CDF/ANAL/EXOTIC/CDFR/9404
Version 4.1
November 10, 2008

Search for Standard Model Higgs Boson Production in Association with a W Boson using Matrix Element and Boosted Decision Trees Techniques with 2.7 fb^{-1} of data

Peter Dong, Bernd Stelzer, Rainer Wallny[1],
Florencia Canelli, Craig Group, Enrique Palencia[2],
Bruno Casal, Alberto Ruiz, Rocío Vilar[3]
Bárbara Álvarez, Javier Cuevas[4]

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

Abstract

We present a search for Standard Model Higgs boson production in association with a W boson using 2.7 fb^{-1} of CDF II data collected between 2002 and 2008. This search is performed using two complementary techniques. A matrix element technique is used to calculate event probability densities for the signal and background hypothesis. A multivariate technique, based on Boosted Decision Trees, is used to combine the event probability densities with kinematic variables to build a final discriminant distribution which is fitted to the data using a binned likelihood approach.

We observe no evidence for a Higgs boson signal and set 95% confidence level upper limits on the WH production cross section times the branching ratio of the Higgs boson to decay to $b\bar{b}$ pairs of $\sigma(p\bar{p} \to WH) \times BR(H \to b\bar{b})/SM < 3.52$ to 107.6 for Higgs boson masses between $m_H = 100$ GeV/c^2 and $m_H = 150$ GeV/c^2. The expected (median) limit estimated in pseudo-experiments is $\sigma(p\bar{p} \to WH) \times BR(H \to b\bar{b})/SM < 4.07$ to 52.1 at 95% C.L.

Contents

1 Introduction

Finding evidence for Higgs boson production in association with a W boson is extremely challenging since it is rarely produced ($\sigma_{WH} \sim 0.1$ pb) in comparison with other processes with the same final state like $W + b\bar{b}$ and top. The signal to background ratio of the analysis is tiny, typically on the order of S/B$\sim$1/100 [1, 2]. In this note we describe two complementary analysis techniques in order to achieve better discrimination of signal and background events by using more information to characterize each event.

This note describes our search for WH production using a matrix element analysis and a boosted decision trees. The implementation of these methods is very similar to our single-top search described in detail here [3, 4]. We combine both techniques using the discriminant from the matrix element analysis as an input variable of the Boosted Decision analysis.

The next Sec. of the note explains the event selection and background estimate for 2.7 fb^{-1} of data. We dedicate Sec. 3.1 to explain the Matrix Element (ME) technique. This section includes the new transfer functions in Sec. 3.2.2. Sec. 4 describes the Boosted Decision Trees (BDT) technique and how we incorporate ME output into BDT. The rest of the note discusses the systematic uncertainties and the results on the data.

The matrix element analysis described here is very similar to the previous one described in [5] and blessed for winter conferences 2008. There are only a few minor changes noted here and described in more detail later on:

- We have slightly changed the event selection in order to synchronize with the WH neural network analysis [6]. More details can be found in Sec. 2. This will help for future combinations.
- We have a new b-jet transfer function (the transfer functions for light and gluon jets did not change). This new TF uses the output of a neural network as input variable instead of the energy of the jets. More details can be found in Sec. 3.2.2.
- We have only one b-jet transfer function for jets in all η region.
- We have increased the number of Higgs mass points. We have from 100 to 150 GeV/c^2 in 5 GeV/c^2 steps.
- We have added the latest data available, data up to period 17 ($\sim$2.7 fb^{-1}).

2 Event Selection and Expected Number of Events

We use very similar event selection as in the single top analysis [3]. The only differences with respect to that analysis are that we relax here the Missing Transverse Energy (MET) cut and that we reject forward jets. Candidate events for this analysis are

selected by requiring a W + 2 jet event topology where the W decays leptonically, $W \to e\nu_e$ or $W \to \mu\nu_\mu$. Muon events are triggered by the high p_T CMUP and CMX trigger as well as through MET+2 jets triggered events as described in [8]. Electron events are triggered by the high p_T CEM trigger as well as through the MET+PEM20 trigger. We require events to contain an isolated electron or muon with offline E_T or $p_T > 20$ GeV and MET > 20 GeV (25 for Phoenix). Jets are clustered with a cone size of $\Delta R < 0.4$ and are required to have $E_T > 20$ GeV after jet corrections at Level5 have been applied and have $|\eta_{detector}|$ <2.0. One or both of the two jets should be identified as a b-jet using the secondary vertex tag (SECVTX) requirement. We veto dilepton, Z-boson, conversion and QCD multi-jet events. We use a rather tight QCD veto implementation (single top QCD veto) which limits the amount of QCD non-W. The implementation of the veto is described in [9] and it is only applied to single tag events and all Phoenix events. The background estimate methodology is identical to our single top analysis and is summarized in a dedicated CDF note [1]. Chapter 1-5 in [1] summarizes the procedure to obtain the signal and background yield for events with a W+jets event topology including b-tagged jets (a.k.a Method2). Chapter 6 in [1] shows the details of the (single-top style) background estimate as used in this analysis.

In this note, we provide a summary with the new selection changes (looser MET, central jets), and Table 1 lists the expected event yield, in the 2 jet bin, in 2.7 fb^{-1} of data. For the MC based background estimations, we have used the theoretical cross sections shown in Table 2. We have added the expected contribution from WH signal using the SM cross sections and branching ratios shown in Table 3. The event yield for all jet multiplicities and subdetectors can be found in Appendix B.

Figure 1 shows the acceptance of WH events as a function of jet-multiplicity in the event.

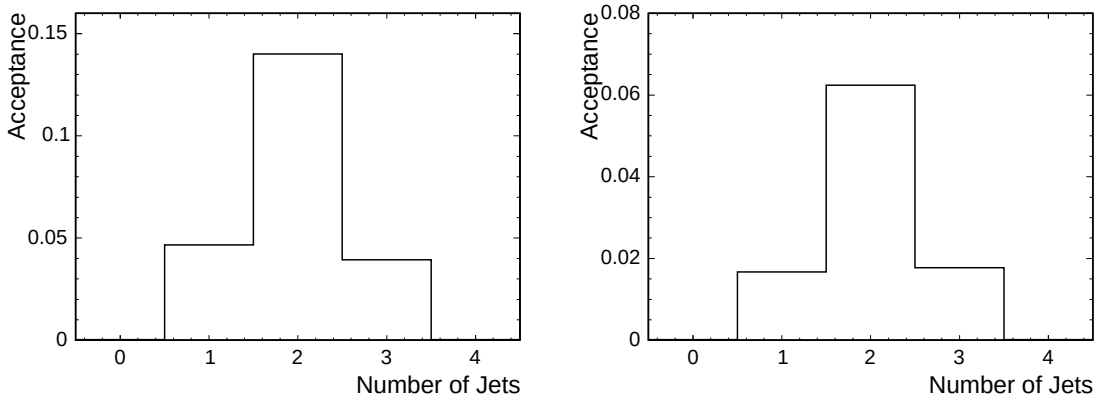

Figure 1: Acceptance of WH events as function of jet-multiplicity. The plot on the left shows the sample before b-tagging requirements. The plot on the right shows the distribution with exactly one tagged jet.

Process	1 tag	2 tags	MC sample
All Pretag Cands.	50644.0 ± 0.0	57174.0 ± 0.0	
WW	56.2 ± 6.2	0.4 ± 0.1	itopww
WZ	23.0 ± 1.7	4.8 ± 0.5	itopwz
ZZ	0.8 ± 0.1	0.2 ± 0.0	itopzz
TopLJ	121.3 ± 17.1	23.8 ± 3.9	ttop75
TopDil	48.8 ± 6.8	14.1 ± 2.3	ttop75
STopT	64.0 ± 9.3	1.8 ± 0.3	stopm0
STopS	40.6 ± 5.7	12.8 ± 2.1	stop00
Z+jets	37.4 ± 5.5	2.1 ± 0.3	ztopXY
Total MC	392.0 ± 35.0	59.9 ± 7.5	—
Wbb	538.7 ± 162.5	70.3 ± 22.5	btopXp
Wcc/Wc	489.1 ± 150.9	6.8 ± 2.3	ctopXw, stopwX
Total HF	1027.8 ± 312.3	77.1 ± 24.7	—
Mistags	458.0 ± 57.9	2.2 ± 0.6	ptopXw, utopXw
Non-W	135.5 ± 54.2	9.0 ± 3.6	non-W
Total Prediction	2013.3 ± 324.1	148.2 ± 26.1	—
WH100	9.5 ± 0.8	2.9 ± 0.3	chgt2k
WH105	8.6 ± 0.7	2.7 ± 0.3	chgt2l
WH110	7.6 ± 0.6	2.4 ± 0.3	chgt1a
WH115	6.3 ± 0.5	2.0 ± 0.2	chgt1b
WH120	4.9 ± 0.4	1.6 ± 0.2	chgt1c
WH125	4.0 ± 0.3	1.3 ± 0.2	chgt2m
WH130	3.1 ± 0.3	1.0 ± 0.1	chgt1d
WH135	2.3 ± 0.2	0.7 ± 0.1	chgt2n
WH140	1.5 ± 0.1	0.5 ± 0.1	chgt1e
WH145	1.0 ± 0.1	0.3 ± 0.0	chgt2o
WH150	0.7 ± 0.1	0.2 ± 0.0	chgt1f
Observed	1998.0 ± 0.0	156.0 ± 0.0	—

Table 1: Number of expected signal and background events, in the 2 jet bin, in 2.7 fb^{-1} of CDF data, passing all the event selection requirements.

3 Matrix Element Analysis

The matrix element analysis relies on the evaluation of event probability densities for signal and background processes based on the Standard Model differential cross-section calculation.

In general a differential cross-section is given by [7]:

$$d\sigma = \frac{(2\pi)^4 |M|^2}{4\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2}} \, d\Phi_n(q_1 + q_2; p_1, .., p_n) \tag{1}$$

Process	Theoretical Cross Section
s-channel	0.884 ± 0.11
t-channel	1.980 ± 0.25
WW	12.4 ± 0.25
WZ	3.96 ± 0.06
ZZ	1.58 ± 0.05
$t\bar{t}$	6.7 ± 0.8
Z+jets	787.4 ± 85.0

Table 2: Theoretical cross sections used for the MC based background estimation.

Higgs Mass (GeV/c^2)	BR($H \to b\bar{b}$)	σ (pb)	σ x BR($H \to b\bar{b}$) (pb)
100	0.812	0.286	0.232
105	0.796	0.253	0.201
110	0.770	0.219	0.169
115	0.732	0.186	0.136
120	0.679	0.153	0.104
125	0.610	0.136	0.083
130	0.527	0.120	0.063
135	0.436	0.103	0.045
140	0.344	0.086	0.030
145	0.256	0.078	0.020
150	0.176	0.070	0.012

Table 3: SM branching ratios ($H \to b\bar{b}$) and cross sections for all Higgs masses.

where $|M|$ is the Lorentz invariant matrix element; q_1, q_2 and m_{q_1}, m_{q_2} are the four momenta and masses of the incident particles; and $d\Phi_n$ is the n-body phase space given by [7]:

$$d\Phi_n(q_1+q_2;p_1,..,p_n) = \delta^4(q_1+q_2-\sum_{i=1}^{n} p_i)\prod_{i=1}^{n}\frac{d^3p_i}{(2\pi)^3 2E_i} \tag{2}$$

The CDF detector would be 'ideal' if we could measure all four momenta of the initial and final state particles very precisely. In this case we could use this formula without modification and normalize it to the total cross section to define the event probability:

$$P_{evt} \sim \frac{d\sigma}{\sigma}.$$

However, several effects have to be considered: (1) the initial state interaction is initiated by partons inside the proton and antiproton, (2) neutrinos in the final state are not identified directly, and (3) the energy resolution of the detector can not be ignored. To address the first point, the differential cross section is folded over the parton distribution functions. To address the second and third points, we integrate

over all particle momenta which we do not measure (e.g. p_z of the neutrino), or do not measure very well, due to resolution effects (e.g. jet energies). The integration reflects the fact that we want to sum over all possible particle variables (y) leading to the observed set of variables (x) measured with the CDF detector. The mapping between the particle variables (y) and the measured variables (x) is established with the transfer function, $W(y,x)$. After incorporating the effects mentioned above, the event probability takes the form:

$$P(x) = \frac{1}{\sigma} \int d\sigma(y) dq_1 dq_2 f(y_1) f(y_2) W(y,x) \tag{3}$$

where $d\sigma(y)$ is the differential cross section in terms of the particle variables; $f(y_i)$ are the PDFs, with y_i being the fraction of the proton momentum carried by the parton ($y_i = E_{q_i}/E_{beam}$); and $W(y,x)$ is the transfer function. Substituting Equation 1 and 2 into Equation 3, and considering a final state with four particles (n=4), the event probability becomes:

$$P(x) = \frac{1}{\sigma} \int 2\pi^4 |M|^2 \frac{f(y_1)}{|E_{q_1}|} \frac{f(y_2)}{|E_{q_2}|} W(y,x) d\Phi_4 dE_{q_1} dE_{q_2} \tag{4}$$

where the masses and transverse momenta of the initial partons are neglected (i.e. $\sqrt{(q_1 \cdot q_2)^2 - m_{q_1}^2 m_{q_2}^2} \simeq 2E_{q_1}E_{q_2}$).

As it will be explained in the next sections, in this analysis we calculate event probabilities with the signal hypothesis for WH and with background hypothesis with the major contributions.

In Sec. 3.1 we show the different matrix element used in the event probabilities. In Section 3.2.2 we describe the neural network used to reproduce the jet energies and derive the transfer function and make certain assumptions on the mapping between particles and measured objects. Section 3.3 shows the calculation of the phase space factor. Section 3.4 adds all these pieces into event probabilities. We explain the construction an event probability discriminant using the signal and background probabilities in Sec. 3.6.

3.1 Matrix Elements for signal and backgrounds

In this analysis, we calculate event probability densities for the WH signal, as well as for the s-channel and t-channel single top, $t\bar{t}$, $Wb\bar{b}$, Wcj, Mistags (Wgg) and diboson background processes. The $Wc\bar{c}$ and non-W events are assumed to be represented fairly well by the background probability density.

We calculate the matrix element ($|M|^2$) for the event probability density at leading order perturbation theory by using the HELAS (HELicity Amplitude Subroutines for Feynman Diagram Evaluations) package [10]. The correct subroutines for a given process are automatically generated by the MadGraph program [11]. We use different subroutines for calculating event probabilities for the WH signal and s-channel, t-channel, $Wb\bar{b}$, Wcj, Wgg, diboson and $t\bar{t}$ background hypotheses. Figures 2 to 6 show the different Feynman diagrams used for each channel.

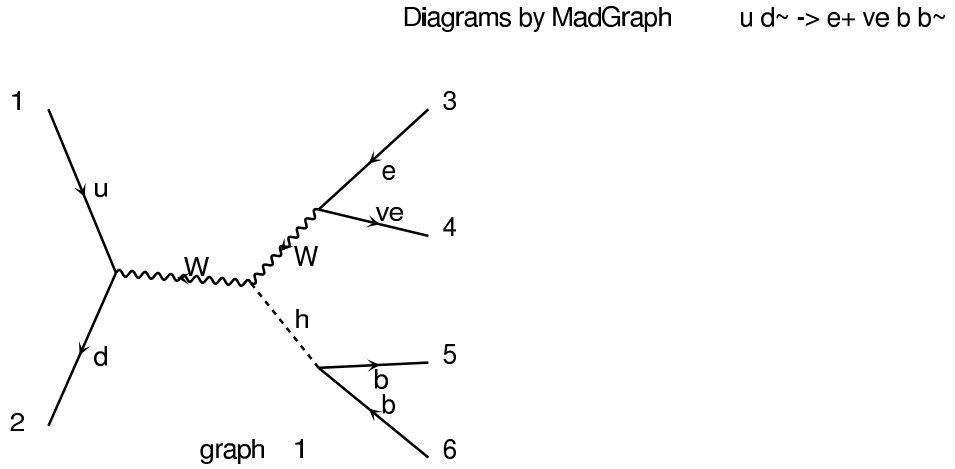

Figure 2: Leading order Feynman diagram for Higgs production in association with a W boson. The Higgs boson is set to decay into a pair of bottom quarks.

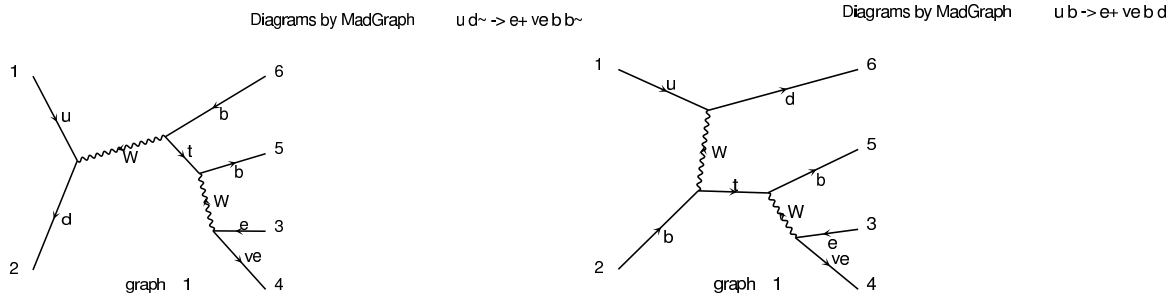

Figure 3: Leading order Feynman diagram for s-channel (left) and t-channel (right) single top quark production and decay used in the calculation of the single top event probability.

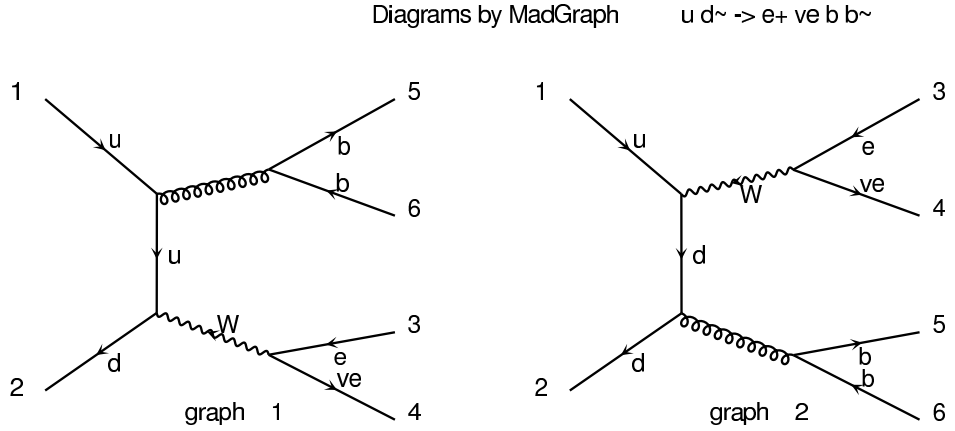

Figure 4: Leading order Feynman diagram for $Wb\bar{b}$ production and decay used in the calculation of the $Wb\bar{b}$ event probability.

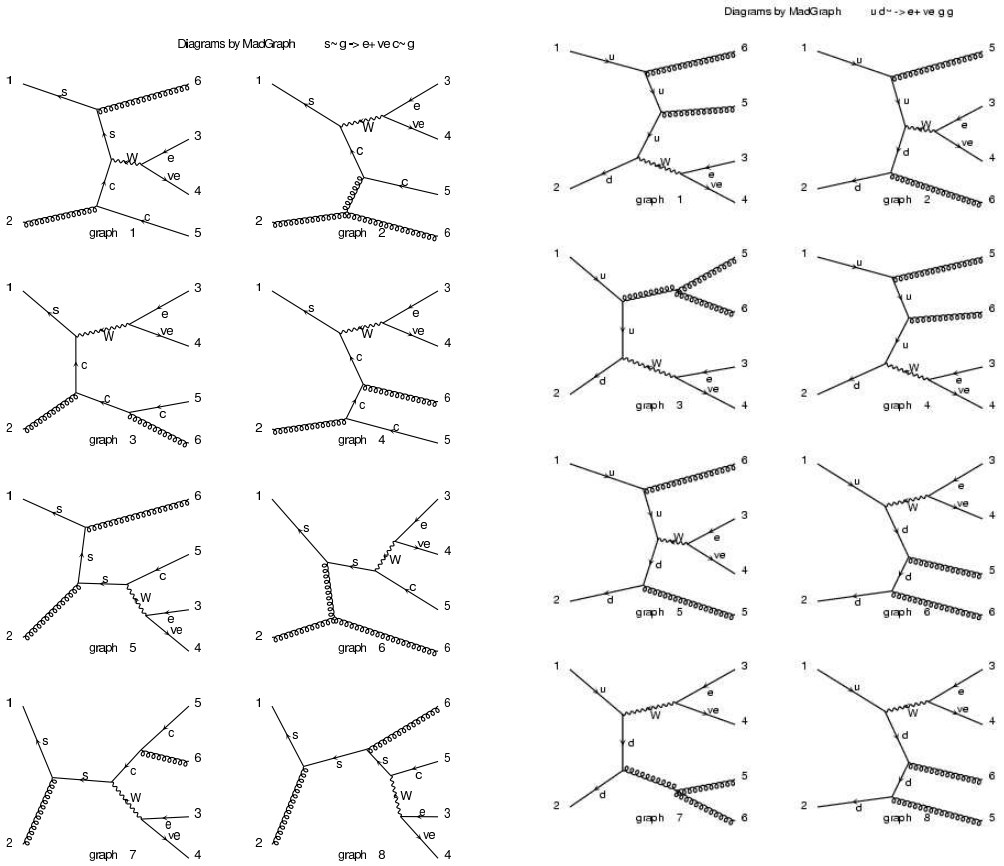

Figure 5: Leading order Feynman diagrams for Wcj (left) and Wgg (right) production and decay used in the calculation of the event probability density.

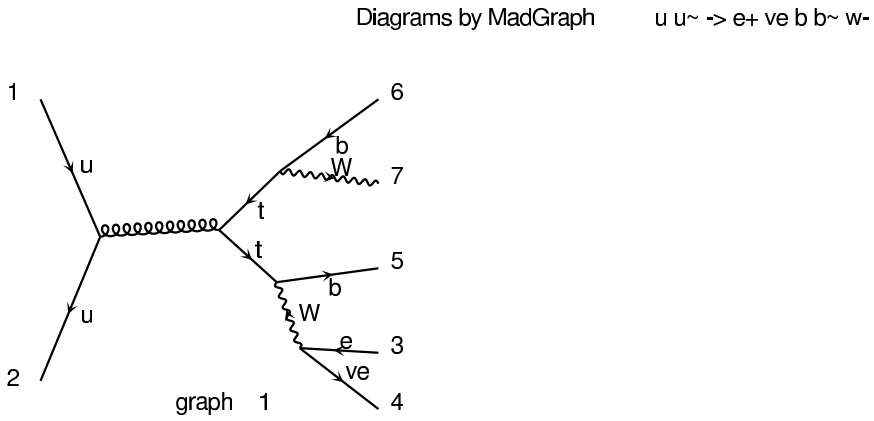

Figure 6: Leading order Feynman diagrams for $t\bar{t}$ production and decay used in the calculation of the $t\bar{t}$ event probability.

3.2 Transfer Function: TF(E_p, E_j)

The transfer function, $W(y,x)$, provides the probability of measuring the set of observable variables (x) that correspond to the set of production variables (y). The set (y) represents all final state particle momenta at the particle level, while the set (x) represents the measured momenta (of the corresponding object) with the CDF detector. In the case of well-measured objects, $W(y,x)$ is taken as a δ-function (i.e. the measured momenta are used in the differential cross section calculation). When the detector resolution cannot be ignored, $W(y,x)$ is taken as a Gaussian-type function. For unmeasured quantities, like the momenta of the neutrino, the transfer function is unity (the transverse momenta of the neutrino, however, can be inferred from energy and momentum conservation).

Lepton momenta are well-measured with the CDF detector and we will assume δ-functions for them (first factor of Equation 5). The jet angular resolution of the calorimeter is also good (on the order of $\sigma_{\Delta_R} \simeq 0.07$) and we assume δ-functions for the transfer function of the jet directions (second factor of Equation 5). The resolution of the measured jet energies, however, is not negligible and the transfer function needs to be derived. Using these assumptions, $W(y,x)$ takes the following form for the four particle final state we consider in the WH search (lepton, neutrino and two jets):

$$W(y,x) = \delta^3(\vec{p_l}^{\,y} - \vec{p_l}^{\,x}) \prod_{i=1}^{2} \delta^2(\Omega_i^y - \Omega_i^x) \prod_{j=1}^{2} W_{jet}(E_{parton_j}, E_{jet_j}) \tag{5}$$

where $\vec{p_l}^{\,y}$ and $\vec{p_l}^{\,x}$ are the produced and measured lepton momenta, Ω_i^y and Ω_i^x are the produced quark and measured jet angles, and E_{parton_j} and E_{jet_j} are the produced quark and measured jet energies.

The transfer between parton and jet energies is determined by the transfer function $W_{jet}(E_{parton}, E_{jet})$. The standard CDF jet energy corrections correct the energies of jets in a way that the means of the corrected jet energies and the original parton energies are equal. Such corrections, however, do not account for the shape of the difference in energies: the shape of the $\delta_E = (E_{parton} - E_{jet})$ distribution. This distribution is asymmetric and features a significant tail at positive δ_E, as shown in Figure 7.

We parameterize the δ_E distribution as a sum of two Gaussian functions: one to account for the sharp peak and one to account for the asymmetric tail:

$$W_{jet}(E_{parton}, E_{jet}) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)} (exp\frac{-(\delta_E - p_1)^2}{2p_2^2} + p_3\ exp\frac{-(\delta_E - p_4)^2}{2p_5^2}) \tag{6}$$

where the parameters p_i have a linear dependence on E_{parton}, i.e.

$$p_i = a_i + b_i E_{parton}$$

A total of 10 parameters $(a_1, b_1, .., a_5, b_5)$ are therefore required to specify $W_{jet}(E_{parton}, E_{jet})$.

After specifying the transfer function, we can apply the general event probability of Equation 4 to the case of the WH analysis.

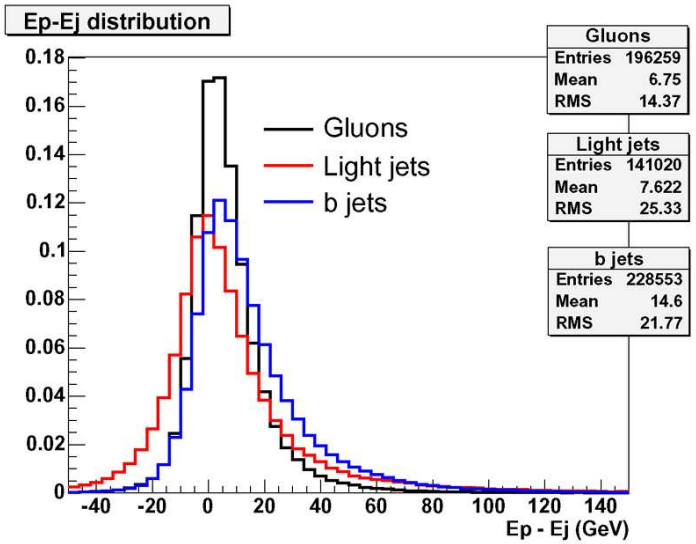

Figure 7: Distribution of $\delta_E = (E_{parton} - E_{jet})$ for matched jets to partons in s-channel (b-jets), t-channel (light-jets), Wc (gluon-jets) Monte Carlo events (passed through full detector simulation).

3.2.1 Parameters for $W_{jet}(E_{parton}, E_{jet})$ for light and gluon jets

We use the same transfer functions used in the matrix element analysis for the winter 2008 (2.2 fb^{-1}). We determine the parameters of the transfer function $W_{jet}(E_{parton}, E_{jet})$ for light jets using the light jets in the t-channel single top sample. These transfer functions are also applied to the background probabilities with light-jets. We parameterize the transfer functions of jet energies inside a cone of R_{cone}=0.7. This can be done only for events where we can match our selected R_{cone}=0.4 jets with R_{cone}=0.7 jets which is about 84% of the events. (Tight R_{cone}=0.7 jets are defined in the same way as R_{cone}=0.4, p_T >20 GeV and $|\eta|$ <2.0.) For the cases where we can't match two R_{cone}=0.4 jets to two R_{cone}=0.7 jets we parameterize a transfer function with jet energies of R_{cone}=0.4 jets. A different transfer function is obtained for R_{cone}=0.4 and R_{cone}=0.7 jets in different $|\eta|$ regions defined as: 0-0.9, 0.9-1.2, and 1.2-2.8.

We also implemented transfer functions obtained from gluons which are later applied to all the probabilities with gluons in the final state. (This transfer functions were obtained only in one region of η and derived from the Wc MC samples.)

We apply all event selection requirements and match reconstructed jets to their corresponding quarks. The quark is required to be aligned within a cone of $\Delta R < 0.4$ around the reconstructed jet-axis in order to be considered 'matched'. We correct the matched jet up to level 5 corrections and write out the jet energy together with the energy of the original particle. As an example, the distribution of measured jet energies for R_{cone}=0.4 versus the original parton energy is shown in the plot of Fig. 8. The parameters of the transfer function are then derived by performing a maximum likelihood fit to these events. If $\mathrm{n}(E_{jet}, E_{parton})dE_{jet}dE_{parton}$ is the number of jets with

jet energies between E_{jet} and $E_{jet} + dE_{jet}$, and particle energies between E_{parton} and $E_{parton} + dE_{parton}$ in this sample, then:

$$n(E_{jet}, E_{parton})dE_{jet}dE_{parton} = n(E_{parton})dE_{parton}W_{jet}(E_{parton}, E_{jet})dE_{jet} \quad (7)$$

where $n(E_{parton})dE_{parton}$ is the number of particles with an energy between E_{parton} and $E_{parton} + dE_{parton}$. The parameters of $W_{jet}(E_{parton}, E_{jet})$ are determined such to maximize the agreement in Equation 7.

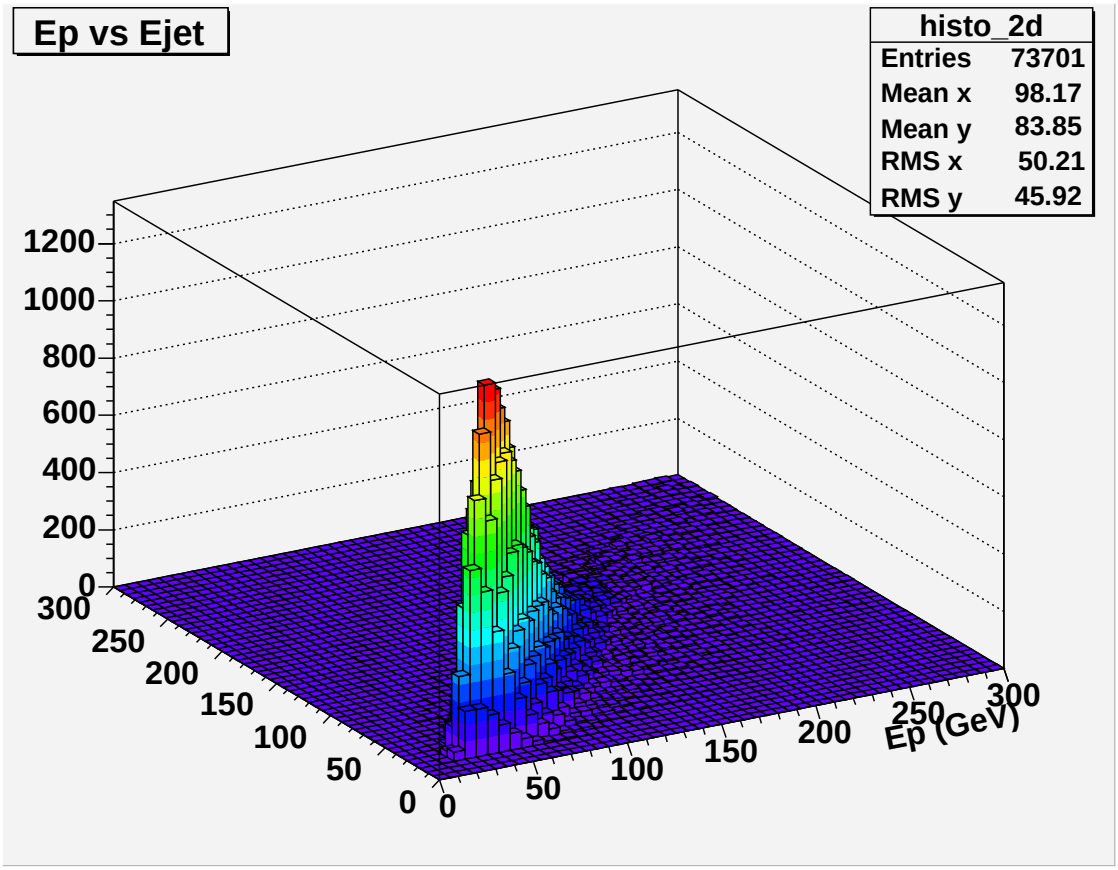

Figure 8: Lego plot of E_{parton} vs E_{jet} (passed through full GEANT detector simulation) for a sample of matched jets of R_{cone}=0.4 to partons in WH Monte Carlo events.

The performance of the parametrization is best shown in Fig. 9, which shows the $\delta E = (E_{parton} - E_{jet})$ distribution (histogram) compared to the prediction from the transfer function (solid line).

3.2.2 Parameters for $W_{jet}(E_{parton}, E_{jet})$ for b and c jets

In order to better reproduce the real parton energy for b and c jets we train a Neural Network (NN) with the Stuttgart Neural Network Simulator (SNNS) using the ROOT interface ROOTSNNS v3.0 [12]. The NN used has one hidden layer with 13 hidden

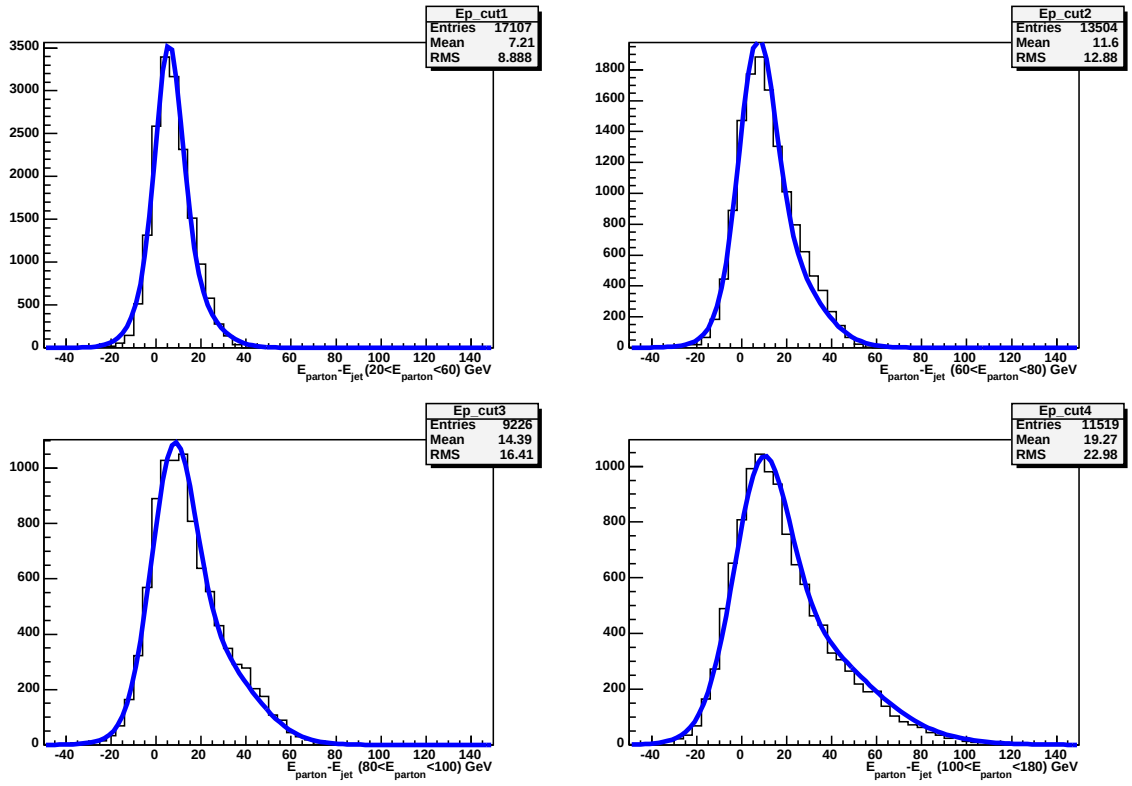

Figure 9: Distributions of $\delta E = (E_{parton} - E_{jet})$ for different ranges of parton energy of matched jets to partons (20-60, 60-80, 80-100, 100-180 GeV). The histograms are WH Monte Carlo events after full detector simulation and R_{cone}=0.4 jet (level 5) corrections. The solid line corresponds to the transfer function using the parameters of R_{cone}=0.4.

nodes. We use 900 epochs for the training. The learning function used in the training is called Scaled Conjugate Gradient descent (SCG). The NN is trained using only WH signal events of several Higgs masses (from 100 to 150 GeV/c^2). We use 1,023,000 jets from WH events with the same amount of jets (93,000) of each Higgs mass.

For the training we use 8 input variables related to the jet kinematics: the total energy of the jet corrected at level 5 (E_j), the sum of the transverse momentum of the tracks in the jet[1] (SumE), the transverse momentum of the jet (p_T), the ϕ and η of the jet, the fraction of energy deposited for the jet in the electromagnetic calorimeter (EMF), the raw (measured) energy of the jet (RawE_j), and the energy of the jets with cone size 0.7[2] (Ejcone7). The distributions of these variables for data and MC are shown in Appendix D.

As for the light and gluon transfer functions the selection applied for the training is exactly the same one as the pretag selection of the analysis with an additional

[1] SumE $= \Sigma \frac{p_{T\ track}}{sin(2 \cdot atan(e^{-\eta_{track}}))}$, $\eta = -\frac{1}{2} \cdot ln(tan\ \theta) \Rightarrow$ SumE $= \Sigma \frac{p_{T\ track}}{sin(\theta)}$.

[2] 16% of the times there are no jets of cone size 0.7 available. In this case, we use the energy of the jet of cone size 0.4.

requirement: the ΔR between the jet and the partons has to be less than 0.4 (ΔR (parton, jet) < 0.4). We remove events that have 2 partons (only b-jets) that match to the same jet of cone size 0.4 (only $\sim$2% of the total).

The NN has one only output variable that is an estimate of the energy of the parton (E_p). The output of the NN for data and MC is shown in Fig. 10.

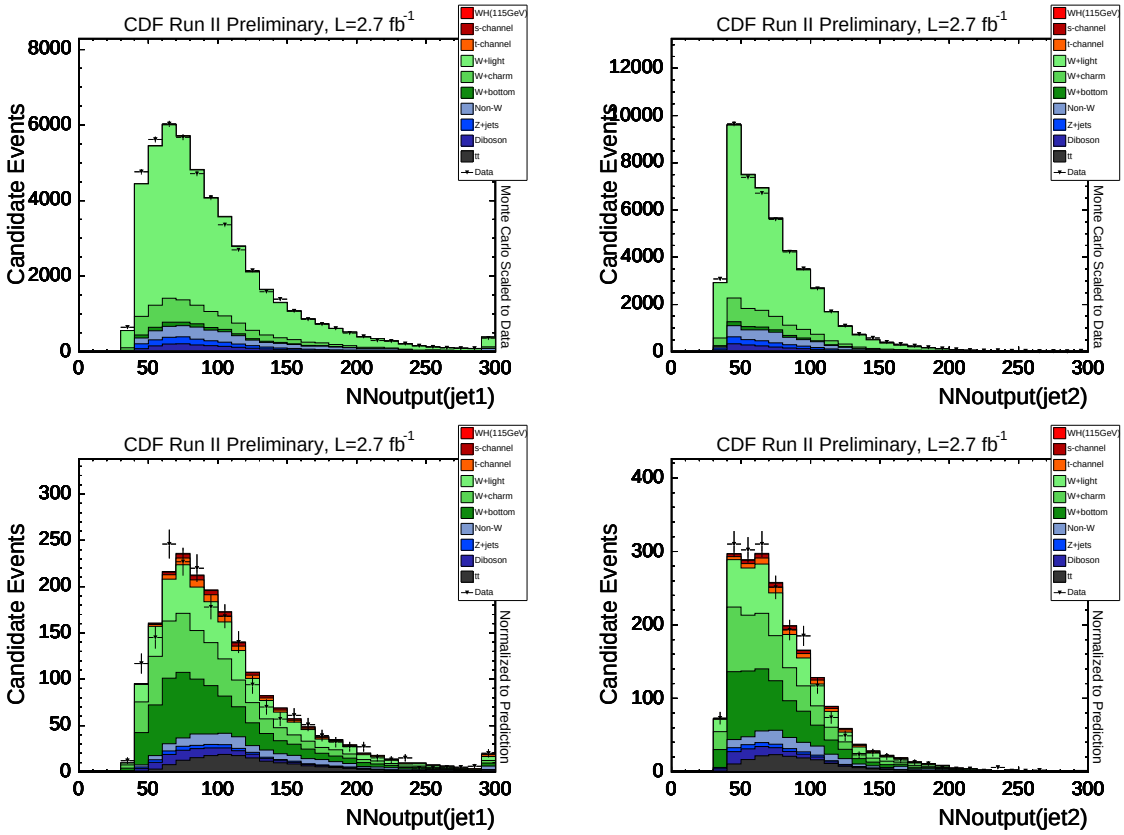

Figure 10: Left (Right): Distributions of the output of the NN for the first (second) jet. The row on top (bottom) is for the untagged (single tag) sample.

The goal of using a NN to reproduce the parton energy is to use the output of the NN (instead of the jet energy, E_j) as input of the transfer function. Figure 11 shows the difference between the parton energy and the jet energy and the NN output. It is clear that the NN output is closer to the parton energy than the level 5 corrected jet energies and that the distribution is also narrower. Therefore, since the NN output provides a better jet resolution, using it as input of the transfer function should help to improve the performance of the transfer function.

We proceed to parameterize the transfer function in the same way done for the light and gluon jets. In this case, we parameterize the $\delta_E \equiv E_p - NN$ instead of $E_p - E_j$ distribution as was done in the previous section. The distribution of the NN output versus the original parton energy of the b quark is shown in Fig. 12.

The functional form used to parameterize δ_E is the same one described in Sec. 3.2.1,

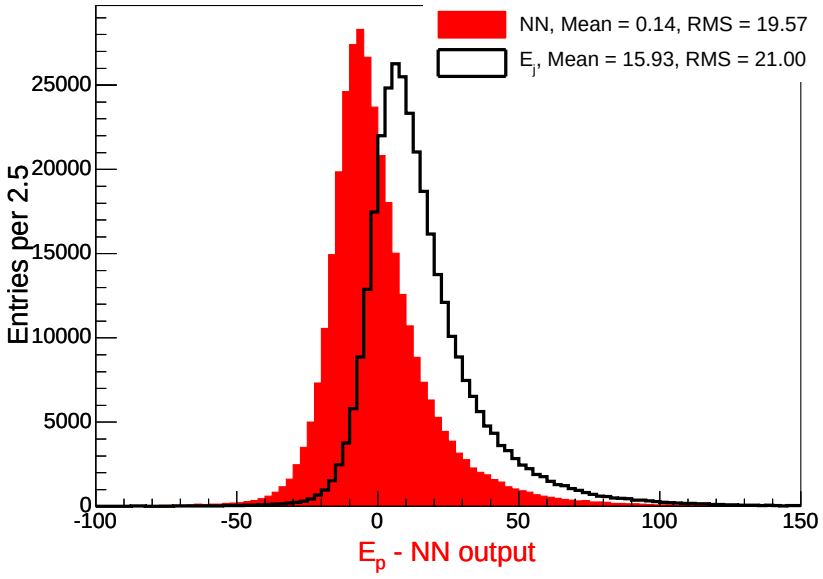

Figure 11: Difference between the parton energy and the jet energy (black) and the NN output (red).

that is a sum of two Gaussian functions given by Eq. 6. The performance of this parameterization is shown in Fig. 13, which shows a good agreement between the $\delta_E \equiv E_p - NN$ distribution (black histogram) compared to the prediction from the new TF (solid blue line).

Finally, we show in the upper plot of Fig. 14 the event probability discriminant for signal and some background events obtained using the new (using NN) and old (using E_j) transfer functions. The bottom plot shows the significance as a function of the cut in the event probability discriminant for both transfer functions. We find there is no evident improvement using the new transfer function.

3.3 Phase Space

The integration of the differential cross section has to be performed over 14 variables corresponding to the momentum vectors of the four final state particles (12 variables) and the longitudinal momenta of the initial state partons (2 variables). There are 11 δ-functions inside the integrals: four for total energy and momentum conservation (part of the phase space factor, see Equation 2) and seven for the transfer function (three for the lepton momentum vector and four for the jet angles, see Equation 5). The calculation of the event probability therefore involves a three dimensional integration. The integration is performed numerically over the absolute value of the quark momenta ($\rho_i = |\vec{p}_i|$) and the longitudinal momentum of the neutrino ($p_{\nu,z}$). The phase space for WH events is derived in Appendix C and has been expressed as a function of the

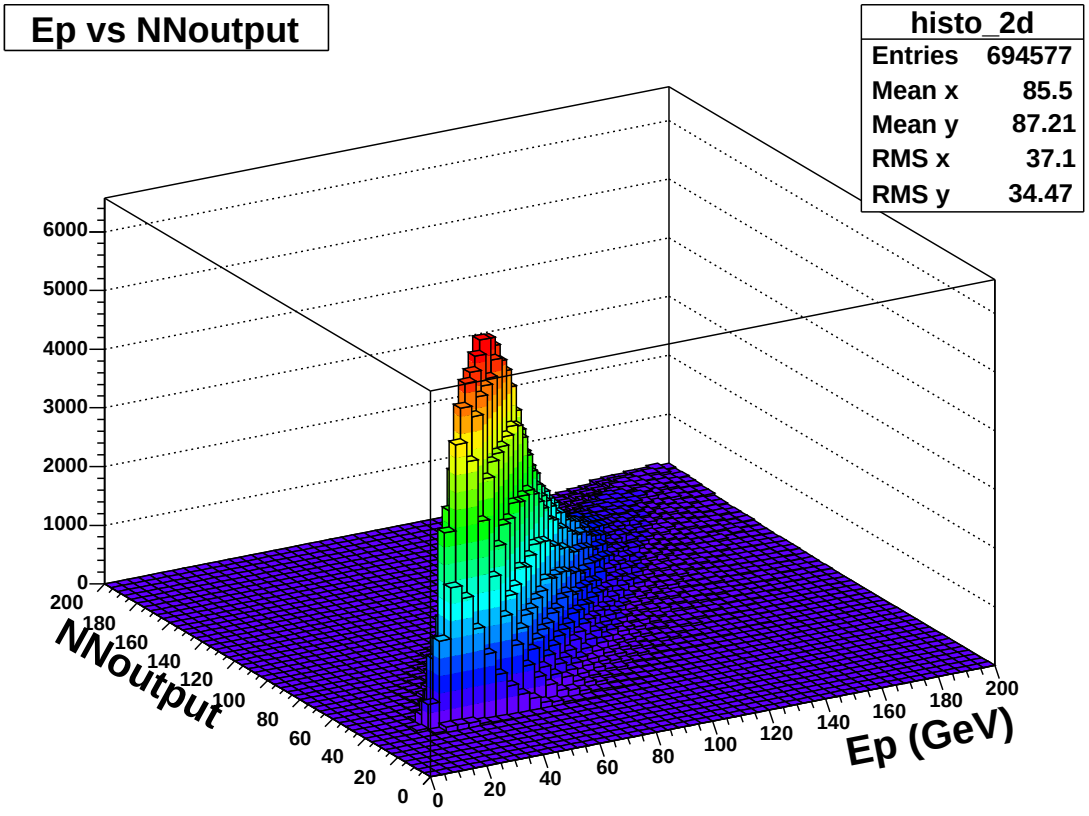

Figure 12: Lego plot of E_{parton} vs NNoutput (passed through full GEANT detector simulation) for a sample of matched jets of R_{cone}=0.4 to partons in WH Monte Carlo events. .

variables $(\rho_1, \Omega_1, \rho_2, \Omega_2, \vec{p}_l, m_W^2)$. The result is:

$$\begin{aligned} d\Phi_4 &= \delta(E_{q_1} + E_{q_2} - \sum_{i=1}^{4} E_i)\ \delta(p_{q_1,z} + p_{q_2,z} - \sum_{i=1}^{4} p_{i,z}) \\ &\times \frac{dm_W^2}{|2E_l \frac{p_{\nu,z}}{E_\nu} - 2p_{l,z}|} \frac{d^3\vec{p}_l}{(2\pi)^3 2E_l} \frac{1}{(2\pi)^3 2E_\nu} \prod_{i=1}^{2} \frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i} \end{aligned} \quad (8)$$

3.4 Event Probability Densities

Substituting the phase space factor (Equation 8) and the transfer function (Equation 5) into the expression for the event probability (Equation 4), we obtain:

$$\begin{aligned} P(x) = \frac{1}{\sigma} \int |M|^2 \frac{f(y_1)}{|E_{q_1}|} \frac{f(y_2)}{|E_{q_2}|} \delta^3(\vec{p}_l{}^y - \vec{p}_l{}^x) \prod_{i=1}^{2} \delta^2(\Omega_i^y - \Omega_i^x) \prod_{j=1}^{2} W_{jet}(E_{parton,j}, E_{jet,j}) \Phi_4 \\ \times \delta(E_{q_1} + E_{q_2} - \sum_{i=1}^{4} E_i) \delta(p_{q_1,z} + p_{q_2,z} - \sum_{i=1}^{4} p_{i,z}) d^3\vec{p}_l dm_W^2 dE_{q_1} dE_{q_2} \prod_{i=1}^{2} d\rho_i d\Omega_i \end{aligned} \quad (9)$$

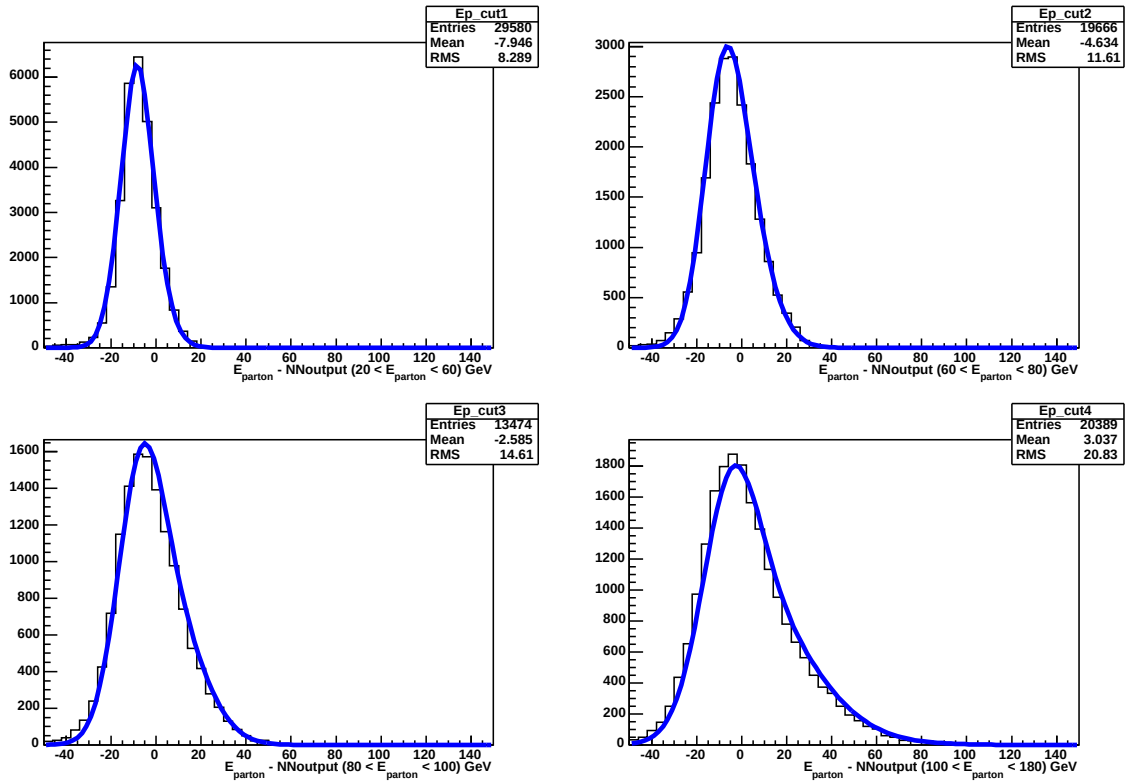

Figure 13: Distributions of $\delta E = (E_{parton} - NN)$ for different ranges of parton energy of matched jets to partons (20-60, 60-80, 80-100, 100-180 GeV). The histograms are WH MC events after full detector simulation and R=0.4 jets with level 5 corrections. The solid line corresponds to the transfer function using the parameters of R=0.4 jets.

where,

$$\Phi_4 = \frac{2\pi^4}{(2\pi)^9}\frac{1}{2E_l}\frac{1}{2E_\nu}\frac{\rho_1^2}{2E_1}\frac{\rho_2^2}{2E_2}\frac{1}{|2E_l\frac{p_{\nu,z}}{E_\nu} - 2p_{l,z}|}.$$

The integration over E_{q_1} and E_{q_2} eliminates the two δ-functions in the second line of Equation 9. The integration over the lepton momenta and the quark solid angles eliminate the δ-functions in the first line of Equation 9 associated with $W(y,x)$. The final event probability takes the form:

$$P(x) = \frac{1}{\sigma}\int d\rho_1 d\rho_2 dm_W^2 \sum_{comb,\nu} |M|^2 \frac{f(q_1)}{|q_1|}\frac{f(q_2)}{|q_2|}\Phi_4 \prod_{i=1}^{2} W_{jet}(E_{parton,i}, E_{jet,i}) \qquad (10)$$

For events with two b-jets (like WH or s-channel single top), we don't know which jet came from the Higgs boson or top quark decay so we have to calculate the probability for both possible jet-parton assignments. The total probability is summed over both combinations. We also try both combinations when there are two tags.

The $t\bar{t}$ matrix element is a special case, because its final state is not the same as WH and candidate events only come into the WH analysis when final state particles

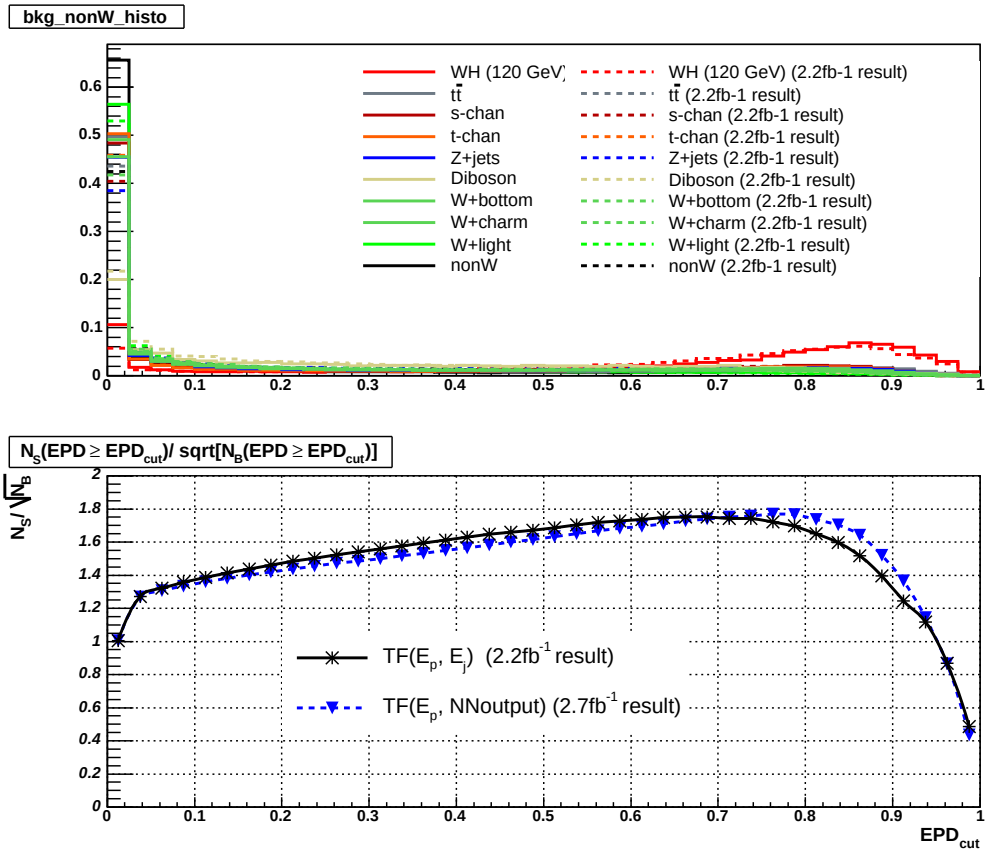

Figure 14: Top: Comparison of the discriminant for signal and some backgrounds using the new TF with NN (dotted lines) and the previous TF with E_j (solid lines). Bottom: Significance ($S/\sqrt{B}$) as a function of the cut in the discriminant using the new (green) and old (black) TFs.

are lost. In this analysis, we assume that one final-state W boson is undetected while the other W boson decays leptonically. Then we integrate over all three components of the momentum of the missing particle.

3.5 Numerical Integration

We perform the three-dimensional integrations with a C++ version of the CERN library function DADMUL [16]. This algorithm is a fully deterministic adaptive quadrature technique that works well for up to three integrations. However, it becomes prohibitively slow for more than three integrals, so we moved to a Monte Carlo integration technique. We use the CERN library algorithm DIVONNE as implemented and improved in the CUBA integration package [17]. This is a Monte-Carlo-based integration based on stratified sampling and aided by methods from numerical optimization. It gives consistent answers with adaptive quadrature for three-integral matrix elements

and gives answers that are self-consistent to at least five significant figures.

We ask each algorithm to converge to 1% estimated error in ten million function calls. Calculating all probabilities (seven for the two-jet bin and four for the three-jet bin) takes from five to ten minutes per event, most of which time is used by the $t\bar{t}$ matrix element calculation.

3.6 Event Probability Discriminant

The event probability density makes use of all measured quantities[3] to specify each event. This should provide good discrimination between signal and background. It uses both possible jet combinations in the event so that the right jet-parton association is always included.

We use the event probability densities as ingredients to build an event probability discriminant, i.e. a distribution which separates signal from background which we can use to fit the data. Perhaps the most intuitive discriminant is the ratio of signal over signal + background probability $EPD = P_s/(P_s + P_b)$. This discriminant is close to zero for ratios dominated by P_b and close to unity for ratios dominated by P_s. This is the discriminant we will use in this analysis.

3.6.1 Including Secondary Vertex Information

Several of the sizable backgrounds in the WH analysis listed in Table 1 don't actually have a b-quark in the final state, but are falsely identified as such. This happens either because a light quark jet is falsely identified to have a displaced secondary vertex (mistags) due to tracking resolution or because real heavy charm quark decays happen to have a sufficiently long life-time to be tagged by the secondary vertex tagger. Therefore, it would be useful to have better separation of b-quark jets from charm or light quark jets. The matrix element based event-probabilities use all event kinematic information to characterize signal and background but can not distinguish b quark jets from charm or light quark flavor jets. On the other hand, the Karlsruhe Neural Net b-tagger uses secondary vertex information, like the secondary vertex mass, the number of tracks, the decay length of the b quark and 22 more variables to distinguish b-quark jets from charm or light quark flavor jets. [13].

Figure 15 shows the neural network b-tagger output for b jets and non-b jets background (left). Both distributions are very well separated. In this analysis we can include the neural network b-jet information by using the network output as a b-jet probability in our discriminant. This is motivated by the right plot of Fig. 15 which shows the network output versus b-jet purity. We define the b-jet probability as $b = 0.5 \cdot (network\ output + 1)$. Since WH events always feature at least one b quark

[3]The only exception is the unclustered energy, the energy deposited in the calorimeter from all sources except from leptons and jets.

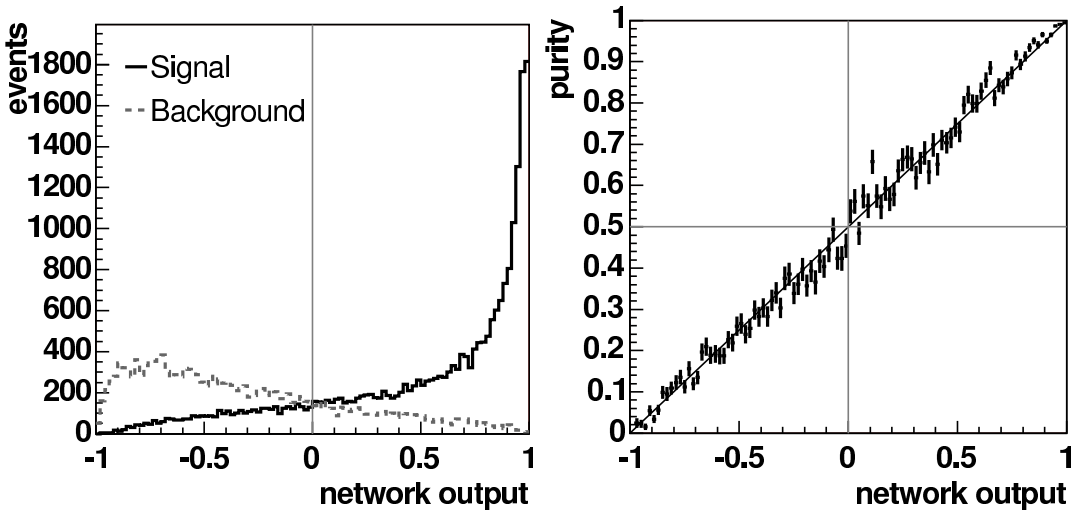

Figure 15: Neural network b-tagger output for b jets and non b jets background (left). Neural Network b-tag output vs purity (right) [13].

in the final state we re-write the event probability discriminant as:

$$EPD = \frac{b \cdot P_{WH}}{b\,(P_{WH} + P_{singletop} + P_{Wb\bar{b}} + P_{t\bar{t}}) + (1-b)\,(P_{Wc\bar{c}} + P_{Wcj} + P_{Mistag} + P_{diboson})} \tag{11}$$

In addition, each probability can receive a normalization constant to adjust the relative weight for that probability. We tune these constants in MC to achieve the greatest expected sensitivity, while making sure every bin is sufficiently populated in Monte Carlo statistics. We also create separate discriminants for the single-tag and double-tag bins. This gives us the ability to tune the discriminants differently in the two tag bins.

There is a small additional gain from weighting the two possible t-channel diagrams separately. One case has an u quark and a b quark in the initial state while the other has a $\bar{u}$ quark and a b. The matrix elements are identical, but the PDFs are different, so we gain sensitivity from treating them separately. We also gain some sensitivity from the Wg + jet matrix element for mistags, which is identical to the Wc + jet matrix element with different PDFs, transfer functions, and quark masses.

3.6.2 Multiple Fit Regions

To increase the sensitivity of the analysis, we split our data sample into different 'analysis channels'. We currently separate the data for single and double tagged events, as well as into old (CEM,PHX,CMUP,CMX) and new leptons (these are all un-triggered muons, CMU, CMP, BMU, CMIO, SCMIO, CMXNT). We fit for one common cross section across in all four channels.

3.6.3 Distributions of the Event Probability Discriminant

In principle, we could define separate discriminants for each Higgs mass point. However, we use the same functional form for all discriminants and only replace the calculation of the WH event probability density corresponding to the proper Higgs mass.

As can be seen from Figures 16 and 17 the separation between signal and background is better for the double-tag category compared to the single-tag one.

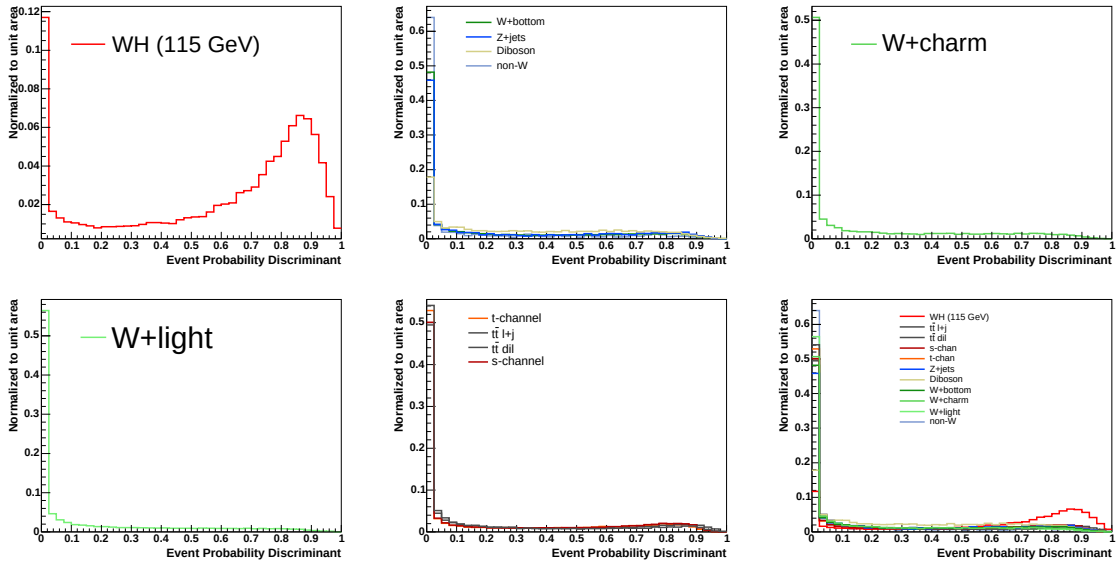

Figure 16: Template histograms used in the WH search for all processes in the single b-tag channel. All histograms are normalized to unit area The bottom right plot shows all the contributions.

As a cross-check we compare the predicted and observed EPD distributions in the untagged sample as shown in Fig. 18 for both, the single and double tag EPDs. We also compare the single and double tag EPDs in the single and double tag samples, respectively. The agreement is good in all cases.

4 BDT Methodology

In order to search for a WH production we develope a multivariate technique based on Boosted Decision Trees (BDT). To build the BDTs we make use of the ROOT-integrated package TMVA [25]. This technique has already been applied at CDF for the sinlge top search as described in [26]. The basics of a BDT is described in the following section.

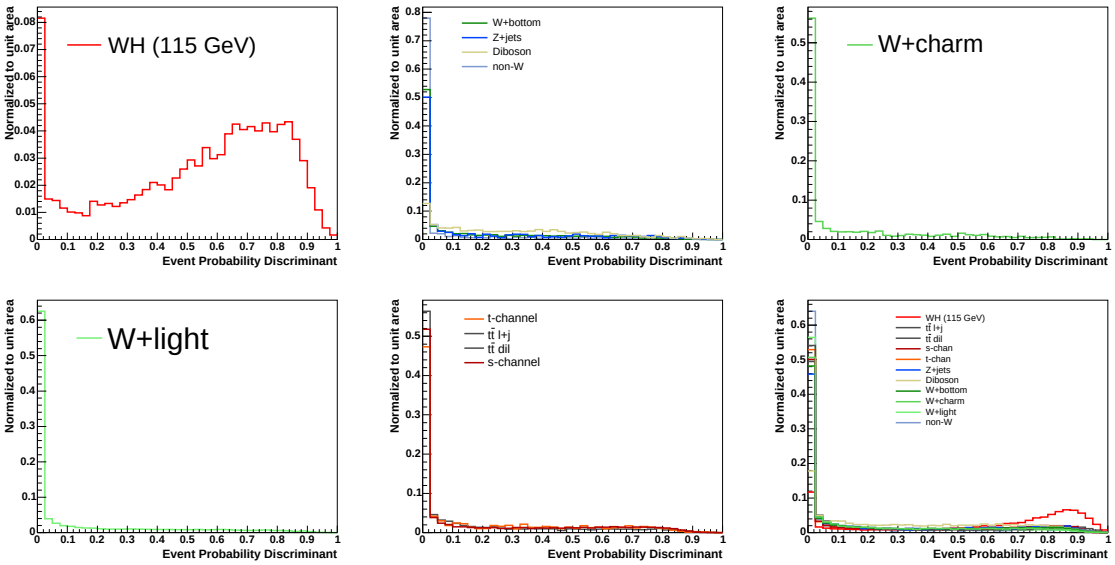

Figure 17: Template histograms used in the WH search for all processes in the double b-tag channel. All histograms are normalized to unit area. The bottom right plot shows all the contributions.

4.1 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 19. 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

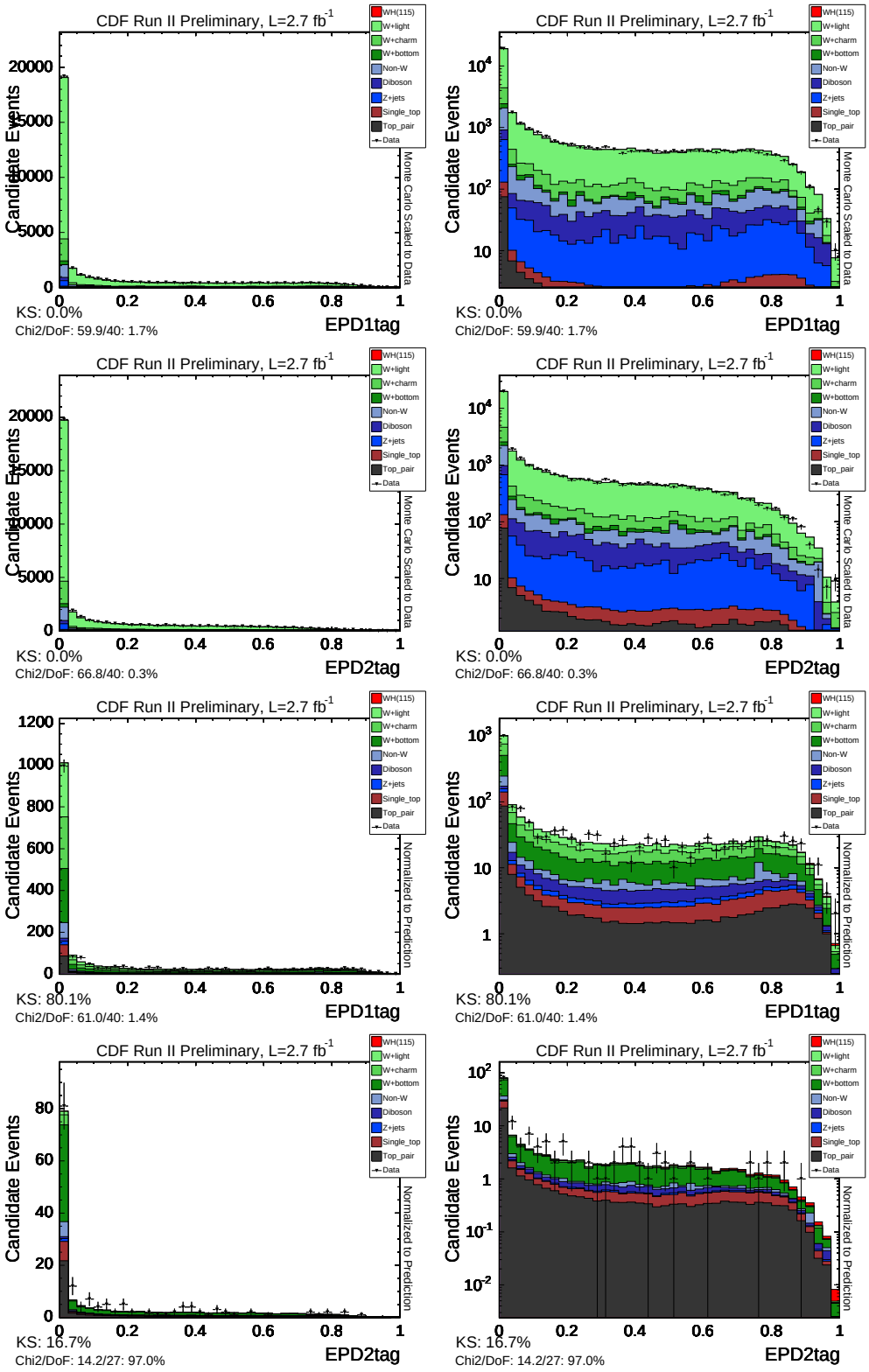

Figure 18: EPD distributions ($m_H = 115\ GeV/c^2$). From top to bottom: single tag and double tag EPDs applied in the untagged sample, single tag EPD applied in the single tag sample and double tag EPD applied in the double tag sample.

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.

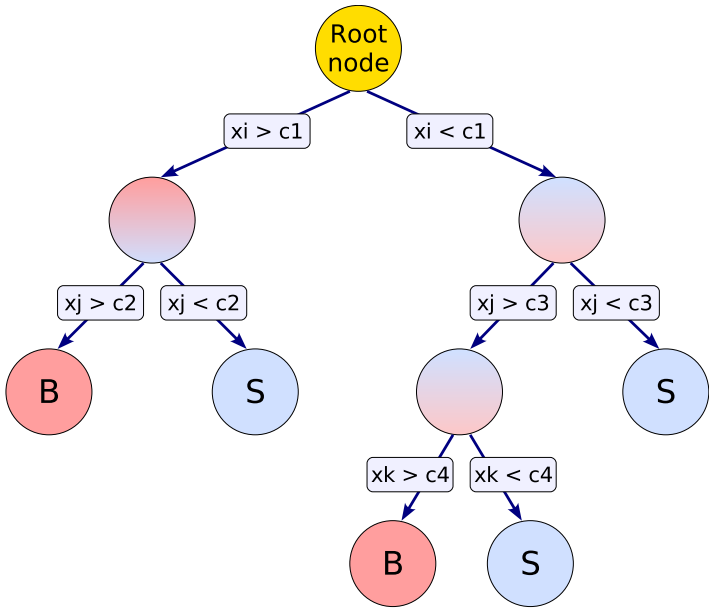

Figure 19: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is performed. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus 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.

4.2 Training of the BDTs

Two different BDTs optimized for the WH search in two different signal regions are trained:

- 2 jet bin, 1 tag
- 2 jet 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.

4.2.1 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 [4]. 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.

4.2.2 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* [27] (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 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 missclassified events over total events) of the previous tree,

$$\alpha = \frac{1 - \text{err}}{\text{err}}. \tag{12}$$

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

$$y_{BDT}(x) = \sum_{i \in forest} \ln(\alpha_i) \cdot h_i(x), \tag{13}$$

[4] 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*

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. 13 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.

4.2.3 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 [28] 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 to 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

$$\rho = \frac{R(\text{node}) - R(\text{subtree below node})}{\#nodes(\text{subtree below node}) - 1} \tag{14}$$

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 analysis until overtraining is completely avoided[5]. Figure 20 shows a check that the BDT response for the test sample is similar in shape to the traingin sample. This is a sign that the BDT is not overtrained.

4.2.4 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. TMVA has the ability of introducing event-by-event weights into the training sample, in this way we can use as much statistics as we have available[6] 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

[5] 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.

[6] At least half of the events in each sample are not used in the trainig but instead are kept for a subsequent test of the training.

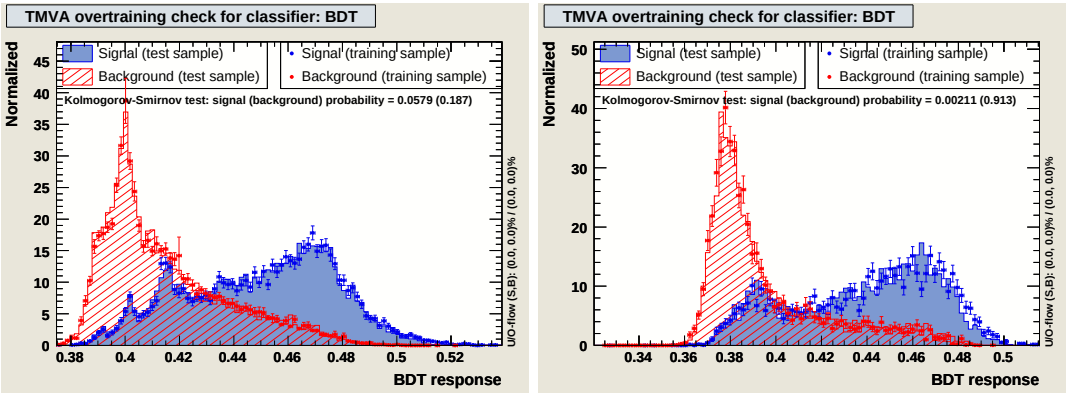

Figure 20: Example of a BDT trained in the 2-jets and 1-tag sample with enough PruneStrength to avoid overtraining. Blue are signal events and red are background events. Points are events from the training sample, and histogram are events from test sample. 1tag is shown on the left and 2tag is shown on the right.

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 s-channel single top, $t\bar{t}$ and $W+bottom$ for the double tag channels, plus $W+charm$, t-channel single top and $W+light$ for the single 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.

4.2.5 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 21 variables used for the training of the BDTs in the 2 jet bin channels are:

- the event probability discriminant based on matrix element probablities
- the ratios between the signal event probability and each one of the background event probabilities
- the invariant mass of the di-jet system, m_{j1j2}
- the E_T of both jets

- the $\Delta\phi$ between the jets and the $\not\!\!E_T$
- the $\Delta\phi$ between the the lepton and the $\not\!\!E_T$
- the p_T and the η of the lepton
- the scalar sum of the transverse energies, $H_T = \sum_{jets} E_T + p_T + \not\!\!E_T$
- the cosine of the angle between the lepton and jets
- the transverse mass of the W boson, $m_T(W)$
- the KIT NN flavor separator [13, 29]
- the missing transverse energy in the event, $\not\!\!E_T$

All the energies involved are level-5 corrected. Validation plots of all the above variables in the two signal regions, as well as in the control regions are shown in appendix F.

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. Table 4 shows the variable ranking for the trained channel.

4.3 Distributions of the BDT Outputs

We use the output of the BDT trained in the two channels as the discriminant for a WH search. The raw output of the BDTs are always in a range inside [-1, 1]; however, 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 background events in order not to make MCLIMIT code confused by the absence of background.

Figure 21 includes the final templates, for the Higgs mass of 115 GeV/c^2 (the templates for the rest of the masses are shown in Appendix G), we use in the cross section fit and to throw pseudo-experiments. For display purposes we group templates with similar shapes in categories, so W+jets-like is the composition of $Wb\bar{b}$, $Wc\bar{c}$, Wcj, $Z + jets$ and non-W weighted by *Method 2* predictions; top-like is composed by $t\bar{t}$ and single top, diboson is WZ, ZZ and WW (in the double tag templates, WW is included in W+jets-like).

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

	2-jets, 1-tag		2-jets, 2-tag	
Rank	Variable	Importance	Variable	Importance
1	EPD	2.428e-01	EPD	9.952e-02
2	Ht	8.334e-02	log(ProbWH/Wc)	9.199e-02
3	J2Et	6.797e-02	log(ProbWH/tchan)	8.354e-02
4	log(ProbWH/Wjg)	5.456e-02	log(ProbWH/ttbar)	7.425e-02
5	log(ProbWH/schan)	5.202e-02	log(ProbWH/schan)	7.367e-02
6	log(ProbWH/Wc)	4.960e-02	log(ProbWH/tchan2)	6.198e-02
7	met	4.872e-02	log(ProbWH/Wgg)	5.751e-02
8	J1Et	4.373e-02	log(ProbWH/Wjg)	4.353e-02
9	log(ProbWH//WccWbb)	4.006e-02	MetLepDPhi	4.174e-02
10	mJ1J2	3.932e-02	mJ1J2	4.145e-02
11	LepPt	3.901e-02	LepEta	4.134e-02
12	log(ProbWH/tchan)	3.723e-02	J1Et	3.968e-02
13	log(ProbWH/ttbar)	3.461e-02	Ht	3.460e-02
14	MetLepDPhi	2.893e-02	KaNN	3.358e-02
15	log(ProbWH/tchan2)	2.842e-02	MetJ1DPhi	3.062e-02
16	log(ProbWH/Wgg)	2.726e-02	J2Et	2.821e-02
17	KaNN	2.715e-02	log(ProbWH//WccWbb)	2.664e-02
18	wmt	1.785e-02	met	2.656e-02
19	LepEta	1.284e-02	cosLepJ1	2.587e-02
20	cosLepJ1	1.246e-02	LepPt	2.431e-02
21	MetJ1DPhi	1.209e-02	wmt	1.939e-02

Table 4: Variable ranking result for both 2-jets channels for $M_H = 115$. Top variable is best ranked.

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 the ratio between the fitting cross-section and the standard model one, $\beta_j = \sigma_j^{Fit}/\sigma_j^{SM}$, 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:

$$\mathcal{L} = \prod_{j=2}^{N_b+1} G_j(\beta_j; \Delta_j) \prod_{k=1}^{B} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \tag{15}$$

where N_b is the number of background channels, and Δ_j is the uncertainty on the fraction β_j.

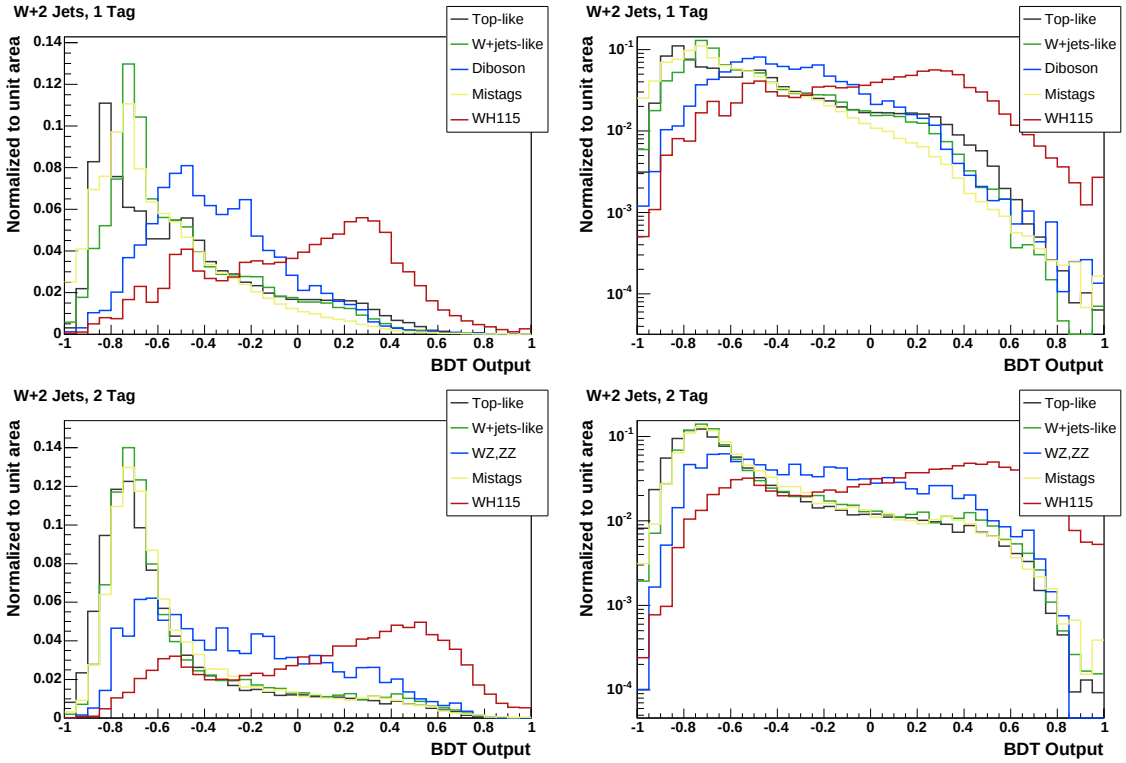

Figure 21: Templates of the BDT outputs in the two-jet bin for the $m_H = 115\ GeV/c^2$ training. Single tag discriminants are on top, while double tag discriminants are on bottom. All histograms are normalized to unit area. Linear (log) scale in the left (right) side.

The Gaussian constraints to the backgrounds are given by:

$$G_j(\beta_j; \Delta_j) = \frac{1}{\sqrt{2\,\pi \cdot \Delta_j^2}} \exp\left[-\frac{1}{2} \cdot \left(\frac{\beta_j - 1.0}{\Delta_j}\right)^2\right] \tag{16}$$

and

$$\mu_k = \beta_1 \cdot T_{1k} + \sum_{j=2}^{N_b+1} (\beta_j \cdot T_{jk}) \tag{17}$$

where T_{1k}, T_{jk} are, respectively, the template entries at bin k of the signal and the background channel j.

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, $\sum_{k=1}^{B} T_{jk} = N_j^{pred}$.

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

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

6 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) luminosity and (6) b tagging SF 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 WH normalization uncertainty is the difference between the systematically shifted acceptance and the default one and are shown in Table 5.

Channel	Lepton ID	Luminosity	$b-$ tagging SF	ISR/FSR + PDF	JES
single-tag	$\sim 2\%$	6%	3.5%	3.1%	2.0%
double-tag	$\sim 2\%$	6%	8.4%	5.6%	2.0%

Table 5: Rate systematic uncertainties for each channel.

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 [14]. 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 the 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).

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:

$$\mathcal{L}(\beta_1,\dots,\beta_{N_b+1};\delta_1,\dots,\delta_{N_s}) = \underbrace{\prod_{k=1}^{B}\frac{e^{-\mu_k}\cdot\mu_k^{n_k}}{n_k!}}_{Poisson\ term}\cdot\underbrace{\prod_{j=2}^{N_b+1}G(\beta_j|1,\Delta_j)}_{Gauss\ constraints}\cdot\underbrace{\prod_{i=1}^{N_s}G(\delta_i,0,1)}_{Systematics} \qquad (18)$$

where

$$\mu_k = \sum_{j=1}^{N_b+1} \beta_j \cdot \underbrace{\left\{ \prod_{i=1}^{N_s} [1 + |\delta_i| \cdot (\epsilon_{ji+} H(\delta_i) + \epsilon_{ji-} H(-\delta_i))] \right\}}_{Normalization\ Uncertainty} \tag{19}$$

$$\underbrace{\cdot\ T_{jk}}_{Shape\ P.} \cdot \underbrace{\left\{ \prod_{i=1}^{N_s} (1 + |\delta_i| \cdot (\kappa_{jik+} H(\delta_i) + \kappa_{jik-} H(-\delta_i))) \right\}}_{Shape\ Uncertainty} \tag{20}$$

and N_s is the number of systematic effects included.

All systematic normalization and shape uncertainties are incorporated in the analysis into the likelihood as nuisance parameters, conform with a fully Bayesian treatment [18]. We take the correlation between normalization and shape uncertainties for a given source into account [19]. 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 18). The $\pm 1\sigma$ 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 19). The $\pm 1\sigma$ changes in bin k of the templates for process j due to the i^{th} source of systematic uncertainty are quantified by κ_{jik+} and κ_{jik-} (see Equation part 20). $H(\delta_i)$ 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 19) and the shape uncertainty (Equation 20), which is how correlations between both effects are taken into account.

We marginalize the likelihood function by integrating $\mathcal{L}(\beta_1, ..., \beta_{N_b+1}, \delta_1, ..\delta_{N_s})$ over all nuisance parameters for many possible values of the WH cross-section $\beta_1 = \beta_{WH}$. The resulting reduced likelihood $\mathcal{L}(\beta_{WH})$ is a function of the WH cross-section β_{WH} only. We use the MCLIMIT package for our statistical treatment [23].

The event detection efficiency includes uncertainties on the lepton ID, trigger efficiencies and b-tagging scale-factors. The uncertainties on the data derived backgrounds (W+bottom, W+charm, Mistags and non-W) are taken from the event yield in Table 1.

7 Result with CDF II Data

We apply the analysis to 2.7 fb^{-1} of CDF Run II data. We compare the BDT output distribution, for a Higgs mass of 115 GeV/c^2, of our candidate events with the sum of predicted WH signal and background distributions as shown in Figure 22. The BDT outputs, in the untagged and tag samples, for the rest of Higgs masses can be found in Appendix G.

In order to extract the most probable WH signal content in the data we perform the maximum likelihood method described in Section 5. We perform marginalization

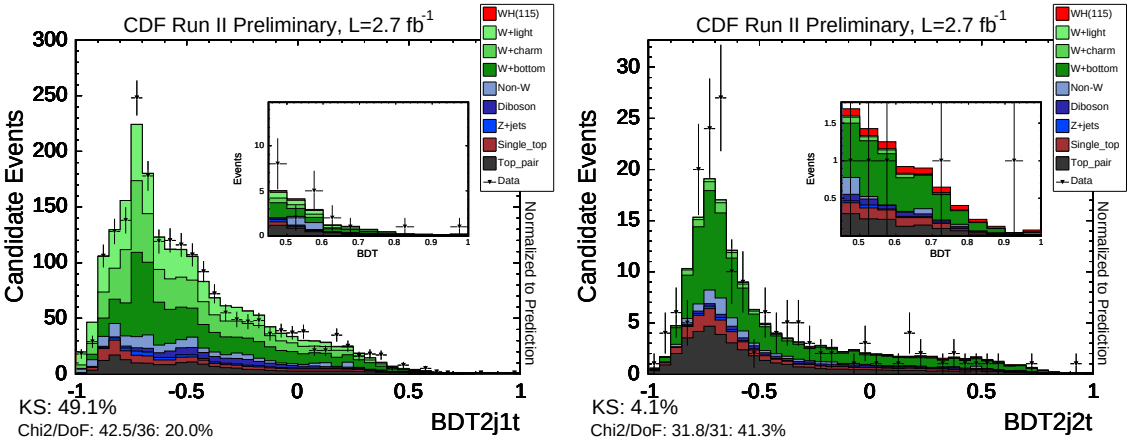

Figure 22: Comparison of the BDT output for lepton + 2 jets data compared to the Monte Carlo prediction for WH ($m_H = 115\ GeV/c^2$) signal and background. The left-hand plot shows the single SecVtx tagged events and the right-hand plot shows the double SecVtx tagged events.

using the likelihood function of Equation 18 with all systematic uncertainties included in the likelihood function. The posterior p.d.f is obtained by using Bayes' theorem:

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

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

To set an upper limit on the WH production cross-section, we integrate the posterior probability density to cover 95% [7]. The observed and expected results are shown in Table 7 and in Figure 23.

σ	100	105	110	115	120	125	130	135	140	145	150
Expected	0.94	0.86	0.82	0.77	0.73	0.69	0.67	0.64	0.62	0.59	0.63
Observed	0.82	0.84	1.13	0.78	0.83	0.69	0.74	0.92	0.78	1.09	1.29

Table 6: Expected and observed upper limit cross sections, in pb, for different Higgs mass points.

σ / SM	100	105	110	115	120	125	130	135	140	145	150
Expected	4.07	4.28	4.82	5.64	7.02	8.32	10.7	14.2	20.8	29.4	52.1
Observed	3.52	4.19	6.67	5.75	7.98	8.36	11.7	20.5	26.1	54.5	107.6

Table 7: Expected and observed upper limit cross sections, in SM units, for different Higgs mass points.

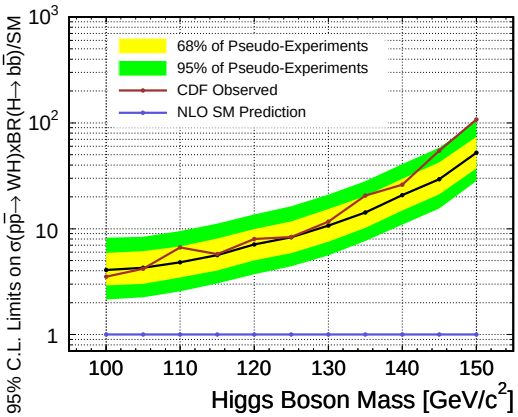

Figure 23: 95 % C.L. upper limits on the WH production cross-sections times branching ratio for $H \to b\bar{b}$ for Higgs boson masses between $m_H = 100$ GeV/c^2 to $m_H = 150$ GeV/c^2. The plot shows the limit normalized to the predictions from the Standard Model.

8 Conclusions

We have used the matrix element analysis technique in a direct search for Higgs boson production in association with a W boson. To extract the most probable WH 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. We observe no evidence for a Higgs boson signal and set 95% confidence level upper limits on the WH production cross section times the branching ratio, in SM units, of the Higgs boson to decay to $b\bar{b}$ pairs of $\sigma(p\bar{p} \to WH) \times BR(H \to b\bar{b})/SM < 3.52$ to 107.6 for Higgs boson masses between $m_H = 100$ GeV/c^2 and $m_H = 150$ GeV/c^2. The expected (median) sensitivity estimated in pseudo experiments is $\sigma(p\bar{p} \to WH) \times BR(H \to b\bar{b})/SM < 4.07$ to 52.1 at 95% C.L.

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 für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Ba-

sic Research; the Ministerio de Educación y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland. We would like to thank Tom Junk (FNAL) for many extremely useful discussions.

References

[1] J. Adelman *et al.*, *Method II For You*, CDF Note 9185.

[2] C. Cully *et al.*, *Calibration of Heavy-Flavor Production in W +1Jet Data*, CDF Note 9187.

[3] P. Dong, B. Stelzer, R. Wallny, F. Canelli, *Search for Single Top Quark Production using the Matrix Element Technique*, CDF Note 9117 (2008).

[4] C. Group, A. Ruiz, B. Casal, R. Vilar, P. Dong, B. Stelzer F. Canelli, *Search for Single Top Quark Production using Boosted Decision Trees*, CDF Note 9117 (2008).

[5] P. Dong *et al.*, *Search for Standard Model Higgs Boson Production in Association with a W Boson using the Matrix Element Technique and 2.2 fb*, CDF Note 9263 (2008).

[6] A. Buzatu *et al.*, *Search for Standard Model Higgs Boson Production in Association with a W Boson using 2.7 fb*, CDF Note 9401 (2008).

[7] Particle Data Group, *The Review of Particle Physics*, http://pdg.lbl.gov (2006).

[8] F. Canelli, B. Casal Larana, C. Group, A. Ruiz, B. Stelzer, R. Vilar, *Increasing Muon Acceptance with MET Plus Jet Triggers*, CDF Note 9105 (2008).

[9] T. Peiffer, *Search for the Higgs Boson in the WH channel with the CDF II Experiment*, FERMILAB Master Thesis, 2008-1, Document Server `http://lss.fnal.gov/archive/masters/fermilab-masters-2008-01.shtml`, (2008).

[10] E. Murayama, I. Watanabe and K. Hagiwara, *HELAS: HELicity Amplitude Subroutines for Feynman Diagram Evaluations*, KEK Report 91-11 (1992).

[11] T. Stelzer and W. F. Long, *HELAS: Automatic Generation of Tree Level Helicity Amplitudes*, Phys. Commun. **81**, 357-371 (1994).

[12] Dmitri Smirnov, Konstantin Anikeev, *ROOTSNNS v3.0, Root Interface to SNNS*, CDF Note 7999 (2006).

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

[14] 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).

[15] F. James, *MINUIT, Function Minimization and Error Analysis Reference Manual*, CERN Program Library Long Writeup, D506 (1994).

[16] A.C. Genz and A.A. Malik, *Remarks on algorithm 006: An adaptive algorithm for numerical integration over an N-dimensional rectangular region*, J. Comput. Appl. Math. 6 (1980) 295-302.
A. van Doren and L. de Ridder, *An adaptive algorithm for numerical integration over an n-dimensional cube*, J.Comput. Appl. Math. 2 (1976) 207-217.

[17] T. Hahn, *CUBA—A library for multidimensional numberical integration*, Comput. Phys. Commun. 168 (2005) 78.

[18] 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.

[19] 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).

[20] 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).

[21] 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).

[22] Guimaraes da Costa, Sherman, *Study of SecVtx Fake Rates with 1/fb*, CDF Note 8263 (2006). *Loose and Tight SecVtx Tag Matrices with 1/fb*, CDF Note 8264 (2006).

[23] 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).

[24] B.W. Harris *et. al.*, *Phys. Rev.* D **66**, 054024 (2002)˙ Z. Sullivan *Phys. Rev.* D **70**, 114012 (2004).

[25] 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).

[26] F Canelli *et al.*, *Search for Single Top Quark Production using Boosted Decision Trees in 2.2fb*$^{-1}$, CDF Note 9263 (2008).

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

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

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

[30] G. Gomez *et al.*, *Measurement of the ttbar Production Cross Section in the Jet Probability Tagged Sample with Gen5*, CDF Note 7697 (2006).

[31] K. Lannon *et al.*, *Combined Search for Higgs Produced in Association with a W Boson with 2.7 fb*$^{-1}$, CDF Note 9583 (2008).

Figure 25: Non-W fits in the pretag sample (QCD veto is not applied)

Figure 27: Non-W fits in the double tag region (QCD veto is not applied).

B Event Yields

Tables 8 and 9 show the expected number of signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements, in the single and double tag case, respectively. Tables 10-14 (15-19) show the event yields in the single (double) tag category for each subdetector type.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	293943.0 ± 0.0	50644.0 ± 0.0	8895.0 ± 0.0	2003.0 ± 0.0	418.0 ± 0.0
Tagged WW	19.8 ± 2.2	56.2 ± 6.2	17.8 ± 1.9	4.5 ± 0.5	1.0 ± 0.1
Tagged WZ	11.1 ± 0.9	23.0 ± 1.7	6.3 ± 0.5	1.4 ± 0.1	0.3 ± 0.0
Tagged ZZ	0.2 ± 0.0	0.8 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged TopLJ	11.8 ± 1.7	121.3 ± 17.1	318.9 ± 44.5	329.8 ± 45.7	90.0 ± 12.5
Tagged TopDil	8.5 ± 1.2	48.8 ± 6.8	35.4 ± 4.9	8.0 ± 1.1	1.3 ± 0.2
Tagged STopT	44.1 ± 6.5	64.0 ± 9.3	11.8 ± 1.7	1.8 ± 0.3	0.2 ± 0.0
Tagged STopS	12.0 ± 1.7	40.6 ± 5.7	12.4 ± 1.7	2.6 ± 0.4	0.5 ± 0.1
Tagged Z+jets	41.9 ± 6.6	37.4 ± 5.5	13.9 ± 2.0	3.5 ± 0.5	0.7 ± 0.1
Tagged Wbb	825.1 ± 250.1	538.7 ± 162.5	148.0 ± 44.8	34.0 ± 10.4	7.9 ± 2.5
Tagged Wcc/Wc	1187.7 ± 370.2	489.1 ± 150.9	122.2 ± 37.7	28.1 ± 8.8	6.4 ± 2.0
Tagged Total HF	2012.8 ± 618.8	1027.8 ± 312.3	270.3 ± 82.1	62.1 ± 19.0	14.3 ± 4.4
Tagged Total MC	149.4 ± 15.3	392.0 ± 35.0	416.8 ± 48.4	351.6 ± 46.4	93.9 ± 12.6
Tagged Mistags	1148.8 ± 150.7	458.0 ± 57.9	127.0 ± 16.7	27.9 ± 4.4	6.1 ± 1.2
Tagged Non-W	122.9 ± 49.1	135.5 ± 54.2	51.5 ± 20.6	6.3 ± 2.5	5.8 ± 2.3
Total Prediction	3433.9 ± 639.0	2013.3 ± 324.1	865.6 ± 98.9	447.9 ± 50.4	120.2 ± 13.6
Tagged WH100	3.7 ± 0.3	9.5 ± 0.8	2.3 ± 0.2	0.4 ± 0.0	0.1 ± 0.0
Tagged WH105	3.0 ± 0.3	8.6 ± 0.7	2.2 ± 0.2	0.4 ± 0.0	0.1 ± 0.0
Tagged WH110	2.4 ± 0.2	7.6 ± 0.6	2.0 ± 0.2	0.3 ± 0.0	0.0 ± 0.0
Tagged WH115	1.8 ± 0.2	6.3 ± 0.5	1.7 ± 0.1	0.3 ± 0.0	0.0 ± 0.0
Tagged WH120	1.3 ± 0.1	4.9 ± 0.4	1.4 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged WH125	1.0 ± 0.1	4.0 ± 0.3	1.2 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged WH130	0.7 ± 0.1	3.1 ± 0.3	1.0 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged WH135	0.4 ± 0.0	2.3 ± 0.2	0.7 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH140	0.3 ± 0.0	1.5 ± 0.1	0.5 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH145	0.2 ± 0.0	1.0 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH150	0.1 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	3789.0 ± 0.0	1998.0 ± 0.0	826.0 ± 0.0	479.0 ± 0.0	131.0 ± 0.0

Table 8: Number of expected single tagged signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	328564.0 ± 0.0	57174.0 ± 0.0	10218.0 ± 0.0	2282.0 ± 0.0	472.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.4 ± 0.1	0.5 ± 0.1	0.3 ± 0.0	0.1 ± 0.0
Tagged WZ	0.0 ± 0.0	4.8 ± 0.5	1.3 ± 0.2	0.3 ± 0.0	0.1 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	23.8 ± 3.9	93.4 ± 15.3	129.2 ± 20.9	37.4 ± 6.0
Tagged TopDil	0.0 ± 0.0	14.1 ± 2.3	12.5 ± 2.0	3.0 ± 0.5	0.6 ± 0.1
Tagged STopT	0.0 ± 0.0	1.8 ± 0.3	1.4 ± 0.2	0.4 ± 0.1	0.1 ± 0.0
Tagged STopS	0.0 ± 0.0	12.8 ± 2.1	4.2 ± 0.7	0.9 ± 0.2	0.2 ± 0.0
Tagged Z+jets	0.0 ± 0.0	2.1 ± 0.3	1.4 ± 0.2	0.4 ± 0.1	0.1 ± 0.0
Tagged Wbb	0.0 ± 0.0	70.3 ± 22.5	24.1 ± 7.8	6.6 ± 2.2	1.5 ± 0.5
Tagged Wcc/Wc	0.0 ± 0.0	6.8 ± 2.3	3.3 ± 1.1	1.3 ± 0.4	0.3 ± 0.1
Tagged Total HF	0.0 ± 0.0	77.1 ± 24.7	27.4 ± 8.9	7.8 ± 2.6	1.9 ± 0.6
Tagged Total MC	0.0 ± 0.0	59.9 ± 7.5	114.8 ± 17.0	134.5 ± 21.3	38.5 ± 6.1
Tagged Mistags	0.0 ± 0.0	2.2 ± 0.6	1.5 ± 0.4	0.5 ± 0.2	0.2 ± 0.1
Tagged Non-W	0.0 ± 0.0	9.0 ± 3.6	4.8 ± 1.9	4.7 ± 1.9	0.7 ± 0.3
Total Prediction	0.0 ± 0.0	148.2 ± 26.1	148.5 ± 19.3	147.5 ± 21.5	41.2 ± 6.2
Tagged WH100	0.0 ± 0.0	2.9 ± 0.3	0.8 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	2.7 ± 0.3	0.7 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	2.4 ± 0.3	0.7 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	2.0 ± 0.2	0.6 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	1.6 ± 0.2	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	1.3 ± 0.2	0.4 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	1.0 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.7 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.5 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	156.0 ± 0.0	174.0 ± 0.0	144.0 ± 0.0	46.0 ± 0.0

Table 9: Number of expected double tagged signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	133126.0 ± 0.0	21622.0 ± 0.0	3608.0 ± 0.0	805.0 ± 0.0	167.0 ± 0.0
Tagged WW	9.3 ± 1.0	22.9 ± 2.5	7.1 ± 0.8	1.8 ± 0.2	0.3 ± 0.0
Tagged WZ	4.9 ± 0.4	8.8 ± 0.7	2.2 ± 0.2	0.5 ± 0.0	0.1 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	5.9 ± 0.9	47.2 ± 6.6	120.9 ± 16.9	124.3 ± 17.2	34.3 ± 4.7
Tagged TopDil	4.0 ± 0.6	17.8 ± 2.5	12.4 ± 1.7	2.9 ± 0.4	0.4 ± 0.1
Tagged STopT	21.0 ± 3.1	26.0 ± 3.8	4.6 ± 0.7	0.7 ± 0.1	0.1 ± 0.0
Tagged STopS	5.9 ± 0.8	16.3 ± 2.3	4.9 ± 0.7	1.0 ± 0.1	0.2 ± 0.0
Tagged Z+jets	5.4 ± 0.8	7.5 ± 1.1	3.4 ± 0.5	1.1 ± 0.2	0.2 ± 0.0
Tagged Wbb	387.9 ± 117.7	223.5 ± 67.9	58.5 ± 18.1	14.0 ± 4.5	3.3 ± 1.1
Tagged Wcc/Wc	549.0 ± 171.3	205.5 ± 63.9	47.7 ± 15.0	11.5 ± 3.8	2.7 ± 0.9
Tagged Total HF	936.9 ± 288.1	429.1 ± 130.8	106.2 ± 32.6	25.4 ± 8.0	5.9 ± 1.9
Tagged Total MC	56.4 ± 5.5	146.8 ± 13.0	155.5 ± 18.2	132.3 ± 17.5	35.6 ± 4.8
Tagged Mistags	520.6 ± 69.3	187.2 ± 26.3	51.3 ± 8.8	10.9 ± 2.6	2.5 ± 0.8
Tagged Non-W	78.6 ± 31.4	83.2 ± 33.3	35.2 ± 14.1	1.0 ± 0.4	0.0 ± 0.0
Total Prediction	1592.5 ± 298.1	846.2 ± 138.2	348.1 ± 40.9	169.6 ± 19.4	44.1 ± 5.2
Tagged WH100	1.8 ± 0.2	3.9 ± 0.3	0.9 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged WH105	1.5 ± 0.1	3.5 ± 0.3	0.9 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged WH110	1.2 ± 0.1	3.1 ± 0.3	0.8 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH115	0.9 ± 0.1	2.6 ± 0.2	0.7 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged WH120	0.6 ± 0.1	2.0 ± 0.2	0.6 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH125	0.5 ± 0.0	1.6 ± 0.1	0.5 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH130	0.3 ± 0.0	1.3 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH135	0.2 ± 0.0	0.9 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH140	0.1 ± 0.0	0.6 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.1 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	1658.0 ± 0.0	843.0 ± 0.0	341.0 ± 0.0	175.0 ± 0.0	46.0 ± 0.0

Table 10: Number of expected single tagged signal and background events with a CEM electron in 2.7 fb^{-1} of CDF data, passing all event selection requirements CEM.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	56369.0 ± 0.0	8415.0 ± 0.0	1446.0 ± 0.0	240.0 ± 0.0	51.0 ± 0.0
Tagged WW	3.4 ± 0.4	7.6 ± 0.8	2.1 ± 0.2	0.6 ± 0.1	0.1 ± 0.0
Tagged WZ	2.3 ± 0.2	3.7 ± 0.3	1.0 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	1.2 ± 0.2	10.5 ± 1.5	25.8 ± 3.6	26.4 ± 3.7	7.0 ± 1.0
Tagged TopDil	1.0 ± 0.1	4.4 ± 0.6	2.9 ± 0.4	0.5 ± 0.1	0.1 ± 0.0
Tagged STopT	6.5 ± 1.0	6.0 ± 0.9	1.1 ± 0.2	0.2 ± 0.0	0.0 ± 0.0
Tagged STopS	1.5 ± 0.2	3.4 ± 0.5	1.0 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged Z+jets	2.2 ± 0.3	1.6 ± 0.2	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged Wbb	138.6 ± 42.0	82.7 ± 25.1	23.8 ± 7.3	4.4 ± 1.4	1.1 ± 0.4
Tagged Wcc/Wc	201.8 ± 62.9	73.6 ± 22.9	19.6 ± 6.1	3.6 ± 1.2	0.9 ± 0.3
Tagged Total HF	340.4 ± 104.7	156.3 ± 47.7	43.4 ± 13.3	8.0 ± 2.5	2.0 ± 0.6
Tagged Total MC	18.1 ± 1.8	37.4 ± 3.3	34.5 ± 3.9	28.2 ± 3.7	7.4 ± 1.0
Tagged Mistags	223.8 ± 29.8	79.2 ± 10.9	21.4 ± 3.5	4.1 ± 0.9	1.0 ± 0.3
Tagged Non-W	22.1 ± 8.8	22.6 ± 9.0	6.7 ± 2.7	5.3 ± 2.1	3.7 ± 1.5
Total Prediction	604.4 ± 109.2	295.5 ± 49.8	106.0 ± 14.6	45.6 ± 5.0	14.1 ± 1.9
Tagged WH100	0.5 ± 0.0	0.9 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.4 ± 0.0	0.8 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.3 ± 0.0	0.7 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.2 ± 0.0	0.6 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.1 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	691.0 ± 0.0	307.0 ± 0.0	110.0 ± 0.0	45.0 ± 0.0	13.0 ± 0.0

Table 11: Number of expected single tagged signal and background events with a PEM electron in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	61147.0 ± 0.0	10687.0 ± 0.0	1770.0 ± 0.0	418.0 ± 0.0	70.0 ± 0.0
Tagged WW	4.6 ± 0.5	12.2 ± 1.3	3.7 ± 0.4	0.9 ± 0.1	0.2 ± 0.0
Tagged WZ	2.5 ± 0.2	4.6 ± 0.3	1.2 ± 0.1	0.3 ± 0.0	0.1 ± 0.0
Tagged ZZ	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	3.1 ± 0.4	27.2 ± 3.8	72.0 ± 10.1	75.2 ± 10.4	20.6 ± 2.9
Tagged TopDil	2.4 ± 0.3	11.7 ± 1.6	8.5 ± 1.2	1.7 ± 0.2	0.3 ± 0.0
Tagged STopT	10.6 ± 1.6	14.8 ± 2.2	2.8 ± 0.4	0.4 ± 0.1	0.0 ± 0.0
Tagged STopS	3.0 ± 0.4	9.5 ± 1.3	2.9 ± 0.4	0.6 ± 0.1	0.1 ± 0.0
Tagged Z+jets	19.7 ± 3.1	12.4 ± 1.8	3.6 ± 0.5	0.8 ± 0.1	0.2 ± 0.0
Tagged Wbb	178.8 ± 54.2	120.8 ± 36.4	31.1 ± 9.4	7.1 ± 2.3	0.7 ± 0.4
Tagged Wcc/Wc	256.2 ± 79.9	107.7 ± 33.2	26.3 ± 8.1	6.0 ± 2.0	0.6 ± 0.3
Tagged Total HF	435.1 ± 133.7	228.5 ± 69.4	57.3 ± 17.4	13.0 ± 4.1	1.3 ± 0.5
Tagged Total MC	46.0 ± 5.0	92.6 ± 8.3	94.9 ± 11.0	80.0 ± 10.6	21.5 ± 2.9
Tagged Mistags	233.7 ± 30.9	93.5 ± 12.5	24.6 ± 3.6	6.1 ± 1.5	0.5 ± 0.4
Tagged Non-W	10.4 ± 4.2	13.5 ± 5.4	5.1 ± 2.0	0.0 ± 0.0	2.1 ± 0.8
Total Prediction	725.2 ± 137.4	428.2 ± 71.2	181.9 ± 21.0	99.1 ± 11.4	25.4 ± 3.1
Tagged WH100	0.9 ± 0.1	2.3 ± 0.2	0.5 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH105	0.8 ± 0.1	2.0 ± 0.2	0.5 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH110	0.6 ± 0.1	1.8 ± 0.1	0.5 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH115	0.4 ± 0.0	1.5 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH120	0.3 ± 0.0	1.2 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH125	0.2 ± 0.0	0.9 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH130	0.2 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.1 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.1 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	844.0 ± 0.0	431.0 ± 0.0	160.0 ± 0.0	119.0 ± 0.0	26.0 ± 0.0

Table 12: Number of expected single tagged signal and background events with a CMUP muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	43301.0 ± 0.0	6489.0 ± 0.0	1028.0 ± 0.0	217.0 ± 0.0	43.0 ± 0.0
Tagged WW	2.6 ± 0.4	7.0 ± 1.0	1.9 ± 0.3	0.5 ± 0.1	0.1 ± 0.0
Tagged WZ	1.4 ± 0.2	2.8 ± 0.3	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	1.6 ± 0.3	12.6 ± 2.1	32.9 ± 5.4	34.4 ± 5.6	9.3 ± 1.5
Tagged TopDil	1.1 ± 0.2	5.1 ± 0.8	3.6 ± 0.6	0.9 ± 0.2	0.1 ± 0.0
Tagged STopT	5.9 ± 1.0	7.1 ± 1.2	1.3 ± 0.2	0.2 ± 0.0	0.0 ± 0.0
Tagged STopS	1.6 ± 0.3	4.5 ± 0.7	1.3 ± 0.2	0.3 ± 0.0	0.0 ± 0.0
Tagged Z+jets	14.6 ± 2.6	6.9 ± 1.2	2.2 ± 0.4	0.5 ± 0.1	0.1 ± 0.0
Tagged Wbb	119.8 ± 36.3	68.6 ± 20.7	17.9 ± 5.4	4.0 ± 1.3	0.9 ± 0.3
Tagged Wcc/Wc	180.7 ± 56.3	65.0 ± 20.0	14.7 ± 4.5	3.4 ± 1.1	0.7 ± 0.3
Tagged Total HF	300.4 ± 92.4	133.6 ± 40.6	32.6 ± 9.9	7.4 ± 2.3	1.7 ± 0.5
Tagged Total MC	28.9 ± 4.1	46.2 ± 5.7	44.1 ± 6.3	37.0 ± 5.8	9.7 ± 1.5
Tagged Mistags	170.7 ± 22.5	56.9 ± 7.6	15.1 ± 2.2	3.0 ± 0.7	0.7 ± 0.2
Tagged Non-W	11.8 ± 4.7	9.7 ± 3.9	1.6 ± 0.7	0.0 ± 0.0	0.0 ± 0.0
Total Prediction	511.8 ± 95.3	246.3 ± 41.9	93.4 ± 12.0	47.5 ± 6.3	12.0 ± 1.7
Tagged WH100	0.5 ± 0.1	1.1 ± 0.1	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.4 ± 0.1	1.0 ± 0.1	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.3 ± 0.0	0.8 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.2 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.2 ± 0.0	0.6 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.1 ± 0.0	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.1 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.1 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	596.0 ± 0.0	238.0 ± 0.0	81.0 ± 0.0	48.0 ± 0.0	13.0 ± 0.0

Table 13: Number of expected single tagged signal and background events with a CMX muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	0.0 ± 0.0	3431.0 ± 0.0	1043.0 ± 0.0	323.0 ± 0.0	87.0 ± 0.0
Tagged WW	0.0 ± 0.0	6.5 ± 0.7	3.0 ± 0.3	0.7 ± 0.1	0.2 ± 0.0
Tagged WZ	0.0 ± 0.0	3.0 ± 0.2	1.1 ± 0.1	0.2 ± 0.0	0.1 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	23.8 ± 3.4	67.2 ± 9.4	69.4 ± 9.7	18.9 ± 2.6
Tagged TopDil	0.0 ± 0.0	9.7 ± 1.4	8.1 ± 1.1	1.9 ± 0.3	0.3 ± 0.0
Tagged STopT	0.0 ± 0.0	10.1 ± 1.5	2.0 ± 0.3	0.3 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	6.9 ± 1.0	2.3 ± 0.3	0.5 ± 0.1	0.1 ± 0.0
Tagged Z+jets	0.0 ± 0.0	8.9 ± 1.3	4.0 ± 0.6	1.0 ± 0.1	0.2 ± 0.0
Tagged Wbb	0.0 ± 0.0	43.1 ± 13.0	16.7 ± 5.2	4.5 ± 1.6	1.9 ± 0.6
Tagged Wcc/Wc	0.0 ± 0.0	37.2 ± 11.5	14.0 ± 4.4	3.7 ± 1.4	1.5 ± 0.5
Tagged Total HF	0.0 ± 0.0	80.3 ± 24.4	30.7 ± 9.4	8.2 ± 2.7	3.4 ± 1.1
Tagged Total MC	0.0 ± 0.0	69.1 ± 6.3	87.8 ± 10.3	74.1 ± 9.8	19.8 ± 2.7
Tagged Mistags	0.0 ± 0.0	41.2 ± 5.4	14.7 ± 2.4	3.7 ± 1.3	1.5 ± 0.5
Tagged Non-W	0.0 ± 0.4	6.5 ± 2.6	3.0 ± 1.2	0.0 ± 0.0	0.0 ± 0.0
Total Prediction	0.0 ± 0.0	197.1 ± 25.9	136.1 ± 14.2	86.0 ± 10.3	24.7 ± 2.9
Tagged WH100	0.0 ± 0.0	1.3 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	1.2 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	1.2 ± 0.1	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.9 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.8 ± 0.1	0.3 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.5 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	179.0 ± 0.0	134.0 ± 0.0	92.0 ± 0.0	33.0 ± 0.0

Table 14: Number of expected single tagged signal and background events with a loose muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	156167.0 ± 0.0	26681.0 ± 0.0	4648.0 ± 0.0	1011.0 ± 0.0	210.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.2 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WZ	0.0 ± 0.0	1.8 ± 0.2	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	9.3 ± 1.5	36.6 ± 6.0	50.6 ± 8.2	14.7 ± 2.4
Tagged TopDil	0.0 ± 0.0	5.4 ± 0.9	4.4 ± 0.7	1.1 ± 0.2	0.2 ± 0.0
Tagged STopT	0.0 ± 0.0	0.7 ± 0.1	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	5.3 ± 0.9	1.7 ± 0.3	0.4 ± 0.1	0.1 ± 0.0
Tagged Z+jets	0.0 ± 0.0	0.4 ± 0.1	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Wbb	0.0 ± 0.0	27.4 ± 10.0	9.8 ± 3.8	2.6 ± 1.1	0.6 ± 0.3
Tagged Wcc/Wc	0.0 ± 0.0	2.6 ± 1.0	1.3 ± 0.5	0.5 ± 0.2	0.1 ± 0.1
Tagged Total HF	0.0 ± 0.0	30.0 ± 10.8	11.0 ± 4.1	3.1 ± 1.2	0.7 ± 0.3
Tagged Total MC	0.0 ± 0.0	23.2 ± 2.9	44.5 ± 6.6	52.5 ± 8.3	15.2 ± 2.4
Tagged Mistags	0.0 ± 0.0	0.8 ± 0.3	0.6 ± 0.2	0.2 ± 0.1	0.1 ± 0.0
Tagged Non-W	0.0 ± 0.8	5.9 ± 2.3	3.9 ± 1.5	0.8 ± 0.3	0.4 ± 0.2
Total Prediction	0.0 ± 0.0	59.9 ± 11.4	60.0 ± 7.9	56.6 ± 8.4	16.3 ± 2.4
Tagged WH100	0.0 ± 0.0	1.2 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	1.1 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	1.0 ± 0.1	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.8 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.5 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	68.0 ± 0.0	82.0 ± 0.0	66.0 ± 0.0	16.0 ± 0.0

Table 15: Number of expected double tagged signal and background events with a CEM electron in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	56369.0 ± 0.0	8415.0 ± 0.0	1446.0 ± 0.0	240.0 ± 0.0	51.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WZ	0.0 ± 0.0	0.8 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	1.8 ± 0.3	7.1 ± 1.2	9.7 ± 1.6	2.7 ± 0.4
Tagged TopDil	0.0 ± 0.0	1.4 ± 0.2	1.1 ± 0.2	0.2 ± 0.0	0.0 ± 0.0
Tagged STopT	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	1.0 ± 0.2	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Z+jets	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged Wbb	0.0 ± 0.0	10.5 ± 3.3	3.6 ± 1.1	0.8 ± 0.3	0.2 ± 0.1
Tagged Wcc/Wc	0.0 ± 0.0	1.1 ± 0.4	0.5 ± 0.2	0.1 ± 0.1	0.1 ± 0.0
Tagged Total HF	0.0 ± 0.0	11.6 ± 3.7	4.1 ± 1.3	1.0 ± 0.3	0.3 ± 0.1
Tagged Total MC	0.0 ± 0.0	5.2 ± 0.6	9.0 ± 1.3	10.1 ± 1.6	2.8 ± 0.4
Tagged Mistags	0.0 ± 0.0	0.4 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Non-W	0.0 ± 0.8	0.4 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Total Prediction	0.0 ± 0.0	17.7 ± 3.7	13.4 ± 1.9	11.1 ± 1.6	3.1 ± 0.4
Tagged WH100	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	19.0 ± 0.0	13.0 ± 0.0	11.0 ± 0.0	4.0 ± 0.0

Table 16: Number of expected double tagged signal and background events with a PEM electron in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	71346.0 ± 0.0	11348.0 ± 0.0	1875.0 ± 0.0	441.0 ± 0.0	73.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WZ	0.0 ± 0.0	1.0 ± 0.1	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	5.2 ± 0.9	20.3 ± 3.3	28.2 ± 4.6	8.1 ± 1.3
Tagged TopDil	0.0 ± 0.0	3.1 ± 0.5	2.9 ± 0.5	0.7 ± 0.1	0.1 ± 0.0
Tagged STopT	0.0 ± 0.0	0.4 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	2.9 ± 0.5	1.0 ± 0.2	0.2 ± 0.0	0.0 ± 0.0
Tagged Z+jets	0.0 ± 0.0	0.7 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Wbb	0.0 ± 0.0	16.6 ± 5.2	5.2 ± 1.6	1.4 ± 0.5	0.1 ± 0.1
Tagged Wcc/Wc	0.0 ± 0.0	1.6 ± 0.5	0.7 ± 0.2	0.3 ± 0.1	0.0 ± 0.0
Tagged Total HF	0.0 ± 0.0	18.2 ± 5.7	5.9 ± 1.9	1.7 ± 0.6	0.2 ± 0.1
Tagged Total MC	0.0 ± 0.0	13.3 ± 1.7	25.2 ± 3.7	29.4 ± 4.7	8.4 ± 1.3
Tagged Mistags	0.0 ± 0.0	0.5 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Non-W	0.0 ± 0.8	2.5 ± 1.0	0.5 ± 0.2	0.0 ± 0.0	0.0 ± 0.0
Total Prediction	0.0 ± 0.0	34.5 ± 6.0	31.9 ± 4.2	31.2 ± 4.7	8.5 ± 1.3
Tagged WH100	0.0 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	0.6 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	0.5 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	32.0 ± 0.0	35.0 ± 0.0	24.0 ± 0.0	8.0 ± 0.0

Table 17: Number of expected double tagged signal and background events with a CMUP muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	44682.0 ± 0.0	6848.0 ± 0.0	1080.0 ± 0.0	234.0 ± 0.0	44.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WZ	0.0 ± 0.0	0.5 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	2.5 ± 0.5	9.4 ± 1.7	13.1 ± 2.4	3.9 ± 0.7
Tagged TopDil	0.0 ± 0.0	1.4 ± 0.3	1.2 ± 0.2	0.3 ± 0.0	0.1 ± 0.0
Tagged STopT	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	1.3 ± 0.2	0.4 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Z+jets	0.0 ± 0.0	0.4 ± 0.1	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged Wbb	0.0 ± 0.0	9.3 ± 2.9	3.0 ± 0.9	0.9 ± 0.3	0.2 ± 0.1
Tagged Wcc/Wc	0.0 ± 0.0	0.8 ± 0.3	0.4 ± 0.1	0.2 ± 0.1	0.0 ± 0.0
Tagged Total HF	0.0 ± 0.0	10.2 ± 3.2	3.4 ± 1.1	1.1 ± 0.3	0.2 ± 0.1
Tagged Total MC	0.0 ± 0.0	6.4 ± 1.0	11.6 ± 2.0	13.6 ± 2.4	4.0 ± 0.7
Tagged Mistags	0.0 ± 0.0	0.3 ± 0.1	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged Non-W	0.0 ± 0.8	0.2 ± 0.1	0.5 ± 0.2	0.0 ± 0.0	0.3 ± 0.1
Total Prediction	0.0 ± 0.0	17.1 ± 3.3	15.6 ± 2.3	14.7 ± 2.5	4.5 ± 0.7
Tagged WH100	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	13.0 ± 0.0	15.0 ± 0.0	14.0 ± 0.0	8.0 ± 0.0

Table 18: Number of expected double tagged signal and background events with a CMX muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	1jet	2jets	3jets	4jets	5jets
All Pretag Cands.	0.0 ± 0.0	3882.0 ± 0.0	1169.0 ± 0.0	356.0 ± 0.0	94.0 ± 0.0
Tagged WW	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged WZ	0.0 ± 0.0	0.7 ± 0.1	0.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged ZZ	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged TopLJ	0.0 ± 0.0	4.9 ± 0.8	19.9 ± 3.3	27.7 ± 4.5	8.0 ± 1.3
Tagged TopDil	0.0 ± 0.0	2.8 ± 0.5	2.8 ± 0.5	0.7 ± 0.1	0.1 ± 0.0
Tagged STopT	0.0 ± 0.0	0.3 ± 0.1	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0
Tagged STopS	0.0 ± 0.0	2.3 ± 0.4	0.8 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Tagged Z+jets	0.0 ± 0.0	0.6 ± 0.1	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Wbb	0.0 ± 0.0	6.5 ± 2.1	2.6 ± 0.9	0.9 ± 0.3	0.4 ± 0.1
Tagged Wcc/Wc	0.0 ± 0.0	0.6 ± 0.2	0.4 ± 0.2	0.2 ± 0.1	0.1 ± 0.0
Tagged Total HF	0.0 ± 0.0	7.1 ± 2.3	3.0 ± 1.0	1.0 ± 0.4	0.5 ± 0.2
Tagged Total MC	0.0 ± 0.0	11.7 ± 1.5	24.5 ± 3.6	28.8 ± 4.6	8.2 ± 1.3
Tagged Mistags	0.0 ± 0.0	0.2 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Tagged Non-W	0.0 ± 0.8	0.0 ± 0.0	0.0 ± 0.0	3.9 ± 1.6	0.0 ± 0.0
Total Prediction	0.0 ± 0.0	19.0 ± 2.7	27.7 ± 3.8	33.8 ± 4.8	8.8 ± 1.3
Tagged WH100	0.0 ± 0.0	0.5 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH105	0.0 ± 0.0	0.4 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH110	0.0 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH115	0.0 ± 0.0	0.4 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH120	0.0 ± 0.0	0.3 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH125	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH130	0.0 ± 0.0	0.2 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH135	0.0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH140	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH145	0.0 ± 0.0	0.1 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Tagged WH150	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Observed	0.0 ± 0.0	24.0 ± 0.0	29.0 ± 0.0	29.0 ± 0.0	10.0 ± 0.0

Table 19: Number of expected double tagged signal and background events with a loose muon in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

C Calculation of Phase Space for W+2jet Events

For WH events[7], $qq' \to t\bar{b} \to b\bar{b}l\nu$, the phase space factor can be written as:

$$d\Phi_4(q_1+q_2;p_1,p_2,p_3,p_4) = \delta^4(q_1+q_2-\sum_{i=1}^{4} p_i)\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} \tag{21}$$

where q_1 and q_2 are the four momenta of the initial quarks; p_1, p_2 are the four momenta of the b and $\bar{b}$ quark, respectively; and p_3, p_4 are the four momenta of the lepton and neutrino, respectively. It is convenient to change variables from momenta $(\vec{p}_1,\vec{p}_2,\vec{p}_3,\vec{p}_4)$ to $(\rho_1,\Omega_1,\rho_2,\Omega_2,\vec{p}_l,\vec{p}_{\nu,T},m_W,\ p_{\nu,z})$ where $\rho_i = |\vec{p}_i|$ is the absolute momentum of the quarks, Ω_i are the quark solid angles, and m_W is the mass of the W boson. One way to perform the transformation is to use the recursive character of the phase space [7]:

$$\begin{aligned}\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} &= \delta(p_W^2-m_W^2)dm_W^2\prod_{i=1}^{4}\frac{d^3\vec{p}_i}{(2\pi)^3 2E_i} \\ &= \frac{dm_W^2}{|\frac{\partial p_W^2}{\partial p_{\nu,z}}|}\frac{d^3\vec{p}_l}{(2\pi)^3 2E_l}\frac{d\vec{p}_{\nu,T}}{(2\pi)^3 2E_\nu}\prod_{i=1}^{2}\frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i}\end{aligned} \tag{22}$$

In the last step, the δ-function was integrated with respect to $p_{\nu,z}$ using:

$$\int f(x)\delta[g(x)]dx = \frac{f(a)}{|g'(a)|}\ \ ,at\ g(a)=0 \tag{23}$$

The partial derivative becomes (neglecting neutrino and lepton masses):

$$\begin{aligned}\frac{\partial p_W^2}{\partial p_{\nu,z}} &= \frac{\partial(p_l+p_\nu)^2}{\partial p_{\nu,z}} \\ &= \frac{\partial}{\partial p_{\nu,z}}(m_l^2+m_\nu^2+2E_lE_\nu-2p_{l,z}p_{\nu,z}-2\vec{p}_{l,T}\vec{p}_{\nu,T}) \\ &= 2E_l\frac{p_{\nu,z}}{E_\nu}-2p_{l,z}\end{aligned} \tag{24}$$

Finally, we substitute Equation 22 into Equation 21 and integrate two δ-functions with respect to the transverse momentum of the neutrino $\vec{p}_{\nu,T}$. The remaining two δ-functions are integrated with respect to the initial quark's longitudinal momentum and energy in the event probability. The expression for the phase space for WH events is:

$$\begin{aligned}d\Phi_4 &= \delta(E_{q_1}+E_{q_2}-\sum_{i=1}^{4}E_i)\ \delta(p_{q_1,z}+p_{q_2,z}-\sum_{i=1}^{4}p_{i,z}) \\ &\times\frac{dm_W^2}{|2E_l\frac{p_{\nu,z}}{E_\nu}-2p_{l,z}|}\frac{d^3\vec{p}_l}{(2\pi)^3 2E_l}\frac{1}{(2\pi)^3 2E_\nu}\prod_{i=1}^{2}\frac{\rho_i^2 d\rho_i d\Omega_i}{(2\pi)^3 2E_i}\end{aligned} \tag{25}$$

[7] and similar for other W + 2 jet topologies

D Validation of the NN Transfer Function Input Variables

Figures 28-29 and 30-31 show the data MC comparisons of all the NN input variables in the untagged control regions and in the single tag signal region, respectively.

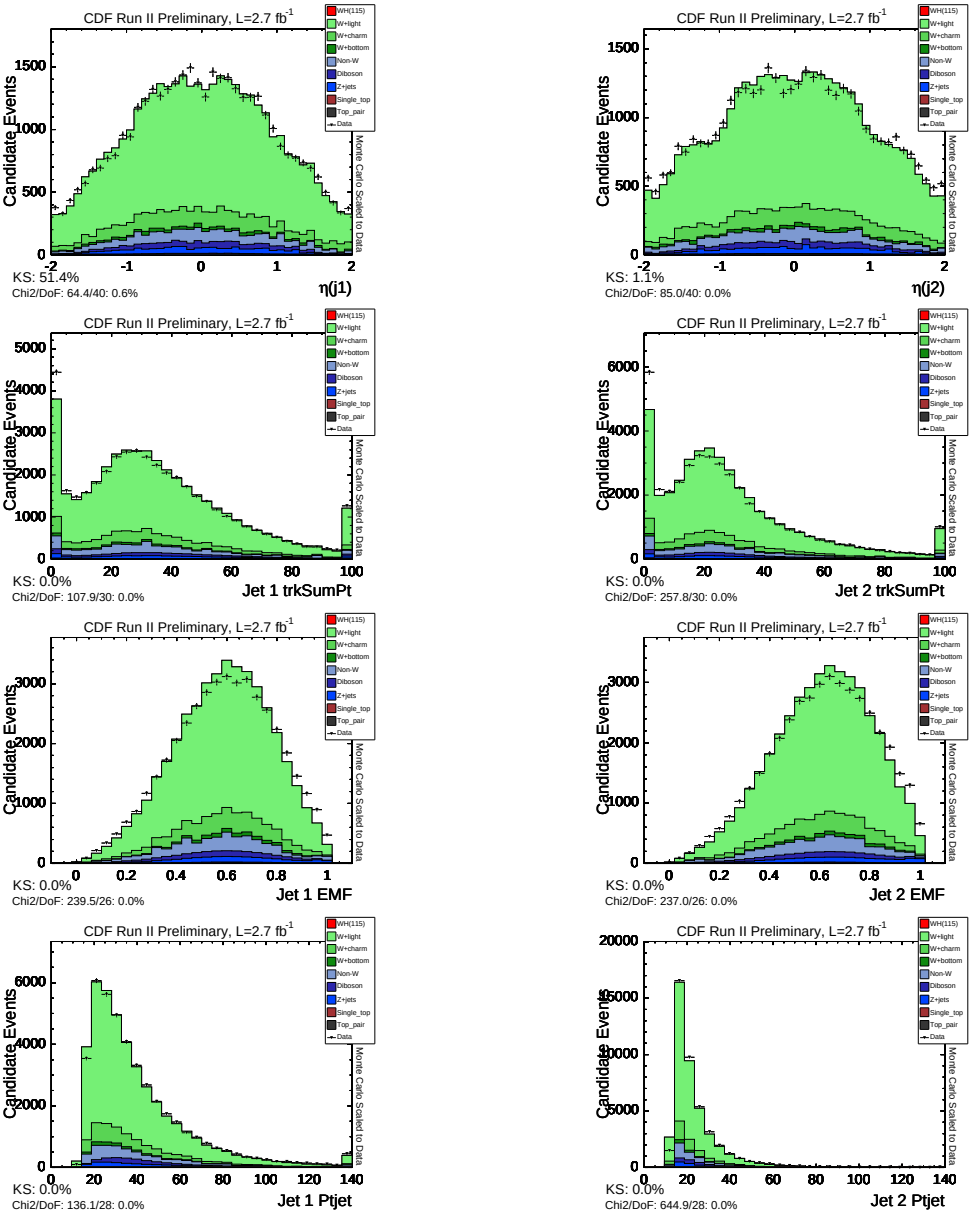

Figure 28: Left (Right): Distributions, in the untagged sample, of the input variables of the first (second) jet used for the training of the NN. From top to bottom: pseudorapidity, SumE, electromagnetic fraction and transverse momentum of the jet.

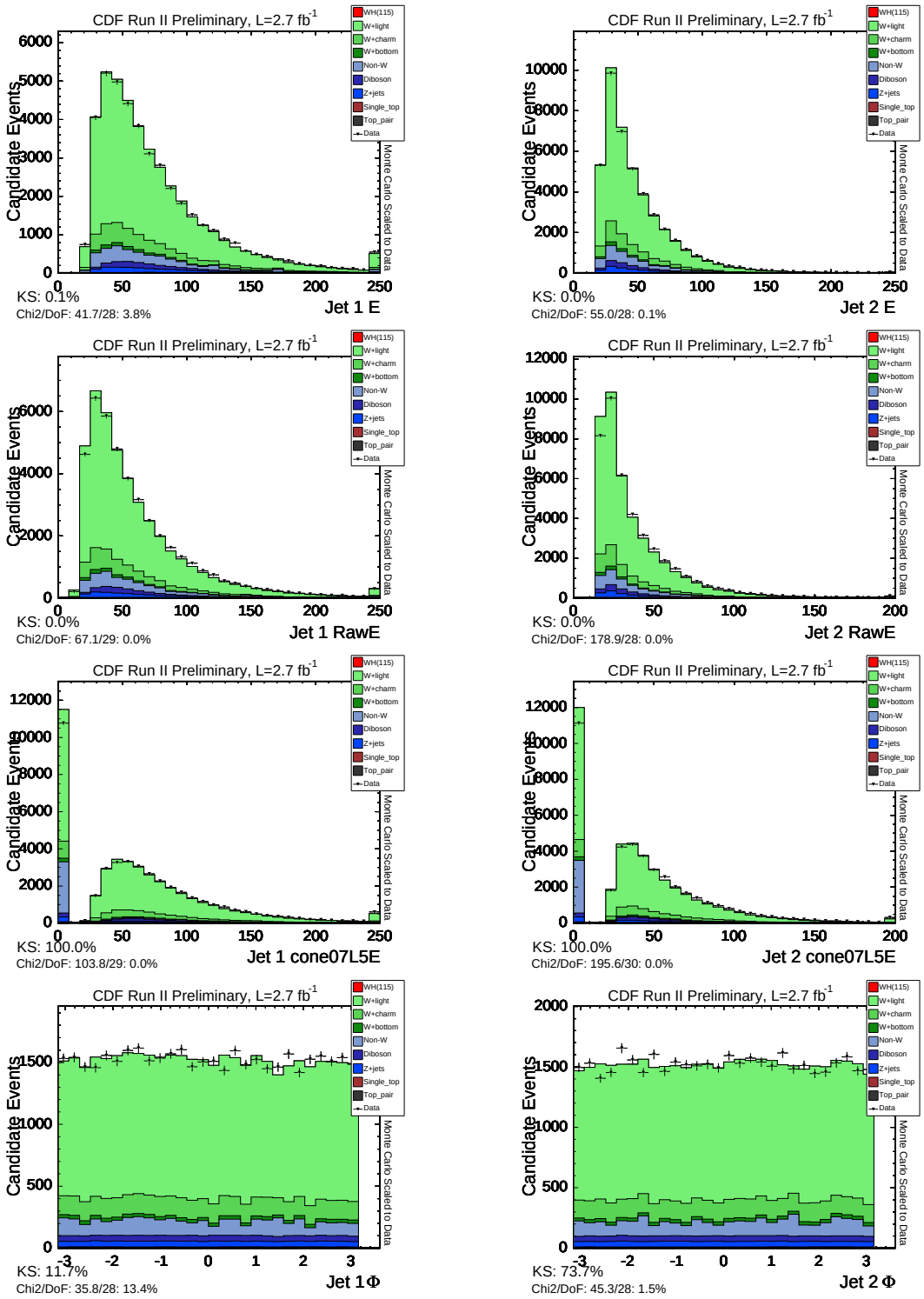

Figure 29: Left (Right): Distributions, in the untagged sample, of the input variables of the first (second) jet used for the training of the NN. From top to bottom: total energy (L5 corrected), raw energy, energy of the jet of R_{cone}=0.7 and ϕ of the jet.

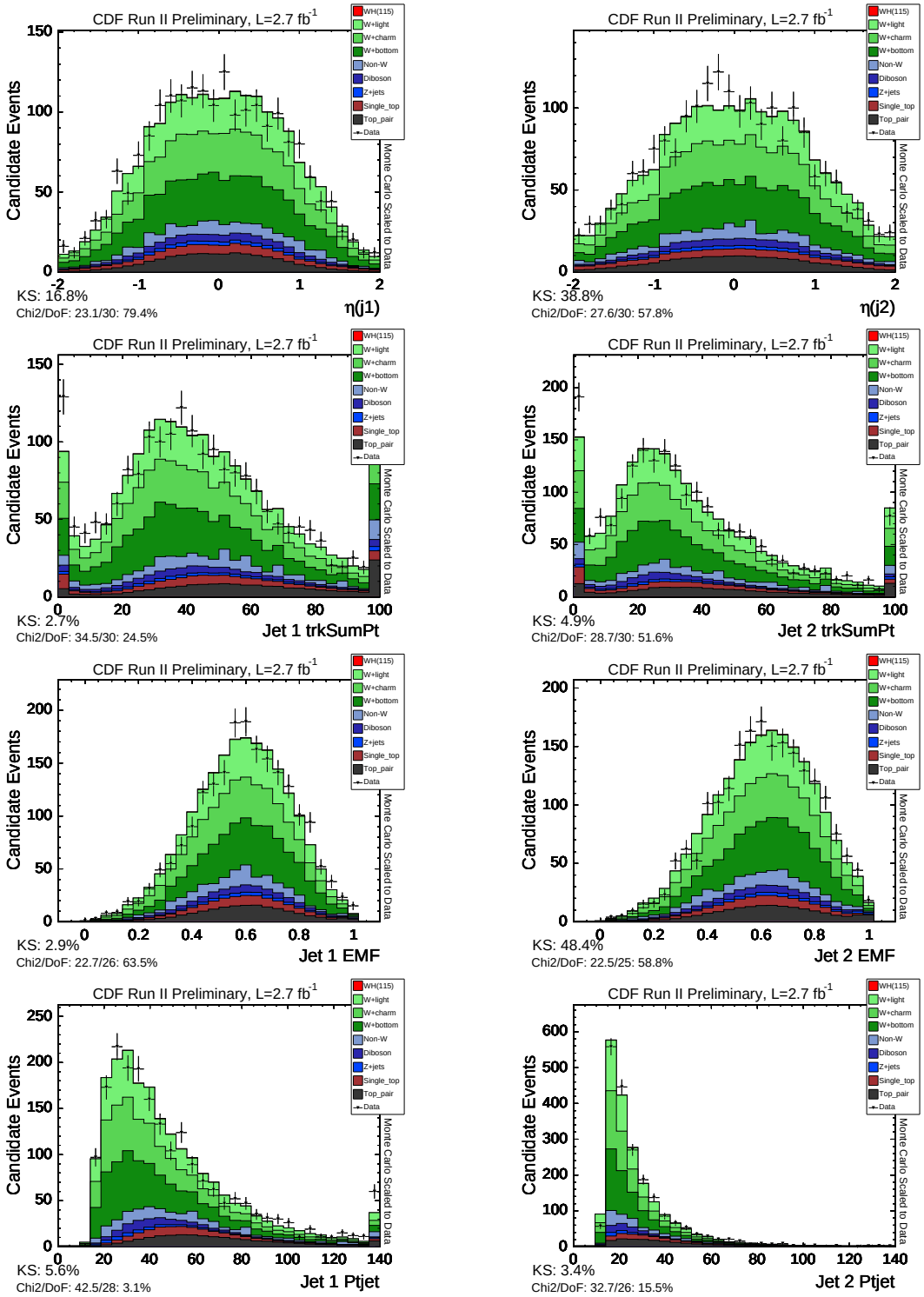

Figure 30: Left (Right): Distributions, in the single tagged sample, of the input variables of the first (second) jet used for the training of the NN. From top to bottom: pseudorapidity, SumE, electromagnetic fraction and transverse momentum of the jet.

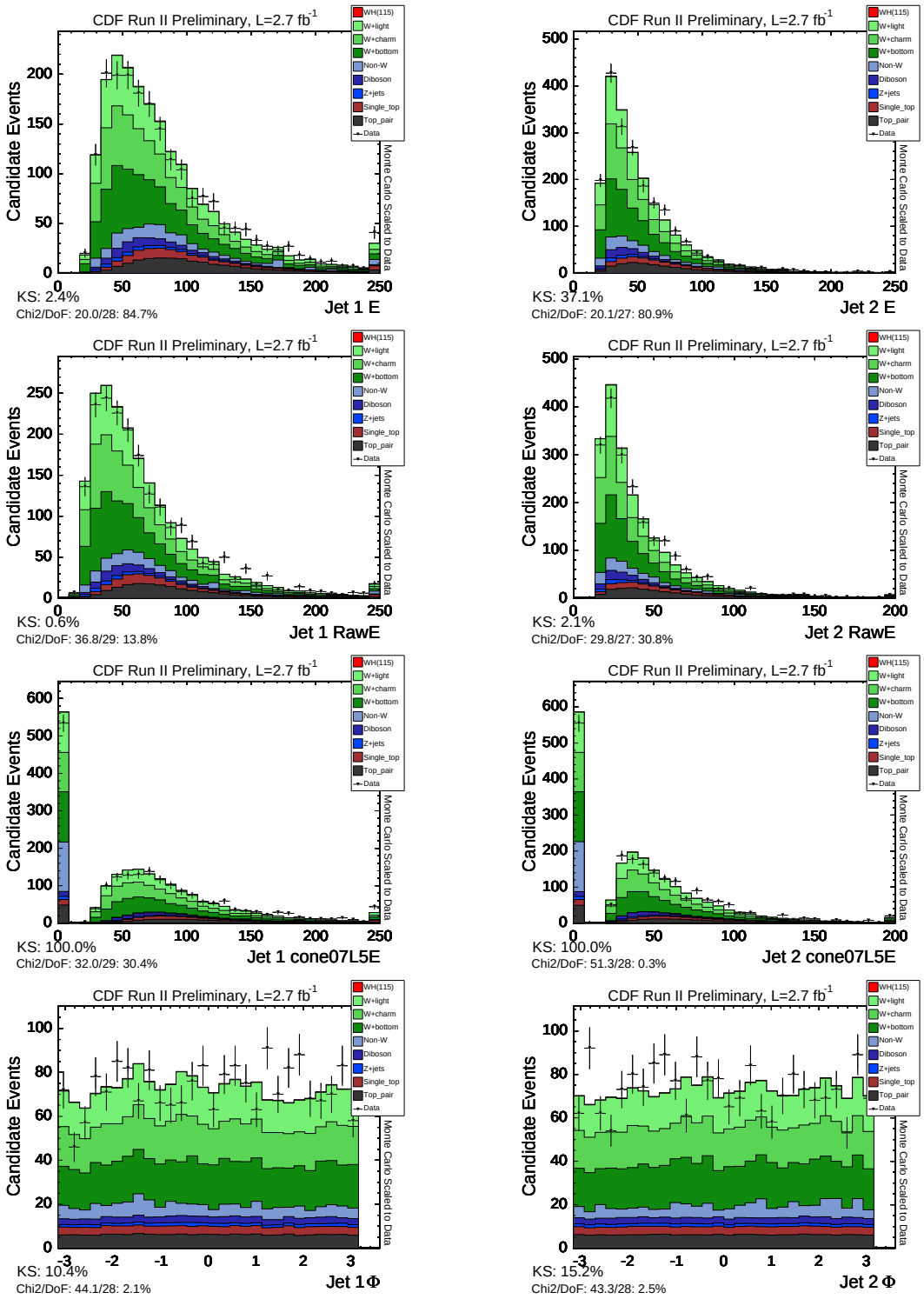

Figure 31: Left (Right): Distributions, in the single tagged sample, of the input variables of the first (second) jet used for the training of the NN. From top to bottom: total energy (L5 corrected), raw energy, energy of the jet of R_{cone}=0.7 and ϕ of the jet.

E Monte Carlo Modeling Checks

In Figs. 32-36, we check the modeling of the Monte Carlo by comparing a few kinematic distributions predicted by the Monte Carlo with data control samples. We choose the "taggable but not tagged" side-band data. That is, we select $W+2$ 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 WH signal region while it still represents a very similar kinematic event topology. Another advantage is that this sample has negligible contribution from signal.

The measured quantities which are input in the probabilities are the four-vectors of the lepton as well as the measured four-vectors of the leading and second leading jet. Figure 37 (39) shows data/Monte Carlo comparisons for the measured four-vector components of the lepton and the two jets in the untagged (single tag) sample.

Figures 38 (40) shows the transverse momentum and pseudo-rapidity distribution for the same reconstructed objects in the untagged (single tag) sample.

Finally, Figs. 43 and 44 show the event probabilities in the untagged and single tag sample, respectively.

The Monte Carlo distributions are normalized to the data. We find generally good agreement between data and Monte Carlo prediction.

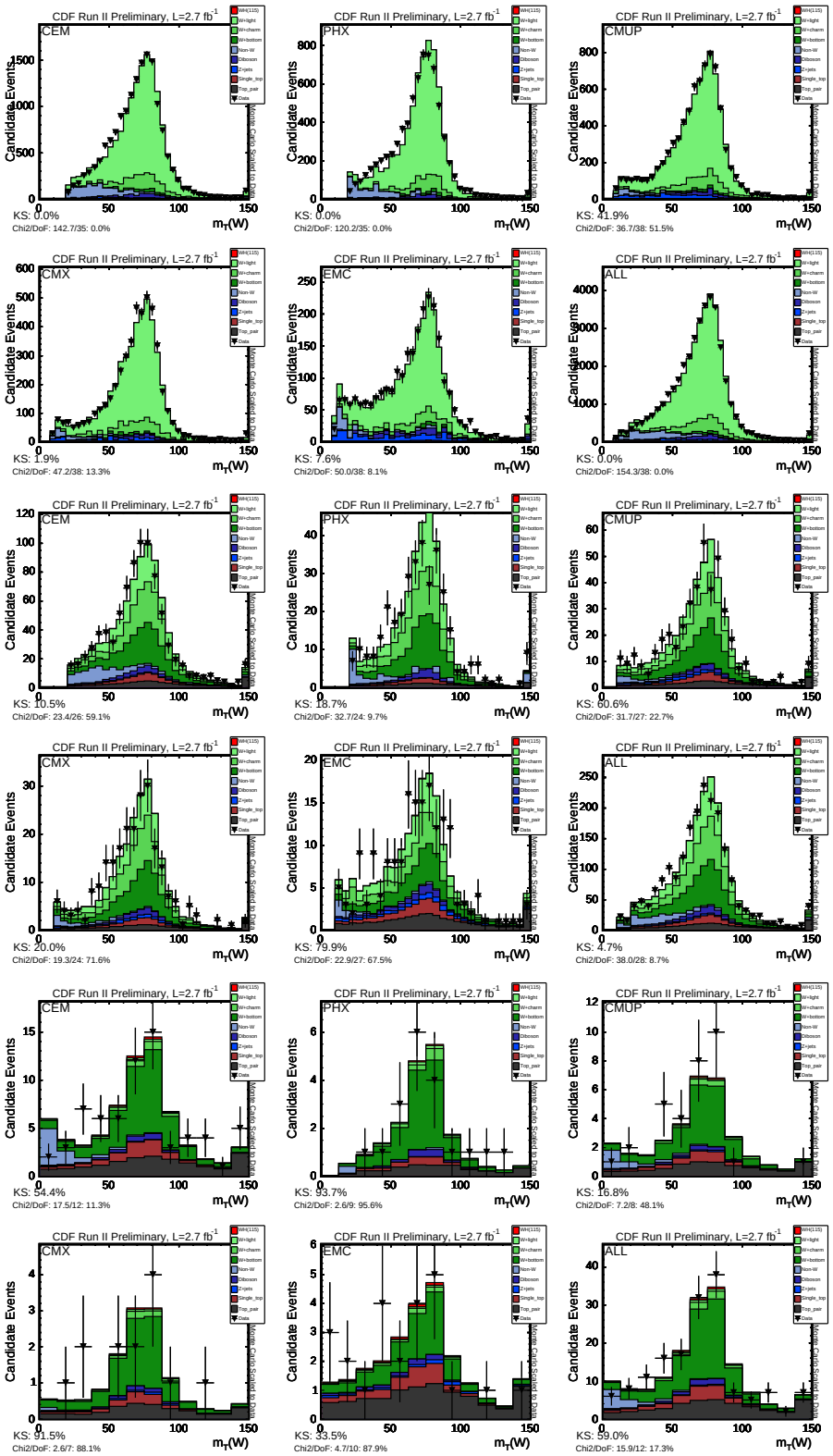

Figure 32: Distribution of the W boson transverse mass in the "taggable but not tagged" control sample (rows 1-2), the 1tag signal sample (rows 3-4) and the 2tag signal sample (rows 5-6). All sets of 2 rows show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

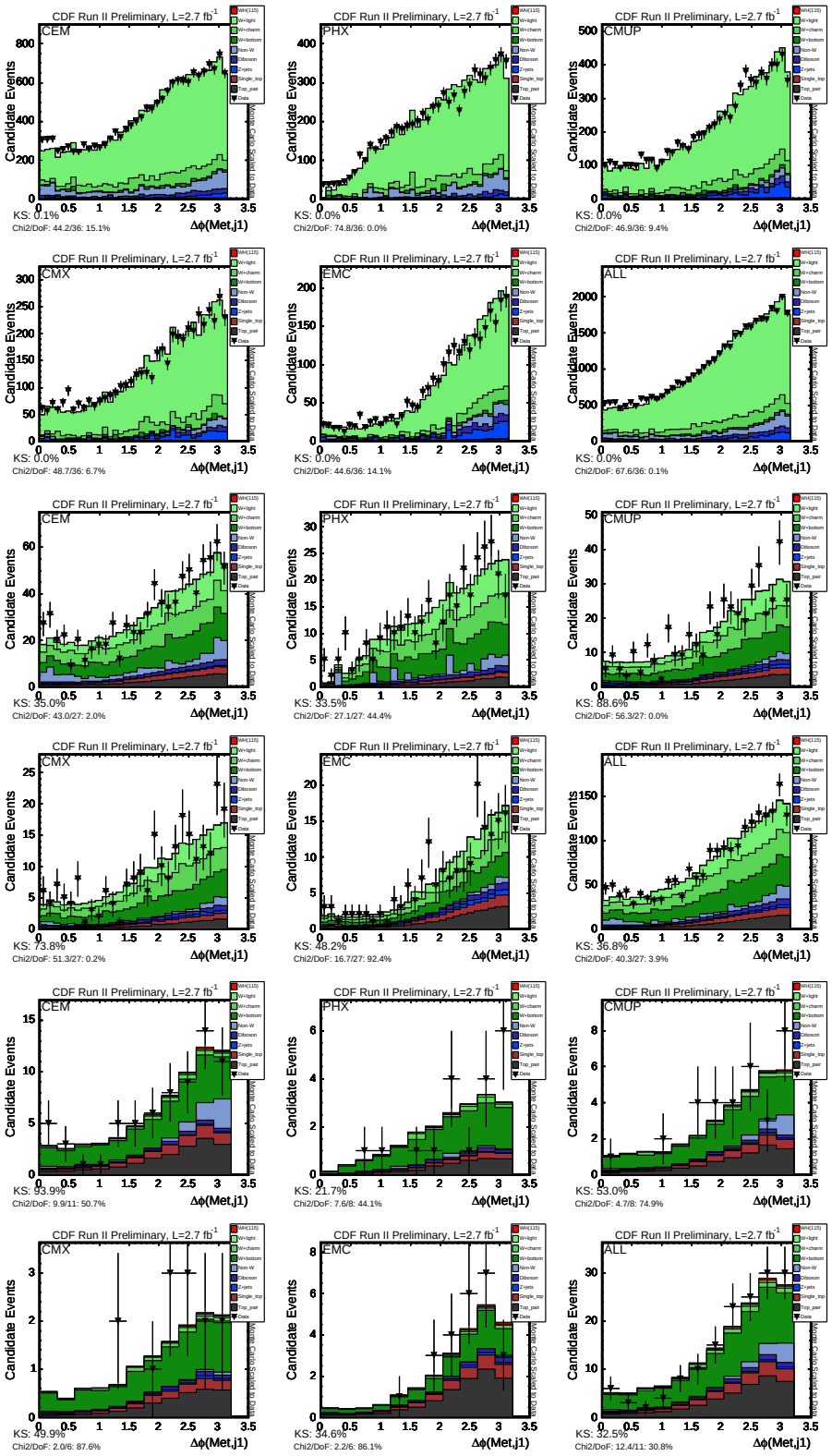

Figure 33: Distribution of the Phi difference of the MET and leading Jet in the "taggable but not tagged" control sample (rows 1-2), the 1tag signal sample (rows 3-4) and the 2tag signal sample (rows 5-6). All sets of 2 rows show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

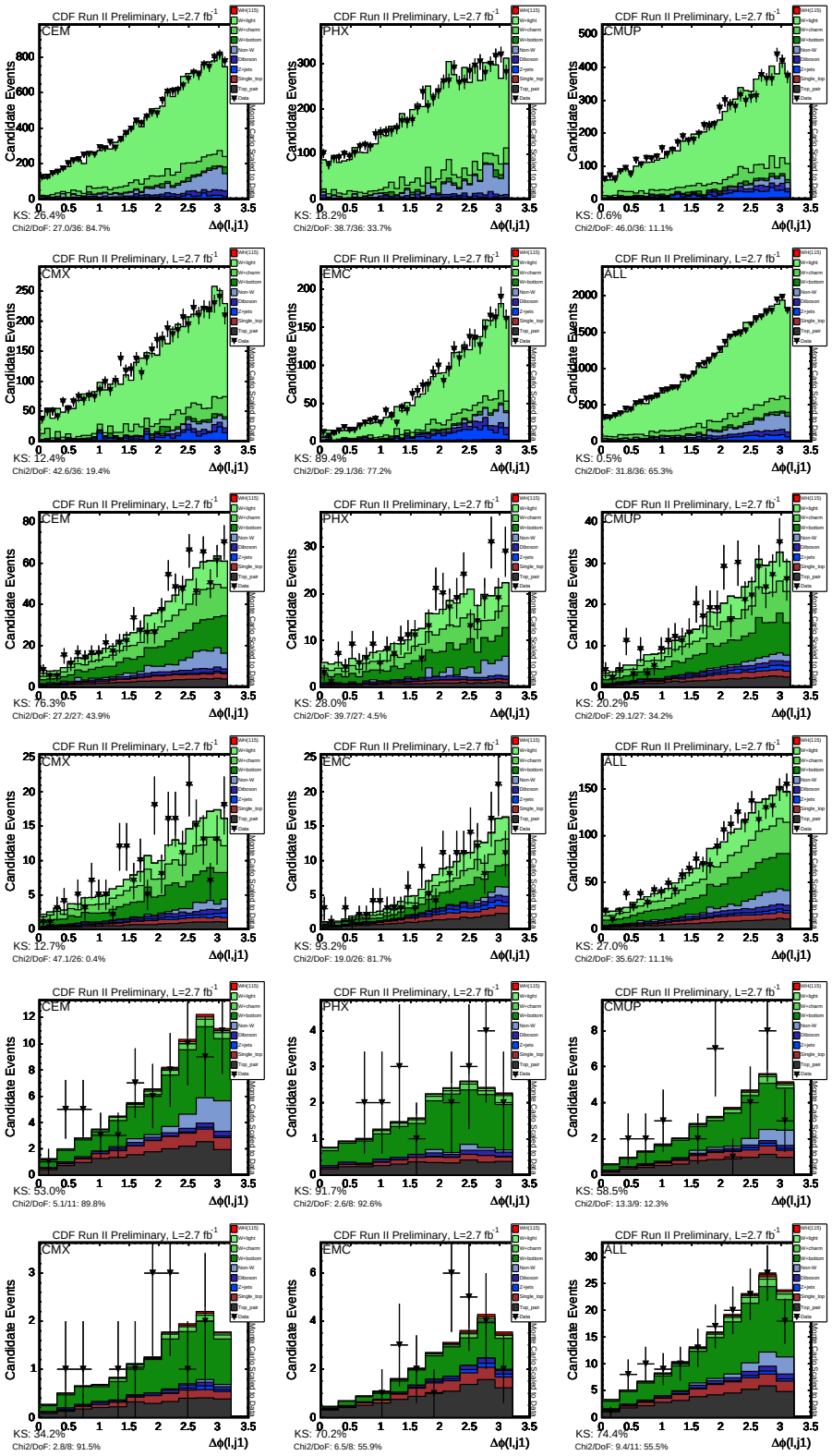

Figure 34: Distribution of the Phi difference of the Lepton and leading Jet in the "taggable but not tagged" control sample (rows 1-2), the 1tag signal sample (rows 3-4) and the 2tag signal sample (rows 5-6). All sets of 2 rows show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

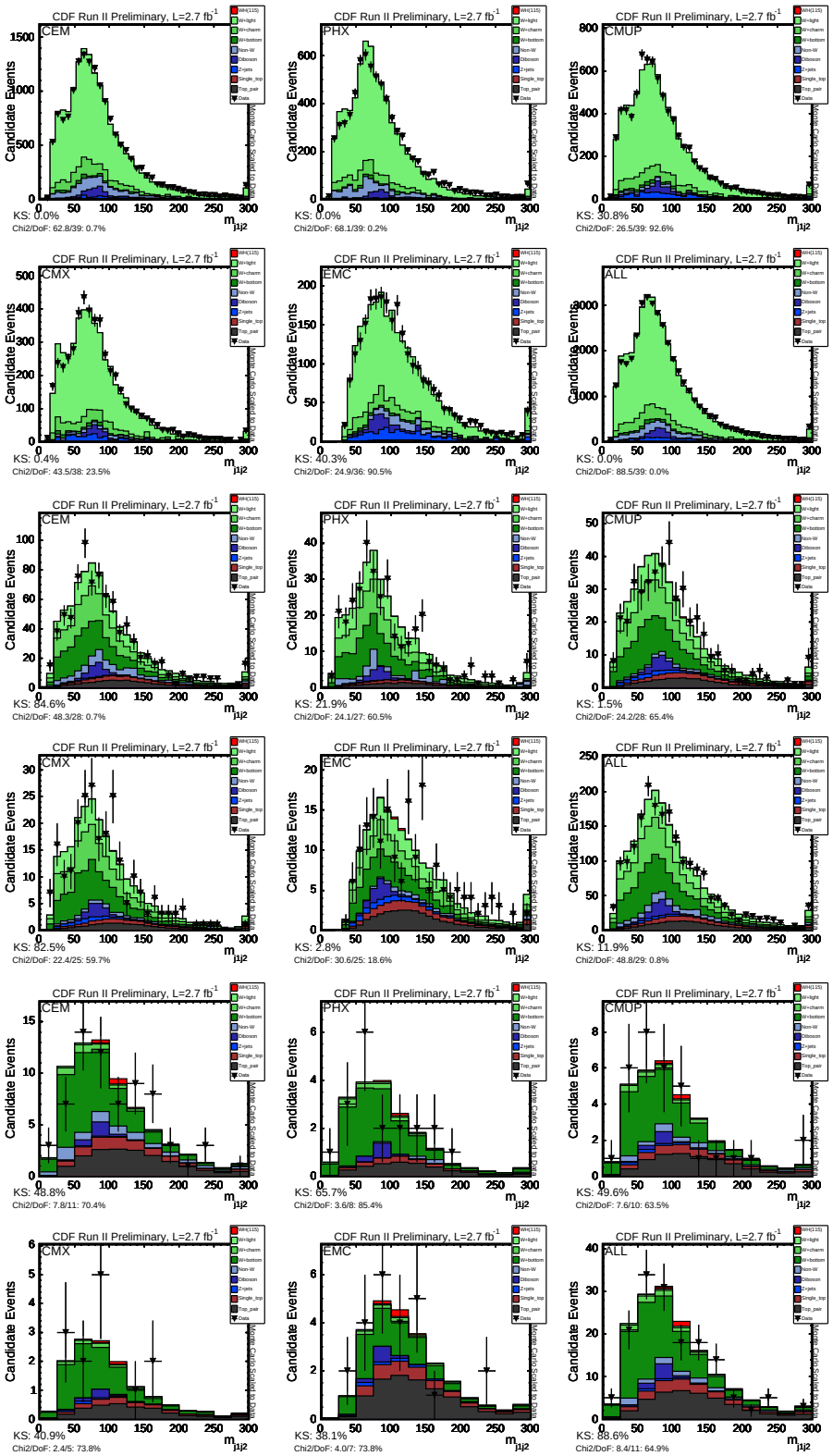

Figure 35: Distribution of the mass of the two leading jets in the "taggable but not tagged" control sample (rows 1-2), the 1tag signal sample (rows 3-4) and the 2tag signal sample (rows 5-6). All sets of 2 rows show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

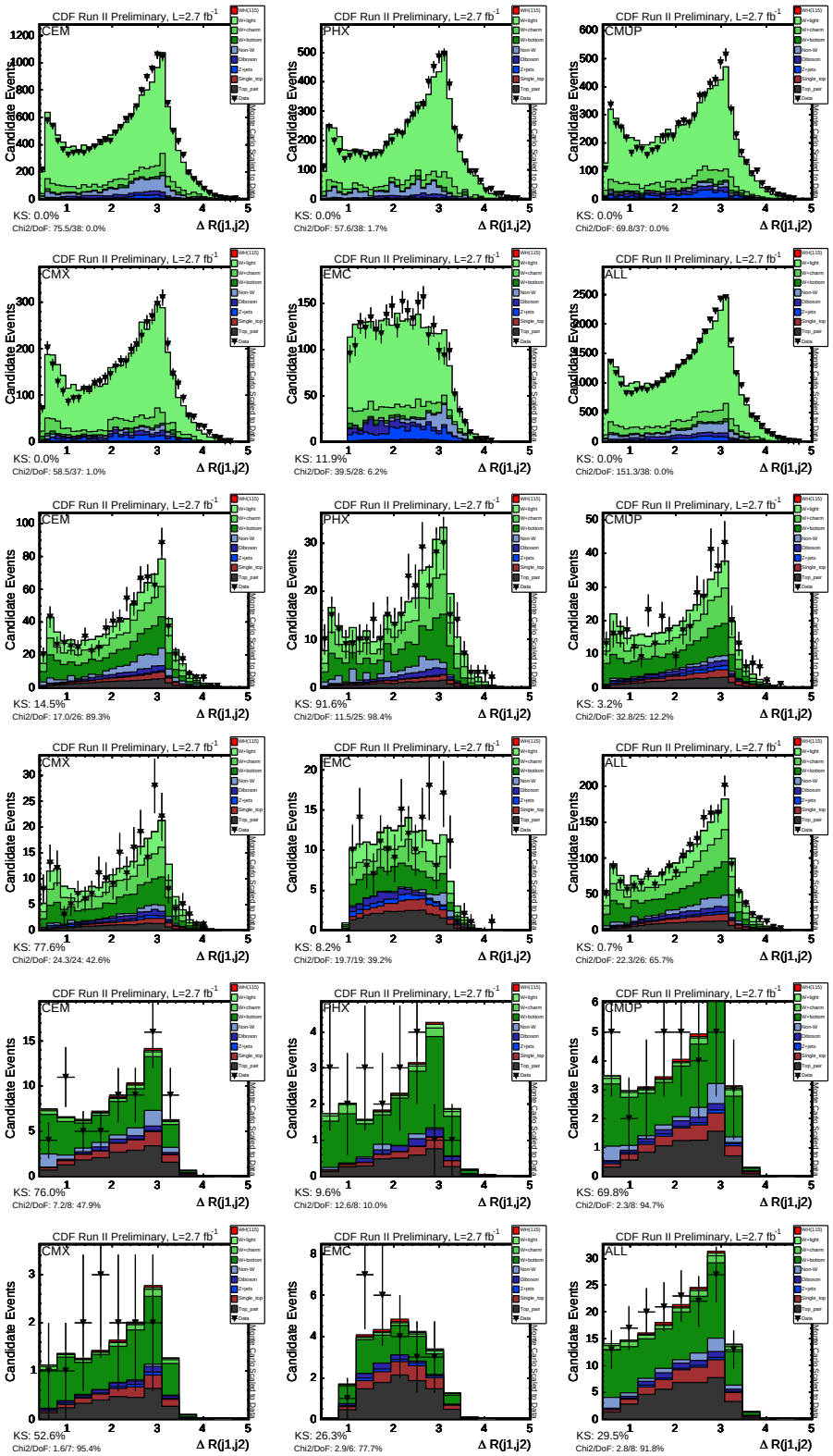

Figure 36: Distribution of the ΔR difference of the two leading jets in the "taggable but not tagged" control sample (rows 1-2), the 1tag signal sample (rows 3-4) and the 2tag signal sample (rows 5-6). All sets of 2 rows show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

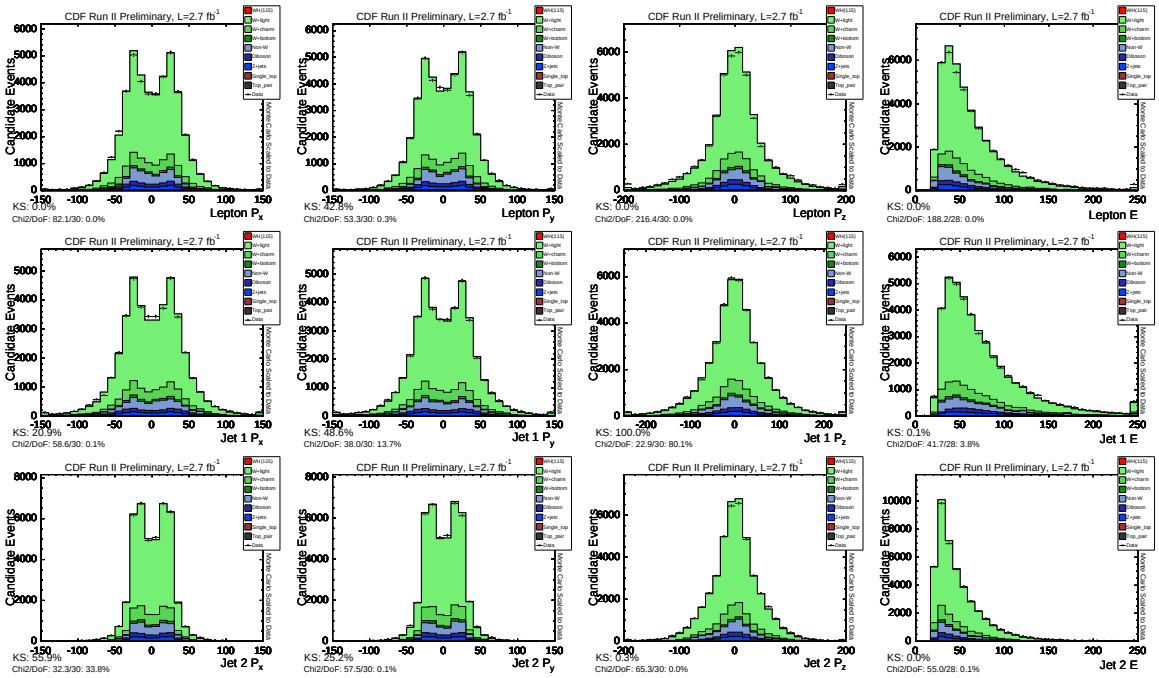

Figure 37: Comparison of the event probability input 4-vectors for untagged W + 2 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

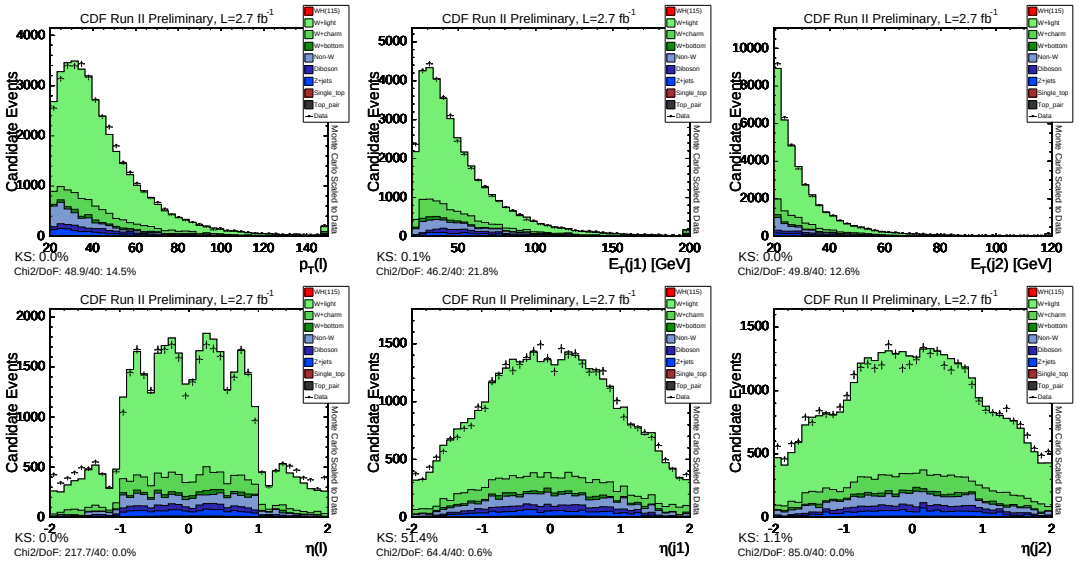

Figure 38: Comparison of the event probability input p_T and pseudo-rapidity distribution for untagged W + 2 jet control data (with at least one taggable jet) compared to the Monte Carlo prediction.

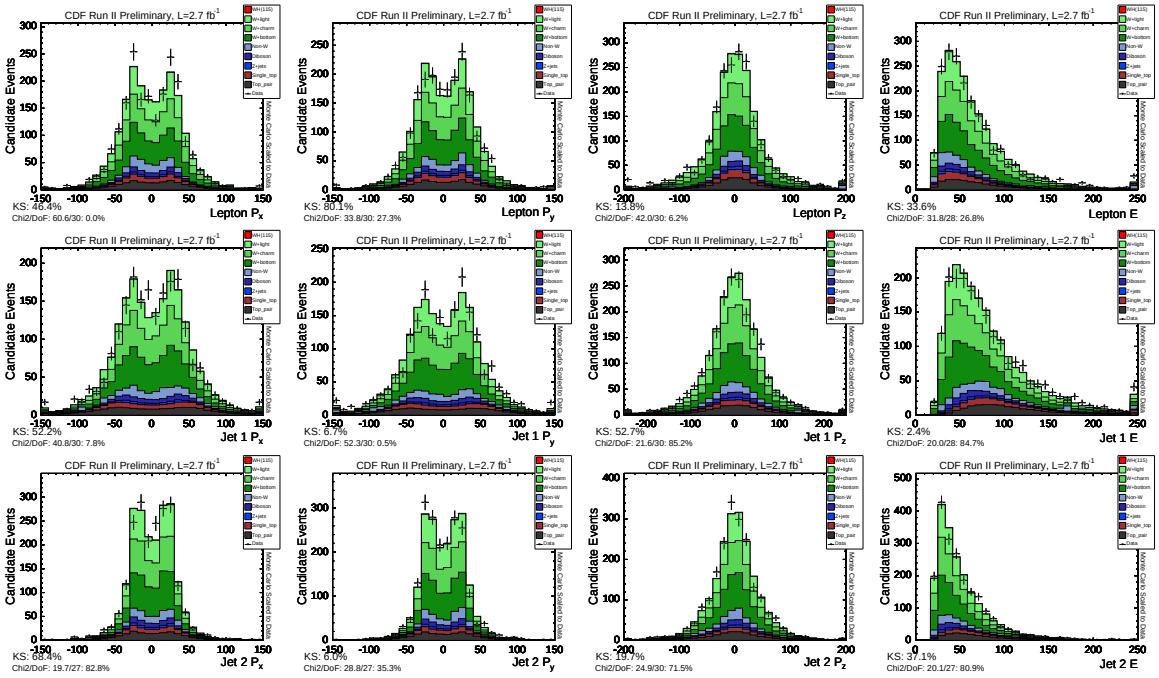

Figure 39: Comparison of the event probability input 4-vectors for single tag W + 2 jet data events compared to the Monte Carlo prediction.

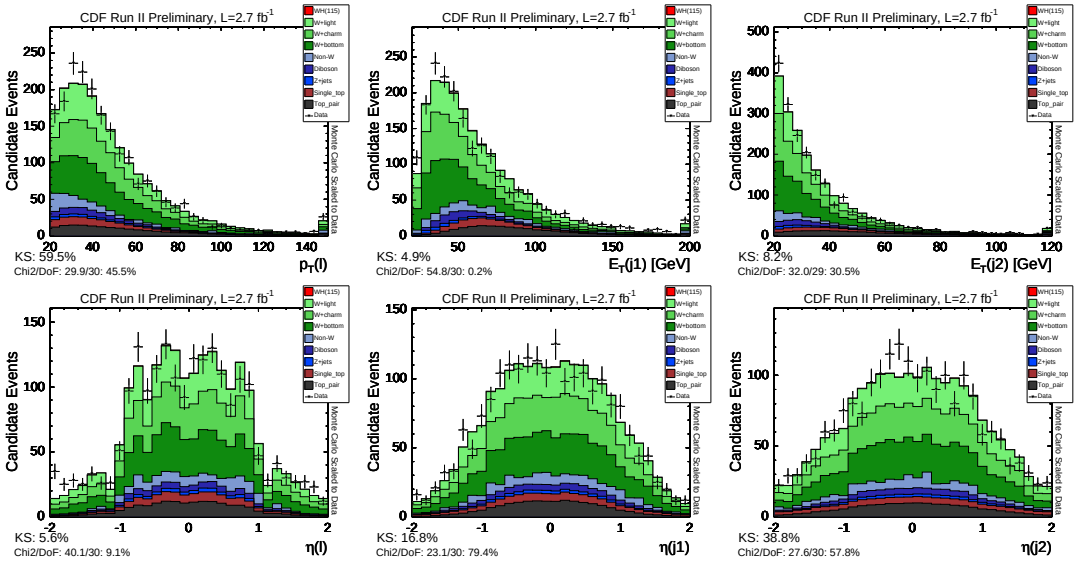

Figure 40: Comparison of the event probability input p_T and pseudo-rapidity distribution for single tag W + 2 jet data events compared to the Monte Carlo prediction.

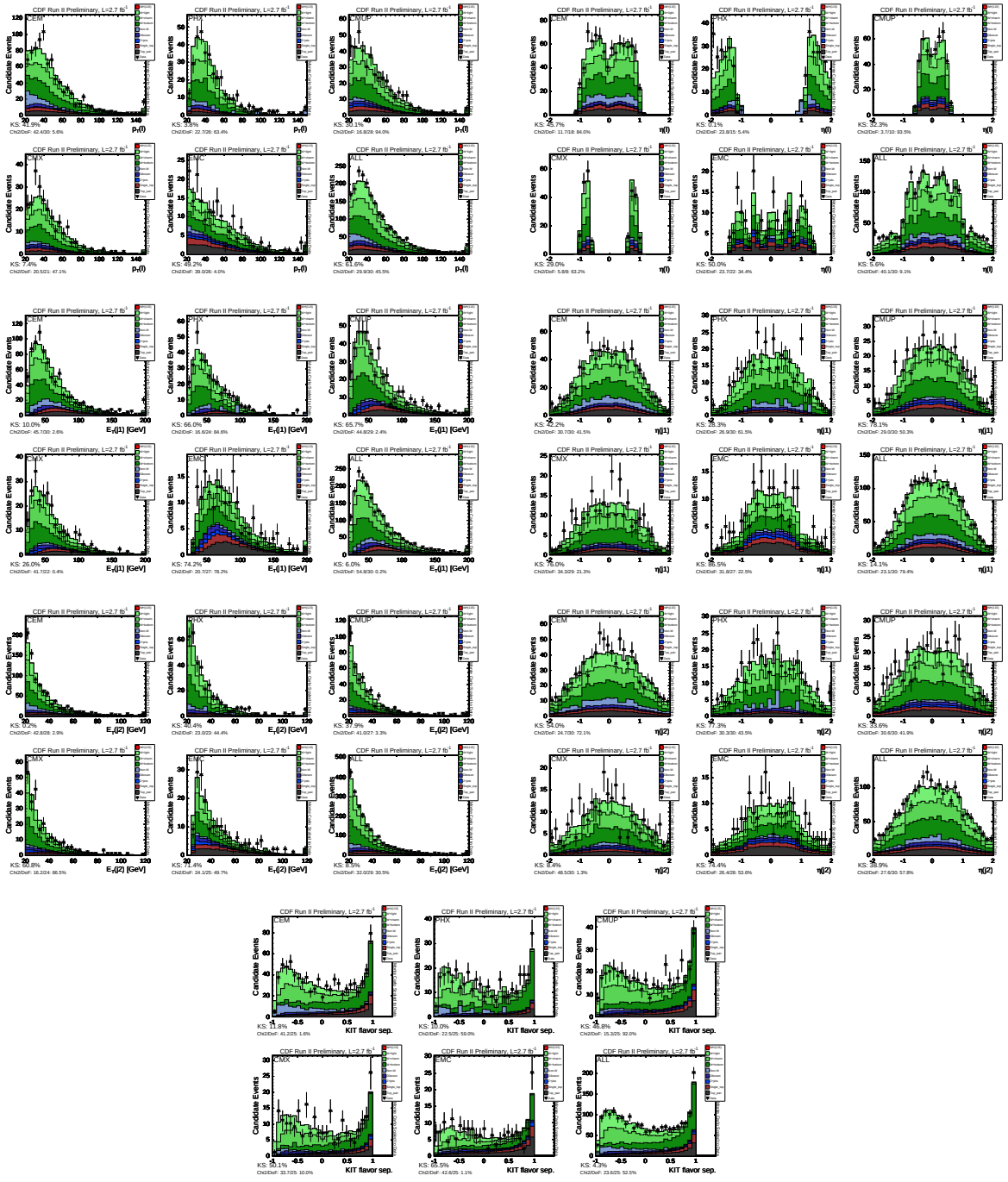

Figure 41: Comparison of the event probability input variables for Monte Carlo prediction and data in the 1tag signal sample. All sets of six plots show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

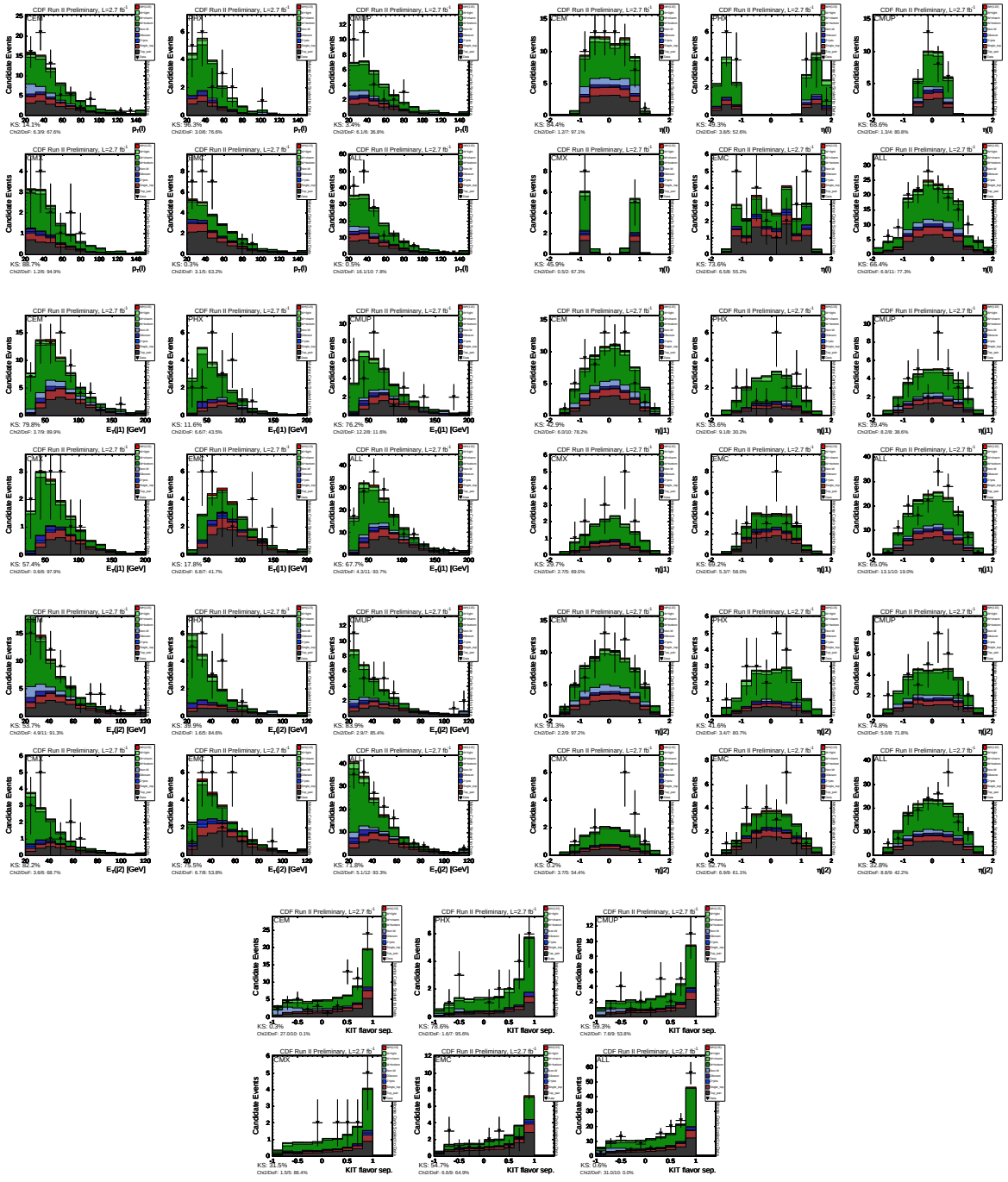

Figure 42: Comparison of the event probability input variables for Monte Carlo prediction and data in the 2tag signal sample. All sets of six plots show, from top left to bottom right: CEM; PHX; CMUP; CMX; newMuons; ALL.

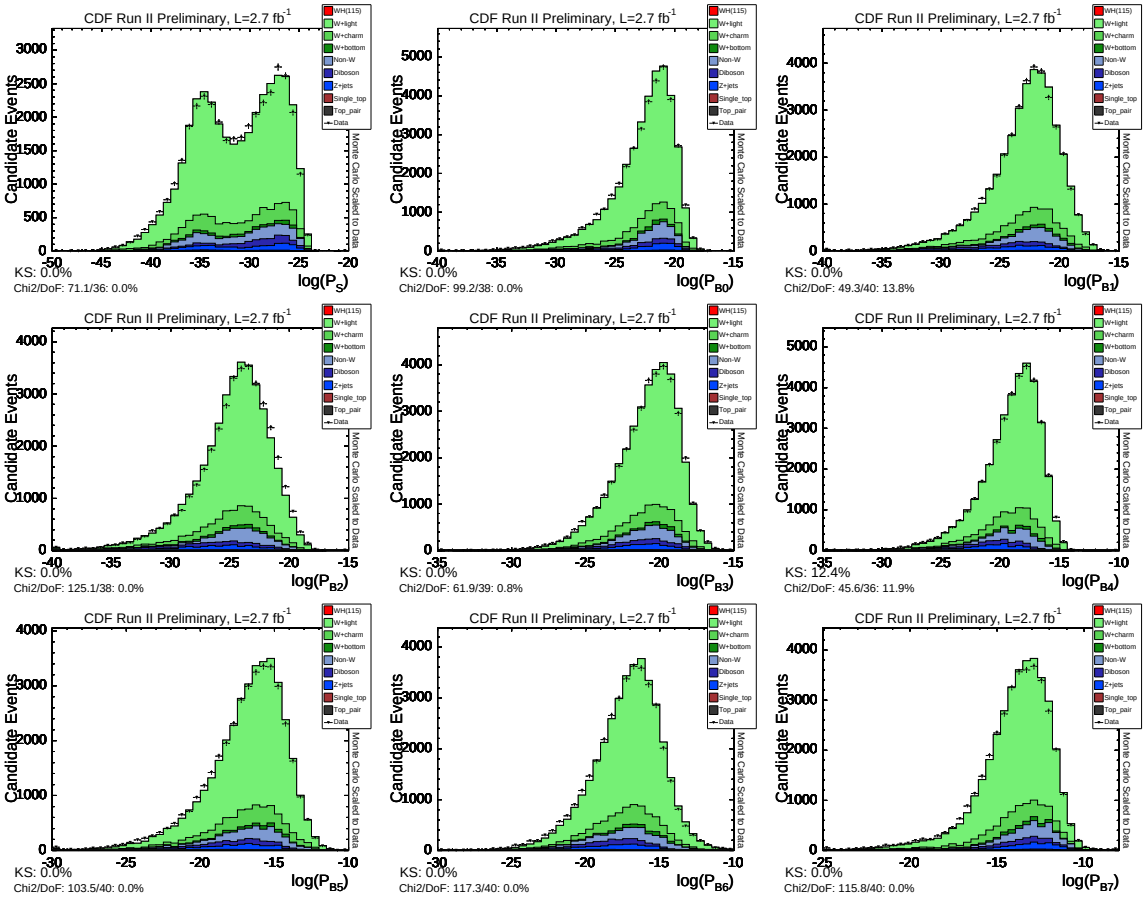

Figure 43: Distribution of the event probabilities in the untagged sample: From top left to bottom right: signal, s-channel, t-channel, t-channel2, W$b\bar{b}$-W$c\bar{c}$, Wc, Wgj, Wgg and $t\bar{t}$.

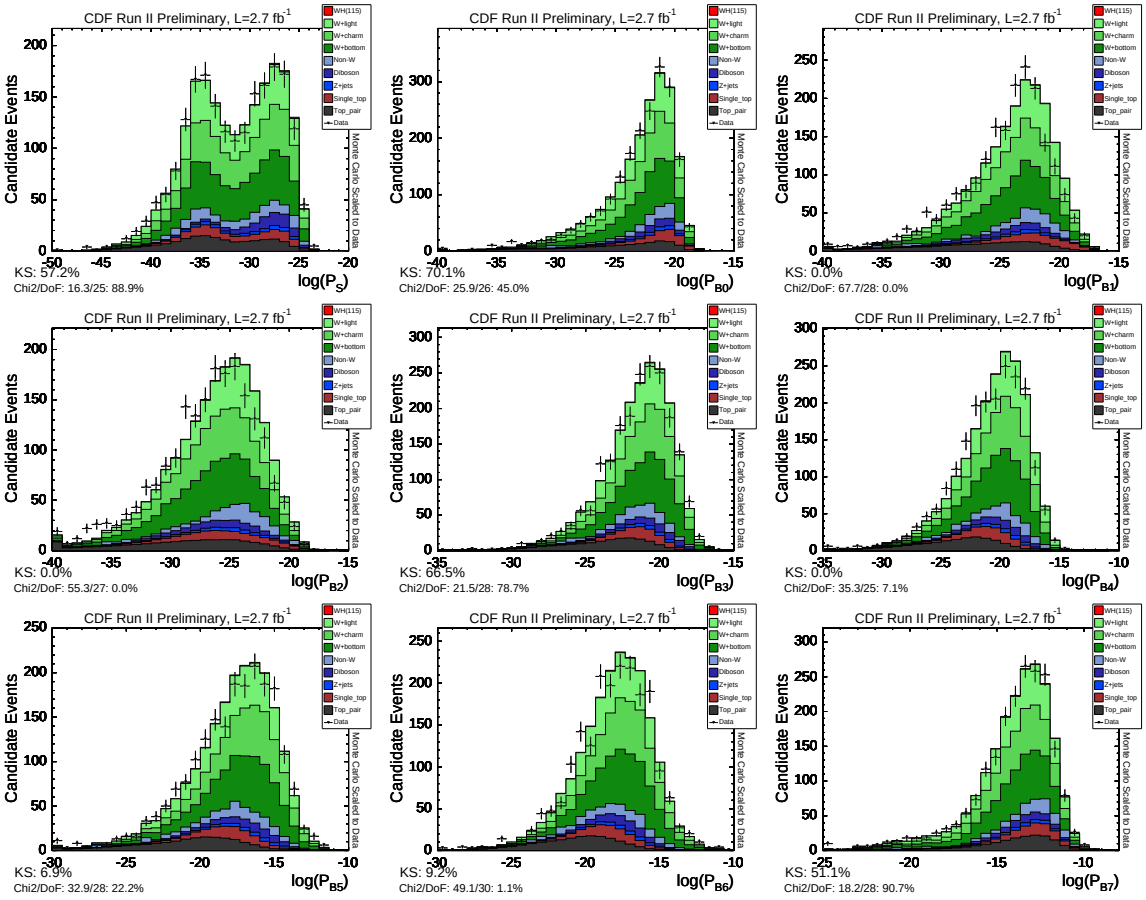

Figure 44: Distribution of the event probabilities in the single tag sample: From top left to bottom right: signal, s-channel, t-channel, t-channel2, W$b\bar{b}$-W$c\bar{c}$, Wc, Wgj, Wgg and $t\bar{t}$.

F Validation of the BDT Input Variables

In this section we cross-check whether the MC prediction of the BDT output represents the data well without looking at the WH candidate events. We choose the "taggable[8] but not tagged" side-band data. That is, we select W+2 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 WH signal region while it still represents a very similar kinematic event topology. Another advantage is that this sample has very little contribution from top (<0.5%). Data-Monte Carlo comparisons of the input variables are shown in this section, where we find generally good agreement between data and Monte Carlo prediction. Figures 45-47 and 48-50 show the data MC comparisons of all the BDT input variables in the untagged control regions and in the single tag signal region, respectively. The agreement is good, which assures us that the BDT output in Monte Carlo is well represented by data.

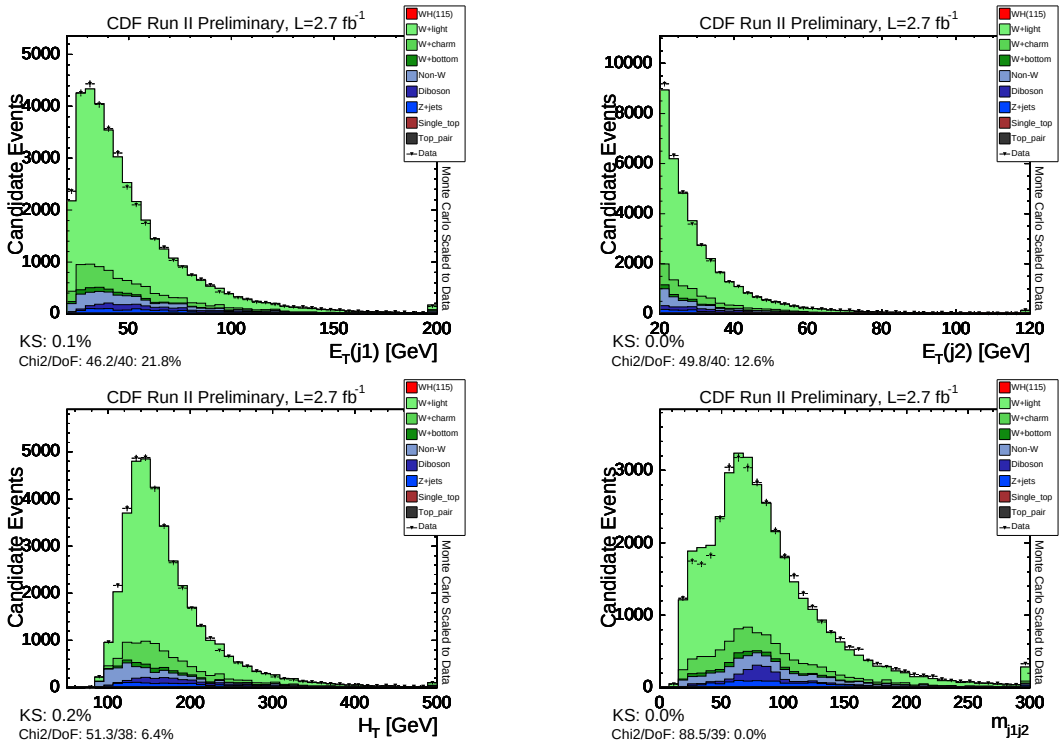

Figure 45: Distributions of the BDT input variables in the untagged sample.

[8]Event with at least one jet with, at least, two good SecVtx tracks

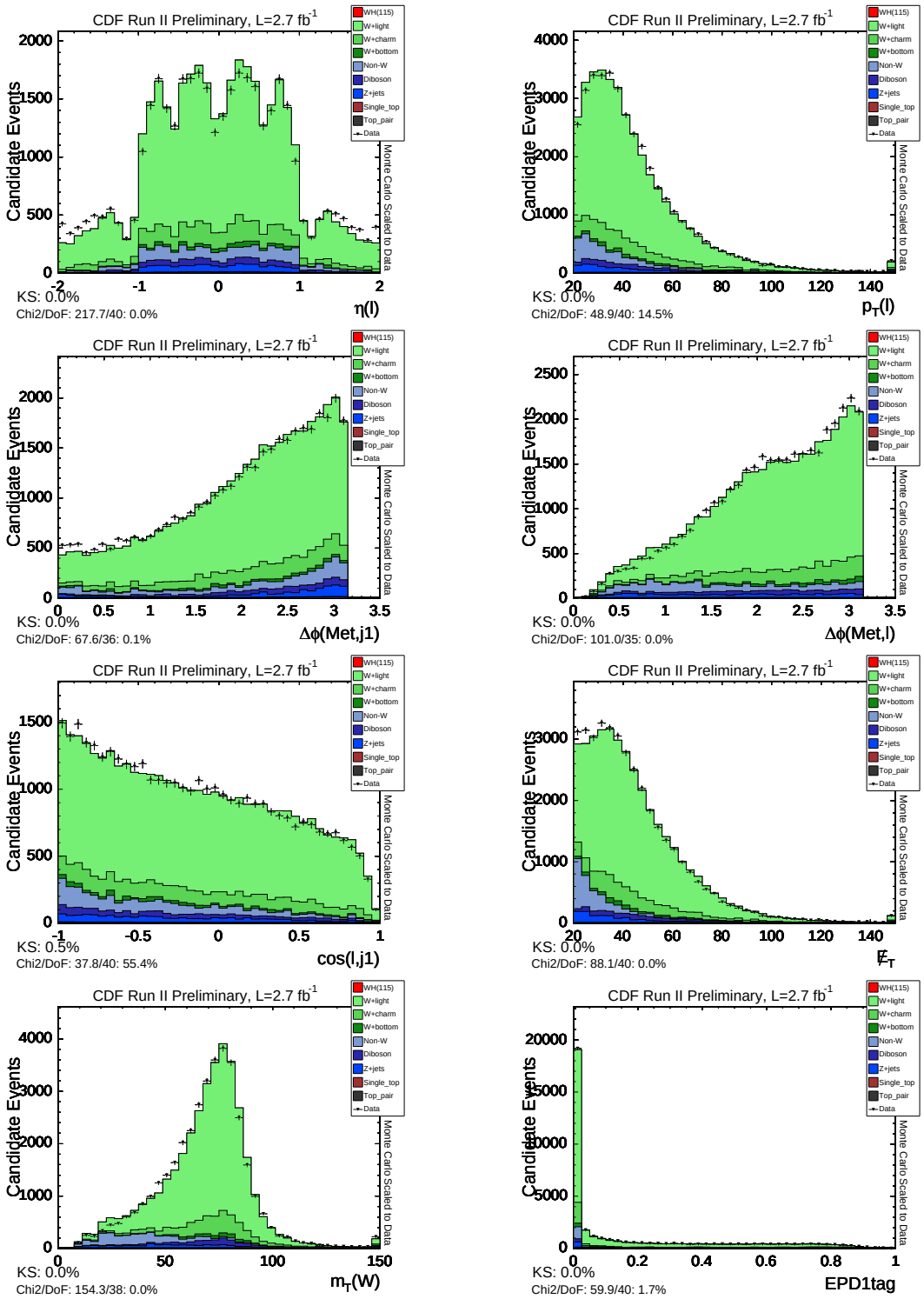

Figure 46: Distributions of the BDT input variables in the untagged sample.

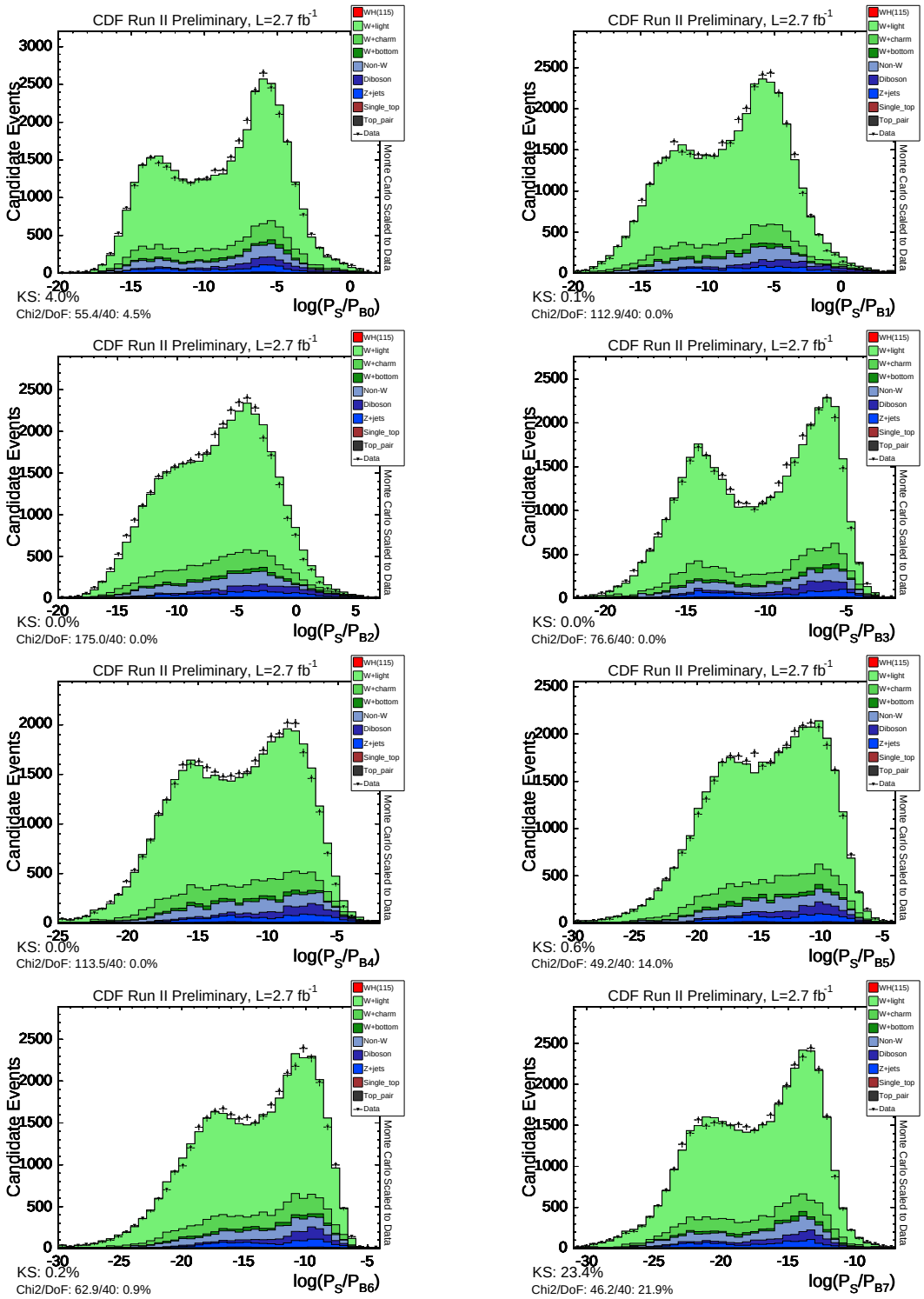

Figure 47: Distributions of the BDT input variables in the untagged sample.

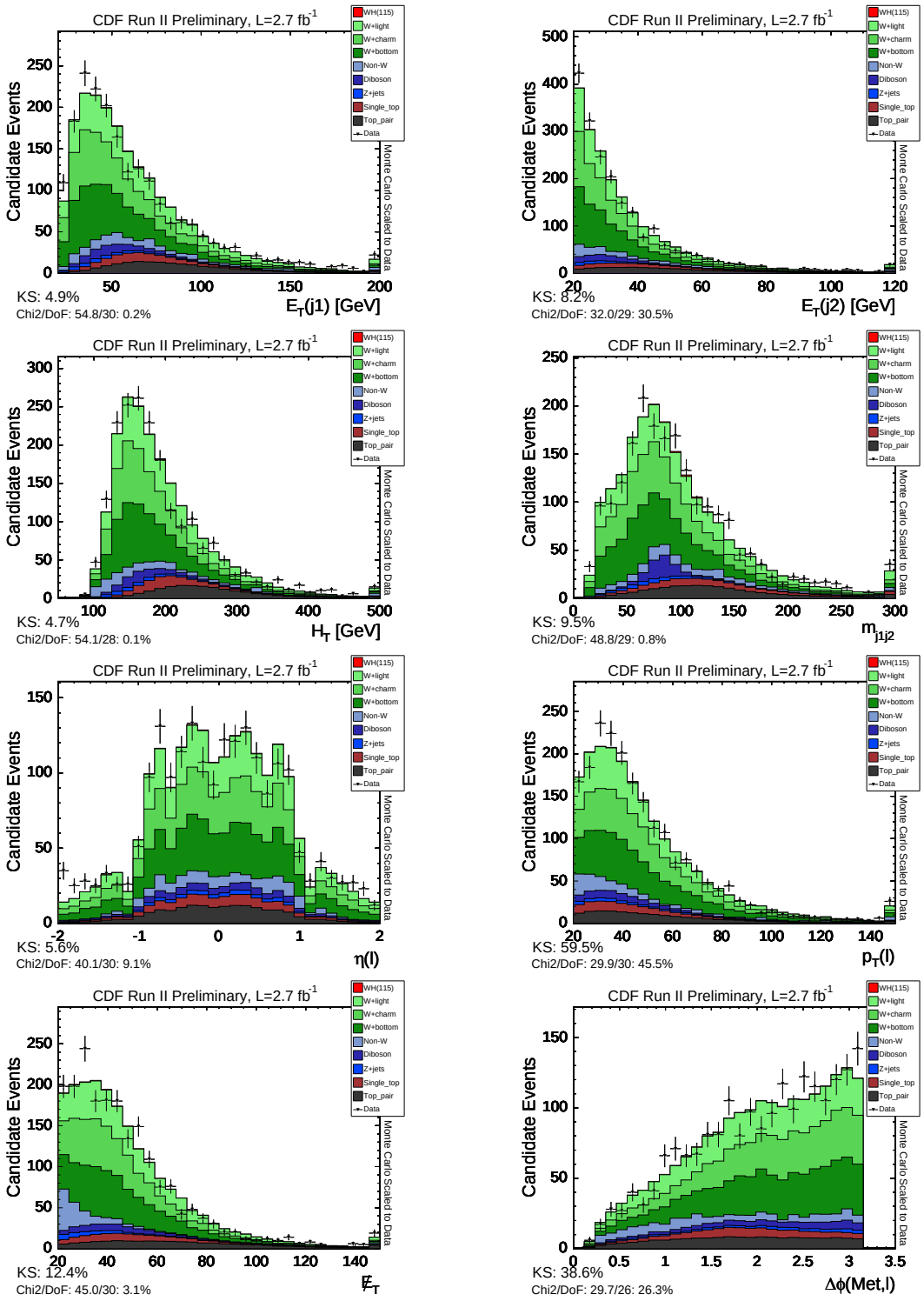

Figure 48: Distributions of the BDT input variables in the single tagged sample.

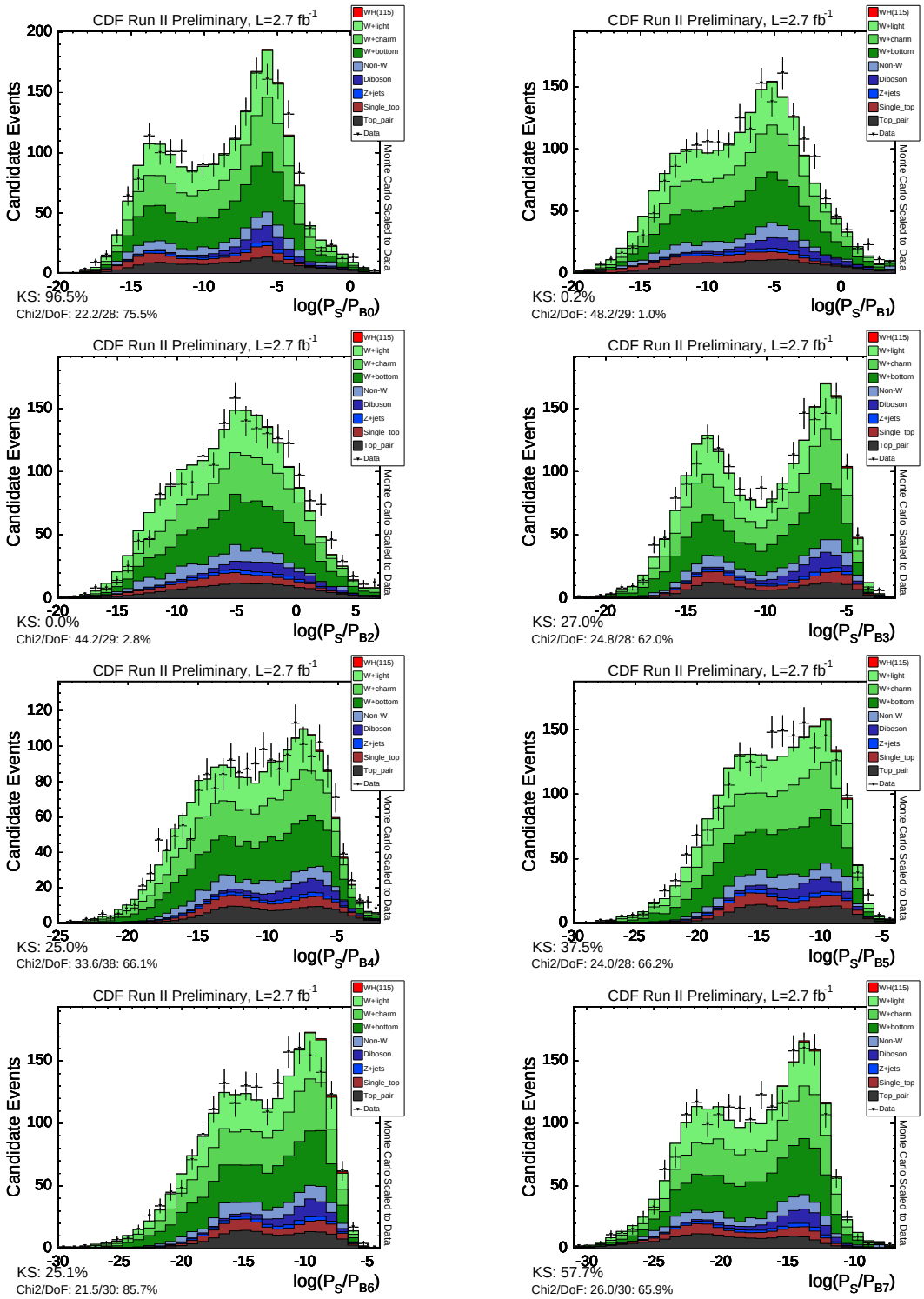

Figure 49: Distributions of the BDT input variables in the single tagged sample.

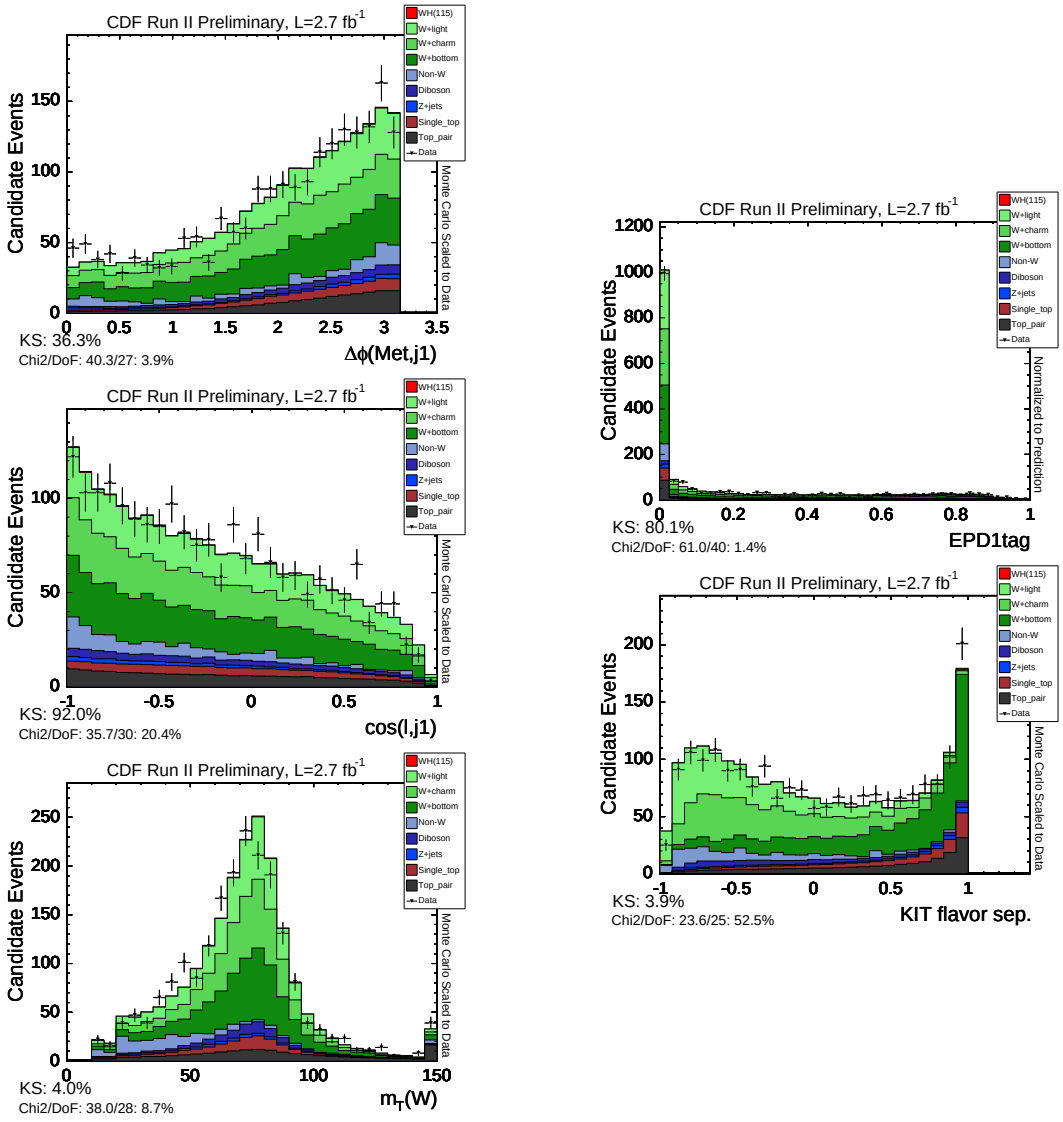

Figure 50: Distributions of the BDT input variables in the single tagged sample.

G BDT Templates and BDT Outputs

Figures 51-56 show the single and double tag templates for all Higgs masses. Figs. 57-58 (59-60) show the single (double) tag optimized BDT output applied over the untagged sample. Figures 61-62 show the single and double tag optimized BDT outputs applied over the single and double tag sample, respectively.

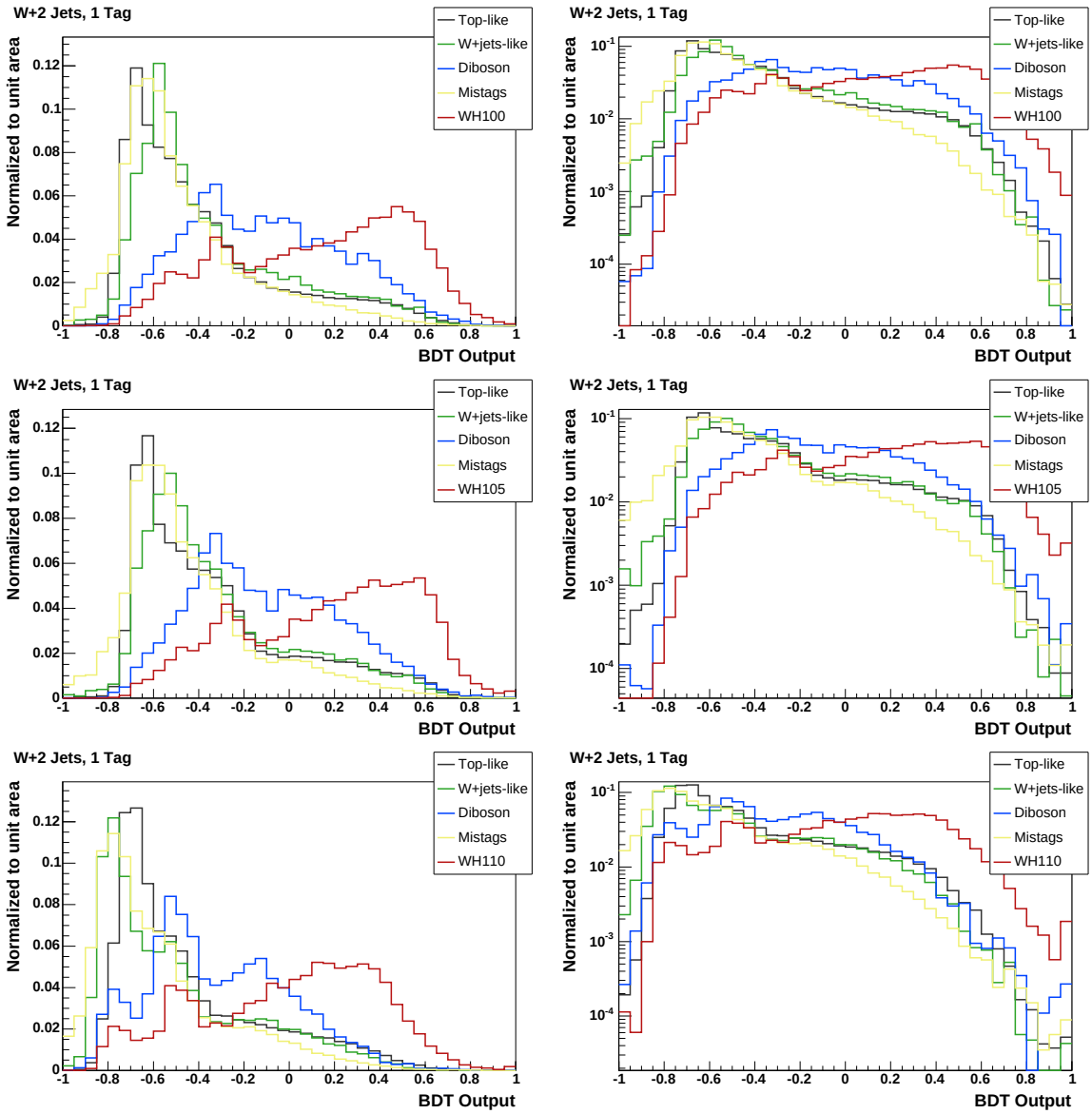

Figure 51: Left (Right): Single tag BDT templates for Higgs masses of 100, 105 and 110 GeV/c^2 in linear (log) scale.

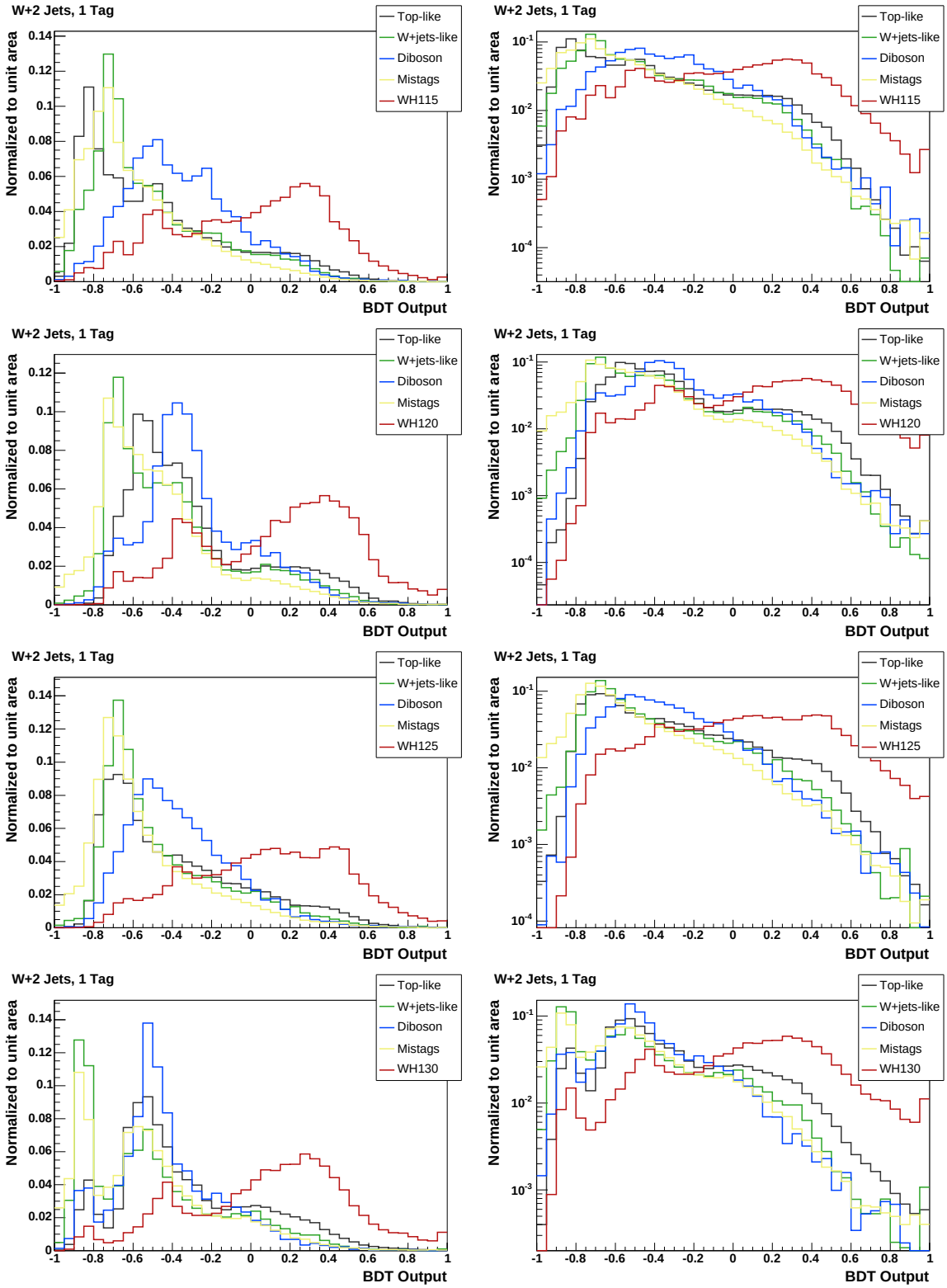

Figure 52: Left (Right): Single tag BDT templates for Higgs masses of 115, 120, 125 and 130 GeV/c^2 in linear (log) scale.

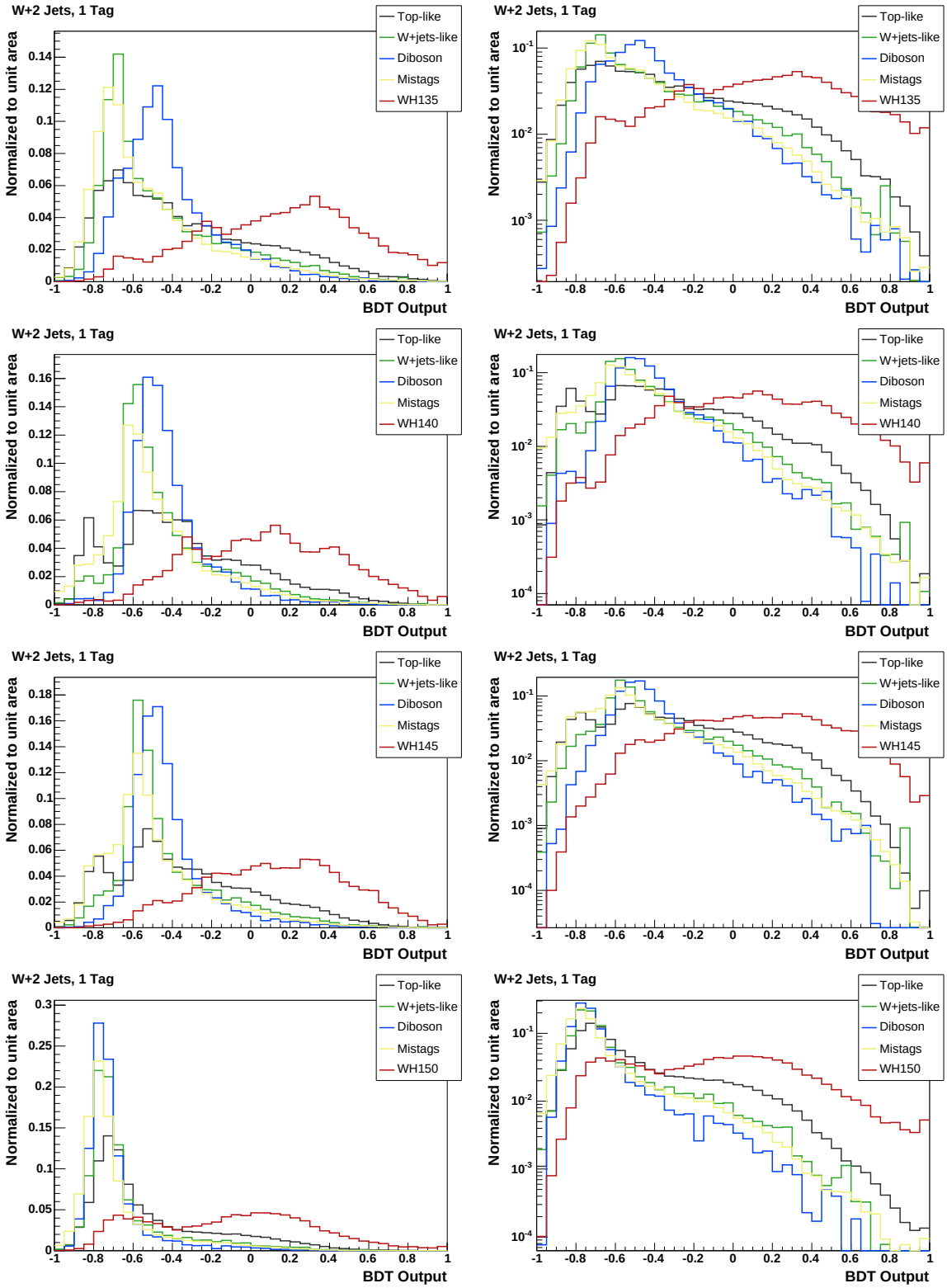

Figure 53: Left (Right): Single tag BDT templates for Higgs masses of 135, 140, 145 and 150 GeV/c^2 in linear (log) scale.

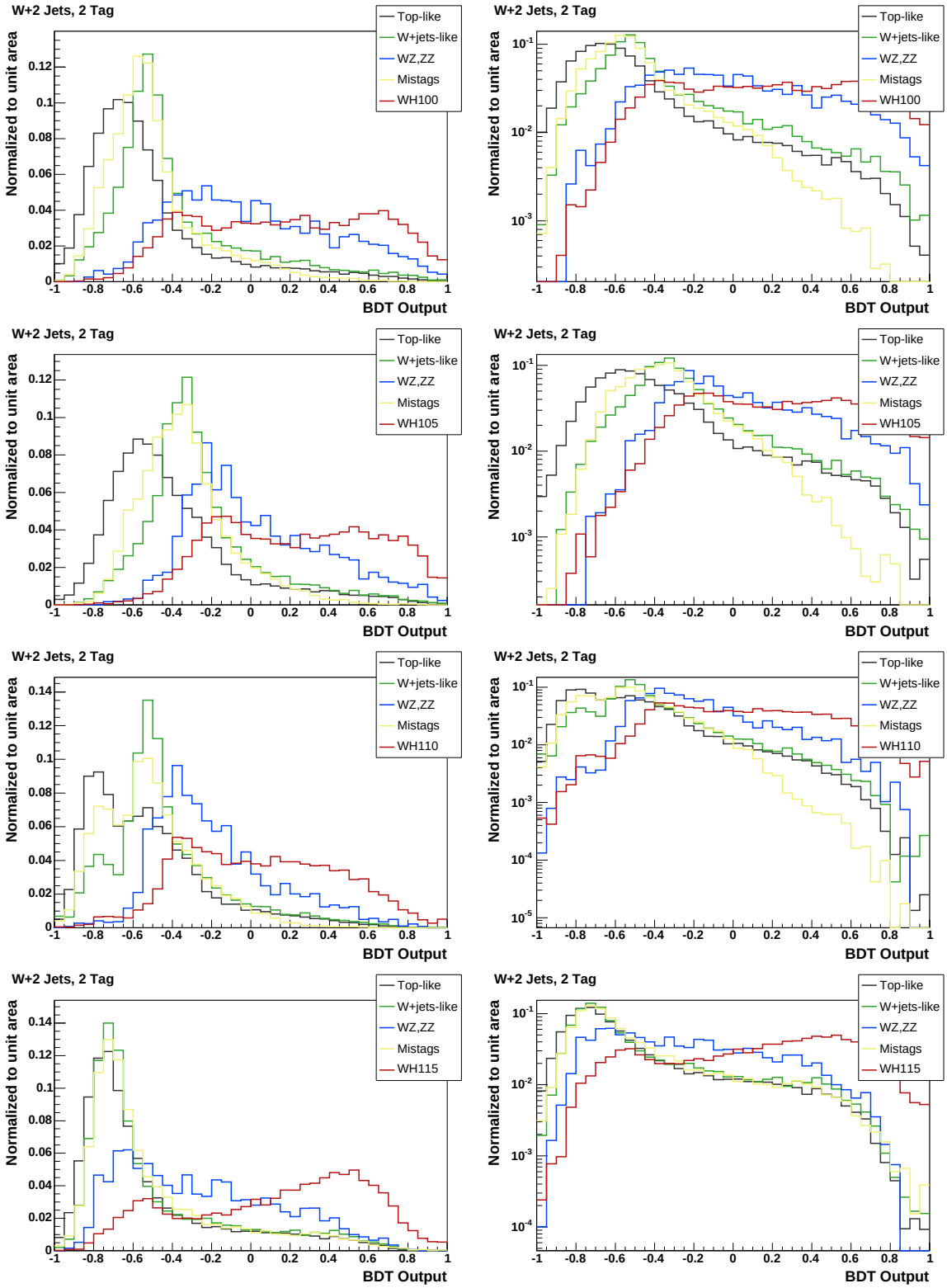

Figure 54: Left (Right): Double tag BDT templates for Higgs masses of 100, 105, 110 and 115 GeV/c^2 in linear (log) scale.

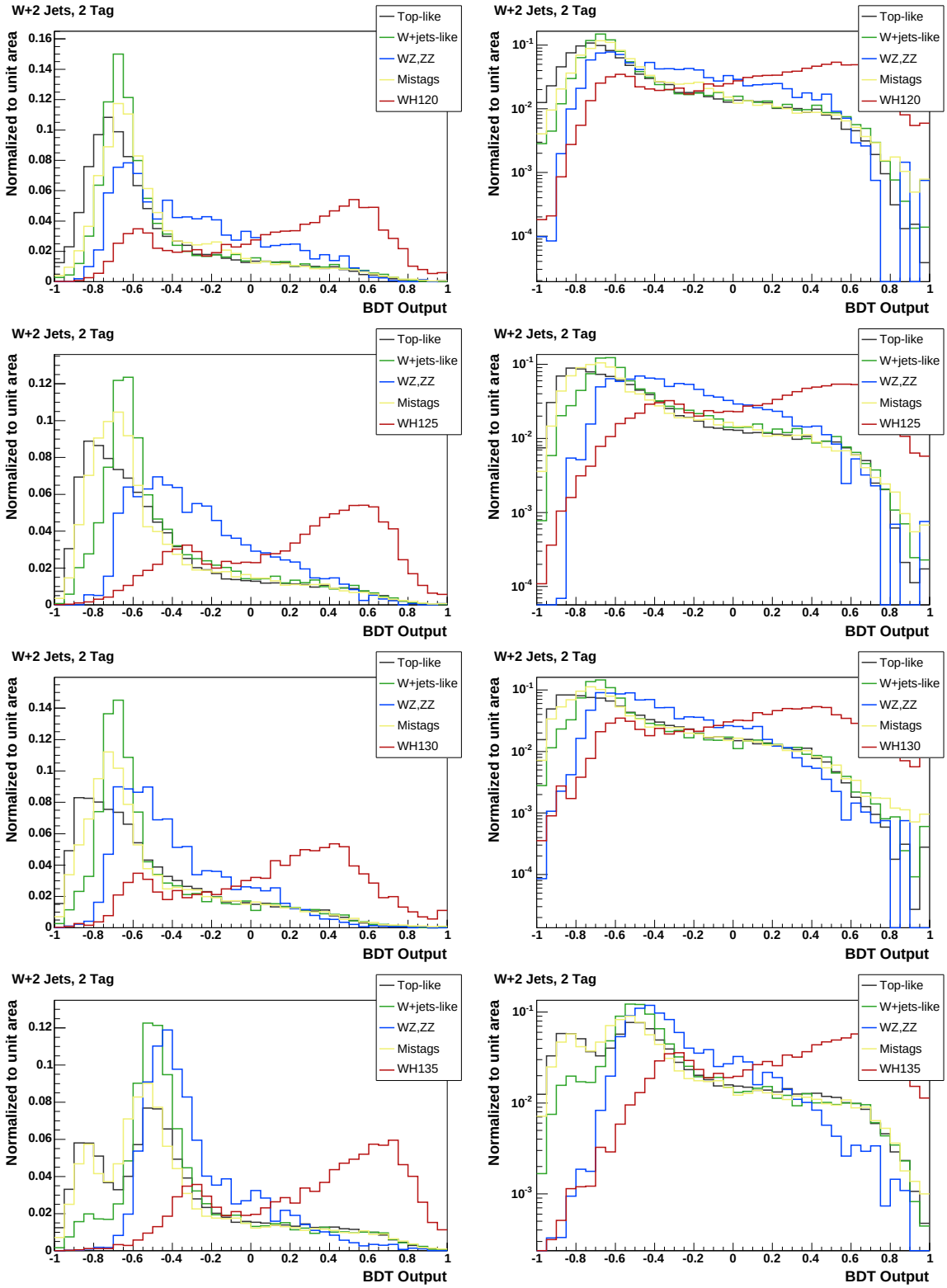

Figure 55: Left (Right): Double tag BDT templates for Higgs masses of 120, 125, 130 and 135 GeV/c^2 in linear (log) scale.

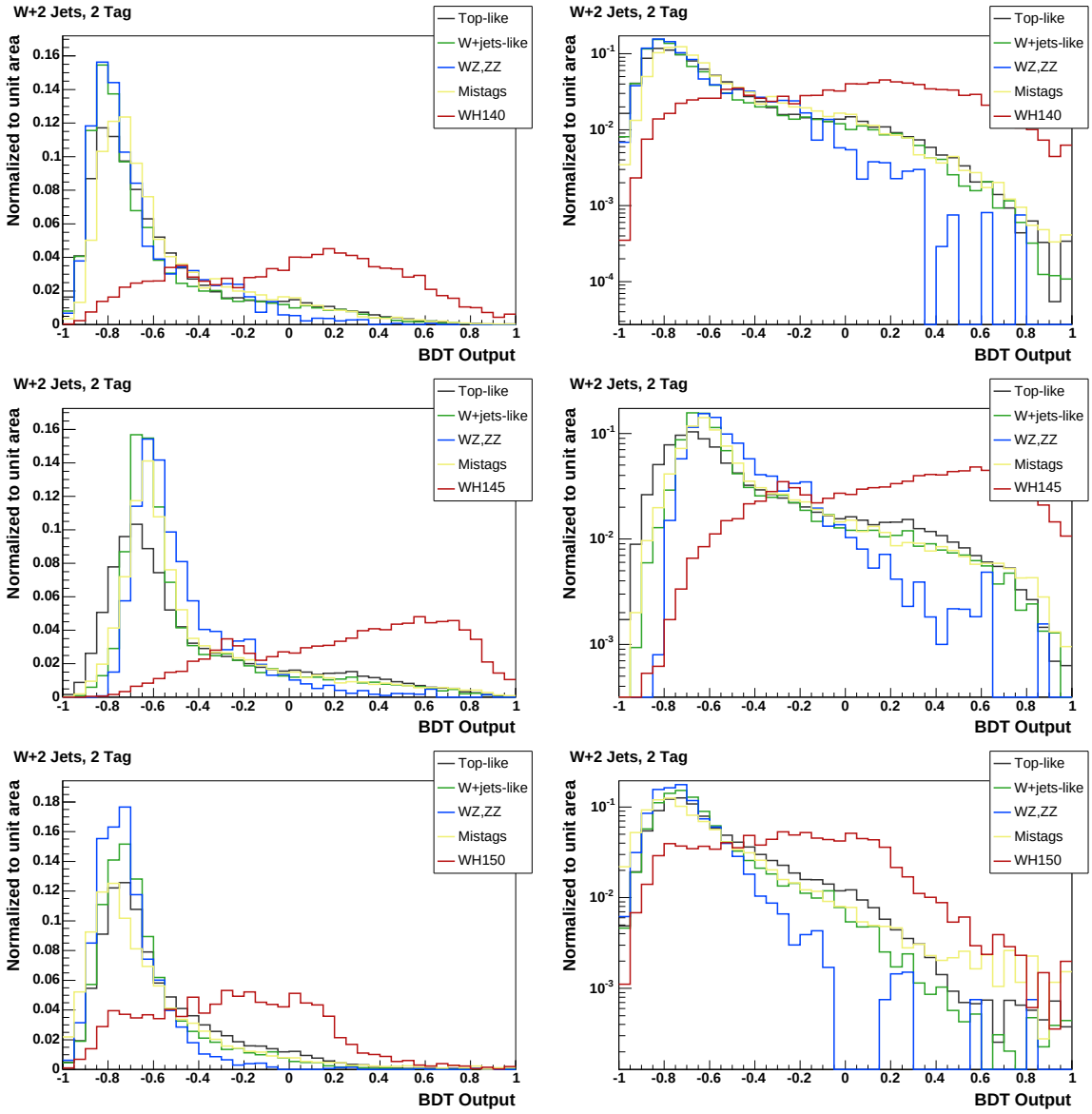

Figure 56: Left (Right): Double) tag BDT templates for Higgs masses of 140, 145 and 150 GeV/c^2 in linear (log) scale.

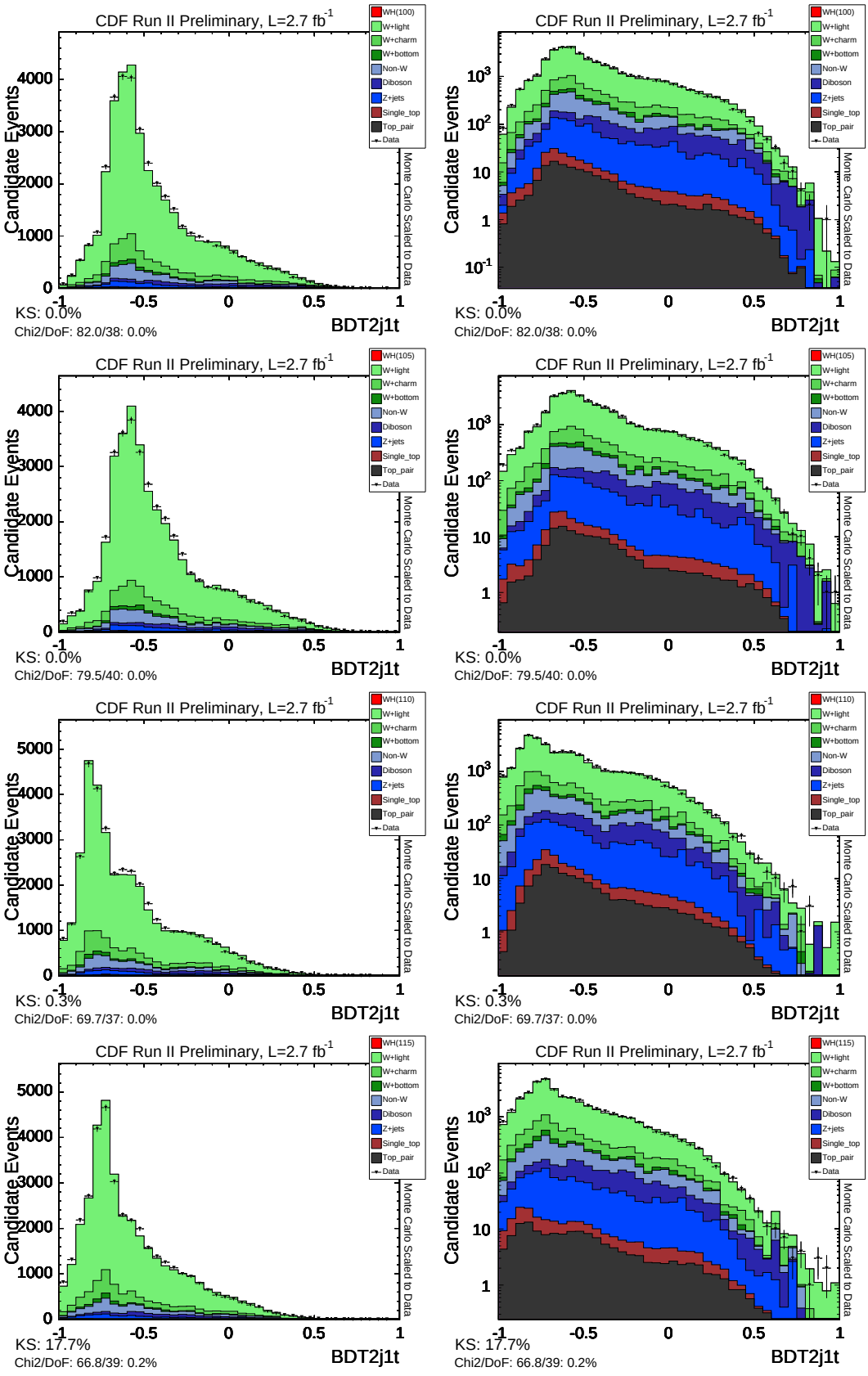

Figure 57: Single tag optimized BDT applied in the untagged sample for Higgs masses of 100, 105, 110, and 115 GeV/c^2. Left (right) column is in linear (log) scale.

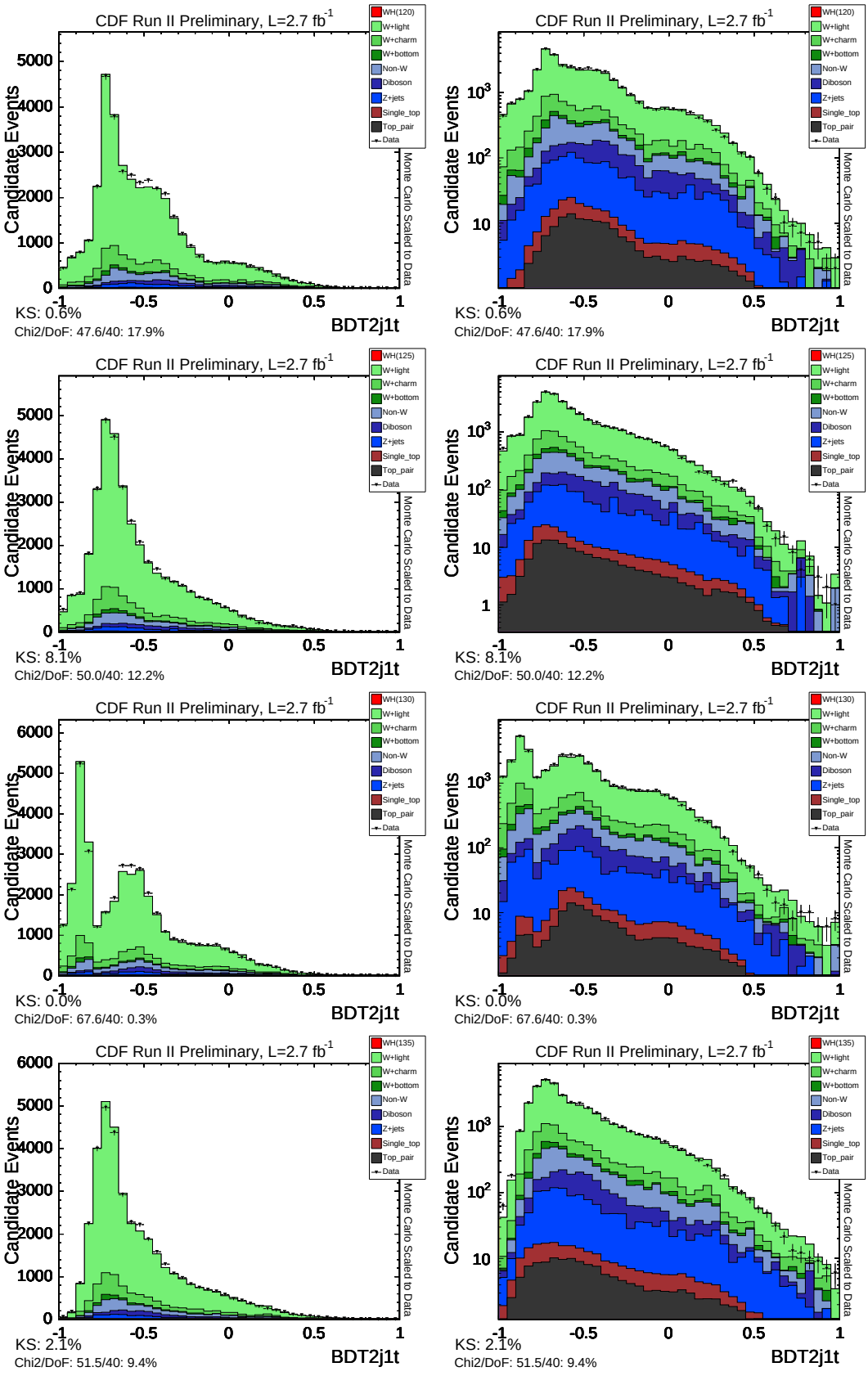

Figure 58: Single tag optimized BDT applied in the untagged sample for Higgs masses of 120, 125, 130 and 135 GeV/c^2. Left (right) column is in linear (log) scale.

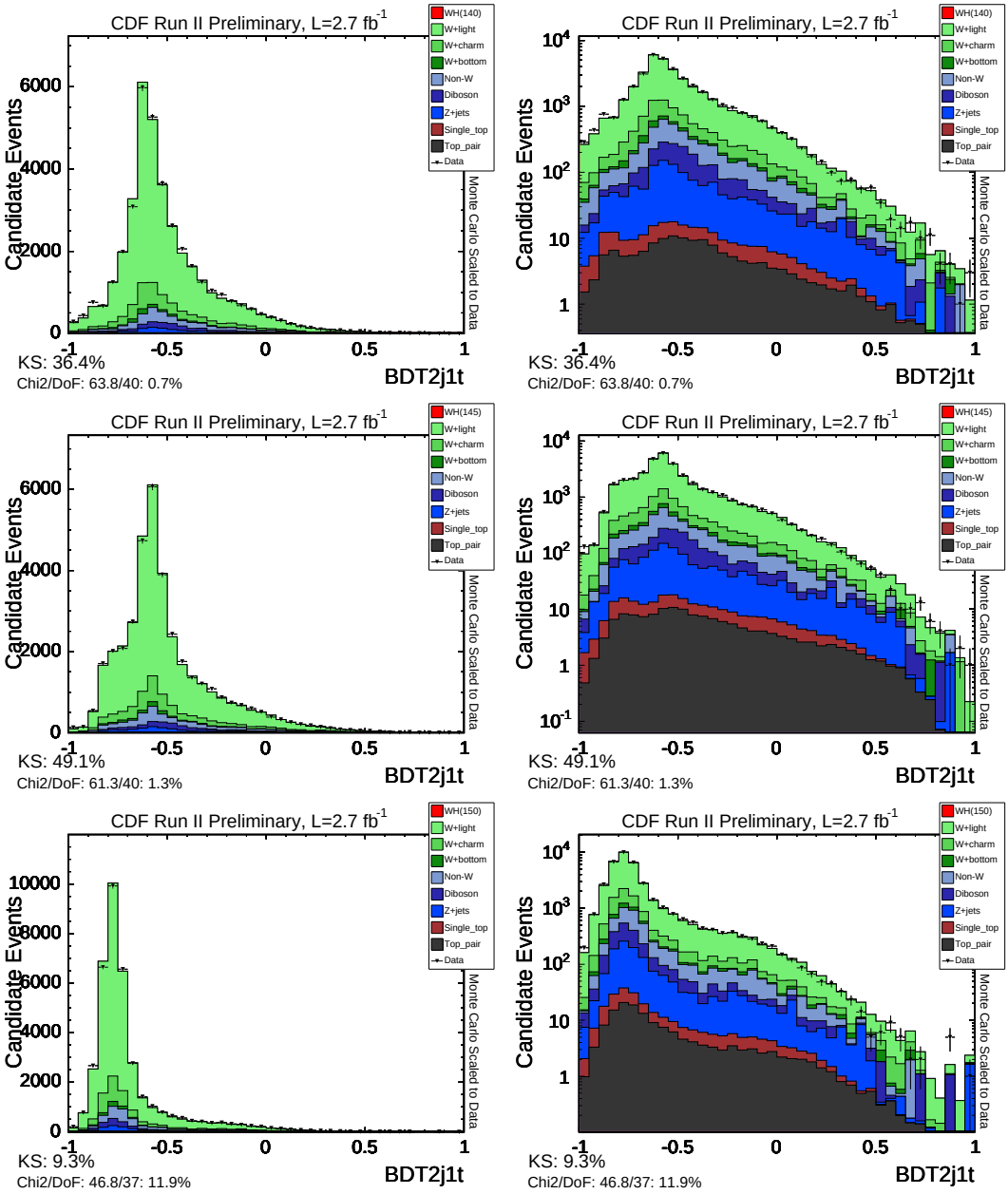

Figure 59: Single tag optimized BDT applied in the untagged sample for Higgs masses of 140, 145 and 150 GeV/c^2. Left (right) column is in linear (log) scale.

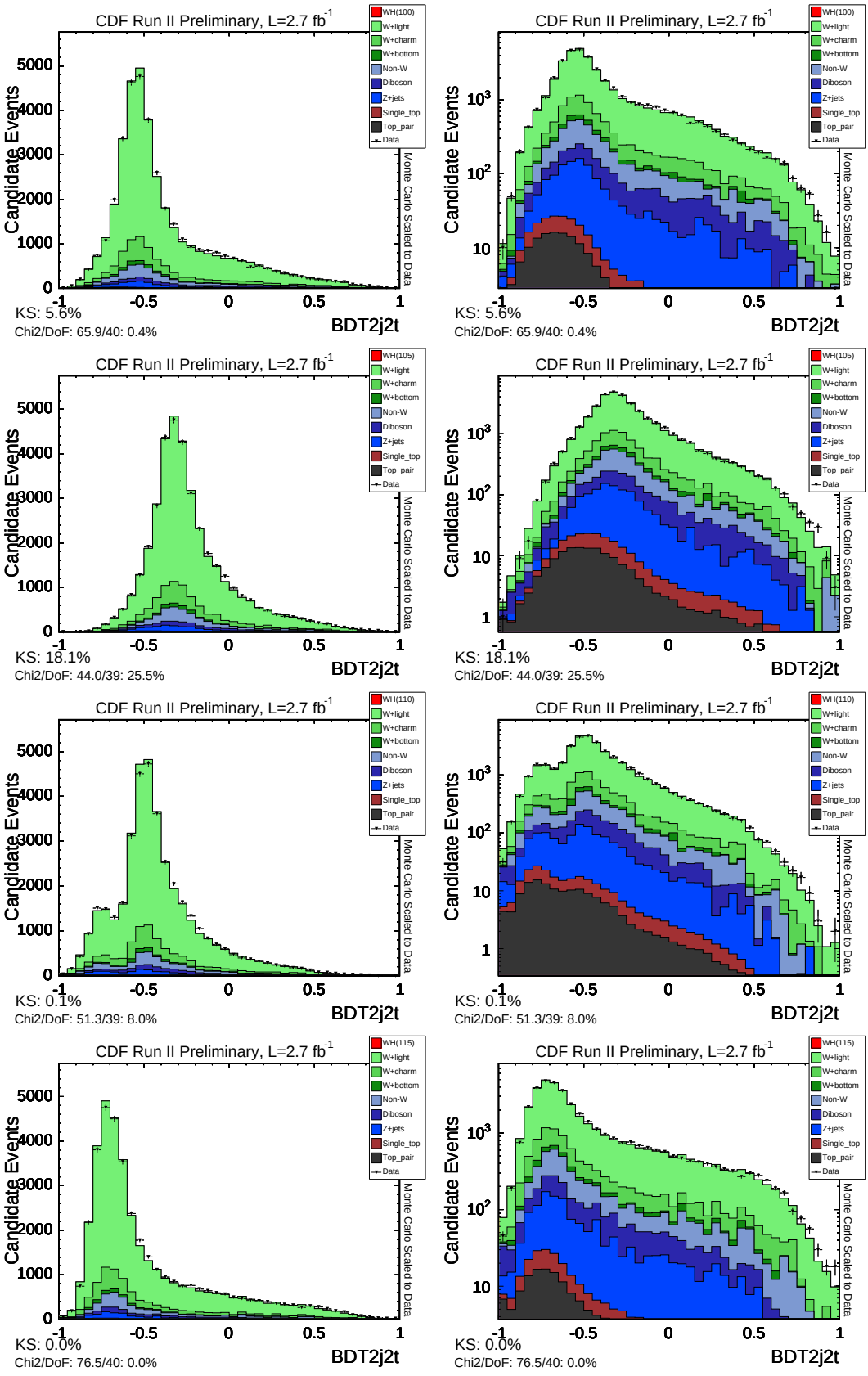

Figure 60: Double tag optimized BDT applied in the untagged sample for Higgs masses of 100, 105, 110 and 115 GeV/c^2. Left (right) column is in linear (log) scale.

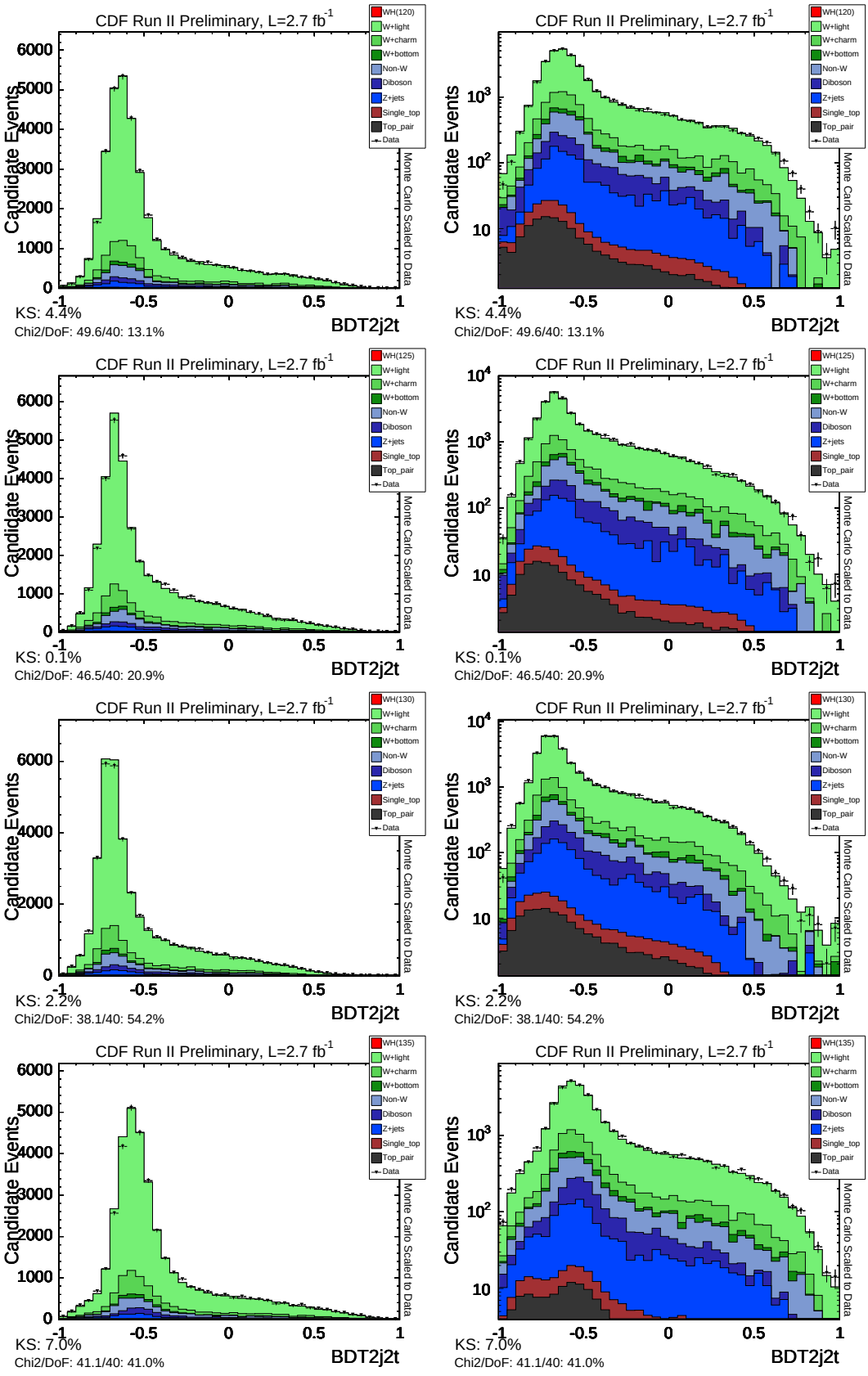

Figure 61: Double tag optimized BDT applied in the untagged sample for Higgs masses of 120, 125, 130 and 135 GeV/c^2. Left (right) column is in linear (log) scale.

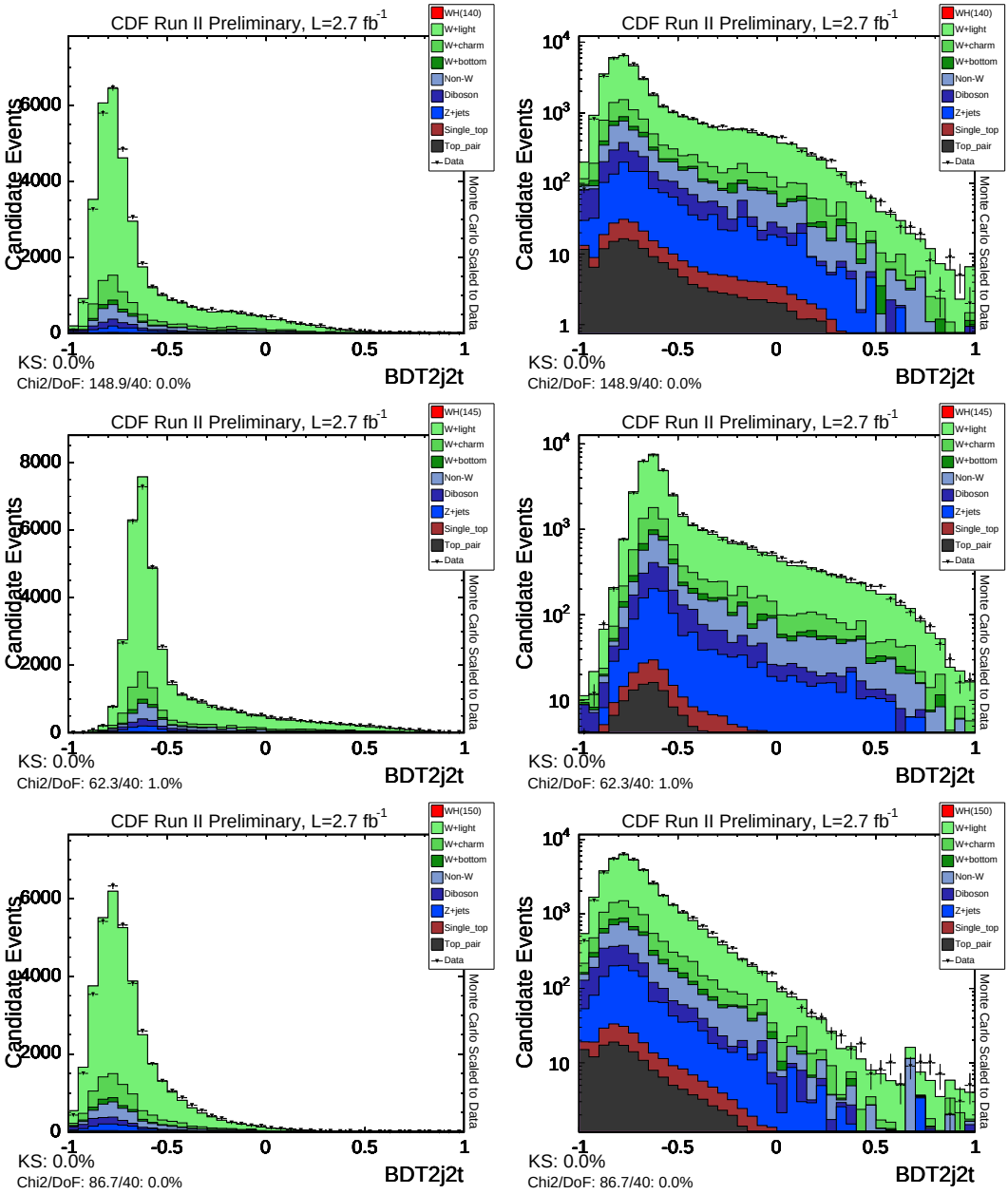

Figure 62: Double tag optimized BDT applied in the untagged sample for Higgs masses of 140, 145 and 150 GeV/c^2. Left (right) column is in linear (log) scale.

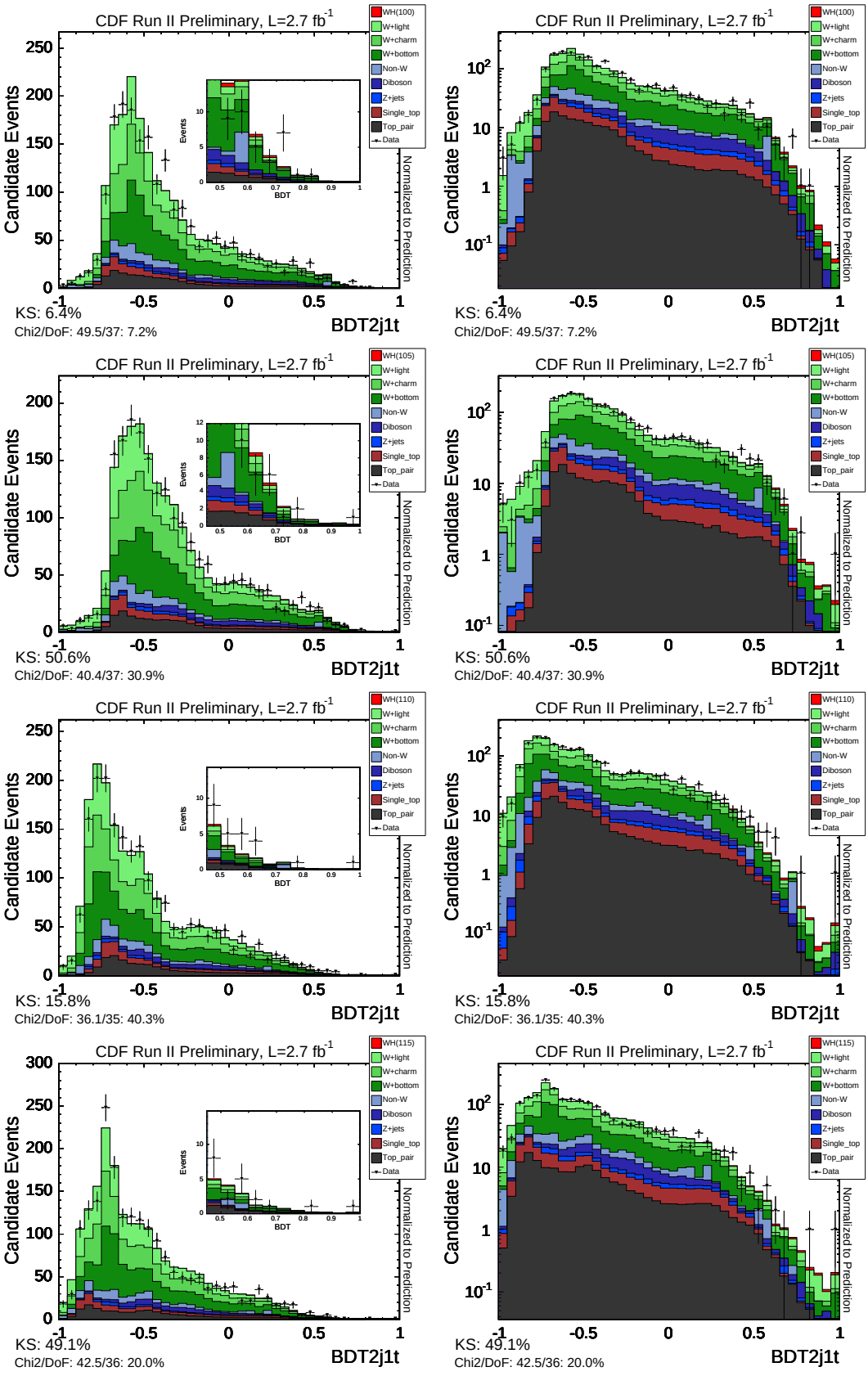

Figure 63: Sinlge tag optimized BDT applied in the single tag sample for Higgs masses of 100, 105, 110 and 115 GeV/c^2. Left (right) is in linear (log) scale.

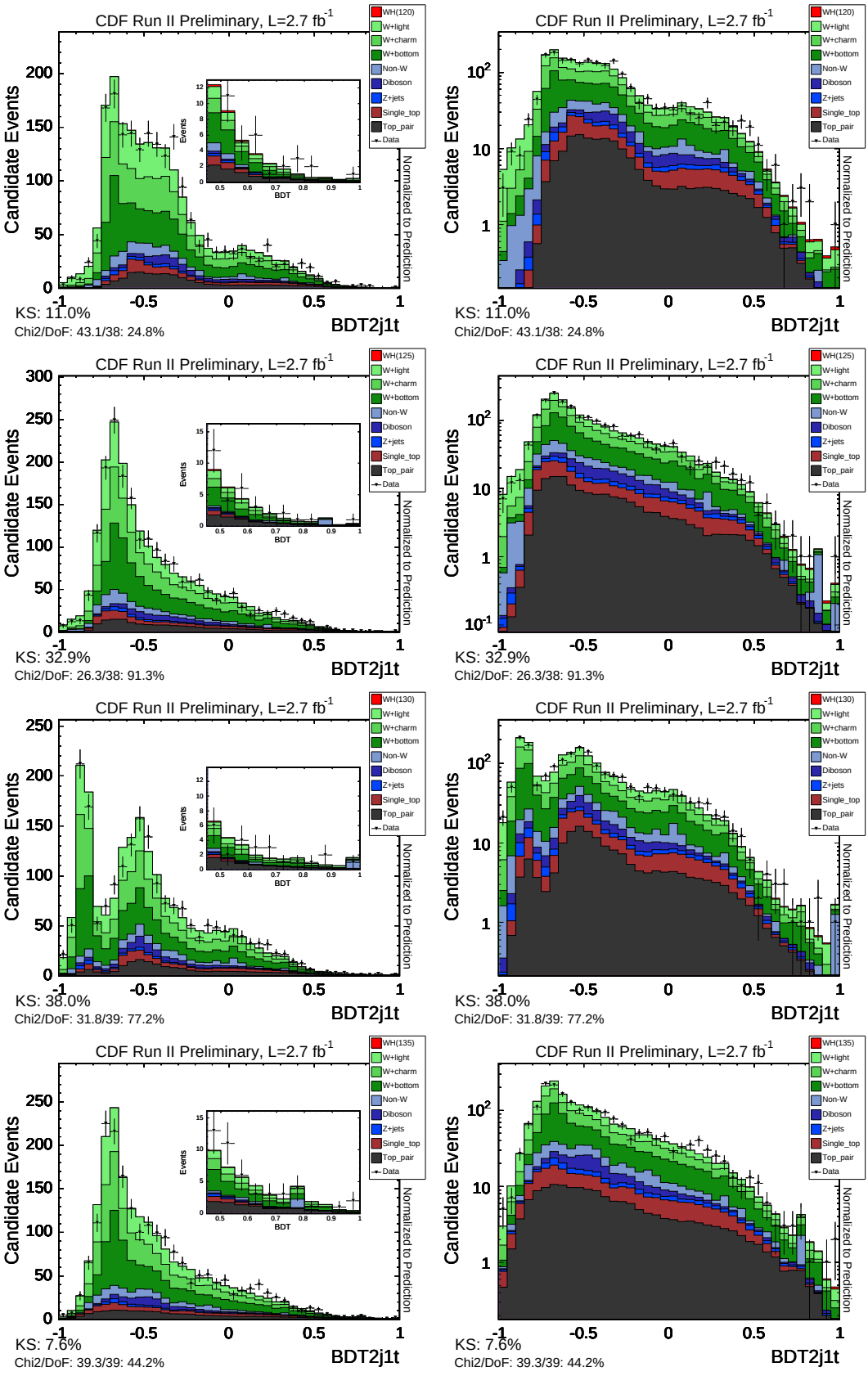

Figure 64: Sinlge tag optimized BDT applied in the single tag sample for Higgs masses of 120, 125, 130 and 135 GeV/c^2. Left (right) is in linear (log) scale.

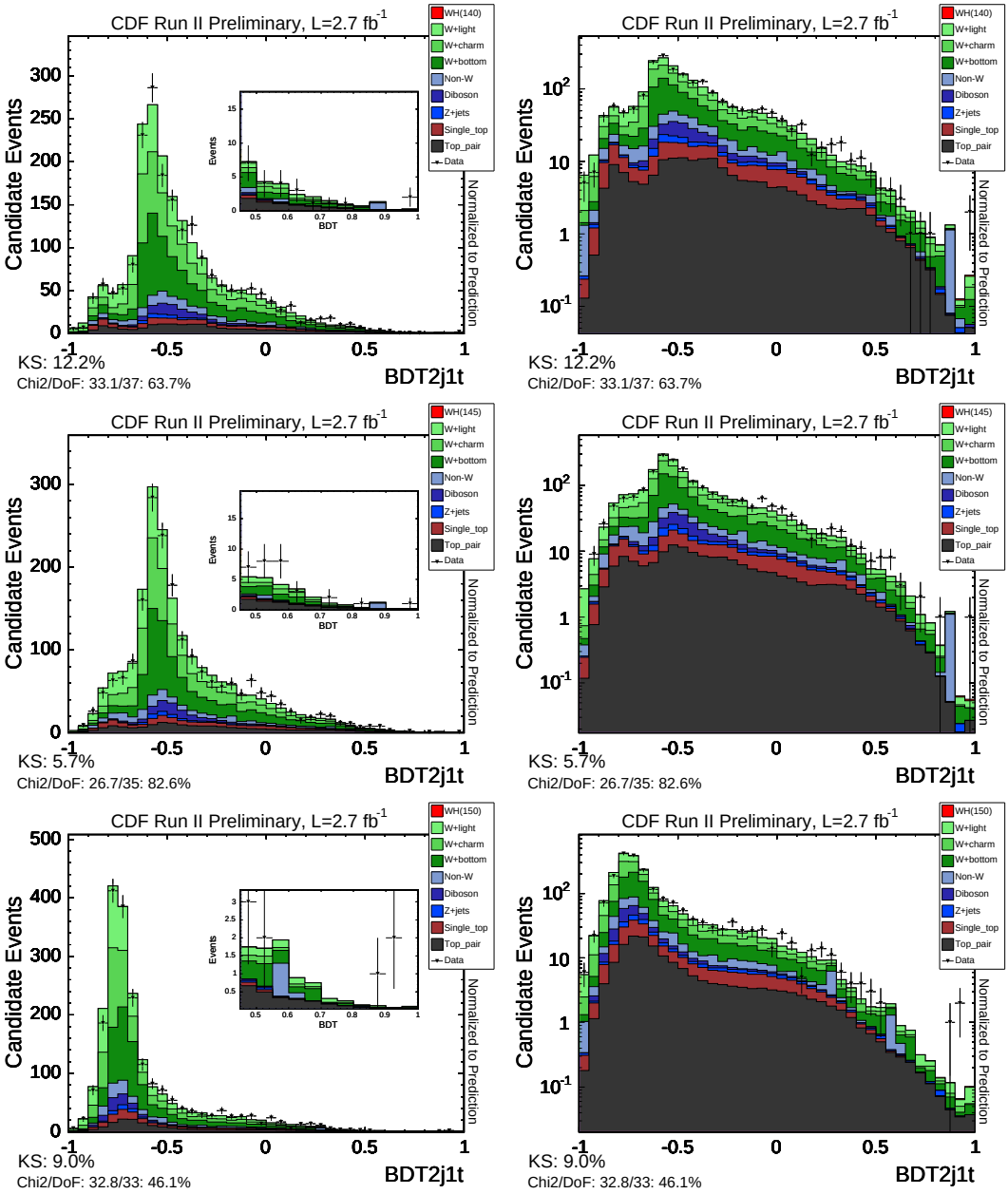

Figure 65: Sinlge tag optimized BDT applied in the single tag sample for Higgs masses of 140, 145 and 150 GeV/c^2. Left (right) is in linear (log) scale.

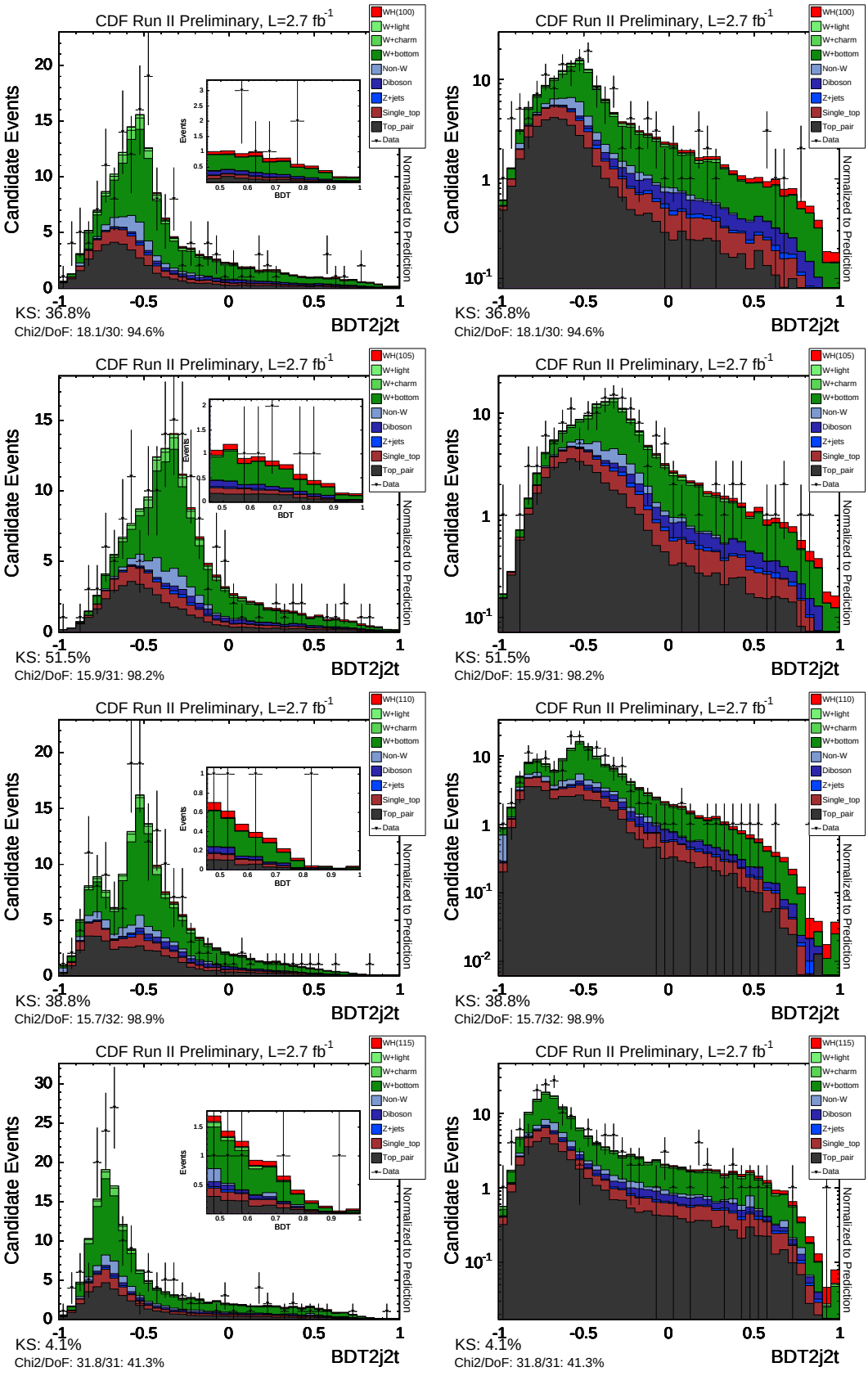

Figure 66: Double tag optimized BDT applied in the double tag sample for Higgs masses of 100, 105, 110 and 115 GeV/c^2. Left (right) is in linear (log) scale.

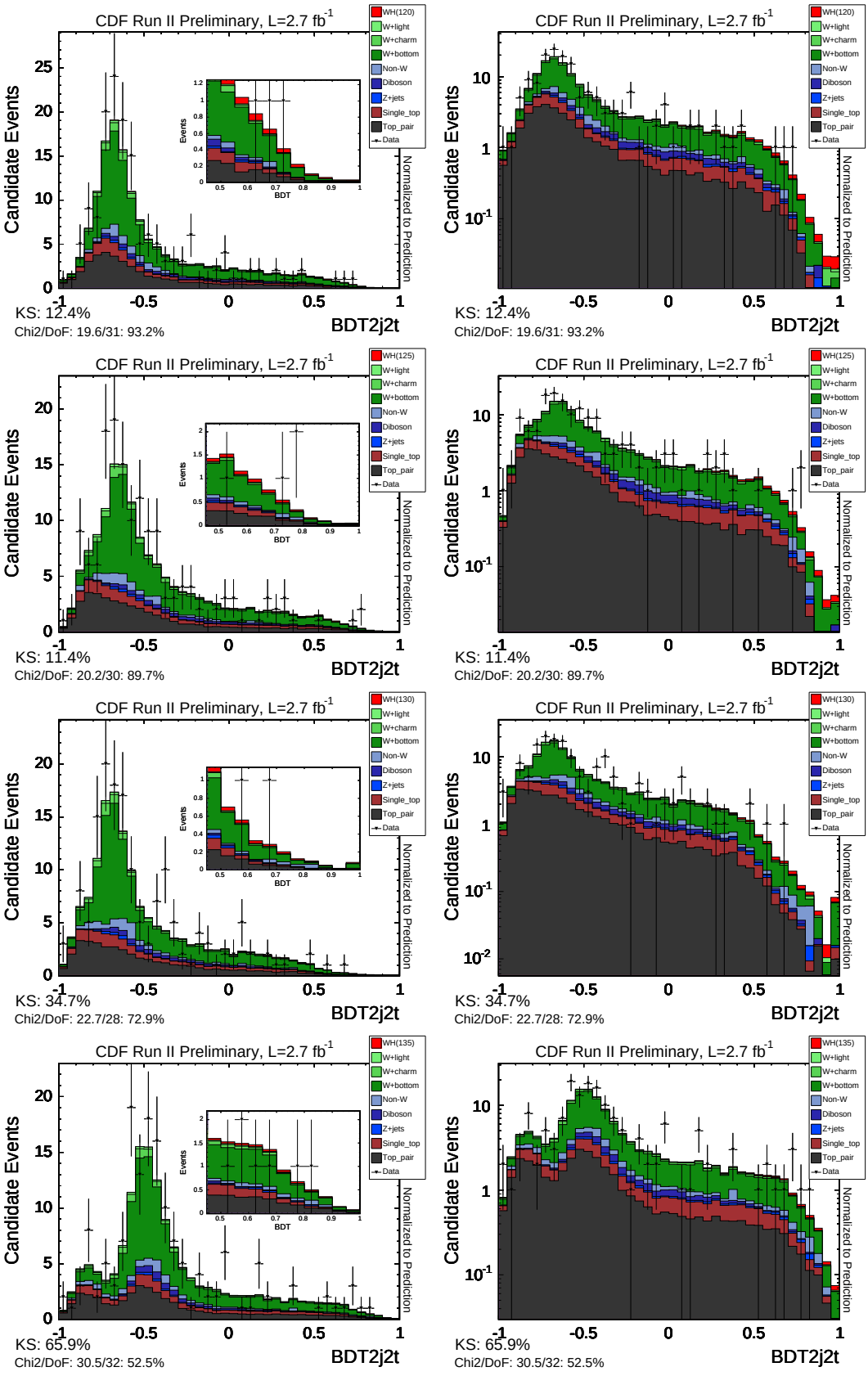

Figure 67: Double tag optimized BDT applied in the double tag sample for Higgs masses of 120, 125, 130 and 135 GeV/c^2. Left (right) is in linear (log) scale.

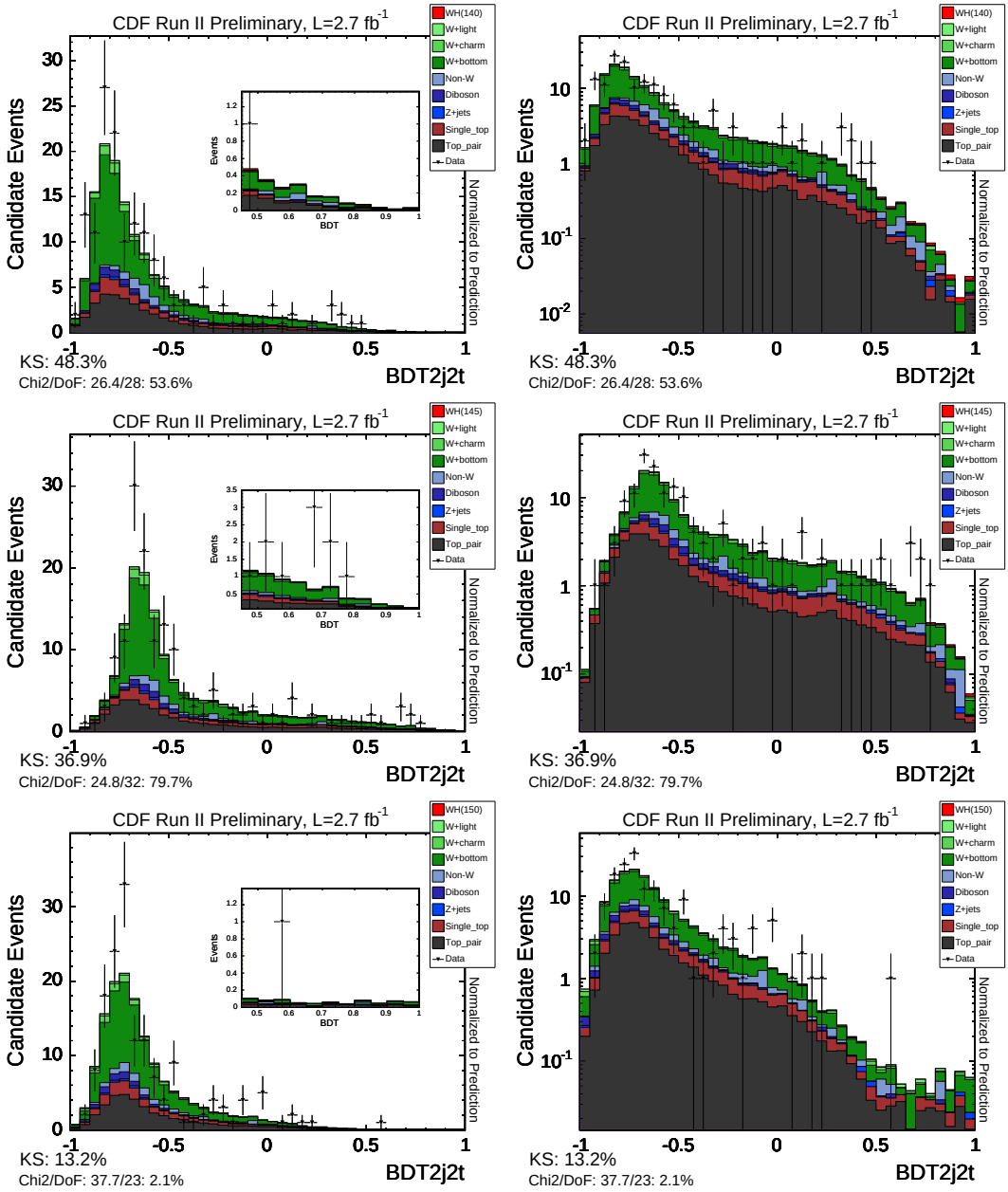

Figure 68: Double tag optimized BDT applied in the double tag sample for Higgs masses of 140, 145 and 150 GeV/c^2. Left (right) is in linear (log) scale.

H Validation of p14-p17: data-to-data comparison

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 for new data (p14-p17) have been compared with old data (prior to period 14). Figures 69-86 show the new data (red) compared with the old data scaled to the new data (black). Reasonable agreement is observed in all distributions studied.

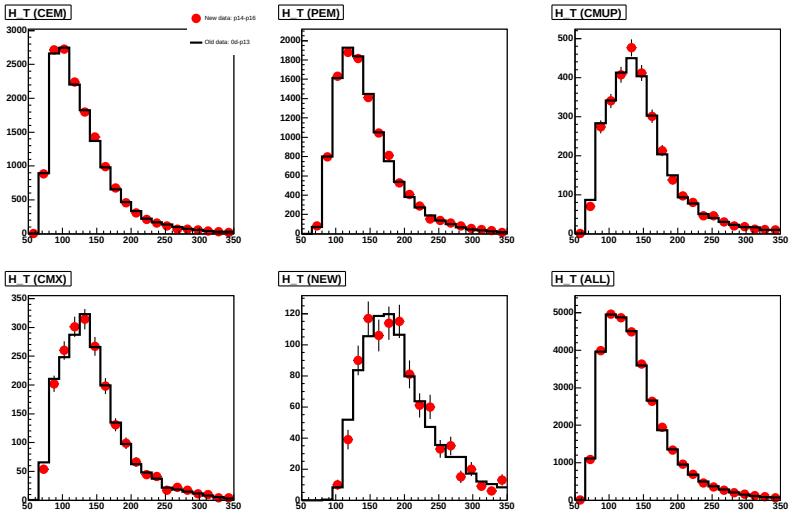

Figure 69: H_T distribution in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

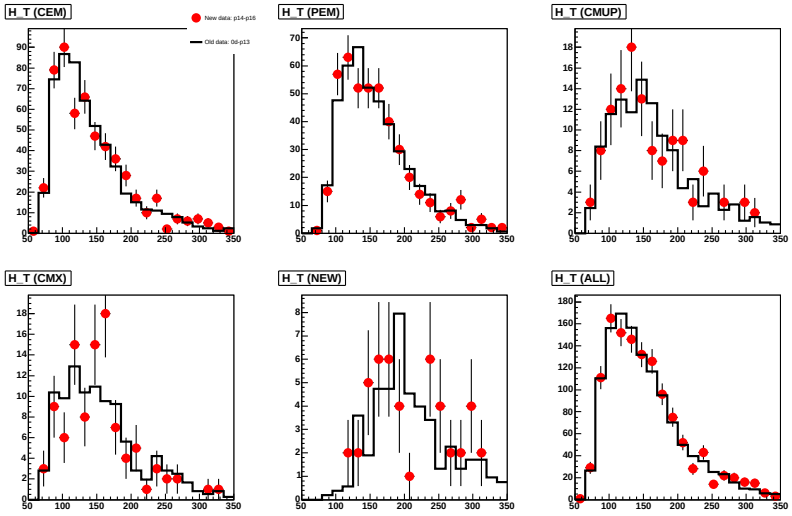

Figure 70: H_T distribution in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

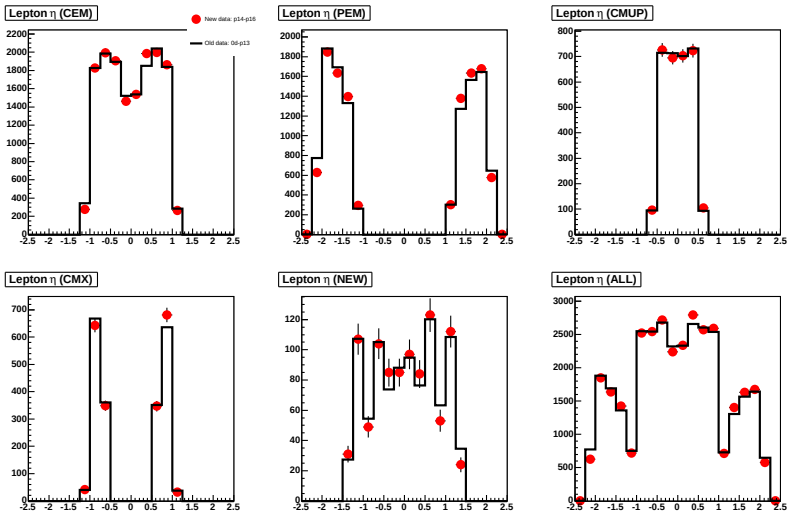

Figure 71: Lepton η distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

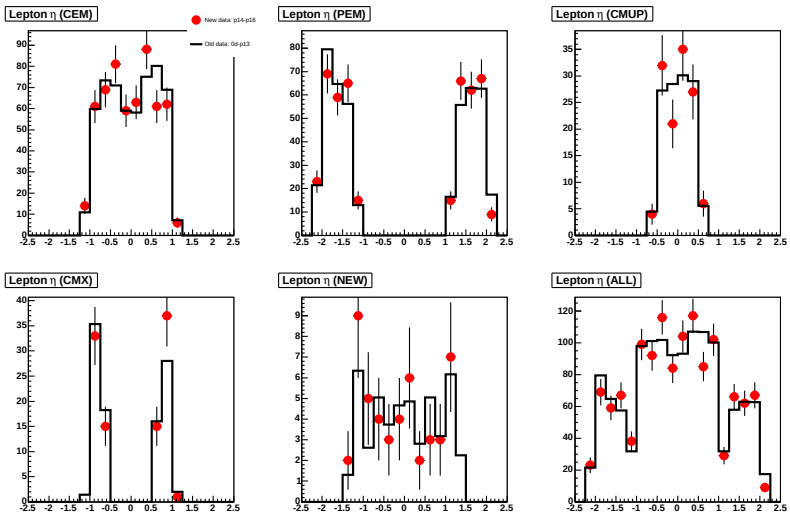

Figure 72: Lepton η distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

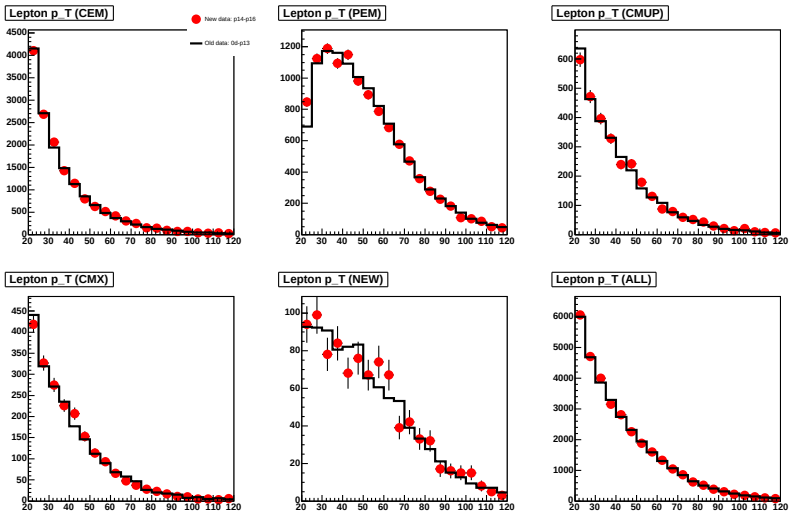

Figure 73: Lepton p_T distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

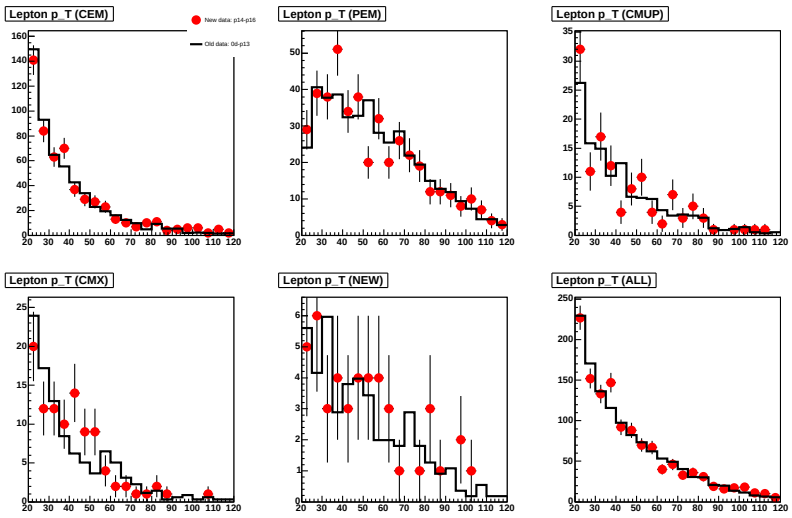

Figure 74: Lepton p_T distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

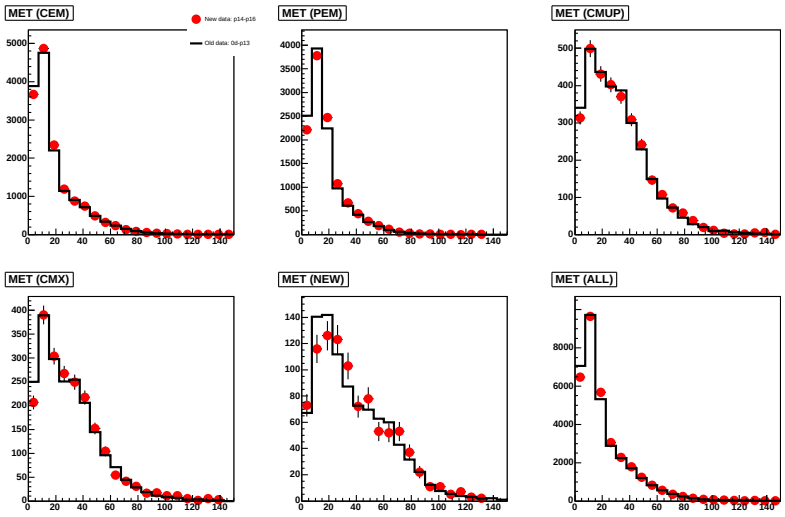

Figure 75: MET distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

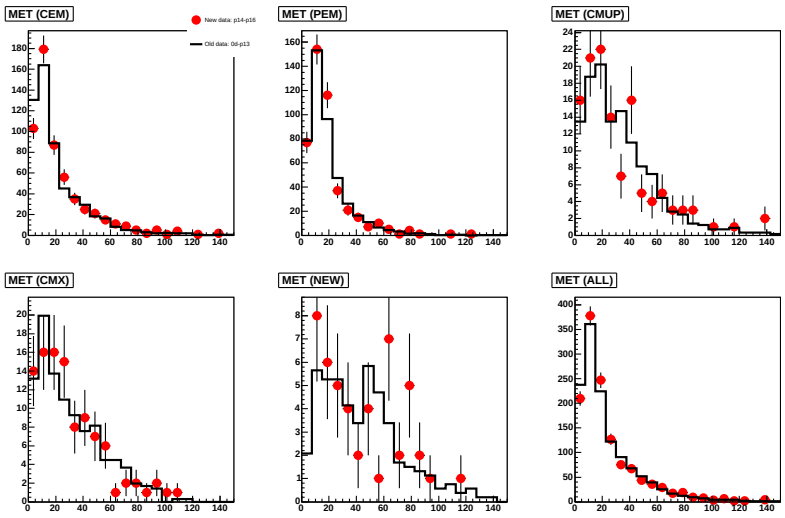

Figure 76: MET distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

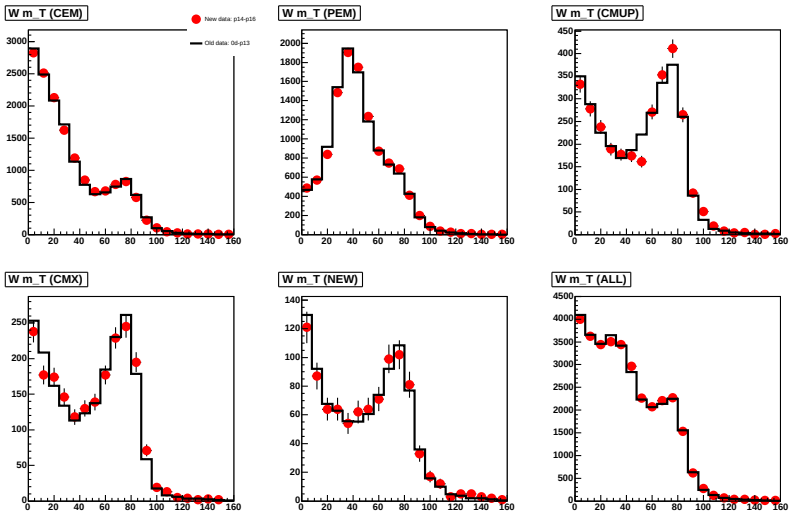

Figure 77: Transverse mass distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

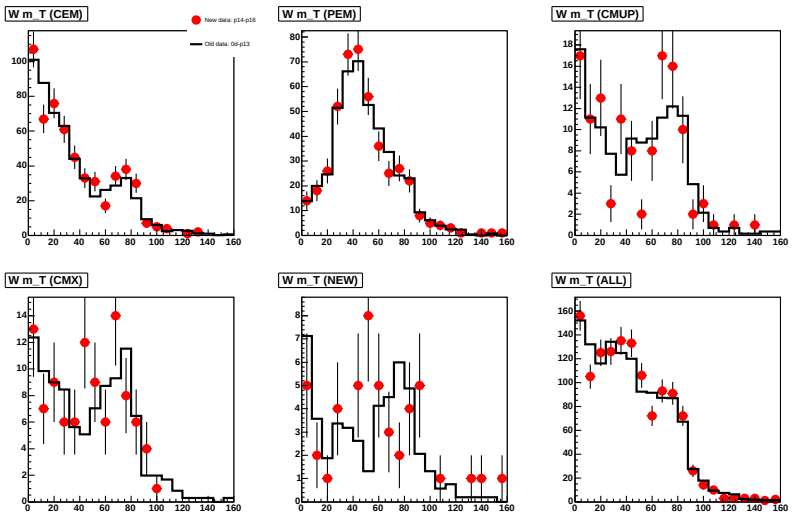

Figure 78: Transverse mass distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

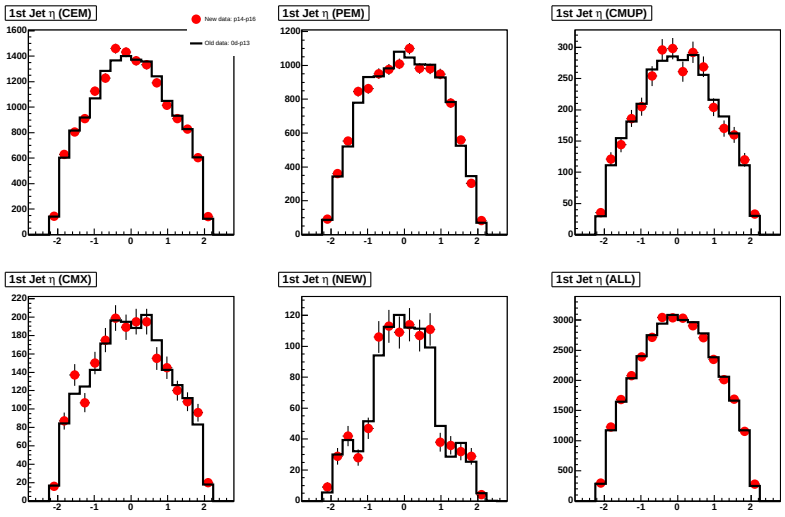

Figure 79: Jet-1 η distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

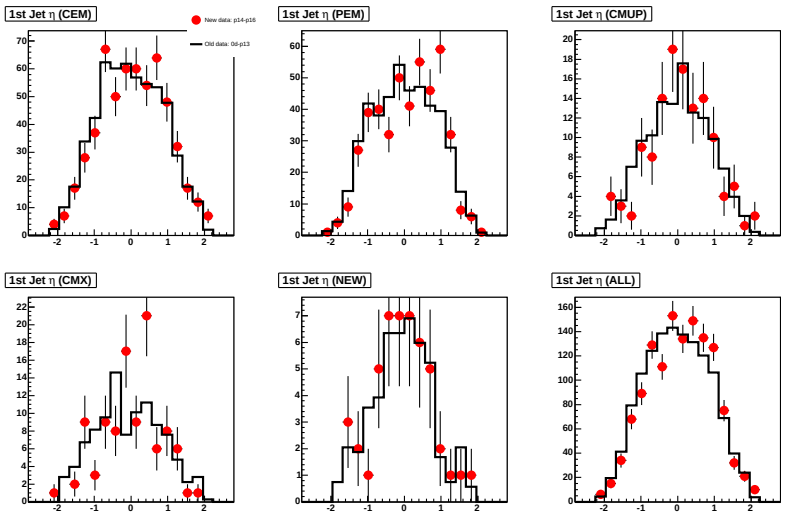

Figure 80: Jet-1 η distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

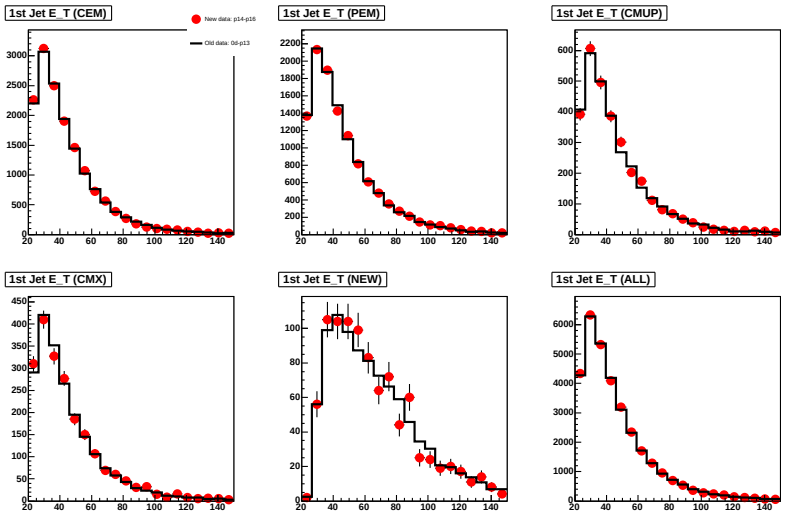

Figure 81: Jet-1 E_T distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

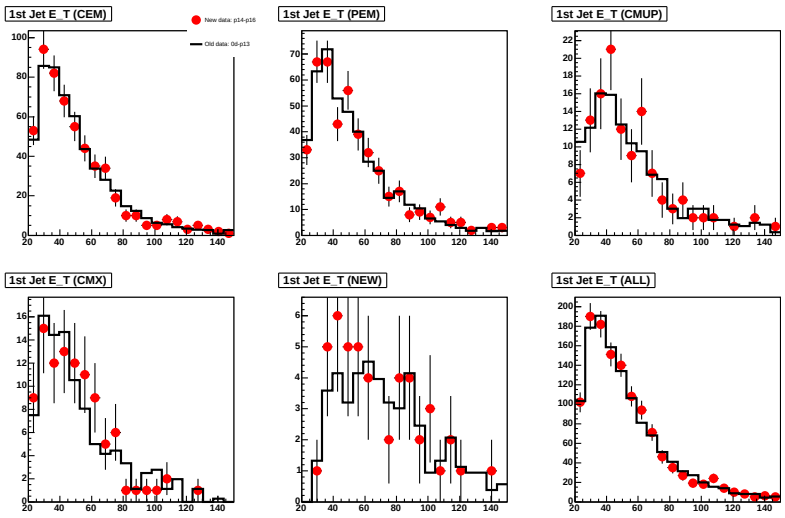

Figure 82: Jet-1 E_T distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

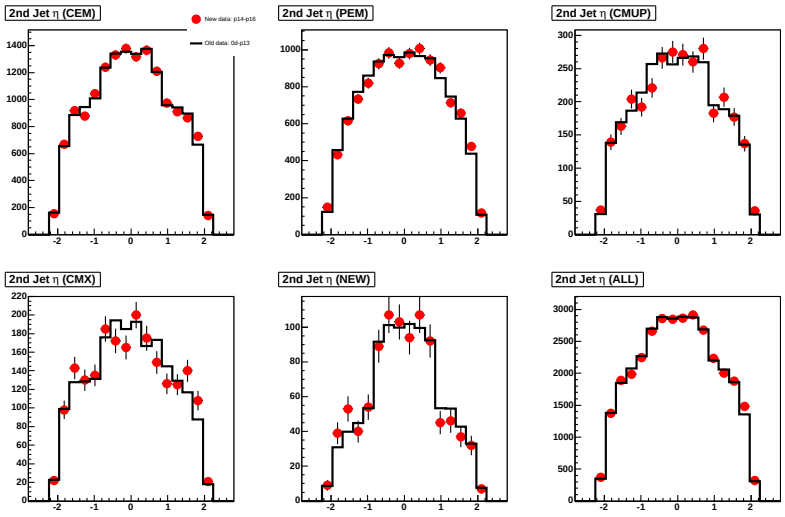

Figure 83: Jet-2 η distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

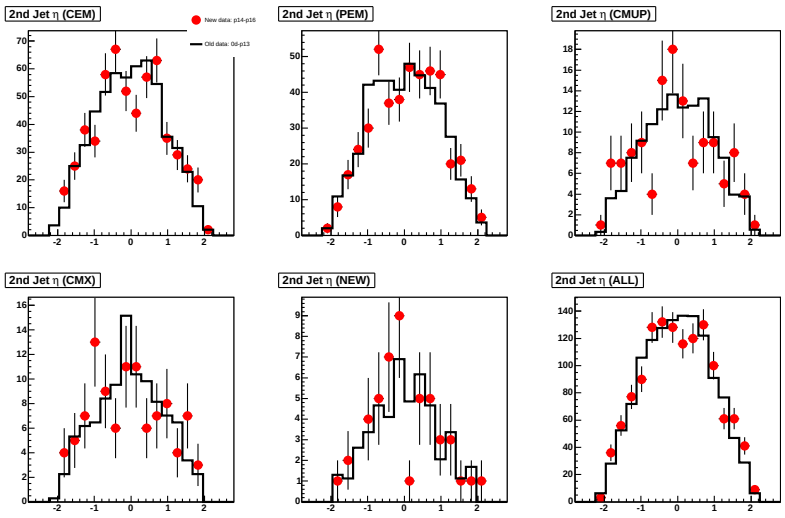

Figure 84: Jet-2 η distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

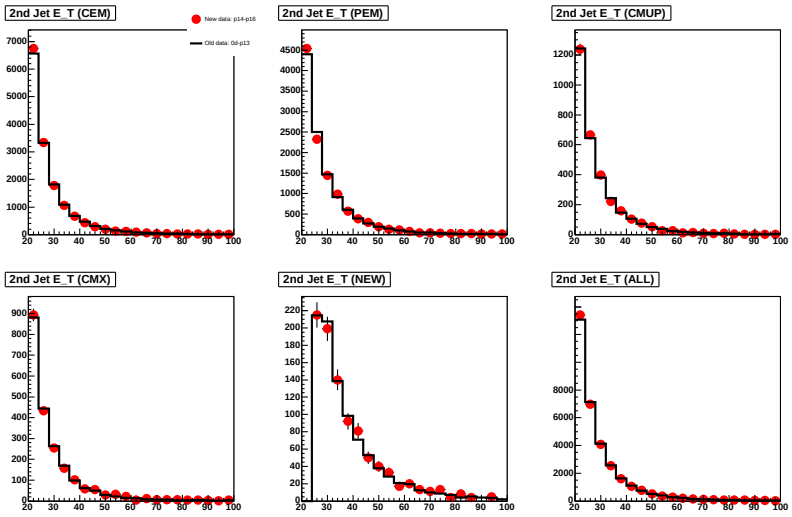

Figure 85: Jet-2 E_T distributions in the no tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

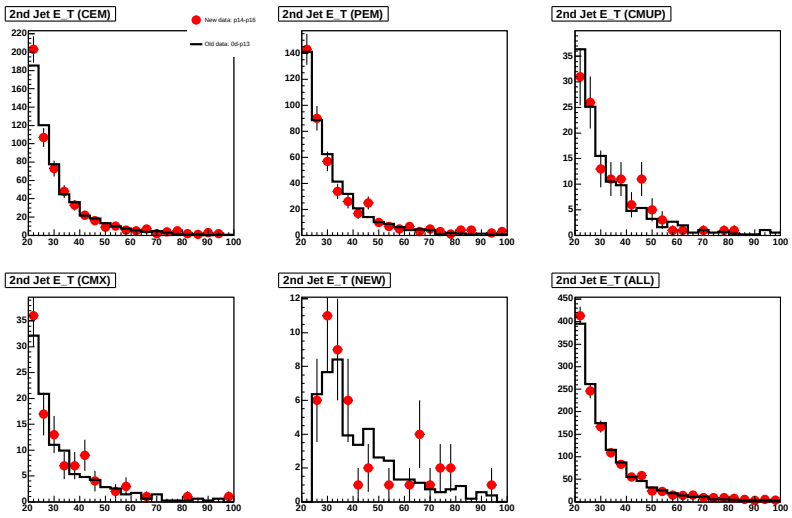

Figure 86: Jet-2 E_T distributions in the 1-tag bin for the new data (p14-p17) compared with old data (prior to p14). The old data has been scaled to the new data.

I Splitting Tagging Categories

In an effort to combine the result described in this note with the one obtained using Neural Network [6], we decided to split the tagging categories in the same way it is done in that analysis. That is, we divide the single tag category (events that have only one of the 2 jets tagged by tight SecVtx) in two orthogonal categories: events where one jet is tagged by tight SecVtx and the other jet (that is not tagged by tight SecVtx) is tagged by JetProbability [30] (using a 5% cut), and events where one jet is tagged by tight SecVtx and the other jet is not tagged by either one of those tagging algorithms. We denote these categories as STJP and ST, respectively. The double tag category (STST) remains the same.

Tables 22, 23 and 24 show the expected number of signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements, in the single (ST), extended double (STJP) and double tag case (STST), respectively.

One important note here is that for events in the STJP category, since both jets are tagged by some tagging algorithm, the QCD veto is not applied.

We apply the BDT trained with the 2 tag sample to the STST events while the BDT trained with the 1 tag sample is applied to the STJP and ST events. We tried to apply the BDT trained with the 2 tag sample to the STJP events but the performance was not so good. The reason is that the BDT trained with the 2 tag sample uses as input an EPD built using the KIT flavor separator of both jets, and this variable is not well defined for jets that are not tagged by SecVtx, as is the case in the STJP sample. Therefore, we decided to use the BDT trained with the 1 tag sample that uses as input an EPD built using only the KIT flavor separator of the jet tagged by SecVtx.

The expected limits for a Higgs mass of 115 GeV/c^2 for tight leptons, loose muons and all of them together, splitting in tagging categories as described above are show in Table 20. The expected limit for the same Higgs mass is 5.64 times the standard model if we classify the events in single and double tag (see Table 7); now, splitting the single tag category, the expected limit is 5.38 times the SM, which is $\sim$4.5% better.

σ/SM	Tight Leptons	Loose Muons	Total
STST	9.71	21.0	—
STJP (BDT 1 tag)	12.29	31.1	—
ST	13.03	25.9	—
Total	6.13	12.7	5.38

Table 20: Expected upper limit cross sections, in SM units, for a Higgs mass of 115 GeV/c^2, for the different tagging categories, for tight leptons (CEM, PEM, CMUP, CMX) and loose muons.

Table 21 shows the final expected limits for all Higgs masses, when splitting tagging categories, and, for comparison, the limits obtained no splitting the tagging categories (as in Table 7). The relative difference is also shown. In average, an improvement of $\sim$5.5% is obtained.

σ/SM	100	105	110	115	120	125	130	135	140	145	150
ST+STST	4.07	4.28	4.82	5.64	7.02	8.32	10.7	14.2	20.8	29.4	52.1
ST+STJP+STST	3.87	4.02	4.62	5.38	6.57	8.07	10.1	13.6	19.5	27.8	48.5
Δ (%)	4.9	6.1	4.1	4.6	6.4	3.0	5.6	4.2	6.3	7.1	6.9

Table 21: Expected upper limit cross sections, in SM units, for all Higgs masses, with and without splitting the tagging categories, for tight leptons (CEM, PEM, CMUP, CMX) and loose muons.

The results shown in this appendix are meant to go into the WH combination [31]. We plan to update the result with other improvements in the near future.

Process	2 jets	3 jets	4 jets	5 jets
All Pretag Cands.	50641.0 ± 0.0	8907.0 ± 0.0	1997.0 ± 0.0	417.0 ± 0.0
Tagged WW	54.07 ± 5.74	16.33 ± 1.62	3.88 ± 0.35	0.77 ± 0.06
Tagged WZ	19.35 ± 1.30	5.11 ± 0.34	1.14 ± 0.08	0.23 ± 0.02
Tagged ZZ	0.67 ± 0.05	0.27 ± 0.02	0.07 ± 0.00	0.02 ± 0.00
Tagged TopLJ	101.97 ± 14.10	239.78 ± 33.50	222.71 ± 32.37	57.34 ± 8.55
Tagged TopDil	38.68 ± 5.36	26.54 ± 3.73	5.70 ± 0.81	0.92 ± 0.13
Tagged STopT	61.53 ± 8.90	10.29 ± 1.46	1.38 ± 0.19	0.16 ± 0.02
Tagged STopS	31.57 ± 4.41	9.12 ± 1.29	1.80 ± 0.26	0.30 ± 0.04
Tagged Z+jets	34.70 ± 5.03	12.22 ± 1.71	3.01 ± 0.41	0.53 ± 0.07
Tagged Wbb	473.91 ± 142.88	123.28 ± 37.28	26.16 ± 8.06	5.81 ± 1.82
Tagged Wcc/Wc	464.68 ± 143.04	110.65 ± 33.95	23.60 ± 7.33	5.26 ± 1.65
Tagged Total HF	938.59 ± 284.96	233.94 ± 70.91	49.76 ± 15.20	11.06 ± 3.40
Tagged Total MC	342.54 ± 29.05	319.66 ± 36.71	239.68 ± 33.05	60.28 ± 8.67
Tagged Mistags	447.21 ± 54.89	120.27 ± 14.99	25.14 ± 3.83	5.47 ± 1.01
Tagged Non-W	126.86 ± 50.74	43.65 ± 17.46	8.15 ± 3.26	4.07 ± 1.63
Total Prediction	1855.20 ± 296.03	717.52 ± 83.10	322.73 ± 36.72	80.88 ± 9.51
Tagged WH100	7.43 ± 0.60	1.72 ± 0.14	0.30 ± 0.03	0.04 ± 0.00
Tagged WH105	6.72 ± 0.54	1.63 ± 0.13	0.26 ± 0.02	0.03 ± 0.00
Tagged WH110	5.89 ± 0.47	1.46 ± 0.12	0.25 ± 0.02	0.03 ± 0.00
Tagged WH115	4.91 ± 0.40	1.28 ± 0.11	0.23 ± 0.02	0.03 ± 0.00
Tagged WH120	3.84 ± 0.31	1.03 ± 0.09	0.17 ± 0.01	0.02 ± 0.00
Tagged WH125	3.18 ± 0.26	0.89 ± 0.07	0.16 ± 0.01	0.02 ± 0.00
Tagged WH130	2.41 ± 0.20	0.72 ± 0.06	0.13 ± 0.01	0.02 ± 0.00
Tagged WH135	1.75 ± 0.14	0.55 ± 0.05	0.10 ± 0.01	0.01 ± 0.00
Tagged WH140	1.18 ± 0.10	0.38 ± 0.03	0.07 ± 0.01	0.01 ± 0.00
Tagged WH145	0.80 ± 0.06	0.26 ± 0.02	0.05 ± 0.00	0.01 ± 0.00
Tagged WH150	0.50 ± 0.04	0.17 ± 0.01	0.03 ± 0.00	0.00 ± 0.00
Observed	1851.00 ± 0.00	682.00 ± 0.00	344.00 ± 0.00	85.00 ± 0.00

Table 22: Number of expected single tagged (STnoJP) signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	2 jets	3 jets	4 jets	5 jets
All Pretag Candidates	57171.0 ± 0.0	10230.0 ± 0.0	2276.0 ± 0.0	472.0 ± 0.0
Tagged WW	1.79 ± 0.50	1.45 ± 0.38	0.65 ± 0.15	0.22 ± 0.05
Tagged WZ	3.60 ± 0.59	1.18 ± 0.20	0.29 ± 0.05	0.07 ± 0.01
Tagged ZZ	0.15 ± 0.02	0.07 ± 0.01	0.03 ± 0.01	0.01 ± 0.00
Tagged TopLJ	20.27 ± 4.09	83.98 ± 16.34	113.91 ± 21.16	34.76 ± 6.39
Tagged TopDil	10.50 ± 1.99	9.21 ± 1.74	2.36 ± 0.44	0.42 ± 0.07
Tagged STopT	2.40 ± 0.56	1.50 ± 0.31	0.37 ± 0.07	0.05 ± 0.01
Tagged STopS	9.39 ± 1.78	3.41 ± 0.64	0.81 ± 0.15	0.17 ± 0.03
Tagged Z+jets	2.60 ± 0.60	1.90 ± 0.44	0.64 ± 0.15	0.15 ± 0.04
Tagged Wbb	62.89 ± 20.76	24.97 ± 8.28	7.44 ± 2.52	1.97 ± 0.69
Tagged Wcc/Wc	23.86 ± 8.68	11.70 ± 4.25	4.22 ± 1.55	1.16 ± 0.44
Tagged Total HF	86.75 ± 28.99	36.67 ± 12.27	11.66 ± 3.95	3.13 ± 1.08
Tagged Total MC	50.71 ± 8.74	102.71 ± 18.80	119.06 ± 21.79	35.84 ± 6.52
Tagged Mistags	10.72 ± 3.55	6.61 ± 2.17	2.34 ± 0.79	0.67 ± 0.26
Tagged Non-W	14.63 ± 5.85	8.43 ± 3.37	4.46 ± 1.79	3.55 ± 1.42
Total Prediction	162.81 ± 31.05	154.41 ± 22.81	137.53 ± 22.23	43.19 ± 6.77
Tagged WH100	2.14 ± 0.33	0.62 ± 0.09	0.12 ± 0.02	0.02 ± 0.00
Tagged WH105	1.92 ± 0.29	0.61 ± 0.09	0.12 ± 0.02	0.02 ± 0.00
Tagged WH110	1.73 ± 0.26	0.53 ± 0.08	0.11 ± 0.02	0.01 ± 0.00
Tagged WH115	1.44 ± 0.22	0.48 ± 0.07	0.11 ± 0.02	0.02 ± 0.00
Tagged WH120	1.15 ± 0.18	0.40 ± 0.06	0.08 ± 0.01	0.01 ± 0.00
Tagged WH125	0.91 ± 0.14	0.33 ± 0.05	0.07 ± 0.01	0.01 ± 0.00
Tagged WH130	0.76 ± 0.12	0.26 ± 0.04	0.06 ± 0.01	0.01 ± 0.00
Tagged WH135	0.54 ± 0.08	0.20 ± 0.03	0.04 ± 0.01	0.01 ± 0.00
Tagged WH140	0.37 ± 0.06	0.15 ± 0.02	0.03 ± 0.00	0.01 ± 0.00
Tagged WH145	0.25 ± 0.04	0.10 ± 0.01	0.02 ± 0.00	0.00 ± 0.00
Tagged WH150	0.16 ± 0.02	0.07 ± 0.01	0.02 ± 0.00	0.00 ± 0.00
Observed	159.00 ± 0.00	156.00 ± 0.00	151.00 ± 0.00	50.00 ± 0.00

Table 23: Number of expected extended double tagged (STJP) signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

Process	2 jets	3 jets	4 jets	5 jets
All Pretag Cand.s	57171.0 ± 0.0	10230.0 ± 0.0	2276.0 ± 0.0	472.0 ± 0.0
Tagged WW	0.39 ± 0.08	0.47 ± 0.08	0.26 ± 0.04	0.09 ± 0.01
Tagged WZ	4.79 ± 0.53	1.33 ± 0.15	0.25 ± 0.03	0.05 ± 0.01
Tagged ZZ	0.15 ± 0.02	0.10 ± 0.01	0.03 ± 0.00	0.01 ± 0.00
Tagged TopLJ	23.74 ± 3.92	93.19 ± 15.25	128.74 ± 20.84	37.30 ± 6.02
Tagged TopDil	14.12 ± 2.30	12.49 ± 2.03	3.02 ± 0.49	0.59 ± 0.10
Tagged STopT	1.78 ± 0.31	1.36 ± 0.23	0.36 ± 0.06	0.05 ± 0.01
Tagged STopS	12.73 ± 2.08	4.20 ± 0.68	0.92 ± 0.15	0.17 ± 0.03
Tagged Z+jets	2.10 ± 0.34	1.35 ± 0.22	0.40 ± 0.07	0.10 ± 0.02
Tagged Wbb	70.41 ± 22.54	24.14 ± 7.84	6.55 ± 2.17	1.55 ± 0.54
Tagged Wcc/Wc	6.80 ± 2.30	3.34 ± 1.14	1.28 ± 0.45	0.35 ± 0.13
Tagged Total HF	77.21 ± 24.70	27.48 ± 8.88	7.83 ± 2.57	1.90 ± 0.64
Tagged Total MC	59.79 ± 7.52	114.49 ± 16.94	133.98 ± 21.22	38.36 ± 6.10
Tagged Mistags	2.14 ± 0.57	1.47 ± 0.42	0.54 ± 0.16	0.16 ± 0.06
Tagged Non-W	8.90 ± 3.56	4.81 ± 1.92	4.96 ± 1.98	0.34 ± 0.14
Total Prediction	148.03 ± 26.07	148.25 ± 19.23	147.30 ± 21.46	40.76 ± 6.13
Tagged WH100	2.87 ± 0.34	0.79 ± 0.09	0.13 ± 0.02	0.02 ± 0.00
Tagged WH105	2.64 ± 0.31	0.73 ± 0.08	0.13 ± 0.02	0.02 ± 0.00
Tagged WH110	2.34 ± 0.27	0.66 ± 0.08	0.12 ± 0.01	0.02 ± 0.00
Tagged WH115	2.00 ± 0.23	0.57 ± 0.07	0.11 ± 0.01	0.01 ± 0.00
Tagged WH120	1.62 ± 0.19	0.49 ± 0.06	0.09 ± 0.01	0.01 ± 0.00
Tagged WH125	1.28 ± 0.15	0.43 ± 0.05	0.08 ± 0.01	0.01 ± 0.00
Tagged WH130	1.03 ± 0.12	0.34 ± 0.04	0.07 ± 0.01	0.01 ± 0.00
Tagged WH135	0.74 ± 0.09	0.25 ± 0.03	0.05 ± 0.01	0.01 ± 0.00
Tagged WH140	0.50 ± 0.06	0.18 ± 0.02	0.04 ± 0.00	0.01 ± 0.00
Tagged WH145	0.34 ± 0.04	0.13 ± 0.01	0.03 ± 0.00	0.00 ± 0.00
Tagged WH150	0.21 ± 0.02	0.08 ± 0.01	0.02 ± 0.00	0.00 ± 0.00
Observed	157.00 ± 0.00	174.00 ± 0.00	144.00 ± 0.00	46.00 ± 0.00

Table 24: Number of expected double tagged (STST) signal and background events in 2.7 fb^{-1} of CDF data, passing all event selection requirements.

J Isotracks

In a further effort to combine the result described in this note with the one obtained using Neural Network [6], we decided to use the so called "isotracks" in the same way it is done in that analysis. In order to be sure that we select the same events, the combination analysis [31] checks the matching between the isotracks provided by both groups and finds excellent agreement. Figs. 87- 90 show validation plots for the isotracks used in this analysis. The agreement is reasonable.

We analyze these kind of events almost in the same way we treat tight leptons or loose muons. The only difference is that the Neural Network TF for b and c jets described in Sec. 3.2.2 is not used, instead we use the standard TF used for the winter 2008 (2.2 fb^{-1}) analysis.

Table 25 shows the expected limits, for a Higgs mass of 115 GeV/c^2, using only isotracks and, for comparison, using only loose muons. The isotracks provide a limit $\sim 8\%$ lower than the loose muons.

σ/SM	Isotracks	Loose muons	Δ (%)
STST	18.5	21.0	11.9
STJP	27.6	31.1	11.3
ST	24.4	25.9	5.8
Total	11.7	12.7	7.9

Table 25: Expected upper limit cross sections, in SM units, for a Higgs mass of 115 GeV/c^2, for the different tagging categories, for loose muons and isotracks only.

If we put together the isotracks, or loose muons, with the tight leptons, we obtain the expected limits shown in Table 26 (these limits are obtained after splitting the tagging categories as explained in the previous Appendix). The improvement respect to using loose muons (Table 21) and with respect to the results in Table 7 are also shown. In average, using isotracks instead of loose muons provides a $\sim$1.5% lower expected limit. The final improvement after splitting in tagging categories and using isotracks instead of loose muons is $\sim$7%.

σ / SM	100	105	110	115	120	125	130	135	140	145	150
Loose muons	3.87	4.02	4.62	5.38	6.57	8.07	10.1	13.6	19.5	27.8	48.5
Isotracks	3.80	3.98	4.53	5.24	6.34	7.97	9.97	13.4	19.2	27.0	48.7
Δ (%)	1.8	1.0	1.9	2.6	3.5	1.2	1.3	1.5	1.5	2.9	-0.4
TOTAL Δ (%)	6.6	7.0	6.0	7.1	9.7	4.2	6.8	5.6	7.7	8.2	6.5

Table 26: Total expected upper limit cross sections, in SM units, for all Higgs masses, using tight leptons and loose muons or isotracks.

Again, the results shown in this appendix are meant to go into the WH combination [31]. We plan to update the result with other improvements in the near future.

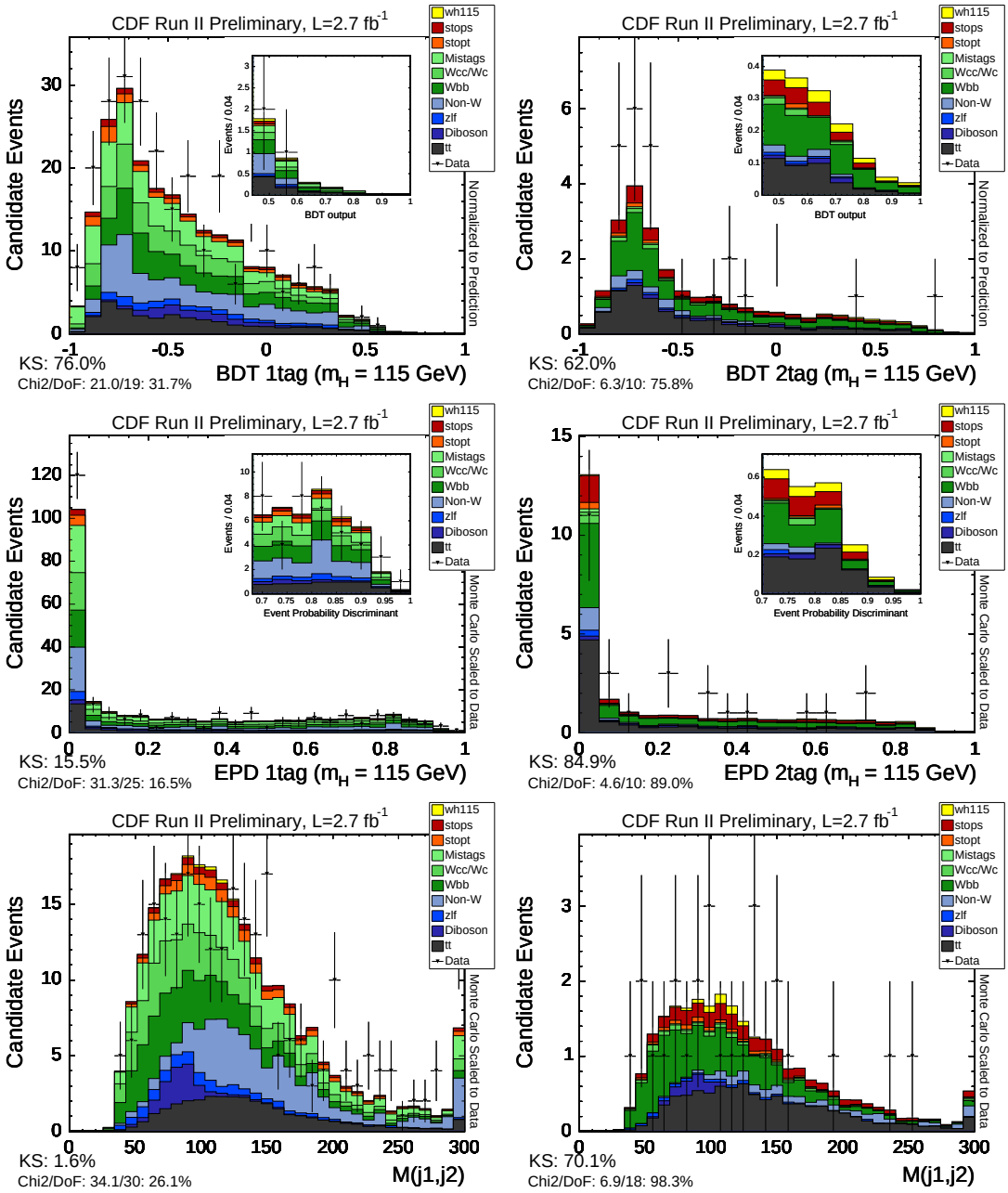

Figure 87: Left (right): distributions of different variables, for isotrack events, in the single (double) tagged sample.

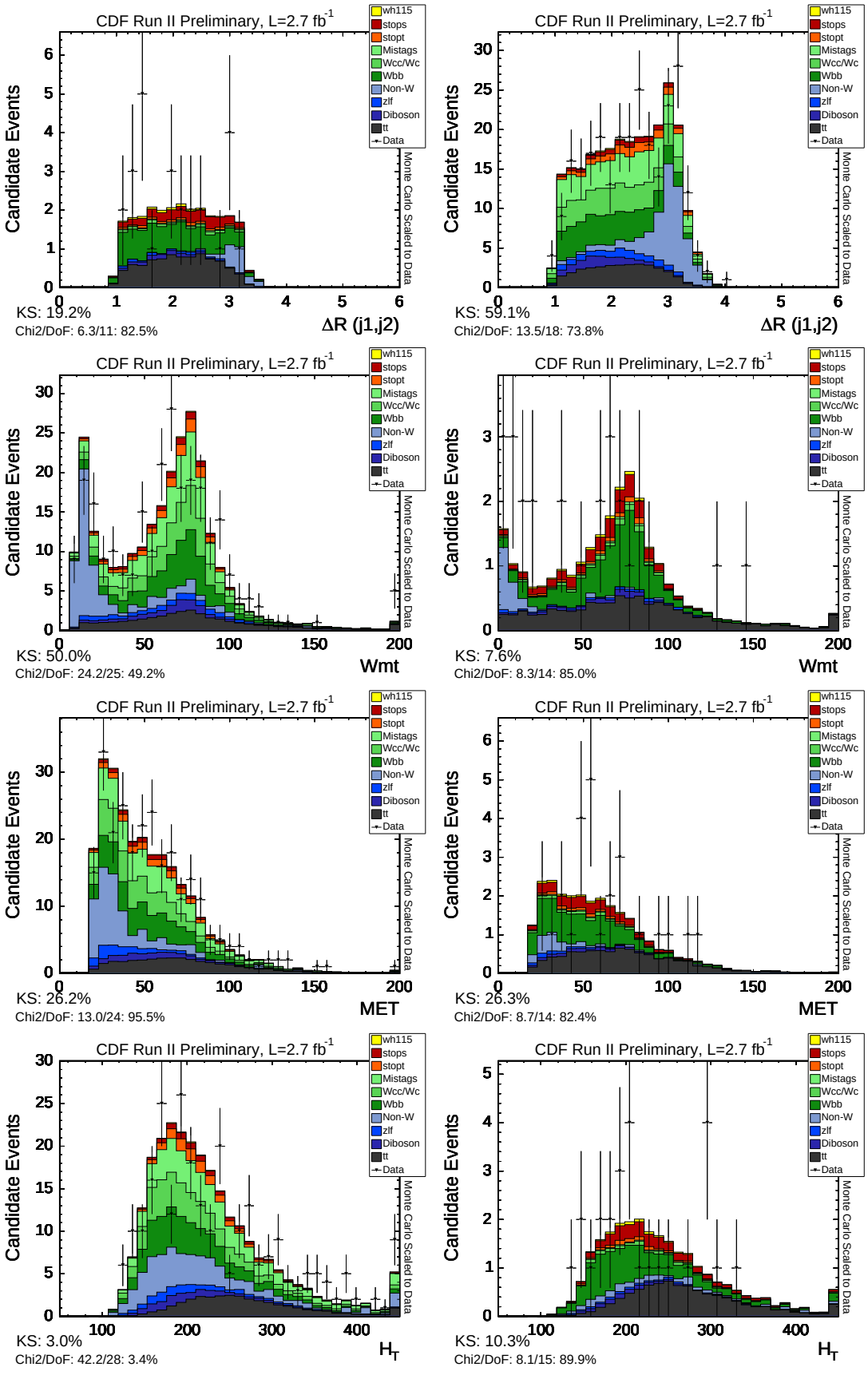

Figure 88: Left (right): distributions of different variables, for isotrack events, in the single (double) tagged sample.

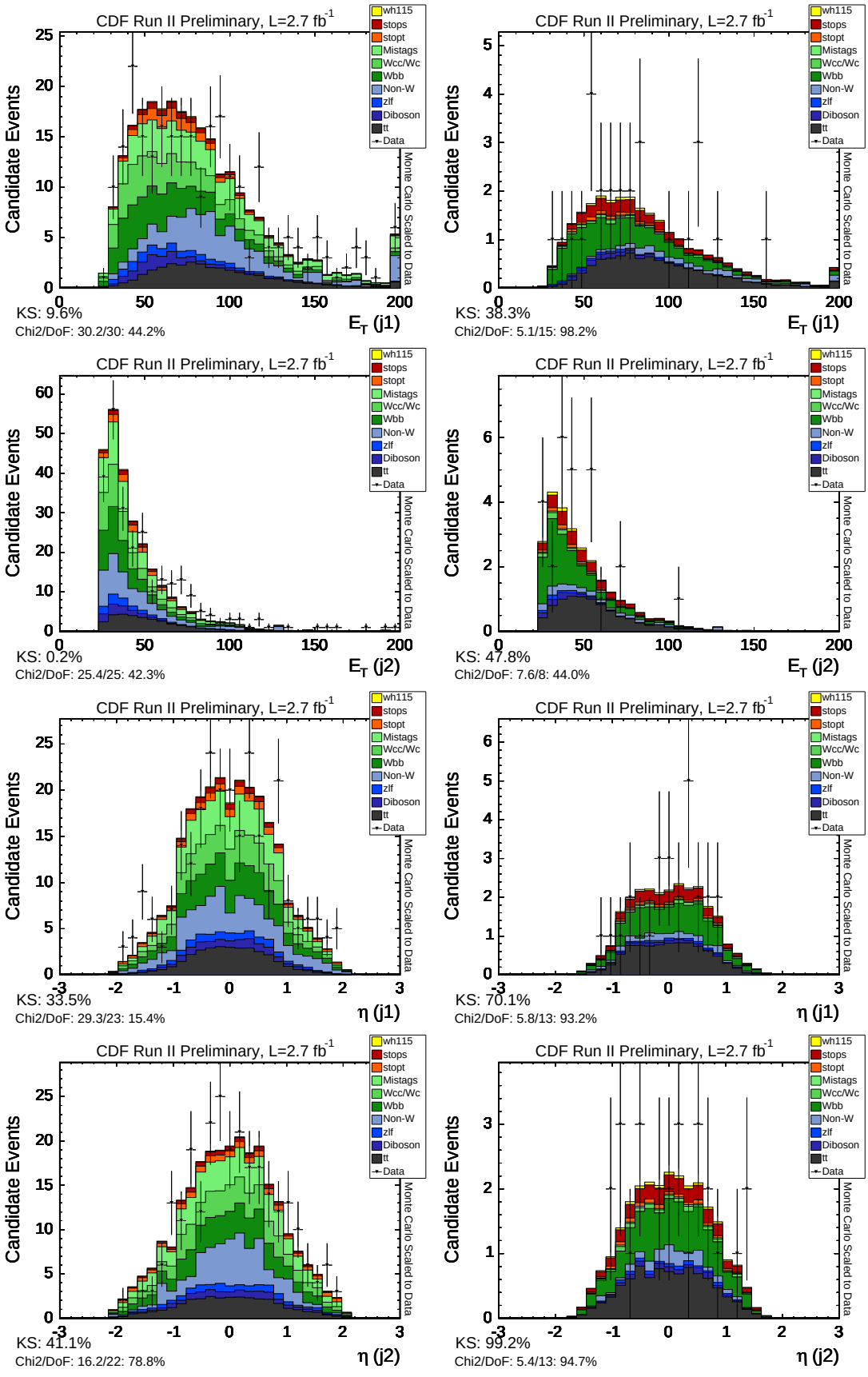

Figure 89: Left (right): distributions of different variables, for isotrack events, in the single (double) tagged sample.

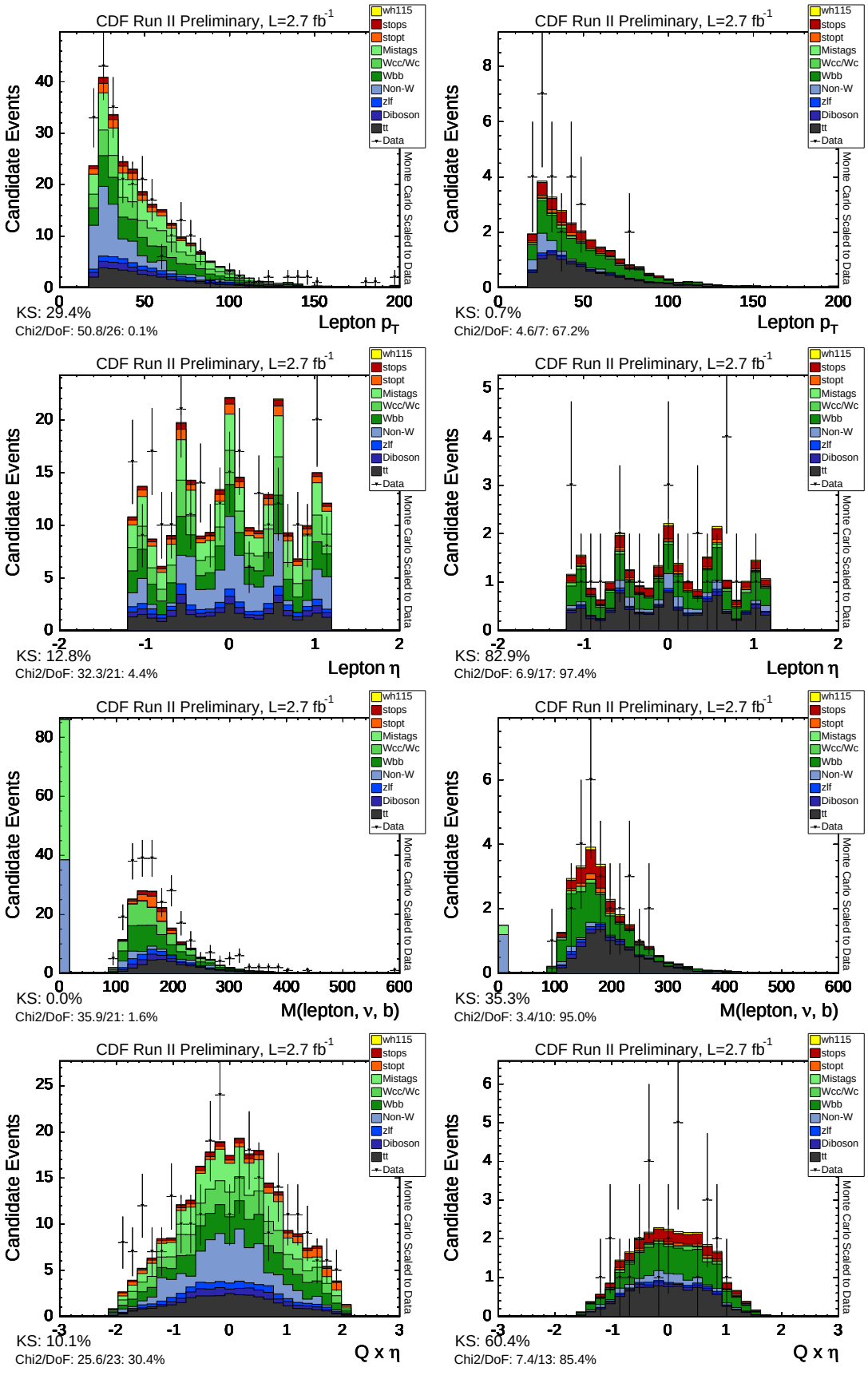

Figure 90: Left (right): distributions of different variables, for isotrack events, in the single (double) tagged sample.