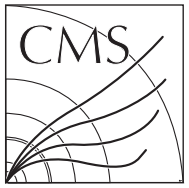


The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only

February 5, 2010

Version 0.1

Studies of Jet Id in 2009 *MinimumBias* Collision Data

Jason St. John
Boston University, Boston, MA, USA

Sertac Ozturk
Cukurova University, Adana, Turkey

Robert M. Harris, Kalanand Mishra, Seema Sharma
Fermi National Accelerator Laboratory, Batavia, IL, USA

Chiyoung Jeong, Sungwon Lee
Texas Tech University, Lubbock, TX, USA

Marek Zielinski
University of Rochester, Rochester, NY, USA

Abstract

The aim of this study is to explore the chosen jet identification criteria for CMS calorimeter jets used in the inclusive jet and dijet analyses of *MinimumBias* collision data from December 2009 LHC run at $\sqrt{s} = 900$GeV. We present distributions of jet ID variables and measures of their correlations using both collision data and Monte Carlo simulation of the *MinimumBias* events. We present the survival fractions for the jets passing the selection criteria of the inclusive jet and dijet analyses, as a function of jet p_T.

Contents

1 Introduction

This study examines the cuts made on jet variables as studied in CMS AN-2010/019 "Inclusive Jet Studies with $\sqrt{s}$ = 900GeV Collisions" and CMS AN-2010/009 "Jet Commissioning using Di-Jet Events from 900 GeV and 2360 GeV Collision Data". In support of the cuts used in those notes, we present the fraction of jets and of dijet events to pass these cuts as a function of corrected jet p_T in the CMS collision data of 2009 at $\sqrt{s} = 900$ GeV and in Monte Carlo simulation. In passing, we also note a few features of the correlations among these cut variables.

1.1 Some definitions

We use the following technical terms throughout the text

1. p_T: transverse momentum of the jet (corrected, unless specified otherwise)
2. η: pseudo-rapidity of the jet calculated with respect to reconstructed primary vertex in the event
3. ϕ: azimuthal angle of the jet calculated with respect to reconstructed primary vertex in the event and in the range $-\pi$ to π
4. *EMF*: electromagnetic energy fraction of the jet
5. *HPD*: a hybrid photo diode collects signal from channels (η segments) of uniform ϕ in the hadronic calorimeter's barrel and endcap partitions.
6. *RBX*: a readout box housing HPDs of four adjacent ϕ-slices.

1.2 Survival fractions

1. *Jet survival fraction for a selection cut* = $\frac{\text{Number of jets passing basic cuts and the given selection cut}}{\text{Number of jets passing basic cuts}}$

Survival fractions defined in this way are the observed survival fractions of the cuts in the sample. When the survival fraction in data is very much less than the MC survival fraction, the cut rejects something reconstructed as jets that is not present in the Monte Carlo simulation of the 900 GeV *MinimumBias* events. Potentially, it is rejecting noise or other backgrounds not present in the simulated events.

2 Event selection and the Jet ID Variables

2.1 Basic event selection

Event selection follows exactly the practice detailed in CMS AN-2010/009 §3, and is restricted to the 900 GeV data and Monte Carlo.

Throughout this document, we use Anti-KT CaloJets with cone size $R = 0.5$ and the standard L2 (relative) and L3 (absolute) jet energy corrections derived from Monte Carlo for collisions with $\sqrt{s} = 900$ GeV.

The term $leading\ jet$ everywhere refers to the jet with the highest corrected p_T in the event.

- Basic event selection criteria
  - must pass beam-crossing technical trigger ("BPTX" / "bit 0")
  - must pass Beam Scintillation Counter technical triggers ("bit 40 OR 41")
  - must have a "physics declared" trigger bit set
  - should not contain beam halo (*i.e.,* veto on technical trigger bits 36 to 39)
  - must contain a primary vertex constructed from at least 4 tracks
  - must contain a primary vertex with $|PVz| < 15$ cm
- Inclusive jet candidate selection criteria:
  - each jet in the event must have corrected $p_T > 10$ GeV
  - each jet must have $|\eta| < 2.6$
- Dijet candidate selection criteria:
  - both leading jets must have corrected $p_T > 10$ GeV
  - both leading jets must have $|\eta| < 3$
  - the two jets should be back-to-back in azimuthal angle within a tolerance window of 1 radian,*i.e.,* $|\Delta\phi - \pi| < 1.0$

2.2 Definition of jet identification variables of interest

We study the following eight jet ID variables of interest for both inclusive jets and for dijet events:

1. *fHPD*: fraction of the jet's energy measured by the "hottest" (or most energetic) HPD
2. *fRBX*: fraction of the jet's energy measured by the most energetic readout box
3. *sigmaEta*:
$$\sigma_\eta = \sqrt{\left[\sum_{\rm tow} \frac{E_T^{\rm tow}}{E_T}\eta_{\rm tow}^2\right] - \left[\sum_{\rm tow} \frac{E_T^{\rm tow}}{E_T}\eta_{\rm tow}\right]^2}, \quad (1)$$
where E_T is the transverse energy of the jet, while $E_T^{\rm tow}$ and $\eta_{\rm tow}$ denote respectively the transverse energy deposited in the calorimeter towers belonging to the jet and the pseudorapidity of those calorimeter towers.
4. *sigmaPhi*:
$$\sigma_\phi = \sqrt{\left[\sum_{\rm tow} \frac{E_T^{\rm tow}}{E_T}\phi_{\rm tow}^2\right] - \left[\sum_{\rm tow} \frac{E_T^{\rm tow}}{E_T}\phi_{\rm tow}\right]^2}, \quad (2)$$
where $\phi_{\rm tow}$ denotes the azimuthal angle of calorimeter towers belonging to the jet.
5. *n90*: minimum number of calorimeter towers carrying 90% of the jet energy
6. *n90hits*: minimum number of calorimeter *rechits* carrying 90% of the jet energy
7. *EMF*: electromagnetic energy fraction of the jet
8. *nTrkCalo*: number of tracks associated to a jet at the calorimeter face

2.3 Inclusive jet sample: distribution of jet ID variables of interest

2.3.1 Distributions in data and Monte Carlo

Distributions for eight jet ID variable are shown below, in Fig. 1, as a reminder of the same distributions shown in CMS AN-2010/019 and CMS AN-2010/009. The cut values used are indicated on the plots, for the "loose" and "tight" jet identification cuts, discussed below.

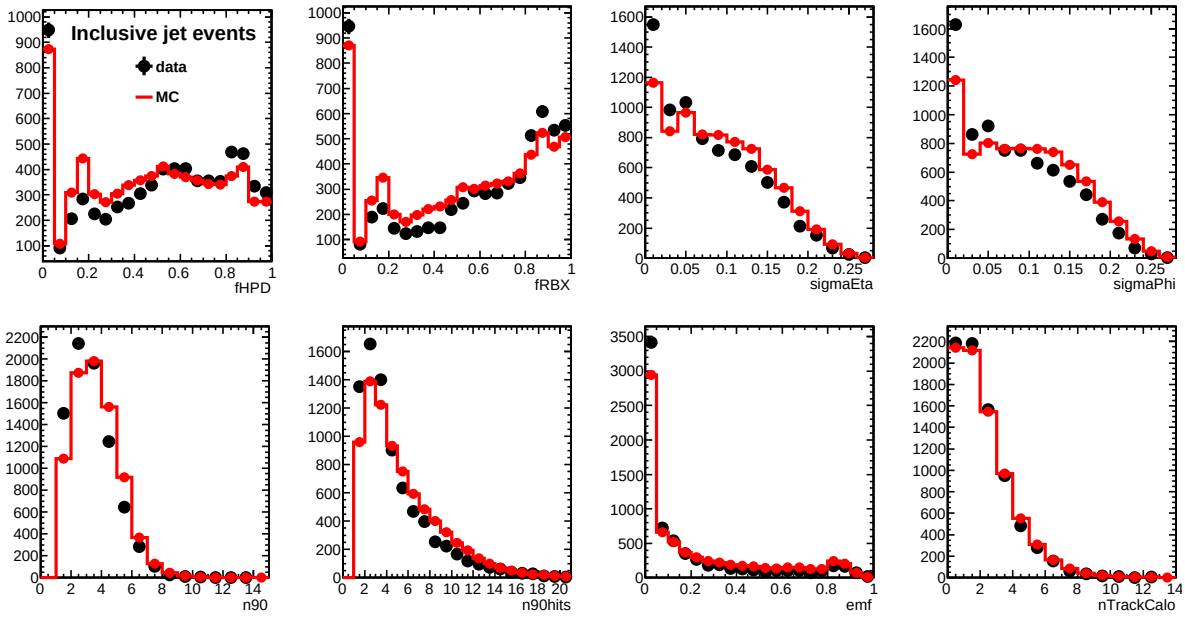


Figure 1: Jet variable distributions among inclusive jets in data and in MC.

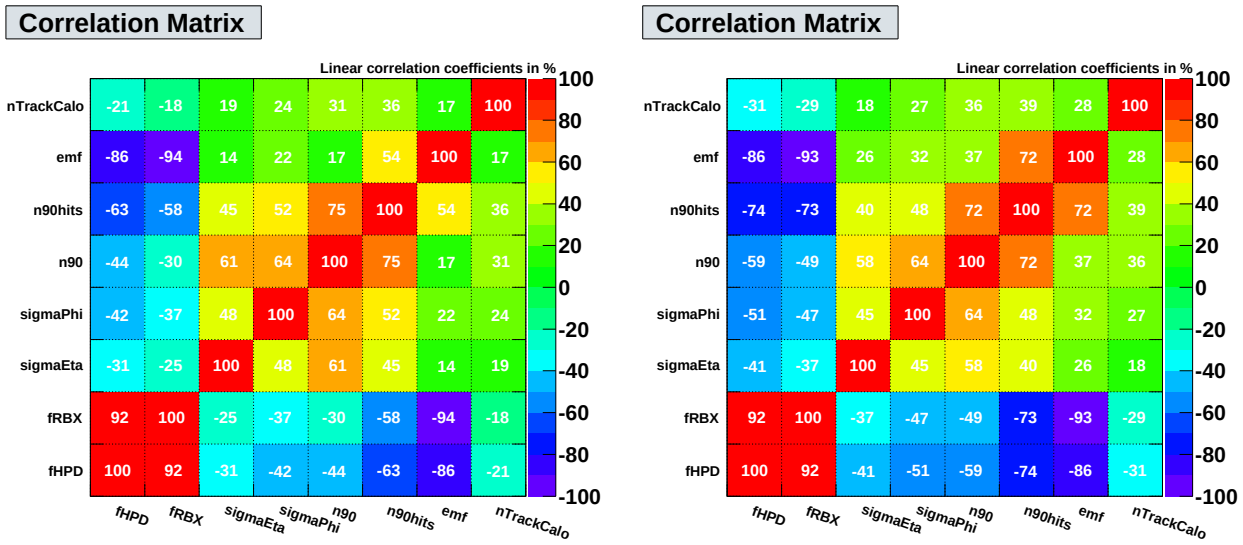
The correlation matrix for all eight of these variables is shown in Fig. 2. Inclusive jets in data give the result on the left, and in Monte Carlo, the result on the right. The MC CaloJets used here are not required to match GenJets, a procedure discussed below, in section 2.3.2. This places the simulation and the data on the same footing.

A few features are readily evident, such as the expected high correlation between fRBX and fHPD, or the expected anti-correlation between these variables and EMF. A jet with a large fraction of its energy in the HCAL is very likely to have a small fraction of its energy in ECAL.

Between n90 and n90hits lies the distinction of calorimeter towers and calorimeter cells. The measure n90 draws from all calorimeter cells, both in ECAL and in HCAL. The contrasting degrees of correlation between them and EMF recalls the presence of single cell noise spikes recorded by the ECAL and often reconstructed as jets in these runs. The opposite case obtains in the HCAL, where multi-cell noise within an HPD or an entire RBX constitutes an important component of noise. This idea goes some way toward explaining the anti-correlation of n90 and n90hits with fHPD and fRBX.

Finally, the general principle that broader jets will require more cells or towers to contain 90% of their deposited energy may be at work behind the slight correlation between the spatial-spread quantities σ_η and σ_ϕ with n90hits and, more significantly, n90.

Of the remaining correlations it would be difficult to draw firm conclusions. It is reassuring to see such close correspondence between Fig. 2a and Fig. 2b, giving confidence in the quality of the Monte Carlo.



(a) (b)

Figure 2: Correlation matrix of jet variables among inclusive jets: (a) data, (b) MC.

2.3.2 Distributions in Monte Carlo: using *MC truth* information

In Monte Carlo, CaloJets, jets reconstructed from the simulated calorimeter response, are matched *a posteriori* to GenJets, which are reconstructed from simulated hadronization of scattered partons. The matching parameter is purely angular; a match is declared when $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.5$. This is the analog of "MC truth" information when used for jets. This matching is necessarily not a simple one-to-one correspondence, in good approximation to the known limitations of jet calorimetry. MC CaloJets which fail to find a match are a small minority. GenJets below 5GeV are not stored in the simulated event, so they are unavailable to be matched, and provide one likely source of this nonideal behavior.

For the same eight jet quantities, distributions are shown in Fig. 3 for MC CaloJets which were matched to GenJets, and for those which failed to match.

Finally, in Fig. 4 are shown the correlation matrices using matched and non-matching CaloJets from this MC sample. Thus, the sample of simulated jets which give rise to Fig. 3a are thought to be dominated by "good" jets, and those behind Fig. 3b may be more influenced by noise. Although there are certainly differences, it is difficult to ascribe significance to them.

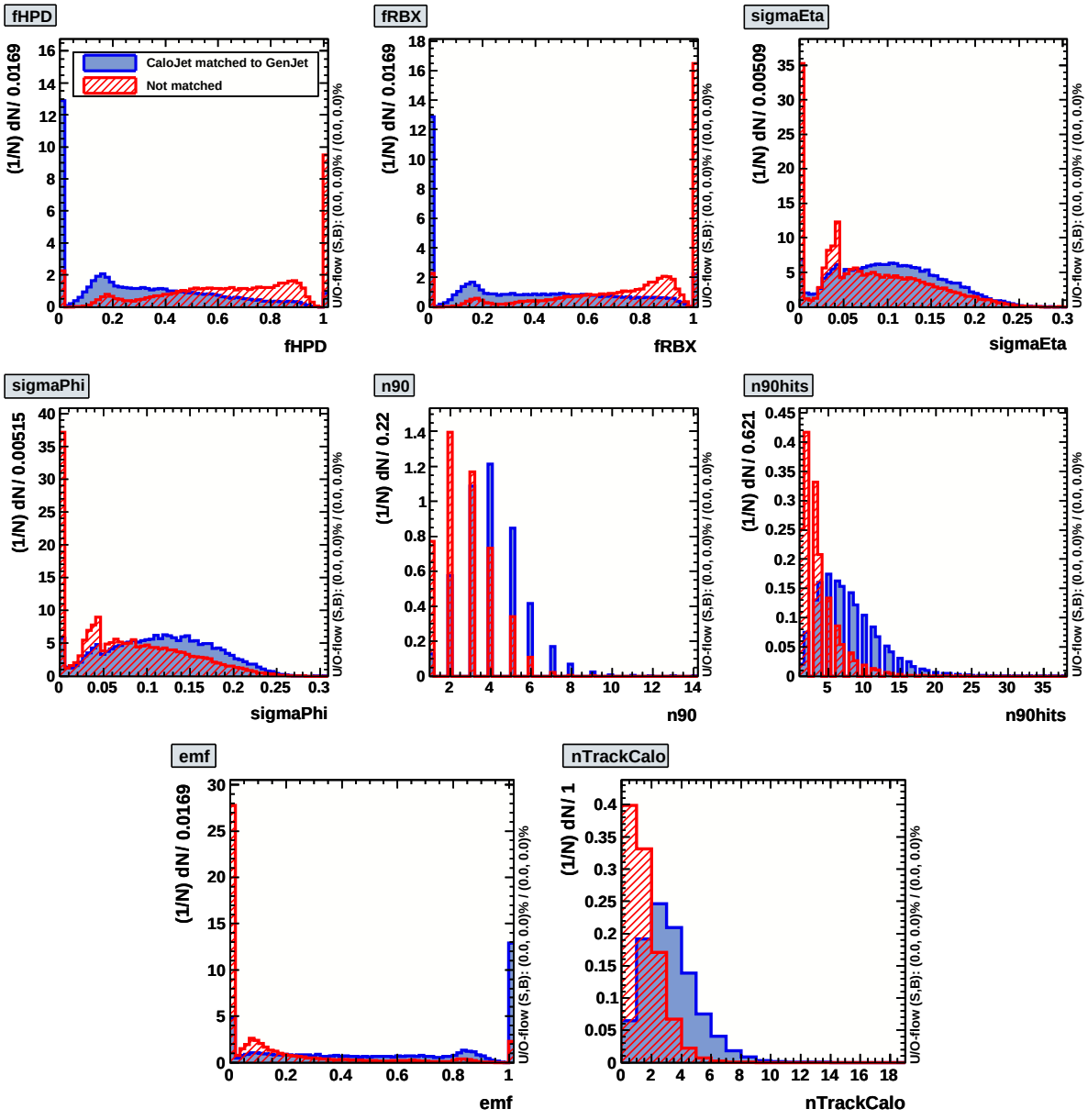
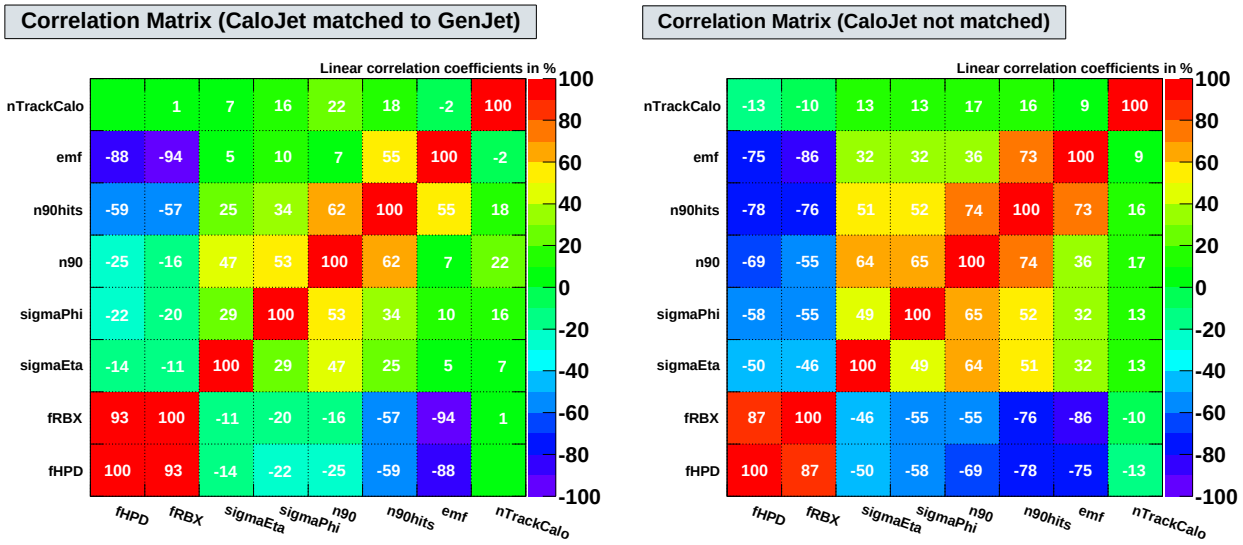


Figure 3: Distribution of some interesting variables in Monte Carlo, among inclusive jets. Distributions are normalized to have the same areas.



(a) (b)

Figure 4: Jet ID variables in inclusive MC CaloJets: (a) matched, (b) unmatched.

2.4 Dijet sample: distribution of jet ID variables of interest

2.4.1 Distributions in data and Monte Carlo

Within these sample sets of data and simulation, we define a dijet topology as one in which the two jets leading in transverse momentum meet a loose cut of being “back-to-back” in azimuth to within 1 radian. This is done to extract the set of events likely to contain the parton- parton scattering events.

For dijets meeting the dijet selection criteria (see §1.2), the agreement for jet quantity distributions and correlations are shown in Fig. 5 and Fig. 6, respectively. Here, the issue of matching CaloJets to GenJets is ignored, and all MC calorimeter jets are allowed as input to the dijet selection.

Within this dijet sample, all agreement between data and Monte Carlo is quite good, strongly suggesting that differences lie more in the representation of noise than in physics.

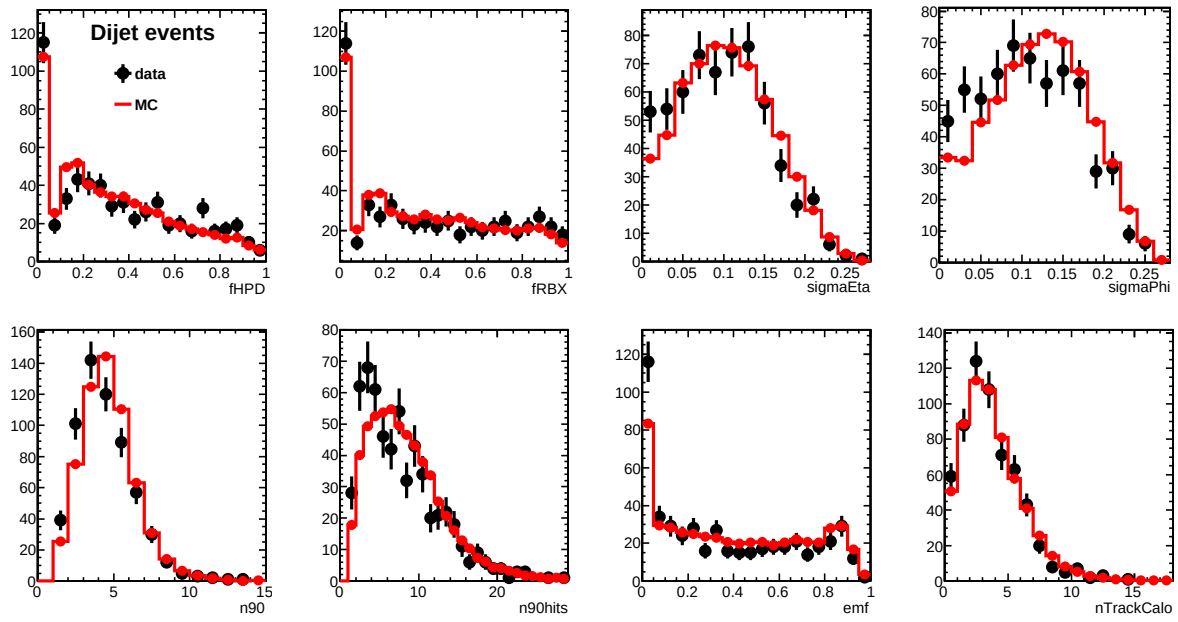
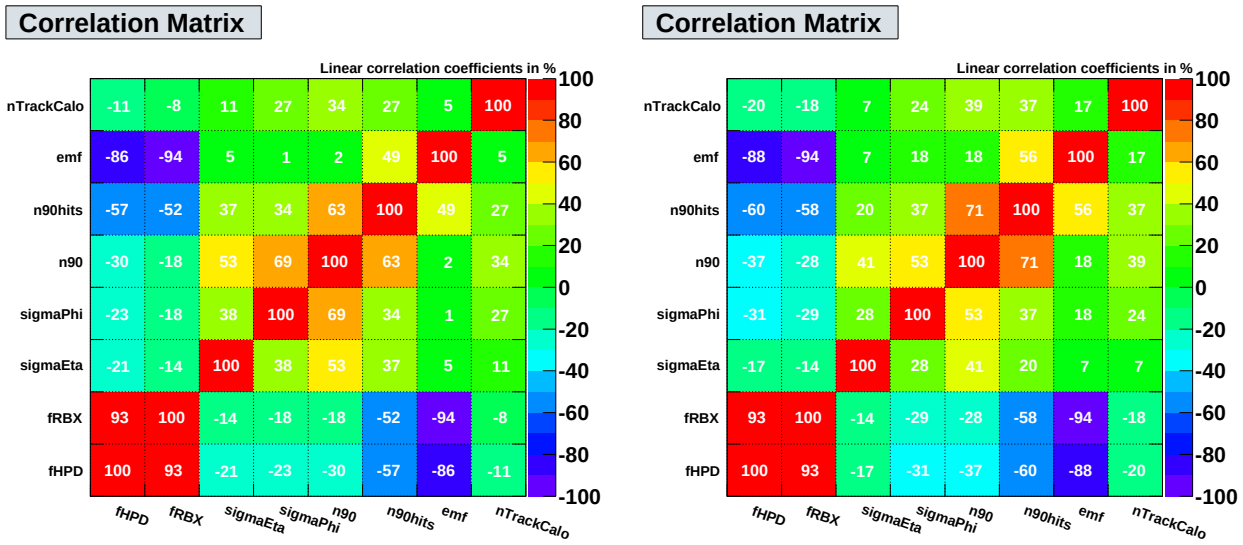


Figure 5: Jet ID variables in dijets: data and MC.



(a) (b)

Figure 6: Jet ID variable correlations in dijets: (a) data, (b) MC.

2.4.2 Distribution in Monte Carlo: using *MC truth* information

Finally, we explore the issue of matching in the dijets found in the simulation. The simulations and correlations of the jet ID variables are for the most part quite similar to those made over the inclusive jets of the sample. Compare Figs. 7 and 8 to Figs. 3 and 4. The fair agreement provides no evidence that there are noise jets in inclusive MC from mismatching.

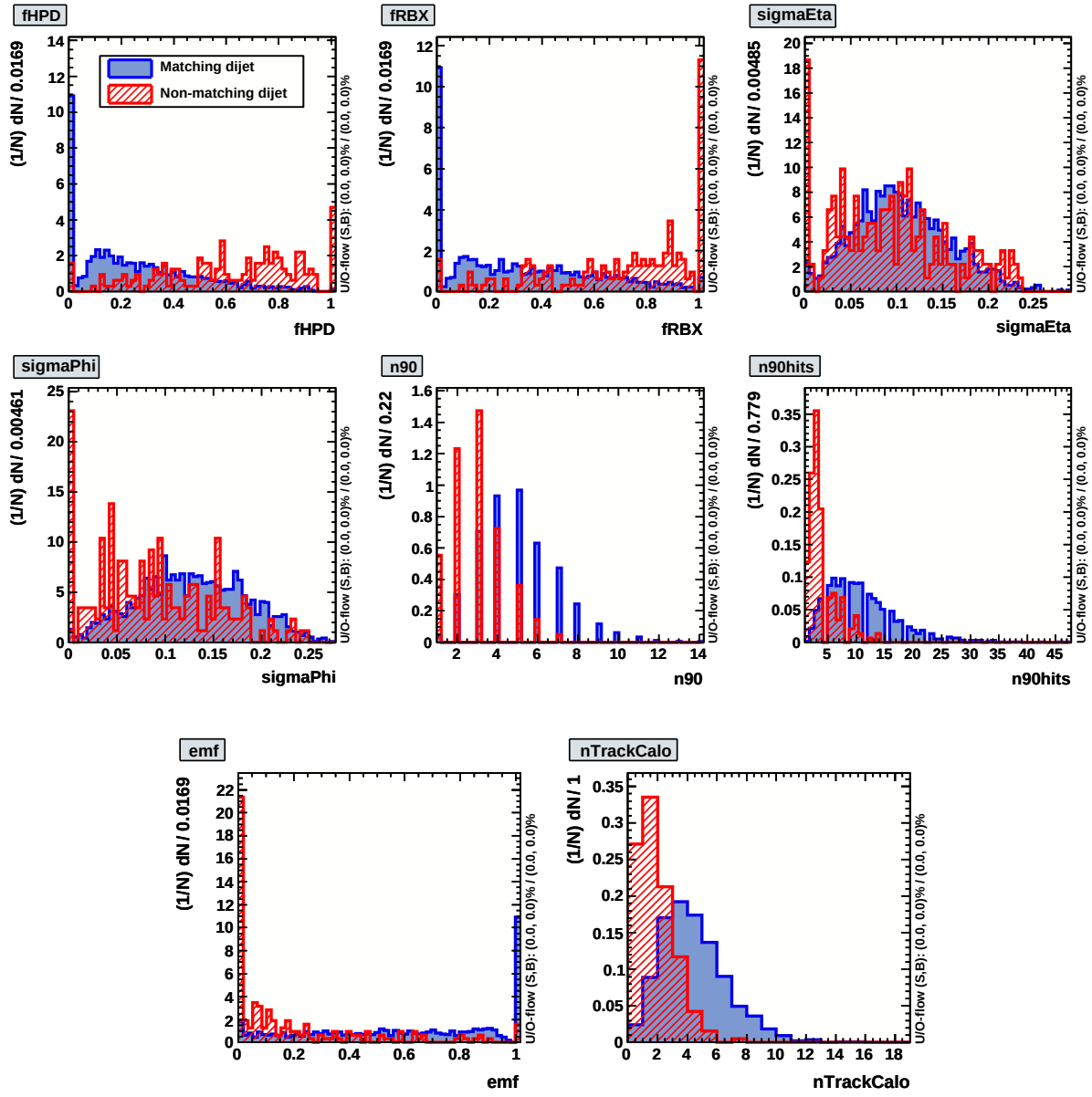
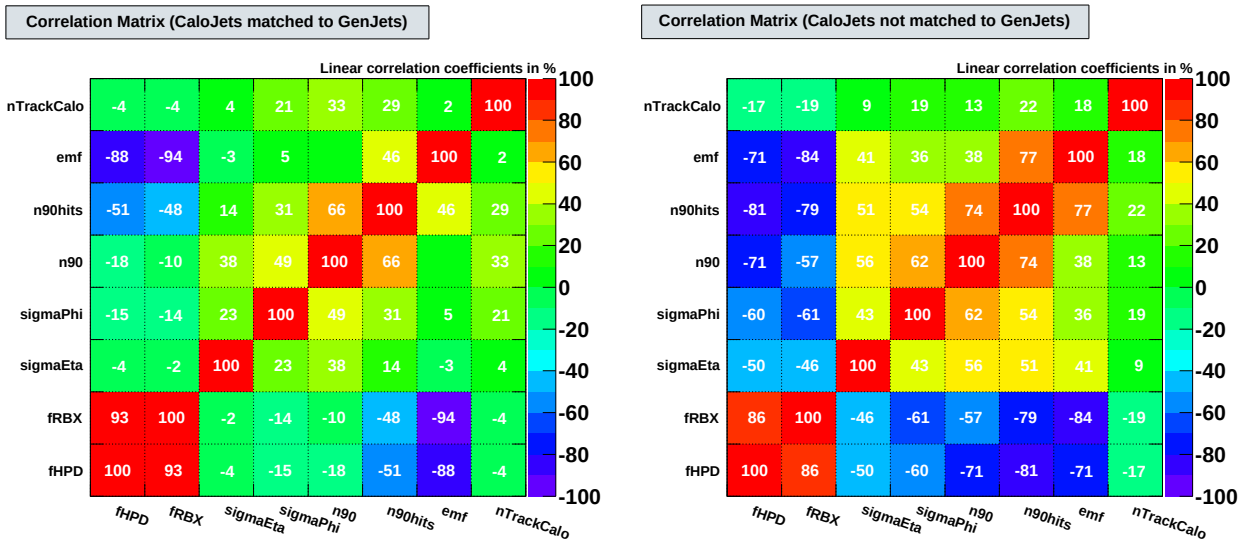


Figure 7: Distribution of some interesting variables in Monte Carlo, among dijets.



(a) (b)

Figure 8: MC DiJets: (a) matched, (b) unmatched.

3 Survival fraction for loose and tight jet ID selection criteria in inclusive jet events

In the inclusive jet analysis, only jets meeting the jet candidate criteria listed in § 2.1 are considered. The cuts used to select jets from these inclusive jets are n90hits > 4, EMF > 0.01 or $|\eta| > 2.6$, fHPD < 0.98, σ_η, $\sigma_\eta > 0.01$. Collectively, these are termed the "tight" jet ID selection criteria. The fractional survival fractions as a function of corrected jet p_T for these four criteria each applied alone are shown in Fig. 9 through 11.

In § 4 we examine the looser cuts applied to jets in our dijet sample, the more relaxed selection criterion of n90hits > 1 is considered. For ease of comparison, that cut's survival fraction is presented here in Fig. 9 alongside n90hits > 4.

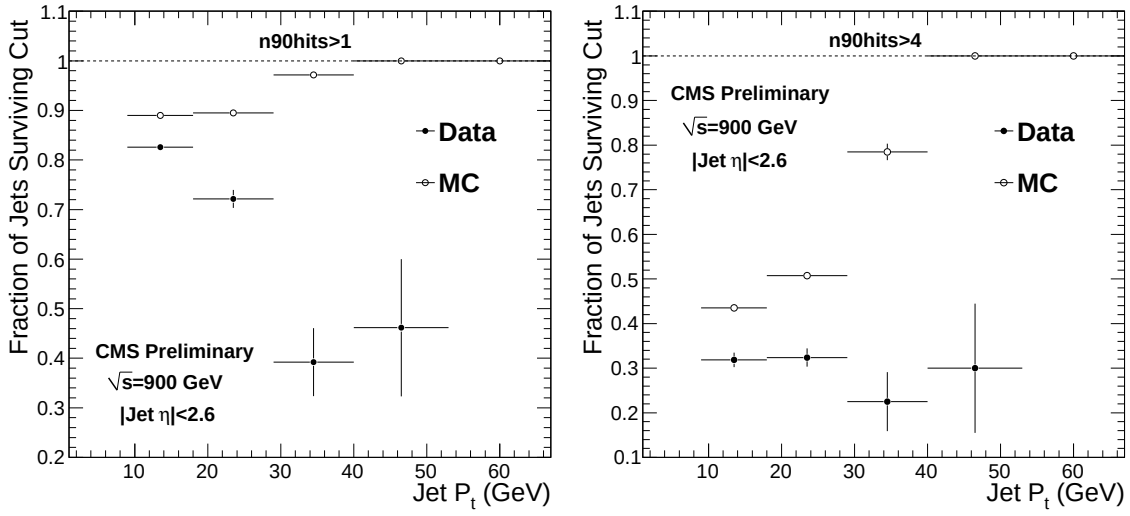


Figure 9: *Left:* Inclusive jet n90Hits> 1 cut survival fraction as a function of corrected jet p_T. *Right:* Inclusive jet n90Hits> 4 cut survival fraction as a function of corrected jet p_T.

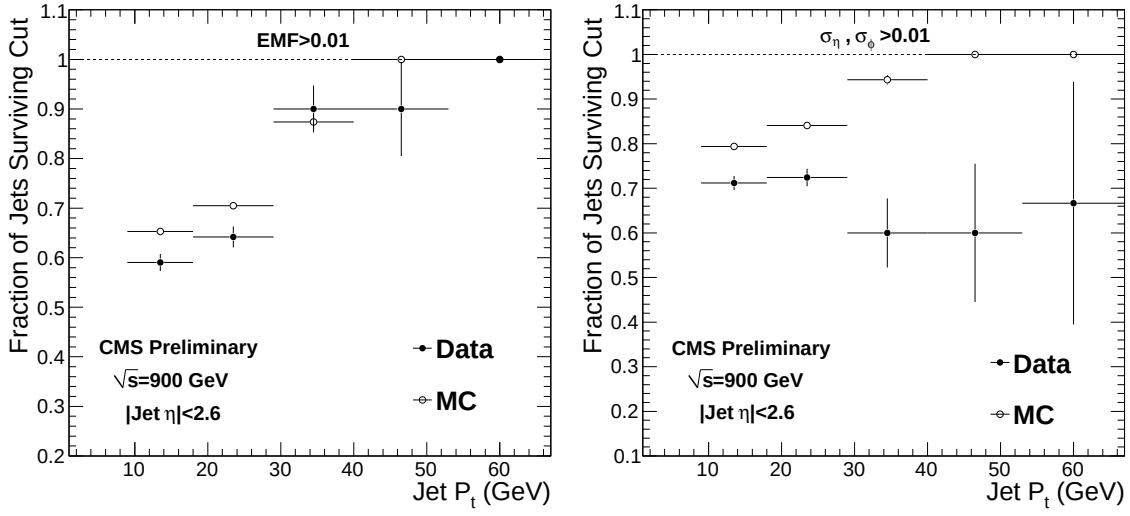


Figure 10: *Left:* Inclusive jet Minimal EMF cut survival fraction as a function of corrected jet p_T. *Right:* Inclusive jet σ_η, $\sigma_\phi > 0.1$ cut survival fraction as a function of corrected jet p_T.

The survival fraction for the "tight" cuts is shown in Fig. 12, on the left. To its right is the resulting inclusive jet p_T spectrum.

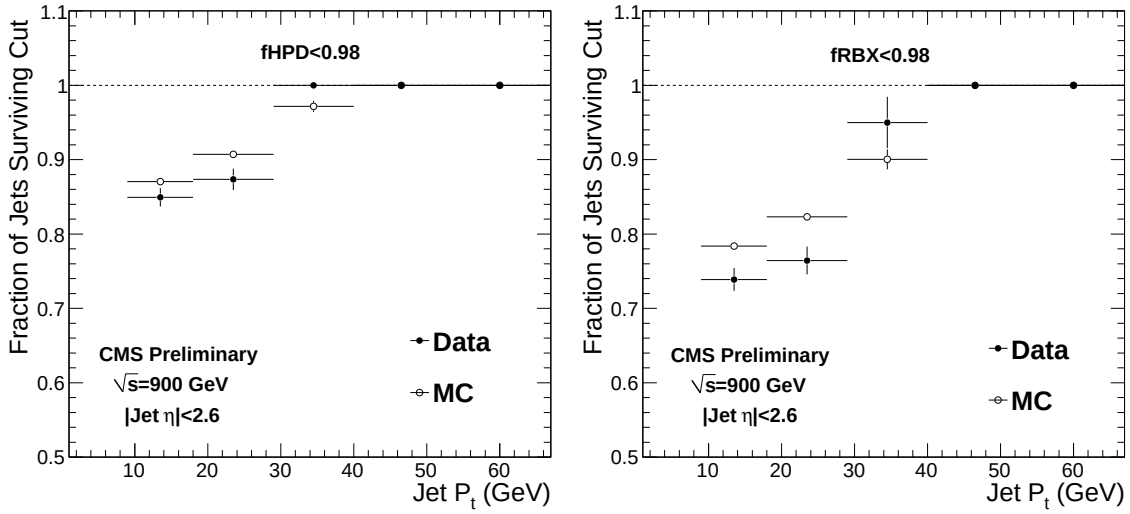


Figure 11: *Left:* fHPD cut survival fraction as a function of corrected jet p_T. *Right:* fRBX cut survival fraction as a function of corrected jet p_T.

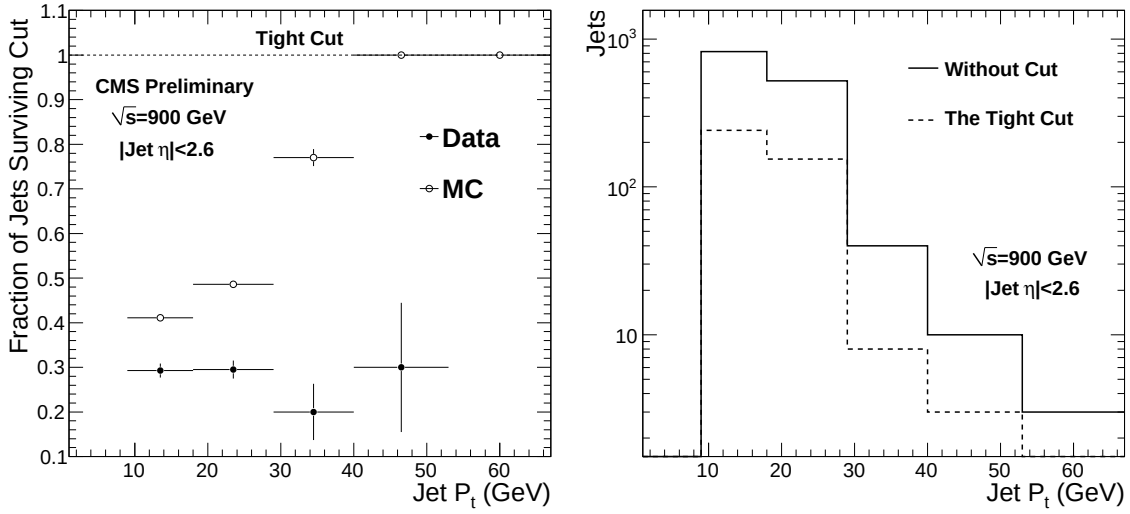


Figure 12: *Left:* The tight jet identification cut survival fraction as a function of corrected jet p_T among inclusive jets. *Right:* The inclusive jet p_T spectrum, with and without this set of cuts, for data.

4 Survival fraction for loose jet ID selection criteria in dijet events

Events having dijet topology are greatly enriched in real jets. For this reason, the same cuts on jet variables pass a much higher fraction of reconstructed hadronic jets in these events than they do in the super-sample of inclusive jets. Here we present the fractional survival fractions for the two leading jets in data and in Monte Carlo.

Within this sample set, we define a dijet topology as one in which the two jets leading in transverse momentum meet a loose cut of being back-to-back in azimuth to within 1 radian. This is done to extract the set of events most likely to contain the parton- parton scattering events. Some such di-jet events include more than two jets, but they are ignored in this section to focus on the di-jet itself.

For the leading two jets comprising a dijet the following four figures show the fractional survival fractions as a function of corrected jet p_T to pass cuts of Fig. 13 n90hits > 1, Fig. 14 EMF > 0.01, and Fig. 15 fHPD < 0.98, separately.

The same survival fraction measure for all cuts simultaneously is given in Fig. 16.

In all four cases:

Left: For the numerator we count the number of jets for which both jets in the dijet event pass the cut, and for the the denominator we count the number of jets in all dijets before cuts.

Right: For the numerator we count the number of jets which pass the cut, and for the denominator we count the number of jets in all dijets before cuts.

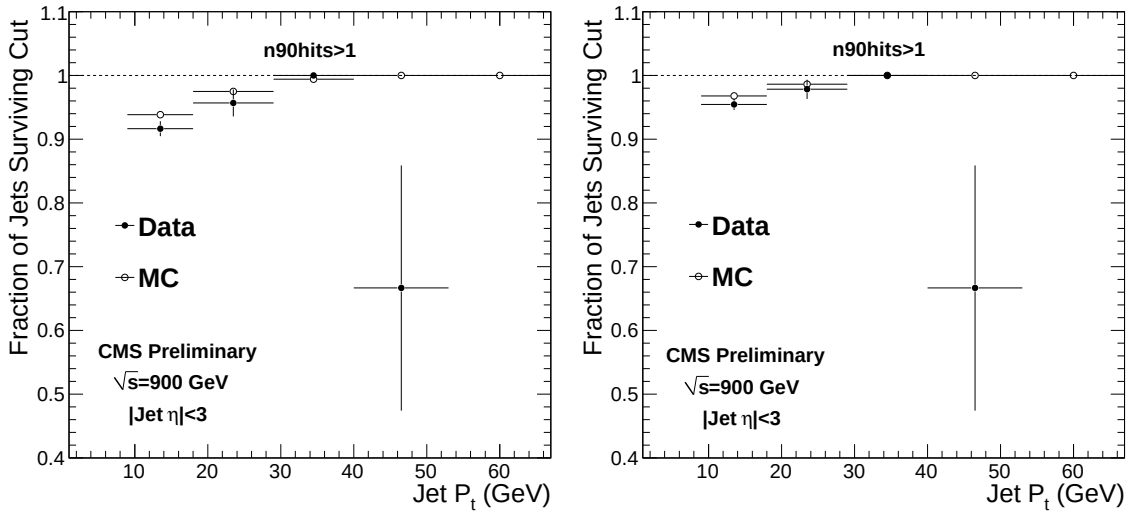


Figure 13: Dijet N90Hits> 1 cut survival fraction as a function of corrected jet p_T. *Left:* The cut is required for both leading jets. *Right:* The cut is required jet-by-jet for the two leading jets..

Comparing the survival fractions for jets comprising dijets to those of inclusive jets we see that in the dijet case, survival fractions are much improved across the whole pT range for which statistics are available.

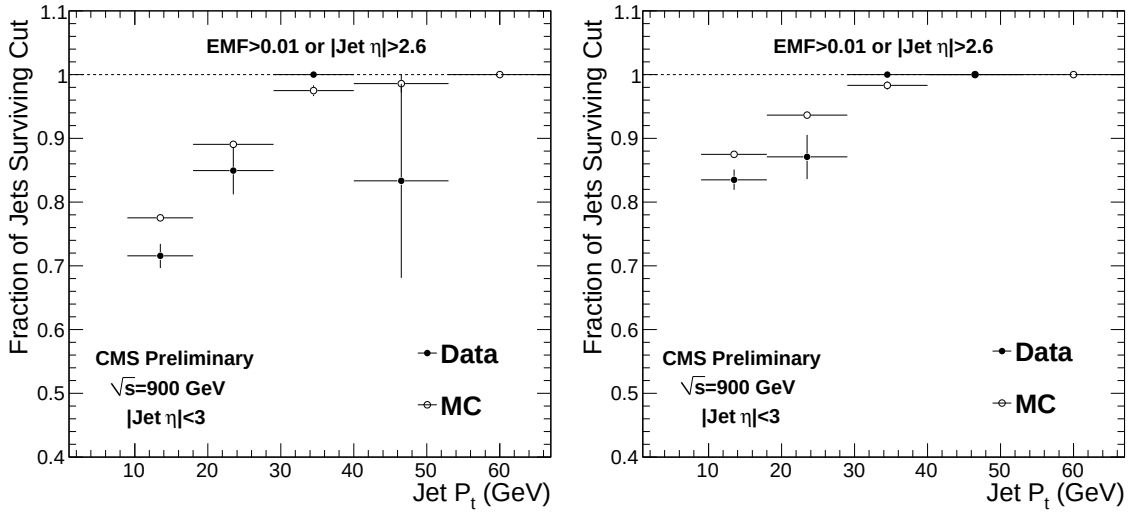


Figure 14: Dijet Minimal EMF cut survival fraction as a function of corrected jet p_T. *Left:* The cut is required for both leading jets. *Right:* The cut is required jet-by-jet for the two leading jets.

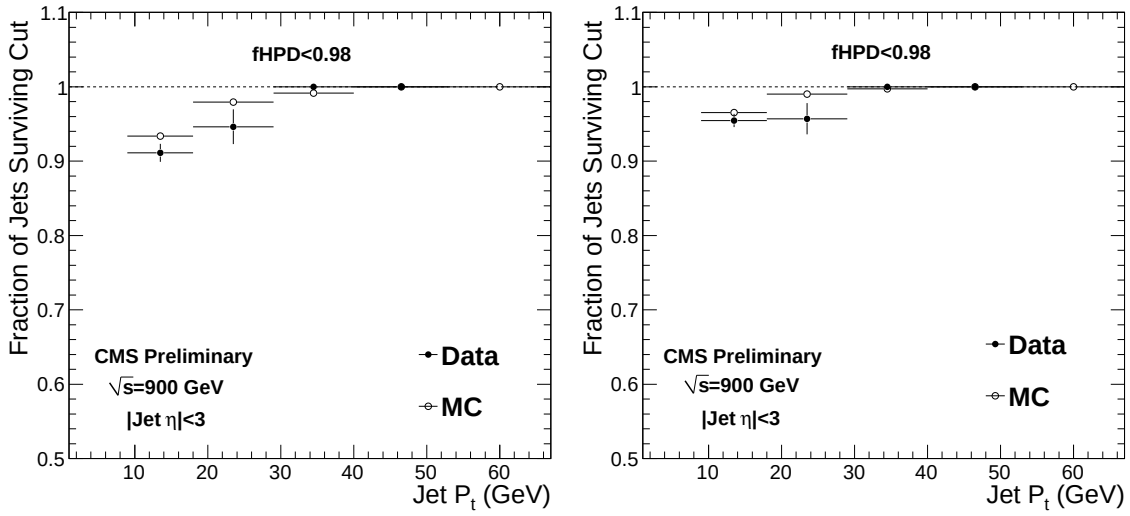


Figure 15: Dijet fHPD cut survival fraction as a function of corrected jet p_T. *Left:* The cut is required for both leading jets. *Right:* The cut is required jet-by-jet for the two leading jets.

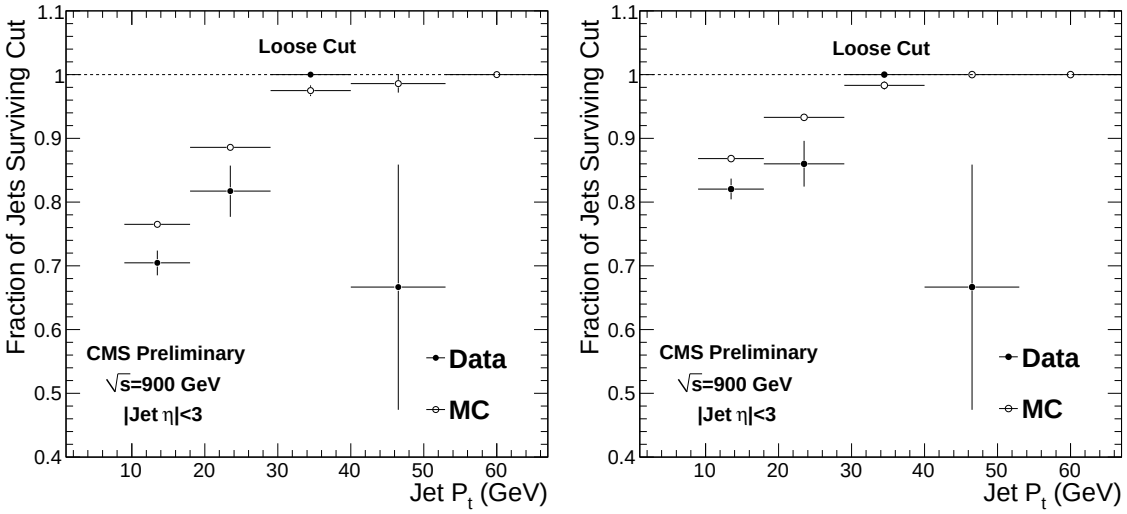


Figure 16: The loose cut survival fraction as a function of corrected jet p_T. *Left:* The cut is required for both leading jets. *Right:* The cut is required jet-by-jet for the two leading jets.

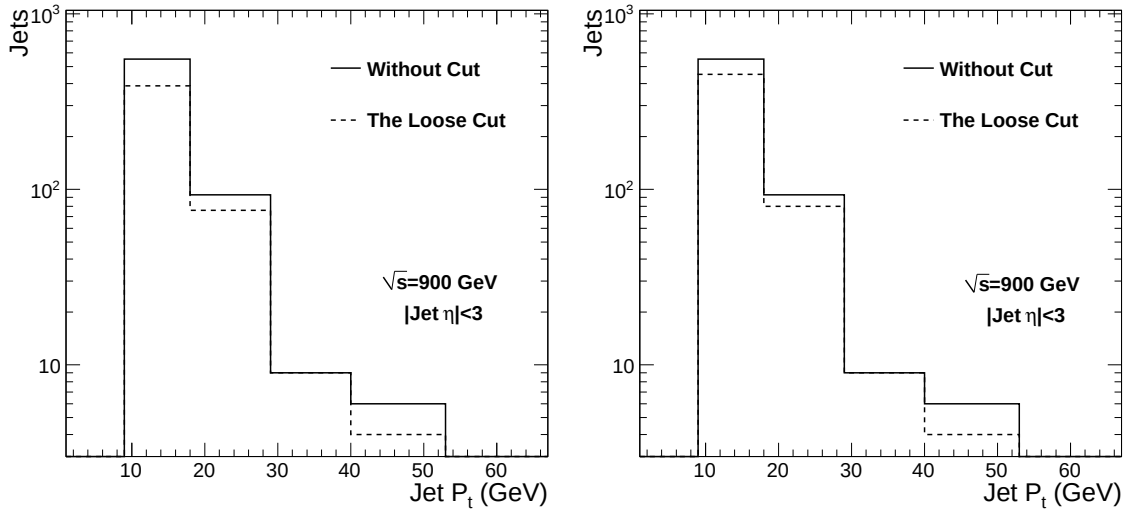


Figure 17: The p_T spectrum among jets comprising dijets where the loose cut is required *Left:* on both jets together and *Right:* on each jet independently.

5 Conclusions

We have presented the cuts made on jet variables as studied in CMS AN-2010/019 "Inclusive Jet Studies with $\sqrt{s}$ = 900GeV Collisions" and CMS AN-2010/009 "Jet Commissioning using Di-Jet Events from 900 GeV and 2360 GeV Collision Data". In support of the cuts used in those notes, we presented the fraction of jets and of dijet events to pass these cuts as a function of corrected jet p_T in the CMS collision data of 2009 at $\sqrt{s} = 900$ GeV and in Monte Carlo simulation. We also noted a few features of the correlations among these cut variables.

The dijet sample we examined is a pure one of physical jets and only uses the loose jet ID cuts for safety and to kill ECAL noise. Jet ID cuts rejected similar numbers of jets in the data and the MC, as a function of p_T.

Inclusive jet analysis has more background and uses appropriately tighter cuts than does dijet analysis. The jet ID cuts reject many more jets in data than they do in simulation, and they have different behavior as a function of p_T. The survival fractions in MC approach one with increasing p_t for both inclusive and dijet analyses. These results are encouraging–all the more so with the LHC 2010 run imminent.