Data Analysis of Drell-Yan Process in the CMS Experiment at the LHC

Department of Physics and Astrophysics
University of Delhi, Delhi-110007

Anuj Raghav

M.Sc. Physics, Semester-4

University Roll Number: 23222762018

Supervisor: Dr. Arun Kumar

2024-25

Outline

- Overview
- Signal and background
- Data analysis chain
- Working algorithm
- Kinematic variables
- Analysis strategy
- Yield table
- Result

Overview

- An analysis of the Drell-Yan(DY) process in the electron-positron (di-electron) final state using the CMS detector at the LHC, CERN.
- Suppress background contributions from top quark production, single-top production, di-boson processes.
- Improve signal purity through optimized event selection and invariant mass window techniques.

Recap

- Studied the theoretical foundations of the DY process and radiation interactions with matter.
- Reviewed the CMS detector components, focusing on ECAL for electron identification.
- Conducted preliminary analysis using Monte Carlo simulations from CMS Open Data, generating key lepton kinematic distributions.
- Identified main background processes (top quark pairs, single top, di-bosons) to prepare for signal optimization this semester.

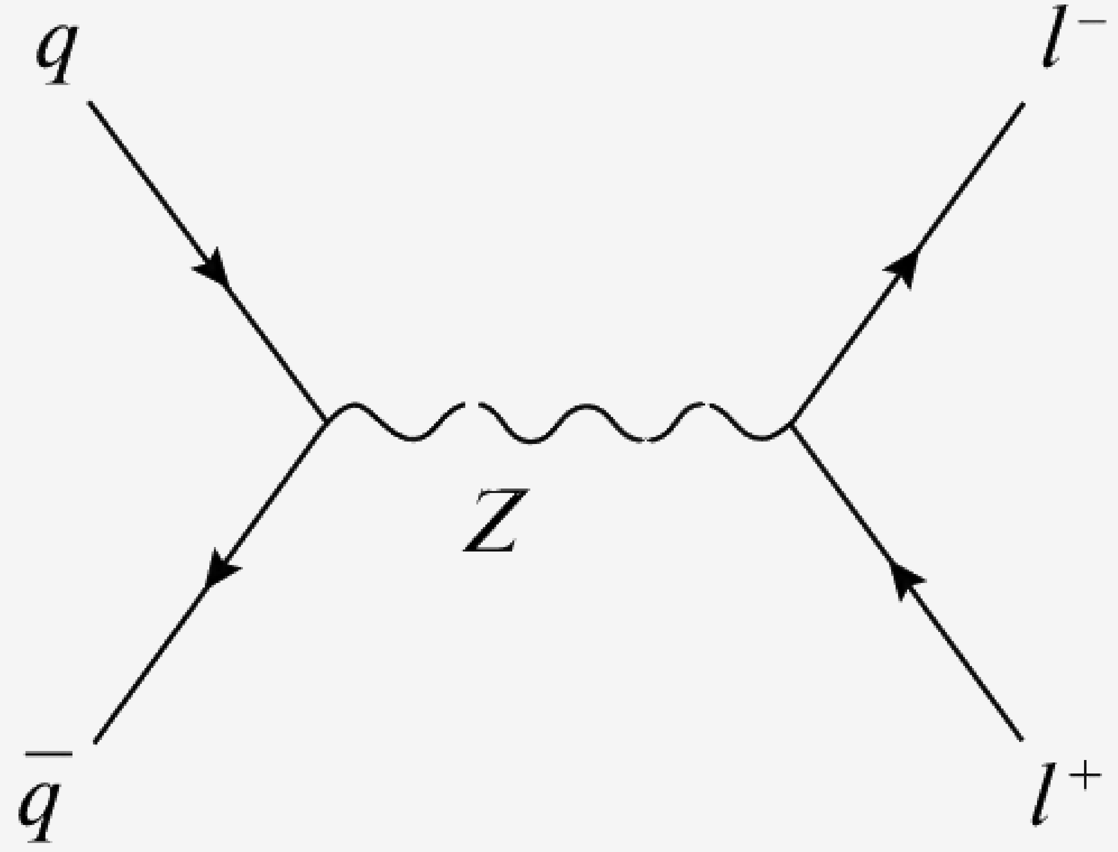
Current Semester Objectives

- Characterize signal and background contributions using CMS opendata and Monte Carlo simulations.
- Define and optimize event selection criteria to improve Drell-Yan signal purity.
- Apply cross-section weighting and luminosity scaling for accurate event yield normalization.
- Enhance data-driven analysis techniques to improve overall signal significance.

Signal

Drell-Yan process:

- Quark from one proton and an antiquark from another proton annihilate to form a *Z boson,* which subsequently decays into a lepton-antilepton pair.
- This study focuses on the DY process in the di-electron final state.
- Cross-section: 6189.39 pb



$$q\bar{q} \longrightarrow Z \longrightarrow e^+e^-$$

Backgrounds-1/2

Top quarks production: (87.31 pb)

- Occurs via quark-antiquark annihilation, producing a top and antitop quark.

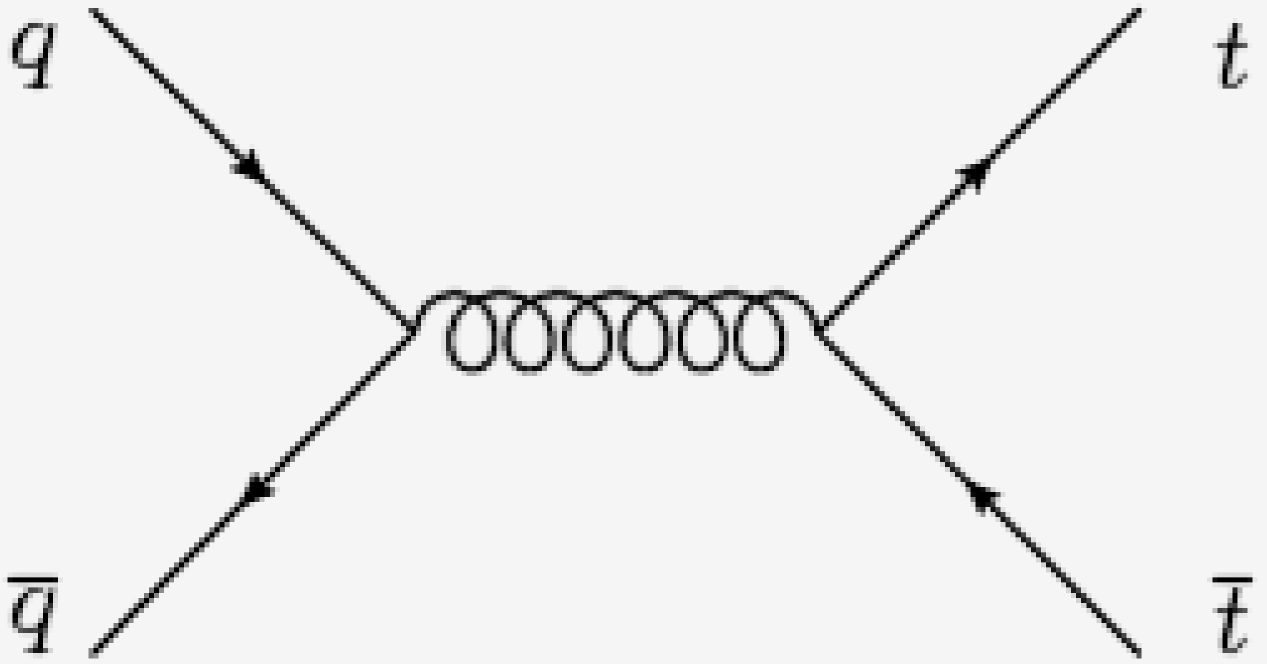
$$q\bar{q} \longrightarrow t\bar{t}$$

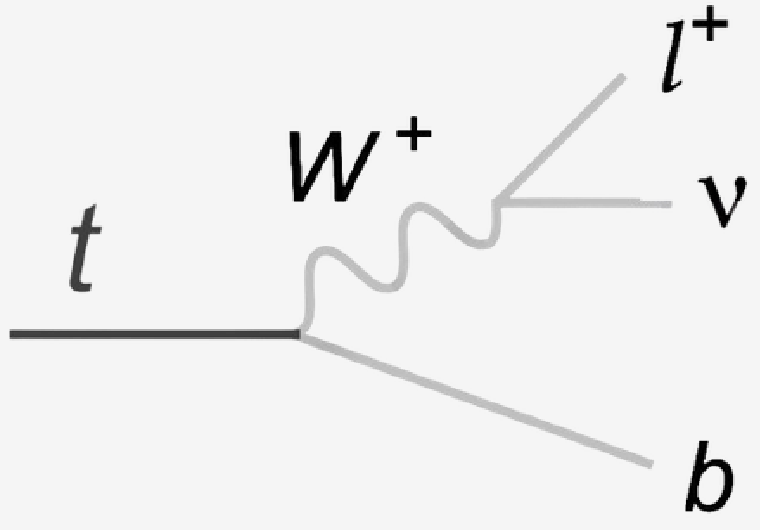
- Each top quark decays almost exclusively into a *W boson* and a *bottom quark*

$$t \longrightarrow W^+b \qquad \bar{t} \longrightarrow W^-\bar{b}$$

- Leptonic decays of W bosons can mimic the di-electron signature.

$$W^- \longrightarrow e^-\bar{\nu}_e \qquad W^+ \longrightarrow e^+\nu_e$$

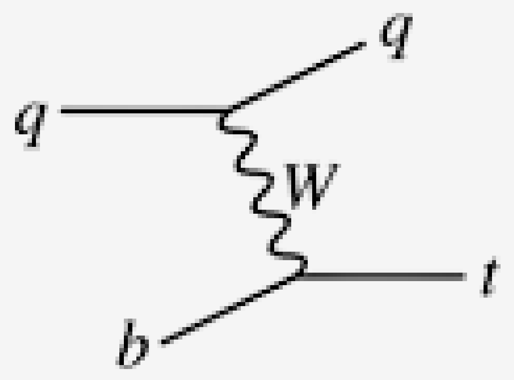


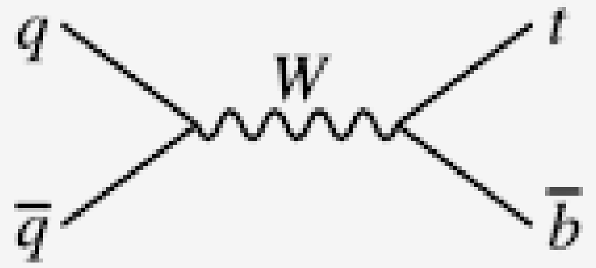


Backgrounds-2/2

Single top quark production

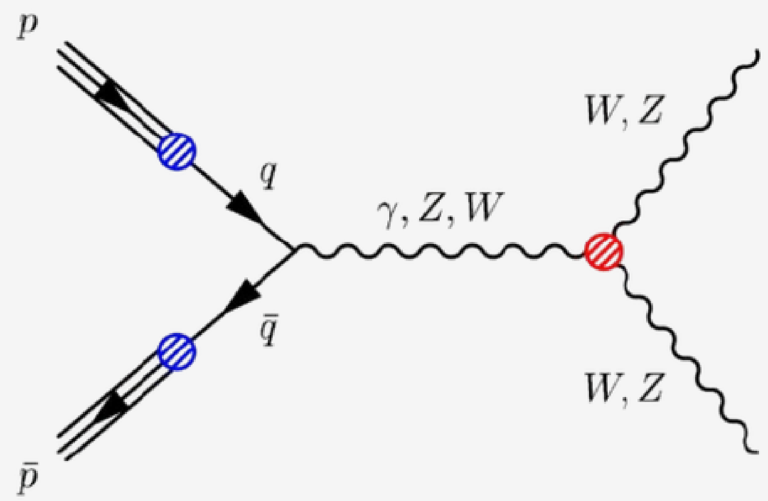
- Single top production (145.27 pb) is a less significant background.
- Produces a top quark, which decays to a W boson and a b-jet; W can decay leptonically.
- Contributes background mainly when jets or photons are misidentified as electrons.



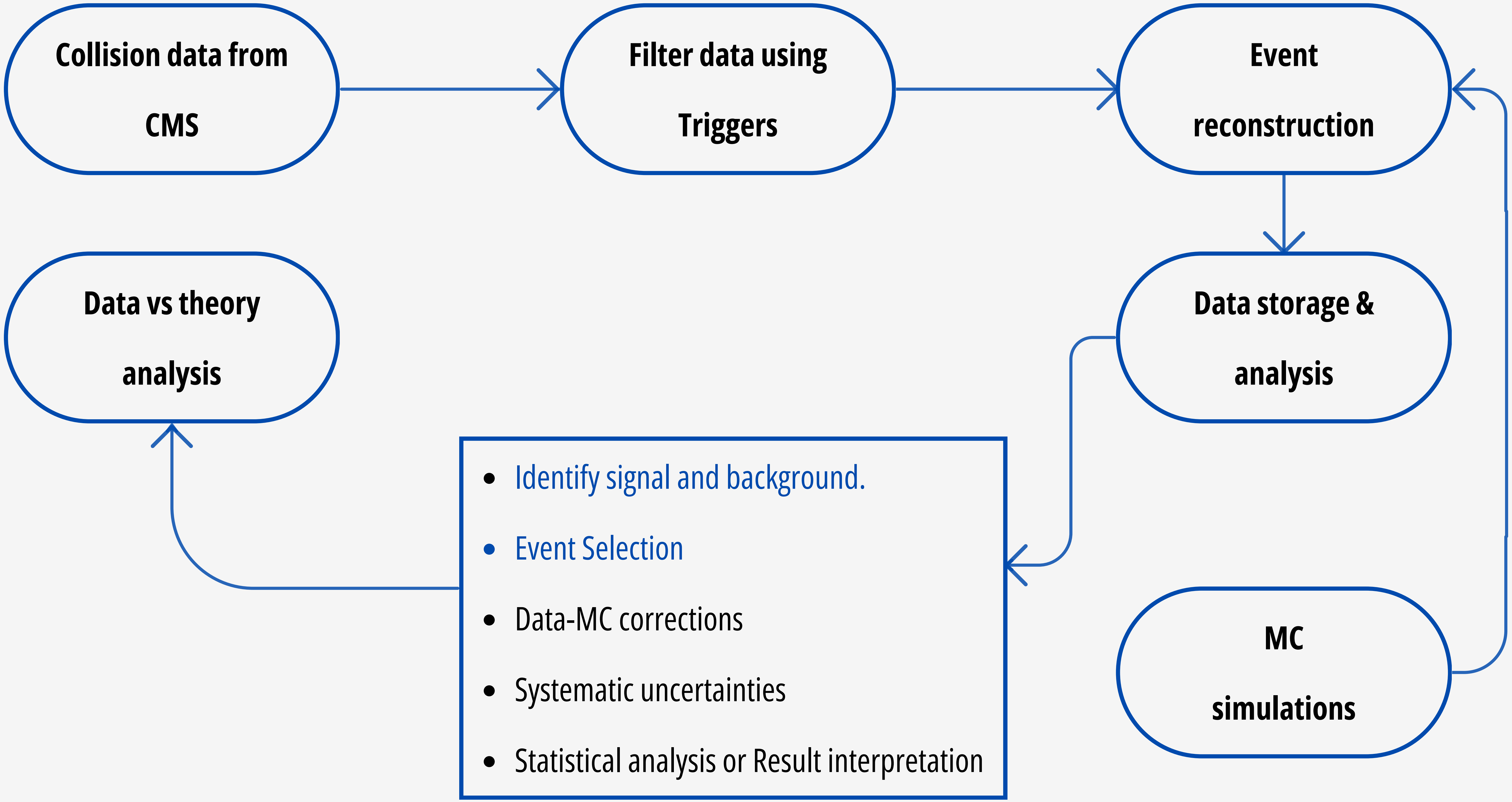


Di-boson production

- Processes like WW (47.13 pb), WZ (16.52 pb) , and ZZ (12.18 pb) can produce lepton pairs, mimicking the di-electron final state.
- Events with additional neutrinos or missing energy can complicate event selection.

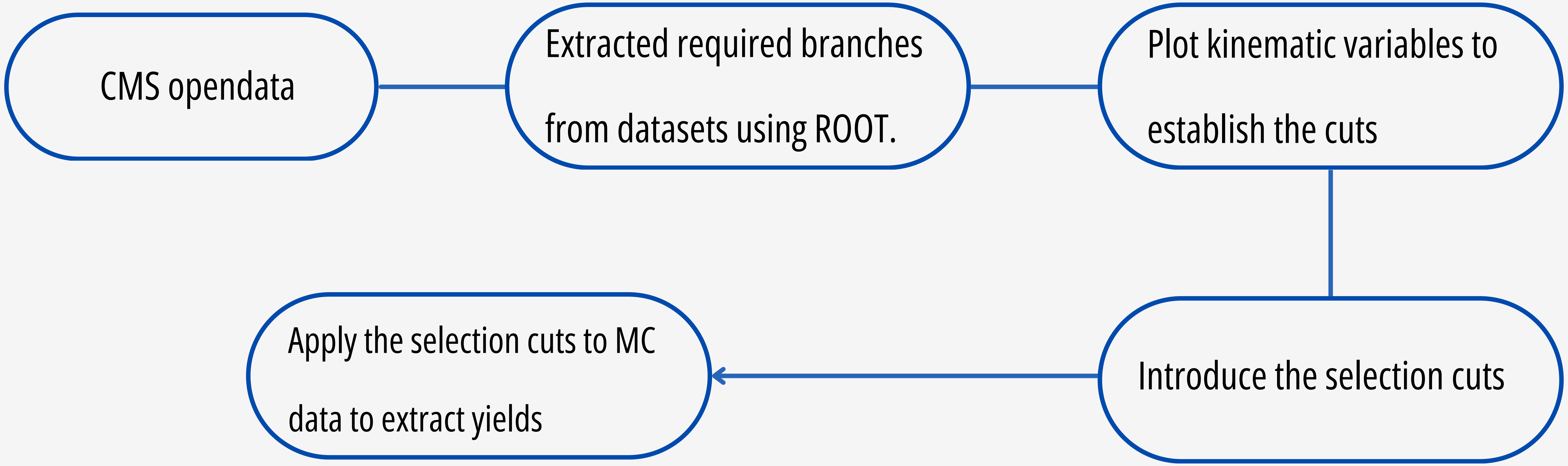


Data Analysis Chain



Working Algorithm

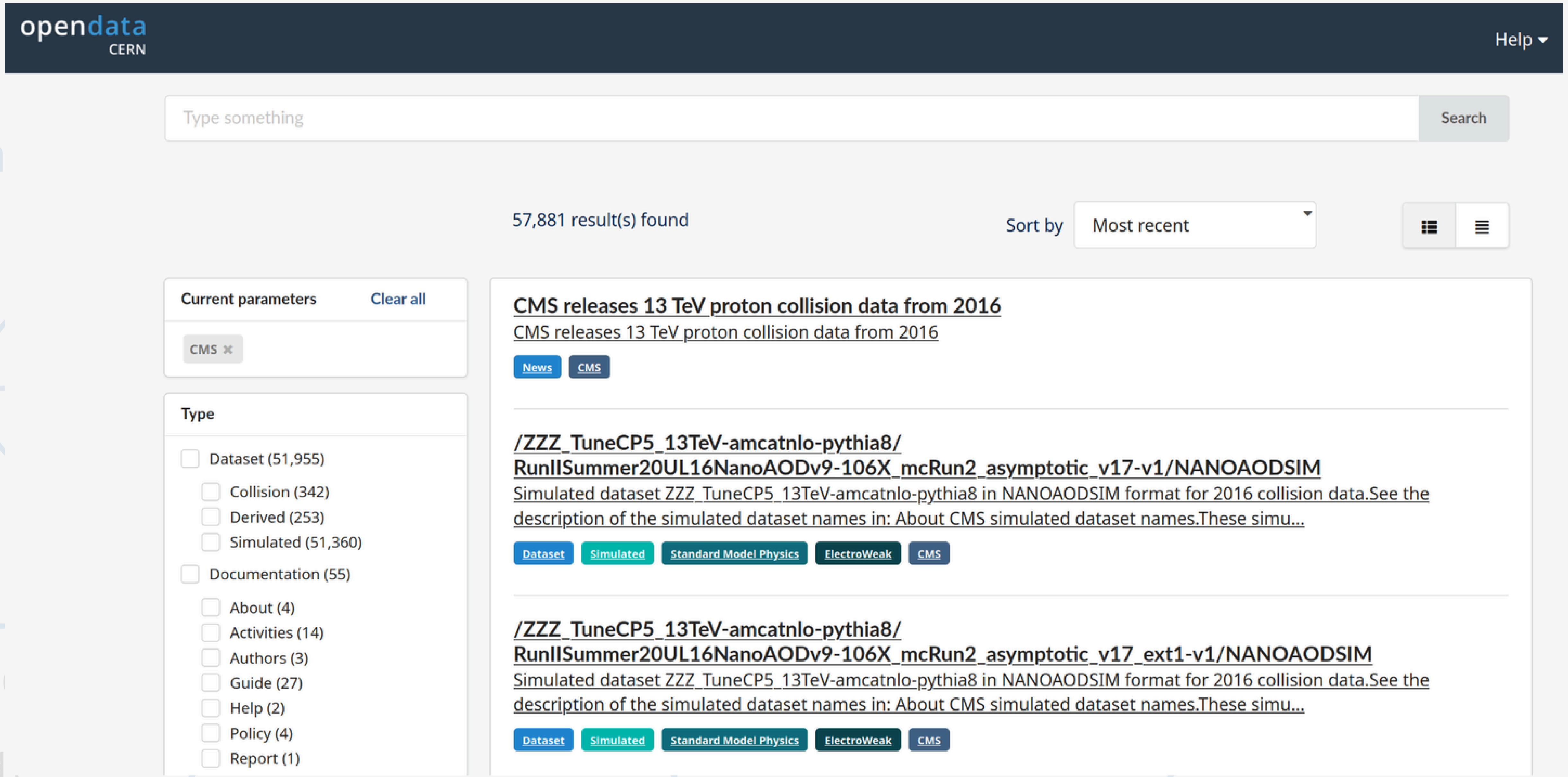
- The analysis is performed using the ROOT data analysis framework developed by CERN.
- Datasets are obtained from the CMS Open Data portal.
- The study is based on MC simulations corresponding to pp collisions at $\sqrt{s} = 13$ TeV recorded in 2016.



Working Algorithm

- The analysis is performed using the ROOT data analysis framework developed by CERN.
- Datasets are obtained from the CMS
- The study is based on MC simulation

CMS opendata



Apply the selection
data to extract yield.

Working Algorithm

2. `branch_extractor.C`

- Extracts a subset of branches from multiple ROOT files (`DYtoLL1.root` to `DYtoLL61.root`) containing CMS Open Data or Monte Carlo events.
- Disables all branches by default, selectively enables relevant physics branches.
- Creates a slimmed ROOT file (`DYtoLL_ext1.root`) with only needed variables.
- Optimizes memory and speeds up subsequent analyses.

eloped by CERN.

√s = 13 TeV recorded in 2016.

CMS opendata

Extracted required branches from datasets using ROOT.

Plot kinematic variables to

Apply the selection cuts to MC data to extract yields

```
void branch_extractor() {
    // Define branches to keep
    std::vector<std::string> branches_to_keep = {
        "run", "luminosityBlock", "event",
        "Electron_mvaFall17V2Iso_WP80", "Electron_mvaFall17V2Iso_WP90", "Electron_mvaFall17V2Iso_WPL",
        "Electron_mvaFall17V2noIso_WP80", "Electron_mvaFall17V2noIso_WP90", "Electron_mvaFall17V2noIso_WPL",
        "Electron_charge", "Electron_cutBased", "Electron_jetIdx", "Electron_pdgId", "Electron_photonIdx",
        "Electron_tightCharge", "Electron_phi", "Electron_pt", "Electron_r9", "Electron_scEtOverPt",
        "Electron_mass", "Electron_dxy", "Electron_dxyErr", "Electron_dz", "Electron_dzErr", "Electron_eCorr",
        "Electron_eInvMinusPInv", "Electron_energyErr", "Electron_eta", "Electron_hoe", "nElectron",
        "Electron_dEscaleDown", "Electron_dEscaleUp", "Electron_dEsigmaDown", "Electron_dEsigmaUp", "Electron_deltaEtaSC",
        "Electron_dr03EcalRecHitSumEt", "Electron_dr03HcalDepth1TowerSumEt", "Electron_dr03TkSumPt", "Electron_dr03TkSumPtHEEP",
        "CaloMET_phi", "CaloMET_pt", "GenMET_phi", "GenMET_pt", "MET_MetUnclustEnUpDeltaX", "MET_MetUnclustEnUpDeltaY",
        "MET_covXX", "MET_covXY", "MET_covYY", "MET_phi", "MET_pt", "MET_significance", "MET_sumEt", "MET_sumPtUnclustered",
        "Pileup_sumLOOT", "PuppiMET_phi", "PuppiMET_phiJERDown", "PuppiMET_phiJERUp", "PuppiMET_phiJESDown", "PuppiMET_phiJESUp",
        "PuppiMET_phiUnclusteredDown", "PuppiMET_phiUnclusteredUp", "PuppiMET_pt", "PuppiMET_ptJERDown", "PuppiMET_ptJERUp",
        "PuppiMET_ptJESDown", "PuppiMET_ptJESUp", "PuppiMET_ptUnclusteredDown", "PuppiMET_ptUnclusteredUp", "PuppiMET_sumEt",
        "Electron_genPartIdx", "Flag_goodVertices", "Pileup_nTrueInt", "Pileup_pudensity", "Pileup_gpudensity", "Pileup_nPU",
        "Pileup_sumEOOT", "Pileup_sumLOOT", "fixedGridRhoFastjetCentralChargedPileUp",
        "Jet_area", "Jet_btagCSVV2", "Jet_btagDeepB", "Jet_btagDeepCvB", "Jet_btagDeepCvL",
        "Jet_btagDeepFlavB", "Jet_btagDeepFlavCvB", "Jet_btagDeepFlavCvL", "Jet_btagDeepFlavQG",
        "Jet_eta", "Jet_phi", "Jet_pt", "Jet_mass", "Jet_electronIdx1", "Jet_electronIdx2", "Jet_jetId"
    };
```

Working Algorithm

5. `Superimposed_plots.C`

- This ROOT macro (`superimposed_plots.C`) processes multiple CMS Open Data Monte Carlo samples to overlay **normalized kinematic distributions** of electron pairs (e.g., from Drell-Yan, $t\bar{t}$, VV backgrounds).
- Applies **tight electron selection** (`mvaFall17V2Iso_WP90`) and requires **exactly two opposite-charge electrons** to reconstruct dilepton quantities.
- Plots distributions of `m_{ll}`, `p_T^{ll}`, projected MET, `Δϕ_{ll}`, and leading/subleading electron `p_T` and `η`, comparing across samples.
- Saves each overlaid distribution as a `.png` file, labeled and styled for presentation.

```
TLorentzVector e1, e2;
e1.SetPtEtaPhiM(Electron_pt[i1], Electron_eta[i1], Electron_phi[i1], 0.000511);
e2.SetPtEtaPhiM(Electron_pt[i2], Electron_eta[i2], Electron_phi[i2], 0.000511);
TLorentzVector dilepton = e1 + e2;

double mll = dilepton.M();
double ptll = dilepton.Pt();
double dphill = fabs(e1.DeltaPhi(e2));

hist_mll->Fill(mll, scale);
hist_ptll->Fill(ptll, scale);
hist_met->Fill(projected_MET, scale);
hist_dphill->Fill(dphill, scale);

// Determine leading/subleading by pT
int lead = (Electron_pt[i1] > Electron_pt[i2]) ? i1 : i2;
int sub  = (lead == i1) ? i2 : i1;

hist_pt_lead->Fill(Electron_pt[lead], scale);
hist_pt_sub->Fill(Electron_pt[sub], scale);
hist_eta_lead->Fill(Electron_eta[lead], scale);
hist_eta_sub->Fill(Electron_eta[sub], scale);
```

Plot kinematic variables to establish the cuts

```
// Create histograms
TH1F *h_mll     = new TH1F(Form("h_mll_%zu", i), "m_{ll} of the electron-positron pair; m_{ll} [GeV]; Normalized Events", 50, 50, 150);
TH1F *h_ptll    = new TH1F(Form("h_ptll_%zu", i), "p_{T}^{ll} of the electron-positron pair; p_{T}^{ll} [GeV]; Normalized Events", 25, 0, 100);
TH1F *h_met     = new TH1F(Form("h_met_%zu", i), "Projected MET for the electron-positron pair; Projected MET [GeV]; Normalized Events", 25, 0, 100);
TH1F *h_dphill  = new TH1F(Form("h_dphill_%zu", i), "#Delta#phi_{ll} between the electron-positron pair; #Delta#phi_{ll}; Normalized Events", 30, 0, 3.14);
TH1F *h_pt_lead  = new TH1F(Form("h_pt_lead_%zu", i), "Leading Electron p_{T}; p_{T}^{lead} [GeV]; Normalized Events", 25, 0, 100);
TH1F *h_pt_sub   = new TH1F(Form("h_pt_sub_%zu", i), "Subleading Electron p_{T}; p_{T}^{sub} [GeV]; Normalized Events", 25, 0, 100);
TH1F *h_eta_lead = new TH1F(Form("h_eta_lead_%zu", i), "Leading Electron #eta; #eta^{lead}; Normalized Events", 30, -3, 3);
TH1F *h_eta_sub  = new TH1F(Form("h_eta_sub_%zu", i), "Subleading Electron #eta; #eta^{sub}; Normalized Events", 30, -3, 3);
```

Working Algorithm

```cpp
if (tightElectrons.size() != 2) continue;

int i1 = tightElectrons[0];
int i2 = tightElectrons[1];

if (Electron_pdgId[i1] * Electron_pdgId[i2] < 0) {
    eventAfterstage1++;

    int lead = (Electron_pt[i1] > Electron_pt[i2]) ? i1 : i2;
    int sublead = (lead == i1) ? i2 : i1;

    if ((Electron_pt[lead] > 25) && (Electron_pt[sublead] > 20) &&
        (fabs(Electron_eta[lead]) < 2.5) && (fabs(Electron_eta[sublead]) < 2.5)) {

        eventAfterstage2++;

        TLorentzVector el1, el2;
        el1.SetPtEtaPhiM(Electron_pt[lead], Electron_eta[lead], Electron_phi[lead], 0.000511);
        el2.SetPtEtaPhiM(Electron_pt[sublead], Electron_eta[sublead], Electron_phi[sublead], 0.000511);
        TLorentzVector dilepton = el1 + el2;

        double mll = dilepton.M();
        double ptll = dilepton.Pt();
        double dphill = fabs(el1.DeltaPhi(el2));
```

```cpp
if (mll > 60 && mll < 120) {
    eventAfterstage3++;

    if (projected_MET < 25) {
        eventAfterstage4++;

        if (ptll < 40) {
            eventAfterstage5++;

            if (dphill > 2.5) {
                eventAfterstage6++;
```

Apply the selection cuts to MC data to extract yields

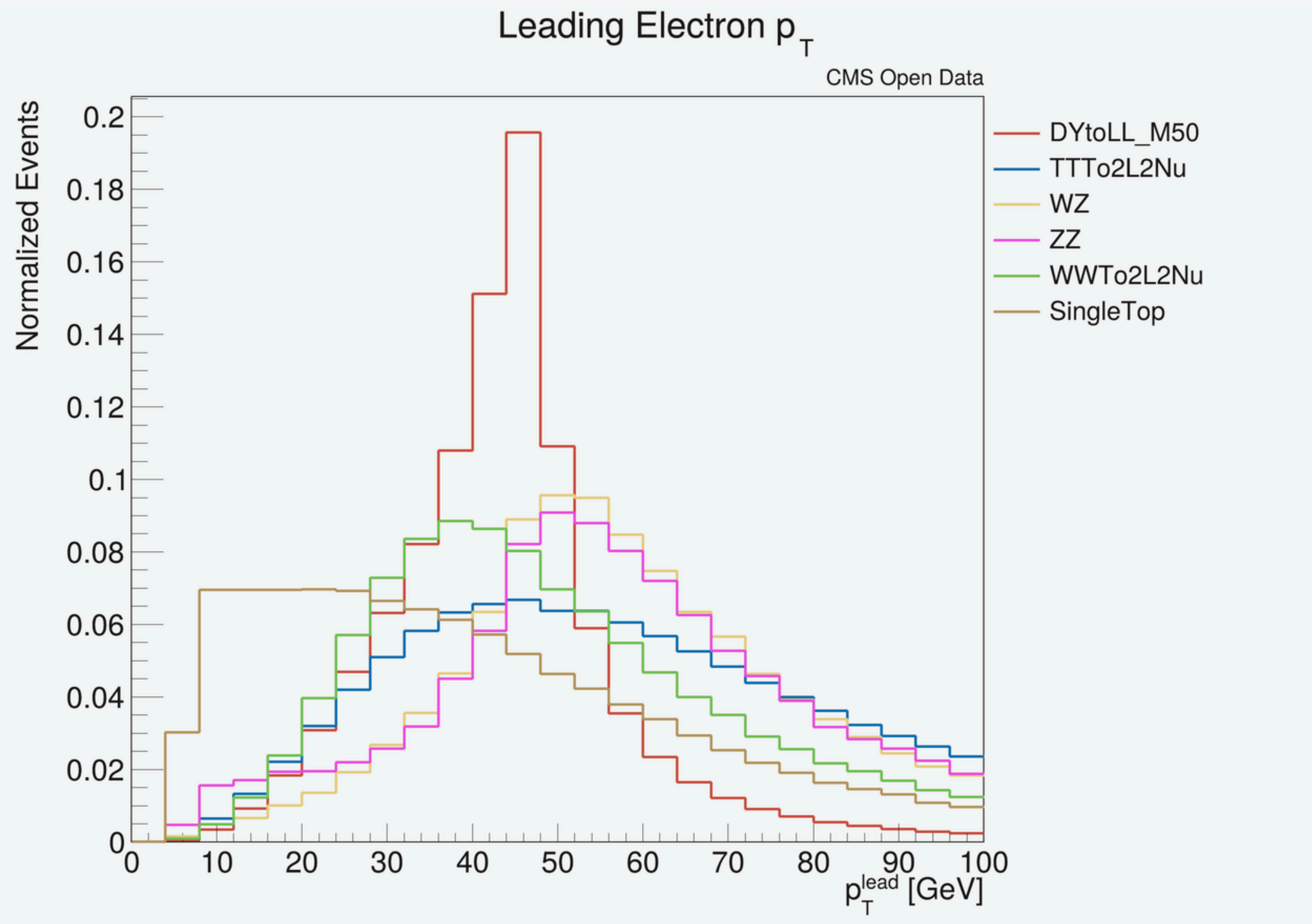
```cpp
cout << "\n\nCut Flow results for DYtoLL_M50.root:\n" << endl;
cout << "Total events in file "<<nEntries<<endl;
cout << "Events after nElectron == 2 and opposite charge: " << eventAfterstage1 << endl;
cout << "Events after |η| < 2.5 and leading pₜ > 25 GeV, subleading pₜ > 20 GeV: " << eventAfterstage2 << endl;
cout << "Events after dilepton mass 60 < mₗₗ < 120 GeV: " << eventAfterstage3 << endl;
cout << "Events after projected MET < 25 GeV: " << eventAfterstage4 << endl;
cout << "Events after pₜ^ll < 40 GeV: " << eventAfterstage5 << endl;
cout << "Events after |Δφ_ll| > 2.5: " << eventAfterstage6 << endl;
cout << "\nFinal events passing all cuts: " << eventAfterstage6 << endl;
cout << "\n\n";
```

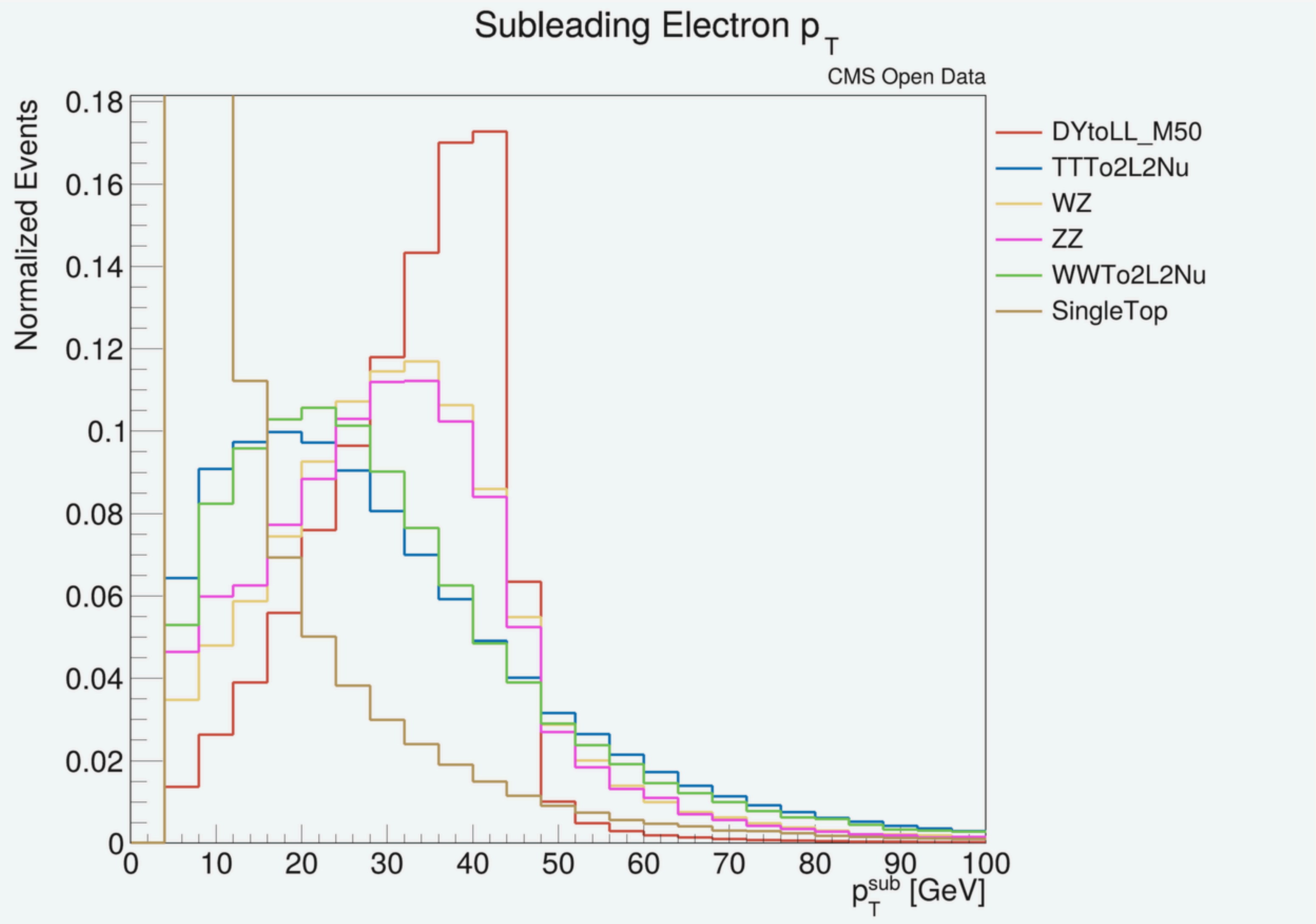
Kinematic Variables-1/4

Some relevant kinematic variables useful for our analysis are :

- **Transverse momentum**

$$p_T = \sqrt{p_x^2 + p_y^2}$$



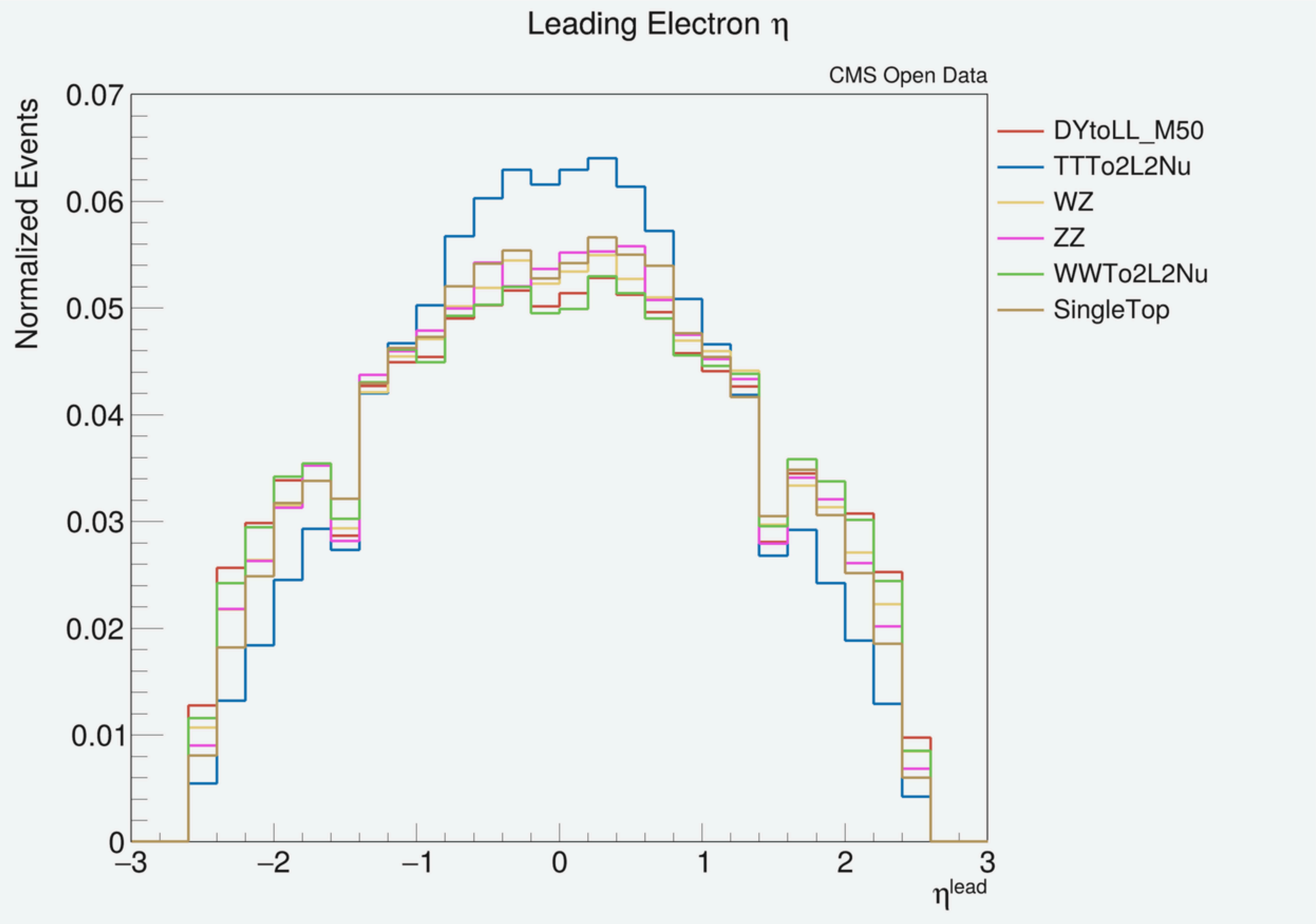


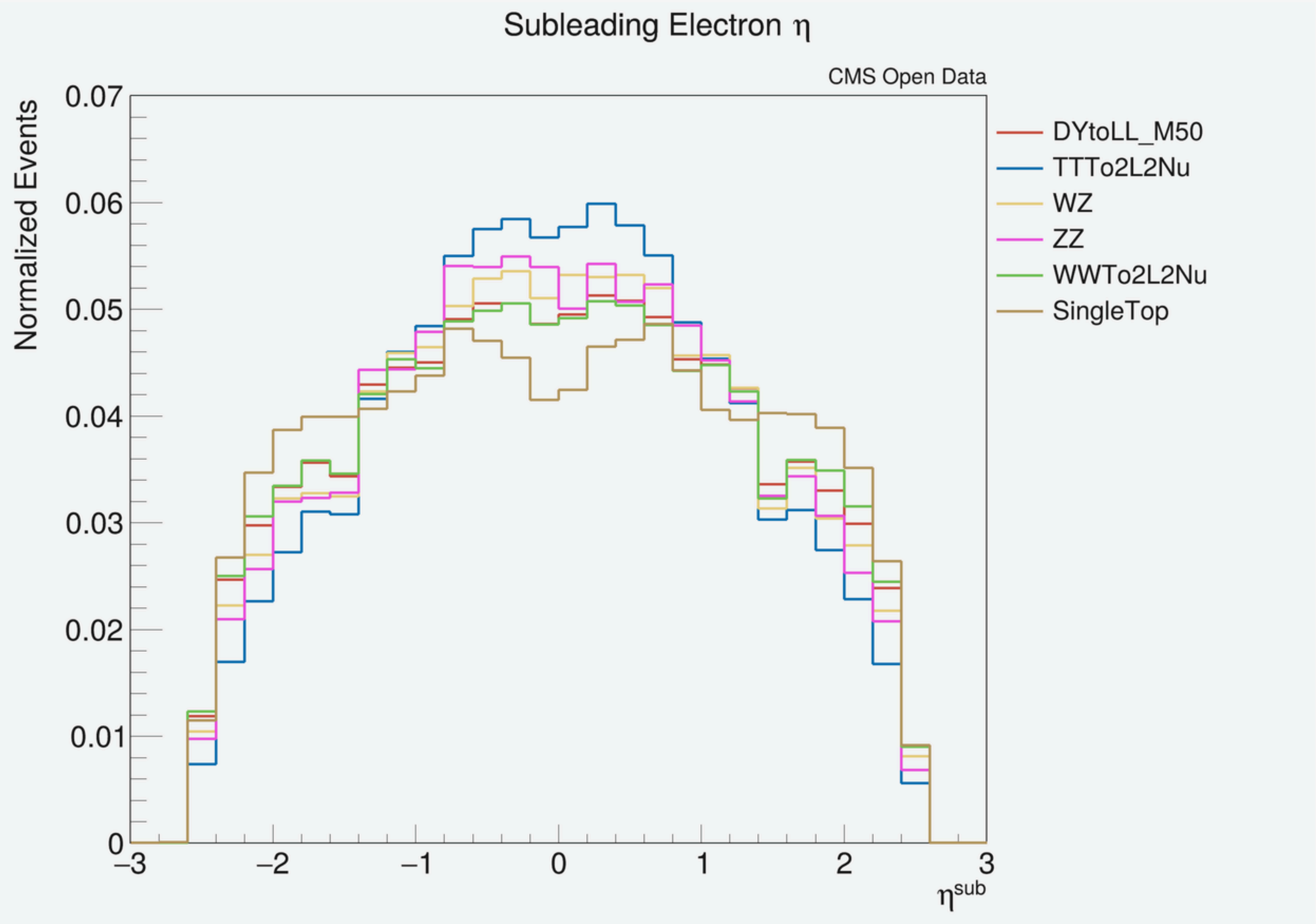
Kinematic Variables-2/4

Some relevant kinematic variables useful for our analysis are :

- **Pseudorapidity**

$$\eta = -\ln\left(tan\frac{\theta}{2}\right)$$



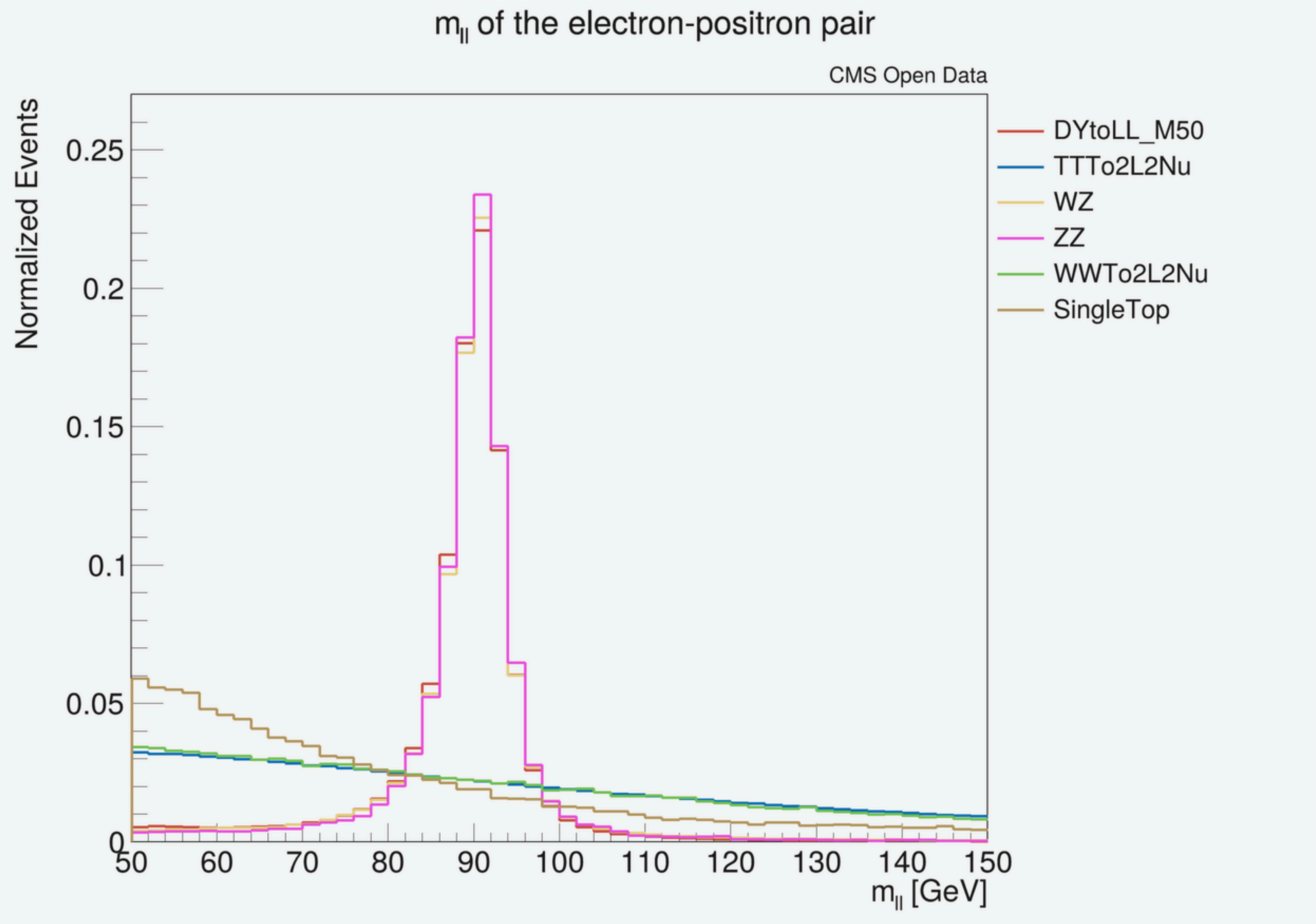


Kinematic Variables-3/4

Some relevant kinematic variables useful for our analysis are :

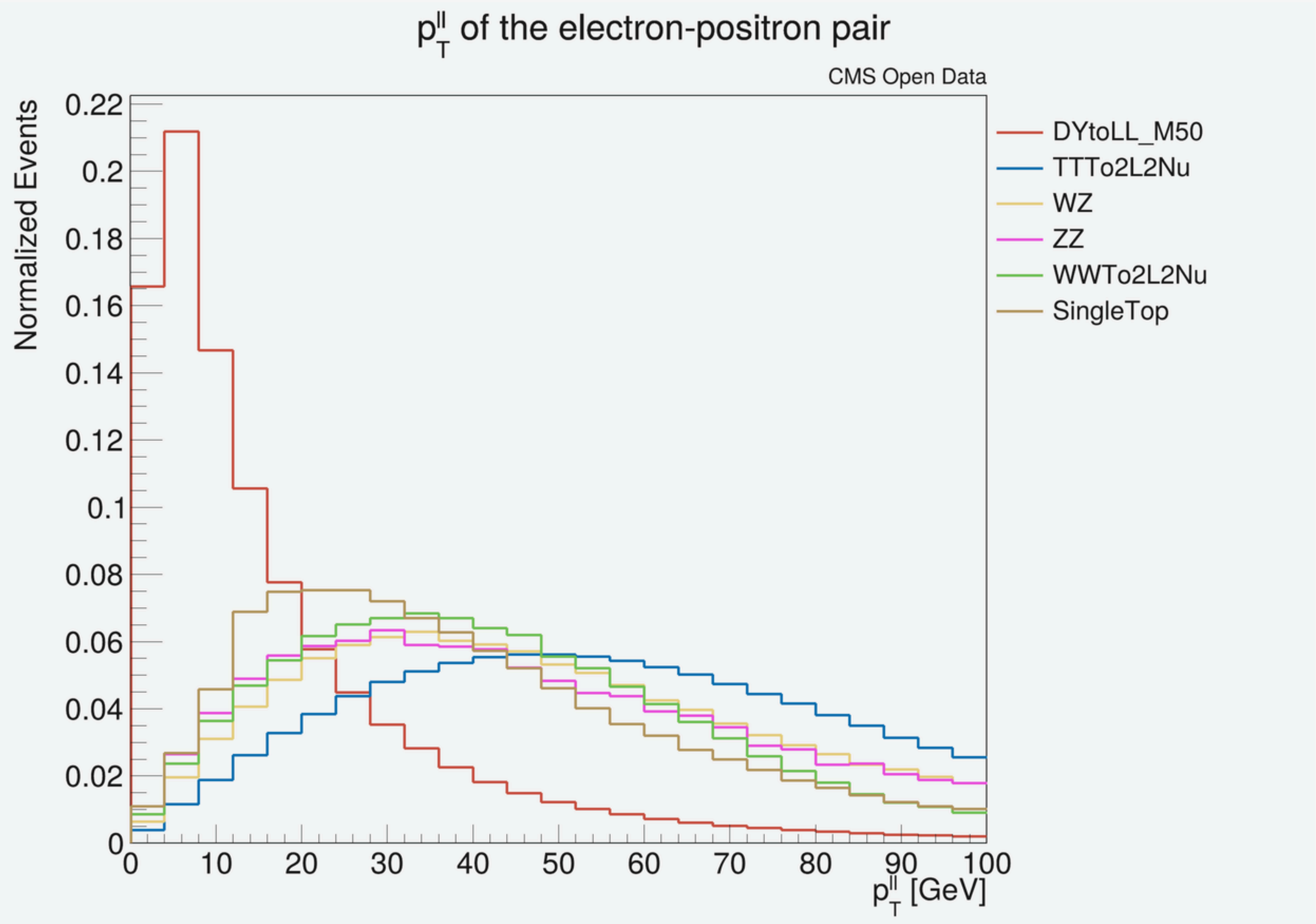
- **Invariant mass of the di-electron**

$$m_{\ell\ell} = \sqrt{(E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2}$$



- **Transverse momentum of di-electron**

$$p_T^{\ell\ell} = \sqrt{p_{T1}^2 + p_{T2}^2 + 2p_{T1}p_{T2}\cos(\Delta\phi)}$$

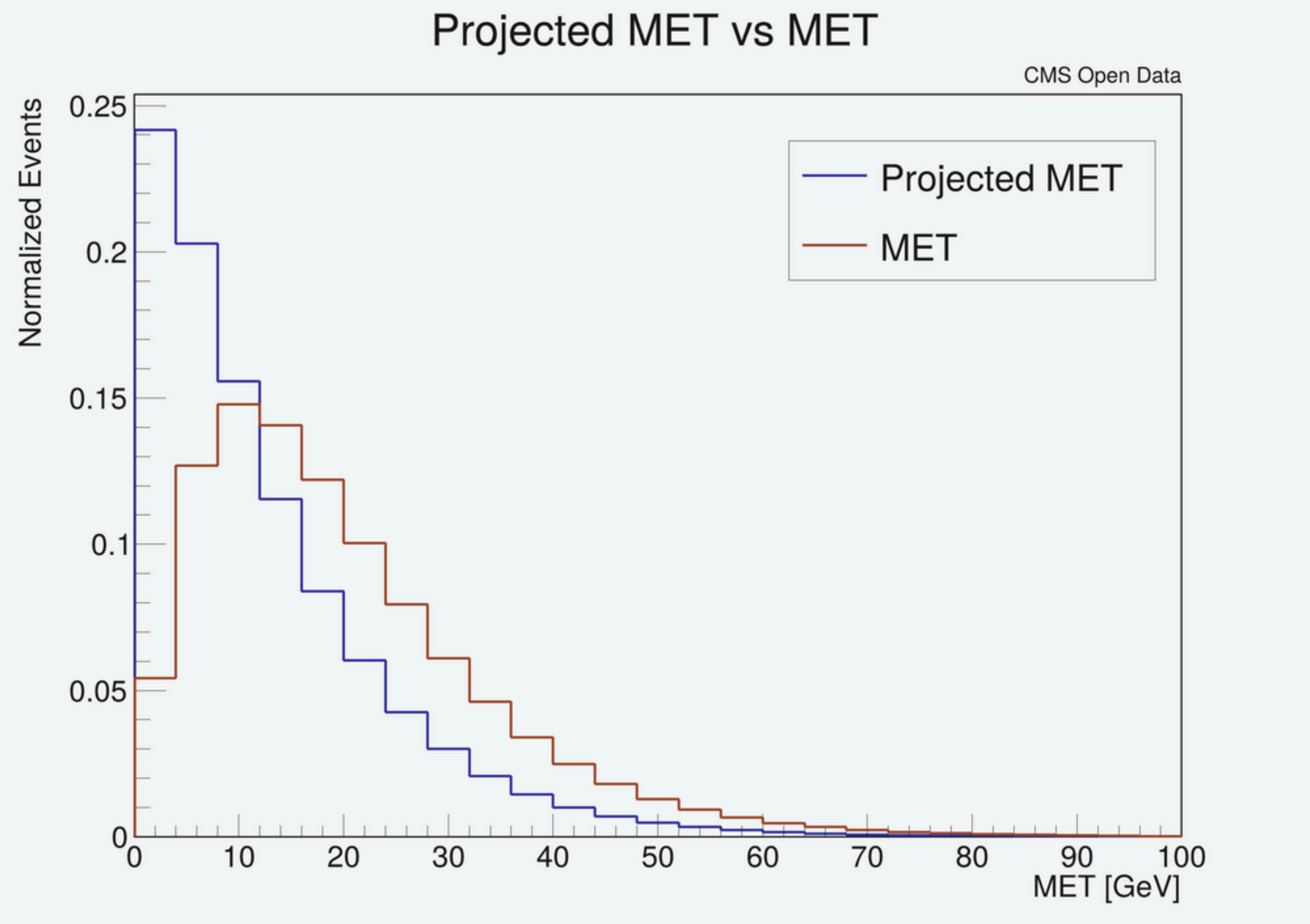


MET and projected MET comparision

- MET is the negative of the vector sum of the pT of all final-state particles reconstructed in the detector.
- MET measures the imbalance in transverse momentum, often due to undetected particles like neutrinos or mismeasurement.
- Projected MET modifies MET by considering its direction relative to the leptons.
- This projection reduces the impact of MET aligned with leptons, which is often a sign of fake MET from detector effects.

$$\text{projected\_MET} = \begin{cases} p_T^{\text{miss}} \cdot \sin(\Delta\phi_{\min}) & \text{if } \Delta\phi_{\min} < \pi/2 \\ p_T^{\text{miss}} & \text{otherwise} \end{cases}$$

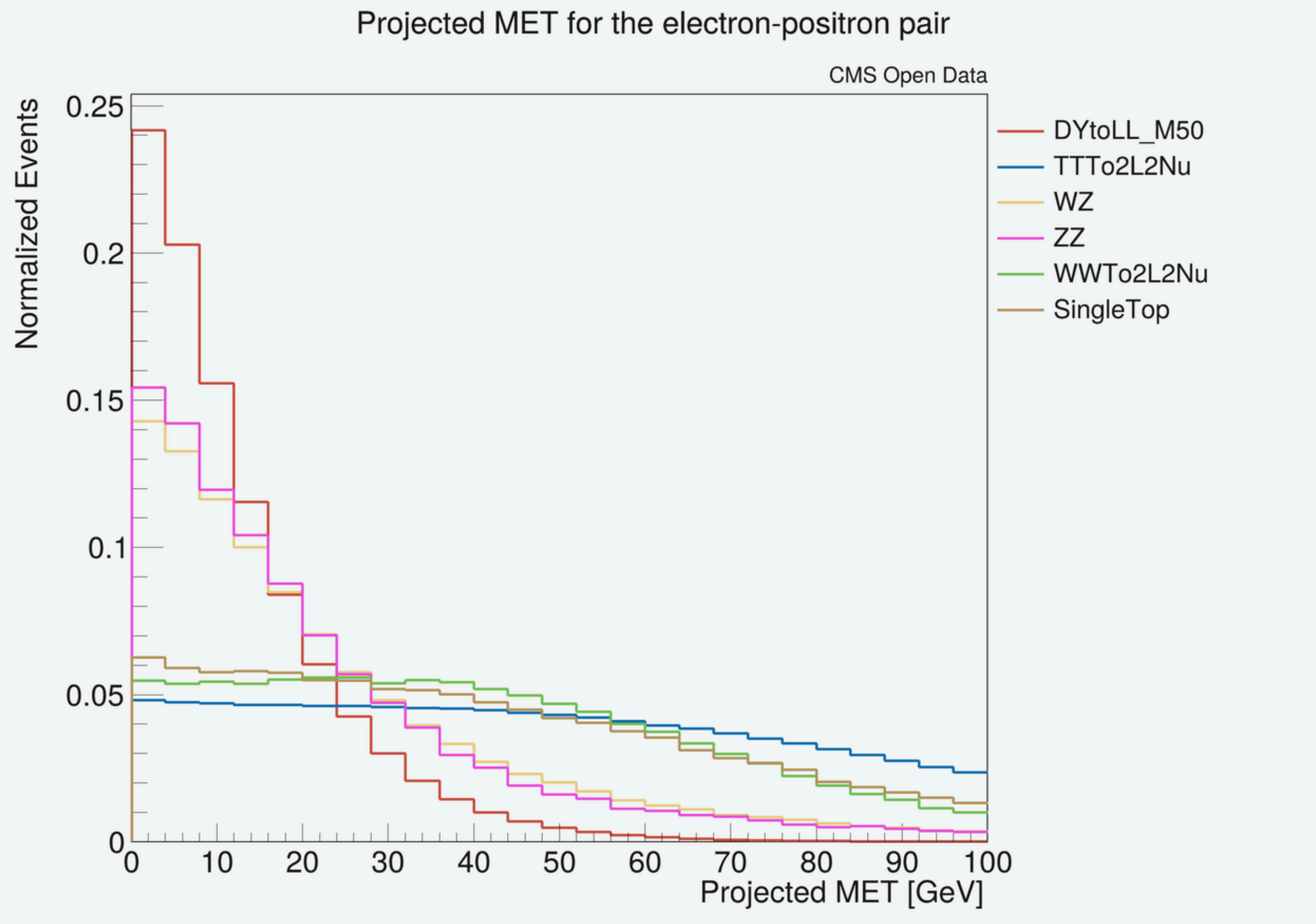
$$\Delta\phi_{\min} = \min(\Delta\phi(\ell_1, \vec{p}_T^{\text{miss}}), \Delta\phi(\ell_2, \vec{p}_T^{\text{miss}})).$$



Kinematic Variables-4/4

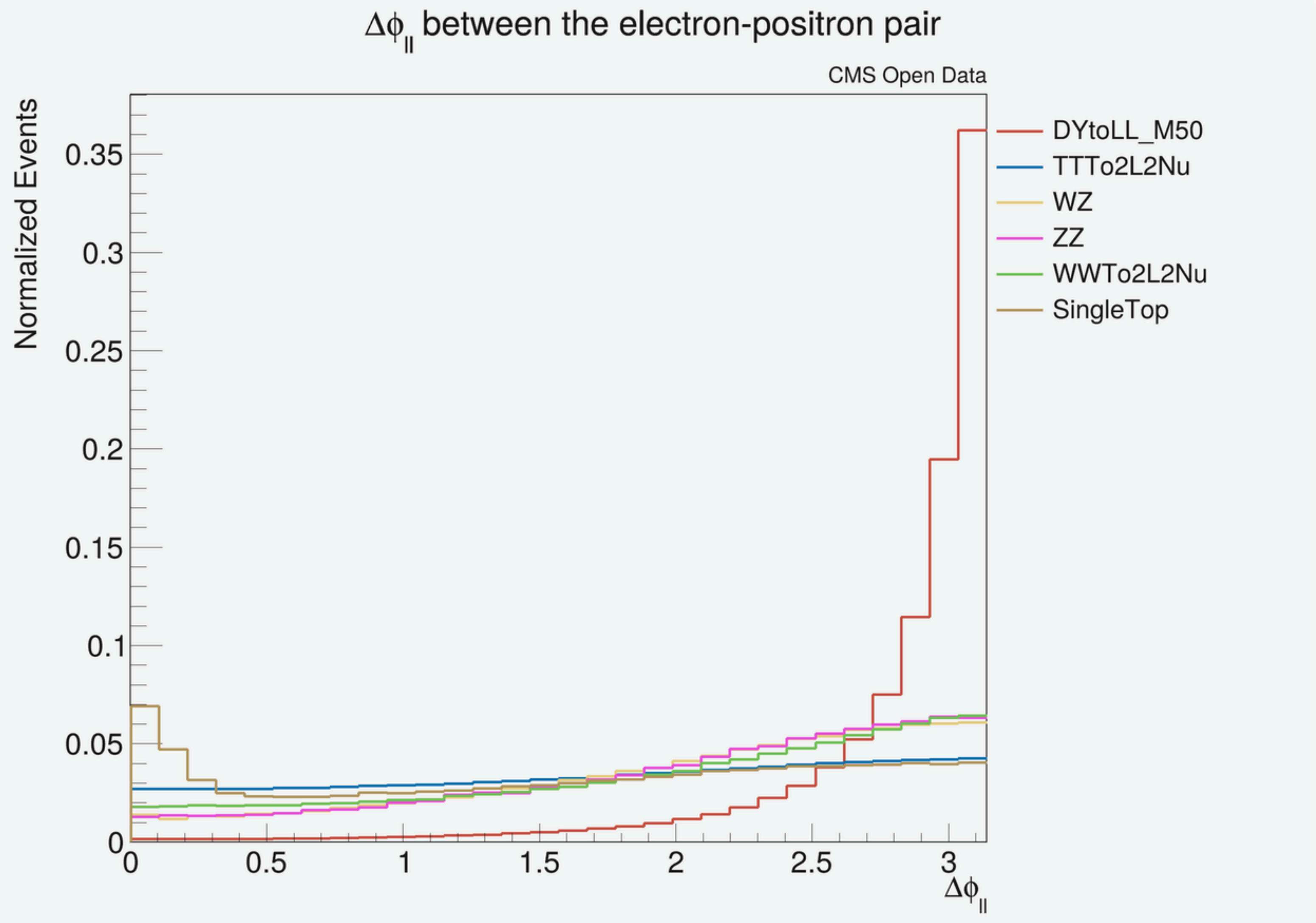
Some relevant kinematic variables useful for our analysis are :

- **Projected Missing Transverse Energy**



- **Azimuthal angle difference**

$$\Delta\phi_{\ell\ell} = |\phi_1 - \phi_2|$$



Selection Cuts & Event Selection

Variable	Selection Cuts
Number of electrons	nElectron == 2
Opposite-sign requirement	$\ell_1 \times \ell_2 = -1$

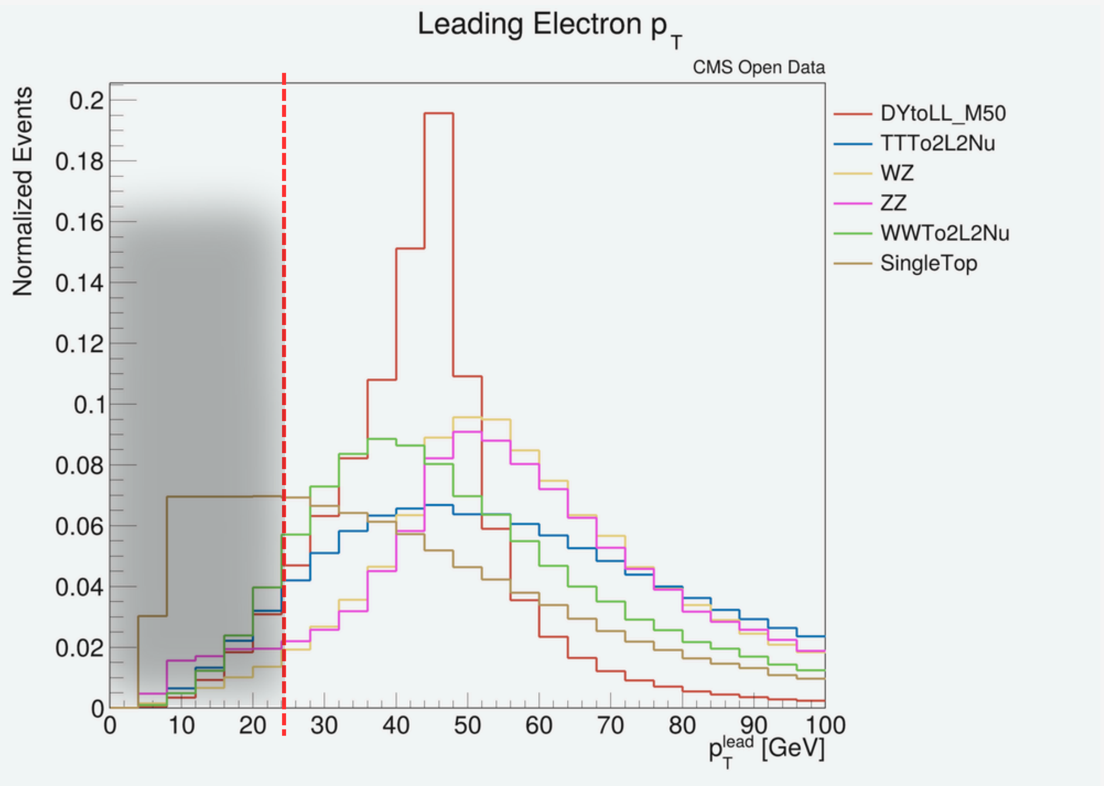
Pre-Selection Criteria

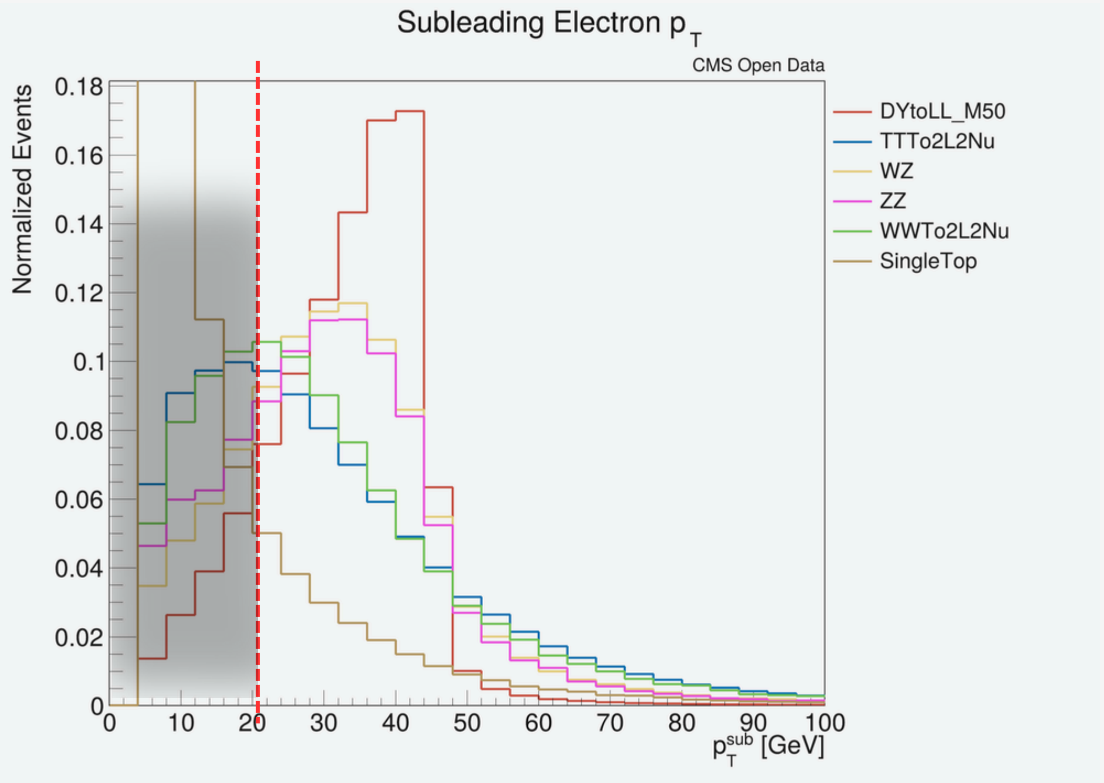
Transverse momentum requirements	Leading > 25 GeV, Subleading > 20 GeV
Pseudorapidity requirement	$\lvert\eta\rvert < 2.5$
Invariant mass requirements	60 GeV < $m_{\ell\ell}$ < 120 GeV
Transverse momentum of di-electrons requirements	$p_T\ell\ell < 40$ GeV
MET requirement	MET < 25 GeV
Azimuthal Seperation requirements	$\Delta\phi_{\ell\ell} > 2.5$

Final selection Criteria

Selection Cuts & Event Selection

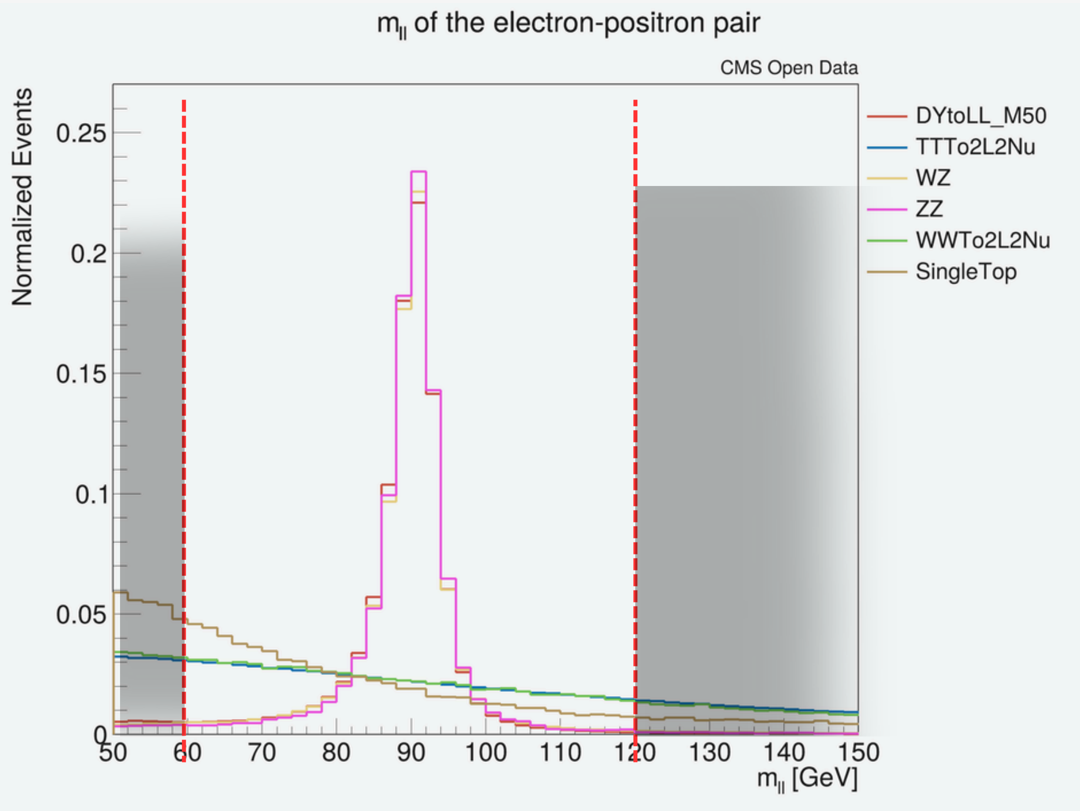
Transverse momentum requirements	Leading > 25 GeV, Subleading > 20 GeV





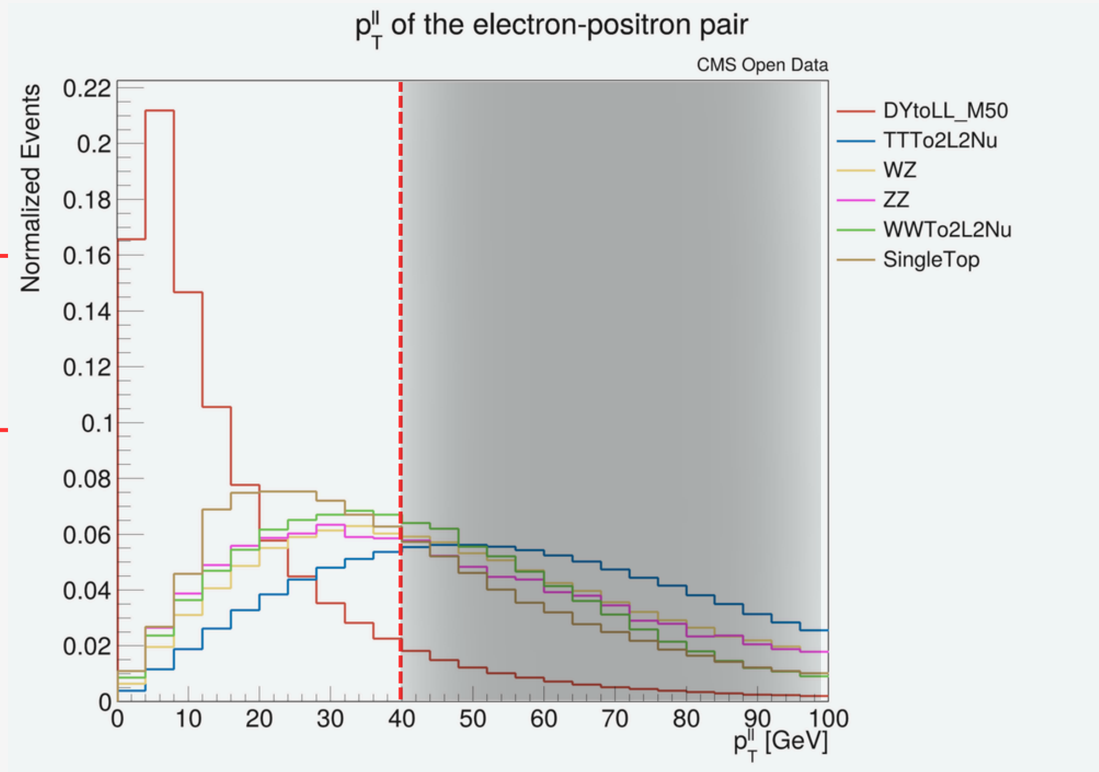
Selection Cuts & Event Selection

Invariant mass requirements	60 GeV < mℓℓ < 120 GeV



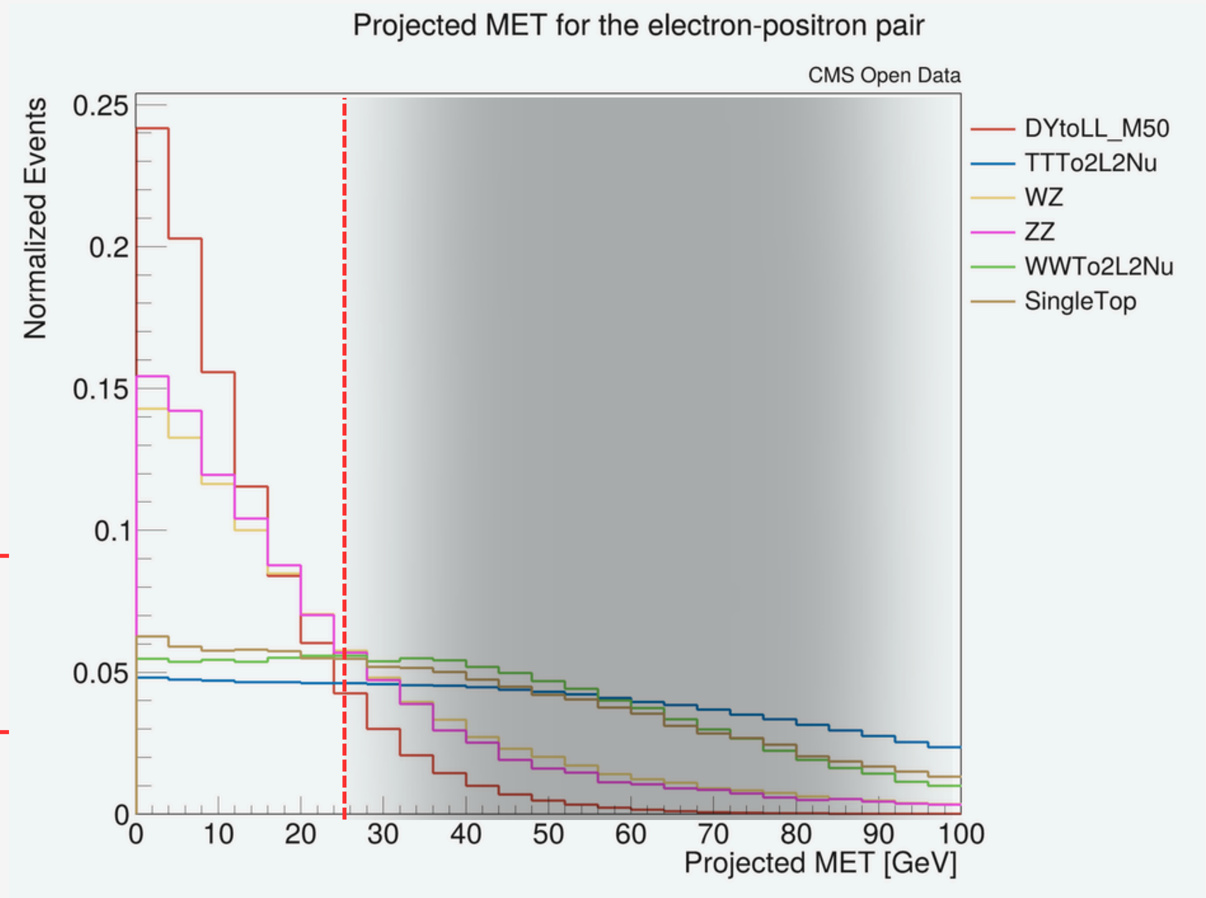
Selection Cuts & Event Selection

Transverse momentum of di-electrons requirements	$p_T\ell\ell \ < \ 40\ \mathrm{GeV}$



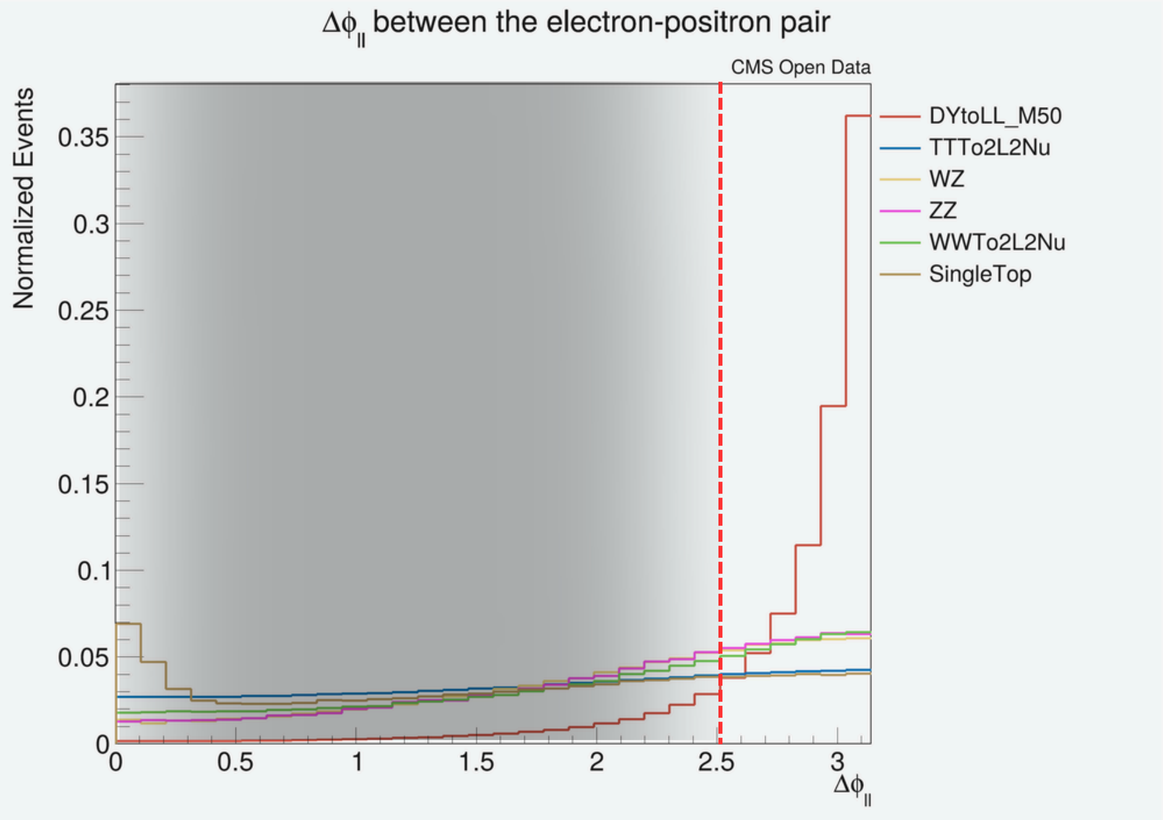
Selection Cuts & Event Selection

MET requirement	MET < 25 GeV



Selection Cuts & Event Selection

Azimuthal Seperation requirements	$\Delta\phi_{\ell\ell} > 2.5$



Analysis Strategy

- The analysis involves all the cuts applied in an sequential manner, a given cut is imposed on all the previous cuts.
- To keep the analysis reasonable, all the number are normalized by an normalization factor developed for a luminosity of 1 pb^{-1} .

$$\omega_{\text{norm}} = \frac{\sigma \cdot \mathcal{L}}{N_{\text{events}}}$$

- The next slide shows the result of all the cuts applied in a sequential manner.
- Shown are yield of the process along with the retention efficiency with respect to pre-selection yield.

Analysis Strategy

```
for (Long64_t i = 0; i < nEntries; i++) {
    tree->GetEntry(i);

    // Select tight electrons
    vector<int> tightElectrons;
    for (UInt_t j = 0; j < nElectron; j++) {
        if (Electron_mvaFall17V2Iso_WP90[j]) {
            tightElectrons.push_back(j);
        }
    }

    if (tightElectrons.size() != 2) continue;

    int i1 = tightElectrons[0];
    int i2 = tightElectrons[1];
```

```
if (Electron_pdgId[i1] * Electron_pdgId[i2] < 0) {
    eventAfterstage1++;

    int lead = (Electron_pt[i1] > Electron_pt[i2]) ? i1 : i2;
    int sublead = (lead == i1) ? i2 : i1;

    if ((Electron_pt[lead] > 25) && (Electron_pt[sublead] > 20) &&
        (fabs(Electron_eta[lead]) < 2.5) && (fabs(Electron_eta[sublead]) < 2.5)) {

        eventAfterstage2++;

        TLorentzVector el1, el2;
        el1.SetPtEtaPhiM(Electron_pt[lead], Electron_eta[lead], Electron_phi[lead], 0.000511);
        el2.SetPtEtaPhiM(Electron_pt[sublead], Electron_eta[sublead], Electron_phi[sublead], 0.000511);
        TLorentzVector dilepton = el1 + el2;

        double mll = dilepton.M();
        double ptll = dilepton.Pt();
        double dphill = fabs(el1.DeltaPhi(el2));
```

```
if (mll > 60 && mll < 120) {
    eventAfterstage3++;

    if (projected_MET < 25) {
        eventAfterstage4++;

        if (ptll < 40) {
            eventAfterstage5++;

            if (dphill > 2.5) {
                eventAfterstage6++;
            }
```

- Code snippets for getting yields of process with sequential cuts

Yield Table

The quantity in the brackets represents the retention efficiency with respect to the pre-selection yield

Process	Pre-Selections	pT, η cuts	mℓℓ cuts	MET cuts	pTℓℓ cuts	Azimuthal separation cuts
Drell–Yan	702,364 ± 231	591,809 ± 212 (84.25%)	576,077 ± 209 (82.02%)	503,826 ± 195 (71.73%)	453,184 ± 185 (64.52%)	431,586 ± 181 (61.45%)
Top quarks	6,651 ± 4	4,289 ± 3 (64.48%)	1,780 ± 2 (26.76%)	458 ± 1 (6.89%)	139 ± 1 (2.09%)	97 ± 1 (1.46%)
Single top	225 ± 1	52 ± 1 (23.11%)	24 ± 1 (10.67%)	7 ± 1 (3.11%)	3 ± 1 (1.33%)	2 ± 1 (0.89%)
WZ	702 ± 2	546 ± 2 (77.78%)	518 ± 2 (73.79%)	344 ± 1 (49.00%)	139 ± 1 (19.80%)	109 ± 1 (15.52%)
ZZ	511 +3	382 ± 2 (74.75%)	361 ± 2 (70.64%)	241 ± 2 (47.16%)	104 ± 1 (20.35%)	67 ± 1 (13.11%)
WW	707 ± 2	465 ± 1 (65.77%)	208 ± 1 (32.25%)	73 ± 1 (10.32%)	54 ± 1 (7.64%)	45 ± 1 (6.36%)

Result

- The analysis analyze the kinematics of DY process as the signal and many background processes like top quark production, single top quark production, and di-boson productions.
- The final signal and background retention yield along with their respective efficiencies are:

Process	Pre-selection Yield	After Analysis Yield	Retention efficiency [%]	Rejection efficiency [%]
Signal	702,364 ± 231	431,586 ± 181	61.45	38.55
Backgrounds	8,796 ± 5	321 ± 1	3.64	96.35

- The scripts used in this analysis can be found on the provided github repository: [anujraghav252/msc-diss]

Thank you for your time!

Back-up

Optimization of projected MET

Threshold	30 GeV	25 GeV	20 GeV
Signal retention (%)	75.00	71.00	65.80
Background retention (%)	14.00	12.68	10.66

- Optimized for three different thresholds.
- Selected the most optimal threshold.

Optimization of $p_T^{\ell\ell}$

Threshold	**30 GeV**	**40 GeV**	**50 GeV**
Signal retention (%)	63.00	67.70	70.34
Background retention (%)	2.27	3.44	4.60

- Optimized for three different thresholds.
- Selected the most optimal threshold.

The Compact Muon solenoid

- The ECAL is cruicial for studying electrons and photons.
- Use $PbWO_4$ crystal due to high density (8.28 g/cm^3), short radiation length (0.89 cm), and small Moliere radius (2.2 cm)