2022年5-8月考核报告

张镇轩

导师：陈明水

2,Sept 2022

Outline:

① Physics Analysis

1. Resonant $HH \to WW\gamma\gamma$
   - HiggsDNA framework:
     - Object selection & Event selection & GEN study
     - Preliminary limit result

2. $H \to \gamma\gamma$ mass measurement with full Run2 dataset
   - Diphoton BDT Retraining

② HGCAL project:

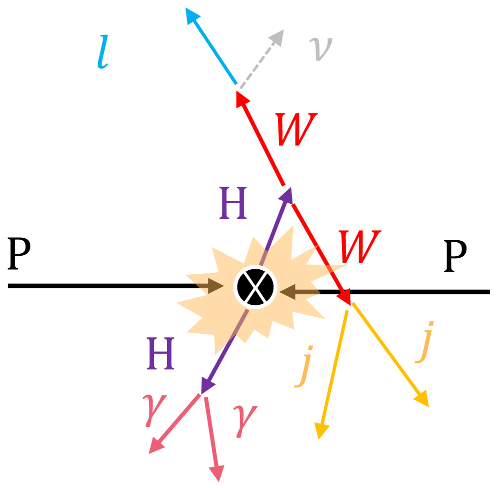
- HGCAL gantry assemble works
  - Vacuum system update
  - Glue system update

③ Plans for near future

Resonant $HH \rightarrow WW\gamma\gamma$

Physics Motivation

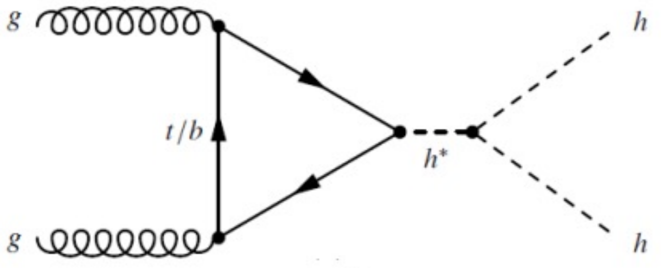
- Search for resonant di-Higgs production at LHC
  - Many BSM theories predict direct or indirect production of new resonances with enhanced cross-section : such as Warped extra dimension (WED), Next-to-minimal supersymmetric standard model(NMSSM) with Radion(spin0) or KK-Graviton(spin2)
  - direct coupling with Higgs boson
- Resonant HH → $WW\gamma\gamma$ has not been studied yet
- $WW\gamma\gamma$ final state:
  - WW has second large BR in Higgs decay
  - $\gamma\gamma$ has good mass resolution, easy to trigger and good efficiency



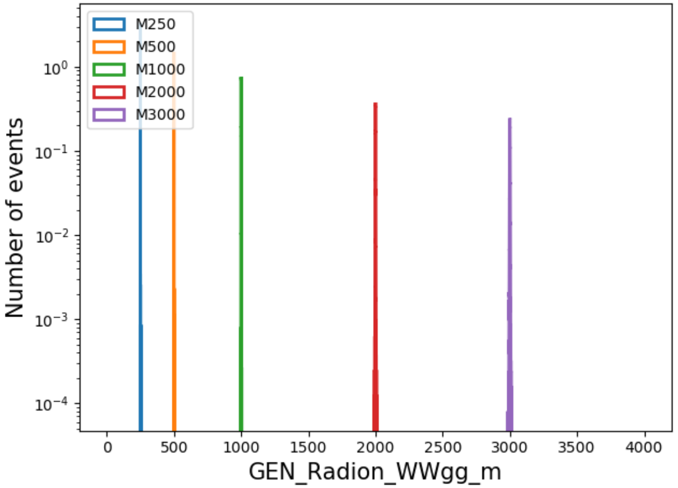
- W decay modes:
  - $W \rightarrow e\overline{\nu_e}$(BR = 10.7%)
  - $W \rightarrow \mu\overline{\nu_\mu}$(BR = 10.63%)
  - $W \rightarrow \tau\overline{\nu_\tau}$(BR = 11.38%)
  - $W \rightarrow q\overline{q}$(BR = 67.41%)

leptonic

hadronic



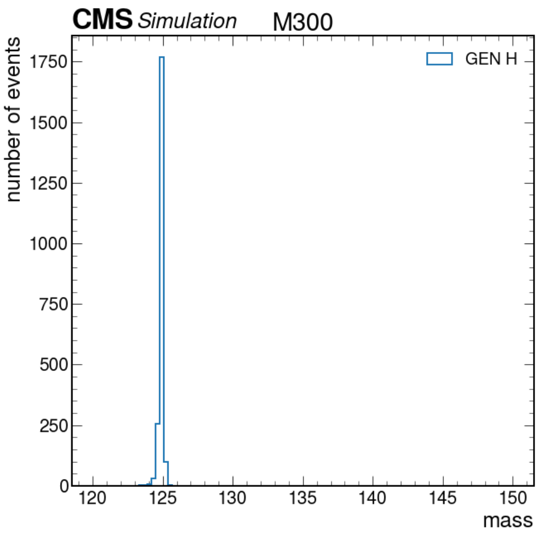
Resonance

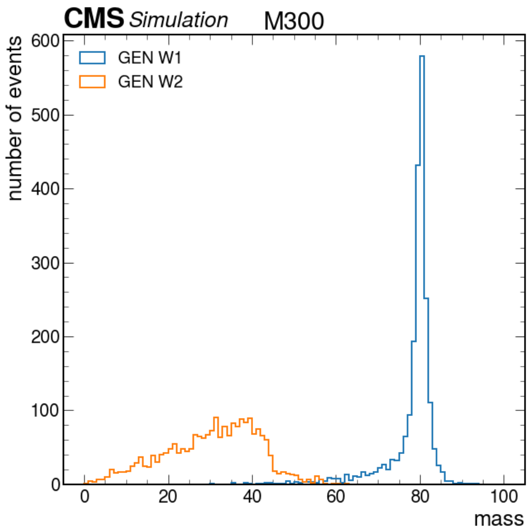


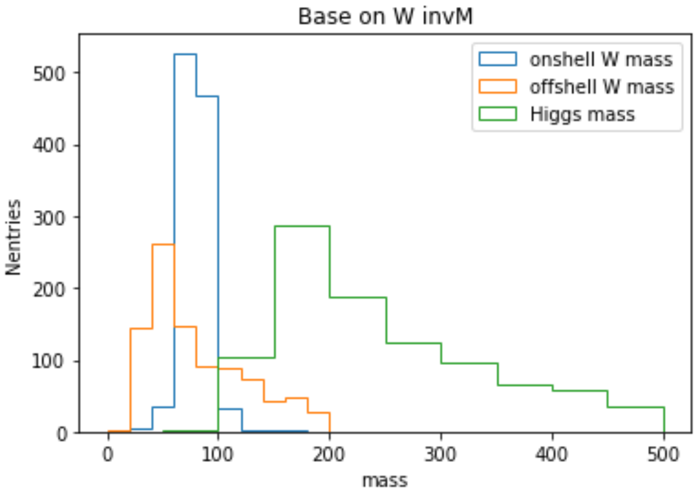
Object selection & Event selection & GEN study

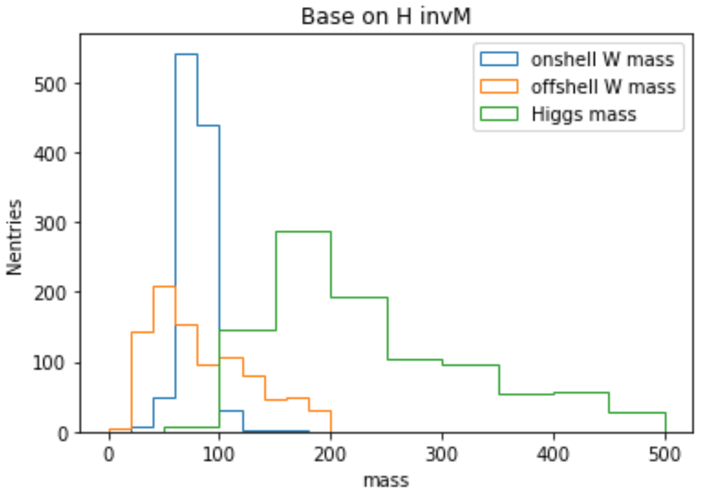
- Add our object selection and event selection in HiggsDNA
  - Created a tagger for $HH \rightarrow WW\gamma\gamma$ preliminary selection(same as previous presentation（2022.04.15） selection criteria)
- Add GEN study in HiggsDNA
  - Created a gen_selection function based on numba to get the gen information for the DNN training
  - Form the 4 gen quarks (reco jets) as W boson based on invariant mass based on awkward array

```
"variables_of_interest" : [
    ["GenW1_qq","pt"],["GenW1_qq","eta"],["GenW1_qq","phi"],["GenW1_qq","mass"],
    ["GenW2_qq","pt"],["GenW2_qq","eta"],["GenW2_qq","phi"],["GenW2_qq","mass"],
    ["GenHWW_qqqq_Higgs","pt"],["GenHWW_qqqq_Higgs","eta"],["GenHWW_qqqq_Higgs","phi"],["GenHWW_qqqq_Higgs","mass"]
```



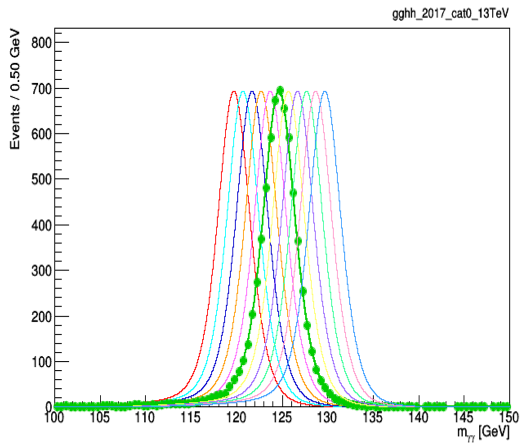


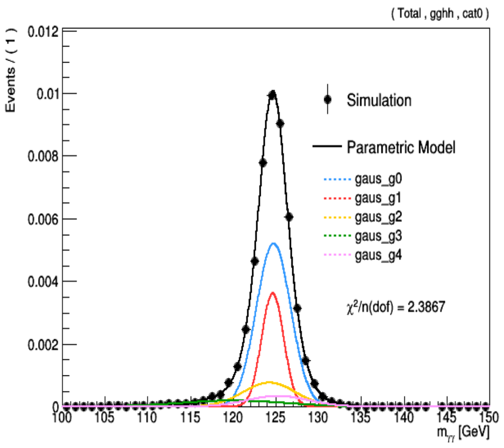


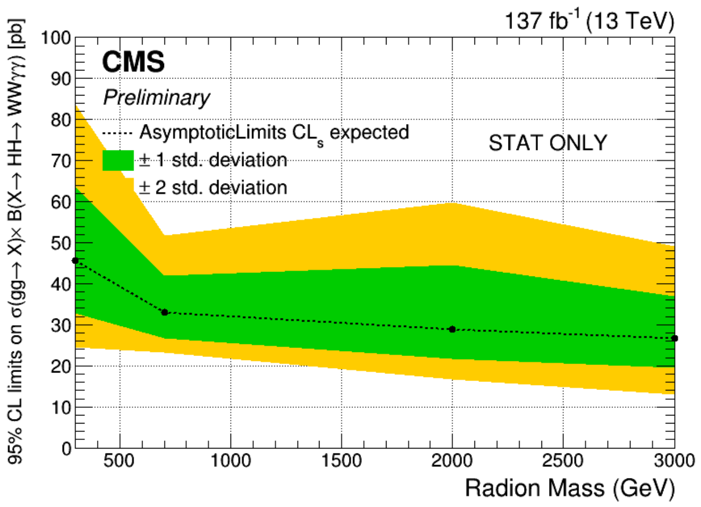


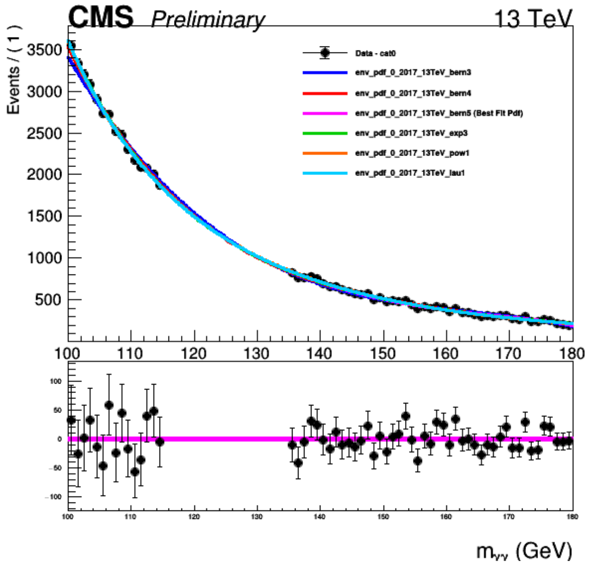
Preliminary limit result

- Flashgg final fit framework
- FullyHadronic channel
- plots mass point 300,700,1250,3000 manually, will try to write a script to run all mass points at once.
- Signal modeling:
  - only 125 GeV mass point with the best $\chi^2/n(dof)$ score
- Background modeling:
  - Perform f-Test to decide which functions(I require (weak) goodness-of-fit criteria)
  - choice of bkg pdf treated as additional discrete nuisance parameter
- Limit result:









$H \rightarrow \gamma\gamma$ mass measurement with full Run2 dataset

Introduction:

- Diphoton BDT used in HIG-19-015 and currently implemented in flashgg, was trained with
  - XGBoost
  - **MC samples** from Legacy16 + UL17 + ReReco18
    - pp from diphoton+jets MC
    - pf + ff from GJets and QCD MC
    - QCD samples **down-weighted** by a factor of 25 to prevent overtraining due to the high weights of QCD dijet samples
  - so-called "**default**" training in the following slides
- Trying to retrain diphoton BDT for each UL16, UL17 and UL18 with
  - Data-driven QCD+Gjets (ff+pf)
  - Two API:
    - TMVA BDT in ROOT: so-called '**TMVA**' in the following slides
    - XGBoost : so-called '**XGBoost**' in the following slides
- Present the retraining results for UL17&UL18 today

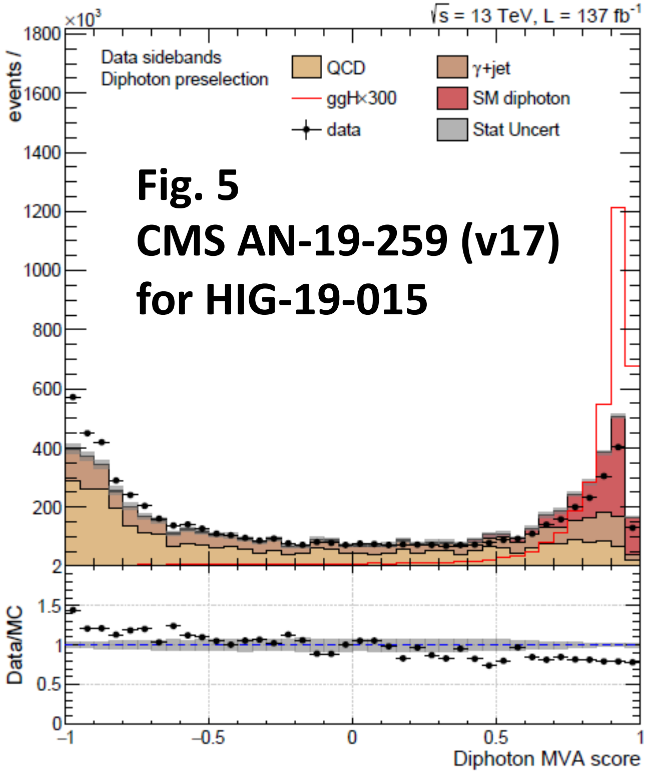
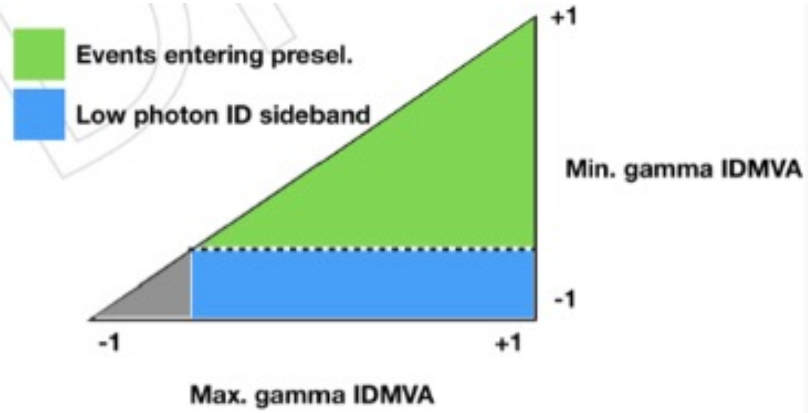


Fig. 5
CMS AN-19-259 (v17)
for HIG-19-015

Data-Driven QCD

- Similar strategy as [HIG-19-013] (ttH($\gamma\gamma$))
- Events which fail the pre-selection cut on minimum photon ID MVA **in data**:
  - "low photon ID sideband" : maxID>-0.9 and -0.95 < minID < -0.9
- Generating a new value of the minimum photon ID MVA in the range [-0.9, max photon ID MVA] randomly, with probability for each value dictated by the **fake PDF from fake photons in γ+jets MC** ("pf" with maxID>-0.9 and minID>-0.95)
- To have the proper shape for the maximum photon ID MVA distribution, events are assigned an additional, individual weight

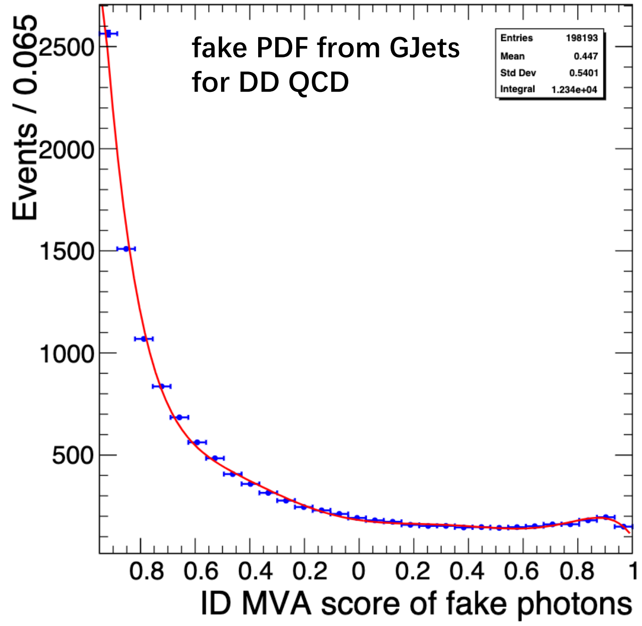


$$w = \frac{\int_B^{\max \gamma\, ID} \text{fake PDF}}{\int_A^B \text{fake PDF}}$$

$A \equiv$ minimum value of photon ID MVA in the low photon ID sideband $= -0.9$

$B \equiv$ preselection cut value on minimum photon ID MVA $= -0.7$

$$W = \frac{\int_{-0.9}^{maxID} fake\ PDF}{\int_{-0.95}^{-0.9} fake\ PDF}$$



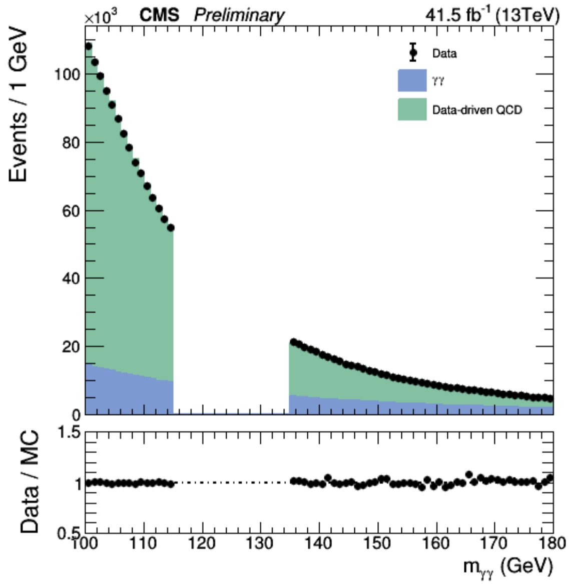
Normalization and reweighting

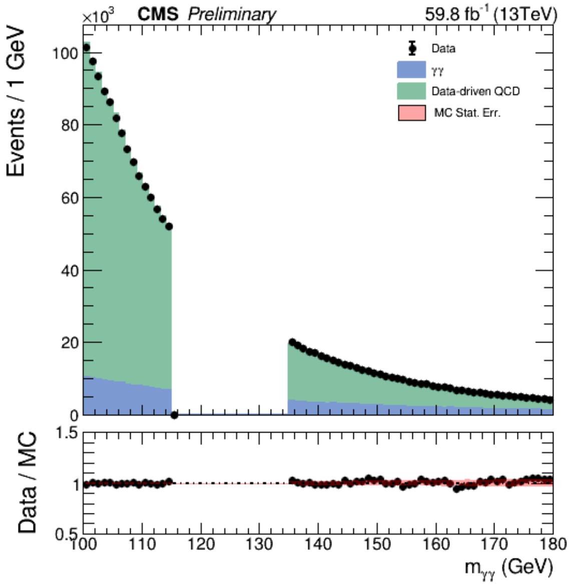
- **Normalization SFs** of MC pp and data-driven pf+ff: 2D fit on min. and max. ID MVA

- **Min ID MVA photon σ_E/E reweighting**: σ_E/E distribution in data-driven pf+ff reweighted to that of "**data-MC $\gamma\gamma$**", to obtain better data/MC agreements on the mass resolutions (σ_{rv} and σ_{wv})

- **Photon PT reweighting**: 2D pT of (min ID-MVA γ, max ID-MVA γ) in data-driven pf+ff reweighted to "data-MC $\gamma\gamma$", ", to obtain better data/MC agreements on pT/mass for each photon candidate

- Checked data/MC comparisons for all input training variables, with data/MC mass-sideband events : next slide
  - Events passing H→γγ preselection





Diphoton BDT Retraining(UL17 & UL18):

- **Samples :**
  - Signal: 125 GeV ggH, VBF, VH, ttH signal events
  - Backgrounds: MC pp (Diphoton+jets) plus data-driven pf+ff
- **Training strategy closely follows that in HIG-16-040/HIG-19-015**
  - Same selections and 10 input variables to train the classifier
  - Signal events is weighted, to assign a higher score for the events with a better mass resolution

$$w_{sig} = \frac{p_{vtx}}{\sigma_{rv}} + \frac{1 - p_{vtx}}{\sigma_{wv}}$$

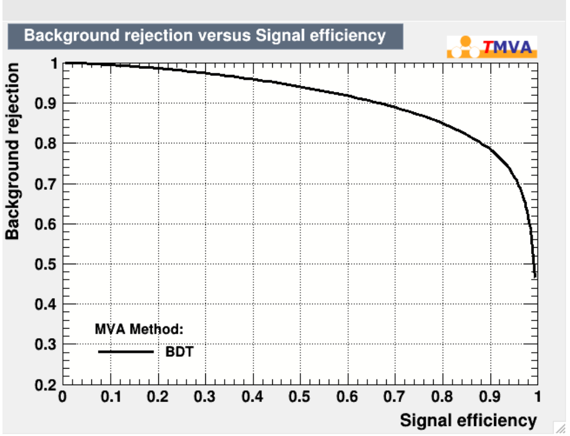
  - Also re-weight events to equalize the sum of weights of the signal and background for XGBoost
- Retraining with two different API:
  - TMVA BDT : ROOT api, identical configurations as in HIG-16-040, used for the cross check of the retraining performance
  - XGBoost : python library, sklearn api
- 70% of events for training and the rest 30% of events for testing, for both sig and bkg

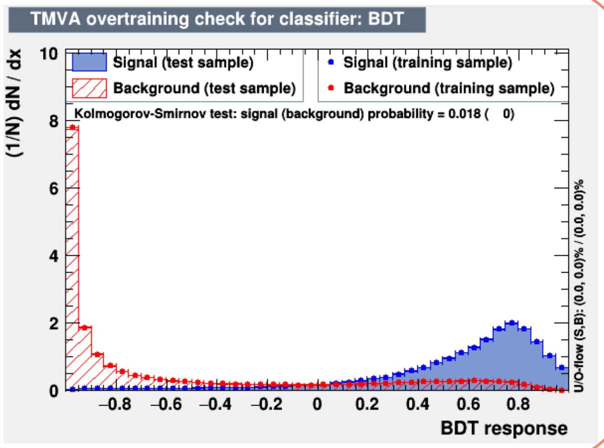
Training results

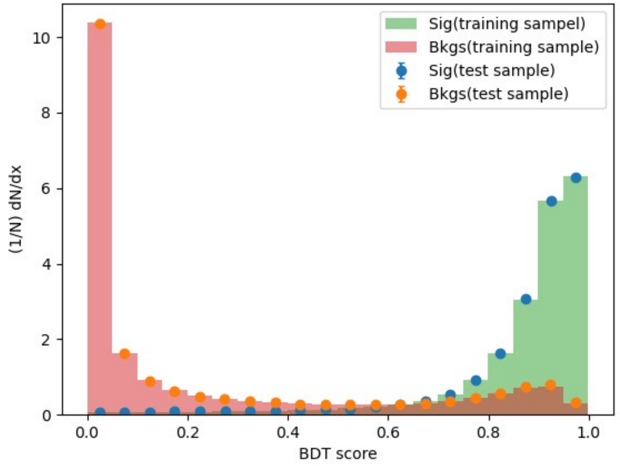
➢ TMVA BDT:ROC curve & score distributions

!H:!V:!CreateMVAPdfs:BoostType=Grad:UseBaggedBoost: :GradBaggingFraction=0.6:SeparationType=GiniIndex:nCuts=2000:MinNodeSize=0.125:Shrinkage=0.1:NTrees=2000:!UseYesNoLeaf:MaxDepth=3

Same as in HIG-16-040





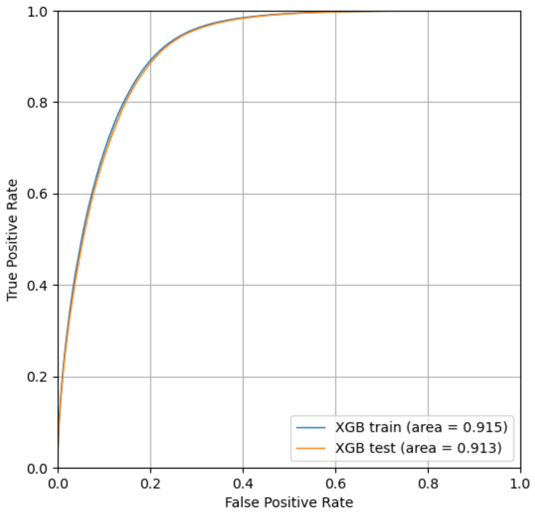


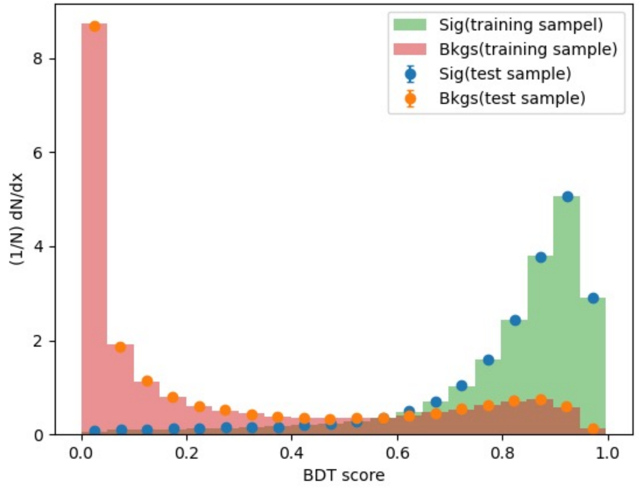
➢ XGBoost: ROC curve & score distributions

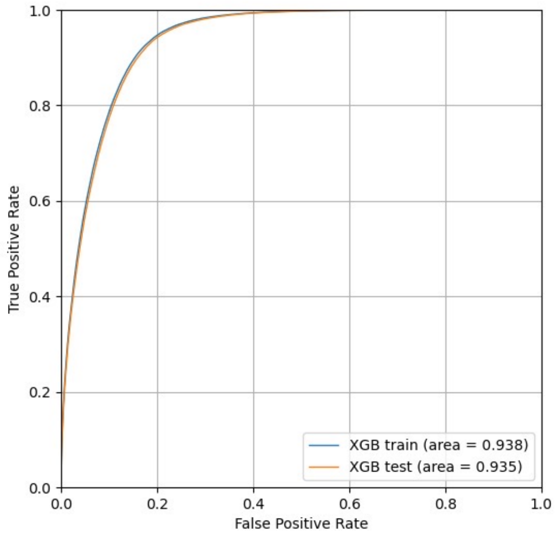
➢ **Bayesian optimization:**

tune the hyperparameters to make sure ROC score for training set always within ~0.5% of test set(no overtraining)

have the best test AUC at the same time







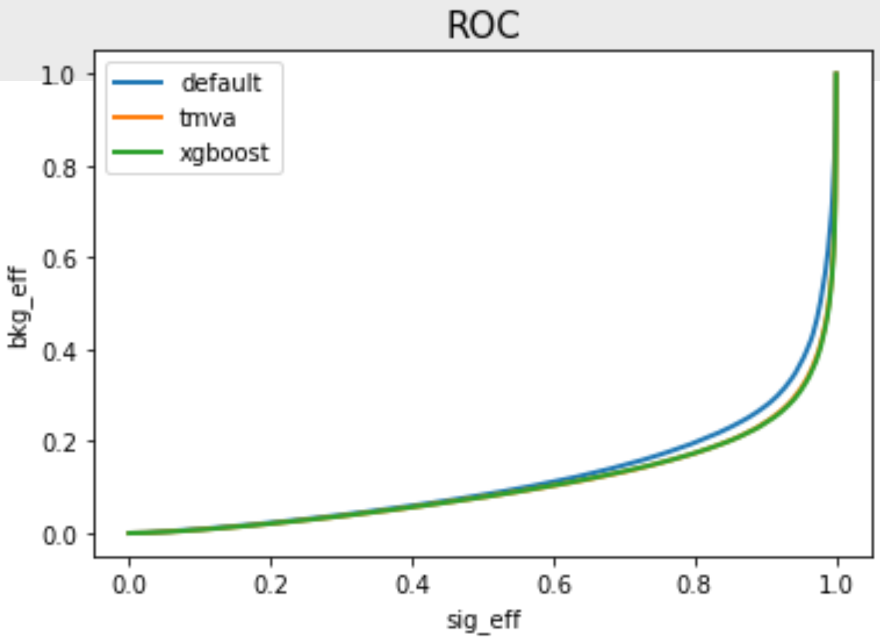
Performance comparisons

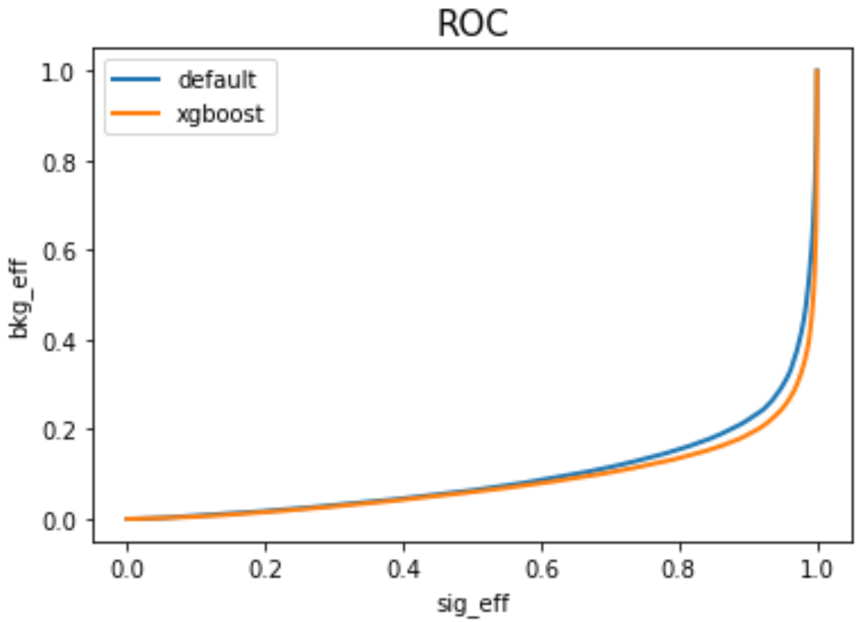
- Signal: all 125 GeV signals events
- Bkg: MC pp plus data-driven pf+ff
- AUC values:

UL17	UL18
Default AUC: 0.87807	Default AUC: 0.9025
TMVA AUC: 0.89251	XGB better than TMVA, so UL18 didn't cross check with TMVA
XGB AUC: 0.89284	XGB AUC: 0.9159

- Better performance from the retraining compared to the exist "default" training

- Almost identical performance between "TMVA" and "XGBoost" retraining
  - A tiny better performance with "XGBoost"



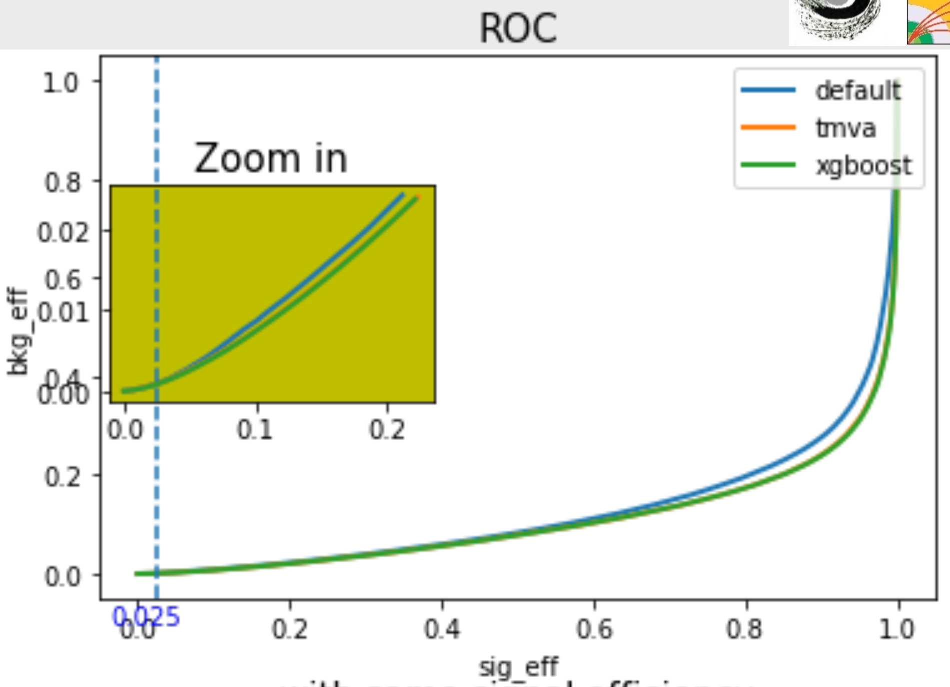


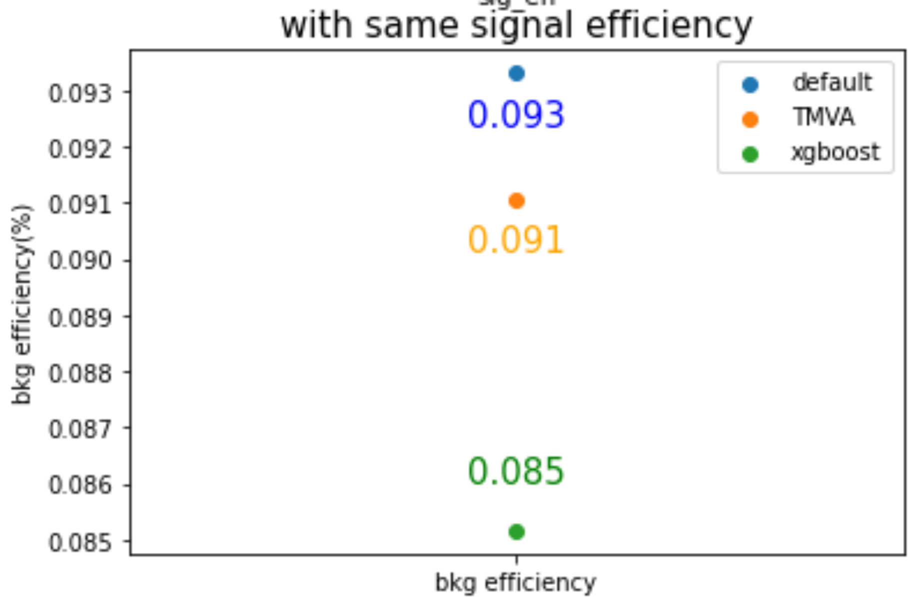
Performance comparisons for UL17:

- For the best event category of the mass measurement, as the preliminary results presented by Neil (link) (with the **transformed** "default" diphoton BDT > **0.932**), the sig eff is **~2.5%** with respect to the preselection

- With the same sig eff, compared bkg eff:
  - "default" training: 0.093%
  - "TMVA" retraining: 0.091%
  - "XGBoost" retraining: 0.085%

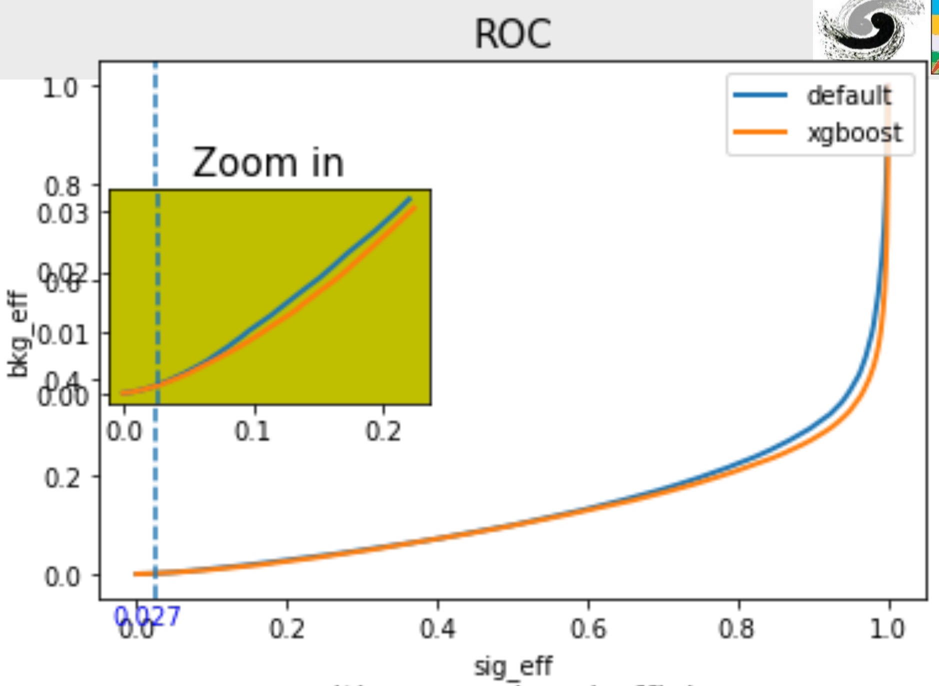
- Compared to the "default" training, "XGBoost" retaining can improve S/$\sqrt{B}$ by ~5%



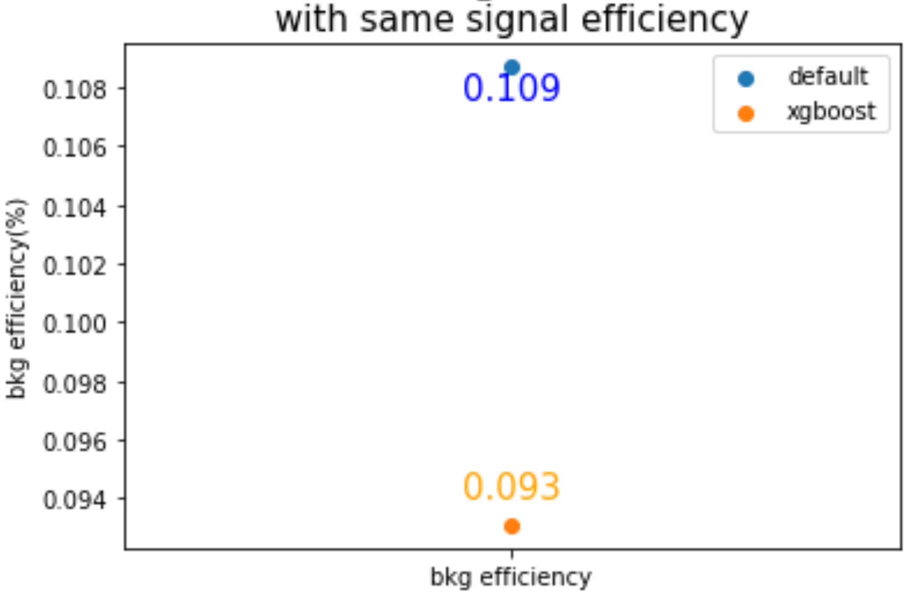


Performance comparisons for UL18

- For the best event category of the mass measurement, as the preliminary results presented by Neil (link) (with the **transformed** "default" diphoton BDT > **0.926(common boundary)**, the sig eff is **~2.7%** with respect to all events passing the preselection



- With the same sig eff, compared bkg eff:
  - "default" training: 0.109%
  - "XGBoost" retraining: 0.093%

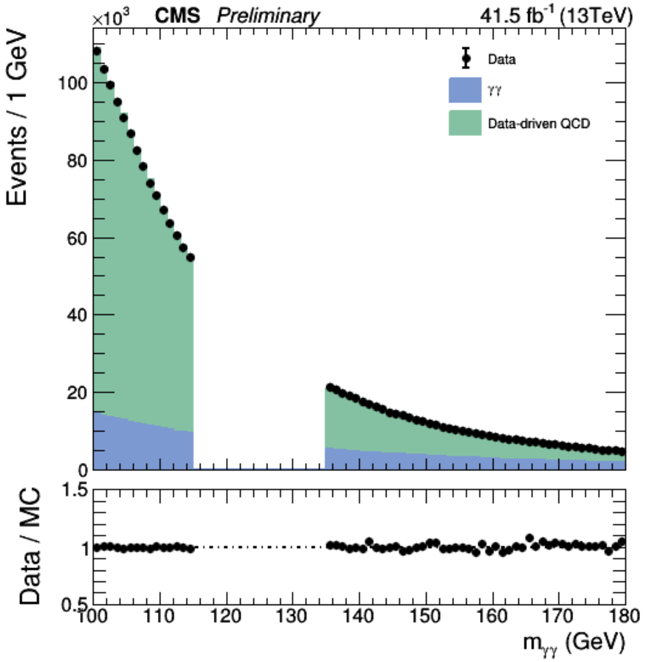


- Compared to the "default" training, "XGBoost" retaining can improve $S/\sqrt{B}$ by ~8% and AMS($=\sqrt{2((S+B)\ln\left(1+\frac{S}{B}\right)-S)}$) by ~5%

Validations

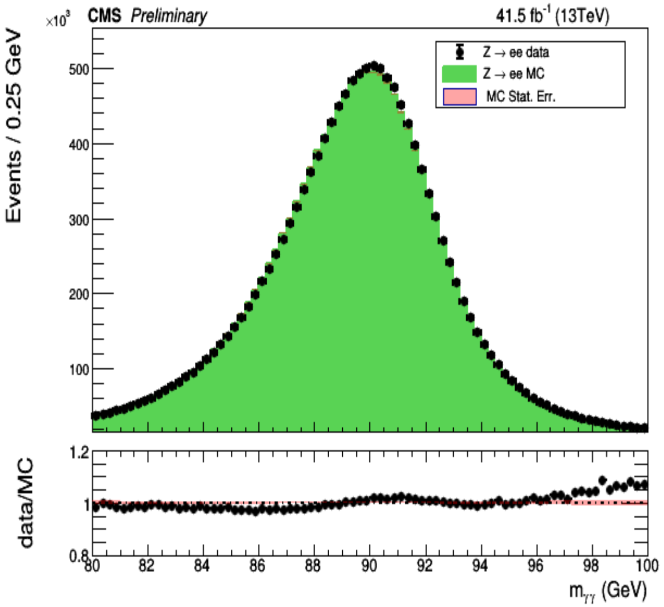
➢ Validations with Data/MC events in mass side-band

- UL2017 "DoubleEG" data
- MC: MC pp plus data-driven pf+ff
- Events passing H→γγ preselection



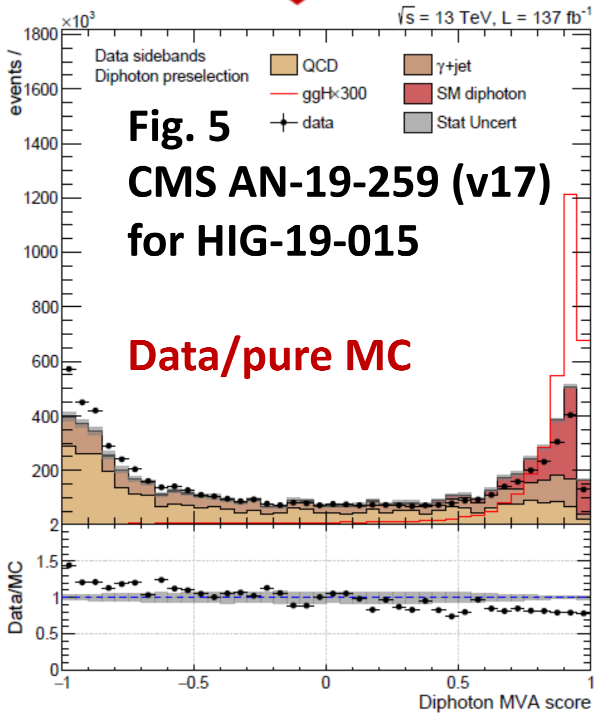
➢ Validations with Z→ee events

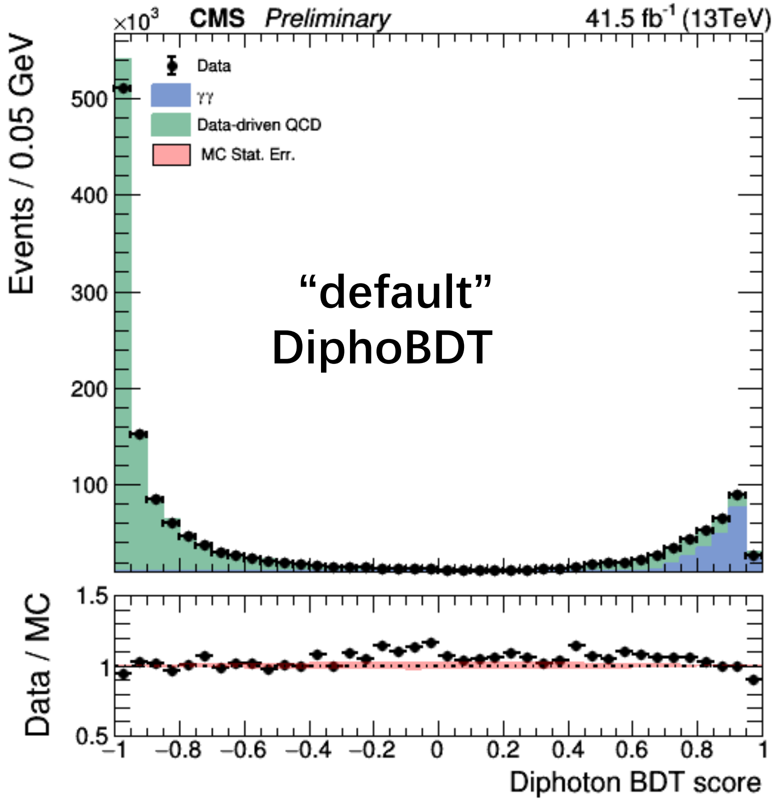
- Samples: UL2017 "SingleElectron" data triggered with "HLT_Ele32_WPTight_Gsf*", and DY MC
- Events passing H→γγ preselection, but with inverted electron veto requirement
- Mass of diphoton candidates: [80,100] GeV

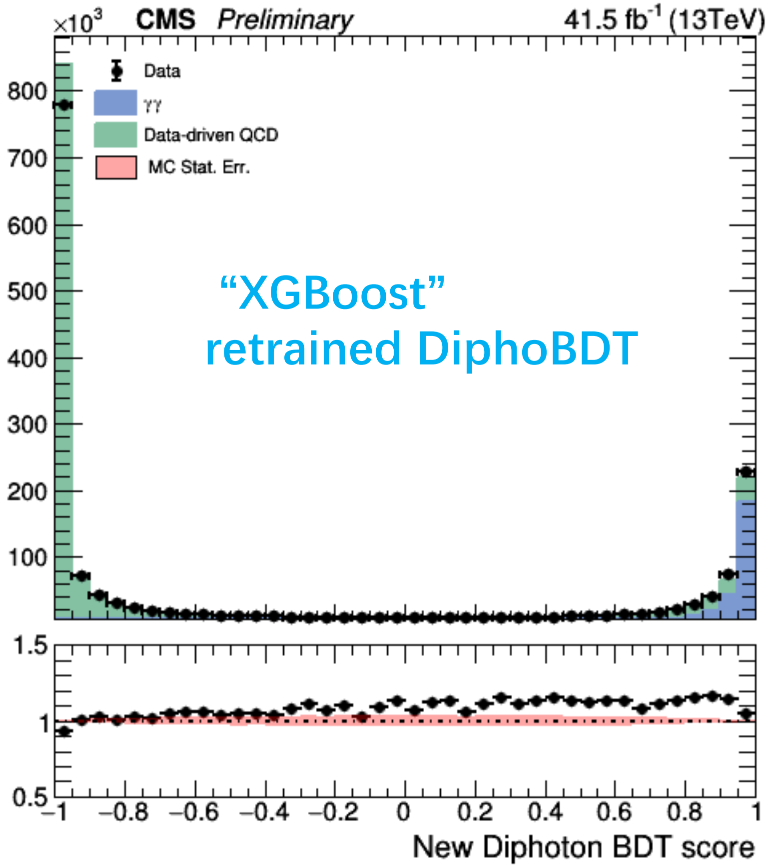


Data/MC Validations UL17

- ➢ **Good data/MC agreement**
- ➢ **Agreements better than the data/pure MC comparison**



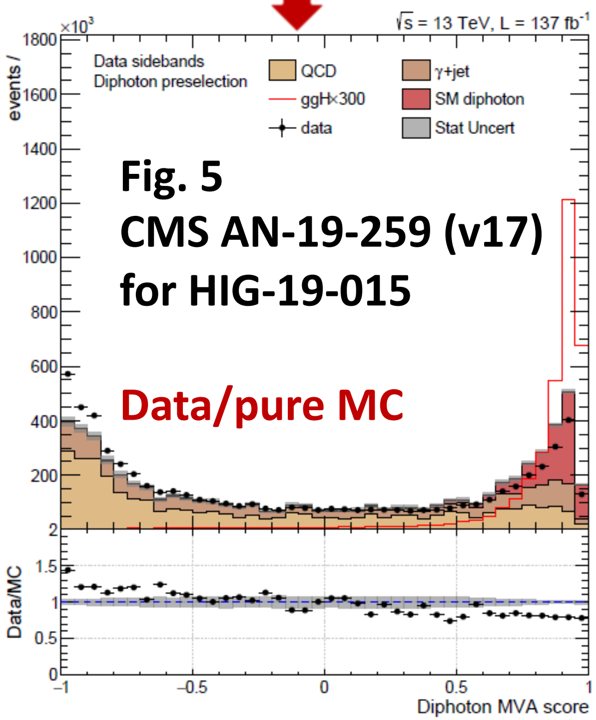




Events in mass side-band
MC: MC pp plus data-driven pf+ff

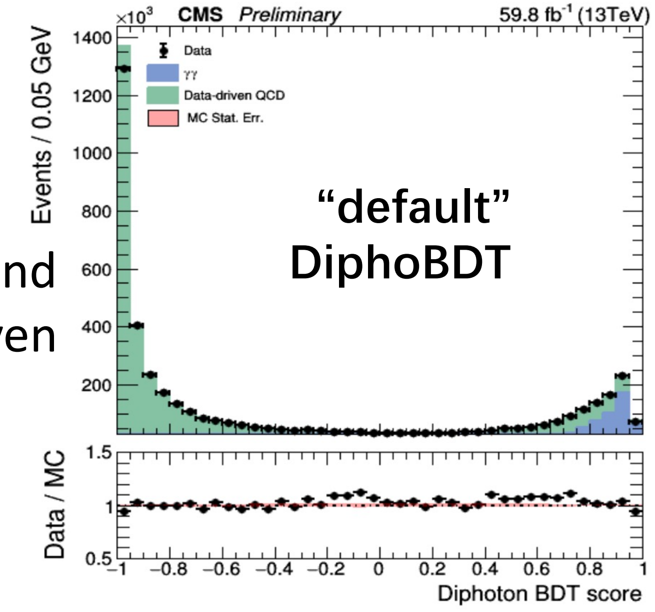
Data/MC Validations

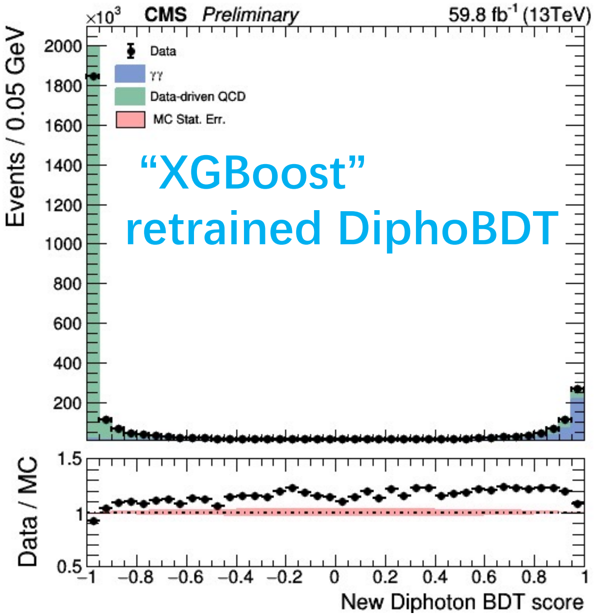
- **Good data/MC agreement**
- **Agreements better than the data/pure MC comparison**

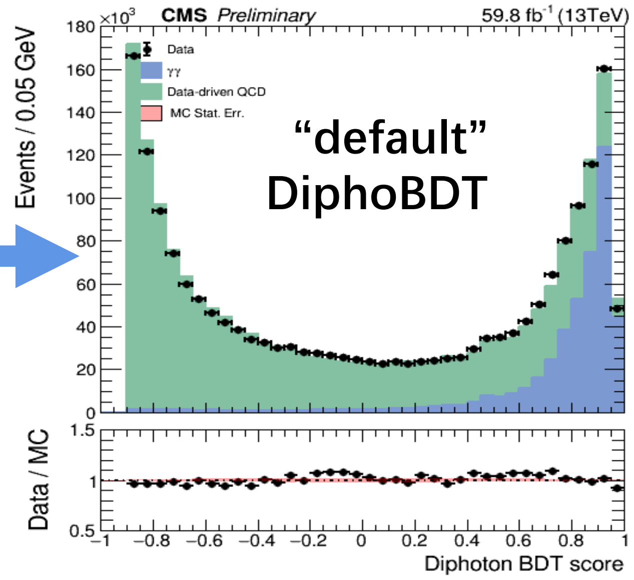


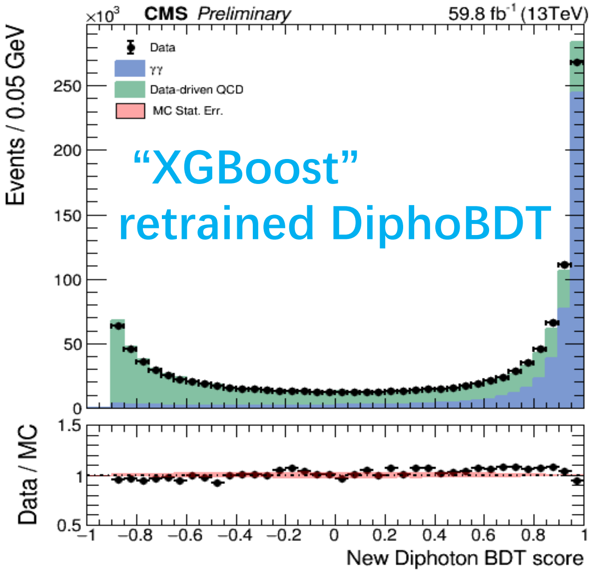
Events in mass side-band MC: MC pp + data-driven pf+ff

A looser cut with Diphoton BDT > -0.9 to reject lots of bkg events then redo the MC normalization according to Ndata





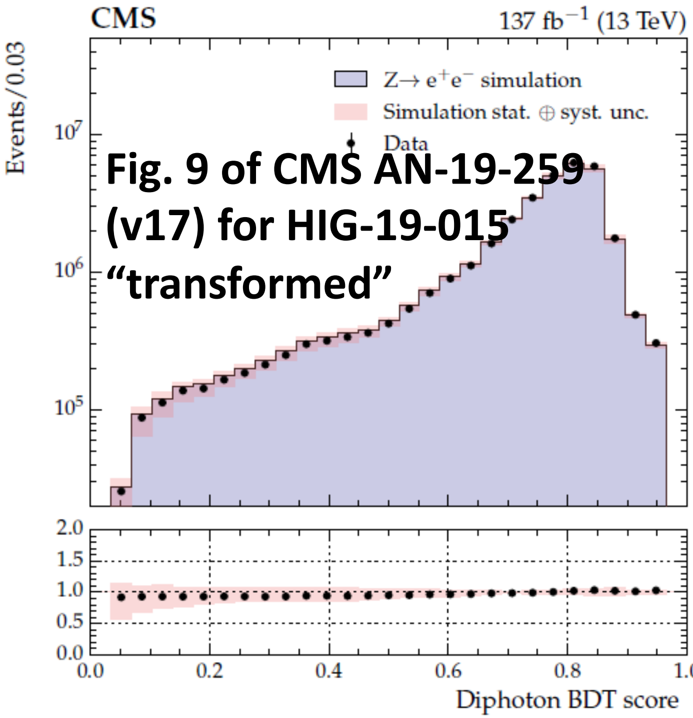




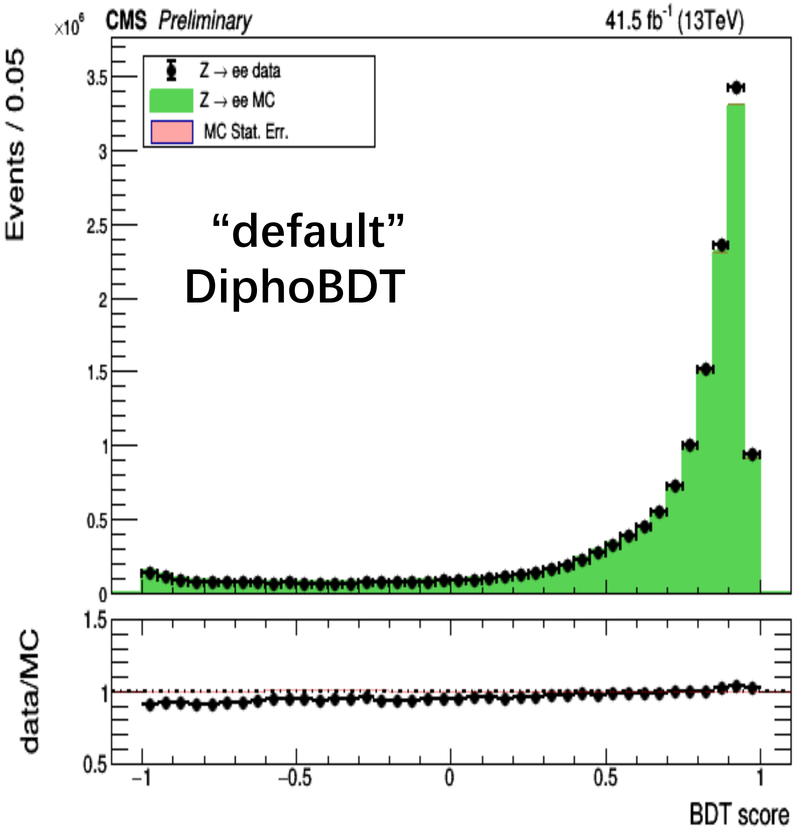
Validations with Z→ee UL17

- **Good data/MC agreement**

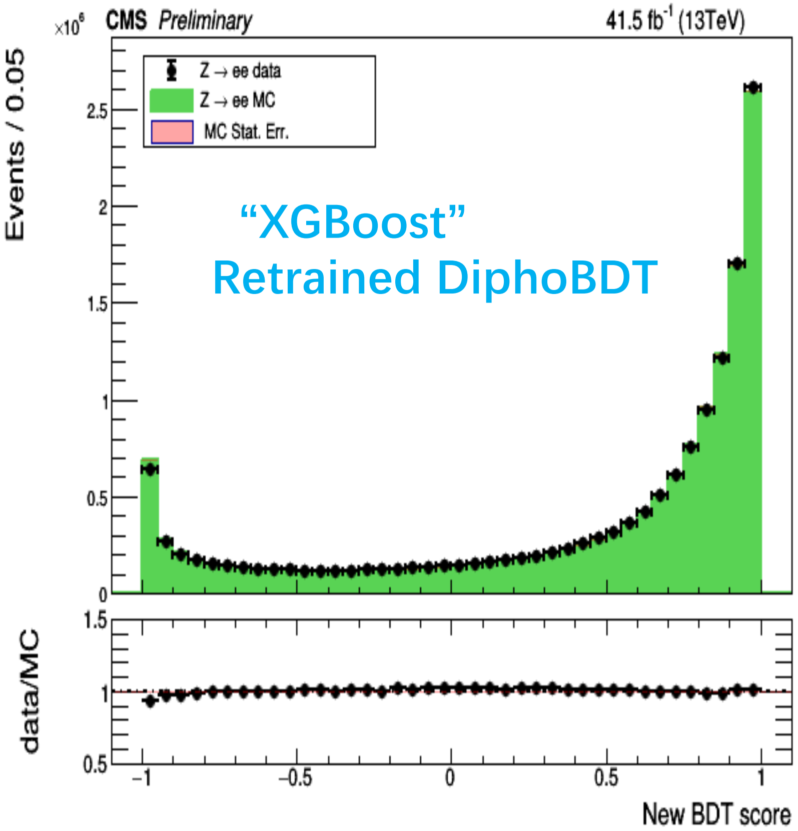
Fig. 9 of CMS AN-19-259 (v17) for HIG-19-015 "transformed"



"default" DiphoBDT



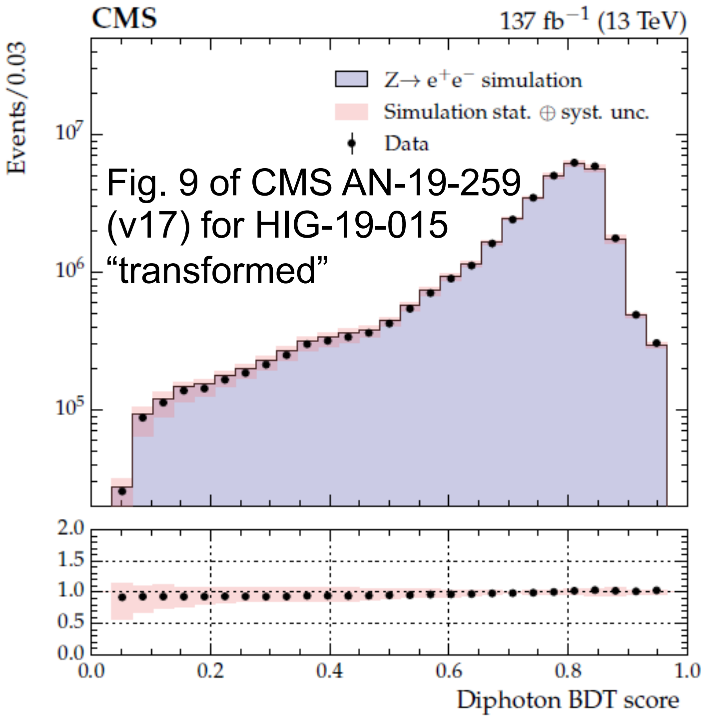
"XGBoost" Retrained DiphoBDT

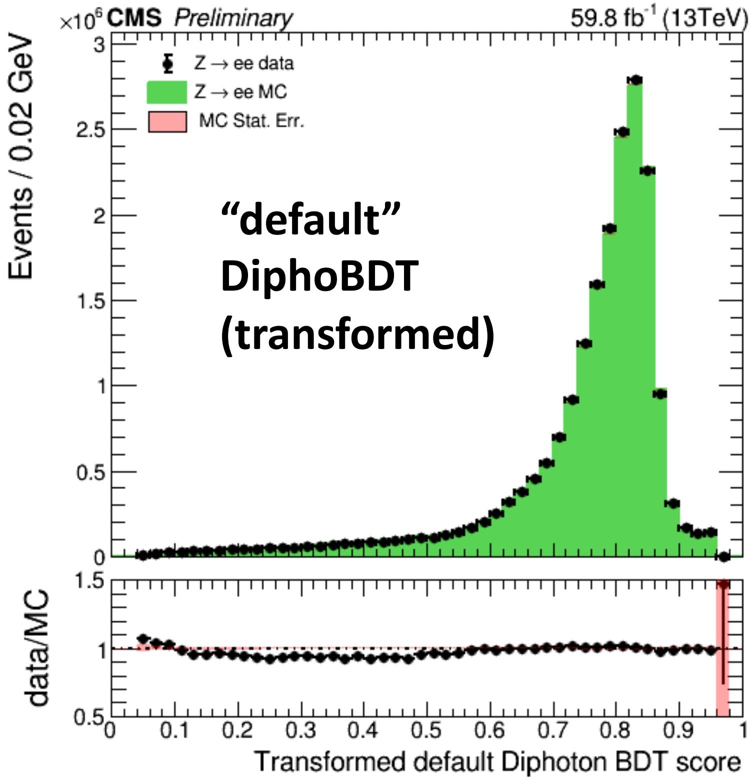


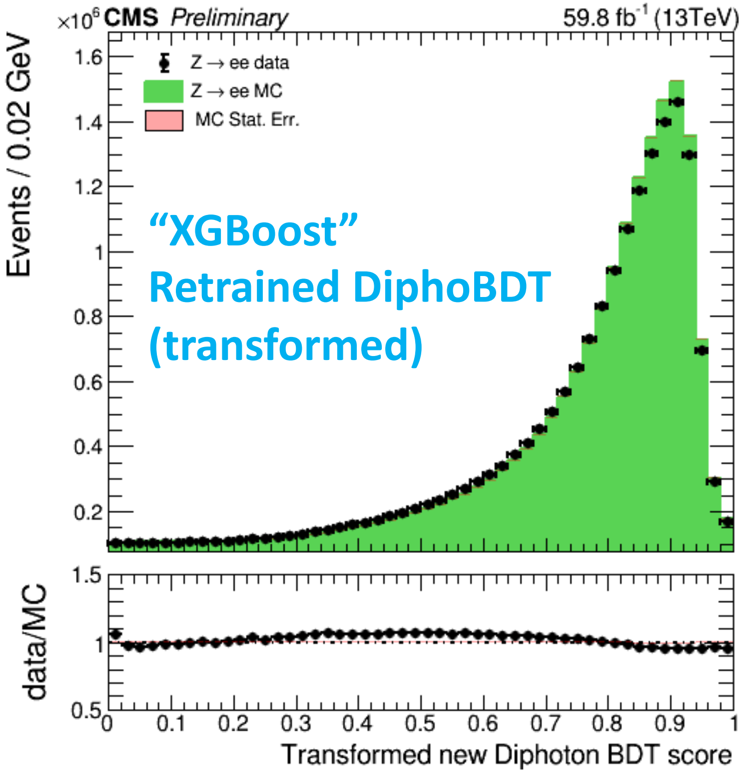
Not yet include the **systematics** from photon ID MVA and photon energy resolution σ_E/E

Validations of the inputs with UL18 Z→ee events (cont.)

- **Good data/MC agreement**





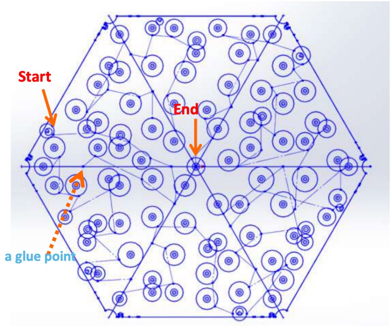


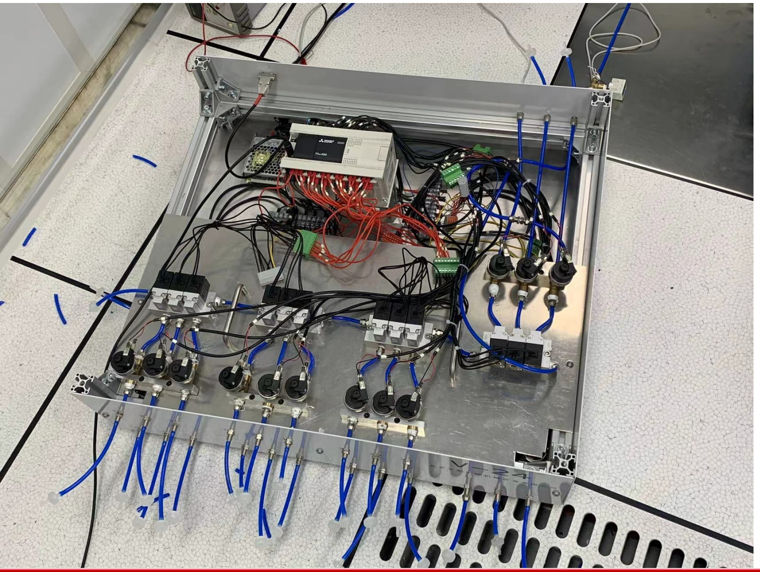
Not yet include the **systematics** from photon ID MVA and photon energy resolution σ_E/E

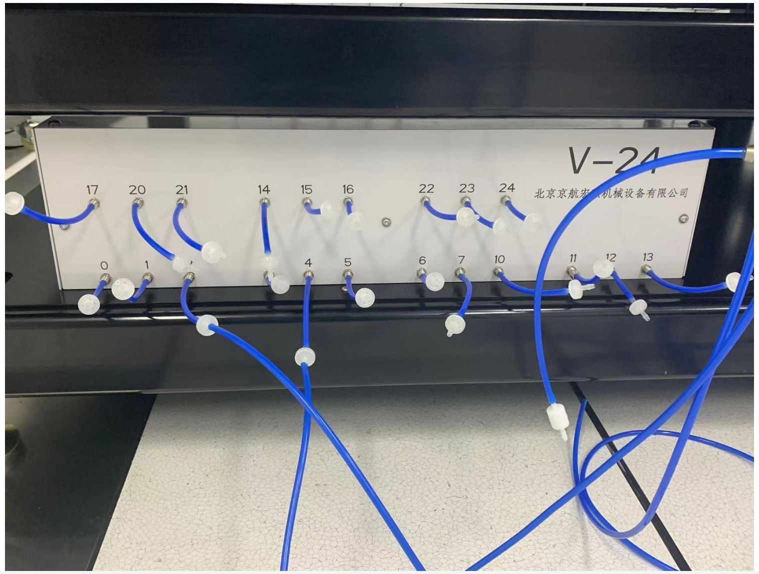
Hardware work

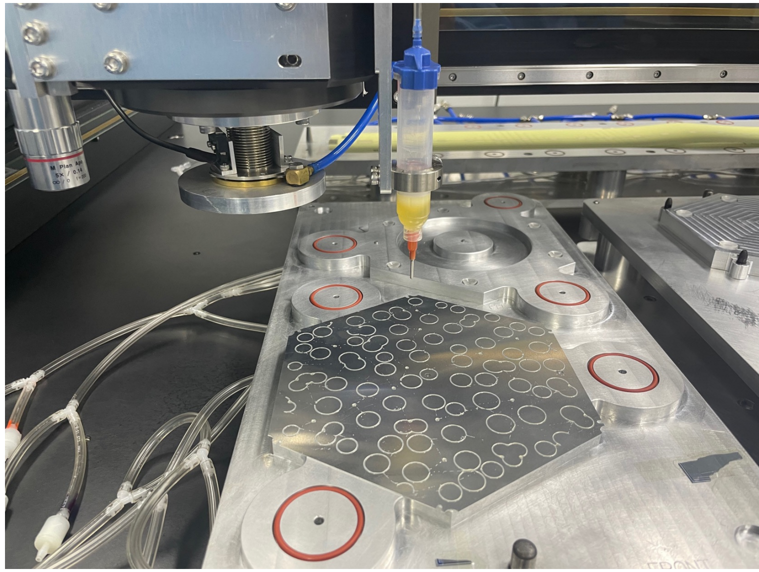
HGCAL gantry works:

- **Vacuum system update:**
  - In order to satisfy the official requirement of multiple assemble task, and out of the consideration of the stability and reliability of the vacuum system, we improve our vacuum system by:
    - Add more pipeline so that we can control more channel at the same time to meet the need of multiple assembling task
    - Quantify the rate of vacuum by using vacuum detector and check pipeline vacuum one by one to make sure the vacuum system work fine
    - Package the whole vacuum system to a box to avoid of accidentally touch
- **Glue system update**
  - Since the module patter have some update, we need to update our glue pattern too
  - Add a new function to add a glue point in the module
  - Check the new glue pattern and it works fine









Summary and plans

Summary

- **Resonant $HH \rightarrow WW\gamma\gamma$:**
  - Transform object selection, event selection, gen selection to HiggsDNA framework
  - Have a preliminary limit result
- **$H \rightarrow \gamma\gamma$ mass measurement with full Run2 dataset:**
  - Retrain diphoton BDT for UL17 & UL18 and presented in CMS Hgg working meeting (07/07/2022 & 28/07/2022)
  - Improve the significance by 5% & 8% for UL17 & UL18
- **Hardware works:**
  - Update vacuum system
  - Update glue system

Plans:

- Resonant $HH \rightarrow WW\gamma\gamma$:
  - follow the previous result to make a complete version and have a presentation in Higgs meeting

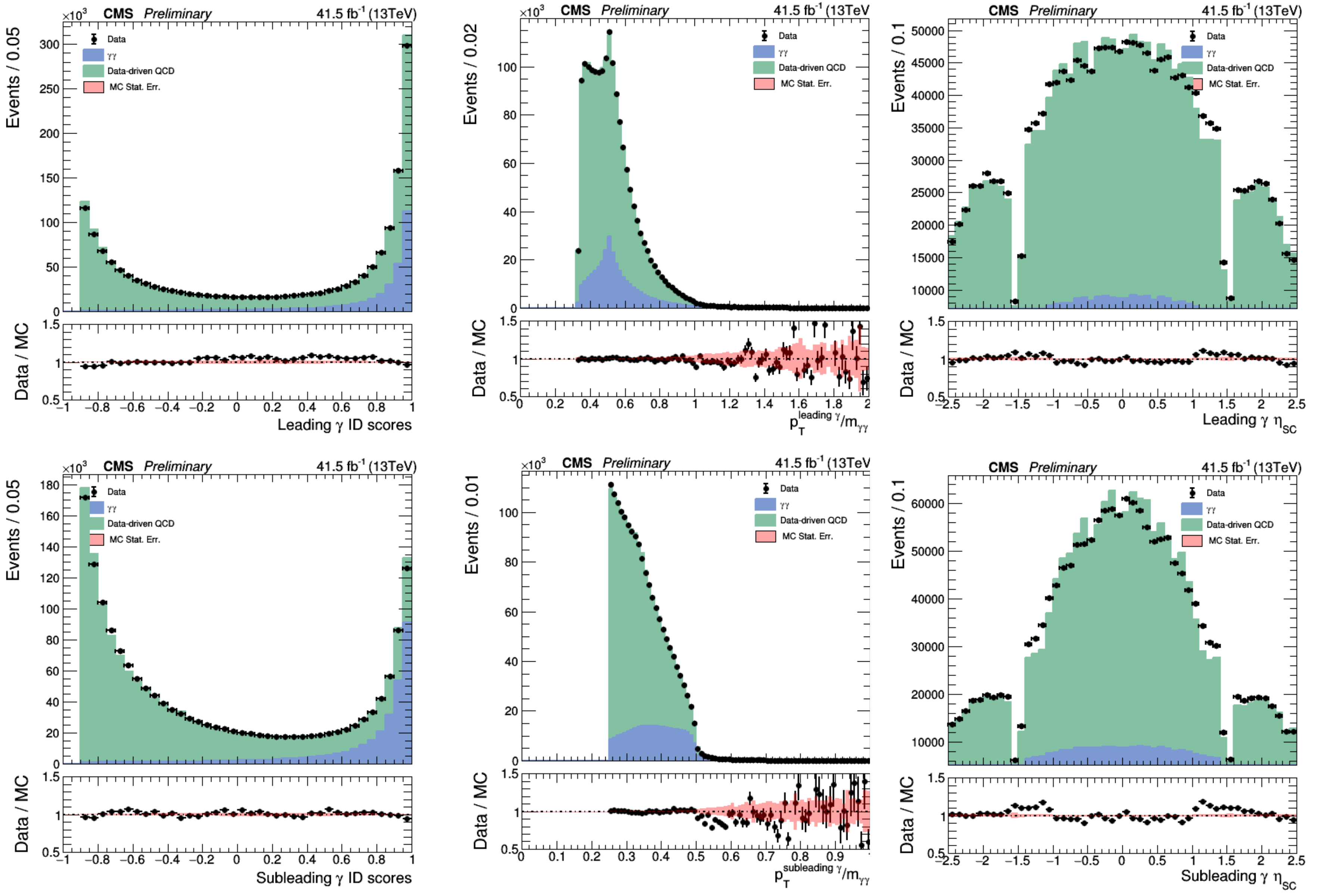
- $H \rightarrow \gamma\gamma$ mass measurement with full Run2 dataset:
  - retrain diphoton BDT for UL2016 preVFP and postVFP, with the same strategy (data-driven QCD and XGboost)
  - focus more on the systematic part

- Hardware part:
  - ready for the assemble works
  - focus more on teaching new students

BACK UP

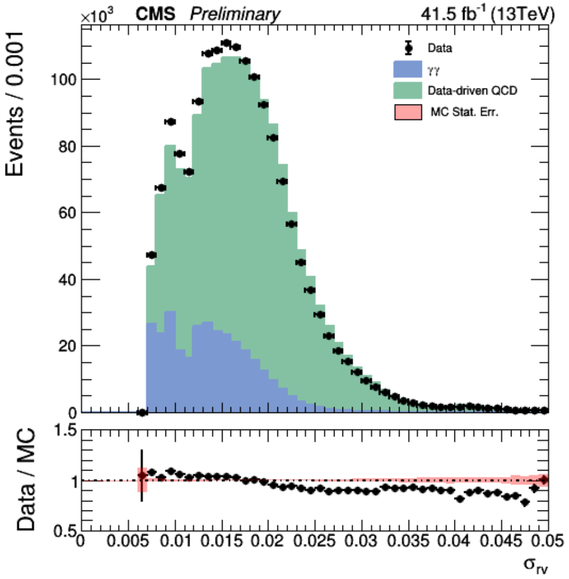
Data/MC comparisons of the 10 input features UL17

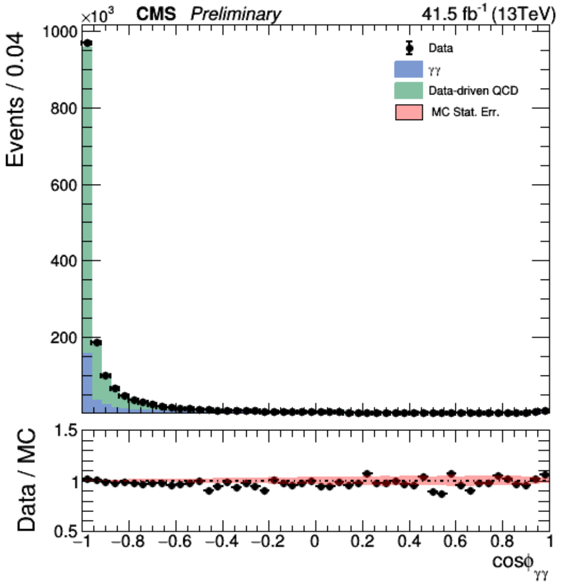
- ➢ Events in mass side-band, after H→γγ preselection
- ➢ MC: MC pp plus data-driven pf+ff
- ➢ **Good data/MC agreement**

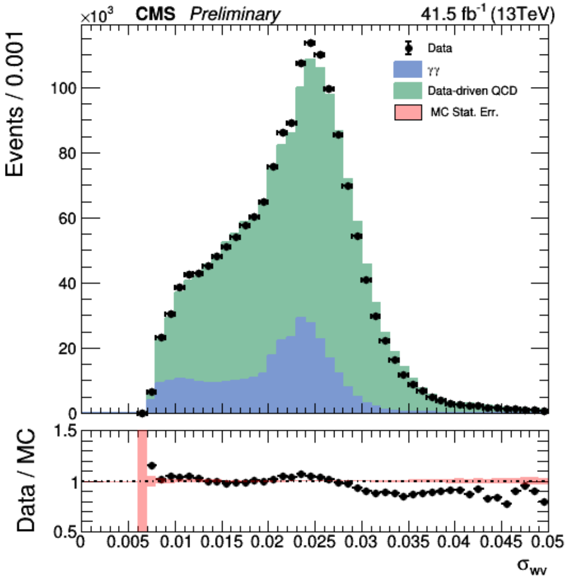


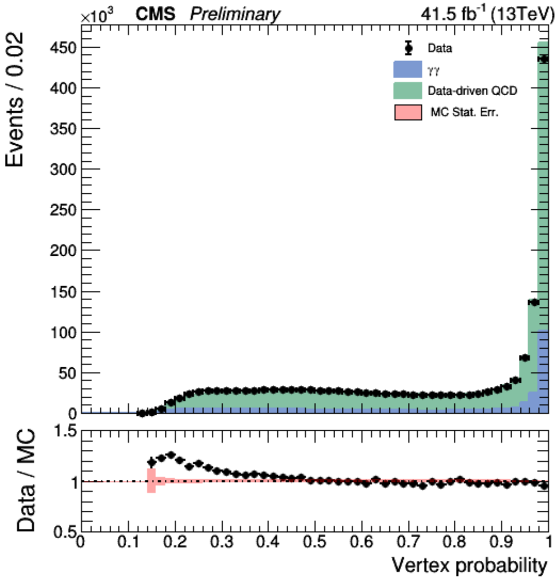
Data/MC comparisons of the 10 input features (UL17 cont.)

- Events in mass side-band, after H→γγ preselection
- MC: MC pp plus data-driven pf+ff
- **Good data/MC agreement**



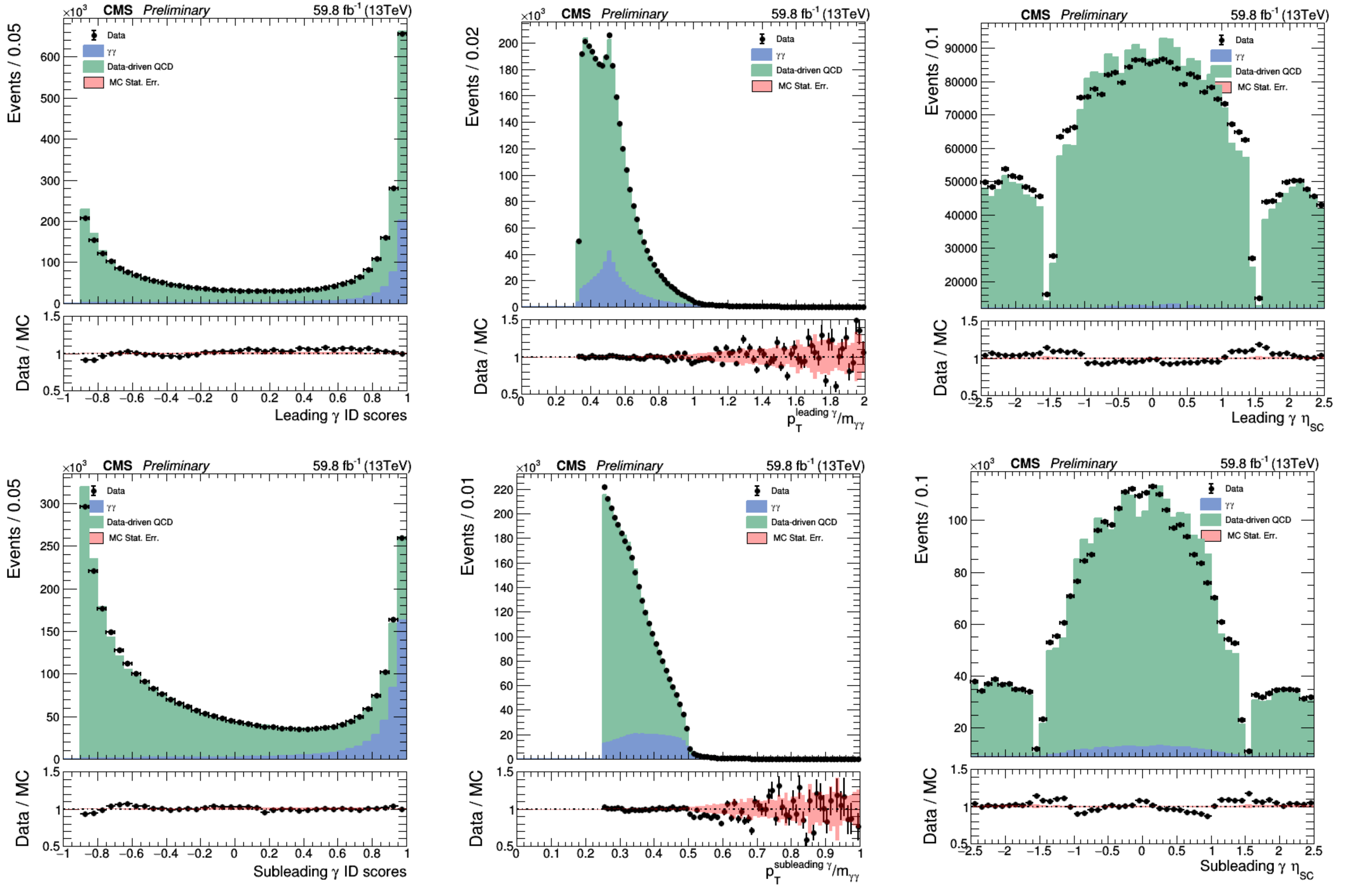






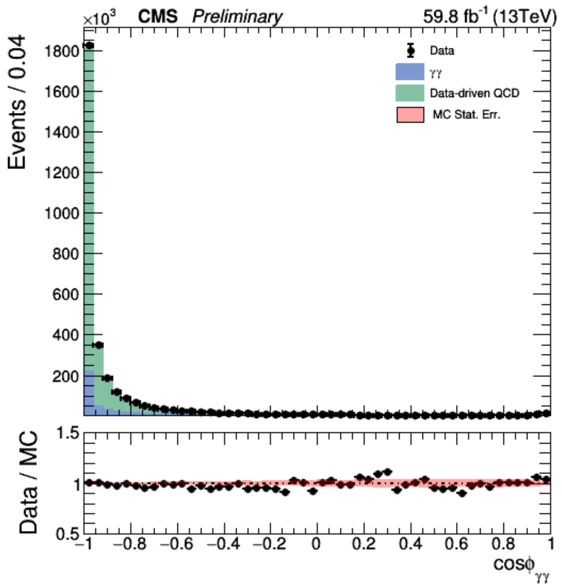
Data/MC comparisons of the 10 input features UL18

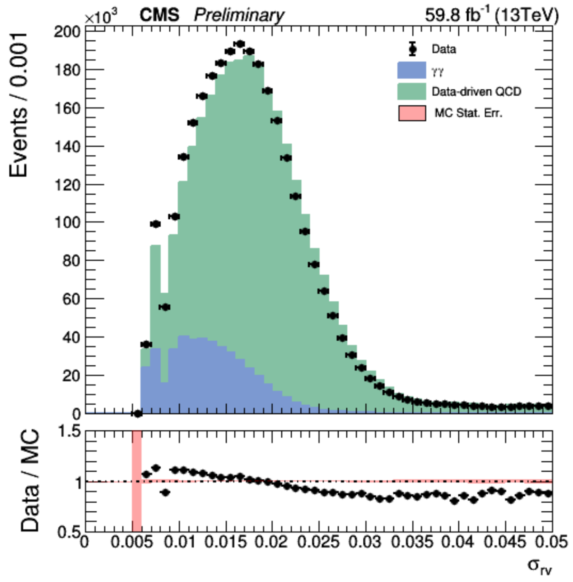
- ➢ Events in mass side-band, after H→γγ preselection
- ➢ MC: MC pp plus data-driven pf+ff
- ➢ **Good data/MC agreement**

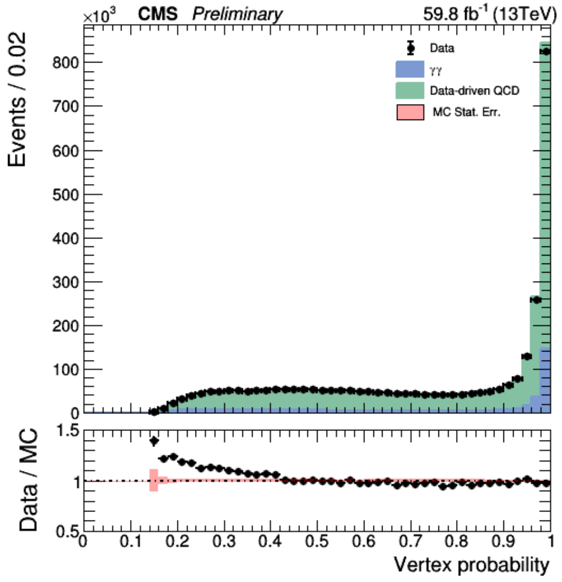


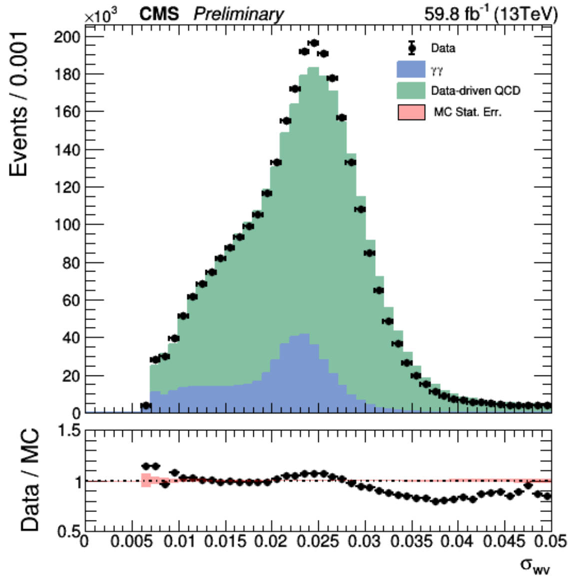
Data/MC comparisons of the 10 input features (UL18 cont.)

- Events in mass side-band, after H→γγ preselection
- MC: MC pp plus data-driven pf+ff
- **Good data/MC agreement**



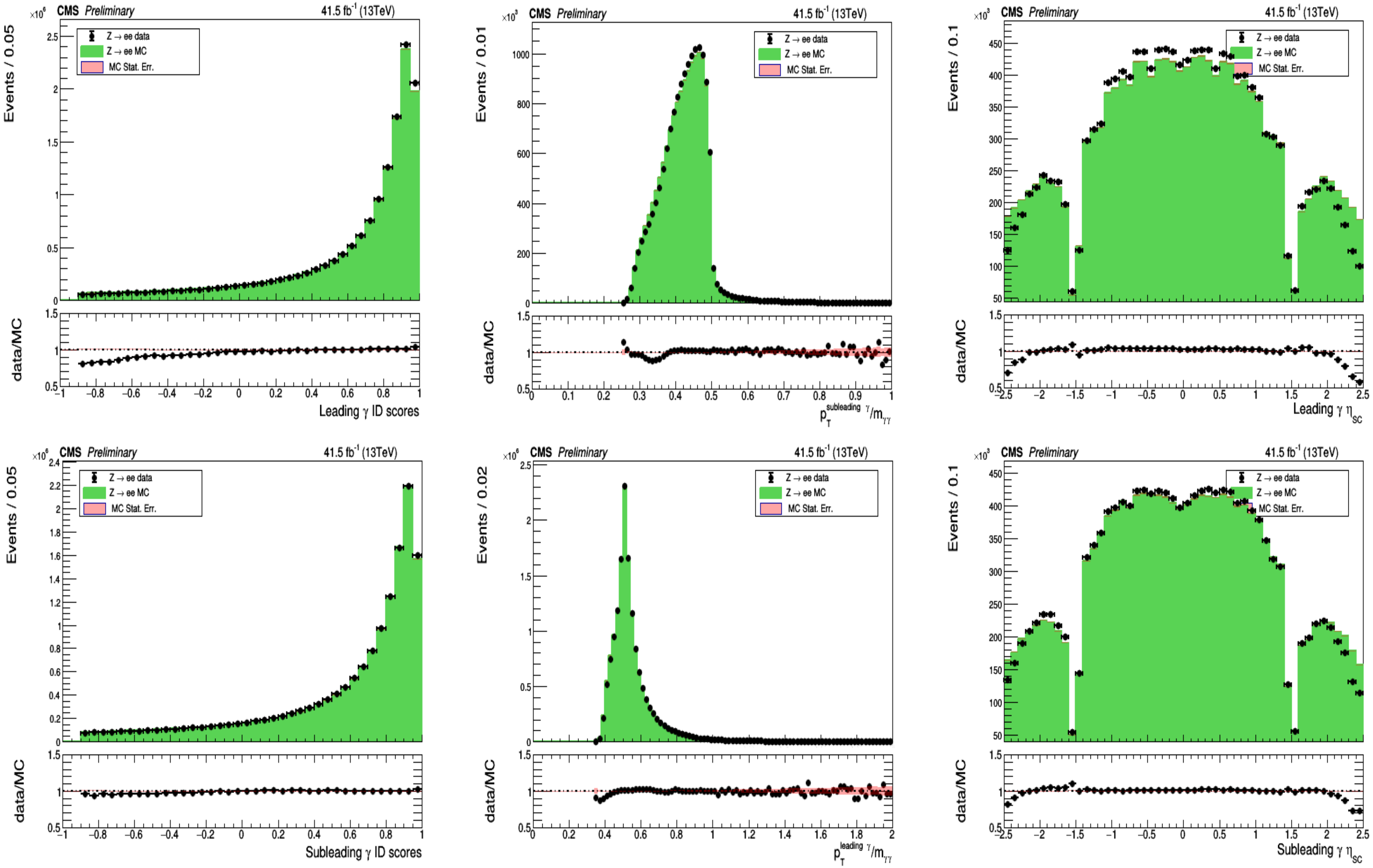






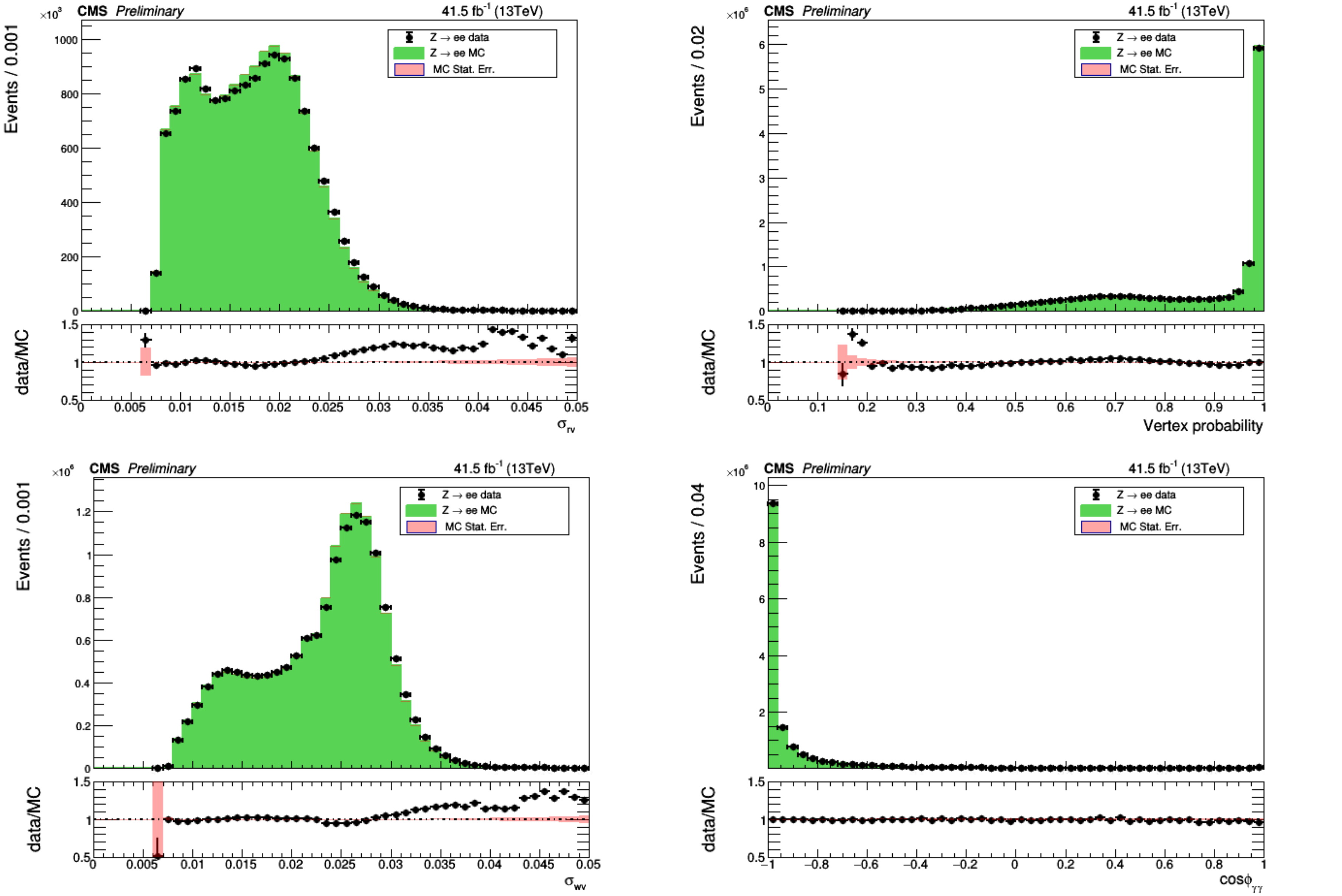
Validations of the inputs with Z→ee UL17

- **Good data/MC agreement**



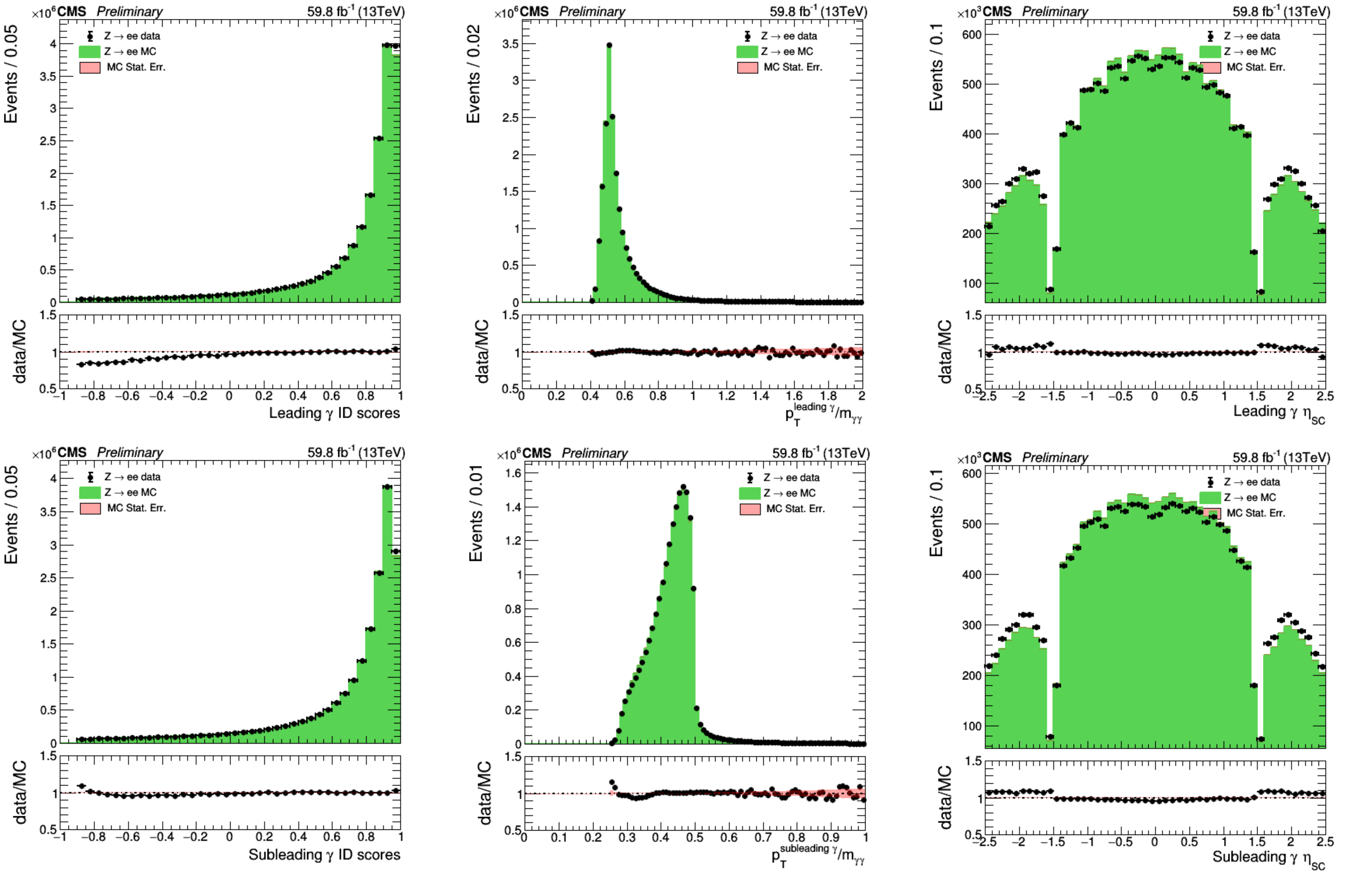
Validations of the inputs with Z→ee UL17 (cont.)

➢ **Good data/MC agreement**



Validations of the inputs with Z→ee UL18

- **Good data/MC agreement**



Validations of the inputs with Z→ee UL18 (cont.)

- **Good data/MC agreement**