eigenvectors denote additional degrees of freedom and the maximal error thus must be a monotonically decreasing function of the number of eigenvectors.

Another point is the stagnancy of error reduction between 100 and 260 eigenvectors. Considering Fig. 5.21 which displays the median of the maximal absolute errors of the single curves, one can see that this stagnancy is not occurring there. Thus, the approximation for a lot of curves steadily improves in the range from 100 to 260 eigenvectors, but for some curves eigenvectors are important which get included in the linear combination at a much later point.

Notably, very often the maximal absolute deviation occurs for the zero-th order Legendre coefficient. Since this coefficient denotes an overall shift for the complete differential cross section curve, it is very important to get its error sufficiently small. What sufficiently small means depends on the actual case. Regarding the experimental data which will be presented in the next chapter, an error in the magnitude of 0.1 millibarn appears acceptable. Fig. 5.22 shows the quantiles of the maximal absolute error within the sample for different number of eigenvectors. Using a linear combination of the first 400 eigenvectors, the maximal absolute error is not greater than 0.1 millibarn. Furthermore, for 90% of cross section curves the error is less than 0.05 millibarn.

At first glance the requirement of 400 base functions (eigenvectors) seems quite high to describe each differential cross section curve within the sample well. Though, these basis functions refer to incident energies and Legendre coefficients at once. If one would regard the differential cross section curve for each of the 75 incident energies individually, it would be equivalent to have found a generic function – the same function for every incident energy – that relies on 5 parameters for a satisfactory approximation of the differential CS at all considered incident energies.

However, one small doubt remains about the generality of this result. A number of 16 basis functions was enough in the case of the total cross section to adequately describe 1000 functions represented by vectors within the sample. In contrast to that, 400 base functions are necessary to approximate 1000 differential cross section curves. The ratio for the latter is a lot closer to one. To consider an extreme case as an example, to be able to describe well 1000 functions by 1000 other functions may not be regarded as a big surprise. Still there are reasons to believe that the number of required eigenvectors would not change drastically by considering a sample of 10000 cross section vectors.
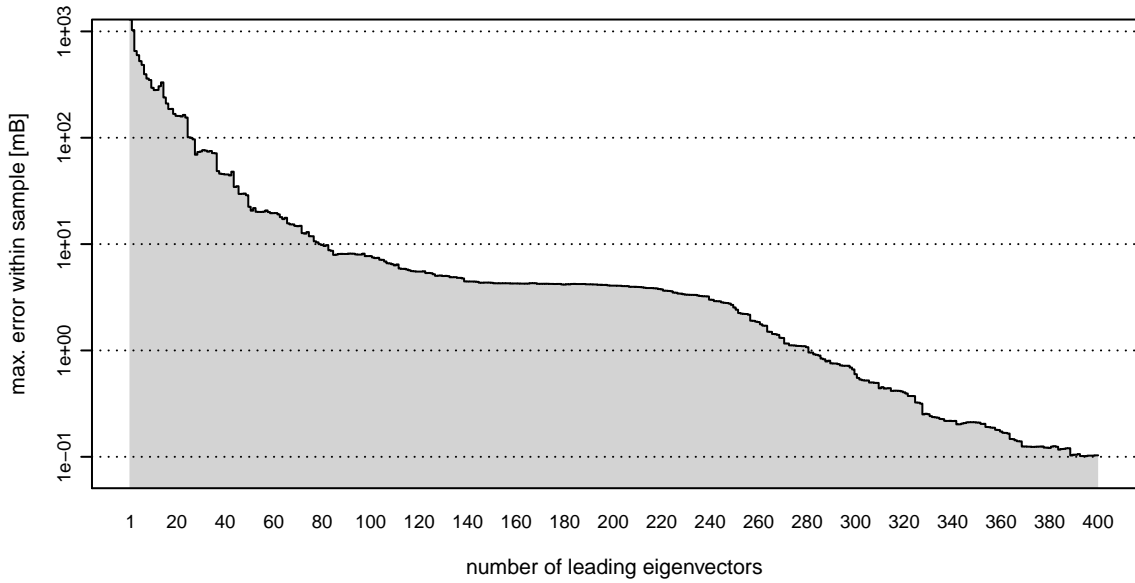
Figure 5.20: The maximal absolute deviation of an approximation from the original curve as a function of the number of considered eigenvectors. The order in which eigenvectors are included in the linear combination is determined by their associated variances. Eigenvectors associated with high variances come first.
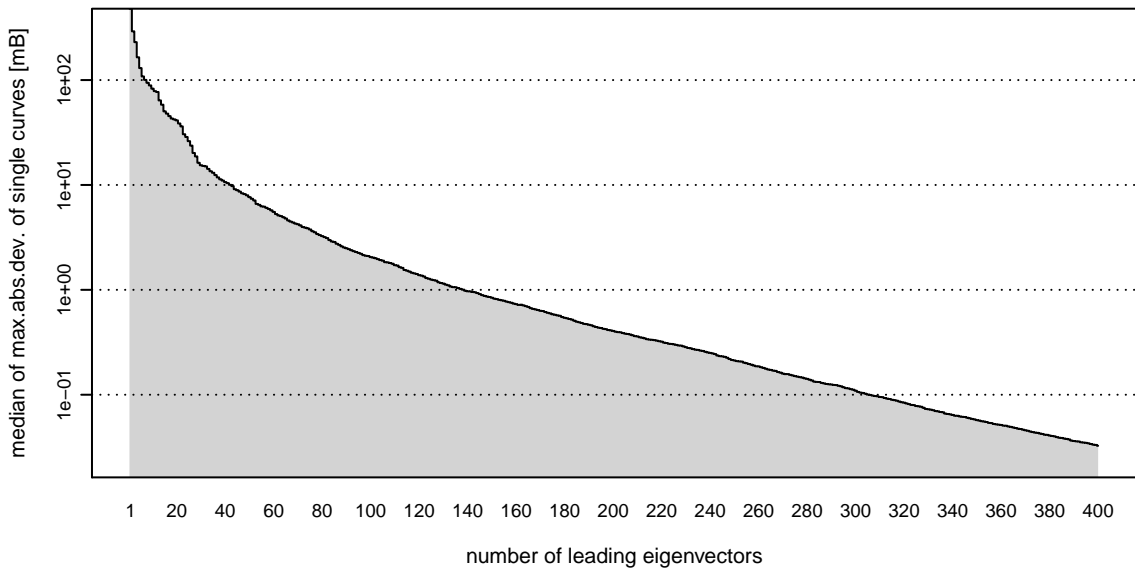

Figure 5.21: The median of the single curve errors as a function of the number of considered eigenvectors. The error for a single curve is given by the maximal absolute deviation of the approximation from the original curve.

64
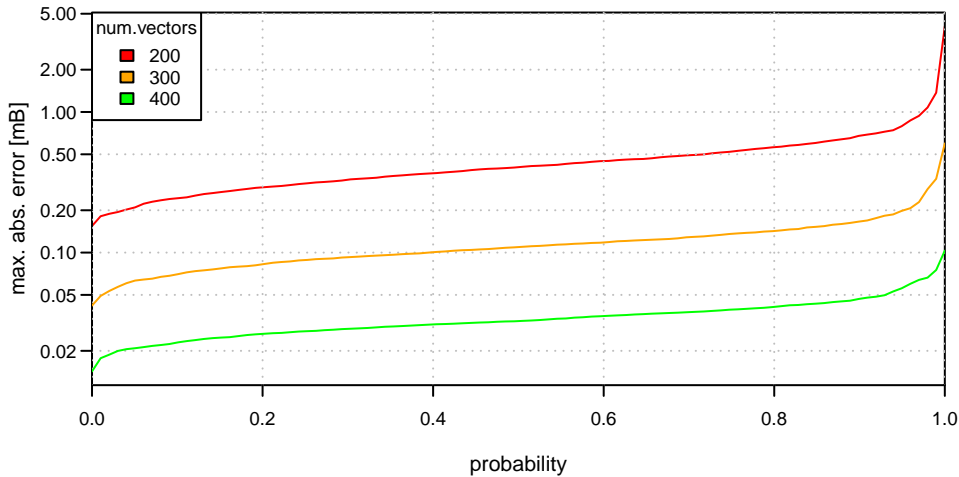
Figure 5.22: The percentage of curves within the sample whose approximation deviates not more than the value stated at the y-axis. For instance, for 80% of the curves the approximations constructed by 400 eigenvectors deviate not more than 0.05 mB.

In order to understand this, it is better to drop the conceptual framework of functions and return to that of vectors again. Assume 1000 points were selected randomly in a restricted domain of a 4575-dimensional vector space. The probability for all of them to lie approximately in a 400-dimensional linear subspace is near to zero. Therefore, having found this to be true for the differential elastic cross section vectors is indeed significant.

Correlations. On the next pages the correlations between different Legendre coefficients are visualized. Fig. 5.23 shows the correlations for Legendre coefficients of different order at the same incident energy. There are strong correlations and they are usually positive. Only coefficients of very high-order are negatively correlated with low-order coefficients. Some diagrams feature a blue square-like structure in the bottom left. Thus, for the leading order Legendre coefficients strong correlations persist over several orders. Then at a particular Legendre order the blue square is continued by a tail-like structure in the diagonal. This means that the range of strong correlations over several orders is abruptly reduced. With increasing incident energy also the size of the blue square in the bottom left increases. Therefore, the relationship between differential Legendre coefficients becomes more and more linear. In consequence also differential cross section values for different angles become linear functions of each other. Explanations for the other diagrams are given in the figure descriptions.
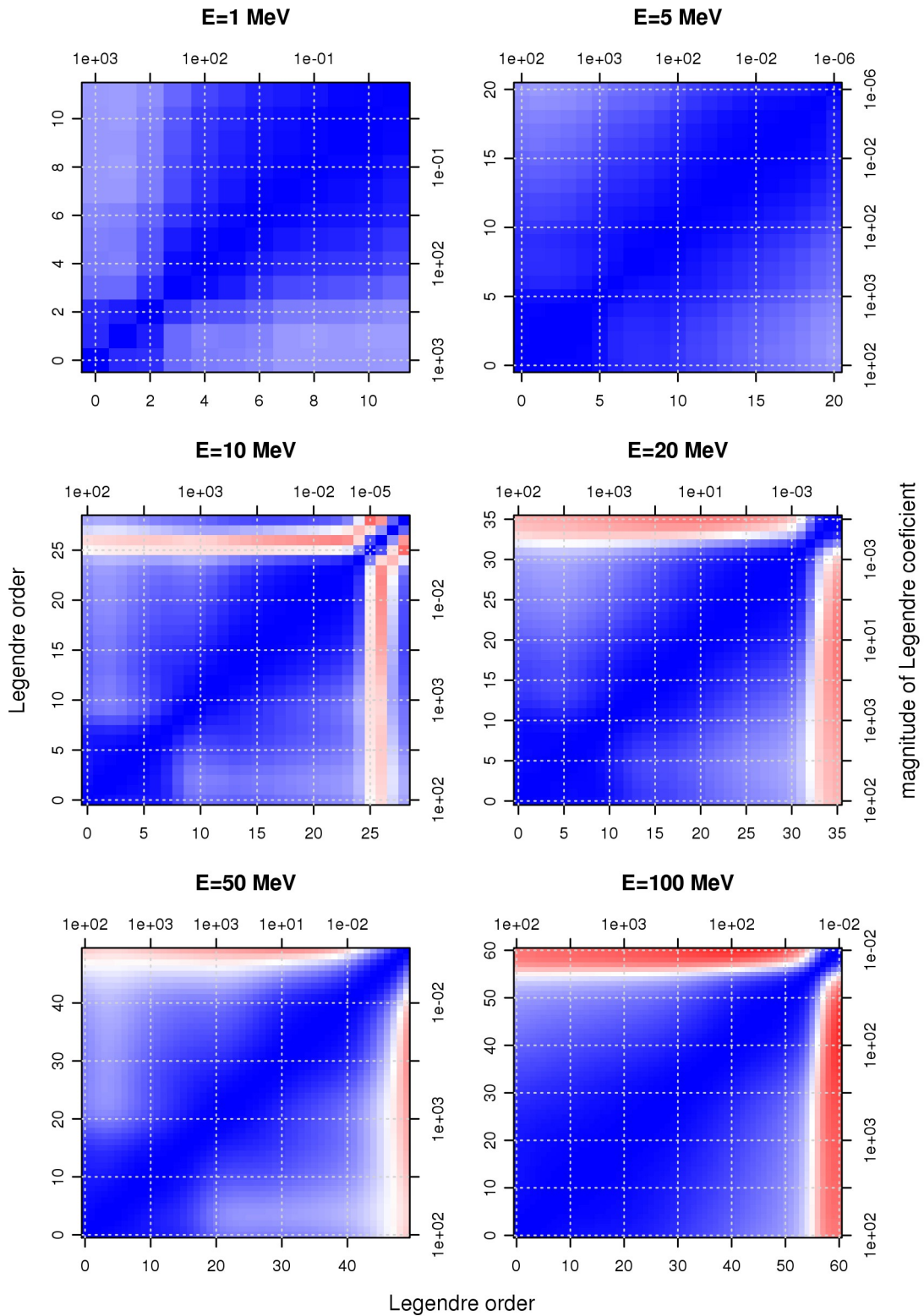
Figure 5.23: Correlations between Legendre coefficients for different energies. The deepest blue in the diagonal indicates a perfect positive correlation; a 'paper white' indicates no correlation; and red negative correlations. The colour code is the same as for Fig. 5.15
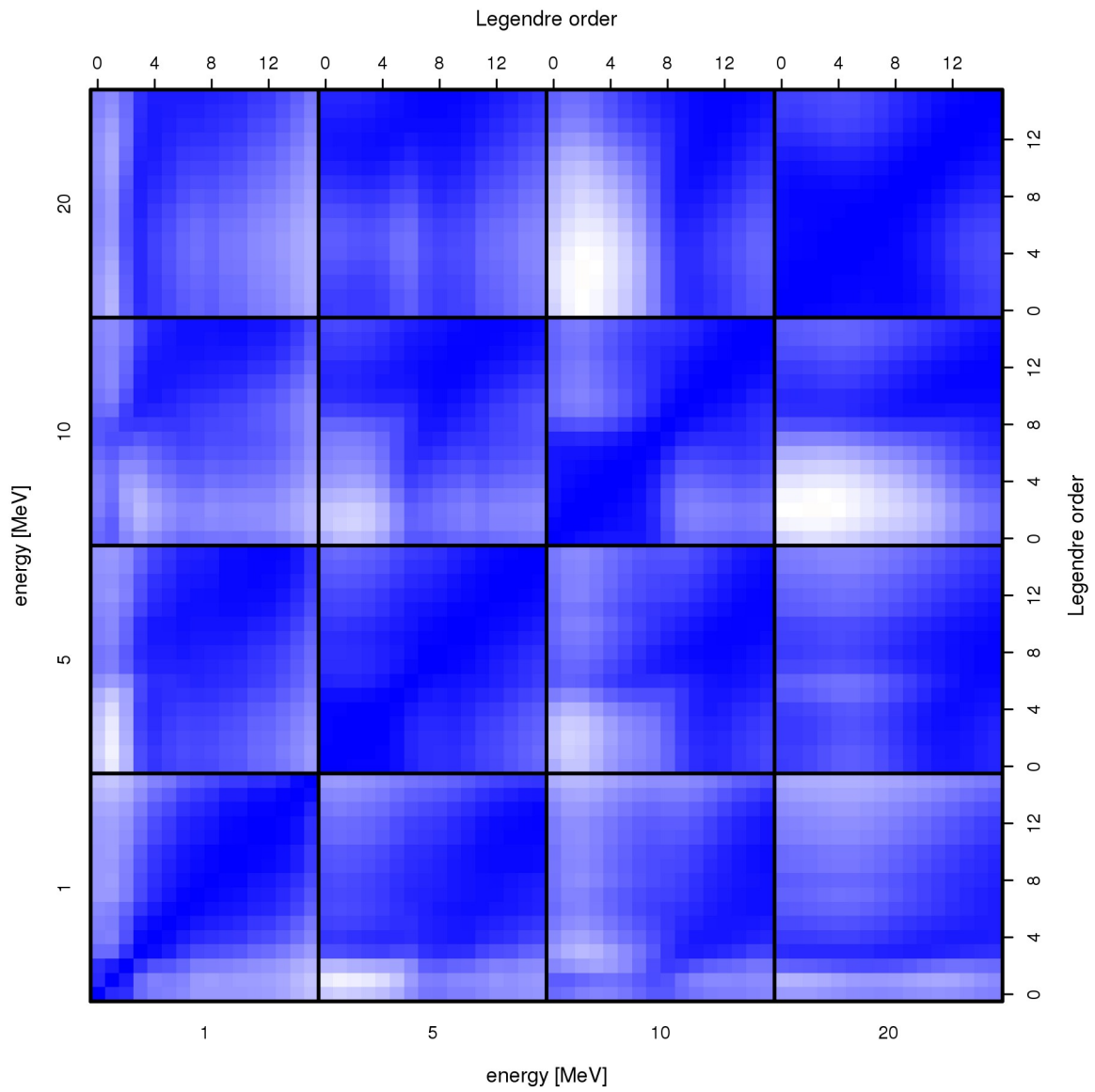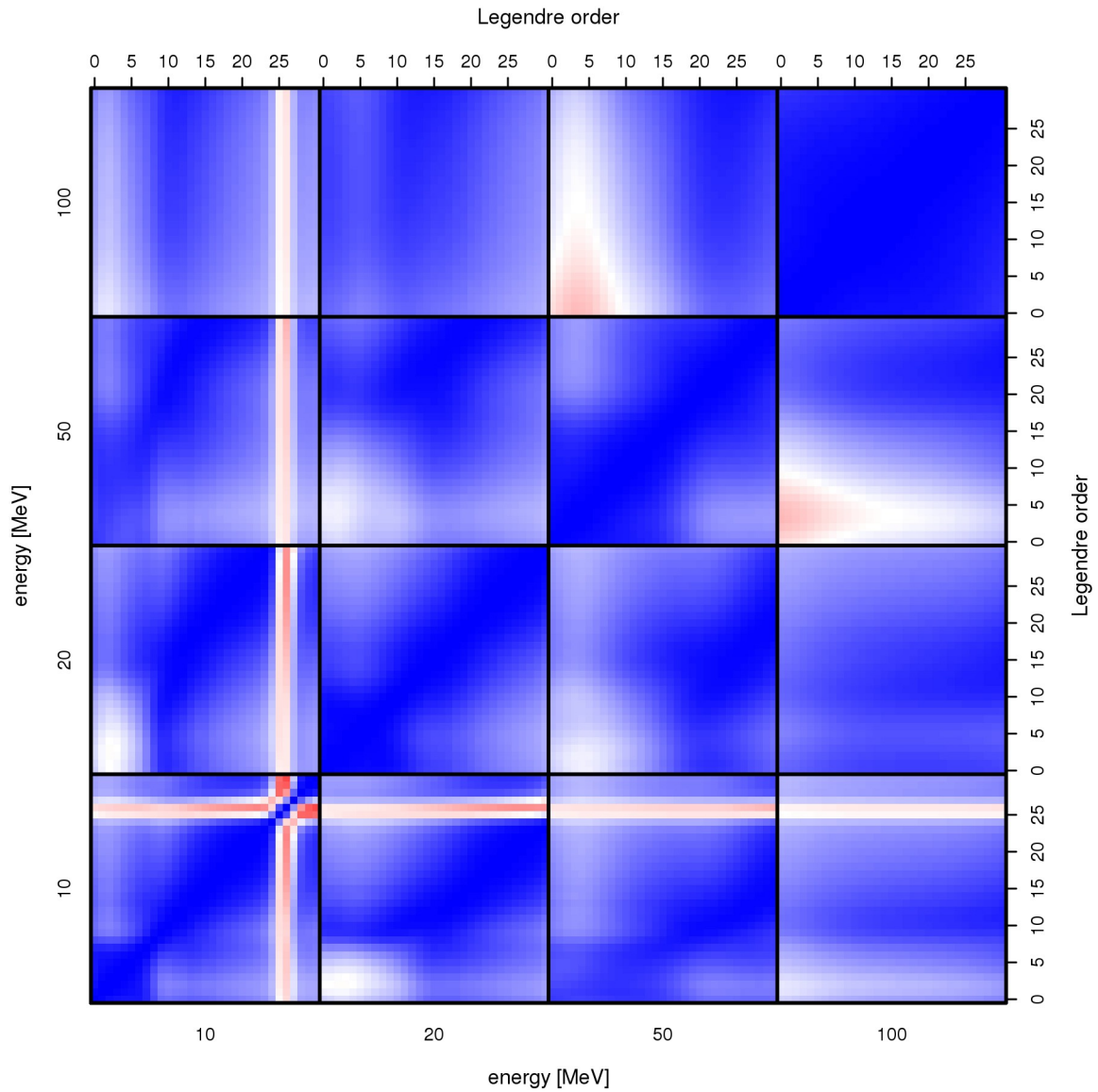
66

Figure 5.24: Correlations between Legendre coefficients up to the order of 15. This diagram includes incident energies where experimental data are usually available. All correlations are positive. The colour code is the same as for Fig. 5.15

Figure 5.25: Correlations between Legendre coefficients up to the order of 30. This diagram includes a broad range of incident energies from 10 MeV to 100 MeV thereby demonstrating how strong the model couples cross section data of different energies together. Notably, correlations still exist between 10 MeV and 100 MeV and even strong ones between 10 MeV and 50 MeV. Low order Legendre coefficients at 50 MeV are weakly negatively correlated to Legendre coefficients at 100 MeV. Hence, if the integral cross section is above the average at 50 MeV, the probability is slightly increased that it is below the average for 100 MeV.
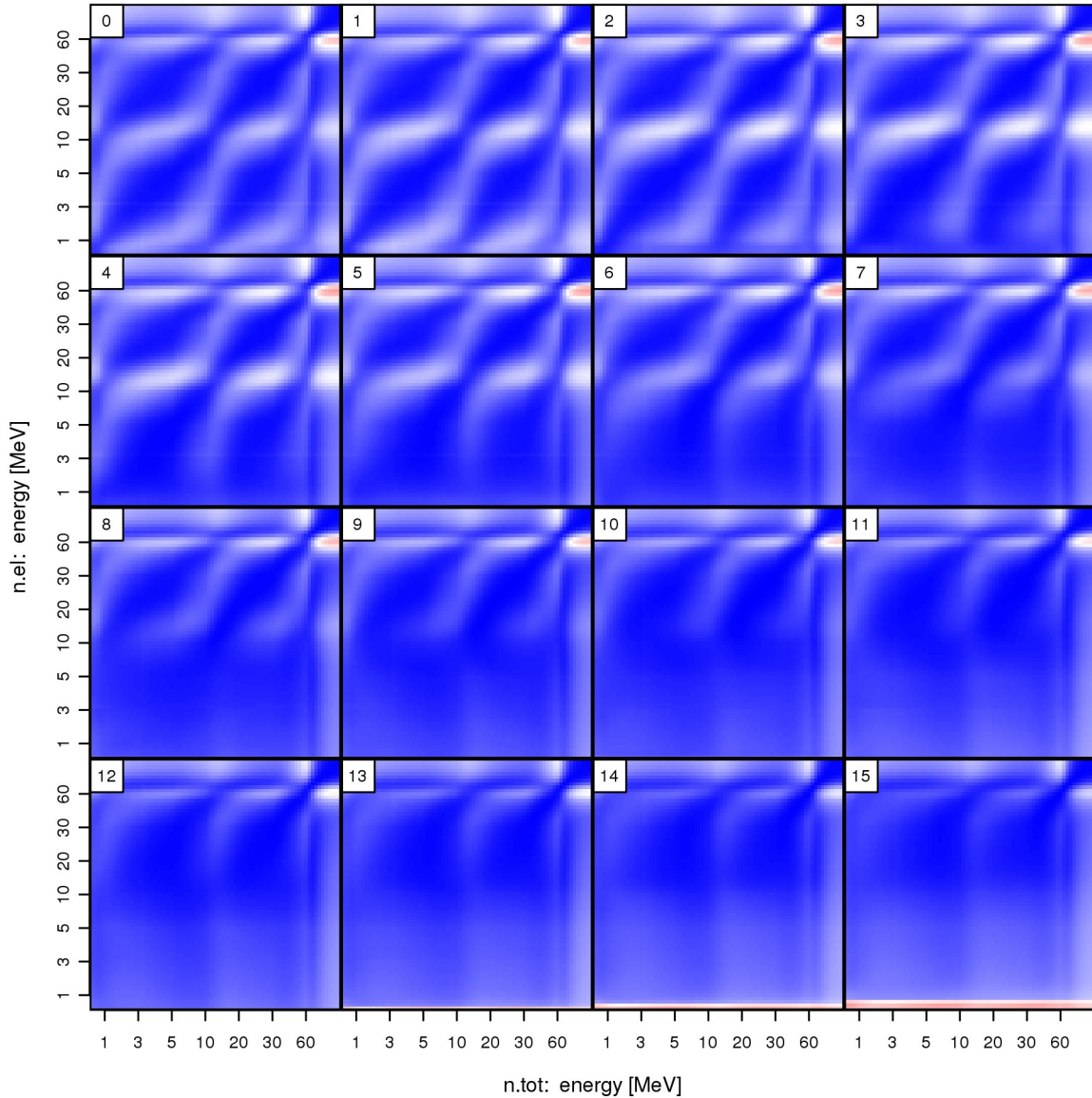
Figure 5.26: Correlations between the total cross section and Legendre coefficients for the elastic channel up to the order of 15. The 0-order Legendre coefficient is strongly correlated to the total CS at the same energy visible by the blue diagonal in the top-left diagram. Thus, there exists in good approximation a linear relationship between the integral elastic and the total CS. Notably, with increasing energy difference correlations are dying away but at some point rise again which can be seen by the white wave-like stripes. Remarkably, the total CS in the range from 2 MeV to 5 MeV is strongly correlated to the integral elastic CS between 20 MeV and 30 MeV. All diagrams contain a little blue square in the top right corner. Therefore, Legendre coefficients for high energies (above 80 MeV) are strongly correlated to the total CS at high energies but only weakly to the total CS at lower energies.

69

5.3 Presentation of the utilized experimental data

This work considers experimental data for the total neutron cross section and the angle differential elastic neutron cross section for tantalum-181. The data was taken from the EXFOR database [1]. The following table summarises the suitable data for which the papers were available:

experimenter	year	E1 [MeV]	E2 [MeV]	num.pts	ref
total neutron cross section data					
Finlay et al.	1993	5.0	600	474	[9]
Foster et al.	1971	2.5	15.0	243	[10]
Islam et al.	1972	1.0	2.0	101	[18]
angle differential elastic neutron data					
Smith	2005	0.3	10.0	714	[36]
Hansen et al.	1985	14.6	14.6	15	[15]
Ferrer et al.	1976	11.0	11.0	26	[8]
Benenson et al.	1973	14.8	14.8	8	[4]
Rosen et al.	1957	14.0	14.0	8	[31]
Remund	1956	3.3	3.3	9	[30]

Table 5.3: Suitable experimental data for the total neutron CS and for the differential elastic neutron CS available in the EXFOR database [1].

For the total CS, the data from R. W. Finlay et al. [9] and D. G. Foster et al. [10] were selected because they offer measurements over a broad range of energies. Furthermore, they state coherent estimates in the common energy range from 4.5 MeV to 15 MeV. Since Finlay et al. provide total CS up to 600 MeV – an energy domain where the optical model is certainly inadequate – only the subset of their data featuring incident energies below 160 MeV was included in the Bayesian update procedure. An overall normalization error of 0.5% for the Finlay data and 3% for the Foster data was assumed. The experimental data and the prior is visualized in Fig. 5.28.

Concerning the angle differential elastic data, only estimates from A. B. Smith [36] from 5 MeV to 10 MeV were taken into account. An amount of 480 values had been measured automatedly on a regular angle-energy grid. At each energy, estimates were obtained for 40 angles between 15° and 160°. Here, an overall normalization error of 5% was assumed. The differential elastic data is presented in Fig. 5.29.

The paper of Smith [36] explicitly denotes that due to the limited energy resolution of 360 keV for the scattered neutrons, obtained differential 'elastic' cross section curves contain inelastic contributions up to and including scattering at the fifth level at about
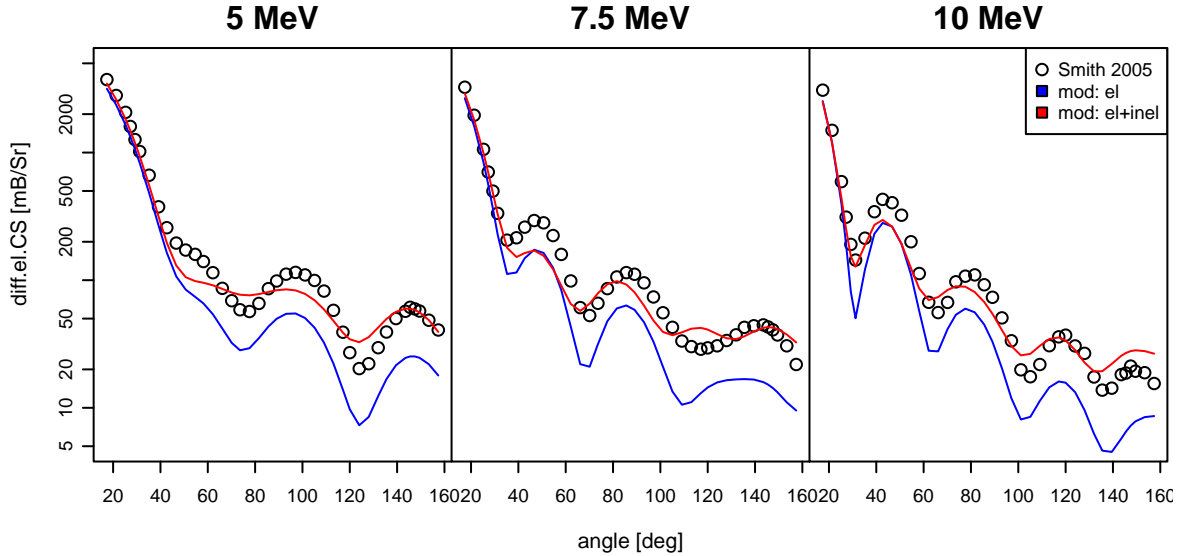
Figure 5.27: The blue curve corresponds to the model sample vector which approximates the experimental data best if only elastic scattering is considered. The red curve corresponds to another sample vector which approximates the experiment best if also inelastic contributions are included. The measure for goodness of fit for a sample curve is given by the maximal relative deviation from the experimental data.

338 keV. In order to include these contributions into the prior, Legendre coefficients α_{ij} of each order i for different levels j were summed up, so $\alpha_{\mathrm{inel},i} = \sum_{j=1}^{5} \alpha_{ij}$. The resulting Legendre coefficients were treated as belonging to another type of differential cross section channel. Therefore, the covariance matrix was extended by another block to include associated covariance elements. The mapping from model space to the experiment space, where the latter includes these inelastic contributions, was performed by calculating both the sensitivity matrix S_{el} to map the elastic channel of the model to the experiment and S_{inel} to map the inelastic channel. The overall sensitivity matrix was given by $S = S_{\mathrm{el}} + S_{\mathrm{inel}}$.

Transferring the samples calculated by TALYS [23] to the experiment space and comparing them to the measured cross sections of Smith indicated the inelastic contributions to be indeed important to describe curve features at higher angles adequately. Fig. 5.27 illustrates the sample vectors whose maximal relative deviation from the experimental data is the smallest of all sample vectors. The red curves of all three sub-figures are associated with the same sample vector which is closest to the experimental data if mapped by S; the blue curves are associated with the sample vectors which is closest to the experimental data if data is mapped by S_{el}.

Treating the experiment as only the elastic cross section had been measured, one would use S_{el} to map from model space to experiment space. However, as can be seen by the blue curve, then the model systematically underestimates the cross sections. Additionally, this deviation cannot be explained by a normalization error of the experiment.
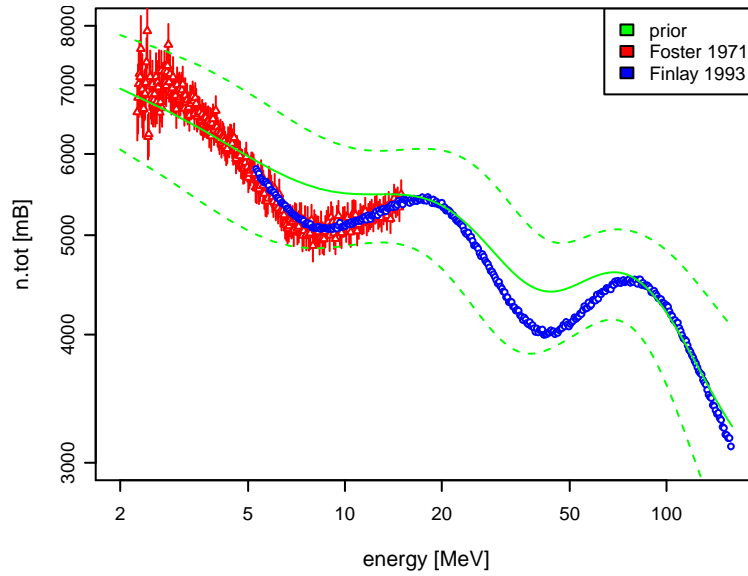
Figure 5.28: The prior in comparison to the experimental data for the total cross section. The 1-σ-confidence interval is marked by a green dashed line for the prior and by vertical bars for the experiments. Since the total error in the Finlay data is below 1%, error bars are not discernible from the points.
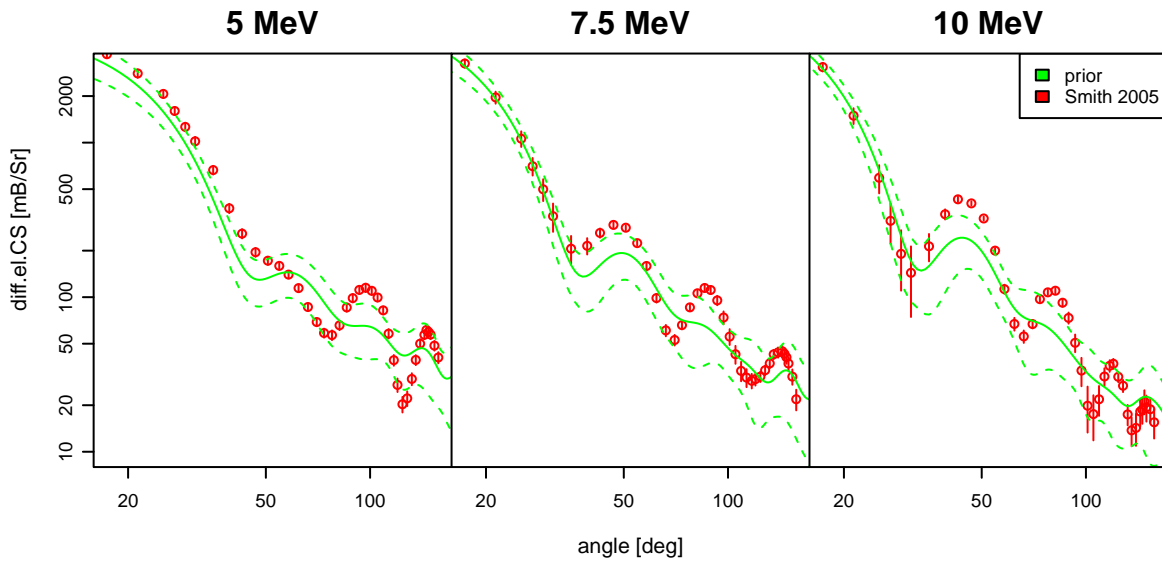

Figure 5.29: The prior in comparison to the experimental data for the differential 'elastic' cross section. The experimental data is contaminated by inelastic contributions from the first five levels at 6, 136, 159, 302 and 338 keV. These contributions are also included in the prior curve.

Applying a factor of about 1.6 to the model curve minimizes the maximal relative deviation from the experiment. However, a normalization error of 60% seems very unlikely for a modern experiment; and even then the experimental curve is still slightly above the model curve. Consequently, the difference between elastic model curve (blue) and experiment is due to also measured inelastic contributions and they have to be taken into account in the Bayesian update procedure.

5.4 Performing the Bayesian update procedure

The correlation diagrams (Fig. 5.23 - 5.25) demonstrated very high correlations between Legendre coefficients of different orders and different energies. Furthermore, strong correlations between the Legendre coefficients and the total CS exist at all energies (see Fig. 5.26). Thus, updating with total CS data should lead to significant changes for the different elastic channel; reversely, updating with differential elastic data should lead to significant changes in the total CS channel. The quality of these modifications due to updating is studied in the following sections.

5.4.1 Updating with total CS data

The update with total cross section data from Finlay et al. [9] and Foster et al. [10] leads to the evaluated curves presented in Fig. 5.30 for the total CS and Fig. 5.31 for the differential elastic CS.

The evaluated curve for the total CS resembles well the shape of the experimental data but is slightly displaced. The associated 1-σ-confidence interval is very close to 0.5% in the whole range and the bulk of experimental points is not within its boundaries. The magnitude of the evaluated uncertainty results from the smallest normalization error of 0.5% given by the Finlay data. The strong correlations between different energies and an overwhelming amount of data points with uncertainties of 1% or less cause the statistical errors to disappear. The reason for this behaviour of the Bayesian update procedure will be elaborated on in section 5.7.

Interestingly, the prior curve for the total CS (Fig. 5.28) tends to overestimate the total CS, whereas the posterior curve underestimates it over a broad energy range. At first glance, the location of the prior suggests that the normalization error of the experiment should lead to an evaluated curve somewhere between the prior curve and the experimental data – hence it should lie above and not below experimental data. The Bayesian update procedure can also be regarded as generalised least-squares fit to experimental data, in which the data of the model is considered as stemming from an artificial experiment. In this framework, the observed effect is known as 'Peelle's Pertinent Puzzle' [28] – the occurrence of evaluated estimates below the estimates of the experiments which entered the evaluation.

This seemingly paradox behaviour is understandable by recalling the eigenvectors of the prior covariance matrix (cf. Fig. 5.6). The model restricts the shape of the evaluated curves to be representable by a linear combination of these eigenvectors. The prior curve

can be obtained by multiplying every eigenvector by the associated mean (also given in Fig. 5.28) and summing these products up. By projecting the experimental curve onto the directions given by these eigenvectors, one obtains the 'mean' values the experiment would assign to these directions. Furthermore, by projecting the experimental covariance matrix onto the same eigenvectors, one obtains the variances along the eigenvectors the experimental data is associated with. If for the representation of the shape of the experimental curve, eigenvectors are important which are associated with very small variances in the prior, and if additionally, the eigenvectors which caused the prior curve to lie above the experimental data are very unlikely by the latter, the observed posterior can occur. The given argument – reduced to its very core – says that the prior regards the experimental data as rather unlikely and the experimental data reversely renders the prior curve very unlikely. The resulting behaviour of the Bayesian update procedure in this case has already been demonstrated qualitatively in the right part of Fig. 3.2.

Although no differential elastic data entered the evaluation, the evaluated differential elastic CS curve at high angles is closer to the experimental data than the prior curve (cf. 5.31 and 5.29) and the experimental data lies within the 1-σ-confidence bands.

At low angles, the evaluated curve is significantly shifted downwards compared to the experimental curve and evaluated confidence bands render the experimental data as highly unlikely.
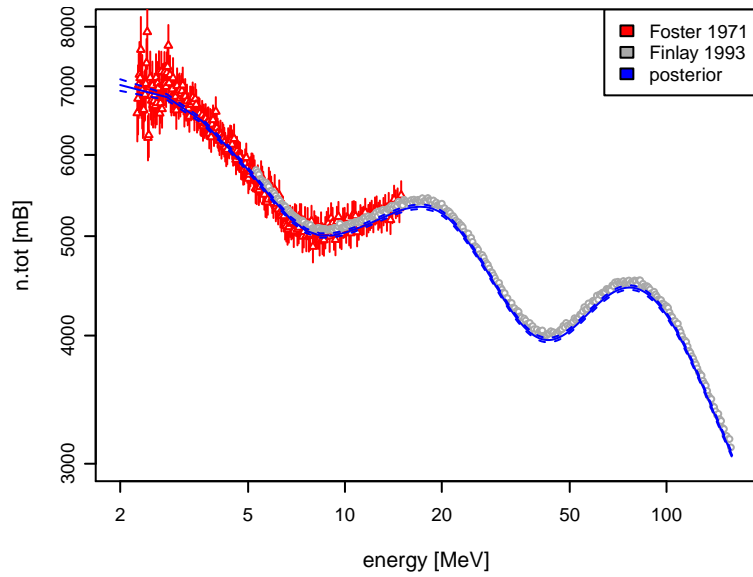
Figure 5.30: The total CS experimental data in comparison to the posterior. Only total CS experimental data entered into the evaluation. The evaluated uncertainty is about 0.5%, which is also the normalization error of the Finlay data. The posterior systematically lies below the experimental data and the latter is not within the 1-σ-confidence interval of the posterior.
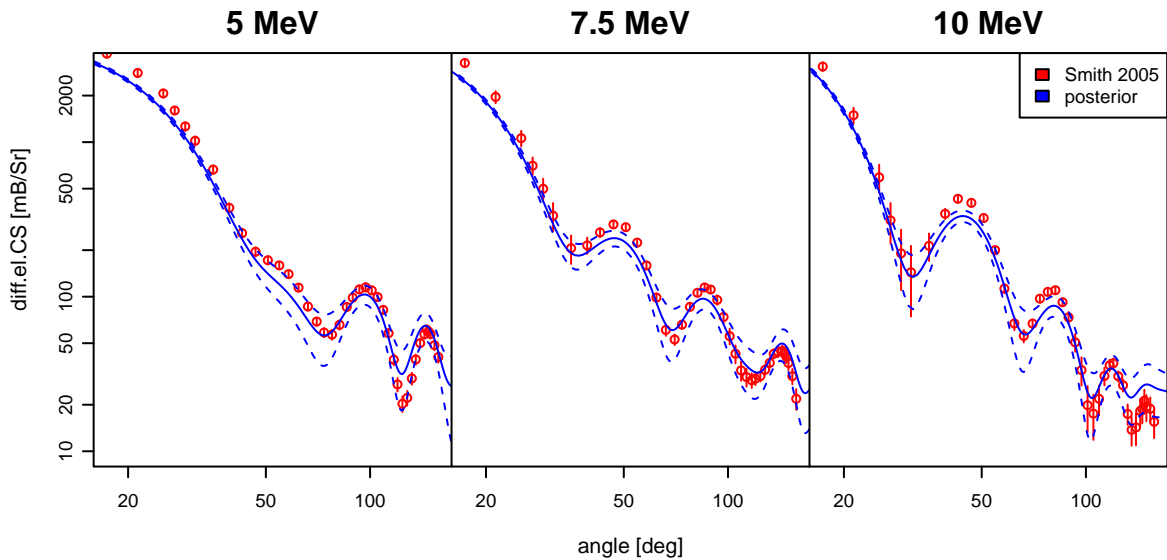

Figure 5.31: Although only total CS data entered into the evaluation, the posterior resembles the differential elastic data far better than the prior. However, the differential CS in the forward-scattering domain are systematically to low. Thus, the model favours a higher portion of the reaction CS on the total CS than the Smith data ascertain.

5.4.2 Updating with differential elastic CS data

The following pages feature evaluated curves in comparison with the experiments where only the differential elastic data from Smith [36] entered into the evaluation. Fig. 5.32 and Fig. 5.33 show the results when all the data from 4.5 MeV to 10 MeV is entered into the evaluation with a normalization error of 5%. The evaluated curves from Fig. 5.34 and Fig. 5.35 arise if data for angles greater than 50° is ignored. Finally, using the data for all angles but without regarding a normalization error results in the evaluated curves of Fig. 5.36 and Fig. 5.37.

In the update scenarios in which a normalization error (5%) was considered, the evaluated uncertainties are in the magnitude of the normalization error (2% - 9%), but the evaluated curves do not match the experimental data in the forward-scattering region. This is due to the fact that the model is biased towards lower cross section values. Ideally, the model should be ignorant of possible CS curves and only impose its systematics. That this is not the case is already visible when considering the prior in Fig. 5.29. Quite some of the experimental data points lie outside the 1-σ-confidence interval. As the posterior covariance matrix is completely determined by the covariance matrices of experiment and prior (cf. Eq. 3.43), the bias could only be removed by adjusting the prior covariance matrix – in which way remains an open question.

Furthermore, the predicted curve for the total CS reacts sensitively to different choices of the normalization error; even differential CS values at large angles which are usually small quantities have a significant influence on the shape of the total CS curve over the whole energy range. It is questionable if the model is really accurate enough to let it allow to modify its prediction for the total CS curve at energies above 100 MeV because some differential data at 5 MeV in the large angle domain had been included.

However, disregarding the too small evaluated uncertainties, the posterior curves generally resemble the experimental data much better than the prior curves.
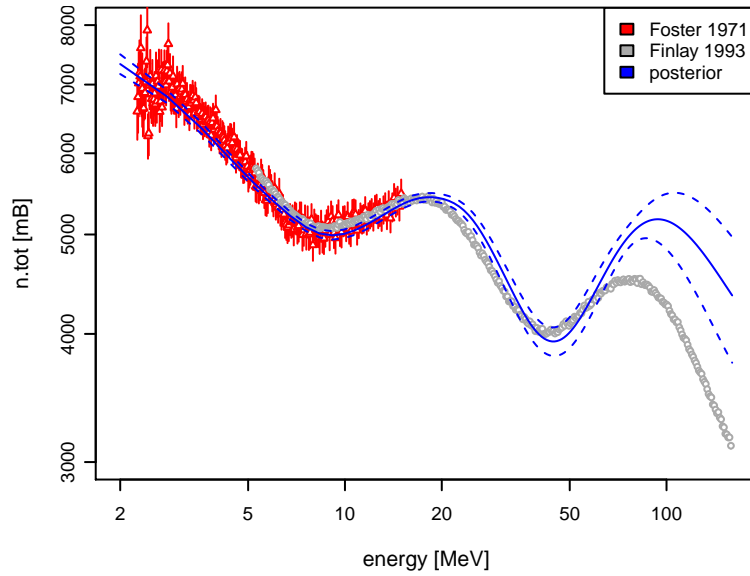
Figure 5.32: The posterior in comparison to total cross section data if only differential elastic experimental data enter into the evaluation. The posterior resembles the experimental data at energies below 50 MeV far better than the prior. Above 100 MeV, it substantially deviates from the experiment and the data of the latter in this energy region is not contained in the 1-σ-confidence interval of the posterior.
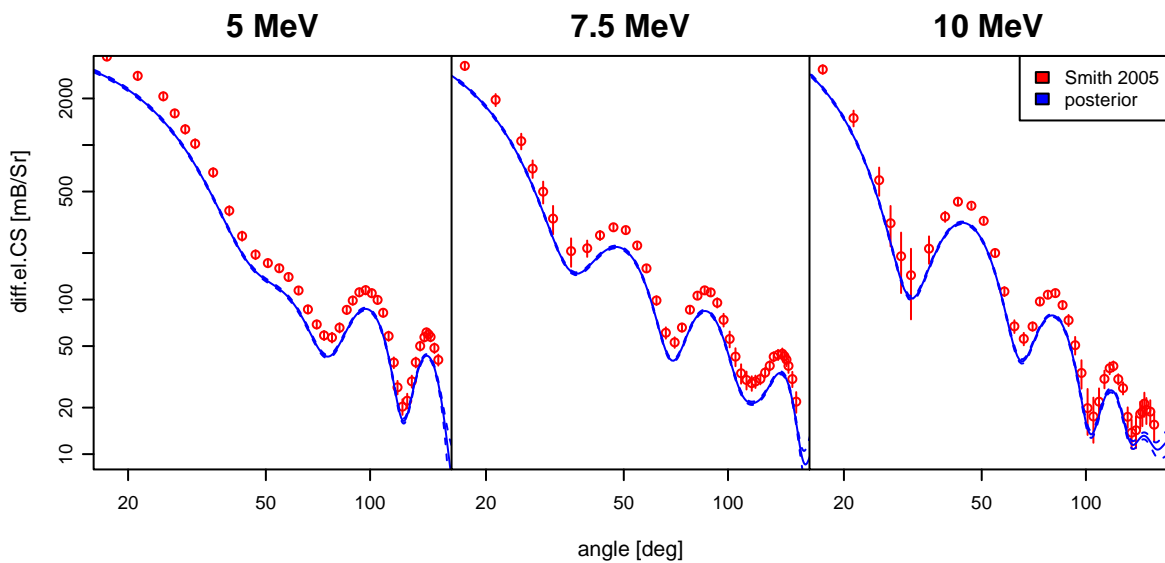

Figure 5.33: The posterior in comparison to differential elastic experimental data if only the latter enter into the evaluation. The posterior approximates the structure of the experimental data well but is shifted. The evaluated uncertainties are between 2% and 9%. Thus, the minimal evaluated error goes below the assumed 5% normalization error of the Smith experiment.
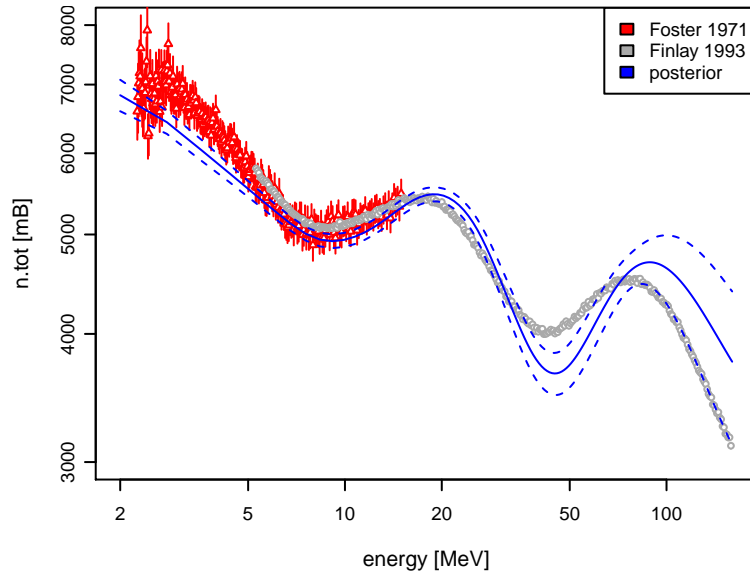
77

Figure 5.34: The posterior in comparison to the experimental data for the total CS if only differential elastic experimental for angles below 50° enter into the evaluation. Comparing with Fig. 5.32, one sees that even small differential CS at angles above 50° have a significant impact on the prediction for the total CS at energies above 100 MeV.
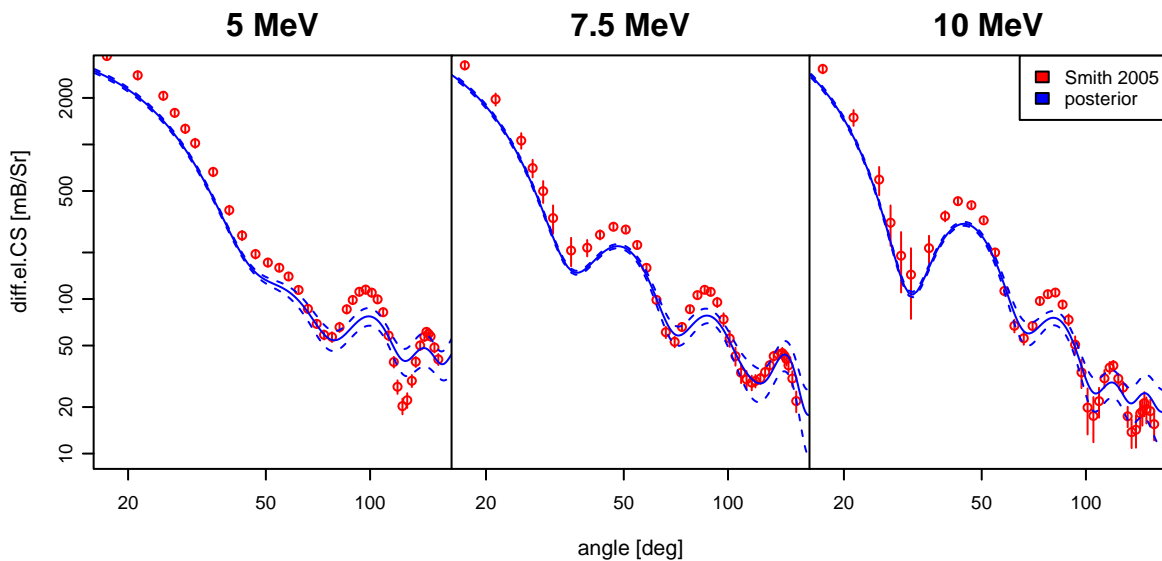

Figure 5.35: The posterior in comparison to the experimental data for the differential elastic CS if only differential elastic experimental for angles below 50° enter into the evaluation. Comparing with Fig. 5.33, visual judgement does not reveal any significant difference in the forward-scattering domain. The posterior resembles the structure of the CS curve at angles greate 50° where no experimental data was included better than the prior. Expectedly, uncertainty bounds are wider than in Fig. 5.33.
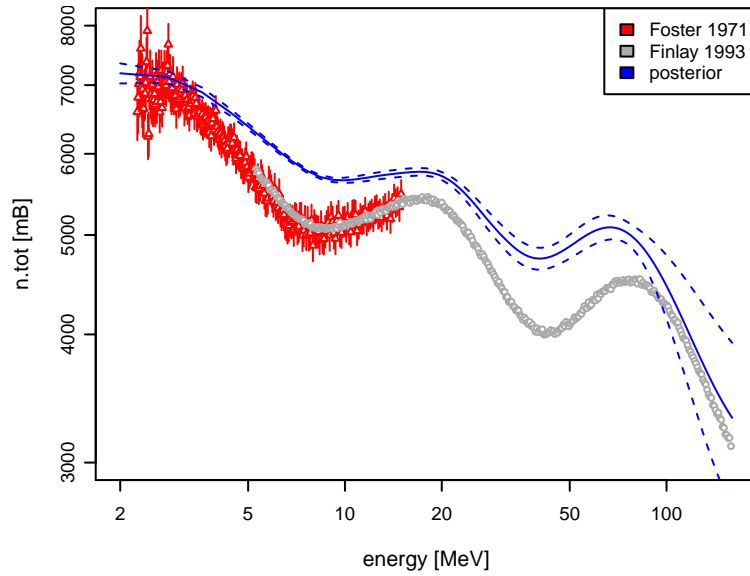
Figure 5.36: The posterior in comparison to the experimental data for the total CS if only differential data enter into the evaluation and no normalization error is assumed. The posterior differs significantly from Fig. 5.32 and Fig. 5.34. The posterior overestimates the CS and evaluated uncertainties do not include the bulk of the experimental data.
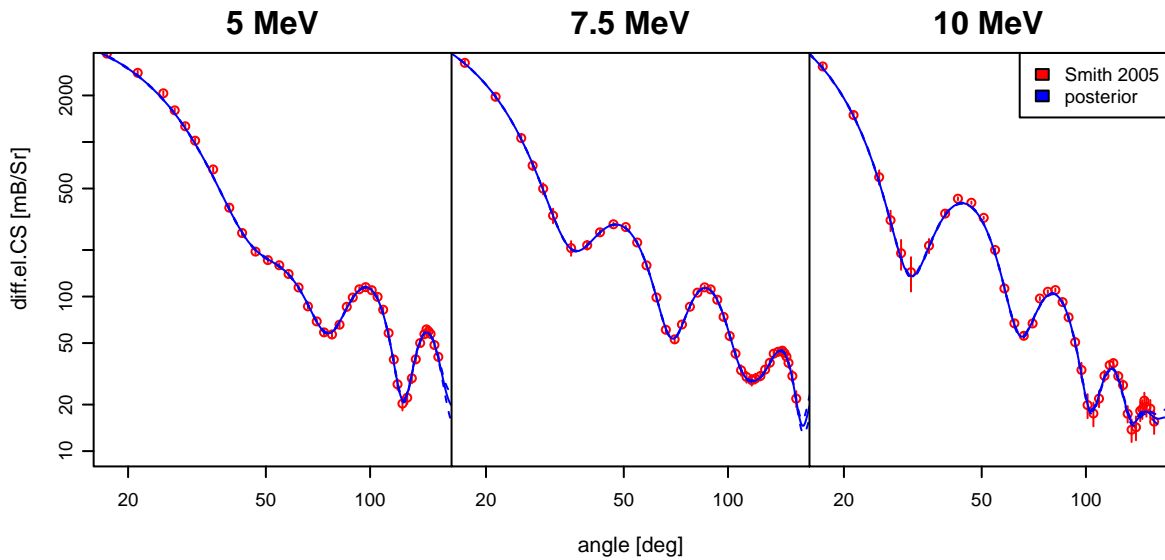

Figure 5.37: The posterior in comparison to the experimental data for the differential CS if only differential data enter into the evaluation and no normalization error is assumed. The posterior perfectly fits the experimental data that entered the evaluation.

5.4.3 Updating with both experimental data types

The inclusion of both total CS and differential elastic experimental data leads to the posterior curves in Fig. 5.41 and Fig. 5.42. The result for the total CS is almost the same as if only total CS data were included (cf. Fig. 5.30). The same is true for the differential elastic CS (cf. Fig. 5.33).

The structure of the posterior covariance matrix has been altered completely by the experimental data. The correlation diagram for the total CS channel (Fig. 5.38) shows that nearly all elements corresponding to energies where experimental data were included are strongly positively correlated with each other.

High correlations over the whole energy range indicate that the model is very certain about the shape of the cross section curve. Thus, updating again the obtained posterior with experimental should only lead to significant changes if the experimental data supports a shape which is regarded as likely by the posterior.

Concerning the differential elastic CS channel, the inclusion of experimental data weakens the correlations and, especially in the case of correlations between the differential elastic CS and the total CS, the covariance matrix loses most of its complicated structural properties.
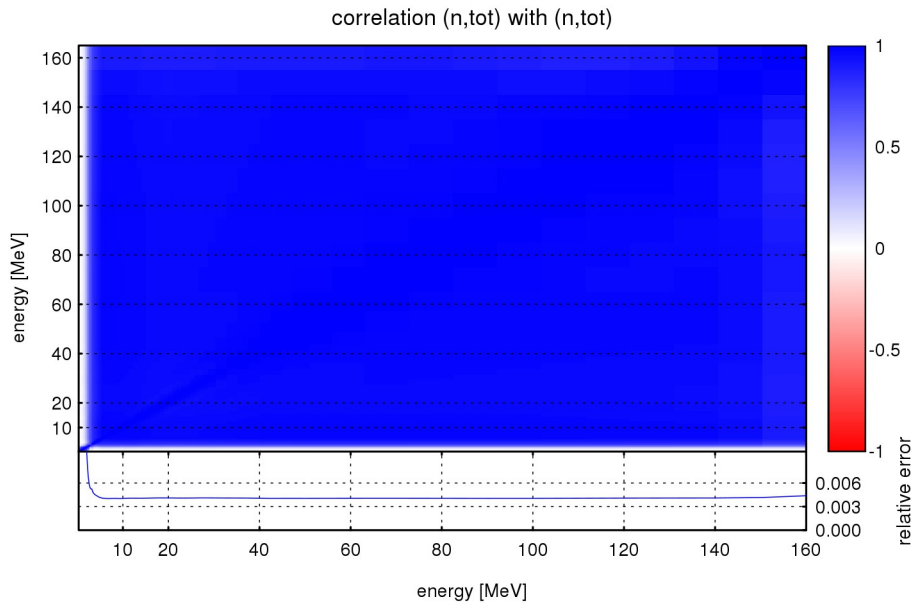

Figure 5.38: Correlations between different incident energies within the total CS channel after a Bayesian update with all experimental data. There is an almost perfect linear dependence between arbitrarily selected energies – one known CS value at any energy almost completely determines the shape of the whole curve.
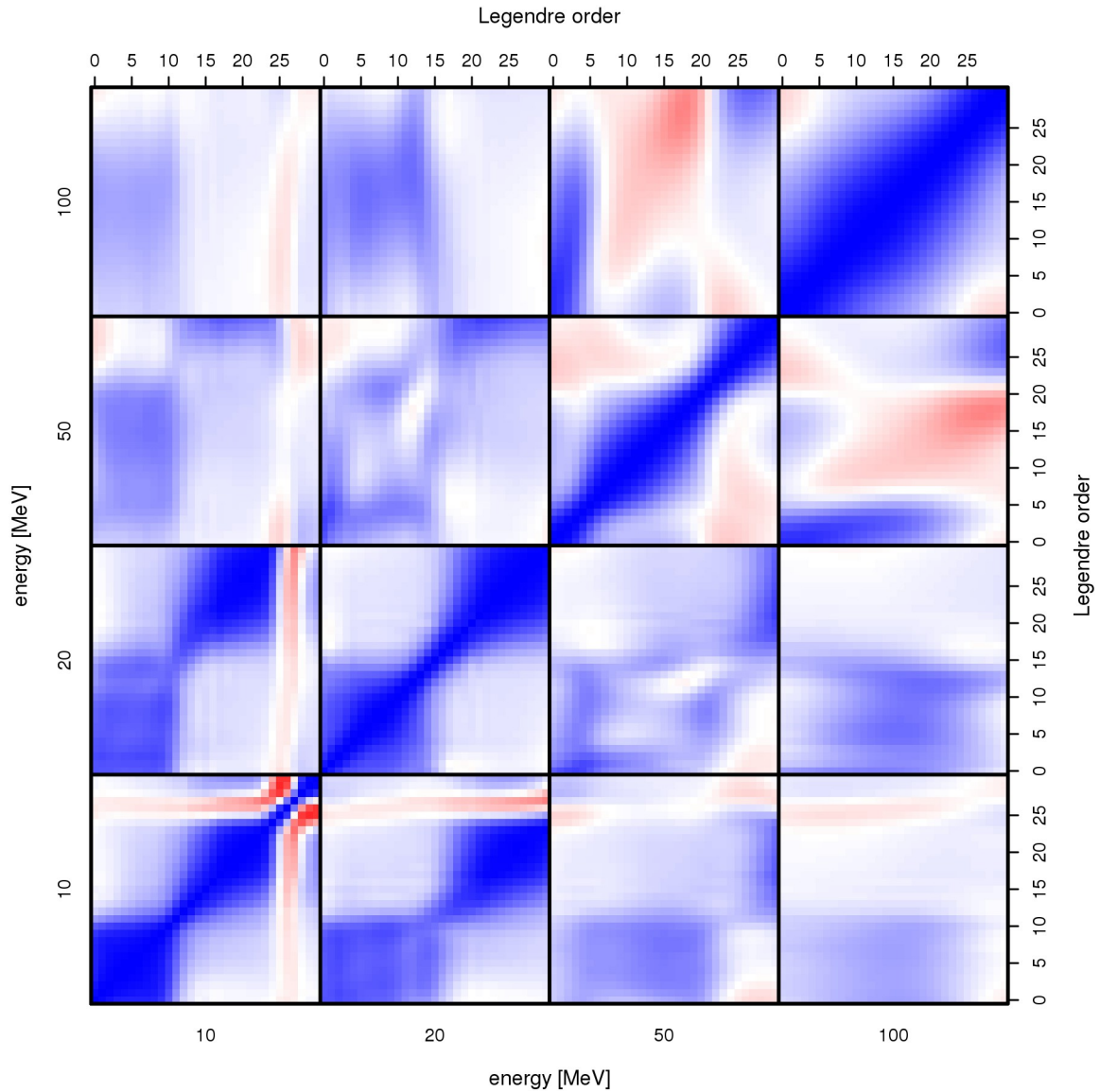
Figure 5.39: The correlation matrix for the differential elastic CS channel after a Bayesian update with all (Foster, Finlay, Smith) experimental data. The structural properties are significantly altered compared to the prior (cf. Fig. 5.25). Especially, the correlations are much less pronounced. This probably means that the posterior does not possess the comfortable feature of fast decreasing eigenvalues any more. Many differently structured eigenvectors contribute considerably to the evaluated covariance matrix. The colour code is the same as in Fig. 5.38
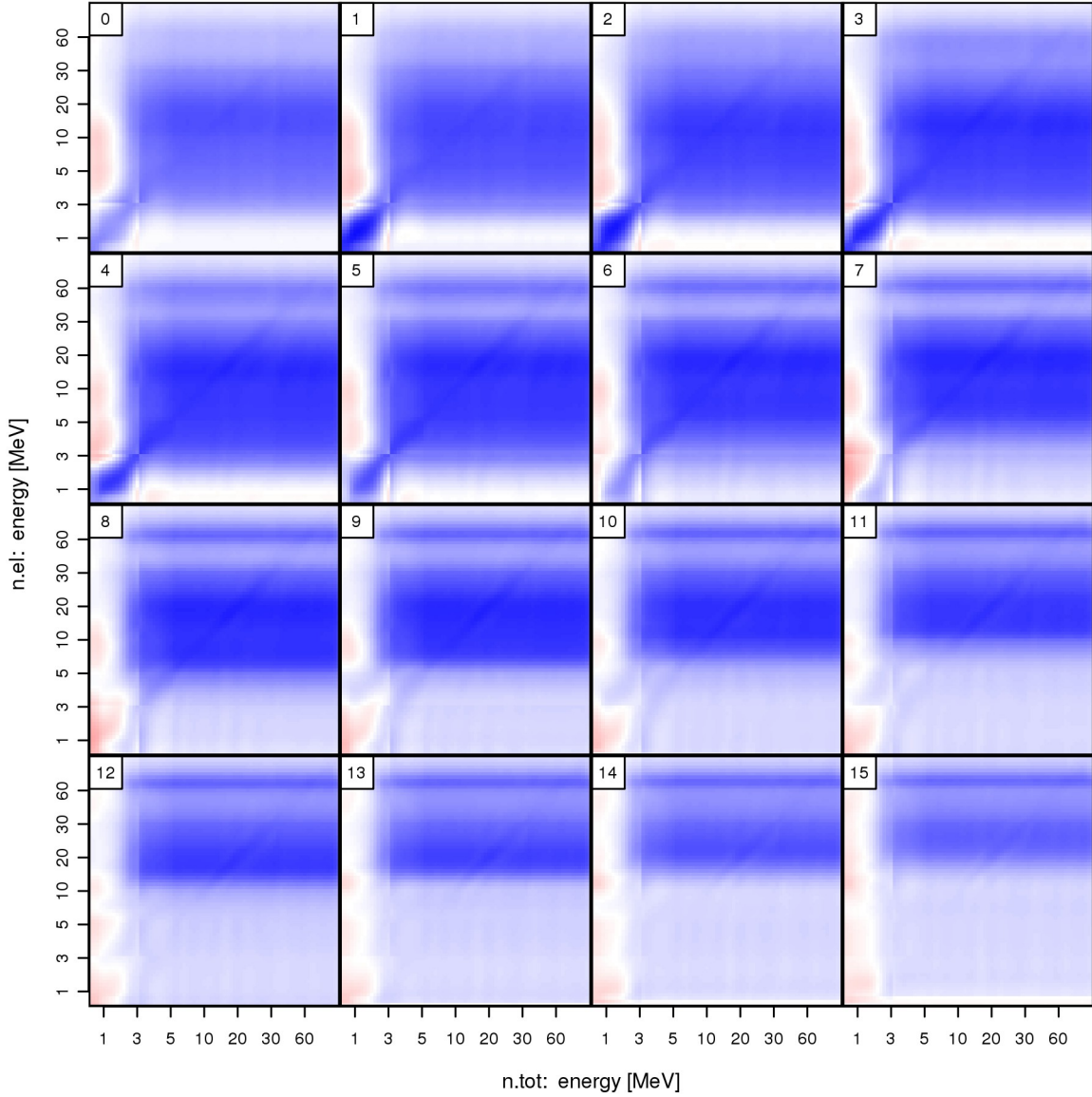
81

Figure 5.40: Correlations between the total cross section and Legendre coefficients for the elastic channel up to the order of 15 after the Bayesian update with all experimental data. The white wave-like structures observable for the prior in Fig. 5.40 vanished. The correlations are weaker than for the prior and more uniform. The correlations between low order Legendre coefficients and the total CS, where one of these quantities is associated with an energy below 2.5 MeV and the other to an energy above 2.5 MeV, are not or slightly negatively correlated. Via the correlations, the exp. estimates for the total CS lead to predictions for energies below 2.5 MeV for the diff. el. CS which are at odds with the model expectation. Thus, the associated correlations are cancelled out by the Bayesian update procedure.
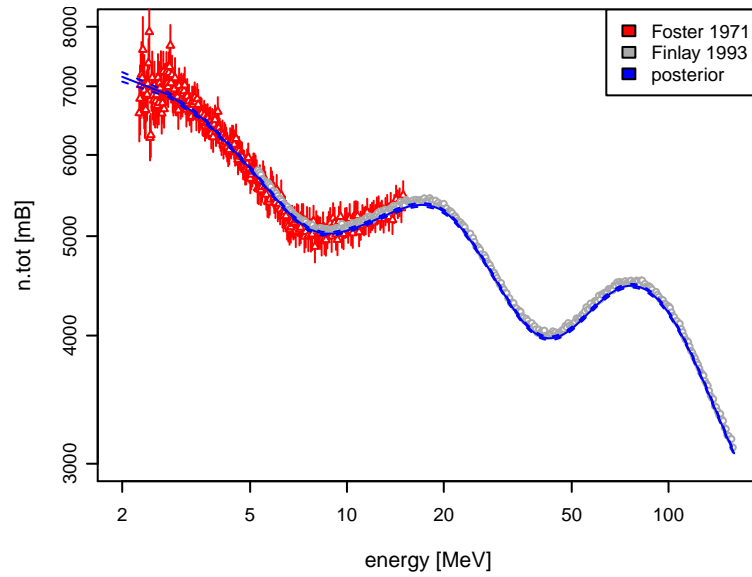
Figure 5.41: The posterior in comparison to the experimental data for the total CS if all experimental data (Foster, Finlay, Smith) enter the evaluation. The result is optically the same as when only total CS data would have been entered into the evaluation (cf. Fig. 5.30). The evaluated uncertainty is about 0.5% and the same for every incident energy.
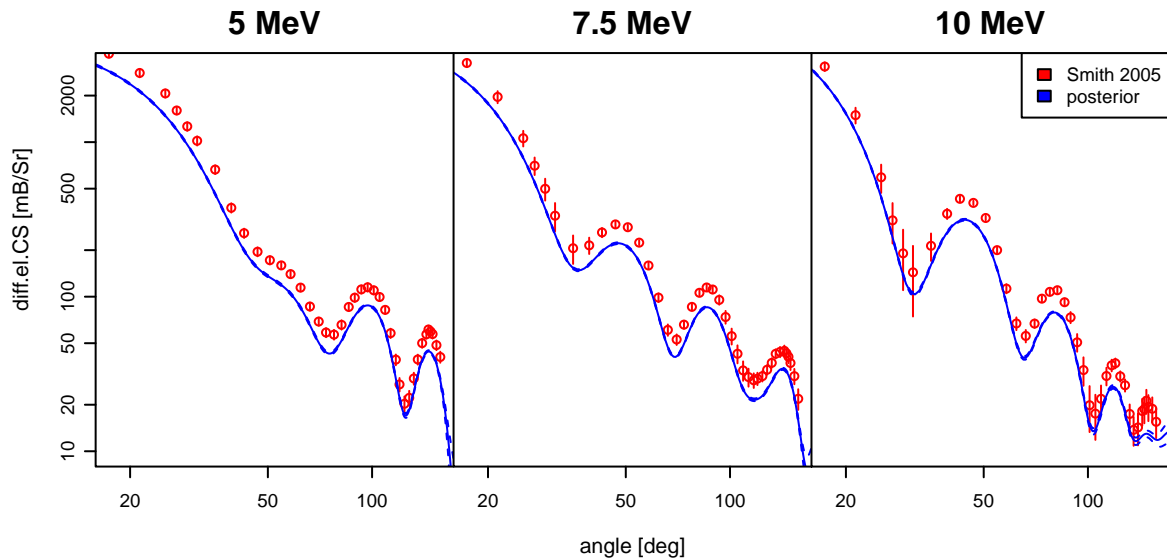
Figure 5.42: The posterior in comparison to the experimental data for the differential CS if all experimental data had been included in the evaluation. The evaluated uncertainties are between 0.9% and 7.5% which is slightly less than in the case of only updating with differential elastic data (1.8% and 8.6%). The median of the relative deviation of the prediction from the exp. estimates is 24.2% – only updating with diff. el. data, it is 25.1%.

5.5 Prior smoothing by linear dimension reduction (PCA)

In the past the Full Bayesian Evaluation Technique had been criticised for being sensitive to experimental outliers. For some energy regions, the total cross section curve is to be expected to have a smooth shape without any oscillatory features. However, performing the Bayesian update procedure with displaced data, i.e. their location in comparison to the whole curve is unphysical, the evaluated curve possesses kinks or oscillations at the energies of the outliers. The result of such an evaluation is illustrated in the left image of Fig. 5.44.

A possible solution to this problem is to diagonalise the prior covariance matrix, set all eigenvalues which fall below a certain value to zero and afterwards transform the truncated matrix back to the original basis. Diagonalising the covariance matrix pictorially means to align the coordinate axes along the principal axes of the ellipsoids representing the isolines of probability density associated with the Gauss distribution. For instance, after a rotation by 45° of the coordinate system plotted in Fig. 5.47, the coordinate axes are aligned to the principal axes of the green ellipse assigned to the prior.

Since the eigenvalues represent the variances along the principal axes, setting these to zero which already have a small value, means to remove degrees of freedom in directions for which the prior makes sharp predictions, i.e. estimates associated with a small uncertainty. If the removed eigenvalues are order of magnitudes smaller than the leading ones, the structure of the correlation matrix is retained.

As the eigenvalues of the prior covariance matrix for the total CS indeed decrease rapidly (cf. Fig. 5.5), it is sufficient to keep the ten leading eigenvectors. Visual inspection of the correlation matrix of the reduced prior (Fig. 5.43) reveals hardly any differences compared to the unaltered prior (Fig. 5.15).

Performing the Bayesian update procedure again on the same data which caused the kinks, this time using the reduced prior, yields the evaluated curve pictured in the right part of Fig. 5.44. No unphysical oscillations are visible there. Furthermore, the outliers hardly affect evaluated uncertainties. The evaluated confidence bands are even at the energies of the outliers in the magnitude of the experimental normalization error.

Though, one should be aware that the model becomes more restrictive by truncating the covariance matrix. If the model is deficient, i.e. it regards the 'true' curve highly unlikely or maybe completely excludes it, truncation worsens the situation. Then, independent of how overwhelming experimental evidence indicates a particular curve shape, the evaluated curve will never resemble it adequately, but nevertheless evaluated uncertainties tend towards zero.

In the case of tantalum-181, the model seems to be reliable. Fig. 5.44 shows evaluated curves resulting from calculations with truncated prior covariance matrices in comparison to the used experimental data (Finlay [9]). In order to reduce the influence of the model bias on the result to a minimum, the experimental normalization error was disregarded for the evaluation. Fig. 5.45 demonstrates that already the truncation to 5 eigenvalues

results in an evaluated curve resembling the experimental data quite well. The evaluated curves which arise by using 10 and 15 eigenvalues are hardly discernible from each other and optically fit the experimental data perfectly.

Apart from truncating for the purpose of smoothing evaluated curves, truncating may also be reasonable for a statistical reason. The prior covariance matrix is estimated from a limited number of sample vectors. Therefore, the obtained sample correlations are only estimates for the true correlations. Conceivably, fluctuations of only a few percent in their values lead to significant modifications of the eigenvectors associated with eigenvalues magnitudes of order smaller than the biggest eigenvalues. Supposedly modifications are significant, these eigenvectors tell little to nothing about the model systematics but are only artefacts of the sampling procedure. In this case, it might be advantageous to remove them as well.
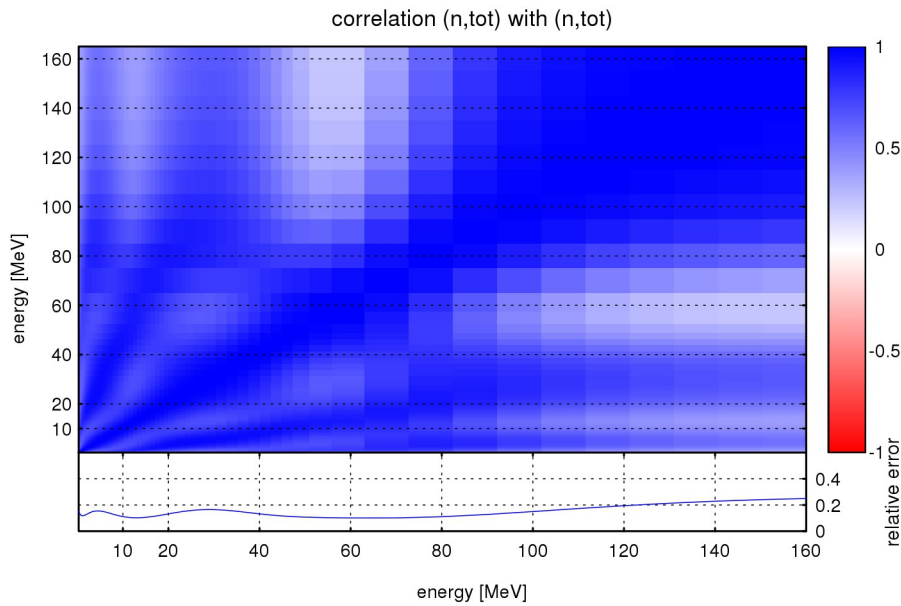

Figure 5.43: The correlations implemented in the reduced prior where the covariance matrix for the total CS is truncated to the 10 leading eigenvalues. Comparison with the correlations established by the unaltered prior shown in Fig. 5.15 reveals hardly any differences.

85
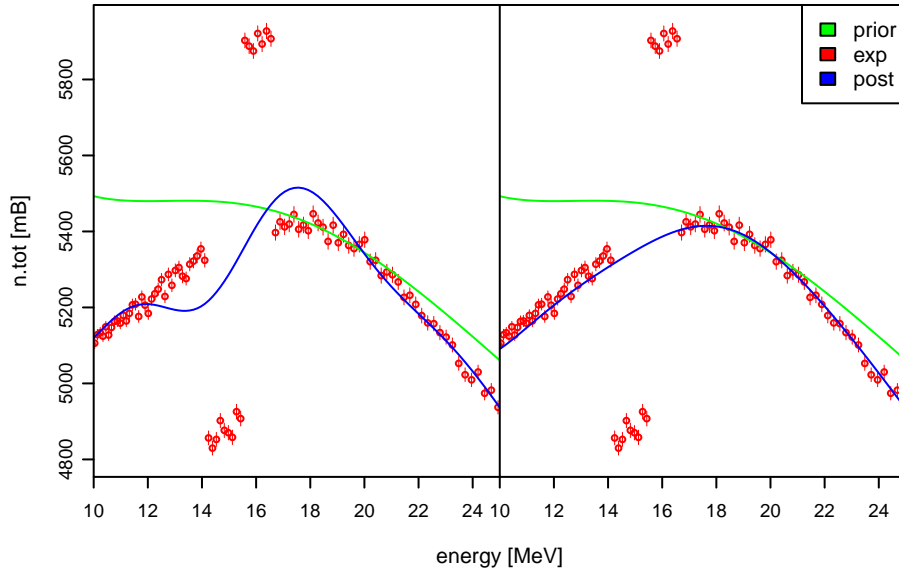
Figure 5.44: The left diagram features an evaluation with the unaltered prior and experimental data containing unphysical outliers. The evaluated curve in the right diagram results from a calculation with a prior truncated to 10 eigenvalues. The experimental data from Finlay [9] was taken and artificial outliers were inserted.

Figure 5.45: The resulting curves of evaluations with prior covariance matrices truncated to different numbers of eigenvalues. The number in the box denotes the number of kept eigenvalues. The scale is given for the lowest curve. Each other curve is shifted away by 500 mB from the lower neighbour curve.

5.6 Discussion of the results

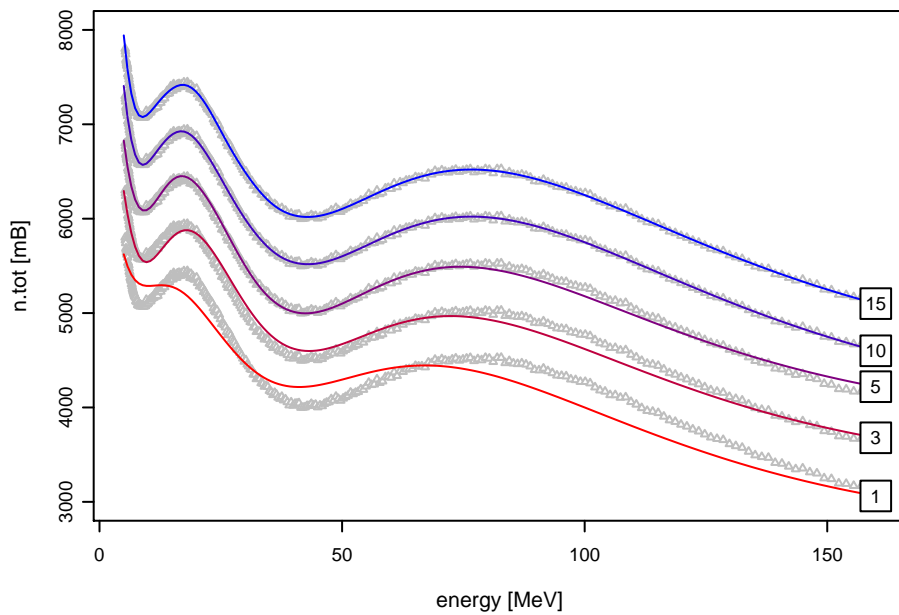The previous sections demonstrated that an update with the total CS data leads to a posterior curve which resembles the total CS data well (Fig. 5.30). However, the assumption of an overall normalization error of 0.5% causes the posterior to be located slightly but visible below the experimental data. The evaluated uncertainty is approximately 0.5% at every energy. Updating with the total CS data, but without considering a normalization error, the evaluated total CS curve differs nowhere more than 1% from the experimental curve and the evaluated uncertainties are between 0.052% and 0.17% – seemingly far too low to be a plausible result.

Regardless of the choice for the normalization error, an update with the total CS data leads to a blatant underestimation of the differential elastic CS at low angles (Fig. 5.31). Furthermore, evaluated error boundaries in forward direction become very narrow and render the differential elastic CS data from Smith [36] highly unlikely. This means either Smith measured the differential elastic CS completely inaccurate – relative errors of 60% have to be assumed to justify the evaluated model curve – or the model predicts the portion of the reaction CS at the total CS seriously wrongly. A wrong prediction for a channel without data is acceptable in the first place, because it is only a guess for the true value. However, the small uncertainties exclude possibilities for the differential elastic CS curve which do not follow the model curve closely.

Thus, the model possesses defects concerning the ratio of reaction CS to total CS. Further indication for this statement to be true is given by Fig. 5.32. If the model is forced to exactly fit the differential elastic CS curve by setting the normalization error to zero, the total CS curve is moved upward.

Updating only with the differential elastic CS data from Smith with a normalization error of 5% leads to the evaluated differential elastic CS curves in Fig. 5.33. The error boundaries at all angles are narrow and do not include the experimental data. The maximal observed deviation in forward direction of the model curve from the experimental data is 985 mB.

If the optical potential implemented in TALYS [23] is believed to describe the total CS and the differential elastic CS of tantalum-181 adequately, there has to be a problem with the chosen parameter boundaries for the optical model, which have been used for sampling and constructing the prior covariance matrix. If the optical model is in principle not able to describe the CS features of tantalum-181, a statistical method based on the wrong physical model is pointless. Before further efforts are undertaken to incorporate differential elastic experimental data into the method, this issue must be resolved.

In all performed calculations the evaluated uncertainties were in the magnitude of the assumed overall normalization error of the experiment, e.g. updating with the differential elastic CS data of Smith [36] and an assumed normalization error of 5% led to evaluated uncertainties between 2% and 8%.

Ensuring evaluated uncertainties not to drop significantly below the normalization error of the experiment is a necessary (but not sufficient) criterion stated in the Ph.D. thesis of D. Neudecker [26] to obtain a valid evaluation. Another criterion could be that the evaluated uncertainties for channels without experimental data should not drop to

very low values as it was observed for the differential elastic channel by updating with total CS data.

5.7 Discussion of the method

The symmetric prior covariance matrix is only guaranteed to be positive semidefinite, hence does not need to have full rank. A reduced rank indicates directions in which no deviation is allowed. The concerned vector space is spanned by vectors with cross section values for different channels, incident energies and angles as elements. The directions in which cross sections are allowed to vary are given by the eigenvectors of the prior covariance matrix associated with eigenvalues greater than zero. It has been shown that the Bayesian update formulas Eq. (3.42) and Eq. (3.43) implement the projection into the subspace spanned by the eigenvectors with non-zero eigenvalues. Thus, the prediction resulting from a Bayesian update step is restricted to lie in this subspace. Consequently, if the eigenvectors are explicitly known, there is no need to consider the whole cross section space. The Bayesian update formulas can be stated for all quantities reduced to this subspace. In order to update with experimental data, the experimental data then must be projected into this subspace before the Bayesian update step can be performed.

In section 5.2.1 the eigenvalue decomposition of the block in the prior covariance matrix associated with the total CS revealed a lot of directions to be negligible. And indeed, the Bayesian update procedure with a trimmed prior covariance matrix, where all except for the 16 biggest eigenvalues were set to zero, yielded optically the same evaluated curve for the total CS as if the full prior covariance matrix would have been used (Fig. 5.45). This finding is of great use for speeding up the method, because the covariance matrix block for the total CS can be reduced to 16 dimensions without losing important features of the model.

However, at the same time it is questionable if real measured cross section curves can always be described accurately by a linear combination of a very limited set of functions which are allowed by the physical model. Likely, some features of the experimental data cannot be described, as it would mean the model can be fitted perfectly to the experimental data. In order to understand what happens when the model is not able to predict all features of the experimental curve, a toy model will be studied in the following.

Assume that the model exclusively allows the same cross section value at each incident energy. Thus, the model restricts cross section curves to straight horizontal lines and the only degree of freedom is an overall shift. Further assume that the experimental data points follow a diagonal line and are given for the same energies as the model data points. In this case, the sensitivity matrix to map from model to experiment space is the identity matrix which will simplify the subsequent considerations. The situation is pictured in Fig. 5.46.

One degree of freedom means that the prior covariance matrix has rank 1, therefore only one eigenvector is associated with a non-zero eigenvalue. Every element of this
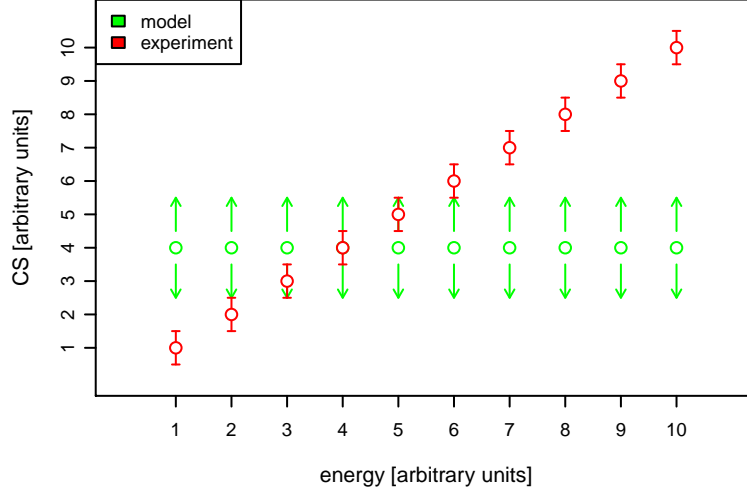
Figure 5.46: Visualization of the toy model: the model exclusively allows the same CS value at every energy. The measured values of the experiment resemble a diagonal line. Thus, the model is deficient.

eigenvector is given by the same value, which implements the restriction to a a horizontal line. The eigenvalue v associated with this eigenvector represents the variance $n\delta^2_{\text{mod}}$ along the latter, where n is the number of energy grid points and δ^2_{mod} the variance at every energy grid point assumed by the model. The variance $n\delta^2_{\text{mod}}$ along the eigenvector can also be regarded as the 1×1 matrix A_{red} obtained by projecting the original $n \times n$ prior covariance matrix A to the 1-dimensional subspace defined by v. The factor n is necessary because otherwise the variance of the model at every energy grid point would be dependent on the number of grid points. This is due to the fact that back projecting the variance associated with the eigenvector into the whole vector space leads to an additional factor $1/n$ caused by the normalization of the eigenvector. The formulas for reducing the prior covariance matrix A to the subspace of allowed directions and extending the reduced matrix to the original representation are given by Eq. (3.35).

The important quantities needed for the Bayesian update step are:

$$v^T = \frac{1}{\sqrt{n}} \overbrace{(1, 1, \cdots, 1)}^{\text{n elements}} \qquad A_{\text{red}} = n\delta^2_{\text{mod}}$$

$$A = vA_{\text{red}}v^T = J_n\delta^2_{\text{mod}} \qquad B_{\text{red}} = v^T B v = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}$$

$$x^T = c\overbrace{(1, 1, \cdots, 1)}^{\text{n elements}} \qquad x_{\text{red}} = v^T x = c\sqrt{n}$$

$$y_{\text{red}} = v^T y = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} y_i = \overline{y}\sqrt{n}$$

x is the prior and y the experimental estimate for the true CS vector; the factor c denotes the y-value of the prior estimate for the horizontal line and $\overline{y}$ is the mean value of the

experimental CS values; the matrix A is the prior and B the experimental covariance matrix; v is the normalized eigenvector of the matrix A associated with the only non-zero eigenvalue $n\delta^2_{\mathrm{mod}}$; B_{ij} denotes the element at row i and column j in the matrix B; analogously y_i denotes the i-th element in the experimental estimate vector y, so each y_i is the CS value at a certain incident energy E_i; J_n is the $n \times n$ matrix of ones – every element is one. All 'red' quantities are obtained from the original ones by projecting them into the 1-dimensional subspace defined by v.

In order to obtain the posterior covariance matrix $\tilde{A}_{\mathrm{red}}$, one can use for convenience formula (3.34) instead of update formula (3.43) which was derived from the former by the Woodbury matrix identity (3.28). Since the sensitivity matrix S is the identity matrix, one finds the expression

$$\tilde{A}_{\mathrm{red}} = \left(A_{\mathrm{red}}^{-1} + B_{\mathrm{red}}^{-1}\right)^{-1} = \frac{A_{\mathrm{red}} B_{\mathrm{red}}}{A_{\mathrm{red}} + B_{\mathrm{red}}} = \frac{\delta^2_{\mathrm{mod}} \sum B_{ij}}{n\delta^2_{\mathrm{mod}} + \frac{1}{n}\sum B_{ij}}. \tag{5.11}$$

For the sake of simplicity, counter variables are not written under the sum symbols which denotes the summation over all n^2 elements B_{ij}.

A simple experimental covariance matrix may include a statistical error Δ^2_{stat} in the diagonal and an overall normalization error Δ^2_{sys} affecting every element:

$$B_{ij} = \Delta^2_{\mathrm{stat}}\delta_{ij} + \Delta^2_{\mathrm{sys}}, \tag{5.12}$$

where δ_{ij} is the Kronecker-Delta, being 1 for $i = j$ and 0 otherwise. Inserting Eq. (5.11) into Eq. (5.11) yields

$$\tilde{A}_{\mathrm{red}} = \frac{\delta^2_{\mathrm{mod}}\, n\Delta^2_{\mathrm{stat}}}{n\delta^2_{\mathrm{mod}} + \Delta^2_{\mathrm{stat}} + n\Delta^2_{\mathrm{sys}}} + \frac{\delta^2_{\mathrm{mod}} n^2 \Delta^2_{\mathrm{sys}}}{n\delta^2_{\mathrm{mod}} + \Delta^2_{\mathrm{stat}} + n\Delta^2_{\mathrm{sys}}}. \tag{5.13}$$

Finally, this result can be transferred back to obtain the complete evaluated covariance matrix $\tilde{A}$:

$$\tilde{A} = v\tilde{A}_{\mathrm{red}}v^T = \frac{\tilde{A}_{\mathrm{red}}}{n}\, J_n = \left(\frac{\delta^2_{\mathrm{mod}}\, \Delta^2_{\mathrm{stat}}}{n\delta^2_{\mathrm{mod}} + \Delta^2_{\mathrm{stat}} + n\Delta^2_{\mathrm{sys}}} + \frac{\delta^2_{\mathrm{mod}} n\Delta^2_{\mathrm{sys}}}{n\delta^2_{\mathrm{mod}} + \Delta^2_{\mathrm{stat}} + n\Delta^2_{\mathrm{sys}}}\right) J_n \tag{5.14}$$

The first important observation is that the evaluated uncertainties are completely determined by the prior and experimental uncertainties. The prior and experimental estimates have no influence. Furthermore, all elements of the evaluated covariance matrix $\tilde{A}$ are equal, which indicates a perfect correlation for arbitrary pairs of CS values at different energies. This is a consequence of the fact that the toy model only permits an overall shift of the horizontal line.

Setting the normalization error Δ_{sys} to zero and assuming $\delta_{\mathrm{mod}} \gg \Delta_{\mathrm{stat}}$ results in $\tilde{A}_{ij} \approx \Delta^2_{\mathrm{stat}}/n$. Thus, the uncertainties decrease by a factor $1/\sqrt{n}$. This proportionality remains valid even if the assumption $\delta_{\mathrm{mod}} \gg \Delta_{\mathrm{stat}}$ does not hold.

Setting the statistical error Δ_{stat} to zero and assuming $\delta_{\mathrm{mod}} \gg \Delta_{\mathrm{sys}}$, the evaluated uncertainty is approximately given by the normalization error Δ_{sys} of the experiment. If on the other hand $\delta_{\mathrm{mod}} \ll \Delta_{\mathrm{sys}}$, the evaluated uncertainty is given by δ_{mod}.

For an increasing amount of experimental data, the influence of the statistical error Δ_{stat} on the evaluated uncertainties tends to vanish and the latter is exclusively determined by the model uncertainty δ_{mod} and the normalization error Δ_{sys}.

To obtain the expression for the evaluated estimate, update formula (3.42) reduced to the allowed subspace is applied:

$$\begin{aligned}\tilde{x}_{\text{red}} &= x_{\text{red}} + A_{\text{red}}(A_{\text{red}} + B_{\text{red}})^{-1}(y_{\text{red}} - x_{\text{red}}) \\ &= \frac{x_{\text{red}}B_{\text{red}} + y_{\text{red}}A_{\text{red}}}{A_{\text{red}} + B_{\text{red}}} = \frac{c\sqrt{n}\Delta_{\text{stat}}^2 + c\,n^{3/2}\Delta_{\text{sys}}^2 + \overline{y}\,n^{3/2}\delta_{\text{mod}}^2}{n\delta_{\text{mod}}^2 + n\Delta_{\text{sys}}^2 + \Delta_{\text{stat}}^2}.\end{aligned} \tag{5.15}$$

The transfer back to the original representation yields

$$x = v\,x_{\text{red}} = \frac{c\,\frac{1}{n}\Delta_{\text{stat}}^2 + c\,\Delta_{\text{sys}}^2 + \overline{y}\,\delta_{\text{mod}}^2}{\delta_{\text{mod}}^2 + \Delta_{\text{sys}}^2 + \frac{1}{n}\Delta_{\text{stat}}^2}\,\vec{1}, \tag{5.16}$$

where $\vec{1}$ is a column vector with n elements, all equal one. If the normalization error Δ_{sys} is zero and the prior model uncertainty δ_{mod}^2 large compared to the quantity Δ_{stat}^2/n, the posterior estimate for the CS value at each energy grid point is given by the mean value of the experimental CS values. Therefore, the toy model completely ignores the information at which energies the data were obtained.

If the systematic error Δ_{sys}^2 is not vanishing and enough experimental data has been collected to render the statistical error negligible, the evaluated estimate is a compromise between the model prediction c and the mean value of the experimental data $\overline{y}$. In the limiting case $\Delta_{\text{sys}} \ll \delta_{\text{mod}}$, the posterior estimate is given by $\overline{y}$ and in the case $\Delta_{\text{sys}} \gg \delta_{\text{mod}}$ by c. Independent of these assumptions, the evaluated CS curve will always resemble a horizontal line perfectly.

These considerations yield an important conclusion: If the model strongly correlates CS values at different energies, but simultaneously excludes the shape of the 'true' CS curve as a possible solution, the prediction can become arbitrarily bad. Furthermore, if a huge amount of experimental data is entered into an evaluation, the evaluated uncertainties are determined by either the experimental normalization error or the model uncertainty depending on which quantity is smaller. Hence, there is no guarantee that the experimental curve lies anywhere close or within the confidence bands of the evaluated curve. A case for cross sections at two different energies where this behaviour could occur is schematically visualized in Fig. 5.47.

Based on the knowledge obtained by the studies on the toy model, the results of section 5.4 can be interpreted as follows: The eigenvalue decomposition of the block in the prior covariance matrix associated with the total CS proved the bulk of directions to be negligible (Fig. 5.5). This fact is also expressed by strong correlations illustrated in Fig. 5.15. Thus, the model is very restrictive about the shape of the evaluated total CS curve. Nonetheless, it seems to be a valid description of reality as a linear combination of the 16 leading eigenvectors associated with highest variances is able to reproduce the experimental curve really well – optically it is a perfect fit (Fig. 5.45). However, at about the 10th leading eigenvector, the uncertainties along the eigenvectors get into the

magnitude of the experimental normalization error. Therefore, $\Delta_{\mathrm{sys}} \ll \delta_{\mathrm{mod}}$ does not hold and the bias given by the a-priori model estimate has a perceivable influence on the posterior estimate – the evaluated total CS curve is visibly shifted away from the experimental CS curve (Fig. 5.30).

The same behaviour – though more extreme – is observed for the differential elastic CS channel. An evaluation with the differential elastic data of Smith [36], without considering a normalization error, resulted in a perfect fit of the evaluated curve to the experimental data (Fig. 5.37). Thus, in principle, the experimental curve is allowed by the model, but the experimental curve is also regarded as highly unlikely by the model and consequently the posterior estimate is visible biased towards the prior estimate. The evaluated uncertainty is in the magnitude of the normalization error. This indicates eigenvectors associated with high variances mainly contribute because otherwise the evaluated uncertainty would not lie in the magnitude of the normalization error. Nonetheless, many more eigenvectors with much lower variances are needed to resemble the shape of the experimental curve. For these eigenvectors, the model uncertainty is much less than the experimental uncertainty and the evaluated estimates along these are strongly biased by the model expectation.
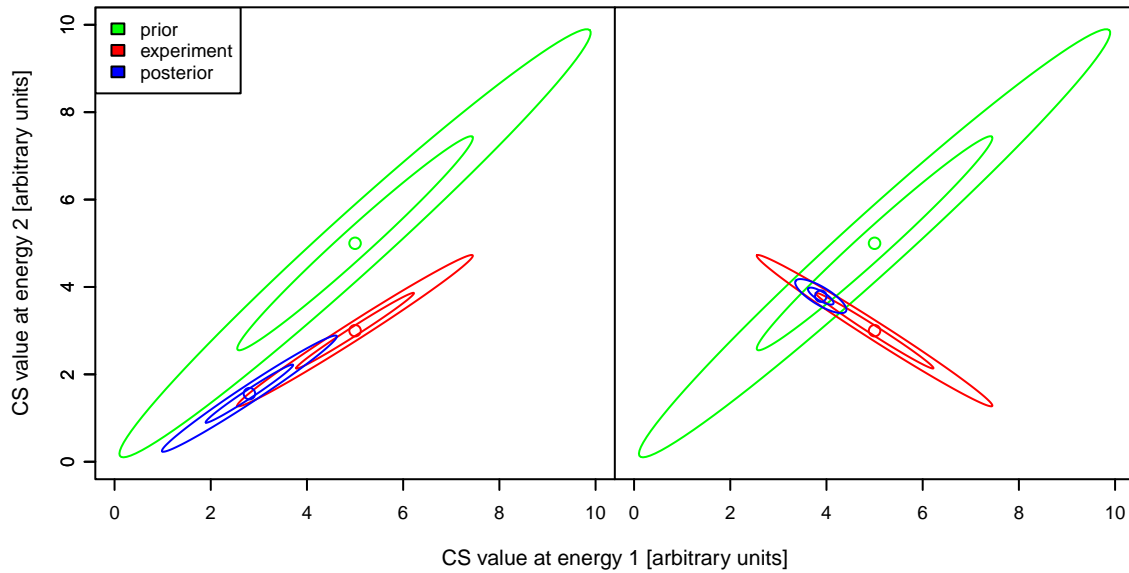
Figure 5.47: Two-dimensional Gaussian prior (green) and experimental (green) density for two cross sections x_1 and x_2 at energies E_1 and E_2 respectively. According to the model, x_1 and x_2 are strongly positively correlated perceivable by the high eccentricity of the green ellipses. The inner ellipses represent the 1-σ-confidence area and the outer the 2-σ-confidence area. The CS x_1 and x_2 are also strongly correlated according to the experiment which means the normalization error is much greater than the statistical error. The left diagram features a situation where the normalization error is in the direction of high model uncertainty. The evaluated uncertainty is determined by the normalization error. In the right diagram the experimental uncertainty is in a direction (i.e. specified by an eigenvector) of low model uncertainty. Then, the evaluated uncertainty is determined by the a-priori uncertainty of the model. Furthermore, the experimental estimates are not included in the evaluated 2-σ-confidence area.

6 Summary

The aim of this thesis was to extend the Full Bayesian Evaluation Technique (FBET) to allow the inclusion of experimental angle differential elastic cross section data into evaluations and test its applicability to this new data type.

In order to test the applicability, the extent to which the utilized multi-dimensional Gauss distribution assumed in the actual FBET is able to capture the systematics of the model, i.e. the real probability density function, was investigated. Although a cluster analysis has shown that the assumption of a multi-dimensional Gauss distribution is not the most appropriate choice, the latter is still capable to capture the restrictions imposed by the model, which was indicated by the existence of strong correlations: In both the total CS channel and the differential elastic CS channel, cross sections at different energies were found to be strongly correlated. In addition, strong correlations for differential elastic CS at different angles were observed. Also, strong correlations between the total CS and the differential elastic CS channel exist.

An eigenvalue decomposition of the block in the prior covariance matrix associated with the total CS demonstrated that about 16 basis functions are sufficient to mimic the majority of the total CS sample curves with an accuracy of about some tenth of a percent. For the differential elastic CS, about 400 basis functions are required to resemble the differential elastic CS sample curves with an accuracy of about $0.1 \, \mathrm{mB/Sr}$. This finding represents not only additional support for the applicability of covariance matrices to capture the systematics of the model, but also suggests a way for speeding up the method, i.e. the reduction of the prior covariance matrix by removing negligible dimensions. In the implementation of this work, differential CS data are associated with a block in the prior covariance matrix of dimension 5000×5000, and therefore the reduction of dimension is an important step to make the FBET feasible to work with several differential CS channels.

Since the included experimental differential elastic CS data contain contributions from inelastic scattering due to the limited energy resolution of the experiment, the effect of these contributions was also studied. At large angles, where differential elastic cross sections are small quantities, the inelastic contributions indeed have a non-negligible impact on the shape of the angular distribution. Consequently, they must be considered in evaluations of differential elastic CS data.

Evaluations of tantalum-181 demonstrated the strong coupling between the total CS and the differential elastic CS channel. Updating with experimental total CS data resulted in significant changes of the evaluated estimates for the differential elastic CS; updating with experimental differential elastic CS data lead to great changes in the evaluated total CS. Even small differential elastic CS values for large angles at energies below $10 \, \mathrm{MeV}$ had a great impact on the evaluated total CS over the complete energy

range from 0.3 MeV to 160 MeV. Due to strong correlations and the large number of experimental data points available for both channels, the evaluated uncertainties dropped to the magnitude of the normalization error of the experiments. For both channels, the evaluated curves did not match the included experimental data points; neither was the latter contained within the evaluated 1-σ-confidence bands. In the case of the total CS, the evaluated CS values were slightly below the experimental data; in the case of the differential elastic CS, the posterior underestimated the experimental data more pronounced. Therefore, the expectation of the prior about the cross sections has a significant influence on the result of an evaluation. The deviations of nearly one barn observed for the differential elastic CS at low angles indicate that the prior regards the included experimental differential elastic CS data as highly unlikely. The underestimation at low angles even occurs if only total CS data entered into the evaluation. This finding suggests that the prior is based on an expectation about the ratio of the integral elastic CS to the reaction CS, which is in disagreement with the experimental data. Therefore, the parameter boundaries for the optical potential used for constructing the prior may be inadequate. Further research on this issue is necessary.

Furthermore, the evaluation of tantalum-181 showed that the result of an evaluation reacts sensitively to the inclusion of experimental differential elastic CS data at large angles. If the physical model is not valid in this angle domain, even the evaluated total CS is inappropriately altered. Thus, future investigation of the validity of the physical models, especially with regard to the large angle domain, is required. If the models will turn out to be deficient, the Full Bayesian Evaluation has to account for that. Otherwise, unreasonable evaluations have to be expected.

Apart from studies of the behaviour of FBET extended to angular differential CS, the issue of unphysical kinks and oscillations in the evaluated curves caused by experimental outliers was tackled. It was shown that it is possible to reduce the sensitivity to outliers by truncating the prior covariance matrix, i.e. eliminating eigenvalues smaller than a critical value. By this procedure, directions in cross section space which are associated with low uncertainty are removed. In these directions, the model bias significantly influences the results of evaluations. A huge amount of experimental evidence is necessary to overcome it. By truncating the prior covariance matrix, the degree of freedom along the concerned directions is completely removed and the estimate for these directions irrevocable frozen to the model expectation. Therefore, the application of this procedure requires high trust in the underlying physical models.

Albeit the main goal of this work, the extension of the Full Bayesian Techniques to angular differential elastic data, was achieved, its applicability to routine evaluations requires still further investigation. Considering the influence of the prior expectation on the results, the question arises whether the uniform distribution in the parameter space of the optical model is the most practicable choice. An uniform distribution in the cross section space which also imposes the restrictions of the model may be more appropriate, thereby eliminating the model bias. How this can be done is a technical question – if it should be done a philosophical one.

Bibliography

[1] I. A. E. Agency. EXFOR database. http://www-nds.iaea.org/exfor/exfor.htm.

[2] N. E. Agency. JEFF Nuclear Data Library. http://www.oecd-nea.org/dbdata/jeff/.

[3] E. Bauge, R. Capote, U. Fisher, A. Y. Konobeyev, P. E. Pereslavtsev, M. Herman, P. Oblozinsky, M. T. Pigni, T. Kawano, P. Talou, I. Kodeli, A. Trkov, A. Koning, D. Rochman, H. Leeb, D. Neudecker, and D. L. Smith. Covariance data in the fast neutron region. Technical Report NEA/NSC/WPEC/DOC(2010)427, Nuclear Science NEA/WPEC-24, OECD, Nuclear Energy Agency, 2011.

[4] R. E. Benenson, K. Rimawi, E. H. Sexton, and B. Center. Small-angle elastic scattering of 14.8 MeV neutrons. *Nuclear Physics A*, 212:147–156, 1973.

[5] M. B. Chadwick, M. Herman, P. Obložinský, M. E. Dunn, Y. Danon, A. C. Kahler, D. L. Smith, B. Pritychenko, G. Arbanas, R. Arcilla, R. Brewer, D. A. Brown, R. Capote, A. D. Carlson, Y. S. Cho, H. Derrien, K. Guber, G. M. Hale, S. Hoblit, S. Holloway, T. D. Johnson, T. Kawano, B. C. Kiedrowski, H. Kim, S. Kunieda, N. M. Larson, L. Leal, J. P. Lestone, R. C. Little, E. A. McCutchan, R. E. MacFarlane, M. MacInnes, C. M. Mattoon, R. D. McKnight, S. F. Mughabghab, G. P. A. Nobre, G. Palmiotti, A. Palumbo, M. T. Pigni, V. G. Pronyaev, R. O. Sayer, A. A. Sonzogni, N. C. Summers, P. Talou, I. J. Thompson, A. Trkov, R. L. Vogt, S. C. van der Marck, A. Wallner, M. C. White, D. Wiarda, and P. G. Young. ENDF/B-VII.1 Nuclear Data for Science and Technology: Cross Sections, Covariances, Fission Product Yields and Decay Data. *Nuclear Data Sheets*, 112(12):2887–2996, 2011.

[6] R. T. Cox. Probability, frequency, and reasonable expectation. *Am. J. Phys.*, 14(1):1–13, 1946.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series (B)*, 39(1):1–38, 1977.

[8] J. C. Ferrer, D. Carlson, and J. Rapaport. Neutron elastic scattering at 11 MeV and the isospin dependence of the neutron-nucleus optical potential. *Nuclear Physics A*, 275:325–341, 1976.

[9] R. W. Finlay, W. P. Abfalterer, G. Fink, E. Montei, T. Adami, P. W. Lisowski, G. L. Morgan, and R. C. Haight. Neutron total cross sections at intermediate energies. *Phys. Rev. C*, 47:237–247, Jan 1993.

[10] D. G. Foster and D. W. Glasgow. Neutron total cross sections, 2.5-15 mev. i. experimental. *Phys. Rev. C*, 3:576–603, Feb 1971.

[11] J. Fox and H. S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition, 2011.

[12] C. Fraley and A. E. Raftery. Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, 97:611–631, 2002.

[13] C. Fraley and A. E. Raftery. MCLUST version 3 for R: Normal mixture modeling and model-based clustering. Technical Report 504, University of Washington, Department of Statistics, (2006, revised in 2009)).

[14] M. H. G. Brys and P. Rousseeuw. A robustification of independent component analysis. *J. Chemometrics*, 19:1–12.

[15] L. F. Hansen, F. S. Dietrich, B. A. Pohl, C. H. Poppe, and C. Wong. Test of microscopic optical model potentials for neutron elastic scattering at 14.6 MeV over a wide mass range. *Phys. Rev. C*, 31:111–119, Jan 1985.

[16] T. Hayfield and J. S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 2008.

[17] P. E. Hodgson. The neutron optical potential. *Reports on Progress in Physics*, 47(6):613, 1984.

[18] E. Islam, M. Hussain, N. Ameen, M. Enayetullah, and N. Islam. Total cross sections of Y, Zr, La, Ce, Pr, Gd, Ho, Ta and Hg for 1.0 to 2.0 MeV neutrons. *Nuclear Physics A*, 209:189–201.

[19] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev. II*, 106(4):620–630, 1957.

[20] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[21] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Pub. Co., 1956.

[22] A. J. Koning. Local and global nucleon optical models from 1 keV to 200 MeV. *Nucl. Phys. A*, 713:231, 2003.

[23] A. J. Koning, S. Hilaire, and M. Duijvestijn. *TALYS 1.2, A Nuclear Reaction Program*. NRG Petten, December 2009.

[24] H. Leeb, D. Neudecker, and T. Srdinko. Consistent procedure for nuclear data evaluation based on modeling. *Nuclear Data Sheets*, 109:2762–2767, 2008.

[25] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30.

[26] D. Neudecker. *The Full Bayesian Evaluation Technique and its Application to Isotopes of Structural Materials*. PhD thesis, Technical University Vienna, April 2012.

[27] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901.

[28] R. Peelle. R. Peelle's Pertinent Puzzle. Technical report, Oak Ridge National Laboratory Memorandum, Washington DC, USA, 1987.

[29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[30] A. E. Remund. *Winkelverteilung und Polarisation gestreuter Neutronen von 3,3 MeV an Kupfer, Tantal, Blei und Wismut*. PhD thesis, ETH Zürich.

[31] L. Rosen and L. Stewart. Neutron emission probabilities from the interaction of 14-MeV neutrons with Be, Ta, and Bi. *Phys. Rev.*, 107:824–829, Aug 1957.

[32] P. Rousseeuw, C. Croux, T. V. M. K. Valentin Todorov A. Ruckstuhl, M. Salibian-Barrera, and M. Maechler. *robustbase: Basic Robust Statistics*, 2012. R package version 0.9-4.

[33] G. E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[34] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[35] D. S. Sivia. *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 1996.

[36] A. B. Smith. Fast-neutrons incident on rotors: Tantalum. *Annals of Nuclear Energy*, 32:1926–1952, 2005.

[37] M. A. Woodbury. Inverting modified matrices. *Memo. Rep. 42, Princeton University*, 1950.