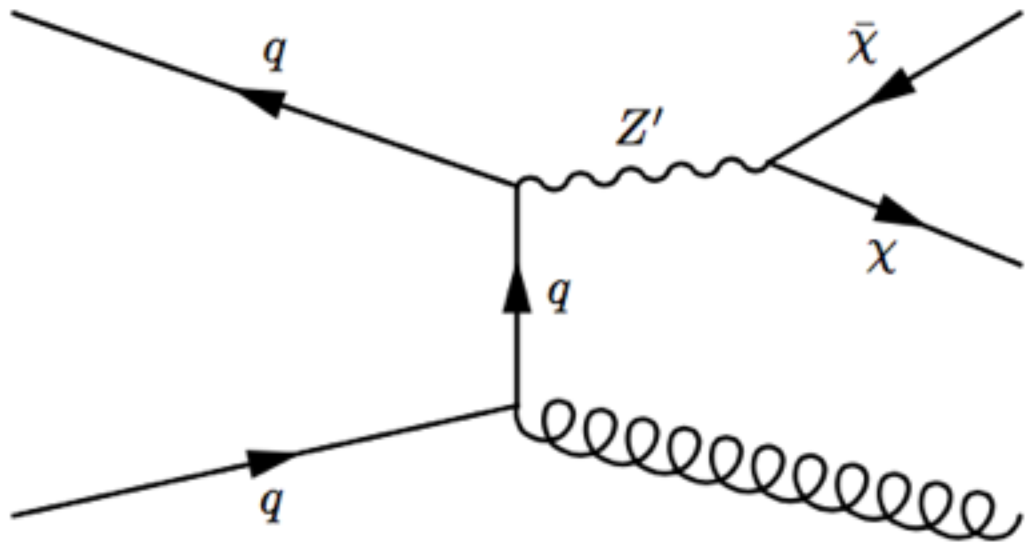
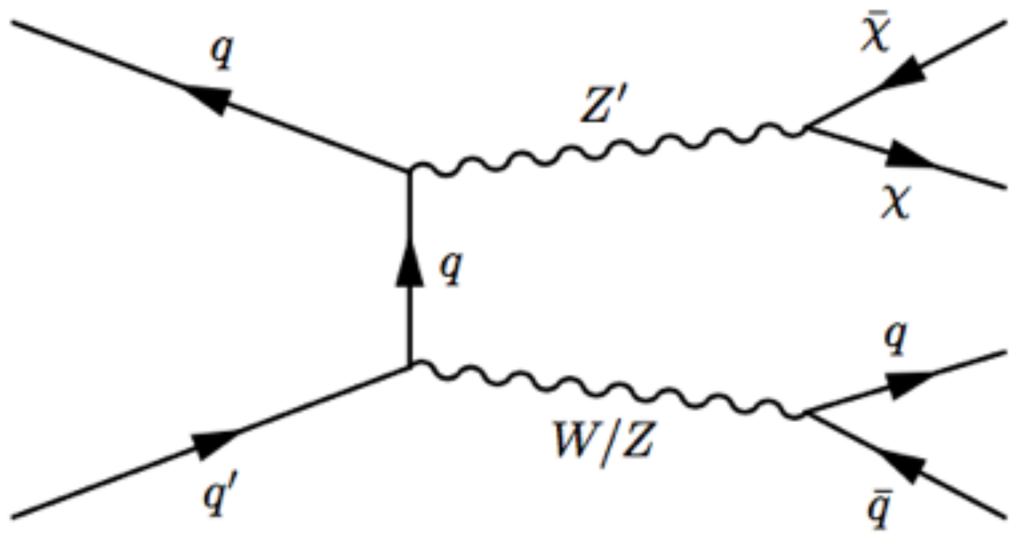
EWK Background Estimation and Uncertainty Prescription used in the Monojet analysis in CMS

Zeynep Demiragli
on behalf of CMS (Monojet group)

Introduction

Search for the pair production of DM in association with a jet from initial-state radiation, which is used to tag and/or trigger the event. Focusing on simplified models with Vector / Axial-vector / Scalar / Pseudo-scalar mediators



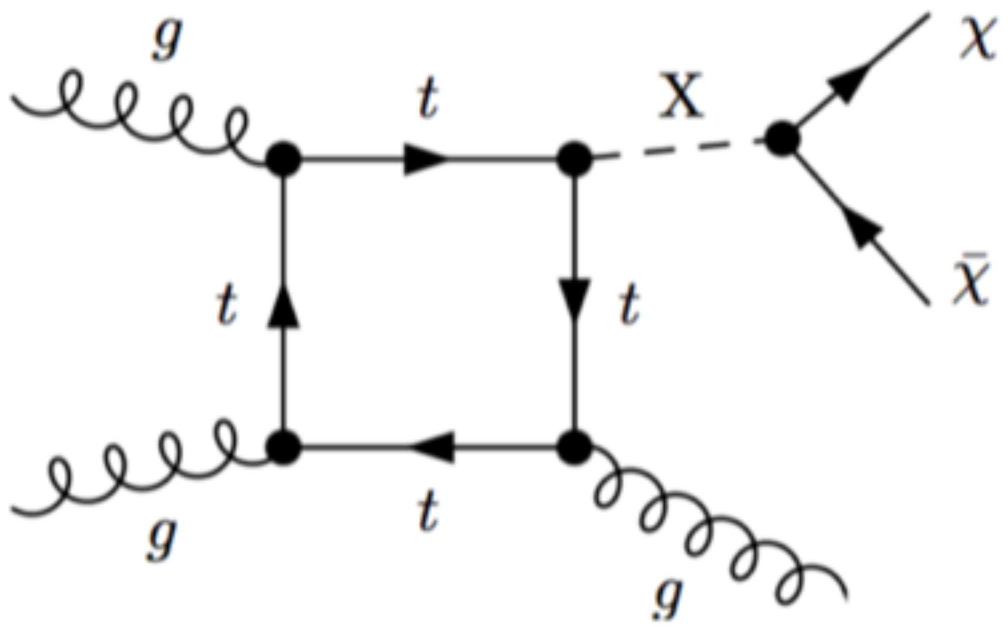


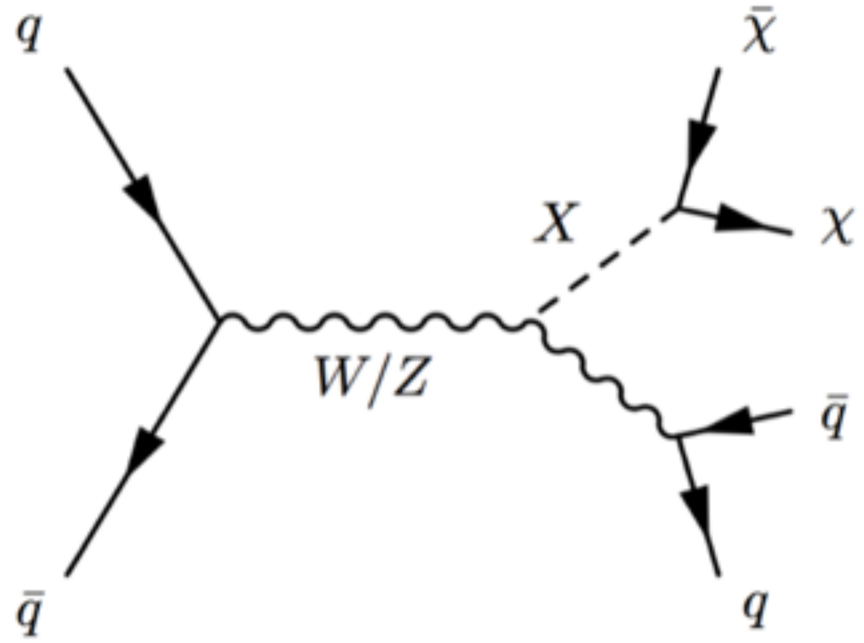
Vector & Axial-vector Mediator

σ(mono-jet) ≥ 100 x σ (mono-W)

Scalar & Pseudo-scalar Mediator

σ(mono-jet) ~ 10 x σ (mono-W)



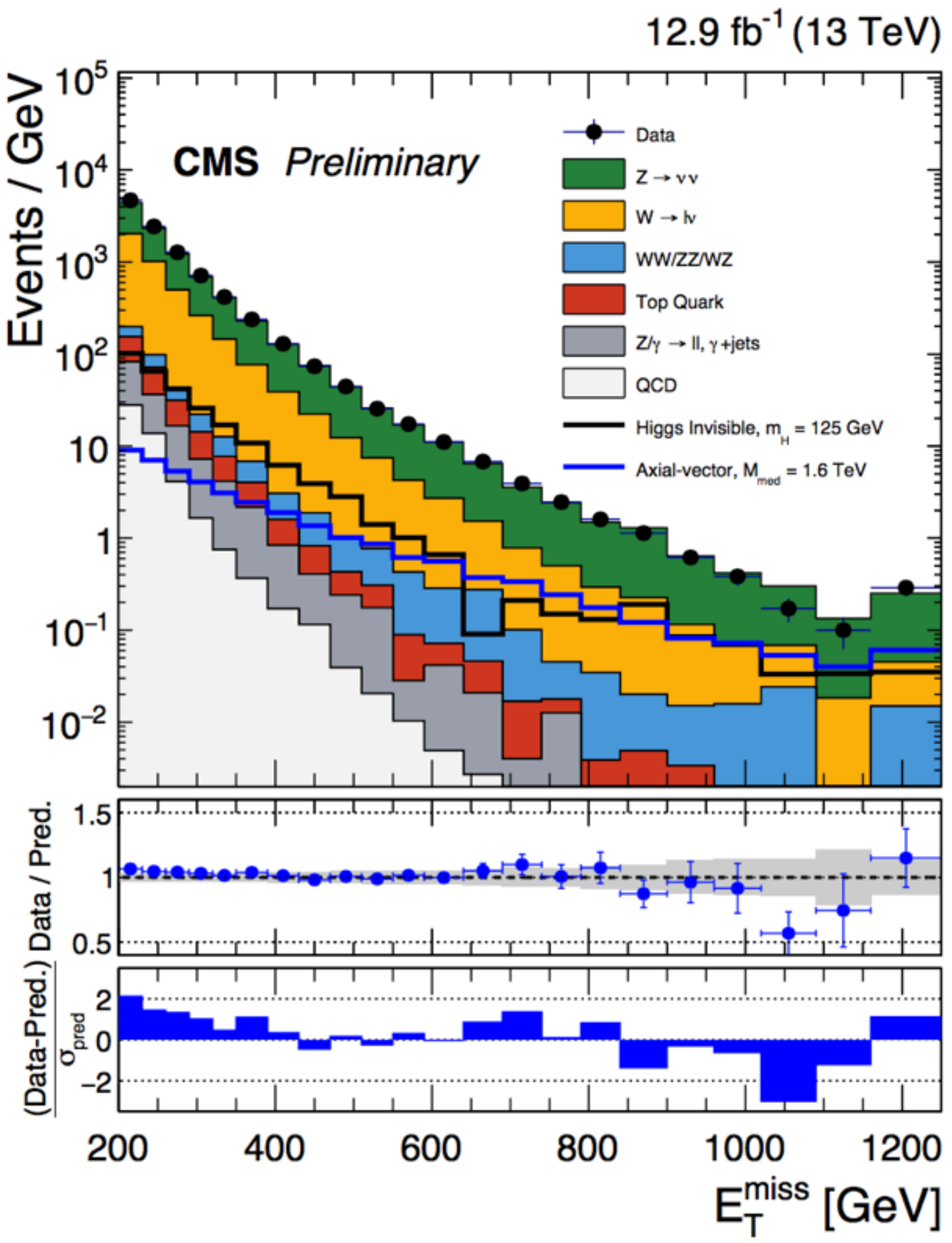


Signal extraction is based on **MET** distribution, fitting 1 parameter in each bin

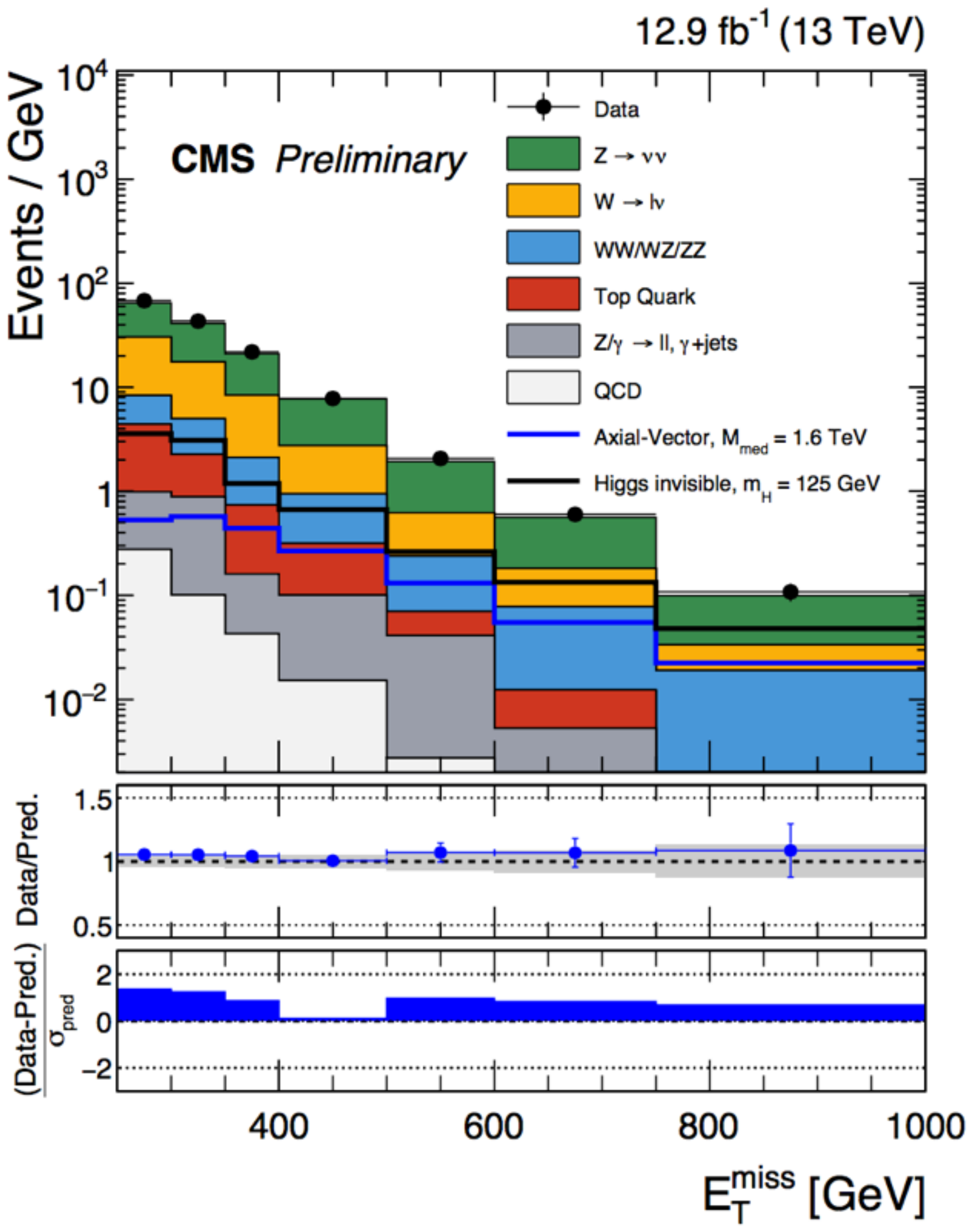
5 Control Regions (Zee, Wen, Zmm, Wmn, GJets) to estimate major backgrounds. **Fit performed simultaneously** in different categories.

ICHEP Monojet Results

Monojet Category



MonoV Category



Z(νν)+jets & W(lν)+jets
> 90% of the total background

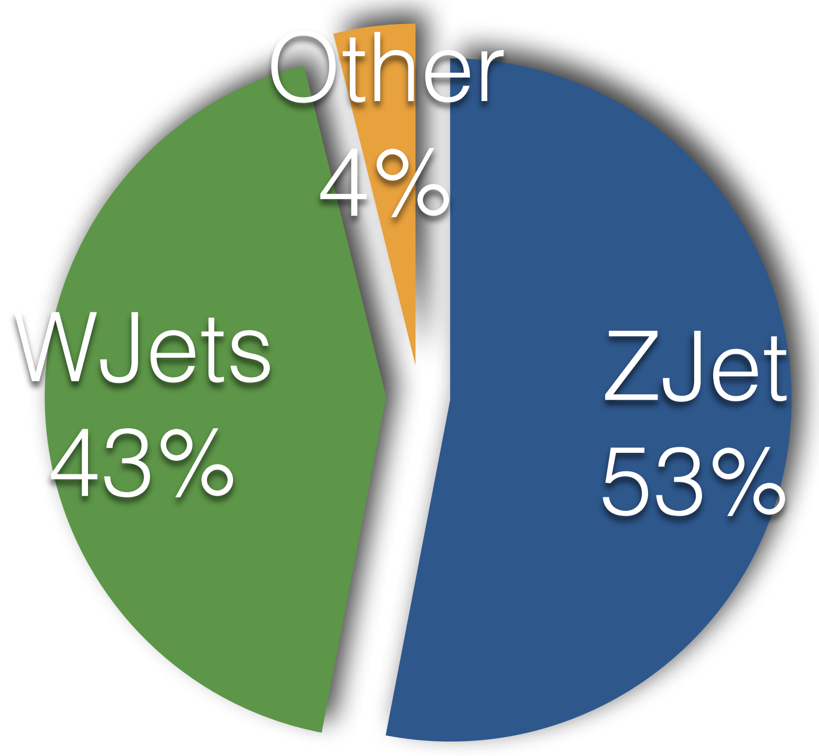
Remaining Backgrounds: DiBoson, Top (ttbar +single top), Z(ll)+j, QCD

Current Analysis Mediator Mass Reach (observed):

- **V & AV Reach: ~ 1.9 TeV**
- **Scalar Reach: ~ 100 GeV**
- **PS Reach: ~ 400 GeV**

Zeynep Demiragli

Background Composition Overview

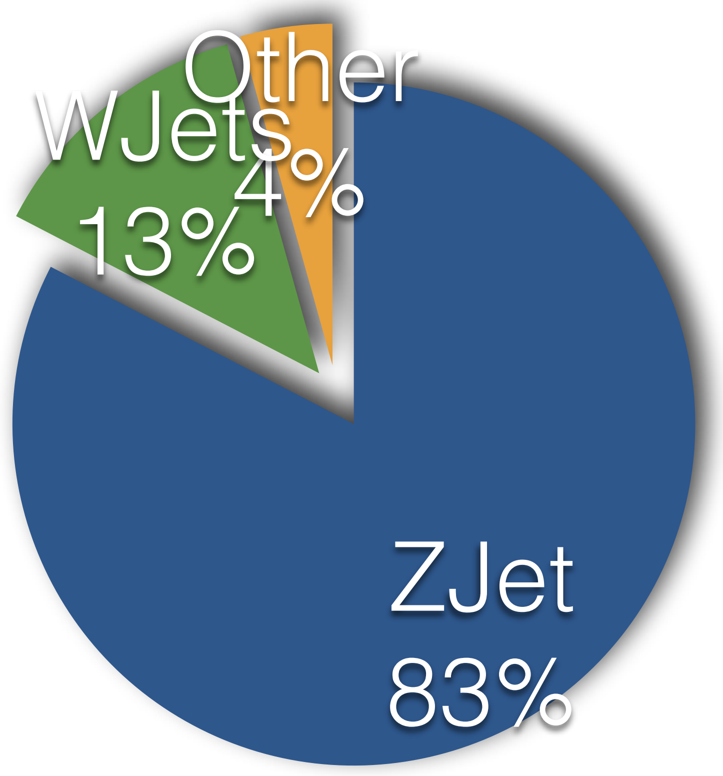


Low MET

Dominant Backgrounds (Data Driven Estimation)

Z(vv)+jets: Irreducible background. Generated at LO in QCD and are corrected to NLO in QCD and EWK.

W(lv)+jets: Becomes irreducible in the case where the leptons are lost out of the detector acceptance. Generated at LO in QCD and are corrected to NLO in QCD and EWK.



High MET

Sub dominant Backgrounds (MC Based)

Z(ll)+jets: Passes the selection if both leptons are lost. Generated at LO in QCD in bins of HT and are corrected to NLO in QCD and EWK.

G+jets: Generated at LO in QCD in bins of HT and are corrected to NLO in QCD and EWK.

Top: tt and single top generated at NLO in QCD using aMC@NLO

Dibosons: Decays of diboson pairs (WW, WZ, ZZ) generated using pythia8

QCD: LO in QCD using the MadGraph (data driven templates used for signal region)

Zeynep Demiragli

Background Generation Settings

Z(vv)+jets: Main irreducible background.

W(lv)+jets: Becomes irreducible in the case where the lepton is lost or out of the detector acceptance.

Z(ll)+jets: Passes the selection if both leptons are lost. Also used in the control regions.

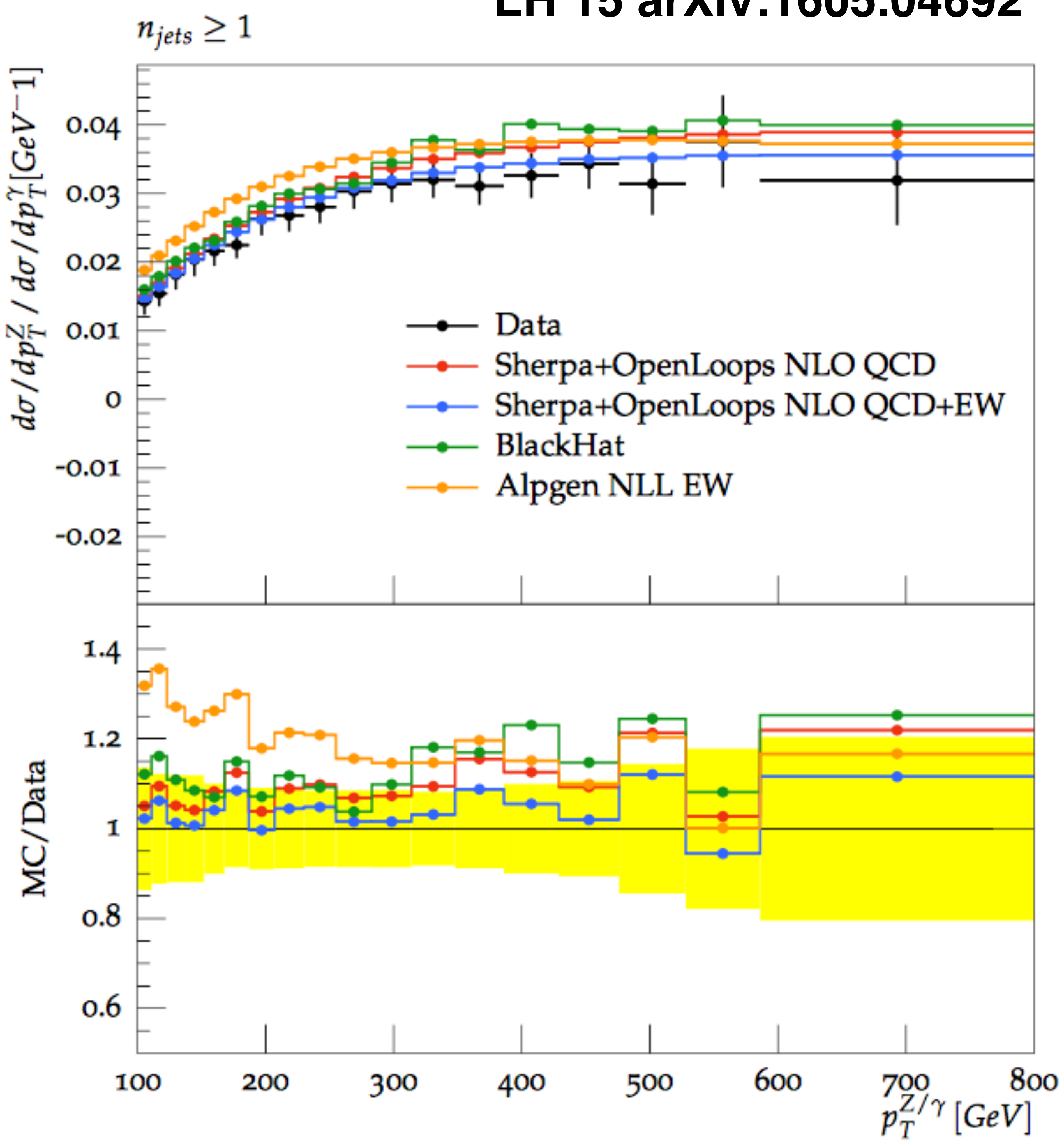
G+jets: Used in the control regions.

Sample	Generator	Order in QCD/ EWK	Parton level jet multiplicity	Ren & Fac Scale	qCut	Parton Matching	PDF
Z + Jets (HT Bin)	Madgraph	LO / LO	4	Dynamic m_T^2	19 GeV	MLM	NNPDF 3.0
W + Jets (HT Bin)	Madgraph	LO / LO	4	Dynamic m_T^2	19 GeV	MLM	NNPDF 3.0
G + Jets (HT Bin)	Madgraph	LO / LO	4	Dynamic m_T^2	19 GeV	MLM	NNPDF 3.0

All generated at LO in QCD in bins of HT showered by Pythia8 and are corrected to NLO in QCD and EWK.

Why is it important to correct for NLO in QCD and EWK?

LH'15 arXiv:1605.04692



CMS Z/gamma ratio (8 TeV) measurement compared to different generators

It is all about how well the simulation describes the data

NLO QCD corrections for the EWK bkgs are derived from:

- pT binned aMC@NLO samples :
  - Z+jets and W+jets generated up to 2 additional partons at ME
  - γ+jets generated with no additional partons at Born level

NLO EWK corrections obtained from:

- arxiv:1511.08692 —> Sherpa Open Loops
  - The corrections are **direct taken from the paper**
  - Corrections were derived by the authors in a phase space very close to our analysis.

The correction factors range from:

QCD: ~ +40% (at 200 GeV) to +2% (at 1 TeV)

EWK: ~ -3% (at 200 GeV) to -20% (at 1 TeV)

as a function of boson transverse momentum.

Background Estimation Method: Overview

Gamma Jets

W (ev) Jets

$$\mu_i^{W\to l\nu} \to f_i(\boldsymbol{\theta}) \cdot \mu_i^{Z\to\nu\nu}$$

Z (vv) Jets

W (lv) Jets

5 Control Regions (Zee, Wen, Zmm, Wmn, GJets) to estimate major backgrounds. **Fit performed simultaneously** in different categories.

Z (mm) Jets

Z (ee) Jets

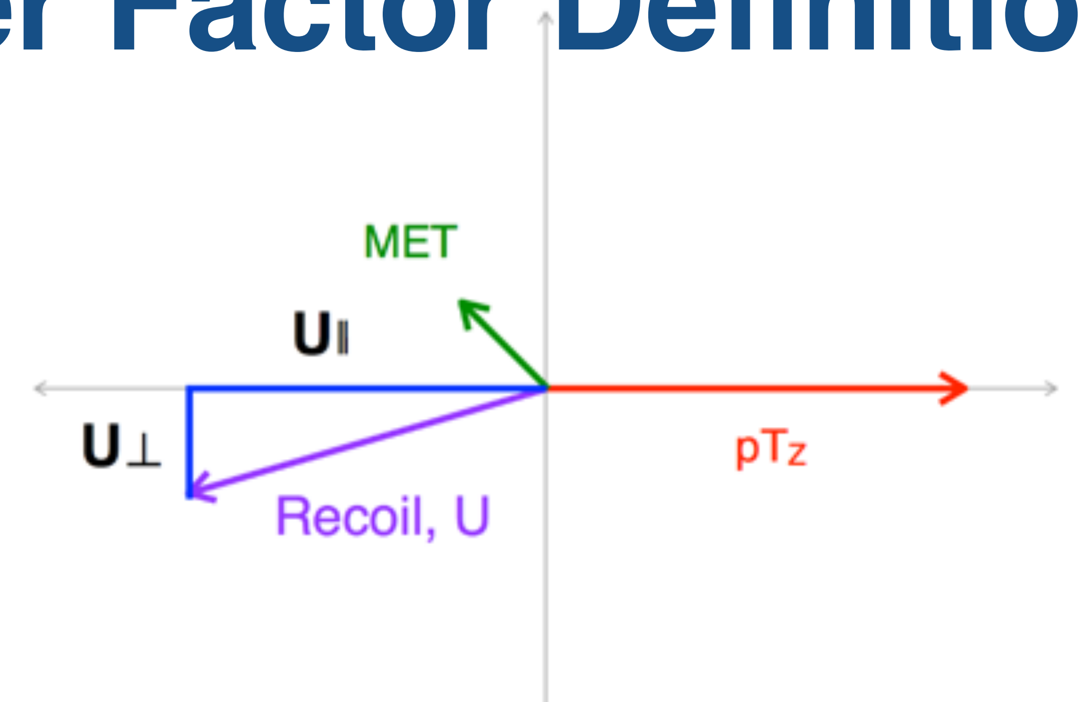
W (mv) Jets

Zeynep Demiragli

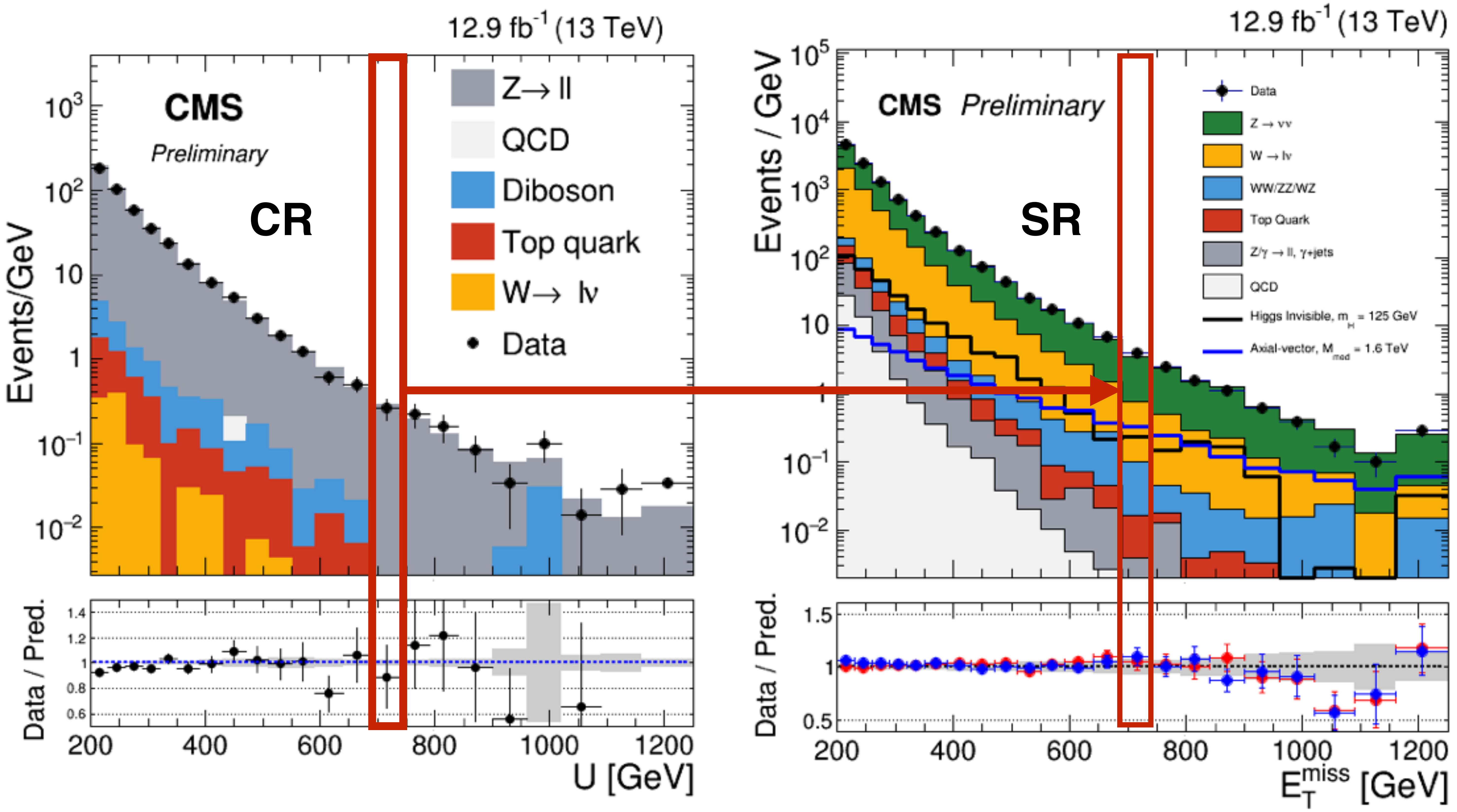
Background Estimation Method: Transfer Factor Definition

Procedure

- **Step 1:** Compute a "Recoil" Variable (U) in the Control Regions (CRs)
  - U = Met + $Pt^{\mu\mu/ee}$ or Met + $Pt^{\mu/e}$ or Met + P_T^{γ}



- **Step 2:** Compute "Transfer Factors" for each bin of recoil to translate between CRs to Signal Region (SR):
  - R_i^{γ} or R_i^{Z} or R_i^{W}



$$R_i^Z = \frac{N_{i,MC}^{Z\to\mu^+\mu^-}}{N_{i,MC}^{Z\to\nu\nu}}$$

N_i is the number of events in bin i of the recoil distribution

- **Step 3:** Embed uncertainties (θ) in the likelihood as constrained additive perturbations to the transfer factors $R^{\gamma/Z/W}$

Background Estimation Method: Likelihood

Objective: Define a partial likelihood for each event category as the product over Poisson likelihoods for each bin in recoil, in each of the control regions

$$\mathcal{L}_c(\mu, \boldsymbol{\mu}^{Z\to\nu\nu}, \boldsymbol{\theta}) = \prod_i \text{Poisson}\left(d_i^\gamma | B_i^\gamma(\boldsymbol{\theta}) + \frac{\mu_i^{Z\to\nu\nu}}{R_i^\gamma(\boldsymbol{\theta})}\right)$$
$$\times \prod_i \text{Poisson}\left(d_i^Z | B_i^Z(\boldsymbol{\theta}) + \frac{\mu_i^{Z\to\nu\nu}}{R_i^Z(\boldsymbol{\theta})}\right)$$
$$\times \prod_i \text{Poisson}\left(d_i^W | B_i^W(\boldsymbol{\theta}) + \frac{f_i(\boldsymbol{\theta})\,\mu_i^{Z\to\nu\nu}}{R_i^W(\boldsymbol{\theta})}\right)$$
$$\times \prod_i \text{Poisson}\left(d_i | B_i(\boldsymbol{\theta}) + (1 + f_i(\boldsymbol{\theta}))\mu_i^{Z\to\nu\nu} + \mu S_i(\boldsymbol{\theta})\right)$$

Number of observed events, in a particular bin in CR

Number of background events in a particular CR bin. Uncertainties incorporated via nuisances $\boldsymbol{\theta}$ (for eg. photon efficiency)

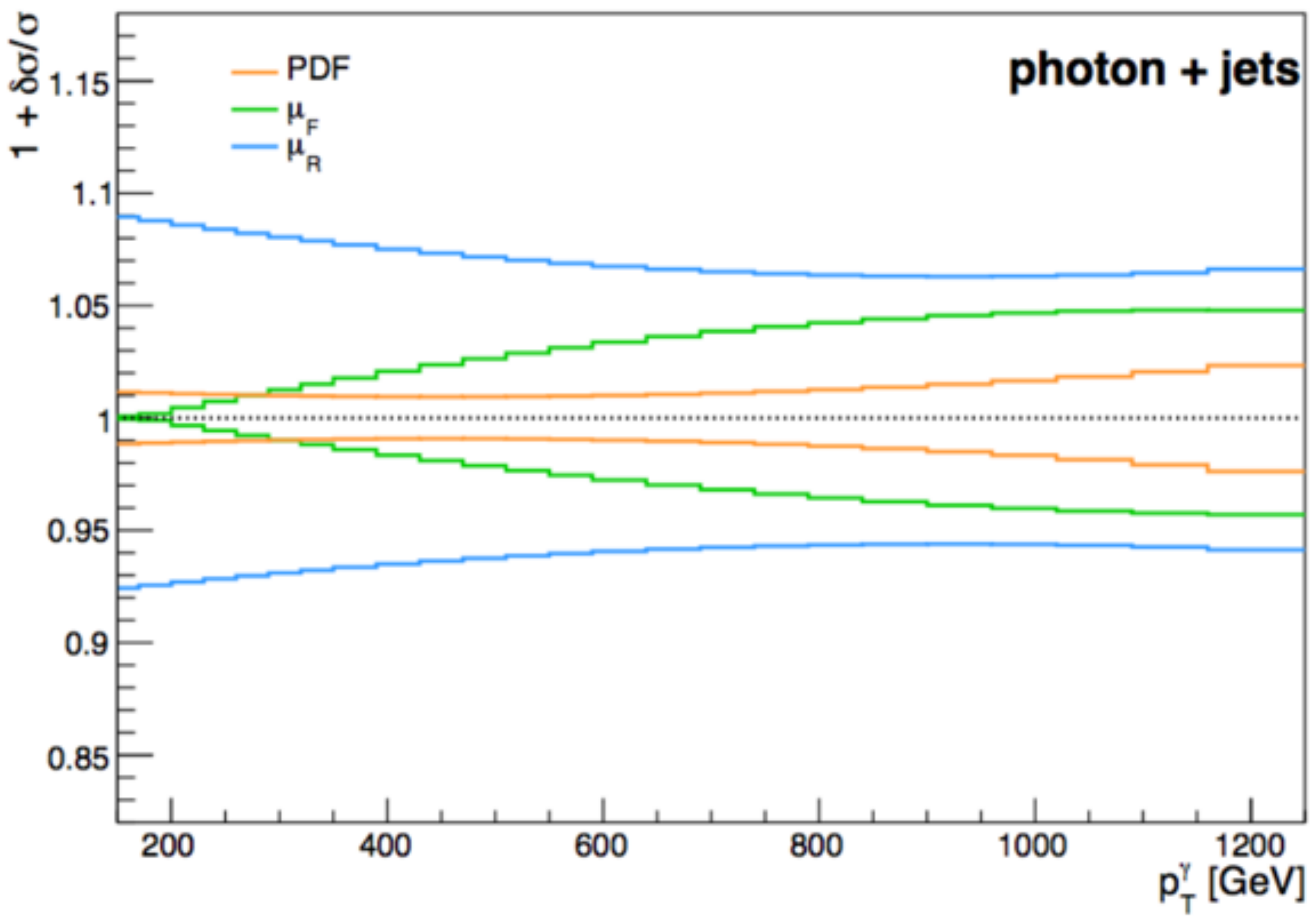
$$\mu_i^{W\to l\nu} \to f_i(\boldsymbol{\theta}) \cdot \mu_i^{Z\to\nu\nu}$$

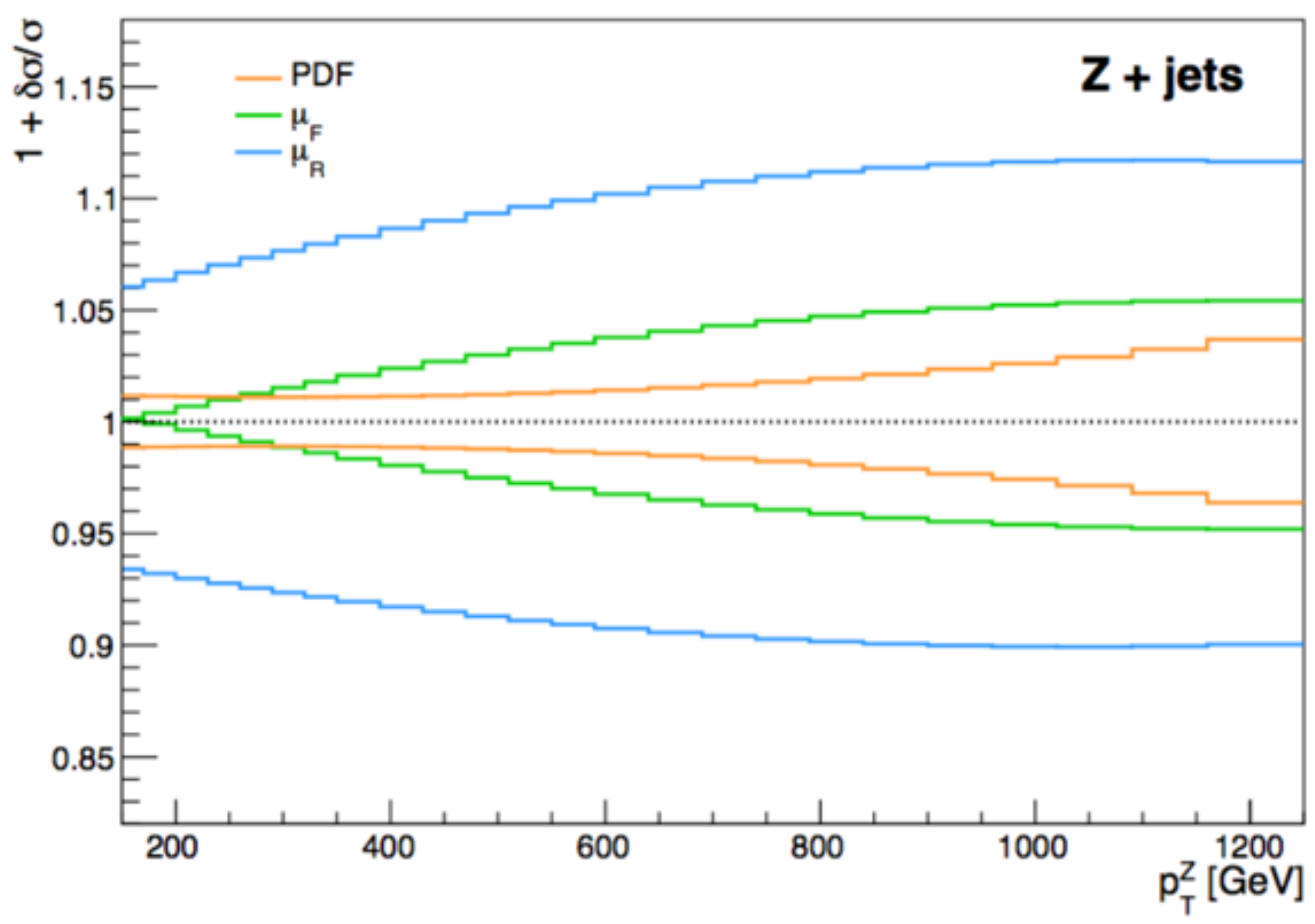
f_i ratio in the signal region :

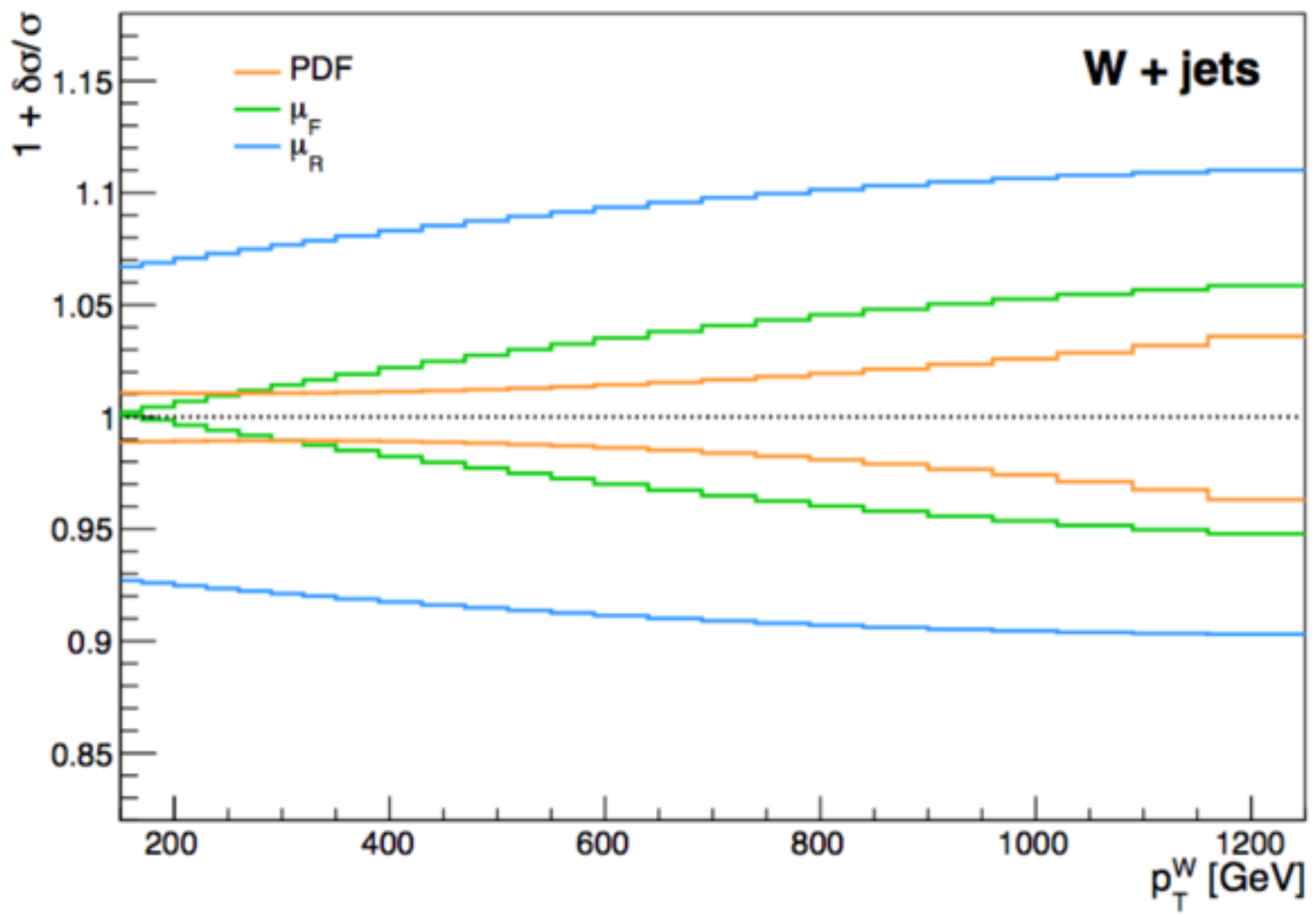
- Relies on theoretical prediction for differential xsec and lepton acceptance

Zeynep Demiragli

Theoretical Uncertainties





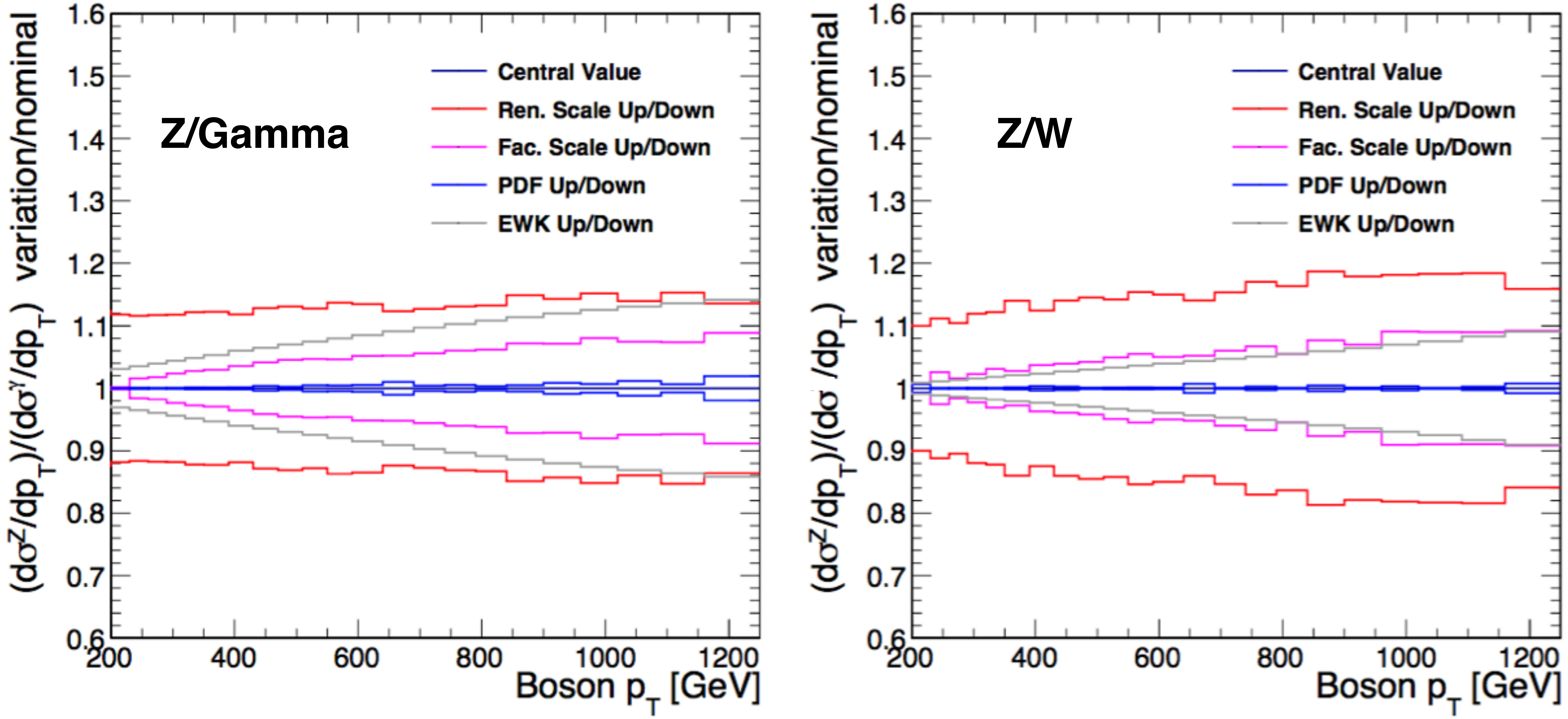


For Each Process Individually

- **QCD Scale Uncertainty:**
  - Vary renormalization and factorization scales by x2 and x1/2 for each process
- **PDF Uncertainty:**
  - NNPDF 3.0 uncertainty
- **NLO EWK Uncertainty:**
  - Use the full correction as uncertainty (Very Conservative Approach)

Full Theoretical Uncertainty Treatment on the Ratios

- **PDF Uncertainty:**
  - 100% correlated on the Z/G & Z/W ratio
    - Mostly cancels out.
    - Final uncertainty ~ 1 %
    - **Correlated across MET/Recoil Bins**
- **NLO EWK Uncertainty:**
  - 100% correlated on the Z/G & Z/W ratio
    - Leads to uncertainty of:
      - 3-15% on the Z/G, 1-10% on Z/W
    - **Uncorrelated across MET/Recoil Bins**



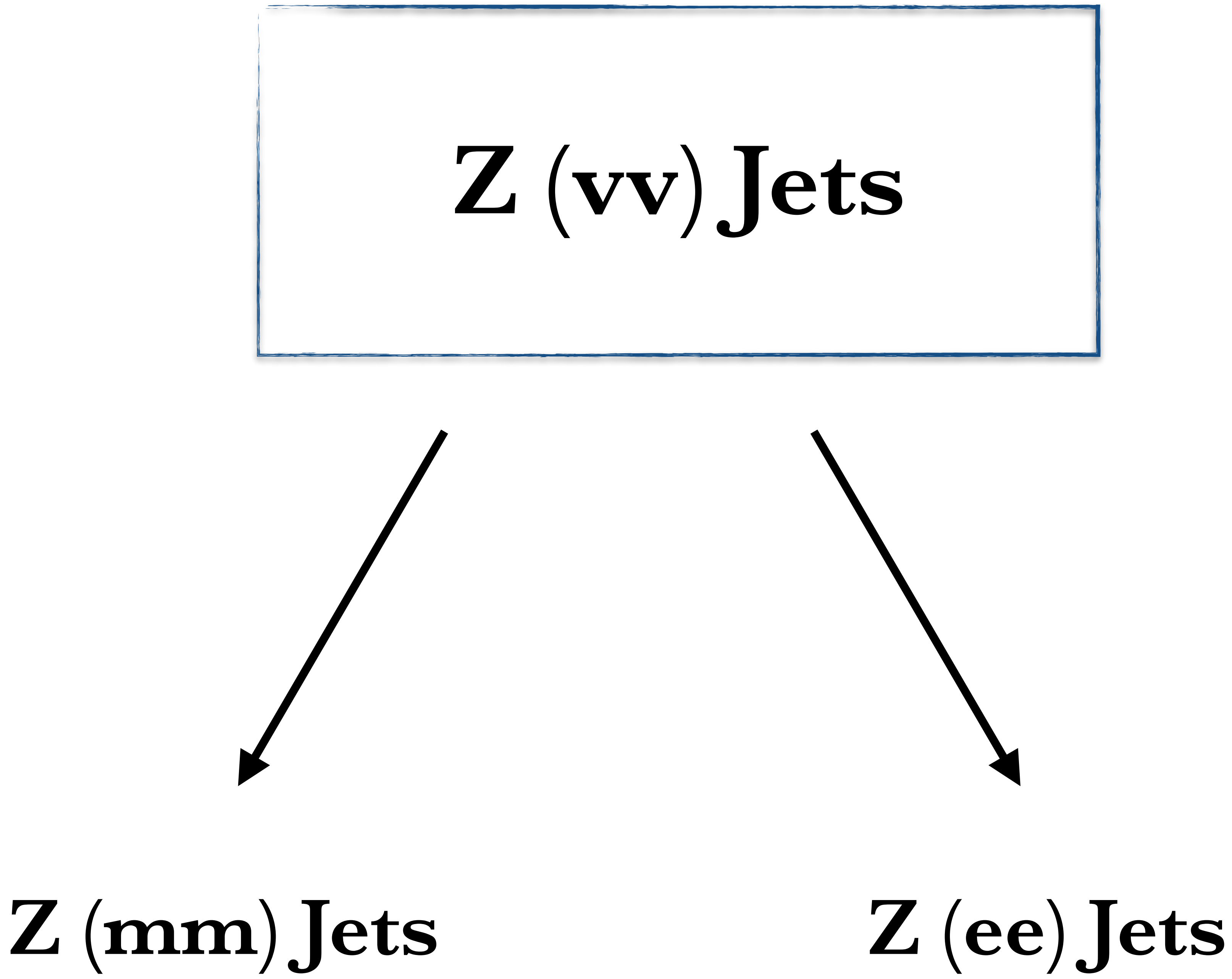
- **QCD Scale Uncertainty:**
  - Partially un-correlated on the Z/G & Z/W ratio
    - partial un-correlation is chosen to achieve an uncertainty on the ratio similar to single process uncertainty.
    - Leads to 10 - 15% renormalization, 1-10% factorization scale uncertainty
    - **Correlated across the MET/Recoil Bins**

Zeynep Demiragli

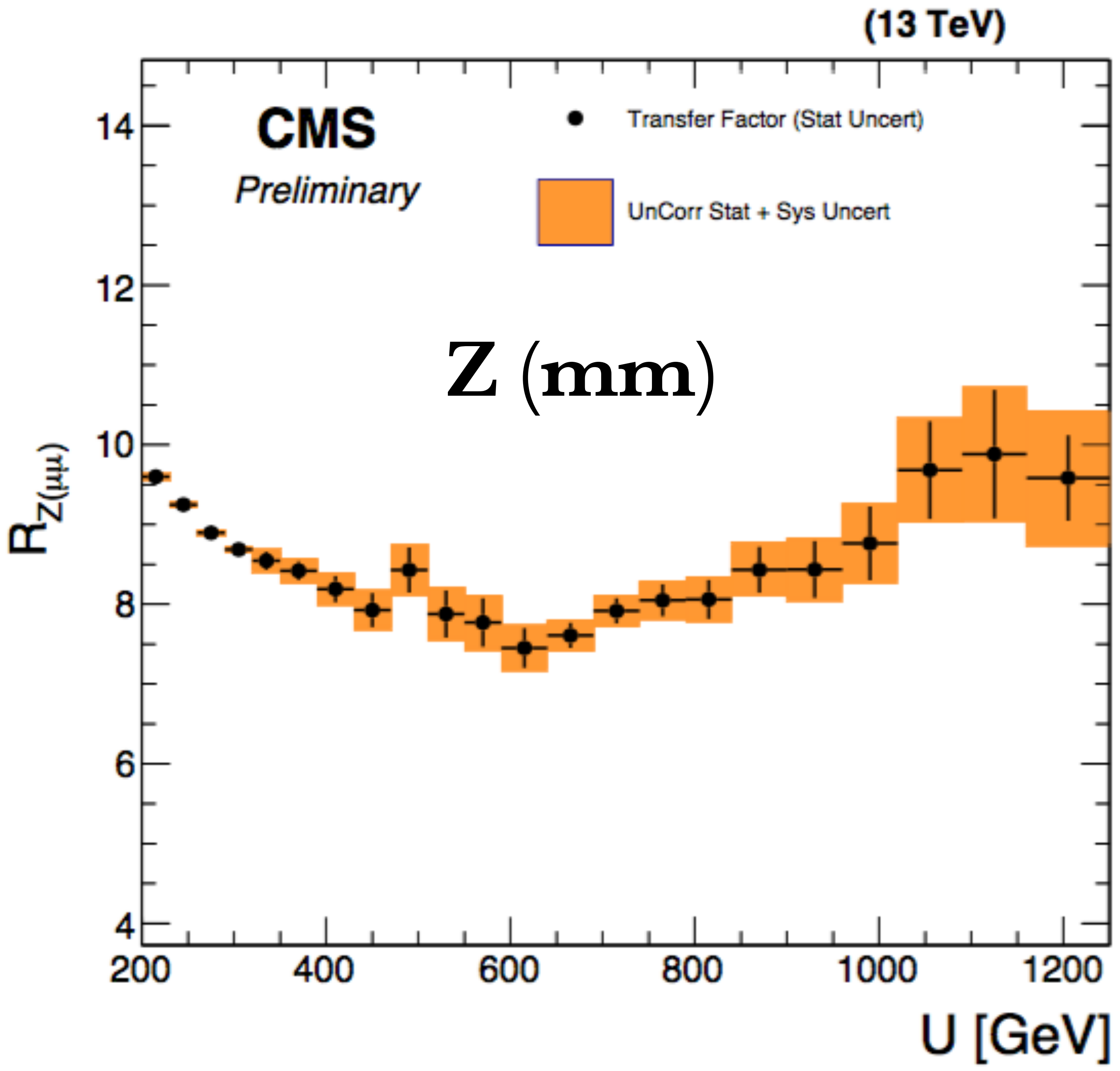
Background Estimation Method: Transfer Factors

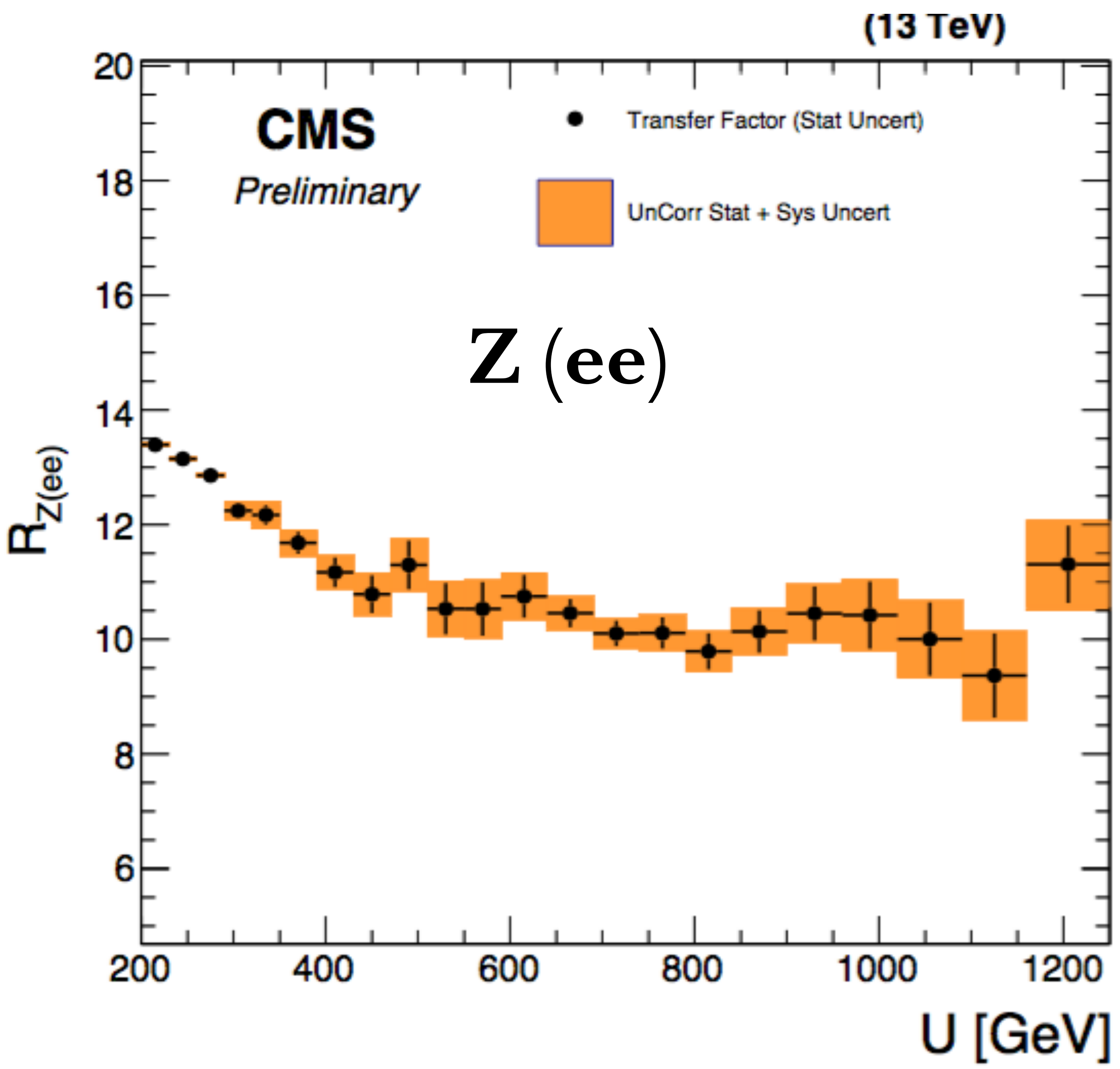
Why is this estimation useful?

- **Advantages:**
  - No theoretical uncertainty on the ratio and most of the experimental uncertainties will cancel too.
- **Shortcomings:**
  - Z(mm) and Z(ee) control regions will run out of events in the high recoil regions



Leading Unc : lepton scale factor + bin by bin statistical uncertainty





Zeynep Demiragli

Background Estimation Method: Transfer Factors

Gamma Jets

Z (vv) Jets

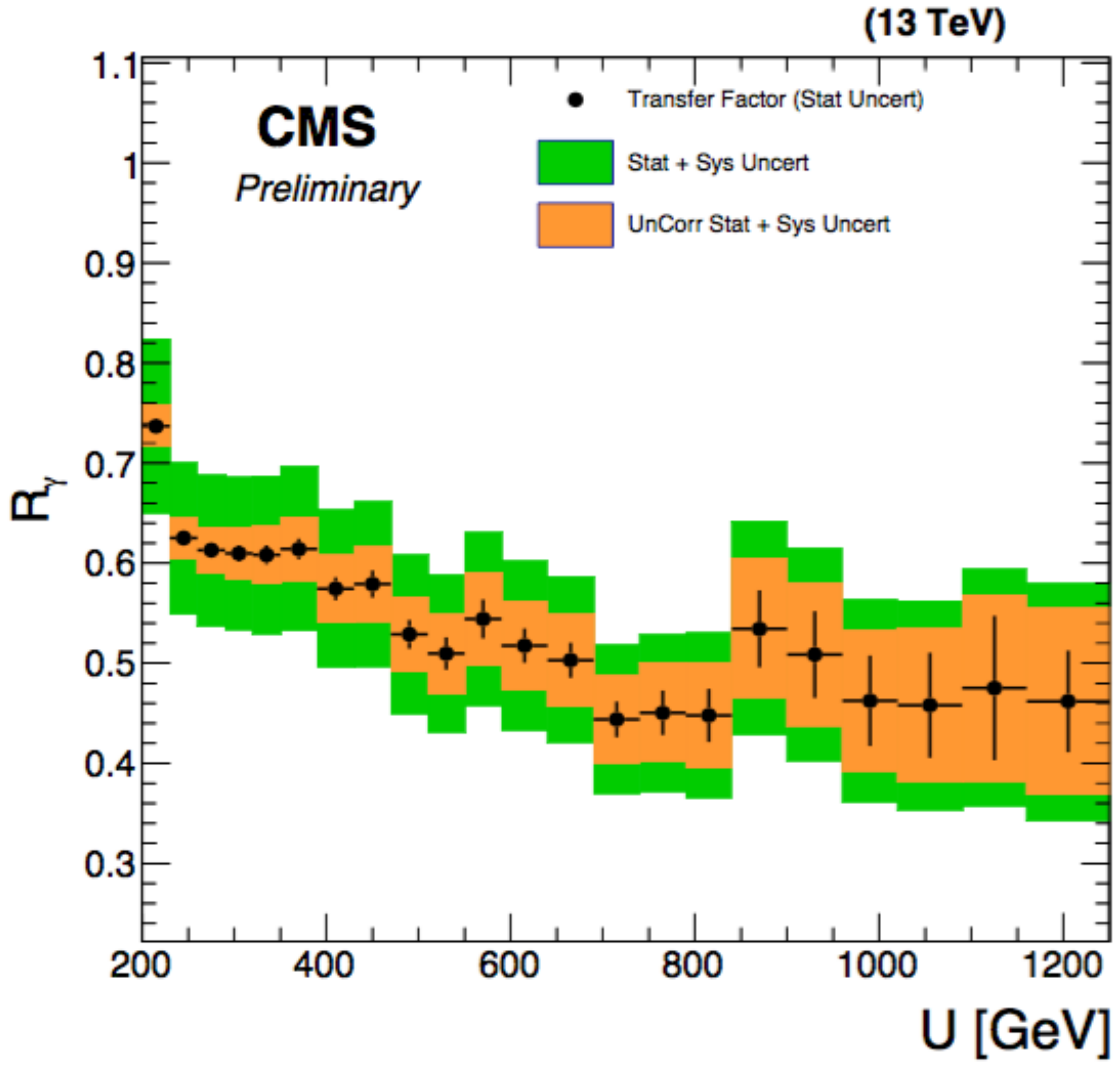
Z (mm) Jets

Z (ee) Jets

Why is this estimation useful?

- **Advantages:**
  - x20 more statistical power compared to Z CRs in the MET Tails
- **Shortcomings:**
  - Considerable theoretical uncertainties (qcd / ewk scale) but when combined with Z(ll) CRs, these uncertainties are constrained

Leading Unc : photon scale factor + ewk uncertainty + bin by bin statistical uncertainty + factorization and renormalization scale uncertainty



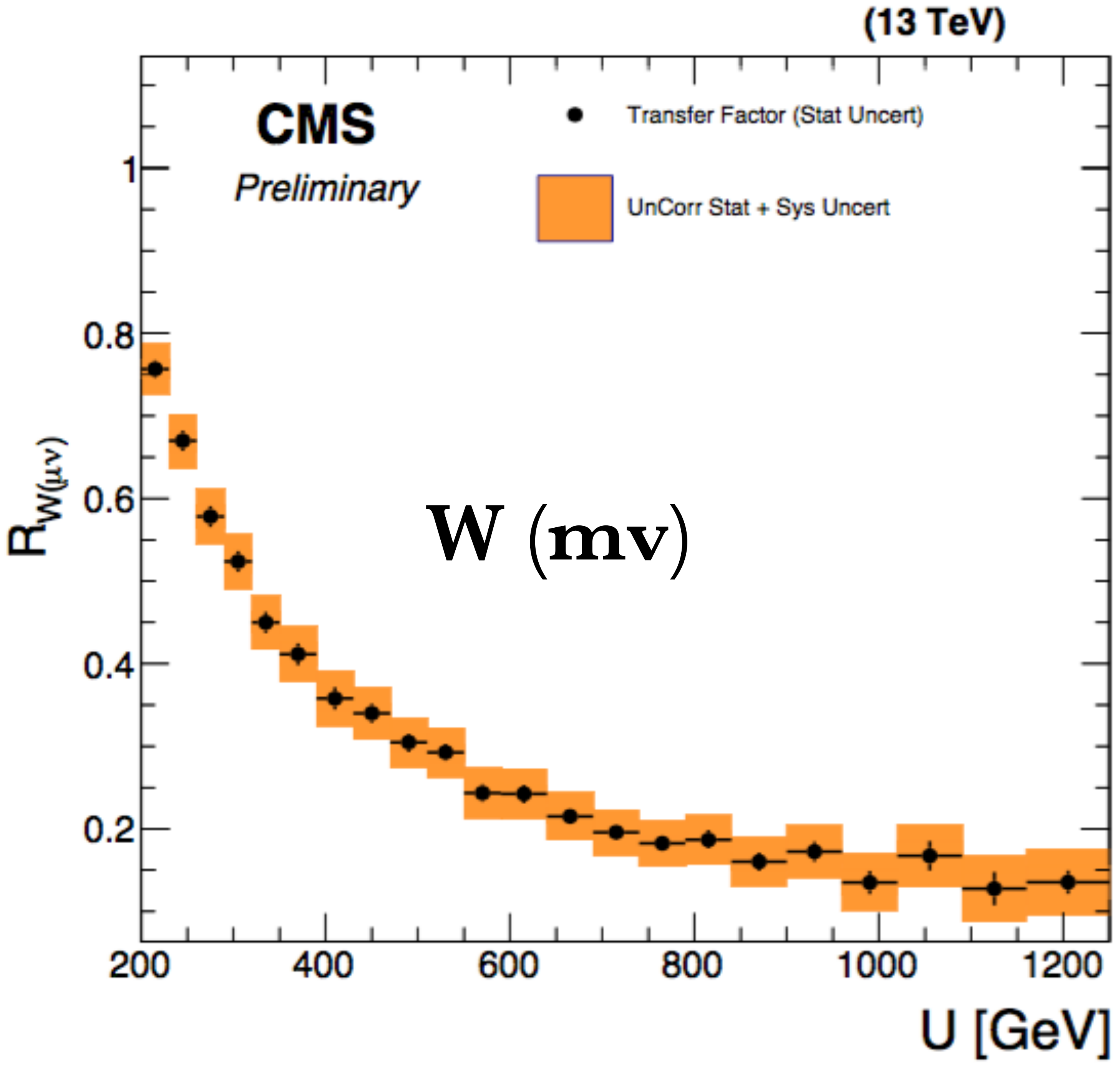
In addition to Zmm & Zee & W estimation improves the limits by ~10 %

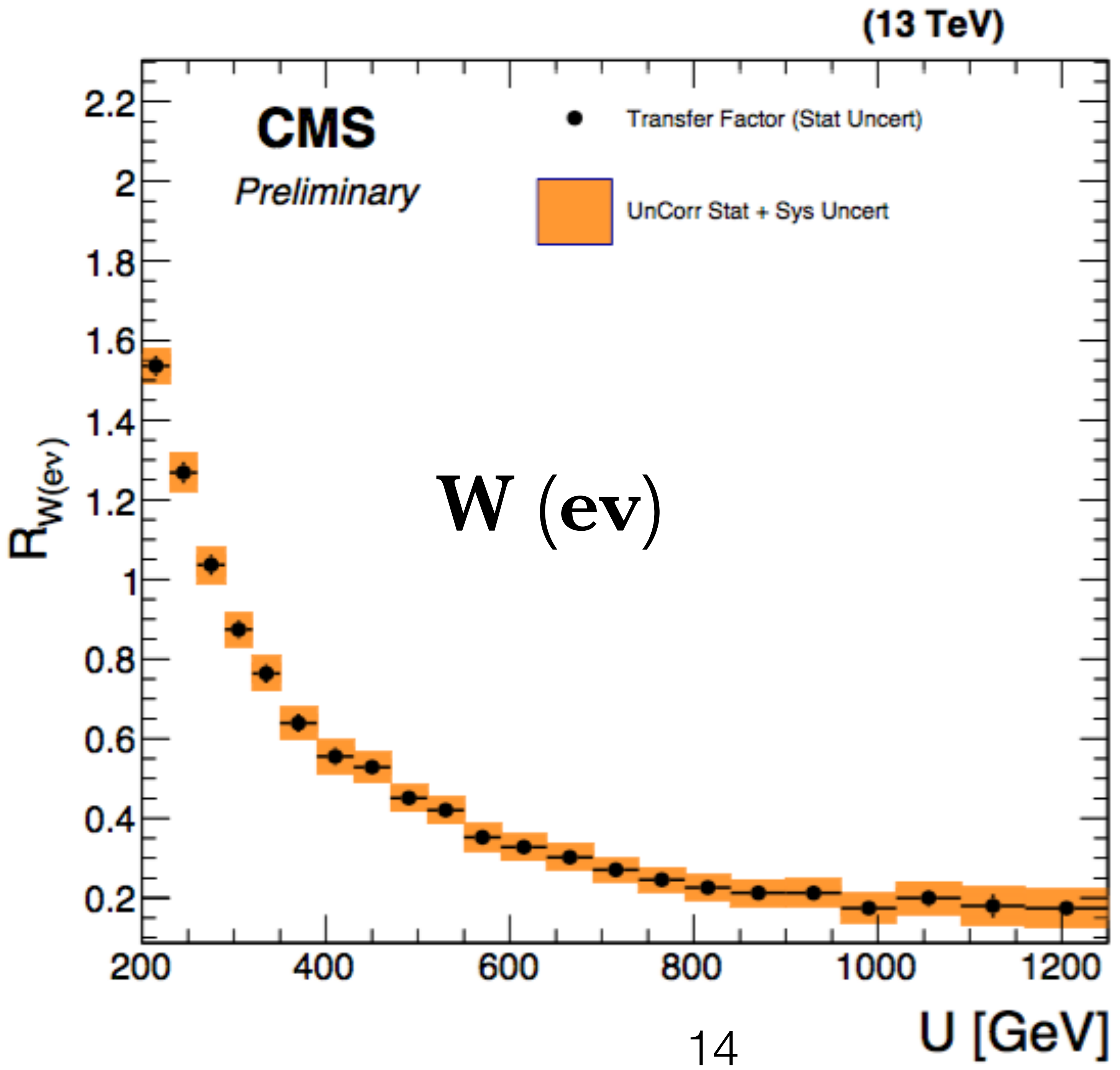
Background Estimation Method: Transfer Factors

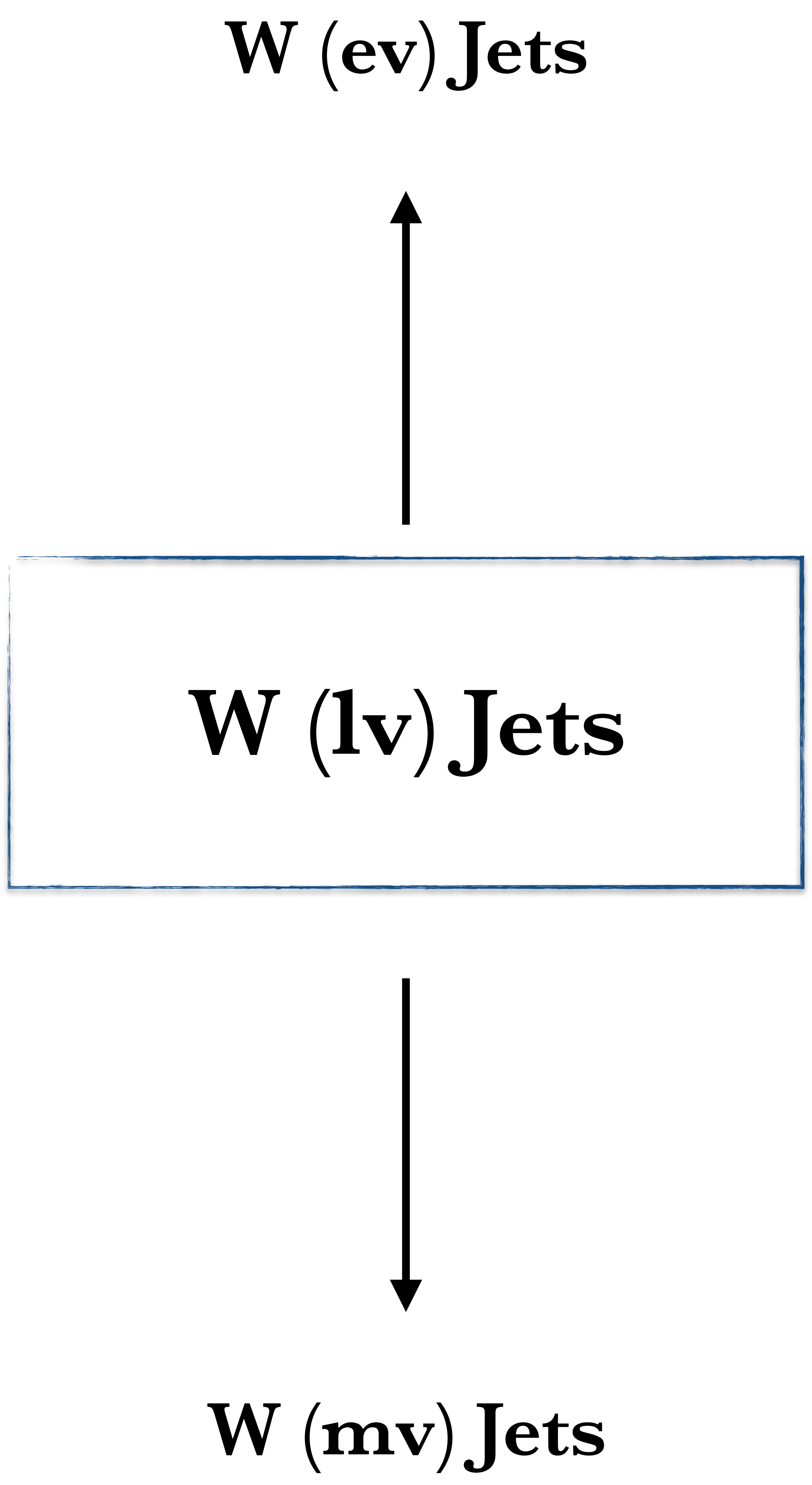
Why is this estimation useful?

- **Advantages:**
  - No theoretical uncertainty on the ratio
  - Most of the experimental uncertainties will cancel too.
- **Shortcomings:**
  - None (at least can’t think of one, this CR is pretty good)

Leading Unc : lepton scale factor + bin by bin statistical uncertainty





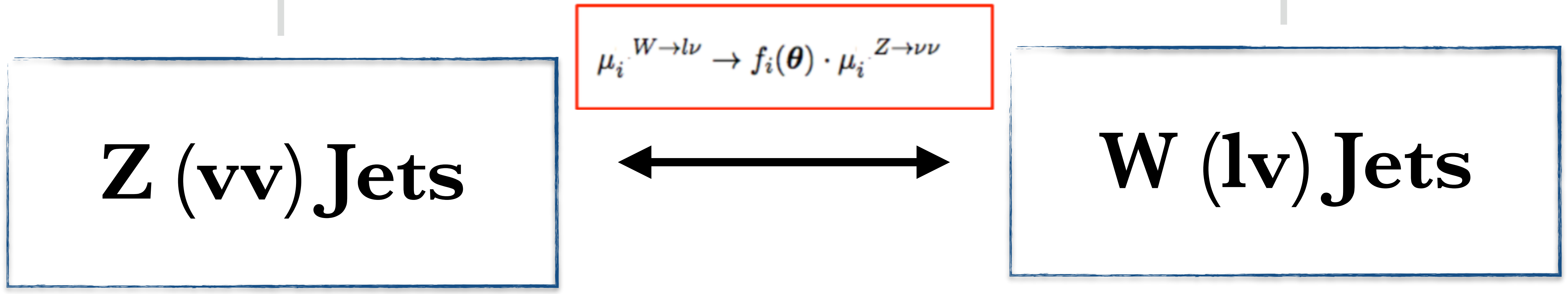


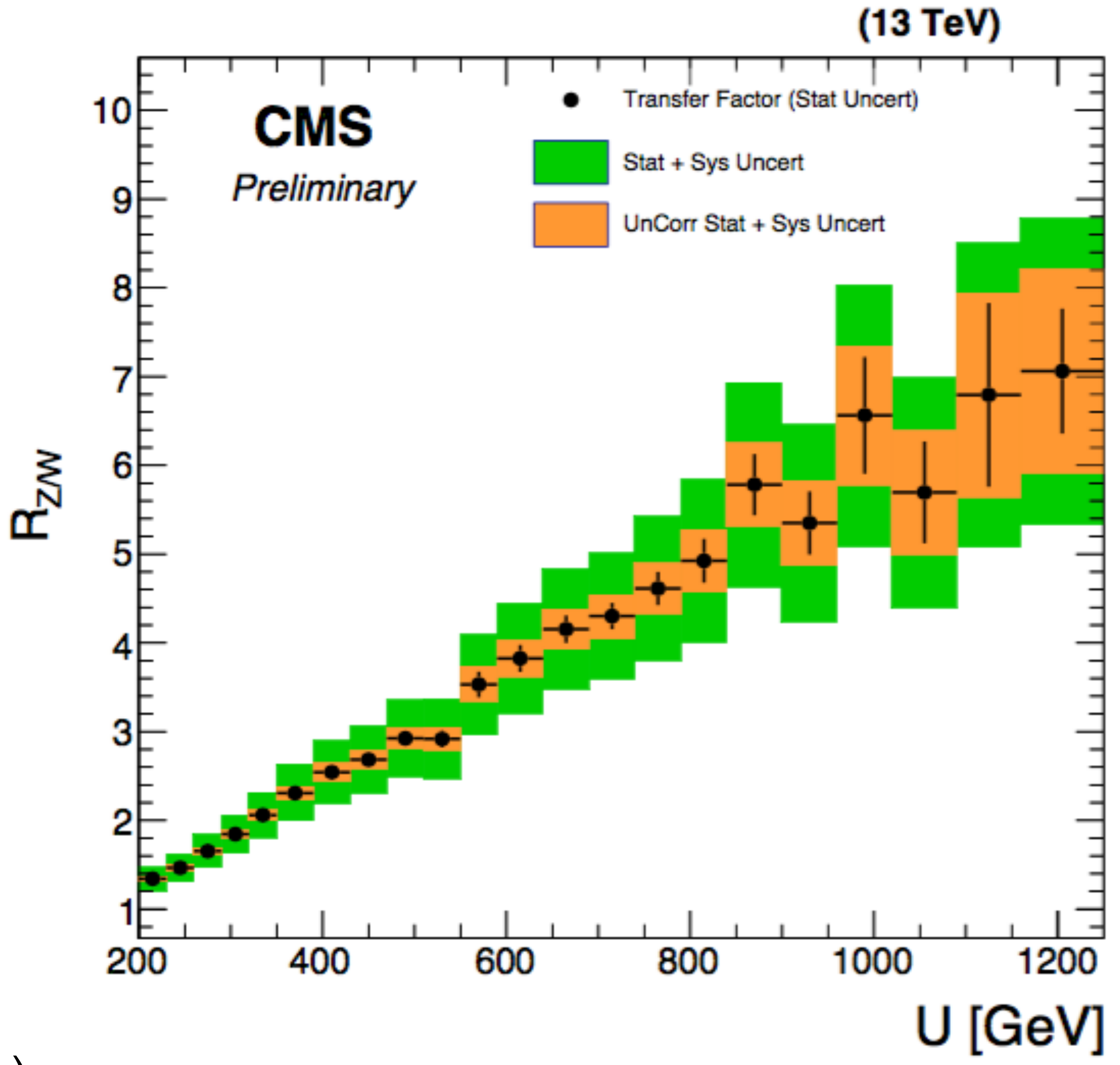
Zeynep Demiragli

Background Estimation Method: Transfer Factors

In addition to Zmm& Zee & Gamma Jet estimations improves the limits by ~7 %

Leading Unc : ewk uncertainty + bin by bin statistical uncertainty + factorization and renormalization scale uncertainty



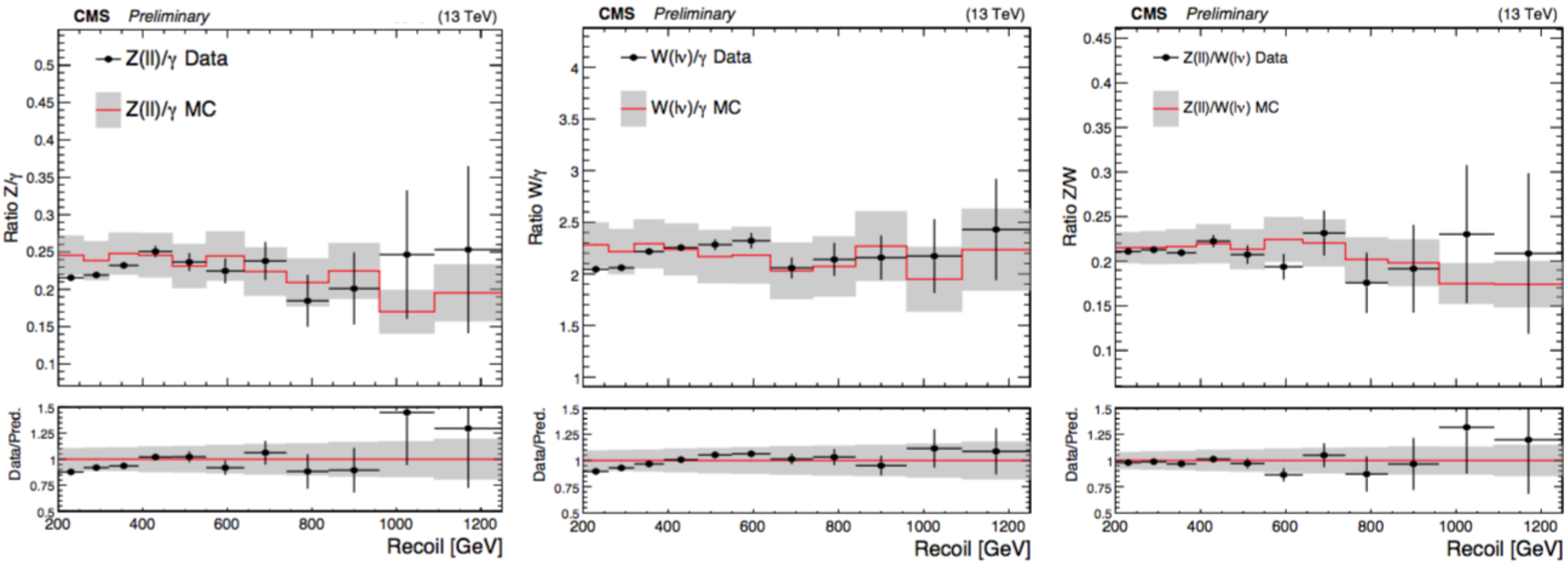


Why is this estimation useful?

- **Advantages:**
  - Additional statistical power at the MET tails
  - Lower theoretical uncertainties compared to Z/Gamma ratio (EWK uncertainties compared to Z/Gamma ratio is roughly halved)
- **Shortcomings:**
  - Would still need to include theoretical uncertainties (qcd / ewk scale)

Zeynep Demiragli

Data Validation For Transfer Factors (Pre Fit)



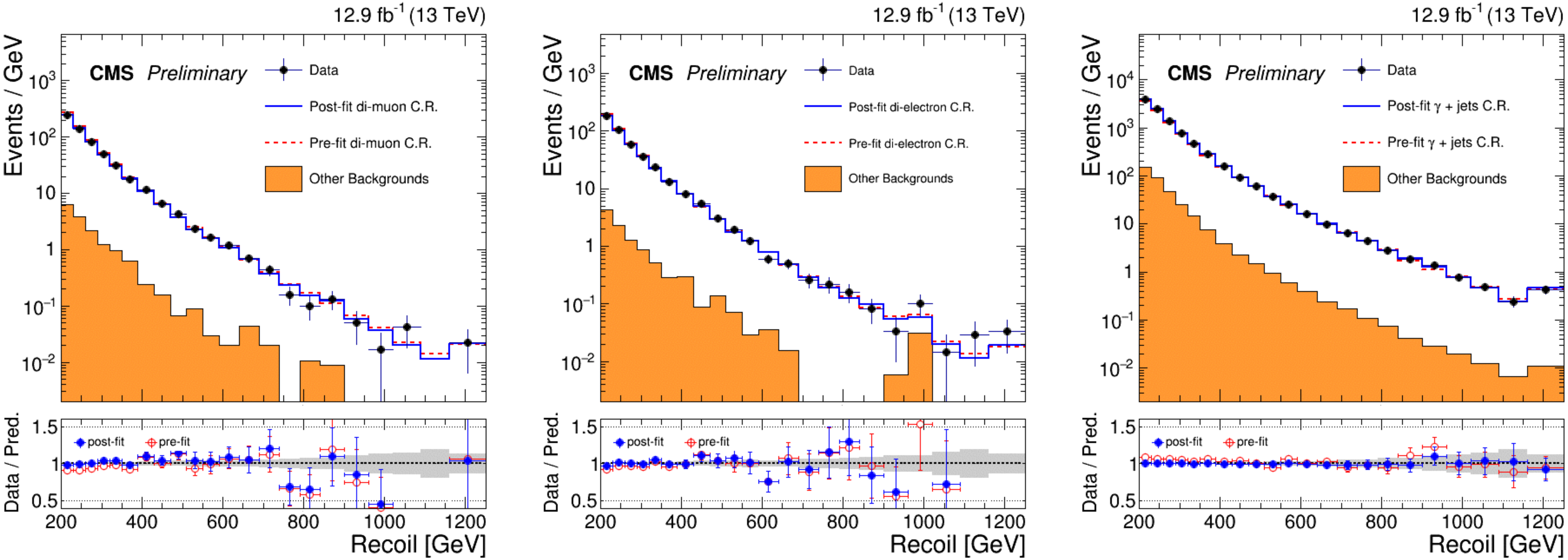
Black ratio from data and statistical uncertainties / Red from MC

Grey band is stat. + sys uncertainty on MC. Sys uncertainty includes theoretical uncertainties

Difference between data / simulation TF is covered by stat+sys uncertainty along the full recoil range

Simultaneous Fit Results …

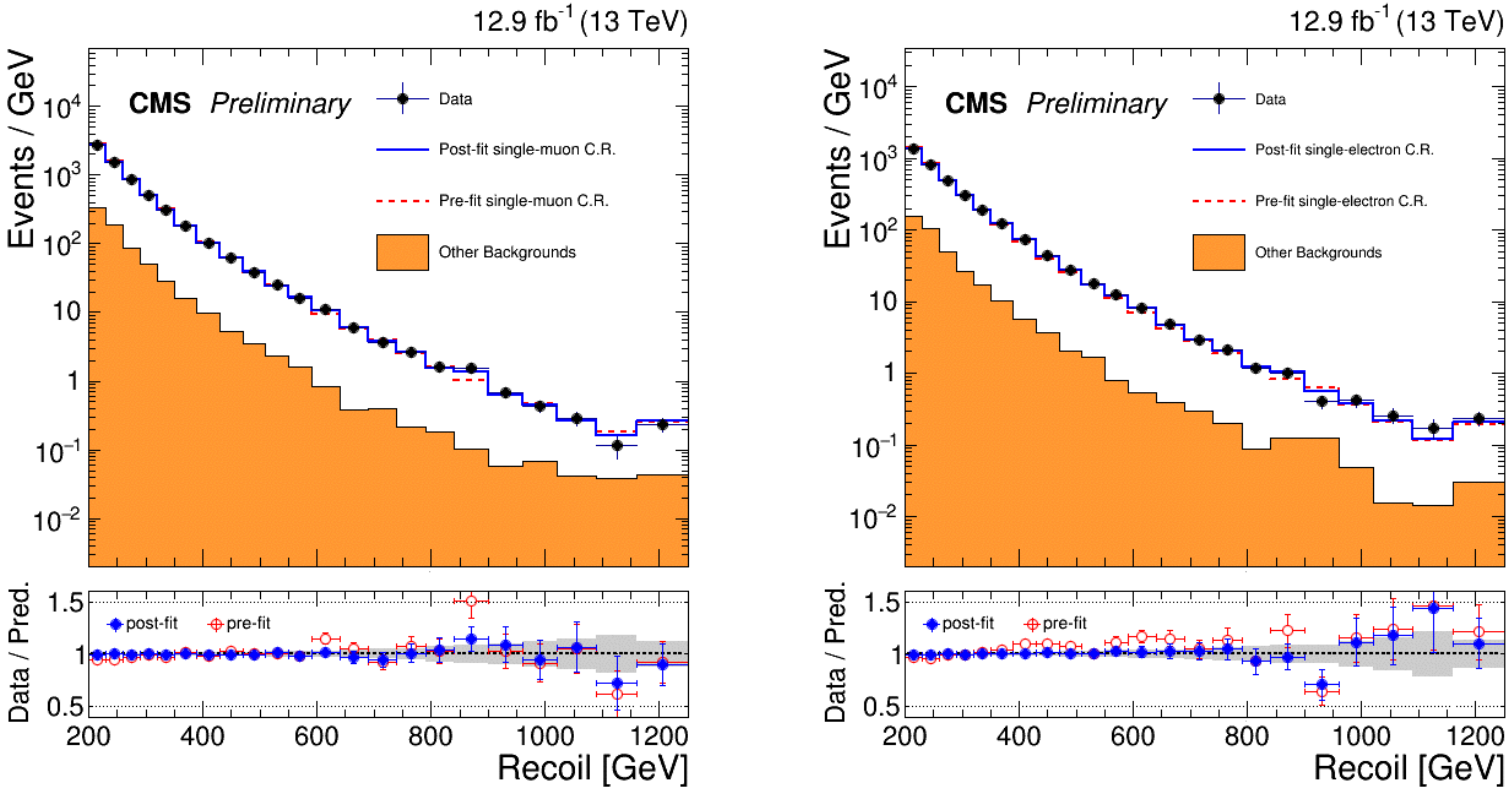
Post Fit in Monojet



Post-fit predictions match well data in all control regions. γ+jet statistically dominates the combined fit

Post-fit uncertainty in the high recoil bins ~ 15%

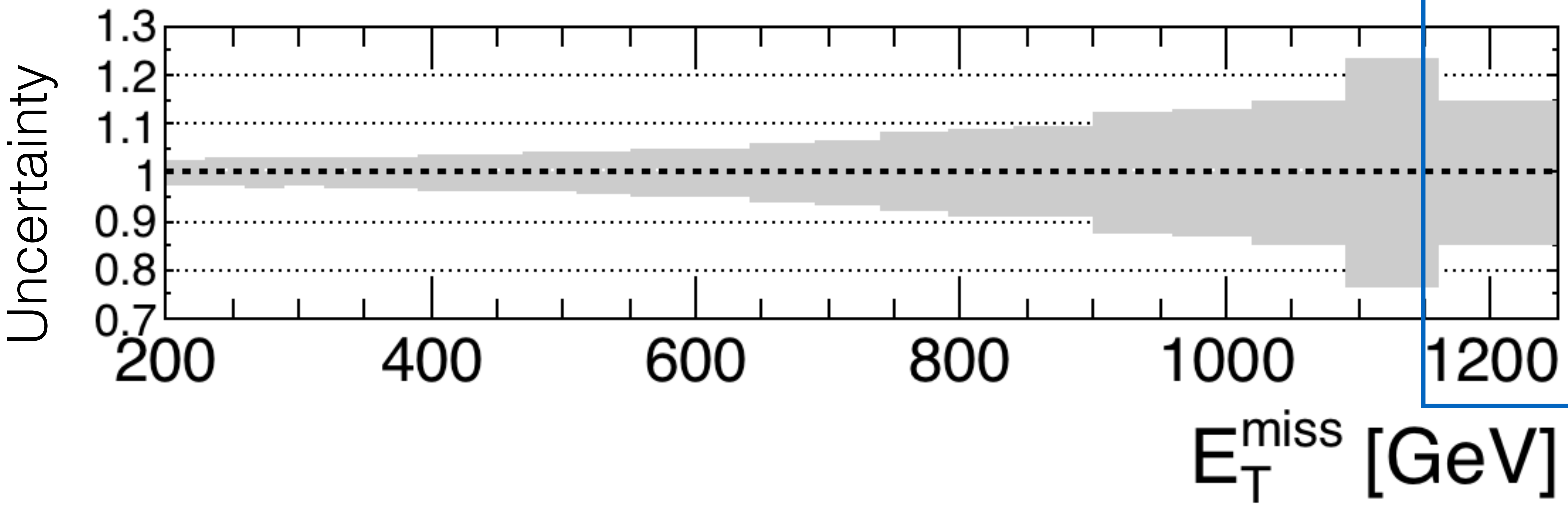
Post Fit in Monojet



Post-fit predictions match well data in all control regions. Statistical fluctuations at the tails of recoil.

Post-fit uncertainty in the high recoil bins ~ 15%

Total Uncertainty on the SR after the CR OnlyFit



- Last bin is overflow and has reduced statistical uncertainty

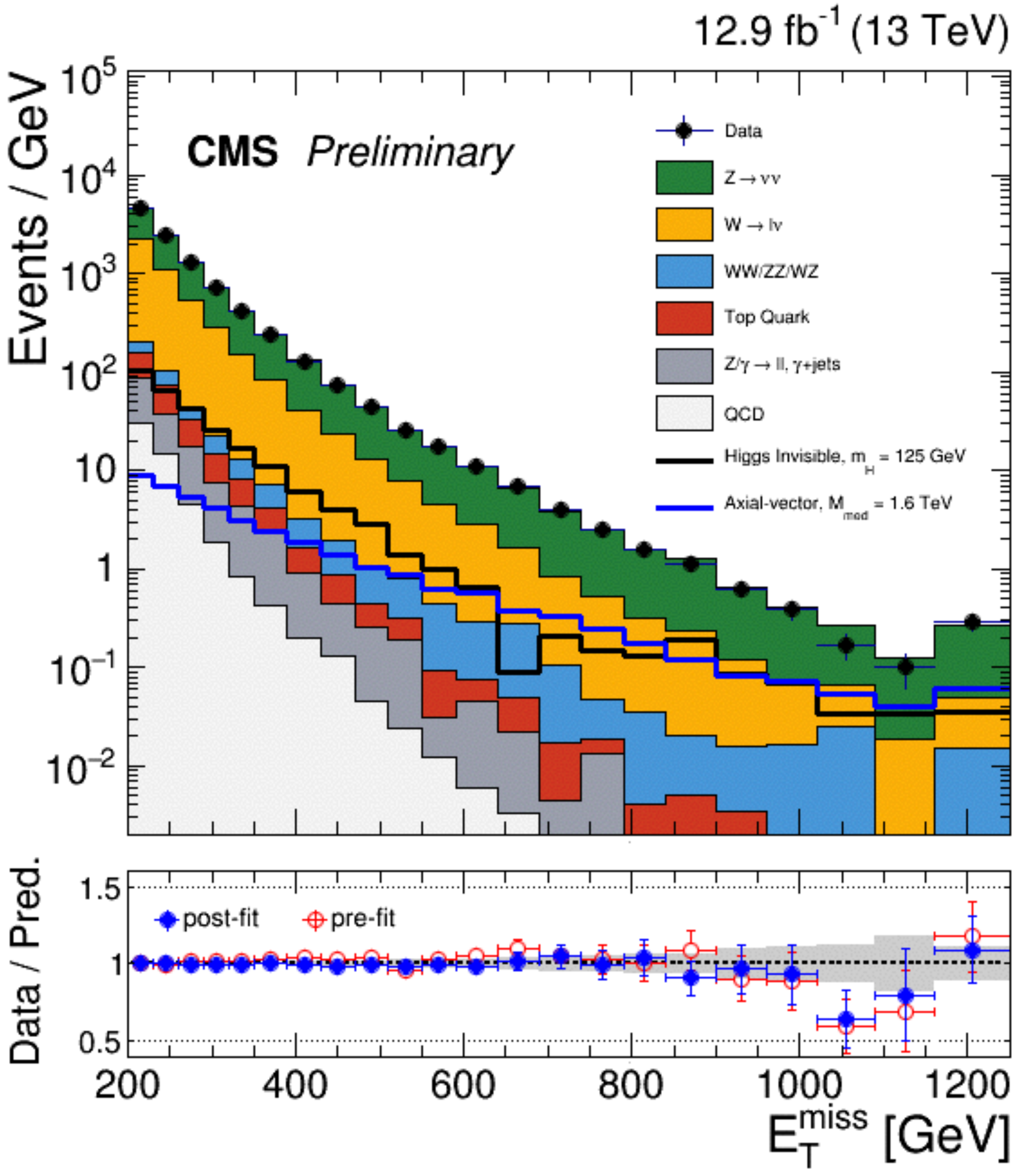
In the low MET region

- Z control regions drive the background fit
- Dominant systematic:
  - uncertainty on electron/muon efficiency

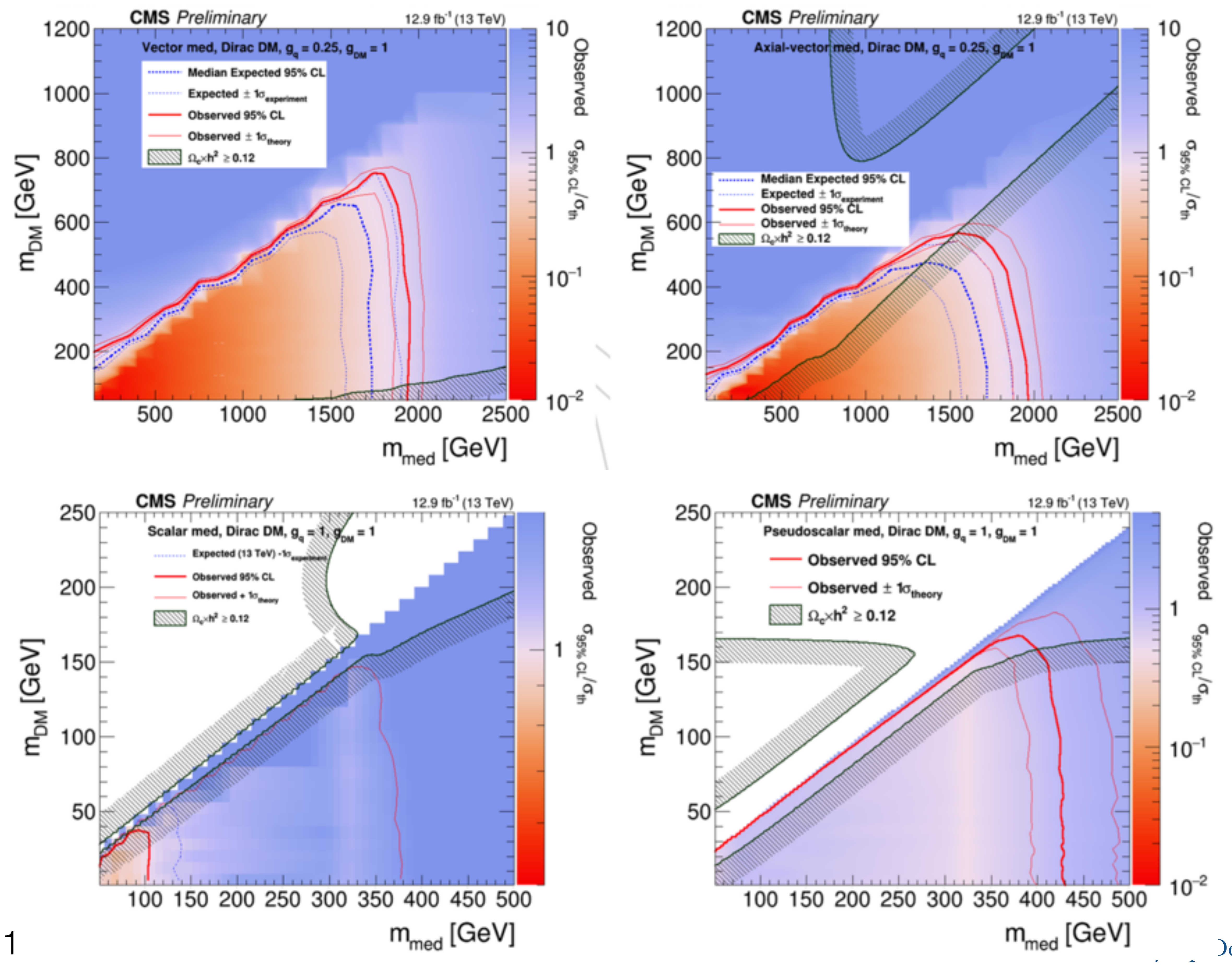
In the high MET region

- ɣ, W control regions drive the background fit
- Dominant systematic:
  - Theory uncertainties on Z/ɣ, Z/W ratios

Results & Conclusion



In CMS we have performed a 2 category (monojet / monoV) analysis where the leading backgrounds are estimated from 5 different control regions throughout a simultaneous fit.



Demiragli

wish list

1) Can we have some guidelines regarding the correlation of the QCD scale uncertainties in the Z/V ratios.

2) Should we correlate the theory uncertainties (QCD scale, EWK) across the pT spectrum. Is it reasonable to allow the low V pT region to constrain the theory uncertainties in the high pT region.

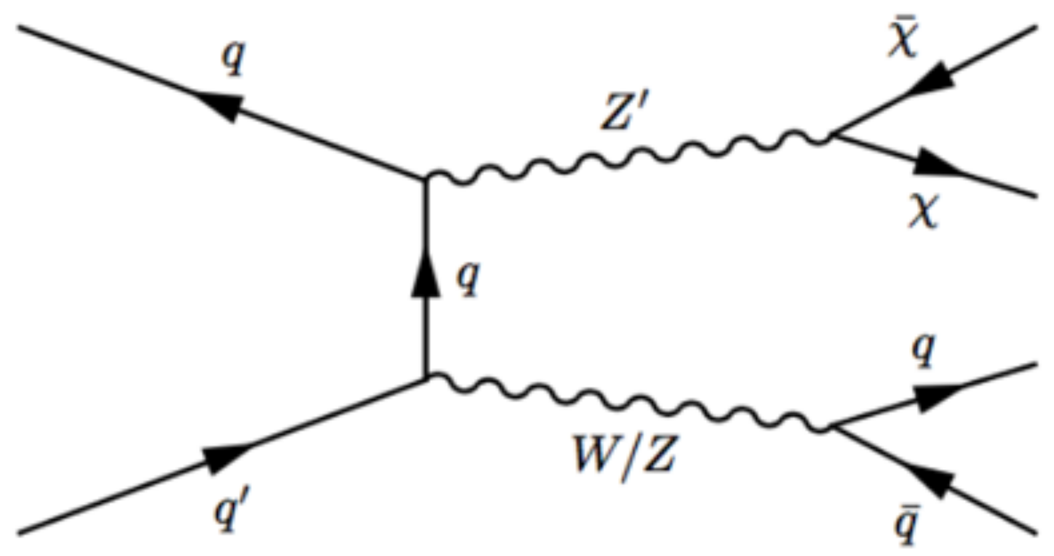
3) How can we compute the electroweak uncertainty? We currently conservatively take the full NLO EWK correction as the uncertainty.

4) Will we have a full integration of the NLO EWK corrections with the parton shower?

Back Up

MonoV Signal Generation Details

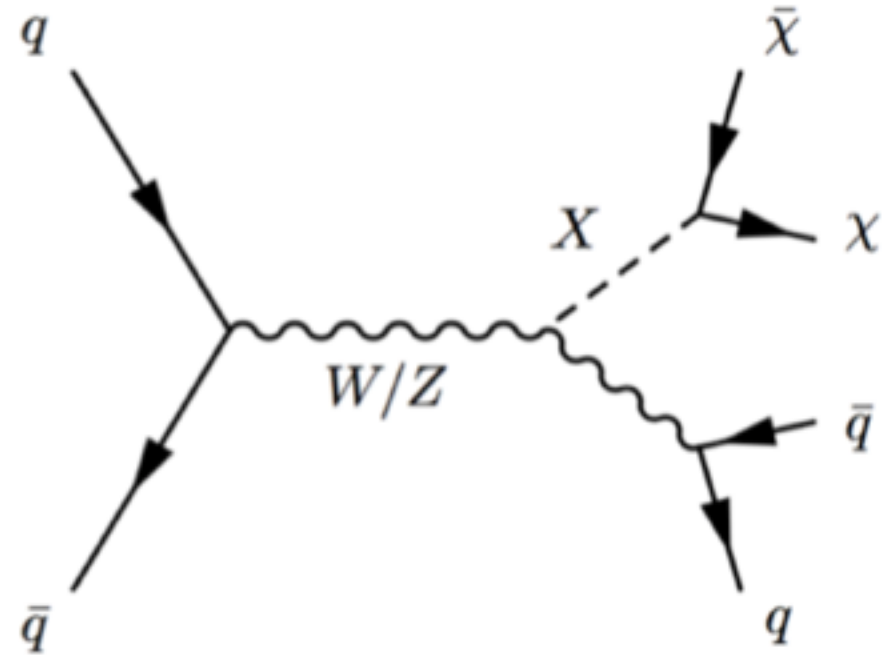
Axial Vector & Vector Mediators



Generator	Order in QCD/ EWK	Parton level jet multiplicity	Ren & Fac Scale	PDF
Madgraph	LO / LO	2	$m^2 + p_T^2$	NNPDF 3.0

Scalar Mediators

Generator	Order in QCD/EWK	Scale
JHUGEN	LO / LO	m^2

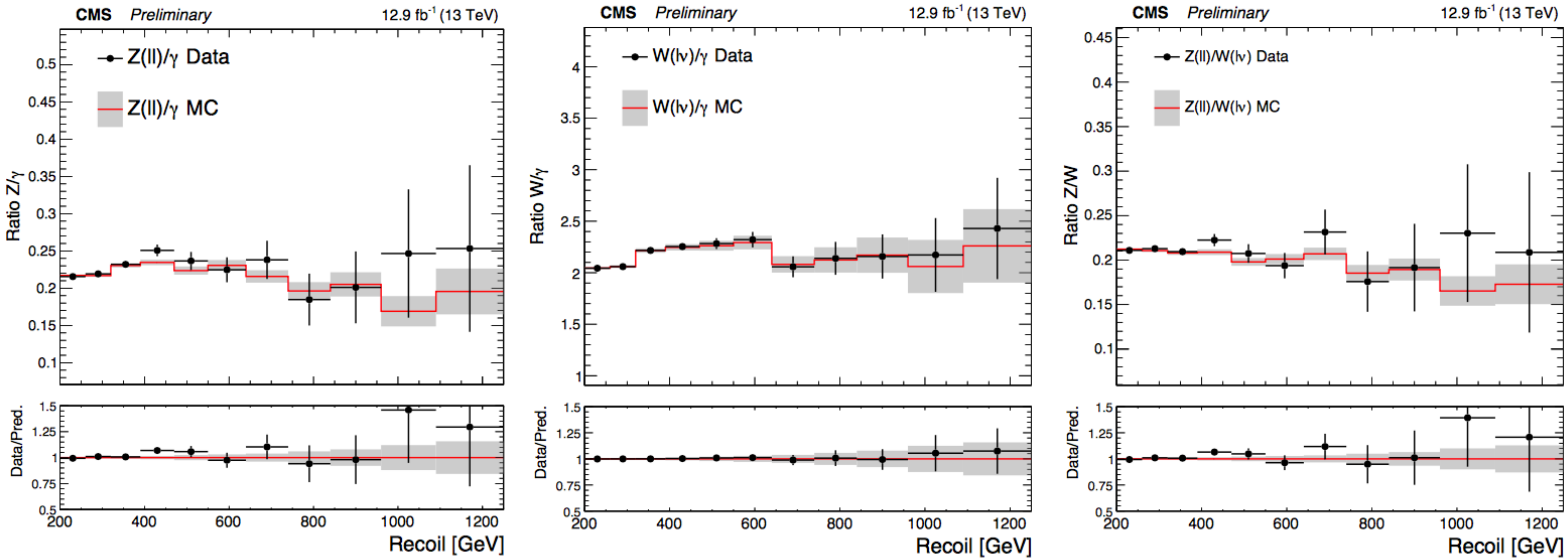


Cross sections are taking form the Higgs yellow report for the BSM Higgs:

- No xsec variation in DM Mass
- Using Higgs coupling
- Computing the BR to DM for a given model (~ 30% at high Mass)

Zeynep Demiragli

Data Validation For Transfer Factors (Post Fit)



Black ratio from data and statistical uncertainties / Red from MC

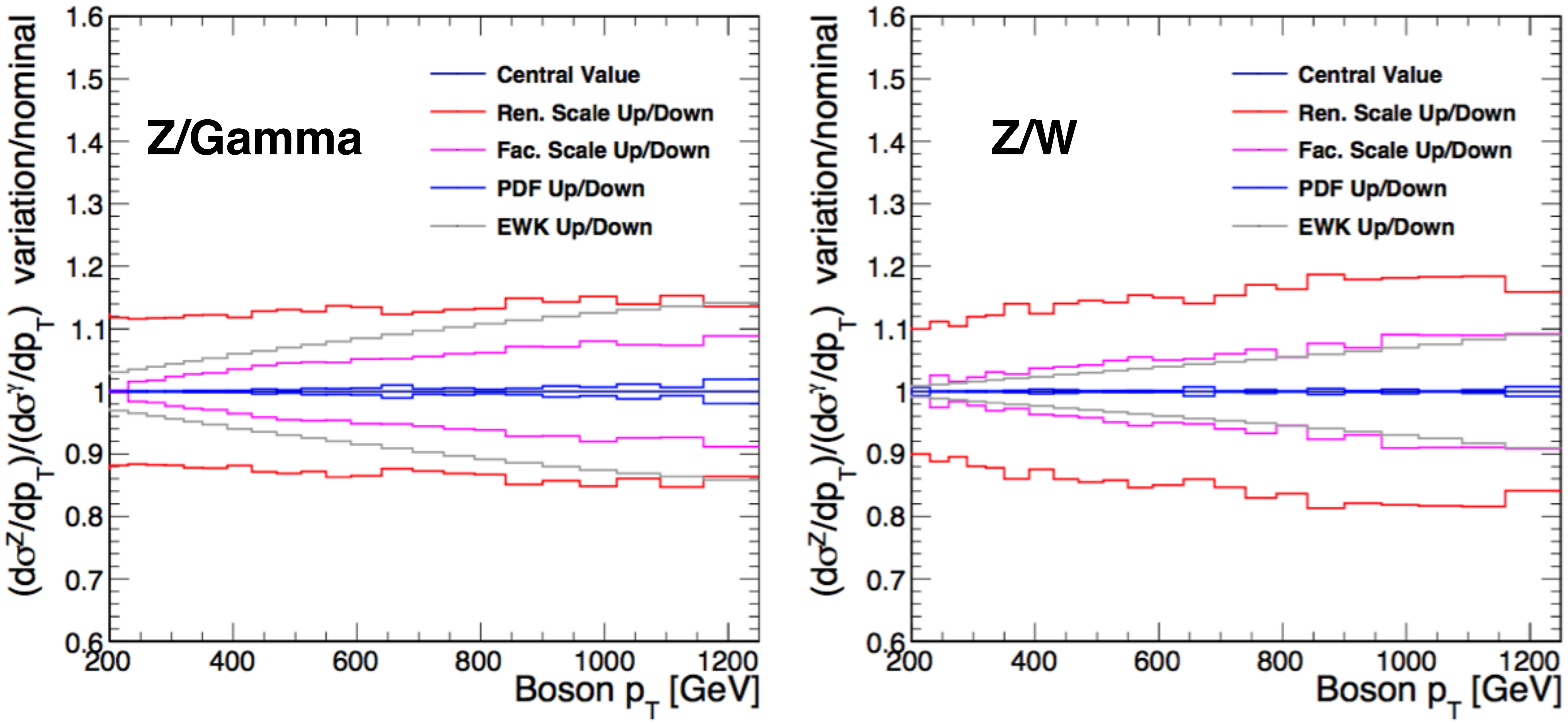
Grey band is stat. + sys uncertainty on MC. Sys uncertainty includes theoretical uncertainties

Difference between data / simulation TF is covered by stat+sys uncertainty along the full recoil range/

NLO corrections used in the analysis

- Z+jets and W+jets generated up to 2 additional partons at ME
- γ+jets generated with no additional partons at Born level
  - The NLO γ+jets sample is generated with the Frixione isolation on photons with parameters:
    - $\epsilon_\gamma = 1$, $\delta_0 = 0.4$, and $n = 1$
  - Base γ+jets MC, a simple cone isolation where no quarks or gluons are allowed within $\Delta R < 0.05$ of the photon
  - To derive consistent set of weights, identical Frixione isolation was applied to the base sample using the matrix-element level information.

Theoretical & Experimental Uncertainties: Overview



Experimental Uncertainties:

Source	Process	Uncertainty
Luminosity	All except V+jets	5%
Electron trigger	Electron CR	2%
Photon trigger	Photon CR	2%
E_T^{miss} trigger	Signal and muon CR	2%
E_T^{miss} recoil	All except V+jets	5%
Muon-ID efficiency	Muon CR per leg	1%
Electron-ID efficiency	Electron CR per leg	2%
Photon-ID efficiency	Photon CR	2%
Tau veto	$W(\ell\nu)+$jets(SR)	3%
Photon purity	QCD in Photon CR	40%
b-jet veto	Top	6%
	All remaining	2%
V-tag efficiency	Top (Mono-V)	13%
	Dibosons (Mono-V)	13%
Top p_T reweight	Top	10%
Top norm	Top	10%
Dibosons norm	Dibosons	20%
$Z(\ell\ell)+$jets norm	$Z(\ell\ell)+$jets (SR)	20%
$\gamma+$jets norm	$\gamma+$jets (SR)	20%
QCD norm	QCD (SR)	100%

Theory Uncertainties:

- QCD Scale: vary renormalization and factorization scales
- PDF: NNPDF 3.0 uncertainty
- NLO EWK: the size of the correction used as uncertainty

Event Selection : Control Region Overview

One or more jets in the event passing the cleaning requirements && min$\Delta\phi$(met, jets) > 0.5

Monojet: Leading jet pT > 100 GeV && Recoil > 200 GeV

MonoV: Leading AK8 Jet > 250 GeV && Recoil > 250 GeV && V-Tagging

Dimuon/Dielectron Control Region

Two two opposite-charged muons/electrons with tight id leading muon/electron

60GeV < $m_{\mu\mu/ee}$ < 120GeV

Recoil = $|Met + Pt^{\mu\mu/ee}|$

Electron/Muon veto & Photon veto & btag veto

Single Muon/Electron Control Region

Leading muon/electron passes the tight lepton selection

True MET > 50 GeV (electrons) && **MT < 160 GeV**

Recoil = $|Met + Pt^{\mu/e}|$

Electron/Muon veto & Photon veto & btag veto

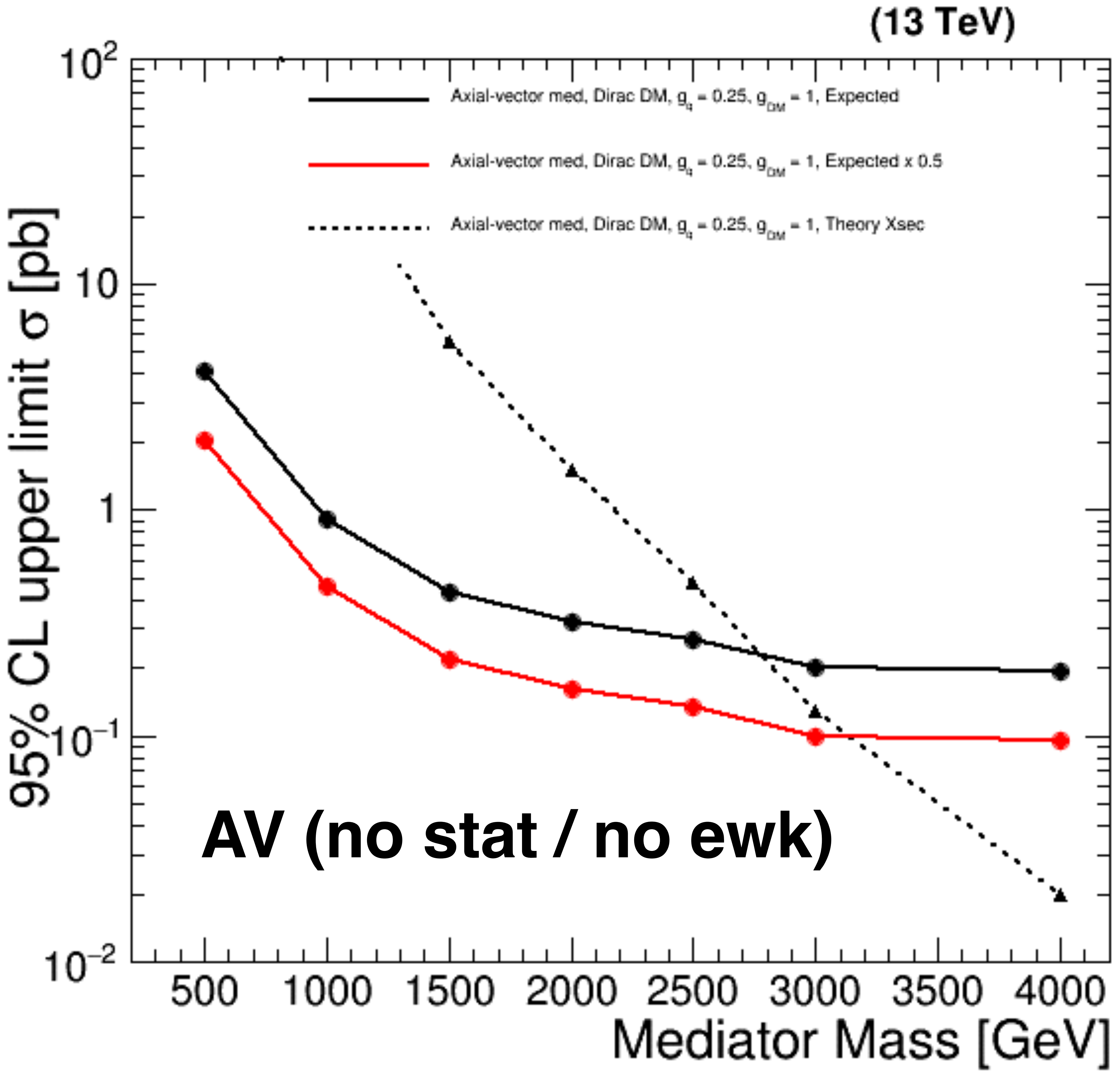
Photon + Jet Control Region

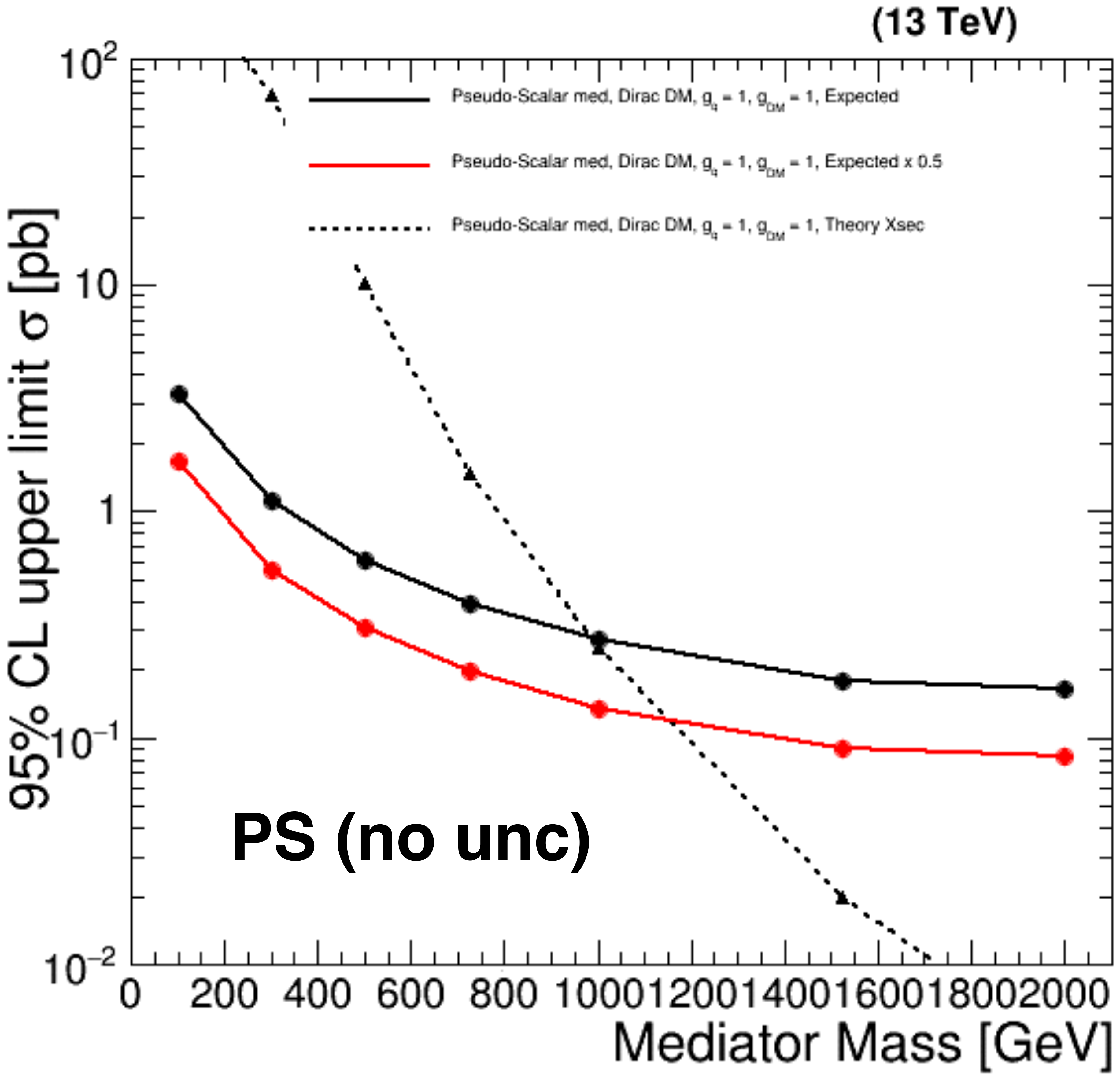
One well-identified photon in the event with pT >175GeV

$|Met + P_{T\gamma}|$

Lepton veto & btag veto

3000/fb projection: XSec Limits for AV & PS





Both for the Axial Vector & PS case the xsec limit is flatting out after 3 TeV and 1.5 TeV respectively. This implies that we are approaching the center of mass energy limit. No matter what we improve (redline is a x0.5 Expected limit) in the analysis cannot give us significance enhancement in our mass reach after these energy ranges.