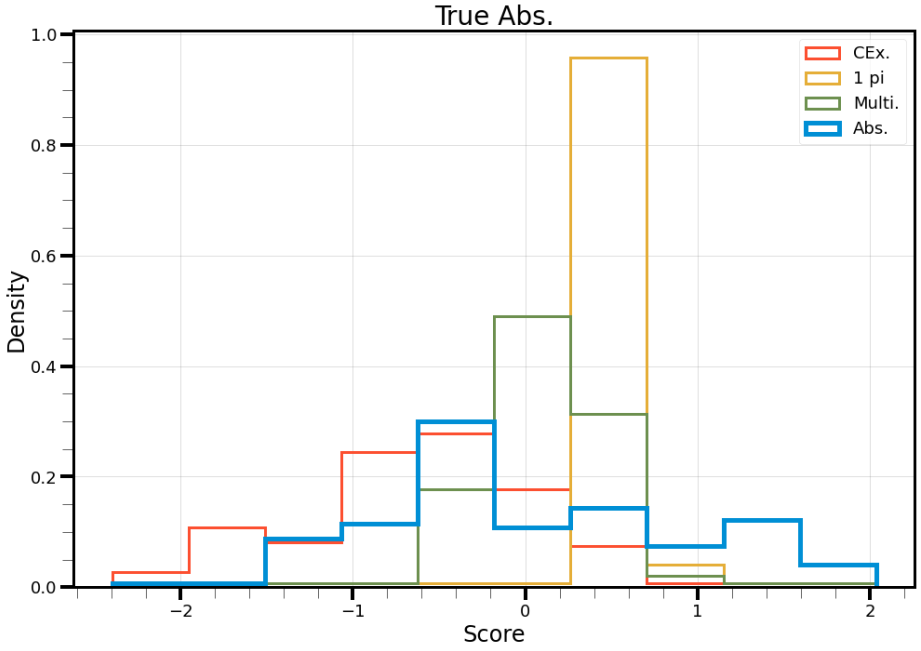
GNN score fitting tests

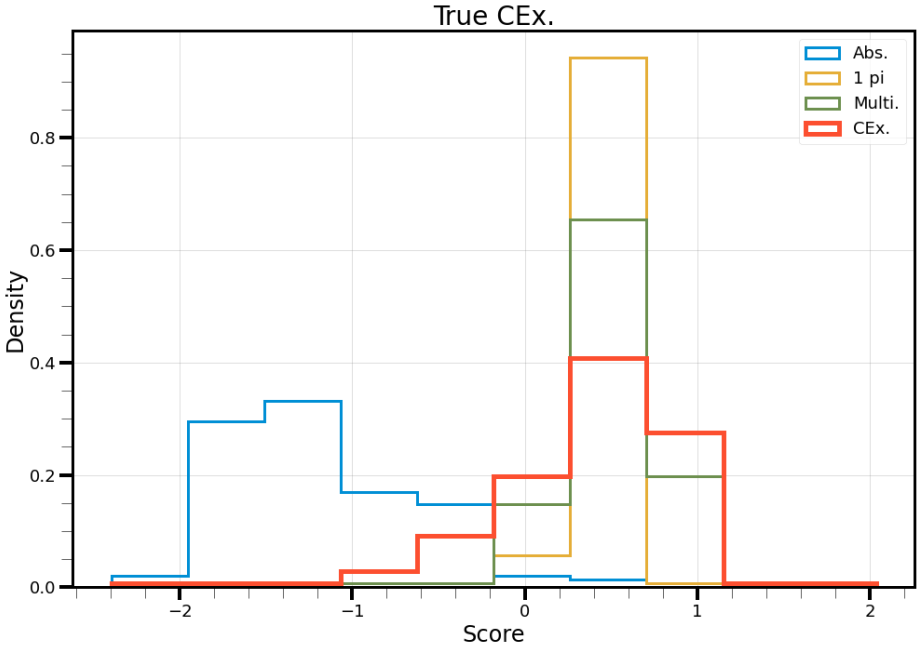
02.08.24

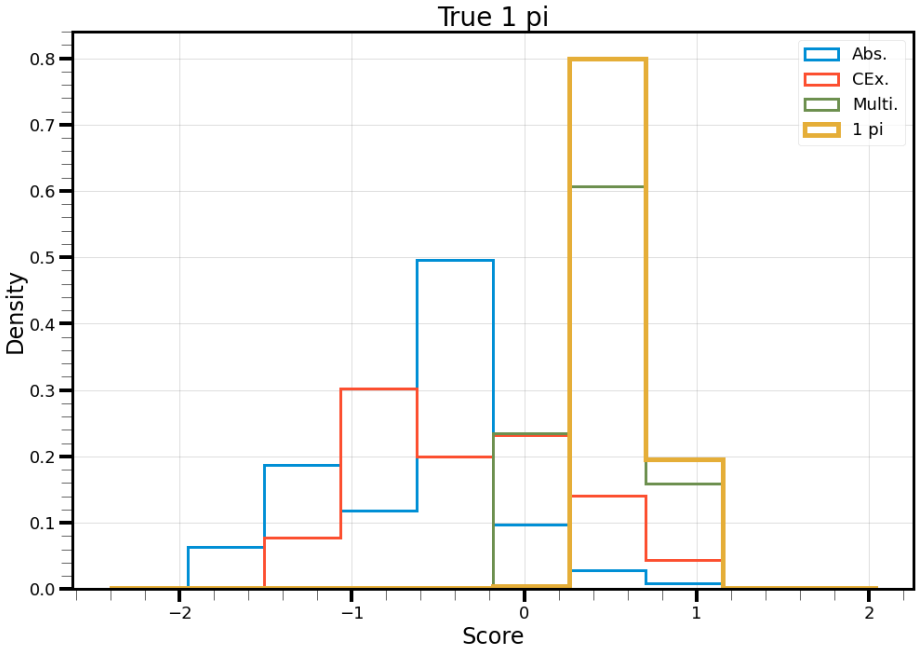
What data

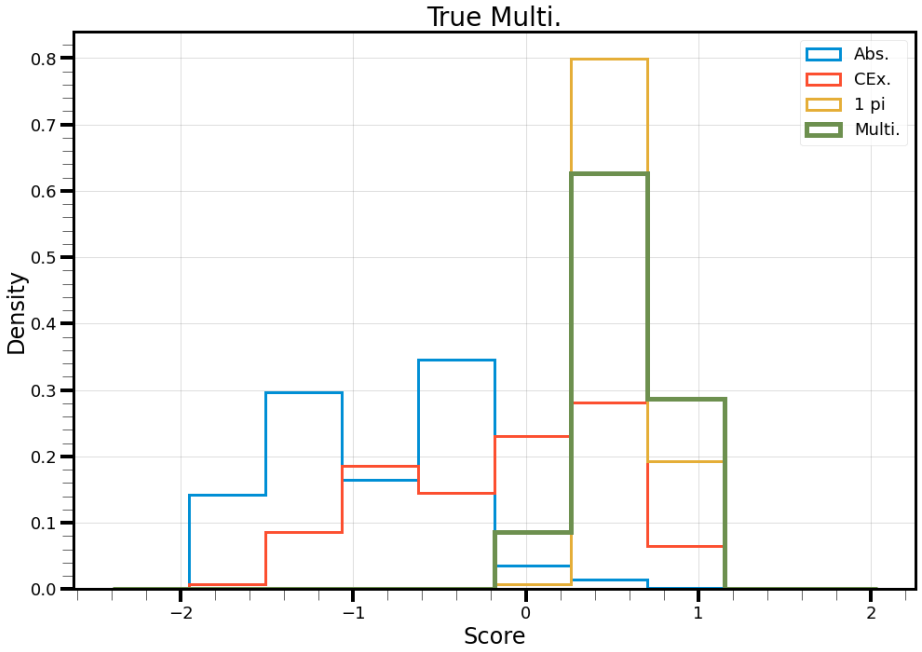
- Using ~80% of full 3GeV MC data set.
  - Meant to be using 100%!
- GNN training using Set00:
  - 80% training (->90%)
  - 10% validation (->10%)
- GNN fit templates using Set01:
  - 80% (->100%)
- Data to be fitted using Set02:
  - 80% (->100%)

Template data:



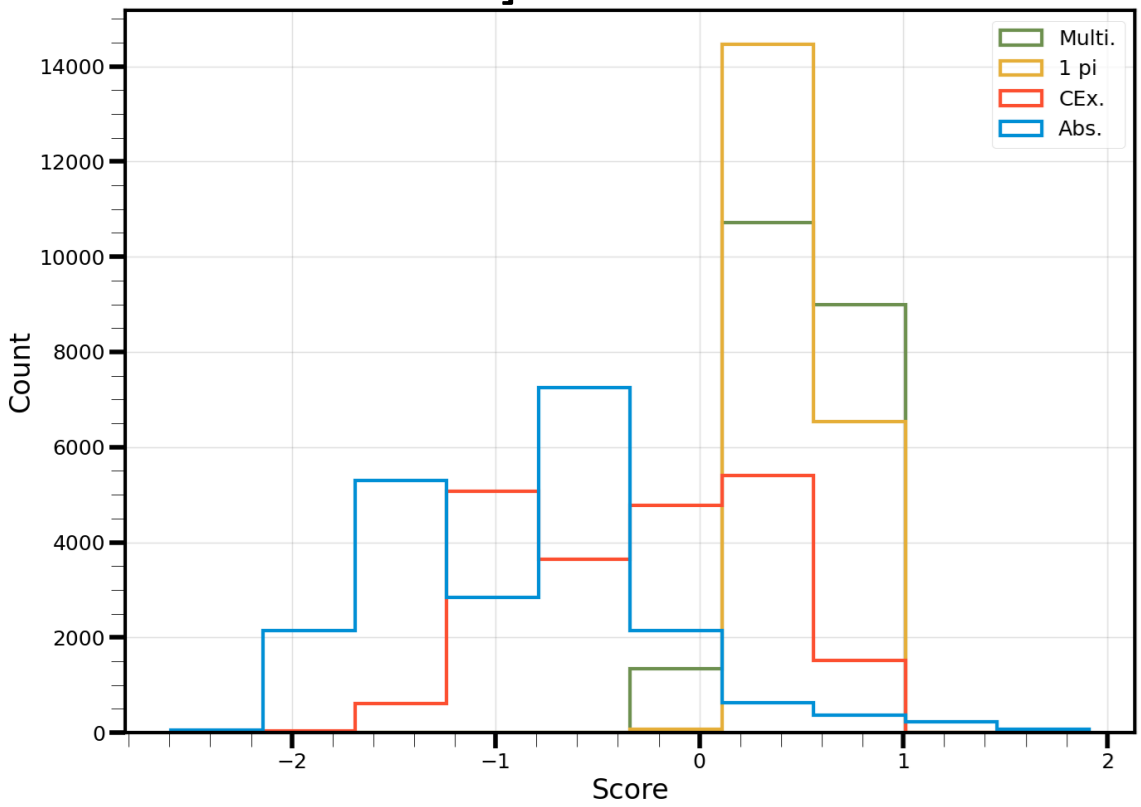






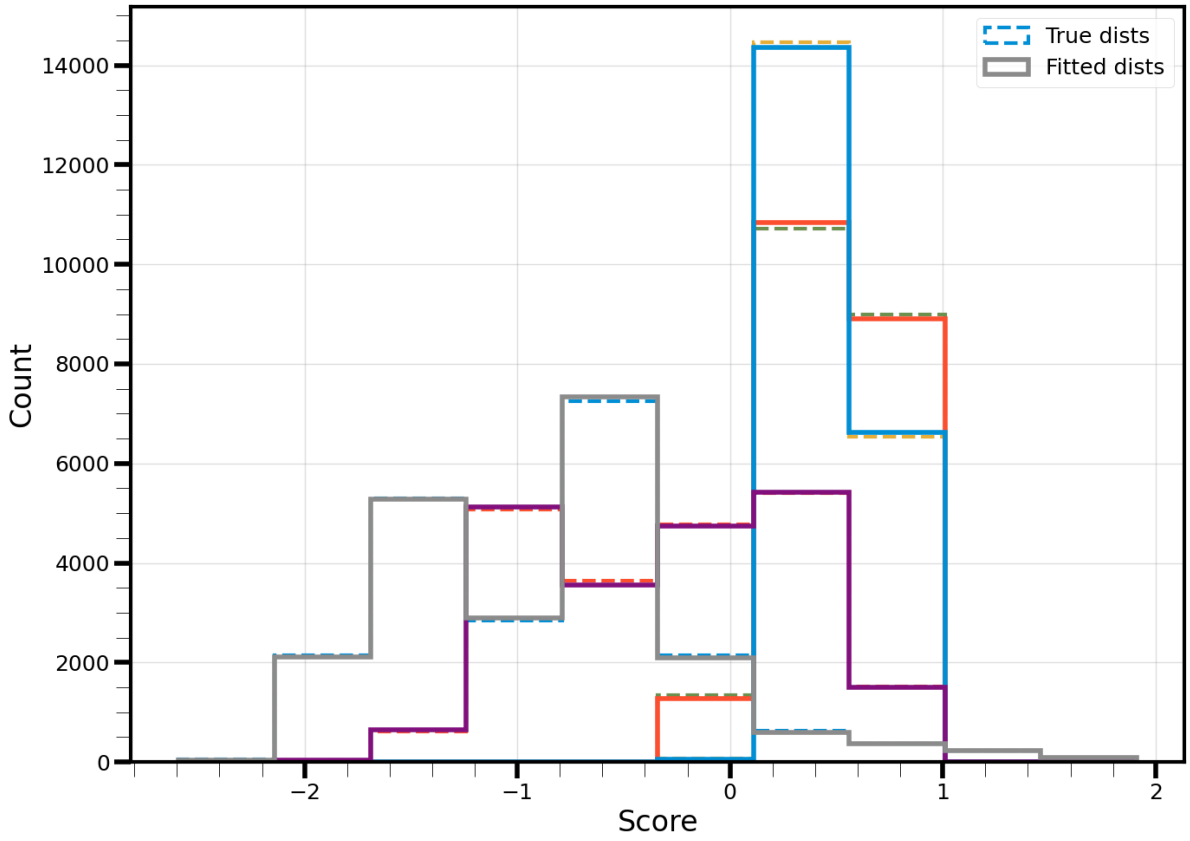
Unseen data

- True composition:
  - [Abs   CEx. 1 pi   Multi.]
  - [1330   991 4891 13844]



Fit

- Uncorrelated
- Assumes templates are Poisson.
  - Wrong! Should be multinomial.



- True vs. predicted vs. initial prediction (GNN class):
  - [1330 991 4891 13844]
  - [1322.5 675.0 4741.1 14291.6]
  - [747 2854 9884 7571]

Improved fit

- Correlate bins – template is now 1x 4D histogram with 3^4=81 bins
- Treat it properly as a multinomial distribution:
  - Should be easier to use uncertainties to deal with empty bins
  - Set p_i as $1/N$? Or $1/2N$?
  - N is the number of evts in template construction.
- Implement with Minuit (or similar) for better uncertainty handling.
  - Will require more faffing with environments…

Other goals

- Run analysis with top GNN score as selected region
- Presentation to Physics Analysis group
  - Talk from DUNE-UK with a bit more detail