Too many trees in RF regressors?

A. Moralejo

Test of RF performance on MC

- Random forest models:

`/fefs/aswg/data/models/AllSky/20240918_v0.10.12_allsky_nsb_tuning_0.00/dec_3476/`

- `Test file:`

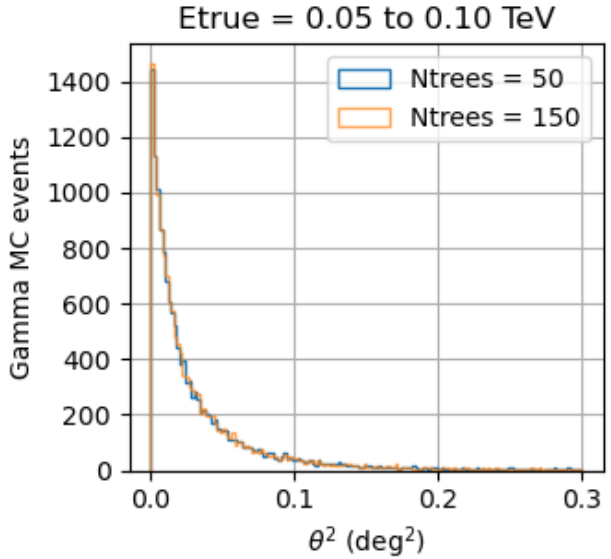
`/fefs/aswg/data/mc/DL2/AllSky/20240918_v0.10.12_allsky_nsb_tuning_0.00/TestingDataset/Gamma/dec_3476/node_theta_10.0_az_248.117_/dl2_20240909_allsky_nsb_tuning_0.00_Gamma_test_node_theta_10.0_az_248.117__merged.h5`

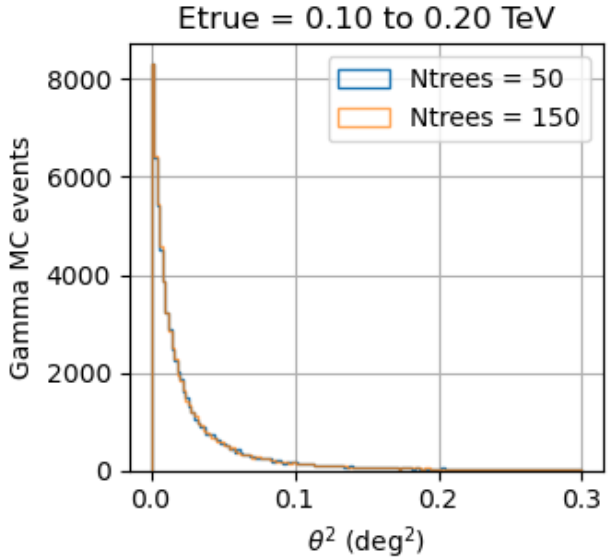
(tested also on a high zenith file, with the same results)

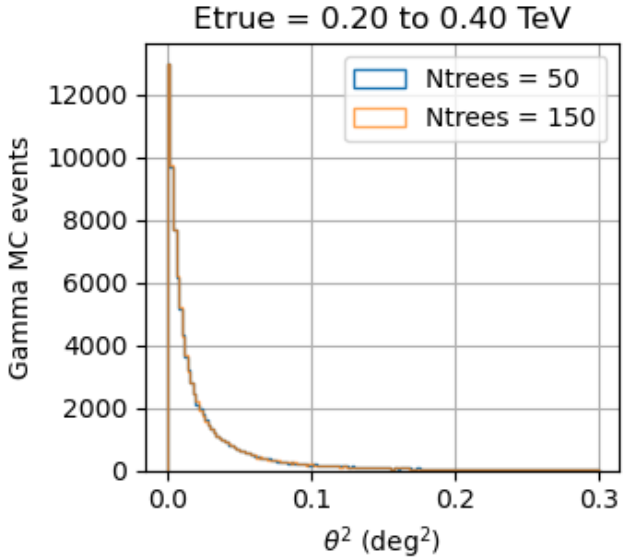
- For the test I simply removed trees from the list “estimators_” of the regressors. It is exactly like growing fewer trees when creating the model
- Shown here is just the result for 50 trees (vs. the original 150). This is a safe choice, one has to go below 5 trees to see a really meaningful degradation of performance
- I only tested the regressors (energy and disp_norm), because they are the most bulky. The classifiers are lighter – plus it is more difficult to test their performance, so I would not touch them

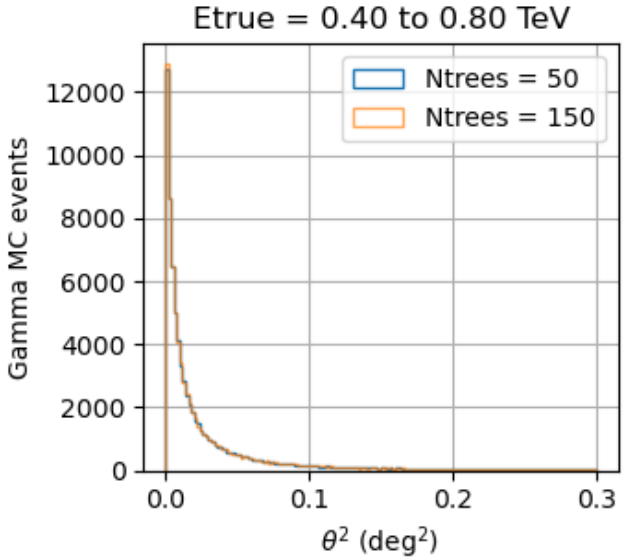
50 vs. 150 disp_norm RF trees θ^2, same disp_sign RF

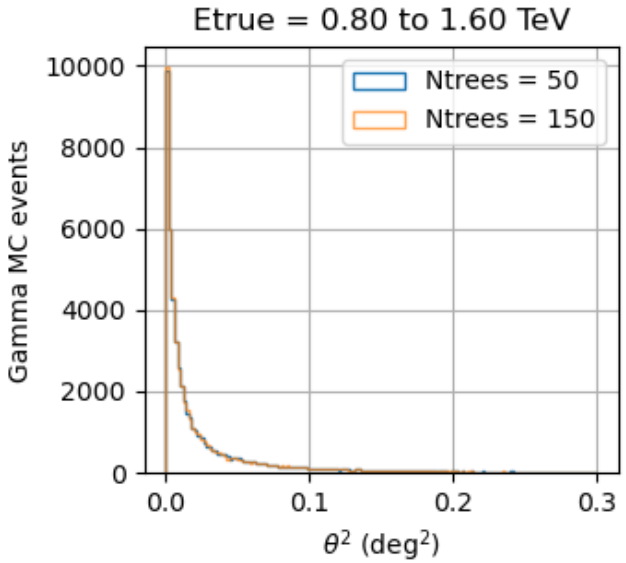
gammaness > 0.7

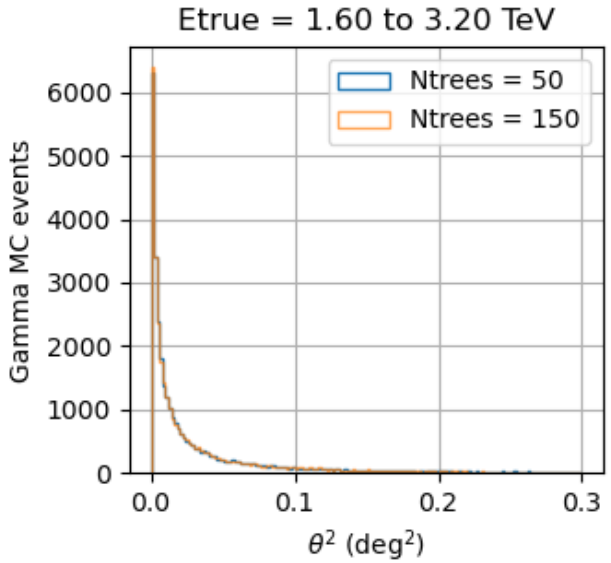


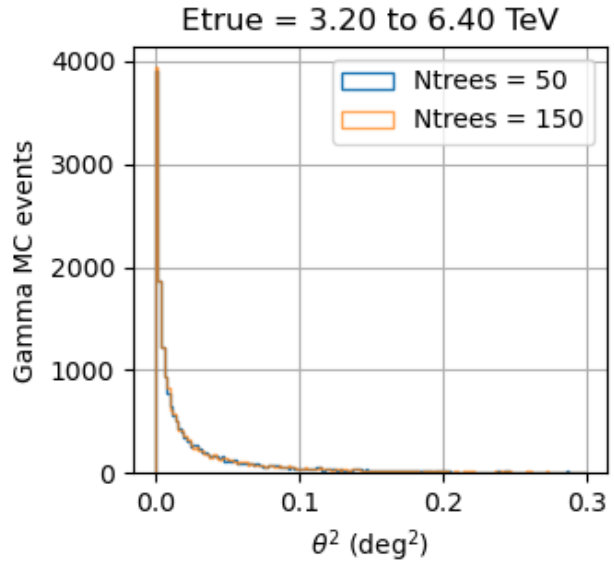


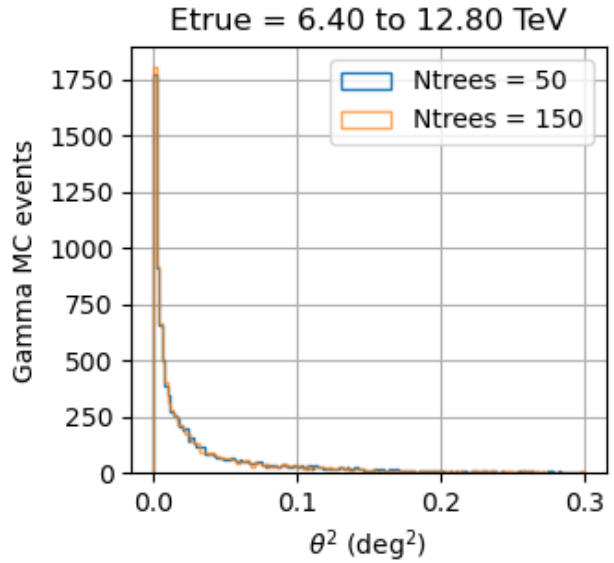






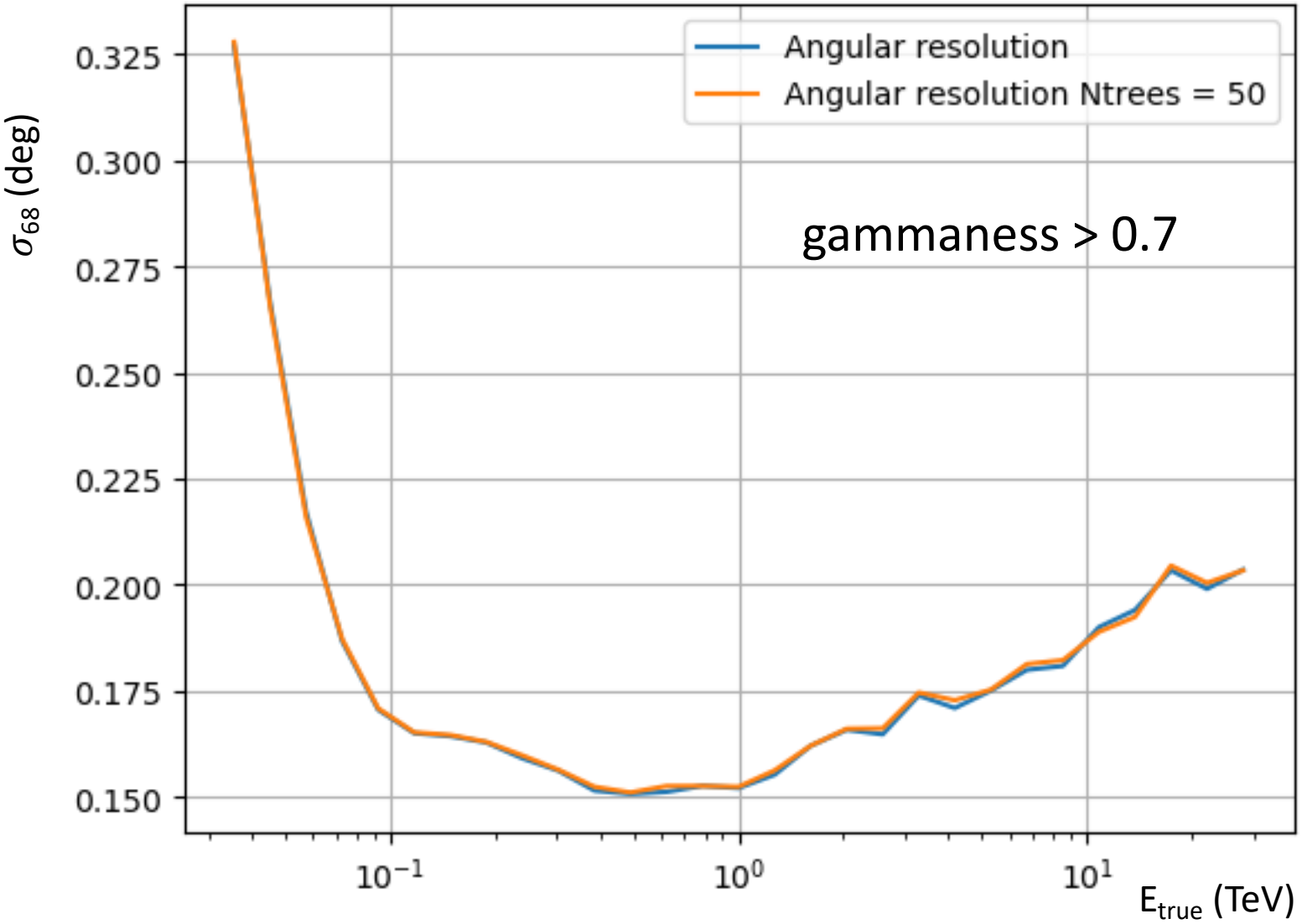






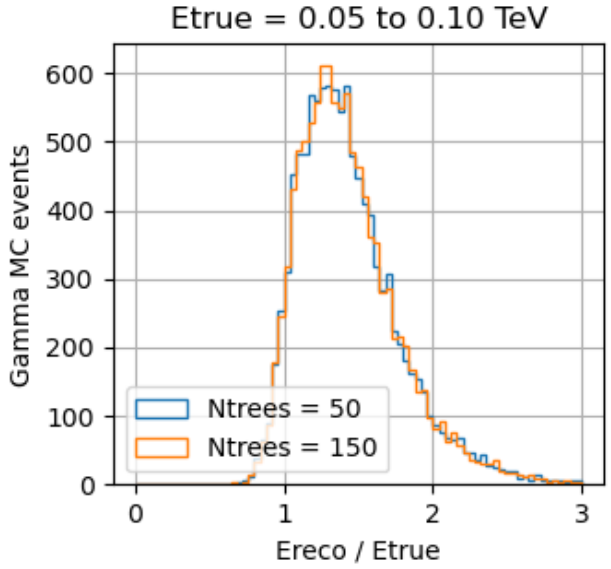
50 vs. 150 disp_norm RF trees, σ_{68} angular resolution

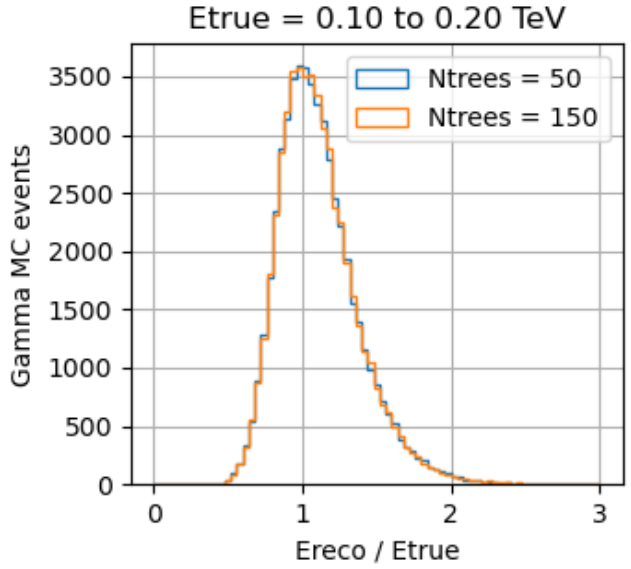
Note: Like in the performance paper, σ_{68} is computed on the population ig events with correct head-tail assignment

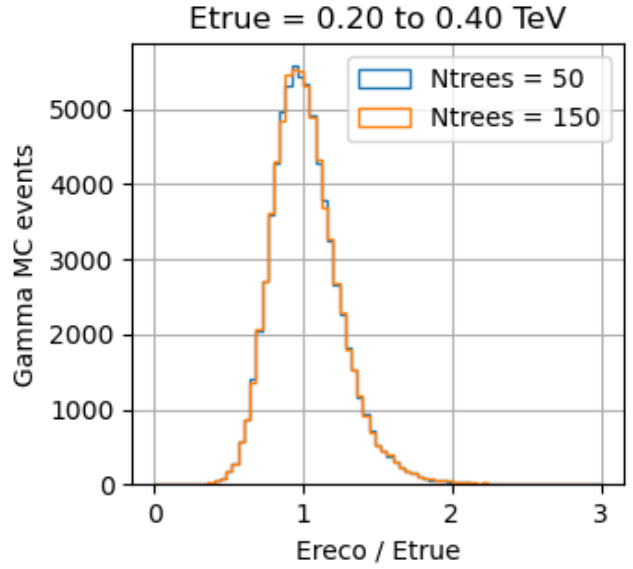


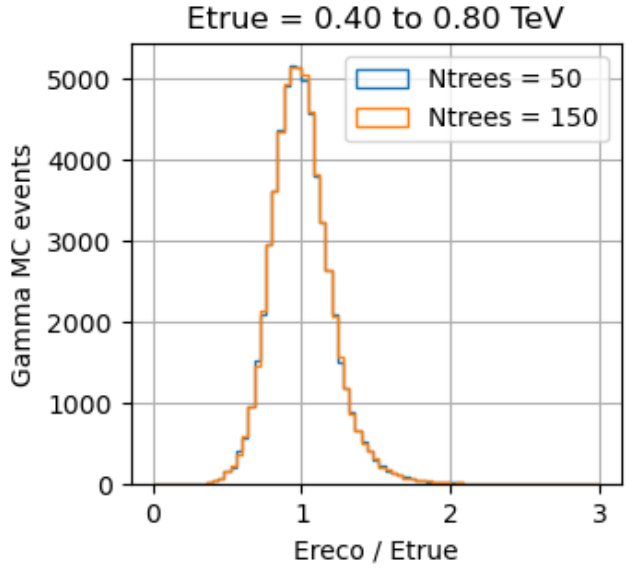
50 vs. 150 energy RF trees, E_{reco} / E_{true}

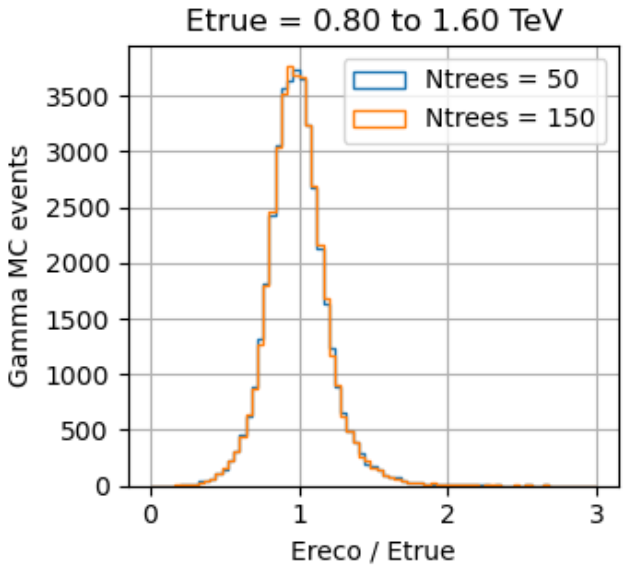
gammaness > 0.7

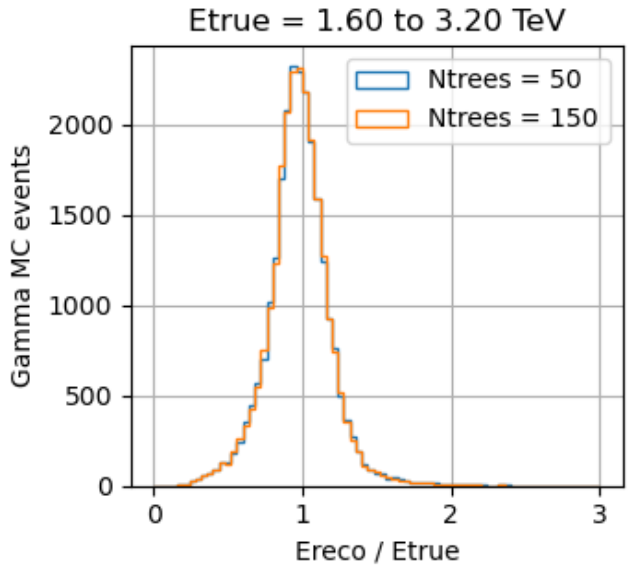


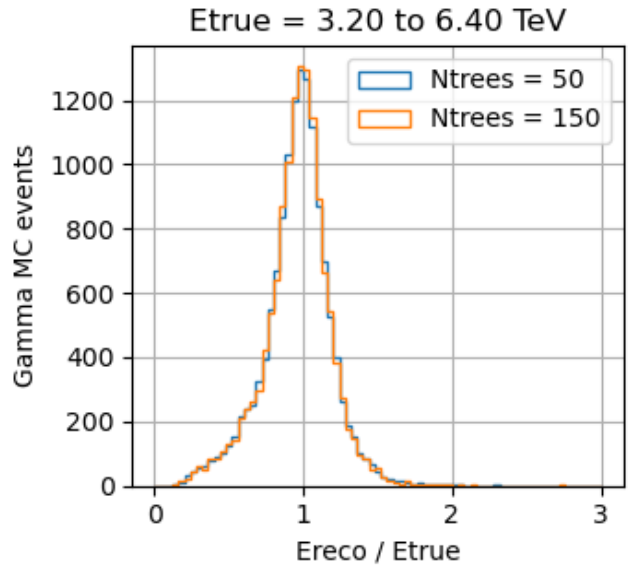


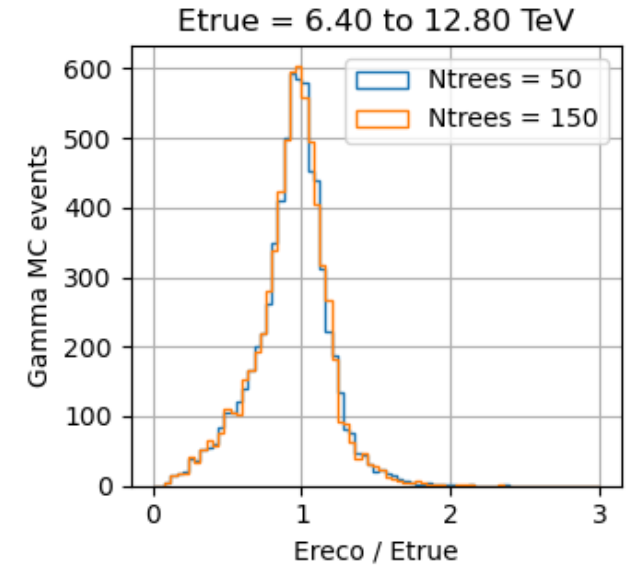




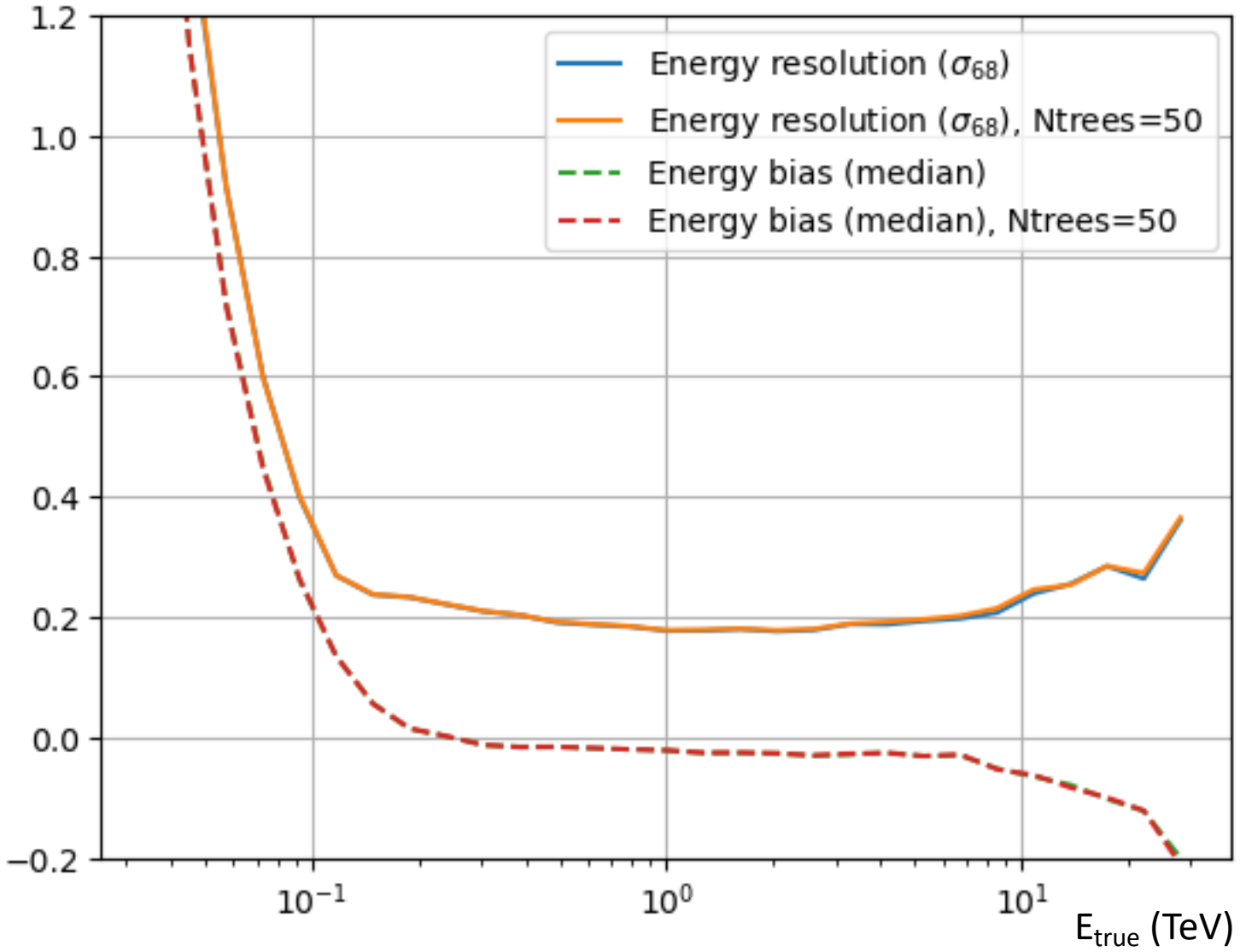








50 vs. 150 energy RF trees, E-resolution & bias



Proposal:

- Change our default “number of estimators” (=trees) for RF regressors from 150 to 50
- This will reduce the memory requirements both in training and application, and also make the dl1 to dl2 step faster