Final method proposal

Where we left it

- Demonstrated that the fit is (just) within acceptable statistics.
- Unfolding was being slightly problematic:
  - What system to use?
  - Multi-dimensional unfolding
  - Unfolding each process
  - How to include systematics

Cross-section calculation

- What is a binned cross-section?
  - Using the expectation value of the bin interval, *assuming the probability of drawing each energy of the bin is uniform* (i.e. arithmetic mean)
- $\sigma_{E_1-E_2} = \frac{1}{E_2-E_1}\int_{E_1}^{E_2}\sigma dE$
- Equation for an infinitesimal energy slice:
- $$\sigma_p = f_p(E)\frac{dE}{dx}\frac{1}{dE}(-1)\ln\left(1+\frac{1}{N(E)}\frac{dN}{dE}dE+O(dE^2)\right)$$
  $$= -f_p(E)\frac{dE}{dx}\frac{1}{N(E)}\frac{dN}{dE}$$

Cross-section calculation

- What is a binned cross-section?
  - Using the expectation value of the bin interval, *assuming the probability of drawing each energy of the bin is uniform* (i.e. arithmetic mean)
- $\sigma_{E_1-E_2} = \frac{1}{E_2-E_1}\int_{E_1}^{E_2}\sigma dE$
- Equation for an infinitesimal energy slice:
- $$\sigma_p = f_p(E)\frac{dE}{dx}\frac{1}{dE}(-1)\ln\left(1+\frac{1}{N(E)}\frac{dN}{dE}dE+O(dE^2)\right)$$
  $$= -f_p(E)\frac{dE}{dx}\frac{1}{N(E)}\frac{dN}{dE}$$
- $f_p(E)$ is the fraction of events of process p at E.

Cross-section calculation

- Integration by parts, assuming dE/dx is constant:
- $\sigma_{p,E_1-E_2} = \frac{1}{E_2-E_1}\frac{dE}{dx}\left(\left[-f_p(E)\ln(N)\right]_{E_1}^{E_2} + \int_{E_1}^{E_2}\ln(N)\frac{df_p}{dE}\,dE\right)$
- This recovers the standard formula if the process fraction, $f_p(E)$, is constant over the bin.
  - If not, let's assume df_p/dx is constant, correction is $O(\frac{dN}{dE}(N\ln(N)-N))$
- For now, assume there's no correction term.

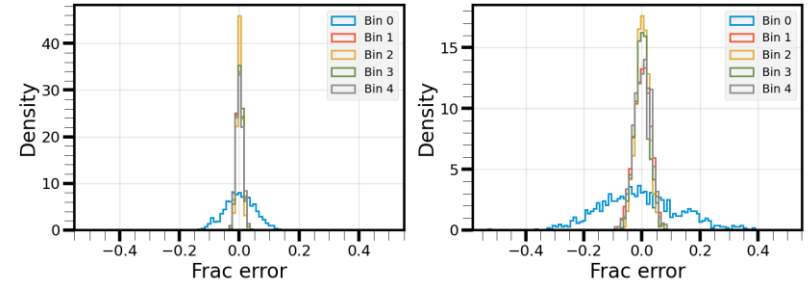
Incomplete slices

- $\sigma_{p,E_1-E_2} = \frac{1}{E_2-E_1}\frac{dE}{dx}\left[-f_p(E)\ln(N)\right]_{E_1}^{E_2}$
- Only has boundary terms.
- An incomplete slice (start and end in the same slice) does not affect this.
- **But**, f_p, the fraction of processes interacting with process p will still observe the interaction of the incomplete slice.
- I propose we treat the cross-section calculation and fractions separately

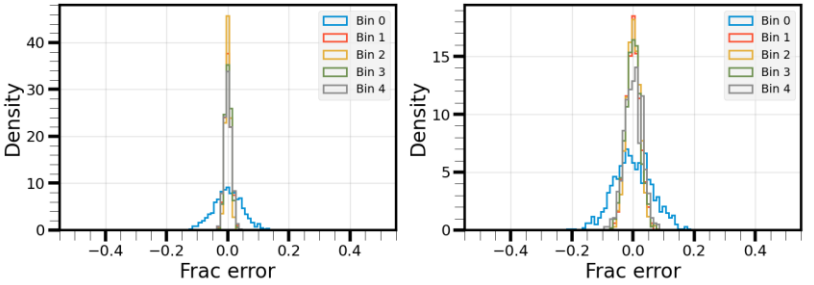
Splitting fractions

- Treating fractions and cross-sections separately has advantages:
  - Improved statistics – no events missing
  - Do not need to track initial energies
- Created a toy to test if this works:

Init energy: Gaussian, 3.0 +/- 0.05 GeV
Bins: 2.9-3.4 (0.1 width)
Flat cross-sections



Fraction from events in cross-section only



Fraction from all interactions in bin

Fit unfolding

- Can do multi-dimensional on energy only without worrying about the GNN fit.
- Still need unfold the fractions in each interacting bin onto the true interaction bins.
- I see 3 solutions:
  - Redo the unfolding with interactions only
  - Project the energy unfolding into the interaction dimension only
  - Project the energy unfolding, and iterate to account for cross-section shape

Iteratively fit unfolding

- The detector's energy response is independent of process.
- The interaction response by itself does not include any cross-section dependence (though the initial+interaction correlated response does).
- If the process fractions were independent of energy, we would not need to unfold the fractions
  - Since the binned cross-section definition is working under this assumption, one could try to justify not doing this.

Iteratively fit unfolding

- If the processes fractions do change in each bin, we should unfold
- Ideally the cross-section dependence is separate from detector energy dependence.
- If the ratio between bins changes, then based on the predicted cross-section change, the amount we unfold by should also change.
- We could try this process iteratively.
- I propose we project the interaction unfolding from energy, for simplicity with no iterative changes.

Alternate unfolding

- Probably not something we should do, but as an idea I considered:
- Fit 2D probability distributions to the *offset* of the initial and interaction energies.
- Throw many toys drawing from these distributions
- Gives central value and uncertainties.

Uncertainties

- Thin properly after getting central value sorted.
- RooUnfold uncertainties for 2D unfolding seem tricky.
- Consider creating weights based on uncertainties and running many toys.