1 Studies of systematics in the Wjj cross section

This section summarizes the current state of studies into some of the systematic uncertainties in the calculation of the Wjj and Wbb production cross section.

We generated Monte Carlo samples for Wjj and Wbb with high statistics (the initial sample sizes were 2×10^9 for Wjj and 2×10^8 for Wbb). The hard scatters were generated using ALPGEN 1.3. Following reweighting, showering was done with PYTHIA, and events were then simulated using PMCS (`p14br-113` vintage). For the Wjj samples, the maximum weights used for reweighting were adjusted downwards a bit to increase the statistics; earlier studies showed that this does not bias the results.

In all cases, the W boson was forced to decay via $W \to e\nu$.

In previous studies, the mass of the b-quark used in PYTHIA was 4.8 GeV, while the mass used in ALPGEN was 4.75 GeV. For these results, the mass used in PYTHIA has been changed to 4.75 GeV.

The following generation parameters were used as the baseline. These are intended to match the samples that have been processed through full simulation.

- CTEQ5L PDFs.
- Scale: $Q^2 = M_W^2 + \sum_j m_T(j)^2$.
- $p_T(j) > 5\,\mathrm{GeV}$.
- $|\eta(j)| < 3$.
- $\Delta R(jj) > 0.4$.

The following cuts were then made on the PMCS objects:

- At least one PMCS electron candidate with $p_T > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. No other ID cuts were applied (note that PMCS doesn't simulate fake electrons).
- $\not{E}_T > 25\,\mathrm{GeV}$.
- At least two jets with $E_T > 20\,\mathrm{GeV}$, $|\eta| < 2.5$. Jets within $\Delta R < 0.4$ of electrons are removed.

For all parameter sets, we will calculate:

- The Wjj and Wbb cross section for the above selection cuts.
- The efficiency for the additional cut $100\,\mathrm{GeV} < m(jj) < 120\,\mathrm{GeV}$ (for the leading two jets).
- The $\sigma(Wbb)/\sigma(Wjj)$ cross section ratio, and the same ratio for the $m(jj)$ cut efficiency.

Note that there is no matching done with the original partons.

We also calculate the results using the matching prescription due to M. Mangano ('MLM') as described in hep-ph/0312274. The particles in the final state following the parton shower and fragmentation stages are clustered into jets using a cone algorithm. In fact, the same algorithm is used as is used to form the input to pmcs, except that the cone radius is set to 0.4 and the E_T threshold is set to be the same as the parton E_T cutoff used in generating the given sample. Then, we require that each of the original ALPGEN partons be *uniquely* contained within one of the particle jets. If this fails, the event is rejected.

Note, however, that the results presented here are based only on $W + 2$ jet ALPGEN samples. Probably we need to combine samples of different multiplicities.

For the above ('baseline') parameters, we get the following results:

	Cross section	$m(jj)$ cut eff
Wjj	$33.9 \pm 0.4\,\mathrm{pb}$	0.095 ± 0.004
Wbb	$0.189 \pm 0.001\,\mathrm{pb}$	0.086 ± 0.002
Wbb/Wjj	$(5.57 \pm 0.08) \times 10^{-3}$	0.90 ± 0.04

(unless otherwise stated, uncertainties are from MC statistics only). Here are the results with MLM matching:

	Cross section	$m(jj)$ cut eff
Wjj	$15.3 \pm 0.3\,\mathrm{pb}$	0.106 ± 0.006
Wbb	$0.129 \pm 0.001\,\mathrm{pb}$	0.089 ± 0.003
Wbb/Wjj	$(8.42 \pm 0.16) \times 10^{-3}$	0.84 ± 0.05

The absolute cross section numbers given here are, of course, not meaningful, as this study doesnt use realistic particle ID or the full analysis cuts. We will, however, look at ratios of cross sections; further numbers will be quoted as ratios to these baseline numbers.

1.1 PDF variations

We generated samples wiht the following alternate PDFs. Recall that the baseline used CTEQ5L, which is a LO PDF.

- CTEQ6L1, the LO version of the CTEQ6 PDFs.
- There is a full error analysis available for the NLO CTEQ6M PDF set. This provides 41 different PDFs: a central value, and the results of going out $\pm 1\sigma$ on each of the 20 different eigenvectors of the fit. Resources did not permit generating 40 samples to explore all eigenvectors. However, CTEQ reports that they saw the biggest change in results by varying eigenvector 15; accordingly, we chose the PDF sets corresponding to the central value and to changes along that eigenvector: CTEQ61.00, CTEQ61.29, and CTEQ61.30.
- The LO version of the MRST01 PDFs, MRSTLO.
- The NLO MRST01 PDFs. There is a central value with $\alpha_s = 0.119$, and two more with $\alpha_s = 0.117$ and 0.121. These are MRST01-1, MRST01-2, and MRST01-3, respectively.

Here are results, tabulating the ratios with respect to the baseline. First, for the cross sections, without matching:

	Wjj	Wbb	Wbb/Wjj
CTEQ61.00	0.949 ± 0.016	0.949 ± 0.010	1.001 ± 0.020
CTEQ61.29	0.917 ± 0.015	0.908 ± 0.009	0.990 ± 0.019
CTEQ61.30	0.925 ± 0.016	0.919 ± 0.009	0.994 ± 0.020
CTEQ6L1	1.020 ± 0.017	1.010 ± 0.010	0.990 ± 0.019
MRST01-1	0.963 ± 0.016	0.926 ± 0.009	0.961 ± 0.019
MRST01-2	0.939 ± 0.016	0.895 ± 0.009	0.953 ± 0.019
MRST01-3	1.008 ± 0.017	0.945 ± 0.010	0.938 ± 0.018
MRSTLO	1.075 ± 0.017	1.003 ± 0.010	0.933 ± 0.018

And here with matching:

	Wjj	Wbb	Wbb/Wjj
CTEQ61.00	0.948 ± 0.024	0.951 ± 0.012	1.003 ± 0.028
CTEQ61.29	0.898 ± 0.023	0.914 ± 0.011	1.018 ± 0.029
CTEQ61.30	0.907 ± 0.023	0.929 ± 0.012	1.024 ± 0.029
CTEQ6L1	1.028 ± 0.026	1.020 ± 0.013	0.992 ± 0.028
MRST01-1	0.970 ± 0.025	0.916 ± 0.012	0.945 ± 0.027
MRST01-2	0.947 ± 0.024	0.895 ± 0.011	0.945 ± 0.027
MRST01-3	1.020 ± 0.025	0.945 ± 0.012	0.927 ± 0.026
MRSTLO	1.083 ± 0.027	1.004 ± 0.013	0.927 ± 0.026

Here the results for the $m(jj)$ cut efficiencies, first without matching. Again, these numbers are ratios relative to the baseline efficiencies.

	Wjj	Wbb	Wbb/Wjj
CTEQ61.00	1.057 ± 0.056	1.036 ± 0.035	0.980 ± 0.061
CTEQ61.29	1.046 ± 0.056	1.085 ± 0.036	1.038 ± 0.065
CTEQ61.30	1.003 ± 0.054	1.037 ± 0.036	1.034 ± 0.066
CTEQ6L1	1.057 ± 0.055	1.066 ± 0.037	1.009 ± 0.063
MRST01-1	0.979 ± 0.053	1.064 ± 0.036	1.087 ± 0.069
MRST01-2	1.008 ± 0.054	1.060 ± 0.036	1.051 ± 0.067
MRST01-3	1.081 ± 0.056	1.075 ± 0.037	0.994 ± 0.062
MRSTLO	1.049 ± 0.054	1.078 ± 0.037	1.027 ± 0.063

And again with matching.

	Wjj	Wbb	Wbb/Wjj
CTEQ61.00	1.045 ± 0.078	1.024 ± 0.041	0.980 ± 0.083
CTEQ61.29	1.011 ± 0.077	1.074 ± 0.043	1.062 ± 0.091
CTEQ61.30	0.974 ± 0.075	1.022 ± 0.042	1.049 ± 0.091
CTEQ6L1	1.026 ± 0.075	1.052 ± 0.043	1.025 ± 0.086
MRST01-1	0.912 ± 0.070	1.050 ± 0.043	1.152 ± 0.100
MRST01-2	0.949 ± 0.073	1.040 ± 0.043	1.095 ± 0.095
MRST01-3	1.028 ± 0.075	1.069 ± 0.043	1.040 ± 0.087
MRSTLO	1.027 ± 0.074	1.089 ± 0.044	1.061 ± 0.088

These data are plotted in Figure 1.

Comparing baseline with CTEQ6L1 and MRSTLO should give an indication of variation between PDF sets, and comparing CTEQ61.00 with 29 and 30 and MRST01-1 with 2 and 3 should give an indication of the uncertainties claimed for a given PDF set. The latter comparisons were done by taking the half the largest variation seen, relative to the central value.

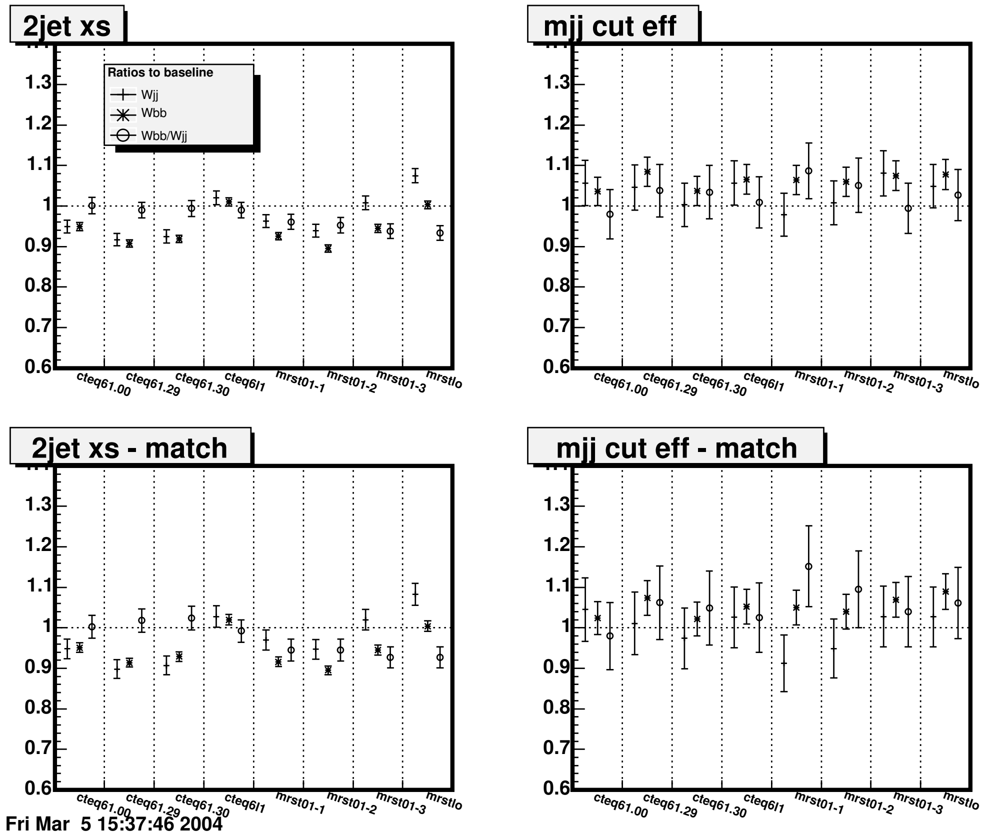

Figure 1: PDF variations

Estimated uncertainties due to PDF choice, first without matching:

	Wjj	Wbb	Wbb/Wjj
baseline vs. CTEQ6L1	0.020	0.010	0.010
baseline vs. MSRTLO	0.075	0.003	0.066
CTEQ61.*	0.017	0.022	0.005
MRST01-*	0.036	0.027	0.012
total	0.087	0.036	0.068

And here with matching:

	Wjj	Wbb	Wbb/Wjj
baseline vs. CTEQ6L1	0.028	0.020	0.008
baseline vs. MSRTLO	0.083	0.004	0.073
CTEQ61.*	0.026	0.019	0.010
MRST01-*	0.038	0.027	0.010
total	0.099	0.039	0.075

The totals are summed in quadrature. (This is probably an overestimate, double-counting some effects.) So this gives effects of about 9–10%, 4%, and 7% for Wjj, Wbb, and Wbb/Wjj, respectively, regardless of whether or not matching is used.

The PDF choice does not seem to have a strong effect on the $m(jj)$ cut efficiency, at least at this level of MC statistics.

1.2 Scale choice

For the baseline, the factorization/renormalization scale Q was set to $Q^2 = m_W^2 + \sum_j m_T(j)^2$. We generated samples with this Q scaled by factors of 0.5, 0.75, 1.5, and 2, as well as with the $Q^2 = M_W^2$ choice. Here are the results for the cross sections, without matching.

	Wjj	Wbb	Wbb/Wjj
$Q^2 = M_W^2$	1.036 ± 0.017	1.067 ± 0.010	1.030 ± 0.019
$Q = 0.5Q_0$	0.971 ± 0.016	1.074 ± 0.012	1.107 ± 0.022
$Q = 0.75Q_0$	1.032 ± 0.017	1.044 ± 0.011	1.012 ± 0.020
$Q = 1.5Q_0$	0.909 ± 0.015	0.882 ± 0.009	0.971 ± 0.019
$Q = 2.0Q_0$	0.828 ± 0.014	0.820 ± 0.008	0.990 ± 0.020

And with matching:

	Wjj	Wbb	Wbb/Wjj
$Q^2 = M_W^2$	1.099 ± 0.027	1.114 ± 0.013	1.014 ± 0.027
$Q = 0.5Q_0$	1.345 ± 0.032	1.264 ± 0.016	0.940 ± 0.025
$Q = 0.75Q_0$	1.166 ± 0.028	1.105 ± 0.014	0.947 ± 0.026
$Q = 1.5Q_0$	0.826 ± 0.022	0.847 ± 0.011	1.026 ± 0.030
$Q = 2.0Q_0$	0.757 ± 0.020	0.775 ± 0.010	1.024 ± 0.031

Here the results for the $m(jj)$ cut efficiencies, without matching:

	Wjj	Wbb	Wbb/Wjj
$Q^2 = M_W^2$	1.033 ± 0.054	1.067 ± 0.034	1.033 ± 0.063
$Q = 0.5Q_0$	1.023 ± 0.054	0.973 ± 0.036	0.952 ± 0.062
$Q = 0.75Q_0$	1.026 ± 0.053	0.994 ± 0.035	0.968 ± 0.061
$Q = 1.5Q_0$	1.064 ± 0.057	1.136 ± 0.038	1.067 ± 0.067
$Q = 2.0Q_0$	1.152 ± 0.061	1.136 ± 0.038	0.986 ± 0.062

And again with matching:

	Wjj	Wbb	Wbb/Wjj
$Q^2 = M_W^2$	0.980 ± 0.072	1.057 ± 0.039	1.078 ± 0.088
$Q = 0.5Q_0$	0.995 ± 0.069	0.970 ± 0.041	0.976 ± 0.079
$Q = 0.75Q_0$	1.016 ± 0.072	0.978 ± 0.041	0.962 ± 0.080
$Q = 1.5Q_0$	0.978 ± 0.077	1.126 ± 0.045	1.152 ± 0.102
$Q = 2.0Q_0$	0.966 ± 0.078	1.076 ± 0.044	1.114 ± 0.101

These data are plotted in Figure 2. The cross section ratio results for the four $Q = xQ_0$ points are also plotted in Figure 3. The point $(0, 1)$ for the baseline has also been added. The x-axis is the base-2 log of the Q-factor x.

In principle, one could try to use data to constrain Q. However, at this point, we simply take the largest variation seen over the range 0.5–2 as the full width of an uncertainty. This results in the following relative uncertainties:

	No matching	Matching
Wjj	0.102	0.294
Wbb	0.127	0.244
Wbb/Wjj	0.068	0.042

For this variable, applying the matching prescription increases the observed variation; however, the slope is similar in the Wjj and Wbb samples, so it largely cancels in the ratio. The difference between the non-matched and matched cases occurs at low values of Q^2. In the non-matched case,

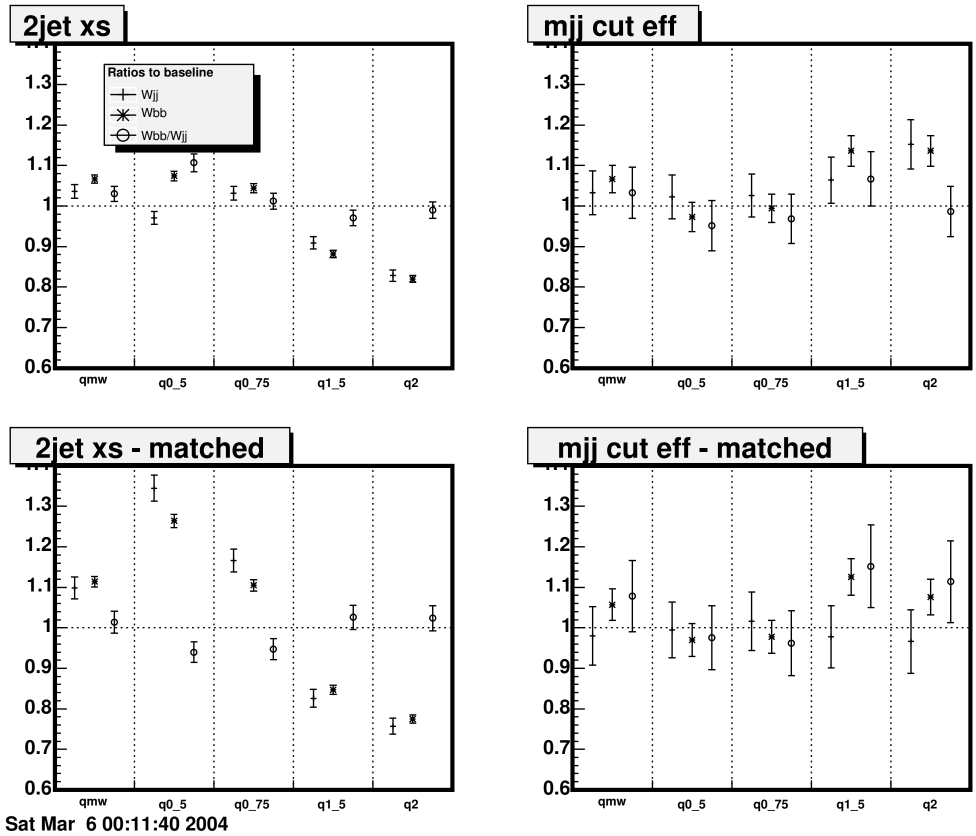

Figure 2: Scale variations

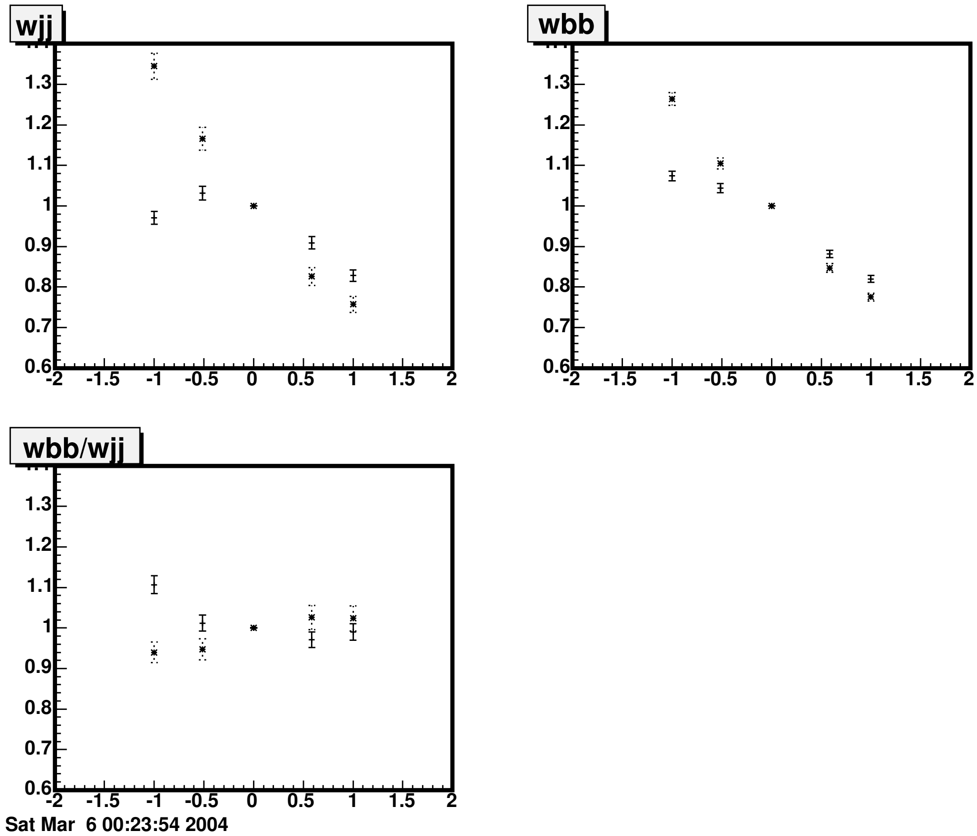

Figure 3: Scale variations vs. $\log_2 x$, where $Q = xQ_0$. Crosses with solid bars are without matching; diamonds with dashed bars are with matching.

there seems to be a saturation seen as Q^2 gets small; this is not seen in the matched case. Their behavior is similar for samples with Q^2 larger than the baseline.

Again, the variation in the $m(jj)$ cut efficiency with respect to this this parameter is not large compared to the MC statistics.

1.3 Shower MC dependence

The baseline sample was also run using HERWIG to do the showering instead of PYTHIA. Here are the results, as ratios from the baseline:

	No match		Match	
	Cross section	$m(jj)$ eff	Cross section	$m(jj)$ eff
Wjj	0.804 ± 0.014	1.212 ± 0.064	0.978 ± 0.025	1.054 ± 0.078
Wbb	0.851 ± 0.009	1.141 ± 0.040	0.905 ± 0.012	1.057 ± 0.044
Wbb/Wjj	1.058 ± 0.022	0.941 ± 0.060	0.926 ± 0.026	1.003 ± 0.085

Here, applying the matching prescription reduces the variation seen in the Wjj and Wbb samples, for both the cross section and $m(jj)$ cut efficiency. The ratio of the cross sections, though, is somewhat different with and without matching.

[Earlier work showed a much smaller difference between the two generators. This seems to have been due to the bug in the DØ interface to PYTHIA that turned off the modeling of the underlying event. This bug was present in the earlier studies but not in this one; the effect of that bug was about the proper magnitude and direction to cancel the effect seen here.]

1.4 Generator jet E_T cut

For the baseline, the ALPGEN jet E_T cut was set to 5 GeV. We also tried doing the generation with the cut set to 2.5, 4, 7, and 10 GeV. Here are the ratios relative to the baseline, first for the cross sections, without matching:

	Wjj	Wbb	Wbb/Wjj
> 2.5 GeV	1.356 ± 0.021	1.039 ± 0.012	0.766 ± 0.015
> 4 GeV	1.110 ± 0.018	1.014 ± 0.011	0.913 ± 0.018
> 7 GeV	0.851 ± 0.015	0.939 ± 0.009	1.103 ± 0.022
> 10 GeV	0.686 ± 0.012	0.855 ± 0.008	1.247 ± 0.025

And here with MLM matching:

	Wjj	Wbb	Wbb/Wjj
$> 2.5\,\mathrm{GeV}$	1.248 ± 0.030	1.049 ± 0.015	0.841 ± 0.023
$> 4\,\mathrm{GeV}$	1.061 ± 0.026	1.019 ± 0.014	0.961 ± 0.027
$> 7\,\mathrm{GeV}$	0.924 ± 0.024	0.934 ± 0.011	1.011 ± 0.028
$> 10\,\mathrm{GeV}$	0.783 ± 0.021	0.874 ± 0.009	1.116 ± 0.032

Here the results for the $m(jj)$ cut efficiencies, without matching:

	Wjj	Wbb	Wbb/Wjj
$> 2.5\,\mathrm{GeV}$	0.974 ± 0.048	1.036 ± 0.039	1.063 ± 0.066
$> 4\,\mathrm{GeV}$	1.026 ± 0.053	1.094 ± 0.041	1.066 ± 0.068
$> 7\,\mathrm{GeV}$	0.992 ± 0.055	1.048 ± 0.034	1.056 ± 0.068
$> 10\,\mathrm{GeV}$	1.100 ± 0.062	1.076 ± 0.032	0.978 ± 0.063

And now the $m(jj)$ cut efficiencies, with MLM matching:

	Wjj	Wbb	Wbb/Wjj
$> 2.5\,\mathrm{GeV}$	0.898 ± 0.065	1.056 ± 0.047	1.177 ± 0.100
$> 4\,\mathrm{GeV}$	1.039 ± 0.075	1.055 ± 0.047	1.015 ± 0.087
$> 7\,\mathrm{GeV}$	1.009 ± 0.076	1.040 ± 0.040	1.031 ± 0.087
$> 10\,\mathrm{GeV}$	1.009 ± 0.080	1.080 ± 0.038	1.071 ± 0.093

These data are plotted in Figure 4 and 5.

Both Wjj and wbb depend smoothly on the parton E_T cut; there's no real sign of saturation at low E_T. However, the dependence is much less for Wbb than for Wjj. This also implies that the parton E_T cut dependence of the Wbb/Wjj ratio is much larger than that for just Wbb.

Figure 6 plots the parton E_T spectra for the Wjj and Wbb samples. The spectrum in Wbb rolls over later; this is why the cross section dependence on the parton E_T cut is less in Wbb.

Here is a summary of the variation seen, over the full range of parton E_T cut from 2.5 to 10 GeV. It can be seen that applying the matching prescription reduces the dependence:

	No matching	Matching
Wjj	0.335	0.233
Wbb	0.092	0.088
Wbb/Wjj	0.240	0.138

The variation could perhaps be reduced further in the matched case if we sum over different generated multiplicities.

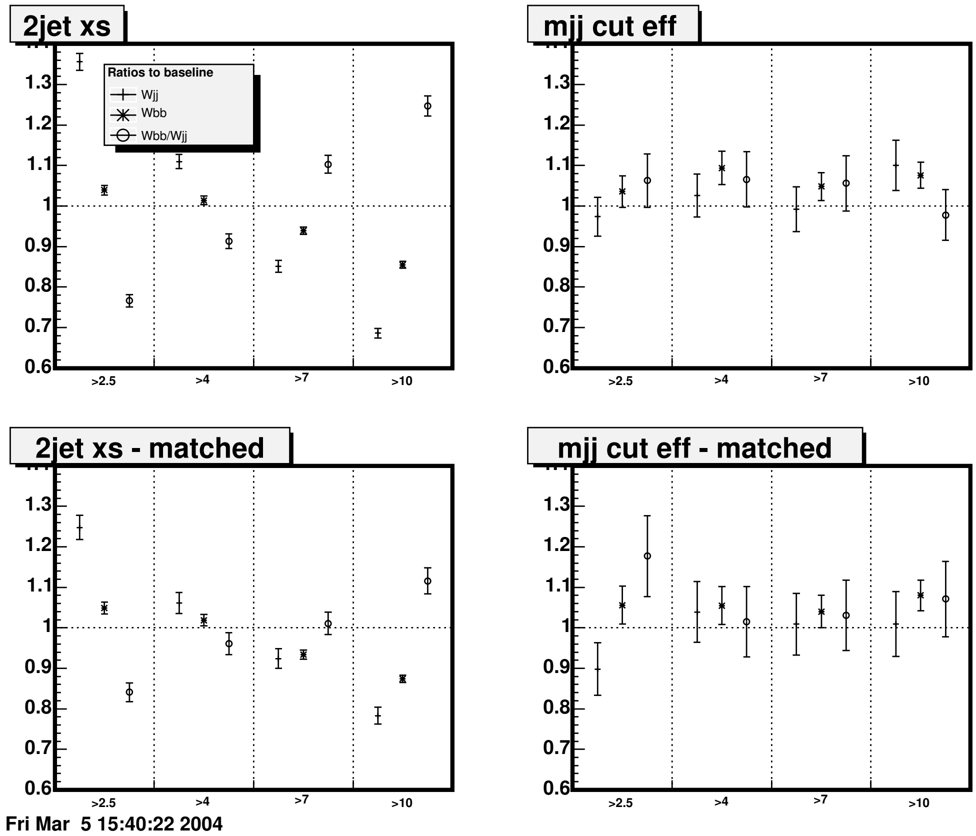

Figure 4: Baseline ratios for different generator-level jet cuts.

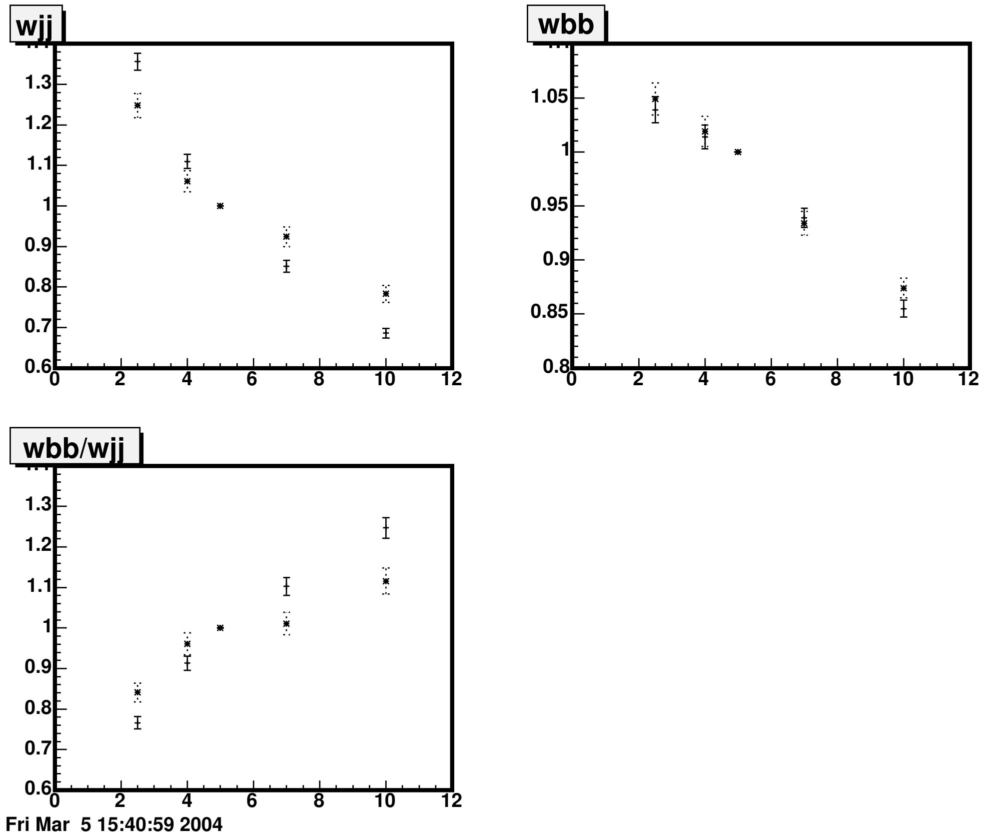

Figure 5: Baseline ratios vs. generator-level jet cut. Crosses with solid bars are without matching; diamonds with dashed bars are with matching.

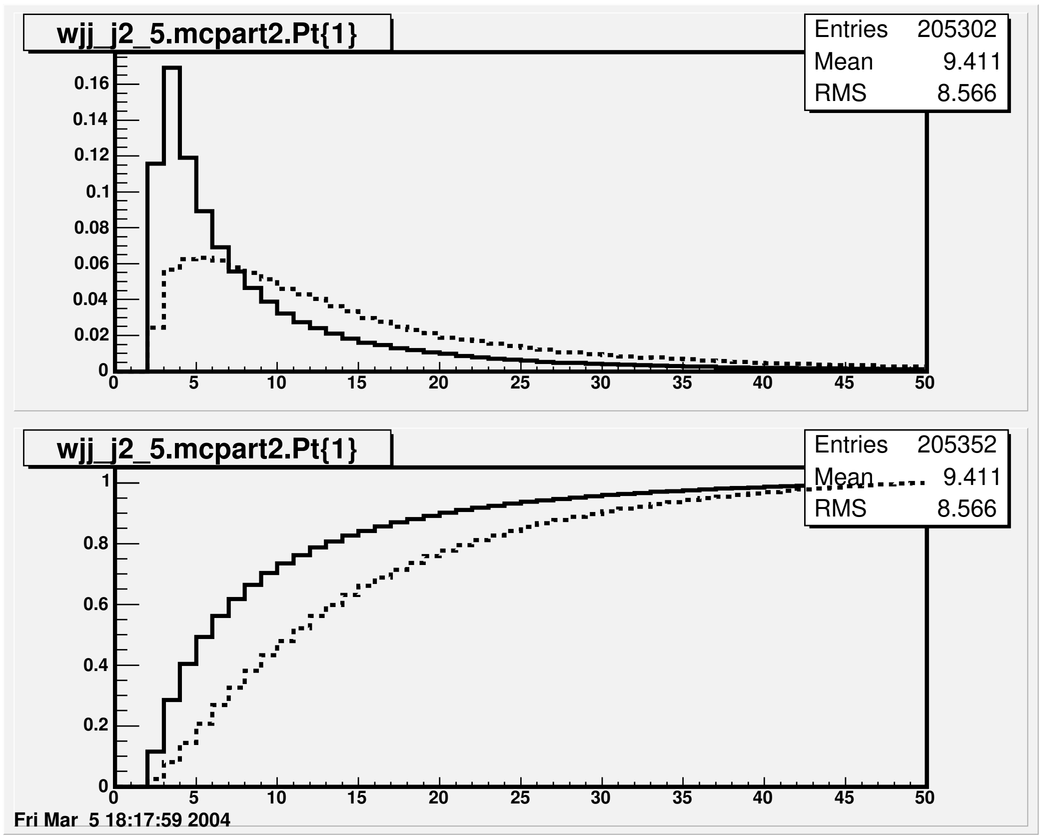

Figure 6: Parton E_T spectra for the Wjj (solid) and Wbb (dashed) samples (top). The E_T of the parton with the smallest E_T is plotted. The bottom plot is a cumulative plot of the distributions in the upper plot (normalized to 1).

1.5 Summary

Summing up the effects considered here (keeping in mind that some of them are essentially arbitrary at this point), we get for the cross sections:

	No matching	Matching
Wjj	41%	39%
Wbb	21%	36%
Wbb/Wjj	26%	18%

Applying the matching prescription causes the variation in Wbb to increase due to the behavior in the low Q^2 region. Further investigation is in progress to understand this. However, the variation with respect to Q^2 after matching is nearly the same in Wjj and Wbb, so this effect largely cancels in the Wbb/Wjj ratio, which does improve when matching is used. Also remember that these results are based on using only $W + 2$ jet ALPGEN samples, which may not be correct for the matching prescription.

The dependence of the $m(jj)$ cut on these effects is rather less. The statistical uncertainties here are about 5%. The only systematic effect seen here that clearly seems to be larger than the statistical noise is the HERWIG/prognamepythia comparison, which is about 7%. Probably claiming $\pm 10\%$ for this would be safe.