

# First look at $ttHH$ Run 3 data & Data-MC comparison

Yang Tianyi

# Data

- Checking begins with the di-muon region
  - **Muon0** and **Muon1** primary datasets contain **all** the single muon and double muon triggered events.
  - Other primary datasets *MuonEG* and *EGamma0/1* will be considered later for **2e**,  **$e\mu$**  and **multi-lepton** channels.
- Muon0 and Muon1 record different events in the same run and luminosity blocks.
- **No double counting** of events between Muon0 and Muon1.
- Using Run2023D era for the checking
  - Luminosity: **9.451 fb<sup>-1</sup>**

# MC normalization and cut scheme

- **genWeight** applied for generator level nominal weights.
- Total events is normalized to the luminosity before preselection:

$$w_{lumi} = \frac{\text{lumi} \times \text{cross-section}}{\sum \text{genWeight}}$$

- The genWeight distribution for each MC dataset is checked as well.

- Cross-section:

|                      | Cross-section /pb | Additional branching ratio                                |
|----------------------|-------------------|---|
| Drell-Yan 2L         | 6221.3            |   |
| $t\bar{t}$ DL        | 96.9              |   |
| $t\bar{t}$ SL        | 404.0             |   |
| $W \rightarrow l\nu$ | 63199.9           |   |
| WW                   | 173.4             | $W \rightarrow l\nu$ : 21.75%                             |
| WZ                   | 54.3              | $W \rightarrow qq$ : 67.41%, $Z \rightarrow ll$ : 6.729%  |
| ZZ                   | 16.7              | $Z \rightarrow qq$ : 69.91%, $Z \rightarrow \nu\nu$ : 20% |

- Cross-section from AN2022\_103
- Additional branching ratio from PDG

RECAP

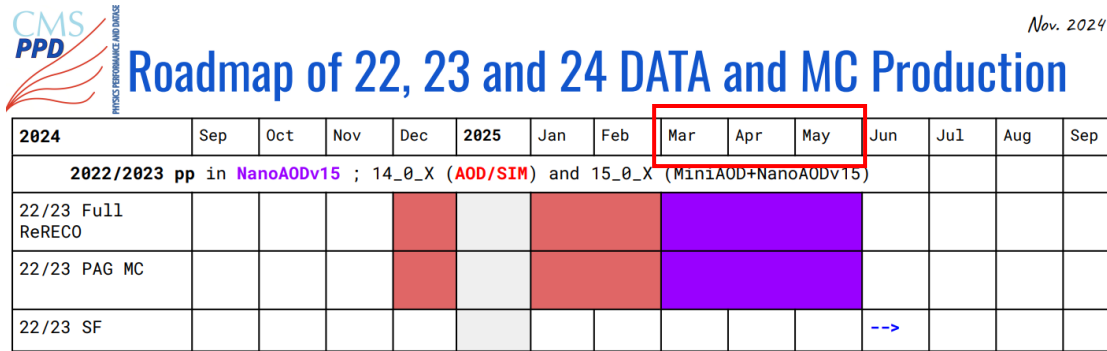
# Dataset on htop machine

- Data:
  - Run2023C: /data1/common/NanoAOD/data/Run2023C 895GB 17.794 fb<sup>-1</sup>
  - Run2023D: /data1/common/NanoAOD/data/Run2023D 454GB 9.451 fb<sup>-1</sup>
- MC:
  - 2023C: /data1/common/NanoAOD/mc/Run3Summer23NanoAODv12/ 895GB
  - 2023D: /data1/common/NanoAOD/mc/ Run3Summer23BPixNanoAODv12/ 426GB
- Primary dataset:
  - EGamma0, EGamma1, Muon0, Muon1
- MC channels:
  - DYto2L, WtoLNu, TTto2L2Nu, TTtoLNu2Q, WWto2L2Nu, WZto2LQ, ZZto2L2Q, ZZto2L2Nu
- Estimated total size: ~1TB 10 fb<sup>-1</sup>, then total Run3 in future would need ~20TB, for future signal MC generation and more dataset, better to have more than 30TB for data storage.

# V15 Nano?

- V12 Nano I downloaded following Pdmv twiki recommendation.
- V15 Nano mentioned in the CMS week.
- The PPD group report shows NanoAODv15 MC and reReco data seems to be ready next year:

This version would still be only for 2023C but not 2023D (with BPix)?



- 2022 and 2023:
  - MC Campaigns: **RunIII2022Summer24** and **RunIII2023Summer24**
    - Ideal detector (no BPix, no EE+)
    - ECAL PF Calib ongoing, PFHC to follow
    - Planned to be launched by December;
  - Data Reprocessing:
    - Pending ECAL PF, PFHC (same as above)
  - Preparations: [\[gitlab\]](#)

- I will need to check with them to make sure I correctly understand the case.

# Object and event preselection

RECAP

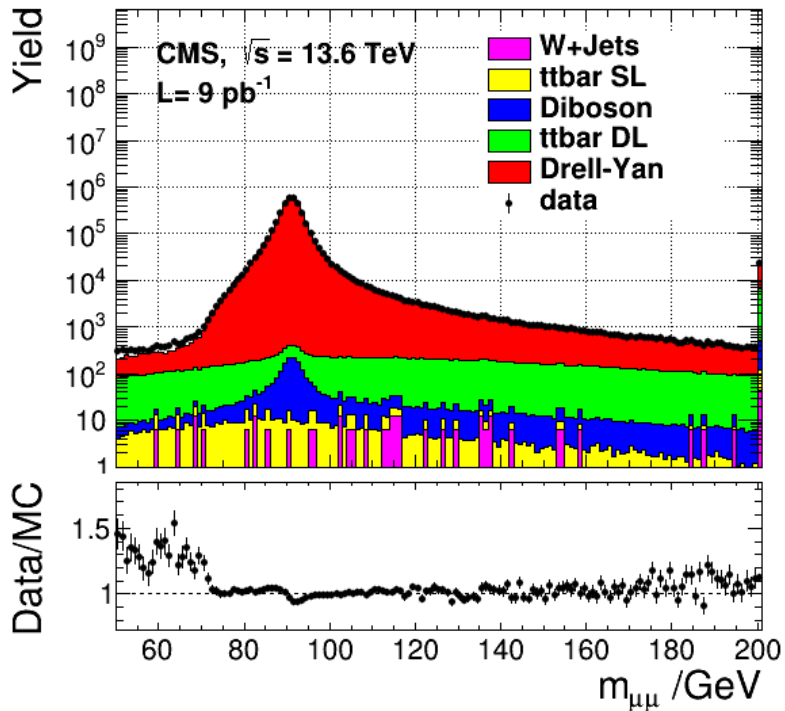
- Muon requirement:
  - Tight ID
  - Tight Iso ( $I_{0.4} < 0.15$ )
  - Kinematic cut:  $p_T > 35\text{GeV}$ ,  $|\eta| < 2.4$
- Events required:
  - Select events with exactly 2 muons.
- **Test multi-lepton scenario and electron channels in the next step.**

# rDataFrameLight package

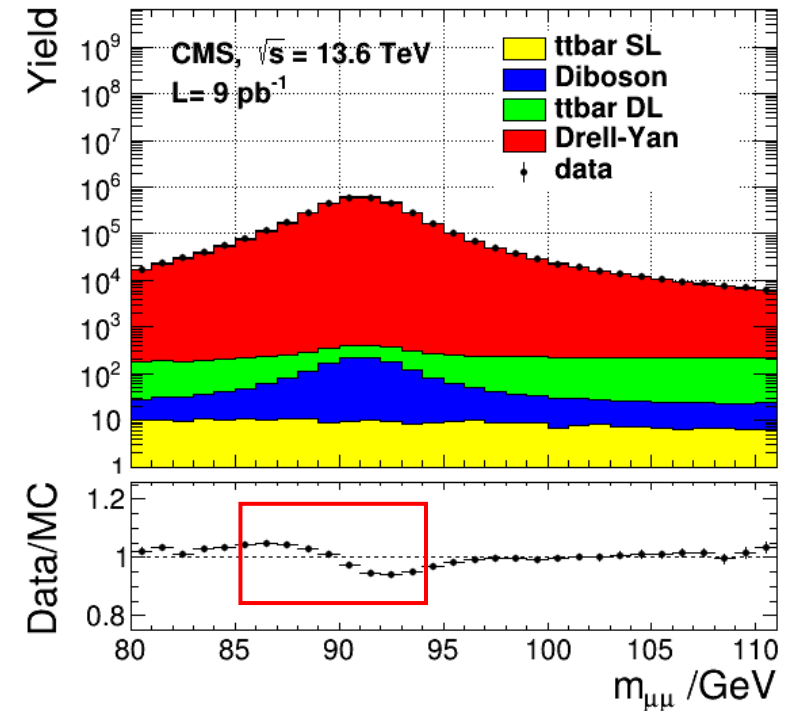
- Aim: providing a light package with RDataFrame to analyze *ttHH*. Welcome to use and give feedback!
- Package on htop machine: `/home/tiyang/public/rDataFrameLight/source`
- Design:
  - Dataset fetching by *SampleControl*
  - RDataFrame processing set by *CutControl*
  - Histogram IO by *HistControl*
  - Draw histograms with Hanyang style by *PlotControl*
- Json header from:
  - `wget https://github.com/nlohmann/json/releases/download/v3.11.3/json.hpp`
- Some example operations for getting and plotting histograms in `utils/`.

# $m_{\mu\mu}$ data-MC comparison

$m_{\mu\mu}$  of data & MC



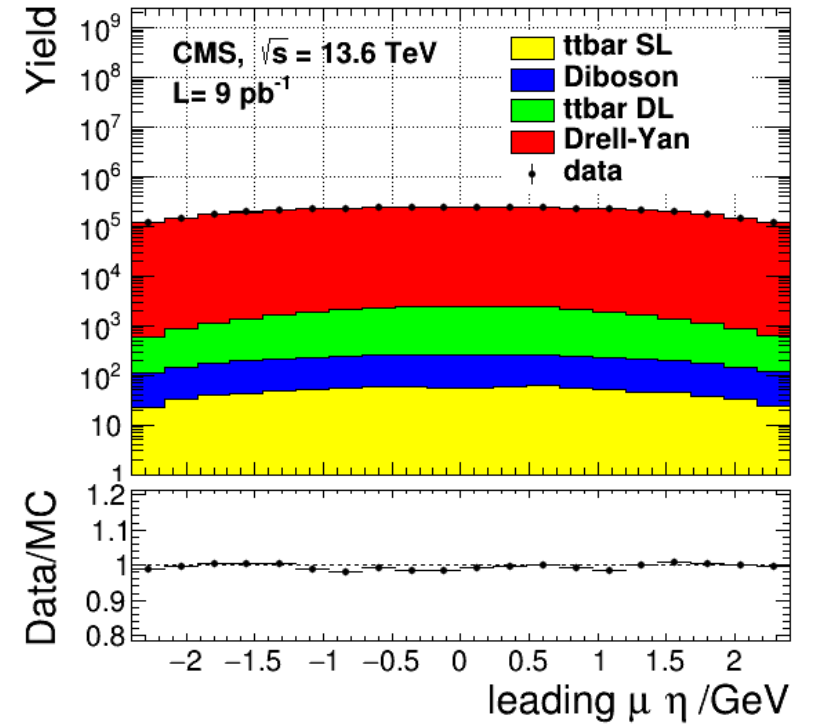
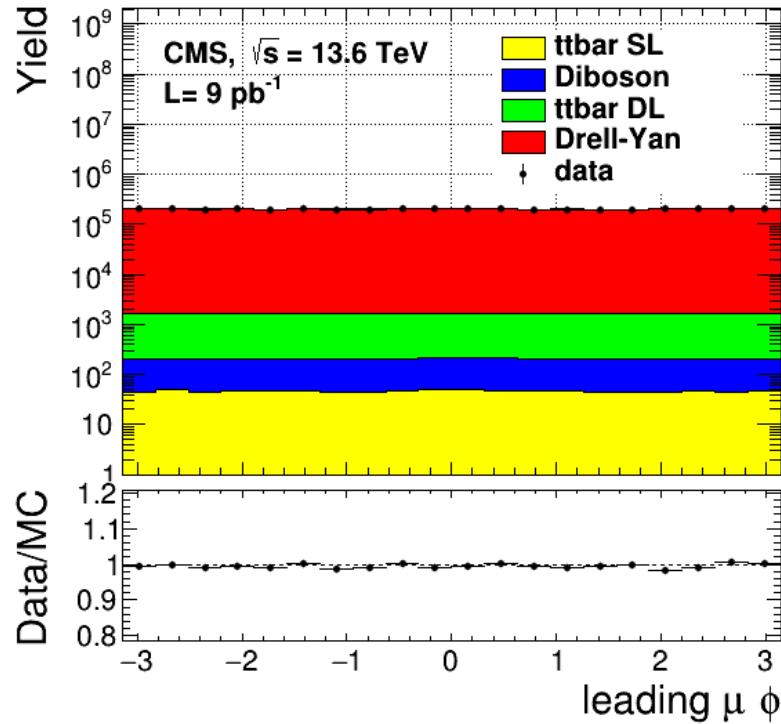
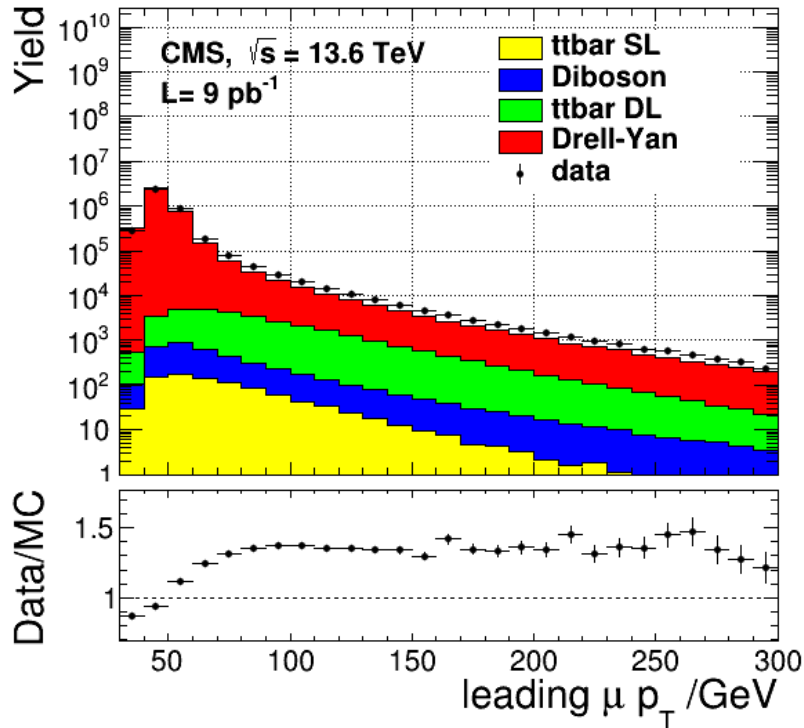
Zoom-in



- W+jet and  $t\bar{t}$  SL contribution very low, ignore in later plots.
- Missing QCD contribution in the low mass region.
- Zoom-in plot shows clearly the  $m_{\mu\mu}$  shift between data and MC.

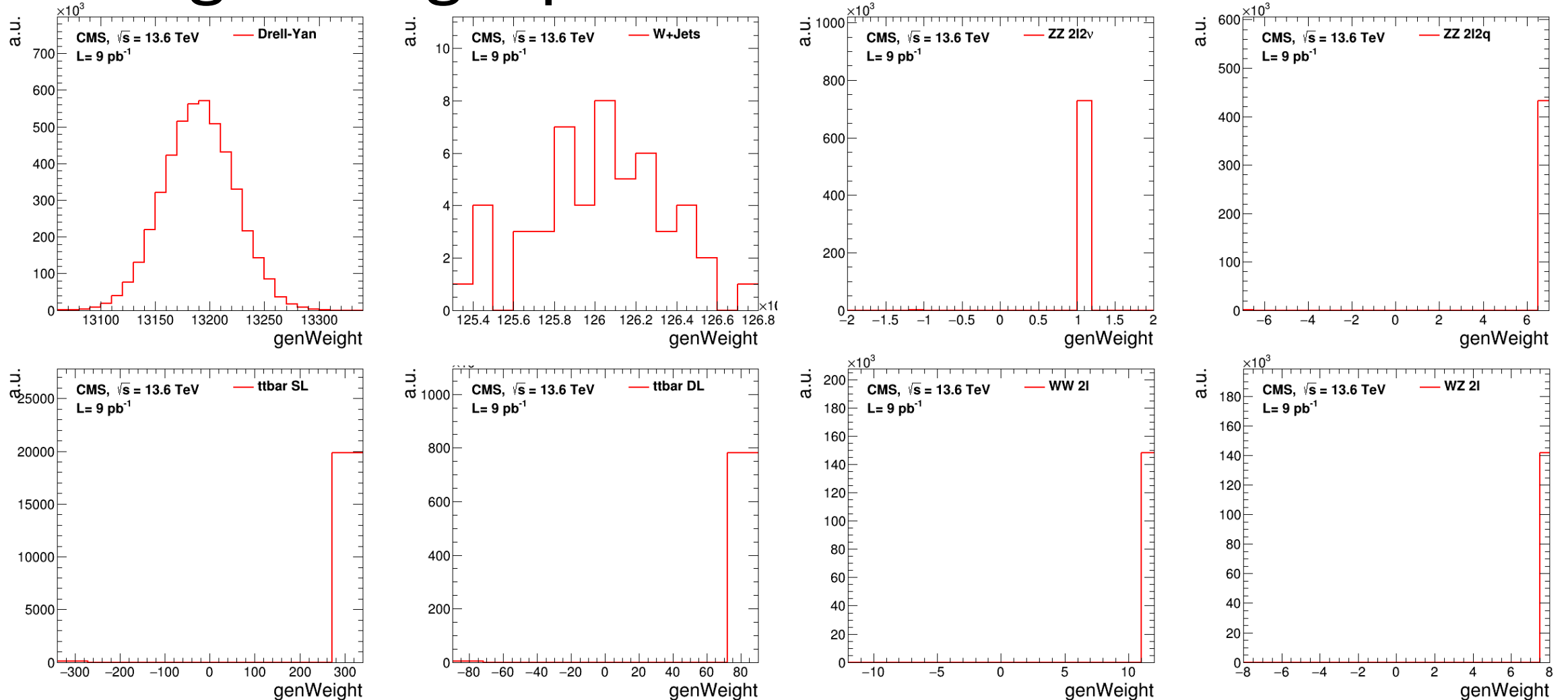


# Leading muon kinematic variables



- Difference exists in  $p_T$
- Good matching in  $\phi$  and  $\eta$ .
- Might be solved by Rochester momentum scale correction.

# MC genWeight plot



- Drell-Yan and W+jets genWeight are all positive, with small variations.
- Diboson and  $t\bar{t}$  have a small number of events with negative weight.

# Todo next

- Contact Pdmv group for the MC sample version recommendation.
- Check the unmatching behavior in kinematic distributions with Muon POG for confirmation.
- Check electron channels and multi-lepton channels.