

Genie

Comparisons

...



Marco Roda, Julia Tena-Vidal

Generator Tool workshop
17 March 2023
Fermilab

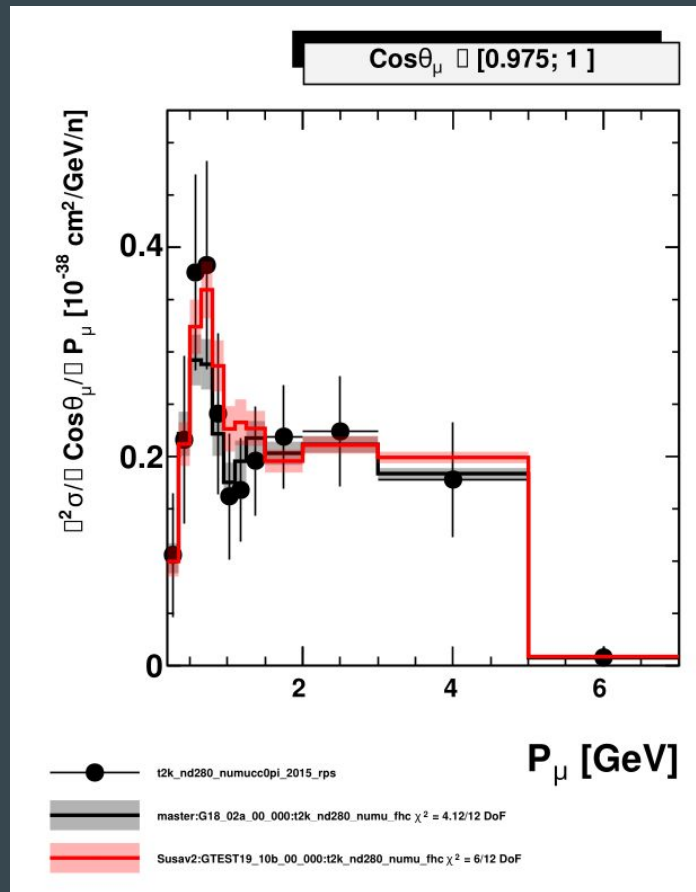


Intro

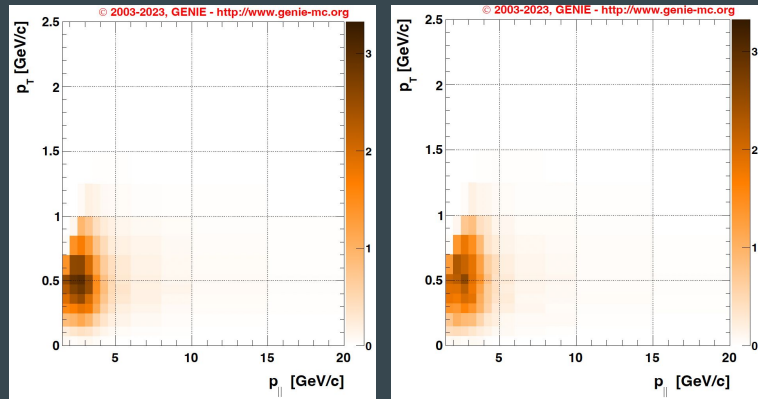
- Overview of the comparisons code
 - It is our internal **database**
 - It is an integrated **part of the GENIE global analysis software**
 - The code is in **active development**
 - A first public release will become available once the global analysis software is stable, fully exploited and documented
- Developed mostly to **benchmark the predictions against data**
 - Eventually it became the **foundation for our tuning system**
 - It provides the comparisons with a format compatible with Professor
 - The bulk of the work is related to **add new datasets** when they are available
 - Make sure we understand the data well to create the right predictions
- **It's not just a database**
 - It also contains code to massively parallelise productions (Splines, event generation, etc)
 - Available for multiple batch systems
 - GRID is not there yet but we are working on it
 - Dedicated scripts to run events and splines are now available

Data formats

- Plots with metrics - including correlations, slices etc
 - The output is not just a plot to emulate a paper plot
 - We have a standard format so we navigate the plots
 - Create chi-squares for portions of the plot
 - We can add systematics on the data
 - For instance, we can add systematics based on the available information on the papers



Example of data-MC comparison

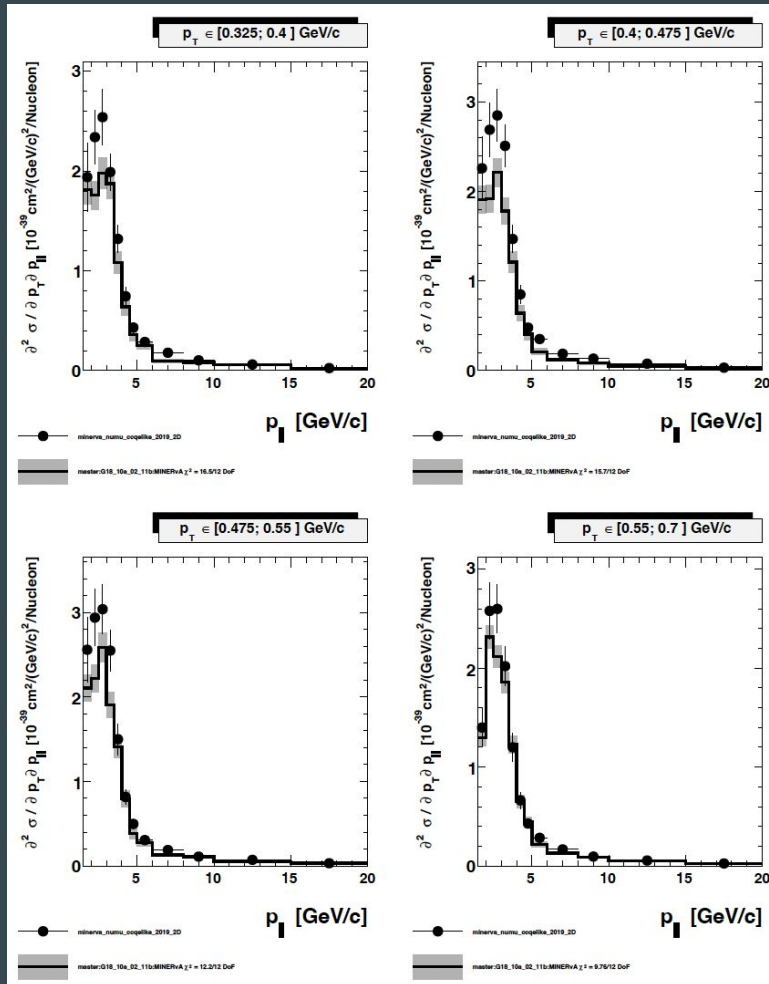
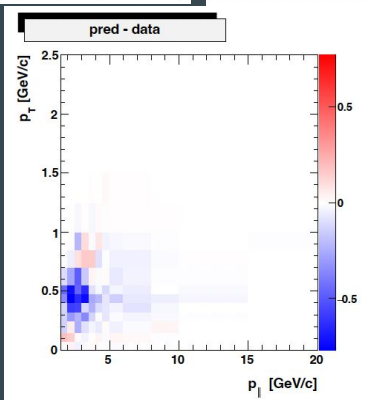


$$\frac{\partial^2 \sigma}{\partial p_T \partial p_{\parallel}} [10^{-39} \text{cm}^2/(\text{GeV}/c)^2/\text{Nucleon}]$$

Data: minerva_numu_ccqelike_2019_2D

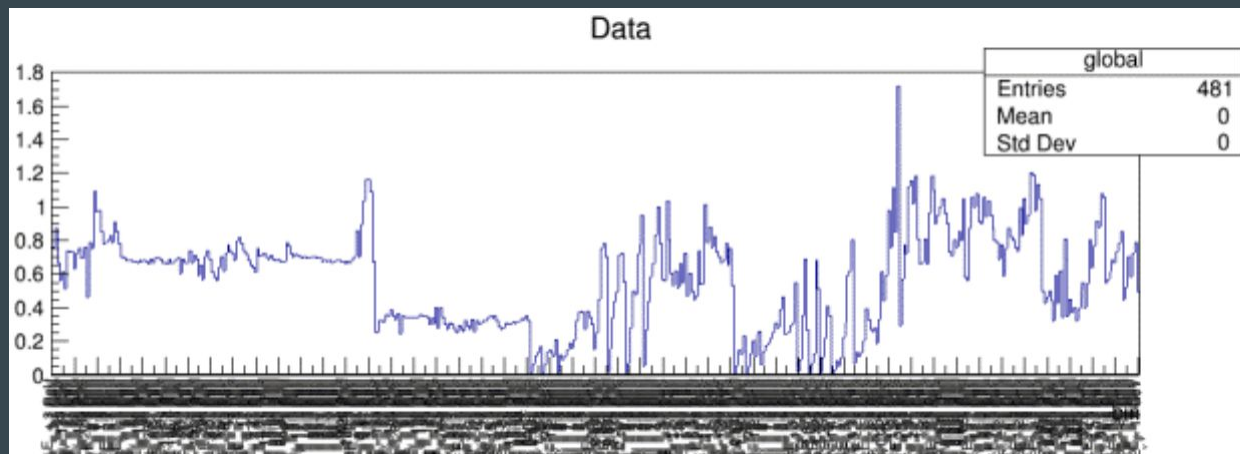
$$\frac{\partial^2 \sigma}{\partial p_T \partial p_{\parallel}} [10^{-39} \text{cm}^2/(\text{GeV}/c)^2/\text{Nucleon}]$$

Pred: master:G18_10a_02_11b:MINERvA



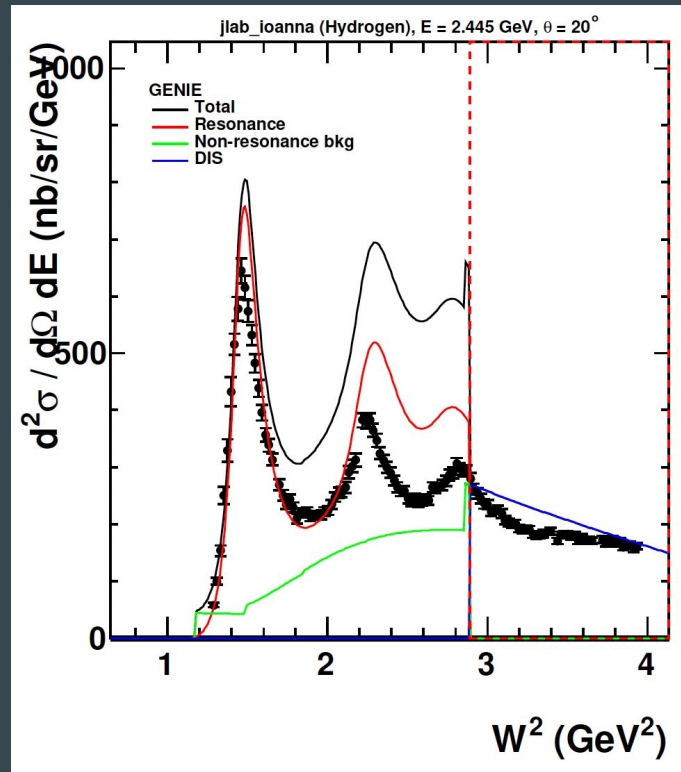
Data formats

- Fully linearised output
 - For the plots for which we have high level information we can obtain data and predictions in the form of N-tuples
 - This allows more detailed analysis if needed
 - This is the format we feed into your tuning machinery



Data formats

- We have multiple output formats
 - The type depends on the usage case
 - Some output are not available for all datasets
 - This reflects the history of the database itself
- Plots to simply compare data and predictions
 - Lowest level of details, oldest format
 - No metrics on the agreement
 - Views usually limited to one plot for datasets



Available datasets and usage case

- Lots of data
 - Neutrino interaction data from bubble chamber data, MiniBooNE, MicroBooNE, MINERvA, minos, T2K ND280
 - Integrated cross sections
 - Flux integrated, differential cross section
 - **N-Dimensional distributions allowed**
 - Hadronization data
 - Electron scattering data
 - Hadron scattering data to validate FSI
- High quality formats are available for
 - Integrated neutrino cross sections
 - Hadronisation data
 - CC datasets from modern experiments (MINERvA, T2K and MicroBooNE)
- We are working to incorporate electron scattering data

Available datasets and usage case

- Regardless of the data format the production is usually the same
 - First we generate a dedicated MC production
 - We can use dedicated flux files, target combinations, etc
 - The MC production is highly parallelized
 - The MC productions are fed into a software the computes the data predictions
 - That is the part that requires time and expertise from the experiments to be done properly
 - Finally, we can construct the comparisons plots and higher level information if possible
 - Ideally we would like to arrive at the point where all the data have all this high level information
 - But it takes time and we are definitely not there

Usability

- The code is private
 - But we are considering opening to some experimental collaborations
 - Collaborations with a lot of GENIE members
 - The idea is train people to do validation and test on model configurations
 - GENIE members would probably use these tools anyway
- There might be some refactoring happening here in the future
 - Part of what is private might become public
 - For instance, spline and event generation scripts
 - We need to understand what resources we have to support public usage



Thank you



Backup