DE LA RECHERCHE À L'INDUSTRIE

Irfu - CEA Saclay
Institut de recherche sur les lois fondamentales de l'Univers

Towards a holistic framework for global assessments of nuclear models

Georg Schnabel

Café SPhN, 20.11.2017

What is it about?

Situation

Models (computationally expensive)

Experimental data (possibly a lot)

Research question

How can we use statistical methods and methods of machine learning in combination with modern computer infrastructure to improve our knowledge about nuclear models and experimental data?

Parameter estimation, <u>uncertainty quantification</u>, uncertainty propagation

Publications

G. Schnabel "**Estimating model bias over the complete nuclide chart with sparse Gaussian processes at the example of INCL/ABLA and double-differential neutron spectra**", submitted to EPJ-N

G. Schnabel "**Fitting and Analysis Technique for Inconsistent Nuclear Data**" Proc. of M&C 2017

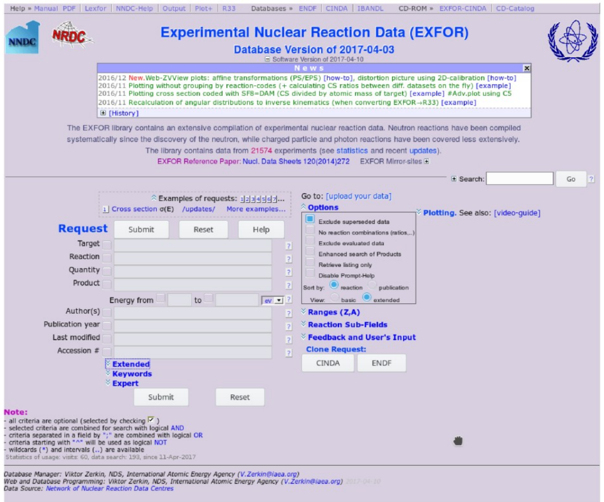
G. Schnabel, and H. Leeb. "**A Modified Generalized Least Squares Method for Large Scale Nuclear Data Evaluation**" NIMA Jan 2017

G. Schnabel "**Adaptive Monte Carlo for Nuclear Data Evaluation**" Proc. of ND 2016

G. Schnabel, and H. Leeb. "**Differential Cross Sections and the Impact of Model Defects in Nuclear Data Evaluation**" Proc. of Wonder 2015.

"Toy" scenario

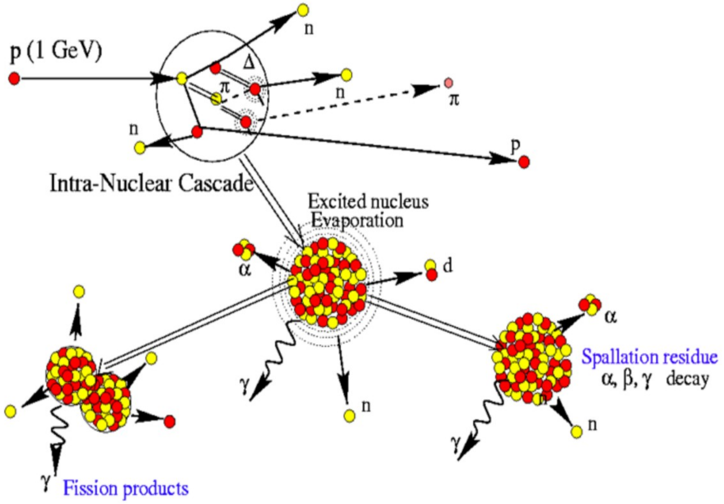
Data: EXFOR database



Database as of: 2017-04-03

Number of ENTRY	21 574	experimental works
Number of SUBENT	150 976	data tables
Number of Datasets	167 857	data tables of reactions
Number of Datapoints	**14 739 297**	total number of data points

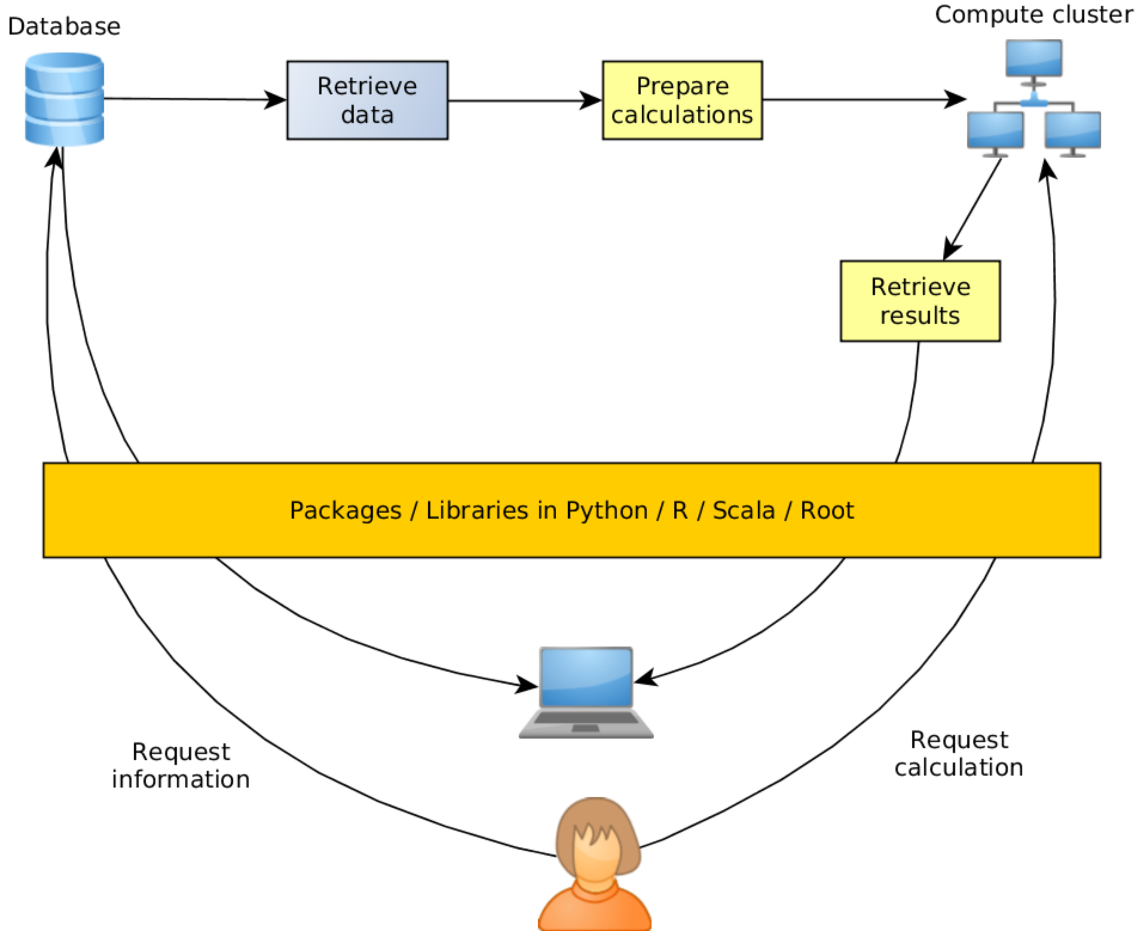
Model: INCL/ABLA



Features:

- Stochastic output
- Computational expensive
- Many parameters
- Large output

Automatization

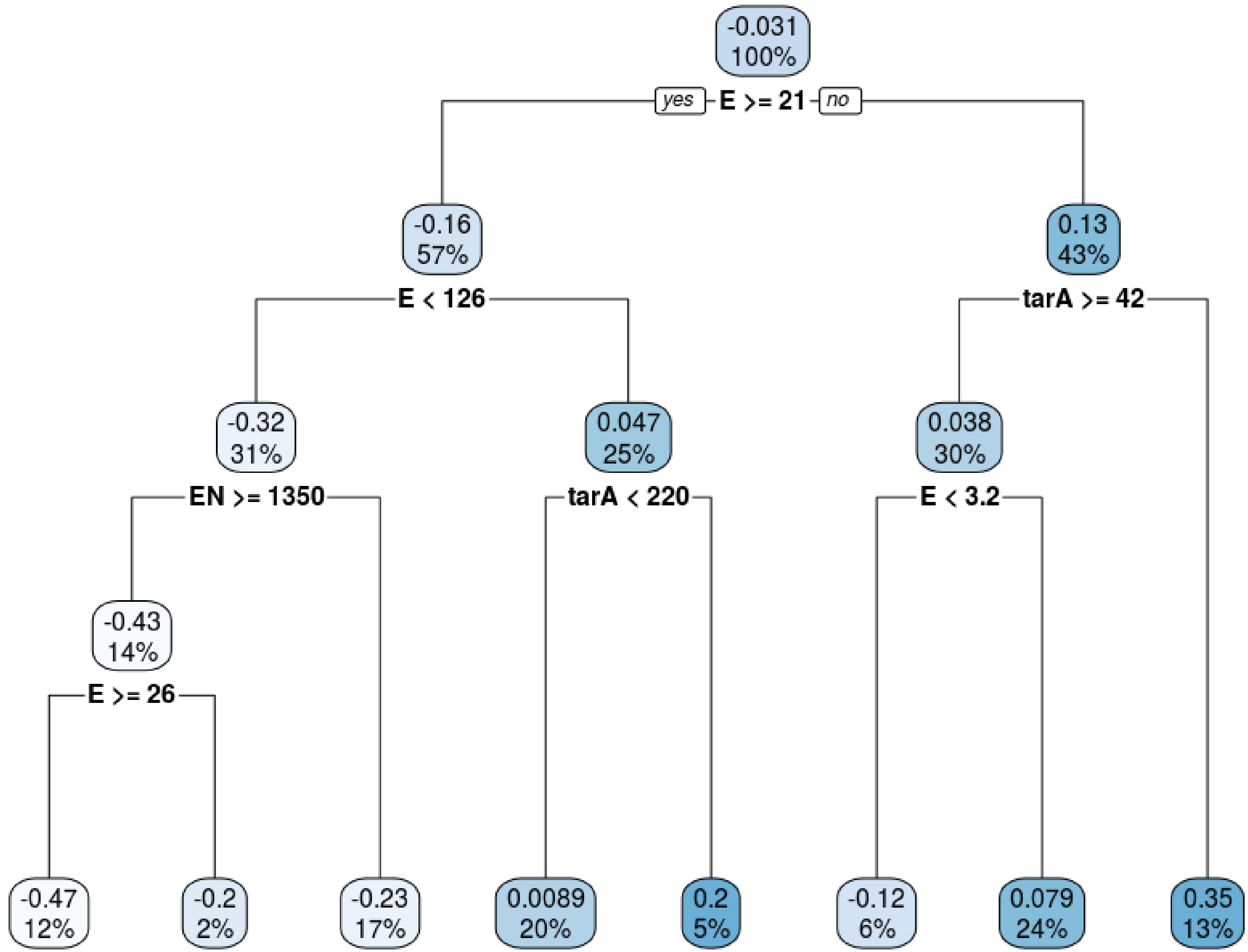


	IDX	REAC	EN	ACCNUM	SUBACCNUM	ANG	E	DATA	PROJ.SYM	TAR.SYM	EJECT.SYM	QUANT	A	Z	MODXS	MODERR	tarA	tarZ	RELDIFF	EXPERR
1:	77048	(13-AL-27(P,X)0-NN-1,,DA/DE)	800	E1762	020	15	1.25	10.020	p	Al27	n	DDX	1	0	6.54215291	0.517202600	27	13	-0.53160590	1.0020
2:	77049	(13-AL-27(P,X)0-NN-1,,DA/DE)	800	E1762	020	15	1.75	5.963	p	Al27	n	DDX	1	0	6.74659519	0.525221723	27	13	0.11614676	0.5963
3:	77050	(13-AL-27(P,X)0-NN-1,,DA/DE)	800	E1762	020	15	2.50	6.791	p	Al27	n	DDX	1	0	6.29682218	0.717589485	27	13	-0.07848051	0.6791
4:	77051	(13-AL-27(P,X)0-NN-1,,DA/DE)	800	E1762	020	15	3.50	5.278	p	Al27	n	DDX	1	0	5.88793762	0.693900103	27	13	0.10359105	0.5278
5:	77052	(13-AL-27(P,X)0-NN-1,,DA/DE)	800	E1762	020	15	4.50	4.833	p	Al27	n	DDX	1	0	5.64260689	0.679290043	27	13	0.14348100	0.4833

9283:	109738	(40-ZR-0(P,X)0-NN-1,,DA/DE)	1600	O1170	008	150	81.00	0.190	p	Zr	n	DDX	1	0	0.13737011	0.030716891	91	40	-0.38312475	0.0190
9284:	109739	(40-ZR-0(P,X)0-NN-1,,DA/DE)	1600	O1170	008	150	92.00	0.130	p	Zr	n	DDX	1	0	0.22666068	0.039456560	91	40	0.42645545	0.0130
9285:	109740	(40-ZR-0(P,X)0-NN-1,,DA/DE)	1600	O1170	008	150	106.00	0.096	p	Zr	n	DDX	1	0	0.10962135	0.006222730	91	40	0.12425817	0.0096
9286:	109741	(40-ZR-0(P,X)0-NN-1,,DA/DE)	1600	O1170	008	150	124.00	0.064	p	Zr	n	DDX	1	0	0.07555356	0.005270431	91	40	0.11553561	0.0064
9287:	109742	(40-ZR-0(P,X)0-NN-1,,DA/DE)	1600	O1170	008	150	147.00	0.053	p	Zr	n	DDX	1	0	0.04361501	0.004144504	91	40	-0.09384990	0.0053

S. V. Gleyzer et al, Development of Machine Learning Tools in ROOT, J. Phys.: Conf. Ser. 762 012043, 2016

Example of ML approach



Terry Therneau, Beth Atkinson and Brian Ripley (2017). rpart: Recursive Partitioning and Regression Trees. R package version 4.1-11. https://CRAN.R-project.org/package=rpart

DE LA RECHERCHE À L'INDUSTRIE

Bayesian statistics

$$P(H\,|\,O) = \frac{P(O\,|\,H)P(H)}{P(O)}$$

H hypothesis
O observation

P(H) probability of hypothesis to be true
P(O) probability of observation to occur

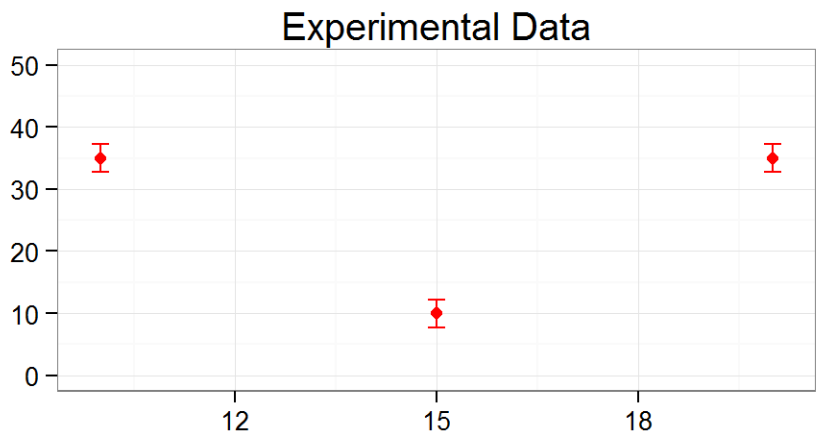
P(O|H) probability of observation O to occur if hypothesis H is true

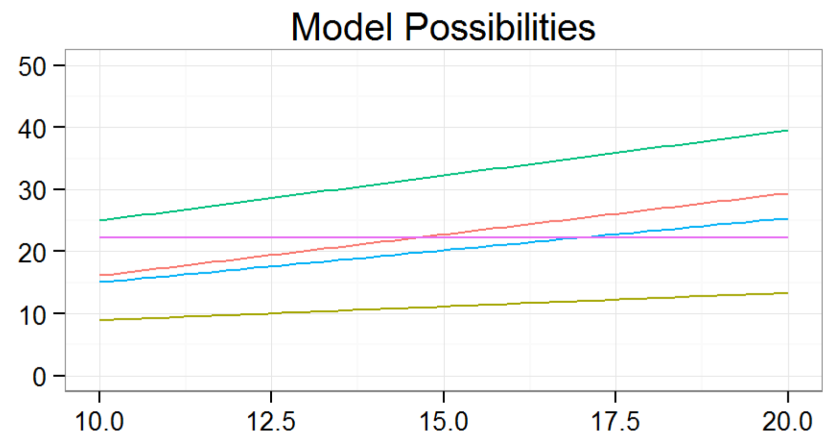
P(H|O) probability of hypothesis after we observed O

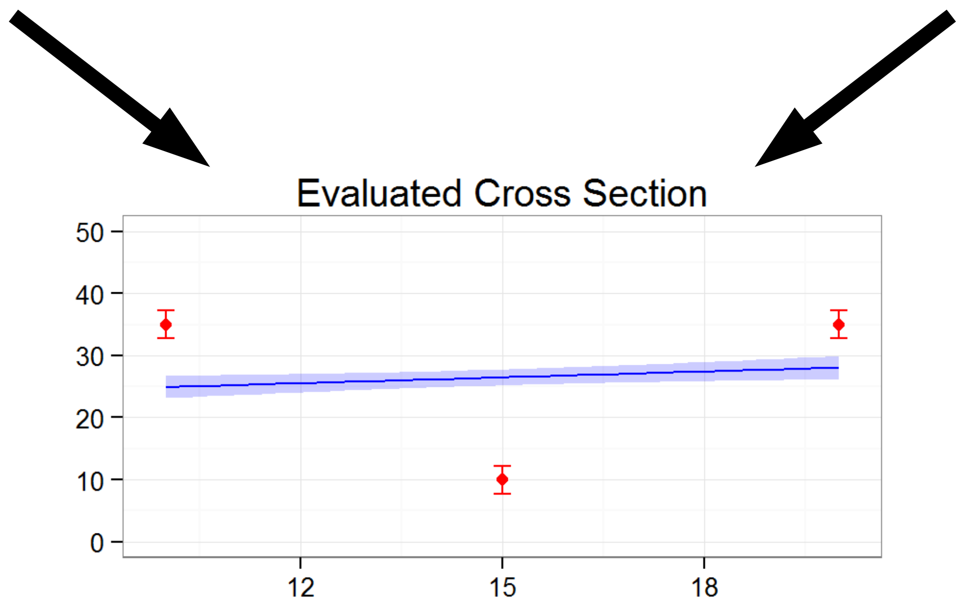
Consistent with Aristotelian logic
Consistent with principles of common sense

R. T. Cox, "Probability, Frequency and Reasonable Expectation" American Journal of Physics, vol. 14, no. 1, p. 1, 1946.

Inappropriate assumptions







GiGo principle
Garbage in,
Garbage out
But also
Good stuff in,
Good stuff out

In practice

Negative cross sections in linearized evaluation methods

Uncertainty reductions beyond experimental limits

Model predictions in disagreement with experiment data

Reasons

Inappropriate prior for model parameters

Imperfect model / Not completely confident in the model

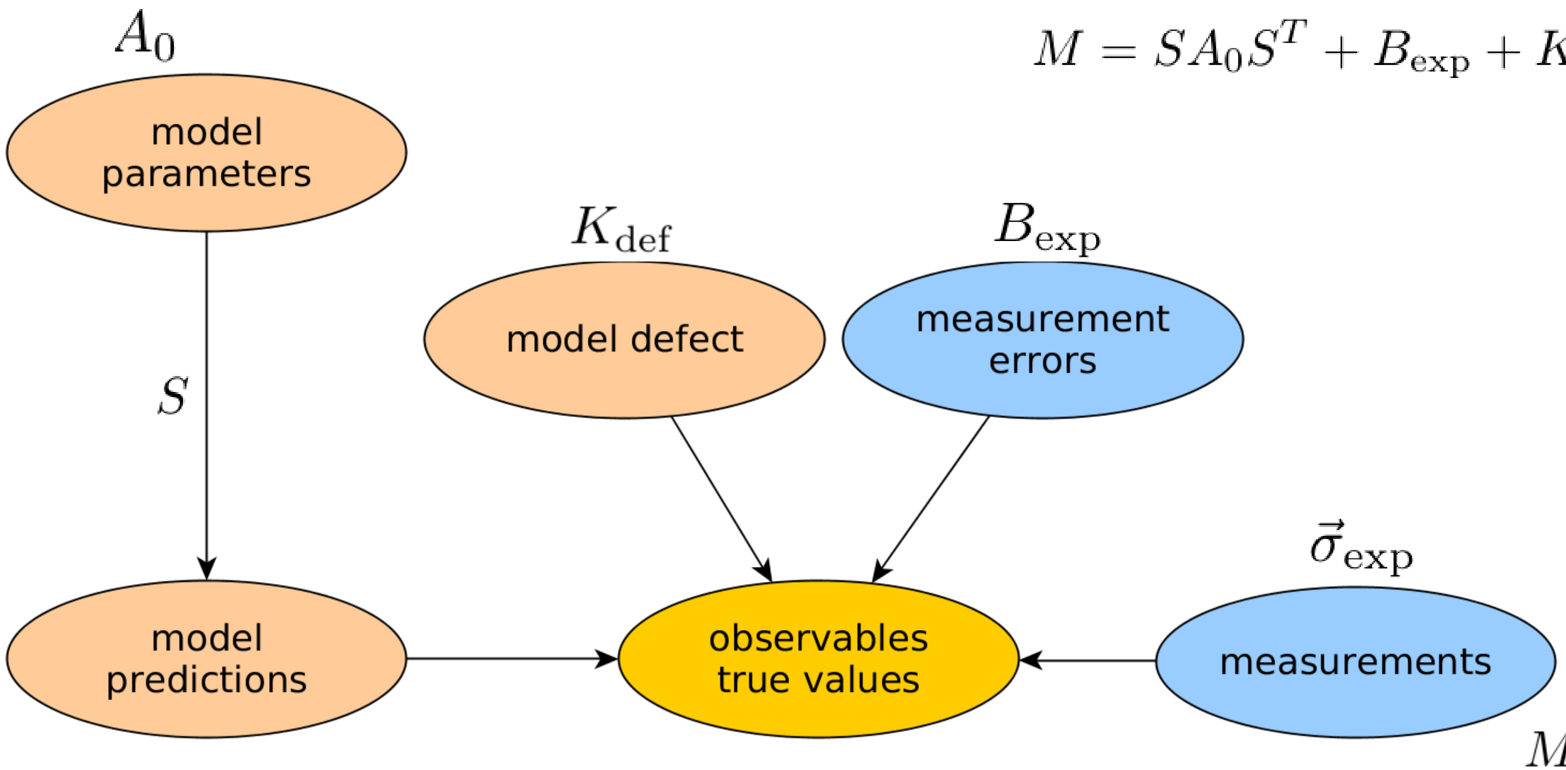
Inaccurate likelihood specification for the data

Solutions

Prior rescaling, likelihood broadening, model defects, removing suspicious experimental data sets

Bayesian network

$$M = S A_0 S^T + B_{\text{exp}} + K_{\text{def}}$$



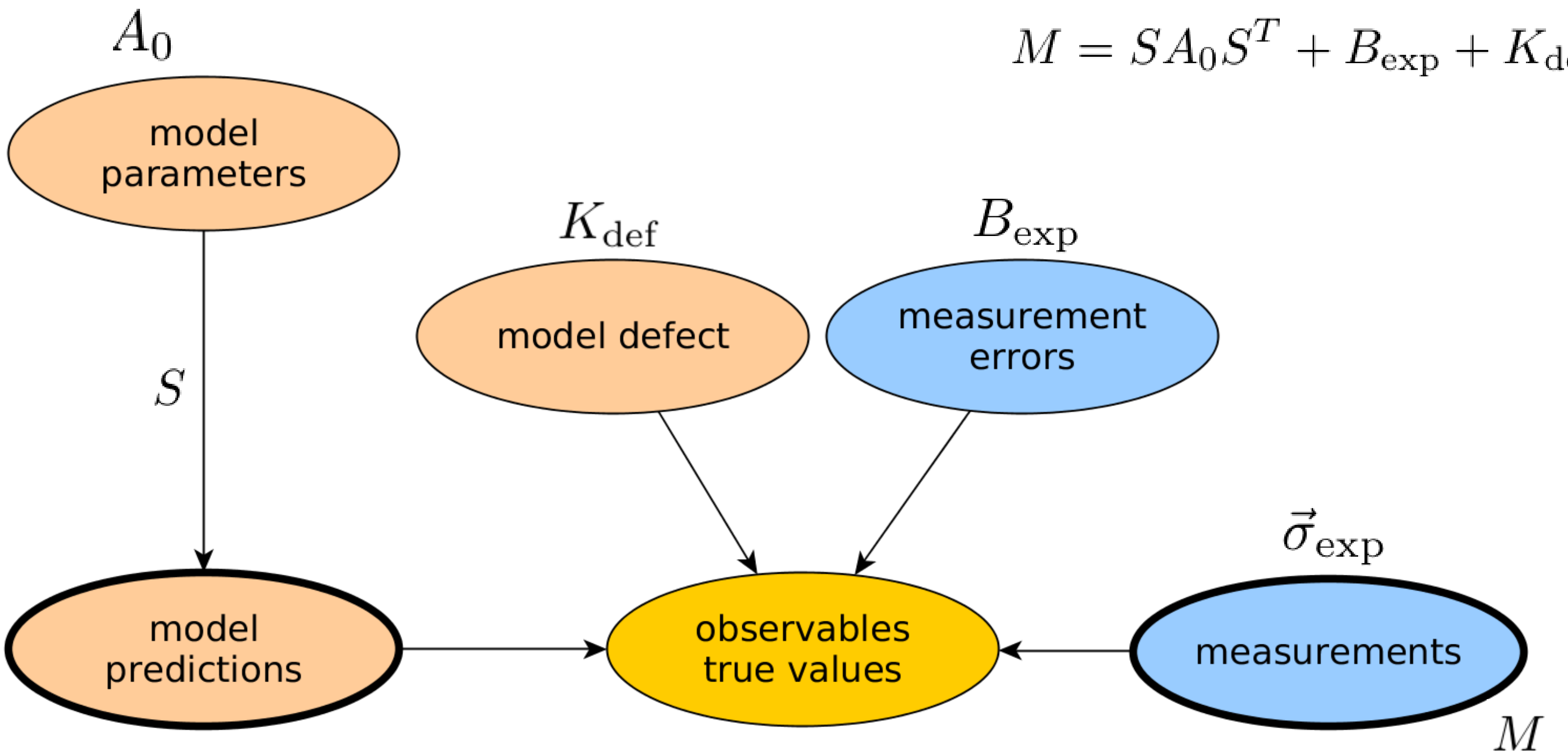
$$\vec{p}_1 = \vec{p}_0 + A_0 S^T M^{-1} (\vec{\sigma}_{\text{exp}} - S\vec{p}_0)$$

$$\vec{\varepsilon}_{\text{exp},1} = B_{\text{exp}} M^{-1} (\vec{\sigma}_{\text{exp}} - S\vec{p}_0)$$

$$A_1 = A_0 - A_0 S^T M^{-1} S A_0$$

$$B_{\text{exp},1} = B_{\text{exp}} - B_{\text{exp}} M^{-1} B_{\text{exp}}$$

Deterministic codes



$$M = SA_0S^T + B_{\mathrm{exp}} + K_{\mathrm{def}}$$

$$\vec{p}_1 = \vec{p}_0 + A_0 S^T M^{-1}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

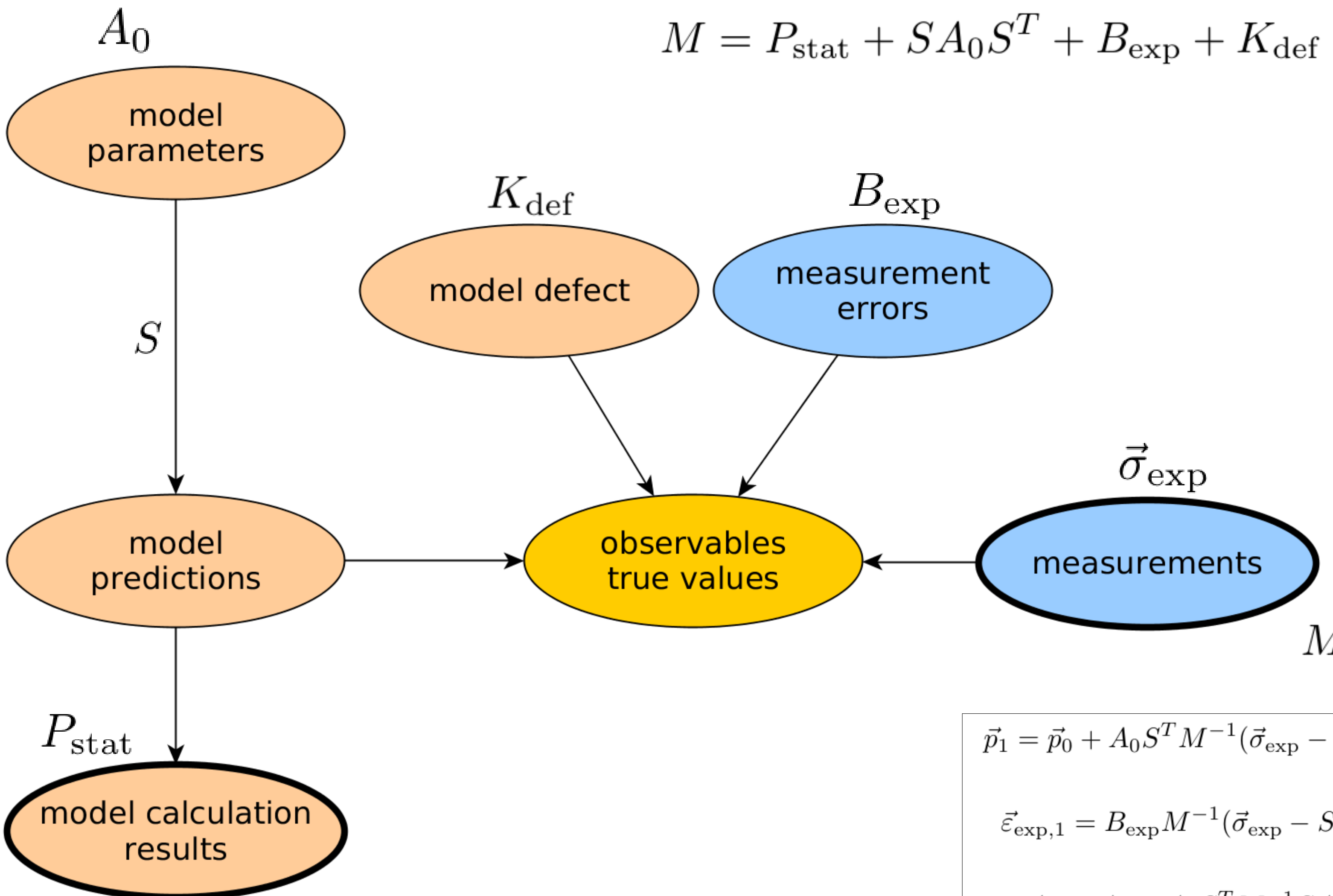
$$\vec{\varepsilon}_{\mathrm{exp},1} = B_{\mathrm{exp}} M^{-1}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

$$A_1 = A_0 - A_0 S^T M^{-1} S A_0$$

$$B_{\mathrm{exp},1} = B_{\mathrm{exp}} - B_{\mathrm{exp}} M^{-1} B_{\mathrm{exp}}$$

Stochastic codes

$$M = P_{\text{stat}} + S A_0 S^T + B_{\text{exp}} + K_{\text{def}}$$



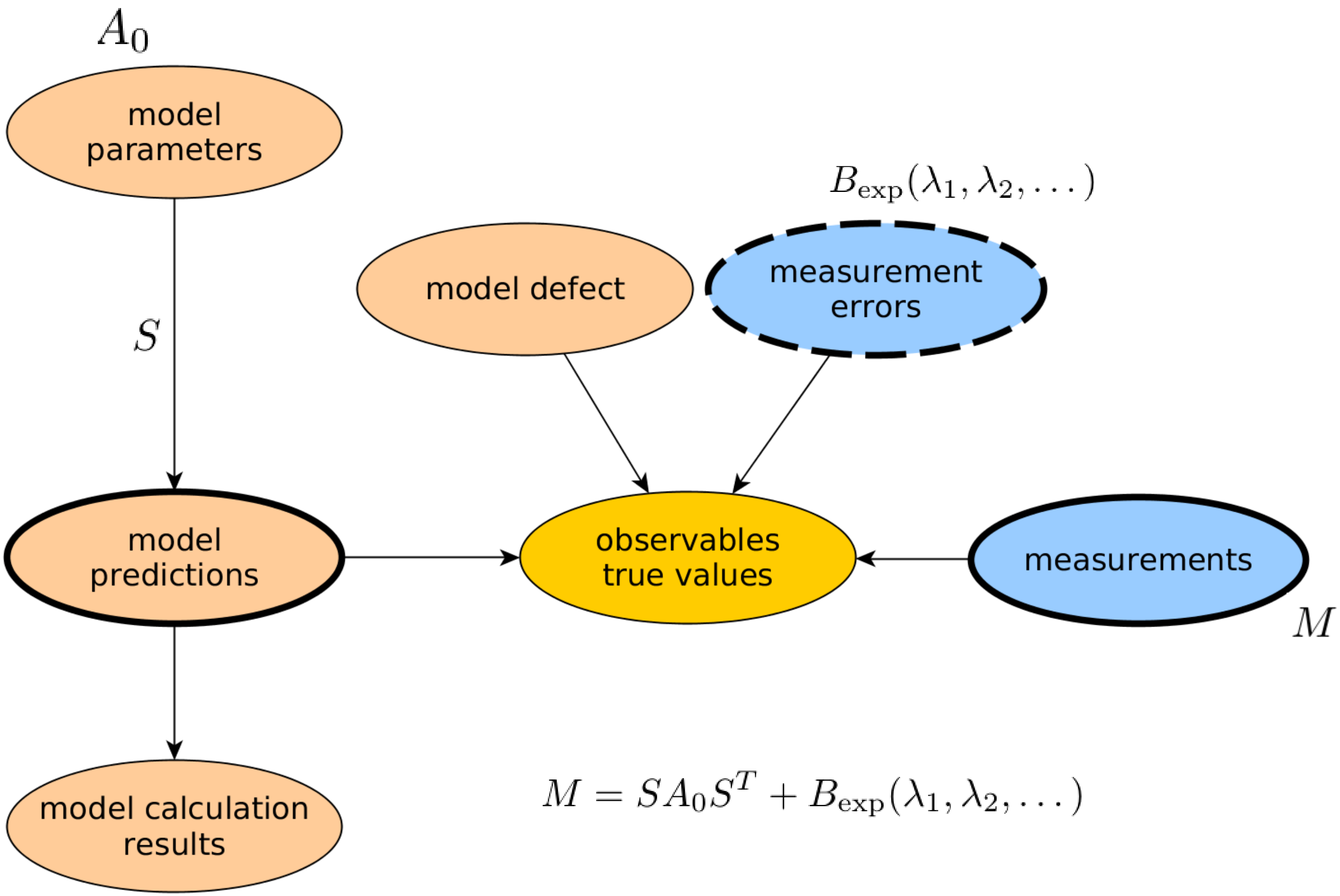
$$\vec{p}_1 = \vec{p}_0 + A_0 S^T M^{-1} (\vec{\sigma}_{\text{exp}} - S\vec{p}_0)$$

$$\vec{\varepsilon}_{\text{exp},1} = B_{\text{exp}} M^{-1} (\vec{\sigma}_{\text{exp}} - S\vec{p}_0)$$

$$A_1 = A_0 - A_0 S^T M^{-1} S A_0$$

$$B_{\text{exp},1} = B_{\text{exp}} - B_{\text{exp}} M^{-1} B_{\text{exp}}$$

Inconsistent data



$$M = S A_0 S^T + B_{\text{exp}}(\lambda_1, \lambda_2, \ldots)$$

G. Schnabel, **Fitting and Analysis Technique for Inconsistent Nuclear Data**
Proc. Int. Conf. on Mathematics & Computational Methods Applied to Nuclear Science & Engineering, April 2017

Marginal likelihood

$$\log \rho(\vec{\sigma}_{\mathrm{exp}} \,|\, \vec{p}_0, S, M) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|M| - \frac{1}{2}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)^T M^{-1} (\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

Entropy

χ^2 term

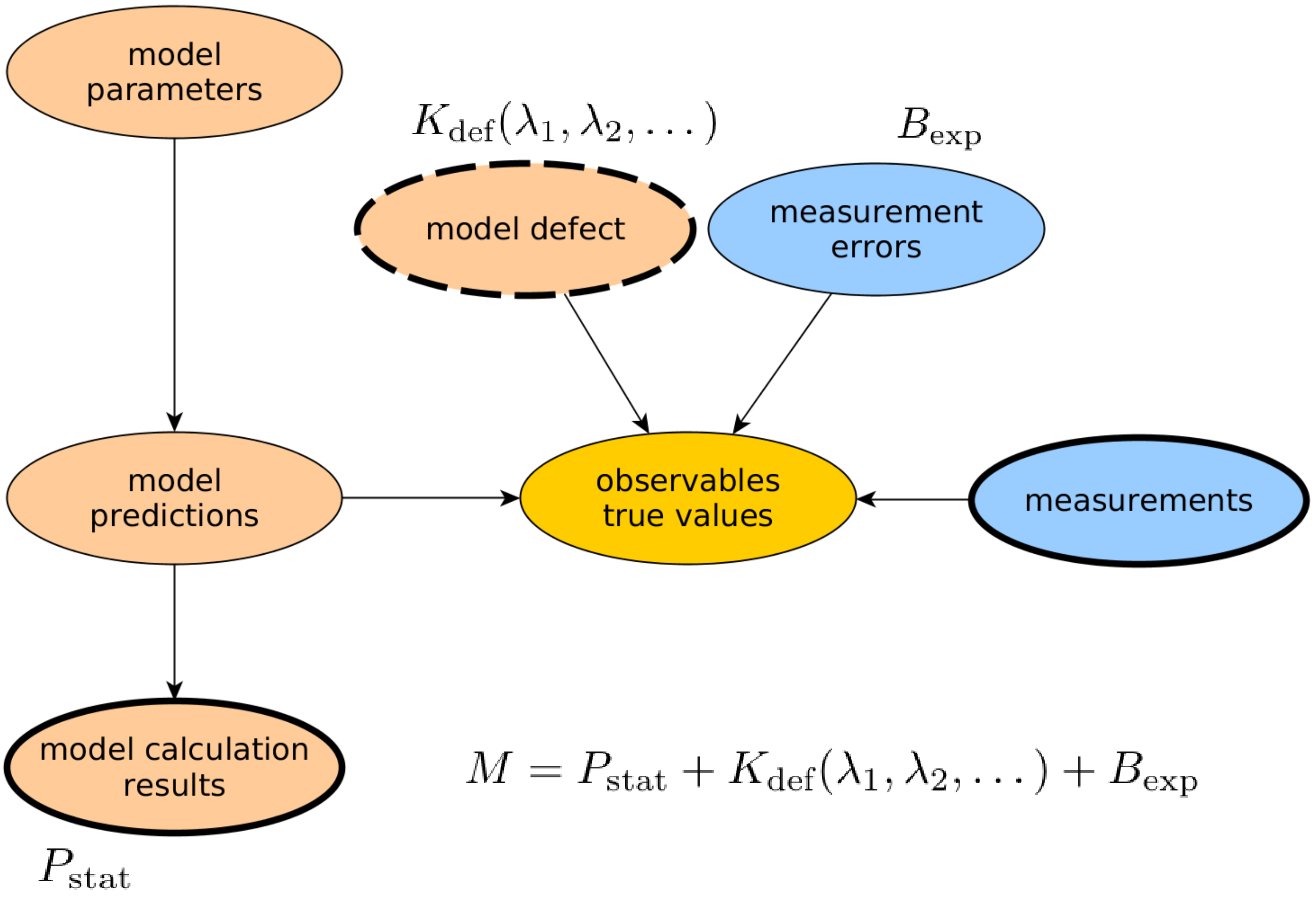
$$M = S A_0 S^T + B_{\mathrm{exp}}(\lambda_1, \lambda_2, \ldots)$$

Questions

Is it computationally feasible?

Can we efficiently maximize this expression?

Imperfect model



$$M = P_{\mathrm{stat}} + K_{\mathrm{def}}(\lambda_1, \lambda_2, \dots) + B_{\mathrm{exp}}$$

(In)finite Covariance Matrix

Covariance matrices can represent a variety of things such as normalization uncertainties, linear trends, splines, Fourier series, polynomial expansions, white noise, etc.

$$y(x) = kx + d, \quad k \sim \mathcal{N}(0, \delta_k^2),\ d \sim \mathcal{N}(0, \delta_d^2)$$

$$\text{Observations} \quad (\vec{y}_{\text{exp}}, \vec{x}_{\text{exp}})$$

$$\vec{p} = \begin{pmatrix} k \\ d \end{pmatrix} = AS^T \left(SAS^T + B\right)^{-1} \vec{y}_{\text{exp}}$$

$$\vec{y}_{\text{pred}} = S_{\text{pred}}\vec{p} = \left(\vec{x}_{\text{pred}}, \vec{1}\right) \begin{pmatrix} k \\ d \end{pmatrix}$$

$$S_{\text{exp}} = \left(\vec{x}_{\text{exp}}, \vec{1}\right)$$

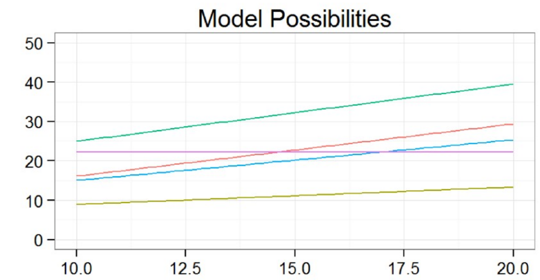
$$A = \begin{pmatrix} \delta_k^2 & 0 \\ 0 & \delta_d^2 \end{pmatrix}$$

$$B = \begin{pmatrix} \varepsilon_1^2 & 0 & 0 \\ 0 & \varepsilon_2^2 & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

$$\kappa(x_1, x_2) := Cov[y(x_1), y(x_2)] = \delta_k^2 x_1 x_2 + \delta_d^2$$

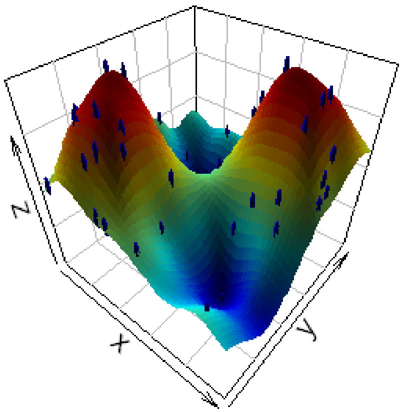
$$K_{\text{pred,exp}} = \kappa(\vec{x}_{\text{pred}}, \vec{x}_{\text{exp}}) \quad K_{\text{exp,exp}} = \kappa(\vec{x}_{\text{exp}}, \vec{x}_{\text{exp}})$$

$$\vec{y}_{\text{pred}} = K_{\text{pred,exp}} K_{\text{exp,exp}}^{-1} \vec{y}_{\text{exp}}$$



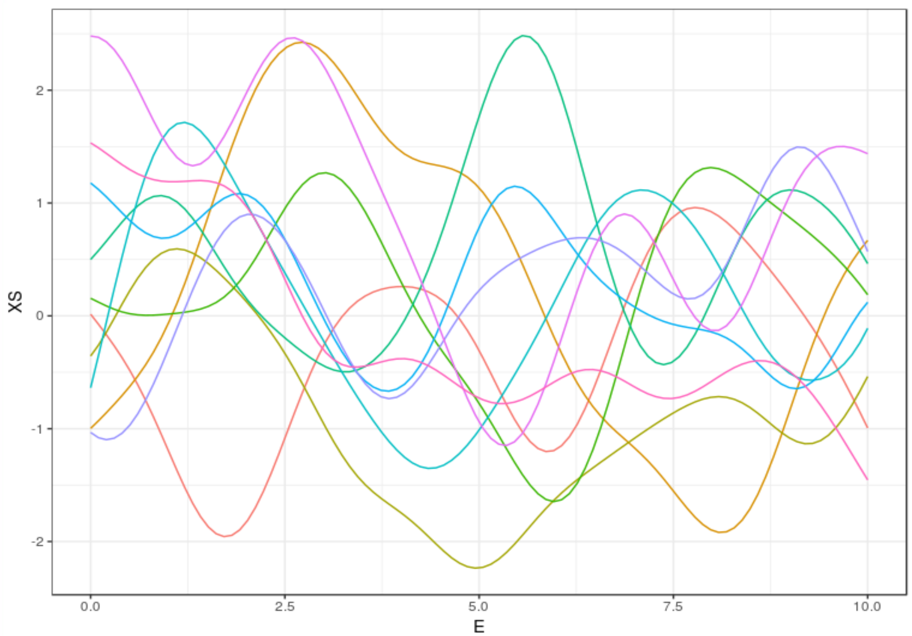
Gaussian processes

Powerful concept
Directly parametrize covariance matrix and work implicitly with an infinite number of parameters/basis functions!

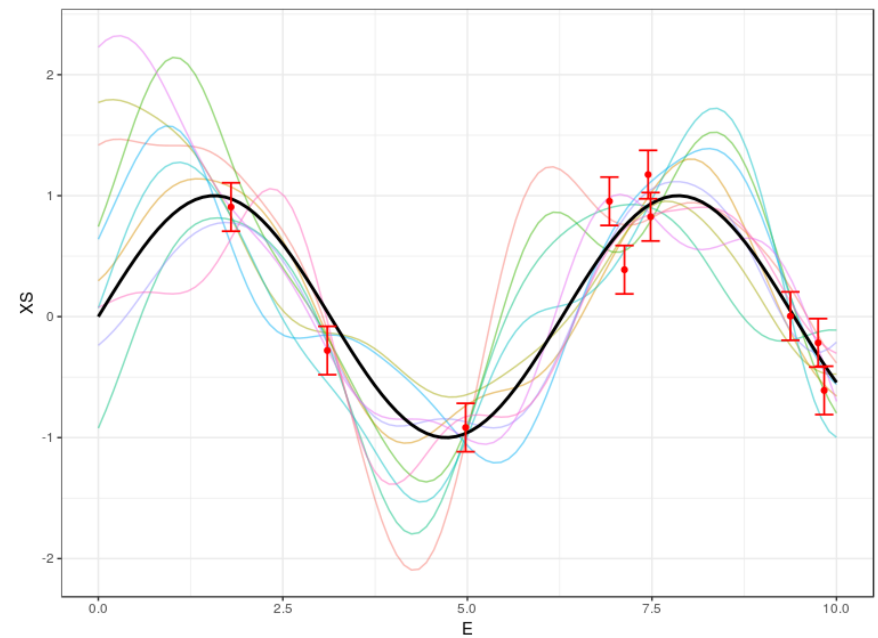


$$\kappa(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta^2 \exp\left(-\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^2}{2\lambda^2}\right)$$

Sample from prior (δ=λ=1)

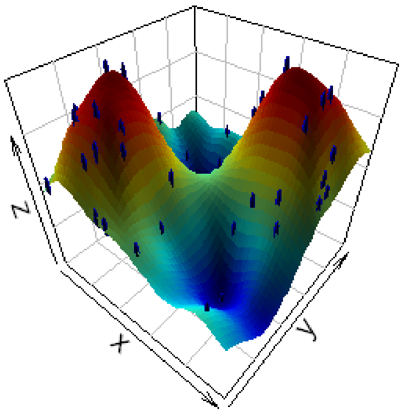


Sample from posterior



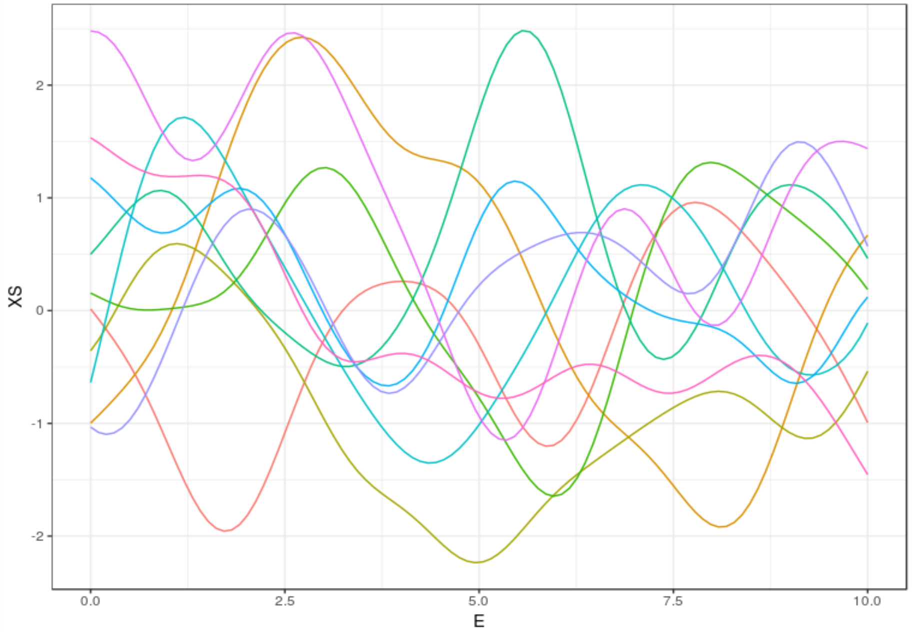
Gaussian processes

Powerful concept
Directly parametrize covariance matrix and work implicitly with an infinite number of parameters/basis functions!

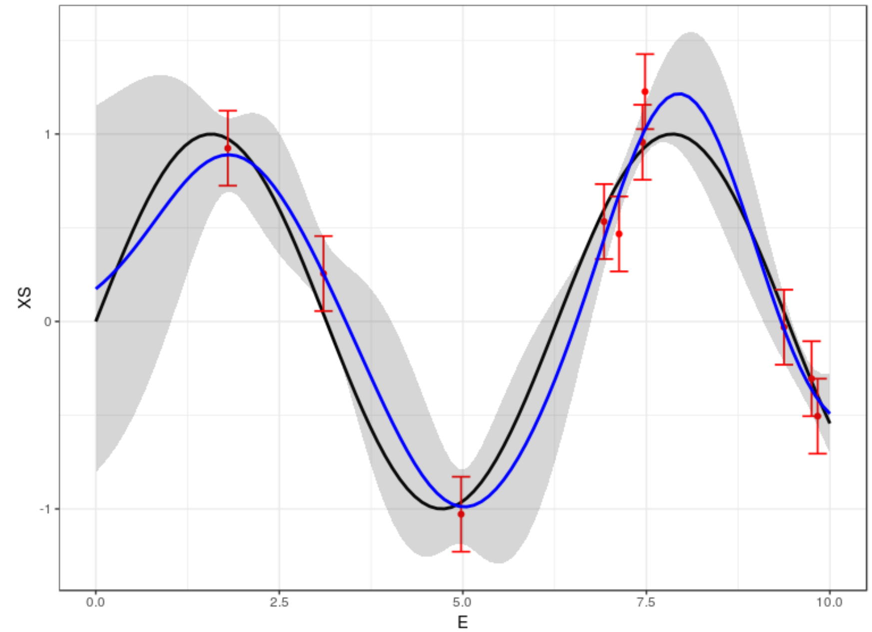


$$\kappa(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta^2 \exp\left(-\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^2}{2\lambda^2}\right)$$

Sample from prior (δ=λ=1)

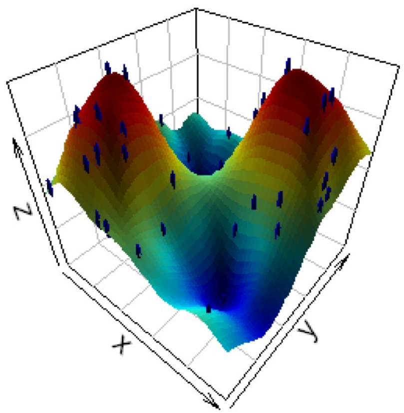


Posterior uncertainty

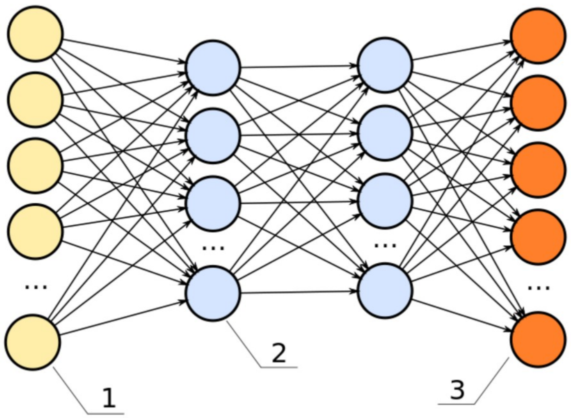


Comparison to neural networks

GP processes



artificial neural networks



Both approaches …

… are methods for classification and regression

… are universal function approximators

Neural networks …

… scale better to large data sets

… are able to capture non-local features

… are difficult to interpret

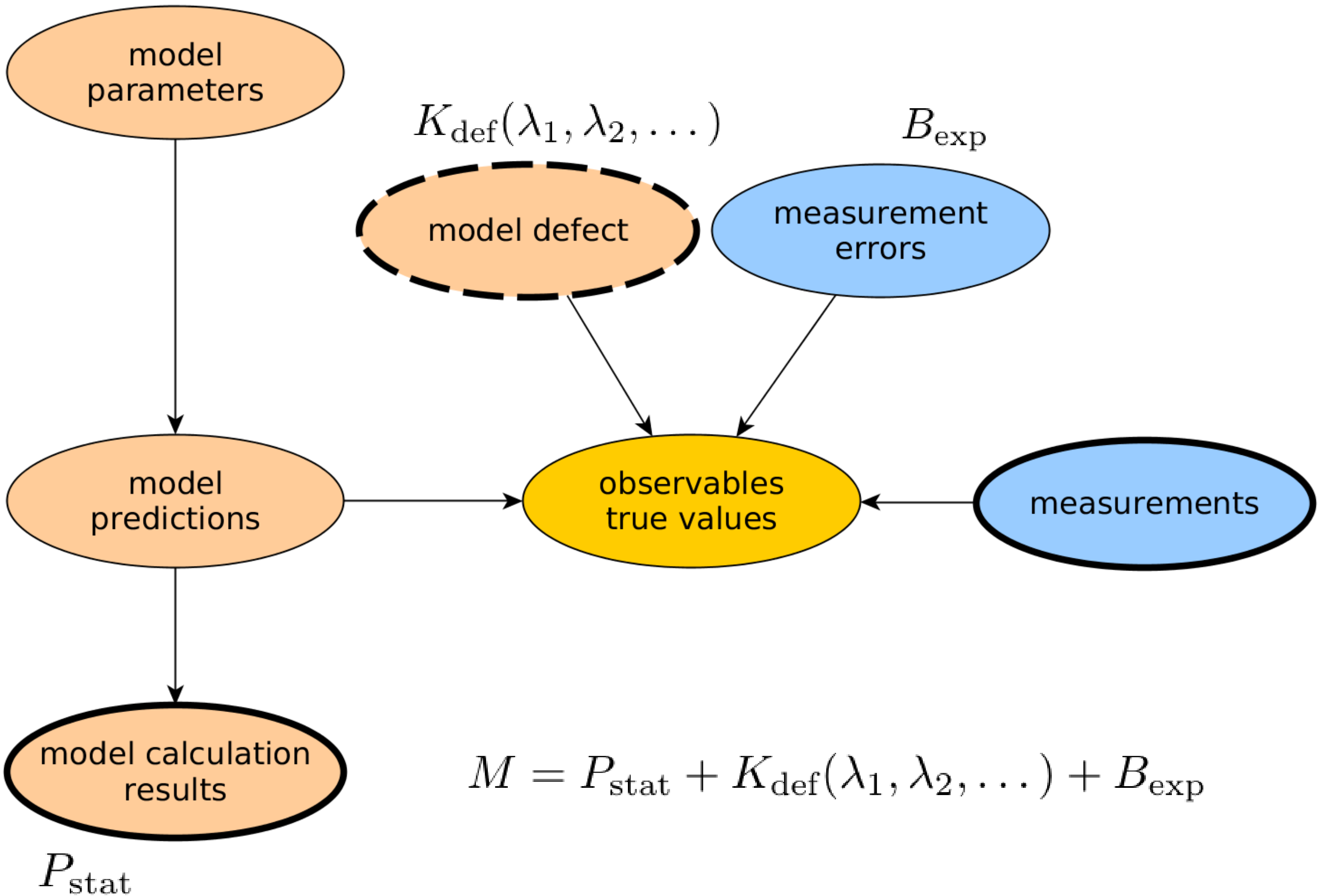
GP processes …

… are statistical methods from the ground up (uncertainties)

… facilitate the incorporation of prior assumptions

… interface well with existing nuclear data evaluation methods

Model bias estimation



$$M = P_{\mathrm{stat}} + K_{\mathrm{def}}(\lambda_1, \lambda_2, \dots) + B_{\mathrm{exp}}$$

(p,X)n above 100 MeV

Isotope	En	E_{min}	E_{max}	θ_{min}	θ_{max}	NumPts
C	800	1.2	700	15	150	189
C	1500	1.2	1250	15	150	245
C	3000	1.2	2500	15	150	128
Na23	800	3.5	266	30	150	84
Al27	800	1.2	700	15	150	119
Al27	1000	2.5	280	15	150	223
Al27	1200	2.0	1189	10	160	404
Al27	1500	1.2	1250	15	150	129
Al27	1600	2.5	280	15	150	226
Al27	3000	1.2	2500	15	150	132
Fe	800	1.2	771	10	160	505
Fe	1200	2.0	1171	10	160	417
Fe	1500	1.2	1250	15	150	129
Fe	1600	2.0	1572	10	160	460
Fe	3000	1.2	2500	15	150	133
Cu	1000	2.5	280	15	150	227
Cu	1600	2.5	280	15	150	231
Zr	1000	2.5	280	15	150	229
Zr	1200	2.0	1189	10	160	423
Zr	1600	2.5	280	15	150	229
In	800	1.2	700	15	150	116
In	1500	1.2	1250	15	150	128
In	3000	1.2	2500	15	150	133
W	800	3.1	333	30	150	110
W	1000	2.5	280	15	150	231
W	1200	2.0	1189	10	160	413
W	1600	2.5	280	15	150	231
Pb	318	5.4	356	7	7	53
Pb	800	1.2	771	10	160	624
Pb	1000	2.5	280	15	150	231
Pb	1200	2.0	1189	10	160	563
Pb	1500	1.2	1250	15	150	249
Pb	1600	2.0	1591	10	160	691
Pb	3000	1.2	2500	15	150	131
Pb208	2000	0.4	402	30	150	170
Th232	1200	2.0	1189	10	160	351

Input space:
A, Z, En, E, θ

Covariance brewing

Combination rules for covariance functions

$$\kappa_{1+2}(x_1, x_2) = \kappa_1(x_1, x_2) + \kappa_2(x_1, x_2)$$

$$\kappa_{1\times 2}(x_1, x_2) = \kappa_1(x_1, x_2) \times \kappa_2(x_1, x_2)$$

Squared exponential covariance functions

$$\kappa_1(x_1, x_2) = {\color{blue}\delta_1}^2 \exp\left(\frac{1}{2{\color{blue}\lambda_1}^2}(x_1 - x_2)^2\right)$$

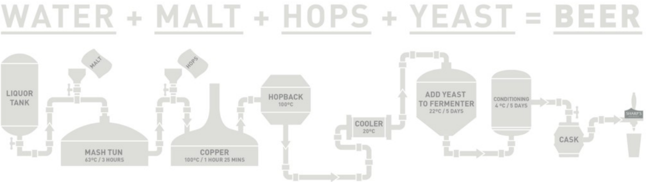
$$\kappa_2(x_1, x_2) = {\color{blue}\delta_2}^2 \exp\left(\frac{1}{2{\color{blue}\lambda_2}^2}(x_1 - x_2)^2\right)$$

Transition kernel

$$\tau_1(x_1, x_2) = \sigma(x_1)\sigma(x_2)$$

$$\tau_2(x_1, x_2) = (1 - \sigma(x_1)(1 - \sigma(x_2))$$

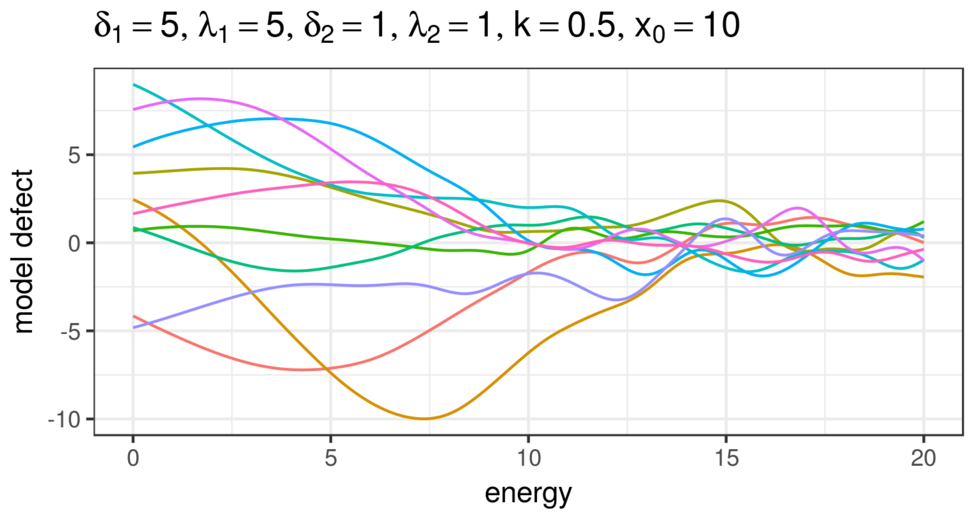
$$\sigma(x) = \frac{1}{1 + \exp(-{\color{blue}k}(x - {\color{blue}x_0}))}$$

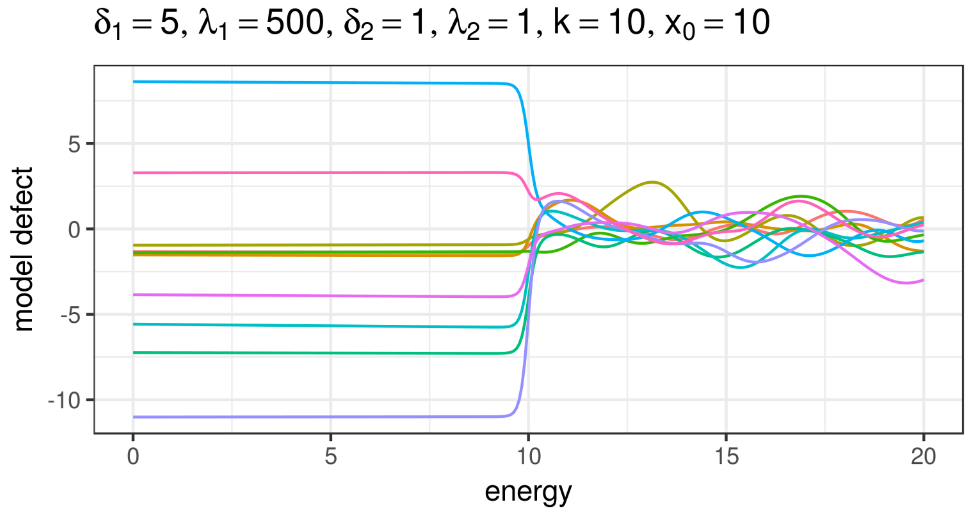


$$\boxed{\kappa_{\text{comp}}(x_1, x_2) = \tau_1(x_1, x_2) \times \kappa_1(x_1, x_2) + \tau_2(x_1, x_2) \times \kappa_2(x_1, x_2)}$$

Prior visualization

$$\kappa_{\text{comp}}(x_1, x_2) = \tau_1(x_1, x_2) \times \kappa_1(x_1, x_2) + \tau_2(x_1, x_2) \times \kappa_2(x_1, x_2)$$





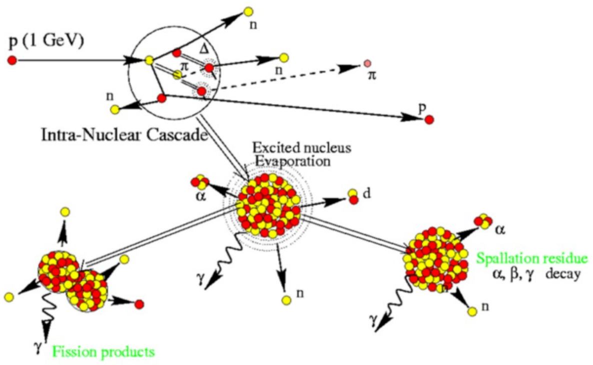
DE LA RECHERCHE À L'INDUSTRIE

cea

Real case

INCL vs experiment data

Final goal: Inclusive DDX data over the complete nuclide chart projectile + target -> ejectile + X (~100 000 data points above 100 MeV, INCL gives predictions for ~40 000)



Inclusive DDX for **p + target -> X + n**

9287 data points, 11 targets, incident energies ranging from 300 to 3000 MeV)

5 dimensional space (A, Z, EN, ANG, E)

κ_1: δ_1, $\lambda_{11}, \ldots, \lambda_{15}$

κ_2: δ_2, $\lambda_{21}, \ldots, \lambda_{25}$

T: k, x_0, Φ

$$M = P_{\mathrm{stat}} + K_{\mathrm{def}}(\lambda_1, \lambda_2, \ldots) + B_{\mathrm{exp}}$$

$$\log \rho(\mathcal{D} \,|\, \vec{p}_0, \vec{\sigma}_{\mathrm{exp}}, M) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|M| - \frac{1}{2}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)^T M^{-1}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

$$\boxed{\kappa_{\mathrm{comp}}(x_1, x_2) = \tau_1(x_1, x_2) \times \kappa_1(x_1, x_2) + \tau_2(x_1, x_2) \times \kappa_2(x_1, x_2)}$$

Introduce sparsity

Assume latent variables (**ps**eudo-**i**nputs)

$$\vec{y}_{\mathrm{obs}} = \overbrace{K_{\mathrm{obs,psi}} K^{-1}_{\mathrm{psi,psi}}}^{S} \vec{y}_{\mathrm{psi}} \qquad \vec{y}_{\mathrm{psi}} \sim \mathcal{N}(\vec{0}, K_{\mathrm{psi,psi}})$$

$$K_{\mathrm{sparse}} = S K_{\mathrm{psi,psi}} S^T = K_{\mathrm{obs,psi}} K^{-1}_{\mathrm{psi,psi}} K_{\mathrm{psi,obs}}$$

Diagonal correction (essential for continuous optimization)

$$K_{\mathrm{sparse}} = diag[K_{\mathrm{obs,obs}} - K_{\mathrm{obs,psi}} K^{-1}_{\mathrm{psi,psi}} K_{\mathrm{psi,obs}}] + K_{\mathrm{obs,psi}} K^{-1}_{\mathrm{psi,psi}} K_{\mathrm{psi,obs}}$$

E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs", Advances in Neural Information Processing Systems 18, Cambridge, Massachussets, 2006

J. Quinonero-Candela and C.E. Rasmussen, "A Unifying View of Sparse Approximate Gaussian Process Regression", Journal of Machine Learning 6 (2005), 1939-1959

Joint optimization

Efficient computation of objective function:

O(m^2n) instead of **O**(n^3) with **m** pseudo-inputs and **n** observations

$$\log \rho(\mathcal{D} \,|\, \vec{p}_0, \vec{\sigma}_{\text{exp}}, M) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|M| - \frac{1}{2}(\vec{\sigma}_{\text{exp}} - S\vec{p}_0)^T M^{-1}(\vec{\sigma}_{\text{exp}} - S\vec{p}_0)$$

<u>Scenario</u>

300 pseudo-input points (1500 parameters)

15 parameters in covariance function (a.k.a hyperparameters)

9287 experiment data points

<u>Timings</u>

Objective function: 1.3 sec (4 cores: 0.5 sec)

Gradient wrt hyperpars & pseudo-inputs: 50 sec (4 cores: 17 sec)

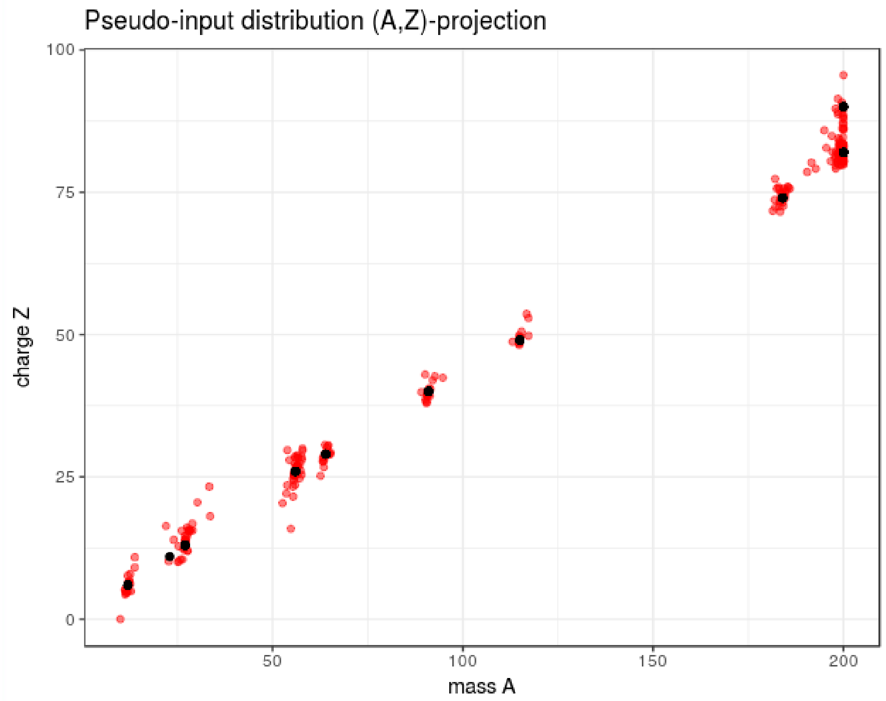
<u>Optimization on cluster</u>

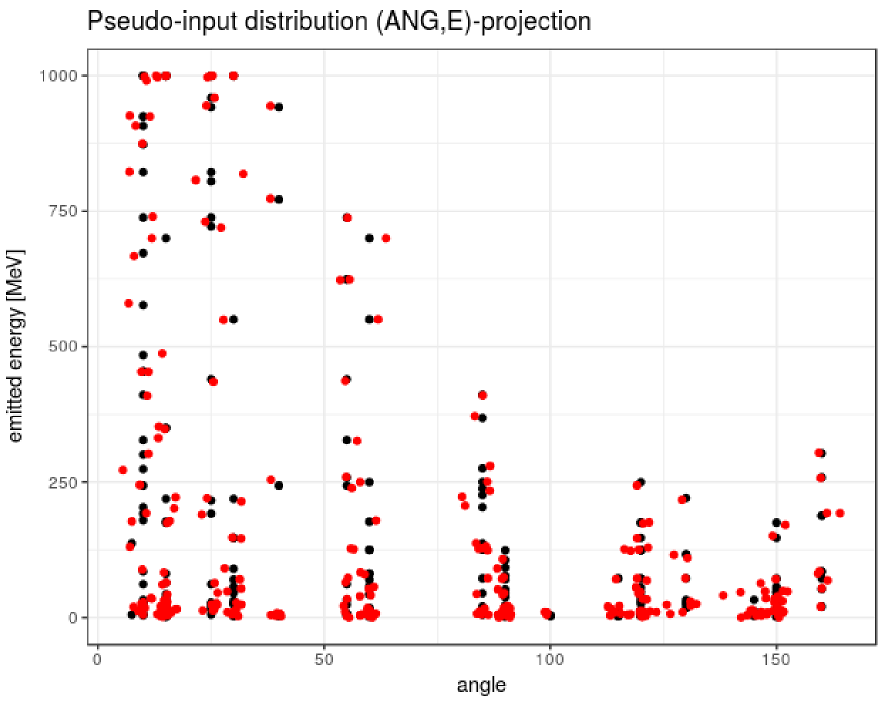
3500 iterations with L-BFGS-B algorithm in 10 hours

using 25 cores (inefficiency: distributed memory)

X^2 / n = 1.03

Pseudo-Inputs & Hyperpars



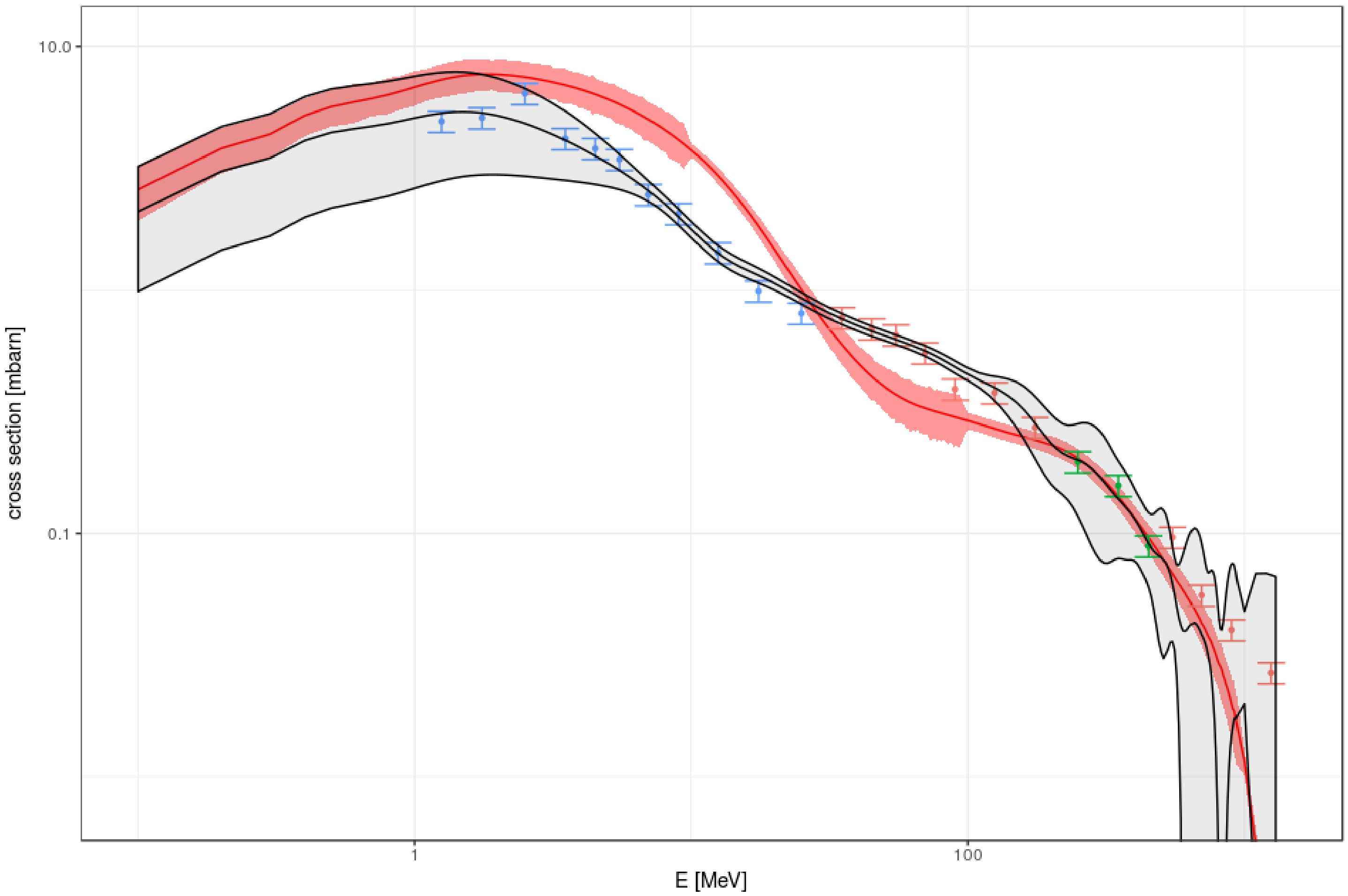


	δ	λ_{EN}	λ_A	λ_Z	λ_{ANG}	λ_E
κ_1	0.5	99	**103**	**41**	68	5
κ_2	0.3	272	**115**	**49**	64	42

τ: $k = 0.3$, $x_0 = 2.7$

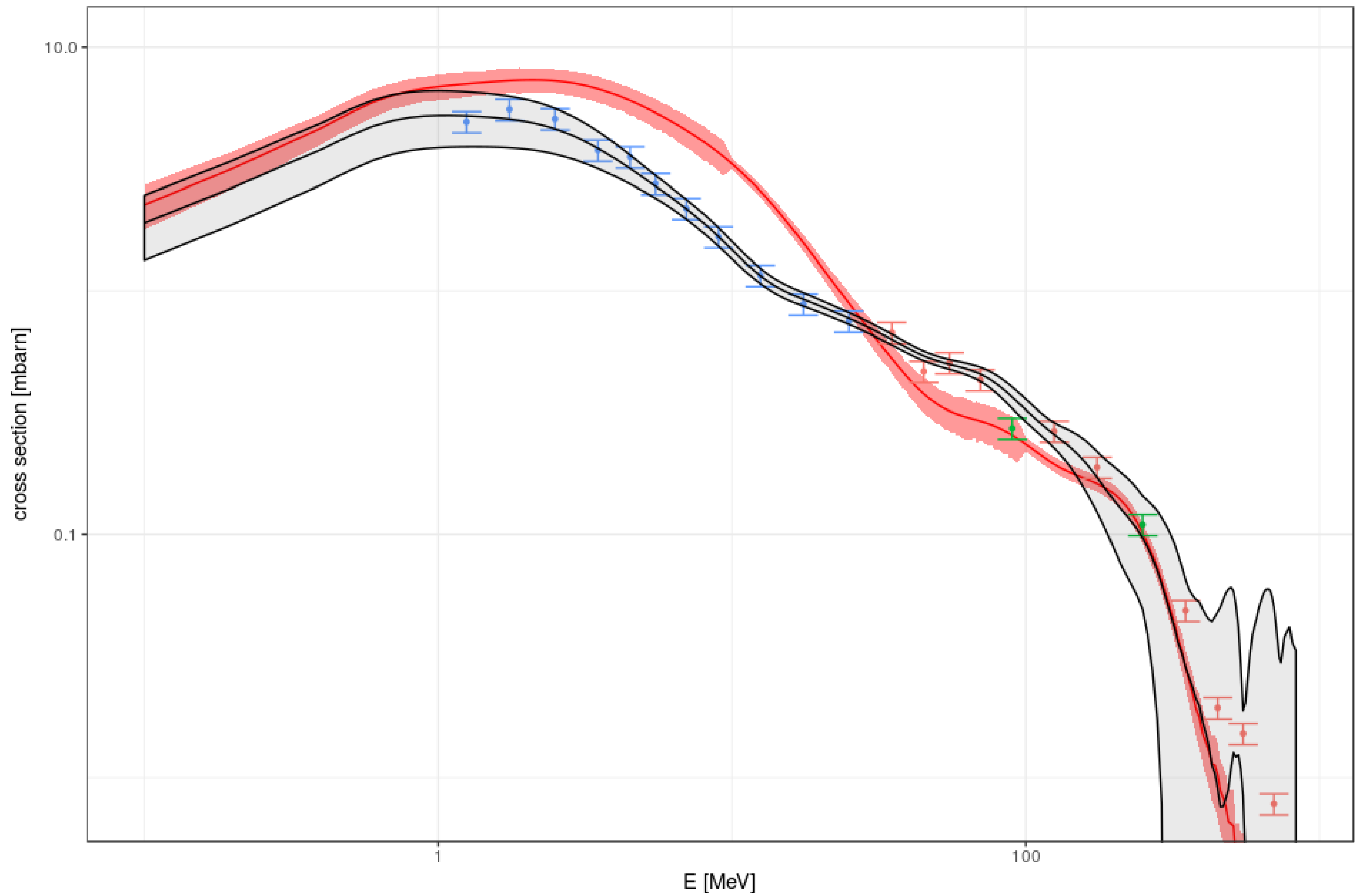
GP prediction

P + Al27 → X + n EN: 1500 MeV ANG: **30** deg



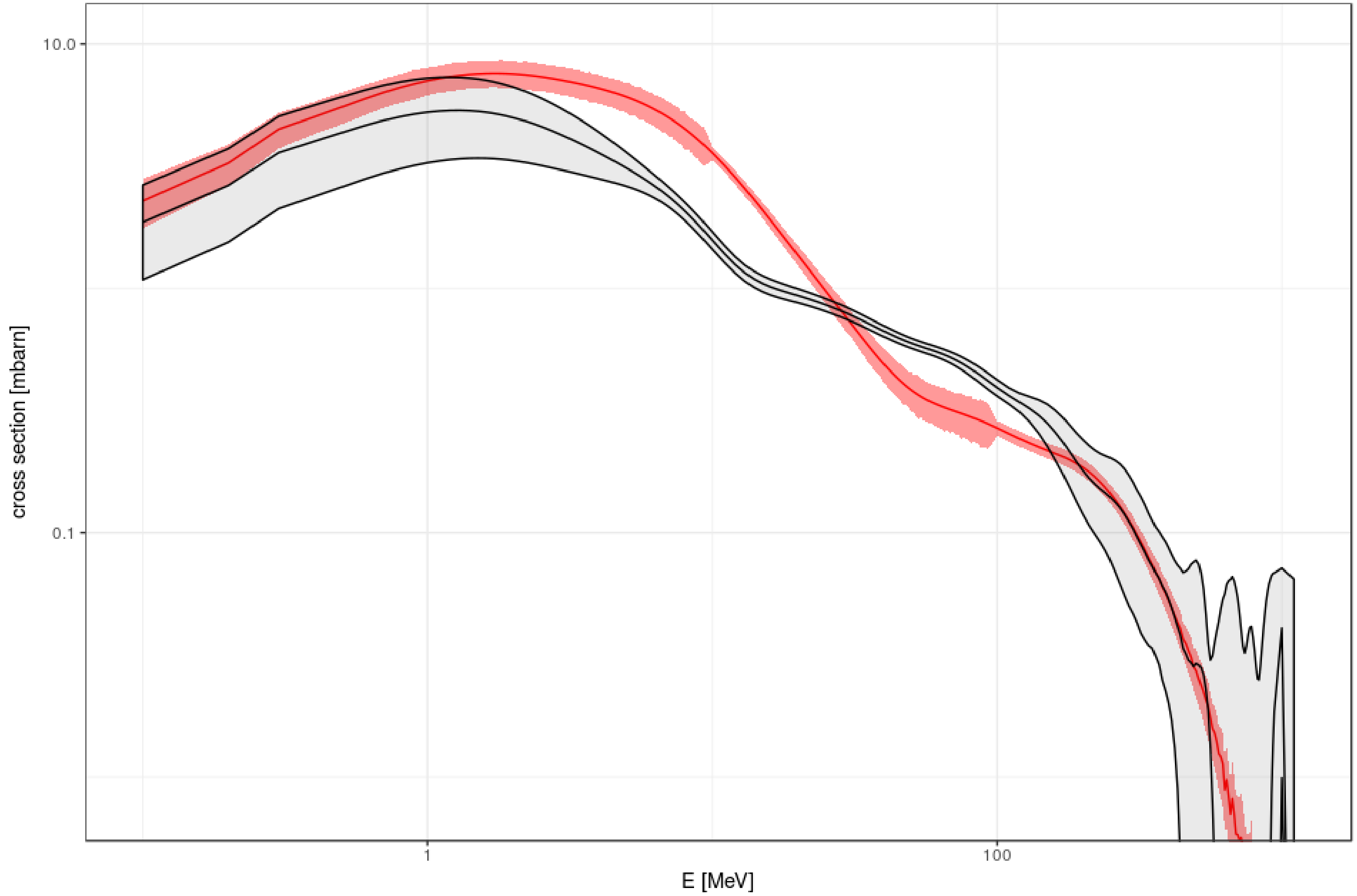
GP prediction

P + Al27 → X + n   EN: 1500 MeV   ANG: 6**0** deg

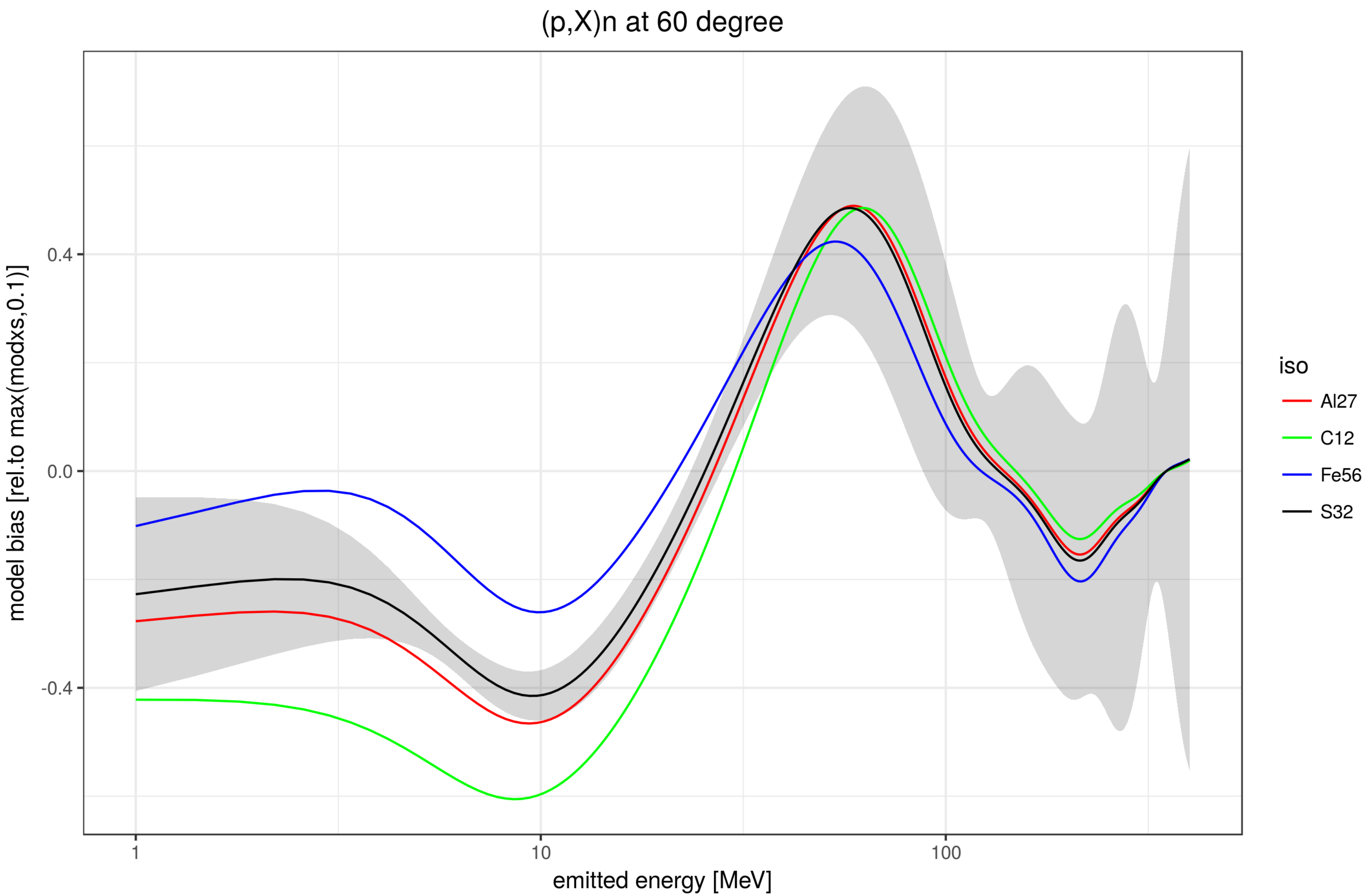


Interpolation between angles

P + Al27 → X + n   EN: 1500 MeV   ANG: **45** deg

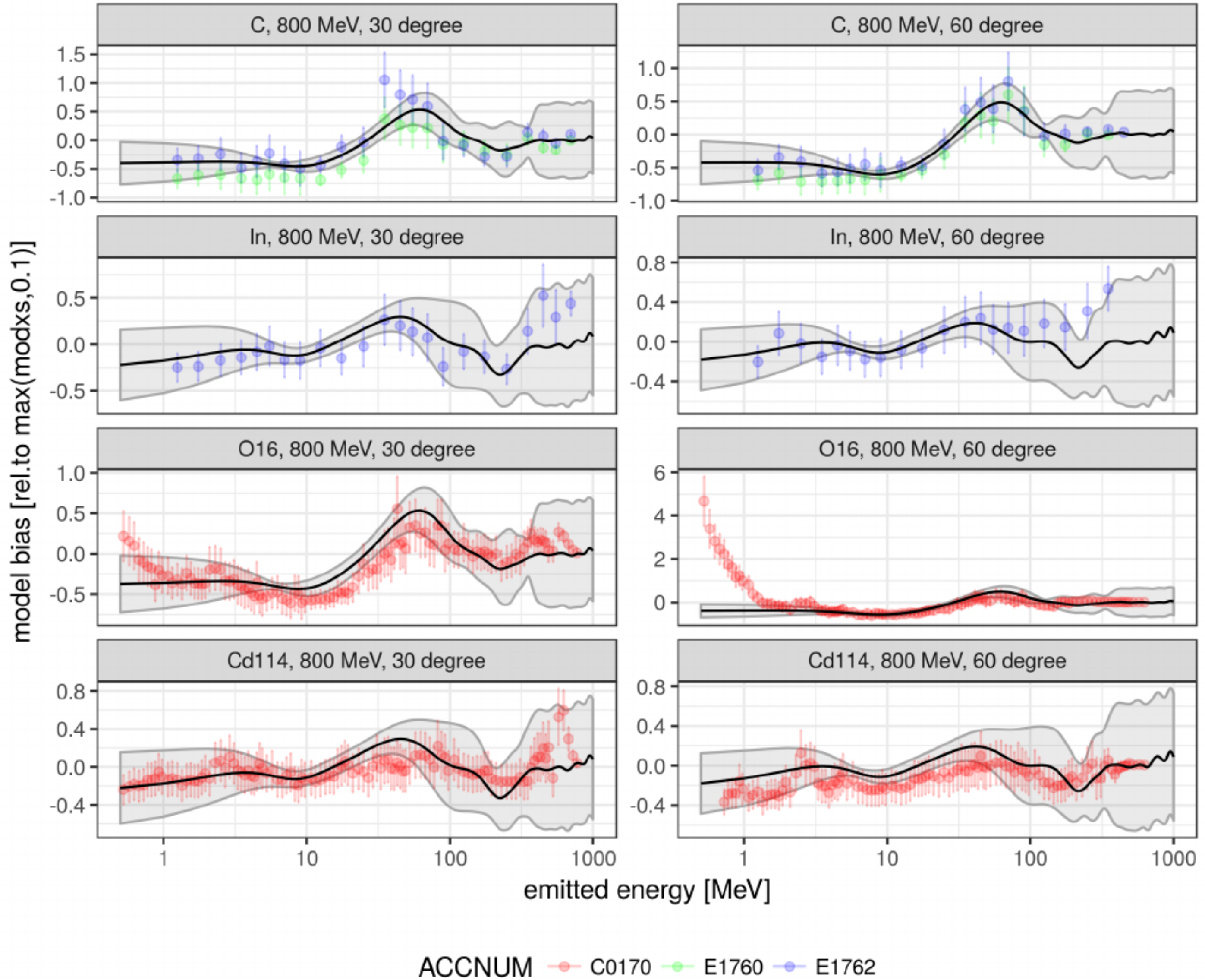


Extrapolation to other isotopes

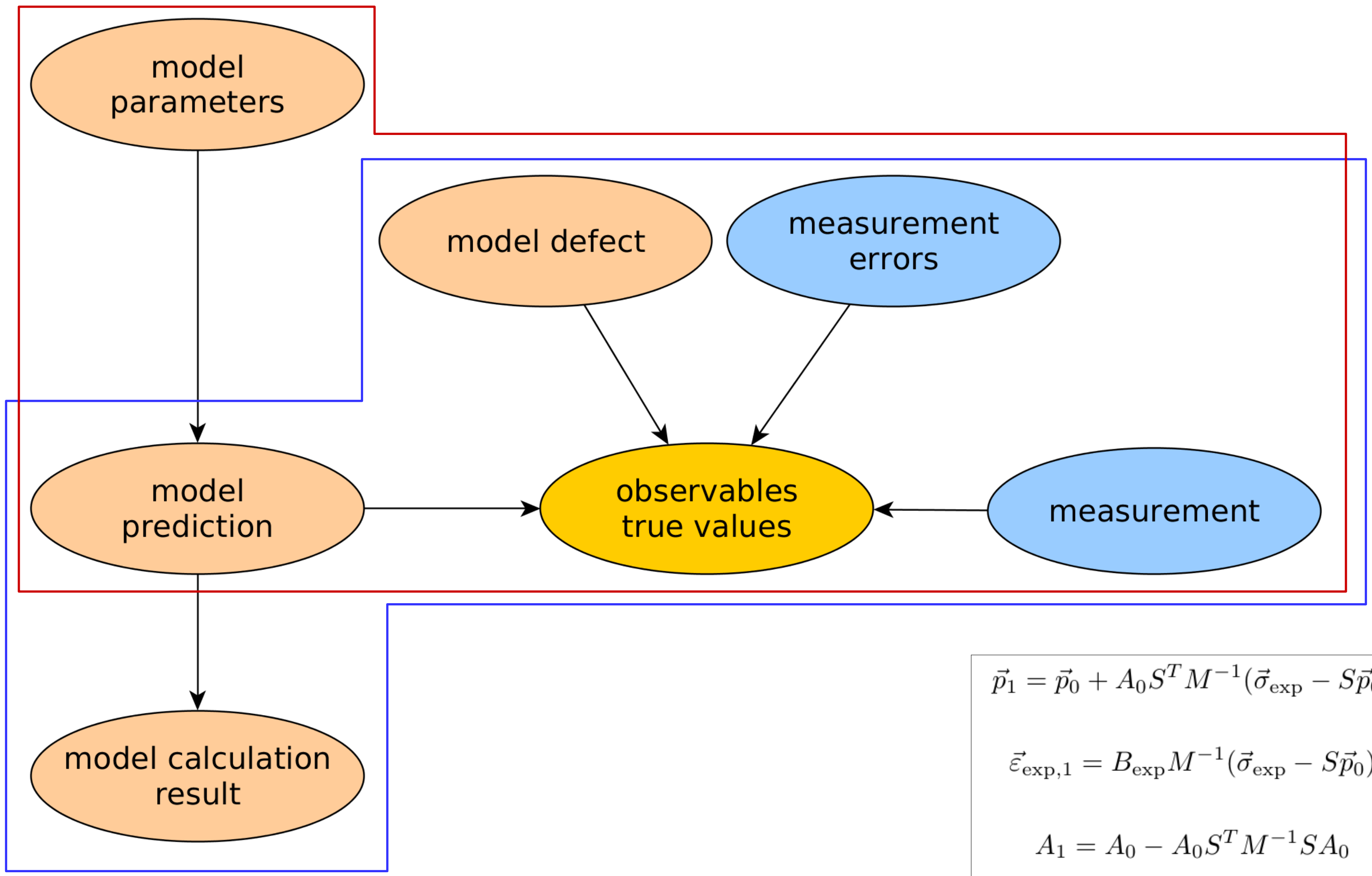


Discussion of extrapolation

Fig. 2. Model bias of INCL in the (p,X)n double differential spectra for 800 MeV incident protons and different isotopes as predicted by GP regression. A missing mass number behind the isotope symbol indicates natural composition. The uncertainty band of the prediction and the error bars of the experiment data denote the 2σ confidence interval. Carbon and indium were taken into account in the GP regression but not cadmium and oxygen. The experiment data is colored according to the associated access number in the EXFOR database. This shows that all displayed data come from just three experiments.



Status quo



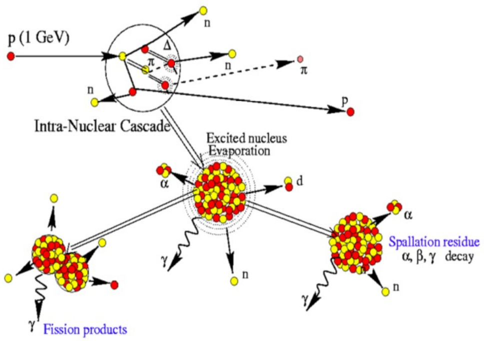
$$\vec{p}_1 = \vec{p}_0 + A_0 S^T M^{-1}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

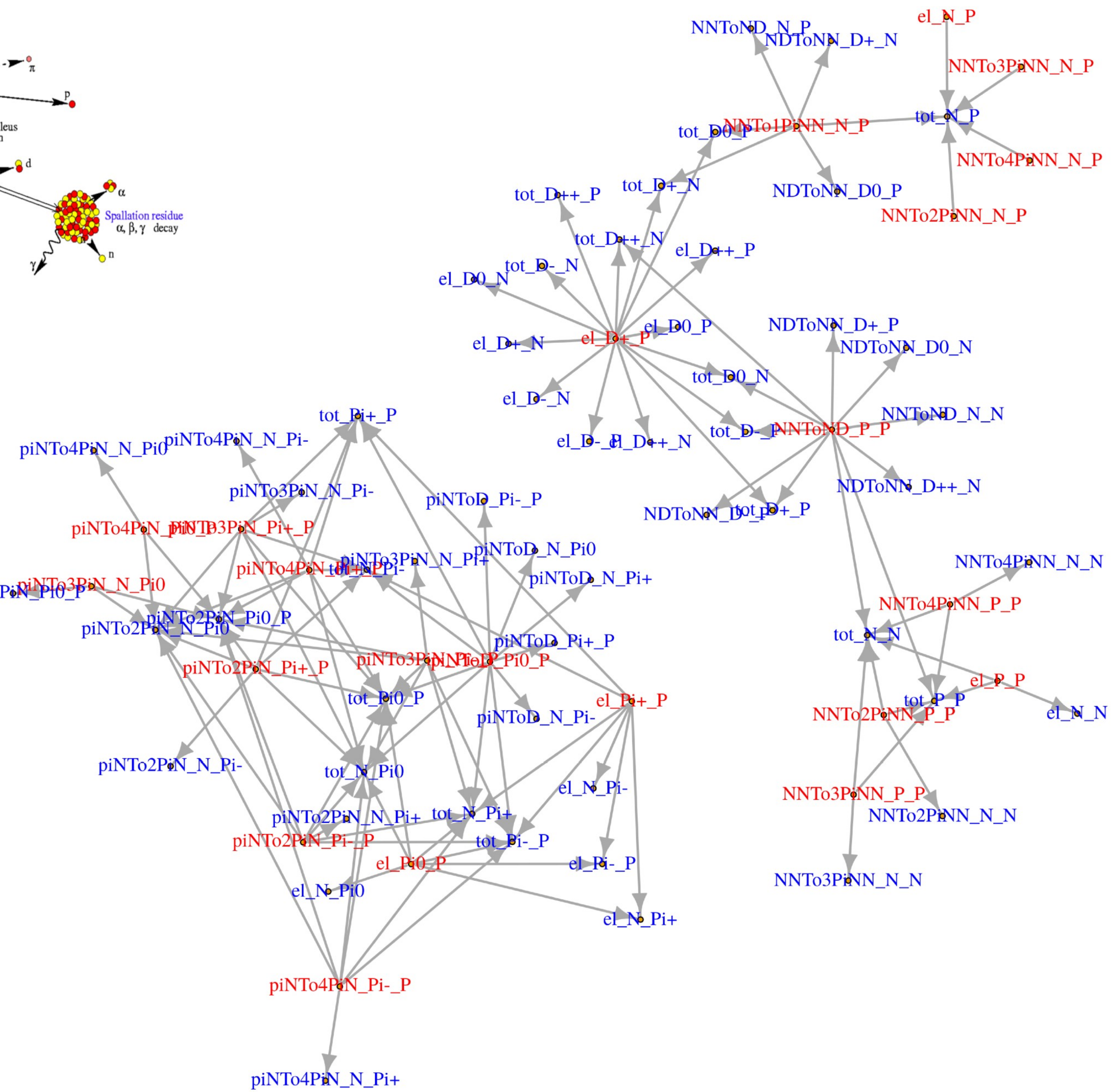
$$\vec{\varepsilon}_{\mathrm{exp},1} = B_{\mathrm{exp}} M^{-1}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

$$A_1 = A_0 - A_0 S^T M^{-1} S A_0$$

$$B_{\mathrm{exp},1} = B_{\mathrm{exp}} - B_{\mathrm{exp}} M^{-1} B_{\mathrm{exp}}$$

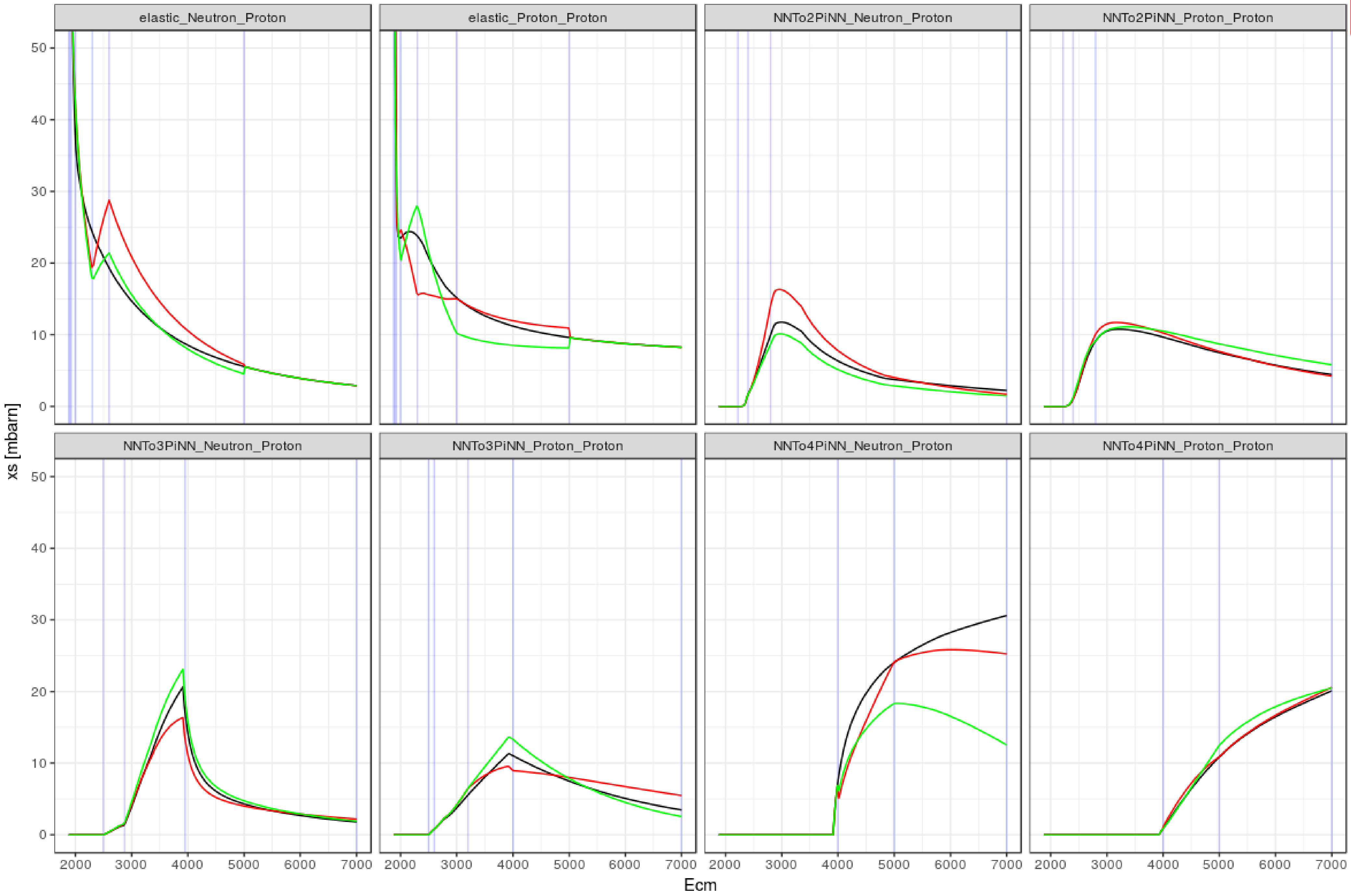
Cross sections & isospin





$$M = SA_0(\delta_1, \delta_2, \ldots)S^T + K_{\mathrm{def}}(\lambda_1, \lambda_2, \ldots) + B_{\mathrm{exp}}$$

Pulling the strings



From deterministic to stochastic

Marginal Likelihood for deterministic linear model

$$\log \rho(\vec{\sigma}_{\mathrm{exp}} \,|\, \vec{p}_0, S, M) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|M| - \frac{1}{2}(\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)^T M^{-1} (\vec{\sigma}_{\mathrm{exp}} - S\vec{p}_0)$$

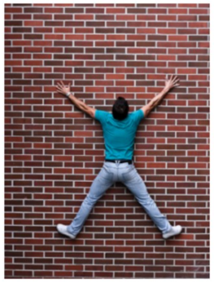
e.g. 10^4 x 10^4 matrix

Marginal Likelihood for stochastic linear model

$$\rho(\vec{\sigma}_{\mathrm{exp}} \,|\, \vec{p}_0, M) = \int \rho(\vec{\sigma}_{\mathrm{exp}} \,|\, \vec{p}_0, S, M)\rho(S)\, dS$$

Challenge

Likely, no analytic solution of integral



Size of N x M matrix **S** with

N … number of experimental data points
M … number of model parameters

equals number of integration variables

For inclusive neutron DDX: 200.000 integration variables

Work ahead and outlook

<u>Methodological</u>

- Complete framework for stochastic linear models
- Investigate the propagation of <u>model bias</u> through simulations
- Conceive a Monte Carlo algorithm for non-linearity

<u>Practical</u>

- Include other reaction data from EXFOR (e.g. isotope production, cumulative xs)
- Use the approach on other model parameters (e.g. potentials)
- Propagate found uncertainties through a transport code

Common sense inference

Hypothesis

- Compute cluster maintenance
- Network cable not plugged in
- Irfu intranet down
- Password expired

Observation

- Can connect to compute cluster
- Cannot connect to compute cluster

$$\frac{\text{If } A \text{ is true, then } B \text{ is true} \qquad B \text{ is true}}{\text{Therefore, } A \text{ becomes more plausible}}$$

CC-IN2P3