Uncertainty evaluation in theoretical calculations of cross sections and rate coefficients

V. Kokoouline

Department of Physics, University of Central Florida, Orlando, FL

Source of uncertainties in theoretical calculations

1. Due to applicability limits of the chosen model.
2. Withing a model, uncertainties of model parameters result in uncertainties of final data, such as cross sections (~random uncertainty).
3. If uncertainties of experimental and theoretical data are known, for the purpose of data evaluation and recommendation, one should combine the two data sets to produce the best guess data with the smallest possible uncertainty.

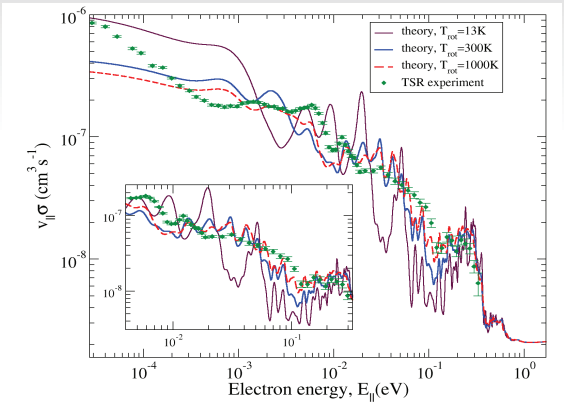
Evaluation of uncertainties within a chosen model

1. If there are not that many parameters, one can simply vary them within a reasonable interval and see the effect of the variation on the final cross sections. It is a sort of sensitivity test.

   Systematic approaches
2. Least squares method
3. Monte-Carlo methods

$e^-+H_3^+$ collisions

Theoretical results are sensitive to certain parameters in the model. Here, the dependence of the rate constant on rotational temperature of the target is assessed in a qualitative way. There are several other parameters in the model.



Least-squares method

Given two sets of data with uncertainties on the same grid of energies (E_i, i=1,2,..), the best guess (recommended) cross section σ_i^{eva} is obtained by minimizing the following function

$$F=\sum_i\left(\frac{(\sigma_i^{rec}-\sigma_i^{\exp})^2}{\Delta\sigma_i^{\exp}}+\frac{(\sigma_i^{rec}-\sigma_i^{the})^2}{\Delta\sigma_i^{the}}\right)$$

Uncertainty $\Delta\sigma_i^{rec}$ of evaluated data can be obtained in a statistical approach.

The approach is good if results (σ_i^{the} or σ_i^{exp}) obtained for different energy data points are not correlated.

Serious disadvantage: Usually, theoretical results are strongly correlated: changing a parameter in the theoretical model changes cross sections for all energy points E_i. Experimental cross sections measured at different E_i may also be correlated.

Some terminology

1. The *mean value*. For example, calculating (measuring) σ at energy E_i

$$\langle\sigma_i\rangle=\int\sigma_i P(\sigma_i)\,d\sigma_i$$

2. In theoretical calculations , several theoretical parameters p_j are described are uncertain. For example, if a potential energy surface on j=1,N_{pes} grid point is used, it introduces N_{pes} parameters p_j, each having its own uncertainty.
3. Strictly speaking, one has to calculate an N-dimensional integral if there are several variables:

$$\langle p_j\rangle=\int p_j P(p_1,p_2,\cdots,p_j,\cdots)\,d\vec{p}$$

4. *Uncertainty* (*standard deviation*)

$$\langle\sigma_i\rangle=\int\sigma_i P(\sigma_1,\sigma_2,\cdots,\sigma_i,\cdots)\,d\vec{\sigma}$$

$$\Delta\sigma_i=\sqrt{\langle(\sigma_i-\langle\sigma_i\rangle)^2\rangle}=\sqrt{\langle\sigma_i^2\rangle-\langle\sigma_i\rangle^2}$$

Covariance matrix and uncertainty

1. *Covariance matrix* describes correlation between variables: Changing a parameter in a model (or in experiment) changes cross sections at all energy points:

$$V_{ii'}=\langle(\sigma_i-\langle\sigma_i\rangle)(\sigma_{i'}-\langle\sigma_{i'}\rangle)\rangle=\langle\sigma_i\sigma_{i'}\rangle-\langle\sigma_i\rangle\langle\sigma_{i'}\rangle$$

2. When i=i':

$$V_{ii}=(\Delta\sigma_i)^2$$

it is *variance*.

Least-squares method with correlations

If experimental and theoretical covariance matrices V_{exp} and V_{the} are known, then one has to minimize F

$$F=(\vec{\sigma}^{rec}-\vec{\sigma}^{\exp})V_{\exp}^{-1}(\vec{\sigma}^{rec}-\vec{\sigma}^{\exp})+(\vec{\sigma}^{rec}-\vec{\sigma}^{the})V_{the}^{-1}(\vec{\sigma}^{rec}-\vec{\sigma}^{the}),$$

written as a matrix product here with cross sections represented in a vector form

$$\vec{\sigma}^{rec}=(\sigma_1^{rec},\sigma_2^{rec},\cdots),\ \vec{\sigma}^{the}=(\sigma_1^{the},\sigma_2^{the},\cdots),\ \vec{\sigma}^{\exp}=(\sigma_1^{\exp},\sigma_2^{\exp},\cdots).$$

If there are several parameters in the theoretical model, finding minimum of F is difficult. Covariance matrix or/and uncertainties of the evaluated data are (easily) not available.

A better way to determine the evaluated cross section is the Bayesian statistics approach.

Bayesian statistics methods

1. An alternative to the least-squared method is to apply a Monte-Carlo/Bayesian approach. We used a modified UMC approach by Capote *et al.*

2. Monte-Carlo approaches account for the uncertainty propagation in theoretical calculations in a systematic way.

3. Combines theoretical and experimental data (more generally, data from two independent sources) and produces recommended data. The two data sources should have uncertainties or covariance matrices. The "evaluation" procedure is based on the Bayesian statistics.

Bayes' theorem for data evaluation

1. Bayes formula gives the probability $P(W|L)$ of an event W at the condition that event L was observed. If $P(L)$, $P(W)$, and $P(L|W)$ are given then

$$P(W|L)=\frac{P(L|W)P(W)}{P(L)}$$

$$P(L|W)=0.85;$$
$$P(L|M)=0.1;$$

$$P(W|L)=\frac{P(L|W)P(W)}{P(L)}=\frac{0.85\cdot 0.5}{0.5\cdot 0.85+0.5\cdot 0.1}=0.89$$

Bayes' theorem for data evaluation

1. Bayes formula gives the probability $P(W|L)$ of an event W at the condition that event L was observed. If $P(L)$, $P(W)$, and $P(L|W)$ are given then

$$P(W|L)=\frac{P(L|W)P(W)}{P(L)}$$

2. For data evaluation: Consider cross section σ at energy E. A theoretical model and uncertainties of its parameters produce an uncertainty in σ (by the means of uncertainty propagation) and the corresponding probability distribution $P(\sigma)\sim P(W)$.

3. If experimental data is considered as additional data (information "L"), one can use Bayes' formula to calculate probability $P(W|L)$ that the "actual" cross section has value σ.

4. This approach combines data and uncertainties from two sources.

Bayesian theorem for data evaluation

Prob. distribution for exp. data for given σ

Prob. distribution for theoretical data

$$P(\sigma|\sigma^{\exp})=\frac{P(\sigma^{\exp}|\sigma)P(\sigma)}{P(\sigma^{\exp})}$$

Prob. distribution for evaluated data

Prob. distribution for exp. data summed over all outcomes σ

$$P(W|L)=\frac{P(L|W)P(W)}{P(L)}$$

Formulas for UMC

$$P(\sigma)=N_{the}\exp\left\{-\frac{1}{2}\left[(\vec{\sigma}-\vec{\sigma}^{the})(\hat{V}^{the})^{-1}(\vec{\sigma}-\vec{\sigma}^{the})\right]\right\}$$

$$P(\sigma^{\exp}|\sigma)=N_{\exp}\exp\left\{-\frac{1}{2}\left[(\vec{\sigma}-\vec{\sigma}^{\exp})(\hat{V}^{\exp})^{-1}(\vec{\sigma}-\vec{\sigma}^{\exp})\right]\right\}$$

$$P(\sigma^{\exp})=\int d\sigma\, P(\sigma^{\exp}|\sigma)P(\sigma)$$

$$V_{ii'}=\langle(\sigma_i-\langle\sigma_i\rangle)(\sigma_{i'}-\langle\sigma_{i'}\rangle)\rangle=\langle\sigma_i\sigma_{i'}\rangle-\langle\sigma_i\rangle\langle\sigma_{i'}\rangle$$

N_{the} and N_{exp} are normalization constants, but they need not to be calculated.
I will use simpler notations:

$$P(\sigma|\sigma^{\exp})\to P_{rec}(\sigma)=\frac{P_{\exp}(\sigma)P_{the}(\sigma)}{\int P_{\exp}(\sigma)P_{the}(\sigma)\,d\sigma}$$

Procedure of determination of theoretical uncertainties and evaluated cross sections

1. Choice of the energy grid, E_i (i=1,...,N): For evaluation, the grid should be the same as in experiment.

2. Varying parameters within an interval of a few Δp_j, and calculating σ_i^k for each set k of parameters. An average over p (summation over k) is performed: mean values $\langle\sigma_i\rangle$ and covariance matrix are obtained

$$\langle\sigma_i\rangle=\int\sigma_i P(p_1,p_2,\cdots,p_j,\cdots)d\vec{p}$$
$$V_{ii'}^{the}=\langle\sigma_i\sigma_{i'}\rangle-\langle\sigma_i\rangle\langle\sigma_{i'}\rangle=\int\sigma_i\sigma_{i'}P(p_1,p_2,\cdots,p_j,\cdots)d\vec{p}-\langle\sigma_i\rangle\langle\sigma_{i'}\rangle$$

3. One can evaluate above multidimensional integrals on a uniform grid or use the Gaussian quadrature. It is slow if the model has many parameters

Monte-Carlo integration

1. One can use a Monte-Carlo approach for the integration as in UMC. For this purpose, one chooses randomly a set k of parameters p_j^k, such that p_j^k is within the interval $\langle p_j\rangle\pm\psi\Delta p_j$, with ψ~4–6, i.e. $p_j^k=\langle p_j\rangle+(2\gamma-1)\psi\Delta p_j$, where γ is a random number between 0 and 1.

2. Then, the mean value and covariance matrix are

$$\langle\sigma_i^{the}\rangle=\frac{\sum_k\sigma_i P(p_1^k,p_2^k,\cdots,p_j^k,\cdots)}{\sum_k P(p_1^k,p_2^k,\cdots,p_j^k,\cdots)}$$

$$\langle\sigma_i^{the}\sigma_{i'}^{the}\rangle=\frac{\sum_k\sigma_i\sigma_{ii'}P(p_1^k,p_2^k,\cdots,p_j^k,\cdots)}{\sum_k P(p_1^k,p_2^k,\cdots,p_j^k,\cdots)}$$

$$V_{ii'}^{the}=\langle\sigma_i^{the}\sigma_{i'}^{the}\rangle-\langle\sigma_i^{the}\rangle\langle\sigma_{i'}^{the}\rangle$$

Taking into account experimental data

1. Experimental data and uncertainties are needed. The most likely, the covariance matrix V_{exp} is unknown. Uncertainties should be used to construct diagonal V_{exp}.

2. One should use a Monte-Carlo approach again because number of variables (σ_i) are now determined by the number of grid points. Now $\sigma_i^k=\langle\sigma_i\rangle+(2\gamma-1)\psi\sqrt{V_{ii}^{the}}$ i.e. σ_i is in the interval $\langle\sigma_i\rangle\pm\gamma\Delta\sigma_i$.

3. Then, the mean value and the covariance matrix are

$$\langle\sigma_i^{eva}\rangle=\frac{\sum_k\sigma_i P_{\exp}(\vec{\sigma})P_{the}(\vec{\sigma})}{\sum_k P_{\exp}(\vec{\sigma})P_{the}(\vec{\sigma})};\qquad\langle\sigma_i^{eva}\sigma_{i'}^{eva}\rangle=\frac{\sum_k\sigma_i\sigma_{i'}P_{\exp}(\vec{\sigma})P_{the}(\vec{\sigma})}{\sum_k P_{\exp}(\vec{\sigma})P_{the}(\vec{\sigma})}$$

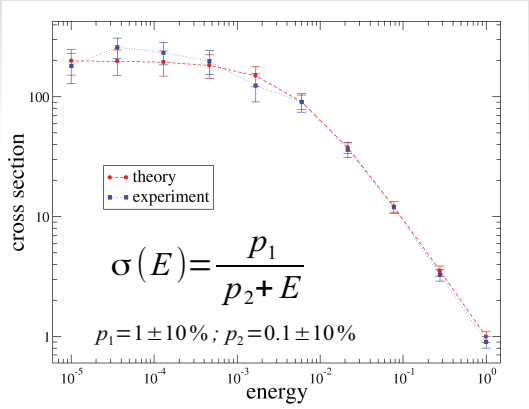
$$V_{ii'}^{eva}=\langle\sigma_i^{eva}\sigma_{i'}^{eva}\rangle-\langle\sigma_i^{eva}\rangle\langle\sigma_{i'}^{eva}\rangle$$

$$P_{the}(\vec{\sigma})=\exp\left\{-\frac{1}{2}\left[(\vec{\sigma}-\vec{\sigma}^{the})(\hat{V}^{the})^{-1}(\vec{\sigma}-\vec{\sigma}^{the})\right]\right\}$$

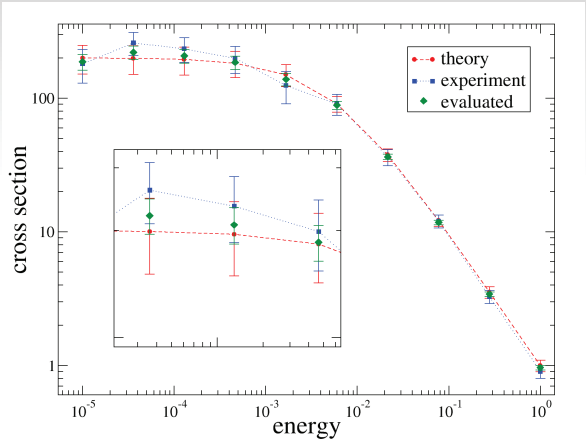
$$P_{\exp}(\vec{\sigma})=\exp\left\{-\frac{1}{2}\left[(\vec{\sigma}-\vec{\sigma}^{\exp})(\hat{V}^{\exp})^{-1}(\vec{\sigma}-\vec{\sigma}^{\exp})\right]\right\}$$

Example

1. Experimental and theoretical cross sections



Evaluated cross section



Conclusions

1. In many situations, it is possible to assess the accuracy of theoretical calculations because theoretical models usually rely on parameters that are uncertain.

2. If there are one or several such parameters, even if some or all parameters are correlated, the discussed approach gives a conceptually straightforward way to calculate uncertainties of final cross sections (uncertainty propagation).

3. Numerically, the statistical approach to the uncertainty propagation could be computationally expensive. However, in situations, where uncertainties are considered to be as important as the actual cross sections (for data validation or benchmark calculations, for example), such a numerical effort is justified.

Conclusions

4. Having data from different sources (say, from theory and experiment), a systematic statistical approach allows one to compare the data and produce "unbiased" evaluated data with improved uncertainties, if uncertainties of initial data from different sources are available.

5. Without uncertainties, the data evaluation/validation becomes impossible. This is the reason why theoreticians should assess the accuracy of their calculations in one way or another. A statistical approach is preferable.

Thank you for your attention