# Varying Model Parameters

V. Elvira, K. L. Genser, R. Hatcher, V. Ivanchenko, M. Kelsey, T. Koi, S. Y. Jun, A. Ribon, V. Uzhinsky, D. H. Wright, J. Yarba

in collaboration with

L. Fields and G. N. Perdue (Fermilab)

The 24th Geant4 Collaboration Meeting, Jefferson Lab

Step. 24, 2019, Plenary 4 (Physics Updates II)

# Contents


- Motivation
- Configurable model parameters and API
- Datasets and analysis framework
- Professor: a parameterization-based tuning tool
- Estimation of systematic uncertainties
- Ongoing work
- Summary and Future work

# Motivation

- User requirement: UR-29 (created Jan 21, 2016)
  - Reweightable uncertainties for systematic uncertainties estimation
- Description
  - Suggested based on features of GENIE Neutrino MC Generator
  - To allow estimation of model uncertainties with a single MC sample
- Initial scope
  - Introduce capabilities to change model parameters
  - Develop a tool kit for comparing Geant4 (with varied parameters) to thin target datasets
  - Develop a fitter that will optimize parameters to data and provide systematic uncertainties

# Configurable Model Parameters

- Number of configurable model parameters in FTFP_BERT
  - Since Geant4 (version) 10.4, an interfaces to configure some hadronic model parameters is provided to developers

| Model | Energy Range | Switches | Variables | Example parameter |
|---|---|---|---|---|
| Fritiof (FTF) | 3 GeV - 10 TeV | 4 | 18 | Baryon $\langle P_T^2 \rangle$ |
| Bertini Cascade | 0 - 12 GeV | 7 | 11 | Fermi Energy |
| Precompound | | 9 | 10 | Level Density |

  - More configurable parameters have been added for the 10.6 and potentially can be extended further (see Julia's talk at 3A)
- Documented in Book For Toolkit Developers: URL

[Changing Internal Parameters of an Existing Hadronic Model]

**Warning**

Changing these parameters without the guidance of the model developers may significantly alter or even degrade the model's physics performance. Any publication based on varied parameters must explicitly state what those values are, along with the physics list used and the GEANT4 version.

# Data Sets used for Optimization

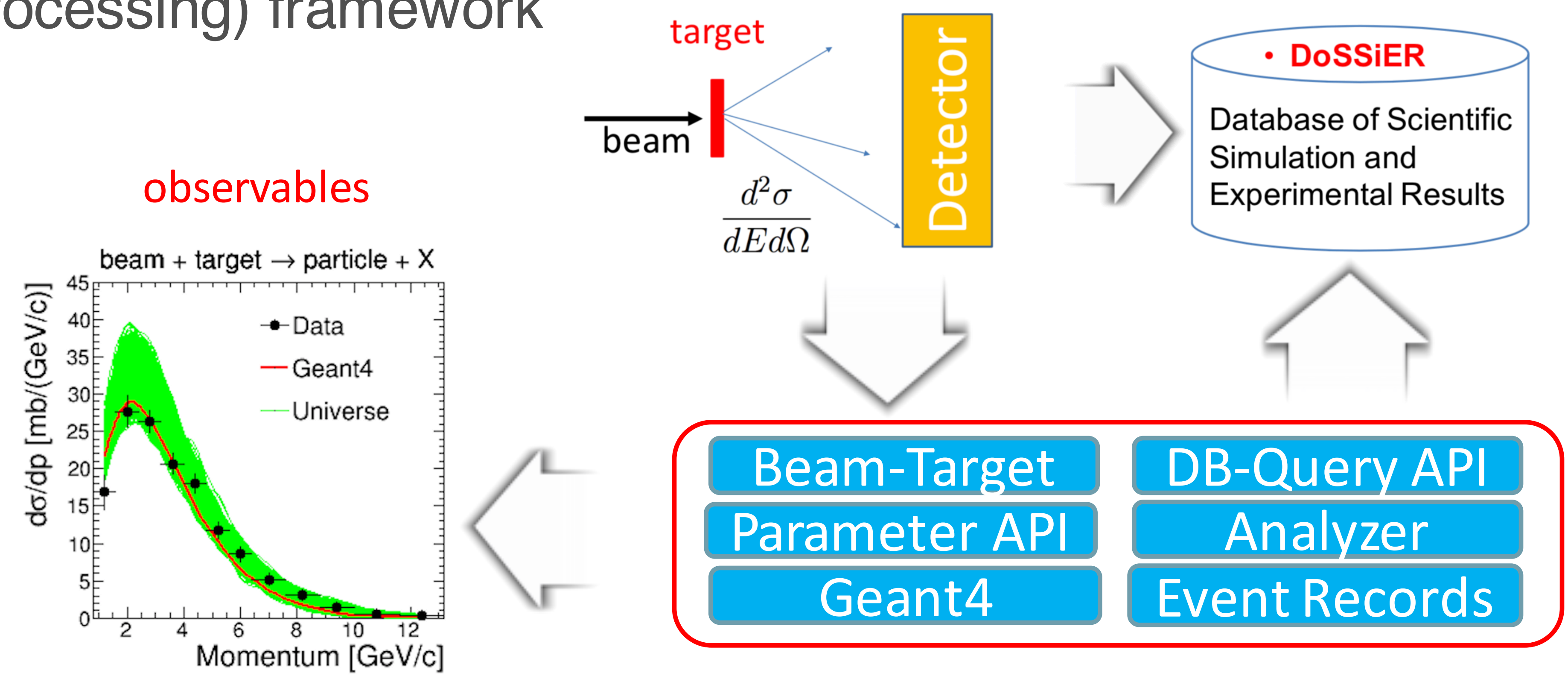
- A variety of experimental data sets
  - Thin target data sets
  - Data sets from general purpose detectors (isolated tracks)
  - Test beam data (*E/p*, energy resolution, shower profiles)
- Example of thin target data sets (publications at backup)

| Experiment | p-Beam (GeV) | Target | Production | Observables |
|---|---|---|---|---|
| IAEA | 0.8, 1.5, 3 | C, Al, Fe, In, Pb | $n$ | $d^2\sigma/dE_{KIN}d\theta$ |
| HARP | 5 | C, Pb | $\pi^\pm$ | $d^2\sigma/dpd\Omega$ |
| ITEP | 1,2,3,4,5,6,7,8.25 | C, Cu, Pb, U | $p$ | $d^3\sigma/d^3p$ |
| ITEP | 7.5 | Be, C, A, Ti, Fe, Cu Nb, Sn, Ta, Pb, U | $p$ | $d^3\sigma/d^3p$ |
| BNL E802 | 14.5 | Be, Al, Cu, Au | $p, \pi^\pm, K^\pm$ | $d^2\sigma/dM_Tdy$ |
| NA61 | 31 | C | $\pi^\pm, \pi^o, K, \Lambda, p$ | $d^2\sigma/dpd\theta$ |
| MIPP | 56.8, 57.3, 82.6, 120 | H, Be, C, Bi, U | $n$ | $d^2\sigma/dpdx_F$ |
| NA49 | 158 | C | $\pi^\pm, p, \bar{p}, n$ | $dN/dx_F$, $\langle p_T \rangle$ vs $d\theta$ |

| Experiment | $\pi^\pm$-Beam (GeV) | Target | Production | Observables |
|---|---|---|---|---|
| ITEP | 1,1.4,2,3,4,5,6 | C Cu Pb U | $\pi^\pm$ | $d^2\sigma/dpdpd\theta$ |
| HARP | 3,5,8,12 | Be, C, Al, Cu, Ta, Pb | $\pi^\pm$ | $d^2\sigma/dpd\theta$ |

# Simulation and Analysis

- J. Yarba has developed various analysis modules to compare Geant4 simulations to thin target data using the art (an event processing) framework




  - The event processing chain can be executed in one step, or it can be subdivided into multiple steps (parallel process)
  - (DoSSiER will need to be replaced)

# Optimization: Objectives and Prerequisites

- Optimization Objectives
  - Determination of values of parameters that describe datasets best
  - Estimation of model parameters uncertainties
- Prerequisites and items to consider
  - API for configurable model parameters
  - Optimization tools
  - Parameters to be optimized
    - sensitivities and stability ($\chi^2$ assumption, normal behavior)
    - avoid the curse of dimensionality
  - Data sets to be used
    - minimize experimental complication (thin target data)
    - cover sufficient phase space (no single representative data)
    - beam type, energy, target material
  - How to propagate uncertainties of parameters to observables?

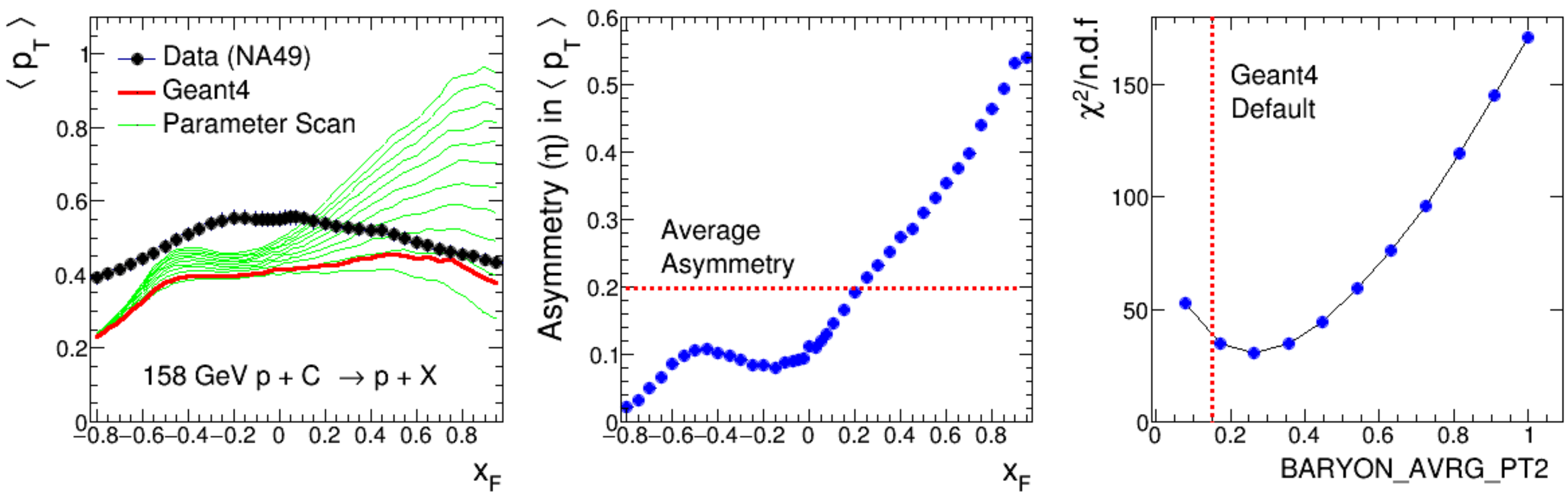
# Model Parameter Sensitivity and (Dependence) Stability

- Scan changes of observables within the allowed range
  - Quantify the derivatives of observables w.r.t. a parameter: the average max-min asymmetry for an approximation

$$\eta = \frac{1}{N_b N_o} \sum_{i}^{N_b} \sum_{o}^{N_o} \eta_i^o$$

where $\eta_i^o = \dfrac{\upsilon_{max} - \upsilon_{min}}{\upsilon_{\mathrm{min}} + \upsilon_{min}}$ for each bin (*b*) of each observable (*o*)

- $\chi^2$ scan: a good probe for the Gaussian prior probability




- Select a set of parameters ($N_p$) to optimize

# Professor (Procedure for estimating systematic errors)

- https://professor.hepforge.org/
- Interpolation: fits observables of the universe to the *n*-degree of polynomial to produce the response function($f$) that predicts data distribution for any set of parameters

$$\text{parameters } (\vec{p}) \overset{f}{\Longleftrightarrow} \text{observables } (\mathcal{O}_i)$$

- Optimization: use response functions to obtain parameters which minimize the goodness of fit (e.g., $\chi^2$) with respect to datasets

$$\frac{\partial}{\partial \vec{p}} \sum_{\mathcal{O}_i=1}^{datasets} w_{\mathcal{O}_i} \sum_{b \in \mathcal{O}_i}^{nbin} \frac{(f(\vec{p}) - v_d)^2}{\epsilon_b^2} = 0 \longrightarrow (\vec{p}_o, \sigma, \rho)$$

- $v_d$ $(\epsilon_b)$ : value (error) of the reference bin
- $\sigma(\rho)$: error (correlation) of the tuned parameter vector $(\vec{p}_o)$

# Generating Tuning Samples

- Select data sets relevant for the energy range of a selected model or models
- Select $N_p$ parameters that significantly affect observables
- Generate $N_s$ samples by varying $N_p$ parameters randomly within allowed limits

$$\vec{p}_i = \vec{p}_{min} + (\vec{p}_{max} - \vec{p}_{min}) \cdot \vec{u}(0,1)$$

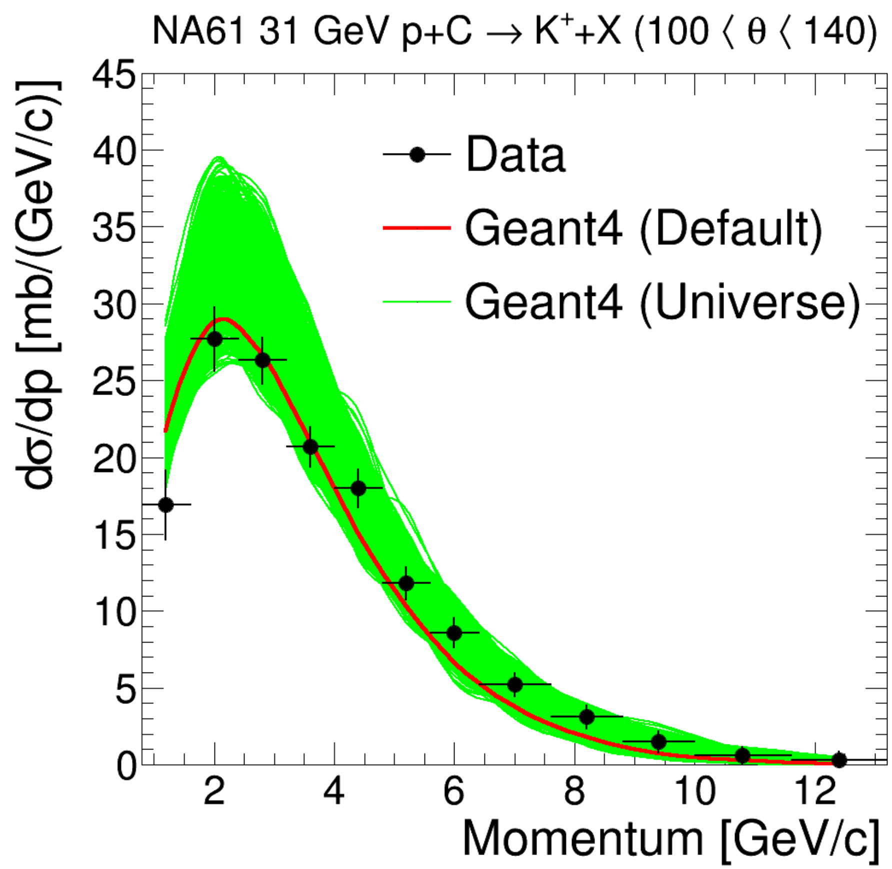
  - Statistical uncertainty of simulation << data error
  - The minimum number of samples ($N_s$) for the $n^{th}$ order polynomial interpolation of $N_p$ number of parameters

$$N_S > N^{(n)}(N_p) = 1 + \sum_{i=1}^{n} \frac{1}{i!} \prod_{j=0}^{i-1} (N_p + j)$$

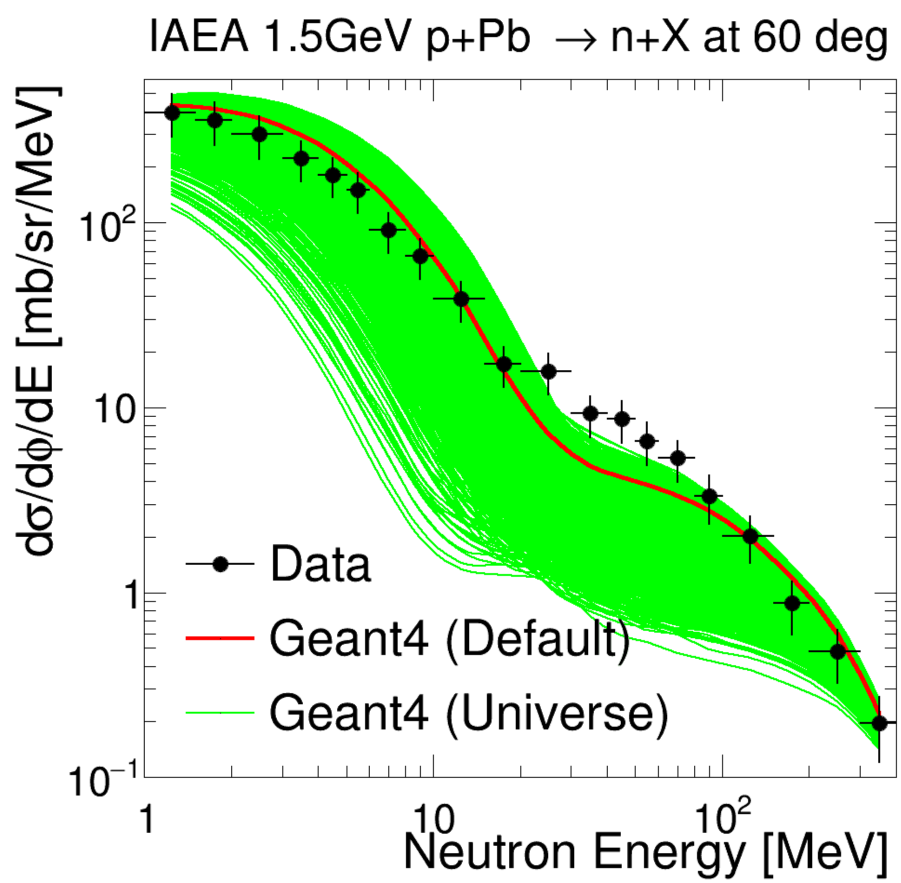
| $N_P$ | $n = 2$ | $n = 3$ |
|---|---|---|
| 4 | 15 | 35 |
| 8 | 45 | 165 |
| 16 | 153 | 969 |

# Generating MC Samples:

- Examples of the spread (green lines) of 1000 samples
  - universe: simulation varying parameters within allowed limits




Variation of 10 FTF parameters

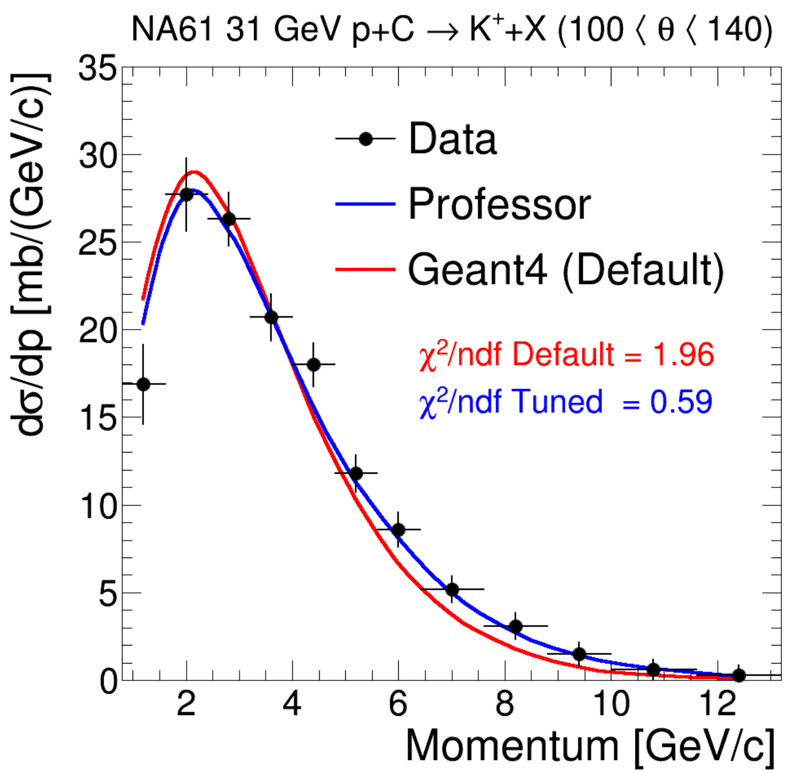



Variation of 4 Bertini parameters

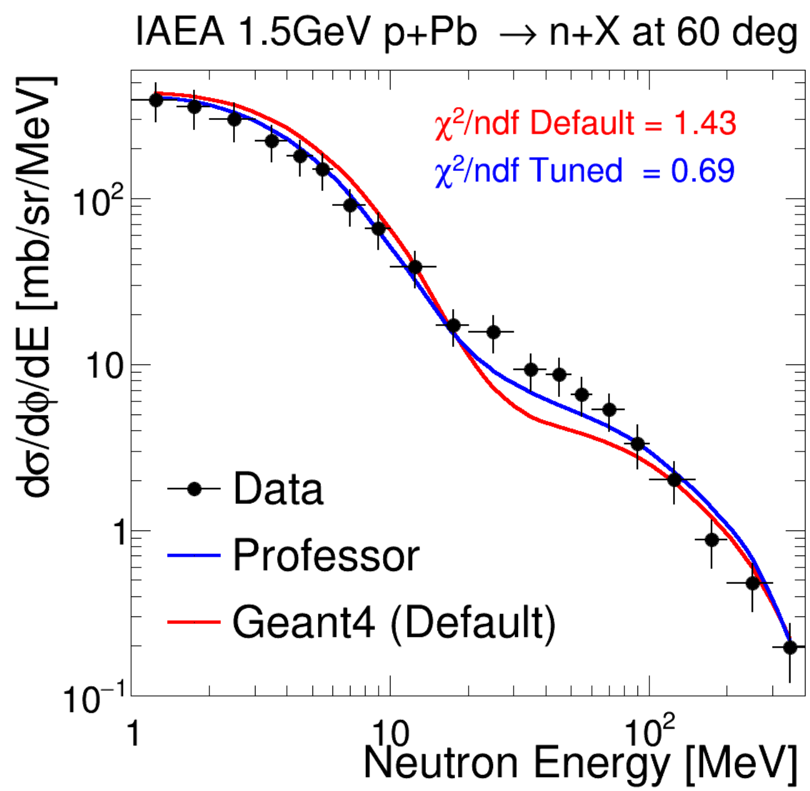
# Professor Fit Result

- Examples of Professor predictions based on thin target data

Best fit using one data set




Global fit using multiple datasets




  - Visible improvement in $\chi^2$
  - A general question is how to choose a set of model parameters and datasets for an optimization study in an unbiased way

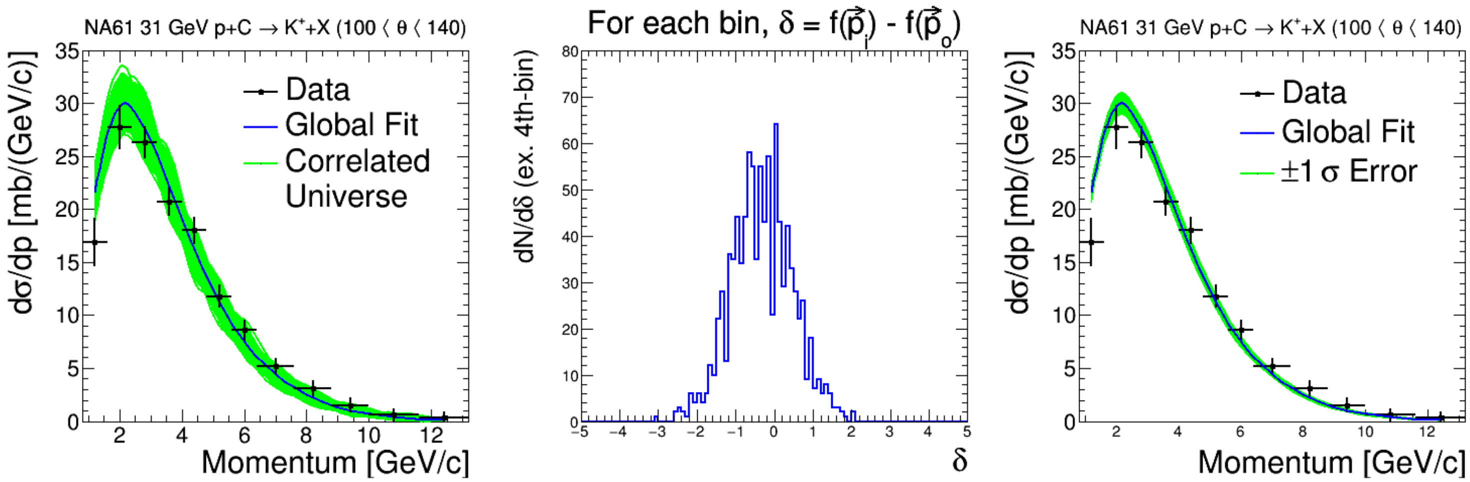
# Estimation of Uncertainty (An example method)

- Generate samples with a set of correlated parameters according to optimized ($\boldsymbol{p}_o$), error($\boldsymbol{\sigma}$) and correlation ($\boldsymbol{\rho}$)

$$\vec{p}_i = \vec{p}_o + \sigma \mathcal{C} \vec{z}_i$$

- C: Cholesky decomposition of $\rho$
- $z_i$: normal random numbers

- Estimate the RMS error for each bin of observables
  - Propagating parameter uncertainty to observables



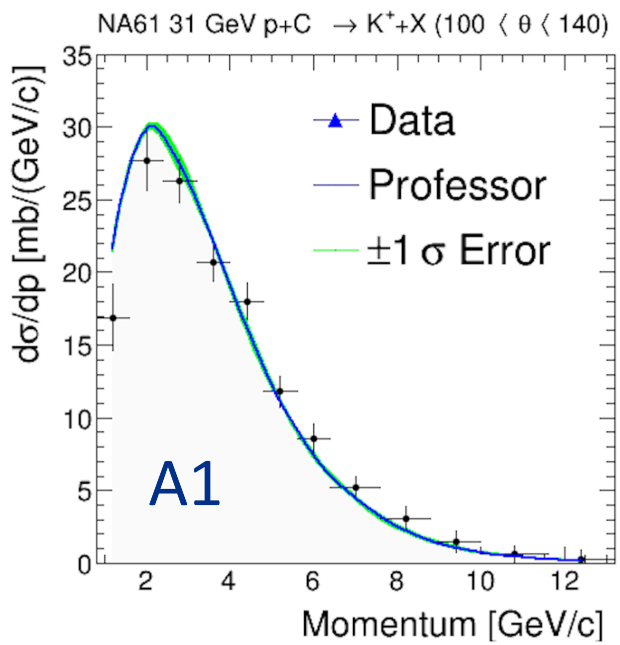

  - This includes the statistical fluctuations due to the divergence of random sequences

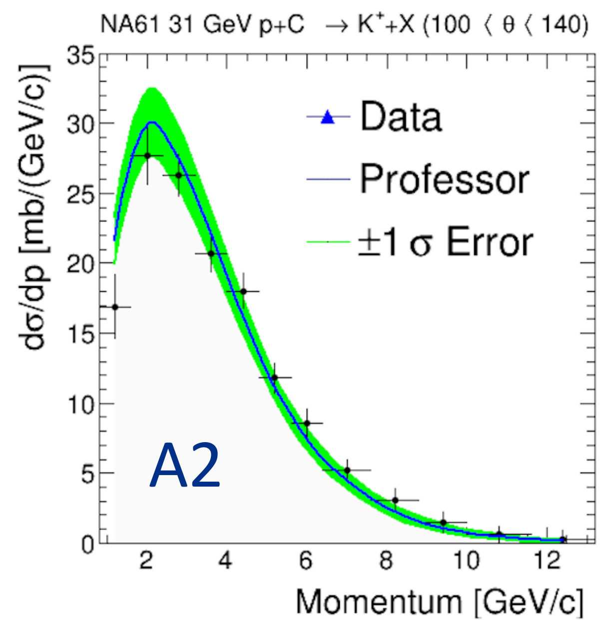
# Estimation of Uncertainty: Alternative Methods

- A1: Professor fit → mean($\boldsymbol{p}_o$), error($\boldsymbol{\sigma}$) and correlation ($\boldsymbol{\rho}$)
  - Use $f(\boldsymbol{p}_i)$ (response function) to predict values of observables

$$\boldsymbol{p}_i = \boldsymbol{p}_o + \boldsymbol{\sigma}\ \boldsymbol{C}\ \boldsymbol{z}_i$$

  - Estimate the RMS error for each bin of an observable



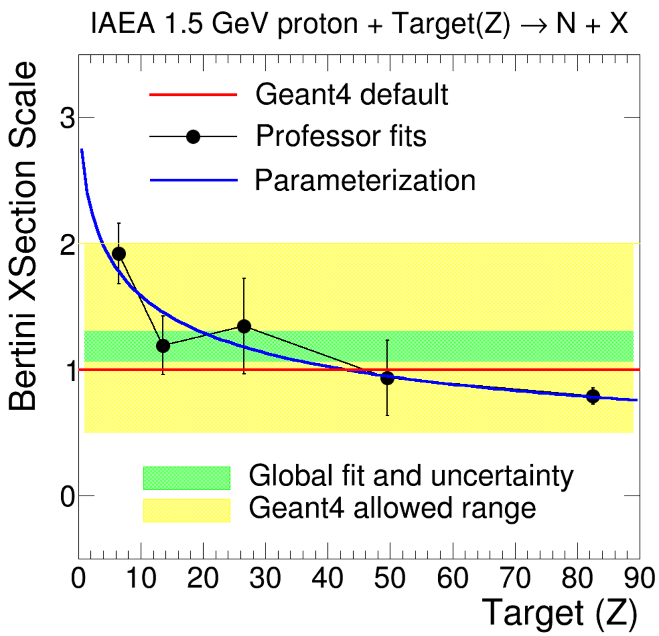




- A2: use mean($\boldsymbol{p}_o$), correlation ($\boldsymbol{\rho}$) from the professor fit, and the default ranges of parameters defined in Geant4 models

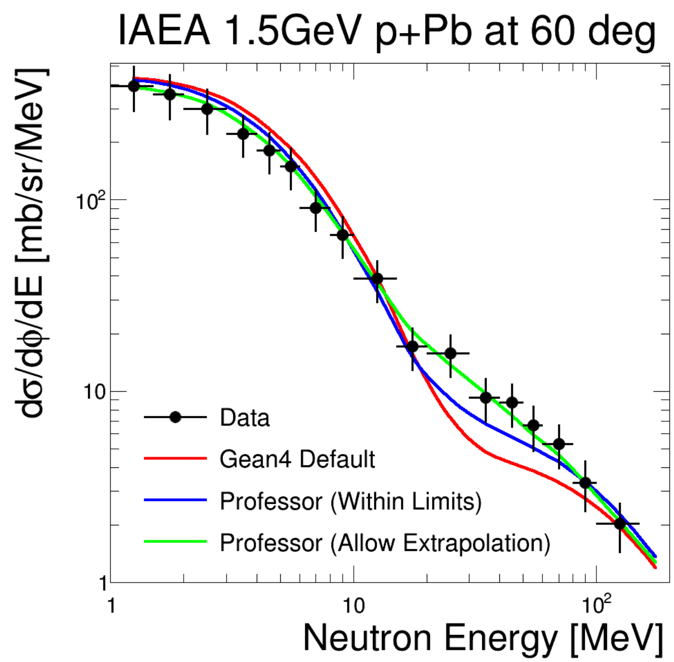
# Ongoing and Future Work

- Ongoing work: going beyond fit results: examples

Target Z-dependence?




Extend parameter limits?




- Future work
  - Enabling tuning of more model parameters
  - Adding more data sets and treatment of correlated errors in data
  - Inclusion of cross sections in tuning
  - Identification of reweightable parameters

# Summary

- An interfaces to configure some hadronic model parameters was provided to developers and is being extended
- Demonstrated benefits of model parameters optimization using Professor
- Discussed how to propagate uncertainties of parameters to experimental observables
- This is work in progress, much more can be done
- Recent publications
  - ACAT19 proceedings: submitted to J. Phys. Conf. Series
  - A detailed description of the study is ready to be submitted to JINST

- See more detailed talks in session 3A!

# Experimental Datasets used for this study

- HARP: M. Apollonio et al., Nucl. Phys. A821 118, 2009
  M. Apollonio et al., Phys.Rev.C80 065207, 2009
  M. Apollonio et al., Phys.Rev.C80 035208, 2009
  M.G. Catanesi et al., Phys.Rev.C77 055207, 2008
- IAEA/Ishibash: JNST, Vol. 34, No. 6, p. 529-537 (June 1997)
- ITEP771: Yu.D.Bayukov et al, Preprint ITEP-148-1983
- Sov.J.Nucl.Phys. 42 116, 1985
- BNL E-802: T. Abbott et al, Phys.Rev. D45, 3906 (1992)
- NA61: N. Abgrall et al., Eur.Phys.J. C76 (2016) no.2, 84
- SAS M6E: D.S. Barton et al., Phys. Rev. D27, 2580 (1983)
- NA49: [http://spshadrons.cern.ch/spshadrons](http://spshadrons.cern.ch/spshadrons)