Scouting Data

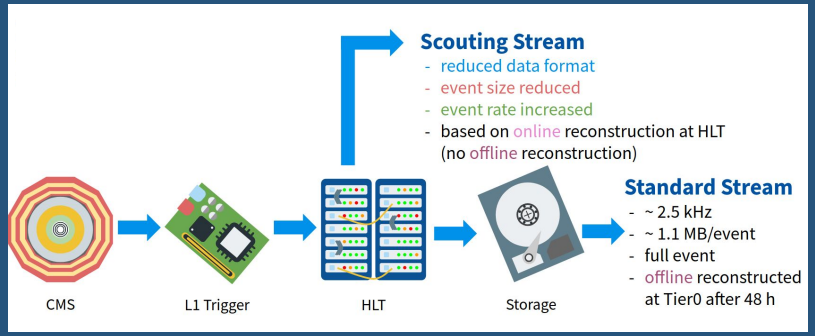
DY-TMD Workshop

Eliott Ducarme

Scouting Intro

As a reminder :

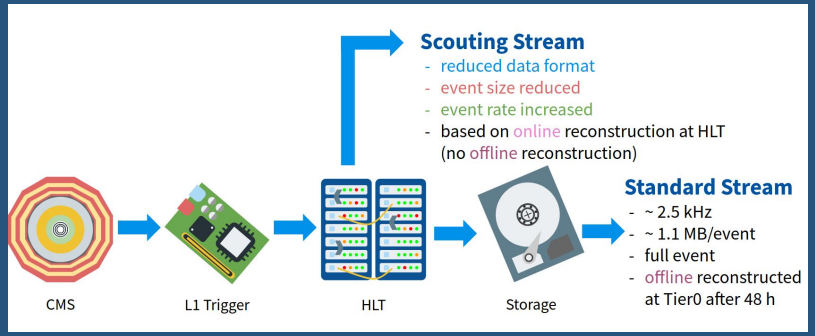
- Scouting is a high rate, small size format



Scouting Intro

As a reminder :

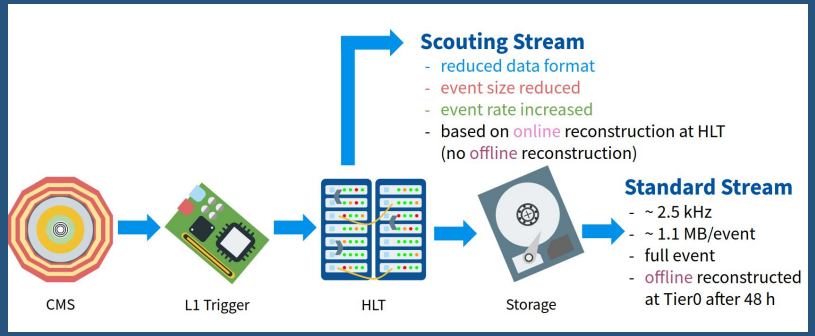
- Scouting is a high rate, small size format
  - Rate is ~25 times higher than the regular stream of data taking
    - O(10^{10}) events in 2024



Scouting Intro

As a reminder :

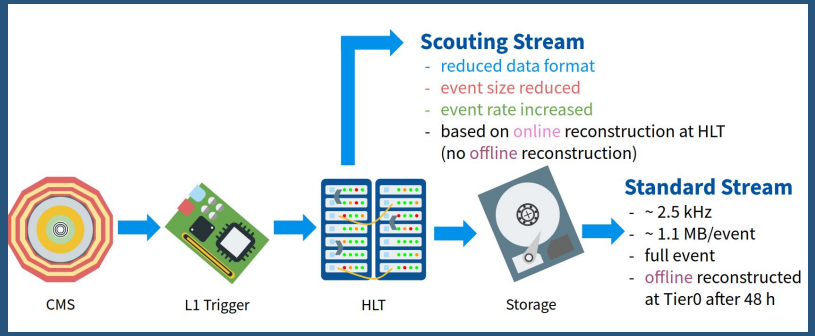
- Scouting is a high rate, small size format
  - Rate is ~10 times higher than the regular stream of data taking
  - Threshold is much lower
    - HLT triggers at muon of $p_T > 3$ GeV/c
    - Unprescaled L1s start $p_T > 4.5$ GeV/c



Scouting : What is it ?

As a reminder :

- Scouting is a high rate, small size format
  - Rate is ~25 times higher than the regular stream of data taking
  - Threshold is much lower
  - Trade-off is the lower information and no offline reconstruction
    - Not storing all objects information
    - Decreased numerical precision on tracks and PF candidates.



Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
- Electrons
- Photons

Scouting : What is saved ?

- PF Candidates
  - Reconstructed by standard PF algorithm
  - $p_T > 0.6$GeV & $|\eta| < 3$
  - Reclustering of Jets
    - AK4 CHS jets
    - AK8 fat jets
- PF jets
- Tracks and Vertices
- Muons
- Electrons
- Photons

Scouting : What is saved ?

- PF Candidates
- PF jets
  - $p_T > 20$ GeV, $|\eta| < 5$
  - Includes content in particles' energies and multiplicities
- Tracks and Vertices
- Muons
- Electrons
- Photons

Scouting : What is saved ?

- PF Candidates
- PF jets
  - $p_T > 20$ GeV, $|\eta| < 5$
  - Includes content in particles' energies and multiplicities
- Tracks and Vertices
- Muons
- Electrons
- Photons

from TriggerTutorial_Scouting_February2025

- **HLT Filters**: None
- **HLT corrections**: Available in the scouting path but uncorrected jets are stored

➢ Does this mean that any jet present in the event is registered ?

Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
  - Standard & PF tracking
  - $p_T > 1$ GeV
  - pixel track information
  - covariance matrix of track fit
- Muons
- Electrons
- Photons

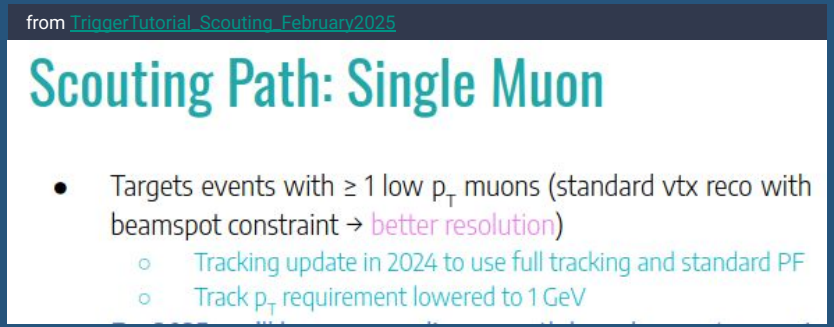
Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
  - Double : NoVtx, p_T > 3 GeV
  - Single : Vtx, No p_T threshold
- Electrons
- Photons

Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
  - Double : NoVtx, p_T > 3 GeV
  - Single : Vtx, No p_T threshold
- Electrons
- Photons

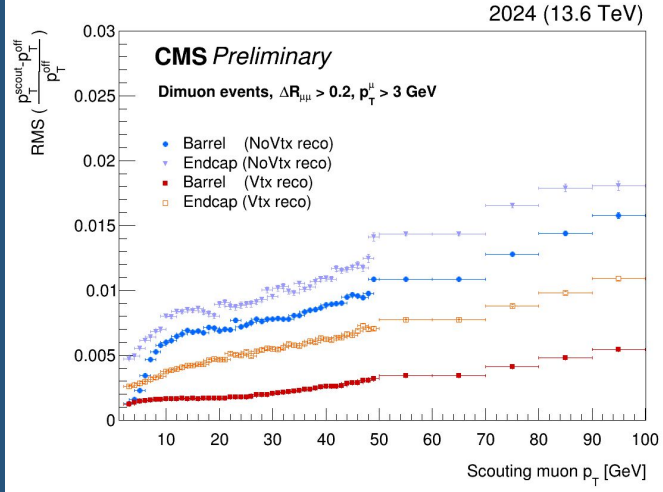
from TriggerTutorial_Scouting_February2025



Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
  - Double : NoVtx, p_T > 3 GeV
  - Single : Vtx, No p_T threshold
- Electrons
- Photons

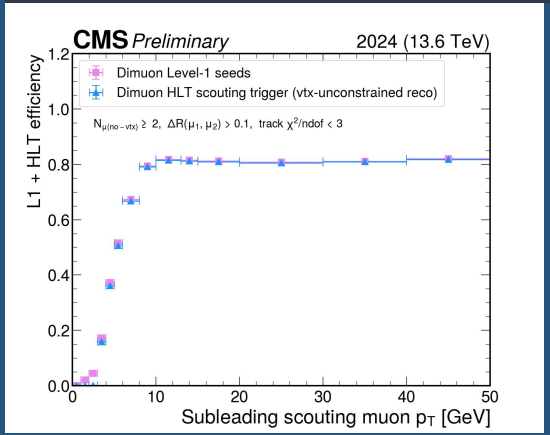
Resolution comparison of Vtx and NoVtx muons

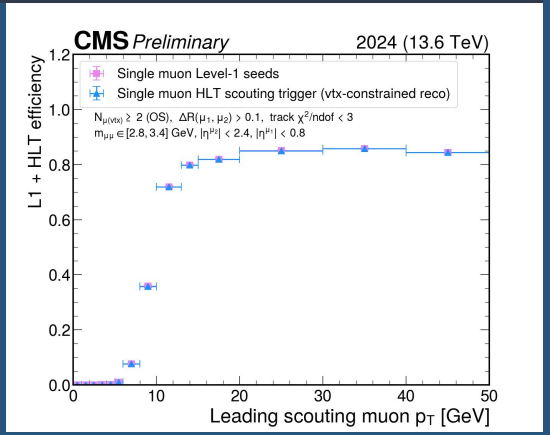


Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
  - Double : NoVtx, p_T > 3 GeV
  - Single : Vtx, No p_T threshold
- Electrons
- Photons

Efficiencies of HLT/L1-triggers for Single and Double stream, from [here]





Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
  - Double : NoVtx, p_T > 3 GeV
  - Single : Vtx, No p_T threshold
- Electrons
- Photons

Scouting : What is saved ?

- PF Candidates
- PF jets
- Tracks and Vertices
- Muons
- E/γ
  - Double : E_T > 12 GeV, SC H/E < 0.2
  - Single : E_T > 30 GeV , SC H/E < 0.2

Run3 2024 : Datasets

- In 2024, there is a total of 109.08 /fb of data in GoldenJSON.
  - ➢ We start with era I of 2024, thus 11.417 /fb
    - ■ Actually 1 run with 0.839 /fb in the plots here, even if marked otherwise

ERA	Delivered by LHC [/fb]	Recorded by CMS [/fb]	Golden JSON [/fb]
B	0.74	0.66	0.13
C	7.95	7.43	7.24
D	8.90	8.30	7.96
E	12.24	11.43	11.32
F	30.47	28.12	27.76
G	42.84	39.40	37.77
H	6.81	6.20	5.44
I	12.85	11.79	11.47
Total	122.81	113.32	109.08

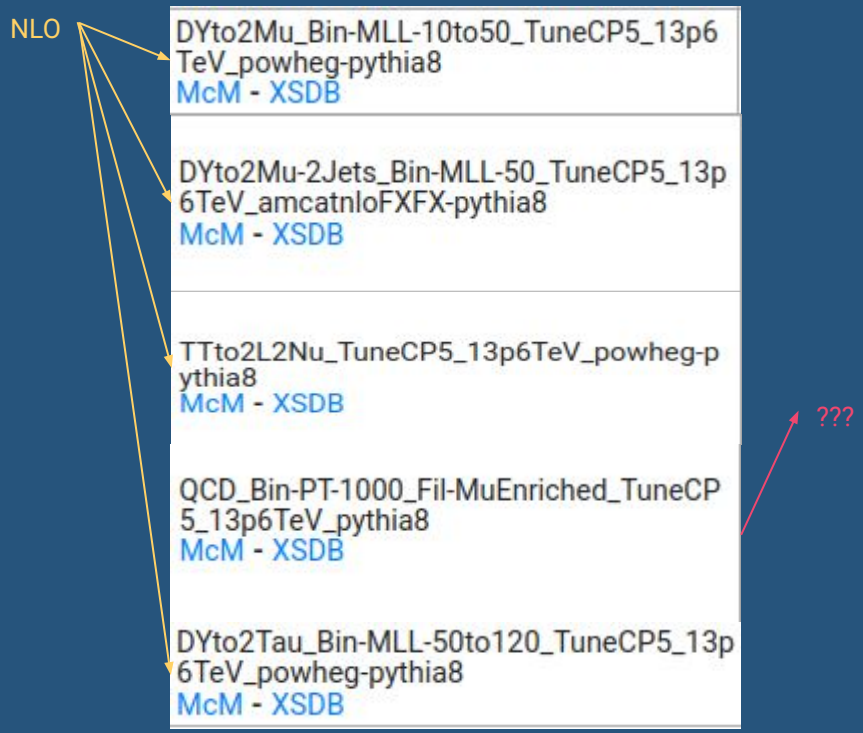
Run3 2024 : Datasets

- In 2024, there is a total of 109.08 fb^{-1} of data in GoldenJSON.
- The following MC samples are used
  - Signal :
    - DYto2Mu
  - Background :
    - QCD, MuEnriched
    - TTto2L2Nu
    - DYto2Tau

Sample	Bin	Xsec
DYto2Mu_Bin-MLL-10to50_TuneCP5_13p6TeV_powheg-pythia8	10to50	6744.0
DYto2Mu-2Jets_Bin-MLL-50_TuneCP5_13p6TeV_amcatnloFXFX-pythia8	50	2124.08
DYto2Tau_Bin-MLL-50to120_TuneCP5_13p6TeV_powheg-pythia8	50to120	2219.0
TTto2L2Nu_TuneCP5_13p6TeV_powheg-pythia8	Incl	98.04
QCD_Bin-PT-*_Fil-MuEnriched_TuneCP5_13p6TeV_pythia8	15to20	2982000.0
	20to30	2679000.0
	30to50	1465000.0
	50to80	409500.0
	80to120	96200.0
	120to170	22980.0
	170to300	7754.0
	300to470	699.1
	470to600	68.24
	600to800	21.37
	800to1000	3.913
	1000	1.323

Run3 2024 : First look

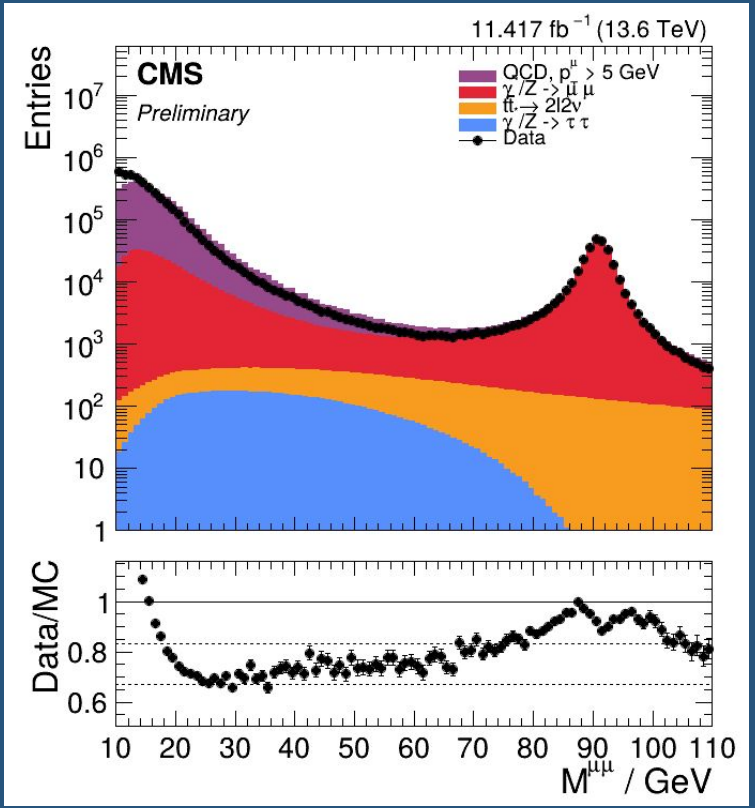
- In 2024, there is a total of 109.08 fb^{-1} of data in GoldenJSON.
- The following MC samples are used
  - Signal :
    - DYto2Mu
  - Background :
    - QCD, MuEnriched
    - TTto2L2Nu
    - DYto2Tau



Run3 2024 : First look

- In 2024, there is a total of 109.08 /fb of data in GoldenJSON.
- The following MC samples are used
  - Signal :
    - DYto2Mu
  - Background :
    - QCD, MuEnriched
    - TTto2L2Nu
    - DYto2Tau
- These cuts are applied as a base, if not stated otherwise :

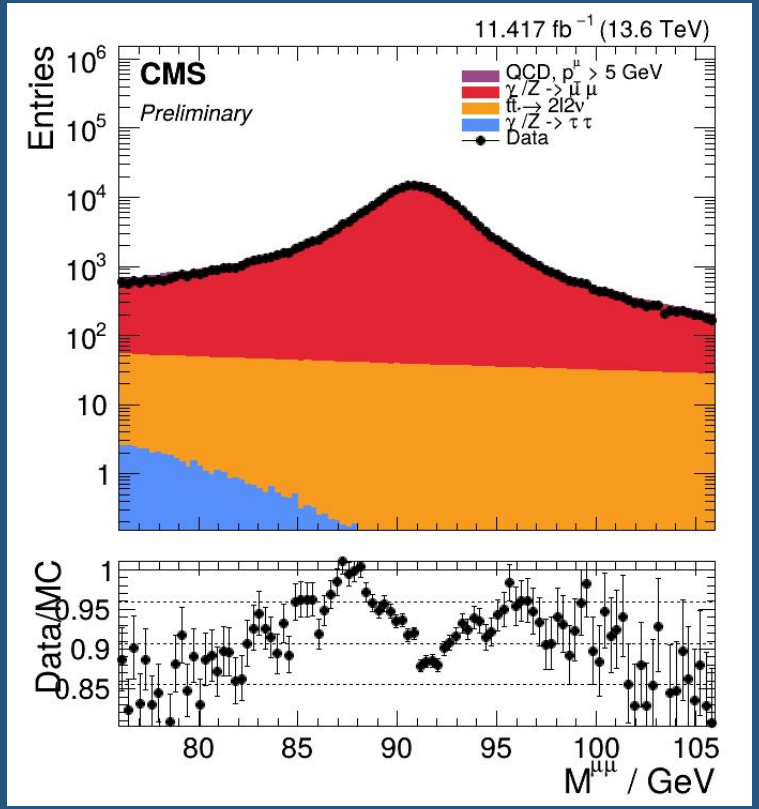
```
pt_mu > 5 GeV/c, |η| < 2.4
relativeTrackerIsolation < 0.03
trk_chi2/tkr_ndof < 10
L1_DoubleMu4p5er2p0_OS_SQ_MassMin7
DST_PFScouting_DoubleMuon
sub_pt/mass_pait < 0.25
lead_pt/mass_pair < 0.45
```



Run3 2024 : First look

- In 2024, there is a total of 109.08 /fb of data in GoldenJSON.
- The following MC samples are used
  - Signal :
    - DYto2Mu
  - Background :
    - QCD, MuEnriched
    - TTto2L2Nu
    - DYto2Tau
- These cuts are applied as a base, if not stated otherwise :

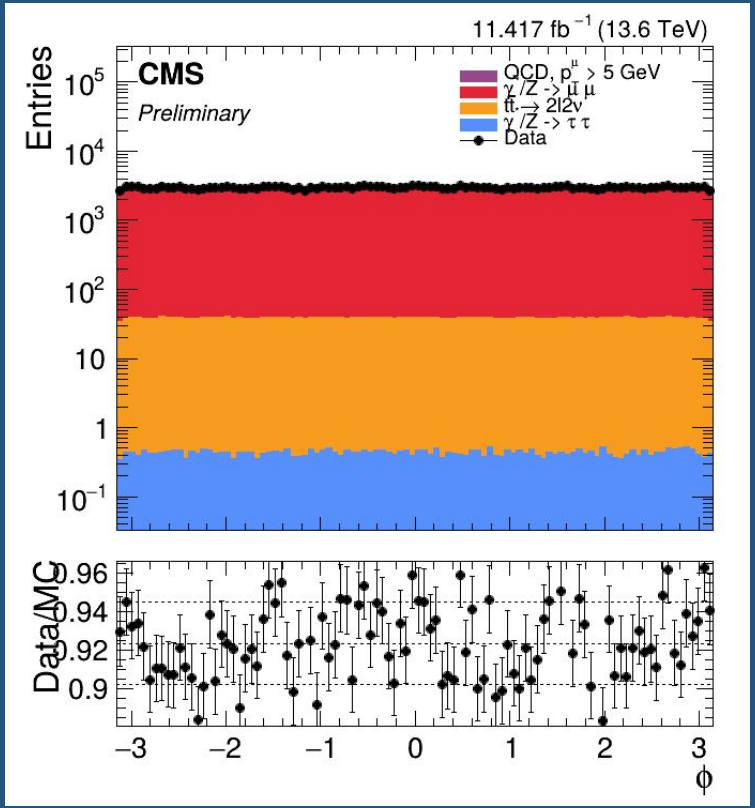
```
pt_mu > 5 GeV/c, |η| < 2.4
relativeTrackerIsolation < 0.03
trk_chi2/tkr_ndof < 10
L1_DoubleMu4p5er2p0_OS_SQ_MassMin7
DST_PFScouting_DoubleMuon
sub_pt/mass_pait < 0.25
lead_pt/mass_pair < 0.45
```



Run3 2024 : First look

- In 2024, there is a total of 109.08 /fb of data in GoldenJSON.
- The following MC samples are used
  - Signal :
    - DYto2Mu
  - Background :
    - QCD, MuEnriched
    - TTto2L2Nu
    - DYto2Tau
- These cuts are applied as a base, if not stated otherwise :

```
pt_mu > 5 GeV/c, |η| < 2.4
relativeTrackerIsolation < 0.03
trk_chi2/tkr_ndof < 10
L1_DoubleMu4p5er2p0_OS_SQ_MassMin7
DST_PFScouting_DoubleMuon
sub_pt/mass_pait < 0.25
lead_pt/mass_pair < 0.45
```

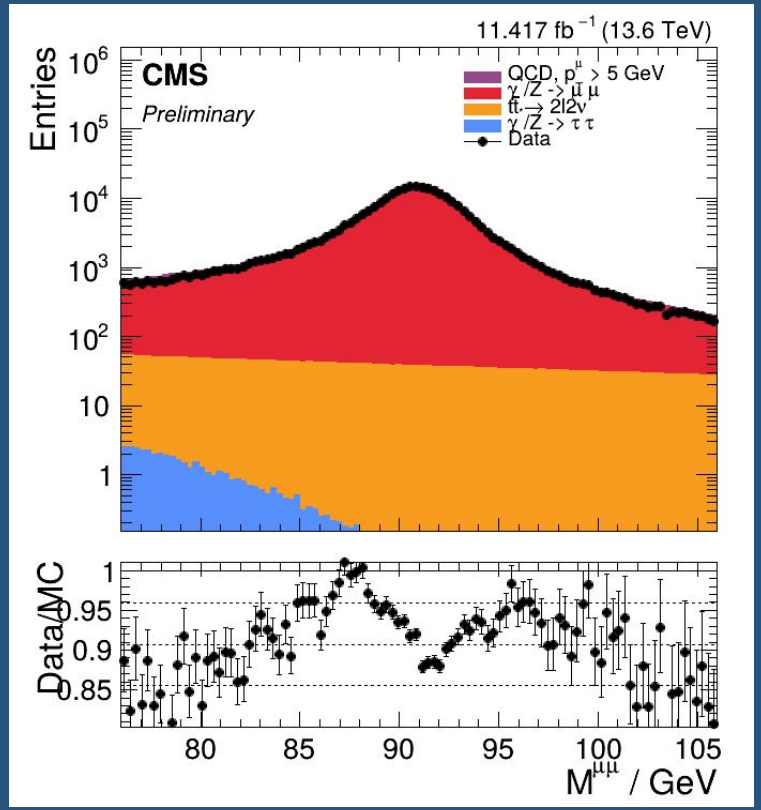


Run3 2024 : First look

- Need to compute efficiency using other HLT trigger at Z-peak
  - Take HLT_IsoMu24 events with 2 muons of p_T > 30 GeV
  - Look into these the fraction that trigger HLT/L1 that we use

$$\varepsilon = \frac{HLT\_IsoMu24 + 2\mu + p_T^{1,2} > 30GeV + HLT/L1}{HLT\_IsoMu24 + 2\mu + p_T^{1,2} > 30GeV\Gamma}$$

- This should show us if there is an issue with the triggers we are using

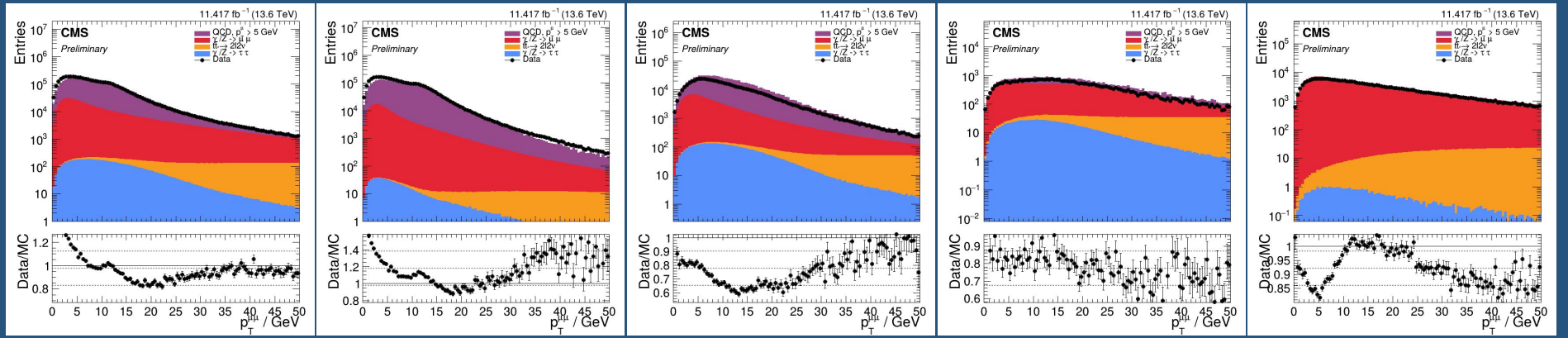


Run3 2024 : Mass Windows

Plots are going to be separated in 5 different mass windows :

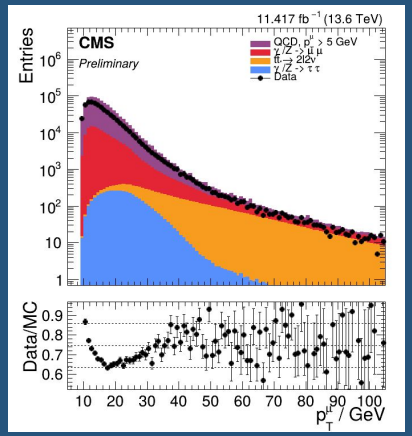
1. Upsilon window : [10, 20] GeV
2. QCD Window : [20, 50] GeV
3. W window : [50, 76] GeV
4. Z window : [76, 106] GeV

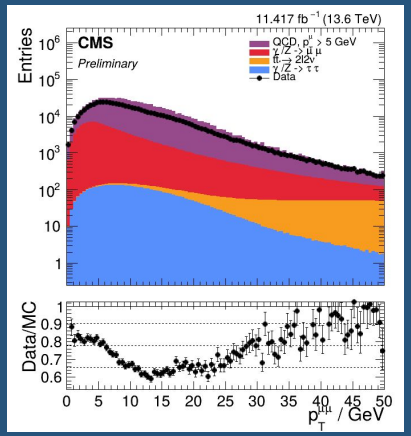
Pair transverse momentum

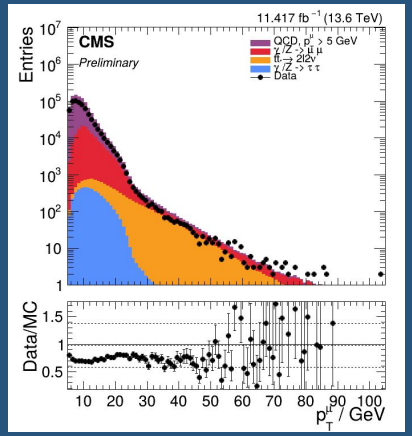


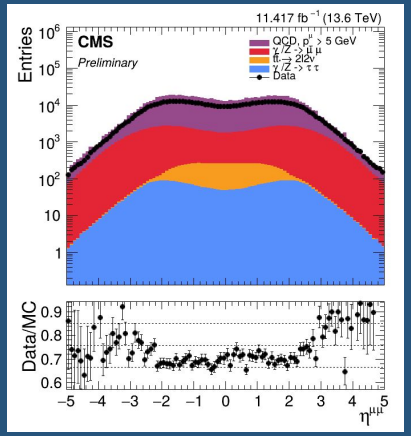
Run3 2024 : Background

- The data/mc ratio is good
  - Still, the big issue is the very high background
    - On the left, distribution in the [20, 60] GeV window





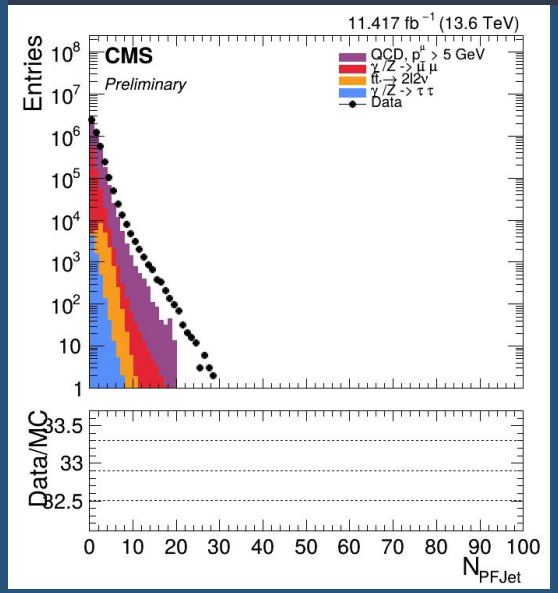


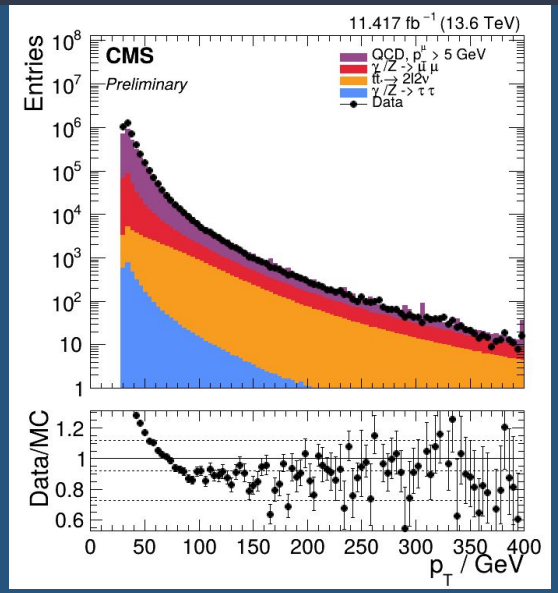


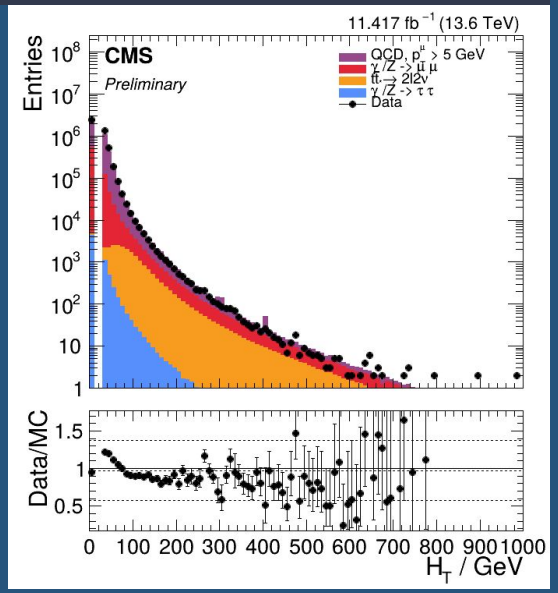
Run3 2024 : Background

- The data/mc ratio is good
  - Still, the big issue is the very high background
- Things we considered doing :
  - Use hadronic activity H_T, DY events are supposed to be relatively clean of hadronic activity
  - Cuts on PFJet : $p_T > 30$ GeV, $|\eta| < 2.5$

Jets variable for inclusive mass



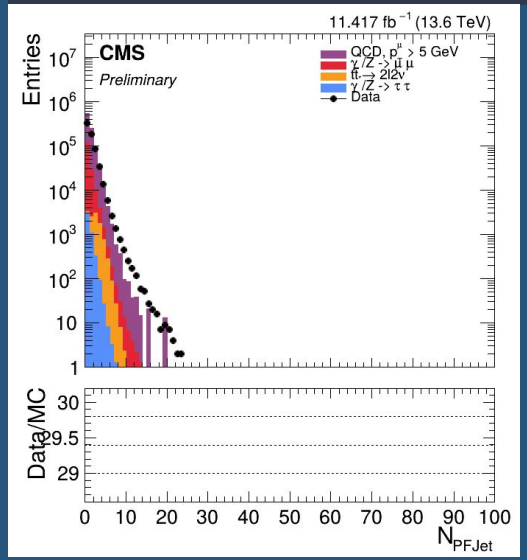


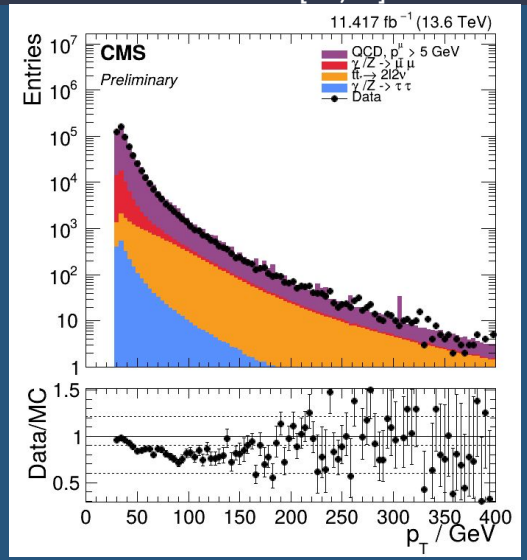


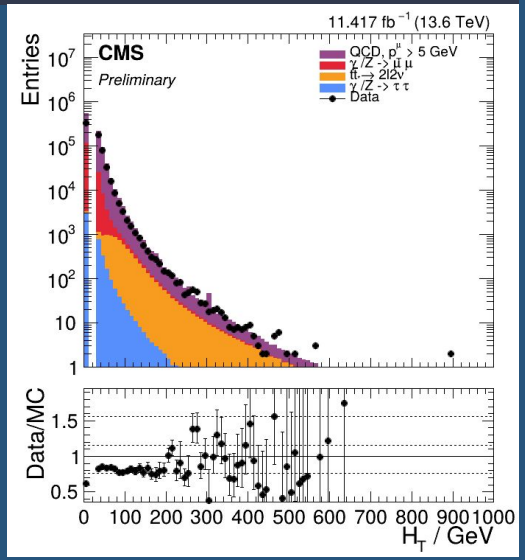
Run3 2024 : Background

- Things we considered doing :
  - Use hadronic activity H_T, DY events are supposed to be relatively clean of hadronic activity
  - Cuts on PFJet : $p_T > 30$ GeV, $|\eta| < 2.5$

Jets variable for [20,60] GeV



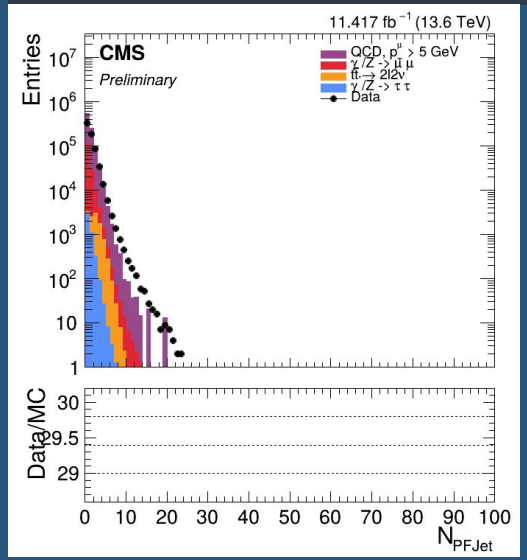


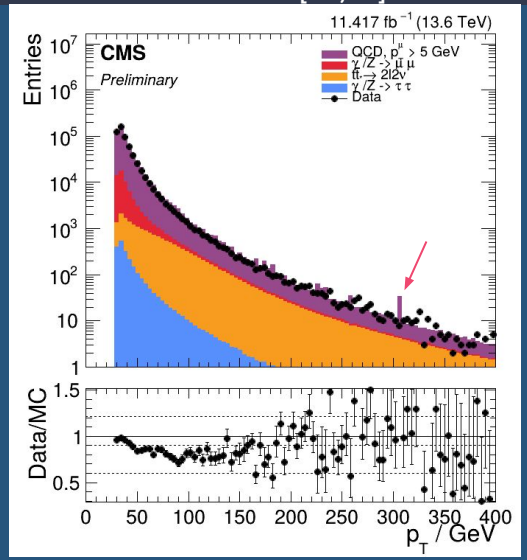


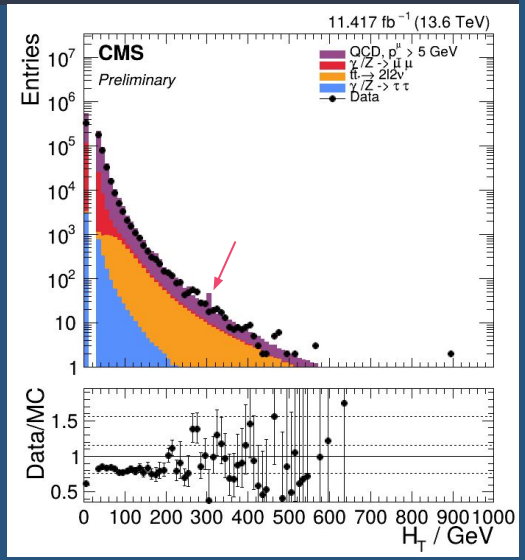
Run3 2024 : Background

- Things we considered doing :
  - Use hadronic activity H_T, DY events are supposed to be relatively clean of hadronic activity
  - Cuts on PFJet : $p_T > 30$ GeV, $|\eta| < 2.5$

Jets variable for [20,60] GeV



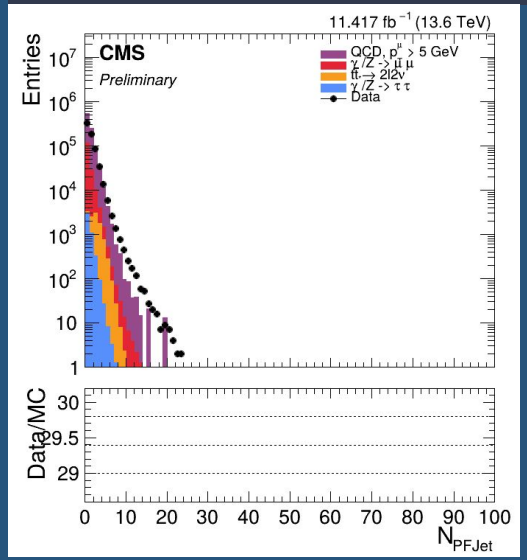


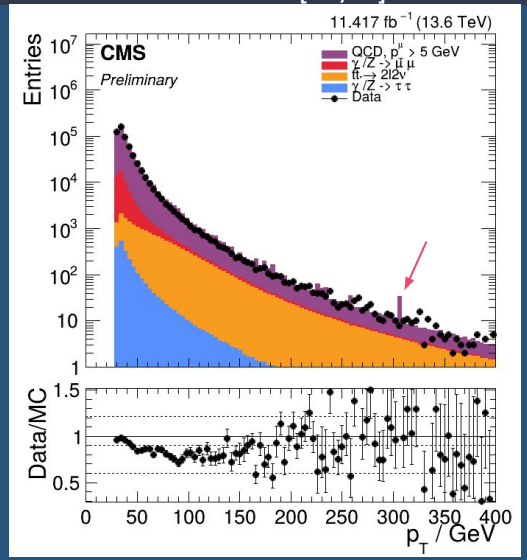


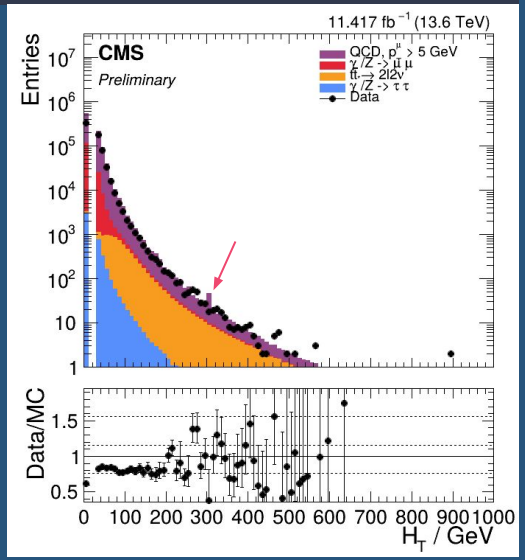
Run3 2024 : Background

- Things we considered doing :
  - Use hadronic activity H_T, DY events are supposed to be relatively clean of hadronic activity
  - Cuts on PFJet : $p_T > 30$ GeV, $|\eta| < 2.5$

Jets variable for [20,60] GeV

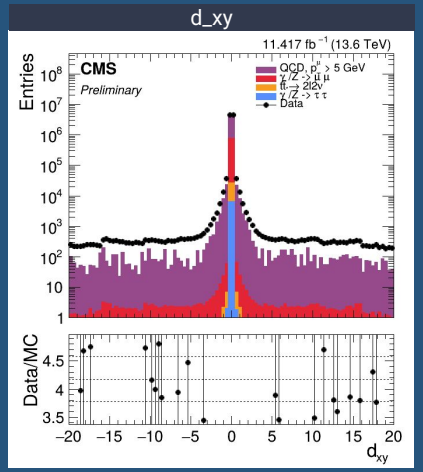


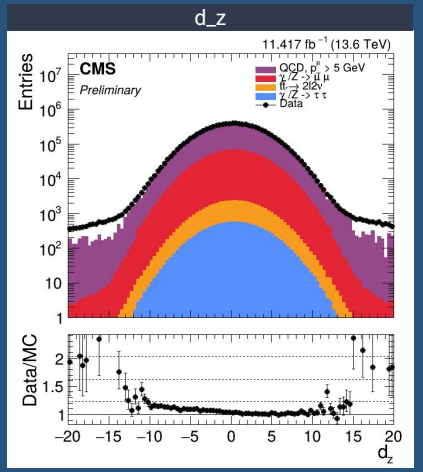




Run3 2024 : Background

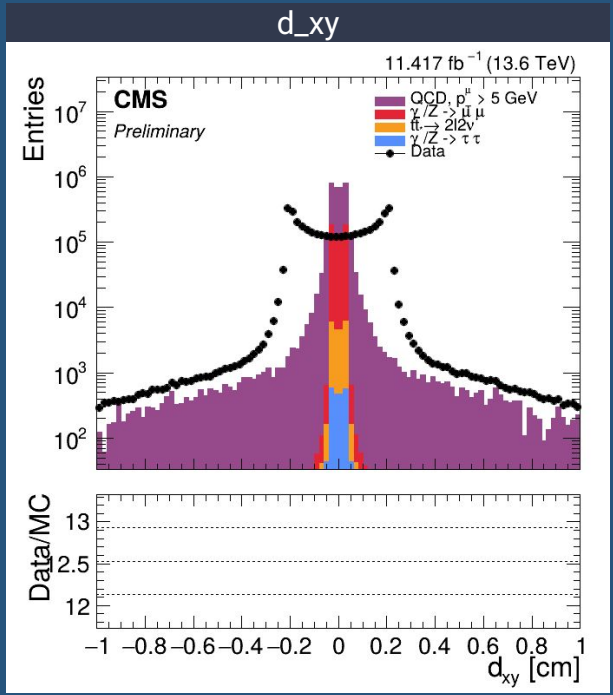
- Impact parameter :
  - could restrain to a small interval from the PV
    - a loose selection is 0.2 cm in d_xy and 0.5 cm in d_z
    - This could be also be put in a ScoutingMuonID similar to the PF tight ID





Run3 2024 : Background

- Impact parameter :
  - could restrain to a small interval from the PV
    - a loose selection is 0.2 cm in d_xy and 0.5 cm in d_z
    - This could be also be put in a ScoutingMuonID similar to the PF tight ID
  - Cutting at 0.1 cm could help a lot
    - we could look even closer



already infamous cowboy/volcano plot

Run3 2024 : Background

- Impact parameter :
  - could restrain to a small interval from the PV
    - a loose selection is 0.2 cm in d_xy and 0.5 cm in d_z
    - This could be also be put in a ScoutingMuonID similar to the PF tight ID
- Can we trust QCD samples ?
  - Could run a validation of it
    - Same-sign region
    - fake estimation methods

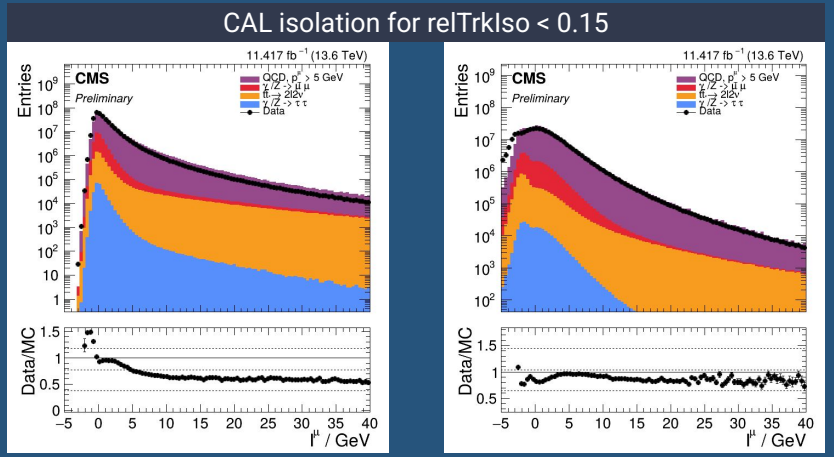
Run3 2024 : Background

- Impact parameter :
  - could restrain to a small interval from the PV
    - a loose selection is 0.2 cm in d_xy and 0.5 cm in d_z
    - This could be also be put in a ScoutingMuonID similar to the PF tight ID
- Can we trust QCD samples ?
  - Could run a validation of it
    - Same-sign region
    - fake estimation methods

Run3 2024 : Background

- Impact parameter :
  - could restrain to a small interval from the PV
    - a loose selection is 0.2 cm in d_xy and 0.5 cm in d_z
    - This could be also be put in a ScoutingMuonID similar to the PF tight ID
- Can we trust QCD samples ?
  - Could run a validation of it
    - Same-sign region
    - fake estimation methods
- Expand isolation cone
  - Use a 0.8 cone and PFCandidates to compute track isolation
  - Could hit even harder
- Use some absolute cut for isolation in ECAL and HCAL

CAL isolation for relTrkIso < 0.15



Conclusion

- Understanding of the Scouting format improved
  - What are the different collections and what they represent
  - In particular, ScoutingMuon has been searched extensively
- Good agreement between the Scouting dataset and the simulation
  - There are still some issues, but we have a good baseline
  - Still QCD should be better understood and the efficiency should be explained
- Last points is what will be the focus for coming future

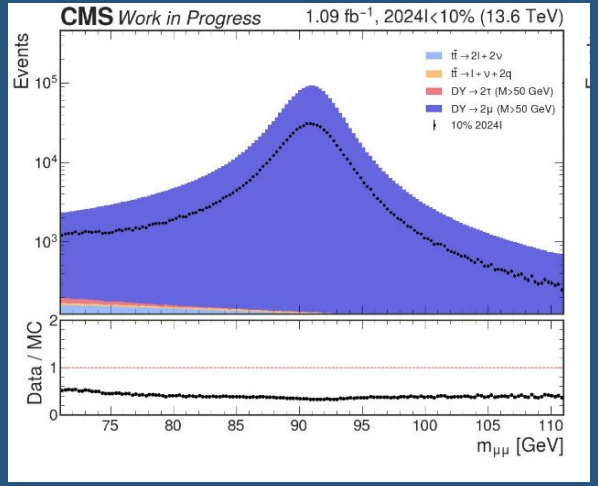
Conclusion

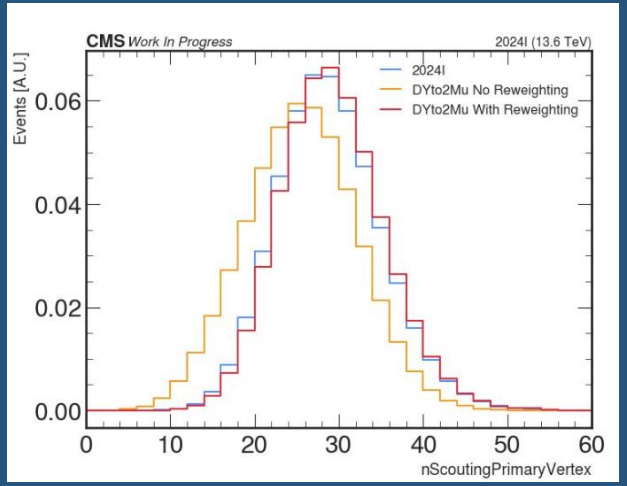
- Understanding of the Scouting format improved
  - What are the different collections and what they represent
  - In particular, ScoutingMuon has been searched extensively
- Good agreement between the Scouting dataset and the simulation
  - There are still some issues, but we have a good baseline
  - Still QCD should be better understood and the efficiency should be explained
  - Defining a ScoutingMuonID
- Last points is what will be the focus for coming future
- Something I have not discussed, low PU have not been explored yet

BACKUP

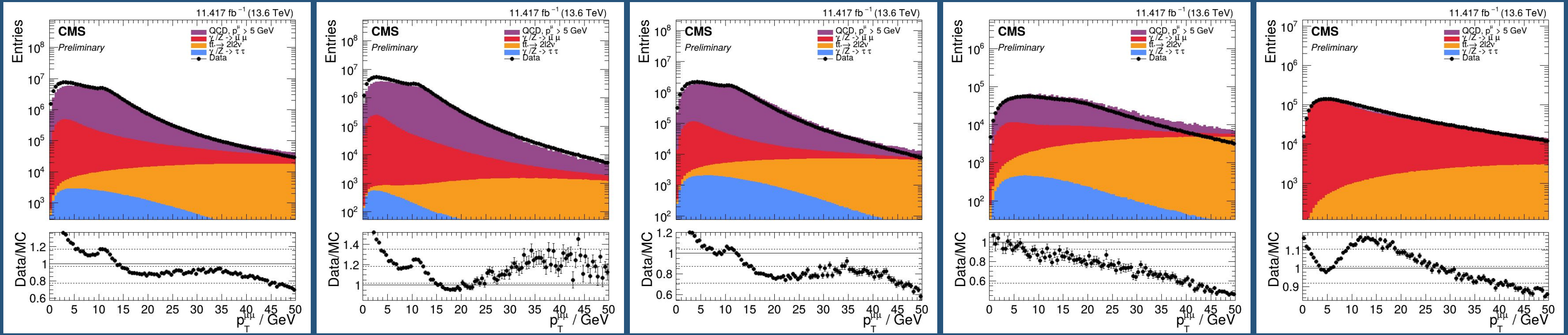
Run3 2024 : HLT Scouting Comparison

- 2 issues, actually the same as in 2022 :
  - Normalisation is off ~20%, but it is also seen by HLT Scouting group.
    - Simple normalisation to the Z peak to alleviate the issue
  - Z peak is not aligned in data and simulation, not seen by Scouting group
    - Difference with our case is they do the pileup reweighting and do not have the same MC
    - Triggers are also different, but it shouldn't make a big difference

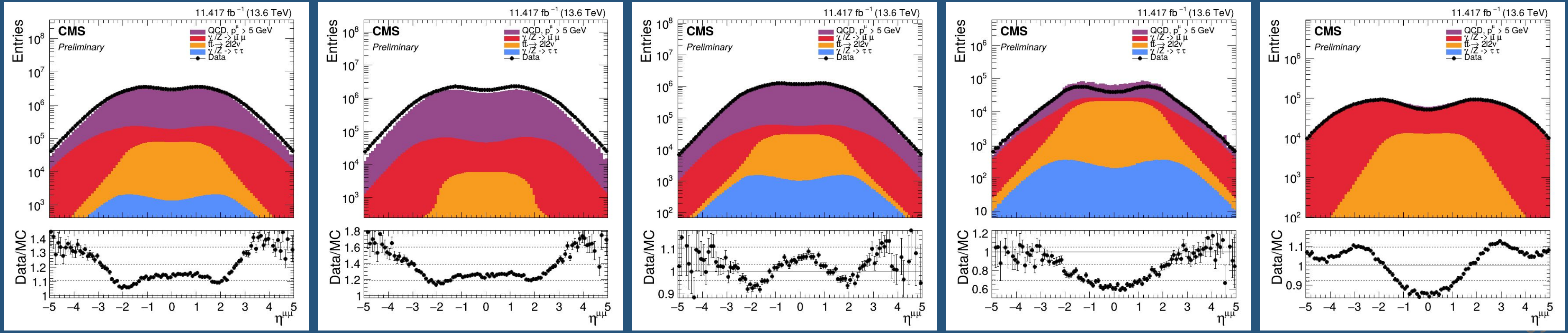




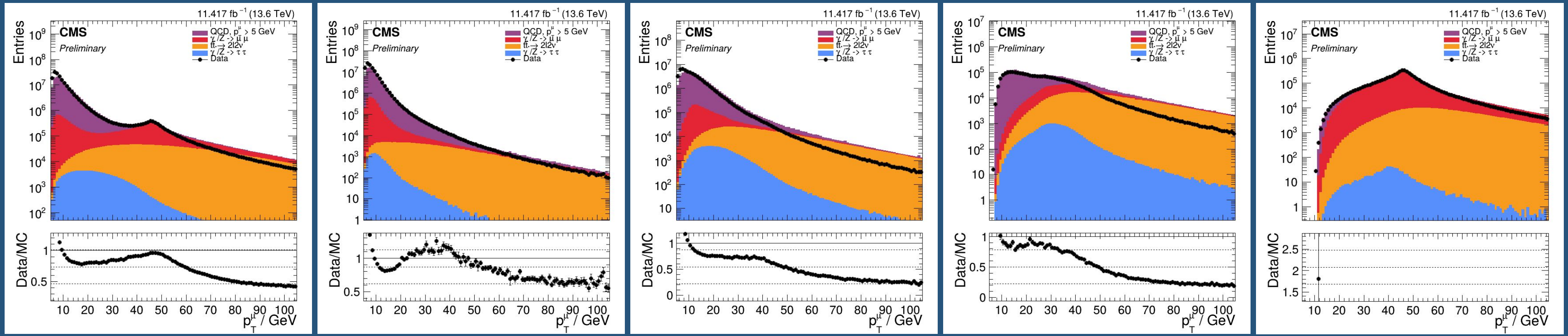
Muon pair p_T



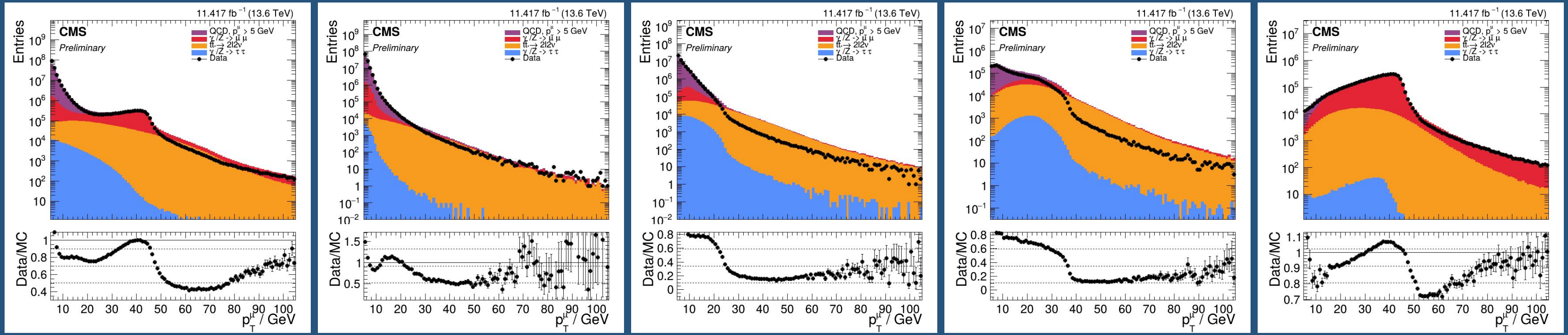
Muon pair eta



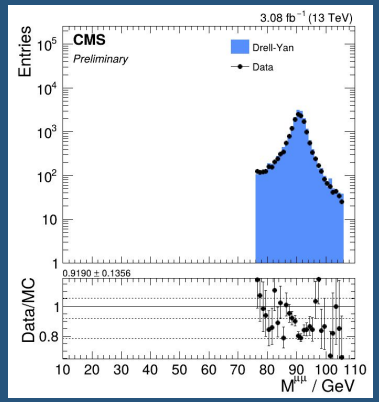
Leading muon p_T

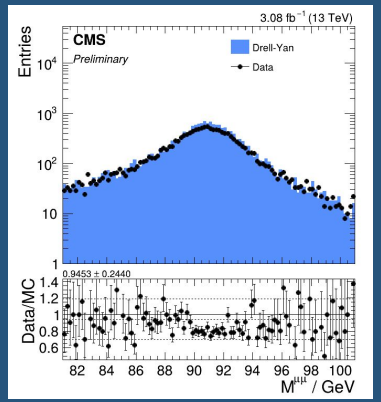


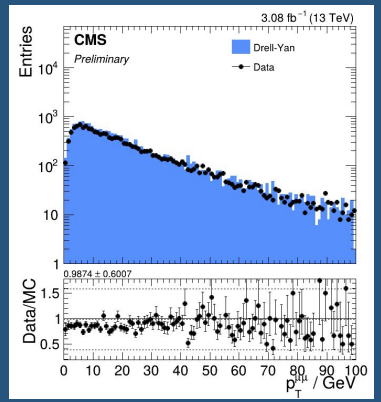
Subleading muon p_T

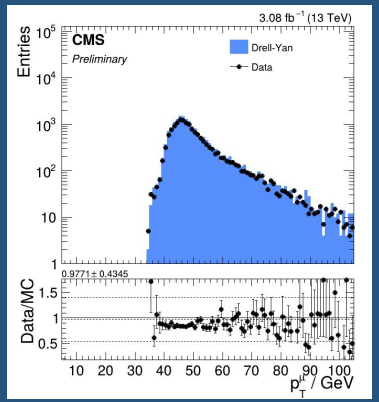


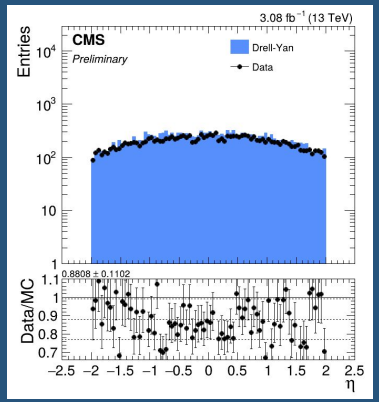
with p_T > 5.0 + correlation cuts MllVsPt_l,s

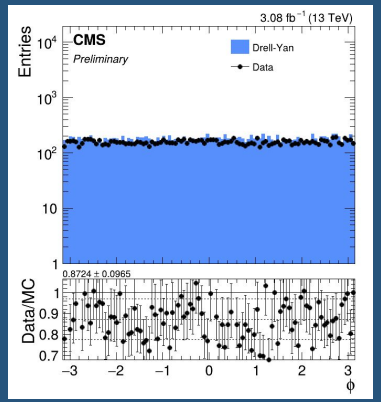


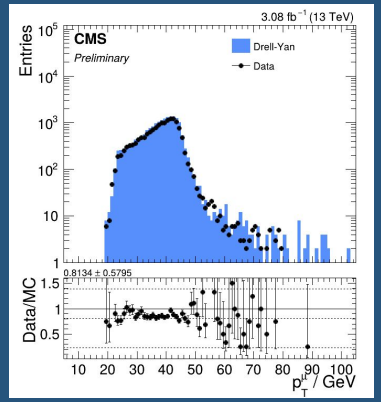


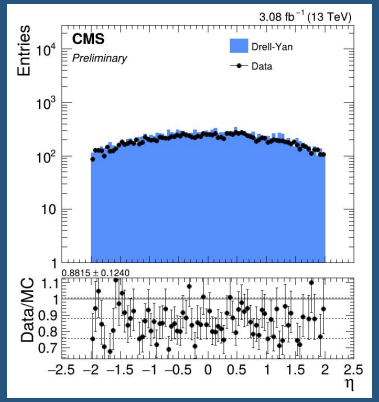












Issue with 2022 Scouting Data

2 big issue with the previous look at the Z peak :

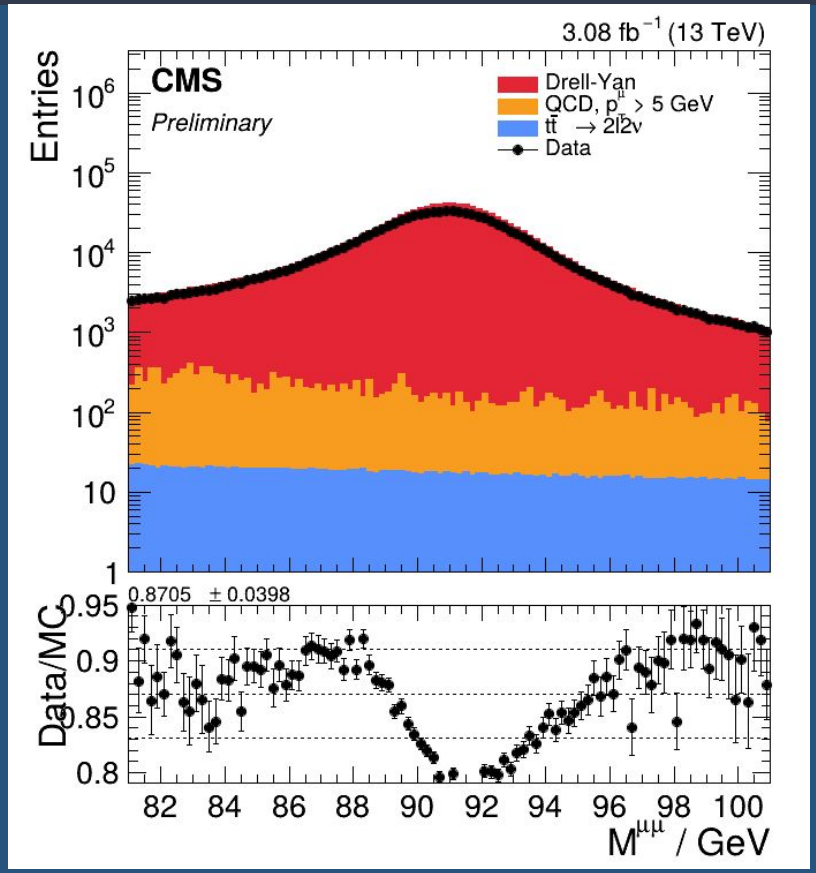
1. Normalisation was off ~15%
2. The Z peak was shifted by a few GeV

Easy check were performed :

- Add ttbarr samples
- Restrict η-range
- Look at different triggers

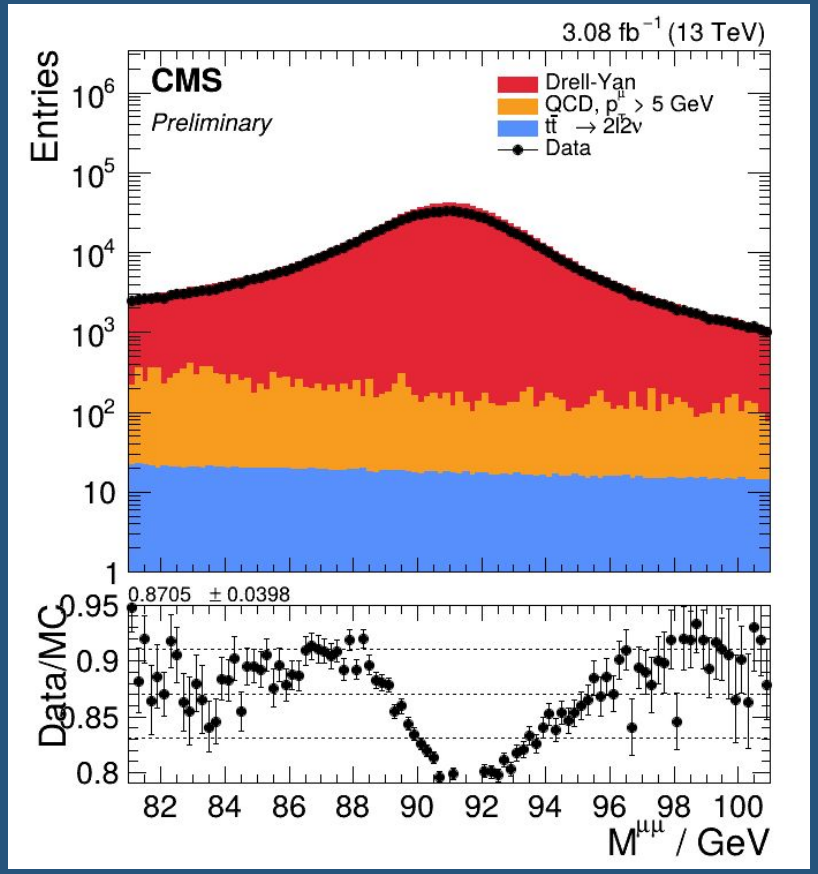
Maybe the tool used to produce ScoutingNanoAOD format was the issue :

- Tool was mostly in development when I produced the samples
- I actually don't remember if the same version of CMSSW was used to produce data and MC…

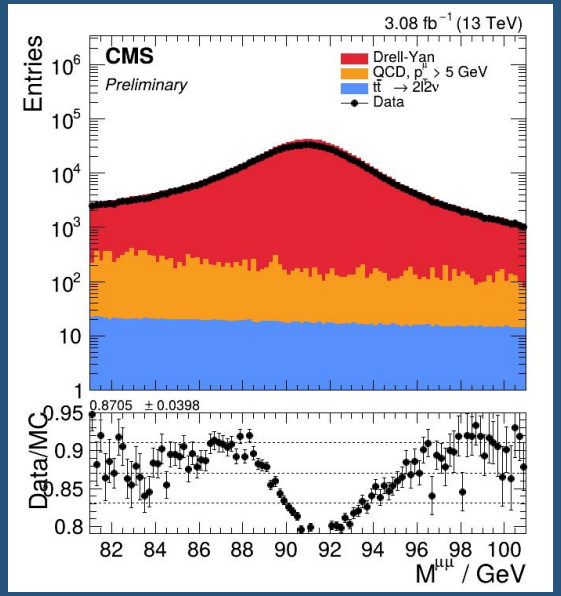


1. Z peak issue : previous result

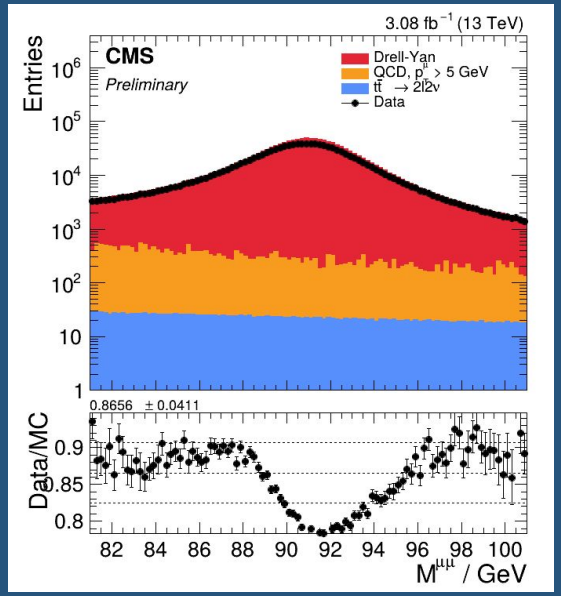
- Cuts used :
  - $|\eta| < 2.0$ + $p_T > 5.0$ GeV
  - 76 < Mll < 106 GeV
  - p_T^{mu2}/Mll > 0.25 + p_T^{mu1}/Mll > 0.45
  - HLT scouting path and L1 fired
    - `DST_Run3_PFScoutingPixelTracking`
    - `L1_DoubleMu4p5er2p0_SQ_OS_Mass_Min7`



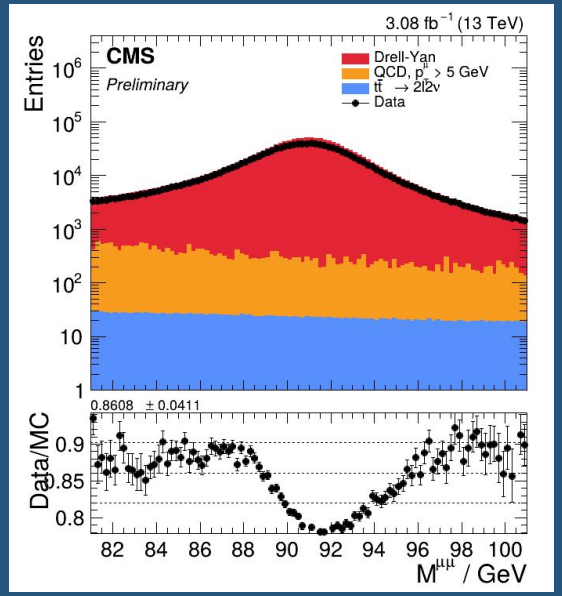
1. Z peak issue : $|\eta| < 2.4$



L1_DoubleMu4p5er2p0_OS_SQ_Mass_Min7

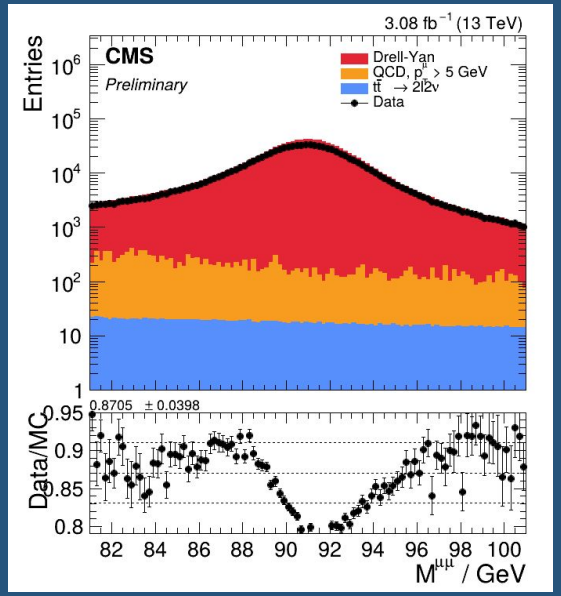


L1_DoubleMu8_SQ

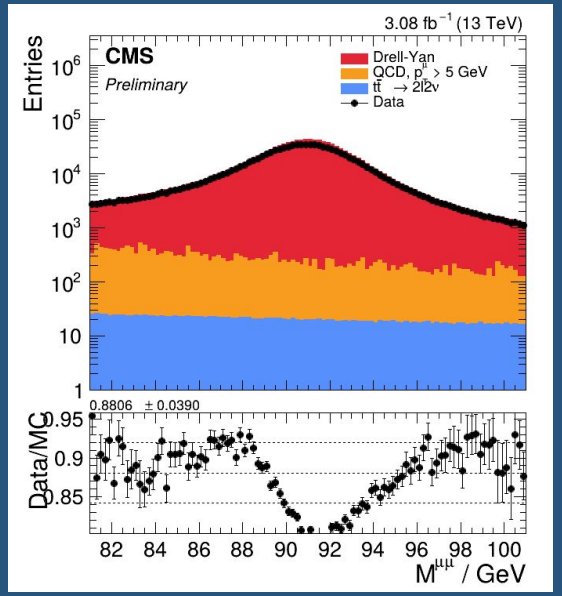


L1_DoubleMu_15_7

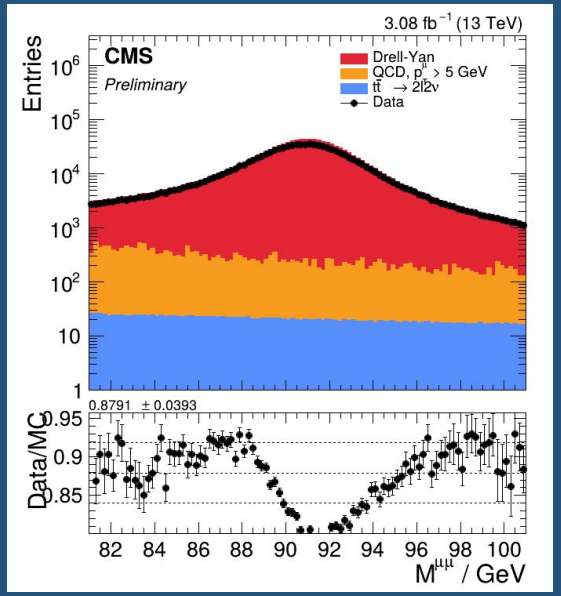
1. Z peak issue : $|\eta| < 2.0$



L1_DoubleMu4p5er2p0_OS_SQ_Mass_Min7

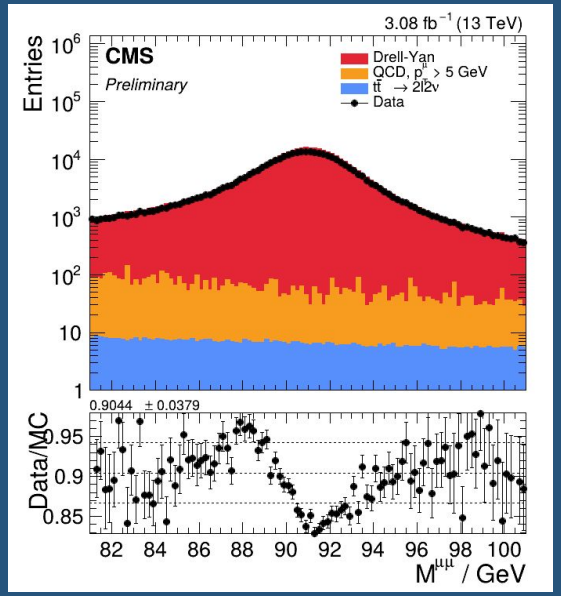


L1_DoubleMu8_SQ

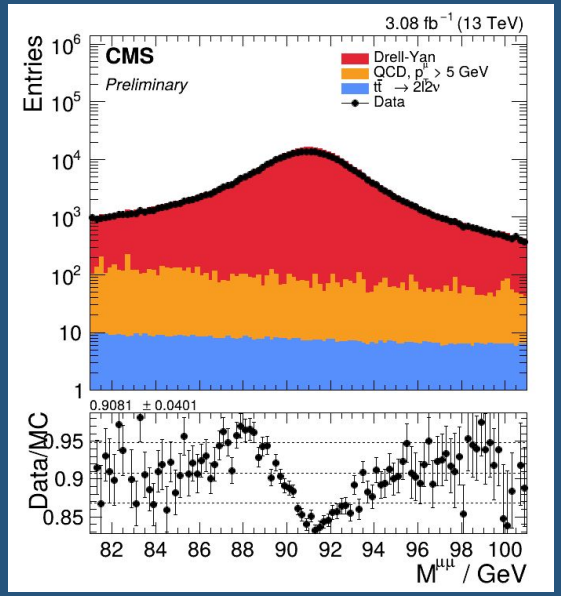


L1_DoubleMu_15_7

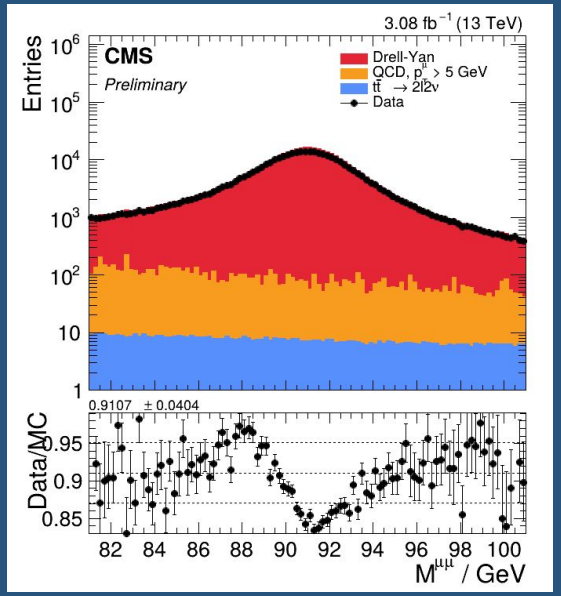
1. Z peak issue : $|\eta| < 0.9$



L1_DoubleMu4p5er2p0_OS_SQ_Mass_Min7



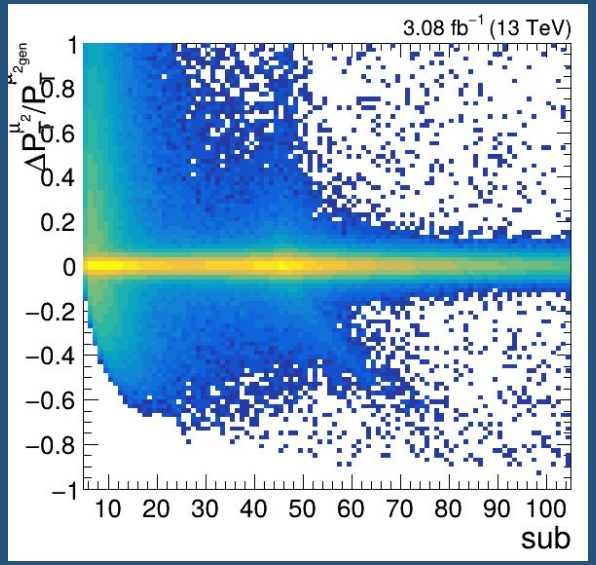
L1_DoubleMu8_SQ

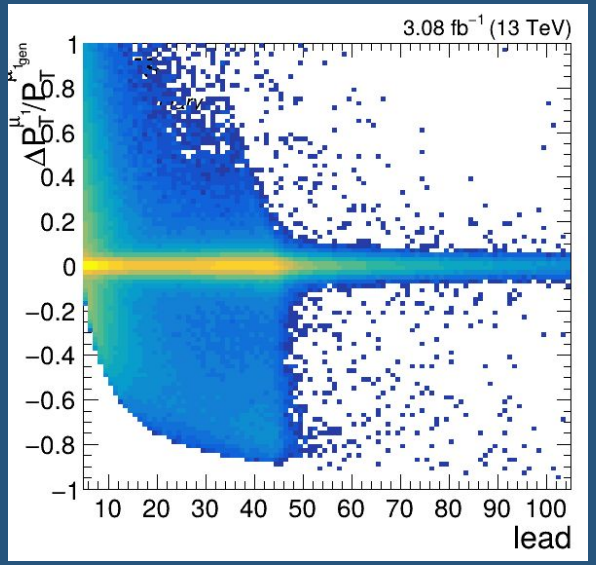


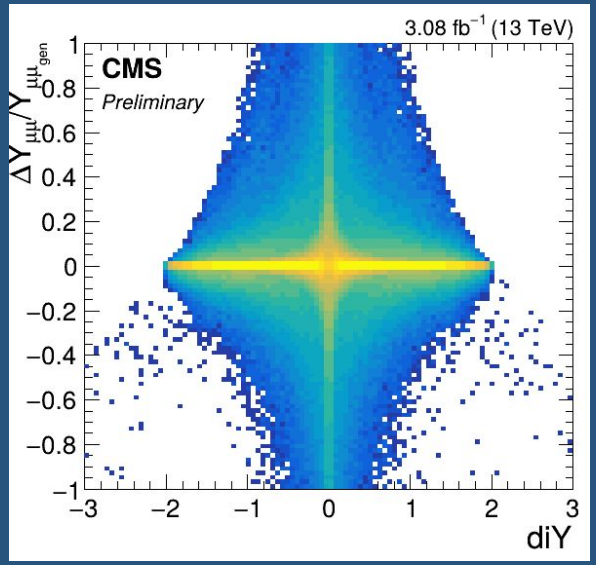
L1_DoubleMu_15_7

2. Resolution matter : Relative-difference

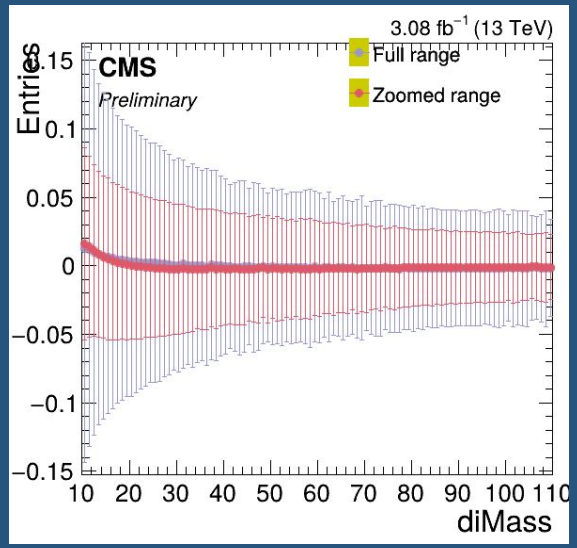
- relative difference : difference between gen-level and rec-level divided by gen-level
  - Now looks fine for leadin and trailing pt and the comtation was corrected for $Y_{\mu\mu}$

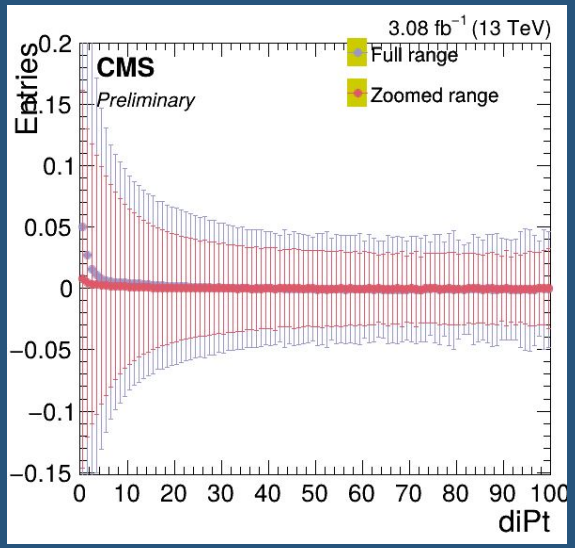


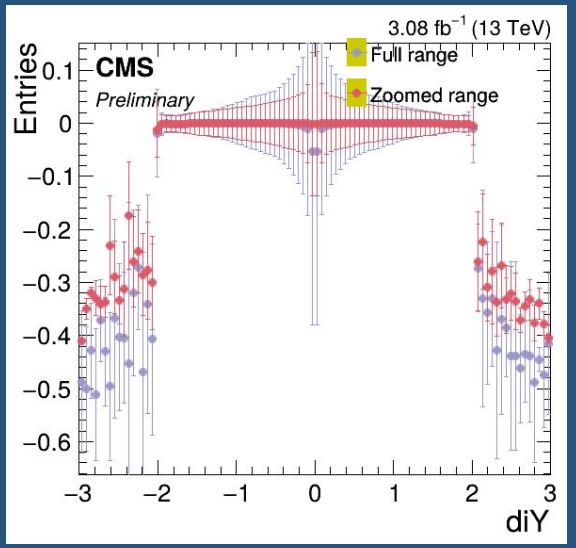




2. Resolution matter : resolution distribution







2. Resolution matter : resolution distribution