DUNE-PRISM FAKE DATA

DUNE NEAR DETECTOR WORKSHOP

FERMILAB

MARCH 22ND, 2018

Stony Brook University

Cristóvão Vilela

FAKE DATA MOTIVATION

- Let's entertain the possibility that our neutrino – nucleus scattering models do not reflect nature accurately…
  - … specifically in ways that affect our understanding of $E_\nu \rightarrow E_{rec}$ …
  - … differently for ν and anti-ν?

- How degenerate are effects that can be reweighted out with effects that distort $E_\nu \rightarrow E_{rec}$ assuming an on-axis near detector?

$$E_\nu^{cal} = E_\ell + \epsilon_n + \sum_{i=1}^{n} \left(E_{p'_i} - M\right) + \sum_{j=1}^{m} E_{h'_j}$$

Sum over knock-out nucleons:
- Neutrons!
- How many?
- How is energy shared?

Sum over mesons:
- If undetected, $\sim m_m$ bias!
- How many?
- How is energy shared?

FAKE DATA MOTIVATION

- We will tune our models to Data / MC discrepancies observed at the near detector.
- In the wide on-axis beam, Data / MC differences induced by $E_\nu \rightarrow E_{rec}$ mis-modelling could be mistaken as a cross section effect.
  - And reweighted out with model tuning.
- DUNE-PRISM should resolve this degeneracy as it allows us to directly probe nature's $E_\nu \rightarrow E_{rec}$ in the range of interest for DUNE oscillations.

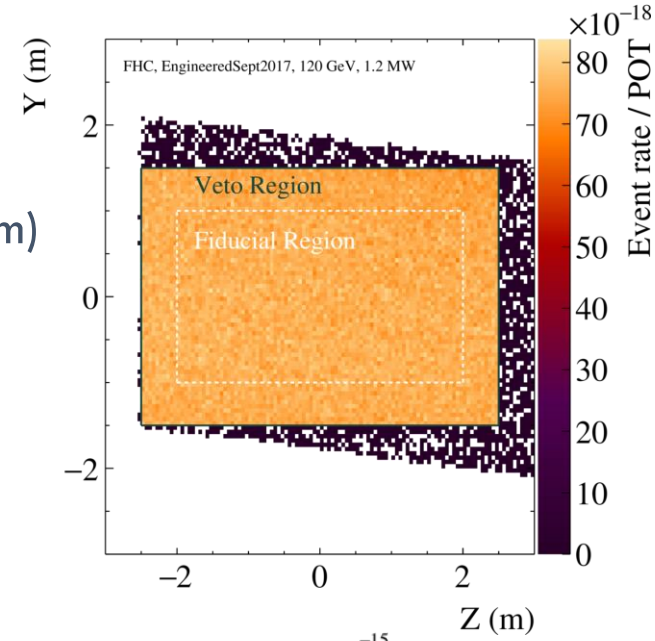
$$E_\nu^{cal} = E_\ell + \epsilon_n + \sum_{i=1}^{n} \left(E_{p'_i} - M\right) + \sum_{j=1}^{m} E_{h'_j}$$

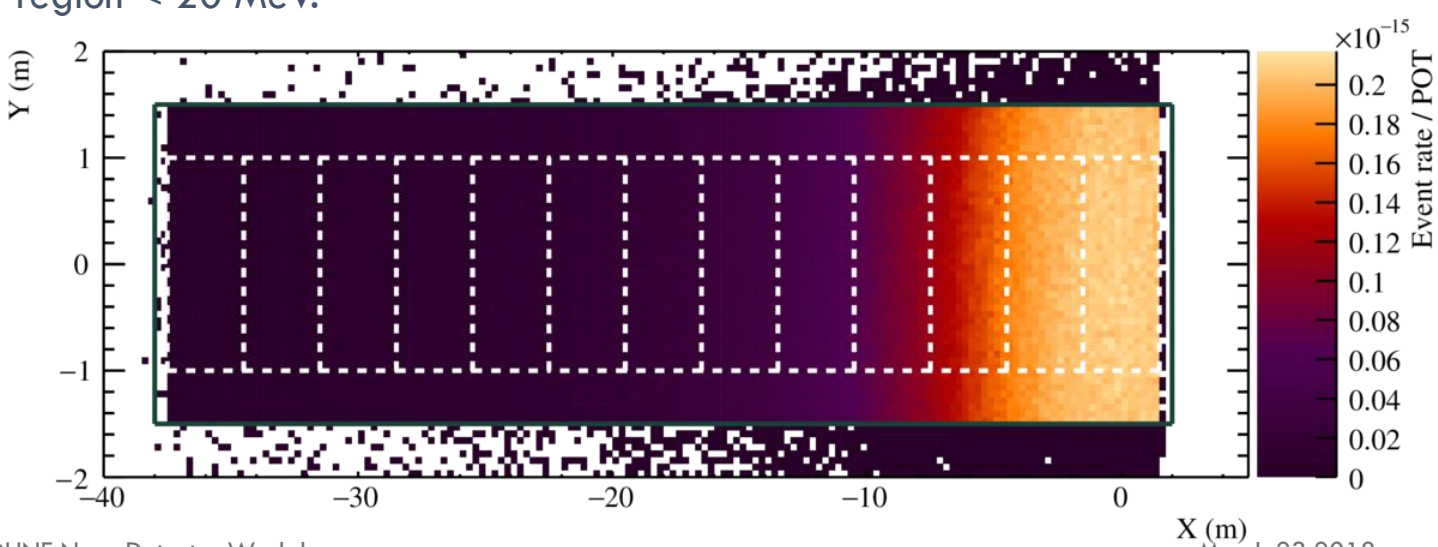
Sum over knock-out nucleons:
- Neutrons!
- How many?
- How is energy shared?

Sum over mesons:
- If undetected, $\sim m_m$ bias!
- How many?
- How is energy shared?

DUNE-PRISM SIMULATION

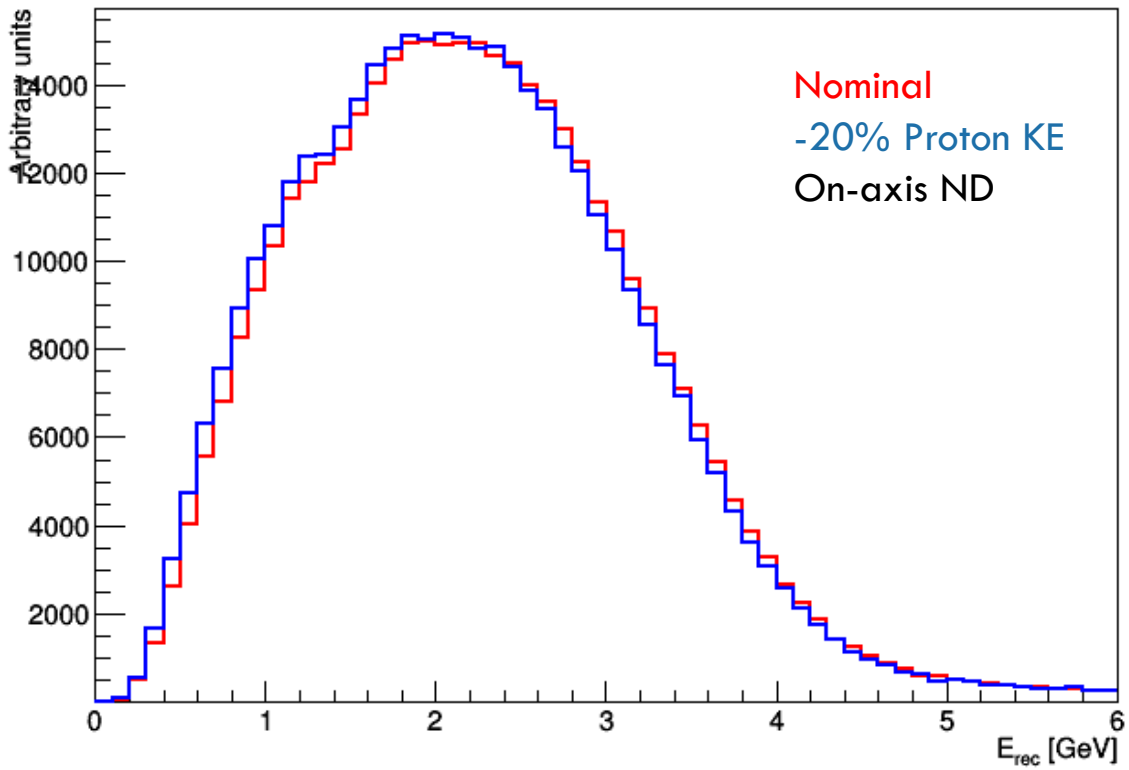
- Simulate GENIE events in a large liquid argon volume
  - 39 x 3 x 5 m.
- Divide large volume into 13 detector-sized (3 x 2 x 4 m) chunks, mimicking "stops" of a moveable detector.
- Define a veto region 50 cm from the detector edges in all directions.
  - Use this region to require hadronic system containment in active volume: non-primary-lepton energy deposits in veto region < 20 MeV.





20% MISSING PROTON ENERGY

- For each event, scale proton energy deposits in the detector by 80%.
- Affects Erec and containment cut at the near detector



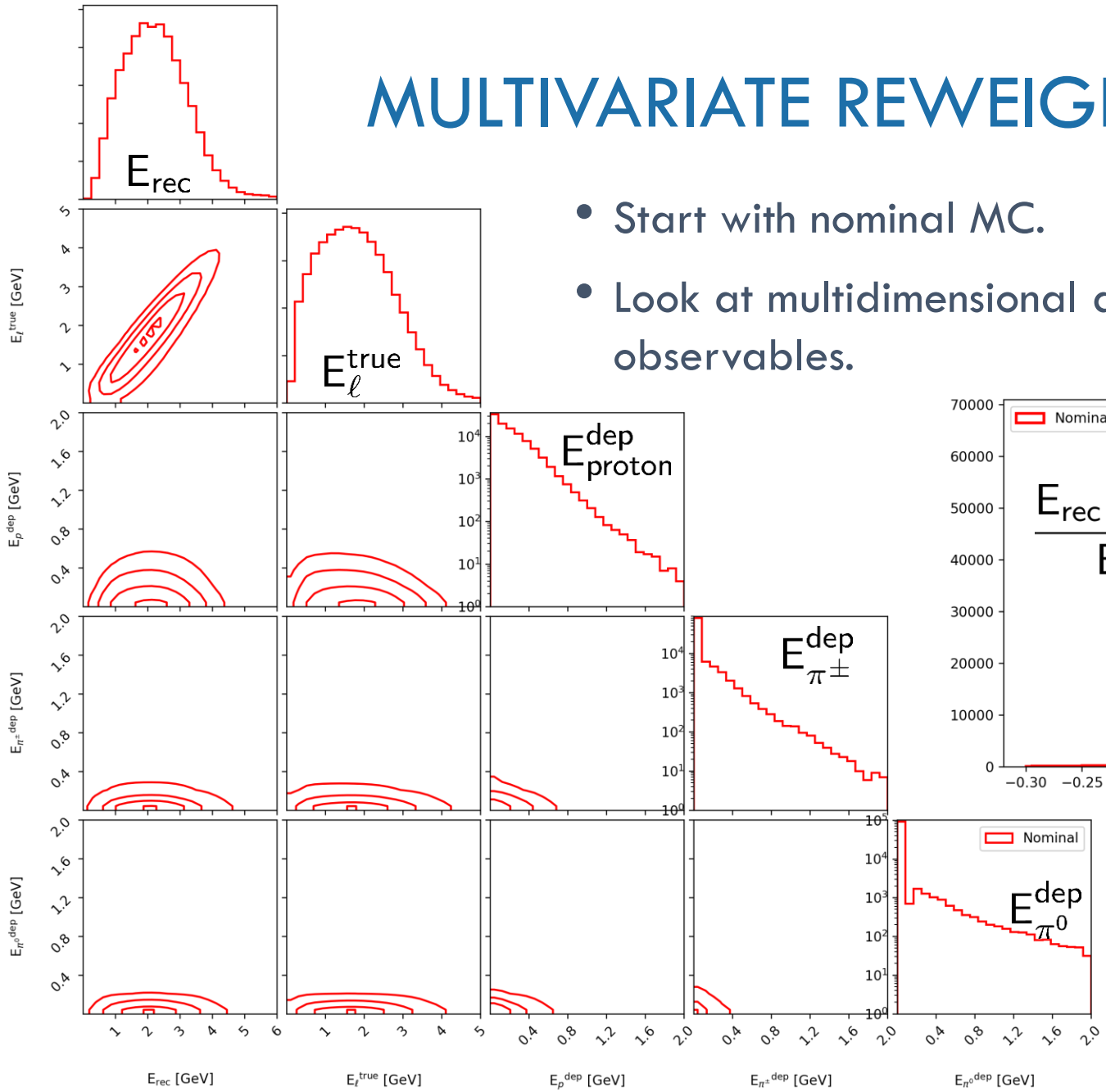
MULTIVARIATE REWEIGHTING

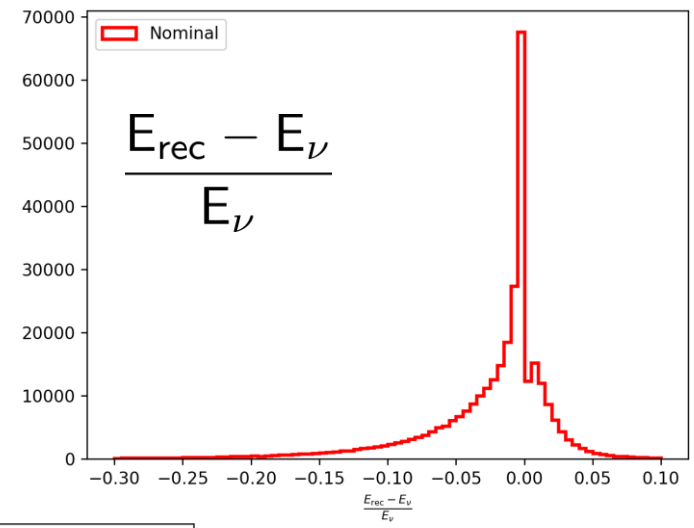
- Use multivariate method* to reweight distributions of observables back to nominal.
  - Train BDT to learn differences between shifted and nominal MC, and produce event weights from output.
  - Observables considered:
    - E_{rec}
    - Primary lepton energy
    - Proton deposited energy
    - Charged pion deposited energy
    - Neutral pion deposited energy

*A. Rogozhnikov arXiv:1608.05806

MULTIVARIATE REWEIGHTING

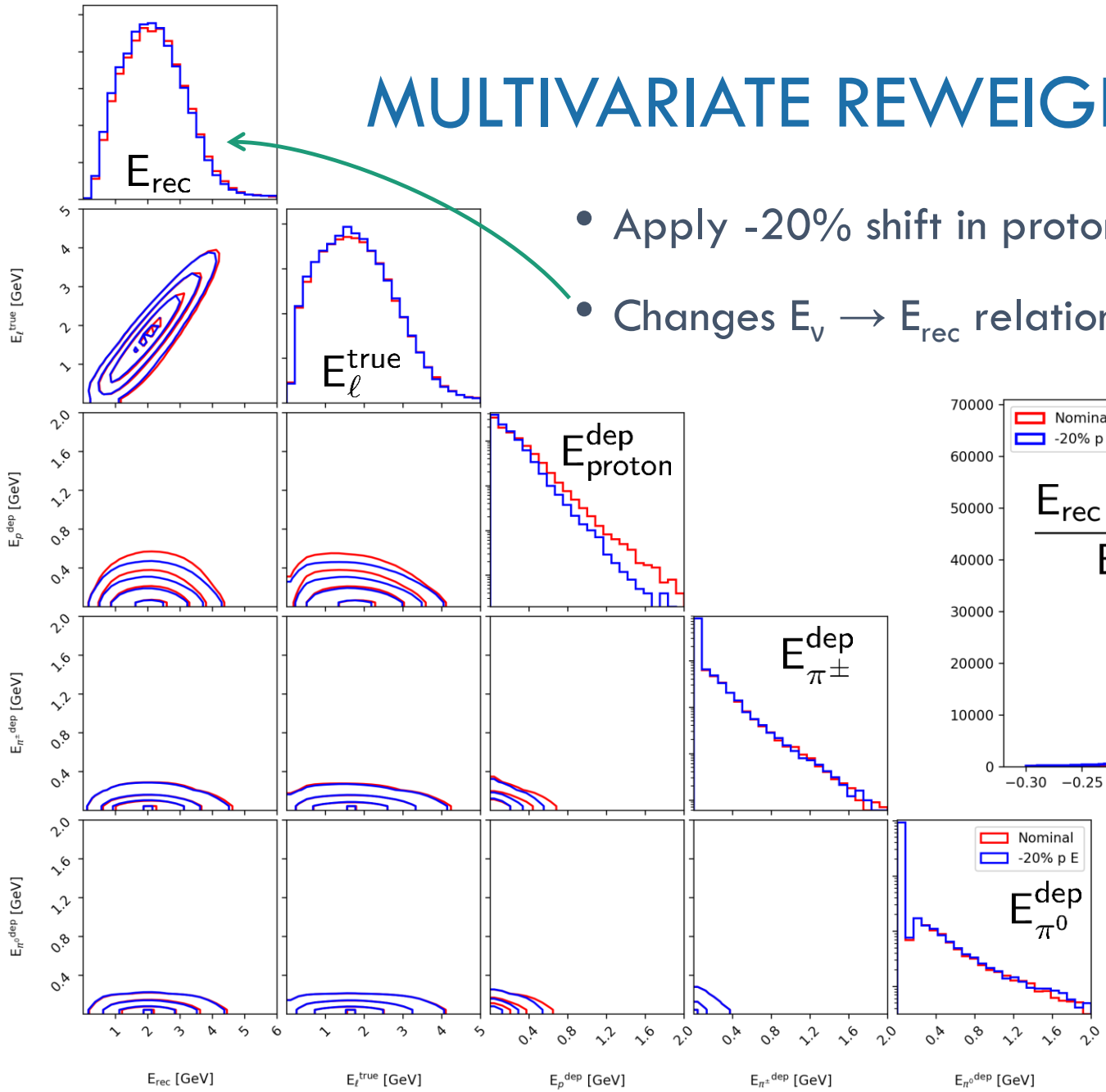
- Start with nominal MC.
- Look at multidimensional distribution of observables.

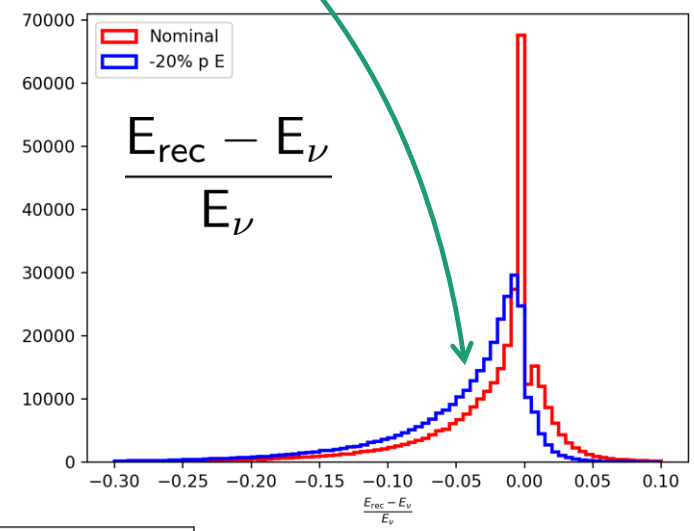




MULTIVARIATE REWEIGHTING

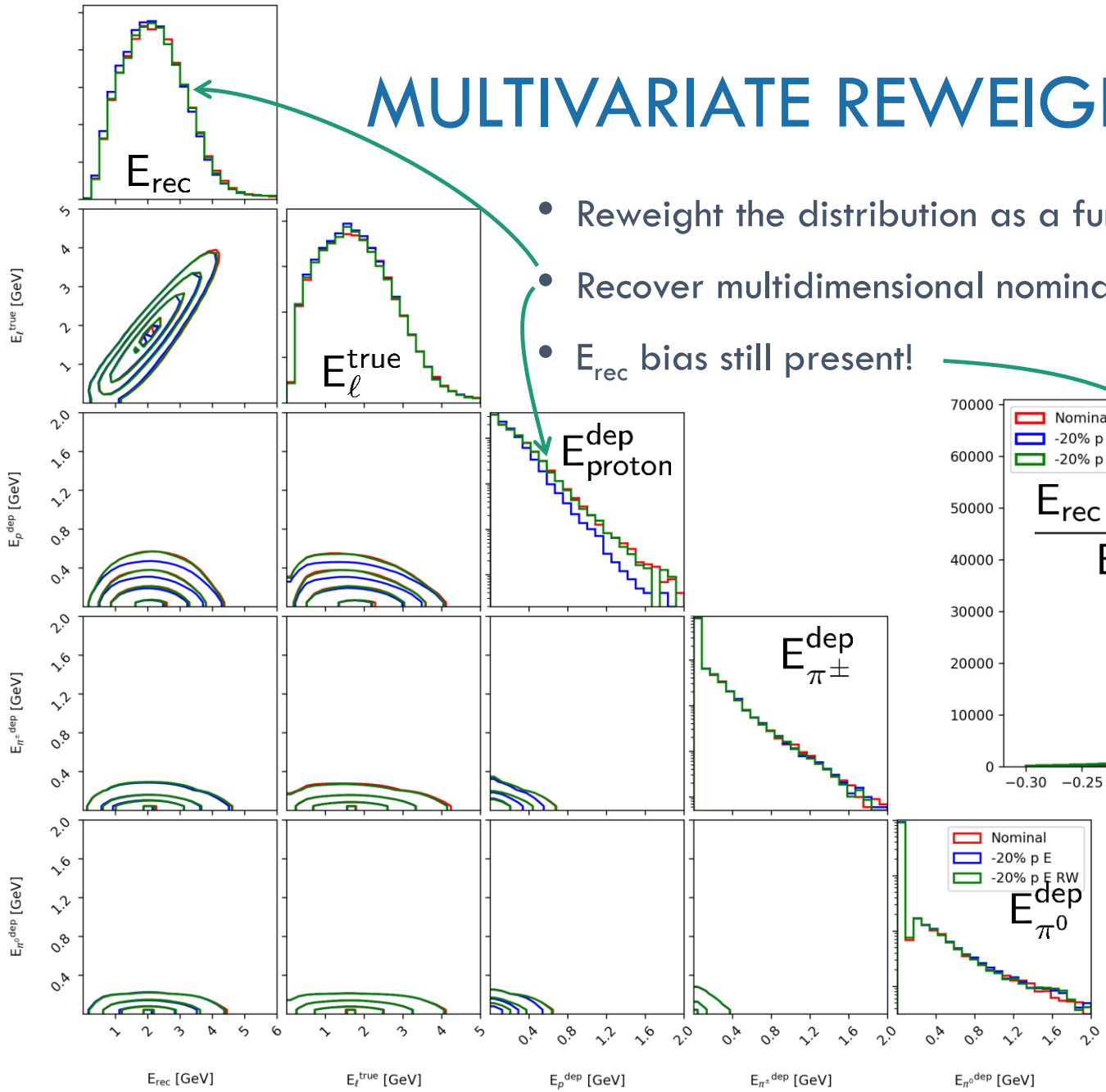
- Apply -20% shift in proton deposited energy.
- Changes $E_\nu \rightarrow E_{rec}$ relation.

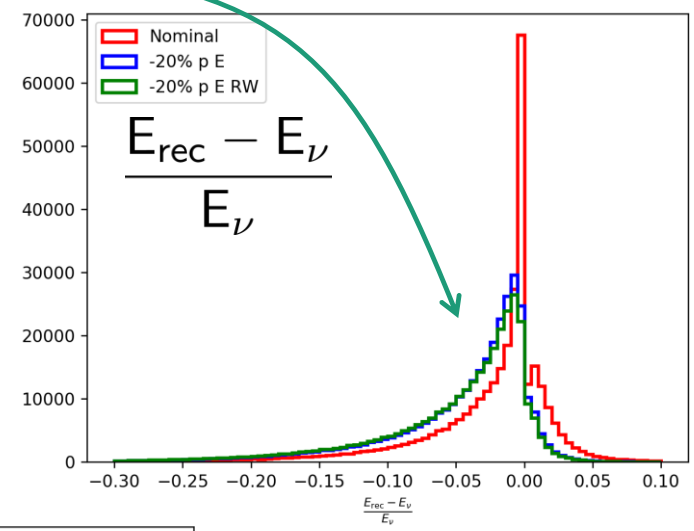




MULTIVARIATE REWEIGHTING

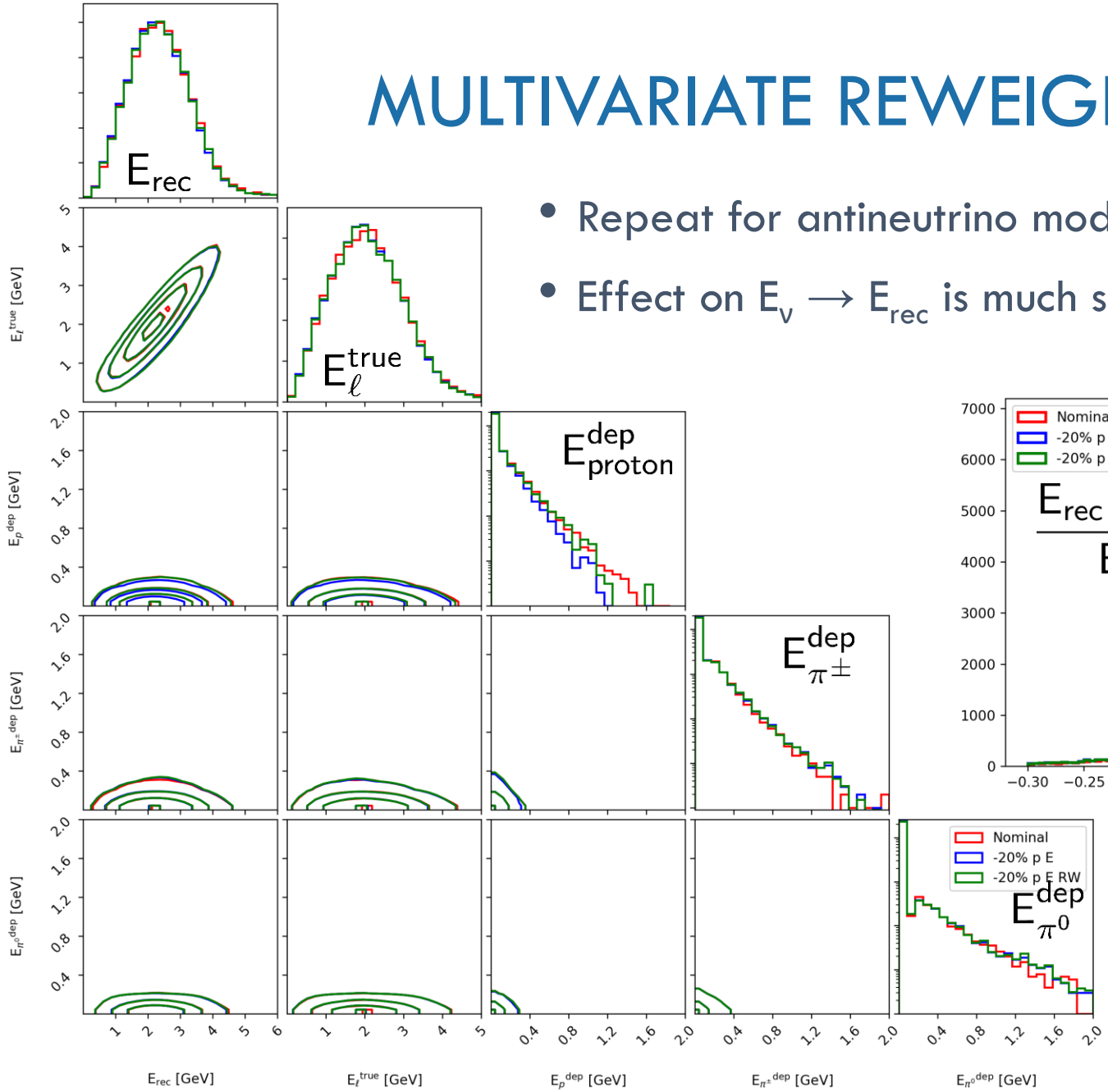
- Reweight the distribution as a function of the observables.
- Recover multidimensional nominal distribution.
- E_{rec} bias still present!

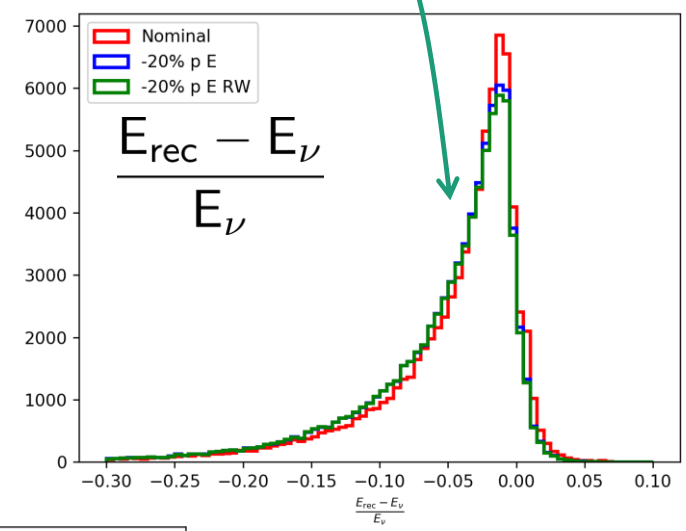




MULTIVARIATE REWEIGHTING

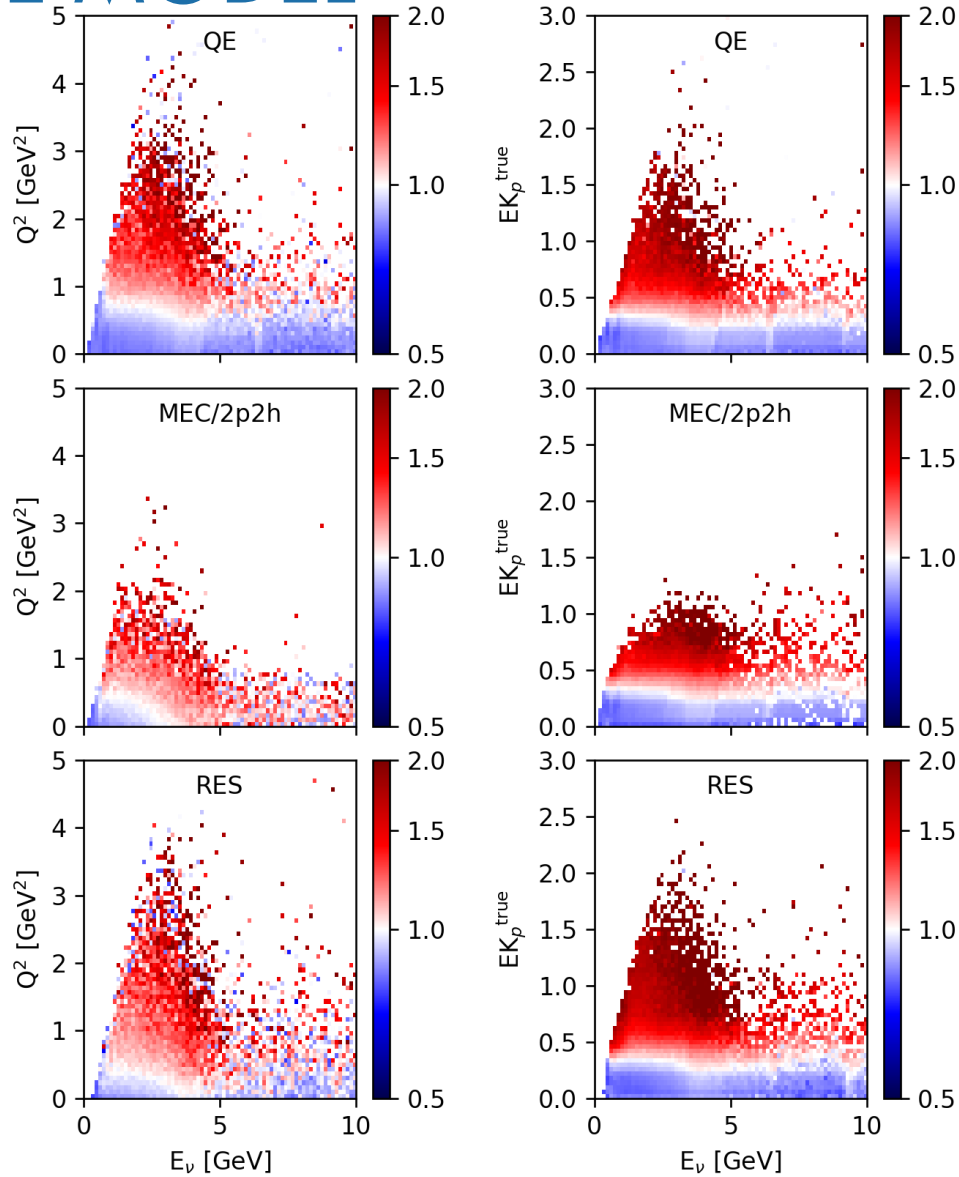
- Repeat for antineutrino mode.
- Effect on $E_\nu \longrightarrow E_{rec}$ is much smaller.





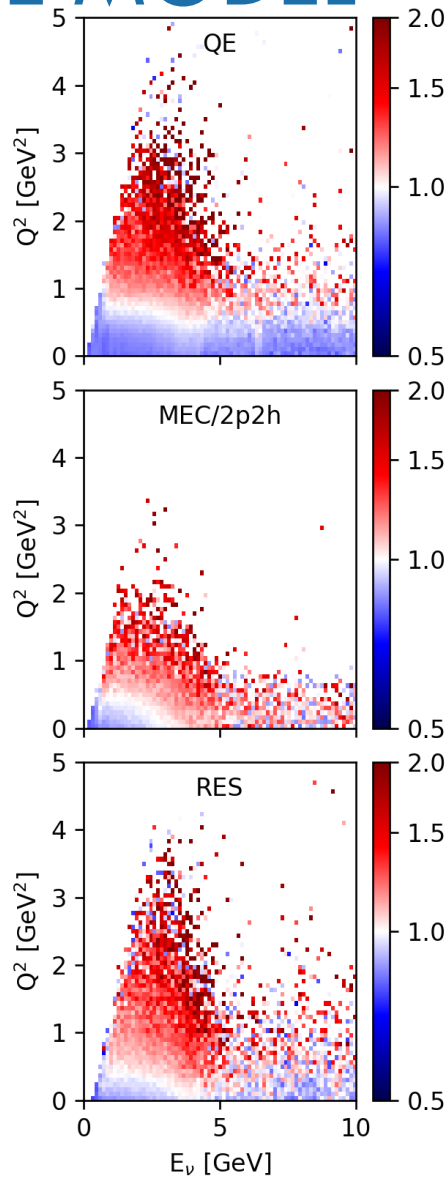
PROPAGATING THE MODEL

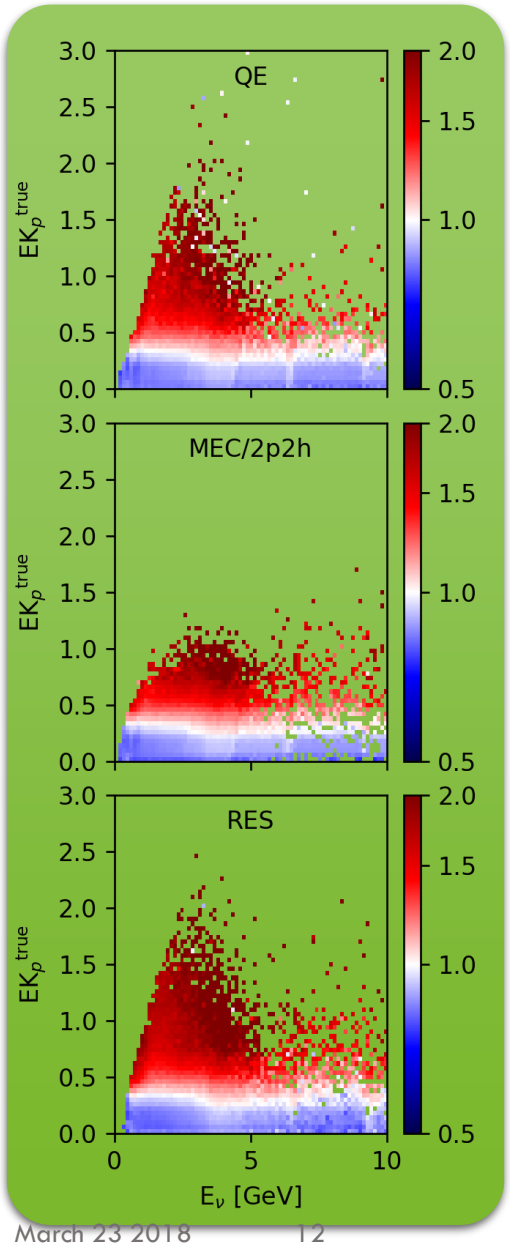
- To study the effect on oscillation fits, we need to propagate this model to far detector.
  - Also to off-axis near detector stops, to demonstrate the PRISM technique.
- Bin event weights in true variables typically used to describe interaction models.
  - Get smoothly varying functions!
  - MVA treats interaction modes differently.
    - Even though it doesn't "know" about them!



PROPAGATING THE MODEL

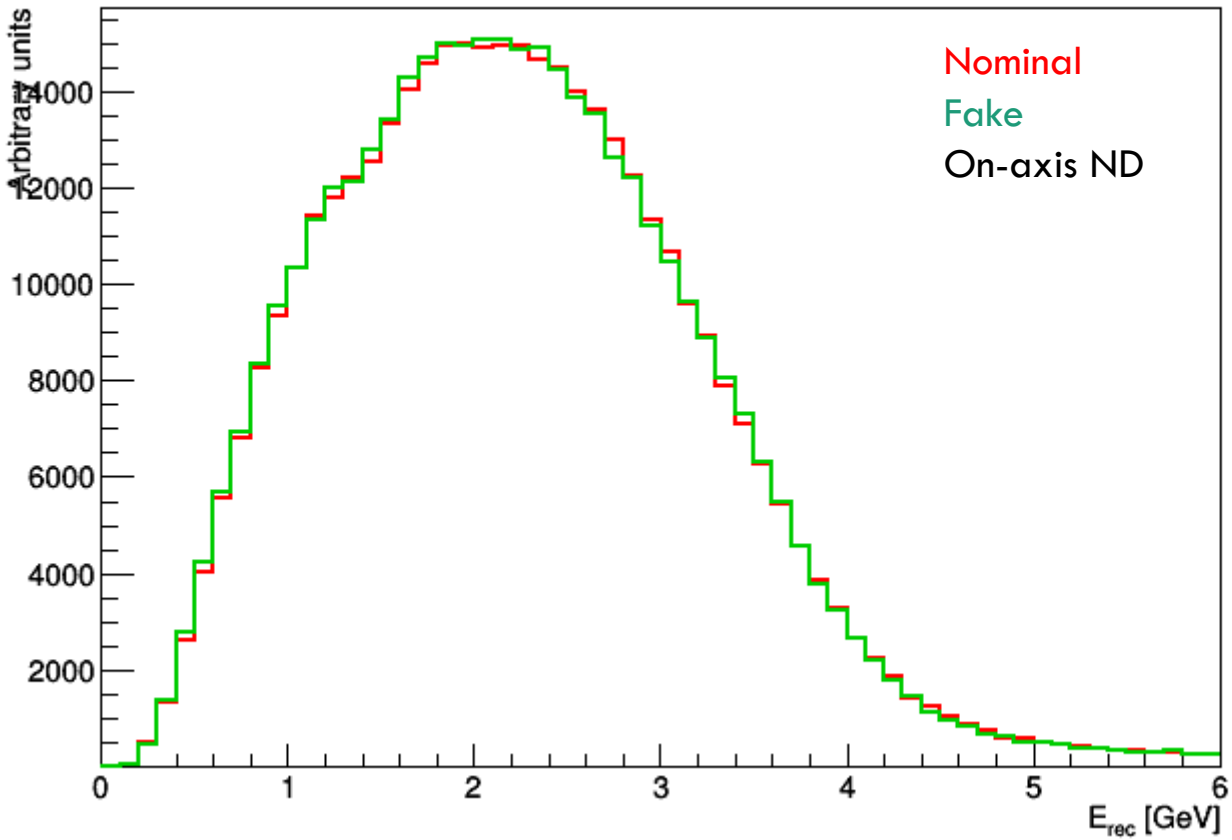
- For this study, use E_ν vs true proton kinetic energy.
- Extract weights separately for ν and anti-ν using FHC and RHC on-axis near detector data.
  - Assume perfect charge separation.
- Do not reweight regions of the space that fall outside of the ND acceptance.
  - These events get weight = 1, but 20% proton deposited energy removed.





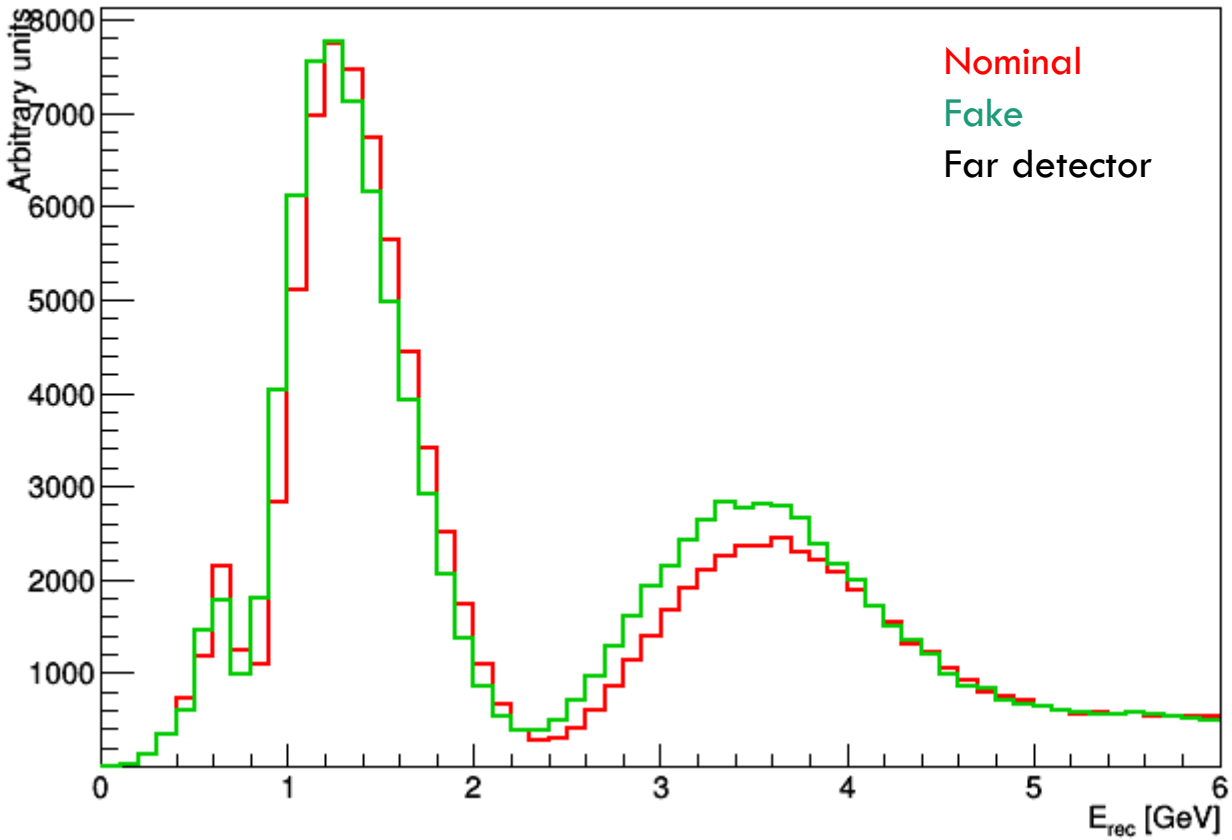
ON-AXIS NEAR DETECTOR

- Very little difference between nominal and fake data sets at on-axis near detector.



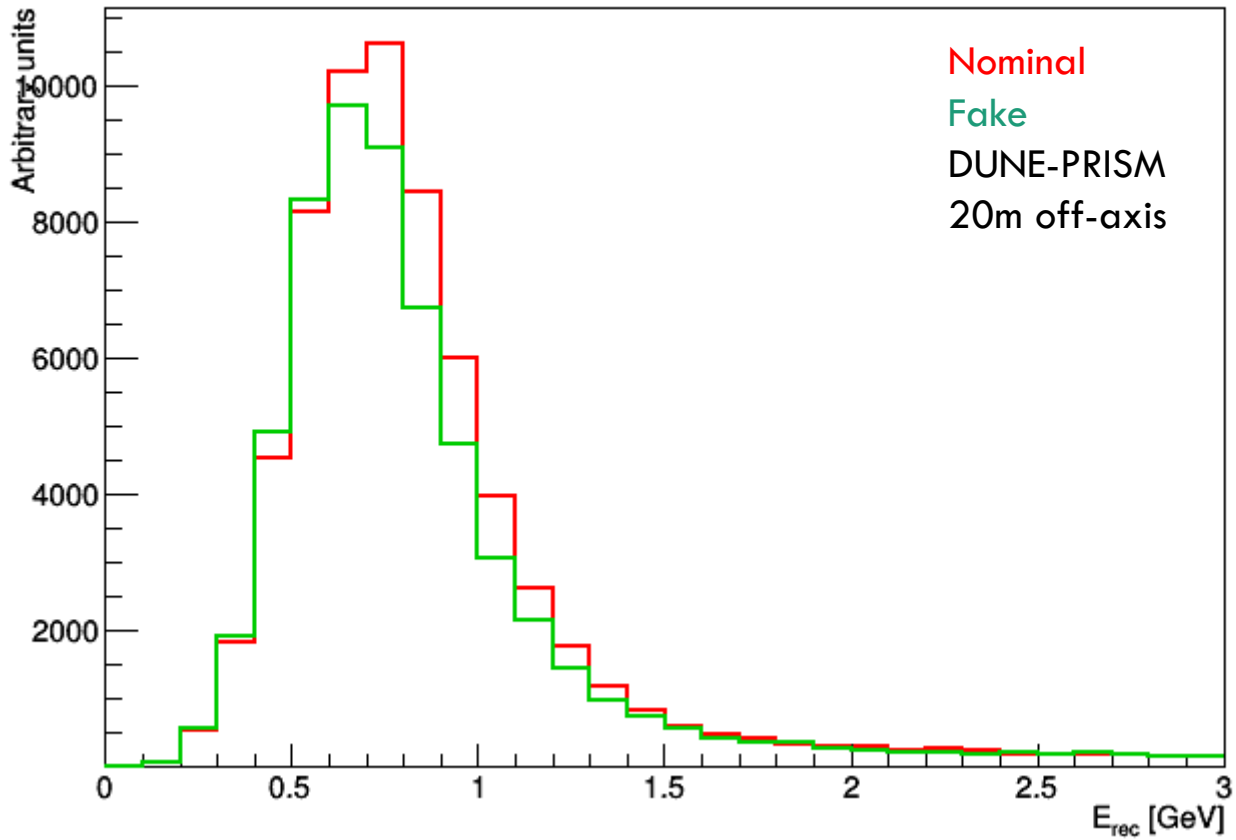
FAR DETECTOR

- Different $E_\nu \rightarrow E_{rec}$ significantly distorts far detector oscillated spectrum.
  - This will induce bias in estimation of oscillation parameters!



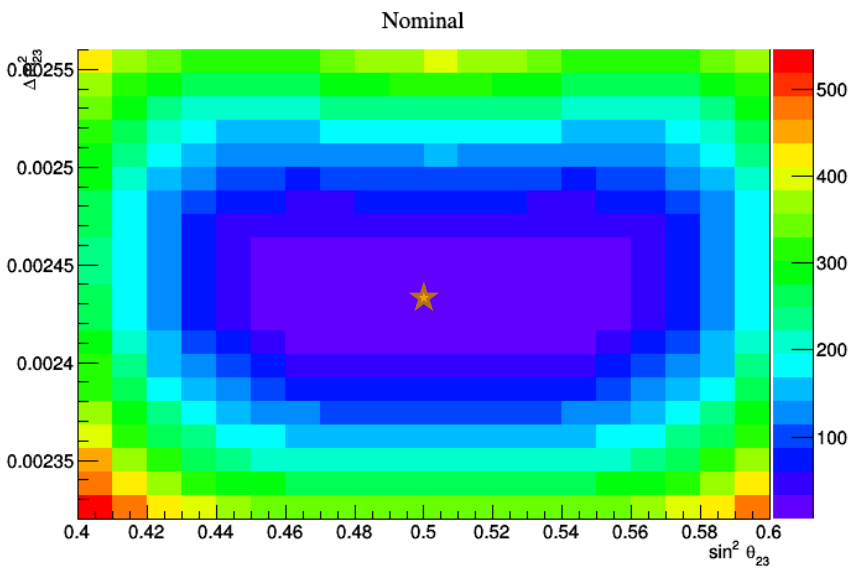
DUNE-PRISM 20 METRES OFF-AXIS

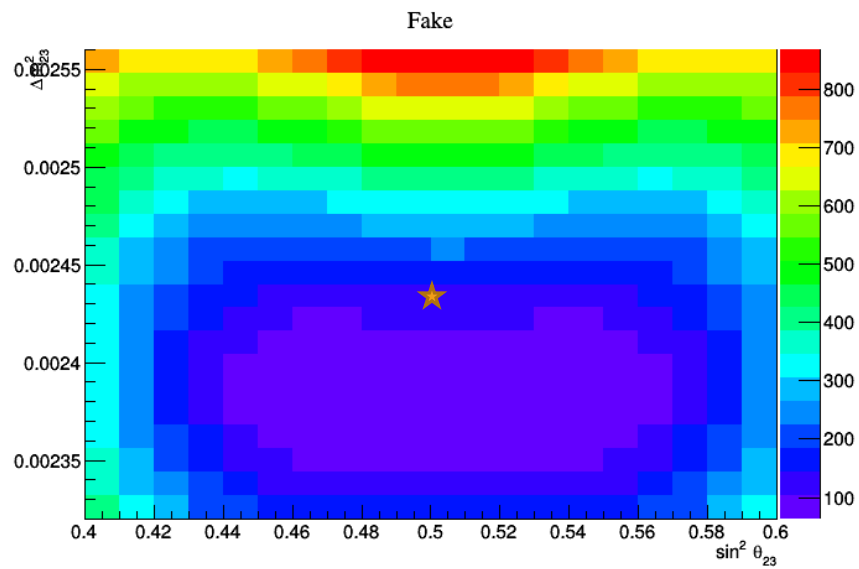
- Fake and nominal data look different when looking at a narrow flux at off-axis positions.



IMPACT ON OSCILLATION ANALYSIS

- Simple χ^2 scan with two-flavour oscillation probabilities, looking only at ν_μ FHC far detector data.
  - No systematics.
  - Statistics equivalent to running for 7 year.





- Detailed CAFAna fit in progress…

SUMMARY

- We have shown that $E_\nu \rightarrow E_{rec}$ mis-modelling is degenerate with reweightable effects (e.g., cross section modelling) looking only at on-axis near detector observables.
  - With optimistic reconstruction assumptions.
- This type of mis-modelling will bias oscillation fits.
- Taking data with DUNE-PRISM off-axis resolves the degeneracy.

MULTIVARIATE REWEIGHTING

- Use Gradient Boosted Decision Tree event reweighting technique*.
- Hyperparameters:
  - Tree splitting criterion: mean squared error
  - Number of estimators: 200
  - Maximum tree depth: 3
  - Minimum samples per leaf: 1000
  - Learning rate: 0.1
  - Loss regularization: 1
- Split MC sample in two: one half will be “Nominal” and the other “Fake”.
- For training, use 75% of the Nominal and Fake samples, and check result on the rest.

*arXiv:1608.05806