

# Full spectrum differential $t\bar{t}$ cross sections in $e/\mu + \text{jets}$



O. Hindrichs  
On behalf of the CMS collaboration



University of Rochester

– LHCtopWG –

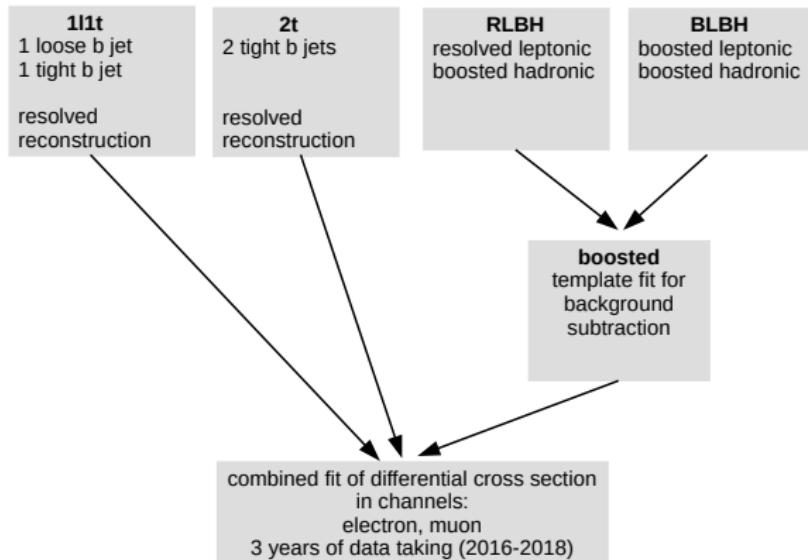
19.05.2021

# Introduction

Documentation: CMS-PAS-TOP-20-001

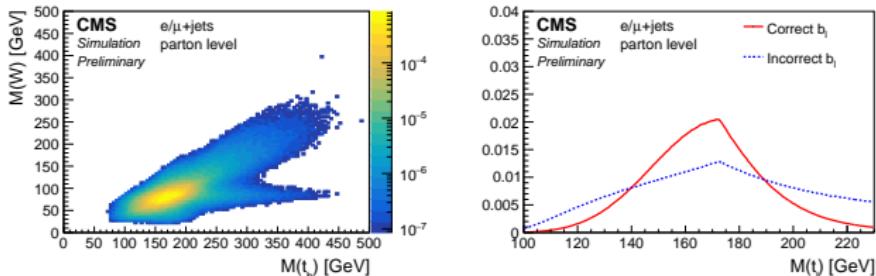
- high precision measurements of differential and double differential cross sections of  $t\bar{t}$  production using CMS Run 2 data ( $137 \text{ fb}^{-1}$ )
- including resolved and boosted reconstruction methods
- results at the parton level in full phase space
- results at particle level, where similar to the detector-level reconstruction a combination of resolved and boosted  $t\bar{t}$  events are defined as signal.
- precise inclusive  $t\bar{t}$  production cross section

Overview:



## Selection

- Exactly 1 isolated electron or muon and at least 4 jets
- 2 b-tagged jets or 1 loose and 1 medium tagged jet



## Reconstruction

For each permutation of jets calculate  $M(t_l)^2 = (p(b_l) + p(l) + p(\nu))^2$ :  
 $p_z(\nu)$  is calculated using  $m_W^2 = (p(l) + p(\nu))^2$  constraint.

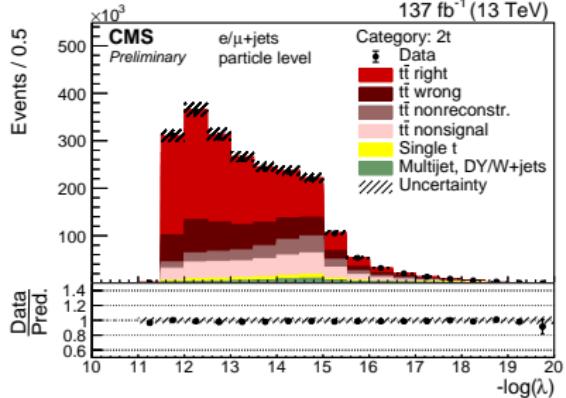
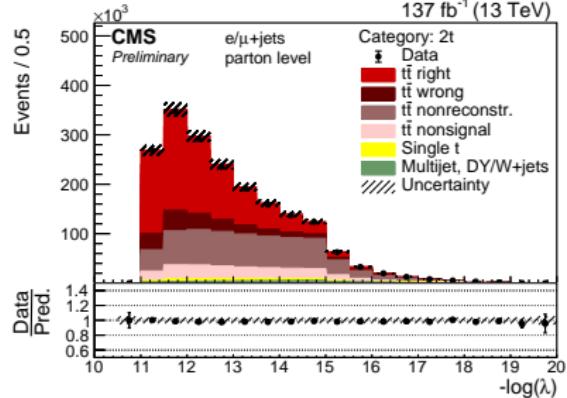
- 0 solutions: find minimal modification of  $p_T^{\text{miss}}$  to obtain a single solution.
- 2 solutions: select with minimum  $|M(t_l) - m_{\text{top}}|$ .

with  $m_2$  and  $m_3$  the invariant masses of the jets tested as W and  $t_h$  on hadronic side calculate:

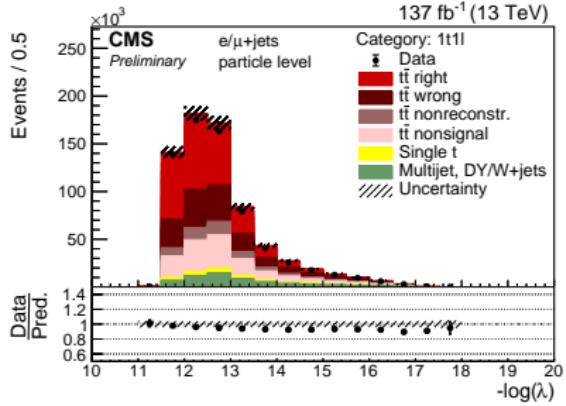
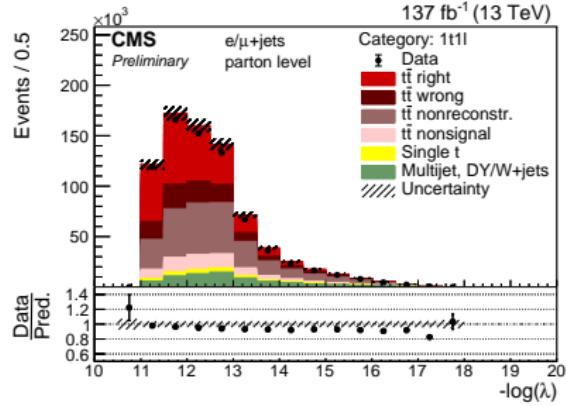
$$\begin{aligned} -\log(\lambda) &= -\log(\lambda_m) - \log(\lambda_\nu) \\ \lambda_m &= P_m(m_2, m_3) \\ \lambda_\nu &= P_\nu(M(t_l)) \end{aligned}$$

→ select permutation minimizing  $-\log(\lambda)$

## Two tight b-tagged jets (2t)



## One loose and one tight b-tagged jet (1t1l)



## Boosted $t_1$ reconstruction

- 1 nonisolated  $\mu/e$   $p_T > 50$  GeV  
1 b jet (loose criterion) within  $\Delta R(\ell, b) = 0.6$
- Remove lepton from jet if constituent; reconstruct neutrino as in resolved categories
- $p_T(t_1 = \ell + b + \nu) > 400$  GeV,  $100 < M(t_1) < 230$  GeV

**Use artificial neural network (NN) to separate from background (QCD multijet and W+jets events):**

$$M(\ell, j) \quad p_T(\ell)/p_T(\ell + j) \quad M(j)/M(\ell, j)$$

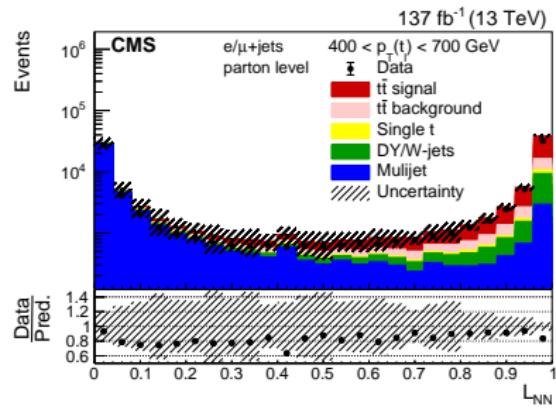
$$I_{\text{near}}/I_0 \quad I_{\text{far}}/I_0$$

$$\text{where } I = \sum_{\text{PF}} p_T(\text{PF}) \Delta R^q(\ell, \text{PF})$$

with  $q = -2, 0, 2$  for  $I_{\text{near}}$ ,  $I_0$ , and  $I_{\text{far}}$ , respectively.

Particle flow (PF) objects in  $\Delta R < 0.4$  around

lepton.



Require:  $L_{\text{NN}} > 0.7$ , efficiency 90% in whole  $p_T$  range

Very efficient in removing multijet events, but W boson background very similar to  $t\bar{t}$

*Efficiency measured in  $t\bar{t}$   $e\mu$ -events with one isolated and one nonisolated lepton.*

## Boosted $t_h$ reconstruction

Boosted  $t_h$  identified as anti- $k_T$  (size =0.8) PUPPI jet with  $p_T > 400 \text{ GeV}$ ,  $|\eta| < 2.4$ , and  $m_{\text{jet}} > 120 \text{ GeV}$ .

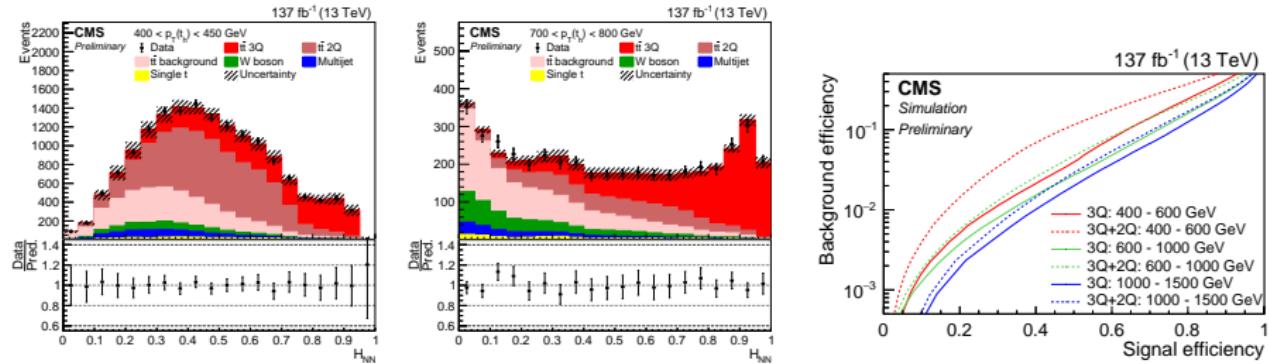
NN used for identification of boosted  $t_h$ .

Boost all constituents into jet center-of-mass frame and cluster with anti- $k_T$  (size =0.5) algorithm. Use 21 input variables for  $H_{\text{NN}}$ :

two and three sub-jets invariant masses, energies of leading sub-jets, n-jettiness ratios, sphericity of sub-jets, ... (*full list in backup*)

4 NN trained in  $p_T$  regions 400–500–700–1000–2000 GeV

All input variables are well modeled in the  $p_T$  regions and years.



- 3Q: 3 quarks from  $t_h$  decay point towards the candidate ( $\Delta R < 0.8$ ).
- 2Q: 2 quarks from  $t_h$  decay point towards the candidate ( $\Delta R < 0.8$ ).

The definitions of 3Q and 2Q are arbitrary. We repeat analysis with ( $\Delta R < 0.7/0.9$ ) final cross sections are the same.

# Event reconstruction and categorization

## Resolved:

- $-\log(\lambda_m) < 11(9)$   
for  $2t(1t1l)$

## BHRL

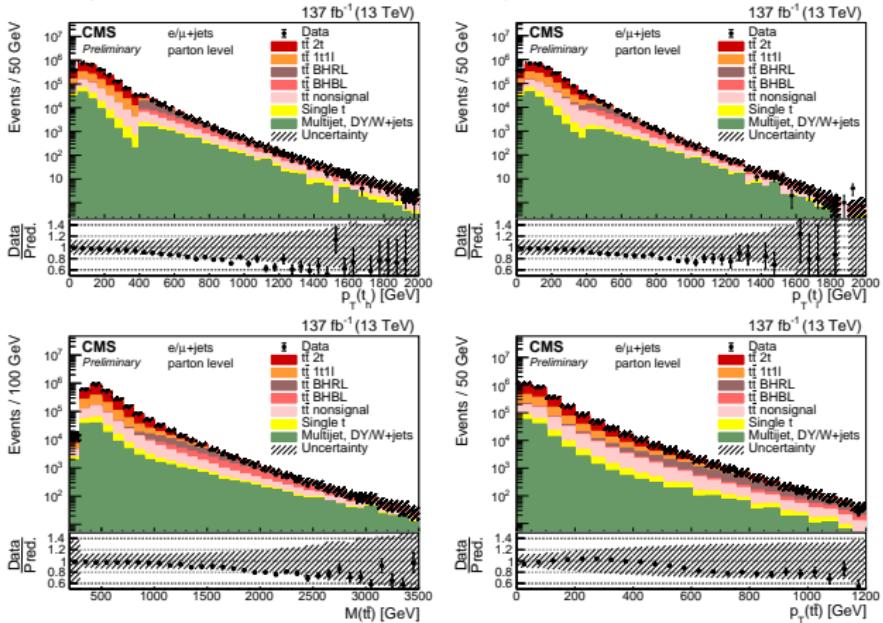
- 1 iso.  $\mu/e + 1 b$  jets  
(resolved reco.)
- 1 boosted  $t_h$  candidate

## BHBL

- 1 boosted  $t_l$  candidate
- 1 boosted  $t_h$  candidate

- Priority:  $2t$ , BHRL/BHBL,  $1t1l$

- If there is more than 1 boosted  $t_h$  candidate, distributions are filled for each candidate. The yields of  $2Q$  and  $3Q$  (never more than one per event) are extracted using  $H_{NN}$  fit.



# Background fit in boosted categories

$H_{NN}$  template fit performed in 9 bins of  $p_T(t_h)$  vs. variable of interest.

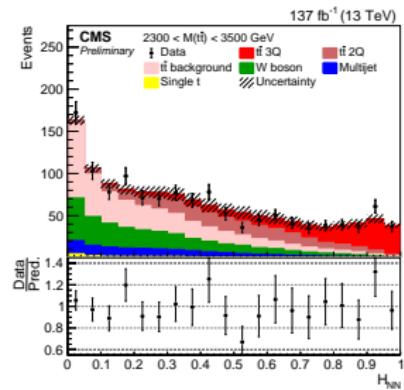
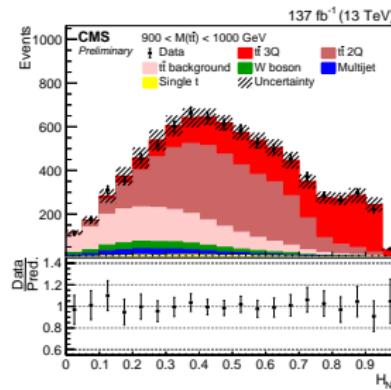
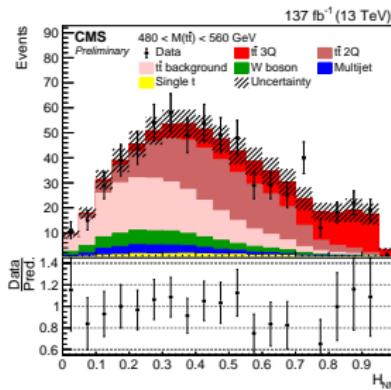
$p_T(t_h)$ -bins: 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1500 GeV (not all bins always filled)

- $t\bar{t}$  signal: 2Q and 3Q fitted separately ; unconstrained from MC
- $t\bar{t}$  background: one or no quark from  $t_h$  decay ; unconstrained from MC
- W/DY background: template from quark enriched jet data ;  $\pm 50\%$  MC prediction
- multijet background: template from b enriched jet data ;  $\pm 50\%$  MC prediction
- single top plus W: can contribute  $t_h$ , template from MC ; MC uncertainties (scale, JES, ...)

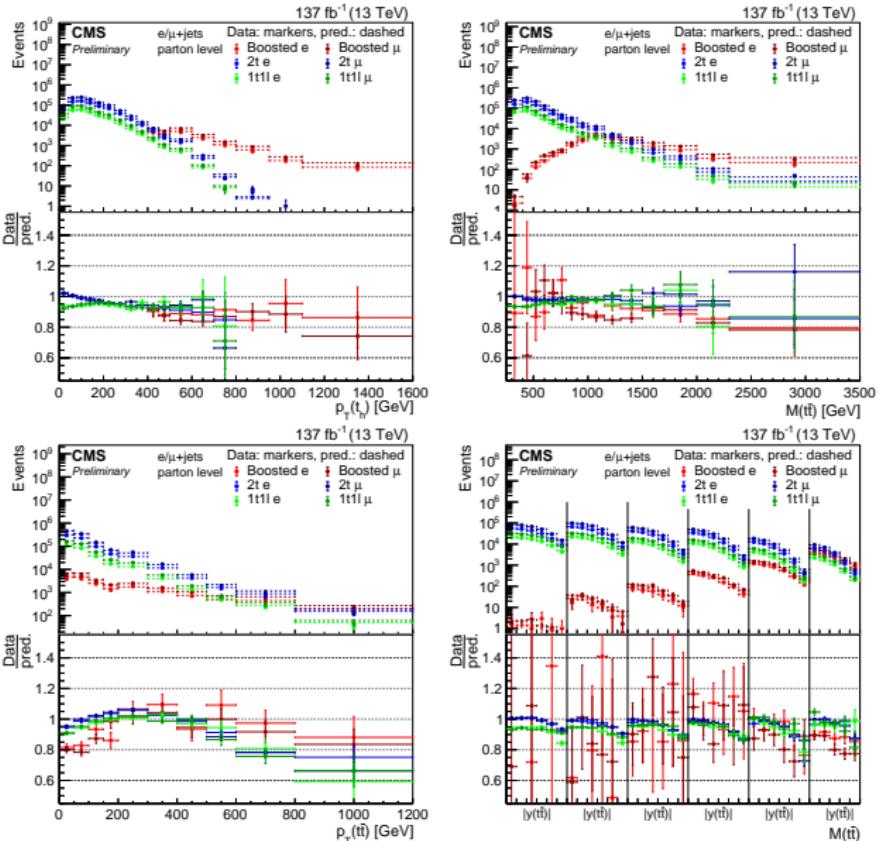
MC template shape uncertainties considered: shower scale, pileup, tune, jet energy, and subjet energy scale.

Uncertainties are externalized and propagated to the fit of the cross sections.

## Fit $M(t\bar{t})$ distribution integrated over $p_T$ bins and added for all categories



# Extracted signal yields



Ratio plots of  $p_T(t_h)$  show smooth transition from resolved to boosted reconstruction methods.

# Common unfolding of all categories

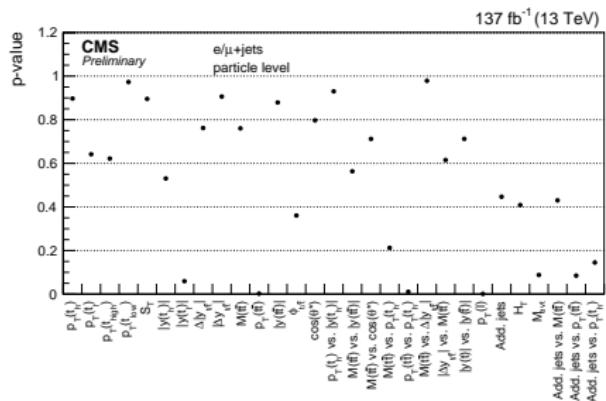
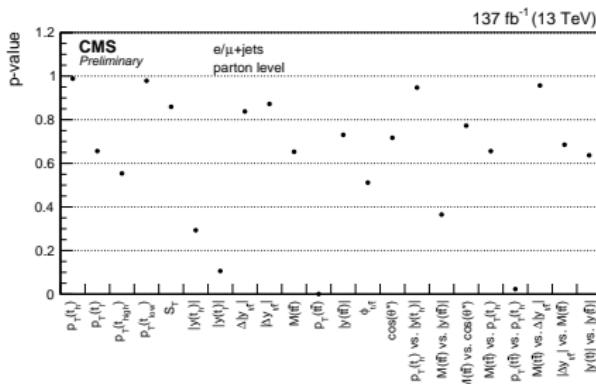
Perform  $\chi^2$ -fit to extract the cross sections:

$$\chi^2 = \sum_{\ell} \sum_y \sum_r (\mathbf{m}_{yrl} - \mathbf{b}_{yrl}(\nu_{\alpha}) - M_{yrl}(\nu_{\alpha})\boldsymbol{\sigma})^T C_{yrl}^{-1} (\mathbf{m}_{yrl} - \mathbf{b}_{yrl}(\nu_{\alpha}) - M_{yrl}(\nu_{\alpha})\boldsymbol{\sigma}) + \kappa(\nu_{\alpha})$$

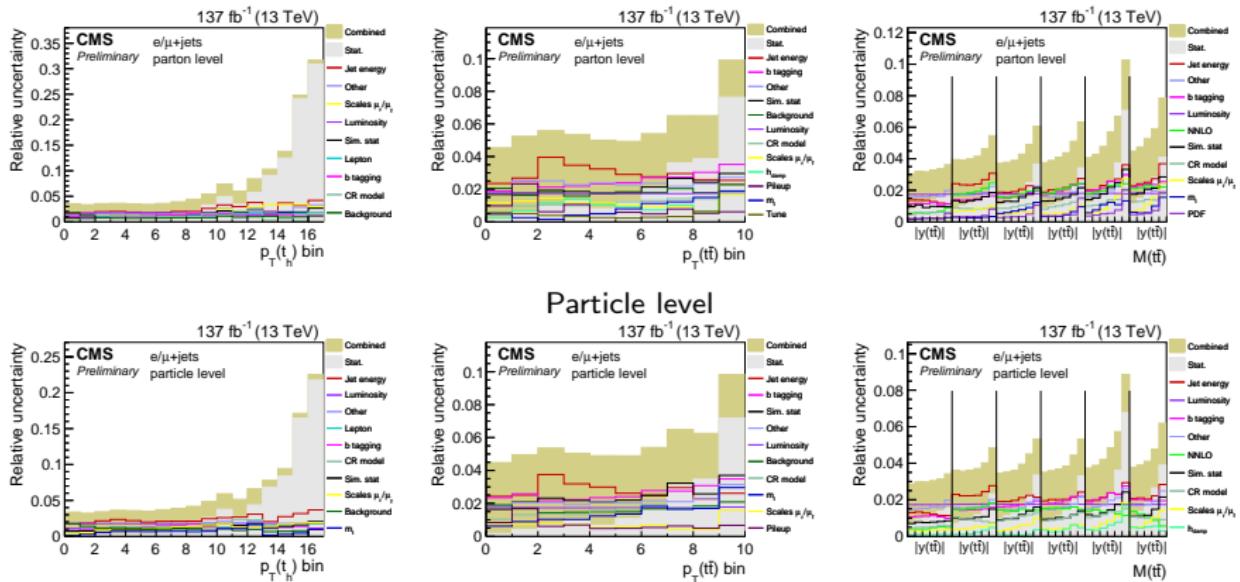
- $\mathbf{m}$  measured distribution of events with the covariance matrix  $C$ ; per year  $y$ , reconstruction  $r$  method (res. 1t1l, res. 2t, boosted), and lepton channels  $\ell$  (18 categories)
- $\boldsymbol{\sigma}$  vector of cross sections (free parameters of interest)
- $M(\nu_{\alpha})$  response matrices that map  $\boldsymbol{\sigma}$  to the  $t\bar{t}$  event yields at detector level.
- $\mathbf{b}(\nu_{\alpha})$  non  $t\bar{t}$  background (only unc. in boosted cat., since background already subtr.).
- $\nu_{\alpha}$  nuisances representing the uncertainties. These are constrained in  $\kappa(\nu_{\alpha})$  taking into account year-by-year correlations.

– No regularization condition is used

Goodness of fits



# Uncertainties

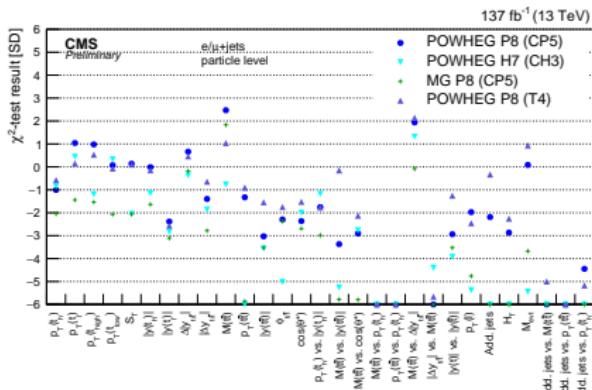
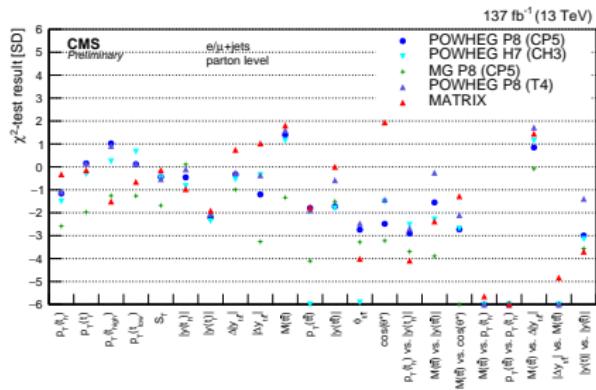


- “Sim. stat”: repeat cross section fits 100 times with response matrices and their uncertainties varied randomly by the statistical uncertainty of the simulation.
- “NNLO”: difference in top quark  $p_T$  between POWHEG+P8 and NNLO prediction.
- Dominated by JES and b-tagging uncertainties at both levels
- More diagonal response matrices at particle level reduce stat. effect in unfolding.

# Results

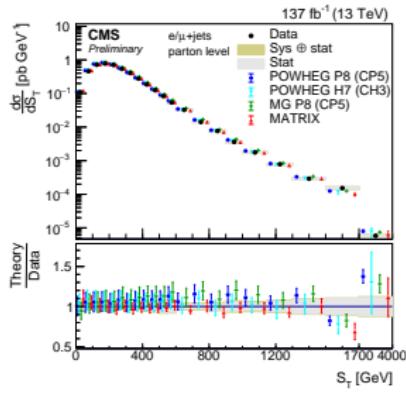
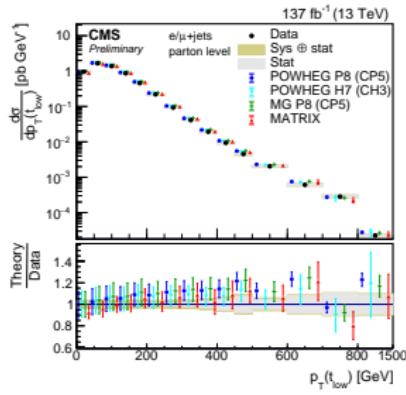
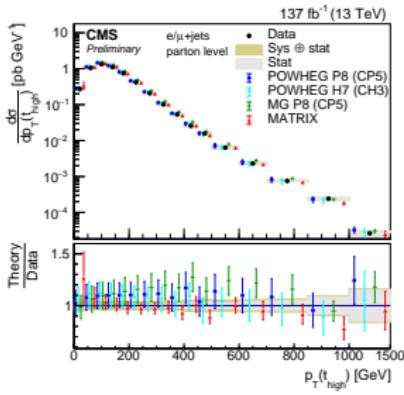
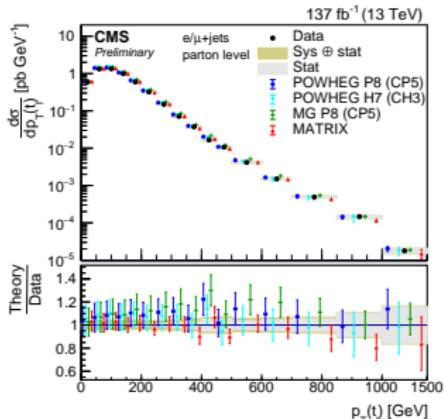
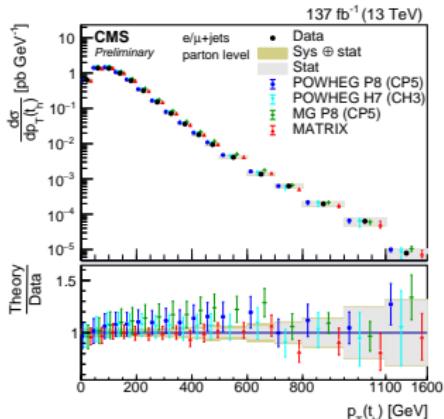
Comparison of measurements to various predictions using  $\chi^2$ -tests.

- uncertainties in measurements and predictions are included.

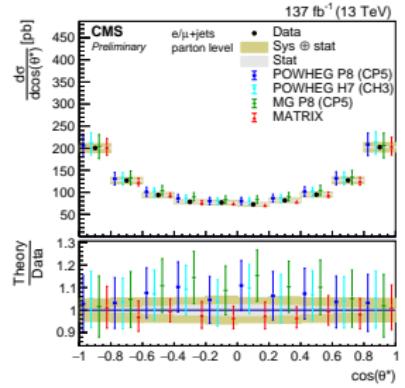
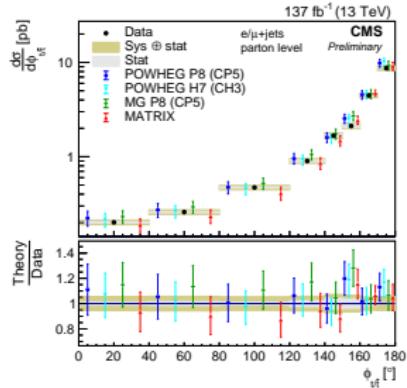
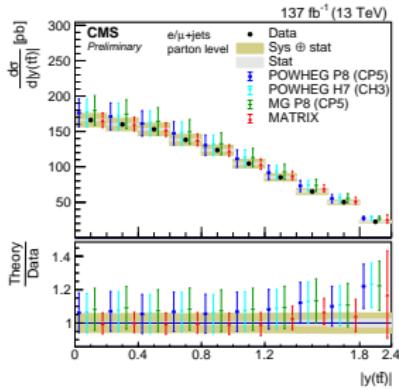
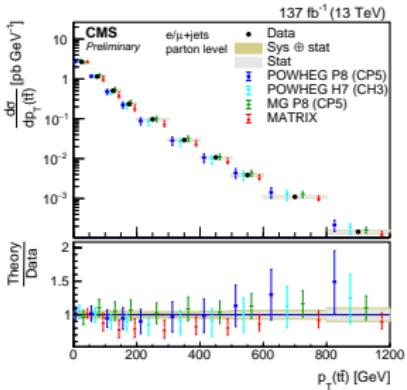
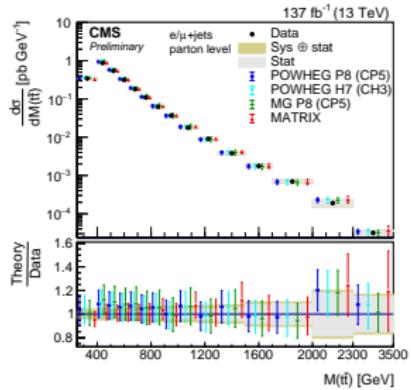


Most of the predictions are in good agreement with the measurement — with a few exceptions:

- $M(t\bar{t})$  vs.  $p_T(t_h)$  and  $p_T(t\bar{t})$  vs.  $p_T(t_h)$  shows largest disagreements.
  - At particle level add. jets vs. kinematic observable are difficult to describe.
- uncertainty in NNLO calculation (MATRIX) is smaller.

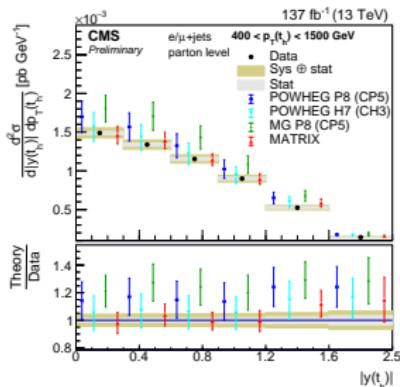
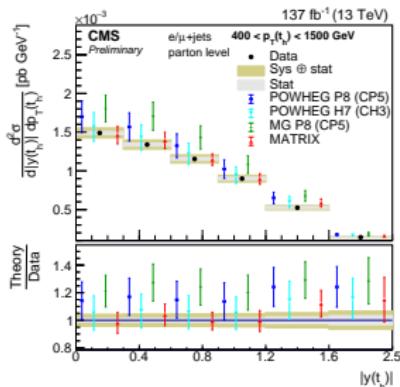
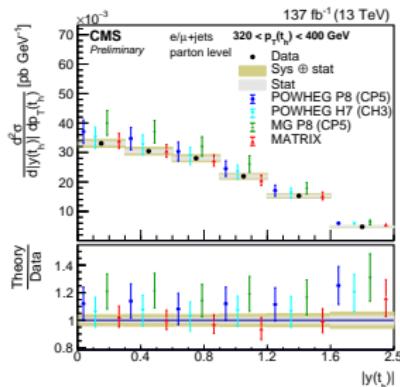
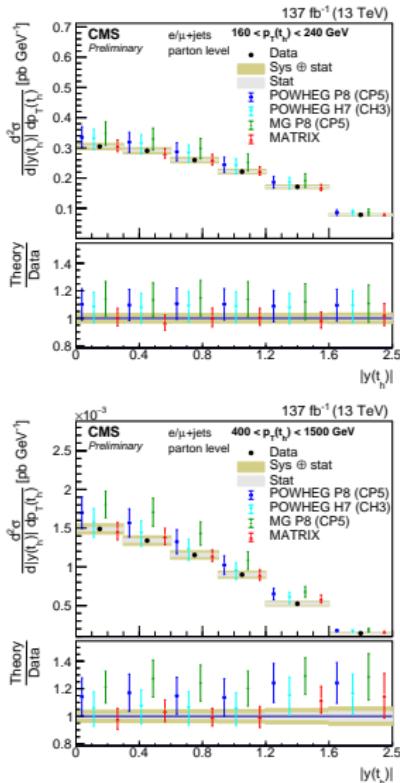
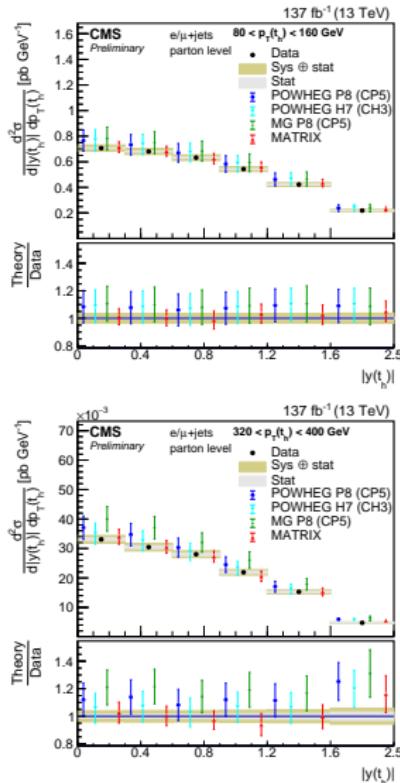
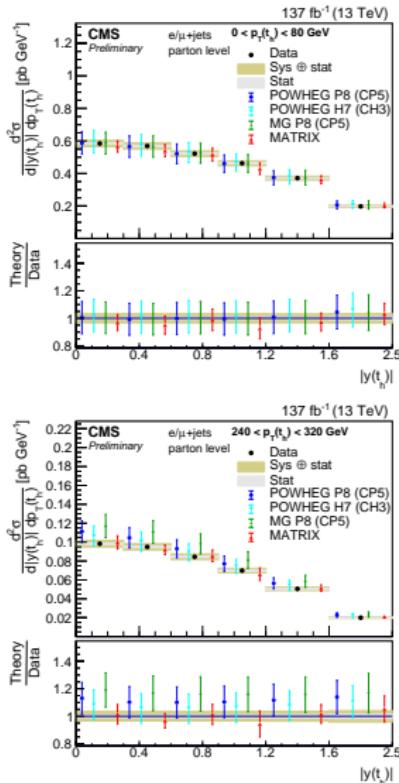


- $p_T$  and  $S_T = p_T(t_h) + p_T(t_l)$  well described by NNLO calculation.
- trend of harder spectrum in NLO calculations disappears above 600 GeV

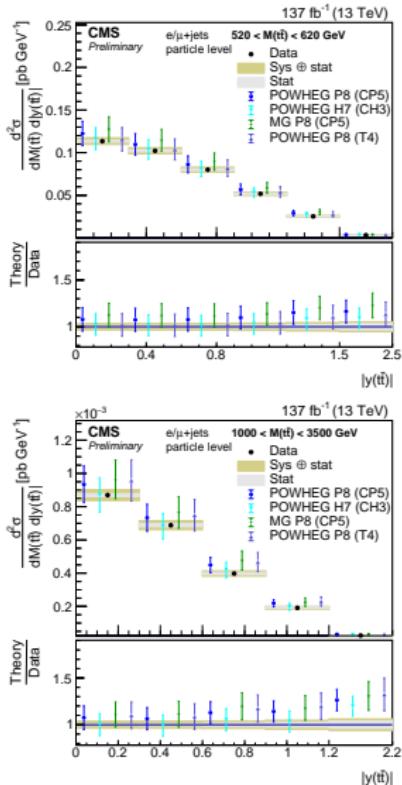
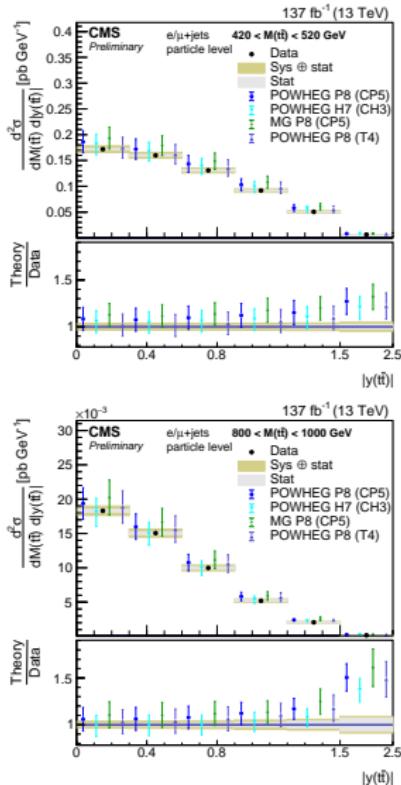
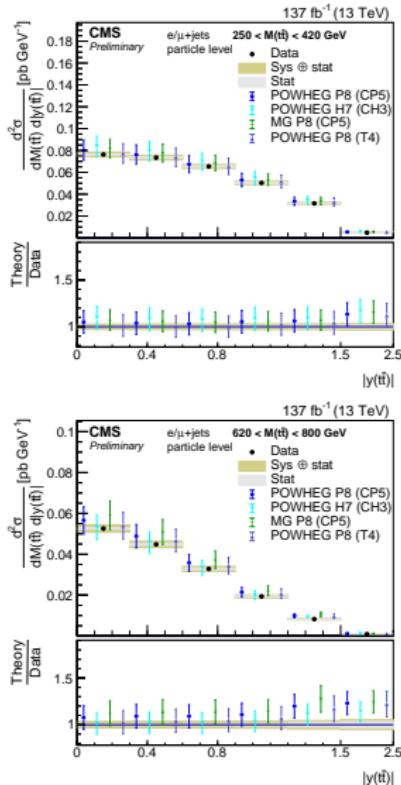


$-M(t\bar{t})$ ,  $p_T(t\bar{t})$ ,  $|y(t\bar{t})|$ ,  $\Delta\phi_{t/\bar{t}}$ , and  $\cos(\theta^*)$  (angle between t and direction of  $t\bar{t}$  system in  $t\bar{t}$  rest-frame): distributions of  $t\bar{t}$  system are well predicted.

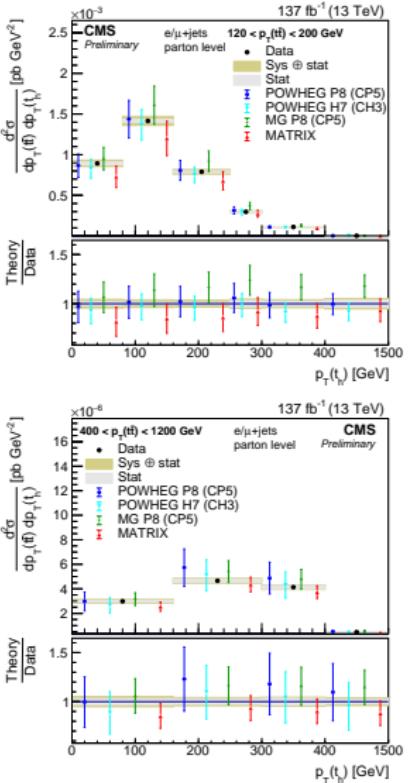
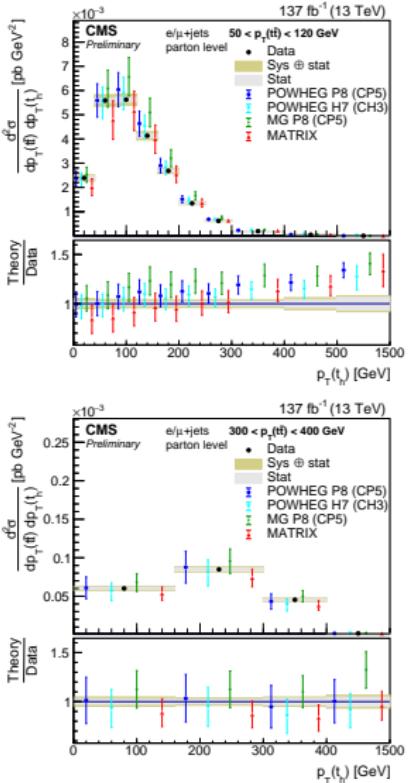
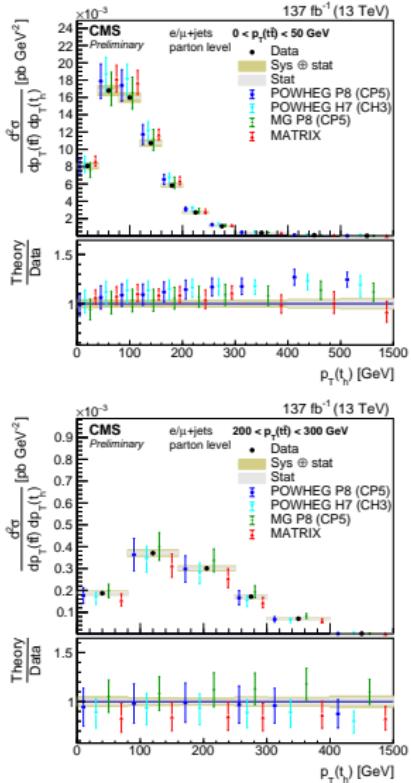
# Double differential cross sections



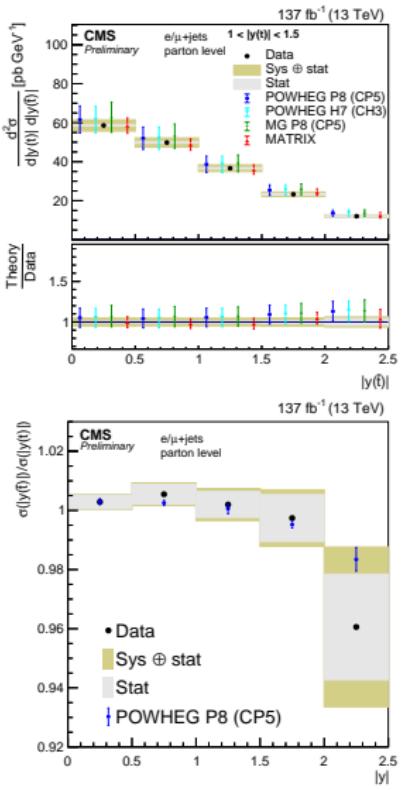
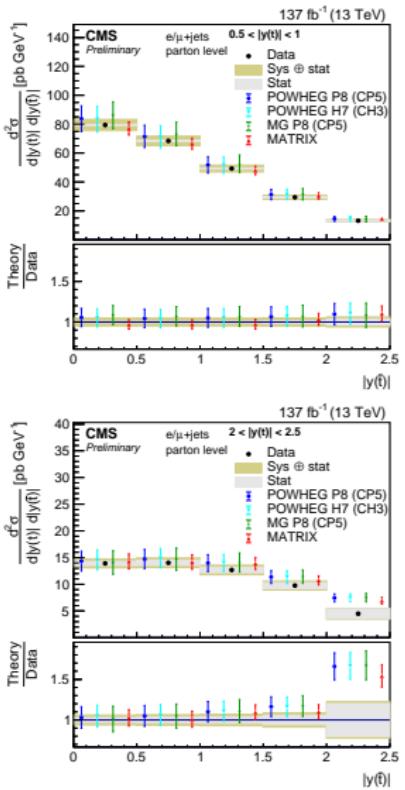
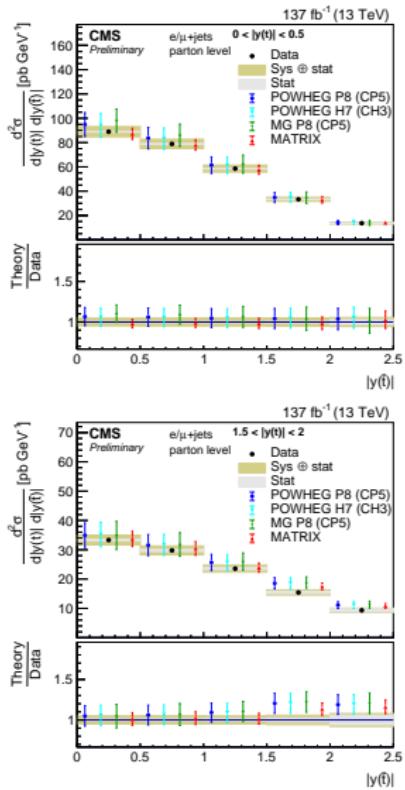
Well modeled  $|y(t_h)|$  in all  $p_T(t_h)$ -bins.



Well modeled  $|y(t\bar{t})|$  in all  $M(t\bar{t})$ -bins, with small overestimation in highest  $|y(t\bar{t})|$ -bins.



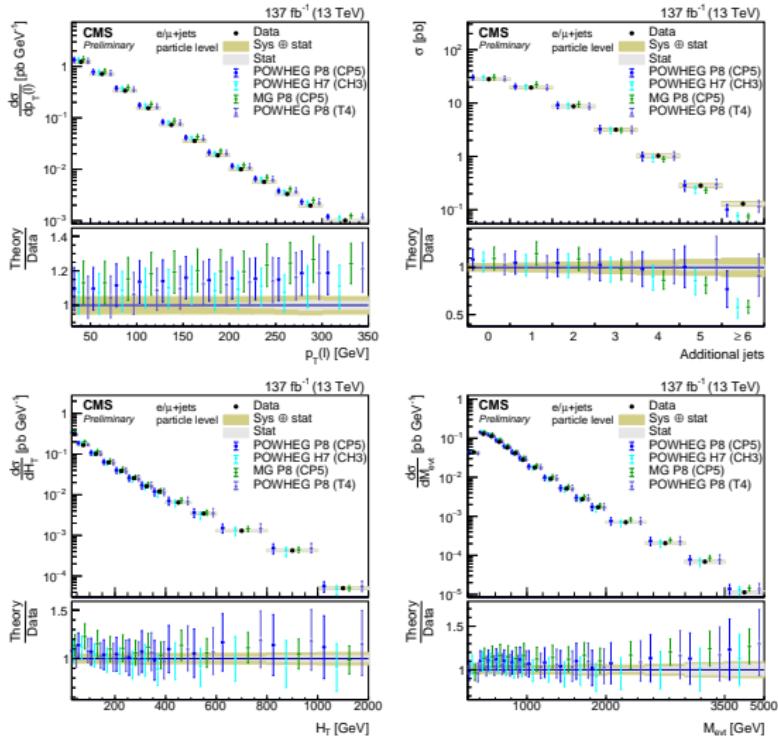
Mismodeling of spectrum in low  $p_T(t\bar{t})$ -bins, while the description at high  $p_T(t\bar{t})$  is better.



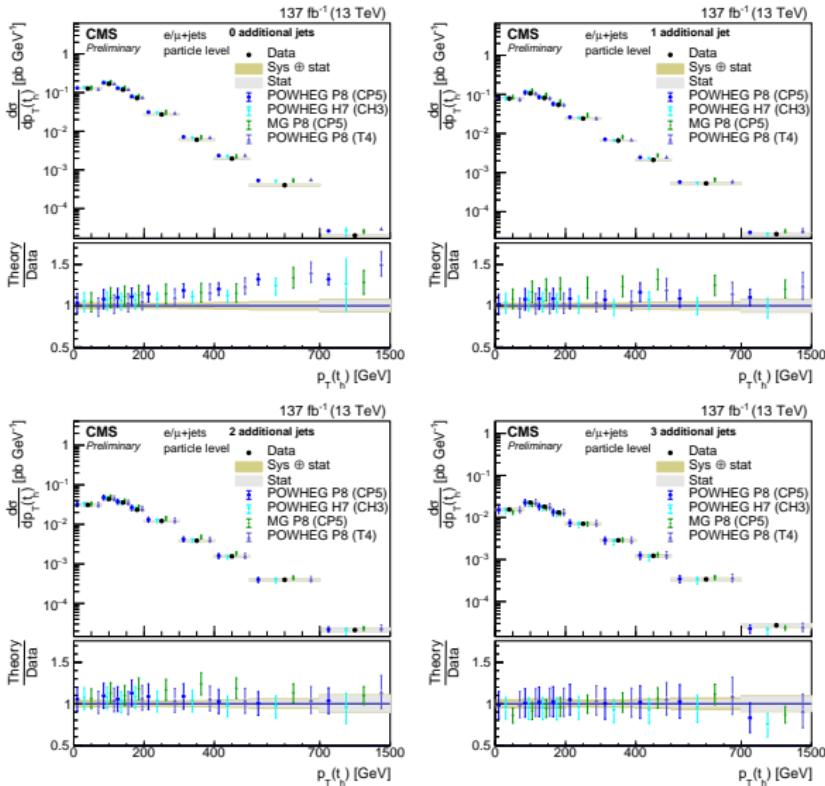
From the double differential cross section  $|y(t)|$  vs  $|y(\bar{t})|$  the ratio  $|y(\bar{t})|/|y(t)|$  is calculated taking into account bin-by-bin correlations.

Although this ratio can be measured very precisely, no significant observation of a charge asymmetry is expected and observed.

# Particle level results



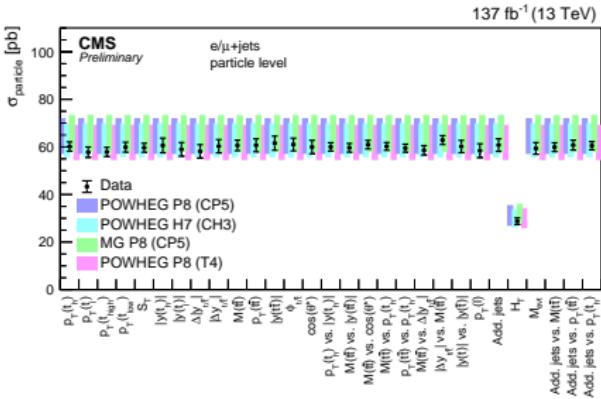
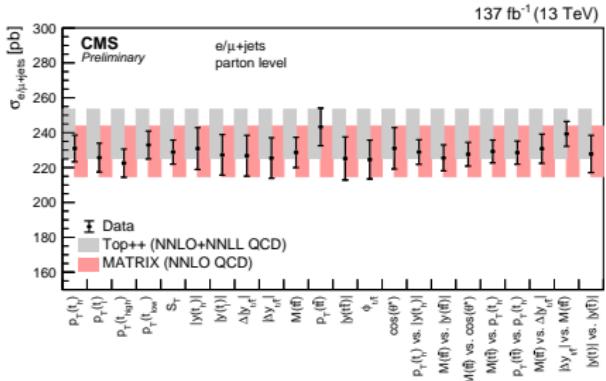
- $H_T$  scalar  $p_T$  sum of the additional jets ( $p_T > 30 \text{ GeV}$ )
- $M_{\text{evt}}$  invariant mass of the top quarks and the additional jets



Only in events without an additional jet we observe a softer  $p_T$  spectrum.

# Inclusive Cross Section

Obtain inclusive cross section from integration of differential distributions:



## Parton level:

- Best (expected) measurement  $M(t\bar{t})$  vs.  $\cos(\theta^*)$ :  $\sigma_{e/\mu+\text{jets}} = 227.6 \pm 6.8 \text{ pb}$ .
- With a branching fraction of  $28.77 \pm 0.32\%$  to  $e/\mu+\text{jets}$ :

$$\sigma_{t\bar{t}} = 791 \pm 25 \text{ pb } (\pm 0.6(\text{stat}) \pm 21(\text{sys}) \pm 14(\text{lumi}) \text{ pb})$$

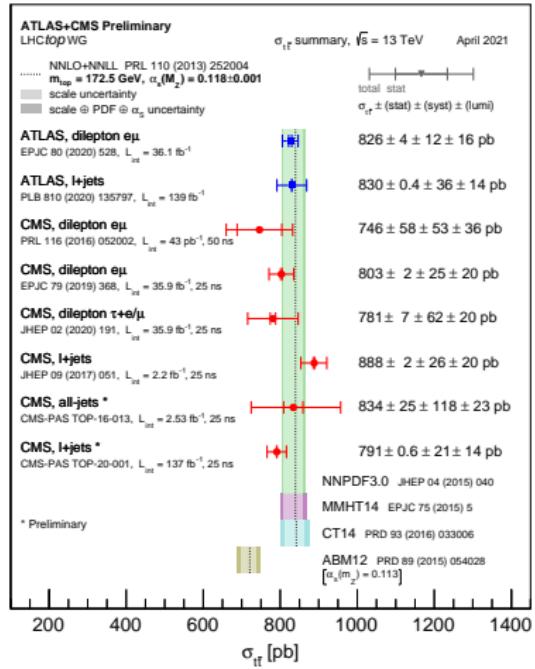
## Particle level:

$$\sigma_{\text{particle}} = 61.0 \pm 1.8 \text{ pb } (\pm 0.04(\text{stat}) \pm 1.5(\text{sys}) \pm 1.1(\text{lumi}) \text{ pb})$$

# Inclusive cross sections uncertainties

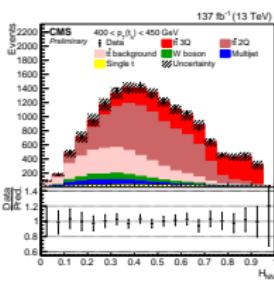
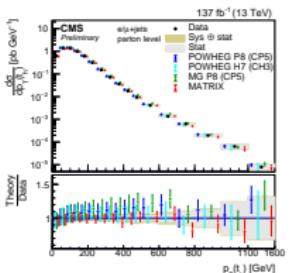
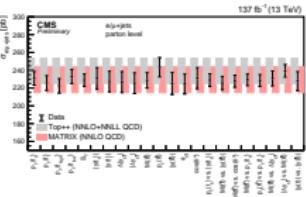
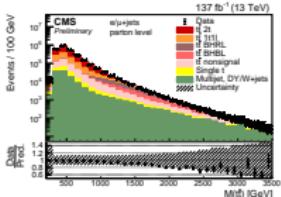
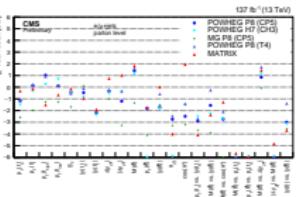
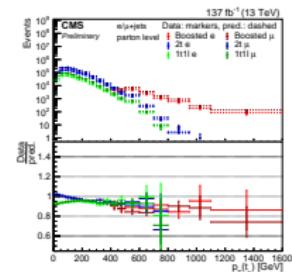
## Uncertainties in $\sigma_{t\bar{t}}$

Source	Uncertainty [pb] (%)
Jet energy	10.9 (1.38)
Branching fraction	8.80 (1.11)
Lepton	7.78 (0.98)
NNLO	7.56 (0.96)
b tagging	6.96 (0.88)
Sim. event count	6.46 (0.82)
Background	6.10 (0.77)
CR model	5.45 (0.69)
Jet energy resolution	3.36 (0.43)
Scales $\mu_r/\mu_f$	3.24 (0.41)
Initial-state PS scale	3.19 (0.40)
Final-state PS scale	2.71 (0.34)
Subjet energy	2.42 (0.31)
b mistagging	2.20 (0.28)
UE tune	2.16 (0.27)
$m_t$	2.08 (0.26)
PDF	1.94 (0.25)
$h_{damp}$	1.51 (0.19)
L1 trigger	0.53 (0.07)
Pileup	0.40 (0.05)
Syst	21.1 (2.66)
Stat	0.56 (0.07)
Int. luminosity	13.8 (1.75)



With 3.2% uncertainty one of the most precise  $t\bar{t}$  cross section measurements.

## Summary



- high precision measurements of differential and double differential cross sections of  $t\bar{t}$  production
  - first combination of resolved and boosted reconstruction → full spectrum measurement
  - combined fit of cross sections in several categories reduces systematic uncertainties
  - results at parton and particle levels
  - precise (3.2% unc.) inclusive  $t\bar{t}$  production cross sections:

$$\sigma_{t\bar{t}} = 791 \pm 25 \text{ pb} (\pm 0.6(\text{stat}) \pm 21(\text{sys}) \pm 14(\text{lumi}) \text{ pb})$$

# BACKUP

**Objects:**

- exactly one dressed muons or electrons, not from hardron decay with  $p_T > 30 \text{ GeV}$  and  $|\eta| < 2.4$
- $p_\nu$ :  $p_T^{\text{miss}}$  taken as sum of all  $\nu$ . Calculated using detector level algorithm
- jets are clustered with a distance parameter of 0.4. Jets with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 2.4$  are selected if there is no lepton within  $\Delta R = 0.4$
- b jets are jets containing a b hadron (ghost matching)
- boosted  $t_l$  are b jets with a selected lepton within  $\Delta R < 0.4$ ,  $p_T(\ell) > 50 \text{ GeV}$ . The momentum is calculated as sum of jet, lepton, and  $p_\nu$ .  $p_T > 380 \text{ GeV}$ ,  $|\eta| < 2.4$ , and  $100 < M(t_l) < 240 \text{ GeV}$  are required
- boosted  $t_h$  are b jets with a distance parameter of 0.8,  $M_{\text{jet}} > 120 \text{ GeV}$ ,  $p_T > 380 \text{ GeV}$ , and  $|\eta| < 2.4$

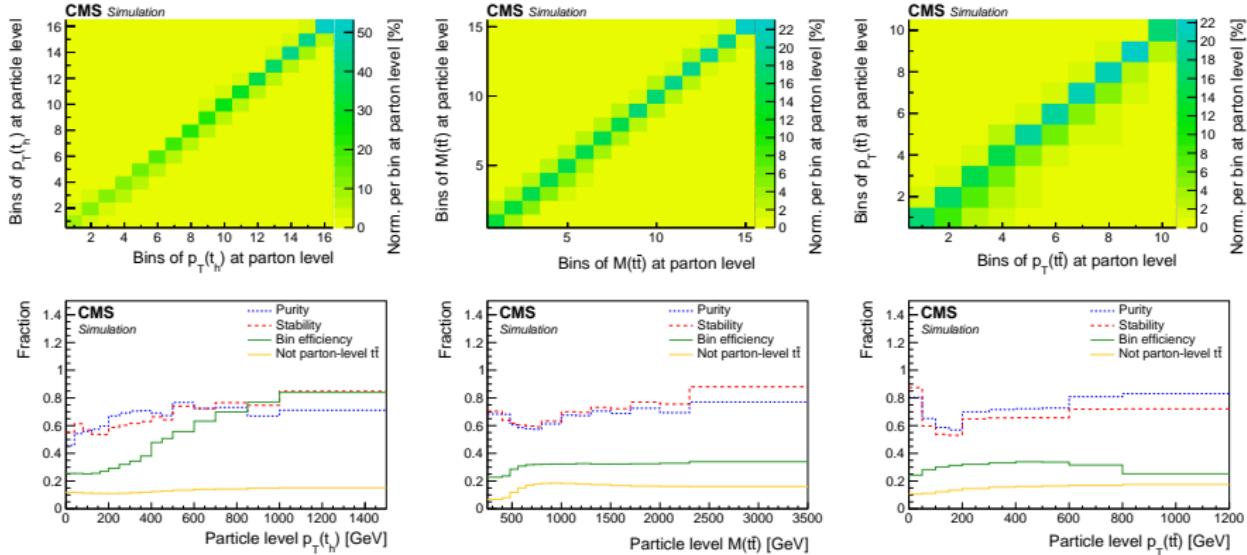
 **$t\bar{t}$  event reconstruction (by priority):**

- ❶ boosted  $t_l +$  boosted  $t_h$
- ❷ minimize:

- $[M(p_\nu + p_\ell + p_{b_\ell}) - m_t]^2 + [M(p_{j_W1} + p_{j_W2}) - m_W]^2 + [M(p_{j_W1} + p_{j_W2} + p_{b_h}) - m_t]^2$
- $[M(p_\nu + p_\ell + p_{b_\ell}) - m_t]^2 + [M_{\text{jet}} - m_t]^2$

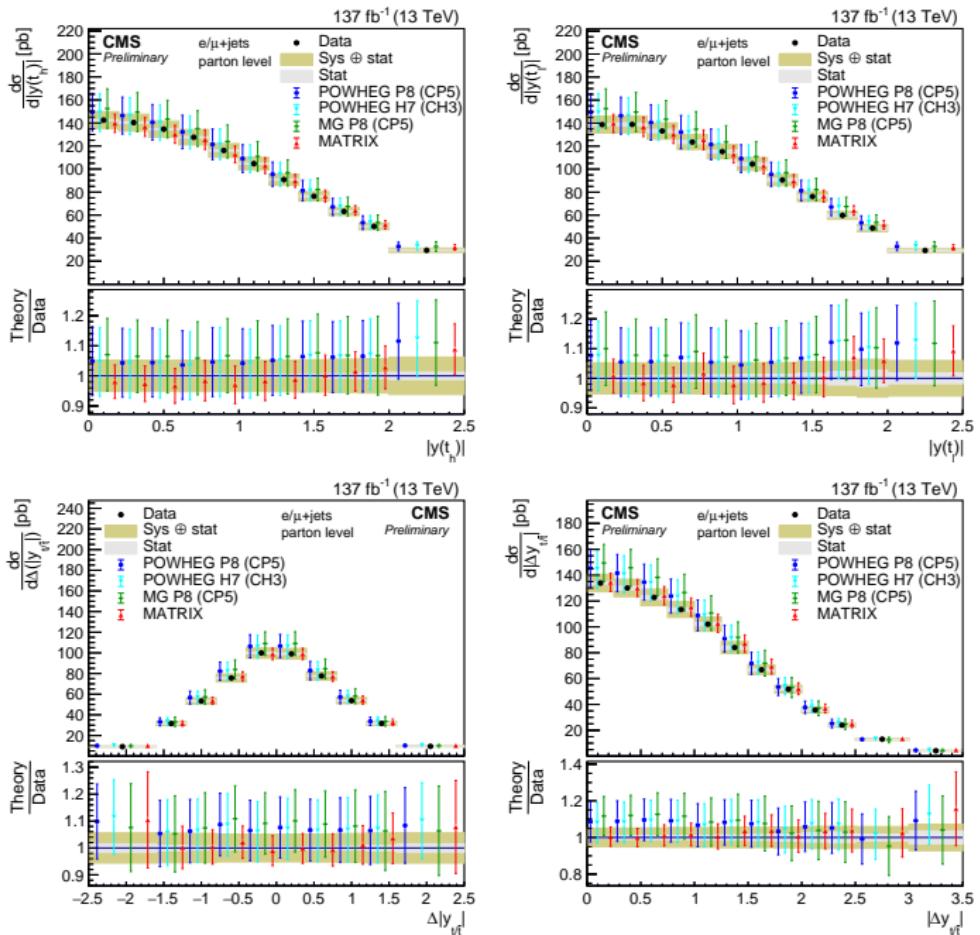
$100 < M(t_h) < 240 \text{ GeV}$  required. If both possible, select reconstruction with  $M(t_h)$  closer to  $m_t$

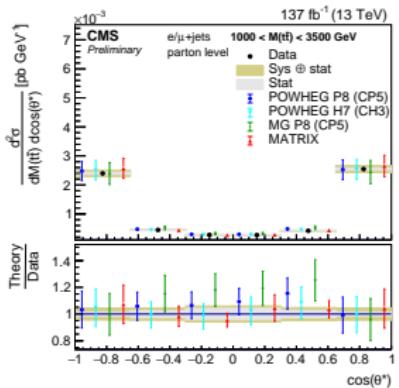
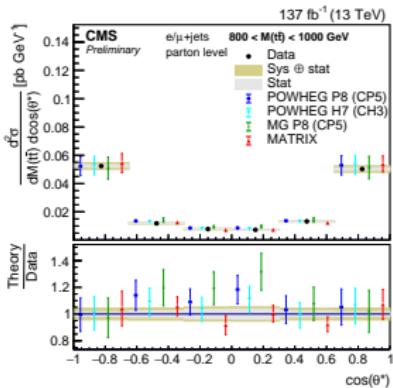
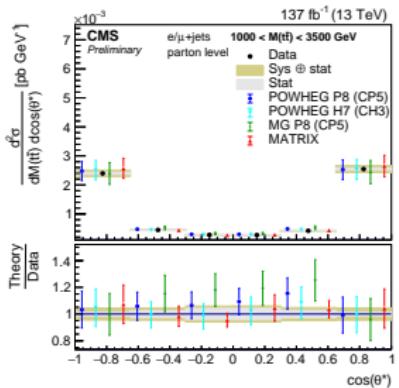
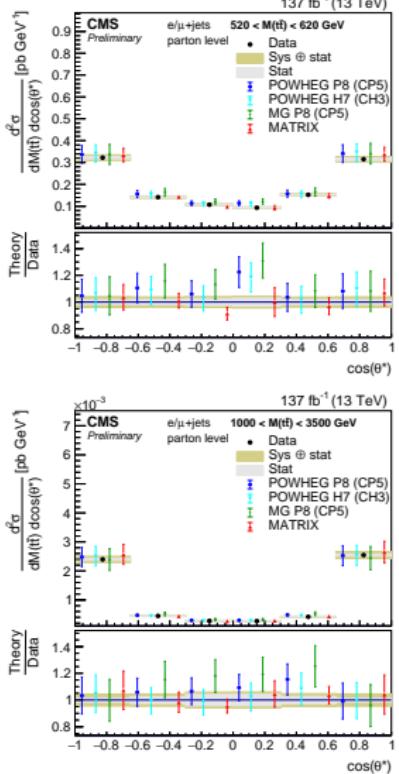
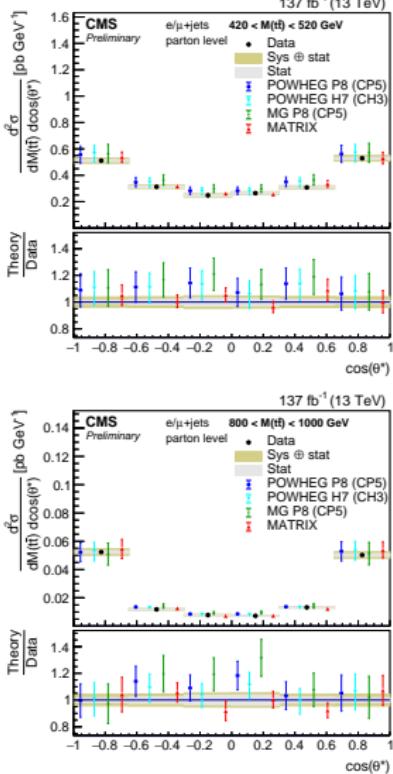
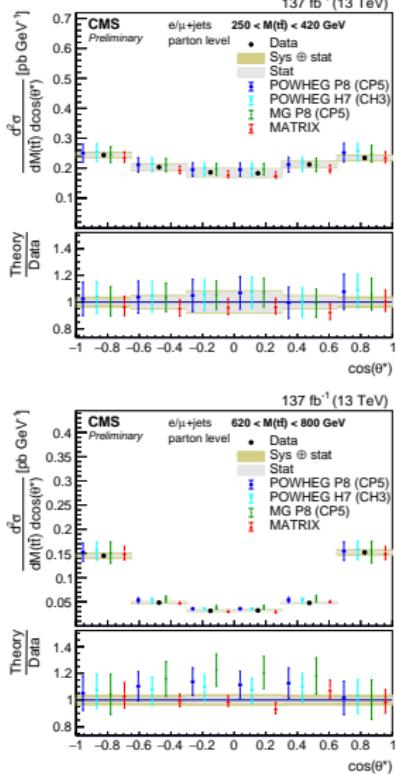
# Parton vs. Particle levels

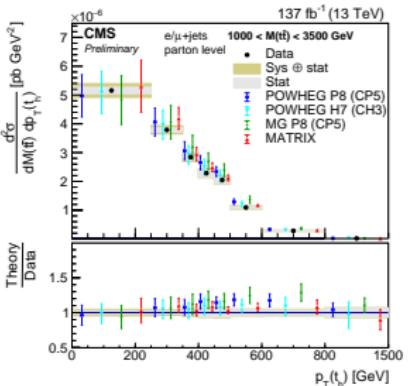
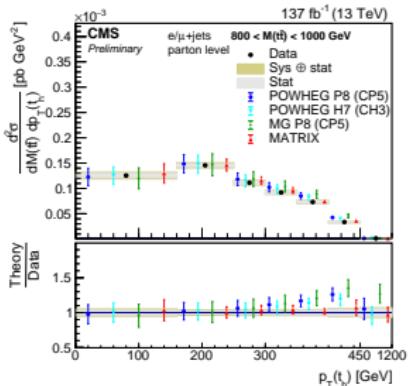
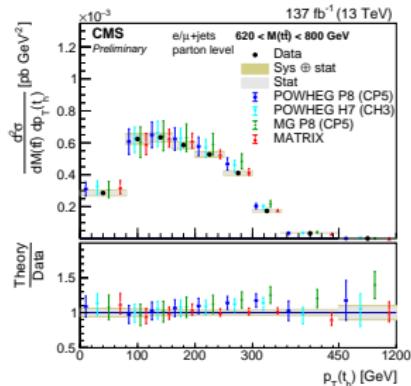
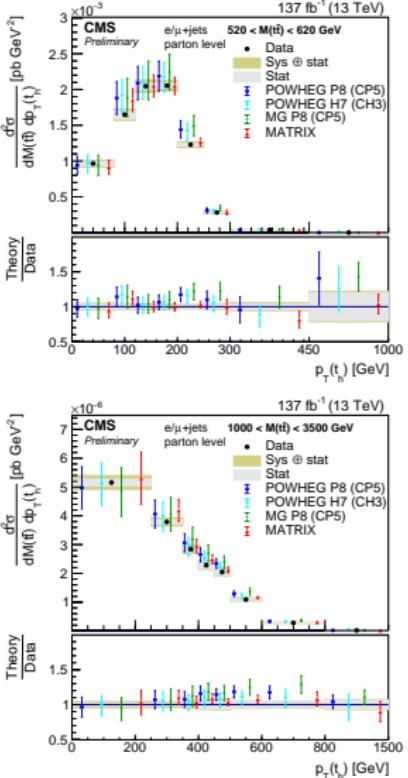
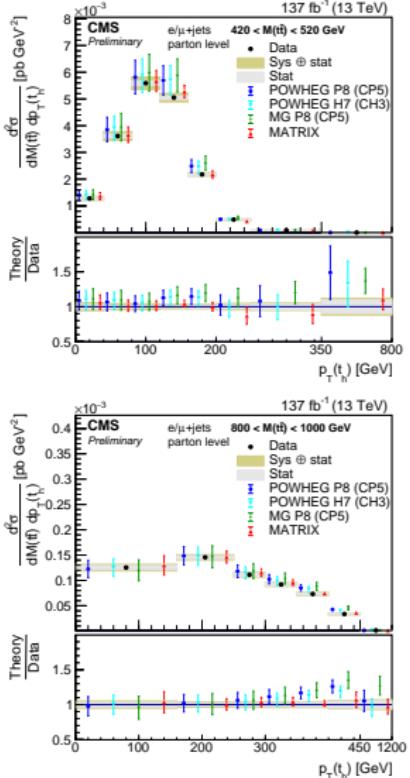
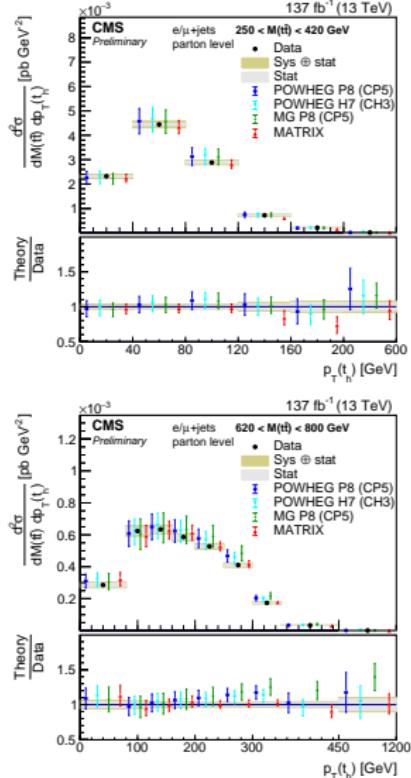


- Particle level shows strong correlation with parton level top quarks.

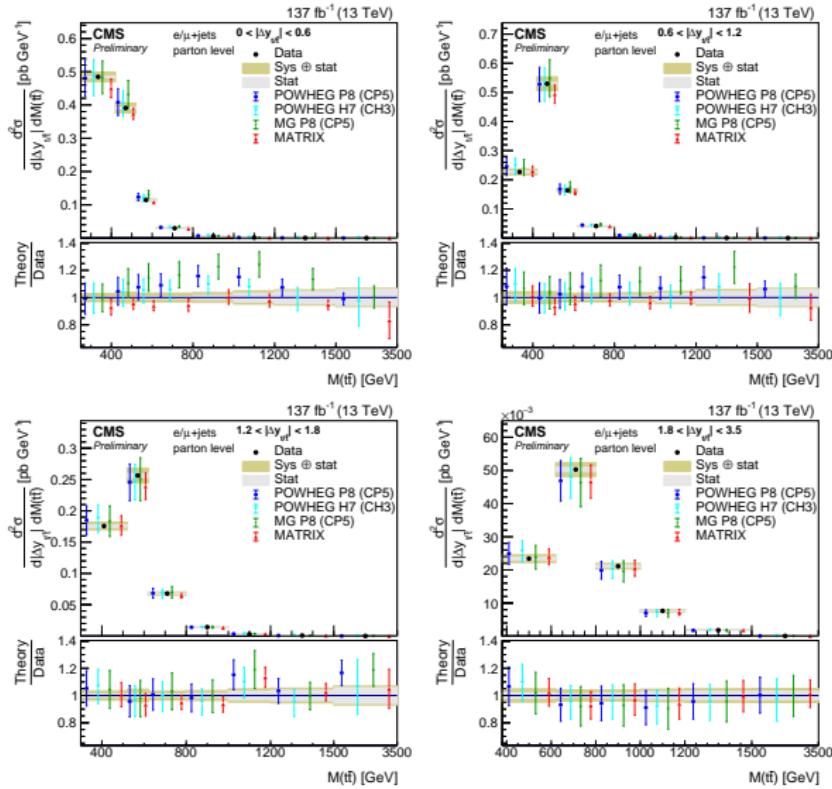
- From all combinations of two subjets calculate the invariant mass and use as input variables the three masses with highest values.
- The number of combinations of two subjets with invariant masses exceeding 40 GeV.
- From all combinations of three subjets calculate the invariant mass and use the two masses with highest values.
- The ratio of the highest invariant mass of three subjets over the invariant mass of all constituents.
- The ratios of  $N$ -jettiness  $\tau_2/\tau_1$ ,  $\tau_3/\tau_2$ ,  $\tau_4/\tau_3$ , and  $\tau_5/\tau_4$  with  $\tau_N = \sum_k \min(q_1 \cdot p_k, q_2 \cdot p_k, \dots, q_N \cdot p_k)$ , where  $q_i$  with  $1 \leq i \leq N$  are the momenta of the  $N$  leading subjets and  $p_k$  are the momenta of all constituents in the rest frame.
- The energy of the four most energetic subjets.
- The value of  $|(\mathbf{p}_1 \times \mathbf{p}_2) \cdot \mathbf{p}_3|$ , where  $\mathbf{p}_1$ ,  $\mathbf{p}_2$ , and  $\mathbf{p}_3$  are the three momenta of the most energetic subjets normalized to unity.
- The sphericity  $s$  of all subjets, where  $s = \frac{3}{2}(\lambda_2 + \lambda_3)$  with  $\lambda_2$  and  $\lambda_3$  the second and third highest eigenvalues of the tensor  $S^{\alpha\beta} = \frac{\sum_i p_i^\alpha p_i^\beta / |p_i|}{\sum_i |p_i|}$ , where  $p_i$  are the momenta of the subjets in the rest frame and  $\alpha, \beta$  are the spatial indices.
- Boost the three leading subjets back to the laboratory rest frame and calculate their momentum fractions relative to the AK8 jet.

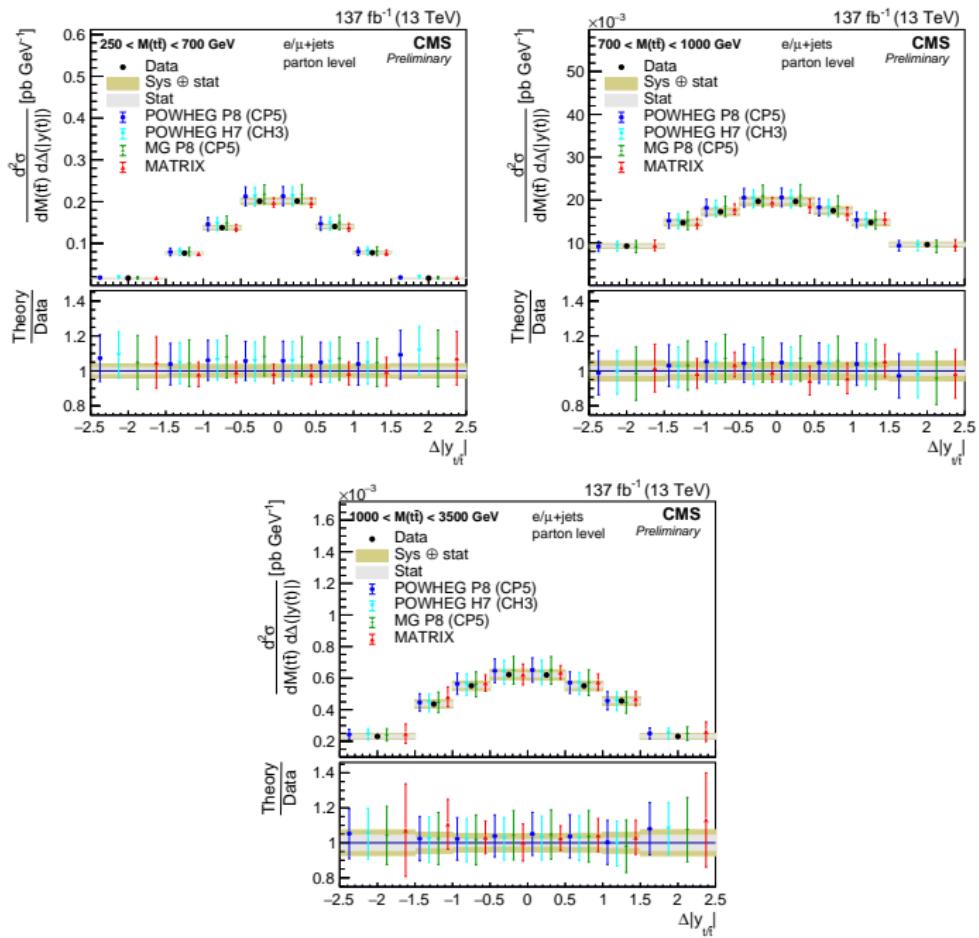




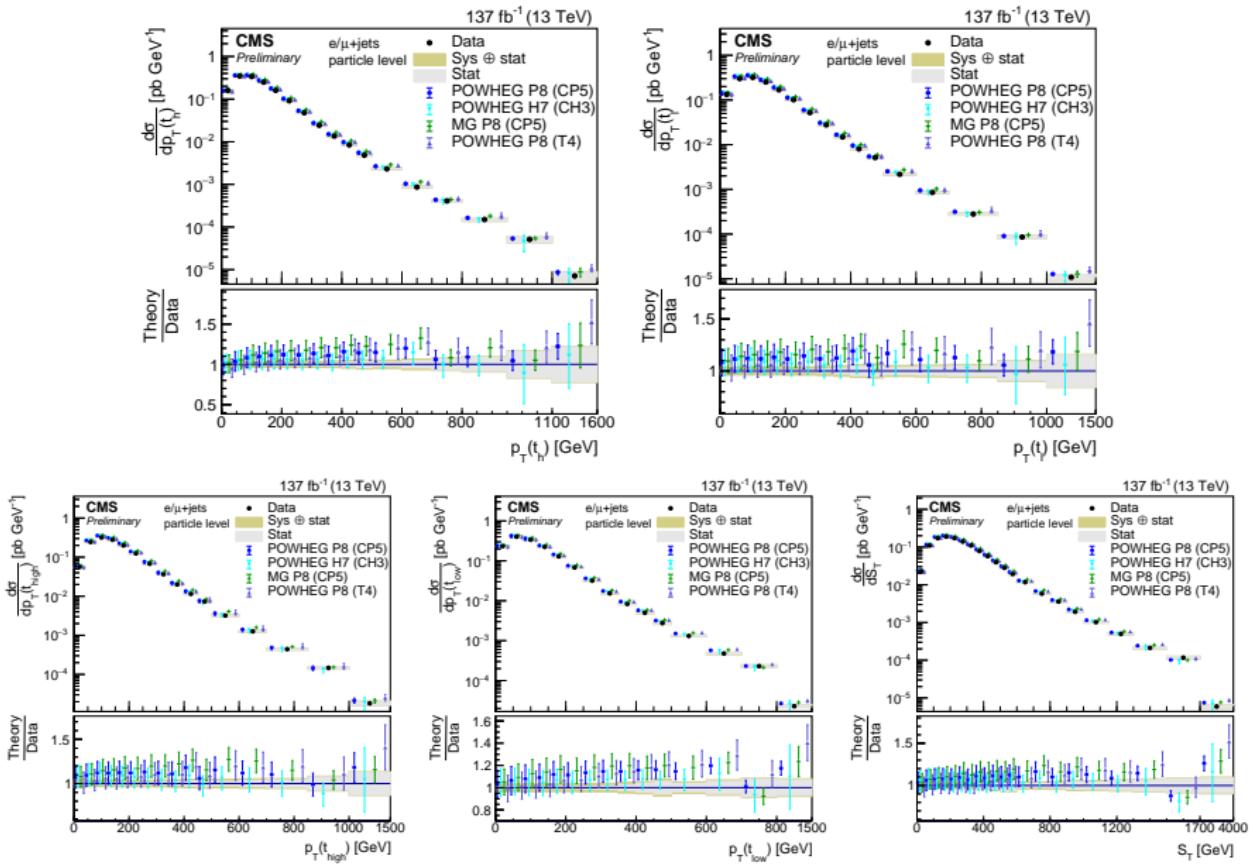


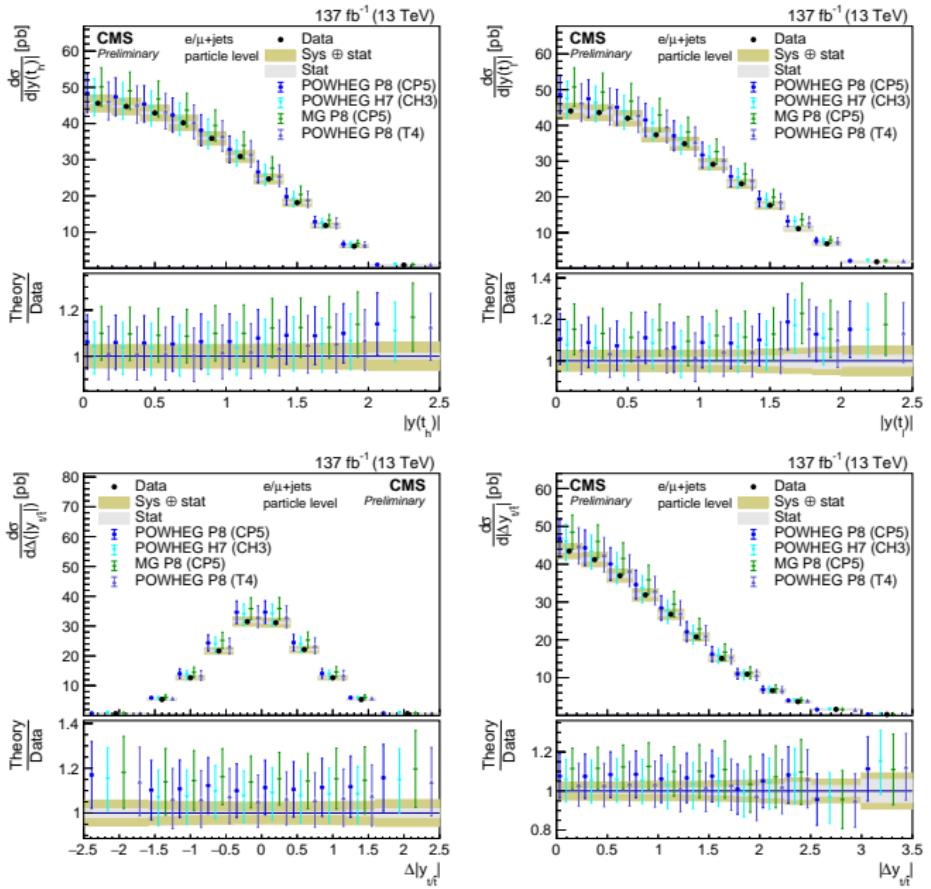
At high  $M(t\bar{t})$  a harder  $p_T$  spectrum is predicted.  
 NNLO slightly better, but has softer spectrum at low  $M(t\bar{t})$ .

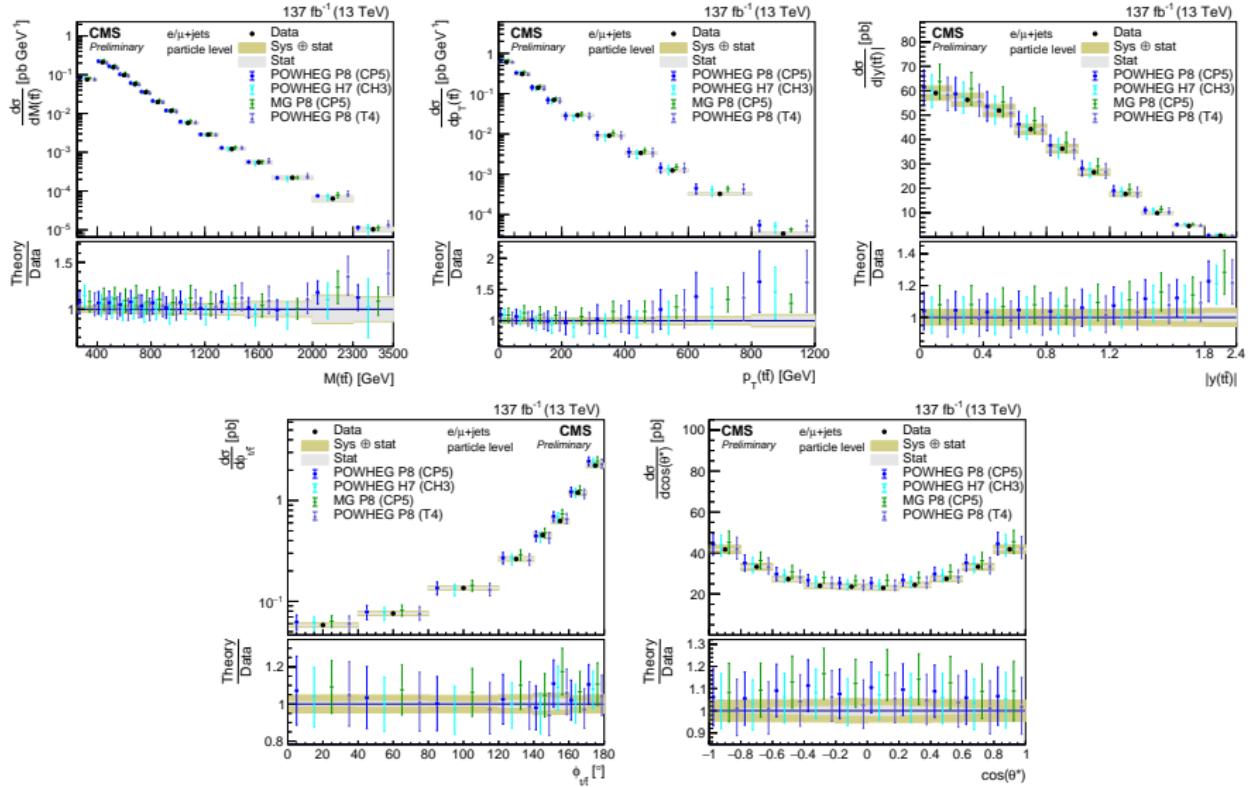




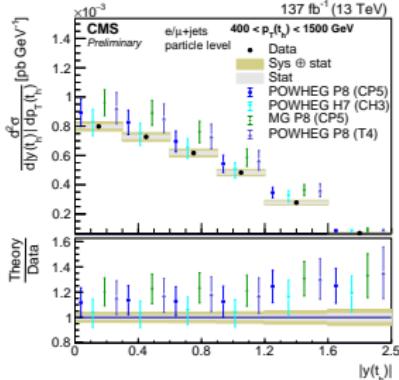
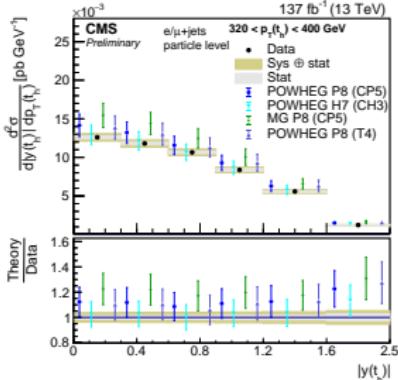
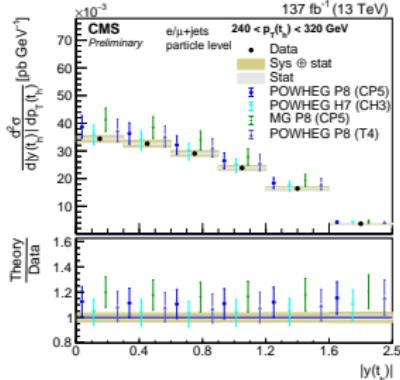
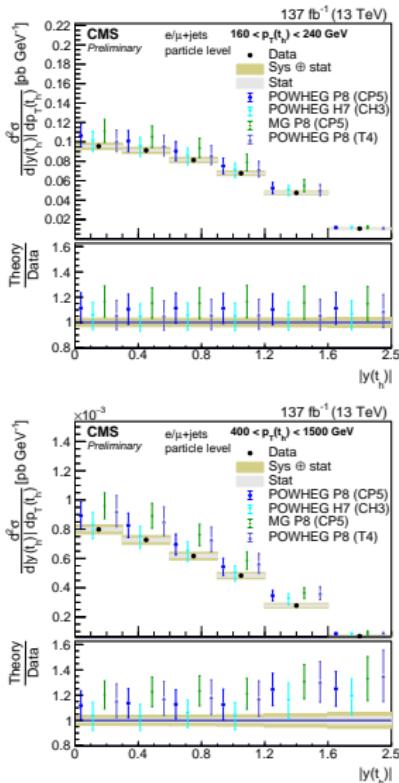
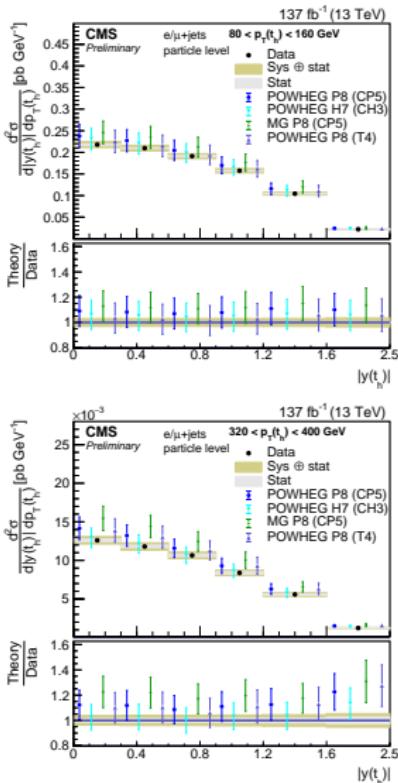
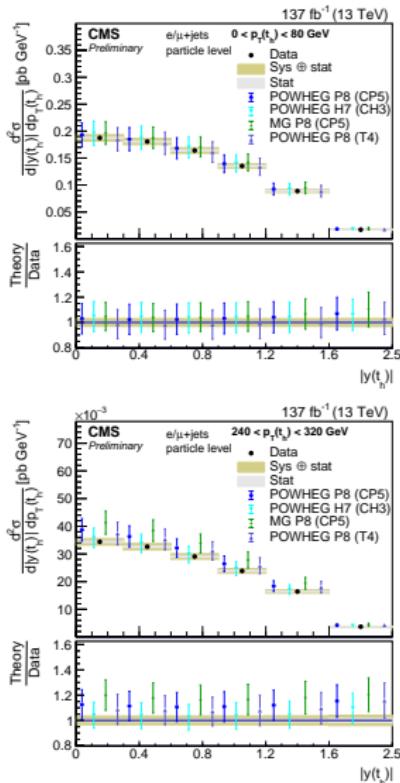
# Particle level cross sections

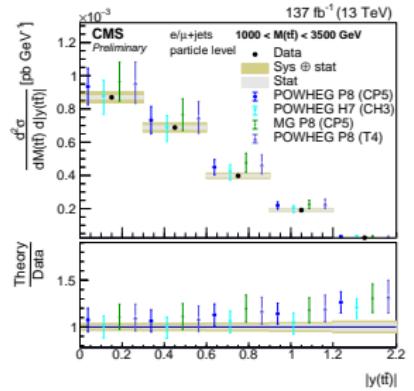
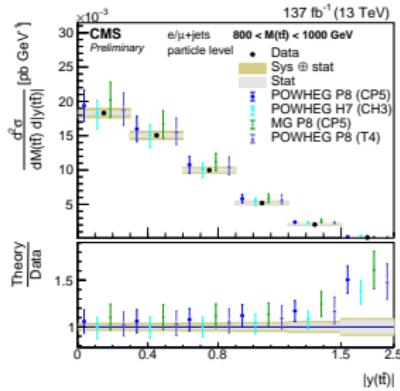
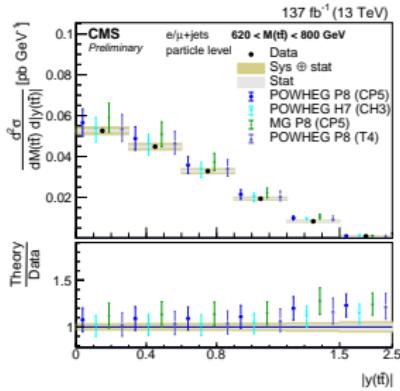
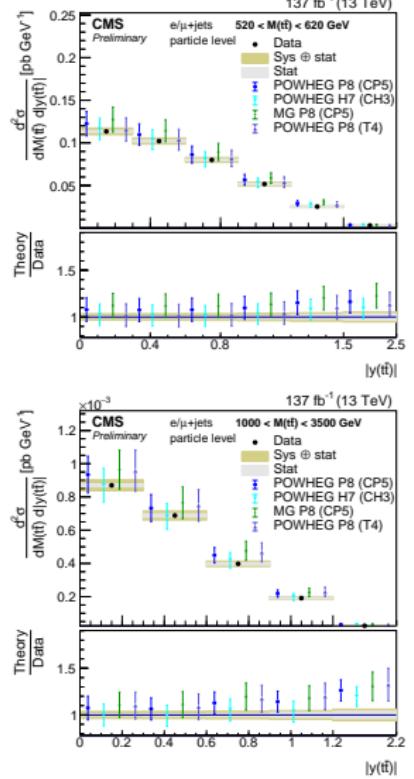
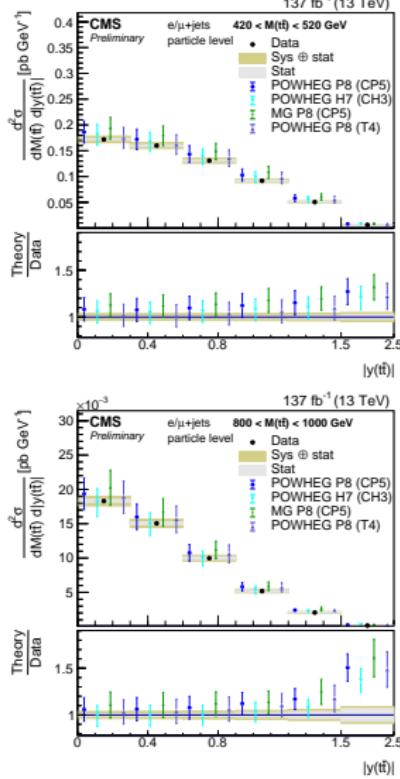
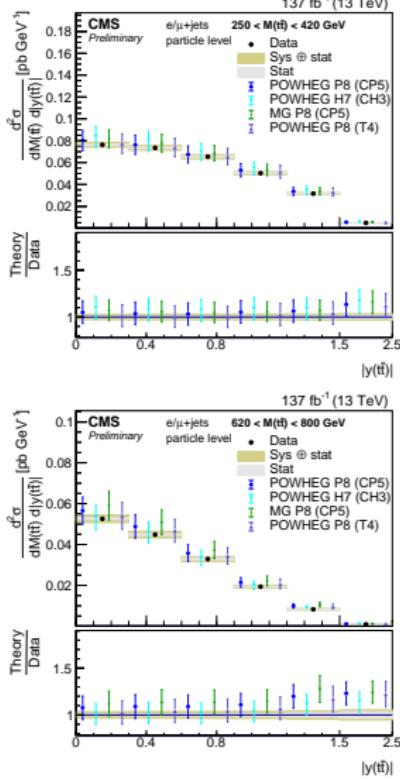


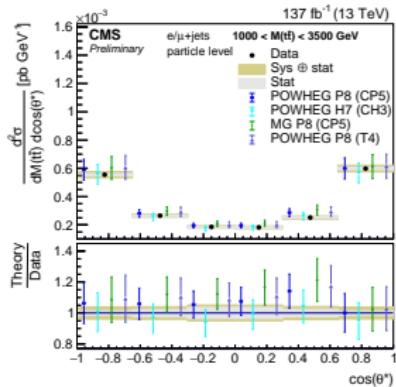
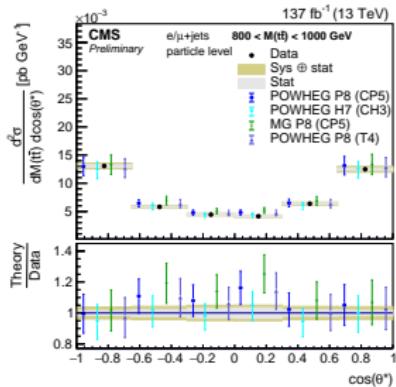
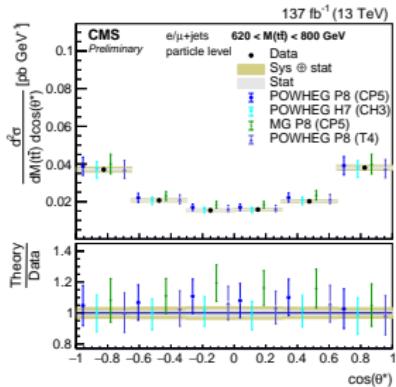
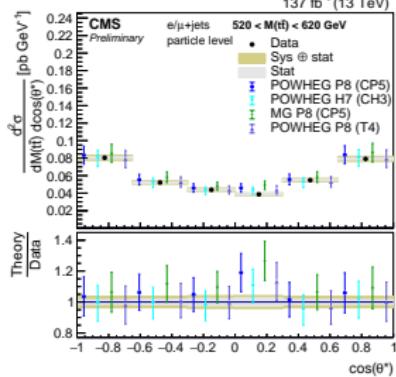
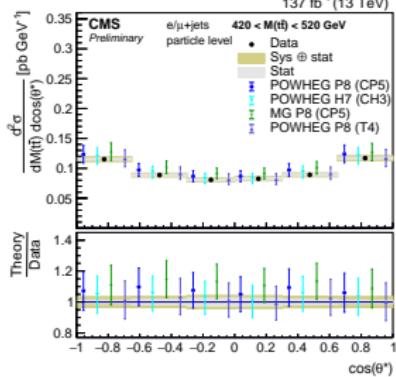
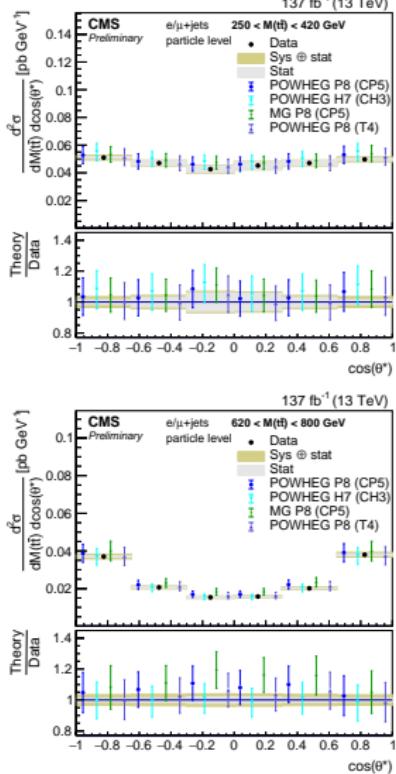


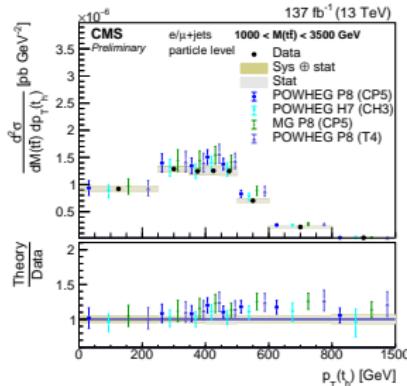
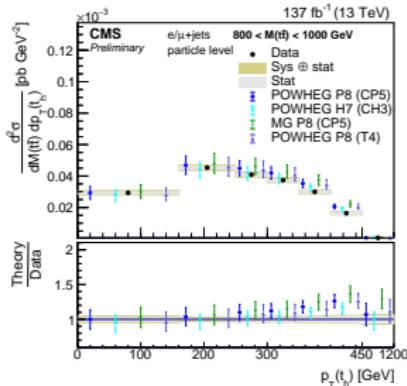
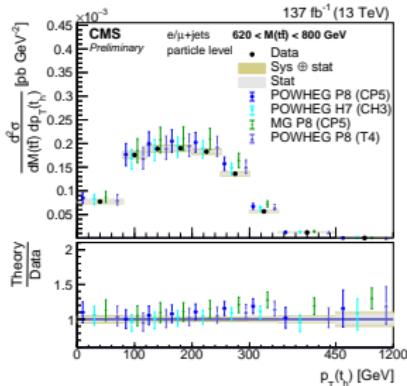
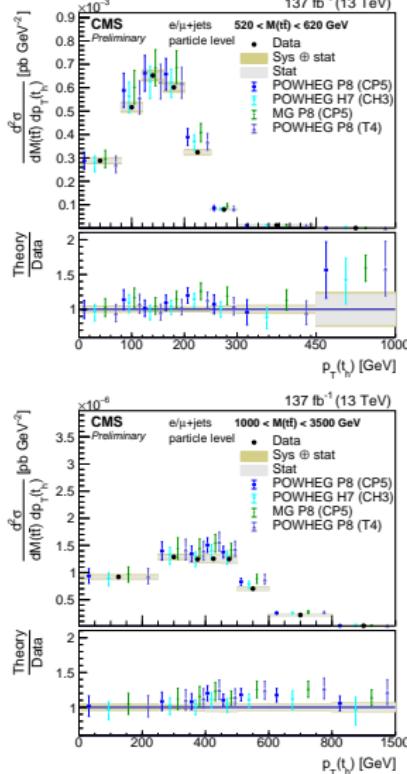
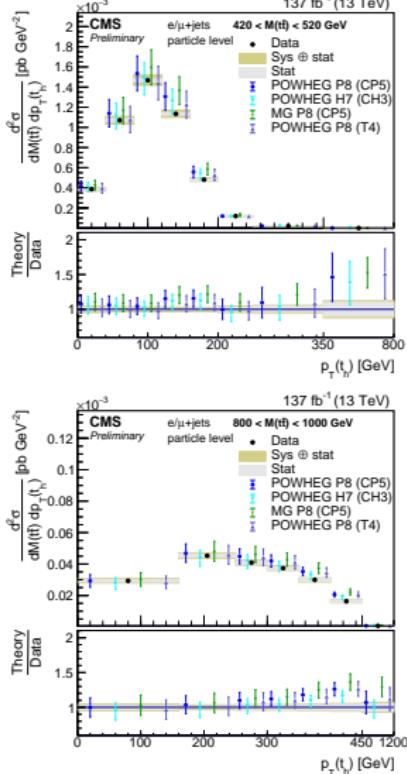
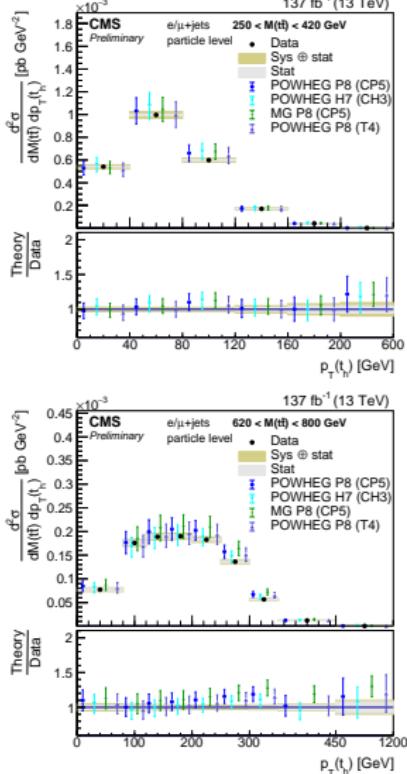


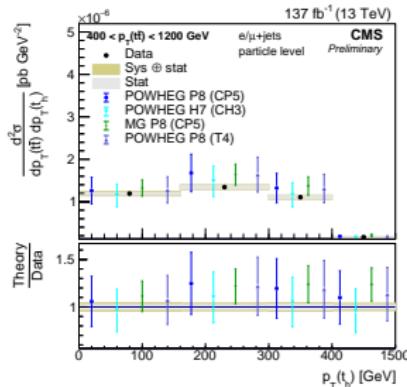
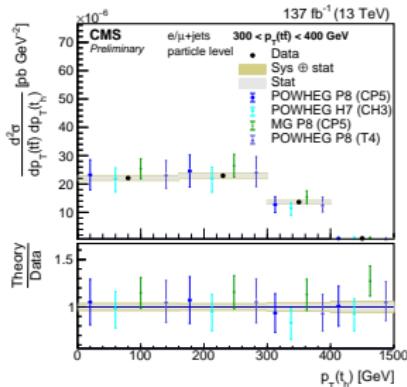
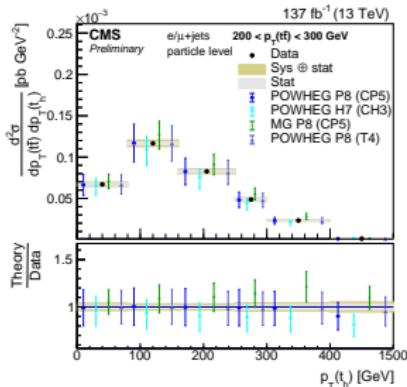
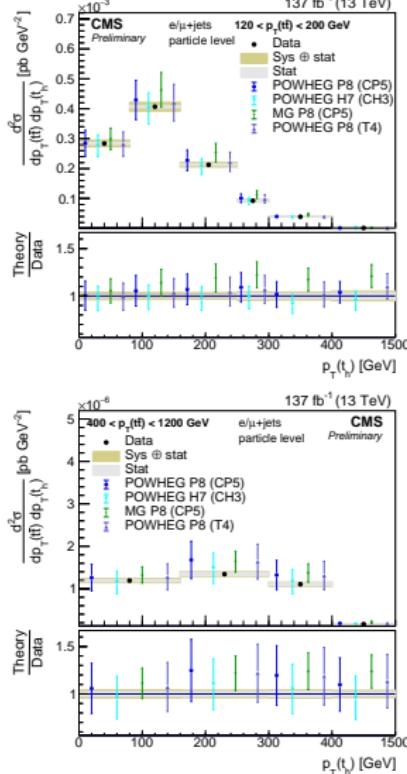
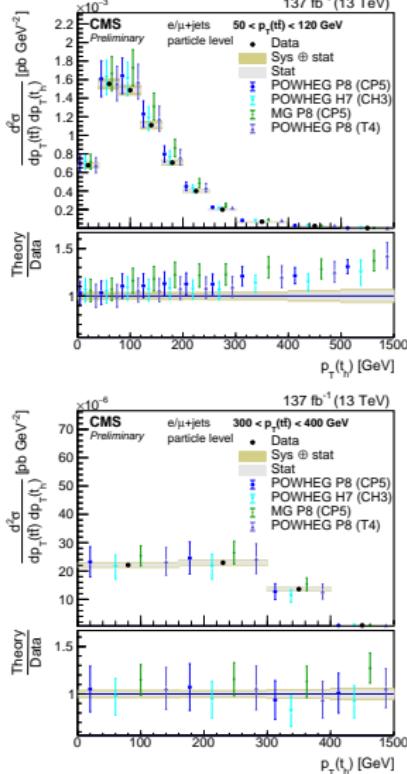
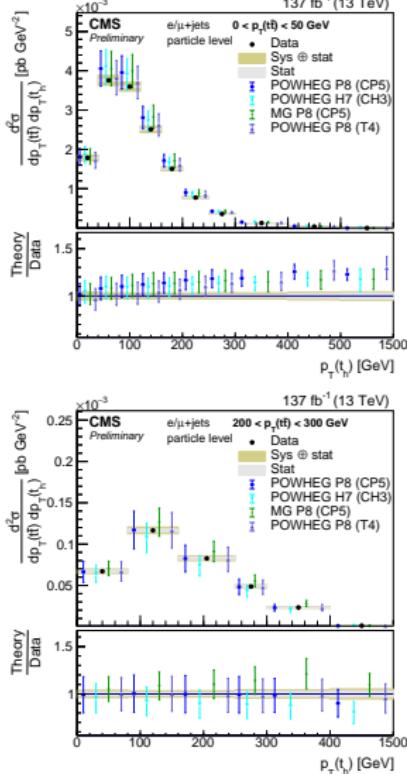
# Double differential cross sections

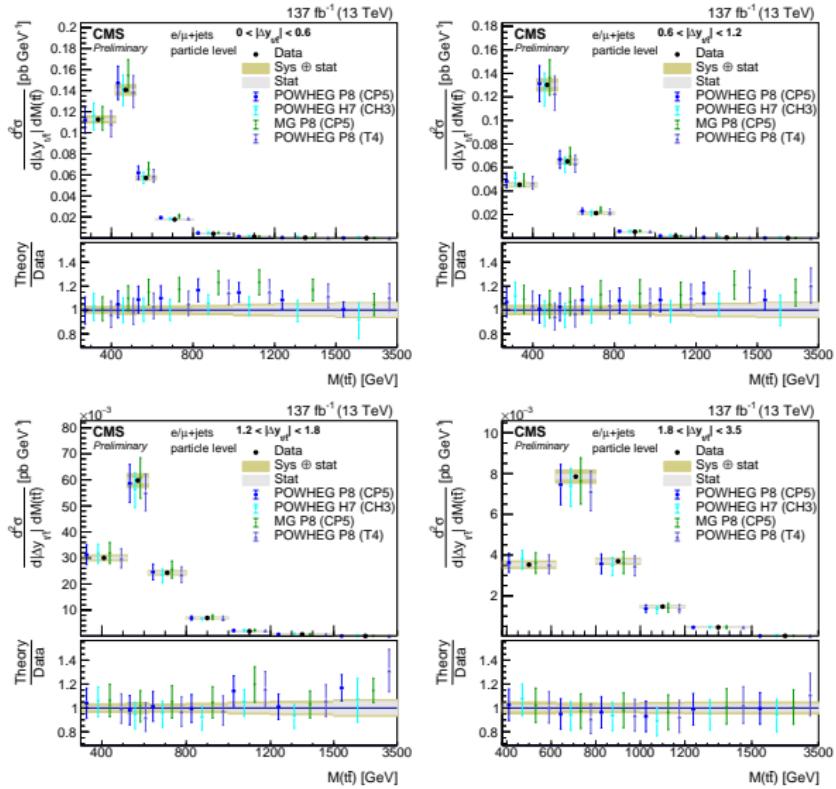


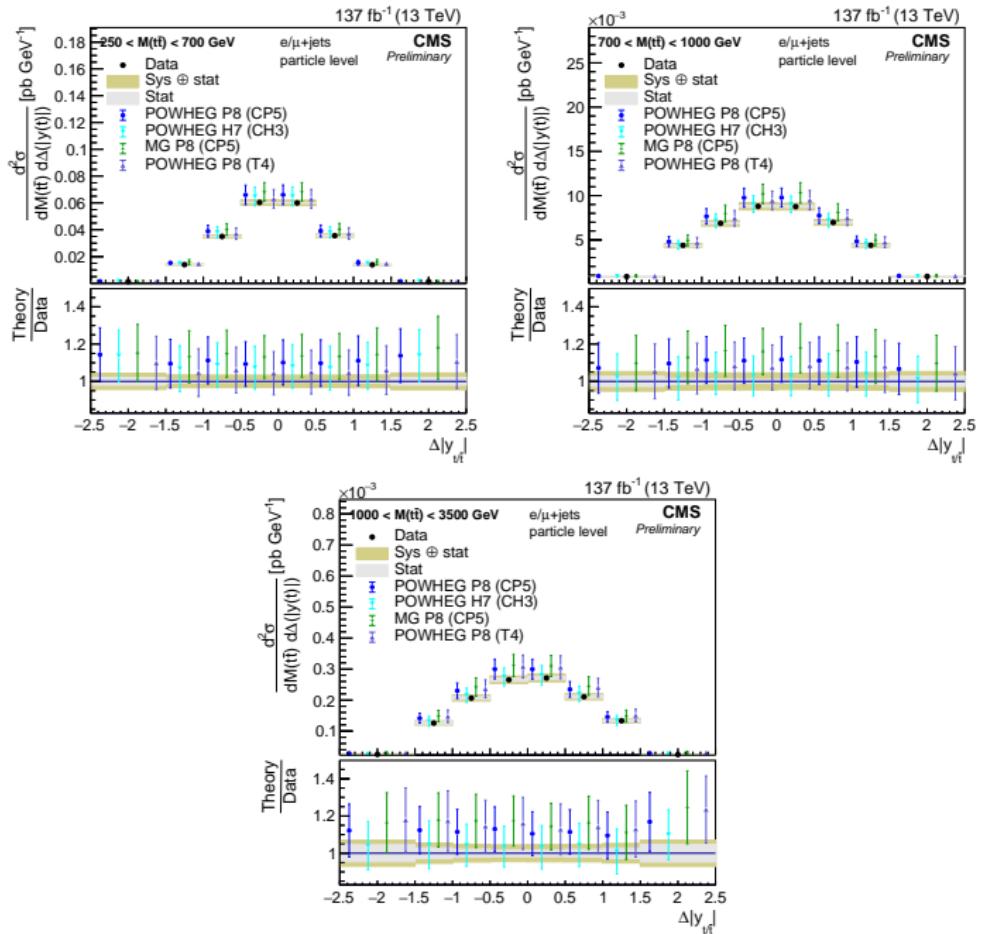


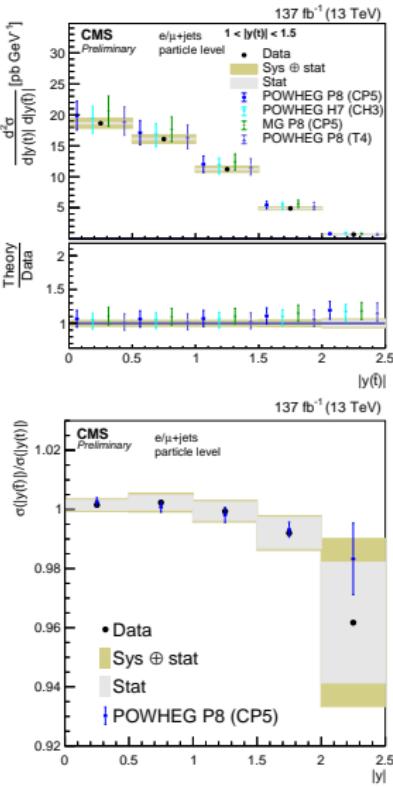
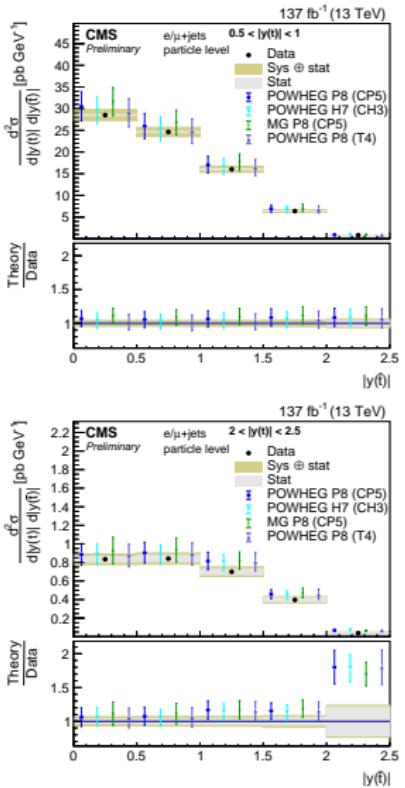
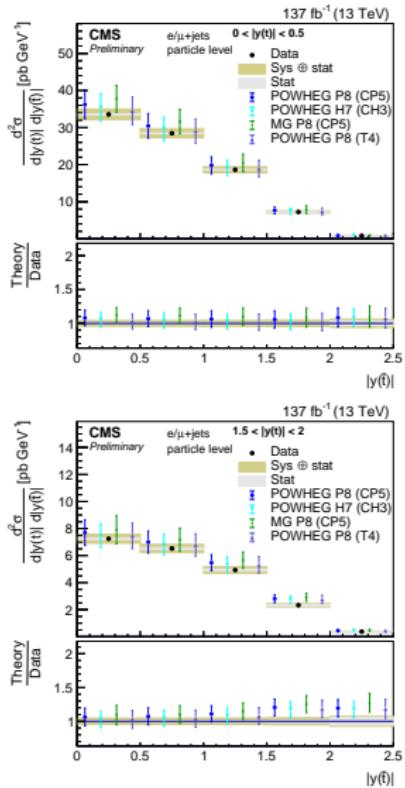






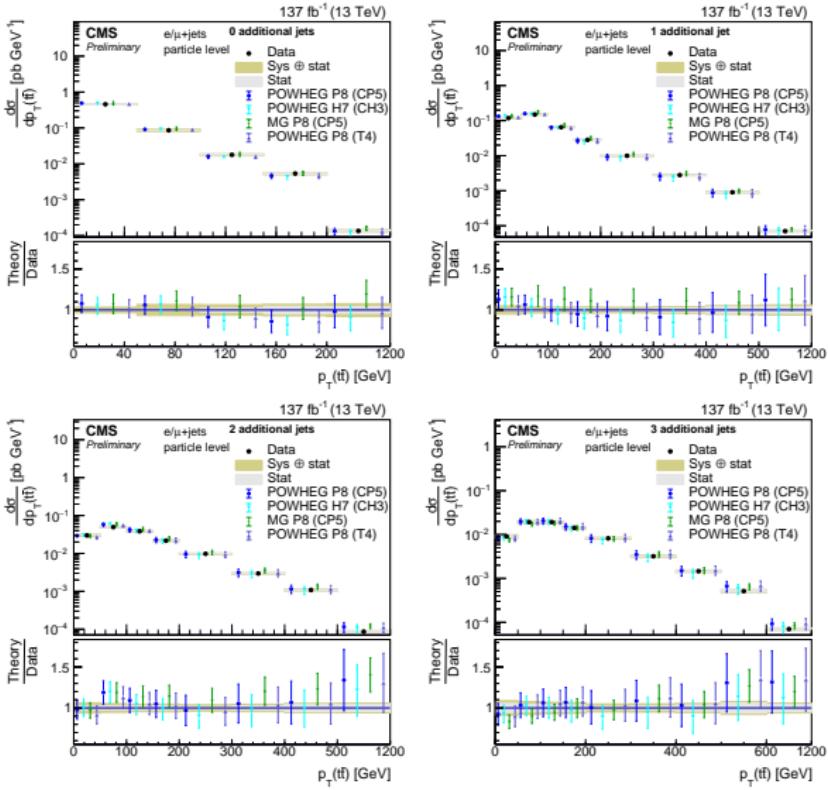


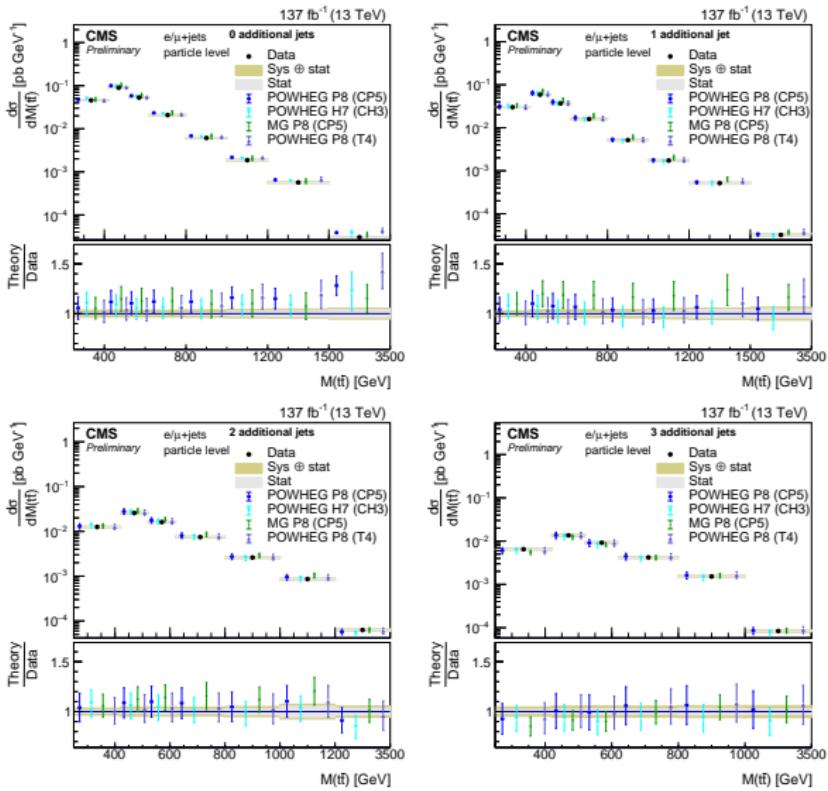




From the double differential cross section  $|y(t)|$  vs  $|y(\bar{t})|$  the ratio  $|y(\bar{t})|/|y(t)|$  is calculated taking into account bin-by-bin correlations.

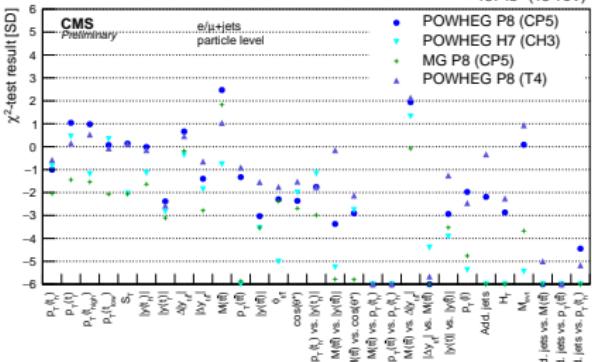
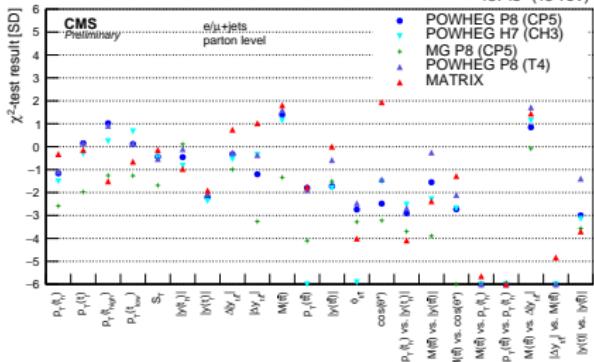
Though this ratio can be measured very precisely, no significant observation of a charge asymmetry is expected and observed.



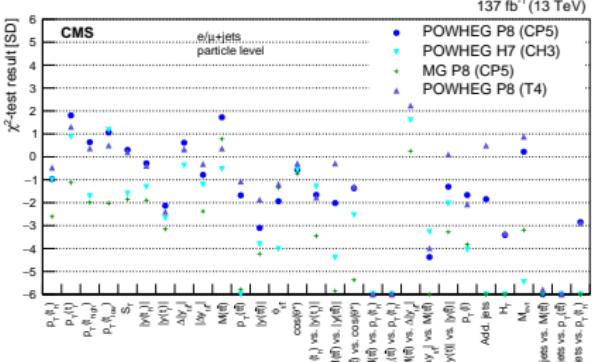
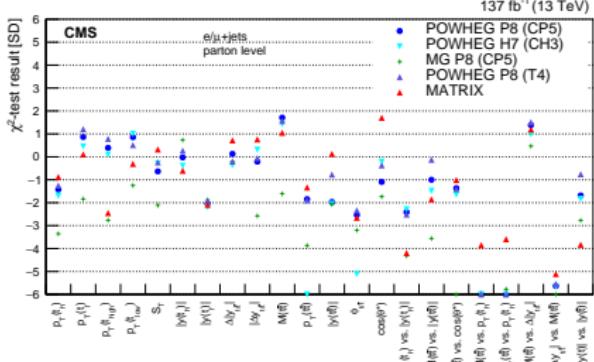


## Normalized differential cross sections

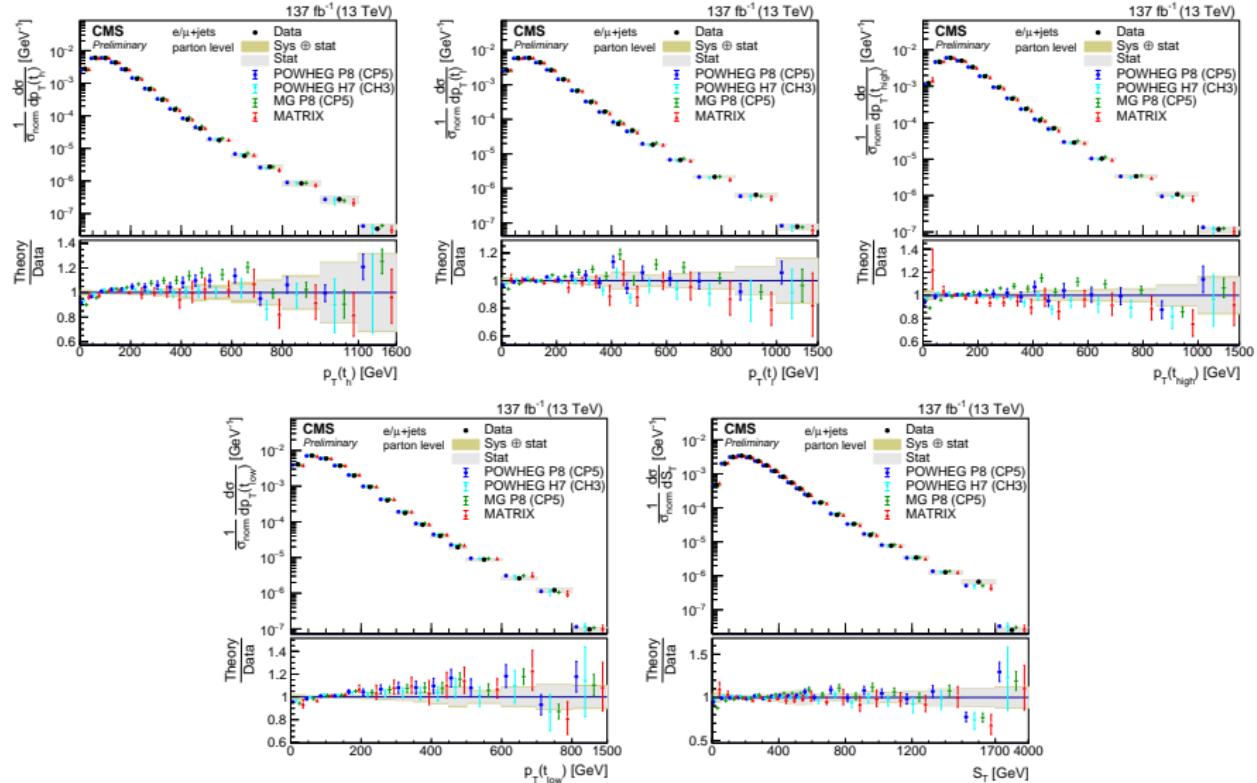
absolute

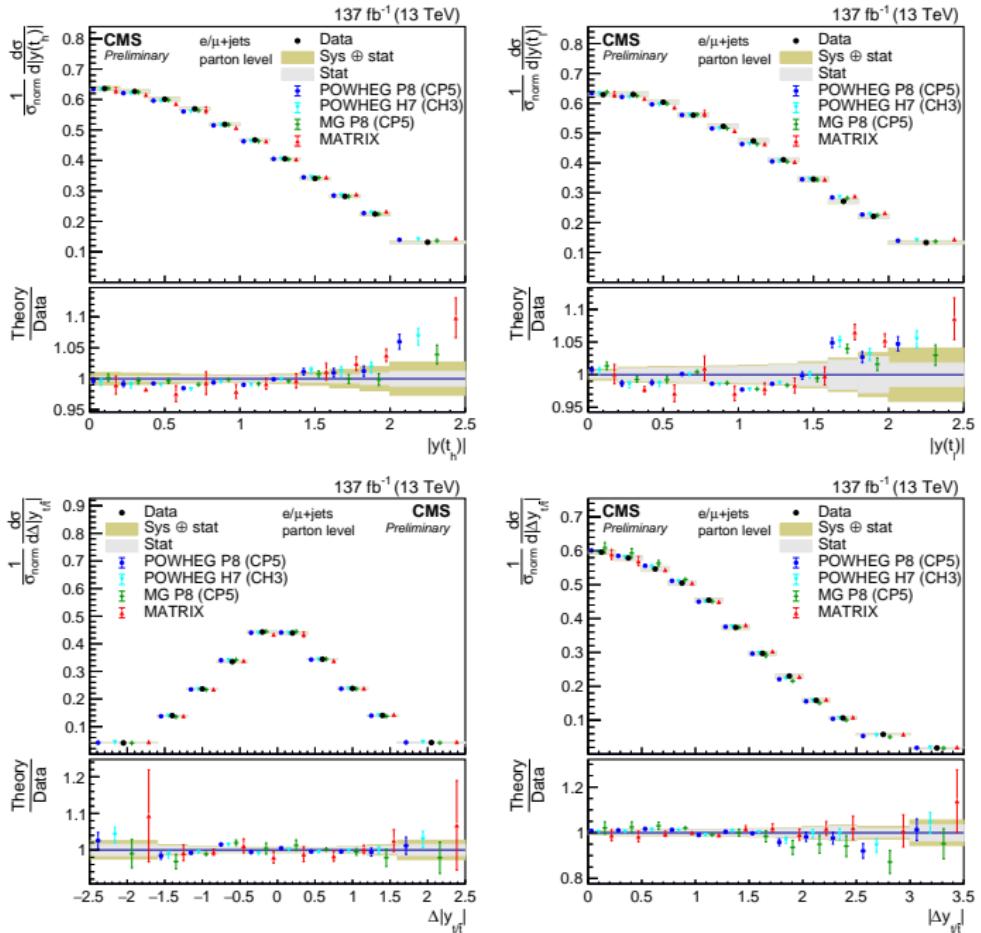


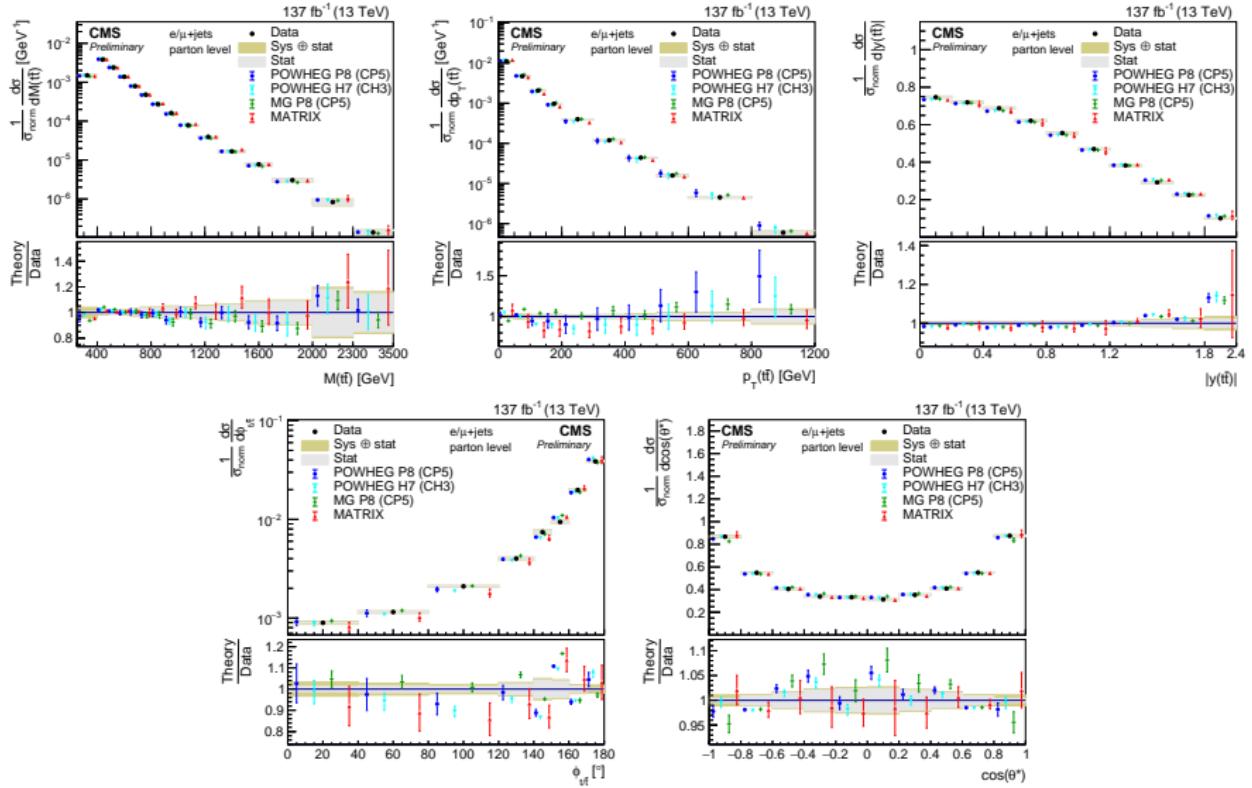
normalized

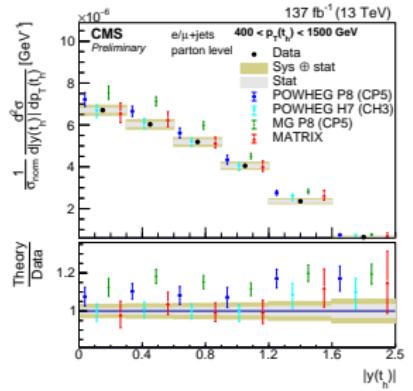
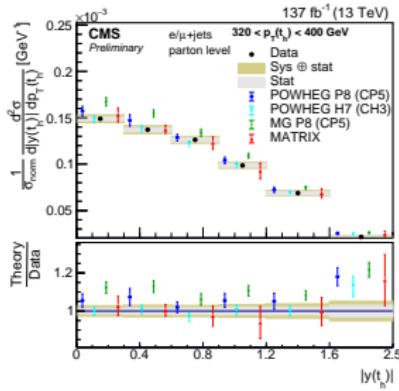
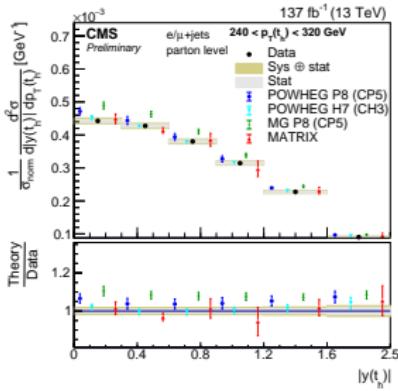
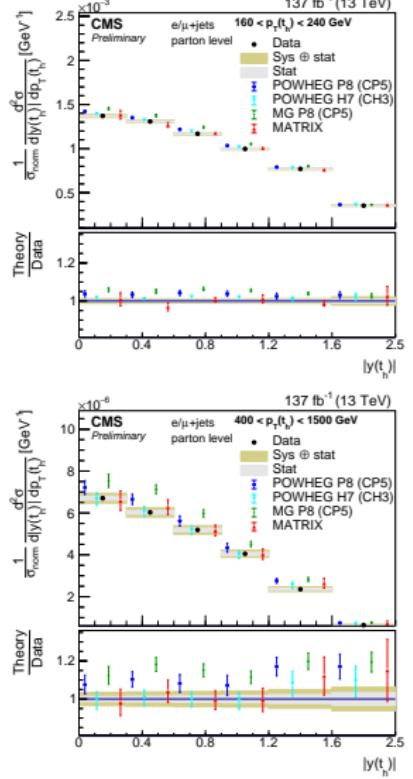
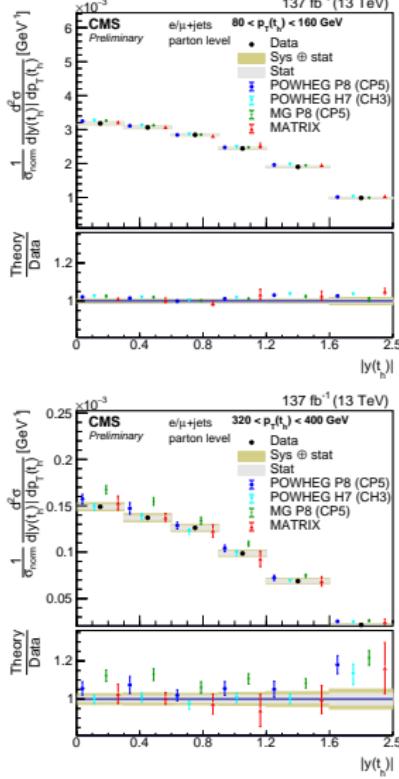
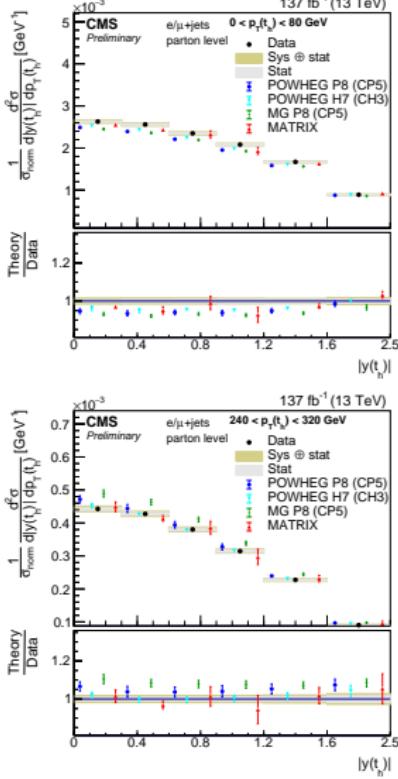


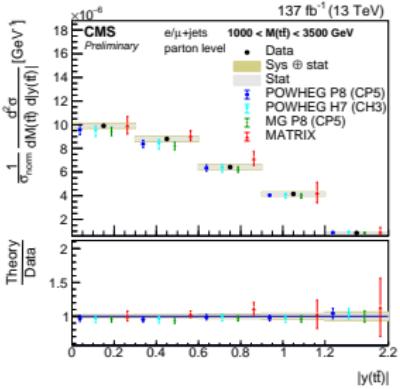
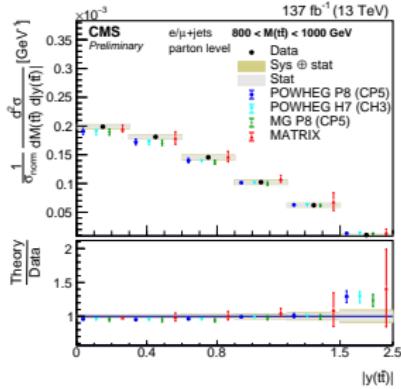
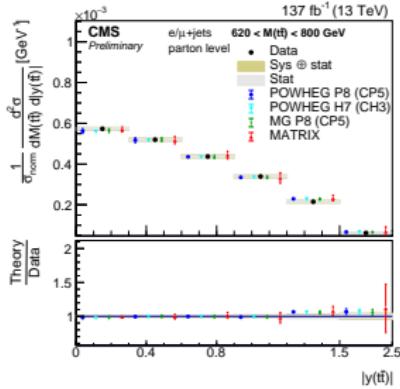
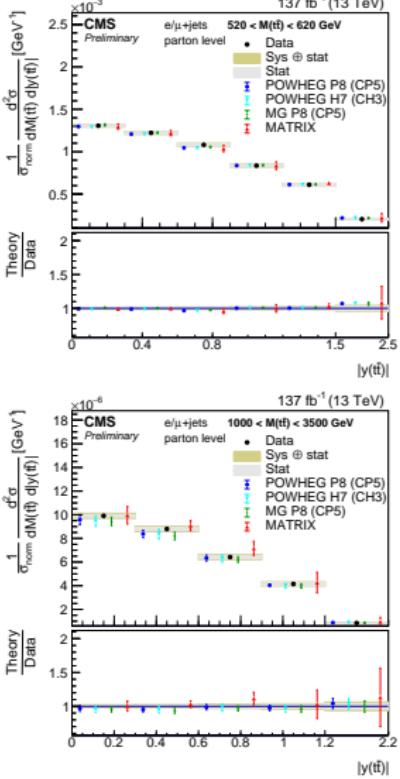
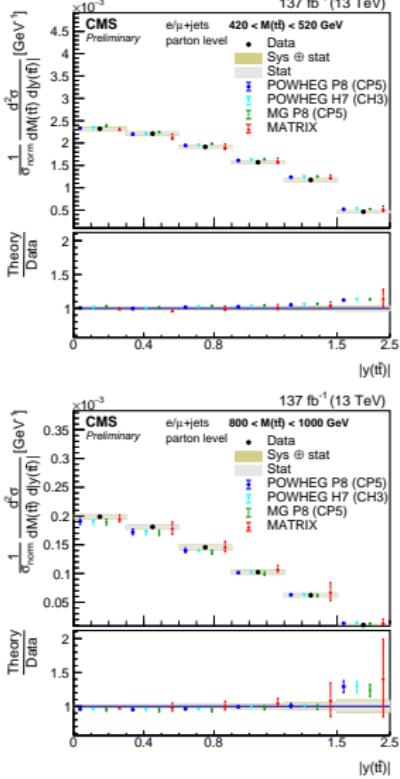
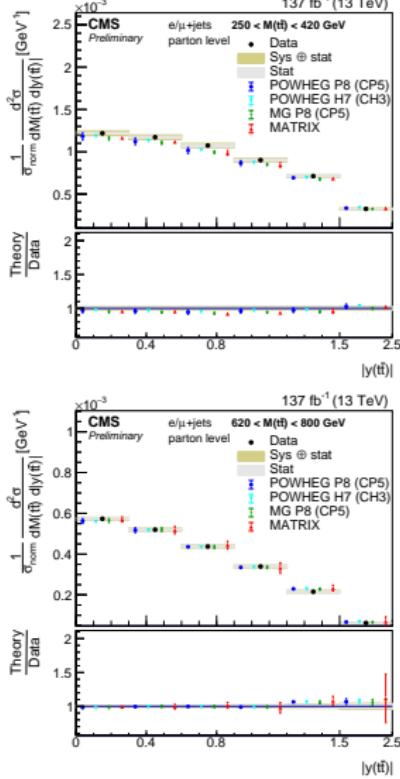
# Normalized differential cross sections at particle level

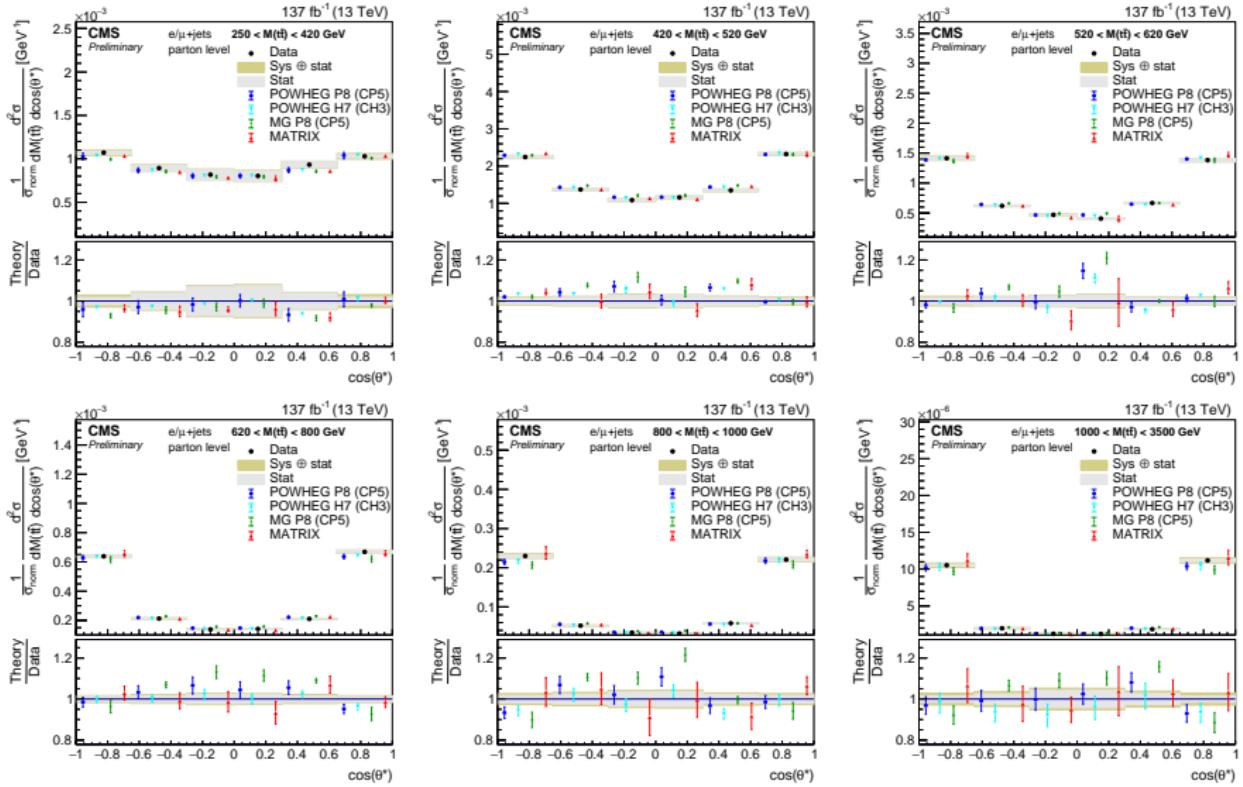


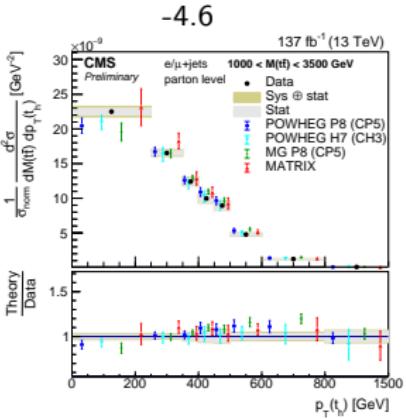
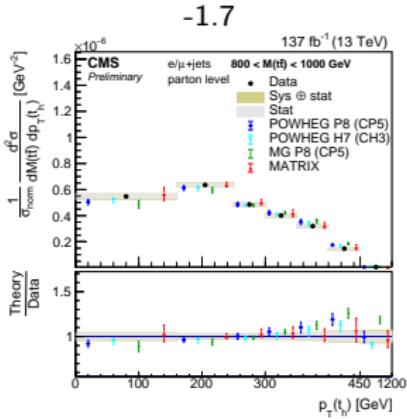
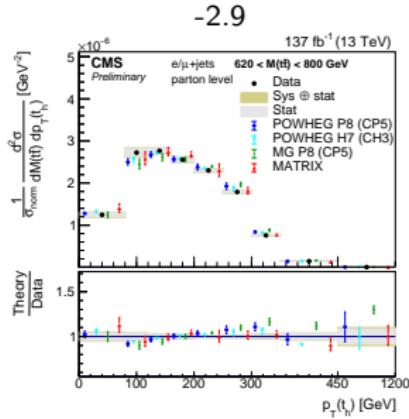
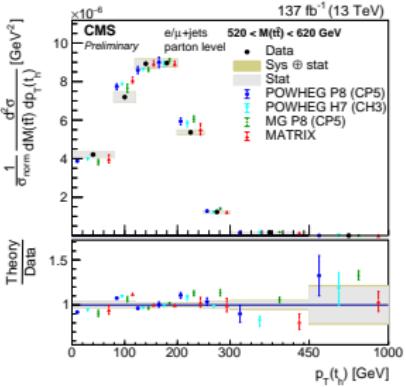
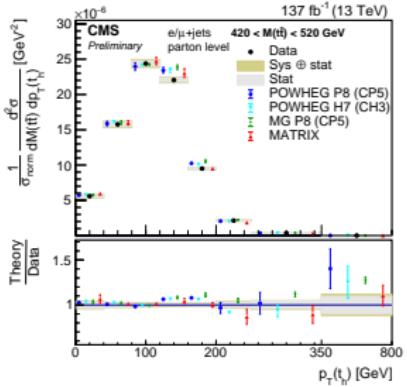
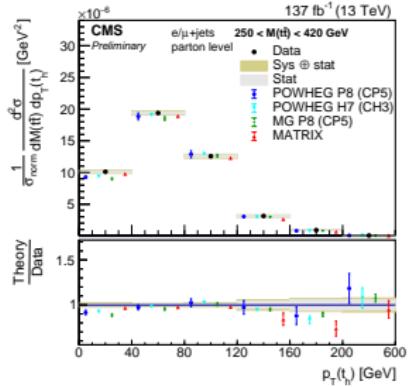








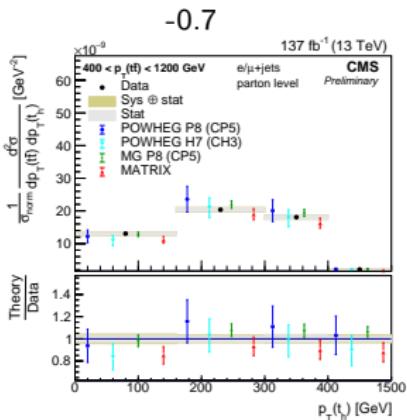
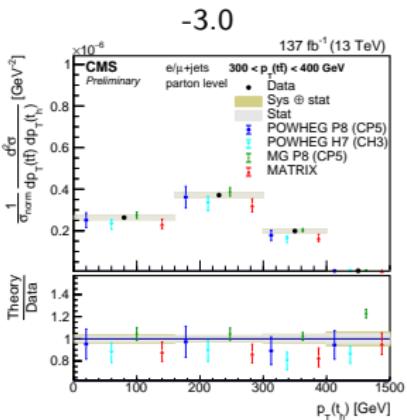
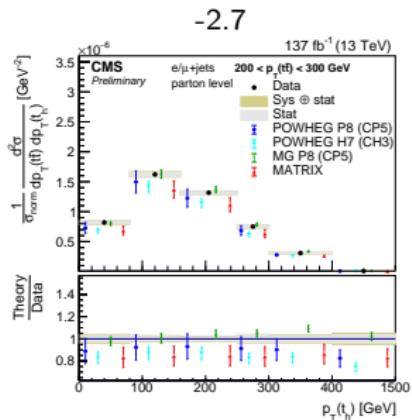
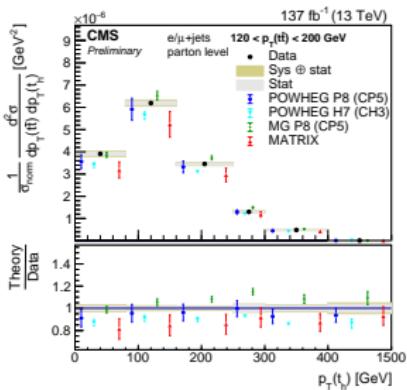
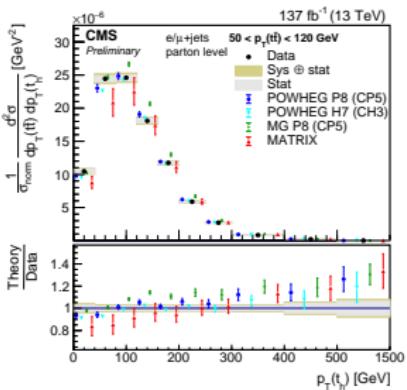
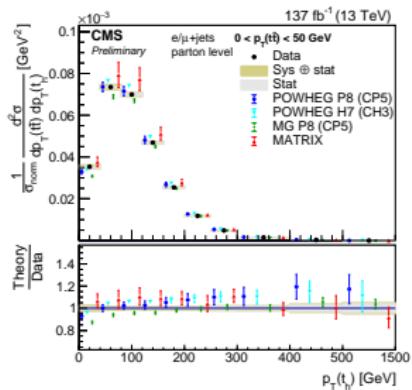




-4.1

-3.5

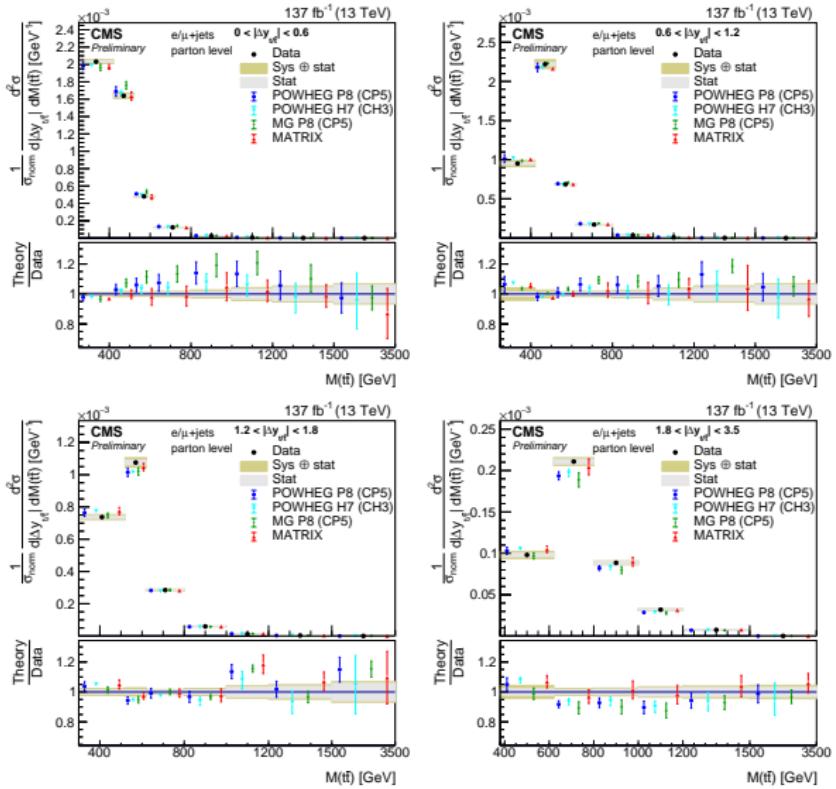
-2.6

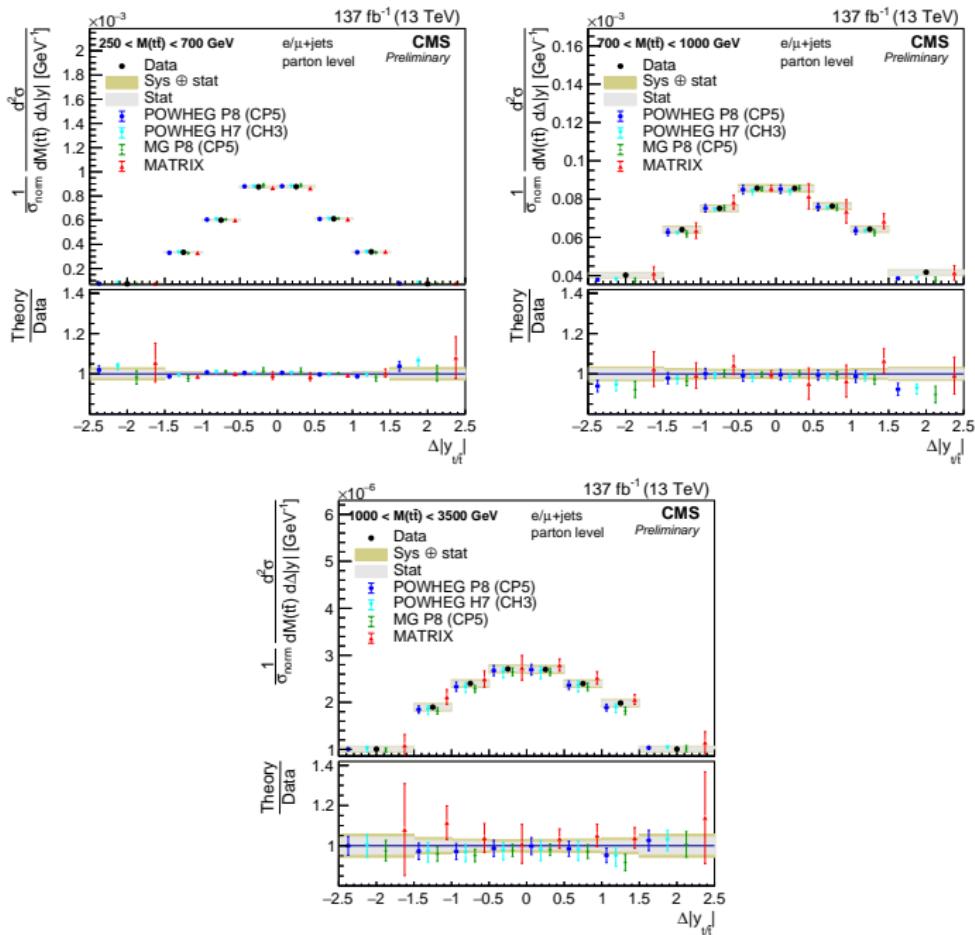


-0.4

0.3

-2.5





# Normalized differential cross sections at particle level

