Treating model defects with a Gaussian Process prior for the parameters

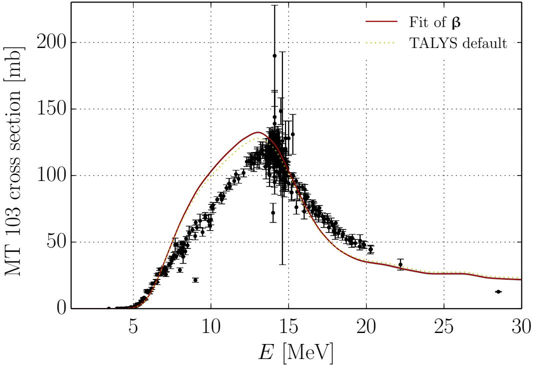
Petter Helgesson [1,2] Henrik Sjöstrand [1]

[1]Uppsala University, Uppsala, Sweden
[2]Nuclear Research and Consultancy Group NRG, Petten, The Netherlands

November 14, 2017

UPPSALA UNIVERSITET

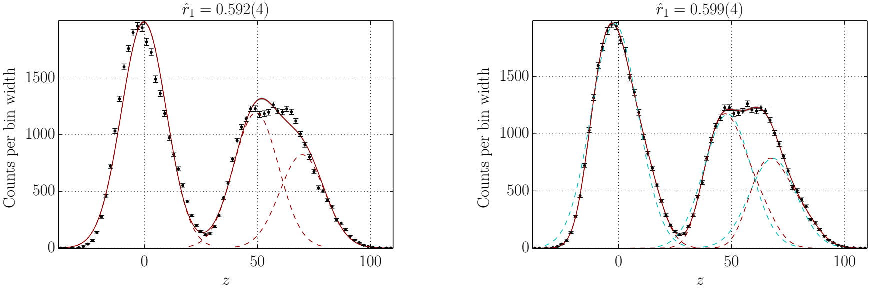
Model defects



- Our model (*e.g.*, TALYS) can not exactly reproduce underlying reality, whatever parameters we choose
- Can lead to biased fits and strongly underestimated uncertainties
  - Especially for evaluations in parameter space [1]

[1] P. Helgesson *et al.*, "Assessment of Novel Techniques for Nuclear Data Evaluation", Accepted for publication in the Proceedings of the 16th International Symposium on Reactor Dosimetry (2017)

Gaussian processes: “normal” usage



- Model in parameter space + defect in observable space:

$$Y_i = f(x_i; \boldsymbol{\beta}) + \mathcal{E} + \mathcal{E}_{\mathrm{m}}(x_i)$$

  - Y_i: ith data point (random variable)
  - $f(x_i; \boldsymbol{\beta})$: model with parameters $\boldsymbol{\beta} = (\beta_1, ..., \beta_l)^{\mathrm{T}}$
  - $\mathcal{E}$: measurement error (its distribution describes the uncertainty)
  - $\mathcal{E}_{\mathrm{m}}(x_i) \sim \mathcal{GP}(0, \kappa(x_i, x_j))$: error in model
- Used for ND evaluation by Schnabel/Leeb [2]

[2] G. Schnabel, “Large Scale Bayesian Nuclear Data Evaluation with Consistent Model Defects”, Ph.D. thesis, Technishe Universität Wien (2015)

UPPSALA UNIVERSITET

So, what *are* Gaussian processes?

- Generalization of Gaussian random vector:
  - *A collection of random variables, any finite number of which have a multivariate normal distribution* [3]
  - Can be evaluated at "any x"!
- Defined by mean function $\mu(x)$ and covariance function $\kappa(x, x')$:
  - $Q(x) \sim \mathcal{GP}(\mu(x), \kappa(x, x'))$

$\Rightarrow$ $\mathbf{Q}(\mathbf{x}) = (Q(x_1), Q(x_2), ..., Q(x_n))^{\mathrm{T}} \sim \mathbf{N}(\boldsymbol{\mu}(\mathbf{x}), \mathbf{K})$

- $\boldsymbol{\mu}(\mathbf{x}) = (\mu(x_1), \mu(x_2), ..., \mu(x_n))^{\mathrm{T}}$
- $(\mathbf{K})_{ij} = \kappa(x_i, x_j)$

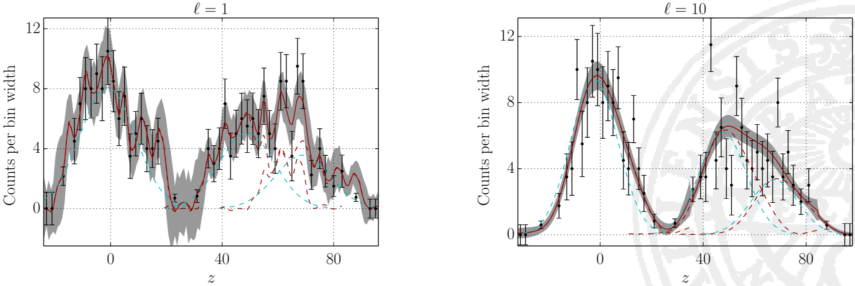
[3] C. Rasmussen and C. Williams, "Gaussian Processes for Machine Learning", (The MIT Press, Cambridge, MA, USA 2006), 2nd ed

The covariance function

- Common (and good) choice of cov. function:

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

  - σ^2: marginal variance
    - Magnitude of $\mathcal{E}_{\mathrm{m}}(x)$ on the order of σ
  - ℓ: correlation length
    - x_i, x_j close/far w.r.t. $\ell \Rightarrow$ strongly/weakly correlated
- Physical constraints $\Rightarrow \kappa(x_i, x_j)$ may need to be more complicated

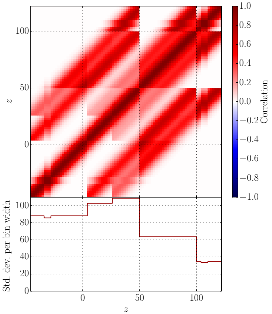


The covariance function

- Common (and good) choice of cov. function:

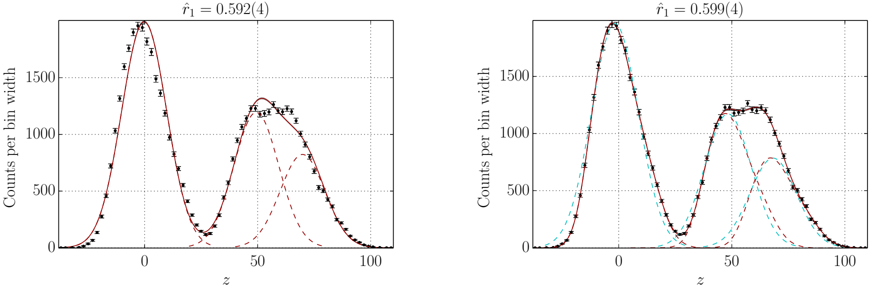
$$\kappa(x_i, x_j) = \sigma^2 e^{-\frac{(x_i - x_j)^2}{2\ell^2}}$$

  - σ^2: marginal variance
    - Magnitude of $\mathcal{E}_{\mathrm{m}}(x)$ on the order of σ
  - ℓ: correlation length
    - x_i, x_j close/far w.r.t. $\ell \Rightarrow$ strongly/weakly correlated
- Physical constraints $\Rightarrow \kappa(x_i, x_j)$ may need to be more complicated



← A semi-complicated $\kappa(x_i, x_j)$

Gaussian processes: “normal” usage (2)



- Simultaneous fit of model parameters $\boldsymbol{\beta}$ and defect $\mathcal{E}_{\mathrm{m}}(x)$
  - σ and ℓ can be determined from data using cross-validation
- Rigorous statistical study [4]: GP shows very good performance
- Possible downsides in ND evaluation:
  - Complexity of covariance functions (to conserve sum rules)
  - Parameters loose meaning (in an uncontrolled fashion)

[4] 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 **88** (2017)

This work: GP in parameter space

- Energy dependent model parameters around "global" parameters $\boldsymbol{\beta}$

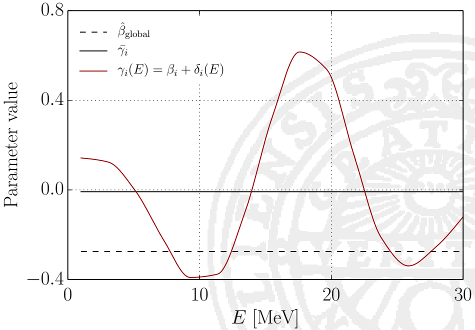
$$\boldsymbol{\gamma}(E) = \boldsymbol{\beta} + \boldsymbol{\delta}(E)$$

- GP prior for deviation from $\boldsymbol{\beta}$

$$\boldsymbol{\delta}(E) = \left(\delta_1\left(E\right), \delta_2\left(E\right), ..., \delta_l\left(E\right)\right)^{\mathrm{T}}$$

$$\delta_j(E) \sim \mathcal{GP}(0, \kappa_j(E, E')).$$

- Ensures consistency at each E
- Parameters can still be used for, *e.g.*, angular distributions
- Improves extrapolation to other nuclides?
- Easier to understand than complicated $\kappa(x, x')$?



This work: GP in parameter space (2)

- For chosen grid $\mathbf{E} = (E_1, E_2, ..., E_{n_E})^{\mathrm{T}}$:

$$\boldsymbol{\gamma}_{\mathbf{E}} = \begin{pmatrix} \beta_1 + \delta_1(E_1) \\ \beta_1 + \delta_1(E_2) \\ \vdots \\ \beta_1 + \delta_1(E_{n_E}) \\ \beta_2 + \delta_2(E_1) \\ \vdots \\ \beta_l + \delta_l(E_{n_E}) \end{pmatrix} = \mathbf{T}_{\mathbf{E}}\boldsymbol{\beta} + \boldsymbol{\delta}_{\mathbf{E}},$$

Simultaneous fit of global and local parameters

- First, a global fit $\hat{\beta}$ is obtained with Levenberg-Marquardt
- As of now, only linearization around $\hat{\beta}$ is used for local fit:

$$\hat{\gamma}_{\mathbf{E}} = \mathbf{T}_{\mathbf{E}}\hat{\beta} + \left(\mathbf{J}_{\gamma_{\mathbf{E}}}^{\mathrm{T}}\mathbf{\Omega}^{-1}\mathbf{J}_{\gamma_{\mathbf{E}}} + \mathbf{P}^{-1}\right)^{-1}\left[\mathbf{J}_{\gamma_{\mathbf{E}}}^{\mathrm{T}}\mathbf{\Omega}^{-1}\left(\mathbf{y} - \hat{\mathbf{f}}\right) + \mathbf{P}^{-1}\mathbf{T}_{\mathbf{E}}\left(\mathbf{p} - \hat{\beta}\right)\right]$$

  - $\mathbf{T}_{\mathbf{E}}\hat{\beta}$: starting guess *only*
  - β also updated in local fit (included in $\gamma_{\mathbf{E}}$)
  - Prior covariance $\mathbf{P}$: combination of GP and prior for β:

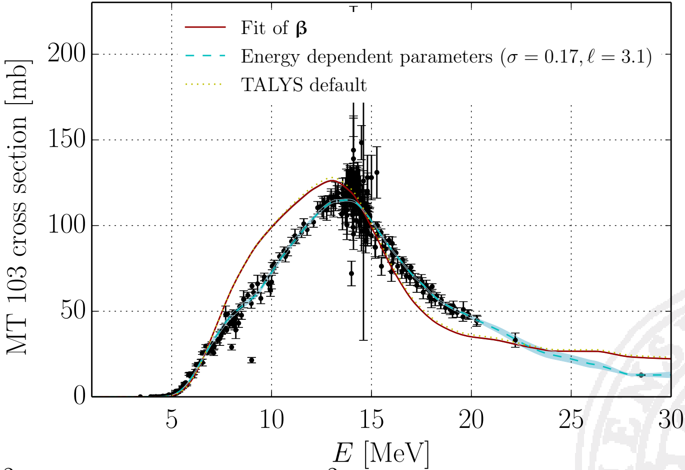
$$\mathbf{P} = \mathbf{K} + \mathbf{T}_{\mathbf{E}}\mathbf{P}_0\mathbf{T}_{\mathbf{E}}^{\mathrm{T}}$$

    - GP cov. matrix $\mathbf{K}$ block diagonal
    - $\mathbf{P}_0$ (prior for β) diagonal $\Rightarrow$ $\mathbf{P}$ block diagonal with easily inverted blocks

- σ and ℓ determined by cross-validation

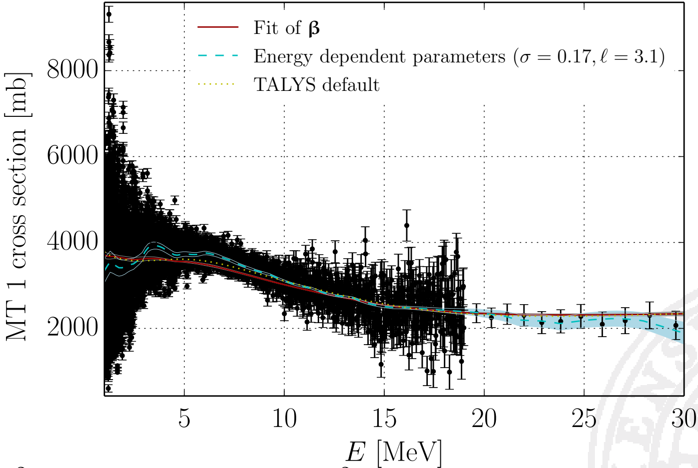
UPPSALA UNIVERSITET

Applied to ^{56}Fe cross sections with "Pseudo-TALYS"



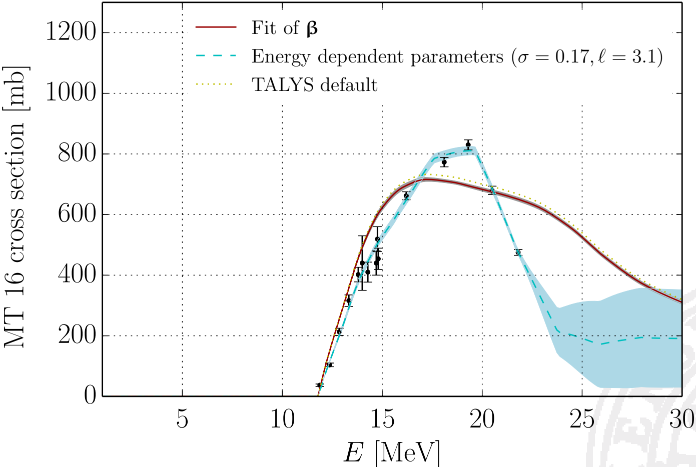
- $\chi^2_{\mathrm{CV}}/\nu = 1.22$ (global fit: $\chi^2/\nu = 8.9$)
- Some problems; related to too strong (blind) belief in exp. data?

Applied to ^{56}Fe cross sections with "Pseudo-TALYS"



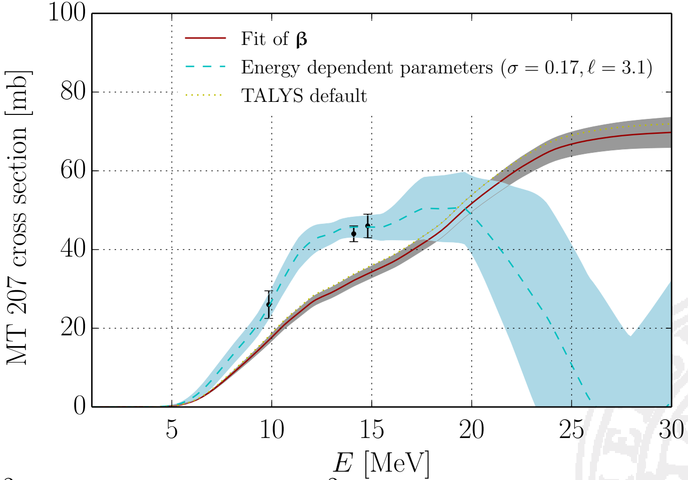
- $\chi^2_{\mathrm{CV}}/\nu = 1.22$ (global fit: $\chi^2/\nu = 8.9$)
- Some problems; related to too strong (blind) belief in exp. data?

Applied to ^{56}Fe cross sections with "Pseudo-TALYS"



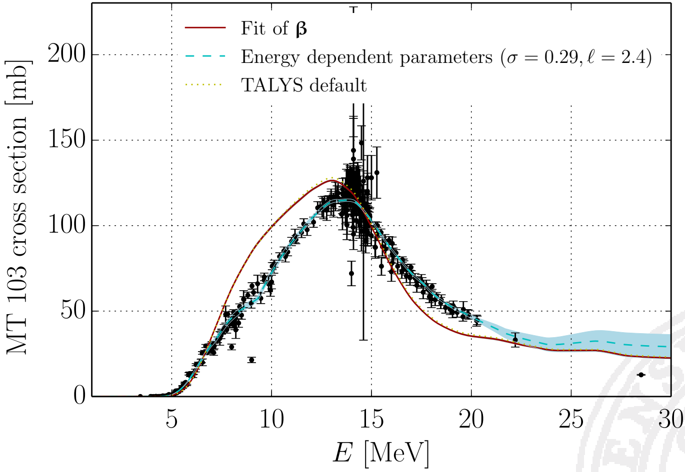
- $\chi^2_{\mathrm{CV}}/\nu = 1.22$ (global fit: $\chi^2/\nu = 8.9$)
- Some problems; related to too strong (blind) belief in exp. data?

Applied to ^{56}Fe cross sections with "Pseudo-TALYS"



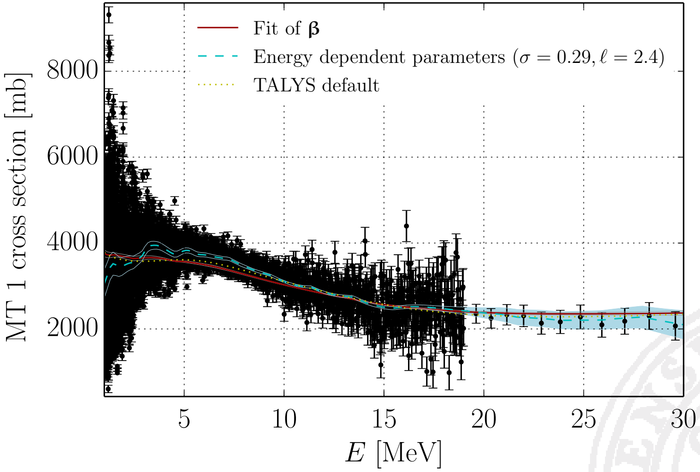
- $\chi^2_{\mathrm{CV}}/\nu = 1.22$ (global fit: $\chi^2/\nu = 8.9$)
- Some problems; related to too strong (blind) belief in exp. data?

^{56}Fe without one dubious point



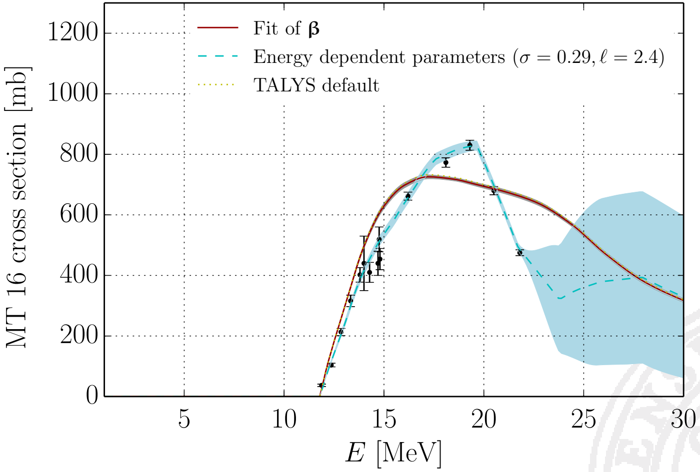
- $\chi^2_{\mathrm{CV}}/\nu = 1.20$

^{56}Fe without one dubious point



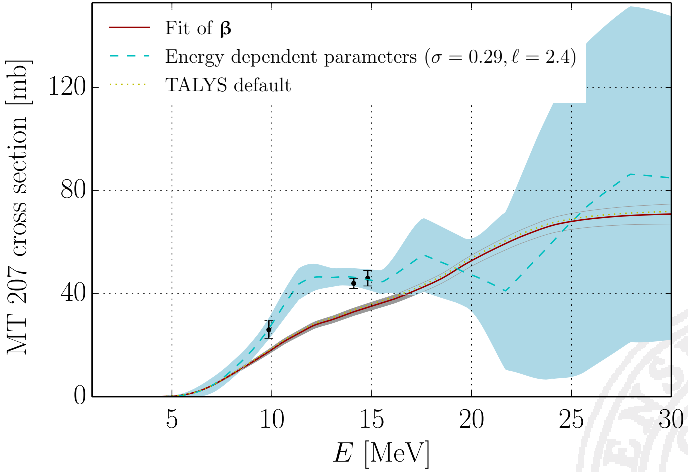
- $\chi^2_{\mathrm{CV}}/\nu = 1.20$

^{56}Fe without one dubious point



- $\chi^2_{\mathrm{CV}}/\nu = 1.20$

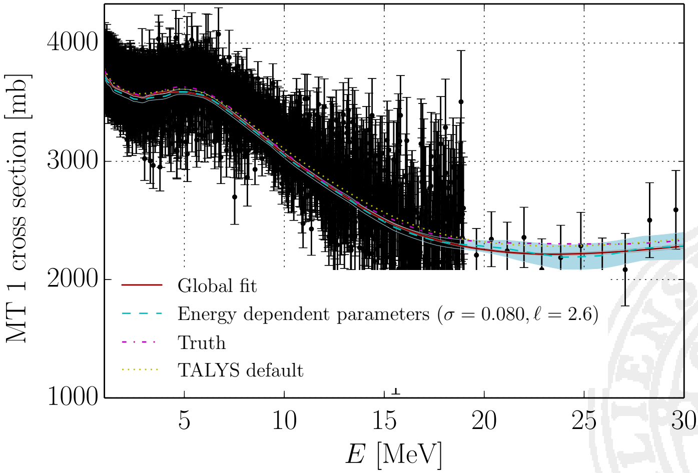
^{56}Fe without one dubious point



- $\chi^2_{\mathrm{CV}}/\nu = 1.20$

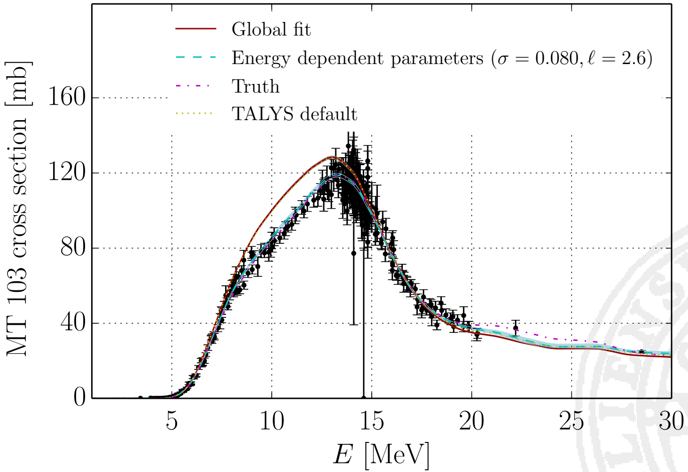
Synthetic data, ^{56}Fe

- Data generated around "Truth" = *Pseudo-TALYS with certain parameters + defect*



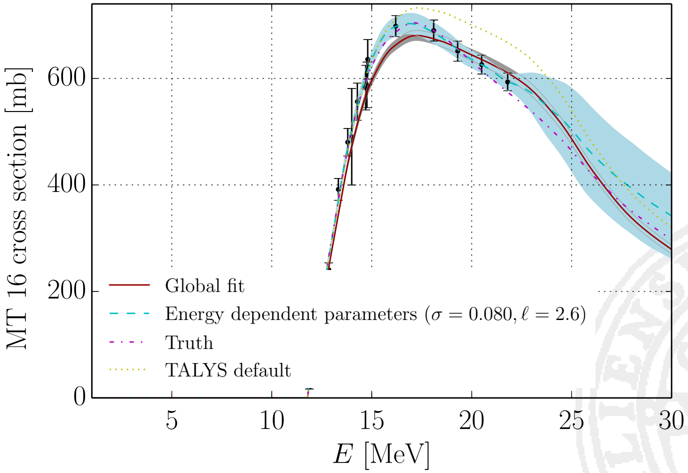
Synthetic data, ^{56}Fe

- Data generated around "Truth" = *Pseudo-TALYS with certain parameters* + *defect*



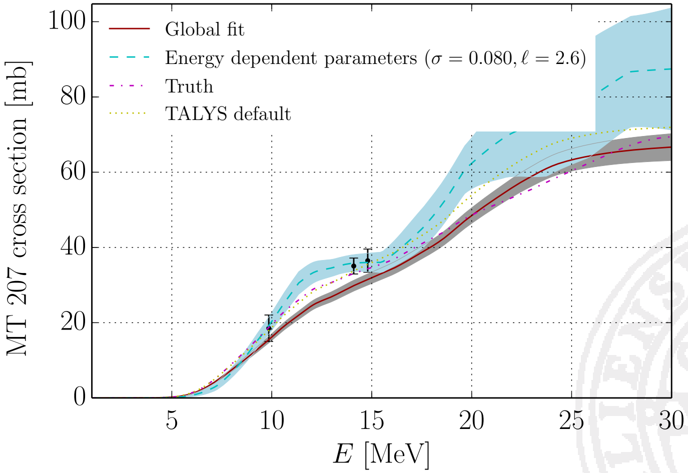
Synthetic data, ^{56}Fe

- Data generated around "Truth" = *Pseudo-TALYS with certain parameters + defect*



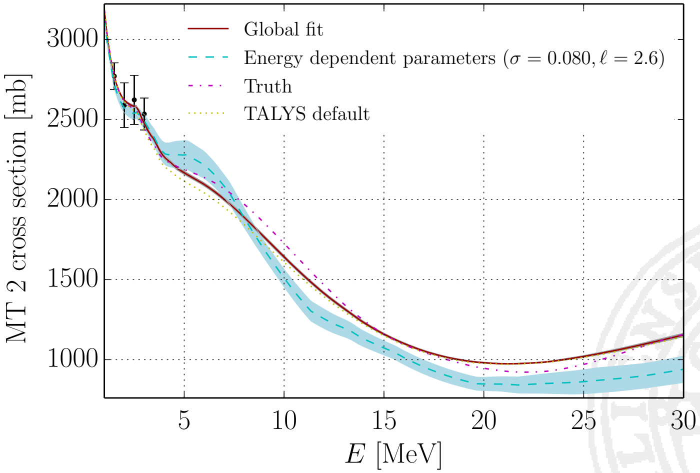
Synthetic data, ^{56}Fe

- Data generated around "Truth" = *Pseudo-TALYS with certain parameters + defect*



Synthetic data, ^{56}Fe

- Data generated around "Truth" = *Pseudo-TALYS with certain parameters + defect*



UPPSALA UNIVERSITET

Conclusions

- We suggest to treat model defects by energy-dependent parameters $\boldsymbol{\gamma}(E) = \boldsymbol{\beta} + \boldsymbol{\delta}(E)$
  - Energy-variation $\boldsymbol{\delta}(E)$ limited by Gaussian processes
- Preliminary results promising
- To do:
  - Treat inconsistent experiments (and reduce trust)
  - Investigate if Levenberg-Marquardt necessary for local fit
  - Move from Pseudo-TALYS to TALYS
  - Proper synthetic data study as in [4]
  - ...
  - ?

[4] 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 **88** (2017)

Conclusions

- We suggest to treat model defects by energy-dependent parameters $\boldsymbol{\gamma}(E) = \boldsymbol{\beta} + \boldsymbol{\delta}(E)$
  - Energy-variation $\boldsymbol{\delta}(E)$ limited by Gaussian processes
- Preliminary results promising
- To do:
  - Treat inconsistent experiments (and reduce trust)
  - Investigate if Levenberg-Marquardt necessary for local fit
  - Move from Pseudo-TALYS to TALYS
  - Proper synthetic data study as in [4]
  - ...
  - ?

Thank you!

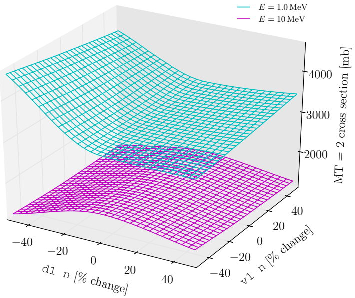
[4] P. Helgesson and H. Sjöstrand, "Fitting a defect non-linear model with or without prior, distinguishing nuclear reaction products as an example", Review of Scientific Instruments **88** (2017)

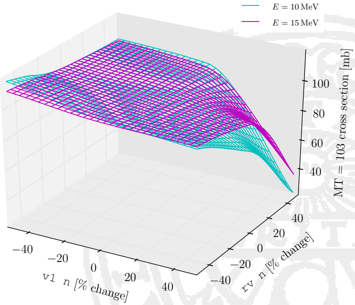
Pseudo-TALYS (^{56}Fe)

- Not intended for producing real results
  - Primarily intended as testing tool in method development
- Includes cross sections with exp. data for $1\,\mathrm{MeV} \leq E \leq 30\,\mathrm{MeV}$ in EXFOR/newbase for ^{56}Fe (16 MTs)

Pseudo-TALYS (^{56}Fe)

- Based on two sets of TALYS runs:
  1. Computing sensitivity at exp. points near default for all parameters varied in TENDL-2015
  2. For each included MT, TALYS is run at a 2D grid with the two parameters contributing most to the uncertainty
- For (pseudo-)exclusive MTs (w.r.t. each E), Pseudo-TALYS returns PCHIP interpolated results from the 2D grid plus linear dependence on other parameters.
- Other MTs are obtained as sums of pseudo-exclusive MTs, *e.g.*, (n, tot) $=$ (n, el) $+$ (n, non-el) $=$ (n, el) $+$ ... (depends on E)





Pseudo-TALYS (^{56}Fe)