CDF/ANAL/TOP/CDFR/6970
Version 1.3
May 12, 2004

Top Mass Measurement in the Lepton + Jets Channel using a Multivariate Template Method

Igor Volobouev, Pedro Movilla Fernandez, John Freeman, Angela B. Galtieri, and Jeremy Lys, *Lawrence Berkeley National Laboratory*
with the Top Mass group and the Jet Corrections group

Abstract

We report a measurement of the top quark mass using Run II data, obtained from a sample of 162 pb^{-1} of top candidate events in the lepton plus jets channel. For each event energy-momentum conservation is used to reconstruct the top quark mass. A multivariate template method is used in a likelihood procedure that evaluates the probability for a certain mass value based on expectation from Monte Carlo. The procedure allows to fit simultaneously events with a single b-tagged jet, two b-tagged jets, and events with no tagged jets at all. We report for now an analysis that uses only events with jets tagged by SECVTX. We find a mass value of $M_{top} = 179.6^{+6.4}_{-6.3} \pm 6.8$ GeV.

Contents

1 Introduction

We report a measurement of the top quark mass obtained with a multivariate template method using top candidate events in the lepton+jets channel with one or more jets tagged as b-jets by SECVTX.

The "template method" used in Run I [1] employs a likelihood procedure to extract a top mass value from the reconstructed top masses of the candidate events. Both signal (at a number of top masses) and background reconstructed mass distributions are used in the likelihood. These distributions are called "templates".

The method discussed here differs from the traditional "template method" in that it attempts to favor the correct combination over the other combinations. In addition to the top mass template it uses templates for other variables to help distinguish signal from background. It also attempts to reduce the systematics from the jet energy scale (JES) uncertainty by performing a W mass fit for each event.

In Section 2 we discuss the data sample and the selection criteria. In Section 3 we discuss the top mass reconstruction method and the jet corrections used. In Section 4 we discuss the KDE method for templates generation. In Section 5 we discuss the likelihood method used to extract a mass from the candidate events. In Section 6 we discuss the choice of variables used to distinguish background from signal. In Section 7 we apply the method on Monte Carlo events, i.e., use pseudo-experiments to determine the expected statistical and systematic uncertainties. In Section 8 we discuss the top mass measurement and its systematic errors. Finally, Section 9 gives the conclusions.

2 Data and Monte Carlo samples

The data sample includes 162 pb^{-1}, i.e., "good silicon" data taken up to the September 2003 shutdown. The data used were obtained from the V4.11.1 REMAKE sample. The silicon tracks were refit with the recent alignment. Recent calorimeter tower corrections were taken into account. See the top group web page for details [2].

The Monte Carlo samples used are those generated for the top group. These are documented on the web [2]. The code used for generating the events was V4.9.1, the reconstruction was also done with V4.9.1.

2.1 Selection of data sample

The selection criteria are those used by the top group for the lepton+jets analyses (obtained using TopEventModule described by Thomson et al [3]). These have been used by Bachacou et al. [4] and by Maruyama et al. [5] for the measurement of the top cross section using jets tagged as b by SECVTX, by Brubaker et al. [6] and by Yorita et al. [7] for the top mass measurement that uses b-tagged jets with SECVTX.

Details of the event selection criteria used are shown in Table 1 for completeness and are summarized below.

Table 1: Cuts used to select events. Jet corrections to level 4 were used.

	Cut Variable	Cut Value
Global Event	Z and cosmic veto	applied
	$\not{E}_{\mathrm{T}}$	$\geq$ 20 GeV
	Number of tight leptons	1
	$\lvert\Delta z_0\rvert$ of COT lepton track	$\leq$ 60 cm
	$\lvert z_{lepton} - z_{eventvertex}\rvert$	$\leq$ 5 cm
Jets	E_T of the 3 highest jets	$\geq$ 15GeV
	E_T of jet-4	$\geq$ 8 GeV
	$\lvert\eta\rvert_{detector}$	$\leq$ 2
Electrons	detector	CEM in fiducial region
	E_T	$\geq$ 20 GeV
	P_T	$\geq$ 10 GeV/c
	E/P (for $P_T \leq$ 50 GeV/c)	$\leq$ 2.0
	E_{HAD}/E_{EM}	$\leq$ 0.055 + 0.00045*E
	L_{shr}	0.2
	$\lvert\Delta z\rvert$ (track z match in CES)	$\leq$ 3 cm
	Q $\times\Delta x$ (track $r-\pi$ match in CES)	-1.5 to +3.0 cm
	χ^2_{strip}	$\leq$ 10
	Number of COT axial segments	$\geq$ 3
	Number of COT stereo segments	$>$ 3
	Isolation ratio in cone of 0.4	$\leq$ 0.1
	Not a conversion	$\Delta\Lambda \geq 0.03$ or $\Delta r \leq 0.3$
Muons	detector	CMUP or CMX
	P_T	$\geq$ 20 GeV/c
	E_{EM}	$\leq$ max(2,2 + 0.0115*(P-100))
	E_{HAD}	$\leq$ max(6,6 + 0.0280*(P-100))
	$\lvert\Delta x\rvert_{CMU}$	$\leq$ 3.0 cm
	$\lvert\Delta x\rvert_{CMP}$	$\leq$ 5.0 cm
	$\lvert\Delta x\rvert_{CMX}$	$\leq$ 6.0 cm
	$\lvert d_0\rvert$ (if no Si hits)	$\leq$ 0.2 cm
	$\lvert d_0\rvert$ (if Si hits)	$\leq$ 0.02 cm
	Number of COT axial segments	$>$ 3
	Number of COT stereo segments	$>$ 3
	Isolation ratio in cone of 0.4	$\leq$ 0.1
	COT exit radius (CMX only)	$\geq$ 140cm

- Only one tight lepton with E_T (P_T) > 20 GeV for electrons (muons).
- The tight lepton is the trigger lepton.
- At least one jet tagged as b by SECVTX [8].
- Jets are corrected using level 4 corrections before event selections are applied. They are corrected to level 7 for top mass reconstruction (see Section 3.3). The jet corrections were provided by the jet correction group [9].
- Two samples are used: 3.5-jets and 4-jets. The criteria are those used in Run I with a different $|\eta|$ cut, as done by Brubaker et al. [6]. For the 3.5-jet sample, we use $E_T(jet4) > 8$ GeV and $|\eta| \leq 2.0$, whereas for the 4-jet samples, we require all 4 jets to have $E_T > 15$ GeV and the same $|\eta|$ cut.
- For the top mass fit we use top specific corrections developed by us.

The number of candidate events for topologies with different number of jets is shown in Table 2, as obtained for a Jet Energy Scale (JES) constraint of 0.035 (see Section 3.1). The last row of Table 2 shows the events left after additional requirements made before

Table 2: Candidate events left after our event selection, for a JES constraint of 0.035.

Number of tags	Number of candidate Events		
	lepton+3-jets	lepton+3.5-jets	lepton+$\geq$ 4-jets
pre-tag sample	387	144	109
$\geq$ 1 SECVTX tag		11	29
after fit+χ^2 cuts		10	21

and after kinematic fitting:

- Number of b-tags $\leq$ 2.
- Events with an electron and a conversion pair close to it (trident events) are removed because the electron E_T is not well measured.
- In the event fitting (Section 3.1), we require a neutrino solution and at least one combination with correct assignment for the b-tagged jet.
- Fit $\chi^2 < 10$.

These criteria are also used on the background events to determine the fraction of event surviving the fit, f_{fit} shown in Tables 4 and 5.

2.2 Background

For the event selection we have used the same cuts as the top cross section measurement with at least one jet tagged as a b jet by SECVTX [4]. We use the background calculations estimated with method 2 for that analysis and apply the additional criteria used for the top mass measurement. Their results (Table 24 in [4]) are shown in Table 3. The W+3-jets and the W+≥4-jets bins contain top events in addition to background, so we have to subtract this contribution. The fraction of non-top events has been calculated by the authors of Ref [4] to be $f_{nt}= 0.927$ for the W+3-jet bin and $f_{nt}=0.658$ for the W+≥4-jets (Henri/Jason private communication).

Table 3: Background summary for an integrated luminosity of 162 pb^{-1}, taken from [4]. The top contamination contribution is subtracted in column 3 and 5 where it applies. The number of observed events is from reference [4].

Background	W+3jets	W+3jets(no top)	W+≥4jets	W+≥4jets(no top)
Events before tagging	387		107	
mistags	5.0±0.8	4.6 ±0.8	2.6±0.4	1.7±0.3
W $b\bar{b}$	3.5±1.2	3.2 ±1.1	1.3±0.5	0.9±0.3
W $c\bar{c}$	0.3±0.5	1.2 ±0.5	0.5±0.2	0.3±0.1
W c	1.7±0.6	1.6 ±0.6	0.4±0.2	0.3±0.1
WW/WZ/ZZ,Z→ $\tau\tau$	0.6±0.1	0.6±0.1	0.1±0.07	0.1±0.07
non-W (QCD)	5.0±0.9	5.0 ±0.9	2.1±0.4	2.1±0.4
single top	1.1±0.2	1.1 ±0.2	0.2±0.03	0.2±0.03
Total	18.2±2.3	17.4 ±2.3	7.1±1.0	5.5±1.0
Observed ≥ 1 tag	29		28	

To evaluate the background for the top mass sample, we have to use only part of the W+3-jet bin as we require $E_T(jet4) > 8$ GeV. We determine the fraction of 3.5-jets events from the ratio 144/387=0.372 of observed events in the pretag sample, shown in Table 2. In addition, as discussed earlier, in fitting the top mass we make additional requirements on the events to select well measured candidates. For the top mass determination we will construct background templates (see Section 4) with shapes which are typically different from those of the signal events. At this time we construct separate templates for three different types of backgrounds: the mistags, the W + b-tagged events and the non-W (QCD) background. The mistags template is obtained by running the mistag matrix on the W+n jets (light flavors) Monte Carlo sample. The W+btag template is obtained from the ALPGEN Monte Carlo + HERWIG. (W+$b\bar{b}$+2jets). The non-W template is obtained from the data, requiring a non-isolated lepton (Iso > 0.2, $\not{E}_T > 20$ GeV). At this time we use the W+btag template for all the other sources of backgrounds. The single top shape is likely to be different from those of the W $b\bar{b}$ background, but it has a small contribution. We plan to use a separate template for single top in the future updates of this note. The background sources, what we use to determine their shapes, and the estimated values with the additional requirements made before and after kinematic fitting, are shown in Tables 4 and 5. The number of events is shown for the JES constraint of 0.07 (see Section 3.1).

Table 4: Background summary for an integrated luminosity of 162 pb^{-1}. The sample with 3.5-jets is corrected for the fraction of 3.5-jets in the 3-jet sample. An additional correction, f_{fit}, is made to account for several requirements made before and after the fit procedure, including the $\chi^2 < 10$ cut (see text). The JES constraint used here is 0.035.

Background	W+3jets	$f_{3.5j}$	W+3.5jets	f_{fit}	W+3.5jets final
mistags	4.64±0.79	0.372	1.73±0.29	0.521	0.90±0.15
W $b\bar{b}$,W $c\bar{c}$	4.45±1.57	0.372	1.66±0.58	0.500	0.83±0.29
Wc,WW/WZ etc.	2.18±0.66	0.372	0.81±0.25	0.500	0.41±0.13
Single top	1.1±0.2	0.372	0.41±0.07	0.500	0.21±0.04
non-W (QCD)	5.0±0.9	0.372	1.86±0.33	0.667	1.24±0.22

Table 5: Summary of backgrounds for the two samples, see Table 4 for explanation of symbols. The f_{fit} value was calculated for JES constraint of 0.035. The number of events with a mass fit refer to a value of the JES constraint of 0.07.

Background	W+3.5jets(final)	W+≥4jets	f_{fit}	W+≥4jets(final)	Total
mistags	0.90±0.15	1.71±0.26	0.498	0.85±0.13	1.75±0.28
W $b\bar{b}$, W $c\bar{c}$	0.83±0.29	1.25±0.46	0.480	0.60±0.22	1.43±0.51
Wc,WW/WZ etc.	0.41±0.13	0.36±0.15	0.480	0.17±0.07	0.58±0.20
Single top	0.21±0.04	0.20±0.04	0.464	0.09±0.02	0.30±0.06
Sum of above 3	1.45±0.34			0.86±0.24	2.31±0.58
non-W (QCD)	1.24±0.22	2.1±0.4	0.628	1.32±0.25	2.56±0.47
Total Backg	3.59±0.43			3.03±0.37	6.62±0.80
Events Observed	11			29	40
Have a mass fit	10			23	33

2.3 Monte Carlo samples used

These are the standard ones: HERWIG and PYTHIA for the top sample, ALPGEN and HERWIG for backgrounds and several sets generated for the evaluation of systematic uncertainties. The samples are listed in Table 6.

3 Top mass reconstruction

The reconstruction of the top mass is performed using a special-purpose kinematic fitter. The difference between this fitter and the standard least-squares kinematic constraint fitting method applied in the LBNL Run 1 top mass analysis [11] lies in the use of the overall calorimeter jet energy scale factor as a fit variable when the W mass constraint is applied. It will be shown that adjusting the jet energy scale to get the best W mass in each event leads to a reduction in the jet systematic error. The price for this reduction is the increase in the statistical uncertainty which occurs because the fitted energy scale is allowed to fluctuate from event to event.

3.1 Kinematic Fitter

In our top quark mass calculation we consider only those parton assignments to the four leading jets which are consistent with the b tagging information. For each parton assignment we attempt to improve the resolution by integrating the top mass over the W mass Breit-Wigner distributions. For each of the two W bosons in the event we use a set of seven m_W^2 values equidistant in the cumulative Breit-Wigner probability. These values cover 96.7% of the cumulative probability range, corresponding to a range in mass-squared from approx. 0.5 to 1.5 times the square of the central m_W value. The central m_W is taken to be 80.423 GeV and the width 2.118 GeV.

For each jet to parton assignment and each mass assignment for the hadronically-decaying W boson we reconstruct the $W \to q\bar{q}$ decay using a least-squares kinematic fit. General kinematic fitting theory is discussed, for example, in Ref. [12]. Here, we describe the algorithm which includes the jet energy scale variable in the fit. In the formulae below, the following definitions are used: $m_{q\bar{q}}^2 = (P_q + P_{\bar{q}})^2$, $P_q \equiv (E_q, \mathbf{p}_q)$, $E_q = \sqrt{\mathbf{p}_q^2 + m_q^2}$. $\mathbf{p}_q$ is the 3-momentum of one of the light quarks. We use $m_q = 0.5$ GeV in all calculations. The masses of b quarks are set to 5 GeV. The quark 3-momenta are related to the corrected 3-momenta of the reconstructed jets, $\tilde{\mathbf{p}}_q$ and $\tilde{\mathbf{p}}_{\bar{q}}$, by

$$\mathbf{p}_q = s\tilde{\mathbf{p}}_q \text{ and } \mathbf{p}_{\bar{q}} = s\tilde{\mathbf{p}}_{\bar{q}} \tag{1}$$

where s is the jet energy scale variable. Our method works by iterative minimization of the constrained χ^2 defined by

$$\chi_{i+1}^2 = (\mathbf{y} - \mathbf{y}_i)^{\mathrm{T}} V^{-1} (\mathbf{y} - \mathbf{y}_i) + 2\lambda \mathrm{Lin}(m_{q\bar{q}i}^2 - m_W^2) \tag{2}$$

Table 6: Monte Carlo samples based on processing version 4.9.1 used for the signal templates, the background templates, and for various systematics studies. See `http://www-cdf.fnal.gov/internal/physics/top/mc/mc_samples_491.shtml` for further reference.

MC Type		Dataset Name
	Signal Samples	
HERWIG	M_{top}=130-155 GeV in 5.0 GeV steps	ttopgk—ttoplk
	M_{top}=160-200 GeV in 2.5 GeV steps	ttopmk—ttopxk, ttopli, ttop5l—ttop8l
	M_{top}=200-230 GeV in 5.0 GeV steps	ttopyk—ttopzk, ttop1l—ttop4l
	Background Samples	
HERWIG	W+4 light partons, $W \to e\nu/\mu\nu$	atop04, atop0a
HERWIG	W+$c\bar{c}$+2 light partons, $W \to e\nu/\mu\nu$	atop45, atop4b
HERWIG	W+$b\bar{b}$+2 light partons, $W \to e\nu/\mu\nu$	atop32, atop48
PYTHIA	single top (s channel, t channel)	ttop0s, ttop1s
	Samples Used for Systematic Studies	
PYTHIA	Tune A	ttopei
	ISR systematics:	
PYTHIA	Tune B	ttop5e
PYTHIA	less ISR ($\Lambda_{\mathrm{QCD}} = 100$ MeV)	ttopbe
PYTHIA	more ISR ($\Lambda_{\mathrm{QCD}} = 384$ MeV)	ttopce
PYTHIA	ISR off	ttop0e
	FSR systematics:	
PYTHIA	lower α_S ($\Lambda_{\mathrm{QCD}} = 190$ MeV)	ttop6e
PYTHIA	no hard gluon from b quark	ttopae
PYTHIA	less FSR	ttopde
PYTHIA	more FSR	ttopee
	Spin systematics:	
HERWIG	spin correlation off	ttopge
	PDF systematics:	
PYTHIA	MRST72 ($\Lambda_{\mathrm{QCD}} = 300$ MeV)	ttop3e
PYTHIA	MRST75 ($\Lambda_{\mathrm{QCD}} = 228$ MeV)	ttop4e
PYTHA	CTEQ6M Run I setting	ttop7e

where $\mathbf{y} \equiv (\tilde{\mathbf{p}}_q, \tilde{\mathbf{p}}_{\bar{q}}, s)$ is a vector to be determined in the 7-dimensional space of reconstructed jet momenta and jet energy scale. $\mathbf{y}_i$ is the solution from the previous iteration, and $\mathbf{y}_0 = (\tilde{\mathbf{p}}_{q\,\mathrm{measured}}, \tilde{\mathbf{p}}_{\bar{q}\,\mathrm{measured}}, 1)$. V is the jet measurement covariance matrix. $\mathrm{Lin}(m^2_{q\bar{q}i} - m^2_W)$ is the linearized version of the W mass constraint $m^2_{q\bar{q}} - m^2_W = 0$. λ is the Lagrange multiplier for the constraint. Factor 2 is included for convenience (obviously, the Lagrange multiplier can be rescaled by an arbitrary non-zero factor).

The covariance matrix V is derived assuming no correlations between jets apart from the common energy scale. Constant η and ϕ resolutions are used: $\sigma_\eta = \sigma_\phi = 0.05$. p_T jet resolution, σ_{p_T}, is set by the top-specific jet energy corrections. It is assumed to be independent from η and ϕ resolutions. With these assumptions, the measurement covariance matrix for each jet is $V_q = diag(\sigma^2_{p_T}, \sigma^2_\phi, \sigma^2_\eta)$ in the (p_T, ϕ, η) space. We propagate the errors to the Cartesian coordinates of the 3-momenta using the standard error propagation technique, with the Jacobian given by

$$J = \frac{\partial(p_x, p_y, p_z)}{\partial(p_T, \phi, \eta)} = \begin{pmatrix} \cos\phi & -p_y & 0 \\ \sin\phi & p_x & 0 \\ p_z/p_T & 0 & |\mathbf{p}| \end{pmatrix} \tag{3}$$

The covariance matrix is then block-diagonal:

$$V = \left(\begin{array}{c|c|c} J_q V_q J_q^{\mathrm{T}} & 0 & 0 \\ \hline 0 & J_{\bar{q}} V_{\bar{q}} J_{\bar{q}}^{\mathrm{T}} & 0 \\ \hline 0 & 0 & \sigma_s^2 \end{array}\right) \tag{4}$$

σ_s is the jet energy scale constraint parameter. It can be assigned the narrow meaning of the error on the jet energy scale measurement from calorimeter calibration. However, we choose to interpret this parameter more widely as the amount of prior knowledge about the jet energy scale in the top mass measurement (in the sense of Fisher information which is proportional to σ_s^{-2}). By assuming less prior knowledge about the jet energy scale, the dependence of the final result on the jet energy scale systematics can be reduced.

The linearized W mass constraint used in Eq. 2 is just

$$\mathrm{Lin}(m^2_{q\bar{q}i} - m^2_W) \equiv \left.\frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\right|_i (\mathbf{y} - \mathbf{y}_i) + m^2_{q\bar{q}\,i} - m^2_W \tag{5}$$

where the $m^2_{q\bar{q}}$ gradient is calculated at the solution obtained during the previous iteration. With the linearized constraint, solutions of the minimization equations

$$\frac{\partial \chi^2_{i+1}}{\partial \mathbf{y}} = \mathbf{0}, \quad \frac{\partial \chi^2_{i+1}}{\partial \lambda} = 0 \tag{6}$$

can be obtained by simple linear algebra manipulations. These solutions are given by

$$\lambda = \kappa_i\,(m^2_{q\bar{q}i} - m^2_W), \quad \mathbf{y}_{i+1} = \mathbf{y}_i - \lambda V \left.\frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\right|_i \tag{7}$$

where $\kappa_i = \left(\frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\bigg|_i^{\mathrm{T}} V \, \frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\bigg|_i \right)^{-1}$. Solutions are iterated until $|m_{q\bar{q}} - m_W| < 0.02$ GeV, so that the constrained χ^2 minimization is performed, essentially, by gradient descent. Covariance matrix V is kept constant throughout the iteration procedure. At the end of the fit, we determine the overall χ^2 of the W mass constraint as

$$\chi^2_W = \kappa_j \, (m^2_{q\bar{q}\,\mathrm{measured}} - m^2_W)^2. \tag{8}$$

where index j refers to the last iteration. The constrained covariance matrix used in the subsequent calculations is estimated as

$$V_c = V \left(I - \kappa_j \, \frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\bigg|_j \left(V \, \frac{\partial m^2_{q\bar{q}}}{\partial \mathbf{y}}\bigg|_j \right)^{\mathrm{T}} \right) \tag{9}$$

After the W mass fit is complete (for a given hadronically-decaying W mass assignment), momenta of all jets in the event (including non-leading jets) are multiplied by the fitted jet energy scale. The masses of the t and $\bar{t}$ quarks, m_t and $m_{\bar{t}}$, are then determined together with their errors, σ_{m_t} and $\sigma_{m_{\bar{t}}}$, and correlation coefficient, ρ. This procedure is carried out for each of the seven mass assignments for the leptonically-decaying W boson.

For leptons we use Run 1 resolutions. Error on the transverse momentum is set for electrons to $\sigma_{p_T}/p_T = \sqrt{0.02^2 + 0.135^2/p_T}$ and for muons to $\sigma_{p_T}/p_T = 0.0011 p_T$. For both lepton types angular errors are set to $\sigma_\phi = 6 \times 10^{-4}$, $\sigma_\lambda = 5 \times 10^{-3}$ (λ is the tangent of the track dip angle). The errors of the transverse unclustered energy components are set to $\sigma_{U_X} = \sigma_{U_Y} = \sqrt{(0.7\, U_T)^2 + 0.1^2}$.

Jet, lepton, and unclustered energy measurement errors are propagated to t and $\bar{t}$ masses using the standard small error approximation formulae. There is one exception: if the quadratic equation for the z component of the neutrino momentum has a small determinant, D, then the small error approximation breaks down because of the $1/\sqrt{D}$ term in the Jacobian (this term can be arbitrarily large). Therefore, for the error propagation purposes we replace $1/\sqrt{D}$ with $1/\sqrt{\max(D, \sigma_D)}$ in this Jacobian. Here σ_D is the error on D as calculated by standard error propagation. Out of the two neutrino solutions we choose the one which results in a smaller value of $|m_t - m_{\bar{t}}|$. W mass assignments without neutrino solutions are discarded. Finally, we impose the $m_t = m_{\bar{t}}$ constraint on the measurement and calculate the top quark mass $M_{t,W}$, its error $\sigma_{M_{t,W}}$, and chi-square for the top mass constraint $\chi^2_{t,W}$ as

$$M_{t,W} = \frac{\sigma^2_{m_{\bar{t}}} m_t + \sigma^2_{m_t} m_{\bar{t}} - \rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}} (m_t + m_{\bar{t}})}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{10}$$

$$\sigma^2_{M_{t,W}} = \frac{\sigma^2_{m_t} \sigma^2_{m_{\bar{t}}} \,(1 - \rho)}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{11}$$

$$\chi^2_{t,W} = \frac{(m_t - m_{\bar{t}})^2}{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}} - 2\,\rho\, \sigma_{m_t} \sigma_{m_{\bar{t}}}} \tag{12}$$

The top mass for each jet to parton assignment is determined by summing over the points in the W mass integration grid weighted by the combined χ^2 of the mass constraints:

$$M_t = \frac{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2} M_{t,W}}{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}} \tag{13}$$

We define the overall jet permutation chi-square as

$$\chi^2 = -2\log\left(\frac{1}{N}\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}\right) \tag{14}$$

where N is the total number of points in the integration grid (in the current analysis $N = 49$). The top mass error is approximated for each permutation by[1]

$$\sigma^2_{M_t} = \sigma^2_{\text{fit}} + \sigma^2_{\text{scatter}} \tag{15}$$

where

$$\sigma_{\text{fit}} = \frac{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2} \sigma_{M_{t,W}}}{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}} \tag{16}$$

and

$$\sigma^2_{\text{scatter}} = \frac{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2} (M_{t,W} - M_t)^2}{\sum e^{-(\chi^2_{t,W}+\chi^2_W)/2}} \tag{17}$$

The C++ code for the top mass kinematic fitter with the jet energy scale constraint can be obtained from the CDF Run 2 code repository. The W mass constraint, error propagation, and top mass constraint are performed by classes "WMassJesFit" and "TopMassJesFit" in the "Top" package. The calculations are organized and results are saved in an ntuple by module "TopMassJesModule" in the "TopMods" package.

3.2 On the Jet Energy Scale Constraint

In our method, the jet energy scale is adjusted separately for every jet permutation in every event. On average, for correct permutation this adjustment partially compensates for the scale shift induced by various sources of systematic errors described in Section 7.3.

Fluctuations in jet fragmentation are considerably larger than the suspected systematic shift in energy scale. Therefore, fluctuations in the $q\bar{q}$ system are imprinted on all other jets in the event through the scale determined in the W mass fit. This results in the increase of top mass statistical error. The amount by which the systematic error is reduced and statistical error is increased can be varied by changing the jet energy scale constraint. Therefore, we believe that our technique provides a valuable tool for balancing statistical and systematic errors. This balancing results in a reduction of the total error of the top mass measurement, and will be especially useful when CDF accumulates more luminosity.

[1] A better error estimate could have been obtained by analyzing correlations between $M_{t,W}$ values on the integration grid. Such an analysis, however, would be very expensive computationally. We assume that $M_{t,W}$ values are almost completely correlated.

The dependence of the total error on the jet energy scale constraint is discussed in detail in Section 7.

Use of the dijet mass in top events to fit the *global* jet energy scale has been attempted by others [13]. Determination of the global scale in top events is difficult because the dijet invariant mass spectrum includes contributions from background and jet permutation combinatorics. This technique will become more useful when CDF registers more top candidate events with two b tags.

3.3 Jet Corrections

We use the "level 7" generic jet corrections provided by the jet correction group [9]. Since these corrections are derived for a mixture of light quark and gluon jets, we apply additional, top-specific corrections on top of the generic corrections in order to take into account the differences between light jets and b jets and in order to assign uncertainties on the jet p_T measurements.

We do not apply additional corrections to the transverse momenta of light jets. For b jets, we multiply the jet momentum by the constant factor 0.981. This factor results in a slight improvement of the reconstructed top mass for the correct jet permutation in HERWIG MC events. Due to this correction, the slope of the median reconstructed top mass for the correct permutation (plotted versus the true top mass) changes from 1.012 to 1.002, and the median reconstructed top mass moves from 175.9 to 174.6 GeV for a true mass of 175 GeV.

We model jet p_T uncertainties for both light and b jets as $\sigma_{p_T} = p_{T\mathrm{jet}}(ae^{-b\,p_{T\mathrm{jet}}} + c)$. These uncertainties are independent from both ϕ and η. Coefficients a, b, and c are given in Table 7. These coefficients are derived by fitting jet samples from $t\bar{t}$ Monte-Carlo events. We combine MC samples in the top mass range from 140 to 210 GeV with 5 GeV step, and consider p_T distributions of reconstructed, level-7 corrected jets in 5 GeV wide bins of HEPG jet[2] p_T. We consider only those events for which we are able to establish an unambiguous match between HEPG jets and reconstructed jets. Reconstructed jets are required to be within the radius of 0.3 in the η-ϕ space from the HEPG jet. In addition, the overall χ^2 of the angular match between four leading jets in the reconstructed event and the daughter jets from the $t\bar{t}$ decay is required to be less than 100 ($\sigma_\eta = \sigma_\phi = 0.05$ is assumed). The mass of the HEPG b jet is required to be between 4.5 and 5.5 GeV. The mass of the HEPG light quark jet is required to be below 2 GeV. Note that the HEPG jet mass requirements effectively turn off the FSR. The effect of FSR on the jet energy is already compensated by the generic corrections (and the width of level-7 corrected p_T distribution is widened appropriately), therefore it is already partially included into the resolution. Additional FSR compensation is due to the W mass constraint with JES.

We characterize the jet resolution by the distance between 25$^{\mathrm{th}}$ and 75$^{\mathrm{th}}$ percentiles of the reconstructed jet p_T ratio distribution multiplied by 0.7413. The fits of this quantity are illustrated in Figs. 3.3 and 3.3 for light and b jets, respectively.

[2] HEPG jets are HERWIG objects with idhep = 94. These objects are used by HERWIG for simulation of the final state radiation.

Table 7: Constants used for modeling the jet p_T resolution

Quark type	a	b	c
b	0.231	0.0446	0.156
light	0.195	0.0463	0.115

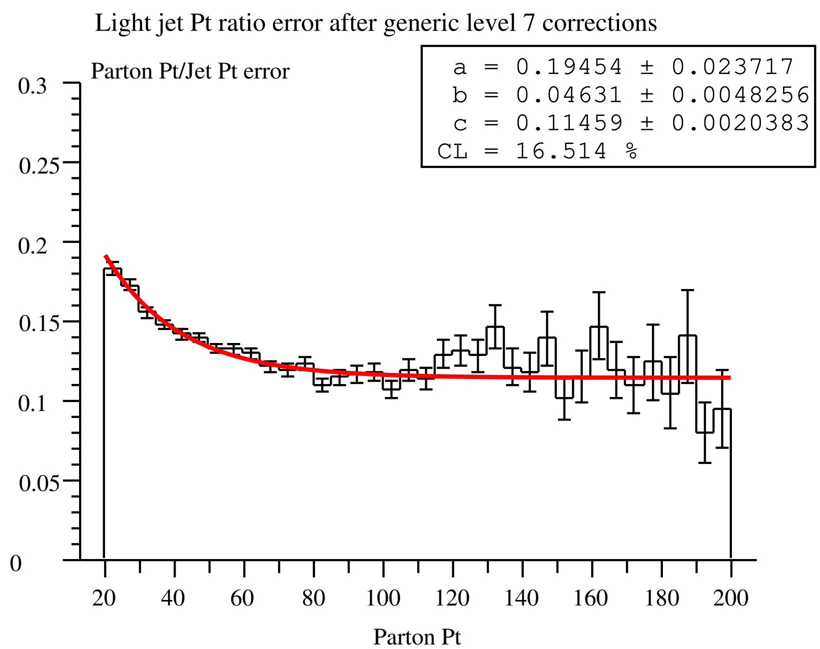


Figure 1: Relative p_T errors assigned to light jets

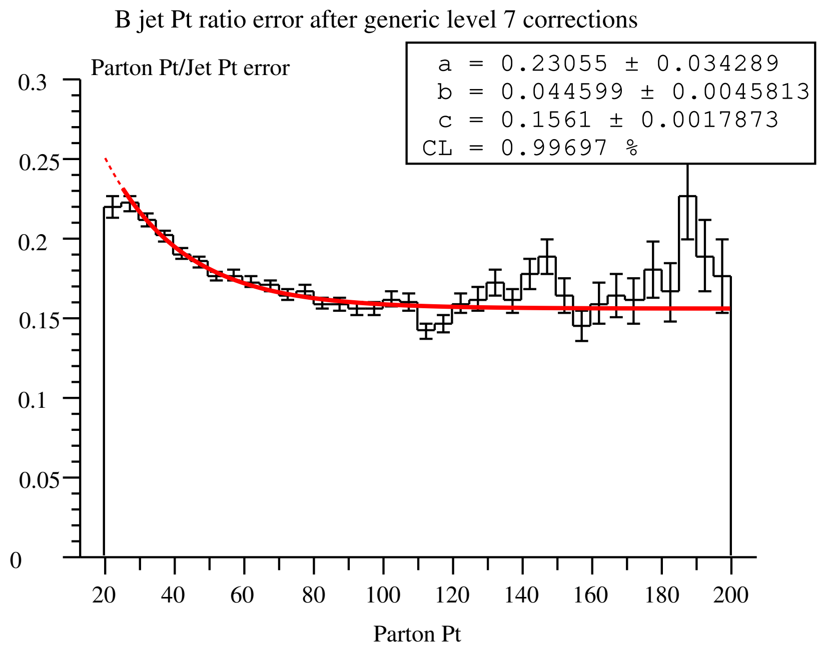


Figure 2: Relative p_T errors assigned to b jets

3.4 Reconstructed Top Mass

For each event passing the selection criteria described in Section 2.1 we apply the method described in Section 3 to reconstruct the top mass. In the following, we will refer to the top specific corrections discussed in Section 3.2 as "LBL" corrections. We have also tried the "JF" corrections [10] which were developed for the Summer 2003 measurement of the top mass with the Run I method.

From our fitter we take the solutions found for those jet to parton assignments ("permutations") which agree with the SVX tag constraint. For a given permutation we consider only the solution for the neutrino momentum for which the top mass from the leptonic side is closest to the top mass of the hadronic side. Among the six remaining permutations we pick the fitted value for the top mass with the smallest χ^2 solution, χ^2_{min}, which we call the reconstructed invariant top mass M_{top} of the event. In case of the LBL corrected events, we try various values for the jet energy scale (JES) constraint parameter ranging from 0.01 up to 0.30. For comparison, we fitted JF corrected events using a JES constraint value of 0.035.

A list of the candidate events with the reconstructed top quark masses obtained using the two top specific corrections and JES constraints between 0.035 and 0.07 is given in Table 8. Also the corresponding χ^2_{min} values and the fitted JES values are shown. In general, the fitter finds a solution fulfilling the demands described in Section 3. In four cases the fitter fails for certain JES constraints. In two events no solution was found for any JES constraint and correction type considered. We found that for a given JES

Table 8: List of candidate top events after standard selection cuts with the run and event numbers, lepton type l, jet multiplicity label n_{j}, and the values of the template variables (H_{T}, $p_{\mathrm{T}}^{(1-4)}$, M_{ij}^{min}, in GeV, and Aco) calculated at jet correction level 7. Also shown are the best χ^2 solutions for the fitted jet energy scale (JES) and the top mass M_{top} (in GeV) obtained from fits using JF and LBL corrections, resp., and for selected JES constraints (values in brackets). Events marked with "*" have two SVX tags among the four leading jets. "—" denotes that the fit has failed. The bottom of the list contains events failing the χ^2 cuts for all cases listed.

								JF (0.035)			LBL (0.035)			LBL (0.05)			LBL (0.07)		
Run	Event	l	N_{j}	H_{T}	$p_{\mathrm{T}}^{(1-4)}$	M_{ij}^{min}	Aco	χ^2	JES	M_{top}	χ^2	JES	M_{top}	χ^2	JES	M_{top}	χ^2	JES	M_{top}
141597	1353293	μ	3.5	272.5	201.6	34.6	0.176	5.62	0.98	136.7	8.80	0.98	133.3	8.80	0.96	132.3	8.61	0.92	130.7
144674	1782954	e	3.5	298.6	196.5	60.8	0.211	4.10	1.01	144.7	6.13	1.01	129.2	6.03	1.01	129.1	5.63	1.06	145.4
145036	245760	μ	4	380.9	274.2	58.9	0.107	3.31	1.02	176.1	4.70	1.02	179.4	4.09	1.04	181.1	3.36	1.08	183.6
148153	6088	e	3.5	345.7	204.4	23.3	0.082	0.16	1.00	148.0	1.38	1.00	145.7	1.35	0.99	145.7	1.28	0.99	145.8
148824	3002985	μ	3.5	158.2	100.8	25.2	0.147	1.90	1.01	146.8	3.27	1.01	123.4	3.07	1.03	123.9	2.79	1.05	124.7
149387	2551061	e	3.5	213.8	134.5	38.2	0.255	2.30	1.01	124.7	5.30	1.02	122.5	4.63	1.04	123.4	3.82	1.07	124.7
152266	3554	μ	4	276.4	203.0	49.0	0.354	1.33	0.99	148.0	2.18	0.99	155.5	2.08	0.97	154.5	1.90	0.96	153.0
153693	799494	μ	4	459.0	346.0	65.9	0.176	0.49	0.99	179.8	2.16	0.99	182.0	2.04	0.97	180.7	1.87	0.96	178.9
153694	1694029	e	4	389.1	330.3	47.0	0.100	3.48	1.01	154.7	3.25	1.01	153.6	3.08	1.03	154.2	2.86	1.04	155.1
153738	2083102	e	4	426.9	327.9	79.0	0.118	2.21	1.02	191.8	1.12	1.01	176.6	1.09	1.01	177.0	1.16	1.02	177.5
154175	1630925	e	4	475.6	250.3	11.6	0.075	0.76	0.99	139.6	2.80	0.99	164.7	2.67	0.97	163.6	2.48	0.95	161.6
155145	132579	e	3.5*	370.1	261.5	63.3	0.312	0.22	1.00	190.0	0.20	1.00	177.6	0.19	0.99	177.3	0.17	0.99	176.9
155345	3194866	e	3.5	282.4	164.0	30.9	0.002	4.66	1.01	168.5	9.75	0.97	122.8	8.84	1.05	170.8	7.35	1.10	174.2
160153	1270879	e	4	453.7	256.1	66.5	0.137	0.72	1.00	168.7	1.73	0.99	170.8	1.65	0.98	169.9	1.50	0.96	168.4
160437	280173	μ	4	346.0	271.0	44.0	0.095	0.94	0.99	175.1	1.35	0.99	186.4	1.31	0.98	185.2	1.21	0.97	183.9
160591	894406	μ	4	683.3	391.1	62.4	0.019	10.90	0.99	226.9	11.34	0.98	230.0	10.70	0.97	226.7	9.69	0.95	221.9
161013	111162	e	4	576.6	321.5	21.3	0.373	11.10	1.03	216.8	10.13	1.03	222.1	8.62	1.06	226.2	6.93	1.10	232.7
161379	494836	e	4	509.8	253.8	39.4	0.188	2.65	1.00	159.4	2.45	0.99	151.0	2.40	0.99	150.4	2.33	0.98	149.5
161633	1571961	e	4	359.6	259.3	24.2	0.149	7.47	0.98	200.3	7.61	0.98	127.7	7.58	0.96	126.5	7.29	0.92	124.7
161788	361577	μ	4	588.5	444.5	86.7	0.268	0.61	1.01	176.6	0.28	1.00	182.1	0.28	0.99	181.8	0.26	0.99	181.4
162423	261933	e	4	253.3	167.8	34.2	0.071	0.35	1.01	135.4	1.94	1.00	125.4	1.92	1.00	125.6	1.88	1.01	125.8
162519	931923	μ	3.5	235.4	147.1	33.3	0.155	1.86	1.01	135.7	6.57	1.02	124.8	6.02	1.03	125.9	5.32	1.06	127.4
162631	7109631	μ	4	358.6	203.2	37.7	0.041	1.34	1.00	132.1	4.41	1.00	132.7	4.38	1.00	132.7	4.35	1.01	132.6
163012	2249546	μ	4	472.7	281.3	26.9	0.191	3.70	1.02	204.7	1.95	1.01	196.1	1.78	1.03	197.5	1.56	1.05	199.6
164274	2932602	μ	3.5	273.2	153.3	22.4	0.311	2.36	0.99	152.5	4.80	0.98	146.2	4.91	0.96	145.1	5.27	0.94	144.1
164819	2297394	e	4	261.9	160.1	32.8	0.287	1.25	0.99	156.5	2.01	0.99	158.9	1.93	0.98	158.0	1.77	0.96	156.5
164844	2309600	μ	4	521.1	290.5	29.6	0.188	1.38	1.01	182.2	0.23	1.00	189.2	0.20	1.01	189.6	0.17	1.01	190.2
165314	1155563	μ	3.5	226.8	154.6	30.7	0.260	1.68	0.99	134.8	4.19	0.98	124.3	4.06	0.97	123.8	3.74	0.94	123.1
166007	498553	e	4	415.0	272.0	39.5	0.170	4.92	0.98	267.9	8.75	0.98	272.6	8.78	0.95	268.3	8.59	0.92	261.4
166567	11615607	μ	3.5*	438.2	316.0	70.0	0.079	10.56	1.02	237.8	9.68	1.02	244.0	9.47	1.03	245.5	9.22	1.05	247.6
166614	804529	e	4	488.1	301.4	86.8	0.132	0.55	1.01	191.0	0.71	1.00	181.6	0.70	1.00	181.2	0.68	0.99	180.7
166805	2534588	μ	4	418.1	286.0	50.0	0.097	6.34	1.00	157.2	5.69	1.02	158.0	5.39	1.03	158.7	5.05	1.05	159.7
167139	1191211	μ	4*	351.4	230.1	60.8	0.196	8.59	1.02	172.8	9.81	1.03	163.6	8.91	1.05	165.4	7.90	1.09	168.0
155320	480816	e	3.5	329.9	175.6	34.8	0.138		—		25.44	1.03	147.5	20.84	1.07	150.2	16.03	1.14	154.7
156116	6116596	e	4*	327.5	264.6	28.5	0.170		—			—			—			—	
162986	1538897	e	4*	543.2	327.5	45.4	0.129	138.06	0.92	119.5	180.30	0.92	209.4	199.28	0.84	199.6	210.41	0.71	184.4
165313	1770456	μ	4	396.8	204.5	29.7	0.100	15.24	1.02	139.2	21.27	1.03	148.1	18.57	1.06	149.8	14.56	1.13	157.7
166717	3530653	e	4	283.3	199.6	59.5	0.099		—		10.35	0.98	158.0	11.39	0.96	156.0		—	
167551	7969376	e	4*	469.5	250.9	40.9	0.251		—			—			—			—	
167629	400451	e	4	369.4	264.5	57.1	0.226	12.03	0.97	211.1	23.46	0.96	207.9	23.66	0.93	204.0	22.99	0.88	157.4

constraint value, the $\chi^2_{\rm min}$ from the fits based on JF corrections are generally smaller than those based on LBL corrections. Furthermore, looser JES constraint yield better $\chi^2_{\rm min}$.

For the top mass analysis we consider events with $\chi^2_{\rm min} < 10$. The number of remaining events based on LBL corrections and a JES constraint value of 0.035 are listed in Table 2 for various event categories. Note that the numbers slightly change with releasing JES constraint.

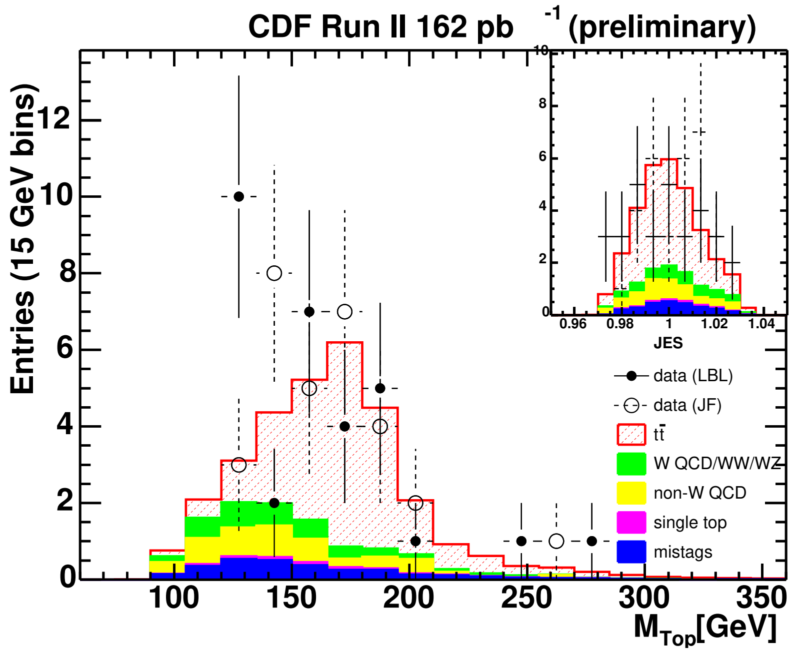
Figure 3 shows the reconstructed invariant top mass distribution extracted from the data with both the LBL top specific corrections (solid points) and JF corrections (empty points) for JES constraint values 0.035 and 0.07. Also the distributions of the fitted JES values are shown in the inset histogram of the same plots.

In addition, the data are compared with the sum of the Monte Carlo signal (for a generated top quark mass of 175 GeV) and the various background contributions discussed in Section 3, in order to demonstrate the performance of the MC samples used in the present analysis. The signal MC is normalized to the LBL corrected data, whereas the background normalization is based on the constraints taken from [4] using the fractions for the samples having a mass fit according to Table 5. Despite of the modest data statistics, the basic features of the shape of the measured distributions are well reproduced by the MC prediction, supporting a maximum at $M_{\rm top}$ ~170-180 GeV mainly caused by signal events, with a steeply falling edge at higher $M_{\rm top}$ values, and a flank at decreasing top mass values falling off more slowly due to the presence of background events at lower values for $M_{\rm top}$.

4 Templates: KDE method

Better discrimination between signal and background and, perhaps, improvement in the top mass sensitivity can be achieved by incorporating information from other kinematical variables into the top mass likelihood. Use of histograms for building templates becomes impractical in more than one dimension because the number of events needed to populate a histogram rises very quickly as more variables are added. It is also desirable to try a variety of variables and their combinations in order to optimize the expected top mass error. Therefore, it appears that one of the nonparametric multivariate density reconstruction methods should be used to prepare the templates from a set of Monte Carlo points. Although a wide variety of such techniques is available (neural nets, support vector machines, wavelet thresholding, *etc.*) we have chosen to use the kernel density estimation (KDE) method [22] due to its maturity, conceptual simplicity, and straightforward software implementation.

In the KDE method, the density at any given point in the multivariate space receives contributions from Monte Carlo events in such a way that events are weighted more heavily when they appear closer to the desired point. The distances between points are determined by the metric which is usually chosen as the inverse of the sample covariance matrix. We use radially symmetric kernels for which the density estimation formula can



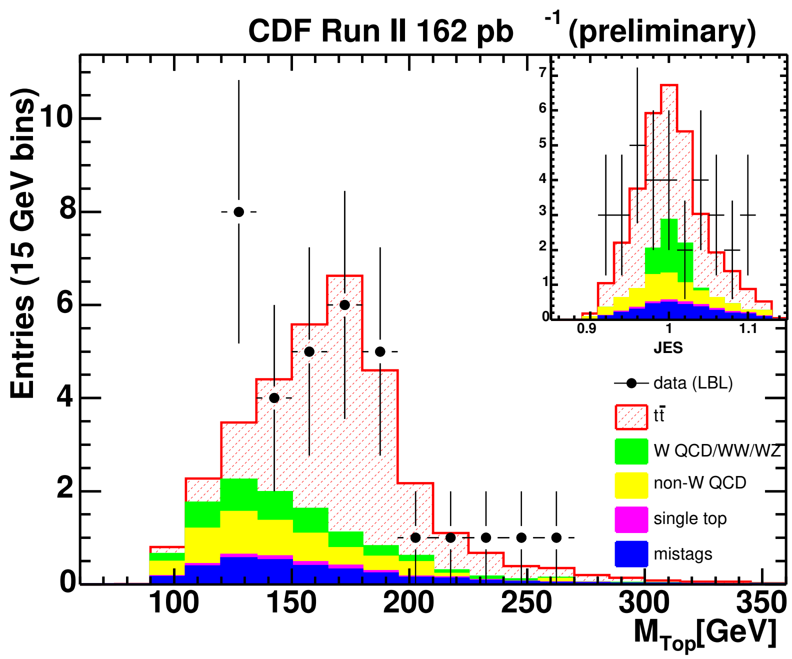


Figure 3: Top mass distribution extracted from the data with LBL and JF corrections, respectively, using JES constraint values of 0.035 (top, 31 events) and 0.07 (bottom, 33 events). Overlaid are the "stacked" MC expectations for the signal with $M_{\rm top} = 175\,\rm GeV$ as well as the various background contributions using the normalization constraints from the cross section analysis [4] and the fitted total background fraction (Section 8.1). The inset histograms show the distributions of the fitted JES values.

be written as

$$\hat{f}(\mathbf{y}) = \frac{1}{Nh^d\sqrt{|\det V|}} \sum_{i=1}^{N} K\left(\frac{(\mathbf{y}-\mathbf{x}_i)^T V^{-1}(\mathbf{y}-\mathbf{x}_i)}{h^2}\right) \tag{18}$$

Here, $\mathbf{y}$ is the point in a d-dimensional space for which we're estimating probability density, the $\mathbf{x}$'s are the N training points which provide us with event density information, V is a d x d matrix whose inverse is used as metric, $K(...)$ is known as a kernel function, and h is the scale parameter (often called "bandwidth"). The integral of the kernel function over the whole space should be 1:

$$\frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int K(r^2) r^{d-1} dr = 1 \tag{19}$$

We use kernel function which we named "Huber kernel"[3]. This function is defined by

$$K_H(r^2) = Norm(a,d) \times \begin{cases} e^{-r^2/2} & \text{if } r \le a \\ e^{-ra+a^2/2} & \text{if } r > a \end{cases} \tag{20}$$

where $Norm(a,d)$ is the normalization term which depends on parameter a and dimensionality of the variable space. Huber kernel behaves like a Gaussian probability density near the origin but decays slower at infinity. In this manner we could ensure that outlying points (typically from background) would contribute to the likelihood value but not have such low density that they would unduly affect the likelihood peak position. The value of parameter a was always chosen so that the Gaussian part of the distribution contains 90% of the density.

In order to reduce the effect of outliers on the metric, we calculate the sample covariance matrix using the robust minimum covariance determinant (MCD) algorithm [18]. The MCD breakdown point was set at 25%.

To increase computing speed, all training and data points are first transformed to the coordinate system in which the robust covariance matrix is unity using transformation matrix $Q = V^{-1/2}$. At the end, densities are transformed back to the original space.

We use the "plug-in" global bandwidth which minimizes asymptotic mean squared integrated error (AMISE) of the density estimator for a sample of points drawn from the multivariate Gaussian distribution:

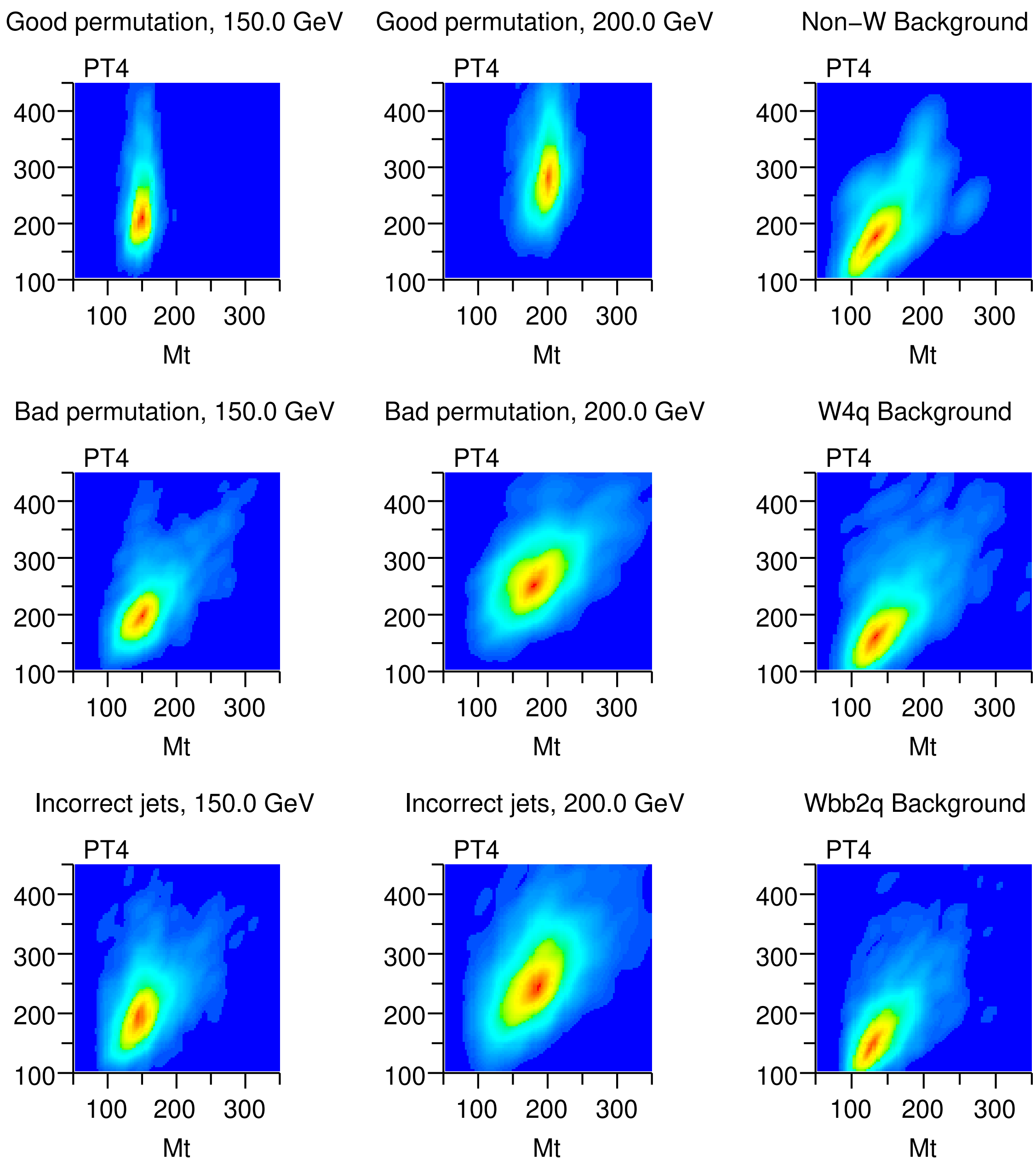
$$h = R_c \left(\frac{4}{d+2}\right)^{1/(d+4)} N^{-1/(d+4)} \tag{21}$$

where R_c is the canonical bandwidth ratio [24] between Gaussian and Huber kernels.

Example plots of 2-dimensional templates obtained using the reconstructed top mass and the sum of the four leading jet p_T values are shown in Figure 4.

[3]To recognize statistician Peter Huber who used this function in his classic paper about M-estimates [23].

Figure 4: Two-dimensional templates obtained using the reconstructed mass and the scalar sum of the four leading jet p_T values. The color maps are not synchronized between the plots.



5 Extracting a top mass from the data

We have discussed a method for performing a kinematic fit on events with one lepton and 4 jets. This topology allows 12 possible ways to obtain a $t\bar{t}$ final state (actually 24 if both neutrino solutions are considered). The number of possible jet permutations is reduced to six if one jet is tagged as a b and it is assigned a b only identity. It is further reduced to two combinations if both b jets are tagged. Clearly the mass measurement resolution depends on which permutation has the best χ^2. The correct combination is expected to have a much better resolution.

For each event we compute the probability that solution with the smallest χ^2 is the correct one, and weight the event accordingly. The weighting is performed by using different template fractions for the different types of permutations.

In addition, to distinguish top signal from background we use more than just the reconstructed mass variable. We use also templates for kinematic and angle variables. We distinguish between variables that are mass dependent and variables that are not mass dependent.

5.1 Different types of signal templates

We split our signal MC events into three subsamples:

1. Events in which the four leading jets correspond to the four daughter quarks in the $t\bar{t} \to l\nu bq\bar{q}\bar{b}$ process, and the jet permutation with the smallest χ^2 results in the correct jet to parton assignment. We call this GP (good permutation) subsample.

2. Events in which the four leading jets correspond to the four daughter quarks in $t\bar{t} \to l\nu bq\bar{q}\bar{b}$, but the smallest χ^2 permutation does not result in the correct jet to parton assignment. This is the BP (bad permutation) subsample. Later we will refer to the union of GP and BP subsamples as the CJ (correct jets) subsample.

3. Events in which one or more leading jets do not come from $t\bar{t} \to l\nu bq\bar{q}\bar{b}$. This subsample includes all $t\bar{t}$ MC events which do not belong to the first two subsamples. This is the IJ (incorrect jets) subsample.

The reconstructed mass distribution for these subsamples is illustrated in Figure 5 for 175 GeV generated mass and jet energy scale constraint value 0.035. The dependence of the median and width of these distributions on the generated mass is shown in Figures 6 and 7.

Knowledge of the subsample to which the data event belongs would clearly result in the improvement of the mass resolution. The importance of this knowledge can be illustrated by the following observation. Let's consider a very simple model in which there is no background, and the reconstructed mass distribution in each subsample is a Gaussian. Let's also assume that the the dependence of the mean of each Gaussian on the top mass is linear, and the standard deviation does not depend on the top mass. In this model

Figure 5: Reconstructed top mass distribution for the three signal subsamples, for the generated mass of 175 GeV

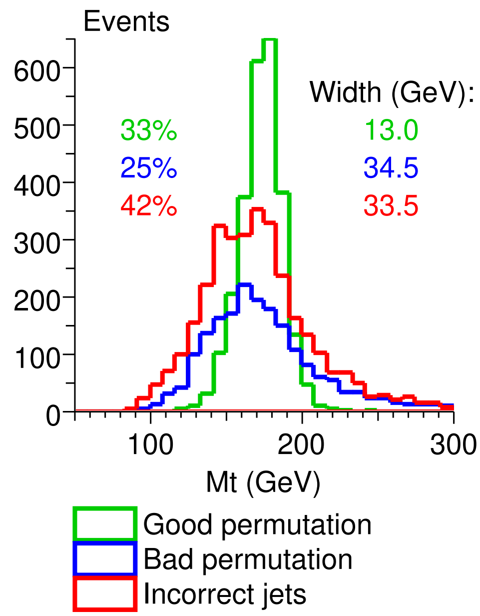


Figure 6: Dependence of the median of the reconstructed top mass distribution on the generated mass for the three signal subsamples. The slope of the GP subsample curve is 1.00. The average slopes for the BP and IJ subsamples are 0.67 and 0.68, respectively.

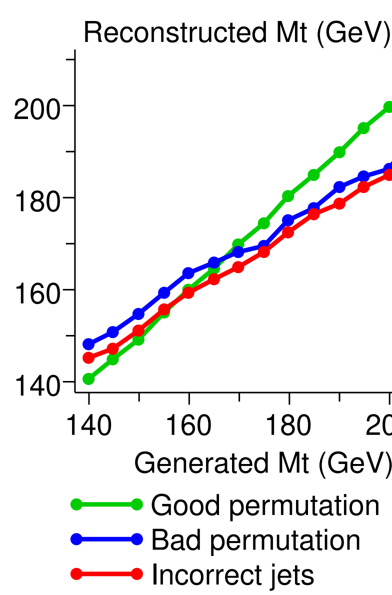
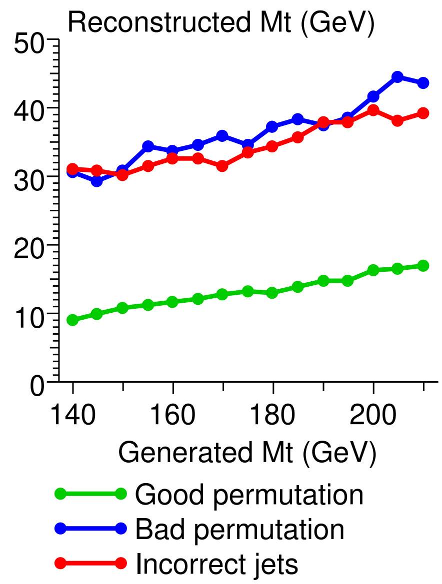


Figure 7: Dependence of the width of the reconstructed top mass distribution on the generated mass for the three signal subsamples. In this plot, width is defined as half of the distance between 16^{th} and 84^{th} percentiles of the distribution.



the Fisher information about the top mass, $F(m_t)$, can be easily evaluated. In case of a perfectly correct assignment of each $t\bar{t}$ event to one of the three subsamples

$$F_1(m_t) = N \sum_{X=\text{GP,BP,IJ}} \frac{f_X k_X^2}{\sigma_X^2},$$

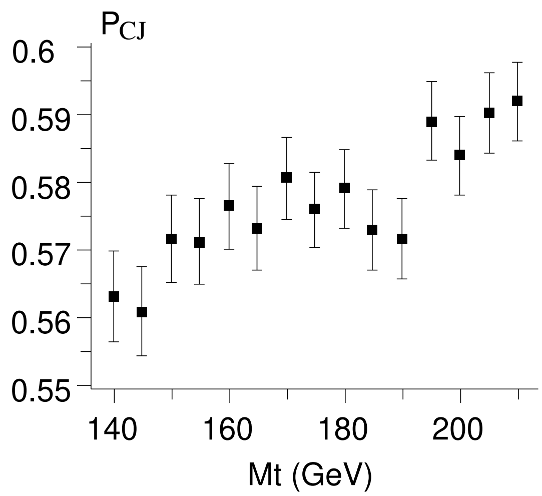
where N is the total number of events, f_X is the fraction of events in the subsample X, k_X is the slope of the mean, and σ_X is the width. In case we can not distinguish between the subsamples at all

$$F_2(m_t) \sim N \frac{k^2_{combined}}{\sigma^2_{combined}} \leq N \frac{\left(\sum f_X k_X\right)^2}{\sum f_X \sigma_X^2}.$$

Using subsample widths, fractions, and slopes as in Figures 5, 6, and 7, $F_1(m_t)/F_2(m_t) > 3.0$. Therefore, in this model perfect subsample assignment improves the top mass resolution by factor of ~ 1.7.

Of course, perfect $t\bar{t}$ subsample separation can not be achieved in practice. Instead, to each data event we assign probabilities to belong to each subsample. These probabilities are calculated using variables which are independent of or only weekly dependent on the top mass, such as χ^2 of the jet permutations, number of b tags in the event, *etc.* At the end, this strategy allows us to use a likelihood formula with a simple normalization condition.

Figure 8: Dependence of the probability to find four leading jets correctly on the generated top mass in the HERWIG $t\bar{t}$ MC samples. At least one b tag is required.



In the current implementation of our analysis, we do not attempt to improve the separation between IJ and the other subsamples[4]. The probability to select correct jets is fixed at its value expected from the signal MC. The dependence of this probability on top mass is weak, as illustrated in Figure 8.

The simple idea of predicting the probability of GP subsample using just the χ^2 of the best permutation does not work. This is illustrated in Figure 9 in which the best permutation χ^2 is plotted for the three signal subsamples (a more detailed study can be found in Ref. [14]). Instead, we employ a more complicated model in which we consider differences between χ^2 values of the best permutation and all other permutations, as well as kinematic properties of $t\bar{t}$ events. This model is inspired by viewing permutation selection as a "diffusion" process in the space of χ^2 values in which events can "drift" from the region in which best χ^2 permutation selects correct jet to parton assignment into the regions where this is no longer true. This "diffusion" is illustrated in Figures 10 and 11. The fraction of incorrectly assigned permutations appears to decrease exponentially as a function of the difference between permutation χ^2 values. Therefore, we make the initial GP probability assignment assuming that the jets were correctly chosen as

$$P(GP|CJ) = \frac{a_b}{\sum_{i=1}^{12} a_i \exp\left(-\frac{\chi_i^2 - \chi_b^2}{w_e(\chi_i^2 + \chi_b^2)}\right)}, \text{ where} \tag{22}$$

$$w_e(y) = \exp(b_e + c_e y + d_e y^2) \tag{23}$$

[4]We intend to improve this separation in the future by studying differences between shapes of gluon and quark jets.

Figure 9: χ^2 of the signal subsamples for 175 GeV generated mass. The plots are normalized to the same number of events in the range shown.

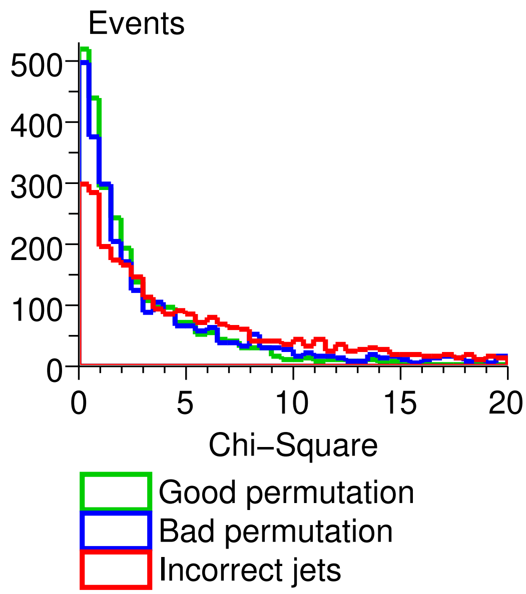


Table 9: Permutation numbers are defined in terms of parton assignments to the four leading jets arranged in the order of decreasing p_T.

Permutation #	Parton sequence
0	$b_h, b_l, q, \bar{q}$
1	$b_l, b_h, q, \bar{q}$
2	$b_h, q, b_l, \bar{q}$
3	$b_l, q, b_h, \bar{q}$
4	$b_h, q, \bar{q}, b_l$
5	$b_l, q, \bar{q}, b_h$
6	$q, b_h, b_l, \bar{q}$
7	$q, b_l, b_h, \bar{q}$
8	$q, b_h, \bar{q}, b_l$
9	$q, b_l, \bar{q}, b_h$
10	$q, \bar{q}, b_h, b_l$
11	$q, \bar{q}, b_l, b_h$

Figure 10: This plot shows events from 177.5 GeV $t\bar{t}$ MC with one b tag in which both the jet permutation with the smallest χ^2 and the correct jet permutation have numbers 0 or 1. Jet permutation numbering follows the convention from Table 9. Events in which permutation 0 is correct are shown with blue dots, and events in which permutation 1 is correct are shown with red dots. If the permutation with the smallest χ^2 was always correct, there would be no red points above the $\chi^2_0 = \chi^2_1$ boundary, and no blue points below. Instead, we observe diffusion-like mixing of the red and blue points near the boundary.

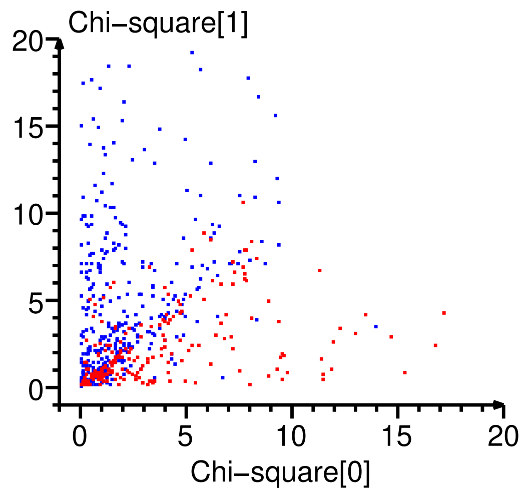


Figure 11: Fraction of the events with correct permutation 0 (blue) and correct permutation 1 (red) as a function of $\chi^2_1 - \chi^2_0$. This is a projection of Figure 10 onto the diagonal which runs from the lower right corner to the upper left, in which the total number of events in each bin is normalized to 1.

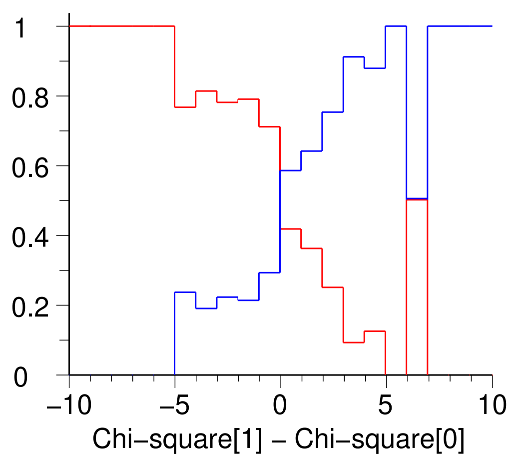
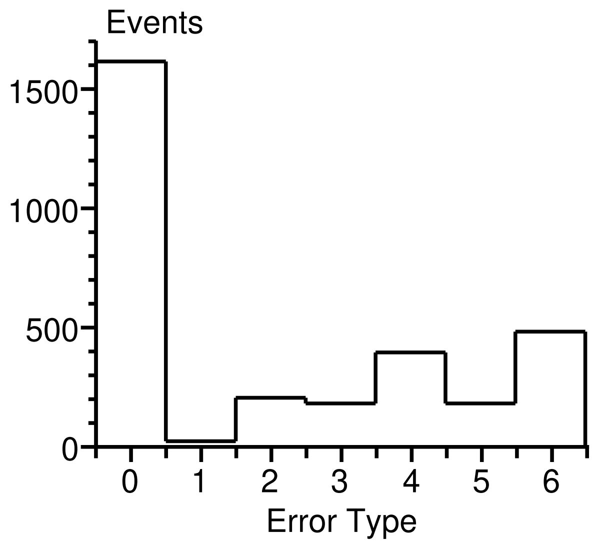


Figure 12: Frequency of jet permutation errors in the CJ subsample with one b tag in the 177.5 GeV $t\bar{t}$ MC events. The error types are listed in Table 10. Error type 1 is highly suppressed because this error can only happen when a light jet is misreconstructed as a tagged b jet.



Coefficient a_i models the event density in the correct region for jet permutation number i. Index b denotes the number of the permutation with smallest χ^2. The width of the transition regions is modelled by the function $w_e(y)$. Since there are 12 possible jet permutations[5], there are 66 pairwise transition regions. We assume that the "transparency" of each transition region is the same in both direction, and the width depends only on the type of error made when the incorrect permutation is chosen. Therefore, the index e in the above formula refers to the error type and not the transition region number. The error types are listed in Table 10, and frequencies of various errors observed in the $t\bar{t}$ signal MC are shown in Figures 12 and 13.

We also include additional kinematical information with weak dependence on m_t into the $P(GP|CJ)$ definition. This information is included by sequential application of the "approximate Bayesian update" formula. Suppose, X is some kinematic variable whose value may be used to improve separation between the GP and BP subsamples. Then formula

$$P(GP|CJ, X) = \frac{\kappa P(GP|CJ)}{\kappa P(GP|CJ) + (1 - P(GP|CJ))} \tag{24}$$

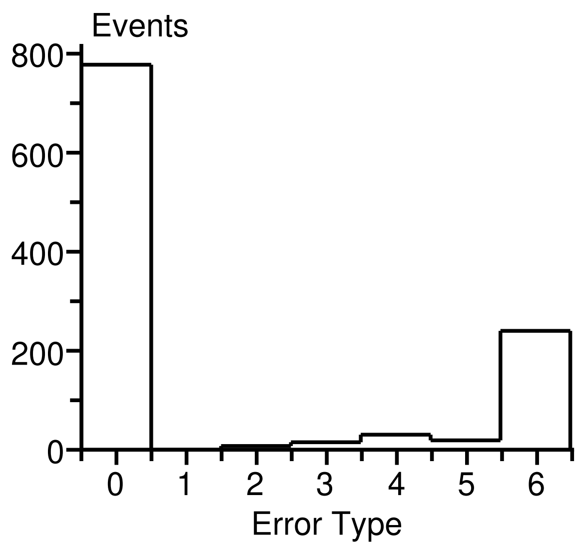
follows from the Bayes theorem when $\kappa = \frac{P(X|GP)}{P(X|BP)}$ and $P(GP|CJ) + P(BP|CJ) = 1$. Rather than calculating κ from MC, we treat it as a decision surface to be optimized together with the constants in the initial $P(GP|CJ)$ assignment.

[5]In the most general case of no b tags

Table 10: Error type numbering scheme. Suppose, sequence $q, \bar{q}, b_l, b_h$ represents the correct assignment of partons to the four leading jets arranged in the order of decreasing p_T. Here, b_l is the b quark which makes a top candidate together with the lepton and the neutrino, and b_h is the other b quark. We do not attempt to distinguish q and $\bar{q}$ in our analysis, therefore we make error type assignments as shown in the right table column. Only permutations in which q precedes $\bar{q}$ are listed. Swap $q \leftrightarrow \bar{q}$ in either correct or incorrect permutation does not change the error type of the incorrect permutation. The meaning of error types does not depend on which parton sequence is correct. Thus, error type 1 always means that the incorrect permutation has both of its b jets swapped with both light jets of the correct permutation, error type 2 means that b_h is picked instead of b_l and one of the light quarks is picked instead of b_h, *etc.*

Parton sequence	Error type
$b_h, b_l, q, \bar{q}$	1
$b_l, b_h, q, \bar{q}$	1
$b_h, q, b_l, \bar{q}$	4
$b_l, q, b_h, \bar{q}$	2
$b_h, q, \bar{q}, b_l$	5
$b_l, q, \bar{q}, b_h$	3
$q, b_h, b_l, \bar{q}$	4
$q, b_l, b_h, \bar{q}$	2
$q, b_h, \bar{q}, b_l$	5
$q, b_l, \bar{q}, b_h$	3
$q, \bar{q}, b_h, b_l$	6
$q, \bar{q}, b_l, b_h$	0

Figure 13: Frequency of jet permutation errors in the CJ subsample with two b tags in the 177.5 GeV $t\bar{t}$ MC events.



The first kinematic variable used to update $P(GP|CJ)$ is $\cos(\varphi)$ where φ is the angle between l and b_l in the rest frame of the W which decays into $l\nu$. Distribution of this angle is shown in Figures 14 and 15 for a set of generated m_t values. We hope to improve the GP/BP subsample separation near $\cos(\varphi) = -1$. The correct permutation is suppressed in this region by the top decay matrix element. We model the decision surface for this variable as

$$\kappa_\varphi = A\left(1 - e^{-\frac{\cos(\varphi)-B}{C}}\right) \tag{25}$$

where A, B, and C are constants to be fitted.

The second kinematic variable used to update $P(GP|CJ)$ is the product $\cos(\theta_1)\cos(\theta_2)$. θ_1 is the angle between the lepton momentum and the beam axis in the rest frame of the top quark which decays semileptonically. θ_2 is the angle between the direction of the light quark and the beam axis in the rest frame of the top quark which decays hadronically. We use that light quark which, in the rest frame of the parent W, has smaller angle with the b quark originating from the hadronically decaying top. The $\cos(\theta_1)\cos(\theta_2)$ product is sensitive to the $t\bar{t}$ spin correlations [15], and its top mass dependence is mild. The decision surface for this variable is modeled as

$$\kappa_{\text{spincorr}} = D\,(1 + E\cos(\theta_1)\cos(\theta_2)) \tag{26}$$

where D and E are constants to be fitted[6].

[6]We have also tried to use the off-diagonal axis [16] instead of the beam axis as a top quark spin quantization basis but did not find any sizable increase in the magnitude of fitted E value.

Figure 14: Distributions of the angle between l and b_l in the rest frame of the W which decays into $l\nu$ for the GP subsample with one b tag. The function overlayed on top of each plot has the same shape for each generated mass, only its height is adjusted.

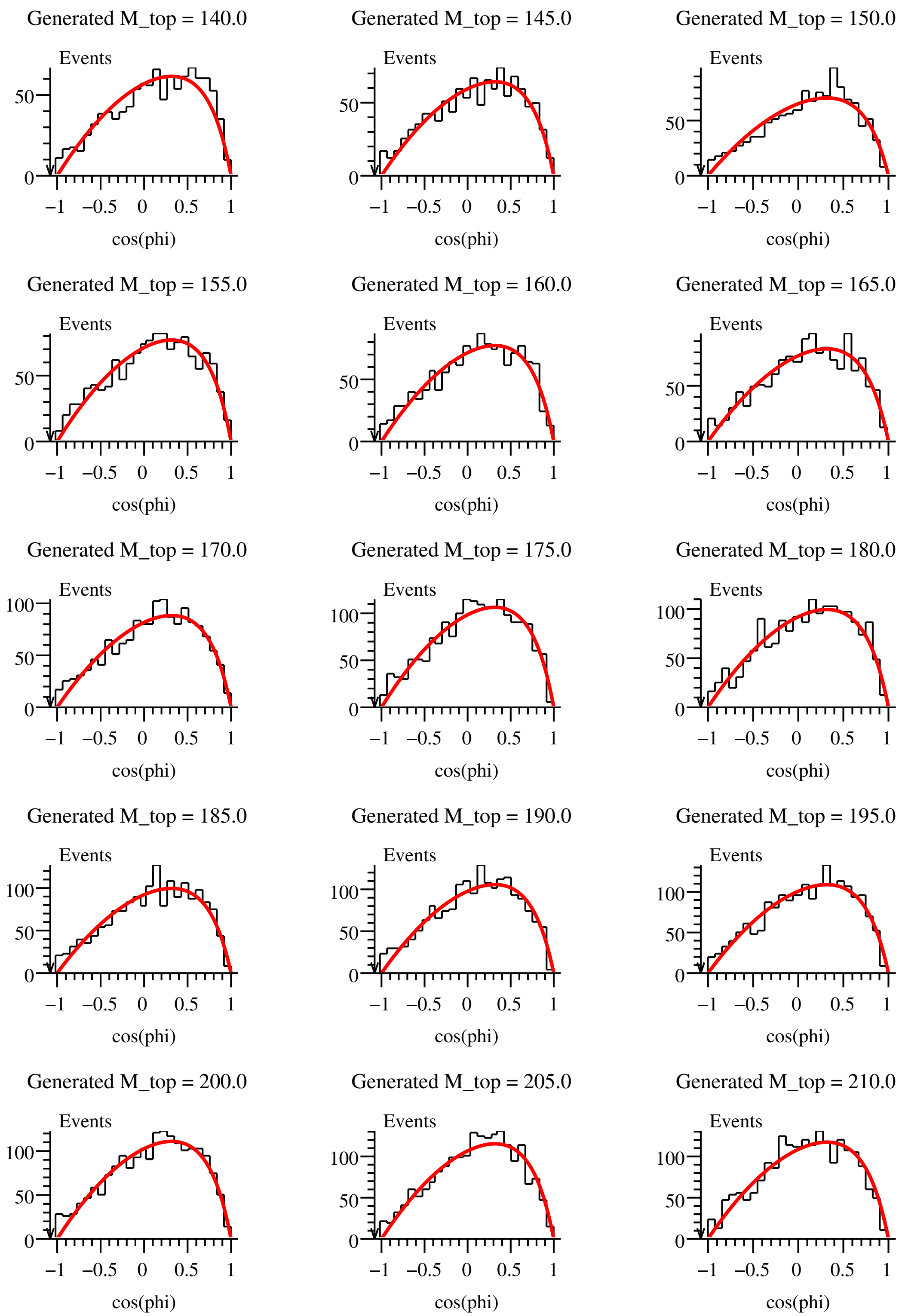


Figure 15: Distributions of the angle between l and b_l in the rest frame of the W which decays into $l\nu$ for the BP subsample with one b tag. The function overlayed on top of each plot has the same shape for each generated mass, only its height is adjusted.

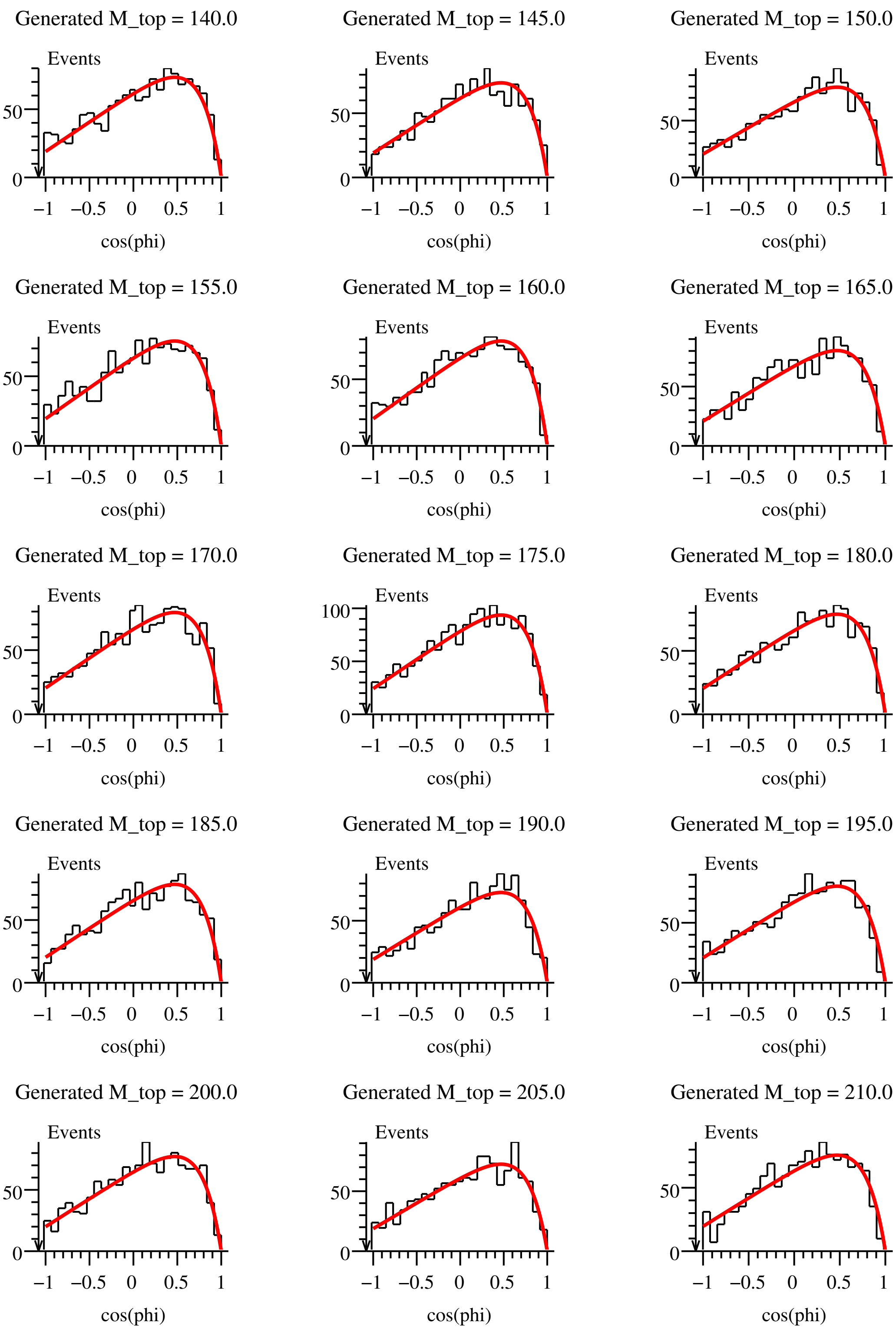
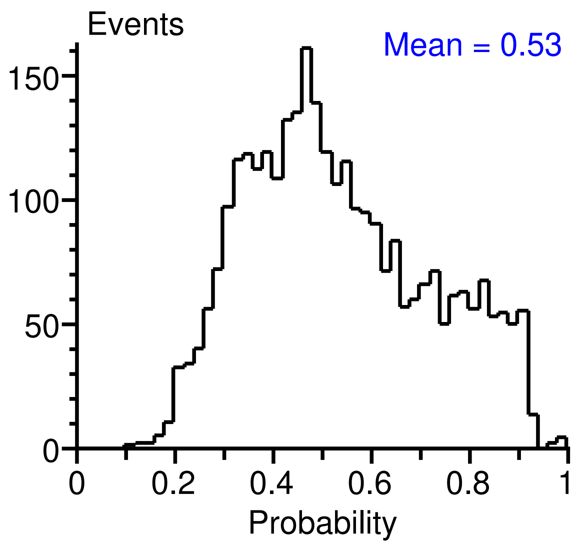


Figure 16: Distribution of the $P(GP|CJ, X)$ probabilities in the CJ subsample with one b tag in the 177.5 GeV $t\bar{t}$ MC events. The fraction of GP events in this subsample was found by counting to be 52%.



Coefficients a_i, b_e, c_e, d_e from Eqs. 22, 23 and decision surface constants A, B, C, D, and E from Eqs. 25, 26 are found by maximum likelihood method. The fit is performed separately for events with one and two b tags. Each MC signal event from the GP subsample is assigned probability $P(GP|CJ, X)$, and each event from the BP subsample is assigned probability $1 - P(GP|CJ, X)$. Then the likelihood of the whole CJ subsample is maximized. For this optimization, we have mixed together signal MC events generated in the m_t range from 140 to 210 GeV, with 5 GeV steps. Equal number of events was used for each generated mass.

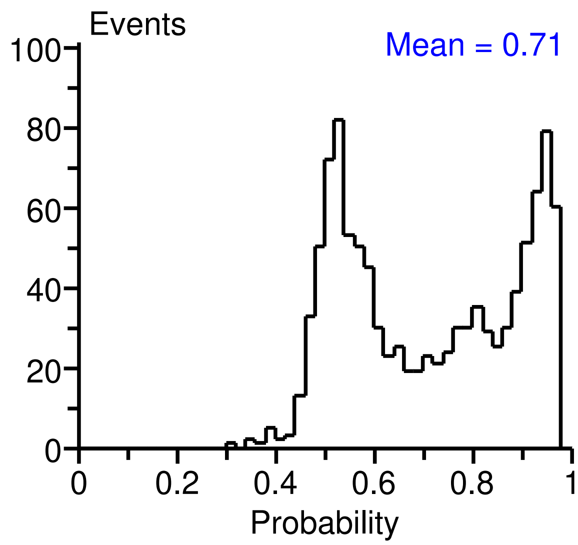
The distribution of $P(GP|CJ, X)$ in the 177.5 GeV $t\bar{t}$ MC events is shown in Figures 16 and 17 for the case of one and two b tags, respectively. Note that the 177.5 GeV mass sample was not used in tuning of the model coefficients. We check the quality of $P(GP|CJ, X)$ modeling by combining CJ subsample events in bins of $P(GP|CJ, X)$ and in each bin comparing the average $P(GP|CJ, X)$ value with the fraction of the events that are GP events. If the model is correct, the fraction found in a sample of N events should be distributed according to the generalized binomial distribution with mean value

$$f = N^{-1} \sum_{i=1}^{N} P(GP|CJ, X_i) \tag{27}$$

and variance

$$\sigma_f^2 = N^{-2} \sum_{i=1}^{N} P(GP|CJ, X_i)(1 - P(GP|CJ, X_i)) \tag{28}$$

Figure 17: Distribution of the $P(GP|CJ, X)$ probabilities in the CJ subsample with two b tags in the 177.5 GeV $t\bar{t}$ MC events. The fraction of GP events in this subsample was found by counting to be 72%.



The distribution of GP fractions predicted by the fit vs. those found in the 177.5 GeV $t\bar{t}$ MC events is shown in Figures 18 and 19 for one and two b tags, respectively. The agreement between predicted and observed fractions is satisfactory.

Since Eq. 22 heavily uses the differences between χ^2 values of different jet permutations, we check that distributions of these differences in the data can be reasonably well described by Monte Carlo. Figures 20, 21, and 22 show the results of KS tests performed on the best χ^2 value, the difference between second best and best χ^2 values, and the difference between third best and best χ^2 values. Data and simulation exhibit good agreement within available statistics.

5.2 Likelihood

We define the likelihood of the observed data sample as

$$L(m_t) = \Omega(f_b, N) \prod_{i=1}^{N} \left(f_b P_b(m_i, x_i) + (1 - f_b) P_s(m_i, x_i, m_t)\right) \tag{29}$$

where N is the number of observed events, m_i is the top mass in the ith event (determined for the jet permutation with lowest χ^2), x_i symbolizes all template variables besides measured mass, P_s and P_b are the signal and background densities in the (m, x) space, f_b is the background fraction, and $\Omega(f_b, N)$ is the background constraint function (penalty for deviation from predicted value). In the current version of our analysis $\Omega(f_b, N) \equiv 1$,

Figure 18: Comparison between predicted and observed fractions of GP events in the CJ subsample with one b tag. Each one of the 20 points on the plot was calculated using 153 events. The confidence level is calculated for the hypothesis that the observed fraction is equal to the predicted one.

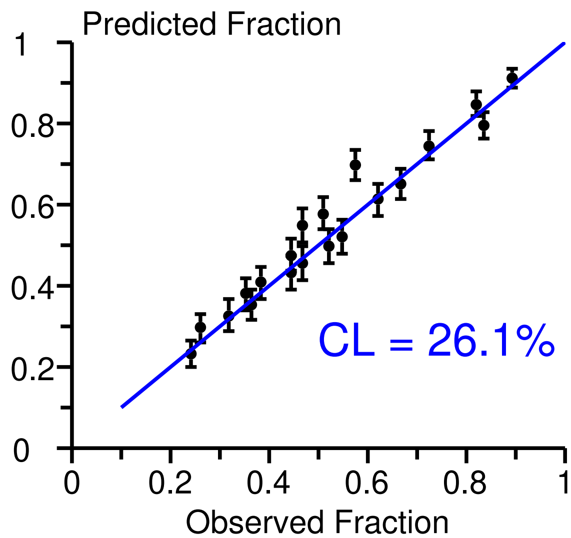


Figure 19: Comparison between predicted and observed fractions of GP events in the CJ subsample with two b tags. Each one of the 20 points on the plot was calculated using 54 events.

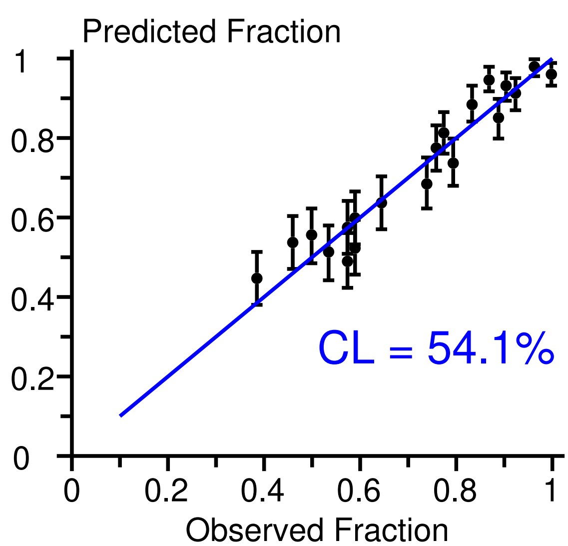


Figure 20: χ^2 of the best permutation. JES constraint value 0.05, background fraction of 35%, and 175 GeV $t\bar{t}$ signal MC were used to make this plot and the two subsequent plots.

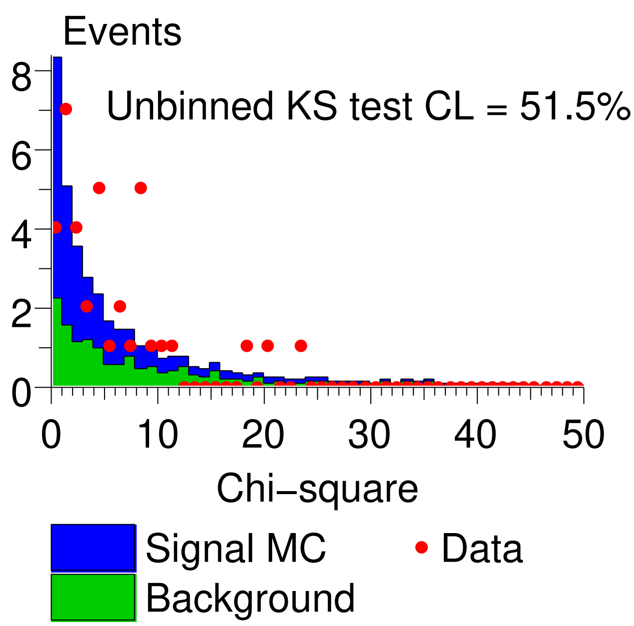


Figure 21: Difference between second best and best jet permutation χ^2.

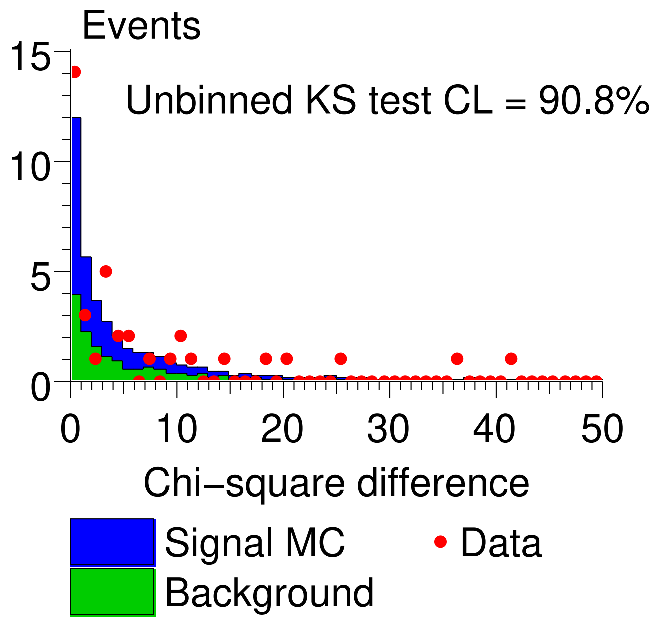
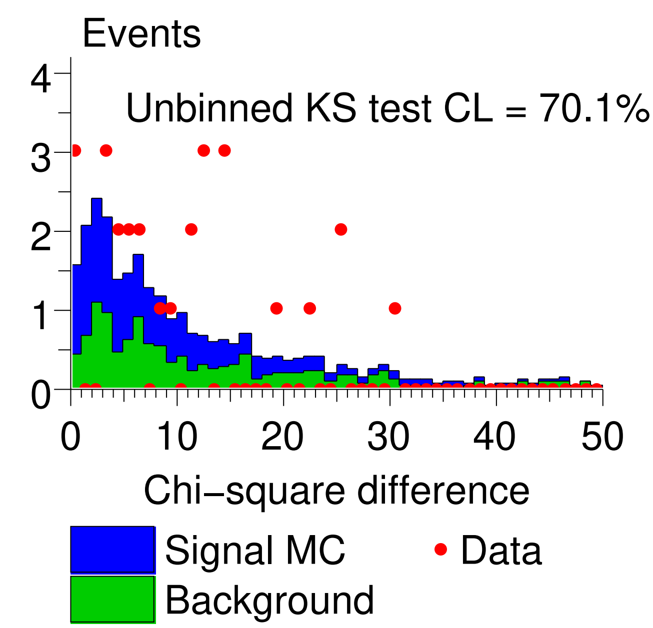


Figure 22: Difference between third best and best jet permutation χ^2.



and the background fraction is allowed to float freely. The signal density is composed as follows:

$$P_s(m, x, m_t) = (p_{\mathrm{GP}} S_{0,m_t}(m, x) + (1 - p_{\mathrm{GP}}) S_{1,m_t}(m, x))\, p_{\mathrm{CJ}} + (1 - p_{\mathrm{CJ}}) S_{2,m_t}(m, x) \quad (30)$$

S_{0,m_t}, S_{1,m_t}, and S_{2,m_t} are the three signal templates for the given generated m_t. p_{GP} is just a short notation for $P(GP|CJ, X)$ defined in Eq. 24, and p_{CJ} is the probability to pick correct jets. We model the background density as

$$P_b(m, x) = \sum_{\text{bg types}} a_j B_j(m, x), \qquad \sum_{\text{bg types}} a_j = 1 \quad (31)$$

where B_j are the templates for different background types, and a_j are the background composition coefficients obtained from the $t\bar{t}$ cross section studies.

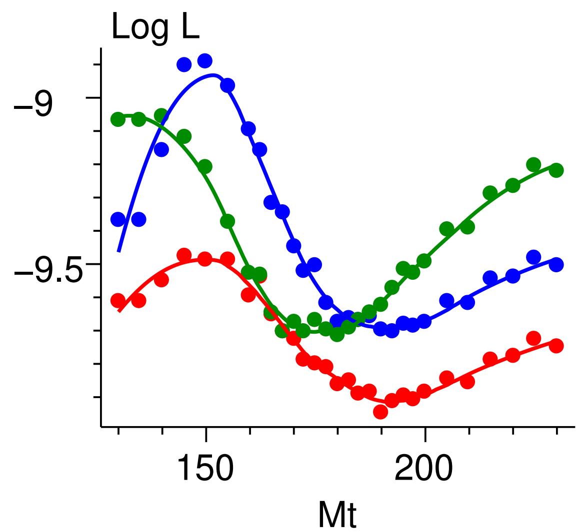
The likelihood in Eq. 29 is not normalized. Instead of trying to impose an explicit normalization condition, we note that the normalization factor would be unnecessary if p_{CJ} and p_{GP} probabilities were completely independent from m and x. We maintain an approximate normalization by estimating p_{CJ} and p_{GP} in a manner which introduces very little dependence on m. However, we do expect that, in general, the width of the pull distribution obtained from pseudo experiments will not be equal to 1, and the pull mean will not be exactly 0.

The likelihood in Eq. 29 is defined only for a discrete set of m_t values for which signal templates were generated. We expect that dependence of the likelihood on m_t should be continuous and smooth for each event. However, our signal templates do not guarantee likelihood continuity because each template is generated using an independent set of Monte Carlo events with finite statistics. To introduce the assumption of continuity and smoothness into the likelihood formula, we use local quadratic polynomial least squares regression with m_t as a predictor and $\log L$ as a response. Interpolation is performed for each event separately, and resulting log likelihood curves are then added to obtain the likelihood of the whole sample. Polynomial regression calculations are performed using routines DPOLFT and DP1VLU from the SLATEC software package [17]. At each interpolation point m, template likelihoods are weighted in the polynomial fit according to the formula $w_i = (1 - ((m_{t,i} - m)/h)^2)^4$ if $|m_{t,i} - m| < h$, and $w_i = 0$ if $|m_{t,i} - m| \geq h$. The bandwidth h is set to 25 GeV. The effect of local regression is illustrated in Fig. 23 for three randomly selected data events.

6 Kinematic Variables

In view of finding a set of variables useful for the multivariate template method, we calculate the numerical values of various global event quantities for each event passing the selection criteria described in Section 2.1. Their differential distributions are intended to have a substantial dependence in the top quark pole mass and/or the discriminative power to distinguish signal from background events.

Figure 23: Smoothing and interpolation of the likelihood with local polynomial regression. Raw likelihood values are shown with circles. Continuous lines display regression results.



Our primary variable, the reconstructed invariant top mass, seems to be a natural choice. In addition to it, we consider further kinematic variables which e.g. are based on the transverse energies of jets and leptons seen in an event. The various variable candidates are discussed in 6.1. We anticipate that we finally try in turn the following ones as a second kinematic variable for our mass analysis.

- $p_{\mathrm{T}}^{(1-4)}$, the sum of the transverse momenta of the four leading jets;
- H_{T}, the sum of the transverse momenta of all jets, plus the transverse momentum of the charged lepton, plus the transverse missing energy;
- M_{ij}^{min}, the minimum invariant di-jet mass among the four leading jets;
- The acoplanarity Aco which quantifies the deviation of the shape of an event from an ideal planar configuration (an exact definition of this quantity is given in Section 6.2).

We calculate all variables at jet correction level 7. Where applicable, all jets with $p_{\mathrm{T}} > 8\,\mathrm{GeV}$ and $\eta < 2$ at the jet correction level relevant for the event selection are included. For the H_{T} variable we use the level 7 corrected missing transverse energy, which differs from the missing transverse energy used for the event selection.

The numerical values of H_{T}, $p_{\mathrm{T}}^{(1-4)}$, M_{ij}^{min} and Aco extracted from the top candidate events are listed in Table 8 together with the values of there constructed top mass M_{top}. The distributions of these variables are shown in Figure 24 and compared with the combined Monte Carlo predictions for signal (with $M_{\mathrm{gen}} = 175\,\mathrm{GeV}$) and for the background

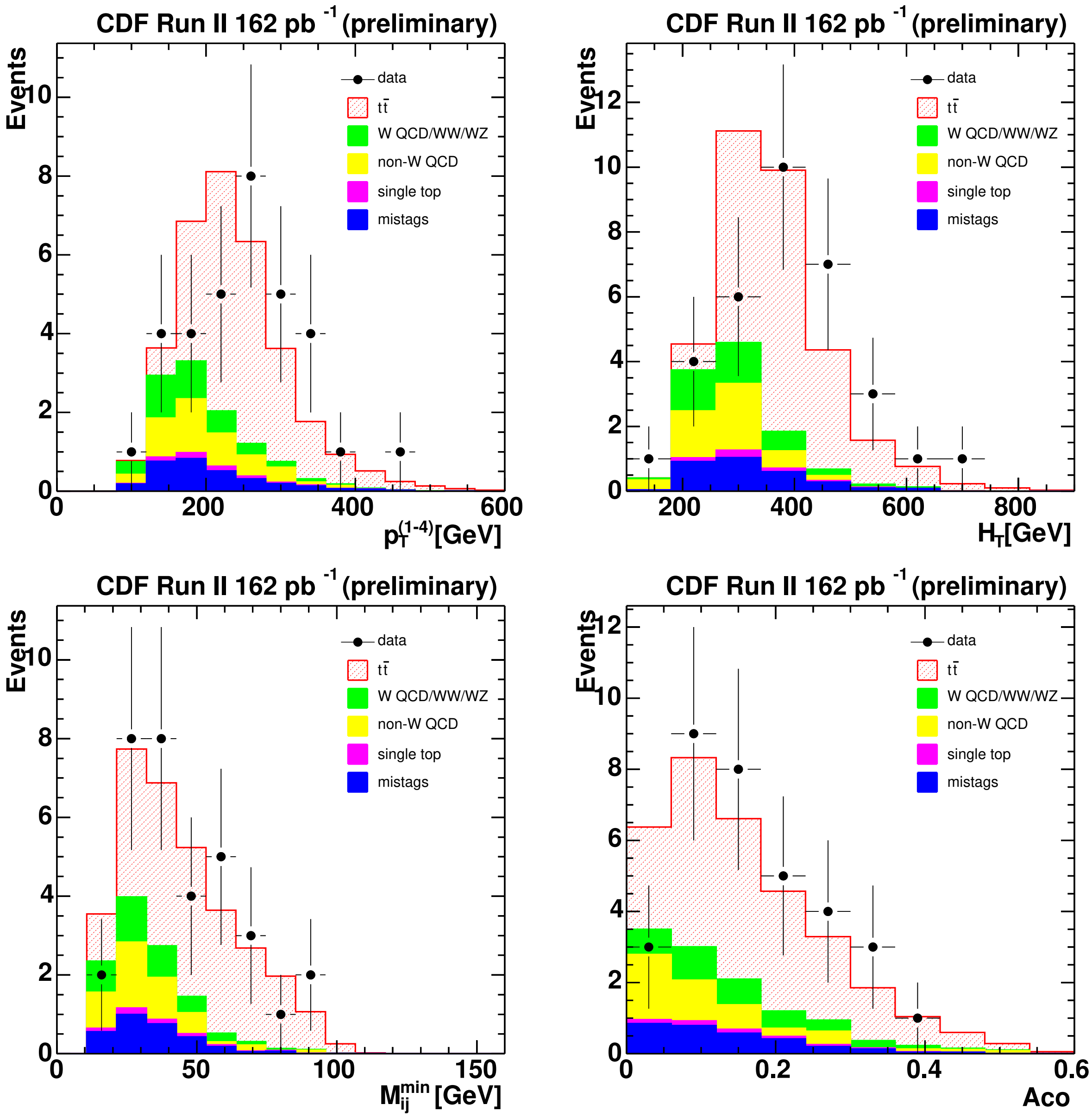


Figure 24: The data distributions of the variables H_{T}, $p_{\mathrm{T}}^{(1-4)}$, $\min(M_{ij})$, and Aco (points). Overlaid are the "stacked" MC expectations for the signal contribution as well as the various background contributions using the normalization constraints from the cross section analysis [4] and the fitted total background fraction (Section 8.1).

contributions according to [4]. Similar to what was found for M_{top}, the plots show reasonable agreement between the data and the Monte Carlo simulation.

6.1 On the Choice of Kinematic Variables

The goal of the multivariate template method is to add more event information into the likelihood formula used for the extraction of the top quark mass by means of probability density distributions of useful event observables. The major objective is to find a set of "good" variables which increase the sensitivity of the likelihood to the top quark mass and improve the discrimination between signal and background. In the following we introduce a set of variables of different types which may be appropriate for our purposes. It is desirable to make a reasonable pre-selection among them without being dependent on the results given by the (CPU) time consuming machinery of the whole top mass analysis. Therefore we seek for promising candidates by following means: 1. We study the correlation of a given set of kinematic variables reconstructed from MC signal samples with the underlying generated top mass (Section 6.4). 2. We quantify the differences between signal and background distributions of the variables by means of appropriate histogram distance measures (Section 6.5).

6.2 Candidate Variables

Due to the large mass of the top quark, the transverse energies observed in $t\bar{t}$ events are generally larger than in non-top background events, like e.g. W plus multihadronic jets from QCD processes which presents a major background contribution. Likewise, signal events are expected to be more spherical. These and further properties specific to $t\bar{t}$ events give reason to consider the following list of energy variables, angular quantities, and event topology observables as candidates variables. Unless otherwise noticed, the quantities are calculated at correction level 7.

- Kinematic energy variables:
  - The transverse momenta of the four leading jets: $p_{\mathrm{T}}^{(i)}$ $(i = 1 \dots 4)$.
  - Various combinations of sums of the transverse jet momenta:
    $p_{\mathrm{t}}^{(i+j)} \equiv p_{\mathrm{T}}^{(i)} + p_{\mathrm{T}}^{(i+1)}$ $(i = 1 \dots 3)$,
    $p_{\mathrm{t}}^{(1-3)} \equiv \sum_{i=1}^{3} p_{\mathrm{T}}^{(i)}$,
    $p_{\mathrm{t}}^{(1-4)} \equiv \sum_{i=1}^{4} p_{\mathrm{T}}^{(i)}$.
  - The transverse energy of the charged lepton: $E_{\mathrm{T}}(l\pm)$.
  - The transverse missing energy: $\not{E}_{\mathrm{T}}$.
  - The scalar sum of the missing transverse energy, the transverse momentum of the charged lepton, and the transverse momenta of all jets:
    $H_{\mathrm{T}} \equiv \sum_{\mathrm{jets}} p_{\mathrm{T}}(i) + E_{\mathrm{T}}(l^{\pm}) + \not{E}_{\mathrm{T}}$.
  - The minimum di-jet mass among the four leading jets: $M_{ij}^{\min}$.
  - The transverse momentum of the reconstructed top quark and of the reconstructed $t\bar{t}$ system: $p_{\mathrm{T}}^{(\mathrm{top})}$, $p_{\mathrm{T}}^{(t\bar{t})}$ (calculated from the fitted momenta of the $\chi^2_{\min}$ solution).

- The invariant mass of the reconstructed $t\bar{t}$ system: $M_{t\bar{t}}$ (calculated from the fitted momenta of best $\chi^2_{\rm min}$ solution).

- Angular and variables:
  - The angle between the reconstructed charged lepton and the reconstructed b quark from the leptonic side in the $l^{\pm}\nu$ rest frame: α^* (calculated from the fitted momenta of the best $\chi^2_{\rm min}$ solution).
  - The minimum separation $\Delta R^{\rm min}_{ij}$ between the jets times the ratio of lowest $p_{\rm T}$ jet from the jet pair used to calculate $\Delta R^{\rm min}_{ij}$, and the charged lepton $p_{\rm T}$: $R^{\rm min} \equiv \Delta R^{\rm min}_{ij} \times \min(p_z^{(i,j)})/p_{\rm T}(l^{\pm})$.
  - The ratio of the sum of the longitudinal and transverse momenta of jets and leptons: $(\sum_{i=1}^{4} |p_z(i)| + p_z(l^{\pm})/(\sum_{i=1}^{4} p_{\rm T}(i) + p_{\rm T}(l^{\pm}))$.

- Shape variables (calculated from the momenta $\vec{p_i}$ of the jets and the charged lepton):
  - The sphericity $S = \frac{3}{2}(\lambda_1 + \lambda_2)$ calculated from the eigenvalues $\lambda_1 < \lambda_2 < \lambda_3$ of the momentum tensor $\Theta_{ab} = \sum_i p^i_a p^i_b / \sum_i |\vec{p_i}|^2$.
  - The acoplanarity $Aco = 4\min(\sum_i |\vec{p_i}|^{\rm out} / \sum_i |\vec{p_i}|)^2$. $\vec{p_i}^{\rm out}$ is the component of p_i perpendicular to the event plane. The plane has to be varyied until the expression in brackets is minimized.
  - The thrust $T = \max_{\vec{n}}(\sum_i |\vec{p_i} \cdot \vec{n}| / \sum_i |\vec{p_i}|)$. $\vec{n}$ has to be chosen such that the expression in brackets is a maximum.

6.3 Check of Basic Kinematics Distributions

Since the present studies are completely relying upon the MC samples introduced in Section 2, a good understanding of both measured and simulated objects such as leptons and jets is required. A mandatory check is to verify if basic event quantities used for the calculation of the kinematic variables mentioned above are reasonably reproduced by the Monte Carlo simulation.

A comparison of the measured transverse momenta $p_{\rm T}$ of the four leading jets, the charged lepton $p_{\rm T}^{\rm (lep)}$, and the corrected transverse missing energy $E_{\rm T}^{\rm (miss)}$ with the combined signal plus background predictions is shown in Figure 25. Likewise, Figure 24 shows the data versus MC prediction for $p_{\rm T}^{(1-4)}$, $H_{\rm T}$, $M^{\rm min}_{ij}$, and $p_{\rm T}^{(1-4)}$. We have performed these comparisons at different jet correction levels and also for various quantities calculated from the best $\chi^2_{\rm min}$ solution of the top mass fit. We generally found that the Monte Carlo modeling of the underlying physical processes as well as the simulation of the detector response is consistent with the data, thus supporting that the MC samples are appropriate for our purposes.

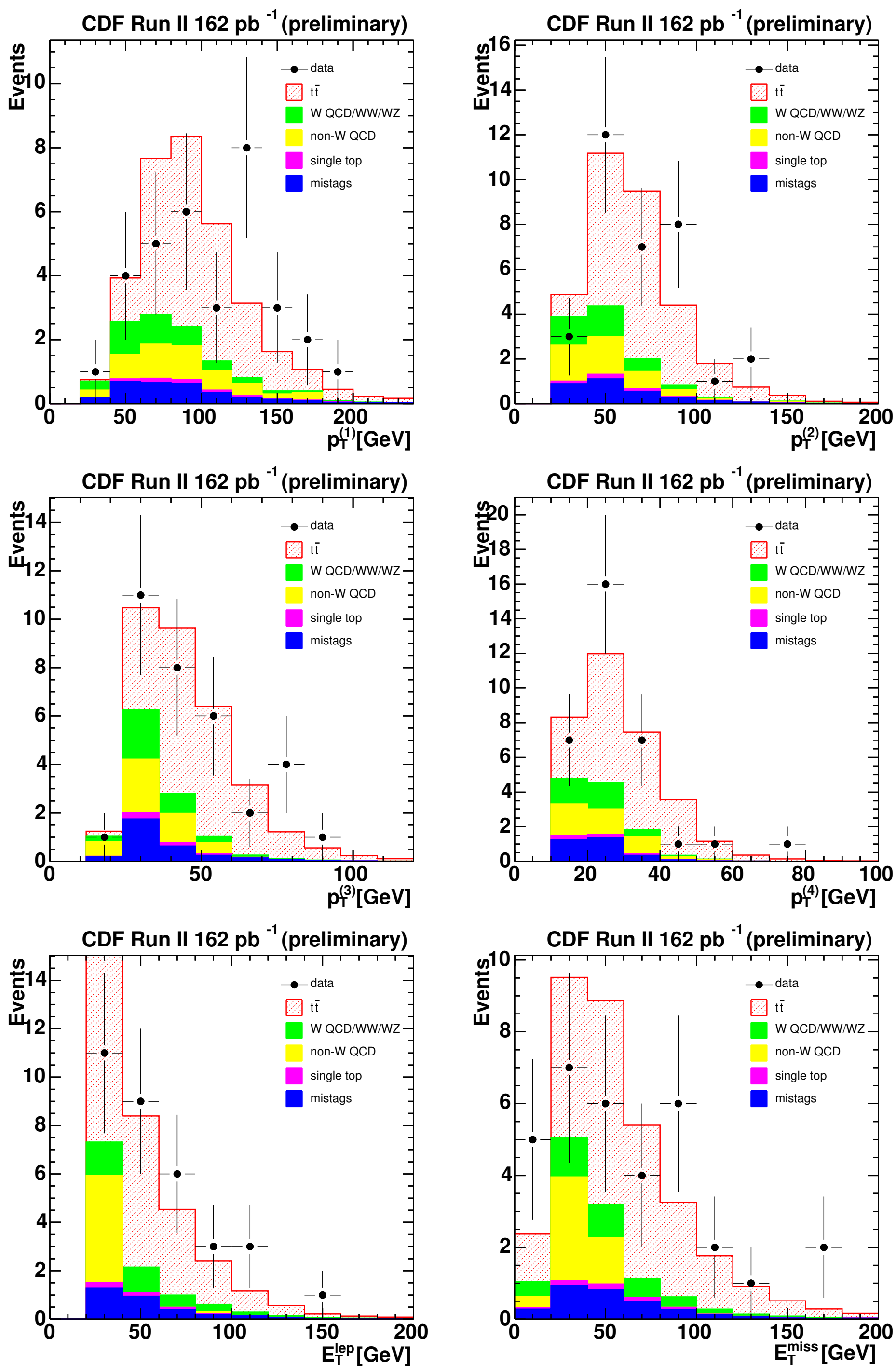


Figure 25: The data distributions for the transverse energies p_{T} of the four leading jets, the charged lepton $p_{\mathrm{T}}^{(\mathrm{lep})}$, and the missing transverse energy $E_{\mathrm{T}}^{(\mathrm{miss})}$ calculated at jet correction level 7 (points). Overlaid are the "stacked" MC expectations for the signal and the background contributions.

6.4 Sensitivity on the Top Quark Mass

We examine the sensitivity of our second kinematic variable x_2 on the top quark mass by calculating the covariance matrix σ_{ij} for a set of variables $\{x_0, x_1, x_2\}$ which include the reconstructed top mass $x_1 = M_{\text{top}}$ and the nominal top quark mass $x_0 = M_{\text{gen}}$ used for the generation of the signal samples. We consider a range of generated top masses between $M_{\text{gen}} = 165\,\text{GeV}$ and $185\,\text{GeV}$. We try both a straightforward calculation, $\sigma_{ij}^{\text{simple}}$, and the minimal covariance determinant (MCD) method [18], σ_{ij}^{MCD}. The latter result is based on a subset of observations for $\{M_{\text{gen}}, M_{\text{top}}, x_2\}$ for which the determinant of their covariance matrix is a minimum. The MCD method provides a more robust estimate of the covariance matrix since it depends to a lesser extend on the presence of outliers in the variable space e.g. caused due to mismeasurement of event quantities.

From the covariance matrices we get the correlation coefficients $\rho_{ij}^{\text{simple}}$ and ρ_{ij}^{MCD}, respectively, whereby the absolute value for ρ_{ij}^{MCD} explicitely depends on the given fraction f of data points typically chosen to be 70-90%. Furthermore, the numbers vary for different ranges of generated top masses taken under consideration. However, the conclusion we draw from the comparison of the correlation coefficients of different variables should not depend on either the value for f or the M_{gen} range. We use $f = 0.9$ for the numbers and plots shown in the following.

A second useful variable is expected to have a ρ_{02} value as large as possible. Since this does not necessarily mean that it adds significant information about the top mass of an event compared to what M_{top} already provides, we also determine the global correlation coefficient $R_i = (1 - 1/\rho_{ii}(\rho^{-1})_{ii})^{\frac{1}{2}}$ which measures the strongest correlation between a variable i and a linear combination of all other variables.

The top left hand side of Figure 26 shows the correlation coefficients $\rho_{02} = \rho(M_{\text{gen}}, x_2)$ obtained using the two different methods. The values for ρ_{02} range between -0.022 and $+0.127$ for the simple calculation and -0.017 and $+0.168$ for the MCD method. The variables most highly correlated to M_{gen} are (sorted by decreasing order): H_{T}, ($M_{t\bar{t}}$ in MCD case), $p_{\text{T}}^{(1-4)}$, and $p_{\text{T}}^{(1-3)}$. The conclusion is that a given quantity is the more sensitiv to M_{gen} the more objects of an event it comprises, according to the naive expectation. Furthermore, a look to the $p_{\text{T}}^{(i)}$ results tells us that higher energetic jets seem to carry more top mass information than lower energetic jets. The information provided by the missing transverse energy and the lepton momentum are less useful. Angular quantities and shape variables have negligible correlation with M_{gen}.

To compare, the correlation coefficient for M_{top} and M_{gen} is $\rho_{01}^{\text{simple}} = 0.138$ and $\rho_{01}^{\text{MCD}} \sim 0.21$, respectively (the latter value slightly depends on the second variable). This means that the top performers mentioned above are almost equally competitive with our primary template variable.

One might expect that the second variable should only be moderately correlated with the reconstructed top mass M_{top} in order to provide additional top mass information of an event. The corresponding correlation coefficients ρ_{12} are shown in the top right-hand side plot of Figure 26. They indeed vary between $+0.18$ to $+0.48$ for the pure hadronic observables and H_{T}, thus supporting that these variables may be useful. The

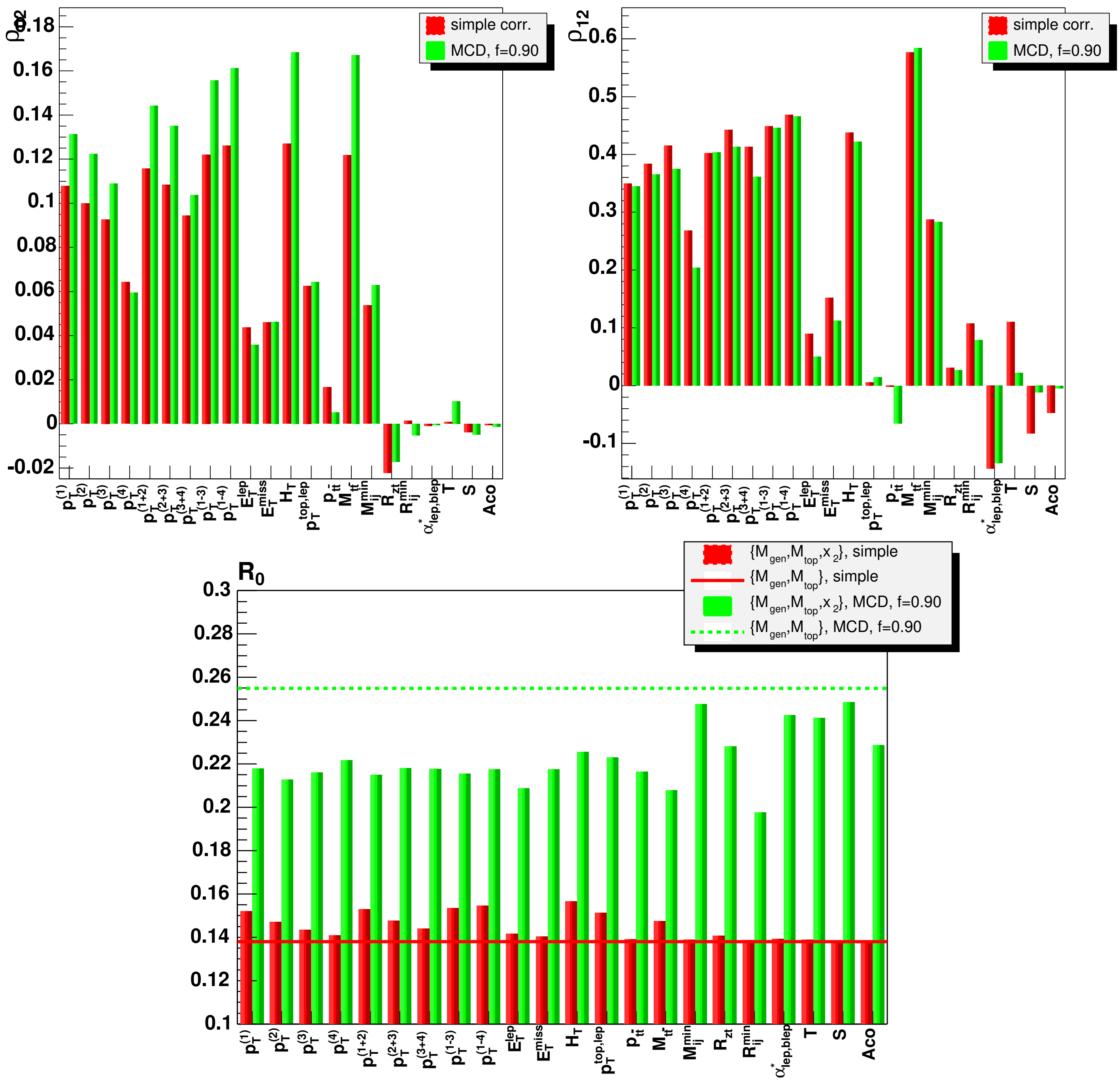


Figure 26: The correlation coefficients $\rho_{02} = \rho(M_{\rm gen}, x_2)$ (top, left-hand side) and $\rho_{12} = \rho(M_{\rm top}, x_2)$ (top, right-hand side) calculated using the simple method and the MCD method, respectively, see text for details. The plot at the bottom shows the global correlation coefficient R_0 for $\{M_{\rm gen}, M_{\rm top}, x_2\}$ (bars) and for $\{M_{\rm gen}, M_{\rm top}\}$ alone (lines).

leptonic and missing transverse energy as well as the angular and shape variables are almost uncorrelated with $M_{\rm top}$. $M_{t\bar{t}}$ reveals (not surprisingly) the strongest correlation with $M_{\rm top}$.

Finally, we show the global correlation coefficients R_0 in the bottom plot of Figure 26. We also compare the values obtained from the scenario with two template variables, $\{M_{\rm gen}, M_{\rm top}, x_2\}$, with the scenario based on only one template variable, $\{M_{\rm gen}, M_{\rm top}\}$. According to the R_0 results obtained with the simple covariance method, the second variable adds only marginal top mass information to what $M_{\rm top}$ already provides. Again, there is some preference for $H_{\rm T}$ and $p_{\rm T}^{(1-4)}$ with $R_0^{\rm simple} = +0.157$ and $R_0^{\rm simple} = +0.154$, which is only slighty larger than the value $+0.138$ obtained with $M_{\rm top}$ alone. The angle and shape variables do not add further sensitivity on the top mass.

In contrast, the comparison of the results based on the MCD method suggest that adding a further variable adds rather "noise" than useful signal information. This observation does not depend on the fraction f choosen for the MCD method. The $\{M_{\rm gen}, M_{\rm top}\}$ scenario shown here yields $R_0^{\rm MCD} = +0.255$, whilst adding a second variable decreases the global correlation coefficient down to $+0.197$. The top performer here is $M_{ij}^{\rm min}$ with $R_0^{\rm MCD} = +0.248$.

Further scenarios were examined in order to maximize the global correlation coefficient R_0, which we do not document here in detail. We have tried all possible combinations $\{M_{\rm gen}, x_1, x_2\}$ without restricting on x_1 as our primary variable. Similarily we investigated the 4-dimensional case $\{M_{\rm gen}, M_{\rm top}, x_2, x_3\}$ and tried various combinations for x_2 and x_3. Summing up, it may be said that there seem to be no way to significantly optimize R_0 w.r.t. the $\{M_{\rm gen}, M_{\rm top}\}$ case.

6.5 Signal/Background Discrimination

We further investigate the capability of the variables mentioned above to distinguish signal events from background events. In particular we evaluate various "divergence" measures motivated by the information theory which can be used to quantify the differentness of probability density distributions, like for example histograms obtained from signal and background Monte Carlo.

Given two probability distributions $p_1(x)$, $p_2(x)$ of a variable x, one can calculate the quantity

$$d(p_1, p_2) = \int \mathrm{d}x \; p_1(x) \log \frac{p_1(x)}{\alpha \cdot p_2(x) + (1-\alpha) \cdot p_1(x)} \tag{32}$$

or, if x is discrete in n bins,

$$d(p_1, p_2) = \sum_{i=1}^{n} \; p_1(x_i) \log \frac{p_1(x_i)}{\alpha \cdot p_2(x_i) + (1-\alpha) \cdot p_1(x_i)} \tag{33}$$

where α is a free parameter. The special case $\alpha = 1$ is known as the *Kullback-Leibler divergence* KL [19], a non-negative but unsymmetric measure which is 0 only for identical probability distributions. The KL mesasure makes only sense for absolutely continuous distributions: it is not defined for bins with $p_2(x) = 0$, as is e.g. often the case when comparing histograms with limited data statistics. In order to cope with this problem, further divergence measures were proposed, like e.g. the *α-skew divergence* [20] given by Eqn. 32 and 33 for $0 < \alpha < 1$. Here the distribution p_2 is being smoothed by the distribution p_1 by an amount controlled by the parameter α. In [20], a value $\alpha = 0.99$ is favoured as an approximation to the KL distance with various advantageous features, which we call *$K_{.99}$ divergence* in the following. Another value $\alpha = 0.5$ is proposed in [21] referred to as *K divergence*.

To obtain a symmetric distance measure, one can define

$$d_s(p_1, p_2) = d(p_1, p_2) + d(p_2, p_1) \tag{34}$$

as suggested in [20, 21]. There is no accepted gold standard for similarity measures. See the references cited for further discussion about the various distributional similarity measures.

In the following, we measure both the K and $K_{.99}$ distances according to Eq. 34 between the signal and background distributions of the variables introduced in Section 6.2. Thus we check that the conclusions we draw do not depend on the amount of smoothing introduced by the parameter α.

An appropriate way to evaluate Eq. 33 for variables whose numerical values x_i span different ranges $[x_i^{\min}, x_i^{\max}]$ of values and have different dimensions, we map the distributions by $\hat{x} = (x_i - x_i^{\max})/(x_i^{\min} - x_i^{\max})$ and then determine the distances from the probability densities $p(\hat{x}_i) = N_i/(N\Delta\hat{x}_i)$, with N the total numbers of events and N_i the number of events in bin i of the distribution. Since the distances depend on the number of bins, we ensure that the $p(\hat{x}_i)$ distributions for the different variables have the same binning. We also try different total numbers of bins.

For comparison, we also calculate the *Kolmogorov-Smirnov* distance KS, which is basically the maximum distance between the cumulative distributions of $p_1(x)$ and $p_2(x)$. Note that KS is primarily sensitive to the difference of the global shape of two distributions.

We consider signal distributions extracted from the Herwig $t\bar{t}$ Monte Carlo for various generated top masses between $M_{\rm top}$=130 GeV to 230 GeV. For the background, we take the standard W+4 jets samples obtained from the ALPGEN+HERWIG Monte Carlo, assuming that the shapes of the other major background distributions are not significantly different.

As an example, the series of plots at the top of Figure 27 compares the normalized signal and background distributions for the $H_{\rm T}$ variable at generated top masses $M_{\rm top} =$ 150 GeV, 180 GeV and 210 GeV. The plot at the bottom shows the numerical values of the three distance measures considered here for top masses ranging from $M_{\rm top}$ =130 GeV to 230 GeV. Note that for a given top mass, the absolute values of the various distance measures are rather different. However, in the scaled representation shown in the plot, the measures agree reasonably in their relative dependence on $M_{\rm top}$, and reflect quantitatively

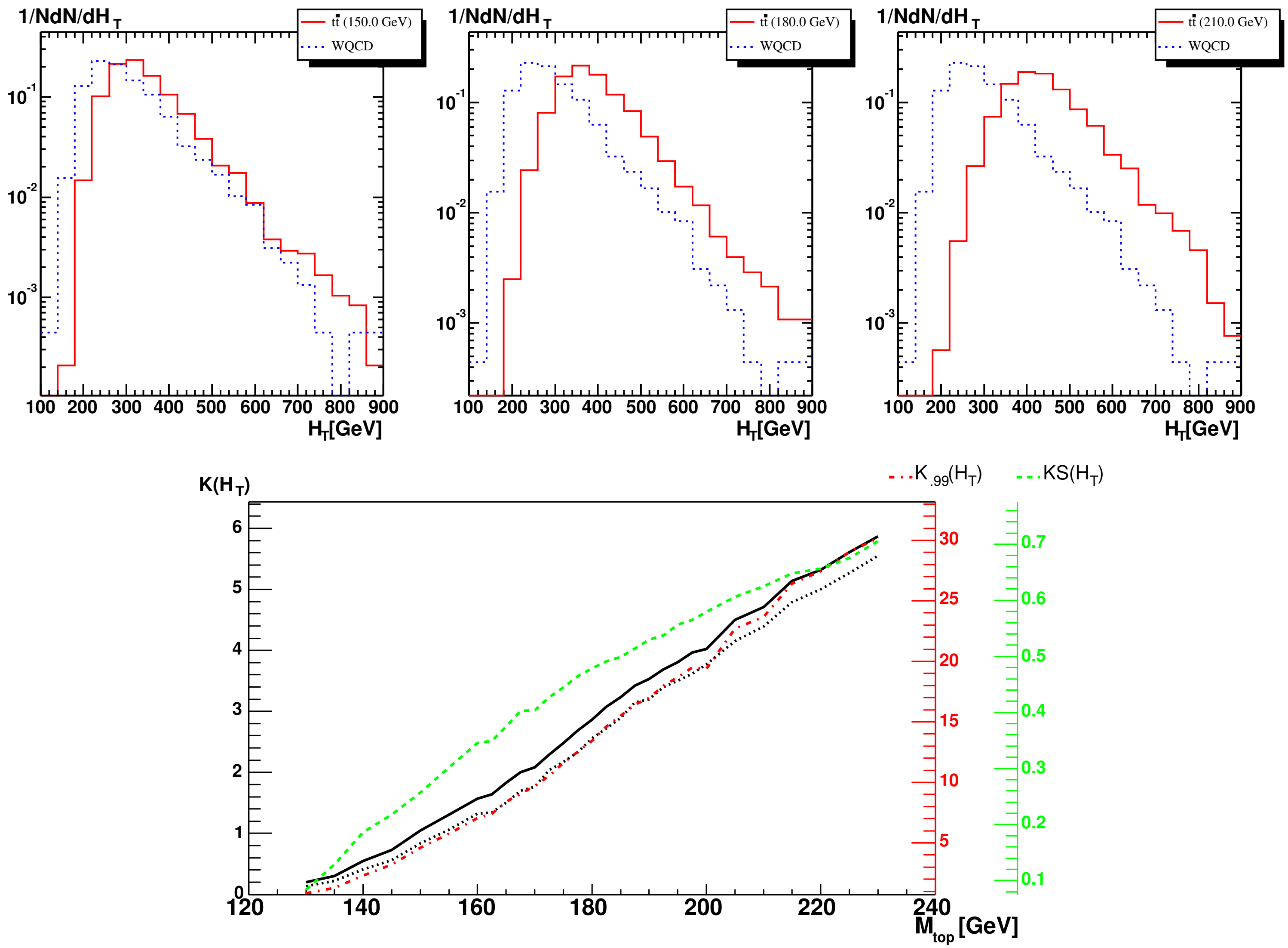


Figure 27: Top: The normalized H_{T} distributions for a $t\bar{t}$ signal at $M_{\mathrm{top}} = 150\,\mathrm{GeV}$, $180\,\mathrm{GeV}$ and $210\,\mathrm{GeV}$ (red solid lines) and for the background W+4jets (blue dotted lines). Bottom: The symmetrized K divergence (black solid curve), the 0.99-skew divergence ($K_{.99}$, red dashed curve), and the Kolmogorov-Smirnow distance KS (green dashed-dotted curve) between signal and background distributions for H_{T} as a function of M_{top}. Also shown is the symmetrized K divergence extracted from the the raw H_{T} distributions (black dotted curve).

the fact that the signal to background separation for the H_{T} variable increases with higher top masses.

In Figure 28 we compare the numerical values of the distance measures for all variables for $M_{\mathrm{top}} = 150\,\mathrm{GeV}$, $180\,\mathrm{GeV}$ and $210\,\mathrm{GeV}$. The left-hand side of each plot contain the results for the kinematic energy variables, the right-hand side the results for the angle and shape variables. The following conclusions can be drawn.

- The K and $K_{.99}$ measures give a similar picture. The pattern of these measures agree also roughly with the KS distance. Changing the number of bins in the probability distributions expectedly changes the absolute scale but not the pattern.

- Looking to the absolute scales, the distances grow with increasing top masses in particular for the kinematic energy variables. Quantities like H_{T} and $p_{\mathrm{T}}^{(1-4)}$ show the strongest dependence on M_{top}, with distance values increasing from $M_{\mathrm{top}} = 150\,\mathrm{GeV}$ to $210\,\mathrm{GeV}$ by a factor of about 5. In contrast, the distances extracted from angular and shape variables like the acoplanarity Aco do generally not significantly change for different top masses. This observations are in agreement with the findings of our correlation studies.

- The distances extracted from level 7 corrected variables are generally not very different from those extrated from the raw quantities, see e.g. Fig. 27 for the H_{T} case.

- Kinematic energy quantities discriminate signal from background better than angle or shape variables, in particular at higher values for M_{top}. Pure leptonic energy variables have almost no discriminative power.

- Our primary variable M_{top} is only average performer in distinguishing signal and background.

- The variable $p_{\mathrm{T}}^{(1-4)}$ is top performer among the variables for values of $M_{\mathrm{top}} \geq 160\,\mathrm{GeV}$.

6.6 Conclusions

From the correlations studies it comes out that we do not add significant sensitivity on the top mass by adding more template variables to our primary variable, the reconstructed invariant top mass M_{top}, which seems to be already an optimal carrier of top mass information of an event. The interpretation of the different behaviour of the global correlation obtained from the straightforward calculation and the MCD method is somehow difficult. Since in the limit $f \to 1$ the results based on the MCD method converge to the result obtained from the straightforward calculation, the (seeming) contradiction observed is coupled to the question of how to handle outliers in the variable space correctly. Therefore we consider the results from both methods. Looking to the global correlation

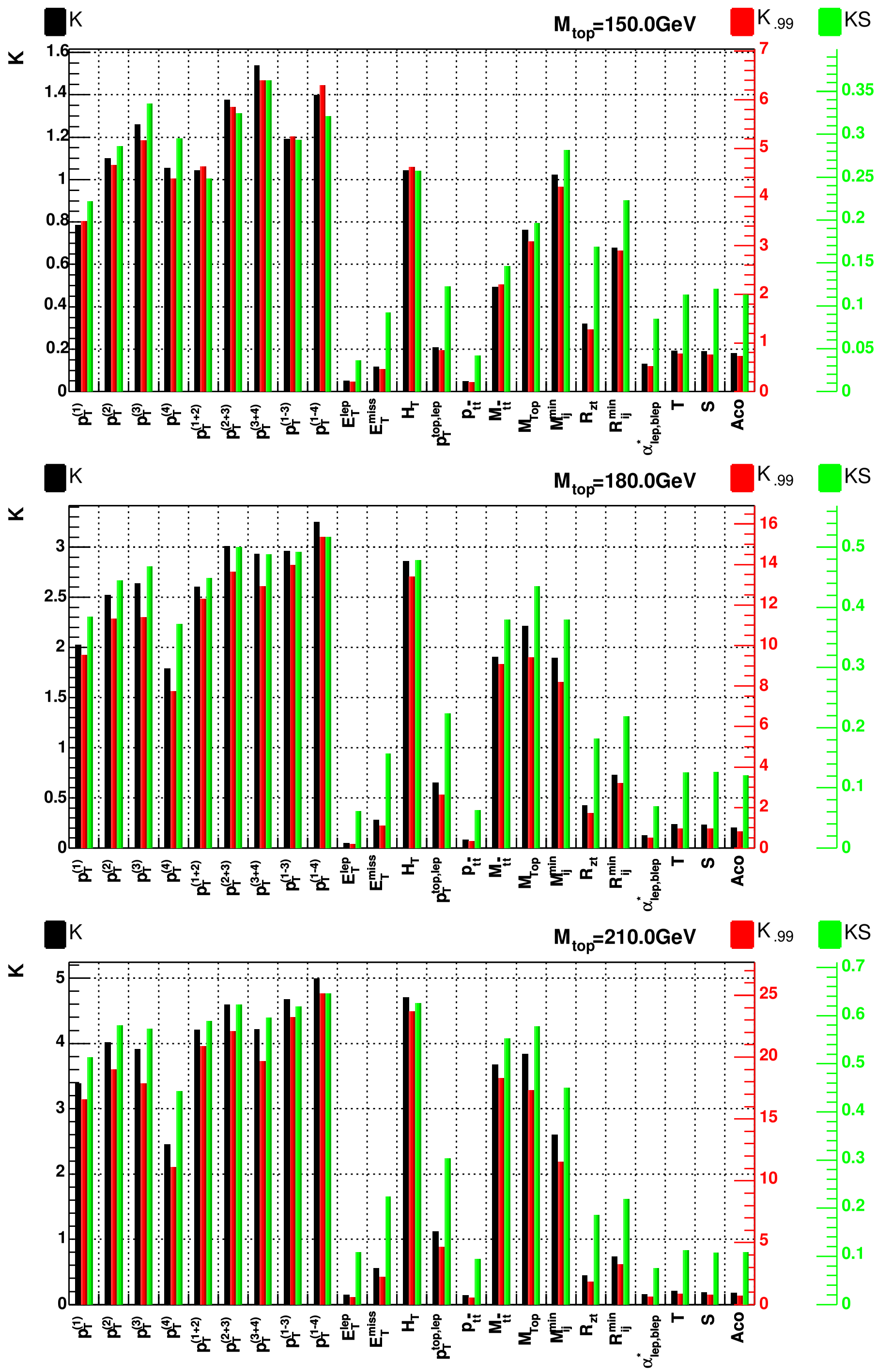


Figure 28: The symmetrized K divergence (black bars), the 0.99-skew divergence ($K_{.99}$, red bars), and the Kolmogorov-Smirnow distance KS (green bars) between signal ($t\bar{t}$) and background (W+4jets) distributions of various variables for generated top masses $M_{\rm top} = 150\,\mathrm{GeV}$, $180\,\mathrm{GeV}$ and $210\,\mathrm{GeV}$.

coefficient obtained using the simple method, there is some preference for the variable H_T and $p_\mathrm{T}^{(1-4)}$, whereas the MCD method favors e.g. M_{ij}^{min}.

A clearer motivation for the choice of variables is given by their capability to discriminate signal from background events. The various divergence measures extracted from the signal and background Monte Carlo samples support kinematic energy variables based on jets. Top performer is the sum of the transverse momenta of the four leading jets $p_\mathrm{T}^{(1-4)}$. Topology variables have less discriminative power, however it might be interesting to try also this class of variables for the multivariate method as complement to the pure energy variables, in particular in view of studying the systematics uncertainties of a top mass measurement. Shape variables behave differently from kinematic energy variables because they provide background information but reveal a much weaker dependence on the given value of a top mass. In this respect, the acoplanarity Aco used as classical shape variable in various previous top quark studies may be useful for these studies.

7 Testing the method on Monte Carlo events

Before measuring the top mass from the data sample, we subject our method of measurement to study by using it on numerous pseudo experiments (PEs). Our goal in this analysis is to find optimal settings for a top mass measurement which minimize the total (systematic and statistical) error. More specifically, we would like to find two things: the best combination of variables to calculate our likelihoods with, and the ideal jet energy scale constraint.

7.1 Creating the Pseudo-Experiments

We use four different event pools to construct our pseudo experiments: 175.0 GeV signal MC (using independent events from those in the template), $Wb\bar{b}2j$ MC, $W4j$ MC sample which, together with the mistag matrix, is used to model SVX mistags, and QCD data identified by the requirement Iso > 0.2. Due to the limited size of background MC and data samples, events which simulate the background are taken from templates themselves (these events are removed from the templates during the mass calculation to avoid overfitting). Our preliminary fit of the templates to the data indicated that the actual background fraction in the sample is 34%. The number of events taken from each sample is allowed to fluctuate around its expected average according to the Poisson distribution. The total number of events in the sample was taken to be 32 — this is the number of events which passes our χ^2 cut at the jet energy constraint value of 0.05.

The average number of background events was taken to be 10.9. These events were partitioned according to the relative fractions presented in Table 4.

7.2 Estimation of the Statistical Error

We calculate positive and negative statistical errors for each pseudo experiment as the product of the errors returned by the fitting procedure and the width of the pull distribution. Assuming that the result is returned in the form $m_{rec} + \sigma_R - \sigma_L$ for generated mass m_{gen}, the pulls are built as histograms filled with $(m_{gen} - m_{rec})/\sigma_R$ values weighted by $\sigma_L/(\sigma_L + \sigma_R)$ when $m_{rec} < m_{gen}$, and filled with $(m_{gen} - m_{rec})/\sigma_L$ values weighted by $\sigma_R/(\sigma_L + \sigma_R)$ when $m_{rec} > m_{gen}$. Pull width is defined as half of the width between the 16th and 84th percentiles of the histogram.

We estimate the expected symmetrized statistical error from a set of pseudo experiments by finding the peak position in the observed distribution of symmetrized errors. The peak position is estimated by smoothing the sample of observed errors with a kernel density estimator, and scanning the range from 0 to 19 GeV with a small step. The peak position is then multiplied by the pull width to get the expected peak error.

7.3 Jet Systematic Error Calculation

The calculation of the systematic uncertainty due to the jet energy measurements is discussed here. The sources of systematics are described in CDF-6565 [9]. In all there are 7 level of jet corrections and 8 systematic uncertainties on the jet energies as follows:

- Relative corrections, i.e., the energy scale as a function of η is corrected to that of the central calorimeter. These systematic uncertainties are shown in Table 11. For this type of corrections, we apply both data and Monte Carlo uncertainties to the simulated events in pseudo experiments.
- Time dependent corrections for the plugs.
- Raw jet energy scale: takes into account differences between run I and run II. Both data and Monte Carlo errors are used.
- Multiple interactions.
- Absolute energy scale, still using the Run I values.
- Underlying event.
- Out of Cone corrections: still using the Run I values.
- Splash out correction: takes into account the energy outside the cone with R=1.0. The correction is 0 GeV, the systematic uncertainty is 1 GeV.

These are the standard jet-related systematics sources used in many top analyses. We add all systematic errors in quadrature. The top mass systematic shift due to the jet systematic errors is determined by running pseudo experiments with jets shifted by +1 and -1 sigma in the simulated data events. Templates used for fitting do not change, and the same event is not allowed to appear both in the templates and in the simulated data sample. The mass shift is determined as the difference between medians of reconstructed mass distributions in two sets of pseudo experiments. We believe that simultaneous adjustment of all jets in the same direction is a conservative procedure, and should not result in a reduction of systematic error on top mass compared to using jet systematic errors with other directional correlation patterns.

Table 11: Systematic uncertainties on the relative correction on jets from Ref [9].

η range	data	simulation
$\|\eta\| < 0.2$	3%	1%
$0.2 < \|\eta\| < 0.6$	0.5%	1%
$0.6 < \|\eta\| < 1.0$	2%	1%
$1.0 < \|\eta\| < 1.4$	4%	7%
$1.4 < \|\eta\| < 2.0$	2%	6%
$\|\eta\| > 2.0$	7%	7%

7.4 Results of Pseudo Experiments

For the purpose of analysis optimization, we define the total top mass error as the sum in quadrature of expected statistical error and expected jet energy systematic error. Figures 29, 30, and 31 show the expected statistical, systematic, and total error as a function of jet energy scale constraint, for several sets of template variables. Although the reconstructed top mass by itself appears to perform better than other variable combinations, we choose to employ the 2-dimensional templates of reconstructed top mass vs. the scalar sum of p_T of the four leading jets. This decision is motivated by our perception that the difference between total expected errors is not significant in comparison with expected error scatter, while presence of the second variable in the template with good signal/background separating power makes the analysis method less sensitive to problems in the background modeling. Also, the study of Section 6.5 shows that the sum of p_T of the four leading jets has better discriminating power than other variables, as seen in Figure 28. For this variable set, we choose the JES constraint value of 0.07 which provides the lowest expected total error. At this constraint value, the systematic error on the top mass from jet reconstruction is 6.7 GeV.

8 Results

8.1 Top mass measurement

The logarithm of the top mass likelihood obtained from data events is shown in Figure 32. The curve peaks at $m_t = 179.6$ GeV. We obtain positive and negative errors from this curve by finding such top mass values where the likelihood becomes smaller than its value at the peak by 0.5. The errors from the curve are +5.8 GeV and -5.7 GeV.

We check the pull distributions using pseudo experiments. The pseudo experiments were performed for JES constraint value of 0.07. We made 5000 pseudo experiments for each generated mass value from 160 to 200 GeV, with 5 GeV step. We generate 33 events per pseudo experiment on average. The average total background fraction was assumed

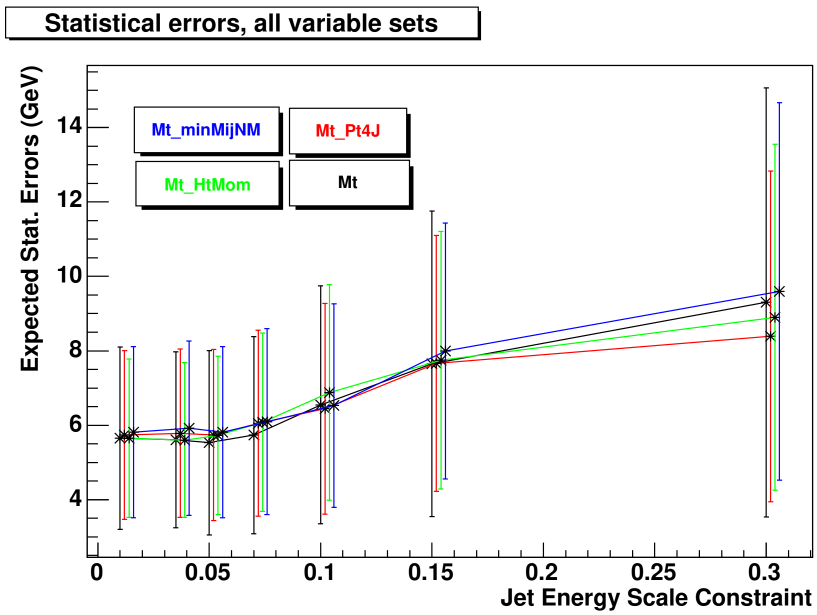


Figure 29: Expected top mass statistical uncertainty. Vertical bars show the range of expected statistical uncertainty from one pseudo experiment.

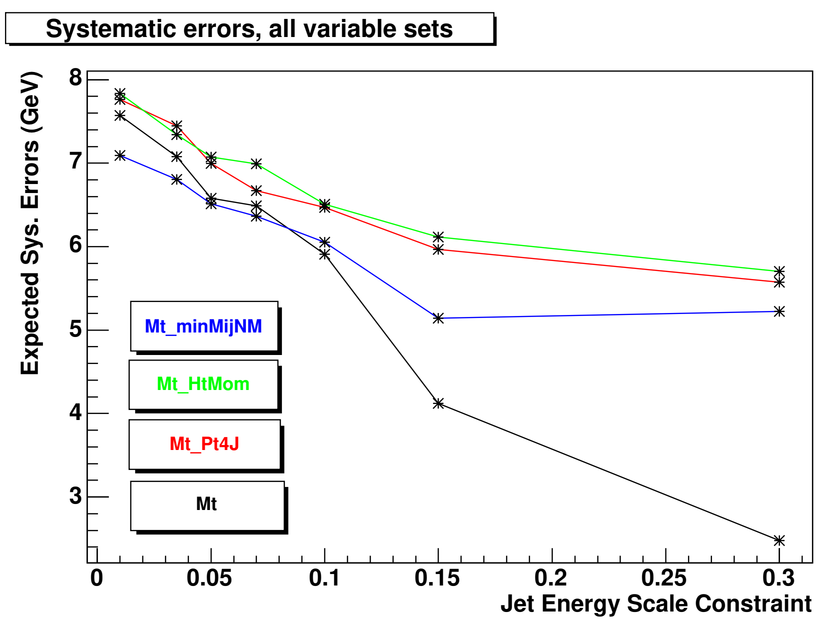


Figure 30: Expected top mass systematic uncertainty from jet reconstruction. At high values of JES constraint, top mass alone results in a smaller systematic shift than other variable sets because it is less sensitive to jet transverse energy cuts applied in the event selection procedure.

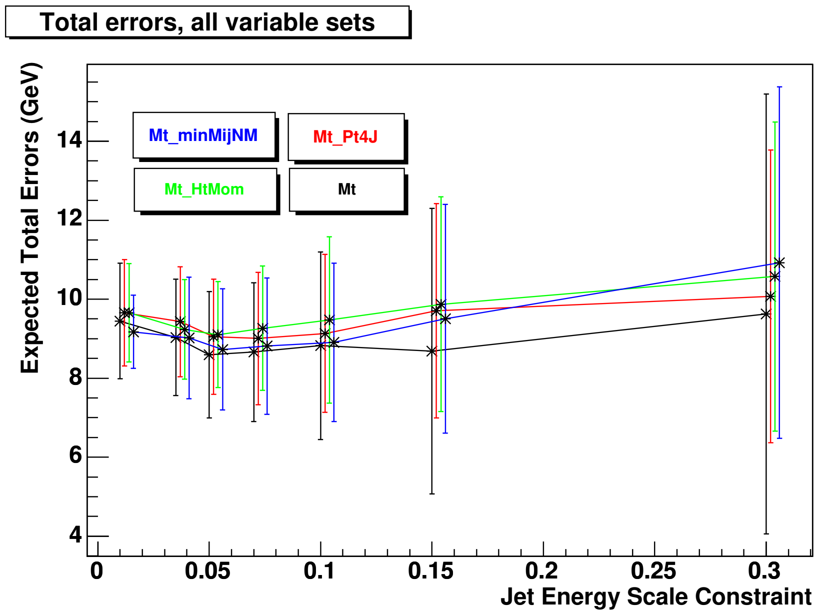


Figure 31: Total expected top mass reconstruction error. Vertical bars show the range of expected error from one pseudo experiment.

to be 0.335 The numbers of events in all subsamples (signal, $Wb\bar{b}$, mistag, and QCD) were allowed to fluctuate around their expected values according to the Poisson distribution. The pull distributions from these pseudo experiments are shown in Fig. 33. We find that the pull width is, on average, equal to 1.10, as illustrated in Figure 34. Therefore, we multiply the errors obtained from the likelihood curve by the pull width, and obtain the result $m_t = 179.6^{+6.4}_{-6.3}$ GeV.[7] The pull distributions exhibit small mass-dependent biases on the order of 1 GeV. We assign a systematic error to our measurement due to these biases as discussed in Section 8.2.

The plot of expected errors from pseudo experiments is shown in Fig. 35.

In addition to the best mass value, we can extract the most probable background fraction. The fractions returned by the fitter are shown in Fig. 36. By interpolating the background fraction between the two nearest template points we obtain the most probable background fraction value of 0.34 ± 0.14.

Using a similar procedure, we also estimate the top mass value for the ≥ 4 jets sample as a crosscheck. Likelihood, pull width, and fitted background fraction for this sample are shown on Figures 37, 38, and 39, respectively. Figure 40 illustrates that the JES constraint value of 0.07 is also optimal for the 4-jets sample, and that the jet energy systematic error for this sample is found to be 6.8 GeV. From this sample we obtain the result $m_t = 179.9^{+5.5}_{-5.3}$ GeV. The plot of expected errors from pseudo experiments is shown

[7] The error scale factor determined by requiring 68.27% coverage of the generated mass by the interval estimate in pseudo experiments is 1.106. It results in the same value of the statistical error after rounding to two significant digits.

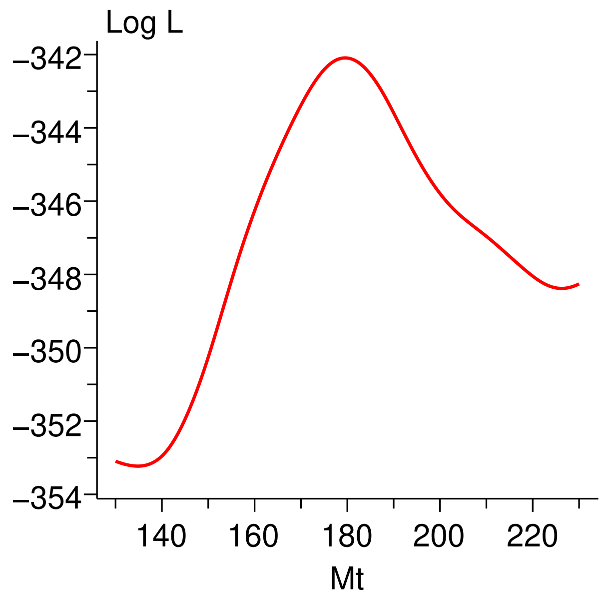


Figure 32: Top mass likelihood from data in the $\geq$ 3.5 jets sample.

in Fig. 41.

8.2 Systematic Uncertainties

The largest systematic error on the top mass is due to the uncertainty in the jet energy determination. This error was estimated in the course of the optimization studies, as discussed in Section 7.4 and illustrated in Fig. 30.

Using the pseudo-experiments technique described earlier, we evaluate the contributions to the top mass systematic uncertainty from generator dependence, initial and final state radiation, and parton distribution function. These are shown in Table 12 for $m_t = 175$ GeV. The PDF systematic uncertainty was estimated using the MC reweighting technique described in [25]. The shifts which correspond to the 20 CTEQ6M eigenvectors are presented in Fig. 42.

Another source of systematic uncertainties is due to the uncertainty on the background composition. We use three background templates whose default contributions are shown in Table 5. We vary each of the backgrounds fractions by $\pm 1\sigma$ and obtain the results shown in Table 13. We add largest systematic shifts for each template in quadrature and obtain 0.3 GeV systematic error due to background composition.

The uncertainty due to variations in the Q^2 used in the background simulation is estimated using $Wb\bar{b}$ samples generated with four different Q^2 values. These samples were used as a source of $Wb\bar{b}$ background events in pseudo experiments, while default $Wb\bar{b}$ templates were used in the reconstruction procedure. The median reconstructed top masses are summarized in Table 14. We take the largest difference of 0.2 GeV between observed

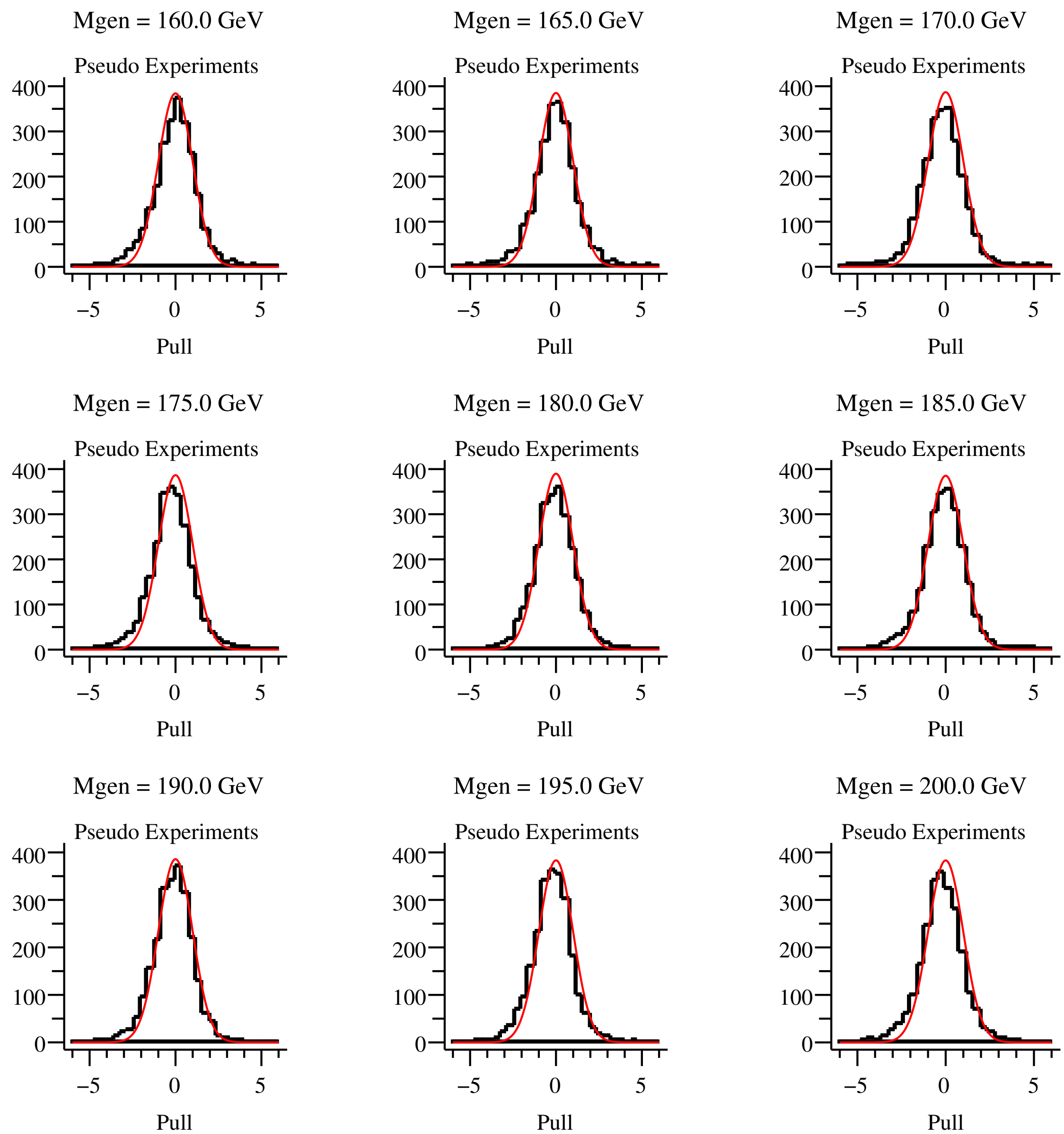


Figure 33: Pull distributions for the ≥ 3.5 jets sample. A Gaussian with mean 0, standard deviation 1, and equivalent area is plotted on top of each pull.

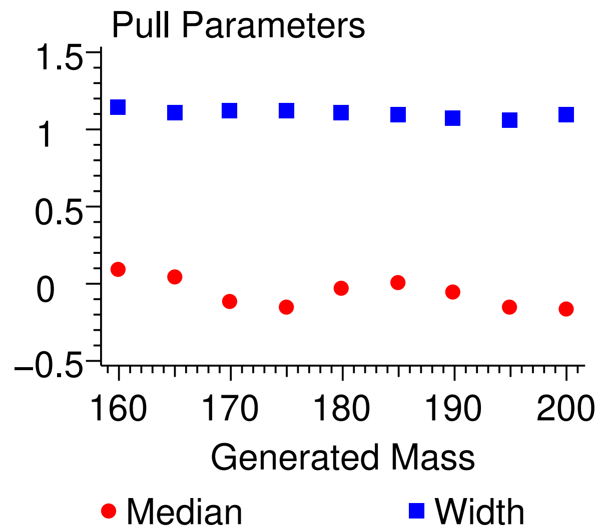


Figure 34: Parameters of the reconstructed pull distributions in the ≥ 3.5 jets sample. Width is defined as half of the distance between 16^{th} and 84^{th} percentiles.

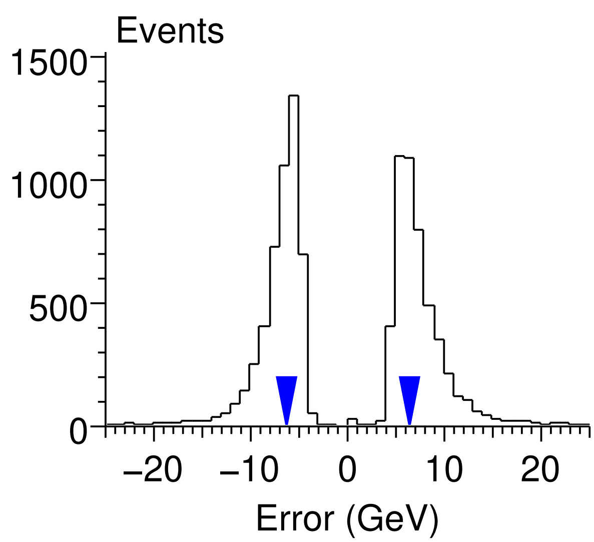


Figure 35: Distribution of the expected top mass statistical errors from pseudo experiments in the ≥ 3.5 jets sample. The horizontal scale has been widened according to the pull width. Blue triangles show the observed error values.

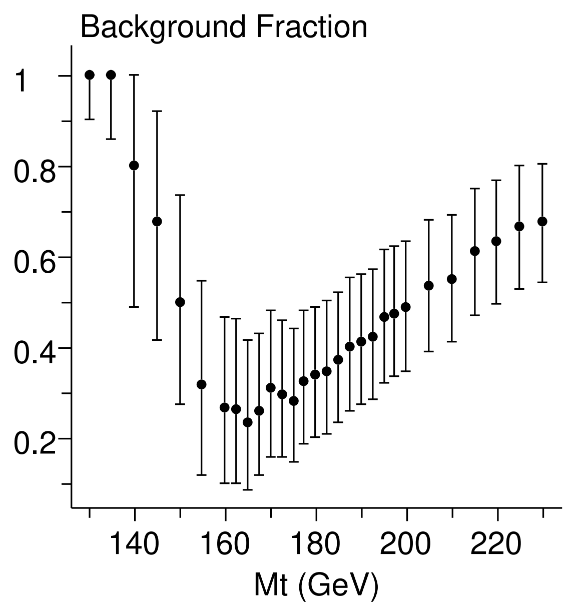


Figure 36: Most probable background fractions for all template masses in the ≥ 3.5 jets sample.

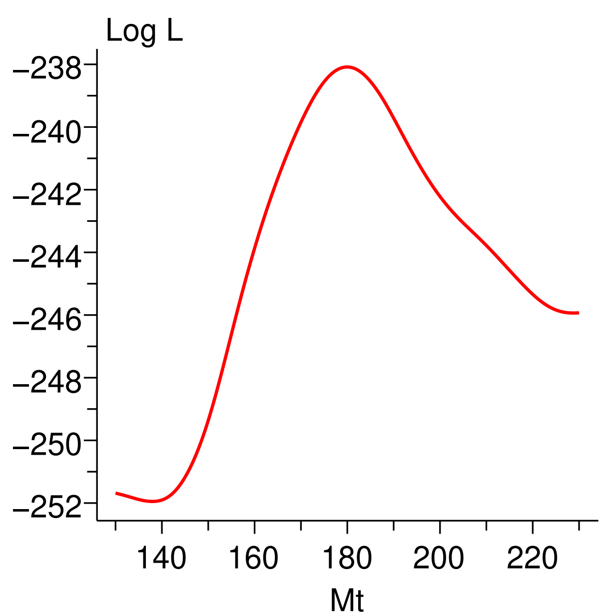


Figure 37: Top mass likelihood from data using the 4-jets sample. This likelihood results in $179.9^{+5.3}_{-5.1}$ GeV measured top mass value (not scaled for pulls).

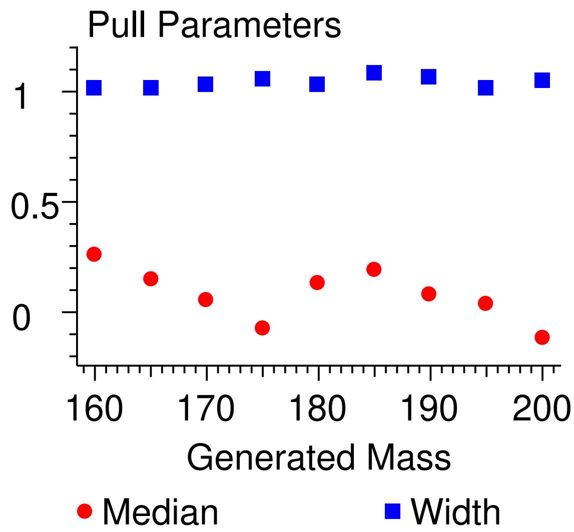


Figure 38: Pull distributions for the 4-jets sample. For these pseudo experiments we have assumed that there is no background. Average symmetrized pull width is 1.04.

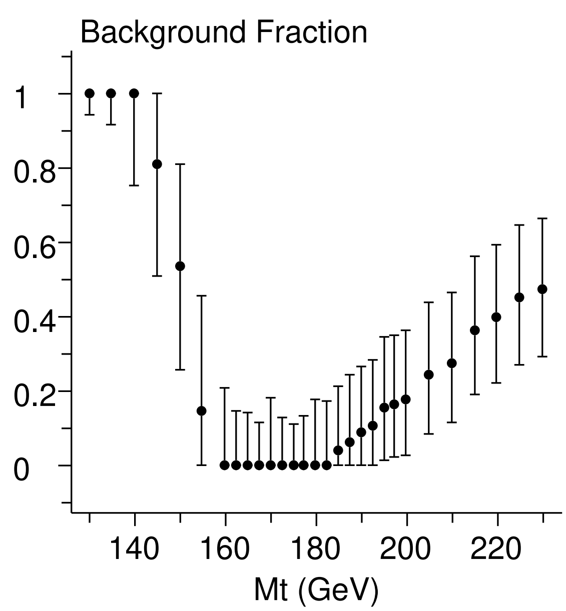


Figure 39: Most probable background fractions for all template masses in the 4-jets sample.

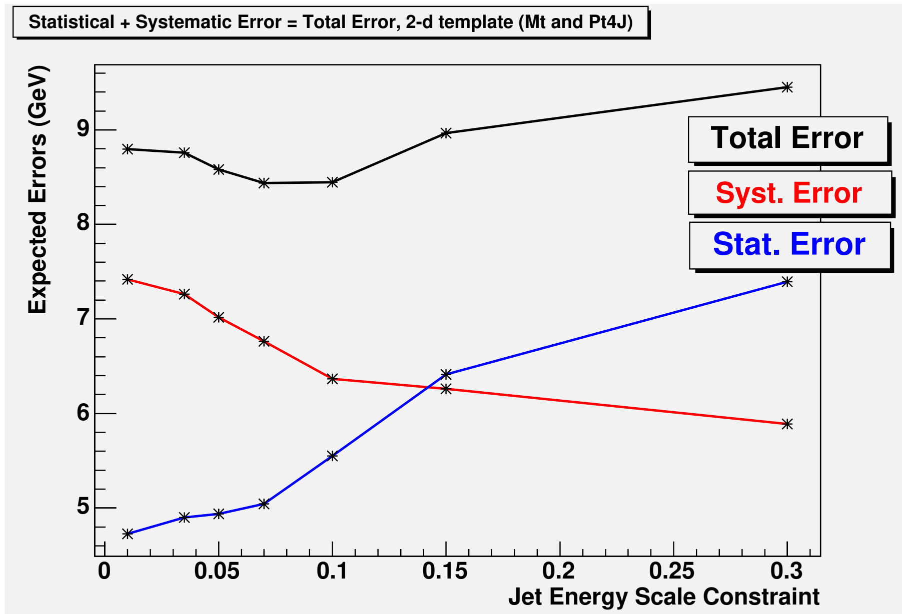


Figure 40: Dependence of the total expected error on the jet energy scale constraint in the 4-jets sample. For these pseudo experiments we have assumed that there is no background.

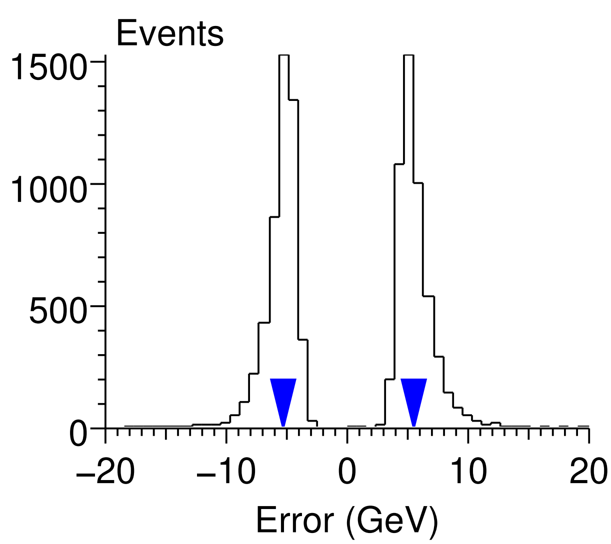


Figure 41: Distribution of the expected top mass statistical errors from pseudo experiments in the 4-jets sample. The horizontal scale has been widened according to the pull width. Blue triangles show the observed error values.

Table 12: Additional systematics for the top mass. In all cases the ΔM_{top} is calculated from the value obtained with PYTHIA tune A sample. The "Is used" column tells whether this particular shift was used to calculate the final systematic error.

Systematic	How done	M_{top} (GeV)	ΔM_{top} (GeV)	Is used
Generator dependence	PYTHIA tune A (default)	174.5 ± 0.2		
	HERWIG (default)	174.4 ± 0.3	-0.1	yes
	HERWIG (no spin correlations)	174.7 ± 0.2	$+0.2$	yes
	PYTHIA Run 1	173.9 ± 0.2	-0.6	no
ISR	PYTHIA tune B	174.9 ± 0.3	$+0.4$	no
	PYTHIA no ISR	174.0 ± 0.3	-0.5	no
	Less ISR(Λ_{QCD}=100 MeV)	174.7 ± 0.3	$+0.2$	yes
	More ISR(Λ_{QCD}=384 MeV)	174.5 ± 0.3	0.0	yes
FSR	Less FSR	174.9 ± 0.3	$+0.4$	yes
	More FSR	175.1 ± 0.3	$+0.6$	yes
	No hard gluon from b quark	173.5 ± 0.3	-1.0	no
PDF	MRST72	174.7 ± 0.3	$+0.2$	yes
	MRST75	174.8 ± 0.3	$+0.3$	no
	MRST72-MRST75	n.a.	-0.1	yes
	Variation of PDF eigenvectors	n.a.	0.3	yes

Table 13: Systematics due to variations in background fractions

Background Template	Shift Direction (σ)	ΔM_{top} (GeV)
$Wb\bar{b}$	$+1$	$+0.1$
	-1	0.0
mistags	$+1$	0.0
	-1	$+0.2$
QCD	$+1$	$+0.2$
	-1	$+0.1$

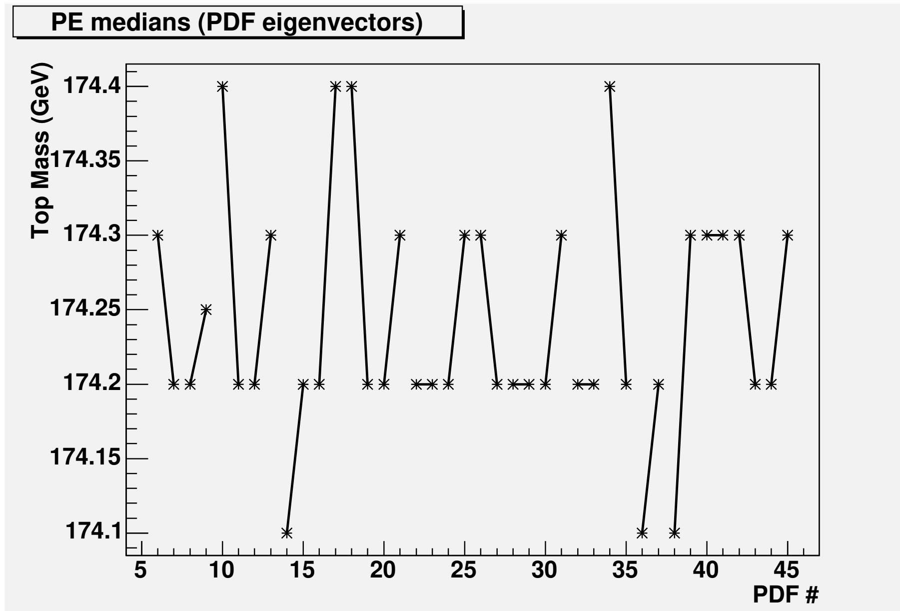


Figure 42: Systematic shifts for CTEQ6M PDF eigenvectors in the ≥ 3.5 jets sample.

medians as the background Q^2 systematic uncertainty. Combined with the uncertainty on background composition, this gives 0.4 GeV systematic error due to background shape.

Table 14: Systematics due to variations in Q^2 used to simulate $Wb\bar{b}$.

Q^2	M_{top} (GeV)
$(2M_W)^2$	174.1 ± 0.3
M_W^2	173.9 ± 0.3
$(M_W/2)^2$	174.0 ± 0.3
$M_W^2 + (p_T^W)^2$	173.9 ± 0.3

Our analysis does not use the b tag scale factor (that is, nothing special is done to the ratio of double/single b tagged events for both $t\bar{t}$ signal and $Wb\bar{b}$ background templates obtained from MC). To check that this does not affect our result, we have performed a set of pseudo experiments in which the probability to pick an event with double b tag (either signal or $Wb\bar{b}$) was reduced by factors 0.74, 0.81, and 0.88, according to the b tag scale factor measurement of 0.81 ± 0.07 from Ref [26]. The largest shift of +0.3 GeV in the median reconstructed mass was observed for the scale factor value of 0.74. We use the magnitude of this shift as the b tagging systematic error.

We assign 0.7 GeV systematic error due to the presence of mass-dependent bias in our fitting procedure. This number is obtained as the RMS bias for generated mass values between 160 and 200 GeV (with 5 GeV step), as illustrated in Fig. 43.

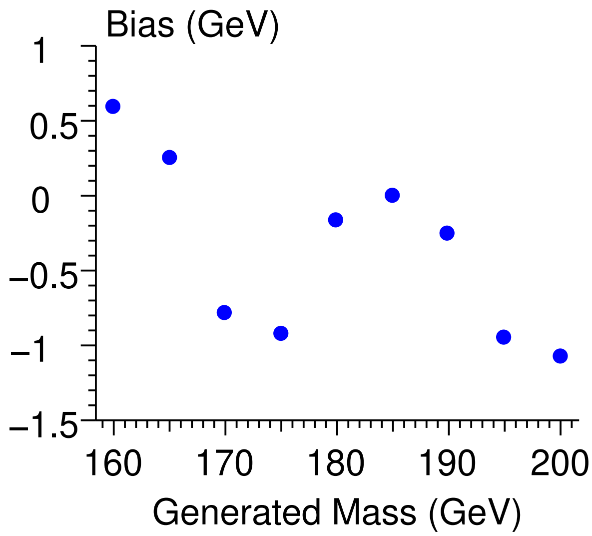


Figure 43: Reconstruction bias values observed in pseudo experiments in the ≥ 3.5 jets sample.

We use the largest systematics from each category in Table 12 and obtain the final values shown in Table 15. We assume that all systematic errors except the jet energy scale are similar for the 3.5 and 4 jet samples, and use systematic errors determined for the 3.5 jet sample in both cases.

9 Conclusions

The top quark mass measured by the multivariate templates method in the ≥ 3.5 jets sample is $179.6^{+6.4}_{-6.3} \pm 6.8$ GeV. This is the first CDF top mass measurement with freely floating background fraction. Our method allows for optimal trading of statistical and systematic errors, and will be especially powerful as the CDF accumulated luminosity increases.

References

[1] T. Affolder et al. *Measurement of the Top Quark Mass with the Collider Detector at Fermilab*, Phys. Rev. D, **63**, 032003 (2001).

[2] http://www-cdf.fnal.gov/internal/physics/top/RunIIWjets/

[3] E. J. Thomson et al. *Top and ElectroWeak Event Classification Module for CDF Run II*, CDF-5947 (October 22, 2003).

Table 15: Top mass systematic errors.

Systematic	ΔM_{top} (GeV)
Jet Energy	6.7 (3.5 jets) / 6.8 (4 jets)
Generators	0.2
ISR	0.2
FSR	0.6
PDF	0.3
Background Shape	0.4
b Tagging	0.3
Fitting Procedure	0.7
Total	6.8 (3.5 jets) / 6.9 (4 jets)

[4] H. Bachacou, J. Nielsen and W-M Yao *Optimized Measurement of the $t\bar{t}$ Cross Section in the SECVTX Tagged Sample*, CDF-6902 (March 2004).

[5] T. Maruyama and M. Shochet *Updated Results from the Determination of the Background Fraction in the b-tagged l+jets $t\bar{t}$ channel using Jet E_T Variables*, CDF-6883, March 2004.

[6] E. Brubaker, A. Gibson, Y-K Kim, Un-Ki Yang, J-F Arguin, P. Sinervo *Updated Measurement of the Top Quark Mass in the lepton+Jets channel*, CDF-6854 (January 2004).

[7] K. Yorita, K. Kondo, *Top quark mass measurement using the DLM method in the lepton + jets channel with SECVTX*, CDF-6929, March 16, 2004.

[8] SECVTX criteria for a b-tag are taken from the working group WEB page: http://www-cdf.fnal.gov/internal/physics/top/RunIIBtag/Offline_Defaults.html

[9] J.-F. Arguin, A. Bhatti, C. Currat, A. Gibson, A. Galtieri, B. Heinemann, G. Latino, and M. Shochet *Jet Corrections for Summer 2003 analyses (offline V4.10.4)*, CDF-6565 (July 2003).

[10] J.-F. Arguin and P. Sinervo. *Revisiting the top-specific jet corrections*, CDF-6404 (December 2003).

[11] L. Galtieri and J. Lys, *The SQUAW Top Mass Fitter*, CDF-3184 (December 1995).

[12] Paul Avery, *Fitting Theory Writeups and References.* http://www.phys.ufl.edu/~avery/fitting.html

[13] J.-F. Arguin and P. Sinervo, *Prospects for using $W \to jj$ decays in $t\bar{t}$ events to constrain the jet energy scale*, talk presented at the Jan. 7 2004 Top Mass Group meeting.

[14] Simon Sabik and Pierre Savard, *Study of the Top Quark Mass χ^2 fitter*, CDF-6306 (February 2003).

[15] G. Mahlon and S. Parke, *Angular correlations in top quark pair production and decay at hadron colliders*, Physical Review D **53**, 4886 (1996).

[16] G. Mahlon and S. Parke, *Maximizing spin correlations in top quark pair production at the Tevatron*, Physics Letters B **411**, 173 (1997).

[17] The SLATEC Common Mathematical Library is a collection of Fortran subroutines for a wide variety of mathematical problems. It is available from http://www.netlib.org.

[18] P.J. Rousseeuw, K. Van Driessen, *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, Technometrics, **41**, 212.

[19] S. Kullback, *Information Theory and Statistics*, Dover Publications, New York, 1968; S. Kullback, R.A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics **22**, 79 (1951).

[20] L. Lee, *Measures of Distributional Similarity*, "37th Annual Meeting of the Association for Computational Linguistics", pp 25–32, 1999.

[21] J. Lin, *Divergence Measures Based on the Shannon Entropy*, IEEE Transactions on information theory, vol 37 no. 1 (1991) 145.

[22] D. Scott, *Multivariate Density Estimation : Theory, Practice, and Visualization* Wiley-Interscience, 1992.

[23] P. Huber, *Robust Estimation of a Location Parameter*, Annals of Mathematical Statistics **35**, 73 (1964).

[24] J. Marron and D. Nolan, *Canonical Kernels for Density Estimation*, Statist. Prob. Lett. **7**, 195 (1988).

[25] Stephen Miller, *How to estimate PDF acceptance uncertainties using Weighted MC*, http://cdfrh0.grid.umich.edu/~miller/pdf/pdf_acceptance.html, April 2004.

[26] H. Bachacou, W. Yao, and T. Wright, *Efficiency of SecVtx B-Tag Used for 2004 Winter Conferences*, CDF-6904 (March 25, 2004).