Beyond MC: Building a Complete Model of Multilepton Events at CDF

UCSD Brownbag Meeting
6th July 2007

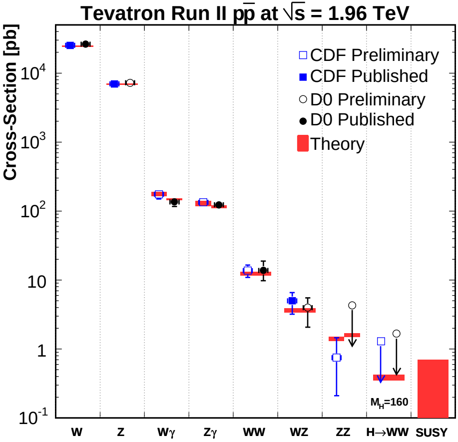
- Introduction
- Correcting and Scaling MC
- Data Driven Fake Calculation

Introduction

Same framework/calculations used for 2, 3, and 4 lepton channels

Physics Yield

- First 5 σ observation of WZ and anomalous couplings ($lll + \not{E}_T$)
- First 3 σ evidence ZZ (4 charged-lepton and $ll + \not{E}_T$)
- Most precise WW cross-section, best $H \rightarrow WW$ limits, working on anomalous couplings ($ll + \not{E}_T$)
- $ZZ \rightarrow lljj$: Limit anamalous couplings at high-p_T, should beat LEP limits



A flexible framework pays off...

Acceptance and Efficiency Maximized

Nine lepton types

- Tight + Loose Central Electron
- Forward Electron with + without a silicon track
- Muon with hits in central+"extended" muon chamber
- Muon as a minimum ionizing particle in the central + forward calorimeter
- Muon or Electron as an isolated track pointing at a calorimeter track

Four trigger paths

- Central Electron
- Forward Electron with $\not{E}_T$
- Central Muon
- Extended Muon

Need to model combinations of 2, 3, and 4 of these

The Master Formula: Calculating an Expected Yield

The contribution of a process ($Z \rightarrow ee$, WW, $t\bar{t}$, ...) to a selection is

$$N_{\text{exp}} = \underbrace{\sigma \times \mathcal{B} \times \epsilon_{\text{mc-filter}} \times \mathcal{L}}_{\text{Expected events } \textit{produced} \text{ in data}} \times \underbrace{\frac{N_{\text{reco}}}{N_{\text{gen}}} \times s_{\text{id}} \times \epsilon_{\text{trg}}}_{\text{Efficiency} \times \text{Acceptance}}$$

where

σ	Process cross-section
$\mathcal{B}$	Branching fraction
$\epsilon_{\text{mc-filter}}$	Presimulation Monte Carlo filter efficiency
$\mathcal{L}$	Data sample luminosity
N_{reco}	Event reconstructed in the MC sample
N_{gen}	Event generated in the MC sample
$s_{\text{lep}} \equiv \frac{\epsilon_{\text{lep},data}}{\epsilon_{\text{lep},MC}}$	Lepton Id efficiency scale factor
ϵ_{trg}	Trigger efficiency (simulation not used in CDF)

- Lots of other things to correct: jet energy scale, track momentum, calorimeter energy (leakage),...

Complication I

s_{lep} and $\epsilon_{\text{trigger}}$ depend on the kinematics of the each event

- In CDF, even Ldepends on the leptons in the event
  - detectors must be labeled good $\Rightarrow$ good run lists (lots of bookkeeping
- Event by event corrections:

$$N_{\text{exp}} = \sum_{i \in \text{MC sample}} sf_i$$

where

$$sf_i = \frac{\sigma \mathcal{B} \epsilon_{\text{mc-filter}} \mathcal{L}_i}{N_{\text{gen},i}} s_{\text{lep,i}} \epsilon_{\text{trg},i}$$

then

$\mathcal{L}_i$	Data sample luminosity for a good run list
$N_{\text{gen},i}$	Event generated in the MC sample for a good run list
s_{lep}	$= \prod_{l \in \text{leptons}} s_{\text{lep},l}(p_T, l, \eta_l)$
ϵ_{trg}	Trigger efficiency calculation more complicated...

Complication II: Trigger Efficiencies

- We combing multiple trigger line together to get the maximum efficiecy
  - The efficiency for a lepton category to fire a particular trigger is determined separately (beyond the scope of this talk)
- Because we use the logical OR of several triggers these efficiencies need to be combined
- Simple example: two leptons, one independant line for each

$$\epsilon_{\text{trg}} = 1 - (1 - \epsilon_1)(1 - \epsilon_2)$$

- We also use a trigger that requires a forward electron and missing energy, no do central muon + two forward electrons

$$\epsilon_{\text{trg}} = 1 - (1 - \epsilon_\mu)\left[1 - \epsilon_{\not{E}_T}(1 - (1 - \epsilon_{e1})(1 - \epsilon_{e2}))\right]$$

- Prescales are similar, but may be correlated between triggers causing more complication

Determining a lepton efficiency scale factor

The General Idea

- measure the ratio between a very loose “probe selection” and the full selection in data and MC
- probe efficiencies are measured elsewhere, and a generally high an well modeled, e.g.
  - the chance for an electron to produced a reconstructed shower in the EM calorimeter
  - the chance for a muon to produce a track (may not be that well modeled in forward region)

Determining a lepton efficiency scale factor

Proceedure

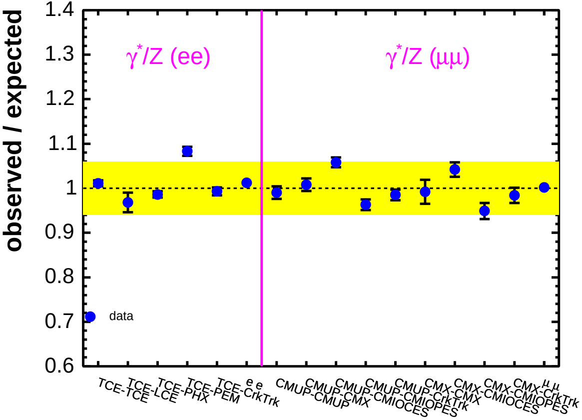
1. Select Z candidates using a “tag” lepton plus a lepton that either passes or fails the selection being studied
   - T = tagged
   - S = selected
   - F = passes probe selection, but not full selection
2. Calculate yields for the passed N_{TS} and failed N_{TF}, use Z mass sidebands to subtract background
3. Calculate efficiencies:
   - if tag is a non-overlapping category with the selected, it's simple:

$$\epsilon_{probe \to selected} = \frac{N_{TS}}{N_{TS} + N_{TF}}.$$

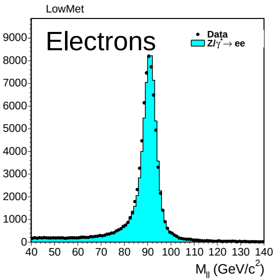
   - if tag and selection are the same (i.e. T=S), then the combinatorics is a bit trickier

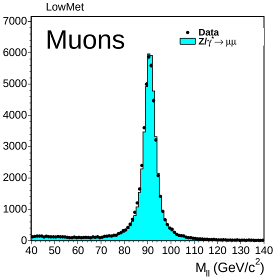
$$\epsilon_{probe \to selected} = \frac{2N_{TT}}{2N_{TT} + N_{TF}}.$$

Cross-check the MC Yield predictions with Drell-Yan



Data/MC ratios for all triggerable combinations





Data driven jet/QCD fake rate: Introduction

Monte Carlo cannot be trusted for fake leptons

- Fakes are due to fluctuations of the shower shapes in the non-perturbative region of QCD (isolation, low multiplicity high-p_T particles,...)

Strategy

- extract a “fake rate” from multi-jet QCD data, has few real leptons
- apply it to a sample with less leptons that the signal mode
  - E.g predict lepton+fake lepton from the single lepton sample applying a fake rate
- It’s an extrapolation from non-lepton like jets to lepton like jets
- Reduce the “distance” of the extrapolation, by extrapolating from “lepton-like jets” instead of all jets
  - The popular wisdom “jet is jet is jet” is not true, some are identified as leptons!

Data driven jet/QCD fake rate: Denominators

We call the objects we extrapolate from “denominators” (i.e. the denominator of the fake rate definition)

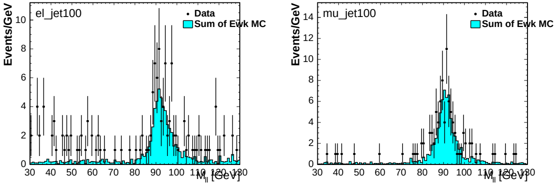
- Based on the same objects as the identified leptons
  - Correlation of the measured energies is straight forward
  - Occurance of the base objects may be subtle
    - E.g. adding a track χ^2 cut to the muon categories supresses mismeasured tracks which are very sample dependent
- Fiducial cuts
- Loose isolation cut

Data driven jet/QCD fake rate: Proceedure Outline

1. Calculate in the jet triggered data

$$\text{Fake Rate} = \frac{\#\text{Identified Leptons}}{\#\text{Denominator Objects}}$$

   - Leading jet vetoed to reduce trigger bias
2. Correct for W and Z contamination using Monte Carlo
   - Slow jet trigger turn-on curve hard to model, check with Z peak in jet triggered data



   - Also number of events with a Z plus a 100 GeV jet may not be that well modeled

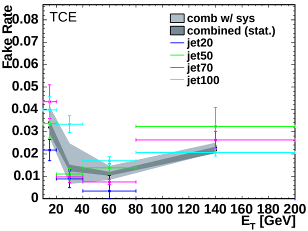
Data driven jet/QCD fake rate: Proceedur (con't)

3 Scale data l+"denominator object" events by measured fake rate
- Loop over l+denom events
- Treat denom as if it is a second lepton for selections and plotting
- If two denoms consider each separtely (unless you want the double fake background)

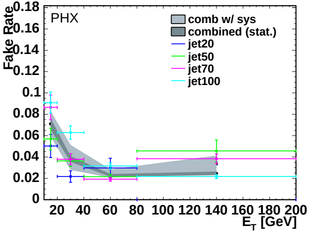
4 Use MC to correct for event that triggered on the fake ...
- An actual fake should fire the trigger at the same rate as a real lepton
  - Otherwise we could use the trigger to supress the fakes
- Denominator objects are less lepton like than actual fakes, so the trigger is different
- Use the ratio of non-triggerable to triggerable from MC to correct:

$$B_{Z\rightarrow \mathrm{nontriggerable}} = B_{Z\rightarrow \mathrm{triggered}} \frac{N^{MC}_{Z\rightarrow \mathrm{nontriggerable\ +\ triggerable\ denominator\ object}}}{N^{MC}_{Z\rightarrow \mathrm{triggerable\ +\ denominator\ object}}}$$

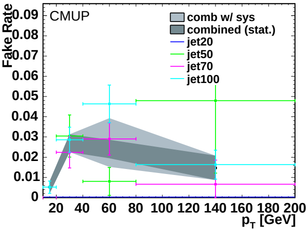
Data driven jet/QCD fake rate: Example Plots



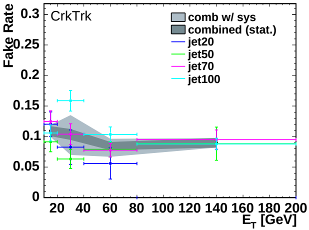
Tight Central Electron



Forward Electron w/Track

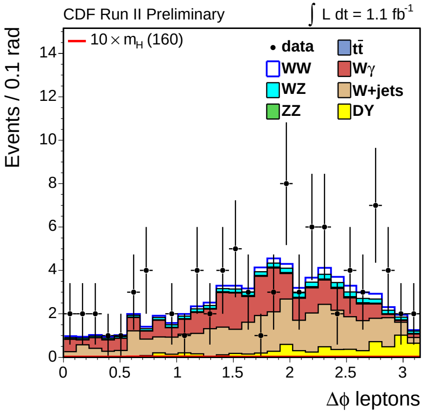


Central Muon

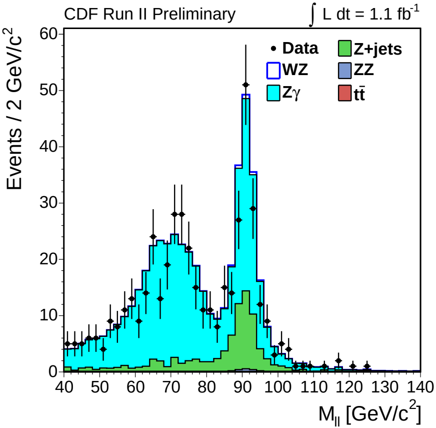


Track in Crack

Cross-checks of fake backgrounds



Same-sign $ll + \not{E}_T$

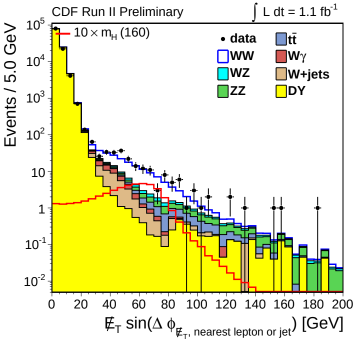


Trileptons low $\not{E}_T$

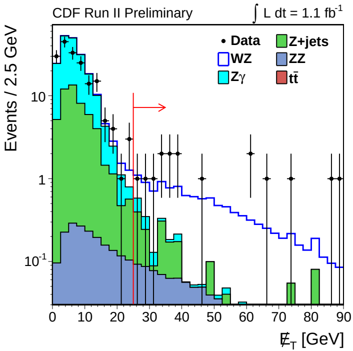
- Systematics on fake calculation are estimated to be $\approx$ 25% base on the variations due to jet samples
- Variation in the isolation cut of the denominator is also studied

Summary

- Now look at the beautiful signals:



WW ... (okay no Higgs)



WZ

- There are lots of details to deal with
- Details can be put in a generalized framework