CDF/ANAL/TOP/CDFR/8490
Version 0.2, September 7, 2006

Mistag Model Used for Single-Top Summer 2006 Analyses

Sarah Budd[b], Matthias Bühler[c], Catalin Ciobanu[b], Peter Dong[a], Thomas Junk[b], Jan Lück[c], Thomas Müller[c], Svenja Richter[c], Bernd Stelzer[a], Wolfgang Wagner[c], Rainer Wallny[a]

[a] *University of California, Los Angeles*
[b] *University of Illinois, Urbana-Champaign*
[c] *Universität Karlsruhe*

Abstract

Mistagged light events represent about 25% of the background events in the searches for single-top quark production. Therefore, it is important to model this background properly. We present a model based on taggable $W + 2p$ events to describe the shape of mistagged light events in the single-top analyses. Furthermore, we introduce a method to estimate the shape uncertainty introduced by this modeling.

Contents

1 Introduction

In searches for single-top quark production, about 25% of the background events are light events with a falsely reconstructed secondary vertex [1], so-called mistagged events. To describe this background correctly, a large amount of this kind of events is needed in simulation. Because of lack of statistics, it is necessary to create a model describing the expected shape of mistagged light events. This is particularly true for multivariate analyses.

2 Mistag Model

In order to describe mistagged light $W + 2p$ events, we use taggable events in the pretag $W + 2p$ sample (ltop2n, ltop2m, atop8t). In doing so, we veto heavy flavor matched events. In those pretag events, we assign taggable jets to be tagged. If an event has two taggable jets, we use both hypotheses for the particular event. Each hypothesis of each event is weighted by the prediction of the mistag matrix [2], taking into account the mistag asymmetry [3]. The weight w for each hypothesis is calculated by:

$$w = \alpha\beta \cdot negrate_{\text{tagged}} \tag{1}$$

with

- $\alpha\beta$: $\sum E_T$ dependent mistag asymmetry
- $negrate_{\text{tagged}}$: negative tag rate prediction of jet which was assigned to be tagged

3 Model Validation

3.1 Comparison to Negatively Tagged Data

To validate the mistag model described in section 2, we compare kinematic distributions of the model with tagged light $W + 2p$ simulated events and negatively tagged W+2jets data. Those comparisons can be seen in figures 1 to 3.

In figure 1 on the upper left-hand side, it is indicated that the E_T spectrum of the model is slightly harder than in negatively tagged data or tagged MC events.

As another cross check, we compared the E_T, η and H_T distributions in the 1 jet bin. Again, the E_T distribution of both tagged $W + 1p$ events and negatively tagged data are slightly softer than the model.

Nevertheless, the agreement between the different descriptions of mistagged events appears to be reasonable.

3.2 Comparison to Weighted Pretag Data

As an additional cross check, the model is compared to pretag data events. Those data events are weighted with the mistag matrix prediction as given by equation 1. In figures 5 to 7, the comparison of some kinematic distributions can be seen.

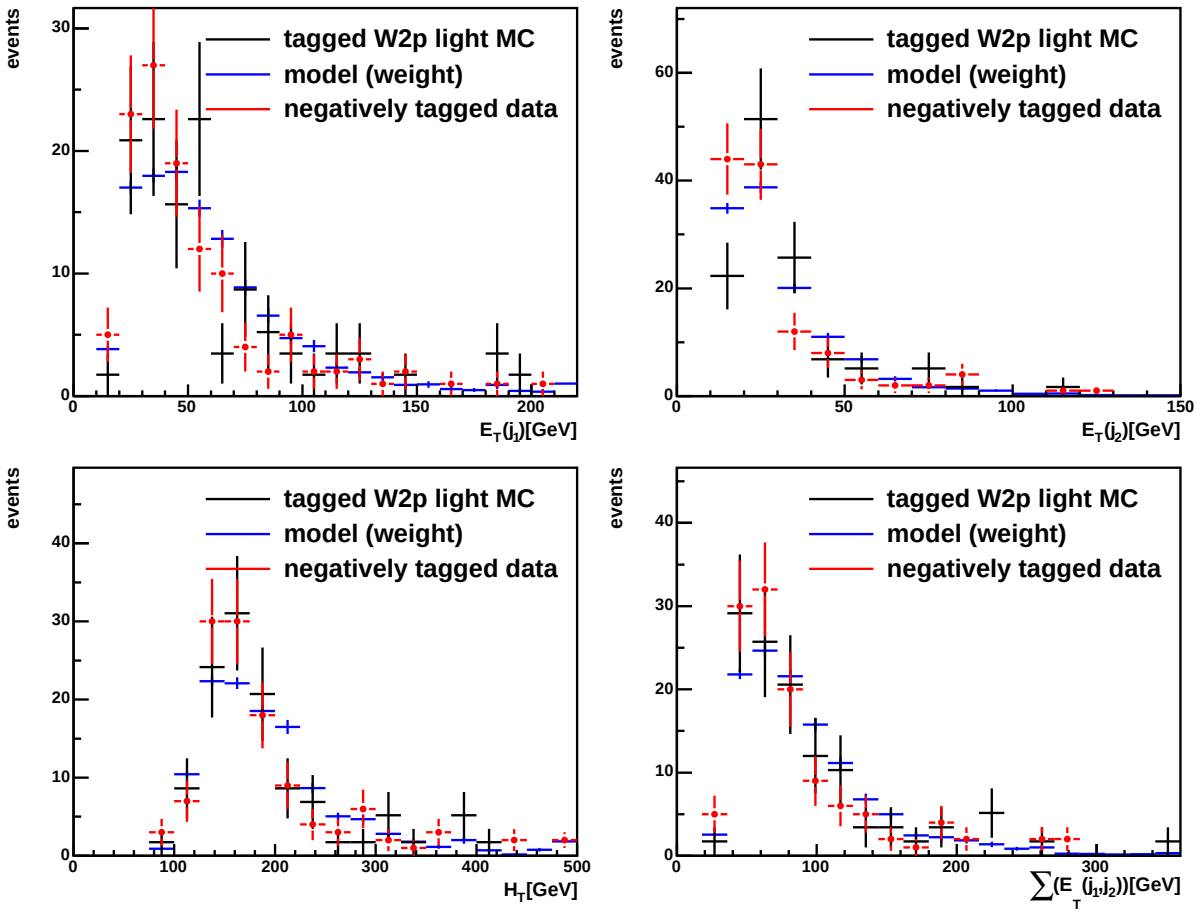

Figure 1: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and negatively tagged W + 2jets data. Upper left: E_T of leading jet (j1); upper right: E_T of second leading jet (j2); lower left: $H_T(s$-channel); lower right: $\sum(E_T)$(j1j2).

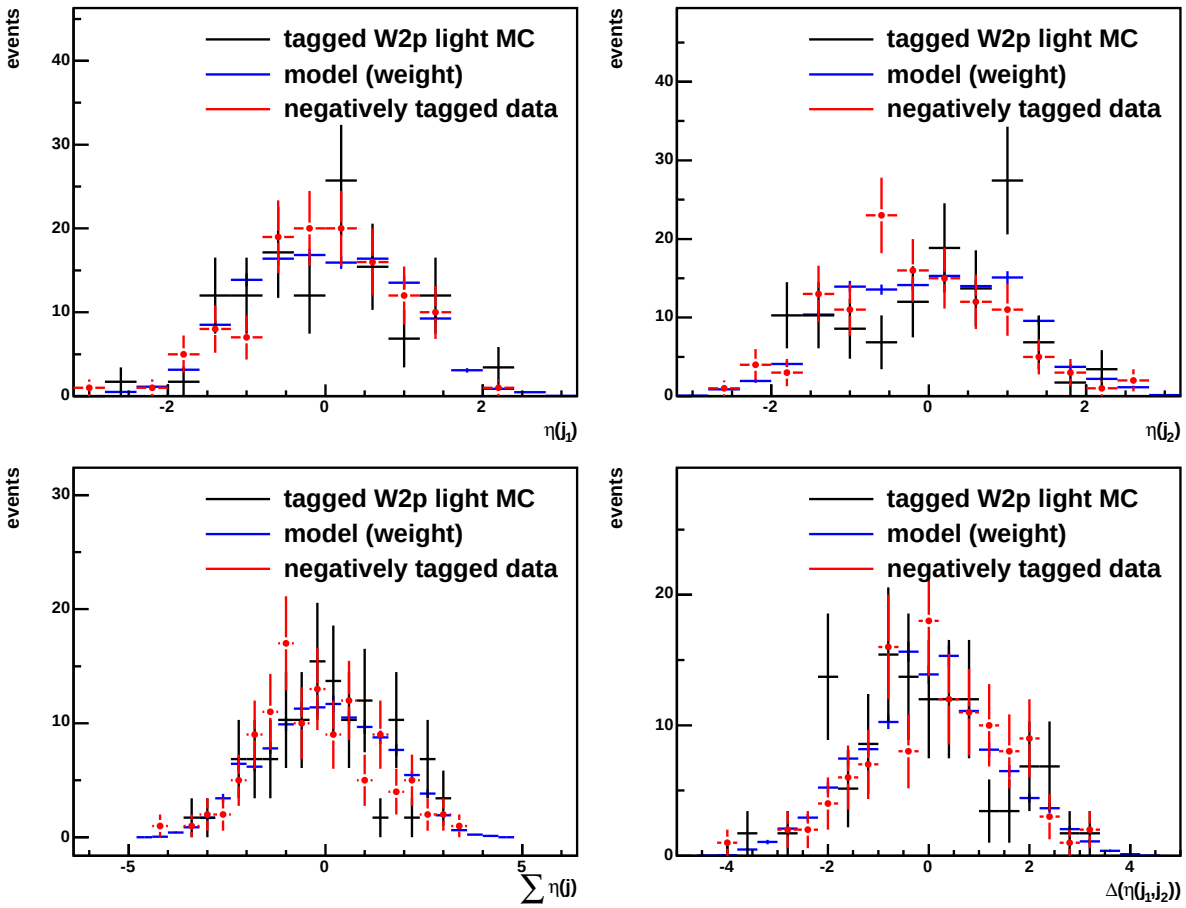

Figure 2: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and negatively tagged W + 2jets data. Upper left: η of leading jet (j1); upper right: η of second leading jet (j2); lower left: $\sum \eta$(j1j2); lower right: $\Delta\eta$(j1j2).

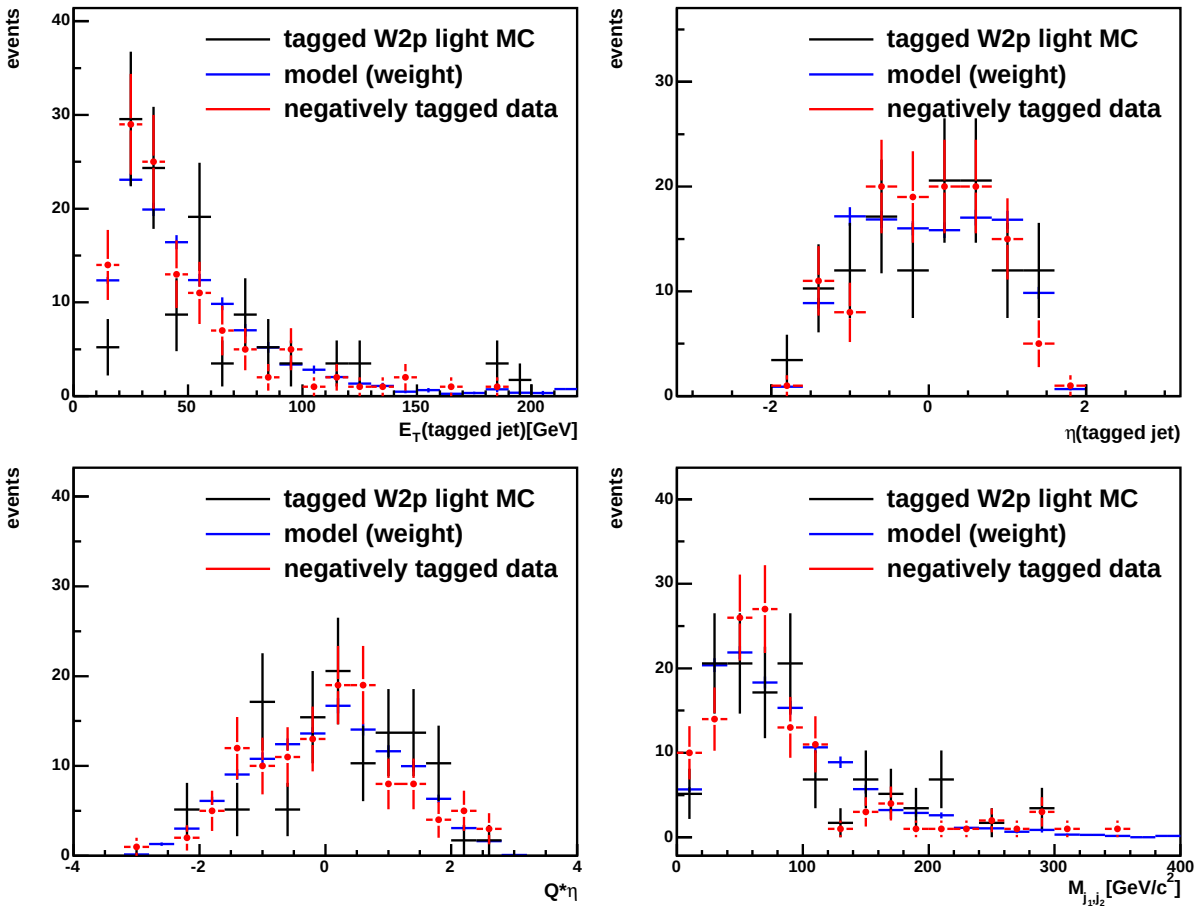

Figure 3: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and negatively tagged W + 2jets data. Upper left: E_T of tagged jet; upper right: η of tagged jet; lower left: $Q \times \eta$; lower right: dijetmass(j1j2).

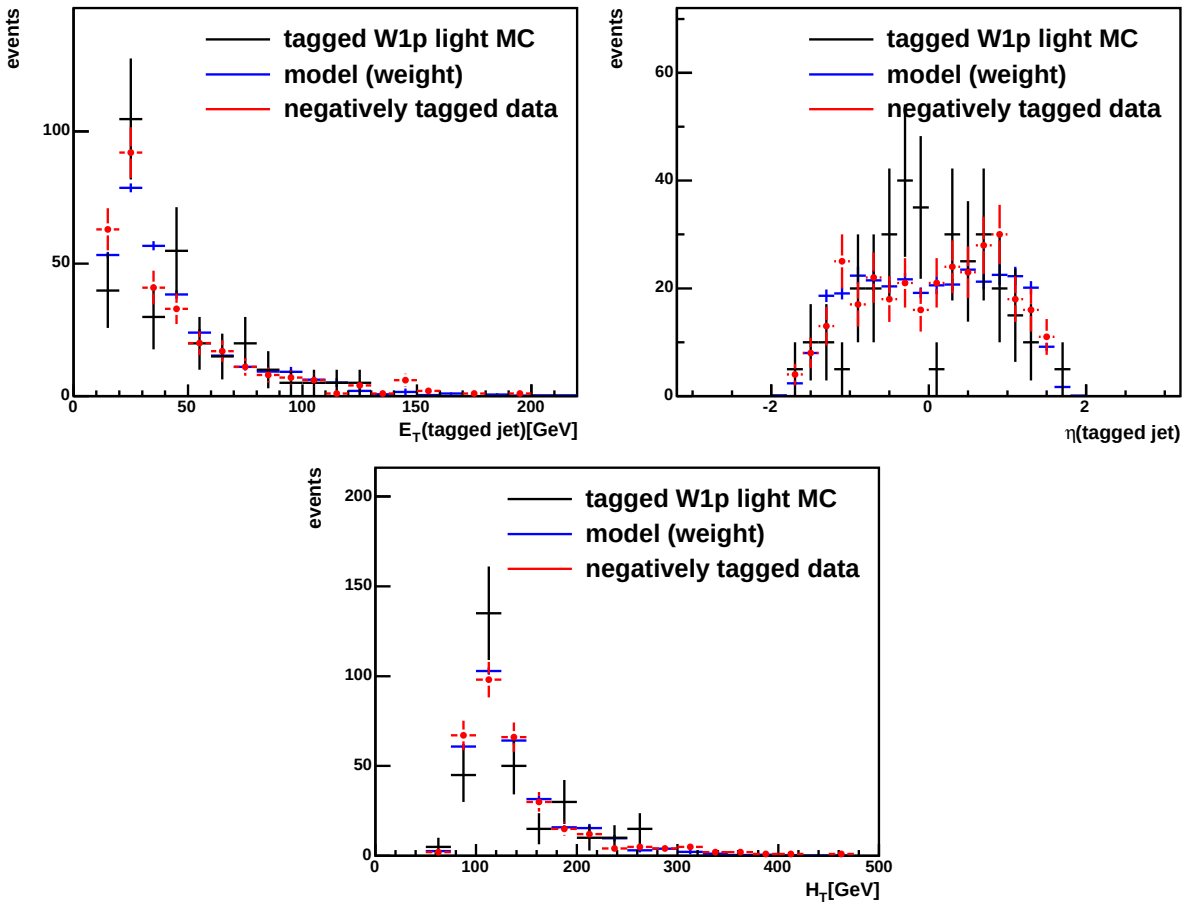

Figure 4: The fake mistag model (blue) in the 1 jet bin is compared to tagged $W + 1p$ light MC events (black) and negatively tagged W + 1jet data. Upper left: E_T of tagged jet; upper right: η of tagged jet; low: $H_T(s\text{-channel})$.

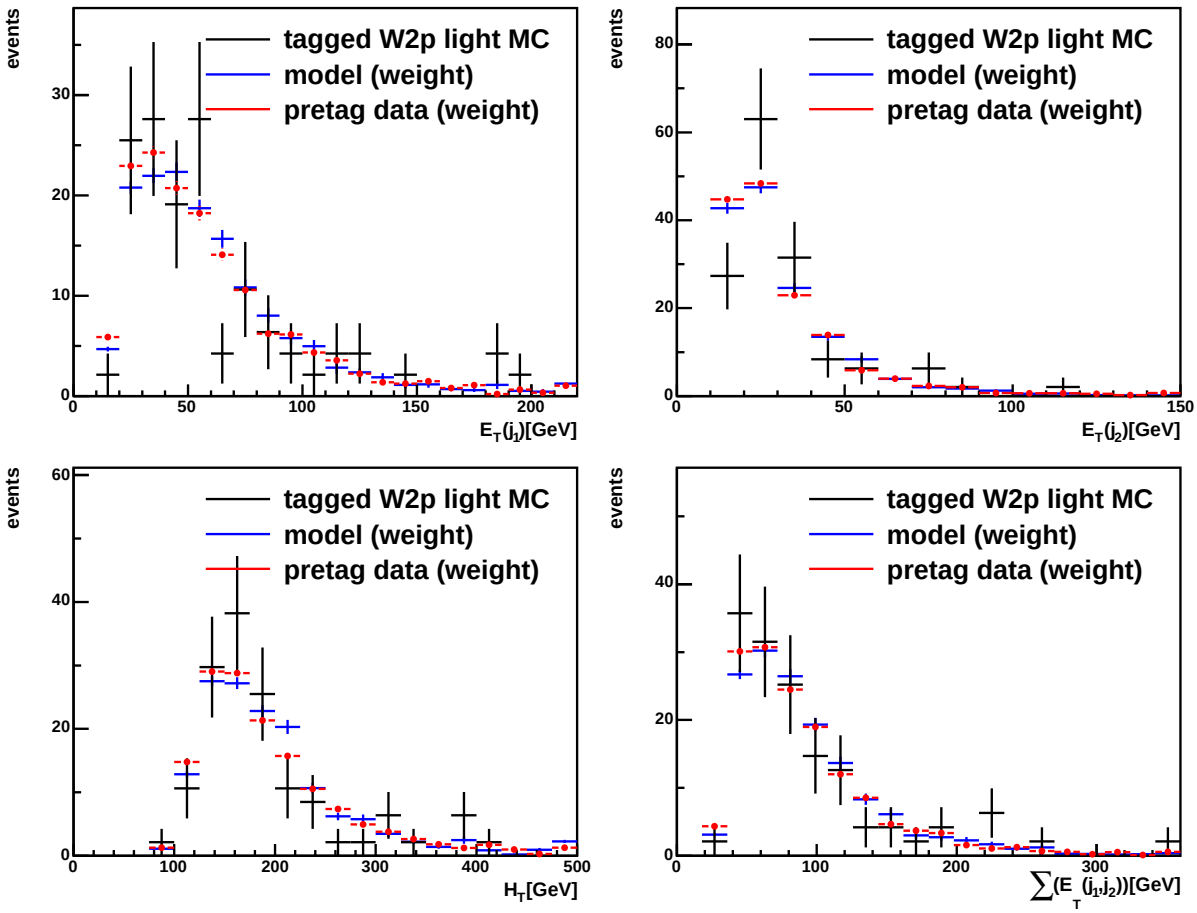

Figure 5: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and weighted pretag W + 2jets data. Upper left: E_T of leading jet (j1); upper right: E_T of second leading jet (j2); lower left: $H_T(s$-channel); lower right: $\sum(E_T)$(j1j2).

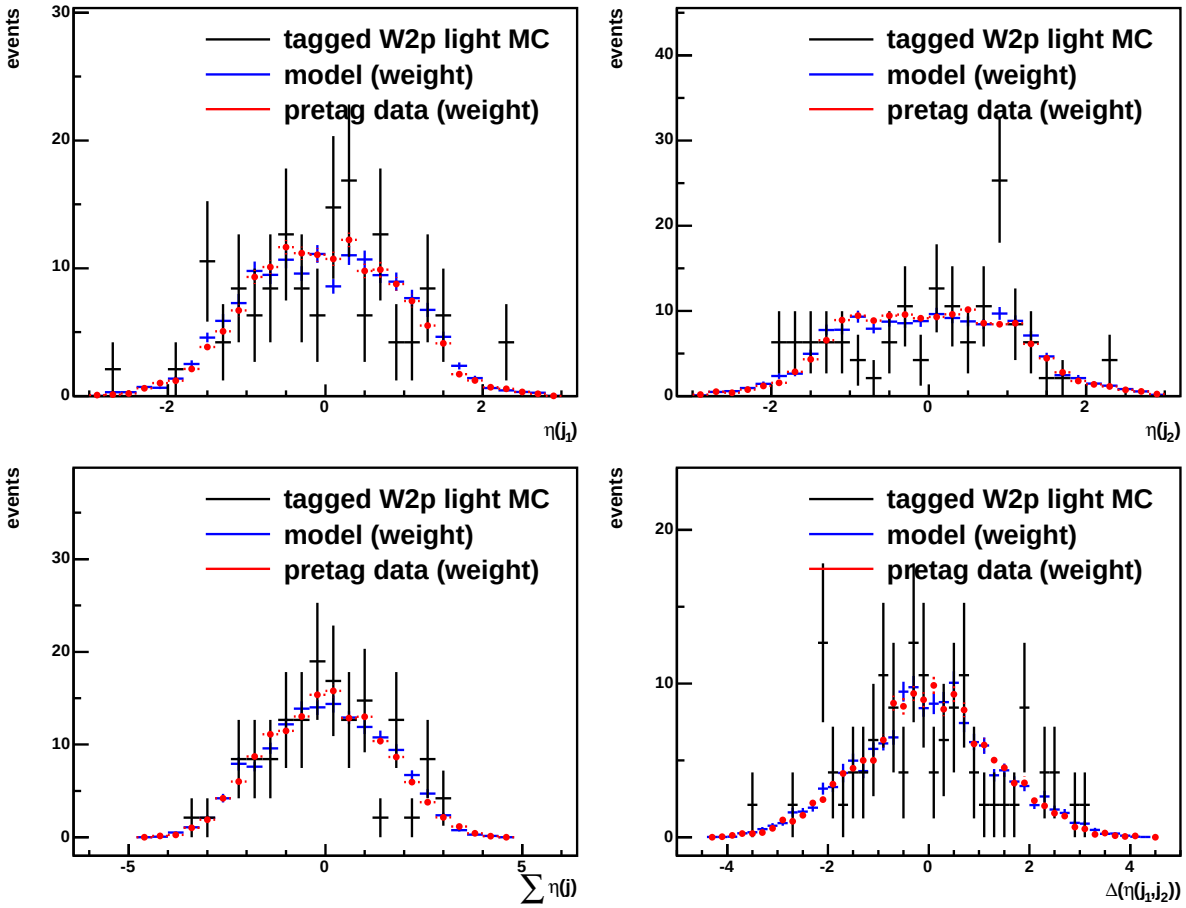

Figure 6: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and weighted pretag W + 2jets data. Upper left: η of leading jet (j1); upper right: η of second leading jet (j2); lower left: $\sum \eta$(j1j2); lower right: $\Delta\eta$(j1j2).

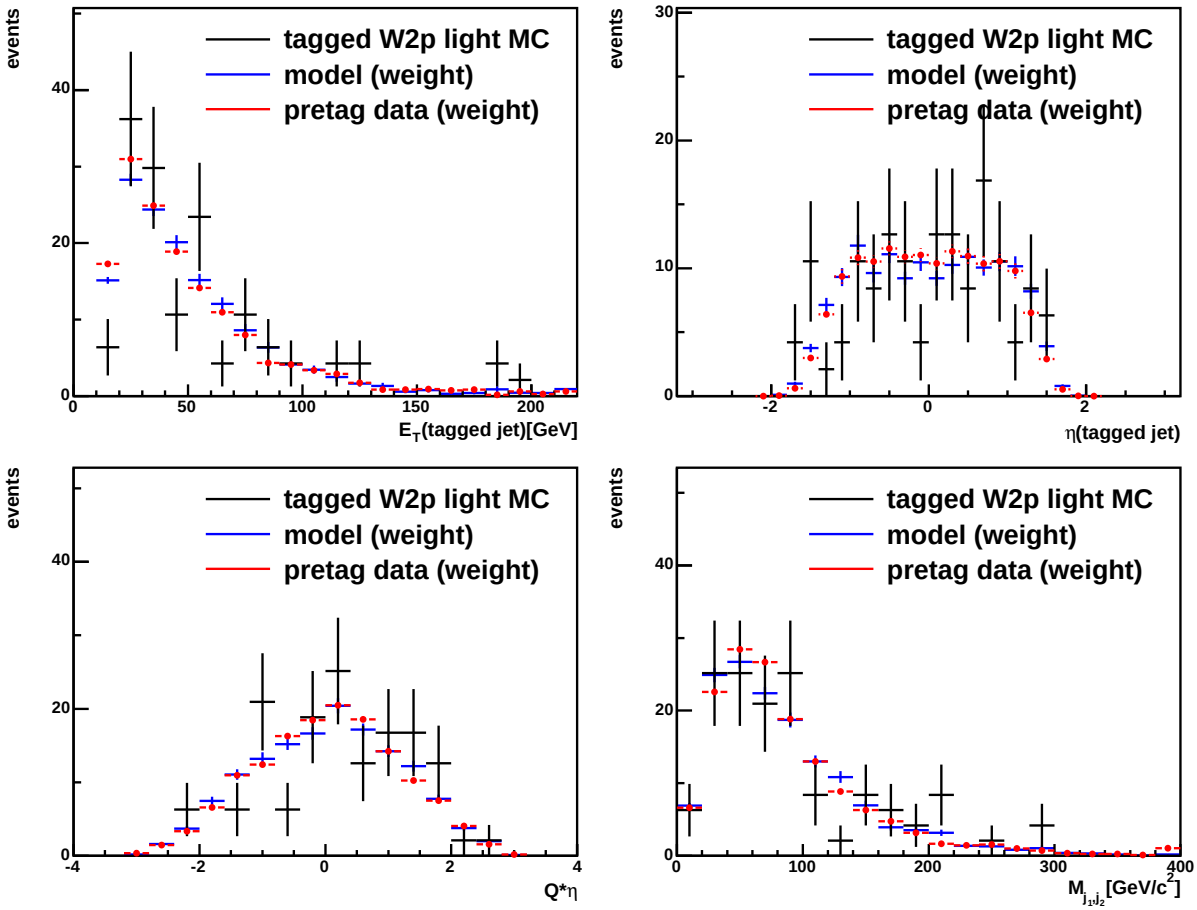

Figure 7: The fake mistag model (blue) is compared to tagged $W + 2p$ light MC events (black) and weighted pretag W + 2jets data. Upper left: E_T of tagged jet; upper right: η of tagged jet; lower left: $Q \times \eta$; lower right: dijetmass(j1j2).

In particular, comparing the E_T distributions of the leading jet (see figure 5, upper left), it can be seen that jets in weighted pretag data events have a harder E_T spectrum than indicated by negatively tagged data events (see figure 1, upper left). The agreement between pretag data and mistag model is reasonable. Looking at the η distributions (see e.g. figure 2), jets in pretag data appear to be slightly more central than in the mistag model. Overall, the kinematic distributions in pretag data and mistag model agree well. This is substantiated by the KS probabilities given in table 1.

Variable	**KS value**	**Variable**	**KS value**
$E_T(j1)$	77.3%	$E_T(j2)$	96.6%
H_T	77.4%	$\sum E_T(j_1, j_2)$	81.9%
$\eta(j_1)$	99.6%	$\eta(j_2)$	99.9%
$\sum \eta(j_1, j_2)$	99.9%	$\Delta\eta(j_1, j_2)$	100%
E_T(taggedjet)	79.4%	η(taggedjet)	99.4%
$Q \times \eta$	99.8%	$M_{j_1 j_2}$	99.1%

Table 1: KS probabilities for pretag data and mistag model. The corresponding distributions are shown in figure 5 to 7.

4 Shape Uncertainty

We estimate the shape uncertainty by varying the event weights given by equation 1 within the errors. For the so-called "plus" scenario, we use $w + \sigma_w$ and for the "minus" scenario, we use $w - \sigma_w$. σ_w is given by equation 2:

$$\sigma_w^2 = (\frac{\partial w}{\partial \alpha\beta})^2 \cdot \sigma_{\alpha\beta}^2 + (\frac{\partial w}{\partial negrate_{\text{tagged}}})^2 \cdot \sigma_{negrate_{\text{tagged}}}^2 \quad (2)$$

with

- $\sigma_{\alpha\beta}$: $0.109 \cdot \alpha\beta$
- $\sigma_{negrate_{\text{tagged}}}$: statistical uncertainty on tag rate and 8% systematic uncertainty

The comparison of the model distributions and the systematic shape distributions are shown in figures 8 to 10.

5 z-Vertex Reweighting

The z-vertex distribution is mismodeled in simulation. To investigate the influence of this mismodeling, we reweight the MC events such that the z-vertex distribution fits the data.

In figure 11, a comparison of the mistag model η distributions is shown for z-vertex reweighted events and events without this additional weight. It is obvious that the z-vertex reweighting does not have a significant influence on the η distributions of the mistag model.

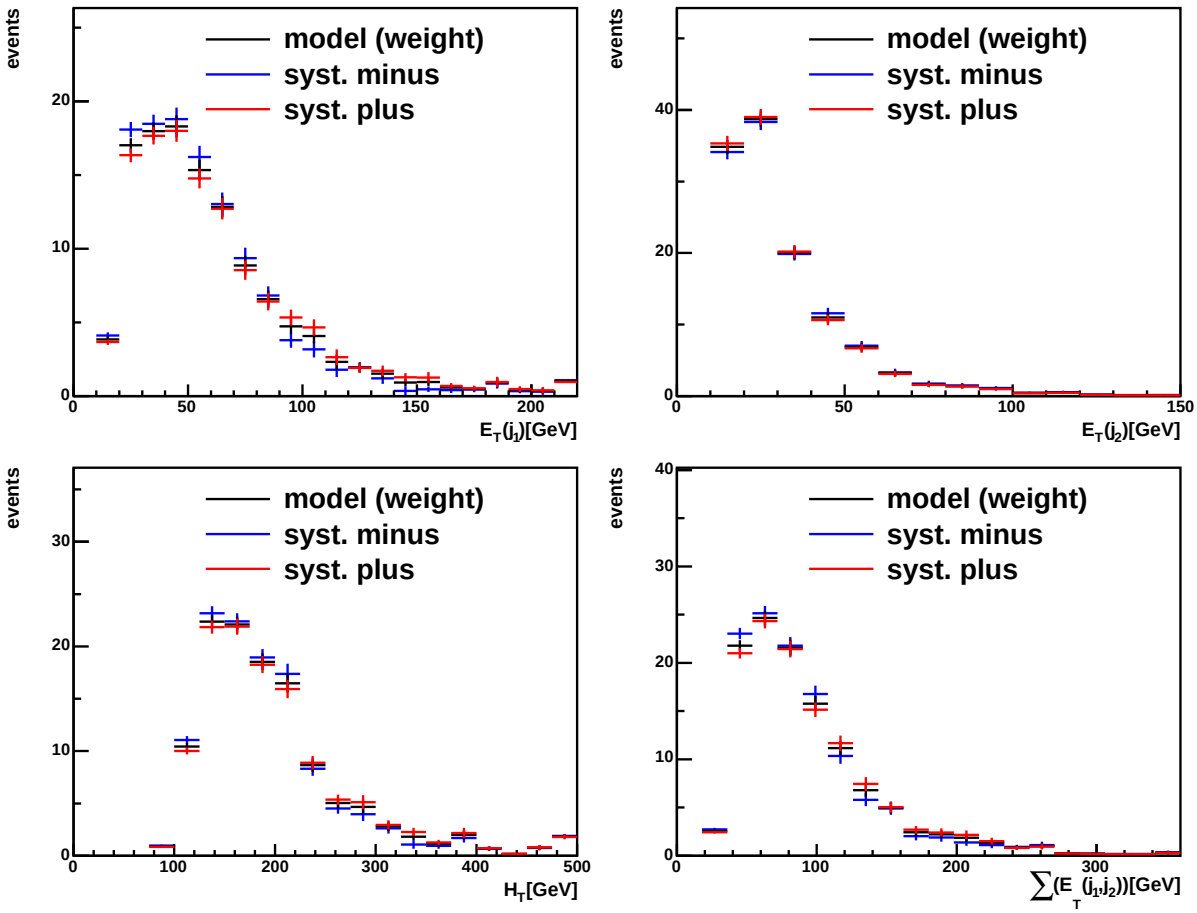

Figure 8: The fake mistag model (black) is compared to the systematics "plus" (red) and "minus" (blue) scenario. Upper left: E_T of leading jet (j1); upper right: E_T of second leading jet (j2); lower left: H_T(s-channel); lower right: $\sum(E_T)$(j1j2).

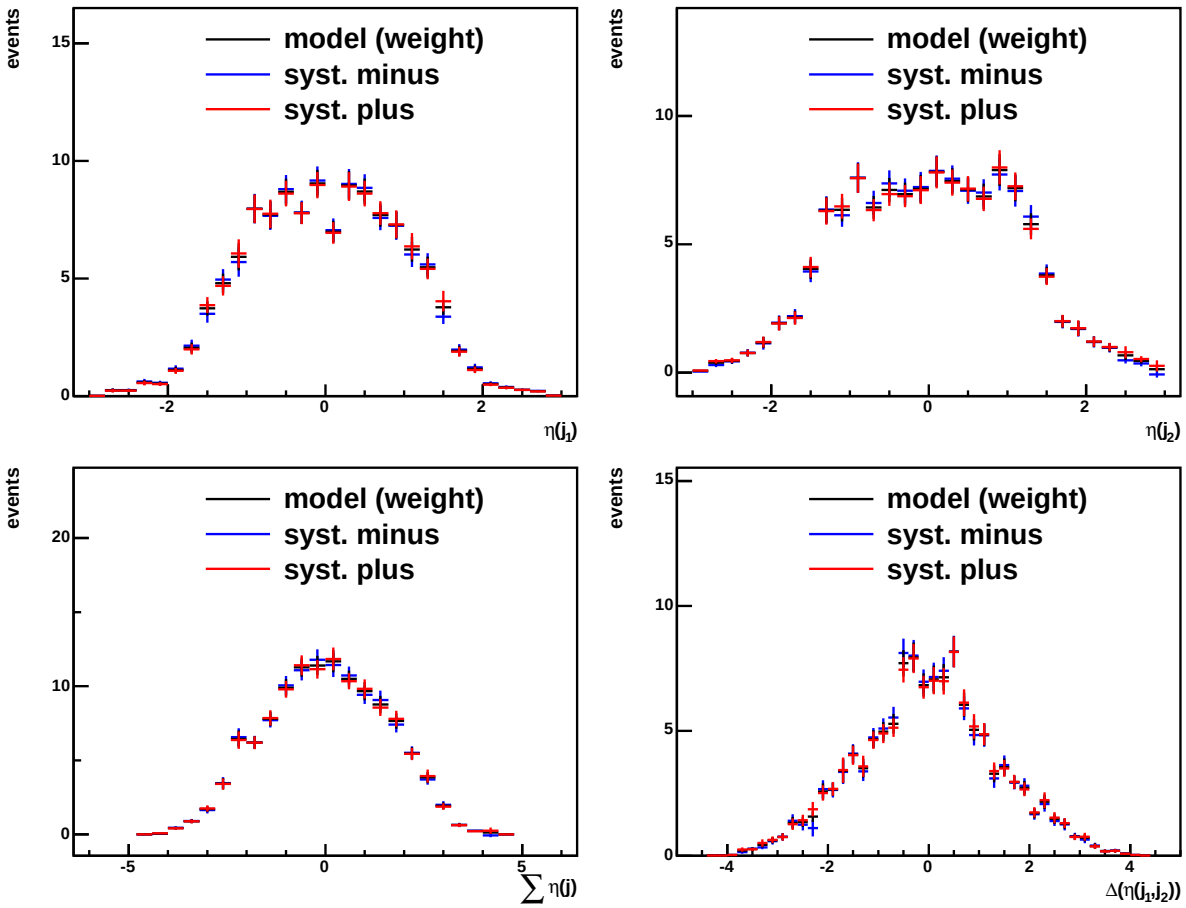

Figure 9: The fake mistag model (black) is compared to the systematics "plus" (red) and "minus" (blue) scenario. Upper left: η of leading jet (j1); upper right: η of second leading jet (j2); lower left: $\sum \eta$(j1j2); lower right: $\Delta\eta$(j1j2).

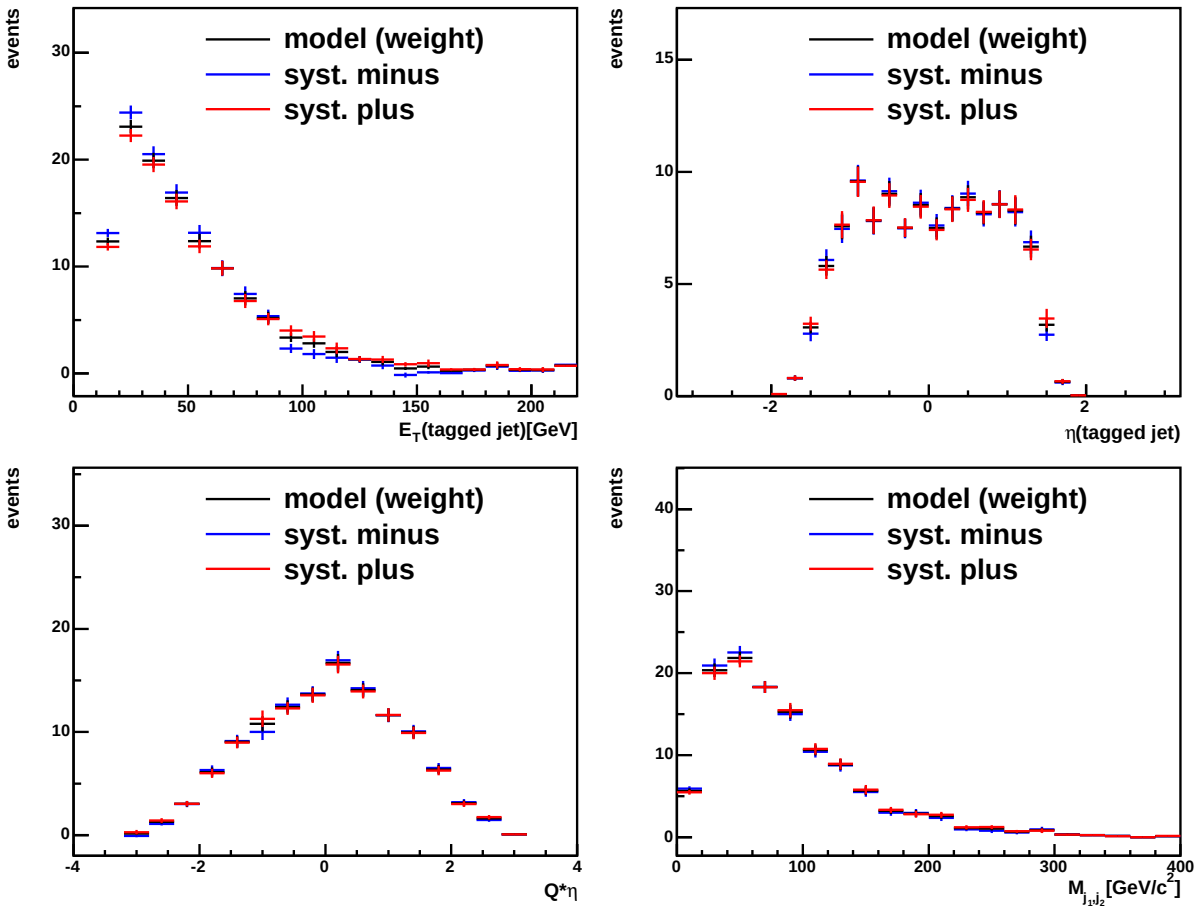

Figure 10: The fake mistag model (black) is compared to the systematics "plus" (red) and "minus" (blue) scenario. Upper left: E_T of tagged jet; upper right: η of tagged jet; lower left: $Q \times \eta$; lower right: dijetmass(j1j2).

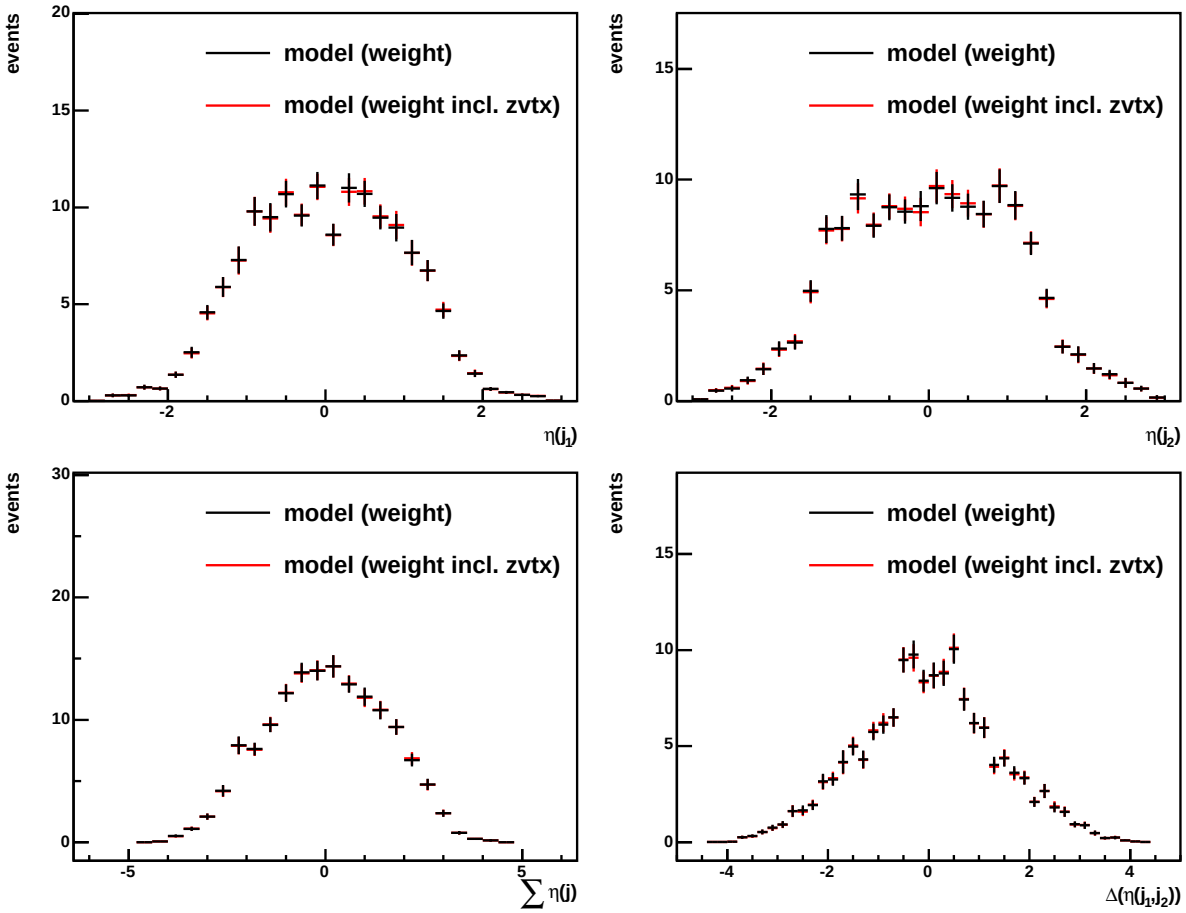

Figure 11: Comparison of mistag model η distributions with (red) and without (black) z-vertex reweighting. Upper left: η of leading jet (j1); upper right: η of second leading jet (j2); lower left: $\sum \eta$(j1j2); lower right: $\Delta\eta$(j1j2).

6 Conclusion

We presented a mistag model based on taggable $W + 2p$ events to model mistagged light events in the single-top analyses. In this model, only events without a heavy flavor match are used. Each event obtains a weight based on the negative tag rate predicted by the mistag matrix.

Comparisons of the model, tagged MC events, negatively tagged data, and weighted pretag data events have shown reasonable agreement, meaning that the model is capable of describing the mistagged light backgrounds. We further investigated the influence of z-vertex mismodeling in simulation. We have shown that reweighting the events in the model to account for this effect does not change the η distributions significantly.

Furthermore, we presented a method to estimate systematic shape uncertainties which include uncertainties in both mistag matrix prediction and mistag asymmetry.

References

[1] S. Budd et al. "Data Based Background Estimate for Summer 2006 Single-Top Search", CDF 8292.

[2] A. Foland et al., "SecVtx Tag Matrices for 5.3.3_nt", CDF 7326.

[3] J. Guimaraes da Costa et al., "Measurement of the SecVtx Mistag Asymmetrie in 5.3.3", CDF 7585.