TOP-12-010: Update and 2D unfolding linearity tests

C. Campagnari, A. George, F. Golf, J. Gran, D. Kovalskyi, V. Krutelyov, R. Villalba
University of California, Santa Barbara
G. Cerati, M. D'Alfonso, D. Evans, R. Kelley, Ian MacNeill, D. Olivito, S. Padhi, Y.Tu, C. Welke, F.Wuerthwein,A.Yagil, J. Yoo
University of California, San Diego
L. Bauerdick, K. Burkett, I. Fisk,Y. Gao, O. Gutsche, B. Hooberman, S. Jindariani, J. Linacre, V. Martinez Outschoorn
Fermi National Accelerator Laboratory

TOP2013 Pre-Approvals
6th August 2013

Paper Draft

- Paper draft was uploaded to CADI on Friday:
  - http://cms.cern.ch/iCMS/jsp/analysis/admin/analysismanagement.jsp?ancode=TOP-12-010
- ready for ARC review

2013/08/02
Head Id:
Archive Id: 200521P
Archive Date:

Charge Asymmetry Measurements in Top Dilepton Events in pp Collisions at $\sqrt{s} = 7$ TeV

The CMS Collaboration

Abstract

We present a measurement of the charge asymmetry in $t\bar{t}$ dilepton events produced in proton-proton collisions at $\sqrt{s} = 7$ TeV. The data sample used for the measurement corresponds to a total integrated luminosity of 5.0 fb^{-1}, collected by the CMS experiment during 2011. The inclusive top–antitop charge asymmetry and lepton charge asymmetry are measured to be $-0.010 \pm 0.017 \pm 0.017$ and $0.009 \pm 0.010 \pm 0.012$, respectively. Both asymmetries are consistent with their standard model expectations.

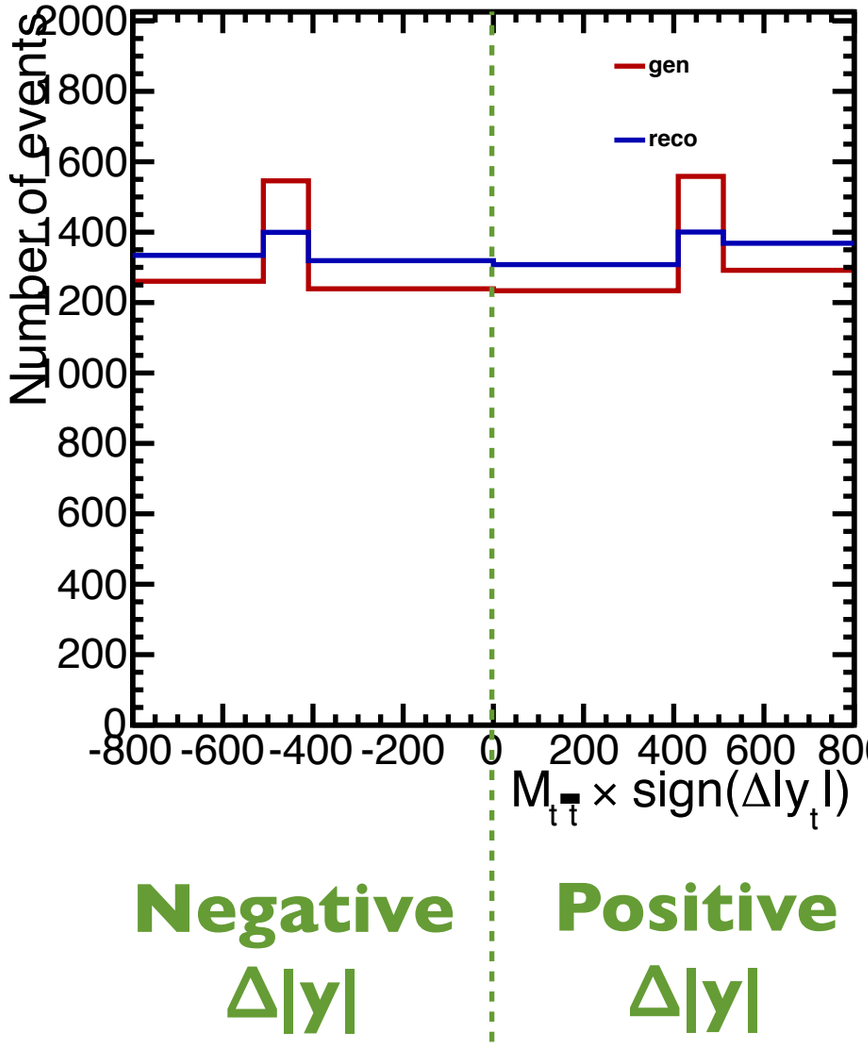
Introduction

- In TOP-12-010 we are measuring the top and lepton charge asymmetries, A_C and A_{lepC}, both inclusively and their differential dependence on M_{ttbar}, $p_{T,ttbar}$, and y_{ttbar} (using the same event selection and method as Yanjun just presented)

$$A_C = \frac{N(|y_t| > |y_{\bar{t}}|) - N(|y_t| < |y_{\bar{t}}|)}{N(|y_t| > |y_{\bar{t}}|) + N(|y_t| < |y_{\bar{t}}|)}.$$ **(requires top reco)**

$$A_{lepC} = \frac{N(|\eta_{l^+}| > |\eta_{l^-}|) - N(|\eta_{l^+}| < |\eta_{l^-}|)}{N(|\eta_{l^+}| > |\eta_{l^-}|) + N(|\eta_{l^+}| < |\eta_{l^-}|)}.$$ **(purely leptonic)**

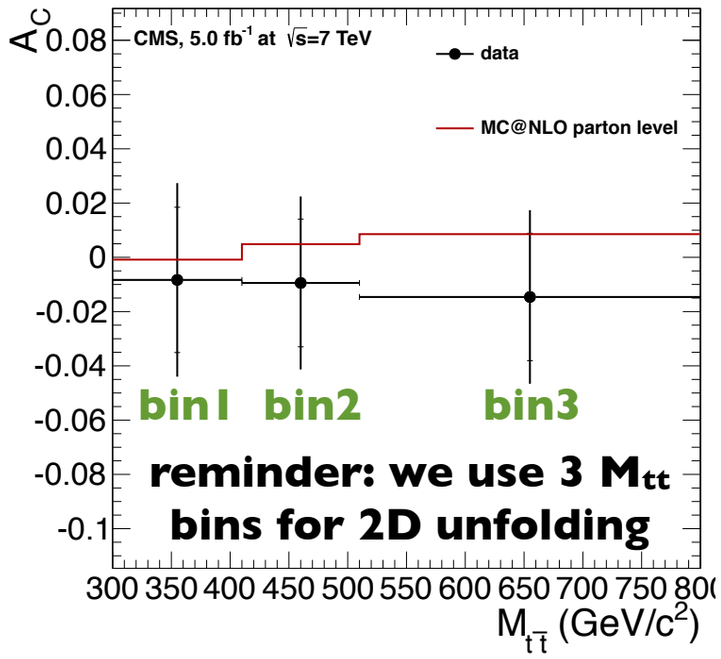
- We've found a problem with the linearity of our unfolding procedure for the A_C differential measurements (A_{lepC} is still OK)
  - it is related to our method of converting the "2D" unfolding into 1D by multiplying M_{ttbar}, $p_{T,ttbar}$, or $|y_{ttbar}|$ by the sign of the asymmetry variable (as done by CDF), meaning we effectively have only 2 bins in the asymmetry variable (see plot on right)



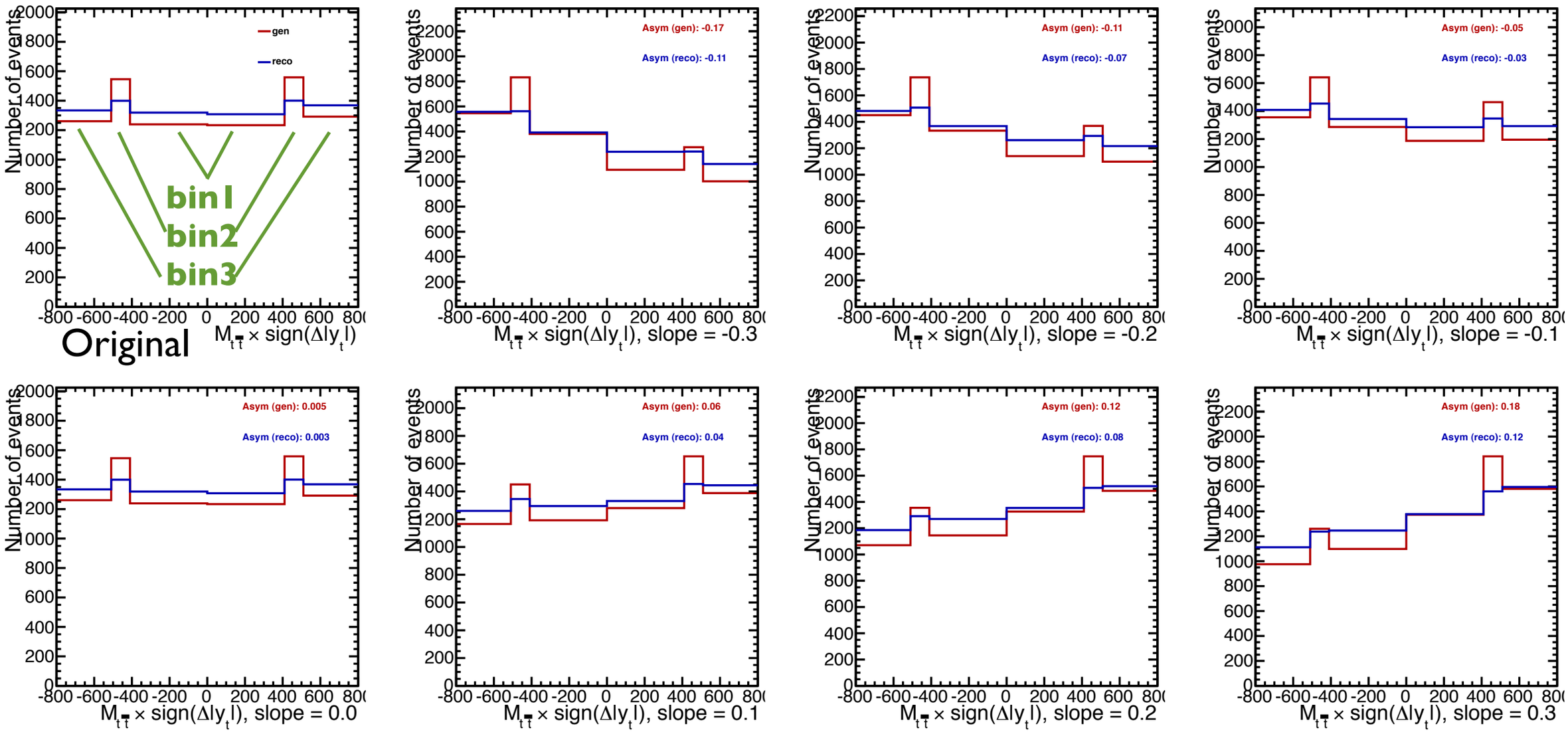
- These slides focus on the M_{ttbar} dependence, but the results and conclusions are the same for the $p_{T,ttbar}$ and y_{ttbar} dependence

Linearity test method

- We add asymmetry to the default MC distribution by applying weight= $1+\text{slope}*\Delta|y_{gen}|$ per event (see plots below), then compare unfolded asymmetry to true asymmetry, inclusively and in the 3 bins we use for the differential measurement

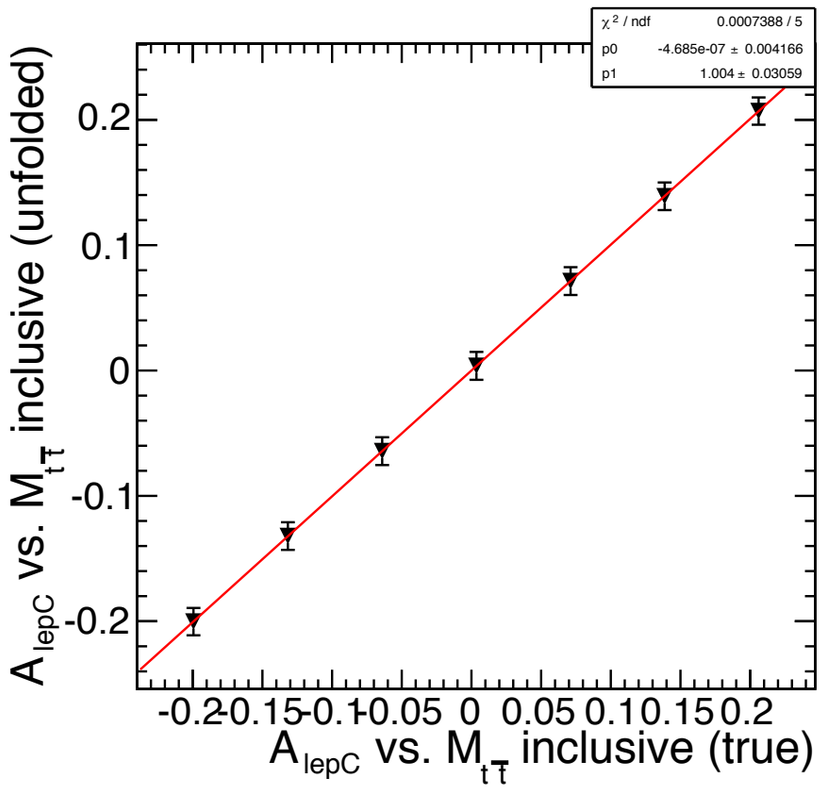


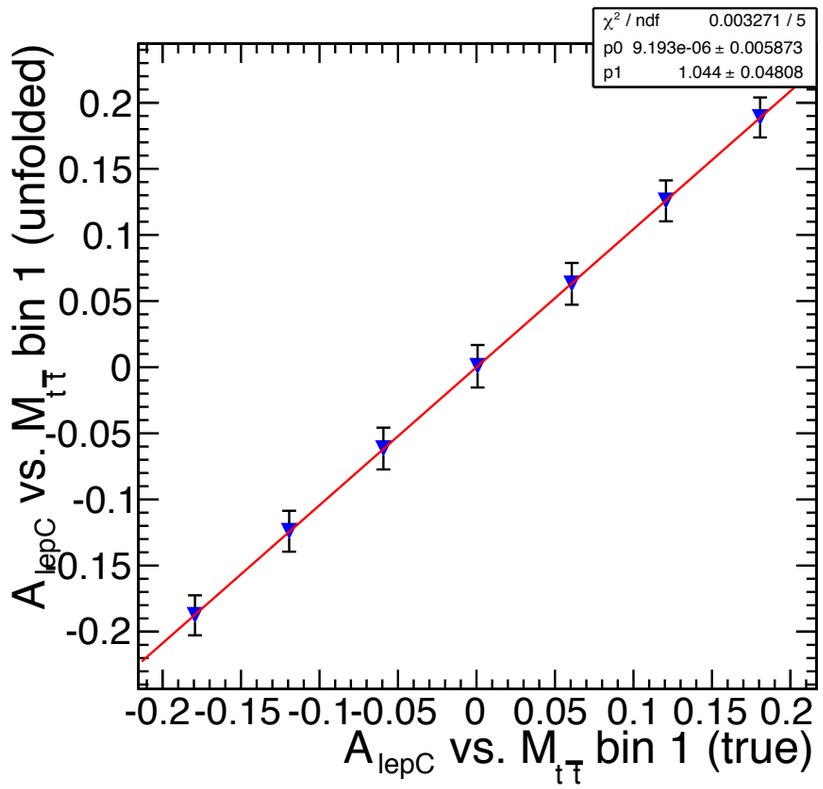
Original and weighted distributions for M_{tt}*sign(Δ|y|), before unfolding:

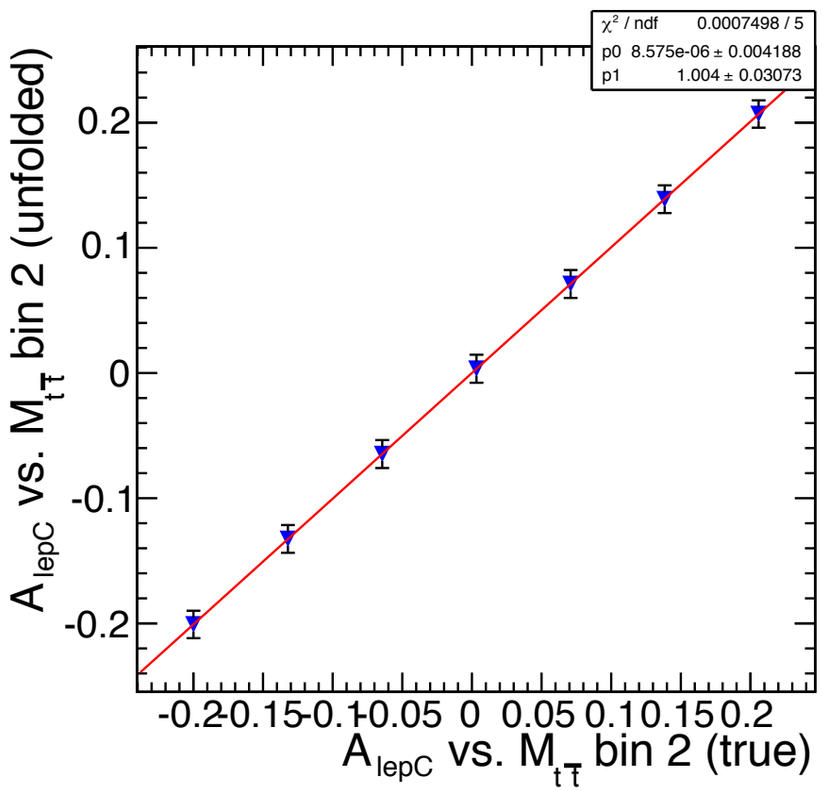


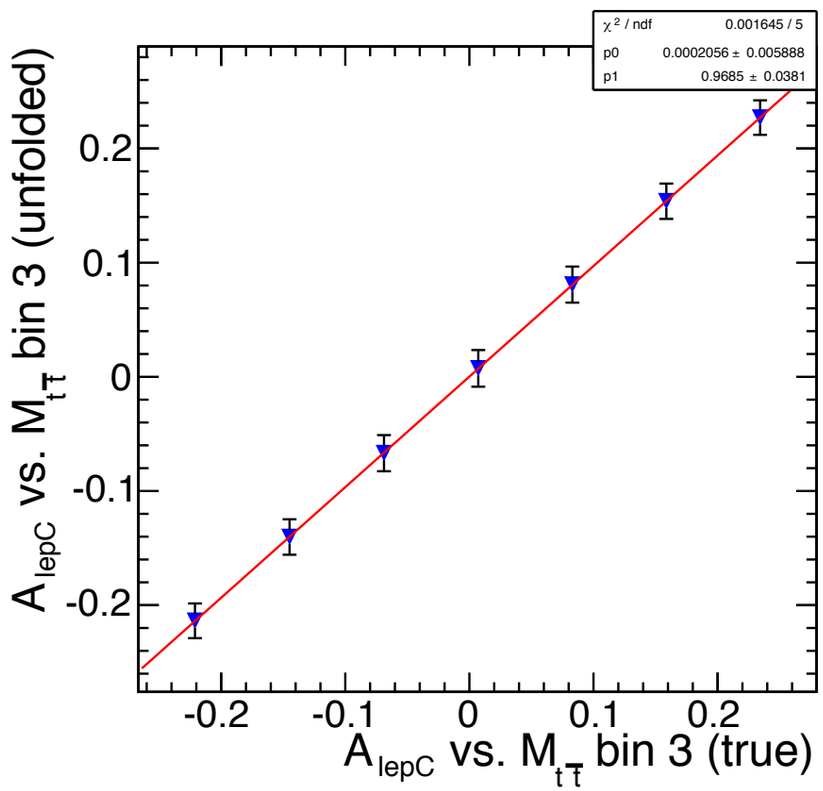
Results for A_{lepC}

- Linearity looks pretty good for all bins (p1~1)
  - inclusive: 1.004
  - bin1: 1.044
  - bin2: 1.004
  - bin3: 0.969



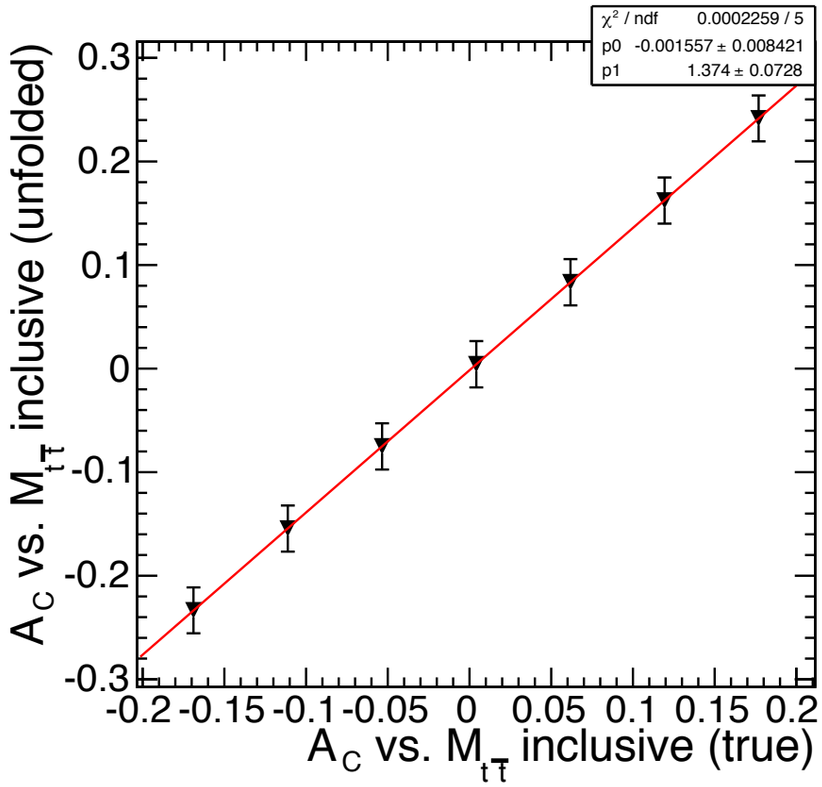


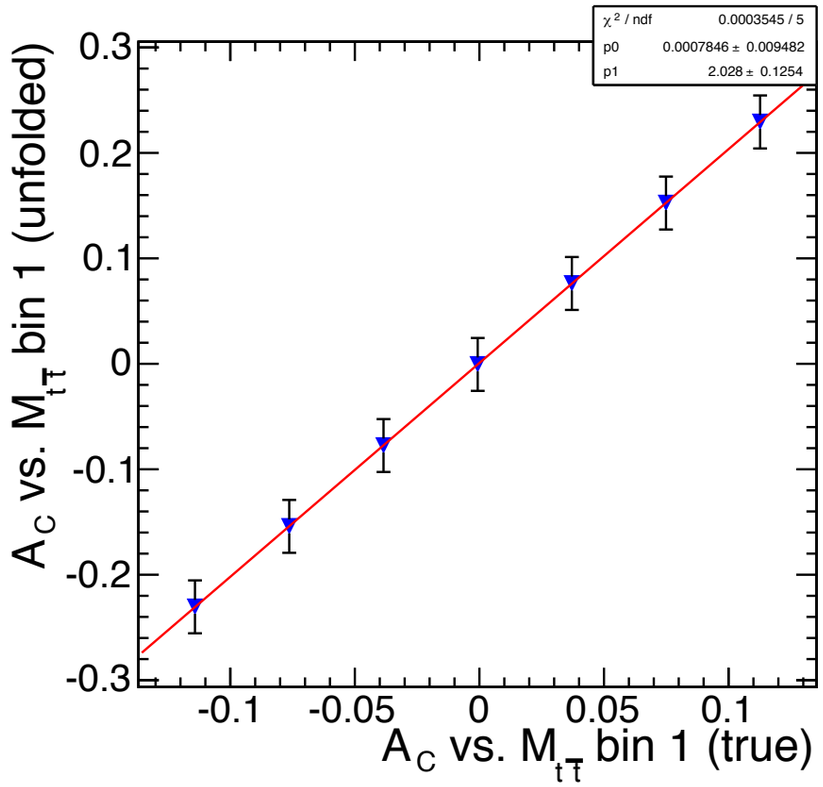


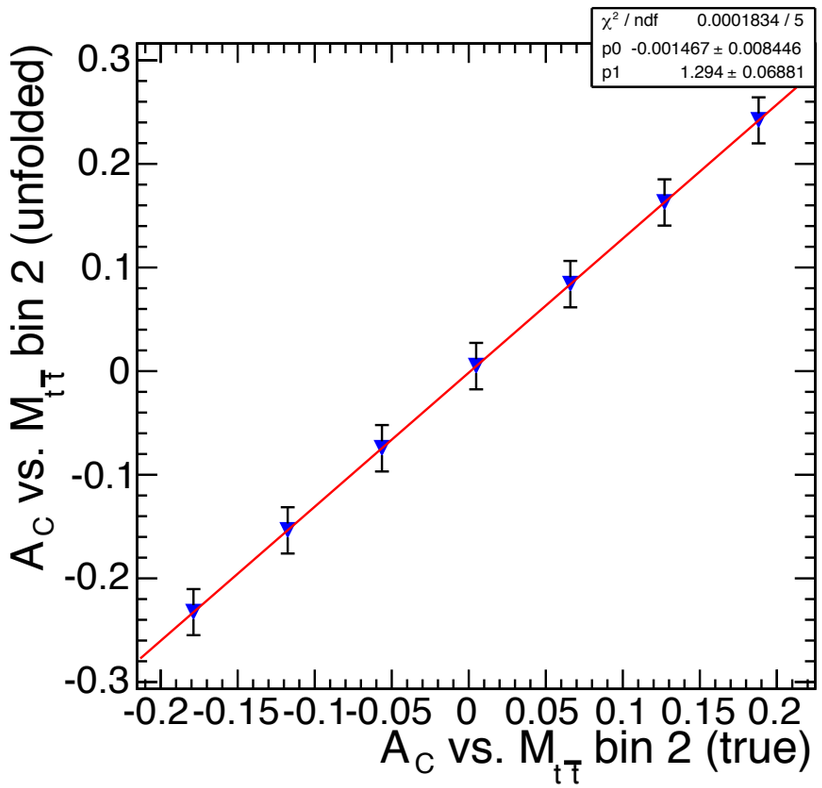


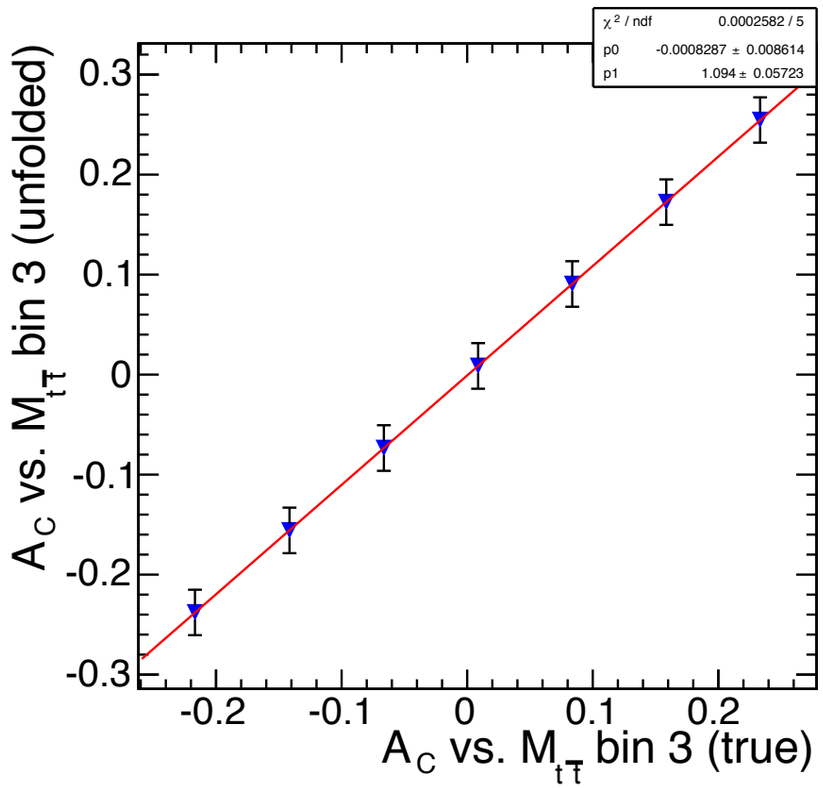
Results for A_C

- Things don't look so good here
  - Slope of linearity plot (p1) significantly away from one for bin1 (2.03) and bin2 (1.29), as well as inclusively (1.37)
- We think this is an inherent bias of using 2 bins in $\Delta|y|$ to unfold distributions with a continuous slope in $\Delta|y_{gen}|$ (as expected for a real physics effect, and which we generate using weight = 1+slope*$\Delta|y_{gen}|$, see slide 4)









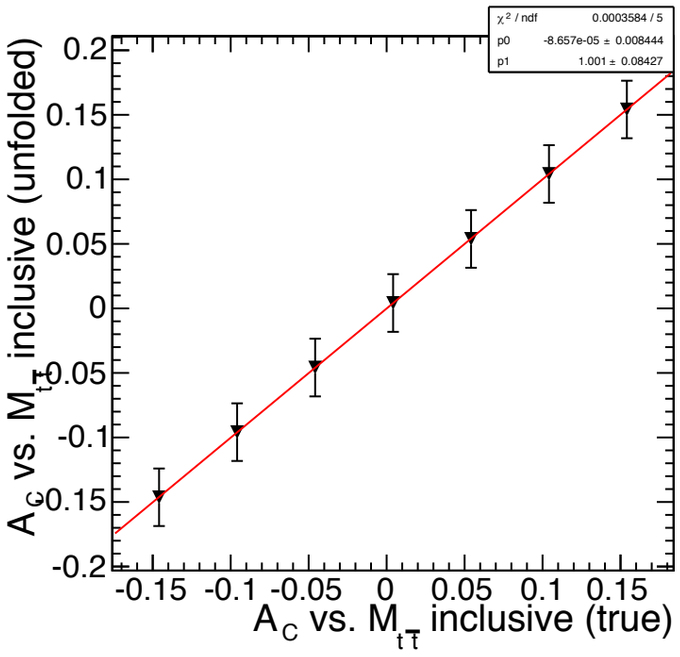
Bias from insufficient Δ|y| bins

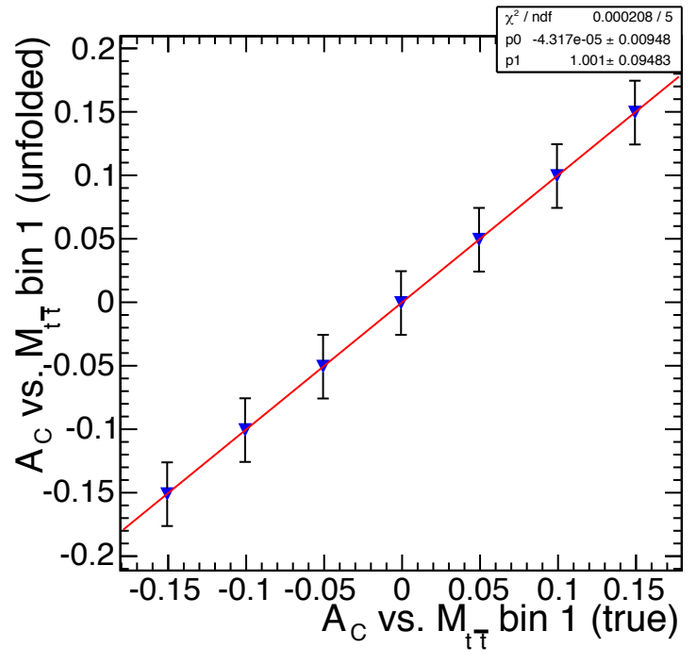
- We add a slope by applying weight = 1+slope*Δ|y_{gen}|
  - => events with large | Δ|y_{gen}| | get the largest change in weight
    - but events with large | Δ|y_{gen}| | are also the events least prone to changing sign of Δ|y_{reco}| through migration
      - however, the response matrix treats them as though they had the average chance of migration (since it has only 2 bins in Δ|y|, the same migration probability is used for all events) => biased unfolding

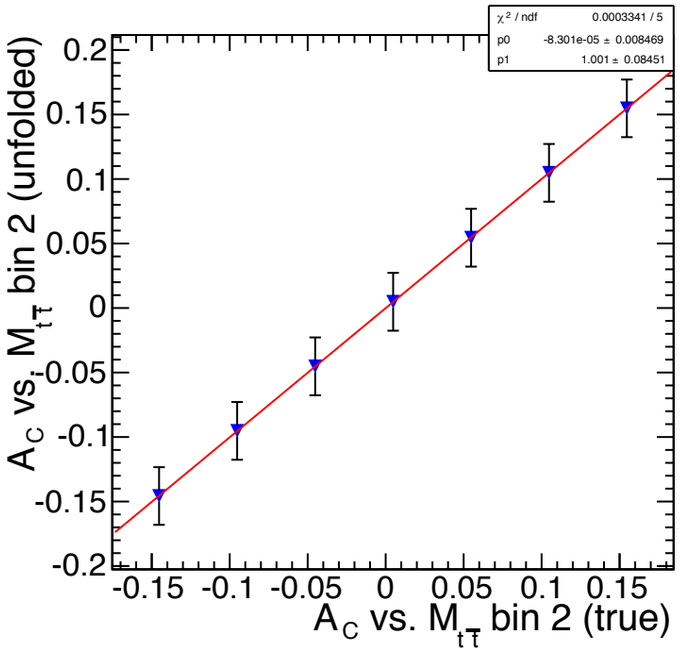
- If instead we add a slope matching the granularity of the response matrix, i.e. weight 1+0.5*slope*sign(Δ|y_{gen}|), the weighting will also affect all events equally
  - => should be no bias
    - this is confirmed on the next slide

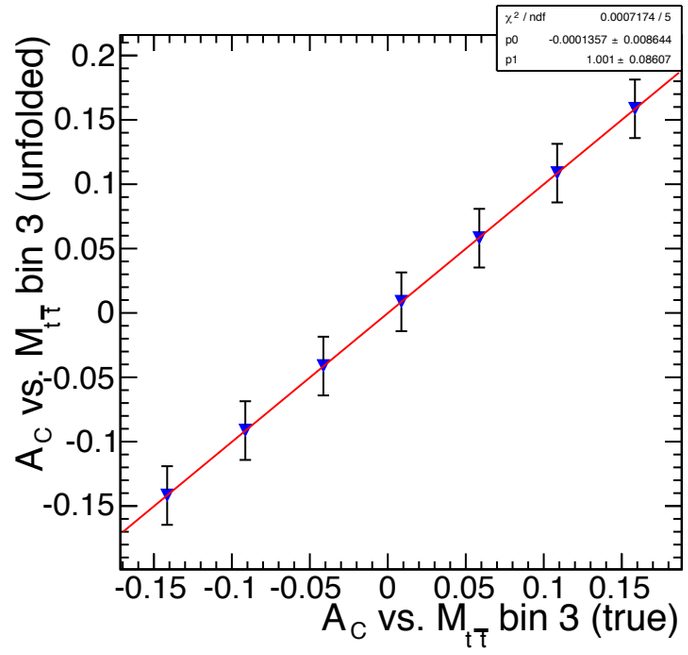
Results for A_C (slope only in sign($\Delta|y_t|$))

- Slope now very close to p1=1 for all bins
  - this result holds for both asymmetry variables, and unfolding vs M_{tt}, $p_{T,tt}$ and y_{tt}
  - but of course the results on slides 5 and 6 are a more accurate estimate of the linearity of the method for a general physics effect
  - The only reason we can get away with using 2 $\Delta|y|$ bins for A_{lepC} (slide 5) is that the response matrix is essentially diagonal (excellent experimental resolution)









Linearity test conclusions

- Linearity of the method is good for the differential dependence of A_{lepC}, but poor for A_C
  - bias comes from using a response matrix with insufficient resolution (only 2 bins in $\Delta|y|$)
- We think showing 2D results for just 1 of the 2 variables might cause problems getting the paper published
- We are therefore considering including only the 1D results in the paper, i.e. the values and distributions for A_C and A_{lepC}

Results in paper

- **These are the inclusive results we definitely want to show**

Asym.	Data (unfolded)	Simulation
A_C	$-0.010 \pm 0.017 \pm 0.017$	0.004 ± 0.001
A_{lepC}	$0.009 \pm 0.010 \pm 0.012$	0.004 ± 0.001

Table 4: Parton-level asymmetries. The uncertainties on the unfolded results are statistical and systematic respectively. The uncertainties on the simulated results are statistical only.

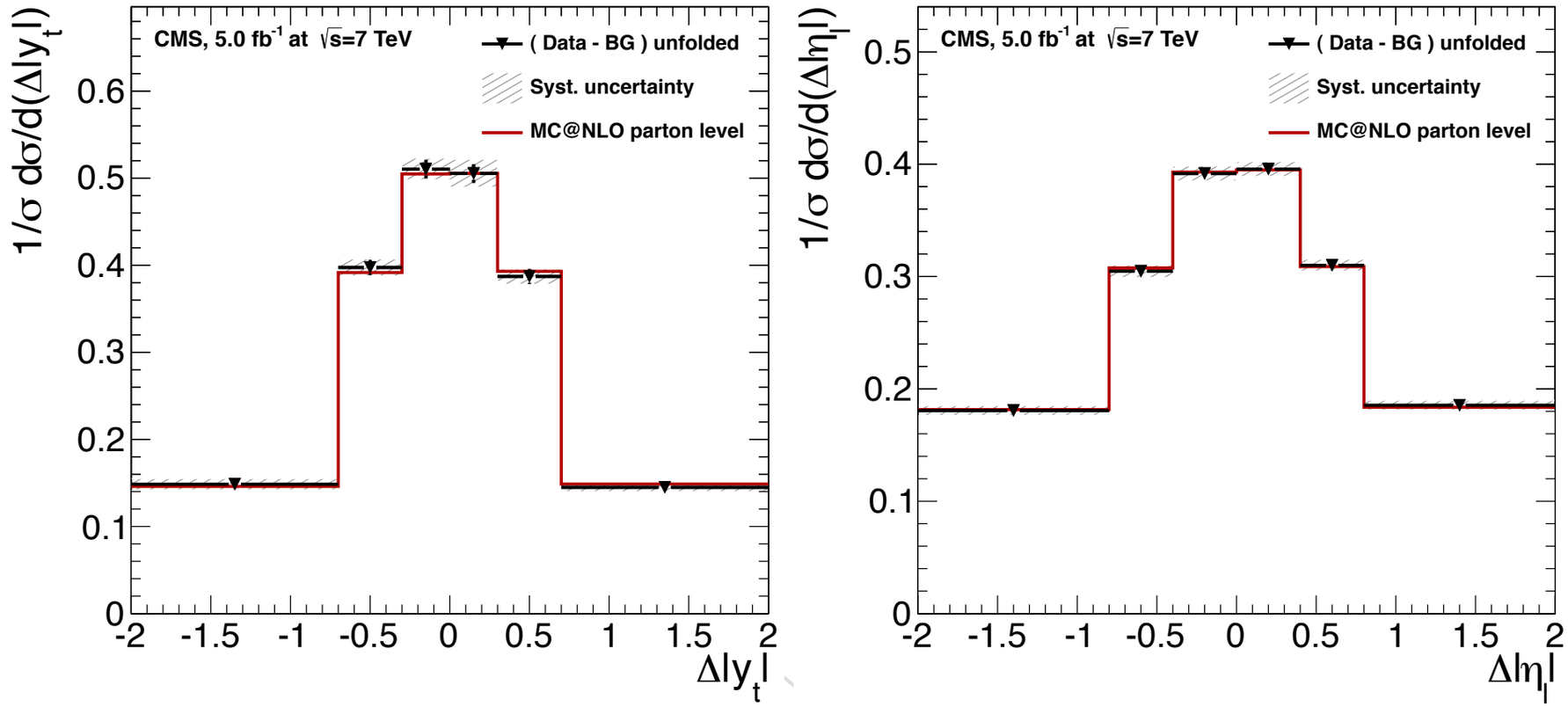
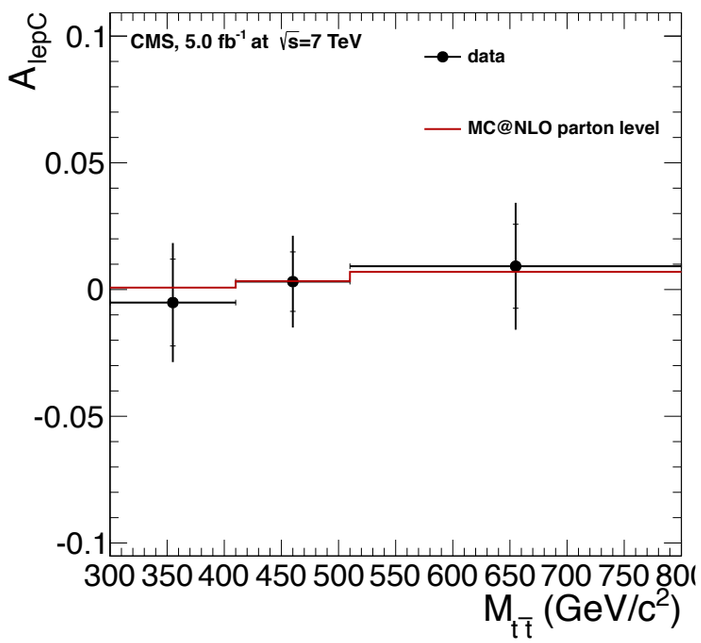
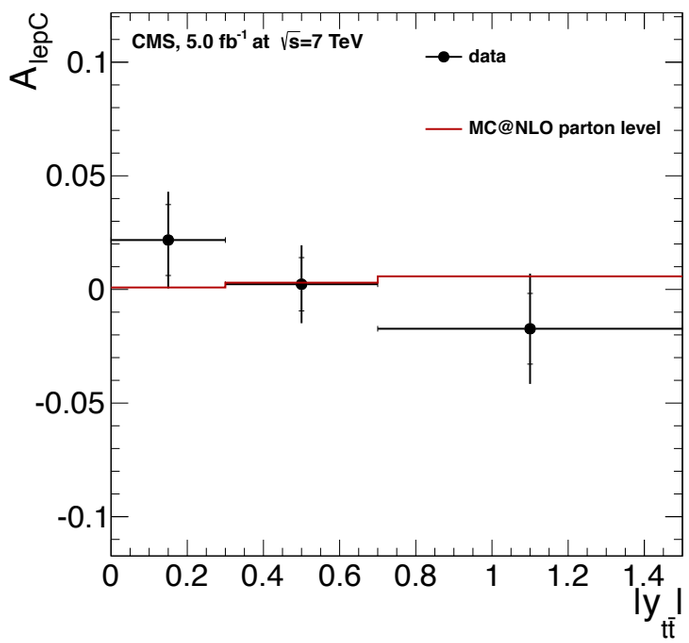


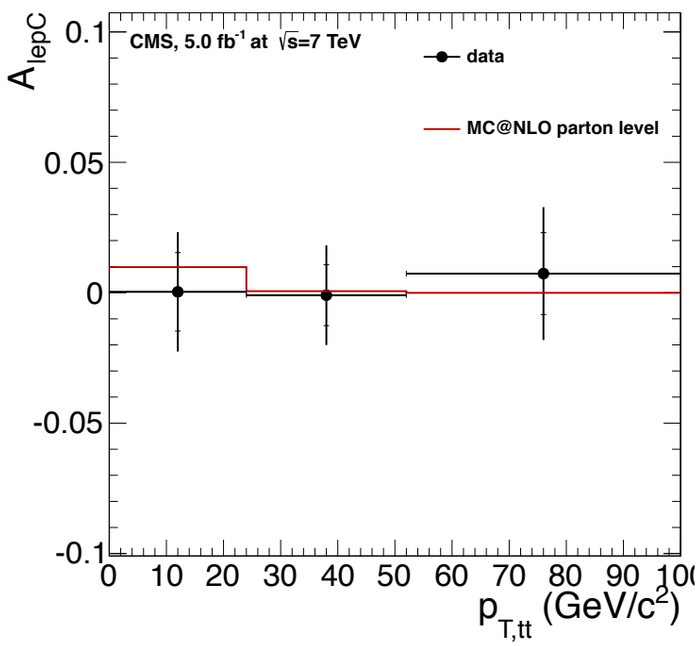
Figure 5: Background-subtracted and unfolded differential cross-sections for $\Delta|y|$ (left) and $\Delta|\eta_l|$ (right). The error bars represent statistical uncertainties only, while the systematic uncertainty band is represented by the hatched area. Note that the bin values are correlated due to the unfolding.

2D results that could be included

- These are the 2D results for A_{lepC}, for which our method demonstrates good linearity (because the response matrix is close to diagonal)
  - we would like to include these in the paper, but are concerned that showing 2D results for only 1 of the 2 variables may cause problems later on



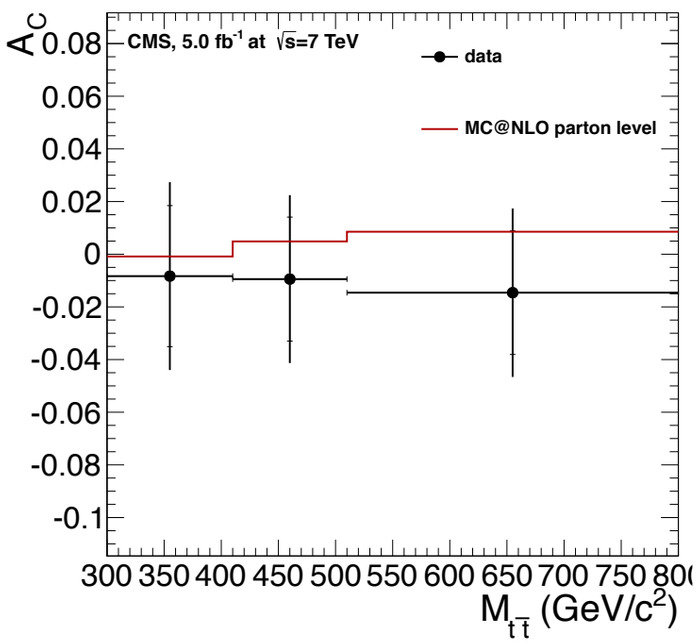


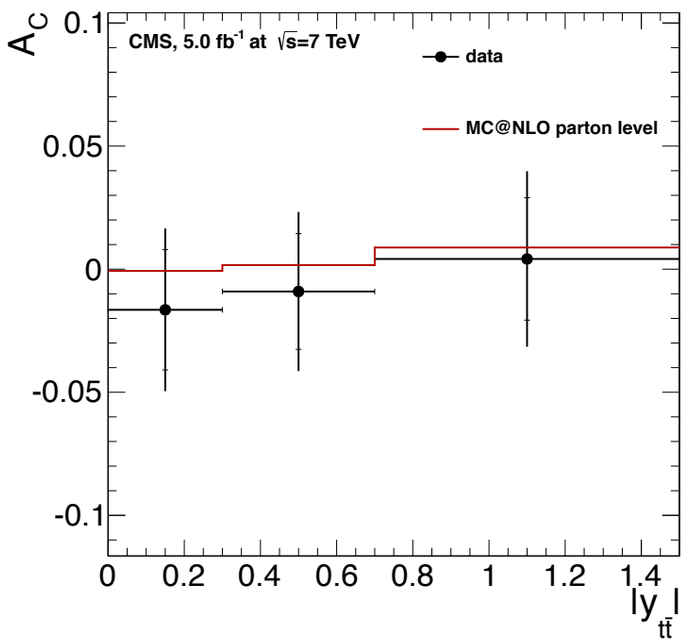


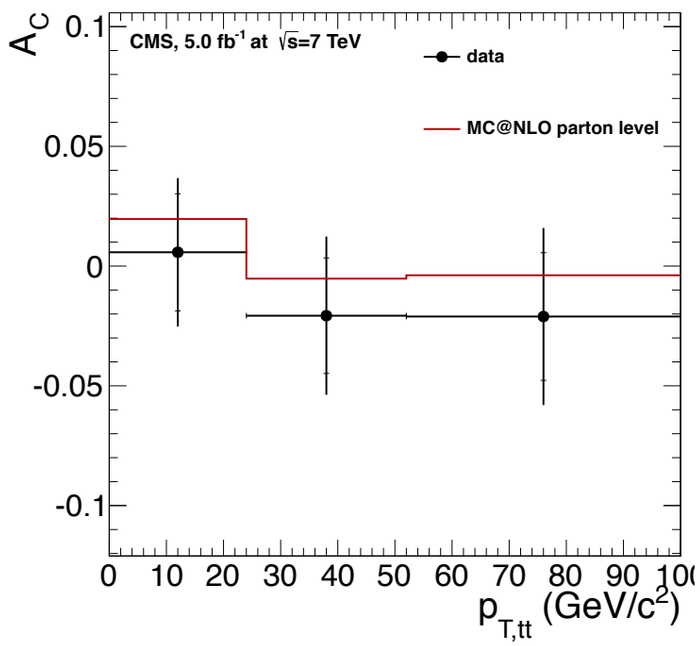
* systematics for the 2D plots are being updated, but are expected change only very slightly

2D results we don't want to show

- These are the 2D results for A_C, where the linearity is poor (because a higher resolution response matrix is needed due to the significant event migration from inaccurate top reconstruction)
  - we don't want to include these in the paper (although since the results are so close to the SM, the bias is actually very small)







* systematics for the 2D plots are being updated, but are expected change only very slightly

Backup

More paper material

Table 1: The observed and simulated yields after the event selection described in the text. Uncertainties are statistical only. The systematic uncertainties on the simulated yields are given in Section 7. Where the simulated yields are zero, upper limits are given based on the weighted yield of a single simulated event passing the selection.

Sample	ee	$\mu\mu$	eμ	all
$t\bar{t} \to \ell^+\ell^-$	1519.53 ± 9.87	1801.10 ± 10.38	5763.73 ± 19.01	9084.37 ± 23.80
$t\bar{t} \to$ other	38.32 ± 1.59	4.02 ± 0.45	91.65 ± 2.40	133.99 ± 2.92
W + jets	< 2	4.72 ± 3.34	11.11 ± 5.10	15.83 ± 6.10
DY$\to \ell^+\ell^-$	30.20 ± 4.39	29.55 ± 4.14	35.02 ± 4.50	94.77 ± 7.53
Di-boson	8.27 ± 0.44	10.20 ± 0.47	27.90 ± 0.81	46.37 ± 1.03
Single top	72.54 ± 2.11	86.77 ± 2.23	289.37 ± 4.20	448.68 ± 5.20
Total (simulation)	1668.86 ± 11.13	1936.36 ± 11.89	6218.78 ± 20.78	9824.00 ± 26.40
Data	1631	1964	6229	9824

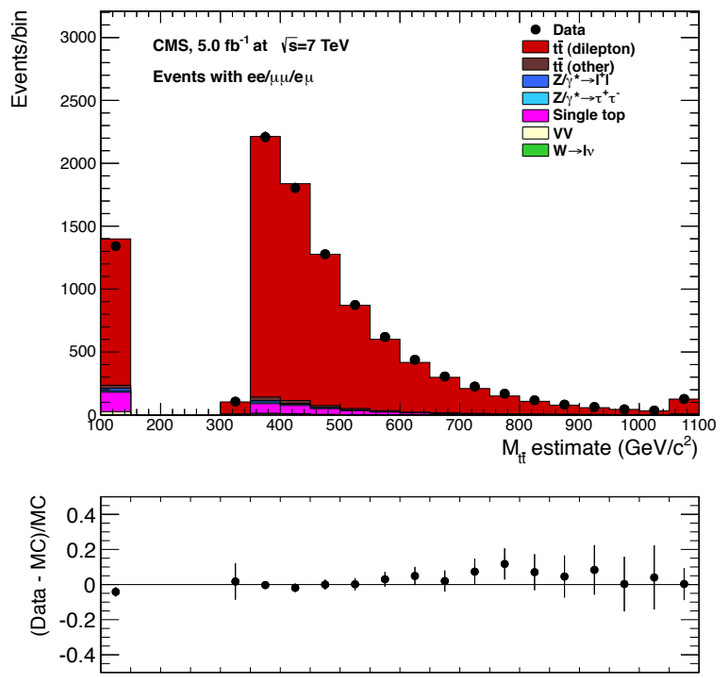


Figure 1: Comparison between data and the simulation for the reconstructed $t\bar{t}$ mass, where events with no solution found are filled in the first bin.

More paper material

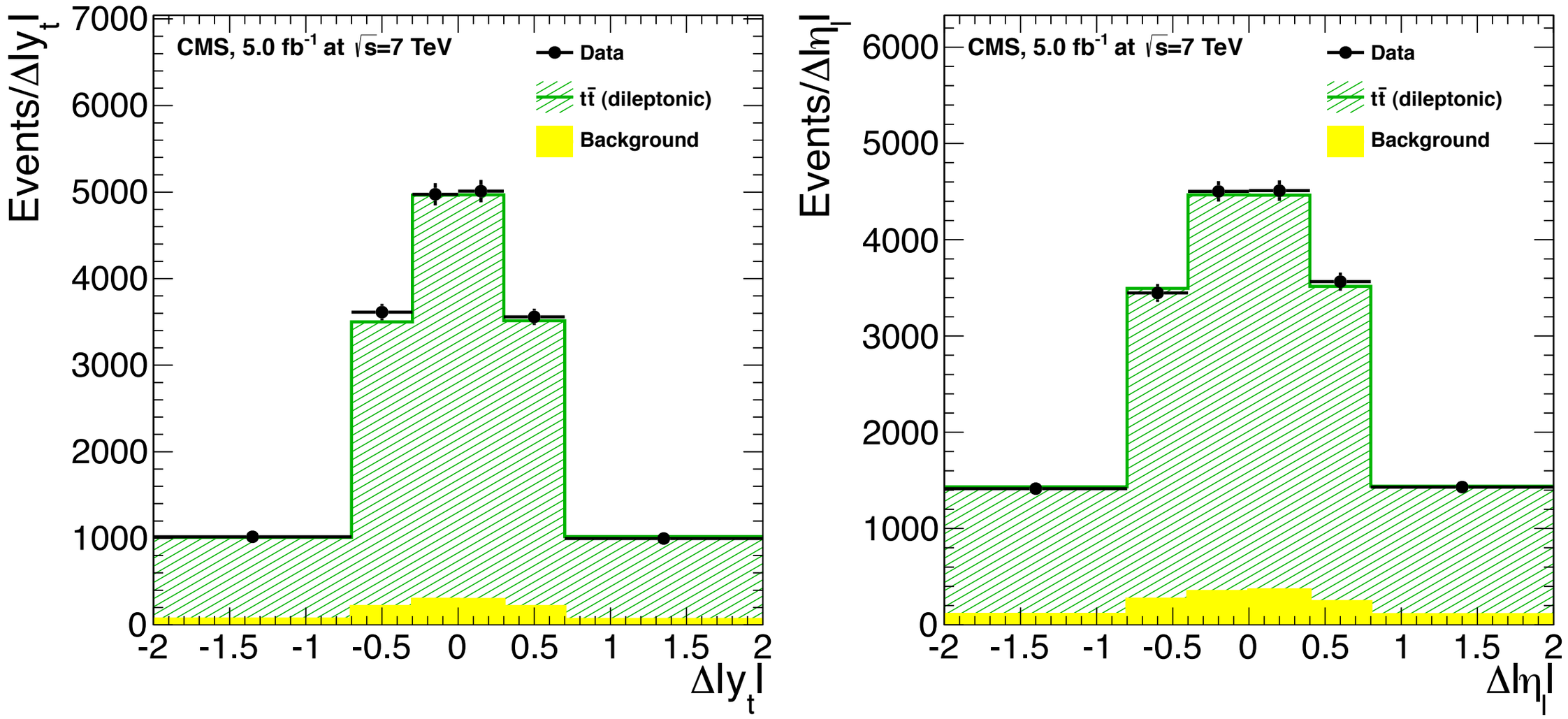


Figure 2: The reconstructed $\Delta|y|$ (left) and $\Delta|\eta_l|$ (right) distributions, for data and the simulation. The simulated events are divided into $t\bar{t} \to \ell^+\ell^-$ and background, where the background consists of the categories other than $t\bar{t} \to \ell^+\ell^-$ in Table 1.

More paper material

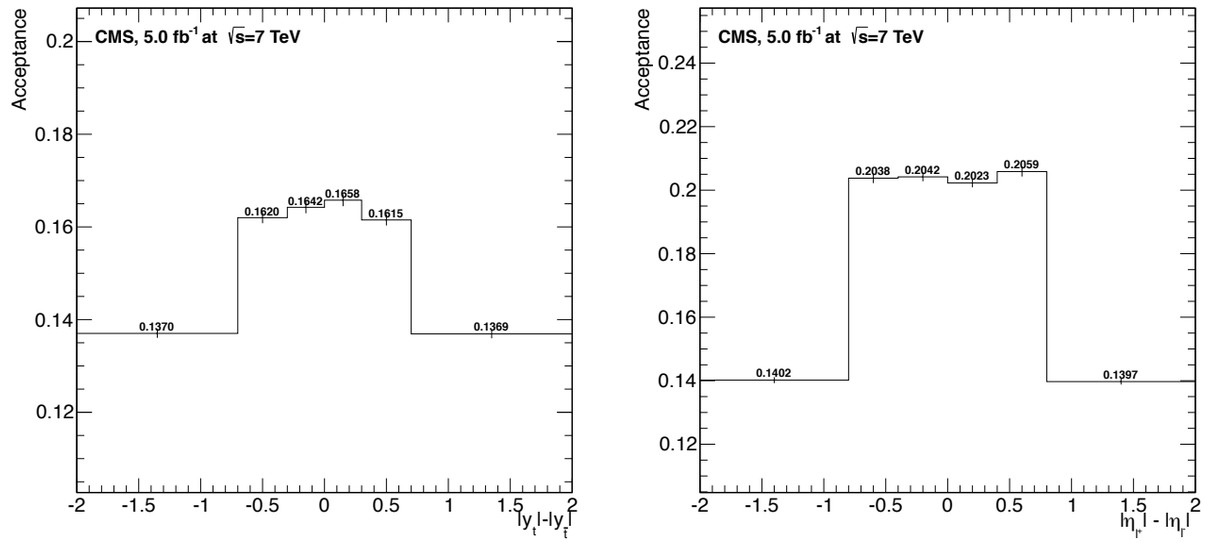


Figure 3: Acceptance matrix bins for $\Delta|y|$ (left) and $\Delta|\eta_l|$ (right).

Table 2: Binning used in the distributions of $\Delta|y|$ and $\Delta|\eta_l|$.

	B1	B2	B3	B4	B5	B6
$\Delta\|\eta_l\|$	[-2.0,-0.8]	[-0.8,-0.4]	[-0.4,-0.0]	[0.0, 0.4]	[0.4, 0.8]	[0.8, 2.0]
$\Delta\|y\|$	[-2.0,-0.7]	[-0.7,-0.3]	[-0.3,-0.0]	[0.0, 0.3]	[0.3, 0.7]	[0.7, 2.0]

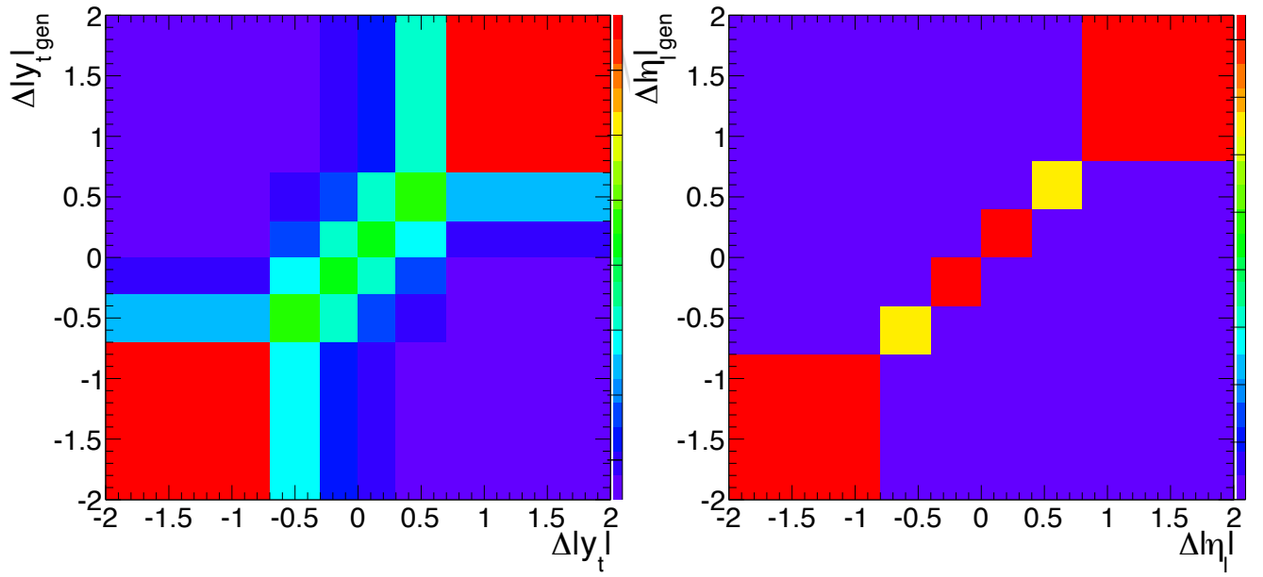


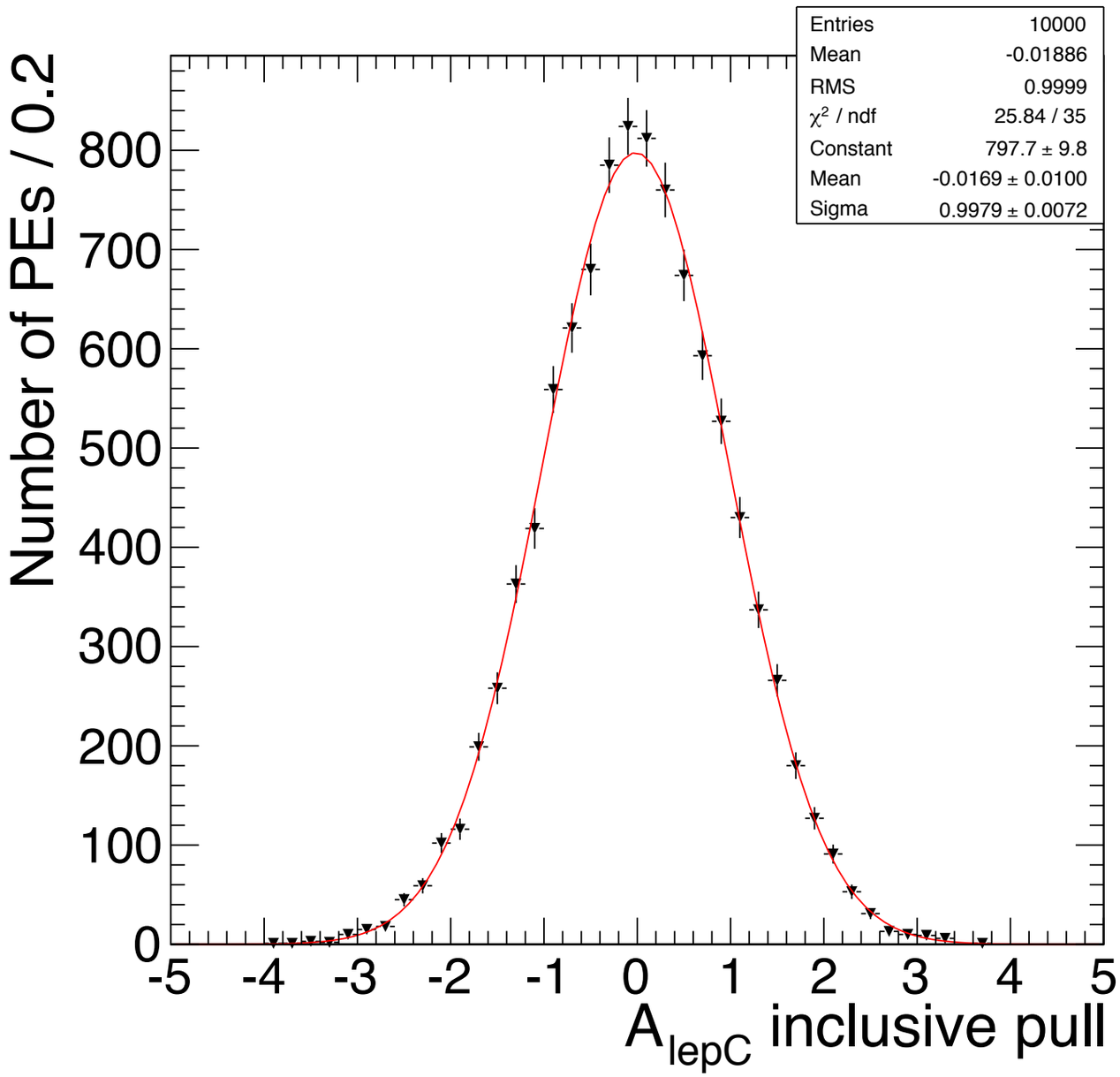
Figure 4: Smearing effects due to the uncertainties of top or lepton reconstructions $\Delta|y|$ (left) and $\Delta|\eta_l|$ (right).

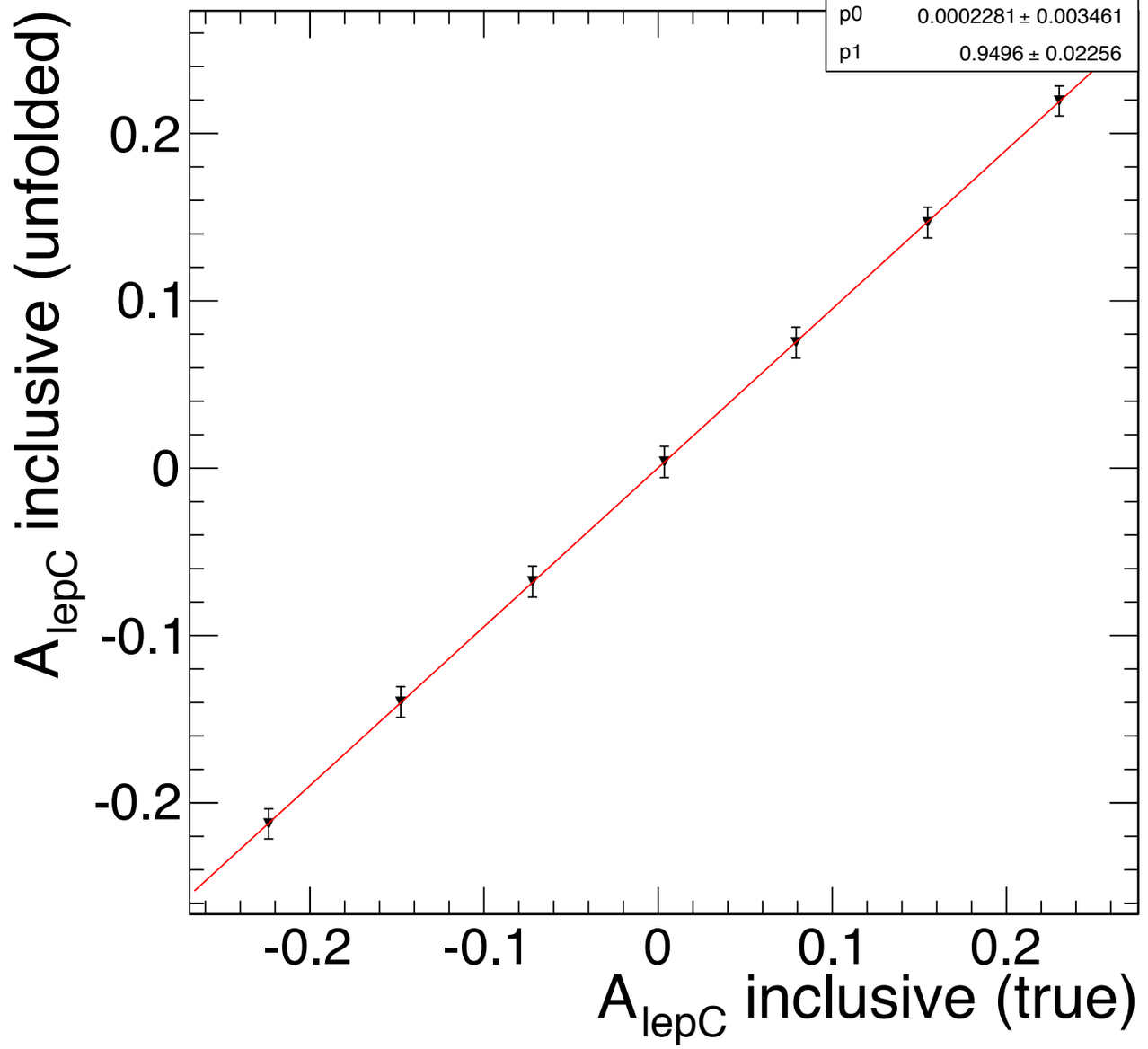
More paper material

Asymmetry	A_C	A_{lepC}
experimental systematic uncertainties		
Jet energy scale	0.003	0.001
Jet energy resolution	0.000	0.000
Lepton energy scale	0.000	0.000
b-tagging SF	0.000	0.000
Lepton selection	0.000	0.000
Pileup	0.000	0.001
Background	0.001	0.001
$t\bar{t}$ modeling uncertainties		
Top mass	0.003	0.002
Fact. and renorm. scales	0.003	0.005
Tau decay	0.000	0.000
PDF	0.000	0.000
Top p_T reweighting	0.001	0.000
Unfolding	0.016	0.010
Total systematic uncertainty	0.017	0.012

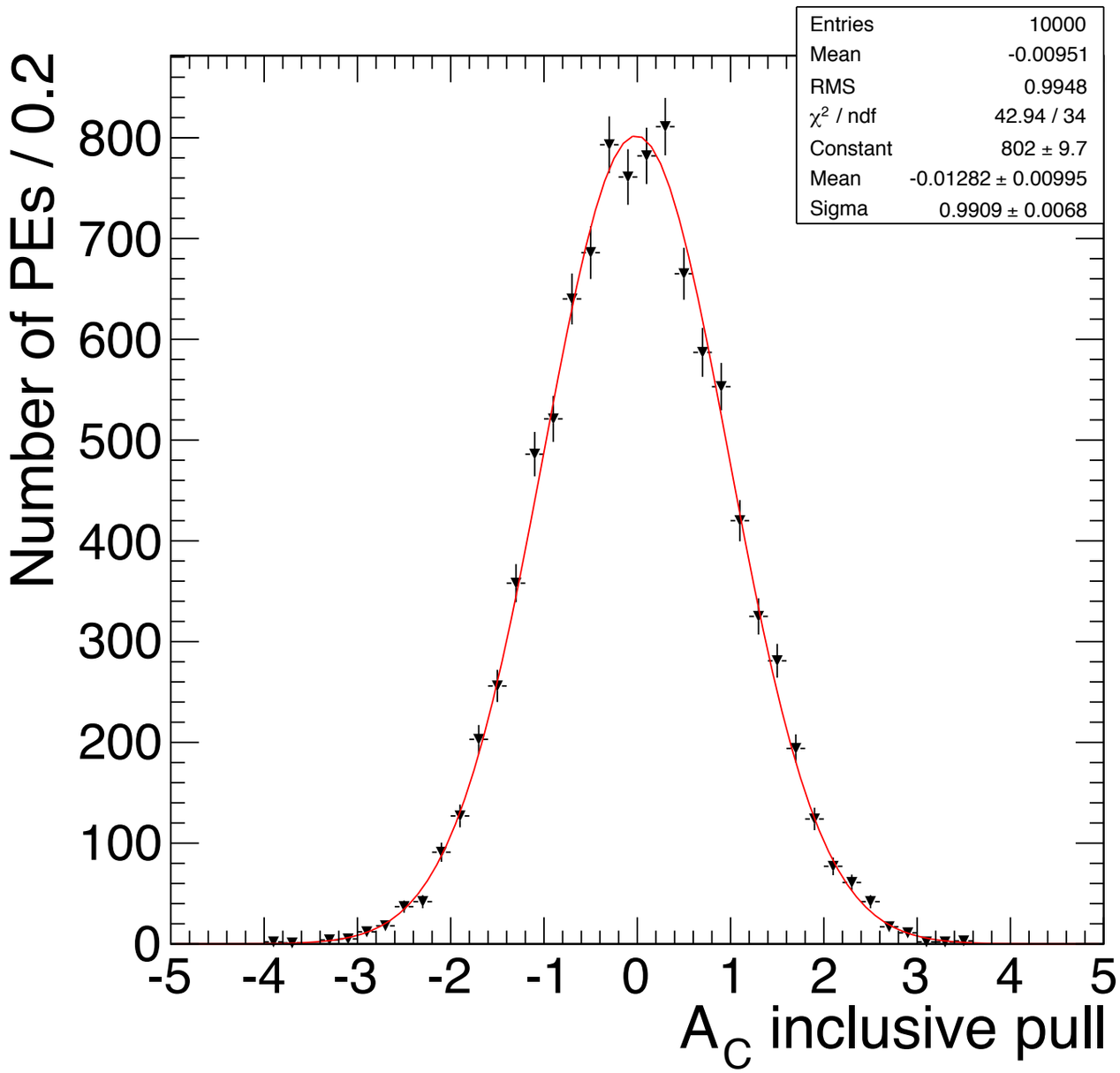
Table 3: Systematic uncertainties on the inclusive unfolded values of A_C and A_{lepC}.

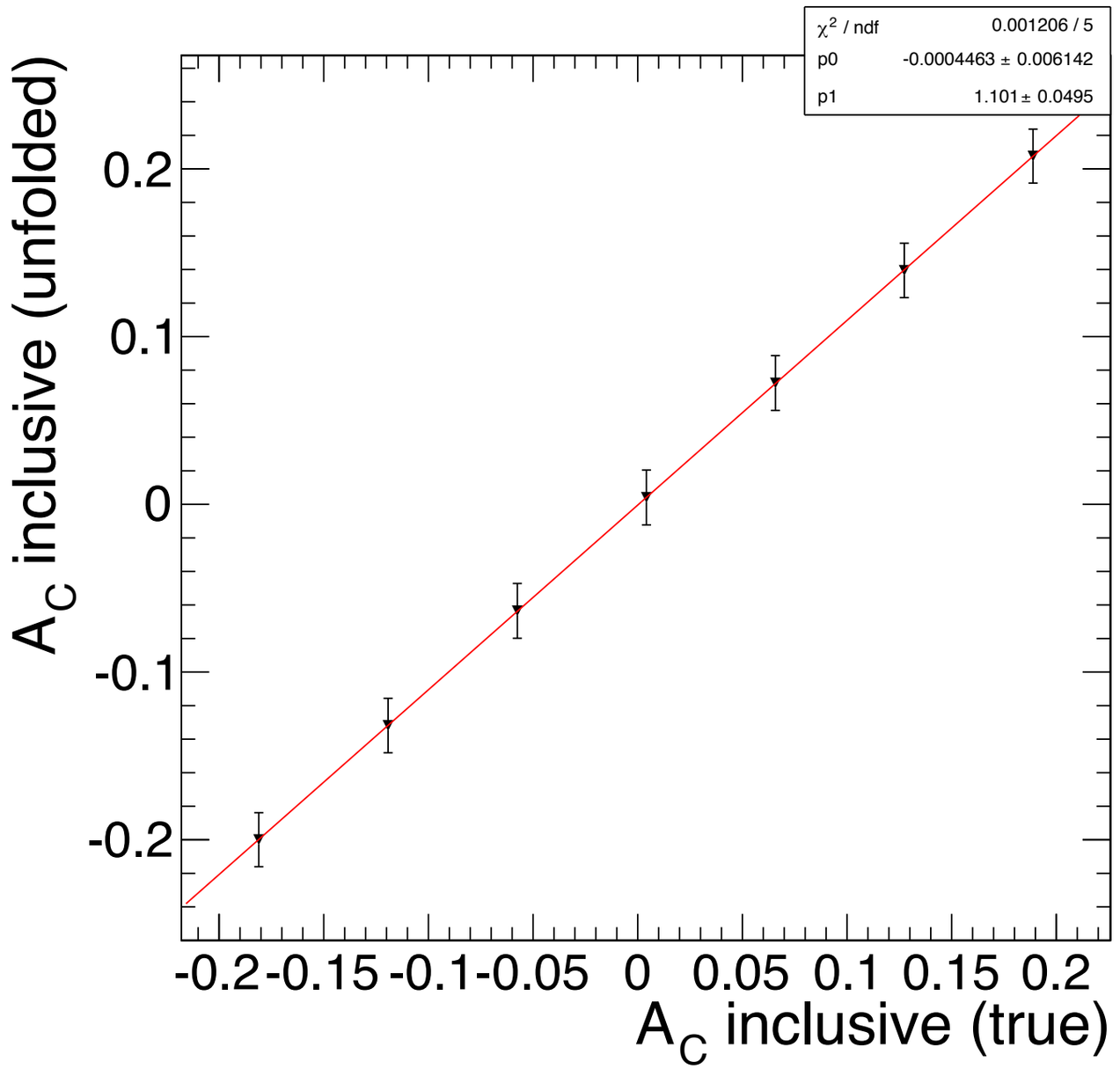
1D linearity (A_{lepC})





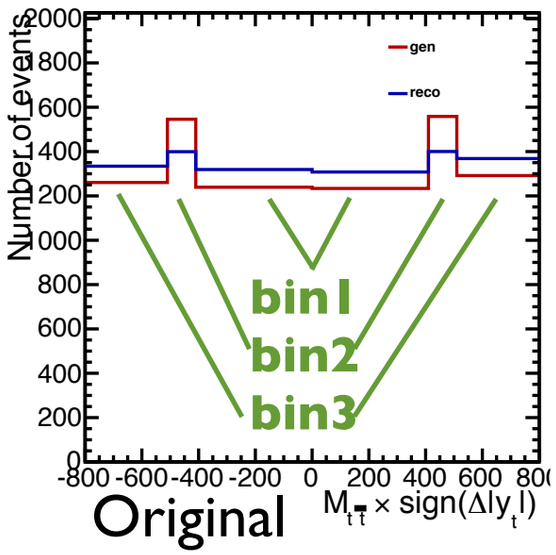
1D linearity (A_C)

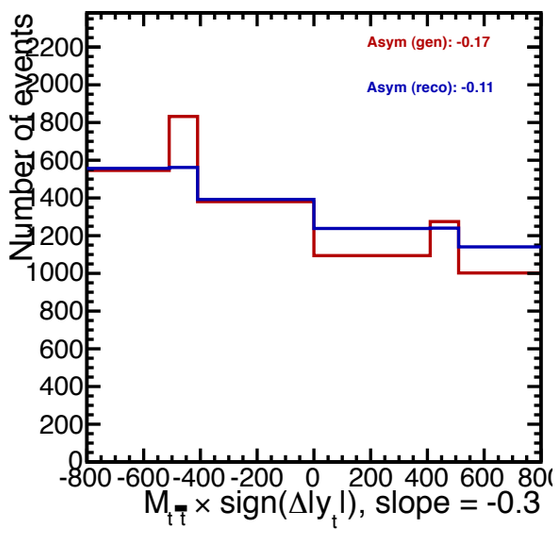


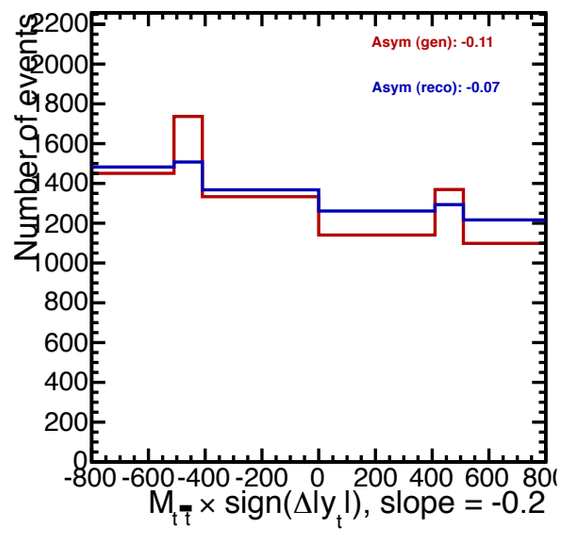


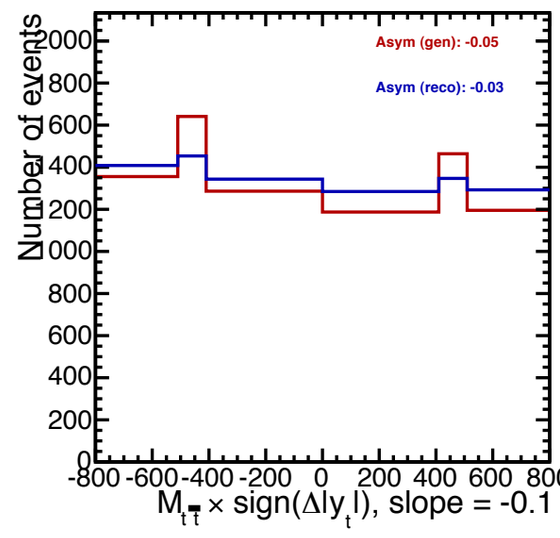
Weighted distributions with continuous slope in $\Delta|y_t|$

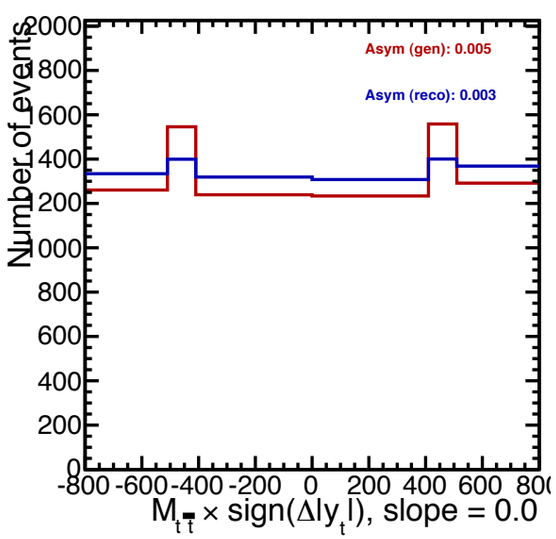
- plots show original results with weight = $1+\text{slope}*\Delta|y_{gen}|$
- Inclusive reco-level asym is **~2/3***gen-level asym, prior to unfolding (see numbers on plots)

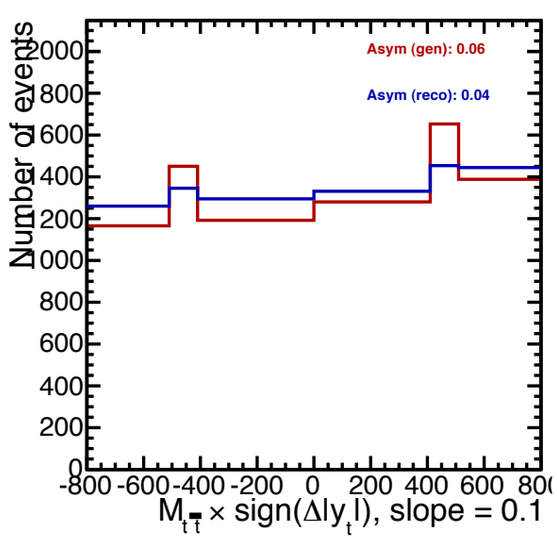


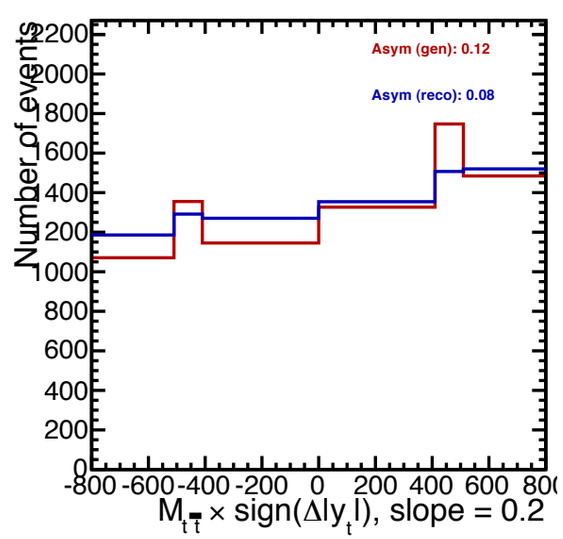


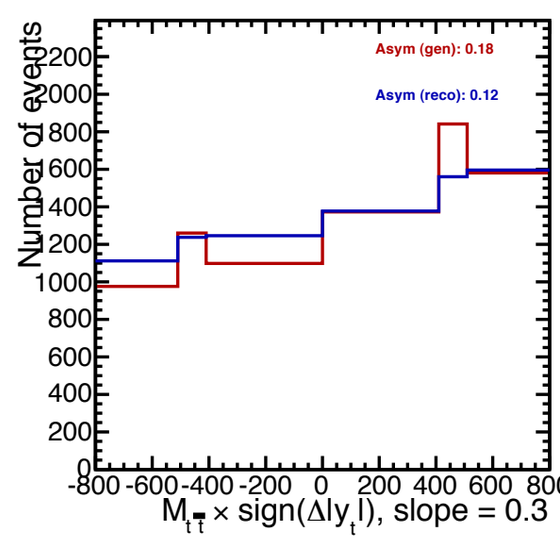












Weighted distributions with slope in sign($\Delta|y_t|$)

- weight 1+0.5*slope*sign($\Delta|y_{gen}|$)
- Reco-level asym is now only ~**1/2***gen-level asym, prior to unfolding
  - => reco distribution responds very differently to the different types of slope, but the effect of applying our "2-bin" response matrix will be the same in both cases
  - => expect p1 (inclusive) to be (1/2)/(2/3)*1.37 ~= 1.0 when we add a slope in sign($\Delta|y_t|$) instead of $\Delta|y_t|$, as is observed on slides 6 and 8