Data/MC Discrepancy in Average Track Multiplicity and Its Impact on SumEt

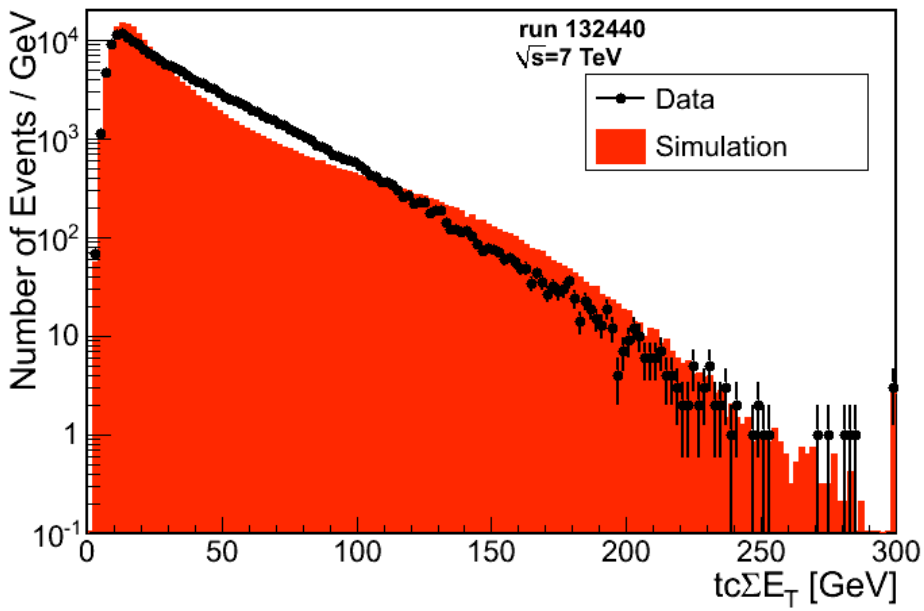
W. Andrews, D. Evans, F. Golf, J. Mulmenstadt, S. Padhi, Y. Tu, F. Wurthwein, A. Yagil -- UCSD
D. Barge, C. Campagnari, P. Kalavase, D. Kovalskyi, V. Krutelyov, J. Ribnik -- UCSB
L. Bauerdick, I. Bloch, K. Burkett, I. Fisk, *Y. Gao*, O. Gutsche, B. Hooberman -- FNAL

1. Release: 356
2. Data (132440/477): /MinimumBias/Commissioning10-*Apr1Skim_GOODCOLL*-v1/RAW-RECO
3. MC: /MinBias/Spring10-START3X_V26A_356ReReco-v1/GEN-SIM-RECO
4. MET values are taking out of the box without additional cleaning

Data/MC Discrepancy in MET SumEt

- Large data/mc discrepancy is observed in *SumEt (not MET)*

http://indico.cern.ch/conferenceDisplay.py?confId=89806



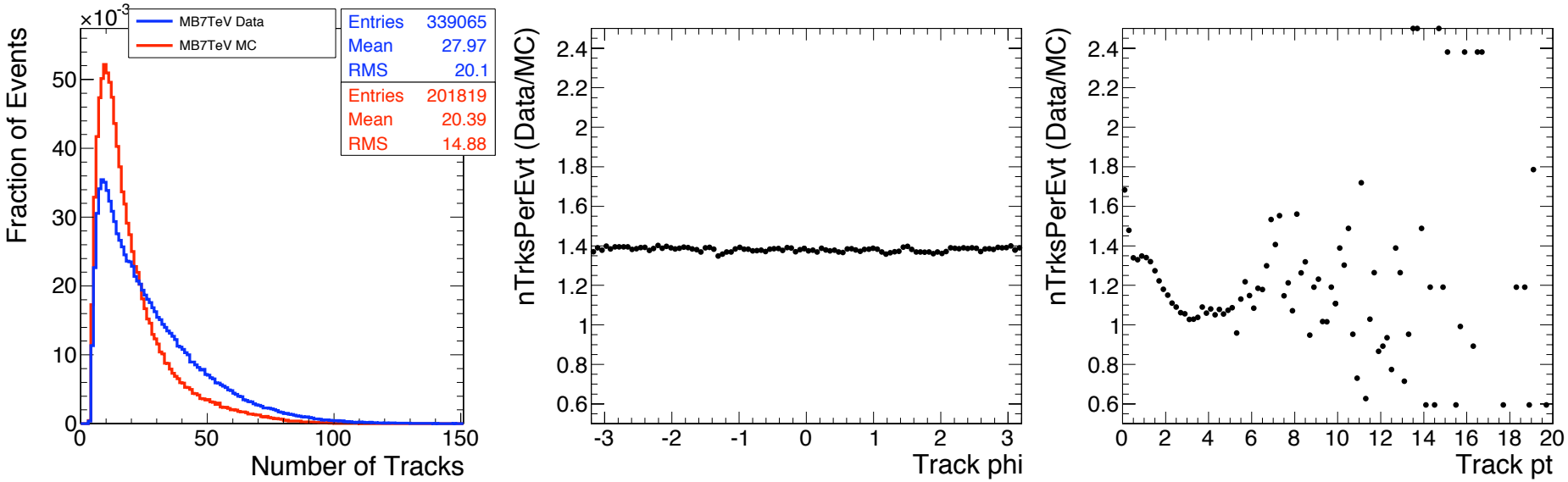
A few facts to consider:

1) It is present in all MET algorithms => not a tracking reconstruction issue
2) First time @ 7TeV - MC has never been tuned at this scale
3) The CMS Pythia MC was seen not well tuned in 900/2360 GeV data. *It simulates less soft particles* and the data excess doesn't have a preferred direction
4) The data/MC discrepancy in nTracks is larger at 7 TeV than at 900 GeV

7TeV Data/MC nTracks Difference (1/2)

- Applied the track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5%*"

selecting primary tracks

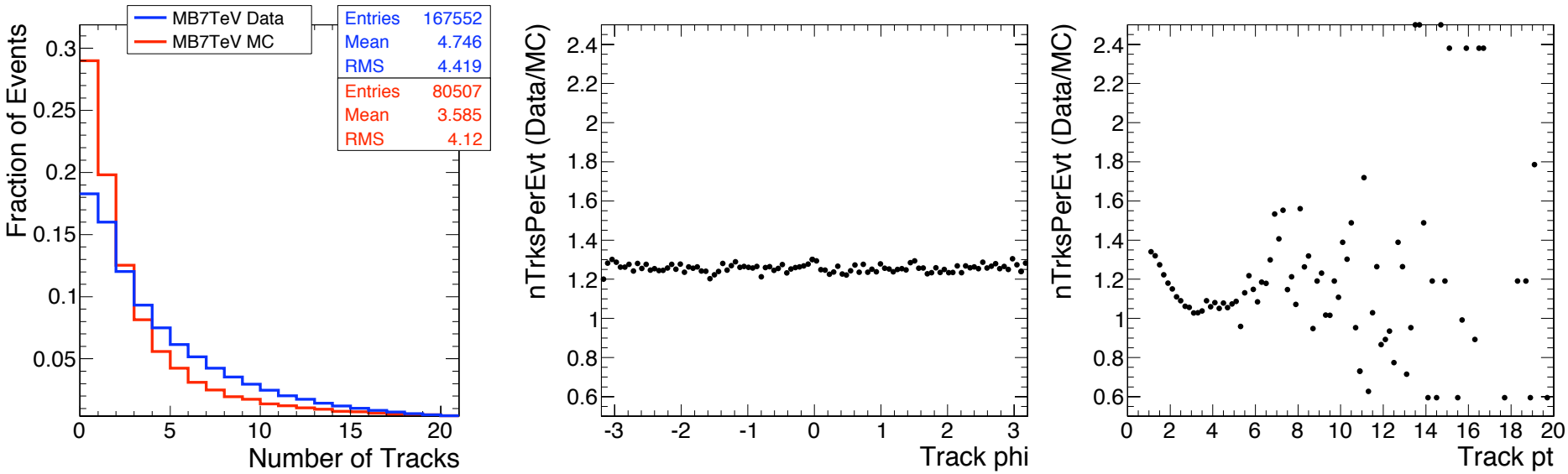


- The kinematic distributions per track are consistent between data/MC
- <nTracks> in the data is on average 37% (*~8 tracks*) larger than the MC
- The data/MC excess is mainly in low pT < 4GeV region
- The nTracks distribution comparison in high pT region is limited by statistics
- Both absolute and relative data/mc discrepancies are larger in 7 TeV than in 900GeV

7TeV Data/MC nTracks Difference (2/2)

- Applied the track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5% && pT > 1GeV*"

  where track reconstruction is better understood



- <nTracks> in the data is ~25%(*1-2 tracks*) more than the MC
- Relative data/mc discrepancy at 7 TeV is roughly the same as at 900GeV
- Absolute data/mc discrepancy is larger at 7 TeV than in 900GeV

Hypothesis: MC has an overall particle deficit compared with nature

Test: Will correcting nTracks distribution improve data/MC agreement of sumEt?

Re-Weight Data nTracks Distribution (1/5)

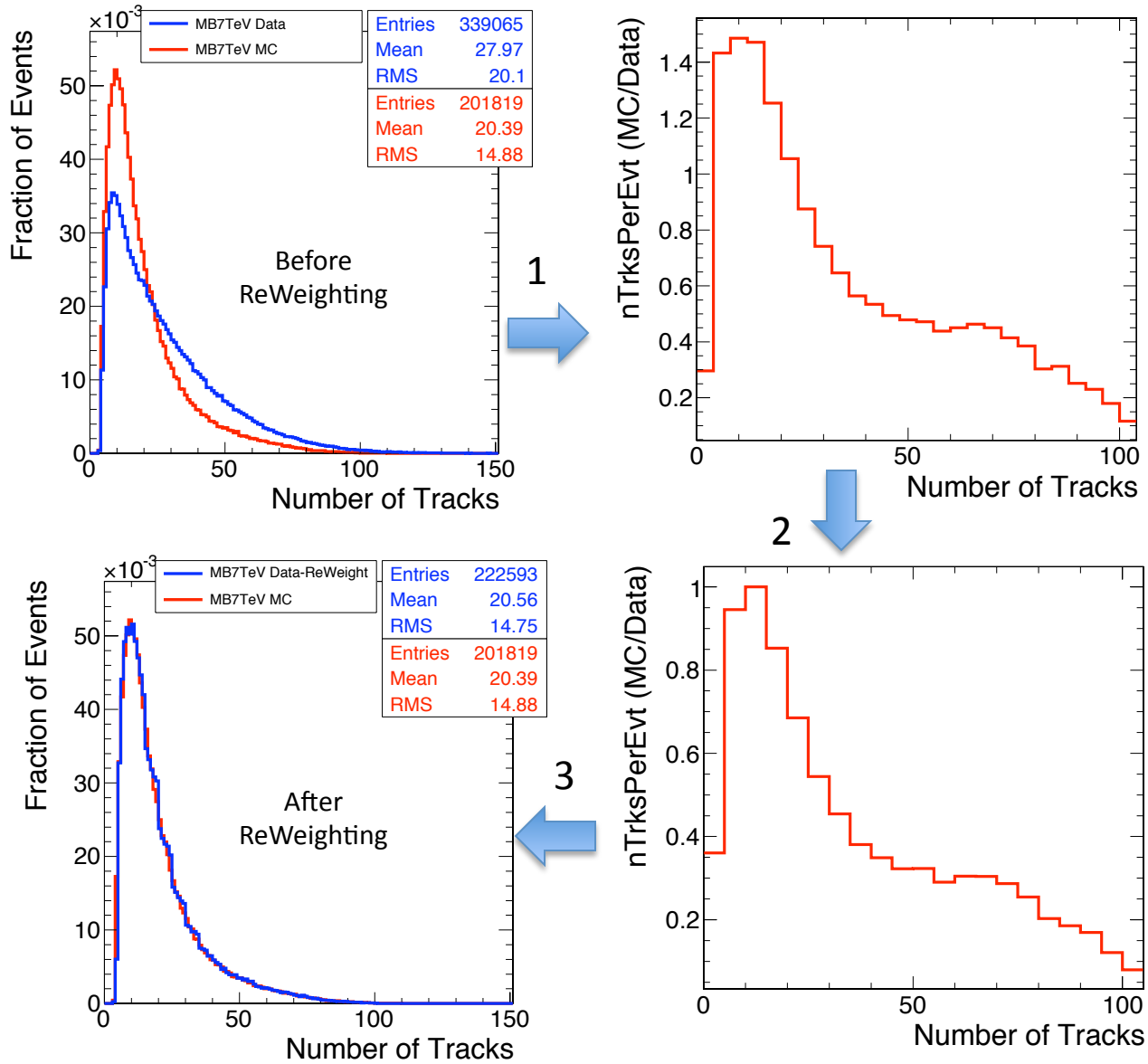
- Since data has a longer tail in nTracks distribution, for this test, *it is simpler to scale the data down to match the mc*

(*results of reweighting the mc to match the data are in the backup slides*)

ReWeighting Procedure

1. Divide the data/MC "nTracks" histograms
2. Scale the largest bin in the data/MC ratio to 1 , and get a response histogram W(nTracks)
3. For each event, generate a random number "rd" btw (0,1)

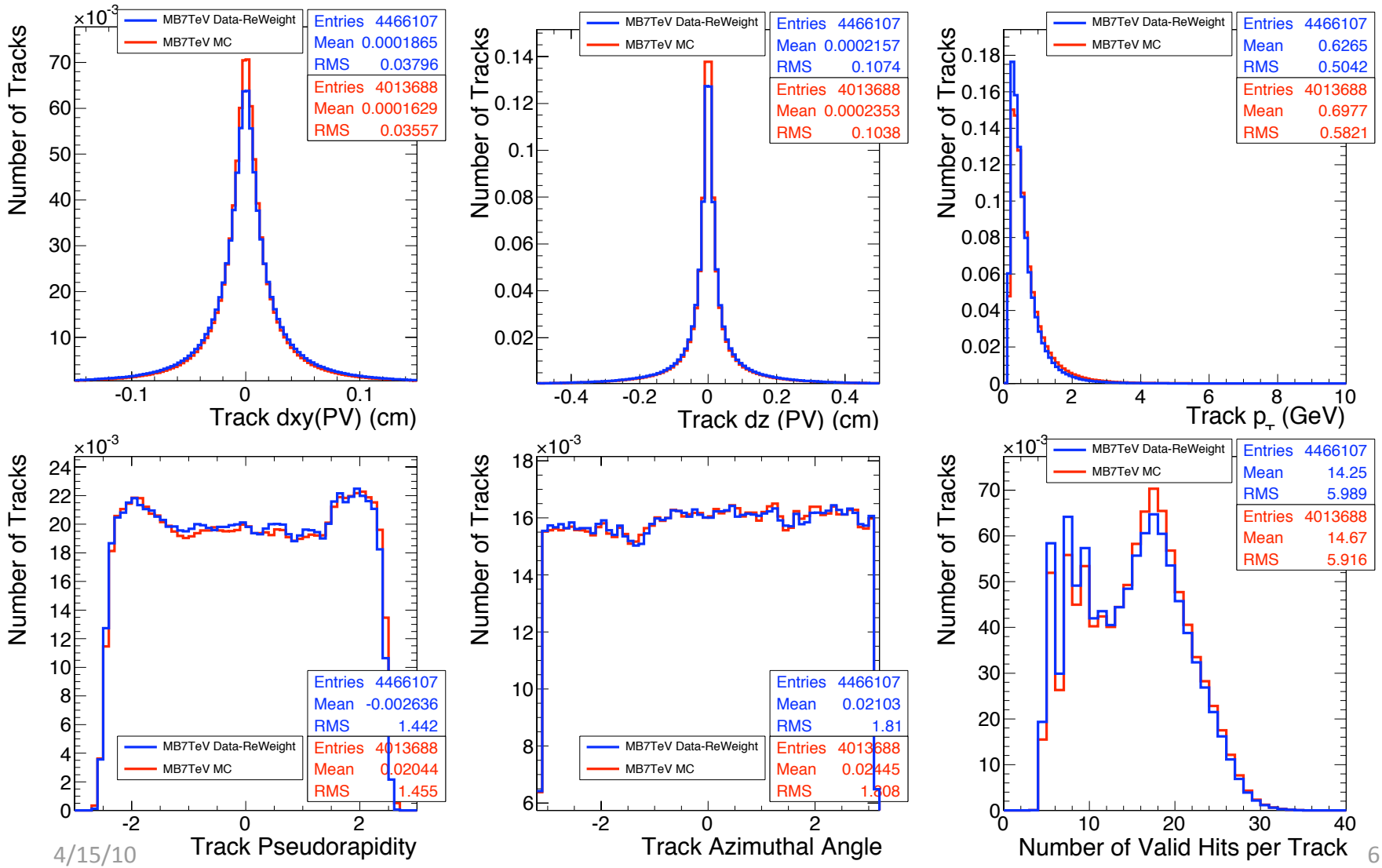
   if rd > W(nTracks), discard the event



Re-Weight Data nTracks Distribution (2/5)

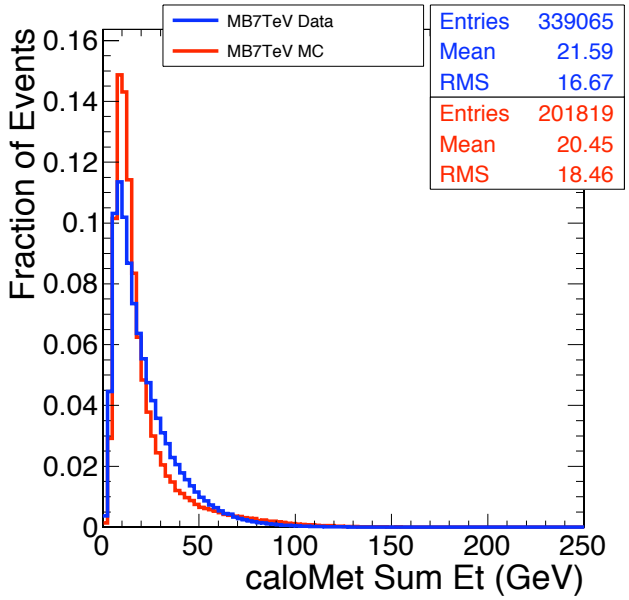
- Applied the track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5%*

The kinematic distributions per track are not affected by the ReWeighting

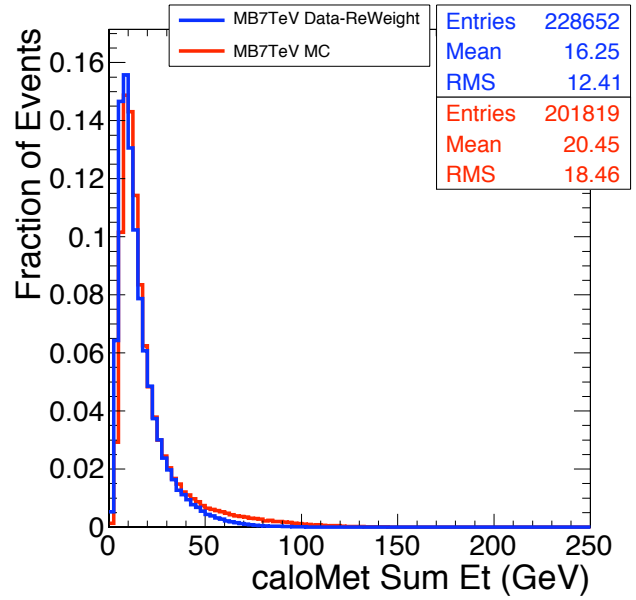


Re-Weight Data nTracks Effect to CaloMET (3/5)

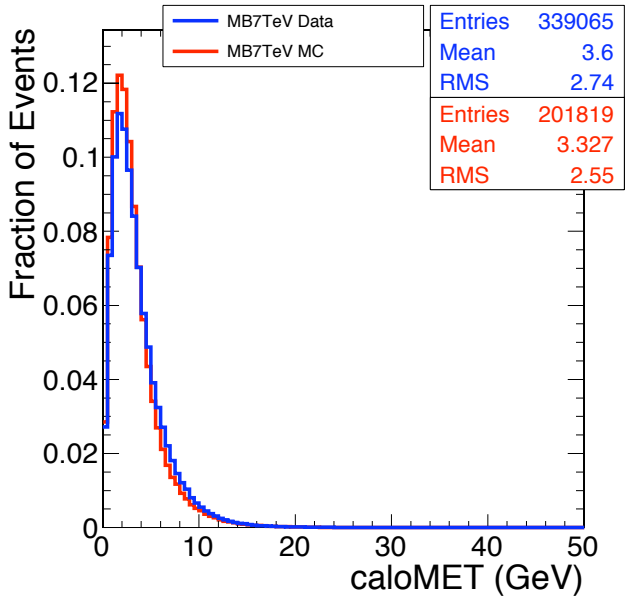
- Sum Et



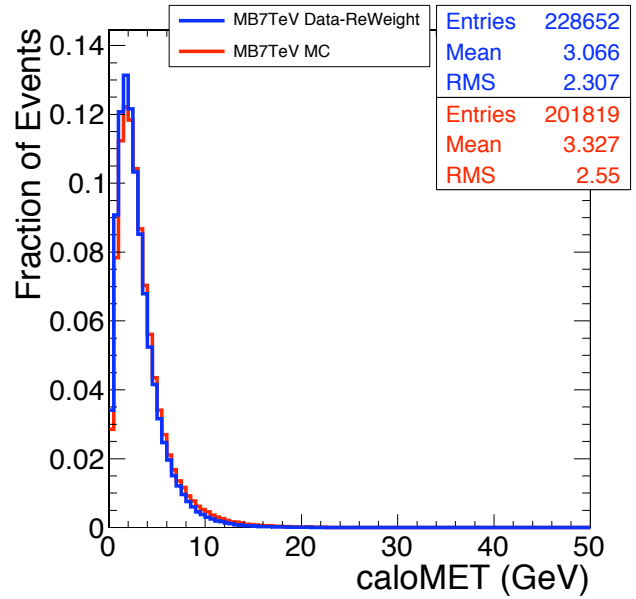
Reweight Data



- MET

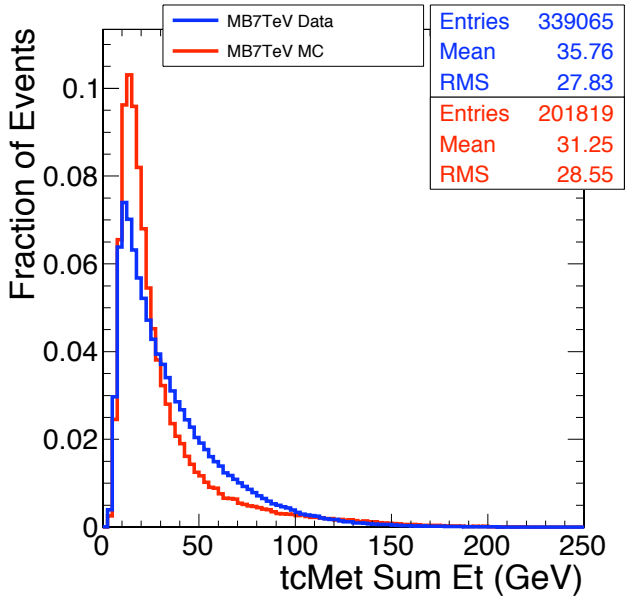


Reweight Data

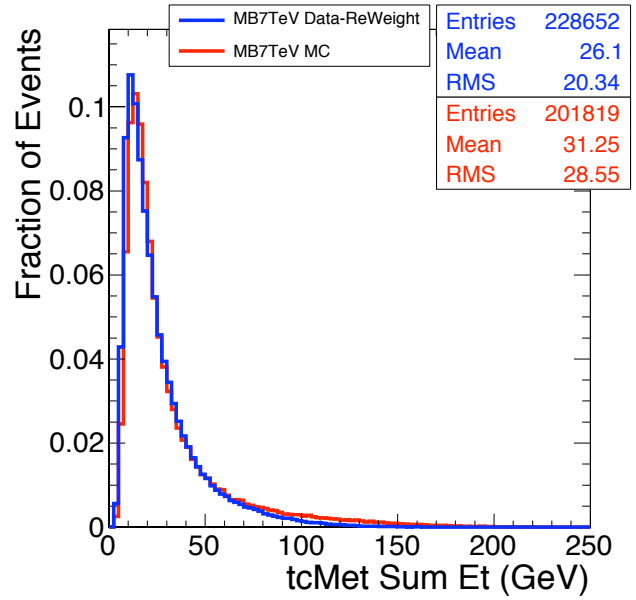


Re-Weight Data nTracks Effect to tcMET (4/5)

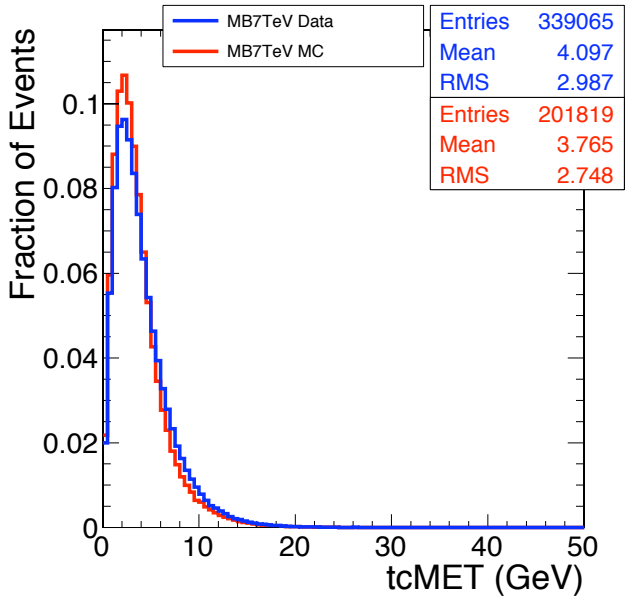
- Sum Et



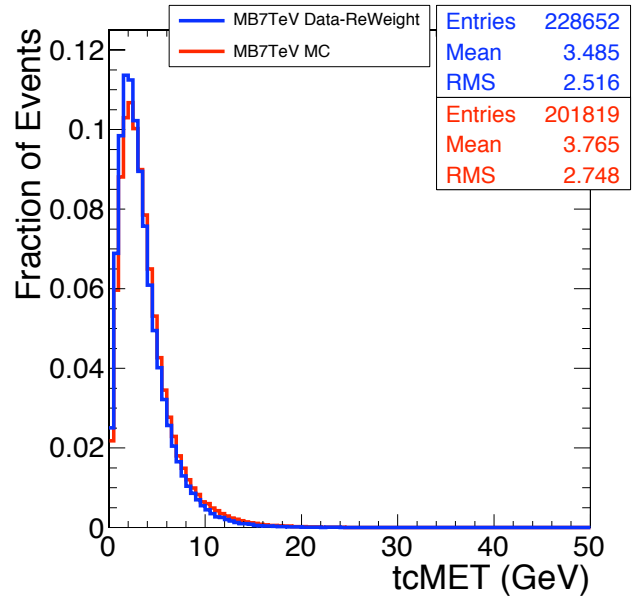
Reweight Data



- MET

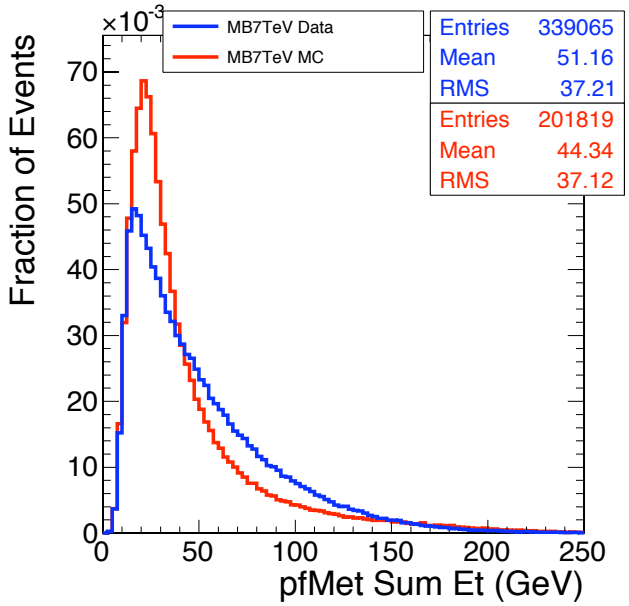


Reweight Data

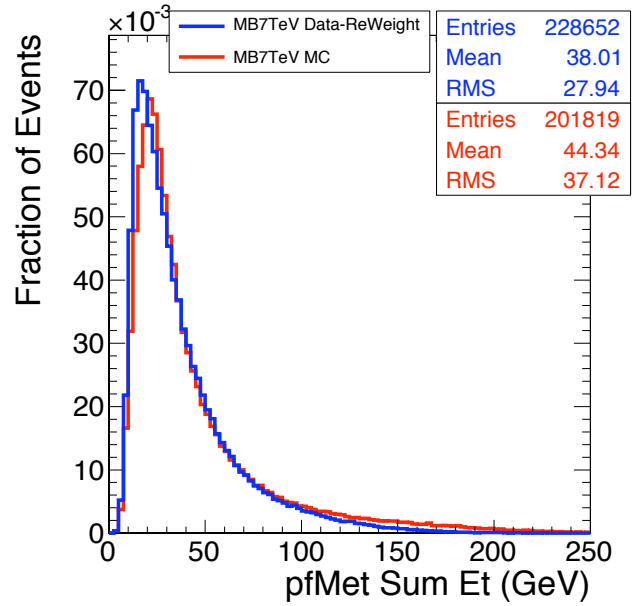


Re-Weight Data nTracks Effect to pfMET (5/5)

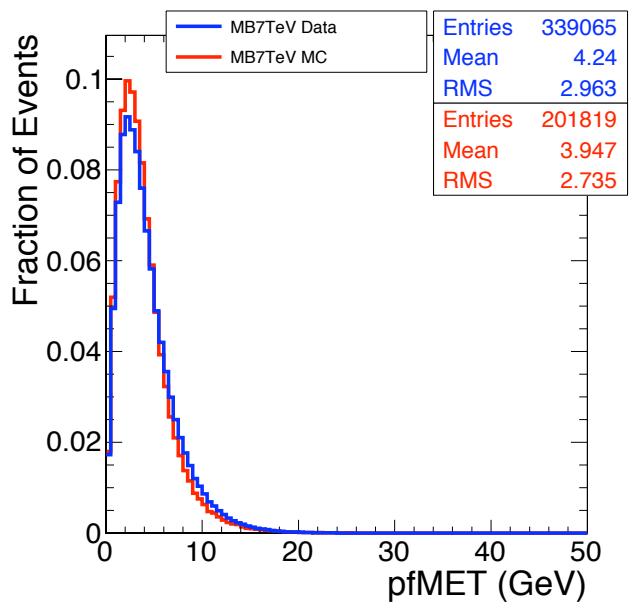
- Sum Et



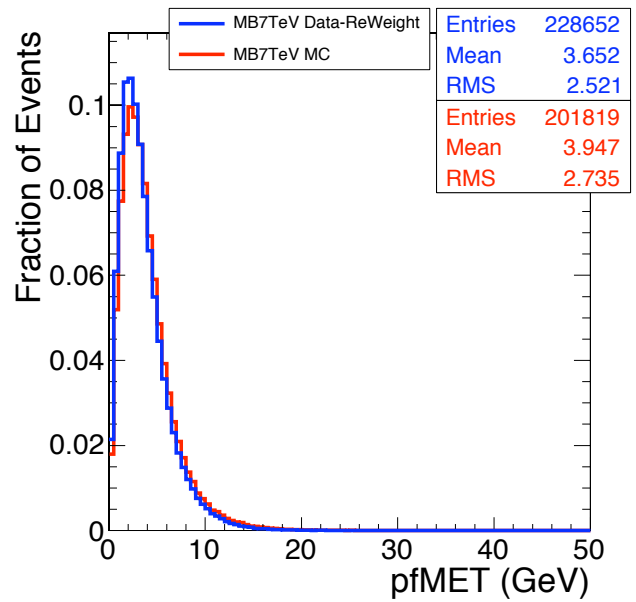
Reweight Data



- MET



Reweight Data



Conclusion

- The data/MC excess in <nTracks> @7 TeV has a larger discrepancy than @900GeV

  ** Data has on average ~ 37 % (~ 8 tracks) more tracks than MC*

  ** Data has on average ~ 25% (> 1 track) more tracks with pT > 1 GeV*

- The effect of the <nTracks> data/MC discrepancy to SumEt was investigated

  ** The data was reweighted to match the MC nTracks distribution*

  ** It was observed that SumEt agree much better between data/MC in lower end*

  ** The MET was not affected by this reweighting*

- Correcting the data/MC discrepancy in # charged particles (nTracks) improves the data/MC correspondence in an independent measurement (SumEt)

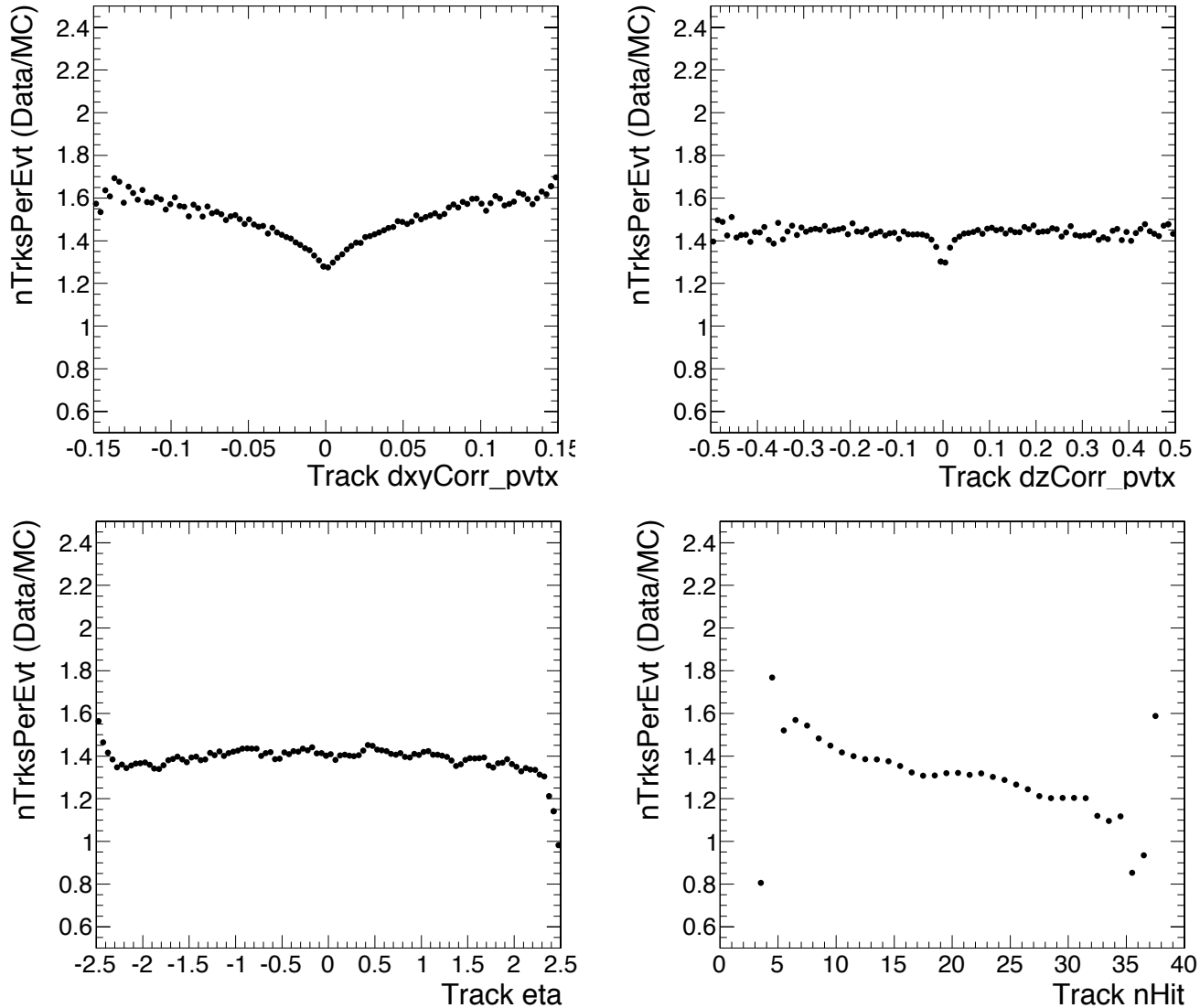
- *Thus, the nTracks and SumEt data/MC discrepancies are a result of the Pythia MC simulating less soft particles with pT < 4 GeV.*

- *To resolve the discrepancy, the MC needs to be retuned!*

BACKUP SLIDES

7TeV Data/MC nTracks Difference (3/3)

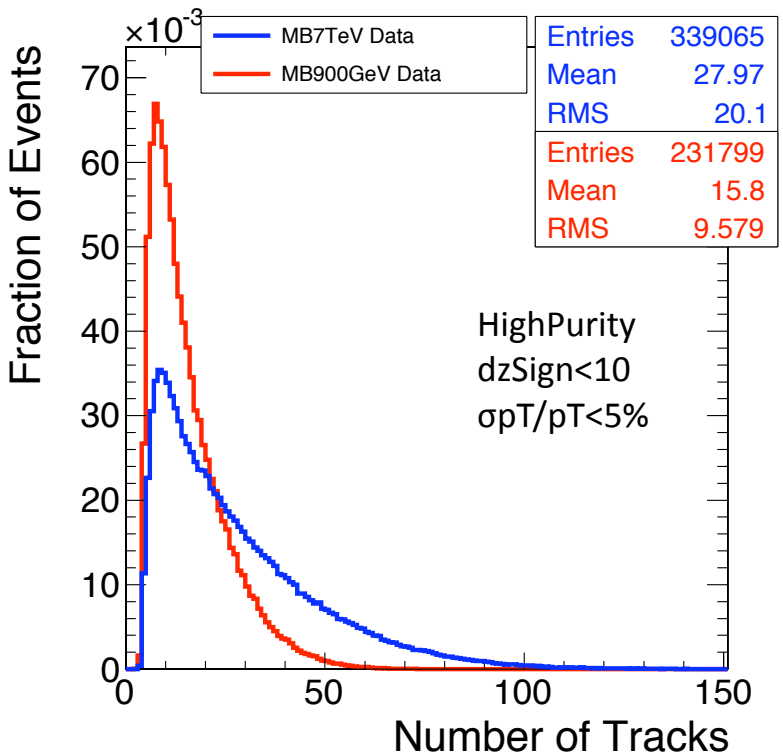
- Track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5%*"

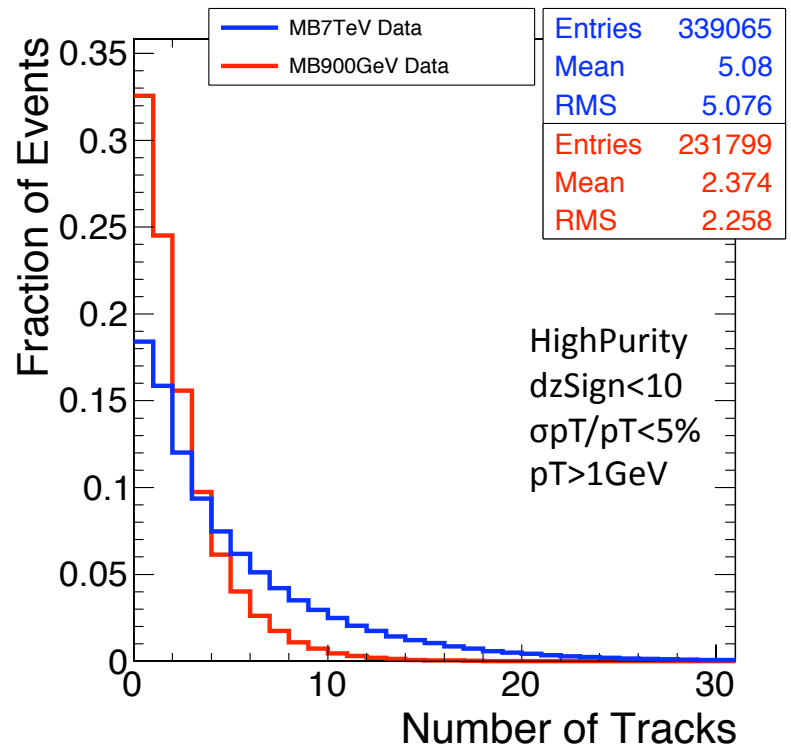


The excess in Data is not limited to any preferred geometrical region

Compare 7TeV and 900 GeV nTracks

- We *expect* <nTracks> increase roughly *(80%-100%)* from 0.9->7 TeV, *observe 80%*





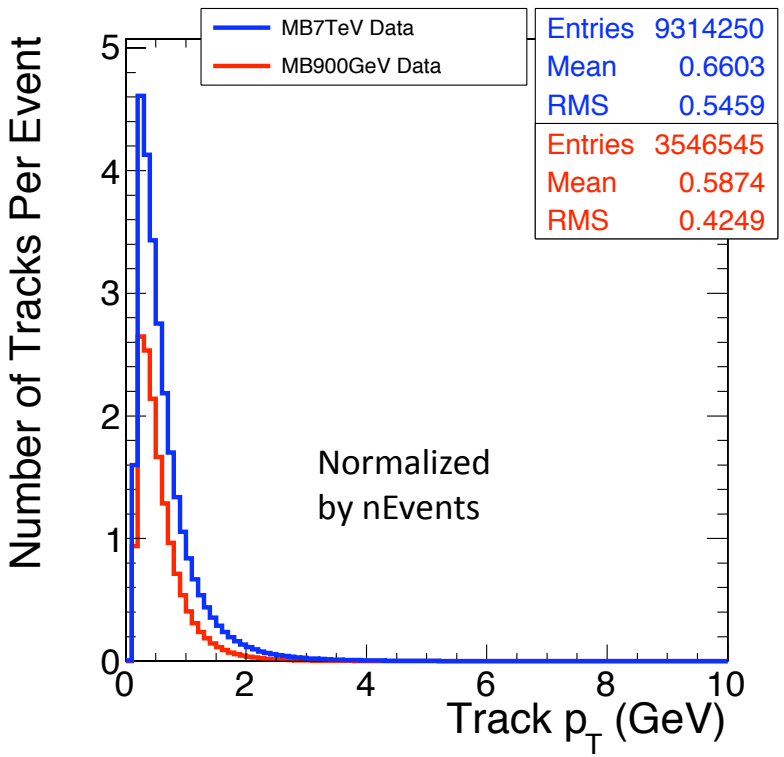
- *<nTracks>(√s) two good fit functions:* *arXiv:0912.0023[hep-ex]*

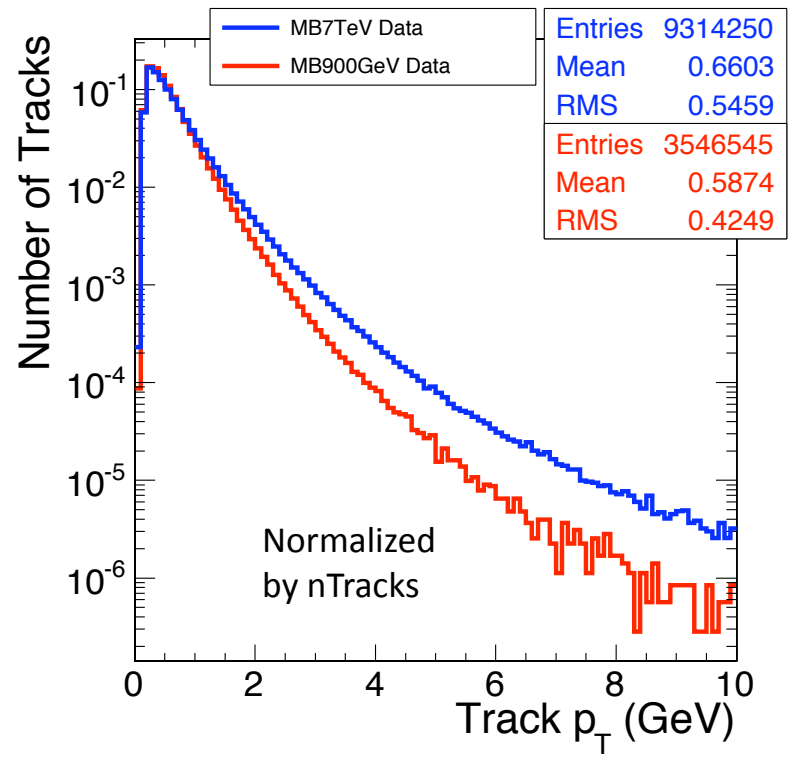
1) $a + 2b\ln(\sqrt{s}) + 4c\ln^2(\sqrt{s}),\ \ (a = 16.65, b = -3.147, c = 0.334)$ <nTracks>(7TeV)/<nTracks>(900GeV) = 1.8

2) $b\sqrt{s}^{2n},\ \ (b = 3.102, n = 0.178)$ <nTracks>(7TeV)/<nTracks>(900GeV) = 2.1

Compare 7TeV and 900 GeV pT

- From 0.9->7 TeV, we also expect more high pT tracks

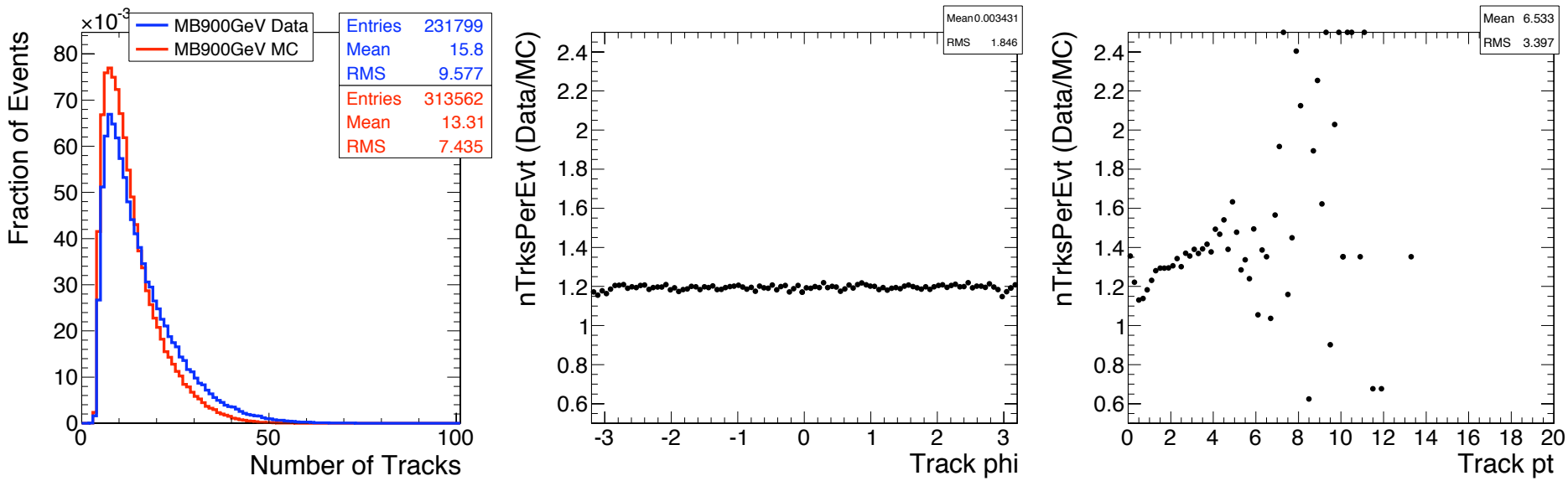




- Track multiplicity and pT at 7 TeV data looks reasonable compared to 900GeV

900GeV Data/MC nTracks Difference (1/2)

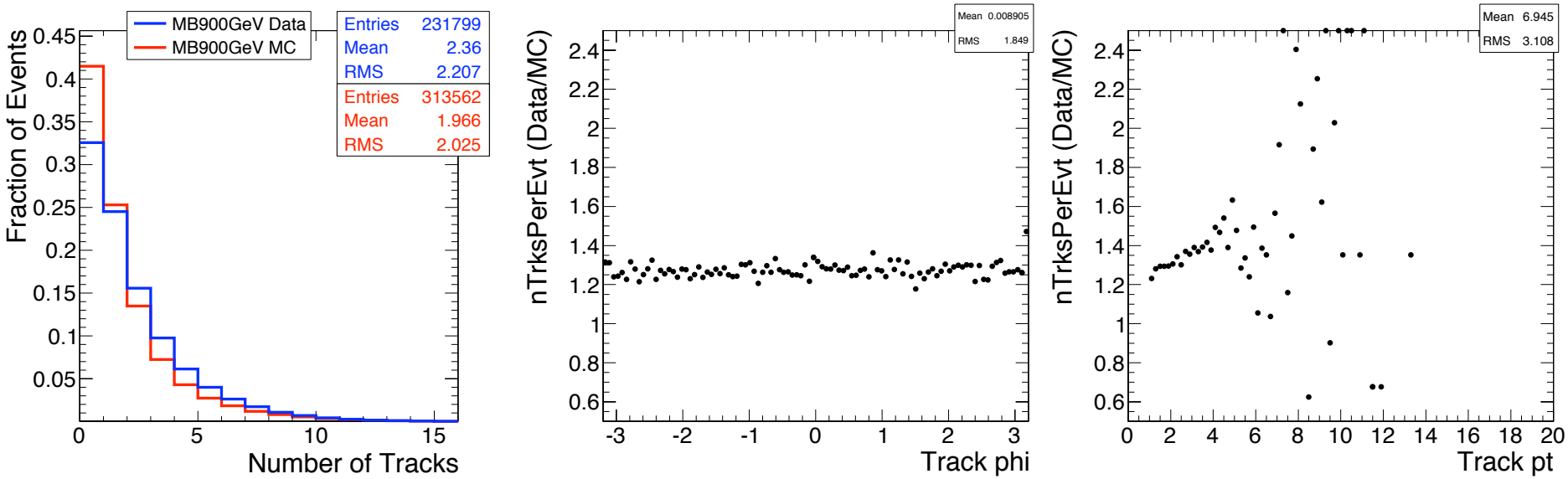
- It has been known that pythia MC used in CMS not tuned well, simulating less soft particles
- Track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5%*"



- The track distributions are consistent on track base (see TRK-10-001)
- <nTracks> in Data is ~ 20% (*~ 2-3 tracks*) more than MC, excess in pT < ~ 4GeV
- High pT (> 4 GeV) tracks are limited by statistics.

900GeV Data/MC nTracks Difference (2/2)

- Track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5% && pT > 1GeV*"

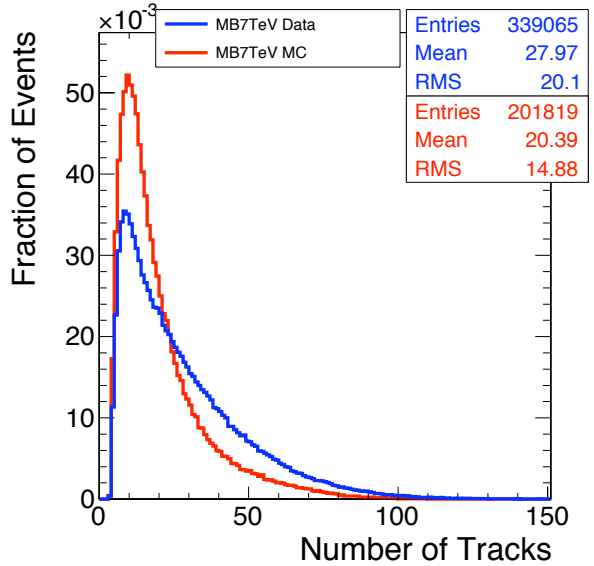


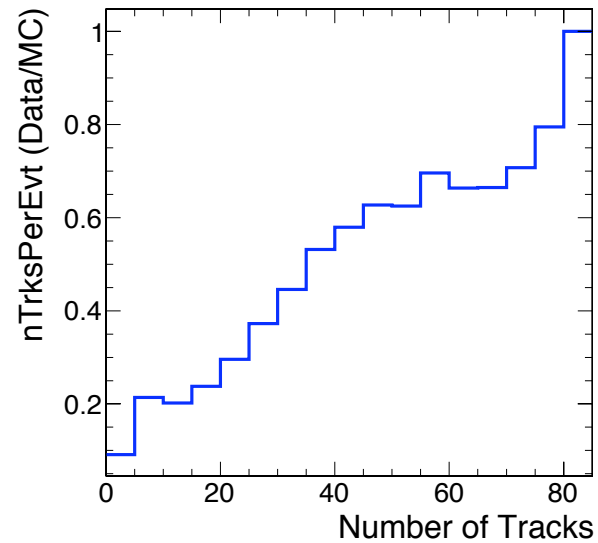
- The track distributions are consistent on track base
- <nTracks> in Data is ~ 25% (*< 1 track*) more than MC, excess in pT < ~ 4GeV
- High pT (> 4 GeV) tracks are limited by statistics.

Re-Weight MC nTracks Distribution (1/5)

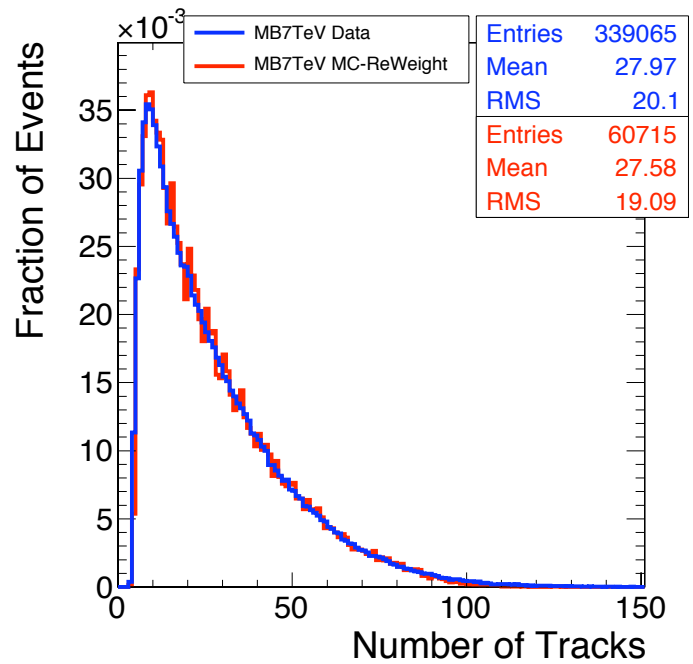
- Divide Data/MC nTracks histograms, and scale the largest Ratio to 1

- *Correct up to nTracks = 80*



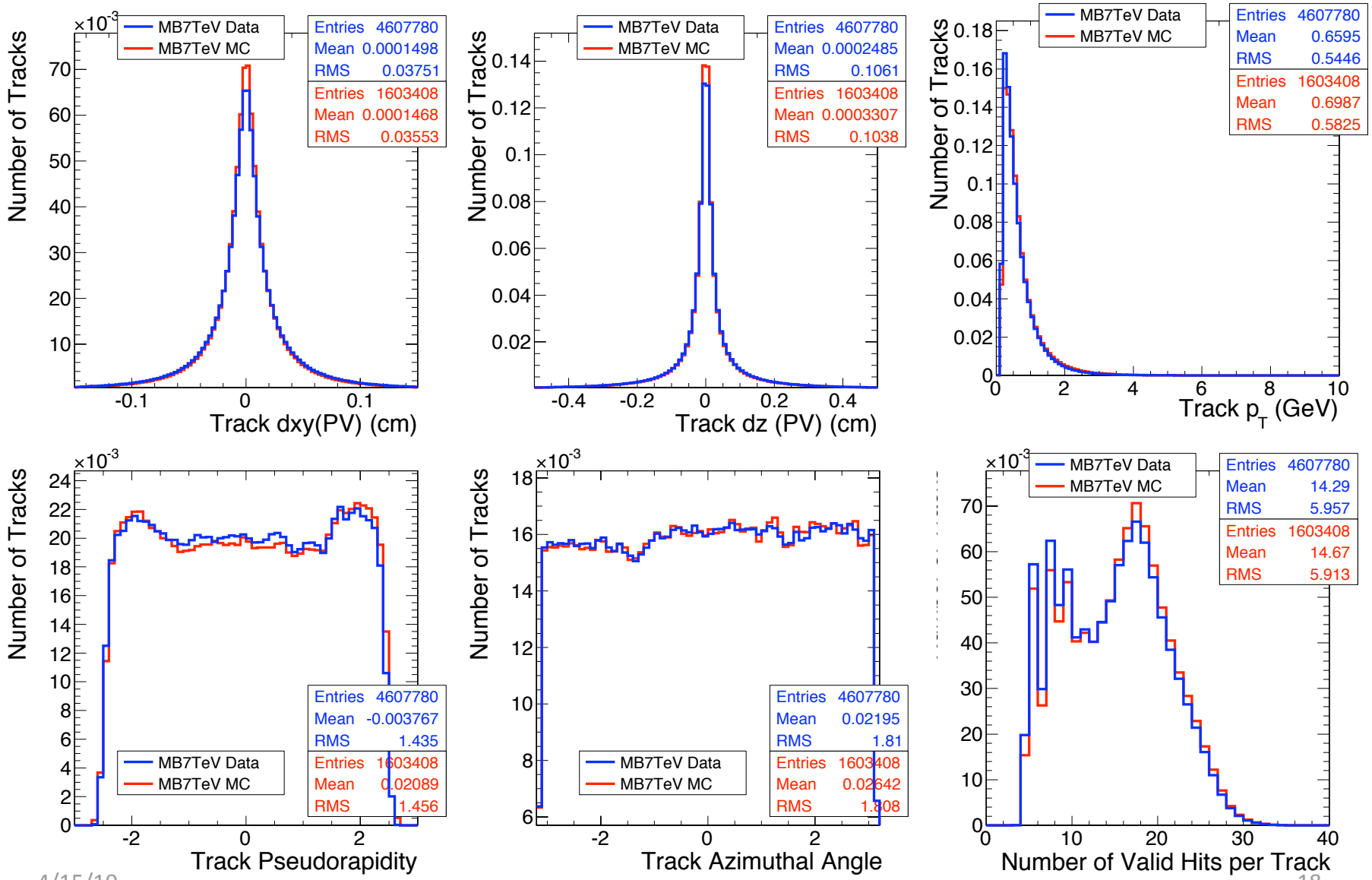


- After reweighting, obtained consistent nTracks distribution



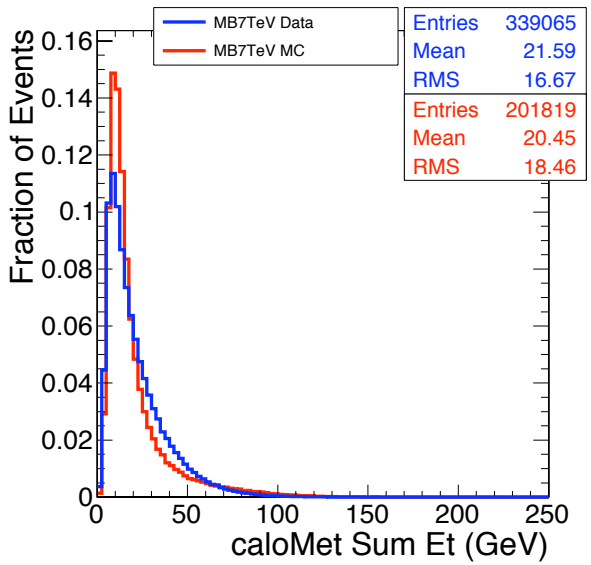
Re-Weight MC nTracks Distribution (2/5)

- Track selection "*HighPurity && |dzSign| < 10 && σpT/pT < 5%*", MC reweighted

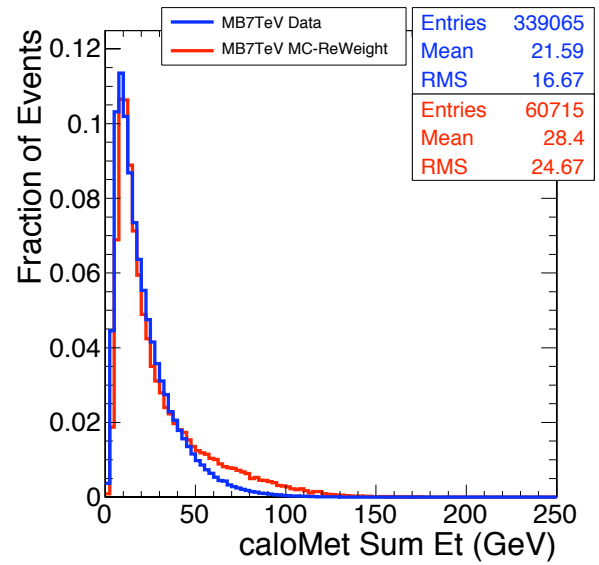


Re-Weight MC Effect to CaloMET (4/5)

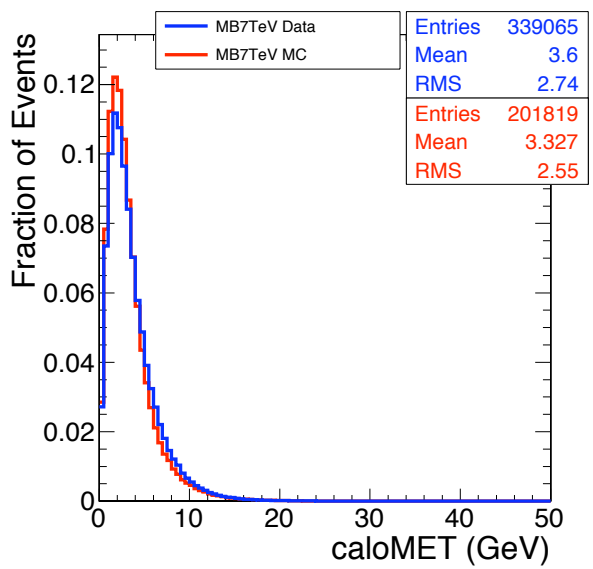
- Sum Et



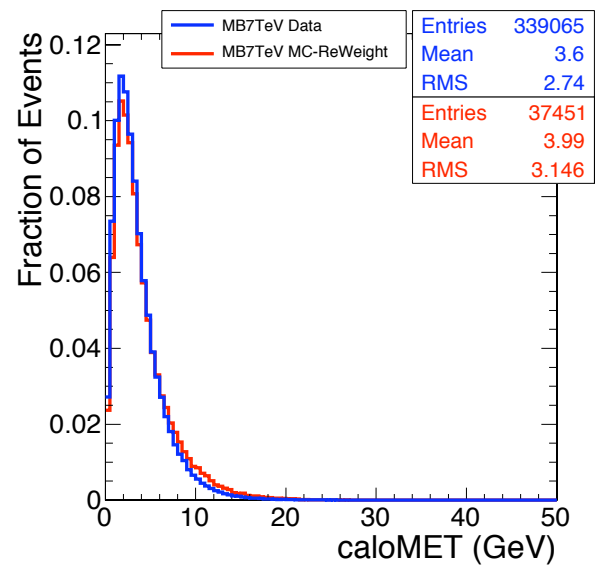
Reweight MC



- MET



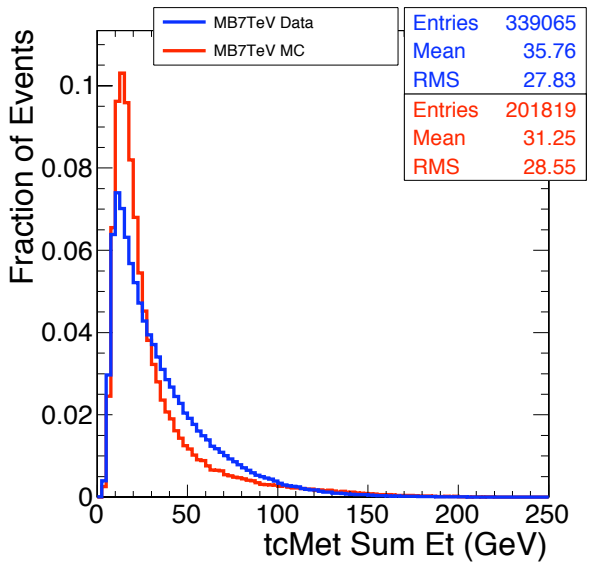
Reweight MC



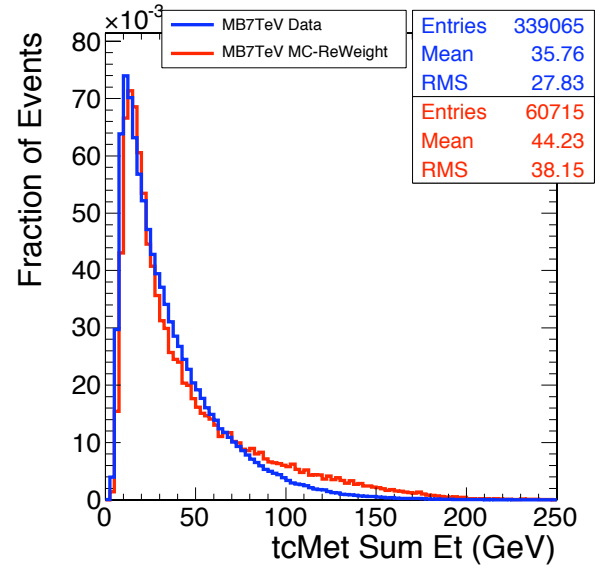
ReWeight up to nTracks = 80, creating artificial long tail in MC

Re-Weight MC Effect to tcMET (3/5)

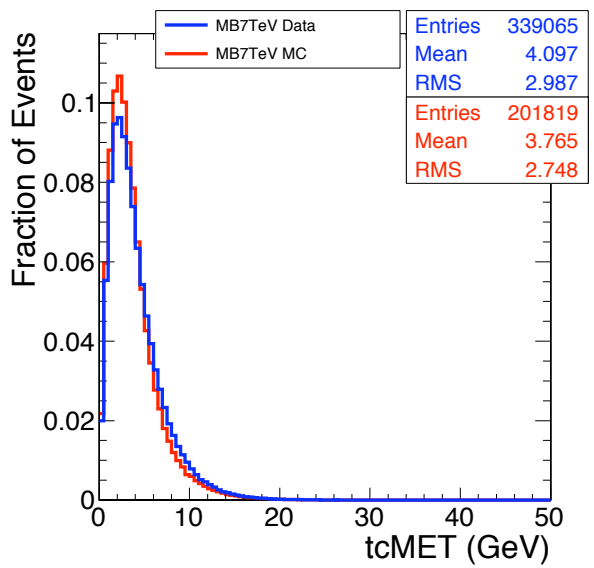
- Sum Et



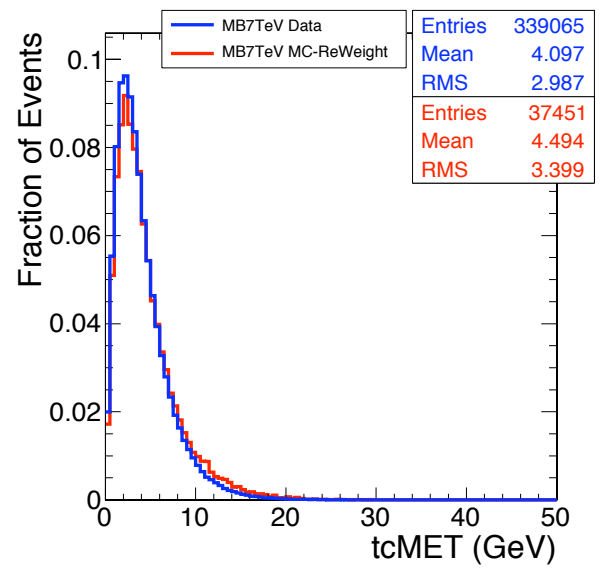
Reweight MC



- MET

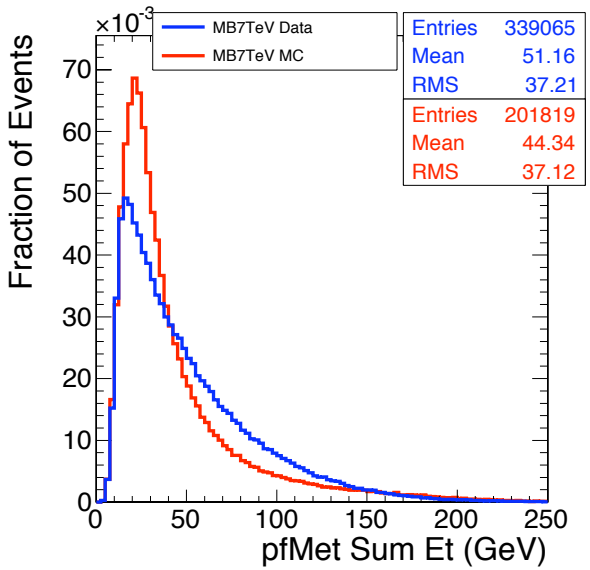


Reweight MC

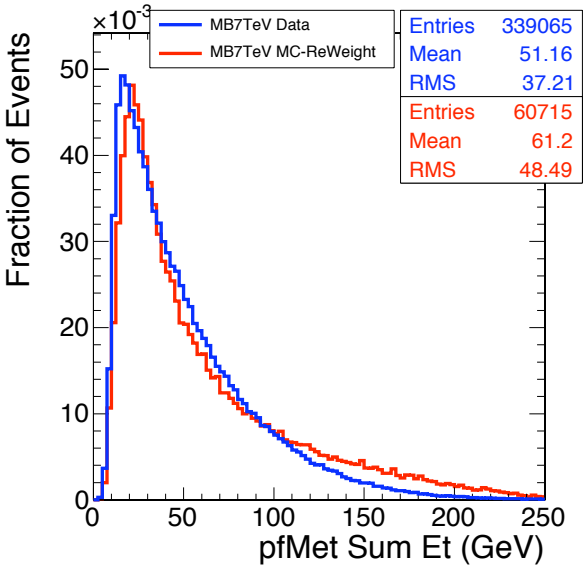


Re-Weight MC Effect to pfMET (5/5)

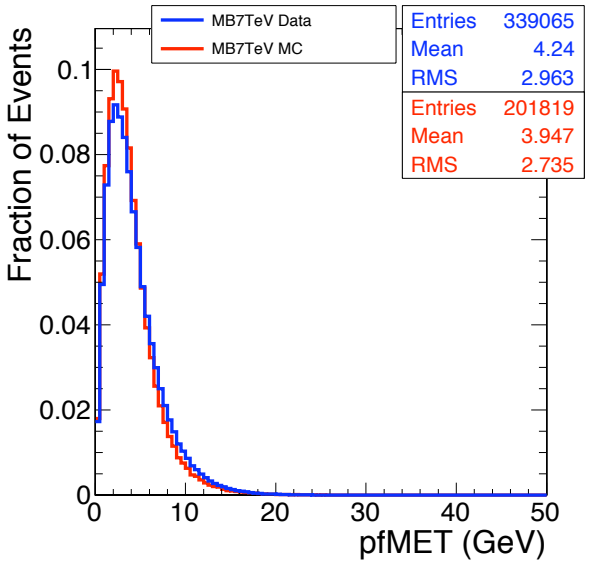
- Sum Et



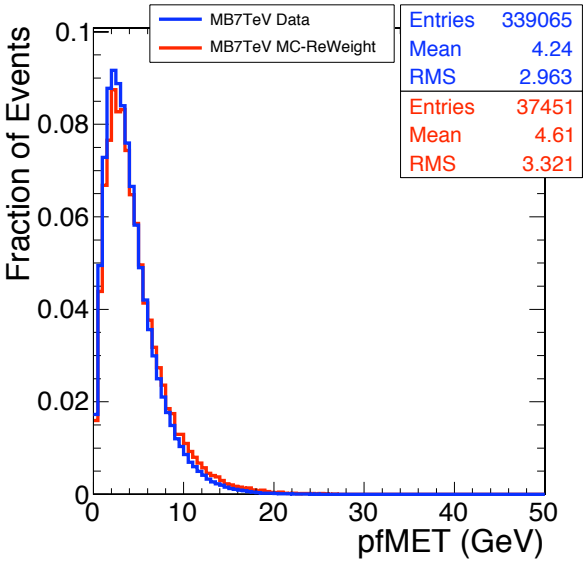
Reweight MC



- MET



Reweight MC



ReWeight up to nTracks = 80, creating artificial long tail in MC