Approaching well-founded comprehensive nuclear data uncertainties

Fitting imperfect models to imperfect data

PETTER HELGESSON
Abstract


Nuclear physics has a wide range of applications; e.g., low-carbon energy production, medical treatments, and non-proliferation of nuclear weapons. Nuclear data (ND) constitute necessary input to computations needed within all these applications.

This thesis considers uncertainties in ND and their propagation to applications such as material damage in nuclear reactors. TENDL is today the most comprehensive library of evaluated ND (a combination of experimental ND and physical models), and it contains uncertainty estimates for all nuclides it contains; however, TENDL relies on an automatized process which, so far, includes a few practical remedies which are not statistically well-founded. A long-term goal of the thesis is to provide methods which make these comprehensive uncertainties well-founded. One of the main topics of the thesis is an automatic construction of experimental covariances; at first by attempting to complete the available uncertainty information using a set of simple rules. The thesis also investigates using the distribution of the data; this yields promising results, and the two approaches may be combined in future work.

In one of the papers underlying the thesis, there are also manual analyses of experiments, for the thermal cross sections of Ni-59 (important for material damage). Based on this, uncertainty components in the experiments are sampled, resulting in a distribution of thermal cross sections. After being combined with other types of ND in a novel way, the distribution is propagated both to an application, and to an evaluated ND file, part of the ND library JEFF 3.3.

The thesis also compares a set of different techniques used to fit models in ND evaluation. For example, it is quantified how sensitive different techniques are to a model defect, i.e., the inability of the model to reproduce the truth underlying the data. All techniques are affected, but techniques fitting model parameters directly (such as the primary method used for TENDL) are more sensitive to model defects. There are also advantages with these methods, such as physical consistency and the possibility to build up a framework such as that of TENDL.

The treatment of these model defects is another main topic of the thesis. To this end, two ways of using Gaussian processes (GPs) are studied, applied to quite different situations. First, the addition of a GP to the model is used to enable the fitting of arbitrarily shaped peaks in a histogram of data. This is shown to give a substantial improvement compared to if the peaks are assumed to be Gaussian (when they are not), both using synthetic and authentic data.

The other approach uses GPs to fit smoothly energy-dependent model parameters in an ND evaluation context. Such an approach would be relatively easy to incorporate into the TENDL framework, and ensures a certain level of physical consistency. It is used on a TALYS-like model with synthetic data, and clearly outperforms fits without the energy-dependent model parameters, showing that the method can provide a viable route to improved ND evaluation. As a proof of concept, it is also used with authentic TALYS, and with authentic data.

To conclude, the thesis takes significant steps towards well-founded comprehensive ND uncertainties.

Keywords: Evaluated nuclear data, uncertainty propagation, uncertainty quantification, model defects, Gaussian processes, TALYS, TENDL, covariances.

Petter Helgesson, Department of Physics and Astronomy, Applied Nuclear Physics, Box 516, Uppsala University, SE-751 20 Uppsala, Sweden.

© Petter Helgesson 2018

ISSN 1651-6214
urn:nbn:se:uu:diva-348553 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-348553)
“[Sannolikhetskalkyler] blir väldigt olika före och efter. [...] Före Harrisburg så var det ju ytterst osannolikt att det som hände i Harrisburg skulle hända, men så fort det hade hänt rakade ju sannolikheten upp till inte minre än hundra procent. Så det var nästan sant att det hade hänt.”

Approximately:

“[Probabilities] become very different before and after. [...] Before Harrisburg, it was extremely unlikely that what happened in Harrisburg should happen, you know. But as soon as it had happened, the probability went up to not less than a hundred percent. Making it almost true that it did happen.”

Tage Danielsson (1928-1985, Swedish writer, actor, debater, initiator of “Sommar i P1”, etc.) during his famous speech “Om sannolikhet” (On probability), 1979.¹

¹The main observation is indeed correct. In terms of conditioned probability, it can be expressed as $P(A|A) = 1$, where $A$ is the event that “Harrisburg happens”.
List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


*My contribution:* I developed and implemented the method, performed the analysis, and wrote the paper.


*My contribution:* I developed and implemented the methods, performed the analysis, and wrote the paper.


*My contribution:* I developed and implemented the methods, performed the analysis, and wrote the paper.

IV P. Helgesson, H. Sjöstrand, and D. Rochman, “Uncertainty-driven nuclear data evaluation including thermal (n,α) applied to $^{59}$Ni,” *Nuclear Data Sheets*, vol. 145, pp. 1–24, 2017.

*My contribution:* I developed and implemented the method, performed the analysis, assembled the ENDF-files, and wrote the paper.


*My contribution:* I wrote most of the scripts, performed the majority of the analysis, and wrote most of the paper. I implemented an adaptive Metropolis algorithm and the
Levenberg-Marquardt algorithm. The work on Gaussian Processes was not implemented by me.

VI  P. Helgesson and H. Sjöstrand, “Fitting a defect non-linear model with or without prior, distinguishing nuclear reaction products as an example,” Review of Scientific Instruments, vol. 88, 2017. The paper was selected as an Editor’s pick.

**My contribution:** I adapted and implemented the method of using Gaussian Processes to treat model defects for the application, developed and implemented the validation framework, performed the analysis, and wrote the paper.


**My contribution:** I developed and implemented the method with smoothly energy-dependent model parameters, developed and implemented the validation framework, performed the analysis, and wrote the paper.

Reprints were made with permission from the publishers.
Contents

1 Introduction ................................................................................................ 11
  1.1 Background ..................................................................................... 11
  1.2 An overview of the thesis .............................................................. 12
  1.3 Used conventions for notation ....................................................... 14

2 Elementary prerequisites ........................................................................... 16
  2.1 Uncertainty quantification ............................................................. 16
    2.1.1 Error vs. uncertainty ........................................................ 16
    2.1.2 Estimating random and systematic uncertainties .......... 17
    2.1.3 Uncertainty propagation .................................................. 20
  2.2 Fitting models to data .................................................................... 23
    2.2.1 Linear models .................................................................. 24
    2.2.2 Non-linear models ........................................................... 28
    2.2.3 Monte Carlo ..................................................................... 29
  2.3 Nuclear data (ND) .......................................................................... 33
    2.3.1 Cross sections and other quantities of interest ............... 33
    2.3.2 Experimental data ............................................................ 37
    2.3.3 Nuclear reaction models and TALYS ..................... 39
    2.3.4 Evaluated nuclear data .................................................... 42
  2.4 Total Monte Carlo (TMC) and TENDL ........................................ 43

3 Fitting techniques in ND evaluation and ND uncertainty propagation .. 47
  3.1 An illustrating toy model ............................................................... 47
  3.2 An overview of selected techniques .............................................. 48
    3.2.1 Observable domain (OD) vs. parameter domain (PD) .. 48
    3.2.2 GLS ................................................................. 50
    3.2.3 UMC-G ............................................................... 51
    3.2.4 UMC-B ................................................................. 52
    3.2.5 Levenberg-Marquardt / GLS-P ....................................... 53
    3.2.6 A few other names and options ...................................... 53
  3.3 Comparing the performance of the techniques ............................ 54
    3.3.1 Generating realities with synthetic data ......................... 54
    3.3.2 The comparison ........................................................... 55
  3.4 More on UMC-B + TMC in practice ............................................ 57
    3.4.1 Interpretation of experiments .......................................... 58
    3.4.2 Some results from Papers I and II .............................. 59
    3.4.3 An alternative computation of the likelihood ............ 62
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENDF</td>
<td>Evaluated Nuclear Data Format</td>
</tr>
<tr>
<td>ENDF/B</td>
<td>American evaluated nuclear data library</td>
</tr>
<tr>
<td>EXFOR</td>
<td>EXchange FORmat (experiment database)</td>
</tr>
<tr>
<td>GLS</td>
<td>Generalized Least Squares (in the OD)</td>
</tr>
<tr>
<td>GLS-P</td>
<td>Generalized Least Squares (in the PD)</td>
</tr>
<tr>
<td>GP(s)</td>
<td>Gaussian Process(es)</td>
</tr>
<tr>
<td>IAEA</td>
<td>International Atomic Energy Agency</td>
</tr>
<tr>
<td>JEFF</td>
<td>OECD’s evaluated nuclear data library</td>
</tr>
<tr>
<td>LM</td>
<td>Levenberg-Marquardt [algorithm]</td>
</tr>
<tr>
<td>MeV</td>
<td>Mega-electronvolt (energy unit)</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>ND</td>
<td>Nuclear Data</td>
</tr>
<tr>
<td>OD</td>
<td>Observable Domain</td>
</tr>
<tr>
<td>OMP</td>
<td>Optical Model Potential</td>
</tr>
<tr>
<td>PCR</td>
<td>Principal Component Regression</td>
</tr>
<tr>
<td>PD</td>
<td>Parameter Domain</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>TENDL</td>
<td>TALYS-based Evaluated Nuclear Data Library</td>
</tr>
<tr>
<td>TMC</td>
<td>Total Monte Carlo</td>
</tr>
<tr>
<td>UMC-B</td>
<td>Unified Monte Carlo–Breakfast</td>
</tr>
<tr>
<td>UMC-G</td>
<td>Unified Monte Carlo–Garage</td>
</tr>
<tr>
<td>URR</td>
<td>Unresolved Resonance Range</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Background

The IPCC (Intergovernmental Panel on Climate Change, [1]) claims that “without additional mitigation efforts”, the global average temperature will have increased between 2.5 °C and 7.8 °C in year 2100 (compared to pre-industrial levels, and with a confidence level of 90%). The panel also states that this will have serious consequences such as raised water levels, floods, increased stress on water supply and reduced food production, to mention a few examples.

Being a low-carbon energy source [1] and not being intermittent (as wind and solar power), nuclear power can be part of the solution to this problem. What makes nuclear power attractive for energy production is essentially the high energy density connected to fission (and fusion); one typically extracts about $50\text{MWd} \approx 4 \cdot 10^6\text{MJ}$ per kilogram of uranium in nuclear reactors of today, to be compared to the heat content of coal, which is $10 - 30\text{MJ}$ per kilogram [2].

There are problems related to nuclear power, though. The reactors of today produce waste both containing fission products, which are highly radioactive, and heavy elements called actinides, which can be hazardous for more than 100 000 years [3]. This time span is hard to grasp, and it necessitates a complex ethical discussion as well as substantial efforts in waste management. A vision for the “fourth generation” of nuclear reactors (Gen. IV) is to use technologies that suppress the production of the long-lived actinides [4], essentially leaving only the fission products, most of which have decayed into stable nuclides after a few hundreds of years [3].

Also, accidents on nuclear power plants have the potential to be severe, with the worst example being the Chernobyl accident in 1986. Tens of people died shortly after the accident while the total number of premature deaths is hard to estimate and strongly debated. The Chernobyl Forum [assembled by, e.g., several UN organs such as the International Atomic Energy Agency (IAEA) and the World Health Organization (WHO)] writes that it “could mean eventually up to several thousand fatal cancers” among the 600 000 most exposed, and among the other 5 million people which to some extent were exposed, “doses are much lower and [...] expected to make a difference of less than 1% in cancer mortality” [5]. However, all energy sources come with a risk; a study in The Lancet [6] estimates that the mortality (deaths per TWh electricity) is 33, 25, 2.8, 18 and 4.6 for lignite, coal, gas, oil and biomass, respectively, compared to 0.074 deaths/TWh for nuclear energy. These numbers include
accident risk but do not include longterm climate effects. The study also points out that the “access to electricity is prerequisite for the achievement of health, and lack of access to it remains one of the principal barriers to the fulfillment of human potential and well-being”.

Unfortunately, yet another application of nuclear physics is nuclear weapons – the same high energy density that makes nuclear fission attractive for energy production can also be enormously destructive. However, applied nuclear physics also plays a major role in nuclear safeguards [7], i.e., in the system of inspections of nuclear facilities which are intended to ensure that no nuclear material gets dissipated. In other words, an application of nuclear physics is to prevent the proliferation of another application, namely, nuclear weapons. There are several other applications of nuclear physics, such as medicine and astrophysics.

In connection with all these applications, computations or simulations are important ingredients in activities such as design, licensing and maintenance. For such computations, knowledge of so-called nuclear data (ND) is necessary for all involved nuclides. In everyday language, the ND can describe, e.g., interaction probabilities, energy release, and decay times. The ND used in applications, evaluated ND, is a combination of experiments and models (see Sec. 2.3.4). Since there are uncertainties in both experiments and models, there are uncertainties in the evaluated ND which will propagate to the results computed for the applications. It is necessary to assess these propagated uncertainties in order to make correct decisions, both in optimizing for economic purposes and when it comes to safety. To phrase it simply, it would be meaningless to spend time and resources on a calculation not knowing how trustworthy the results are. In the climate change case, this is seized on by the IPCC, who consequently quote confidence intervals and levels of confidence for different claims in their report, just as with the temperature increase cited in the very beginning of this section. This thesis studies ND uncertainties, and their propagation through computations. It particularly aims towards providing methods that can strengthen the foundation of the ND uncertainties in the most comprehensive library of evaluated ND, i.e., TENDL.

1.2 An overview of the thesis

The thesis is based on a set of papers, which are listed on pages v-vi and concatenated at the end of the thesis. Before the collection of papers, there is a comprehensive summary of seven chapters out of which you are currently reading the first. Chapter 2 contains some background knowledge on uncertainty quantification and propagation, model fitting, nuclear data, so-called Total Monte Carlo (TMC) and TENDL, which can be helpful for the remainder of the thesis. In Chapters 3-5, the work described in detail in the different articles is summarized to be more easily accessed, and to make the connection
between the papers clearer. These chapters divide the papers into three groups, even if there is some referencing across the group boundaries. Chapter 6 contains some previously unreported work.

Chapter 3 covers Papers I-III and Paper V. Considering both the chronology and the type of study in Paper V, this paper could fit naturally together with Papers VI-VII. However, Paper V studies several different fitting techniques used in ND evaluation, under quite neat circumstances: the techniques are compared with respect to a quite simple toy problem. The problem of fitting ND to experimental data is essential to this thesis and, therefore, Paper V constitutes a good introduction to the other papers. Papers I-III consider one of the above-mentioned fitting techniques, “UMC-B”, but not entirely using its formalism and framework, and with a focus on propagating uncertainties to applications. Since Papers I-III consider authentic observables, experiments, and applications, they are to a large part devoted to practical details, e.g., on how to interpret experiments automatically, and this makes them more difficult to penetrate than Paper V. Also, a lot of their results are difficult to draw conclusions from, much because of a naïve implementation of UMC-B, giving problems that could have been realized on beforehand, for example by performing a study such as that of Paper V.

Paper IV is summarized in its own chapter, namely, Chapter 4. Inspired by some aspects of Papers I-III, and discouraged by others, Paper IV uses sampling of experimental errors to propagate experimental uncertainties to evaluated ND, as well as to applications. That is, rather than fitting the ND to the experiments, the experiments and their uncertainties are used in the forward direction, to produce a distribution of ND. Just as Papers I-III, this paper works with authentic data, and results in the current evaluation for $^{59}\text{Ni}$ in OECD’s library of evaluated ND, namely, JEFF 3.3 [8, 9].

In Papers I-III and Paper V, as well as by other authors [10, 11], a large problem in the fitting of ND to experimental data is identified: model defects. By this, we mean that the used model cannot exactly reproduce the underlying truth, which leads to biased results with underestimated uncertainties. Chapter 5 summarizes Papers VI-VII, which consider two related ways to treat model defects. In both cases, Gaussian Processes (GPs) are used to give more flexibility in the fit. In Paper VI, a more conventional way of using GPs is applied and thoroughly studied, while another approach, using smoothly input-dependent model parameters, is developed and used in Paper VII. Another difference between the two papers is that Paper VI considers the analysis of certain experiments rather than ND evaluation, which is the aim of Paper VII. Just as Paper V, these two papers make use of synthetic data generated around a large set of known “truths” to enable studies of how the results are distributed in comparison to the sought truth, but with an increasing complexity of the problems. Even if there is only synthetic data in Paper VI, Chapter 5 also contains previously unpublished results where the method of Paper VI is applied to authentic data.
As just mentioned, Paper VII studies the performance of its methodology on synthetic data. Also, the used model is not an authentic nuclear reaction code; it is a fast executing model which is based on TALYS [12], and intended to include much of the complexity of TALYS from a fitting point of view. In Chapter 6, the method of Paper VII is further developed and applied to authentic data, and using authentic TALYS. Before using the data in the method, substantial preprocessing is performed. It is also shown that since the model is linearly interpolated, the large set of data can be reduced to a much smaller set without losing any additional information. Further, the chapter includes a preliminary implementation of a method which automatically assigns additional uncertainties to data sets based on their agreement to other sets.

Finally, some general concluding remarks are found in Chapter 7.

For formal reasons, it should be noted that Secs. 1.1, 2.1, 2.3.1-2, 2.4, and 3.4.1-2 are based on the author’s licentiate\(^1\) thesis [13].

1.3 Used conventions for notation

To simplify the reading of the thesis, the following conventions for notation is followed, as much as reasonably achievable:

- Scalars are denoted in italics, e.g., \(x, \beta\).
- Upright and boldface font with serifs (when applicable) is used for vectors, e.g., \(y, Y, \beta\). A few special cases of (short) vectors are denoted as scalars, if vector properties are not exploited.
- Capital letters in upright and boldface font without serifs are used for matrices, e.g., \(X, S, \Omega\).
- All vectors (that aren’t transposed) are column vectors. Vectors are often printed as transposed row vectors, e.g., \(x = (x_1, x_2, \ldots, x_n)^T\), where \(T\) stands for transpose.
- Random variables (or vectors) are most often denoted by capital letters, and a realization of the random variable is the corresponding lowercase letter, e.g., \(x_i\) is a realization of \(X_i\), and \(y\) is a realization of \(Y\). Sometimes a parameter is considered random in some instances, and otherwise not, making this convention somewhat diffuse.
- Probability density functions (PDFs) are most often denoted \(p_X(x)\), where \(X\) is the considered random variable and \(x\) the argument. Conditioned PDFs are therefore normally written as, e.g., \(p_{X|Y}(x)\). In many of the papers, \(f\) takes the place of \(p\).
- Integrals without a specified domain are taken over the whole domain of the integrand, typically \(\mathbb{R}^n\).
- The expected variable of a random variable \(X\) is denoted \(\langle X \rangle\).

\(^1\)In Sweden, the so-called licentiate degree can be achieved approximately halfway through PhD studies.
• The variance of a univariate random variable $X$ is denoted $V(X)$, and the covariance of a random vector $\mathbf{X}$ is denoted $\mathbf{V}(\mathbf{X})$.
• The element on the $i$th row and $j$th column of the matrix $\mathbf{A}$ is denoted $(\mathbf{A})_{ij}$.
• Empty entries in matrices represent appropriately sized blocks of zeros. The above holds for the comprehensive summary but the same conventions apply, to a large degree, also to the papers.
2. Elementary prerequisites

This chapter is intended to introduce the reader to a few concepts which are helpful for the remainder of the thesis. In Sec. 2.1, the terms **error** and **uncertainty** are defined, followed by a short discussion on the basics in the estimation and propagation of uncertainties. Sec. 2.2 contains a few aspects of fitting models to data, which forms the foundation of papers V-VII, and it is also a central topic in Papers I-III. Sec. 2.3 gives an introduction to nuclear data, which is the application of all papers. Finally, Sec. 2.4 gives an introduction to Total Monte Carlo (TMC), an ND propagation methodology which has a certain impact on the thesis, in particular on Papers I-III. The evaluated ND library TENDL is related to TMC and is also introduced in Sec. 2.4.

2.1 Uncertainty quantification

2.1.1 Error vs. uncertainty

The words error and uncertainty are used interchangeably by many authors and experimenters. However, this thesis follows the convention of Ref. [14], i.e., to distinguish between the two. The author believes that this is helpful for the understanding.

To shed light on the difference, consider the conceptually simple situation where the distance between two points on a piece of paper is measured with a steel ruler. There is a true distance \( a \), but it is unknown. Using the ruler, we estimate the distance to \( x \). This is our best guess for \( a \), but we are almost sure that the error \( x - a \) is non-zero\(^1\). It is in the nature of the error that it is unknown; if it were not, it should have been corrected for. However, based on the measurement technique, one can estimate the **uncertainty** of \( x \). In the example of the ruler, a natural limitation is the accuracy of the ruler’s scale; if the ruler shows millimeters, a careful reading could perhaps provide guesses for tenths of millimeters, but there would be a large risk to confuse 0.3 mm with 0.4 mm, for example, which gives an idea about the uncertainty.

The error \( \varepsilon \) can be modeled as a random variable, i.e., such that the estimated distance \( x \) is an observation of the random variable

\[
X = a + \varepsilon,
\]

\(^1\) Almost sure is actually a mathematically well-defined term for an event that has probability 1, and this holds for the event that \( X - a \neq 0 \) (\( x \) is an observation of \( X \)), because distances are continuous quantities.
and define uncertainty as the dispersion of the distribution for $\varepsilon$. The dispersion is often quantified by the standard deviation $\sigma(\varepsilon)$, which is the square root of the variance $V(\varepsilon)$, i.e.,

$$\sigma^2(\varepsilon) = V(\varepsilon) = \int_{-\infty}^{\infty} (\varepsilon - \langle \varepsilon \rangle)^2 p_{\varepsilon}(\varepsilon) \, d\varepsilon,$$

(2.2)

where $p_{\varepsilon}(\varepsilon)$ is the probability density function (PDF) for $\varepsilon$, and $\langle \varepsilon \rangle$ is the expected value of $\varepsilon$, i.e.,

$$\langle \varepsilon \rangle = \int_{-\infty}^{\infty} \varepsilon p_{\varepsilon}(\varepsilon) \, d\varepsilon,$$

(2.3)

which can be interpreted as the mean value one would approach when drawing an increasingly large sample from random variables distributed as $\varepsilon$, or the “center of mass” of the distribution.

Note that the standard deviation only captures one particular aspect of the distribution, see for example Fig. 2.1, where two PDFs (for some random variable $Z$) with the same standard deviation (and expected value) are shown – yet the probability for, e.g., $Z > 0.7$ is quite different for the two distributions. One of the functions follows a normal distribution (a.k.a. Gaussian), a distribution which is often assumed, mainly because it can be justified by the so-called Central Limit Theorem in cases where the considered random variable can be written as the sum of “many” (more or less independent) random variables (which is often the case to some approximation) [15]. Following the principle of maximum entropy [16, 17], one can also show that the least informed choice of distribution is the normal distribution. For the normal distribution, $\sim 68\%$ of the probability mass lies within $\pm$ one standard deviation from the expected value [15]. Despite the above-mentioned limitation of the standard deviation, it is frequently used to quantify the uncertainty, also in this thesis. Alternatives could be to quote confidence intervals (in practice, this is normally just the expected value $\pm$ the standard deviation scaled by a so-called quantile, a number which relies on an assumption of a distribution, and often a normal ditto) or tolerance intervals using order statistics along the line of Wilks’ method [18], or to estimate the full error distribution.

2.1.2 Estimating random and systematic uncertainties

Ideally, one would know the distribution for $\varepsilon$, but this is in practice not the case. In the example with the ruler above one could maybe estimate the uncertainty in the reading with a standard deviation of 0.1 mm, and then one could possibly add uncertainties such as the manufacturing tolerance, and uncertainty due to a possibly unknown temperature for which the ruler was intended

\[2\text{The least informed choice need not be the most conservative.}\]
(and if such a temperature is known, the thermometer has an uncertainty, too, as well as the expansion coefficient for the steel). However, how did the ruler or thermometer manufacturers determine their manufacturing uncertainties in the first place? Moreover, could we do a more rigorous determination of the reading uncertainty?

As long as uncertainties are random (as opposed to systematic), they are in principle simple to estimate, provided that it is possible to repeat the measurement. By random, we mean that if the measurement is repeated, the error will fluctuate randomly with an expected value of zero. If so, it is possible to obtain a random sample \( \mathbf{x} = (x_1, x_2, ..., x_n)^T \) with observations from \( \mathbf{X} = (X_1, X_2, ..., X_n)^T \), where all the \( X_i \) are independent and distributed as \( X \). Then, one can estimate the distribution of \( X \) in Eq. (2.1), or certain properties of this distribution. In particular, it is possible to estimate the standard deviation of \( X \) (which is the same as the standard deviation of \( \varepsilon \)) using the sample standard deviation:

\[
s(\mathbf{x}) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2},
\]

Figure 2.1. Two probability density functions with the same expected value and variance.

\(^3\)It is not always necessary to use the sample standard deviation; an example of the opposite is seen in Sec. 2.3.2.
where \( \bar{x} \) is the mean value of \( x \), i.e.,

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]  

(2.5)

In the example with the ruler, one could let many different people redo the measurement to be able to estimate the uncertainty using Eq. (2.4). Naturally, one would estimate the distance using \( \bar{x} \), rather than just the first observation. The mean value \( \bar{x} \) is an observation of

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,
\]  

(2.6)

which has standard deviation \[15\]

\[
\sigma(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} V(X_i)} = \frac{\sigma(X)}{\sqrt{n}},
\]  

(2.7)

assuming that the \( X_i \) are mutually independent, i.e., that their uncertainties are random only. Thus, by repeating the measurement \( n \) times, we do not only obtain an uncertainty estimate; by using \( \bar{x} \), we also reduce the (random) uncertainty by a factor of \( 1/\sqrt{n} \).

Systematic uncertainties are worse, in both respects. They are not as easy to estimate, and they cannot be reduced by repeating the same measurement, since they are not random, and therefore, one will repeat the same error. In general, it is desired to “make” systematic uncertainties random, by varying more parameters in the measurement setup. In the example of the ruler, there may be an uncertainty in the manufacturing of the ruler which will be manifested as a systematic uncertainty, which one can estimate by using many different rulers (if the manufacturer gives an uncertainty estimate, they may have done something similar). However, the manufacturing uncertainty may also contain systematic components, so it could make sense to use rulers from different manufacturers, or even different measurement techniques, and so on. Such a process can obviously be expensive, and it may be necessary to use experience from similar measurements or somehow derive an uncertainty estimate from underlying principles, a rough example of which is the estimate of 0.1 mm in the very beginning of this section. In Ref. [14], such uncertainty estimates are discussed in more detail (referred to as “Type B” uncertainties).

It should be stressed that even though it may be hard to imagine observing a distribution for systematic errors, they should also be regarded as random variables. The problem is that it is often difficult to obtain multiple observations from its distribution.

A particular concern arises if the same systematic uncertainty occurs in different measurements: their errors will be \textit{correlated}. This fact will have
an important impact when uncertainties are propagated, see Sec. 2.1.3. As a simple example, assume that we want to estimate the sum $a + b$, where both quantities are measured separately. If $X = a + \varepsilon$ and $Y = b + \varepsilon$ (they have a common error $\varepsilon$), then the variance of their sum becomes [15]

$$\text{Var}(X + Y) = \text{Var}(2\varepsilon) = 4\text{Var}(\varepsilon).$$

(2.8)

This can be compared to the case when $X = a + \varepsilon_1$ and $Y = b + \varepsilon_2$, where $\varepsilon_1$ and $\varepsilon_2$ are uncorrelated with the same variance as $\varepsilon$. This results in [15]

$$\text{Var}(X + Y) = \text{Var}(\varepsilon_1) + \text{Var}(\varepsilon_2) = 2\text{Var}(\varepsilon).$$

(2.9)

2.1.3 Uncertainty propagation

When performing a computation or simulation which uses input containing uncertainties, it is of interest how the uncertainties of the input propagate to the results. Modeling the input as a random vector $X = (X_1, X_2, ..., X_m)^T$, a resulting quantity $g(X)$ is a random variable which follows a distribution depending on $X$. The function $g(X)$ could, for example, give the dose rate absorbed by a person at certain position, or the decay heat from radioactive waste, provided certain input information $X$. This section briefly describes two standard methods to estimate the uncertainty in $g(X)$ given information on $X$: linear and Monte Carlo uncertainty propagation. Fig. 2.2 provides a synoptic comparison of the two methods.

**Linear uncertainty propagation**

If the uncertainties are “small enough”, it can be reasonable to keep only the zeroth and the first orders in a Taylor expansion of $g(X)$ about the expected value of $X$, which yields (assuming $g(X)$ has continuous first partial derivatives on its domain) [15]

$$\text{Var}(g(X)) \approx \sum_{i=1}^{m} \sum_{j=1}^{m} \left( \frac{\partial g}{\partial x_i} \right) \left( \frac{\partial g}{\partial x_j} \right) \bigg|_{x = \langle X \rangle} \text{C}(X_i, X_j),$$

(2.10)

where $\text{C}(X_i, X_j)$ is the covariance between $X_i$ and $X_j$, defined as

$$\text{C}(X_i, X_j) = \int_{\mathbb{R}^m} (x_i - \langle X_i \rangle)(x_j - \langle X_j \rangle) p_X(x) \, dx,$$

(2.11)

where $p_X(x)$ is the PDF for $X$. The covariance describes the uncertainty of $X_i$ and $X_j$ but also their correlation – the covariance can be written

$$\text{C}(X_i, X_j) = \sigma(X_i)\sigma(Y_i)\rho(X_i, Y_i),$$

(2.12)

where $\rho$ is the correlation coefficient, which can be shown to satisfy $-1 \leq \rho \leq 1$ [15]. The closer $|\rho|$ is to 1, the greater is the linear dependence, and
Figure 2.2. A comparison of linear (top) and Monte Carlo (bottom) uncertainty propagation, for a univariate case. Before the propagation, the distribution is Gaussian in both cases. After propagation, the distributions are similar close to the mean, but the Monte Carlo propagation gives thicker tails. The distortion originates from the uncertainty being too large for the linear propagation; the tails of the input distribution are outside the region which is well approximated by a linearized model.

$\rho$ is positive if $X_i$ and $X_j$ tend to “vary in the same direction” (covary) and negative in the opposite case.

Note that $C(X_i, X_i) = V(X_i)$, giving that if all different random variables are uncorrelated, Eq. (2.10) simplifies to

$$V(g(X)) \approx \sum_{i=1}^{m} \left( \frac{\partial g}{\partial x_i} \right)^2 \bigg|_{x=\langle X \rangle} V(X_i). \quad (2.13)$$

Quite intuitively, the uncertainty of $g(X)$ thus depends on the uncertainty of the different arguments $X_i$ and on how strongly $g$ depends on the arguments, cf., the upper half of Fig. 2.2. Eq. (2.10) can be interpreted similarly, but it also takes the correlation into account.

Eq. (2.10) can be generalized to the case of multiple output quantities $g(X) = (g_1(X), g_2(X), ..., g_l(X))^T$, and it can be neatly written in matrix form as

$$C_g \approx S^T C_X S, \quad (2.14)$$

where $C_g$ and $C_X$ are the covariance matrices of $g$ and $X$, respectively, i.e., $C(X_i, X_j)$ is found on the $i$'th row in the $j$'th column of $C_X$, and analogously for $C_g$, and the element on the $i$'th row in the $k$'th column of the so-called
The sensitivity matrix $\mathbf{S}$ is

$$
(S)_{ik} = \frac{\partial g_k}{\partial x_i},
$$

(2.15)

The (approximate) variances of the different random variables $g_1(\mathbf{X})$, $g_2(\mathbf{X})$, ..., $g_1(\mathbf{X})$ are found along the diagonal of $\mathbf{C}_g$. Their covariances may, e.g., be used in further propagation.

Eqs. (2.14) and (2.15) have made up the backbone of ND uncertainty propagation since the work of Usachev [19] in the early 1960’s. Hence, the words covariance matrices and sensitivities are frequently occurring in the field of ND uncertainty propagation. The estimation of the sensitivity matrix $\mathbf{S}$ can be performed by perturbing each one of the input variables $x_i$ at a time (simple numerical differentiation). If the number of uncertain inputs is large and the number of outputs of interest is less, one can consider the so-called adjoint method [20] to obtain the sensitivity matrix. However, this is an intrusive method; i.e., internal properties of $g$ (which can be a complicated computer code) must be used.

The main downside of this type of error propagation is that it is an approximation; nonlinear dependence on the input is not captured. For example, at a minimum or maximum of $g(\mathbf{x})$ the first derivatives are zero, leading to a zero estimate of the uncertainty. It is possible to include more terms in the Taylor expansion, but this would make it necessary to estimate higher order derivatives [21]. Another disadvantage is that the methodology characterizes the full distribution of $\mathbf{X}$ by its expected value and covariance$^4$. As a univariate example, the methodology would not differ between the two distributions in Fig. 2.1.

**Monte Carlo uncertainty propagation**

A more direct way to estimate the propagated uncertainty is to generate a random sample $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(n)}$ from $\mathbf{X}$, and to simply evaluate $g$ for all observations of this random sample. In this way, one obtains a random sample from $g(\mathbf{X})$ [15], namely $g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(2)}), ..., g(\mathbf{x}^{(n)})$. It is then straightforward to estimate the standard deviation of $g(\mathbf{X})$ using Eq. (2.4), and it is also possible to estimate the full distribution of $g(\mathbf{X})$, or, say, $P(g(\mathbf{X}) > c)$, i.e., the probability that $g(\mathbf{X})$ exceeds some particular value $c$ (a design criterion may require this probability to be small). It is also possible to estimate the uncertainty in these quantifications of the uncertainty, e.g., as outlined in Ref. [22].

Except the possibility to study more aspects of the distribution of $g(\mathbf{X})$ than its mean and standard deviation, other advantages compared to the linear uncertainty propagation are that non-linearities can be captured (the only approximation in this methodology is the finite number of samples) and that

---

$^4$Higher order moments (the expected value and variance are the two lowest orders) can also be included by including more terms in the Taylor expansion, but the number of necessary higher order derivatives to include would grow even larger [21].
other aspects of the distribution of \( X \) than its expected value and covariance can be taken into account.

The major disadvantage is that more evaluations of \( g(X) \) typically are necessary, which becomes a problem if \( g(X) \) is computationally expensive to evaluate (which often can be considered the case in applied nuclear physics). It may be argued that it is poorly invested time to do such an accurate uncertainty propagation since the knowledge of the distribution of \( X \) may be a larger limitation than the approximation in linear uncertainty propagation. Indeed (in ND uncertainty propagation), it is often the case that the only information on \( X \) consists of (estimates of) the expected value and covariance, and then one will have to assume a distribution. Normally, this will be a (multivariate) normal distribution because of the reasons pointed out at the end of Sec. 2.1.1. It is worth noting that the number of evaluations of \( g \) does not depend on the length of \( X \) [23], while the number of derivatives to be estimated for the sensitivity matrix \( S \) in Eq. (2.14) depends on how many quantities one believes to have an impact on the uncertainty. If propagating, e.g., cross section uncertainties with linear uncertainty propagation, this means that either the adjoint method must be implemented or that the cross sections must be relatively coarsely grouped in the determination of the sensitivity matrix, or otherwise a large number of “function evaluations” (i.e., computer code runs) are necessary.

There are several variance reduction techniques that can be considered in order to reduce the necessary sample size, e.g., Latin Hypercube Sampling [24]. The implementation is not as simple, and the results from these methods are a little less straightforward to interpret; for example, the variance estimator used in Latin Hypercube Sampling has an unknown bias, even if this bias is small [25].

### 2.2 Fitting models to data

Much of the work in this thesis concerns the fitting of physical models to experimental data, both using deterministic and stochastic (Monte Carlo) methods. The fitted models are non-linear, but the used deterministic methods are founded upon the fitting of linear models. Therefore, the basics of fitting linear models are presented in Sec. 2.2.1; in fact, the maximum likelihood estimator of the model parameters is derived, as well as its covariance, because the author often finds that such derivations in the literature are unnecessarily complicated. The resulting estimators are found in Eqs. (2.26) and (2.27) without a prior, and generalized to include a prior in Eqs. (2.33) and (2.34). The linear model fitting is followed by a few comments on the fitting of non-linear models in Sec. 2.2.2, and the basics of Monte Carlo fitting are presented in Sec. 2.2.3.
The general situation considered here is that we observe a data vector \( y = (y_1, y_2, ..., y_n)^T \), which we view as an observation of the random vector \( Y = (Y_1, Y_2, ..., Y_n)^T \). Corresponding to each point in \( Y \), there is a set of inputs \( x = (x_1, x_2, ..., x_n)^T \), which we assume to be perfectly known. The data \( y \) can represent a neutron cross section (see Sec. 2.3) and \( x \) can contain different neutron energies, for example. We assume that \( f(x; \beta) \) is a function of \( x \) with parameters \( \beta = (\beta_1, \beta_2, ..., \beta_l)^T \), such that

\[
Y = f(x; \beta) + \epsilon, \tag{2.16}
\]

where \( f(x; \beta) = (f(x_1; \beta), f(x_2; \beta), ..., f(x_n; \beta))^T \), and \( \epsilon \) is a random vector describing the error in the measurement (the function \( f \) can, e.g., be a complicated model such as TALYS, see Sec. 2.3.3). In other words, the difference between \( y \) and the truth underlying the data is the measurement error \( \epsilon \), which is a realization of \( \epsilon \). We typically assume that \( \epsilon \) is normally distributed around 0 with a covariance matrix \( \Omega \), which we write as \( \epsilon \sim N(0, \Omega) \), but this is not always necessary.

The general goal is to obtain a PDF describing where the true parameters \( \beta \) are, posterior to observing the data. Often, we are satisfied with estimating the expected value of \( \beta \) and the covariance matrix associated to the estimate, i.e., the best estimate for \( \beta \), how uncertain this estimate is, and how correlated the estimates of different parameters are. Note that when we assume a normal distribution for \( \beta \), the distribution is fully characterized by the expected value and covariance matrix.

### 2.2.1 Linear models

In this context, a linear model is linear with respect to its parameters. If the model is linear, Eq. (2.16) can be reformulated as

\[
Y = X\beta + \epsilon, \tag{2.17}
\]

where \( X \) is the design matrix, which includes information on both the model and the setup of the “experiment”. For example, if the model \( f \) is a quadratic polynomial of an input variable \( x \),

\[
f(x; \beta) = \beta_1 + \beta_2 x + \beta_3 x^2, \tag{2.18}
\]

and \( y \) is observed for \( x \)-values \( x = (x_1, x_2, ..., x_n)^T \), the design matrix is

\[
X = \begin{pmatrix}
1 & x_1 & x_1^2 \\
1 & x_2 & x_2^2 \\
\vdots & \vdots & \vdots \\
1 & x_n & x_n^2
\end{pmatrix}. \tag{2.19}
\]
To really make this concrete, if \( x = (0, 0.1, 1, 2) \), the design matrix of Eq. (2.19) becomes

\[
X = \begin{pmatrix}
1 & 0 & 0 \\
1 & 0.1 & 0.01 \\
1 & 1 & 1 \\
1 & 2 & 4
\end{pmatrix}.
\] (2.20)

More generally, a general linear model of one input variable \( x \) can be written

\[
f(x; \beta) = \beta_1 g_1(x) + \beta_2 g_2(x) + \ldots + \beta_l g_l(x).
\] (2.21)

Then, we have

\[
X = \begin{pmatrix}
g_1(x_1) & g_2(x_1) & \cdots & g_l(x_1) \\
g_1(x_2) & g_2(x_2) & \cdots & g_l(x_2) \\
\vdots & \vdots & \ddots & \vdots \\
g_1(x_n) & g_2(x_n) & \cdots & g_l(x_n)
\end{pmatrix}.
\] (2.22)

It is straightforward to generalize this to linear models of more than one variable, by letting \( x \) be a vector.

Assume that \( \mathcal{E} \) follows a (multivariate) normal distribution centered around zero with the covariance matrix \( \Omega \), which we write as \( \mathcal{E} \sim \mathcal{N}(0, \Omega) \). Motivations for this assumption are summarized in Sec. 2.1.1 (the discussion holds for multivariate normal distributions, too). Assume that the covariance matrix is positive definite\(^5\). Per definition, the maximum-likelihood (ML) estimate of \( \beta \) is the choice of \( \beta \) that maximizes, \( p_{\beta|y}(\beta) \), the probability density function (PDF) for \( \beta \), given the observed data \( y \). If there is no prior distribution for \( \beta \), \( p_{\beta|y}(\beta) \) is proportional to the PDF for \( Y \) given \( \beta \) [26]. Using the PDF of a normal distribution [27]

\[
p_{\beta|y}(\beta) \propto p_Y(\beta) \propto e^{-\frac{1}{2}(X\beta - y)^T \Omega^{-1} (X\beta - y)}.
\] (2.23)

Maximizing this is equivalent to minimizing the “generalized chi-squared”

\[
\chi^2(\beta) = (X\beta - y)^T \Omega^{-1} (X\beta - y).
\] (2.24)

Taking the gradient (“differentiating”) with respect to \( \beta \) gives [28]

\[
\nabla_\beta \chi^2(\beta) = \nabla_\beta [\beta^T X^T \Omega^{-1} X\beta - \beta^T X^T \Omega^{-1} y - y^T \Omega^{-1} X\beta + y^T \Omega^{-1} y]
\]

\[
= 2X^T \Omega^{-1} X\beta - 2X^T \Omega^{-1} y,
\] (2.25)

\(^5\Omega \) must be non-negative definite, and if no elements of \( \mathcal{E} \) are fully correlated, it is positive definite. In the applications in this thesis, this can always be assumed.
where it was used that the transpose of a scalar equals itself. This is zero if and only if $\beta = \hat{\beta}(y)$, where

$$\hat{\beta}(y) = \left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}y. \quad (2.26)$$

The Hessian matrix (“the second derivative”) is found by taking the derivative of Eq. (2.25) with respect to $\beta^T$, giving $X^T\Omega^{-1}X$ [28]. If $X$ has full rank, this is positive definite [28], so this must be a minimum [29]. Since it is the only zero of the gradient, it must be the global minimum. Thus, $\hat{\beta}$ is the ML estimate of $\beta$.

Taking the covariance of the corresponding estimator, $\hat{\beta}(Y)$, gives

$$V(\hat{\beta}) = \left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}V(Y)\left[\left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}\right]^T$$

$$= \left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}X\left(X^T\Omega^{-1}X\right)^{-1} = \left(X^T\Omega^{-1}X\right)^{-1}, \quad (2.27)$$

using the rule for the covariance of a matrix product with a random vector [28], and basic rules for matrix operations. The corresponding observable $\hat{y}$ is estimated by inserting $\hat{\beta}$ into the model and setting the error to 0, giving

$$\hat{y} = X\hat{\beta}, \quad (2.28)$$

with covariance matrix (again using the rule from [28])

$$V(\hat{y}) = XV(\hat{\beta})X^T. \quad (2.29)$$

It can be shown [30] that $\hat{\beta}$ is the most efficient unbiased linear estimator of $\beta$. Further, Eq. (2.27) gives the exact covariance given the assumptions of the model. Thus, if the model is linear with Gaussian errors, there is a closed form estimate of the parameters with good properties, and the uncertainties and correlations are easily estimated. The computations involve solving matrix equations, however. If the number of data points is large, this becomes time (and memory) consuming, and there is a risk of numerical problems if the errors are highly correlated (i.e., if $\Omega$ is ill-conditioned). Moreover, the assumptions of the model are often too restrictive.

A standard way to check if the model is reasonable is to make use of the fact that given the model, $\chi^2(\hat{\beta})$ [see Eq. (2.24)] follows a $\chi^2$-distribution with $n - l$ degrees of freedom (the number of data points minus the number of

\[\text{If } X \text{ does not have full rank, we cannot solve for } \hat{\beta}. \text{ If there are more data points than parameters, full rank means that the columns of } X \text{ are linearly independent. If they are not, two parameters are indistinguishable. It would happen, e.g., if one tried to fit the model } f(x, \beta) = \beta_1 + \beta_2x + \beta_3x, \text{ i.e., } \beta_2 \text{ and } \beta_3 \text{ could be replaced by one single parameter.}\]
fitted parameters) [27]. Thus, a $\chi^2(\hat{\beta})$ deviating “too much” from the expected value of such a distribution (i.e., $n - l$) indicates that something is wrong: the model or the data covariance matrix $\Omega$. The opposite, that $\chi^2(\hat{\beta})$ is “close to” $n - l$, does not guarantee that the model is correct, however.

**With a prior distribution**

Now assume that there is prior information about the parameters which we also can model with a normal distribution, i.e., we know *before* observing the data $y$ that $\beta \sim \mathcal{N}(p, P)$. Formally, we can see the prior information as additional data. The prior parameters $p$ are appended to the data $y$, the prior covariance matrix $P$ is appended to the lower right corner of $\Omega$ (making it block-diagonal), and an identity matrix of appropriate size $I$ is appended to the bottom of the design matrix $X$, giving the corresponding matrices

$$y_\dagger = \begin{pmatrix} y \\ p \end{pmatrix}, \quad \Omega_\dagger = \begin{pmatrix} \Omega & 0 \\ 0 & P \end{pmatrix}, \quad \text{and} \quad X_\dagger = \begin{pmatrix} X \\ I \end{pmatrix},$$

(2.30)

where it is assumed that the data are independent of the prior parameters (the prior parameters $p$ must not have been fitted to any data correlated to $y$). Inserting these vectors and matrices in place of their “non-daggered” counterparts in Eq. (2.26) gives

$$\hat{\beta} = \left( X^T I \right) \left( \Omega^{-1} 0 \begin{pmatrix} 0 \\ P^{-1} \end{pmatrix} \right) \left( \begin{pmatrix} X \\ I \end{pmatrix} \Omega^{-1} 0 \begin{pmatrix} 0 \\ P^{-1} \end{pmatrix} \right) \begin{pmatrix} y \\ p \end{pmatrix}$$

(2.31)

using block-wise inversion of block-diagonal matrices [28]. This can be rewritten by first multiplying both sides with $X^T \Omega^{-1} X + P^{-1}$ from the left, giving

$$(X^T \Omega^{-1} X + P^{-1}) \hat{\beta} = X^T \Omega^{-1} y + P^{-1} p$$

(2.32)

$$= X^T \Omega^{-1} y + \left( X^T \Omega^{-1} X + P^{-1} \right) p - X^T \Omega^{-1} X p$$

$$= \left( X^T \Omega^{-1} X + P^{-1} \right) p + X^T \Omega^{-1} (y - X p).$$

Solving for $\hat{\beta}$ by multiplying with $\left( X^T \Omega^{-1} X + P^{-1} \right)^{-1}$ from the left gives

$$\hat{\beta} = p + \left( P^{-1} + X^T \Omega^{-1} X \right)^{-1} X^T \Omega^{-1} (y - X p).$$

(2.33)

This expression may be more pedagogical than Eq. (2.31), since one can see that $\hat{\beta}$ deviates from $p$ in proportion to the deviation between the data and the prior expectation of the data $X p$. The covariance for $\hat{\beta}$ with a prior is obtained
in complete analogy with Eq. (2.27), or as a byproduct of Eq. (2.31):

\[ V(\hat{\beta}) = \left( P^{-1} + X^T \Omega^{-1} X \right)^{-1}. \]  

(2.34)

If there is no prior information, one can view this as if the diagonal entries of \( P \) tend to infinity, giving that \( P^{-1} \to 0 \), and the two occurrences of \( p \) in Eq. (2.33) can be seen to cancel out. Thus, Eq. (2.26) and Eq. (2.27) are retained.

Using the view of the prior as additional data, the \( \chi^2 \)-test briefly described above can be used on Eq. (2.24) with \( y, \Omega, \) and \( X \) replaced by its daggered counterparts from Eq. (2.30). Using the block-diagonality of \( \Omega^\dagger \), the result becomes \( \chi^2_\dagger(\beta) = \chi^2(\beta) + \chi_p^2(\beta) \), where \( \chi^2(\beta) \) is found in Eq. (2.24), and

\[ \chi_p^2(\beta) = (\beta - p)^T P^{-1} (\beta - p). \]  

(2.35)

The sum \( \chi^2(\beta) + \chi_p^2(\beta) \) follows, under the assumptions of the model, a \( \chi^2 \)-distribution with \( n + l - l = n \) degrees of freedom, since we have \( n + l \) “data points” including the prior information.

### 2.2.2 Non-linear models

The condition for the model to be linear is often too restrictive, as for the nuclear reaction models considered in this thesis. In such cases, one may write the model as (again considering one input variable \( x \))

\[ Y = f(x; \beta) + \varepsilon, \]  

(2.36)

where \( f(x; \beta) = (f(x_1; \beta), f(x_n; \beta), ..., f(x_n; \beta))^T \). It may still be adequate to model the errors to follow a normal distribution, i.e., \( \varepsilon \sim \mathcal{N}(0, \Omega) \). If so, it is still desirable to minimize \( \chi^2(\beta) \), but with \( X\beta \) replaced by \( f(x; \beta) \):

\[ \chi^2(\beta) = (f(x; \beta) - y)^T \Omega^{-1} (f(x; \beta) - y). \]  

(2.37)

Unfortunately, this minimization is not as easy with a non-linear model. Most deterministic methods rely on linearizing the model, giving

\[ f(x; \beta) = f(x; \beta^{(0)}) + J \cdot (\beta - \beta^{(0)}), \]  

(2.38)

where \( \beta^{(0)} \) is some guess for \( \beta \), and \( J \) is the Jacobian matrix, containing all the derivatives of \( f(x; \beta) \) with respect to \( \beta \). Using this linearization, Eq. (2.36) can be written in the form of Eq. (2.17), and the basic ideas of linear regression can be applied to iteratively refine the estimate of \( \beta \). In practice, it is necessary to take measures to avoid diverging solutions, e.g., using the Levenberg-Marquardt algorithm [31, 32, 33], which is accessibly presented in Paper VI, and used in Papers V-VII. The algorithm takes more careful steps than a direct implementation; how careful the steps are varies between iterations, depending on whether the \( \chi^2 \) is improved or not. The covariance can
be estimated based on the final linearization. The method converges, under reasonable conditions, to a local minimum \[33\]. However, this minimum need not be the global minimum; it may depend on the starting guess. To be strict, one should consider the possibility of several local minima, and (hopefully) find them using different starting guesses. If several minima with significant likelihood is found, they should ideally be combined to yield a better estimate of the distribution of $\beta$ \[26\].

2.2.3 Monte Carlo

Monte Carlo techniques can be an alternative route for fitting non-linear models. They do not rely on linearizing the model, and have the advantage that any prior distribution can be assumed for $\beta$, and any distribution can be assumed for the experiments $Y$ given $\beta$. Further, even if normal distributions are assumed for these two parts, the posterior distribution may still become non-normal, due to non-linearities in the model. The main downside of Monte Carlo methods is that they can demand a large number of function evaluations, which can lead to infeasible computational times.

**Brute force Monte Carlo assuming a prior**

The perhaps simplest form of Monte Carlo fitting assumes that there is a prior distribution $p_\beta(\beta)$ for the model parameters, from which we can generate a sample $\beta^{(1)}, \beta^{(2)}, ..., \beta^{(N)}$, where $N$ is the sample size. For each of these, we can evaluate the model, giving $f(x; \beta^{(1)}), f(x; \beta^{(2)}), ..., f(x; \beta^{(N)})$. To estimate the posterior distribution of the parameters, or of any quantity derived from the parameters, given the data $y$, the realizations $\beta^{(1)}, \beta^{(2)}, ..., \beta^{(N)}$ can be provided with weights $w_j$ proportional to the likelihood $p_{Y|\beta^{(j)}}(y)$ for $j \in \{1, 2, ..., N\}$. As the notation indicates, the likelihood is the PDF of the data $y$ given the parameters $\beta$.

The motivation for this use of weights follows below. First, let us just note that if Gaussian data covariances are assumed, the likelihood is obtained using the PDF of a multivariate normal distribution \[27\]:

$$p_{Y|\beta}(y) = \frac{e^{-\frac{1}{2} (f(x; \beta) - y)^T \Omega^{-1} (f(x; \beta) - y)}}{(2\pi)^{n/2} \sqrt{\det(\Omega)}} \propto e^{-\chi^2(\beta)/2}, \quad (2.39)$$

where $\chi^2(\beta)$ is defined in Eq. (2.37). Since only $\chi^2(\beta)$ depends on $\beta$, the weights can be set to $w_j = e^{-\chi^2(\beta^{(j)})/2}$.

Paper II includes an unnecessarily complicated motivation for this weighting, and here follows a simpler explanation. We are interested in aspects of the distribution of $\beta$ that can be obtained using integrals such as

$$\langle Q \rangle_{\beta|y} = \int Q(\beta) p_{\beta|y}(\beta) \, d\beta, \quad (2.40)$$

29
where $Q(\beta)$ can be a matrix-, vector-, or scalar-valued function of $\beta$ (the latter two are special cases of matrix valued functions). For example, $Q$ can be $f = f(x; \beta)$ or the outer vector product $ff^T$ (which is a matrix). Using these two, we can obtain the expected value and covariance of $f$, given the data, i.e.,

$$\langle f \rangle_{\beta|y} = \langle ff^T \rangle_{\beta|y} - \langle f \rangle_{\beta|y} \langle f \rangle_{\beta|y}^T .$$

(2.41)

We may also be directly interested in some integral quantity. As an example, the neutron multiplication factor $k_{\text{eff}}$ is considered in Papers I-III, which can be obtained from first computing cross sections (see Sec. 2.3) using model parameters in TALYS (see Sec. 2.3.3), followed by using these cross sections in a neutron transport code such as MCNP [34]. In such a case, we may be interested in the expected value and variance of $k_{\text{eff}} = k_{\text{eff}}(\beta)$, which are

$$\langle k_{\text{eff}} \rangle_{\beta|y} \text{ and } V(k_{\text{eff}}) = \langle k_{\text{eff}}^2 \rangle_{\beta|y} - \langle k_{\text{eff}} \rangle_{\beta|y}^2 .$$

(2.42)

All these expected values can be expressed as Eq. (2.40). Thus, the goal is to compute, or estimate, Eq. (2.40).

Now, Bayes’ theorem [27] states that

$$p_{\beta|y}(\beta) = \frac{p_{Y|\beta}(y)p_{\beta}(\beta)}{p_Y(y)} = \frac{p_{Y|\beta}(y)p_{\beta|y}(\beta)}{\int p_{Y|\beta'}(y)p_{\beta'}(\beta') \, d\beta'},$$

(2.43)

where the prior PDF of the data $p_Y(y)$ is rewritten using the generalized law of total probability [Eq. (A.11) in Paper III]. Inserting this into Eq. (2.40) gives

$$\langle Q \rangle_{\beta|y} = \frac{\int Q(\beta)p_{Y|\beta}(y)p_{\beta}(\beta) \, d\beta}{\int p_{Y|\beta}(y)p_{\beta}(\beta) \, d\beta} = \frac{\langle Q(\beta)p_{Y|\beta}(y) \rangle_{\beta}}{\langle p_{Y|\beta}(y) \rangle_{\beta}},$$

(2.44)

where the final expected values are taken with respect to $\beta$, as indicated by the subscripts. But by the law of large numbers [35], the mean value of a sample drawn from a distribution converges in probability to the expected value of the corresponding random variable, as the sample size $N \to \infty$ (provided that the expected value exists, i.e., that the integral is well defined). In our case, this can be written

$$\hat{Q} = \frac{1}{N} \sum_{j=1}^N Q(\beta^{(j)})p_{Y|\beta^{(j)}}(y) \Rightarrow \langle Q \rangle_{\beta|y} \text{ as } N \to \infty,$$

(2.45)

also using that the ratio of limits equals the limit of the ratio, provided that the limits exist [29]. We can rewrite $\hat{Q}$ as

$$\hat{Q} = \frac{\sum_{j=1}^N w_j Q(\beta^{(j)})}{\sum_{j=1}^N w_j},$$

(2.46)
where, if the errors in the experiments follow a normal distribution,

\[ w_j = e^{-x^2(\beta^{(j)})/2}, \]  

using Eq. (2.39).

Returning to the example from Papers I-III in connection with Eq. (2.42), the mean and variance of \( k_{\text{eff}} \) can thus be estimated using

\[ \hat{k}_{\text{eff}} = \frac{\sum_{j=1}^{N} w_j k_{\text{eff}}(\beta^{(j)})}{\sum_{j=1}^{N} w_j}, \]

\[ \hat{V}(k_{\text{eff}}) = \frac{\sum_{j=1}^{N} w_j k_{\text{eff}}^2(\beta^{(j)})}{\sum_{j=1}^{N} w_j} - \hat{k}_{\text{eff}}^2, \]

respectively. This can be said to be a combination of TMC and UMC-B (see 2.4 and Sec. 3.2.4). UMC-B [basically just applying Eq. (2.46) to nuclear observables] is also used in Paper V but using Metropolis sampling, see below.

**Sampling from another distribution**

It is not necessary to have, or assume, a prior distribution and sample from this. If sampling from a PDF \( h(\beta) \) which is nonzero at least for every \( \beta \) where \( p_{\beta|y}(\beta) \) is nonzero, we can use that

\[ \langle Q \rangle_{\beta|y} = \int \frac{Q(\beta)p_{\beta|y}(\beta)}{h(\beta)}h(\beta) \, d\beta. \]  

(2.49)

The same reasoning as above gives

\[ \hat{Q} = \frac{\sum_{j=1}^{N} Q(\beta^{(j)})p_{\beta|y}(\beta^{(j)})/h(\beta^{(j)})}{\sum_{j=1}^{N} p_{\beta|y}(\beta^{(j)})/h(\beta^{(j)})} \to \langle Q \rangle_{\beta|y} \text{ as } N \to \infty. \]  

(2.50)

It may not be that easy to compute \( p_{\beta|y}(\beta^{(j)}) \). However, assume that we can compute a function \( g(\beta) \propto p_{\beta|y}(\beta) \), e.g., \( p_{y|\beta}(y)p_{\beta}(\beta) \) or, if there is no prior distribution for \( \beta \), \( p_{y|\beta}(y) \). Then, the proportionality constant cancels and we can use Eq. (2.46) with the weights

\[ w_j = \frac{g(\beta^{(j)})}{h(\beta^{(j)})}. \]

(2.51)

This is often referred to as (self-normalized) **importance sampling**, due to the possibility to sample in regions that are “important”, which depends on \( Q \). If \( Q \) is a scalar \( q \), the best choice is \( h(\beta) = |q(\beta)|p_{\beta|y}(\beta) \) [35], but this is normally not known on beforehand (this would normally mean that we already knew the answer).

**Markov chain Monte Carlo: the Metropolis algorithm**

Eq. (2.46) converges to the desired expected value as \( N \to \infty \), using the weights of Eq. (2.47) or Eq. (2.51). However, nothing is said about how fast
this convergence is. In relative terms, the uncertainty of a Monte Carlo estimate is proportional to $1/\sqrt{N}$ if the sample is not autocorrelated and the target distribution has finite variance [35]. In absolute terms, the convergence depends on the sampling distribution, and this can be of great importance.

The Metropolis algorithm [36, 37], a type of Markov chain Monte Carlo algorithm, is often more efficient than the suggestions above since it is often difficult to choose a good sampling distribution for the importance sampling. Loosely speaking, the Metropolis algorithm chooses the sampling distribution for you. In fact, it provides a way to obtain a sample from a PDF which approaches a PDF of choice, e.g., $p_{β|y}(β)$. It is only necessary to be able to compute $p_{β|y}(β)$ up to a constant. This serves our purpose well; we can consider the function $π(β) = p_{y|β}(y)p_{β}(β)$ and do not need to bother about the denominator in Eq. (2.43). One downside is that the sample becomes autocorrelated, i.e., the samplings are not independent of one another.

There are a few somewhat different versions of the Metropolis algorithm; here we describe Random-walk Metropolis. At each moment, the algorithm is in a so-called state, with a certain value $β^{(k)}$ for the variable of interest, where $k \in \{1, 2, ..., N\}$. It randomly walks between the states in such a way that it is more likely to “spend much time” being in states where the PDF is relatively large. It relies on two fundamental choices: a starting state $β^{(0)}$, and a proposal distribution $Q$, e.g., a normal distribution centered about zero with a certain covariance. The algorithm can be outlined as follows:

1. Given a state $β^{(k)}$, the algorithm randomly selects a new state $β^{\text{proposal}}$ as $β^{(k)} + δ$, where $δ$ is sampled from $Q$.
   - If $π(β^{\text{proposal}}) > π(β^{(k)})$: set $β^{(k+1)} = β^{\text{proposal}}$.
   - Else: set $β^{(k+1)} = β^{\text{proposal}}$ with probability $π(β^{\text{proposal}})/π(β^{(k)})$.
   - Otherwise: set $β^{(k+1)} = β^{(k)}$.
2. The state $β^{(k+1)}$ is saved as a (potential) part of the sample.
3. Return to 1, for the state $β^{(k+1)}$.

Informally, step 1 translates to

1. From the current state, the algorithm randomly selects a proposed state in the surroundings. The proposed state is considered below.
   - If the sought PDF is greater in the proposed state than in the current, move to the proposed state.
   - If the sought PDF is not greater in the proposed state than in the current, consider moving anyway, and let chance decide. The probability for a move is large if the decrease in the sought PDF is small.
   - If we did not make any move, let the next state equal the current.

If the proposal distribution is symmetric\(^7\), such that it would be as likely to go from $β^{(k)}$ to $β^{\text{proposal}}$ as the other way around, the sample will be obtained

---

\(^7\)The Metropolis-Hastings algorithm [38] provides a generalization which can handle asymmetric proposal distributions.
from a distribution which asymptotically (as $N \to \infty$) reaches the PDF proportional to $\pi(\beta)$, i.e., $p_{\beta|y}(\beta)$. To decrease the dependence on the starting guess, a number of observations (hundreds or thousands) is normally discarded from the beginning of the algorithm; this is called burn-in. A sample obtained with the Metropolis algorithm is autocorrelated since each observation in the sample depends on the previous. If properly converged, this does not introduce a bias to the distribution, but it means that the uncertainties of resulting estimates are greater than if the sample would be independent [39]. If the sample is used in a time-consuming computation (e.g., in neutron transport simulations), it is therefore advisable to discard a quite large number of states between every recorded observation.

The choice of proposal distribution is essential. If the proposed steps in general are too long, it is likely that most proposed steps will be rejected by the algorithm, and the algorithm gets stuck in a state. If, on the other hand, the proposed steps are too short, it will take a very large number of samples until the algorithm has moved substantially. Both the choice of proposal distribution and the number of discarded states between every recorded observation are discussed a little more in Paper V, for the situation considered there.

2.3 Nuclear data (ND)

By nuclear data (ND) we mean numerical representations of nuclear physics processes. This thesis focuses on interactions between neutrons and nuclides, due to their outstanding importance in many applications (e.g., in nuclear power), so this introduction to ND only regards reactions with a neutron impinging on a nuclide, even if many concepts are valid for other incoming particles.

2.3.1 Cross sections and other quantities of interest

A very important type of ND is the cross section, which provides a way to quantify interaction probability between an incoming particle (neutron) and a nuclide. Perhaps counter-intuitively, the cross section has the dimension of an area; it can be seen as the “effective area” of a nuclide, seen in the perspective of the approaching neutron. It is normally given in barns (b), where $1 \text{ b} = 10^{-28} \text{ m}^2$, which is on the same order of magnitude as the geometrical area of a medium sized nuclide (although such an area is not well defined).

To be a bit more precise, consider a neutron with energy $E$, traveling through a medium with a concentration $N$ (1/volume) of a nuclide of interest, which has a cross section $\varsigma(E)$. The probability $\text{d}P(E)$ that the particle interacts
(with this type of nuclide) when traveling the distance $dx$ is

$$dP(E) = N\varsigma(E)\,dx,$$  \hspace{1cm} (2.52)

which can be rearranged to give

$$\varsigma(E) = \frac{\,dx}{1} \frac{dP(E)}{N}.$$  \hspace{1cm} (2.53)

Thus, the cross section is the interaction probability per distance traveled by the neutron, normalized by the concentration of nuclides (or multiplied by the “volume per particle”). To explicitly write out the energy dependence of a cross section is an attempt to stress this particular fact: cross sections indeed depend on the energy of the incoming neutron, which is discussed further below.

The cross section is usually given the symbol $\sigma$ (sigma), which is unfortunate in ND uncertainty quantification, because another very strong convention is to use $\sigma$ for the standard deviation (see Sec. 2.1.1), used a lot in uncertainty quantification. In this text, as well as in most of the papers, $\sigma$ stands for standard deviation. Another variant of sigma, $\varsigma$, is used to denote cross sections.

After the interaction, the particle moves in another direction, with a new energy. Both the interaction probability and the probability for different angles or combinations of energies and angles is often quantified by so-called differential and double-differential cross sections, respectively. In the ND field, however, “cross section” is often reserved for angle and energy integrated cross sections, i.e., it only quantifies the probability for a certain type of interaction and does not specify the probability density for a particular angle or energy of the recoiling particle. The angular and energy-angle distributions are then treated as other types of ND, and the product of these yields differential and double-differential cross sections, respectively. In this text, this convention of separating cross sections and angular/energy-angle distributions is followed, and the focus is directed towards cross sections.

Even if we focus on this relatively simple form of cross sections, there are several types of nuclear interactions, each quantified by its own cross section. The total cross section is the cross section for any interaction; this is conventionally denoted (n,tot) for incoming neutrons (“n” denotes neutron). In reactor physics [40], this total cross section is often divided into the cross section for scattering and absorption, where absorption can be divided into fission (n,f) and capture, where the latter primarily contains the (n,\gamma) reaction.

---

4Frequently, e.g., in Ref. [40], this equation is written as $R_A = \varsigma I_A N_A$, where it is assumed that $I_A$ neutrons per area and time impinge on a “sufficiently thin” target with $N_A$ nuclides per area, and $R_A$ is the number of interactions per area and time. Noting that $R_A/I_A$ is the interaction probability for one neutron, and replacing $N_A$ by $N\,dx$ to handle the requirement of the “sufficiently thin target” mathematically (and consequently, the probability becomes differential, too), gives Eq. (2.52).
in which the nuclide absorbs the neutron and emits a gamma-ray. This is often just called capture, or radiative capture. The fission and capture reactions must be appropriately balanced to sustain a chain reaction. Scattering can be divided into elastic scattering, \((n,el)\) and inelastic scattering \((n,n'1)\), where the former is scattering without kinetic energy loss and the latter is scattering which leaves the nuclide excited. This classification of the cross sections is depicted in Fig. 2.3(a).

For general purpose ND, the cross sections are classified more along the lines of Fig. 2.3(b). Generally, a larger flexibility is needed for different applications, and experimental data and models may also consider smaller subsets of interactions. The total cross section is divided into elastic scattering, \((n,el)\), and interactions that are not elastic scattering, \((n,non-el)\). The \((non-el)\) reactions are divided into various reactions, which in turn often can be sub-classified. For example, one may divide inelastic scattering into different reactions depending on which energy level the nuclide is excited to.

For conventional nuclear reactors, relatively slow neutrons (“thermal” neutrons) are the most abundant, and for such low neutron energies the total cross section is often (almost) only made up of the \((n,el)\), \((n,\gamma)\) cross sections, since other interactions need a certain kinetic energy of the neutron – they are so-called threshold reactions. Important exceptions are fissile nuclides such as \(^{235}\text{U}\) and \(^{239}\text{Pu}\) which are likely to undergo fission after absorbing a thermal neutron. Another interesting exception (studied in Paper IV) is \(^{59}\text{Ni}\), which has significant thermal \((n,\alpha)\) and \((n,p)\) cross sections (neutron capture followed by \(\alpha\)-particle and proton emission, respectively).

As indicated previously, cross sections depend on the energy of the incoming neutron. This dependence is clear from a look at Fig. 2.4, which shows the \((n,\gamma)\) and \((n,el)\) cross sections for \(^{59}\text{Ni}\) as an example. To make the energy scale meaningful, it can be mentioned that neutrons coming from fission have energies of a few MeV. In conventional reactors, the majority of such neutrons are slowed down, and most fissions are induced by neutrons with a few eV (no prefix this time). In other applications, neutrons of a few hundred MeV can be important.
For low enough energies (typically up to $\sim 1$ eV), all non-threshold cross sections except the elastic are nearly proportional to $1/\sqrt{E} \propto 1/v$, where $v$ is the neutron velocity (hence, the cross section appears as a straight line on a log-log plot). An interpretation is that the interaction probability is proportional to the time during which the neutron is in the vicinity of the nuclide. The elastic cross section is constant for such energies, however.

For somewhat higher energies, the energy levels of the compound nuclide becomes evident, and resonant behavior appears; interaction is more likely if the energy the neutron adds to the system matches an excited state of the compound nuclide. The resonances are explicitly included in the data up to a certain energy, above which there is an unresolved resonance range (URR) where there, in reality, still are resonances. The resonances are not included in this region because of lack of information, but there can be so-called URR parameters included in the ND which can be necessary to use in certain computations. The URR parameters basically describe the expected distance between resonances and their expected “size” (resonance widths). Above the unresolved resonance range, there is the fast region. Here, the different levels are so close that they overlap, so much that the resonance structure disappears.

The resonance structure gives rise to a particular group of ND – so-called resonance parameters. Instead of storing the cross section on a very fine energy grid, one uses models that describe the resonance structure using a set of parameters for each resonance.

Except cross sections and energy-angle distributions, there are several other types of ND which are important in different categories of computations. For example: fission yields (the probability for different fission fragments after fis-
sion of a particular nuclide), average neutron multiplicities ($\bar{\nu}$, the number of neutrons released in a fission reaction; a key to maintain a chain reaction in a nuclear reactor), prompt fission neutron spectra (PFNS, the PDF describing the energy of neutrons coming out of a fission reaction), half-lives and branching ratios of radioactive nuclides (the latter describes the probability for different paths of decay), and binding energies [the amount of energy gained by a particular nuclide from being bound together, compared to if all nuclei (protons and neutrons) were free particles; this determines the so-called $Q$-values, the amount of energy released in a particular nuclear reaction]. The list goes on. Although much of the discussion in this thesis could be generalized to other types of ND (more or less easily), the practical examples most often cover cross sections.

2.3.2 Experimental data

Since the 1960’s, a substantial amount of experimental ND has been collected at several data centers [42]. A large fraction of these is compiled and stored into the EXFOR (EXchange FORmat) database [42] maintained by the IAEA – only for neutron cross sections, there are more than 20 000 datasets in EXFOR, in total containing more than 3.6 million data points.

EXFOR contains so-called differential experimental data, i.e., microscopic quantities measured in relatively “clean” experiments, for example, cross sections for a particular nuclide or element at a particular energy and practically all other types of data mentioned in Sec. 2.3.1. The differential data are what is directly interesting as input for computations.

There are also experiments performed on macroscopic systems, often referred to as integral experiments, which can be used to validate computer codes and evaluated ND (see Sec. 2.3.4). Some of these experiments qualify to be classified as benchmarks, and the OECD sub-organization NEA (Nuclear Energy Agency) maintains, e.g., a “handbook” with criticality safety benchmarks [43], devoted to the neutron multiplication factor $k_{\text{eff}}$ in a fission chain reaction.

In differential nuclear physics experiments, one is typically forced to measure something else than the quantity of direct interest, after which one tries to deduce the value of this quantity of interest. For example, a cross section measurement can be carried out by detecting the number of interactions $C$ (“counts”) when a target with $N$ nuclides is subject to a neutron flux $\phi$ during an exposure time $t$. “Knowing” $N$, $\phi$ and $t$ as well as the detection efficiency

---

9There are, however, experiments reported to EXFOR for mixtures of nuclides, and also averaged over relatively wide energy spectra.
\[ \varsigma = \frac{C}{\epsilon N \phi t}, \]  

(2.54)

disregarding background subtraction. In practice, the flux \( \phi \) is easiest to measure by performing the same type of measurement but with a reference material, and rearrange Eq. (2.54) to

\[ \phi' = \frac{C'}{\epsilon' N' \varsigma' t'}, \]  

(2.55)

where the prime denotes that all quantities are related to this reference measurement. Assuming \( \phi' = \phi \), one obtains

\[ \varsigma = \frac{C \epsilon' N' t'}{C' \epsilon N t'} \varsigma'. \]  

(2.56)

Sometimes it is assumed that \( \epsilon' = \epsilon \) and the expression simplifies further. Often, the ratio \( \varsigma / \varsigma' \) is reported rather than the deduced cross section.

Certain cross sections, which are considered to be particularly well-known, are included among the carefully analyzed Neutron cross section standards \[44\], maintained by the IAEA. One such cross section is typically used as the reference cross section \( \varsigma' \) in Eq. (2.56).

A special type of experiment, a transmission experiment, can be used for the total cross section. It relies on carrying out two measurements; one with the target in place, and one without. Assuming constant flux, and if \( C_{\text{in}}/t_{\text{in}} \) is the count rate (of unaffected neutrons) with a target, and \( C_{\text{out}}/t_{\text{out}} \) is the count rate without the target, the transmission is

\[ T = \frac{C_{\text{in}}/t_{\text{in}}}{C_{\text{in}}/t_{\text{out}}}, \]  

(2.57)

i.e., the fraction of neutrons which have passed through the target without interacting at all. Integrating Eq. (2.52) gives that the expected transmission is

\[ \langle T \rangle = e^{-N d \varsigma}, \]  

(2.58)

where \( N \) is the nuclide density and \( d \) is the target thickness. Solving for the cross section, one obtains

\[ \varsigma = -\frac{\ln \langle T \rangle}{N d}. \]  

(2.59)

\[10\]This is in principle obtained from integrating Eq. (2.52), under the assumption that the target is thin enough, such that it is unlikely for a neutron to undergo several interactions, and thereby change energy.
Uncertainties – random and systematic
Since most of the ND describe random processes, measurements naturally contain a random uncertainty. For example, the cross section describes interaction probability, and even if there is a value of $C$ that corresponds to this exact probability, the observed $C$ is an observation from a certain probability distribution\footnote{By definition, it will follow a binomial distribution \cite{15}; every passing neutron will either undergo the interaction or not. If the interaction probability for each passing neutron is small, the binomial distribution is well approximated by a Poisson distribution \cite{15}. With a large enough expected value ($\langle C \rangle \gtrsim 20$ according to Ref. \cite{15}), the Poisson distribution is well approximated by a normal distribution thanks to the central limit theorem.}. It is similar to estimating the probability to get a sum of twelve when rolling two dice by repeating the dice roll a finite number of times – there is an uncertainty in the result even though the probability has a certain value ($1/36$). This natural, random, uncertainty of $C$ is often referred to as “counting statistics”.

There may also be other uncertainties that can be considered random, e.g., an instability in the flux in Eq. (2.54) (or such that $\phi$ only approximately equals $\phi'$ which modifies Eq. (2.56)).

Further, there may be several systematic uncertainties; in the cross section measurement, $N'$ and $N''$ are typical examples. Since the same target (with the same $N$) often is used for several energies, and even reactions, or by different groups of experimenters (examples of all these cases are found in Paper IV), the fact that the error is correlated will be essential. If the correlations are not taken into account, one consequence is that an experiment with a large number of data points practically is treated as a few points with extremely small uncertainty. The experiments in ND range from containing one point up to several thousand.

A particularly interesting systematic uncertainty is that of the reference cross section $\varsigma'$, since a few standard cross sections (see Sec. 2.3.2) are used for the majority of all cross section experiments (and the standard cross sections are also correlated with each other \cite{44}); this introduces correlations between the errors of very many different experiments.

2.3.3 Nuclear reaction models and TALYS
Despite the impressive number of experimental data points mentioned in Sec. 2.3.2, experiments are not enough. The total number of data points in “all nuclear data” is huge; in theory, it is even infinite since many quantities depend continuously on, e.g., energy. Also, even if one would be satisfied with covering a dense energy grid, the variety of different reactions and nuclides (not to mention different types of ND) results in a lot more data points than what reasonably can be measured, so experimental data are not enough for applications. To this end, one uses the help of models for interpolation and extrapolation, in the end producing so-called evaluated nuclear data, which is
discussed a bit more in Sec. 2.3.4. The mentioned interpolation and extrapolation can also be important between different nuclides, such that estimates of data for nuclides without experimental data is possible; the physical models can make us learn more from our observations of nature. Further, the models can be used to introduce rules of physical consistency; some cross sections must sum to another cross section, for example, which is not ensured by experimental data.

As mentioned in Sec. 2.3.1, resonance models can be used in the resonance range (and below). Examples of such are the Single- or Multilevel Breit-Wigner models or the Reich-Moore approximation, all of which are approximations of R-matrix theory [45]. R-matrix theory is derived from quantum mechanics, but the resulting models have parameters which must be determined from experiments, and the parameters differ from resonance to resonance. As such, these models are not very useful when it comes to larger “gaps” in the experimental data. Systematics may however be used to provide some guess about the overall magnitude and to estimate URR parameters: average resonance parameters including expected distance between resonances.

For such systematics, but in particular for energies above the resolved resonance range, nuclear model data can be obtained from nuclear reaction codes such as TALYS [12] or EMPIRE [46], where the former is used directly and indirectly in several parts of this thesis.

TALYS combines several different nuclear reaction models; they have different roles and their importance varies for different neutron energies. The optical model has its name from treating the incoming particle similarly to a light ray impinging on a glass sphere [47] which is not perfectly see-through, i.e., some light is absorbed. It replaces the complicated interaction between an incoming particle and a nucleus by an optical model potential (OMP) $U(r,E)$, which is a function of the distance to the center of the target, $r$, and the energy $E$ of the incoming neutron. The OMP $U(r,E)$ is complex (in the mathematical sense of the word), where the real and imaginary parts are responsible for scattering and absorption (leading to a reaction), respectively.

The Schrödinger equation, making up the foundation of quantum mechanics, is solved (numerically) using this potential with the incoming neutron described as a wave. It yields a set of basic information about the reaction which is further used by TALYS; for example, it computes the total cross section and divides it into a reaction part and a so-called shape elastic part:

$$S_{(n,tot)} = S_{\text{shape-el.}} + S_{\text{reac.}}.$$  \hspace{1cm} (2.60)

Shape elastic interactions, also called direct elastic or potential scattering [47], are elastic interactions where the neutron “bounces off” the nuclide; it is affected by the potential but there is no forming of a compound nucleus (see below). In general, the reaction cross section also includes a part of the elastic cross section, so the (non-el) and reaction cross sections are not equal.
If there is a reaction (as opposed to shape elastic scattering), different reaction mechanisms are possible. The probability for a certain mechanism to occur depends on the energy of the neutron, and several different mechanisms can occur before the process is over; the mechanisms can be divided into compound, pre-equilibrium, and direct reactions.

In compound reactions, which take relatively long time ($\sim 10^{-16}$ s [47]), the neutron and the nuclide forms a new, compound nucleus. This adds energy to the system, such that the compound nucleus is excited, and it de-excites by ejecting gamma-rays, neutrons, protons, or other particles, possibly in combination and in several steps, until the resulting nucleus is in a stable, or metastable, state. At different points is the de-excitation process, some nuclei can also fission. Compound reactions are described by Hauser-Feshbach theory, using so-called transmission coefficients which are partially obtained from the optical model, also using tabulated information about nuclear structure (e.g., energy levels) and level density models. A central part of the theory for compound reactions is that the initial state of the system is forgotten (in many respects) due to a large number of nucleon-nucleon interactions (collisions inside the nucleus).

In direct and pre-equilibrium reactions, there is no formation of a compound nucleus between the incoming neutron and the target. One direct reaction is already mentioned: direct elastic scattering. Other direct reactions typically demand more energy of the incoming neutron to take place; the neutron is then able to interact with only a small part of the nucleons at the surface of the nuclide. For direct reactions, there are only one or two nucleon-nucleon interactions. Pre-equilibrium reactions are somewhere in between compound and direct reactions, in that there are several nucleon-nucleon interactions, but not that many that the initial state is forgotten. For both these mechanisms, the Hauser-Feshbach theory does not apply, and there is a range of other models involved. However, even if the first reaction is direct or pre-equilibrium, we are often left with an excited nucleus which then may be de-excited as in a compound reaction (at least for $E \lesssim 40$ MeV [12]).

All in all, a lot of models, each with their model parameters can be involved in computing cross sections, angular distributions, etc., for a reaction using TALYS: OMP parameters for the target, and level density parameters for each residual nuclide, for example. Even if the incoming particle is a neutron, the OMP is different for different outgoing particles, introducing even more parameters. The number of parameters with a significant impact depends on the nuclide, but it is on the order of 50. However, they are to some extent motivated by physics, and there are systematics that can be used to provide parameter values even for nuclides with little or non-existing experimental information. The results using such systematics are quite uncertain; the uncertainties are coarsely estimated to be about $6 - 10\%$ for the best known interactions such as (n,tot) and (n,el), and approximately $50 - 100\%$ for others [48]. Before closing this section, it can be interesting to note that the compu-
tational time for a TALYS calculation depends on the nuclide and energy grid, but ranges from seconds to minutes on a contemporary laptop.

2.3.4 Evaluated nuclear data

As described in Sec. 2.3.3, the experimental data are not enough for applications, and the models need experimental data to give reasonably precise results. In applications, one therefore normally makes use of so-called evaluated nuclear data, which is a combination of experimental data (which is selected and analyzed) and nuclear model data. For a few nuclides, there are energy ranges and cross sections with so dense experimental data that a direct fit which linearly interpolates between grid points (see Sec. 3.2.2) is sufficient; such evaluations are generally the most trusted and form the Neutron Cross Section Standards [44].

For other nuclides, there is very little or no data. In such cases, one has to rely on the predictive power of the nuclear models. For most of the nuclides that are important to established applications, there is an intermediate amount of data, and some form of fitting models to data is applied. In the resonance range, codes such as SAMMY [49] and REFIT [50] use least squares to fit resonance parameters to cross section data if available. In the fast range, models such as TALYS can be fitted (in some sense), most often using a particular form of Generalized Least Squares (GLS, as described in Sec. 3.2.2). Some other options for this fitting are described in Sec. 3.2.

Integral experiments are also used, primarily for validation of the evaluated ND. In practice, however, there is an adjustment to some of these data. In, e.g., Ref. [52], there is a more explicit adjustment to integral experiments. It should be stressed that any such adjustment relies on a good prior distribution for the differential data; otherwise there are too many degrees of freedom.

There are numerous different ND libraries (in different versions), which contain evaluated ND for large sets of nuclides, e.g., ENDF/B [41], JEFF [8] and TENDL (TALYS-based Evaluated Nuclear Data Library, [53]; see also Sec. 2.4). For much of its content, the libraries use the same evaluations, but in other cases there are different opinions on which data should be recommended.

For the “most important” nuclides and types of reactions, nuclear data evaluators also provide estimates of the covariance matrices for the evaluated nuclear data, for the use in the assessment of nuclear data uncertainties and their propagation. When using these covariances, it is most common to use linear uncertainty propagation, but there are examples of Monte Carlo uncertainty propagation starting from the covariance matrices [54]. A problem with the covariance matrices is that information may be lost by storing only the covariances, cf., Sec. 2.4.

---

12 For (n,tot), transmission data are preferred [see Eq. (2.57)].
As of today\textsuperscript{13}, the evaluated nuclear data are normally stored in the ENDF format (Evaluated Nuclear Data Format [45]), where different types of data and different interactions are numbered according to more or less strict conventions. These conventions are followed by processing codes such as NJOY [56] and PREPRO [57], which can be necessary for making the ND usable in different types of applications. For example, the preprocessing can include converting resonance parameters to cross sections including Doppler broadening (an effect of material temperature which is extremely important for nuclear reactors), and the grouping of data into energy bins (used in deterministic codes for reactor physics).

2.4 Total Monte Carlo (TMC) and TENDL

Total Monte Carlo (TMC) [58, 53] is a methodology for ND uncertainty propagation built up around the nuclear reaction model code TALYS [12], accompanied by a few other codes, e.g., TARES [59, 53] which completes the TALYS results with resonance parameters. Since TMC was presented by Koning and Rochman in 2008, it has been applied to several cases, e.g., for a large set of experimental benchmarks [60], in computing the neutron multiplication for fuel assemblies [61], and in shielding simulations [62], but also for more complicated systems such as neutronics simulations in a full sized reactor core [63] as well as in a transient analysis [64]. The author of this thesis has previously used TMC in a study of differences in propagated ND uncertainties for the different reactor fuels UO\textsubscript{2} and MOX [22], e.g., finding that the propagated uncertainty from thermal scattering was surprisingly large for MOX because of how the cross sections for \textsuperscript{239}Pu and \textsuperscript{241}Pu relate to each other. The method has also been compared to other uncertainty propagation methods [65], sometimes showing large differences depending on the use of propagation method and on the used ND library.

The methodology takes the Monte Carlo approach for the uncertainty propagation (see Sec. 2.1.3). The more rare feature, in ND uncertainty propagation, is that the uncertainties are propagated all the way from the reaction model parameters, i.e., from the input to the TALYS code structure (in principle, other codes filling the same purpose can be used). As illustrated in Fig. 2.5, the methodology can be outlined as follows:

1. Randomly sample the nuclear reaction model parameters $\beta$ (the distribution is discussed below), $n$ times.
2. Use the resulting $\beta^{(1)}$, $\beta^{(2)}$, ..., $\beta^{(n)}$ in TALYS runs, which give $n$ sets of ND, referred to as random ND files (or simply random files) in the following.

\textsuperscript{13}There are plans to replace the ENDF format (it has artifacts of punch-card usage) with a more modern format [55].
Figure 2.5. A schematic view of TMC. The random files can also be used in the production of TENDL. Using feedback from how the random files agree to experimental data, the parameter distribution is obtained using BMC. In principle, it could also be done using, e.g., UMC-B (see Sec. 3.2.4).

3. Use the $n$ random ND files from the previous step in a computation or simulation of a nuclear system.
4. Study the distribution of the output quantities to draw conclusions on the propagated ND uncertainty.

The computation in step 3 can be done using any type of code (or system of codes) making use of ND, as long as the user can access the ND input (in some industrial codes, this can be a practical problem). One may thus use a Monte Carlo transport code such as MCNP [34], and this, “using Monte Carlo uncertainty propagation through Monte Carlo codes”, gave rise to the name Total Monte Carlo. However, the computation can very well be deterministic, an example of which is found in Ref. [64].

The random files in step 3 can also be used for the production of ENDF files including covariances to quantify the uncertainties, as in the TALYS-based Evaluated ND Library, TENDL [53]. TENDL is the most comprehensive ND library in general, and in particular when it comes to covariances; covariance files are available for all the $\sim 2800$ available nuclides at the TENDL homepage [66], and in many cases also random files suitable for TMC. The library is primarily based on TALYS calculations, but it also contains files which are copied from other libraries. In any case, the major part of the uncertainties are based on parameter sampling. Because TMC and the uncertainties of TENDL are so closely related, they are discussed alternately for the remainder of this section.

Compared to linear uncertainty propagation, TMC naturally inherits the advantages with Monte Carlo uncertainty propagation in general, e.g., no linearity assumption and the possibility to capture more aspects of output and input distributions than the expected value and covariance, see Sec. 2.1.3, and it preserves correlations between different parts of the data which are otherwise lost because of limitations in the ENDF format. On top of these advantages, each random file in TMC is physically consistent, giving rise to physically consistent uncertainty estimates (the distribution cannot leave the physically possible
region) with correlations motivated by the physical models. Also, the implementation is simple with a relatively low risk for errors; between different types of computation and simulation codes, a substantial amount of processing of the data is necessary (mentioned in Sec. 2.3.4), and if linear uncertainty propagation is used, the covariance and sensitivity matrices must be processed, too. Finally, the methodology is suitable for large-scale automation, and this is the reason that TENDL is the most comprehensive ND library.

TMC, of course, inherits the main disadvantage of Monte Carlo uncertainty propagation, too, namely the number of needed computational code runs. Another problem, which applies also to TENDL, is that the model-driven approach (as opposed to driven by experimental data) is sensitive to model defects, i.e., that the model cannot exactly reproduce the truth underlying the data. We will return to model defects several times in this thesis. Related to the model defects, the model-driven approach introduces very strong correlations, e.g., for a cross section at different energies [67]. If there is no experimental data present, this may not be a problem since the strong correlations reflect the way the data are produced. However, if the model is imperfect (which it is), the strong correlations will make us underestimate uncertainties obtained from extrapolation. If a lot of experimental data are present, it is probably best to make more use of this data than the models can allow for, and if the data are produced in this way, the random files should be produced similarly, reducing the correlations. Development in this direction was one of the recommendations of Paper II, and it is the aim of Paper VII. A method to combine TALYS calculations with experimental data is also suggested in Ref. [67], also mentioned in the beginning of Chapter 5.

Disregarding the issue of model defects and model-induced correlations, it is still necessary to choose the distribution for $\beta$. The model parameters are determined by experimental data, and, therefore, statistical inference using experimental data should be used to estimate the parameter distribution. TMC and TENDL have been criticized (e.g., in Refs. [51, 68]) for not having a rigorous statistical foundation and no systematic treatment of experimental correlations. During the last few years, much work has been done in removing qualitative and subjective judgment. In particular, Ref. [48] presents Bayesian Monte Carlo$^{14}$ (BMC), which uses a form of importance sampling (see Sec. 2.2.3) to adjust the parameter distributions for TMC random files and TENDL. Still, a number of quite important practical remedies are not readily justified. For example, the $\chi^2$-measure used for the weights is normalized by a “global $\chi^2$” as a remedy for model defects, and the $\chi^2$ itself handles correlations in a non-rigorous, but perhaps practical, way. Schnabel [10] also criticizes the implementation of importance sampling used in BMC. This being said, BMC and TENDL probably constitute the most ambitious attempt to treat models and experimental data in an automatized way.

$^{14}$The name Bayesian Monte Carlo, in its literal meaning, is more general than in Ref. [48].
Papers I-III are to a large degree devoted to the distribution of the model parameters $\beta$ in TMC, and Paper V considers similar problems but for a toy model. Paper IV takes an approach where experimental uncertainties are propagated using TMC-like sampling instead of used for model fitting. Paper VII is focused on treating model defects, in a way which is highly compatible with the TENDL framework.
3. Fitting techniques in ND evaluation and ND uncertainty propagation

In this chapter, we look into existing methods used, or suggested, for an important part of the ND evaluation process: the fitting of models to experimental data. We focus on the fitting of smoothly varying observables, such as cross sections in the fast range (not in the resonance range). Since Paper V compares several different techniques for such fitting in a situation which is easy to overview, this paper is used as a starting point for the chapter. Papers I-III consider one of these techniques (UMC-B), focusing on the propagation to applications (TMC). These papers are discussed in Sec. 3.4.

3.1 An illustrating toy model

Paper V (summarized in Sec. 3.3) uses a simple toy model to compare the fitting techniques briefly described in Sec. 3.2. The toy model can also be used as an example which helps explaining how the techniques work, and therefore, it is introduced already in this section.

The situation at hand is illustrated in Fig. 3.1. A univariate and single-valued observable \( f(x) \), which has a form inspired by a Prompt Fission Neutron Spectrum (PFNS), is considered. The input variable \( x \) is dimensionless, but can be thought of as a neutron energy in MeV, and we are interested in \( f(x) \) at a grid of eleven \( x \)-values between 0.01 and 15. At six of these grid points, there are experimental data available. A model with three parameters \( c = (c_1, c_2, c_3) \), is assumed\(^1\) to describe the observable, namely,

\[
f_M(x; c) = \left( c_1 \sqrt{x} + \frac{c_2}{\sqrt{x}} \right) e^{-x/c_3},
\]

and there is a prior distribution for the parameters \( c \). Note that Fig. 3.1 shows the truth, the prior, and the experimental data (no resulting fits) for one case; in the paper, a large number of similar cases is studied. More details follow in Sec. 3.3.1.

\(^1\)The assumption is incorrect to a varying degree, as described in Sec. 3.3.1.
Figure 3.1. One version of the toy model of Paper V. The “truth” is not discussed in Sec. 3.1; this topic is revisited in Sec. 3.3.

3.2 An overview of selected techniques

This section briefly presents a few methods used, or suggested for the fitting of models to experimental data within ND evaluation. The four methods described in more detail are compared in Paper V, and the comparison is summarized in Sec. 3.3.

3.2.1 Observable domain (OD) vs. parameter domain (PD)

One way to categorize the ND evaluation techniques presented in Secs. 3.2.2-3.2.5 is to divide them into those that work in the observable domain (OD) and those that work in the parameter domain (PD). The difference between the two categories is illustrated in Fig. 3.2, showing flow charts of methods working in the respective domains.

In OD techniques, it is necessary to assume a prior distribution\(^2\), as explained briefly below. The prior for the model parameters, \(p_\beta(\beta)\), is used to generate a prior, \(p_f(f)\), for the observable \(f\) (at some grid\(^3\)). In Fig. 3.2, this is done using Monte Carlo propagation, but it could be obtained using a linearization of the model. In any case, an assumption for the distribution of the prior for the observable is assumed (typically a normal distribution). After \(p_f(f)\) is obtained, the model is abandoned. The distribution for \(f\) is updated

\(^2\)The prior, in particular the correlations it infers, is necessary for interpolation or extrapolation in OD techniques, which becomes somewhat clearer for GLS in 3.2.2. If the experimental data are dense enough, one can directly fit the observable to the data, as in Ref. [44].

\(^3\)If the observable depends continuously on, e.g., energy, it is considered on a grid, such as in Sec. 3.1.
Figure 3.2. Flow charts illustrating the main principles of ND evaluation in the observable domain and the parameter domain. The model parameter vector is denoted $\beta$ and the vector of observables is denoted $f$. The functions $p_\beta(\beta)$ and $p_f(f)$ are the prior PDFs, and $p_\beta|y_{\exp.}(\beta)$ and $p_f|y_{\exp.}(f)$ are the posterior PDFs, of model parameters and observables, respectively.

Using the experimental data, giving the posterior distribution for the observable $p_f|y_{\exp.}(f)$. Thus, the model impacts the posterior exclusively through the prior for the observable, which makes the assumption of a prior necessary if any interpolation or extrapolation is needed. Note that there is not necessarily a distribution of parameters which corresponds to the posterior distribution of the observable.

In PD techniques, however, a posterior distribution for the model parameters $\beta$ is obtained. The model parameters are compared, via the model, to the experimental data. Thus, the model is directly involved in the fit, and it is not necessary to have a prior distribution for interpolation or extrapolation. By sampling the posterior model parameters, a sample from the posterior of the observables can be obtained. Even if the posterior model parameters would be sampled from a normal distribution, non-linearities in the model typically cause the posterior for the observable to be non-normal.

The techniques described in more detail below are summarized in Table 3.1. GLS and UMC-G work in the OD, while Levenberg-Marquardt and UMC-B work in the PD. The UMC-methods are Monte Carlo methods, while the others are deterministic.
### Table 3.1. An overview of the ND evaluation techniques described in Secs. 3.2.2-3.2.5.

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>Observable domain</th>
<th>Parameter domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLS</strong></td>
<td>General Least Squares</td>
<td>Levenberg-Marquardt: Generalized least squares for linearizations of model</td>
</tr>
<tr>
<td></td>
<td>Standard regression allowing for prior and correlated data</td>
<td>Normal distr. assumed for parameters and exp. data</td>
</tr>
<tr>
<td></td>
<td>Normal distr. assumed for observable and exp. data</td>
<td></td>
</tr>
<tr>
<td>Monte Carlo</td>
<td><strong>UMC-G</strong></td>
<td><strong>UMC-B</strong></td>
</tr>
<tr>
<td></td>
<td>Unified Monte Carlo – “Garage”</td>
<td>Unified Monte Carlo – “Breakfast”</td>
</tr>
<tr>
<td></td>
<td>(Any) distr. assumed for prior observable and exp. data</td>
<td>(Any) distr. assumed for prior parameters and exp. data</td>
</tr>
</tbody>
</table>

#### 3.2.2 GLS

GLS stands for Generalized (or General) least squares [69, 70], and it is the method used in most ND evaluations [51, 70]. In the context of pure regression, Generalized least squares [30] normally refers to the use of Eq. (2.26) or Eq. (2.33), with a non-diagonal data covariance matrix $\Omega$ (as opposed to Ordinary least squares, which assumes $\Omega = \sigma^2 I$). In nuclear data evaluation, the term GLS is normally a bit more restrictive: it refers to the usage of Eq. (2.26) or Eq. (2.33), but in the OD. That is, the “parameters” to fit are simply the desired observables, and the starting point of the actual fit is a prior distribution for the observable at a grid. In the example described in Sec. 3.1, the parameters to fit would, thus, *not* be the model parameters $c$, but the observable at the eleven grid points, $f$.

If Eq. (2.33) is rewritten with notation suitable for GLS, we get

$$\hat{f} = f_{\text{prior}} + \left( C_{\text{prior},f}^{-1} + S^T \Omega^{-1} S \right)^{-1} S^T \Omega^{-1} (y - S f_{\text{prior}}), \tag{3.2}$$

where $\hat{f}$ is the estimate of the observable (at the chosen grid), and $f_{\text{prior}}$ and $C_{\text{prior},f}$ are the expected values and the covariance matrix for the observable, respectively. The design matrix $X$ in Eq. (2.33) is here called sensitivity matrix and denoted $S$. It determines how one should translate the observable $f$ to the experimental points; often, this translation is essentially linear interpolation. In Sec. 3.1, the experimental points are located at some of the grid points. Therefore, $S$ becomes a $6 \times 11$ matrix with only ones and zeros. In a more complicated situation, not all quantities observed in the experiments are
a linear function of the observable. If so, it may be advantageous to iteratively improve $S$.

As mentioned in Sec. 3.2.1, OD techniques need a prior distribution for interpolation or extrapolation; this refers to interpolation or extrapolation to grid points which do not have any direct influence on any experimental points. For GLS, a prior is necessary if there are any columns in $S$ with only zeros, for example; this implies that $S^T \Omega^{-1} S$ is singular [28], such that the equation Eq. (3.2) cannot be solved if $C_{\text{prior},f}^{-1} \to 0$. Expressed differently, if the observable at some grid point has no direct impact on the experiments, it is necessary to have a prior which correlates the observable at this grid point to the observable at other grid points. If a prior is unnecessary and not used, the method is often referred to as Simple least squares [69, 51].

Using a matrix manipulation rule [28], Eq. (3.2) can be rewritten

$$\hat{f} = f_{\text{prior}} + \left[ C_{\text{prior},f} - C_{\text{prior},f} S^T \left( SC_{\text{prior},f} S^T + \Omega \right)^{-1} SC_{\text{prior},f} \right] \cdot S^T \Omega^{-1} (y - S f_{\text{prior}}).$$

Depending on the number of data points compared to the number of grid points (i.e., on the sizes of the involved matrices) and on the condition numbers of the involved matrices, it can be a better idea to use this form. If $C_{\text{prior},f}$ is singular (which happens if some points in $f$ are the sums of others, for example), it is even necessary, because Eq. (3.2) uses direct inversion of this matrix.

3.2.3 UMC-G

UMC-G [71] stands for Unified Monte Carlo–Garage\(^4\), and it is a Monte Carlo method working in the OD. Thus, just as in GLS, the model is included by generating a prior distribution for the observable, $p_f(f)$, using the model, a prior distribution for the model parameters, and a distribution assumption for the prior observables. A distribution assumption is also necessary for the likelihood $p_Y|f(y)$, i.e., the PDF of the experiments given the observables. Typically, normal distributions are assumed both for the prior $p_f(f)$ and for the likelihood $p_Y|f(y)$.

With these assumptions in place, the posterior distribution for the observable is obtained from Bayes’ theorem,

$$p_{f|y}(f) = \frac{p_f(f) p_Y|f(y)}{p_Y(y)}.$$  (3.4)

A sample from this distribution can be obtained using the Metropolis algorithm (see Sec. 2.2.3), as in Paper V. Alternatively, quantities such as $\langle f \rangle$ or

\(^4\)The suffix “Garage” stems from that UMC’s main inventor D.L. Smith had his car repaired at the time. It was appended to the original name UMC to distinguish UMC-G from UMC-B (B for Breakfast, which has a similar etymology).
$\mathbf{V}(\mathbf{f})$ can be estimated using “brute force”, sampling from the prior distribution and using Eqs. (2.46) with the weights from Eq. (2.47).

If normal distributions are assumed for both distributions above, and if the expected values of the experimental data are linear functions of the observable (e.g., if the observable is a cross section and the experiment is directly expressed as a cross section, and linear interpolation is used), UMC-G is a Monte Carlo equivalent to GLS as described in 3.2.2, and the results will also follow a normal distribution. However, there may be reasons to assume some other distribution(s), e.g., a Poisson distribution for the likelihood. Moreover, some experiments may be expressed as, e.g., ratio data; i.e., the values provided by the experimenter could be the cross section at an energy relative to the cross section at another energy (or another cross section). Such experiments can be transformed to the observable space, but this calls for an approximative treatment of the uncertainties and correlations. With UMC-G, one can instead compute the quantity corresponding to the experimental data from the sampled $\mathbf{f}$, and compute the likelihood in the domain of the experiments. This can make a substantial difference [72]. It is not clear from Ref. [72] if the sensitivity matrix in GLS is improved iteratively; otherwise, this could remove part of the difference.

3.2.4 UMC-B

UMC-B [73] is the PD analogue to UMC-G. If “brute force” sampling directly from the prior from the model parameters is used, UMC-B is the direct implementation of Eqs. (2.46) and (2.47), i.e., the sampled model parameters are given weights proportional to their likelihood. An alternative is to generate a sample from the posterior distribution for the parameters, $p_{B|Y}(\beta)$, using the Metropolis algorithm, as in Paper V. Just as in UMC-G, some distribution must be assumed for the errors of the experiments, and this is typically a normal distribution. However, instead of assuming a distribution for the observable, a distribution is assumed for the prior parameters. Even if both these distributions are chosen to be Gaussians, the posterior distribution for the parameters need not be Gaussian, because they are compared via a non-linear model.

To make the difference between PD and OD techniques concrete, one can compare UMC-B to UMC-G. For simplicity, assume brute force implementations of both. Both in UMC-B and UMC-G, sets of observables are then given weights in accordance with their agreement to experimental data. In UMC-B, each set of observables originate from a certain set of model parameters, that are sampled from the prior distribution of the parameters. Thus, we may equivalently say that the weights are given to the model parameters. In UMC-G, however, each set of observables is obtained sampling from a distribution in the observable domain, and there is no particular set of model parameters
corresponding to such a set of observables. Therefore, such a set can violate physical rules.

Papers I-III essentially use UMC-B for the direct estimation of the distributions of integral quantities (such as the neutron multiplication factor $k_{\text{eff}}$ in a nuclear reactor). In these papers, a lot of practicalities, such as the interpretation of experimental data, are addressed (some details in Sec. 3.4).

3.2.5 Levenberg-Marquardt / GLS-P

The Levenberg-Marquardt (LM) algorithm [31, 32, 33] is a technique for fitting non-linear models. It was briefly described in Sec. 2.2.2, and it is described with some detail in Paper VI. There, it is also explicitly described how it can be used with a prior distribution for the parameters. It is an iterative algorithm, based on finding gradually improving linearizations of the model. For the final linearization, generalized least squares is used (in the wider, “pure regression”, sense of the word; see the first paragraph of Sec. 3.2.2). Therefore, the use of the LM algorithm is referred to as GLS-P in Paper V, where “P” stands for the PD.

Thus, it is a deterministic method working in the PD, i.e., the actual model parameters are fitted. Implicitly, normal distributions are assumed for the model parameters and for the experimental data. If normal distributions are used in UMC-B, and the final linearization in LM is adequate, the methods should give similar results. Here, the term adequate includes three aspects: that the method has not converged to a local maximum, that there is only one maximum of importance, and that the model is well approximated by the linearization within the uncertainties.

3.2.6 A few other names and options

Plenty of other names for fitting techniques in ND evaluation occur, most of which are similar to, or special cases of, any of the techniques in Secs. 3.2.2-3.2.5.

BMC [48] is mentioned already in Sec. 2.4, as the method to produce the parameter distributions for TENDL. It is a Monte Carlo method working in PD; thus, it resembles UMC-B. There is, however, no way to formally derive the likelihood function used, even if it resembles that of UMC-B: the $\chi^2$ attempts to handle correlations in an intuitive way, and there is a normalization of the $\chi^2$, practically rescaling the experimental uncertainties on a somewhat arbitrary basis. Since the experimental covariances most often are incomplete (at least in the databases used for this large-scale treatment), and the models are imperfect, these practical remedies are understandable, but not rigorously correct. Backward-Forward Monte Carlo (BFMC) [74] is also most similar to UMC-B, and use similar practical remedies as BMC.
The EMPIRE-Kalman method [70] is most similar to GLS-P in that it uses Generalized least squares in the PD, but only using one linearization of the model [10] (which is EMPIRE [46]), i.e., not the iteratively improved linearization as within the LM algorithm. The Full Bayesian Evaluation Technique (FBET, [75]) uses GLS (in the OD), but also includes a treatment of model defects by adding a covariance matrix to the experimental covariance matrix based on the agreement to experiments for neighboring nuclides.

3.3 Comparing the performance of the techniques

Paper V compares the performance of the techniques in Secs. 3.2.2-3.2.5 for the simple model briefly described in Sec. 3.1. To be more precise, the comparison is actually carried out for \( N = 1200 \) different situations related to that of Sec. 3.1; this is done to be able to compare the methods for more than one random case, which would give results highly sensitive to chance. The different cases can be thought of as different potential “realities”, with somewhat different conditions: the underlying truth, the errors in experimental data, and the prior vary between the realities. Since the data are synthetic, the results can be compared to the underlying truth. More details on how this is done in Paper V is found in Sec. 3.3.1. The general idea, generating synthetic data for a large number of realities to test the performance of a method, reappears in Papers VI and VII which are summarized in Chapter 5.

3.3.1 Generating realities with synthetic data

Throughout Paper V, all fits use the model of Eq. (3.1), restated here:

\[
f_M(x; c) = (c_1 \sqrt{x} + \frac{c_2}{\sqrt{x}}) e^{-x/c_3}, \tag{3.5}\]

where the model parameter vector \( c = (c_1, c_2, c_3) \). The underlying truth is generated using a somewhat different expression, namely,

\[
f_T(x; a) = (a_1 \sqrt{x} + a_2 x) e^{-x/a_3}. \tag{3.6}\]

where \( a = (a_1, a_2, a_3) \). Comparing the equations one can see that the truth \( f_T \) can be reproduced by \( f_M \) if \( a_2 = 0 \); otherwise, there will be what we call a model defect present, i.e., it will be impossible to find a set of model parameters \( c \) such that the model reproduces the truth at all \( x \). Thus, if \( a_2 \neq 0 \), there is a model defect.

Therefore, for the \( N = 1200 \) different realities in Paper V, \( a_2 \) is sampled with an expected value of 0; one could perhaps say that the “model defect is sampled around 0”. The other parameters in \( a \) are fixed. Even if they do not
impact the degree of the model defect directly, it could be preferable to sample also these parameters, to cover a wider range of possible situations.

Thus, in each reality, we have an underlying truth. The experimental data are then generated around the truth, with errors sampled from both random and systematic uncertainties. In the fits, two different data covariance matrices are used for each reality. One of them is consistent with the experimental data, i.e., the covariance matrix is generated sampling from the same error distribution as the actual errors. For the other data covariance matrix, a part of the systematic uncertainty is neglected.

Finally, assuming prior uncertainties for the model parameters $c$, the prior expected model parameters, $\langle c \rangle$, are sampled in accordance to these uncertainties.

Thus, Fig. 3.1 only illustrates one of the realities in Paper V, with particular $a_2$ and prior $\langle c \rangle$, and a particular set of experimental data.

### 3.3.2 The comparison

All of the methods described in Secs. 3.2.2-3.2.5 are applied to each of the realities generated as described in Sec. 3.3.1. All distributions that are assumed within the fits (for the experimental data, and for the observable/parameters) are normal. Because of this, and since the studied situation does not include any non-linearities between the observable and the experimental data$^5$, GLS and UMC-G give the same results, up to the Monte Carlo statistics in UMC-G. Also, it turns out that UMC-B converges well enough, and that the final linearizations of LM are adequate enough, such that the results of UMC-B and LM agree well, too. Therefore, all significant differences (in this setup) are between the OD (GLS and UMC-G) and PD (LM and UMC-B) techniques.

Between these two groups of methods, differences are observed. Fig. 3.3 shows some selected results for the case with the data covariance matrix consistent with the data generation. Results are shown for the 11 grid points, but also for two “integral” quantities which take all the grid points into account (end users of the evaluated ND typically use quantities of this form).

Fig. 3.3(a) shows the “evaluation bias”, i.e., the mean relative error obtained in the fits. For both OD and PD, the bias is on the order of 1% for $x \lesssim 5$; it is overall smaller for PD. Then, it increases rapidly in magnitude for larger $x$, reaching almost 100% for PD and more than 1000% for OD. Similar results are seen for the integral quantities, where “Rel.” is more affected by the large $x$ and “Const.” is more in line with results for smaller $x$. Thus, with respect to evaluation bias, the PD techniques perform better.

$^5$Such non-linearities could be, e.g., ratio data, see 3.2.3. During the work with Paper V, ratio data was actually tested. Substantial differences were then seen between GLS and UMC-G (as expected [72]), but barely between LM and UMC-B.
Fig. 3.3. Selected results from the work of Paper V, both for the grid points and for the “benchmarks”. These results are for the case with a data covariance matrix consistent with the data. The uncertainty bars show 1σ sampling uncertainty, caused by the finite number of realities. Fig. 1 in Paper V contains more information.

Fig. 3.3(b) shows \( \chi^2_{\text{truth}}/N \), the reduced \( \chi^2 \) comparing the fit to the underlying truth, including the uncertainty of the fit in the comparison. If the model and the assumptions made in the fits would be correct, these reduced \( \chi^2_{\text{truth}} \)-values should be close to 1. Even though the evaluation bias was somewhat smaller for PD, the \( \chi^2_{\text{truth}} \)-values are, overall, substantially greater for PD than for OD. Also for OD, most \( \chi^2_{\text{truth}} \)-values are greater than 1. Thus, for both groups of methods, the uncertainties are generally underestimated, in particular for PD.

The worse performance for PD can be explained by the PD techniques being more sensitive to model defects. This is reasonable, since the PD techniques are strictly constrained by the model; if the model is incorrect, this constraint will induce an error. The OD techniques are more flexible because they abandon the model after the prior is generated; the prior will impact the results but the results need not follow the model (and will not do so, in practice). Indeed, when studying how \( \chi^2_{\text{truth}} \) depends on \( a_2 \), the parameter controlling the model defect, it is clear that the dependence generally is stronger for PD. An example of this can be seen in Fig. 3.4, which shows the error in “Const.” normalized by the uncertainty\(^{6} \) for each reality, plotted vs. \( a_2 \). Fig. 2 of Paper V is similar (but for \( x = 0.50 \)). One can also see from the figure that the OD techniques also are affected by the model defect; thus, whatever technique is used, model defects need to be addressed if they are present (and it is not controversial to say that they are). It is interesting to compare 3.4 to Fig. 8 of Paper VII, where model defects are treated.

---

\(^{6}\)The square of this quantity equals \( \chi^2_{\text{truth}} \).
Even if the overall performance is worse for the PD techniques, their smaller bias is not without value. Also, if there are no model defects, the PD techniques perform better in all respects (this was also tested during the work with Paper V). This is not surprising; if the model is correct, it is an improvement to make as much use of it as possible. There are also other advantages with the PD, such that the results will follow all physical rules which are included in the model.

For the case with a data covariance matrix which is inconsistent with the data generation, the picture is quite similar for the bias; it is generally a little worse, especially for PD such that the competition between OD and PD with respect to bias evens out. However, the results are more dispersed, while the uncertainty estimates are smaller, giving generally larger $\chi^2_{\text{truth}}$-values. This is not surprising, but quantifies the potential effect of doing an improper analysis of experimental uncertainties.

In one respect, the UMC methods differ to the deterministic methods: the number of function calls. To obtain a reasonably uncorrelated sample of size 200 in UMC-B, the number of function calls varied between approximately $10^4$ and $10^6$, which may be infeasible if a more time consuming model is used. With LM/GLS-P, the number of function calls is approximately 50. Out of this, $3/4$ is used for the estimating the Jacobian, and this part increases linearly with the number of parameters (if not intelligent approximations are made); for a reasonable number of parameters, far less function calls are needed.

3.4 More on UMC-B + TMC in practice

Papers I-III uses weights proportional to the likelihood to adjust the distribution of random ND files used in TMC for ND uncertainty propagation to applications. Assuming that the original random files describe the prior distribution of the model parameters, this is equivalent to using a “brute force”
UMC-B, and propagating the results all the way to applications. This assumption is incorrect; the used random files are affected by the experimental data and is thus inappropriate as a prior. The model parameters used to produce the random files were not available at the time of the work, so it was not possible to compute the PDF for the model parameters and compensate for this, using importance sampling. To test the general idea in a simple way, the authors decided to assume that the files represented the prior distribution.

Unlike in Paper V, the random files in Papers I-III describe actual ND, and the experimental data are real, i.e., not synthetic. This leads to a number of practical considerations within the interpretation of the experimental data, which are summarized in Sec. 3.4.1. This is followed by the discussion of some of the results from Papers I-II in Sec. 3.4.2.

3.4.1 Interpretation of experiments

For UMC-B to work, just as for any of the fitting methodologies, the likelihood must be estimated based on the experimental information, which consists of the observed values and (hopefully) uncertainty information. In EXFOR [76], the database with experimental data considered in Papers I and II, uncertainties can be given in one or more component, but for many experiments only a total uncertainty (or even none at all) is quoted. The original format is not very easy to interpret automatically, there are several different ways to compile the data, using different conventions for missing values etc., and different units. Further, it is often unclear whether an uncertainty is random, systematic, or a combination of the two – and many (not to say most) experiments are reported with incomplete uncertainty information; in many cases, only the uncertainty from counting statistics (see Sec. 2.3.2) is included. Thus, there may be overlooked systematic uncertainties. Moreover, there can be data that are erroneously reported or compiled.

Ideally\(^7\), experts on experimental uncertainties evaluate the available data sets (which, ideally, are well reported), select which are reliable and carefully construct experimental covariance matrices from this. For large scale work as in TENDL (see 2.3.3), this is not feasible, and some kind of automatic treatment is necessary. In the used scheme [48] for TENDL, this leads to a few practical remedies that are not statistically well-founded.

Papers I and II attempt to interpret the EXFOR database automatically, using a set of simple rules, in order to construct an experimental covariance matrix covering the full set of data. Among the more important rules, experimental points with only a total uncertainty quoted get penalized by assuming that the experimental point has both a random uncertainty and a systematic uncertainty of the quoted magnitude. Also, a 1% systematic uncertainty which is common for all experiments measuring the same quantity is added. Otherwise,

\(^7\)It can possibly be discussed if this is truly ideal, as it introduces some human bias.
all uncertainties except one random uncertainty are considered systematic uncertainties which are common for all data points within the same experiment. With all the uncertainty components determined, the data covariance matrix is determined using the assumption that the random variables describing the data points given the model parameters $\beta$ can be written as

$$Y_i|\beta = f(x_i; \beta) + \mathcal{E}_{\text{random}} + \sum_{\ell=1}^{\nu} \sigma_{i\ell} \mathcal{E}_\ell,$$

(3.7)

where $f(x_i; \beta)$ is the model value given $\beta$, $\mathcal{E}_{\text{random}}$ and $\sigma_{i\ell} \mathcal{E}_\ell$ are random variables describing the random error and the error due to the $\ell$th systematic error component, respectively; $\mathcal{E}_\ell \sim N(0, 1)$ and $\sigma_{i\ell}$ is the standard deviation of $Y_i$ due to the $\ell$th systematic contribution. Taking the covariance of $Y_i$ and $Y_j$ gives

$$\text{cov}(Y_i, Y_j) = \delta_{ij} \sigma_i^2 + \sum_{\ell=1}^{\nu} \sigma_{i\ell} \sigma_{j\ell},$$

(3.8)

where $\sigma_i^2 = V(E_i)$ and $\delta_{ij}$ is the Kronecker delta; it equals 1 if $i = j$ and 0 otherwise.

A very influential part of the automatic interpretation of EXFOR is an attempt to automatically identify and reject outliers. In the work of Papers I and II, this is implemented in a practical but scientifically unsatisfactory way, based on the deviation between the experiments and the distribution of the random files. This is unsatisfactory since experiments may be rejected because of an erroneous distribution of the model data, and not the other way around. If an automatic treatment of outliers is possible, it should be based on the deviation between different experiments.

The rules used for the EXFOR interpretation in Paper I and II differ somewhat, because of discussions among the authors initiated by feedback on Paper I from the community of experimenters. The difference gives substantial effect on the comparable results, as seen in Sec. 3.4.2. Otherwise, Paper II is an extension of the work in Paper I, to a large degree. It contains more results, a more detailed description of the methodology, and a more detailed discussion of the results, as well as a study where a group of parameters representing assumptions in the automatic EXFOR interpretation are varied to study how sensitive the methodology is to the choice of these parameters.

3.4.2 Some results from Papers I and II

In Papers I and II, the experimental covariance matrix implied by Eq. (3.8) is computed using the random and systematic uncertainty components obtained according to Sec. 3.4.1, and this covariance matrix is used in the computation of the likelihood weights according to Eq. (2.46) with weights proportional to $e^{-\chi^2/2}$. These weights are then used to weigh the distributions for quantities
computed for several applications using the random files which at the time were publicly available at the TENDL homepage [66], giving weighted estimates for the ND uncertainty propagated to these quantities. The random files of three nuclides, $^{235}\text{U}$, $^{239}\text{Pu}$, and $^{56}\text{Fe}$, are used in both papers, while Paper II also includes $^{238}\text{U}$.

In Paper I, the weights shifted the mean values of the distributions in most cases, while the ND uncertainty was not reduced much in any case. An example is shown in Fig. 3.5(a), concerning the neutron multiplication factor $k_\infty$ (disregarding neutron leakage) in a UO$_2$ pin cell (at “end of life”) in a typical PWR (pressurized water reactor), varying $^{235}\text{U}$ data and using weights from experiments with $E < 5 \text{ eV}$. The prior and posterior distributions (unweighted and weighted) for $k_\infty$ are compared, and one can see how the weights from experimental data generally suppress files giving low $k_\infty$ and amplify files giving greater $k_\infty$, resulting in a shift upwards of the distribution. One can also note the possibility that the prior (unweighted) distribution limits the weighted distribution too much – what if there would have been random files giving even greater $k_\infty$?

In Paper II, the posterior uncertainties can be divided into two groups: for some applications (where particular ND is varied), the propagated ND uncertainty estimates are reduced to practically zero by the weights, and the central values are shifted. For other applications, the ND uncertainty is practically unaffected by the weights, just as the central values. An example of the former is shown in Fig. 3.5(b), which illustrates the distribution of the same content as Fig. 3.5(a) from Paper I, but this time, the distribution is very peaked. The large difference in distribution is caused by the different rules used for the EXFOR interpretation in the two papers.

The very peaked distributions in Paper II can have several quite different reasons, e.g.,
1. The parameters are sampled in a region where the likelihood is low, giving large absolute differences between $\chi^2$-values of different random files, which leads to large relative differences in the likelihood due to the exponential behavior of the likelihood function.

2. The parameters are sampled from a “too” large region resulting in a bad resolution of the region with a high likelihood. This case is not very problematic, since the small uncertainty estimate would be close to the truth.

3. Erroneous interpretation of the experimental data.

4. Model defects.

The two former of these could be remedied by using feedback from the experimental data already in the sampling of the model parameters, e.g., using the Metropolis algorithm, or by leaving the Monte Carlo approach altogether.

In the paper, the third item leads us to conclude that the rules for interpreting EXFOR should be further discussed and refined. In particular, the treatment of outliers needs improvement. Perhaps, full automation may be very difficult. As a consequence, Paper IV contains a careful treatment of the experimental data. Automated interpretation of experimental data is revisited in Chapter 6, but with an entirely different approach.

Finally, the model defects must be treated. Chapter 5, summarizing Papers VI-VII is devoted to this subject.

The unaffected distributions may be due to model defects or limitations of the EXFOR interpretation, too, but it can also be a result of too “narrow” prior distributions; the assumption that the used random files describe the prior is incorrect, as remarked in the beginning of Sec. 3.4. The problem of the possibly too narrow prior should be tackled in the same way as the other problems related to sampling in the wrong region. It can also be that the weak impact of the weights can be caused by that the studied integral quantities may be most sensitive to resonances. A large part of the resonance range is kept fixed in Paper II, but they may nevertheless be so many that it is inefficient to limit them by weighting full random files. Therefore, it is concluded in Paper II that the resonances need special treatment, as they also get in conventional evaluation.

Because the distributions separated into these two groups of either being very affected by the weights or practically not at all, the study of how sensitive the weighted propagated uncertainties are to different choices of the EXFOR parameters did not give very clear results on the original question, i.e., how sensitive the method is to these choices. It does work as a validation of the implementation; the propagated uncertainties behave as one would expect them to when certain uncertainties were increased, and so on. An interesting observation from the sensitivity study which is discussed in some detail in Paper II is that the effect of increasing fully correlated uncertainties becomes saturated, i.e., the weights do not generally approach a uniform distribution (as they do if all random uncertainties are increased) when systematic uncertainties increase. This behavior is explained by that increasing systematic uncer-
tainties do not allow for deficiencies in the shape of the energy dependence of
the experiments compared to the model curves, it only allows for large offsets
between experiments and models. For this reason, it is potentially dangerous
to model all systematic uncertainties as fully correlated.

Despite the difficulty to draw conclusions from the sensitivity study, it is
clear from comparing Papers I and II (see, e.g., Figs. 3.5(a) and 3.5(b)) that
the interpretation of the experiments does matter, indeed. A better commu-
nication of uncertainties, and a database with transparently evaluated experi-
ments would indeed benefit this type of analysis. Another approach to the
interpretation of experimental data is described in Chapter 6.

Finally, there is one major point made in Paper I which is not included
in Paper II. A “Russian Roulette” technique is described, and used to reduce
the number of necessary application simulations. It is similar to the variance
reduction technique called Russian Roulette used in neutron transport codes
such as MCNP [34]. Without introducing a bias, random files can be removed
(“killed”) by chance, with a survival probability proportional to the weight,
after which the surviving files are used with equal weight. A remark one can
make is that it would possible to apply a Russian Roulette procedure only on
files with weights under a certain threshold.

3.4.3 An alternative computation of the likelihood

Paper III is focused on a detail in UMC-B; it considers an alternative way
to compute the likelihood necessary in UMC-B, i.e., sampling of systematic
errors to give a Monte Carlo estimate. This is done instead of computing it
using the general $\chi^2$ including the solving of a matrix equation, and it means
that it is not necessary to assume a particular multivariate distribution for the
likelihood, and it may also be pedagogical. Even if the method works asymp-
totically, it is very inefficient compared to solving the matrix equations. Per-
haps, one should have considered the possibility to use sampling of systematic
errors only for errors which are clearly non-normal and for such errors which
are correlated over a very large number of experimental points, i.e., to use
the generalized $\chi^2$ for limited subsets of the experimental data. Another al-
ternative, mentioned in the conclusions of Paper III, is to use the sampling of
errors for other purposes than for estimating the likelihood. Paper IV (sum-
marized in Chapter 4) follows this general idea, and samples systematic errors
in experiments and propagates the results in the forward direction.

3.5 Conclusions for Chapter 3

The chapter briefly describes four fitting techniques that can be used to merge
ND models and experimental data within ND evaluation. Two of the methods
(LM/GLS-P and UMC-B) work in the parameter domain (PD), i.e., model
parameters are fitted to the data. The two other are observable domain (OD) techniques, i.e., the model is only used to generate a prior distribution in the observable domain, and then, the observable itself is fitted to the data. The reader also finds a summary of how the techniques compare, primarily based on the work of Paper V, which considers synthetic data in a large number of imagined “realities”.

Besides Paper V, Papers I-III are also summarized. They were written before the work of Paper V, and do not make much use of the formalism of the beginning of this chapter. However, they essentially apply a “brute force” UMC-B to practical cases, using real experimental data, and propagate fitted parameter distributions to integral quantities. Therefore, the chapter contains some additional discussion on this use of UMC-B, and practical considerations for the interpretation of experimental data, which would be necessary in any of the techniques.

In Papers I and II, it is seen that the brute force implementation of UMC-B has problems. Even if there is a risk that the coarse interpretation of experiments cause much of the problems, many problems can be related to poor convergence. This could be remedied by using a more intelligent sampling scheme, e.g., the Metropolis algorithm. However, already in the quite simple situation considered in Paper V, even the Metropolis algorithm required a substantial number of function calls, which correspond to a severe computational burden for a model such as TALYS. Also, in working with Paper V, it was seen for the simple case considered there that the difference between UMC-B and LM/GLS-P was insignificant, even for ratio data. Further, it is likely that other uncertainties are more important than those induced by using a deterministic method. Finally, even if deterministic methods suffer the risk to find a local minimum, Monte Carlo methods can end up in similar problems. Thus, it may be a better idea to study the performance of deterministic methods (for the fitting; Monte Carlo methods can still be a good choice for uncertainty propagation), and from Paper VI and on, only deterministic methods are addressed.

An example of something which is likely to be more important than the approximations in deterministic fitting are the model defects. The effect of a particular model defect was quantified in Paper V, and it is seen that it has a substantial impact for all methods, even if the impact is worse for the PD techniques. The effect of model defects is revisited in Papers VI-VII, but for more complicated models (only for LM/GLS-P). Again, the effects are seen to be severe, but these papers also consider how to treat the model defects.

Even if the PD techniques are more sensitive to model defects, they have other advantages. Perhaps most importantly, they guarantee constraints included in the model to hold. Also, it can be practical: the TENDL framework builds thoroughly verified and consistent ENDF files with complete covariance information, using a distribution of parameters. These are the two major reasons motivating the work in Paper VII, in which model defects are treated by modifying the model parameters instead of partially leaving the PD (as in
Paper VI). Moreover, even if OD techniques are used in final evaluations, they rely on prior parameter distributions, which in practice are obtained by fitting parameters to the data on a global scale. Thus, the model parameters have to be considered at some stage.
4. An evaluation with forward propagation of uncertainty components in experiments

As summarized in Chapter 3, Papers I-III attempt to fit ND model parameters by giving weights to random sets of ND based on experimental data (UMC-B). In Paper III, the weighting makes use of the sampling of systematic errors in experimental data, a method which proved to be quite inefficient. In Paper IV, experimental errors are also sampled, but not with the main purpose to fit model parameters. Rather, the uncertainties are propagated from the experiments to the ND, and all the way to applications, in analogy with the forward propagation of model parameter uncertainties in TMC. For the considered nuclide, $^{59}$Ni, the situation is a bit special, making this approach appropriate: the most important experiments are for thermal cross sections (for low neutron energies; in equilibrium with thermal vibrations of nuclides). This makes the model constraints less complicated than otherwise; nevertheless, there are physical constraints to consider. It can be mentioned that limitations in the ENDF format and the reported experimental data give rise to complications in using, e.g., SAMMY [49] or REFIT [50], two of the standard tools for carrying out low energy evaluations, which is one of the reasons for the unconventional approach taken here.

The work of Paper IV lead to an evaluated ND file for $^{59}$Ni, now included as the $^{59}$Ni file in OECD’s nuclear data library JEFF 3.3 [8, 9].

4.1 Motivating work on $^{59}$Ni

$^{59}$Ni is not naturally occurring, because it is radioactive and decays with a half-life of 76,000 years. However, $^{58}$Ni is the most abundant of the nickel isotopes, and neutron capture on $^{58}$Ni leads to the production of $^{59}$Ni if nickel is subject to a neutron flux. Therefore, nickel-containing components in nuclear reactors (such as many stainless steel components) will, after some time, contain $^{59}$Ni. In fact, the $^{59}$Ni content reaches a maximum of approximately 3% of the initial nickel content after a few years [77]. In turn, $^{59}$Ni has a rare property: non-threshold ($n$,α) and (n,p) cross sections\(^1\). This has the result that $^{59}$Ni is one of the largest contributors to the helium and hydrogen production in stainless steel [77].

\(^1\)For most nuclides, the ($n$,α) and (n,p) cross sections are practically zero for low neutron energies; there is a reaction threshold.
Before the work of Paper IV, the evaluations for $^{59}$Ni available in the major ND libraries did not contain any uncertainties. Even in TENDL, there were no uncertainties for the thermal (n,α) and (n,p) cross sections. Therefore, Paper IV had the goal to provide a modern evaluation for $^{59}$Ni, including complete covariance information.

### 4.2 Sampling errors in thermal cross section experiments

Since the thermal cross sections of $^{59}$Ni are identified as the most important for applications; this is the focus of the evaluation. Also, most experiments available for $^{59}$Ni measure thermal cross sections. When found, the publications for the available thermal cross section experiments are studied in some detail. Otherwise, the analysis is based on the corresponding EXFOR entries.

Each of the experiments is assumed to follow either of Eqs. (2.54), (2.56), or (2.59). We try to identify which of the components of the respective equations that were included in the uncertainty analysis by the experimenters. If an uncertainty component is judged to be missing or is not mentioned, an uncertainty is assumed using values based on the other analyzed experiments. These values are intended to be on the conservative side because of the lacking knowledge. Uncertainty components which are the same for different experimental points (even for different experiments) are identified. There are also some corrections made, and renormalization to current reference cross sections [correcting Eq. (2.56) using the currently established estimate of $\varsigma'$].

Using the results from the above analysis, the different uncertainties are used to sample errors for the individual uncertainty components, providing a way to get a sample from the distribution of all the thermal data points. The ten used points range over four interactions: (n,tot), (n,γ), (n,α) and (n,p). For each interaction, and for each realization in the sample, the respective points are weighed together using generalized least squares. If certain constraints are not fulfilled [e.g., if (n,tot) is less than the sum of the others; see Sec. 4.3 for more details], the realization is discarded and replaced by another. In this way, a sample is obtained for these four thermal cross sections, taking the constraints into account following Bayes’ theorem.

The resulting distribution for the thermal cross sections is not normal. Nevertheless, we summarize the distribution using the mean values and standard deviations, which are seen in Table 4.1, where they also are compared to the corresponding values in the earlier JEFF evaluations, and of Mughabghab’s “Atlas of Neutron Resonances” [78]. The results are significantly different (with respect to the sampling uncertainty) to both JEFF and Mughabghab, but compatible (with respect to the ND uncertainty).
Table 4.1. Estimated expected values $\langle \varsigma \rangle$ and standard deviations $\sigma(\varsigma)$ in barns for the thermal cross sections of the current $^{59}$Ni evaluation, compared to the values from Mughabghab [78] and the previous JEFF evaluation. The sampling uncertainties of the standard deviations are determined using the method described in Ref. [22].

<table>
<thead>
<tr>
<th></th>
<th>$\langle \varsigma \rangle$</th>
<th>$\sigma(\varsigma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{(n,\alpha)}$</td>
<td>12.73(4)</td>
<td>.71(3)</td>
</tr>
<tr>
<td>$^{(n,p)}$</td>
<td>1.51(1)</td>
<td>.209(9)</td>
</tr>
<tr>
<td>$^{(n,\gamma)}$</td>
<td>73.7(2)</td>
<td>3.8(2)</td>
</tr>
<tr>
<td>$^{(n,tot)}$</td>
<td>94.8(3)</td>
<td>4.7(2)</td>
</tr>
</tbody>
</table>

This work

Mughabghab

JEF(F) 2.2-3.2

$^{(n,\alpha)}$ 12.3 0.6 1.69 -

$^{(n,p)}$ 2.0 0.5 -

$^{(n,\gamma)}$ 77.7 4.1 -

$^{(n,tot)}$ 80.7 -

98.3 -

Figure 4.1. An illustration of the work flow in Paper IV. The abbreviation “xs” stands for cross section. The labels of the embedded figures are not intended to be visible in print; the figure is just an overview of the work (many of the embedded figures are found in larger scale in Paper IV).
4.3 Combining thermal cross sections with other data

Even if the thermal cross sections are most important, they must be combined with higher energy cross sections and other types of ND to yield a complete evaluation. The procedure is summarized in Fig. 4.1. In the upper left corner, the sampling of the thermal cross sections is illustrated, with the resulting distributions depicted as a few marginal distributions and a correlation matrix. We use sampling from two other sources of information, shown in the two boxes directly below the thermal cross sections: resonance parameters and the parameters of TALYS (version 1.8 [12]). For the TALYS parameters, the parameter distribution of TENDL 2015 is used. TALYS gives cross sections for energies above the resonance range, and data that are not cross sections. It also gives average level spacings and average widths for resonances, which is fed to the sampling of resonance parameters. This is because the resonance parameters are sampled based on two sources of information, one of which is these average quantities. They are used to sample the positions and widths of the resonances at negative energies (a.k.a. bound resonances) and in a part of the unresolved resonance range (URR); the bound resonances are not directly measurable but will impact the thermal cross sections. The other source for the resonance parameters is data reported to EXFOR (reported as resonance parameters). These are also sampled, using the uncertainties reported by the experimenters, and some added uncertainties which are connected to the thermal cross section experiments through experimental correlations.

In R-matrix theory (or, rather, from the used approximation thereof: Multi-Level Breit-Wigner [49]), the thermal cross sections are a result from the resonance parameters. Therefore, the thermal cross sections need to match the resonance parameters, including those of the bound resonances. For each set of sampled thermal cross sections and resonance parameters, the widths of the bound resonances (which follow quite dispersed distributions), are therefore adjusted until the thermal cross sections are reproduced. If the numerical root finder used for this adjustment fails to find a solution, the combination of resonance parameters and thermal cross sections is considered “unphysical”, and the realization is discarded as described in Sec. 4.2.

After this step, we have a sample from the distribution of resonance parameters which is adjusted to match the distribution of the thermal cross sections. This is combined with the TALYS results to yield a distribution of complete sets of ND: “Random ND” in the figure. This random ND can be used directly in an application for Monte Carlo uncertainty propagation. An alternative is to compress the information of the random ND to one single ENDF file with covariances, which is done for JEFF 3.3 in this work. It can be worth noting that the file contains correlations between all interactions, as partially illustrated by the lower left part in the “One ENDF file...” block of Fig. 4.1. It can also be noted (from the right hand part of the same block) that the correlations are nonzero over the full energy range; this may be random noise, but there
are reasons for nonzero correlations, in particular the TALYS parameter’s impact on the resonance parameters. Format limitations result in the removal of some correlations that are included in the random files, e.g., between the URR parameters and the fast range.

4.4 Concluding remarks for Chapter 4

An unconventional approach to ND evaluation is used in Paper IV, resulting in a new $^{59}$Ni evaluation, now included in JEFF 3.3. Instead of fitting resonance parameters to experimental data, both errors in the thermal data and resonance parameters are sampled on an equal footing. Parameters of sampled bound resonances are then adjusted, such that the distribution of thermal cross sections is matched. The approach assumes normal distributions only for the most fundamental quantities in the experiments, and can propagate otherwise arbitrary distributions all the way to and through the applications. Also, correlations between all involved sources of information is considered, reflecting the assumptions that are made.

The approach is also taken as a pragmatic way to tackle the particular situation, with non-threshold (n,\(\alpha\)) and (n,p) reactions, and very little “raw” cross section data except for the thermal cross sections. Ideally, the experimental information would be more complete. If so, it could be preferable to fit the resonance parameters deterministically, taking the thermal cross sections into account at the same time. It is, however, difficult to imagine how the procedure to sample bound resonances could be applied correctly in combination with one single deterministic fit. Therefore, a procedure similar to that of Paper IV could be motivated for more nuclides.
5. Treating model defects using Gaussian Processes (GPs)

In Paper IV (summarized in Chapter 4), we focus on the distribution of thermal cross section data resulting from experiments, so model defects are not considered. In the matching of thermal cross sections and resonance parameters, a model does play a role, but the distribution of the thermal cross sections is not affected much. Of course, for the fast range, entirely driven by models, model defects can be important. Also, in cases with more experimental data, or for a wider range of energies, the model and experiments interact in a more complicated fashion, and some approach fitting the observables to the data is probably necessary. If so, model defects can have severe consequences, as seen in Chapter 3, and as will be seen later in this chapter [see, e.g., Figs. 5.1(b), 5.5(a), and 5.6]. The results become biased towards the model and the errors will not be reflected by the uncertainty estimates, i.e., the uncertainties will be underestimated, often very much so. Thus, model defects need to be treated somehow.

Different suggestions to treat model defects in ND evaluation have been presented, where many are studied in Ref. [10]. The main approach has traditionally been to add fully correlated uncertainties on the observables, based on the difference between the best model fit and experimental data of high quality [51]. Other approaches have been to rescale either the covariance matrix of the fit or the experimental covariances based on the agreement to experimental data. As argued in Ref. [10], this can be justified to treat underestimated experimental data uncertainties, but not really to treat model defects: the results will still be biased towards the incorrect model. In Ref. [67], a Monte Carlo merging of experimental data and TALYS results is suggested, but more as a means to reduce the correlations. This is criticized in Ref. [10] because the method does not include any smoothness criterion. Refs. [10, 79] instead suggest to use Gaussian processes (GPs, [80]) and, seemingly successfully, apply them in an ND evaluation context\(^1\). GPs have previously been used to treat model defects in several situations, e.g., for polynomial regression as early as 1975 [81]. In these cases, a defect term is added to the model, and the model parameters and the defect are fitted simultaneously.

In Sec. 5.1, which summarizes Paper VI, GPs are used in this way to treat model defects for a nuclear physics application which is not directly ND evaluation; it is used to distinguish overlapping peaks in a histogram of data. This is

\(^1\)To the author's knowledge, GPs have not been used in an evaluation accepted for any of the major evaluated nuclear data libraries.
a type of analysis often necessary in nuclear physics experiments; the different peaks can represent the number of reactions with a certain reaction product, for example. Some shape is normally assumed for the peaks, typically a normal distribution. In Paper VI, the framework of treating model defects with GPs is used to fit arbitrarily shaped peaks, by viewing them as peaks with a known shape with a defect added to them.

Paper VII (summarized in Sec. 5.2) looks into model defects for ND evaluation, and develops a different approach to treat model defects using GPs. Instead of adding a defect term outside the model, we let the model parameters depend on an input parameter, namely, the neutron energy, and ensure this dependence to be “smooth” by using GPs. This guarantees certain physical rules to be followed, without the need for constructing complicated covariance functions (terminology used for GPs). It is also relatively easy to incorporate in the TENDL [53] framework by using existing code packages to produce files of the ENDF format.

In fact, a small study using GPs to treat model defects in UMC-B is included already in Paper V. This part was not investigated as systematically as the other techniques, but the results were promising. Both Papers VI and VII use GPs together with the Levenberg-Marquardt (LM) algorithm, following the discussion in Sec. 3.5. As Paper V, both Papers VI and VII make use of synthetic data to enable a comparison to the truth. In this thesis, examples using authentic data are also presented, however.

5.1 Treating model defects with GPs in the observable domain

5.1.1 The studied problem

Paper VI was written to treat an often faced problem in the analysis of nuclear physics experiments: to distinguish peaks in a histogram of data. The situation is illustrated by Fig. 5.1, which is discussed more in the next paragraph. The peaks can represent the number of reactions with a certain reaction product; the concrete example initiating the study is such measurements of isomeric fission yields [82], where different isomers\(^2\) of a fission product are to be distinguished. Imperfect measurement resolution makes the peaks have a certain dispersion, causing the peaks to overlap more or less. The peaks are often assumed to follow normal distributions (Gaussians), which to some extent can be motivated by the central limit theorem. The assumption is often a reasonable approximation, at least.

Fig. 5.1(a) depicts an ideal situation, where data are generated from three Gaussian peaks, and a model with three Gaussian peaks is fitted to the data

\(^2\)Isomers are of the same element and with the same mass number, but with different excitation energy, giving them slightly different mass.
Figure 5.1. The type of problem considered in Paper VI. A model with three Gaussian (normal) peaks is fitted to the data, to determine the probability $r_1$ for a reaction to “end up” in the middle peak, given that it is in any of the two rightmost peaks. In (b), the Gaussian assumption is erroneous, which gives a significant discrepancy between the data and the fit. In both cases, the true probability is $r_1 = 0.60$; it is estimated by $\hat{r}_1$, which is printed above the respective sub-figure. The expected number of total counts is 50 000 for this figure.

using the Levenberg-Marquardt algorithm. One of the parameters in the model is $r_1$, the probability for a reaction to one of the rightmost peaks to be in the middle peak.$^3$ Just as the other parameters in the model, $r_1$ is estimated in the fit.

In reality, the Gaussian assumption is not always a good enough approximation; we may have a situation similar to that of Fig. 5.1(b) [cf., authentic data in Sec. 5.1.4]. For this figure, the data are generated from a non-normal distribution, but the model nevertheless assumes Gaussian peaks. This is another example of a model defect (cf., Secs. 3.3.2 and 3.4.2): the truth underlying the data cannot be reproduced by the model. It is obvious from Fig. 5.1(b) that the model defect causes systematic discrepancies between the fit and the data, and the error is not reflected by the uncertainty in the fit (uncertainty bands are included in the figure, but they are barely distinguishable). In Paper VI (and Sec. 5.1.3), it is also seen that this causes a bias in the estimate of $r_1$, which is not reflected by its uncertainty estimate.

In Paper VI, the idea to use GPs to treat model defects is applied to this situation.

$^3$For isomeric fission yields, it could be that the leftmost peak is one nuclide and the two rightmost peaks represent two isomers of the same nuclide; if so, $r_1$ is the isomeric fission yield of the more excited isomer.
5.1.2 GPs to treat model defects

GPs (defined in the next paragraph) can be used to treat model defects in fitting problems by adding a term to the original model, such that Eq. (2.16) becomes

\[ Y = f(x; \beta) + \mathcal{E}_m(x) + \mathcal{E}, \]  

(5.1)

where the newly added term \( \mathcal{E}_m(x) = (\mathcal{E}_m(x_1), \mathcal{E}_m(x_2), ..., \mathcal{E}_m(x_n))^T \), with \( \mathcal{E}_m(x) \) being a GP with a zero mean function (prior to observing the data). We write this as

\[ \mathcal{E}_m(x) \sim \mathcal{GP}(0, \kappa(x, x')) \],

(5.2)

where 0 is the particular mean function and \( \kappa(x, x') \) is the covariance function of the GP, which defines the covariance between \( \mathcal{E}_m(x) \) and \( \mathcal{E}_m(x') \).

To the definition of GPs:

A Gaussian process (GP) is “a collection of random variables, any finite number of which follows a multivariate normal distribution” [80].

It is characterized by its mean and covariance functions, analogously to how a Gaussian random vector is characterized by its mean vector and covariance matrix. One may think of a GP as a function which takes a vector of inputs \( x \) and returns a Gaussian random vector. In the context of model defects, the main point is that it provides a correlation structure to the model defect term. We get

\[ \mathcal{E}_m(x) \sim \mathcal{N}(0, \mathbf{K}), \]  

(5.3)

where the covariance \( \mathbf{K} \) is determined by the covariance function:

\[ (\mathbf{K})_{ij} = \kappa(x_i, x_j). \]  

(5.4)

Thus, the covariance function is the truly essential part for our purposes. It is often appropriate to assume that the covariance function is a “squared exponential”:

\[ \kappa(x_i, x_j) = \sigma^2 e^{-\frac{(x_i - x_j)^2}{2\ell^2}}. \]  

(5.5)

This expression introduces two hyper-parameters\(^4\): \( \sigma \) and \( \ell \), where \( \sigma \) is the marginal standard deviation, and the correlation length \( \ell \) controls how fast the correlation decreases as the distance between \( x_i \) and \( x_j \) increases. These hyper-parameters can be determined from the data, e.g., using cross-validation as in Papers VI-VII.

In Paper VI, the covariance function is based on several covariance functions such as that of Eq. (5.5). To be able to estimate \( r_1 \), it is necessary to attribute different parts of the defect to the different peaks. Therefore, it is assumed that all peaks have the same shape, but the particular shape is arbitrary.

\(^4\)The prefix “hyper” is inserted to distinguish these parameters from the model parameters in \( \beta \).
Figure 5.2. An example of a covariance function used in the work of Paper VI, which results from the assumption that all peaks have the same shape.
The defect of each peak is modeled to follow Eq. (5.5), giving a specific form for the full covariance function. An example is shown in Fig. 5.2.

With $\epsilon_m(x)$ being a Gaussian vector, the model parameters can be estimated as if the experimental errors followed $\epsilon_m(x) + \epsilon$, which is also Gaussian [i.e., using Eq. (2.26) with $\Omega$ replaced by $\Omega + K$]. Simultaneously, the model defect term $\epsilon_m$ can be fitted to the data at some choice of grid $x^*$, which need not be equal to $x$.

In a usual fit (without the model defect part), we get a posterior distribution for the model parameters and for the errors. The latter is less often considered, but the errors are distributed around the residuals $y - f(x; \hat{\beta})$, posterior to the fit. Now, with the model defect term included, we also get a posterior distribution for the model defect. They are all correlated.

5.1.3 Performance on synthetic data

A situation similar to that shown in Fig. 5.1(b) is seen in Fig. 5.3(a), where a model assuming that the peaks are normal is fitted to non-normal data. Fitting to the same data, but treating the model defect as described in Sec. 5.1.2, the fit in Fig. 5.3(b) is obtained. The fit follows the data much better, and a somewhat better estimate for $r_1$ is obtained (even if the previous estimate for $r_1$ is not remarkable by itself). We can also see, in Fig. 5.3(c), that the shape estimated by the fit with GP is much better than that without; it agrees so well with the truth that the two curves are difficult to distinguish.

For this particular example, the GP treatment of the model defect seems to work well. However, it is more interesting to consider the overall performance for a large number of data sets such as the one above. Similarly to in Paper V, $M = 10^4$ sets of data are generated, and the fits are carried out for each case. This is repeated for varying statistics, quantified by different numbers of expected counts in all the three peaks. Doing so, and focusing on the estimates...
for $r_1$, we get the results that are summarized in Fig. 5.4. The results are compared to results without the GP treatment.

Fig. 5.4(a) shows the bias of $\hat{r}_1$, which is the average deviation from the true value for $r_1$. For the poorest statistics, there is a bias both with and without the GP treatment, of similar magnitude. This bias is, however, much smaller than the uncertainty of $\hat{r}_1$. With the GP treatment, the bias decreases in magnitude to values close to 0, as $\langle N \rangle$ increases. Without the GP treatment, the bias worsens.

Fig. 5.4(b) shows the mean $\chi^2$ comparing the different $\hat{r}_1$ to the true $r_1$, which should be close to 1 for a good method. Even if Fig. 5.4(a) showed a bias for the poorest statistics, the uncertainty of $\hat{r}_1$ is so large for such poor statistics, that the reduced $\hat{\chi}^2_{\text{true},r_1}$-values are close to 1, both with and without the GP treatment. With the GP treatment, the $\hat{\chi}^2_{\text{true},r_1}$-values remain close to 1 as $\langle N \rangle$ increases. Without GP treatment, however, the $\hat{\chi}^2_{\text{true},r_1}$-values increase.

Overall, we can conclude that the fits using the GP treatment allowing for arbitrarily shaped peaks give results that are distributed close to what is desired. Without the GP treatment, improving statistics reveal that the results are not distributed as they should; the uncertainties become underestimated. In
the paper, it is also seen that if one considers the fitted curves, the errors in the fits without GP are revealed already for poor statistics.

In addition to this usage of GPs, Paper VI also studies a few other aspects of a fit in the situation described in Sec. 5.1.1. For example, it is shown to be important to include the uncertainty of prior knowledge, for good statistics. For poor statistics, on the other hand, it is important to iteratively use the fit to update the uncertainties from counting statistics.

5.1.4 An example with authentic data

In Paper VI, only synthetic data are considered, to really be able to study how the method reproduces the truth underlying the data. However, it is interesting to look at an example with real data (from Ref. [83]), and see if there is an impact. The particular data set is chosen because it resulted in relatively poor fits assuming Gaussian peaks, and because it had quite good statistics. The results of this section have not been reported previously.

The data are from a measurement carried out with a Penning trap [84], and they are observed in terms of a “reduced cyclotron frequency” \( z \), which is assumed to be related to the nuclide mass \( m \) following

\[
z = \frac{C}{m} - z_-, \tag{5.6}
\]

where \( C \) and \( z_- \) are constants, i.e., independent of mass (for \( z_- \), this is a first order approximation, and also assumes an ideal Penning trap). The fit then uses the model

\[
f(z; \beta) = B + A_0 f(z - z_0, \sigma) + A_{12} [r_1 f(z - z_1, \sigma) + (1 - r_1) f(z - z_1, \sigma)],
\]

where \( B \) is a constant background and \( f \) describes a peak (based on a Gaussian with mean 0 and variance \( \sigma^2 \)). The reduced cyclotron frequencies defining each peak can be related to the mass of the peaks using Eq. (5.6):

\[
z_k = \frac{C}{m_k} - z_-; \quad k \in \{0, 1, 2\}. \tag{5.8}
\]

Further, \( m_1 = m_2 + E_{\text{exc}}/c^2 \), and the model parameters to be fitted are \( \beta = (B, C, z_-, \sigma, A_0, m_0, A_{12}, m_2, r_1, E_{\text{exc}})^T \). The masses \( m_0 \) and \( m_2 \), as well as the excitation energy \( E_{\text{exc}} \) and the magnetron frequency \( z_- \) are known up to some uncertainty, so they are given a prior distribution following tabulated values for the three former [85, 86], and a reading from the experiment for the latter \((1654 \pm 165) \text{ Hz}\).

Using this model, and a Gaussian assumption, for a certain set of data, gives Fig. 5.5(a). It is obvious that the fit does not agree well with the data, and the
Figure 5.5. Results from fits to authentic data, if Gaussian peaks are assumed, or if GPs are used along the lines of Paper VI. The leftmost peak consists of \(^{96}\text{Sr}\) nuclides, and the two peaks to the right are for two isomers of \(^{96}\text{Y}\). The rightmost peak is the ground state and the middle peak is the first excited state, which is slightly heavier because of its excitation energy.

The estimate for \(r_1\) changes a bit, too. The difference is about twice the uncertainty of one fit, so it is not extreme, but still indicates a systematic error in the first estimate which is greater than the uncertainty estimate. Both the lower \(\chi^2_{\text{LOO}}/\nu\) and the study in Paper VI suggests that we should use the GP treatment. The number of counts is more than 12 000, so Fig. 5.4 indicates that the error can be reduced substantially by applying the method.

5.2 Treating model defects with GPs in the parameter domain

The covariance functions in Paper VI are not as uncomplicated as the squared exponential in Eq. (5.5), because of the constraint that all peaks must have the same shape. In ND evaluation, the model is much more complicated than in
Paper VI, and the constraints that need to be fulfilled are more complicated, too. In particular, there are several cross sections that can be written as sums of other cross sections in certain energy intervals. As a result, the covariance functions in ND evaluation must be constructed carefully, along the lines described in Ref. [10].

As briefly described in the introduction to the chapter, another approach using GPs to treat model defects is developed in Paper VII, where model parameters are allowed to depend smoothly on an input variable. In the case of ND evaluation, the model parameters can depend on the incoming neutron energy $E$. If so, all physical constraints inferred by the model are exactly followed for each $E$. Further, it would be relatively easy to incorporate such a parameter based methodology into one of the success factors for TMC and TENDL [53]: the carefully tested route from a distribution of model parameters to complete ENDF files, either “random files” or one file with covariances. In fact, there is a capability for letting many of the parameters in TALYS to be energy-dependent [12]; the challenge is to construct a reasonable energy dependence.

### 5.2.1 Summary of the method

In Paper VII, the model parameter vector $\beta$ of the considered model $f(x; \beta)$ is replaced by an energy-dependent parameter vector defined as

$$\gamma(E) = \beta + \delta(E), \quad (5.9)$$

with, prior to observing the data, the difference to the “global model parameters” $\beta$ is modeled as

$$\delta(E) = (\delta_1(E), \delta_2(E), \ldots, \delta_l(E))^T, \quad (5.10)$$

where the individual $\delta_j(E)$ are mutually independent GPs with a zero mean function:

$$\delta_j(E) \sim GP(0, \kappa_j(E, E')). \quad (5.11)$$

The covariance functions $\kappa_j(E, E')$ are assumed to follow squared exponentials [see Eq. (5.5)].

In practice, the energy-dependent parameter vector is considered at an energy grid $E^* = (E_1^*, E_2^*, \ldots, E_n^*)^T$, and we define a new parameter vector as
“γ(E) evaluated at the grid E∗”:

\[
\gamma_{E^*} = \begin{pmatrix}
\gamma_1(E_1^*) \\
\gamma_1(E_2^*) \\
\vdots \\
\gamma_1(E_n^*) \\
\gamma_2(E_1^*) \\
\vdots \\
\gamma_l(E_1^*) \\
\vdots \\
\gamma_l(E_n^*) \\
\end{pmatrix} = \begin{pmatrix}
\beta_1 + \delta_1(E_1^*) \\
\beta_1 + \delta_1(E_2^*) \\
\vdots \\
\beta_1 + \delta_1(E_n^*) \\
\beta_2 + \delta_2(E_1^*) \\
\vdots \\
\beta_l + \delta_l(E_1^*) \\
\vdots \\
\beta_l + \delta_l(E_n^*) \\
\end{pmatrix}.
\]

(5.12)

5.2.2 Applied to ND evaluation with synthetic data

For the testing of the methodology on an ND evaluation problem, we use “Pseudo-TALYS”. It is a model based on TALYS, but it executes faster, such that a study with a large number of realities such as in Papers V-VI is achieved with reasonable computational resources. Even if it is a lot simpler than TALYS, it is intended to contain a lot of the complexity of TALYS from a model fitting point of view: it is non-linear, and returns results for 17 different cross sections. Some of these cross sections are constructed as sums of others, and yet others are nevertheless correlated via the model parameters. The nuclide under consideration is $^{56}$Fe.

For each out of 2000 considered “realities”, a true set of cross sections is generated based on Pseudo-TALYS with sampled model parameters. However, as in Paper V, a model defect is “sampled around 0”, and added to the cross sections. Data are then generated around this truth, using an input structure and data covariances based on the experimental data available in EXFOR. The method with energy-dependent model parameters is applied to fit Pseudo-TALYS to this data.

When discussing the results, fits with $E$-dependent parameters are referred to as local fits, and they are compared to global fits, i.e., fits without the $E$-dependent parameters. For a particular reality, both the global and local fits are shown for a few cross sections in Fig. 5.6. The reality is chosen to be representative of cases with larger model defects. The $(n,tot)$ and $(n,n')$ cross sections for another such reality is found in Paper VII. Just as in the example shown in the paper, the local fit is generally closer to the truth, and especially if the
Figure 5.6. The truth, data, and both global and local fits to the data for a few cross sections, from the work of Paper VII. The lower part of each sub-figure shows the fits’ deviations from the truth normalized by their uncertainty; in this part, the axis is linear between $-2$ and $2$, and logarithmic otherwise.
uncertainties are taken into account: the deviation from the truth normalized to the uncertainty is between $-2$ and a little more than $2$ everywhere in the figure. In this case, and in many others, the local fit agrees remarkably well with the truth even where there is no data; see, for example, the $(n,n')$ cross section. This can be explained by the correlation to other interactions such as $(n,tot)$, where there is data. It is possible that similar behavior would be seen if the more conventional type of GP treatment (see Sec. 5.1) was used, but it may also be that the stronger correlation via the model parameters increases predictability.

The global fit may not look that bad where there is data, but we can see that the uncertainties often are strongly underestimated. The underestimation is worst where there is no data, but the uncertainties are poor even in other cases, for example around 10 MeV for the $(n,p)$ cross section. These results can be attributed as another example of poor results when there are untreated model defects. To summarize the particular example illustrated in Fig. 5.4, one could say that the local fit performs very well, and the global fit does not.

Just as in Papers V and VI, the most interesting results are obtained when considering all the different realities as a whole. It turns out that the example discussed above is quite representative: the local fits perform much better than the global fits. For example, the top half of Fig. 5.7(a) shows the distributions for $\chi^2_{\text{true}}$, the $\chi^2$-values comparing the fits including their uncertainties to the truth point by point for all points on a rather dense energy grid, over all interactions. If the fits, and their uncertainties, would be distributed perfectly, the histograms would follow a $\chi^2$-distribution with 1 degree of freedom, and such a distribution is also seen in the figure. One sees that the distribution for the global fits is far from the desired distribution (note the log-scale!), with a substantial amount of $\chi^2_{\text{true}}$-values ranging up to a few thousands. The local fits, on the other hand, give a distribution which resembles the desired, even if the tails are somewhat too thick. Thus, both some underestimation and overestimation of uncertainties occur. Looking at the distributions for individual interactions, some agree better than average, such as $(n,p)$ in the lower half of Fig. 5.7(a). The distributions are summarized by their mean values $\bar{\chi}^2_{\text{true}}$ in the top half of Fig. 5.7(b), interaction by interaction. The global fits show poor results for most interactions, while the local fits have values relatively close to one, even if most values deviate from one with statistical significance.

An analogous comparison, but for an integral over the whole energy range, is seen in the lower half of Fig. 5.7(b). For the “all*” case, a vector including the integrals of all cross sections that are not sums of the others except inelastic scattering to individual excited states $(n,n'_1)$ is considered (and using the generalized $\chi^2$). Overall, both the local and global fits perform better than for the point-by-point comparison. In particular, many of the $\bar{\chi}^2_{\text{true}}$-values are now insignificantly off from one. The improvement for the global fits still does not make them good enough.
Density

(a) The distributions of the logarithm of the pointwise $\chi^2_{true}$ for all points in all interactions (top), and for all points in the (n,p) cross section (bottom), both for the global and local fits.

(b) The mean $\chi^2_{true}$ for the pointwise (top) and integral (bottom) comparisons, for all considered interactions. In the pointwise case, division by zero has lead to that the "rest" and total are missing for the global fits.

Figure 5.7. Selected results from the work of Paper VII considering all 2000 realities.

To conclude, the method with $E$-dependent parameters improves the fits’ ability to reproduce the truth with a remarkable amount compared to usual model fitting. Thus, it provides a very promising route to improving ND evaluation.

5.3 Conclusions for Chapter 5

This chapter summarizes two papers which study two different ways to use GPs to treat model defects. The approach of Paper VI adds a defect term, modeled by a GP, to the original model. In the other approach, used in Paper VII, the defect is added in the parameter domain; the parameters are allowed to vary with energy. The difference to an overall parameter value is modeled by a GP. In this way, flexibility is added to the model without leaving the model, considering one energy at a time. One advantage of such a method is that it would be relatively straightforward to include in the TENDL production framework.

The two approaches are applied to two quite different situations. In each of the two cases, the usage of GPs outperforms the method they are compared to, i.e., not treating the model defects. This may be a somewhat unfair competition, and it would be compelling to compare the two on an equal footing, e.g., applying the approach of Paper VI to the situation studied in Paper VII. The
method of Paper VII could also be compared to other techniques, in particular to GLS which is at least less sensitive to model defects, as seen in Paper V, and to GLS augmented with a GP treatment of model defects.

Paper VI considers a situation in which different peaks are to be distinguished in a histogram of data, often used to separate different reaction products, for example. The GPs enable the fitting of arbitrarily shaped peaks, and it is shown that this can give an essential improvement to the resulting estimates, and we also approach virtually unbiased results with adequate uncertainty estimates as the statistics improve.

The other method is applied to a situation closely related to ND evaluation. A model similar to TALYS from a fitting point of view is fitted, with and without the energy-dependent parameters. It is seen that the distribution of results is improved by orders of magnitude compared to fitting without the energy-dependent parameters.

In both papers, large sets of synthetic data are used, such that it is possible to compare the results to the underlying truth. Specific forms of the model defects are assumed, and it is difficult to say how representative they are. In this chapter, results applying the method of VI to authentic data are also presented. It gives a large improvement to the fit, but not a perfect $\chi^2$, probably indicating that the underlying shapes of the peaks are not entirely equal. This could possibly be treated by adding another model defect term.

To conclude, if model defects are present they need to be treated, and GPs are shown to be a viable option for both of the studied situations. In the next chapter, the method of Paper VII is applied both using authentic data, and using the real TALYS.
6. GPs in the parameter domain with authentic data and TALYS

This chapter stands out from Chapters 3-5, as it is not summarizing any of the papers listed on page v. Instead, it is a continuation of the work in Paper VII, but using a proper model code, TALYS, and authentic data. Since the work is not reported elsewhere, a lot more details are included in the description. Practicalities when using real data are discussed in Sec. 6.1. The additional considerations necessary when using the methodology for TALYS are described in Sec. 6.2, which also reports the resulting fits.

There are several details in this chapter which should be further developed, so the results are of a preliminary nature, and should more be seen as a proof of concept. As in Paper VII, all numerical examples are for $^{56}$Fe.

6.1 Data-driven evaluation and reduction of experimental data

As discussed in connection with Papers I-III, the use of authentic data calls for some practical considerations. The data can be inconsistent, and the uncertainty information is often limited to random uncertainties. For comprehensive work such as TENDL, the data has to be interpreted automatically; the TENDL project currently uses a method which is not entirely well-founded, statistically, as discussed in Secs. 2.4 and 3.2.6. In Papers I and II, attempts were made to use the information in EXFOR, and a set of simple rules, to automatically construct covariance matrices, with limited success.

Here, another approach, inspired by Refs. [87, 88], is taken. It is entirely driven by the data, and assigns random and systematic uncertainties based on the agreement between data points and sets, respectively. We also propose a method which automatically adds uncertainties to data sets which are not confirmed by other sets. First, it is observed that a substantially reduced set of data can be used because the model is interpolated, and the preprocessing incorporates this reduction of data.

6.1.1 Reducing the data

When using a proper nuclear reaction model such as TALYS, it is often not feasible to run TALYS for all energies where there are experimental data, i.e.,
it is necessary to interpolate the results from a computational grid to the energies of the experimental data. Linear interpolation is used because it conserves the summing of individual interactions, even if there are more sophisticated interpolation techniques which give smooth curves with continuous derivatives. Below, we see that we can actually reduce the data to much fewer data points using GLS to a grid based on the interpolation grid, without losing any more information than what is already lost because of the interpolation.

Before showing that it is equivalent to use the reduced data, we present how the data are reduced in practice. The reduced set of data can be obtained by choosing an appropriate reduced grid $X_r$ (see below), and then using GLS (see Sec. 3.2.2) to fit the data onto this grid (without a prior). If $S$ is the matrix interpolating the data from the reduced grid $X_r$ to the original experimental input $X$, the reduced data are obtained through

$$y_r = [S^T \Omega_0^{-1} S]^{-1} S^T \Omega_0^{-1} y,$$

where $\Omega_0$ is the covariance matrix of the original data. The covariance matrix of the reduced data is

$$\Omega_r = [S^T \Omega_0^{-1} S]^{-1}.$$

The reduced data can have much fewer points, so the computational complexity of the problem is reduced by using it. An example is seen in 6.1(a), where 6371 points for the total cross section is reduced into 38 points, assuming a computational grid with 41 equidistant points and a point close to the threshold of $(n,n_1)$. This is the computational grid used in practice all throughout this chapter.

From the figure, we may get the impression that the reduced uncertainties are too small in comparison to the spread of the original data; that they, therefore, may give too small uncertainties in a following model fit. Note, however, that using the original data gives the same uncertainties in the fit of the model. The possibly too small uncertainties are discussed further (and treated) in Sec. 6.1.2.

It remains to choose the reduced grid $X_r$ such that it fulfills two criteria. First, it must be possible to solve for the reduced data, i.e., $S^T \Omega_0^{-1} S$ must be invertible. This means that, for example, if there is only one experimental point, this must also be the reduced data point. Also, the reduced data shall not result in interpolations over unnecessarily long ranges. Therefore, we base the reduced grid on the computational grid $E^\dagger = \left( E_1^\dagger, E_2^\dagger, ..., E_{n^\dagger}^\dagger \right)^T$ for each interaction. If there are at least two distinguished experimental points on an interval $[E_i^\dagger, E_{i+1}^\dagger]$, the points $E_i^\dagger$ and $E_{i+1}^\dagger$ are included in the reduced grid. If there is only one energy $E_{\text{exp}}$ with experimental data in $[E_i^\dagger, E_{i+1}^\dagger]$, the

\footnote{Note that the interpolation is not done between observed data points; it is done between model calculations.}
energy $E_{\text{exp}}$ itself is included in the reduced grid, instead. In this way, $S$ has full rank which gives that $S^T \Omega_0^{-1} S$ is invertible [28], provided that $\Omega_0^{-1}$ is invertible (which it has to be in any case).

Using the reduced data gives a smaller computational burden, which is especially important when inverting proper covariance matrices. It makes the estimation of systematic uncertainties described in Sec. 6.1.2 much more efficient. To the author’s knowledge, this reduction is not a used procedure in, e.g., TENDL.

**Equivalence of fitting to reduced data**

This subsection is devoted to showing that the claim above actually holds: that using the reduced data is equivalent to using the original data. Formally, the proof applies to the LM algorithm, but by renaming the involved matrices and vectors, it should be applicable to GLS, too. An extension to Monte Carlo methods has not been considered, but the author believes that this would work, too.

Under the premises above, and if we fit the interpolated model to the original data set, the data $y$ is compared to $Sf$, where $f$ is the function evaluated at the reduced grid $x_r$ ($f$ can be interpolated if necessary; it does not matter if we linearly interpolate once more). Then, the iterations in the LM algorithm follow (the notation is detailed in Paper VI)

$$
\beta^{\text{proposal}} = \beta^{(k)} + \left( J_0^T \Omega_0^{-1} J_0 + P^{-1} + \lambda D \right)^{-1} \cdot \left[ J_0^T \Omega_0^{-1} (y - Sf) + P^{-1} \left( p - \beta^{(k)} \right) \right],
$$

where the subscript 0 on $\Omega_0$ and $J_0$ denotes that they refer to the original data. In the notation, we omit that the Jacobian and the function depend on $k$. 

Figure 6.1. Original, reduced, and adjusted data for the total cross section.
If we fit the model to the reduced data obtained by Eq. (6.1), we get

\[
\beta^{\text{proposal}} = \beta^{(k)} + \left( J_r^T \Omega_r^{-1} J_r + P^{-1} + \lambda D \right)^{-1} \cdot \left[ J_r^T \Omega_r^{-1} (y_r - f) + P^{-1} \left( p - \beta^{(k)} \right) \right],
\]  

(6.4)

where the subscript \(r\) denotes that \(\Omega_r\) and \(J_r\) refer to the reduced data. Since \(I = \left[ S^T \Omega_0^{-1} S \right]^{-1} \), we can write

\[
y_r - f = \left[ S^T \Omega_0^{-1} S \right]^{-1} S \Omega_0^{-1} y - \left[ S^T \Omega_0^{-1} S \right]^{-1} S \Omega_0^{-1} Sf
\]

(6.5)

where we used Eq. (6.2) in the final step. Inserting Eq. (6.5) and the inverse of Eq. (6.2) into Eq. (6.4), we get

\[
\beta^{\text{proposal}} = \beta^{(k)} + \left( J_r^T S^T \Omega_0^{-1} S J_r + P^{-1} + \lambda D \right)^{-1} \cdot \left[ J_r^T S^T \Omega_0^{-1} (y - Sf) + P^{-1} \left( p - \beta^{(k)} \right) \right]
\]

(6.6)

Finally, because the \(j\)th column of \(J_0\) is

\[
(J_0)_j = \frac{\partial}{\partial \beta_j} Sf = S \frac{\partial}{\partial \beta_j} f = S(J_r)_j,
\]

(6.7)

we have

\[
J_0 = SJ_r.
\]

(6.8)

Replacing each occurrence of \(SJ_r\) by \(J_0\) in Eq. (6.6) yields Eq. (6.3) [using that \((SJ_r)^T = J_r^T S^T\)]. Thus, we can just as well use the reduced data as the original data, and the reduced data gives a smaller computational burden.

### 6.1.2 Estimating unknown uncertainties

In Papers I and II, attempts were made to automatically generate experimental covariances based on a set of simple rules which were based on coarse assumptions on the experimental setups. Even though the results of these papers are difficult to interpret because of many different possible problems with the work, it could be seen from comparing the two papers that the assumptions made are very important for the results. Further, the approach to detect outliers was, to be frank, founded on a circular argument: it was based on the deviation to the prior of the model which we intended to fit.
Here, we consider a method to estimate unknown random and systematic uncertainties which is more related to the work of Refs. [87, 88], adjusting the uncertainties based on the agreement to other data. The method effectively detects outliers and reduces their importance. It is based on the reduction of data described in the previous subsection, and assumptions on the covariance structure of the experiments. Parameters determining the covariance within this structure are then fitted to maximize the likelihood of the observed distribution of data. Since there are resonances reaching into the considered energy range and we consider the average cross sections, the method is also adjusted to take an additional “random” uncertainty into account for lower energies.

Even if the approach here has many advantages over the approach taken in Papers I and II, there are some aspects that need to be investigated further, and the results are preliminary, as stated in the introduction to the chapter. Also, the method should be compared to those in Refs. [87, 88], at least qualitatively.

The estimation of uncertainties proceeds in two steps.

1. First, each experiment is treated separately, for two reasons:
   - To detect underestimated random uncertainties.
   - To reduce the number of data points in each experiment.
2. The next step takes the individually reduced and preprocessed experiments as input. In this step, each interaction is treated separately with two purposes:
   - To estimate systematic uncertainties based on a certain correlation structure.
   - To reduce the data onto a common, final, energy grid.

In Sec. 6.1.3, uncertainties are also added to isolated experiments.

### Reducing each experiment and estimating random uncertainties

In this subsection, the data vector \( \mathbf{y} = (y_1, y_2, \ldots, y_i, \ldots, y_n)^T \) refers to the original data of one single experiment. At this stage, the points are only associated with individual random uncertainties \( \sigma_{i,0} \).

As mentioned in Sec. 6.1.1, the reduced uncertainties may appear too small in comparison to the spread of the original data. For these energies, there are resonances, but we are interested in the average cross section (which is what is modeled by TALYS). In principle, it could be that the seemingly small uncertainties are due to the smaller variance of an average compared to individual observations [cf., Eq. (2.7)]. Among the resonances, however, the variation in the data is much greater than the experimental uncertainties (even if they are correctly estimated). It is, therefore, reasonable to base the uncertainty on the actual dispersion of the data, and not only on the reported experimental uncertainties, i.e., a random uncertainty component is added to account for the presence of resonances. Since there is a risk of underestimated uncertainties in general, we apply the same approach to the whole energy region. To distinguish between random and systematic uncertainty, the procedure is applied to each experimental set at a time. Simultaneously, each set of data is reduced as
described in Sec. 6.1.1, resulting in more manageable data covariance matrices in later steps.

The estimation of random uncertainties is done similarly to how the data variance is estimated in ordinary least squares (OLS [30]). In the \( j \)th interval, the variance of each point \( y_i \) is set to

\[
\sigma_i^2 = \max \left( \sigma_{i,0}^2, \frac{1}{n_j - n_{dof}} \sum_{i'} (y_{i'} - y_{i'}(y_r))^2 \right),
\]  

(6.9)

where \( \sigma_{i,0}^2 \) is the reported variance, \( y_{i'}(y_r) \) is the value corresponding to \( y_{i'} \) obtained by linear interpolation of the reduced data, and \( i' \) ranges over the \( n_j \) data points in the interval. The number of degrees of freedom of the fit, \( n_{dof} \), is assumed to be 1 if there are two neighboring intervals which are used in the fit of the reduced data, and otherwise \( n_{dof} = 2 \). With the new set of uncertainties, the reduced data are reattained, which gives new results from Eq. (6.9), in turn giving a new \( y_r \). The procedure is iterated over until convergence.

A simple example illustrating a maximum likelihood (ML) approach to estimation of additional uncertainties

The second step mentioned in the beginning of 6.1.2 is best introduced by a simple example. How the method is applied to estimate systematic uncertainties for the actual data is presented in the next subsection.

Consider three uncorrelated data points measuring the same quantity, reported as \( y_i \pm \sigma_i^0, \) for \( i \in \{1, 2, 3\} \). Assuming normal distributions, their PDFs are illustrated with black solid curves in Fig. 6.2(a). One of the points is located far away from the others, compared to the uncertainty; it is probably an outlier, and the set of data is highly inconsistent. If, nevertheless, the data are assumed to be reported correctly, the solid red curve is the resulting least squares fit based on this. Even if the uncertainty of the fit is included, the \( \chi^2 \)-value becomes more than 500, making it extremely unlikely that the data are correctly reported.

If it is assumed that there may be uncertainties missing in the reported values, such that the data points actually follow \( y_i \pm \sigma_i^0 \pm \sigma_i^+ \), the likelihood for the observed data follows from the PDF of a multivariate normal distribution:

\[
f_{Y | \{\sigma_i^+\}_{i=1}^3} (y) = \frac{e^{-\chi^2/2}}{(2\pi)^{3/2} \sqrt{\det \left( \Omega_r S^T + \Omega \right)}},
\]  

(6.10)

with

\[
\chi^2 = (y - S y_r)^T \left( S \Omega_r S^T + \Omega \right)^{-1} (y - S y_r),
\]  

(6.11)

where \( y_r = [S^T \Omega^{-1} S]^{-1} S^T \Omega^{-1} y, \ \Omega_r = [S^T \Omega^{-1} S]^{-1}, \ S = (1, 1, 1)^T, \) and \( \Omega \) is the diagonal matrix with \( \sigma_i^{02} + \sigma_i^{+2} \) along the diagonal. Note that \( \Omega \) depends on the \( \sigma_i^+ \), such that \( y_r \) and \( \Omega_r \) also do so.
The $\chi^2$ in the exponent decreases as any of the additional uncertainties $\sigma_i^+$ increases, resulting in an increasing numerator. However, the denominator also increases with increasing $\sigma_i^+$; the PDF becomes more “spread out”. The numerator is limited above, but the denominator is not, so there is a trade-off between the two. There may be several local maxima of Eq. (6.10), each with its own interpretation of the data; one or the other may have underestimated uncertainty. For the preliminary results of this work, we content ourselves with finding one local maximum. Undesired effects of this are, to some extent, treated by adding uncertainties to “isolated” experiments in Sec. 6.1.3.

For the particular example considered here, the found (local) ML estimate of the additional uncertainties are illustrated by showing the corresponding distributions as dashed gray curves in 6.2(a). The uncertainty is increased a lot for the rightmost point (it is practically rejected; it is barely possible to even distinguish the top of the new curve), increased a bit for the leftmost point, and left unchanged for the middle point. The resulting fit is close to the middle point; it is moved only slightly to the left with a somewhat smaller uncertainty than the point itself.

Intuitively, we appreciate the de facto removal of the outlier to the far right, but may not be entirely happy with the unequal treatment of the two mutually consistent points to the left. In fact, there is another local maximum with almost the same likelihood for which the roles of these two points are interchanged. There is even a local maximum where the rightmost point is unaffected and the two leftmost points are given large additional uncertainties; this has much lower likelihood, though. These solutions should be weighed together [26], but this is not done here. As mentioned above, this is treated (to
some extent) by adding uncertainties to “isolated” points in Sec. 6.1.3. The result of such a treatment in this case is seen in Fig. 6.2(b). Here, an uncertainty is added to the previously unaffected point, making the two points approximately equally important (which is not without a portion of “luck”).

For reference, we can mention that instead of searching for a maximum of Eq. (6.10), we search for a minimum of

\[ \chi^2 + \ln \left[ \det \left( S \Omega S^T + \Omega \right) \right], \]

which is obtained by taking negative of the logarithm of Eq. (6.10) and omitting a constant.

**Estimating systematic uncertainties of experimental sets**

After the primary preprocessing of each data set [i.e., the reduction and the use of Eq. (6.9)], the maximum likelihood approach described for the simple example above is applied to estimate systematic uncertainties of the experiments, for one interaction at a time. The goal is to obtain data that are consistent, and that points that are contradicted by other points shall get increased uncertainties. This is done in connection with an additional reduction of data along the lines described in Sec. 6.1.1, giving an even sparser grid than before. A parametrized covariance structure is assumed for each experiment, and an ML estimate of these parameters is sought: the parameters are adjusted such that different experiments agree well enough with each other; at the same time, “unnecessarily” large uncertainties are ruled out because they spread out the PDF of the data too much.

For each experiment \( \nu \), we assume that the covariance follows

\[ \Omega_{\nu} = \Omega_{\nu}^0 + K_{\nu}, \]

where \( \Omega_{\nu}^0 \) is the covariance matrix obtained from the previous step, and the structure of the added covariance matrix \( K_{\nu} \) depends on the number of points \( n_{\nu} \) in the experiment (because more parameters than points would be redundant):

\[
(K_{\nu})_{ij} = \begin{cases} 
  a_{\nu}^2; & n_{\nu} = 1 \\
  a_{\nu}^2 e^{-\frac{(E_i - E_j)^2}{2b_{\nu}^2}}; & n_{\nu} = 2 \\
  a_{\nu}^2 e^{-\frac{(E_i - E_j)^2}{2b_{\nu}^2}} + c_{\nu}^2 f_{\text{ref.}}^2(E, c); & n_{\nu} \geq 3
\end{cases}
\]

where \( a_{\nu}, b_{\nu} \) and \( c_{\nu} \) are parameters to be estimated, and \( f_{\text{ref.}}(E, c) \) is a reference function for the cross section, taken as the default TALYS result. Thus, \( c_{\nu} \) describes an uncertainty which is relative to this reference, and fully correlated over the experiment. Such an uncertainty could result from an uncertainty in a

---

This means that cross-experimental correlations are disregarded, as discussed in Sec. 6.1.4
sample thickness, for example. Often, relative uncertainties are assumed to be relative to the experimental value, but this yields, at least, two problems. First, it leads to a bias towards lower values, because a point which happens to be low gets a lower uncertainty, and vice versa.\footnote{A similar issue, with Poisson distributed uncertainties, is discussed in Paper VI, where the effect of biasing towards smaller observations is quantified, compared to if the covariances are updated using the fit.} Also, this is identified as one of the main reasons for the so-called Peele’s Pertinent Puzzle [89] in Ref. [90].

The term including $\nu$ and $b_{\nu}$ is a squared exponential, the same as the covariance function used for the GPs to treat model defects in this thesis. This term is intended to approximate various systematic components that are not fully correlated over the whole energy range, e.g., treatment of background and detector efficiencies.

Another, random, covariance component is added for points with energy below 6 MeV, because of the resonances present there. The estimation based on one experiment at a time may not be enough; for example, it fails if the points of one experiment are sparse in the region. If this would be included in the systematic covariance structure above, discrepancies in this range can dominate over the actual systematic uncertainties. This component is not modeled by another parameter per experiment or point; instead, there is one parameter per computational grid point in this region, and the uncertainty of a particular point is linearly interpolated between the grid points, if necessary.

As an example, consider the (n,tot) cross section depicted in Fig. 6.1. There are four experiments, two with one point each and two with thousands of points (reduced down to 27 and 38, respectively). There are eight grid points below 6 MeV. All in all, this results in $1 + 1 + 3 + 3 + 8 = 14$ parameters for the additional experimental covariance.

With this parametrized covariance structure, the parameters are optimized along the same lines as in the previously described simple example, by minimizing Eq. (6.12) with respect to all the parameters. Because only one local minimum is sought, the starting guess for the numerical minimization algorithm will impact the result. The starting guess for $a_{\nu}^2$ is set using the analogue of Eq. (6.9), with $i'$ ranging over the points in the $\nu$th experiment, and with $n_{dof} = 0$ for simplicity. A minimum value of 5% of the mean of the considered cross section is also used, where default TALYS is used as the reference. For $c_{\nu}^2$, the same approach is used, but the variance is estimated for the relative values instead of the absolute (again, relative to default TALYS). The correlation length $b_{\nu}$ is given the starting guess 3 MeV. This choice is not particularly motivated, and can definitely be discussed more.
6.1.3 Decreasing the trust in isolated experiments

When there are different experiments for the same interaction and for similar energies, a comparison between the different experiments can give an idea about missing uncertainties. Quantitatively, they can be estimated along the lines of the ML approach described in Sec. 6.1.2. If there are no other experiments to compare to, this is not possible; the most likely additional uncertainty is zero. For example, this is the case for the experiment consisting of the highest energy point for the (n,p) cross section [see Fig. 6.3(b)]. There are even interactions with only one experiment in the whole energy range. We refer to such experiments as “isolated”. A sound general attitude in science towards results that are not confirmed by reasonably independent sources, is to consider them uncertain.

Here, the aim is to implement such an attitude towards isolated experiments into the automatic preprocessing of experiments. This is done in three steps:

1. We quantify the relative impact of the data point \( y_i \) on the point \( (y_r)_j \) by \( \eta_{ij} \) in Eq. (6.17).
2. The degree of isolation for the \( \nu \)th experiment is inferred from the non-zero \( \eta_{ij} \) of all its points, resulting in the measure \( \tilde{\eta}_\nu \).
3. A penalty is given to the experiment, based on its degree of isolation, according to Eqs. (6.19) and (6.20).

To quantify the degree of isolation for each point, we use the reduced data again.

In this section, we consider the experiment-wise reduced data (obtained after the first step in Sec. 6.1.2) as experimental data; \( y = (y_1, y_2, \ldots, y_i, \ldots, y_n)^T \) covers all such points for the considered interaction. The reduced data \( y_r = y_r(y) \) refers to the final reduced data, i.e., as obtained after the final step in Sec. 6.1.2. Defining \( e_i \) as the \( i \)th unit vector of the same length as \( y \), we can find the sensitivity of the reduced data to each data point \( y_i \) as

\[
\frac{\partial y_r}{\partial y_i} = \lim_{h \to 0} \frac{y_r(y + he_i) - y_r(y)}{h} = y_r(e_i),
\]

where it is used that \( y_r \) depends linearly on the data. Using Eq. (6.1), we get

\[
\frac{\partial y_r}{\partial y_i} = \left( S^T \Omega S^{-1} \right)^{-1} S^T \Omega^{-1} e_i = \left[ \left( S^T \Omega S^{-1} \right)^{-1} S^T \Omega^{-1} \right]_{\cdot i}.
\]

where \( S^- \) denotes the left pseudo-inverse of \( S \), and \( \cdot i \) denotes the \( i \)th column of the matrix. Thus, the reduced data’s sensitivity to each experimental point can be obtained from \( S^- \).

Now, we use this sensitivity to quantify how isolated an experiment is, by determining how important each experimental point is for the reduced data. The idea is that if only one experiment is important for the points in \( y_r \) they affect, we consider the experiment to be isolated. To be precise, we consider
the squared relative sensitivity for each point in $y$, i.e.,

$$\eta^2_{ij} = \frac{[S^-]_{ji}^2}{\sum_{i'=1}^n [S^-]_{ji'}^2}.$$  \hfill (6.17)

To quantify the isolation of an experiment, we use

$$\tilde{\eta}^2_\nu = \sum_i \eta^2_{ij},$$  \hfill (6.18)

where $i$ runs over all points in the experiment and the mean value is taken over the nonzero $\sum_i \eta^2_{ij}$. For experiments with more than one point, the least $\sum_i \eta^2_{ij}$ is removed because two experiments can affect one point in between them without confirming each other much.

All $\eta_\nu$ will be in the interval $[0, 1]$. To all points in the $\nu$th experiment, we add a covariance term

$$\xi(\tilde{\eta}_\nu) f_{\text{ref}}(E_i, c) f_{\text{ref}}(E_j, c) e^{-\frac{(E_i - E_j)^2}{2\lambda_0^2}}$$  \hfill (6.19)

where

$$\xi(\tilde{\eta}_\nu) = \begin{cases} 0.30 \cdot \frac{(\tilde{\eta}_\nu - \tilde{\eta}_0)^2}{(1-\tilde{\eta}_0)^2}; & \tilde{\eta}_\nu > \tilde{\eta}_0 \\ 0; & \tilde{\eta}_\nu \leq \tilde{\eta}_0 \end{cases},$$  \hfill (6.20)

with $\tilde{\eta}_0 = 0.9$ and $\lambda_0 = 3$ MeV. Again, the reference is the default TALYS, except for a lower limit of 10 mb. The covariance term is based on personal judgment, and can be subject to discussion. A potential improvement is to base the maximum penalty of 30\% on the systematic uncertainties obtained for non-isolated points in 6.1.2, perhaps for a large range of nuclides.

In many cases, this penalty for isolation applies to experiments with only one point, which is often totally isolated, i.e., $\tilde{\eta}_\nu = 1$. If so, the above simply leads to an additional uncertainty of 30\% of that point (relative to the default TALYS).

To be exact, an experiment with a large $\tilde{\eta}_0$ need not be isolated. It can also be the case that it just dominates the reduced data completely; the other points have so large uncertainty that the point becomes practically isolated. For the simple example studied above, Fig. 6.2(b) shows the impact of this treatment. After the ML estimation of additional uncertainties, the middle point dominates the fit, resulting in a $\tilde{\eta} = 0.990$, giving an additional uncertainty of 24\% for that point, resulting in an intuitively better combination of the different points, even if it is probably not desirable to have a larger uncertainty for the middle point than for the leftmost point. A similar situation, and behavior, is seen for the higher energy experiments for $(n,2n)$, see Fig. 6.3(a), where the ML treatment increases the uncertainty of the lower Corcalciuc set, while this
treatment of isolated experiments increases the uncertainty of the upper Corcalciuc set⁴, resulting in reduced data which is consistent with both sets. Even if the approach seems to work reasonably well for these cases, it would be better to weigh different local maxima of the likelihood together, as discussed above.

6.1.4 Comments on the resulting data

We have already seen examples of the data resulting from the treatment presented in this section, but let us comment the result a bit more.

The total number of points is finally reduced from 7461 points to 129, making the covariance matrix easily manageable even though systematic uncertainties are added.

For many interactions, there are only a few points, and from one single experiment. In such cases, the consequence of the treatment in Sec. 6.1.3 is that a systematic uncertainty of 30% is added, with a correlation length of 3 MeV, because of the decreased trust in isolated experiments.

In cases where experiments can be compared to each other, a coarse ML estimation of systematic error components of a certain structure is performed. It is coarse in the sense that only a local maximum of the likelihood is sought. Often, the combination with increased uncertainties of isolated experiments yields reasonable results in any case, but it should be stressed that it would be better to combine multiple local maxima. For example, for the (n,tot) cross section [see Fig. 6.1(b)], the Boschung data point is practically rejected because it disagrees with two other sets, and the Cornelis data are provided with a systematic component, and this is all very well. However, the Harvey data does not get any systematic uncertainty from the ML treatment, and it is not

---

⁴This is an example of how results can be poorly reported to EXFOR: two experiments by the same author in the same year are reported with entirely inconsistent results.
isolated because it is confirmed by the other data sets up to 20 MeV. Thus, the procedure results in completely uncorrelated uncertainties for all the points above 20 MeV, and it is not likely that there are no significant correlated uncertainties there. In analogy with the simple example illustrated in Fig. 6.2, there is probably another local maximum in which the Harvey data has a systematic uncertainty; these are two different interpretations of the data.

A peculiar behavior sometimes observed, for example in Fig. 6.3(a), is that there is a reduced data point with a huge uncertainty far from any data point of the individual experiments. This is due to a point being very close to a neighboring reduced grid point. Such a point appears peculiar, but it will not impact a later fit much. Otherwise, the reduced data generally appears reasonably good, as the previously discussed (n,tot) in Fig. 6.1(b), and for (n,2n) and (n,p) in 6.3. The correlation matrices for the final data for the latter two cases are seen in Fig. 6.4. For (n,2n), there are two blocks with quite strong correlation, caused by a few experiments dominating in the different regions. The zero correlation between the blocks origins from the assumption of no cross-experimental correlations (discussed a bit more below). For (n,p), the correlations are generally weaker, reflecting that there are many more experiments for this interaction. There is a tendency of anti-correlation between neighboring points; this is explained by original experimental points in between the reduced points.

Some other aspects than multiple local maxima are neglected. Uncertainties with cross-experimental correlations, e.g., resulting from using the same target or monitor cross section, are disregarded. Since different experiments can have correlated uncertainties, the idea to compare them as independent sources of information is not entirely correct. Quite easily, one could at least add some correlated uncertainty approximating the monitor uncertainty. However, there is even a risk that agreeing experiments suffer from confirmation bias, i.e., that
an experiment disagreeing with previous experiments is scrutinized in more
detail than if it confirms previous results.

The approach described here can be used in combination with certain man-
ual considerations. For example, important experiments can be manually eval-
uated, and not be allowed to get automatically assigned uncertainties; espe-
cially if the experiment is isolated, this can reduce the uncertainty a lot. Er-
roneous data, probably with the lower Corcalciuc set for (n,p) as an example, could be removed.

Finally, there is a certain risk for human bias in the choice of some details of
the methodology; if the final data would end up close to the lower Corcalciuc
set for (n,p), details of the method would probably have been revised, for ex-
ample. As another example, one could probably consider 5 MeV or 7 MeV for
the limit of the resonance range, just as well as 6 MeV. During the develop-
ment and testing, 6 MeV was chosen because it gave intuitively better results.
This can serve as a good example of how difficult it can be to completely
eliminate subjective judgment.

6.2 TALYS with energy-dependent parameters

In this section, we describe the additional considerations necessary to apply
the method of Paper VII using authentic TALYS (and not Pseudo-TALYS).
Basically, there are three such considerations:

1. TALYS is able to take energy-dependent values for some of its param-
eters, and not all, which can be more reasonable from a physics point
of view, and there are other reasons to reduce the number of energy-
dependent parameters further, see Sec. 6.2.1. The formalism of Paper
VII therefore needs to be slightly generalized to allow for only having
some energy-dependent parameters.

2. A prior distribution for the model parameters in TALYS has to be con-
sidered.

3. The dimensionality, i.e., the number of parameters involved in the fit,
must be reduced.

These topics are addressed below, section by section. A few additional con-
siderations are briefly discussed in Sec. 6.2.4.

6.2.1 A small generalization: some parameters are not
energy-dependent

Why the generalization is necessary

As mentioned above, the formalism of Paper VII should be generalized to be
able to handle some parameters that are not energy-dependent. One reason
for this can be that it is more reasonable to let some parameters be energy-
dependent from a physics point of view. Moreover, when the Koning-Delaroche
OMP is considered for one energy at a time, there is a certain redundancy among the parameters. For example, the real part of the volume-central well depth has a phenomenological energy dependence, which includes four parameters \( v_i; \ i \in \{1, 2, 3, 4\} \):

\[
V_V(E) = v_1 \left[1 - v_2(E - E_f) + v_3(E - E_f)^2 - v_4(E - E_f)^3\right], \tag{6.21}
\]

where \( E \) is the energy of the incoming neutron and \( E_f \) is the Fermi energy, determined from the neutron separation energies of the involved nuclides. Thus, this well depth is already energy-dependent (as are the other four), but with a certain functional form.

With parameters that are energy-dependent as in Paper VII, we want the energy dependence to be a bit more flexible than to follow the particular forms suggested in the Koning-Delaroche OMP. However, if all the \( v_i \) are allowed to be energy-dependent, we introduce a strong local multicollinearity into the fit, i.e., for each energy, variation of one parameter can be canceled by the variation of another. Even if we are able to constrain the full well depth, the various parameters can have almost unaffected marginal variance (but strong anti-correlation), resulting in lost predictability. Also, because of non-linearities, the fit can be poorly constrained by the data even if the final linearization is well constrained; the large marginal variances of the different parameters can make us move well outside the range where the linearization is a good approximation.

Out of the four \( v_i \), we therefore only let \( v_1 \) be energy-dependent. For the well depths of the other parts of the OMP, we let \( d_1, w_1, v_{so1}, w_{so1}, \) and \( r_C \) be energy-dependent, which all have similar roles for different parts of the OMP. Thus, we let the well depths be energy-dependent with the shape of the Koning-Delaroche OMP as a starting point. The OMP parameters are used for 6 different particles in total (also for outgoing particles), so there are \( 6 \cdot 6 - 3 = 33 \) (\( r_C \) does not apply for neutral particles) parameters that are potentially energy-dependent. In the input, the listed OMP parameters are controlled by \( v_{1\text{adjust}}, d_{1\text{adjust}}, w_{1\text{adjust}}, v_{so1\text{adjust}}, w_{so1\text{adjust}}, \) and \( r_{C\text{adjust}} \), respectively. On top of model parameters, we have also considered to allow energy dependence for one level density parameter, \( s_{2\text{adjust}} \), and one gamma-ray strength function parameter, \( s_{gradjust} \). However, it is explained in 6.2.4 why these two parameters are chosen not to be energy-dependent.

### The generalization in practice

Assume that the number of energy-dependent parameters is \( l_{\text{dep.}} \), and the number of parameters which are fixed with respect to energy is \( l_{\text{fixed}} \), giving the total number of parameters \( l = l_{\text{dep.}} + l_{\text{fixed}} \), where the parameters are ordered with the fixed parameters at the end. The generalization of \( \gamma_{E^*} \) from
Eq. (5.12) becomes

\[
\gamma_{E^*} = \begin{pmatrix}
\gamma_1(E_1^*) \\
\gamma_1(E_2^*) \\
\vdots \\
\gamma_1(E_n^*) \\
\gamma_l_{dep.}(E_{n^*}) \\
\beta_l_{dep. + 1} \\
\vdots \\
\beta_l
\end{pmatrix} = \begin{pmatrix}
\beta_1 + \delta_1(E_1^*) \\
\beta_1 + \delta_1(E_2^*) \\
\vdots \\
\beta_1 + \delta_1(E_n^*) \\
\beta_l_{dep.} + \delta_l_{dep.}(E_{n^*}) \\
\beta_l_{dep. + 1} \\
\vdots \\
\beta_l
\end{pmatrix}.
\]

(6.22)

Just as in Paper VII, this can be written as

\[
\gamma_{E^*} = T_{E^*}\beta + \delta_{E^*},
\]

(6.23)

using generalizations of Eqs. (16) and (18) of Paper VII:

\[
T_{E^*} = \begin{pmatrix}
1_{n^*} \\
1_{n^*} \\
\ddots \\
1_{n^*} \\
1_{l_{fixed}}
\end{pmatrix}, \quad \delta_{E^*} = \begin{pmatrix}
\delta_1(E_1^*) \\
\delta_1(E_2^*) \\
\vdots \\
\delta_1(E_n^*) \\
\delta_2(E_1^*) \\
\vdots \\
\delta_l_{dep.}(E_{n^*}) \\
0 \\
0 \\
\vdots \\
0
\end{pmatrix},
\]

(6.24)

where \(1_{n^*}\) is a length \(n^*\) vector of only ones (there are \(l_{dep.}\) such), and \(I_{l_{fixed}}\) is an \(l_{fixed} \times l_{fixed}\) identity matrix.

In Sec. 6.2.3, we will obtain a prior covariance matrix for the global parameters which is block diagonal:

\[
P_0 = \begin{pmatrix}
P_0^{dep} \\
P_0^{fixed}
\end{pmatrix},
\]

(6.25)

where \(P_0^{dep}\) is \(l_{dep.} \times l_{dep.}\) and \(P_0^{fixed}\) is \(l_{fixed} \times l_{fixed}\). Using this block-diagonality and the structure of \(T_{E^*}\), the covariance of Eq. (6.23) is [28]

\[
P = \begin{pmatrix}
K + T_{E^*}^{dep}P_0^{dep}T_{E^*}^{dep, T} \\
P_0^{fixed}
\end{pmatrix},
\]

(6.26)
where $T_{E^*}^{\text{dep.}}$ is the upper left $n^*l_{\text{dep.}} \times l_{\text{dep.}}$ block of $T^{\text{dep.}}$ and $K$ is a block diagonal matrix, where each block results from a covariance function of the GPs $\delta_j$; $j \in \{1, 2, ..., l_{\text{dep.}}\}$, i.e., just as in Paper VII but with $l_{\text{dep.}}$ blocks instead of $l$.

The inverse of $P$ can be taken block-wise [28], and the upper left block can be inverted efficiently using the analogue to Eq. (23) of Paper VII. In the special case that $P_0^{\text{dep.}}$ is diagonal (not only block diagonal), further simplification of the inversion can be achieved following the steps in the paper.

6.2.2 Prior parameter distribution

We start with the $l^0 = 335$ adjust parameters varied for $^{56}\text{Fe}$ in TENDL 2015. Most of these have little or no impact on the cross sections we consider; nevertheless, there are plenty of parameters. However, there is knowledge on how uncertain the parameters are from other nuclides; one could even say that certain restrictions on the parameters are part of the models themselves. In other words, there is reason to assume a prior distribution for the model parameters.

In Ref. [53], a set of uncertainties for many TALYS parameters is presented by Koning and Rochman, the former being the main author of TALYS. These are individually tuned using expert judgment combined with trial and error, to give distributions of observables that more or less cover most experimental data. In Ref. [48], Koning presents a systematical study aiming to determine a reasonable prior distribution of the model parameters, which is based on the manually selected uncertainties in Ref. [53]. This is done by performing TALYS calculations for all nuclides with existing neutron data, with parameters sampled around the values obtained using the global Koning-Delaroche OMP (the parameters are not adjusted for each particular nuclide). The sampling is done using uniform distributions on the range $[-M\sigma_{[53]}, M\sigma_{[53]}]$, where $\sigma_{[53]}$ is the uncertainty from Ref. [53], and $M$ is an adjustable factor. Then, it is determined what fraction of all experimental data in EXFOR [76] falls inside a $1\sigma$ band for the obtained cross sections.

For $M = 5$, about $2/3$ of the points were in the uncertainty band, and this was chosen as the appropriate value for $M$, probably because $2/3$ are close to $68\%$, the probability for a Gaussian random variable to be within $\pm 1\sigma$. However, for $M = 20$, the fraction is about $0.82$ and the growth rate seems small, i.e., even if the uncertainties are very large, there is a fair amount of points outside the band. This indicates either model defects, or that the data are erroneous, or a combination of the two. We cannot expect these data to be within reach of the model at all; assume that only a fraction of $0.85$ is reachable. Then, one may argue that only $68\% \cdot 0.85 = 0.58$ would suffice to give a good choice for $M$, which is achieved for $M \approx 3$ (reading from a figure in Ref. [48]). The standard deviation of a uniform distribution on
\[-M\sigma_{[53]}, M\sigma_{[53]}\] is [15] \(\sqrt{(2M\sigma_{[53]})^2/12} = M\sigma_{[53]}/\sqrt{3}\). With \(M = 3\), we thus get a standard deviation of \(\sqrt{3}\sigma_{[53]}\). In this work, this is therefore chosen as the prior uncertainty for each parameter.

The \(^{56}\)Fe data used later is also used in the determination of the prior distribution, which in principle violates the assumption of independence between prior and data. Since \(^{56}\)Fe is just one nuclide out of 309 with experimental data, we use the approximation that the \(^{56}\)Fe data are negligible in the full set of data.

6.2.3 Dimensionality reduction

With too many parameters, a model can often fit an extensive range of data but will lose in predictive power [26]. We have a large number of parameters, but also a prior which hopefully restricts the model results to a “reasonable range”, which would reduce this problem. However, the complexity of the numerical computation of the Jacobian grows linearly with the number of parameters, why it is nevertheless desirable to exclude parameters that do not impact the fit. Therefore, it is desirable to reduce the number of dimensions included in the fit; at this stage, we do so by transforming our parameters based on principal component regression (PCR, [91]), and exclude the dimensions with relatively small eigenvalues. In principle, the variability from other dimensions, which may be important where there is no data, could be reattained by merging the fitted dimensions with the prior for distributions for the rest. This is not done in this work, so far. Thus, in this thesis, we reduce the number of dimensions with the PCR-like approach described below and only consider the dimensions that are included in the fit.

Because of non-linearities, the principal components change during the search for the best linearization. However, updating the principal components would make it necessary to compute the Jacobian with respect to all parameters at every step. Hence, the reduction is based on \(J_0\), the Jacobian at the default TALYS values only.

Because some parameters are allowed to be energy-dependent and others not, the dimensionality reduction is carried out separately for the two sets of parameters. Therefore, \(J_0\) is divided into \(J_{0,\text{dep.}}\) and \(J_{0,\text{fixed}}\), the first \(I_{\text{dep.}}^0\) and last \(I_{\text{fixed}}^0\) columns of \(J_0\), respectively. Let \(U_{\text{dep.}}\) and \(U_{\text{fixed}}\) be the matrices whose columns are the orthonormal eigenvectors of

\[
J_{0,\text{dep.}}^T \Omega^{-1} J_{0,\text{dep.}} \quad \text{and} \quad J_{0,\text{fixed}}^T \Omega^{-1} J_{0,\text{fixed}}, \tag{6.27}
\]

respectively. In ordinary PCR, i.e., applied to ordinary least squares, the eigenvectors of \(X^TX\) are the principal components [91], where \(X\) is the mean-centered design matrix. A least squares problem with a general data covariance matrix \(\Omega\) can be transformed into an ordinary least squares problem, in which the original \(X^T \Omega^{-1} X\) plays the role of \(X^TX\) in the transformed problem.
Therefore, for each of the two sets of parameters, and for the considered linearization, taking the eigenvectors of the matrices in Eq. (6.27) is similar to ordinary PCR. The impact of the prior is ignored when obtaining the principal components, and the mean-centering is neglected.

The matrices $U_{\text{dep}}$ and $U_{\text{fixed}}$ only refer to the first $l_{\text{dep}}$ and last $l_{\text{fixed}}$ original parameters, respectively. Therefore, we can transform the original parameters $\beta$ into the full set of transformed parameters $\alpha^0$ using $\alpha^0 = A_0^T \beta$, where

$$A_0 = \begin{pmatrix} U_{\text{dep}} \\ U_{\text{fixed}} \end{pmatrix}. \tag{6.28}$$

For both $U_{\text{dep}}$ and $U_{\text{fixed}}$, we drop dimensions with eigenvalues that contribute with less than a fraction of $10^{-4}$ to the cumulative sum of eigenvalues. This leaves us with 6 dimensions that are energy-dependent and 10 that are not, i.e., a total of $l = 16$ dimensions.

Define $A$ as $A_0$ with the corresponding columns removed. Then, we transform the original parameters using

$$\alpha = A^T \beta. \tag{6.29}$$

Note that $A$ is $l^0 \times l$, so we move from $l^0 = 335$ dimensions to $l = 16$. Given a prior covariance matrix $P_{\beta}$ for $\beta$, we obtain the prior covariance matrix

$$P_\alpha = A^T P_{\beta} A, \tag{6.30}$$

for $\alpha$.

The parameters $\alpha$ can then be fitted using the method of Paper VII, using the generalization of Sec. 6.2.1. This results in the estimated “energy-dependent” vector $\hat{y}_{E^*}^\alpha$, at the grid $E^*$, defined in Eq. (6.23) but with $\beta$ replaced by $\alpha$. After the fit, one could transform the results back to the original parameters. This is not done here, but we can nevertheless study the obtained cross sections, by using the transformed model.

### 6.2.4 A few more considerations

The two hyper-parameters of the Gaussian processes are, of course, essential to the results. They are selected based on cross-validation as in Paper VII, but thanks to the data reduction described in Sec. 6.1.1, it is now feasible to use leave-one-out cross validation, more as in Paper VI. However, the cross validation, and the search for hyper-parameters, uses the final linearization of an available fit for efficiency. That is, the LM algorithm is not properly used for each removed data point, or even for each attempted set of hyperparameters. Without going into much detail, it turns out that the results are quite sensitive to the starting guess, i.e., the set of hyper-parameters for which the fit is actually carried out. Therefore, we systematically tested all different
combinations of $\sigma_0 \in \{1, 2\}$ (relative to the diagonal of the prior), and $\lambda \in \{2, 2.5, 3, 3.5\} \text{MeV}$, based on some initial tests. The minimum of Eq. (36) in Paper VII ($\chi^2 + \text{penalty}$) was obtained for $\sigma_0 = 1.40$ and $\lambda = 2.60 \text{MeV}$, which was obtained starting from $\sigma_0 = 2$ and $\lambda = 2.5 \text{MeV}$. Similar results ($\sigma_0 = 1.34$ and $\lambda = 2.46 \text{MeV}$) were obtained starting from $\sigma_0 = 1$ and $\lambda = 2 \text{MeV}$. Among all the finally attempted values, it is also seen that $\sigma_0 = 1.40$ and $\lambda = 2.60 \text{MeV}$ gives the least value for Eq. (36), and this choice is used for the remainder of this chapter.

The effect of using the linearization in the optimization of hyper-parameters could be studied in more detail in the future, and it is not impossible to carry out the full fit for each tested set of hyper-parameters. Each fit takes about an hour on a contemporary CPU, with TALYS running in parallel on a few tens of cores, and it takes about 20-50 steps for the minimization algorithm to converge.

In the results presented below, only the OMP parameters listed in Sec. 6.2.1 are allowed to be energy-dependent. As also mentioned in Sec. 6.2.1, we also considered to let two other parameters be energy-dependent. To this end, we applied the same systematic testing as above for this selection of model. This introduces some increased flexibility, which actually leads to somewhat smaller $\chi^2$-values, but the penalty term of Eq. (36) in Paper VII increased more, so the simpler model was selected.

6.3 Results and discussion

Fig. 6.5 shows a comparison between the global and local fits for a selection of cross sections. To be more specific, the transformed parameters selected in Sec. 6.2.3 are fitted, after which we have sampled the transformed parameters from the resulting covariances, giving 300 sets of TALYS parameters for which TALYS is run (on a finer energy grid than before). Note that the plotted data points are reduced and preprocessed as described in Sec. 6.1; thus, regions with a lot of independent data are manifested as small uncertainties. Also, for lower energies, the data are intended to represent average cross sections, and the resonances are included as larger uncertainties on these averages.

The global fits by themselves are perhaps more pleasing to the eye, because they are smoother, but we may suspect from Paper VII that the uncertainties are too small; consider, e.g., the (n,2n) or (n,p) cross sections at higher energies [Figs. 6.5(b) and 6.5(e)]. In the (n,2n) case, the very uncertain data points above 15 MeV are also strongly correlated [see Fig. 6.4(a)], so the experimental information is very limited here. There are also examples where the global fit obviously fails to fit the data, as for (n,tot) and (n,n') at low energy. The leave-one-out $\chi^2$ (see Paper VI) is $\chi^2_{\text{LOO}} = 389$, which is substantially more than the number of data points, i.e., 129.
Figure 6.5. A comparison of the global and local fits using TALYS and authentic data for a selection of cross sections.
As expected, the uncertainties are generally greater for the local fits; they are in most cases small where there are plenty of precise data, and because of the flexibility from the energy-dependent parameters, the uncertainty increases relatively fast when moving away from such regions. For many of the examples where the global fit fails to reach the data, the local fit succeeds much better. Using a method which is more driven by experiments, and less by models, one consequence is that seemingly weird behavior can be observed. For example, there are two bumps in the total cross section (for example) that are not predicted by the model. Note however, that the correlations are weaker for the local fit, so random samples from the distribution may very well yield a different shape. The local fit gives $\chi^2_{\text{LOO}} = 209$, so it indeed agrees much better to the data; however, it is far from 129 so we can reject an hypothesis that both the treatment of the data, and the model augmented with energy-dependent parameters, are perfect.

In some cases, such as for (n,xp) in Fig. 6.5(g), even the local fit seems extremely constrained given that there is only one, very uncertain, point for this interaction. However, the (n,xp) cross section is a linear combination of cross sections which produce protons, and for low enough energies, the only possibility to do so is the (n,p) reaction, for which there are plenty of data [cf., Fig. 6.5(e)]. As more possibilities to produce protons open up, the uncertainty increases.

It may be more concerning that a cross section such as (n,α), without any experimental data, is not more uncertain [see 6.5(f)]; the expected prior uncertainty is on the order of 100% [48]. This can to some extent be explained similarly as for (n,xp), because there is data for (n,xα), see 6.5(h). However, it may also be a consequence of the selected principal components; the components are selected based on how much they can be restricted by the data. Thus, if there is a cross section without any data, which is most sensitive to parameters that do not impact interactions with data, the uncertainty may be very underestimated here. Thus, the uncertainty of parameters that are not involved in the fit should be included.

The systematic deviation between data and both fits seen in Fig. 6.5(d) can raise suspicions; it can be argued that this is a model defect for which the energy-dependent parameters do not help. However, the six data points for higher energies are all from the same experiment, they are very strongly correlated after the preprocessing in Sec. 6.1. Thus, a systematic error which affect all points can explain the systematic deviation. However, the uncertainty of this curve appears small.

6.4 Conclusions and outlook for Chapter 6
This chapter presents results using the method of Paper VII applied to TALYS and using authentic data. It is generally more difficult to draw as strong con-
clusions as in Paper VII since we do not know what the truth is; all we can do is compare to the data. The predictability could be investigated with more care, e.g., using validation data, but we use leave-one-out cross-validation by studying the $\chi^2_{\text{LOO}}$, as defined in Paper VI. This gives a substantially lower value for the fit with energy-dependent parameters. Further, it was seen in Paper VII that the energy-dependent parameters have a large potential in outperforming the global fits, so this path is worth pursuing.

Before applying the method to the authentic data, a method to generate experimental covariances automatically, inspired by Refs. [87, 88], is applied to the data. Instead of using a set of rules such as in Papers I and II, this approach estimates uncertainties based on the distribution of data; a certain covariance structure is assumed, and parameters in this structure are estimated to maximize the likelihood of the observed data. The maximization only searches for one local maximum, which is seen to be a too rough approach in this case; the method should be developed with the aim to find all important maxima, and to combine them. The assumed covariance structure could also be subject to discussion, as well as starting values in the optimization algorithm.

On top of this estimation of experimental covariances, the chapter also suggests a method which assigns uncertainties to experiments that are not verified by other experiments. To this end, the degree of “isolation” for an experiment is quantified, and an additional uncertainty is assigned to isolated experiments, based on how isolated they are. The author believes that a feature of this type is desirable for an automatic interpretation of data, but the exact details of the method should be discussed.

As mentioned previously, the work of this chapter should rather be seen as a proof of concept; showing that the method of Paper VII can be applied to TALYS and authentic data. Several details need further consideration, out of which a few already are mentioned. Another example is to decide on exactly how the dimension reduction should be carried out. If a PCR-like method is used, it shall be studied further; for example, it remains to study how many components should be chosen, and how the results appear if the variability in other dimensions is included. It may be necessary to add a prior restriction on the observables, as is done in Ref. [48].

Leaving all the above aside, a few more things must be in place for the methods of this chapter to be applied to a real evaluation. For example, the method should be combined with a proper treatment of the resonance range, including the resonances which reach up to a few MeV.
7. Conclusions

Primarily, this thesis aims to estimate uncertainties in evaluated nuclear data (ND) that fulfill two criteria. First, the uncertainties must be well-founded on statistical grounds, such that their interpretation is clear. Also, the methodology should allow for large-scale automation in order to yield comprehensive ND uncertainties, i.e., ranging over all practically occurring nuclides, and including uncertainties and correlations for all parts of the data. As of now, the far most comprehensive ND library, TENDL, incorporates experimental data in a way which is not entirely well-founded in a statistical sense; the methodology of TENDL should be augmented to remove this issue.

One fundamental ingredient for a more well-founded treatment is proper experimental covariance matrices, i.e., not only containing sound point-by-point uncertainties, but also correlations between the points, arising from unavoidable systematic uncertainties. A problem faced here is that the information on uncertainties is limited, particularly on systematic uncertainties. Further, even for cases with relatively well-reported uncertainties available, they are rarely compiled in a way which is easy to interpret automatically. In Papers I and II, attempts are nevertheless made to automatically read and interpret the available uncertainty information. When information is judged to be missing, uncertainties are assigned based on a few simple rules. The results are not very easy to decipher because of an overly ambitious approach, doing too many things at the same time. However, from the differences between Papers I and II arising from differences in the used rules, it is clear that the rules are very decisive. At the same time, it is difficult to find such rules that are well-founded.

In Chapter 6, the automatic interpretation of experimental data is revisited. Here, missing uncertainties are estimated using the distribution of the data, with some additional conservatism added when experiments are not confirmed by other experiments. Such an approach seems promising, but it needs to be both further developed and systematically examined. Possibly, it can be used in combination with refined rules based on those in Papers I and II. A method to strongly reduce the number of experimental data points without losing any information (which is not lost anyway) is also suggested in Chapter 6, and this approach is recommended for anyone fitting linearly interpolated models to data.

The above-mentioned difficulties to decipher the results of Papers I and II are much because of the naïve implementation of the Monte Carlo fitting technique UMC-B. In Paper V, different fitting techniques suggested for ND evaluation are studied systematically in a controlled environment; the underlying
truths are known, and the data are generated based on this truth. From these results, it is clear that the results in Papers I and II can be poorly converged; the studied Monte Carlo fitting techniques need a vast number of function calls, even though a simple model is used. Further, in the case considered in Paper V, the choice between Monte Carlo or deterministic techniques is unimportant for the results. Therefore, the later papers focus on deterministic methods. Paper V also highlights that model defects can play a major role, especially for techniques working in the parameter domain.

Before addressing model defects, let us mention Paper IV, which takes a detour from the automatized approach. Individual uncertainty components of the experiments on $^{59}$Ni are analyzed and combined with other sources of information using a novel approach. The resulting distribution of ND is propagated to the helium production in stainless steel, as it can be an important contribution to the material damage in a reactor environment. Further, the distribution of ND was condensed into an evaluated ND file with covariances, now part of OECD’s library of evaluated ND, i.e., JEFF 3.3.

Aside from the just mentioned helium production, the thesis also includes the propagation of uncertainties to several other applications, such as the effect of the shielding fuel assemblies used at Ringhals, a Swedish nuclear power plant.

Papers VI and VII, and Chapter 6, are much devoted to the topic of model defects. The two papers take two different approaches to the use of Gaussian processes (GPs) to treat model defects, and Chapter 6 is a continuation of the work in Paper VII. In Paper VI, the “conventional” approach is taken, where a term is added to the original model. This is applied to a situation in which different peaks in a histogram of data are to be separated using a model fit, and GPs are used to allow for arbitrarily shaped peaks, i.e., without a known functional form. In the paper, synthetic data are used to show that the method gives the expected results. If a Gaussian shape is incorrectly assumed, it is also seen that the resulting errors can be substantial. In the thesis, results using authentic data are also included, showing that the difference can be important in reality, too. Thus, a recommendation from this thesis is to use such a treatment in the considered situation, at least if conventional fits are not satisfactory.

Paper VII returns to ND evaluation. A model similar to TALYS is augmented with additional flexibility by allowing energy-dependent parameters, where the energy dependence is controlled by GPs. As in Paper VI, synthetic data are used. The underlying truth varies in such a way that the “degree of model defect” varies, and it is seen that the method improves the handling of model defects by orders of magnitude, compared to a fit without the energy-dependent parameters. Qualitatively speaking, the distribution of results is not far from the desired distribution, even though the experimental data are sparse, so the method constitutes a promising route for ND evaluation in the fast range. Nonetheless, it would be very interesting to use the framework of Paper VII to compare the method to the more conventional way to use GPs (as
Moreover, the comparison should include the standard method in ND evaluation, GLS, which is less sensitive to model defects than the reference method in Paper VII. Chapter 6 shows that the method really can be applied using TALYS, and with authentic data (preprocessed as mentioned above). There are, however, aspects that must be further investigated.

On top of such investigations and the systematic studies of how the method with energy-dependent parameters compare to other methods, it is also mentioned above that the automatic treatment of experimental data suggested in Chapter 6 should be further developed and investigated. To use the method of Paper VII for a full evaluation, it is also necessary to merge it with a proper treatment of the resonance range, and the experiments for other ND than cross sections should be considered. Further, one should revisit an aspect which is neglected after Paper II, namely, the effect of uncertainties in the incoming neutron energies.

To conclude, this thesis includes well-founded ND uncertainties for $^{59}$Ni (now in a well-used library), a comparison of a set of techniques used in ND evaluation, and a validated method which enables the fitting of arbitrarily shaped peaks in, e.g., the analysis of experiments. Further, it suggests a new method for ND evaluation, which constitutes a viable way to treat model defects within the framework of TENDL, and outlines how to treat experiments automatically within such a system. Hence, even if the goal is not yet reached, the thesis takes important steps in the direction towards well-founded comprehensive ND uncertainties.
I greatly appreciate the financial support from the Nuclear Research and Consultancy Group NRG (Petten, the Netherlands) through the SAFETAL project. I also thank the Swedish Center for Nuclear Technology SKC (funded by Ringhals AB, Forsmarks Kraftgrupp AB, Oskarshamns Kraftgrupp and Westinghouse) for their financial support, in particular through the MABIL project.

I would further like to thank my supervisors Henrik Sjöstrand, Stephan Pomp, Arjan Koning, Dimitri Rochman and Klaes-Håkan Bejmer for their feedback, really interesting discussions and a set of highly entertaining evenings. A few extra thanks to Arjan, for answering me about details on TENDL and TALYS that I failed to find in the publications until I already had the answer from you. On top of the supervisors, lots of people have been important for me and my development; for example, Denise Neudecker, who spent a lot of her time (that administration!) to give me the nice opportunity to discuss with her and others for a whole month in New Mexico. Also, Roberto Capote: each time I’ve bumped into you, you’ve always pointed me in an interesting direction.

To Henrik, I have really appreciated the discussions with you during the last few years; both scientific and non-scientific, even if the latter category also tend to take a quite scientific, and interesting, approach in many cases. Also, thank you for all the reading of the thesis; I feel bad for forgetting to tell you that one or the other part already was about to be rewritten. The thesis is a lot better than it would have been without your feedback; not to mention what it would have been with another main supervisor!

A lot of colleagues in Uppsala have made the years more fun; table-tennis breaks with Ali and Peter Andersson’s interesting ability to see life from all possible perspectives, to mention a couple of examples. And thanks to Cecilia who says I can sing!

Finally, thanks to my family; all of them, including those who aren’t biologically related; especially my parents for the unconditional love and support (a couple of matrices were named after you but that part got removed...). Also, thanks to JB, for always being there, and almost always giving reasons for both a good laugh and contemplation. And thanks to Lina, for all the wonderful love, laughs, hugs, and support. You also got a matrix, in the same part as mum and dad. Hence, it was also removed.
På väg mot välgrundade heltäckande osäkerheter i kärndata
- Anpassning av bristfälliga modeller till bristfällig data


Denna avhandling, som bygger på en samling artiklar, handlar till stor del om osäkerheter, och framför allt osäkerheter i evaluerad kärndata. Sådana osäkerheter, och korrelationer mellan olika delar av datan (som kan sammanfattas i kovarianser), är nödvändiga för att kunna uttala sig om osäkerheten i resultat av beräkningarna som nämndes ovan. Idag innehåller de största biblioteken med kärndata mer eller mindre noggrant framtagna osäkerheter för ett antal atomkärnor – de som betraktas som särskilt viktiga. För många andra finns inga osäkerheter alls, och många atomkärnor ingår över huvud taget inte. Det överlägsset mest heltäckande biblioteket, TENDL, inkluderar osäkerhetsuppskattningar för alla de omkring 2800 atomkärnor som det innehåller. Detta bygger på att stora delar av processen för att ta fram dessa evalueringar är automatiserad. Till stor del grundar sig processen på fördelningar av modellparameterar i programmet TALYS, som kombinerar en stor mängd kärnfysikaliska modeller. Ån så länge har den automatiska anpassningen av dessa parametrar till experimentdata krävt vissa lösningar som är praktiska men inte alltid helt välgrundade. Ett långsiktigt mål, som denna avhandling siktar mot, är att komplettera processen i TENDL-projektet med metoder som gör att detta heltäckande bibliotek med osäkerheter också står på en god statistisk grund.

En av de huvudingredienserna på vägen mot mer välgrundade osäkerheter är likaledes välgrundade **kovariansmatriiser** för experimentdata som används. Ofta är den rapporterade datan bara försedd med enkla osäkerheter som gäller
punkt för punkt – det saknas information om systematiska osäkerheter, vilka
blir väldigt viktiga i och med att de leder till korrelationer mellan olika dat-
apunkter. Om informationen finns, är den ofta svårtillgänglig för automatis-
erad tolkning. I Artikel I-II görs i vilket fall försök att tolka den tillgäng-
liga informationen, och att komplettera den med hjälp av en uppsättning en-
kla regler. Resultaten i dessa artiklar är något svårtydda, eftersom flera saker
gjordes samtidigt, men de mindre skillnader som finns mellan den automatiska
tolkningen av experiment i Artikel I och Artikel II leder till stora skillnader i
resultaten. Alltså blir detaljerna i en sådan metod utslagsgivande, och sam-
tidigt är det svårt att bestämma riktigt välgrundade sådana regler.

Avhandlingen återvänder till automatisk behandling av experiment mot slut-
et. Här används ett helt annat angreppssätt, som bygger på hur datan är förde-
lad, både inom experiment och mellan experiment, tillsammans med basala
antaganden om korrelationsstrukturen inom experiment. Detta verkar vara ett
mer lovande angreppssätt, men det finns flera detaljer att utveckla. I samband
med denna typ av automatiserade behandling av experiment görs också iakt-
tagelsen att antalet datapunkter (ofta tusentals) kan reduceras ner till ett bety-
dligt mindre antal – utan att man går miste om någon information som man
inte skulle gå miste om ändå när man använder interpolering mellan beräkn-
ingspunkter. Denna reducering kan vara mycket behändig när man ska lösa
det ganska komplicerade optimeringsproblem som uppstår när man behandlar
datan på det vis som antyds ovan.

Att resultaten i artikel I och II är svårtydda, som nämntes ovan, har delvis
att göra med en nagot naiv implementering av UMC-B, en teknik för att an-
passa modeller till data i kärndatasammanhang. Artikel V studerar flera sådana
tekniker i en situation med syntetisk data vilket gör att det går att jämföra re-
sultaten med sanningen som ligger bakom datan. Från dessa resultat är det
tydligt att resultaten i artikel I och II kan bygga på för dålig statistik. Det blir
också tydligt att **modelldefekter** – att ens modell inte kan reproduera verkligheten vad man än väljer för parametrar – är viktiga, vilket vi återkommer
till nedan. Samtidigt leder bland annat arbetet med artikel V till att arbetet
riktar in sig på deterministiska metoder i fortsättningen.

Artikel IV behandlar $^{59}$Ni, en atomkärna som ger ett viktigt bidrag till ma-
terialskador i rostfritt stål i en reaktormiljö. Artikeln är lite av ett avsteg från
tankarna om automatiserbarhet, och anpassningen av modeller är inte cen-
tral i detta arbete. I stället analyseras flera experiment för termiska tvärsnitt
(tvärsnitt vid låga energier hos neutronen) eftersom dessa är både viktigast
för materialskadorna och för att de utgör huvuddelen av experimenten. Sedan
simuleras osäkerheter i olika komponenter i experimenten för att ge en fördel-
ning av termiska tvärsnitt, som till sist kombineras med andra typer av data på
ett nytt sätt. Arbetet med denna artikel resulterade i vad som nu är $^{59}$Ni-filen i
OECD:s kärndatabibliotek JEFF 3.3.

Artikel VI och VII behandlar till stor del de ovan nämnda modelldefek-
terna. I båda fallen används Gaussiska processer för att öka flexibiliteten hos


I artikel VII används, återigen, syntetisk data, men i avhandlingens kapitel 6 så visas det att metoden fungerar på verklig data för $^{56}$Fe (behandlad som kort beskrivet ovan) och med riktiga TALYS. Viss utveckling krävs dock fortfarande.

Sammanfattningsvis har avhandlingen lett till välgrundade osäkerheter för $^{59}$Ni (som nu är del i ett välanvänd bibliotek), en översikt och jämförelse av ett antal evalueringstekniker, en validerad metod för att anpassa godtyckligt formed toppar i analys av experimentdata, visat på en väg mot en mer välgrundad evalueringsteknik som är lämplig att använda i ramverket kring TENDL, och skisserat på en automatisk behandling av data som skulle ingå i ett sådant system. Även om målet inte är nått än, så tar denna avhandling viktiga steg mot heltäckande osäkerheter i kärndata, som också är välgrundade.
References

[2] “U.S. energy information administration (EIA) homepage.”


2008.


See also: Argonne National Laboratory Report ANL/NDM-166, 2008.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1669

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)