Annals of Nuclear Energy 120 (2018) 35–47

Contents lists available at ScienceDirect

Annals of Nuclear Energy

journal homepage: www.elsevier.com/locate/anucene

Treating model defects by fitting smoothly varying model parameters: Energy dependence in nuclear data evaluation

P. Helgesson [a,b,*], H. Sjöstrand [a]

[a] Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden
[b] Nuclear Research and Consultancy Group NRG, Petten, The Netherlands

ARTICLE INFO

Article history:
Received 21 February 2018
Received in revised form 8 May 2018
Accepted 11 May 2018

Keywords:
Model fitting
Model defects
Covariances
Gaussian processes
Uncertainty quantification
Nuclear data evaluation

ABSTRACT

The fitting of models to data is essential in nuclear data evaluation, as in many other fields of science. The models may be necessary for interpolation or extrapolation, but they are seldom perfect; there are model defects present which can result in severe biases and underestimated uncertainties.

This work presents and investigates the idea to treat this problem by letting the model parameters vary smoothly with an input parameter. To be specific, the model parameters for neutron cross sections are allowed to vary with neutron energy. The parameter variation is limited by Gaussian processes, but the method should not be confused with adding a Gaussian process to the model.

The performance of the method is studied using a large number of synthetic data sets, such that it is possible to quantitatively study the distribution of results compared to the underlying truth. There are imperfections in the results, but the method is seen to readily outperform fits without the energy-dependent parameters. In particular, the estimates of uncertainty and correlations are much better. Hence, the method seems to offer a promising route for future treatment of model defects, both for nuclear data and elsewhere.

1. Introduction

The fitting of various observables to experimental data plays a central role in nuclear data (ND) evaluation. This paper is directed towards ND evaluation, but the idea can be applied to any case where a parametrized model is fitted to data, which is essential in a wide range of science and applications. For the reader unacquainted to ND: an important subset of ND are neutron cross sections, describing the probabilities for different types of interactions between a neutron and a nuclide. There is a large number of important nuclides, plenty of different possible interactions, and the cross sections depend on the neutron energy. This paper considers the simultaneous fitting of several such cross sections to quite sparse experimental data, with the aid of imperfect models of the underlying physics. Figs. 2 and 3 (of the results section) can help illustrating the considered situation.

In regions where there are plenty of data, one can directly use the experimental data together with some assumption on reasonable smoothness, as in the "Neutron cross section standards" (Carlson et al., 2009). For a lot of observables, there are more or less large gaps in the data, which make it necessary to use nuclear reaction models to help in the long-range inter- or extrapolation (with respect to, e.g., energy, observable, and even nuclide).

One way to implicitly include the model is to generate a prior distribution for the observables based on the model, as in what nuclear data evaluators refer to as General(ized) Least Squares (GLS, Smith, 1991; Herman et al., 2011), or in Unified Monte Carlo-Garage (UMC-G, Smith, 2007). This has the disadvantage that, even though the model impacts the results trough the prior, the model is abandoned in the actual fit. This leads to observables which cannot be reproduced by the model, and, thus, important physics may be broken. An alternative is to actually fit the parameters of the model, as in Unified Monte Carlo-Breakfast (UMC-B, Capote et al., 2012), or using a deterministic non-linear fitting algorithm such as Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963; Helgesson and Sjöstrand, 2017. Such methods are naturally compatible with Total Monte Carlo (TMC, Koning and Rochman, 2008, 2012), stochastic uncertainty propagation from model parameters to applications. There are also alternative fitting techniques related to TMC, such as Bayesian Monte Carlo (Koning, 2015) which is used for the production of later TENDL libraries (Koning and Rochman, 2012).

A disadvantage with fitting the parameters of the model is that such methods are more sensitive to *model defects* (Helgesson et al.,

* Corresponding author at: Dpt. of Phys. and Astronomy, Box 516, 751 20 Uppsala, Sweden.
E-mail address: petter.helgesson@physics.uu.se (P. Helgesson).

https://doi.org/10.1016/j.anucene.2018.05.026
0306-4549/

2017), i.e., that the model is not able to exactly reproduce the true physics underlying the data, whatever the parameters are. This typically leads to biased results with strongly underestimated uncertainties (Helgesson et al., 2017; Helgesson and Sjöstrand, 2017). The other class of fitting techniques described above (such as GLS) will also give erroneous results if there are model defects present (Helgesson et al., 2017). Further, it is not controversial to claim that there are, to some extent, model defects in most ND evaluation situations, so the issue should be treated somehow. Schnabel and Leeb have used Gaussian processes (GPs) to address the problem of model defects for some nuclear data problems (Schnabel, 2015; Schnabel and Leeb, 2016), and the results appear successful. The authors of this paper have applied GPs to fit peaks of arbitrary shape to an histogram of data, and by using synthetic (simulated) data also shown that it gives desired results (Helgesson and Sjöstrand, 2017). In both these cases, a term describing the defect is added to the physics model, and this term is modeled by a GP.

In this paper, we present and study the idea to let the *parameters* of the model depend on the neutron energy E (an input variable), and to let this energy dependence be limited by GPs. This has a number of potential advantages over the "usual" use of GPs to treat model defects. These are discussed a little more in Section 3.2, but we could already mention that physical constraints are guaranteed to be conserved at each E, without the need to carefully define the GP to include these constraints. It would also be more straightforward to incorporate this technique into the TMC and TENDL frameworks, to produce complete and consistent nuclear data for a very wide range of nuclides.

The paper starts in Section 2 by introducing some background knowledge, including a more formal description of the problem to address, and a brief introduction to GPs. Section 3 presents the method of using energy-dependent parameters. This is followed by validating the method in Section 4 using synthetic (but rather realistic) data, allowing a comparison of the results to the truth which underlies the data. As customary, the paper ends with conclusions and an outlook for the future development of this work.

2. A quick review of definitions and prerequisites

2.1. The problem at hand + some notation

We consider a part of the nuclear data evaluation process: using experimental data and nuclear reaction models to estimate a set of physical observables and the covariances associated to them. We limit the study to the cross sections of 17 reaction channels (referred to as *channels* in the following) for a specific nuclide, ^{56}Fe, for neutron energies between 1 MeV and 30 MeV. In reality, the ^{56}Fe cross sections show strong resonances up to a few MeV. In this paper, we consider average cross sections and neglect the resonances. Since we work with synthetic data to study the performance of the method, neglecting the resonances enables us to assume that the cross sections vary smoothly with energy. The cross sections of ^{56}Fe were chosen because of their importance for estimating radiation damage in fission and fusion applications.

Formulated mathematically, we have a data vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ which is an observation of a random vector $\mathbf{Y}$. We assume that $\mathbf{Y}$ is normally distributed around the underlying truth $\mathbf{f}_{\text{true}}$:

$$\mathbf{Y} = \mathbf{f}_{\text{true}} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \Omega). \tag{1}$$

The random vector $\boldsymbol{\varepsilon}$ describes the experimental error and, thus, Ω is the experimental data covariance matrix. This is typically *not* diagonal, because of systematic uncertainties.

We also have a function $f(x;\boldsymbol{\beta})$, which is a model used to approximate the truth. In practice, $f(x;\boldsymbol{\beta})$ could be a nuclear reaction code such as TALYS (Koning et al., 2015). In this work, we use a TALYS-*like* model which executes faster, but like TALYS it is non-linear and it is intended to contain much of the complexity of TALYS from a data fitting point of view (details in Appendix A). The function $f(x;\boldsymbol{\beta})$ depends on an input variable x (details in the next sentence) and a vector of model parameters $\boldsymbol{\beta}$. To simplify the notation (in some respects), we include a selection of the channel in the input to f, by defining the input variable x as $x = (E, c)$, where E is the neutron energy and c is a discrete variable determining the channel. All the input data are summarized into the "vector" $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. In a sense, x is a length 2 vector and $\mathbf{x}$ is an $n \times 2$ matrix. However, we will not exploit the matrix structure of $\mathbf{x}$, so we use the notation for scalars and vectors for x and $\mathbf{x}$, respectively (italics and boldface with serifs, respectively). We will sometimes write $E(x)$ to denote the energy associated with x, and define $\mathbf{E}(\mathbf{x}) = (E(x_1), E(x_2), \ldots, E(x_n))^{\mathrm{T}}$.

As a consequence of the definition of x, we consider the model function $f(x;\boldsymbol{\beta})$ to be a single-valued function even if we have several reaction channels. We further define $\mathbf{f}(\mathbf{x};\boldsymbol{\beta}) = (f(x_1;\boldsymbol{\beta}), f(x_2;\boldsymbol{\beta}), \ldots, f(x_n;\boldsymbol{\beta}))^{\mathrm{T}}$.

If the physical model is able to perfectly reproduce the truth, there is a set of parameters $\boldsymbol{\beta}_{\text{true}}$ such that $\mathbf{f}_{\text{true}} = \mathbf{f}(\mathbf{x};\boldsymbol{\beta}_{\text{true}})$, and

$$\mathbf{Y} = \mathbf{f}(\mathbf{x};\boldsymbol{\beta}_{\text{true}}) + \boldsymbol{\varepsilon}; \quad \boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}, \Omega). \tag{2}$$

This is the usual formulation of a non-linear least squares problem, which often can be solved using the Levenberg–Marquardt (LM) algorithm (Levenberg, 1944; Marquardt, 1963; Helgesson and Sjöstrand, 2017), or using, e.g., Markov Chain Monte Carlo (Press et al., 2007).

However, there are often *model defects* present, i.e., the model is *not* able to perfectly reproduce the truth. In other words, there is no parameter set such as $\boldsymbol{\beta}_{\text{true}}$. If so, the consequences of assuming that there are no model defects can be severe (Helgesson et al., 2017; Helgesson and Sjöstrand, 2017; Schnabel, 2015): the deviations from the underlying truth can be much too large compared to the estimated uncertainties.

Gaussian processes (see Section 2.2) have previously been used to treat model defects (see Section 2.3). In this paper, we also use GPs for this purpose, but in a novel way (see Section 3.1).

The model parameters $\boldsymbol{\beta}$ are based on the "`adjust`" parameters of TALYS (Koning et al., 2015), which all have default values of 1 and are allowed to take values in [0.5, 2] or [0.1, 10] depending on the parameter. However, the parameters are transformed using

$$\beta_j = \begin{cases} a_j - 1; & a_j \geqslant 1 \\ 1 - 1/a_j; & a_j < 1 \end{cases}, \tag{3}$$

where a_j is the corresponding `adjust` parameter. In other words, the parameters $\boldsymbol{\beta}$ describe the change from the default, relative to the least of the default and the resulting parameter value ($\beta_j = -1$ means $a_j = 0.5$, for example). In this way, the parameters $\boldsymbol{\beta}$ can be symmetrically distributed around 0, and the ranges for the adjust parameters translate to $[-1, 1]$ and $[-9, 9]$, respectively. The parameters are *approximately* restricted to these bounds by assuming a prior distribution for each β_j which is centered around 0 with standard deviations of 1/3 and 3, respectively. This rather uninformed prior is not expected to impact the results much, except for giving some numerical advantages.

2.2. Gaussian processes (GPs)

This section intends to present the most central concepts of GPs that are necessary for our purposes. In Helgesson and Sjöstrand