# BACKGROUNDS AND UNCERTAINTIES FOR TTH(H—>BB)

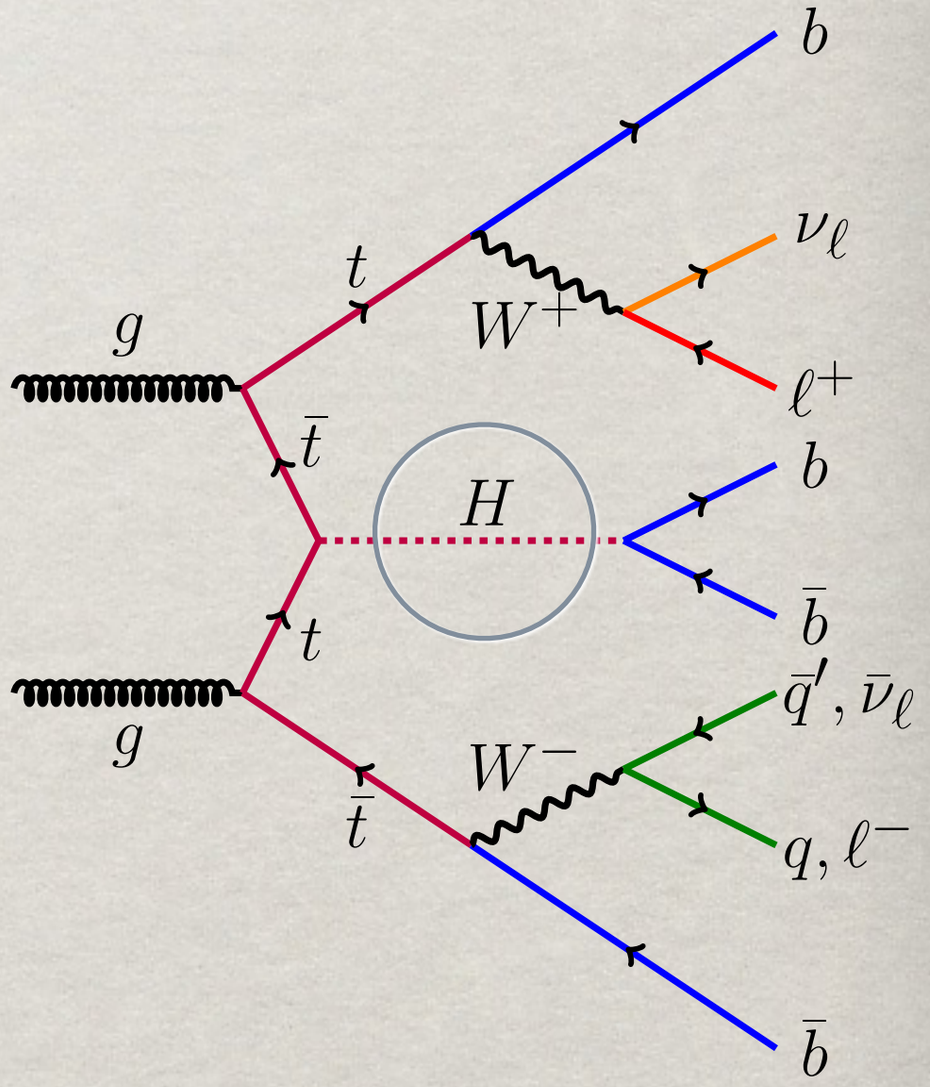

WUMING LUO

ON BEHALF OF TTH@CMS



# BACKGROUNDS FOR TTH

- TTJets is the main background:
- Other relevant bkg MC
  - ZJets, ttbar+Z/W, WJets, WW, WZ, ZZ, single top
- Need accurate background modeling:
  - Small Signal(0.132 pb @125GeV), large Bkg(25.81 pb)
  - Bkg very similar to Signal
    - irreducible: tt+bb



# INCLUSIVE TTJETS SAMPLES @8TEV

- Dedicated ttjets samples for different tt decay mode
- All these samples are generated with MADGRAPH +PYTHIA
  - Inclusive: up to tt+3 extra patrons (@Tree Level)
  - Statistics(~10M) is OK, some limitations

| | XS(pb) | nGen |
|---|---|---|
| tt—> jets | 112.33 | 31111456 |
| tt—> lv + 4jets | 107.66 | 25327478 |
| tt—> lvlv+ 2jets | 25.81 | 12100452 |

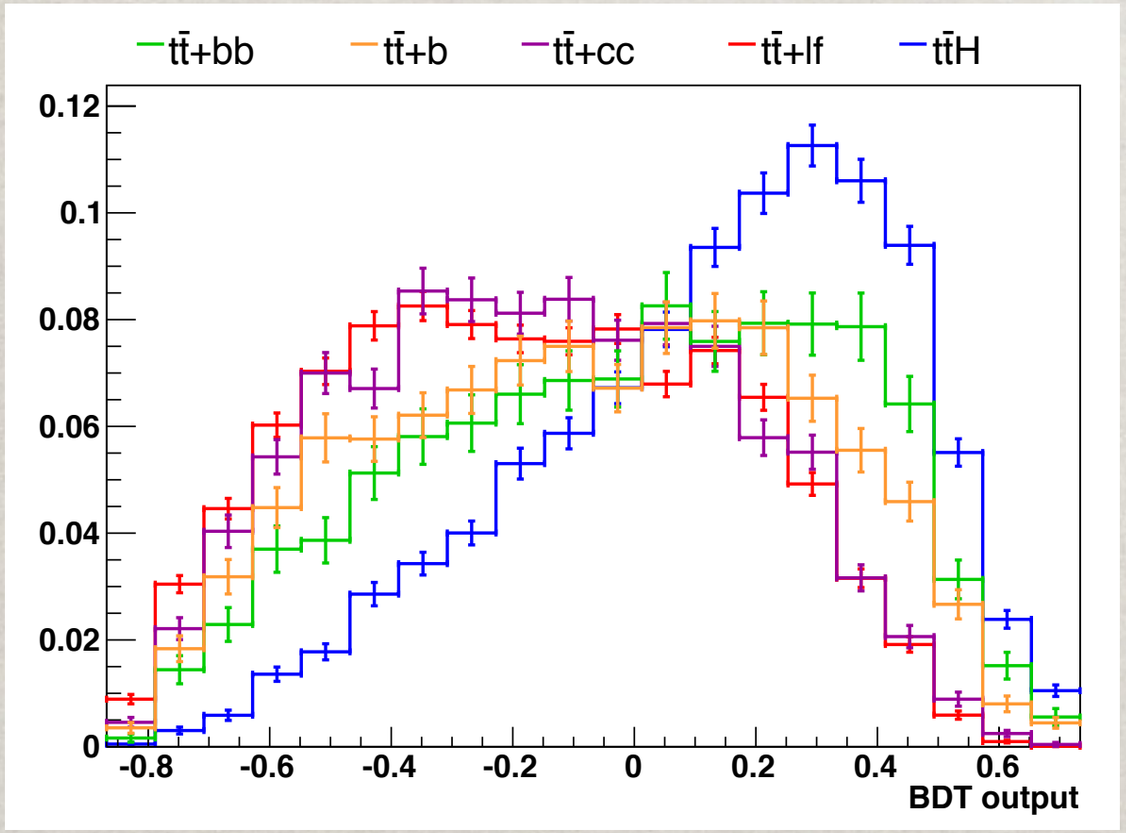
# EXCLUSIVE TTJETS SAMPLES

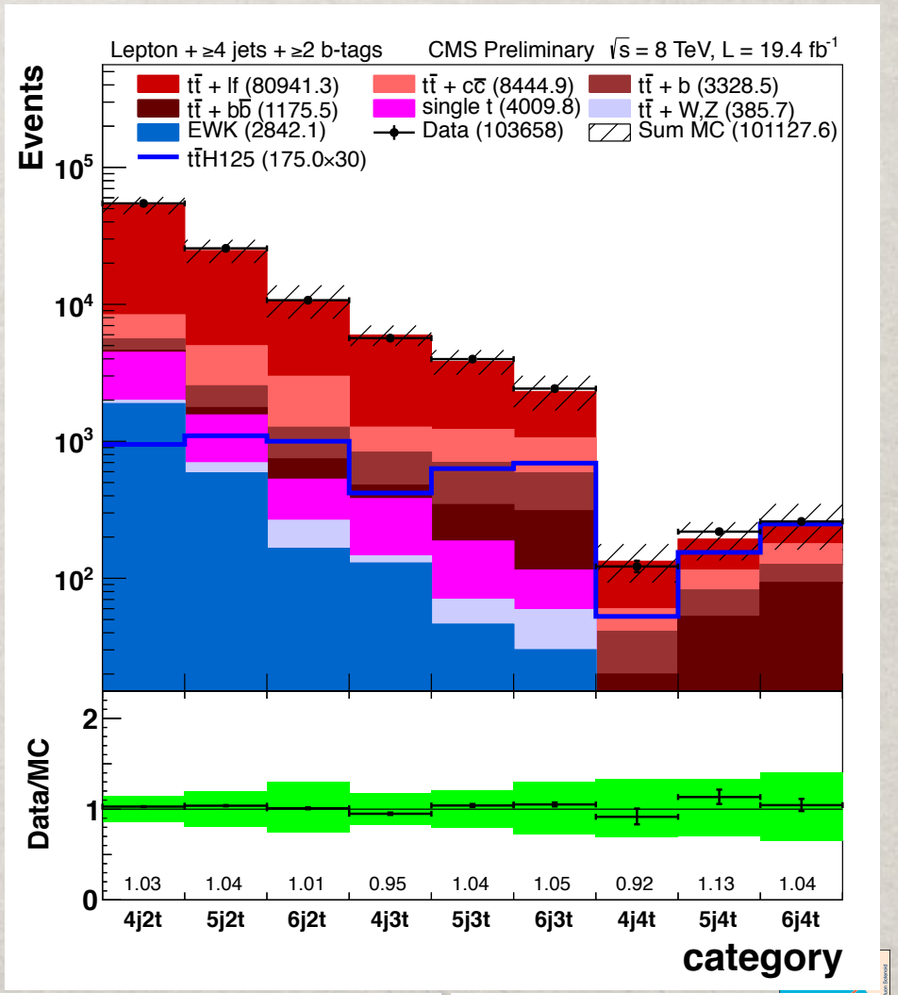
- For ttH (H—>bb), tt+HF(heavy flavor) is more similar to Signal than tt+LF(light flavor)
- Different uncertainty on the production of additional LF jets compared to HF jets(e.g. $Q^2$ scale)
- Split the inclusive ttjets sample, based on the quark flavor associated with the reco jets in the event
  - tt+bb: ≥2 reco jets matched to "extra" b-quarks(non top)
  - tt+b: only one match (soft and collinear)
  - tt+cc: ≥1 reco jets matched to c-quarks
  - tt+LF: the rest

# TT+X COMPARISON

- Different BDT shapes for the different tt+X samples
  - tt+bb most similar to ttH, tt+LF least similar
- Fraction for different tt+X samples change from category to category

LJ+≥6jets+3tags



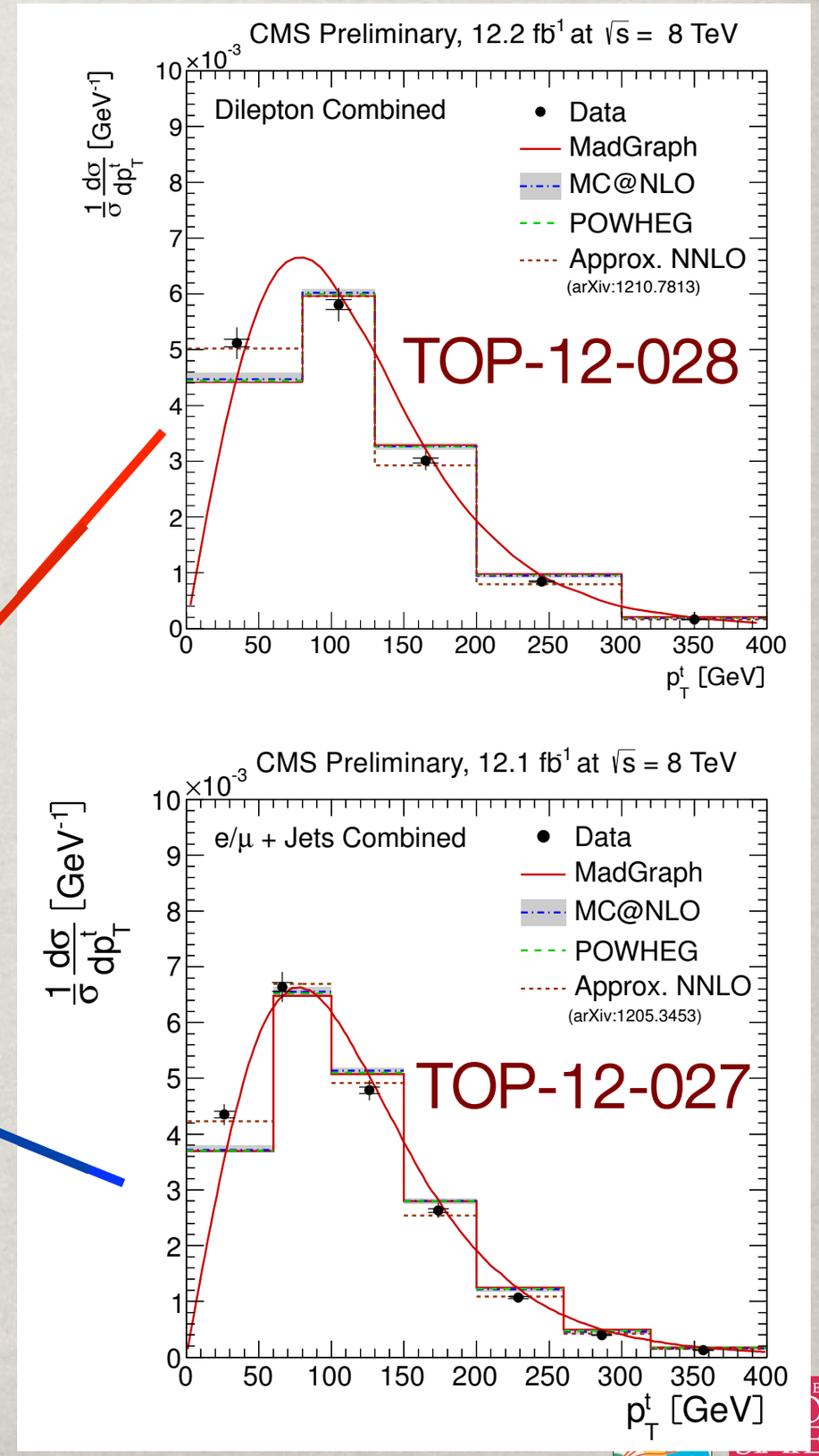



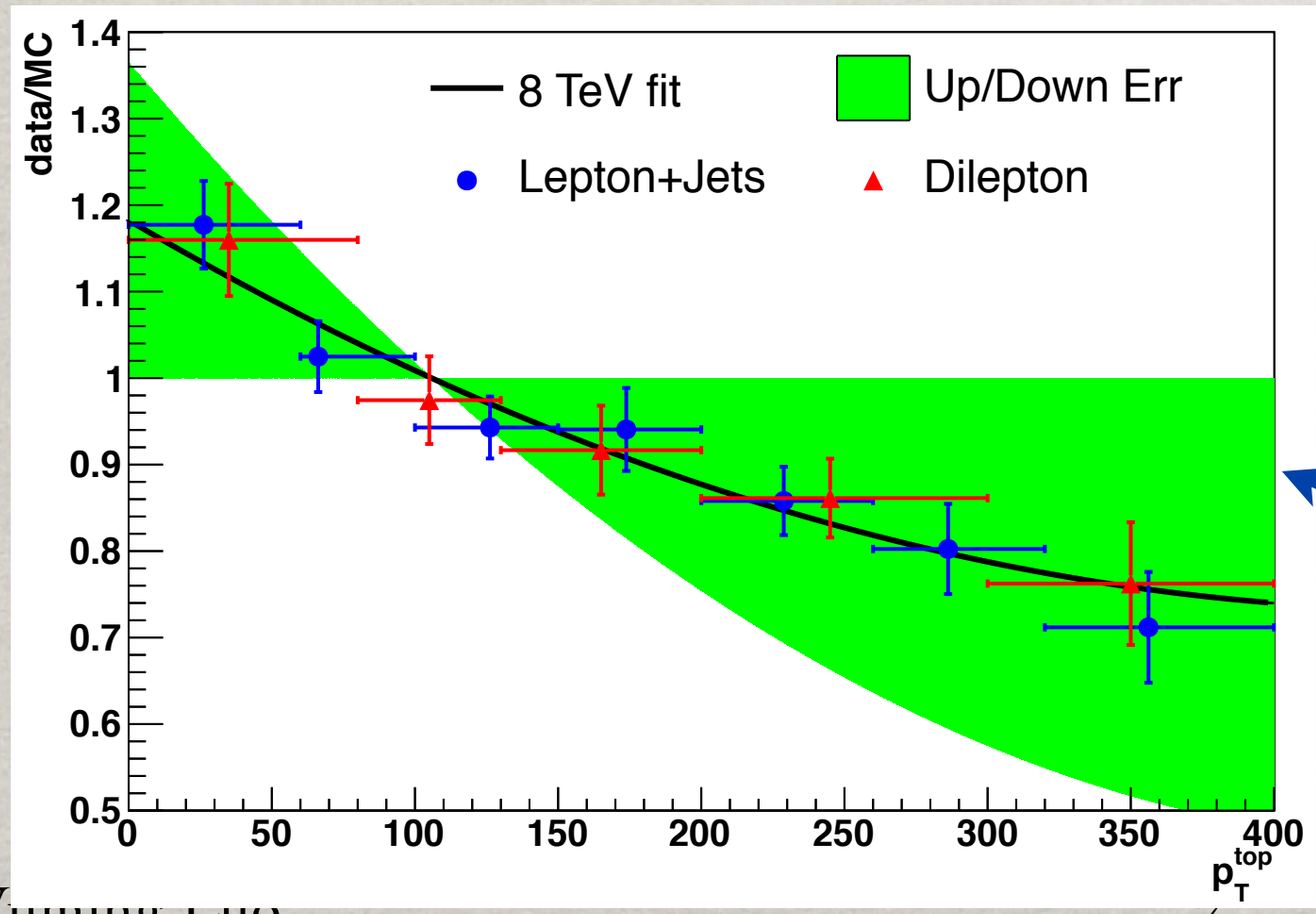
# CORRECTIONS TO MC@8TEV

- PileUp
- Lepton Scale Factor, trigger Scale Factor
- Jet Energy Scale and Jet Energy Resolution
- B-tagging Scale Factor
  - Correct MC jet b-tag discriminator shape to match Data
    - for both heavy flavor(HF) and light flavor(LF) jets
- top Pt reweighting*
  - Data/MC discrepancy for jets and lepton Pt spectrum
  - Due to mis-modeling in the top Pt spectrum

# TOP PT REWEIGHTING

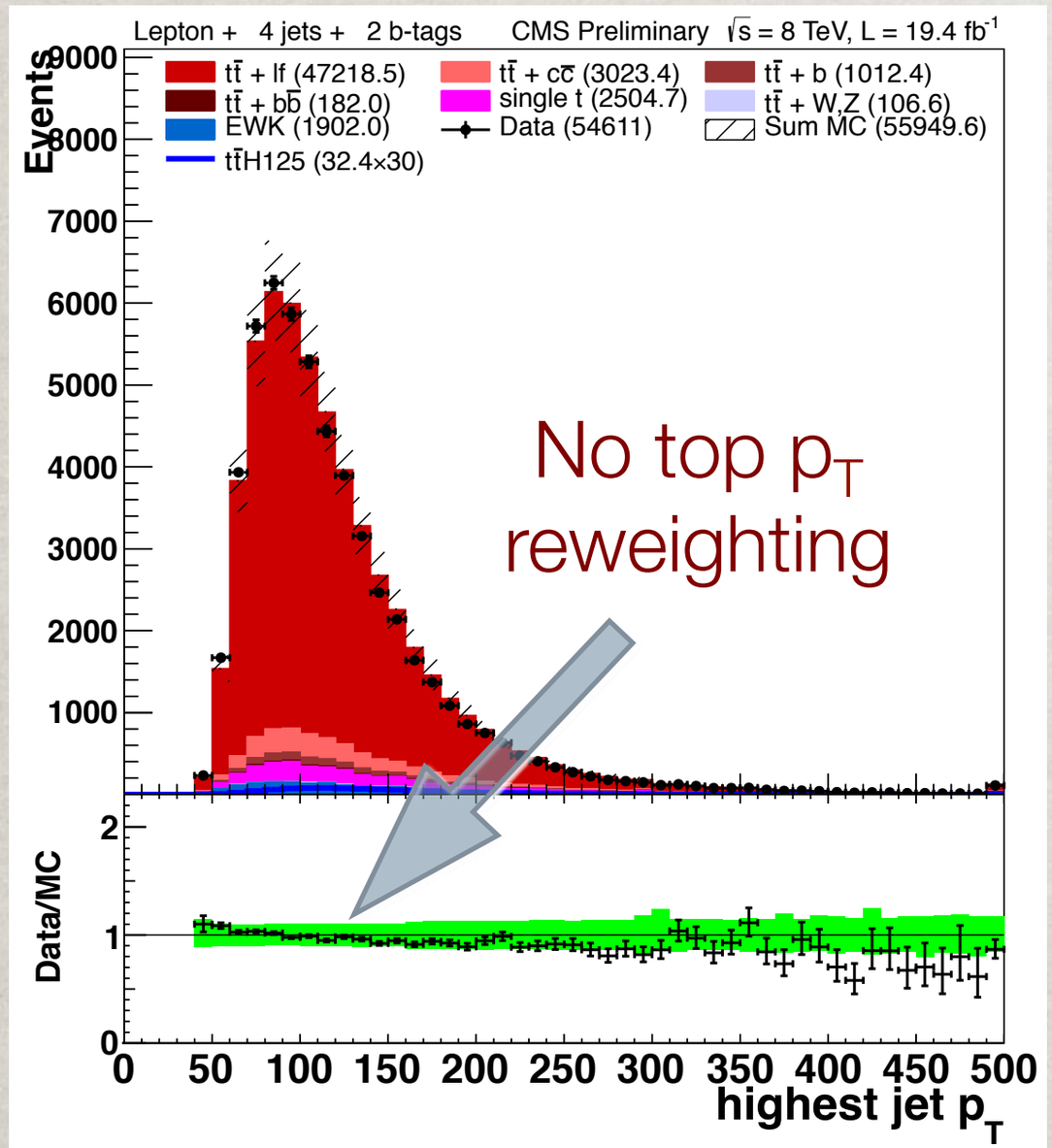
- Top $P_T$ spectrum softer in Data than MC (e.g. Madgraph)
- Top differential cross section measurements provide data/MC SFs
- Uncertainty: no correction (-1σ) and doubling correction (+1σ)

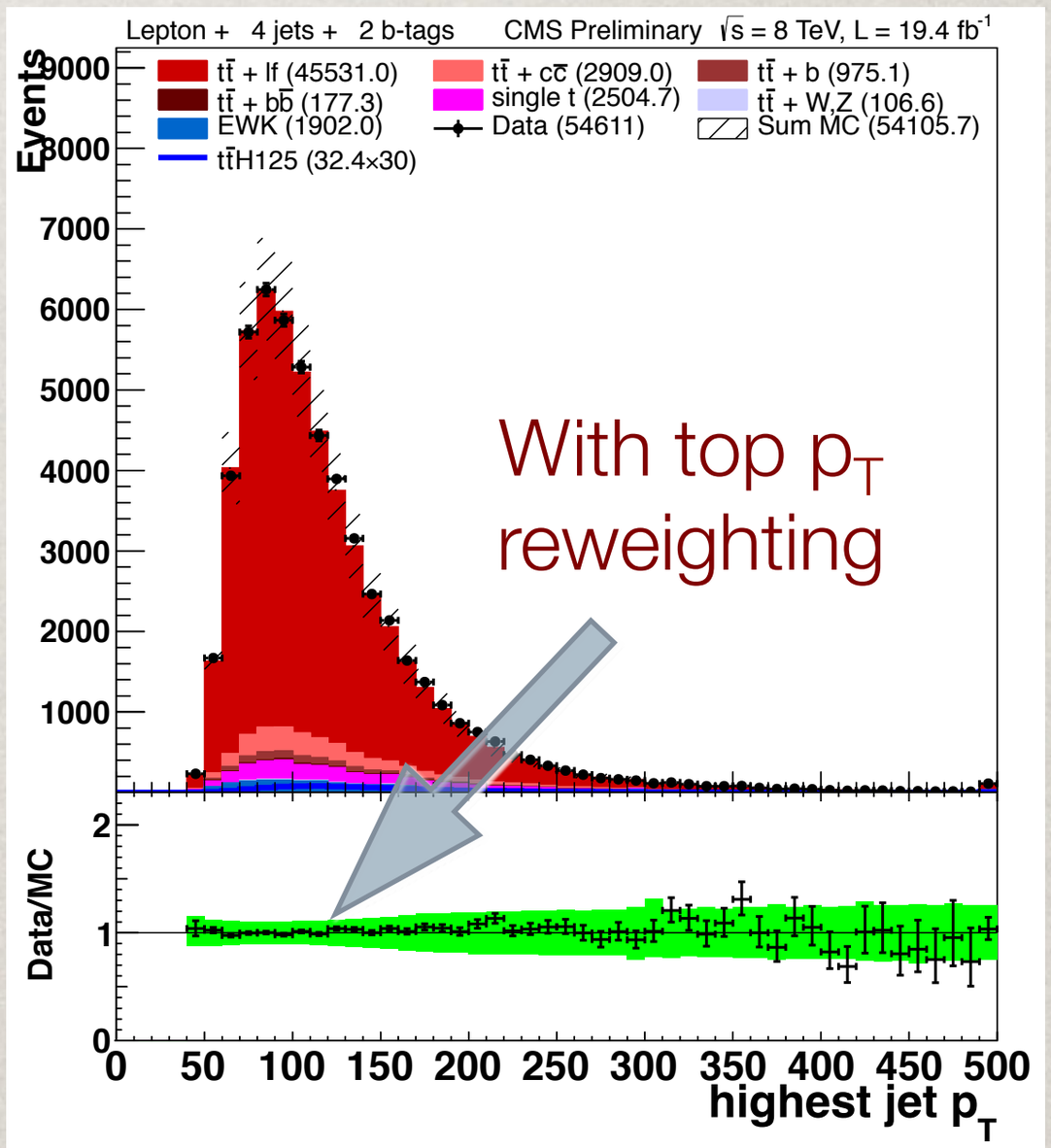


# EFFECT OF TOP $P_T$ REWEIGHTING

- Significant improvement in data/MC agreement



# SYSTEMATIC UNCERTAINTIES

- PileUp, Lepton/Trigger SF
- b-tag SF
- JER and JES
- Luminosity, Cross Section
- top Pt reweighting
- MC statistics
- $Q^2$ scale for MadGraph ttjets
- Extra 50% rate uncertainty for tt+HF

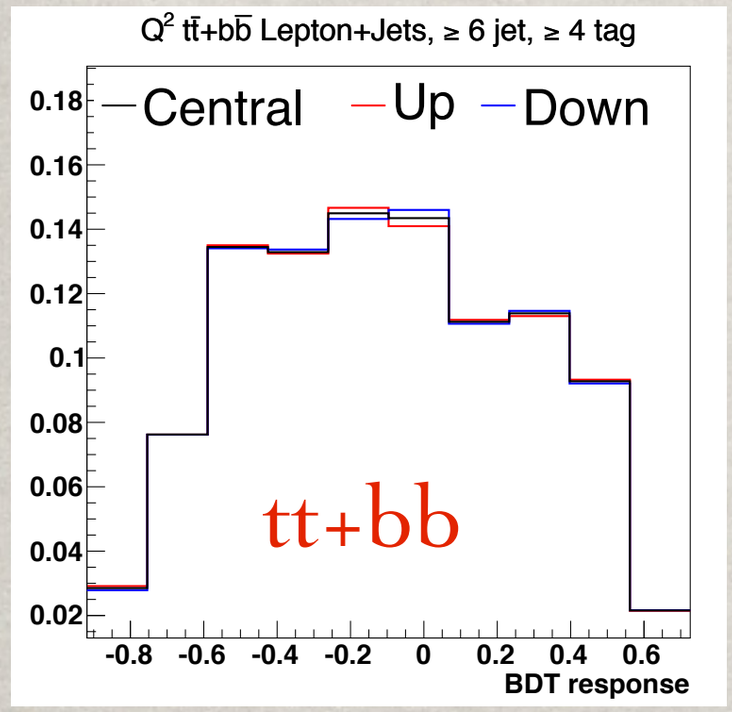
} TTJets

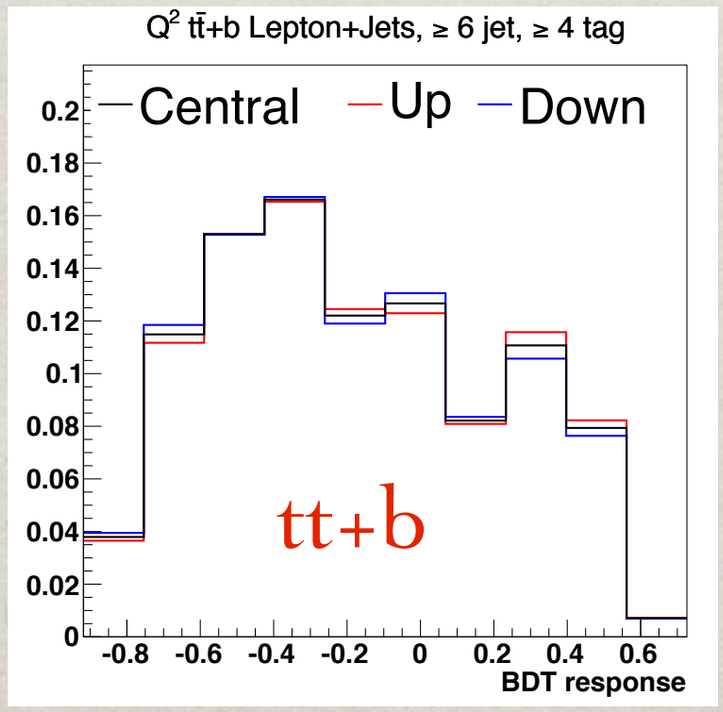
# MC STATISTICS

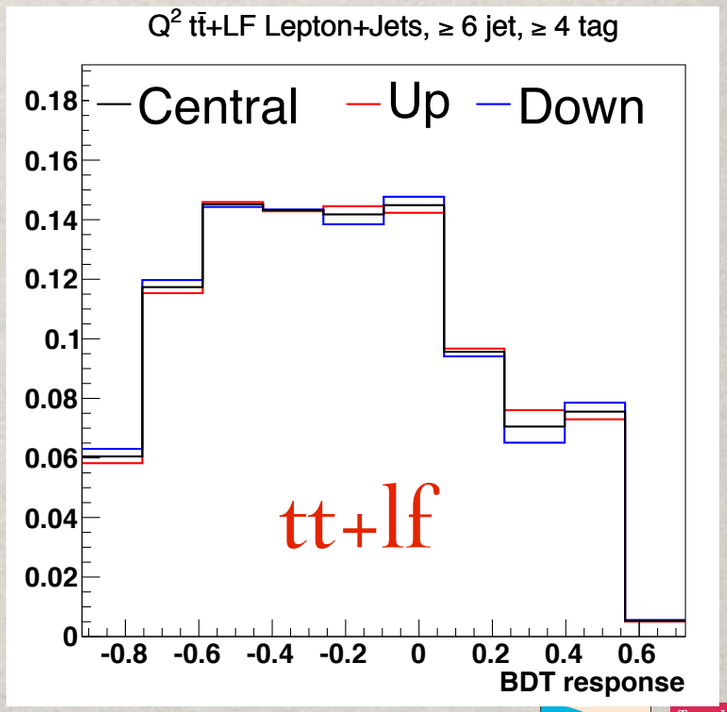
- Basic idea (Barlow-Beeston):
  - Assign separate uncertainty for every MC process in every MC bin to allow MC BDT shapes to float within stat. uncertainty
  - Include MC bin uncertainty if
    - it contributes significantly to total uncertainty
    - MC uncertainty comparable to data uncertainty
    - bin has sufficient signal
  - Gives observed MC stats effect with small number of nuisances: 36 stat. uncertainties

# $Q^2$ SCALE FOR TTJETS

- We use NLO xsec to normalize inclusive ttjets sample and MadGraph to get category normalization(number of jets)
  - At tree-level MC, biggest uncertainty in MadGraph is $Q^2$ scale
- Calculate weights for event based on standard shifts in both renormalization scale and factorization scale
- Split $Q^2$ scale uncertainty based on tt + LF vs HF and based on the number of extra jets in event(0,1,2)



# ADDITIONAL TT+HF UNCERTAINTY

- Contribution from tt+HF very similar to signal
  - uncertainty on rate and shape has a big impact on our ttH search
- Due to lack of more accurate higher order theory predictions, we obtained tt+HF estimate and uncertainty based on the inclusive ttbar sample
- Assign an extra 50% rate uncertainty for tt+bb, tt+b and tt+cc independently on top of other uncertainties
- Cross checks show 50% is a reasonable choice

# IMPACT OF SYSTEMATIC UNCERTAINTIES

- Top 5 uncertainties are related to ttjets backgrounds
- tt+bb rate uncertainty has a very large impact

| Removed Uncertainty | Improvement in limit (%) |
|---|---|
| tt+bb 50% rate uncertainty | 18.7 |
| top Pt reweighting | 5.0 |
| tt+b 50% rate uncertainty | 4.2 |
| tt+cc 50% rate uncertainty | 3.4 |
| MC statistical Uncertainties | 3.4 |

# IMPROVEMENT FOR RUN2

- Higher order calculation for ttjets samples
  - e.g. aMC@NLO+PYTHIA8 from KIT
- Better estimation of tt+HF
  - Dedicated tt+HF sample?
  - ttbb XS measurement from DESY, difficult to use due to the overlap of Signal region
- Discriminating variables ttH vs ttbb
  - Spin correlation, sub-jet structure etc
  - Inputs from theorists?

# SUMMARY

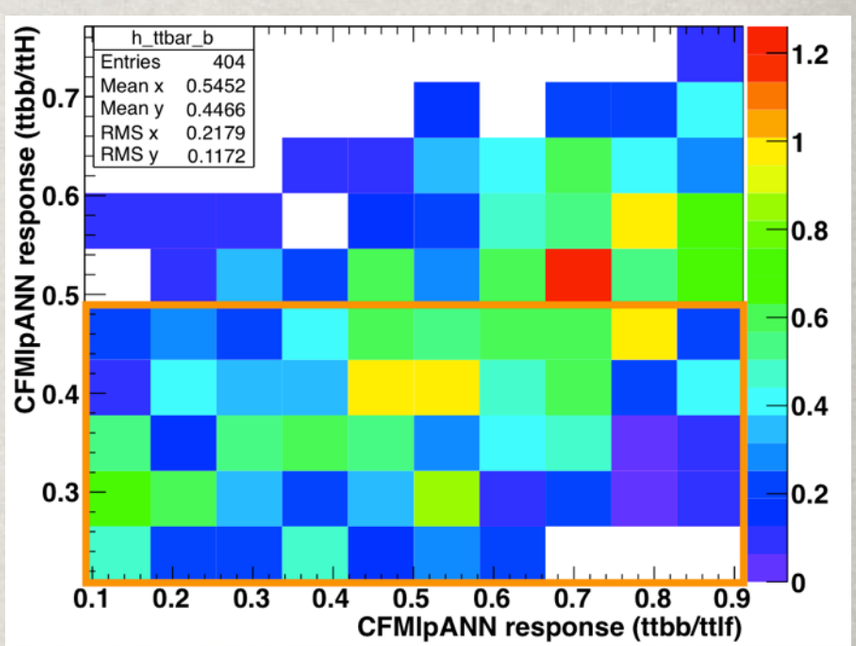
- TTjets is the main background for ttH(H—>bb)
  - irreducible tt+bb difficult to model
  - TTjets MC modeling related uncertainties(Rate or Shape) have large impact on our analysis
    - Especially the rate for tt+bb
- Higher order calculations and better tt+HF estimation will help us for RUN2

# BACK UP

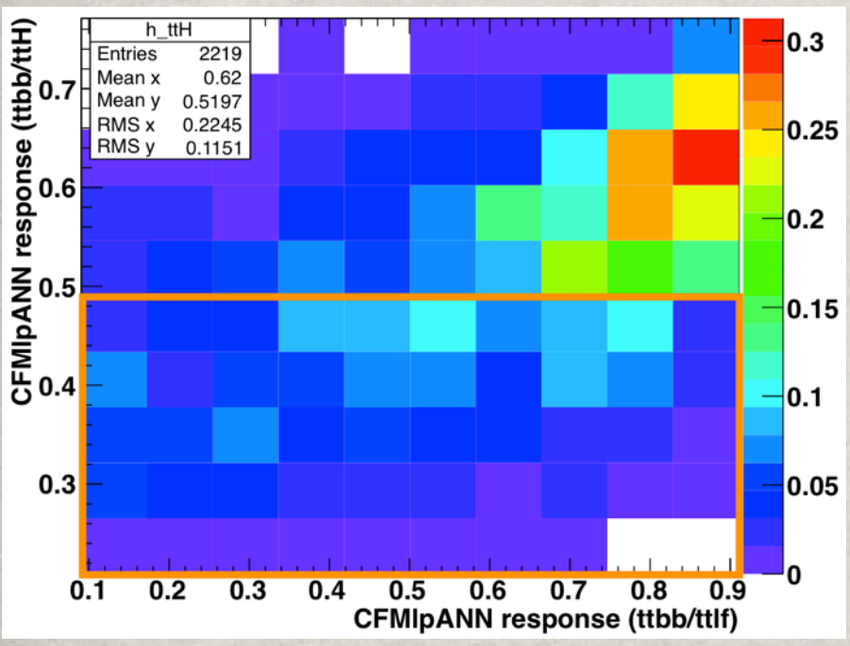
# TTBB/TTB CONSISTENCY CHECK

- Define 2 new MVAs:
  - ttb(b) vs ttlf (x-axis)
  - ttH vs ttb(b) (y-axis)
- Find a region with: (the orange one)
  - low ttH contamination
  - decent ttbb/ttlf ratio

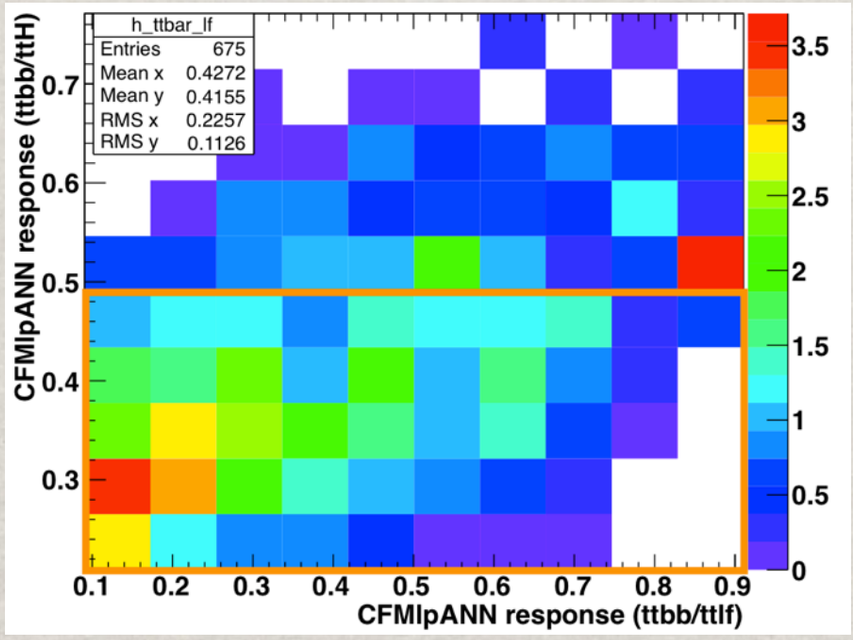
ttb



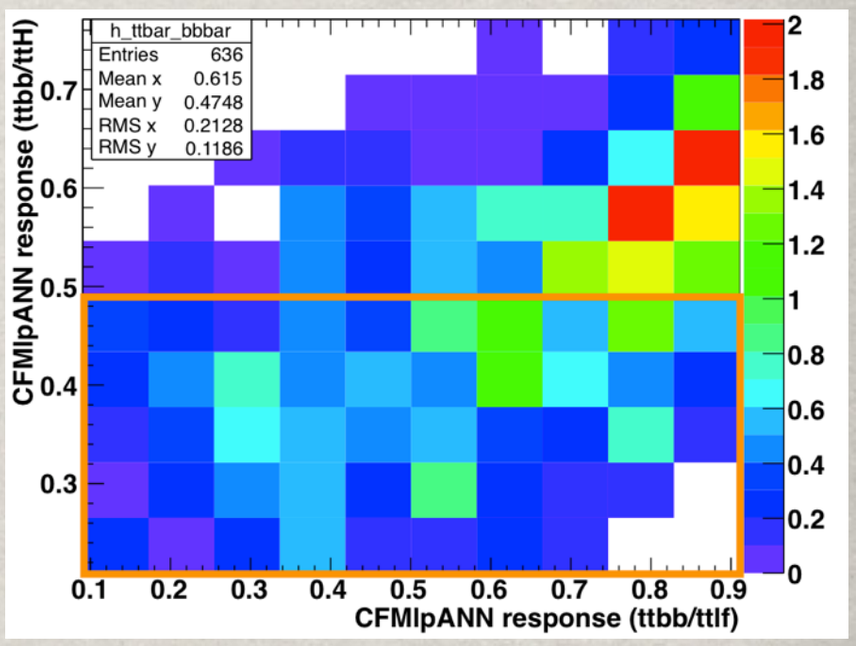

ttH




ttlf



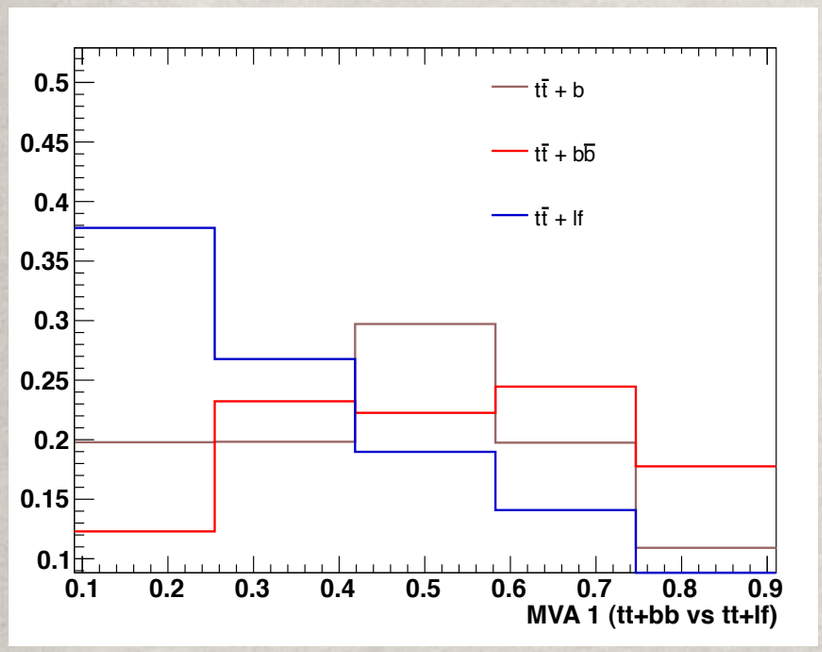

ttbb



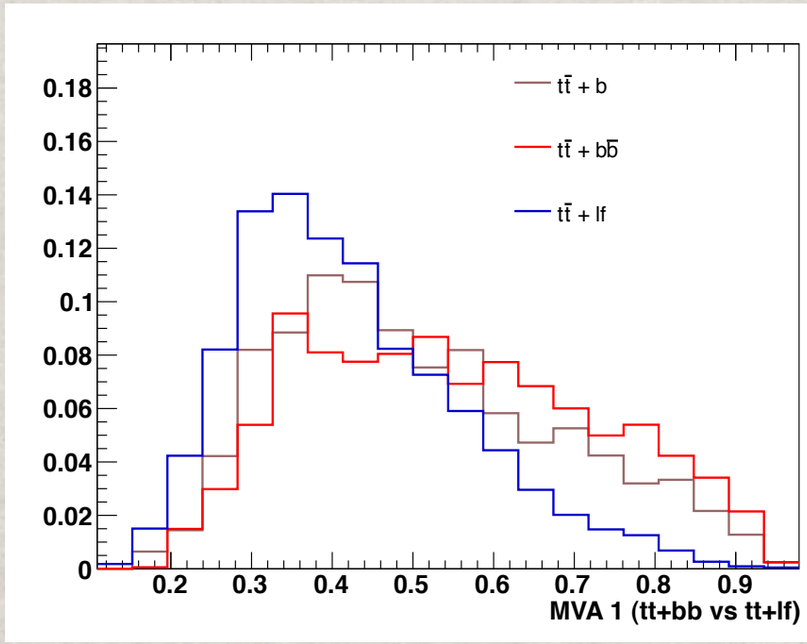
# TTBB/TTB CONSISTENCY CHECK

- Use the shape of the ttb(b) vs ttlf BDT
- Consider tt+bb and tt+b as 2 separate signals
- Find best fit value r while fitting to data using all *7* categories

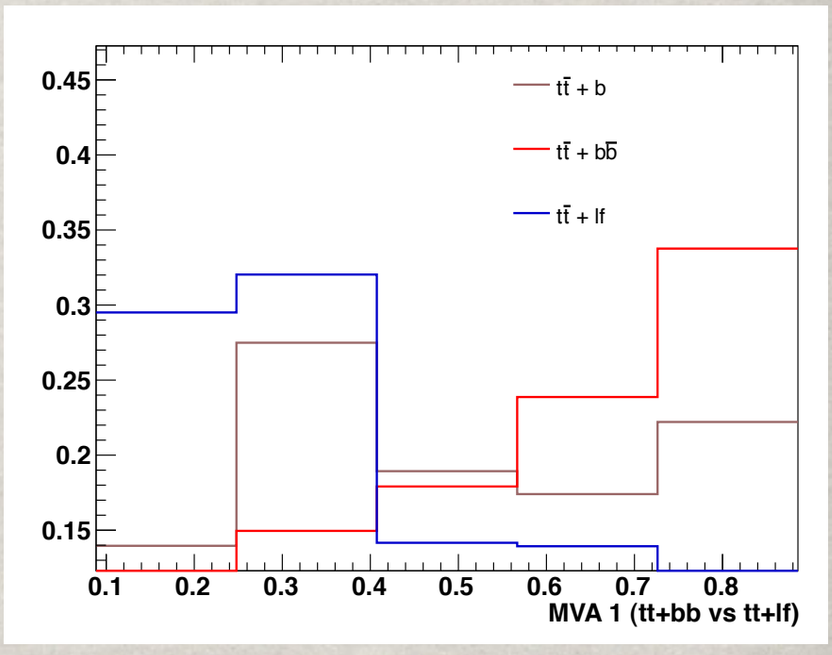
5jets+≥4tags




≥6jets+3tags




≥6jets+≥4tags

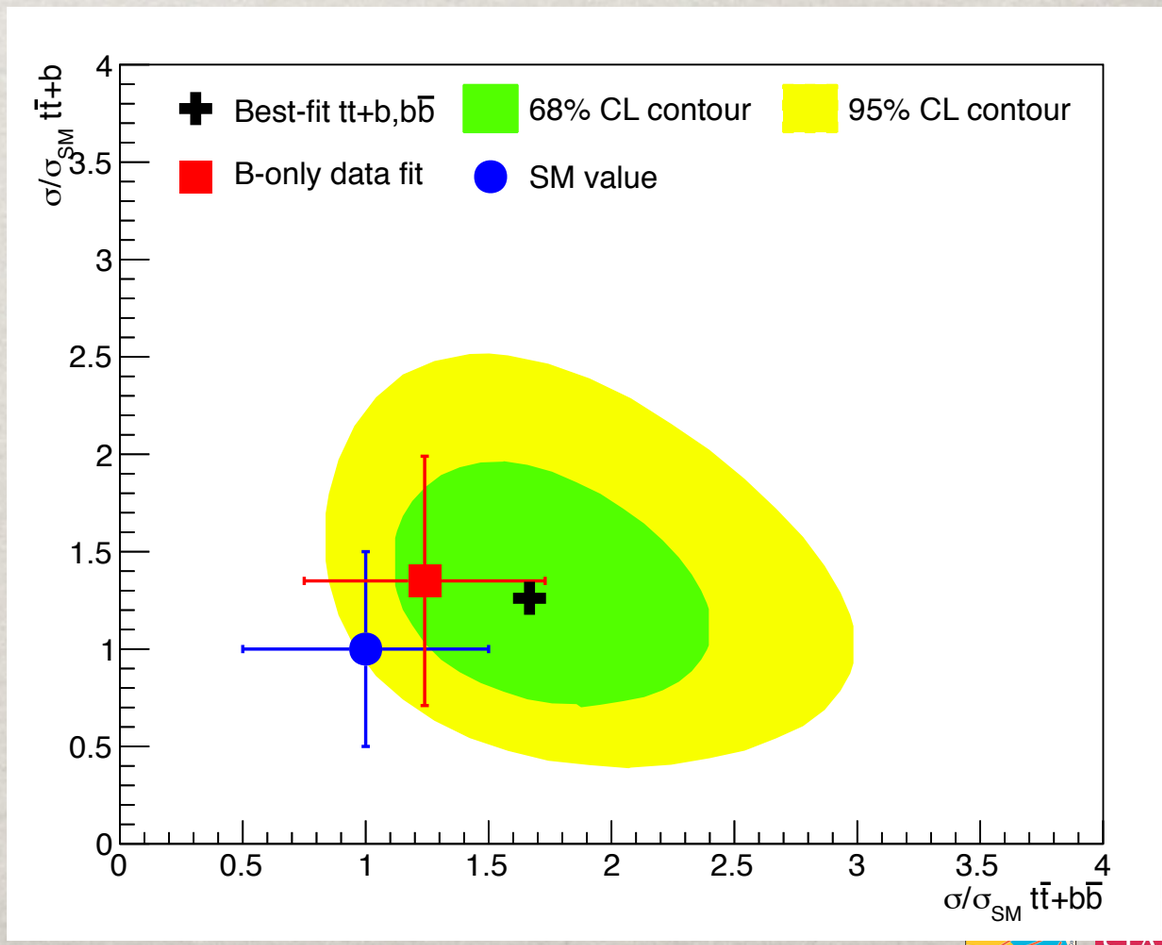


# TTBB/TTB CONSISTENCY CHECK

- The **2D fitting** includes all uncertainties except the extra 50% rate uncertainty for tt+bb and tt+b
- The fit result favors a larger amount of tt+bb and tt+b than the **SM prediction**, but still consistent within uncertainties
- We also tried a **bkg-only fit to data**, the nuisance parameters are pulled in a way consistent with this cross check

Black point: 2D fit for ttbb/ttb
Blue point: SM prediction with 50% uncertainty
Red point: B-only data Maximum Likelihood fit with errors

*both fits are using "combine" tool



# IMPACT OF SYS. UNCERT.

| Impact of removing one systematic uncertainty on the full analysis | |
|---|---|
| Removed uncertainty | Improvement in the limit (%) |
| CMS_ttH_QCDscale_ttbb | 18.7 |
| CMS_ttH_CSVCErr1 | 5.7 |
| CMS_ttH_topPtcorr | 5.0 |
| CMS_ttH_QCDscale_ttb | 4.2 |
| CMS_ttH_QCDscale_ttcc | 3.4 |
| Monte Carlo Statistical Uncertainties | 3.4 |
| CMS_ttH_CSVHF | 3.1 |
| QCDscale_ttH | 2.7 |
| CMS_ttH_CSVHFStats2 | 1.9 |
| Q2scale_ttH_ttbar1p | 1.9 |
| Q2scale_ttH_ttbar_bb | 1.1 |
| CMS_ttH_CSVHFStats1 | 1.1 |
| CMS_ttH_eff_lep | 1.1 |
| Q2scale_ttH_V | 1.1 |
| pdf_qqbar | 1.1 |
| QCDscale_ttbar | 1.1 |
| CMS_ttH_CSVCErr2 | 1.1 |
| CMS_scale_j | 0.4 |
| CMS_ttH_CSVLFStats1 | 0.4 |
| pdf_gg | 0.4 |
| CMS_ttH_CSVLF | 0.4 |
| CMS_ttH_CSVLFStats2 | 0.4 |
| Q2scale_ttH_ttbar2p | 0.4 |