Validation System Upgrade status

Matteo Galli

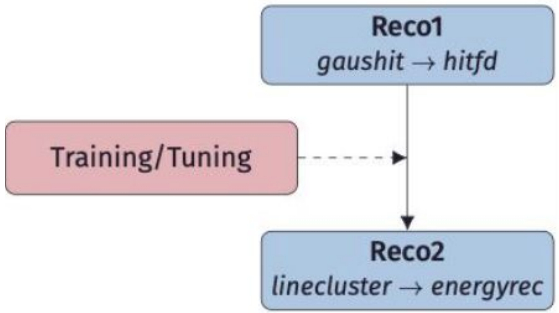
11/03/2024

Validation system

Goal: have a simple way to check the sanity of a sample with respect to a given reference, with a **graphical comparison** of distributions of selected variables.

- Want to be able to check all **stages** of a sample production:
  - **Reco1** level = **Generation** + **Detector** + **Hit** (New)
  - **Reco2** level = **Pandora** / **Higher-level reco** (Already existing)

Status: validation packages for both Reco1 and Reco2 exist, w/ slight differences in implementations (and a few things left to be done/decided)!



Reco1 validation

- Setup:
  - New analyzer **GenRecoValidator** in *duneana*
    - Generation level (eg. true vertex position, nu azimuth angle, nu energy, etc...)
    - Detector level (eg. detected photons for different flavors)
    - Hit level (eg. hit charge, hit peak time, hit width, etc...)
  - A dedicated **fcl file** to run the analyzer in larsoft (`run_GenRecoValidator.fcl`)
  - Two **macros** to get the graphical comparison of distributions (`RecoPlots.C` - `Reco1CompareDataDistribution.C`)
- Status:
  - **Analyzer**:
    - PR in ***duneana*** done and approved
      - It should be **merged** this week.
  - **fcl** + **marcos**:
    - will be added by Andy (thanks) to the **Continuous Integration** (CI) **system**
      - temporary: you can find these files in my git repository here

Reco2 validation

- Setup:
  - A dedicated **fcl file** (`run_pandora_ana.fcl`) to run the AnaTree analyzer in larsoft and keeps just the Reco2/High-level variables in the output file
  - Two **macros** to get the graphical comparison of distributions

- Status:
  - already implemented in **CI system**
  - used already extensively for validation of **LBL** production
  - still need to add a few checks of interest for atmospheric neutrino production:
    - eg. additional Energy and Angle related distributions

Summary status

- Validation of Reco1 and Reco2 implemented slightly different, but both in CI system
  - **Reco1** is ready to be added in the CI system (soon, by Andy).
  - **Reco2** needs additional distributions for **atm nu**
- The developed model, composed of analyzer and macros, can be easily developed or replicated for different reco-flows.

- Moving forward: what should we do?
  - Option 1: stick w/ this setup, to mirror the production steps (**separate analyzers** for reco1, reco2)?
    - and add potentially other individual analyzers for other reco-flows (i.e. another analyzer for PDS checks)
  - Option 2: use a **single analyzer** for all steps that need validation ?
    - i.e. add Reco2 variables to the GenRecoValidator analyzer of Reco1