Data-MC discrepancy study

Yang Tianyi

Datasets

- Run 2022C-G Muon
  - Golden json luminosity 5.0104, 2.9700, 5.8070, 17.7819, 3.0828 fb^{-1} (34.7 fb^{-1} in total)
- Run 2023C-D Muon0 & Muon1
  - Golden json luminosity: 17.794, 9.451 fb^{-1} (27.2 fb^{-1} in total)
- Run 2024C-I Muon0 & Muon1
  - Golden json luminosity: 7.24, 7.96, 11.32, 27.76, 37.77, 5.44, 11.47 fb^{-1} (109 fb^{-1} in total)

- Run 2022 and Run 2023 with corresponding condition MC samples downloaded for direct comparison: **61.9 $\mathbf{fb}^{-1}$**
- From Run 2022 to Run 2024 all data in era for physics analysis: **170.9 $\mathbf{fb}^{-1}$**

Data trigger path and filter

- Trigger menu
  - Single muon trigger `HLT_IsoMu{20|24}{_eta2p1} HLT_IsoMu27 HLT_Mu50`
  - Dimuon trigger `HLT_Mu{17|19}_TrkIsoVVL_Mu{8|9}_TrkIsoVVL_DZ_Mass{3p8|8}`
  - Trimuon triggers `HLT_TripleMu_5_3_3_Mass3p8_DZ HLT_TripleMu_10_5_5_DZ HLT_TripleMu_12_10_5`
- MET noise filter:
  - `Flag_goodVertices, Flag_globalSuperTightHalo2016Filter, Flag_EcalDeadCellTriggerPrimitiveFilter, Flag_BadPFMuonFilter, Flag_eeBadScFilter, Flag_BadPFMuonDzFilter, Flag_hfNoisyHitsFilter`

MC background samples

- Dominant Drell-Yan global background:
  - DYto2L-4Jets_MLL-50_TuneCP5_13p6TeV_madgraphMLM-pythia8/NANOAODSIM/
  - Cross-section (NNLO value): 6345.99 pb
- Other important backgrounds:
  - W+jets, DY with photons.
  - $t\bar{t}$ DL/SL
  - Diboson ($WW2l2\nu, WZ2l2q, ZZ2l2\nu, ZZ2l2q, WZ3l, ZZ4l$), triboson
  - $t\bar{t}t\bar{t}$
  - $t\bar{t}V(V)$
  - Higgs: $t\bar{t}H, t\bar{t}VH, tH$
- Note that in Run3Summer22NanoAODv12, the TTWZ sample is missing.
- The Run2024 corresponding MC sample not finish downloaded yet, currently with rescale the MC of Run2022 and Run2023 by $\mathbf{170.9\ fb^{-1}}$/ $\mathbf{61.9\ fb^{-1}}$=2.76 for Full data comparison.

Object selections

- Muon p_T need to take Rochester corrections
- Run2024 does not have the Rochester correctionlib json released yet, thus Run2024 data use the original Muon p_T before correction.
- Muon selection
  - GoodMuon25: $|\eta| < 2.4$, tight ID, tight Iso, $p_T > 25\ \mathrm{GeV}$
  - GoodMuon15: $|\eta| < 2.4$, tight ID, tight Iso, $p_T > 15\ \mathrm{GeV}$
  - Tight isolation: Muon_pfRelIso04_all<0.15
- Dimuon preselection
  - Leading Muon fulfill $p_T > 25\ \mathrm{GeV}$ condition
  - Other Muons fulfill $p_T > 15\ \mathrm{GeV}$ condition
  - nGoodMuon_15 > 1 && nGoodMuon_25 > 0

MC corrections

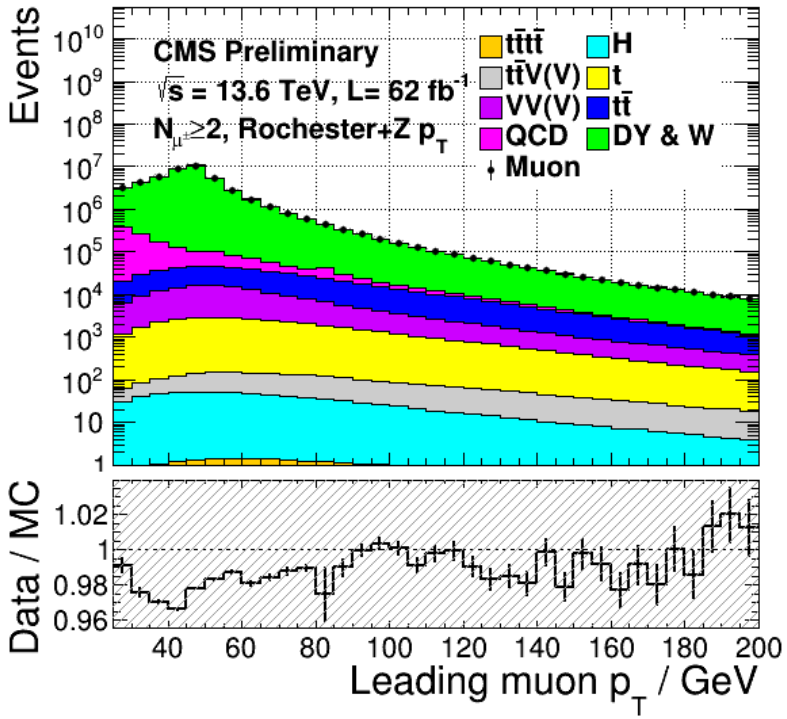
- Pile-up correction
  - From https://gitlab.cern.ch/cms-nanoAOD/jsonpog-integration.git
  - Applying based on truth interactions
  - Adding additional weights
- Muon scale corrections
  - From https://gitlab.cern.ch/cms-nanoAOD/jsonpog-integration.git
  - Applying for the leading and subleading Muon
  - Adding additional scaling factors
- Drell-Yan Z p_T correction
  - From https://gitlab.cern.ch/cms-higgs-leprare/hleprare.git
  - Based on the truth Z boson p_T to fix the Muon momentum
  - Adding additional weights

Cutflow table

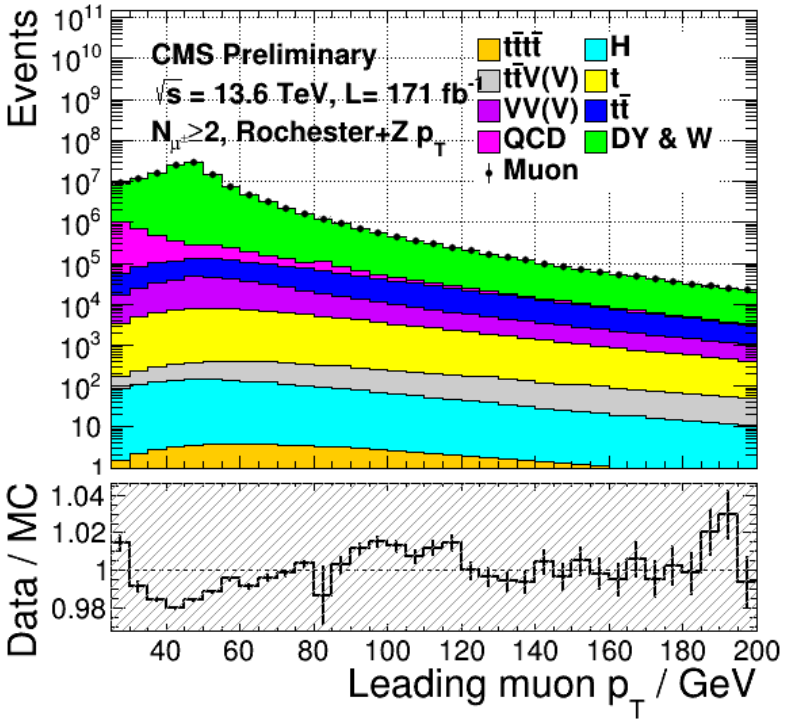
	S0: global CR						
MC campaign	22PreEE		22PostEE			23PreBPix	23PostBPix
Drell-Yan	5514746.6		17781241.1			14627220	7355623
$t\bar{t}$	52990.9		157659.9			121917.1	64944.2
tV	5451.3		16351.4			12464.8	6587.1
$t\bar{t}V(V)$	269.11		828.8			616.6	326.1
Higgs	112.91		337.48			260.69	137.71
W+jets	2000.4		4612.1			5824.6	2937.7
VV	18630.5		55848.6			42865.6	22711.21
$t\bar{t}t\bar{t}$	4		12.2			9.16	4.8
QCD	141505		406849			353025	184440
Era	2022C	2022D	2022E	2022F	2022G	2023C	2023D
data	3334682	2203560	4412469	12549031	2189719	14043412	7639872
Data/MC	96.6%		103.4%			92.6%	100%

Leading muon p_T

Run2022+Run2023



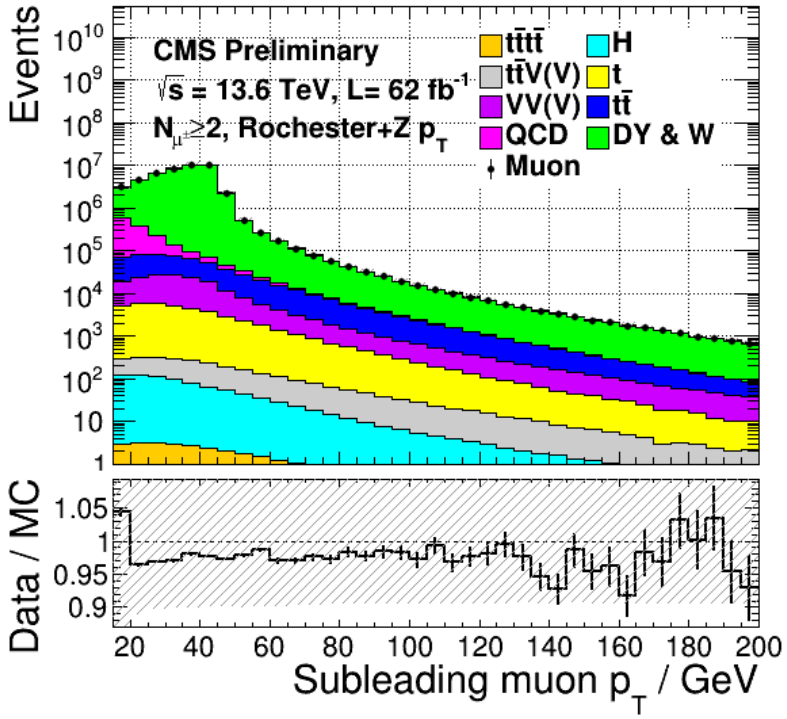
Run2022+Run2023+Run2024



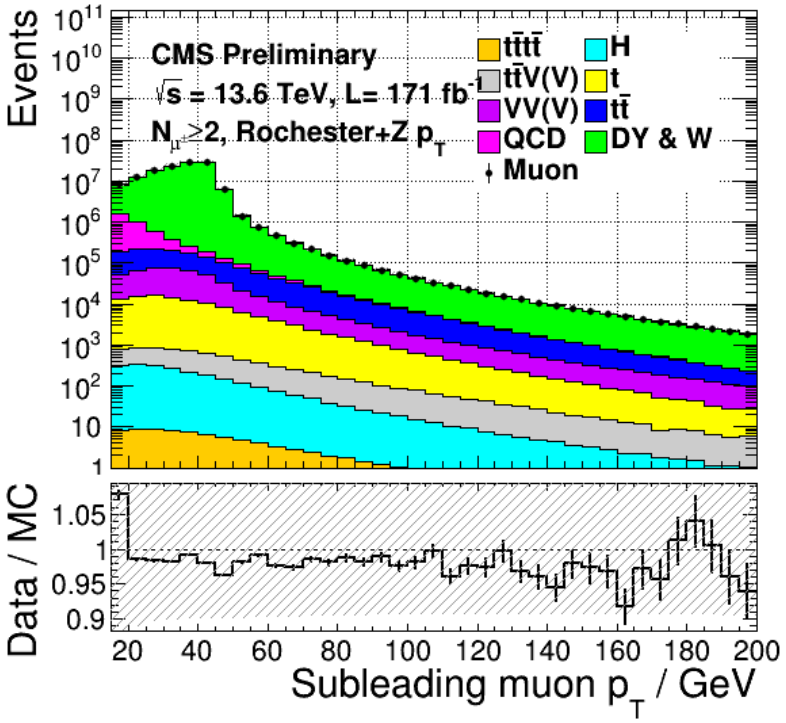
- Good agreement with the Drell-Yan Z p_T corrections applied.

Subleading muon p_T

Run2022+Run2023



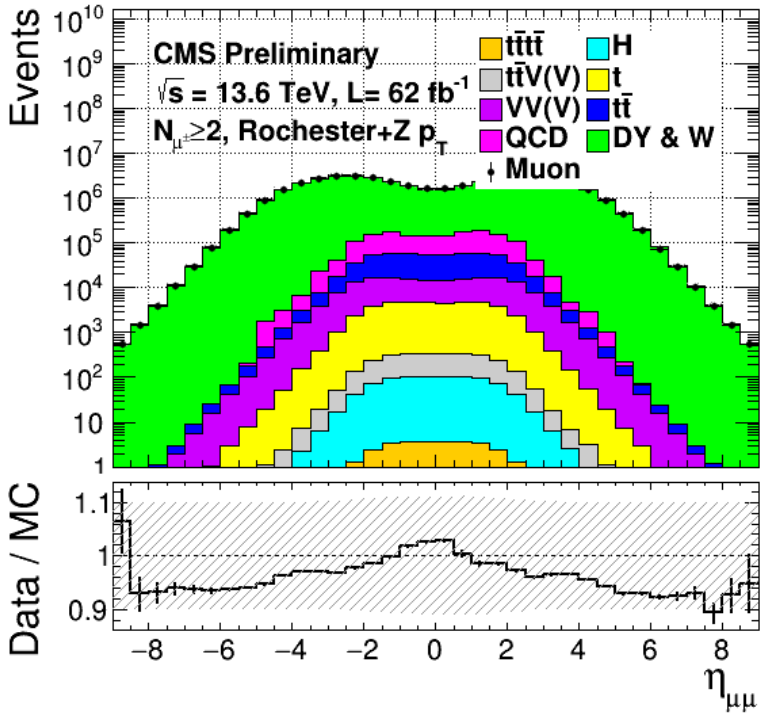
Run2022+Run2023+Run2024



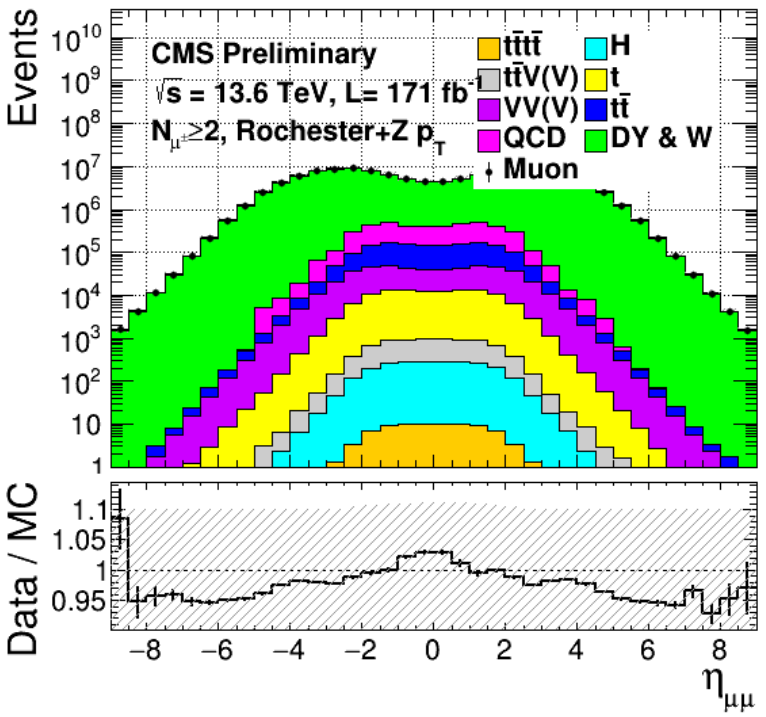
- Good agreement with the Drell-Yan Z p_T corrections applied.

Di-Muon variables $\eta_{\mu\mu}$

Run2022+Run2023



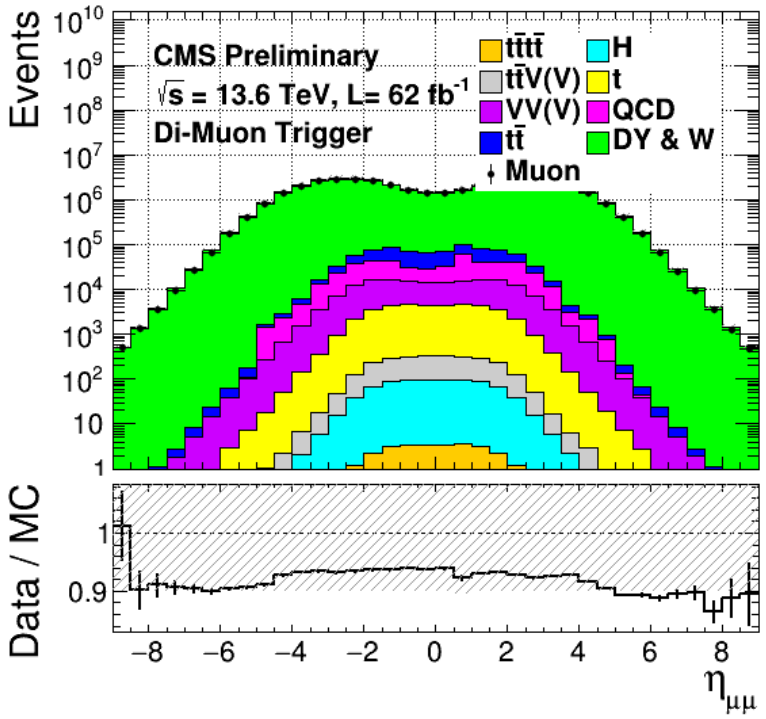
Run2022+Run2023+Run2024



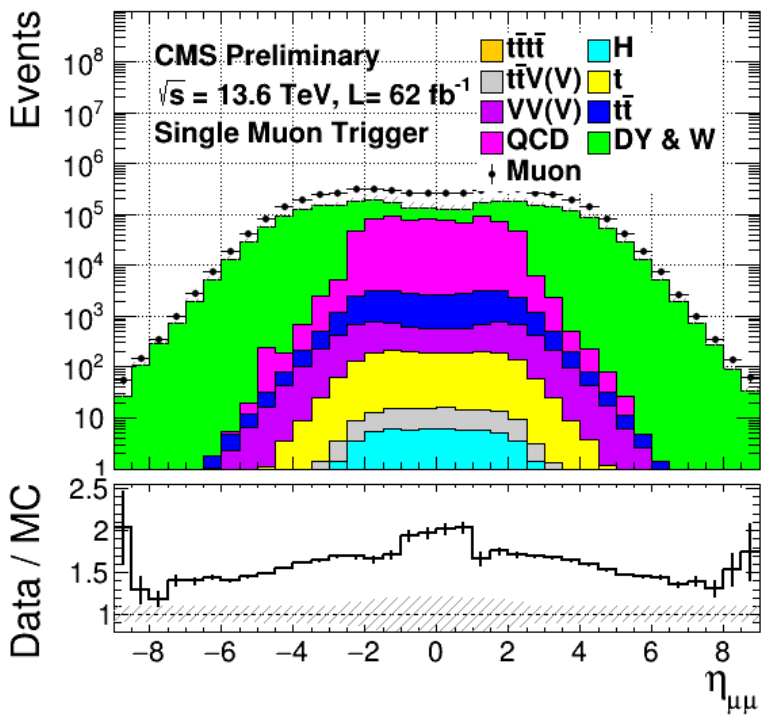
- Have a peak in the central region.

Trigger path checking 2022+2023

Dimuon trigger



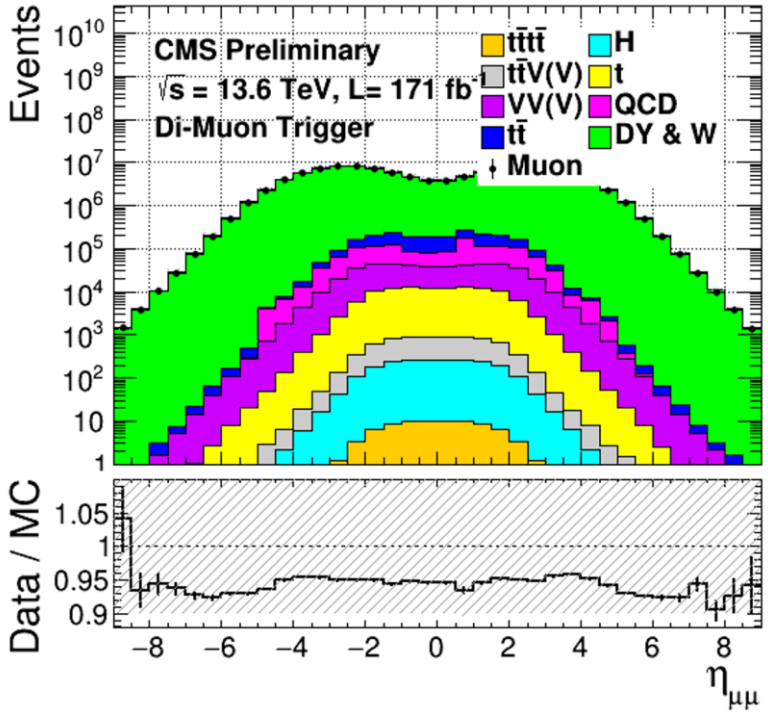
Single muon trigger



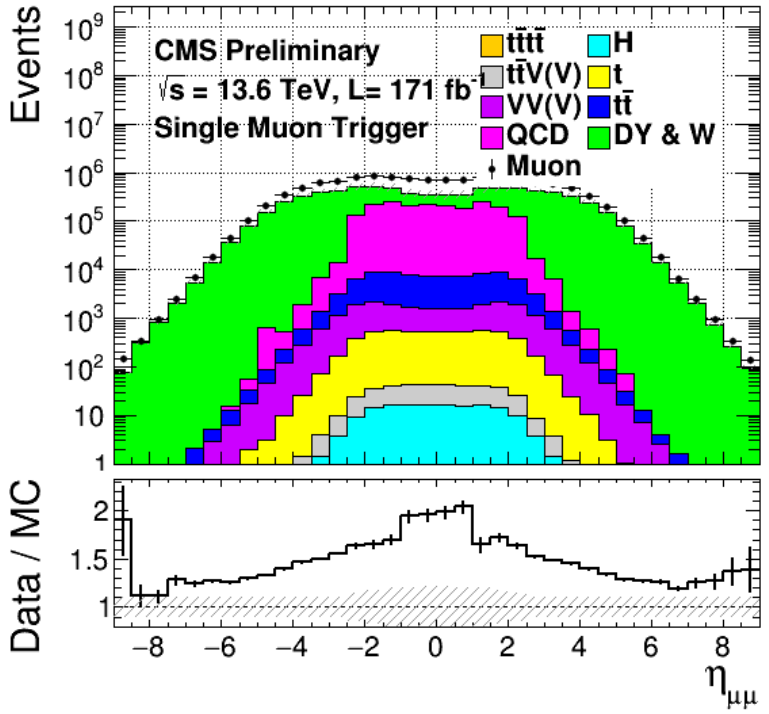
- The dimuon trigger has a flat shape after Z p_T correction, but low normalization.
- The single muon trigger plot has the peak in the middle as seen in the complete trigger scheme plot.

Trigger path checking 2022+2023+2024

Dimuon trigger



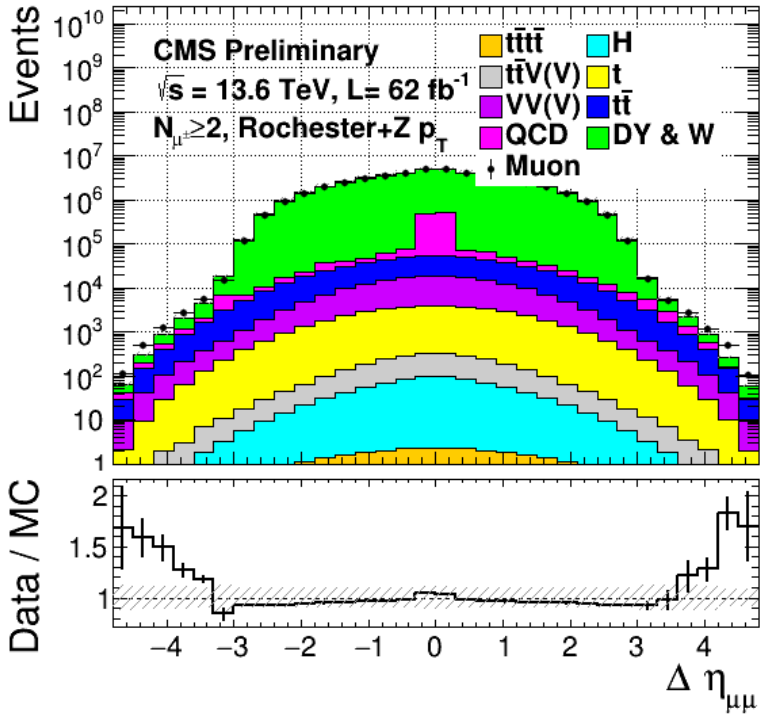
Single muon trigger



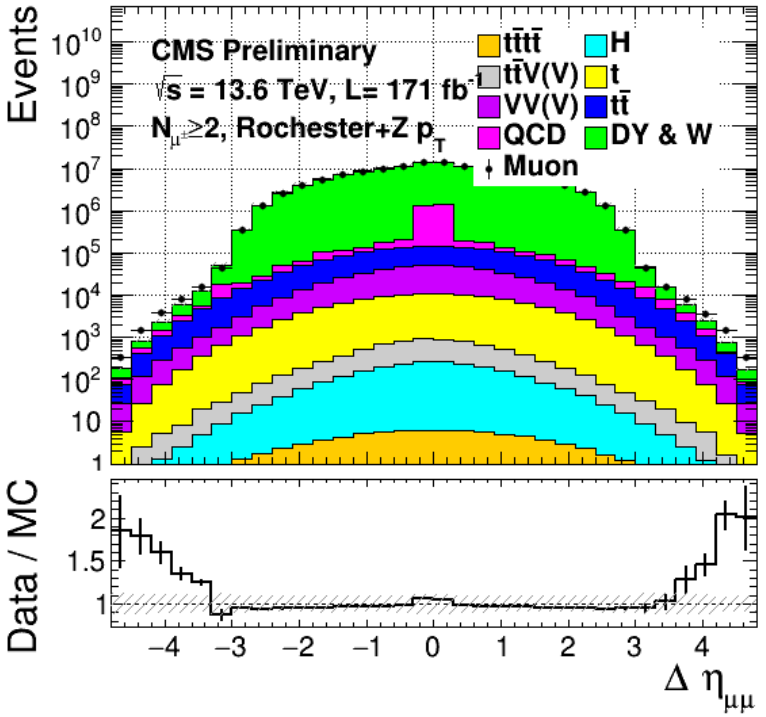
- The Run2024 data makes the normalization deviation smaller.
- The peak in the single muon trigger path remains.

Di-Muon variables $\Delta\eta_{\mu\mu}$

Run2022+Run2023



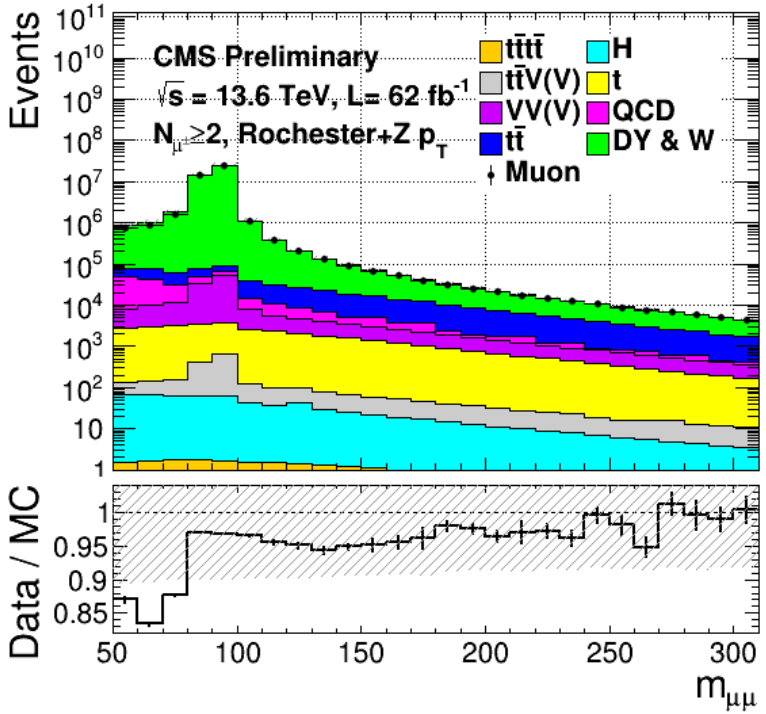
Run2022+Run2023+Run2024



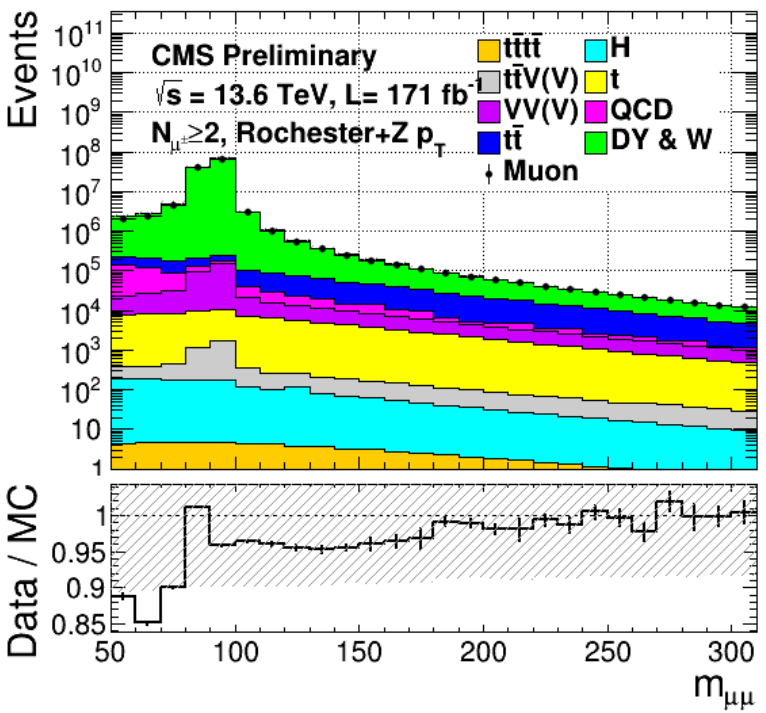
- The side regions have a gap between data and MC.

Di-Muon variables $m_{\mu\mu}$

Run2022+Run2023



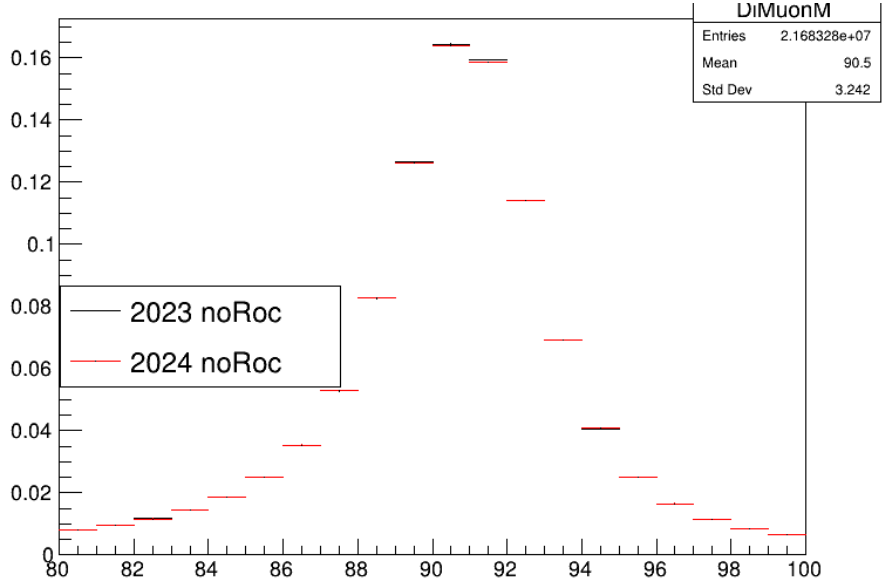
Run2022+Run2023+Run2024

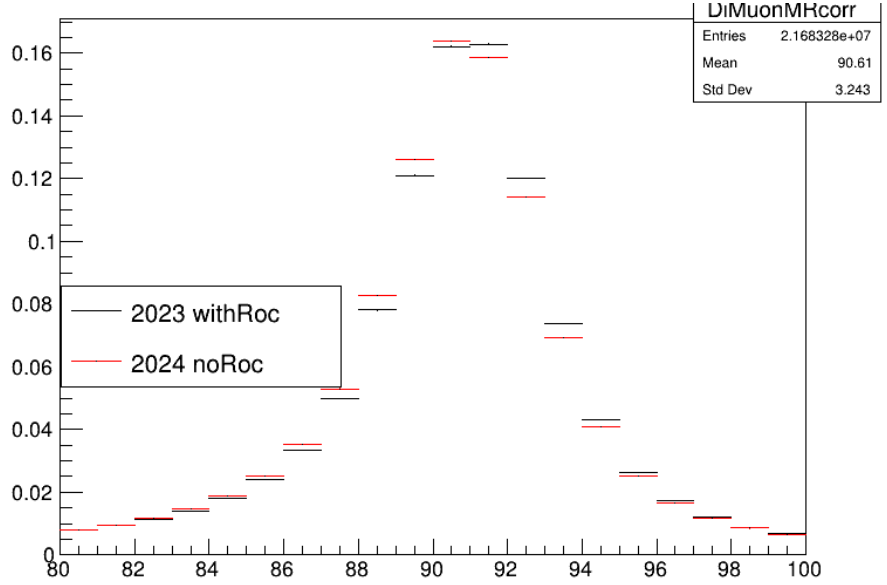


- The $m_{\mu\mu}$ distribution has a low mass overshoot in both case.
- The Run2024 does not have Rochester correctionlib json from Muon POG shared, thus the mass peak shift shows again.

Rochester correction

- Rochester correction 2024 will be finished soon: https://indico.cern.ch/event/1550716/contributions/6535947/attachments/3074969/5441414/pT_calibration_may25.pdf

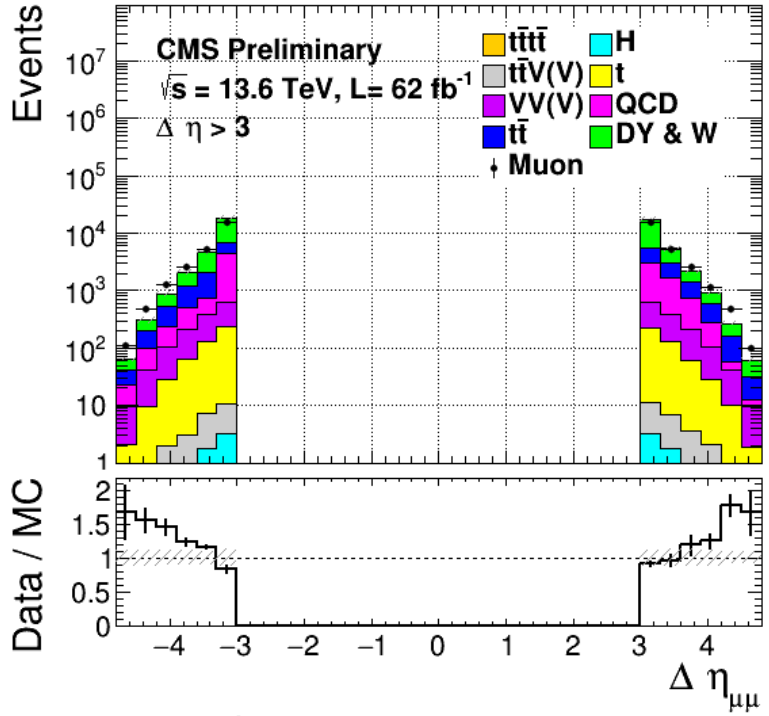




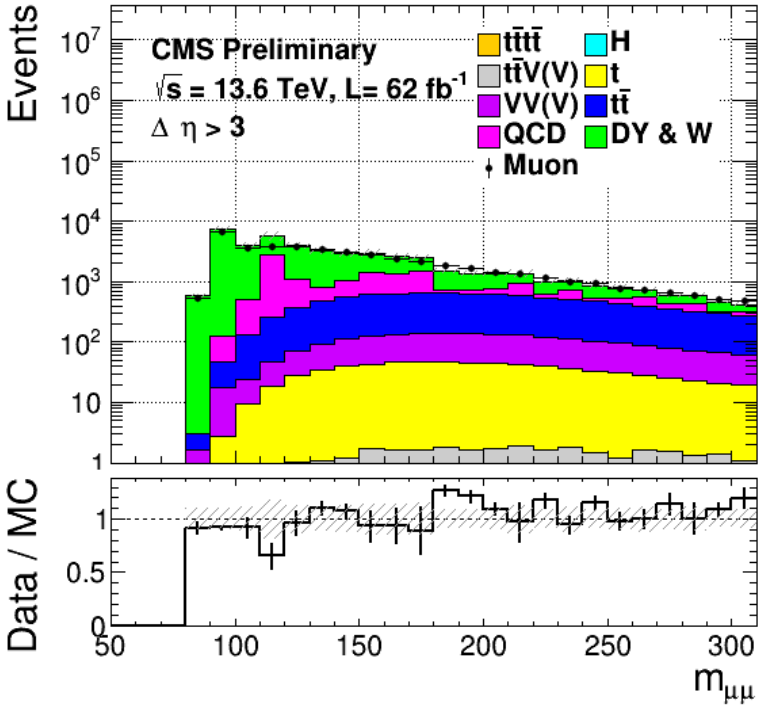
- Normalized shape of $m_{\mu\mu}$ comparison of 2023 and 2024 data.
- Pre-Rochester correction difference is small. The main difference of the previous slide with 2024 data is due to no Rochester correction for 2024 applied.

Large $\Delta\eta_{\mu\mu}$ region check

$\Delta\eta_{\mu\mu}$>3.0



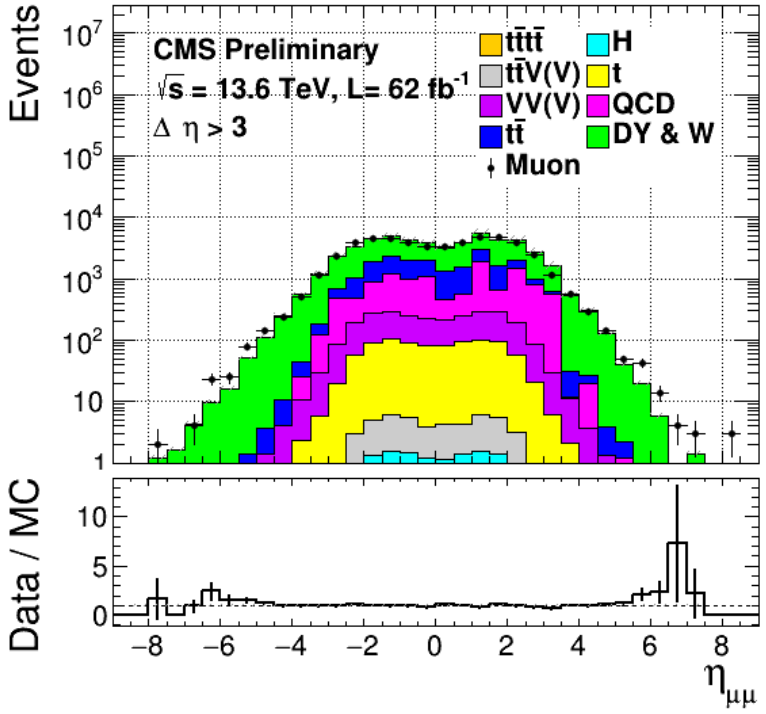
$m_{\mu\mu}$



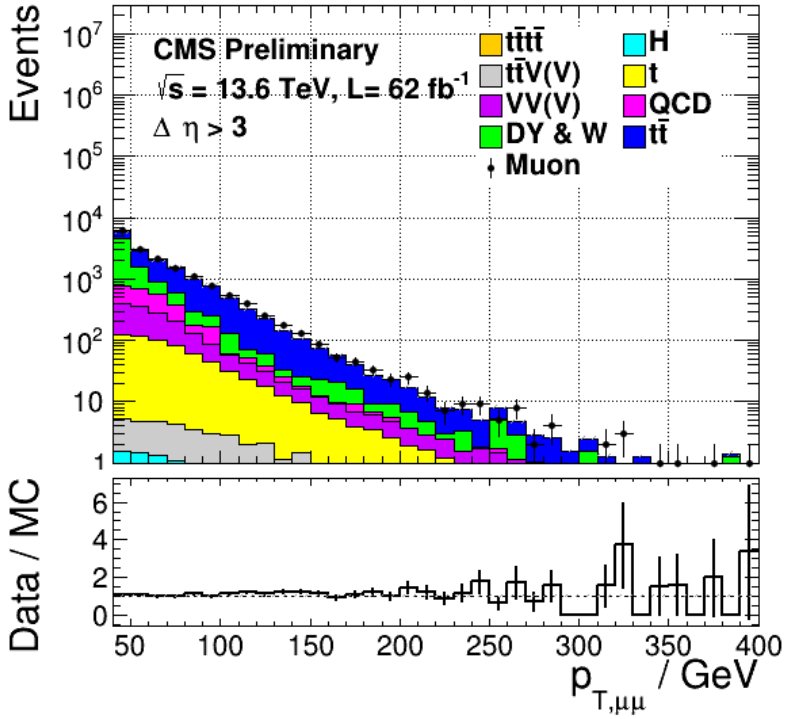
- Take the large $\Delta\eta_{\mu\mu}$ cut where has the gap to dig.
- Good $m_{\mu\mu}$ agreement.

Large $\Delta\eta_{\mu\mu}$ region check

$\eta_{\mu\mu}$



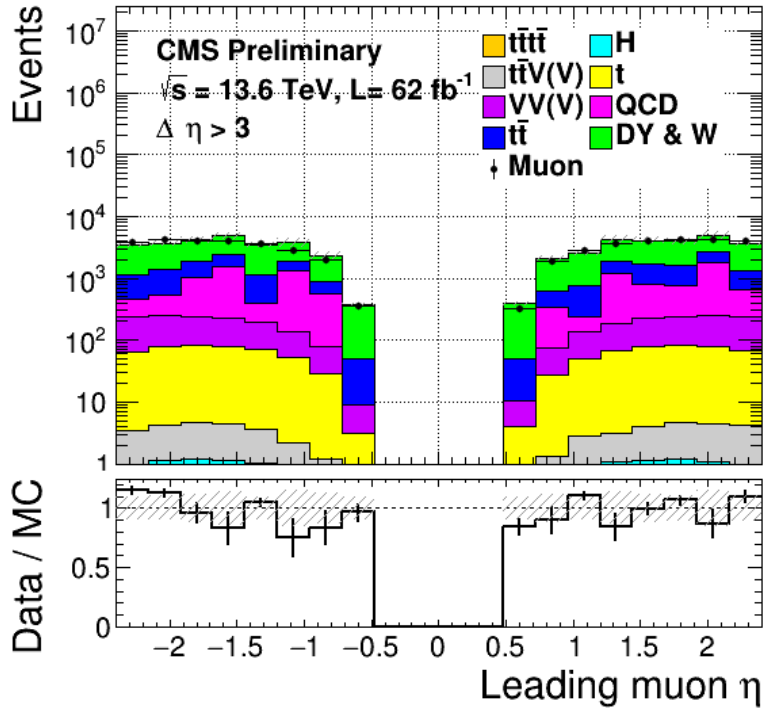
$m_{\mu\mu}$



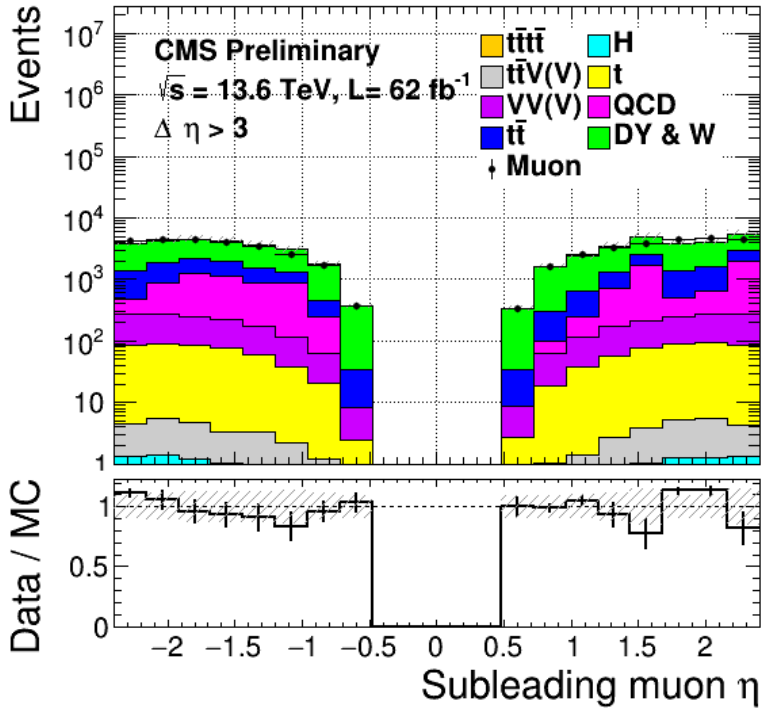
- Dimuon η also gets narrower range with the $\Delta\eta_{\mu\mu}$ cut.

Muon η

Leading Muon



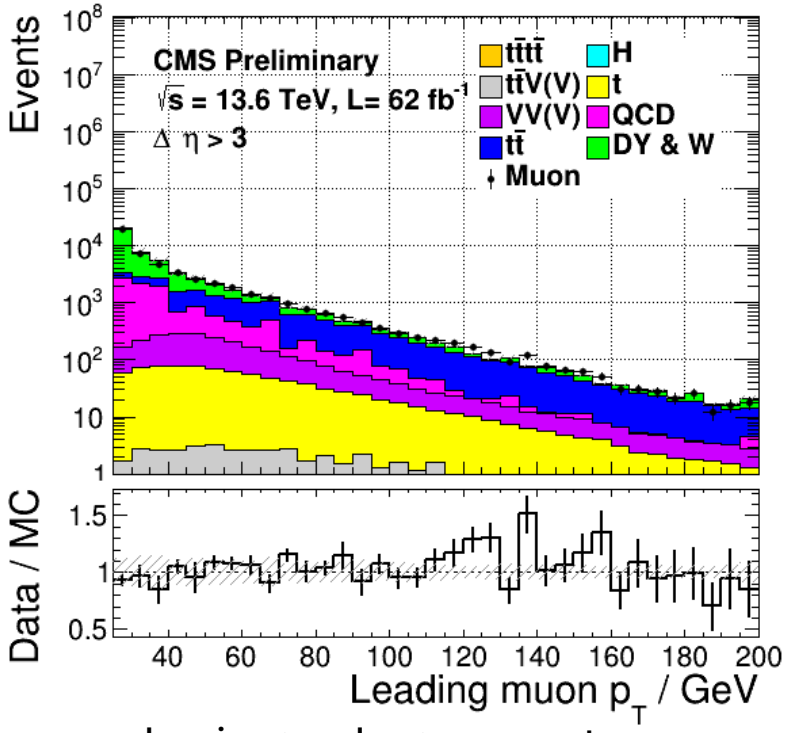
Subleading Muon



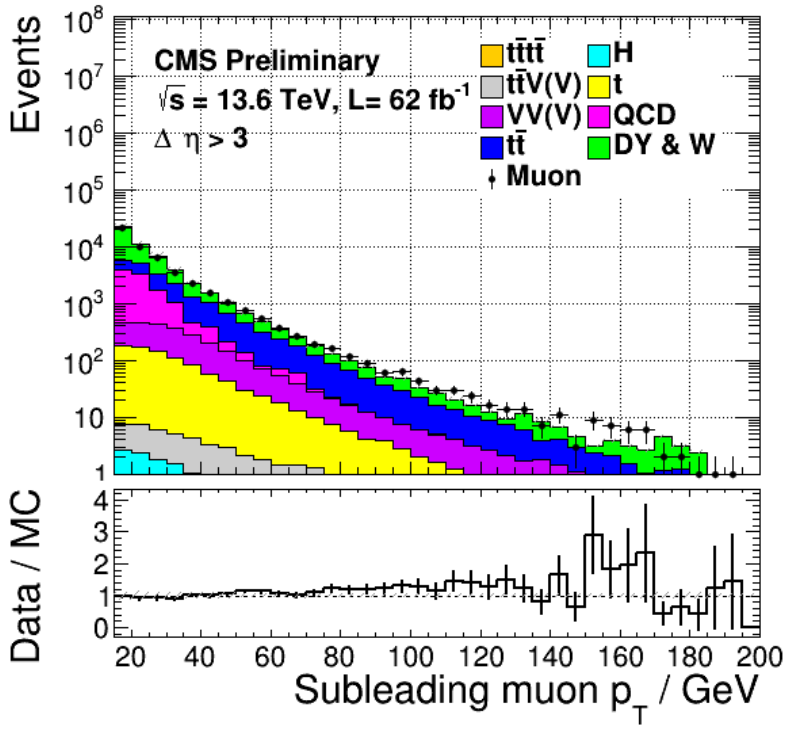
- No central region Muon contributing to large $\Delta\eta_{\mu\mu}$.
- I did not take explicit cut on this.
- Data-MC agreement is also fine.

Muon p_T

Leading Muon



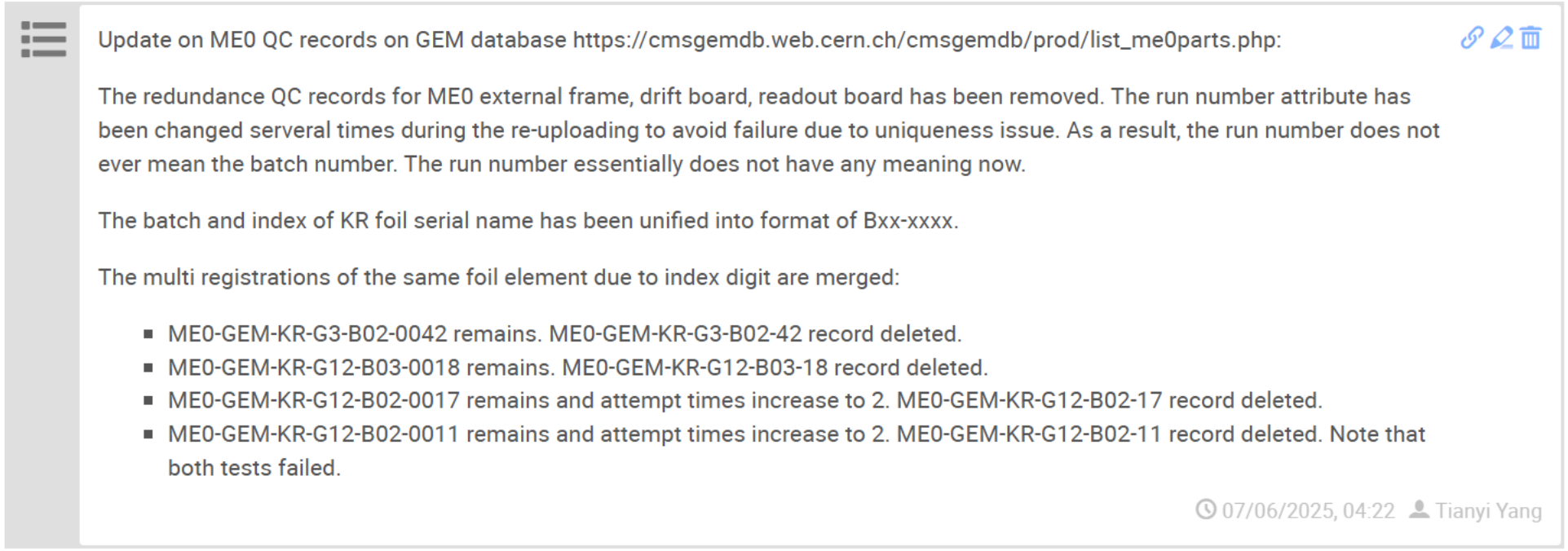
Subleading Muon



- Muon p_T also in good agreement.

GEM database

- Solve all the redundance record and naming convention in oracle database.
- https://indico.cern.ch/event/1555397/



Update on ME0 QC records on GEM database https://cmsgemdb.web.cern.ch/cmsgemdb/prod/list_me0parts.php:

The redundance QC records for ME0 external frame, drift board, readout board has been removed. The run number attribute has been changed serveral times during the re-uploading to avoid failure due to uniqueness issue. As a result, the run number does not ever mean the batch number. The run number essentially does not have any meaning now.

The batch and index of KR foil serial name has been unified into format of Bxx-xxxx.

The multi registrations of the same foil element due to index digit are merged:

- ME0-GEM-KR-G3-B02-0042 remains. ME0-GEM-KR-G3-B02-42 record deleted.
- ME0-GEM-KR-G12-B03-0018 remains. ME0-GEM-KR-G12-B03-18 record deleted.
- ME0-GEM-KR-G12-B02-0017 remains and attempt times increase to 2. ME0-GEM-KR-G12-B02-17 record deleted.
- ME0-GEM-KR-G12-B02-0011 remains and attempt times increase to 2. ME0-GEM-KR-G12-B02-11 record deleted. Note that both tests failed.

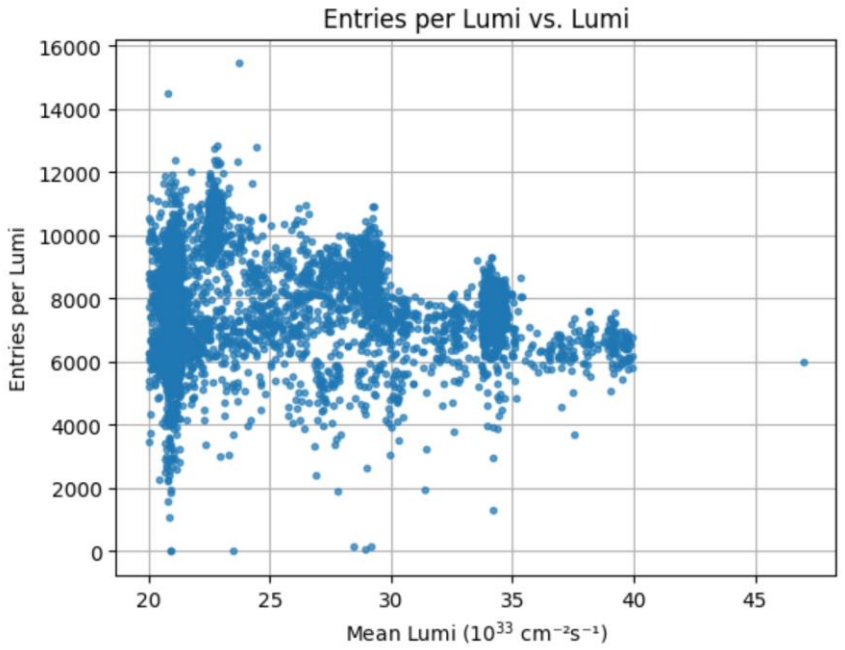
07/06/2025, 04:22 Tianyi Yang

GEM AutoDQM

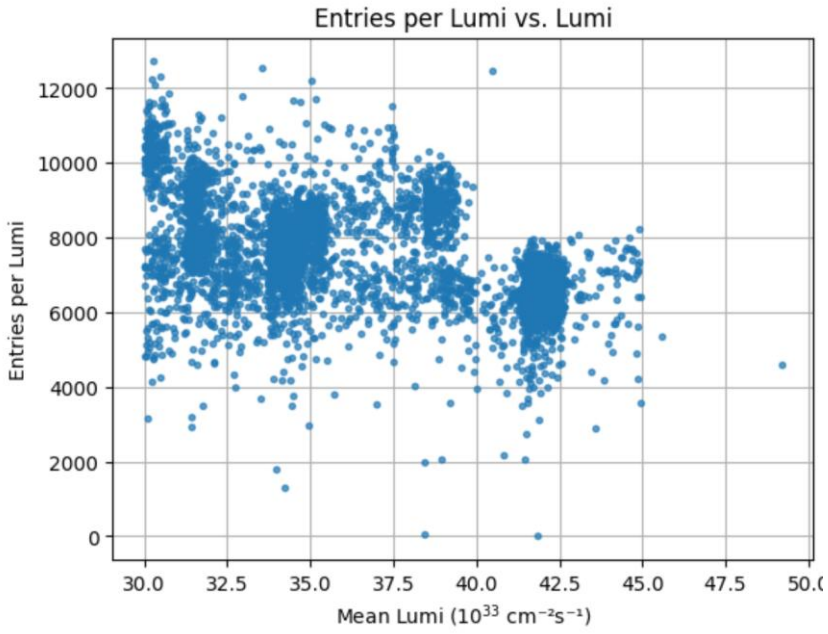
- ME downloaded:
  - `GEM/Digis/occ_GE11-M-L1`
  - `GEM/Digis/occ_GE11-M-L2`
  - `GEM/Digis/occ_GE11-P-L1`
  - `GEM/Digis/occ_GE11-P-L2`
  - From `/StreamExpress/Run2025C-Express-v1/DQMIO` on dial.
- Run2025C selection
  - From Run 392174.
  - At the time I downloaded, Run number up to 392542.
  - Require Collision25 run without GEM standby/bad/empty.

Different merging level test

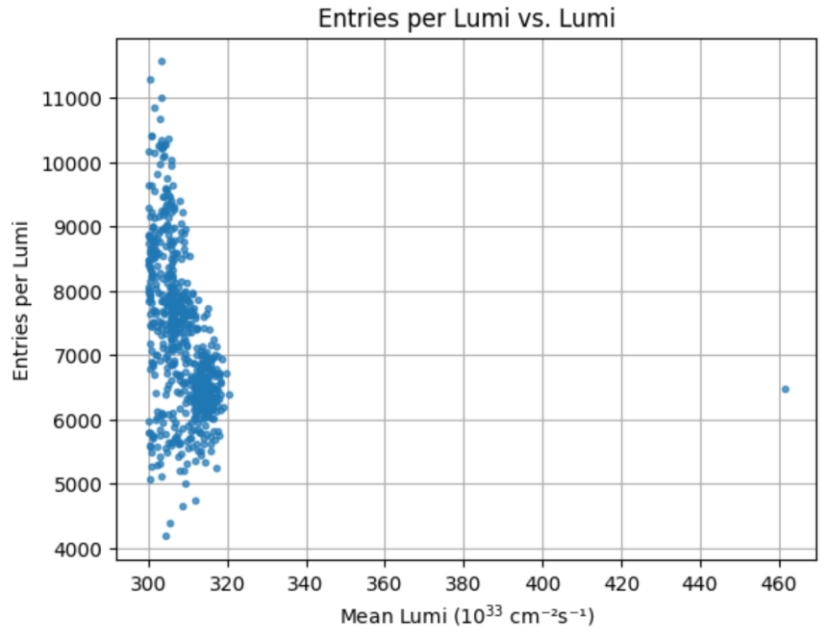
Merge to 20



Merge to 30

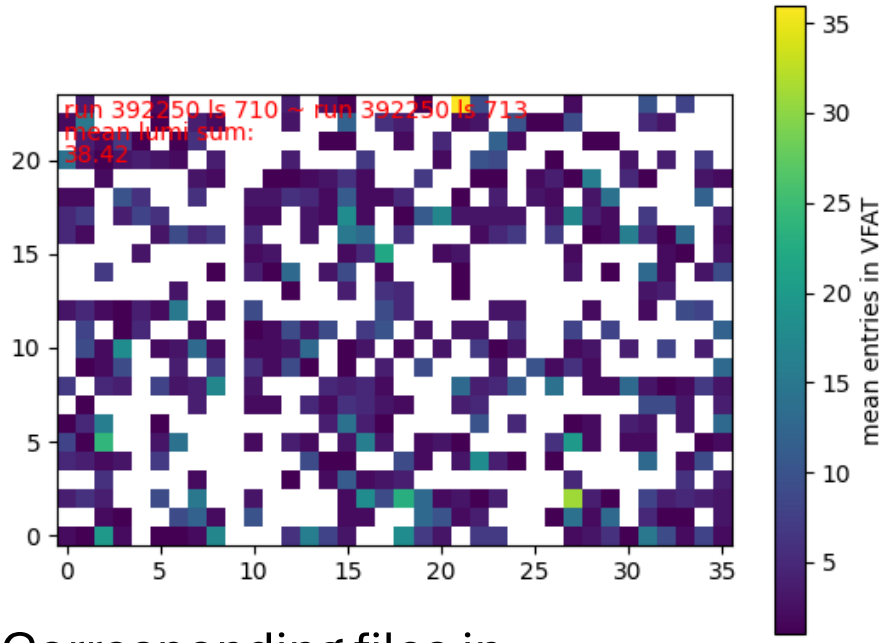


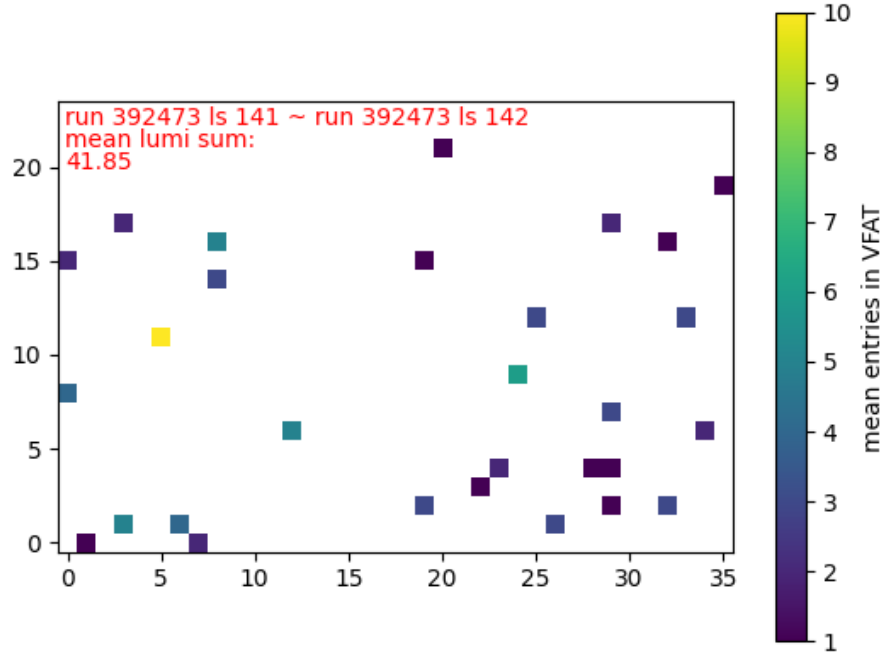
Merge to 300



- Merge to 30 seems already reach a good level of occupancy.
- Entry per lumi similar to up to 300.

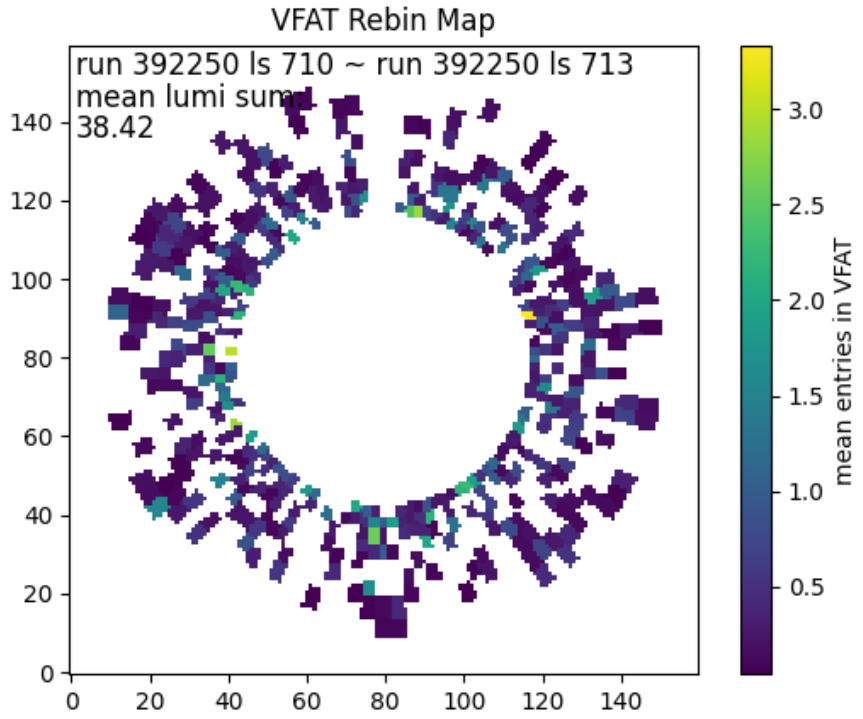
Sample lumi section merging histogram low occupancy example

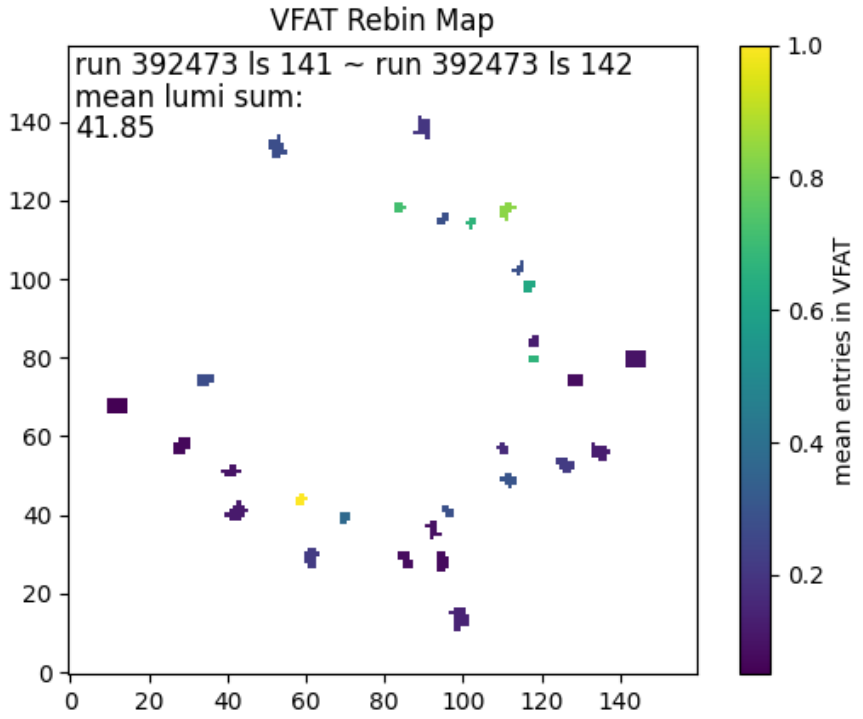




- Corresponding files in **/eos/user/t/tiyang/AutoDQM/data2025/lumi_30_data_GEM_Digis_occ_GE11-P-L1__StreamExpress_Run2025C-Express-v1_DQMIO_392174_392542.parquet.npz**
- Maybe should define a lower bound of entries per lumi to reject those superLS at 2000?

Image view

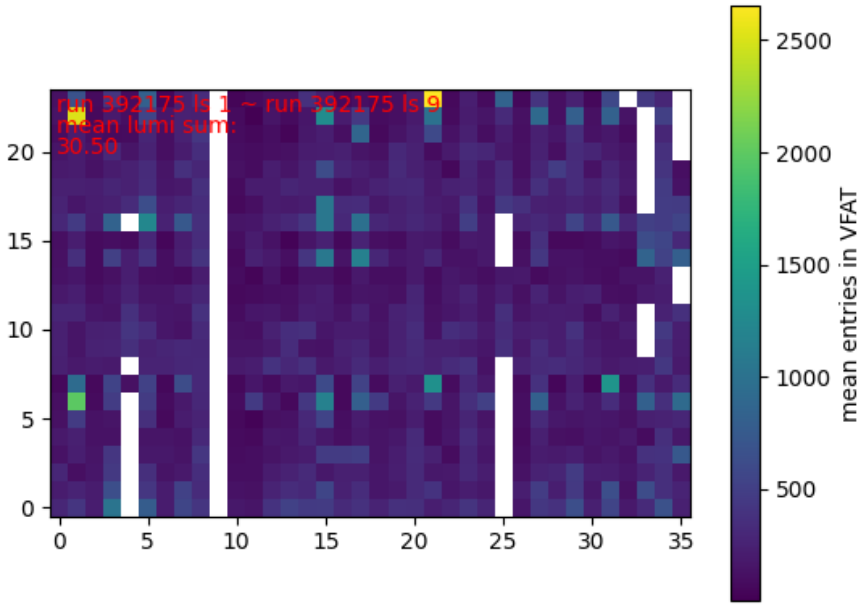




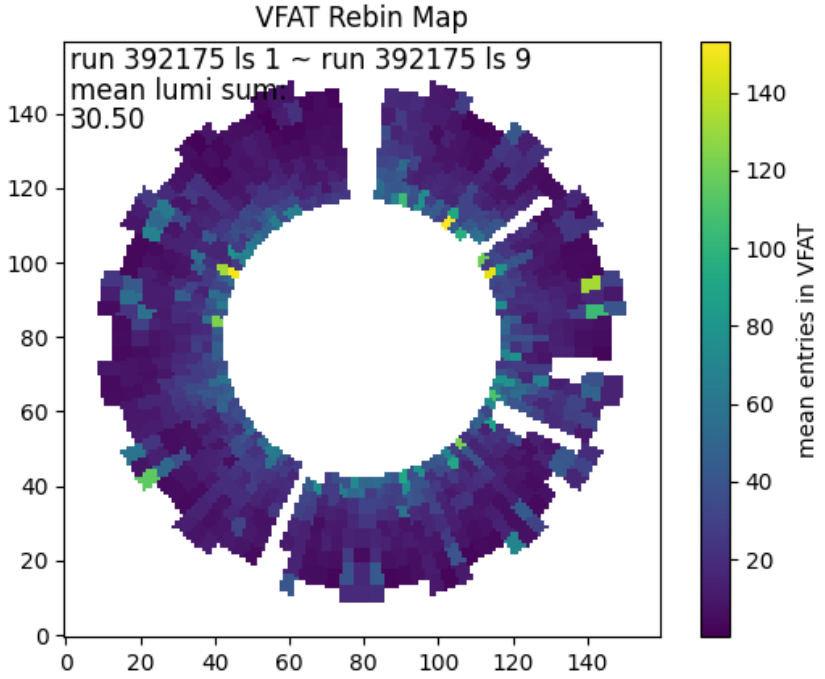
- Corresponding files in **/eos/user/t/tiyang/AutoDQM/data2025/lumi_30_GEM_Digis_occ_GE11-P-L1__StreamExpress_Run2025C-Express-v1_DQMIO_392174_392542.parquet.npz**

Sample lumi section merging ordinary case

Histogram



Image



- Most superLS are with good occupancy and looks like this.
- The off regions are consistent behavior, but can have run-wise difference.

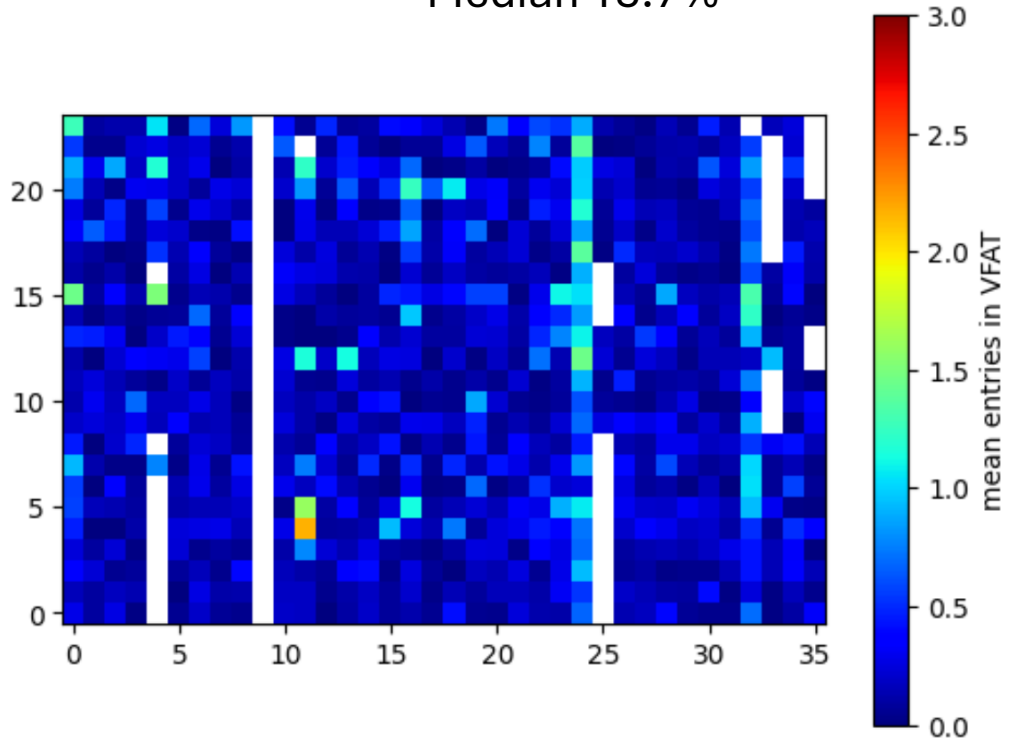
Performance using histogram or image

- Proposed overall $\text{loss} = \sum_{pixel}(|\Delta_{occ}| * \overline{occ})$
  - The points outside the detector will not be counted due to 0 occupancy in mean plot.
  - Whatever the binning (160*160 or 24*36), the total occupancy in the image or histogram is the same.
  - Presented in relative difference in percentage.

Comparison

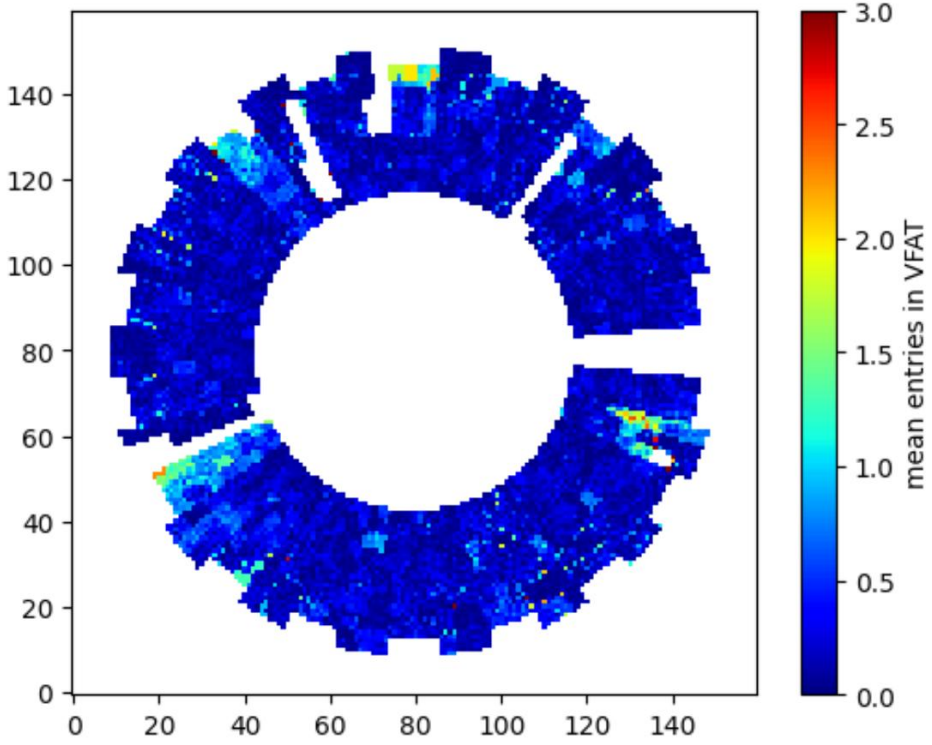
Histogram

Mean 19.8%
Median 15.7%



Image

Mean 19.1%
Median 15.6%



- Comparison of histogram and image validation loss using the same dataset partition for training.
- The mean image training loss is 18.8%, histogram training loss is 19.3%.

Backup

Cutflow table

	S0: global CR						
MC campaign	22PreEE		22PostEE			23PreBPix	23PostBPix
Drell-Yan	5405654.2+109092.4		17459445.4+321795.7			14391153.0+236067.3	7230418.4+125205.0
$t\bar{t}$	50377.8+2613.1		148030.2+9629.7			115171.1+6746.0	61298.1+3646.1
tV	2659.2+2677.9+114.2		8014.3+7990.0+347.1			6112.9+6091.8+260.1	3235.6+3214.1+137.4
$t\bar{t}V(V)$	3.01+20.7+159.6+85.8		10.1+60.7+500.0+258.0			13.6+48.9+363.0+191.1	4.3+25.6+193.3+102.9
Higgs	0.32+0.39+18.82+40.93+33.85+18.6		0+0+57.28+121.73+100.76+57.71			0.91+1.14+43.34+93.96+78.29+43.05	0.24+0.30+23.4+49.92+41.33+22.52
W+jets	2000.4		4612.1			5824.6	2937.7
$VV(V)$	5219.7+4320.4+4141+955.9+1087.9+2815.5+34.1+38.5+17.5		15598.7+12961.0+12376.5+2912.1+3168.2+8529.8+105.6+119.2+53.7+23.8			11986.5+9897.8+9535.0+2205.6+2555.0+6461.3+78.5+88.1+39.9+17.9	6356.9+5248.6+5069.8+1159.2+1325.9+3431.1+41.8+47.3+21.1+9.51
$t\bar{t}t\bar{t}$	4.0		12.2			9.16	4.8
QCD	141505		406849			353025	184440
Era	2022C	2022D	2022E	2022F	2022G	2023C	2023D
data	3334682	2203560	4412469	12549031	2189719	14043412	7639872

DY2L50+ DY2L10-50

DL+SL

$tW^- + \bar{t}W^+$ +TZQB

TTVV+ttZ10-50+ttZ50+ttW

TTWH+TTZH+ttHBB+ttHnonBB+thq+thw

WW2L+WZ2L+ZZ2L2Q+ZZ2L2Nu+ZZ4L+WZ3L+WWW+WWZ+WZZ+ZZZ