

Post talk comments

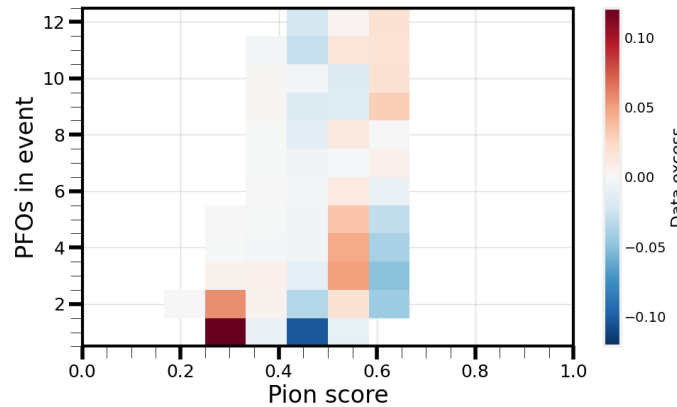
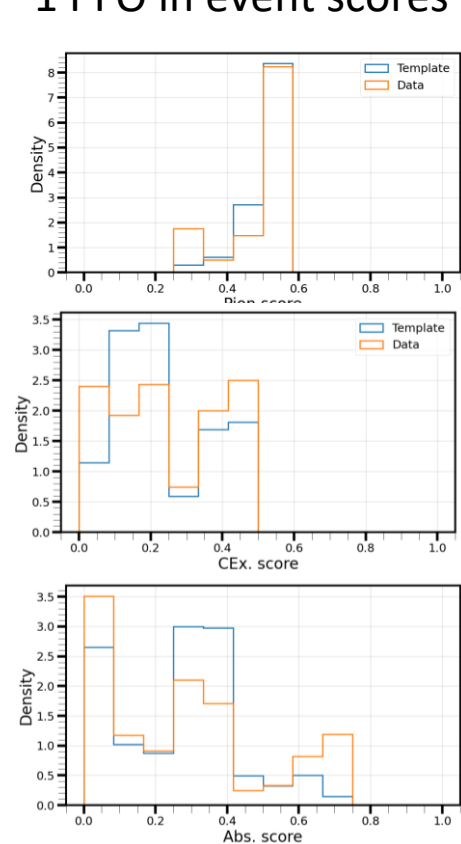
# Comments

- Generally seemed positive.
- Main comments from Jake:
  - Consider dividing templates into energies too.
  - Some magic about the energy slice which might skip unfolding.
  - Potentially some confusion about MC/data discrepancies, still communicating.
- Started looking through tech note, still trying to understand the fit minimisation
- Planning to chat with Jake soon

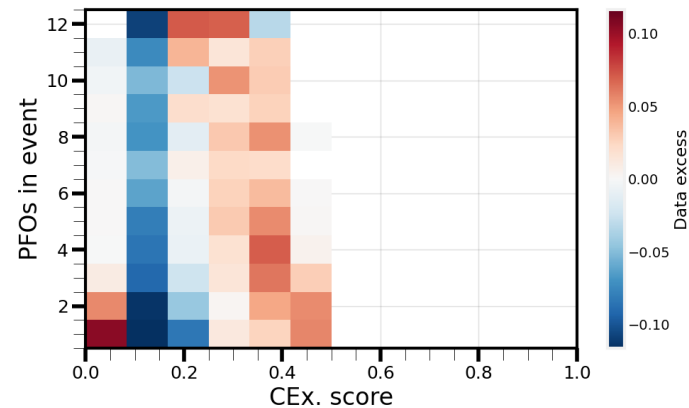
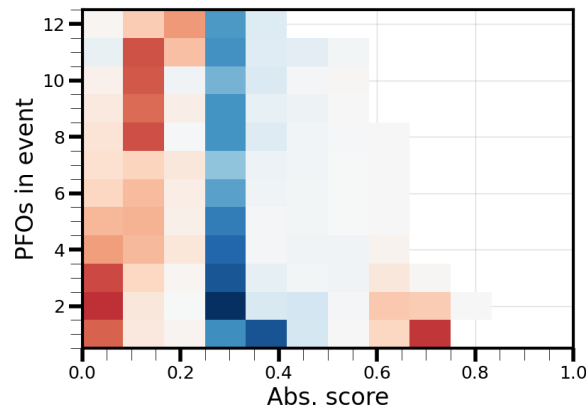
# PFO count variation - comparison

- Plots compare all MC events (not split by true process) vs. data events.

## 1 PFO in event scores

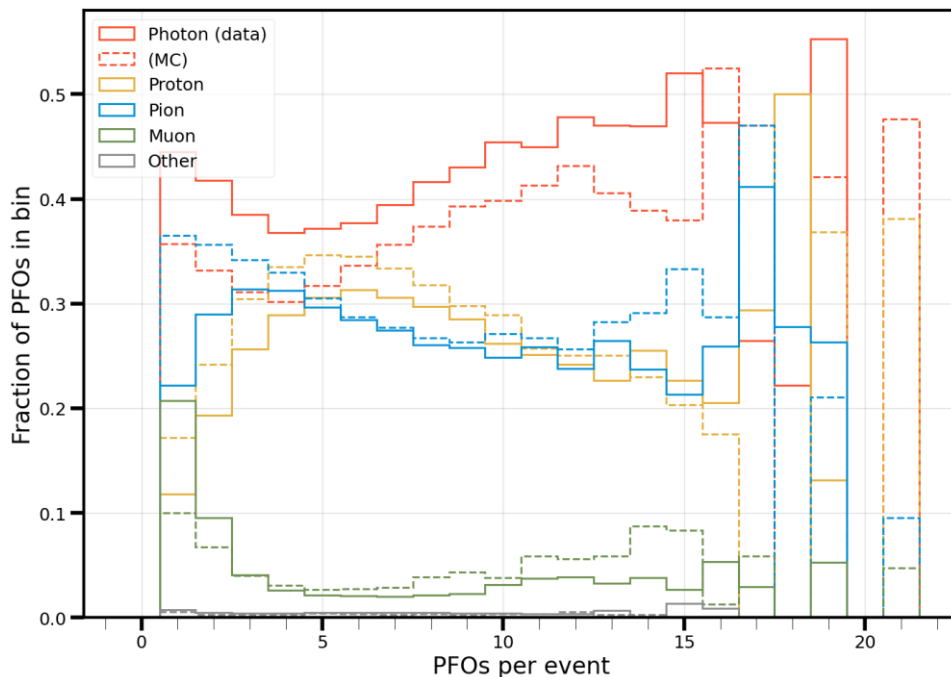


2D histograms: **excess in data** as a function of GNN score over range between 1-12 PFOs per event (13+ PFO excluded)

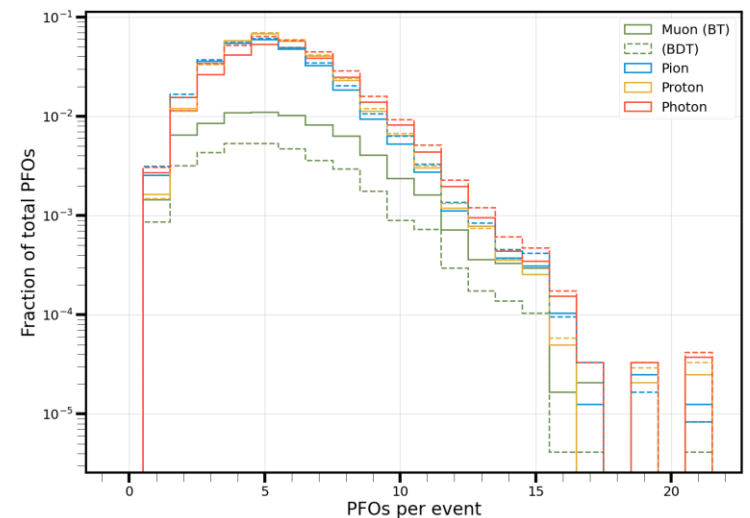


# Particle content

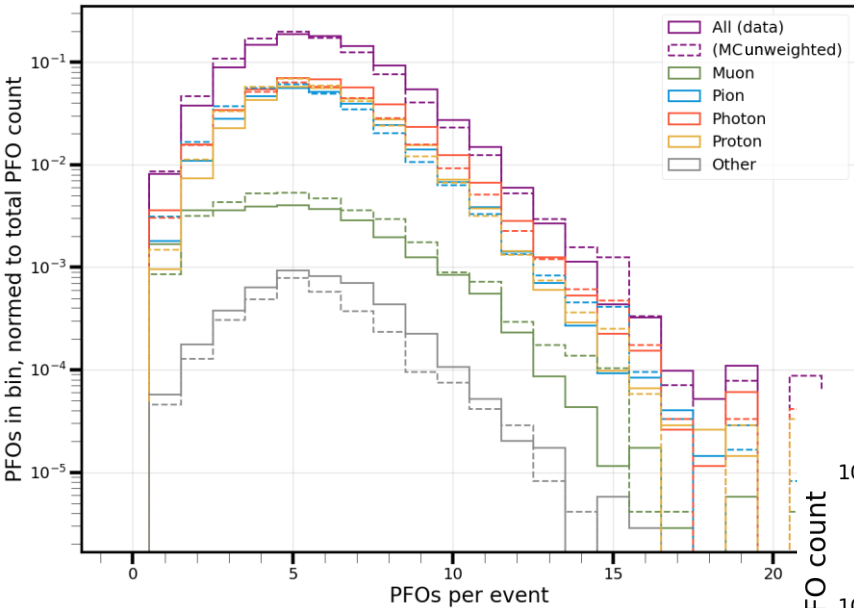
- MC vs. data discrepancy could be caused by mismodelling of the species expected from nuclear events.
- Use a simple BDT (same BDT used for PID in the full network) to estimate proportions of particles in MC vs. data.



BDT classification count (solid) vs. back-tracked classification count (dashed)



# Reweighting events

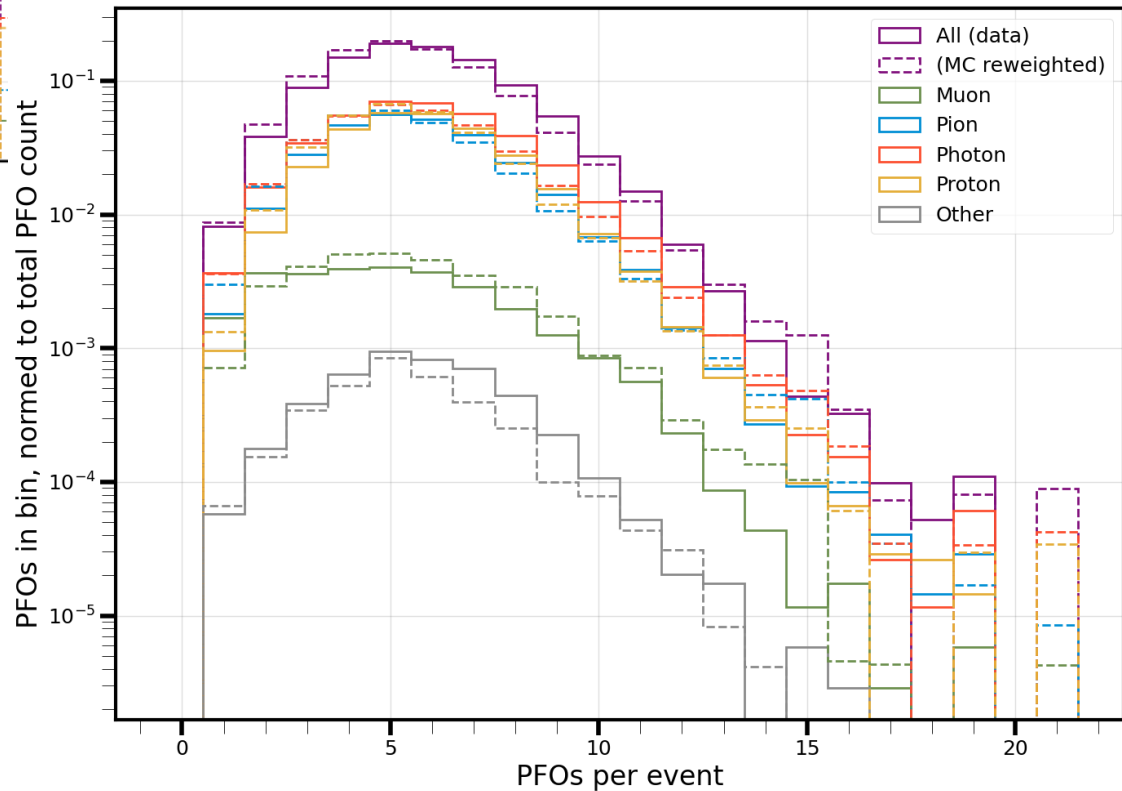


$$w_p = \frac{N_p^{MC}}{N_{tot}^{MC}} \times \frac{N_{tot}^{data}}{N_p^{data}}$$
$$w_{evt} = \sum_p n_p w_p / n_{tot}$$

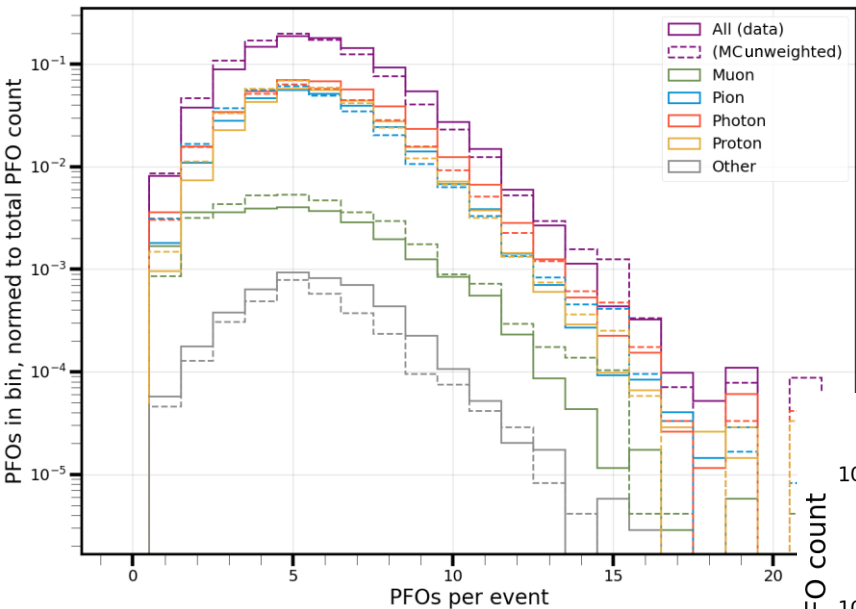
$N$ : all events

$n$ : particular event

$p$ : particle species



# Reweighting events

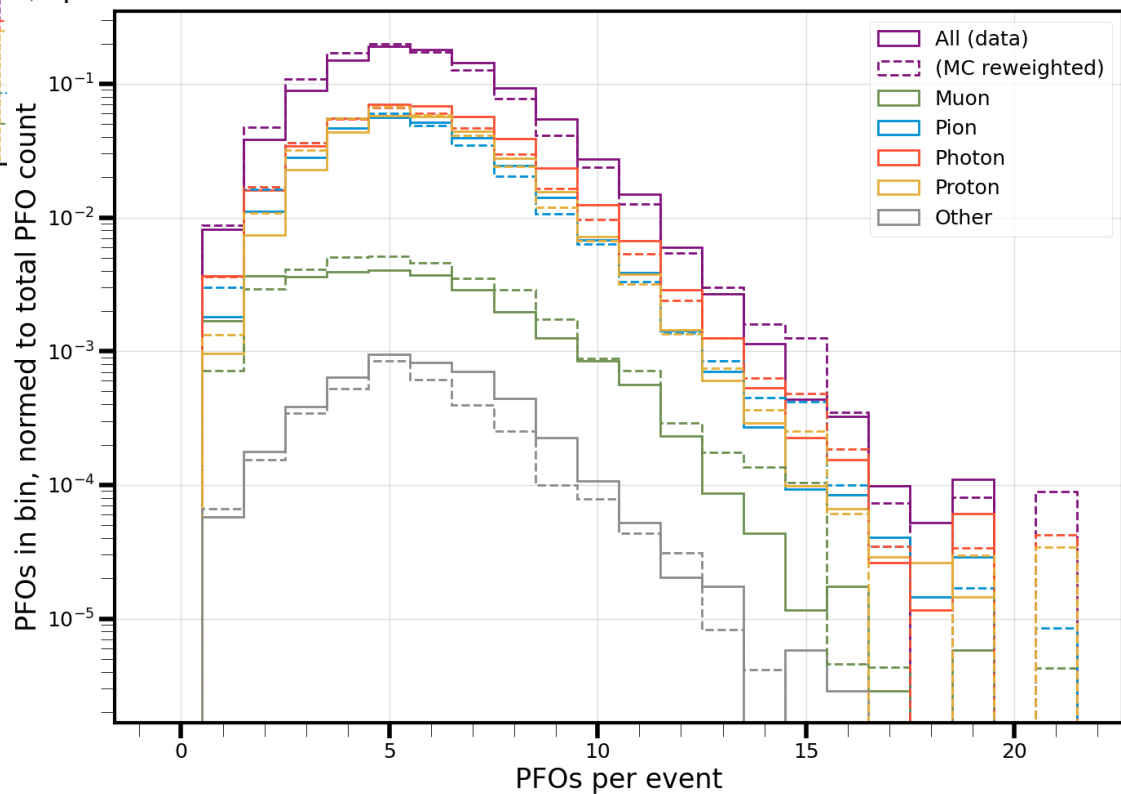


$$w_p = \frac{N_p^{MC}}{N_{tot}^{MC}} \times \frac{N_{tot}^{data}}{N_p^{data}}$$
$$w_{evt} = \sum_p n_p w_p / n_{tot}$$

$N$ : all events

$n$ : particular event

$p$ : particle species



# After weighting

If the re-weighting accounts to the MC/data discrepancy, the MC/reweighted difference should match the MC/data difference.

