Title
Comparison of Cut-based and Matrix Element Method Results for Beyond Standard Model Quarks

Journal
Acta Physica Polonica B, 44(6)

ISSN
0587-4254 1509-5770

Authors
Akyazi, E.
Unel, G.
Yilmaz, M.

Publication Date
2013

DOI
10.5506/APhysPolB.44.1273

COMPARISON OF CUT-BASED AND MATRIX ELEMENT METHOD RESULTS FOR BEYOND STANDARD MODEL QUARKS

E. Akyazi[a†], G. Unel[b‡], M. Yilmaz[a§]

[a]Gazi University, Physics Department, Institute of Science and Technology Ankara, Turkey
[b]University of California at Irvine, Physics Department, USA

(Received October 12, 2012; revised version received March 20, 2013; final version received May 14, 2013)

In this work, two different methods for extracting the mass of a new quark from the (pseudo) data are compared: the classical cut-based method and the Matrix Element Method. As a concrete example, a fourth family up type quark is searched in p–p collisions of 7 TeV center-of-mass energy. We have shown that even with a very small number of events, Matrix Element Method gives better estimations for the mass value and its error. Especially, for event samples in which Signal-to-Background ratio is greater than 0.2, Matrix Element Method reduces the statistical error approximately ten times.

DOI:10.5506/APhysPolB.44.1273
PACS numbers: 14.65.Jk, 12.15.Ff

1. Introduction

In searching for new phenomena at the particle physics experiments, it is very important to extract the values of the unknown parameters with maximal statistical significance from small data samples. At this point, Matrix Element Method (MEM) provides a very powerful tool which gave the most precise value for top quark mass at Tevatron experiments DØ and CDF [1–4]. After the method became more popular, it has also been applied to other analysis such as electroweak single top quark production [5], estimation of the longitudinal W boson helicity fraction in top quark decays [6] and searches for the Higgs boson [7]. It can be applied to any mass analysis which includes exclusive decay channels at hadron colliders for BSM

† akyazi@cern.ch
‡ Gokhan.Unel@cern.ch
§ metiny@gazi.edu.tr

researches. In this paper, a brief description of this method is followed by a comparison of the results of a heavy quark search analysis using a traditional cut-based method, to those from the Matrix Element Method.

1.1. Matrix Element Method

The name Matrix Element Method comes from the fact that probability function, which is used in this method, is driven by the physical matrix element. The matrix Element Method uses both theoretical and experimental information to extract the values of any unknown parameters from the experimental data. Therefore, the essential point of the MEM is that it maximally uses the information contained in the physics of the problem, without trying to extract it from the distributions, as in the case of cut and count method. In this technique, each experimentally measured quantity is associated to a Bayesian probability function $P(x|\alpha)$ which gives the probability to observe this event in a certain theoretical frame α. The probability weight, which is based on square matrix element [8, 9], can be written in the following form [10, 11]

$$P(x|\alpha) = \frac{1}{\sigma} \int d\phi(y) |M|^2 dw_1 dw_2 f_1(w_1) f_2(w_2) W(x,y)\,, \tag{1.1}$$

where x is a set of detector-level kinematic quantities, y is the parton-level 4-vectors, σ is the parton level cross section ($1/\sigma$ factor ensures the normalization of probability), M is the matrix element describing the production and decay process, $f_1(w_1)$ and $f_2(w_2)$ are parton distribution functions, $d\phi(y)$ is phase-space element, and $W(x,y)$ is the transfer function or resolution function which describes the probability density to reconstruct an assumed partonic final state y as a measurement x in the detector.

The probability is derived by integrating over all possible parton states, and each configuration is weighted according to its probability to produce the observed measurement. The weights are then combined together into a likelihood to determine the most probable value of the parameter of interest (top quark mass, W helicity, *etc.*).

The likelihood function for N measured events can be written as

$$L(\alpha) = e^{-N \int \bar{P}(x,\alpha) dx} \prod_{i=1}^{N} \bar{P}(x_i,\alpha)\,, \tag{1.2}$$

where α is any parameter that we want to estimate and $\bar{P}(x_i,\alpha)$ is measured probability density. The derivation of likelihood can be found in [12]. The best value of α is obtained through maximization of the likelihood or, more practically, by minimizing $-\ln L(\alpha)$ with respect to α.

1.1.1. Transfer functions

The determination of transfer functions (TF) is the most important part of Matrix Element Method. As mentioned before, transfer functions map parton level quantities to detector level measured quantities or *vice versa*. The energy resolution of the leptons and the jets is parametrized with transfer functions $W(\Delta E = E_{\rm parton} - E_{\rm jet})$ and they give the probability for a measurement $E_{\rm jet}$ in the detector, if the true object energy is $E_{\rm parton}$. TFs can be decomposed into a product of functions for each external or internal particle, and each part can be handled separately. Although there are different types of TFs that can be found in various analysis, the most used one for jets is Cannelli's double Gaussian formulation [13]: one Gaussian is for the symmetric peak, while the other accommodates the asymmetric tails of the ΔE distribution. In this formulation, jet transfer function is expressed to be a function only of the relative energy difference between the parton and the jet

$$W(\Delta E) = \frac{1}{\sqrt{2\pi}(a_2+a_3a_5)}\left(\exp\left(! - \frac{(\Delta E - a_1)^2}{2a_2^2}\right) + a_3 \exp\left(-\frac{(\Delta E - a_4)^2}{2a_5^2}\right)\right) , \tag{1.3}$$

where the energy dependence of these a_i parameters can be written in following form [14]

$$a_i = a_{i,0} + a_{i,1}\sqrt{E} + a_{i,2}E \,. \tag{1.4}$$

These parameters can be determined by minimizing a likelihood formed by measuring parton energy and matched jet energy in a Monte Carlo sample under consideration and they must be determined in different pseudorapidity regions of the calorimeter to account for resolution differences in the detector.

Theoretically, lepton energies and angles can be parametrized as a Gaussian but, in practice, they are assumed to be almost well-measured by a detector apparatus, so the TFs for lepton energies and all the particle angles can be parametrized by delta functions. This parametrization is also less time-consuming for computation of the weights because of the dimensional reduction it introduces.

2. Analysis

In this work, comparison of the Matrix Element Method and cut-based method for mass reconstruction analysis of fourth family up-type quark [15], u_4, at 7 TeV center-of-mass energy using event samples which include different Signal-to-Background (S/B) ratios has been presented. For simplicity, neither detector resolution effects nor systematic effects have not been considered in this study.

This analysis is based on Monte Carlo events generated with MadGraph/MadEvent [16] and processed through PYTHIA [17] for the parton-shower and hadronization. Finally, detector respons is simulated by PGS [18]. In this study, the mixing between fourth generation and the first SM family is assumed to be 100 percent. Therefore, the decay channel $u_4 \to W + d$ becomes the dominant one. As a signal, the pair production of up type fourth family quark, u_4, at a proton–proton collider at a center-of-mass energy of 7 TeV is considered. The full process for signal events can be written as

$$pp \to u_4\bar{u}_4 \to W^-W^+jj\,, \tag{2.1}$$

where j is a jet originating from a d quark or $\bar{d}$ quark and one W decays leptonically, whereas the other decays hadronically. For simplicity, electronic decay mode of the W is considered. Therefore, the signal is searched in the $4j + 1e +$ MET final state. As the dominant background sample, $t\bar{t}$ events in which the top quark pairs decay semi-leptonically has been taken under consideration. These background events are also produced with MG-ME/PYTHIA-PGS chain with CTEQ6L1 [19] as the PDF set.

The Monte Carlo events have been produced for three different mass values of u_4 quark: 400, 500 and 600 GeV. These events were required to contain the right number of jets and leptons in the final state (*i.e.* 4 jets and 1 electron for this study).

2.1. Cut-based analysis

In the cut-based analysis, leptonically decaying W bosons were reconstructed from the 4-momentum of the lepton and the missing transverse momentum. Assuming a massless neutrino and on-shell W mass, the z component of the neutrino, and its energy are obtained by solving these two equations with two unknowns. If the equations can be solved, the solution providing the smallest $|P_z|$ is selected. The rational behind this selection is to use the smallest estimated value, thus to reduce the error margin. If the equation set cannot be solved ($\Delta < 0$) then, the neutrino four momentum is formed using the collinearity approximation, *i.e.* by assuming the same η for the neutral and charged leptons and again a massless neutrino. Hadronically decaying W bosons were reconstructed using the 4-momentum of two soft jets in each event. The two relevant jets are selected by considering the pairing of all jets, and by selecting the pair which would minimize a χ^2 defined as

$$\chi^2 \equiv \frac{(M_{jj} - M_W)^2}{\sigma_W^2} + \frac{(M_{jjj} - M_{j\nu l})^2}{\sigma_Q^2}\,, \tag{2.2}$$

where M_{jj} is the reconstructed invariant mass from two jets, M_{jjj} is the reconstructed invariant mass from three jets, $M_{j\nu l}$ is reconstructed invariant mass from lepton, MET and jet, σ_W is decay width of W, σ_Q is decay width of new heavy quark. The W-jet association ambiguity is resolved by selecting the combination which yields in the smallest difference between the masses of the two reconstructed u_4 quarks in the same event. The u_4 invariant mass is obtained by taking the average of the hadronically and leptonically decaying u_4 quarks. In the generation step, standard kinematic selection criteria are applied as follows:

$$\begin{aligned} P_{\mathrm{T},e} &> 10 \text{ GeV}\,, \\ P_{\mathrm{T},j} &> 20 \text{ GeV}\,, \\ |\eta_e| &< 2.5\,, \\ \Delta R(e,j) &> 0.4\,, \\ \Delta R(j,j) &> 0.4\,, \\ |\eta_j| &< 5\,, \end{aligned} \tag{2.3}$$

where $P_{\mathrm{T},e}$ ($P_{\mathrm{T},j}$ is transverse momentum of electrons (jets), $|\eta_e|$ ($|\eta_j|$), is the rapidity for electrons (jets) and, $\Delta R\,(e,j)$ is the angular distance between electrons and jets and $\Delta R\,(j,j)$ is angular distance between jets, with $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

In the first reconstruction step, *a* u_4 quark of 500 GeV was used and u_4 invariant mass was extracted from a sample containing only 15 signal events. The reconstructed mass histogram for this case is shown in Fig. 1.

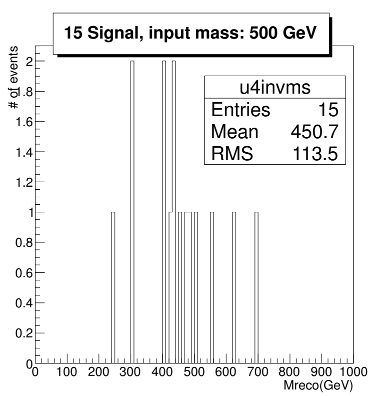

Fig. 1. Invariant mass histogram of u_4 with the cut-based method for an input test mass of 500 GeV. The result is extracted from a pure signal sample which contain only 15 events.

The same procedure has been applied to other samples containing different numbers of signal and background events. In short, the S/B ratio was scanned from a purely signal sample down to a purely background sample keeping the total number of events same, namely, 15. The cases which were scanned are: 13 signal (S) + 2 background (B), 11 S + 4 B, 9 S+ 6 B, 7 S + 8 B, 5 S +10 B, 3 S + 12 B, 15 B. Invariant mass histograms obtained for these cases are shown in Fig. 2.

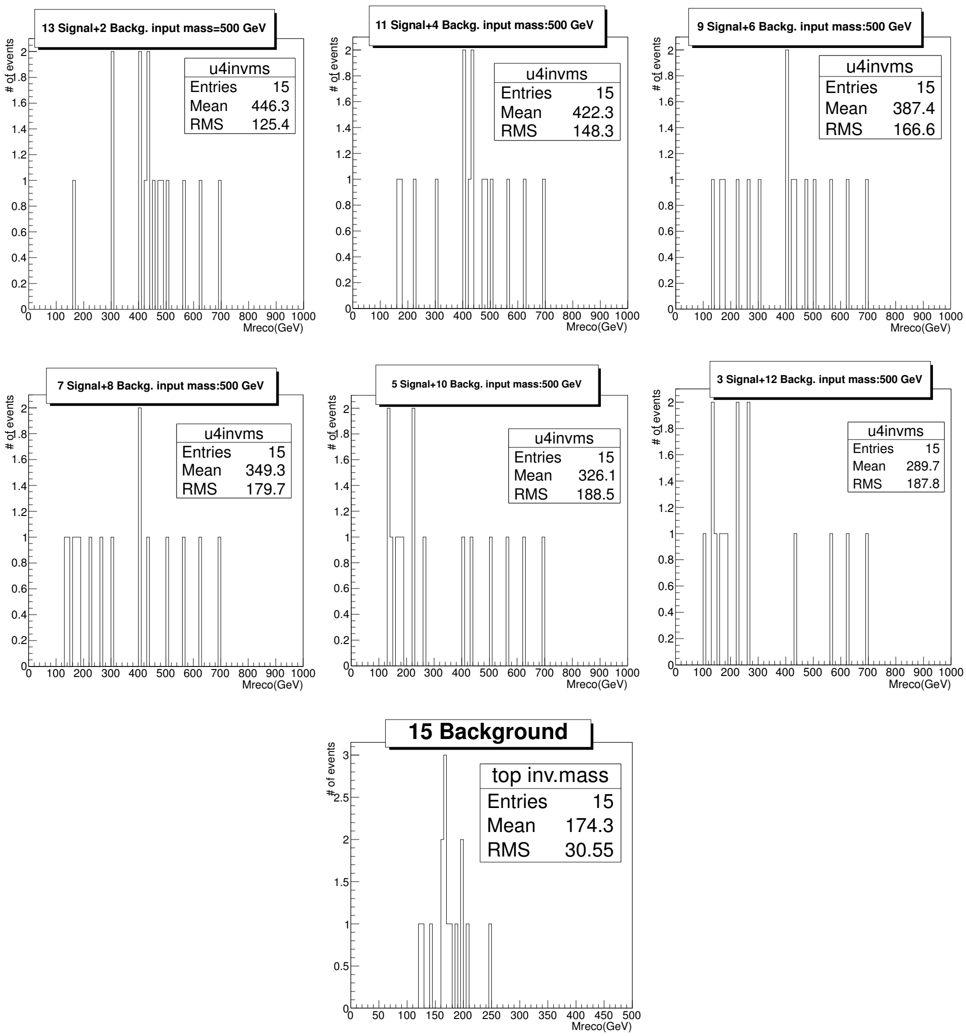

Fig. 2. Invariant mass histograms obtained from cut-based analysis for various event samples with decreasing S/B ratio and an equal signal mass of 500 GeV.

This procedure was also tested with other u_4 masses, namely 400 and 600 GeV. The reconstructed invariant mass histograms for these input masses are shown in Fig. 3 and Fig. 4.

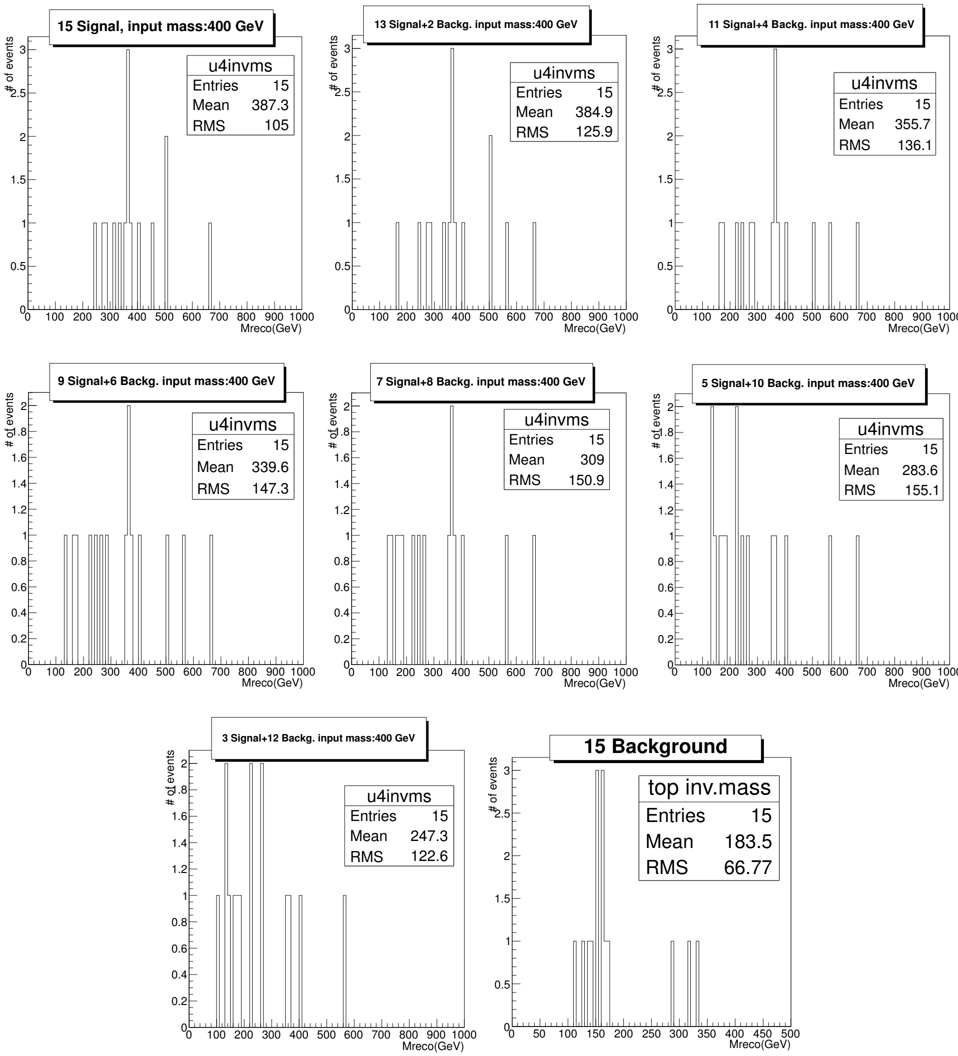

Fig. 3. The same as Fig. 2 but for $m_{u_4} = 400$ GeV.

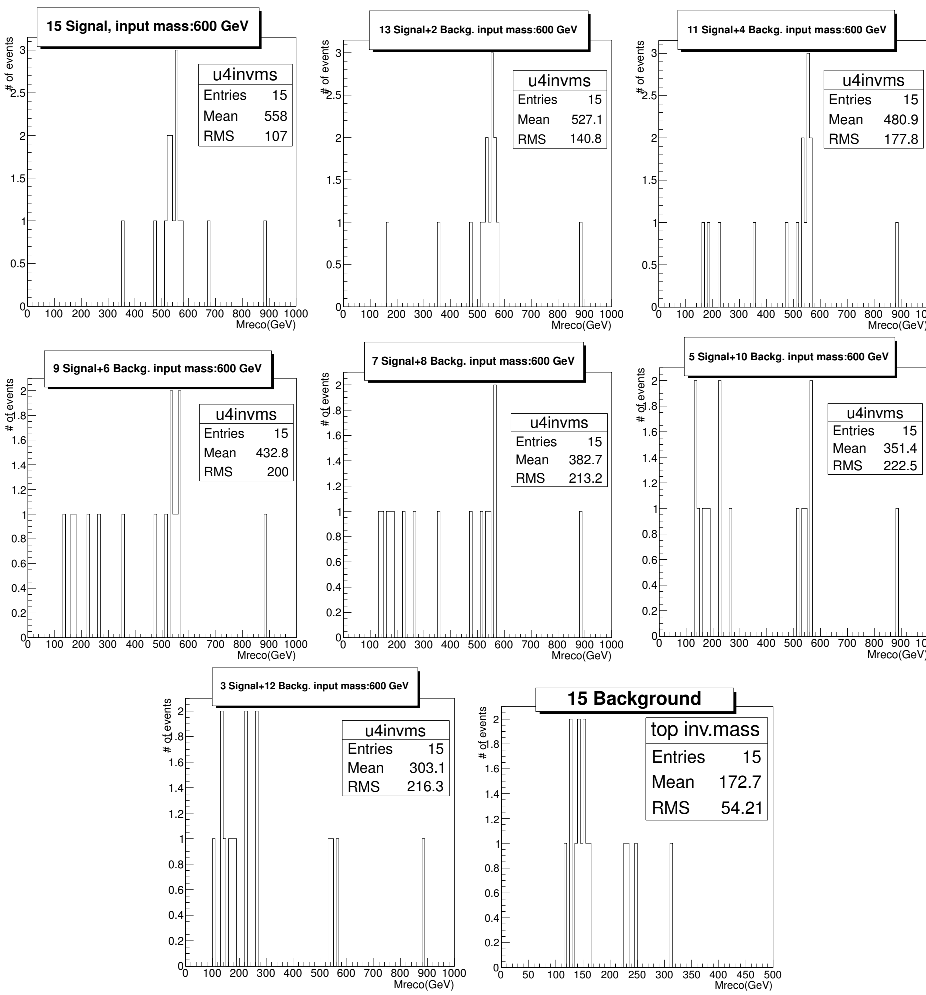

Fig. 4. The same as Fig. 2 but for $m_{u_4} = 600$ GeV.

The input masses and the reconstructed masses using the cut-based technique for the final states with different S/B ratios are shown in Table I.

The error values shown here indicate only statistical errors (systematics effects are not considered at this stage). One can see from Table I that even in the case of pure signal sample, the deviation from input values is large and the most correct result is obtained for 400 GeV input mass. These huge uncertainties seen here came from the low statistics. The second interesting point is that the samples including mostly background events also give new quark mass estimations around u_4 input mass instead of top mass, therefore, this approach is relatively useless for discriminating signal and background events especially with low statistics.

TABLE I

Invariant mass values extracted from the cut-based analysis for various samples which have different S/B ratios and different input values.

Event sample	Output u_4 masses for		
	input mass = 400 GeV	input mass = 500 GeV	input mass = 600 GeV
15 signal	387.3 ± 105.0	450.7 ± 113.5	558 ± 107
13 signal + 2 backg.	384.9 ± 125.9	446.3 ± 125.4	527.1 ± 140.8
11 signal + 4 backg.	355.7 ± 136.1	422.3 ± 148.3	480.9 ± 177.8
9 signal + 6 backg.	339.6 ± 147.3	387.4 ± 166.6	432.8 ± 200
7 signal + 8 backg.	309.0 ± 150.9	349.3 ± 179.7	382.7 ± 213.2
5 signal + 10 backg.	283.6 ± 155.1	326.1 ± 188.5	351.4 ± 222.5
3 signal + 12 backg.	247.3 ± 122.6	289.7 ± 187.8	303.1 ± 216.3
0 signal + 15 backg.	183.5 ± 66.77	174.3 ± 30.55	172.7 ± 54.21

2.2. Matrix Element Method analysis

This method relies on the correct calculation of the weights in Eq. (1.1). To ensure their correct computation, MadWeight [11], which was developed by the MadGraph Team, has been used. MadWeight is a phase space generator which takes lhco files [20] and processes information with data cards and returns likelihood values for the parameter of interest.

In this part, event files for 15 signal, 13 signal+2 background, 11 signal+ 4 background and so forth are used in MadWeight to estimate the signal mass for three input u_4 masses: 400, 500 and 600 GeV. A sample of $N = 15$ events are processed through MadWeight for the evaluation of the weights. The mass of the u_4 quark is extracted through the minimization of $-\ln(L(m_{u_4}))$ with respect to the m_{u_4}.

In this note, the default transfer function in MadWeight has been used. In the default TF set, the jet energy is parametrized by a double Gaussian, and all other quantities such as the angles of visible particles and the energy of leptons are assumed to be well measured. This means that the corresponding transfer functions for lepton energies and angles are given by delta functions. The transfer function associated with a neutrino (MET) is taken to be one.

As in the cut-base approach, the analysis started from event samples which were generated with an input mass of 500 GeV. The likelihood curves obtained for this mass with various signal and background samples are shown in Fig. 5.

Estimated u_4 masses and statistical errors are shown in the legend box of each graph, except the last one, *i.e.* 3S plot in which one finds 167.77 GeV. These estimations are extracted from a parabolic curve fit to $(-\ln L, \mathrm{Mass})$ points obtained from MadWeight. Error values include both standard deviation of likelihoods, evaluated via increasing the minimum likelihood value

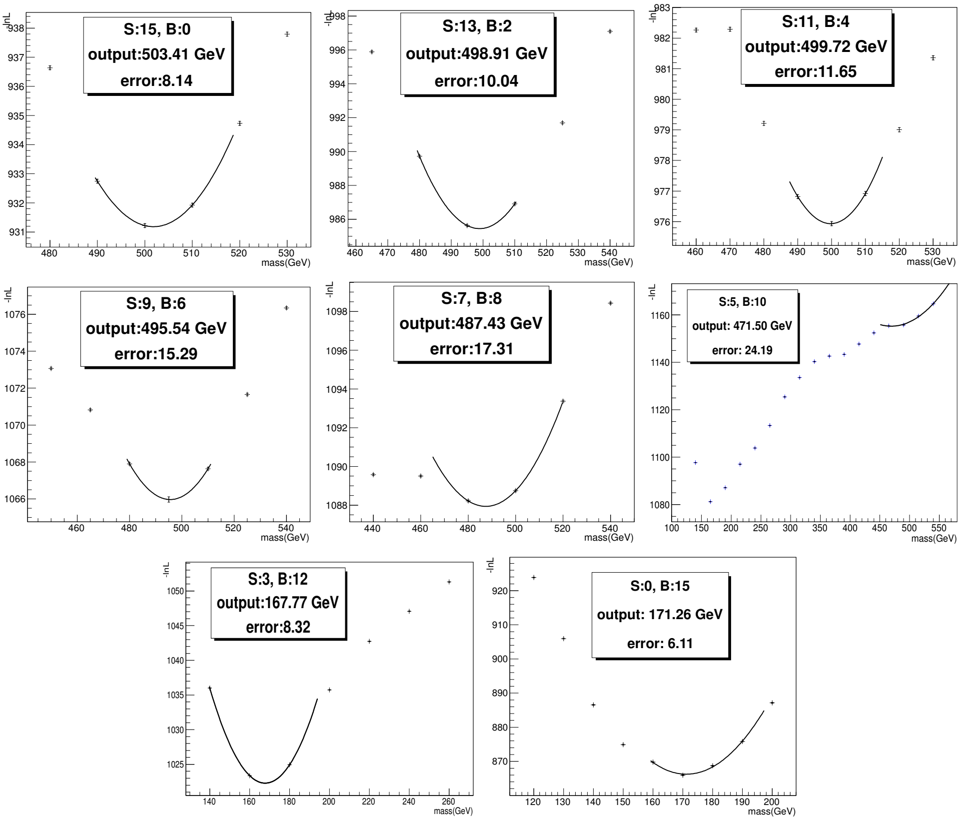

Fig. 5. Plots for likelihoods for samples of 15 events containing different ratios of S/B generated with input u_4 mass of 500 GeV. The mass value of u_4 has been extracted from the parabolic curve fitting of the points around the minima.

by 1/2, which corresponds to a 1σ deviation and also the errors originating from parabola fitting. If a wide mass range is scanned, then two likelihood minima are obtained (top, u_4) except the 3S12B case, where only one value corresponding to the top quark mass is found. As seen in 5S10B plot, there are two local minima between 350–550 GeV interval, and the nearest one to u_4 input mass is chosen.

The same procedure has been applied for event samples produced with input masses of 400 and 600 GeV. The resulting curves are shown in Figs. 6 and 7, respectively.

The input masses and the reconstructed masses with statistical errors using matrix element technique from the final state with different S/B ratios are shown in Table II.

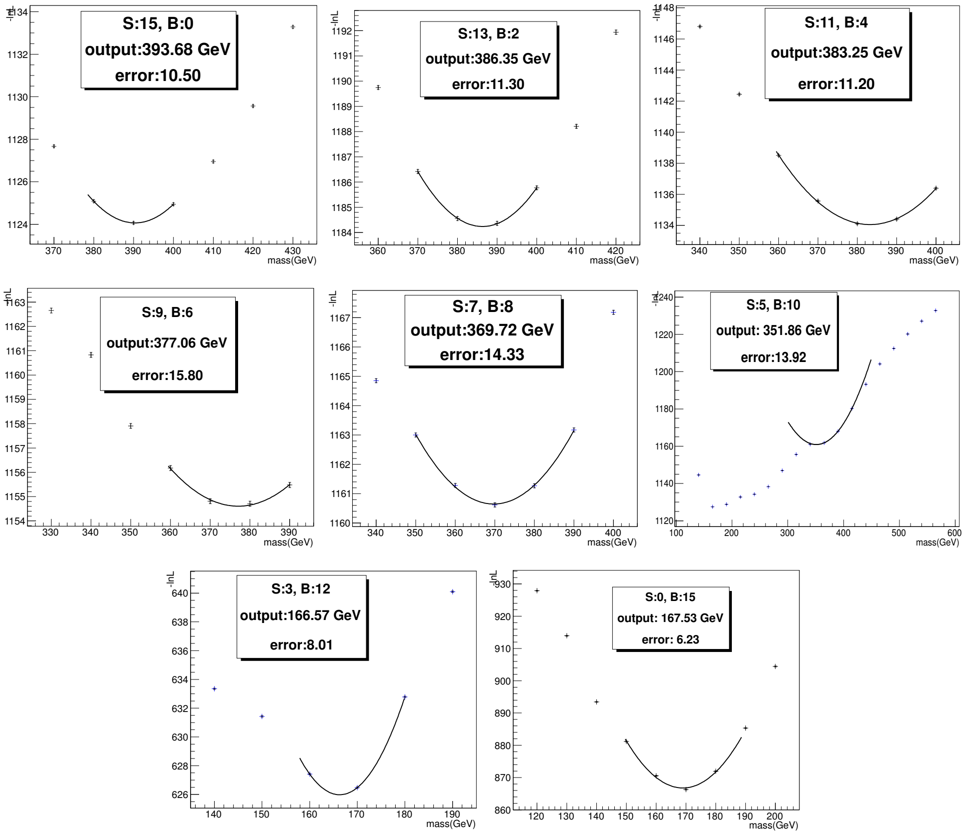

Fig. 6. The same as Fig. 5 but for $m_{u_4} = 400$ GeV.

TABLE II

Matrix Element analysis results obtained for various u_4 input masses and event samples which include various S/B ratios.

Event sample	Output u_4 masses for		
	input mass = 400 GeV	input mass = 500 GeV	input mass = 600 GeV
15 signal	393.68 ± 10.50	503.41 ± 8.14	621.05 ± 10.02
13 signal + 2 backg.	386.35 ± 11.30	498.91 ± 10.04	622.97 ± 12.46
11 signal + 4 backg.	383.25 ± 11.20	499.72 ± 11.65	617.15 ± 12.84
9 signal + 6 backg.	377.06 ± 15.80	495.54 ± 15.29	610.34 ± 13.80
7 signal + 8 backg.	369.72 ± 14.33	487.43 ± 17.31	608.88 ± 14.46
5 signal + 10 backg.	351.86 ± 13.92	471.50 ± 24.19	558.50 ± 18.07
3 signal + 12 backg.	166.57 ± 8.01	167.77 ± 8.32	168.30 ± 7.45
0 signal + 15 backg.	167.53 ± 6.23	171.26 ± 6.11	168.72 ± 7.73

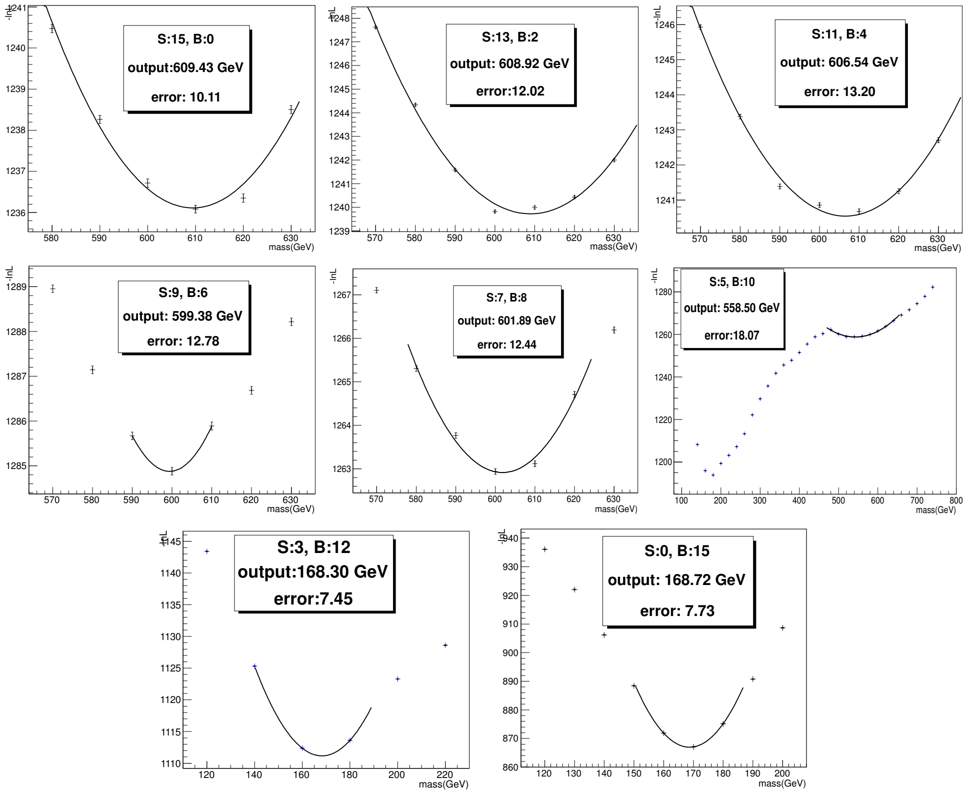

Fig. 7. The same as Fig. 5 but for $m_{u_4} = 600$ GeV.

By comparing Tables I and II, it can be clearly seen that, the MEM gives much smaller deviations from the input values for masses and errors compared to the cut-based analysis. In addition, as the number of background events increases, the resulting value approaches the top quark mass again as opposed to the cut-based results.

Furthermore, when the relative deviation from the true value (True Value–Reconstructed Value/True Value) is plotted against the Signal/Signal +Background ratio, one notices that, the deviations obtained from the Matrix Element Method are much smaller than the ones extracted from the cut-based analysis technique, especially for S/S+B values greater than 0.2.

As shown in Fig. 8, the Matrix Element Method becomes less accurate in the region of S/S+B $<$ 0.2.

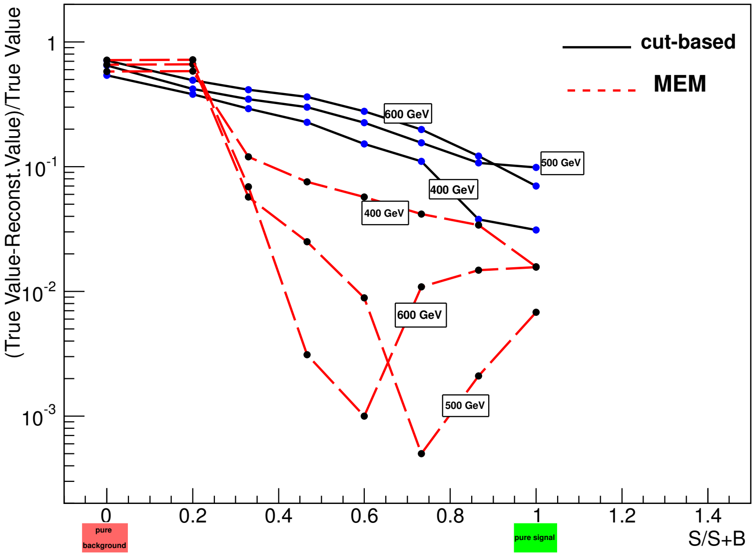

Fig. 8. S/S+B ratios *vs* corresponding relative errors for both cut-based and Matrix Element Method results for different input u_4 masses.

3. Conclusion

This study shows that for data samples containing small number of events with various signal-to-background ratios, the matrix element method gives essentially better values for the parameter of interest (mass of fourth family up type quark, in this analysis) and associated statistical errors. Error values obtained from the MEM are, on the average, ten times lower than cut-based results. As a second result, MEM is, also a powerful tool to discriminate signal and background events even with small statistical data if $S/S+B > 0.2$.

The authors would like to express their special thanks to Saleh Sultansoy for his very valuable comments and Olivier Mattelaer for helpful answers to MadWeight related issues. E.A. acknowledges the support from the Turkish Atomic Energy Authority.

REFERENCES

[1] A. Abulencia *et al.* [CDF Collaboration], *Phys. Rev.* **D75**, 031105 (2007) [arXiv:hep-ex/0612060].

[2] T. Aaltonen *et al.* [CDF Collaboration], *Phys. Rev. Lett.* **99**, 182002 (2007) [arXiv:hep-ex/0703045].

[3] V.M. Abazov *et al.* [DØ Collaboration], *Phys. Rev.* **D74**, 092005 (2006) [arXiv:hep-ex/0609053].

[4] V.M. Abazov *et al.* [DØ Collaboration], *Phys. Rev.* **D75**, 092001 (2007) [arXiv:hep-ex/0702018].

[5] V.M. Abazov *et al.*, *Phys. Rev. Lett.* **103**, 092001 (2009).

[6] V.M. Abazov *et al.*, *Phys. Lett.* **B617**, 1 (2005) [arXiv:hep-ex/0404040].

[7] T. Aaltonen *et al.*, *Phys. Rev. Lett.* **103**, 092002 (2009).

[8] K. Kondo, *J. Phys. Soc. Jpn.* **57**, 4126 (1988).

[9] R.H. Dalitz, G.R. Goldstein, *Phys. Rev.* **D45**, 1531 (1992).

[10] P. Artoisenet, O. Mattelaer, MadWeight: Automatic Event Reweighting with Matrix Elements, Prospects for Charged Higgs Discovery at Colliders, September 16–19, 2008, Upsala, Sweden.

[11] MadWeight, https://server06.fynu.ucl.ac.be/projects/madgraph/wiki/MatrixElement

[12] J.C.E. Vigil, Maximal Use of Kinematic Information for the Extraction of the Mass of the Top Quark in Single-lepton tt events at DØ, Univ. of Rochester, New York 2001.

[13] M.F. Canelli, Helicity of the W Boson in Single-lepton tt^{-} Events, FERMILAB-THESIS-2003-22, 2003.

[14] O. Mattelaer, A New Approach to Matrix Element Re-weighting, Université Catholique de Louvain, Centre for Cosmology, Particle Physics and Phenomenology, January 2011.

[15] B. Holdom *et al.*, *PMC Phys.* **A3**, 4 (2009).

[16] F. Maltoni, T. Stelzer, *J. High Energy Phys.* **0302**, 027 (2003) [arXiv:hep-ph/0208156].

[17] T. Sjostrand, S. Mrenna, P. Skands, *J. High Energy Phys.* **0605**, 026 (2006) [arXiv:hep-ph/0603175].

[18] J. Conway, http://physics.ucdavis.edu/∼conway/research/software/pgs/pgs4-general.htm

[19] J. Pumplin *et al.*, *J. High Energy Phys.* **07**, 012 (2002) [arXiv:hep-ex/0201195].

[20] LHCO wiki page, http://www.jthaler.net/olympicswiki/doku.php