Comparative Analysis of Determination and Account of Unrecognized Source Uncertainty (USU) Covariances in the Data Evaluation Process

V.G. Pronyaev
with contribution from G. Schnabel and S.A. Badikov

October 2023

Detailed report: USU covariances - Methods3.doc

Test1 for intercomparison of of the methods of the USU introducing in the evaluation

To make the presentation of methods more transparent and intercomparable, the specific set of pseudo-experimental data (test1) was prepared.

Test1 case is a set of data with uncertainties for $^{235}U(n,f)$ absolute cross sections.

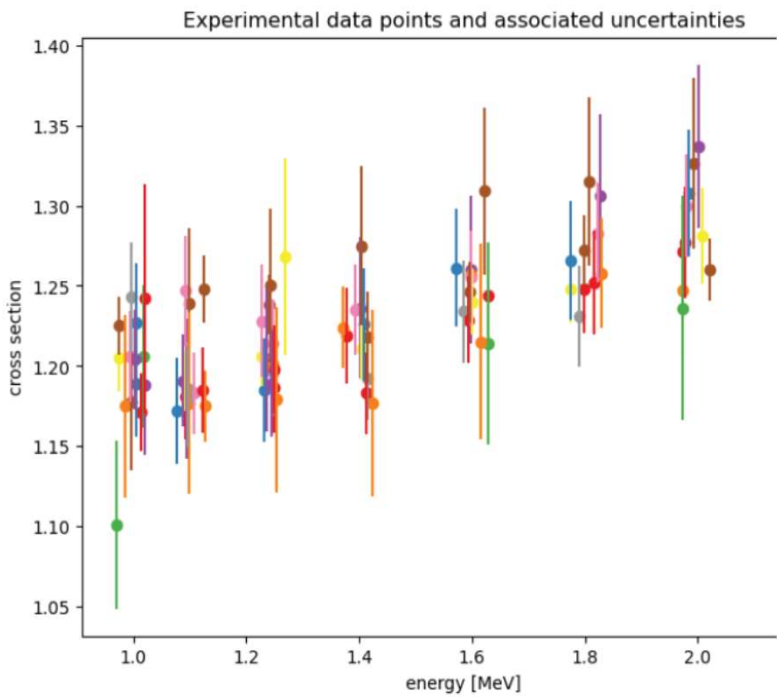
The energy range between 1 and 2 MeV, where cross section is relatively flat, contains 19 data sets with 85 data values at 7 energy nodes.

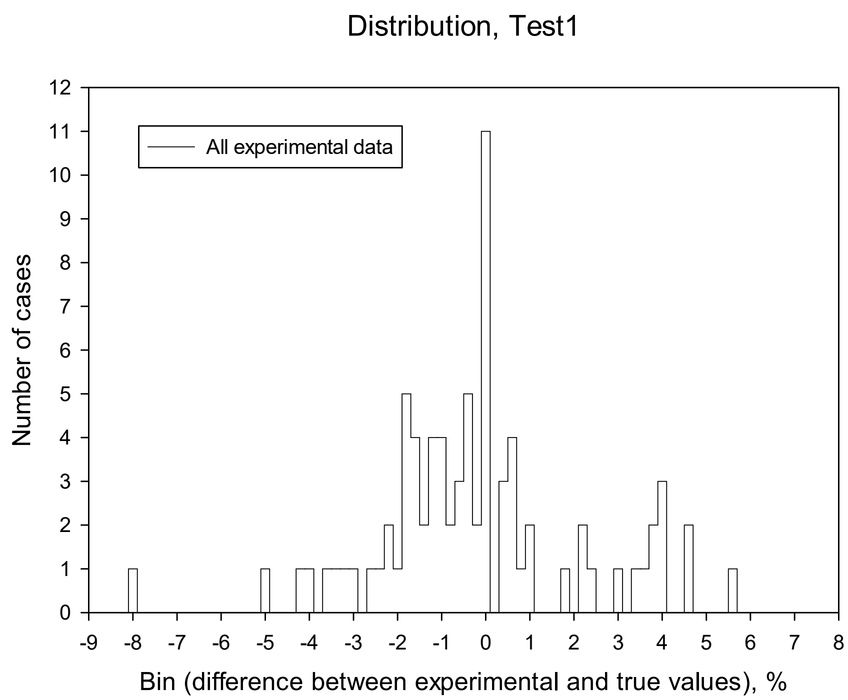
The covariance matrix for each data set include 2 components of uncertainty: statistical and normalization.

The normalization component is substantially higher than the statistical component for most data.

This gives strong cross-energy correlations and large contribution in chi-square if data have different shapes.

Test1 experimental data





Cross sections given at the nodes with artificial spread to make them visible and distribution of the experimental values relative some evaluated ("true") values

Test1 experimental data

Typical low-triangle correlation matrix of experimental data

	1	2	3	4	5	6	7
1	1.00						
2	0.85	1.00					
3	0.85	0.78	1.00				
4	0.90	0.83	0.83	1.00			
5	0.92	0.85	0.85	0.90	1.00		
6	0.93	0.86	0.86	0.92	0.93	1.00	
7	0.93	0.86	0.86	0.92	0.93	0.95	1.00.

High level of cross-energy correlations contributes substantially in the chi-square of the fit when data have different shapes

Approaches for introducing USU covariances

Three approaches for introducing USU covariances were used in test1 comparison:

- **Ad hoc approach for outlaying experimental data in the Standards data base (V.G. Pronyaev);**
- **Statistical approach, based at the analysis of the experimental values spread (S.A. Badikov);**
- **Statistical approach, based at the marginalization of USU variances common at each energy node for all experimental data of given reaction (G. Schnabel).**

Slides below present my understanding of the methods and interpretation of the results. The authors of the approaches will give more detailed discussions of their methods.

Ad hoc approach for outlaying experimental data

Ad hoc approach was used in Standards 2007 and Standard 2017 evaluation for the work with outlaying data. The following procedure was used:

- **GMA fit was done for all data as they presented in the GMA database**
- **The obtained evaluation was considered as an approximation to the true value**
- **Experimental data outlaying from true value more than at 2 sigma at single node and at 1 sigma at 2 or more consecutive nodes for given data set were considered as outlaying data**
- **Additional component of uncertainties were added to the uncertainties in these points nodes**
- **The medium energy range correlation was used for calculations of off-diagonal covariances between the nodes with outlaying data**
- **The fit was repeated with obtaining new true values and iterative procedure was used up to the convergence for true value**

Ad hoc approach for outlaying experimental data: formulation

Such treatment of uncertainties for outliers can be replaced by the approach with creating the ad hoc USU covariances.

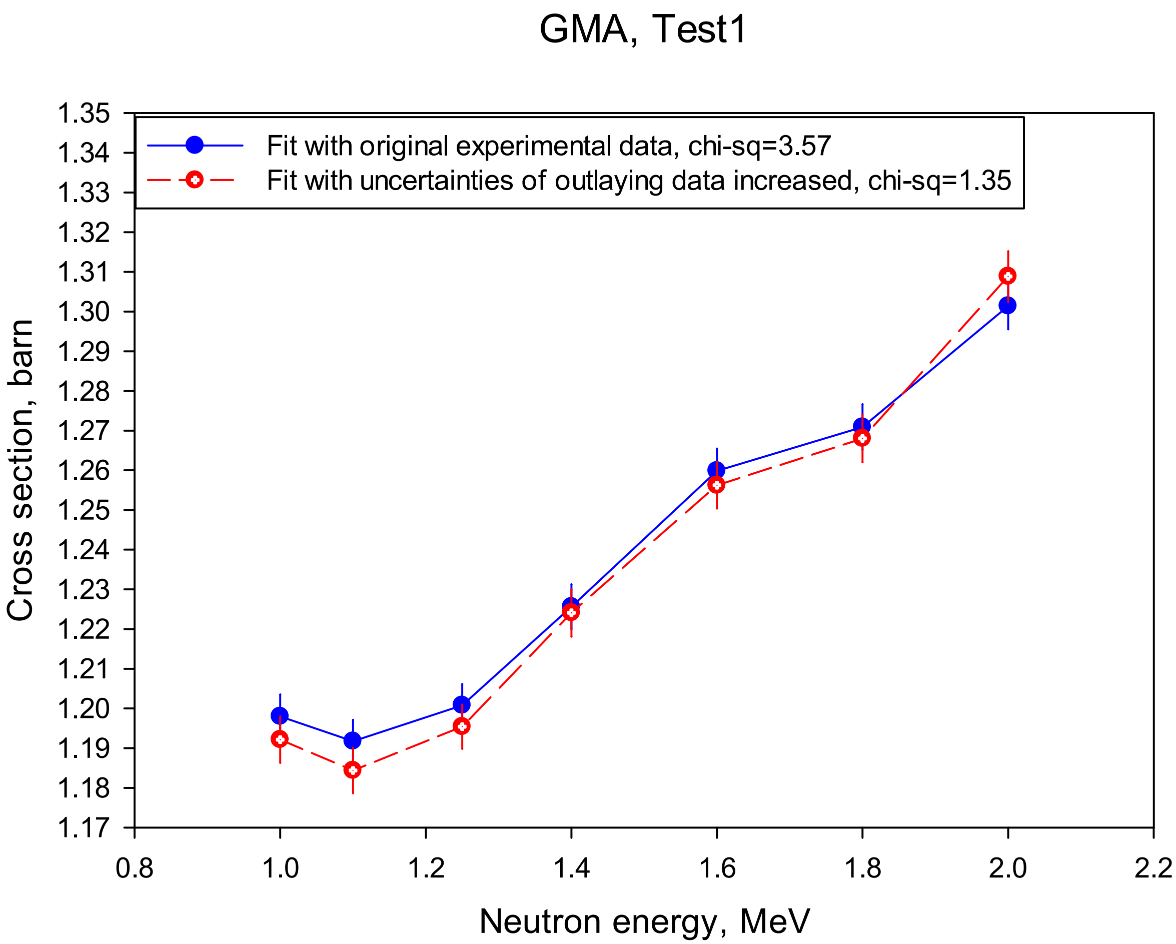
The USU covariances between energy nodes *i* and *j* for any experimental data set can be written as:

$$COV_{ij} = \sqrt{\left(Y_i - Y_{i,true}\right)^2 - \delta Y_i^2} \times \sqrt{\left(Y_j - Y_{j,true}\right)^2 - \delta Y_j^2}$$
$$\times (Y_i - Y_{i,true})/\left|Y_i - Y_{i,true}\right| \times (Y_j - Y_{j,true})/\left|Y_j - Y_{j,true}\right|,$$

where terms in first line define the absolute value of covariances and in second line – the sign of the covariances.

These USU covariances should be added to the experimental covariances for outlaying data

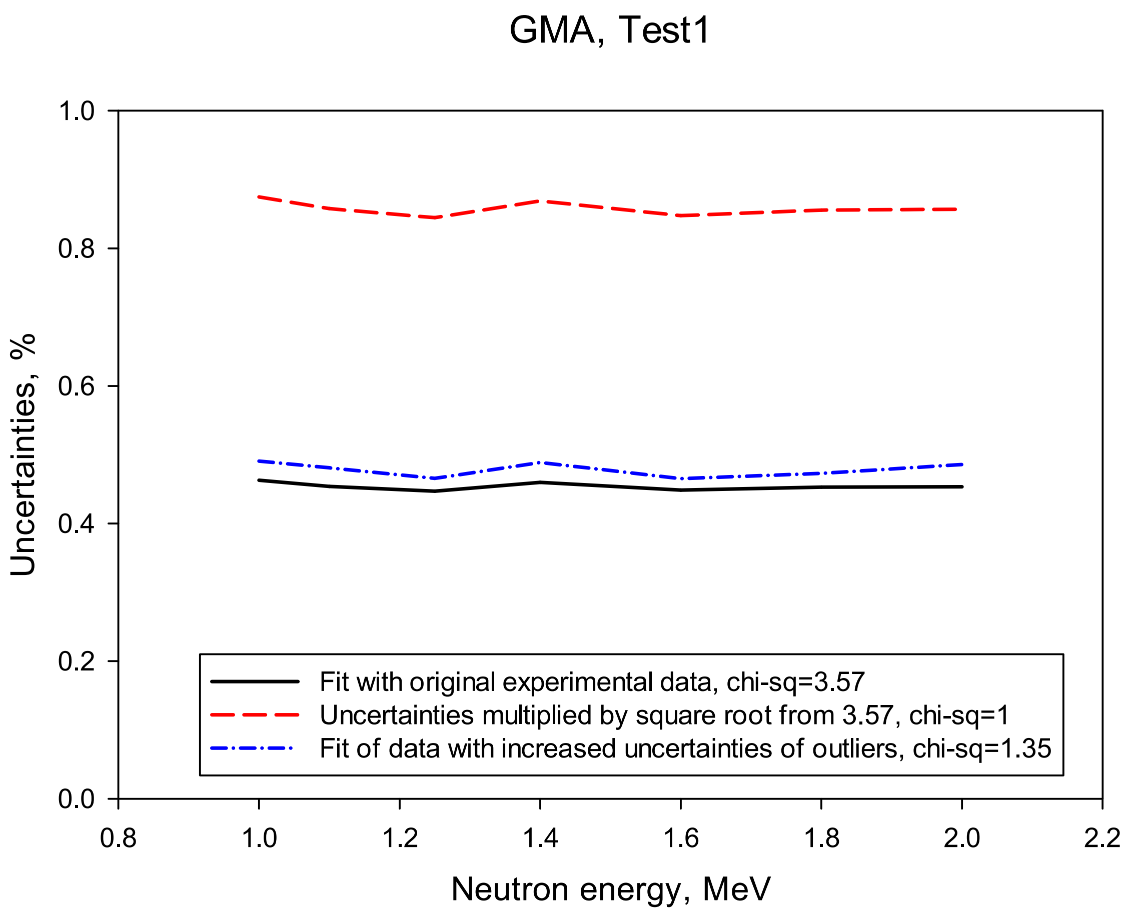
Ad hoc approach for outlaying experimental data: results



Evaluated values obtained with original experimental data are compared with evaluated values obtained in the fit with increased uncertainties of for outliers

(as in Standards2017)

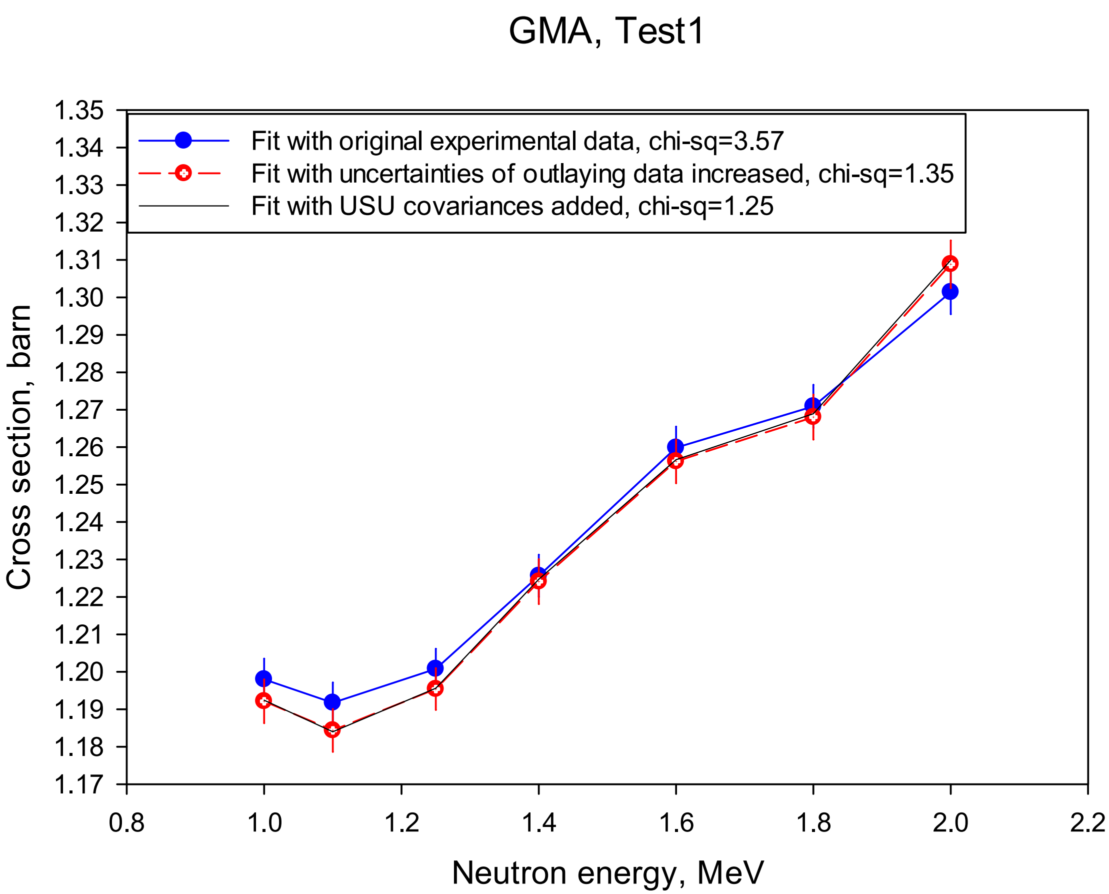
Ad hoc approach for outlaying experimental data: results



Evaluated % uncertainties obtained with original experimental data are compared with evaluated % uncertainties obtained in the fit with increased uncertainties of outliers

Uncertainties in the fit with original data multiplied at square root from chi-square are shown

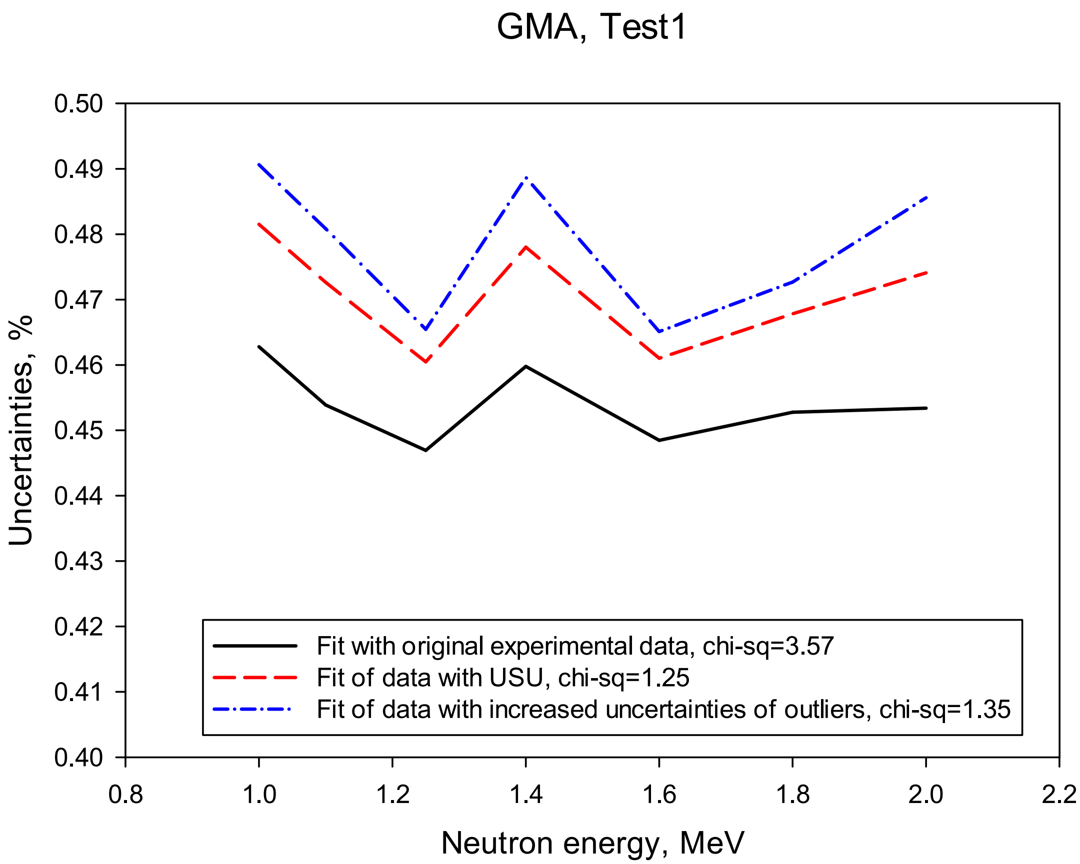
Ad hoc approach for outlaying experimental data: results



Evaluated values obtained in the fit with added USU covariances are compared with evaluated values obtained in the fit with original data and with data with increased uncertainties of for outliers

Evaluations with USU covariances and with increased uncertainties of outlaying data are practically practically indistinguishable but chi-square is reduced from 1.35 to 1.25

Ad hoc approach for outlaying experimental data: results



% uncertainties obtained in the fit with added USU covariances are compared with % uncertainties obtained in the fit with original data and with data with increased uncertainties of for outliers

Percent uncertainties (or variances) with USU covariances and with increased uncertainties of outlaying data are only slightly different but chi-square is reduced from 1.35 to 1.25

Ad hoc approach for outlaying experimental data: results

Node, MeV	Elements of USU covariance matrix, $(\%)^2$						
1.00	5.76	12.48	7.92	0.00	0.00	0.00	-8.16
1.10	12.4	27.04	17.16	0.00	0.00	0.00	-17.68
1.25	7.92	17.16	10.89	0.00	0.00	0.00	-11.22
1.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2.00	-8.16	-17.68	-11.22	0.00	0.00	0.00	11.56

Typical USU covariances for experimental data set covering all energy range.
The correlation matrix consists from elements +1 and -1

Ad hoc approach for outlaying experimental data: results

Without USU						
1.00						
0.81	1.00					
0.83	0.85	1.00				
0.80	0.82	0.85	1.00			
0.83	0.84	0.87	0.85	1.00		
0.83	0.83	0.87	0.85	0.87	1.00	
0.83	0.83	0.87	0.85	0.87	0.87	1.00

With USU						
1.00						
0.78	1.00					
0.82	0.82	1.00				
0.77	0.78	0.83	1.00			
0.81	0.81	0.86	0.83	1.00		
0.81	0.80	0.86	0.81	0.85	1.00	
0.81	0.79	0.85	0.81	0.84	0.84	1.00

All correlation elements of the evaluated covariance uncertainty matrix are decreased

Ad hoc approach for outlaying experimental data: conclusion

We may expect that the ad hoc approach for introducing of USU covariances in the GMA fit of standards data will have small influence at the evaluated values and their uncertainties in comparison with the 2017 standard evaluation

Variance analysis approach

The main points of the approach are the following:

- **a priori estimate of the cross section is not used;**
- **declared measurement uncertainties are not used;**
- **distributions of statistical and systematic (normalization) components of the uncertainties are estimated from the distributions of the experimental values;**
- **model function is used for estimation of the off-diagonal covariances.**

Variance analysis approach

The uncertainties of the systematic errors are assumed to be equal for all the measurements carried out within the group. A systematic error corresponding to an experiment k $(k=1,\dots,K)$ in the group m $(m=1,\dots,M)$ is calculated as an average deviation of measurements $y_{i_k}^{k,m}$ from the values of the model function

$$\eta^{k,m} = \frac{1}{n_k}\sum_{i=1}^{n_k}(y_{i_k}^{k,m} - f(E_{i_k}^{k,m}, \hat{\vec{\theta}}))\,.$$

A diagonal ($L_1 = L_2$) or near diagonal ($L_1 = L_2 - 1$) Pade-approximant

$$f^{[L_1,L_2]}(E,\vec{\theta}) \equiv c + \sum_{i=1}^{I}\frac{a_i}{E - r_i} + \sum_{j=1}^{J}\frac{\alpha_j(E-\zeta_j)+\beta_j}{(E-\zeta_j)^2+\gamma_j^2}$$

was used as a model function ($L_1 + L_2 + 1 = 2I + 4J + 1$).

Variance analysis approach

The systematic errors $\eta^{k,m}$ form a distribution (over experiments) of the systematic errors within the group m. The variance V_{mm} of the distribution is an estimate for the variance of the systematic errors of the measurements carried out within the group

$$V_{mm} = \frac{1}{K-1}\sum_{k=1}^{K}(\eta^{k,m} - a_m)^2 ,$$

where a_m is an estimate of the mean of the distribution

$$a_m = \frac{1}{K}\sum_{k=1}^{K}\eta^{k,m} .$$

Variance analysis approach

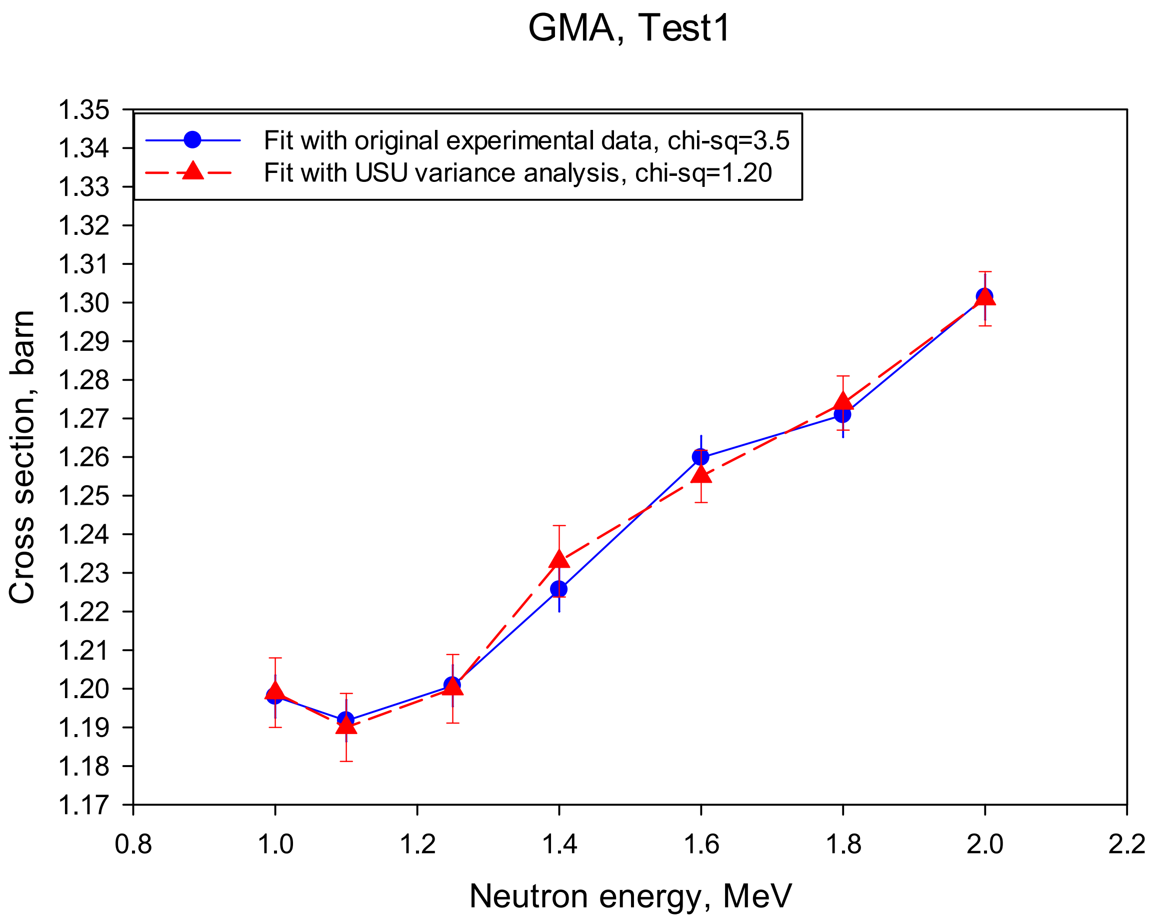
The covariances of the systematic errors are estimated as follows

$$V_{m_1 m_2} = \frac{1}{K}\sum_{k=1}^{K} \eta^{k,m_1}\eta^{k,m_2} - a_{m_1}a_{m_2}$$

The deviations of the residuals $(y_{i_k}^{k,m} - f(E_{i_k}^{k,m}, \hat{\bar{\theta}}))$ from the systematic error $\eta^{k,m}$ of the k-th experiment in the group m form another distribution (over measurements) corresponding to the k-th experiment. The variance of this distribution represents the variance of the statistical errors of the measurements carried out within the the k-th experiment in the group m.

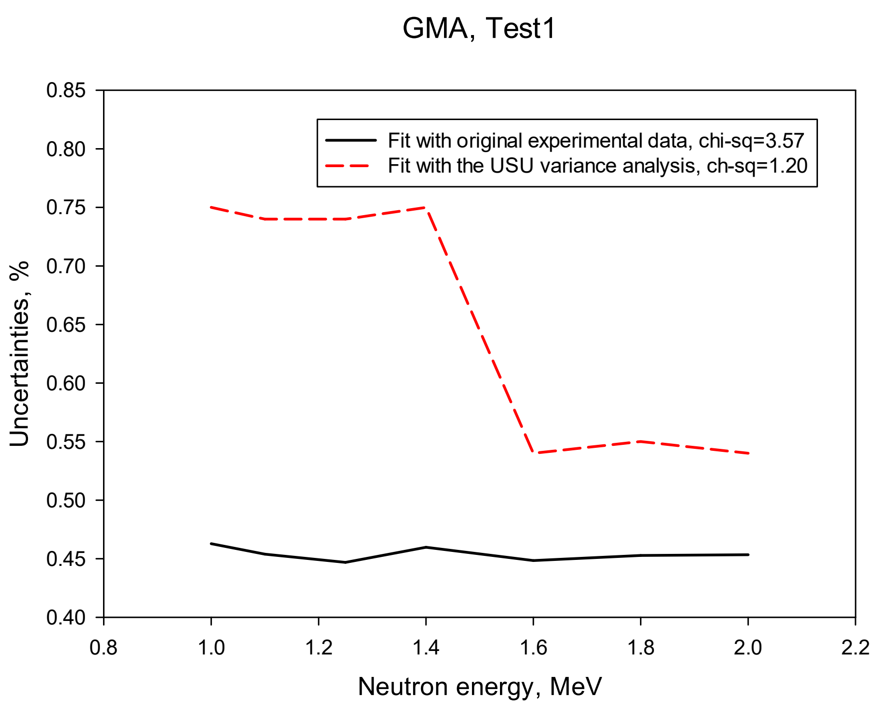
The evaluated values and their covariances are calculated by the least squares method. The calculations are organized in an iterative way. The evaluated curve, calculated in assumption of equal unknown uncertainties of the experimental errors, is used as an initial approximation.

Variance analysis approach: results



The evaluated values with the variance USU account are rather close to the fit with the original data and with ad hoc USU account

Variance analysis approach: results



1.00						
0.84	1.00					
0.84	0.87	1.00				
0.82	0.85	0.85	1.00			
0.83	0.85	0.85	0.85	1.00		
-0.01	-0.27	0.65	-0.37	-0.02	1.00	
0.82	0.85	0.85	0.85	0.85	-0.17	1.00

% uncertainties of the fit with the USU account are substantially larger than without account of the USU. Correlation matrix of USU component

Variance analysis approach: conclusion

The variance analysis approach may satisfy to the CSEWG 1991 requirements that 2/3 of all experimental data are in the limits of the evaluated uncertainties but it is difficult to implement for large GMA database

Bayesian approach with Monte Carlo sampling (MCMC)

The Bayesian approach with MC sampling uses simple model of USU introducing. Bayesian formulation in general form for non-model data fit can be written as:

$T' = T + dT = T + MG^{+}(GMG^{+} + V)^{-1}(R - T)$

$M' = M + dM = M - MG^{+}(GMG^{+} + V)^{-1}GM,$

T' **is a vector of ("posteriori") evaluated data,**

T **is a vector of "priori" evaluated data,**

M' **is a covariance matrix of uncertainties of (posteriori) evaluated data,**

M **is a covariance matrix of uncertainties of (priori) evaluated data,**

R **is a vector of experimental data,**

V **is a covariance matrix of uncertainty of the experimental data,**

G **is a matrix of the sensitivity coefficients of the data reduction or the reaction combination, upper indexes (+) and (-*1*) means the operators of the matrix transposing or inversion.**

Bayesian approach with Monte Carlo sampling (MCMC)

There are indexes which are not shown:

- **energy node,**
- **type of reaction (fission, capture,…),**
- **type of data (absolute, shape, absolute ratio, shape of ratio), data combination (sums, integrals).**

GMA and its updated version GMAPy use the Generalized Least Squares approach which, in case of the first uninformative prior is identical to the Bayesian approach (proof by N. Larson):

$T' = (G^{+}V^{-1}G)^{-1}G^{+}V^{-1}R$

$M' = (G^{+}V^{-1}G)^{-1}$

with the same designations as for Bayesian approach

Bayesian approach with Monte Carlo sampling (MCMC)

Model for introducing of USU for test1 was formulated by replacing the covariance matrix of experimental data

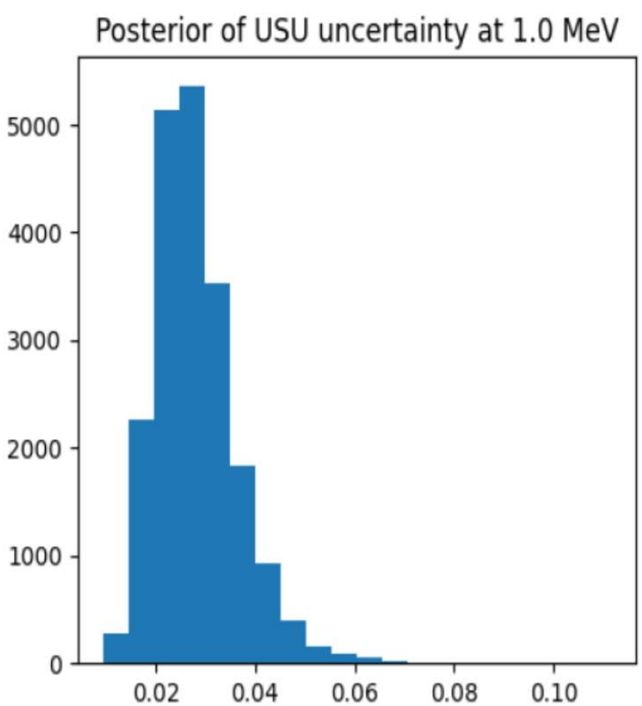
V_{ij}^{k} **on** $V_{ij}^{k} + V_{ij}^{USU} \times \delta_{ij}$,

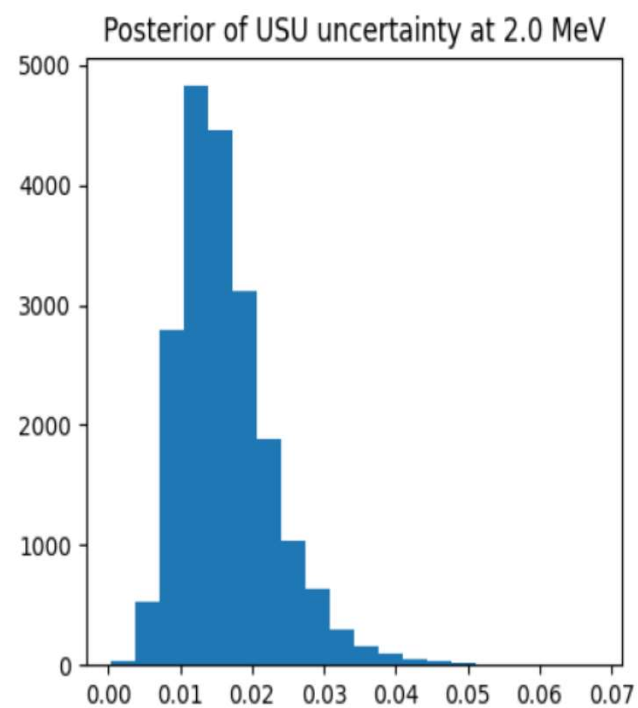
where *i,j* are indexes of the energy nodes, *k* is an index of experimental data set, δ_{ij} **is a Kronecker symbol, and** $V_{ij}^{USU} \times \delta_{ij}$ **is covariance matrix of USU.**

This diagonal matrix (variances) is the same for all experimental data sets.
More complex model will be with index *k* assigned to V_{ij}^{USU}**, and even more complex with** δ_{ij} **replaced at *1.***

20,000 Monte Carlo samplings for USU variances was used with a following Bayesian fit. Flat (uninformative) distribution of USU variances was used as a prior.

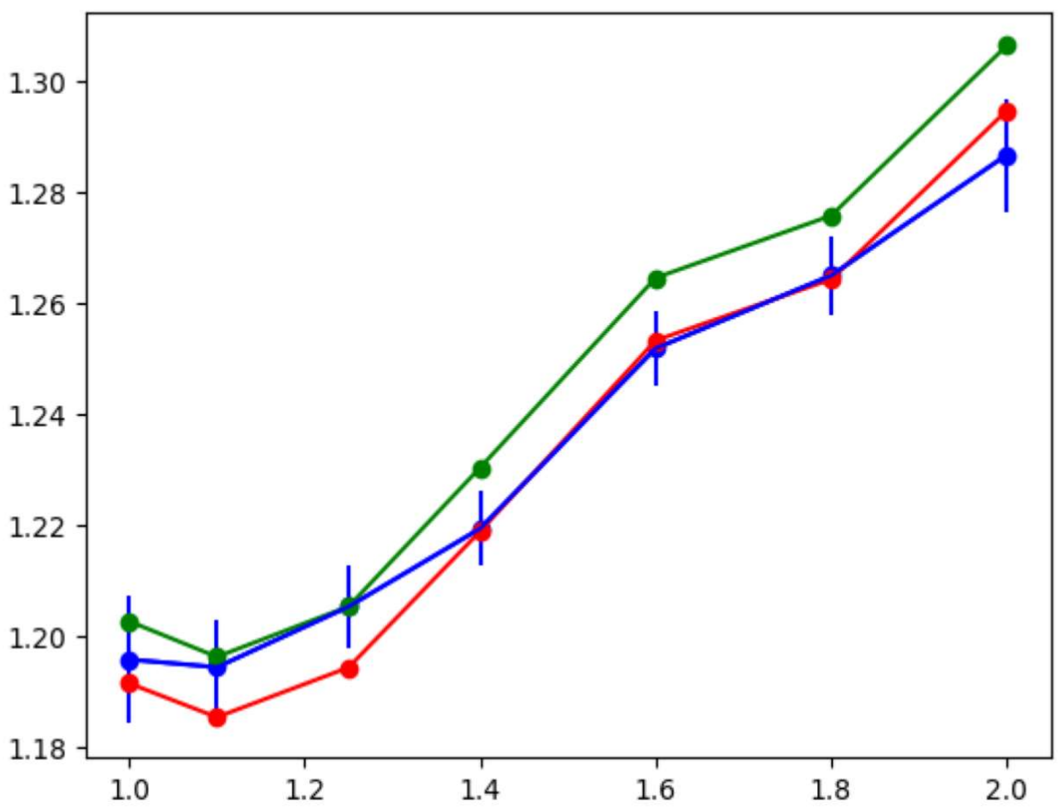
Bayesian approach with Monte Carlo sampling: results





Posterior USU uncertainties are obtained from their distributions, which usually are not normal distributions. In some cases they can be uninformative.

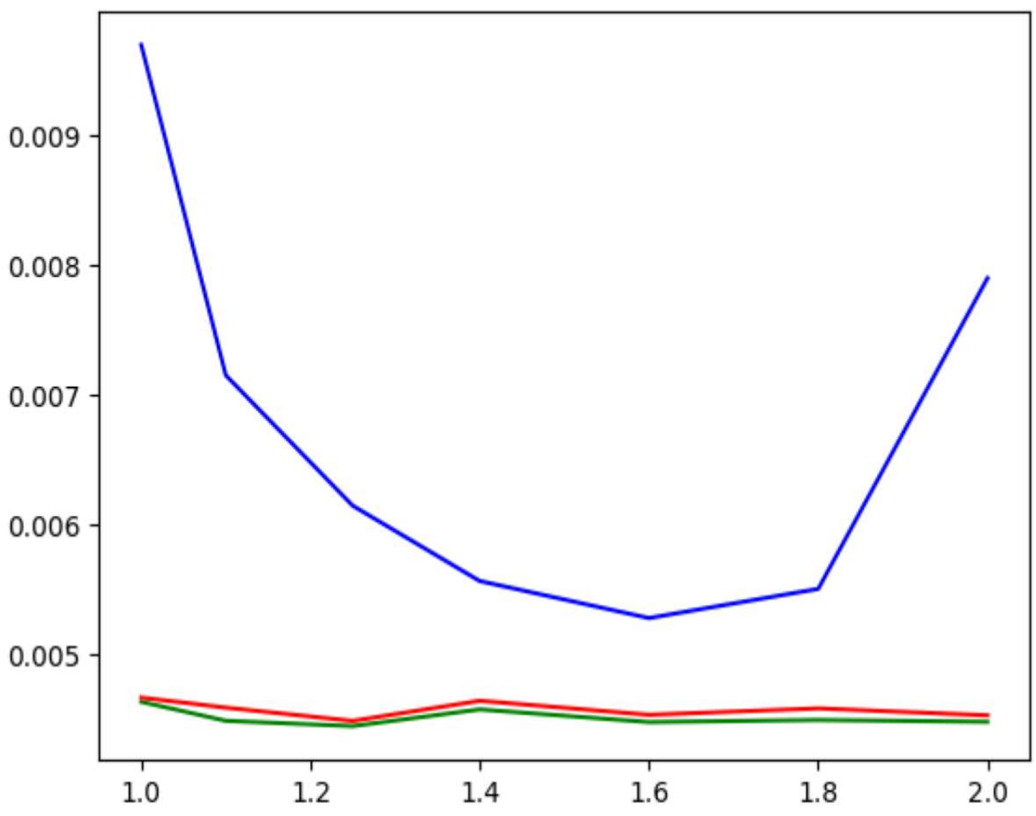
Bayesian approach with Monte Carlo sampling: results



Evaluated cross sections obtained without correction at PPP and without USU (red signs and curve), with correction at PPP but without USU (green signs and curve), with correction at PPP and with USU (blue signs and curve)

Relative uncertainties are used in the fit to exclude PPP
Large influence of USU at the evaluation

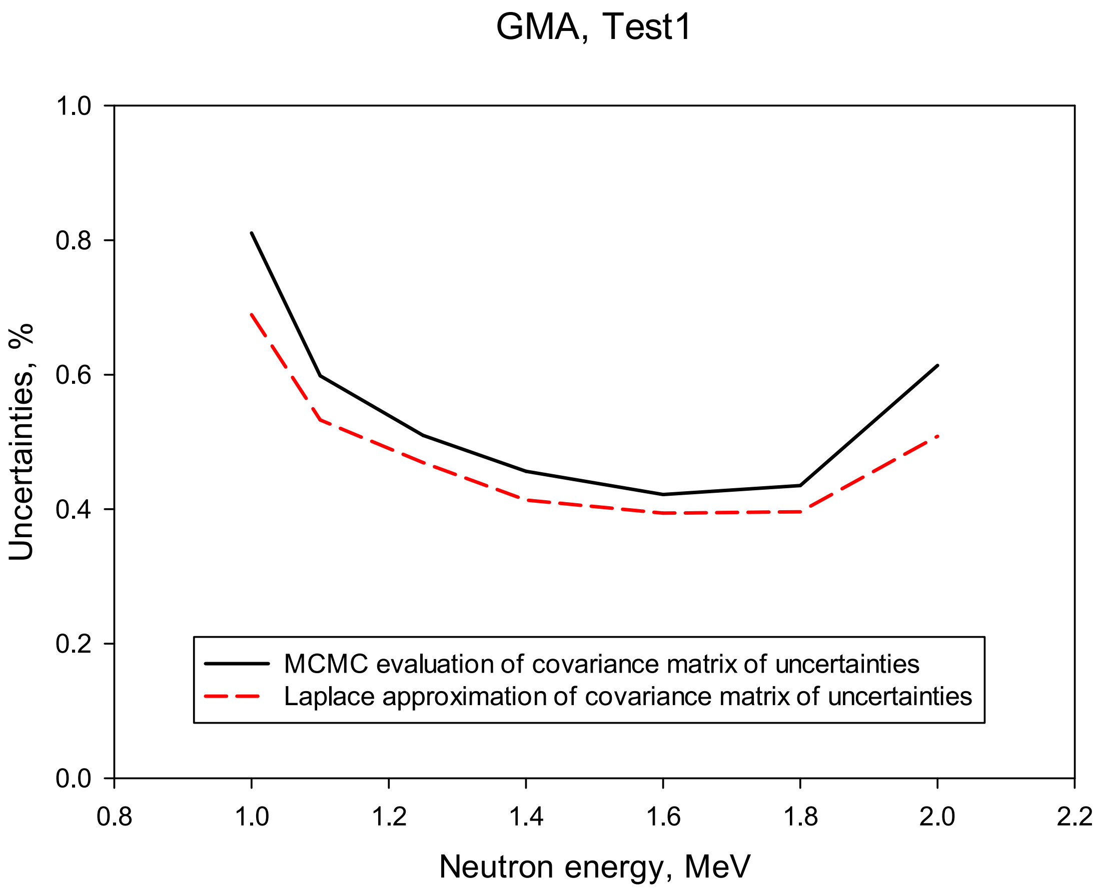
Bayesian approach with Monte Carlo sampling: results



Relative uncertainties of the evaluated data are substantially increased.

Relative uncertainties obtained without correction at PPP and without USU (red signs and curve),
with correction at PPP but without USU (green signs and curve),
with correction at PPP and with USU (blue signs and curve)

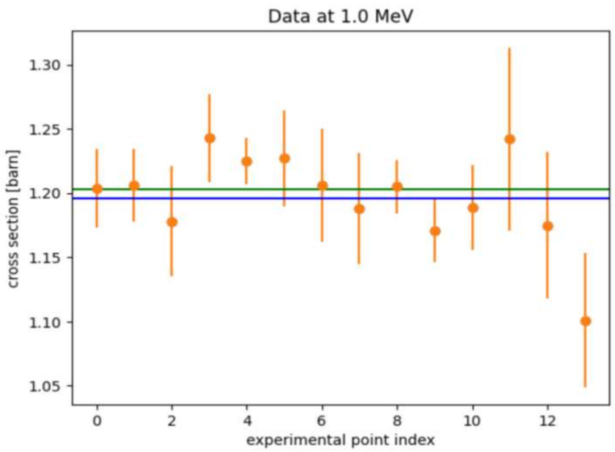
Bayesian approach with Monte Carlo sampling: results

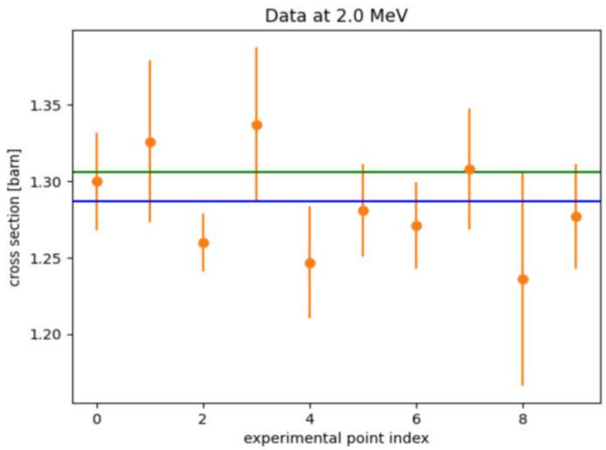


Laplace approximation of covariance matrix of evaluated uncertainties was applied to see the influence of non-linear behavior in the parameter space.

There is an influence, but it is not great.

Bayesian approach with Monte Carlo sampling: results

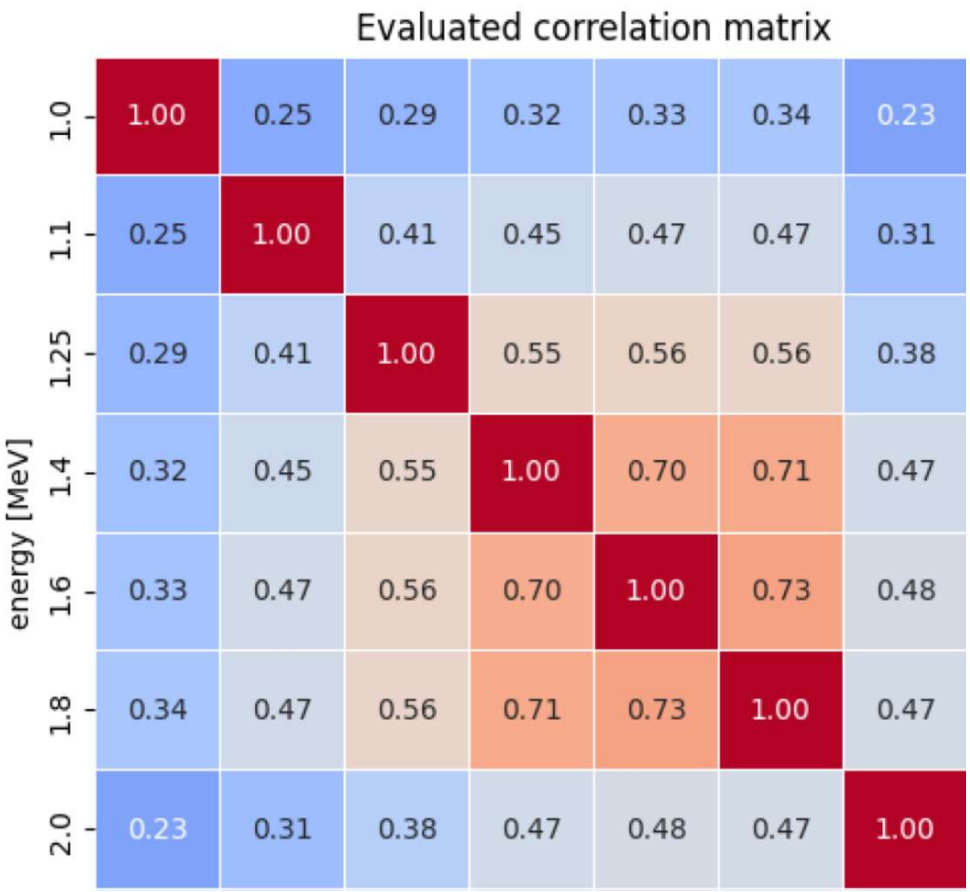




Evaluation at two nodes in comparison with evaluated values: green - with correction at PPP but without USU account, blue - with correction at PPP and with USU account.

Large contribution of USU at 2 MeV decreases the strong cross energy correlations in the original experimental data, what shifts the evaluated value more close to the expected in the absence of correlations. But large contribution of USU at 1 MeV node don't shift practically the evaluated value obtained without account of USU, because this value is close to the expected without account of cross energy correlations.

Bayesian approach with Monte Carlo sampling: results



Correlation matrix of data evaluated with account of USU has much lower cross energy correlations in comparison with the matrix obtaine in the fit with original data.

Without USU						
1.00						
0.81	**1.00**					
0.83	**0.85**	**1.00**				
0.80	**0.82**	**0.85**	**1.00**			
0.83	**0.84**	**0.87**	**0.85**	**1.00**		
0.83	**0.83**	**0.87**	**0.85**	**0.87**	**1.00**	
0.83	**0.83**	**0.87**	**0.85**	**0.87**	**0.87**	**1.00**

Bayesian approach with Monte Carlo sampling: conclusion

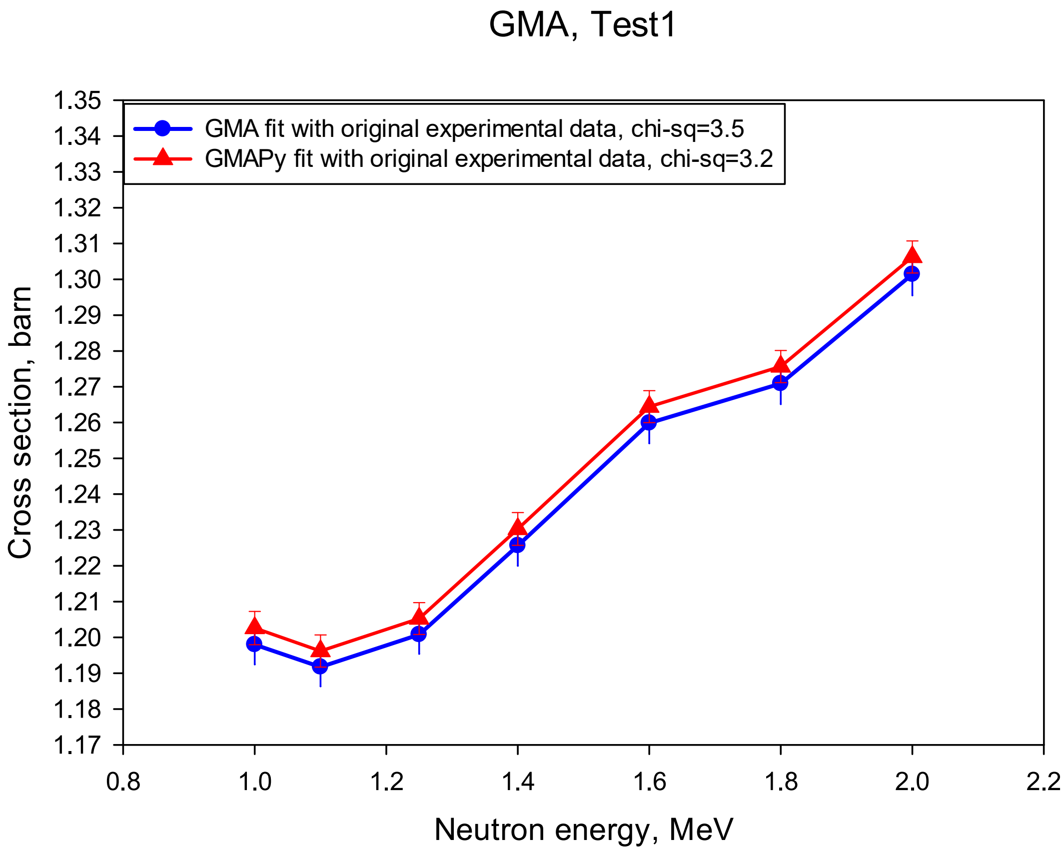
Bayesian approach with MC sampling gives medium increase of the uncertainty evaluated data with the USU account.

It can be practically implemented only using wide energy mesh (small number of USU parameters).

The current implementation of the accounting for the USU may lead to a shift in the evaluated cross sections of the order of the evaluated uncertainties.

High-precision and low-precision measurements are “penalized equally”, which leads to a worsening impact on the evaluation of the results of high-precision measurements. Will experimenters agree with this?

Intercomparison of the results

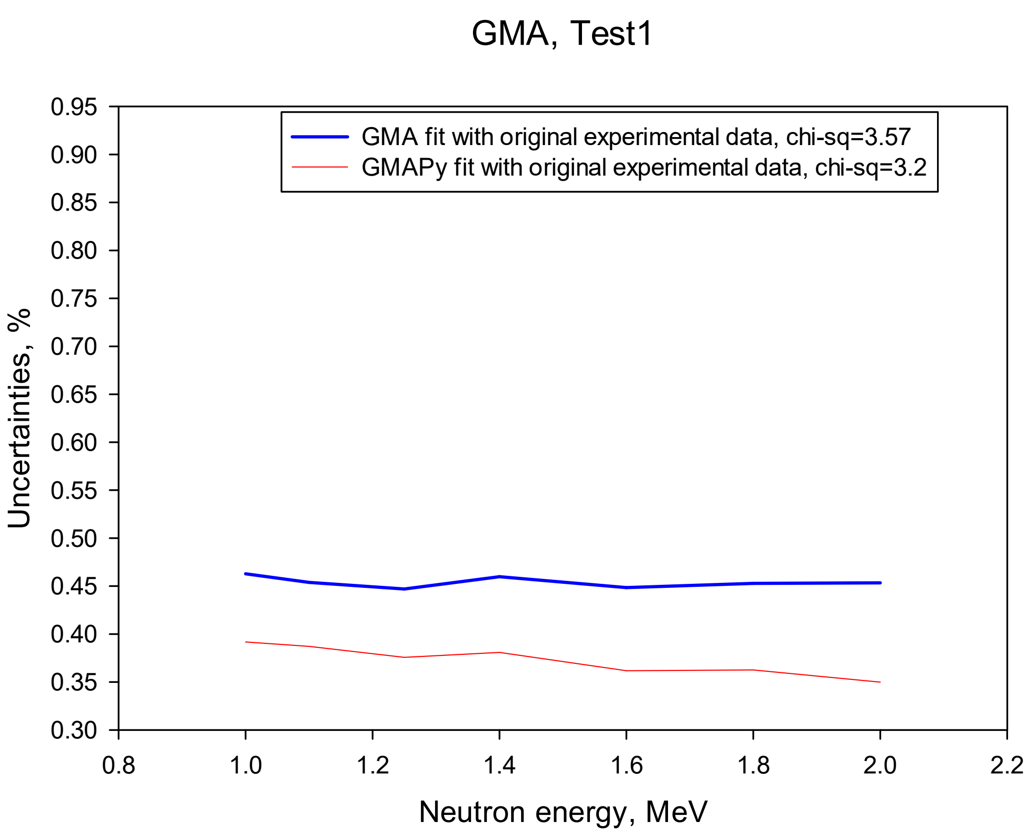


GMA and GMAPy use slightly different technical fixes for uncertainties of experimental datato to exclude the PPP.

GMAPy uses relative covariances in the PPP exclusion, GMA uses covariances with variances replaced at square of relative uncertainties times posterior evaluation.

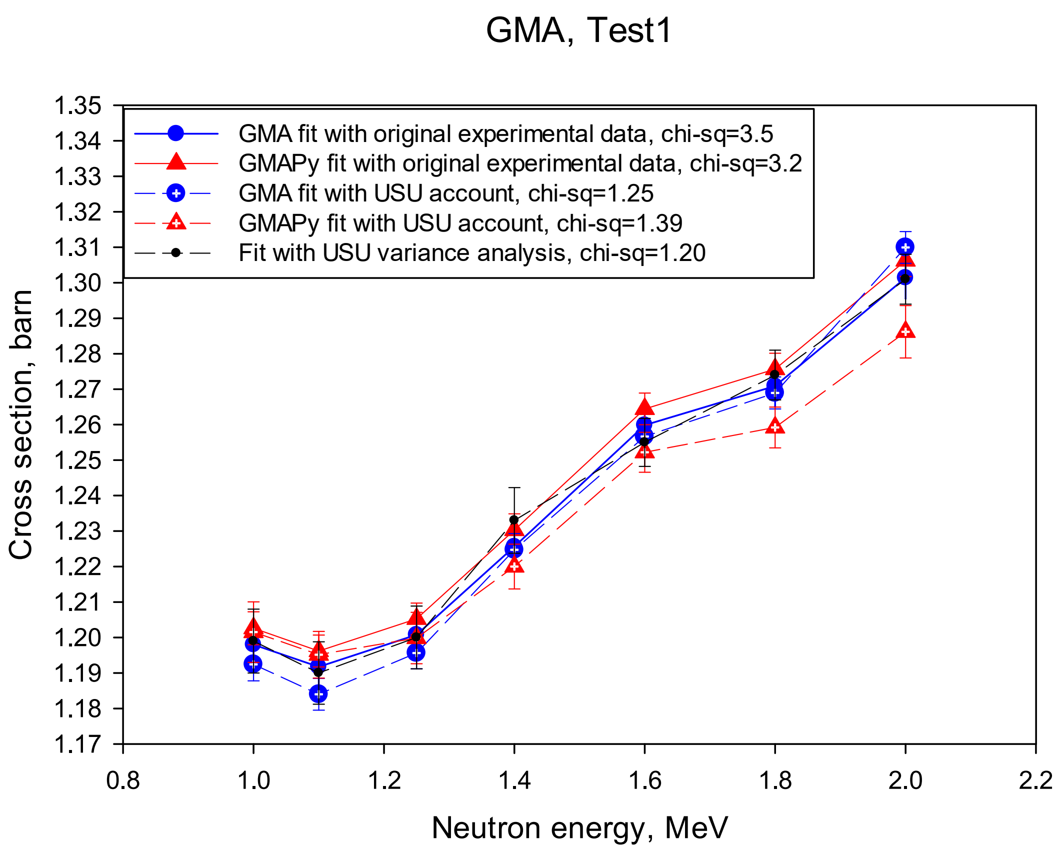
+0.38% higher GMAPy

Intercomparison of the results



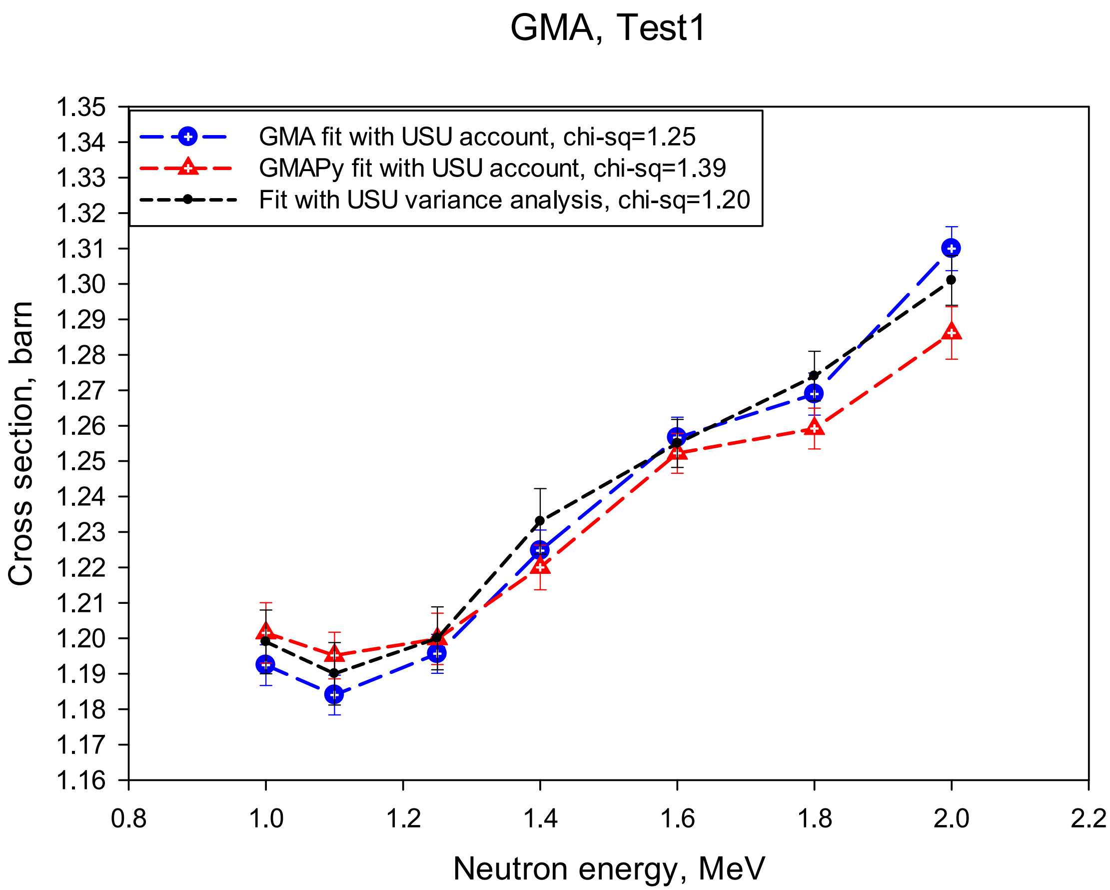
The difference in the PPP account in GMA and GMAPy leads also to the difference in the evaluated uncertainties

Intercomparison of the results



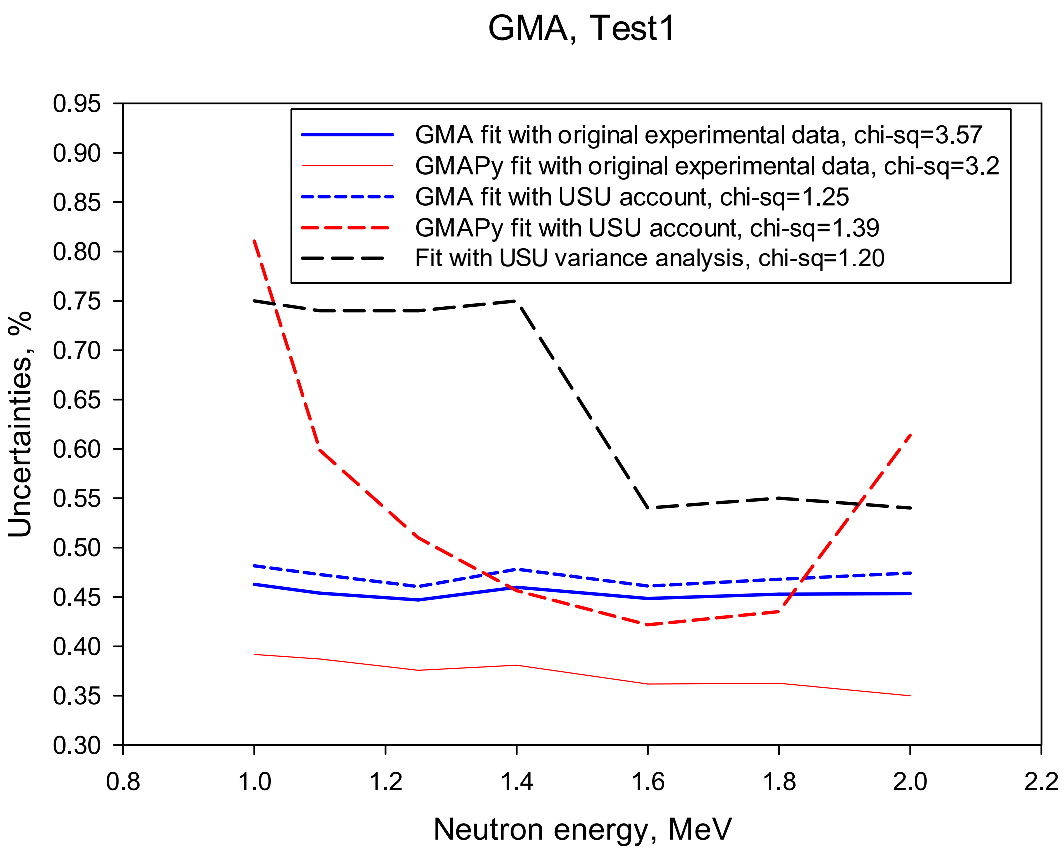
Spread of the evaluated results for different options used with and without USU account

Intercomparison of the results



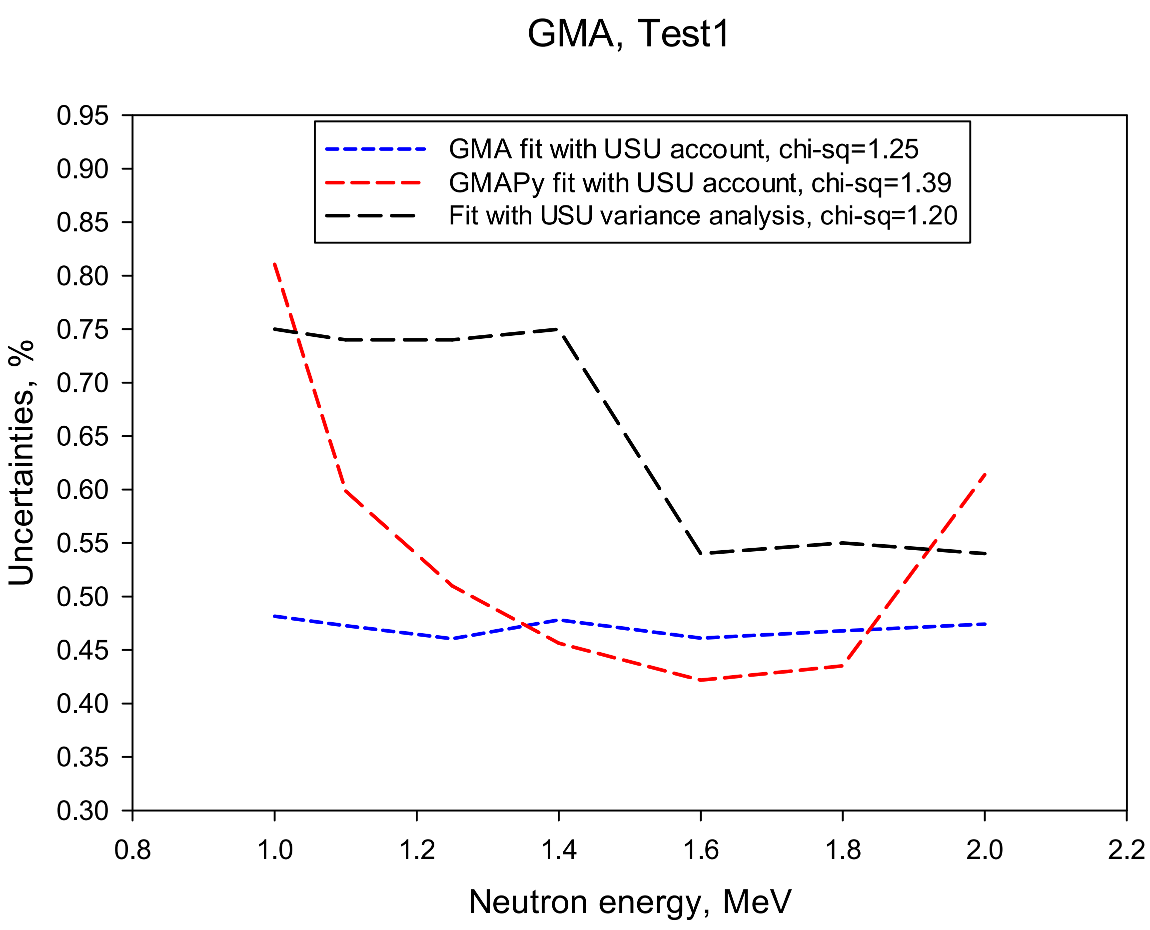
Evaluated values obtained with three methods of USU account

Intercomparison of the results



Evaluated percent uncertainties values obtained for different options used with and without USU account

Intercomparison of the results



Evaluated percent uncertainties obtained with three methods of USU account

Table of main USU account features

	Ad hoc analysis	Variance analysis	Bayesian with MC
USU covariances of experimental data	Estimated for each energy node of each experimental data set.	Is a component of total covariance matrix of uncertainty of evaluated data.	Estimated as diagonal covariance matrix (variances only) with the same values for all experimental data sets
USU covariances of evaluated data	Can be calculated as difference between covariance matrix of evaluated data with and without USU account	Can be calculated as difference between covariance matrix of evaluated data with and without USU account	Can be calculated as difference between covariance matrix of evaluated data with and without USU account

Table of main USU account features

	Ad hoc analysis	Variance analysis	Bayesian with MC
USU determination	From values and uncertainties of outlaying experimental data relative a prior “true” values	From statistical analysis of distributions of the experimental values without account of their uncertainties	From statistical analysis of distribution of experimental values relative prior “true” values with account of experimental uncertainties
Impact of the USU at the experimental data	Only at the outlaying data	At all data without separation of USU component	At all data with the same additional USU variances

Table of main USU account features

	Ad hoc analysis	Variance analysis	Bayesian with MC
Impact of the USU at the evaluated values	minor	minor	large
Impact of the USU at the evaluated covariances	minor	medium	large

Table of main USU account features

	Ad hoc analysis	Variance analysis	Bayesian with MC
Main drawback	Ad hoc procedure	Experimental uncertainties are not accounted	The same USU for all experimental data at given node greatly influences the evaluated values, since it reduces the impact of the declared high-precision data
Possibility of further develop-ment	Practically no	The model can be excluded(?)	The use of different USU for different experimental data at the nodes if not too many parameters(?) Consistent high-precision data can be used without USU addition. But how should be the consistent data defined statistically (without ad hoc approach)?

General conclusion

1. **Test1 is a very specific case with strong cross energy correlations in the experimental data. Uncertainties for each experimental data set include only statistical and normalization (fully correlated on energy) components. This is done intentionally to increase impact of the off-diagonal covariances in the data fit.**
2. **Ad hoc method of USU account has minor influence at the evaluated values and covariance matrix of uncertainties. The results of the evaluation with a full GMA database will be probably close to that obtained with the treatment of outlaying data in the 2017 Standards. It is easy implement in the GMA evaluation procedure for each energy node .**
3. **Variance method without consideration of experimental uncertainties gives evaluation close to the ad hoc method and largest contribution of the USU. It is difficult to use in the standards evaluation with a full GMA database.**
4. **Bayesian with MC method gives probably best estimate of diagonal covariance matrix of uncertainties of experimental data. But because the same diagonal covariances are used for USU component in different experimental data it reduces substantially the influence of the high accuracy data at the evaluation. It has computational limitations at the number of energy meshes.**