

## Monte Carlo Fragmentation Tuning for Jet Energy and Resolution Analysis

Timothy Hessing, Steve Behrends, Paul Tipton, Brian Winer

February 15, 1990

### Abstract

In order to extract both energy corrections and resolution for jets in the CDF Detector, we need to use Monte Carlo to simulate the response of jets in the detector. A major ingredient involved in the jet energy corrections and jet resolution is the non-linearity in the response of the detector to charged particles. We have therefore tuned the fragmentation in the SETPRT Monte Carlo routine to model the charged particle fragmentation observed in the 1988-1989 CDF data set. In addition, the CTC tracking efficiency in jets has been measured and parametrized, and introduced into the QFL tracking simulation.

### Introduction

The jet energy corrections currently being used are based on 1987 data. Since in this data set the highest transverse energy ( $E_t$ ) for the jets is only 250 GeV, the corrections were only calculated to 250 GeV. The 1988-1989 data set extends to 410 GeV; therefore it is necessary to recalculate these corrections to extend the range of validity. At the same time, we would like to extract the jet energy resolution from the simulation.

Much work has been done to tune the simulation to reproduce the results of the test beam.[1] The simulation has been tuned to reproduce the single pion response observed in the test beam as well as the non-linearity observed at lower pion energies. We therefore believe that if we can tune the simulation to reproduce the fragmentation observed in the 1988-1989 data, we can then trust the simulation to give the correct result for the jet energy corrections and jet resolution. This note is the result of that study.

## Data and Simulation

To study the fragmentation we used the same data set currently being used for the inclusive jet cross section and dijet invariant mass measurements. This consists of approximately  $4.31 \text{ pb}^{-1}$  of jet data in the region above  $120 \text{ GeV}$  in  $E_t$ , and approximately  $870.55 \text{ nb}^{-1}$  of jet data in the region from  $30\text{--}120 \text{ GeV}$ . The high  $E_t$  data set came from spin production (QCS02), while the lower  $E_t$  data came from the QCD02 production.

The simulation was performed using the following path:

```
SIMJET SETPRT QFLANA JETCLU QTKEFF CENJET JETFRG.
```

1. SIMJET generated jets. It created two partons in a falling spectrum defined to match the spectrum measured in the inclusive jet cross section. The generated jets were also required to be within  $\pm 0.9$  in  $\eta$ .
2. SETPRT fragmented the jets. This routine has five adjustable parameters which we used to tune the fragmentation. During this study, we found it necessary to modify this routine to incorporate an  $E_t$  dependence into one of these five variables.
3. QFLANA simulated the detector response for the generated jets.
4. JETCLU clustered the simulated jets.
5. QTKEFF incorporated CTC tracking efficiency into the simulation by flagging tracks as "lost", on the basis of a parametrization of tracking efficiency in jet events.
6. CENJET filtered cosmic rays in the data and was therefore used on the simulation also.
7. JETFRG extracted and plotted the various fragmentation quantities we wanted to study.

Since the simulation only supplied two partons, we required our events from the actual data to be dijet events. This was done by requiring the following:

1. There must be at least two jets.
2. There must be no third jet with  $E_t > 20 \text{ GeV}$ .
3. The leading  $E_t$  jet and the second jet must be back-to-back in  $\phi$ , within  $\pm 30$  deg.

4. The Z vertex was required to be within 60 cm of the center of the detector.
5. At least one jet was required to be in the central detector,  $0.1 \leq |\eta| \leq 0.7$ .

The jets were clustered with a cone size of 0.7 and tracks were required to be within the cone of the jet, and to have  $P_t \geq 500 \text{ MeV}/c$ . It should be noted that these same cuts were applied to both the simulation and the data.

## Tracking Efficiency

The fragmentation distributions observed in data includes effects of CTC tracking inefficiency in jets. It was therefore necessary to measure this efficiency and incorporate it into our code (QTKEFF). The procedure for measuring the tracking efficiency is described below.

First, we started with real events having a jet in the central detector. The data came from both spin and production streams. Events were selected using the following criteria:

1. At least one jet in the central detector.
2. A second jet back to back in  $\phi$ , within 30 deg.
3. The average jet energy,  $(E_{t1} + E_{t2})/2$ , greater than 15.0 GeV.
4. No third jet with  $E_t > 15 \text{ GeV}$  or  $E_t > 0.25 * (E_{t1} + E_{t2}) \text{ GeV}$ .
5.  $\eta_{boost}$  and  $\eta_{star} < 1.0$
6. A Z-vertex within 50.0 cm of the center of the detector.

Next, these events had a Monte Carlo track embedded into the central jet. We embedded a different Monte Carlo track into the same real event 30 times (saves some CPU). The Monte Carlo tracks had the following characteristics:

1. The Z distribution for the longitudinal momentum (relative to the jet axis) was flat between 0.0 and 0.4.
2. The transverse momentum (again relative to the jet axis) had an exponential spectrum with a average  $P_t$  of 0.7 GeV/c.
3. The pseudorapidity of the embedded track relative to the jet axis was required to be less than 0.8.

4. The azimuthal angle in the jet was randomized between 0 and 360 deg.
5. The Monte Carlo track exited the CTC at the full radius.
6. The charge of the track was randomized with equal contributions of  $\pm$  charges.

To study the efficiency, we retracked (5.0 tracking) the events and determined how often we found the embedded track. A good track in the events was defined by the standard CTRSEL requirements and the requirement that the  $P_t$  was greater than 500 MeV/c. In each event, the Monte Carlo track was considered found if more than 25% of the hits on a good track matched the hits of the Monte Carlo track (generator level). That track was considered the embedded Monte Carlo track. If more than one track in the event satisfied this requirement, the track with the largest fraction of hits matching the Monte Carlo track was identified as the embedded Monte Carlo track.

We parameterized the tracking efficiency as a function of inter-track distance and jet  $E_t$ . The inter-track distance was broken down into XY ( $r, \phi$ ) and Z components. Table 1 shows the efficiency as a function of inter-track distance for jets with an average  $E_t$  of 80 GeV. The table is broken down into 11 bins in the XY direction (0 - 3 cm, 3 - 6 cm, 6 - 9 cm, ..., 27 - 30 cm, > 30 cm) and 7 bins in the Z direction (0 - 5 cm, 5 - 10 cm, ..., 25 - 30 cm, > 30 cm). These distances are the *sum* of the inter-track distances at three different wire radii (wires 0, 36, 83). The values in the table have been "smoothed" by eye to give a monotonically increasing efficiency with XY and Z distances. We believe the uncertainty on each component is on the order of 5 - 10%.

Table 1											
Track Efficiency in jets with $\langle E_t \rangle = 80 \text{ GeV}$											
$\Delta XY$ (3 cm steps) vs $\Delta Z$ (5 cm steps)											
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	24-27	27-30	>30
0-5	0.62	0.75	0.78	0.80	0.84	0.85	0.88	0.90	0.91	0.91	0.91
5-10	0.65	0.77	0.79	0.82	0.85	0.87	0.89	0.90	0.91	0.91	0.92
10-15	0.67	0.78	0.81	0.84	0.87	0.90	0.90	0.91	0.91	0.92	0.92
15-20	0.69	0.79	0.83	0.85	0.88	0.90	0.90	0.91	0.91	0.92	0.92
20-25	0.71	0.80	0.84	0.86	0.89	0.89	0.91	0.91	0.91	0.92	0.93
25-30	0.73	0.82	0.85	0.89	0.90	0.91	0.91	0.92	0.92	0.93	0.93
>30	0.76	0.85	0.87	0.90	0.91	0.91	0.92	0.92	0.93	0.93	0.93

To determine the dependence on jet  $E_t$ , we repeated the measurement using a sample of jets with  $E_t > 150 \text{ GeV}$ . The results are shown in Table 2. Finally, we parameterize the tracking efficiency as follows:

For  $E_t(\text{jet}) < 80 \text{ GeV}$  :

$$1 - \text{Eff}(I, J) = 1 - \text{Eff}(I, J)[80] - 0.001 \times (80 - E_t(\text{jet}))$$

For  $E_t(\text{jet}) > 80 \text{ GeV}$  :

$$1 - \text{Eff}(I, J) = 1 - \text{Eff}(I, J)[80] - S(I, J) \times (80 - E_t(\text{jet}))$$

where  $\text{Eff}(I, J)[80]$  is an element of Table 1. The slopes  $S(I, J)$  are linear extrapolations between Table 1 and Table 2, element for element.

$$S(I, J) = (\text{Eff}(I, J)[80] - \text{Eff}(I, J)[170])/90.$$

The  $S(I, J) * 100$  are given in Table 3. We have altered a few slope elements to smooth the results but these changes were very small.

Table 2											
Track Efficiency in jets with $E_t > 150 \text{ GeV}$ ( $\langle E_t \rangle = 170 \text{ GeV}$ )											
$\Delta XY$ (3 cm steps) vs $\Delta Z$ (5 cm steps)											
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	24-27	27-30	>30
0-5	0.50	0.62	0.68	0.70	0.72	0.75	0.78	0.79	0.81	0.82	0.84
5-10	0.58	0.64	0.68	0.71	0.74	0.76	0.79	0.81	0.83	0.84	0.86
10-15	0.62	0.66	0.69	0.73	0.76	0.79	0.82	0.84	0.86	0.87	0.89
15-20	0.64	0.68	0.71	0.74	0.78	0.81	0.84	0.87	0.89	0.90	0.90
20-25	0.68	0.72	0.76	0.80	0.82	0.84	0.87	0.88	0.90	0.91	0.91
25-30	0.74	0.78	0.80	0.82	0.84	0.86	0.88	0.89	0.90	0.91	0.92
>30	0.78	0.80	0.81	0.82	0.85	0.87	0.88	0.90	0.91	0.92	0.93

Table 3											
$S(I, J) = \Delta Eff(I, J) / \Delta E_t$											
Values Multiplied by 100											
$\Delta XY$ (3 cm steps) vs $\Delta Z$ (5 cm steps)											
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	24-27	27-30	>30
0-5	0.14	0.14	0.11	0.11	0.12	0.11	0.11	0.11	0.11	0.10	0.08
5-10	0.10	0.13	0.12	0.12	0.12	0.12	0.11	0.10	0.09	0.08	0.07
10-15	0.06	0.13	0.13	0.12	0.12	0.12	0.09	0.08	0.06	0.06	0.03
15-20	0.06	0.12	0.13	0.12	0.11	0.10	0.07	0.04	0.02	0.02	0.02
20-25	0.03	0.09	0.09	0.07	0.08	0.06	0.04	0.03	0.02	0.02	0.02
25-30	0.02	0.04	0.07	0.07	0.07	0.06	0.03	0.03	0.02	0.02	0.01
>30	0.02	0.06	0.07	0.07	0.07	0.04	0.04	0.02	0.02	0.01	0.00

These efficiencies are given as a function of two-track separation. We believe the efficiency should be applied to both tracks. This is a best (and naive) guess and is probably not strictly true. Tables 1 and 2 roughly agree at the lower right (large inter-track distance) and disagree at small inter-track distances. The upper left element (smallest inter-track distance) is changing the fastest (see Table 3). At roughly 523 GeV observed  $E_t$ , our parameterization would predict 0 efficiency. We would expect the efficiency to flatten out at some lower  $E_t$ . We arbitrarily chose 400 GeV as the energy above which the efficiency does not change and this is incorporated in QTKEFF. This certainly enters as a systematic uncertainty. To study jets much greater than 400 GeV, we would need more statistics.

## Fragmentation Variables and Tuning

In order to compare the data and simulation, the following quantities were used to tune the fragmentation:

1. The fragmentation  $Z$ , defined in the following ways:

$$\begin{aligned} Z(P) &= P_{//}(Track)/P(Jet) \\ Z(E) &= P_{//}(Track)/E(Jet) \\ Z(T) &= P_{//}(Track)/\Sigma(P_{in\ cone}(Track)) \end{aligned}$$

$P_{//}$  is the momentum of the track parallel to the jet axis. Each  $Z$  was plotted as  $(1/N_{jets})(dN/dZ)$ .

2. The number of tracks in the cone of the jet (Jet Multiplicity), was plotted as  $(1/N_{jets})(Mult)$ .
3. The number of tracks in the cone 90 deg away in  $\phi$  from the jet, (Underlying Event Multiplicity). was plotted as  $(1/N_{jets})(Mult_{Und})$ .
4. The momentum perpendicular to the jet axis, of tracks in the cone of the jet, was plotted as  $(1/N_{jets})(dN/dP_{\perp})$ , and  $P_{\perp}$  is the momentum of the track perpendicular to the jet axis.
5. The momentum perpendicular to the axis of a cone 90 deg away in  $\phi$  from the jet, of tracks in a cone 90 deg away in  $\phi$  from the jet, was plotted as  $(1/N_{jets})(dN/dP_{\perp\ Und})$ .
6. The  $P_t$  (relative to the beam axis) of the tracks in the cone around the jet, was plotted as  $(1/N_{jets})(dN/dP_t)$ .
7. The  $P_t$  of the tracks in the cone 90 deg away from the jet, was plotted as  $(1/N_{jets})(dN/dP_t\ Und)$ .
8. The  $P_t$  flow of the tracks with respect to the lead jet in  $\phi$ , which is the  $\Delta\phi$  of the track with respect to the lead jet weighted by the  $P_t$  of the track.
9. The  $P_t$  loss flow of the tracks with respect to the lead jet in  $\phi$ , which is the  $\Delta\phi$  of the track with respect to the lead jet weighted by the amount  $P_t(\text{loss})$  by the track due to the non-linearity.
10. The amount of energy loss in the jet due to the non-linearity in the central hadron calorimeter.

All of these variables were plotted in slices of jet  $E_t$ , so as to tune the fragmentation evolution.

SETPRT contains 5 variables which can be used to tune the fragmentation of jets. These variables are labeled in SETPRT as follows: XGEN(1), XGEN(2), SIGQT, SIGQT0, and CON2. XGEN(1) and XGEN(2) are used to describe the Field Feynmann fragmentation parameterization in SETPRT. SIGQT0 is used to define the transverse  $P_t$  for the underlying event. SIGQT is used to define the transverse  $P_t$  for the jet. CON2 represents the transverse fragmentation power. The default values of these variables are listed in Table 4, along with the "Final Tune" values. We tuned these five variables until we saw agreement in the plots, listed above. This means that the means, sigmas and shapes of each plot in the simulation was in agreement with what we observed in the data. We tuned for jets with  $E_t$  ranging from 30–60 GeV and also for jets with  $E_t$  ranging from 120–150 GeV.

Table 4  
SETPRT Tunes

Variable	Default Values	Low Values	High Values	Final Tune
XGEN(1)	0.950	0.950	0.950	0.950
XGEN(2)	7.000	7.500	7.500	7.500
SIGQT	0.275	0.275	0.600	0.250
SIGQT0	0.370	0.370	0.370	0.370
CON2	1.100	1.200	1.200	1.200

After examining the simulation at higher  $E_t$ 's in comparison with the data, we found it necessary to make SIGQT a function of jet  $E_t$ . Knowing both the high (120–150 GeV) and low (30–60 GeV)  $E_t$  tunes, we then modified both SETPRT and JETGEN to scale SIGQT with jet  $E_t$ . These low and high values are also listed in Table 4.

## Results and Conclusions

With the final modifications to SETPRT and JETGEN, we have tuned the fragmentation in SETPRT to agree with what we have observed in the 1988–1989 CDF jet data, over all jet  $E_t$ 's. We then compared these results to the fragmentation observed in jets ranging from 10–20 GeV found in the Photon data sample. As can be seen in figures 1–11, the various fragmentation plots, in both simulation and data, agree at both low and high  $E_t$ .

The main reason for tuning the fragmentation is to model the non-linearity correctly. Figure 12 shows the percent of energy loss due to the non-linearity in the central hadron in both the data and simulation are in agreement. Figure 13 shows this same agreement combined with changing the tracking efficiency, to reflect the systematic errors we have associated with the tracking efficiency. These changes are: changing the efficiency +7% (1ST), changing the efficiency -7% (2ND), associating the tracking efficiency to one track instead of two (3RD), and pinning the efficiency at 250  $GeV$  instead of 400  $GeV$ .

This agreement observed in the simulation with the data for the fragmentation, together with the tuning in QFL of the single pion response, gives us confidence that we will be able to extract both jet energy corrections and jet resolution out to 450  $GeV$  in  $E_t$ , and that the simulation will yield accurate results.

## References

- [1] CDF-1066, "Single Pion Response in the Central Calorimeter".

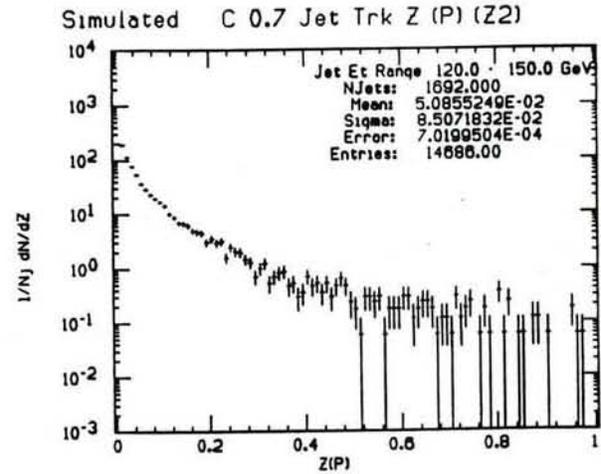
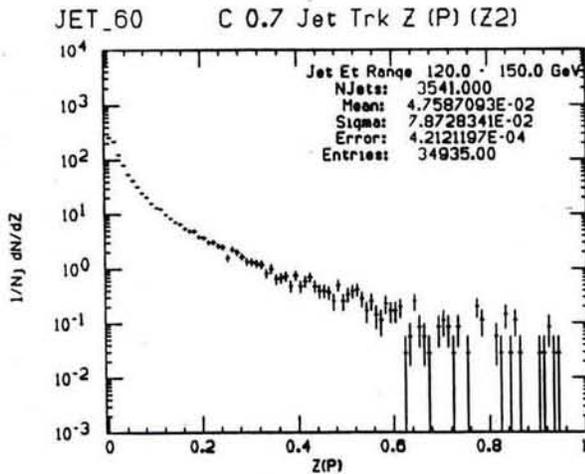
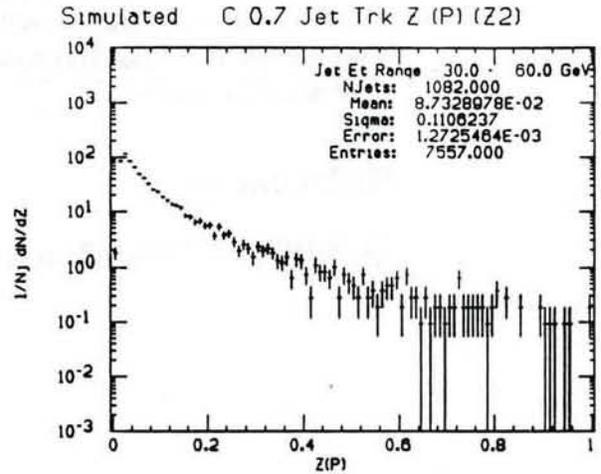
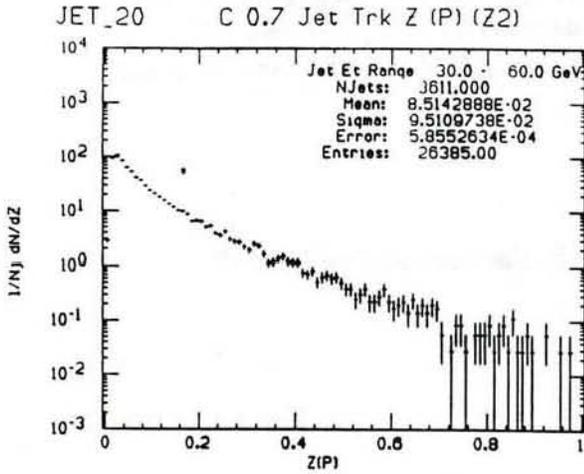
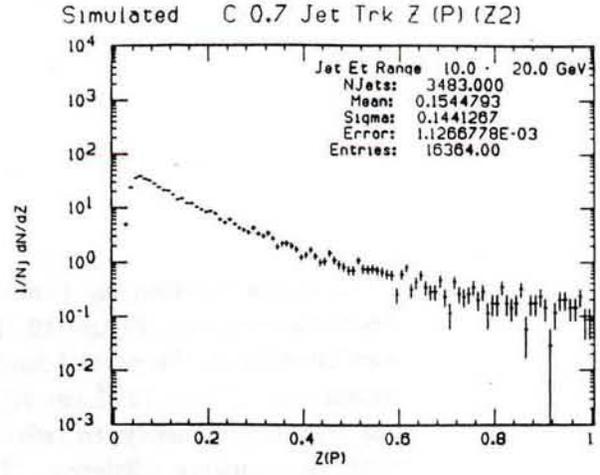
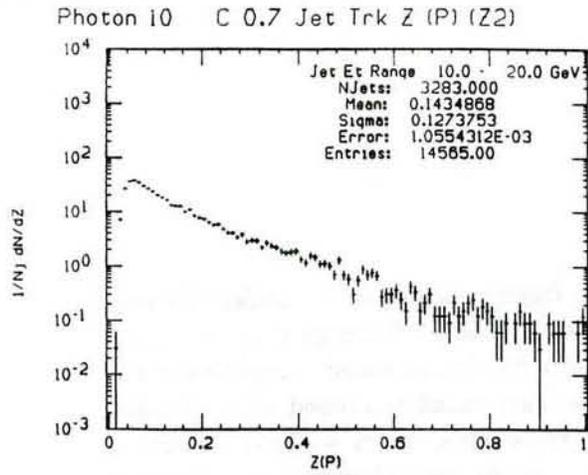


Figure 1

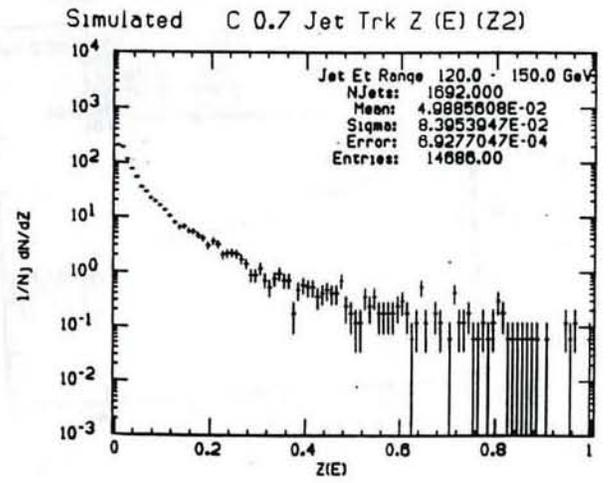
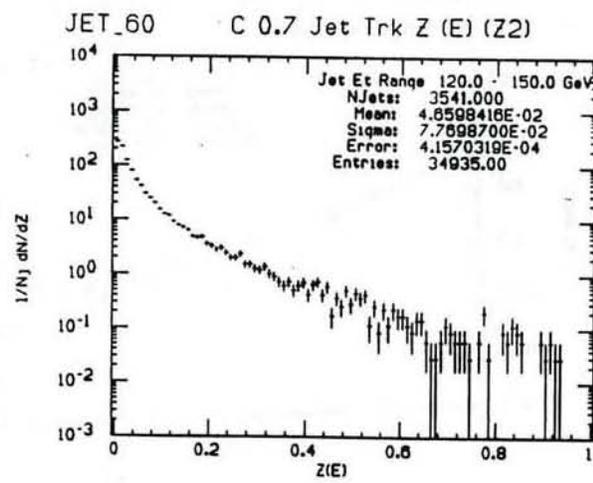
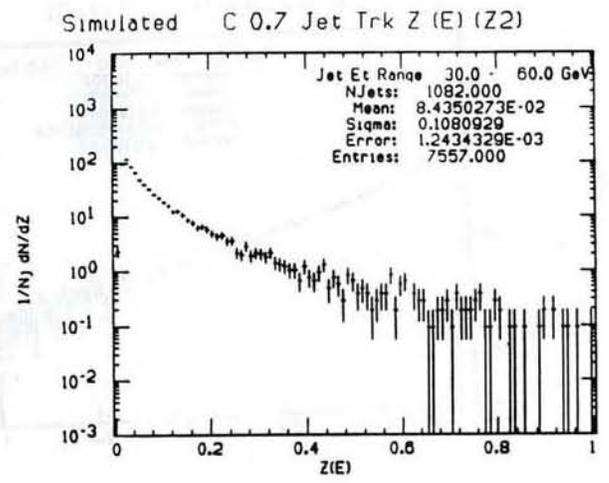
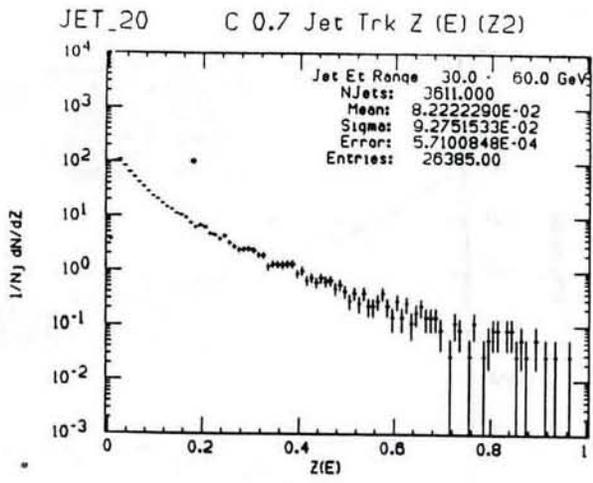
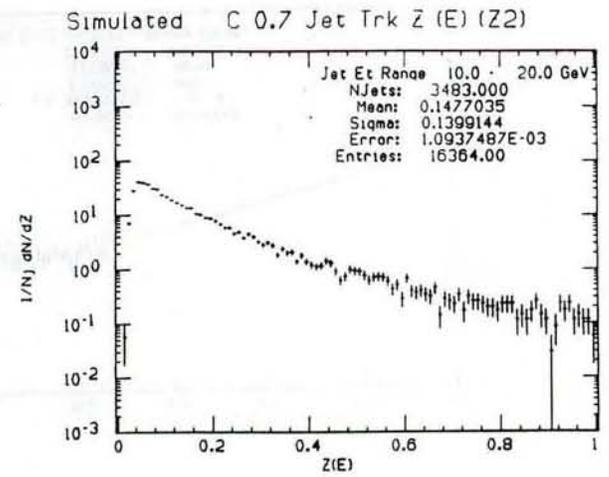
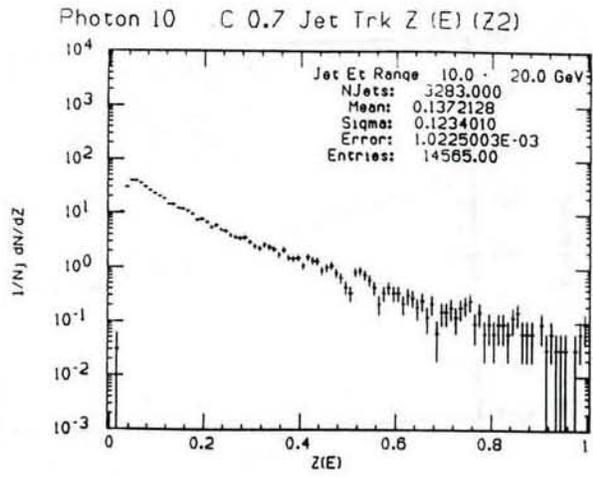


Figure 2

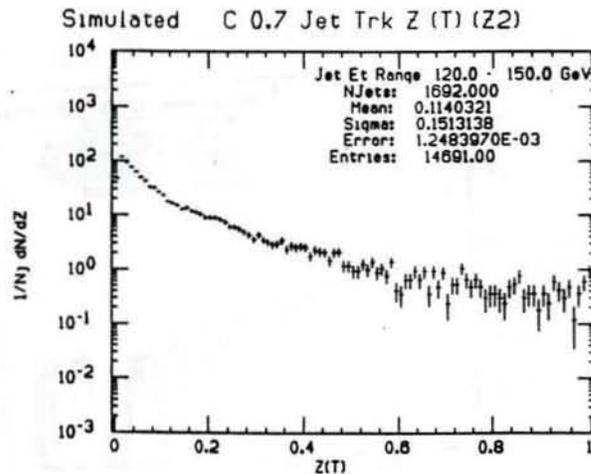
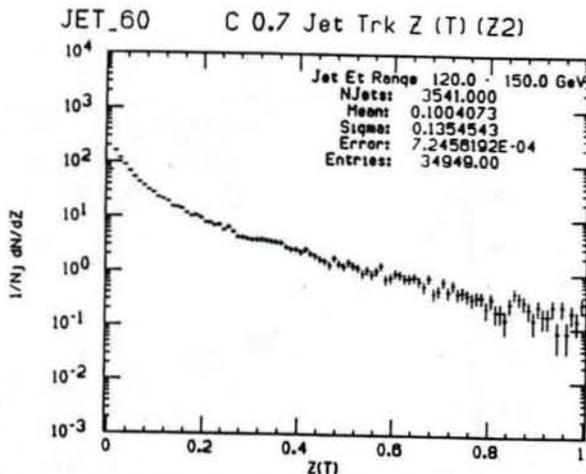
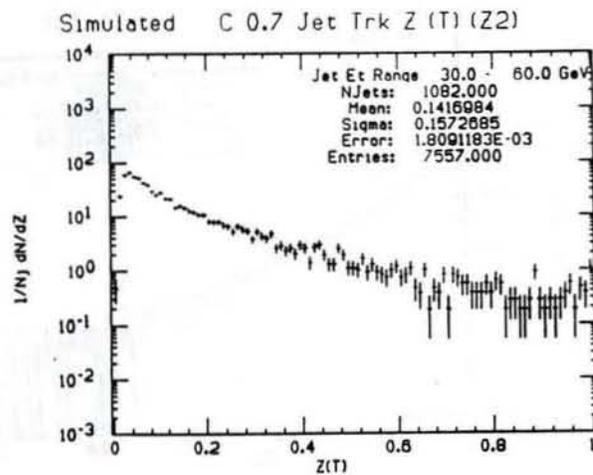
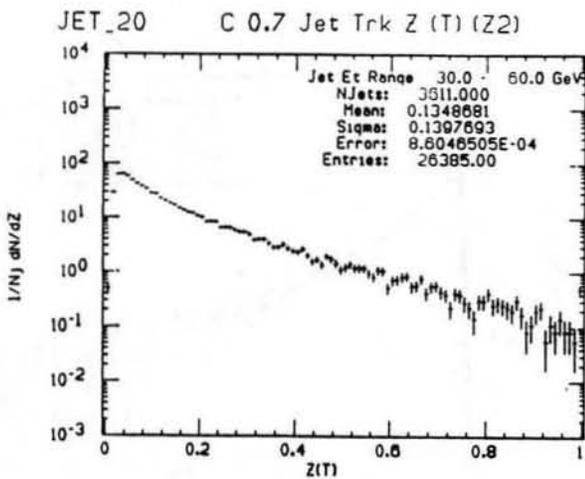
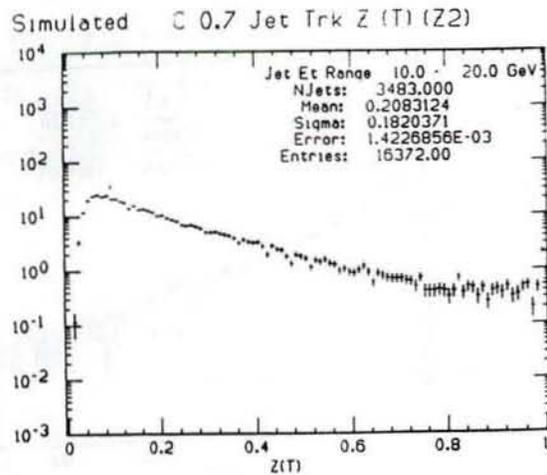
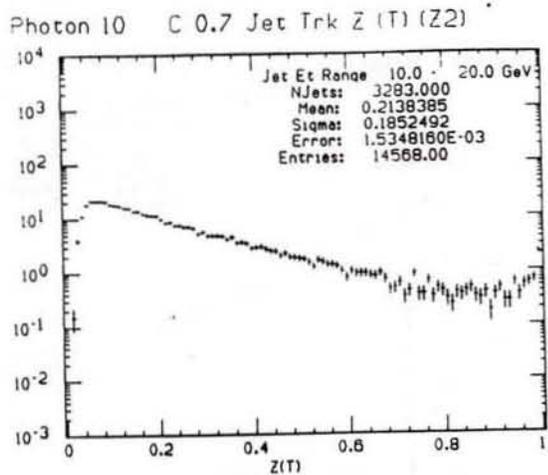


Figure 3

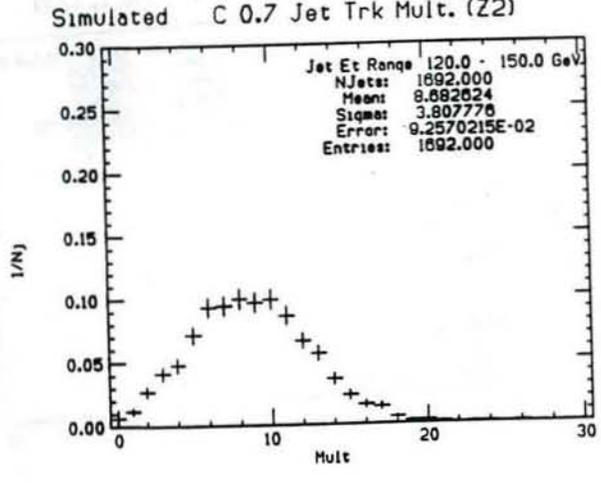
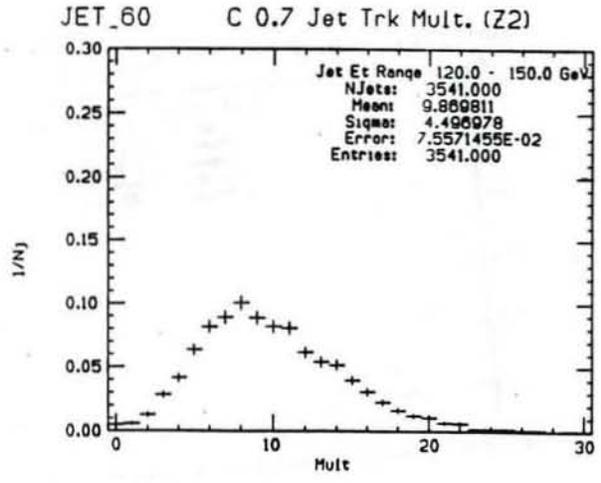
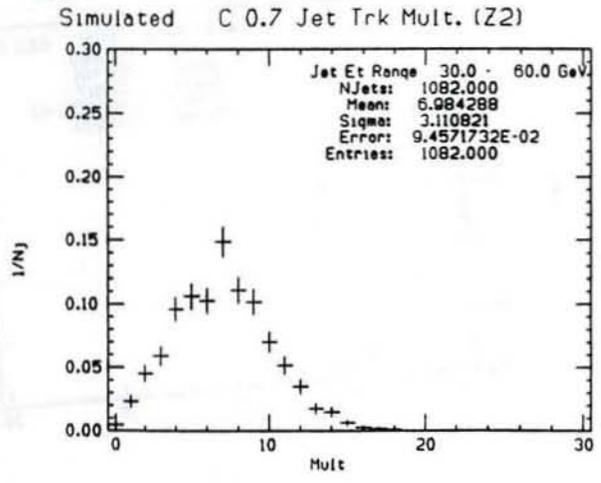
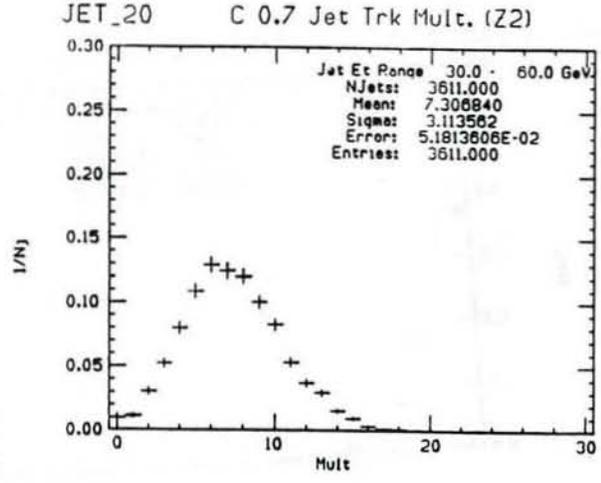
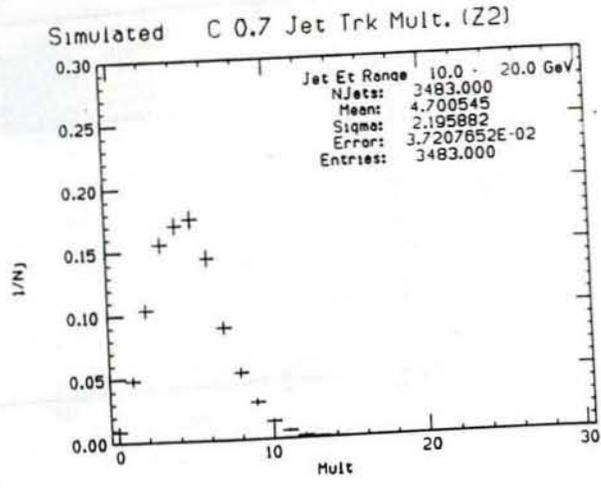
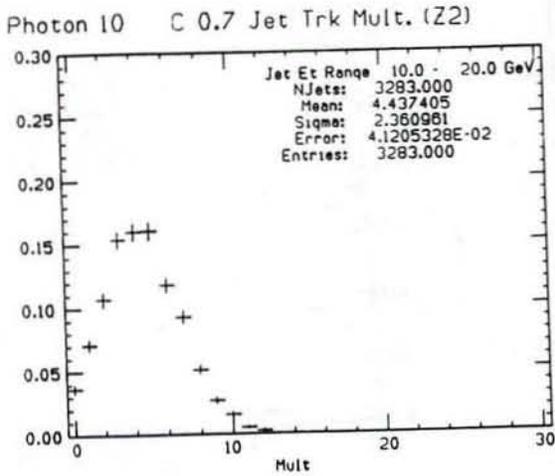
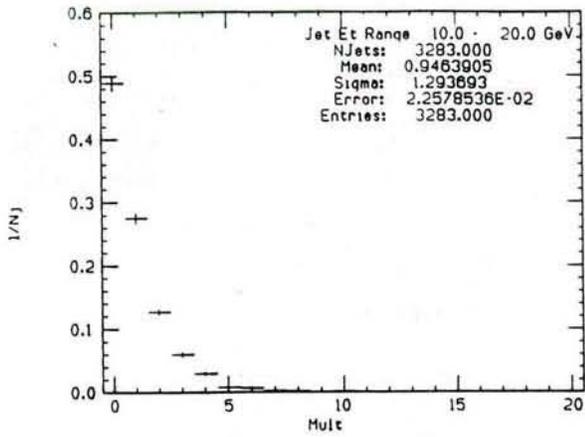
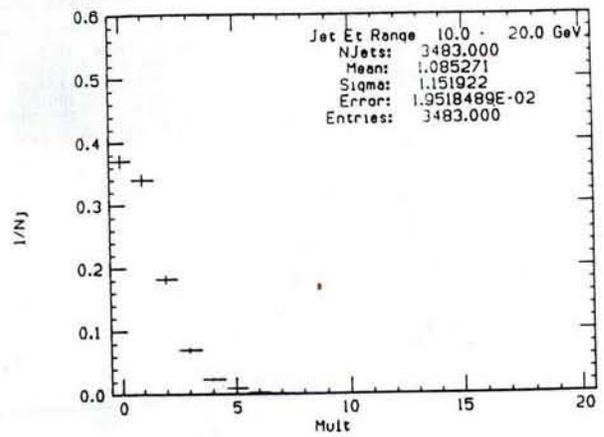


Figure 4

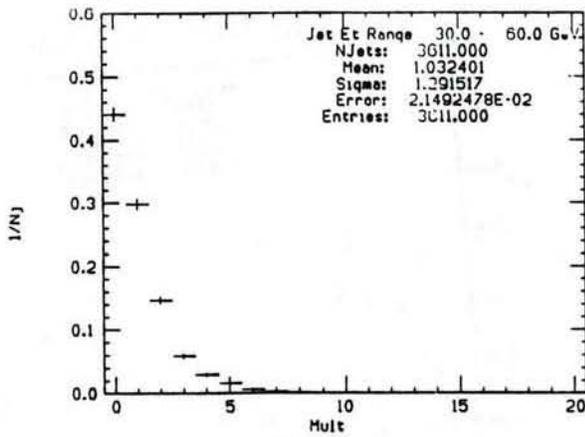
Photon 10 C 0.7 Und Trk Mult. (Z2)



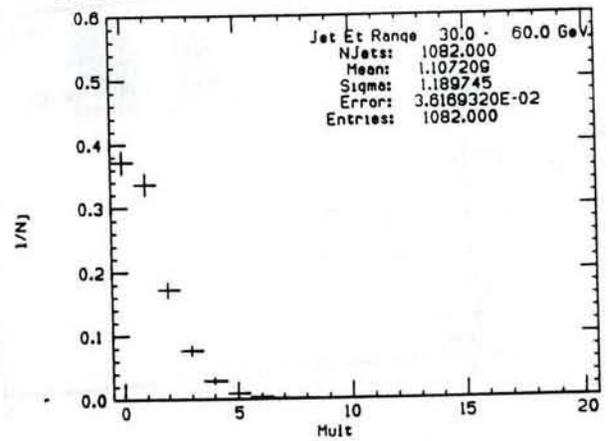
Simulated C 0.7 Und Trk Mult. (Z2)



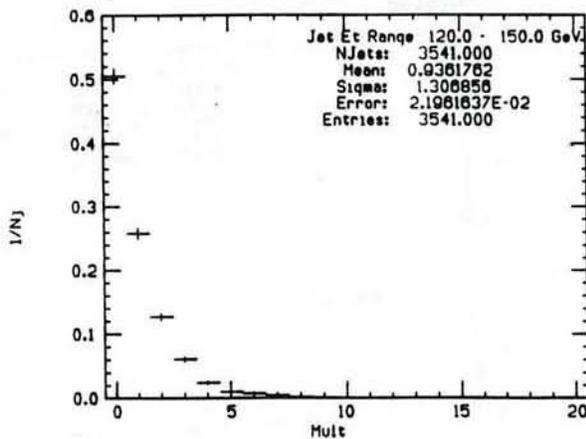
JET\_20 C 0.7 Und Trk Mult. (Z2)



Simulated C 0.7 Und Trk Mult. (Z2)



JET\_60 C 0.7 Und Trk Mult. (Z2)



Simulated C 0.7 Und Trk Mult. (Z2)

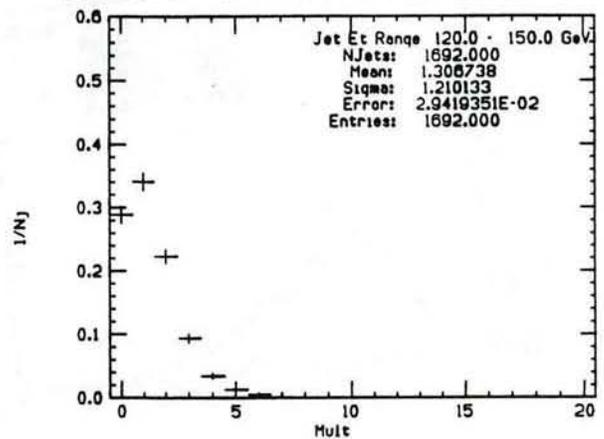
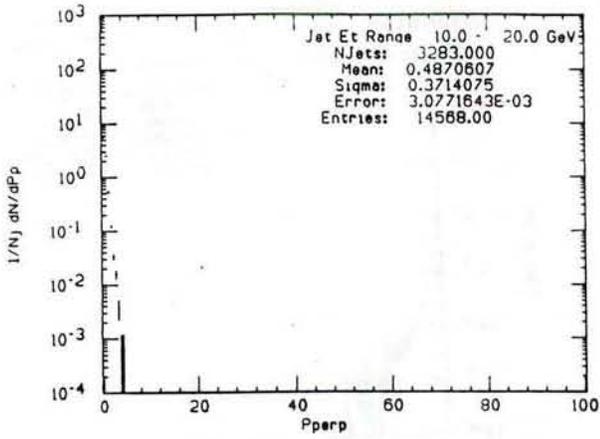
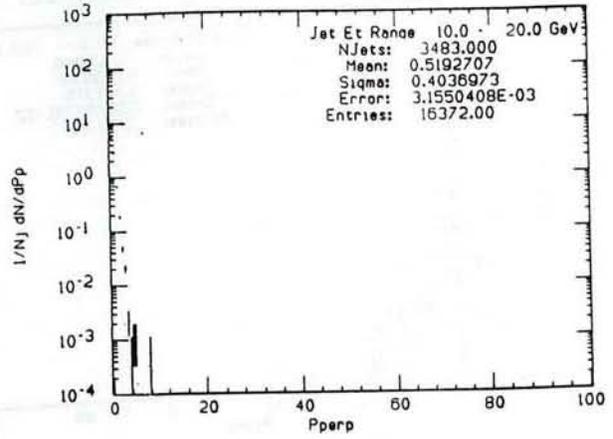


Figure 5

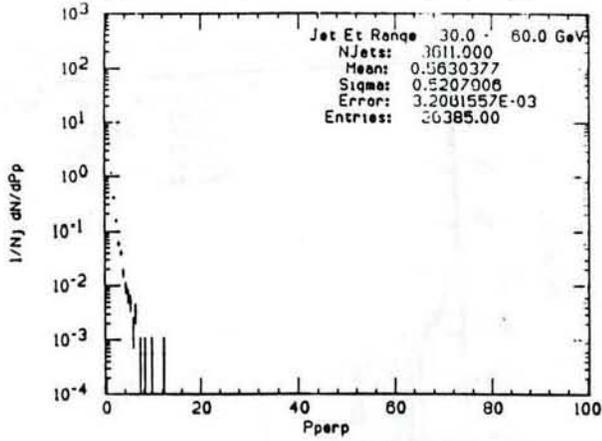
Photon 10 C 0.7 Jet Trk P perp (Z2)



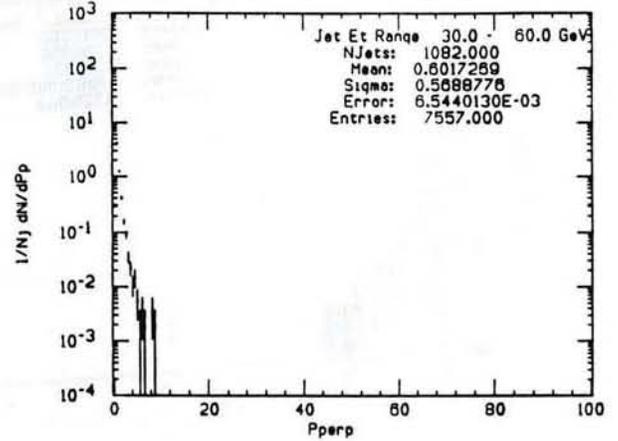
Simulated C 0.7 Jet Trk P perp (Z2)



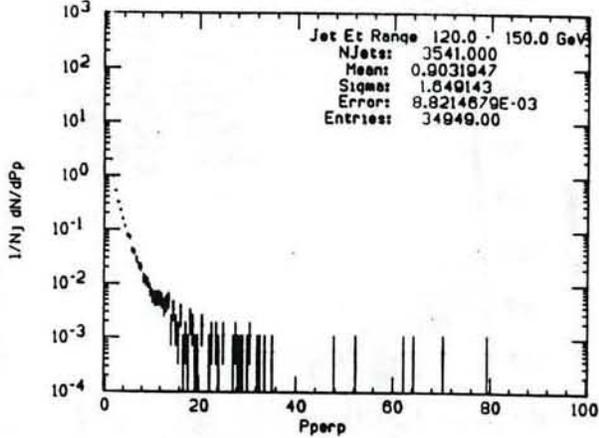
JET\_20 C 0.7 Jet Trk P perp (Z2)



Simulated C 0.7 Jet Trk P perp (Z2)



JET\_60 C 0.7 Jet Trk P perp (Z2)



Simulated C 0.7 Jet Trk P perp (Z2)

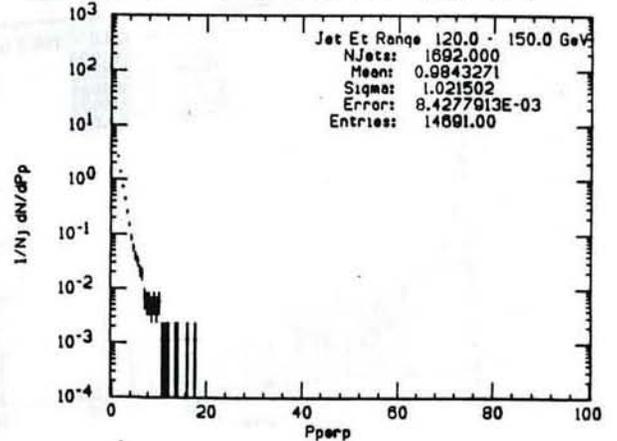


Figure 6

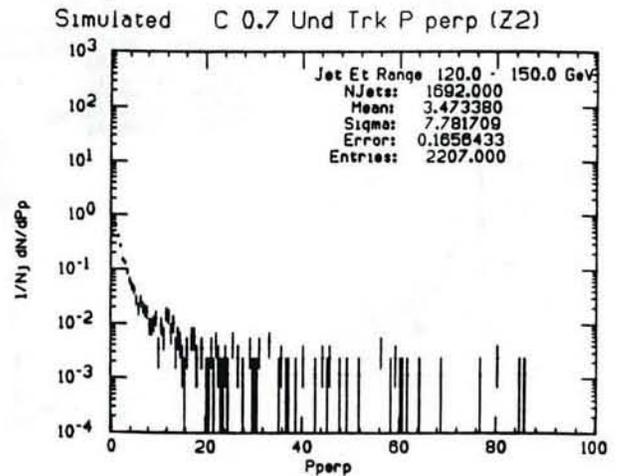
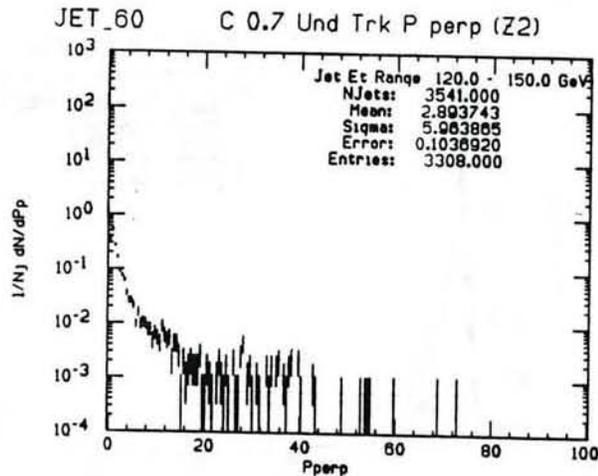
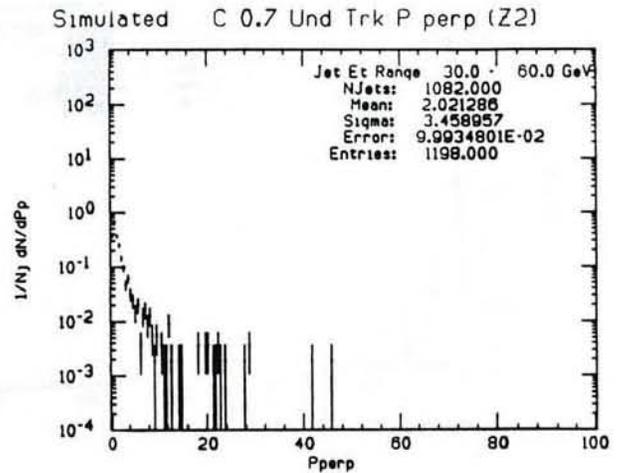
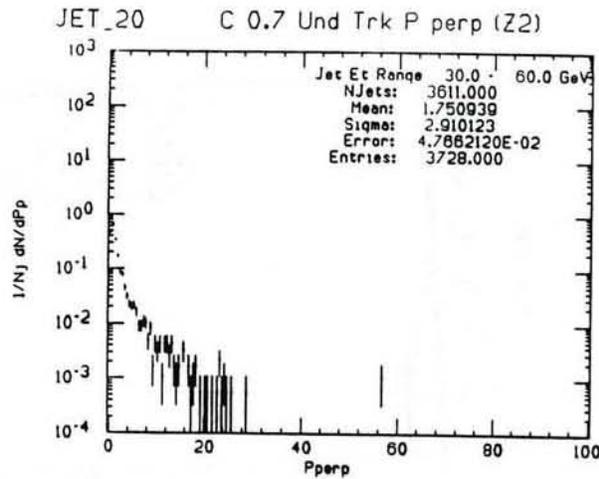
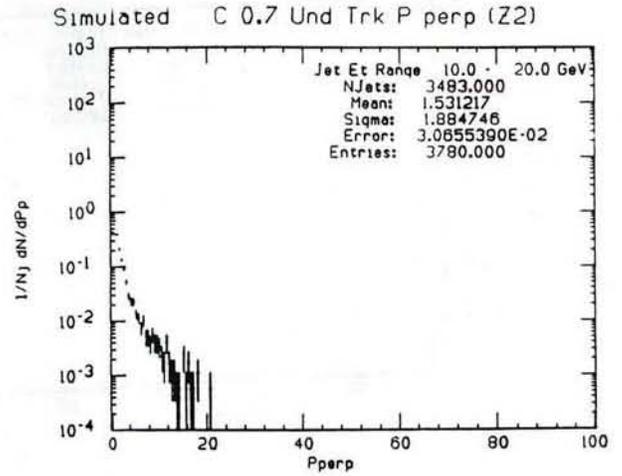
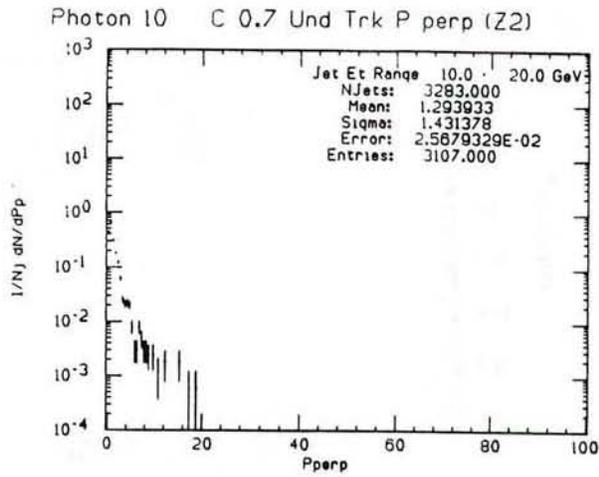


Figure 7

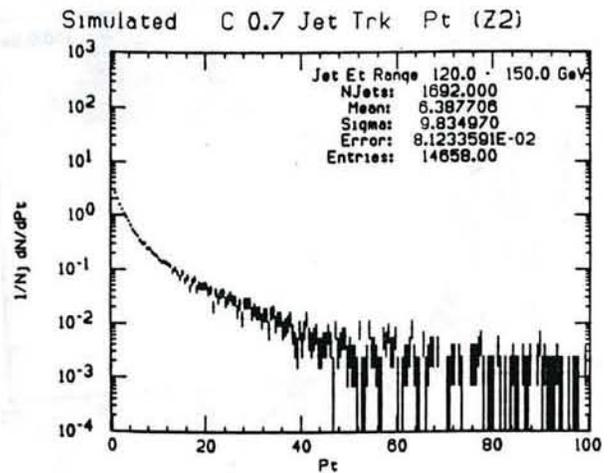
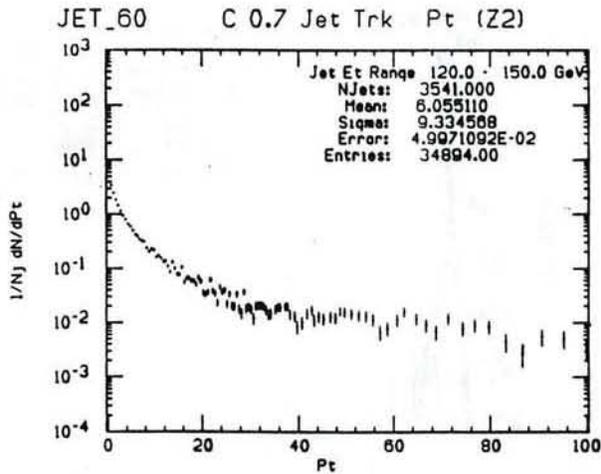
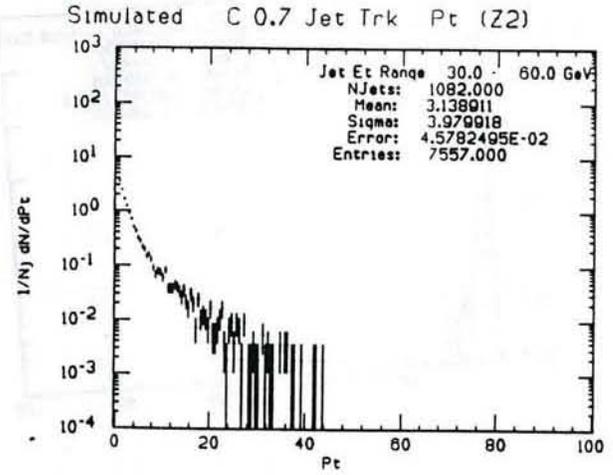
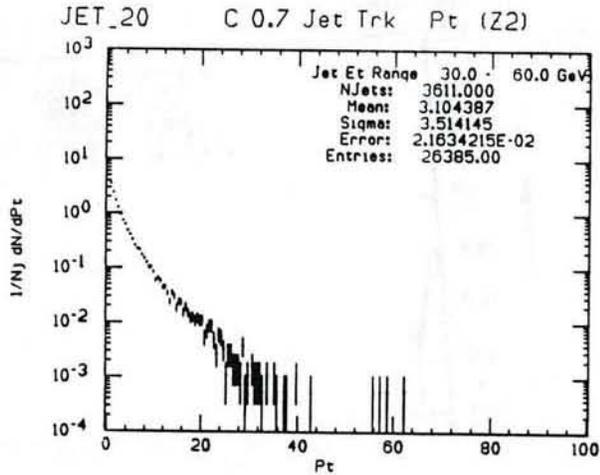
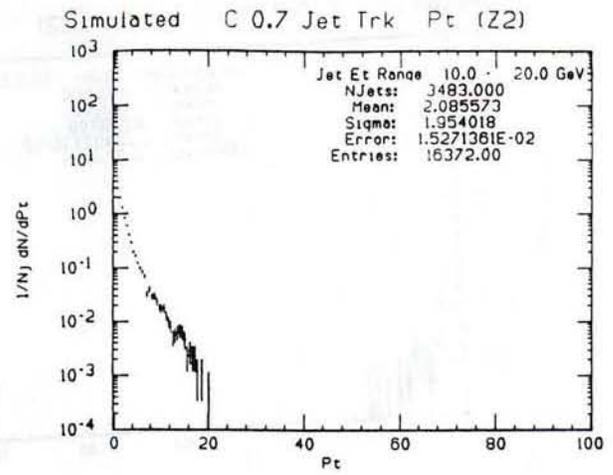
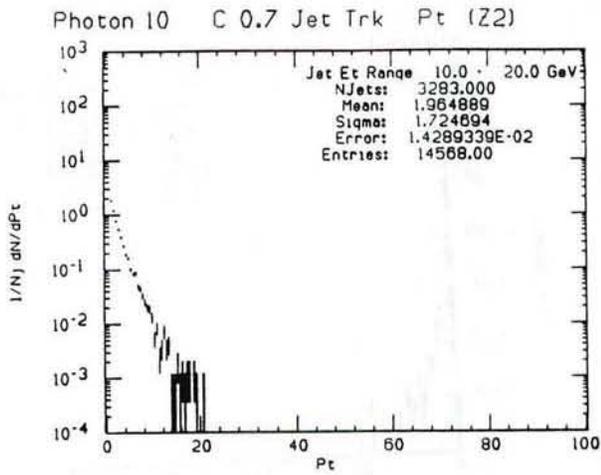


Figure 8

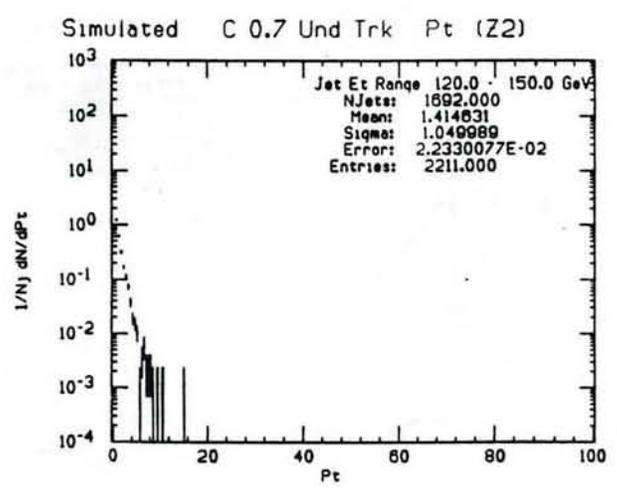
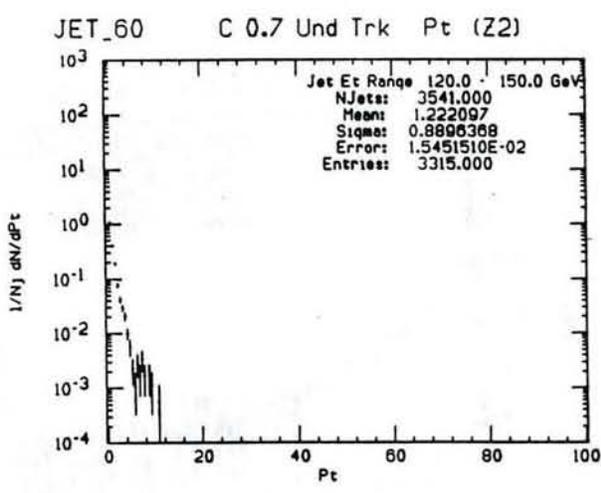
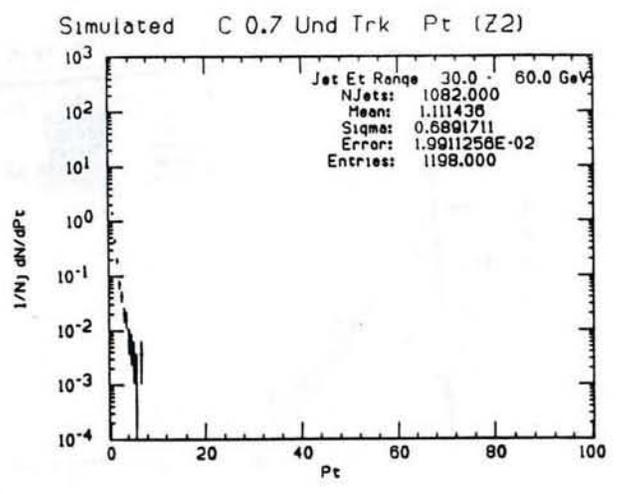
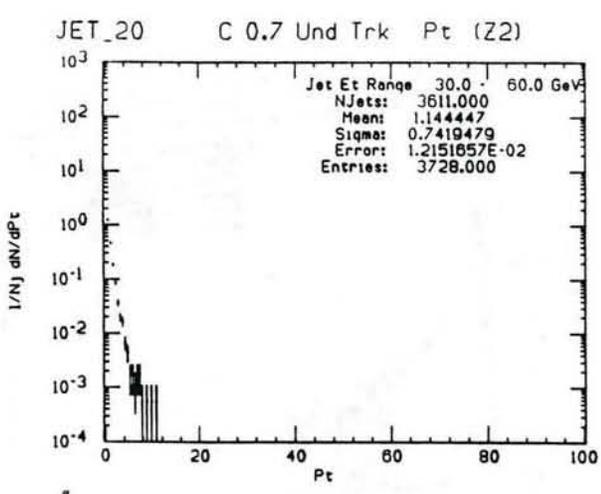
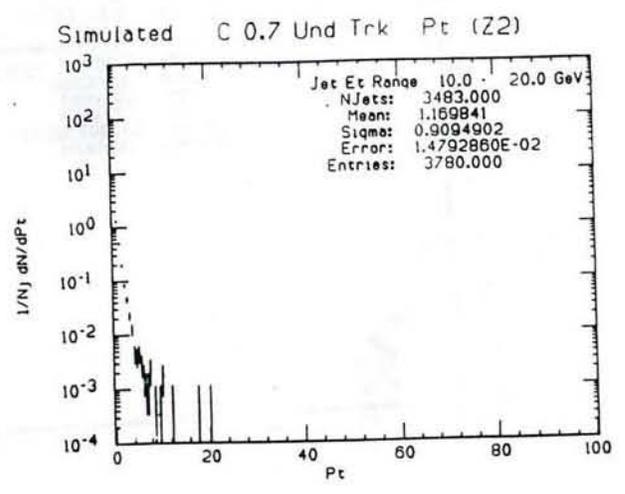
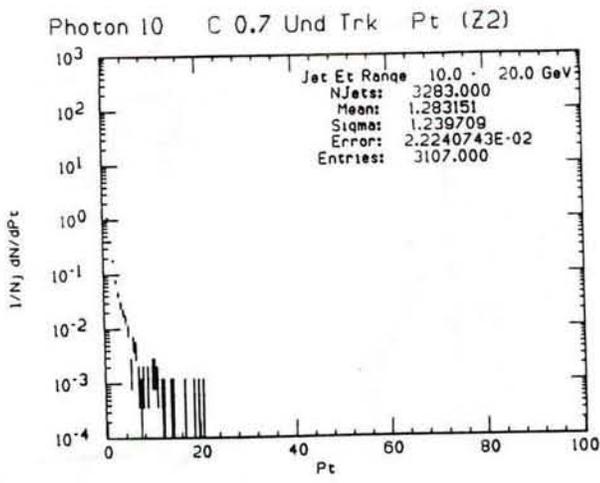
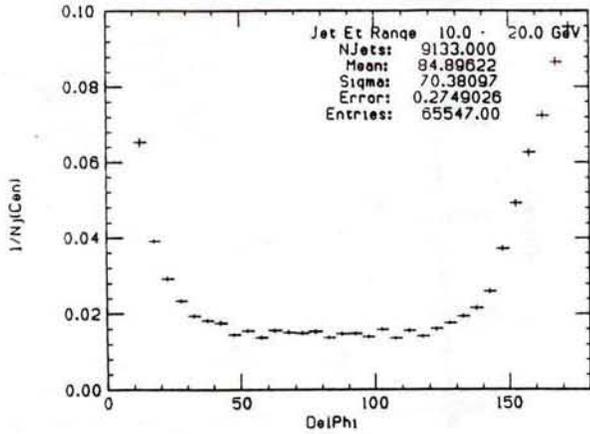
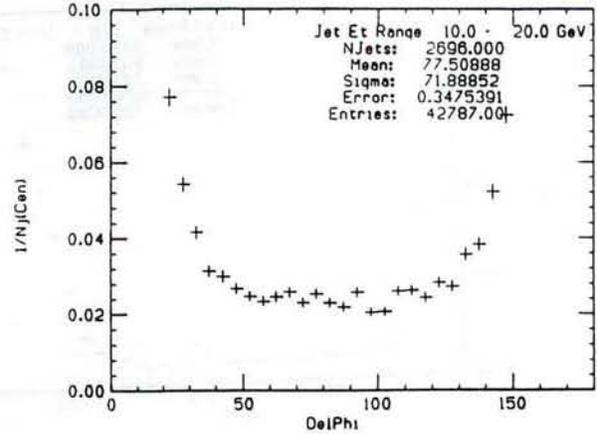


Figure 9

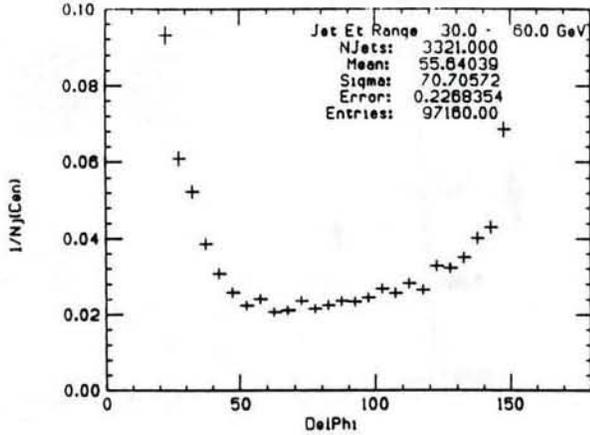
Photon 10 C 0.7 Trk Pt Flow (L)(Z2)



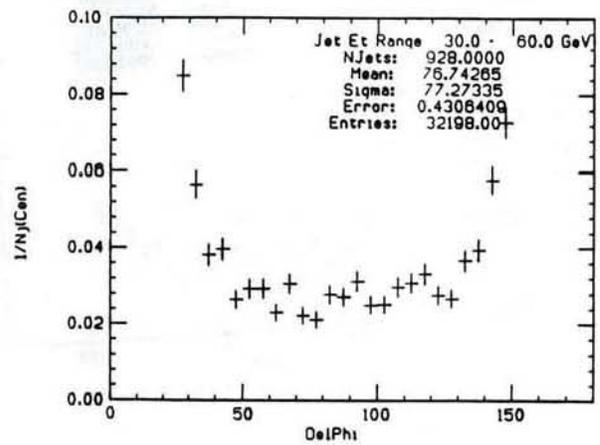
Simulated C 0.7 Trk Pt Flow (L)(Z2)



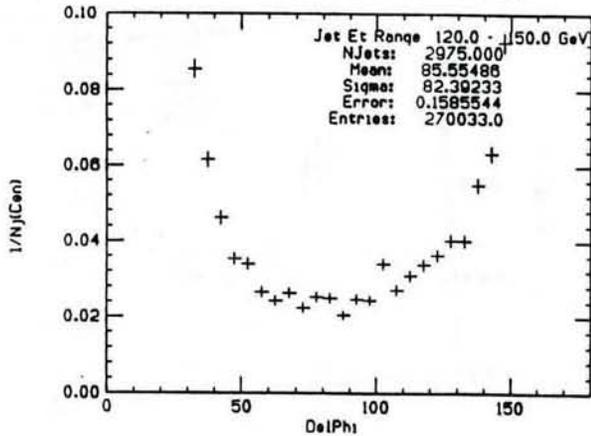
JET\_20 C 0.7 Trk Pt Flow (L)(Z2)



Simulated C 0.7 Trk Pt Flow (L)(Z2)



JET\_60 C 0.7 Trk Pt Flow (L)(Z2)



Simulated C 0.7 Trk Pt Flow (L)(Z2)

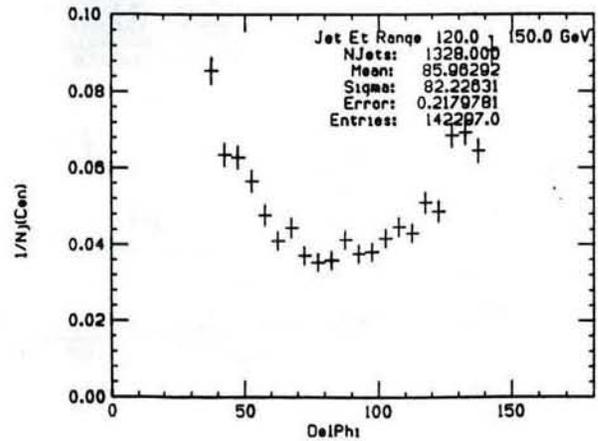
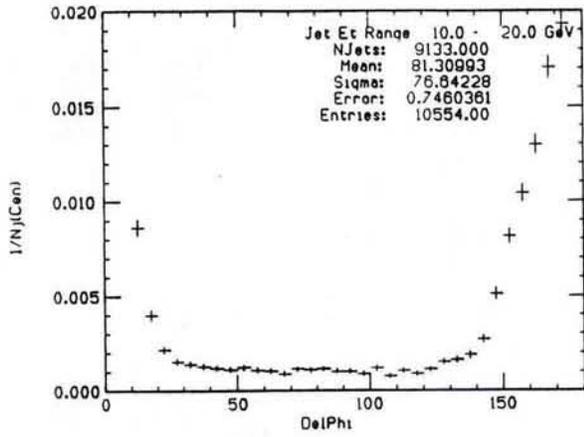
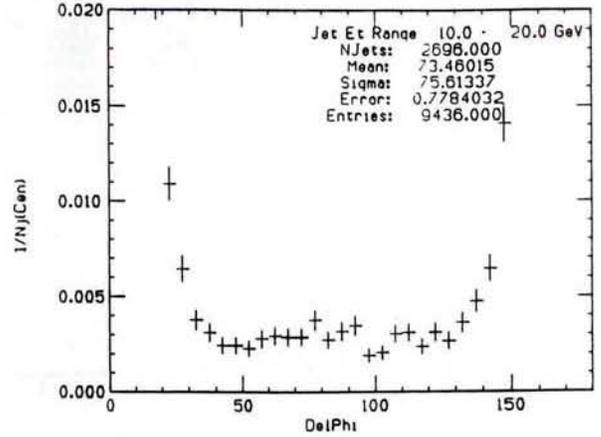


Figure 10

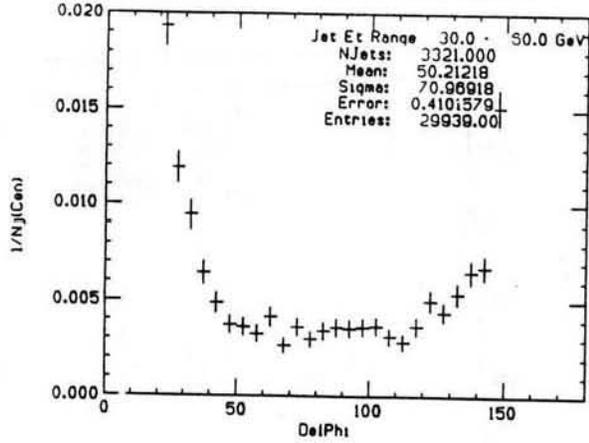
Photon 10 C 0.7 Trk Loss Flow (L)(Z2)



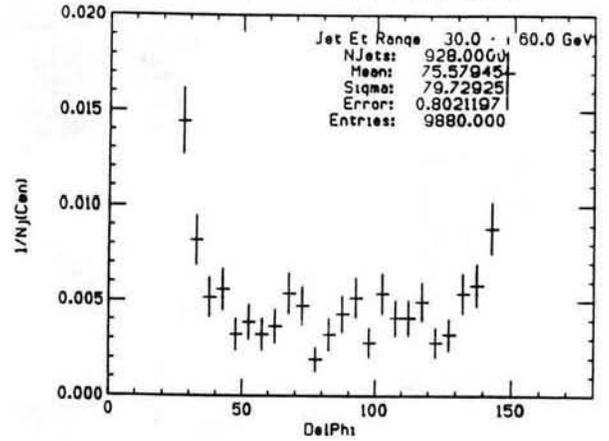
Simulated C 0.7 Trk Loss Flow (L)(Z2)



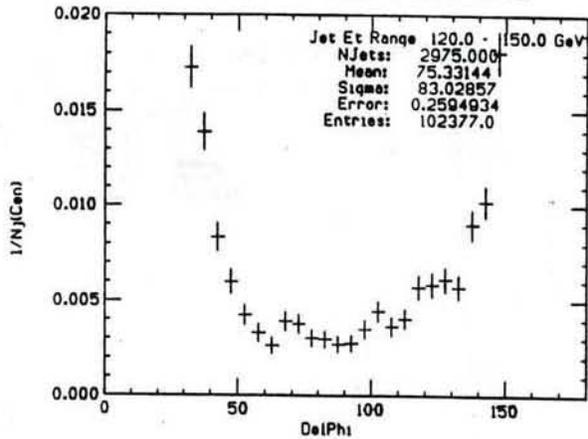
JET\_20 C 0.7 Trk Loss Flow (L)(Z2)



Simulated C 0.7 Trk Loss Flow (L)(Z2)



JET\_60 C 0.7 Trk Loss Flow (L)(Z2)



Simulated C 0.7 Trk Loss Flow (L)(Z2)

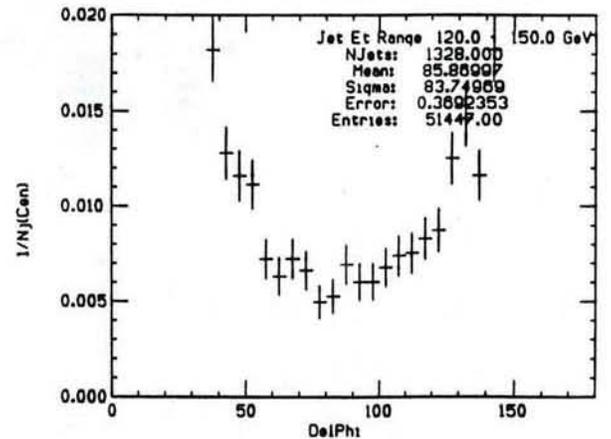
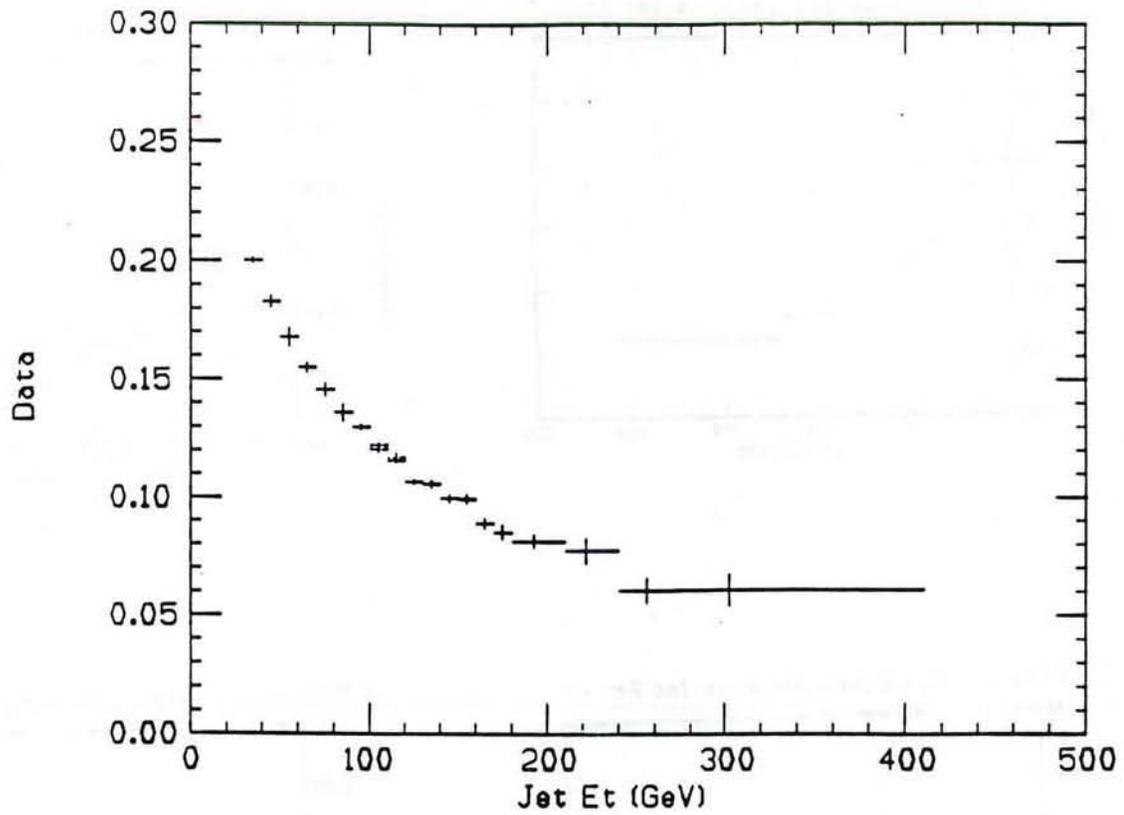


Figure 11

$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet } E \text{ vs Jet } E_t$



$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet } E \text{ vs Jet } E_t$

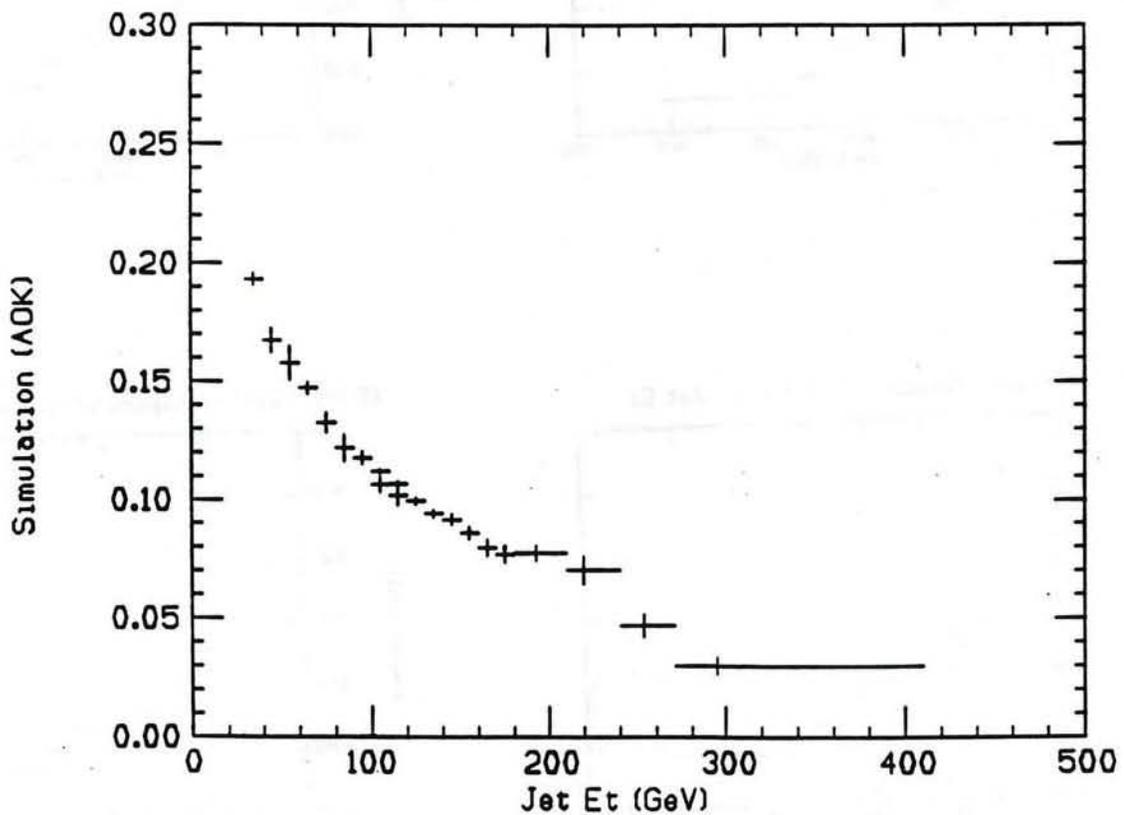
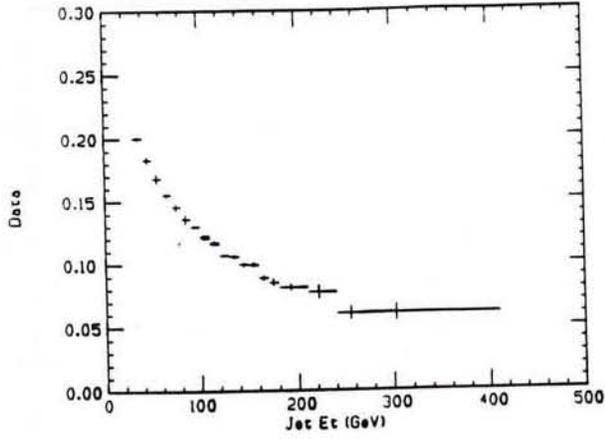
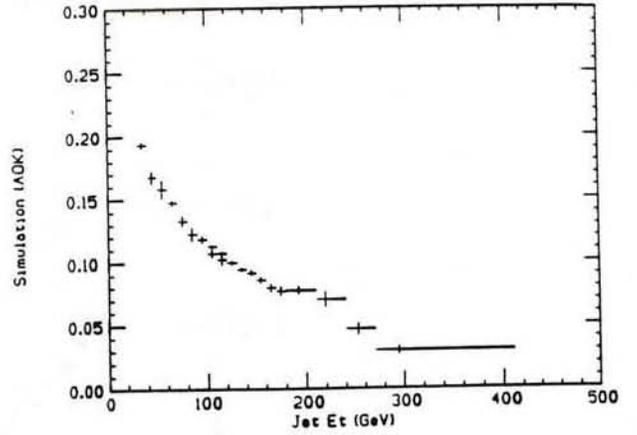


Figure 12

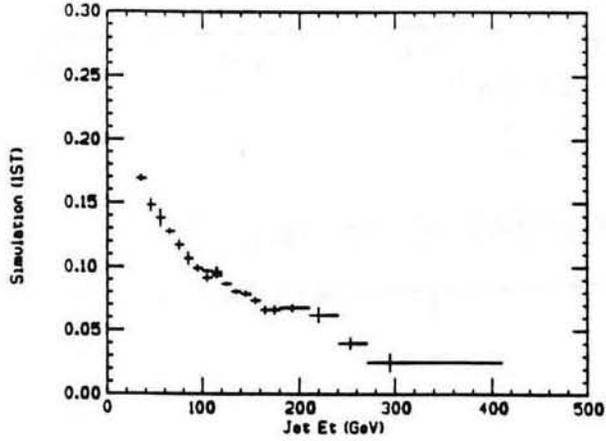
$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$



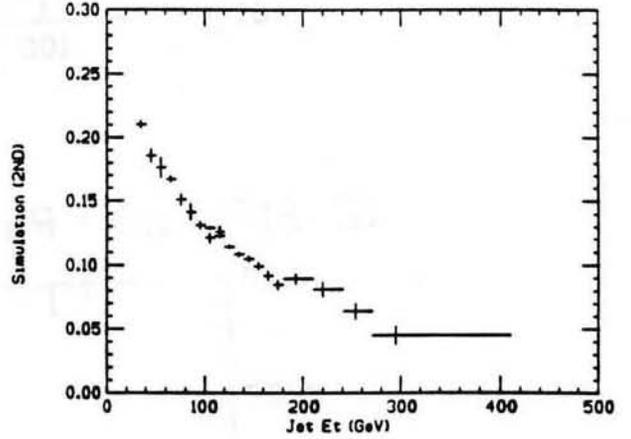
$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$



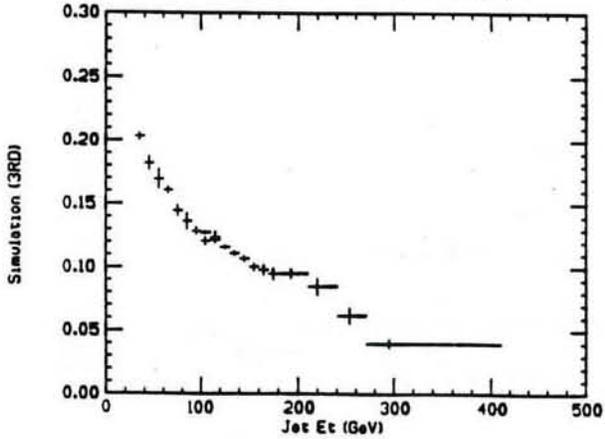
$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$



$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$



$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$



$(\Sigma (P(\text{Track}) - \text{Response}))/\text{Jet E vs Jet Et}$

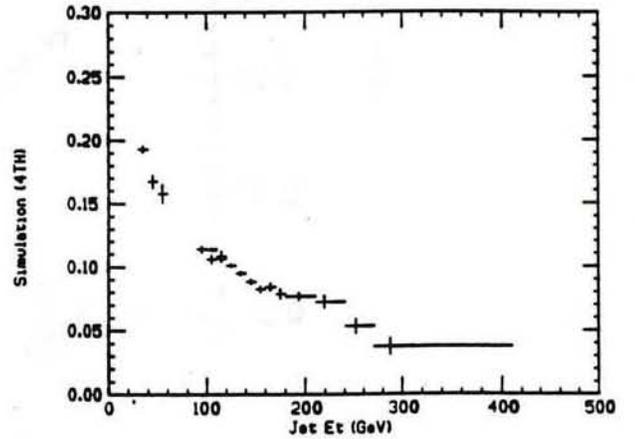


Figure 13