W (eν) + Jets

Data Driven Ttbar Rejection

Maria Fiascaris
University of Oxford

in collaboration with
Alessandro Tricoli (RAL)

Outline

- Data Samples and Event Selection Cuts
- Effect of Missing Et cut on background contamination
- Overview of Method for ttbar background subtraction
- Results
- Technical Details
  - B-Tagging studies
  - B-Jet Veto studies
  - Fitting the Missing Et shape
- Conclusions and Outlook

Samples & Selection Cuts

- **Samples (v. 12 Athena):**

Signal (Alpgen)

Sample	Process	Npartons
6101-6106	W (e ν) + Np	0-5
6280-6283	W (l ν) + bb Np	0-3

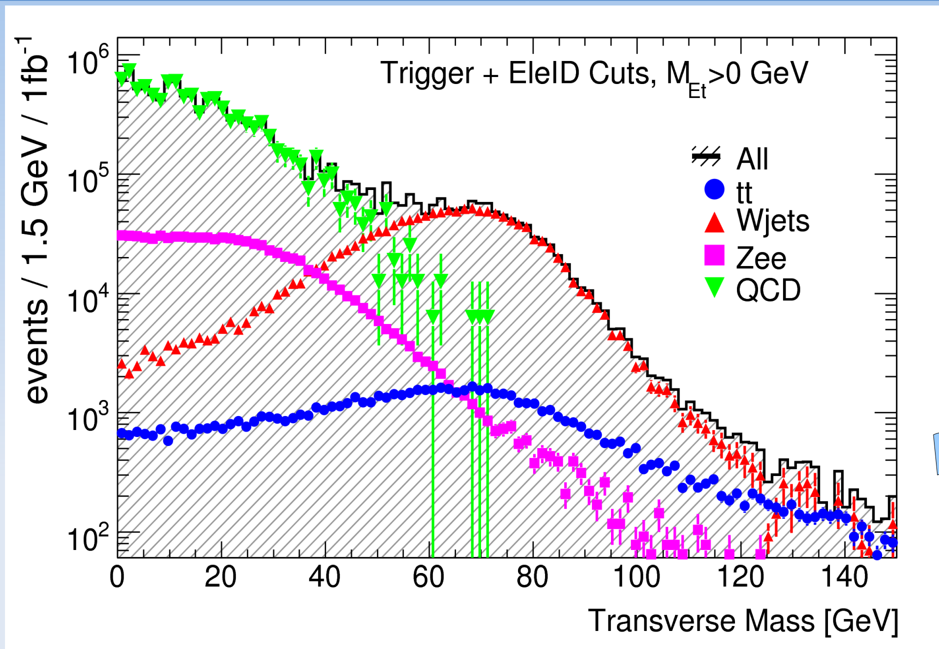
Note: samples 6101-6106 only include light and c jets

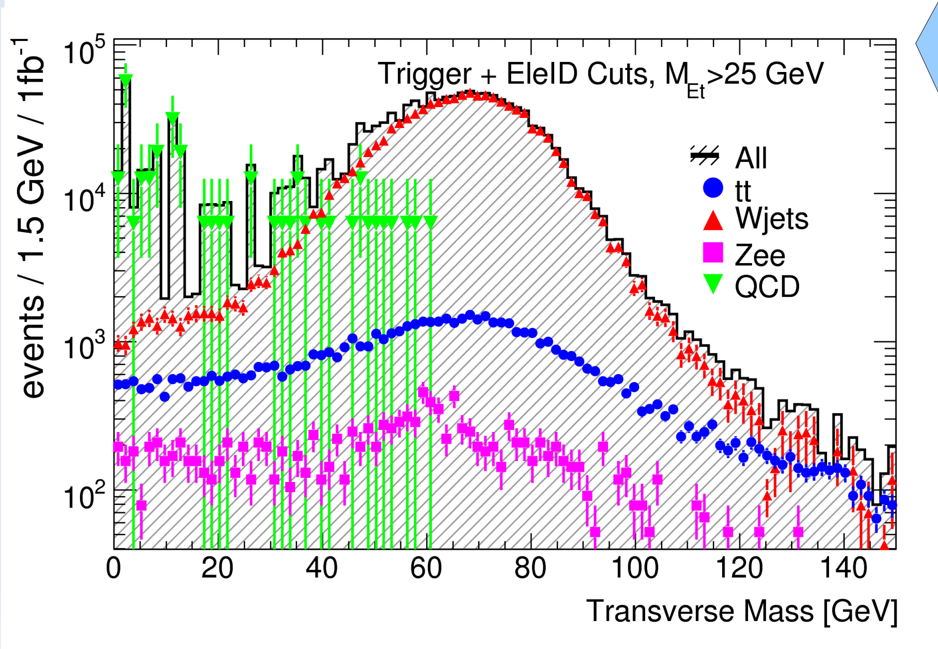
Background (Pythia)

Sample	Process
5802	JF17 (Filtered Dijets)
5568	ttbar
5144	Z $\rightarrow$ ee

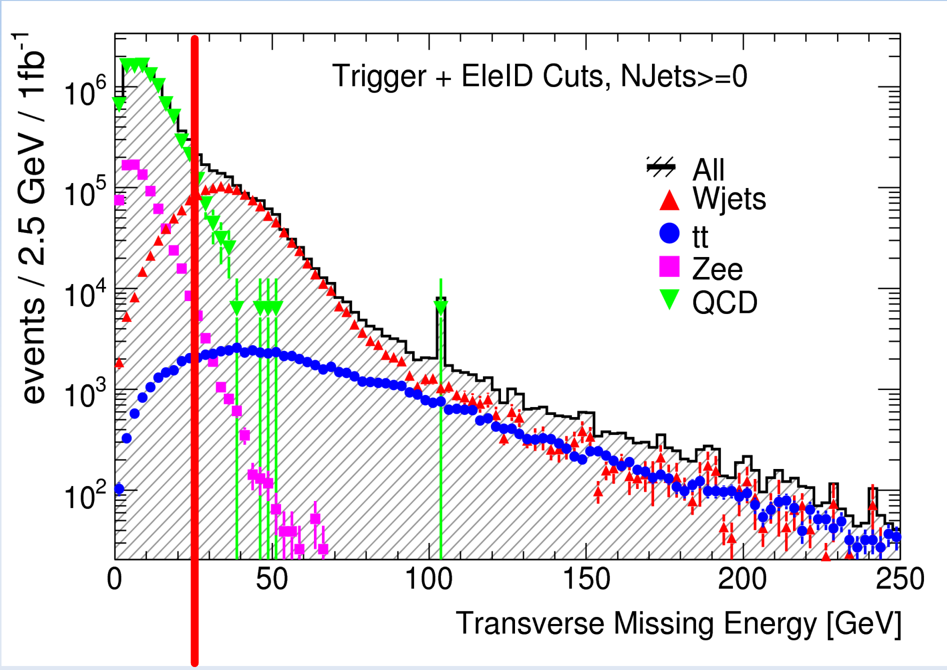
- **Standard Cuts:**
  - **Electrons:**
    - medium IsEM
    - outside cracks (1.37 < |eta| < 1.52)
    - pt > 25 GeV
    - no isolation cut yet
  - **MissingEt Cut:** 25 GeV
  - **Jets:**
    - Tower Cone 04
    - Pt > 20 GeV
    - Electron-Jet overlap Removal: DeltaR=0.4

Background & MET Cut





Apply a Missing Et Cut at **25 GeV**

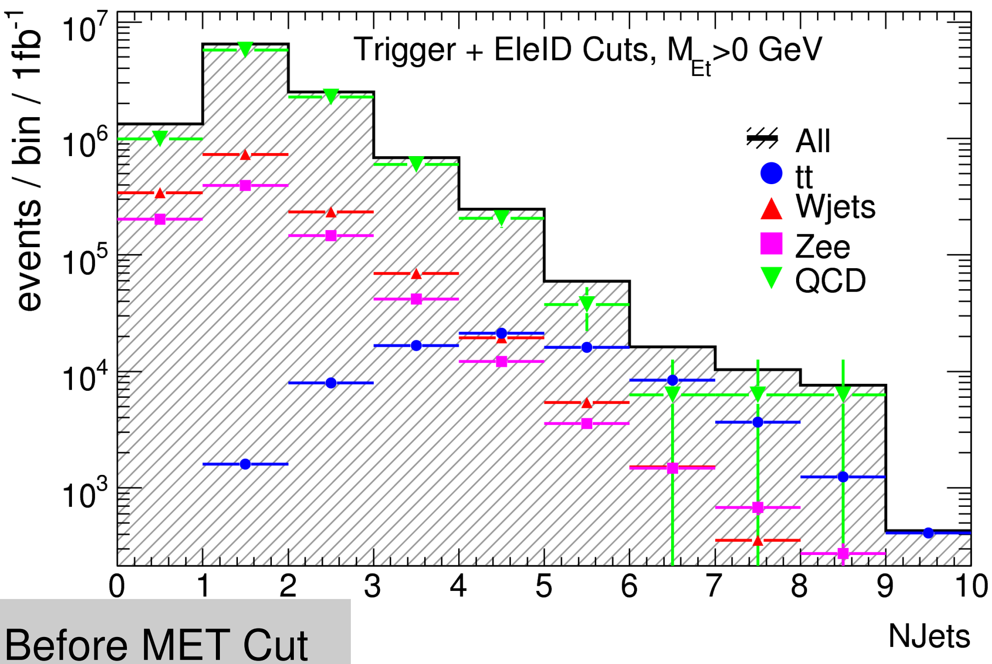


Signal events and background contamination (100* Nbkg / Nsig) with /without MET Cut:

Met Cut	Nev Wjets	Zee	QCD	ttbar
None	$(1.403\pm0.005)\cdot10^6$	57.22±0.43	702.34±20.23	5.53±0.05
25 GeV	$(1.110\pm0.004)\cdot10^6$	1.28±0.04	29.19±4.20	6.05±0.06

➔ ttbar fraction not reduced MET cut!

Background vs Jet Multiplicity



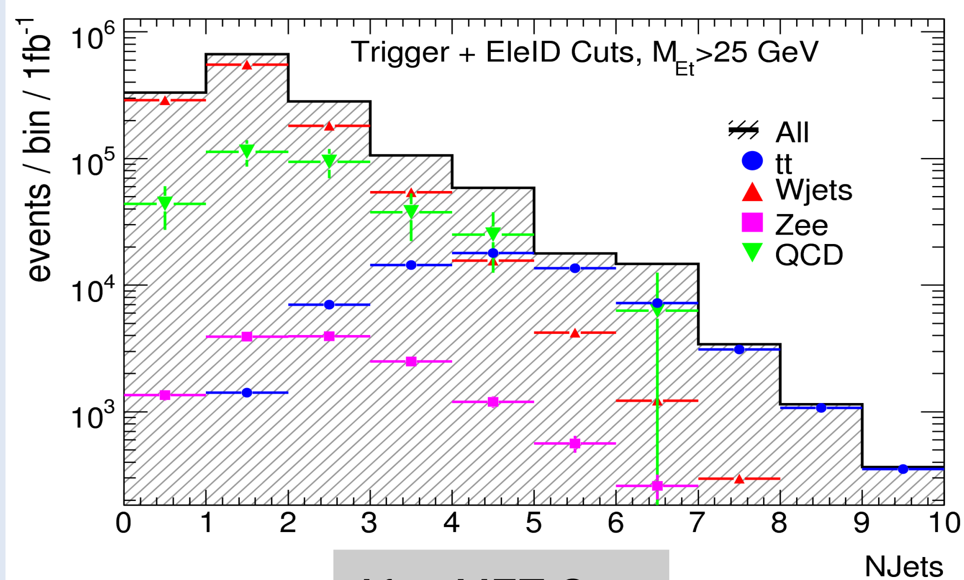
Before MET Cut

ttbar fraction becomes important with high jet multiplicity (before MET cut):

Njets≥	Nev Wjets	$100 \cdot N_{tt}/N_{Wjets}$
0	$(1.403\pm0.005)\cdot10^6$	5.53 ± 0.05
1	$(1.060\pm0.004)\cdot10^6$	7.31 ± 0.07
2	$(3.304\pm0.014)\cdot10^5$	22.97 ± 0.23
3	$(9.604\pm0.057)\cdot10^4$	70.75 ± 0.85
4	$(2.689\pm0.024)\cdot10^4$	190.81 ± 3.08
5	$(7.384\pm0.124)\cdot10^3$	406.90 ± 10.54

From now on ignore QCD and Zee:
already subtracted (see Alessandro Tricoli's talk)
and focus on **data driven methods for ttbar background subtraction**.

NOTE: QCD background already subtracted implies a cut:
MET > 22.5 GeV (because of the photon sample normalization range, refer to Alessandro's talk)



After MET Cut

Data Driven Ttbar Subtraction

Similar method to QCD:

Do not apply the MET cut (with the exception of the 22.5GeV cut after QCD subtraction), but use the information on the MET shape of the signal vs MET shape of the background to estimate tt background.

METHOD:

- Apply All cuts BUT Missing Et (**Full Sample, S0**)
- Select a High Purity **Ttbar** Sample (**ttbar Control Sample, tCS**)
- Select a High Purity **W+jets** Sample (**Signal Control Sample, wCS**)

  (alternatively extract W+jets MET shape from Monte Carlo)

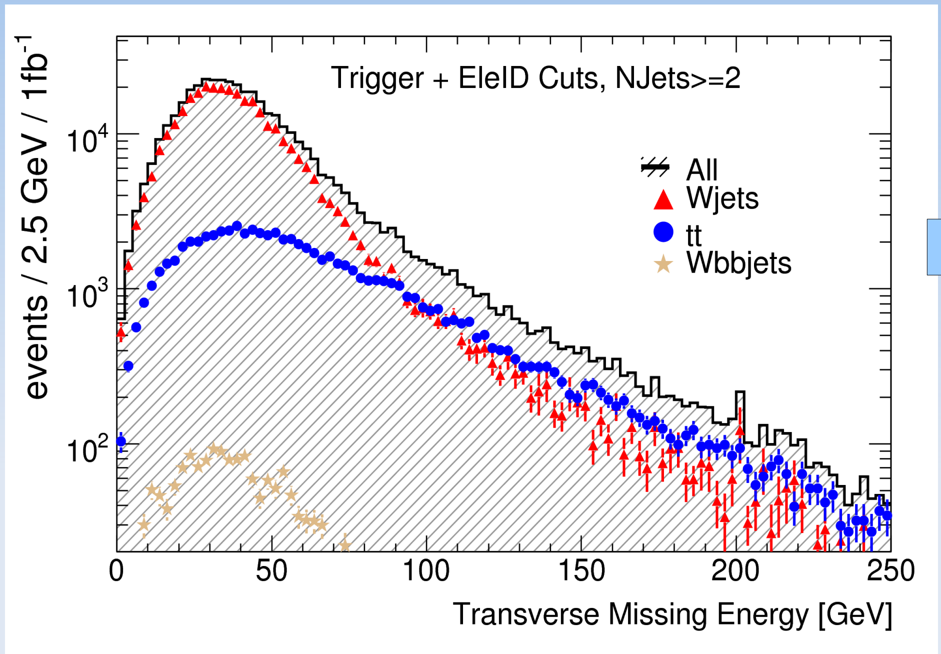
1) From tCS: Fit MET spectrum -> extract shape of ttbar

2) From wCS: Fit MET spectrum -> extract shape of signal (W+jets)

3) From S0: Simultaneous fit Signal (Wjets) + Bkg (ttbar) -> Find Normalization

Apply for **each Multiplicity Bin** (to calculate Cross-Section as a function of Jet Multiplicity)

Ttbar Control Sample Selection



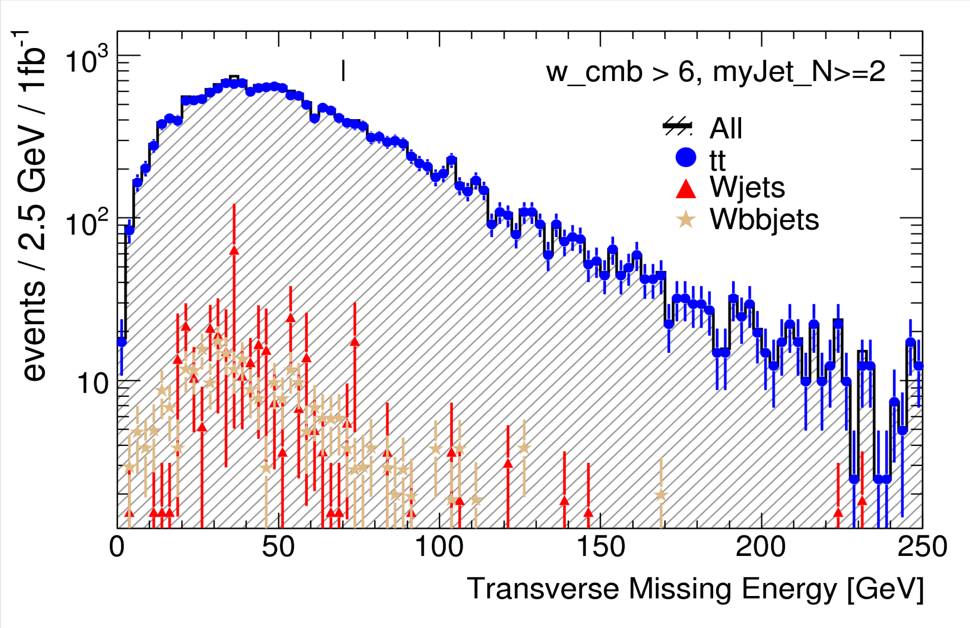
How to select the control sample?

- high rejection power against signal (W+jets)
- should not bias MET distribution

We use **B-tagging:**

- cut: Njets ≥ 2, b-tagging weight > xx

Control Sample



W+jets Contamination to the Control Sample

Njets ≥ 2	Ttbar (%)	Wjets(%)	Wbbjets(%)
w_cmb >5	93.1	5.7	1.2
w_cmb >6	96.0	2.7	1.3

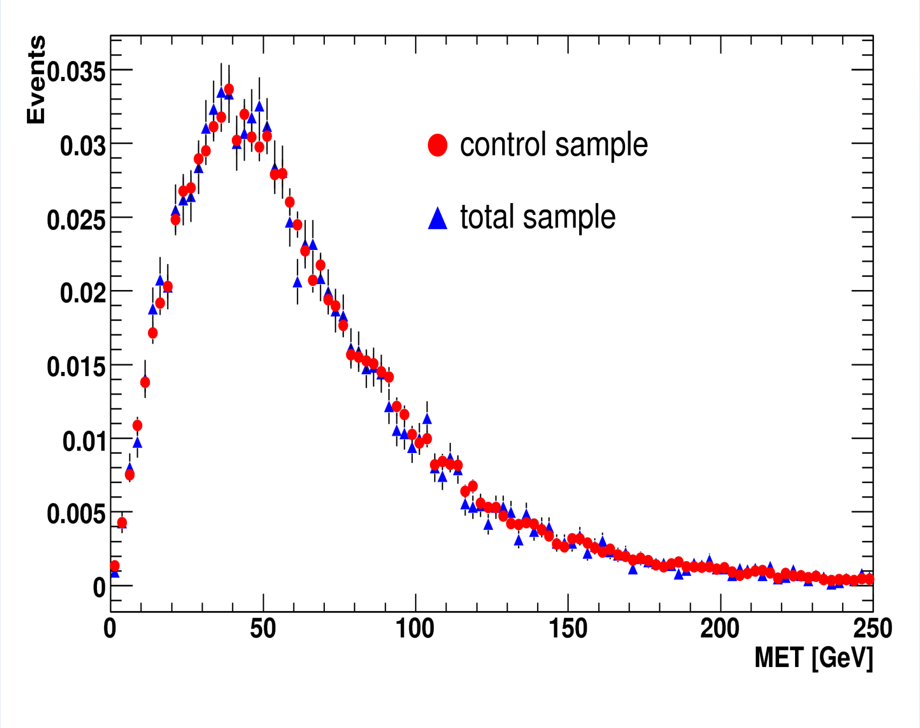
Njets ≥ 5			
w_cmb >5	98.0	1.6	0.4
w_cmb >6	98.8	1.1	0.6

Control Sample: Distortion in MET

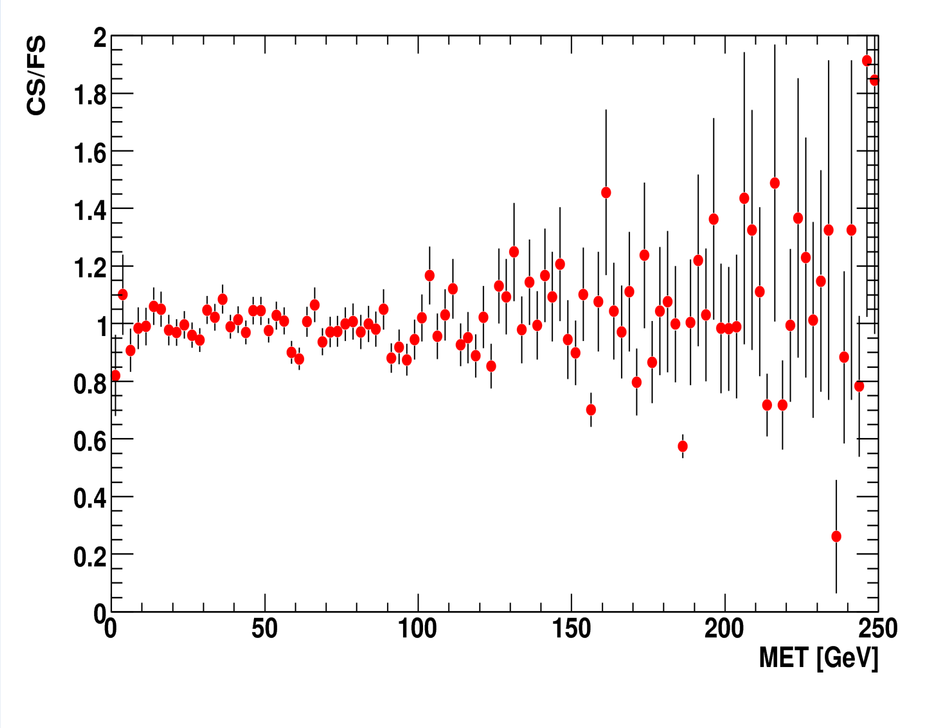
Check that B-tagging cut does not distort the MET shape:

Control Sample: w_cmb > 6, Njets ≥ 2

Total Sample: no btag (all cuts but MET) , Njets ≥ 0



Comparing shape (normalized to 1) between control sample and full sample



Ratio Control Sample/ Total Sample after normalization

Signal Control Sample

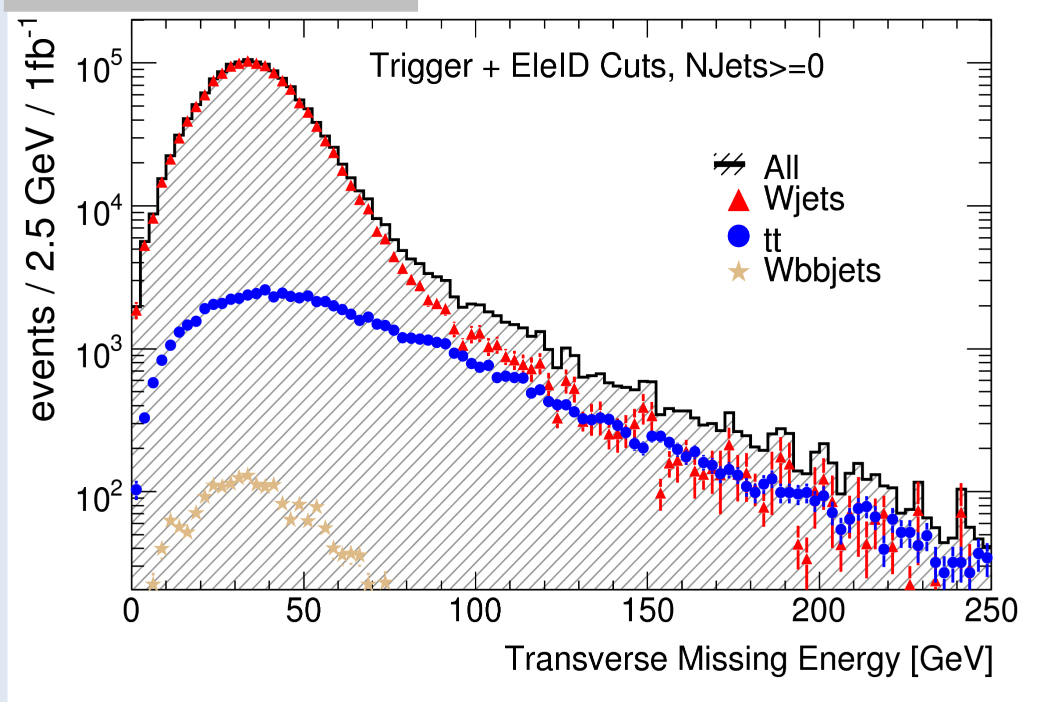
- Use an **Anti B-tagging Cut** to select a pure W+jets control sample:
  - Reject any events with: Njets ≥0, b-tagging weight > xx

B-Jet Veto	**Nev Wjets**	**$100 \cdot N_{tt}/N_{Wjets}$**
None	$(1.403 \pm 0.005) \cdot 10^6$	5.53±0.05
$w_{cmb}>4$	$(1.358 \pm 0.005) \cdot 10^6$	1.10±0.02

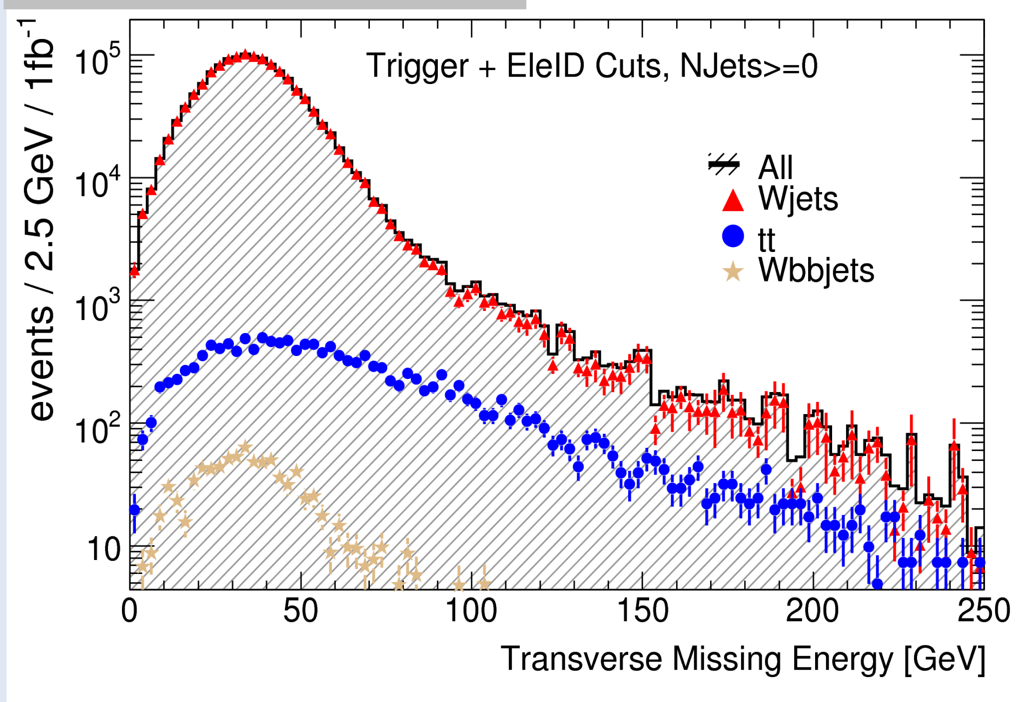
For the inclusive sample ($N_{jets} \geq 0$)

→ Purity could be a problem, especially at high jet multiplicity....

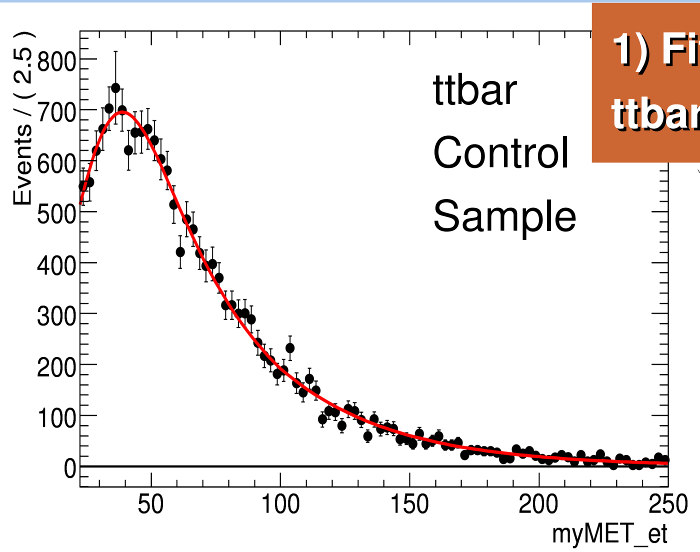
Before B-Jet Veto



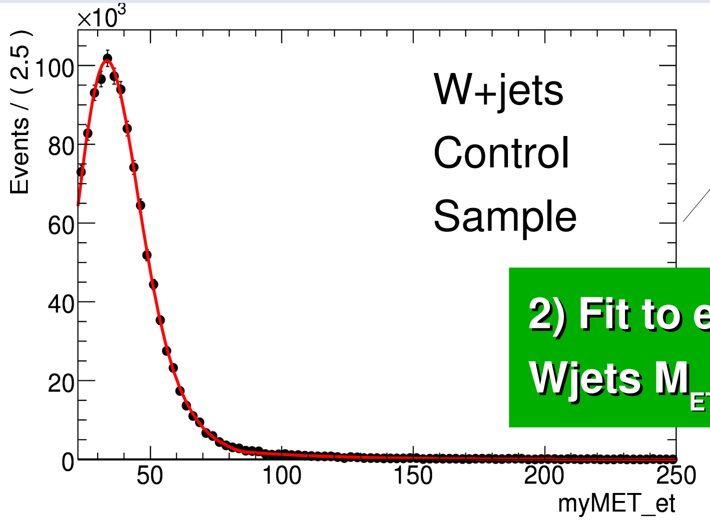
After B-Jet Veto ($w_{cmb}>4$)



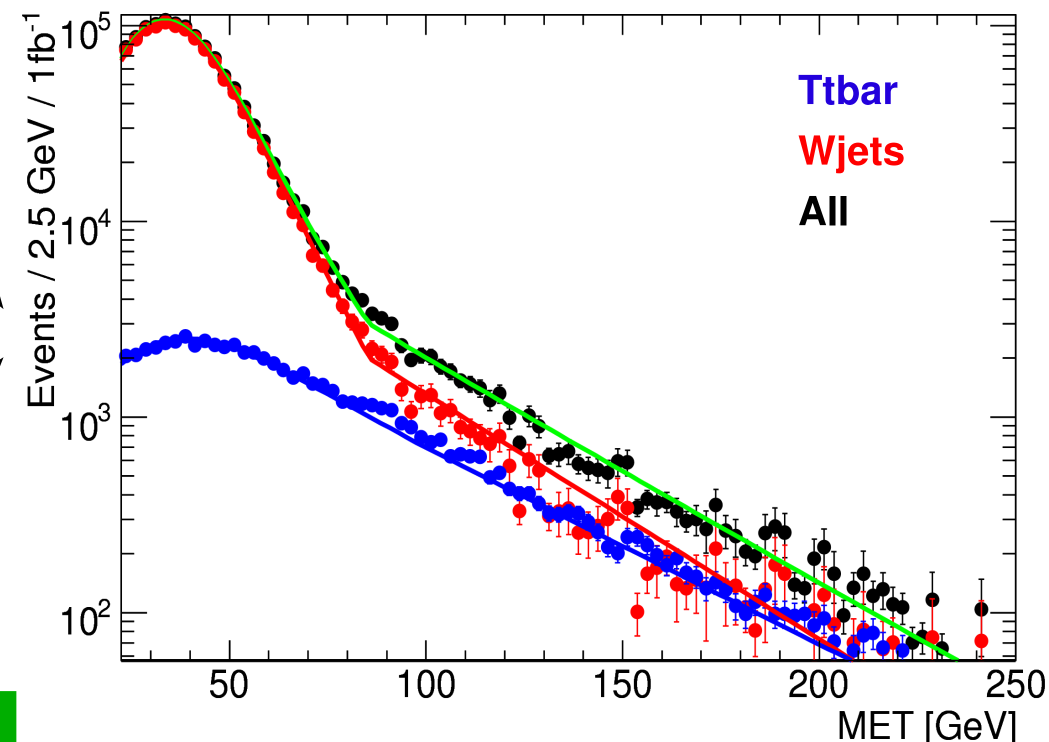
Fitting & Normalizing



1) Fit to extract the ttbar M_{ET} shape



2) Fit to extract the Wjets M_{ET} shape



3) Fix the shape of Wjets and ttbar: do a combined fit of Signal+Background to extract the relative normalization

Results

Number of predicted ttbar events in the MET range 22.5-250 GeV

Computed in 3 steps:

1) **Prediction 1**:

   Assume a pure ttbar Control Sample (no W+jets contamination) and take the W+jets shape from MC

2) **Prediction 2**:

   Include W+jets and Wbb+Jets contamination in the ttbar Control Sample

3) **Prediction 3**:

   All of above plus extract W+jets shape from the Signal Control Sample (including contamination of the ttbar)

Jet Multi.	Ntt (true)	Pred1 Error (%)	Pred2 Error (%)	Pred3 Error (%)
Njets ≥0	$(6.848 \pm 0.026) \cdot 10^4$	4.06	6.56	4.93
Njets ≥1	$(6.837 \pm 0.026) \cdot 10^4$	2.94	5.75	8.33
Njets ≥2	$(6.692 \pm 0.026) \cdot 10^4$	3.11	6.16	10.71
Njets ≥3	$(5.976 \pm 0.024) \cdot 10^4$	4.55	6.18	11.92
Njets ≥4	$(4.489 \pm 0.021) \cdot 10^4$	5.47	6.79	12.49

Error on prediction:
\# Predicted Ev - True Ev
divided by # True Ev

B-Tag cut for ttbar CS: $w_{cmb} > 6$

Bjet Veto cut for signal CS: $w_{cmb} > 4$

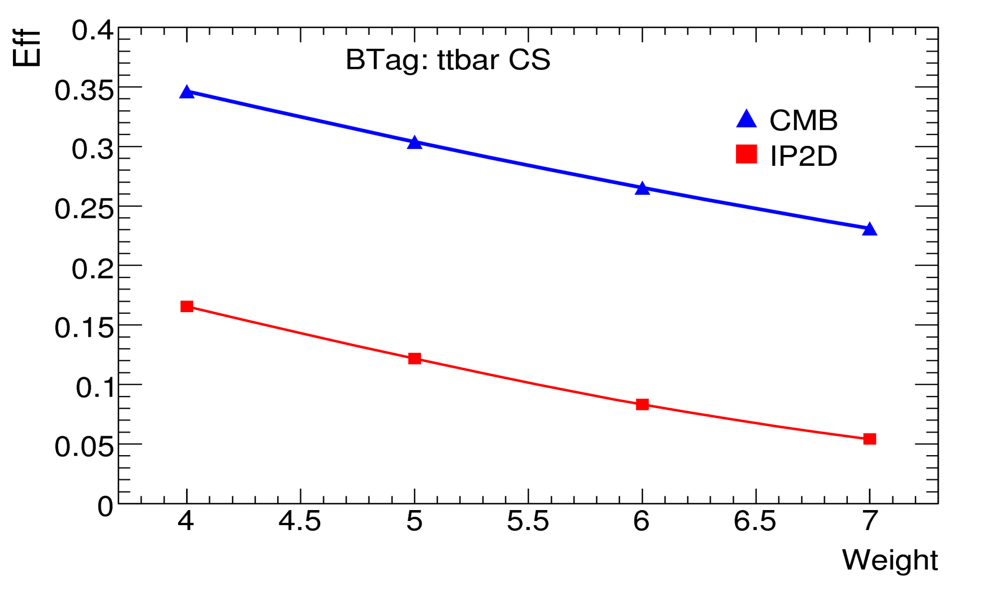
Effect of tt Control Sample Contamination

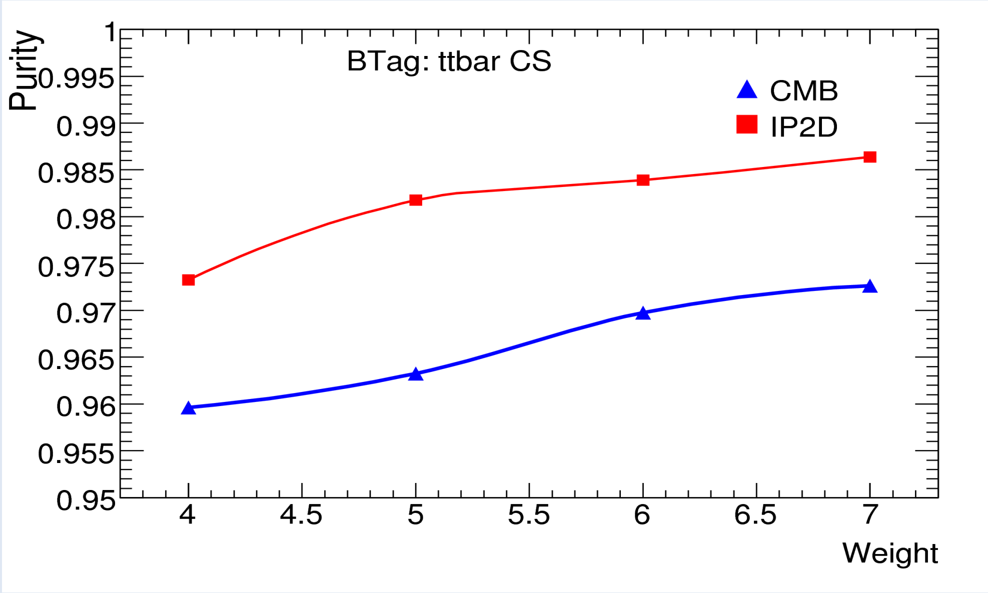
Effect of Signal Control Sample Contamination

B-Tagging studies for Control Sample Selection

- The Selection of the Control Sample is **crucial** for this study! Important to study how we can improve
- There are several **B-Tagging Algorithms** available, I have selected the 2 most used:
  - **IP2D**: uses only the transverse impact parameter
  - **CMB** (combined): a combination of transverse and impact parameter and secondary vertex information
- **Advantages/Disadvantages**:
  - **CMB**: better performance (preferred one if well understood)
  - **IP2D**: more robust, considered more reliable with real data
- Which one to use? Studies on their performance for the ttbar/signal control sample selection on-going... (see next slide)

B-Tagging Algorithms Performance



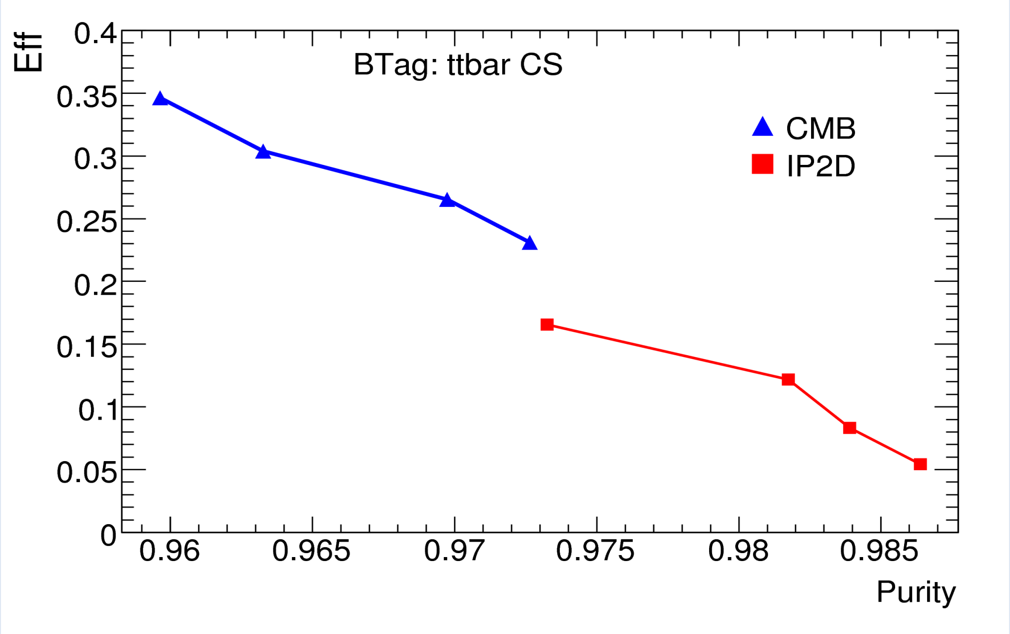


Look at the **efficiency** and **purity** of the ttbar Control Sample using **IP2D** and **CMB** and different cuts on the b-tagging weight:

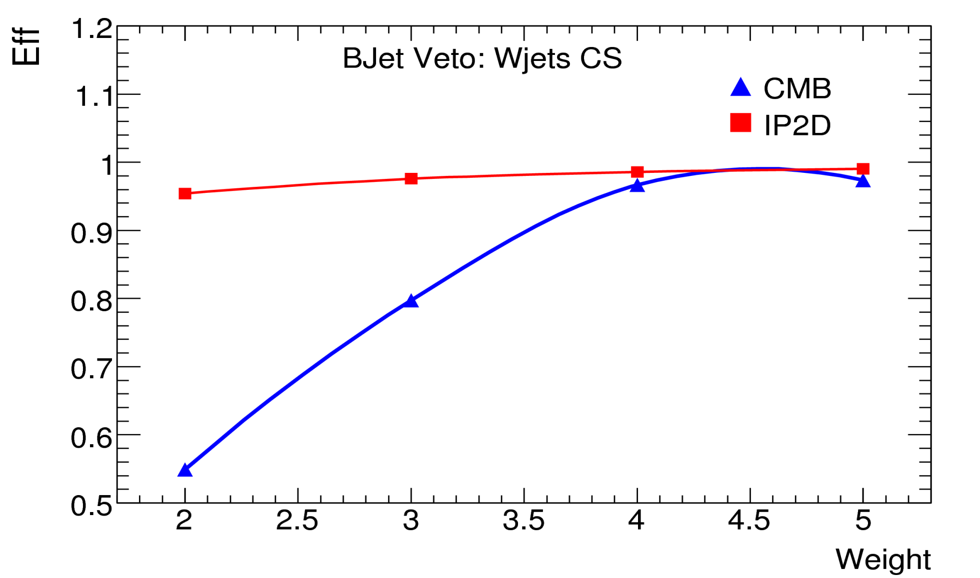
Njets ≥ 2, w > x

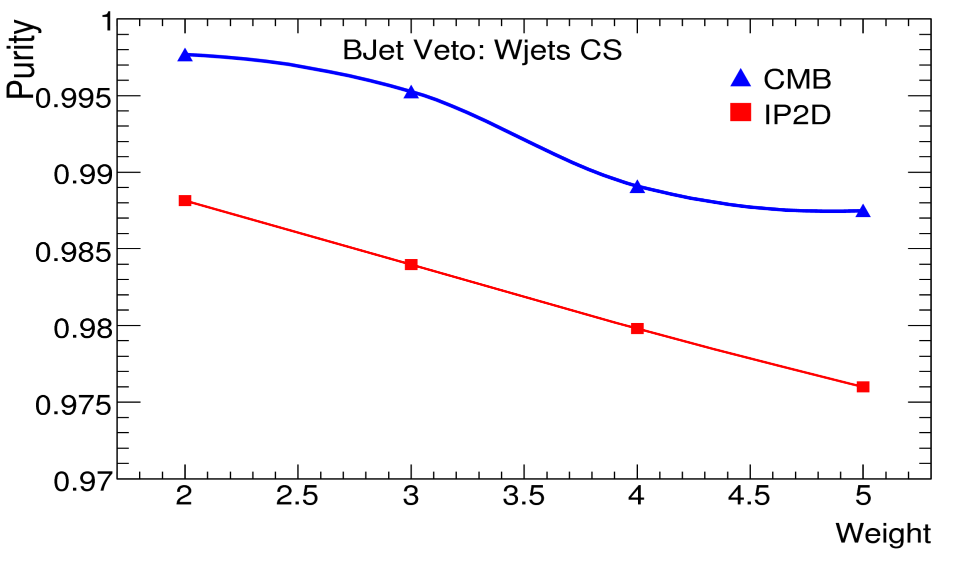
- **CMB** is more efficient (good at high NJets)
- **IP2D** give higher purity (important!)

However we are looking at two different regions for CMB and IP2D -> it is not a proper comparison!



B-Jet Veto Studies



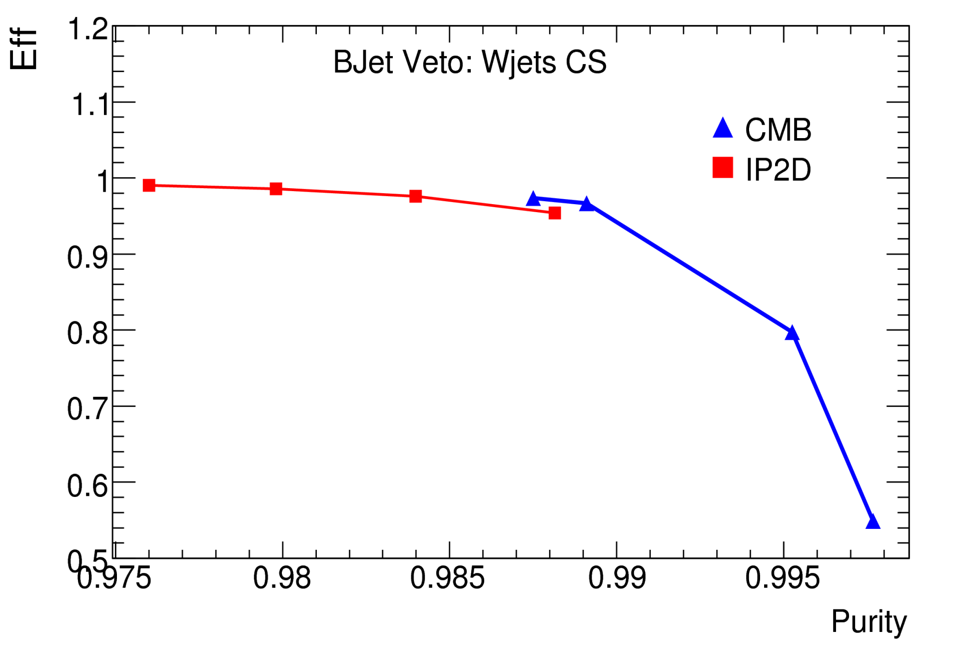


Compare **efficiency** and **purity** of the Signal Control Sample for **IP2D** and **CMB** and different cuts on the b-tagging weight:

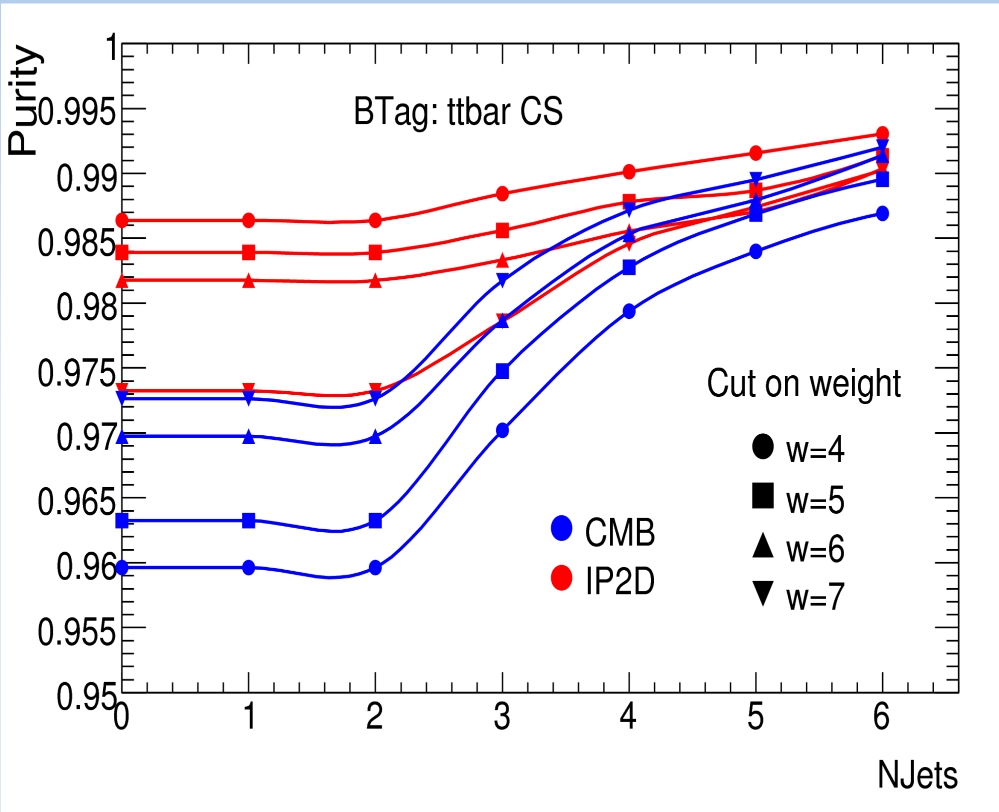
- **IP2D** is more efficient
- **CMB** give higher purity

Reject Events with:
Njets ≥ 0, w > x

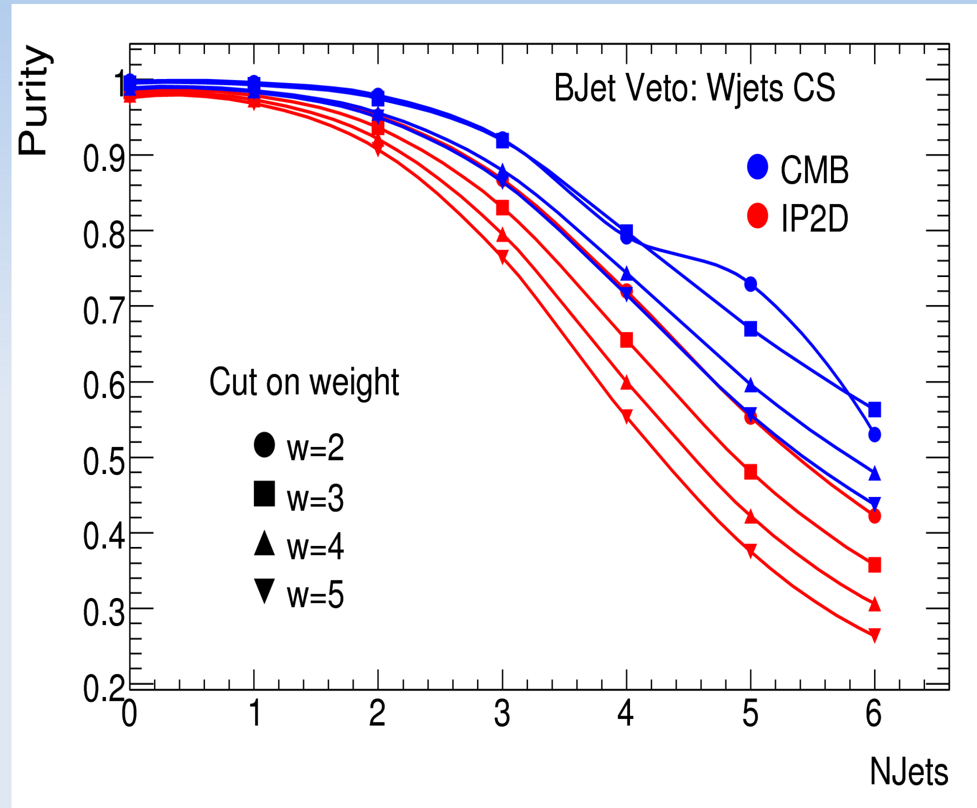
Opposite than before -> we are applying a b-jet veto!



Control Samples: Purity vs NJets



Ttbar Control Sample:
purity is above 95% even at low multiplicity
(where ttbar fraction is small)



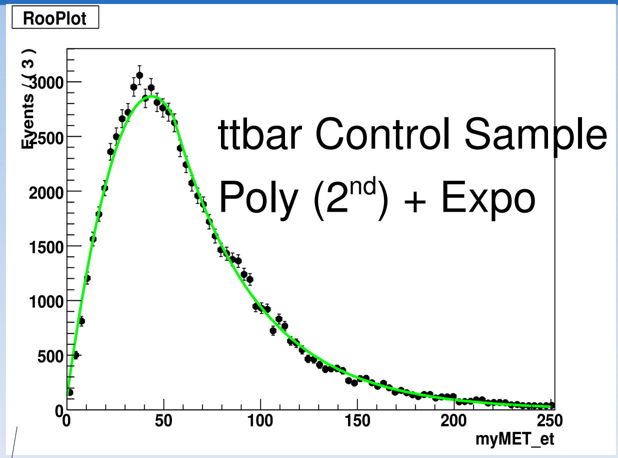
Signal Control Sample:
very low purity at high multiplicity

Fitting MET shape

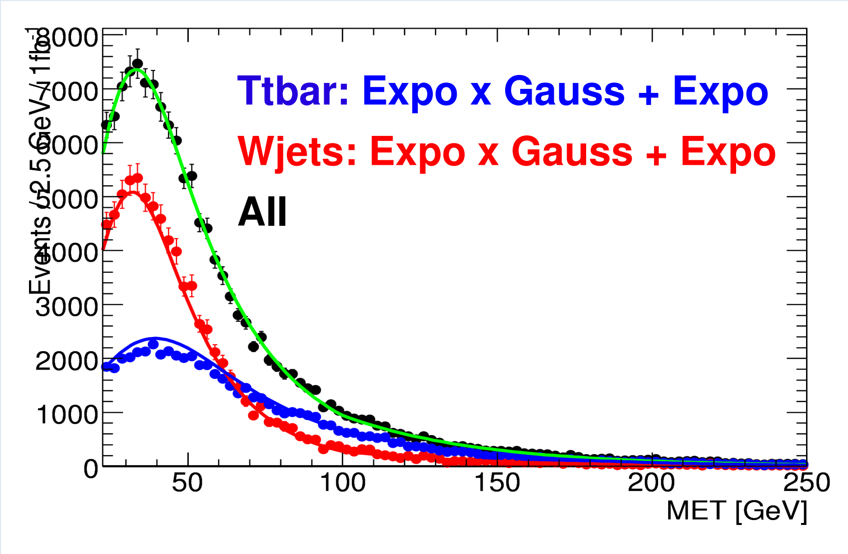
Parametrization for MET spectra not easy.

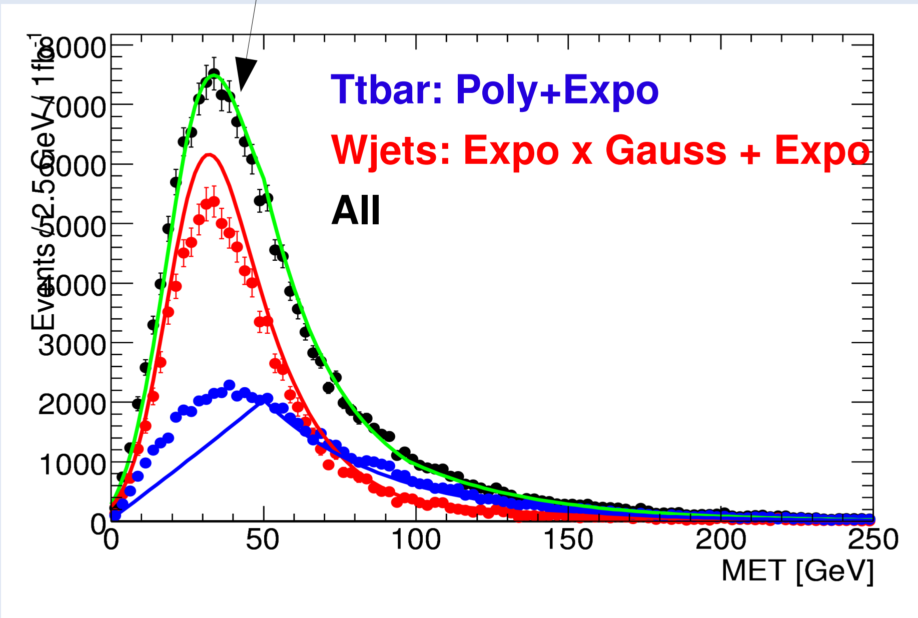
Investigated the following parametrization (using RooFit):

- **Polynomial * Exponential** (like QCD) :
  does not describe the peak
- **Polynomial (0-50 GeV) + Expo (>50GeV)**:
  works but very unstable (problem: glue two PDFs)
- **Gauss convoluted with Expo (0-70/80 GeV)**
  + Expo (> 70/80 GeV) - boundary free parameter:
  best description I could find



Looks ok but for some multiplicity bins it fails: continuity of function

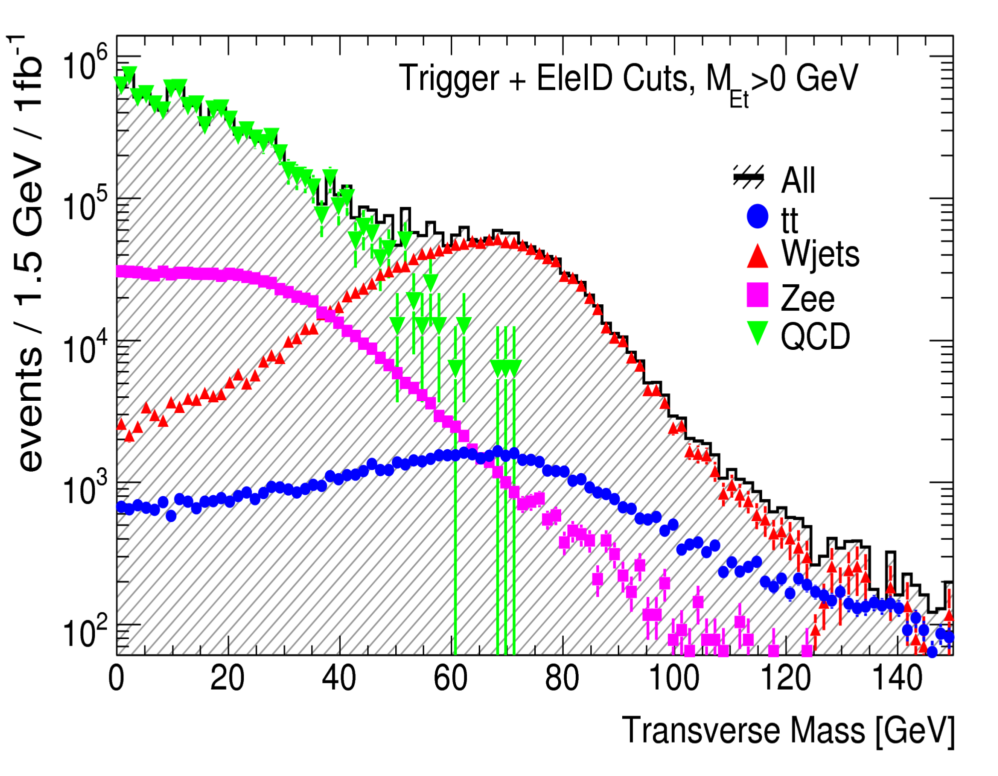




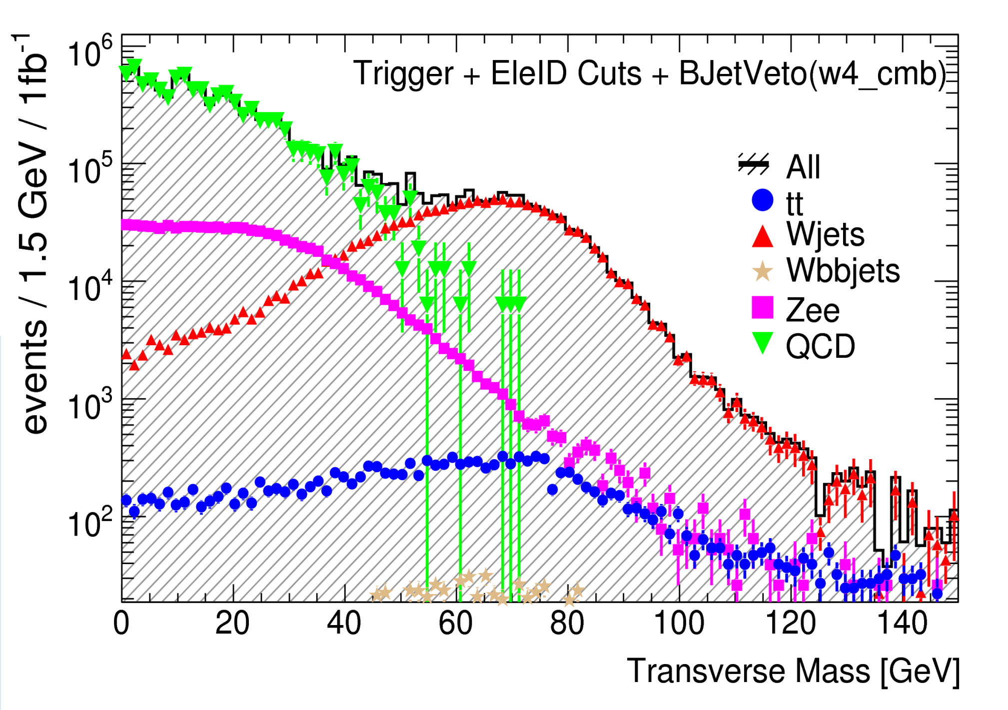
Conclusions and Outlook

- Described a data-driven method for ttbar background subtraction in multiplicity bins:
  - it works -> relative error on number of ttbar events is ~**5-7%** (if we extract the Signal Shape from MC), can go up to 11-12% using a Signal Control Sample
- Having shown the method is applicable, we need to **optimize** it:
  - continue studies on B-tagging, B-jet veto to find the optimal cut for the Control Sample selection (optimize per multiplicity bin)
  - B-jet veto for Signal Control Sample: purity is an issue, need to investigate more cuts
  - could improve the fitting procedure
- Still a lot of work, but it looks promising!

EXTRAS

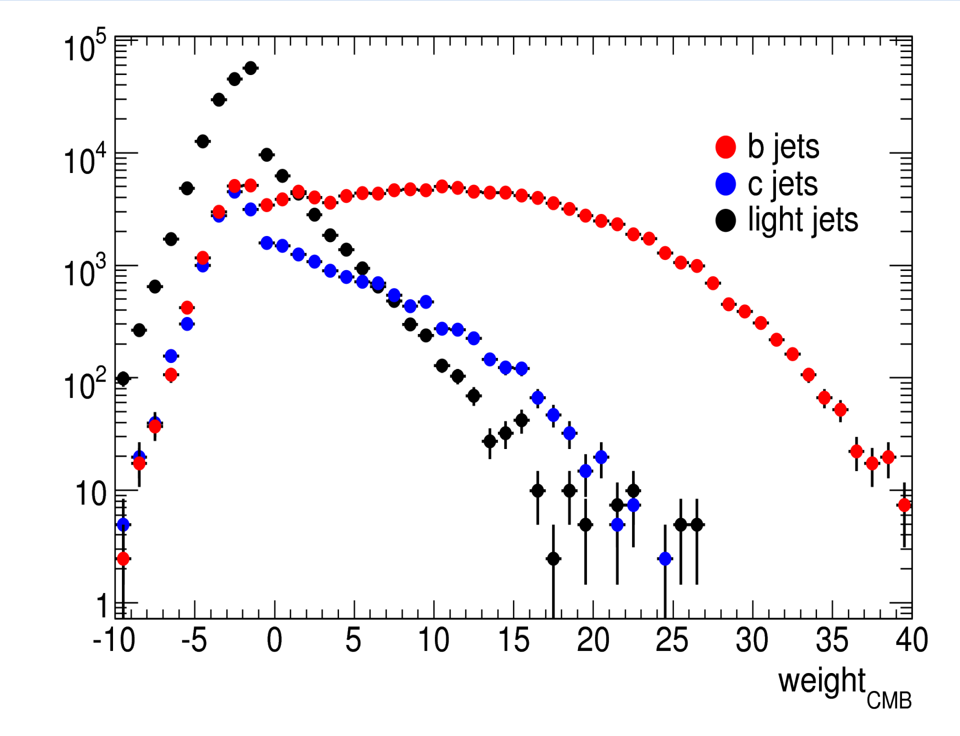


After Jet Veto

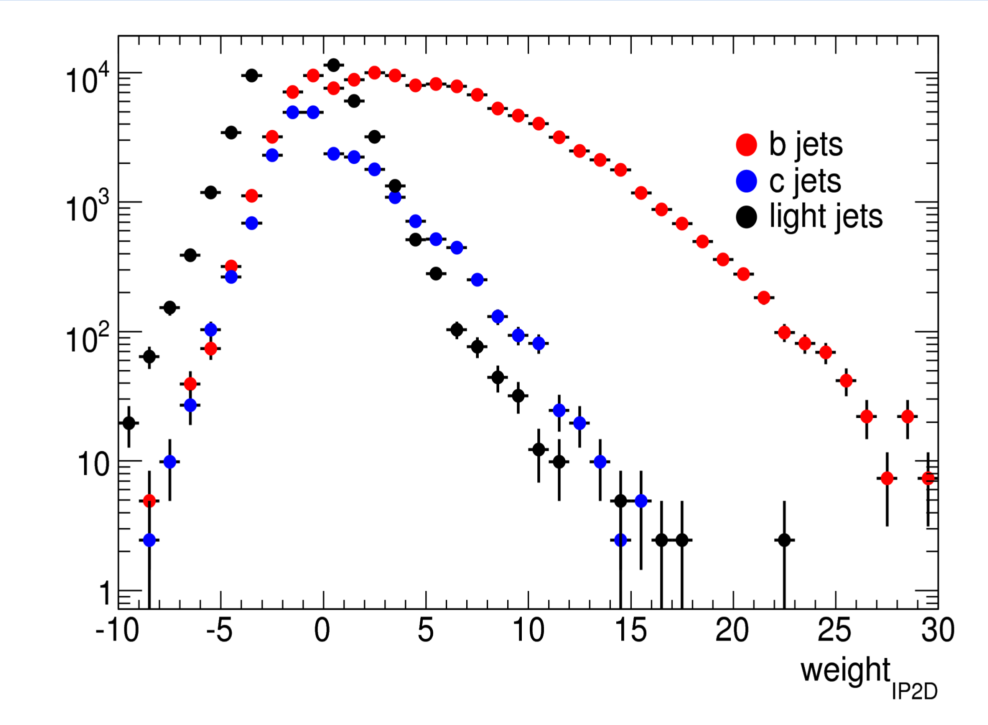


Before Jet Veto

B-Tagging Algorithms



Combined Algorithm



IP2D