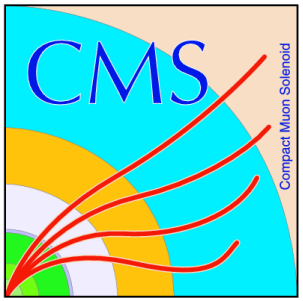


University of Zurich[UZH]

Data-driven background estimates in ttH(bb) and tt+jets at CMS

Daniel Salerno

Higgs Toppings Workshop 2018

29 May 2018, Benasque, Spain

Fully hadronic ttH(bb)

CMS-HIG-17-022
arXiv:1803.06986

Event selection

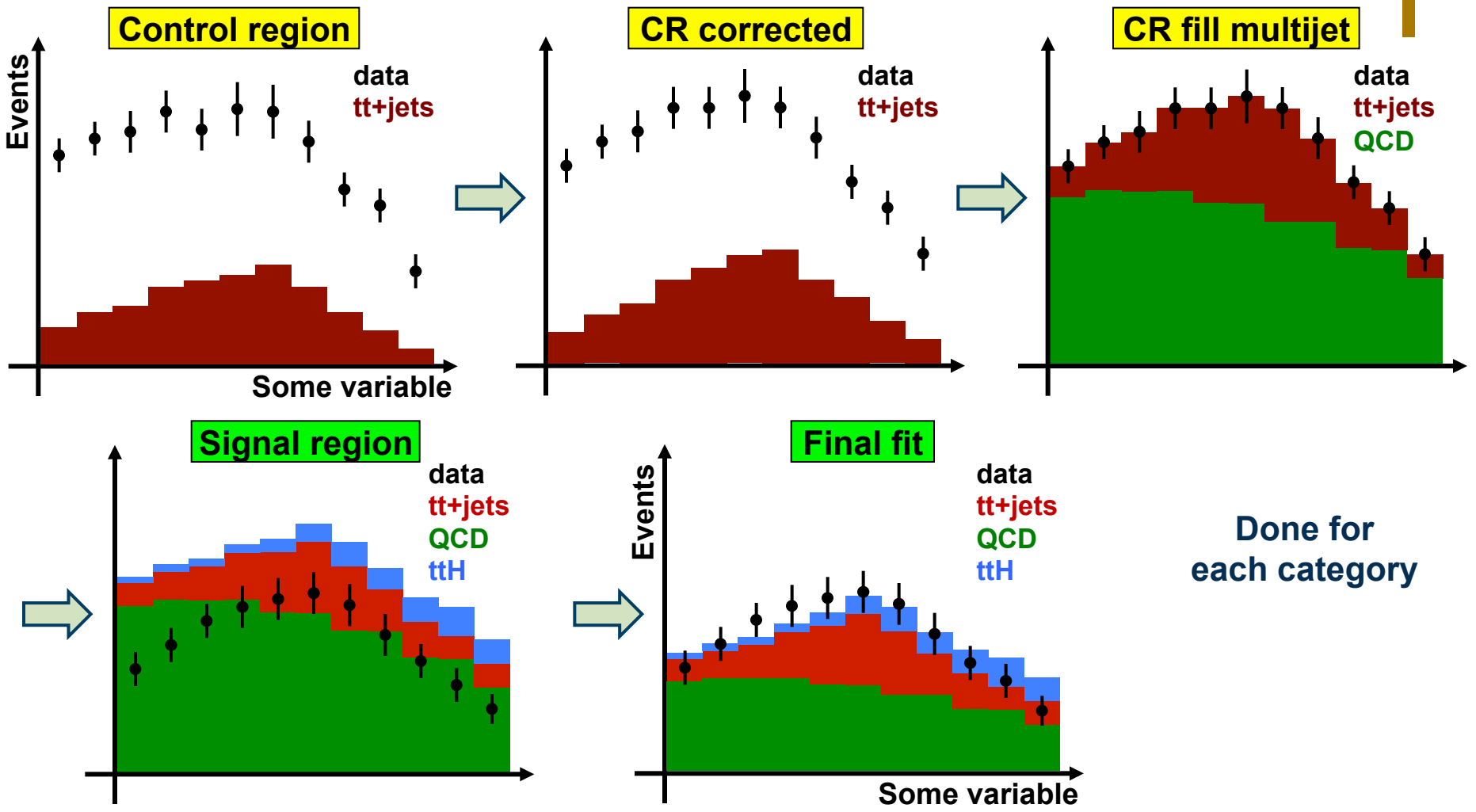
- ≥7 jets
- ≥3 medium b-tagged jets
- H_T > 500 GeV
- Quark-gluon likelihood ratio > 0.5
- Six categories:
  - 7 jets, 3 b-tags
  - 8 jets, 3 b-tags
  - ≥9 jets, 3 b-tags
  - 7 jets, ≥4 b-tags
  - 8 jets, ≥4 b-tags
  - ≥9 jets, ≥4 b-tags

Background estimation

- Control region in each category
  - Same selection except for b-tagging:
  - 2 medium b-tags
  - 3 or ≥4 loose b-tags
- Correction for kinematic differences of b-jets
  - p_T, η, and ΔR_{min} from the 2 medium b-tags
- **QCD** shape as the difference between data and ttbar MC
  - Normalisation left free-floating in final fit
  - Shape uncertainties from differences in validation region

Fully hadronic ttH(bb)

CMS-HIG-17-022
arXiv:1803.06986

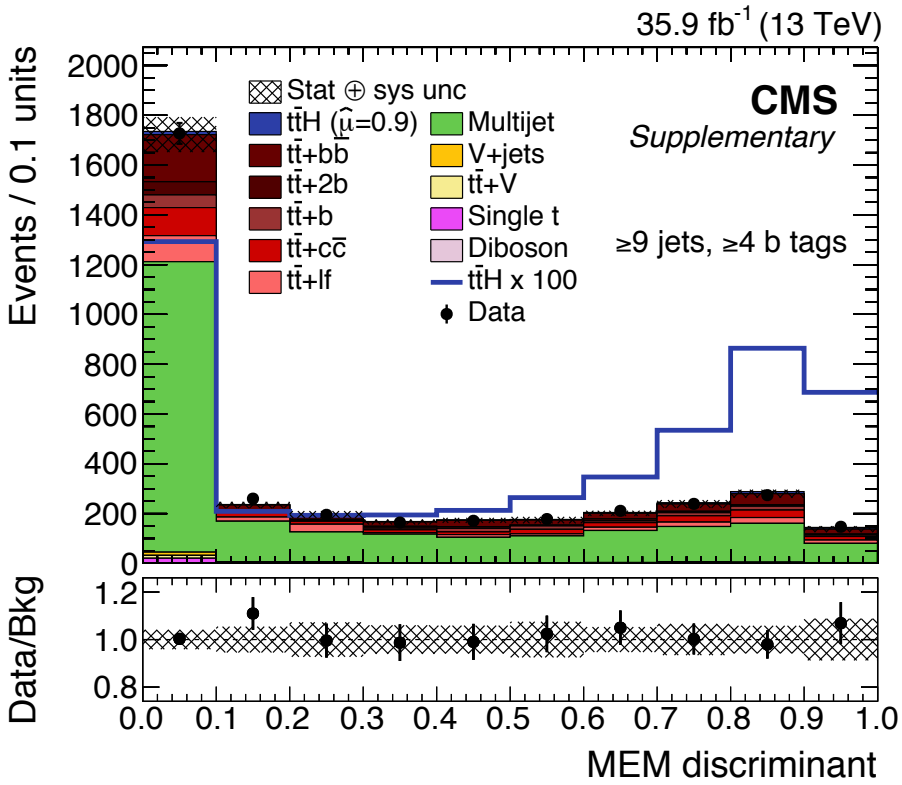


Done for each category

Fully hadronic ttH(bb)

CMS-HIG-17-022
arXiv:1803.06986

Final estimate



1 of 6 categories

Questions

- Can we use information from high-yield, high-background categories to constrain the lower-yield, signal-rich categories?
- Can we get any better estimate of shape in the SR?

Fully hadronic ttbar

CMS-TOP-16-013
CMS-PAS-TOP-16-013

Event selection

- ≥6 jets
- ≥2 medium b-tagged jets
- H_T > 500 GeV
- ΔR_{bb} > 2.0
- Cut on kinematic fit probability
  - Equal top quark masses
  - Fixed W masses (80.4 GeV)
- Boosted analysis:
  - ≥2 fat jets (AK8) with p_T > 200 GeV and soft-drop mass > 50 GeV
  - ≥1 b-tagged subjet in each fat jet
  - Fisher discriminant cut on n-subjettiness ratios (3 vs. 2/1 prong)

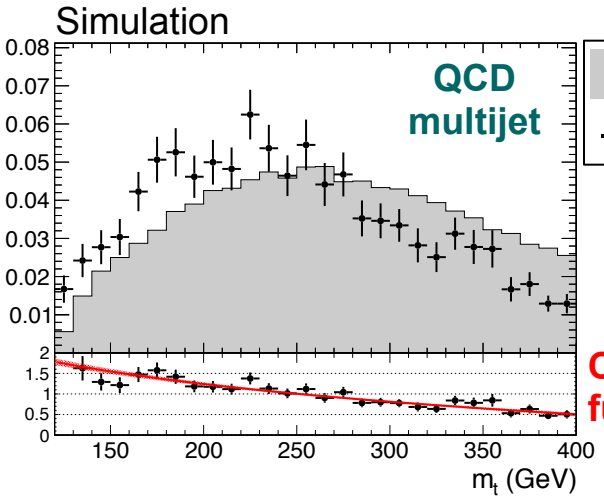
Background estimation

- Control region in each category
  - Same selection except for b-tagging:
  - 0 medium b-tags (jets/subjets)
- Correction for kinematic differences of b-jets
  - Done for the variable of interest using ratio SR/CR from MC
  - m_t for signal extraction
    - From kin. fit or leading fat jet m_{SD}
  - Smooth log-linear function used
- **QCD** shape as corrected data
  - Normalisation left free-floating in final fit
  - Shape unc. from MC transfer fn.

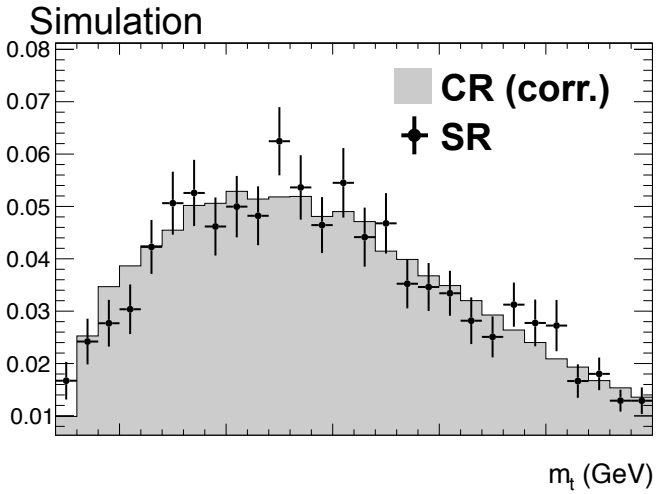
Fully hadronic ttbar

CMS-TOP-16-013
CMS-PAS-TOP-16-013

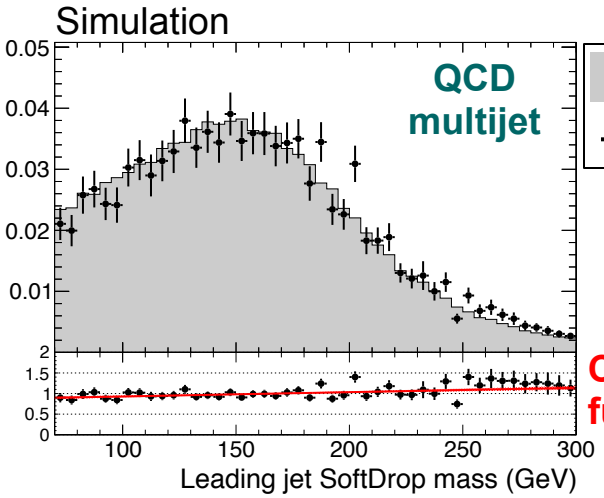
Resolved



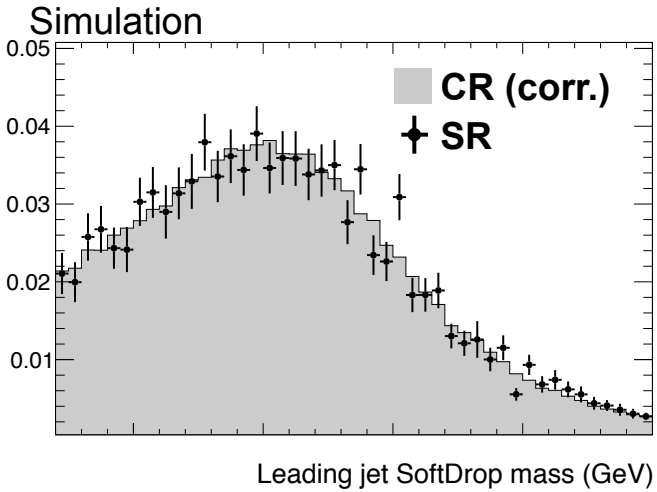
Correction function



Boosted



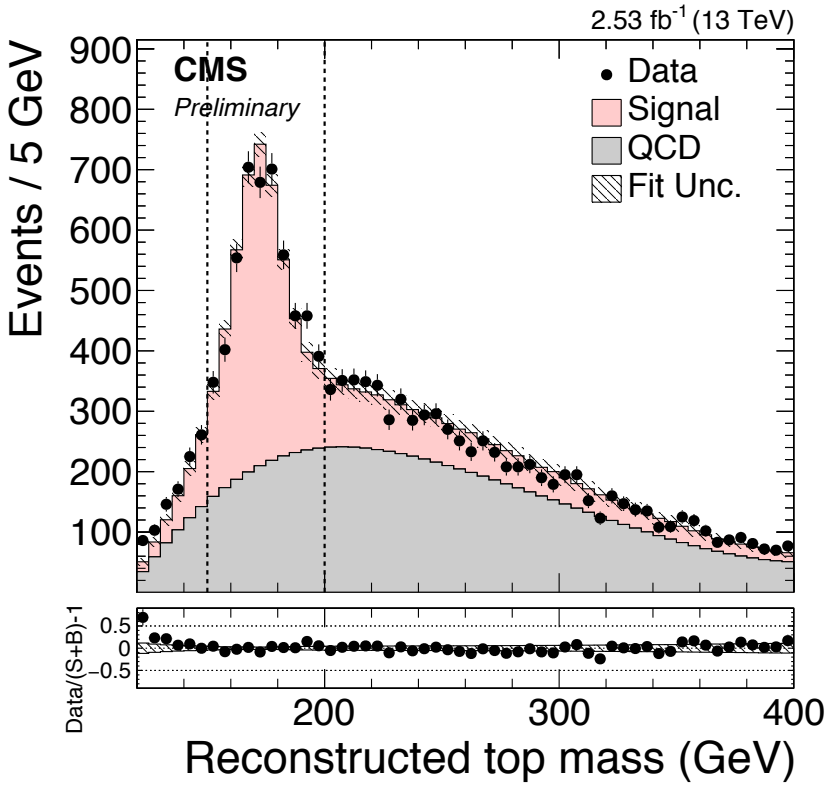
Correction function



Fully hadronic ttbar

CMS-TOP-16-013
CMS-PAS-TOP-16-013

Final estimate



Done for all variables

Question

- Can we remove the reliance on MC for shape differences between CR and SR?
- Can we link the boosted and resolved regions to add some constraints?

Semi-leptonic ttbar

CMS-TOP-17-002
arXiv:1803.08856

Event selection

- 1 lepton (e/μ)
- ≥4 jets
- ≥2 medium b-tagged jets

Backgrounds

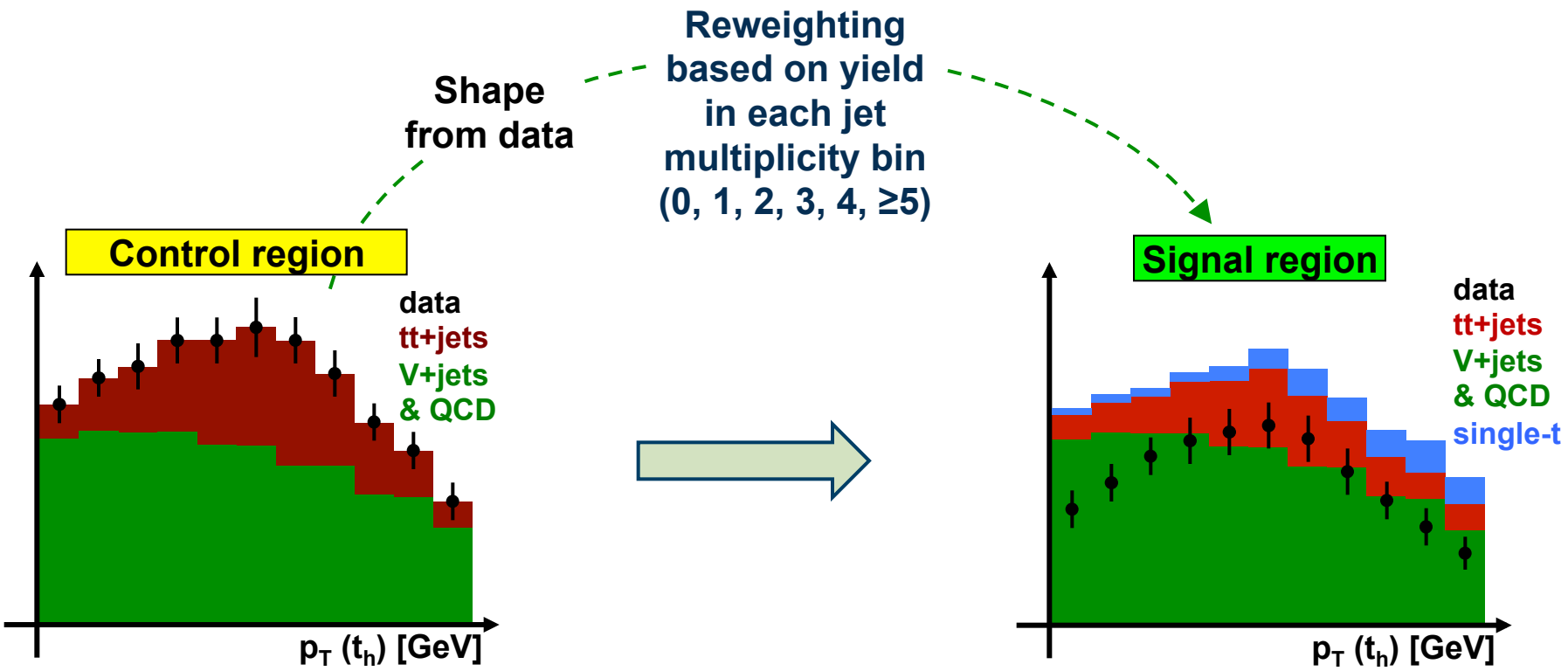
- Single top quark
  - ~3% of selected data
  - Estimate from MC
- V+jets & QCD multijet
  - ~ 2% of selected data
  - Combined estimate from data

Background estimation

- Control region in each category
  - Same selection except for b-tagging:
  - 0 medium b-tags
  - ~ 85% V+jets & QCD
  - ~ 15% ttbar
- Plot p_T distribution
- Scale normalisation to MC yield in SR for each jet multiplicity bin
- **V+jets + QCD** shape as scaled CR data
  - Shape uncertainties from n-jet reweighting and CR deviations
  - 8% normalisation uncertainty

Semi-leptonic ttbar

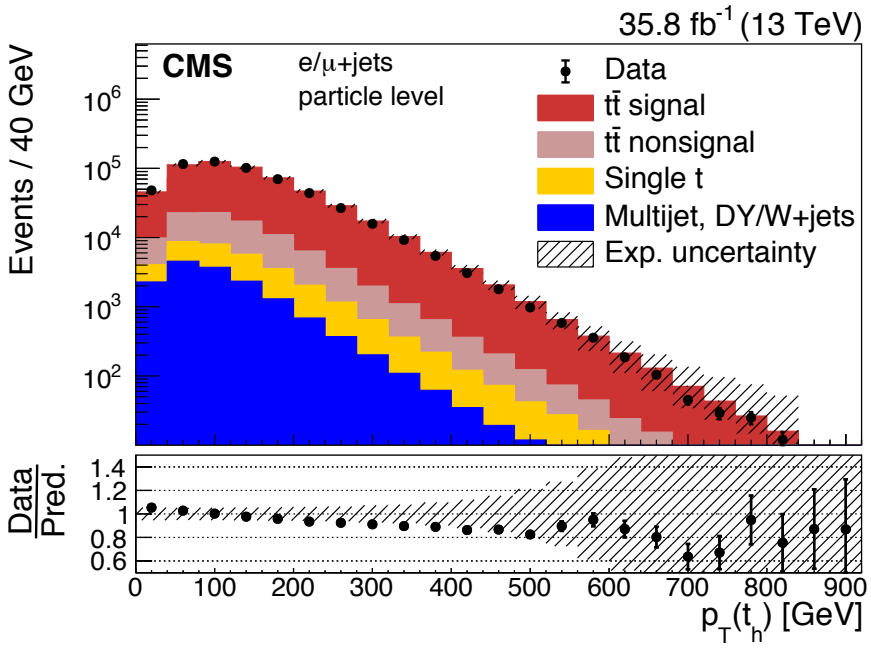
CMS-TOP-17-002
arXiv:1803.08856



Semi-leptonic ttbar

CMS-TOP-17-002
arXiv:1803.08856

Final estimate



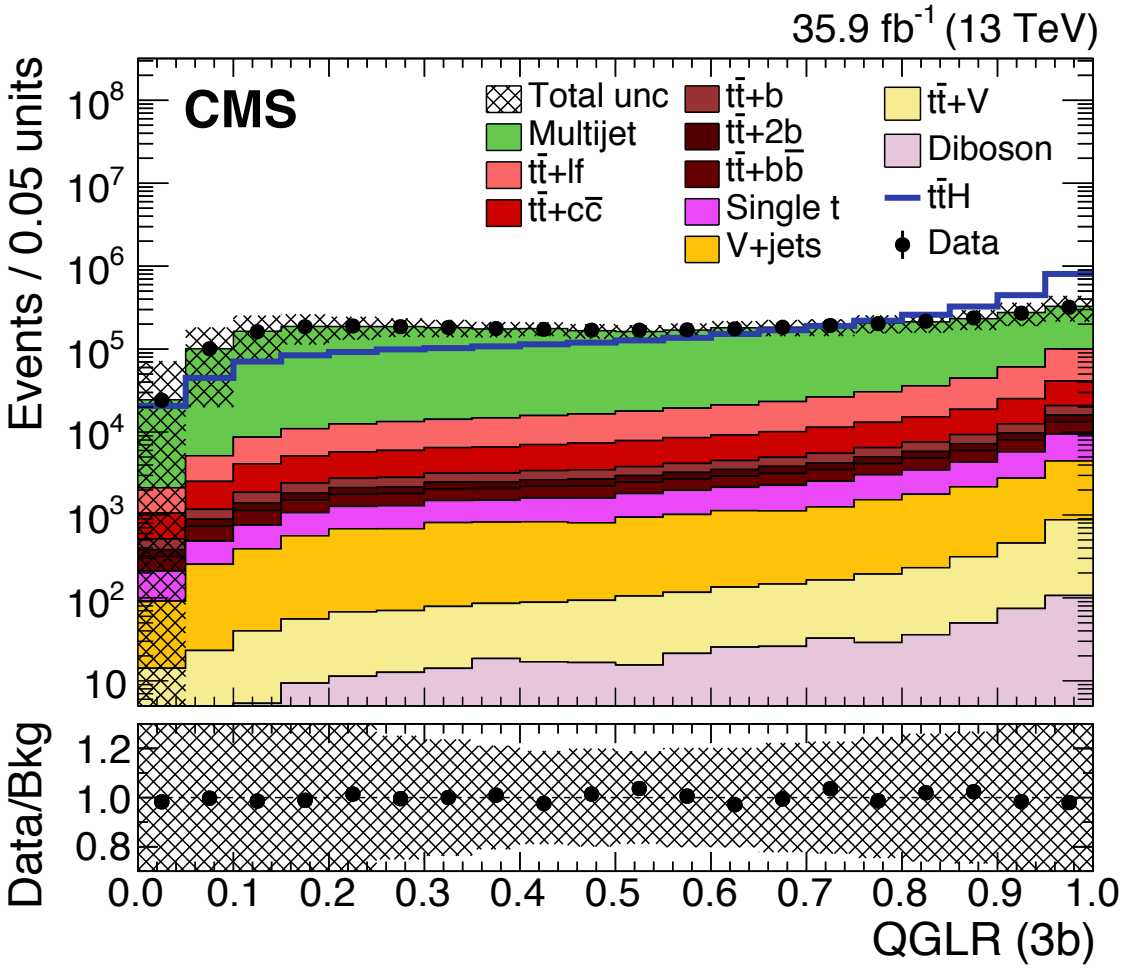
Done for all variables

Issues

- Can we improve by subtracting the 15% of ttbar in the CR?
  - How would this affect systematics?
- Can we remove reliance on MC yield?

Backup

QGLR

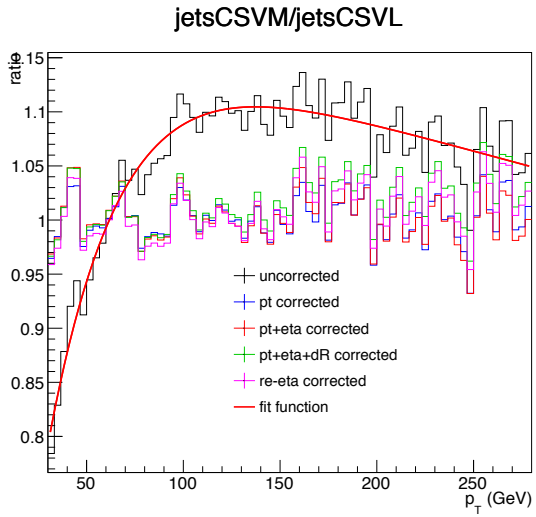

35.9 fb-1 (13 TeV)
CMS
Total unc
Multijet
tt̄+lf
tt̄+cc̄
tt̄+b
tt̄+2b
tt̄+bb̄
Single t
V+jets
tt̄+V
Diboson
tt̄H
Data
Events / 0.05 units
Data/Bkg
QGLR (3b)

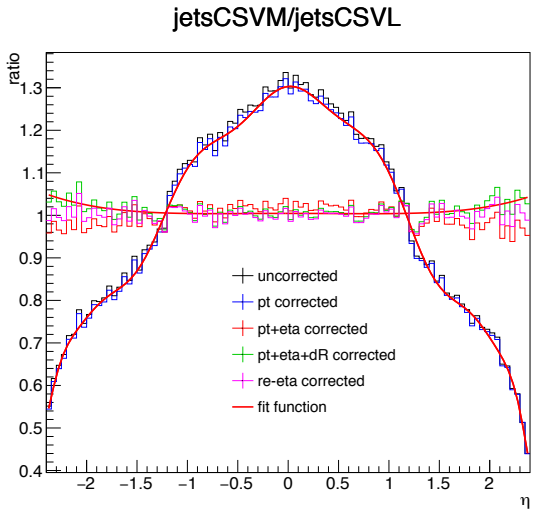
CR corrections

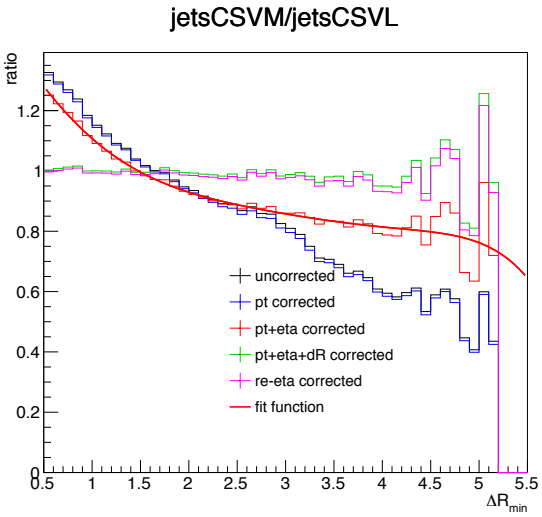
CERN-THESIS-2018-031
CMS-TS-2018-003

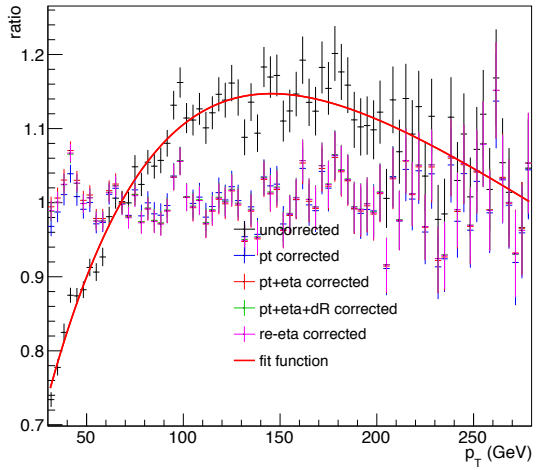
data

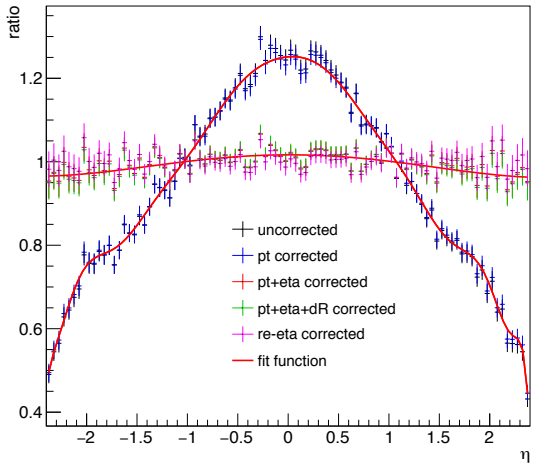
ttbar MC

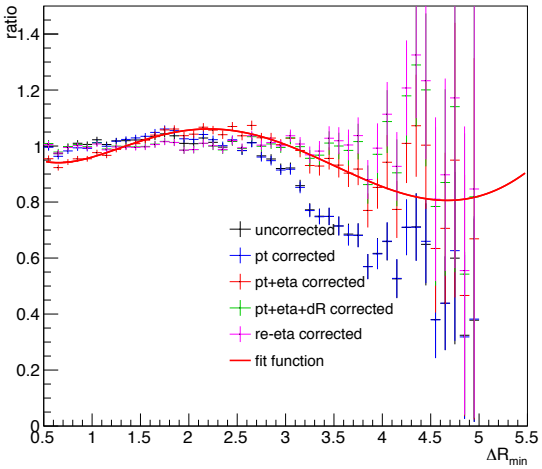




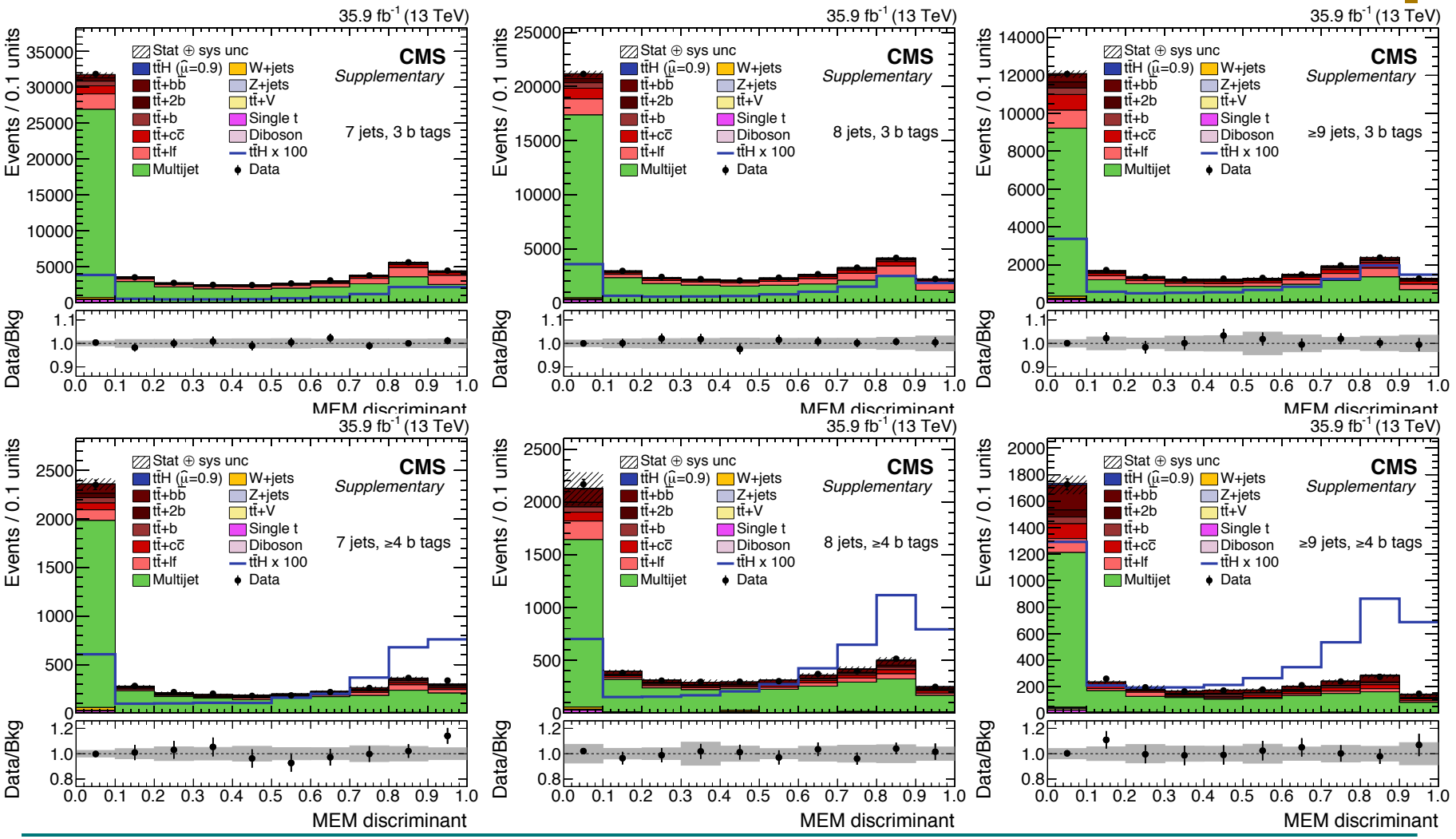






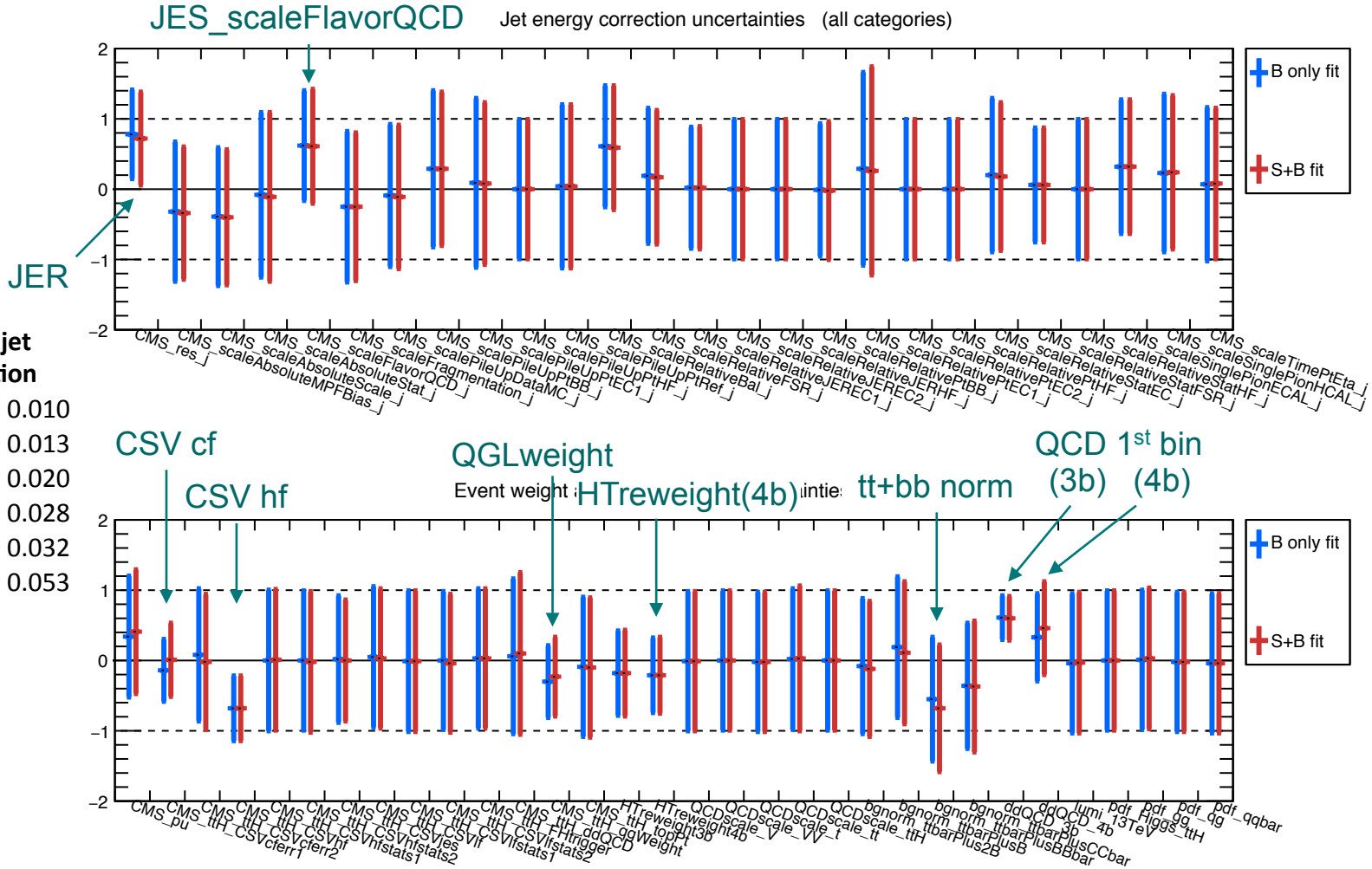


Post-fit MEM distributions



ML-fit pulls

CERN-THESIS-2018-031
CMS-TS-2018-003



QCD multijet normalisation

7j,3b	1.005 ± 0.010
8j,3b	1.005 ± 0.013
9j,3b	1.008 ± 0.020
7j,4b	1.012 ± 0.028
8j,4b	1.010 ± 0.032
9j,4b	1.031 ± 0.053