Data Releases and Preservation

Luke Pickering

Measuring neutrino interactions for next-generation oscillation experiments

ECT*, Trento, 2024/10/25

Science and Technology Facilities Council

Supported by URF\R1\211661

THE ROYAL SOCIETY

Data Releases

Why Release?

- Data is very valuable (and expensive: economic and human time)
- Improving interaction models without high-quality, accessible measurements to predict is both very difficult and ~pointless
- For neutrino oscillations: We need improved models for the headline precision physics goals

What do we want from data releases?

- An unambiguous prescription for making a prediction of a measurement and calculating a goodness of fit for the prediction

Data Releases

Why Release?

- Data is very valuable (and expensive: economic and human time)
- Improving interaction models without high-quality, accessible measurements to predict is both very difficult and ~pointless
- For neutrino oscillations: We need improved models for the headline precision physics goals

What do we want from data releases?

- An unambiguous prescription for making a prediction of a measurement and calculating a goodness of fit for the prediction
- Simple, right?

This Talk

- This talk is mostly my understanding of the **problem domain** of publishing and using scattering data

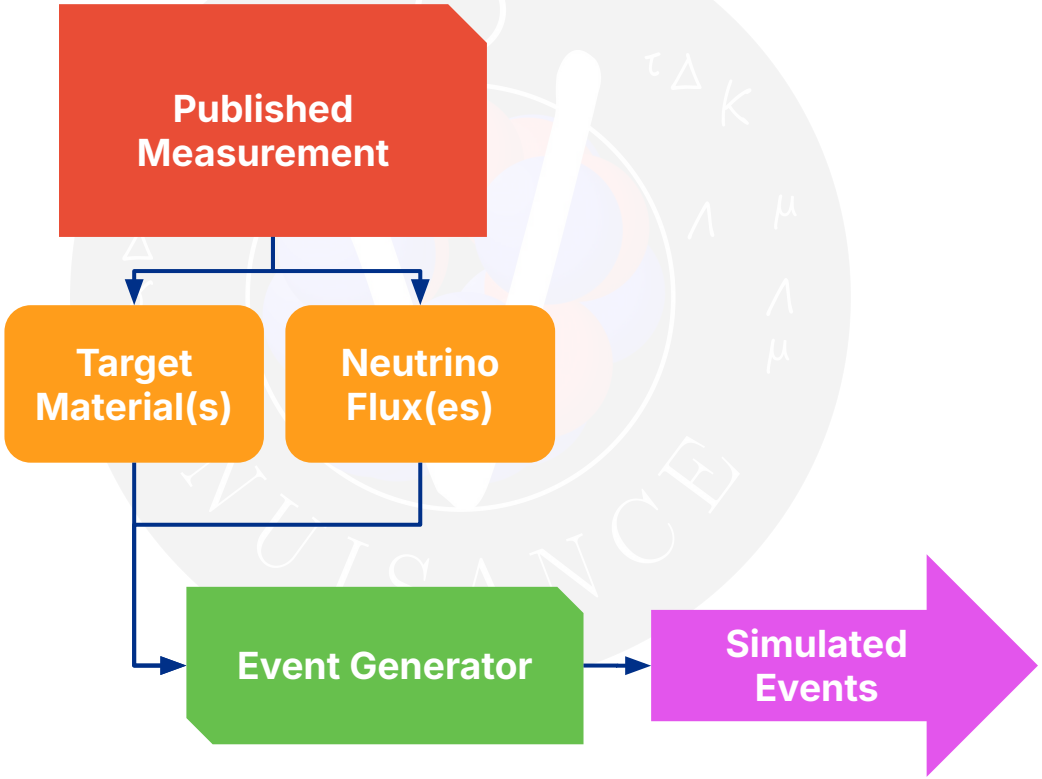
This Talk

- This talk is mostly my understanding of the **problem domain** of publishing and using scattering data
  - Ultimately, the problem is not difficult to solve unlike other problems discussed at this workshop:
    - Whats are best measurements to make?
    - Which statistical meat grinder to use to summarise measurement?
    - How to model neutrino-nucleus interactions in the few-GeV region
    - ...

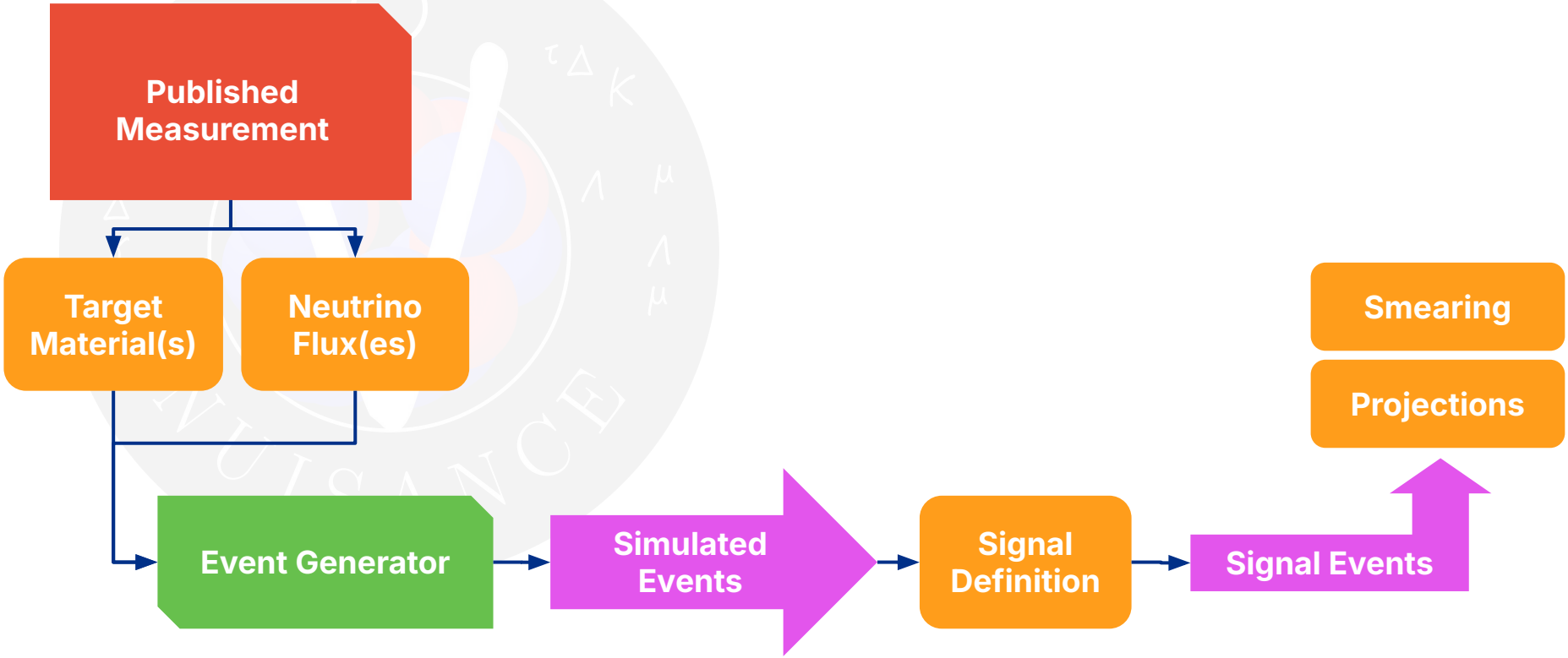
This Talk

- This talk is mostly my understanding of the **problem domain** of publishing and using scattering data
  - Ultimately, the problem is not difficult to solve unlike other problems discussed at this workshop:
    - Whats are best measurements to make?
    - Which statistical meat grinder to use to summarise measurement?
    - How to model neutrino-nucleus interactions in the few-GeV region
    - ...
- I will also show a quick example of a possible solution that we've been working on
- **BUT:** The most useful solution is the one that is used, so this is also a call to engage!

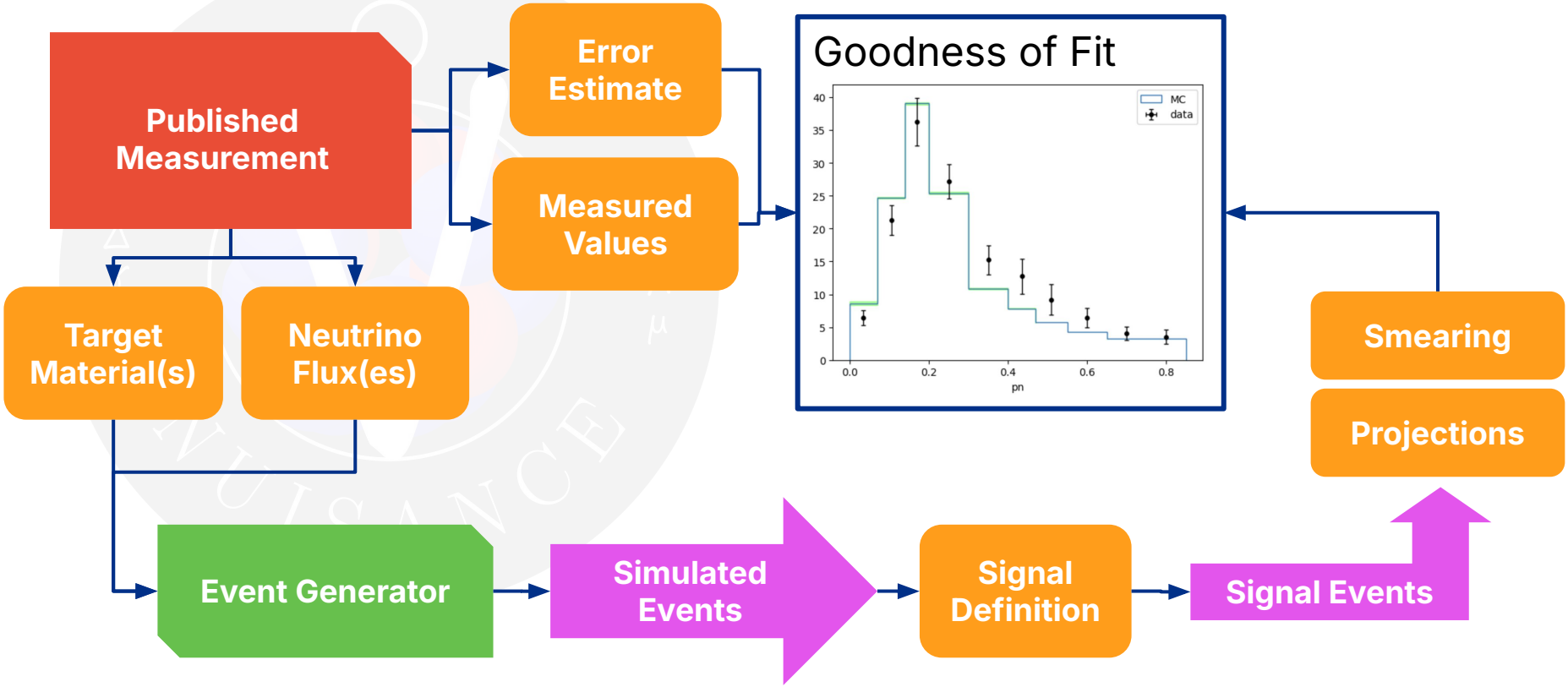
Measurement Predictions: A Workflow



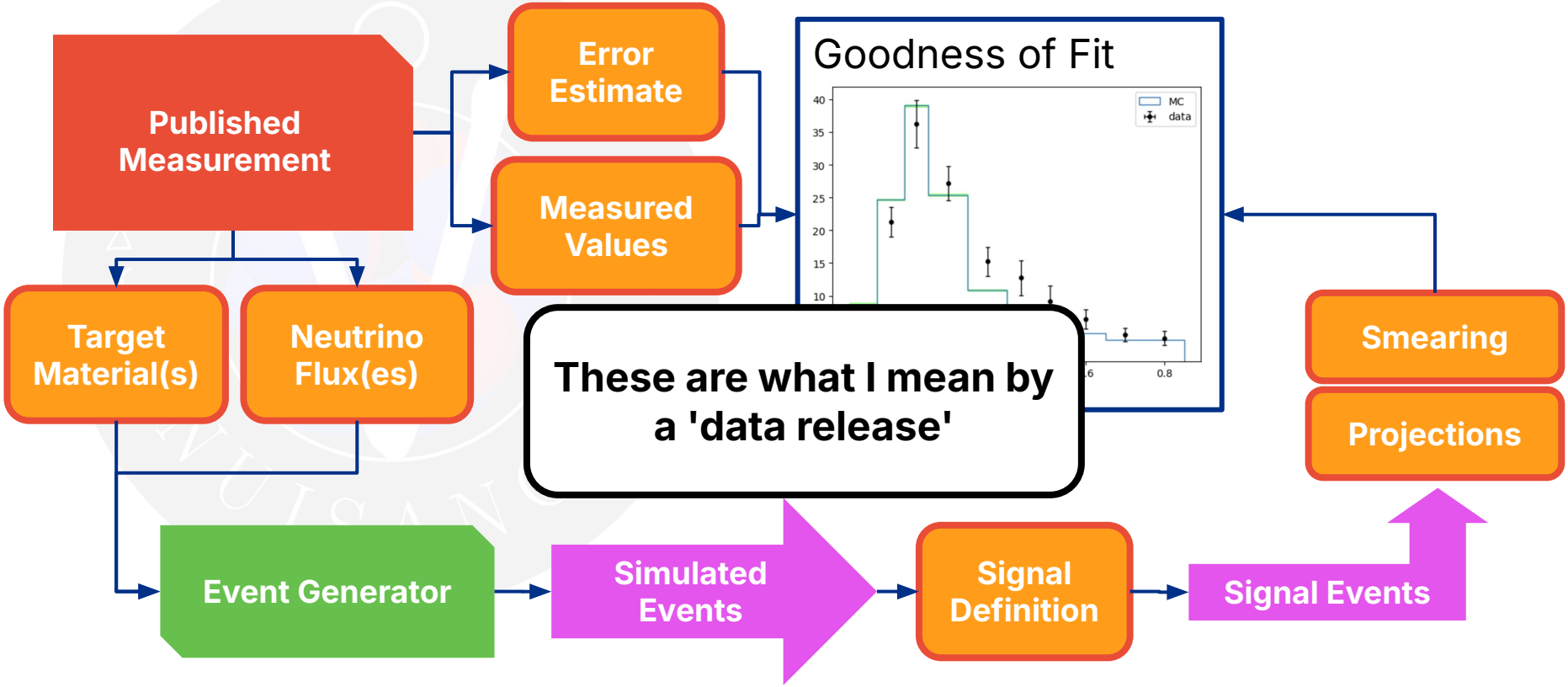
Measurement Predictions: A Workflow



Measurement Predictions: A Workflow

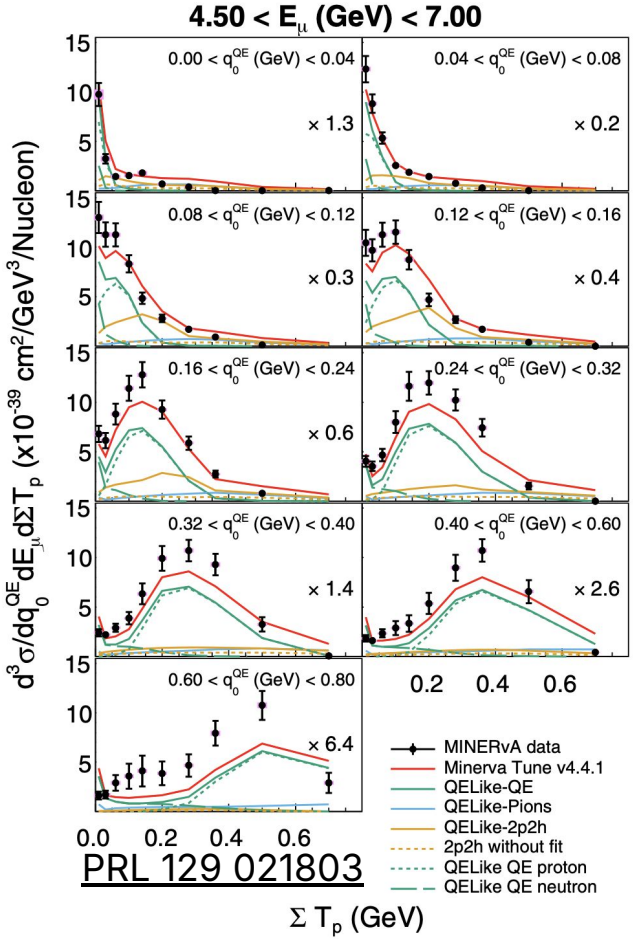


Measurement Predictions: A Workflow



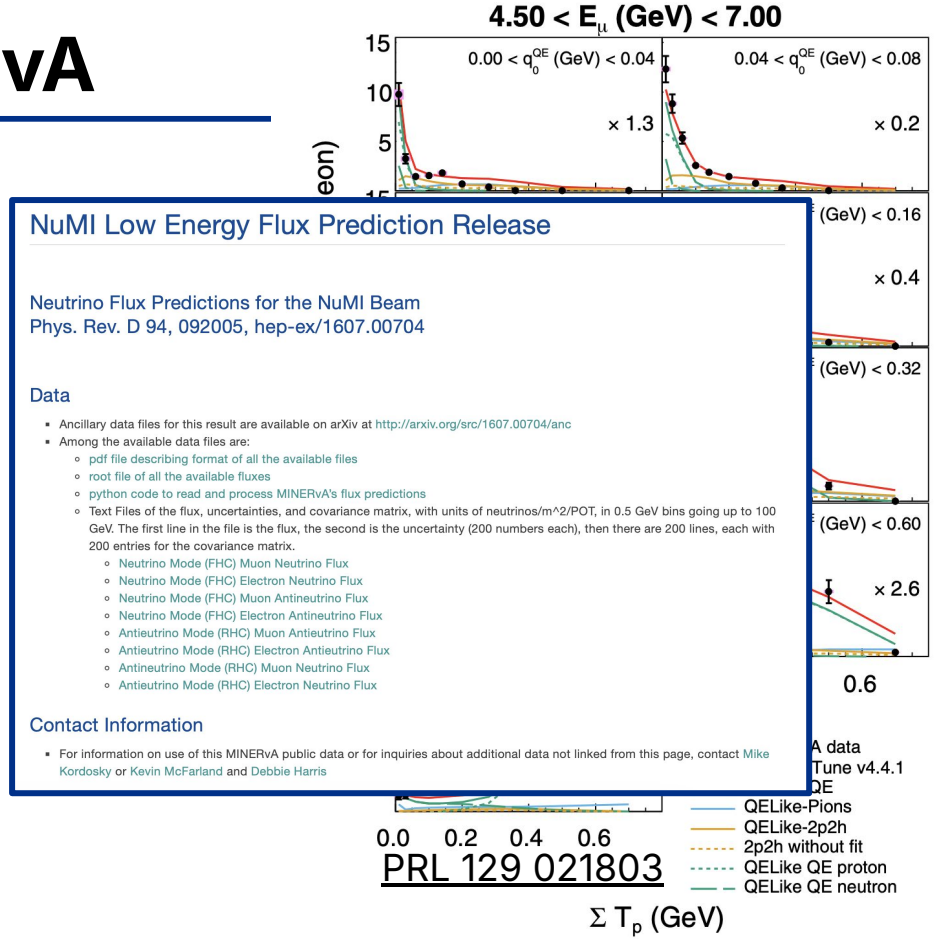
An Example from MINERvA

- Data histogram (bins, values)



An Example from MINERvA

- Data histogram (bins, values)
- Errors (covariance in bin number)
- **Flux shape**



NuMI Low Energy Flux Prediction Release

Neutrino Flux Predictions for the NuMI Beam
Phys. Rev. D 94, 092005, hep-ex/1607.00704

Data

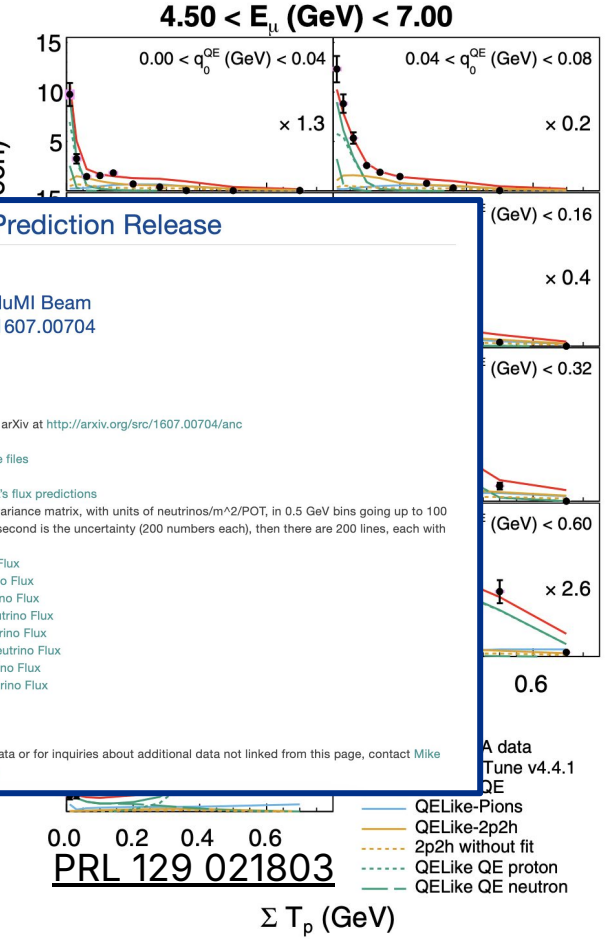
- Ancillary data files for this result are available on arXiv at http://arxiv.org/src/1607.00704/anc
- Among the available data files are:
  - pdf file describing format of all the available files
  - root file of all the available fluxes
  - python code to read and process MINERvA's flux predictions
  - Text Files of the flux, uncertainties, and covariance matrix, with units of neutrinos/m^2/POT, in 0.5 GeV bins going up to 100 GeV. The first line in the file is the flux, the second is the uncertainty (200 numbers each), then there are 200 lines, each with 200 entries for the covariance matrix.
    - Neutrino Mode (FHC) Muon Neutrino Flux
    - Neutrino Mode (FHC) Electron Neutrino Flux
    - Neutrino Mode (FHC) Muon Antineutrino Flux
    - Neutrino Mode (FHC) Electron Antineutrino Flux
    - Antieutrino Mode (RHC) Muon Antieutrino Flux
    - Antieutrino Mode (RHC) Electron Antieutrino Flux
    - Antineutrino Mode (RHC) Muon Neutrino Flux
    - Antieutrino Mode (RHC) Electron Neutrino Flux

Contact Information

- For information on use of this MINERvA public data or for inquiries about additional data not linked from this page, contact Mike Kordosky or Kevin McFarland and Debbie Harris

An Example from MINERvA

- Data histogram (bins, values)
- Errors (covariance in bin number)
- Flux shape
- **Target: CH**



NuMI Low Energy Flux Prediction Release

Neutrino Flux Predictions for the NuMI Beam
Phys. Rev. D 94, 092005, hep-ex/1607.00704

Data

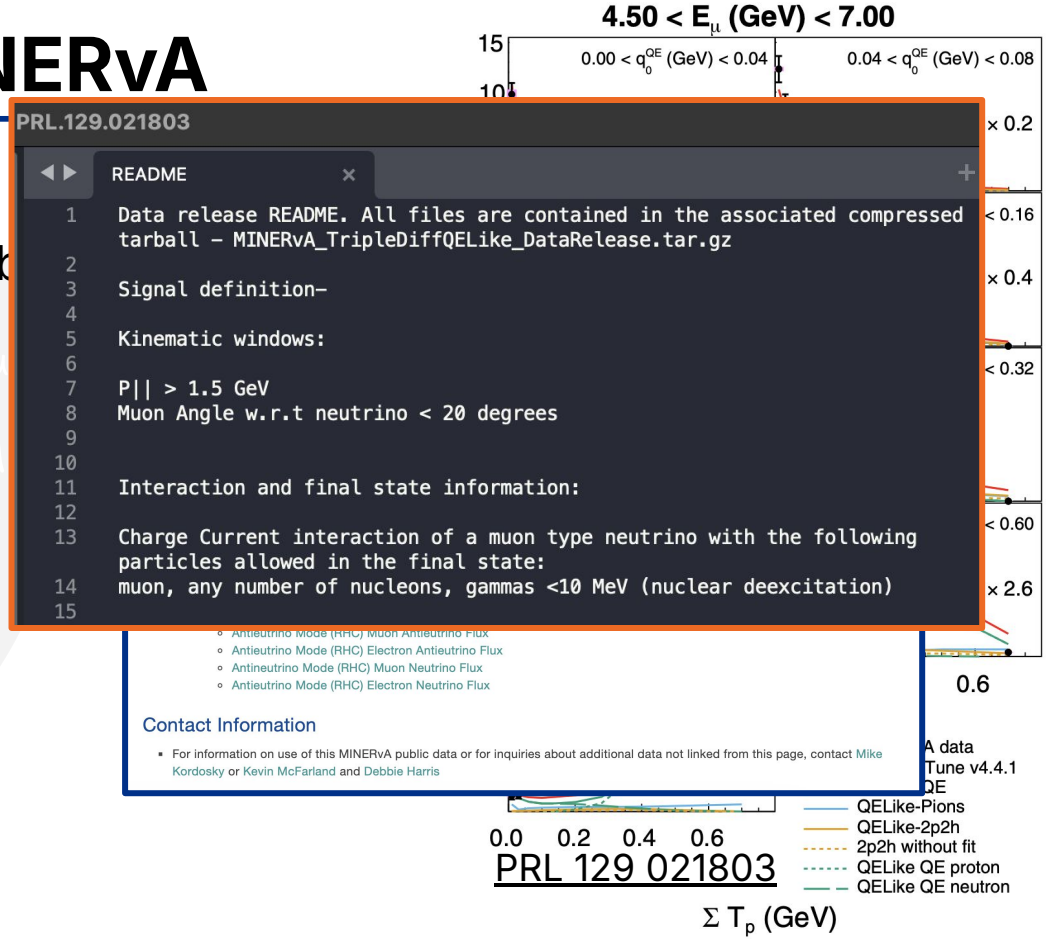
- Ancillary data files for this result are available on arXiv at http://arxiv.org/src/1607.00704/anc
- Among the available data files are:
  - pdf file describing format of all the available files
  - root file of all the available fluxes
  - python code to read and process MINERvA's flux predictions
  - Text Files of the flux, uncertainties, and covariance matrix, with units of neutrinos/m^2/POT, in 0.5 GeV bins going up to 100 GeV. The first line in the file is the flux, the second is the uncertainty (200 numbers each), then there are 200 lines, each with 200 entries for the covariance matrix.
    - Neutrino Mode (FHC) Muon Neutrino Flux
    - Neutrino Mode (FHC) Electron Neutrino Flux
    - Neutrino Mode (FHC) Muon Antineutrino Flux
    - Neutrino Mode (FHC) Electron Antineutrino Flux
    - Antieutrino Mode (RHC) Muon Antieutrino Flux
    - Antieutrino Mode (RHC) Electron Antieutrino Flux
    - Antineutrino Mode (RHC) Muon Neutrino Flux
    - Antieutrino Mode (RHC) Electron Neutrino Flux

Contact Information

- For information on use of this MINERvA public data or for inquiries about additional data not linked from this page, contact Mike Kordosky or Kevin McFarland and Debbie Harris

An Example from MINERvA

- Data histogram (bins, values)
- Errors (covariance in bin numb
- Flux shape
- Target: CH
- **Signal definition**



An Example from MINERvA

- Data histogram (bins, values)
- Errors (covariance in bin numb
- Flux shape
- Target: CH
- Signal definition
- **Projection operators**

PRL.129.021803

README

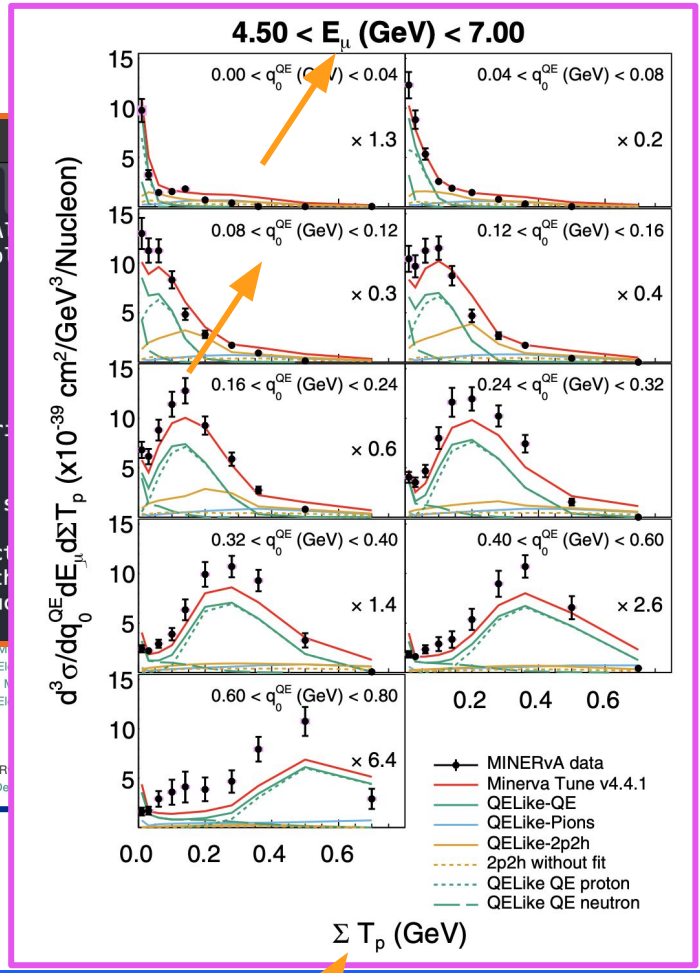
```
1   Data release README. A
    tarball - MINERvA_Trip
2
3   Signal definition-
4
5   Kinematic windows:
6
7   P|| > 1.5 GeV
8   Muon Angle w.r.t neutr
9
10
11  Interaction and final s
12
13  Charge Current interact
    particles allowed in th
14  muon, any number of nu
15
```

Antineutrino Mode (RHC) M
Antineutrino Mode (RHC) El
Antineutrino Mode (RHC) M
Antineutrino Mode (RHC) El

Contact Information

- For information on use of this MINER
  Kordosky or Kevin McFarland and De

```
36
37  ###
38  SumTp vs q0_qe vs Muon E
39  ###
40  ____
```



An Exampl

- Data histogra
- Errors (covar
- Flux shape
- Target
- Signal definit
- Projection op

NuMI Low Energy Flux Prediction Release

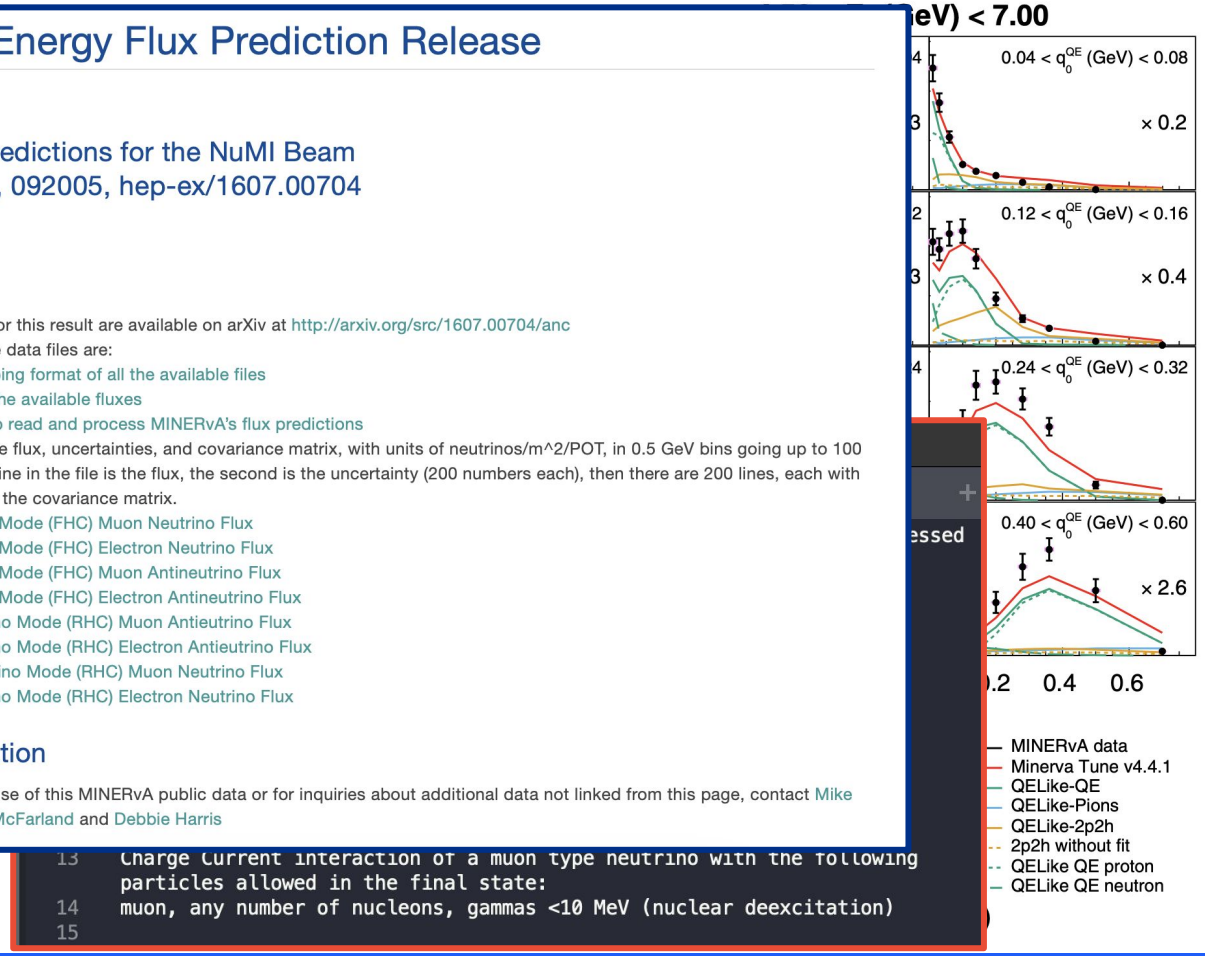
Neutrino Flux Predictions for the NuMI Beam
Phys. Rev. D 94, 092005, hep-ex/1607.00704

Data

- Ancillary data files for this result are available on arXiv at http://arxiv.org/src/1607.00704/anc
- Among the available data files are:
  - pdf file describing format of all the available files
  - root file of all the available fluxes
  - python code to read and process MINERvA's flux predictions
  - Text Files of the flux, uncertainties, and covariance matrix, with units of neutrinos/m^2/POT, in 0.5 GeV bins going up to 100 GeV. The first line in the file is the flux, the second is the uncertainty (200 numbers each), then there are 200 lines, each with 200 entries for the covariance matrix.
    - Neutrino Mode (FHC) Muon Neutrino Flux
    - Neutrino Mode (FHC) Electron Neutrino Flux
    - Neutrino Mode (FHC) Muon Antineutrino Flux
    - Neutrino Mode (FHC) Electron Antineutrino Flux
    - Antieutrino Mode (RHC) Muon Antieutrino Flux
    - Antieutrino Mode (RHC) Electron Antieutrino Flux
    - Antineutrino Mode (RHC) Muon Neutrino Flux
    - Antieutrino Mode (RHC) Electron Neutrino Flux

Contact Information

- For information on use of this MINERvA public data or for inquiries about additional data not linked from this page, contact Mike Kordosky or Kevin McFarland and Debbie Harris



Overview of Encountered Approaches

Cross section data/Flux predictions/Erro

- Plots in papers

Requires manual digitisation, infeasible for many bin measurements

I've tried asking ChatGPT to do it... don't recommend

Overview of Encountered Approaches

Cross section data/Flux predictions/Erro

- Plots in papers
- Tables in papers

ChatGPT is probably better at this one, but it still requires human effort to 'get' the data

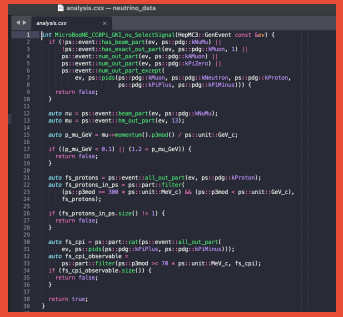
Overview of Encountered

Better! As there is no standard format between or within experiments, requires (sometimes significant) effort on the part of the user to understand the form of the data

Cross section data/Flux predictions/Erro

- Plots in papers
- Tables in papers
- Text tables in supplementary materials
- ROOT files in supplementary materials
- File storage on experiment webspace/third party repository

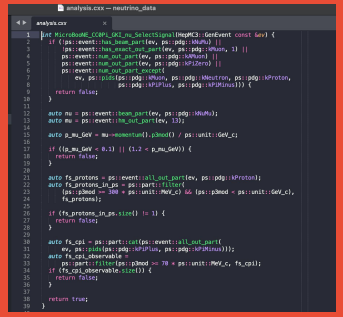
The Ill-posed Analysis Problem



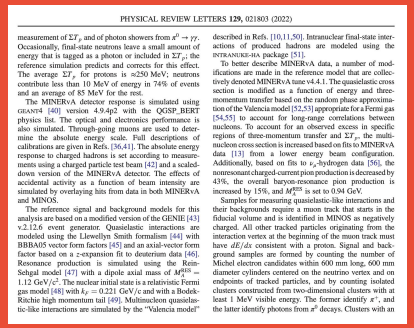
Expert analyser writes code for signal/projection

- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)

The Ill-posed Analysis Problem



Expert analyser writes code for signal/projection


PHYSICAL REVIEW LETTERS 129, 021803 (2022)

measurement of ΣT_p and of photon showers from $\pi^0 \to \gamma\gamma$. Occasionally, final-state neutrons leave a small amount of energy that is tagged as a photon or included in ΣT_p; the reference simulation predicts and corrects for this effect. The average ΣT_p for protons is ≈250 MeV; neutrons contribute less than 10 MeV of energy in 74% of events and an average of 85 MeV for the rest.

The MINERvA detector response is simulated using GEANT4 [40] version 4.9.4p2 with the QGSP_BERT physics list. The optical and electronics performance is also simulated. Through-going muons are used to determine the absolute energy scale. Full descriptions of calibrations are given in Refs. [36,41]. The absolute energy response to charged hadrons is set according to measurements using a charged particle test beam [42] and a scaled-down version of the MINERvA detector. The effects of accidental activity as a function of beam intensity are simulated by overlaying hits from data in both MINERvA and MINOS.

The reference signal and background models for this analysis are based on a modified version of the GENIE [43] v.2.12.6 event generator. Quasielastic interactions are modeled using the Llewellyn Smith formalism [44] with BBBA05 vector form factors [45] and an axial-vector form factor based on a z-expansion fit to deuterium data [46]. Resonance production is simulated using the Rein-Sehgal model [47] with a dipole axial mass of $M_A^{\mathrm{RES}} = 1.12\ \mathrm{GeV}/c^2$. The nuclear initial state is a relativistic Fermi gas model [48] with $k_F = 0.221\ \mathrm{GeV}/c$ and with a Bodek-Ritchie high momentum tail [49]. Multinucleon quasielastic-like interactions are simulated by the "Valencia model" described in Refs. [10,11,50]. Intranuclear final-state interactions of produced hadrons are modeled using the INTRANUKE-hA package [51].

To better describe MINERvA data, a number of modifications are made in the reference model that are collectively denoted MINERvA tune v4.4.1. The quasielastic cross section is modified as a function of energy and three-momentum transfer based on the random phase approximation of the Valencia model [52,53] appropriate for a Fermi gas [54,55] to account for long-range correlations between nucleons. To account for an observed excess in specific regions of three-momentum transfer and ΣT_p, the multinucleon cross section is increased based on fits to MINERvA data [13] from a lower energy beam configuration. Additionally, based on fits to ν_μ-hydrogen data [56], the nonresonant charged-current pion production is decreased by 43%, the overall baryon-resonance pion production is increased by 15%, and M_A^{RES} is set to 0.94 GeV.

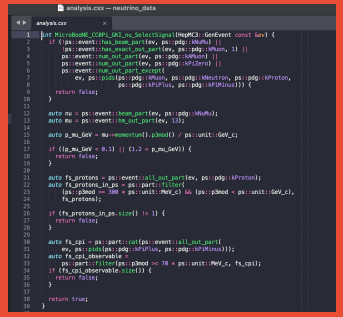
Samples for measuring quasielastic-like interactions and their backgrounds require a muon track that starts in the fiducial volume and is identified in MINOS as negatively charged. All other tracked particles originating from the interaction vertex at the beginning of the muon track must have dE/dx consistent with a proton. Signal and background samples are formed by counting the number of Michel electron candidates within 600 mm long, 600 mm diameter cylinders centered on the neutrino vertex and on endpoints of tracked particles, and by counting isolated clusters constructed from two-dimensional clusters with at least 1 MeV visible energy. The former identify π^+, and the latter identify photons from π^0 decays. Clusters with an

Paper committee describes analyser code in English

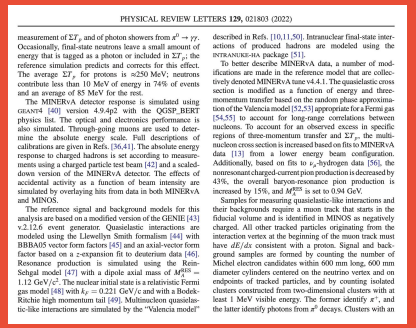
An

- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)

The Ill-posed Analysis Problem



Expert analyser writes code for signal/projection



PHYSICAL REVIEW LETTERS 129, 021803 (2022)

measurement of ΣT_γ and of photon showers from $\pi^0 \to \gamma\gamma$. Occasionally, final-state neutrons leave a small amount of energy that is tagged as a photon or included in ΣT_γ; the reference simulation predicts and corrects for this effect. The average ΣT_γ for protons is ≈250 MeV; neutrons contribute less than 10 MeV of energy in 74% of events and an average of 85 MeV for the rest.

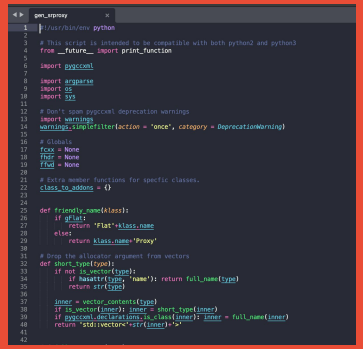
The MINERvA detector response is simulated using GEANT4 [40] version 4.9.4p2 with the QGSP_BERT physics list. The optical and electronics performance is also simulated. Through-going muons are used to determine the absolute energy scale. Full descriptions of calibrations are given in Refs. [36,41]. The absolute energy response to charged hadrons is set according to measurements using a charged particle test beam [42] and a scaled-down version of the MINERvA detector. The effects of accidental activity as a function of beam intensity are simulated by overlaying hits from data in both MINERvA and MINOS.

The reference signal and background models for this analysis are based on a modified version of the GENIE [43] v.2.12.6 event generator. Quasielastic interactions are modeled using the Llewellyn Smith formalism [44] with BBBA05 vector form factors [45] and an axial-vector form factor based on a z-expansion fit to deuterium data [46]. Resonance production is simulated using the Rein-Sehgal model [47] with a dipole axial mass of $M_A^{\mathrm{RES}} = 1.12\ \mathrm{GeV}/c^2$. The nuclear initial state is a relativistic Fermi gas model [48] with $p_F = 0.221$ GeV/c and with a Bodek-Ritchie high momentum tail [49]. Multinucleon quasielastic-like interactions are simulated by the "Valencia model" described in Refs. [10,11,50]. Intranuclear final-state interactions of produced hadrons are modeled using the INTRANUKE-hA package [51].

To better describe MINERvA data, a number of modifications are made in the reference model that are collectively denoted MINERvA tune v4.4.1. The quasielastic cross section is modified as a function of energy and three-momentum transfer based on the random phase approximation of the Valencia model [52,53] appropriate for a Fermi gas [54,55] to account for long-range correlations between nucleons. To account for an observed excess in specific regions of three-momentum transfer and ΣT_γ, the multinucleon cross section is increased based on fits to MINERvA data [13] from a lower energy beam configuration. Additionally, based on fits to ν_μ-hydrogen data [56], the nonresonant charged-current pion production is decreased by 43%, the overall baryon-resonance pion production is increased by 15%, and M_A^{RES} is set to 0.94 GeV.

Samples for measuring quasielastic-like interactions and their backgrounds require a muon track that starts in the fiducial volume and is identified in MINOS as negatively charged. All other tracked particles originating from the interaction vertex at the beginning of the muon track must have dE/dx consistent with a proton. Signal and background samples are formed by counting the number of Michel electron candidates within 600 mm long, 600 mm diameter cylinders centered on the neutrino vertex and on endpoints of tracked particles, and by counting isolated clusters constructed from two-dimensional clusters with at least 1 MeV visible energy. The former identify π^+, and the latter identify photons from π^0 decays. Clusters with an

Paper committee describes analyser code in English



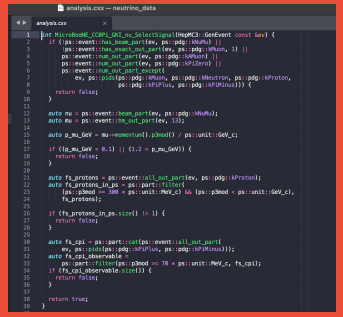
Non-expert user implements prediction code

An

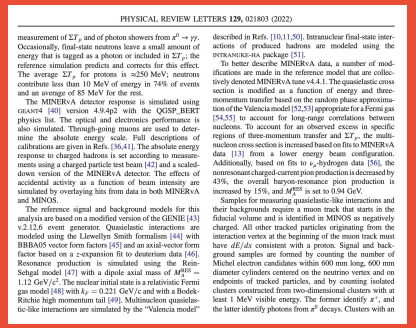
- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)

The Ill-posed Analysis Problem

Lossy projection operator → **Ill-posed unfolding problem**



Expert analyser writes code for signal/projection


PHYSICAL REVIEW LETTERS 129, 021803 (2022)

measurement of ΣT_γ and of photon showers from $\pi^0 \to \gamma\gamma$. Occasionally, final-state neutrons leave a small amount of energy that is tagged as a photon or included in ΣT_γ; the reference simulation predicts and corrects for this effect. The average ΣT_γ for protons is ≈250 MeV; neutrons contribute less than 10 MeV of energy in 74% of events and an average of 85 MeV for the rest.

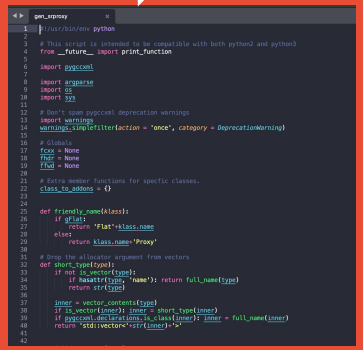
The MINERvA detector response is simulated using GEANT4 [40] version 4.9.4p2 with the QGSP_BERT physics list. The optical and electronics performance is also simulated. Through-going muons are used to determine the absolute energy scale. Full descriptions of calibrations are given in Refs. [36,41]. The absolute energy response to charged hadrons is set according to measurements using a charged particle test beam [42] and a scaled-down version of the MINERvA detector. The effects of accidental activity as a function of beam intensity are simulated by overlaying hits from data in both MINERvA and MINOS.

The reference signal and background models for this analysis are based on a modified version of the GENIE [43] v.2.12.6 event generator. Quasielastic interactions are modeled using the Llewellyn Smith formalism [44] with BBBA05 vector form factors [45] and an axial-vector form factor based on a z-expansion fit to deuterium data [46]. Resonance production is simulated using the Rein-Sehgal model [47] with a dipole axial mass of $M_A^{\mathrm{RES}} = 1.12\ \mathrm{GeV}/c^2$. The nuclear initial state is a relativistic Fermi gas model [48] with $k_F = 0.221\ \mathrm{GeV}/c$ and with a Bodek-Ritchie high momentum tail [49]. Multinucleon quasielastic-like interactions are simulated by the "Valencia model" described in Refs. [10,11,50]. Intranuclear final-state interactions of produced hadrons are modeled using the INTRANUKE-hA package [51].

To better describe MINERvA data, a number of modifications are made in the reference model that are collectively denoted MINERvA tune v4.4.1. The quasielastic cross section is modified as a function of energy and three-momentum transfer based on the random phase approximation of the Valencia model [52,53] appropriate for a Fermi gas [54,55] to account for long-range correlations between nucleons. To account for an observed excess in specific regions of three-momentum transfer and ΣT_γ, the multinucleon cross section is increased based on fits to MINERvA data [13] from a lower energy beam configuration. Additionally, based on fits to ν_μ-hydrogen data [56], the nonresonant charged-current pion production is decreased by 43%, the overall baryon-resonance pion production is increased by 15%, and M_A^{RES} is set to 0.94 GeV.

Samples for measuring quasielastic-like interactions and their backgrounds require a muon track that starts in the fiducial volume and is identified in MINOS as negatively charged. All other tracked particles originating from the interaction vertex at the beginning of the muon track must have dE/dx consistent with a proton. Signal and background samples are formed by counting the number of Michel electron candidates within 600 mm long, 600 mm diameter cylinders centered on the neutrino vertex and on endpoints of tracked particles, and by counting isolated clusters constructed from two-dimensional clusters with at least 1 MeV visible energy. The former identify π^+, and the latter identify photons from π^0 decays. Clusters with an

Paper committee describes analyser code in English



Non-expert user implements prediction code

An

- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)

Overview of Encountered Approaches

Cross section data/Flux predictions/Error matrices/Smearing matrices:

- Plots in papers
- Tables in papers
- Text tables in supplementary materials
- ROOT files in supplementary materials
- File storage on experiment webspace/third party repository

Conceptually simple to release:
Values + minimal metadata (bin definitions, variable units, ...)

Just needs standardisation!

Analysis Steps:

- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)
-

Overview of Encountered Approaches

Cross section data/Flux predictions/Error matrices/Smearing matrices:

- Plots in papers
- Tables in papers
- Text tables in supplementary materials
- ROOT files in supplementary materials
- File storage on experiment webspace/thi

Conceptually simple to release:
Values + minimal metadata (bin definitions, variable units, ...)

Just needs standardisation!

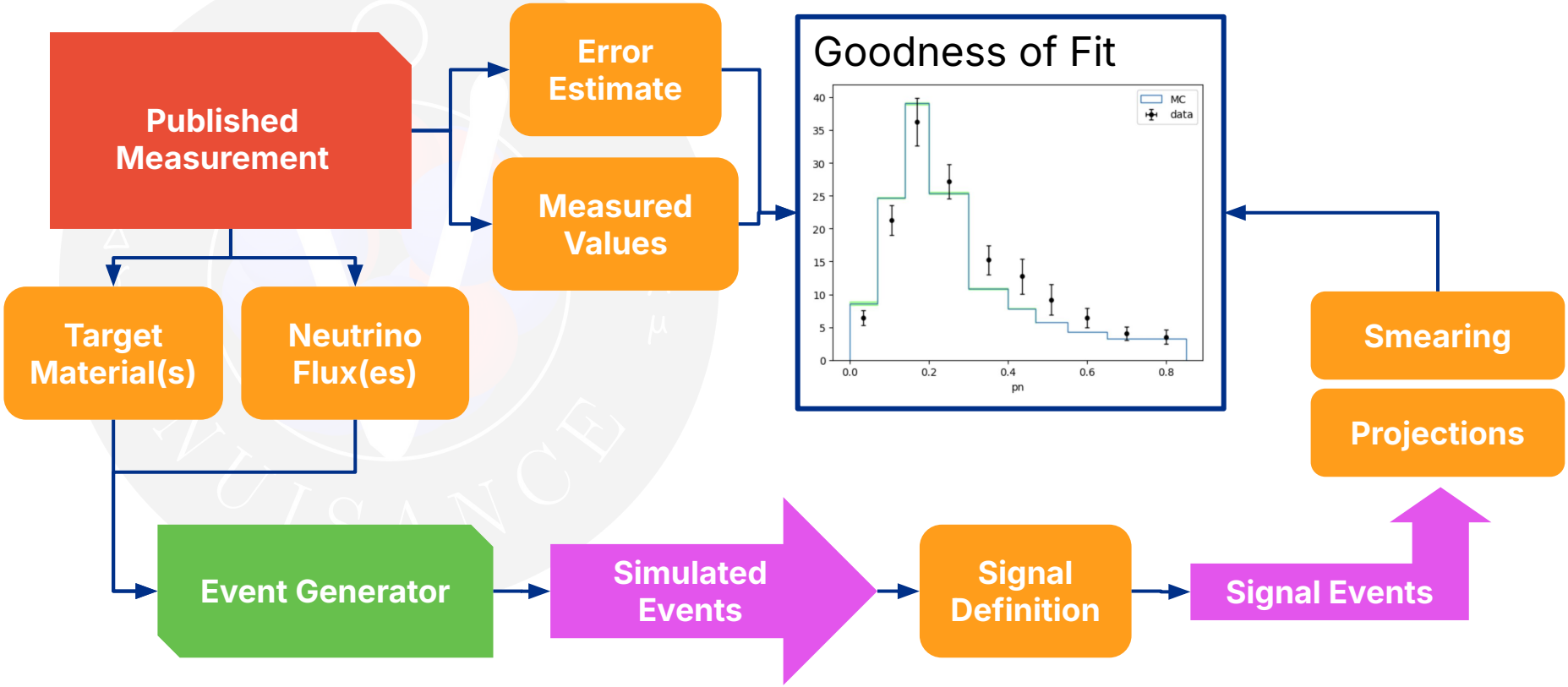
More difficult to release:
How to publish operators that are executable over a simulated event stream?

Just learn NUISANCE?

Analysis Steps:

- Vaguely hinted at throughout paper text
- Clearly signposted signal definition in paper and mathematical representation of independent projections
- Official example implementation (in NUISANCE or elsewhere)

Measurement Predictions: A Workflow



Personal View: The Way Forward

- Common Formats, Standards, Documentation:
  - **Data Release Aphorism:** Most of human progress has been driven by shared "languages"
- We need more community engagement on the 'boring' bits, so we can all do the interesting bits.

- Standardised, machine-readable formats for data with appropriate metadata
- Well-documented, open source, community tools to build/validate standardised data releases
- Measurements are getting more complicated, and more powerful!

A Solution

Use a standard, text-based format: HEPData YAML
Histogram + KV metadata

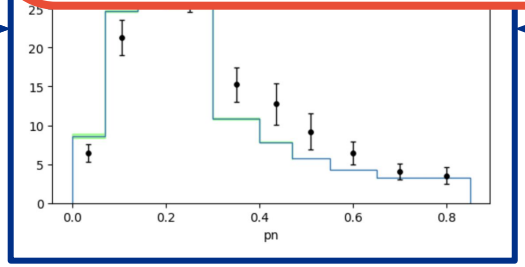
Published Measurement

Error Estimate

Measured Values

Target Material(s)

Neutrino Flux(es)



Smearing

Projections

Event Generator

Simulated Events

Signal Definition

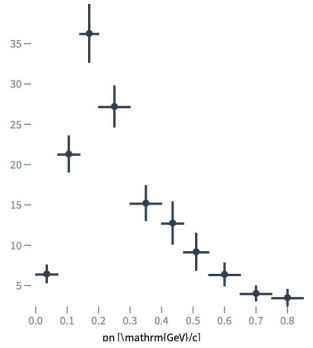
Signal Events

A Solution

Use a standard, text-based format: HEPData YAML
Histogram + KV metadata

variable_type	cross_section_measurement
selectfunc	analysis.cxx:MicroBooNE_CC0Pi_GKI_nu_SelectSignal
pn:projectfunc	analysis.cxx:MicroBooNE_CC0Pi_GKI_nu_pn
target	Ar
pn:prettyname	δp_n
prettyname	$\mathrm{d}\sigma/\mathrm{d}\delta p_n$
cross_section_units	1e-38 cm2|PerTarget|per_bin_width
probe_flux	microboone_flux_numu
errors	covariance-deltapn
smearing	smearing-deltapn
pn [GeV/c]	**cross_section** [cm^2 c/GeV /Ar]
0.0 - 0.07	6.4406 ±1.1679 total
0.07 - 0.14	21.314 ±2.2968 total
0.14 - 0.2	36.266 ±3.6505 total
0.2 - 0.3	27.206 ±2.6118 total
0.3 - 0.4	15.223 ±2.2399 total
0.4 - 0.47	12.758 ±2.6894 total

Visualize



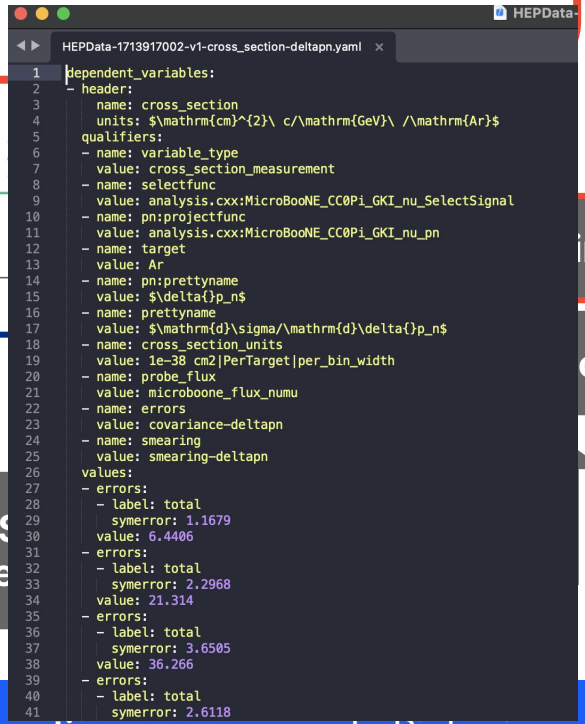
Sum errors ☑ Log Scale (X) ☐ Log Scale (Y) ☐

Deselect variables or hide different error bars by clicking on them.

Variables

cross_section [cm^2 c/GeV /Ar]
variable_type:cross_section_measurement
Summed error

https://www.hepdata.net/record/sandbox/1713917002



A Solution

Use a standard, event record format: NuHepMC, based on HepMC3

Error Estimate

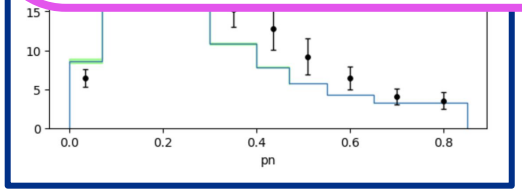
SciPost Physics Codebases Submission

NuHepMC: A standardized event record format for neutrino event generators

Steven Gardiner[1⋆], Joshua Isaacson[1†] and Luke Pickering[2‡]

1 Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA
2 STFC, Rutherford Appleton Laboratory, Harwell Oxford, United Kingdom

⋆ gardiner@fnal.gov, † isaacson@fnal.gov, ‡ luke.pickering@stfc.ac.uk



Smearing

Projections

Event Generator → Simulated Events → Signal Definition → Signal Events

A Solution

Use a standard, event record format: NuHepMC, based on HepMC3

Implementations for: NEUT, NuWro, GENIE, GiBUU, ACHILLES, Marley

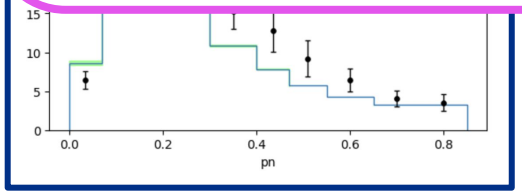
SciPost Physics Codebases Submission

NuHepMC: A standardized event record format for neutrino event generators

Steven Gardiner[1⋆], Joshua Isaacson[1†] and Luke Pickering[2‡]

1 Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA
2 STFC, Rutherford Appleton Laboratory, Harwell Oxford, United Kingdom

⋆ gardiner@fnal.gov, † isaacson@fnal.gov, ‡ luke.pickering@stfc.ac.uk



Error Estimate

Smearing

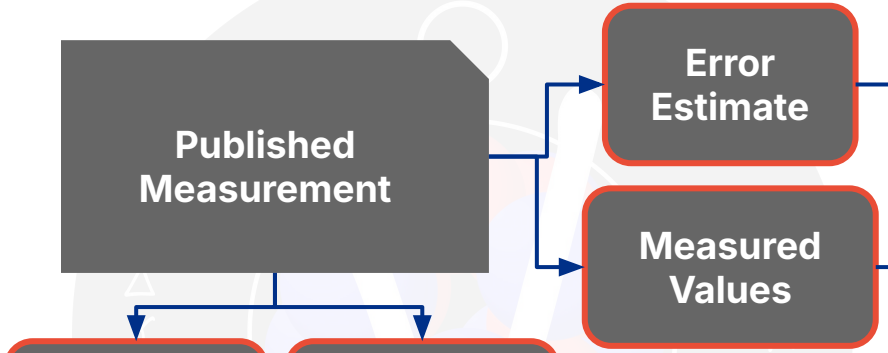
Projections

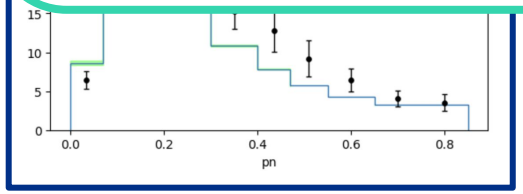
Event Generator → Simulated Events → Signal Definition → Signal Events

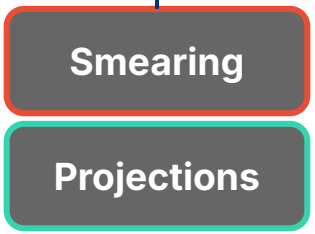
A Solution

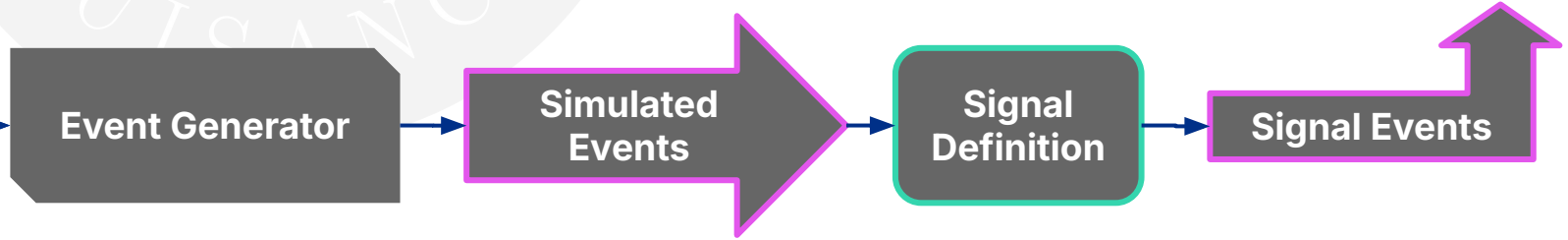
Use a standard, minimally-dependent set of tools for defining event processors in code:

ProSelecta









A Solution

Use a standard, minimally-dependent, declarative set of tools for writing portable event processors:

ProSelecta

Error Estimate

HEPData

Search HEPData

Search

Sandbox About Submission Help

Browse all View all Publication Data ProSelecta File

Hide Publication Information

Resources

Abstract
No abstract available.

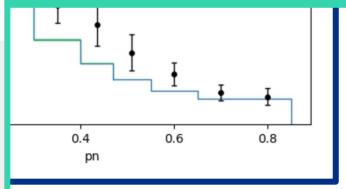
analysis.cxx

License: CC0

Selection and projection function examples. Can be executued in the ProSelecta environment v1.0.

```
int T2K_nueCCPiplus_2024_Select(HepMC3::GenEvent const &ev) {
 using namespace ps::pdg;

 if (!ps::event::has_beam_part(ev, kNuE) ||
    !ps::event::has_at_least_out_part(ev, ps::pids(kElectron, kPiPlus),
                                      {1, 1})) {
  return false;
 }

 return true;
}

double T2K_nueCCPiplus_2024_ElectronMomentum(HepMC3::GenEvent const &ev) {
 using namespace ps::pdg;

 if (!ps::event::has_beam_part(ev, kNuE) ||
    !ps::event::has_out_part(ev, kElectron)) {
  return ps::kMissingDatum<double>;
 }

 return ps::event::hm_out_part(ev, kElectron)->momentum().p3mod() /
     ps::unit::GeV;
}
```

Download



Smearing

Projections

Signal Definition

Signal Events

MicroBooNE GKI in Action

Example notebook

https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.

[2]:
```
import pyNUISANCE as pn

rf = pn.RecordFactory()
hepdata_rec = rf.make_record({"type":"hepdata",
    "recordref":"hepdata-sandbox:1713917002"})
```

MicroBooNE GKI in Action

Example notebook

https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analyses

```
print(hepdata_rec.get_analyses())
```

```
['cross_section-deltapn', 'cross_section-deltaalpha3dq', 'cross_section-deltaphi3d', 'cross_section-deltapnpar', 'cross_section-deltapnperp', 'cross_section-deltapnperpx', 'cross_section-deltapnperpy', 'cross_section-serialdeltapn_deltaalpha3dq', 'cross_section-serialdeltaalpha3dq_deltapn']
```

MicroBooNE GKI in Action

[Example notebook]

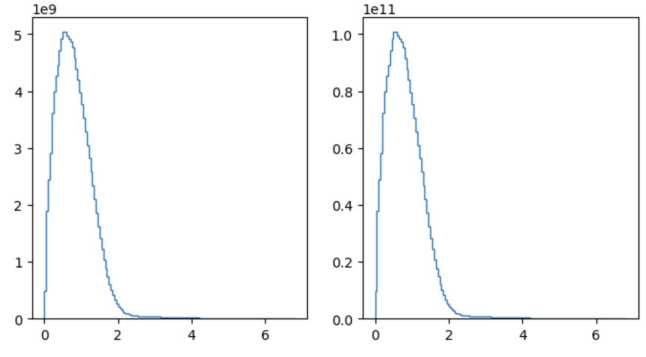
https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analyses
3. Interrogate analysis for flux/target information

```
[6]: ana = hepdata_rec.analysis("cross_section-deltapnperpx")
```

```
[10]: [ str(tgt) for tgt in ana.get_target() ]

[10]: ['{ A = 40, Z = 18, weight_by_mass = 1 }']
```

```
[7]: import matplotlib.pyplot as plt

probe_part, probe_flux_count = ana.get_probe_flux(False)
_, probe_flux_count_density = ana.get_probe_flux(True)

fig, axes = plt.subplots(1,2, figsize=(8,4))

probe_flux_count.mpl().hist(histtype="step",plot_axis=axes[0])
probe_flux_count_density.mpl().hist(histtype="step",plot_axis=axes[1])

plt.show()
```



MicroBooNE GKI in Action

[Example notebook]

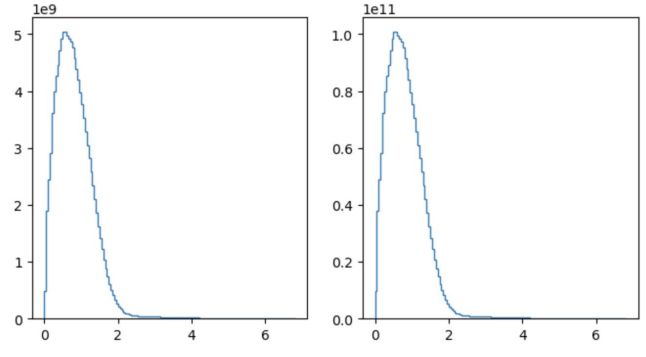
https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analyses
3. Interrogate analysis for flux/target information

```
[6]: ana = hepdata_rec.analysis("cross_section-deltapnperpx")
```

```
[10]: [ str(tgt) for tgt in ana.get_target() ]

[10]: ['{ A = 40, Z = 18, weight_by_mass = 1 }']
```

```
[7]: import matplotlib.pyplot as plt

probe_part, probe_flux_count = ana.get_probe_flux(False)
_, probe_flux_count_density = ana.get_probe_flux(True)

fig, axes = plt.subplots(1,2, figsize=(8,4))

probe_flux_count.mpl().hist(histtype="step",plot_axis=axes[0])
probe_flux_count_density.mpl().hist(histtype="step",plot_axis=axes[1])

plt.show()
```



MicroBooNE GKI in Action

Example notebook

https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analyses
3. Interrogate analysis for flux/target information
4. Load in event stream

[13]:
```
neut_events = pn.EventSource("events_for_inspirehep_2709091v1.hepmc3.gz")
if not neut_events:
    print("Failed to read file")
```

MicroBooNE GKI in Action

Example notebook

https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analyses
3. Interrogate analysis for flux/target information
4. Load in event stream
5. Build comparison

```
[14]: comparison = ana.process(neut_events)
```

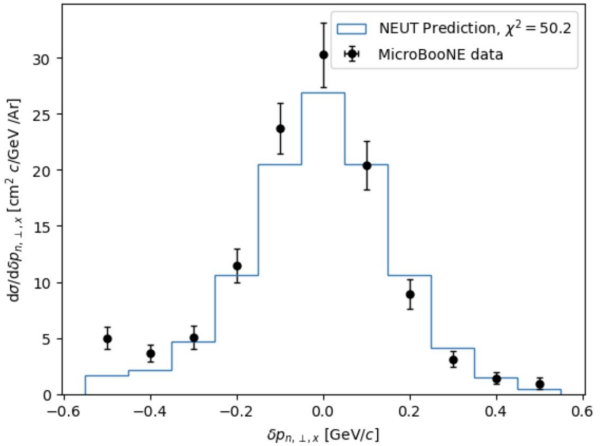
MicroBooNE GKI in Action

Example notebook

https://www.hepdata.net/record/sandbox/1713917002

1. Download and parse HEPData record:
   a. Read data, errors, smearing, compile selection/projection operations. Fully automated.
2. Interrogate for available analy
3. Interrogate analysis for flux/ta
4. Load in event stream
5. Build comparison
6. Make plots

```
[15]: comparison.data[0].mpl().data_hist(label="MicroBooNE data")
      comparison.predictions[0].mpl().hist(histtype="step",label=r"NEUT Prediction, $\chi^{2}=$"f"{comparison.likelihood():
      print(comparison.likelihood())
      plt.legend()
      plt.plot()

      50.22179776047446
[15]: []
```



How to Get Involved

Experimentalists:

- Publish new measurement data releases in a standardised format
  - **Bonus points:** convince your experimental collaboration to use a standard format
- **Be brave:** Read the ProSelecta specification and use it in defining your measurement

Theorists:

- If a measurement you want to use isn't in the common format, convert it! We can help.
- **Be brave**:
  - Use HepMC3, use ProSelecta, use a container if you're worried about building the software
  - Get outside your comfort zone, that's where the useful data is and will continue to be

How to Get Involved

Expe

- P
- B

Theo

- If
- B

```
picker24/pittpaccbox:root $ dnf install root
Last metadata expiration check: 0:00:13 ago on Fri Oct 25 07:55:57 2024.
Dependencies resolved.
================================================================================
 Package                 Architecture  Version                 Repository  Size
================================================================================
Installing:
 root                    aarch64       6.32.06-1.el9           epel       156 k
Installing dependencies:
 annobin                 aarch64       12.31-2.el9             appstream  1.0 M
 dwz                     aarch64       0.14-3.el9              appstream  125 k
 efi-srpm-macros         noarch        6-2.el9_0.0.1           appstream   21 k
 file                    aarch64       5.39-16.el9             baseos      48 k
 fonts-srpm-macros       noarch        1:2.0.5-7.el9.1         appstream   27 k
 gcc-plugin-annobin      aarch64       11.4.1-3.el9.alma.1     appstream   43 k
 ghc-srpm-macros         noarch        1.5.0-6.el9             appstream  7.8 k
 glibc-gconv-extra       aarch64       2.34-100.el9_4.3        baseos     1.6 M
 gnu-free-fonts-common   noarch        20120503-25.el9         epel       122 k
 gnu-free-mono-fonts     noarch        20120503-25.el9         epel       414 k
 gnu-free-sans-fonts     noarch        20120503-25.el9         epel       740 k
 gnu-free-serif-fonts    noarch        20120503-25.el9         epel       1.4 M
 go-srpm-macros          noarch        3.2.0-3.el9             appstream   26 k
 json-devel              aarch64       3.11.3-1.el9            epel       156 k
 kernel-srpm-macros      noarch        1.0-13.el9              appstream   15 k
 lua-srpm-macros         noarch        1-6.el9                 appstream  8.4 k
 ocaml-srpm-macros       noarch        6-6.el9                 appstream  7.7 k
 openblas-srpm-macros    noarch        2-11.el9                appstream  7.3 k
 perl-srpm-macros        noarch        1-41.el9                appstream  8.1 k
 pyproject-srpm-macros   noarch        1.12.0-1.el9            appstream   12 k
 python-srpm-macros      noarch        3.9-53.el9              appstream   17 k
 qt5-srpm-macros         noarch        5.15.9-1.el9            appstream  7.8 k
 redhat-rpm-config       noarch        207-1.el9.alma.1        appstream   66 k
 root-cling              aarch64       6.32.06-1.el9           epel        27 M
 root-core               aarch64       6.32.06-1.el9           epel        68 M
```

  - Get outside your comfort zone, that's where the useful data is and will continue to be

Linky Dinky Dump

- ProSelecta: https://github.com/NUISANCEMC/ProSelecta
- NuHepMC:
  - HepMC3: https://gitlab.cern.ch/hepmc/HepMC3
  - Specification: https://arxiv.org/pdf/2310.13211
  - C++/python tools: https://github.com/NuHepMC/cpputils
- HEPData:
  - Example HEPData Record: https://www.hepdata.net/record/sandbox/1713917002
  - HEPData YAML format: https://hepdata-submission.readthedocs.io/en/latest/submission_yaml.html
  - HEPData python tools: https://github.com/HEPData/hepdata_lib/
  - NUISANCE conventions and tools: https://github.com/NUISANCEMC/HEPData
  - Some example of converted neutrino data sets: https://github.com/NUISANCEMC/neutrino_data
- NUISANCE slack: https://www.nuisance-xsec.slack.com

//What I Didn't Cover

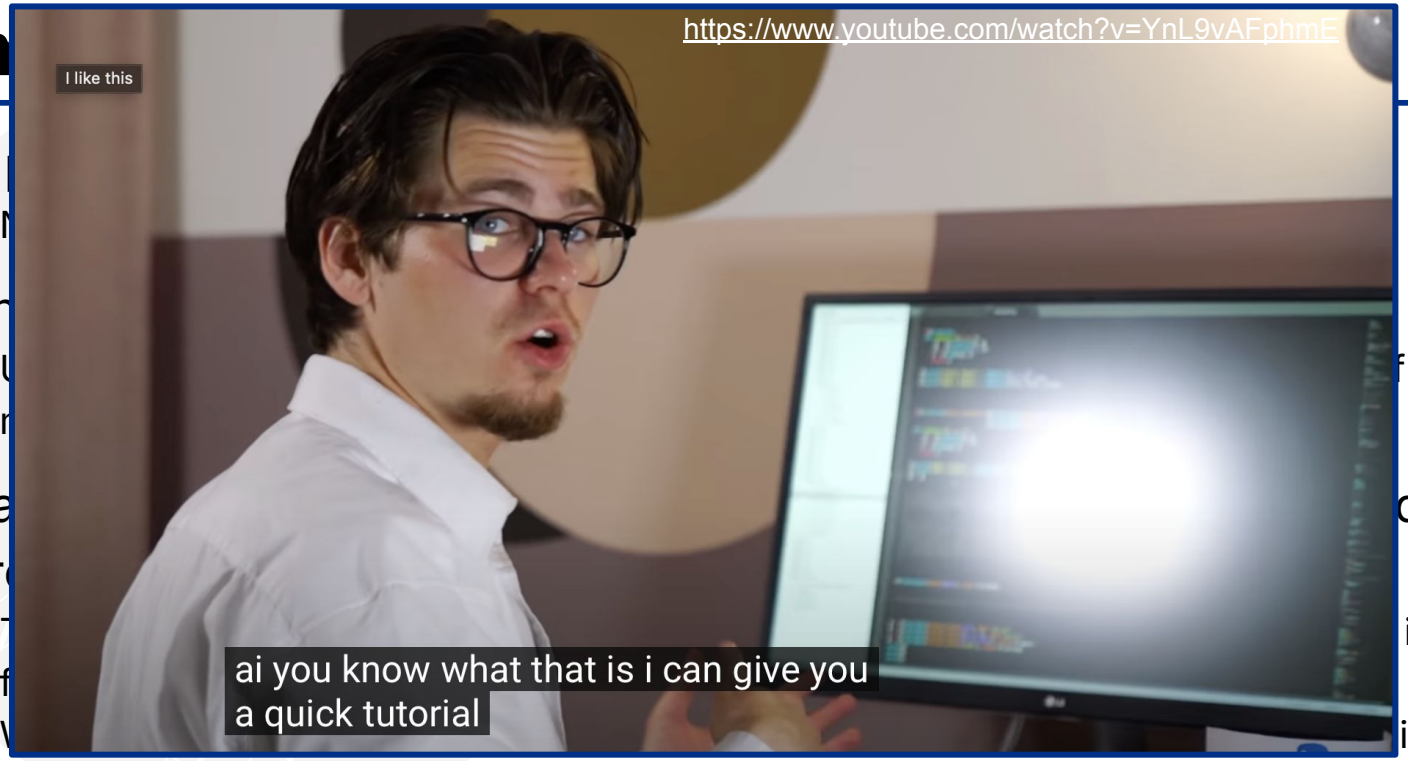
- Standards for releasing and using data and MC events from experiments
  - c.f. MINERvA data-preservation effort
  - Fantastic effort, I think the problem domain isn't yet fully explored, so the solution requires more work. That should be done!
  - Here's we're talking about easy things that we should just agree how to do and do them

Summary

- The problem domain is clear
  - No new research is needed
- We have sketched a solution that seems promising:
  - Unless a better one appears, we will work on converting the historical corpus of measurements to this format for use in NUISANCE3
- Engagement will improve the tools, reduce your cognitive load, and improve your productivity if you produce or use scattering data:
  - There is some upfront cost, but I think the increased productivity and reduction in scope for bugs/misinterpretation is worth it
  - We are committed to writing training material, supporting the tools and users using the tools
    - If you have ideas for training materials, please let us know: API docs, readthedocs, jupyter examples, youtube videos of Luke programing...

Summ

- The
  - N
- We h
  - U
    r
- Enga
  impr
  - T
    f
  - V



d

in scope

ing the tools

  - If you have ideas for training materials, please let us know: API docs, readthedocs, jupyter examples, youtube videos of Luke programing...

Summary

- The problem domain is clear
  - No new research is needed
- We have sketched a solution that seems promising:
  - Unless a better one appears, we will work on converting the historical corpus of measurements to this format for use in NUISANCE3
- Engagement will improve the tools, reduce your cognitive load, and improve your productivity if you produce or use scattering data:
  - There is some upfront cost, but I think the increased productivity and reduction in scope for bugs/misinterpretation is worth it
  - We are committed to writing training material, supporting the tools and users using the tools
    - If you have ideas for training materials, please let us know: API docs, readthedocs, jupyter examples, youtube videos of Luke programing...

Backup