CDF//TOP/CDFR/9428
Version 1.0
July 21, 2008

Search for Single Top Quark Production using Boosted Decision Trees in 2.7 fb^{-1}

Florencia Canelli,[1] Bruno Casal Laraña,[2] Peter Dong,[3] Craig Group,[1] Alberto Ruiz,[2] Bernd Stelzer,[3] Rocio Vilar [1]

[1] *Fermilab*
[2] *Instituto de Física de Cantabria (CSIC - Univ. Cantabria)*
[3] *University of California, Los Angeles*

Abstract

We present a search for electroweak single top quark production using 2.7 fb^{-1} of CDF II data collected between 2002 and 2008. The analysis employes a multivariate technique based on Boosted Decision Trees (BDT), where the output is used to build a discriminant variable which we will fit to the data. We search for a combined single top s- and t-channel and measure a cross section of $2.4^{+0.8}_{-0.7}$ pb assuming a top quark mass of 175 GeV/c^2. We use the MCLIMIT to calculate the signal significance. The expected (median) p-value for this analysis is 3.8×10^{-7} % which corresponds to a 5.0σ signal significance assuming single top quark production at the rate predicted by the Standard Model. The observed p-value is 0.00017 %, corresponding to a 3.6σ excess.

Contents

0 Introduction

Finding single top quark production is challenging since it is rarely produced ($\sigma_{singletop} \sim$ 2.9 pb) in comparison with other processes with the same final state like W+jets and $t\bar{t}$. The signal to background ratio of the analysis is small, typically on the order of S/B$\sim$1/18 [1]. This calls for a better discrimination of signal and background events which can be achieved by using more information to characterize each event.

This note describes a new analysis approach that attempts to make optimal use of information in the data by means of a multivariate technique via Boosted Decision Trees (BDT) . We train four differnt BDTs which are optimezed for the search of single top signatures in four different final states, namely $W + 2$ jets and $W + 3$ jets with exaclty one SecVtx tag or ≥ 2 SecVtx tags, where the W decays leptonically into an electron (CEM and PHX electrons are considered) or a muon (CMUP and CMX, as well as the new extra muons [2] from the MET+JETS trigger are used in this analysis).

In section 1 we present the event selection employed with the expected event yield. Section 2 describes the BDT multivariate method using the TMVA package [3]. The rest of the note discusses the systematic uncertainties and the results on the data. This note has been updated to include p14-p17 data and this new data is validated against the data from p13 and before in Appendix B. Details of our previous result with 2.2 fb^{-1} of CDF data can be found in reference [4].

1 Event Selection and Expected Number of Events

Candidate events for this analysis are selected by requiring a $W + 2$ jet event topology where the W decays leptonically, $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$. One or both of the two jets should be identified as a b-jet using the secondary vertex tag requirement. The detailed event selection and the estimate strategy has been performed and summarized in a separate CDF note [1, 5]. Table 1 lists the expected event yield.

1.1 Non-triggered Muons

In the previous iteration of the analysis [4], we included new non-triggered muons according to the work of [2]. This acceptance gain is equivalent to 50% of the CMUP contribution for the signal sample, while the background acceptance is increased by about 35%. The overall sensitivity gain is estimated to be 9% [6].

2 Methodology

In order to search for a single top quark production we develope a multivariate technique based on Boosted Decision Trees. To Build the BDTs we make use of the ROOT-integrated package TMVA [3]. The basics of a BDT is described in the following section.

Process	Number of Events in 2.7 fb^{-1}		Sample
	W + 2 jets	W + 3 jets	
s-channel	49.3 ± 7.0	16.3 ± 2.3	stop00
t-channel	74.3 ± 10.9	22.3 ± 3.2	stopm0
$Wb\bar{b}$	549.1 ± 165.5	169.8 ± 51.3	btopXp
$Wc\bar{c}, Wcj$	453.5 ± 139.9	126.7 ± 39.0	ctopXw,stopwX
$Mistags$	410.7 ± 51.0	125.5 ± 15.8	ptopXw, utopXw
$non-W$	75.6 ± 30.2	27.4 ± 11.0	anti-electron / jet-electron / non-isolated data
WW	49.4 ± 5.5	18.0 ± 2.0	itopww
WZ	24.4 ± 1.9	7.3 ± 0.6	itopwz
ZZ	0.7 ± 0.1	0.3 ± 0.0	itopzz
$Z + jets$	31.1 ± 4.6	13.4 ± 2.0	ztopXY
$t\bar{t}$ dilepton	58.8 ± 8.4	51.3 ± 7.3	ttop75
$t\bar{t}$ non-dilepton	114.7 ± 16.4	359.2 ± 51.1	ttop75
Total signal	123.6 ± 17.9	38.6 ± 5.5	
Total prediction	1891.1 ± 312.4	934.4 ± 108.0	
Observed in data	1874	902	

Table 1: Number of expected single top and background events in $2.7 fb^{-1}$of CDF data passing all event selection requirements.

2.1 Description of a Boosted Decision Tree

A Decision Tree (DT) is a sequence of cuts using the discriminating variable (from a given set of variables) which gives the best sig-bkg separation in each "node", until some stop criterion is reached. The phase space is split into regions that are eventually classified as signal or background, depending on the majority of training events that end up in the final "leaf" nodes. A schematic view of a DT is shown in Figure 1. A single DT is very similar to rectangular cuts. However, whereas a cut-based analysis is able to select only one hypercube as region of phase space, the decision tree is able to split the phase space into a large number of hypercubes, each of which is identified as either signal-like or background-like . The path down the tree to each leaf node represents an individual cut sequence that selects signal or background depending on the type of the leaf node.

A shortcoming of decision trees is their instability with respect to statistical fluctuations in the training sample from which the tree structure is derived. For example, if two input variables exhibit similar separation power, a fluctuation in the training sample may cause the tree growing algorithm to decide to split on one variable, while the other variable could have been selected without that fluctuation. In such a case the whole tree structure is altered below this node, possibly resulting also in a substantially

different classifier response.

This problem is overcome by constructing a forest of decision trees and classifying an event on a majority vote of the classifications done by each tree in the forest. All trees in the forest are derived from the same training sample, with the events being subsequently subjected to so-called boosting, a procedure which modifies their weights in the sample. Boosting increases the statistical stability of the classifier and typically also improves the separation performance compared to a single decision tree. However, the advantage of the straightforward interpretation of the decision tree is lost. While one can of course still look at a limited number of trees trying to interprete the training result, one will hardly be able to do so for hundreds of trees in a forest. Nevertheless, the general structure of the selection can already be understood by looking at a limited number of individual trees.

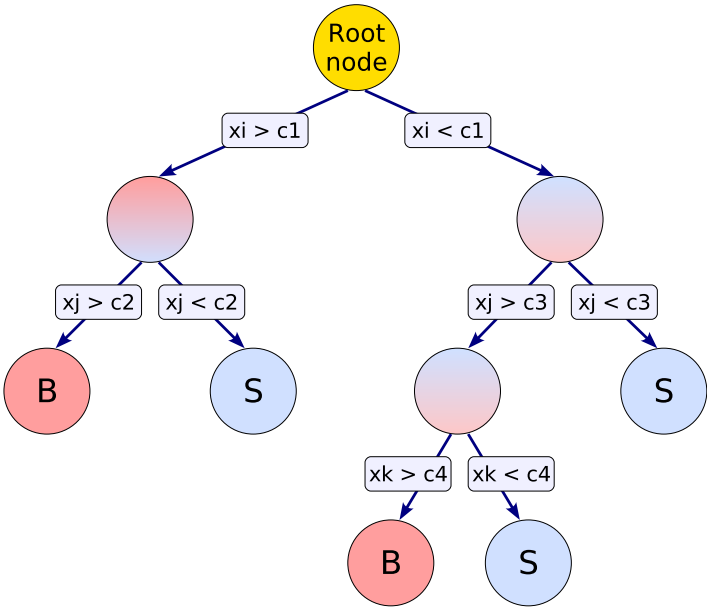

Figure 1: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is performed. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled S for signal and B for background depending on the majority of events that end up in the respective nodes.

2.2 Training of the BDTs

Four different BDTs optimized for the search of single top in four different signal regions are trained:

- 2 jets bin, 1tag
- 2 jets bin, 2 tags
- 3 jets bin, 1 tag
- 3 jets bin, ≥ 2 tags

In the following subsections we summarize the choices made for the building of the BDTs (splitting criteria, boosting algorithm and pruning method), as well as the samples and variables used for the training.

2.2.1 Building a DT

The training or building of a DT is the process that defines the splitting criteria for each node. At each node, the split is determined by finding the variable and corresponding cut value that provides the best separation between signal and background. The node splitting is stopped once it has reached a minimum number of events. The end –or leaf nodes– are classified as signal or background according to the class the majority of events belongs to. Different separation criteria can be configured to assess the performance of a variable and a specific cut requirement. For this analisys we have chosen the so-called *Gini Index* which optimezes the quantity $p\cdot(1-p) = S\cdot B/(S+B)^2$, where $p = S/(S+B)$ is the purity and $S(B)$ is the number of signal (background) events in the node [1]. The cut values are optimised by scanning over the variable range with a granularity given by the parameter $nCuts$. The value of $nCuts = 25$ proved to be a good compromise between computing time and step size. Finer stepping values did not increase noticeably the performance of the BDTs.

2.2.2 Boosting Algorithm

Boosting is a general procedure in which the same classifier is trained several times using a successively boosted (reweighted) training event sample. The final classifier is then derived from the combination of all the individual classifiers. The most popular boosting algorithm is the so-called *AdaBoost*[7] (adaptive boost), where events that were misclassified during the training of a tree are given a higher event weight in the training of the next following tree. Starting with the original event weights when training the first decision tree, the subsequent tree is trained using a modified event sample where the weights of previously misclassified events are multiplied by a common boost weight α. The boost weight is derived from the misclassification rate err (number of missclasified events over total events) of the previous tree,

$$\alpha = \frac{1 - \text{err}}{\text{err}}. \tag{1}$$

[1] Another splitting criteria based on $S/\sqrt{S+B}$ has been investigated for this analyses, however the performance of the trained BDT has been found to be slightly worse than using *Gini Index*

With the result of an individual tree $h(x)$ (x being the tuple of input variables) encoded for signal and background as $h(x) = +1$ and -1, respectively, the resulting event classification $y_{BDT}(x)$ for the boosted classifier is then given by

$$y_{BDT}(x) = \sum_{i \in forest} \ln(\alpha_i) \cdot h_i(x), \tag{2}$$

where the sum is over all trees in the forest. Small (large) values for $y_{BDT}(x)$ indicate a background-like (signal-like) event.

Another possible modification of Eq. 2 is to use the training purity in the leaf node as respectively signal or background weights rather than relying on the binary decision, $h(x) = p$. It has been found that it is preferable to train with the latter option the single tag BDTs, while the double tag BDTs –with smaller statistics samples– perform better when trained with the former option.

2.2.3 Pruning Method

Pruning is the process of cutting back a tree from the bottom up after it has been built to its maximum size. Its purpose is to remove statistically insignificant nodes and thus reduce the overtraining of the tree. It has been found to be beneficial to first grow the tree to its maximum size and then cut back, rather than interrupting the node splitting at an earlier stage. This is because apparently insignificant splits can nevertheless lead to good splits further down the tree.

In this analysis we use the Cost complexity [8] pruning algorithm which relates the number of nodes in a subtree below a node to the gain in terms of misclassified training events by the subtree compared the the node itself with no further splitting. The cost estimate R chosen for the misclassification of training events is given by the misclassification rate $1 - \max(p, 1-p)$ in a node. The cost complexity for this node is then defined by

$$\rho = \frac{R(\text{node}) - R(\text{subtree below node})}{\#nodes(\text{subtree below node}) - 1} \tag{3}$$

The node with the smallest ρ value in the tree is recursively pruned away as long as $\rho <$ PruneStrength, where PruneStrength is a parameter which has to be tuned for each analisys until overtraining is completely avoided[2]. An example of a non-overtrained BDT is shown in Figure 2

2.2.4 Training Samples

A natural way of choose the sample composition for the training of the BDTs is to use a mixture of all the expected proccesses with the correct estimated composition.

[2] A more aggressive approach would be to chose a smaller PruneStrength. Although the overtraining wouldn't be completely avoided, one could remove the events used in the training to build the discriminant templates, so the bias from the overtrainig is avoided and signal to background separation power would be improved.

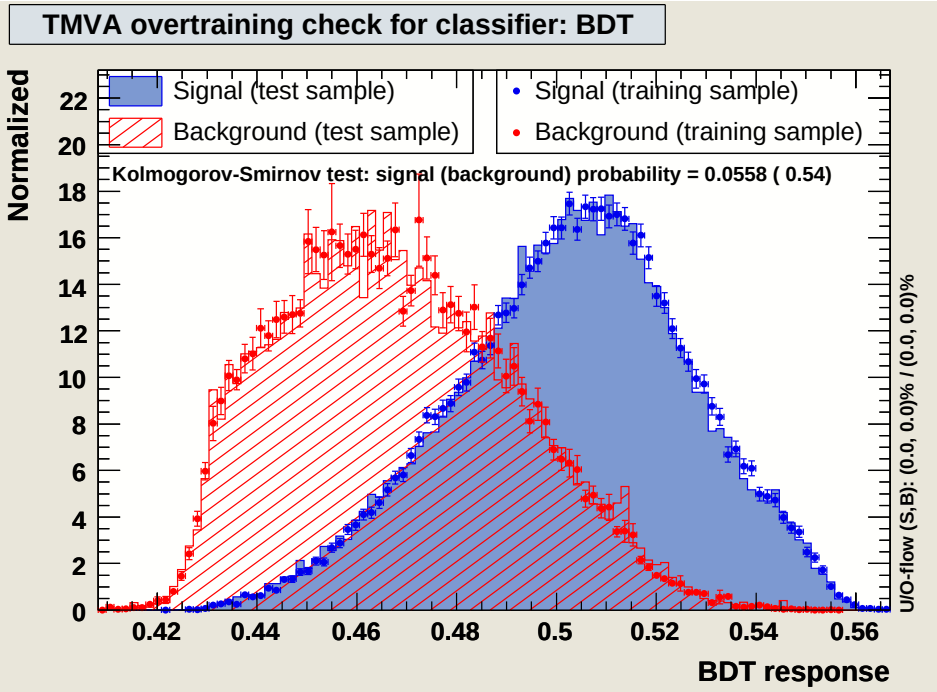

Figure 2: Example of a BDT trained in the 2-jets and 1-tag sample with enough PruneStrength to avoid overtraining. Blue are signal events and red are background events. Points are events from the training sample, and histogram are events from test sample.

TMVA has the ability of introduce event by event weights into the training sample, in this way we can use as much statistics as we have available[3] while keeping the correct predicted composition between samples. Although, one has to be careful with large samples –for instance the mistag sample– since it has been found that composing your training sample with much more background events than signal events can degrade the performance of the trained BDT. Then, it is preferable to cut big samples in such a way that the total number of background events used for the training is smaller than ~ 5 times the total number of signal events while keeping the real background composition via weights in the training.

The backgrounds proccesses included for the training are $t\bar{t}$ and $W + bottom$ for the double tag channels, plus $W + charm$ and $W + light$ for the 1 tag channels. The inclusion of all backgrounds (except non-W) into the training has been investigated, but it has been found that the performance of the BDTs are very similar so the simpler case has been chosen.

[3]At least half of the events in each sample used in the trainig are kept for a subsequent test of the training.

2.2.5 Input Variables

Decision trees are insensitive to the inclusion of poorly discriminating input variables. While for artificial neural networks it is typically more difficult to deal with such additional variables, the decision tree training algorithm will basically ignore non discriminating variables as for each node splitting only the best discriminating variable is used.

The variables used for the training of the BDTs in the 2-jet bin channels are:

- the E_T and the η of both jets
- the p_T and the η of the lepton
- the missing transverse energy in the event $\not\!\!E_T$
- the scalar sum of the transverse energies $H_T = \sum_{jets} E_T + p_T + \not\!\!E_T$
- the invariant mass of the di-jet system m_{j1j2}
- the η and the transverse mass of the W boson $m_T(W)$
- the mass of the reconstructed top $m_{l\nu b}$
- the invariant mass of the lepton, neutrino and both jets $m_{l\nu j1j2}$
- the charge of the lepton times the η of the b-quark jet $Q \times \eta$
- the KIT NN flavor separator [9, 10]
- the $\Delta\phi$ between the jets and the $\not\!\!E_T$
- the $\Delta\phi$ between the jets and the lepton
- the $\Delta\phi$ between the the lepton and the $\not\!\!E_T$
- the cosine of the angle between the lepton and the jets

For the 3-jet channels the following variables are also included:

- the E_T and the η of the 3rd jet
- the three combinations of di-jet system invariant masses
- the invariant mass of the three jets
- the invariant mass of the lepton, neutrino and three jets $m_{l\nu j1j2}$
- the $\Delta\phi$ between the 3rd jet and the $\not\!\!E_T$
- the $\Delta\phi$ between the 3rd jet and the lepton

	2-jets, 1-tag		2-jets, 2-tag	
Rank	Variable	Variable Importance	Variable	Variable Importance
1	KaNN	4.597e-01	mJ1J2	7.453e-02
2	mJ1J2	1.799e-01	Mlnub	6.599e-02
3	QEta	1.077e-01	wmt	6.453e-02
4	Mlnub	4.684e-02	MetJ1DPhi	5.142e-02
5	J1Et	4.570e-02	Mlnuj1j2	4.805e-02
6	wmt	1.890e-02	KaNN	4.656e-02
7	J2Et	1.874e-02	cosLepJ1	4.645e-02
8	LepPt	1.639e-02	J1Eta	4.563e-02
9	cosLepJ1	1.572e-02	J1Et	4.522e-02
10	MetLepDPhi	1.481e-02	J2Et	4.424e-02
11	J2Eta	1.063e-02	LepEta	4.366e-02
12	LepEta	8.744e-03	MetJ2DPhi	4.339e-02
13	Mlnuj1j2	8.106e-03	LepJ2DPhi	4.252e-02
14	Ht	7.755e-03	cosLepJ2	4.249e-02
15	MetJ1DPhi	7.589e-03	MetLepDPhi	4.098e-02
16	LepJ1DPhi	6.232e-03	LepJ1DPhi	3.886e-02
17	cosLepJ2	6.035e-03	LepPt	3.869e-02
18	met	4.676e-03	QEta	3.850e-02
19	MetJ2DPhi	4.174e-03	met	3.840e-02
20	J1Eta	3.922e-03	J2Eta	3.573e-02
21	WEta	3.881e-03	Ht	3.227e-02
22	LepJ2DPhi	3.839e-03	WEta	3.188e-02

Table 2: Variable ranking result for both 2-jets channels. Top variable is best ranked.

All the energies involved are level-5 corrected. When a variables requires a b-quark jet, it is chosen as the b-tagged jet for the single tag samples, and as the jet with largest $Q \times \eta$ for the double tag sample. For the reconstruction of the neutrino, the smallest $|p_z^\nu|$ solution is chosen.

Validation plots of all the above variables in the four signal regions, as well as in the control regions are shown in Appendix A. Validation for the newest data periods is included in B

A ranking of the BDT input variables is derived by counting how often the variables are used to split decision tree nodes, and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node. Tables 2-3 show the variable ranking for each trained channel.

	3-jets, 1-tag		3-jets, 2-tag	
Rank	Variable	Variable Importance	Variable	Variable Importance
1	KaNN	3.846e-01	QEta	7.785e-02
2	mJ2J3	7.124e-02	Ht	5.032e-02
3	wmt	5.449e-02	Mlnub	4.698e-02
4	mJ1J3	5.358e-02	mJ2J3	4.100e-02
5	Ht	4.978e-02	mJ1J2	3.864e-02
6	QEta	4.729e-02	mJ1J2J3	3.735e-02
7	Mlnub	3.924e-02	MetLepDPhi	3.677e-02
8	mJ1J2J3	2.642e-02	wmt	3.632e-02
9	J3Et	2.609e-02	J1Eta	3.573e-02
10	J2Eta	2.549e-02	cosLepJ1	3.447e-02
11	mJ1J2	2.490e-02	J2Et	3.400e-02
12	cosLepJ1	2.370e-02	J3Et	3.324e-02
13	LepPt	1.910e-02	mJ1J3	3.307e-02
14	LepEta	1.774e-02	J2Eta	3.275e-02
15	MetLepDPhi	1.382e-02	cosLepJ3	3.046e-02
16	J1Eta	1.373e-02	J1Et	3.017e-02
17	LepJ1DPhi	1.223e-02	WEta	2.993e-02
18	cosLepJ2	1.069e-02	MetJ2DPhi	2.988e-02
19	cosLepJ3	9.661e-03	LepJ1DPhi	2.898e-02
20	J1Et	8.641e-03	KaNN	2.894e-02
21	MetJ3DPhi	8.278e-03	LepEta	2.873e-02
22	MetJ2DPhi	8.256e-03	LepJ3DPhi	2.855e-02
23	met	8.141e-03	Mlnuj1j2j3	2.828e-02
24	MetJ1DPhi	7.496e-03	MetJ3DPhi	2.692e-02
25	LepJ3DPhi	6.743e-03	cosLepJ2	2.616e-02
26	J2Et	6.133e-03	LepPt	2.518e-02
27	LepJ2DPhi	6.035e-03	MetJ1DPhi	2.508e-02
28	WEta	5.932e-03	LepJ2DPhi	2.204e-02
29	Mlnuj1j2	5.417e-03	met	2.160e-02
30	Mlnuj1j2j3	5.108e-03	Mlnuj1j2	2.060e-02

Table 3: Variable ranking result for both 3-jets channels. Top variable is best ranked.

2.3 Distributions of the BDT Outputs

We use the output of the BDT trained in the four channels as the discriminant for a s-channel and t-channel combined search. The raw output of the BDTs are always in a range inside [-1, 1], howeve we make a transformation which consist in a stretching and a shift in such a way that the output goes from -1 to 1. In that transformation we require the overflow bin to have at least some backgrounds events in order not to make

MCLIMIT code confused by the absence of background. The choice for the number of bins and the overflow bin are optimized performing a 2D scan based on the o-value as figure of merit.

Figure 3 includes the final templates we use in the cross section fit and to throw pseudo-experiments. For showing purposes, we group templates with similar shapes in categories, so $W + bottom$ is the composition of $Wb\bar{b}$, WZ, ZZ, $Z + jets$ and non-W weighted by *Method 2* predictions; $W + charm$ is composed by $Wc\bar{c}$, Wcj and WW. Figure 4 shows the predicted distributions of the BDT output in the four channels.

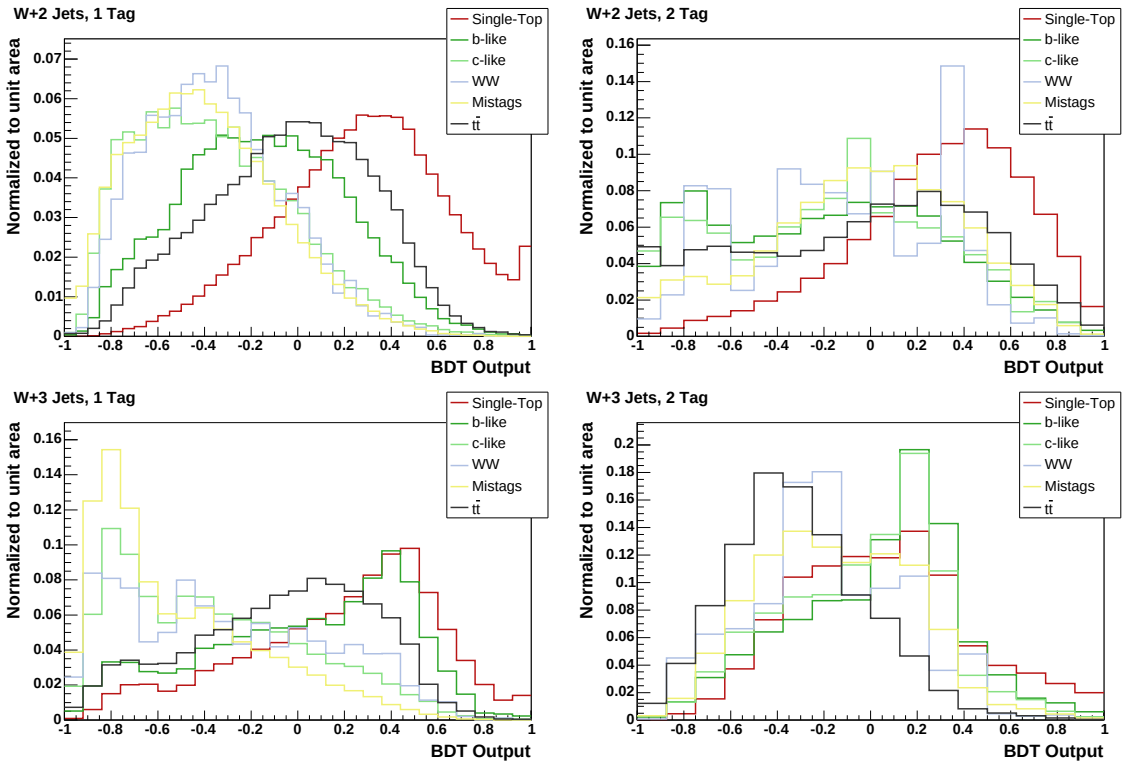

Figure 3: Templates of the BDT outputs. The top two plots show the two-jet bin while the bottom plots show the three-jet bin. Single-tag discriminants are on the left side, while double-tag discriminants are on the right. All histograms are normalized to unit area.

2.4 Cross Check of the Method

In this section we cross-check whether the MC prediction of the BDT output represents the data well without looking at the single top candidate events. We choose the "taggable but not tagged" side-band data. That is, we select W+2/3 jets events according to our nominal event selection and require that at least one jet is taggable but that neither are tagged by the SECVTX algorithm. This event selection is orthogonal to the single top signal region while it still represents a very similar kinematic event topology.

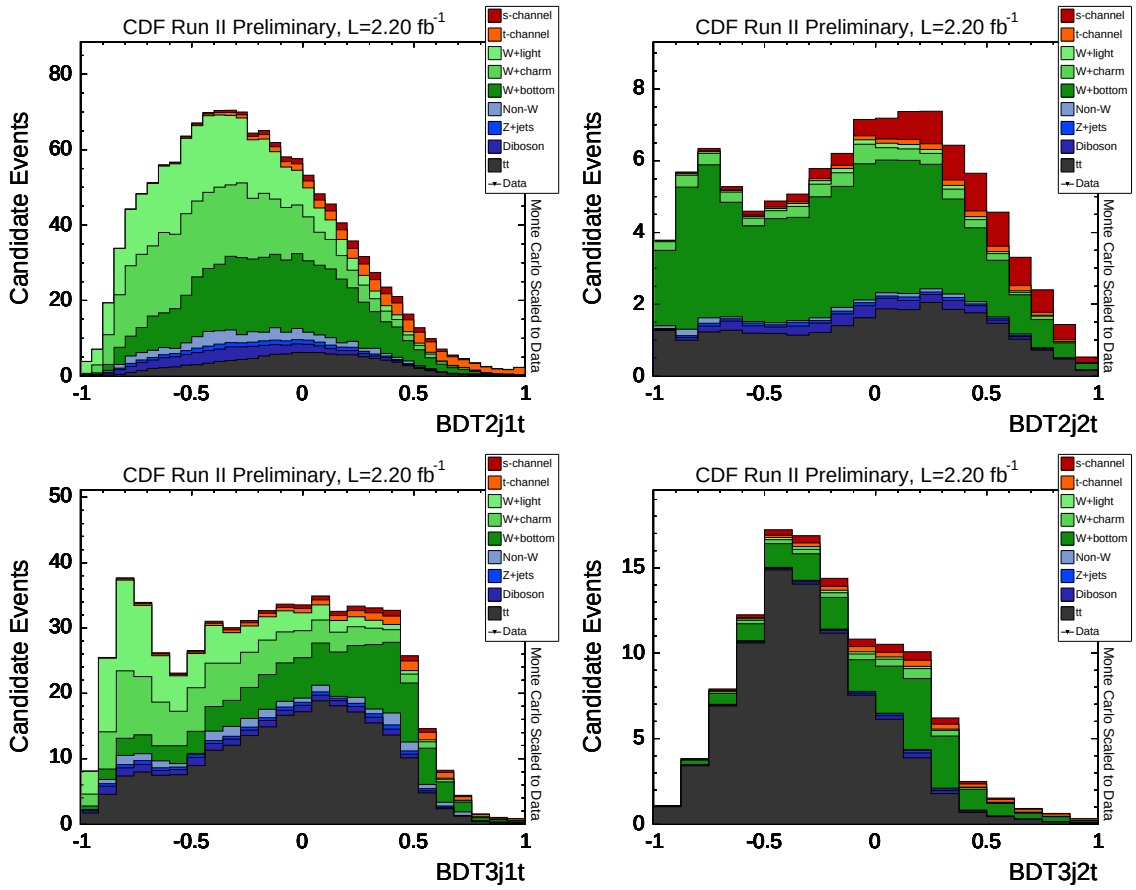

Figure 4: Predicted distributions of the BDT outputs. The top two plots show the two-jet bin while the bottom plots show the three-jet bin. Single-tag discriminants are on the left side, while double-tag discriminants are on the right.

Another advantage is that this sample has very little contribution from top (<0.5%). The input variables for the BDT are the same as in the single top analysis (with some exceptions as the KIT flavor separator and variables requiring a b-jet, these variables for the untagged samples are defined in Appendix A). Data-Monte Carlo comparisons of the input variables are shown in Appendix A, where we find generally good agreement between data and Monte Carlo prediction. We also plot the distribution of the BDTs outputs shown in Figures 5-6. The agreement is good, which assures us that the BDT output in Monte Carlo is well represented by data.

We also cross-checked the shape of our $t\bar{t}$ template by looking in the $t\bar{t}$-enriched data tagged lepton + 4 jets sample. Figure 7 shows the comparison. The lepton + 4 jets sample has an expected $t\bar{t}$ contribution of about 75%. In these plots, the Monte Carlo distributions are normalized to the data. Within statistics, we find good agreement in the data and Monte Carlo shapes.

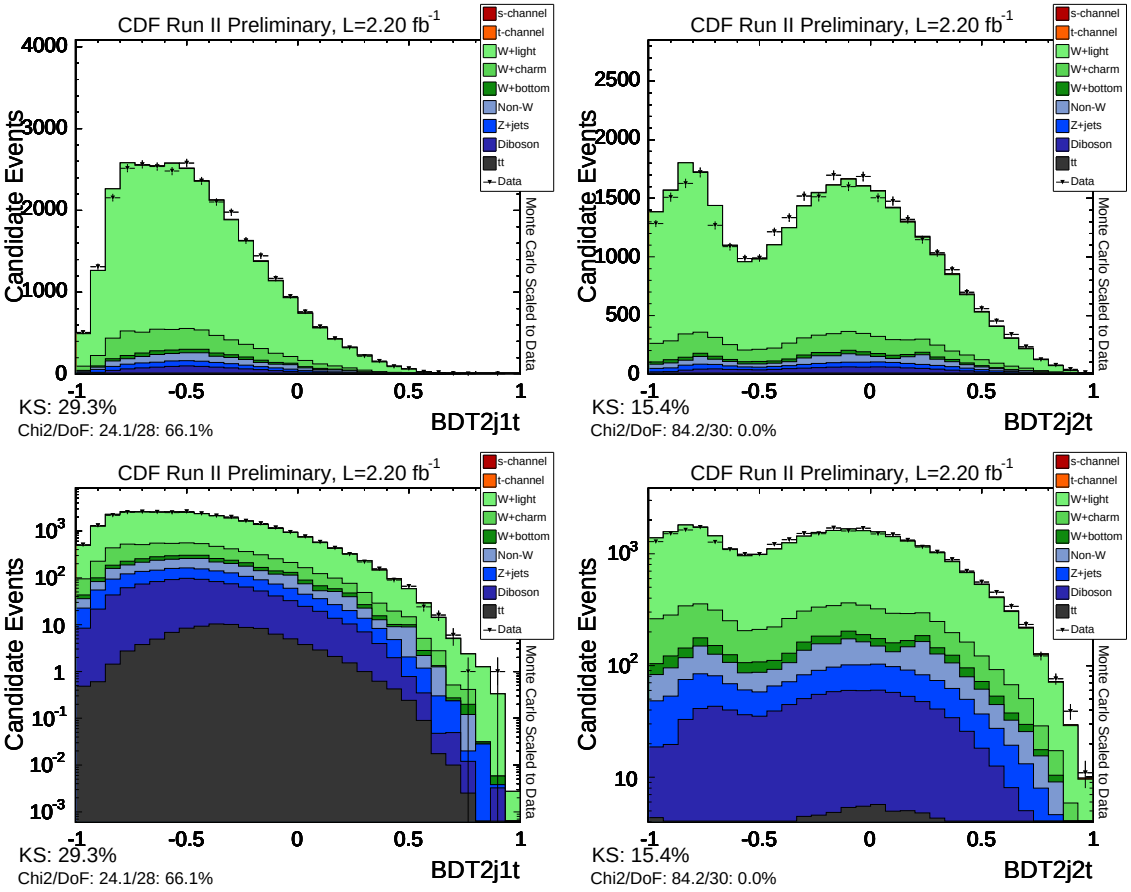

Figure 5: Comparison of the outputs of the BDTs for $W+2$ jet Monte Carlo and data in the taggable but not tagged control sample. The left-hand plots are the single tagged events and the right-hand plots are double-tagged. Linear scale in the top figures, and logarithmic scale in the bottom figures (shown for the $2.2 fb^{-1}$ sample only).

2.5 The Likelihood Function

The likelihood function, $\mathcal{L}$, is a function of the unknown Poisson means for signal and background and is defined such that it expresses the joint probability of observing the N data events at their respective values of the BDT output. The values of the Poisson means at which $\mathcal{L}$ achieves its maximum, corresponds to the most probable estimate for the true signal and background content in the data sample.

We perform a binned likelihood fit to the BDT output. To make it easier to compare the different fit parameters, we define the fit parameter as $\beta_j = \sigma_j^{Fit}/\sigma_j^{SM}$ where β_j is unity when the fit result corresponds to the expected number of events obtained from the independent *Method 2* signal/background estimate:

$$\mathcal{L} = \prod_{j=2}^{5} G_j(\beta_j;\sigma_j) \prod_{k=1}^{B} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \tag{4}$$

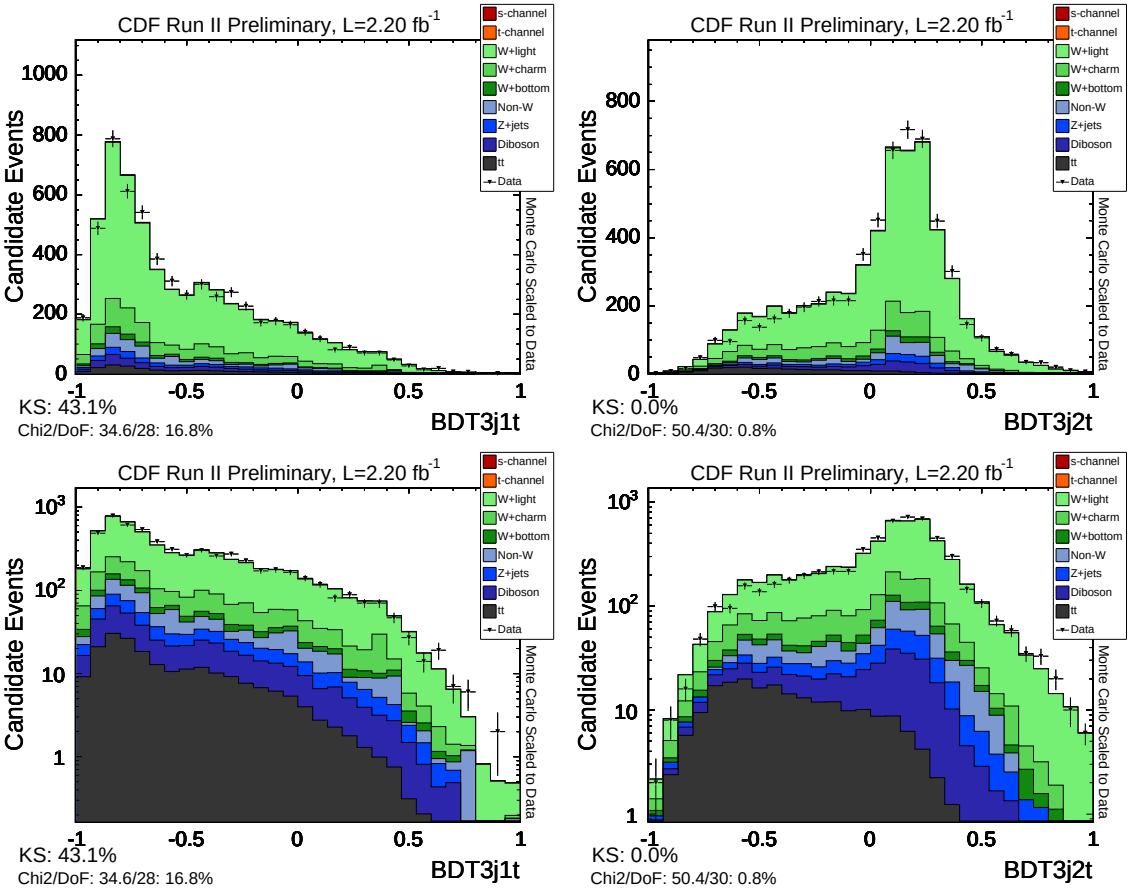

Figure 6: Comparison of the outputs of the BDTs for $W+3$ jet Monte Carlo and data in the taggable but not tagged control sample. The left-hand plots are the single tagged events and the right-hand plots are double-tagged. Linear scale in the top figures, and logarithmic scale in the bottom figures (shown for the $2.2 fb^{-1}$ sample only).

The Gaussian constraints to the backgrounds are given by:

$$G_j(\beta_j;\sigma_j) = \frac{1}{\sqrt{2\,\pi\cdot\sigma_j^2}}\exp\left[-\frac{1}{2}\cdot\left(\frac{\beta_j-1.0}{\sigma_j}\right)^2\right] \tag{5}$$

$$\mu_k = \beta_{\text{single top}}\cdot T_{jk} + \beta_{\text{W+bottom jets}}\cdot T_{jk} + \beta_{\text{W+charm jets}}\cdot T_{jk} + \beta_{\text{mistags}}\cdot T_{jk} + \beta_{\text{t}\bar{\text{t}}}\cdot T_{jk} \tag{6}$$

The index k runs over the bins of the fitted histogram. The template histograms are normalized to the predicted number of events as shown in Table 1. This means, $\sum_{k=1}^{B} T_j k = N_j^{pred}$.

In addition, the prediction in each bin needs an additional Gaussian uncertainty due to the limitations of Monte Carlo statistics. Each bin is allowed to fluctuate according to the total uncertainty in that bin, which is the sum in quadrature of the weight of

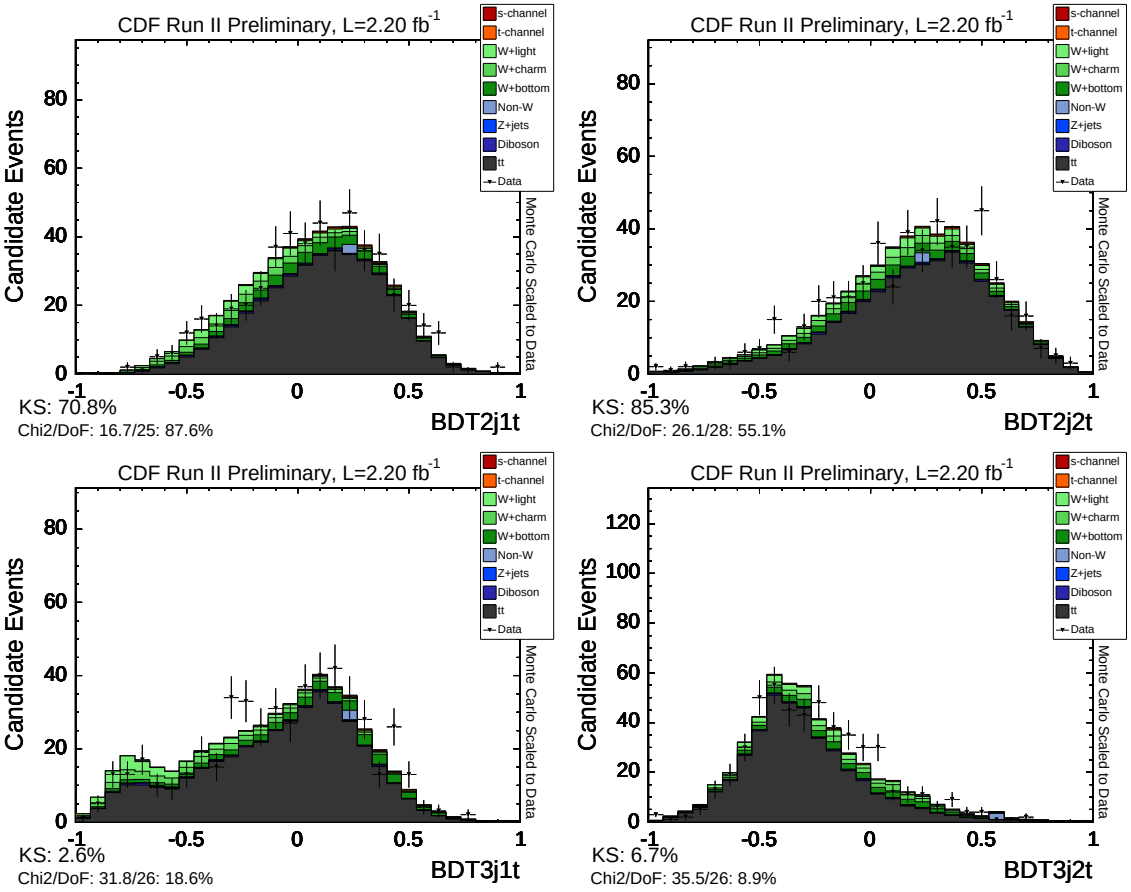

Figure 7: Comparison of the outputs of the BDTs for lepton + four jets data and Monte Carlo, using the leading two jets. The left-hand plots are the single tagged events and the right-hand plots are double-tagged. The top plots are in the two-jet bin and the bottom plots are in the three-jet bin (shown for the $2.2 fb^{-1}$ sample only.)

each event. This prevents us from overestimating our sensitivity due to a fluctuation in Monte Carlo.

3 Incorporating Systematic Uncertainties

Systematic uncertainties can bias the outcome of this analysis and have to be incorporated into the result. We address systematic uncertainty from several different sources: (1) jet energy scale (2) initial state radiation (3) final state radiation (4) parton distribution functions (5) the event generator, the uncertainty in the event detection efficiency and luminosity (6) neural network jet flavor separator uncertainty, (7) ALPGEN Monte Carlo factorization and renormalization scale uncertainty (8) uncertainty on the mistag model, (9) uncertainty on the non-W flavor composition, and (10) uncertainty on the modeling of the Monte Carlo simulation, as estimated from the most

discrepant shapes in the control variables.

Systematic uncertainties can influence both the expected event yield (normalization) and the shape of the discriminant distribution.

Normalization uncertainties are estimated by recalculating the acceptance using Monte Carlo samples altered due to a specific systematic effect. The single top normalization uncertainty is the difference between the systematically shifted acceptance and the default one and are shown in Table 4.

The effect of the uncertainty in the jet energy scale is evaluated by applying jet-energy corrections that describe $\pm 1\sigma$ variations to the default correction factor. Systematic uncertainties due to the modeling of ISR and FSR are obtained from dedicated Monte Carlo samples where the strength of ISR/FSR was increased and decreased in the parton showering to represent $\pm 1\sigma$ variations [11]. To evaluate the uncertainty associated with the specific choice of parton distribution functions, we use the recommendation from the joint physics group and vary the 40 independent eigenvectors of the CTEQ parton distribution functions and compare to the MRST PDFs. We quadratically sum the uncertainty from the CTEQ and MRST PDF uncertainty if the difference between the CTEQ and MRST PDFs is larger than the CTEQ uncertainty.

The effect of event generator, event detection, b-tag scale factor, and luminosity uncertainty is determined from the background estimate (for the signal template only; the background templates have these numbers included in their Gaussian constraints). The neural network jet flavor separator has a systematically shifted outputs that we employ as systematic uncertainty. ALPGEN Monte Carlo systematic uncertainties in shape are estimated from dedicated samples—rate uncertainties are already included in the background normalization uncertainties.

The uncertainty on the mistag model shape is determined by weighting pretag data with the mistag matrix [12]. The non-W flavor composition is determined from studies made using the neural net b-tagger to estimate the flavor composition of the non-W samples. The uncertainty is estimated by comparing the default flavor ratios (45% b, 40% c, 15% light) with an "extreme" estimate (60% b, 30% c, 10% light).

In order to account for possible mis-modeling in Monte Carlo, we assign a systematic to variables that look sufficiently discrepant to suspect a mis-modeling. We do this by re-weighting the templates using the distributions of that variable in the sideband of events which have at least one taggable jet but no tagged jets. Two variables were chosen for this systematic: the pseudo-rapidity of the last jet, which shows an excess in data in the far forward region, asymmetrically on the east side; and the quantity ΔR between the two jets, in which the Monte Carlo comes close but does not match the data perfectly. Figure 8 shows the distribution and the re-weighting approach.

For all backgrounds the normalization uncertainties are represented by the uncertainty on the predicted number of background events, obtained from Method2 and are

incorporated in the analysis as Gaussian constraints in the likelihood function.

$$\mathcal{L}(\beta_1,...,\beta_5;\delta_1,...,\delta_{11}) = \underbrace{\prod_{k=1}^{B} \frac{e^{-\mu_k}\cdot\mu_k^{n_k}}{n_k!}}_{Poisson\ term} \cdot \underbrace{\prod_{j=2}^{5} G(\beta_j|1,\Delta_j)}_{Gauss\ constraints} \cdot \underbrace{\prod_{i=1}^{11} G(\delta_i,0,1)}_{Systematics} \tag{7}$$

$$where,\ \mu_k = \sum_{j=1}^{5}\beta_j \quad \cdot \quad \underbrace{\left\{\prod_{i=1}^{11}[1+|\delta_i|\cdot(\epsilon_{ji+}H(\delta_i)+\epsilon_{ji-}H(-\delta_i))]\right\}}_{Normalization\ Uncertainty} \tag{8}$$

$$\underbrace{\cdot\ \alpha_{jk}}_{Shape\ P.} \quad \cdot \quad \underbrace{\left\{\prod_{i=1}^{11}(1+|\delta_i|\cdot(\kappa_{jik+}H(\delta_i)+\kappa_{jik-}H(-\delta_i)))\right\}}_{Shape\ Uncertainty} \tag{9}$$

All systematic normalization and shape uncertainties are incorporated in the analysis into the likelihood as nuisance parameters, conform with a fully Bayesian treatment [13]. We take the correlation between normalization and shape uncertainties for a given source into account [14]. The relative strength of a systematic effect due to the source i is parameterized by the nuisance parameter δ_i in the likelihood function, constrained to a unit-width Gaussian (last term in Equation 7). The $\pm1\sigma$ changes in the normalization of process j due to the i^{th} source of systematic uncertainty are denoted by $\epsilon_{ji}+$ and $\epsilon_{ji}-$ (see Equation part 8). The $\pm1\sigma$ changes in bin k of the H_T templates for process j due to the i^{th} source of systematic uncertainty are quantified by κ_{jik+} and κ_{jik-} (see Equation part 9). $H(\delta_i)$ represents the Heaviside function, defined as $H(\delta_i) = 1$ for $\delta_i > 0$ and $H(\delta_i) = 0$ for $\delta_i < 0$. The Heaviside function is used to separate positive and negative systematic shifts (for which we have different normalization and shape uncertainties). The variable δ_i appears in both the term for the normalization (Equation 8) and the shape uncertainty (Equation 9), which is how correlations between both effects are taken into account.

We marginalizing the likelihood function by integrating $\mathcal{L}(\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\delta_1,..\delta_{11})$ over the nuisance parameters $\beta_2,\beta_3,\beta_4,\beta_5,\delta_1,..\delta_{11}$ for many possible values of the single top cross-section β_1 from [0..5]. The resulting reduced likelihood $\mathcal{L}(\beta_1)$ is a function of the single top cross-section β_1 only.

We list all systematic acceptance changes due to systematic uncertainties (aside from the uncertainty in the background estimate, which includes luminosity, generator, b-tagging, and heavy flavor uncertainties) in Table 4 and Table 5.

Figures 9 through 16 show the shifts in the distributions caused by shifts of one standard deviation in the different sources of uncertainty.

Systematic	Process	1 tag	2 tag
JES	Single top	−0.6% / +0.3%	+2.2% / −1.1%
JES	b-like	+6.0% / −6.6%	+9.7% / −9.9%
JES	c-like	+6.1% / −5.6%	+10.0% / −10.0%
JES	$t\bar{t}$	+9.9% / −9.4%	+8.5% / −7.6%
ISR	Single top	+1.9% / +2.1%	+0.3% / +6.6%
ISR	$t\bar{t}$	−2.6% / −7.1%	+0.6% / −9.4%
FSR	Single top	+4.8% / −0.7%	+7.5% / +0.8%
FSR	$t\bar{t}$	−5.1% / −2.6%	−8.0% / −1.7%
PDF	Single top	−3.0% / +3.0%	−2.0% / +2.0%
PDF	$t\bar{t}$	−1.8% / +1.8%	−1.7% / +1.7%
m_{top}	Single top	+7.3% / −6.2%	+9.0% / −7.6%
m_{top}	$t\bar{t}$	+7.8% / −8.1%	+7.8% / +8.1%

Table 4: Rate systematic uncertainties used in this analysis for the $W + 2$ jet bin . The m_{top} samples are used not as a systematic but as a different mass point for the measurement.

Systematic	Process	1 tag	2 tag
JES	Single top	−9.4% / +10.3%	−8.3% / +4.2%
JES	b-like	+6.8% / −6.3%	+9.5% / −10.0%
JES	c-like	+4.7% / −5.6%	+15.4% / −11.4%
JES	$t\bar{t}$	+4.6% / −5.1%	+5.5% / −5.1%
ISR	Single top	−3.3% / −4.8%	+5.8% / −5.0%
ISR	$t\bar{t}$	−0.6% / −4.5%	−0.5% / −6.6%
FSR	Single top	−3.3% / −3.8%	+2.2% / −2.3%
FSR	$t\bar{t}$	−3.4% / −2.2%	−3.4% / −2.7%
PDF	Single top	−2.6% / +2.6%	−1.9% / +1.9%
PDF	$t\bar{t}$	−1.8% / +1.8%	−1.7% / +1.7%
m_{top}	Single top	+7.4% / −6.3%	+8.0% / −6.8%
m_{top}	$t\bar{t}$	+7.8% / −8.1%	7.8% / −8.1%

Table 5: Rate systematic uncertainties used in this analysis for the $W + 3$ jet bin. The m_{top} samples are used not as a systematic but as a different mass point for the measurement.

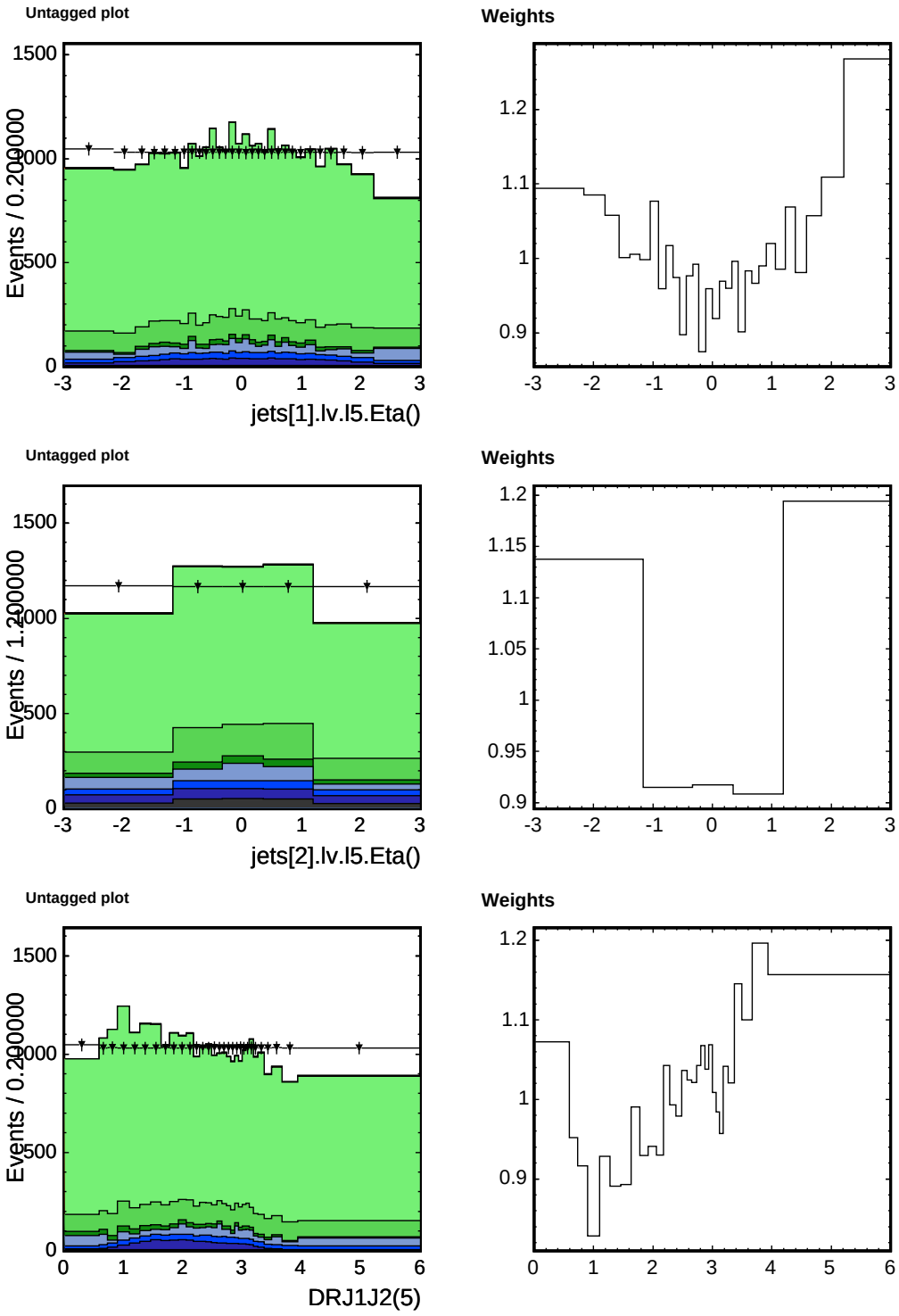

Figure 8: Distributions which suggest a mis-modeling in Monte Carlo, binned so as to be flat in data (to reduce the effect of statistical uncertainties) and the resulting weights assigned: pseudorapidity of the second jet in the two-jet bin, pseudorapidity of the third jet in the three-jet bin, and ΔR between the jets in the two-jet bin.

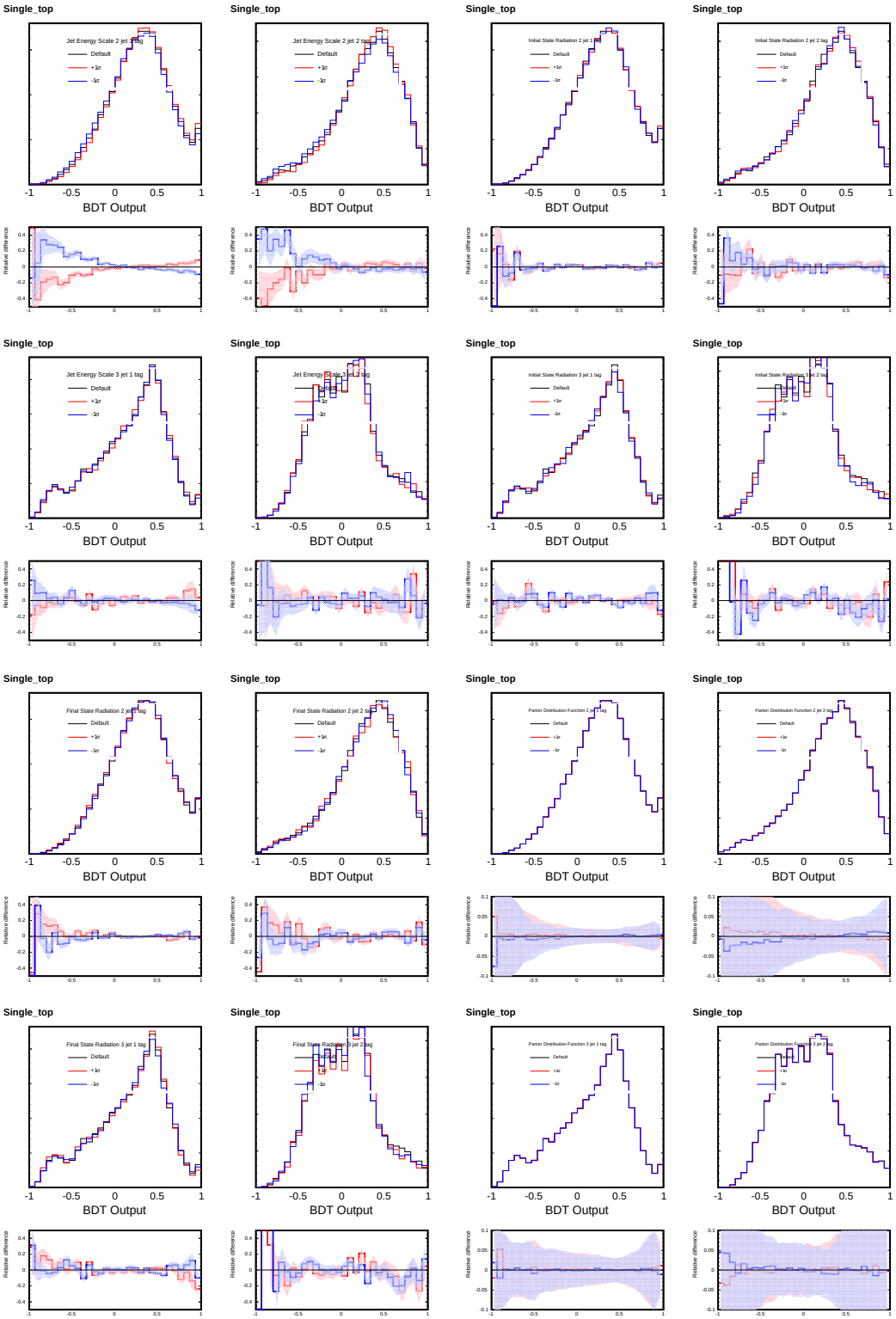

Figure 9: Shape systematics for the single top template.

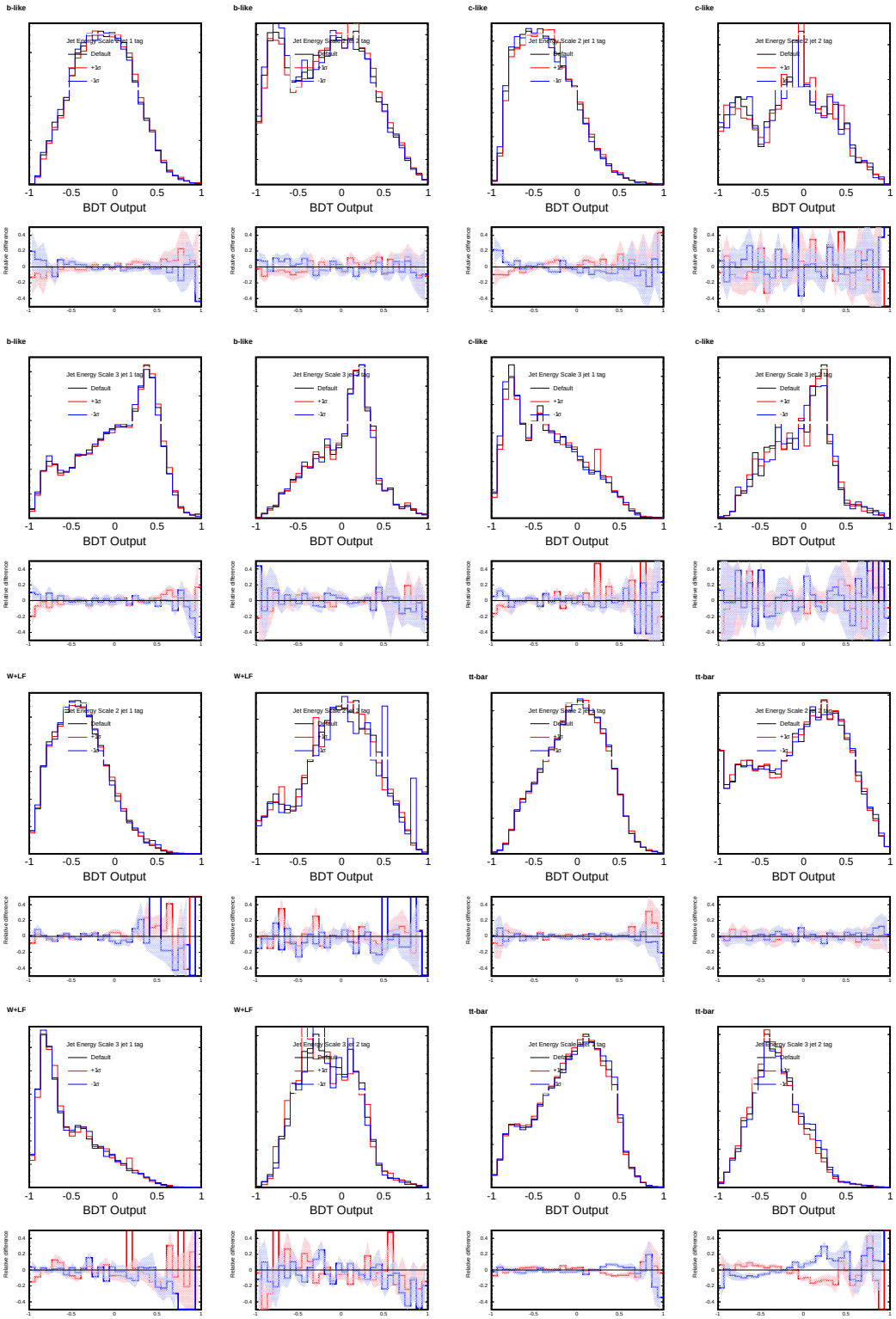

Figure 10: JES systematic uncertainty evaluated in background events.

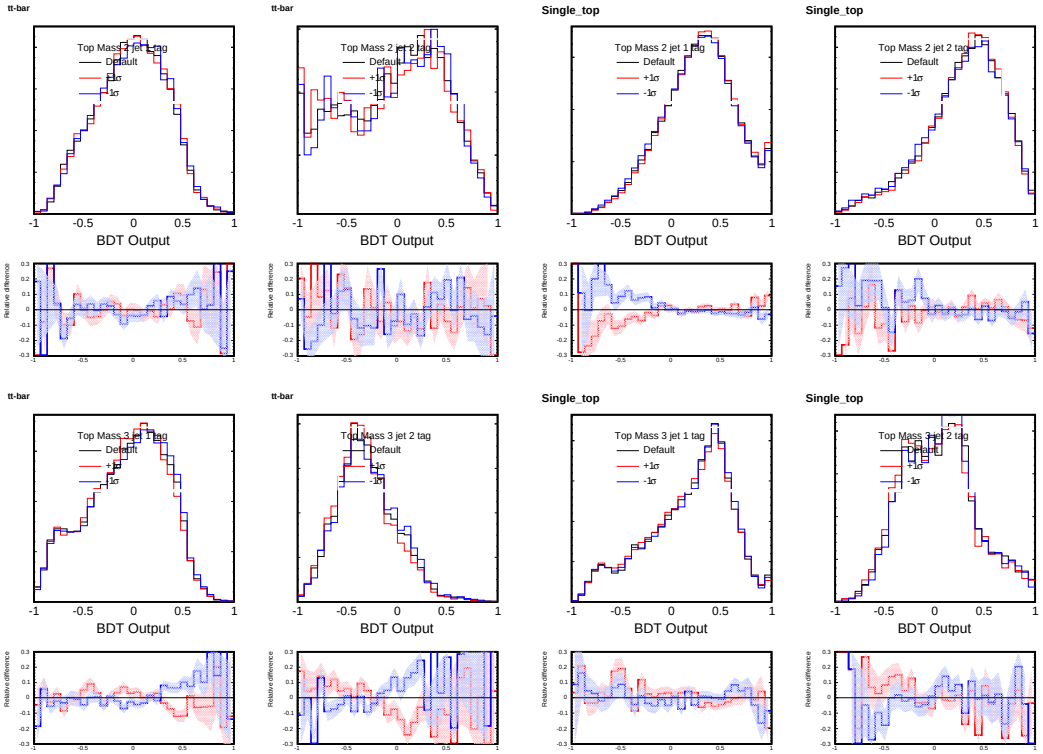

Figure 11: Shape systematics for the single top and $t\bar{t}$ templates due to the top mass uncertainty. Top masses of 170 and 180 GeV/c^2 are taken as 2σ variations.

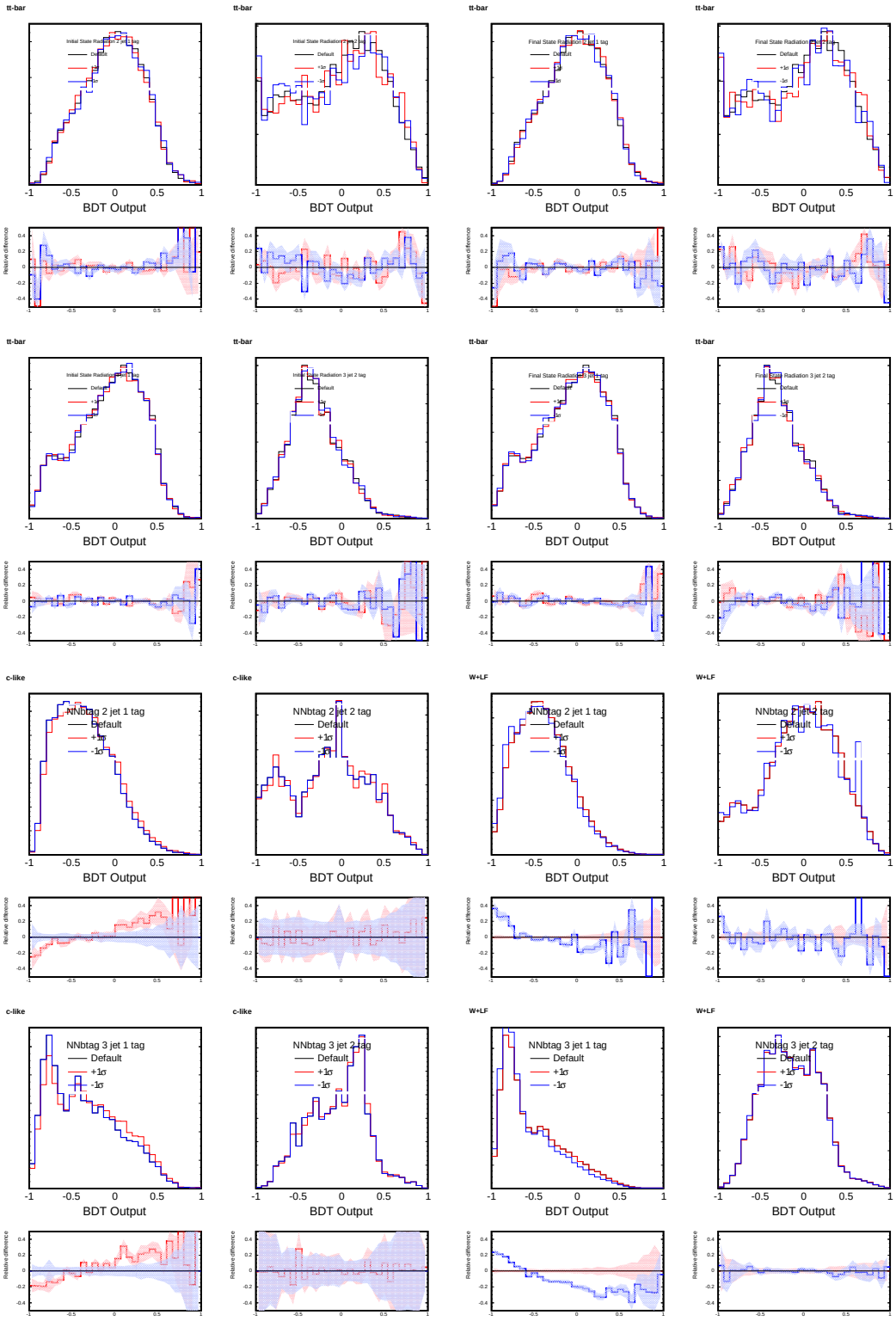

Figure 12: Shape systematics for initial and final state radiation for the $t\bar{t}$ template. And neural net jet flavor separator systematic uncertainty evaluated in background events.

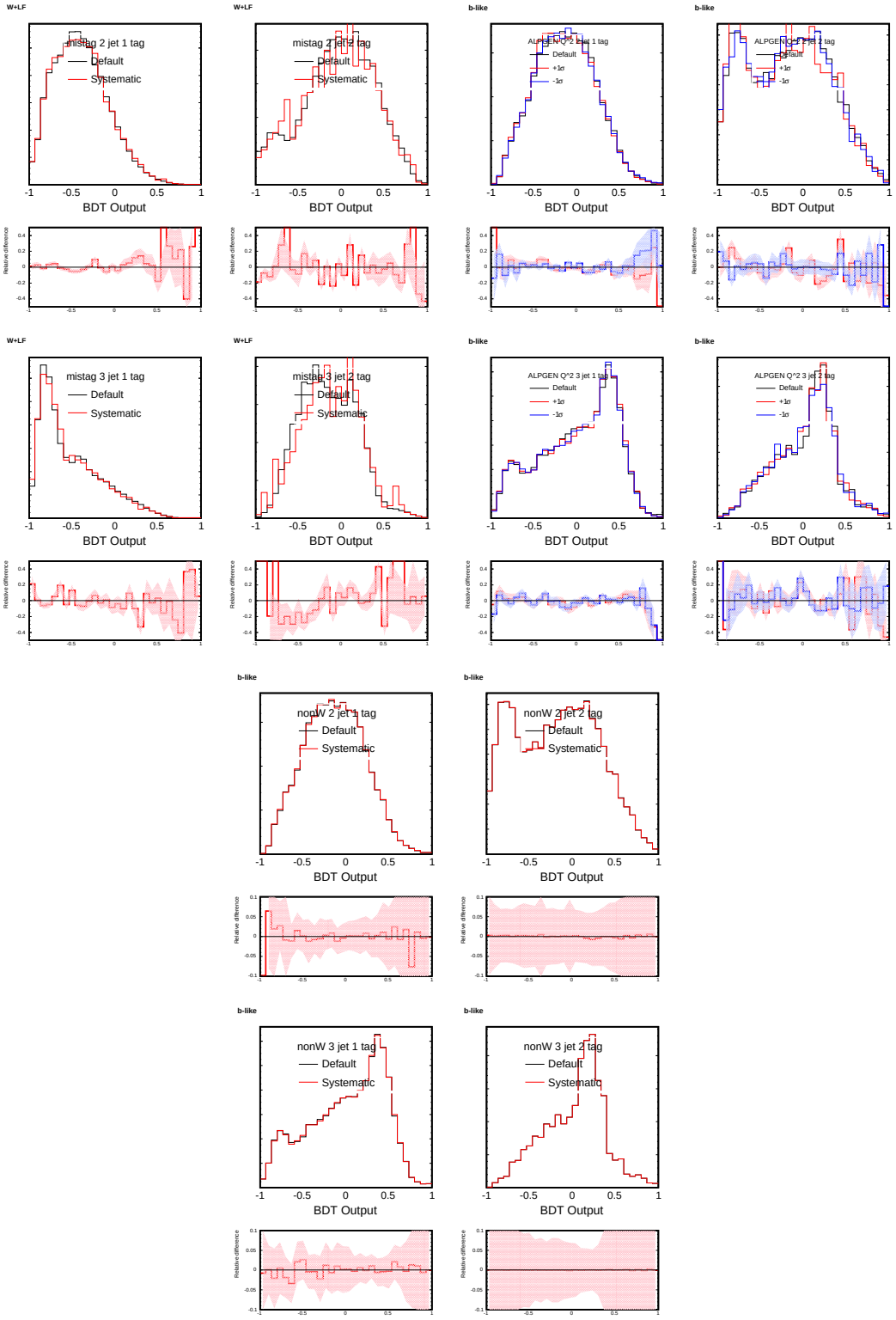

Figure 13: Background systematic uncertainties, including mistag model uncertainty, ALPGEN factorization and renormalization scale uncertainty, and non-W model uncertaintiy.

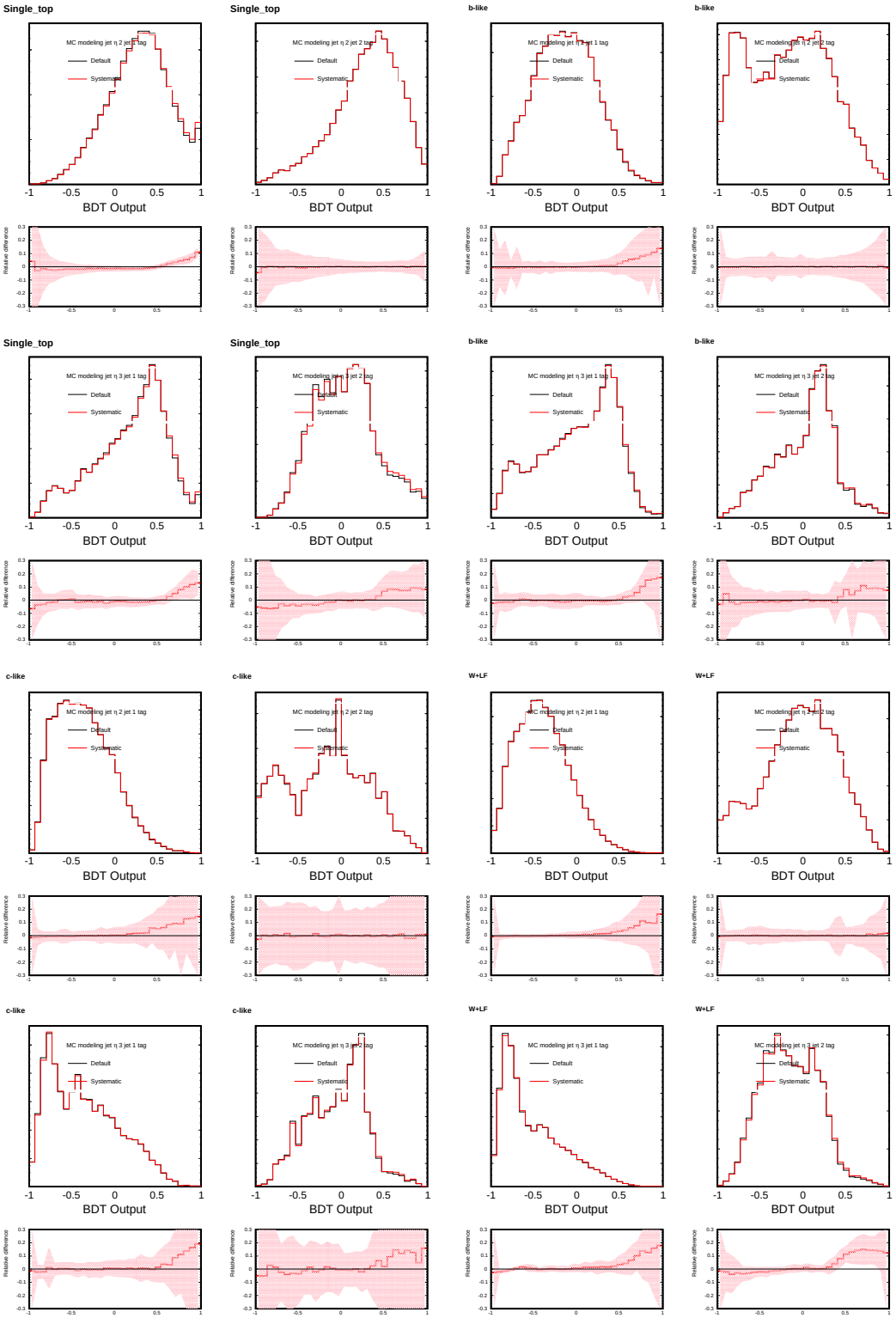

Figure 14: Systematic uncertainty due to mis-modeling of last jet pseudo-rapidity.

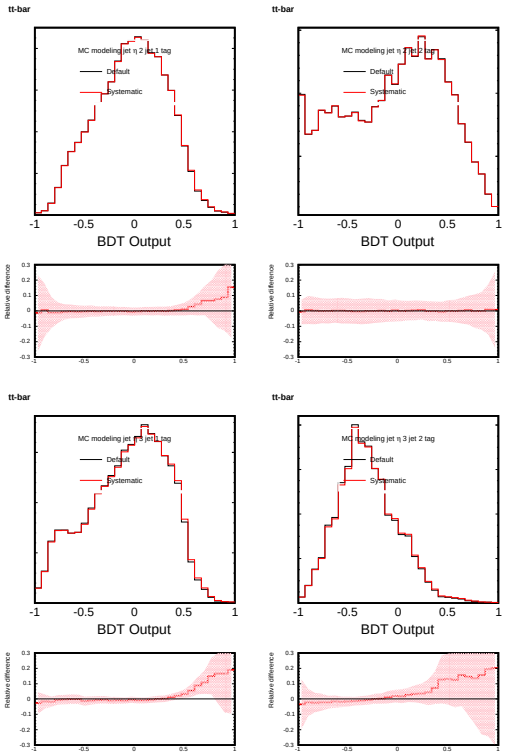

Figure 15: Systematic uncertainty due to mis-modeling of last jet pseudo-rapidity.

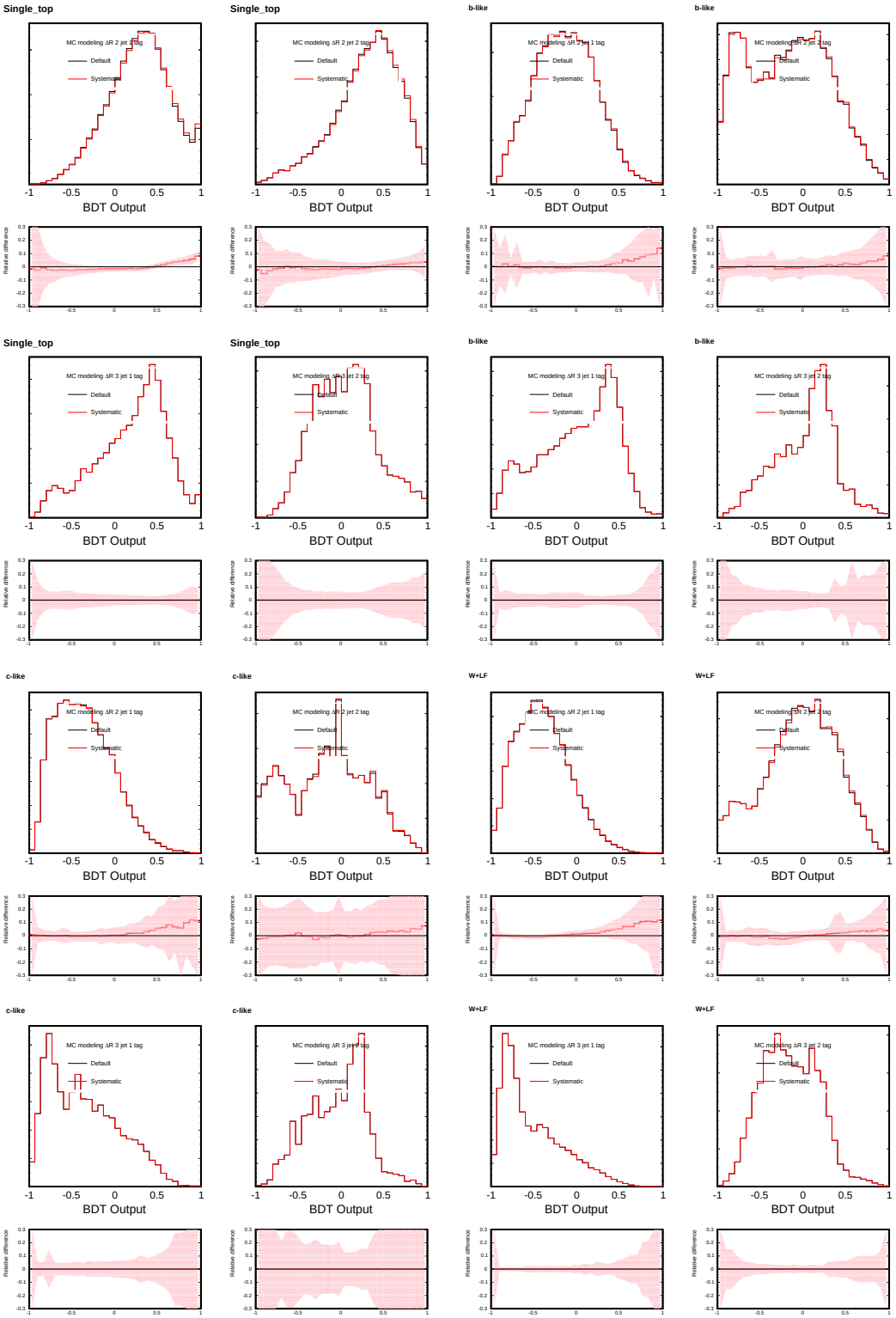

Figure 16: Systematic uncertainty due to mis-modeling of ΔR between the two jets.

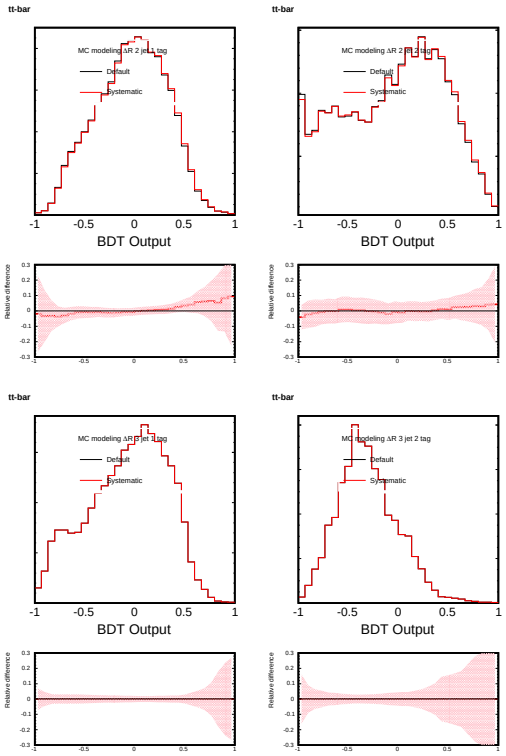

Figure 17: Systematic uncertainty due to mis-modeling of ΔR between the two jets.

3.0.1 Multiple Fit Regions

To increase the sensitivity of the analysis, we create separate templates for the untriggered muons, which have a substantially different background composition from the rest of the subdetectors. We create templates for single and double tags separately, and for the two-jet and three-jet bins separately. We fit for one common cross section across all eight different regions when we perform the final fit.

4 Result with CDF II Data

We apply the analysis to 2.7 fb^{-1} of CDF Run II data. We compare the BDT output distributions of 2247 candidate events with the sum of predicted single top and background templates (see Figure 18 for the unblinded discriminant distributions).

In order to extract the most probable single top content in the data we perform the maximum likelihood method described in Section 2.5. The posterior p.d.f is obtained by using Bayes' theorem:

$$p(\beta_1|data) = \frac{\mathcal{L}^*(data|\beta_1)\pi(\beta_1)}{\int \mathcal{L}^*(data|\beta_1')\pi(\beta_1')d\beta_1'}$$

where $\mathcal{L}^*(data|\beta_1)$ is the reduced likelihood and $\pi(\beta_1)$ is the prior p.d.f. for β_1. We adopt a flat prior, $\pi(\beta_1) = H(\beta_1)$, in this analysis, with H being the Heaviside step function.

The most probable value (MPV) corresponds to the most likely combined single top production cross section given the data. The uncertainty corresponds to the range of highest posterior probability density which covers 68.27% [15].

We perform marginalization using the likelihood function of Equation 7 with all systematic rate and shape uncertainties included in the likelihood function. The most probable value for the single top cross section is obtained at $2.4^{+0.8}_{-0.7}$ pb. The posterior probability density is shown in Figure 19.

5 Expected Sensitivity and Hypothesis Test

We interpret the result using the CLs/CLb method developed at LEP [16]. We compare our data against two models, one asserting that the data is due to background processes only (b) and one which includes Standard Model single top production in addition to the background processes (s+b). We propagate all systematic uncertainties in our statistical method as described in the next section 3. Using the test statistic $Q = \frac{\mathcal{L}(data)|s+b}{\mathcal{L}(data)|b}$ we compute the probability (p-value) hat the background only (b) model fluctuated equal or up to the observed value Q_{obs} in the data (observed p value) and to the median Q value of signal+background (s+b) pseudo-experiments (expected p-value). Figure 20 shows the distribution of the test statistics for pseudo-experiments performed for (b) and (s+b). We expect a p-value of 3.8×10^{-7} % (5.0σ). and observe a p-value of 0.00017% (3.6σ) in the data.

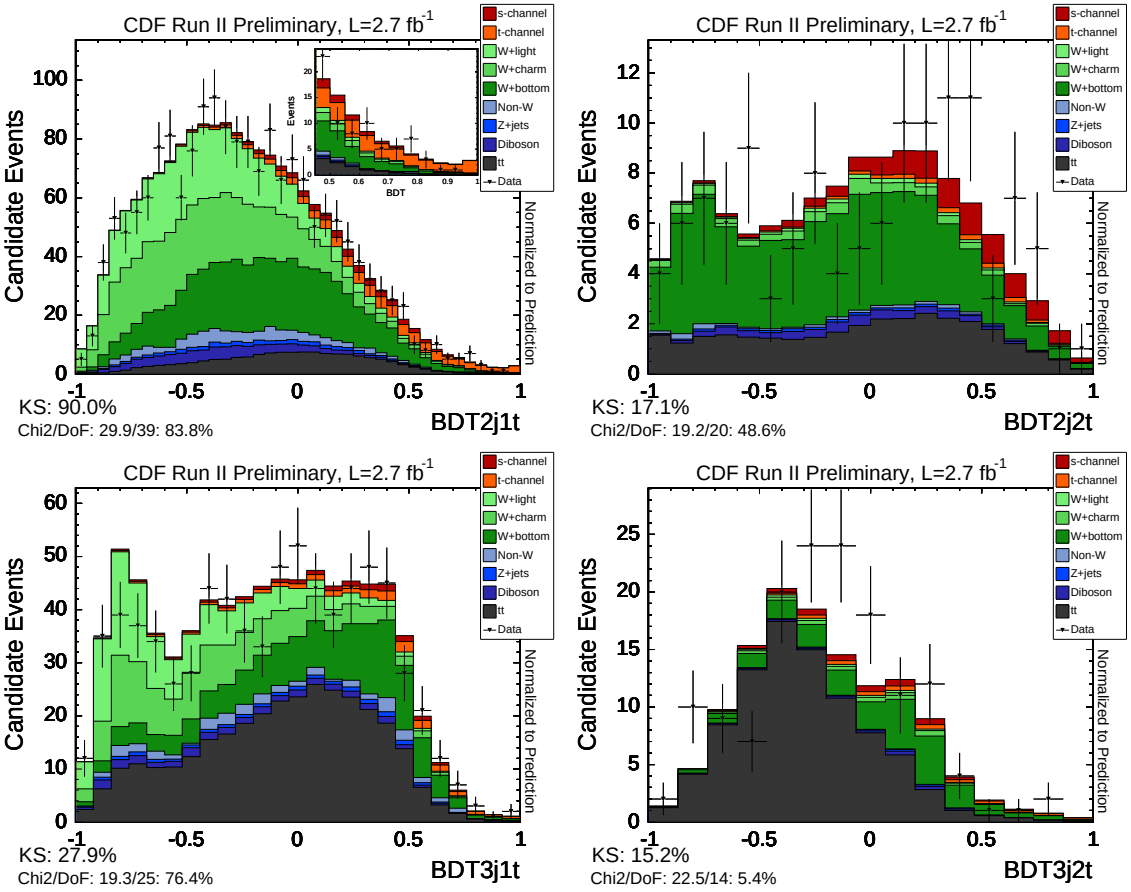

Figure 18: CDF data compared to Monte Carlo prediction for signal and background. The top plots show the two-jet bin while the bottom plot shows the three-jet bin; the left-hand plots show the single-tagged events while the right-hand plots show the double-tagged events.

6 Conclusions

We have used the matrix element analysis technique in a direct search for electroweak single top quark production. Our search was done simultaneously for s-channel and t-channel single top production. To extract the most probable single top content in data, we apply a maximum likelihood technique. All sources of systematic rate and shape uncertainty are included in the likelihood function. We have analyzed 2.7 $fb^{-1}$of CDF Run II data and measure a combined s- and t-channel single top production cross section of:

$$\sigma_{singletop} = 2.4^{+0.8}_{-0.7} \text{ pb (all systematics included)}$$

assuming a top quark mass of 175 GeV/c^2. We use the MCLIMIT program [16] to calculate the signal significance. The observed p-value in 2.7fb^{-1} of CDF data is 0.00017 (3.6σ). The expected (median) p-value in pseudo-experiments is 3.8×10^{-7} % (5.0σ).

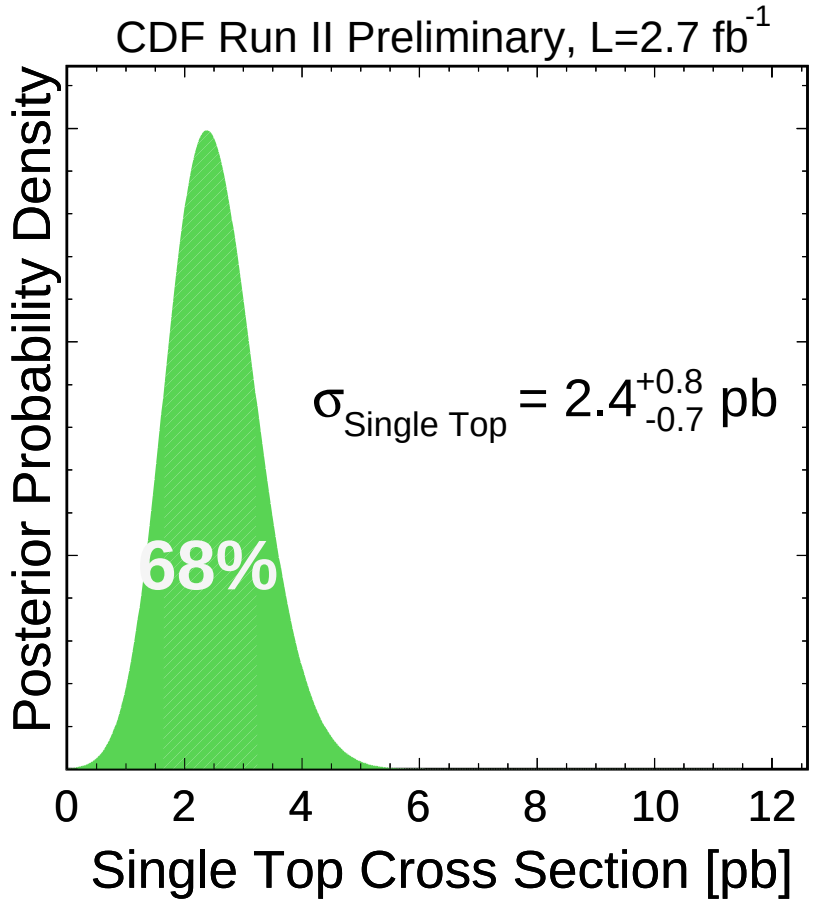

Figure 19: Marginalized likelihood of the single top cross section using $2.7 fb^{-1}$of data. The error band shows the 68% uncertainty (all systematics included) on the measurement.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; and in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

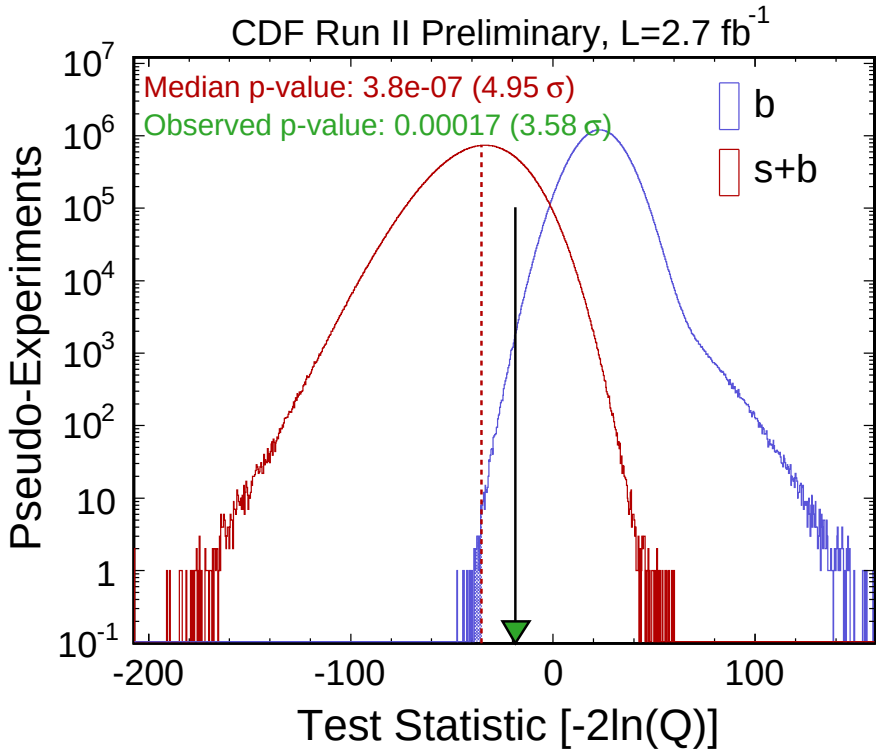

Figure 20: Distribution of the test statistics $Q = \frac{\mathcal{L}(data)|s+b}{\mathcal{L}(data)|b}$ for the single top + background hypothesis (s+b) and the background only hypothesis (b) in 90 million pseudo-experiments. The arrow indicates the observed value in 2.7 fb^{-1} of CDF II data. The dashed line shows the median of the (s+b) hypothesis which is used to determine the expected p-value.

References

[1] J. Adelman *et al. Method II For You*, CDF 9285 (2008)

[2] F. Canelli, B. Casal Laraña, C. Group, A. Ruiz, B. Stelzer, R. Vilar *Increasing Muon Acceptance with MET Plus Jet Triggers*, CDF Note 8489 (2006)

[3] A. Hocker, P. Speckmayer, J. Stelzer, F. Tegenfeldt, H. Voss, K. Voss, A. Christov, S. Henrot-Versille, M. Jachowski, A. Krasznahorkay Jr., Y. Mahalalel, R. Ospanov, X. Prudent, M. Wolter, A. Zemla, *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039v4 (2007)

[4] F. Canelli, B. Casal Laraña, P. Dong, C. Group, A. Ruiz, B. Stelzer, R. Vilar *Search for Single Top Quark Production using Boosted Decision Trees in 2.2fb*, CDF 9263 (2008)

[5] D. Amidei *et al. Calibration of Heavy-Flavor Production in W+1 Jet Data*, CDF 9187 (2008)

[6] S. Budd *et al. Combination of CDF Single Top Quark Searches with 2.2 fb^{-1} of Data*, CDF Note 9251 (2008)

[7] Y. Freund and R.E. Schapire, J. of Computer and System Science 55, 119 (1997)

[8] L. Breiman, J. Friedman, R. Olshen and C. Stone, *Classification and Regression Trees*, Wadsworth (1984).

[9] S. Richter *et al.*, *A Neural Network b Tagger for Single-Top Analyses*, CDF 7816 (2005)

[10] T. Chwalek *et al.*, *Update of the neural network b tagger for single-top analyses*, CDF 8903 (2007)

[11] Y.-K. Kim, U.-K. Yang, *Initial state gluon radiation studies on Drell-Yan data for top-pair production in hadron Collider*, CDF Note 6804 (2003)

[12] Guimaraes da Costa, Sherman *Study of SecVtx Fake Rates with $1fb^{-1}$*, CDF Note 8263 (2006) *Loose and Tight SecVtx Tag Matrices with $1fb^{-1}$*, CDF Note 8264 (2006)

[13] L. Demortier, *Bayesian treatments of Systematic Uncertainties*, Proceedings of Advanced Statistical Techniques in Particle Physics, Grey College, Durham, 18 - 22 March 2002, http://www.ippp.dur.ac.uk/Workshops/02/statistics/proceedings.shtml

[14] C. Ciobanu, T. Junk, T. Müller, P. Savard, B. Stelzer, W. Wagner, T. Walter, *Likelihood Function for Single Top Search with 162 pb^{-1}*, CDF Note 7106 (2004)

[15] Particle Data Group, *The Review of Particle Physics*, http://pdg.lbl.gov (2004)

[16] L. Read, J.Phys G **28**, 2693 (2002) and T. Junk, Nucl. Instum. Meth. **434**, 435 (1999). See also P. Bock *et al.* (The LEP Collaborations), CERN-EP-98-046 (1998) and CERN-EP-2000-055 (2000).

[17] Sarah Budd, Matthias Buehler, Catalin Ciobanu, Peter Dong, Thomas Junk, Jan Lueck, Thomas Muller, Svenja Richter, Bernd Stelzer, Wolfgang Wagner, Rainer Wallny *Mistag Model Used for Single-Top Summer 2006 Analyses*, CDF Note 8490 (2006)

[18] Sarah Budd, Matthias Buehler, Catalin Ciobanu, Peter Dong, Thomas Junk, Jan Lueck, Thomas Muller, Svenja Richter, Bernd Stelzer, Wolfgang Wagner, Rainer Wallny *Estimation and Modeling of Non-W Background for Single-Top Searches*, CDF Note 8489 (2006)

3 3 3 3 3 3 6 7 9 9 17 17 18 18 30 30, 31

A Validation of Input Variables

Figures 21-88 show the data MC comparisons of all the BDT input variables in the untagged control regions as well as in the signal regions. In the untagged samples the b-quarq jet is chosen as the leading jet. The KIT falvor separator is randomized according to a mistag distribution. These distributions are for the 2.2 $fb^{-}1$ data sample (up to period 13). The newest data period (p14-p17) sample compared with the MC prediction is included in Appendix B.

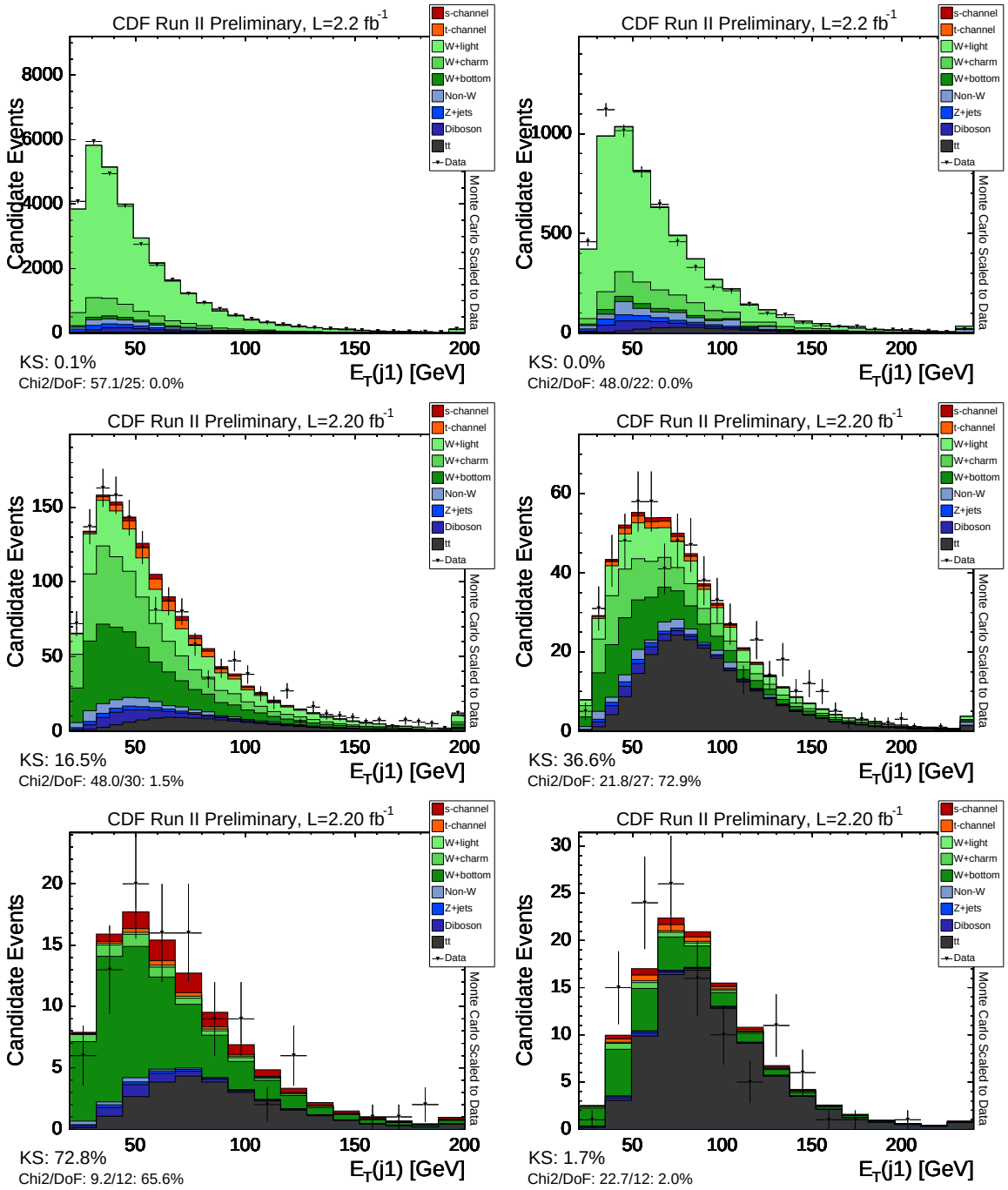

Figure 21: E_T of the leading jet. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

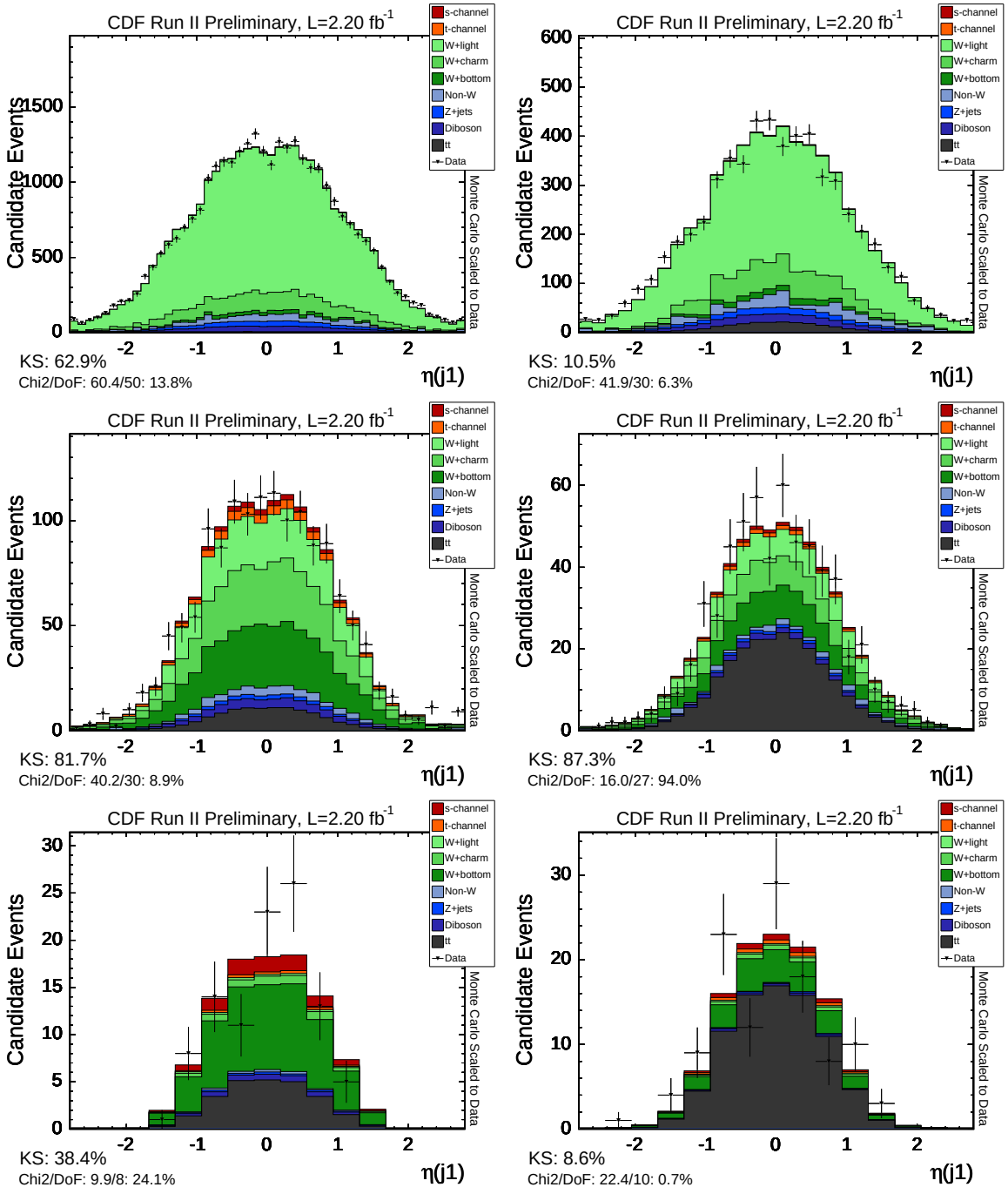

Figure 22: η of the leading jet. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

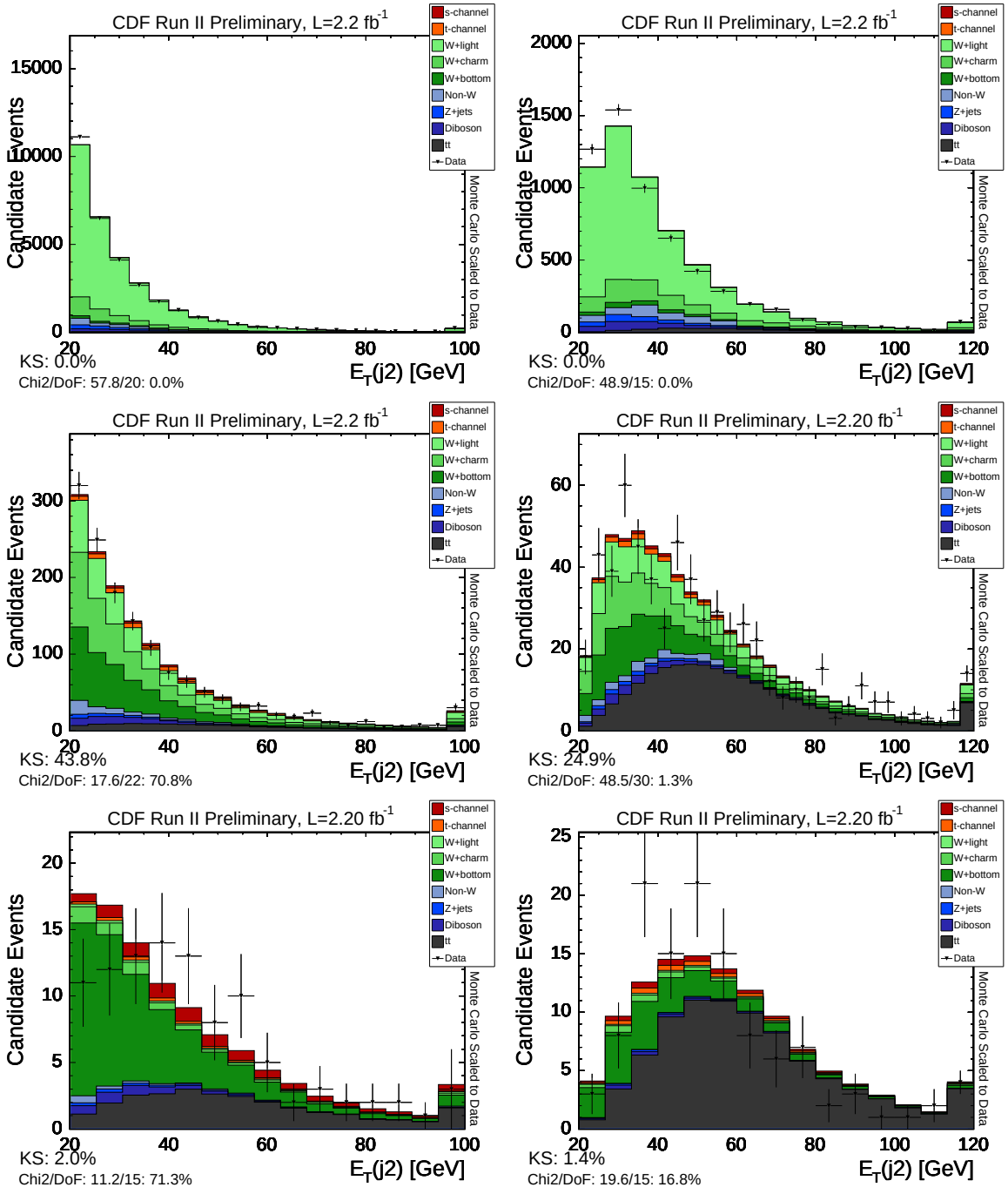

Figure 23: E_T of the 2nd jet. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

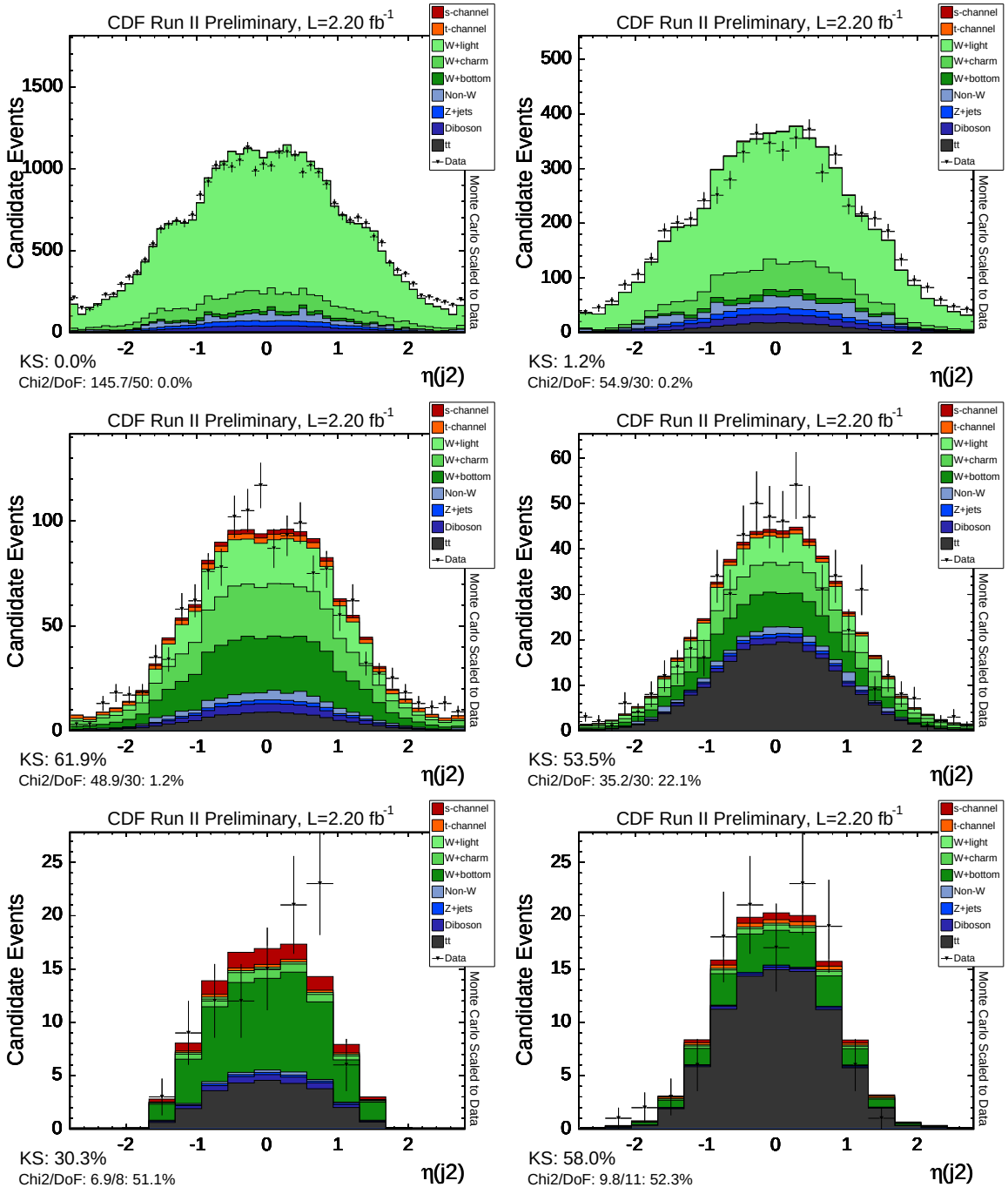

Figure 24: η of the 2nd jet. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

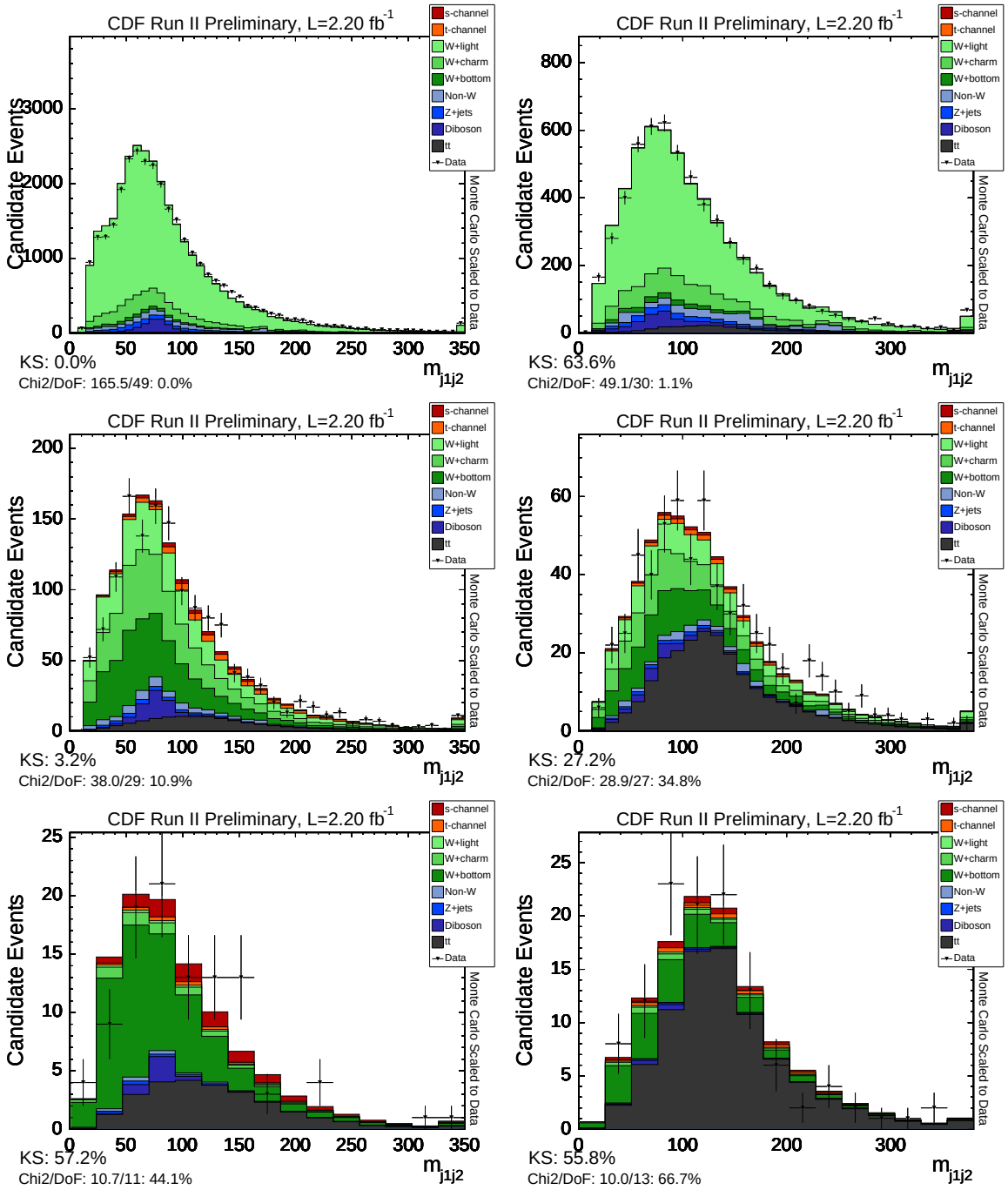

Figure 25: Invariant dijet mass. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

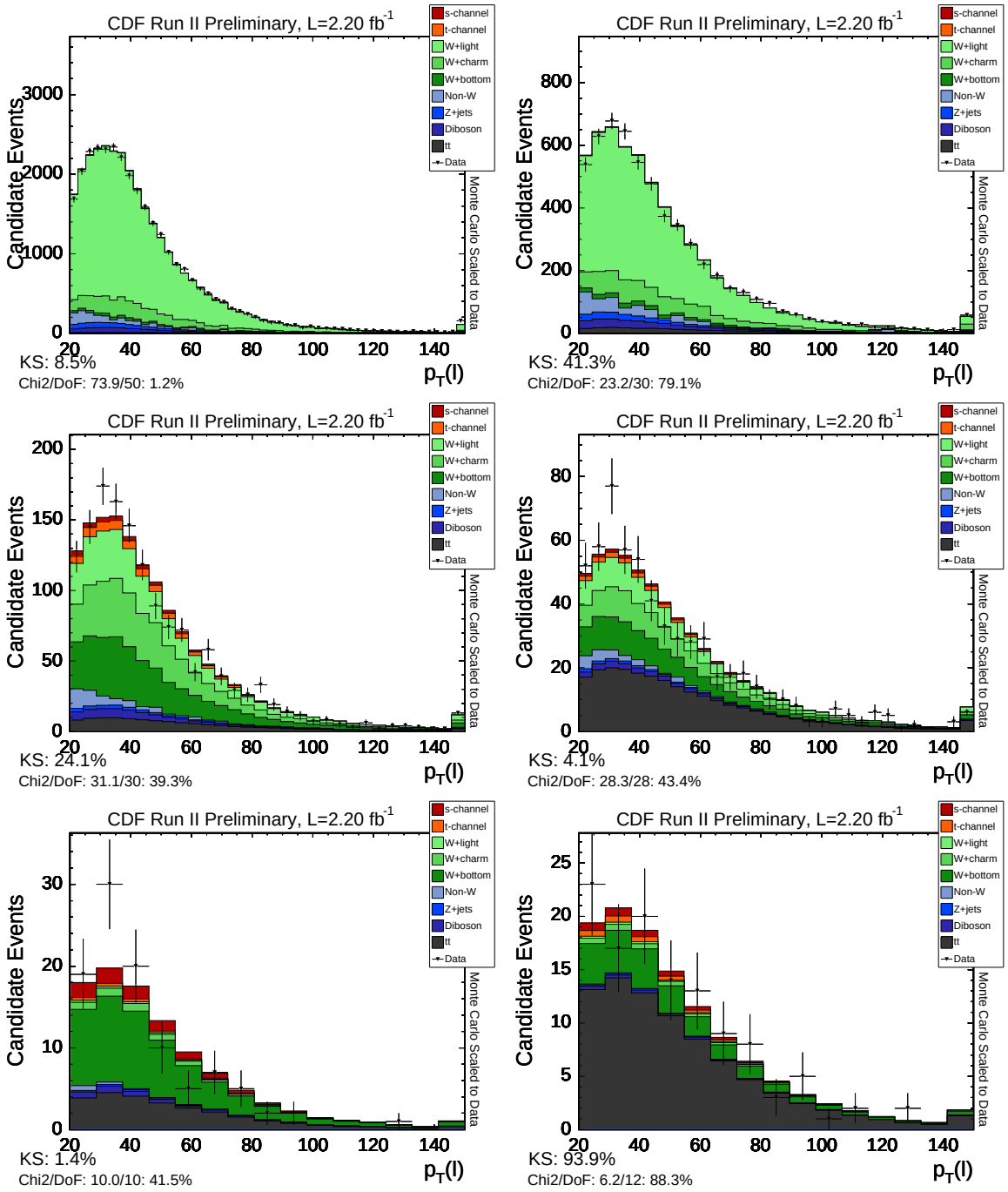

Figure 26: Lepton p_T. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

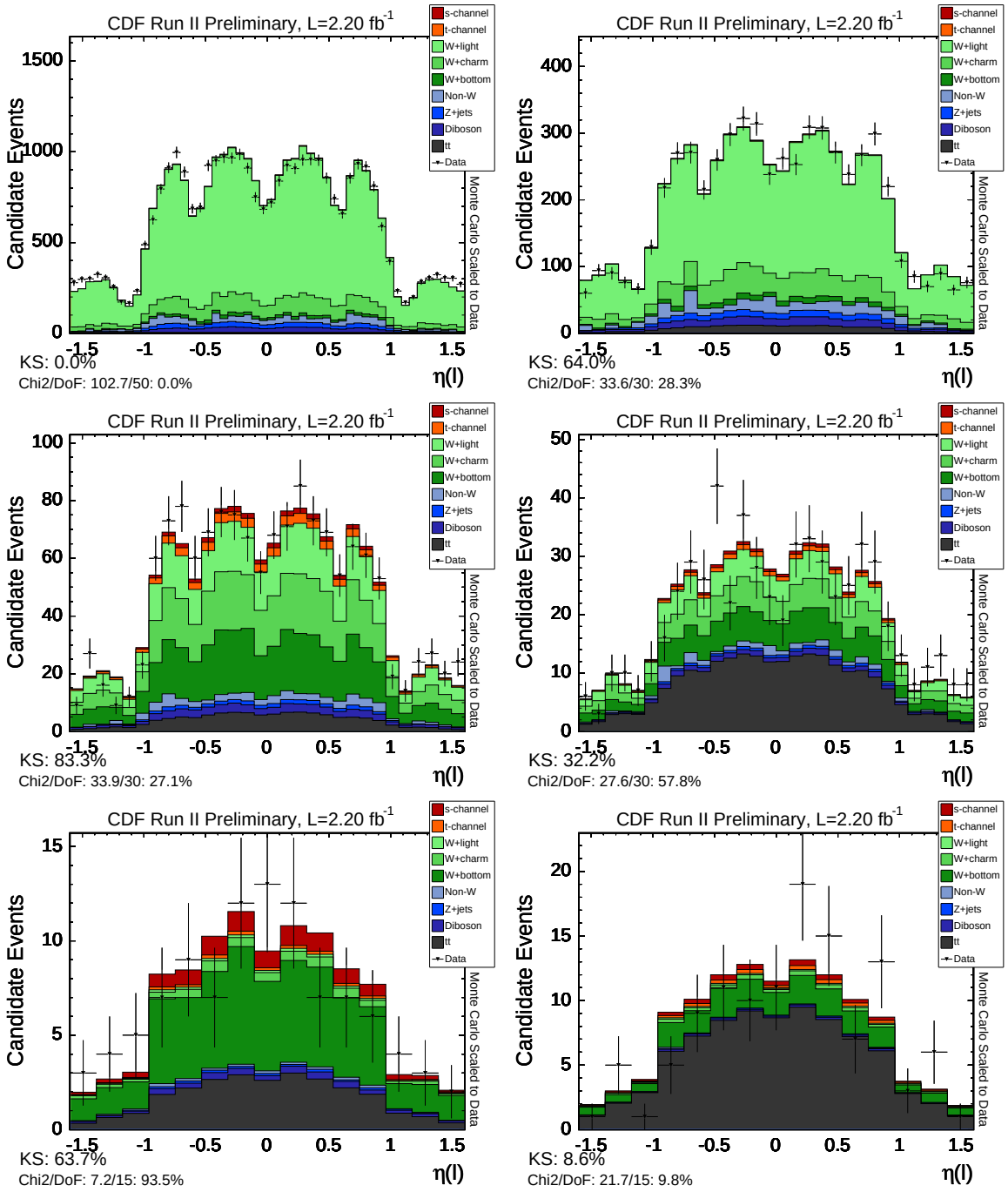

Figure 27: Lepton η. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

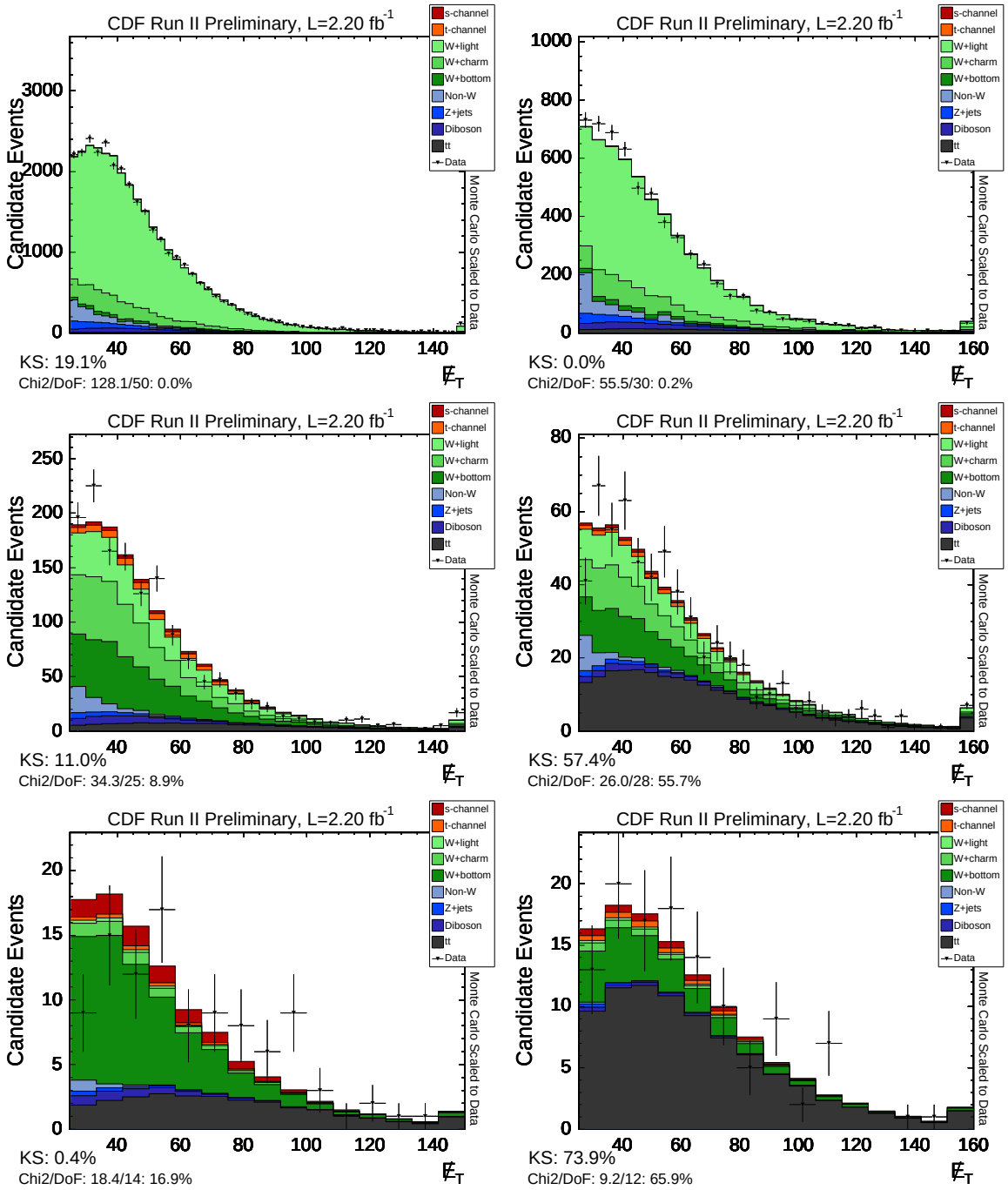

Figure 28: Missing Transverse Energy. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

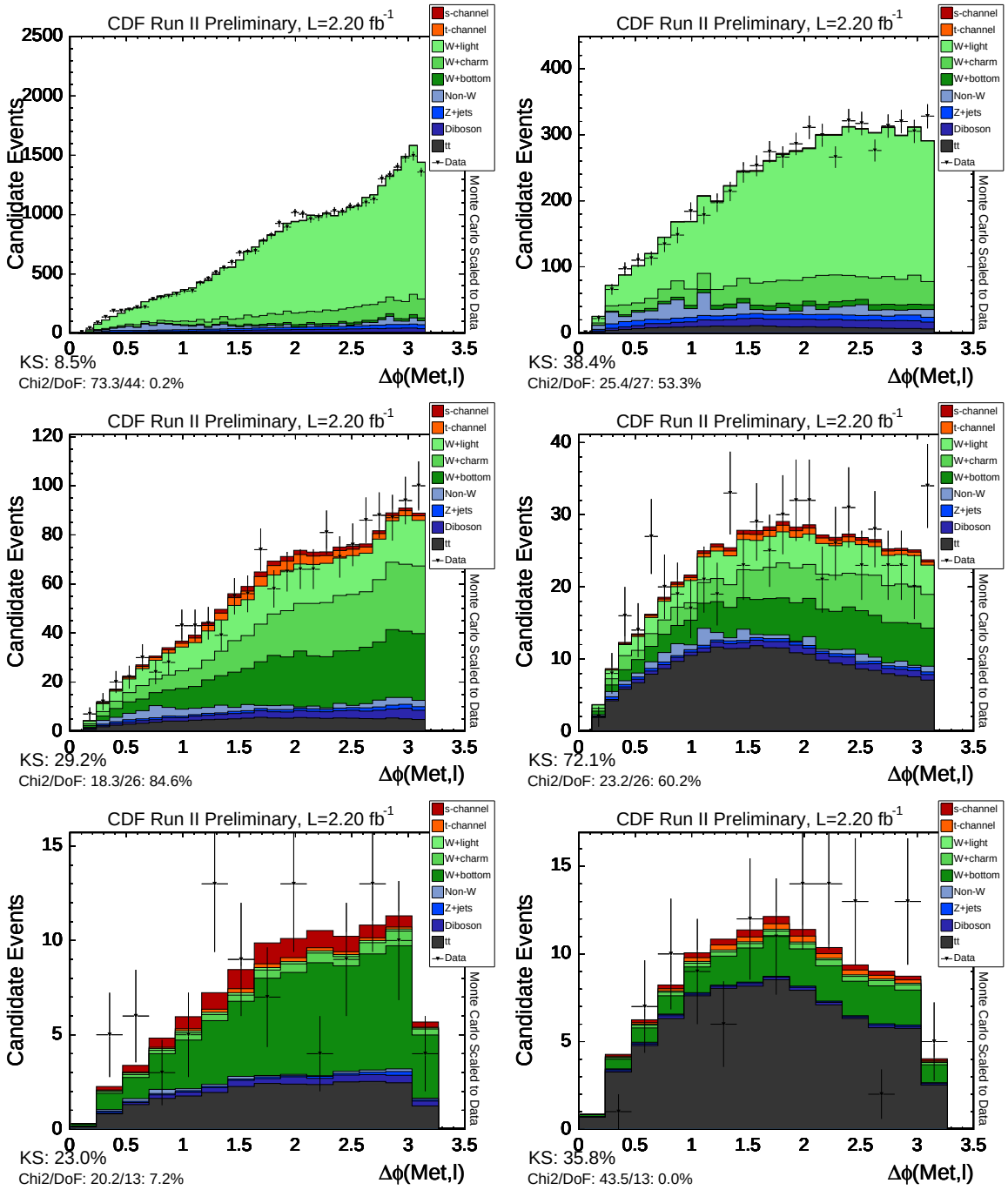

Figure 29: $\Delta\Phi$ between $\not\!\!E_T$ and lepton. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

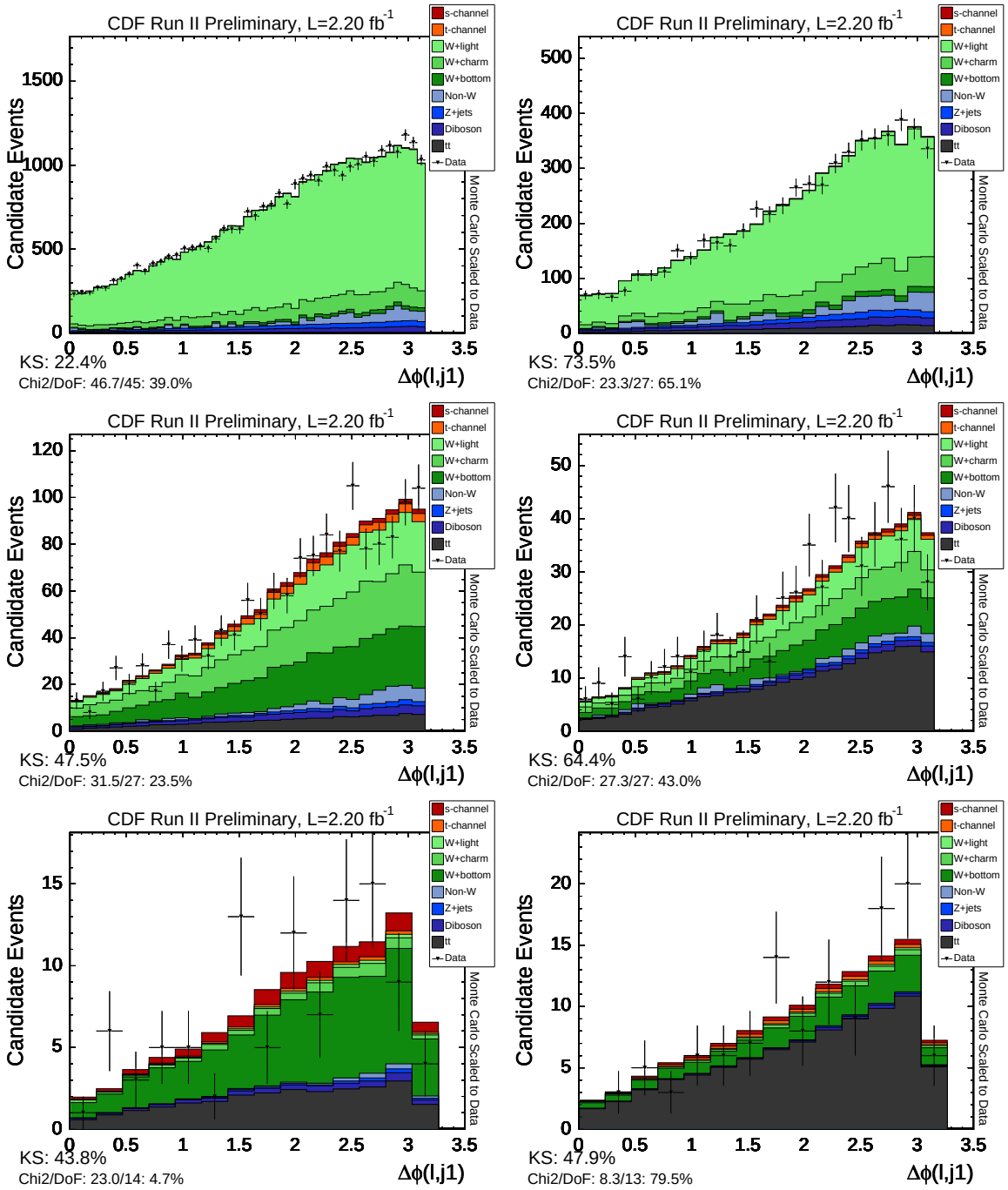

Figure 30: $\Delta\Phi$ between leading jet and lepton. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

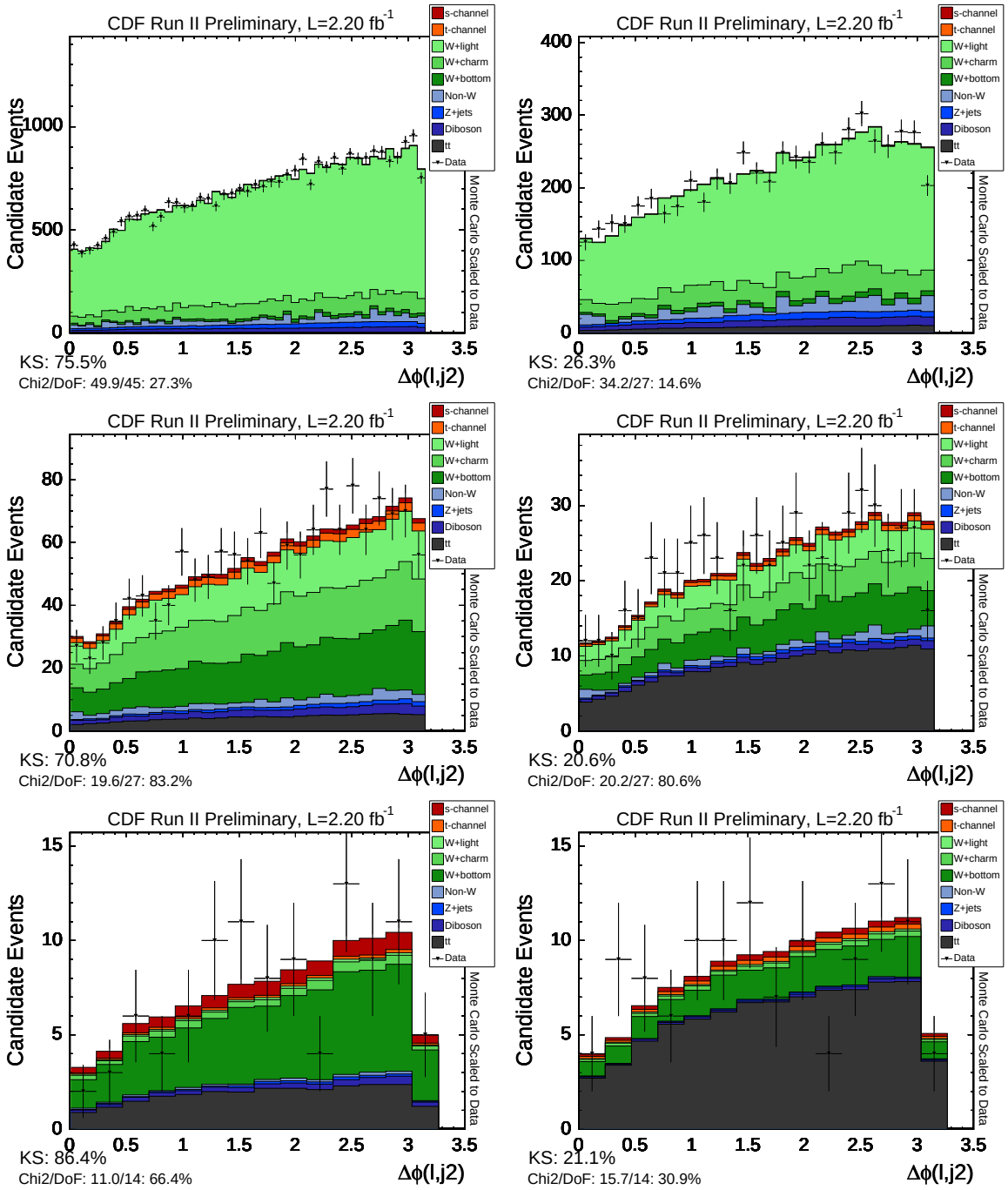

Figure 31: $\Delta\Phi$ between 2nd jet and lepton. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

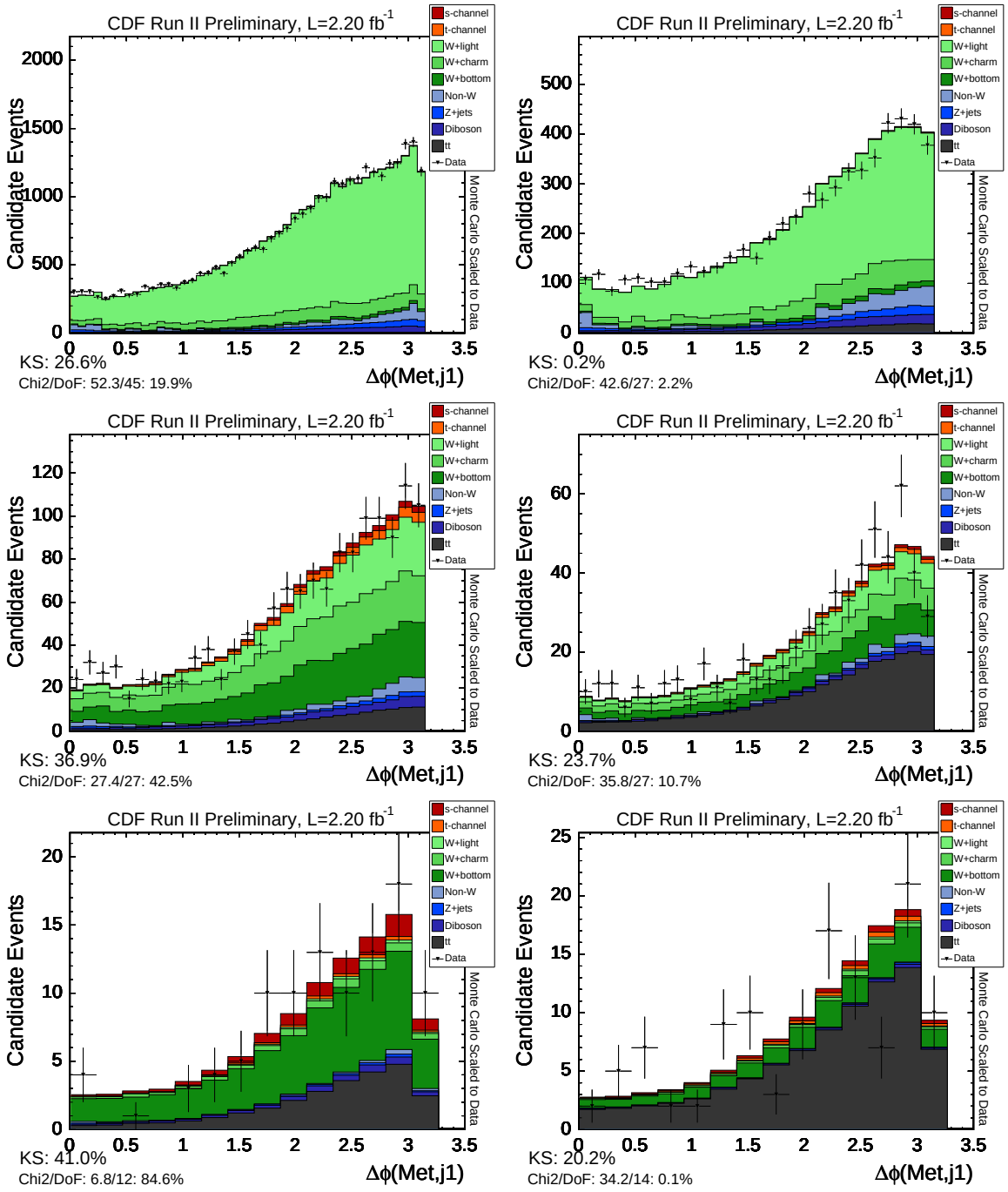

Figure 32: $\Delta\Phi$ between leading jet and $\not\!\!E_T$. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

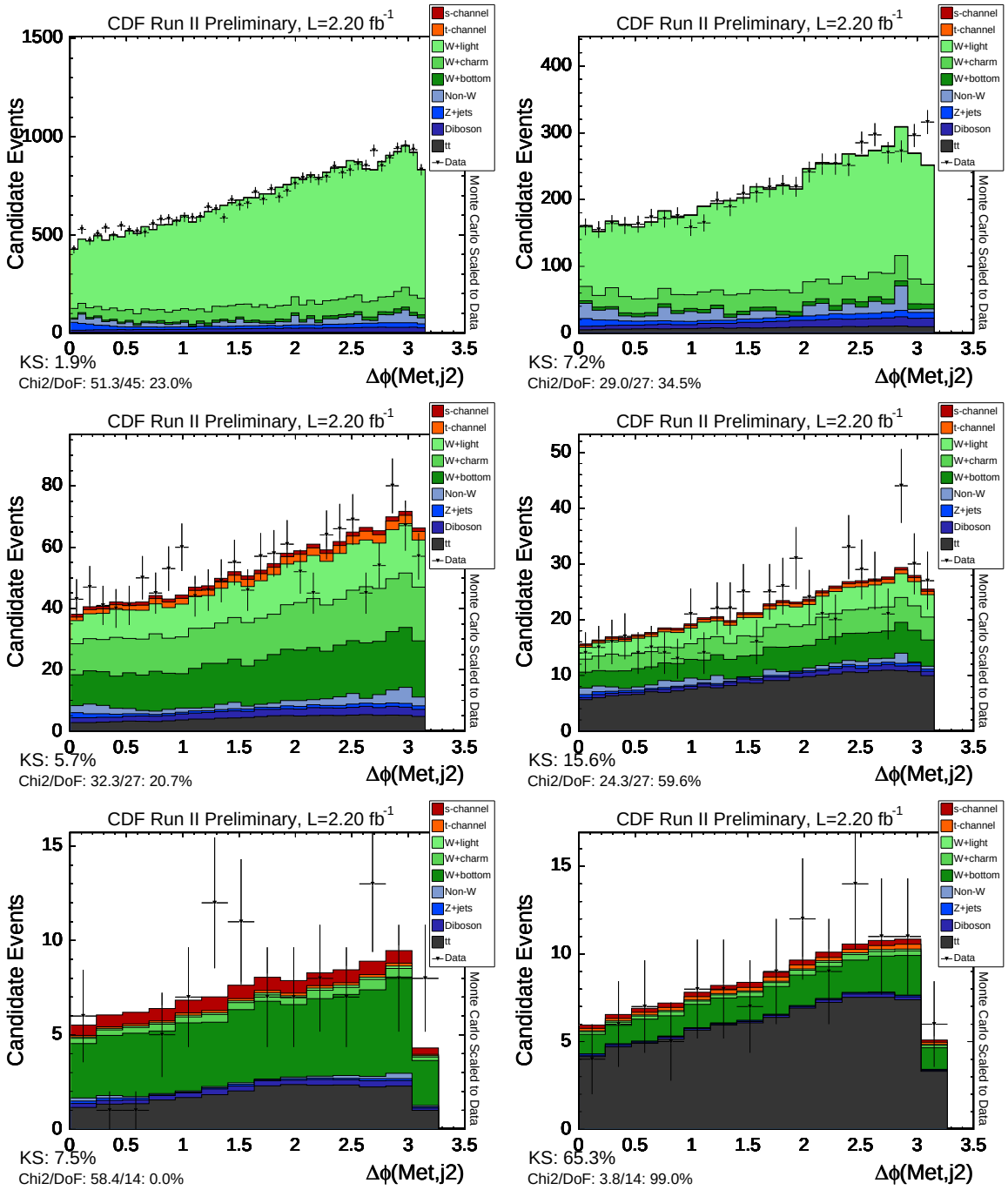

Figure 33: $\Delta\Phi$ between 2nd jet and $\not\!\!E_T$. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

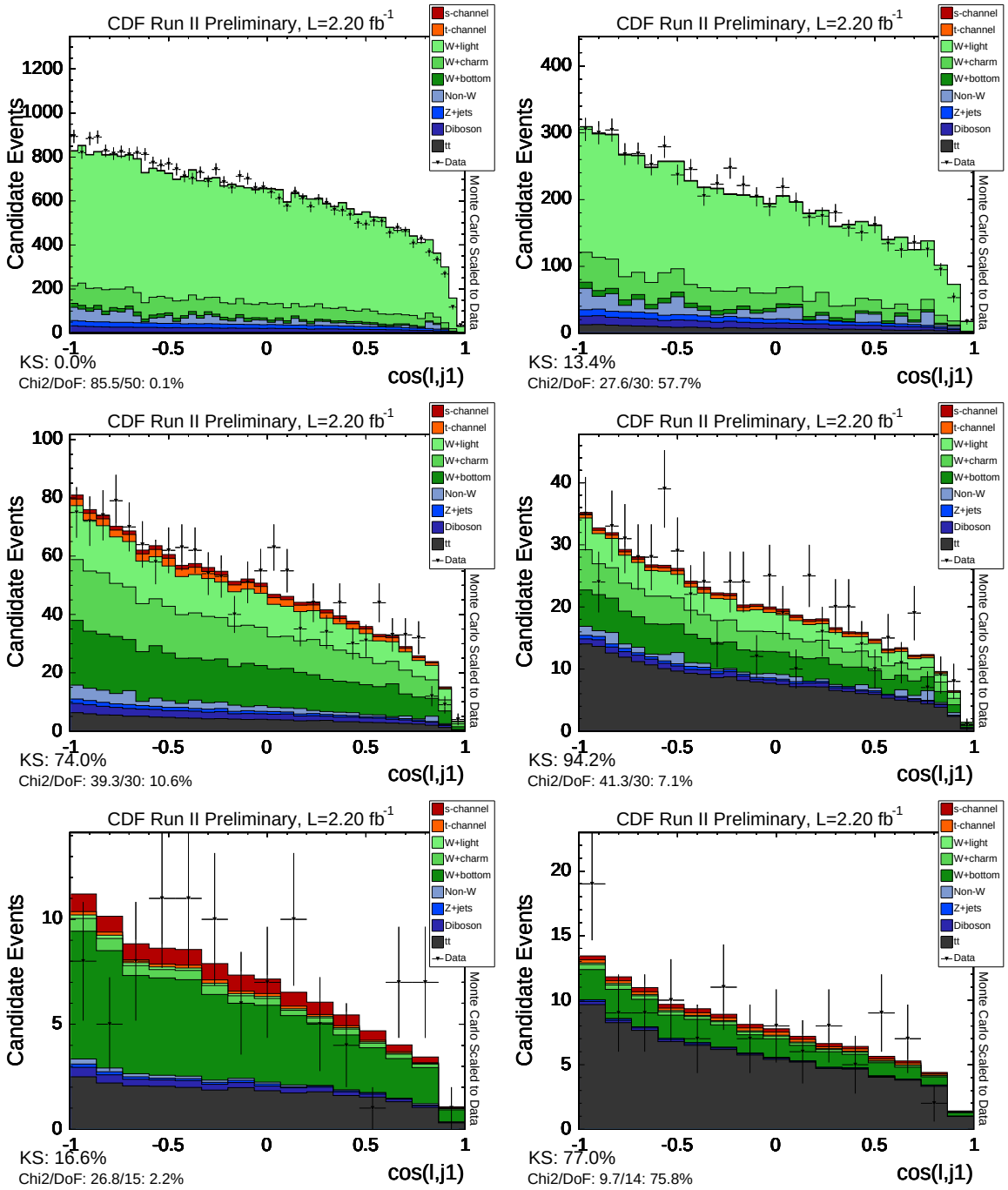

Figure 34: Cosine of the angle between leading jet and lepton. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

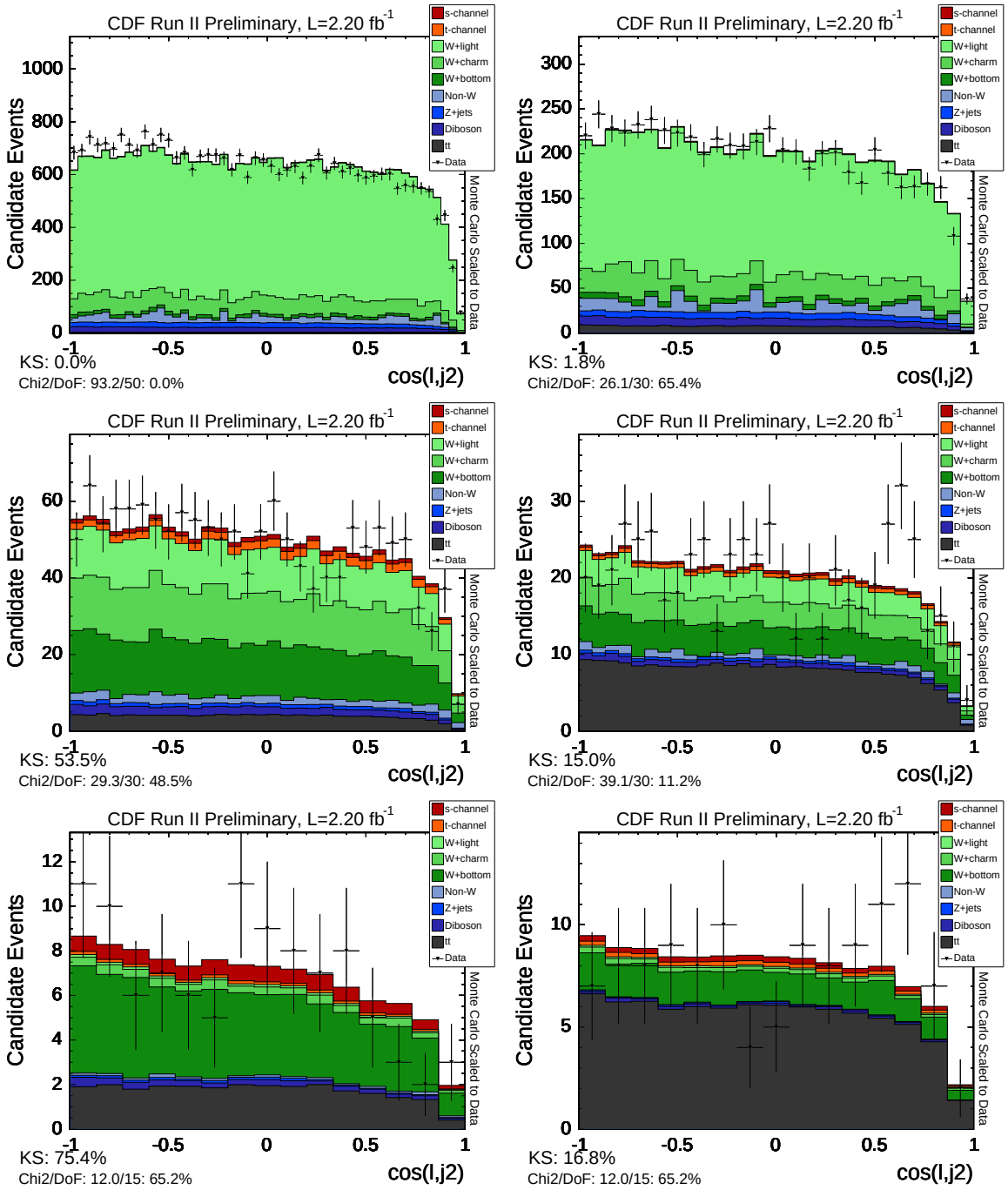

Figure 35: Cosine of the angle between 2nd jet and lepton. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

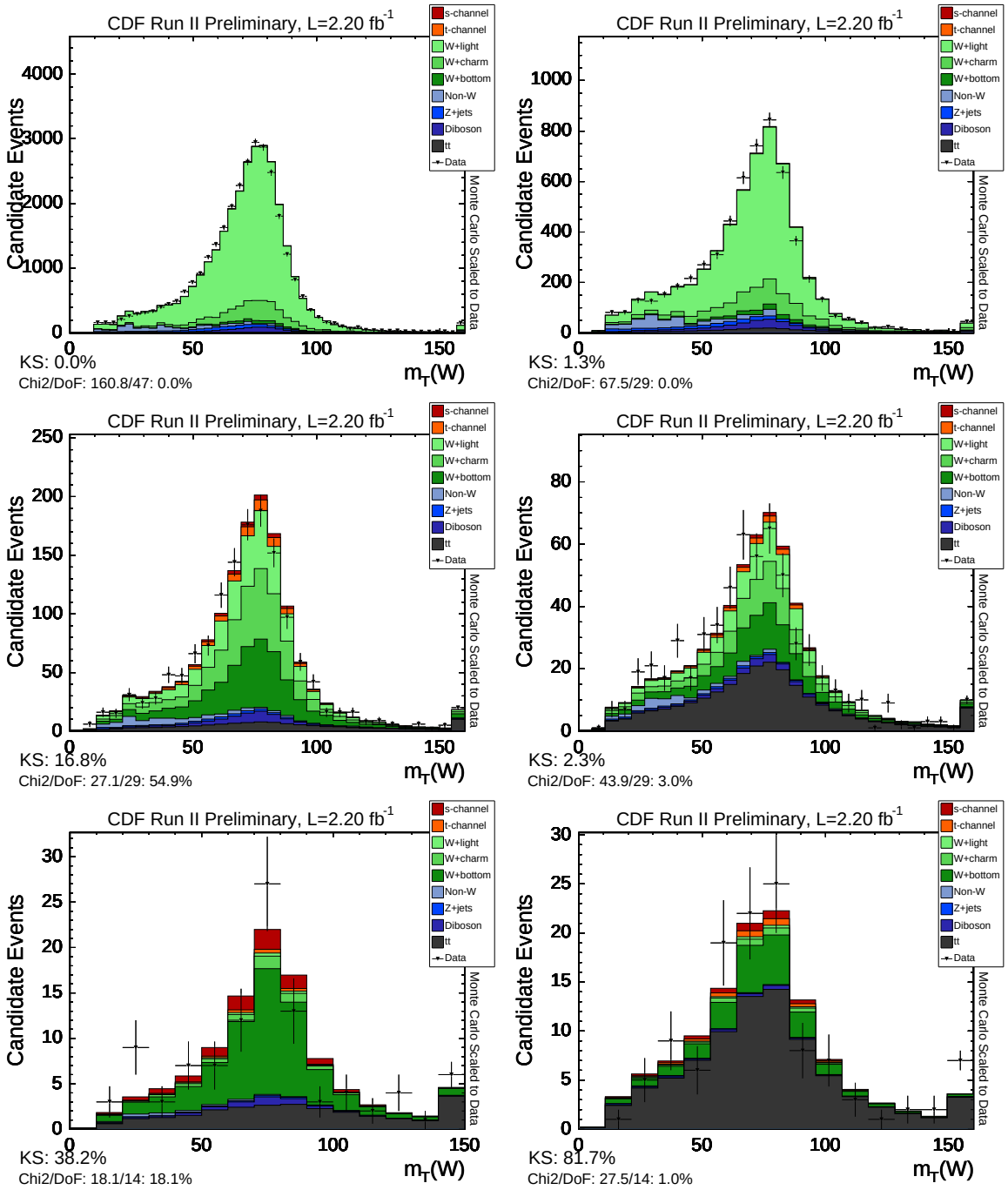

Figure 36: Transverse mass of the W boson. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

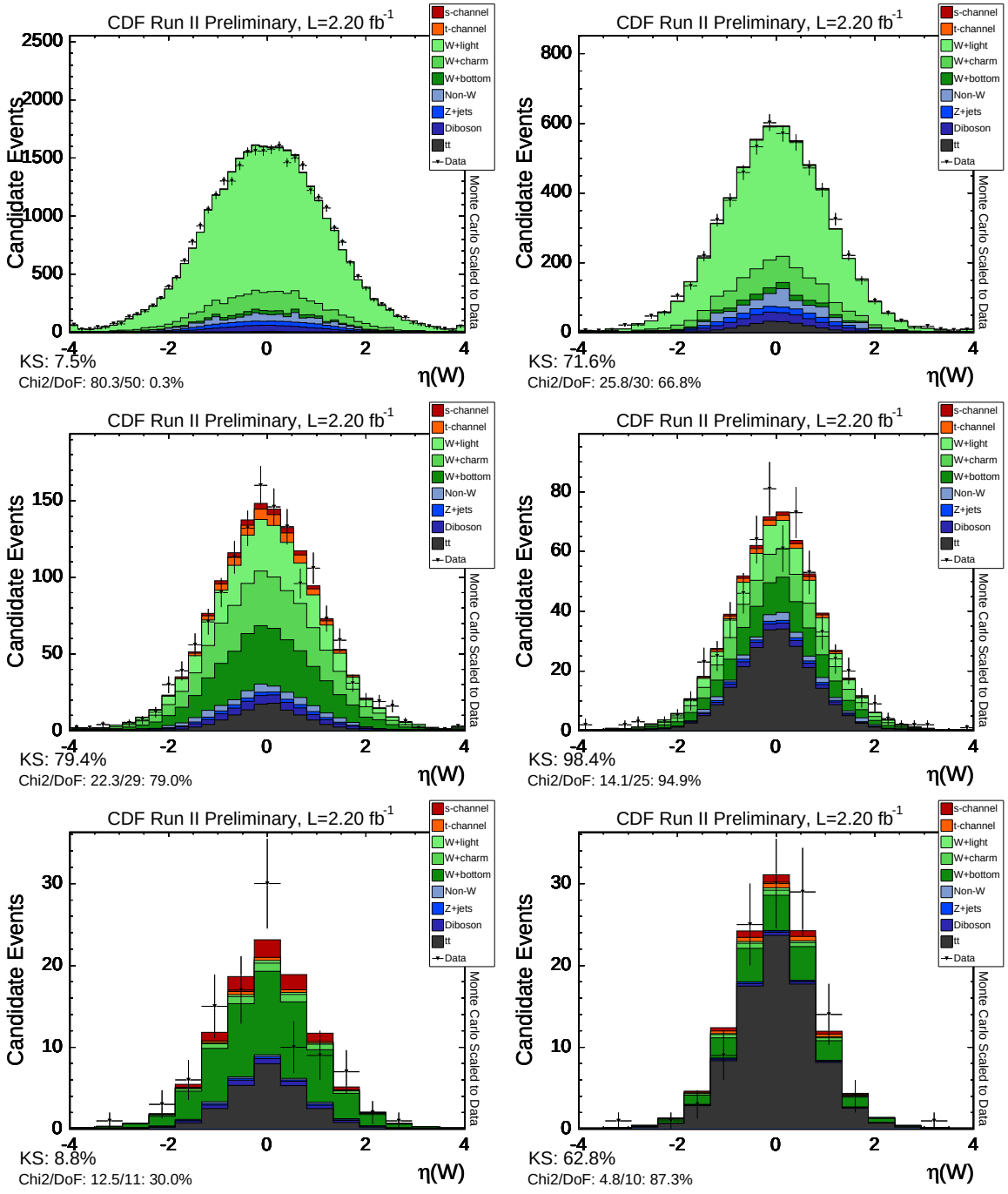

Figure 37: η of the W boson. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

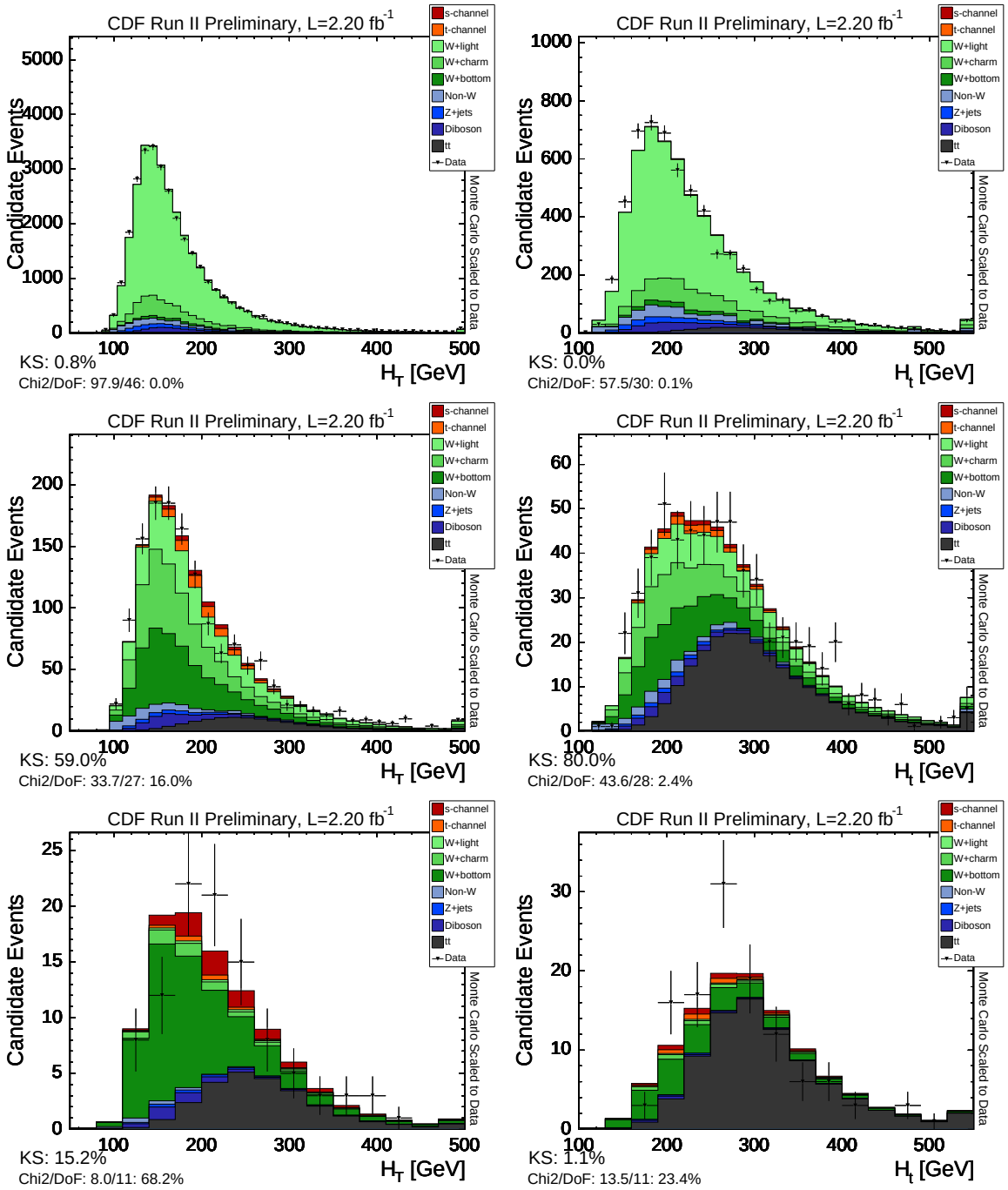

Figure 38: Sum of scalar energies of the event. Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

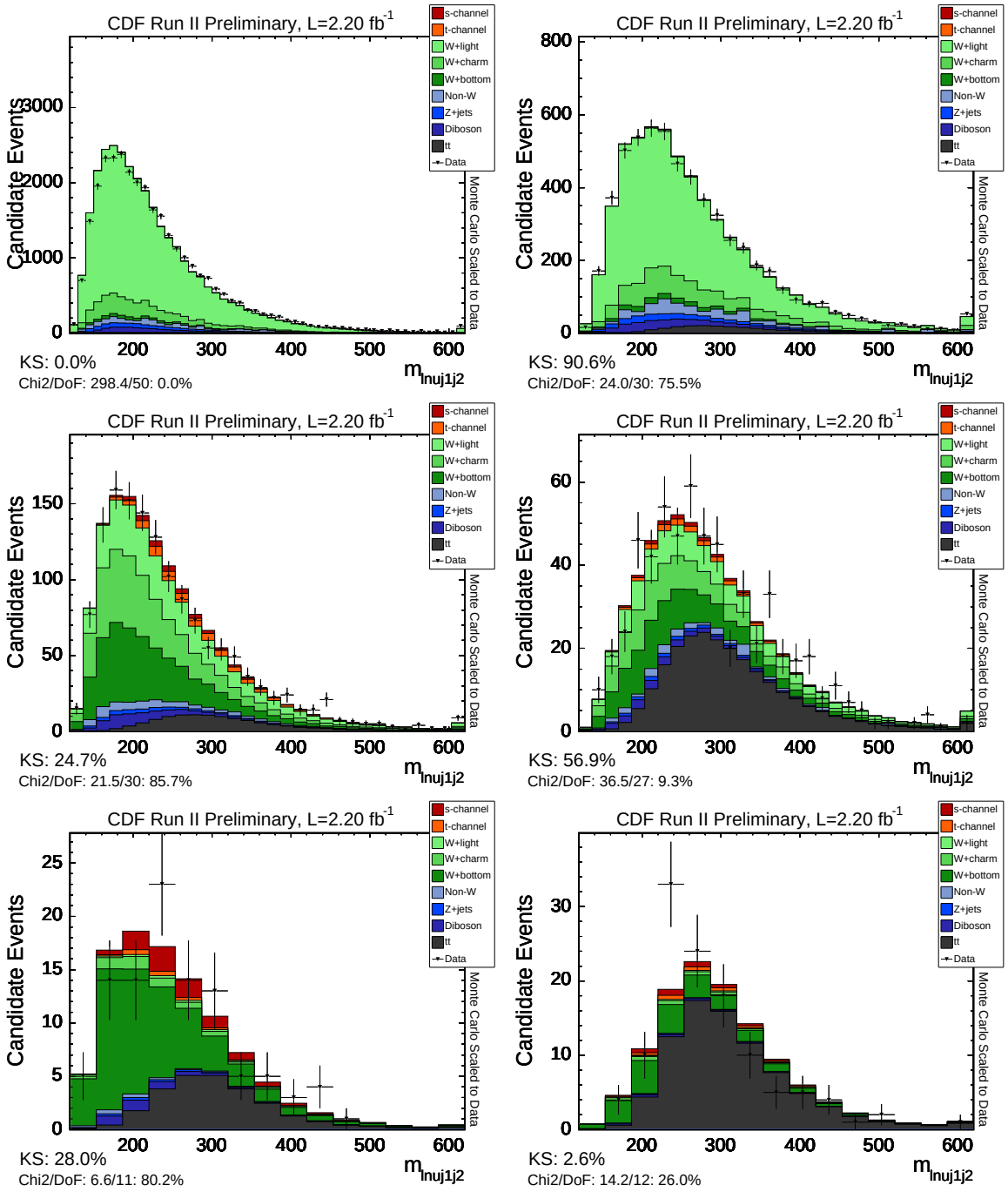

Figure 39: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

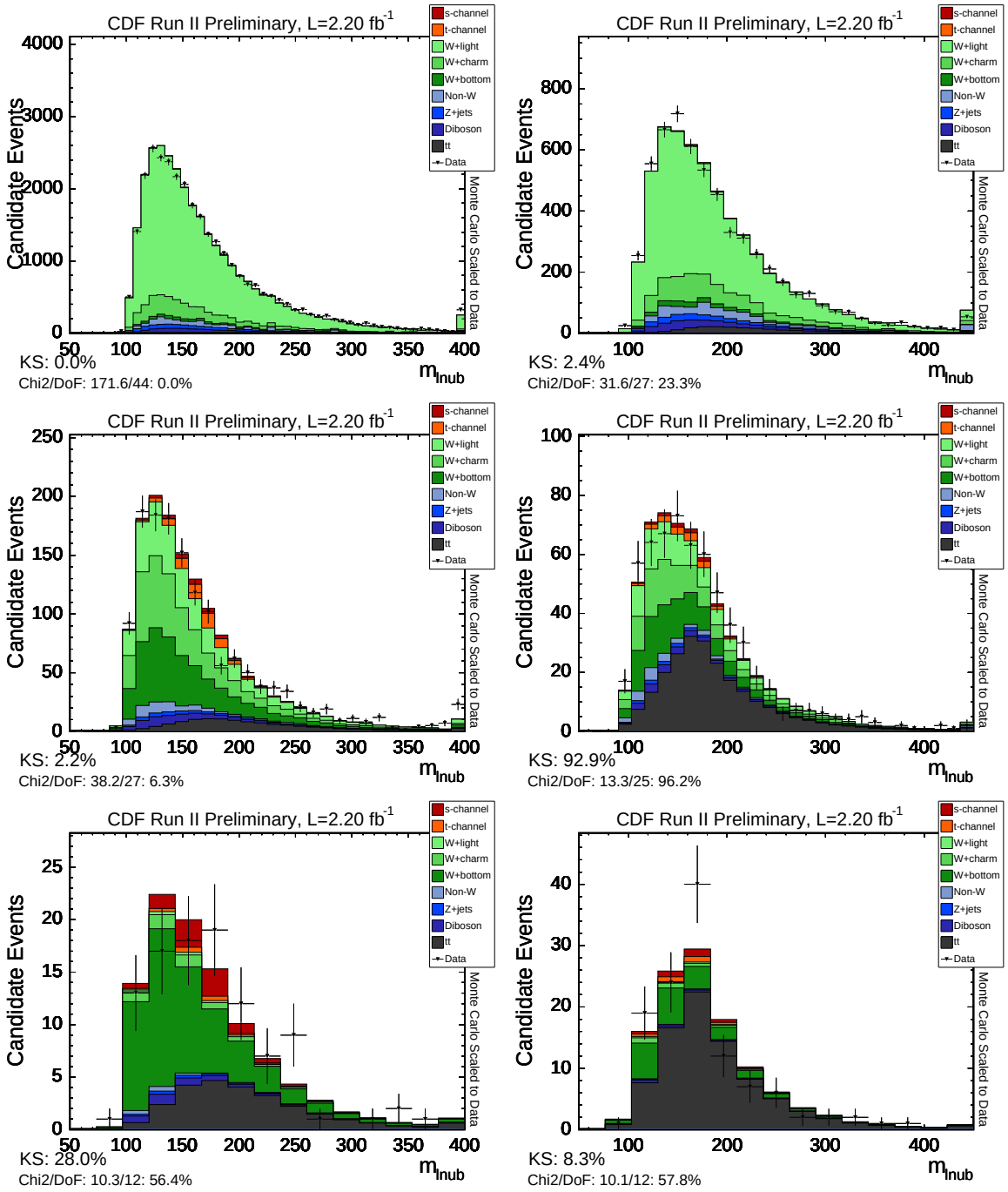

Figure 40: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

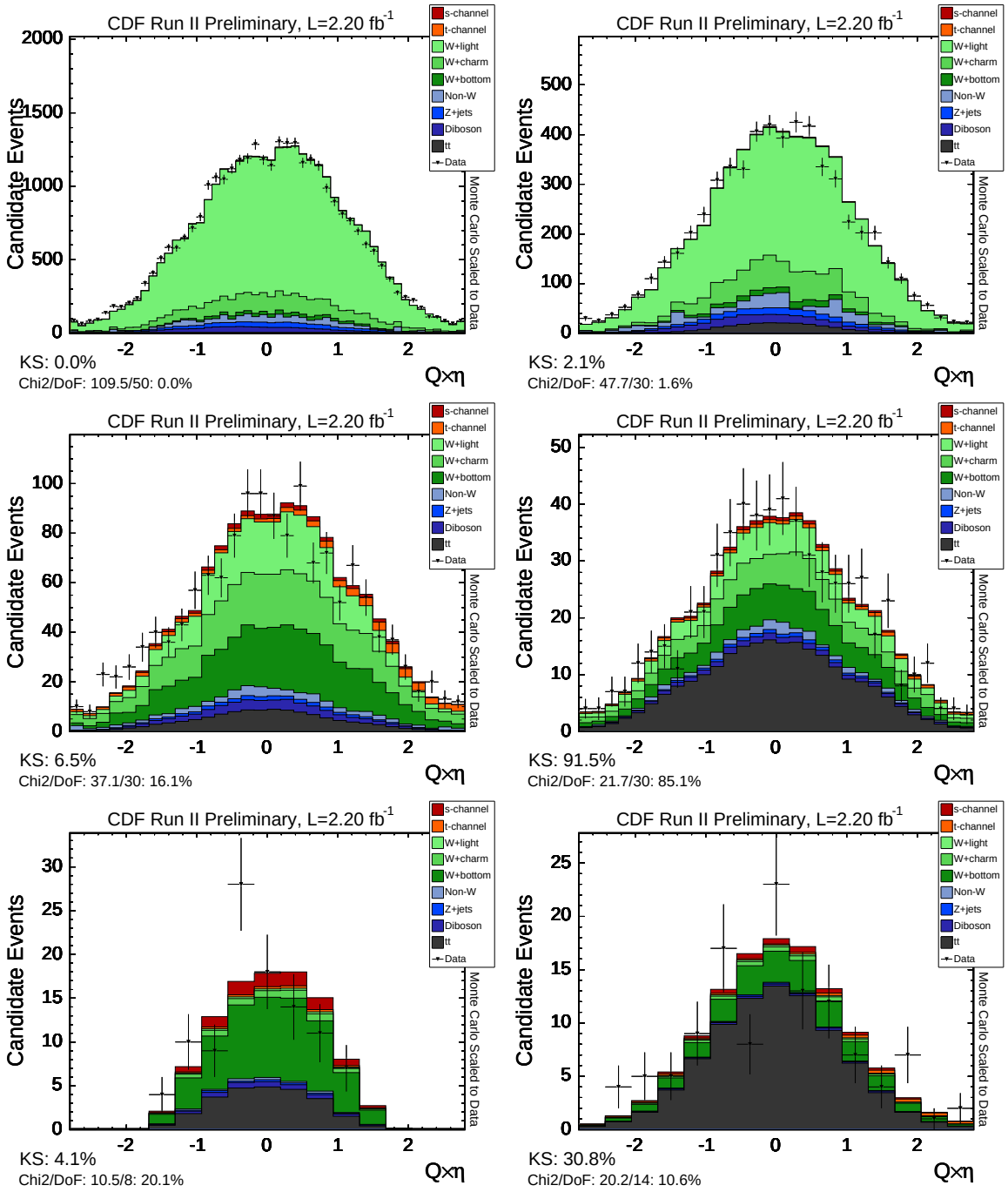

Figure 41: Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

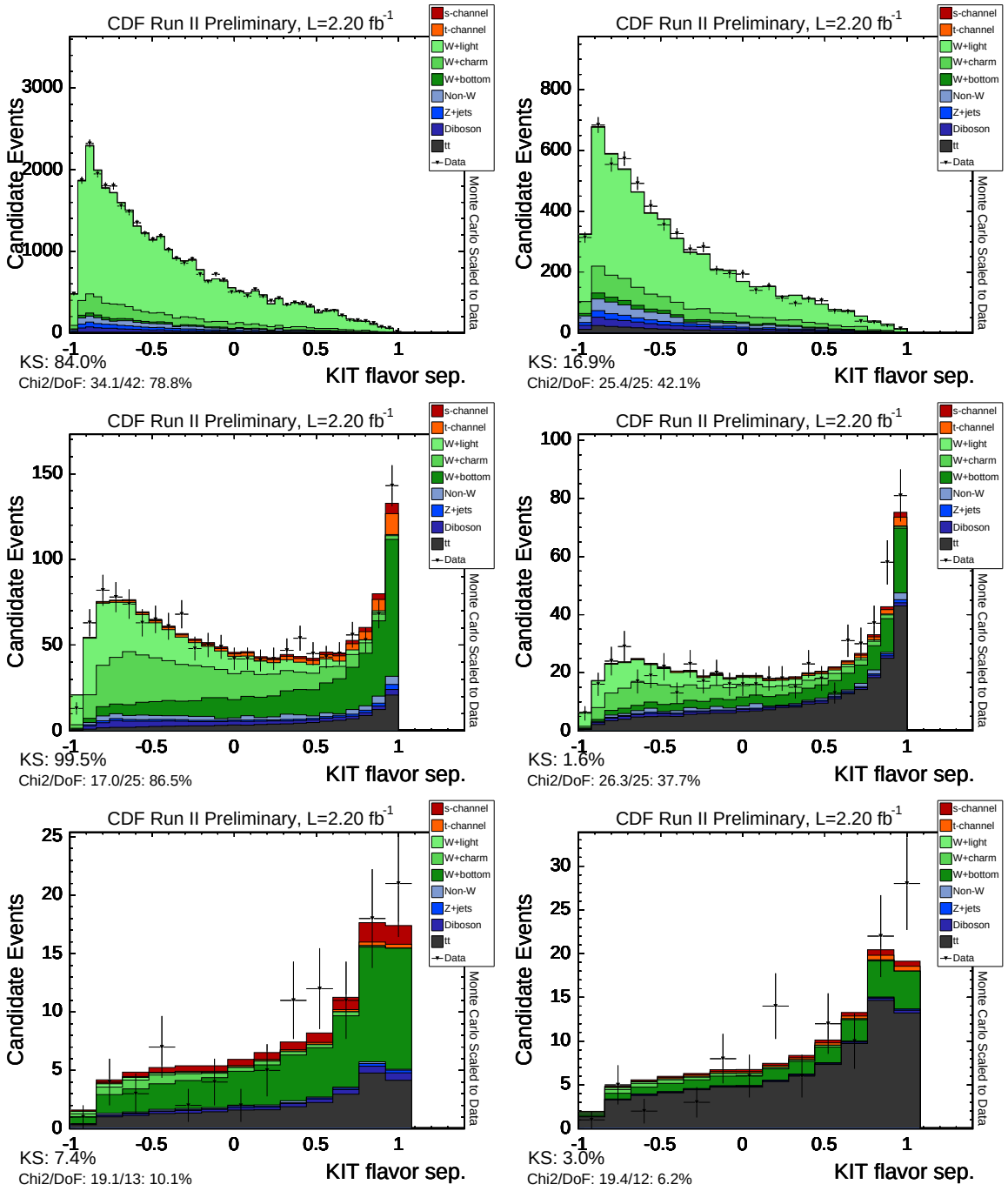

Figure 42: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

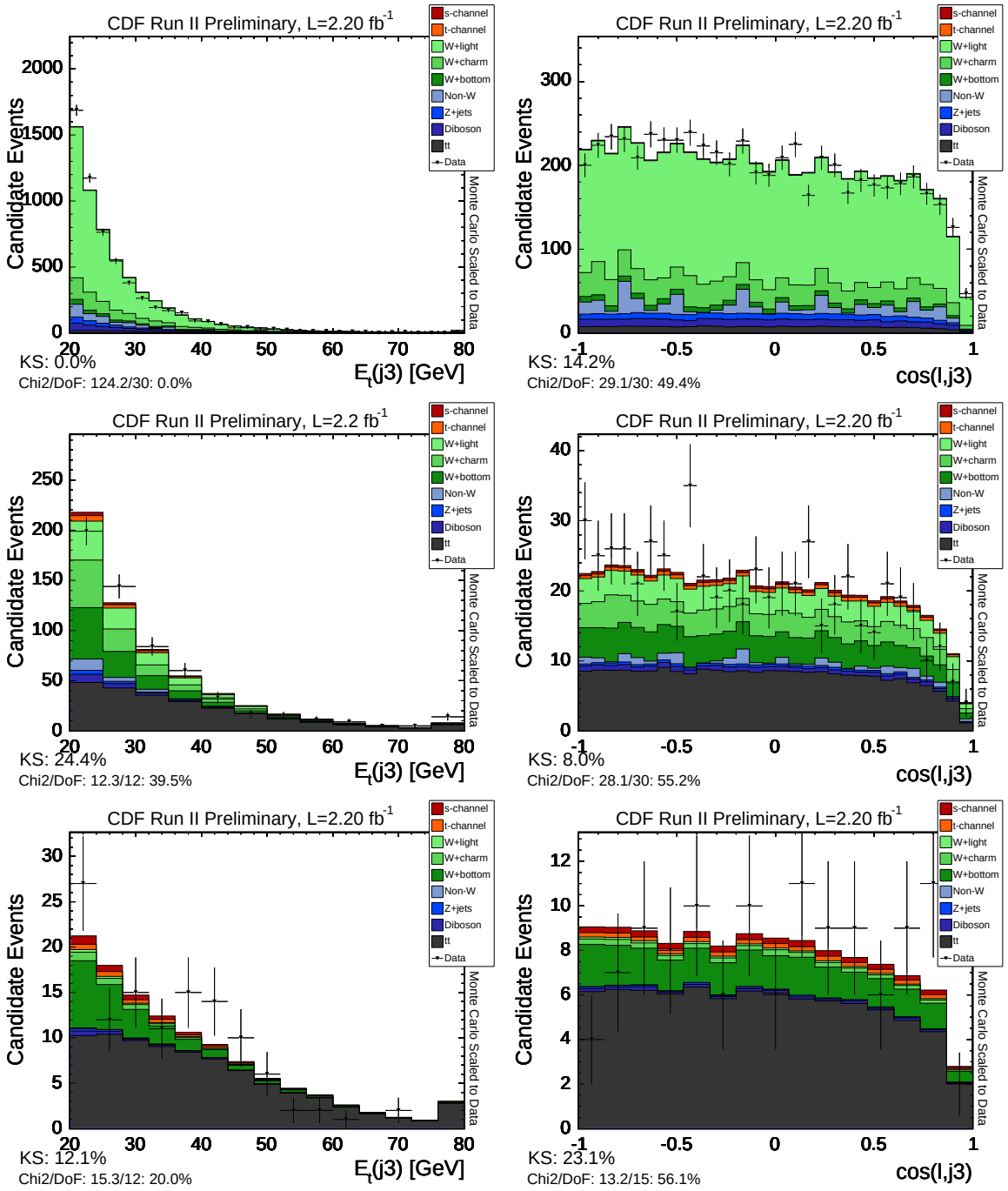

Figure 43: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

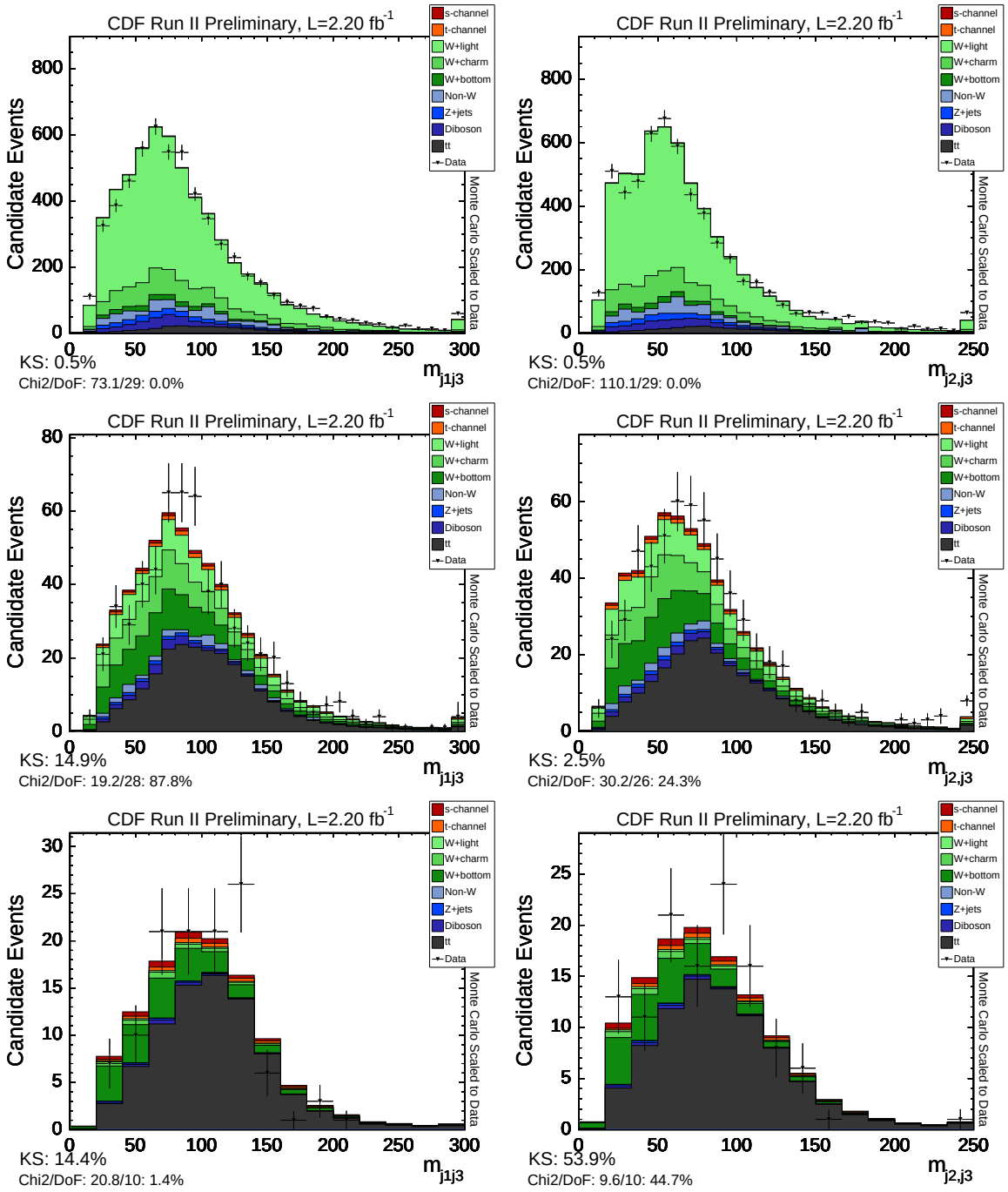

Figure 44: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

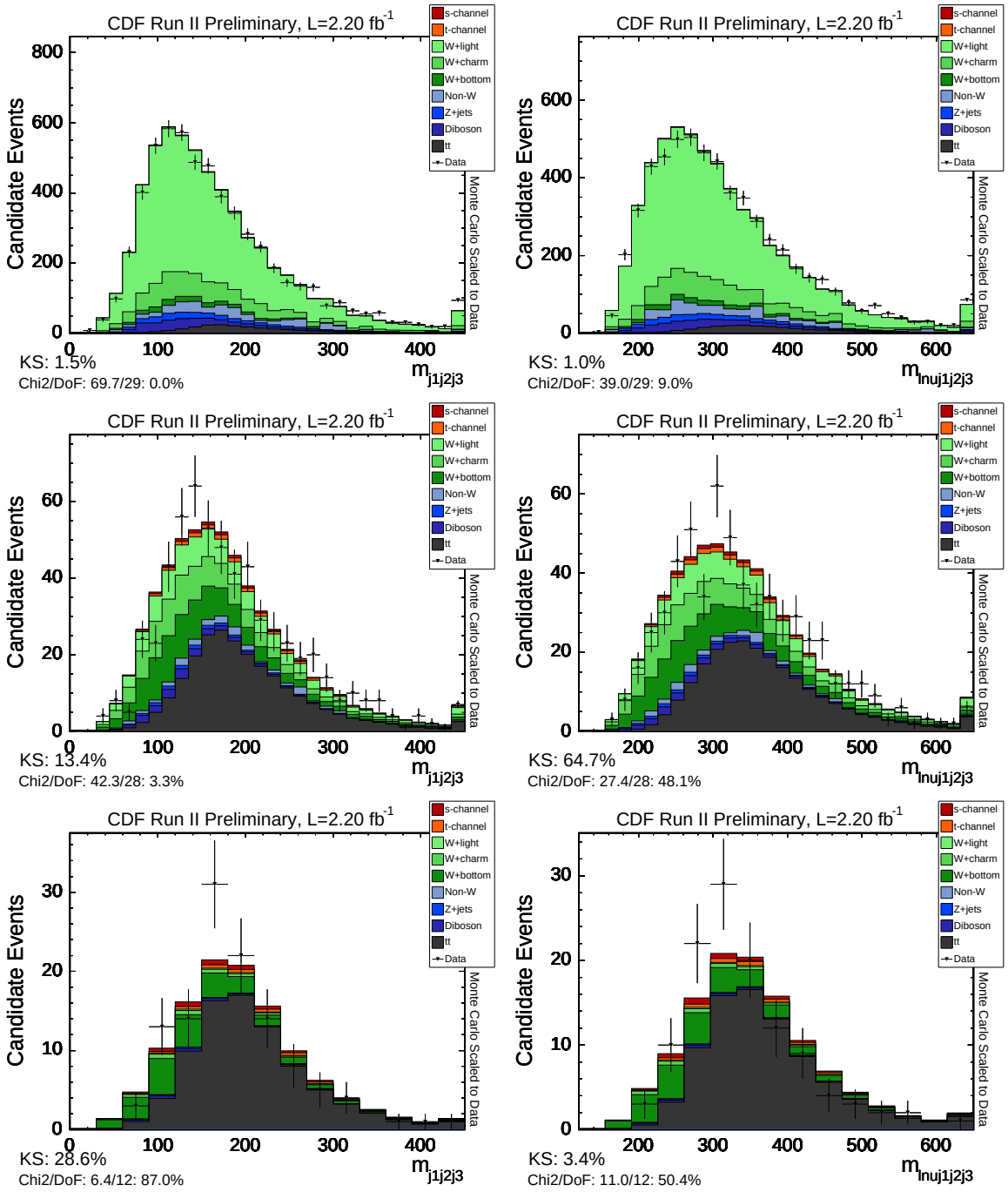

Figure 45: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

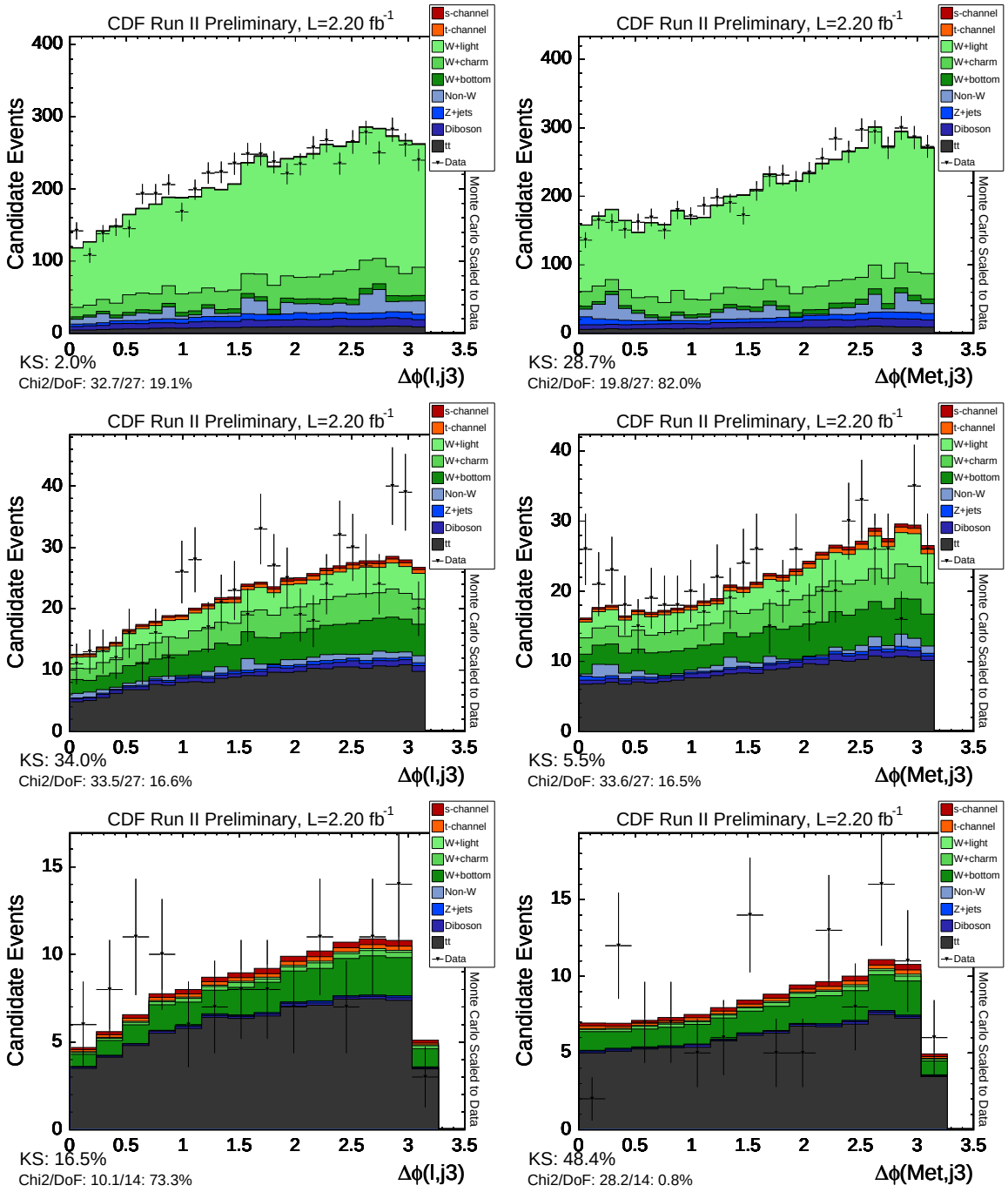

Figure 46: . Top is the untagged sample, middle is the single tagged sample, and bottom is the double tagged sample. Left is 2-jet bin, and right is 3-jet bin.

B Validation of new data sample (p14-p17)

As a quick validation check of the newest data periods to be included in the analysis a comparison of the shapes of some important kinematic variables in the 2-jet bin for new data (p14-p17) have been compared with old data (prior to period 14). Figures 47-62 show the new data (red) compared with the old data scaled to the new data (black). Reasonable agreement is observed in all distributions studied.

For completeness the p14-p17 data sample is also compared with the MC prediction in **??-??**.

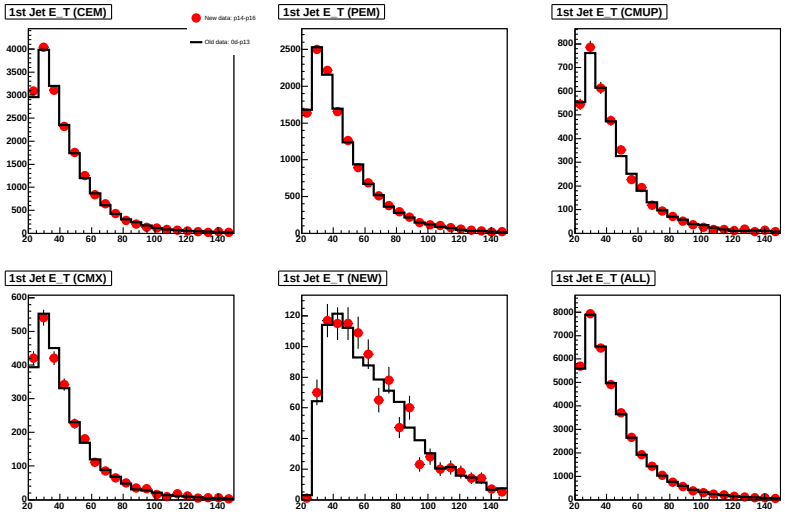

Figure 47: E_T distributions of the first jet in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

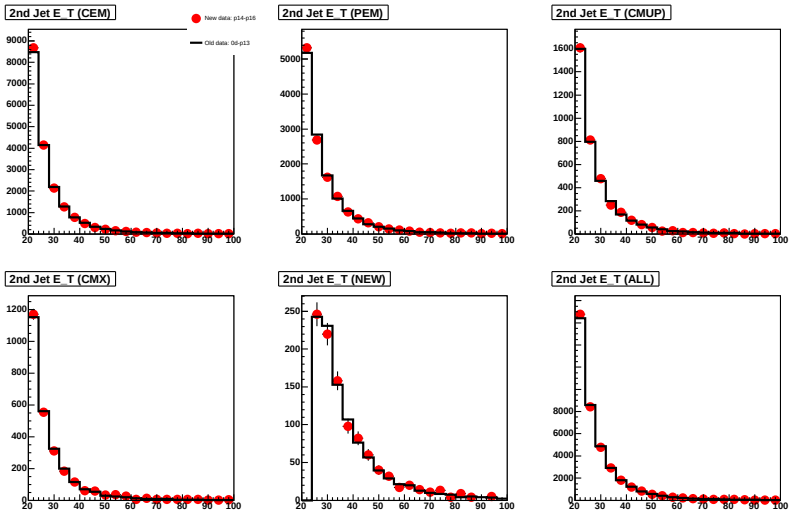

Figure 48: E_T distributions of the second jet in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

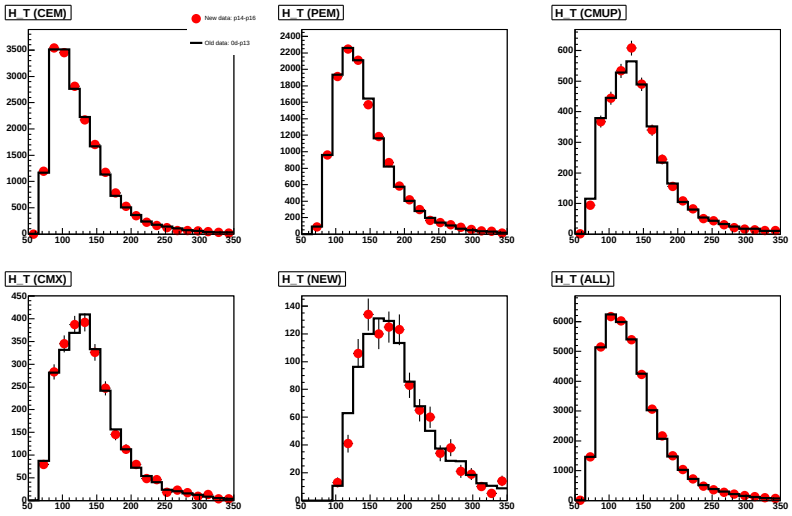

Figure 49: H_T distributions in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

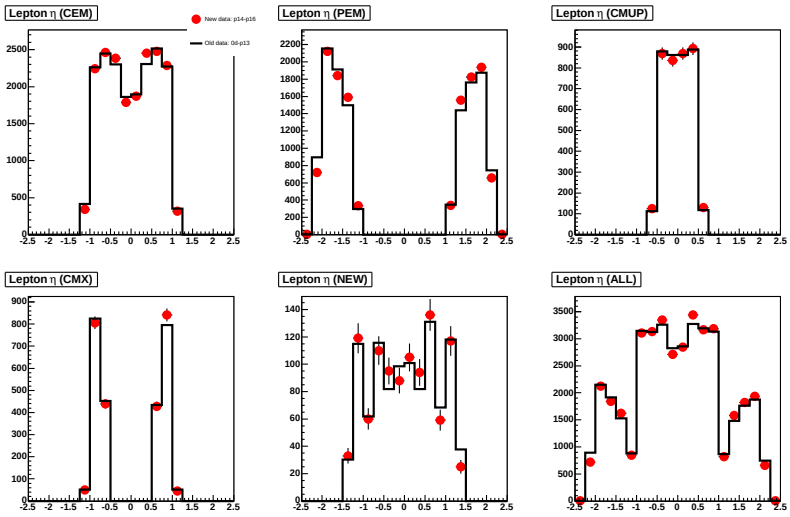

Figure 50: Lepton-η distributions in the >0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

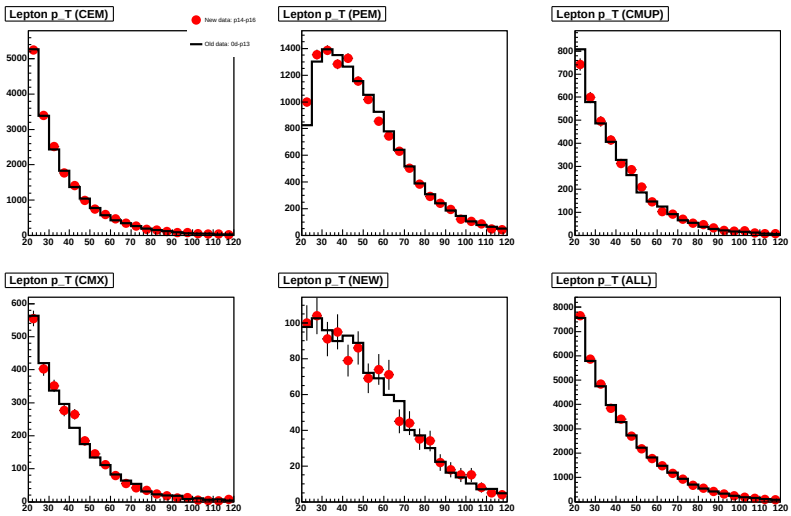

Figure 51: Lepton-P_T distributions in the >0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

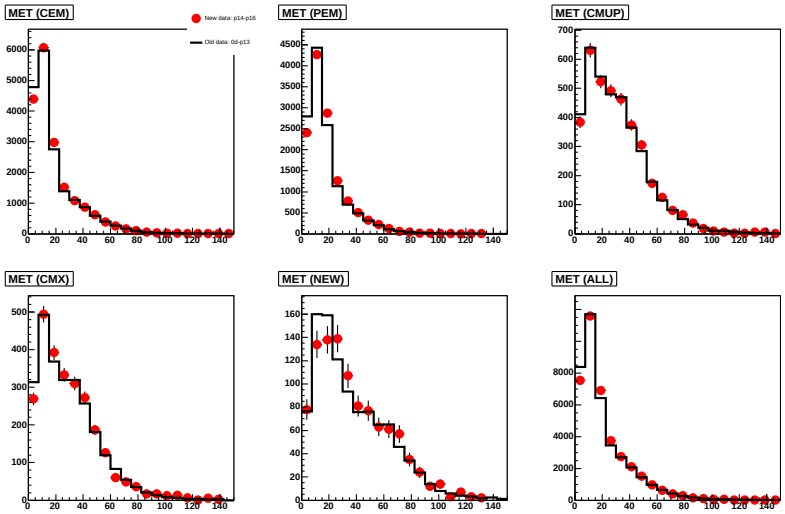

Figure 52: MET distributions in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

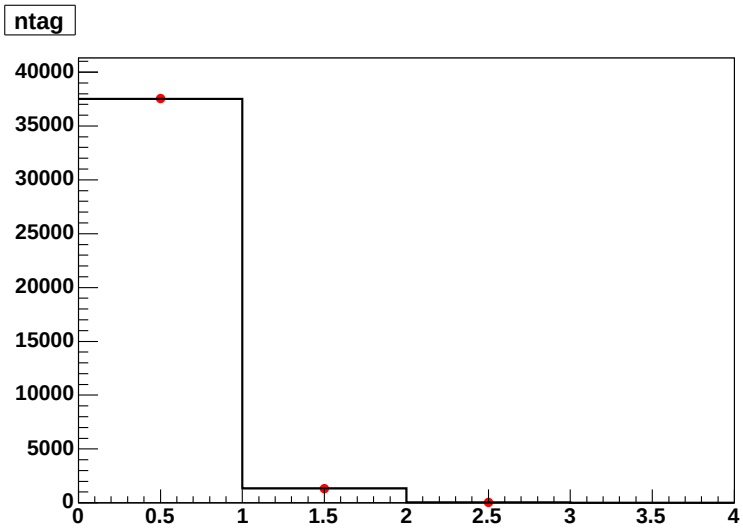

Figure 53: N-tag distribution in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

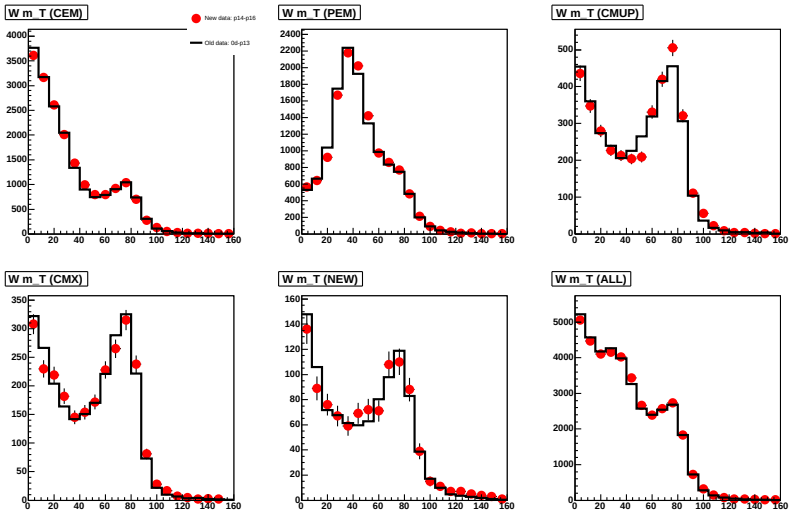

Figure 54: W transverse mass distribution in the > 0 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

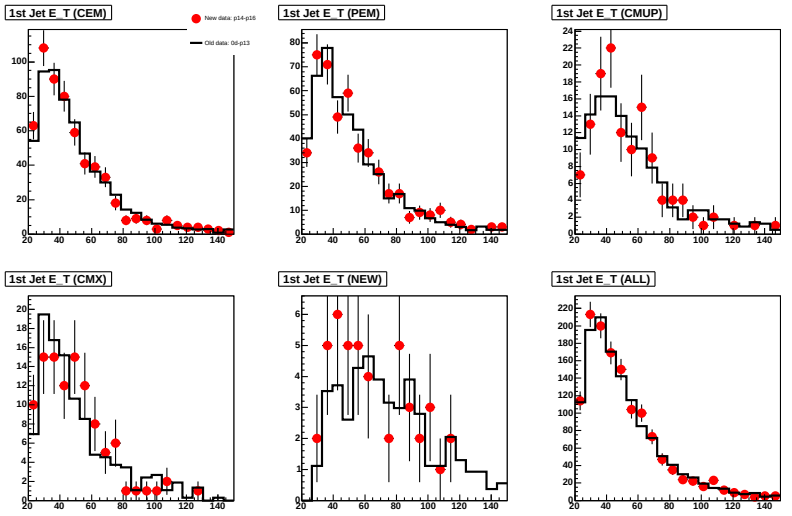

Figure 55: E_T distributions of the first jet in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

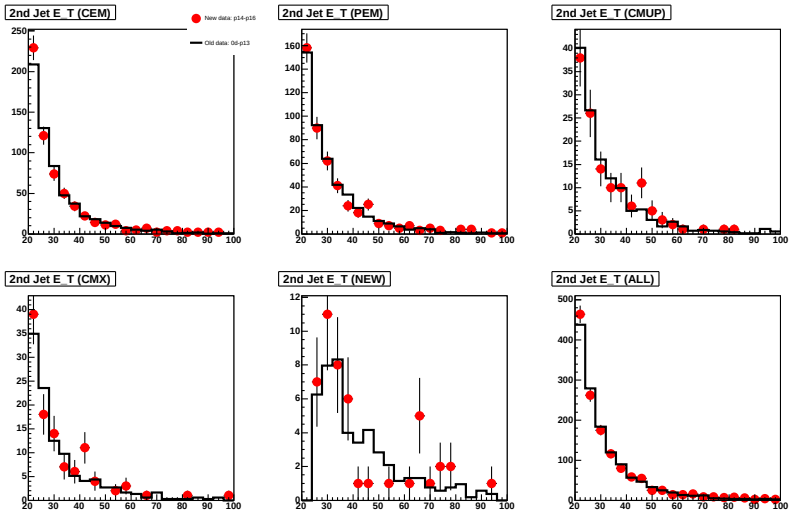

Figure 56: E_T distributions of the second jet in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

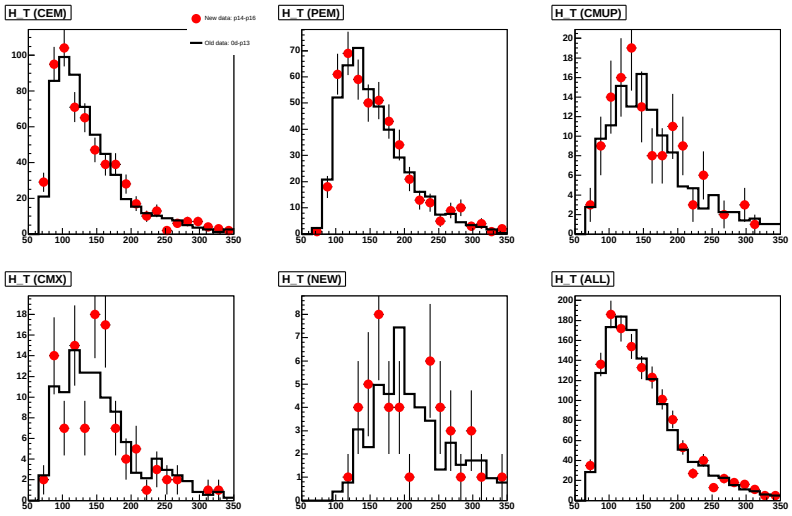

Figure 57: H_T distributions in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

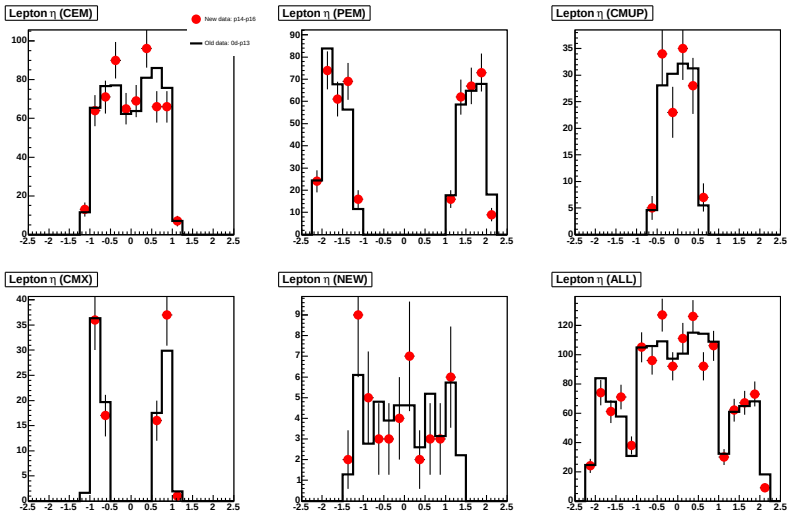

Figure 58: Lepton-η distributions in the $>$ 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

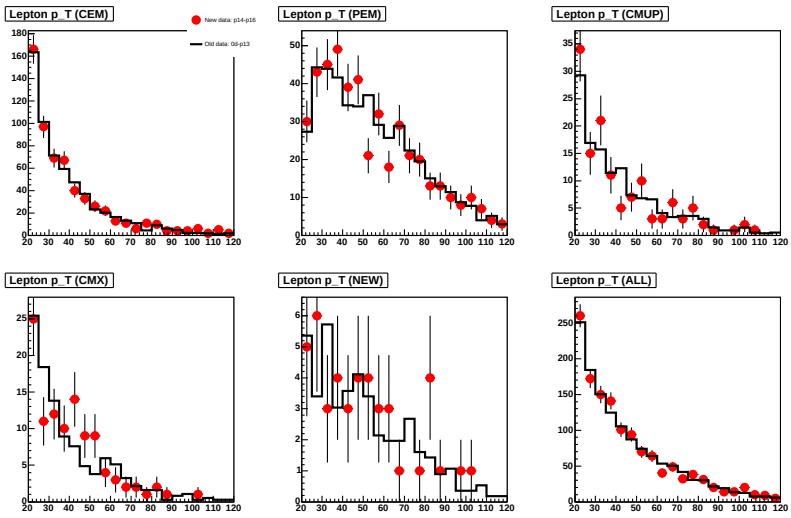

Figure 59: Lepton-P_T distributions in the $>$ 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

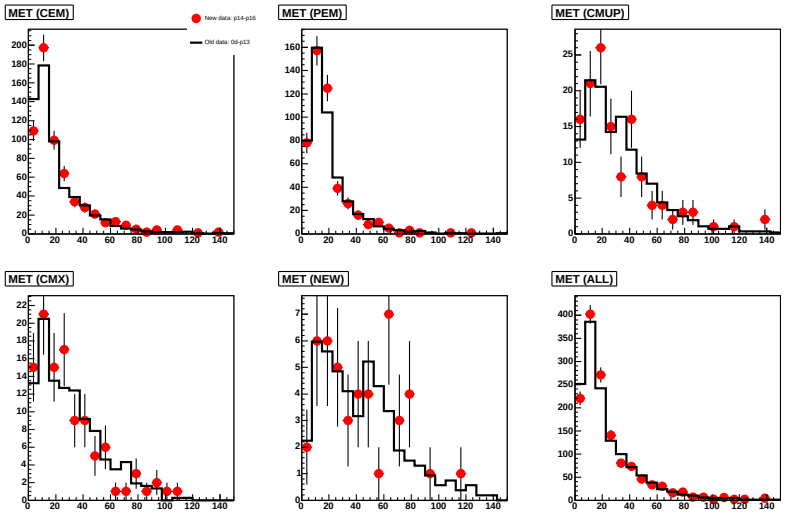

Figure 60: MET distributions in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

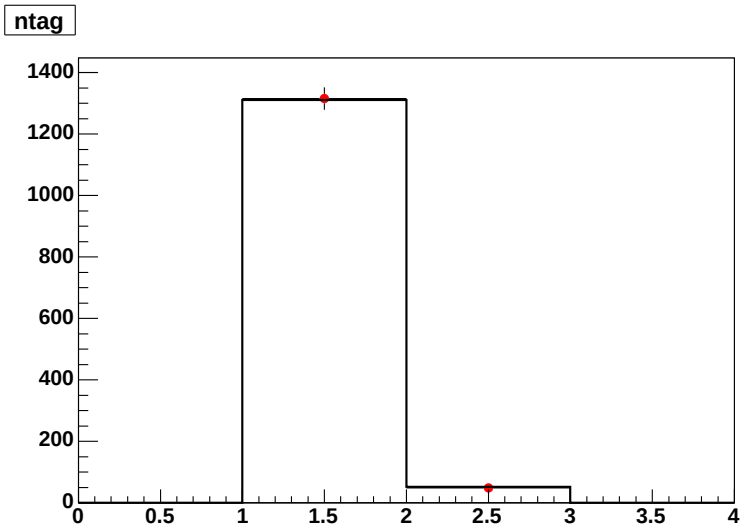

Figure 61: N-tag distribution in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

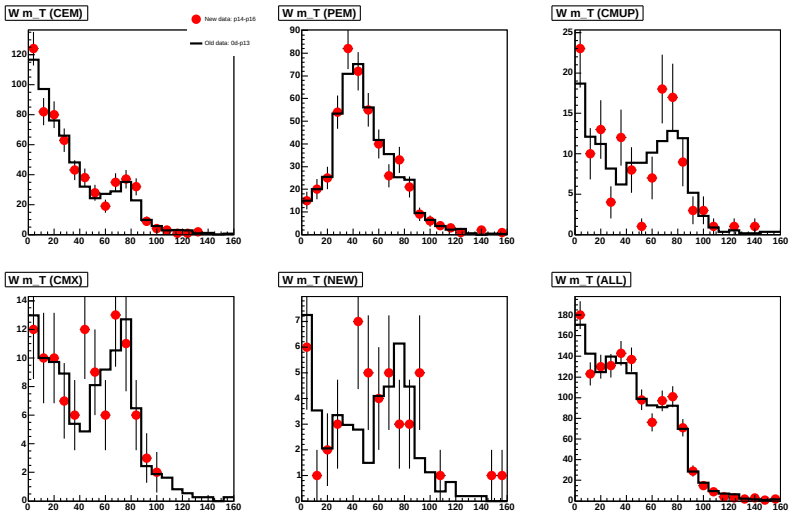

Figure 62: W transverse mass distribution in the > 1 tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

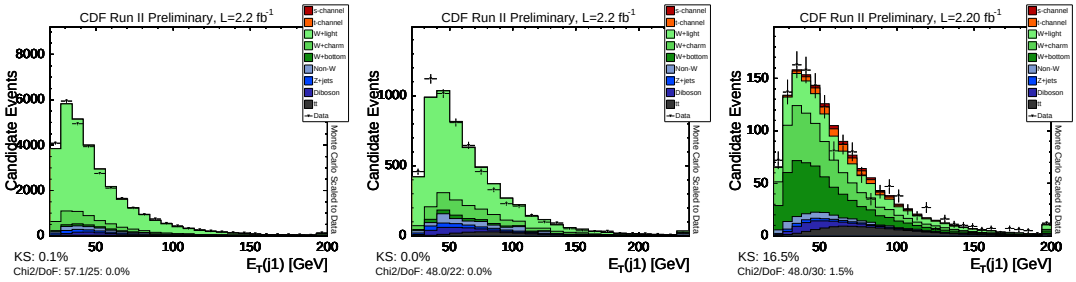

Figure 63: E_T of the leading jet. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

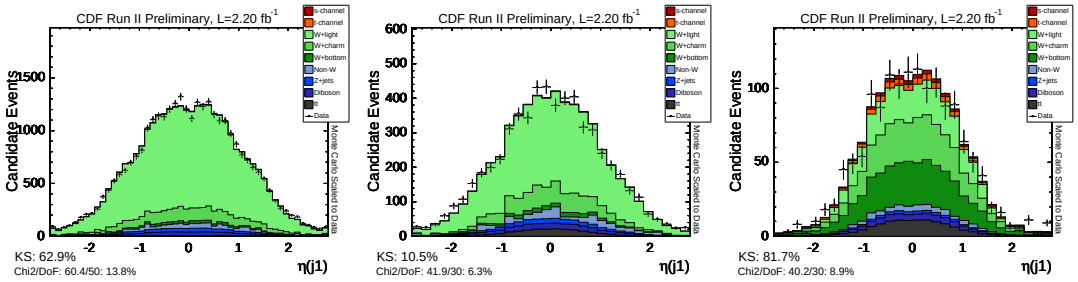

Figure 64: η of the leading jet. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

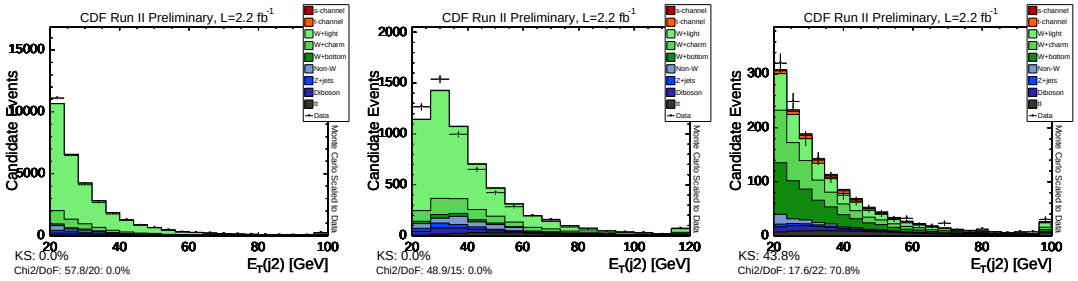

Figure 65: E_T of the 2nd jet. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

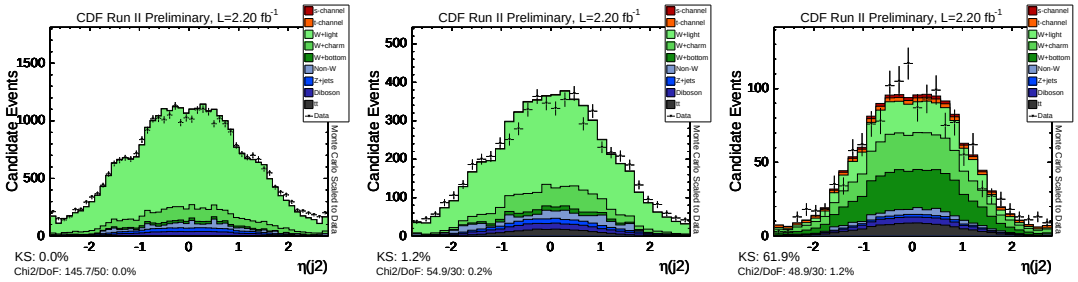

Figure 66: η of the 2nd jet. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

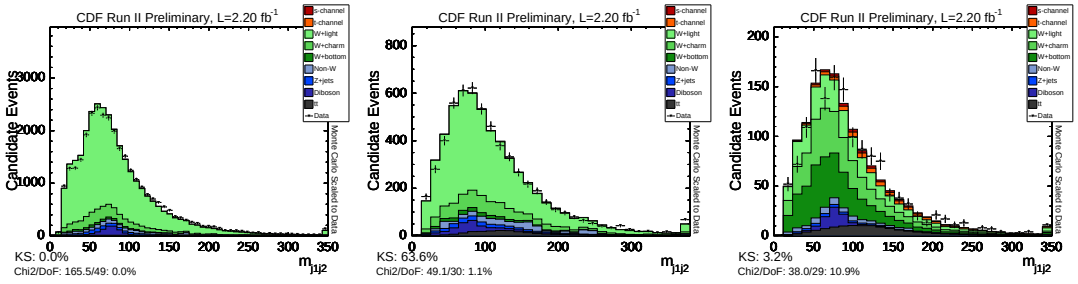

Figure 67: Invariant dijet mass. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

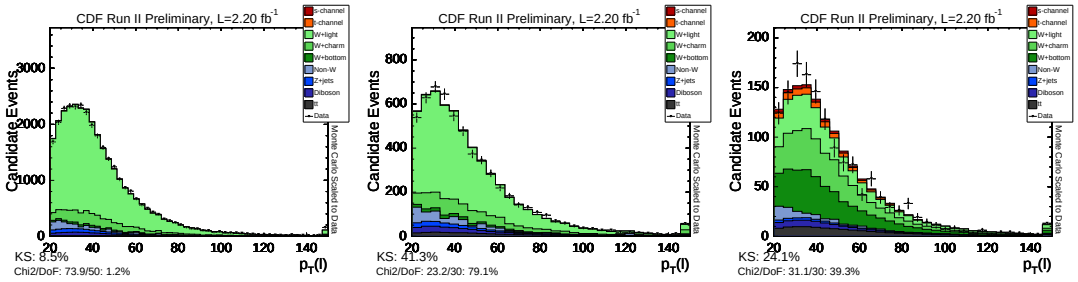

Figure 68: Lepton p_T. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

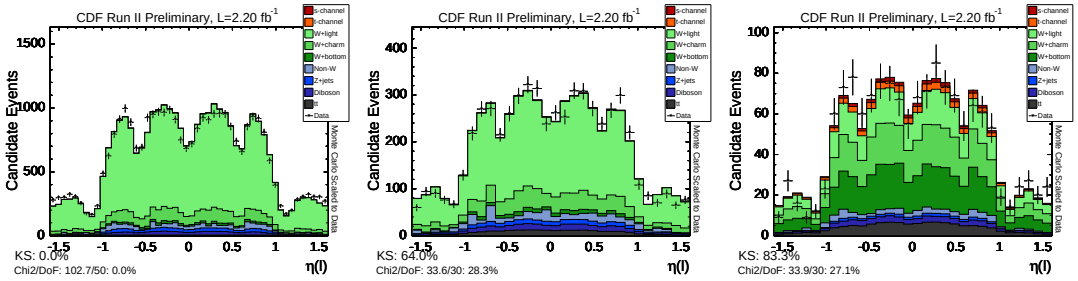

Figure 69: Lepton η. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

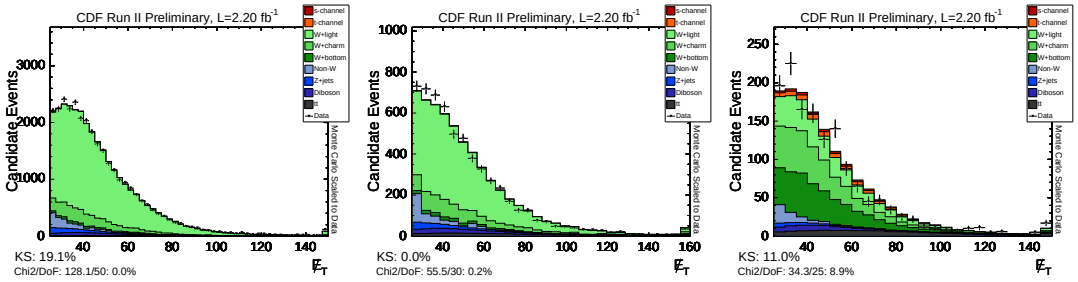

Figure 70: Missing Transverse Energy. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

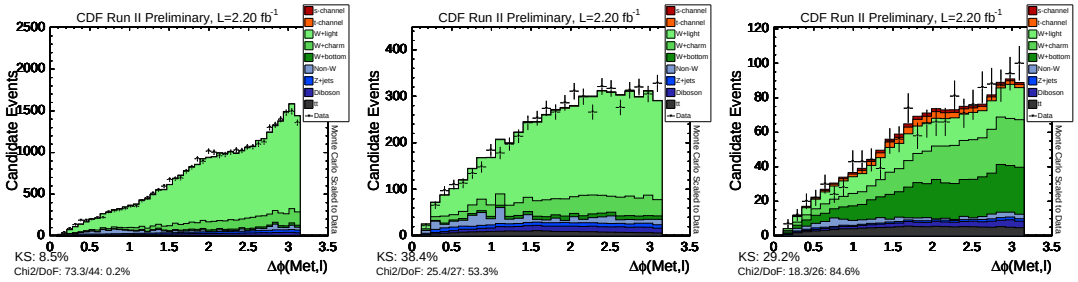

Figure 71: $\Delta\Phi$ between $\not\!\!E_T$ and lepton. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

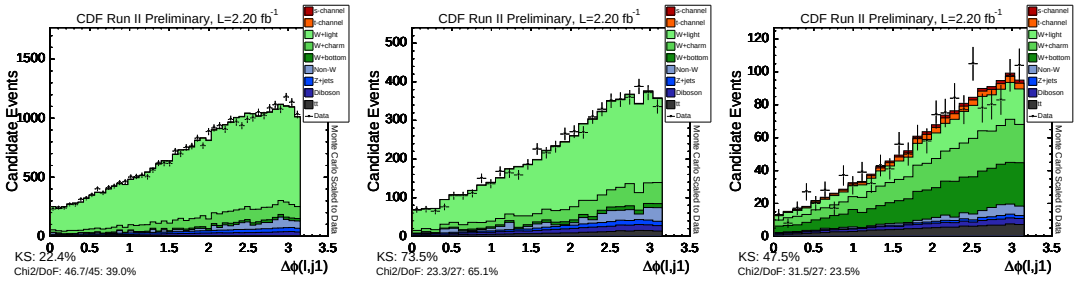

Figure 72: $\Delta\Phi$ between leading jet and lepton. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

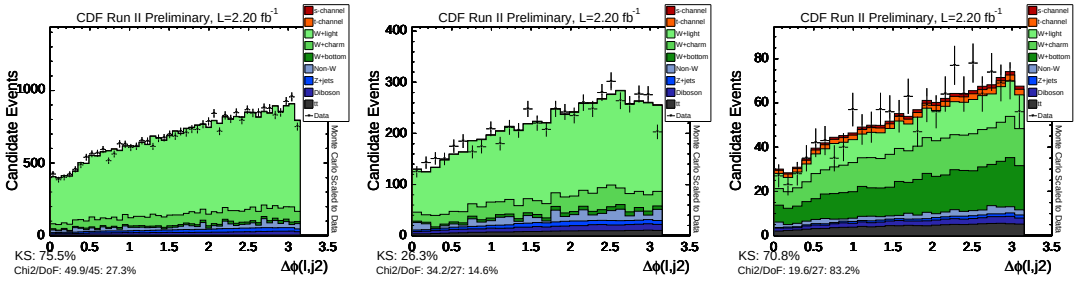

Figure 73: $\Delta\Phi$ between 2nd jet and lepton. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

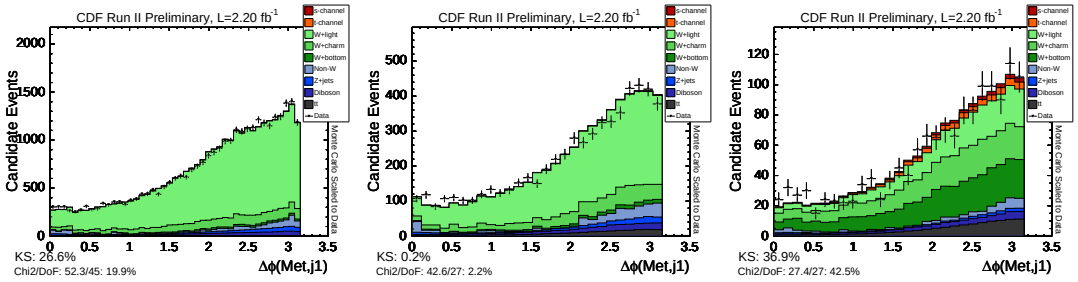

Figure 74: $\Delta\Phi$ between leading jet and $\not\!\!E_T$. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

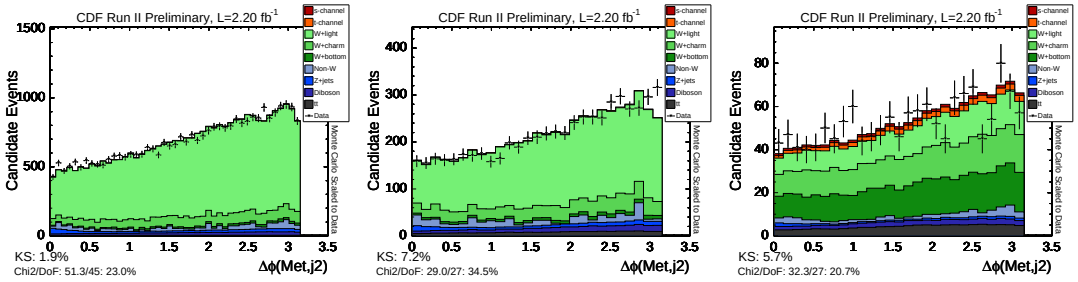

Figure 75: $\Delta\Phi$ between 2nd jet and $\not\!\!E_T$. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

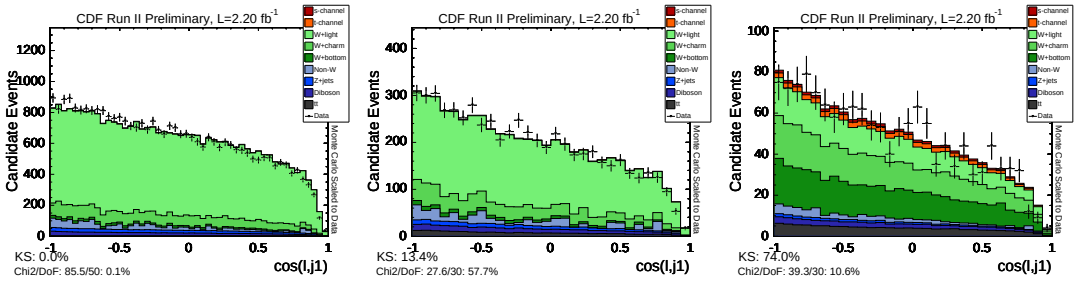

Figure 76: Cosine of the angle between leading jet and lepton. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

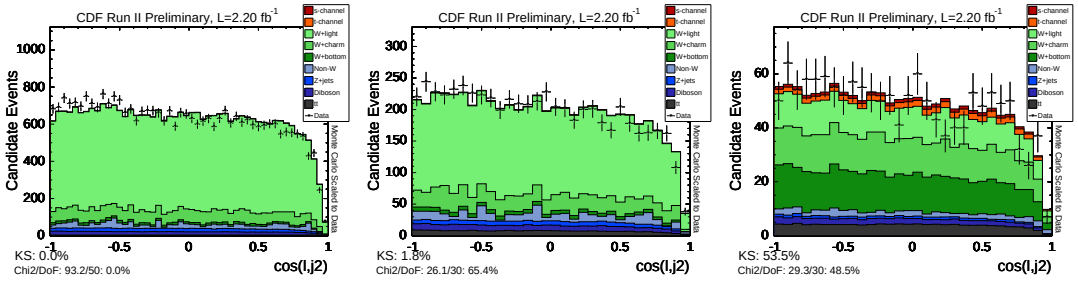

Figure 77: Cosine of the angle between 2nd jet and lepton. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

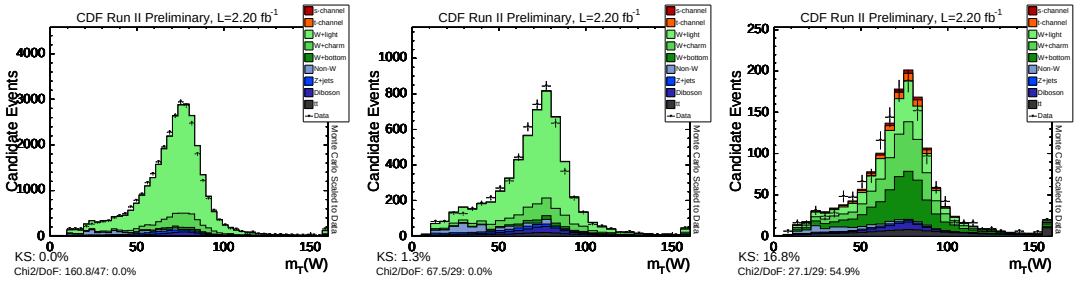

Figure 78: Transverse mass of the W boson. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

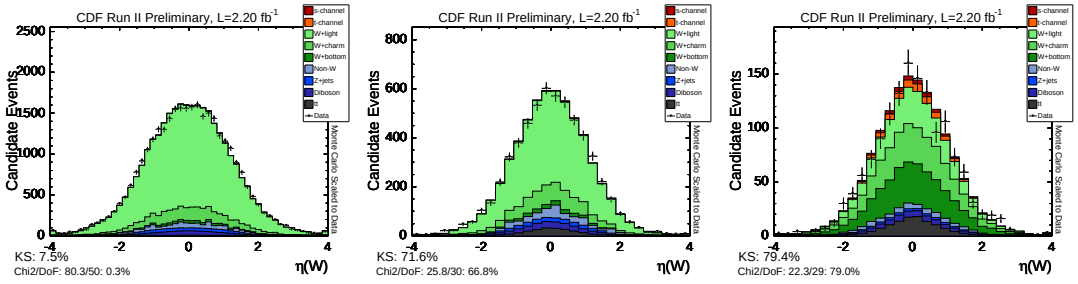

Figure 79: η of the W boson. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

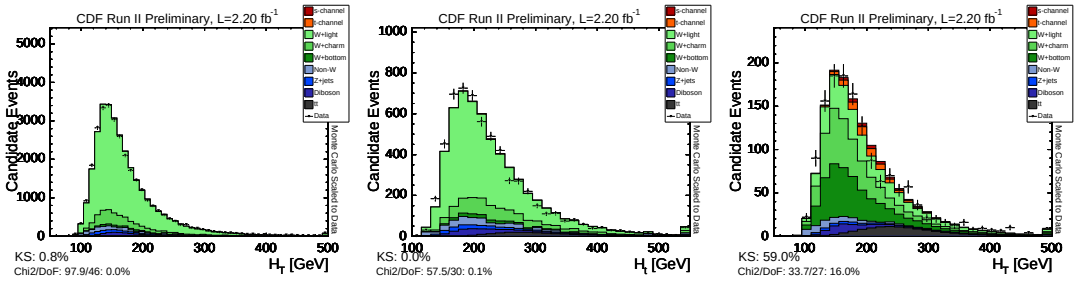

Figure 80: Sum of scalar energies of the event. Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

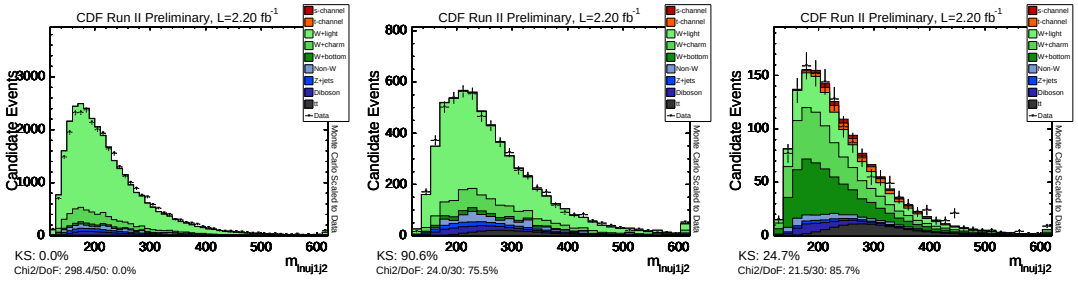

Figure 81: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

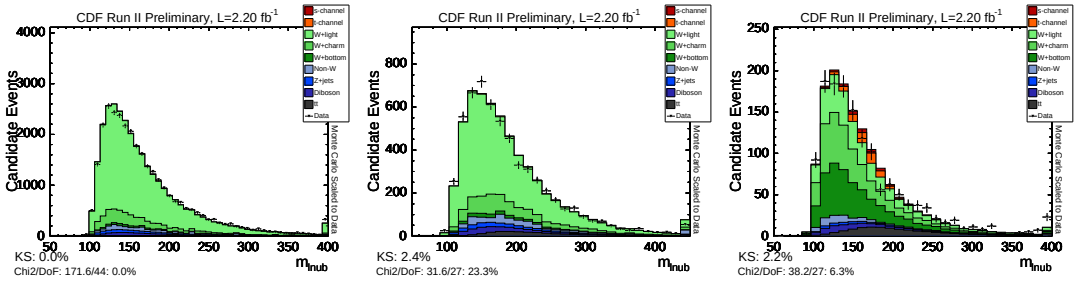

Figure 82: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

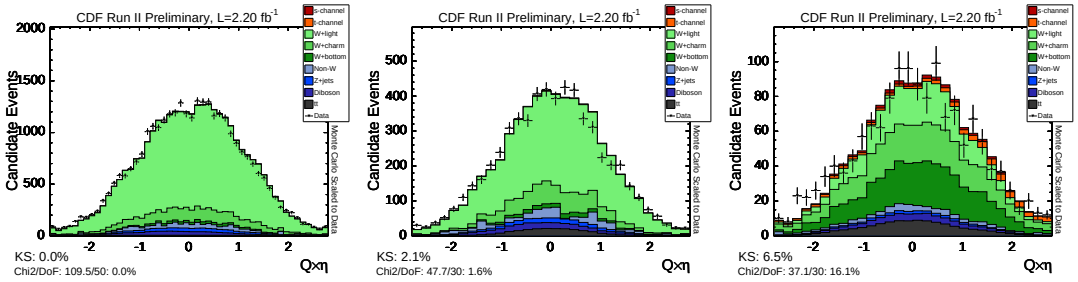

Figure 83: Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

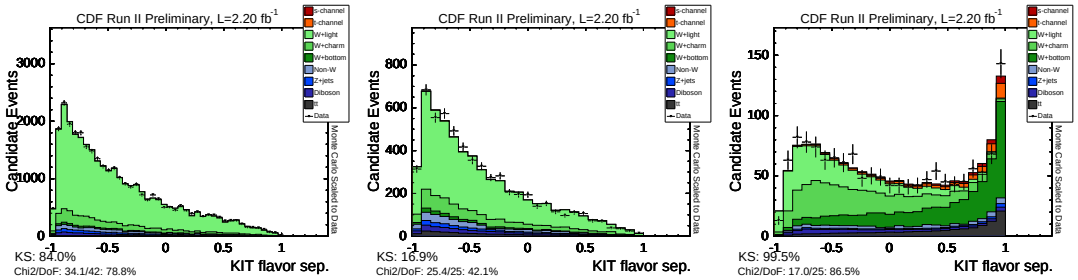

Figure 84: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

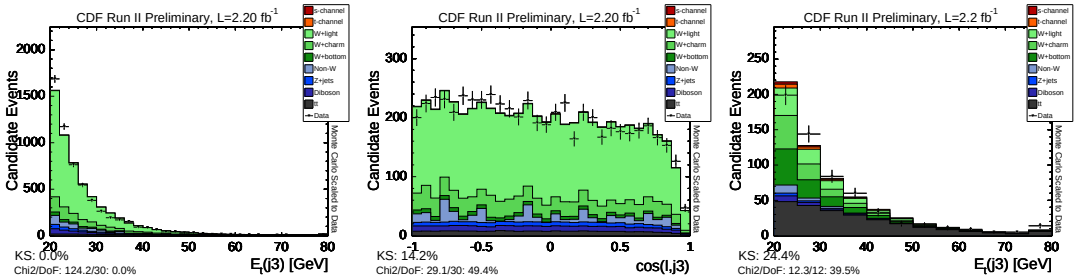

Figure 85: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

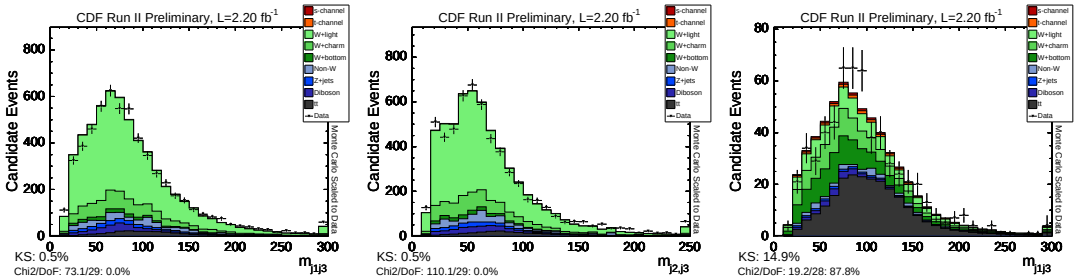

Figure 86: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

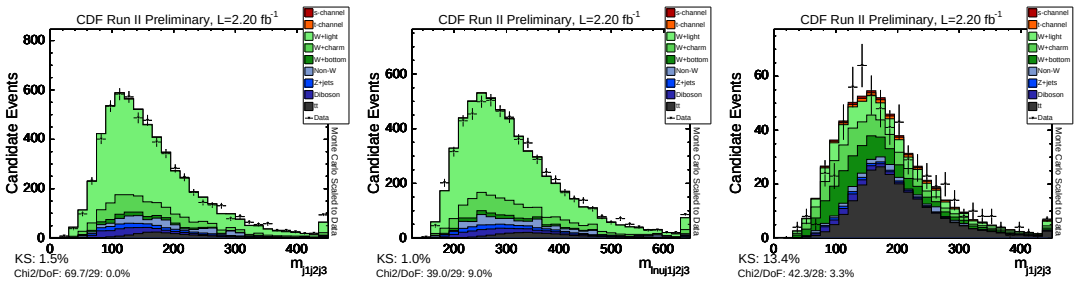

Figure 87: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.

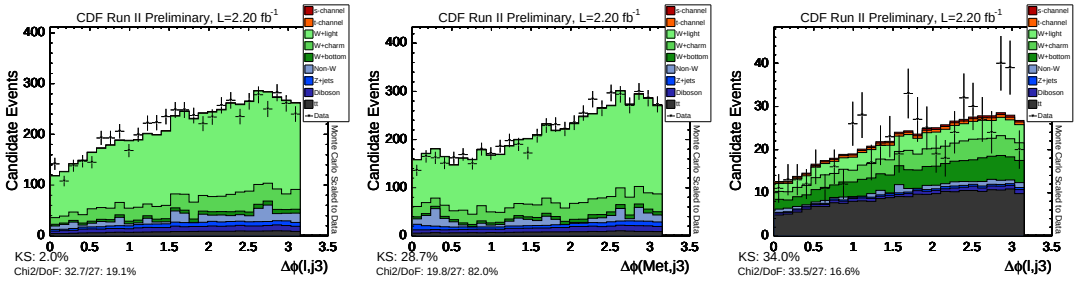

Figure 88: . Left is untagged 2-jet bin, middle is untagged 3-jet bin, and right is single tagged 2-jet bin.