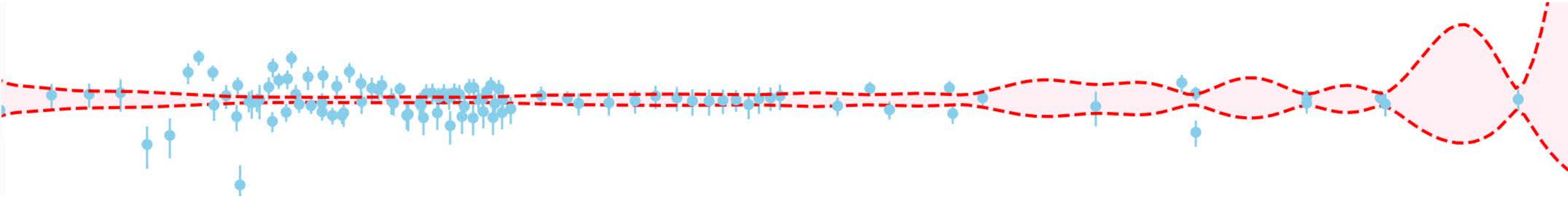
A Comparison of Gaussian Process Regression and GLS Results for ^{235}U(n,f) Cross Section



Hiroki IWAMOTO[1]
[1] JAEA

Contents

Introduction

■ Nuclear data evaluation

- Model-based evaluation

  TALYS, EMPIRE, CCONE, CoH, SUMMY, etc.

- Experimental data-based evaluation (non-model fit)
  - GLS (Generalized Least-Squares method) ☜ Well-established in ND community
  - GPR (Gaussian Process Regression) ☜ Popular in ML

Basic formula of GPR (1/2)

- **Formula**:

$$\Pr(\mathbf{t}_*|\mathbf{t}) = \mathcal{N}(\mathbf{t}_* \mid \boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$$

Mean: $\boldsymbol{\mu}_* = \mathbf{C}_{\mathbf{x}\mathbf{x}_*}^\top \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{t}$

Covariance: $\boldsymbol{\Sigma}_* = \mathbf{C}_{\mathbf{x}_*\mathbf{x}_*} - \mathbf{C}_{\mathbf{x}\mathbf{x}_*}^\top \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\mathbf{x}_*}$

where the partial covariance $\mathbf{C}$ is expressed as

Noise term

$$C(x_i, x_j) = \mathcal{K}(x_i, x_j) + (dt_i)^2 \delta_{ij}$$

☜ **Normal use (white noise, w/o correlation)**

$$\Rightarrow C(x_i, x_j) = \mathcal{K}(x_i, x_j) + \mathrm{cov}(x_i, x_j)$$

☜ **Modified (w/ correlations)**

$$C(x_{*i}, x_j) = \mathcal{K}(x_{*i}, x_j)$$

$$C(x_{*i}, x_{*j}) = \mathcal{K}(x_{*i}, x_{*j})$$

To take into account experimental correlations, a noise term was modified.

Basic formula of GPR (2/2)

- GPR uses a kernel function $\mathcal{K}$ in $\mathbf{C}$ to represent the correlation between training data.
- This study investigated results for two typical kernel functions.

Kernel function

➢ **Radial Basis Function (RBF) kernel**

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

➢ **Matérn 3 kernel**

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$

☜ **Newly incorporated in this work**

where

- ℓ : Length-scale parameter
- σ : Magnitude parameter

☜ **"Hyperparameters"**, which is optimized so that the log-evidence becomes maximum:

$$\mathbf{\Theta}_{\mathrm{opt}} = \underset{\mathbf{\Theta}}{\mathrm{argmax}} \left[\ln p(\mathbf{t}|\mathbf{\Theta})\right]$$

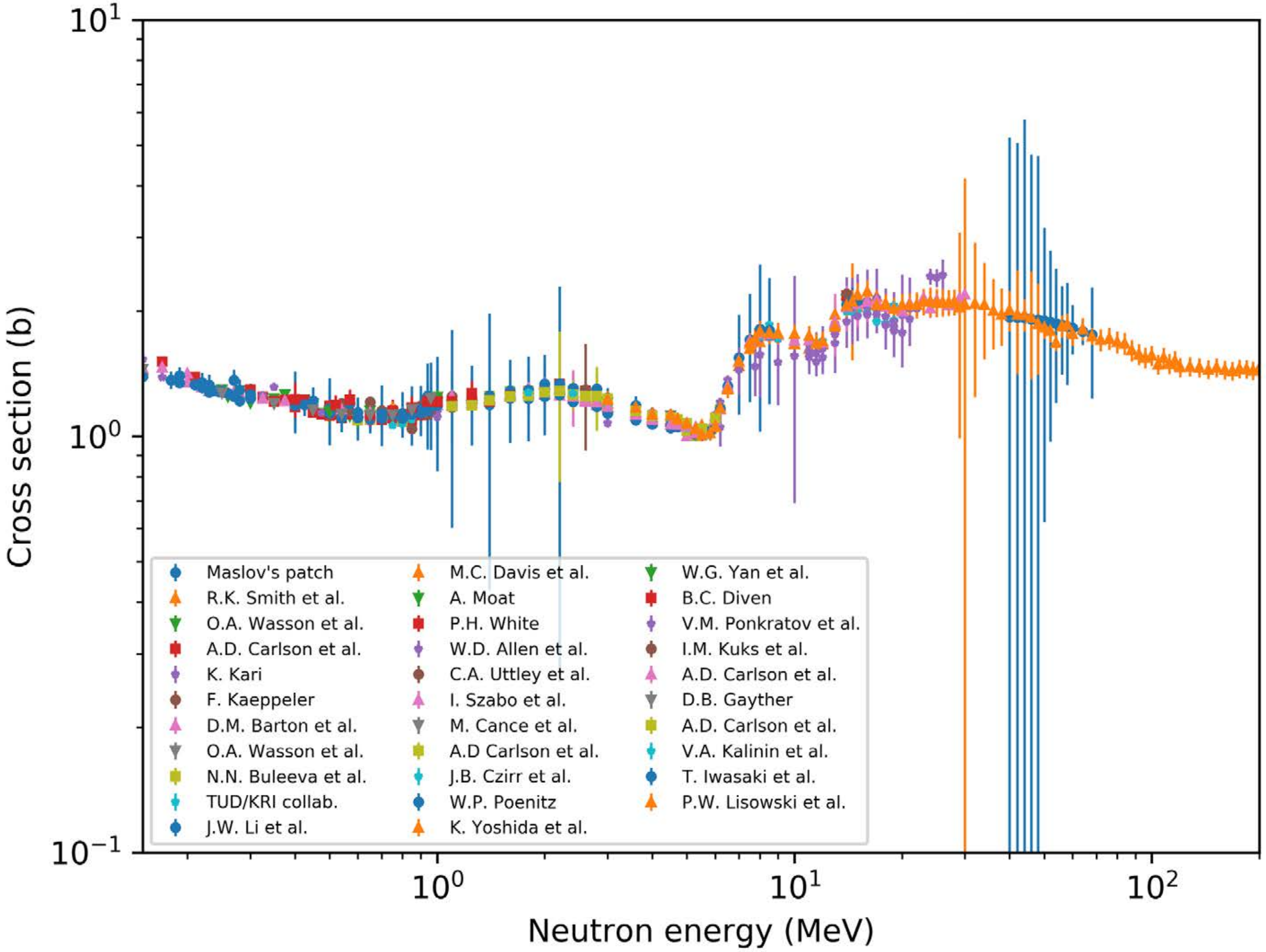
Log-evidence

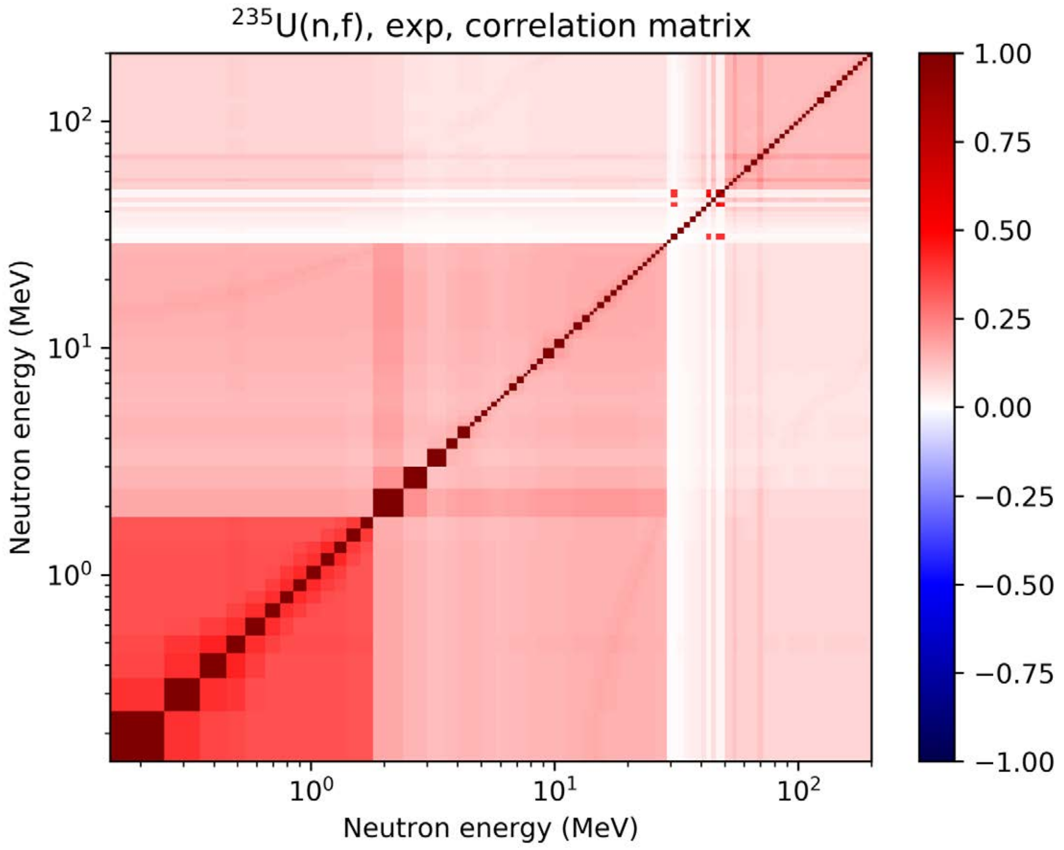
Experimental datasets for $^{235}U(n,f)$ cross section and GLS results

Experimental datasets for $^{235}U(n,f)$ cross section

Given by Georg SCHNABEL
- u5_absexpcov.csv
- u5_exp_datafram.csv

No. of data points: 657; Energy range: 150 keV – 200 MeV

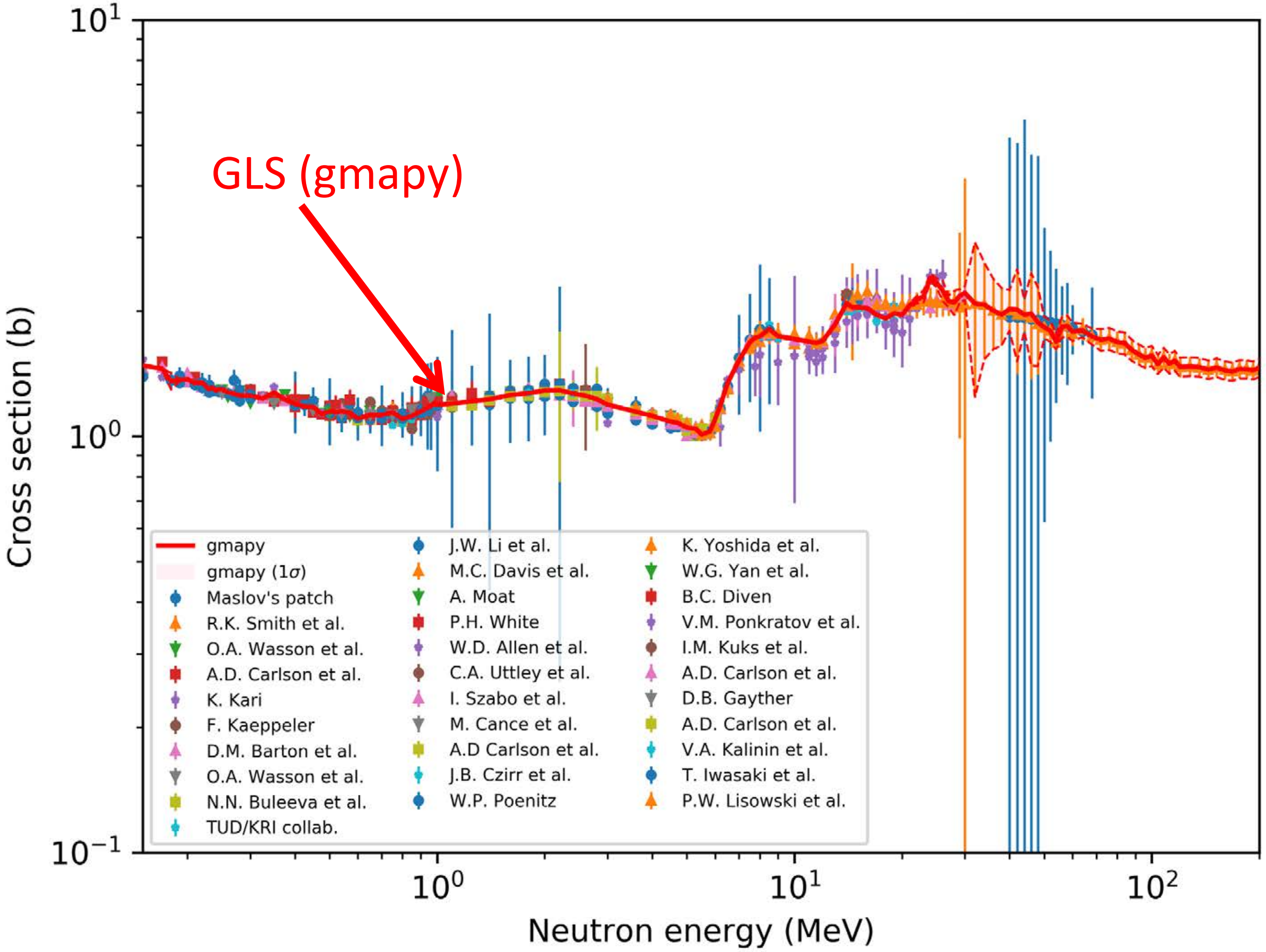


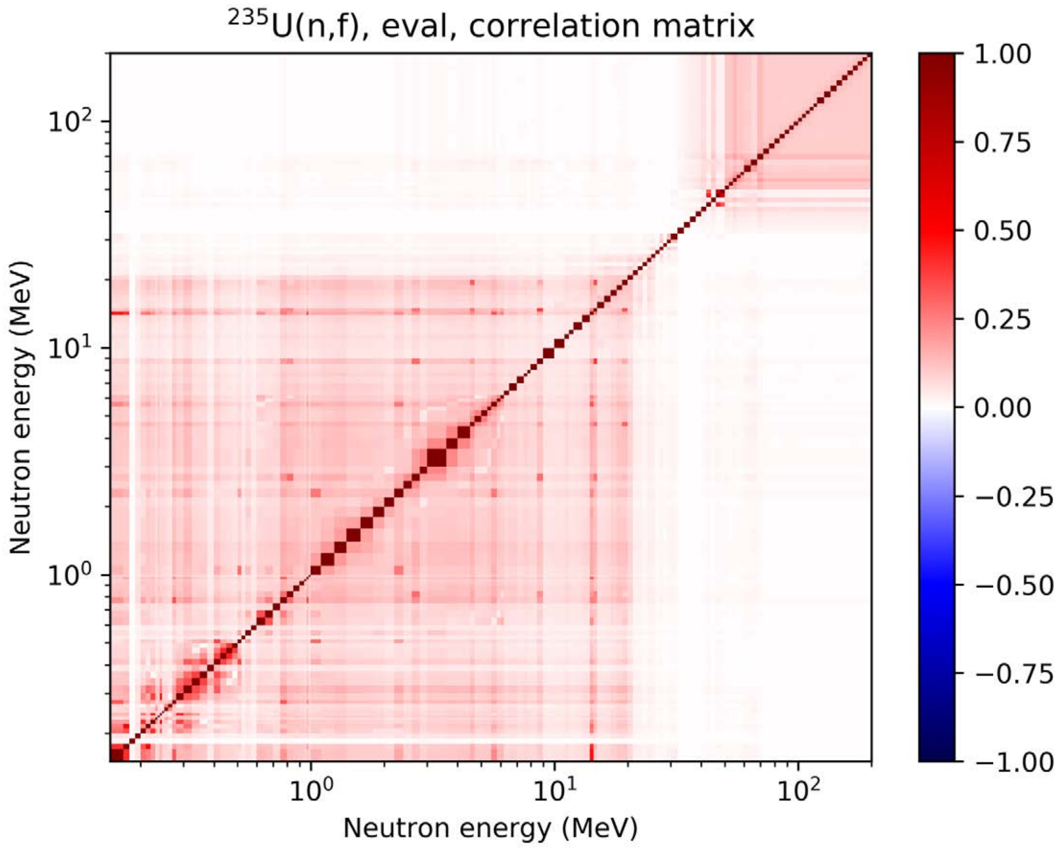


Experimental correlation matrix

GLS results

Given by Georg SCHNABEL
- u5_evalcovmat.csv
- u5_eval_datafram.csv





Evaluated correlation matrix

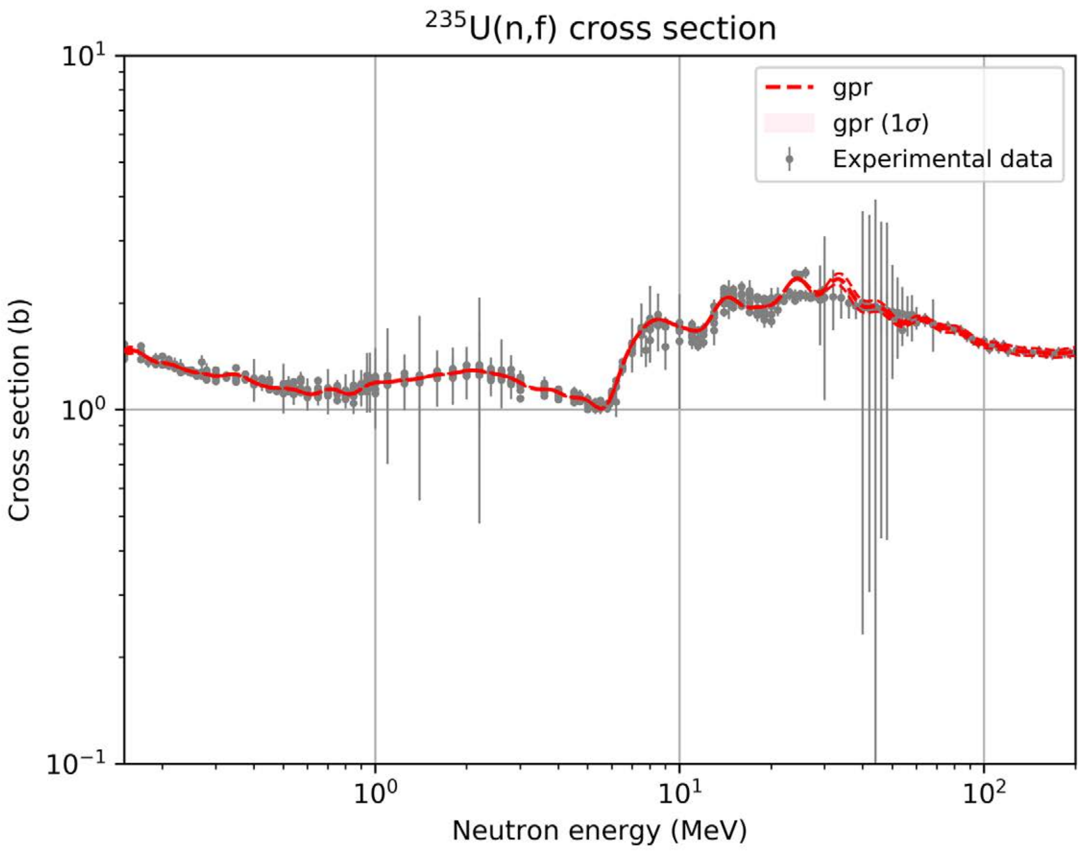
gmapy: https://github.com/IAEA-NDS/gmapy

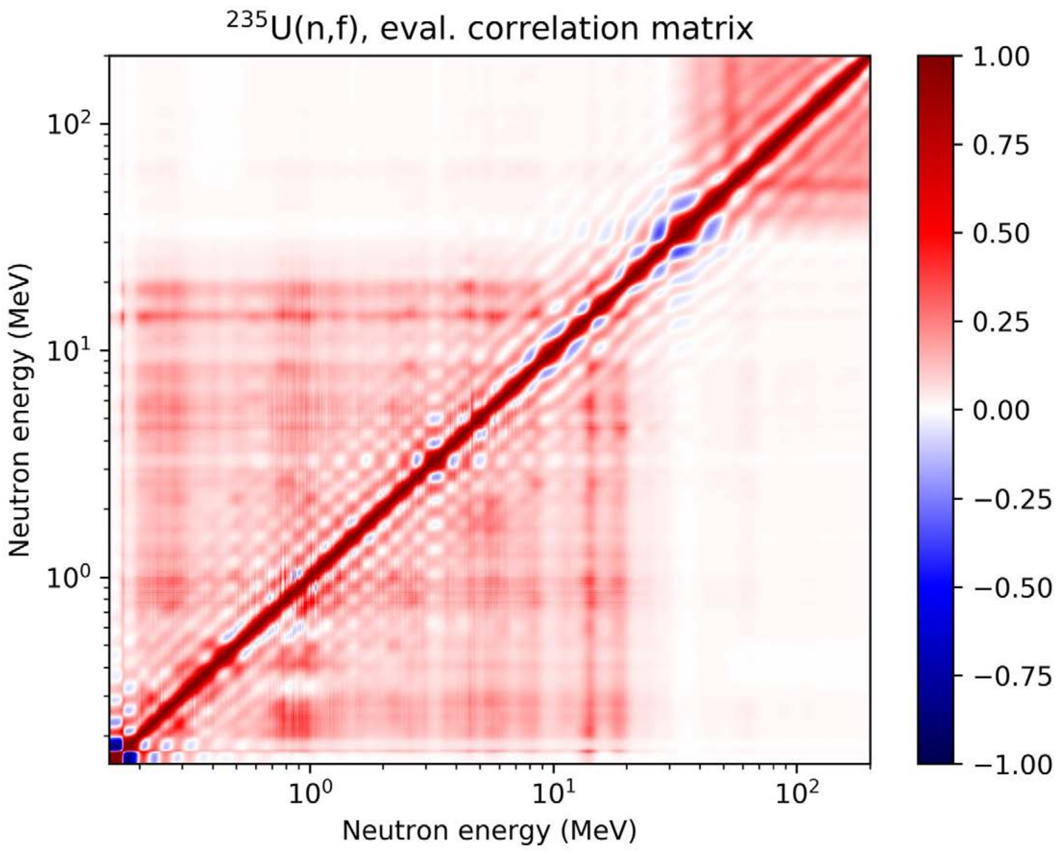
GPR results for RBF kernel

RBF kernel

- $\ell = 0.10$ (Length-scale parameter)

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

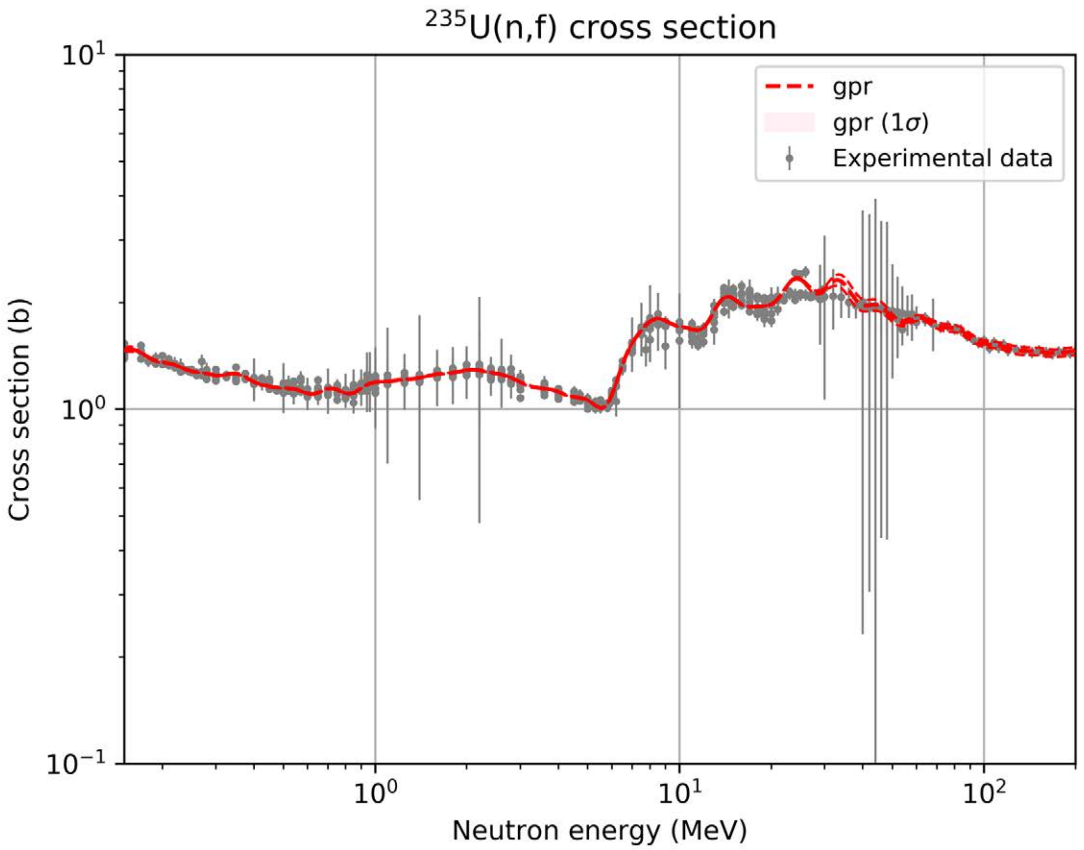


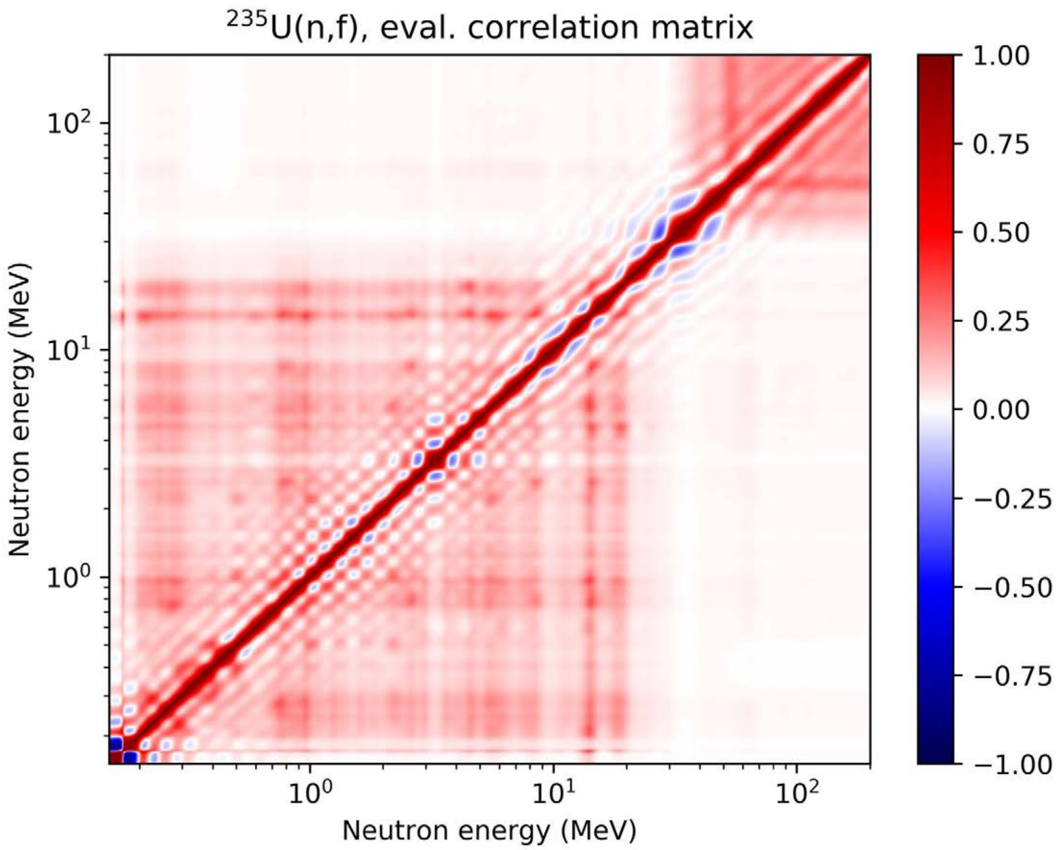


RBF kernel

- $\ell = 0.09$

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

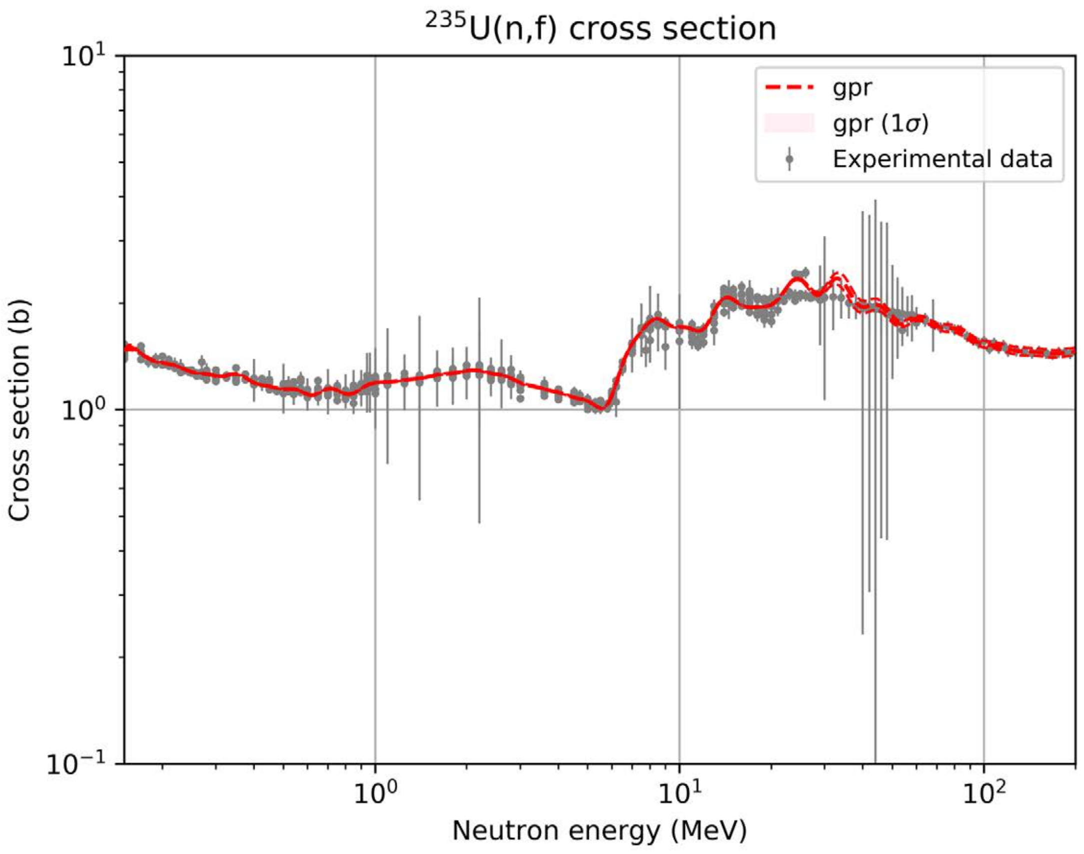


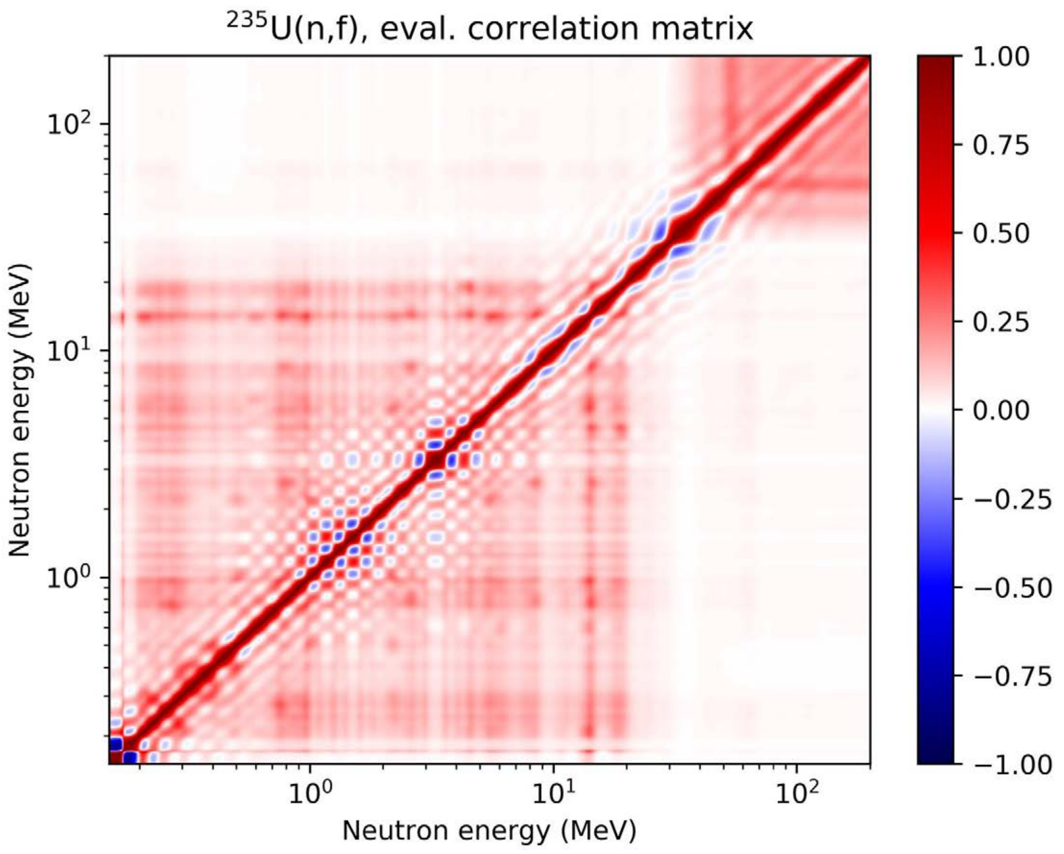


RBF kernel

- $\ell = 0.08$

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

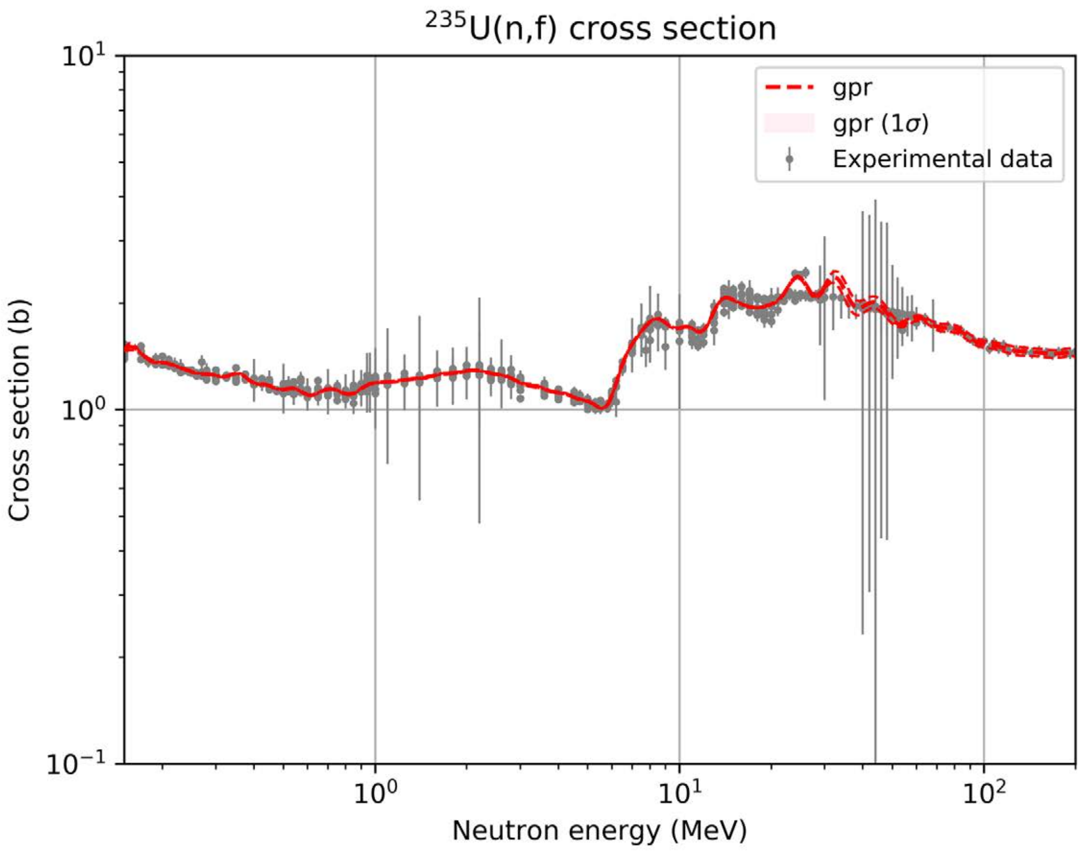


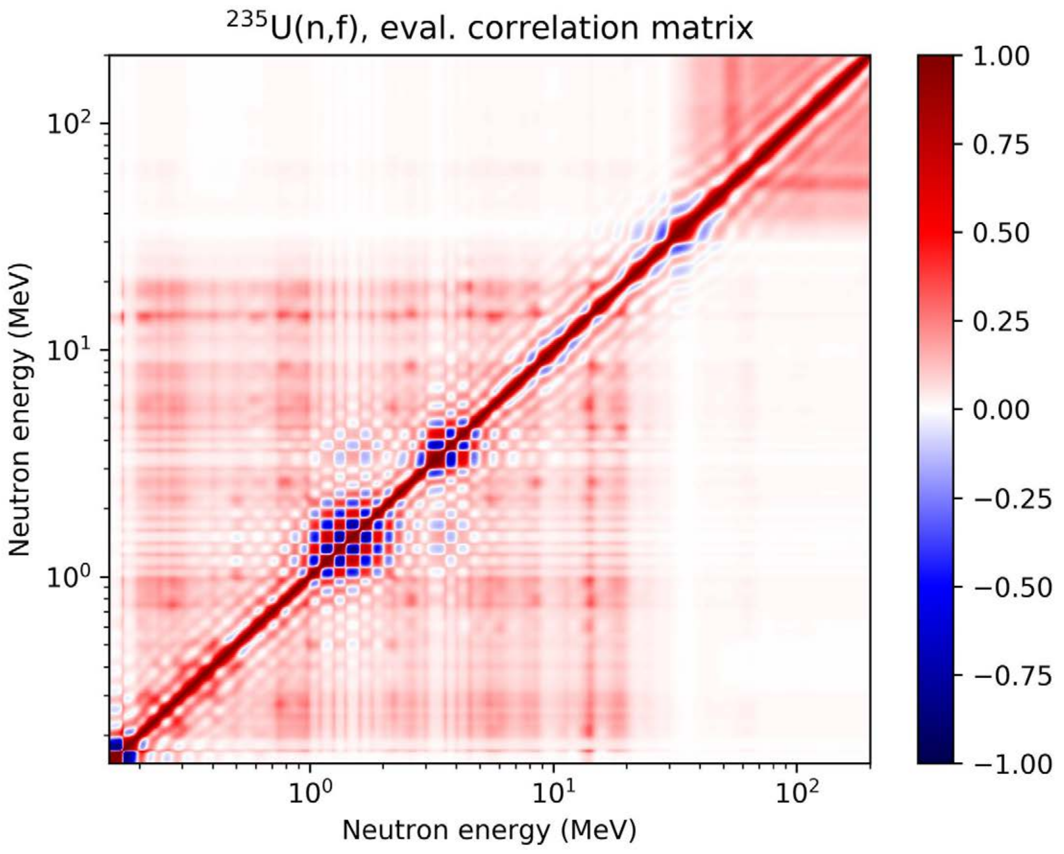


RBF kernel

- $\ell = 0.07$

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$



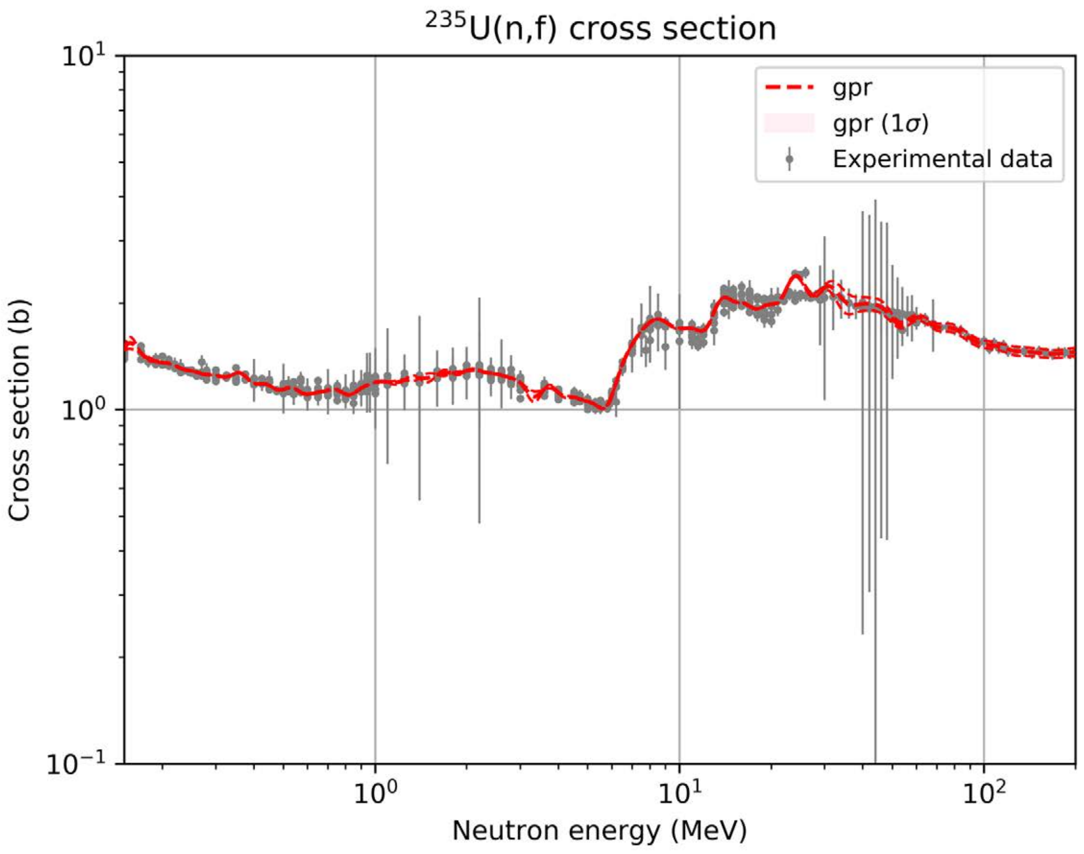


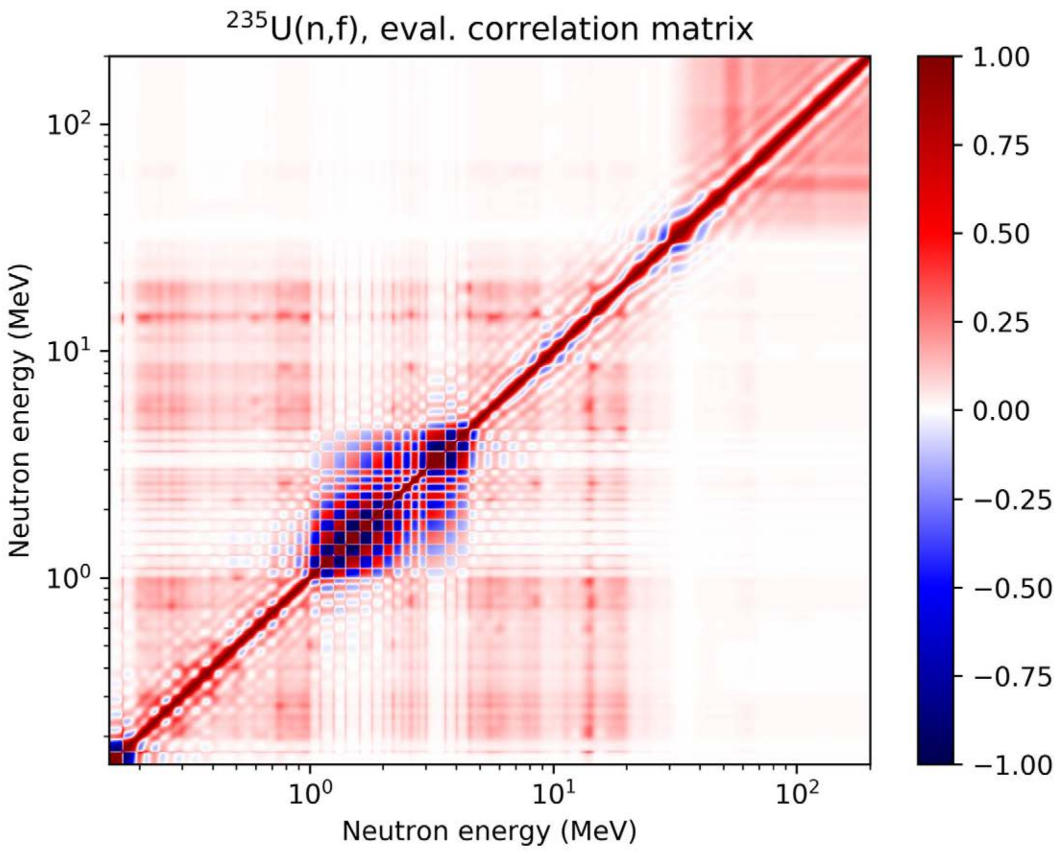
未来へげんき
To the Future / JAEA

RBF kernel

- $\ell = 0.06$

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$

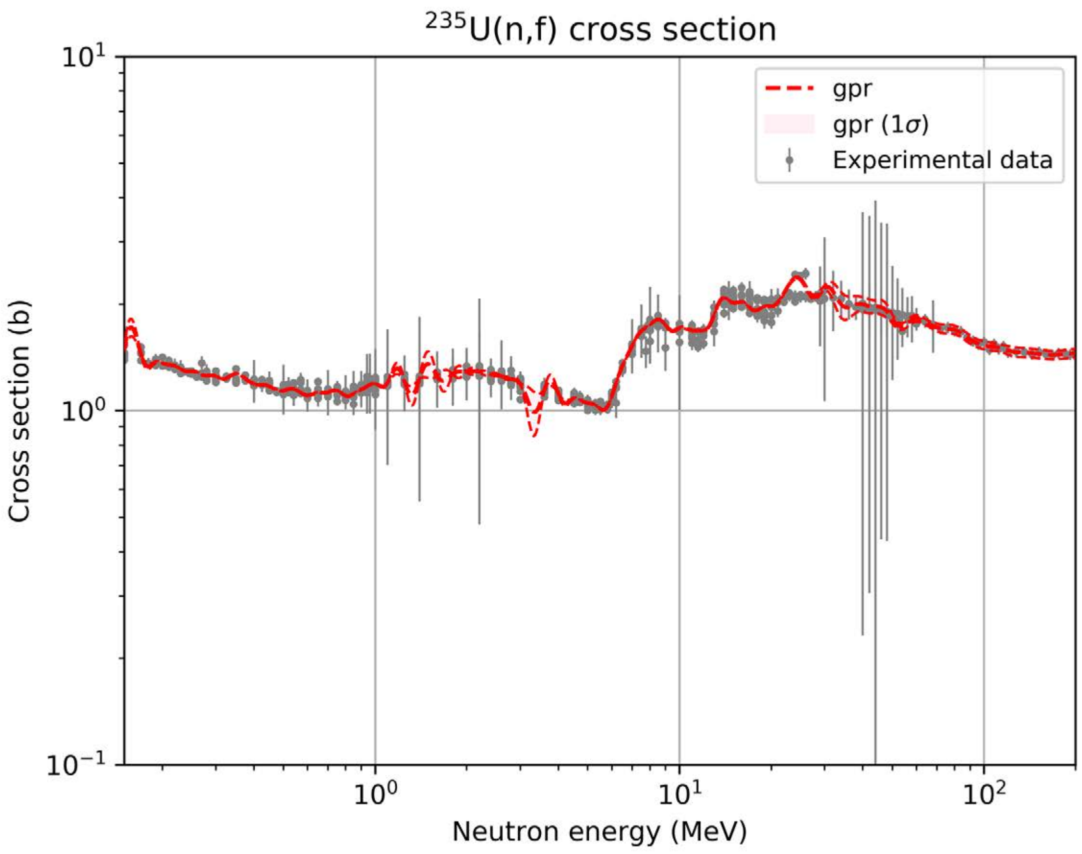


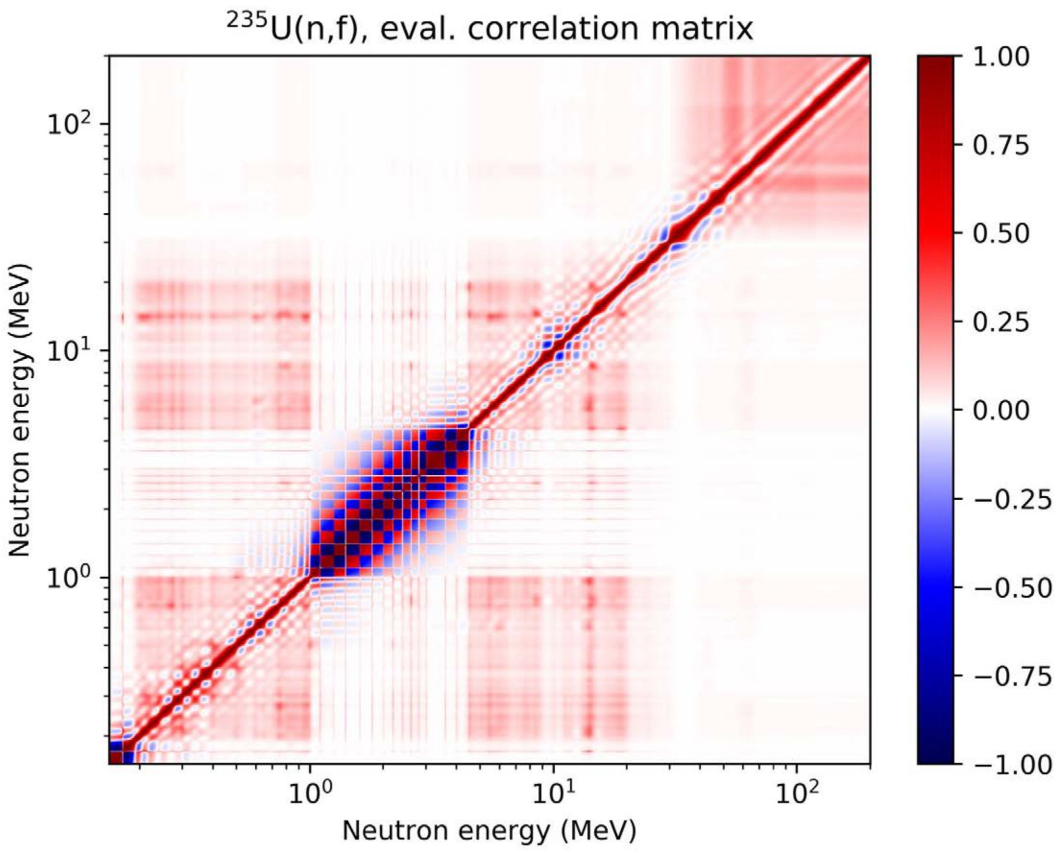


RBF kernel

- $\ell = 0.05$

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$



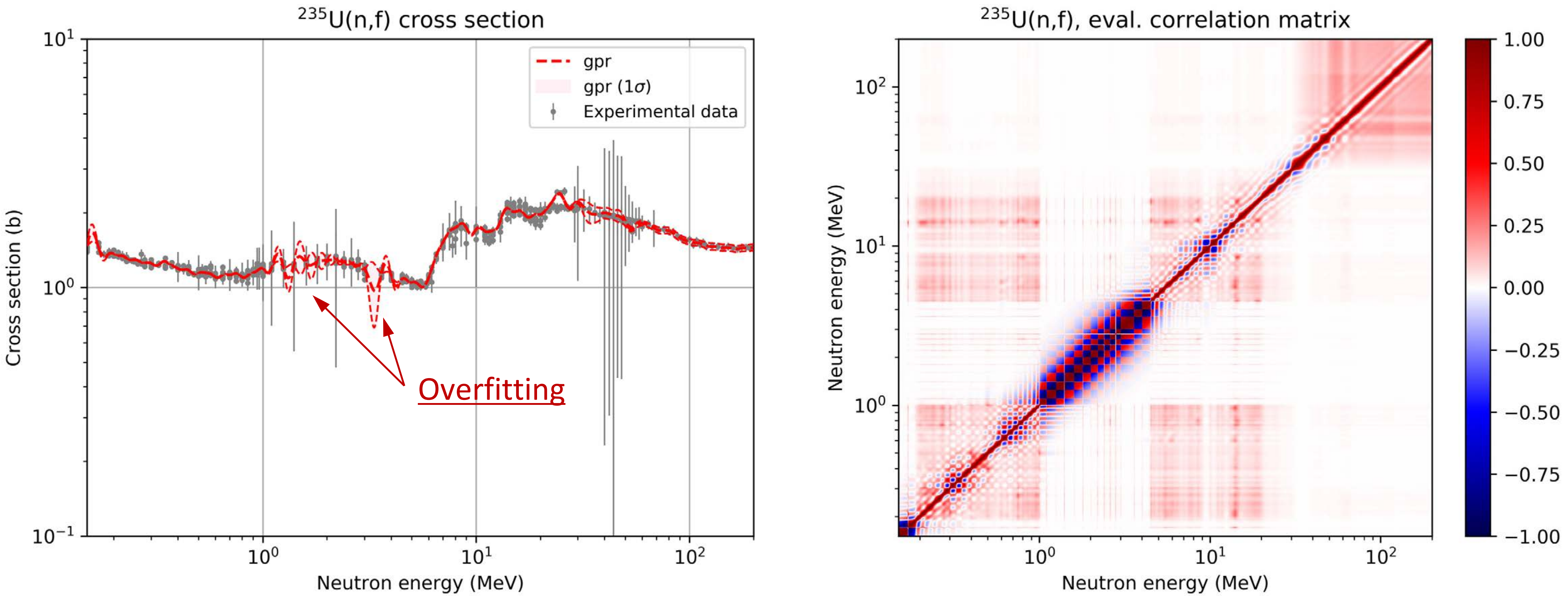


未来へげんき
To the Future / JAEA

RBF kernel

- $\ell = 0.043$ (mathematically optimal)

$$\mathcal{K}_{\mathrm{RBF}} = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2\ell^2}\right)$$



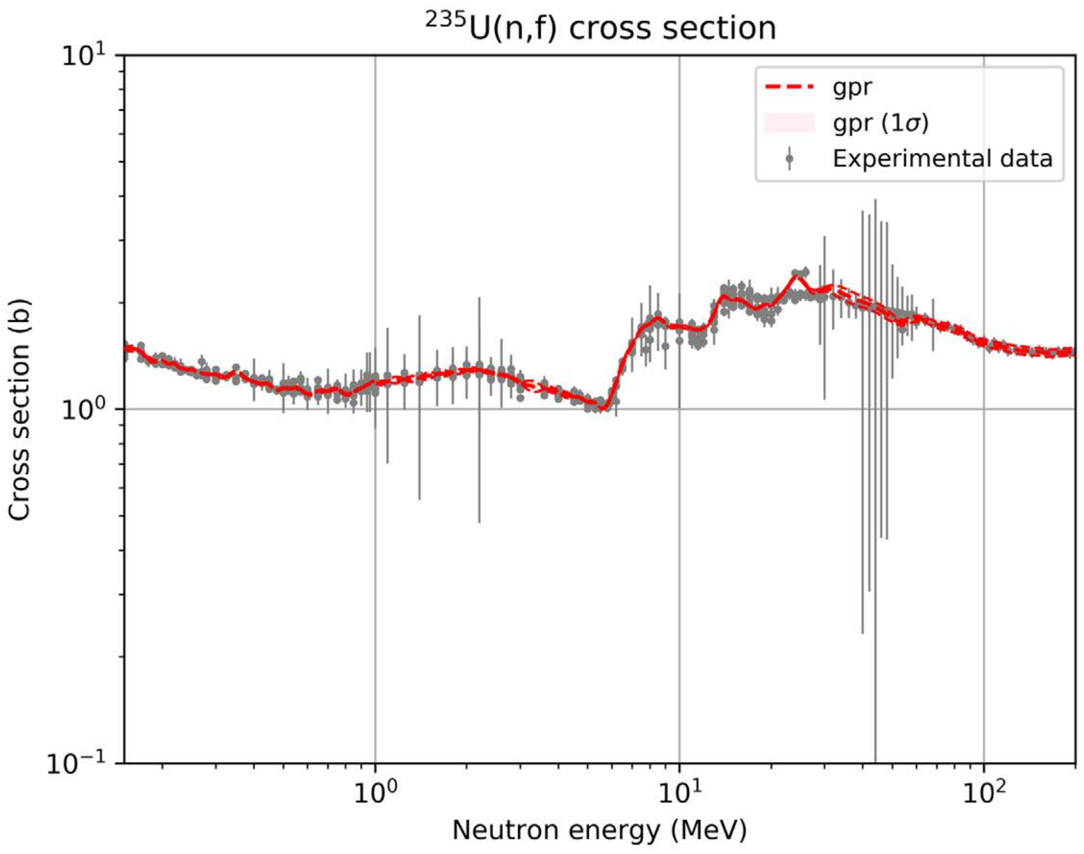
- Results with the RBF kernel are sensitive to hyperparameters.
- Overfitting was observed at the optimal point

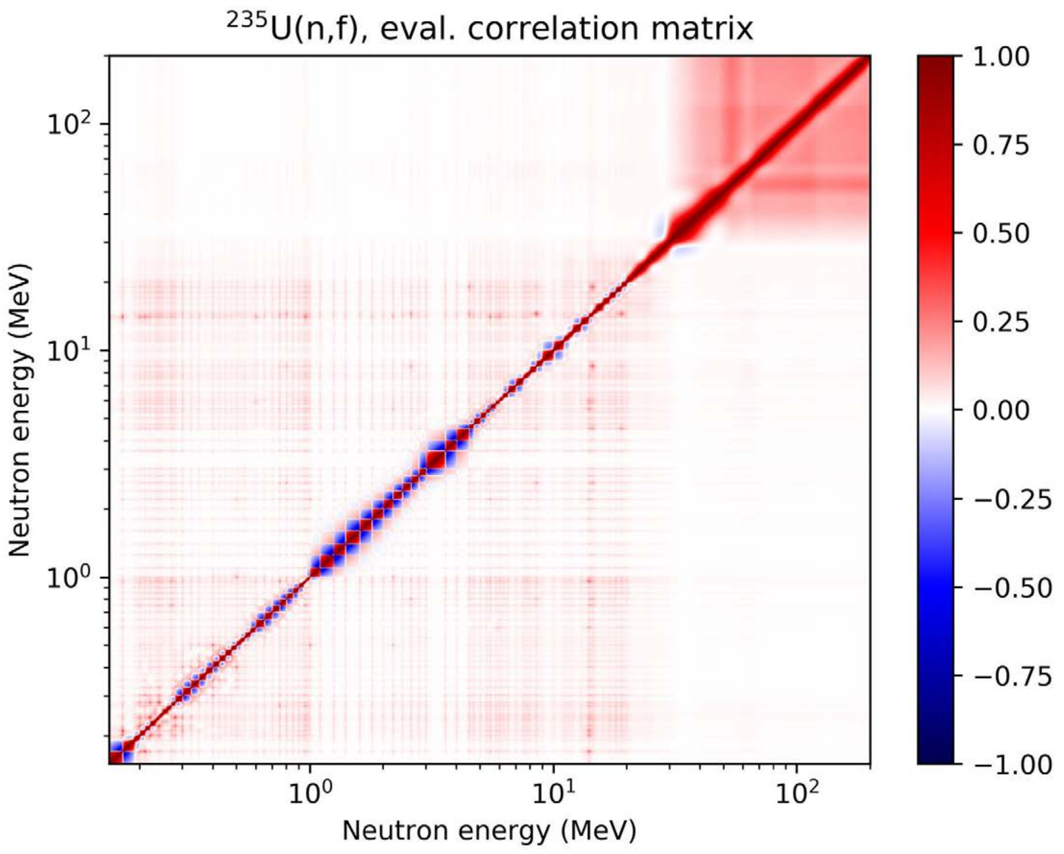
GPR results for Matérn 3 kernel

Matérn 3 kernel

- $\ell = 1.0$

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$

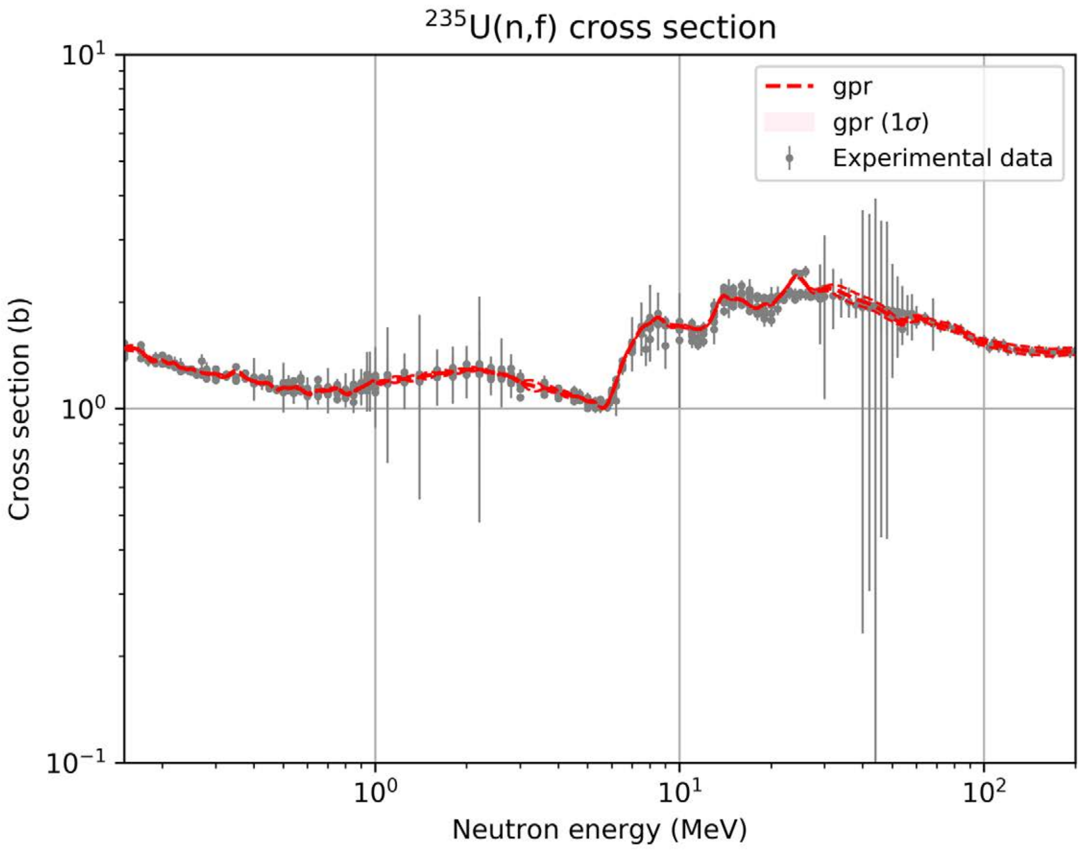


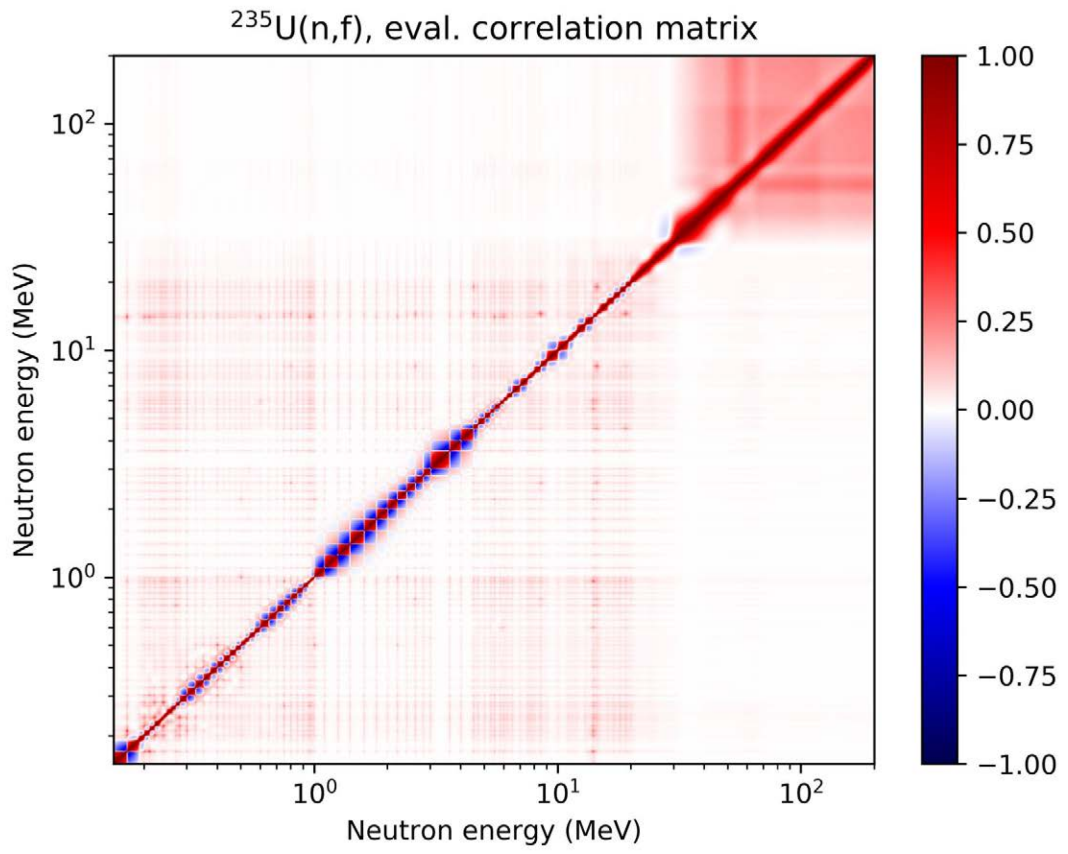


Matérn 3 kernel

- $\ell = 0.9$

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$

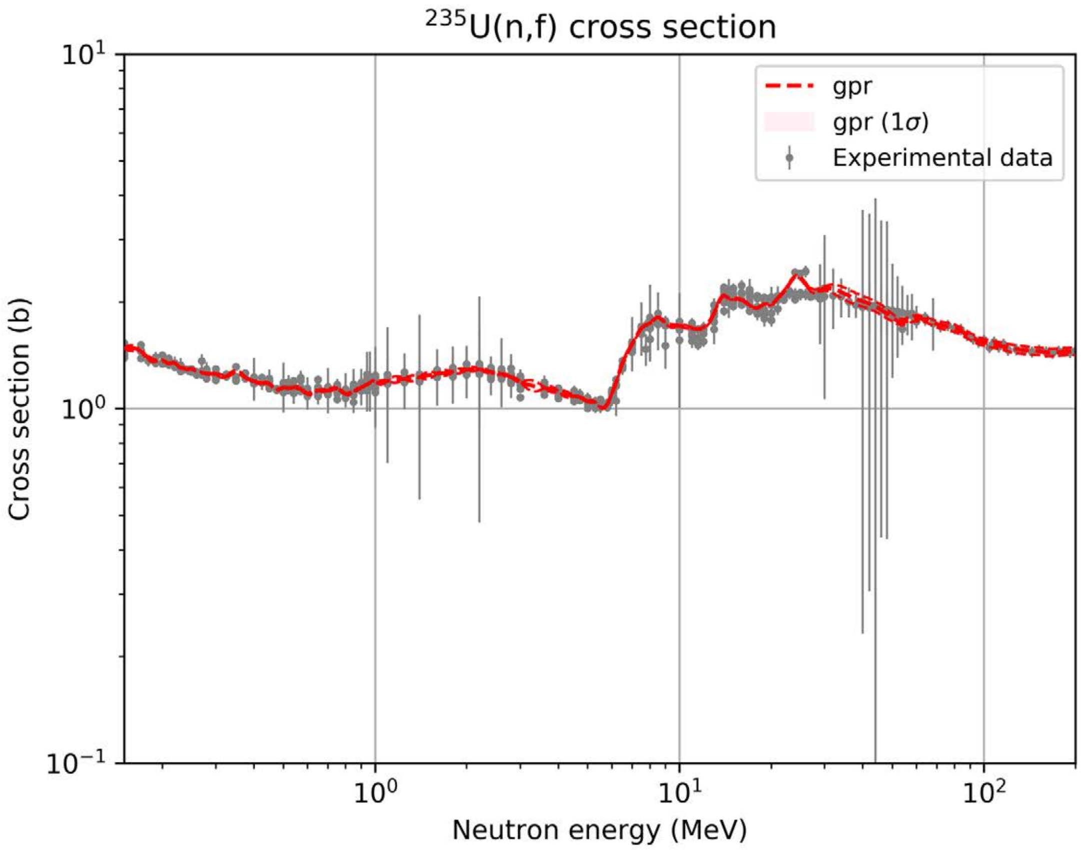


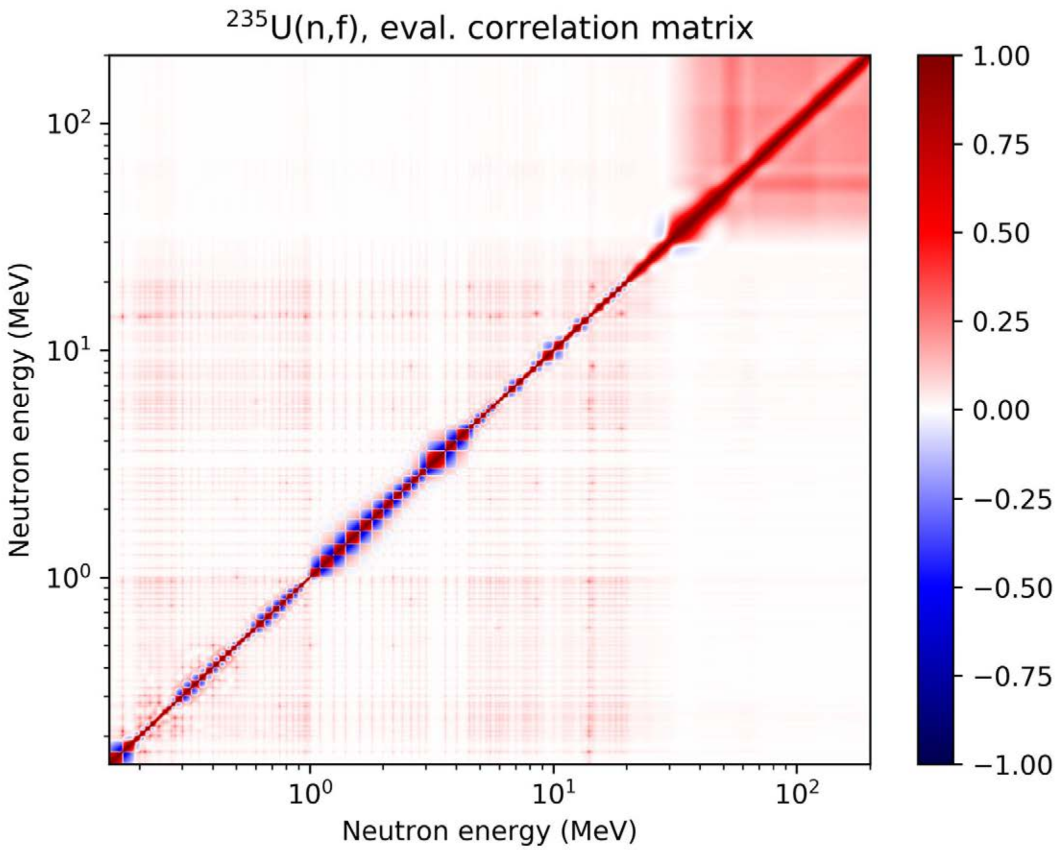


Matérn 3 kernel

- $\ell = 0.8$

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$



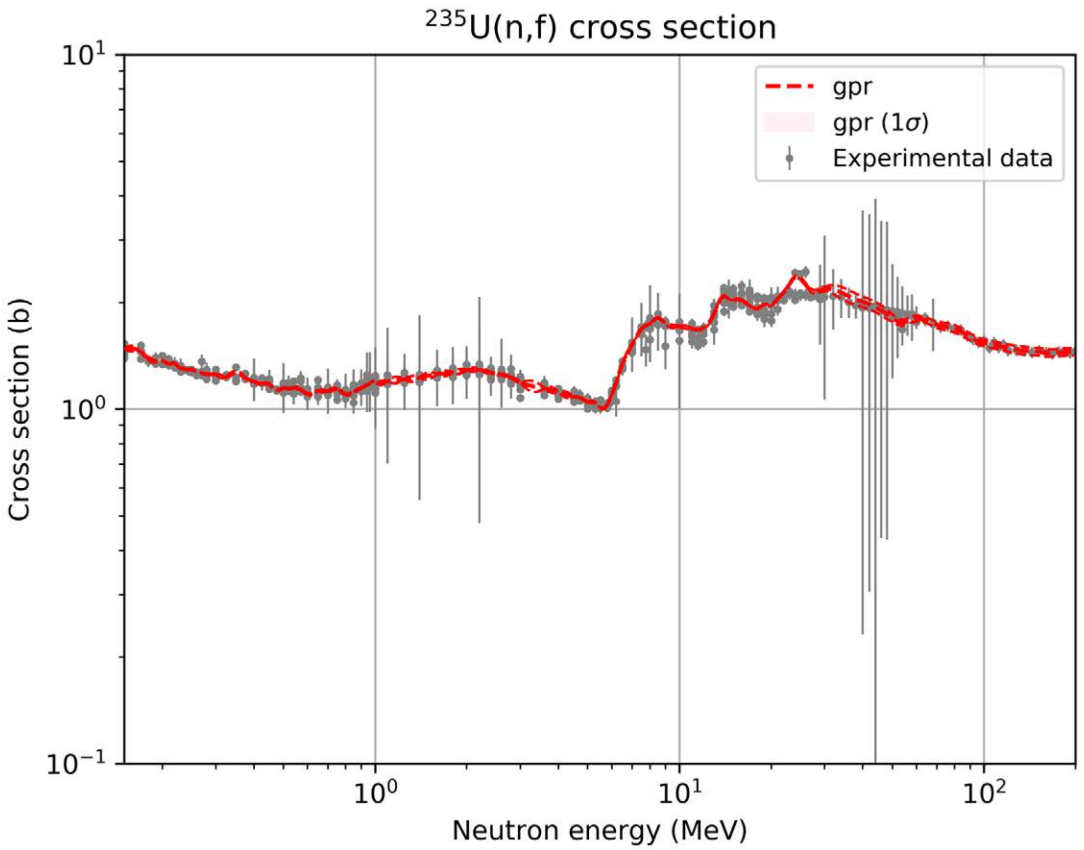


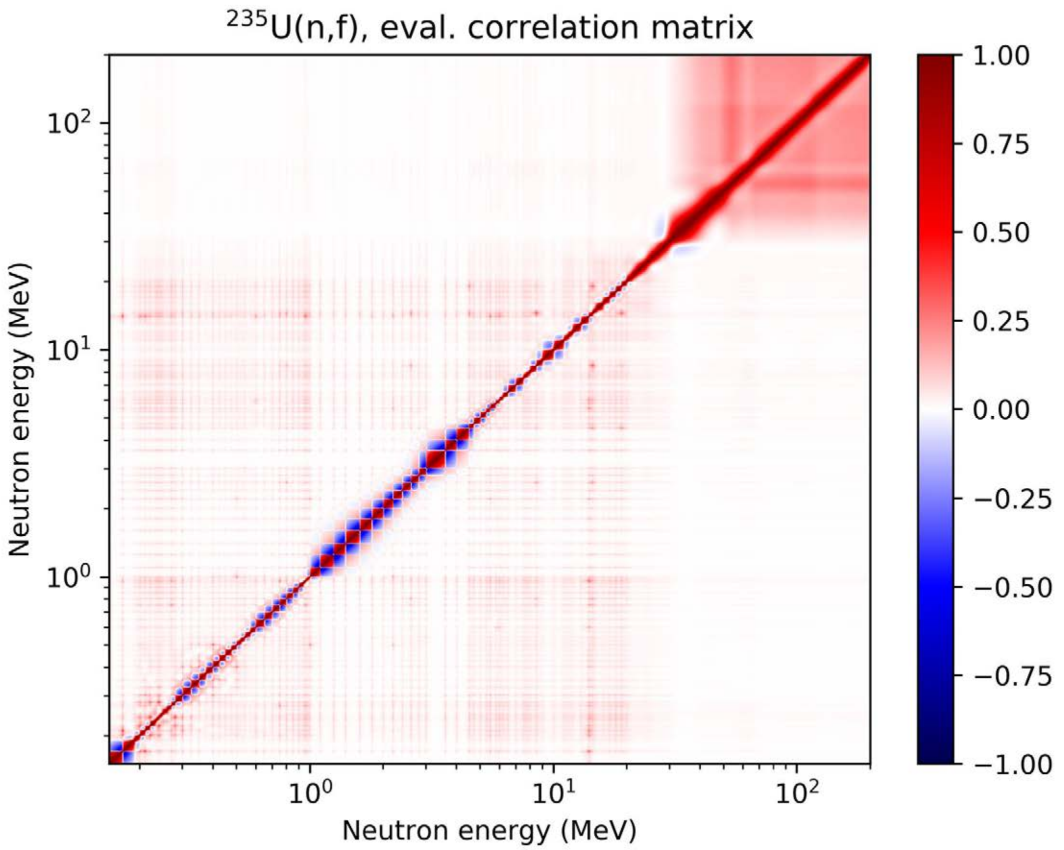
Matérn 3 kernel

未来へげんき To the Future / JAEA

- $\ell = 0.7$

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$

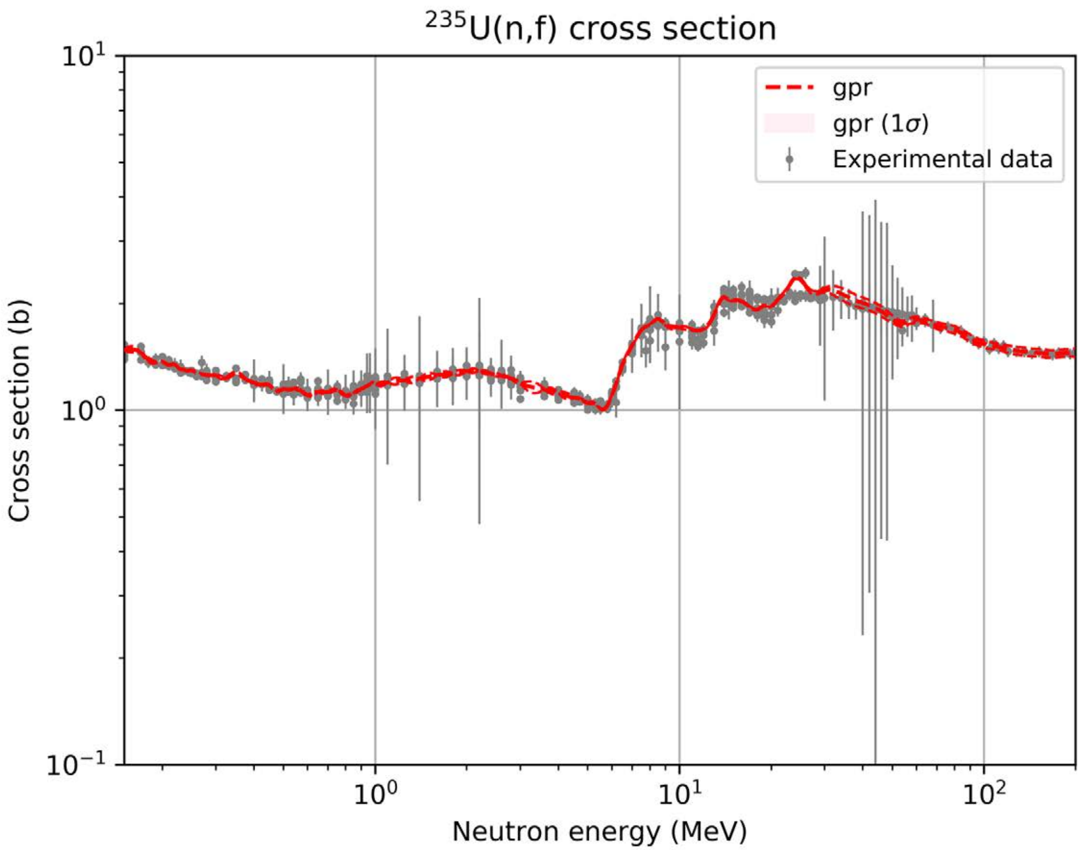


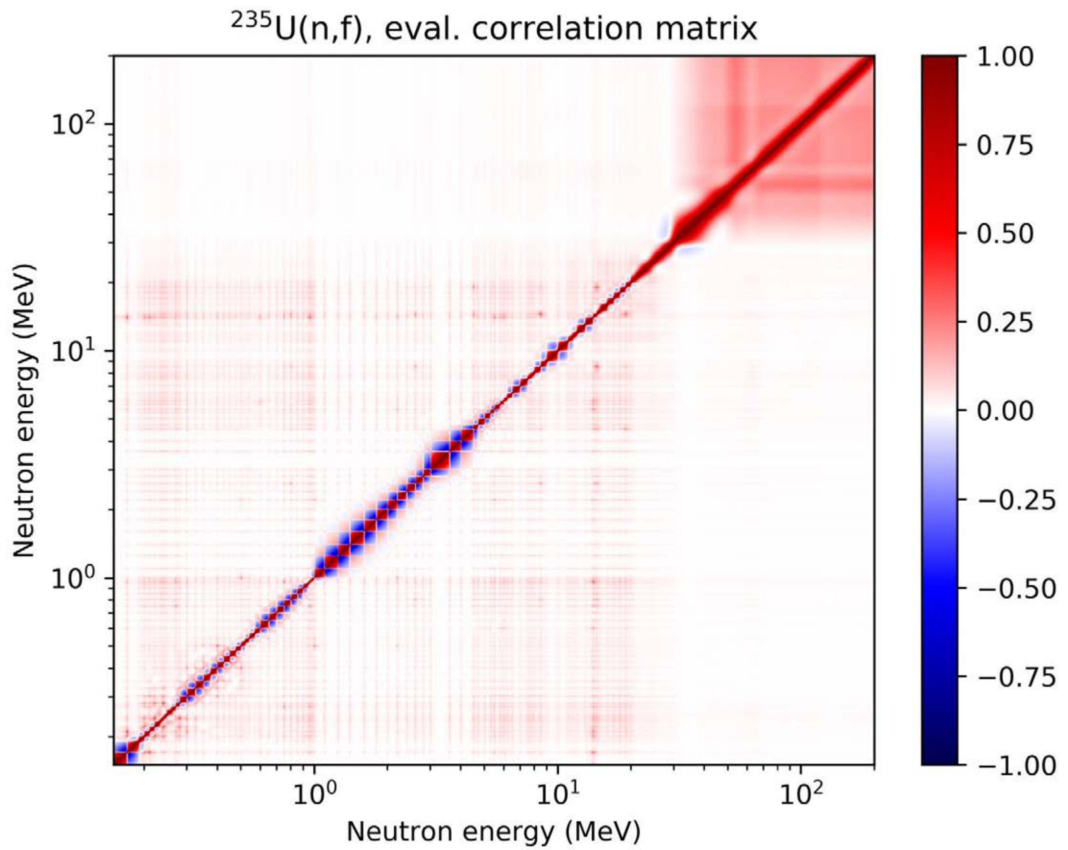


Matérn 3 kernel

- $\ell = 0.41$ (mathematically optimal)

$$\mathcal{K}_{\mathrm{Matern},3/2} = \sigma^2 \left(1 + \frac{\sqrt{3}}{\ell}|x_i - x_j|\right) \exp\left(-\frac{\sqrt{3}}{\ell}|x_i - x_j|\right)$$



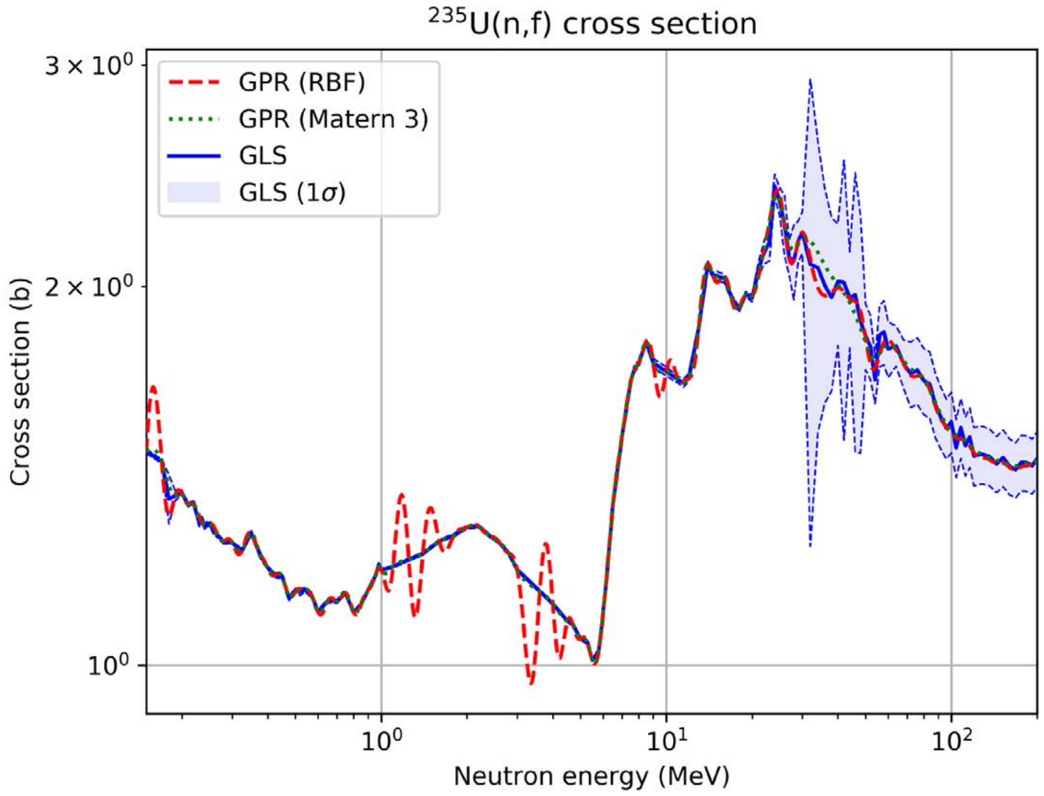


- Results with the Matérn 3 kernel are less sensitive to hyperparameters.

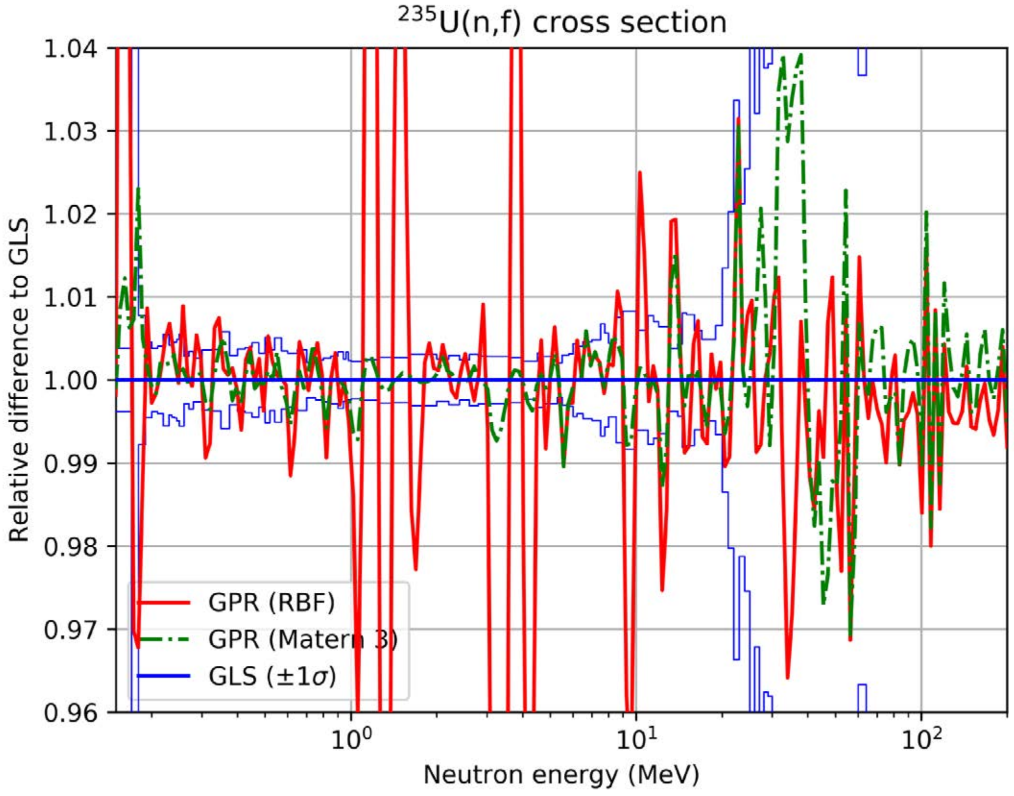
Comparison with GLS results

Comparison with GLS results

- Cross section



- Relative difference

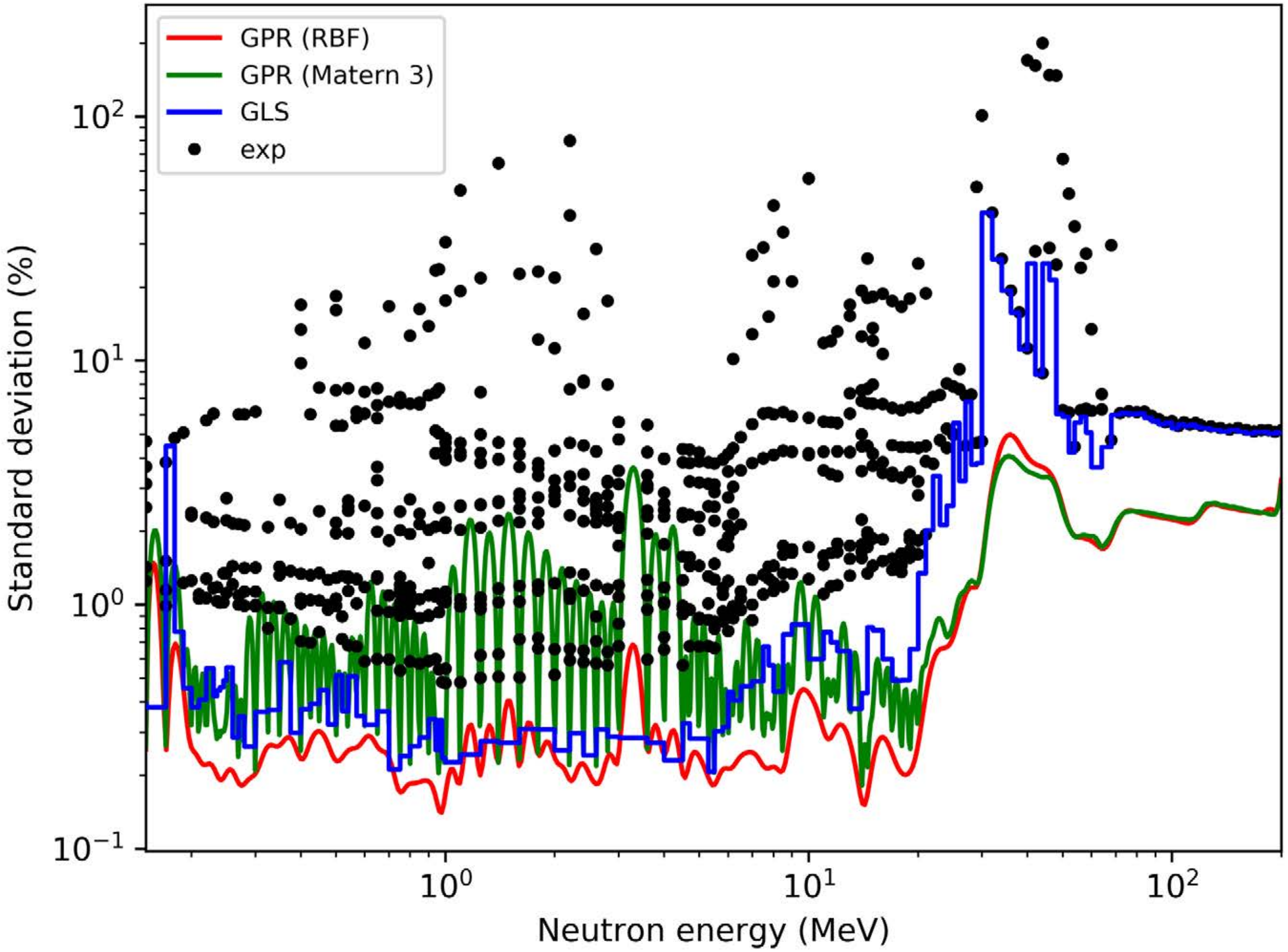


- GPR with the Matérn 3 kernel gives more consistent results with GLS.
  Differences are almost within GLS 1σ uncertainty.

Comparison with GLS results

To the Future / JAEA

● Standard deviation (diagonal components of covariance matrix)



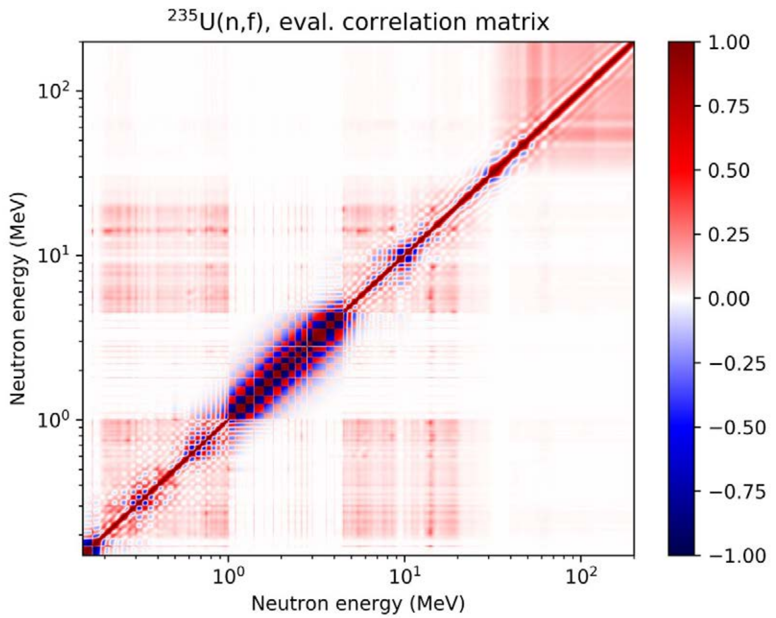
- GPR tends to give large standard deviations between experimental data points.
- Above 20 MeV, GPR gives smaller standard deviations than GLS.
- Instead, correlations are larger than GLS (next slide ☞)

未来へげんき
To the Future / JAEA

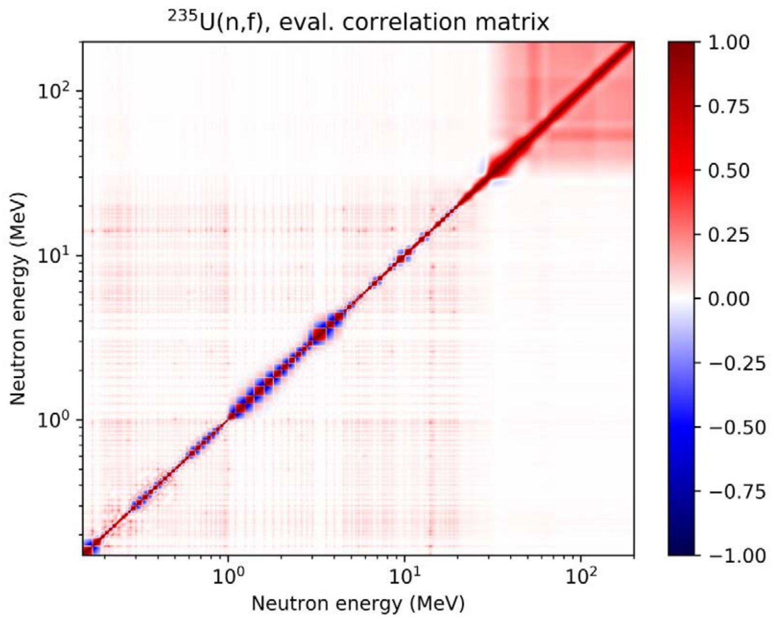
Comparison with GLS results

● Correlation matrix

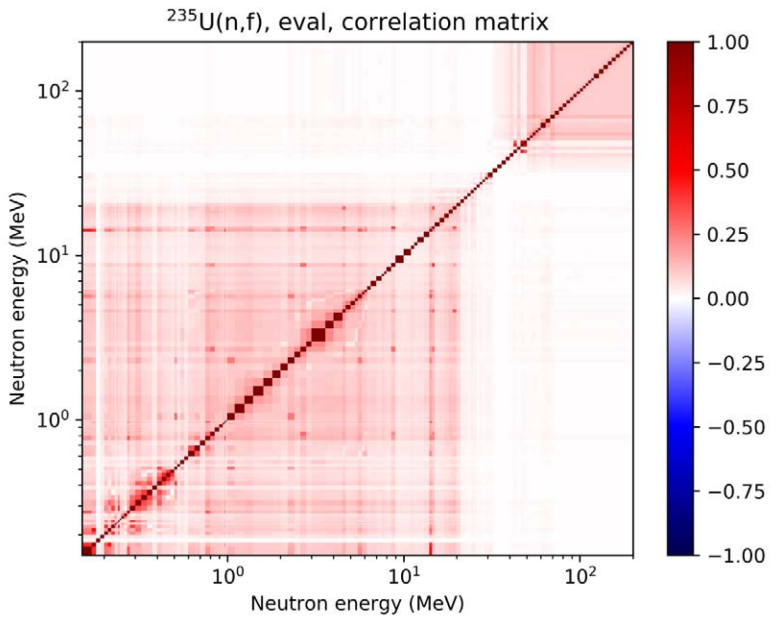
✓ GPR (RBF kernel)



✓ GPR (Matérn 3 kernel)



✓ GLS



- Broadly speaking, GPR results are similar to GLS results, but the strength of correlation differs from each other.
- Positive and negative correlations can be seen in GPR near the diagonal components.

Summary and suggestion

Summary

- GPR is a flexible tool to evaluate nuclear data, whose results could vary depending on the kernel function (and hyperparameters) adopted.

  ➡ For $^{235}U(n,f)$ case, the Matérn 3 kernel gave better results, but the covariance was somewhat different from the GLS results.

Suggestion

- GPR is a promising option for evaluating nuclear data.
- The covariance data evaluated by GPR may help covariance users, for example, in uncertainty quantification of reactor physics parameters.