Discussion panel

...

NuSTEC workshop on Electron Scattering
31 March 2022

Marco Roda - University of Liverpool
for the GENIE collaboration

Questions

1. What is new with GENIE v3.02.00 ?
2. Is there machinery ready or somewhat reusable to implement a full neutrino world data tuning?
    a. Is this inclusive? Exclusive?
    b. What are the dominant differences in inclusive and exclusive tuning?
3. What about for electron world data tuning?
    a. Can this be easily extended for muons? Pions?

- 1 will be short, mostly a list of links so you can find the details
- 2 and 3 are connected
    - I'll start with a general intro about the tuning strategy so that the details can be followed
    - I'll try to give as much details as possible on the tools we used in GENIE while answering to 2
    - 3 will be an easy corollary

What is new in genie version 3.02.00 ?

- Complete list: http://releases.genie-mc.org/
- Some highlights important for the topics of this workshop
  - New cross section models: SuSAv2, ...
  - Correlated Fermi Gas model
  - New FSI models: INCL, Geant4
  - Electron scattering tunes
  - New tunes:
    - free nucleon scattering Phys.Rev.D 104 (2021) 7, 072009
    - first hadronization tune Phys. Rev. D 105 (2022) 1, 012009
  - A number of technical features and improvements

- Left for 3.02.02 or later releases
  - You can see some preliminary results in Eur.Phys.J.ST 230 (2021) 24
  - MK model for single pion production
    - The code is ready but the validation is ongoing
  - COH Gamma production
    - We are working on optimising the configuration and the consistency with the rest of the code

GENIE strategy for tuning

1. Define the data we want to use for the fit
    - they cannot be changed because all the predictions needs to be done in advance

2. Brute force sampling over a phase-space - no reweight
    - In each point we evaluate
        - all the necessary splines
        - Event generation
        - create prediction

3. Once we have all the predictions
    - we interpolate them bin-by-bin for every point used in the fit
    - Polynomial interpolation of a selected order in all the variables of the phase space
        - We call them Professor parameterizations

4. We use those interpolations to do the fit

- Some numbers
    - O(1 K) sampling points depending on the size of the parameter space and of the polynomial order
    - O(5 TB) space on disk to build the predictions. Genie here has the additional burden of the splines, that are part of the production
    - O(< 1 month) to run on a farm

The genie machinery - what is reusable?

- The whole machinery is reusable in its core
  - It was adapted from LHC experiments
    - Most of the building blocks are
  - There are some bits and pieces that are specific, mostly to the generator or the computing infrastructure

- Required machinery, AKA shopping list colour code reusable mixed or not relevant specific
  - Professor code (public)
    - handle the sampling, the parameterisations, some utilities control function
      - We extended privately some of those functionalities
  - automated scheduling for batch production - genie optimised
    - Private, but some of the scripts are available in generator
    - Most likely they are useless if you use a farm which is not one of ours
  - database with data and to build prediction out of genie
    - The data are public but the database is genie related
  - Professor to build parameterisations
  - Minimiser - there are no constraints here. pick your favourite
    - We use Minuit2
    - We developed the minimisation function to include proper correlation, priors, etc was developed for our tunes

The tunes themselves are reusable

- We have a development model in which the tunes are also reusable

- Two main approaches
  a. The results can be used as priors for a further tune
     - Tunes can be built on top of other tunes
       - We implemented multi-dimensional gaussian priors
     - Example
       - We are now working on nuclear tunes using 0pi datasets
       - The free nucleon tune results are used as priors for the nuclear tune

  b. The tuning procedure does not require changes in the generator itself
     - Other collaborations can use the tunes
     - We are trying to be careful and make sure that the tunes are published so the details are clear

Some details on what is not immediately reusable

- batch production scripts
  - Professor will give you a set of directories, one for each point of the parameter space
    - point coordinates
    - configuration files generated according to a template you provided
  - Your jobs is to fill that directory so that eventually there will be your prediction for that point of the parameter space
  - Most likely this is farm dependent, especially for productions of this size: you need some fine tuning

- Database
  - System that can construct the predictions based on
    - configuration of the database AKA the datasets you want to use
    - the parameters of the input
      - In our case genie is public and it's part of the process
      - The generator dependent part is the constructions of the predictions starting from generated events
  - It has be modular so you can swap in and out datasets and treat every point as an entry in an array
    - Professor treats all those points as a bin in a histogram
  - In GENIE this evolved simply from our internal database we used for benchmarking the releases
    - Unfortunately the whole infrastructure is not well maintained for electron data, or hadron data.
    - But the system allows it naturally

- Minimisation function and fitting tools
  - in order to do the fits we had to take into account loads of things
    - datasets correlations
    - weights
    - priors - that are not always gaussian
    - additional nuisance parameters
  - All of this has to be able to interface with the professor API for the parameterizations
    - That's why we started with a simple python and we developed our own fitting system

Inclusive or exclusive fit

- From the machinery point of view inclusive or exclusive does not matter
  - The bottleneck (if there is any) is in your database
    - Can it construct exclusive or inclusive predictions?
    - Are there exclusive or inclusive datasets you can tune against?
  - The number of points (usually higher for exclusive datasets) is not relevant for this approach
    - The computational effort is due to the brute force scanning, 100 or 1000 DoF for the fit will not make any difference
    - Only the fitting might be a bit slower (1 h instead of 10 minutes)
      - compared to 1 month to produce the professor parameterization, it's nothing

- From the statistical point of view
  - exclusive and inclusive datasets from the same experiments are correlated
    - And most likely the correlation is not provided and it's unknown
  - The same goes for different exclusive datasets from the same experiment (or beam)
  - While doing global analyses it becomes clear very quickly that the major barriers are the data releases information
    - Some correlations are obvious to be required
    - But if not provided the best you can do is guess

- From the model point of view
  - They can be tuned as long as there are parameters to tune
  - Inclusive quantities are easier to tune
    - Things like hadronic table are not controlled by parameters so they are not tunable because of their implementation
    - We started discussions with theorists to allow some controls over those tables but so far nothing has come out of it

Electron scattering tunes

- I hope to have made clear that nothing related to the procedure is specific for neutrinos
  - In fact professor was developed for LHC experiments - jet shower MCs
  - We simply built our own machinery around it
    - Based on tools we already had inside GENIE

- If your generator can produce electron (or pion, or whatever) predictions
  - You can use it with this procedure
    - The only requirement is that the software does not require compiling if you change a parameter
  - Assuming that your database is capable of consuming the event and produce the predictions

- The problem becomes statistical:
  - Given the statistics of electron scattering data, combined neutrino-electron tunes might be tricky, yet very desirable
    - The choice of the parameters we want to tune becomes crucial
      - common parameters will be driven by electrons scattering
      - unless a dedicated weight system is put in place
    - I think for electron scattering it's important to keep in mind that electron scattering tunes will be used as a base for neutrino tunes too!
      - This approach will allow it naturally
  - This borders with one of the other questions so I will stop here