Towards a holistic framework for global assessments of nuclear models

Georg Schnabel
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, uncertainty quantification, uncertainty propagation
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


"Toy" scenario

**Data: EXFOR database**

- Database as of: 2017-04-03

<table>
<thead>
<tr>
<th>Number of ENTRY</th>
<th>21,574</th>
<th>experimental works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SUBENT</td>
<td>150,976</td>
<td>data tables</td>
</tr>
<tr>
<td>Number of Datasets</td>
<td>167,857</td>
<td>data tables of reactions</td>
</tr>
<tr>
<td>Number of Datapoints</td>
<td><strong>14,739,297</strong></td>
<td>total number of data points</td>
</tr>
</tbody>
</table>

**Model: INCL/ABLA**

- Features:
  - Stochastic output
  - Computational expensive
  - Many parameters
  - Large output
Automatization

Example of ML approach

Bayesian statistics

\[
P(H \mid O) = \frac{P(O \mid H)P(H)}{P(O)}
\]

H hypothesis
O observation

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

\[P(O \mid H)\] probability of observation O to occur if hypothesis H is true

\[P(H \mid O)\] probability of hypothesis after we observed O

Consistent with Aristotelian logic
Consistent with principles of common sense

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

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

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

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

\[ A_1 = A_0 - A_0 S^T M^{-1} S A_0 \]

\[ B_{\text{exp,1}} = B_{\text{exp}} - B_{\text{exp}} M^{-1} B_{\text{exp}} \]
Deterministic codes

\[ M = SA_0S^T + B_{\text{exp}} + K_{\text{def}} \]

\[ \tilde{p}_1 = \tilde{p}_0 + A_0S^T M^{-1}(\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0) \]

\[ \tilde{e}_{\text{exp,1}} = B_{\text{exp}} M^{-1}(\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0) \]

\[ A_1 = A_0 - A_0S^T M^{-1}SA_0 \]

\[ B_{\text{exp,1}} = B_{\text{exp}} - B_{\text{exp}}M^{-1}B_{\text{exp}} \]
Stochastic codes

\[ M = P_{\text{stat}} + SA_0S^T + B_{\text{exp}} + K_{\text{def}} \]

- Model parameters \( A_0 \)
- Model defect \( K_{\text{def}} \)
- Measurement errors \( B_{\text{exp}} \)
- Observables true values
- Measurements \( \tilde{\sigma}_{\text{exp}} \)
- Model predictions
- Model calculation results

\[
\begin{align*}
\tilde{p}_1 &= \tilde{p}_0 + A_0S^TM^{-1}(\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0) \\
\tilde{\sigma}_{\text{exp},1} &= B_{\text{exp}}M^{-1}(\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0) \\
A_1 &= A_0 - A_0S^TM^{-1}SA_0 \\
B_{\text{exp},1} &= B_{\text{exp}} - B_{\text{exp}}M^{-1}B_{\text{exp}}
\end{align*}
\]
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

\[
\log \rho(\bar{\sigma}_{\text{exp}} \mid \bar{p}_0, S, M) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |M| - \frac{1}{2} (\bar{\sigma}_{\text{exp}} - S\bar{p}_0)^T M^{-1} (\bar{\sigma}_{\text{exp}} - S\bar{p}_0)
\]

Entropy

\[
M = SA_0 S^T + B_{\text{exp}}(\lambda_1, \lambda_2, \ldots)
\]

Questions

Is it computationally feasible?

Can we efficiently maximize this expression?
Imperfect model

\[ M = P_{\text{stat}} + K_{\text{def}}(\lambda_1, \lambda_2, \ldots) + B_{\text{exp}} \]
(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.

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

Observations \((\bar{y}_{\text{exp}}, \bar{x}_{\text{exp}})\)

\[ \bar{p} = \begin{pmatrix} k \\ d \end{pmatrix} = AS^T (SAS^T + B)^{-1} \bar{y}_{\text{exp}} \]

\[ \bar{y}_{\text{pred}} = S_{\text{pred}} \bar{p} = \begin{pmatrix} \bar{x}_{\text{pred}} \\ 1 \end{pmatrix} \begin{pmatrix} k \\ d \end{pmatrix} \]

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

\[ K_{\text{pred,exp}} = \kappa(\bar{x}_{\text{pred}}, \bar{x}_{\text{exp}}) \quad K_{\text{exp,exp}} = \kappa(\bar{x}_{\text{exp}}, \bar{x}_{\text{exp}}) \]

\[ \bar{y}_{\text{pred}} = K_{\text{pred,exp}} K_{\text{exp,exp}}^{-1} \bar{y}_{\text{exp}} \]
Gaussian processes

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

\[ \kappa(x_1, x_2) = \delta^2 \exp\left(-\frac{(x_1 - x_2)^2}{2\lambda^2}\right) \]

Sample from prior (\(\delta=\lambda=1\))

Sample from posterior
Gaussian processes

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

\[ \kappa(x_1, x_2) = \delta^2 \exp \left( -\frac{(x_1 - x_2)^2}{2\lambda^2} \right) \]

Sample from prior ($\delta=\lambda=1$)

Posterior uncertainty
Comparison to 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

\[ M = P_{\text{stat}} + K_{\text{def}}(\lambda_1, \lambda_2, \ldots) + B_{\text{exp}} \]
<table>
<thead>
<tr>
<th>Isotope</th>
<th>En</th>
<th>$E_{\text{min}}$</th>
<th>$E_{\text{max}}$</th>
<th>$\theta_{\text{min}}$</th>
<th>$\theta_{\text{max}}$</th>
<th>NumPts</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>800</td>
<td>1.2</td>
<td>700</td>
<td>15</td>
<td>150</td>
<td>189</td>
</tr>
<tr>
<td>C</td>
<td>1500</td>
<td>1.2</td>
<td>1250</td>
<td>15</td>
<td>150</td>
<td>245</td>
</tr>
<tr>
<td>C</td>
<td>3000</td>
<td>1.2</td>
<td>2500</td>
<td>15</td>
<td>150</td>
<td>128</td>
</tr>
<tr>
<td>Na23</td>
<td>800</td>
<td>3.5</td>
<td>266</td>
<td>30</td>
<td>150</td>
<td>84</td>
</tr>
<tr>
<td>Al27</td>
<td>800</td>
<td>1.2</td>
<td>700</td>
<td>15</td>
<td>150</td>
<td>119</td>
</tr>
<tr>
<td>Al27</td>
<td>1000</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>223</td>
</tr>
<tr>
<td>Al27</td>
<td>1200</td>
<td>2.0</td>
<td>1189</td>
<td>10</td>
<td>160</td>
<td>404</td>
</tr>
<tr>
<td>Al27</td>
<td>1500</td>
<td>1.2</td>
<td>1250</td>
<td>15</td>
<td>150</td>
<td>129</td>
</tr>
<tr>
<td>Al27</td>
<td>1600</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>226</td>
</tr>
<tr>
<td>Al27</td>
<td>3000</td>
<td>1.2</td>
<td>2500</td>
<td>15</td>
<td>150</td>
<td>132</td>
</tr>
<tr>
<td>Fe</td>
<td>800</td>
<td>1.2</td>
<td>771</td>
<td>10</td>
<td>160</td>
<td>505</td>
</tr>
<tr>
<td>Fe</td>
<td>1200</td>
<td>2.0</td>
<td>1171</td>
<td>10</td>
<td>160</td>
<td>417</td>
</tr>
<tr>
<td>Fe</td>
<td>1500</td>
<td>1.2</td>
<td>1250</td>
<td>15</td>
<td>150</td>
<td>129</td>
</tr>
<tr>
<td>Fe</td>
<td>1600</td>
<td>2.0</td>
<td>1572</td>
<td>10</td>
<td>160</td>
<td>460</td>
</tr>
<tr>
<td>Fe</td>
<td>3000</td>
<td>1.2</td>
<td>2500</td>
<td>15</td>
<td>150</td>
<td>133</td>
</tr>
<tr>
<td>Cu</td>
<td>1000</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>227</td>
</tr>
<tr>
<td>Cu</td>
<td>1600</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>231</td>
</tr>
<tr>
<td>Zr</td>
<td>1000</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>229</td>
</tr>
<tr>
<td>Zr</td>
<td>1200</td>
<td>2.0</td>
<td>1189</td>
<td>10</td>
<td>160</td>
<td>423</td>
</tr>
<tr>
<td>Zr</td>
<td>1600</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>229</td>
</tr>
<tr>
<td>In</td>
<td>800</td>
<td>1.2</td>
<td>700</td>
<td>15</td>
<td>150</td>
<td>116</td>
</tr>
<tr>
<td>In</td>
<td>1500</td>
<td>1.2</td>
<td>1250</td>
<td>15</td>
<td>150</td>
<td>128</td>
</tr>
<tr>
<td>In</td>
<td>3000</td>
<td>1.2</td>
<td>2500</td>
<td>15</td>
<td>150</td>
<td>133</td>
</tr>
<tr>
<td>W</td>
<td>800</td>
<td>3.1</td>
<td>333</td>
<td>30</td>
<td>150</td>
<td>110</td>
</tr>
<tr>
<td>W</td>
<td>1000</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>231</td>
</tr>
<tr>
<td>W</td>
<td>1200</td>
<td>2.0</td>
<td>1189</td>
<td>10</td>
<td>160</td>
<td>413</td>
</tr>
<tr>
<td>W</td>
<td>1600</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>231</td>
</tr>
<tr>
<td>Pb</td>
<td>318</td>
<td>5.4</td>
<td>356</td>
<td>7</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>Pb</td>
<td>800</td>
<td>1.2</td>
<td>771</td>
<td>10</td>
<td>160</td>
<td>624</td>
</tr>
<tr>
<td>Pb</td>
<td>1000</td>
<td>2.5</td>
<td>280</td>
<td>15</td>
<td>150</td>
<td>231</td>
</tr>
<tr>
<td>Pb</td>
<td>1200</td>
<td>2.0</td>
<td>1189</td>
<td>10</td>
<td>160</td>
<td>563</td>
</tr>
<tr>
<td>Pb</td>
<td>1500</td>
<td>1.2</td>
<td>1250</td>
<td>15</td>
<td>150</td>
<td>249</td>
</tr>
<tr>
<td>Pb</td>
<td>1600</td>
<td>2.0</td>
<td>1591</td>
<td>10</td>
<td>160</td>
<td>691</td>
</tr>
<tr>
<td>Pb</td>
<td>3000</td>
<td>1.2</td>
<td>2500</td>
<td>15</td>
<td>150</td>
<td>131</td>
</tr>
<tr>
<td>Pb208</td>
<td>2000</td>
<td>0.4</td>
<td>402</td>
<td>30</td>
<td>150</td>
<td>170</td>
</tr>
<tr>
<td>Th232</td>
<td>1200</td>
<td>2.0</td>
<td>1189</td>
<td>10</td>
<td>160</td>
<td>351</td>
</tr>
</tbody>
</table>
Covariance brewing

Combination rules for covariance functions

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

\[ \kappa_{1\times2}(x_1, x_2) = \kappa_1(x_1, x_2) \times \kappa_2(x_1, x_2) \]

Squared exponential covariance functions

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

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

Transition kernel

\[ \tau_1(x_1, x_2) = \sigma(x_1)\sigma(x_2) \]

\[ \tau_2(x_1, x_2) = (1 - \sigma(x_1))(1 - \sigma(x_2)) \]

\[ \sigma(x) = \frac{1}{1 + \exp(-k(x - x_0))} \]

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

\[ \delta_1 = 5, \lambda_1 = 5, \delta_2 = 1, \lambda_2 = 1, k = 0.5, x_0 = 10 \]

\[ \delta_1 = 5, \lambda_1 = 500, \delta_2 = 1, \lambda_2 = 1, k = 10, x_0 = 10 \]
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)

\begin{align*}
k_1: \delta_1, \lambda_{11}, \ldots, \lambda_{15} \\
k_2: \delta_2, \lambda_{21}, \ldots, \lambda_{25} \\
T: k, x_0, \Phi
\end{align*}

\[
M = P_{\text{stat}} + K_{\text{def}}(\lambda_1, \lambda_2, \ldots) + B_{\text{exp}}
\]

\[
\log \rho(D | \tilde{p}_0, \tilde{\sigma}_{\exp}, M) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |M| - \frac{1}{2} (\tilde{\sigma}_{\exp} - S\tilde{p}_0)^T M^{-1} (\tilde{\sigma}_{\exp} - S\tilde{p}_0)
\]

\[
\kappa_{\text{comp}}(x_1, x_2) = \tau_1(x_1, x_2) \times \kappa_1(x_1, x_2) + \tau_2(x_1, x_2) \times \kappa_2(x_1, x_2)
\]
Introduce sparsity

Assume latent variables (pseudo-inputs)

\[
\tilde{y}_{\text{obs}} = K_{\text{obs,psi}} K^{-1}_{\text{psi,psi}} \tilde{y}_{\text{psi}}
\]

\[
\tilde{y}_{\text{psi}} \sim \mathcal{N}(0, K_{\text{psi,psi}})
\]

\[
K_{\text{sparse}} = SK_{\text{psi,psi}} S^T = K_{\text{obs,psi}} K^{-1}_{\text{psi,psi}} K_{\text{psi,obs}}
\]

Diagonal correction (essential for continuous optimization)

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


Joint optimization

Efficient computation of objective function:

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

\[
\log \rho(D \mid \tilde{p}_0, \tilde{\sigma}_{\text{exp}}, M) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |M| - \frac{1}{2} (\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0)^T M^{-1} (\tilde{\sigma}_{\text{exp}} - S\tilde{p}_0)
\]

Scenario

300 pseudo-input points (1500 parameters)
15 parameters in covariance function (a.k.a hyperparameters)
9287 experiment data points

Timings

Objective function: 1.3 sec (4 cores: 0.5 sec)
Gradient wrt hyperpars & pseudo-inputs: 50 sec (4 cores: 17 sec)

Optimization on cluster

3500 iterations with L-BFGS-B algorithm in 10 hours
using 25 cores (inefficiency: distributed memory)

\( X^2 / n = 1.03 \)
Pseudo-Inputs & Hyperpars

<table>
<thead>
<tr>
<th></th>
<th>$\delta$</th>
<th>$\lambda_{EN}$</th>
<th>$\lambda_A$</th>
<th>$\lambda_Z$</th>
<th>$\lambda_{ANG}$</th>
<th>$\lambda_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_1$</td>
<td>0.5</td>
<td>99</td>
<td><strong>103</strong></td>
<td><strong>41</strong></td>
<td>68</td>
<td>5</td>
</tr>
<tr>
<td>$K_2$</td>
<td>0.3</td>
<td>272</td>
<td><strong>115</strong></td>
<td><strong>49</strong></td>
<td>64</td>
<td>42</td>
</tr>
</tbody>
</table>

$\tau$: $k = 0.3$, $x_0 = 2.7$
GP prediction

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

Diagram showing the cross section as a function of energy (E [MeV]), with data points and error bars. The graph illustrates the predicted behavior of the reaction under specified conditions.
GP prediction

\[ P + Al^{27} \rightarrow X + n \quad EN: 1500 \text{ MeV} \quad ANG: 60 \text{ deg} \]
Interpolation between angles

P + Al27 → X + n

EN: 1500 MeV

ANG: 45 deg
Extrapolation to other isotopes

$(p,X)n$ at 60 degree
**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

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

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

\[
A_1 = A_0 - A_0 S^T M^{-1} S A_0
\]

\[
B_{\text{exp},1} = B_{\text{exp}} - B_{\text{exp}} M^{-1} B_{\text{exp}}
\]
Cross sections & isospin
Pulling the strings

\[ M = SA_0(\delta_1, \delta_2, \ldots)S^T + K_{\text{def}}(\lambda_1, \lambda_2, \ldots) + B_{\text{exp}} \]
From deterministic to stochastic

Marginal Likelihood for deterministic linear model

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

Marginal Likelihood for stochastic linear model

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

Challenge

Likely, no analytic solution of integral

Size of $N \times M$ matrix $S$ with

$N \ldots$ number of experimental data points
$M \ldots$ number of model parameters

equals number of integration variables

For inclusive neutron DDX: $200,000$ integration variables
Work ahead and outlook

Methodological

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

Practical

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

If A is true, then B is true

\[ B \text{ is true} \]

Therefore, A becomes more plausible

CC-IN2P3