Tuning the GENIE neutrino event generator with Professor

Julia Tena Vidal at Tel Aviv University
on behalf of the GENIE Collaboration

Workshop on Neutrino Event Generators
15^{th}-16^{th} March 2023

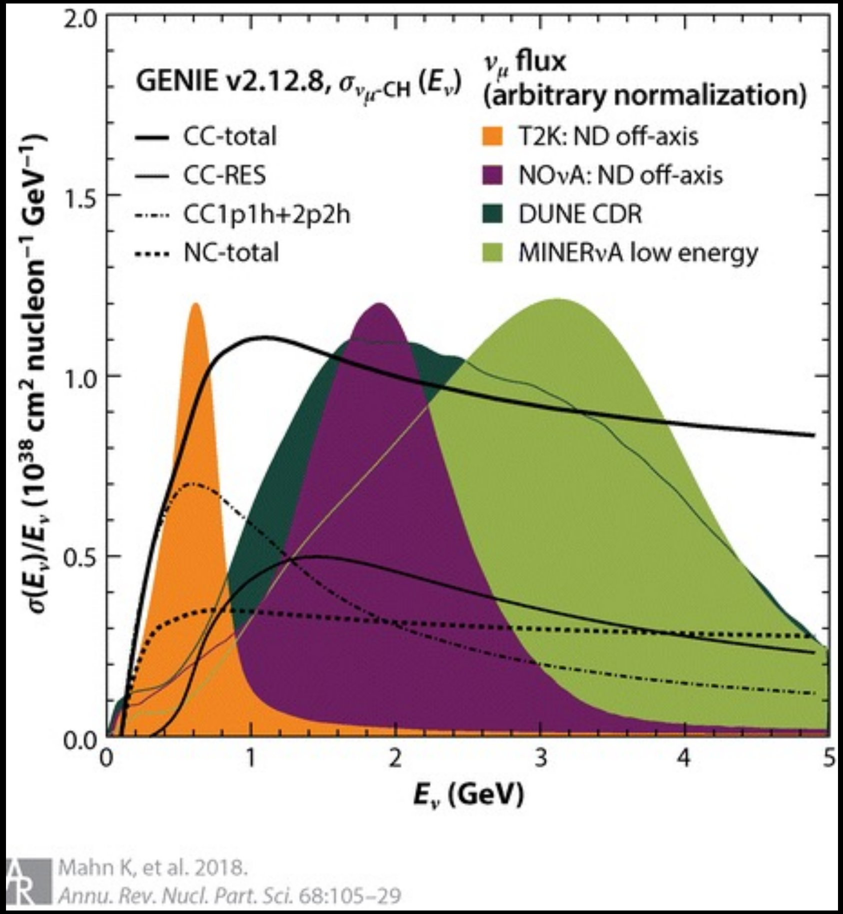
TEL AVIV UNIVERSITY אוניברסיטת תל אביב

Genie

Simulating neutrino interactions for accelerator experiments

Neutrino experiments rely on simulations to **reconstruct the neutrino energy** and **estimate systematic uncertainties**

- **Models are not complete**
  - Limited phase space coverage
  - Do not handle transition between different kinematic regions
  - Can fail to describe hadronic final states
  - Nuclear effects are factorized out
- **Lack of model systematic uncertainties estimates from theorists**



The GENIE Event Generator

- GENIE seeks to model neutrino interactions for all probes and targets across the kinematic region of interest for neutrino experiments
- It provides a number of comprehensive model configurations (CMCs)
  - Each configuration consists of a self consistent set of interaction models
  - Theoretical inputs are used to simulate specific processes at specific parts of the phase space
  - **Relies on empirical approaches to extrapolate in other kinematic regions and provide exclusive final-state information**

Examples of empirical models in GENIE CMC’s

- **Parameterization of vector and axial form factors**
  - Electron and neutrino scattering data
- **Simulation of pion production on free nucleon interactions**
  - Hydrogen and deuterium data from bubble chamber experiments
- **Neutrino induced hadronization (AGKY model)**
  - Hadronization data from bubble chamber experiments
- **Final-state interaction models**
  - Hadron-nucleus data

Lack of MC/Data description

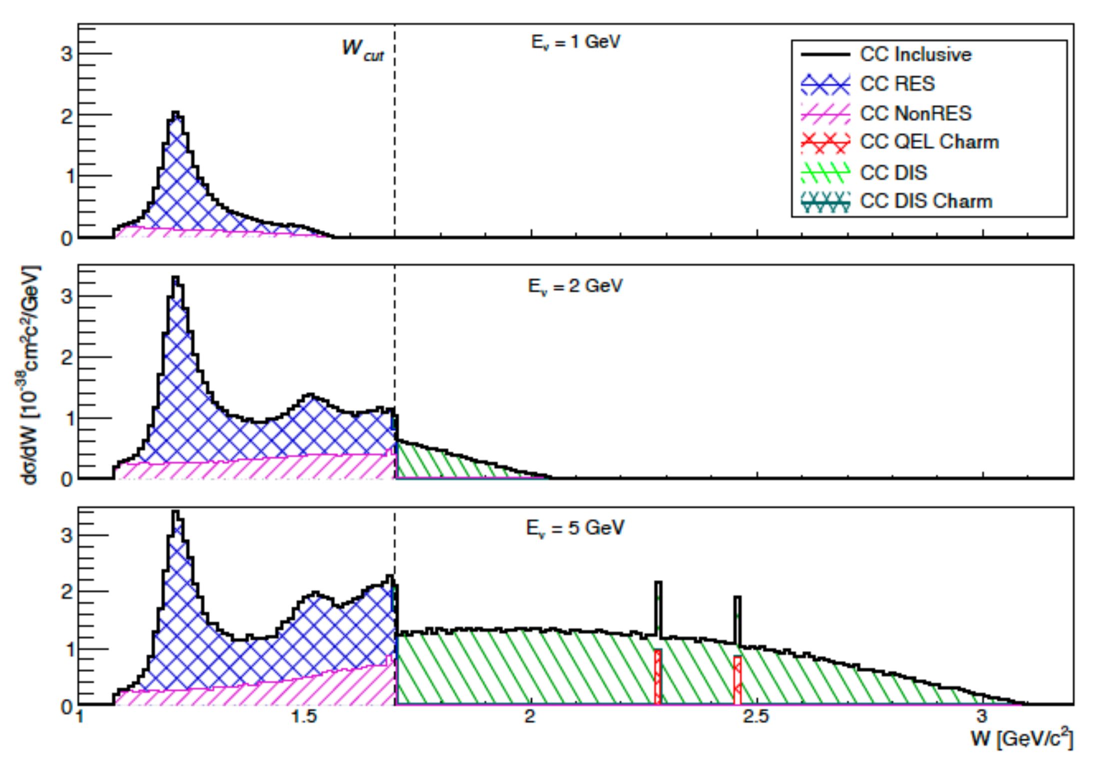
- The Shallow Inelastic Scattering (SIS) region is modelled as:

PhysRevD.104.072009

$$\frac{d^2\sigma^{inel}}{dQ^2dW} = \begin{cases} \frac{d^2\sigma^{RES}}{dQ^2dW} + \frac{d^2\sigma^{Non-RES}}{dQ^2dW} & for\ W < \boldsymbol{W_{cut}} \\ \frac{d^2\sigma^{DIS}}{dQ^2dW} & for\ W \geq \boldsymbol{W_{cut}} \end{cases}$$

$$\frac{d^2\sigma^{Non-RES}}{dQ^2dW} = \frac{d^2\sigma^{DIS}}{dQ^2dW} \cdot \Theta(\boldsymbol{W_{cut}} - W) \cdot \sum_m \boldsymbol{f_m}(Q^2, W)$$

Free parameters



- The Resonant contribution is modelled from Rein-Sehgal or Berger-Sehgal
- The non-resonant background is modelled using the Deep-Inelastic-Scattering model (Bodek and Yang) extrapolated at $W < W_{cut}$

PhysRevD.104.072009

Lack of MC/Data description

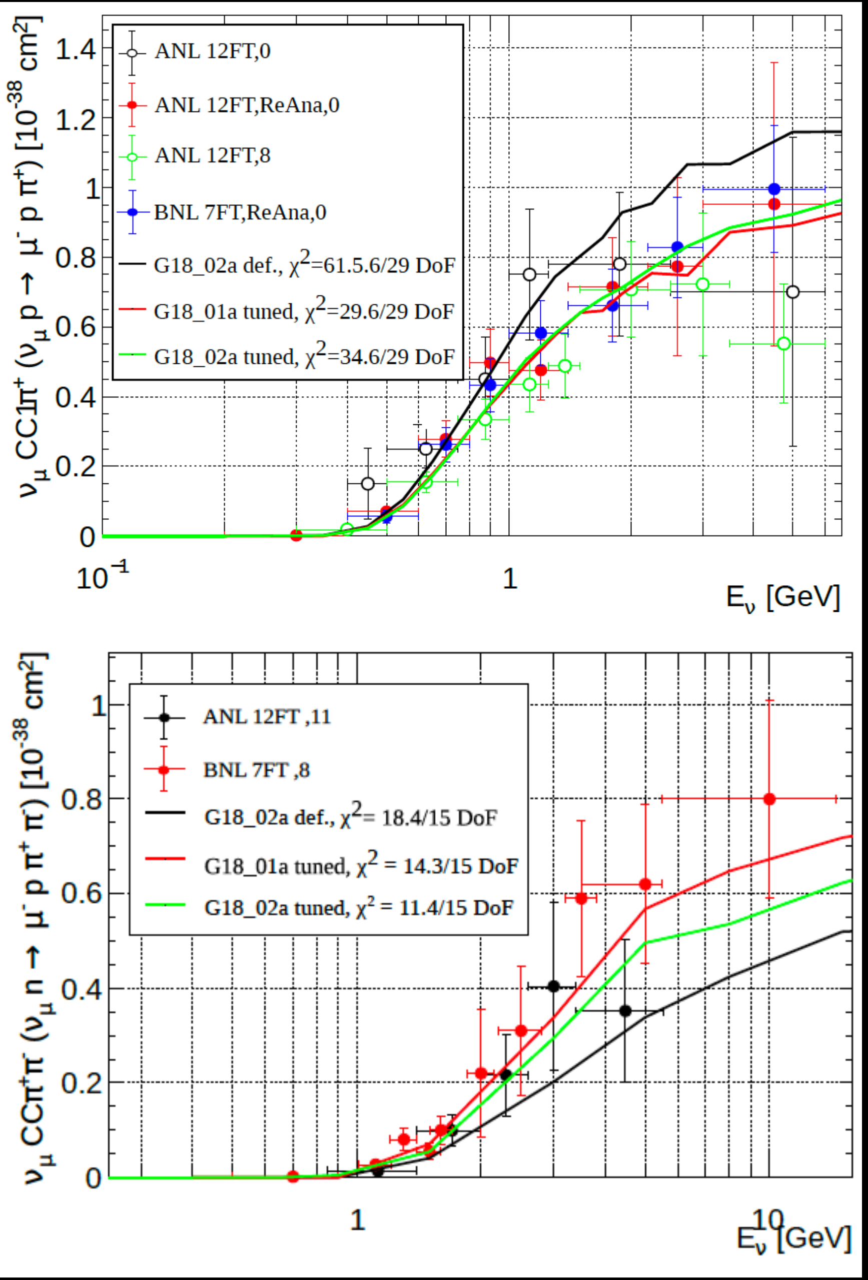
- The Shallow Inelastic Scattering (SIS) region is modelled as:

$$\frac{d^2\sigma^{inel}}{dQ^2dW} = \begin{cases} \frac{d^2\sigma^{RES}}{dQ^2dW} + \frac{d^2\sigma^{Non-RES}}{dQ^2dW} & for\ W < \boldsymbol{W_{cut}} \\ \frac{d^2\sigma^{DIS}}{dQ^2dW} & for\ W \geq \boldsymbol{W_{cut}} \end{cases}$$

$$\frac{d^2\sigma^{Non-RES}}{dQ^2dW} = \frac{d^2\sigma^{DIS}}{dQ^2dW} \cdot \Theta(\boldsymbol{W_{cut}} - W) \cdot \sum_m \boldsymbol{f_m}(Q^2, W)$$

Free parameters

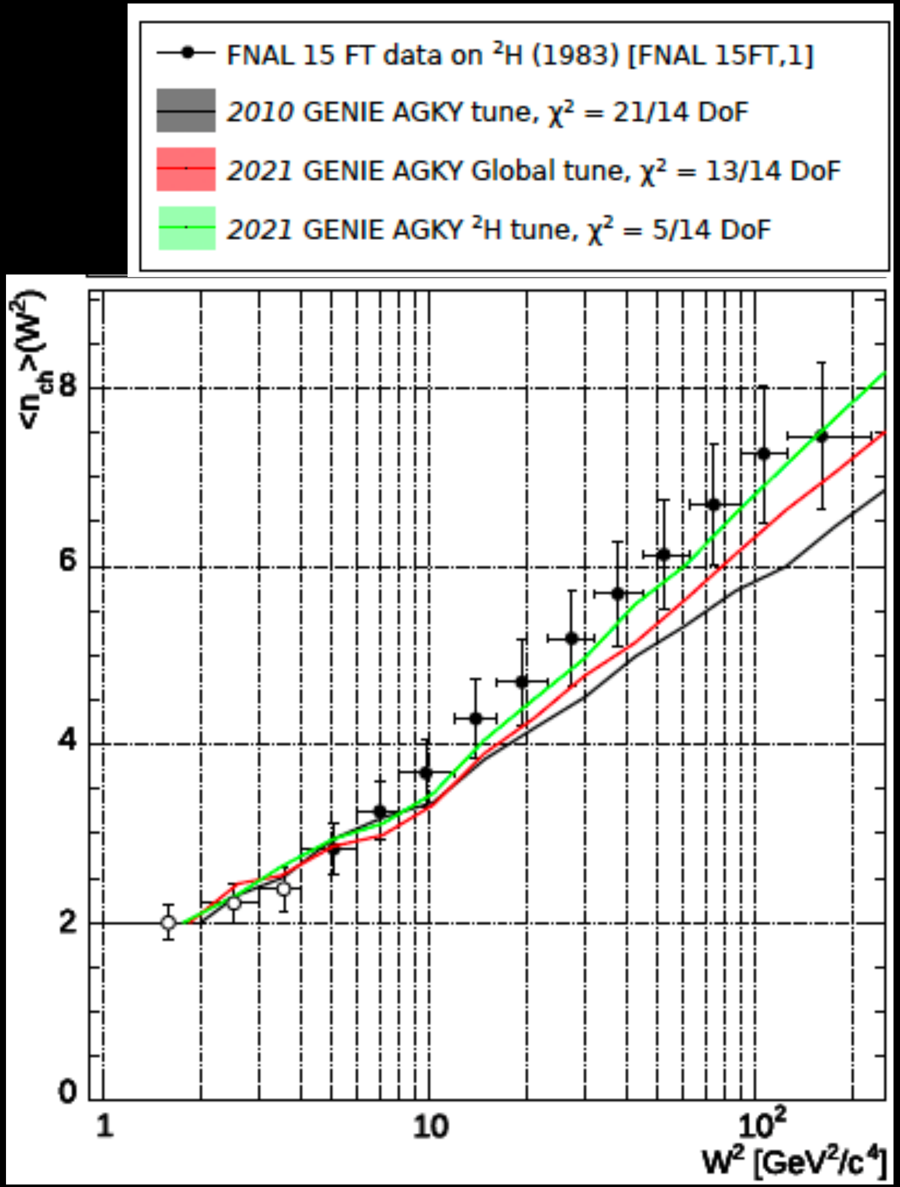
- **Tuning is essential to address double counting**
- Untuned GENIE (G18_02a def) fails to describe one and two pion production data on free nucleon



PhysRevD.105.012009

Lack of MC/Data description

- Hadronization is handled with the AGKY model [Eur.Phys.J.C63:1-10,2009]
  - Low-W empirical model for SIS/DIS events at W<2.3 GeV/c^2
  - PYTHIA 8 for events with W>3 GeV/c^2
  - Transition region in the 2.3<W<3 GeV/c^2 region
    - Probability of using Low-W with respect to PYTHIA changes linearly with W
- **The untuned model fails to describe neutrino hadronization data and it does not provide with model uncertainties**
  - AGKY can't be tuned using standard Reweighting procedures



GENIE's global analysis with Professor

GENIE develops a global analysis of scattering data for tuning and uncertainty characterization of comprehensive neutrino interaction models

- **Goals:**
  - Perform a global tune that improves data/MC agreement
  - Avoid double-counting in transition region
  - Estimate data-driven uncertainties
- **Requirements:**
  - Tune all aspects of event generators, **including non-reweightable processes**
  - The software can accommodate neutrino, hadron nucleon and electron scattering data
- **Summary of previous GENIE tunes:**
  - Neutrino-Nucleon Cross-Section Model Tuning in GENIE v3 [PhysRevD.104.072009]
  - Hadronization Model Tuning in GENIE v3 [PhysRevD.105.012009]
  - Neutrino-nucleus CC0π cross-section tuning in GENIE v3 [PhysRevD.106.112001]

This talk focuses on the global analysis software

GENIE’s global analysis with Professor

- The GENIE global analysis software is based on the Professor concept
  - Professor is extensively used in the collider community
- GENIE’s global analysis software is adapted to the needs of the neutrino community
  - Has diversified from the standard Professor software
- The GENIE Generator package interfaces with GENIE tuning software via GENIE Comparisons software
- Professor is not available in UPS but it can be installed on the FNAL grid
- **The Professor concept can be used to tune other event generators**

https://professor.hepforge.org

GENIE’s global analysis with Professor

- Professor’s concept:

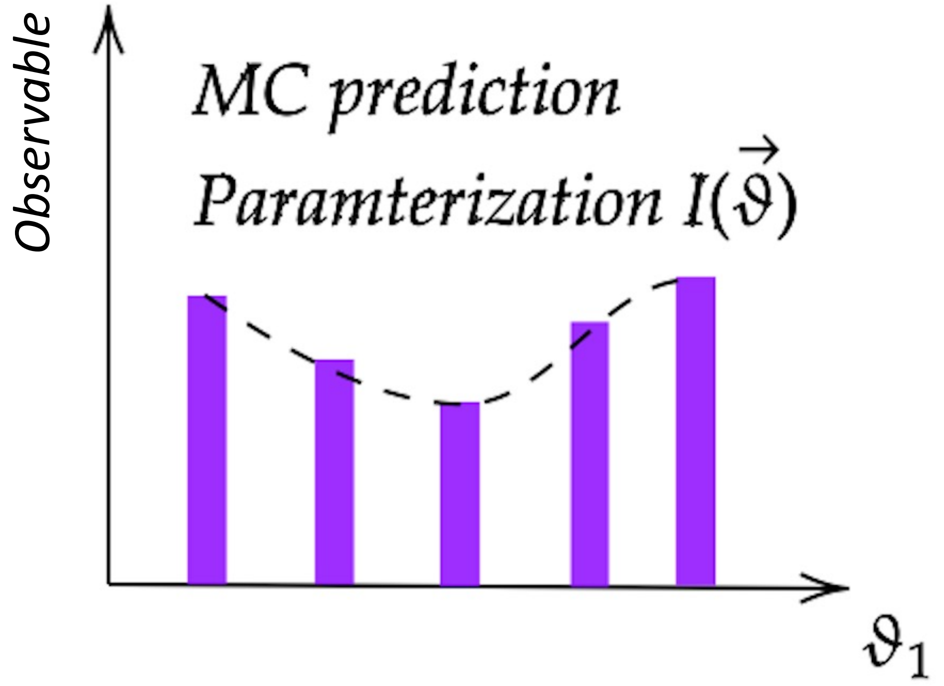
  Parameterization instead of full MC

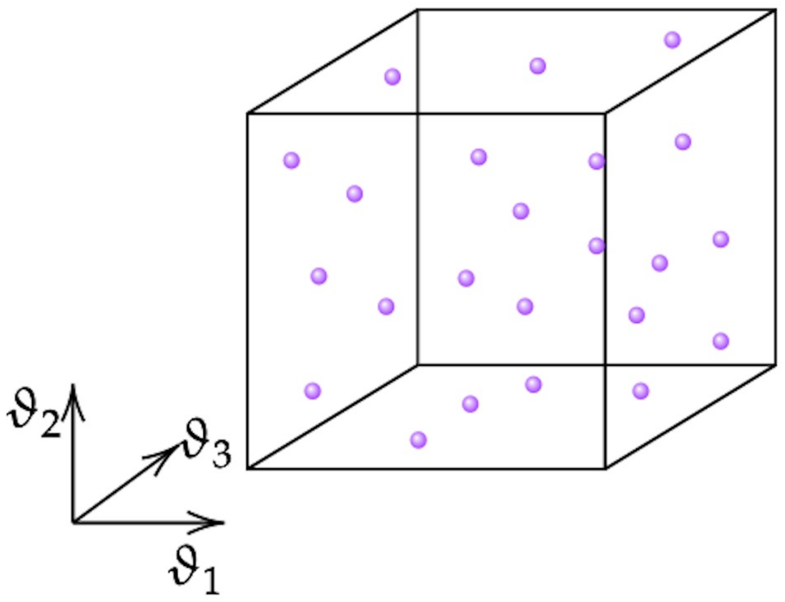
- Main take-aways:
  - The MC response for a given observable is parameterized as a function of the parameters phase-space
    - We call this the response function
  - The minimization of the likelihood is done with the parameterized response function

https://professor.hepforge.org

General Tuning Workflow

1. Select the parameters to tune and define physical phase-space
2. Define your observable in terms of bins
3. Evaluate observable bin's behaviour with brute force
4. Use sampled points to parameterize bin's behaviour with an N-Dimensional polynomial
5. Minimize using parameterization instead of full MC





Selection of parameters of interest

- The selection of parameters will depend on the interaction process you want to tune
- **The parameters to tune are directly used in the cross-section computation**
- For instance, we might want to tune M_A^{QEL}:

CommonParam.xml used in spline generation:

```
<param_set name="QuasiElastic">
  <!--
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Value of axial form factor at Q2=0
  -->
  <param type="double" name="QEL-FA0"> -1.2670 </param>

  <!--
       ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
       Axial and vector masses for quasi-elastic scattering
  -->
  <param type="double" name="QEL-Ma"> 0.961242 </param>
  <param type="double" name="QEL-Mv"> 0.840 </param>

</param_set>
```

Structure used for Professor-based sampling of phase-space:

<param type=”double” name = ”QEL-Ma”> {QEL-Ma} </param>

Selection of parameters of interest

- The selection of parameters will depend on the interaction process you want to tune
- **The parameters to tune are directly used in the cross-section computation**

- **Benefits of this approach:**
  - Any parameter involved in the cross-section or event generation algorithms can be tuned
  - Can tune non-reweightable parameters
    - Hadronization, final-state interaction (including cascade FSI models), etc…
  - New parameterizations are added directly in the GENIE Generator
  - The results of the tune can be easily included in GENIE CMC's to be run out-of-the-box by users
    - Complex configurations are handled with tune tags: Example of nuclear tune configuration (GPRD18_10a)

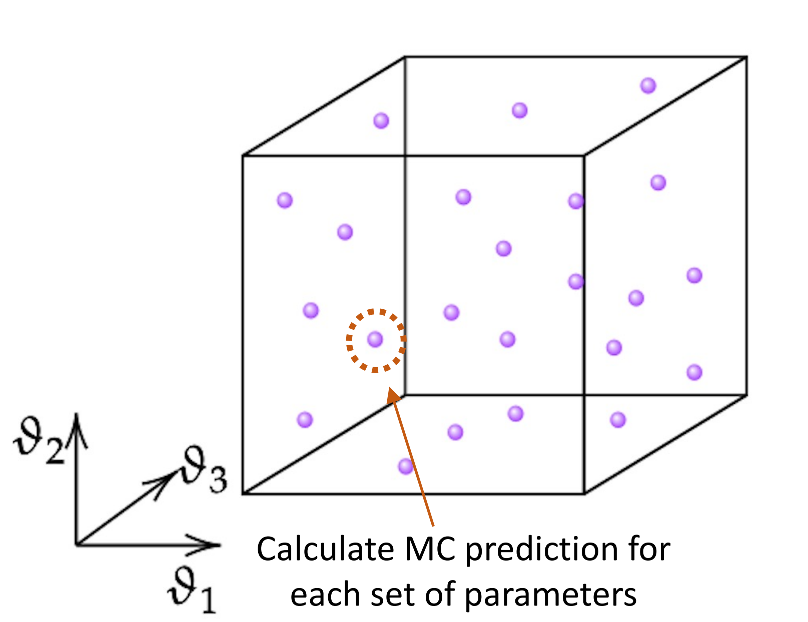
Sampling of phase-space

- Once the set of parameters is selected $(\vartheta_1, \vartheta_2, \ldots, \vartheta_{N_\vartheta})$, the next step is to define the parameters phase-space
  - Ideally, the best-fit result should lie around the middle of the phase-space
- In order to parameterize the response-function with an N-dimensional polynomial, we uniformly sample the phase space with

$$N_{MC\ Samples} = \frac{(N_\vartheta + N)!}{N_\vartheta!\,N!} \cdot 1.5$$

N_ϑ	4th order polynomial	5th order polynomial
2	22	31
5	189	378
10	1500	4500
13	3570	12852

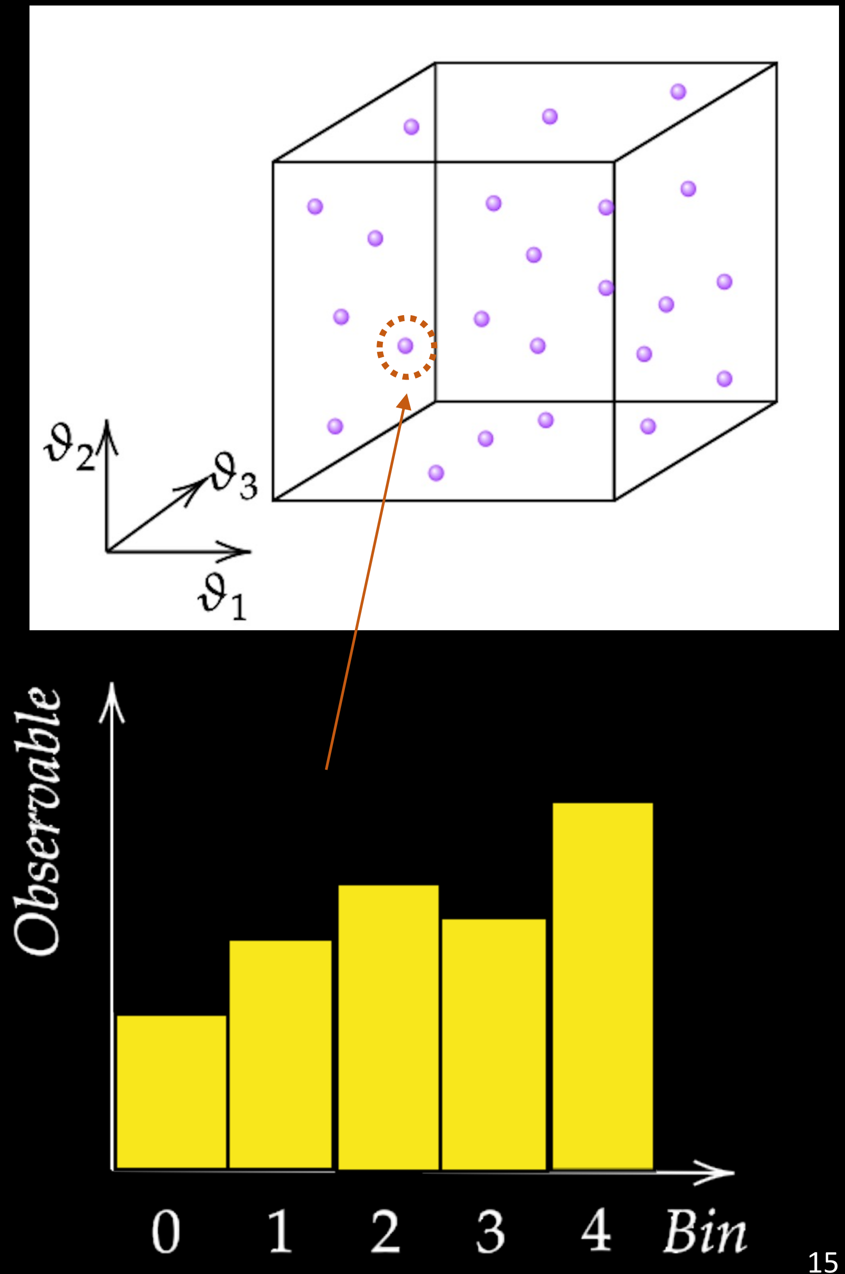
N_ϑ dimensions phase-space



The generation of all the samples is the most expensive CPU expensive step
It can be easily parallelized to minimize computing time
It happens before the actual fit (which takes few minutes to run)

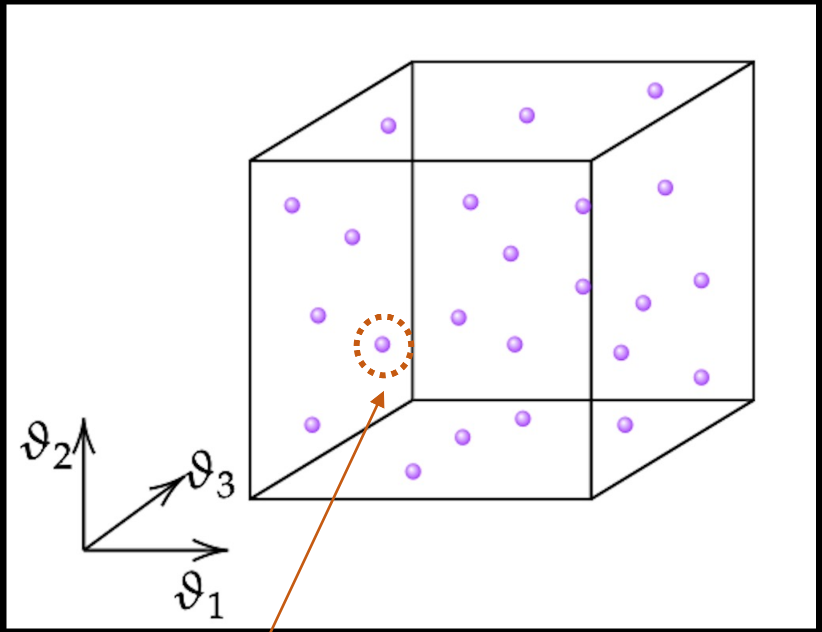
Definition of observable

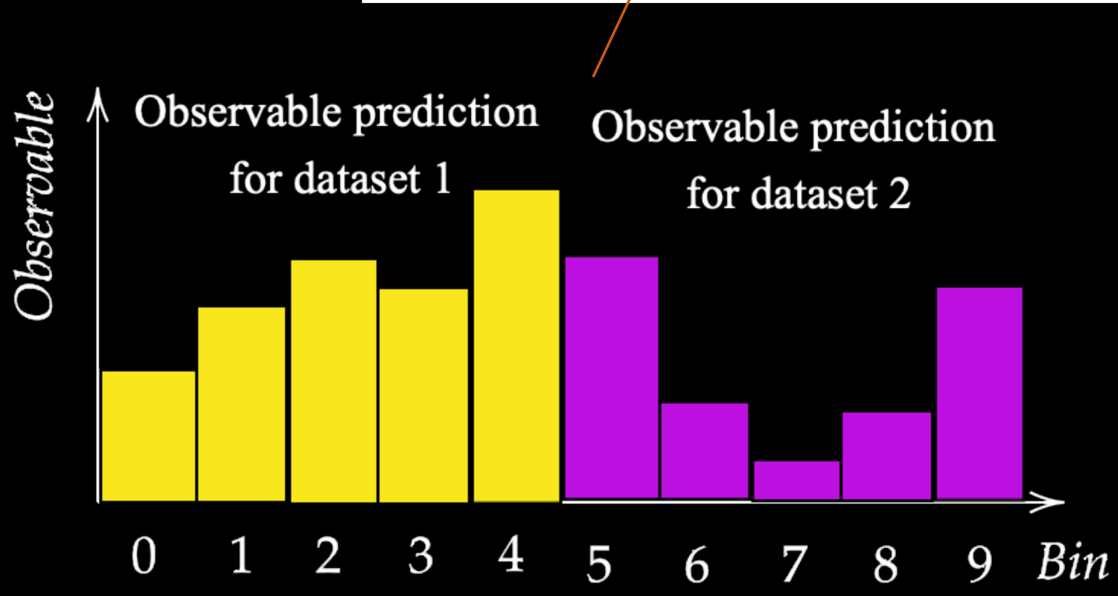
- The observable is defined as a histogram
  - The most common observable is a differential cross section
  - The observable can correspond to single- and double- differential cross-section measurements
  - Other observables are also compatible, such as hadron multiplicities
  - In each bin, we assume a gaussian distribution
  - **The histogram definition is handled by the internal GENIE database**
    - **The observable definition is associated to a data release**
- We have a histogram per set of parameters in our N_{ϑ}-dimensional phase-space



Definition of observable for multiple datasets

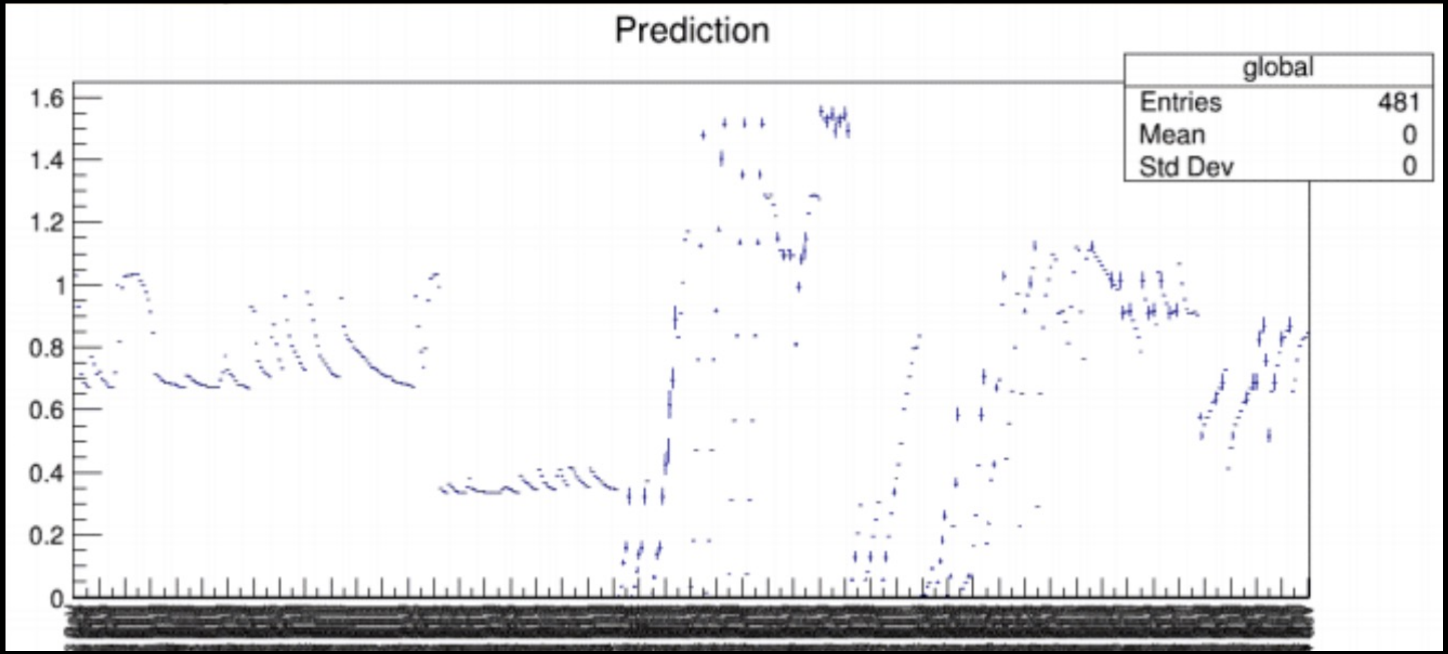
- The observable is defined as a histogram
  - It is possible to use predictions associated to different data releases
  - Each dataset can have a different observable definition
  - All the information is stored in the same 1D-histogram





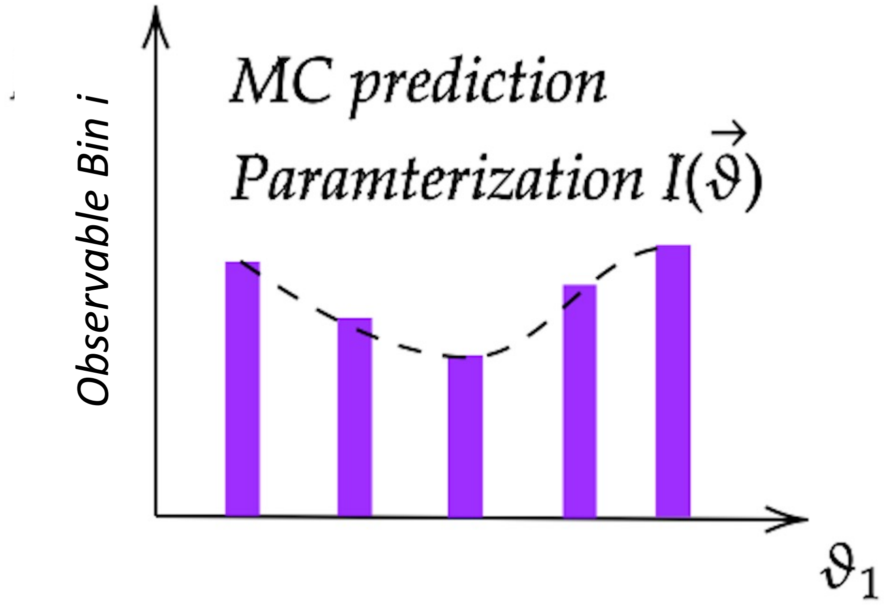
Definition of Observable
Example from GENIE's free nucleon tune

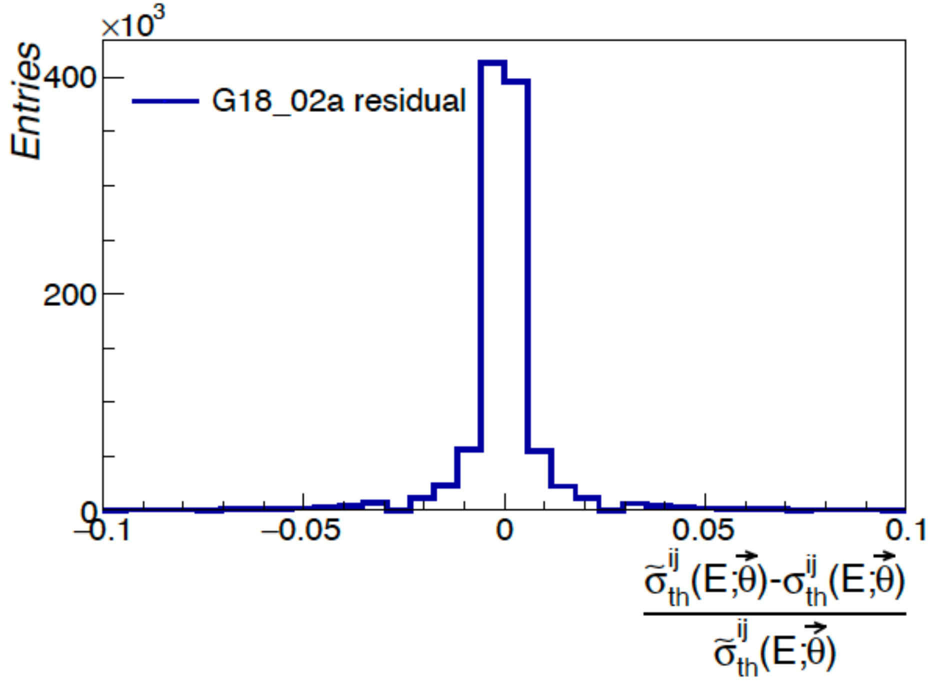
- Prediction histogram associated to thirty-three datasets [PhysRevD.104.072009]
  - The observable corresponds to a series of GENIE Predictions for ν_μ and anti- ν_μ CC inclusive, QEL, single-pion and two-pion production associated to ANL 12 ft, BNL 7ft, BEBC and FNAL bubble chamber data
- This prediction is computed with a single parameter set of our sampled phase space



Parameterization of the MC response function

- For each bin, we parameterize the observable mean value and error dependency on the parameters
- The parameterization is fit against the brute force scan
- The parameterization is an **approximation**
- It is possible to quantify its accuracy with the residual:
  - True prediction - parameterization bin-by-bin





Minimization from parameterization

- **The final step is to minimize the parameterisation against the data using `minuit`**
- The GENIE tuning software allows us to consider:
  - Correlation between data bins and datasets
  - Bin weights
    - We can use the weights to remove data from the fit
  - Proper treatment for strongly correlated data releases with Peelle's Pertinent Puzzle resolution proposed by:
    - Jaafar Chakrani et. al. at PoS NuFact2021
    - Nucl. Instrum. Meth. A 346, 306 (1994)
  - Gaussian priors on parameters ($\boldsymbol{\theta}_0$)
    - Uncorrelated and correlated
  - Nuisance parameters $\phi(\boldsymbol{f})$
    - These can also be constrained with priors

Example from the free-nucleon tune [PhysRevD.104.072009]:

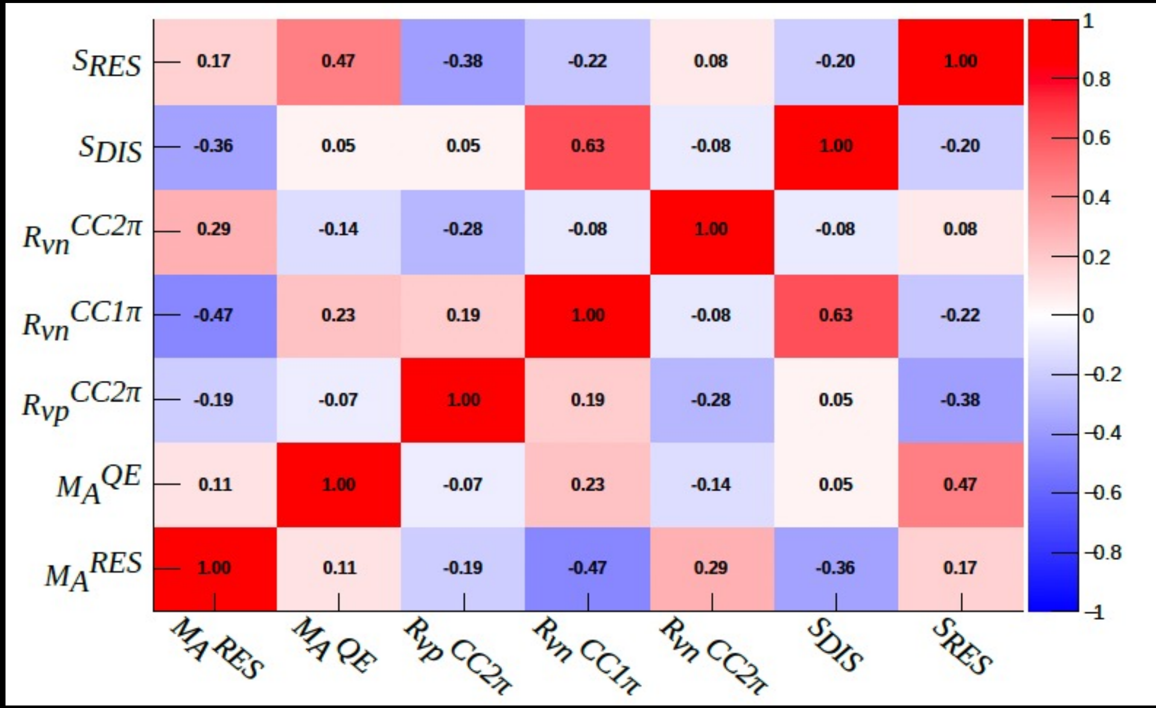
$$\chi^2(\boldsymbol{\theta},\boldsymbol{f}) = \sum_{i,j,k} w^{ijk} \frac{(\phi_j(\boldsymbol{f})\tilde{\sigma}_{\mathrm{th}}^{ij}(E_k|\boldsymbol{\theta}) - \sigma_d^{ijk})^2}{(\delta\sigma_{\mathrm{stat}}^{ijk})^2} + (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T \Sigma_\theta^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + \sum_j \frac{(f_j-1)^2}{(\delta f_j)^2}$$

Most of these features are not present in the Professor software as these were developed by the GENIE Collaboration to tune GENIE against neutrino data

Result of the tune (out-of-the-box)

- Out-of-the-box, we obtain an estimation for the best-fit parameter mean values, uncertainties and correlation between parameters
- Example from the free nucleon tune paper: PhysRevD.104.072009

Parameter	G18_01a(/b)	G18_02a(/b)
W_{cut}	1.94	1.81
M_A^{QEL}	1.00 ± 0.01	1.00 ± 0.013
M_A^{RES}	1.09 ± 0.02	1.09 ± 0.014
$R_{\nu p}^{\text{CC}1\pi}$	0.06 ± 0.03	0.008
$R_{\nu p}^{\text{CC}2\pi}$	1.1 ± 0.2	0.94 ± 0.075
$R_{\nu n}^{\text{CC}1\pi}$	0.14 ± 0.03	0.03 ± 0.010
$R_{\nu n}^{\text{CC}2\pi}$	2.8 ± 0.4	2.3 ± 0.12
S_{RES}	0.89 ± 0.04	0.84 ± 0.028
S_{DIS}	1.03 ± 0.02	1.06 ± 0.01
$\chi^2/157$ DoF	1.84	1.64

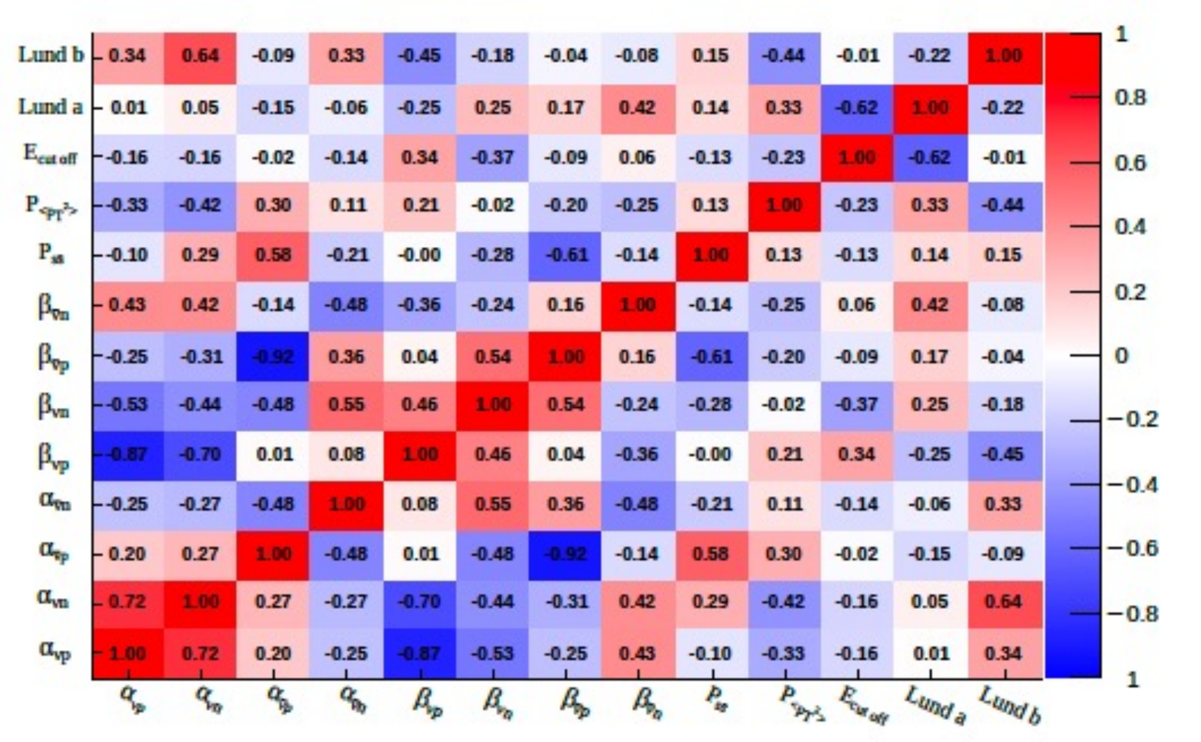


Result of the tune (out-of-the-box)

- Out-of-the-box, we obtain an estimation for the best-fit parameter mean values, uncertainties and correlation between parameters
- Example from the hadronization tune paper: [PhysRevD.105.012009]

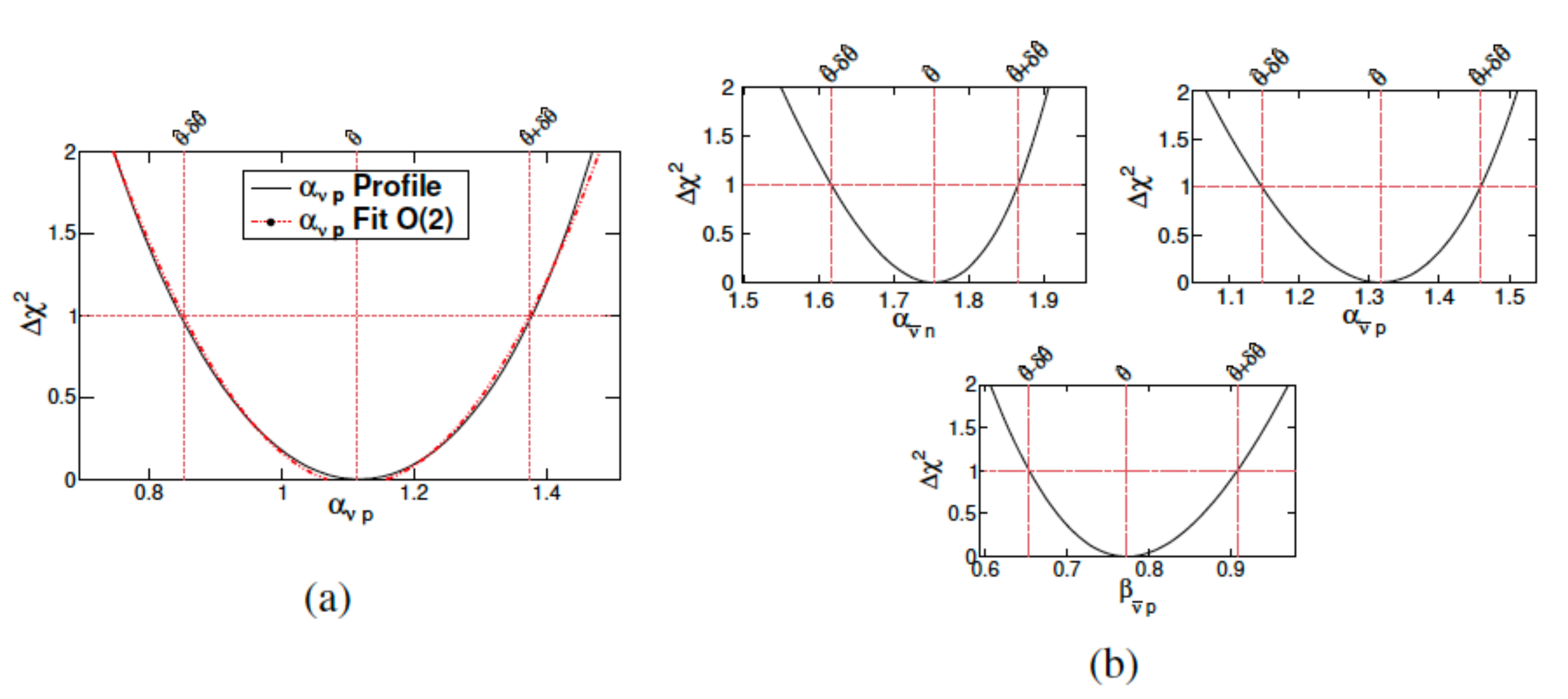
We did tunes of as many as 13 parameters and extract data-driven uncertainties

Parameter	*2021* Global Fit
$\alpha_{\nu p}$	1.1 ± 0.3
$\alpha_{\nu n}$	$1.75^{+0.14}_{-0.11}$
$\alpha_{\bar{\nu} p}$	$1.32^{+0.16}_{-0.14}$
$\alpha_{\bar{\nu} n}$	1.11 ± 0.09
$\beta_{\nu p}$	0.79 ± 0.15
$\beta_{\nu n}$	0.5 ± 0.1
$\beta_{\bar{\nu} p}$	0.8 ± 0.1
$\beta_{\bar{\nu} n}$	$0.88^{+0.09}_{-0.08}$
P_{ss}	0.27 ± 0.04
$\langle p_\perp^2 \rangle$ [GeV$^2/c^2$]	0.46 ± 0.05
E_{CutOff} [GeV]	0.30 ± 0.04
Lund *a*	1.53 ± 0.13
Lund *b* [c^4/GeV^2]	1.16 ± 0.09
	87.9/62 DoF



Parameter error estimation

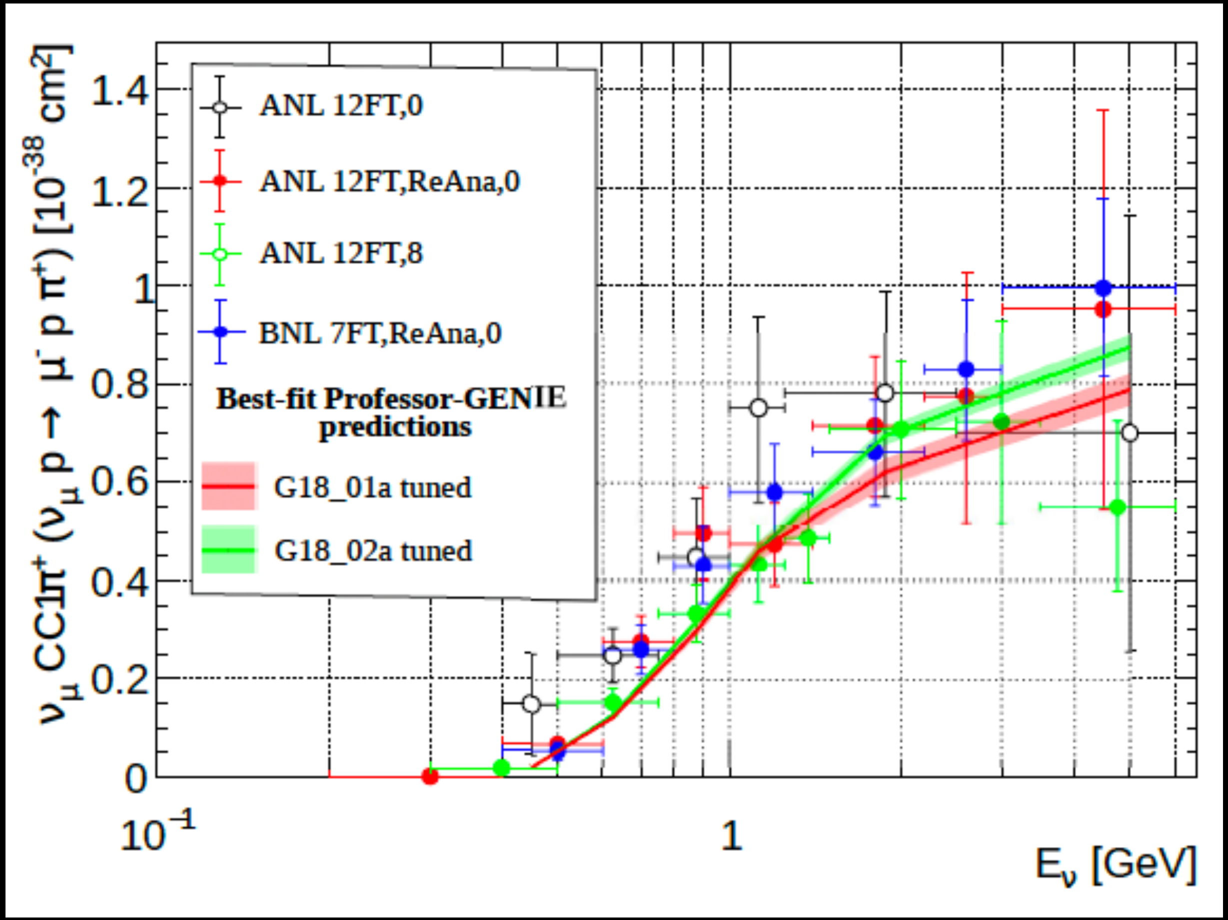
- It is possible to obtain 1D and 2D contours out of the fit
- Asymmetric errors can be extracted from profiles under the $\Delta\chi^2$ condition
- Example from the hadronization tune paper: PhysRevD.105.012009



Propagation of parameter uncertainties to GENIE Predictions

PhysRevD.105.012009

- We can propagate the post-fit uncertainty into the parameterization
- It effectively creates a confidence belt out of the tune parameterizations



Conclusions

- GENIE is developing a global analysis software for tuning and uncertainty characterization of comprehensive interaction models
- The software, based on the Professor concept, is based on the minimization of parameterized MC response functions
- **Main features of GENIE's global analysis software:**
  - All event generation models can be tuned, including non-reweightable models such as hN INTRANUKE or the AGKY hadronization model
  - New parameters can be easily added in the MC Event Generator itself
  - The most CPU consuming step (cross-section and event generation of the phase-space) can be parallelized
  - The tuning step itself its fast and can easily be reproduced
  - This tool can be used for other MC event generators

Thank you for your interest

The GENIE Collaboration

Backup slides

Lack of MC/Data description

- G18_10a_02_11b fails to describe MINERvA ν_μCC0π data at $0.15 < p_T < 0.7$ GeV/c
  - Valencia model for 1p1h and 2p2h interactions
  - hA FSI model, Berger-Sehgal model for RES
- This has a big consequences on the neutrino energy reconstruction

MINERvA $\nu_\mu CC0\pi$ data