IAEA
60 Years
Atoms for Peace and Development

Probabilistic methods for nuclear data

Georg Schnabel
g.schnabel [at] iaea.org

Nuclear Data Section
Division of Physical and Chemical Sciences NAPC
Department for Nuclear Sciences and Applications
IAEA, Vienna

Workshop on Computational Nuclear Science and Engineering

13 July 2021

Short bio

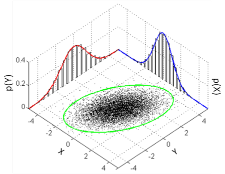
- Studied physics at TU Vienna
- PhD in nuclear data evaluation 2015
- Postdoc at CEA Saclay (2015-2018) and Uppsala University (2018-2019)
- Since 2020 nuclear physicist in Nuclear Data Section at IAEA dealing with nuclear data library projects and code development

Outline

About nuclear data

Bayesian statistics

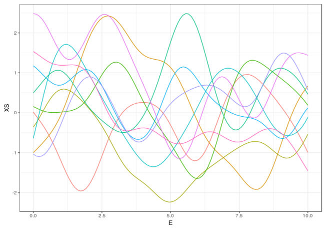
Multivariate normal distribution



Approaches to solve Bayes equation in nuclear data evaluation

Generalized Least Squares

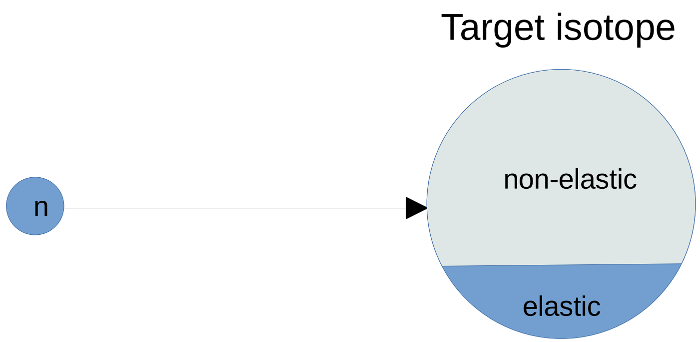
Gaussian processes



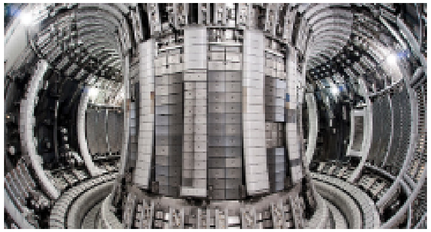
Nuclear data

PSI Gantry 2 facility

Probabilities of various nuclear interactions involving the atomic nuclei, e.g., cross sections.



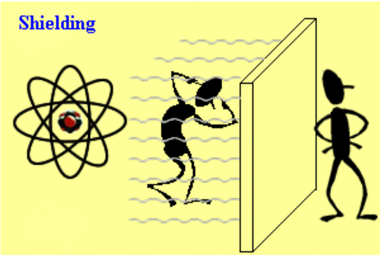
Palisades Nuclear Generating Stations

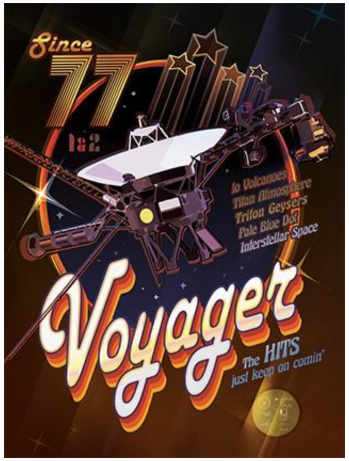


Joint European Torus

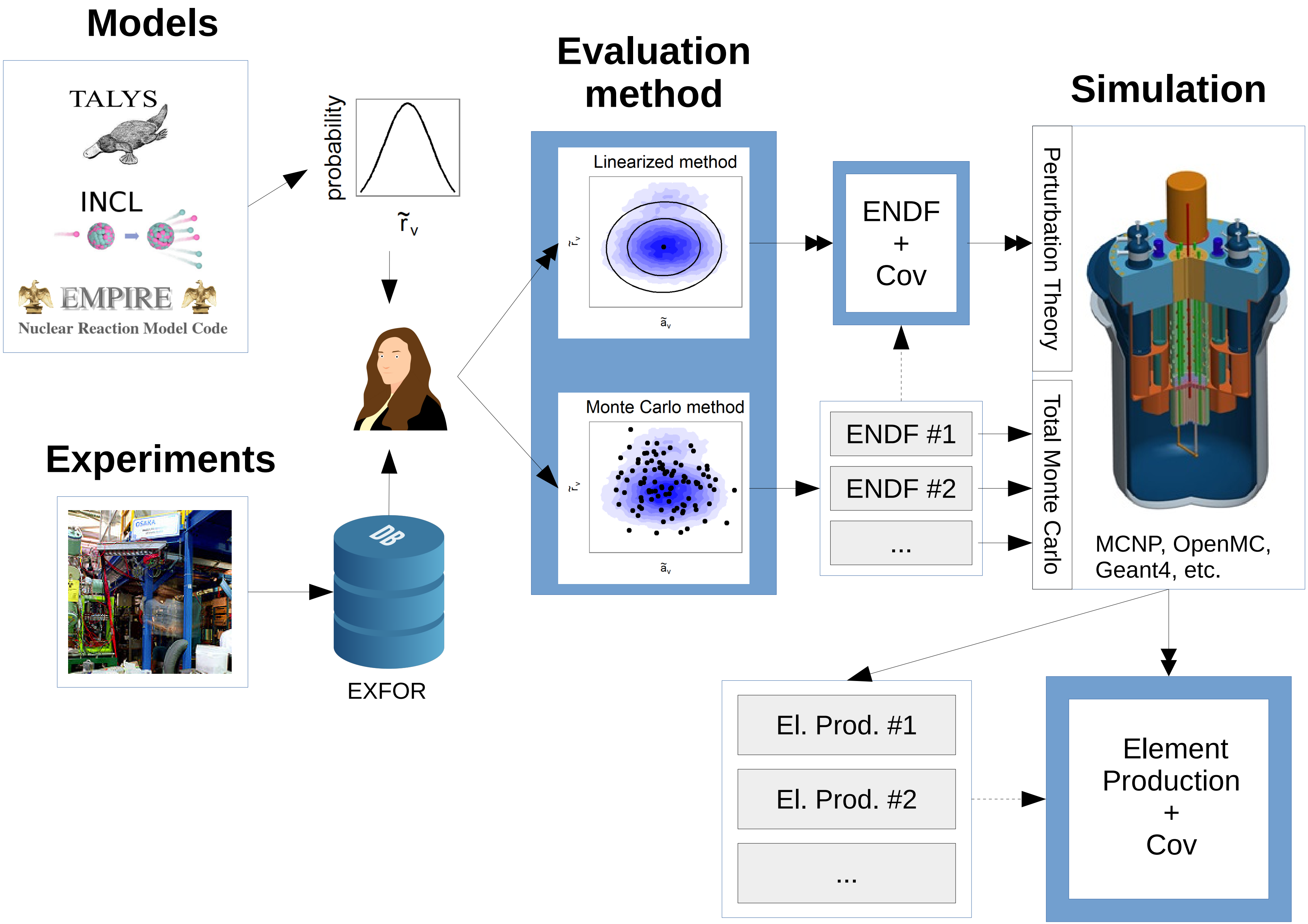
Relevant for:

- Reactor physics
- Radiation dosimetry
- Radiation protection
- Radioactive waste management
- Astrophysics
- Nuclear medicine
- Fusion research
- ...

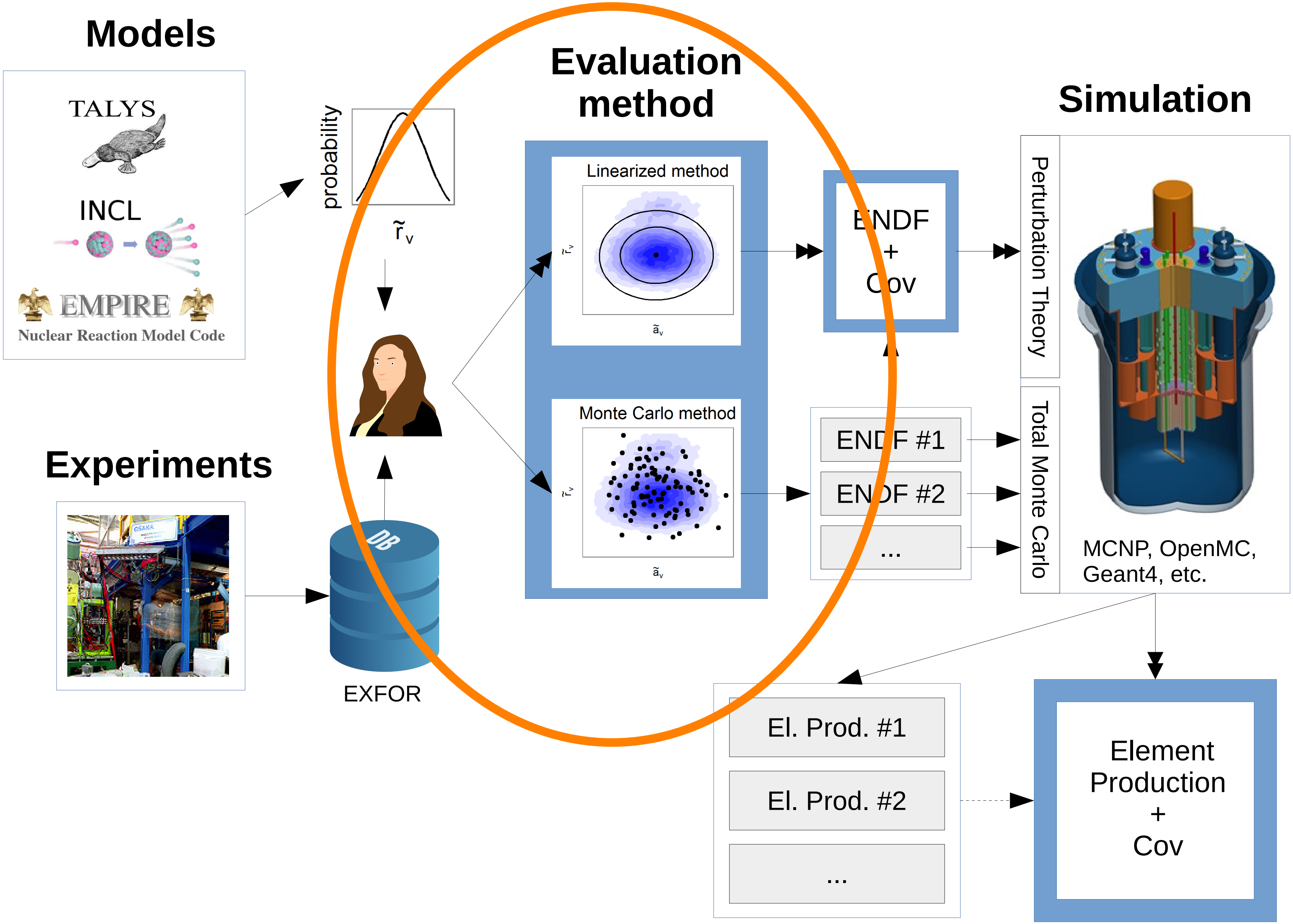




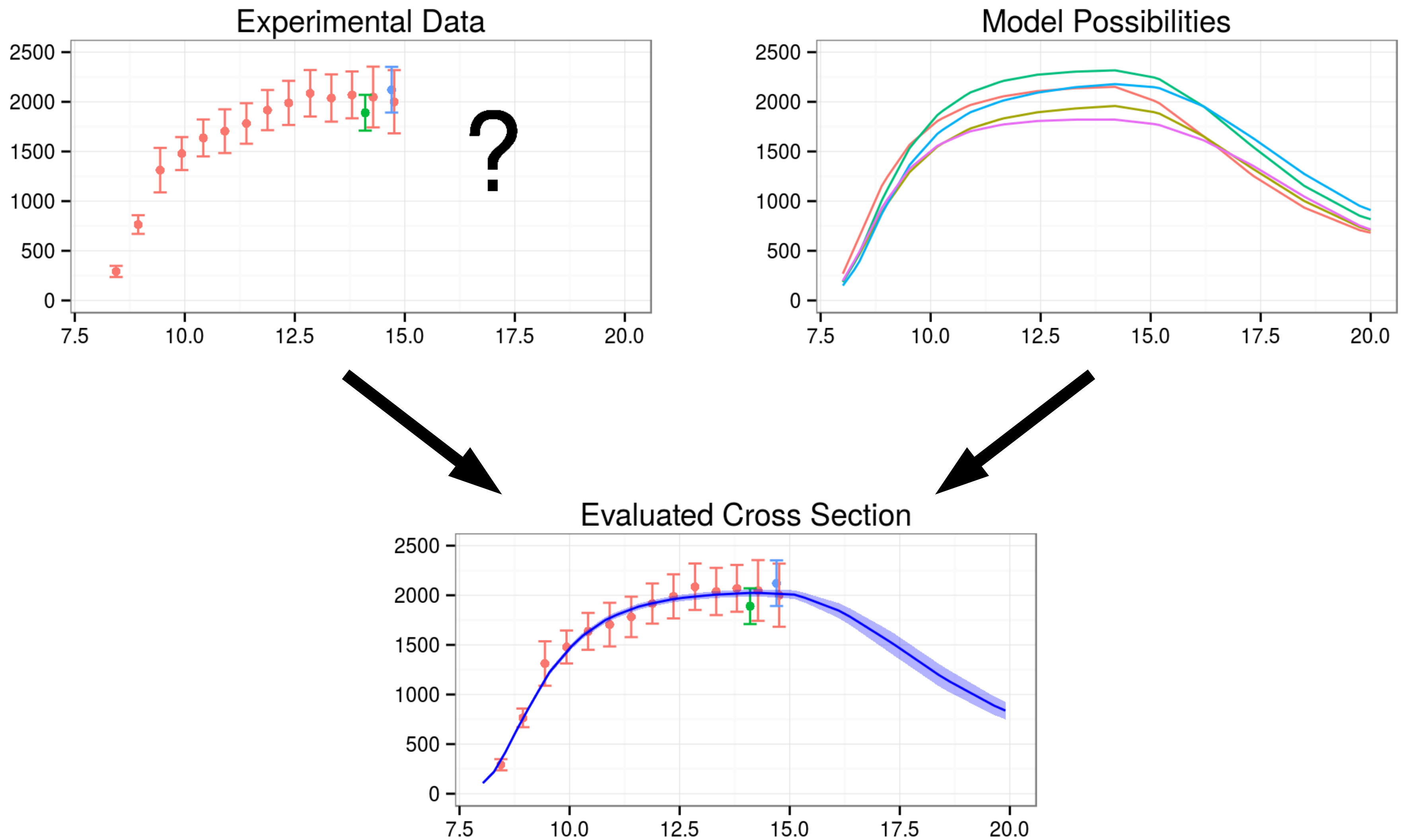
Nuclear data evaluation



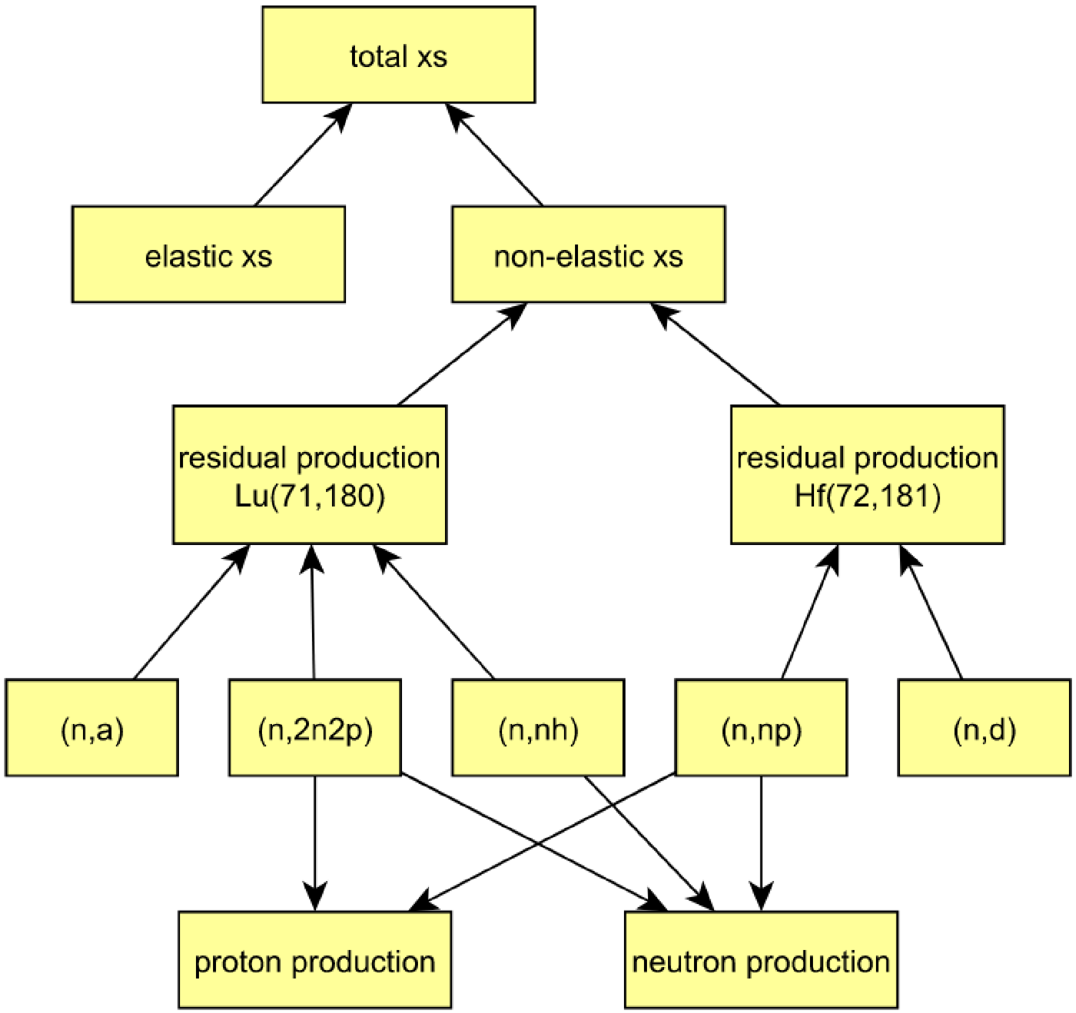
Nuclear data evaluation



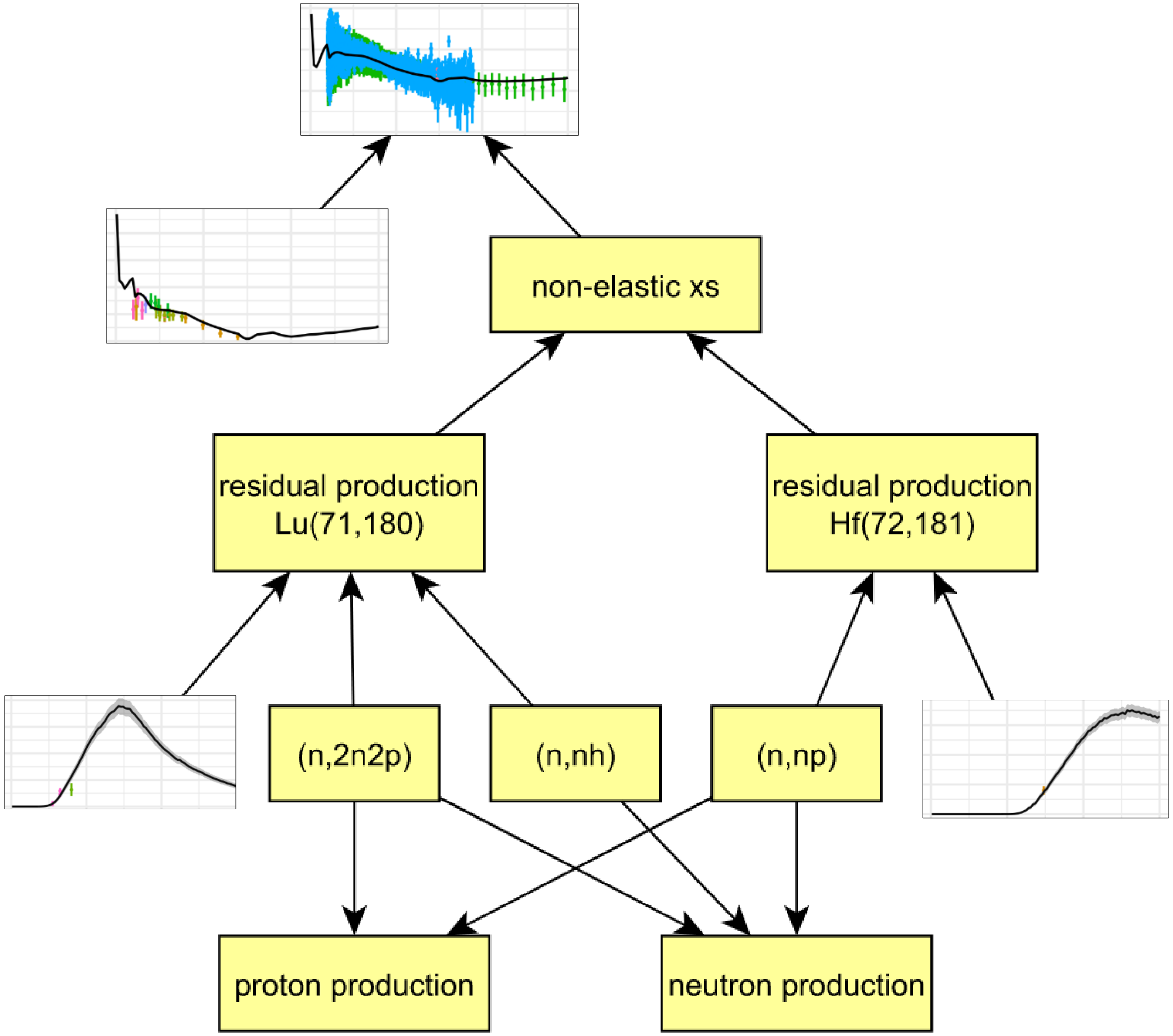
General task setting



System of reactions



System of reactions



Jaynes Robot

Hypotheses

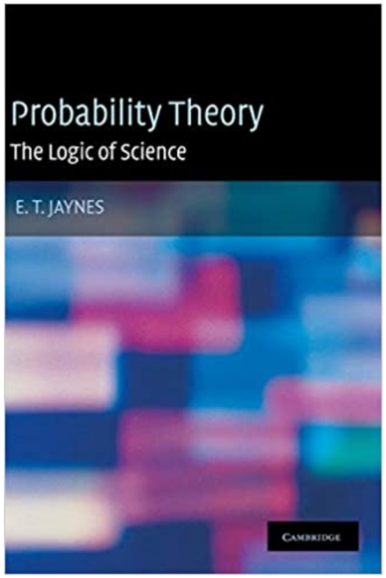
H1: It rained
H2: It did not rain

Observations

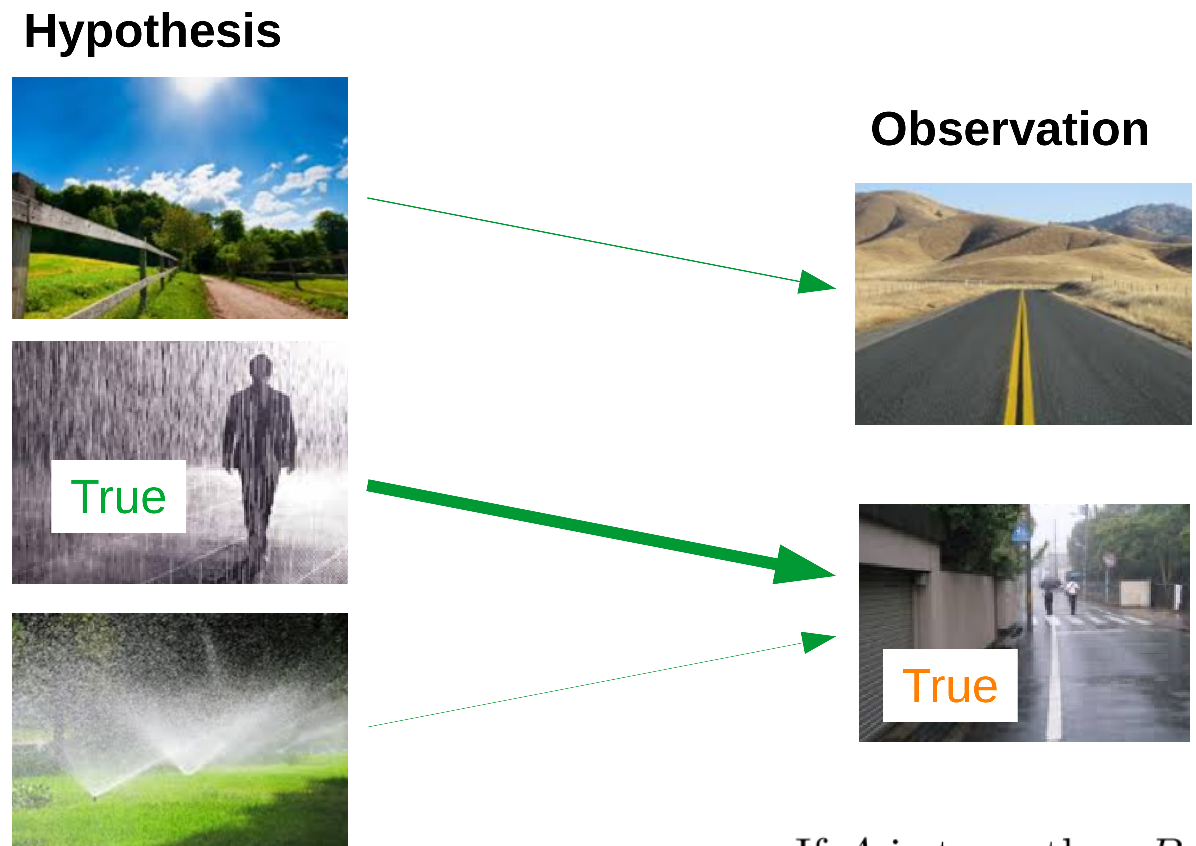
O1: The ground is wet
O2: The ground is dry

Which hypothesis is true?

Edwin Thompson Jaynes
1922-1998



Consistency with Aristotelian logic

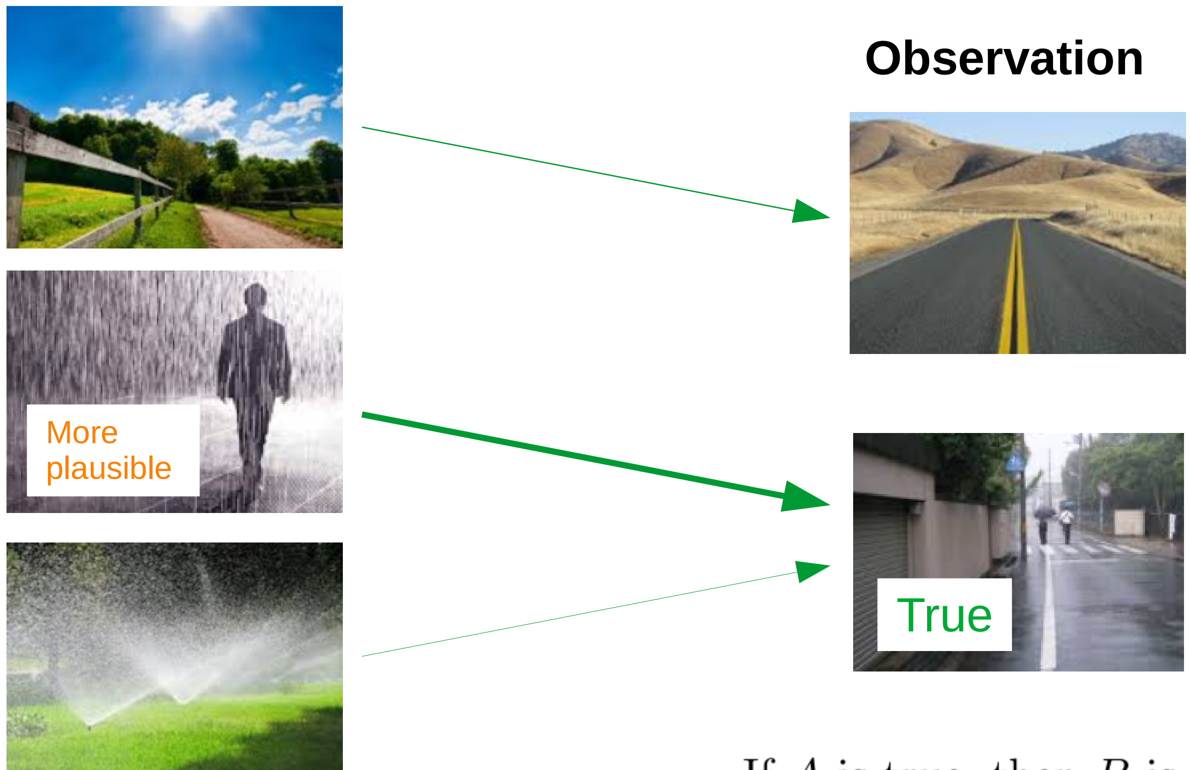


$$\frac{\begin{array}{c}\text{If } A \text{ is true, then } B \text{ is true}\\ A \text{ is true}\end{array}}{\text{Therefore, } B \text{ is true}}$$

Consistency with common sense

Hypothesis

Observation



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

Desiderata

(I) Degrees of Plausibility are represented by real numbers

(II) Qualitative Correspondence with common sense

(IIIa) If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.

(from E.T. Jaynes, Probability Theory: The Logic of Science)

Richard Threlkeld Cox
1898-1991

Desiderata / Cox theorem

(I) Degrees of Plausibility are represented by real numbers

(II) Qualitative Correspondence with common sense

(IIIa) If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.

(from E.T. Jaynes, Probability Theory: The Logic of Science)

Richard Threlkeld Cox
1898-1991

→ **Computation rules of probability theory follow**

e.g., product rule

$$P(H, O) = P(O \mid H) P(H)$$

Bayesian update formula

Thomas Bayes
1701-1761

$$P(H \mid O) \propto P(O \mid H) P(H)$$

Pierre Simon Laplace
1749-1827

H hypothesis ("It rained")
O observation ("Floor wet")

P(H) probability of hypothesis to be true

P(O|H) probability to make observation O given hypothesis H is true

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

Bayesian update formula

Thomas Bayes
1701-1761

$$P(H\,|\,O) \propto P(O\,|\,H)P(H)$$

Pierre Simon Laplace
1749-1827

H hypothesis ("It rained")
O observation ("Floor wet")

"Prior"	**P(H)**	probability of hypothesis to be true
"Likelihood"	**P(O|H)**	probability to make observation O given hypothesis H is true
"Posterior"	**P(H|O)**	probability of hypothesis H given we observed O

Bayesian update formula

Thomas Bayes
1701-1761

$$P(H \mid O) \propto P(O \mid H) P(H)$$

Pierre Simon Laplace
1749-1827

H hypothesis
O observation

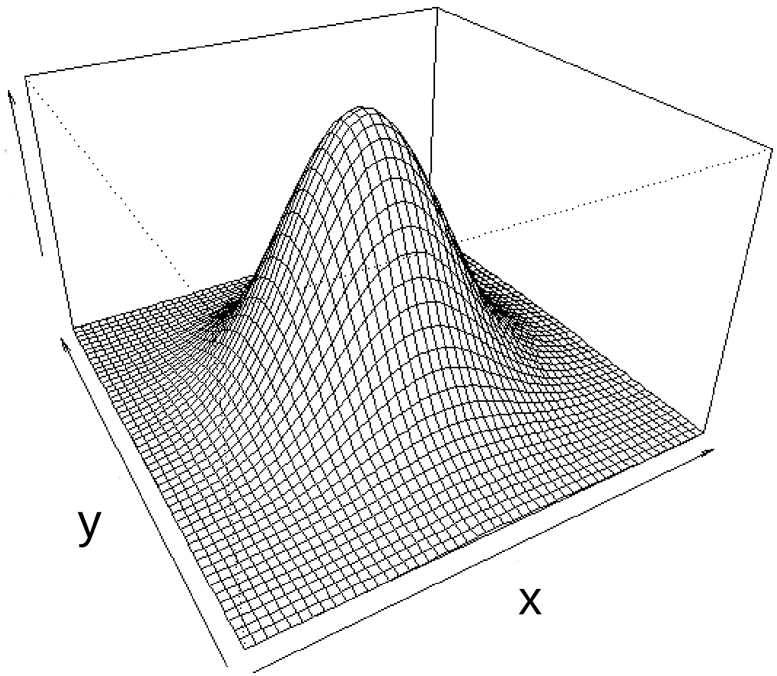
e.g., H(x) := “The cross section at 5 MeV is x mBarn”

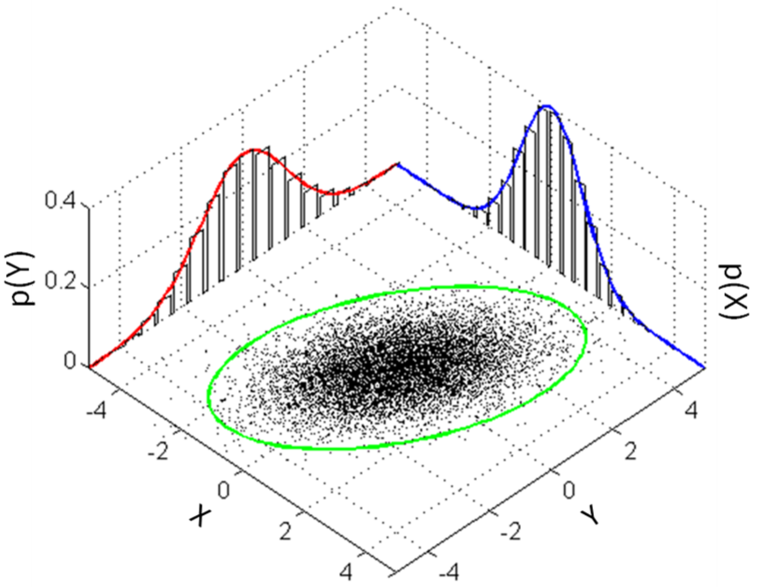
P(H(x)) → P(x)

Which probability distributions for P(O|H) and P(H)?

Multivariate normal distribution

$$\mathcal{N}(\vec{x}\,|\,\vec{x}_0, \mathbf{A}_0) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{A}_0|}} \exp\left(-\frac{1}{2} (\vec{x} - \vec{x}_0)^T \, \mathbf{A}_0^{-1} \, (\vec{x} - \vec{x}_0) \right)$$





$$\vec{x} = \begin{pmatrix} \text{altitude} \\ \text{temperature} \end{pmatrix} \qquad \vec{x}_0 = \begin{pmatrix} 1000 \\ 10 \end{pmatrix}$$

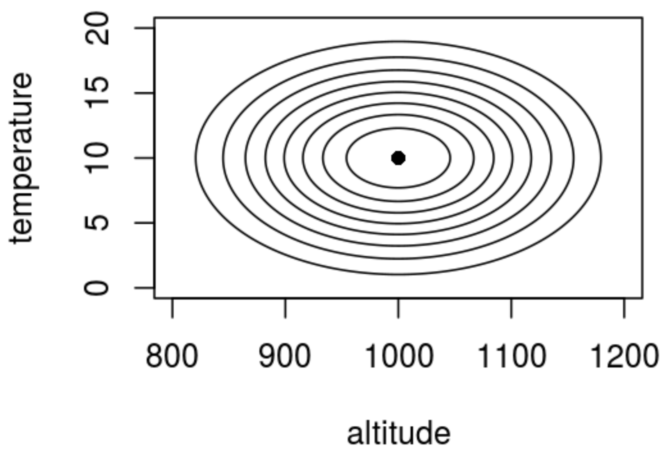
$$\vec{x} = \begin{pmatrix} \text{altitude} \\ \text{temperature} \end{pmatrix}$$

Covariance matrix

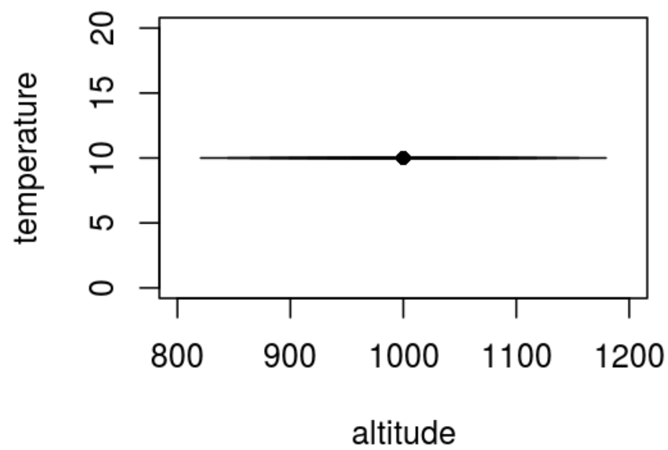
Variance / uncertainty squared

covariance

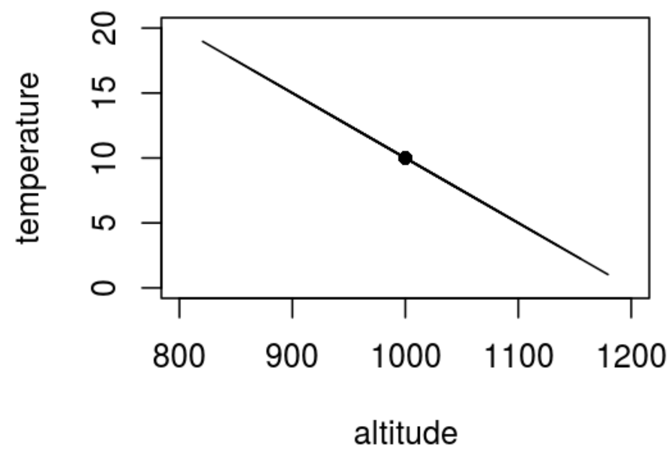
$$\mathbf{A}_0 = \begin{pmatrix} 100^2 & 0 \\ 0 & 5^2 \end{pmatrix}$$

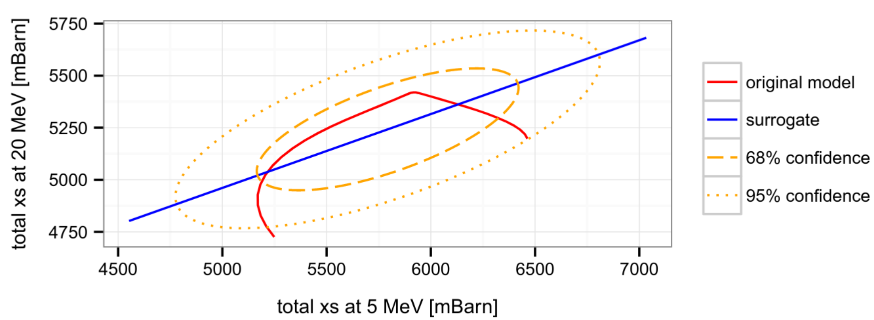


$$\mathbf{A}_0 = \begin{pmatrix} 100^2 & 0 \\ 0 & 0 \end{pmatrix}$$



$$\mathbf{A}_0 = \begin{pmatrix} 100^2 & -100 \cdot 5 \\ -100 \cdot 5 & 5^2 \end{pmatrix}$$



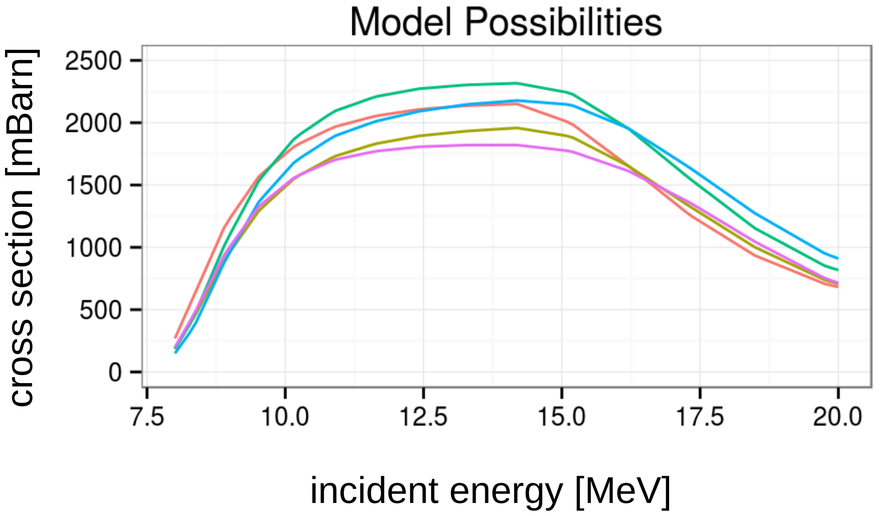


A covariance matrix captures linear relationships and uncertainties

Bayesian Inference in Nuclear Data

Basic assumption

Observables can be predicted by nuclear model M based on parameter set p



$\pi(\vec{p}\,|\,M)$ prior distribution on model parameters

$f(\vec{\sigma}\,|\,\vec{p}, M)$ likelihood: probability to measure cross sections σ if model parameter set p is true

$\pi(\vec{p}\,|\,\vec{\sigma}, M)$ posterior: refined probability distribution for parameters taking into account observations σ

$$\pi(\vec{p}\,|\,\vec{\sigma}, M) \propto f(\vec{\sigma}\,|\,\vec{p}, M)\pi(\vec{p}\,|\,M)$$

Which probability distributions?

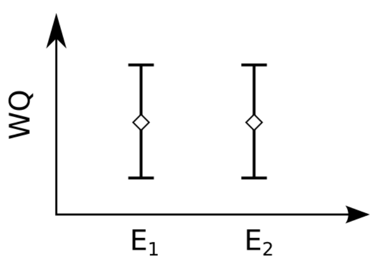
Prior distribution (model parameters)

$$\pi(\vec{p}\,|\,M) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{A}_0}} \exp\left(-\frac{1}{2}(\vec{p}-\vec{p}_0)^T \mathbf{A}_0^{-1}(\vec{p}-\vec{p}_0)\right)$$

supported by
- principle of maximum entropy
- easy to work with!

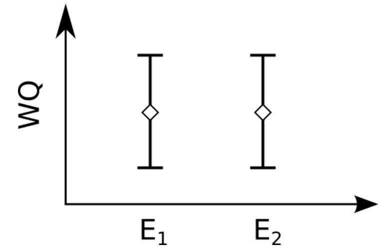
Likelihood (experimental data)

$$f(\vec{\sigma}\,|\,\vec{p}, M) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{B}}} \exp\left(-\frac{1}{2}(\vec{\sigma}-M[\vec{p}])^T \mathbf{B}^{-1}(\vec{\sigma}-M[\vec{p}])\right)$$



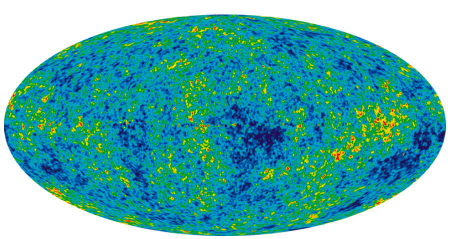
supported by
- principle of maximum entropy,
- limiting distribution,
- central limit theorem
- easy to work with!

Cooking a covariance matrix



detector

sample thickness*



background noise

“statistical” (counting) uncertainty

$$\mathbf{B}_{\text{stat}} = \begin{bmatrix} \sigma_1^2\delta_1^2 & 0 \\ 0 & \sigma_2^2\delta_2^2 \end{bmatrix}$$

detector calibration uncertainty

$$\mathbf{B}_{\text{cal}} = \begin{bmatrix} \sigma_1^2\gamma_1^2 & \sigma_1\sigma_2\gamma_1\gamma_2 \\ \sigma_1\sigma_2\gamma_1\gamma_2 & \sigma_2^2\gamma_2^2 \end{bmatrix}$$

uncertainty about sample thickness
uncertainty about background noise
...

$$\mathbf{B} = \mathbf{B}_{\text{stat}} + \mathbf{B}_{\text{cal}} + \mathbf{B}_{\text{thick}} + \ldots$$

Quantities of interest

- Expectation of parameters according to posterior distribution

$$\mathbb{E}\left[\vec{\sigma}\right] = \int\int\cdots\int M(\vec{p})\pi(\vec{p}\,|\,\vec{\sigma}, M)dp_1 dp_2 \ldots dp_N$$

- Mode (maximum) of posterior distribution

$$\text{find } \vec{p}_{\text{max}} \text{ that maximizes } \pi(\vec{p}\,|\,\vec{\sigma}, M)$$

- Standard deviation parameters according to posterior distribution
- Correlations of distribution

Nuclear data challenges

Regarding nuclear models

- Non-linear (nuclear physics is complicated)
- Not analytic (differential equations, simulation)
- Computationally expensive (minutes to hours)

Regarding linear algebra

- Computing with large covariance matrices

Regarding statistical models

- Imperfect physics model
- Multivariate normal distribution may be not always appropriate
- Uncertainties wrong or unknown

$$\pi(\vec{p} \,|\, \vec{\sigma}, M) \propto f(\vec{\sigma} \,|\, \vec{p}, M)\pi(\vec{p} \,|\, M)$$

$$f(\vec{\sigma} \,|\, \vec{p}, M) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{B}}} \exp\left(-\frac{1}{2}(\vec{\sigma} - M[\vec{p}])^T \mathbf{B}^{-1} (\vec{\sigma} - M[\vec{p}])\right)$$

Nuclear data challenges

Regarding nuclear models

- Non-linear (nuclear physics is complicated)
- Not analytic (differential equations, simulation)
- Computationally expensive (minutes to hours)

Regarding linear algebra

- Computing with large covariance matrices

Regarding statistical models

- Imperfect physics model
- Multivariate normal distribution may be not always appropriate
- Uncertainties wrong or unknown

$$\pi(\vec{p} \,|\, \vec{\sigma}, M) \propto f(\vec{\sigma} \,|\, \vec{p}, M)\pi(\vec{p} \,|\, M)$$

$$f(\vec{\sigma} \,|\, \vec{p}, M) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{B}}} \exp\left(-\frac{1}{2}(\vec{\sigma} - M[\vec{p}])^T \mathbf{B}^{-1} (\vec{\sigma} - M[\vec{p}])\right)$$

How to evaluate?

$$\pi(\vec{p}_{\text{true}} \,|\, \vec{\sigma}_{\text{exp}}, M) \propto f(\vec{\sigma}_{\text{exp}} \,|\, \vec{p}_{\text{true}}, M)\pi(\vec{p}_{\text{true}} \,|\, M)$$

Three possibilities

a) Linearization

$$M_{\text{lin}}[\vec{p}] = M[\vec{p}_0] + \mathbf{S}(\vec{p} - \vec{p}_0)$$

b) Surrogate approach

$$\pi(\vec{p}_{\text{true}} \,|\, M) \rightarrow \pi(\vec{\sigma}_{\text{true}} \,|\, M_{\text{sur}})$$

c) Monte Carlo procedure

simplified models

exact model

Monte Carlo approach

$$\pi(\vec{p}_{\rm true}\,|\,\vec{\sigma}_{\rm exp}, M_{\rm lin}) \propto \exp\left[-\frac{1}{2}(\vec{\sigma}_{\rm exp} - M[\,\vec{p}_{\rm true}\,])^T \mathbf{B}^{-1}(\vec{\sigma}_{\rm exp} - M[\,\vec{p}_{\rm true}\,])\right]$$

$$\times \exp\left[-\frac{1}{2}(\vec{p}_{\rm true} - \vec{p}_0)^T \mathbf{A}_0^{-1}(\vec{p}_{\rm true} - \vec{p}_0)\right]$$

1) generate ensemble $\vec{p}_1, \vec{p}_2, \vec{p}_3, \ldots$ from prior

2) calculate $M[\,\vec{p}_1\,], M[\,\vec{p}_2\,], M[\,\vec{p}_3\,], \ldots$

3) Calculate statistics from posterior distribution, e.g., best estimates, correlations, uncertainties, etc. by likelihood weighting $f(\sigma_{\rm exp}\,|\,\vec{p}_i, M)$

$$\mathbb{E}[\,\vec{\sigma}_{\rm true}\,] = \frac{\sum_{i=1}^n M[\,\vec{p}_i\,] f(\sigma_{\rm exp}\,|\,\vec{p}_i, M)}{\sum_{i=1}^n f(\sigma_{\rm exp}\,|\,\vec{p}_i, M)}$$

Linearization

$$\pi(\vec{p}_{\rm true}\,|\,\vec{\sigma}_{\rm exp}, M) \propto f(\vec{\sigma}_{\rm exp}\,|\,\vec{p}_{\rm true}, M)\pi(\vec{p}_{\rm true}\,|\,M)$$

$$\pi(\vec{p}_{\rm true}\,|\,\vec{\sigma}_{\rm exp}, M_{\rm lin}) \propto \exp\left[-\frac{1}{2}(\vec{\sigma}_{\rm exp} - M_{\rm lin}[\vec{p}_{\rm true}])^T\mathbf{B}^{-1}(\vec{\sigma}_{\rm exp} - M_{\rm lin}[\vec{p}_{\rm true}])\right] \times\exp\left[-\frac{1}{2}(\vec{p}_{\rm true} - \vec{p}_0)^T\mathbf{A}_0^{-1}(\vec{p}_{\rm true} - \vec{p}_0)\right]$$

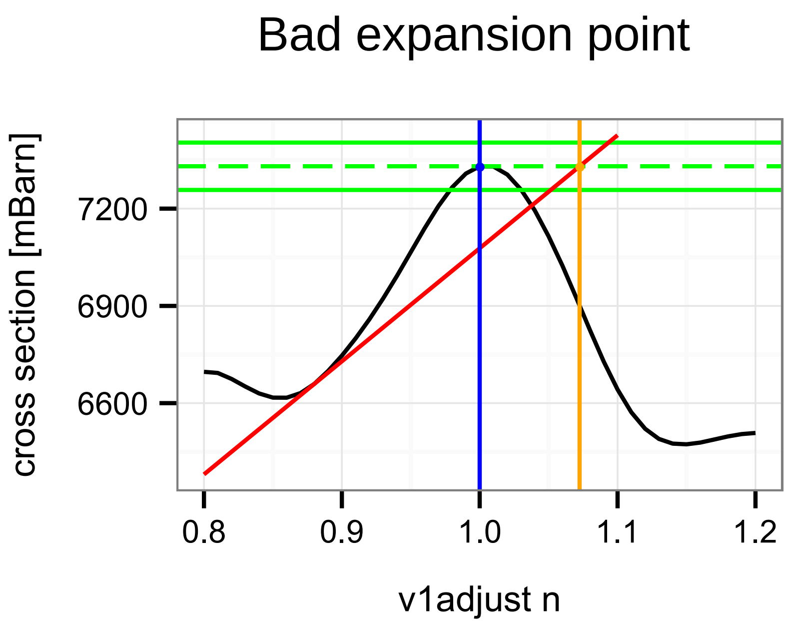
$$M_{\rm lin}[\vec{p}] = M[\vec{p}_0] + \mathbf{S}(\vec{p} - \vec{p}_0)$$

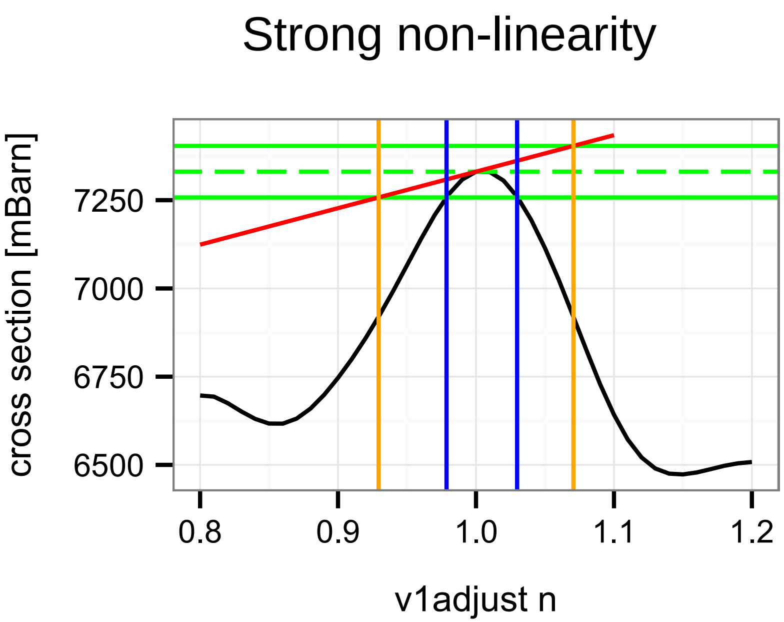
Analytical update formulas

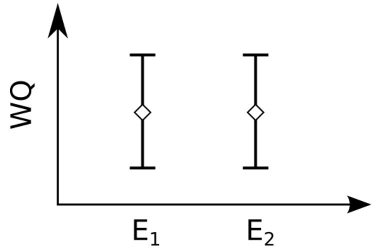
$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}\mathbf{S}\mathbf{A}_0$$

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}(\vec{\sigma}_{\rm exp} - M_{\rm lin}[\vec{p}_0])$$

Linearization: Possible issues

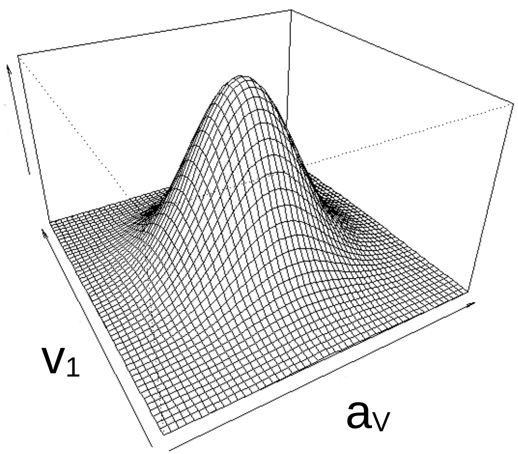






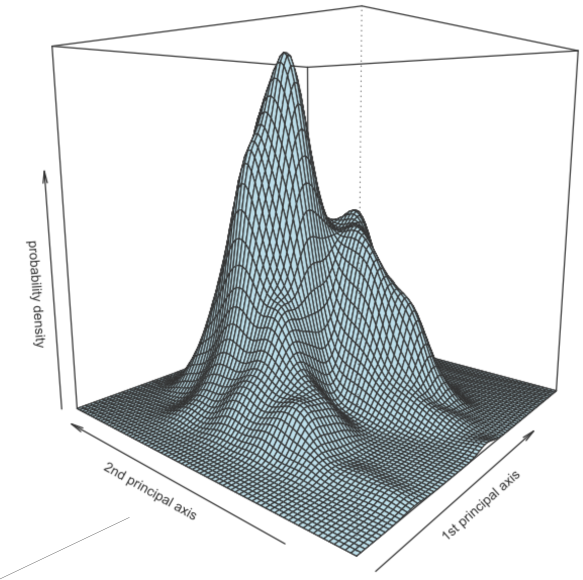
Surrogate approach
basic idea

$$\pi(\vec{p}_{\rm true}\,|\,M)$$



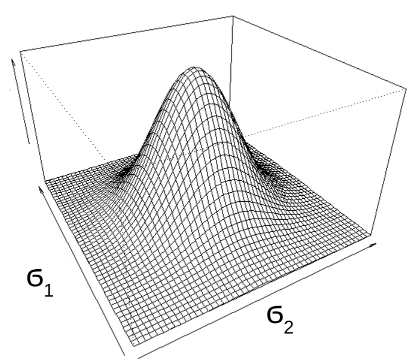
is linked to

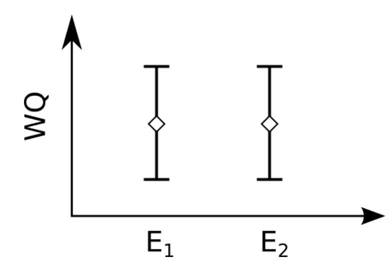
$$\pi(\vec{\sigma}_{\rm true}\,|\,M)$$



replaced by

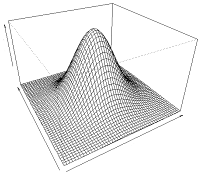
$$\pi(\vec{\sigma}_{\rm true}\,|\,M_{\rm sur})$$



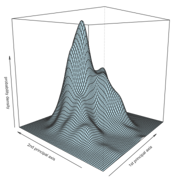


$$\pi(\vec{p}_{\rm true}\,|\,\vec{\sigma}_{\rm exp}, M) \propto f(\vec{\sigma}_{\rm exp}\,|\,\vec{p}_{\rm true}, M)\,\pi(\vec{p}_{\rm true}\,|\,M)$$

Surrogate approach
construction of M_{sur}

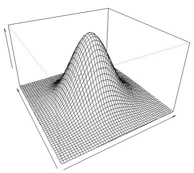


1) draw ensemble $\vec{p}_1\, \vec{p}_2, \vec{p}_3, \ldots, \vec{p}_n$ from prior distribution $\pi(\vec{p}_{\mathrm{true}} \,|\, M)$



2) use nuclear model to calculate $\vec{\sigma}_1 = M[\,\vec{p}_1\,], \vec{\sigma}_2 = M[\,\vec{p}_2\,], \ldots$

3) estimate multivariate normal distribution in observation space



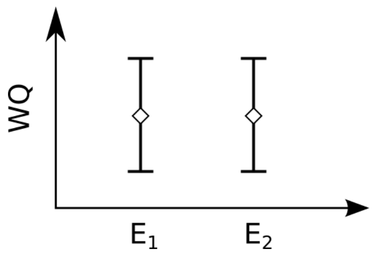
$$\vec{\sigma}_0 = \frac{1}{n}\sum_{i=1}^{n} \vec{\sigma}_i \qquad \mathbf{A}_0 = \frac{1}{n}\sum_{i=1}^{n} (\vec{\sigma}_i - \vec{\sigma}_0)(\vec{\sigma}_i - \vec{\sigma}_0)^T \qquad M_{\mathrm{sur}}[\,\sigma_{\mathrm{true}}\,] = \mathbf{S}\vec{\sigma}_{\mathrm{true}}$$

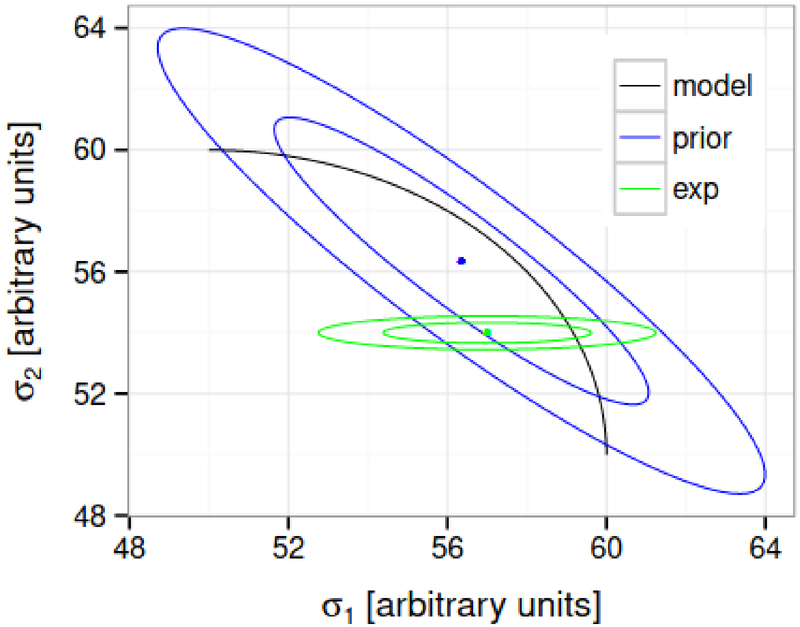
$$\pi(\vec{p}_{\mathrm{true}} \,|\, \vec{\sigma}_{\mathrm{exp}}, M_{\mathrm{lin}}) \propto \exp\left[-\frac{1}{2}(\vec{\sigma}_{\mathrm{exp}} - M_{\mathrm{sur}}[\,\vec{\sigma}_{\mathrm{true}}\,])^T \mathbf{B}^{-1} (\vec{\sigma}_{\mathrm{exp}} - M_{\mathrm{sur}}[\,\vec{\sigma}_{\mathrm{true}}\,])\right]$$
$$\times \exp\left[-\frac{1}{2}(\vec{\sigma}_{\mathrm{true}} - \vec{\sigma}_0)^T \mathbf{A}_0^{-1} (\vec{\sigma}_{\mathrm{true}} - \vec{\sigma}_0)\right]$$

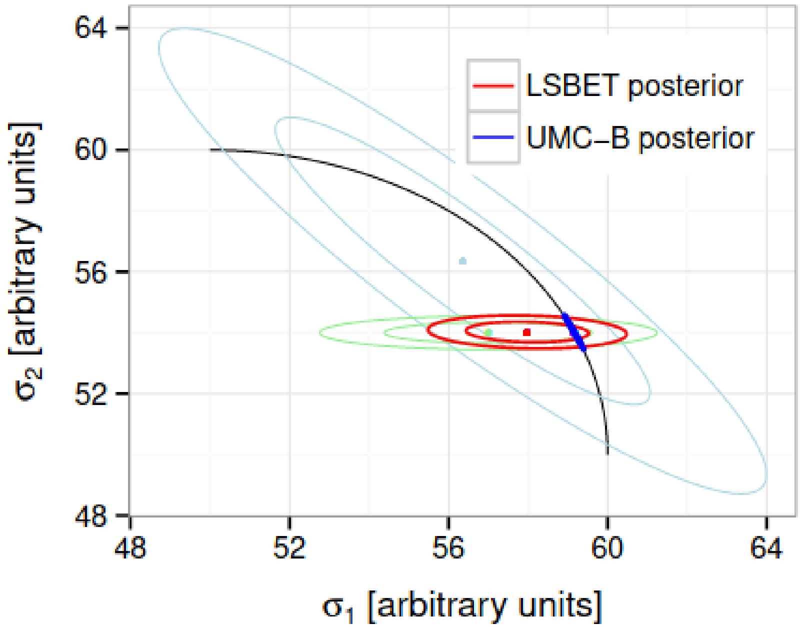
Analytic update formulas

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}\mathbf{S}\mathbf{A}_0$$
$$\vec{\sigma}_1 = \vec{\sigma}_0 + \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}(\vec{\sigma}_{\mathrm{exp}} - M_{\mathrm{sur}}[\,\vec{\sigma}_0]\,)$$

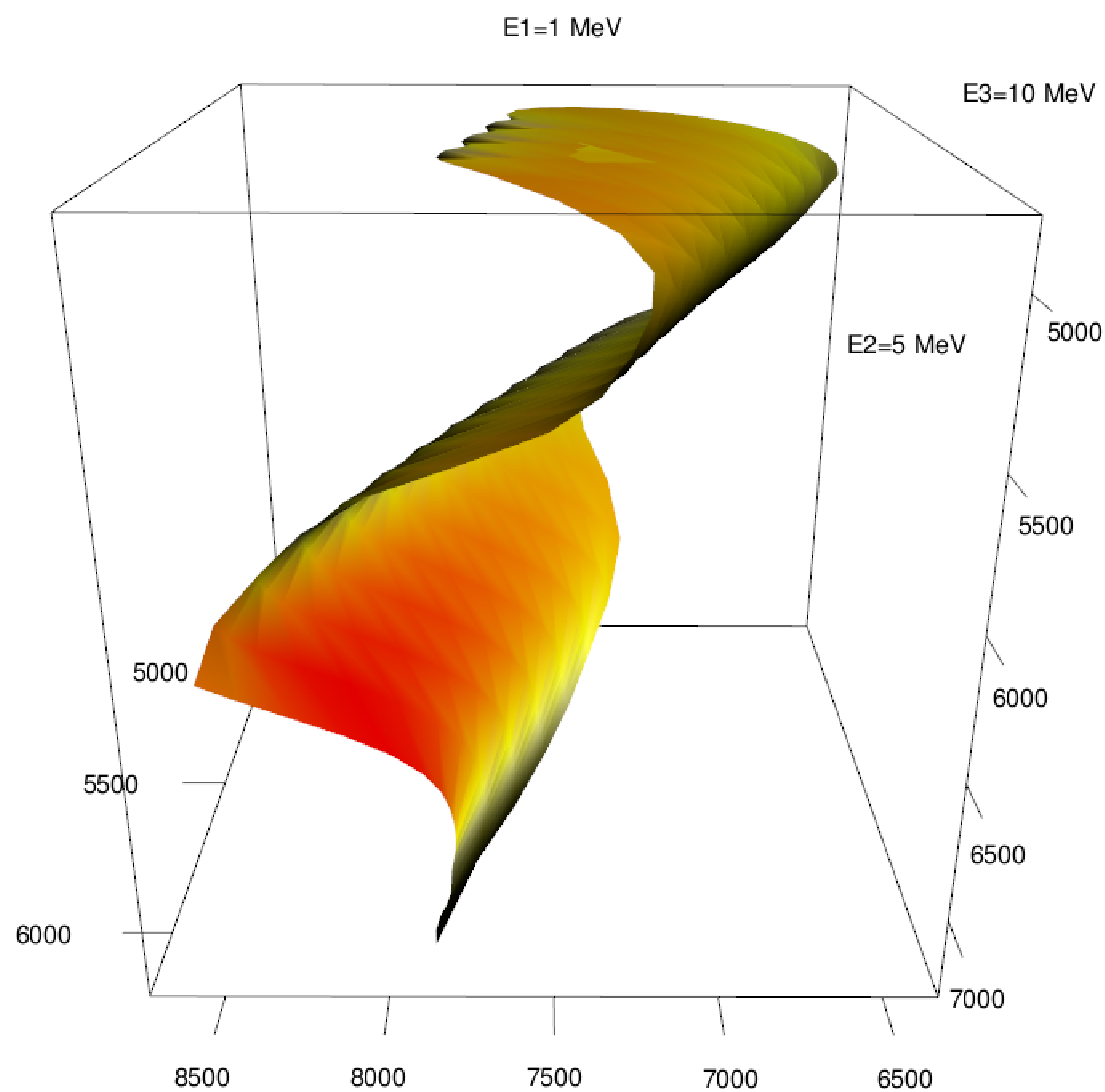
Surrogate approach: Behavior







Visualization of non-linearity



Generalized Least Squares (GLS)

in a nutshell

Prior	$\pi(\vec{p}) = \mathcal{N}(\vec{p}\,\vert\,\vec{p}_0, \mathbf{A}_0)$
Likelihood	$f(\vec{\sigma}_{\text{exp}}\,\vert\,\vec{p}) = \mathcal{N}(\vec{\sigma}_{\text{exp}}\,\vert\,M_{\text{lin}}[\vec{p}], \mathbf{B})$

$M_{\text{lin}}[\vec{p}] = \vec{\sigma}_{\text{ref}} + \mathbf{S}\vec{p}$

Bayesian update $\vec{\pi}(\vec{p}\,\vert\,\vec{\sigma}_{\text{exp}}) \propto f(\vec{\sigma}_{\text{exp}}\,\vert\,\vec{p})\,\pi(\vec{p}\,\vert\,\vec{\sigma}_{\text{exp}})$

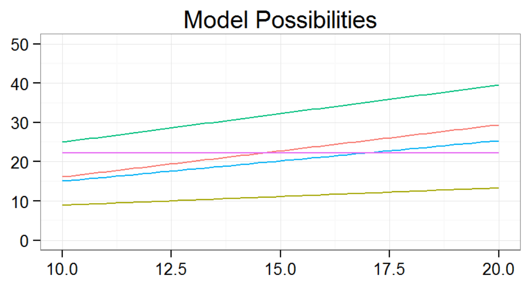
Posterior	$\pi(\vec{p}\,\vert\,\vec{\sigma}_{\text{exp}}) = \mathcal{N}(\vec{p}\,\vert\,\vec{p}_1, \mathbf{A}_1)$

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}(\vec{\sigma}_{\text{exp}} - M_{\text{lin}}[\,\vec{p}_0]\,)$$

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0\mathbf{S}^T(\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1}\mathbf{S}\mathbf{A}_0$$

"GLS formulas"

Simple example: straight line model



$$y(x) = {\color{green}k}x + {\color{green}d}$$

Prior

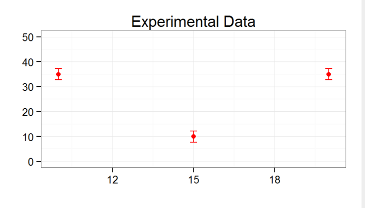
$$k \propto \mathcal{N}(0, \delta_k^2)$$

$$d \propto \mathcal{N}(0, \delta_d^2)$$

$$\vec{p}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \mathbf{A}_0 = \begin{pmatrix} \delta_k^2 & 0 \\ 0 & \delta_d^2 \end{pmatrix}$$

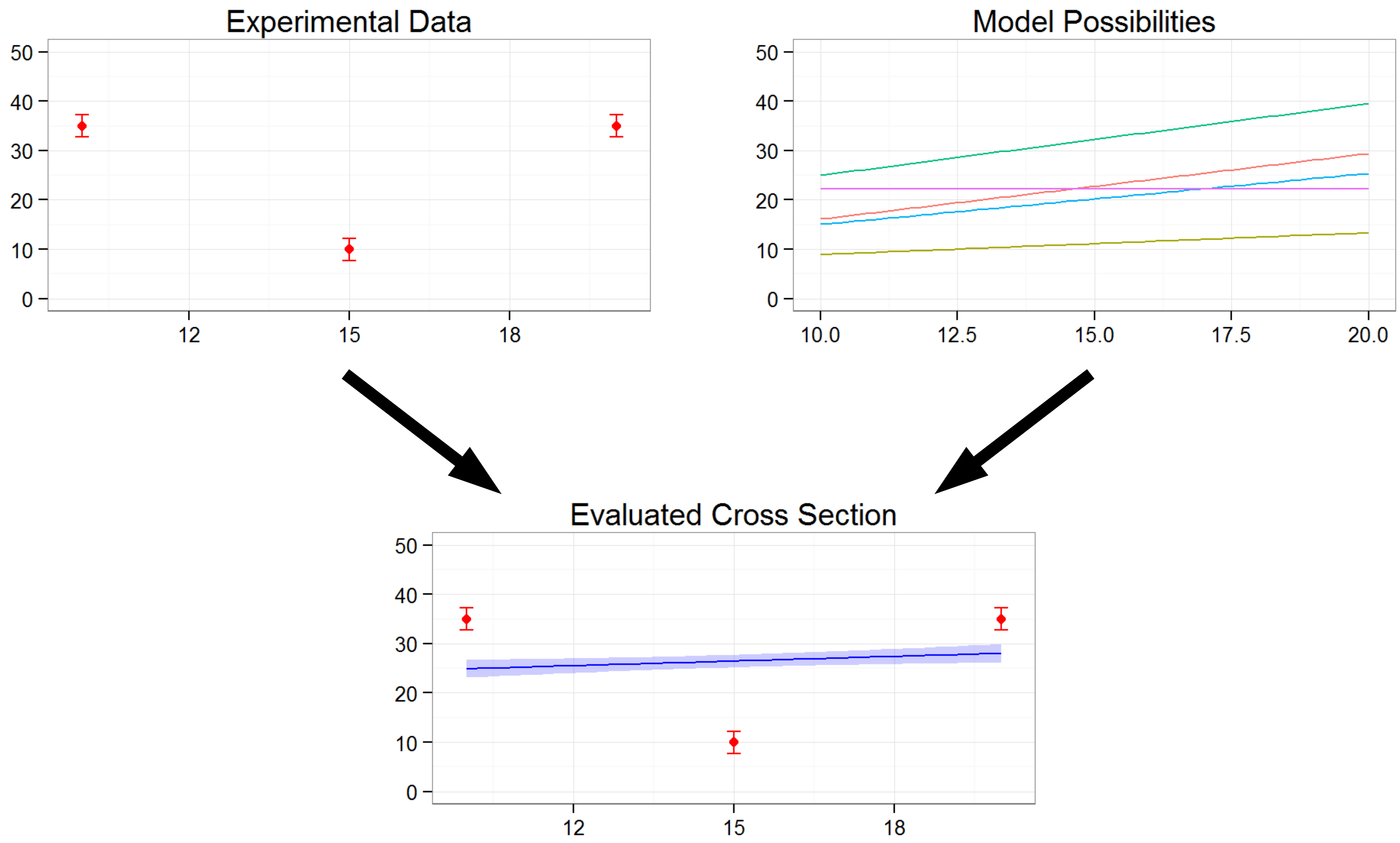
Likelihood

$$\vec{\sigma}_{\text{exp}} = \begin{pmatrix} \sigma_{\text{exp},1} \\ \sigma_{\text{exp},2} \\ \vdots \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

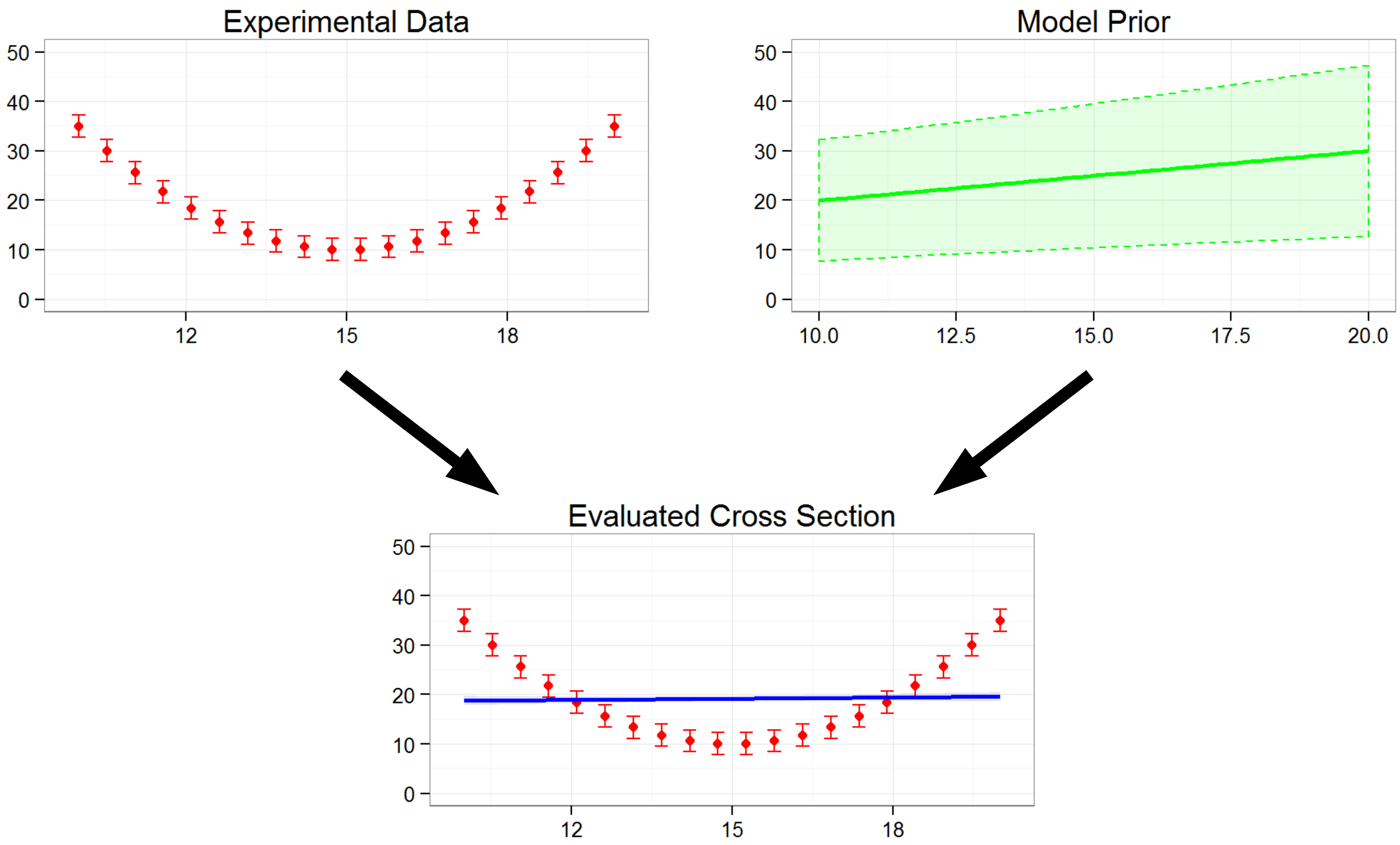


$$M_{\text{lin}}[\vec{p}] = \mathbf{S}\vec{p} \qquad \mathbf{S} = \begin{pmatrix} E_1 & 1 \\ E_2 & 1 \\ \vdots & \end{pmatrix}$$

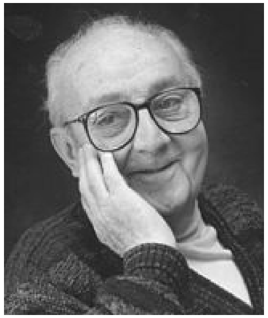
Straight line model



Straight line model
(garbage in, garbage out)



Important insight about models



"All models are wrong but some are useful"

- George E. P. Box

Ice cream or pizza?

Reality

More realistic modeling

Experimental Observation = Model Prediction + Measurement Error

$$\sigma_{Exp}(E) = \sigma_{Mod}(E) + \varepsilon_{Exp}(E)$$

$$\sigma_{True} = \sigma_{Mod}$$

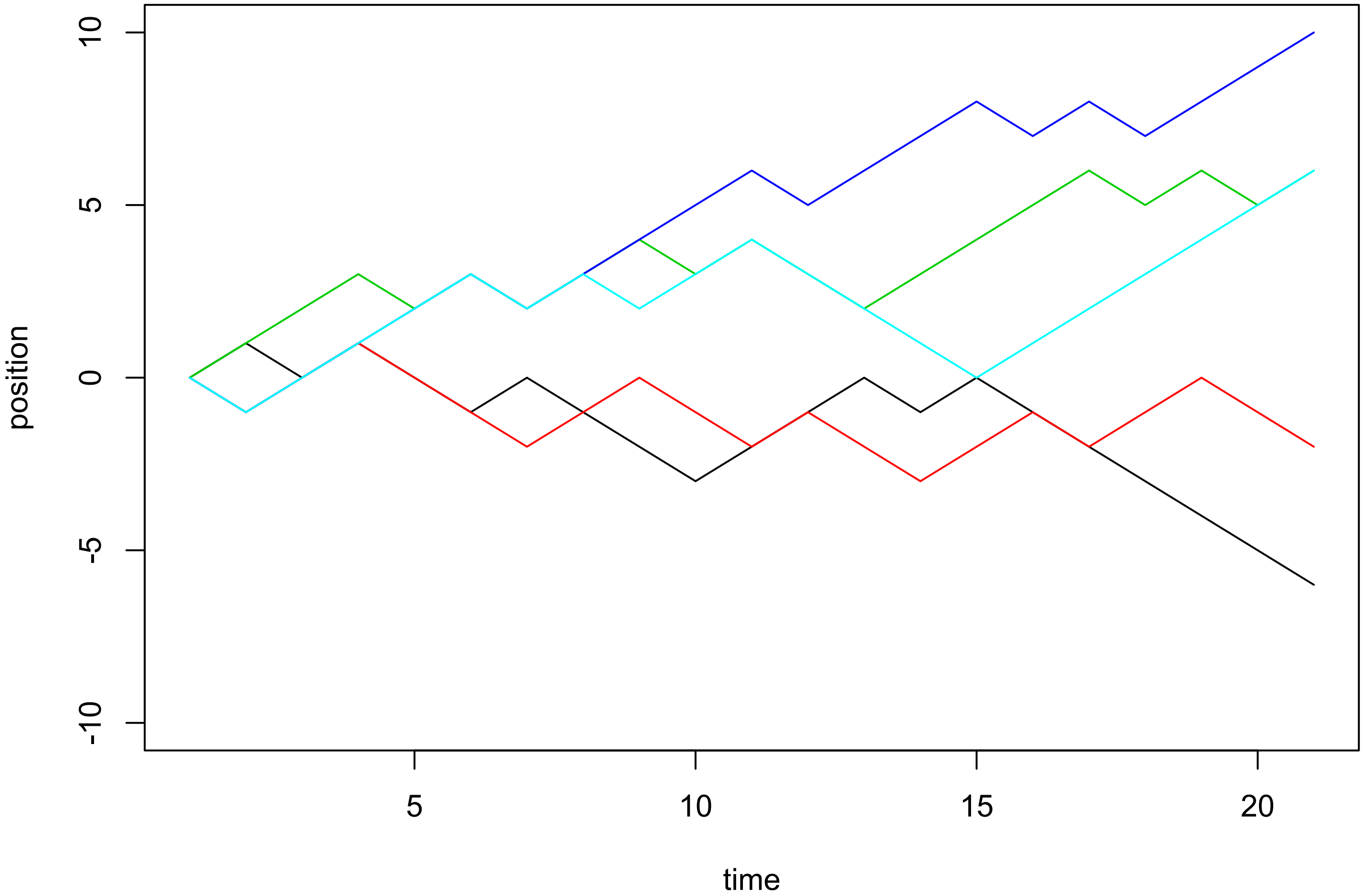
Experimental Observation = Model Prediction + Model Error + Measurement Error

$$\sigma_{Exp}(E) = \sigma_{Mod}(E) + \varepsilon_{Mod}(E) + \varepsilon_{Exp}(E)$$

$$\sigma_{True} = \sigma_{Mod} + \varepsilon_{Mod}$$

Stochastic process

e.g., stock price evolution



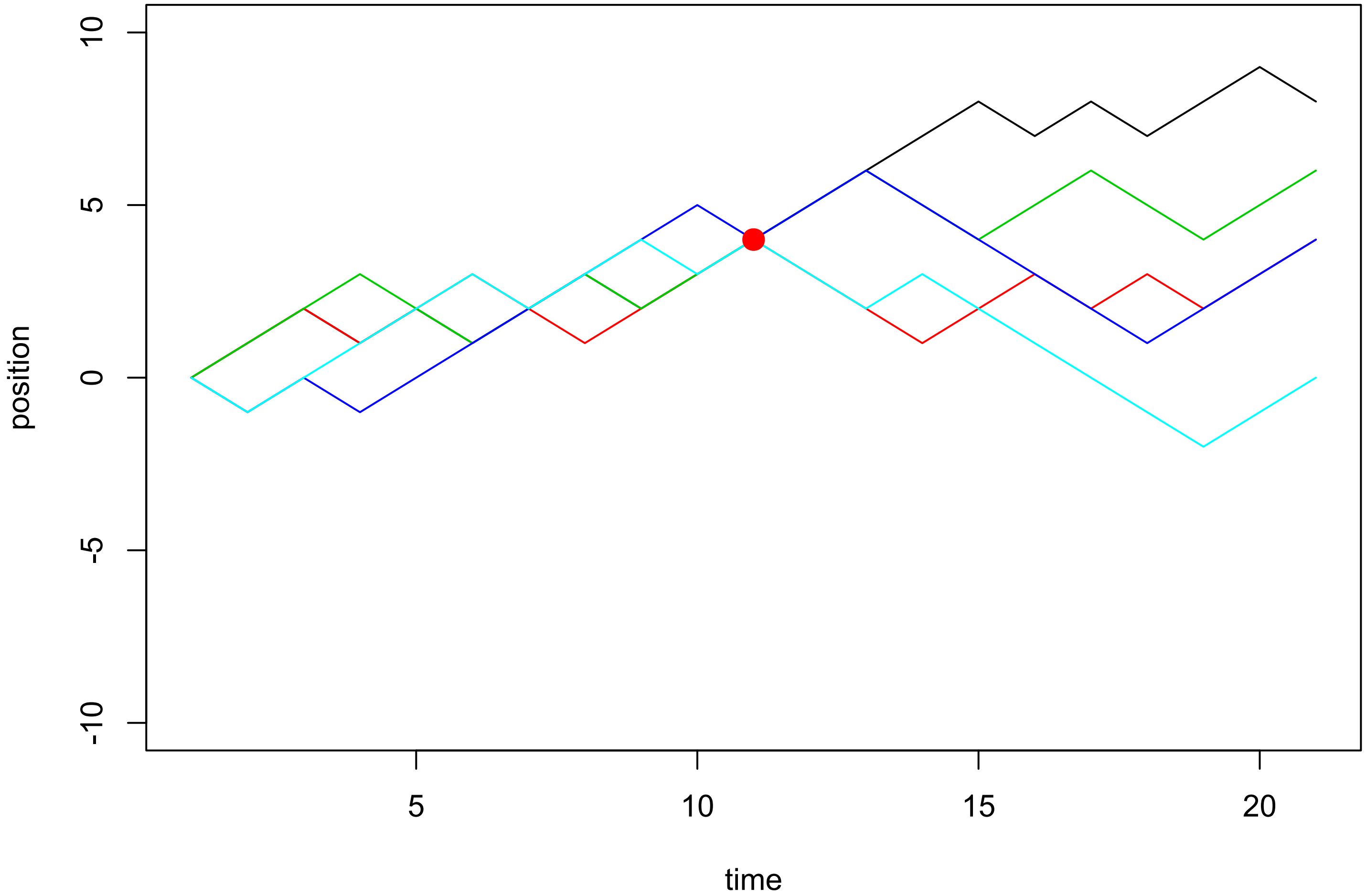
1 jump/second

jumps: +1 / -1

stochastic process is associated with a random function

How can we mathematically define a process?

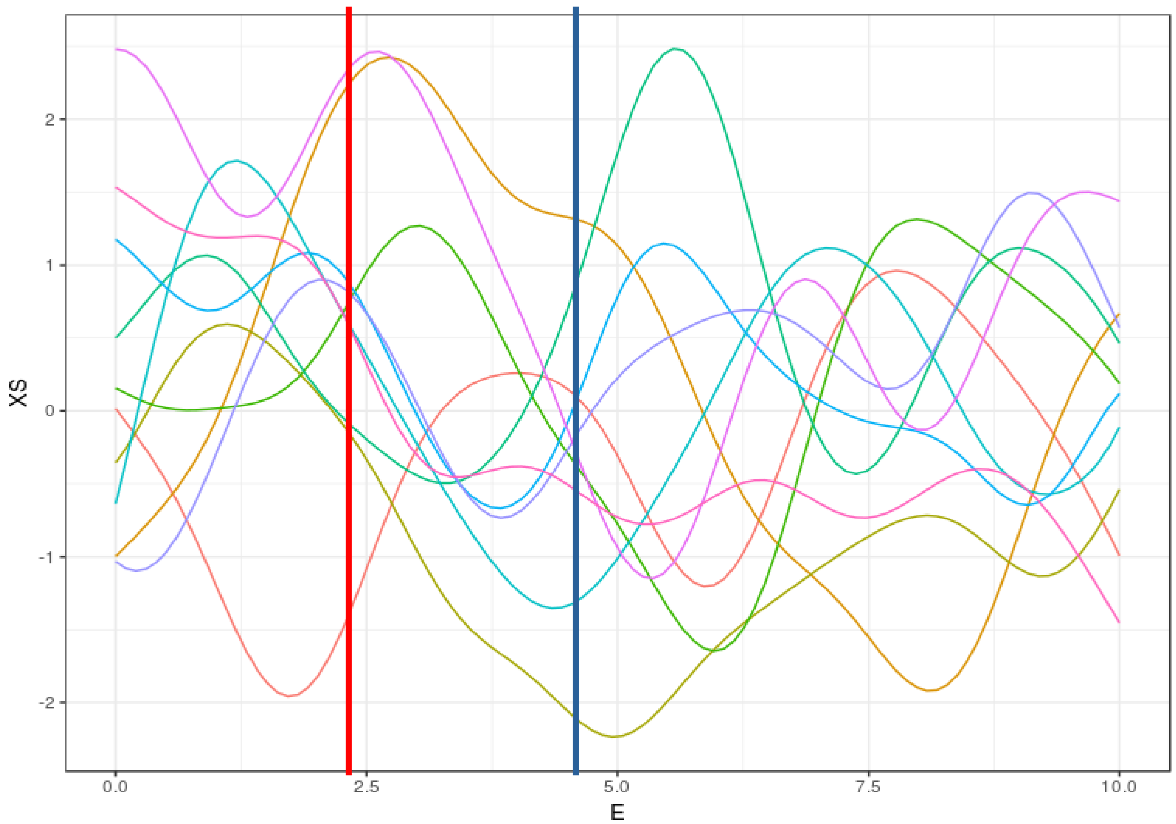
Stochastic process
with observation

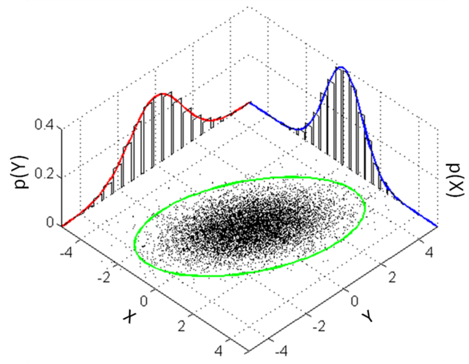


observation
rules out
a lot of paths

How can we constrain a process with observations?

Characterization of a Gaussian process





All finite sets of function values at different locations are distributed according to a multivariate normal distribution
→ <u>Gaussian process</u>

Specification of a Gaussian process

Specify a function that yields the covariance between function values f($\mathbf{x}_1$) and f($\mathbf{x}_2$) for any possible pair $\mathbf{x}_1$ and $\mathbf{x}_2$. This function is called a **covariance function**.

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

Specify another function μ(**x**) that yields the center value of the process at location **x.** This function is called a **mean function**.

$$\mu(\mathbf{x}) = 0$$

Sampling from a Gaussian process

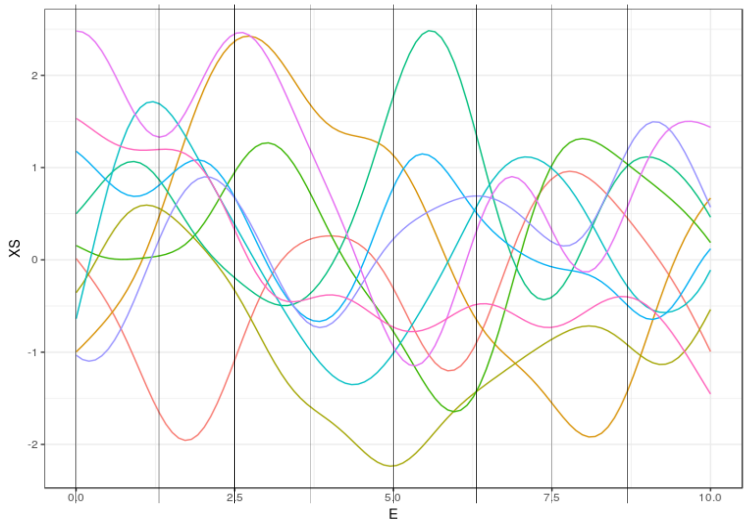
$$\kappa(\boldsymbol{x}_1, \boldsymbol{x}_2) = \delta^2 \exp\left(-\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^2}{2\lambda^2}\right)$$

$$\mu(\mathbf{x}) = 0$$

$$\mathbf{K} = \begin{pmatrix} \kappa(x_1, x_1) & \kappa(x_1, x_2) & \kappa(x_1, x_3) & \cdots \\ \kappa(x_2, x_1) & \kappa(x_2, x_2) & \kappa(x_2, x_3) & \cdots \\ \kappa(x_3, x_1) & \kappa(x_3, x_2) & \kappa(x_3, x_3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$\vec{m} = \begin{pmatrix} \mu(x_1) \\ \mu(x_2) \\ \mu(x_3) \\ \vdots \end{pmatrix}$$

$$\mathbf{y} \sim \mathcal{N}(\vec{m}, \mathbf{K})$$

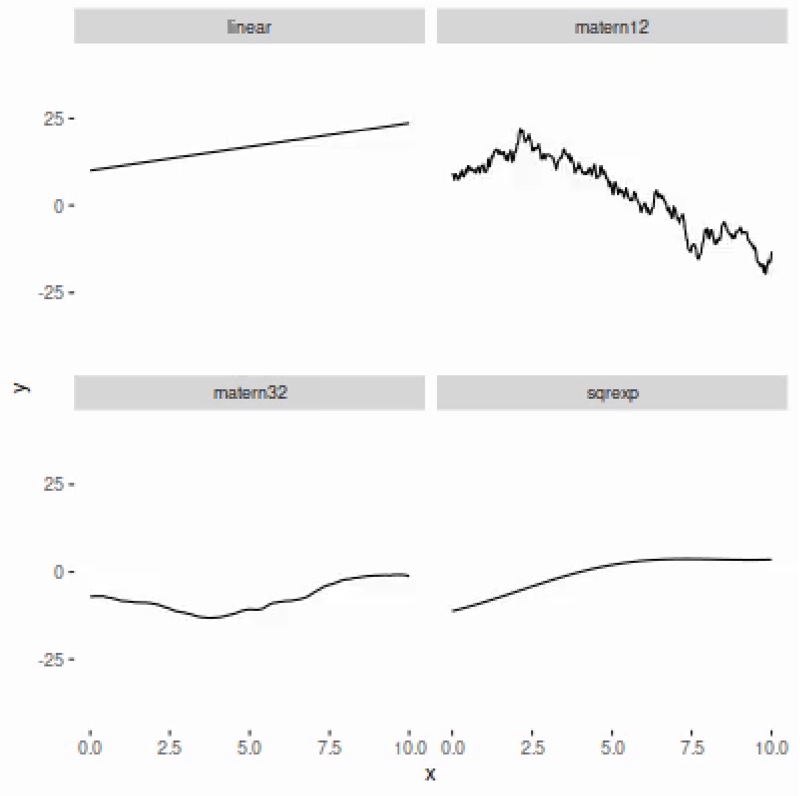


$X_1 \quad X_2 \quad X_3 \quad \ldots$

Examples of Gaussian processes

$$\kappa_{\text{lin}}(x_1, x_2) = x_1 x_2 {\delta_k}^2 + {\delta_d}^2$$

$$\kappa_{\text{mat},1/2}(x_1, x_2) = \delta^2 \exp\left(-\frac{|x_1 - x_2|}{\rho}\right)$$



Play video

$$\kappa_{\text{mat},3/2}(x_1, x_2) = \delta^2 \left(1 + \frac{\sqrt{3}|x_1 - x_2|}{\rho}\right) \exp\left(-\frac{\sqrt{3}|x_1 - x_2|}{\rho}\right)$$

$$\kappa_{\text{sqrexp}}(x_1, x_2) = \delta^2 \exp\left(-\frac{1}{2\lambda^2}(x_1 - x_2)^2\right)$$

Constraining with observations / Model training

Use GLS formulas:

$$\vec{p}_1 = \vec{p}_0 + \mathbf{A}_0 \mathbf{S}^T (\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1} (\vec{\sigma}_{\text{exp}} - M_{\text{lin}}[\,\vec{p}_0]\,)$$

$$\mathbf{A}_1 = \mathbf{A}_0 - \mathbf{A}_0 \mathbf{S}^T (\mathbf{S}\mathbf{A}_0\mathbf{S} + \mathbf{B})^{-1} \mathbf{S}\mathbf{A}_0$$

Introduce computational mesh **U** and experimental mesh **V** and identify:

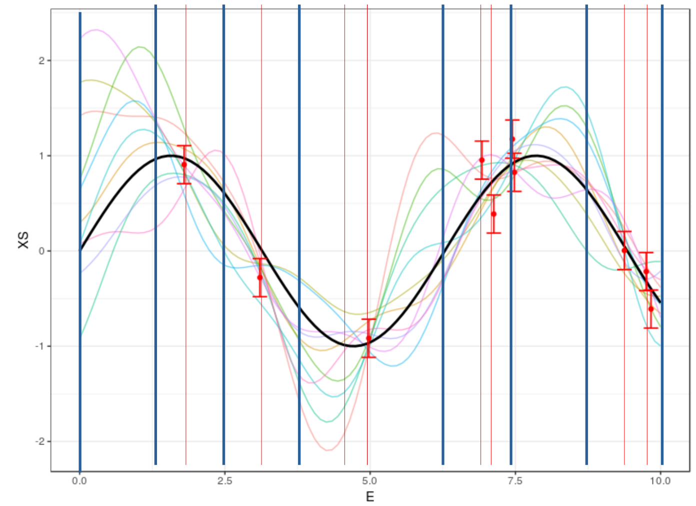
$$\mathbf{A}_0 \rightarrow \mathbf{K}_{UU}$$

$$\vec{p}_0 \rightarrow \vec{m}_U$$

$$M_{\text{lin}}[\vec{p}_0] \rightarrow \vec{m}_V$$

$$\mathbf{S}\mathbf{A}_0\mathbf{S}^T \rightarrow \mathbf{K}_{VV}$$

$$\mathbf{S}\mathbf{A}_0 \rightarrow \mathbf{K}_{VU}$$

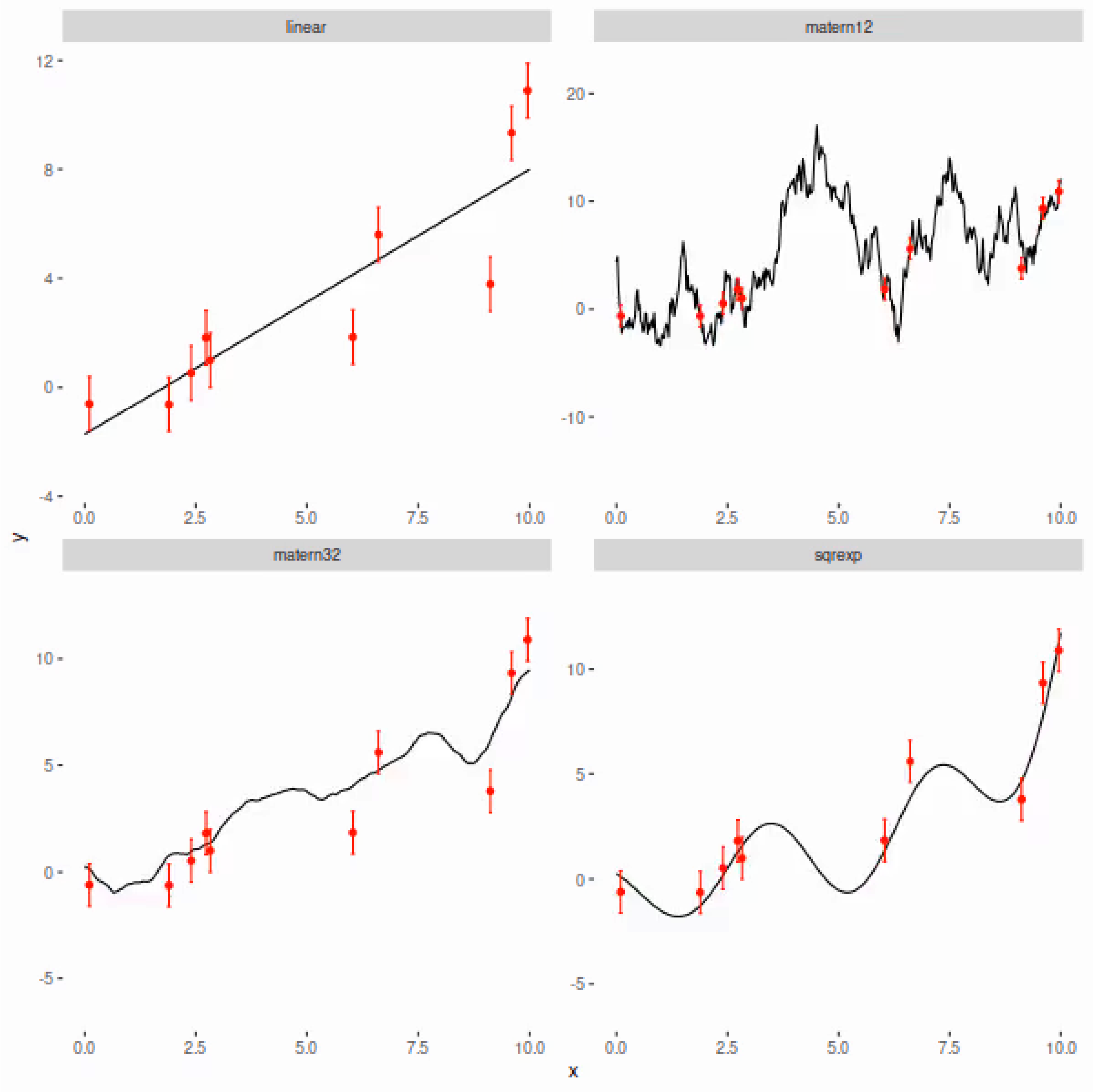


Resulting vector p_1 contains posterior predictions on mesh **U** and $\mathbf{A}_1$ is the associated covariance matrix

$$\vec{p} \sim \mathcal{N}(\vec{p}_1, \mathbf{A}_1)$$

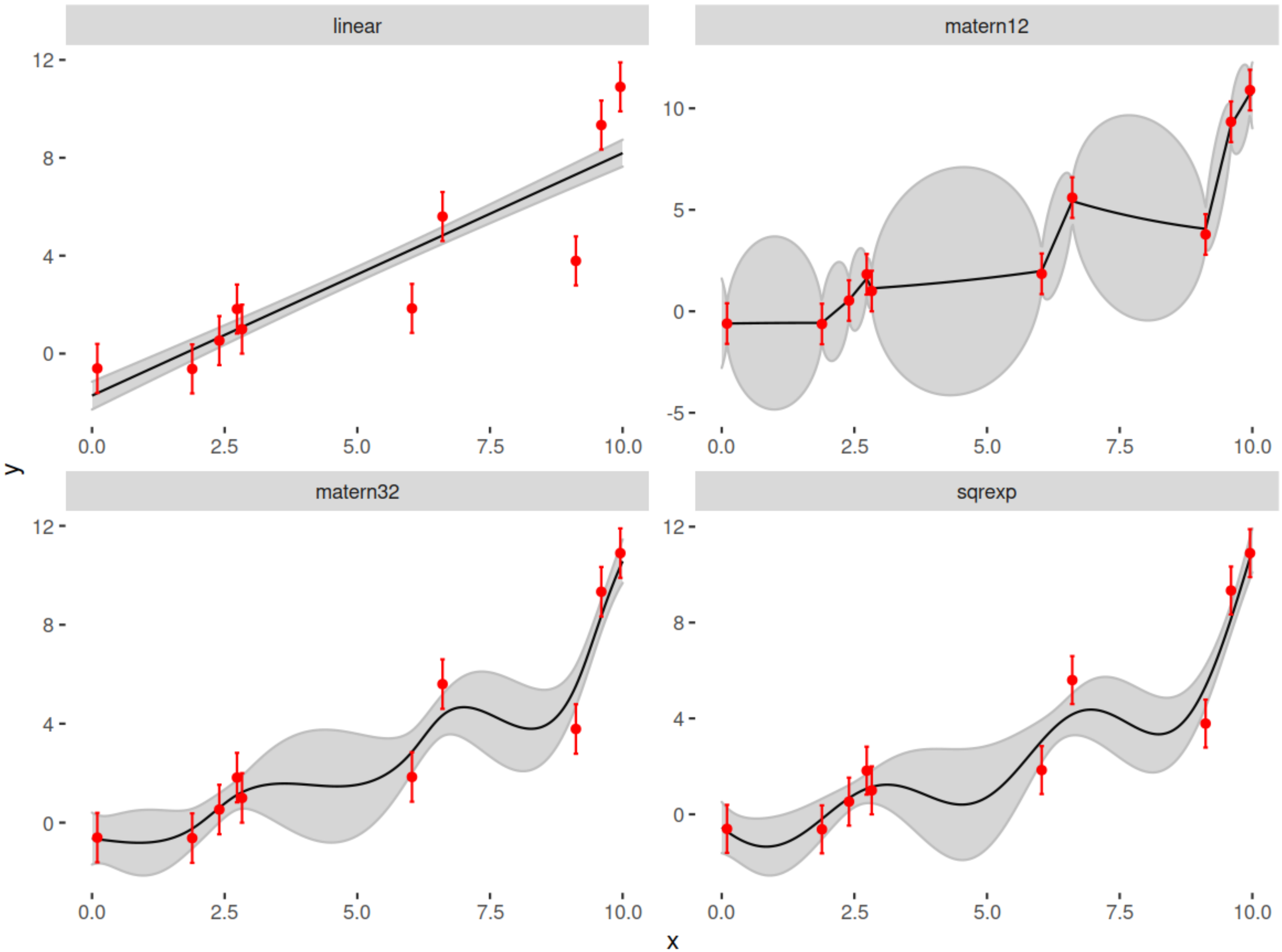
(Other terms: Gaussian process regression, Kriging)

GPs constrained with observations



Play video

GPs constrained with observations

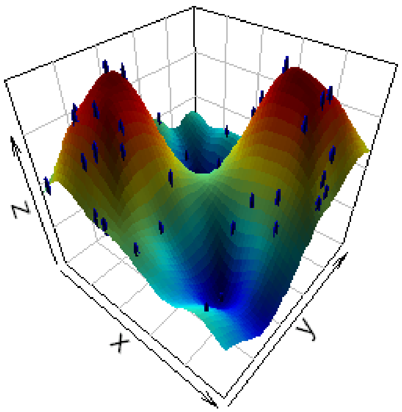


The posterior covariance matrix encodes more than just uncertainties!

Machine learning view on GPs

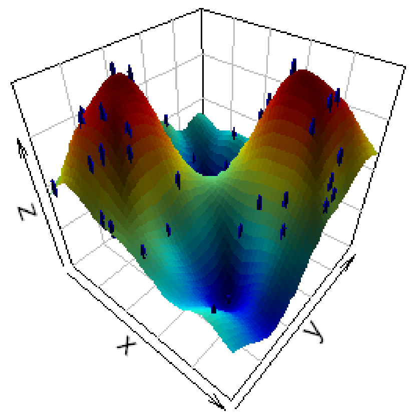
- Gaussian process regression is a method to learn a non-linear function **f(x)** from observations $\{(x_i, y_i)\}_{i=1..N}$
- It can be used for regression and classification problems

→ It is an alternative to other ML techniques, e.g., random forests and neural networks with its own pros & cons

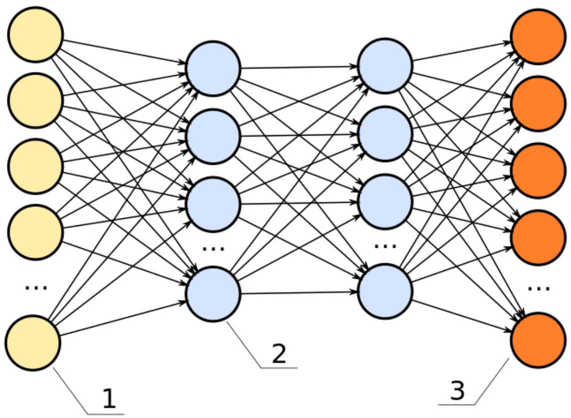


Comparison to neural networks

GP processes



artificial neural networks



<u>Both approaches</u> …

… are methods for classification and regression

… are universal function approximators

<u>Neural networks</u> …

… scale better to large data sets

… are able to capture non-local features

… are difficult to interpret

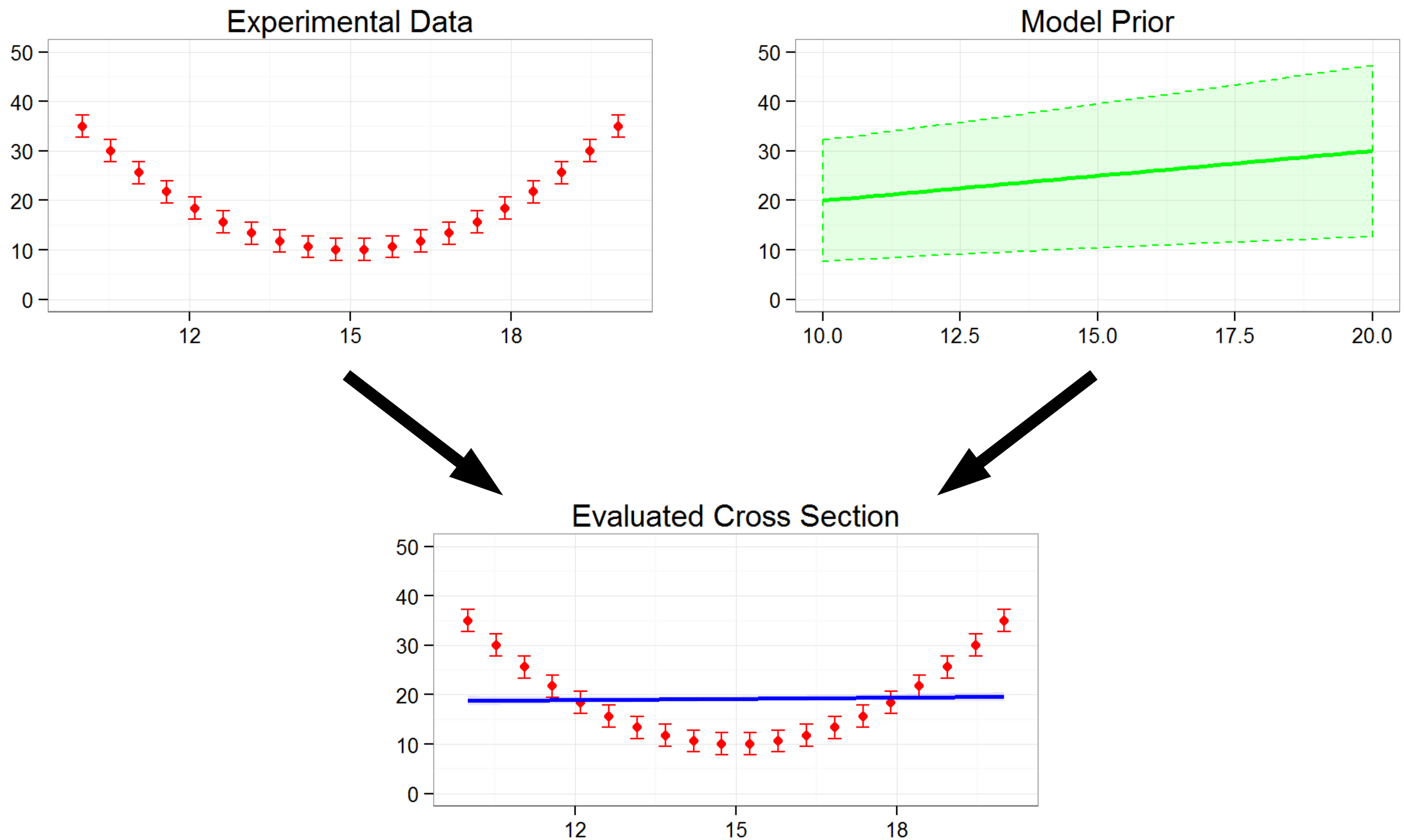
<u>GP processes</u> …

… are statistical methods from the ground up (uncertainties)

… facilitate the incorporation of prior assumptions

… interface well with existing nuclear data evaluation methods

Why we started talking about GPs



Linear model as Gaussian process

$$y(x) = kx + d \qquad \vec{p}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \mathbf{A}_0 = \begin{pmatrix} \delta_k^2 & 0 \\ 0 & \delta_d^2 \end{pmatrix}$$

$$M_{\text{lin}}[\vec{p}] = \mathbf{S}\vec{p} \qquad \mathbf{S} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \end{pmatrix}$$

$$\mathbf{S}(x) = \begin{pmatrix} x & 1 \end{pmatrix}$$

$$\kappa_{\text{mod}}(x_1, x_2) = \mathbf{S}(x_1)\mathbf{A}_0\mathbf{S}(x_2)^T$$

Relationship: linear models and GPs

- All linear models (splines, Fourier series, other polynomials, …) can be represented as a Gaussian process if one assigns a prior covariance matrix to the model parameters/coefficients
- Some Gaussian processes can be represented as a linear model with a finite number of parameters and a multivariate normal prior on the parameters
- All Gaussian processes can be represented as a linear model with an infinite number of parameters

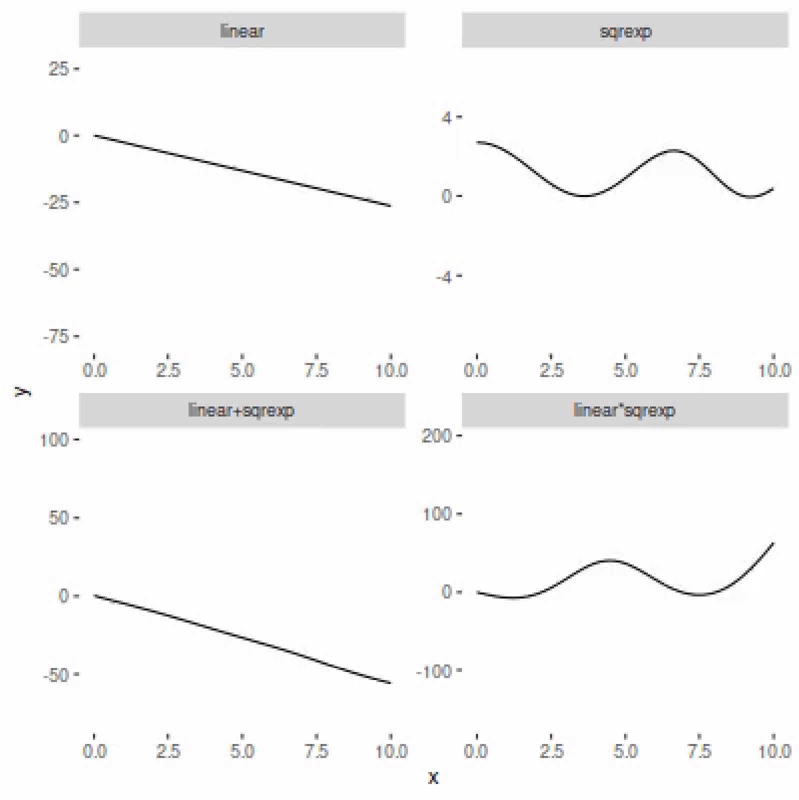
Clarification: e.g., $y(x) = ax^2 + bx$ is a linear model in our context because the predictions at all x are a linear function of the parameters a and b

Combination of Gaussian processes

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

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

Behavior of combined GPs



Play video

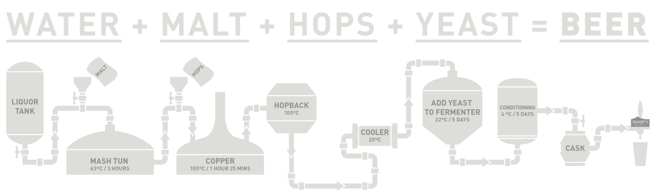
$$\kappa_{1+2}(x_1, x_2) = \kappa_1(x_1, x_2) + \kappa_2(x_1, x_2)$$ → Sum of random functions $\mathbf{f_1(x) + f_2(x)}$

$$\kappa_{1\times 2}(x_1, x_2) = \kappa_1(x_1, x_2) \times \kappa_2(x_1, x_2)$$ → Some kind of modulation

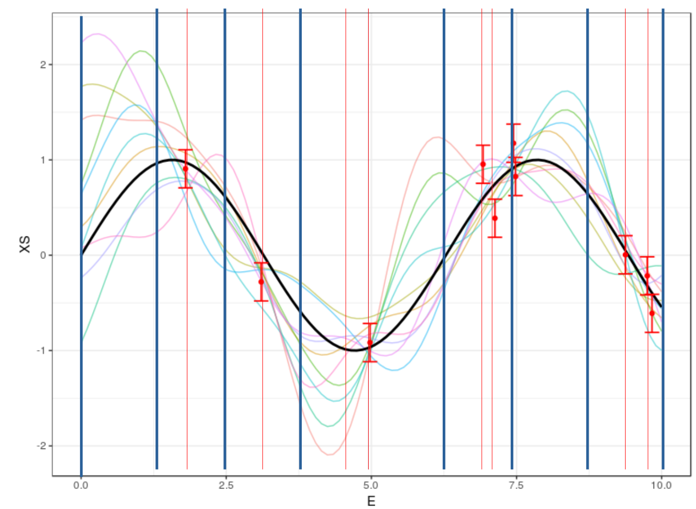
Disentangling Gaussian processes

$$\kappa(x_1, x_2) = \kappa_{\text{mod}}(x_1, x_2) + \kappa_{\text{def}}(x_1, x_2)$$

$$\mu(x) = \mu_{\text{mod}}(x) + \mu_{\text{def}}(x)$$



$$f(x) = f_{\text{mod}}(x) + f_{\text{def}}(x)$$



computational mesh **U** and experimental mesh **V**

Compound prediction on computational mesh **U** and associated covariance matrix:

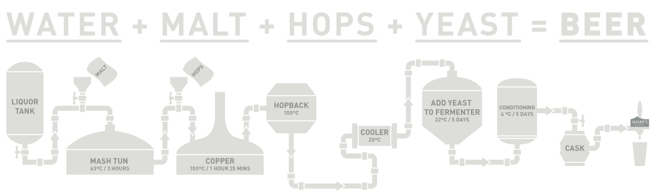
$$\vec{m}'_U = \vec{m}_U + \mathbf{K}_{UV} \left(\mathbf{K}_{VV} + \mathbf{B}\right)^{-1} \left(\vec{\sigma}_{\text{exp}} - \vec{m}_V\right)$$

$$\mathbf{K}'_{UU} = \mathbf{K}_{UU} - \mathbf{K}_{UV} \left(\mathbf{K}_{VV} + \mathbf{B}\right)^{-1} \mathbf{K}_{UV}^T$$

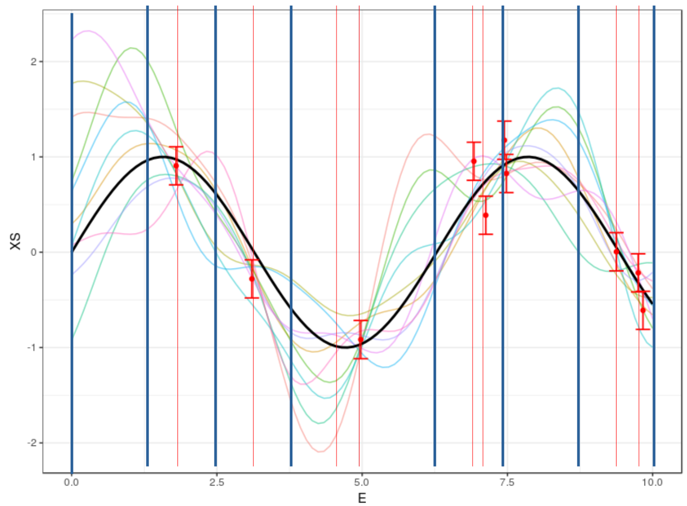
Disentangling Gaussian processes

$$\kappa(x_1, x_2) = \kappa_{\text{mod}}(x_1, x_2) + \kappa_{\text{def}}(x_1, x_2)$$

$$\mu(x) = \mu_{\text{mod}}(x) + \mu_{\text{def}}(x)$$



$$f(x) = f_{\text{mod}}(x) + f_{\text{def}}(x)$$



computational mesh **U** and experimental mesh **V**

Model prediction on computational mesh **U** and associated covariance matrix:

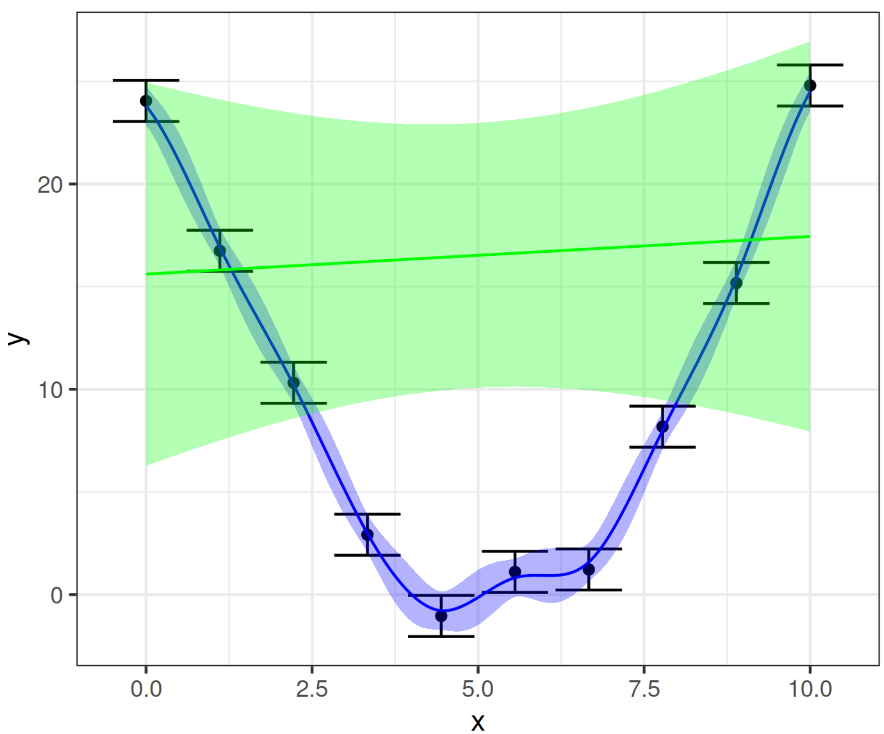
$$\vec{m}_U^{\text{mod}\prime} = \vec{m}_U^{\text{mod}} + \mathbf{K}_{UV}^{\text{mod}} \left(\mathbf{K}_{VV} + \mathbf{B}\right)^{-1} \left(\vec{\sigma}_{\text{exp}} - \vec{m}_V\right)$$

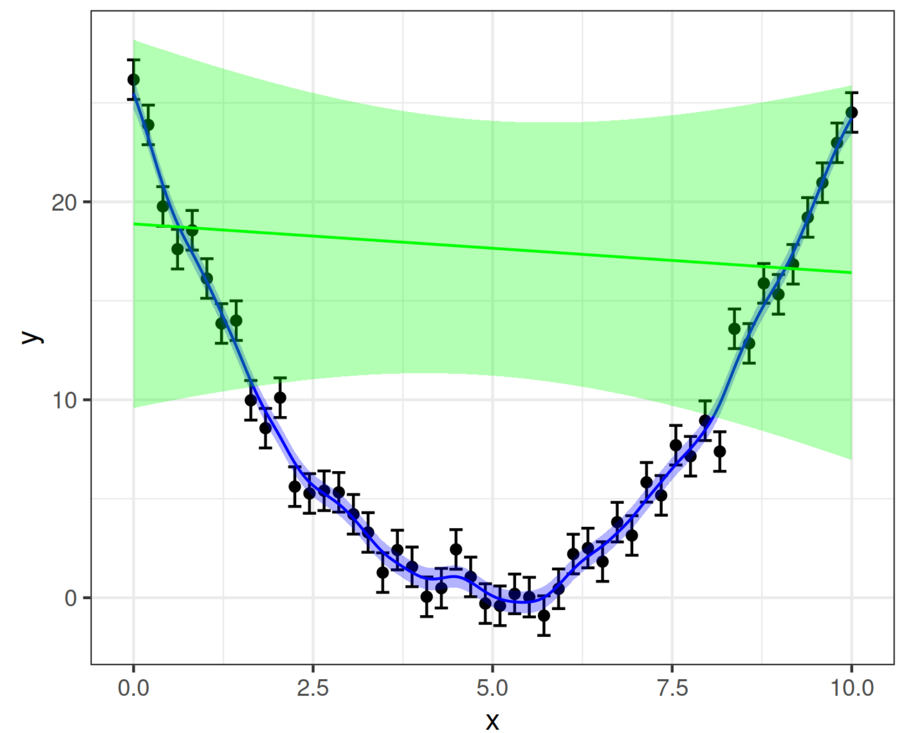
$$\mathbf{K}_{UU}^{\text{mod}\prime} = \mathbf{K}_{UU}^{\text{mod}} - \mathbf{K}_{UV}^{\text{mod}} \left(\mathbf{K}_{VV} + \mathbf{B}\right)^{-1} \left(\mathbf{K}_{UV}^{\text{mod}}\right)^T$$

Example of disentangled compound GP

$$\kappa_{\text{mod}}(x_1, x_2) := \kappa_{\text{lin}}(x_1, x_2) = x_1 x_2 {\delta_k}^2 + {\delta_d}^2$$

$$\kappa_{\text{def}}(x_1, x_2) := \kappa_{\text{sqrexp}}(x_1, x_2) = \delta^2 \exp\left(-\frac{1}{2\lambda^2}(x_1 - x_2)^2\right)$$





More conservative model prediction uncertainties

Recap

- Introduction to Bayesian statistics with multivariate normal model
- Common approaches in Nuclear Data to (approximately) evaluate the Bayesian update equation (GLS method)
- Gaussian processes as flexible modeling framework to fit functions which is mathematically compatible with the GLS method (Gaussian process regression = GLS with linear models with possibly infinitely many parameters)

Bayesian inference is a large research field and it is applied in many domains, e.g., natural language processing, image analysis, time series prediction, etc. and there is a vast world of possibilities beyond the GLS method, also for nuclear data!

Nuclear data challenges & perspectives

Regarding nuclear models

- Non-linear (nuclear physics is complicated)
- Not analytic (differential equations, simulation)
- Computationally expensive (minutes to hours)

Regarding linear algebra

- Computing with large covariance matrices

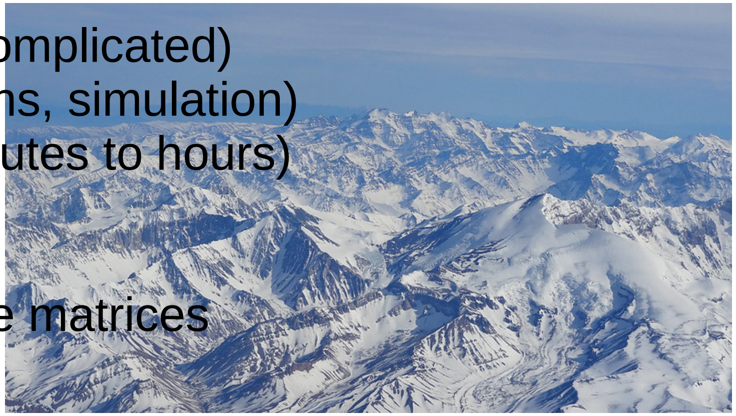
Regarding statistical models

- Imperfect physics model
- Multivariate normal distribution may be not always appropriate
- Uncertainties wrong or unknown

Machine Learning techniques

- Robust outlier detection
- Global approaches to cross section predictions over the nuclide chart
- Enhance traditional Bayesian evaluation techniques with ML methods

Nuclear data challenges & perspectives



Regarding nuclear models

- Non-linear (nuclear physics is complicated)
- Not analytic (differential equations, simulation)
- Computationally expensive (minutes to hours)

Regarding linear algebra

- Computing with large covariance matrices

Regarding statistical models

- Imperfect physics model
- Multivariate normal distribution may be not always appropriate
- Uncertainties wrong or unknown

Machine Learning techniques

- Robust outlier detection
- Global approaches to cross section predictions over the nuclide chart
- Enhance traditional Bayesian evaluation techniques with ML methods

References

- E. T. Jaynes “Probability Theory: The Logic of Science” (first chapters available online for free)
- C. E. Rasmussen & C. K. I Williams, “Gaussian processes for Machine Learning”, http://www.gaussianprocess.org/gpml/
- Presentation, videos and code at github.com/gschnabel/compnuc-workshop-2021