CDF/ANAL/TOP/CDFR/7816
Version 4.0, June 16, 2006

A Neural Network b Tagger for Single-Top Analyses

Michael Feindt, Svenja Richter, Wolfgang Wagner
Institut für Experimentelle Kernphysik, Universität Karlsruhe

Abstract

We present a neural network b tagger that exploits a broad range of variables to identify b quark jets with high purity. Our tagging algorithm is applied to jets that are already selected by the secondary vertex tagger (SecVtx). In addition to the long lifetime of b hadrons the network tagger utilizes variables that describe the high b mass, the high decay multiplicity and the decay into leptons. The additional application of this network in the single–top analysis leads to a reduction of more than 60% of non-b vertices but keeps about 82% of real b vertices.

Contents

1 Introduction **2**

2 Data and Monte Carlo Samples **3**

3 Discriminating Variables **4**

4 Functionality of the Neural Network **14**
4.1 The training process 16
4.2 Preprocessing of the variables 16
4.3 Automatic variable selection 16

5 Training of the neural network **17**
5.1 Determination of the cut on the significance of the input variables 24

6 Determination of the working point **26**

7 Samples used for data–MC comparison **28**
7.0.1 Parameterization of the 8 GeV L2 lepton trigger 28

8 Comparison of distributions of training variables **29**

9 Preparation of the samples **42**

10 Composition of the away jet sample **43**

11 Determination of the correction function **45**
11.1 Accumulating the distributions 46
11.2 E_T dependence of the correction function 49
11.3 Electron Jet 49

12 Mistags **52**
12.1 Samples 54
12.2 Comparison of output distributions 54
12.3 Mistag Matrix 55
12.3.1 Correction functions 57

13 Influence of kinematic jet variables **57**
13.1 Data–MC comparison 63

14 Correction of the W + light jets sample **66**
14.1 Network using kinematic jet variables 66

15 Cross check of determination of the composition of the inclusive electron data **67**

16 Cross check with other samples **68**
16.1 Network using kinematic jet variables 71
16.2 Network without jet E_T and jet η 71
16.2.1 Discussion of influence of electron enhancement to the network using jet kinematics 71

17 Determination of systematics **77**

18 Conclusion **79**

List of Figures **79**

Bibliography **82**

1 Introduction

The identification of jets that include a b hadron is essential for measurements concerning top quarks. Since the lifetime of top quarks ($\tau \approx 4 \cdot 10^{-25}$ s) is shorter than the hadronization time, top quarks decay essentially as free quarks without forming hadrons. According to the Standard Model the top quark decays predominantly into a W boson and a b quark with a branching ratio close to 100%. The identification of b jets is an important method to reduce the large $W +$ jets background and obtain a data sample enriched in top quark events. This is particularly true for single-top searches. In previous single-top analyses [2] b jets were identified by the standard secondary vertex b tagger, SecVtx [3, 4].

Due to the signature of single-top events, the main backgrounds are $t\bar{t}$ and W+jets production as described in [2]. However, after the requirement of a SecVtx b tag, about 50% of the remaining non–top background are still from c or light quarks, as can be seen in table 1.

The secondary vertex tag method exploits mainly one characteristic feature of b hadrons, their long lifetime of about 1.6 ps. The goal of this analysis is to increase the purity of the b jet identification by further suppressing the backgrounds that do not contain a b jet but have a reconstructed secondary vertex. Further features of b hadrons are used: the large mass of 5 GeV/c^2, a high decay multiplicity, and a certain probability of decaying into leptons. An additional b tagging algorithm is applied on top of the secondary vertex tag. We use a neural network to combine several variables into one test statistic which discriminates between b jets and light or c jets.

2 Data and Monte Carlo Samples

The samples that are used for the training of the neural network b tagger described in this note are the same as for the 5.3.3 single-top analysis. The single-top samples were generated with MadEvent (s–channel: mtopya; t–channel: mtopua + mtopta). The matching of the t–channel samples was performed as described in Ref. [1]. The $t\bar{t}$ sample (ttopel) was generated with Pythia, the W+jets samples with Alpgen: $W + b\bar{b} + 0p$ (atopub, atopwb, atop0t), $W + c\bar{c} + 0p$ (atoprb, atopkb), $W + c + 1p$ (atopic), and $W + 1p$ (atopaa, atopcb), $W + 2p$ (atop5a, atopdb, atop8t), $W + 3p$ (atop6a, atopeb) and $W + 4p$ (atop7a, atopfb). In these samples the W boson is forced to decay into one of its leptonic decay modes: $W \to e\nu$, $W \to \mu\nu$, or $W \to \tau\nu$.

For the training we use events where exactly one electron or muon is identified according to tight criteria.

For the network training we use a mixture of several processes to model the background

$Wb\bar{b}$	34.6%
$Wc\bar{c}$	13.2%
Wc	13.1%
Mistags	29.4%
non-W	7.4%
Di-Boson	2.3%

Table 1: Non–top background composition for the single top analysis after requiring a secondary vertex tag.

sample as realistically as possible. The composition of the background mixture is based on the single-top background estimate.

3 Discriminating Variables

The goal of this analysis is the suppression of events with reconstructed secondary vertices that are not related to the decay of a b hadron, i.e. to improve the tagging purity by an additional rejection of c and light jets with a reconstructed secondary vertex.

For this purpose, it is necessary to extend the b jet identification by variables with additional discriminating power that exploit all characteristics of b hadrons like lifetime, mass, decay multiplicity and leptonic decays.

The variables used for the identification of b jets are defined for jets tagged by SecVtx. The requirements made for those jets are:

- jet collection 3, i.e. JetClu 0.4
- tight SecVtx tag = 1
- corrected jet $\mathrm{p_T} > 15$ GeV, correction level 4 (jetCorr04b)

One set of useful variables is given by the SecVtx algorithm for reconstructed secondary vertices. Those variables are shown in figures 1, 2 and 3. Since b hadrons are more massive than hadrons containing only charm or light quarks, the mass at the reconstructed secondary vertex should be larger in real b jets than in c or light ones. In figure 1 one can see the characteristic cut off of the non–b distribution at the charm quark mass. In addition, the transverse momentum of the secondary vertex includes information about the difference between b vertices and non–b ones. Apart from that, the 2– and 3–dimensional decay lengths of the secondary vertex and their significances are an additional information about the origin of the reconstructed vertex, especially when combined with other discriminating variables. One possibility to combine the information given by the SecVtx algorithm is the so–called pseudo lifetime of the secondary vertex: $c\tau = L_{xy} \times M_{\mathrm{vtx}}/P_T^{\mathrm{vtx}}$ [5].

An additional information is the pass of the secondary vertex fit. In pass 1, the fitting algorithm tries to reconstruct a vertex candidate that consists at least of three tracks. If this can not be achieved, a second pass is performed. In pass 2, tighter quality cuts on the tracks are required. A vertex reconstructed in pass 2 must contain at least two tracks. For details see Refs. [3, 4].

The high decay multiplicity of b hadrons affects the number of tracks associated to the secondary vertex. On average, this number should be larger for real b jets than for falsely reconstructed vertices or charm vertices. In figure 2, the number of *good SecVtx tracks*[1] in the jet is shown, here *good* means that the track must pass certain quality cuts [4]. Figure 2 also contains the number of pass 1 tracks, and the number of used tracks.

Another information about the reconstructed secondary vertex is the χ^2 of the vertex fit per degree of freedom and the net charge at the secondary vertex as displayed in figure 3.

[1] defTracks w/ $\mathrm{p_T} > 0.5$ GeV + veto for algorithms 16,2,21 (→ branch in TopNtuple)

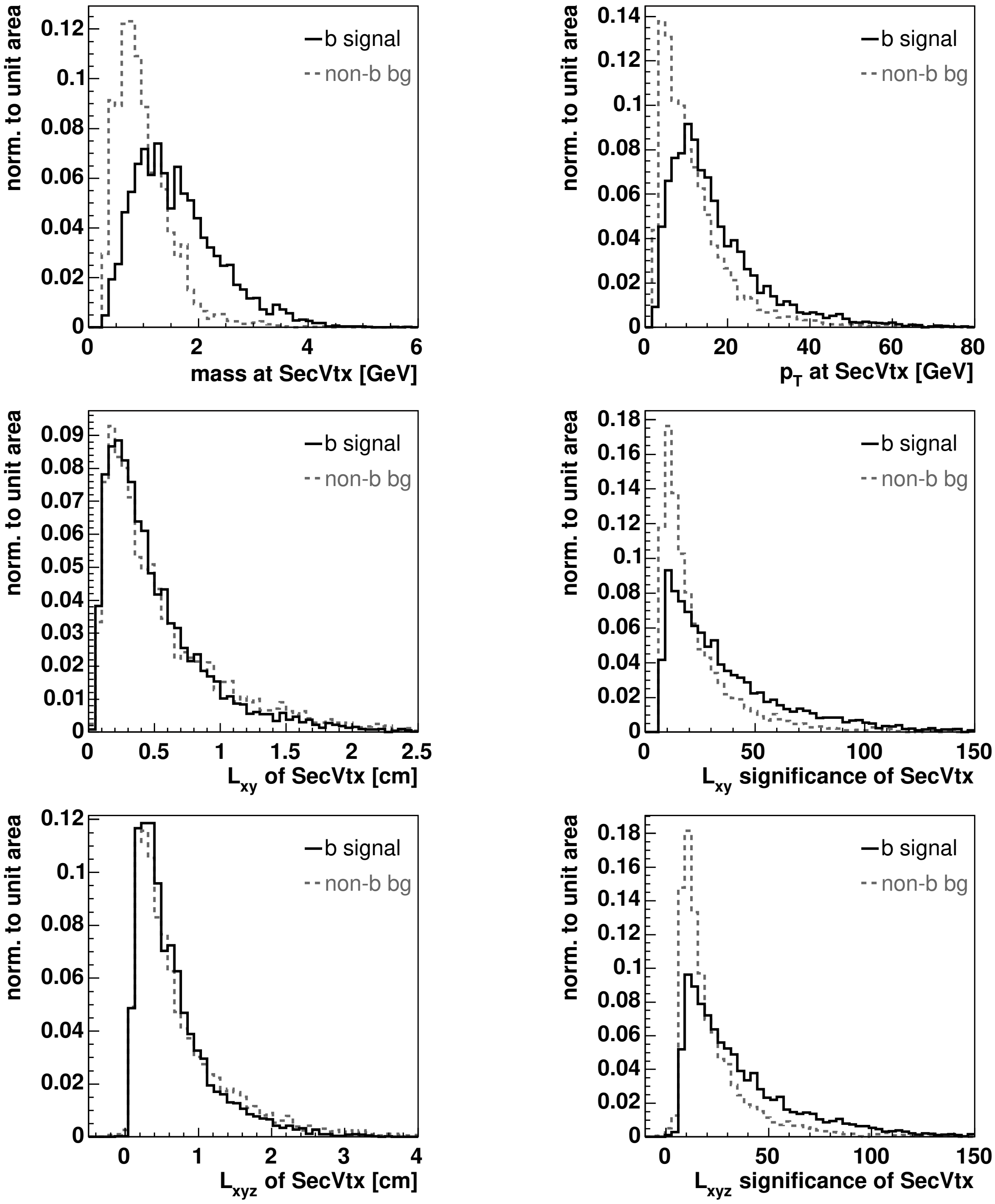
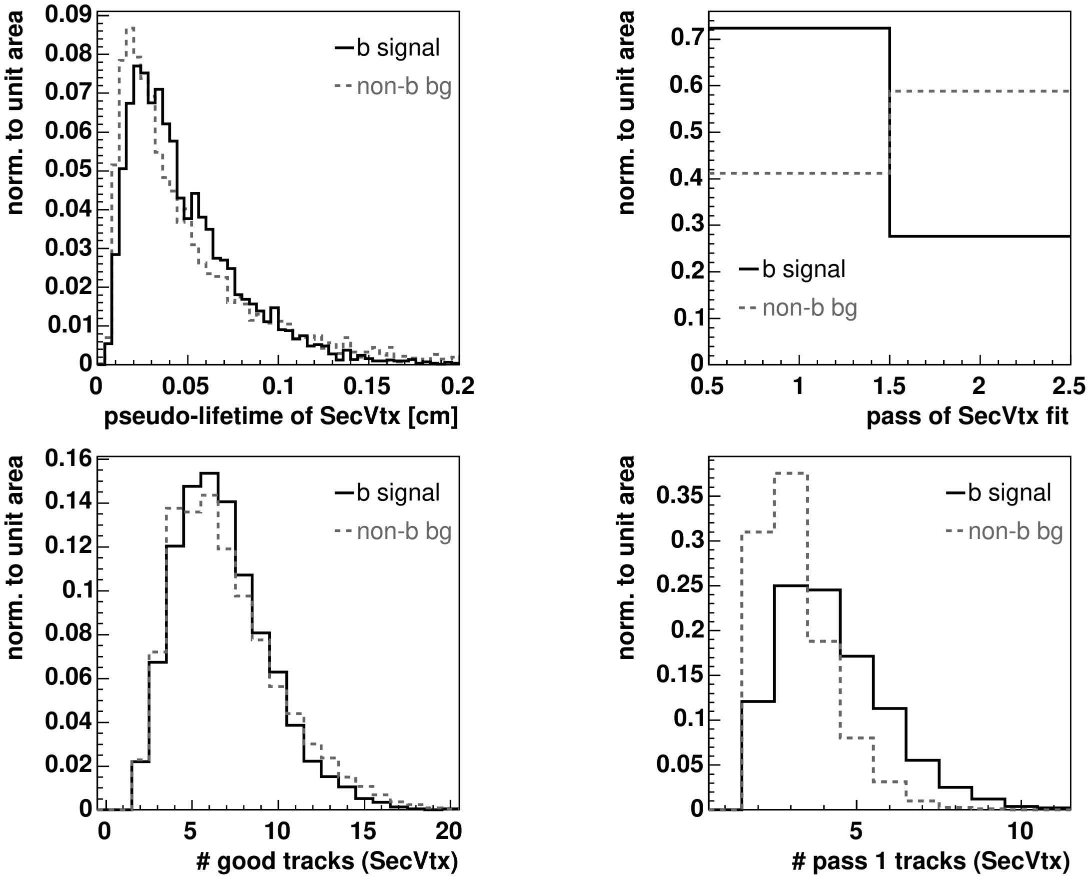


Figure 1: **Distributions of discriminating SecVtx variables.** These variables are given by the SecVtx algorithm. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.



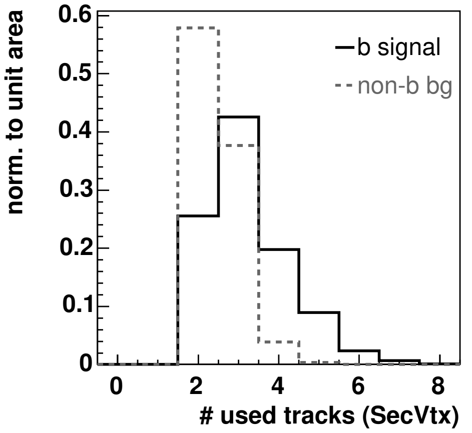
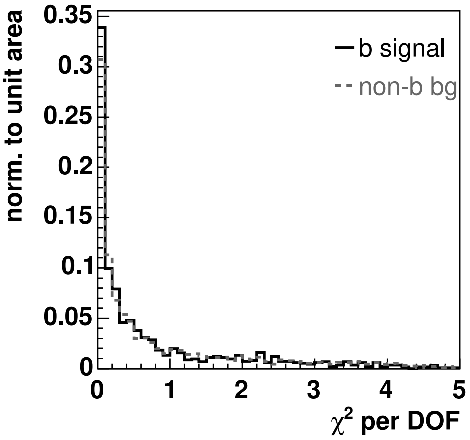


Figure 2: **Distributions of more discriminating SecVtx variables.** These variables are given by the SecVtx algorithm. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.



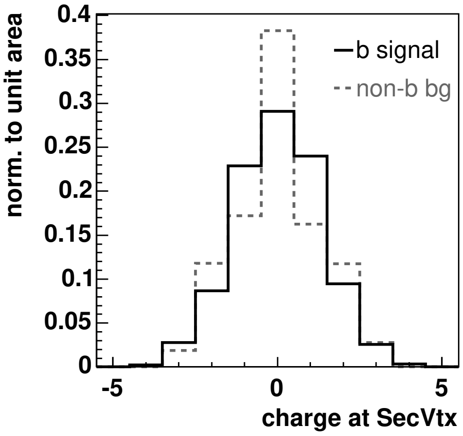


Figure 3: **Distributions of more discriminating SecVtx variables.** These variables are given by the SecVtx algorithm. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

In addition to the vertex related variables discussed above track based variables can be taken into account. We use SecVtx tracks marked as good for this purpose. The good SecVtx tracks in each tagged jet are sorted by their momentum. The three tracks with the largest momenta are taken as *1st track, 2nd track and 3rd track* mentioned in the plots. The SecVtx track variables used in this analysis are shown in figures 4, 5, 6, 7 and 8. For those tracks, their total momenta, rapidities and transverse momenta with respect to the jet axis are used for discrimination.

In addition, the impact parameter d_0, its significance, a lifetime–signed impact parameter and its significance are taken. The lifetime–signed impact parameter is calculated via

$$\frac{\epsilon^{zik} \cdot p^i_{\mathrm{jet}} \cdot p^k_{\mathrm{track}}}{|\epsilon^{zik} \cdot p^i_{\mathrm{jet}} \cdot p^k_{\mathrm{track}}|} \cdot d_0 \ ,$$

its significance analogically.

Furthermore, in each tagged jet, the number of tracks is counted that have an impact parameter significance of more than 1, 2 and 3 σ. Those numbers are taken as discriminating variables as well.

It can be seen that the tracks in tagged b jets have a smaller rapidity along the jet axis and a higher transverse momentum with respect to the jet axis than tracks in tagged non–b jets (see figure 5). It is obvious that the lifetime–signed impact parameters of tracks in b jets are shifted to higher values compared to the tracks in non–b jets (see figure 7). Additionally, the number of displaced tracks is higher in b jets than in non–b jets (figure 8).

In addition to those variables taking into account the large mass, the high decay multiplicity and the long lifetime of b hadrons, we want to exploit that b hadrons have a relatively high probability of decaying semi–leptonically. That is why the number of leptons and their momenta are considered as well (figure 9).

Kinematic jet quantities namely the transverse energy and the pseudo–rapidity of

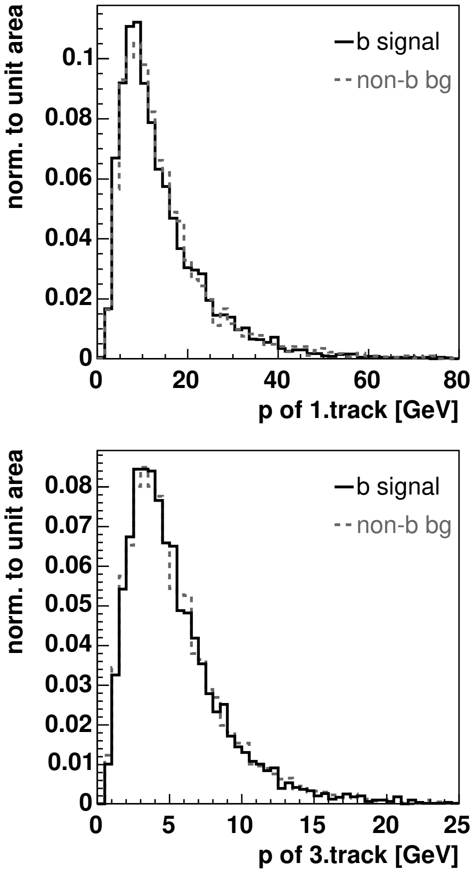


Figure 4: **Distributions of kinematic discriminating track variables.** In each tagged jet, all the SecVtx tracks were sorted by their momentum. Thus, "1. track" means the track with the highest momentum in the jet and so on. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

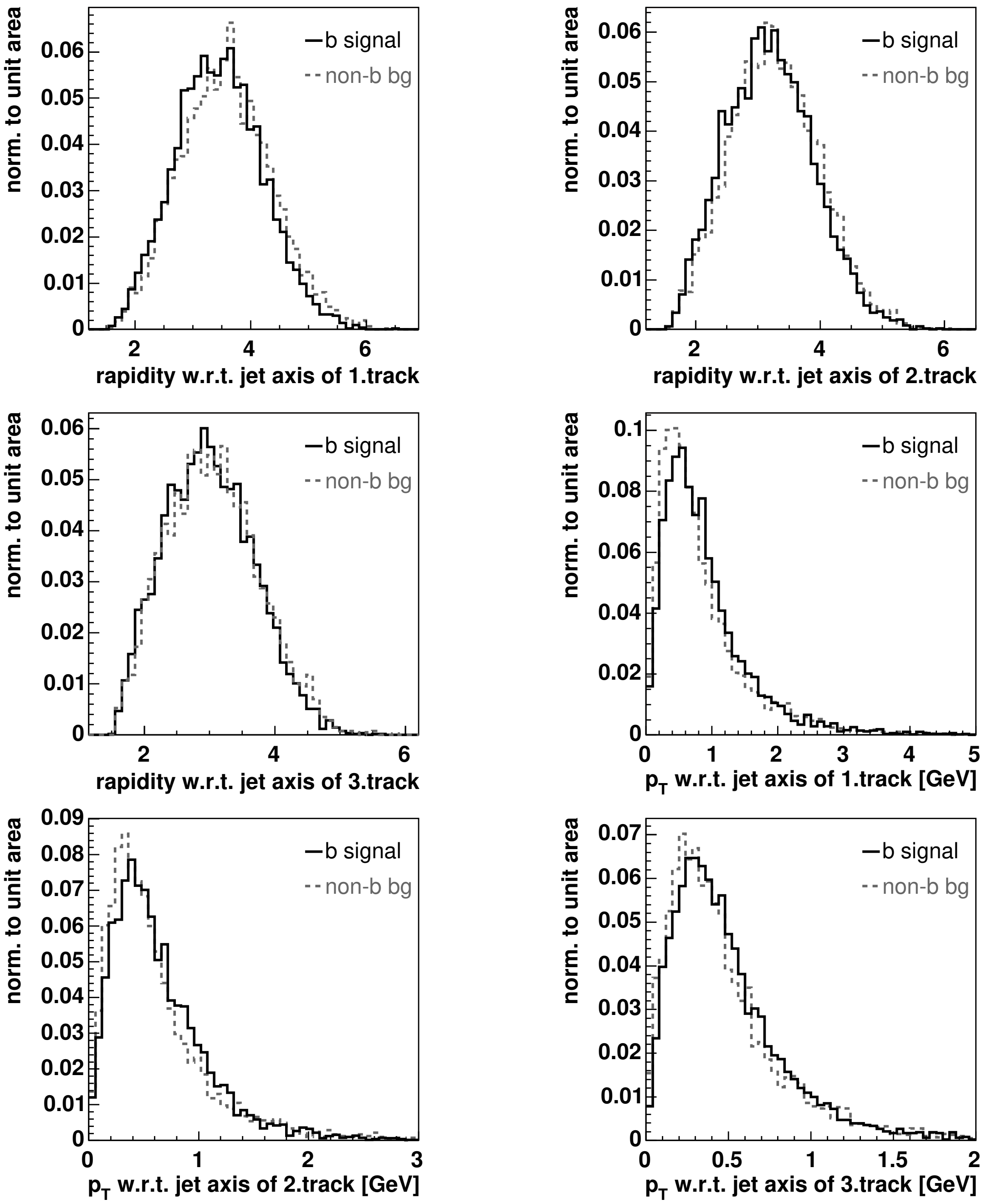


Figure 5: **Distributions of more kinematic discriminating track variables.** In each tagged jet, all the SecVtx tracks were sorted by their momentum. Thus, "1. track" means the track with the highest momentum in the jet and so on. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

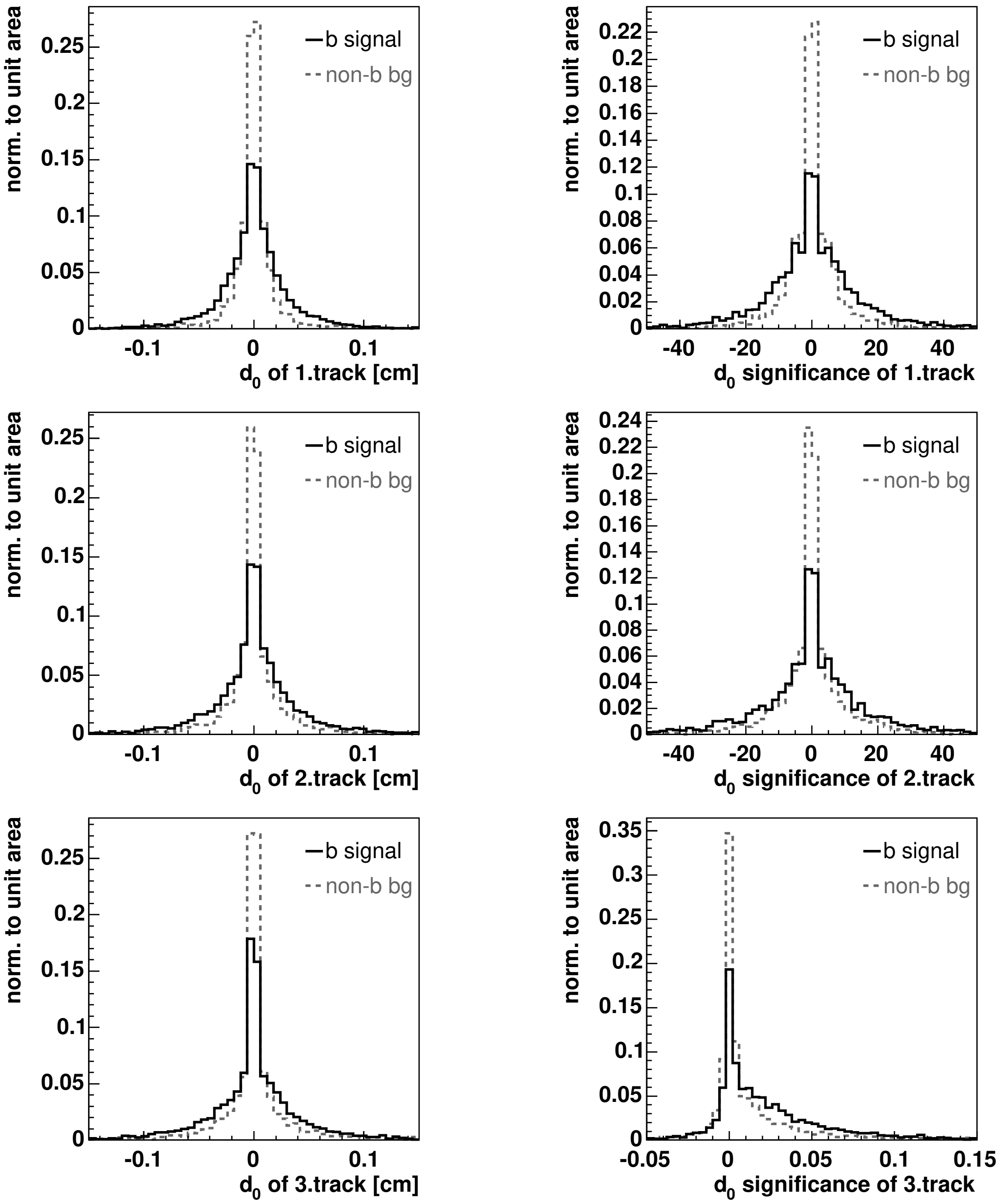


Figure 6: **Distributions of discriminating track variables.** In each tagged jet, all SecVtx tracks are sorted by their momentum. Thus, "1. track" means the track with the highest momentum in the jet and so on. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

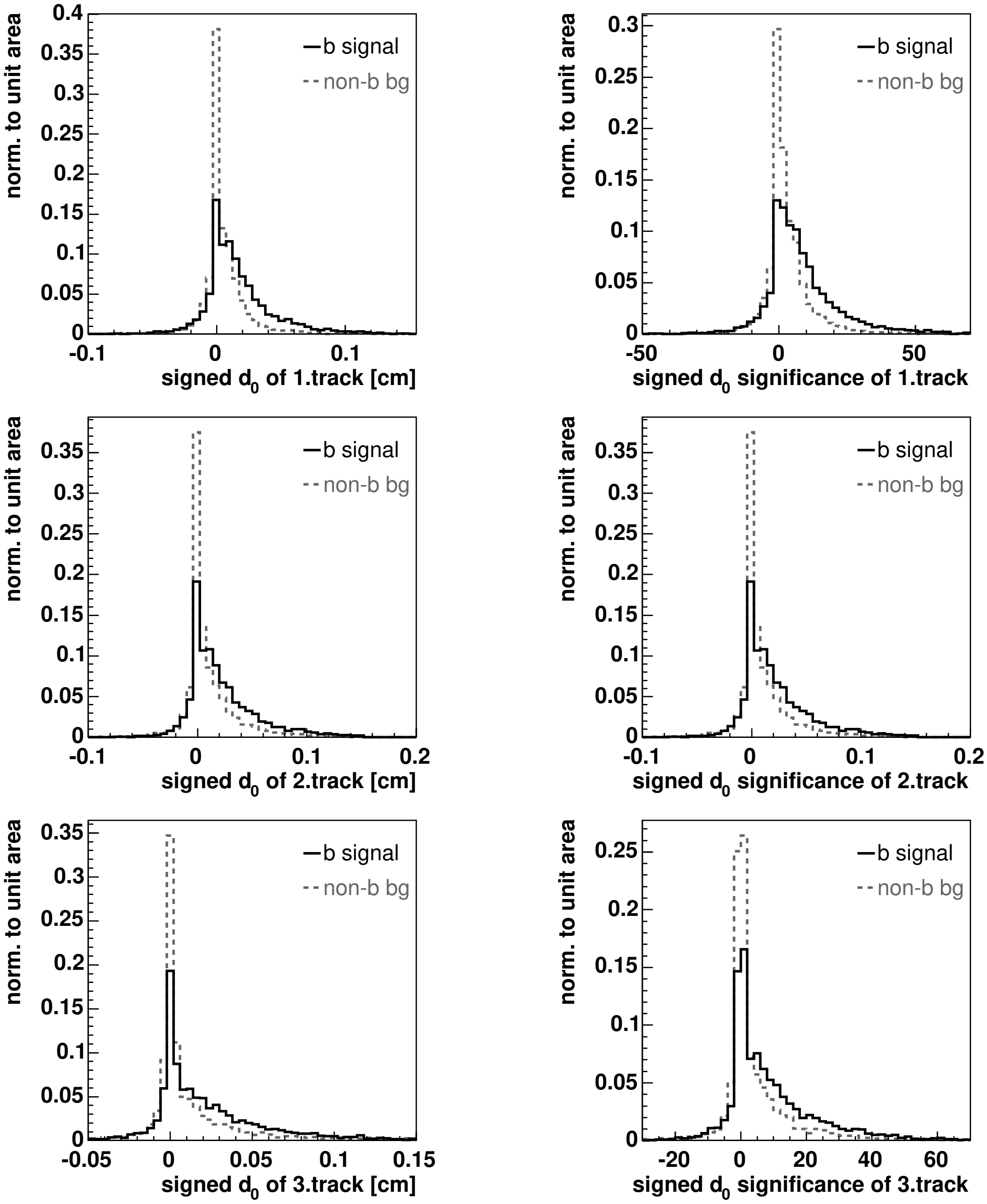


Figure 7: **Distributions of discriminating track variables.** In each tagged jet, all SecVtx tracks are sorted by their momentum. Thus. "1. track" means the track with the highest momentum in the jet and so on. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

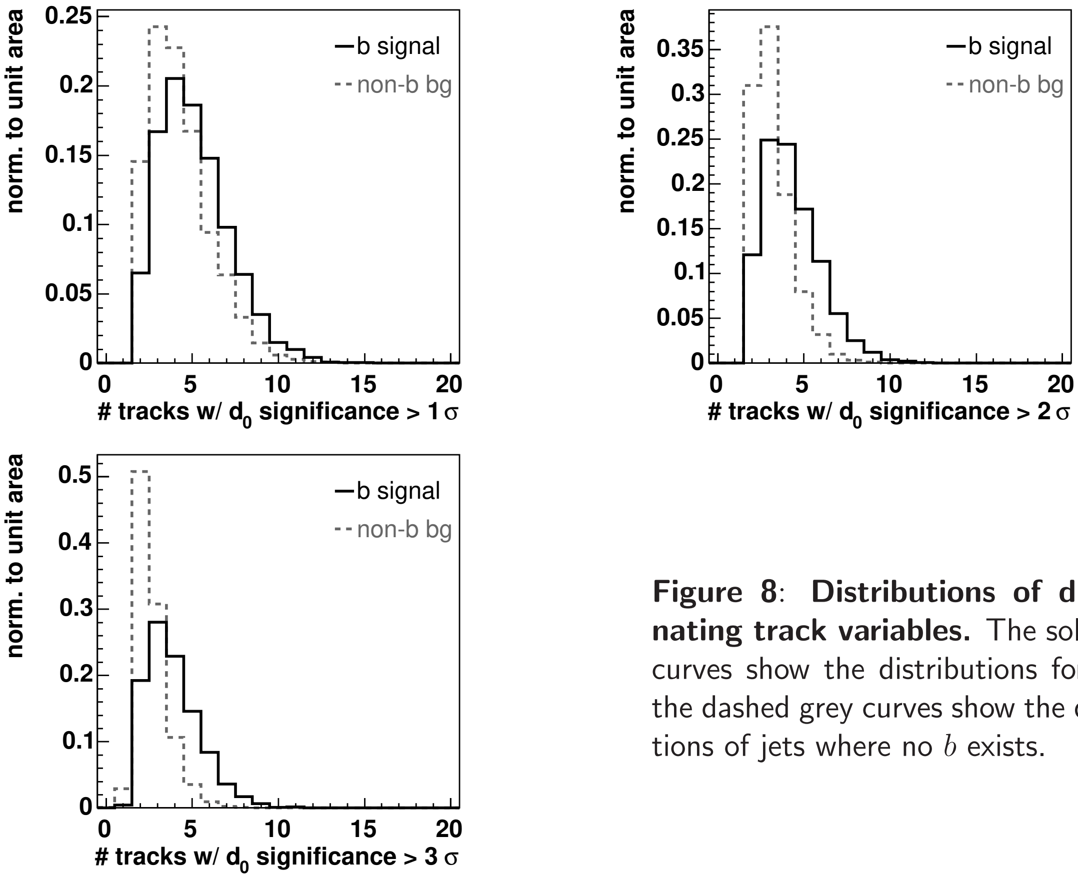


Figure 8: **Distributions of discriminating track variables.** The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

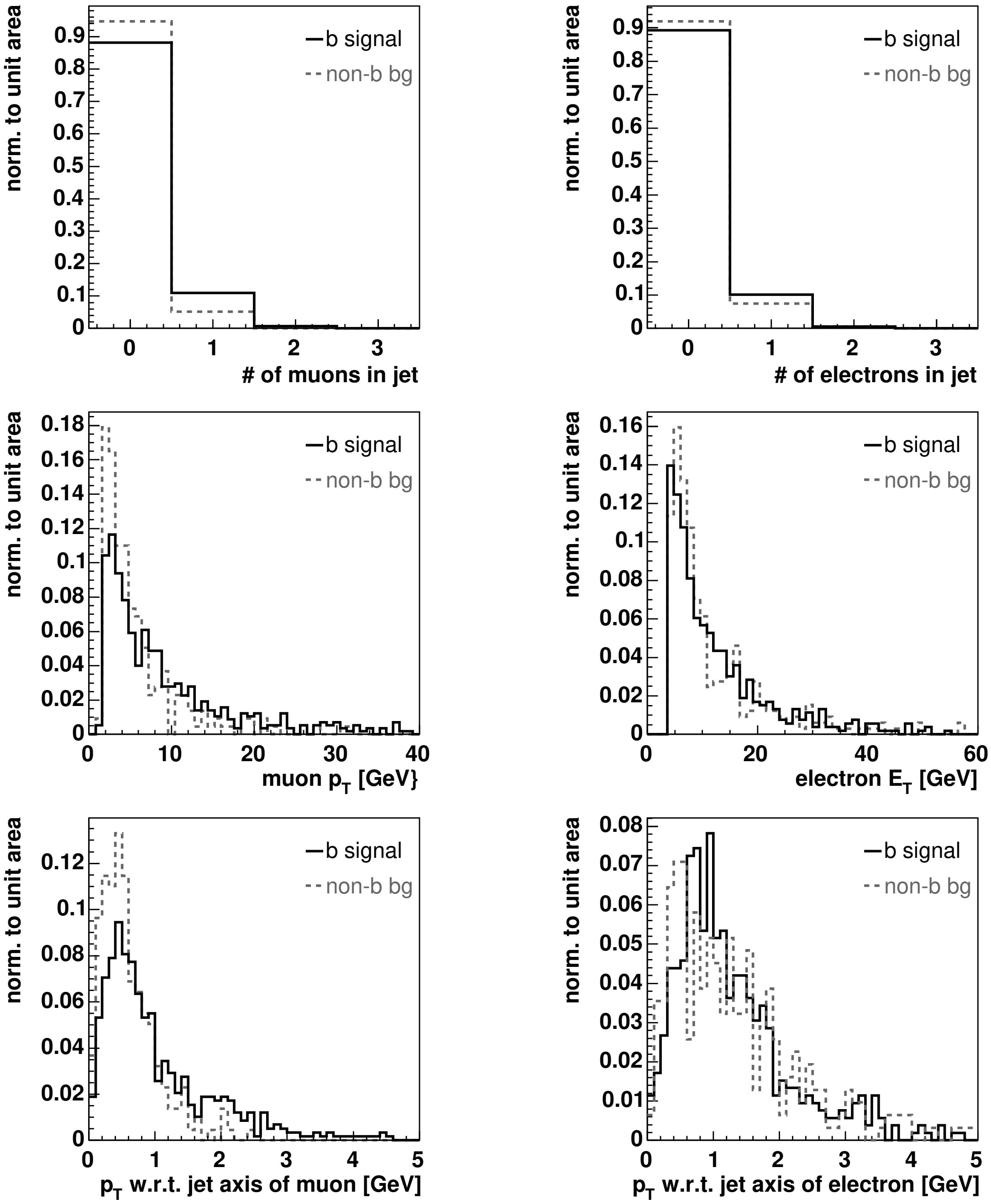


Figure 9: **Distributions of discriminating lepton variables.** If available, the information of leptons in each tagged jet is displayed. In solid black, the leptons in b jets are plotted, in dashed grey the leptons in non–b jets.

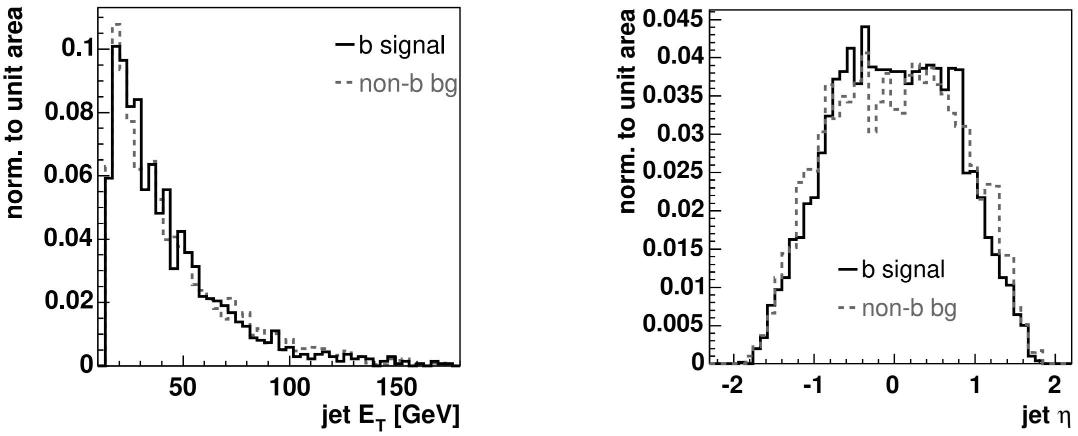


Figure 10: **Distributions of discriminating jet variables.** The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists..

the jet can contain valuable information. From figure 10 it can be seen that the $\mathrm{E_T}$ distributions for b and non–b jets are very similar. We will include this variable only to take into account the correlations to other characteristic quantities.

Many variables show only a small difference between b and non–b and are highly correlated. A cut–based analysis is therefore not an optimal approach. In order to incorporate all the information and all the correlations, the combination of those variables into one test statistic is performed by a neural network.

4 Functionality of the Neural Network

The neural network used for the b tagger described in this note is the NeuroBayes® package [8] provided by the company Phi-T. NeuroBayes® combines a three–layer feed forward neural network (see figure 11 as an example) with a complex robust preprocessing. This preprocessing is performed before the input variables are fed to the neural network (see section 4.2). Bayesian regularization techniques are used for the training process (see section 4.1). There is one input node for each input variable plus one bias node. The number of nodes in the hidden layer can be freely chosen by the user. There is one output node which gives a continuous output in the interval [-1,1].

The nodes of two consecutive layers are connected with variable weights. For each node j, a biased weighted sum of the values of the previous layer x_i is calculated

$$a_j(\mathbf{x}) = \sum_i \omega_{ij} x_i + \mu_{0,j} \tag{1}$$

and passed to the transfer function which gives the output of the node. The bias $\mu_{0,j}$ (which is calculated for each input) implements the thresholds of the several nodes: if the input to a node is larger than its threshold, the node will send an input to the next layer. The output of each node is determined by a transformed sigmoid function

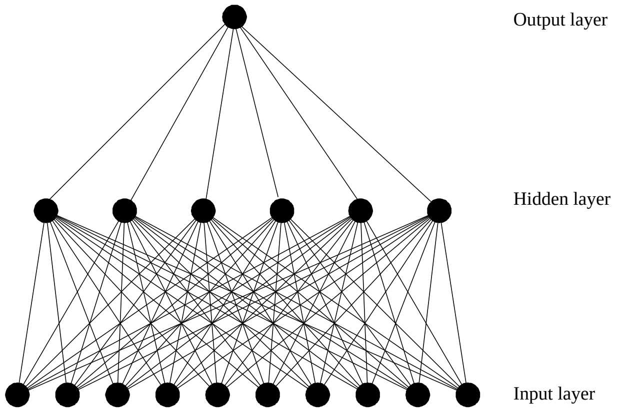


Figure 11: **General geometry of a 3 layer neural network.** There is one input node for each variable plus one bias node, an arbitrary number of hidden nodes and one output node. The nodes in two consecutive layers are catenated with variable connections.

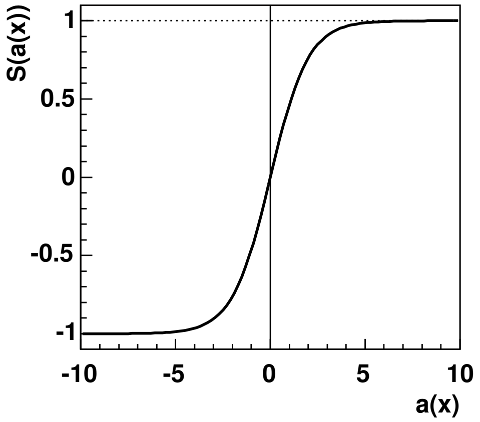


Figure 12: **The transformed sigmoid activation function S(a(x))** as given by equation 2.

$$S(\mathbf{x}) = \frac{2}{1 + e^{-a(\mathbf{x})}} - 1 \qquad (2)$$

which gives an output of -1 for background and +1 for signal. As can be seen in figure 12, the sigmoid function is only sensitive to a relatively small range around zero. By this transformation, the interval $[-\infty, \infty]$ is mapped to the interval [-1,+1]. For very large ($x \to \infty$) or very small ($x \to -\infty$) values, a saturation effect is reached. The bias mentioned above shifts the mean of the input data distribution to the linear part of the sigmoid function.

The output of the neural network for output node k is calculated by

$$o_k = S(\sum_{j=0}^{M} \omega_{jk} \cdot S(\sum_{i=0}^{d} \omega_{ij} x_i + \mu_{0,j})) \qquad (3)$$

where d is the number of input and M the number of hidden nodes. ω_{ij} denotes the weights from the input to the hidden layer, ω_{jk} the weights from the hidden to the output layer. $\mu_{0,j}$ is the weight that connects the bias node with the hidden nodes.

4.1 The training process

The training of the neural network is done by minimizing the deviation between the true output and the one calculated by using the actual weights. The error function minimized in this neural network is the entropy error function, which is essentially given by the sum of the logarithms of the output values. The aim of the training of the neural network is to find the minimum in the multidimensional structure of the error function with many peaks and valleys. As this task can be difficult to solve, the training process is done by the combined method of gradient descent and backpropagation. The neural network is trained with regularization techniques to improve generalization performance and to avoid overtraining. During the training process, weights (and even nodes) that have become completely insignificant are pruned away. This reduces the number of free parameters and hence improves the signal–to–noise ratio by removing the cause of the noise. This leads again to an improved generalization ability. For details of the above mentioned features see Refs. [8, 9].

4.2 Preprocessing of the variables

To find the optimal starting point for minimizing the error function, the input variables are preprocessed. This preprocessing is done in a completely automatic way. To reduce the influence of extreme outliers, the input distributions are equalized to lie between -1 and 1. Those flattened distributions are then converted into a Gaussian distribution, centered at zero with standard deviation 1. This avoids saturation of the nodes due to the above mentioned shape of the activation function (see figure 12). After this transformation, the input variables are linearly decorrelated by diagonalization and rotation of the covariance matrix into a unit matrix.

The above mentioned transformation to a Gaussian distribution may be altered by individual variable preprocessing like fitting the flattened distribution with a spline if this is considered to be sensible. In addition, discrete variables can be treated as members of classes. The preprocessing of those kinds of variables can also deal with a certain order of values if this is important, e.g. the number of tracks in a jet. The preprocessing is also able to deal with variables that are not given for every event by assigning the missing values to a δ–function.

Of course, those preprocessed input variables do not have any physical meaning, unlike the original ones.

4.3 Automatic variable selection

The significances of the training variables are determined automatically during the preprocessing in NeuroBayes®.

The correlation matrix of all preprocessed input variables is calculated including the correlation of all variables to the target. One variable after the other is omitted to determine the loss of total correlation to the target caused by its removal. The variable with the smallest loss of correlation is discarded leading to an (n-1)–dimensional correlation matrix. The same procedure is repeated with the reduced correlation matrix to find the least important of the (n-1) remaining variables. At the end, we get a list of ordered variables as it is shown in tables 3 and 4.

The significance of each variable is calculated by dividing the loss of correlation induced by its removal at the relevant point in the procedure described above by the square root of the sample size, i.e those significances are relative numbers in terms of the reduced correlation matrices.

After the preprocessing process, it is possible to cut on the significance of the variables. Only those variables are taken into account that include relevant information that is not already incorporated by other variables.

5 Training of the neural network

The neural network is trained with single–top Monte Carlo samples and Monte Carlo samples of the main backgrounds. Besides the t– and s–channel of single-top quark production, the $t\bar{t}$ and the $Wb\bar{b}$ backgrounds are used as b signal. As non–b background, the $Wc\bar{c}$, the Wc and the mistags from $W + light$ jets production are taken. From the single–top, $t\bar{t}$ and $Wb\bar{b}$ samples we use tagged jets that contain at least one track in the OBSP bank identified as a b daughter. The mistags are obtained by the $W1p$, $W2p$, $W3p$ and $W4p$ samples by rejecting all events that do inclose a b or charmed hadron using the Monte Carlo truth information given by the HEPG bank. For the Wc and $Wc\bar{c}$ samples we required at least one track in the tagged jet identified as a c daughter by the Monte Carlo truth information.

Those samples are mixed with fractions close to the background estimation of the single-top quark production[2] (see table 2). As the statistics of the mistag samples are very low, we weighted those samples with a weight of 4 for the training process.

In a first approach, 50 variables were investigated. Those variables are shown in tables 3 and 4 sorted by their significance (determined as described in section 4.3). Some of the variables introduced in section 3 are very different for pass 1 and pass 2 secondary vertices due to the vertex fitting procedure (see Refs. [3, 4] for details). Even if the neural network can deal with those kind of dependences, we decided to teach it explicitly the differences between the two kinds of secondary vertices by splitting up some of the most "pass specific" variables. For those distributions, we used two separate input variables. One is only filled for pass 1 vertices, the other one only for pass 2 vertices. The comparison of the distributions of those variables for pass 1 and pass 2 can be seen in figures 13 and 14. In addition, we decorrelated those variables from the binary variable "pass 1 or pass 2" such that the neural network only takes into account the additional correlation to the target.

The number of input variables was reduced to 25 by a cut on the significance. Only variables with a significance larger than 3 σ are taken for the training of the network (see table 3). Those are some of the variables plotted in figures 1, 2, 3, 4, 5, 6, 7, 8 and 10. For the lepton information (shown in figure 9), only the transverse momenta of the muons

[2]If the *a priori* probability is correct, the output of a well–trained neural network can be interpreted as Bayesian *a posteriori* probability.

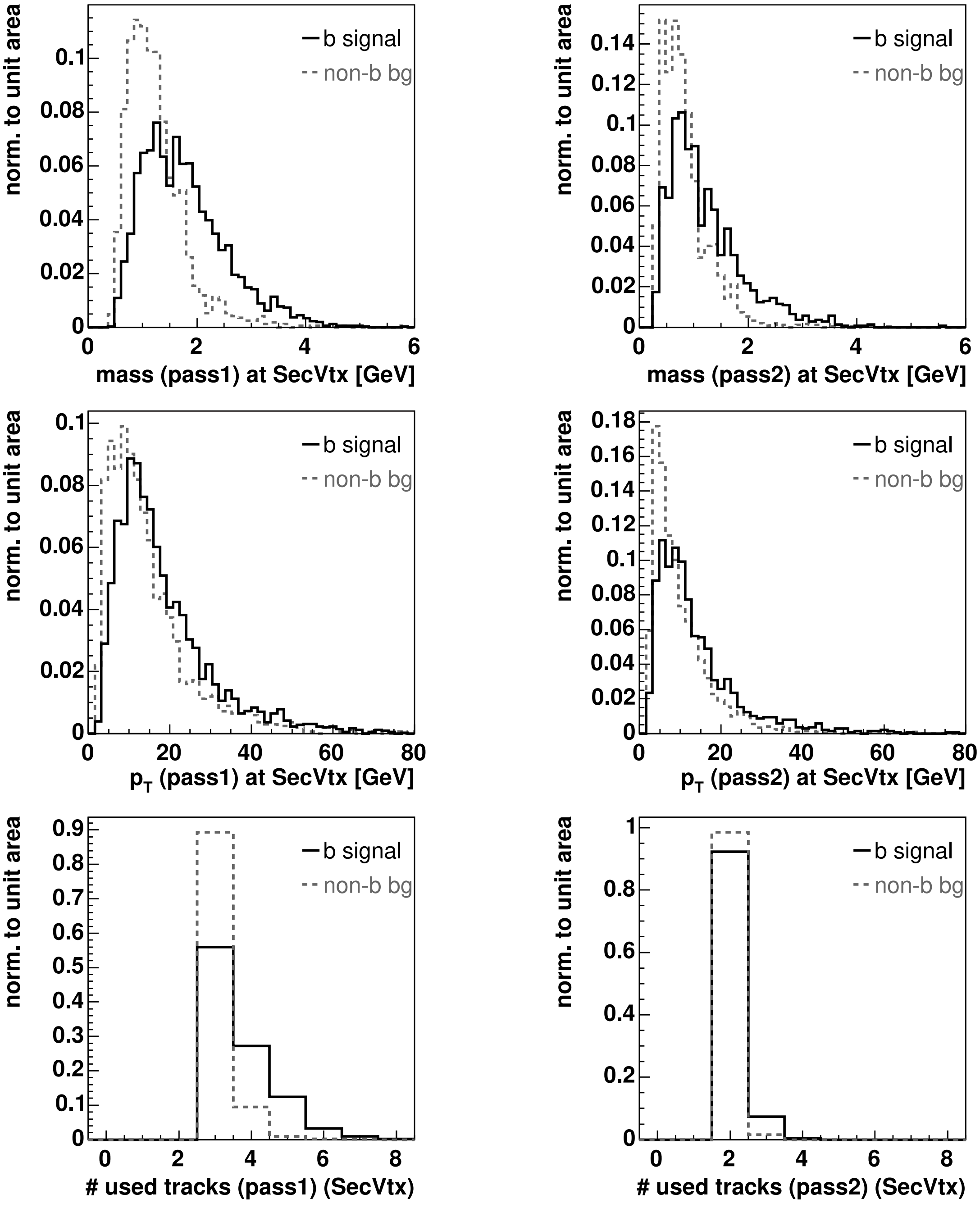


Figure 13: **Distributions of discriminating SecVtx variables for pass 1 and pass 2.** Distributions of some SecVtx variables split up for the 2 passes of the vertex fit. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

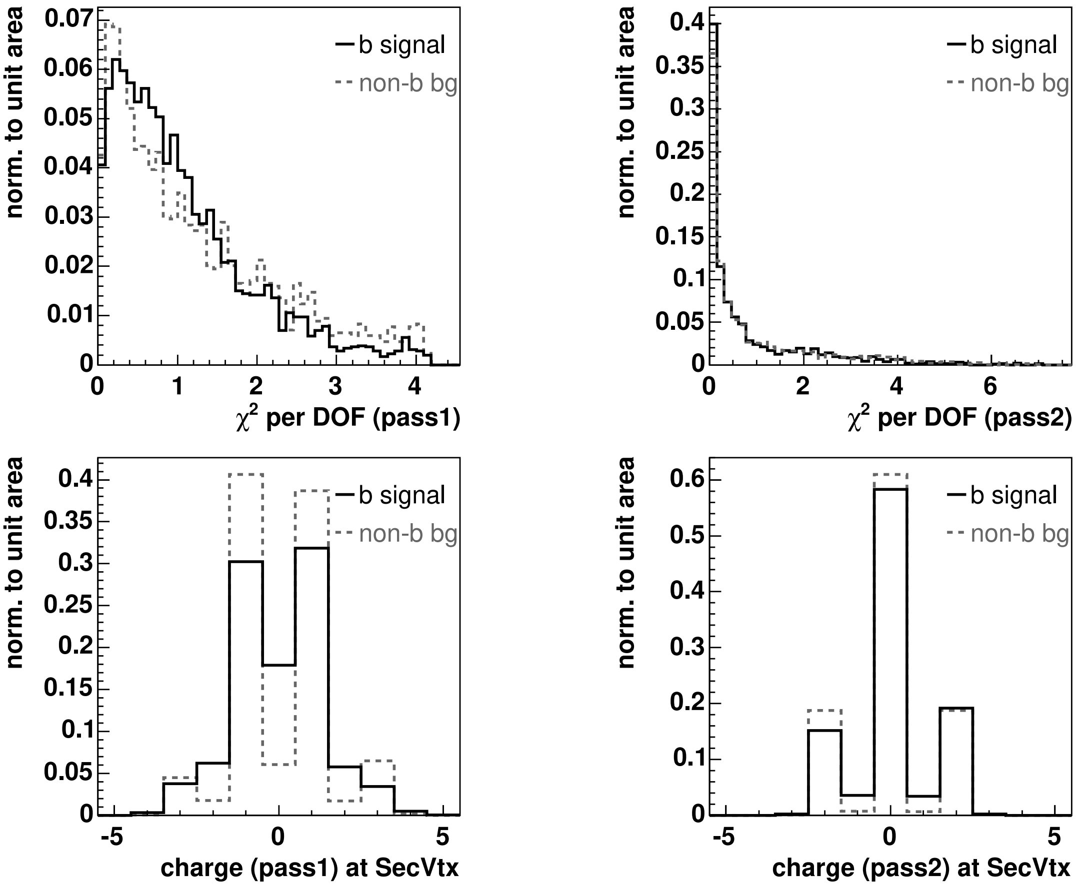


Figure 14: **More distributions of discriminating SecVtx variables for pass 1 and pass 2.** Distributions of some SecVtx variables split up for the 2 passes of the vertex fit. The solid black curves show the distributions for b jets, the dashed grey curves show the distributions of jets where no b exists.

b signal	Fraction of the complete sample
t–channel	2.4%
s–channel	2.6%
$t\bar{t}$	24.7%
$Wb\bar{b}$	28.9%
Σ b signal	58.6%
non–b background	Fraction of the complete sample
$Wc\bar{c}$	10.1%
Wc	9.7%
Mistags	21.6%
Σ non–b background	41.4%

Table 2: **Ratios of the training samples.** This table shows the mixing ratios of the training samples for the neural network. The numbers are obtained by the background estimation for the single-top quark analysis [6, 7]. As the statistics for the mistags are very low, we obtained the according ratio by giving them a weight of 4 in the training.

and electrons w.r.t. the jet axis survived the significance cut. The way the cut on the significance was determined is described in section 5.1.

The complete set of variables used for the training are:

- pass of the SecVtx fit
- mass at the SecVtx for both passes
- p_T of the SecVtx for pass 1 vertices
- pseudo $c\tau$ of the SecVtx
- χ^2 per DOF of SecVtx fit for both passes
- number of good tracks (SecVtx)
- number of used tracks (SecVtx) for both passes
- L_{xy} of the SecVtx and its significance
- L_{xyz} significance of the SecVtx
- number of tracks in the jet with d_0 significance > 1 and 3 σ
- lifetime–signed d_0 significance of track 1, 2 and 3*
- rapidity w.r.t the jet axis of 1.track
- p of 1.track

*According to the ordering by their momentum.

Rank	Variable	Significance (in σ)
1	# tracks with d_0 significance > 3 σ	80.5
2	signed d_0 significance of 1.track	42.3
3	signed d_0 significance of 2.track	27.1
4	signed d_0 significance of 3.track	22.7
5	mass at SecVtx (pass 2)	21.1
6	mass at SecVtx (pass 1)	20.6
7	pass of SecVtx fit (1 or 2)	19.7
8	p_T of muon w.r.t jet axis	14.5
9	$c\tau$ of SecVtx	13.3
10	L_{xy} significance	2.7
11	L_{xyz} significance	10.4
12	p_T of SecVtx (pass 1)	9.3
13	jet E_T	10.1
14	# used tracks in SecVtx fit (Nvtrk) (pass 1)	9.7
15	rapidity w.r.t jet axis of 1.track	6.5
16	# good tracks in jet (Ntrk)	6.1
17	L_{xy} of SecVtx	6.0
18	p_T of electron w.r.t jet axis	5.8
19	# tracks with d_0 significance > 1 σ	5.2
20	# used tracks in SecVtx fit (Nvtrk) (pass 2)	4.0
21	χ^2 per DOF of SecVtx fit (pass 2)	4.2
22	χ^2 per DOF of SecVtx fit (pass 1)	3.9
23	p of 1.track	3.6
24	p_T w.r.t jet axis of 3.track	3.3
25	jet η	3.2

Table 3: **Used set of discriminating variables sorted by significance.** These most significant variables are taken for the training of the neural network. Only variables with significances larger than 3 σ are used. The variables which didn't survive the significance cut are listed in table 4. The significances are determined as described in section 4.3.

Rank	Variable	Significance (in σ)
26	rapidity w.r.t jet axis of 2.track	2.7
27	d_0 of 1.track	2.4
28	signed d_0 of 3.track	2.3
29	d_0 of 3.track	2.4
30	rapidity w.r.t jet axis of 3.track	2.1
31	signed d_0 of 2.track	1.7
32	d_0 of 2.track	2.0
33	signed d_0 of 1.track	1.7
34	p_T w.r.t jet axis of 1.track	1.8
35	p of 3.track	1.4
36	electron E_T	1.3
37	# electrons in jet	1.2
38	# tracks with d_0 significance > 2 σ	0.3
39	# pass 1 tracks in jet (Ndtrk)	1.1
40	d_0 significance of 1.track	1.1
41	p_T of SecVtx (pass 2)	1.0
42	p of 2.track	0.9
43	d_0 significance of 3.track	0.8
44	p_T w.r.t jet axis of 2.track	0.2
45	muon p_T	0.3
46	d_0 significance of 2.track	0.2
47	Charge at SecVtx (pass 2)	0.1
48	# muons in jet	0.1
49	L_{xyz} of SecVtx	0.1
50	Charge at SecVtx (pass 1)	0.0

Table 4: **Set of discarded variables sorted by significance.** These variables didn't survive the cut on the significance > 3 σ. The significances are determined as described in section 4.3.

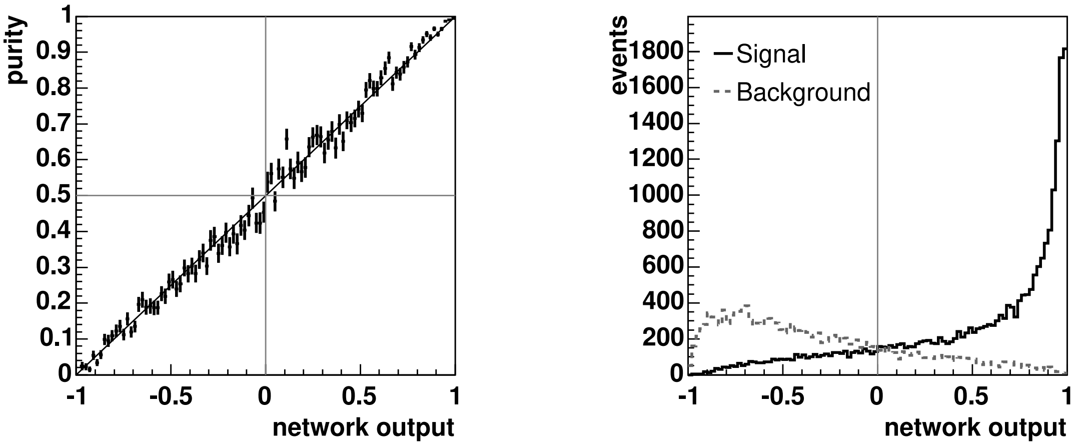


Figure 15: **Result of the training.** On the left hand side, the purity in each output bin is plotted versus the output. On the right hand side, the output distribution of the trained neural network is shown for signal and background.

- p_T w.r.t. the jet axis of 3.track
- p_T of muon w.r.t. the jet axis
- p_T of electron w.r.t. the jet axis
- E_T and η of the jet

All the other variables thought to be useful (listed in table 4) include only little extra information that is not already provided by another variable. Therefore, they are discarded.

The training of this neural network is done as described in section 4; the network structure is the following: 26 input nodes, 10 hidden nodes and one output node. Different numbers of hidden nodes were tested but did not result in better discrimination.

The quality of the training can be checked by plotting the signal purity in each output bin versus the output (see figure 15, left). If the network is well trained, all the points should lie on the diagonal (see NeuroBayes® documentation [8] for details). One can see that this condition is fulfilled. That means that the output of the net can be interpreted as the probability that a pattern is signal or background under the assumption that the network was trained with a realistic mixture of signal and background (i.e. correct *a priori* probability).

The output of the trained network is shown in figure 15 on the right hand side. One can see a very good separation between signal and background. The b signal events are cumulated at output 1, the non–b background events at output -1 as expected according to section 4.

The output distributions for the different samples are shown in figure 16. The Monte Carlo truth is used as described before. One can see that there is no difference for the distributions of the b signal samples. That indicates that the network is completely insensitive to the origin of a b jet. The same is true for the charm samples.

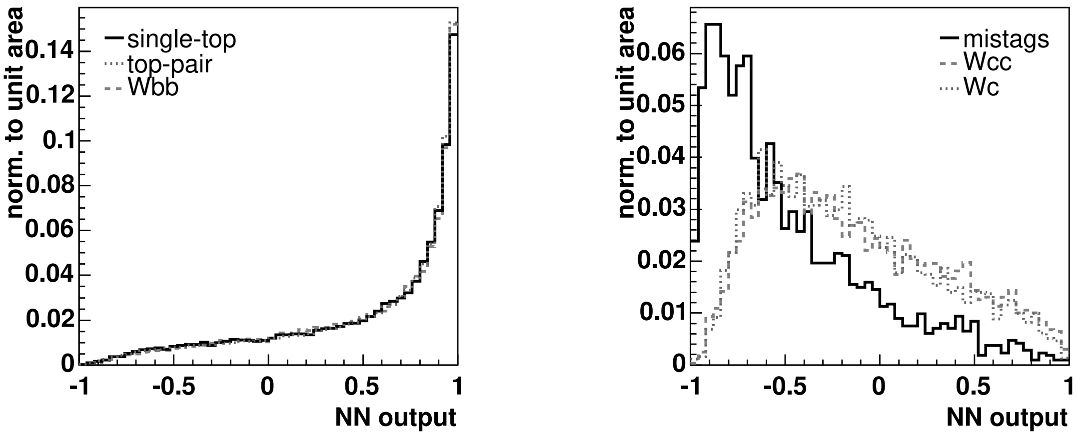


Figure 16: **Output distributions for the different samples.** On the left hand side, the output distributions for single–top, $t\bar{t}$ and $Wb\bar{b}$ is shown. On the right hand side, the output distributions are shown for $Wc\bar{c}$, Wc and mistags. MC truth is used as described in the text.

In figure 17 the output distributions of the training samples are shown split up into pass 1 and pass 2 secondary vertices. Again, they agree perfectly well. In addition, it can be seen that pass 1 vertices are in general more signal–like than pass 2 vertices. This feature is mainly attributed to the fact that a pass 1 vertex consists at least of three tracks while a pass 2 vertex mainly has only two tracks. This behavior was already visible when we trained the neural network without teaching it the differences between pass 1 and pass 2 vertices explicitly.

If one plots the signal purity versus the signal efficiency (shown in figure 18), one can see the potential improvement of the b tagging purity by the application of the neural network. After the requirement of a SecVtx tag, only about half of the secondary vertices in the training samples are related to the decay of a b hadron. With the application of the neural network, this value can be improved up to more than 95% but this is at the expense of losing about 50% of the signal events (see plot 18). This shows that one has to make a compromise between purity and efficiency. How this is done for the single–top analysis is explained in section 6.

5.1 Determination of the cut on the significance of the input variables

The cut on the significance of the input variables was determined by training several networks and reducing the number of input variables step by step. The plots showing the b signal purity in dependence of the b signal efficiency for trainings with different cuts on the significance are shown in figure 19. The training with a cut on the significance of the input variables $> 0.5\ \sigma$ excludes only variables that do not include any significant information but statistical fluctuations. Therefore, we used this training for the comparison with trainings with harder cuts on the significance of the input variables.

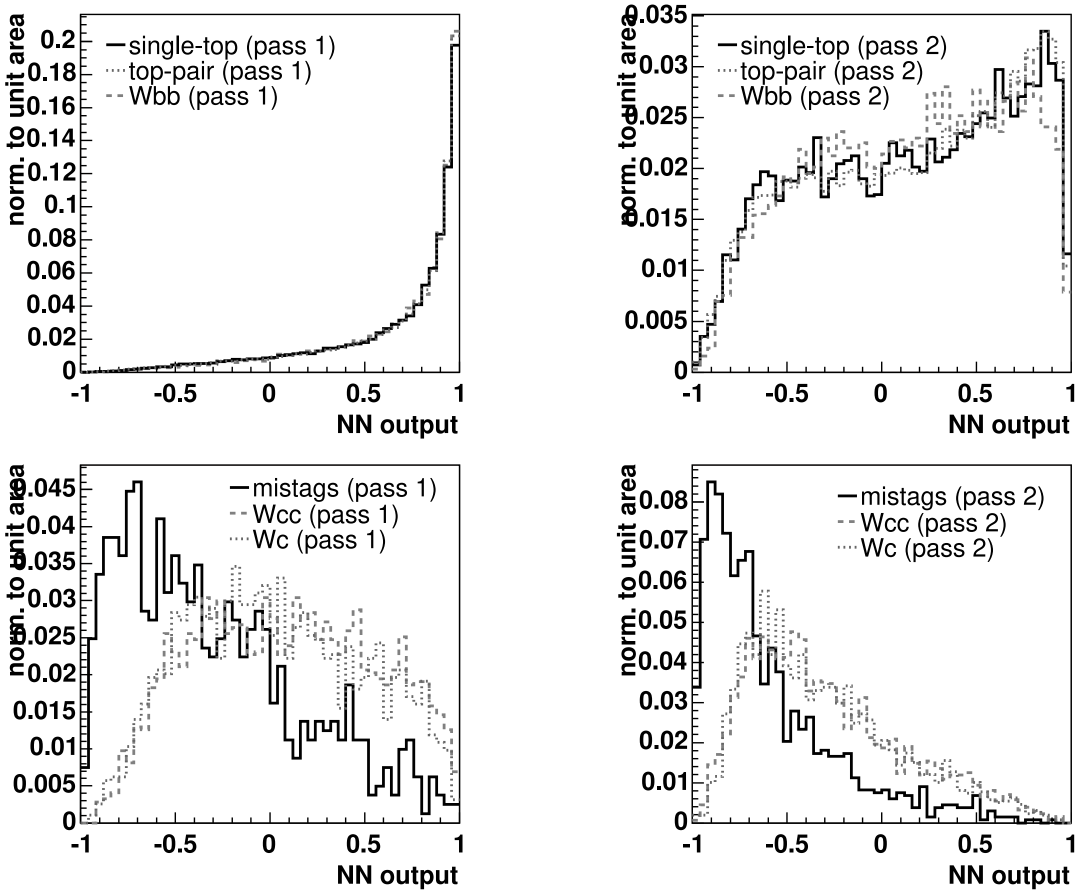


Figure 17: **Output distributions for the different samples.** The upper plots show the output distributions for single–top, $t\bar{t}$ and $Wb\bar{b}$ for pass 1 (left) and pass 2 (right) vertices. The lower plots show the output distributions for $Wc\bar{c}$, Wc and mistags for pass 1 (left) and pass 2 (right) vertices. MC truth is used as described in the text.

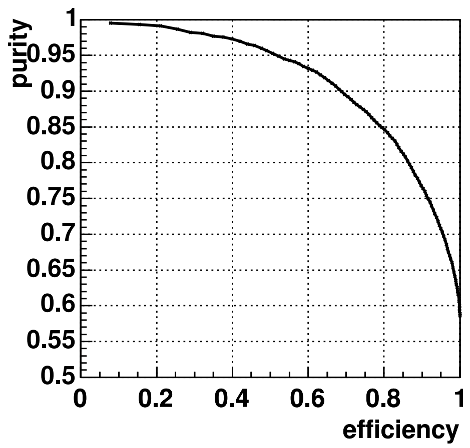


Figure 18: **Signal purity vs. signal efficiency.** The purity of the sample is plotted against the signal efficiency.

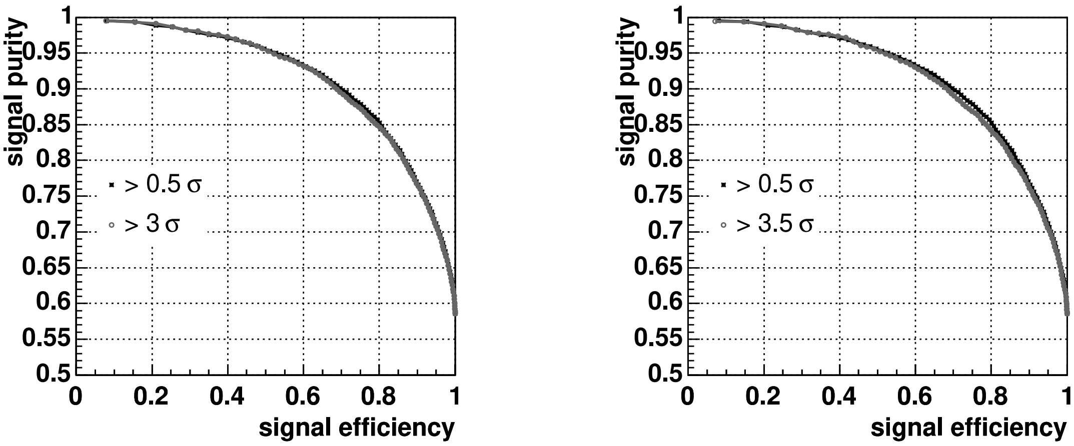


Figure 19: **Signal purity vs. signal efficiency for the different trainings.** The black curves describe the training w/ the cut $> 0.5\ \sigma$, the grey ones describe the trainings w/ harder cuts on the significance.

The training with a cut at 3 σ (figure 19, upper left hand side) overlaps almost completely with the training using the 43 variables listed in tables 3 and 4 with a significance $> 0.5\ \sigma$. For harder cuts on the significance, the deviations are getting larger. This fact led to the decision of taking the most important 25 variables to train the neural network by using only variables with a significance $>3\ \sigma$.

6 Determination of the working point

To find the optimal compromise between tagging purity and tagging efficiency one has to determine an analysis specific cut on the network output. This working point for the single–top analysis was found by optimizing the ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top.

The number of events after the requirement of a SecVtx tag ($\mathrm{N}_{\mathrm{exp}}^{\mathrm{SecVtx}}$) is obtained by the background estimation for the single–top analysis [6, 7] and is shown in table 5 after at least one SecVtx b tag was required.

The efficiency $\epsilon_{\mathrm{sample}}^{\mathrm{NN}}(\mathrm{NNcut})$ for a specific cut on the neural network output is evaluated for the different samples. The expected number of events for each sample after the additional application of the neural network can be calculated via:

$$\mathrm{N}_{\mathrm{exp}}^{\mathrm{NN}} = \epsilon_{\mathrm{sample}}^{\mathrm{NN}}(\mathrm{NNcut}) \times \mathrm{N}_{\mathrm{exp}}^{\mathrm{SecVtx}}$$

The ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top is calculated with both s– and t–channel as signals. For the non–W and diboson backgrounds, we assume the NN efficiency $\epsilon_{\mathrm{non-W}}^{\mathrm{NN}}(\mathrm{NNcut})$ and $\epsilon_{\mathrm{Di-Boson}}^{\mathrm{NN}}(\mathrm{NNcut})$ to be the average efficiency of the b signal samples $\epsilon_{\mathrm{t-channel}}^{\mathrm{NN}}(\mathrm{NNcut})$, $\epsilon_{\mathrm{s-channel}}^{\mathrm{NN}}(\mathrm{NNcut})$, $\epsilon_{\mathrm{t\bar{t}}}^{\mathrm{NN}}(\mathrm{NNcut})$ and $\epsilon_{\mathrm{Wb\bar{b}}}^{\mathrm{NN}}(\mathrm{NNcut})$ to consider the worst case.

A cut on the output at 0.0 is found to give the optimal ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top. The calculated optimized ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ is shown in table 6 together with the one before the

sample	expected number of events
t–channel	10.7
s–channel	7.2
$t\bar{t}$	65.4
$Wb\bar{b}$	101.4
$Wc\bar{c}$	38.7
Wc	38.3
Mistags	86.2
non–W	21.6
Di–Boson	6.79

Table 5: **Expected number of events according to single–top background estimation.** The expected number of events are shown after requiring at least one b tag. Those numbers are estimated for 320 pb^{-1} for the 2– and 3–jet bin together [6, 7].

sample	$\frac{S}{\sqrt{B}}$
before cut	0.95
w/ NN-output > 0.0	1.04
improvement	9.5%

Table 6: **Ratio of $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top.** The ratio is shown before the application of the neural network and after the application of the neural network. For the determination of the ratio with the cut on the NN output, the efficiencies shown in table 7 are taken. For non–W and di–boson, the average b signal efficiency is assumed. Both s– and t–channel are taken as signal.

application of the network. The improvement is 9.5%. This improvement has the same order of magnitude as achieved by other cuts applied in the single–top analysis (see [2]).

Table 7 shows the percentages of remaining events after the additional application of the neural network for the optimized cut on the output. In addition to the optimal cut, two more cuts are presented that lead ratios $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ with a deviation from the optimal ratio only at the per mille level. One can see that about 87% of b signal events have an output > -0.2 while more than 50% of the charmed background events are rejected by this cut. For the mistags, the rejection efficiency is about 78%. Having a look at figure 18 shows that the b tagging purity is improved from 57% to 77%.

When cutting on the output at 0.0, we keep about 82% of the b signal events but reject more than 60% of the charm and more than 85% of the light jets with a secondary vertex. This cut would lead to a b purity of about 85% as can be seen in figure 18.

In addition, table 7 indicates that the network is not very sensitive to the origin of b jets since the efficiencies are almost the same for the different b samples (t–channel, s–channel, $t\bar{t}$ and $Wb\bar{b}$). That is expected according to the output distributions shown in figure 16.

In summary, a large improvement of the b–tagging purity is achieved using our neural

Sample	output > -0.2	output > -0.1	output > 0.0
t–channel	87.4%	84.6%	82.1%
s–channel	87.1%	84.2%	81.4%
$t\bar{t}$	85.7%	82.9%	80.1%
$Wb\bar{b}$	88.1%	85.4%	82.6%
$Wc\bar{c}$	49.7%	43.5%	37.2%
Wc	48.2%	40.8%	34.5%
Mistags	22.2%	17.5%	14.0%

Table 7: **Efficiencies of the best cuts for the single–top analysis.** The percentages of remaining events for the three best cuts on the neural network output are shown for the single–top analysis.

network. Before it can be applied to data, a determination of a correction function is necessary to correct for different efficiencies in data and simulation in analogy to what it is done by the SecVtx scale factor.

7 Samples used for data–MC comparison

For the comparison of data and MC, the same control samples are taken as for the calculation of the SecVtx scale factor. A more detailed description of those samples can be found in the corresponding CDF notes [10, 11].

Inclusive 8 GeV Electron Data We use the 8 GeV inclusive electron sample from the blpc0d dataset. This sample has been stripped for an 8 GeV electron and two 10 GeV jets.

Monte Carlo We are using some generic $2 \to 2$ scattering QCD events with a minimum outgoing transverse momentum of 15 GeV where all flavors are generated (btop5a, btop6a, btop7a, btop8a). A filter requiring a 7 GeV electron or muon in generator parton list discards most light flavor events. In this sample, unlike data, every event contains a lepton that was not created by material interactions. The events are accepted or rejected according to a trigger parameterization of the 8 GeV L2 lepton trigger (see section 7.0.1). This sample was generated with HERWIG 6.4.

One problem that can occur using these Monte Carlo samples is that they only include leading order processes. Monte Carlo studies [12] showed that leading order processes do not provide an adequate description of $b\bar{b}$ production, especially the ϕ–correlations.

In the future, we will check whether this is relevant for the kinematic region of events selected in the single–top analysis.

7.0.1 Parameterization of the 8 GeV L2 lepton trigger

The Monte Carlo events are accepted or rejected according to a trigger parameterization of the 8 GeV L2 lepton trigger. These parameterizations describe the trigger's efficiency

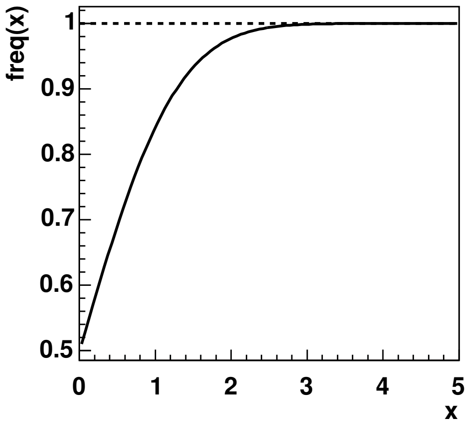


Figure 20: The normal frequency function: $freq(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$

to identify an electron based on its fully reconstructed E_T, p_T and, in part, calorimeter isolation. The parameterization we have chosen is the one that corresponds to the moderate isolation cut (Isol$_{0.4}$<4) [10], although the shape of the function does not vary strongly with isolation. For such electrons, the trigger efficiency is given by (4) [13]:

$$freq\left((\sqrt{E_T} - \sqrt{E_{T0}})/(2\sigma_E)\right) \times freq\left((p_T - p_{T0})/(2\sigma_p p_T p_{T0})\right) \times \left(1 - e^{(2.0 - E_T)/(2\sigma_{E2})}\right) \tag{4}$$

with

$E_{T0} = 7.93$
$\sigma_E = 0.096$
$\sigma_{E2} = 1.10$
$p_{T0} = 7.56$
$\sigma_p = 0.0139$

The base unit is GeV [13].

The normal frequency function freq(x)

$$freq(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

is shown in figure 20.

8 Comparison of distributions of training variables

A first step is to compare the distributions of the training variables in simulation and data.

The distributions of all training variables (see tables 3 and 4) are plotted in figures 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 and 32 for data and MC. For most variables one can observe an almost perfect agreement. The only deviations can be seen in the

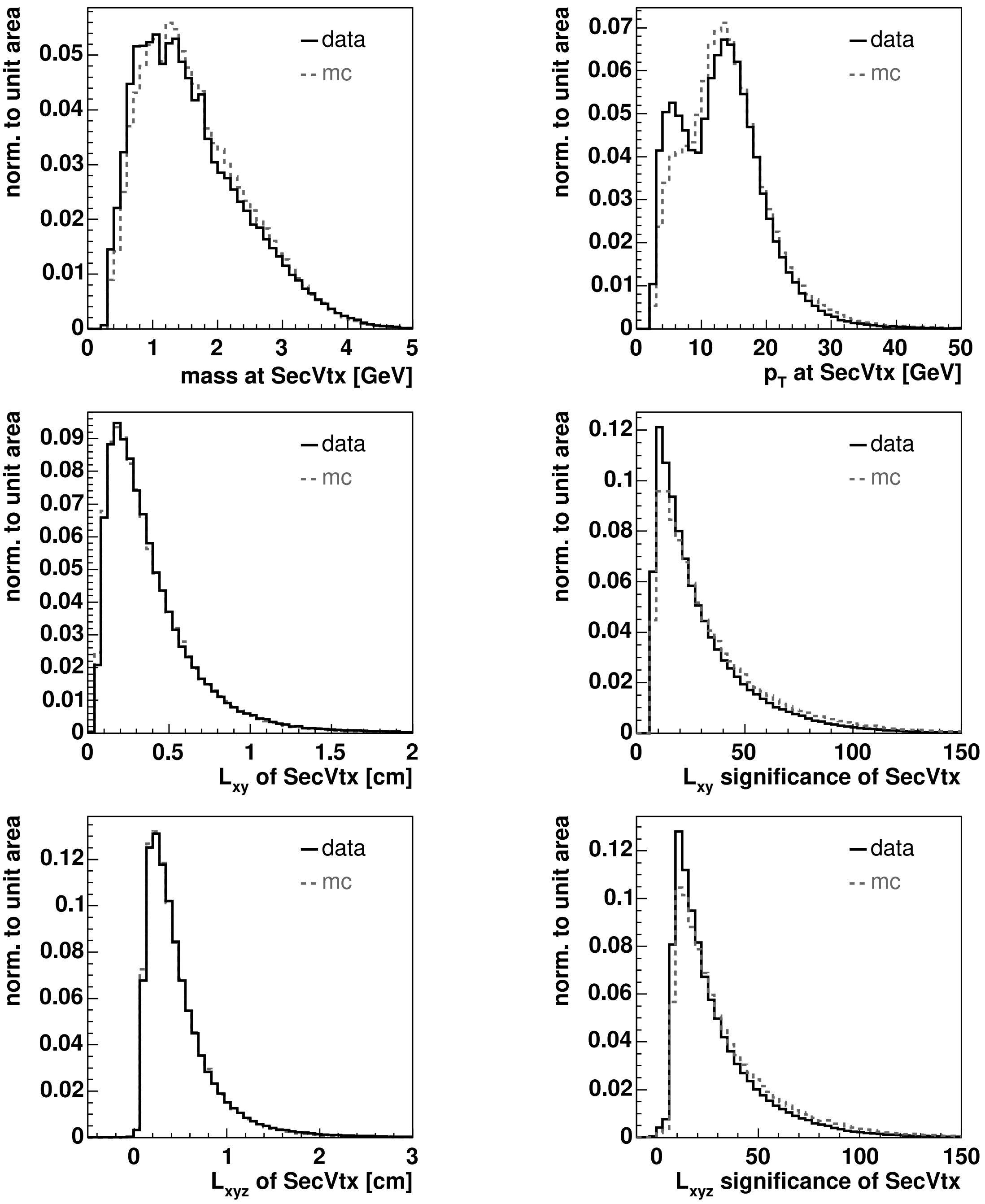


Figure 21: **Comparison of discriminating SecVtx variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 1.

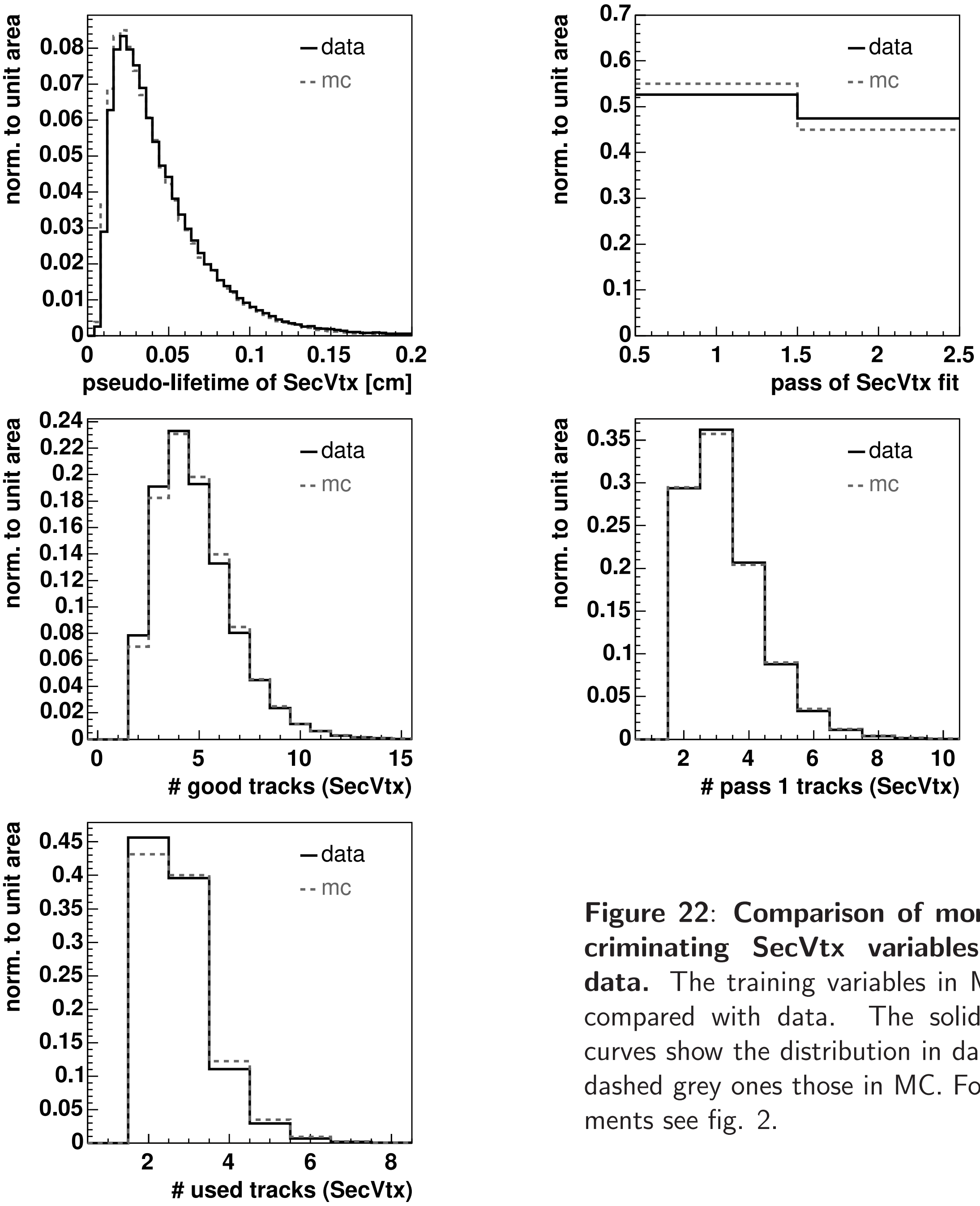


Figure 22: **Comparison of more discriminating SecVtx variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 2.

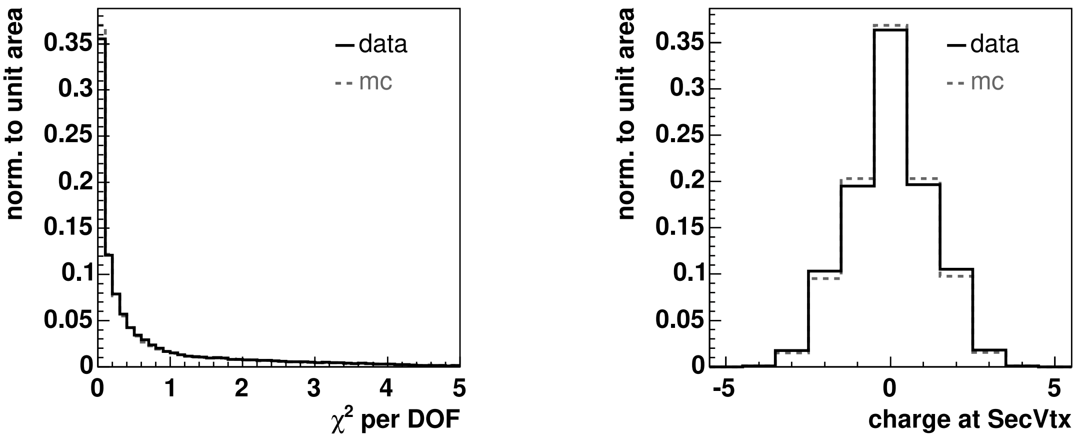
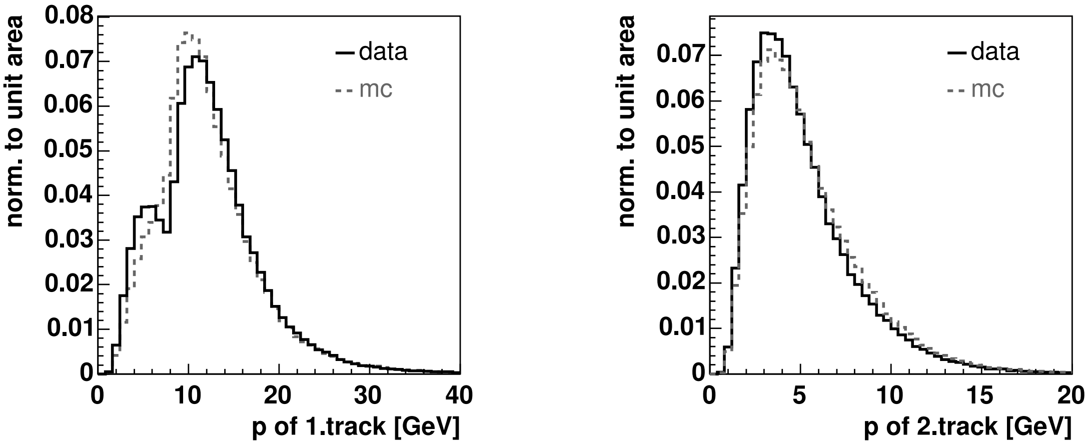


Figure 23: **Comparison of more discriminating SecVtx variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 3.



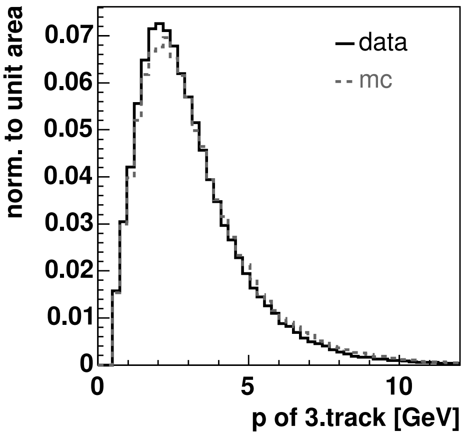


Figure 24: **Comparison of kinematic discriminating track variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 4.

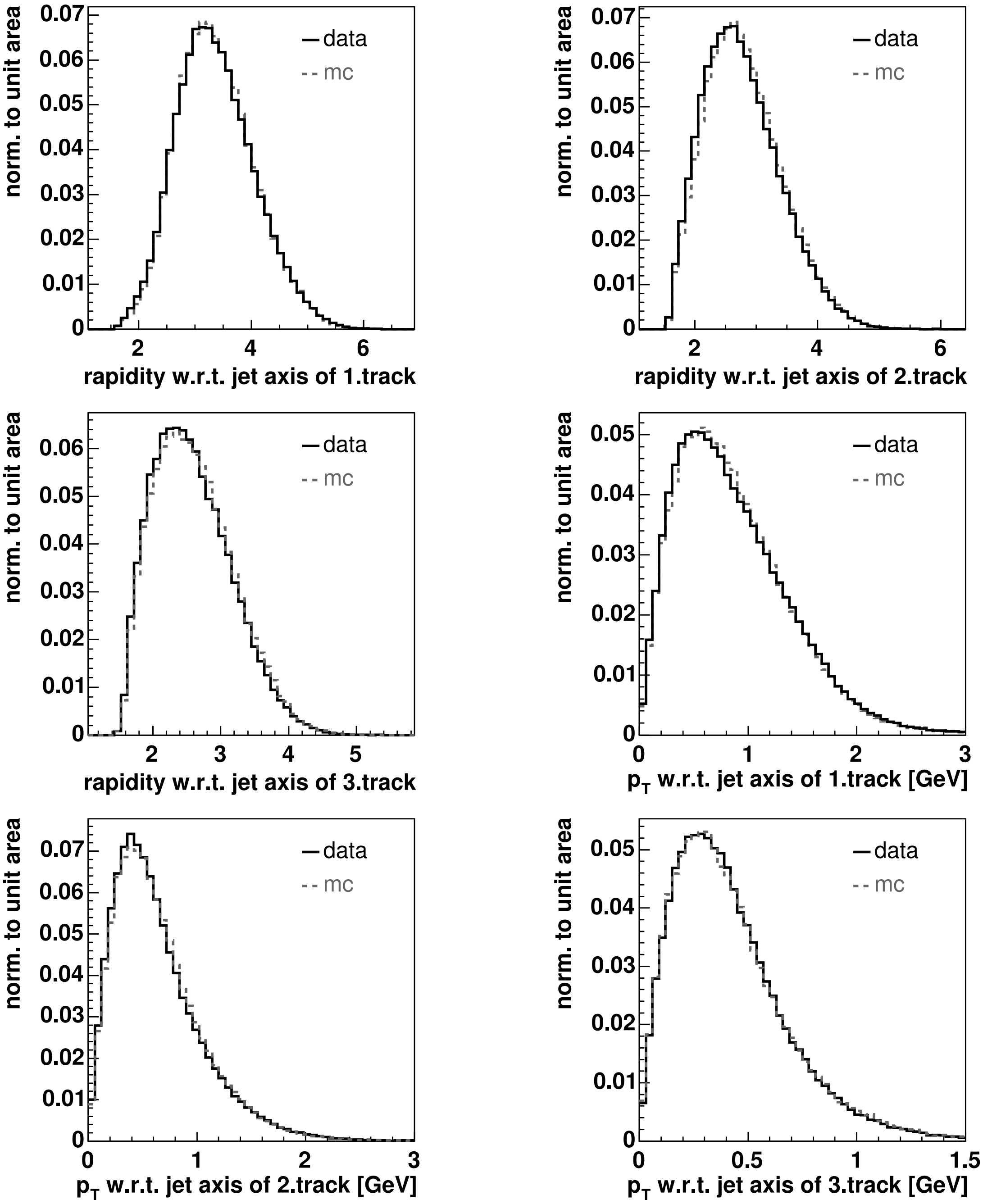


Figure 25: **Comparison of more kinematic discriminating track variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 5.

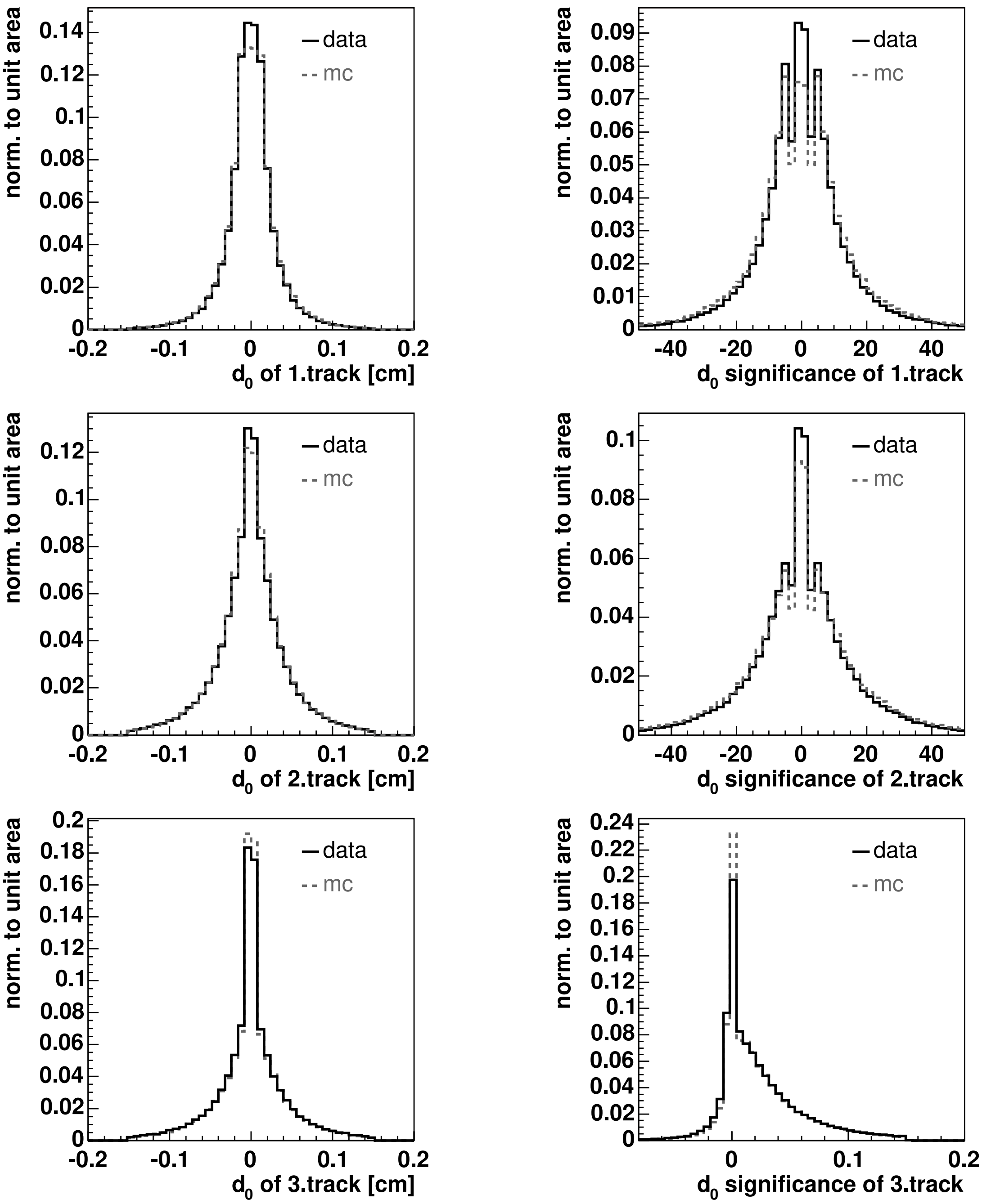


Figure 26: **Comparison of discriminating track variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 6.

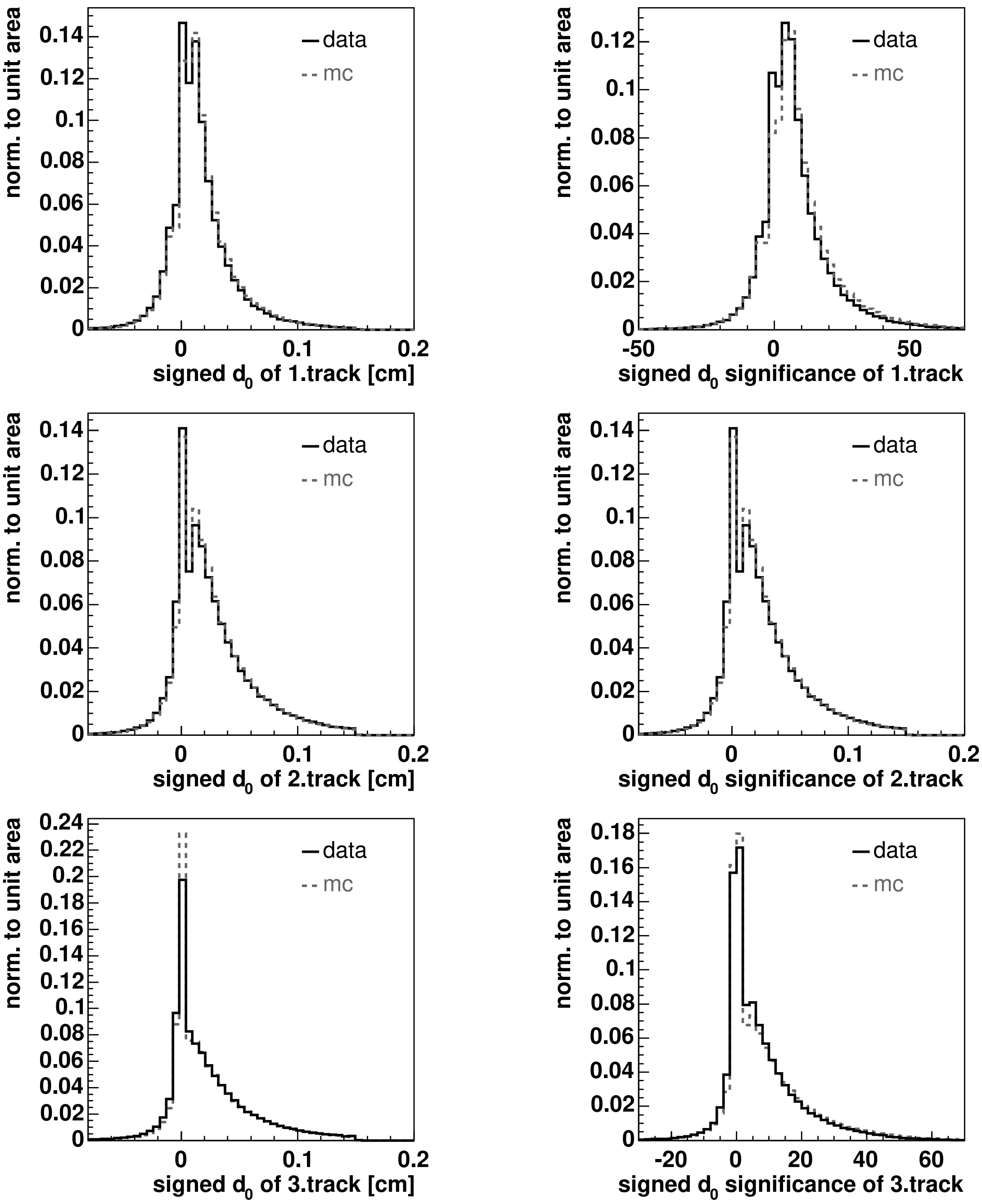


Figure 27: **Comparison of discriminating track variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 7.

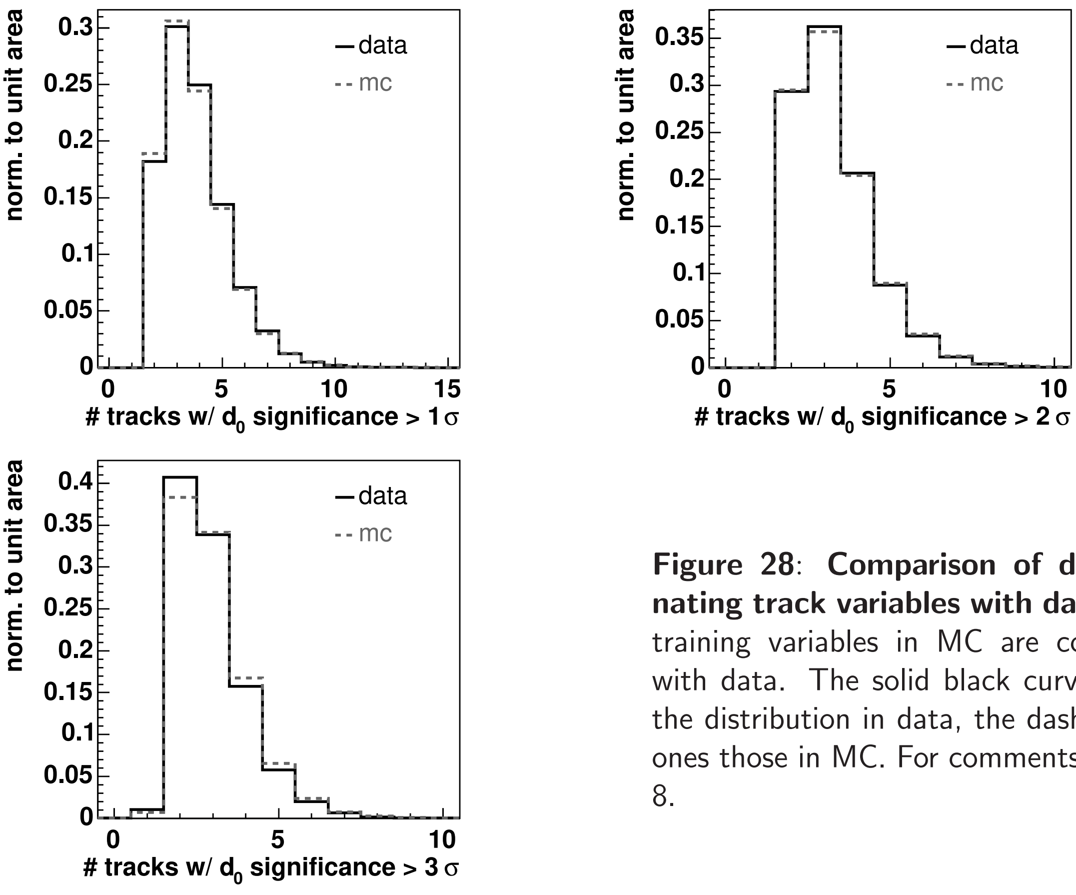


Figure 28: **Comparison of discriminating track variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 8.

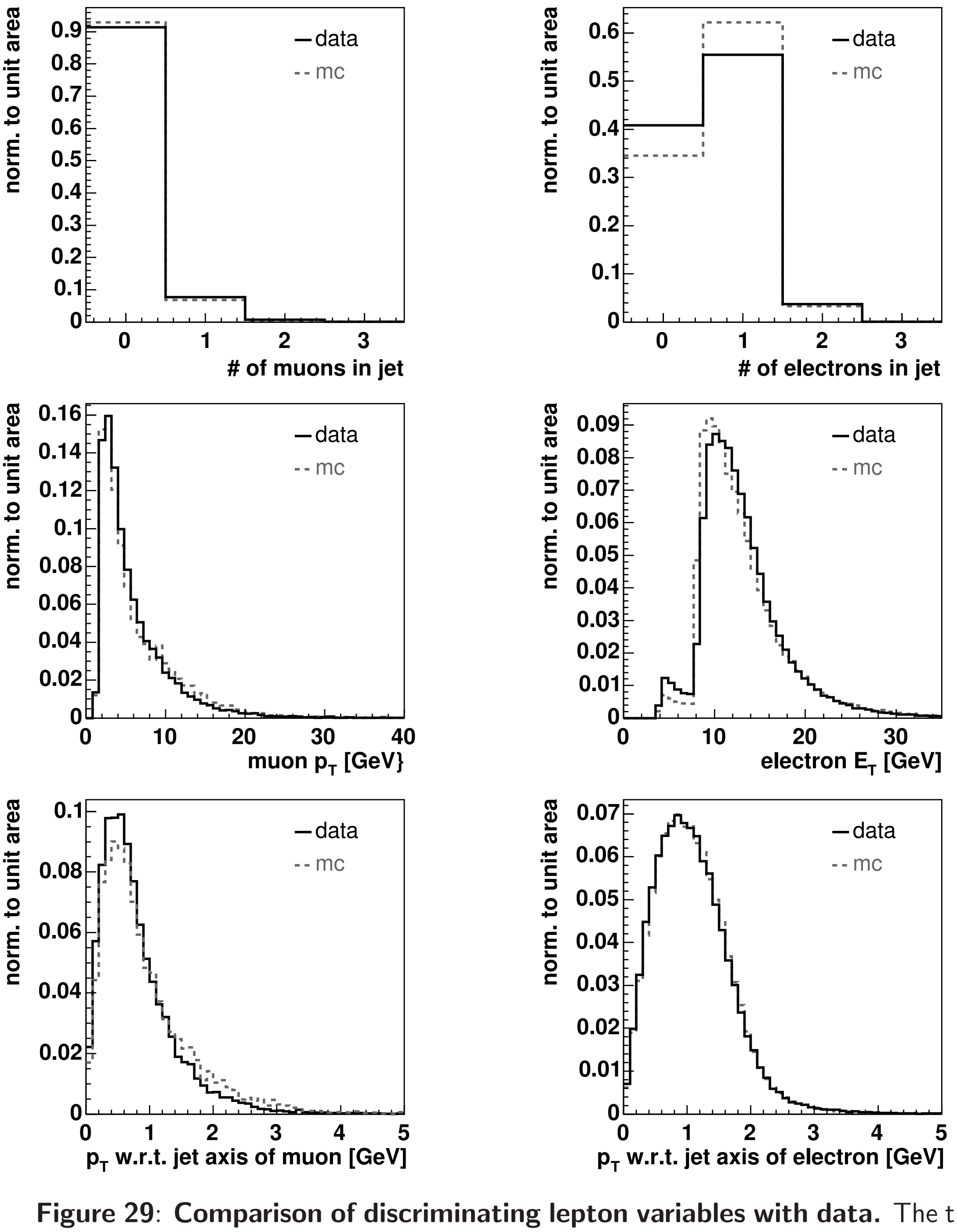


Figure 29: **Comparison of discriminating lepton variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 9.

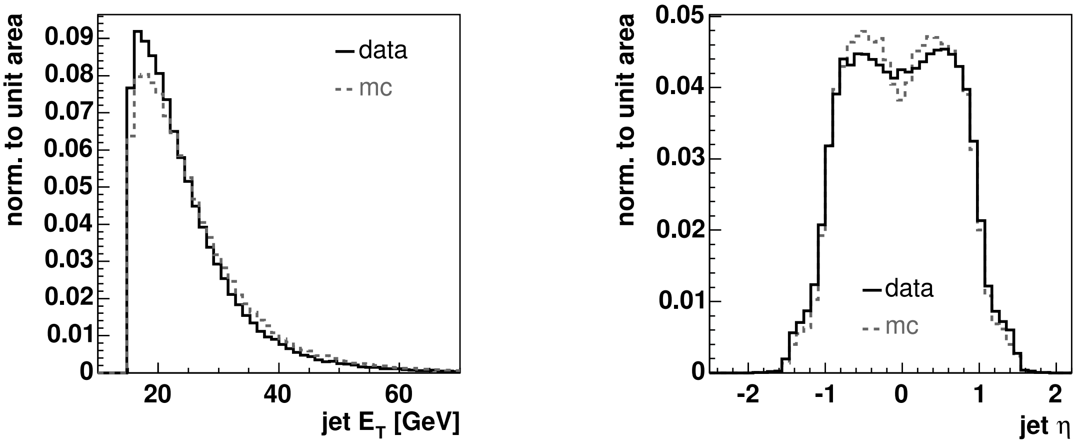


Figure 30: **Comparison of discriminating jet variables with data.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 10.

distributions of the transverse momentum at the secondary vertex (shown in figure 21 on the upper right hand side and in the middle two plots in figure 31) and the distributions of the momentum of the track with the highest momentum in the jet (shown in figure 24 on the upper left hand side). In both cases, the data distributions contain a bump at low momenta which cannot be seen in simulation.

We suggest that these bumps are due to the track of the high–p_T electron. In data, the electron track fails more often to be a good track than in simulation. If one requires the trigger electron (i.e. the electron fulfilling the cuts described in section 9) to be in the same jet as the considered track and in addition that the track fulfills the appropriate requirements to be a SecVtx track (i.e. $p_T > 0.5$ GeV, see section 3), the distributions of the momentum of the track with the highest momentum in the jet for data and MC fit perfectly which can be seen in figure 33 on the lower left hand side.

In addition, the track of the electron seems to be used less often for the secondary vertex fit in data than in simulation. This can be proven by requiring the electron track to be used in the fit and compare the transverse momentum at the secondary vertex for data and MC. With these requirements, those distributions fit quite well (see figure 33, upper plots).

The discussed differences are thus specific to our control sample which contains a high–p_T electron. In genuine b jet events these differences are not to be expected.

The final conclusion is that the simulated distributions describe the data well. Therefore, it is justified to train the neural network with simulated events. We further use the inclusive lepton data to quantify the residual difference between data and Monte Carlo. This analysis is described in the next sections.

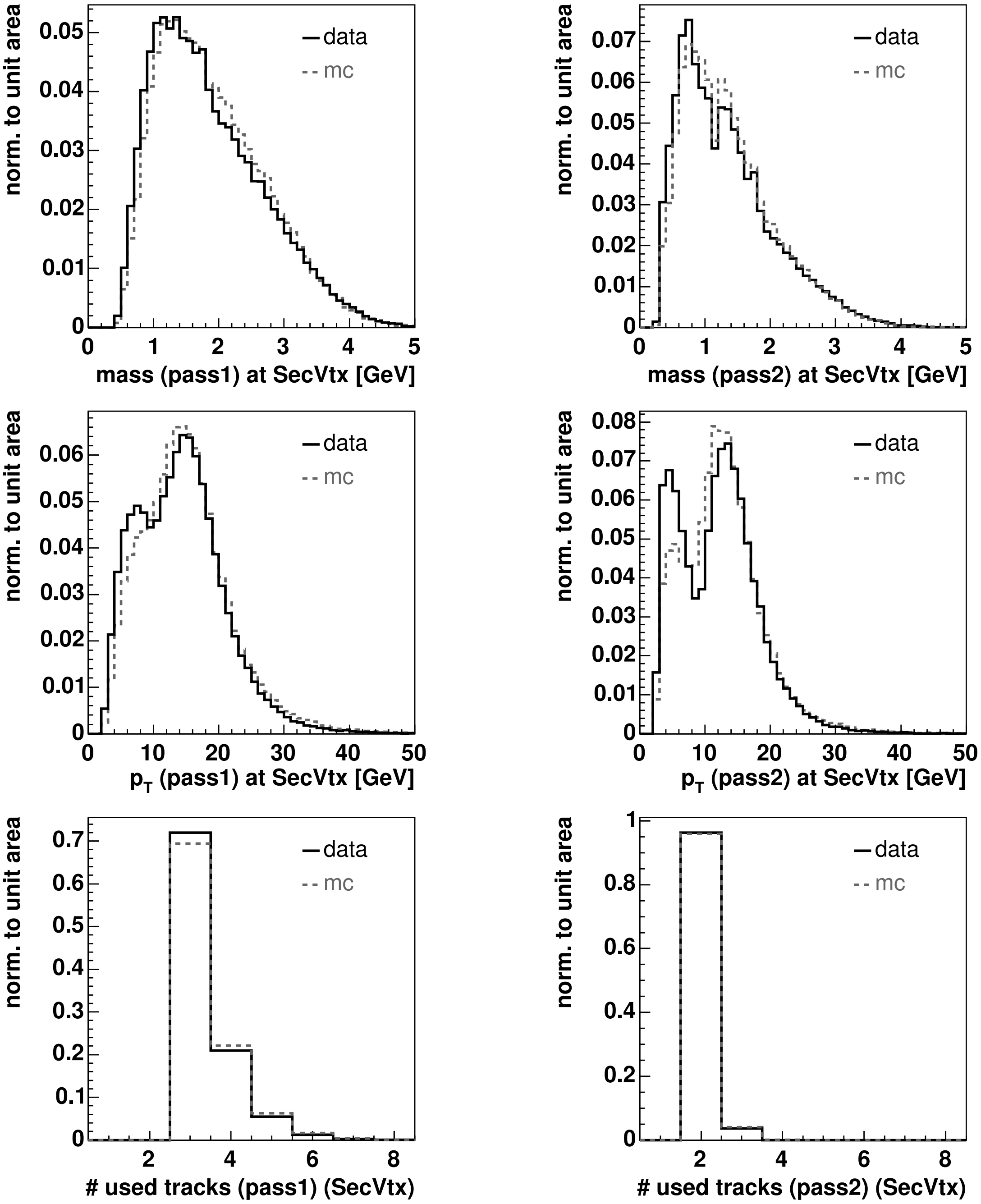


Figure 31: **Comparison with data for variables that are used separately for pass 1 and pass 2.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 13.

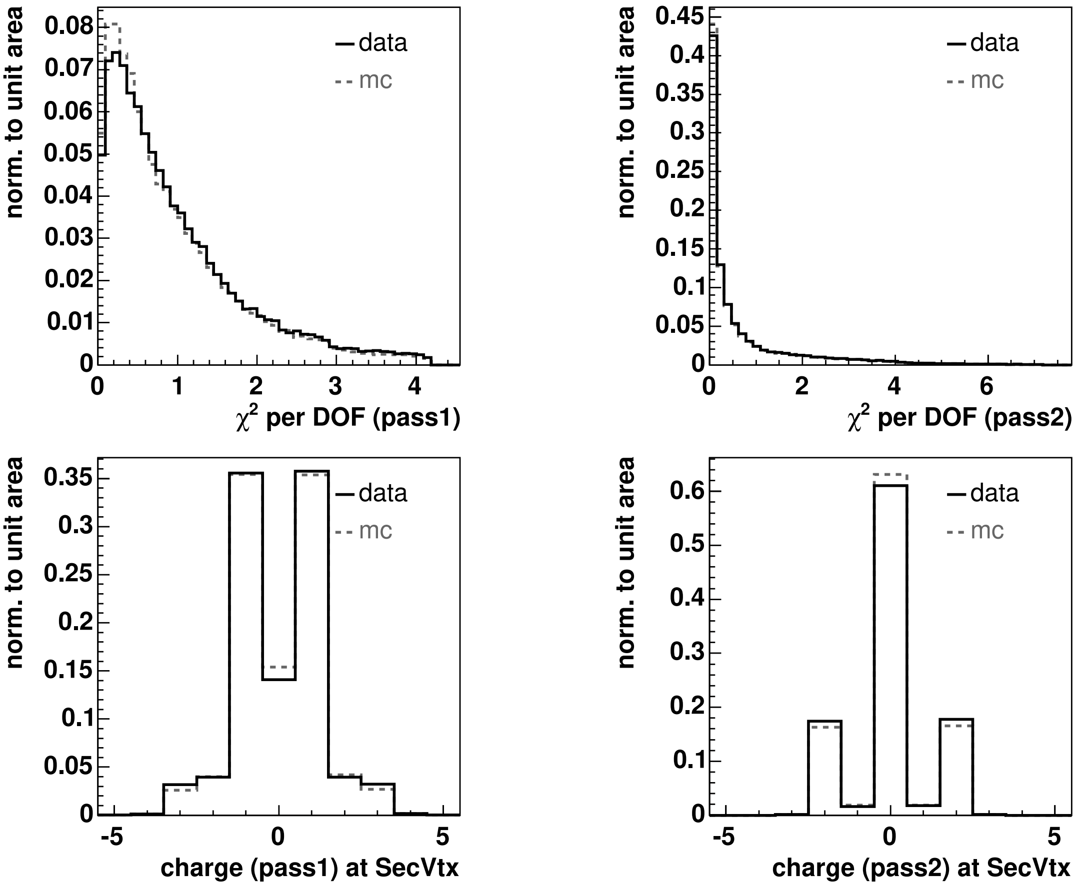


Figure 32: **Comparison with data for variables that are used separately for pass 1 and pass 2.** The training variables in MC are compared with data. The solid black curves show the distribution in data, the dashed grey ones those in MC. For comments see fig. 14.

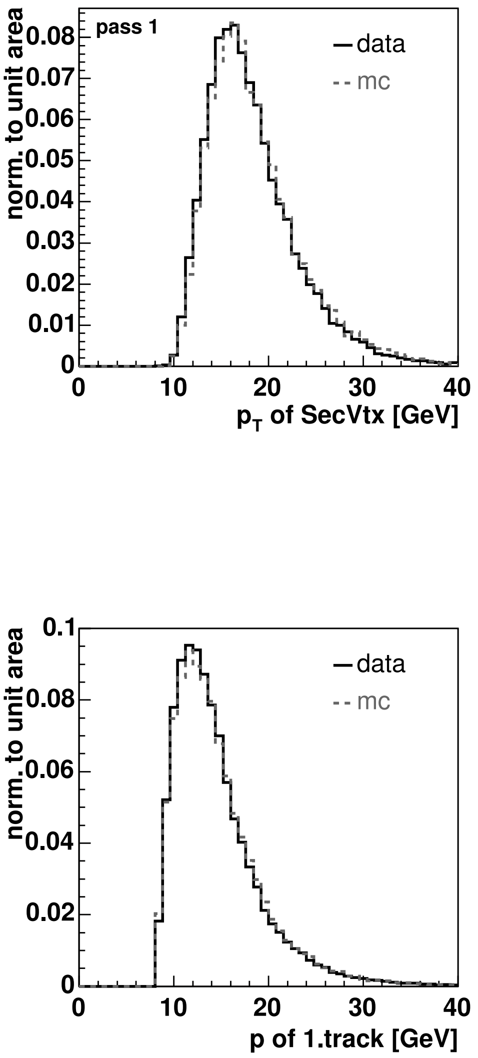


Figure 33: **Distributions of the transverse momentum of the secondary vertex and the momentum of the 1.track taking into account the differences concerning the electron.** The upper plots show the distributions of the transverse momentum of the secondary vertex with the requirement that the electron track was used in the SecVtx fit. The plot on the upper left hand side shows the distribution of pass 1 vertices, the one on the right hand side the distribution of pass 2 vertices .On the lower left hand side, the distribution of the momentum of the SecVtx track with the highest momentum in the jet is plotted with the requirement that the electron track is a SecVtx track. The solid black curves show the distribution in data, the dashed grey ones those in MC.

9 Preparation of the samples

We decided to determine the correction function similarly like the SecVtx scale factor [10, 11]. We require an energetic electron within a jet. This electron, often the signal of a semi-leptonic heavy quark decay, enhances the b and c content of the jets that are to be tagged. Hence the precision of the determination of the correction function is increased compared to not requiring this electron.

Since heavy flavor particles are mostly produced in pairs, we require an additional jet opposite in ϕ to increase the purity of the samples.

The requirements for the electron and the two jets are the following [10, 11]:

Electron cuts:

- $|z_{\text{electron track}}|$ at $r = 10.6$ cm < 43.5 cm
- $|z_{\text{electron track}}|$ at $r = 2.44$ cm < 43.5 cm
- the reconstructed electron must be within the fiducial region of the calorimeter, as it is given by the TopNtuple structure (fidele 1)
- $E_T > 9$ GeV (calorimeter)
- $P_T > 8$ GeV (track)
- $0.5 < \frac{E}{P} < 2.0$
- $\frac{had}{em} < 0.05$
- $L_{shr} < 0.2$
- $|\Delta_{x_{CES}}| < 3$ cm
- $|\Delta_{z_{CES}}| < 5$ cm
- $\chi^2_{CES_{strip}} < 10$
- $|z_e - z_0^{event}| < 5$ cm
- Non–isolated: $I = (E^{0.4}_{EM} + E^{0.4}_{Had})/E_{electron} > 0.1$
- if several electrons are present, pick the electron with the highest E_T

Selection of the electron jet:

- corrected $E_T > 15$ GeV (correction level 4)
- nearest jet in $\eta - \phi$–space to electron
- $\Delta R(electron - jet) < 0.4$

Selection of the away jet:

- corrected $E_T > 15$ GeV (correction level 4)
- $|\eta| < 1.5$
- $|\phi_{ejet} - \phi_{ajet}| > 2$ rad
- if several jets pass the cuts above, pick the one with the largest $\Delta\phi$ w.r.t the electron jet

In contrast to the determination of the SecVtx scale factor, we direct our attention to the away jet. The reason is the following: The neural network is very sensitive to the presence of leptons. This is not only the case due to the use of lepton specific variables (p_T of the electron w.r.t. the jet axis) but also due to the use of track based variables such as the momentum and the rapidity of the 1.track w.r.t. the jet axis. The determination of the correction function using the electron jet would introduce a very strong bias. The correction factor has to be valid for generic b jets not for a specific subsample.

As the neural network is based on the reconstruction of a secondary vertex, the away jet has to be tagged by SecVtx.

We further reject events in which the electron is identified as coming from a conversion.

10 Composition of the away jet sample

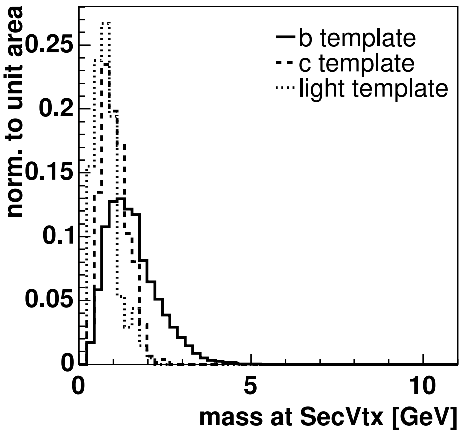
To be able to compare the tagging efficiencies, one has to be sure that the composition of the tagged away jets is comparable in data and simulation.

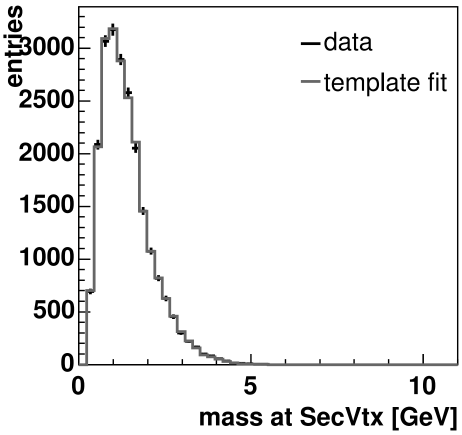
The composition in MC is determined using the Monte Carlo truth. We looked for heavy flavor quarks with a minimum p_T of 5 GeV in a cone of 0.4 around the jet axis. If we found one, the jet is assigned as b or c jet, depending on the flavor. If we found none, the jet is assigned as light.

The composition in data is estimated by fitting the mass at the secondary vertex as described in Ref. [14]. The templates which are to be fitted to the data distribution were obtained by the inclusive electron MC samples btop5a, btop6a, btop7a and btop8a. The templates were generated by requiring both electron and away jet as described above. Only the away jet had to be tagged by SecVtx, no matter if the electron jet was tagged or not. Those templates can be seen in figure 34 on the upper left hand side.

We have two possibilities for the determination of the correction function using the tagged away jet: the electron jet can be required to be tagged or not. The first possibility has the advantage that the heavy flavor content is increased, the second one has higher statistics. The results of fitting the templates to the distribution of the mass at the secondary vertex can be seen in figure 34. In both cases the fitted distributions describe the real one satisfactorily.

If we require only the away jet to be tagged by SecVtx, we obtain the compositions shown in table 8. The b fraction in data (72.9%) seems to be much less than in the simulation (88.9%). If we require both jets to be tagged by SecVtx, we do not only increase the b fraction as shown in table 9 but we also diminish the deviations between data (87.2%) and simulation (95.0%).





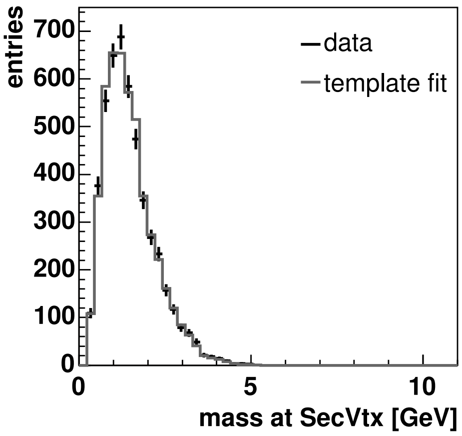


Figure 34: **Estimation of the composition of the away jet in data.** The templates shown on the upper left hand side are obtained from the btop5a, btop6a, btop7a and btop8a samples. The requirements are the same as described in section 9. Only the away jet was required to be tagged. On the upper right hand side, the result of the fit of those templates to the data distribution is shown for the case that only the away jet is required to be tagged by SecVtx, on the lower left hand side for the case that both electron and away jet are tagged by SecVtx.

sample	b fraction	c fraction	light fraction
MC	88.9%	8.8%	2.3%
data	72.9% ± 1.9%	17.2% ± 2.8%	9.9 ± 1.9%

Table 8: **The fractions of b, c and light away jets in Monte Carlo and data, only away jet tagged.** The composition in the simulation was obtained by using the Monte Carlo truth. The fractions in data were estimated by fitting templates of distributions of the mass at the secondary vertex to the data distribution. For these numbers, only the away jet was required to be tagged by SecVtx, no matter if the electron jet is tagged or not. Events where the electron is from a conversion are excluded.

sample	b fraction	c fraction	light fraction
MC	95.0%	3.2%	1.8%
data	87.2% ± 3.0%	9.9% ± 3.8%	2.9 ± 2.0%

Table 9: **The fractions of b, c and light away jets in Monte Carlo and data, both jets tagged.** The composition in the simulation was obtained by using the Monte Carlo truth. The fractions in data were estimated by fitting templates of distributions of the mass at the secondary vertex to the data distribution. For these numbers, both the electron and the away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.

To be sure that the mass fit is not influenced by the similarity of the light and the c template, we performed another fit taking light and c jets together. The result we obtained for the b fraction was essentially the same as before. That is why we assume that our result (at least for the b fraction) is as trustworthy.

The conclusion we draw from these numbers is that it is the best to require both the electron and the away jet to be tagged by SecVtx to make sure that we obtain jet samples with mostly comparable composition. However, our analysis indicates that even if we require both jets to be tagged a residual difference in the sample composition between data and Monte Carlo remains.

Another caveat is that our comparison between data and Monte Carlo is based on $b\bar{b}$ events that have a back–to–back kinematic. These are events predominantly due to $b\bar{b}$ production by flavor creation ($gg \to b\bar{b}$, $q\bar{q} \to b\bar{b}$). However, the conclusions are largely applied to $Wb\bar{b}$ events which are mainly due to gluon splitting ($W + g$, $g \to b\bar{b}$). The assumption here is that the selected sample represents generic b jets.

11 Determination of the correction function

We compare the output distributions for data and Monte Carlo for two cases: (1) only away jet tagged and (2) both electron and away jet tagged. Those plots are shown in figure 35. One can see that the distributions fit quite well. For the estimation of the systematic uncertainties contained in those distributions we use the integrated output

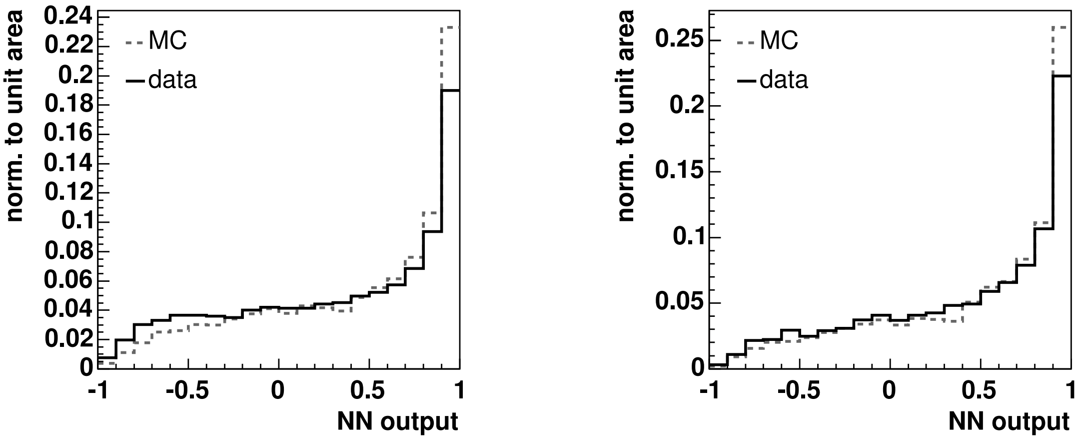


Figure 35: **Comparison of the network output of the away jet for data and Monte Carlo.** On the left hand side, the distributions for data and Monte Carlo are shown for the requirement that the away jet is tagged, no matter if the electron jet is tagged or not. On the right hand side, the distributions are shown for the requirement that both electron and away jet are tagged. Events where the electron is from a conversion are excluded. The data distributions are plotted in solid black, the MC ones in dashed grey.

distributions as described in the next section.

11.1 Accumulating the distributions

For the determination of a correction function, we accumulate the normalized output distributions (see figure 35). This accumulation is done by integrating the output distribution from right to left. For example, the bin at output -1 in the accumulated output distribution contains the complete integral under the output distribution. The bin with output 0 contains the integral of the original distribution starting at output 0. These accumulated distributions are shown on the left hand sides in figures 36 and 37.

On the right hand side of figures 36 and 37, the ratio of the accumulated data and MC distributions is shown.

The accumulated distributions for the sample where we required only the away jet to be tagged show larger deviations between data and Monte Carlo. The ratio between data and MC is smaller than 1. This can be understood taking the composition estimates (see table 8) into account.

As we showed in section 10, the composition of the away jet in data corresponds more to the simulated one if we require both electron and away jet to be tagged. That's why we use this two–tag sample for the estimation of the systematic uncertainty.

For this sample, we redo the comparison separately for pass 1 and pass 2 vertices. The according ratios of the accumulated output distributions are shown in figure 38. One can see that the deviations for pass 2 vertices are negligible.

For a cut–based use of the neural network tagger, no correction in addition to the SecVtx scale factor is needed for pass 2 secondary vertices. For pass 1 vertices, the

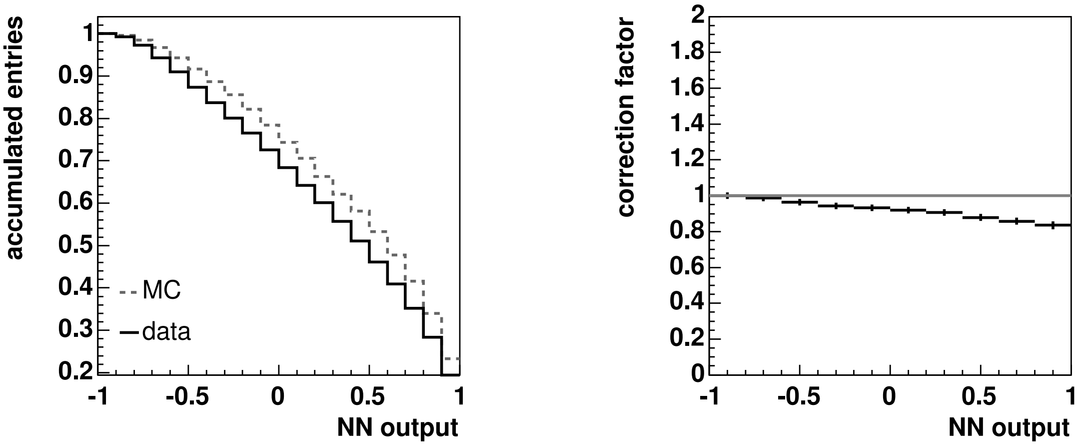


Figure 36: **Comparison of the accumulated network output for data and Monte Carlo if only the away jet is tagged.** The accumulated normalized output distributions of the away jet in data (solid black) and MC (dashed grey) are shown on the left hand side. On the right hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, only the away jet is required to be tagged by SecVtx, no matter if the electron jet is tagged or not. Events where the electron is from a conversion are excluded.

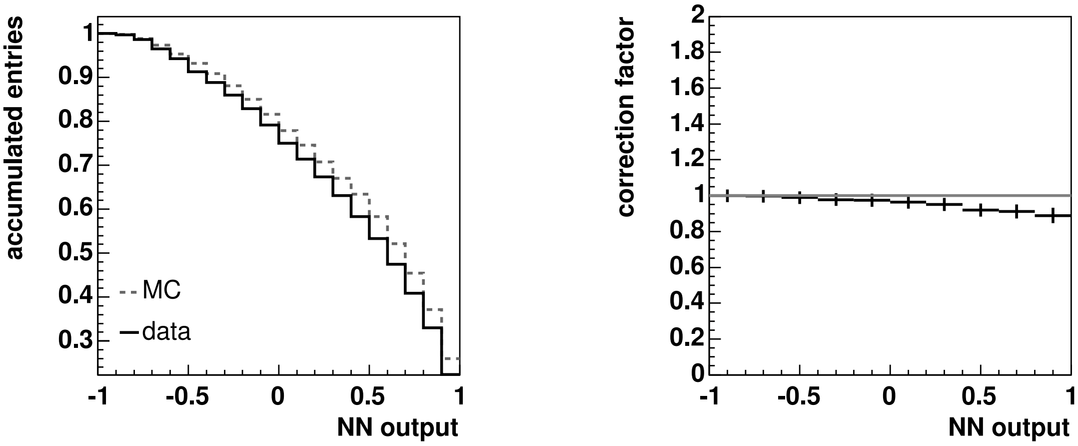
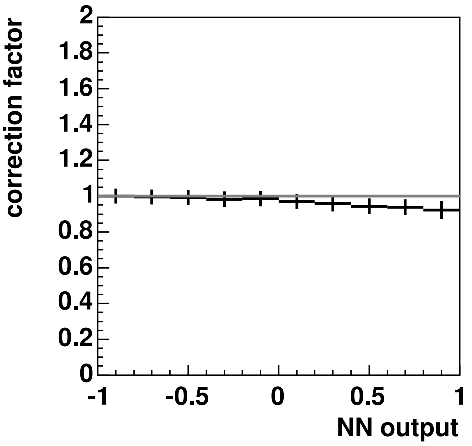


Figure 37: **Comparison of the accumulated network output for data and Monte Carlo.** The accumulated normalized output distributions of the away jet in data (solid black) and MC (dashed grey) are shown on the left hand side. On the right hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, both electron and away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.



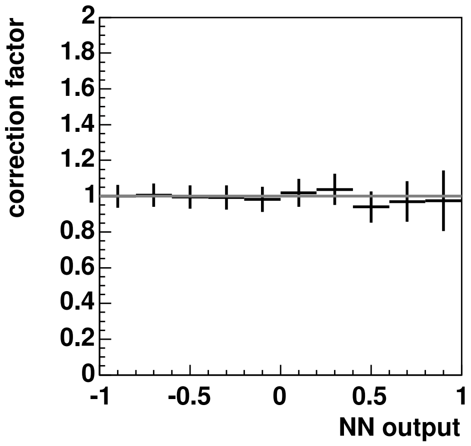


Figure 38: **Correction factor for pass 1 and pass 2 vertices.** On the left hand side, the ratio of the accumulated data distribution divided by the accumulated MC distribution is shown for pass 1 vertices, on the right hand side for pass 2 vertices.

Cut	>-0.9	>-0.7	>-0.5	>-0.3	>-0.1
SF	1.000 ± 0.042	0.995 ± 0.042	0.992 ± 0.042	0.983 ± 0.042	0.987 ± 0.042
Cut	>0.1	>0.3	>0.5	>0.7	>0.9
SF	0.969 ± 0.042	0.958 ± 0.042	0.943 ± 0.043	0.938 ± 0.045	0.922 ± 0.048

Table 10: **Correction factor for pass 1 vertices for the cut–based use of the neural network.** These factors are read out from the plot on the left hand side in figure 38. The table has to be read from left to right. The last bin which fully includes the interval of the chosen cut determines the correction factor SF.

correction factor can be read out of the plot shown in figure 38 on the left hand side.

In table 10, the correction factor for the pass 1 vertices is shown for the 10 bins displayed in figure 38 on the left hand side. For a given cut, the bin has to be chosen which fully includes the interval specified by the cut. I.e. for the cut >0.0 (which was found to give the best ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top, see section 6) the correction factor would be $SF = 0.987 \pm 0.042$. We recommend not to use this factor as a real correction but as a measure of the systematic uncertainty. The reason is that it is not possible to determine the effects on the observed differences induced by the disagreement between the compositions as indicated in table 9 and the effects induced by real discrepancies between data and simulation.

For the use of the tagger as an input to another neural network for the single–top analysis we recommend to estimate the systematic uncertainty by shifting the simulated accumulated distribution shown in figure 37 to the data one and investigating the influence of the shifted output to the result of the training of the single–top neural network. For this purpose, we treat pass 1 and pass 2 vertices the same way.

In figure 39 (left), the corrected output of the away jet in the simulation is plotted versus the original output. This corrected output will be used to estimate the systematic

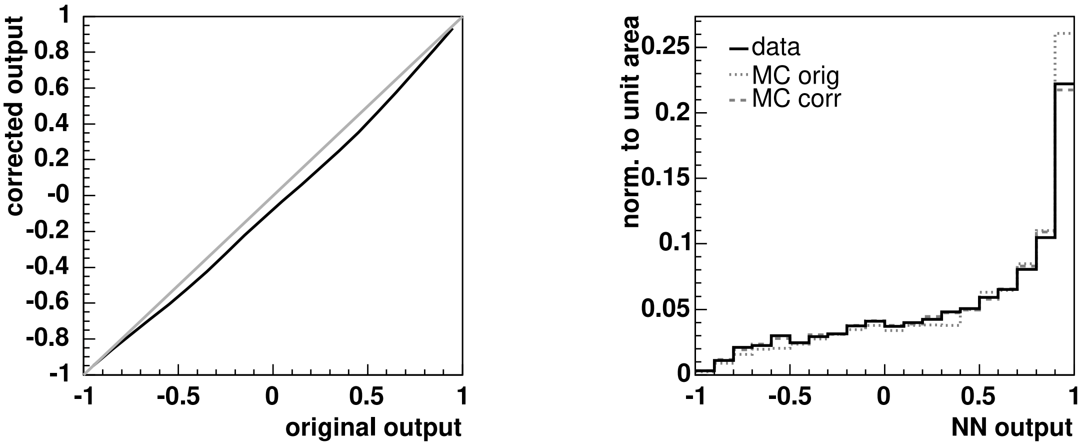


Figure 39: **Comparison of original and corrected MC output.** The plot on the left hand side shows the corrected MC output versus the original output of the away jet. The plot on the right hand side shows the comparison of the corrected and the original MC distributions with data. The dotted histogram shows the original MC output distribution. The corrected output distribution is shown in dashed grey. The solid black histogram represents the data distribution. Both passes are taken together.

uncertainties for the use of the tagger as a continuous variable. The comparison of the original and the corrected output of the away jet in the simulation with the data distribution can be seen in figure 39 on the right hand side. It can be seen that the corrected MC distribution is very alike the data.

11.2 E_T dependence of the correction function

We also investigated the E_T dependence of the correction function. In figure 40, the ratios of the accumulated output distributions are shown for two different jet E_T bins. On the upper left hand side, the ratios of the accumulated output distributions of pass 1 vertices are shown, on the right hand side for pass 2. Within the error bars, we cannot see any dependence of the E_T of the jet.

11.3 Electron Jet

As we mentioned before, the electron jet is a more specific kind of jet and can therefore not be used for the estimation of systematic uncertainties. For the sake of completeness, we had a look at it as well.

In tables 11 and 12 one can see the composition of the electron jet in MC and the estimated one in data. Again, the estimation of the data composition was performed by fitting templates of the secondary vertex mass to the mass distribution in data. Those templates were obtained by the inclusive electron samples btop5a, btop6a, btop7a and btop8a. For the creation of the vertex mass templates for the electron jet, only the electron jet has to be tagged by SecVtx, no matter if the away jet is tagged or not. The templates as well as the fit results are shown in figure 41.

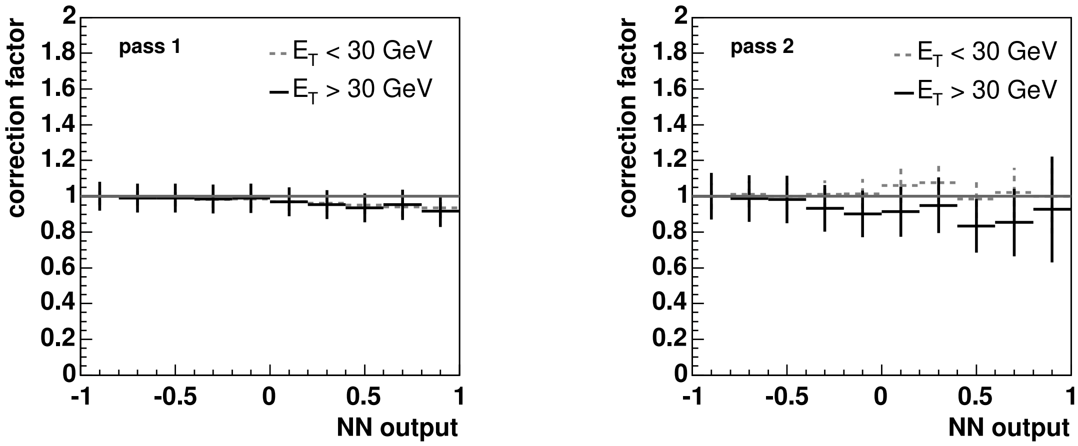


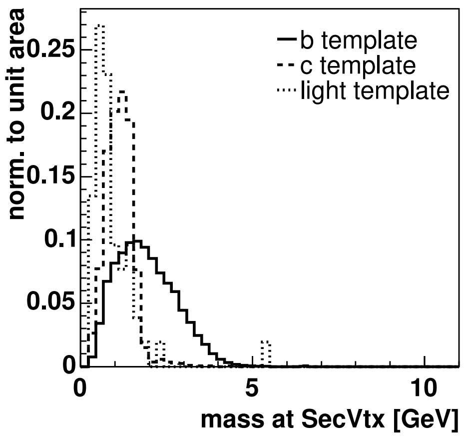
Figure 40: $\mathrm{E_T}$ **dependence of the correction function.** On the left hand side, pass 1 vertices are shown, on the right hand side for pass 2. Jets with $\mathrm{E_T}<$ 30 GeV are displayed in dashed grey, jets with $\mathrm{E_T} >$ 30 GeV in solid black.

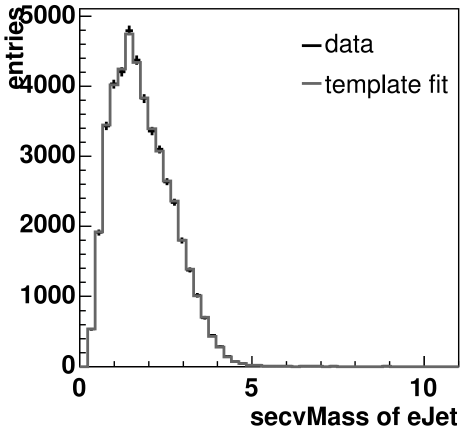
sample	b fraction	c fraction	light fraction
MC	92.7%	7.2%	0.1%
data	90.7% ± 1.1%	6.2% ± 1.2%	3.1 ± 1.1%

Table 11: **The fractions of b, c and light electron jets in Monte Carlo and data, only electron jet tagged.** The composition in the simulation was obtained by using the Monte Carlo truth. The fractions in data were estimated by fitting templates of distributions of the mass at the secondary vertex to the data distribution. For these numbers, only the electron jet was required to be tagged by SecVtx, no matter if the away jet is tagged or not. Events where the electron is from a conversion are excluded.

sample	b fraction	c fraction	light fraction
MC	98.0%	1.9%	0.04%
data	97.6% ± 2.1%	0.9% ± 1.7%	1.5 ± 1.0%

Table 12: **The fractions of b, c and light electron jets in Monte Carlo and data, both jets tagged.** The composition in the simulation was obtained by using the Monte Carlo truth. The fractions in data were estimated by fitting templates of distributions of the mass at the secondary vertex to the data distribution. For these numbers, both the electron and the away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.





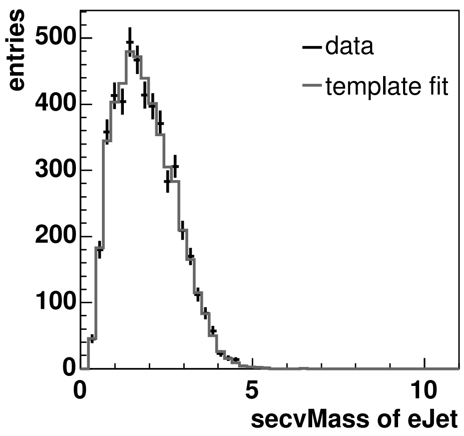


Figure 41: **Estimation of the composition of the electron jet in data.** The templates shown on the upper left hand side are obtained from the btop5a, btop6a, btop7a and btop8a samples. The requirements are the same as described in section 9. Only the electron jet was required to be tagged. On the upper right hand side, the result of the fit of those templates to the data distribution is shown for the case that only the electron jet is required to be tagged by SecVtx, on the lower left hand side for the case that both electron and away jet are tagged by SecVtx.

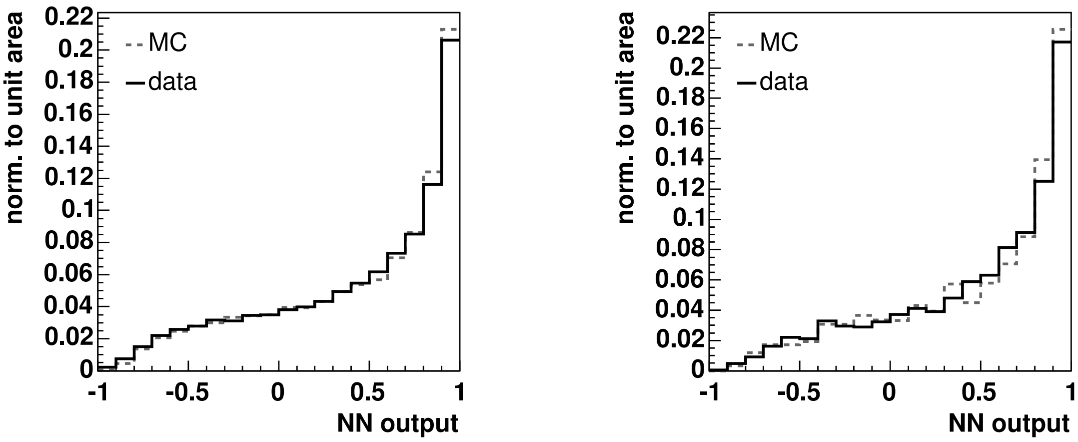


Figure 42: **Comparison of the network output of the electron jet for data and Monte Carlo.** On the left hand side, the distributions for data and Monte Carlo are shown for the requirement that the electron jet is tagged, no matter if the away jet is tagged or not. On the right hand side, the distributions are shown for the requirement that both electron and away jet are tagged. Events where the electron is from a conversion are excluded. The data distributions are plotted in solid black, the MC ones in dashed grey.

One can see that the composition of the electron jet in data and MC seems to be almost completely comparable, no matter if the away jet is required to be tagged or not.

The output distributions for the electron jet for both cases are shown in figure 42. Those distributions were again accumulated as described in section 11.1 for the away jet. Those distributions as well as the ratios of data divided by MC are shown in figure 43 for the sample where only the electron jet is required to be tagged and in figure 44 for the two–tag sample.

We can see that the agreement between the output distributions for data and simulation is almost perfect for the electron jet. This confirms our suggestion that most of the differences seen for the away jet (see figures 36 and 37) are due to the different compositions in data and MC.

12 Mistags

The estimation of the systematic uncertainties described before is done using signal–like samples. As a cross check, we want to compare the output distributions for more background–like samples. As we require the jets to be tagged by SecVtx, this is a difficult exercise. For getting background–like samples, we are looking at jets with a negative secondary vertex tag. Those vertices are exclusively reconstructed out of misreconstructed tracks. The assumption we make is that negatively tagged vertices are a valid description for positively mistagged vertices.

A caveat is that those events have per default a negative decay length and therefore different kinematics than positively tagged jets. For the decay length, we take the absolute

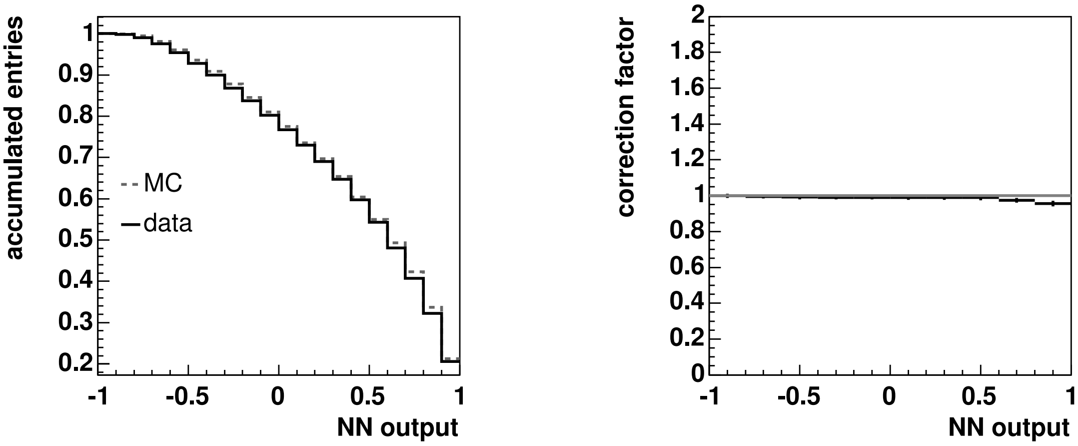


Figure 43: **Comparison of the accumulated network output of the electron jet for data and Monte Carlo for only electron jet tagged.** The accumulated normalized output distributions of the electron jet in data (solid black) and MC (dashed grey) are shown on the left hand side. On the right hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, only the electron jet is required to be tagged by SecVtx, no matter if the away jet is tagged or not. Events where the electron is from a conversion are excluded.

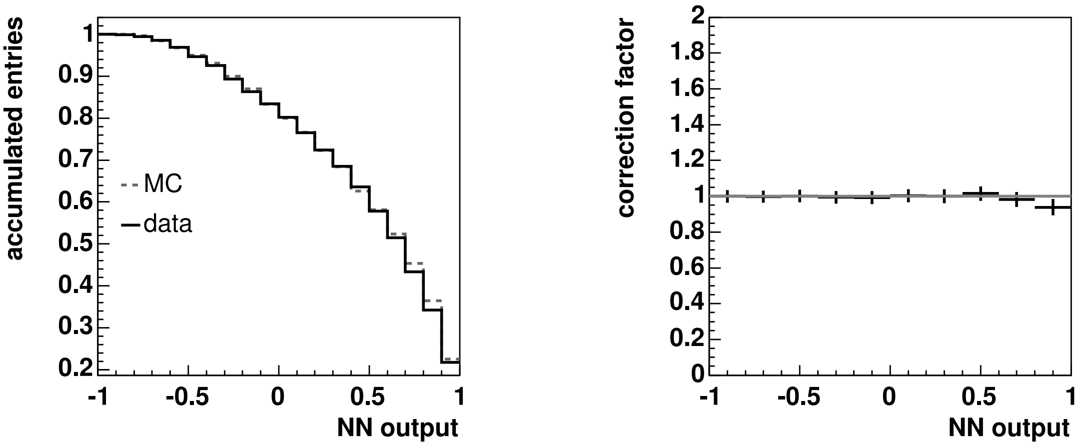


Figure 44: **Comparison of the accumulated network output of the electron jet for data and Monte Carlo.** The accumulated normalized output distributions of the electron jet in data (solid black) and MC (dashed grey) are shown on the left hand side. On the right hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, both electron and away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.

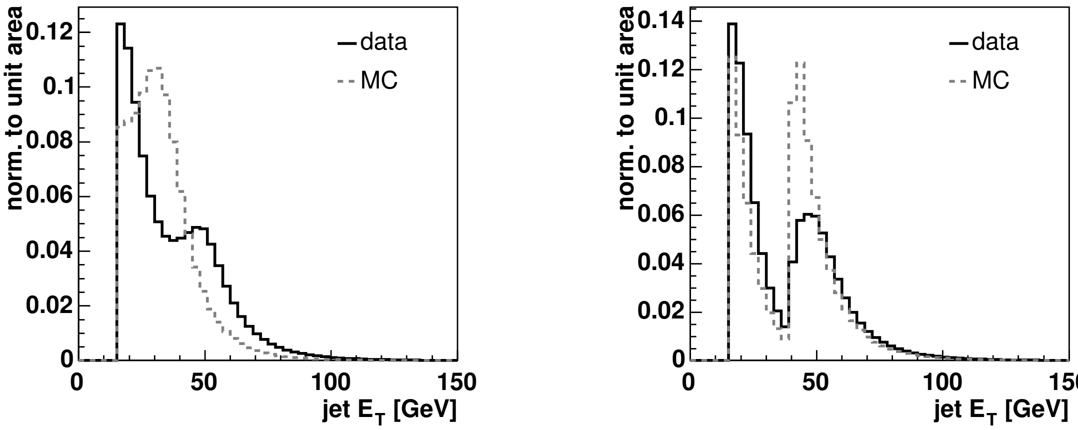


Figure 45: **Jet $\mathrm{E_T}$ spectra for jet samples in data and simulation.** The plot on the left hand side shows the $\mathrm{E_T}$ spectra in data and MC. The plot on the right hand side shows the spectra with $\mathrm{E_T} > 40$ GeV. The data distributions are plotted in solid black, the MC ones in dashed grey.

value. But the tracks are still pointing to the negative hemisphere. This causes the correlations between the variables to be very different from those used to train the neural network.

12.1 Samples

We are using generic dijet samples for this comparison. For data, we use $\mathrm{E_T} > 20$ GeV (gjt10d) and $\mathrm{E_T} > 50$ GeV (gjt20d) jet triggers. For the simulated events, we are using Pythia dijet samples with $\mathrm{p_T} > 18$ GeV/c (jqcd1f) and $\mathrm{p_T} > 40$ GeV/c (jqcd2f) as well as Herwig dijet samples with $\mathrm{p_T} > 18$ GeV/c (jqcd2h) and $\mathrm{p_T} > 40$ GeV/c (jqcd3h). One caveat using those samples is that the jet trigger is not simulated in the MC sample as can be seen in figure 45 on the left hand side. If one cuts at jet $\mathrm{E_T} > 40$ GeV, this leads to comparable spectra for data and MC as it is shown in figure 45 on the right hand side.

In addition, we require at least 2 jets with $\mathrm{E_T} > 15$ GeV and investigate the two most energetic ones.

12.2 Comparison of output distributions

First, we have a look only at the simulated events. We compare the negatively tagged light jets to the positively tagged light jets to all negatively tagged jets including heavy flavor. This comparison is shown in figure 46 on the left hand side. One can see that the curves are very similar. This leads to the conclusion that mistagged light jets are described consistently in simulation.

The comparison of the negatively tagged MC events to the negatively tagged data events is shown in figure 46 on the right hand side. One can see some differences between the distributions. This indicates that mistags are not well modeled in the simulation,

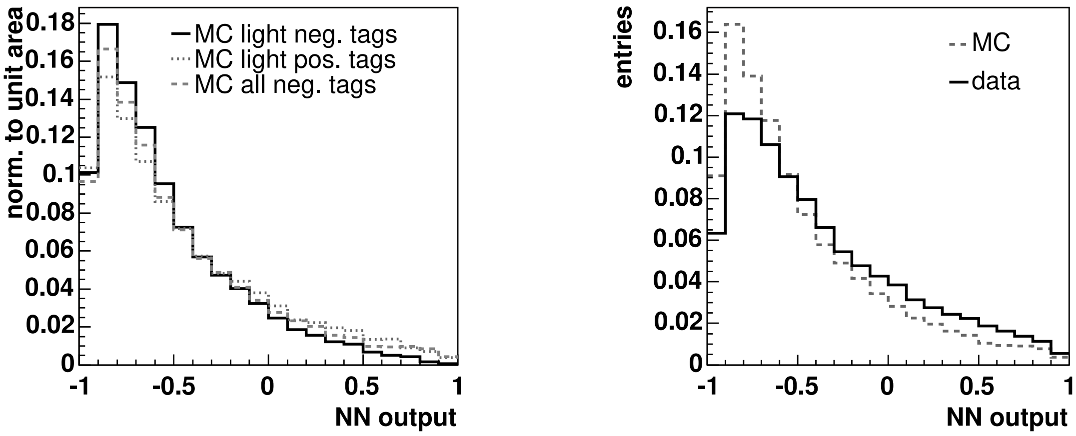


Figure 46: **Comparison of negatively tagged events.** On the left hand side, the negatively tagged light MC events (solid black), the positively tagged light MC events (dotted grey) and all negatively tagged MC events (dashed grey) are shown. On the right hand side, the comparison of the negatively tagged events in data and MC is shown. Both passes are taken together.

even though the agreement is still reasonable taking into account the caveats mentioned before.

Therefore, we will use a correction function to take those differences into account. This correction will be applied to the $W1p$, $W2p$, $W3p$ and $W4p$ MC samples. We accumulate the output distributions (see figure 46, right) for data and Monte Carlo as described in section 11.1. The accumulated distributions and the ratio of those distributions can be seen in figure 47.

It is obvious that the deviations between data and simulation are larger than for signal–like events. To obtain a correction taking those discrepancies into account, we examine the dependence of the correction function on the same parameters as for the SecVtx mistag matrix [15]. This is described in the next section.

In figure 48, the output distributions are shown separately for pass 1 and pass 2 secondary vertices. One can see that the distributions are quite different. As expected, the output distributions for pass 2 are more background–like than the ones for pass 1 vertices. Therefore, we will treat those two kinds of secondary vertices independently.

12.3 Mistag Matrix

In this section, we evaluate the dependence of the deviation between data and simulation on jet $\mathrm{E_T}$, $\sum \mathrm{E_T}$ of all taggable jets, # of good tracks in the jet, $|\eta|$ and ϕ. Here, taggable jets means that uncorrected $\mathrm{E_T} > 10$ GeV, $|\eta| <2.4$ and # good tracks ≥ 2 [15].

For the first approach, we examine the dependences completely inclusive by neglecting possible correlations between them.

In figure 49, the ratios of the accumulated output distributions are shown for pass 1 vertices in dependence on the parameters mentioned above. Within the given statistics,

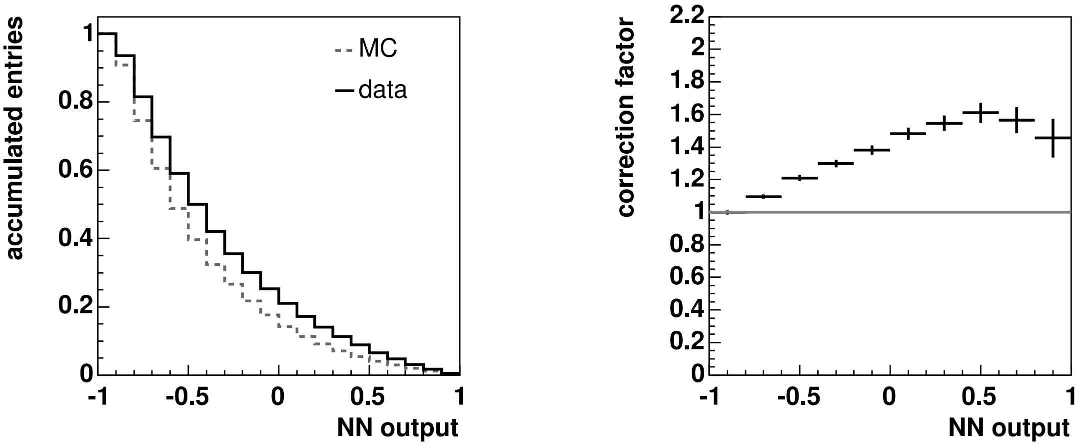


Figure 47: **Comparison of accumulated output distributions for data and Monte Carlo for negatively tagged jets.** The accumulated normalized output distributions of the two most energetic jets in data (solid black) and MC (dashed grey) are shown for negative tags on the left hand side. On the right hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, the 2 most energetic jets are required to be negatively tagged by SecVtx. Both passes are taken together.

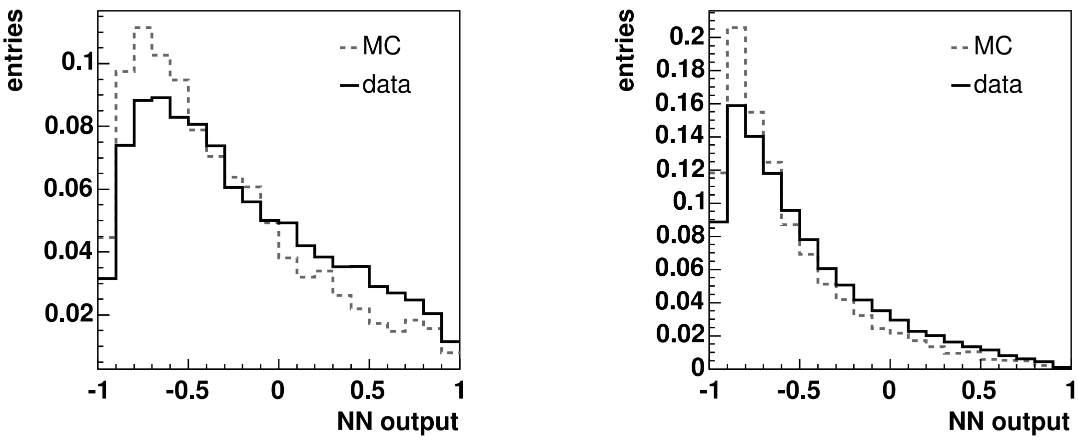


Figure 48: **Comparison of negatively tagged events separately for pass 1 and pass 2.** On the left hand side, the comparison of the output distributions for negatively tagged pass 1 vertices in data (solid black) and MC (dashed grey) are shown. On the right hand side, the negatively tagged pass 2 vertices are compared.

we see only dependences on jet E_T for pass 1 vertices. Therefore, we will parameterize the correction function for pass 1 vertices only in dependence on jet E_T.

In figure 50, the ratios of accumulated output distributions for pass 2 vertices are shown in dependence on the same parameters. Here, one can see that the deviations between data and simulation do not only depend on jet E_T, but also on $\sum E_T$ and the number of good tracks in the jet. Again, angular dependences can be neglected. Therefore, we will parameterize the correction function for pass 2 vertices in differential bins of jet E_T, $\sum E_T$ and number of good tracks in the jet.

For the sake of statistics, we use the original E_T spectra (see figure45, left) for the differential description of the correction functions.

12.3.1 Correction functions

As motivated in the previous section, the correction function of pass 1 vertices is parameterized in dependence on jet E_T. The parameterization of the correction function of pass 2 vertices is done in bins of E_T, $\sum E_T$ and the number of good tracks in the jet.

T he correction functions are displayed in figure 51. It can be seen that there are mainly two different kinds of correction functions for pass 2 vertices: For some bins, i.e. $\sum E_T \geq 60$ GeV, $E_T \geq 30$ GeV, Ntrk ≥ 6 as well as $\sum E_T \geq 60$ GeV, $E_T < 30$ GeV, Ntrk ≥ 6 and $\sum E_T < 60$ GeV, $E_T \geq 30$ GeV, Ntrk ≥ 6, a relatively large correction is needed. For the other bins, the correction is mainly negligible. For pass 1 vertices, the correction needed for jets with $E_T \geq 30$ GeV is relatively large while the correction function for low energetic jets is almost compatible with the diagonal.

To demonstrate that the correction works, the data output distributions as well as the simulated original and simulated corrected output distributions are shown in figure 52 for pass 1 and pass 2 vertices. One can see, that the corrected simulated output distributions are very alike the data.

13 Influence of kinematic jet variables

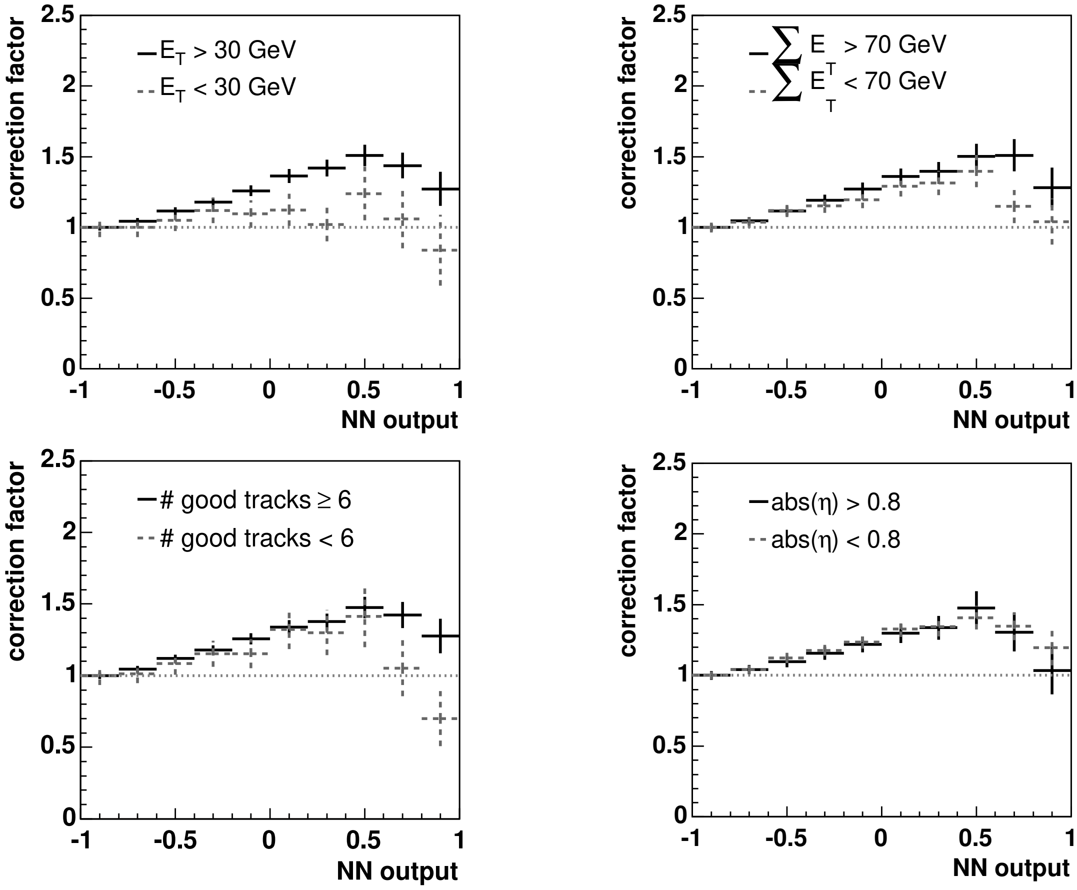
Table 7 indicates that the neural network is not very sensitive to the origin of the b jets. As it should be possible to use the neural network tagger in other analyses than single-top, we examine the influence of the kinematic jet variables[3] to the training result.

We trained another neural network without kinematic jet variables. For this neural network, all variables (except jet E_T and jet η) listed in tables 3 were used to train the neural network.

The plot showing the purity in each bin versus the neural network output for this training is shown on the left hand side in figure 53. One can see that the points are lying on the diagonal. That indicates that the network is well trained. The output of the neural network can be seen in figure 53 on the right hand side.

The output distributions of the samples used to train the network are shown in figure 54. One can see that the distributions for b jets are very similar. The same is true for the charm samples.

[3] jet E_T and jet η, see section 5



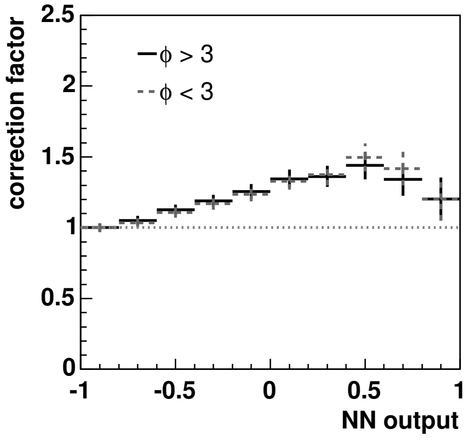
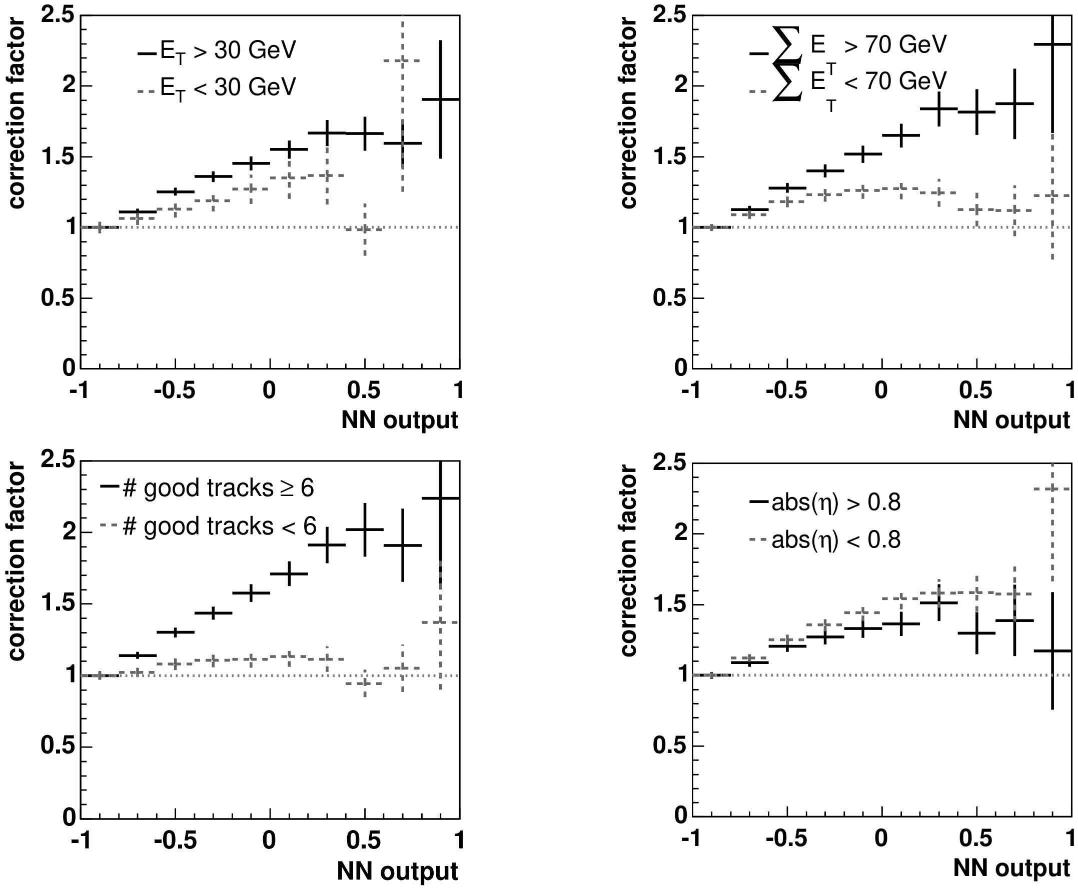


Figure 49: **Dependences of the deviations between data and simulation for pass 1 vertices.** The ratios of the accumulated output distributions of pass 1 vertices are shown. On the upper left hand side, the dependence of the correction on jet $\mathrm{E_T}$ is shown, on the upper right hand side the dependence on $\sum \mathrm{E_T}$. The plot on the middle left hand side shows the ratio for two bins of # good tracks in the jet. On the middle right hand side, two bins of abs(η) are shown. The plot on the lower left hand side shows the dependence on ϕ.



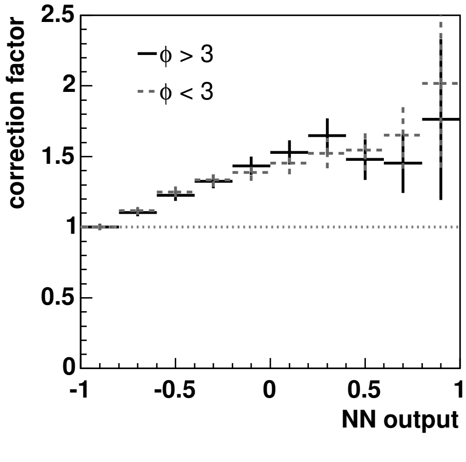
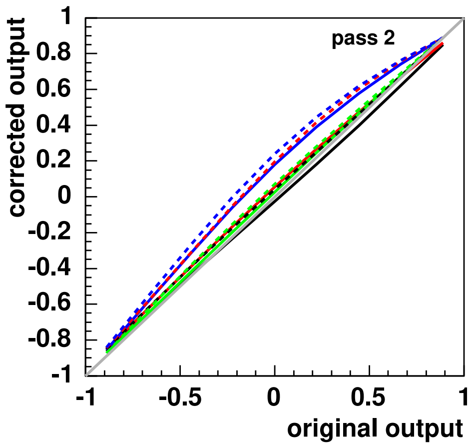
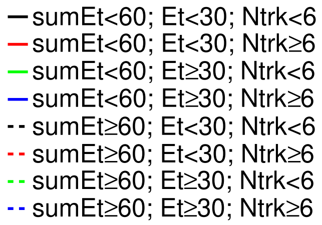


Figure 50: **Dependences of the deviations between data and simulation for pass 2 vertices.** The ratios of the accumulated output distributions of pass 2 vertices are shown. On the upper left hand side, the dependence of the correction on jet $\mathrm{E_T}$ is shown, on the upper right hand side the dependence on $\sum \mathrm{E_T}$. The plot on the middle left hand side shows the ratio for two bins of $\#$ good tracks in the jet. On the middle right hand side, two bins of abs(η) are shown. The plot on the lower left hand side shows the dependence on ϕ.





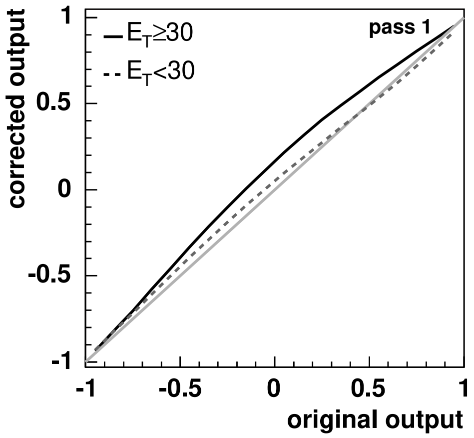


Figure 51: **Correction functions for pass 1 and pass 2 vertices.** The plot on the upper left hand side shows the correction functions for pass 2 vertices parameterized in different bins of jet $\mathrm{E_T}$, $\sum \mathrm{E_T}$ and number of good tracks in the jet. Besides it, its legend is shown. The plot on the lower left hand side shows the correction functions of pass 1 vertices in two bins of jet $\mathrm{E_T}$.

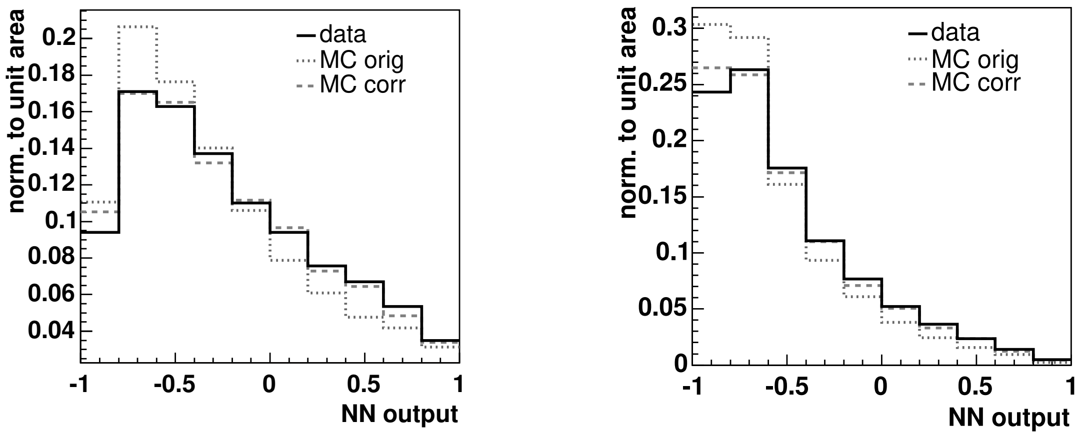


Figure 52: **Comparison of data, MC and corrected MC output distributions.** On the left hand side, the output distributions for pass 1 vertices are shown, on the right hand side for pass 2 vertices.

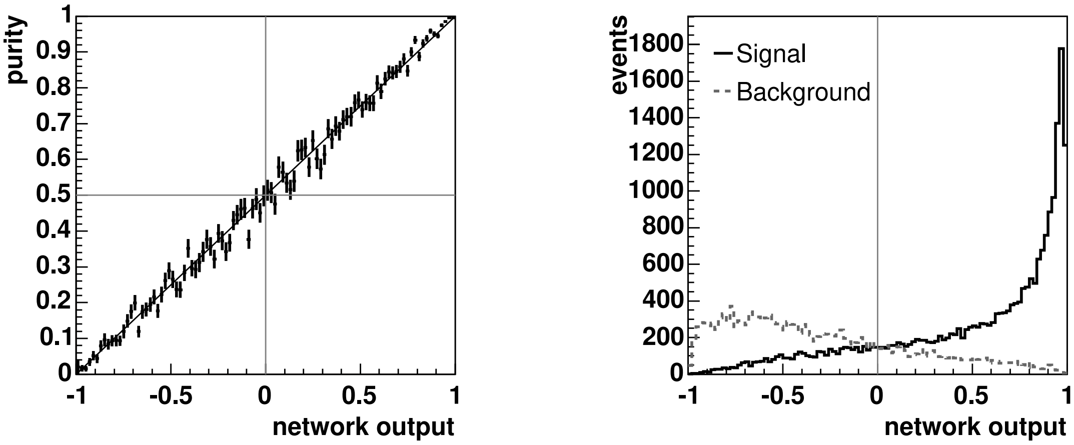


Figure 53: **Result of the training without jet $\mathrm{E_T}$ and jet η.** On the left hand side, the purity in each output bin is plotted versus the output. On the right hand side, the output distribution of the trained neural network is shown for signal and background.

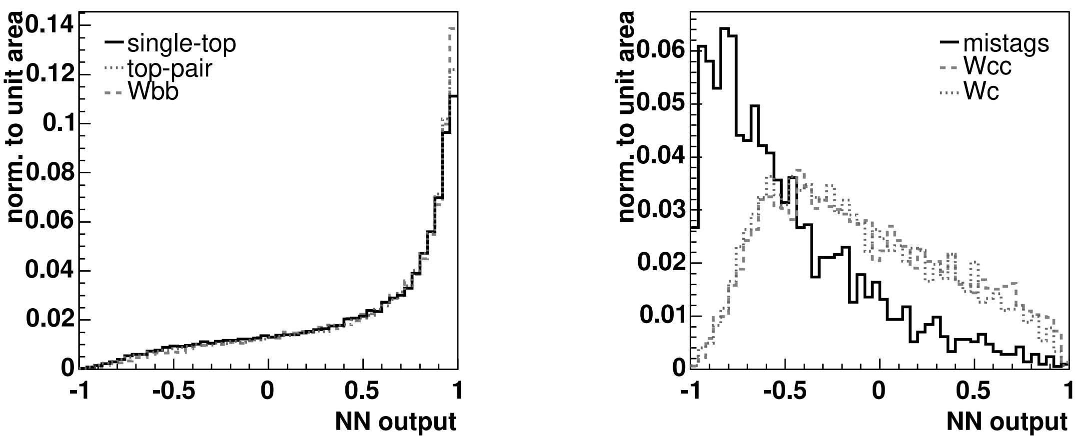


Figure 54: **Output distributions for the different samples for the training without jet $\mathrm{E_T}$ and jet η.** On the left hand side, the output distributions for single–top, $t\bar{t}$ and $Wb\bar{b}$ is shown. On the right hand side, the output distributions are shown for $Wc\bar{c}$, Wc and mistags. MC truth is used as described in section 5.

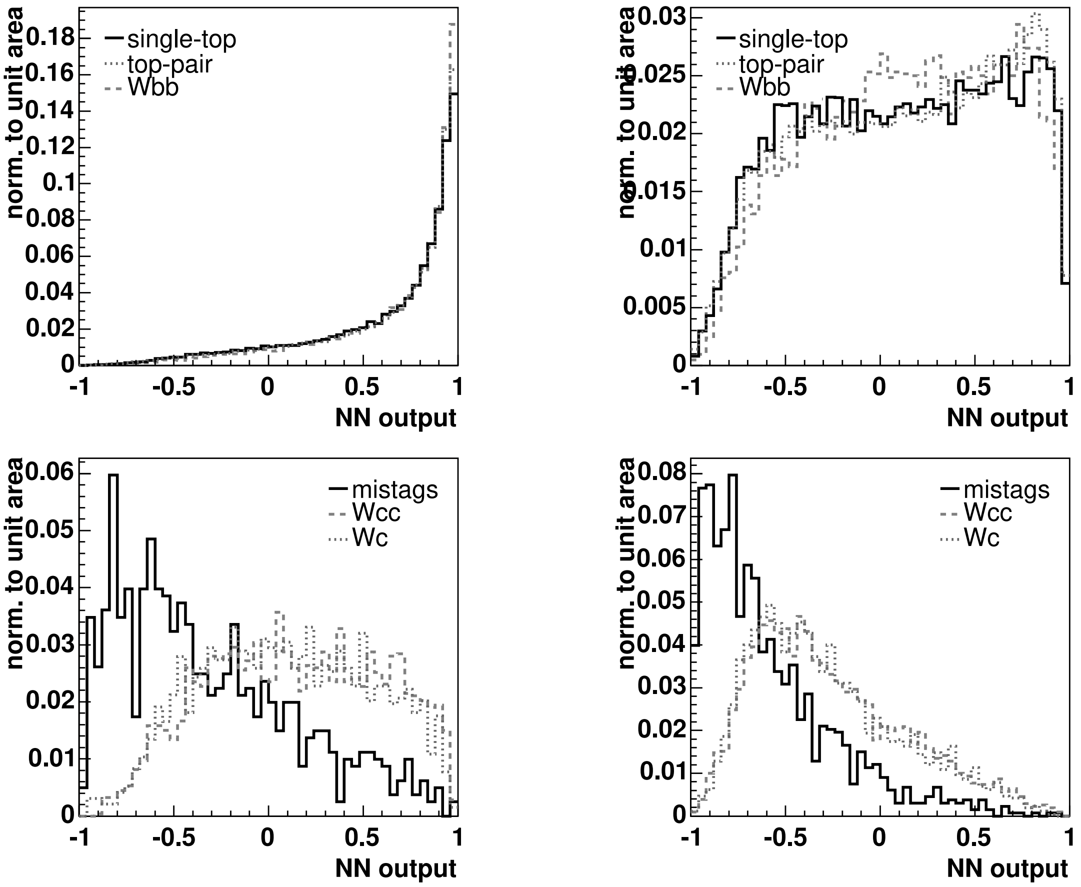


Figure 55: **Output distributions for the different samples.** The upper plots show the output distributions for single–top, $t\bar{t}$ and $Wb\bar{b}$ for pass 1 (left) and pass 2 (right) vertices. The lower plots show the output distributions for $Wc\bar{c}$, Wc and mistags for pass 1 (left) and pass 2 (right) vertices. MC truth is used as described in section 5.

In figure 55, one can see the output distributions of the training samples split up into pass 1 and pass 2 secondary vertices. One can see the same behavior as described in section 5 for the neural network using jet $\mathrm{E_T}$ and jet η.

The comparison of the purity versus efficiency plots for the neural network using jet $\mathrm{E_T}$ and jet η and the one without those variables is shown in figure 56. One can see that the performance of the network without those kinematic jet variables is worse.

Again, we optimized $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top to find the best cut on the neural network output. The best cut for the single–top analysis would be NN–output > -0.2. In table 13 one can see that the improvement of the ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top is only 7.4% while it is 9.5% for the neural network using jet $\mathrm{E_T}$ and jet η.

As before, we found three cuts leading to almost the same ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$. The differences are on the per mille level. The efficiencies for those cuts can be found in table 14. If one compares those numbers with the efficiencies of the neural network using jet $\mathrm{E_T}$ and jet η (see table 7), one can see that the background suppression is smaller for a given signal efficiency for the network without using jet kinematic variables.

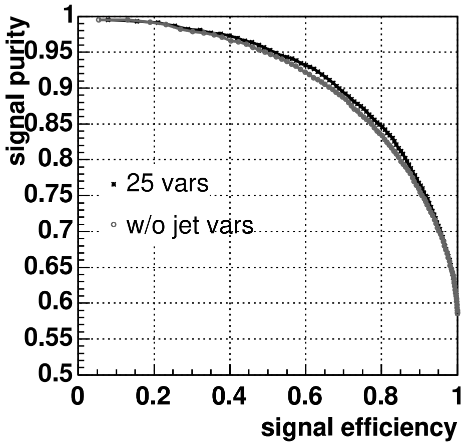


Figure 56: **Comparison of the trainings w/ and w/o jet $\mathrm{E_T}$ and η** In black, the training using the jet kinematic variables is shown as described before, in grey the one without jet $\mathrm{E_T}$ and η.

sample	$\frac{S}{\sqrt{B}}$
before cut	0.95
w/ NN-output > -0.2	1.02
improvement	7.4%

Table 13: **Ratio of $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top for the training without jet $\mathrm{E_T}$ and jet η.** The ratio is shown before the application of the neural network and after the application of the neural network. For the determination of the ratio with the cut on the NN output, the efficiencies shown in table 14 are taken. For non–W and di–boson, the average b signal efficiency is assumed. Both s– and t–channel are taken as signal.

From these numbers we conclude that it is advantageous to use the kinematic variables, $\mathrm{E_T}$ and η of the jet, for the single–top analysis. But also $t\bar{t}$ analyses would benefit from using the neural network including the kinematic variables.

13.1 Data–MC comparison

As described in section 11.1, we compare the output distributions of the neural network without jet kinematic variables in data and MC. The output distributions, the accumulated distributions and the ratio of the accumulated distributions can be seen in figure 57.

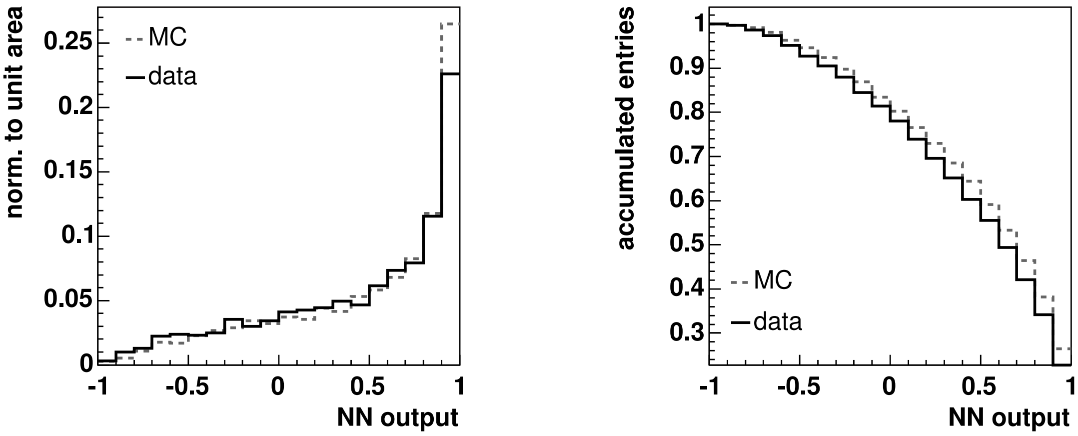
In figure 58, the ratios of the accumulated output distributions are shown separately for pass 1 and pass 2 vertices.

We draw the same conclusions as for the network using jet $\mathrm{E_T}$ and η (see section 11.1): For a cut-based use, a correction factor will be determined separately for pass 1 and pass 2 vertices using figure 58. Again, one can see that a correction is only necessary for pass 1 vertices.

The correction factors for pass 1 vertices for the cut–based use of the neural network without jet kinematic variables can be read out of table 15. For the best cut on the neural network output for this neural network >-0.2, the correction factor would be 0.985 ± 0.042

Sample	output > -0.3	output > -0.2	output > -0.1
t–channel	88.8%	86.0%	83.0%
s–channel	89.1%	86.2%	83.2%
$t\bar{t}$	88.0%	85.2%	82.2%
$Wb\bar{b}$	91.4%	88.8%	86.0%
$Wc\bar{c}$	61.5%	53.8%	46.9%
Wc	58.9%	51.2%	43.9%
Mistags	27.0%	21.6%	17.3%

Table 14: **Efficiencies of the best cuts for the single–top analysis for the network w/o jet kinematics.** The percentages of remaining events for the three best cuts on the neural network output are shown for the single–top analysis.



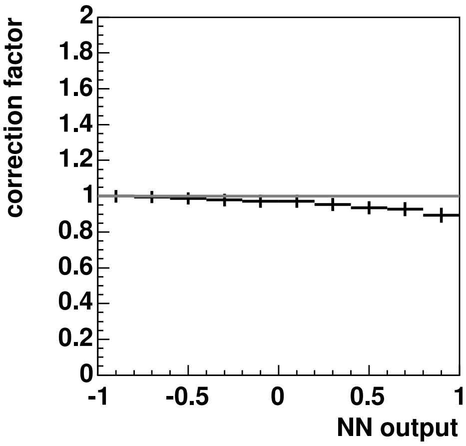
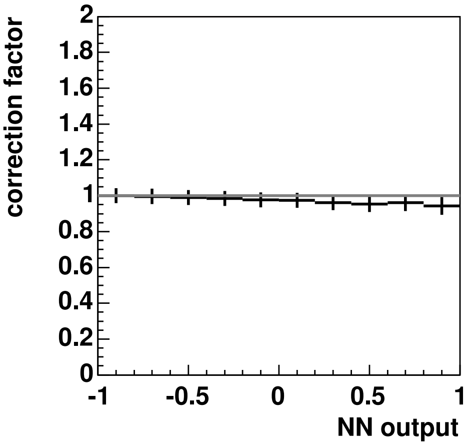


Figure 57: **Comparison of the network output for data and Monte Carlo for the network without jet kinematic variables.** The output distributions for data (solid black) and MC (dashed grey) are shown on the upper left hand side. The accumulated normalized output distributions of the away jet in data (solid black) and MC (dashed grey) are shown on the upper right hand side. On the lower left hand side, the ratio of the accumulated output distribution in data and in Monte Carlo is shown. For these plots, both electron and away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.



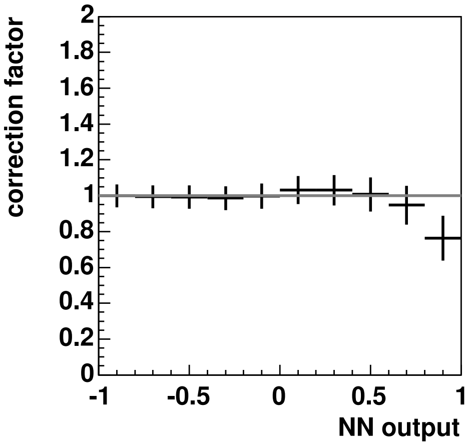


Figure 58: **Correction factor for the network without jet kinematic variables for pass 1 and pass 2 vertices.** On the left hand side, the ratio of the accumulated data distribution divided by the accumulated MC distribution is shown for pass 1 vertices, on the right hand side for pass 2 vertices.

Cut	>-0.9	>-0.7	>-0.5	>-0.3	>-0.1
SF	1.000 ± 0.042	0.996 ± 0.042	0.990 ± 0.042	0.985 ± 0.042	0.978 ± 0.042
Cut	>0.1	>0.3	>0.5	>0.7	>0.9
SF	0.975 ± 0.042	0.962 ± 0.042	0.952 ± 0.043	0.960 ± 0.045	0.943 ± 0.049

Table 15: **Correction factor for pass 1 vertices for the cut–based use of the neural network without jet kinematic variables.** These factors are read out from the plot on the left hand side in figure 58. The table has to be read from left to right. The last bin which fully includes the interval of the chosen cut determines the correction factor SF.

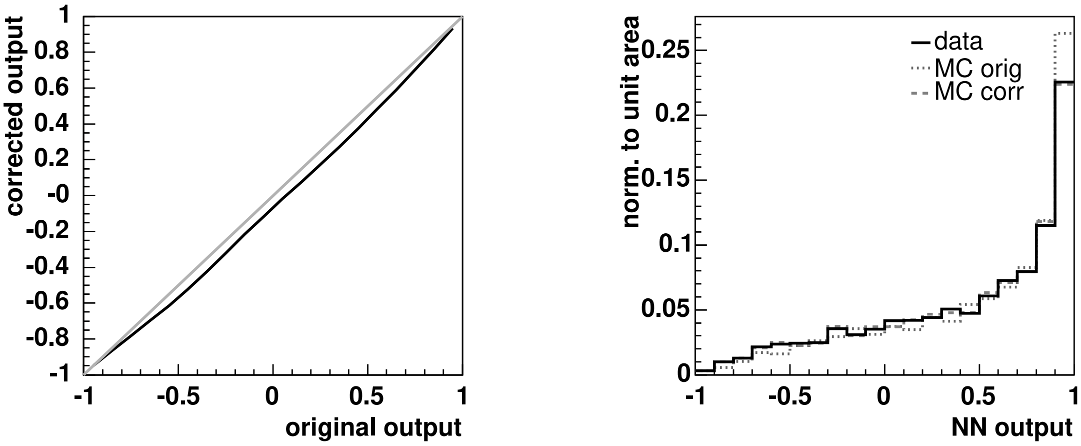


Figure 59: **Comparison of original and corrected MC output for the neural network without jet kinematic variables.** The plot on the left hand side shows the corrected MC output versus the original output of the away jet. The plot on the right hand side shows the comparison of the corrected and the original MC distributions with data. The dotted histogram shows the original MC output distribution. The corrected output distribution is shown in dashed grey. The solid black histogram represents the data distribution.

For the use of the tagger as continuous variable, e.g. as input to another neural network, the systematic uncertainty will be estimated the same way as described in section 11.1 by shifting the accumulated MC distribution to fit the data (see figure 57, upper right).

The correction function for the tagger without jet kinematic variables can be seen in figure 59 on the left hand side. The corrected output will be used to estimate the systematic uncertainty for the use of the tagger as a continuous variable. The comparison of the original and the corrected output with the data is shown in figure 59 on the right hand side.

14 Correction of the W + light jets sample

14.1 Network using kinematic jet variables

Assuming that mistagged light jets are described by negative tagged events, we correct the output distribution of the W+light jets sample for the single–top analysis by using the correction functions derived in section 12 (see figure 51). The corrected output distribution of the W +light jets sample can be seen in figure 60 in comparison to the original one. As expected, the corrected distribution is less background–like than the original one.

As described in section 6, we determined the best cut on the neural network output when using the corrected output distribution for the mistags described by the W + light jets sample. We found three cuts to lead to the same ratio $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ at the per mille level. The application of the neural network leads to an improvement in $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ of 8.4% if we use the

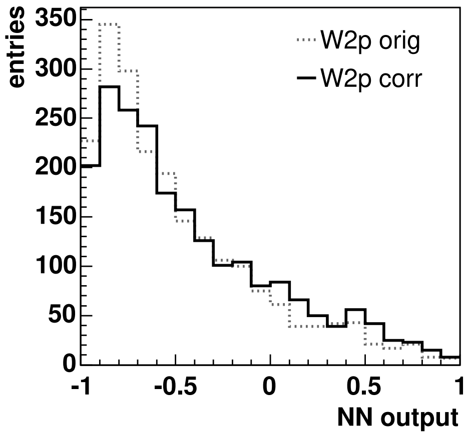


Figure 60: **Comparison of corrected and original output for W + light jets.** For the correction of the output, the function shown in figure 51 on the left hand side was used.

sample	$\frac{S}{\sqrt{B}}$
before cut	0.95
w/ NN-output > 0.0	1.03
improvement	8.4%

Table 16: **Ratio of $\frac{S}{\sqrt{B}}$ for single–top when W + light jets is corrected.** The ratio is shown before the application of the neural network and after the application of the neural network. For the determination of the ratio with the cut on the NN output, the efficiencies shown in table 17 are taken. For non–W and di–boson, the average b signal efficiency is assumed. Both s– and t–channel are taken as signal. The output of the W + light jets sample was corrected as described in section 12.

corrected distribution for the W + light jets sample (see table 16). The related efficiencies can be found in table 17. It is obvious that the mistag suppression is slightly worse due to the corrected output distribution.

15 Cross check of determination of the composition of the inclusive electron data

In sections 10 and 11.3, the composition of the inclusive electron data was determined by fitting templates of the mass at the secondary vertex to the data distribution. In this section, we determine the composition by using the output distribution of the neural network tagger itself. As the output distributions for b, charm and light jets are very different, it should be possible to determine the contributions of those templates in data.

In figure 61, the templates used to fit the NN output distribution of the away jet are shown as well as the results of the fit to the data distribution. Within the given errors, the data distribution can be well described by fitting templates of the NN output distribution. The estimated composition of the away jet is shown in table 18 for the case that only the away jet is tagged by SecVtx and in table 19 for the case that both electron and away jet

Sample	output > -0.2	output > -0.1	output > 0.0
t–channel	87.4%	84.6%	82.1%
s–channel	87.1%	84.2%	81.4%
$t\bar{t}$	85.7%	82.9%	80.1%
$Wb\bar{b}$	88.1%	85.4%	82.6%
$Wc\bar{c}$	49.7%	43.5%	37.2%
Wc	48.2%	40.8%	34.5%
Mistags	27.7%	22.9%	19.1%

Table 17: **Efficiencies of the best cuts for the single–top analysis when W + light jets is corrected.** The percentages of remaining events for the three best cuts on the neural network output are shown for the single–top analysis. The output distribution of the W + light jets sample was corrected with the correction function displayed in figure 51 on the left hand side.

sample	b fraction	c fraction	light fraction
MC	88.9%	8.8%	2.3%
data (secvMass)	72.9% ± 1.9%	17.2% ± 2.8%	9.9 ± 1.9%
data (NN output)	74.9% ± 1.5%	20.0% ± 1.8%	5.1 ± 1.0%

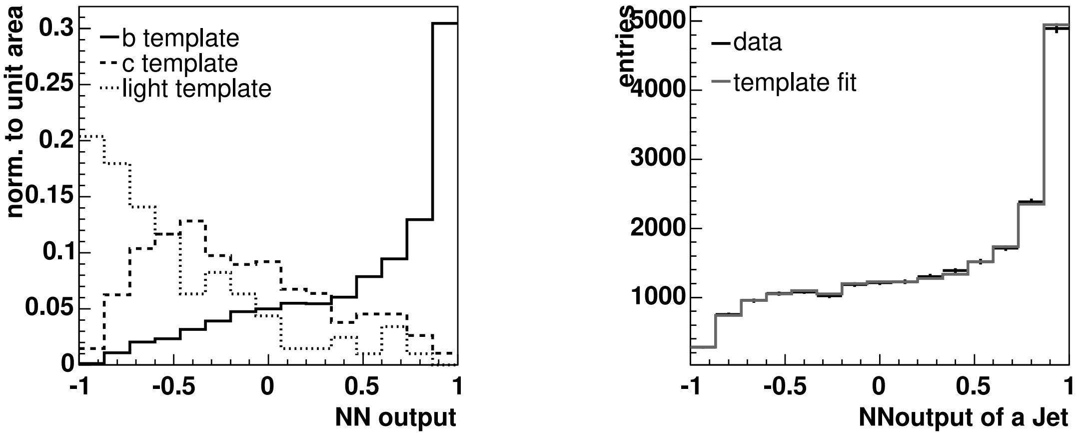
Table 18: **Composition of the away jet in data determined by NN b tag template fit; only away jet tagged** The composition in the simulation as well as the result obtained by the secondary vertex mass fit are taken from table8. The fractions in the third row were determined by fitting templates of the NN b tag output distributions to the data. For these numbers, only the away jet was required to be tagged by SecVtx, no matter if the electron jet is tagged or not. Events where the electron is from a conversion are excluded.

are tagged by SecVtx. The corresponding results for the electron jet can be seen in figure 62 and tables 20 and 21.

In general, it can be seen that the estimated composition in data is the same for fitting templates of the mass at the secondary vertex and for fitting templates of the NN output distribution. These results indicate that fitting templates of the NN output distribution is not only a valid cross check but also a promising tool for estimating data composition and even for background estimation.

16 Cross check with other samples

In this section, we compare the output distributions of the samples used to train the neural network with output distributions of completely independent samples. For this cross check, we use the btopXa samples and some W+$Higgs$ (WH115_CTEQ5L_ISRon_FSRon, WH120_CTEQ5L_ISRon_FSRon, WH130_CTEQ5L_ISRon_FSRon) samples taking into account the MC truth.



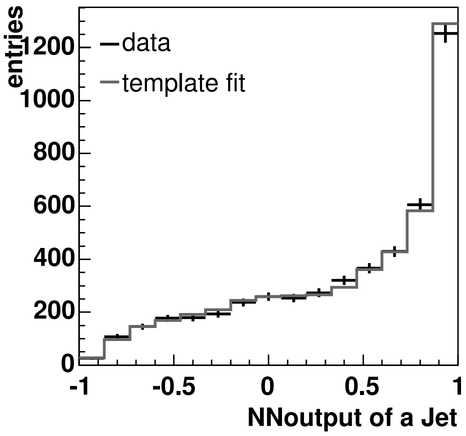
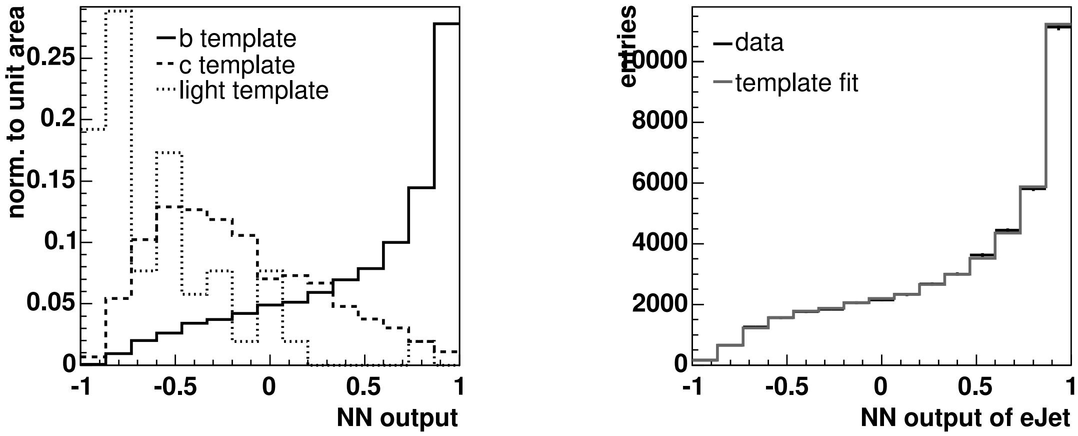


Figure 61: **Cross check of estimation of the composition of the away jet in data.** The templates shown on the upper left hand side are obtained from the btop5a, btop6a, btop7a and btop8a samples with only the away jet required to be tagged. On the upper right hand side, the result of the fit of those templates to the data distribution is shown for the case that only the away jet is required to be tagged by SecVtx, on the lower left hand side for the case that both electron and away jet are tagged by SecVtx.

sample	b fraction	c fraction	light fraction
MC	95.0%	3.2%	1.8%
data (secvMass)	87.2% ± 3.0%	9.9% ± 3.8%	2.9 ± 2.0%
data (NN output)	88.9% ± 2.2%	9.3% ± 1.9%	1.9 ± 0.8%

Table 19: **Composition of the away jet in data determined by NN b tag template fit; both jets tagged.** The composition in the simulation as well as the result obtained by the secondary vertex mass fit are taken from table9. The fractions in the third row were determined by fitting templates of the NN b tag output distributions to the data. For these numbers, both the electron and the away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.



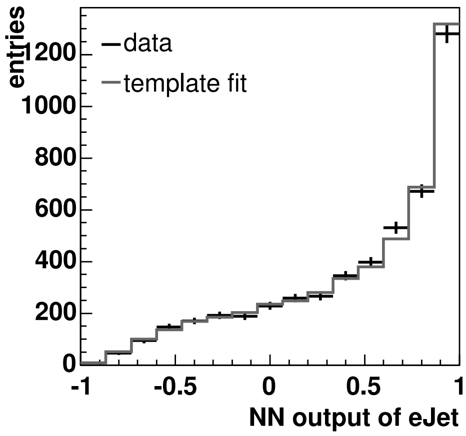


Figure 62: **Cross check of estimation of the composition of the electron jet in data.** The templates shown on the upper left hand side are obtained from the btop5a, btop6a, btop7a and btop8a samples with only the electron jet required to be tagged. On the upper right hand side, the result of the fit of those templates to the data distribution is shown for the case that only the electron jet is required to be tagged by SecVtx, on the lower left hand side for the case that both electron and away jet are tagged by SecVtx.

sample	b fraction	c fraction	light fraction
MC	92.7%	7.2%	0.1%
data (secvMass)	90.7% ± 1.1%	6.2% ± 1.2%	3.1 ± 1.1%
data (NN output)	92.0% ± 0.9%	6.9% ± 0.7%	1.1 ± 0.3%

Table 20: **Composition of the electron jet in data determined by NN b tag template fit; only electron jet tagged.** The composition in the simulation as well as the result obtained by the secondary vertex mass fit are taken from table11. The fractions in the third row were determined by fitting templates of the NN b tag output distributions to the data. For these numbers, only the electron jet was required to be tagged by SecVtx, no matter if the away jet is tagged or not. Events where the electron is from a conversion are excluded.

sample	b fraction	c fraction	light fraction
MC	98.0%	1.9%	0.04%
data (secvMass)	97.6% ± 2.1%	0.9% ± 1.7%	1.5 ± 1.0%
data (NN output)	98.8% ± 2.2%	0.8% ± 1.4%	0.4 ± 0.3%

Table 21: **Composition of the electron jet in data determined by NN b tag template fit; both jets tagged.** The composition in the simulation as well as the result obtained by the secondary vertex mass fit are taken from table11. The fractions in the third row were determined by fitting templates of the NN b tag output distributions to the data. For these numbers, both the electron and the away jet are required to be tagged by SecVtx. Events where the electron is from a conversion are excluded.

The jet E_T spectra of the different samples can be seen in figure 63.

16.1 Network using kinematic jet variables

In figure 64, one can see the output distributions of the several samples for the training using jet E_T and jet η together for pass 1 and pass 2 vertices. In figure 65, the output distributions for pass 1 and pass 2 vertices are shown separately. One can see that the distributions of the different samples are very well comparable.

16.2 Network without jet E_T and jet η

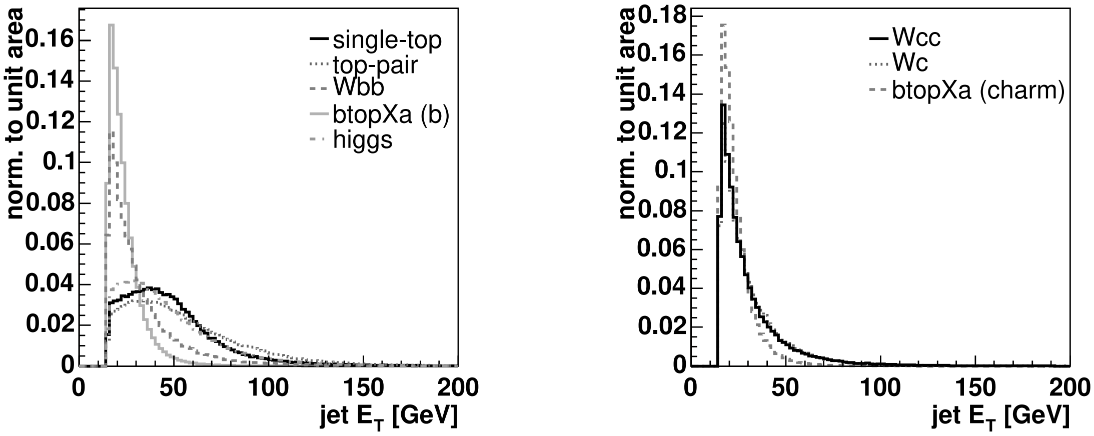
In figure 66, one can see the output distributions of the training not explicitly using jet kinematic variables. All distributions look similar. But if one splits up the output distributions for pass 1 and pass 2 vertices as shown in figure 67, it is obvious that the output distribution of the btopXa sample looks different especially for b jets with a pass 2 secondary vertex (see figure 67, upper right). For those kinds of jets, the output distribution of b jets in the btopXa sample is more signal–like than for the other b samples.

The reason for that difference is that the btopXa samples are stripped for an electron at HEPG level. This electron enhancement leads to more signal–like output distributions. In figure 68, the output distributions of the different b samples are shown separately for jets with no electron at all and for jets with at least one electron. Now the distributions look completely comparable.

16.2.1 Discussion of influence of electron enhancement to the network using jet kinematics

As we have shown in section 16.1, we do not see the influence of the btopXa electron enhancement to the output distribution if we use jet E_T and jet η to train the neural network.

In figure 69, we compare the output distribution of the pass 2 btopXa vertices for the training using jet E_T and η and for the one without. It is obvious that the inclusion of jet kinematic variables to the training makes a very small difference for the jets without any electrons (see figure 69, left). For the jets that do contain electrons, the output



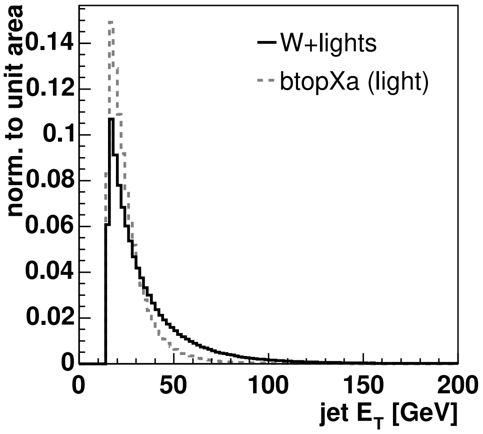
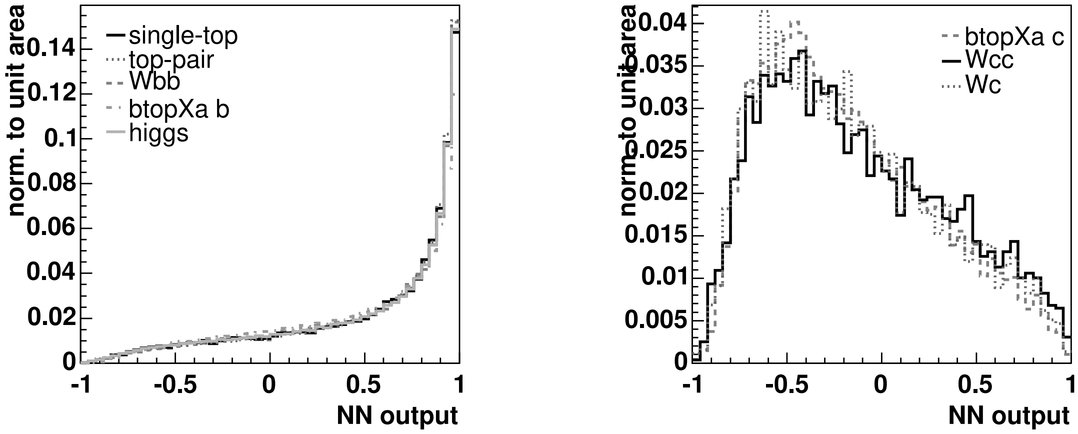


Figure 63: $\mathrm{E_T}$ **spectra of the different samples.** The jet $\mathrm{E_T}$ spectra of the different b, charm and light samples are shown. The b samples can be seen on the upper left hand side, the charm samples on the upper right hand side and the light samples on the lower left hand side.



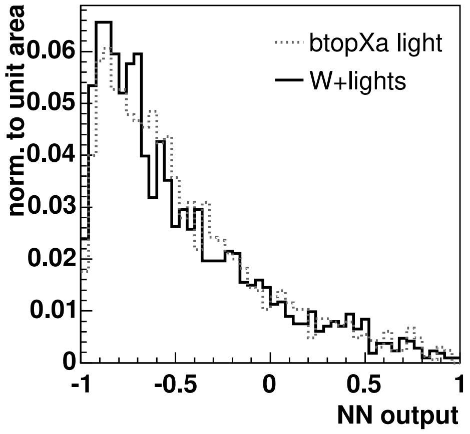


Figure 64: **Output distributions for the different samples.** The plot on the upper left hand side shows the comparison of the different b samples, the plot on the upper right hand side the comparison of the different charm samples and the plot on the lower left hand side the different light samples.

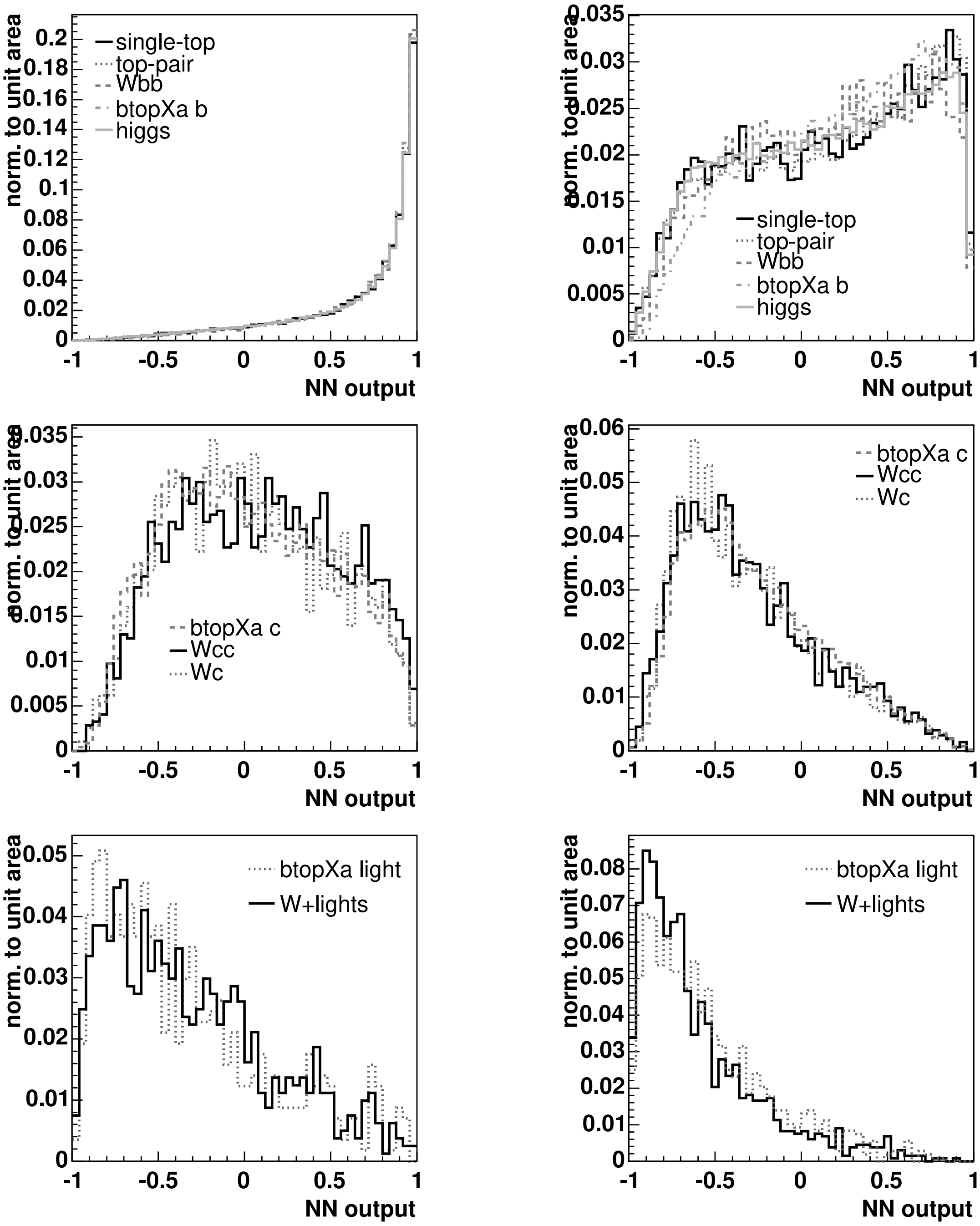
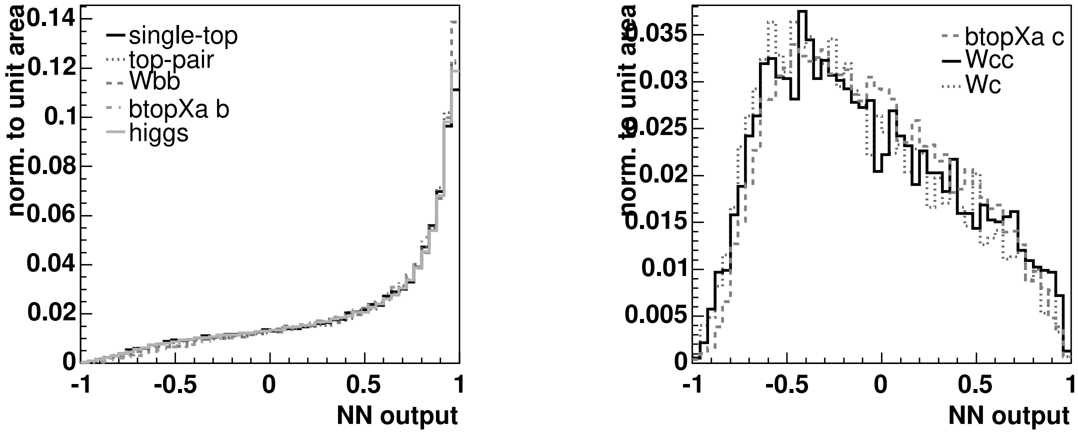


Figure 65: **Output distributions for the different samples split up for pass 1 and pass 2 vertices.** The upper plots show the distributions of the b samples, the plots in the middle the distributions of the charm samples and the lower plots the distributions of the light samples. The plots on the left show the distributions for pass 1 vertices, the plots on the right the distributions for pass 2 vertices.



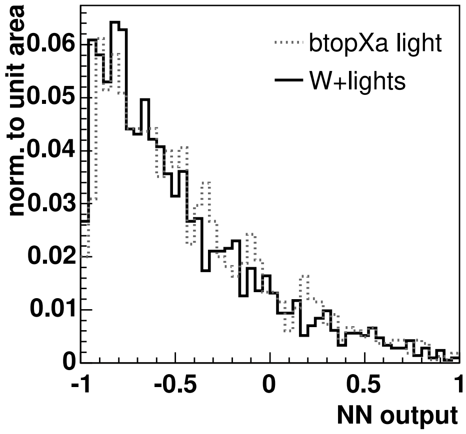


Figure 66: **Output distributions for the different samples for the training without jet $\mathrm{E_T}$ and η.** The plot on the upper left hand side shows the comparison of the different b samples, the plot on the upper right hand side the comparison of the different charm samples and the plot on the lower left hand side the different light samples.

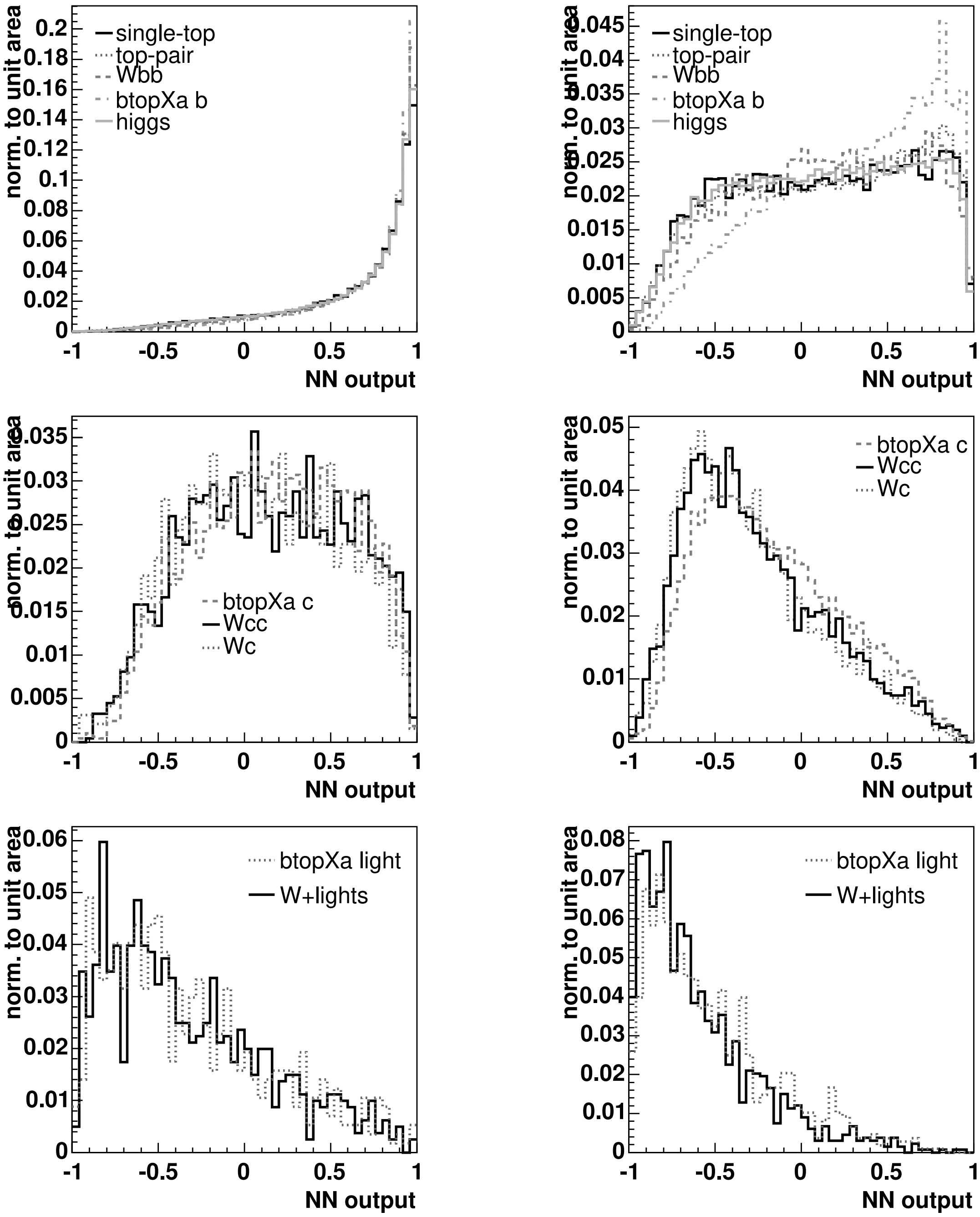
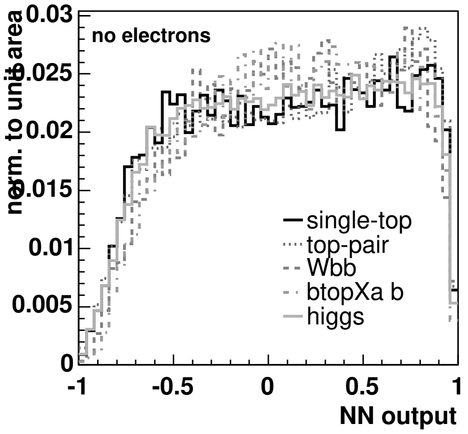


Figure 67: **Output distributions for the different samples for the training without jet $\mathrm{E_T}$ and η split up for pass 1 and pass 2 vertices.** The upper plots show the distributions of the b samples, the plots in the middle the distributions of the charm samples and the lower plots the distributions of the light samples. The plots on the left show the distributions for pass 1 vertices, the plots on the right the distributions for pass 2 vertices.



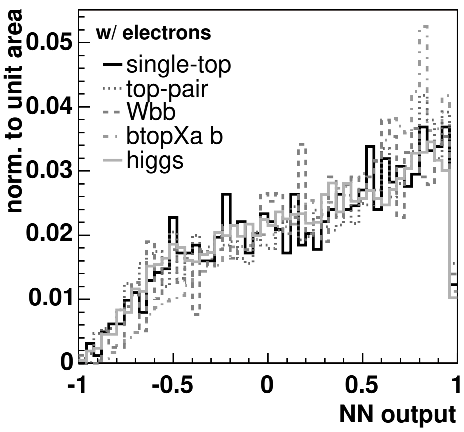


Figure 68: **Output distributions using the training without jet kinematic variables for pass 2 b vertices in dependence of presence of electrons.** The plot on the left hand side shows the output distributions of pass 2 b vertices where no electron is in the tagged jet. The plot on the right hand side shows the output distributions of b samples with at least one electron in the tagged jet.

distributions are slightly more background–like if the neural network is trained with jet kinematic variables.

But this feature is not the only reason for the similarity of the output distributions of the b samples for the training using jet kinematics. In addition to the btopXa distribution looking slightly more background–like, the top and Higgs samples look more signal–like, at least for jets which do not contain electrons.

In figure 70, the output distributions of the top and Higgs samples are compared for the trainings with and without jet kinematic variables for jets without any electron. One can see that the output distributions of the top and Higgs samples are more signal–like if the neural network was trained with jet E_T and jet η. Again, this behavior leads to more similar output distributions of btopXa and the other b samples when using jet kinematic variables as shown in figure 65.

For pass 2 vertices in jets containing electrons, the inclusion of jet kinematic variables does not make any difference for top and Higgs samples [1]. For $Wb\bar{b}$, the inclusion of jet kinematics to the training does not change the output distributions of pass 2 vertices[1], no matter if there is an electron in the jet or not. The same is true for the pass 1 vertices that are related to a real b hadron[1].

In conclusion, we consider the usage of the kinematic variables E_T and η as an advantage and we recommend the usage of the neural network using these variables.

17 Determination of systematics

To use the neural network b tagger as a continuous variable, e.g. as input to another neural network like for the single–top analysis, the systematic shape uncertainty of the

[1]Comparison not shown.

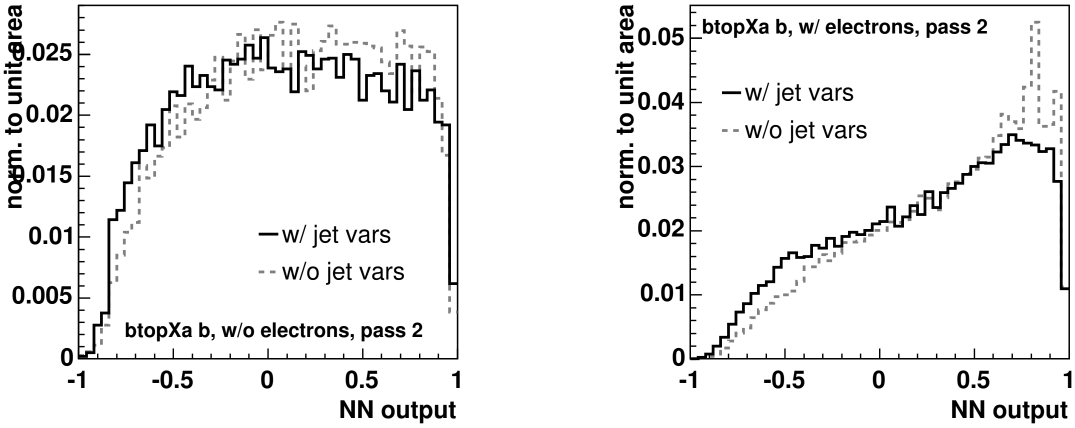


Figure 69: **btopXa output distributions for the trainings with and without jet kinematic variables.** The output distributions of jets with a pass 2 secondary vertex are shown. On the left hand side, jets with no electron are shown, on the right hand side, jets with at least one electron.

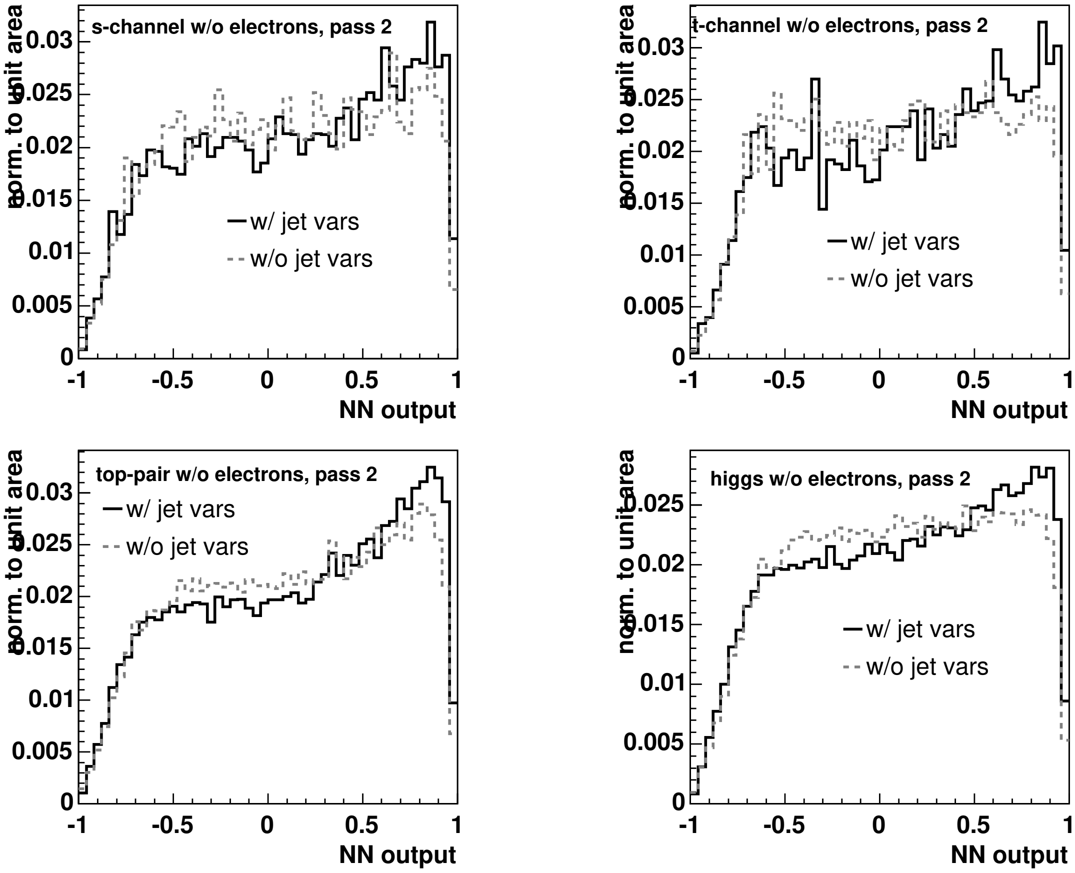


Figure 70: **Top and Higgs output distributions for the trainings with and without jet kinematic variables.** Jets that do not contain electrons and that have a pass 2 secondary vertex are shown.

output distribution has to be determined.

In the single–top analysis, this is done by applying the correction functions to the neural net output and thus creating skewed templates of the output distributions. With those templates, we investigate two scenarios to determine the systematic shape uncertainty:

1. pessimistic:
   apply b shift to b, mistag shift to charm and mistags.

2. optimistic:
   don't correct b and mistags, apply b shift to charm.

18 Conclusion

We presented a neural network b tagger applicable to events with a secondary vertex tag. The network includes information on the vertex mass, decay multiplicity and leptonic decays into the b jet identification, in addition to the lifetime of B hadrons. The network b tagger can be used to improve the purity of the tagged sample. Correcting the light jets for the differences between data and simulation and cutting on the network output at 0.0 keeps 82% of the tagged b jets while rejecting 64% of the tagged charmed jets and 81% of the tagged light jets. The gain in $\frac{\mathrm{S}}{\sqrt{\mathrm{B}}}$ for single–top is 8.4% using those numbers.

We have made detailed comparisons between data and Monte Carlo. The input distributions to the neural network show in general a good agreement between data and Monte Carlo. Residual differences are partially due to the difficulty of deriving adequate control samples in data and Monte Carlo which feature the identical flavor composition (b, c, light).

We observe differences in the neural network output distribution in signal–like and background–like samples. We derived a correction method to account for these differences. If a cut is applied to the neural network output, a simple correction factor for the cut efficiency is sufficient. If the neural network output is used further as a continuous variable, for example as an input to another neural network, the b tag network output is shifted by a correction function.

For signal–like events, we recommend to use the uncorrected output distribution and to estimate the systematic uncertainty by using the shifted distribution or the correction factor. For the simulated mistags, we will use an output distribution which is corrected for the differences seen between the mistag output distributions in data and Monte Carlo.

The neural network tagger is implemented in the single top analysis.

It is shown that the tagger is not very sensitive to the origin of the b jets. Therefore our network tagger is more generally applicable and can be used in other analyses like top–quark–pair and Higgs searches as well, even it is not completely optimized.

For analyses that deal with b jets with softer $\mathrm{E_T}$ spectra, it could be best to use the neural network that was trained without jet kinematic variables because we saw a small influence in the btopXa control samples.

List of Figures

1 Distributions of discriminating SecVtx variables. 5
2 Distributions of more discriminating SecVtx variables. 6
3 Distributions of more discriminating SecVtx variables. 7
4 Distributions of kinematic discriminating track variables. 8
5 Distributions of more kinematic discriminating track variables. 9
6 Distributions of discriminating track variables. 10
7 Distributions of discriminating track variables. 11
8 Distributions of discriminating track variables. 12
9 Distributions of discriminating lepton variables. 13
10 Distributions of discriminating jet variables. 14
11 General geometry of a 3 layer neural network. 15
12 The transformed sigmoid activation function S(a(x)) 15
13 Distributions of discriminating SecVtx variables for pass 1 and pass 2. . . . 18
14 More distributions of discriminating SecVtx variables for pass 1 and pass 2. 19
15 Result of the training. 23
16 Output distributions for the different samples. 24
17 Output distributions for the different samples. 25
18 Signal purity vs. signal efficiency. 25
19 Signal purity vs. signal efficiency for the different trainings. 26
20 The normal frequency function: 29
21 Comparison of discriminating SecVtx variables with data. 30
22 Comparison of more discriminating SecVtx variables with data. 31
23 Comparison of more discriminating SecVtx variables with data. 32
24 Comparison of kinematic discriminating track variables with data. 32
25 Comparison of more kinematic discriminating track variables with data. . . 33
26 Comparison of discriminating track variables with data. 34
27 Comparison of discriminating track variables with data. 35
28 Comparison of discriminating track variables with data. 36
29 Comparison of discriminating lepton variables with data. 37
30 Comparison of discriminating jet variables with data. 38
31 Comparison with data for variables that are used separately for pass 1 and pass 2. 39
32 Comparison with data for variables that are used separately for pass 1 and pass 2. 40
33 Distributions of the transverse momentum of the secondary vertex and the momentum of the 1.track taking into account the differences concerning the electron. 41
34 Estimation of the composition of the away jet in data. 44
35 Comparison of the network output of the away jet for data and Monte Carlo. 46
36 Comparison of the accumulated network output for data and Monte Carlo if only the away jet is tagged. 47
37 Comparison of the accumulated network output for data and Monte Carlo. 47
38 Correction factor for pass 1 and pass 2 vertices. 48

39 Comparison of original and corrected MC output. 49
40 E_T dependence of the correction function. 50
41 Estimation of the composition of the electron jet in data. 51
42 Comparison of the network output of the electron jet for data and Monte Carlo. 52
43 Comparison of the accumulated network output of the electron jet for data and Monte Carlo for only electron jet tagged. 53
44 Comparison of the accumulated network output of the electron jet for data and Monte Carlo. 53
45 Jet E_T spectra for jet samples in data and simulation. 54
46 Comparison of negatively tagged events. 55
47 Comparison of accumulated output distributions for data and Monte Carlo for negatively tagged jets. 56
48 Comparison of negatively tagged events separately for pass 1 and pass 2. . 56
49 Dependences of the deviations between data and simulation for pass 1 vertices. 58
50 Dependences of the deviations between data and simulation for pass 2 vertices. 59
51 Correction functions for pass 1 and pass 2 vertices. 60
52 Comparison of data, MC and corrected MC output distributions. 60
53 Result of the training without jet E_T and jet η. 61
54 Output distributions for the different samples for the training without jet E_T and jet η. 61
55 Output distributions for the different samples. 62
56 Comparison of the trainings w/ and w/o jet E_T and η 63
57 Comparison of the network output for data and Monte Carlo for the network without jet kinematic variables. 64
58 Correction factor for the network without jet kinematic variables for pass 1 and pass 2 vertices. 65
59 Comparison of original and corrected MC output for the neural network without jet kinematic variables. 66
60 Comparison of corrected and original output for W + light jets. 67
61 Cross check of estimation of the composition of the away jet in data. . . . 69
62 Cross check of estimation of the composition of the electron jet in data. . . 70
63 E_T spectra of the different samples. 72
64 Output distributions for the different samples. 73
65 Output distributions for the different samples split up for pass 1 and pass 2 vertices. 74
66 Output distributions for the different samples for the training without jet E_T and η. 75
67 Output distributions for the different samples for the training without jet E_T and η split up for pass 1 and pass 2 vertices. 76
68 Output distributions using the training without jet kinematic variables for pass 2 b vertices in dependence of presence of electrons. 77
69 btopXa output distributions for the trainings with and without jet kinematic variables. 78

70 Top and Higgs output distributions for the trainings with and without jet kinematic variables. 78

References

[1] C. Ciobanu et al., MadEvent Signal Samples used in the Run II Single–Top Search, CDF Note 7020, 2004.

[2] D. Acosta et al., Search for Electroweak Single Top Quark Production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$TeV, Phys.Rev.D71:012005, hep-ex/0410058, 2005.

[3] W. Yao et al., A Seed Vertex b-Tag Algorithm for Top, CDF Note 2716, 1994.

[4] C. Neu et al., SecVtx Optimization Studies for 5.3.3 Analyses, CDF Note 7578, 2005.

[5] H. Bachacou et al., Heavy Flavor Contributions to the SECVTX–tagged W + Jets Sample, CDF Note 7007, 2004.

[6] S. Budd et al., Event detection efficiency for single top events in CDF II, CDF Note 7863, 2005.

[7] W. Wagner, Talk in CDF Top Properties Meeting 10/14/2005.

[8] M. Feindt, A Neural Bayesian Estimator for Conditional Probability Densities, e-Print Archive physics/0402093, 2004.

[9] M. Feindt and U. Kerzel, The NeuroBayes Neural Network Package, Proceedings of the ACAT 05, to be published.

[10] A. Holloway et al., SecVtx Tagging Scale Factor for 5.3.3 Analysis, CDF Note 7333, 2004.

[11] P. Lujan et al., Efficiency of SecVtx B–Tag Used for 2005 Winter Conferences, CDF Note 7343, 2005.

[12] G. J. Barker et al., Monte Carlo Study of lepton SVT sample, CDF Note 7318, 2004.

[13] S. Baroiant et al., Measurement of Electron Trigger Efficiencies for Level1 and Level2 8 GeV Triggers, CDF Note 6257, 2003.

[14] M. Baumgart et al., Determining the b and c content of jet samples using the SECVTX vertex mass, CDF Note 6675, 2003.

[15] A. Foland et al., SecVtx Tag Matrices for 5.3.3_nt, CDF Note 7326, 2004.