Graph Neural Networks for pion event classification

Dennis Lindebaum

ProtoDUNE-SP Hadron Analysis

14.08.24

University of BRISTOL

Content

Context - what processes?

- 3GeV ProtoDUNE-SP pion beam data
- Measuring π^+ - Ar cross-sections
- We consider 4 possible interaction channels
  - Absorption: $\pi^+ + \mathrm{Ar} \rightarrow \mathrm{X}$
  - Charge exchange: $\pi^+ + \mathrm{Ar} \rightarrow \pi^0 + \mathrm{X}$
  - Single pion production: $\pi^+ + \mathrm{Ar} \rightarrow \pi^\pm + \mathrm{X}$
  - Multiple pion production: $\pi^+ + \mathrm{Ar} \rightarrow A\,\pi^\pm + B\,\pi^0 + \mathrm{X},$ $A + B \geq 2$

	$\pi^\pm$s	π^0s
Abs.	0	0
CEx.	0	1
1 pi	1	0
Multi.	otherwise	

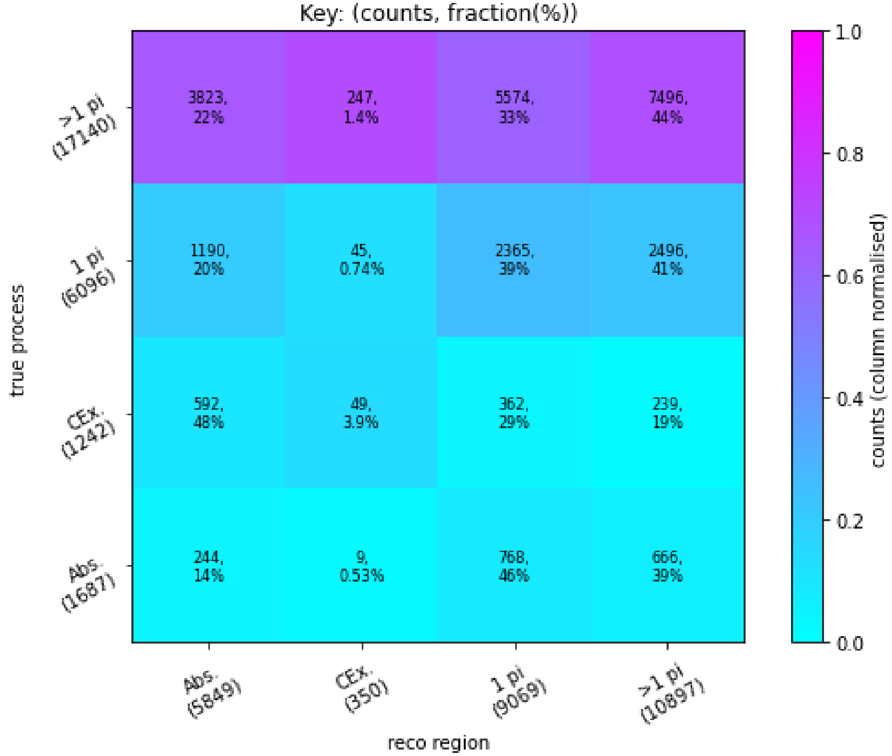
Note this combines:
Quasi elastic scattering: $\pi^+ + \mathrm{Ar} \rightarrow \pi^+ + \mathrm{X}$
Double charge exchange: $\pi^+ + \mathrm{Ar} \rightarrow \pi^- + \mathrm{X}$
(can't distinguish π^+ / π^-)

Context - previous work

- Using analysis framework developed by Shyam Bhuller in Bristol.
  - Find his most recent talk here: [ProtoDUNE-SP Hadron Analysis Bi-Weekly Meeting (May 1, 2024) · INDICO-FNAL (Indico)]
  - This work – use a GNN to perform process selection.
  - Other aspects unchanged (selection, energy slice, unfolding, etc.).
- Copies the beam selection used by Shyam, not yet done any 3GeV optimisation.
  - Not yet implemented the fiducial cut on first 30cm.
  - Not yet made a cut on low energy pions.

Context - why graphs?

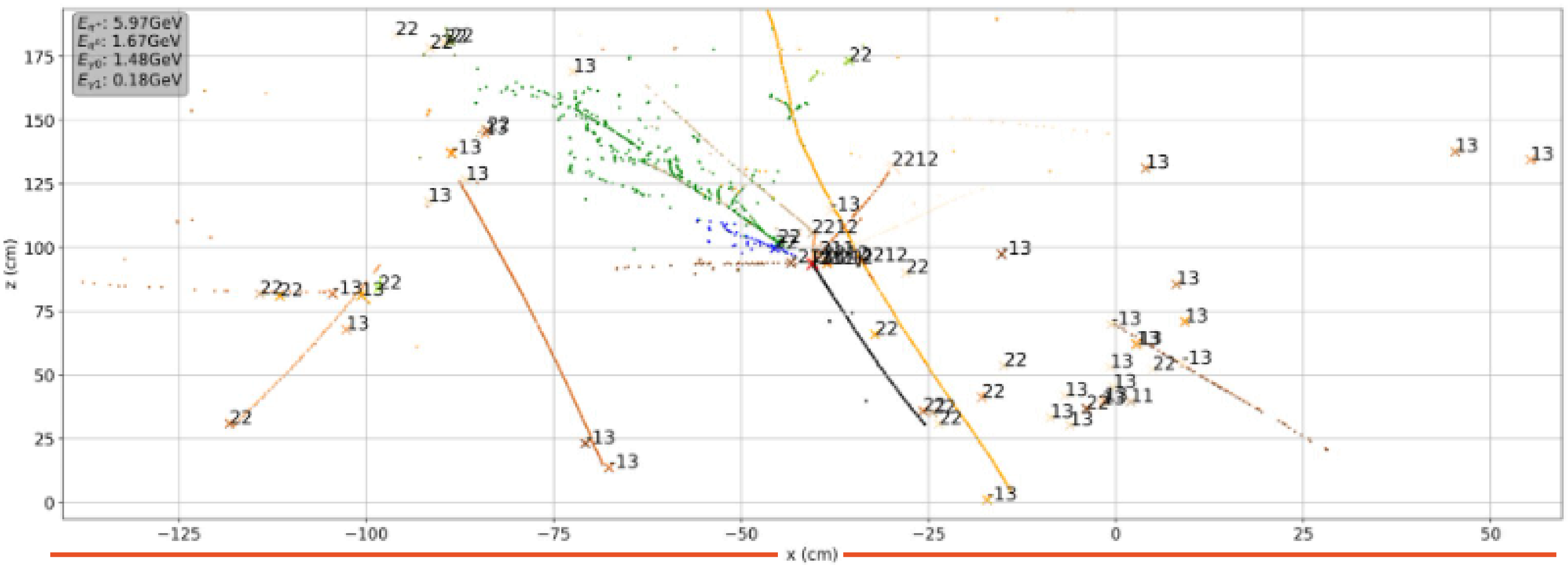
- Simple cuts-based method performs poorly.
- Identify PFOs as particles using cuts on the PFO properties.
- Count the identified PFOs to classify the event.



- *x*-axis is classification predicted by a method
- *y*-axis is the true classification (label)
- Colour shows purity: n_cell / sum_column
- Percentage shows efficiency: n_cell / sum_row

Context - why graphs?

- We don't know how many particles will appear in an event!
- A strategy is required to deal with an unknown number of inputs.
- Graph Neural Networks apply the same function all connected points, then perform *aggregation* to reduce to a known size.

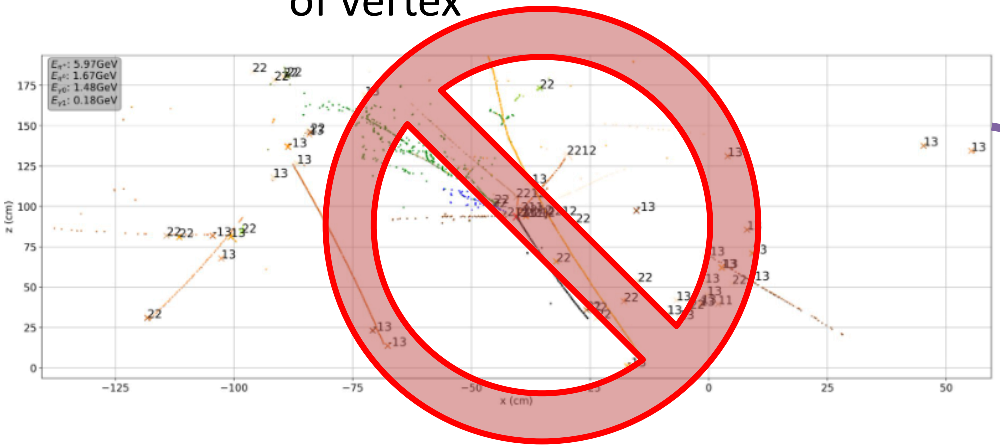


ProtoDUNE events as graphs

- Convert events to heterogeneous, undirected graph

Multiple types of vertex

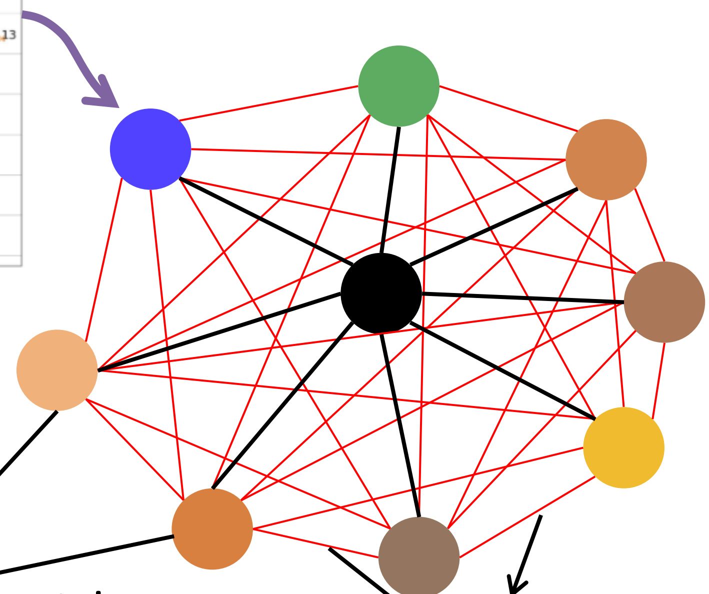
Edges are the same in both directions



Not looking at hits

Instead, properties calculated post-reconstruction

- Track score
- Pion/proton χ^2 fits
- Etc.



Vertices contain kinematic data for particle identification

Edges contain geometric information to determine relationships

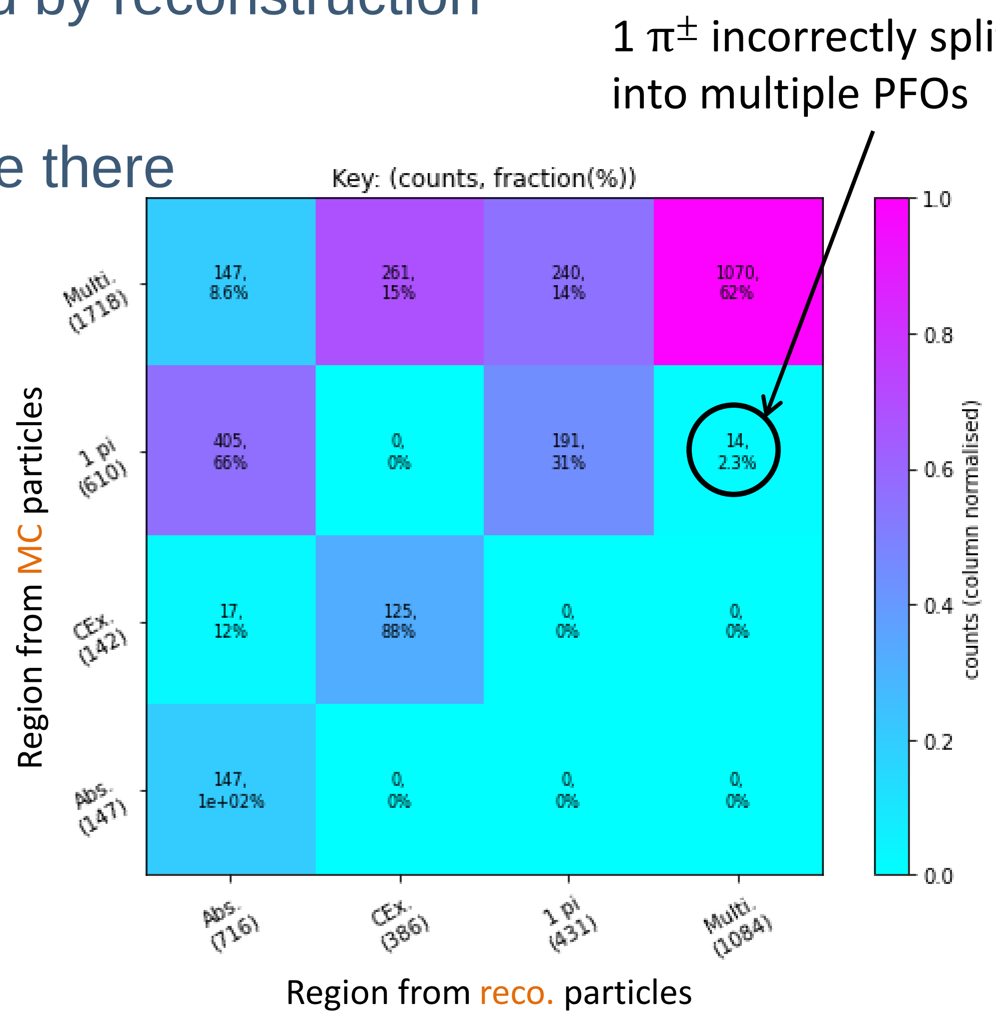
What data?

- Three levels of data used to test performance:

  1. Monte-Carlo (MC):
     - Contains all particles from the Geant4 generator.
     - Takes truth values (i.e. particle relations, particle species) from Geant4 particles.
     - GNN achieves perfect performance.
  2. Back-tracked (BT) (aka. cheated):
     - Contains only reconstructed objects (generated by pandora clustering).
     - Relates reco. PFOs to MC particles by tracking the proportion deposited energy, take truth information from the related MC particles
  3. Reconstructed data (reco.):
     - Reconstructed objects with reconstructed properties

Reconstruction effects

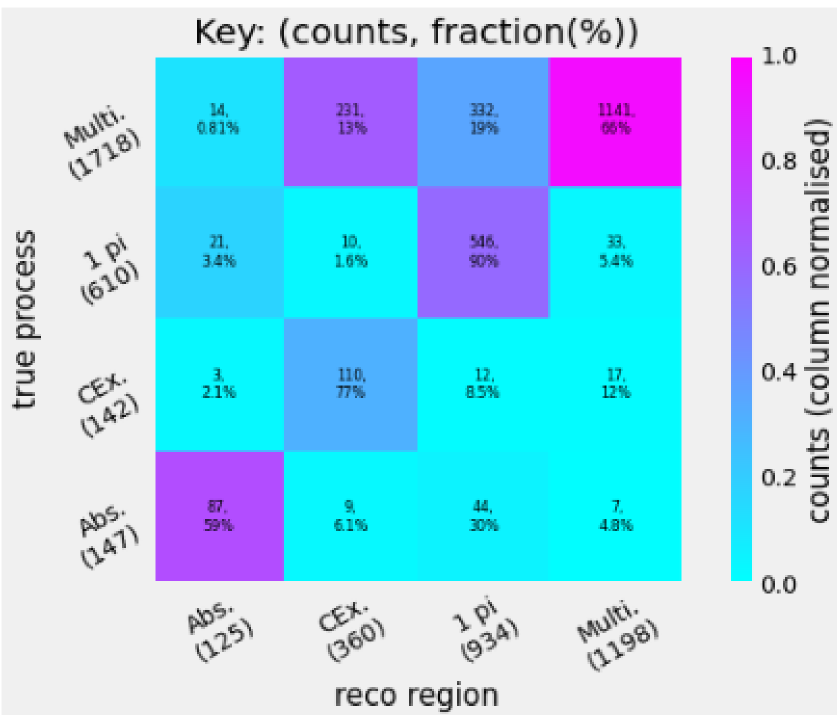
- Back-tracked data is affected by reconstruction efficiency.
- PFOs are only created where there is energy deposition in the detector *AND* the PFO clustering algorithms find this energy.
  - One of the failure modes is clustering hits from multiple particles as a single PFO.
  - Fewer PFOs per event than Monte-Carlo truth.



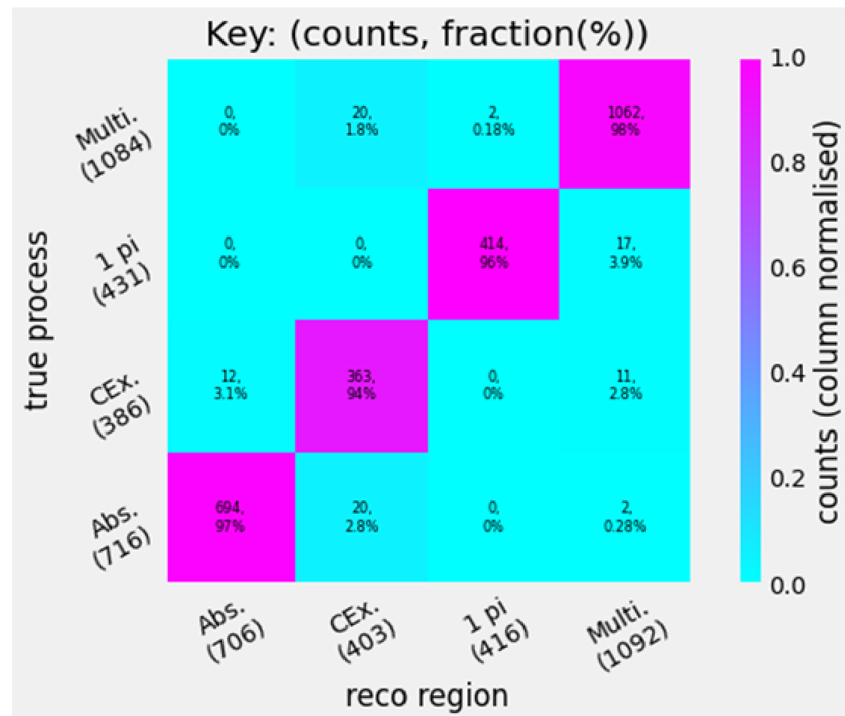
Back-tracked performance

- Can apply weightings to prefer efficiency over purity
  - Real test – which will produce lower uncertainty in the analysis

MC classified regions as true process



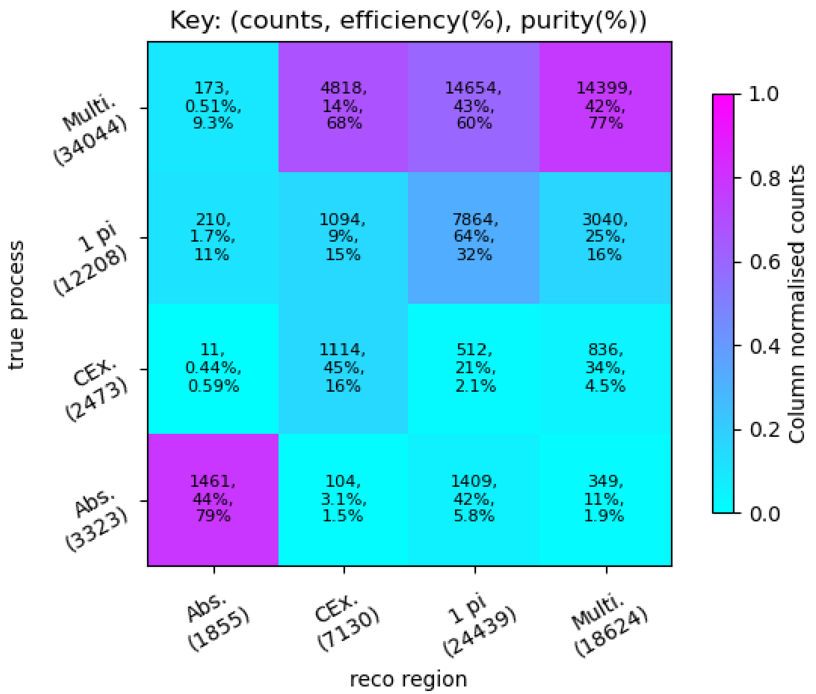
Reco. classified regions as true process



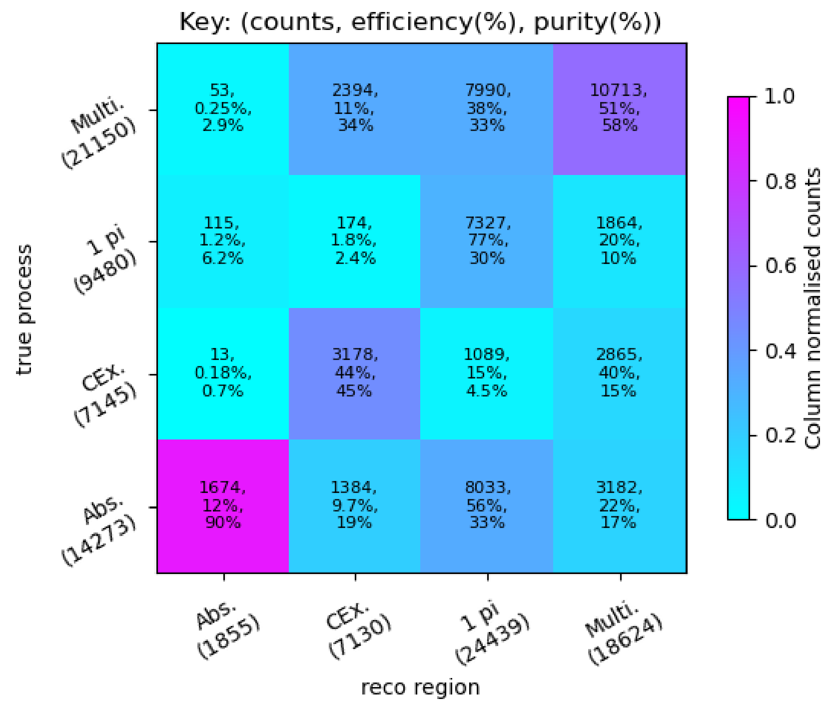
Reconstructed performance

- Performance further diminished using reconstructed properties
  - Biggest culprit seems to be $\pi^{\pm}$ identification.

MC classified regions as true process



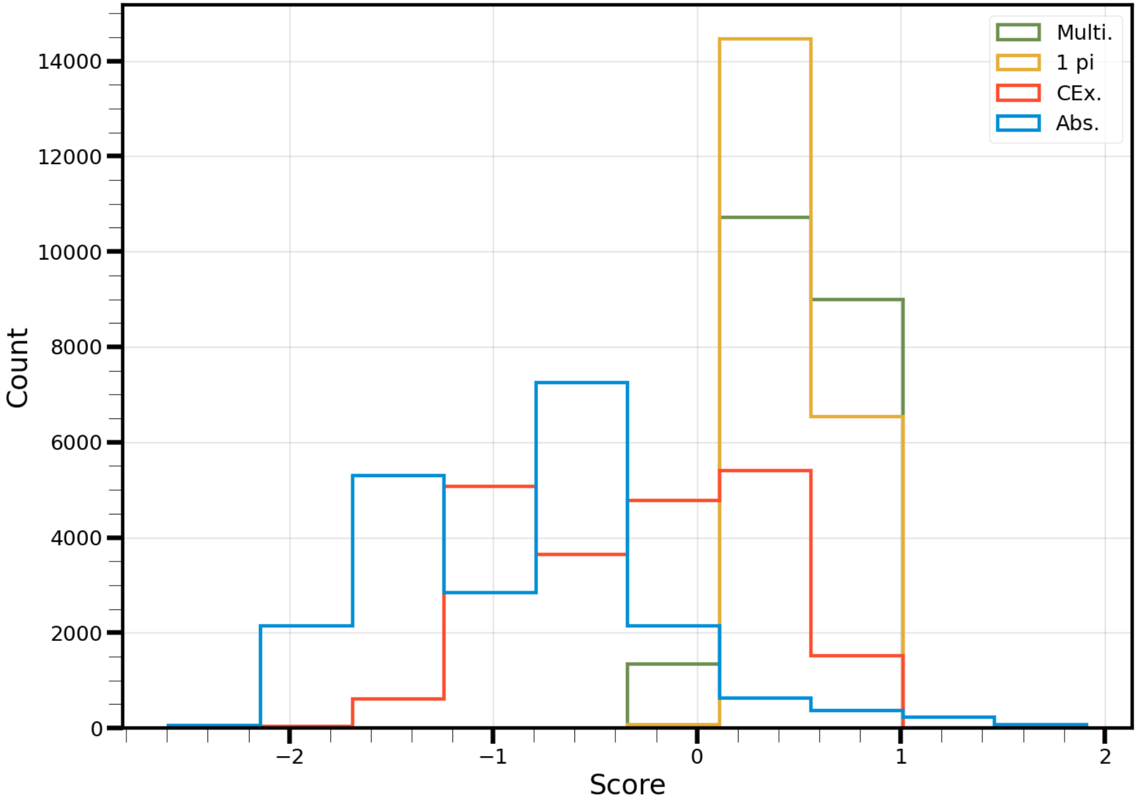
Reco. classified regions as true process



Use in analysis

- Instead of classifying individual events, fit the measured score distributions with true classification MC template distributions.

Distribution of GNN scores per channel across all test events

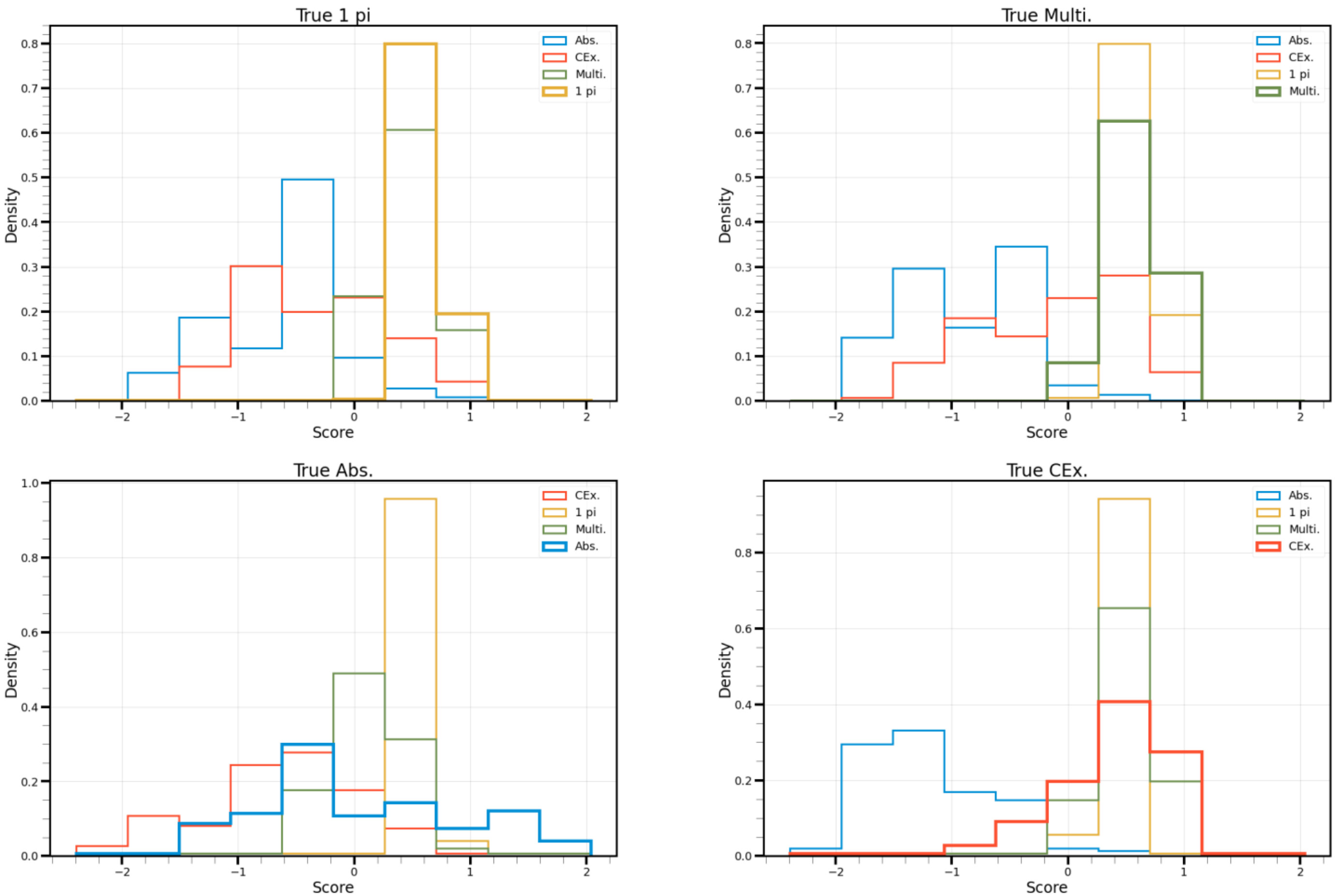


Fit - templates

- Split half the MC data as a template (actual template have 4D correlation)

MC classified single pion production

MC classified multi. pion production



MC classified absorption

MC classified charge exchange

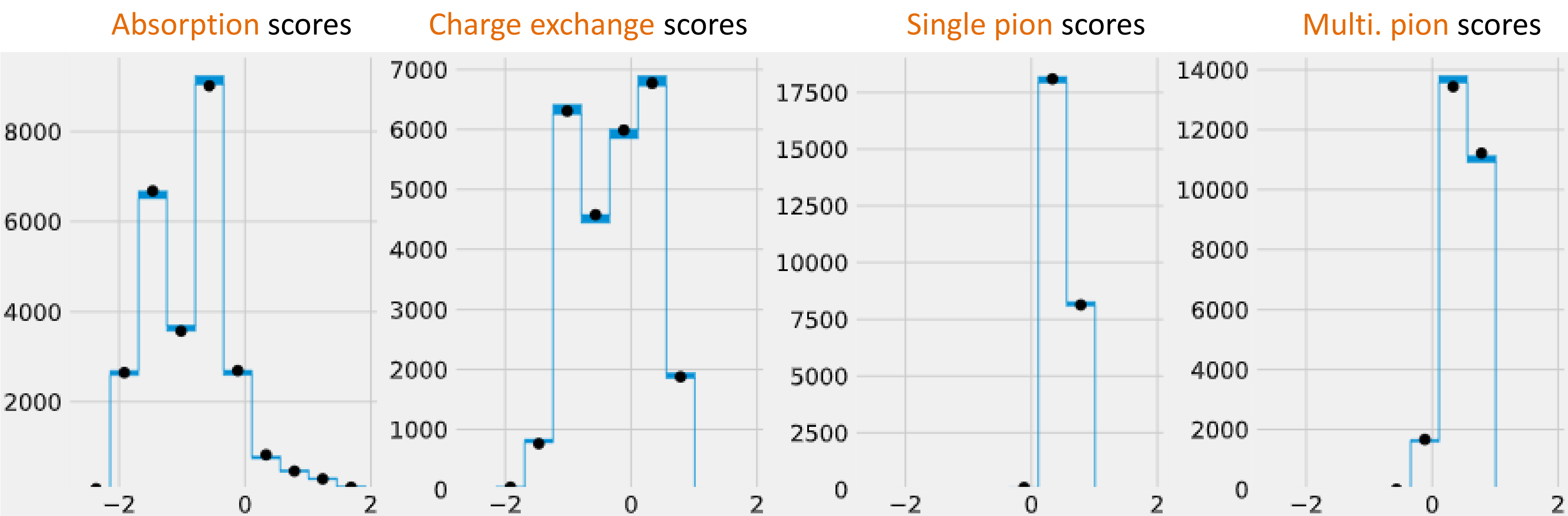
Fit – Minuit template fit

- The fit uses (python) Minuit's template fit, using Dembinski and Abdelmotteleb method.
- Data, D_{c_i}: count in bin i of channel c
  - $c \in$ GNN score channels (abs., CEx., single pion, multi. pion)
  - $i \in$ bin indices
- Template, $t_{c_i,r}$: count in bin c_i given true event region r
  - $r \in$ Regions (abs., CEx., single pion, multi. pion)
- A full template, $T_{c_i}(\mu_r)$ is constructed as:
  - $T_{c_i}(\mu_r) = \sum_r \mu_r \frac{t_{c_i,r}}{\sum_{c_i} t_{c_i,r}}$
  - With μ_r is parameter to be fit, the actual count of each region.
- The fit maximises the binned likelihood $\mathcal{L}\left(D_{c_i} \middle| T_{c_i}(\mu_r)\right)$, accounting for the template nature of the fit to find μ_r.

Fit – Minuit template fit

- Use Minuit to fit the remaining MC from the templates

	Absorption	Charge exchange	Single pion production	Multiple pion production
True count	1652	1244	6120	17305
Fit count	1602.14	1166.81	6325.83	17225.99
Fit error	73.10	170.20	369.92	449.06

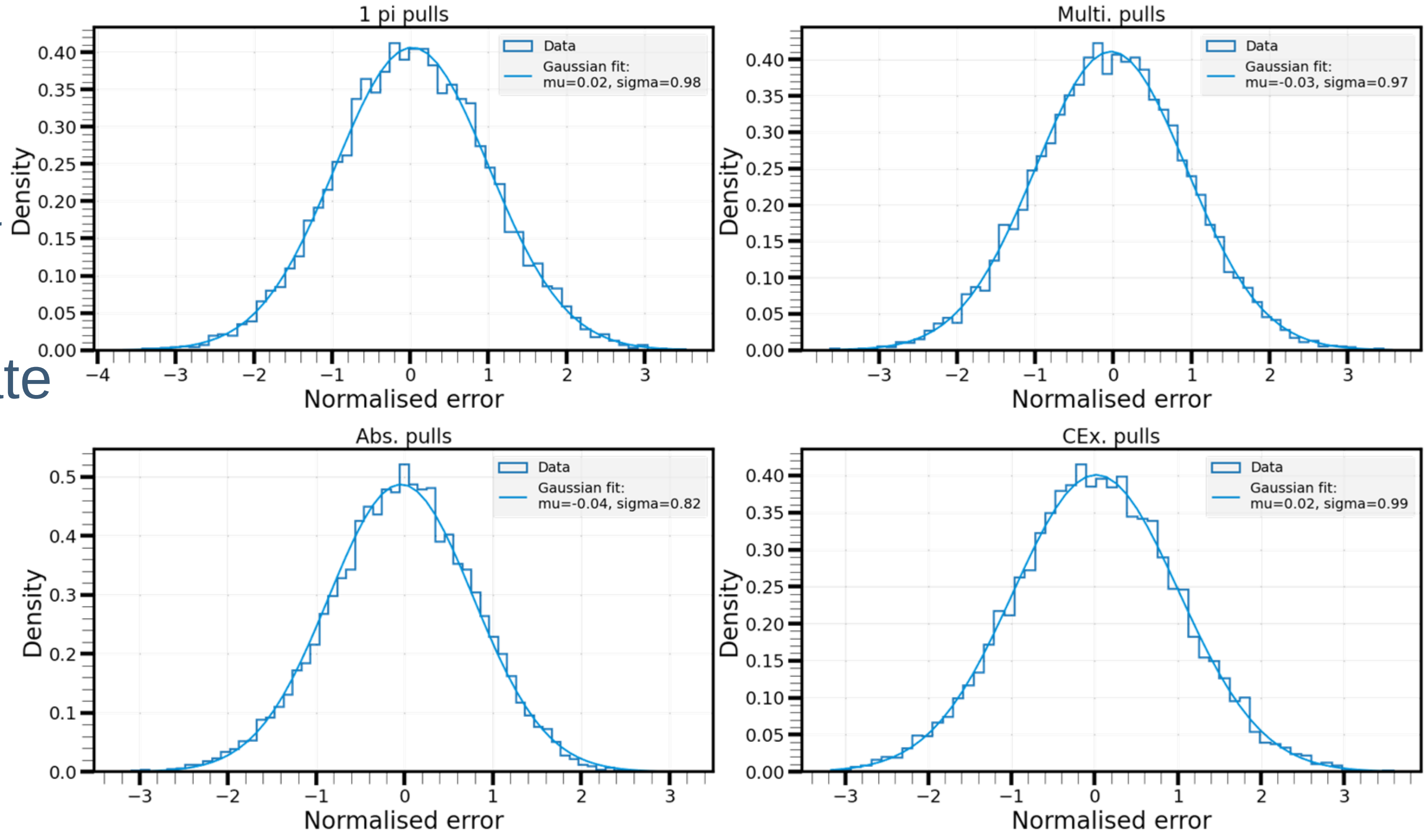


University of BRISTOL
DUNE

Checks – Pulls

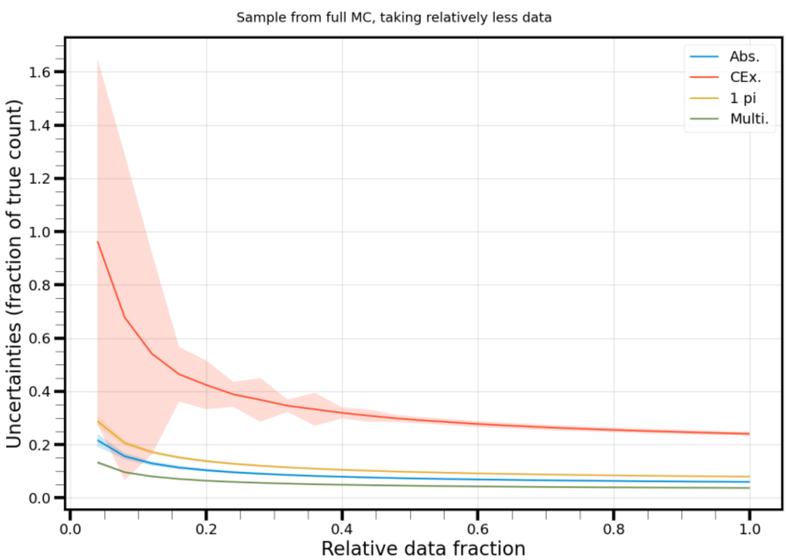
$$\text{pull} = \frac{\text{fit pred.} - \text{true count}}{\text{fit uncertainty}}$$

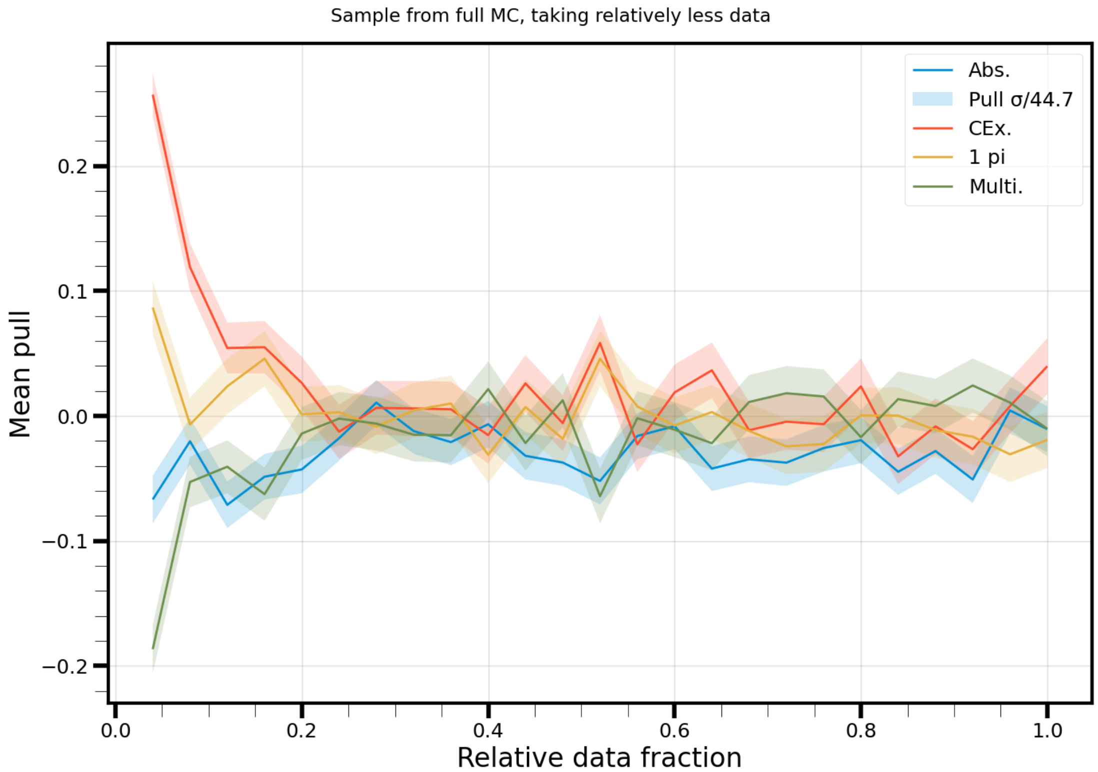
- Want to check the robustness of the fit:
  - Can it recover the actual “data” process fractions, even when the templates are mismodelled?
- The MC data is shuffled, split into data /templates
- Some template property may be edited.
- Shown: 10,000 pulls



Checks – Data statistics

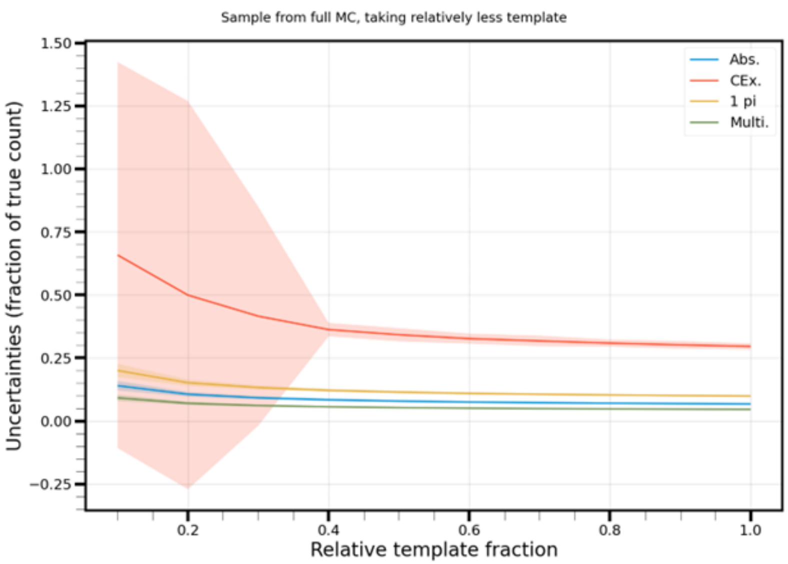
- Are the pulls consistent as the number of events in the data changes?
- “Relative data fraction” runs from 500 at 0.04 to 25,000 at 1.0
- For each test: 2,000 pulls

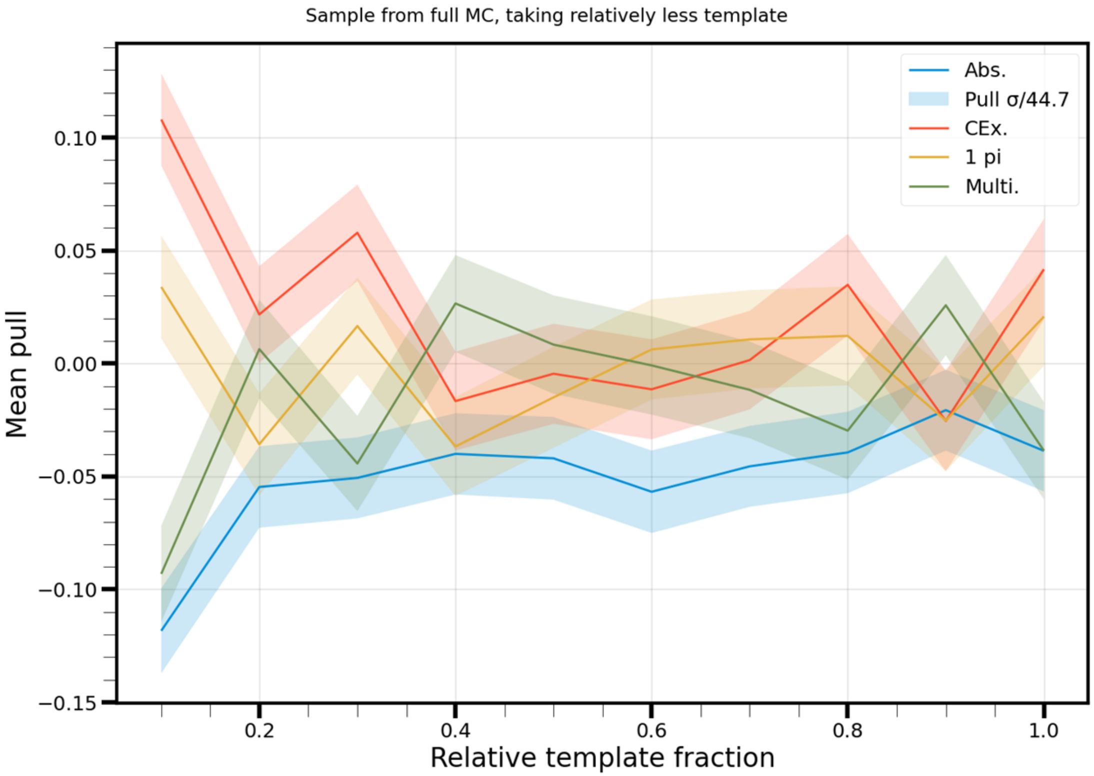




Checks – Template statistics

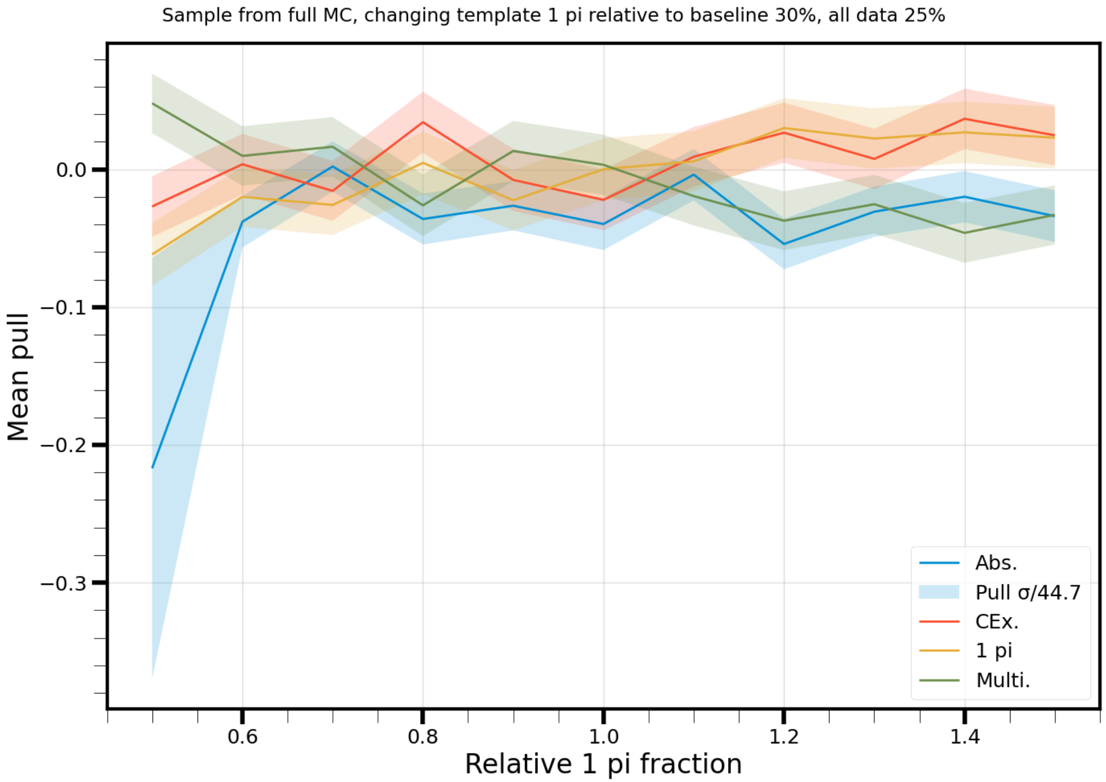
- Are the pulls consistent as the number of events in the template changes?
- "Relative data fraction" runs from 2,500 at 0.1 to 25,000 at 1.0
- For each test: 2,000 pulls



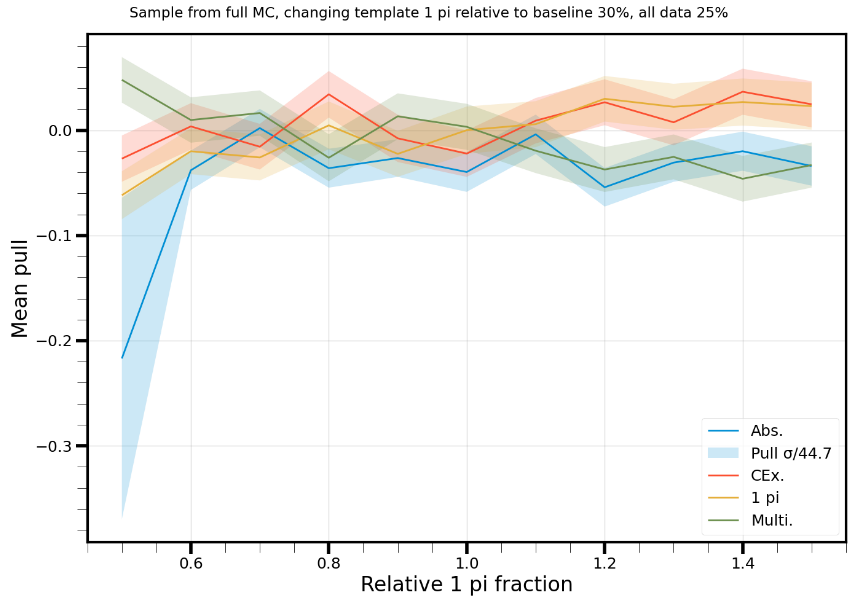


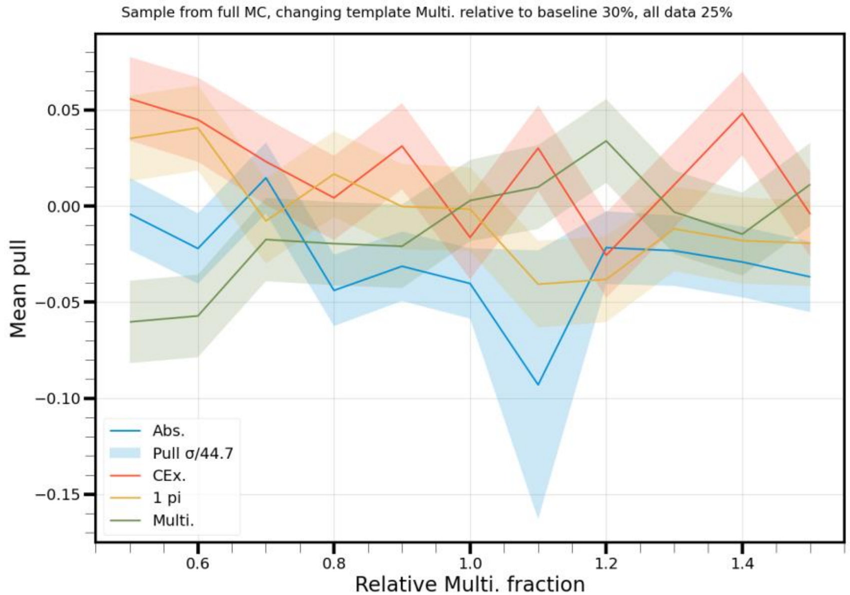
Checks – Template process fracs.

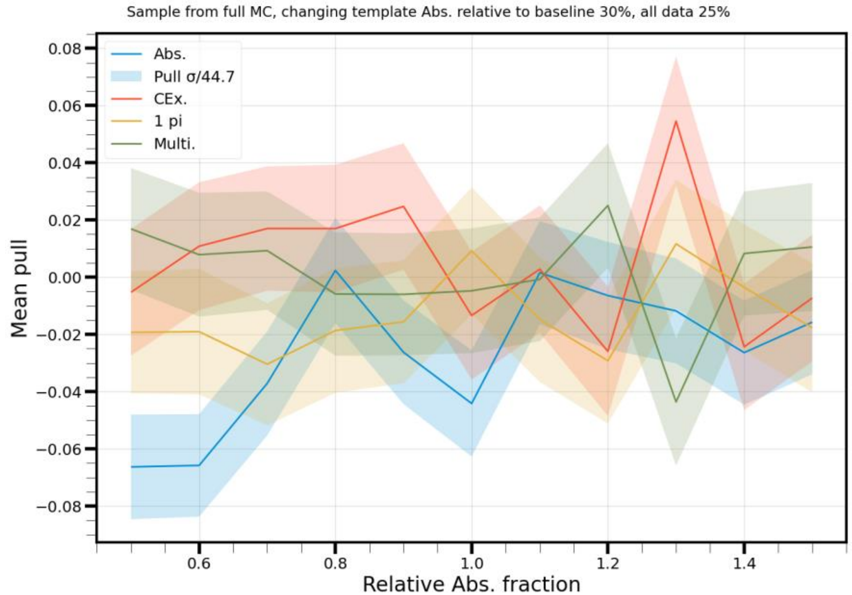
- If the fraction of processes (abs., cex. etc.) change in the template, does the fit still follow the data.
- 12,500 data, 15,000 baseline template, weighted 7,500 - 22,500
- A bad method would show a positive gradient (under-estimate when template has a deficit)

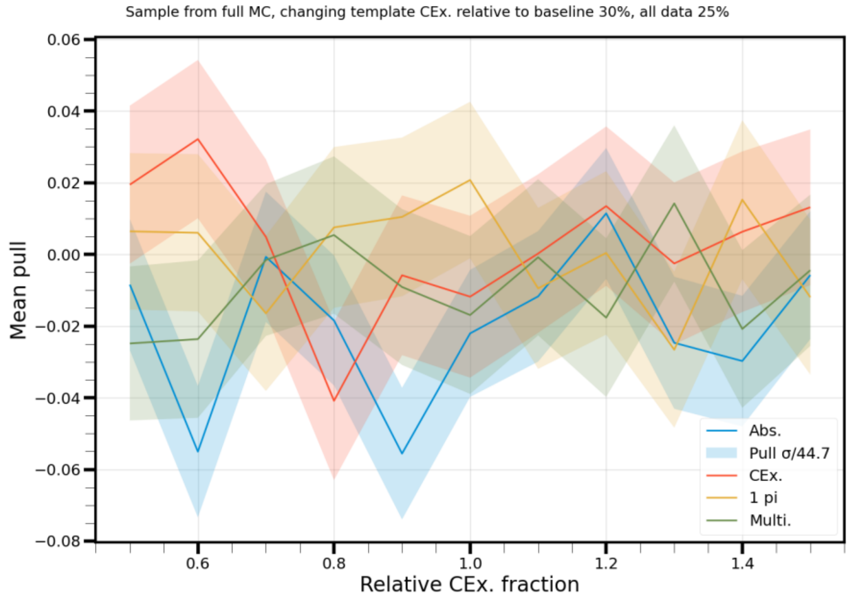


Checks – Template process fracs.







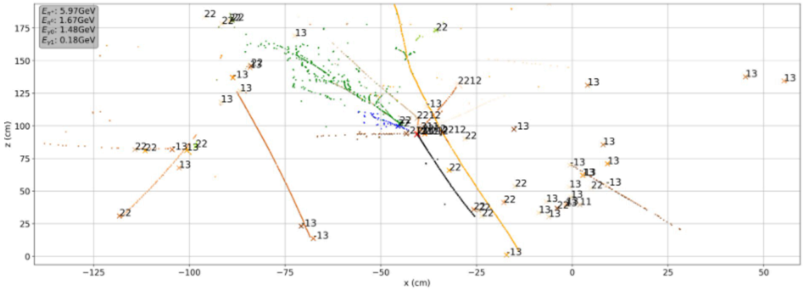


Checks – Comments

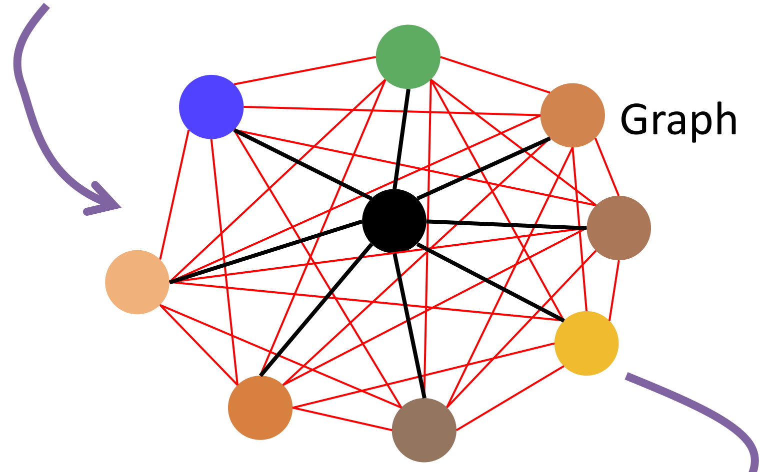
- The single/multiple pion production channels appear to have slight dependency on the template fractions.
- Charge exchange struggles with low statistics.
- Check still to do:
  - “Smearing” the templates (shuffling the template bin occupancies)
  - Reweight the template as a function of:
    - Beam particle energy (energy correlations)
    - Number of PFOs in event (reconstruction effects)
  - Make tweaks to the MC before running the GNN.

Summary

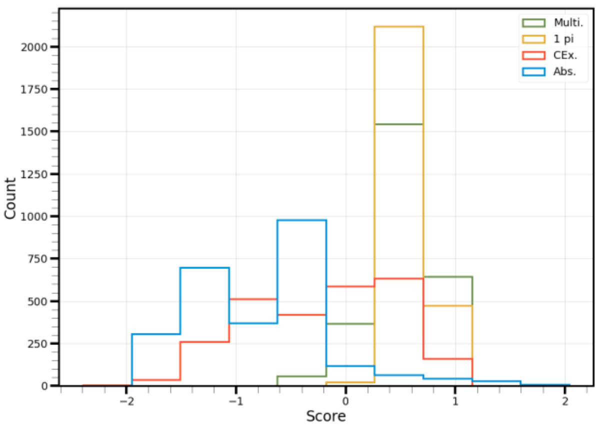
PD-I events



Graph



Distribution fits

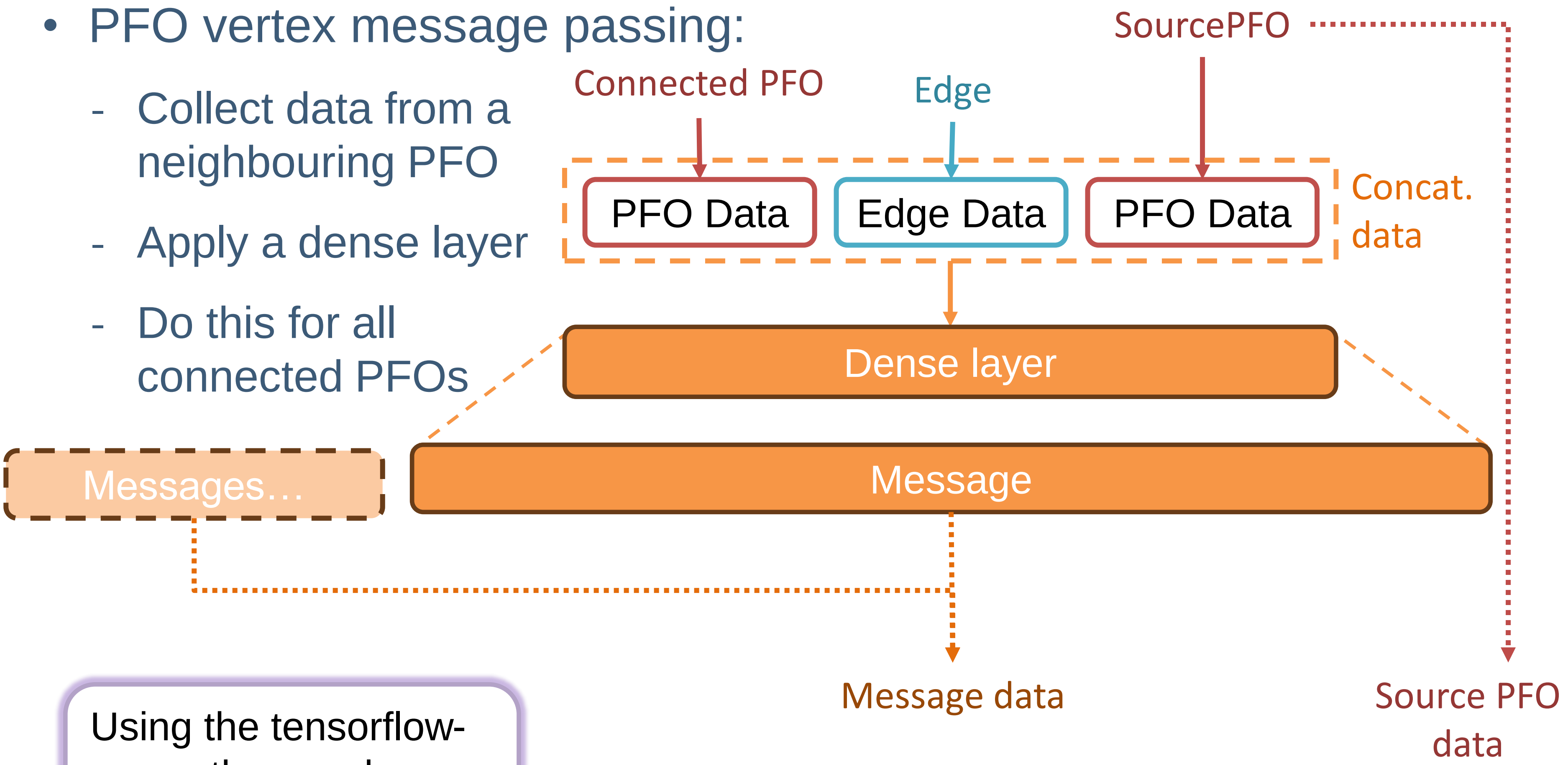


- Graph neural networks are appropriate for data with an unknown number of inputs.
- Extra losses give extra handles to understand GNN performance.
- Per event classification is mediocre.
- Fitting score distributions looks promising for improving accuracy.

Back-up

Graph Neural Networks

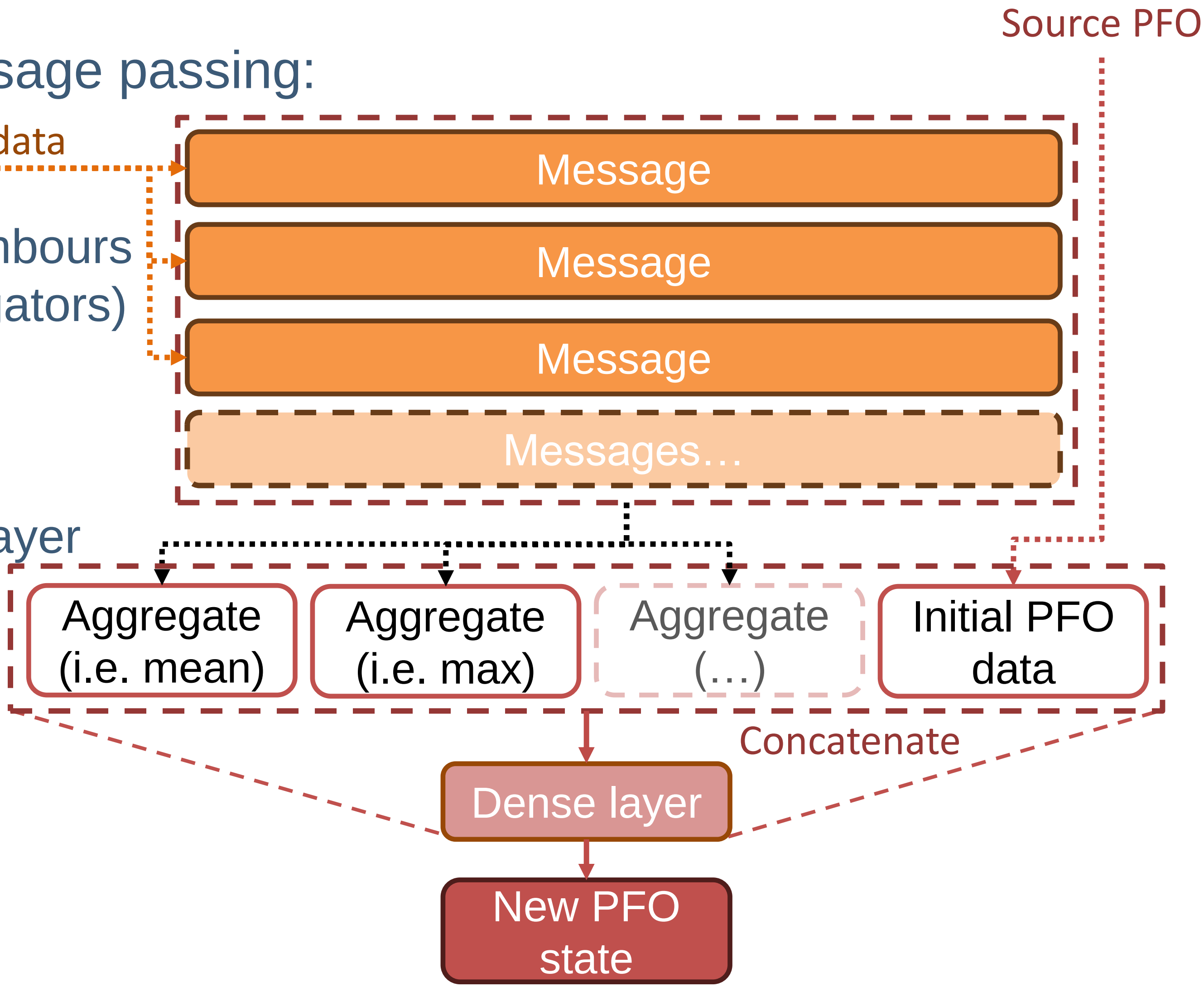
- PFO vertex message passing:
  - Collect data from a neighbouring PFO
  - Apply a dense layer
  - Do this for all connected PFOs



Using the tensorflow-gnn python package. [arXiv:2207.03522](https://arxiv.org/abs/2207.03522)

Graph Neural Networks

- PFO vertex message passing:
  - …
  - Aggregate neighbours (multiple aggregators)
  - Concatenate aggregations
  - Apply a dense layer
  - Update state

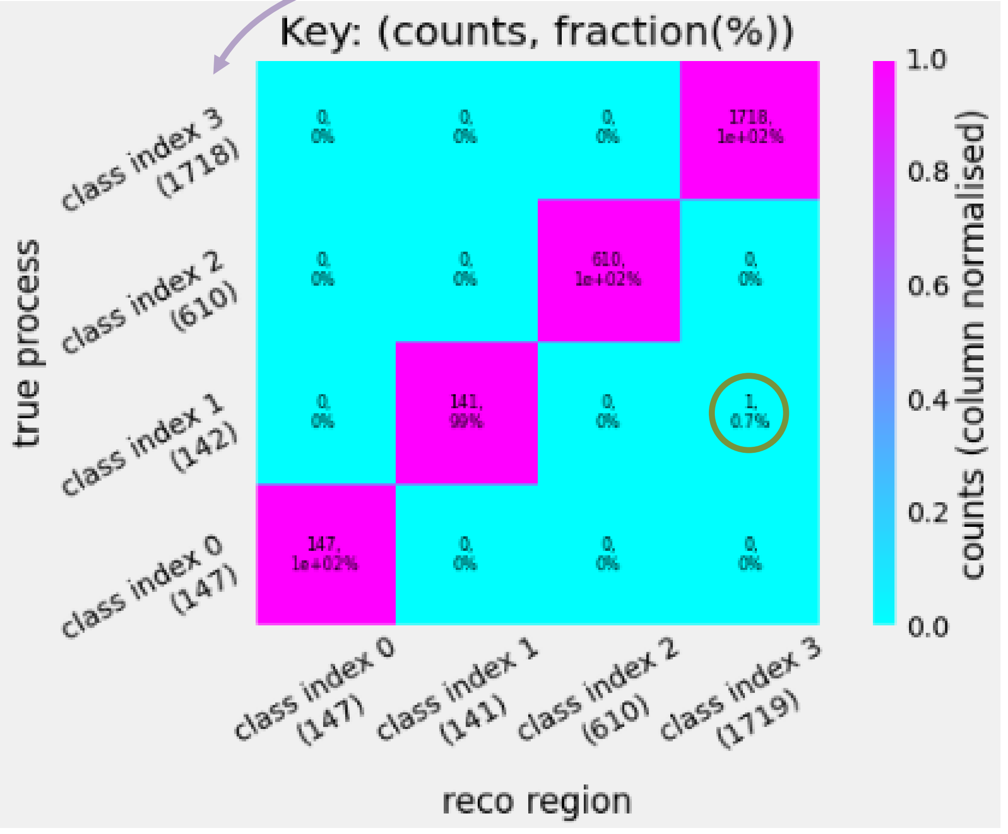


Principle Neighbour Aggregation: arXiv:2004.05718

Monte-Carlo performance

- Perfect (MC) information can achieve perfect classification (1 misclassification here).
- With MC data, this can event be achieved without any message passing steps.
  - Look once at the PFOs surrounding the beam.
  - Don't have to infer any information.

3: Multi-pion production
2: Single pion production
1: Charge exchange
0: Absorption

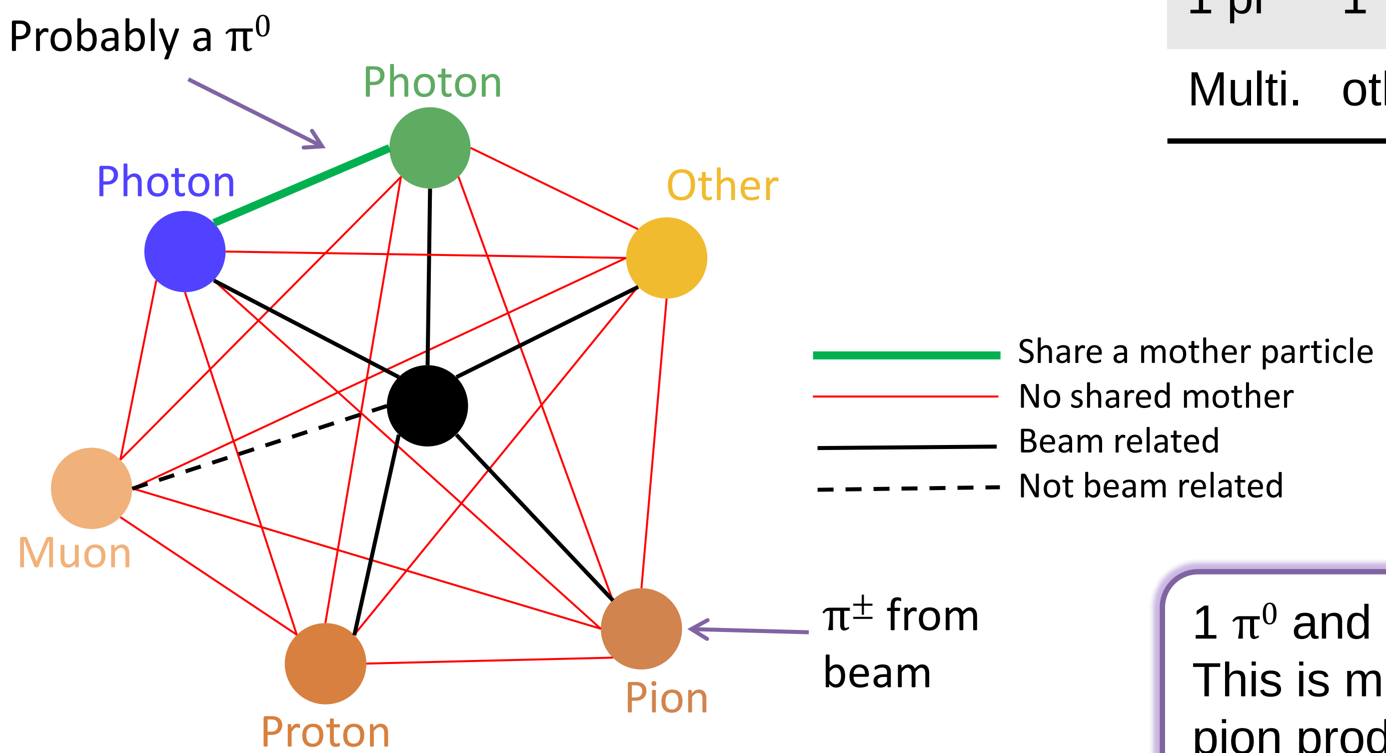


University of BRISTOL

Example graph

- Example Monte-Carlo event graph.
- Can we determine the classification?

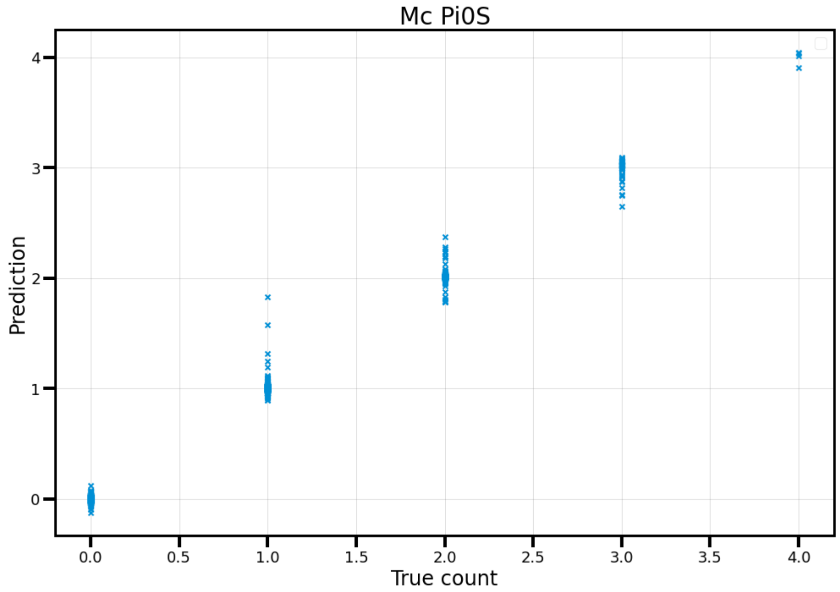
	$\pi^{\pm}$s	π^{0}s
Abs.	0	0
CEx.	0	1
1 pi	1	0
Multi.	otherwise	

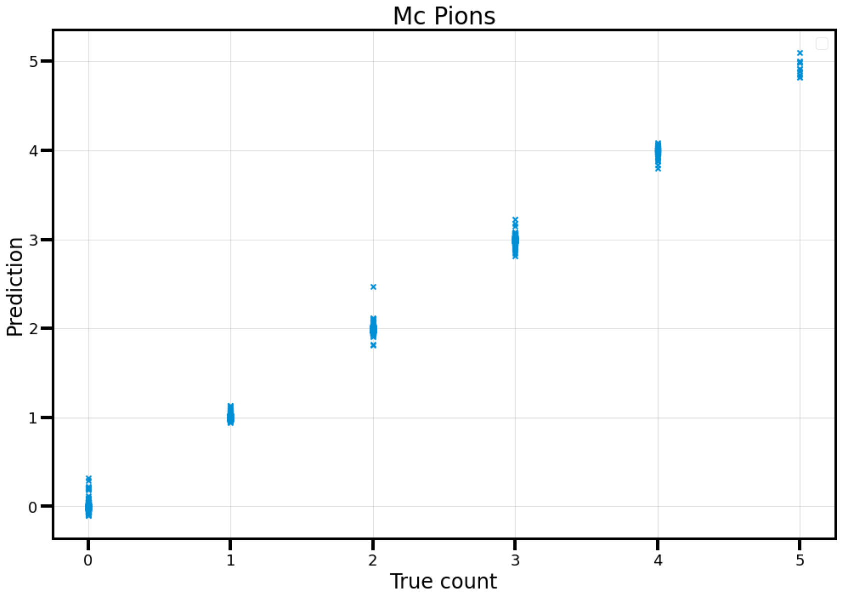


1 π^0 and 1 $\pi^{\pm}$:
This is multiple pion production

Interpretability

- Add an output to the network which aims to count the $\pi^{\pm}/\pi^{0}$ s.
- These are trained with a regression loss (mean absolute error) against the true $\pi^{\pm}/\pi^{0}$ count in the event.
- Can this help us understand the network's weaknesses



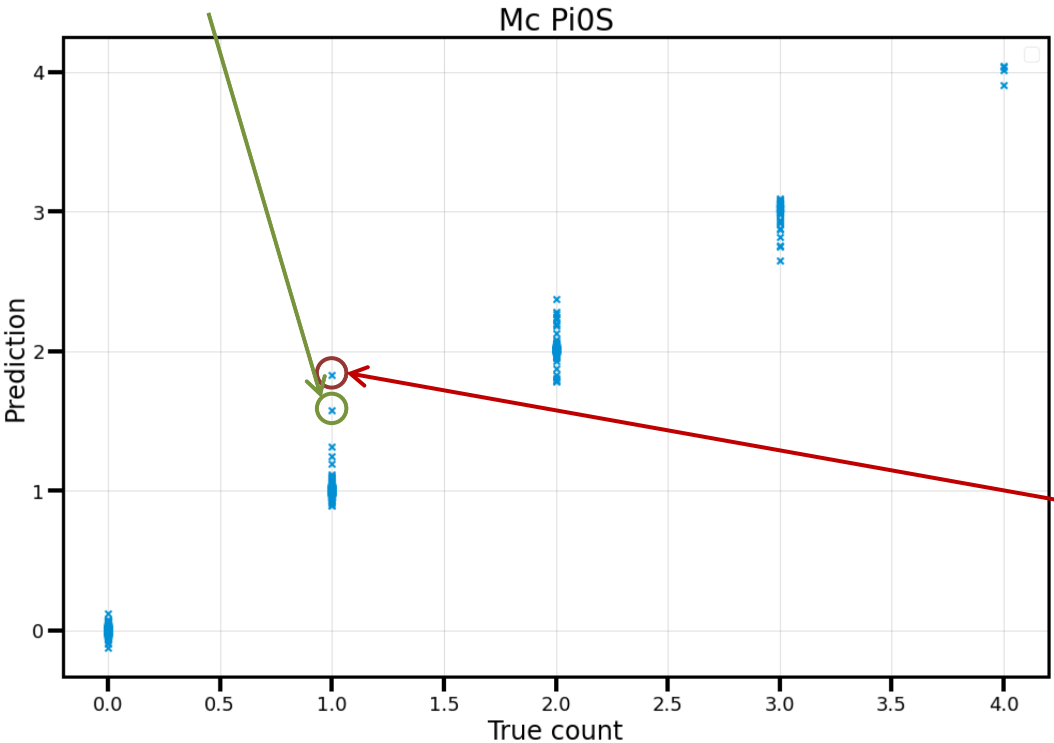


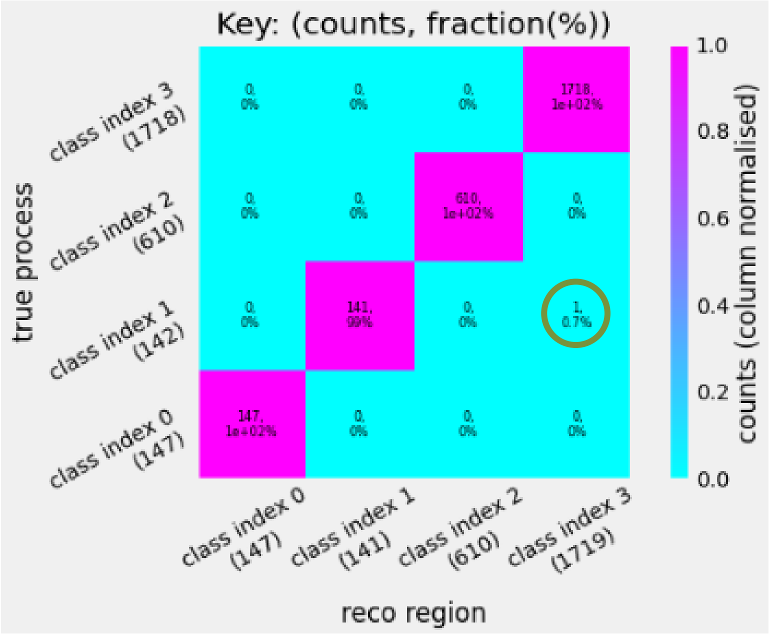
Interpretability

3: Multi-pion production
2: Single pion production
1: Charge exchange
0: Absorption

- Recall the trained model had a misclassified event.
- Which event is this in the particle counts?

This is the misclassified event!





This event was already multiple pion production *(the number of pions is already >2, so an additional pion doesn't change region)*

University of BRISTOL

Network structure

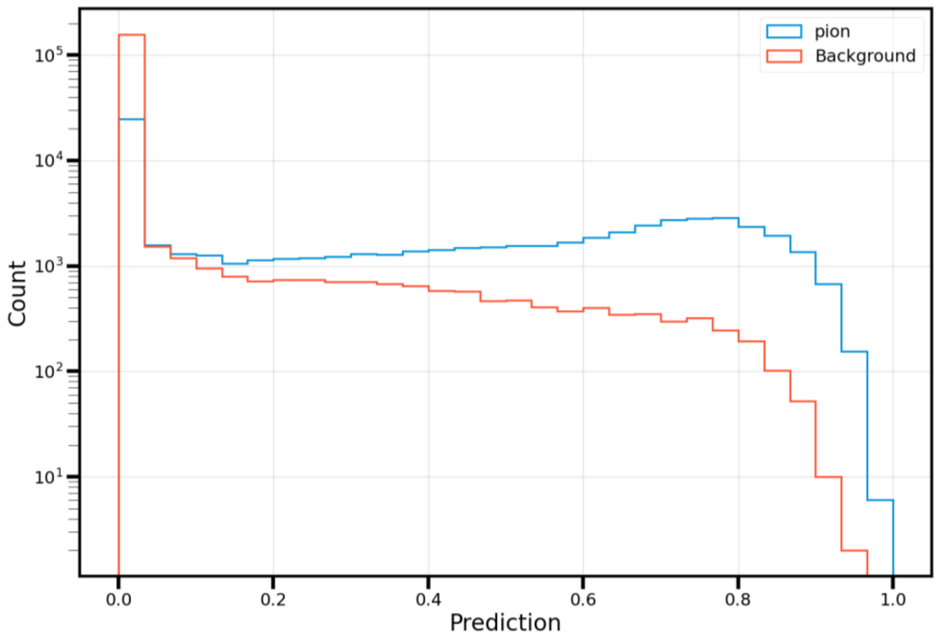
- Message passing step:
  1. PFO update (update PFO states from neighbour PFOs)
  2. Neighbour update (update edges based on connected PFOs)
  3. Beam collection (update the beam vertex with PFO information)

1. Set initial state (apply a dense layer to input features)
2. Beam collection
3. Message passing (x2)
4. Readout beam state (get the data in the beam vertex)
5. Classifier layers (dense layers to make a 4-component output)
   - Loss: focal categorical crossentropy

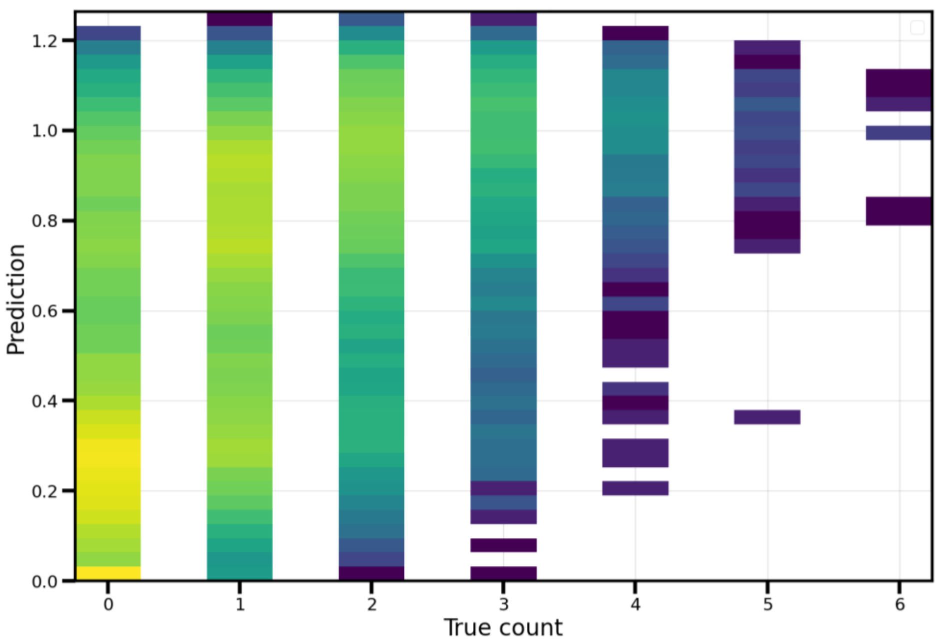
Reconstructed performance

- Primary failure modes seem to be:
  - Bad $\pi^{\pm}$ classification
  - Distinguishing 1 and >1 particles present

Pion identification on graph nodes

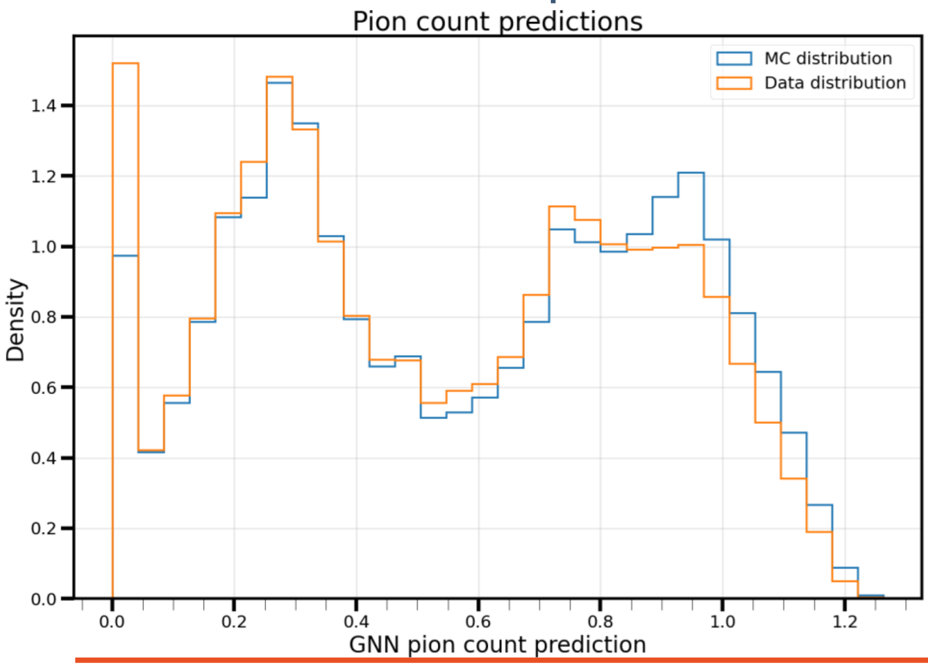


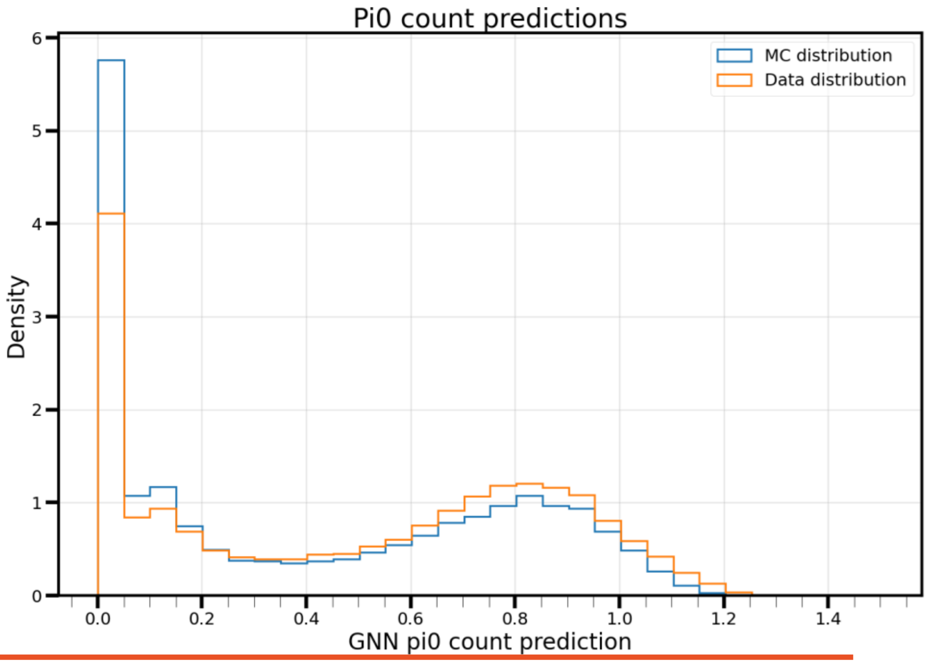
Reconstructed pion count predicted per event



Checks - data

- The predictions of $\pi^{\pm}$ / π^{0} counts are separate outputs from the same GNN which performs classification.
- Can investigate similarity by comparing these outputs in data and MC – will the templates be valid?
  - Want to find a quantitive test…

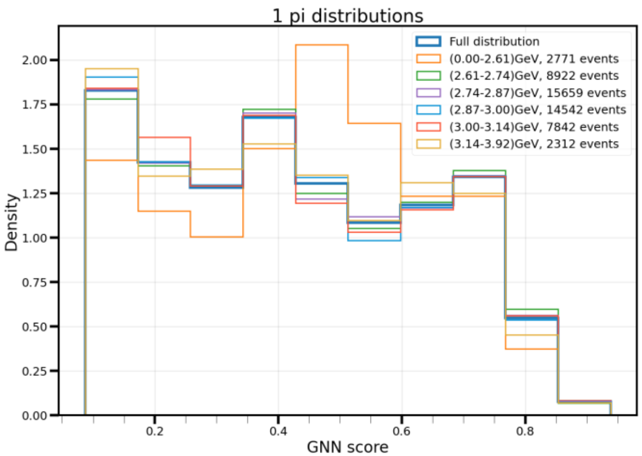


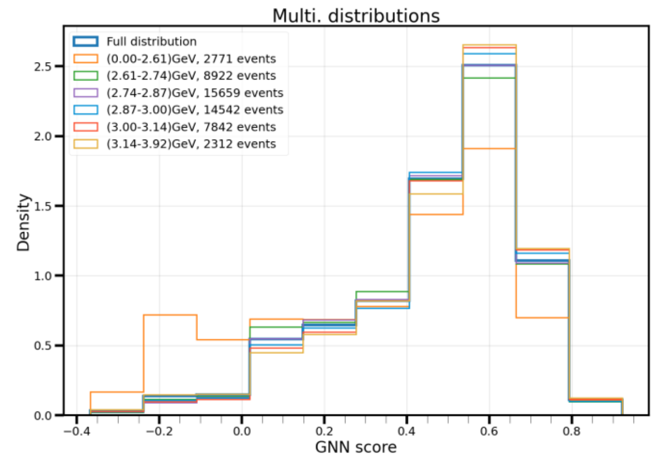


Checks - energy dependance

- The templates should not depend on the energy of the interaction.

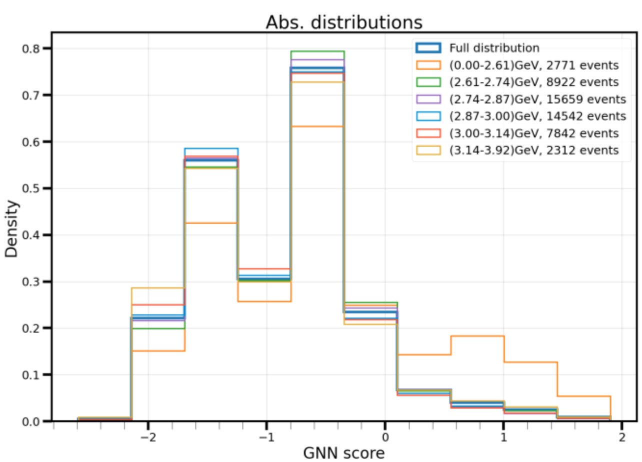
All MC
single pion
scores

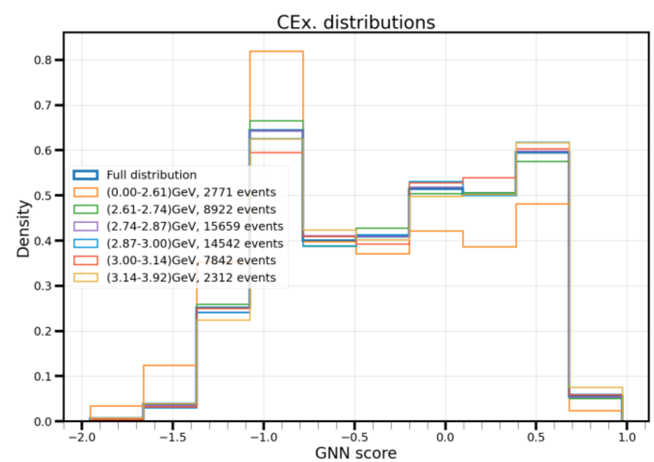




All MC
multi. pion
scores

All MC
absorption
scores





All MC
charge
exchange
scores