$t\bar{t}$ Production Cross Section in the All-Hadronic Decay Channel p17

Version 3.0

Michael Begel
Jeroen Hegeman
Hendrik Hoeth
David Lam
Gianluca Petrillo
Karel Soustruz

June 9, 2009

Abstract

The $t\bar{t}$ production cross section has been measured in the all-hadronic decay channel using $\mathcal{L} \approx 1$ fb^{-1} of Run IIa data (p17). Events with at least six high-p_T jets were used in this analysis. At least two of the jets were required to be b-tagged using the neural network b-tagger. All jets were required to be taggable to suppress the effects of additional $p\bar{p}$ interactions. The ALPGEN Monte Carlo generator was used to generate the $t\bar{t}$ signal samples. The heavily QCD-multi-jet-dominated background was modeled using a specially selected data sample. Signal was separated from background using a topological likelihood. The resulting $t\bar{t}$ production cross section was

$$\sigma_{t\bar{t}}(m_t = 170\,\mathrm{GeV}/c^2) = 7.9\ \pm 1.5\ (\mathrm{stat})\ \pm 1.6\ (\mathrm{sys})\ \pm 0.5\ (\mathrm{lum})\ \mathrm{pb}$$
$$\sigma_{t\bar{t}}(m_t = 175\,\mathrm{GeV}/c^2) = 6.9\ \pm 1.3\ (\mathrm{stat})\ \pm 1.4\ (\mathrm{sys})\ \pm 0.4\ (\mathrm{lum})\ \mathrm{pb}$$

The cross section agrees with theoretical expectations.

Contents

1 Introduction **3**
1.1 Top quark pair decay modes . . . 3
1.2 Measurement of the top quark pair production cross section . . . 5

2 Event Selection **7**
2.1 Samples . . . 7
2.2 Data quality requirements . . . 8
2.3 Integrated luminosity . . . 9
2.4 Trigger requirements . . . 9
2.4.1 Jet triggers: combinatorics . . . 10
2.4.2 Trigger Weight . . . 12
2.5 Jets . . . 12
2.6 Multi-Jet Selection . . . 13
2.7 Primary Vertex . . . 13
2.8 b-Jet identification . . . 14
2.9 Preselection Efficiency . . . 14

3 Background modeling **19**
3.1 Introduction . . . 19
3.2 Background generation procedure . . . 20
3.3 Background validation . . . 21
3.4 Systematic studies of the background modeling . . . 24
3.4.1 Phase space of the acceptor sample . . . 24
3.4.2 Variations of the shaping cuts . . . 27

4 Separating signal from background **31**
4.1 Likelihood input variables . . . 31
4.2 Likelihood performance . . . 33

5 Results **38**
5.1 Signal Purity . . . 38
5.2 Systematic uncertainties . . . 38
5.2.1 Signal model . . . 38
5.2.2 Jet energy, resolution and ID . . . 43
5.2.3 b-Jet related uncertainties . . . 43
5.2.4 Luminosity . . . 44
5.2.5 Primary vertex position and reweighting . . . 44
5.3 Result . . . 44

A Trigger Details **46**

B Likelihood control plots **49**
B.1 Control plots after cutting on likelihood output 58
B.1.1 Likelihood output less than 0.2 58
B.1.2 Likelihood output more than 0.8 66

C Figures for $m_{\mathbf{t}} = 175$ GeV/c^2 **74**
C.1 Control plots after cutting on likelihood output 82
C.1.1 Likelihood output less than 0.2 82
C.1.2 Likelihood output more than 0.8 90

Chapter 1

Introduction

At the Tevatron, with $\mathrm{p\overline{p}}$ collisions at $\sqrt{s} = 1.96$ TeV, the dominant $t\bar{t}$ quark production mechanism is $q\bar{q}$ annihilation, followed by gluon fusion. At leading order in α_s the contributing processes are shown in figure 1.1. Flavour excitation and gluon splitting processes become important at next-to-leading order. (figure 1.2). Since the quark-gluon initiated processes only contribute starting at next-to-leading order, their contribution is small compared to the other two cases. At next-to-leading order the $\mathrm{q\overline{q}}$ and gg initiated processes contribute approximately 85% and 15%, respectively. The absolute $\mathrm{q\overline{q}}$ contribution is well known, but due to the relatively large uncertainties on the gluon PDFs the gg contribution can vary from 11%–21% [1]. Assuming a top quark mass of $m_\mathrm{t} = 173$ GeV/c^2, the standard model predicts a next-to-next-to-leading order[1] $\mathrm{t\overline{t}}$ production cross section of 7.37 ± 0.56 pb [2]. This prediction is based on the CTEQ6.5M PDF set. The uncertainties contain contributions from PDF uncertainties and renormalisation/factorisation scale variations.

1.1 Top quark pair decay modes

In the standard model top quarks decay almost exclusively into a bottom quark and a W vector boson. Cataloging top-antitop pair decays based on the decay products of the Ws, the following channels are distinguished:

Both Ws decay into a $\mathrm{q\overline{q}}$ pair: all-hadronic (or all-jets)
While this channel has by far the largest branching ratio of all, $\approx 46\%$, its lack of distinguishing features (e.g. isolated leptons) makes it hard to distinguish from QCD multi-jet production. More importantly the QCD multi-jet cross section is several orders of magnitude larger than that for $\mathrm{t\overline{t}}$ production. On the other hand the absence of leptons, particularly in combination with the dominance of QCD multi-jet production, renders all other typical backgrounds, like W+jets, negligible.

Both Ws decay into e/μ and a neutrino: di-lepton
The presence of two isolated leptons both aids in triggering and in offline event selection, compensating somewhat for the small branching fraction of $\approx 6\%$. However, each lepton is associated by a neutrino, leaving the events kinematically underconstrained. This complicates mass analyses in the di-lepton channel.

One decays into e/μ, one into a $\mathrm{q\overline{q}}$ pair: lepton+jets
With a branching fraction of $\approx 34\%$ and a clear, isolated lepton to trigger and select on this can be considered the 'golden' $\mathrm{t\overline{t}}$ decay channel. Two of the three momentum components of the missing neutrino momentum can be determined using kinematic constraints, leaving only an ambiguity on the sign of the z-component of the neutrino momentum.

[1] More precisely: 'an approximate NNLO cross section which is exact to logarithmic accuracy' [2].

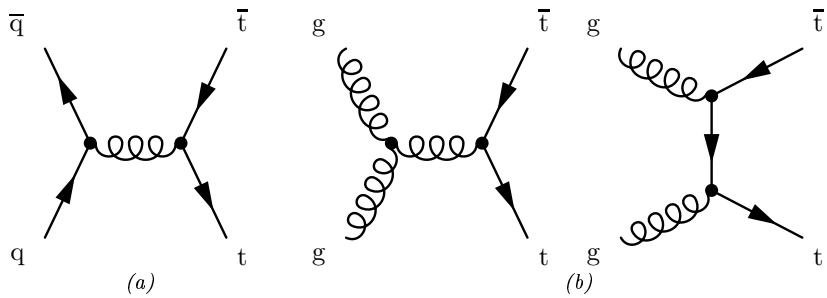


Figure 1.1: Leading order ($\mathrm{O}(\alpha_s^2)$) top-antitop pair production at the Tevatron: (a) quark-antiquark annihilation and (b) gluon fusion. Charge conjugated and 'swapped' diagrams are implicitly included.

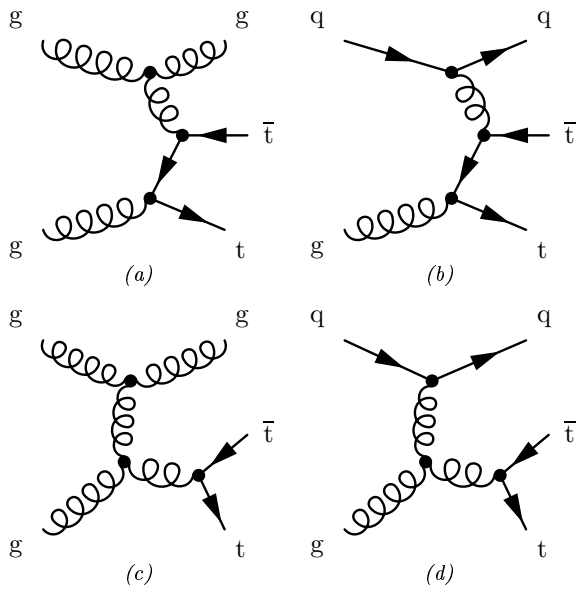


Figure 1.2: Next-to-leading order ($\mathrm{O}(\alpha_s^3)$) processes contributing to $t\bar{t}$ production at the Tevatron. (a), (b) Flavour excitation diagrams, (c), (d) gluon splitting diagrams. The quark-gluon initiated processes ((b), (d)) only contribute at next-to-leading order and do not form a significant contribution to the overall $t\bar{t}$ production cross section. Charge conjugated and 'swapped' diagrams are implicitly included.

One or both decay into a τ lepton

The decay channels involving τ leptons suffer from several different complications. First of all more than a third ($\approx 36\%$) of τs will decay into electrons or muons and be lost to this channel. Secondly, hadronically decaying τs produce jets which are hard to distinguish from parton jets and require a dedicated energy calibration.

By far the largest branching fraction contributes to the all-jets channel. The experimental signature of those events (see also figure 1.3) is:

- At least six jets.
- Two of these jets are, in general, high p_{T} b-jets from the top quark decays.
- The four other quarks should in principle form two pairs with the W boson mass as invariant mass.
- Absence of any isolated leptons or missing transverse energy ($\not{E}_{\mathrm{T}}$).

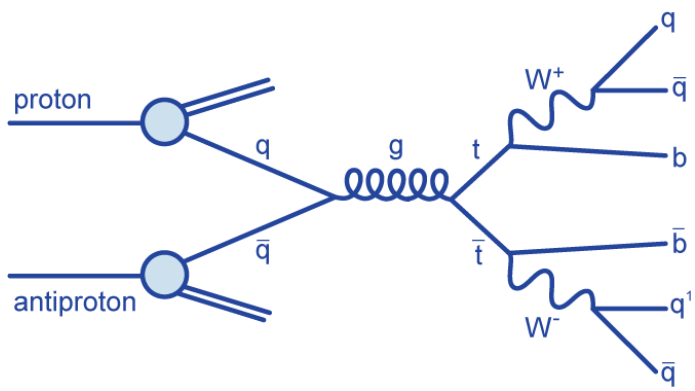


Figure 1.3: All-jets $t\bar{t}$ decay event signature: two high p_T b-jets and at least four other jets in the absence of any isolated leptons or missing transverse energy.

The $V - A$ nature of the $t \to Wb$ coupling in the standard model leads to a predominantly longitudinal polarization of the W bosons: $\approx 70\%$ [3, 4, 5] for values of the top quark mass in the 170–175 GeV/c^2 range [6]. Any polarization of the W boson directly affects the angular distributions of its decay products. Whereas longitudinal polarization results in a distribution proportional to $\sin^2\theta^*$, where θ^* is the angle between the W boson's momentum in the top quark rest frame and the down-type quark momentum in the W boson rest frame, transverse polarization results in a $1 + \cos^2\theta^*$ behavior. In the former case both decay products favor the direction perpendicular to that of the W while in the latter case the thrust axis tends to be aligned with the W direction. The relative strength of the longitudinal contribution leads to a more even distribution of energy between the two jet daughters of W bosons of hadronic top quark decays. This can be used as an additional part of the experimental signature.

The absence of isolated leptons is a disadvantage from the trigger point of view. Although the high jet multiplicity makes it relatively easy to design high efficiency multi-jet triggers for this channel, it provides another challenge: avoiding too high trigger rates on QCD multi-jet events.

The multi-jet $t\bar{t}$ signal is extremely small compared to general QCD multi-jet production. The cross section for hard di-jet production (which can increase in jet multiplicity from radiative jets) is more than six orders of magnitude larger. For $b\bar{b}$ di-jet production the cross section is already a thousand times smaller, while for six-jet events the QCD cross section is only a hundred times larger than the expected $t\bar{t}$ cross section. Moreover, the cross section for QCD events with both a $b\bar{b}$ pair and four additional jets is only approximately six times larger than the expected signal. This suggests a strong preselection of data events on the presence of at least six jets, at least two of which are tagged as b-jets. Since approximately half of the W decays will produce c-quarks, which tend to be tagged as b-quarks, one should allow for the presence of more than the expected two b-jets.

1.2 Measurement of the top quark pair production cross section

A cross section being a production rate, the basic ingredients are the number of signal events in the data sample N_{events}, the total data taking efficiency ϵ and the total luminosity $\mathcal{L}$ to which the data sample corresponds:

$$\sigma = \frac{N_{\text{events}}}{\mathcal{L}\epsilon}.$$

The overall flow of the analysis, shown in figure 1.4, is as follows. First calibration and (MC) event weights are applied to the data and Monte Carlo samples. A series of preselection cuts is applied to each sample to improve the signal-to-background ratio. A specially preselected data sample is used to model the background

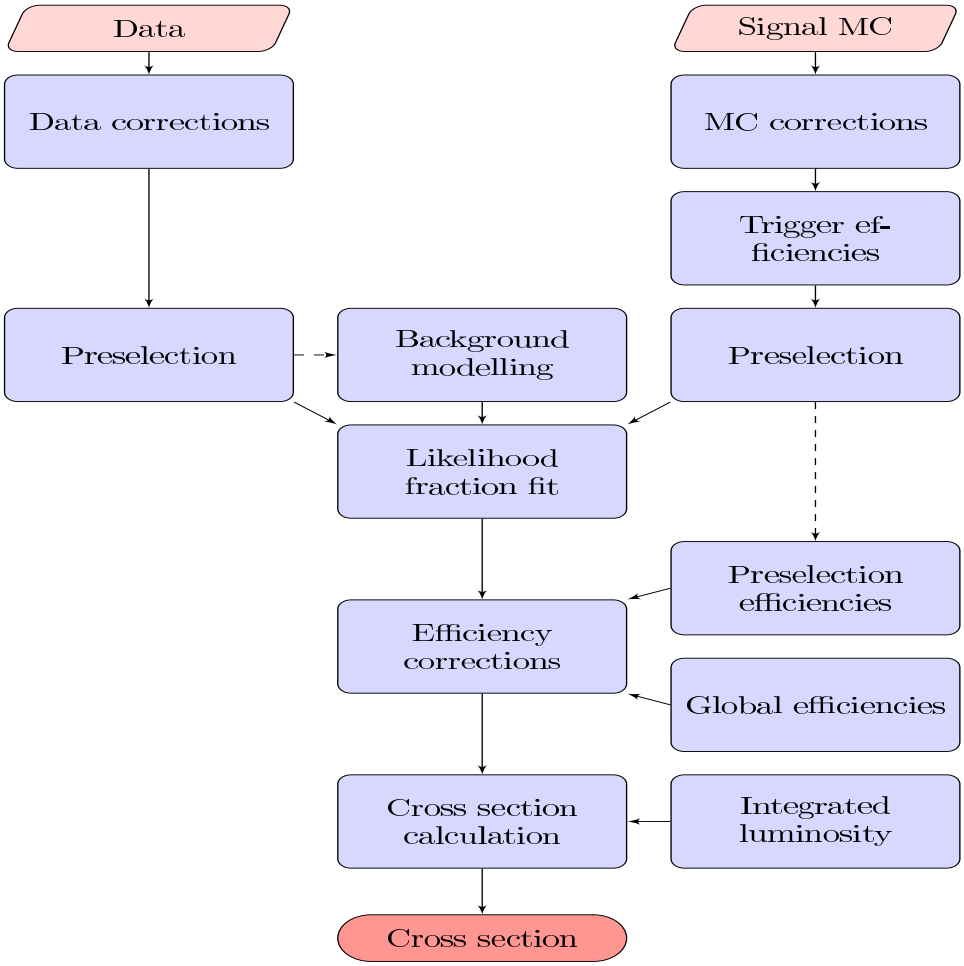


Figure 1.4: Overview of the analysis work flow.

and a likelihood analysis is used to separate signal from background and estimate the number of signal events present in the data sample. After correcting for the trigger and preselection efficiencies this gives the estimated number of signal events produced. Combining this with the total, integrated luminosity to which the triggered data sample corresponds allows to calculate the production cross section.

Chapter 2

Event Selection

This section describes the selection of the data samples used in the analysis as well as the choice of Monte Carlo samples and corrections applied. These criteria are applied to the data prior to using multivariate techniques to separate signal from background and are termed 'preselection'. Preselection cuts and corresponding efficiencies are discussed in Section 2.9. The preselection removes events with low data quality (mostly due to hardware problems) and selects events based on the hadronic $t\bar{t}$ signature. After preselection the sample is still dominated by background.

2.1 Samples

The data sample is the Run IIa data set, spanning the trigger list versions v8 through v14 and corresponding to approximately 1 fb^{-1} of recorded luminosity. Since there are no QCD models available that describe the data for six-jet events, instead of relying on Monte Carlo, the data-based background model described in Chapter 3 was used in this analysis.

The Monte Carlo $t\bar{t}$ samples were generated using the ALPGEN generator v2.05 [7] for the hard interaction, followed by PYTHIA v6.319A[8] for fragmentation, shower development and hadronisation. All MC samples were processed through the default GEANT based D0 detector simulation and reconstructed using the default D0 reconstruction software using release p17.09.08. $t\bar{t}$ samples are available for several assumed values of the top quark mass. All ALPGEN samples were generated for the all-jets, lepton+jets and di-lepton channels separately, each in turn split into subsamples for 0, 1 and 2 additional light partons. Table 2.1 shows the request ids for all used ALPGEN samples, sorted by channel and light parton multiplicity, together with the matching cross sections.

The Monte Carlo samples are corrected for several imperfections of the event generation and/or detector simulation:

- The central value of the Monte Carlo samples was reweighted to correspond the CTEQ6.5M [10]. (All MC samples were generated using CTEQ6L1 [11].) The LHAPDF package [12] v5.3.1 was used to interface to the CTEQ PDF sets.
- The default top group b-fragmentation reweighting [13] was applied.
- The primary vertex z-position (z_{PV}) distribution was reweighted to match the one in data between $|z_{\mathrm{PV}}| < 35$ cm (the rest is cut out in the preselection). See below for more details.
- The instantaneous luminosity profile in Monte Carlo was reweighted to match that of the selected data sample. See below for more details.

Table 2.1: Request id numbers and matching cross sections for all relevant ALPGEN *subsamples. The 'two light partons' bin in the p17.09.06 samples is affected by the 'inclusive bin bug' [9]. The matching cross sections shown (marked by †) are corrected by* 1.52, 1.43, *and* 1.54, *respectively, for this problem.*

Subsample	Request ids	Cross section
tt_170_Alpgen_alljet+0lp	70115, 70116	1.506
tt_170_Alpgen_alljet+1lp	70117	0.635
tt_170_Alpgen_alljet+2lp	70132	0.314†
tt_175_Alpgen_alljet+0lp	84772, 84773, 84774, 84775	1.297
tt_175_Alpgen_alljet+1lp	84792, 84793, 84794	0.544
tt_175_Alpgen_alljet+2lp	84812, 84813	0.267†

(a) ALPGEN all-jets samples

Subsample	Request ids	Cross section
tt_170_Alpgen_leptonjet+0lp	38689, 69333, 69335, 72854, 72855, 72856, 82637, 82638	1.505
tt_170_Alpgen_leptonjet+1lp	38691, 69337, 72857, 72859, 82639, 82640	0.635
tt_170_Alpgen_leptonjet+2lp	38692, 69339, 72860, 82641	0.309†
tt_175_Alpgen_leptonjet+0lp	29249, 29250, 29266, 29267, 29268, 37340, 37341	1.299
tt_175_Alpgen_leptonjet+1lp	29251, 29269	0.541
tt_175_Alpgen_leptonjet+2lp	29888, 29890	0.259†

(b) ALPGEN lepton+jets samples

Subsample	Request ids	Cross section
tt_170_Alpgen_dilepton+0lp	69332, 69334, 85252, 85253	0.376
tt_170_Alpgen_dilepton+1lp	69336, 85254	0.159
tt_170_Alpgen_dilepton+2lp	69338, 85272	0.084†
tt_175_Alpgen_dilepton+0lp	29193, 29243, 29244, 29262, 29263	0.324
tt_175_Alpgen_dilepton+1lp	29245	0.135
tt_175_Alpgen_dilepton+2lp	29885, 29886, 29889, 37322	0.070†

(c) ALPGEN di-lepton samples

2.2 Data quality requirements

Not all recorded data is suited for analysis. Occasionally problems arise with subdetectors or with the overall detector- and/or trigger configuration. The data quality selection removes all data not suited for a given analysis. Data can be rejected per run, per luminosity block, or per event. Runs are the 'control room units' of data taking: periods generally ranging from two to four hours during which the detector and trigger configuration is kept unchanged. Bad runs reflect the presence of longer periods of impaired data taking. Each run is graded from the point of view of each of the subdetector, allowing analyzers to select only those runs with good grades for all subdetectors of interest. This system is augmented by offline expert studies

Table 2.2: The trigger choice and luminosity for the different run ranges. The luminosity shown is the integrated recorded luminosity after data quality selection.

Run range	Trigger	$\mathcal{L}$ (pb^{-1}).
v8	4JT10	24.54
v9	4JT10	24.07
v10	4JT10	10.22
v11	4JT10	64.68
v12	4JT12	210.23
v13a	JT2_4JT12L_HT	49.34
v13b	JT2_4JT12L_HT	257.39
v14a	JT2_4JT20	189.83
v14b	JT2_4JT20	143.55
Total $\mathcal{L}$ (pb^{-1})		973.84

evaluating the shifter grades. Bad luminosity blocks are often due to problems with the trigger or luminosity systems or to short periods of deteriorated calorimeter performance. To reject events affected by transient calorimeter issues like coherent noise all events are rejected in which any of the calorimeter quality flags [14] were raised. For this analysis the standard data quality selection [15, 16] was applied. This ensures proper operation of all subdetectors, luminosity normalizability and only selects runs taken with official trigger lists.

Since luminosity information is not available on an event-by-event basis, the event based data quality is treated as a preselection inefficiency. The fractional data loss due to events removed based on the calorimeter event quality flags has been estimated in zero-bias data. The luminosity-weighted average loss was estimated to be $2.702 \pm 0.001\%$ [17].

2.3 Integrated luminosity

The integrated recorded luminosity for good quality data for all run ranges is shown in Table 2.2. A detailed explanation and justification of the luminosity choices can be found in ref. [18].

To simulate the effects of noise and pile-up in the Monte Carlo, all MC samples are overlaid with zero-bias data. The instantaneous luminosity distribution in this zero-bias data is not representative for the whole Run IIa data set and is reweighted to achieve better agreement between the instantaneous luminosity distributions in data and MC. Figure 2.1 shows the luminosity profiles (recorded luminosity as a function of instantaneous luminosity) for all individual triggers used. For details on the triggers please refer to Chapter 2 and Appendix A. The luminosity profiles were combined together according to their respective integrated luminosities (Table 2.2).

2.4 Trigger requirements

The trigger used in the data selection required at least four jets (4JT). The exact trigger definition changed from time to time depending on instantaneous luminosity and developments in the underlying algorithms. The run ranges associated with each trigger definition are labeled based on te version of the trigger list as presented in Table 2.2. Please refer to Appendix A for a detailed listing of all trigger terms making up each trigger. This appendix also contains the full definitions of all run ranges in terms of run numbers.

Single jet turn-on curves for the three levels of the triggers were taken from the standard DØ SPC files detailed in DØ note 5093.

To create data sets of workable size the full data set is 'skimmed' into subsets using a rough preselection based on both the triggers that fired and the reconstructed objects in each event. The input to this analysis

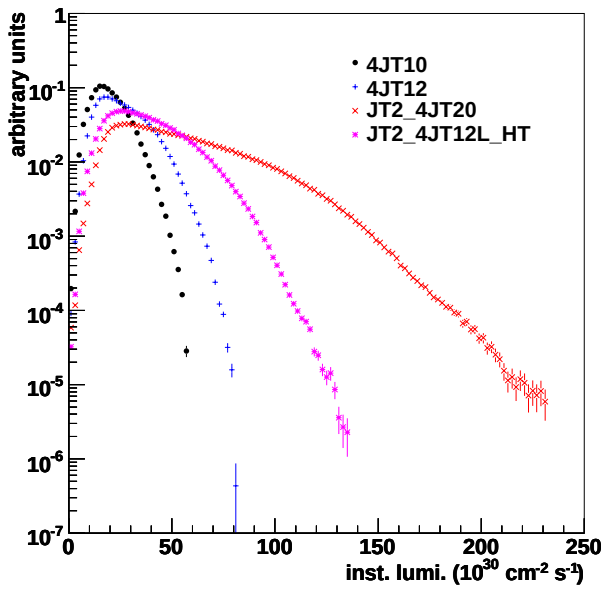


Figure 2.1: Luminosity profiles for the individual four-jet triggers. All profiles are scaled to unit area.

is the so called '3JET' skim. This skim requires at least three $\mathcal{R}_{\text{cone}} = 0.5$ jets within $|\eta_{\text{d}}| < 2.6$,[1] one with (JES uncorrected) $p_{\text{T}} > 20$ GeV/c and two others with $p_{\text{T}} > 15$ GeV/c. The skimming efficiency is treated as a fourth trigger level. Selecting all events from the MUinclusive skim[2] that fire at least one of the muon triggers (MUW_*, MUH*_*, etc.) the single-jet '$p_{\text{T}} > 15$ GeV/c' and '$p_{\text{T}} > 20$ GeV/c' turn-on curves are estimated, parameterized as a function of JES corrected jet p_{T}, and binned in bins of $\Delta\eta_{\text{d}} = 0.1$. For the Level 4 turn-ons no distinction is made between the different run ranges. Figure 2.2 shows the two skimming turn-on curves for both p_{T} thresholds. The skimming turn-ons have sharp flanks around $p_{\text{T}} \approx 25$ GeV/c (33 GeV/c) for the '$p_{\text{T}} > 15$ GeV' ('$p_{\text{T}} > 20$ GeV/c') curves.

2.4.1 Jet triggers: combinatorics

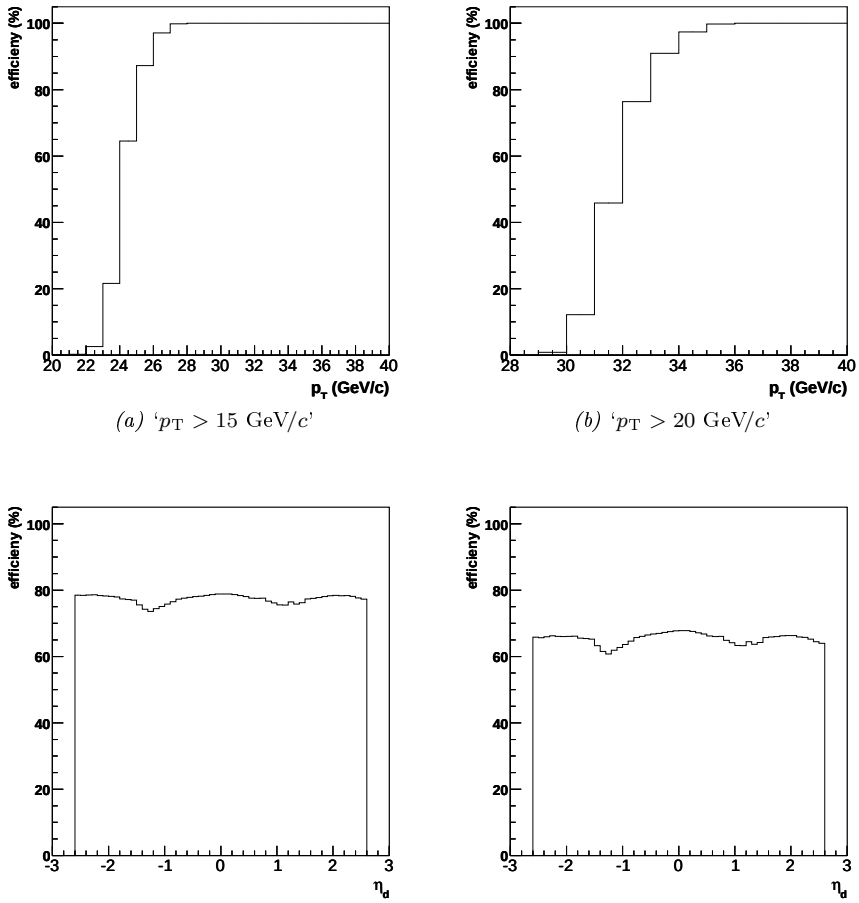
In combining the single object trigger turn-ons some care has to be taken to correctly handle the correlations between different jet multiplicity terms with different thresholds. For example a jet satisfying a 25 GeV/c threshold will automatically also satisfy all terms with lower thresholds. We follow the prescription detailed in ref. [19].

Using a simplified notation JET(m, x) for a generic multi-jet trigger term requiring m jets with $p_{\text{T}} > x$ GeV/c any multi-jet trigger with M terms can be written as

$$\prod_{m=1}^{M} \text{JET}(m, p_{\text{T}}(m)),$$
$$p_{\text{T}}(m+1) \geq p_{\text{T}}(m).$$

For example, a trigger requiring four jets with $p_T > 30\,\text{GeV}/c$ would be written as JET$(4, 30)$. The trigger

[1] η_{d} is detector η where the η is defined with respect to (0,0,0) instead of the nominal interaction.
[2] The MUinclusive skim only requires a single loose muon with a central $p_{\text{T}} > 8$ GeV/c.



(a) '$p_{\mathrm{T}} > 15$ GeV/c' *(b)* '$p_{\mathrm{T}} > 20$ GeV/c'

(c) '$p_{\mathrm{T}} > 15$ GeV/c' *(d)* '$p_{\mathrm{T}} > 20$ GeV/c'

Figure 2.2: Skimming turn-on curves as a function of offline, JES corrected, jet p_T ((a), (b)) and vs. jet position η_d ((c), (d))

probability for a multi-jet event can be written as one minus all possible ways not to trigger on the event:

$$P(\text{event}) = 1 - \sum_{j=0}^{M-1} P_j$$

where P_j represents the probability that j jets in the event satisfy the trigger. The trigger probability P_j for an event with N offline jets to fire a term requiring j jets can be written as the product of j jets satisfying the trigger and $N - j$ jets failing, properly summed over all permutations of j out of N jets:

$$P_j = \textsc{jet}(j, p_{\mathrm{T}}(j)) = \sum_{\text{permutations}} \left(\prod_{k=1}^{j} P(k; p_{\mathrm{T}}(j)) \prod_{l=j+1}^{N} 1 - P(l; p_{\mathrm{T}}(j)) \right),$$

with $P(k; p_{\mathrm{T}}(j))$ the probability that jet k satisfies the threshold $p_{\mathrm{T}}(j)$. P_0, the term requiring all jets in

the event to fail the trigger, can be written as

$$\begin{aligned} P_0 &= 1 - P(\text{at least one jet satisfies}) \\ &= \prod_{i=1}^{N} \left(1 - P(i; p_{\mathrm{T}}(1))\right). \end{aligned}$$

Here $P(i; p_{\mathrm{T}}(1))$ represents the probability for jet i to satisfy the trigger term with the highest threshold. If any one of the jets would satisfy this threshold, it would also satisfy all other terms and thus satisfy the trigger. A general term P_i $(i > 0)$ contains i terms:

$$P_i = \sum_{j=1}^{i} P_{i,j}$$

with

$$\begin{aligned} P_{i,j} &= P_i^{i-j+1, p_{\mathrm{T}}(1); i-j+2, p_{\mathrm{T}}(2); \ldots; i-j+j, p_{\mathrm{T}}(j)} \\ &= \frac{1}{(i-j+1)!} \sum_{j\text{-jet permutations } j_1, \ldots j_j} P(j_1; p_{\mathrm{T}}(1)) \\ &\quad \times \left(P(j_2; p_{\mathrm{T}}(2)) - P(j_1; p_{\mathrm{T}}(1))\right) \\ &\quad \times \ldots \\ &\quad \times \left(P(j_j; p_{\mathrm{T}}(j)) - P(j_{j-1}; p_{\mathrm{T}}(j-1))\right) \\ &\quad \times \prod_{k=1, k \neq j_1, \ldots, j_j}^{N} \left(1 - P(k; p_{\mathrm{T}}(j))\right). \end{aligned}$$

Here $P(i; p_{\mathrm{T}}(j))$ is again the probability for jet i to pass threshold $p_{\mathrm{T}}(j)$. The sum runs over all permutations of j jets out of the N jets in the event. The terms containing the $j_1 \ldots j_j$ outside the product represent the probability of the j jets satisfying the trigger. To prevent double-counting between the different thresholds each of these terms consists of the probability difference of the jet satisfying thresholds i and $i+1$. The product represents the remaining $N - j$ jets failing the trigger.

2.4.2 Trigger Weight

The trigger efficiency determination is based on single-object trigger turn-on curves. For each trigger level the single-object curves are combined into a single turn-on curve. The different trigger level turn-ons are multiplied to obtain the overall trigger efficiency. The overall trigger efficiencies are folded into the Monte Carlo samples as event weights. Effects of trigger changes over the different run ranges are taken into account by using the luminosity averaged average trigger efficiency:

$$\epsilon = \frac{\Sigma_i \mathcal{L}_i \epsilon_i}{\Sigma_i \mathcal{L}_i}$$

where the sum runs over all run ranges i, ϵ_i is the trigger efficiency for run range i and the $\mathcal{L}_i$ are integrated luminosity values from the third column of Table 2.2. The weight is the inverse of the efficiency. The luminosity weighted average trigger efficiency as a function of $H_{\mathrm{T}} = \sum_i p_{T\,i}$ is shown in Figure 2.3.

2.5 Jets

Standard DØ jet identification criteria were used to select good jets. Additionally, all jets had to have $|\eta_d| < 2.5$ where η_d is the jet η with respect to the center of the detector rather than the primary interaction vertex.

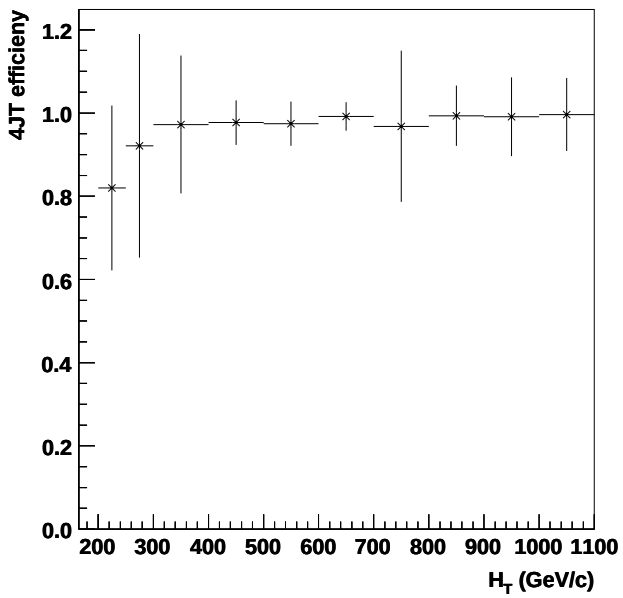


Figure 2.3: Luminosity weighted average trigger efficiency vs. H_T.

To calibrate all jet energies the final p17 jet energy scale (as implemented in `jetcorr` version `v07-02-67`), including muon corrections, was applied. The muon corrections correct for the jet energy 'missed' in the calorimeter and carried away by muons inside a jet. High p_{T} muons, having nearly straight tracks, are often reconstructed with arbitrarily high transverse momenta. To prevent these mismeasured muons from deteriorating the jet resolution through the muon corrections the size of the JES muon corrections is 'capped' at 60 GeV [20].

The final p17 Monte Carlo jet energy scale is applied to the Monte Carlo, including muon correction capping as for data. In addition jet 'smearing, shifting and removal' [21] (SSR) as implemented in `caf_mc_utils` version `p18-br-81` is applied. Following the common procedure, only smearing is applied by default, smearing plus shifting is applied to provide a systematic uncertainty (see Section 5.2.2).

2.6 Multi-Jet Selection

An event sample called '6+15' was created by selecting events with at least six jets with $p_{\mathrm{T}} > 15$ GeV/c. A tighter sub-sample was used for the cross section measurement. This sample, referred to as '4+40', required that at least four of the selected jets satisfy $p_{\mathrm{T}} > 40$ GeV/c. The loose '6+15' sample was used whenever larger statistics were necessary for specific studies.

2.7 Primary Vertex

The primary interaction vertex, reconstructed with at least three associated tracks, was required to be within 35 cm of the center of the detector along the Z axis. The track multiplicity cut is part of the standard D0 primary vertex selection. Since high jet multiplicity events also tend to have high track multiplicity this cut is *very* efficient for the sample at hand. The primary vertex acceptance corrects for the fraction of events missed due to the vertex position cut. Figure 2.4 shows the $|z_{\mathrm{PV}}|$ distribution for the '4+40' data sample.

The distribution is fitted within the $|z_{\mathrm{PV}}| < 35$ cm range with the sum of two Gaussians. The range is limited to within the SMT fiducial range $|z_{\mathrm{PV}}| \lesssim 38$ cm since for vertices outside that range slightly looser track requirements are used. The fit extrapolation outside the fit range is also shown. The agreement between fit and data points in the fit range is excellent. Integrating the fit result out to $|z_{\mathrm{PV}}| = 1000$ cm shows a primary vertex acceptance of 79.5% with an uncertainty of 2.0% based on the effects of variations of the jet p_{T} cuts. The (tiny) fit uncertainty is negligible compared to those variations.

The vertex acceptance fits to the z_{PV} distributions for the '4+40' and '6+15' preselection samples yield statistically consistent results. Since the primary vertex distribution does not depend on the jet selection (it is determined by the incoming beams) the '6+15' sample was used to increase statistics in the MC to data comparison. Figure 2.5 shows the z_{PV} distributions of the '6+15' sample in data and MC. The distribution in Monte Carlo is slightly shifted and marginally narrower than the one in data. The right figure in Figure 2.5 shows the ratio of data/MC for the z_{PV} distributions. The relative shift of the two distributions results in a 'double bump' structure. This can be fitted using a fourth order polynomial (black line in Figure 2.5) but instead it was chosen to use the ratio of the data- and MC fit results to the z_{PV} distributions for the MC reweighting (red line in Figure 2.5). Both approaches show good agreement. The distribution of the primary vertex z-position z_{PV} in Monte Carlo was reweighted to correspond to that in data within the selected $|z_{\mathrm{PV}}|$ range.

2.8 b-Jet identification

The neural network tagger [22, 23] is used to identify b-jets. The NN > 0.65 operating point was chosen to suppress the fake rate to manageable levels. The efficiencies for tagging b-jets and c-jets as well as the fake rate for the neural network tagger are described in ref. [23] and implemented in the `btags_cert` package. The required tag rate functions (TRFs) were read from the file `NN_p17_trf_cert_fixed_v0.5.root`. The b-tagging in Monte Carlo was implemented as an actual tag based on a random number. For each taggable jet a random number is thrown and the corresponding TRF is evaluated. If the TRF value exceeds the random number the jet is tagged.

Jet taggability requires the spatial matching of a calorimeter jet with a track jet. Imperfections in the simulation of tracking and/or shower development can lead to taggability differences between data and Monte Carlo. This taggability difference was studied using samples with at least six jets within $|\eta_{\mathrm{d}}| < 2.5$, three with $p_{\mathrm{T}} > 40$ GeV/c the others with $p_{\mathrm{T}} > 15$ GeV/c. Since the flavour content of this data sample differs significantly from the (signal) Monte Carlo, b-jet selection criteria were applied to connaturalize both samples. Since this biases the sample towards higher taggability, a cross check was performed to study the effects. At the p_{T}-plateau above ≈ 40 GeV/c, where no strong flavour differences are expected, the effect of applying b-id is the same in data and Monte Carlo. Figure 2.6 shows a comparison between data and Monte Carlo taggabilities based on the b-jet selection used in the analysis: $p_{\mathrm{T_b}} > 40$ GeV/c and NN > 0.65. Figure 2.7 shows the corresponding data-to-MC taggability ratio.

On average, the MC properly reproduces the taggability of the data, so by default no scale factors were applied to the MC events. The overall differences between data and MC are of the order of 2%. Loosening the b-jet criteria, lowering the p_{T} cut from 40 to 15 GeV/c or the neural network tagger cut from 0.65 to 0.2, results in slightly lower relative taggabilities, well within the quoted systematic uncertainty. This MC-to-data differences was treated by uniformly decreasing taggability in MC by 2% (i.e. randomly making two in a hundred taggable jets not taggable) and using that as a systematic 'down' variation. The difference was symmetrised to represent the corresponding 'up' variation (see also Section 5.2).

2.9 Preselection Efficiency

The typical event signature of at least six jets, two high p_{T} b-jets in the absence of any isolated leptons is used to increase the signal-to-background ratio by applying a series of preselection cuts.

Events are preselected as follows:

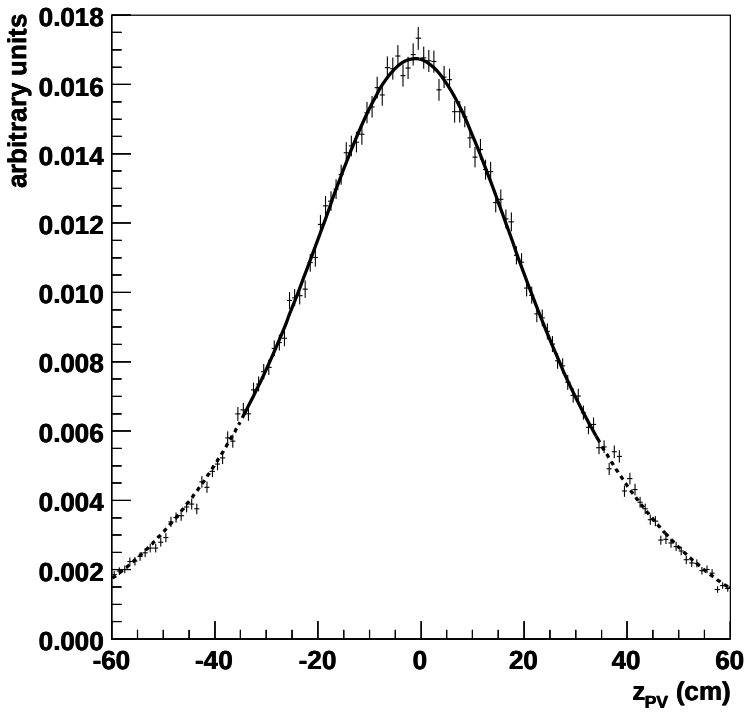
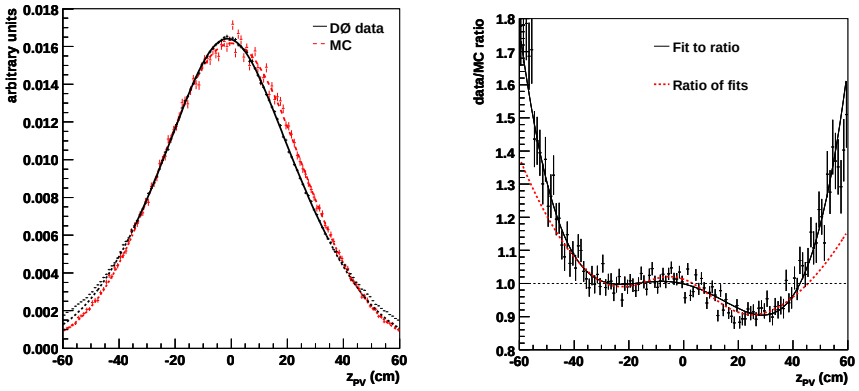
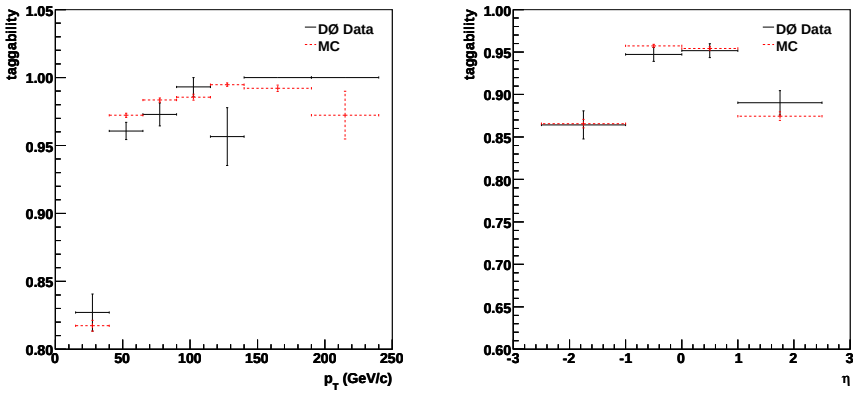


Figure 2.4: The primary vertex z-position distribution (markers) in data (scaled to unit area) fitted with the sum of two Gaussians (solid line). The dotted line shows the fit extrapolation outside the ±35 cm fit range.



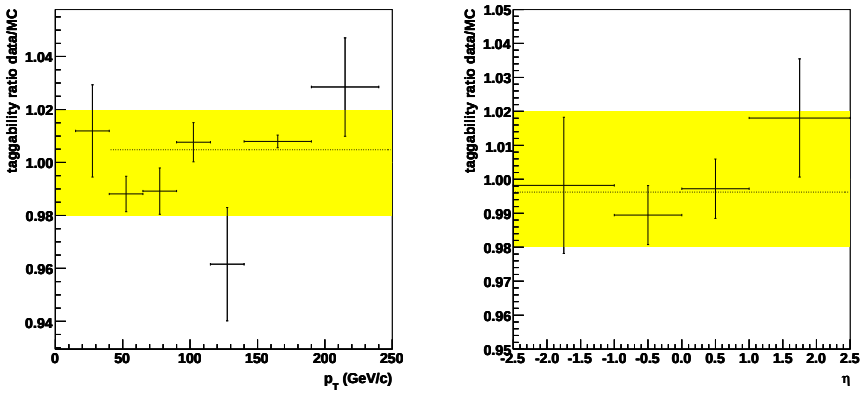
(a) *(b)*

Figure 2.5: Primary vertex z-position distributions for the '6+15' samples in data (black markers) and MC (red markers), both scaled to unit area. Also shown are the fits to these distributions (solid black and dotted red lines respectively) and the fit prediction outside the $|z_{PV}| < 35$ cm fit range for data (black dotted line). The right plot shows the ratio of data/MC distributions (markers), a fourth order polynomial fit to the ratio (black dotted line) and the ratio of the two fit results in the top figure (red dotted line). The latter is used for the z_{PV} reweighting applied to the Monte Carlo.



(a) vs. jet p_{T} *(b)* vs. jet η

Figure 2.6: Taggability comparison between data and Monte Carlo, as a function of jet (a) p_T and (b) η.



(a) vs. jet p_{T} *(b)* vs. jet η

Figure 2.7: Relative Monte Carlo-to-data taggability scale factor as a function of jet (a) p_T and (b) η. The shaded bands indicate the (1.00 ± 0.02) scale factor.

- Events should contain at least one good (i.e. at least three associated tracks) primary vertex within $|z_{\mathrm{PV}}| < 35$ cm.
- The trigger requirement has to be fulfilled.
- No isolated electrons/positrons or muons are allowed. This cut assures orthogonality with the lepton+jets cross section analyses.
- At least six taggable jets are required with $p_{\mathrm{T}} > 15$ GeV/c and within $|\eta_{\mathrm{d}}| < 2.5$. The η_{d} cuts mainly improve the signal fraction since QCD multi-jet production tends to be more forward than the $\mathrm{t\bar{t}}$ signal. The taggability requirement ensures all jets point back to the primary vertex. It should be noted that no explicit veto is applied on the presence of jets from additional interactions in the same bunch crossing. All analysis steps consider only taggable jets assumed to belong to the reconstructed hard scatter. An explicit veto on additional vertices and/or jets would hardly be more effective but would remove a significant fraction of the available statistics.
- At least four of the jets have to pass the p_{T} threshold of $p_{\mathrm{T}} > 40$ GeV/c. The requirement of two $p_{\mathrm{T}} > 40$ GeV/c jets is connected with the expectation of two high p_{T} b-jets and also ensures the tagger operates in an efficient range of its parameter space. The requirement of two additional high p_{T} jet drastically improves the signal content while being fairly efficient for the $\mathrm{t\bar{t}}$ signal.
- At least two of those high p_{T} jets have to be b-tagged with $\mathtt{NN} > 0.65$.

Based on the 'four jets with $p_{\mathrm{T}} > 40$ GeV/c' requirement, this sample is called the '4+40' sample. Unless explicitly stated otherwise, all results in this note refer to this '4+40' sample.

The efficiencies for the above preselection cuts are summarized in Table 2.3a (Table 2.3b) assuming a top quark mass of $m_{\mathrm{t}} = 170$ GeV/c^2 ($m_{\mathrm{t}} = 175$ GeV/c^2). The number of events for each subsample (num.) have been adjusted so that the total number of events is equal to the number of candidates selected in the data. A signal to background of 12.5% was assumed for the final preselection in these tables. Overall, the preselection efficiency for $\mathrm{t\bar{t}}$ is $\approx 2\%$. Of the preselected $\mathrm{t\bar{t}}$ events, approximately 7.5% are from leptonic decay channels. Most are from the lepton+jets decay channel ($\approx 60\%$) or τ+jet ($\approx 40\%$). Only a few dilepton events survive the preselection ($\approx 0.05\%$ of all $\mathrm{t\bar{t}}$).

	$t\bar{t}$ all hadronic			$t\bar{t}$ leptonic			$t\bar{t}$ inclusive			background			S:B
	num.	excl.	incl.	num.	excl.	incl.	num.	excl.	incl.	num.	excl.	incl.	
total	3061	1.00	1.00	3814	1.00	1.00	6876	1.00	1.00				
vertex & DQ	2367	0.77	0.77	2953	0.77	0.77	5320	0.77	0.77				
triggered	1659	0.70	0.54	795	0.27	0.21	2454	0.46	0.36				
≥ 4 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	1642	0.99	0.54	760	0.96	0.20	2402	0.98	0.35	18856297	1.00	1.00	1 : 7850.0
≥ 5 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	1454	0.89	0.47	451	0.59	0.12	1905	0.79	0.28	6514404	0.35	0.35	1 : 3420.2
≥ 6 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	890	0.61	0.29	171	0.38	0.045	1062	0.56	0.15	1734613	0.27	0.092	1 : 1633.7
lepton veto	890	1.00	0.29	135	0.79	0.035	1025	0.97	0.15	1474734	0.85	0.078	1 : 1438.9
all jets taggable	612	0.69	0.20	62	0.46	0.016	675	0.66	0.098	506291	0.34	0.027	1 : 750.6
≥ 2 b-tagged jets with $p_T > 15$ GeV	222	0.36	0.072	22	0.35	0.0057	243	0.36	0.035	11795	0.023	0.00063	1 : 48.5
≥ 2 b-tagged jets with $p_T > 40$ GeV	138	0.62	0.045	14	0.64	0.0036	152	0.63	0.022	2573	0.22	0.00014	1 : 16.9
≥ 3 jets with $p_T > 40$ GeV	135	0.98	0.044	13	0.95	0.0034	149	0.98	0.022	2070	0.80	0.00011	1 : 13.9
≥ 4 jets with $p_T > 40$ GeV	112	0.83	0.037	9	0.71	0.0025	121	0.82	0.018	929	0.45	0.000049	1 : 7.7
Efficiency	0.03652 ± 0.00041			0.002464 ± 0.000042			0.01687 ± 0.00019						

(a) $m_{\mathrm{t}} = 170$ GeV/c^2

	$t\bar{t}$ all hadronic			$t\bar{t}$ leptonic			$t\bar{t}$ inclusive			background			S:B
	num.	excl.	incl.	num.	excl.	incl.	num.	excl.	incl.	num.	excl.	incl.	
total	2800	1.00	1.00	3437	1.00	1.00	6238	1.00	1.00				
vertex & DQ	2165	0.77	0.77	2661	0.77	0.77	4826	0.77	0.77				
triggered	1555	0.72	0.56	747	0.28	0.22	2302	0.48	0.37				
≥ 4 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	1540	0.99	0.55	716	0.96	0.21	2256	0.98	0.36	18856444	1.00	1.00	1 : 8359.9
≥ 5 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	1371	0.89	0.49	417	0.58	0.12	1788	0.79	0.29	6514520	0.35	0.35	1 : 3643.2
≥ 6 jets with $p_T > 15$ GeV & $\|y\| < 2.5$	846	0.62	0.30	153	0.37	0.045	999	0.56	0.16	1734675	0.27	0.092	1 : 1736.5
lepton veto	846	1.00	0.30	120	0.79	0.035	966	0.97	0.15	1474793	0.85	0.078	1 : 1526.9
all jets taggable	582	0.69	0.21	56	0.46	0.016	637	0.66	0.10	506328	0.34	0.027	1 : 794.7
≥ 2 b-tagged jets with $p_T > 15$ GeV	213	0.37	0.076	19	0.34	0.0055	231	0.36	0.037	11807	0.023	0.00063	1 : 51.0
≥ 2 b-tagged jets with $p_T > 40$ GeV	139	0.65	0.049	12	0.64	0.0035	151	0.65	0.024	2574	0.22	0.00014	1 : 17.1
≥ 3 jets with $p_T > 40$ GeV	136	0.98	0.049	12	0.95	0.0034	147	0.98	0.024	2071	0.80	0.00011	1 : 14.0
≥ 4 jets with $p_T > 40$ GeV	113	0.83	0.040	8	0.70	0.0024	121	0.82	0.019	929	0.45	0.000049	1 : 7.7
Efficiency	0.04038 ± 0.00022			0.002359 ± 0.000081			0.01859 ± 0.00011						

(b) $m_{\mathrm{t}} = 175$ GeV/c^2

Table 2.3: Preselection efficiencies based on the ALPGEN *Monte Carlo samples for $m_t = 170$ GeV/c^2 (a) and $m_t = 175$ GeV/c^2 (b). The background samples are partly preselected (see Chapter 3) so only the additional cuts are shown. Signal and background numbers have been adjusted, assuming 12.5% signal, to sum to the number of candidate events selected in the data.*

Chapter 3

Background modeling

3.1 Introduction

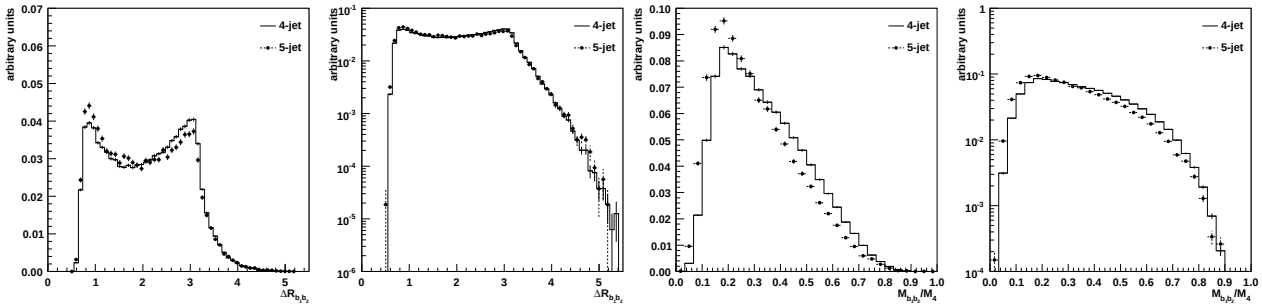
Since there are no QCD models available that describe the data for six-jet events, instead of relying on Monte Carlo, a data-based background model is used. Starting from a guaranteed QCD multi-jet dominated data sample similar to the preselected analysis sample a dedicated background sample is generated. The following points are important to consider when building a data-based background model:

- The starting sample used in the background generation procedure should contain a negligible amount of signal contamination. As the signal signature contains at least six jets, lowering the jet multiplicity should decrease the signal fraction. The cross section for QCD multi-jet production decreases strongly with increasing jet multiplicity. Together with the fact that even in the preselected analysis sample the signal fraction around ten percent, the above guarantees that any sample preselected similarly to the analysis sample but with lower jet multiplicity will contain a negligible fraction of signal events ($\approx 0.5\%$ in five-jet and $\approx 0.15\%$ in four-jet).
- The modeling procedure should not introduce significant biases. It implies special care should be taken to verify that mass-related variables like H_{T} are modeled correctly.
- All variables used in the multivariate analysis to separate signal and background should also be modeled correctly.

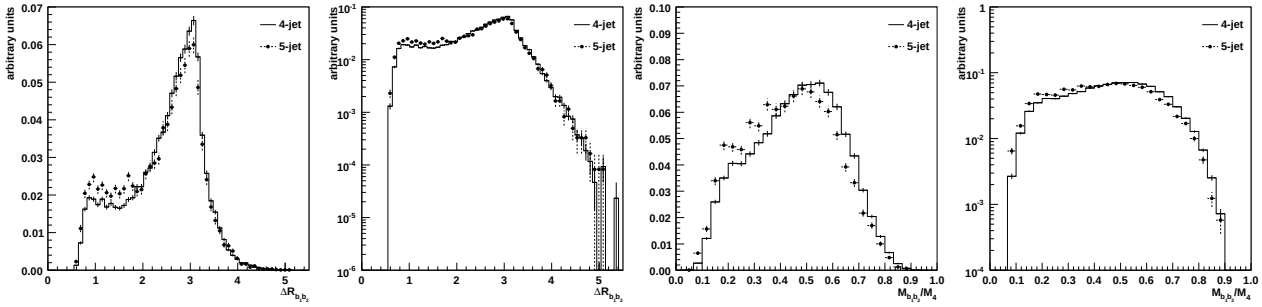
The basic idea is to use four-jet and five-jet versions of the preselected analysis sample to model six-jet background events. This is achieved by matching four-jet (or five-jet) events to events containing six (or more) jets with similar phase space characteristics and donating all 'extra' jets to the event with lower jet multiplicity. This way 'fake' events with high jet multiplicity can be generated which are guaranteed not to contain any significant signal contributions. (In this chapter, unless explicitly stated otherwise, 'six-jet' implies 'at least six jets'.)

The preselected data events are double tagged meaning that at least two jets have been identified using the b-id. The two most common QCD diagrams that produce $\mathrm{b\overline{b}}$ events are direct production where the two b quarks are produced in the hard scatter, and gluon splitting ($g \to \mathrm{b\overline{b}}$). In the first case, the b-jets typically have a large angular separation (peaks at π) and large dijet mass, while the b-jets arising from gluon splitting typically have small angular separation (peaks at twice the jet cone radius, $2\mathcal{R}_{\mathrm{cone}}$ =1.0) and smaller dijet mass. The relative proportion of directly produced $\mathrm{b\overline{b}}$ to gluon-split $\mathrm{b\overline{b}}$ is a strong function of the p_{T} of the b-jets but not should not be a strong function of the overall jet multiplicity (for events with four or more jets).

Figure 3.1 shows the results for the angular difference between the two leading b-jets, $\Delta\mathcal{R}_{\mathrm{b_0b_1}} = \sqrt{(\Delta\eta_{b_0b_1})^2 + (\Delta\phi_{b_0b_1})^2}$, as well as for the invariant mass of these two b-jets relative to the mass of the four leading jets, $M_{\mathrm{b_0b_1}}/M_4$, for two choices of the b-jet p_{T} threshold: $p_{\mathrm{T}} > 15$ GeV/c and $p_{\mathrm{T}} > 40$ GeV/c.



(a) $\Delta\mathcal{R}_{\mathrm{b_0 b_1}}$, $p_{\mathrm{T}} > 15$ GeV/c *(b)* $M_{\mathrm{b_0 b_1}}/M_4$, $p_{\mathrm{T}} > 15$ GeV/c



(c) $\Delta\mathcal{R}_{\mathrm{b_0 b_1}}$, $p_{\mathrm{T}} > 40$ GeV/c *(d)* $M_{\mathrm{b_0 b_1}}/M_4$, $p_{\mathrm{T}} > 40$ GeV/c

Figure 3.1: $\Delta\mathcal{R}_{b_0 b_1}$ *(left) and* $M_{b_0 b_1}/M_4$ *(right) for two choices of b-jet* p_T.

The peak at low $\Delta\mathcal{R}_{\mathrm{b_0 b_1}}$ is representative of $g \to \mathrm{b\bar{b}}$ while the peak near π is representive of direct production. The relative heights of the two peaks is strongly dependent on p_{T}, but there is little difference between the four-jet and five-jet events. This study shows that even though these distributions do depend on the p_{T} criteria used to select b-jets, they depend only weakly on the number of jets in the event. This increases confidence in the use of four-jet and five-jet events to generate six-or-more-jet events for the background sample.

3.2 Background generation procedure

The background generation procedure takes as input two preselected data samples:

The acceptor sample is an additional data sample, preselected exactly as the analysis sample, but containing only five-jet (or four-jet) events. This sample contains a negligible amount of $\mathrm{t\bar{t}}$ events and is the basis for the background sample. There are two different acceptor samples available: one containing four-jet- and one containing five-jet events.

The donor sample is the preselected analysis sample of data candidate events. Jets from events in this sample will be donated to events with lower jet multiplicity in the acceptor sample.

The acceptor and donor samples use preselection criteria very close to those used for the analysis sample (see section 2.9), using similar jet multiplicity (six or more jets), but with slightly looser jet p_{T} requirements. The double p_{T} threshold used in the analysis preselection (15 and 40 GeV/c) strongly shape the phase space occupied by the selected sample. This manifests itself in bumps at 40 GeV/c in the p_{T} spectra of the fourth and softer jets. To facilitate the phase space matching procedure in the background generation all p_{T} cuts were lowered to 15 GeV/c. After generation and shaping the background samples are tightened by increasing the p_{T} thresholds.

In the matching procedure, events from the acceptor sample are matched to randomly selected donor candidates. The goal of the matching is to combine events with similar phase space configurations. The p_{T} of the leading jet is strongly correlated with the momentum transfer Q^2 of the hard interaction. This momentum transfer strongly influences the phase space characteristics of the event. Matching events with similar leading jet p_{T}'s can be used to select events with similar values for Q^2. The phase space of an event is mostly determined by that hard interaction. QCD multi-jet production at the Tevatron is dominated by $2 \to 2$ and $2 \to 3$ processes. An additional matching is required between the third acceptor jet and the fourth donor jet to more strongly select events with similar phase space characteristics. The aim is to select events with similar characteristics between the softest acceptor jet and the (soft) jets to be donated.[1] These jets are referred to as 'match jets'.

The first step of the background generation matches events with similar transverse momenta of the match jets ($\Delta p_{\mathrm{T}} < 1$ GeV/c) and similar leading jet p_{T}'s. First a leading jet p_{T} cut of $\Delta p_{\mathrm{T}} < 1$ GeV/c is tried, subsequently increased in steps of 1 GeV/c up to $(\Delta p_{\mathrm{T}})_{\mathrm{max}} = 20$ GeV/c in case no match is found.[2] If no match is found the acceptor event is rejected. Matches which would generate events with unphysical configurations (i.e. any donor jet has a higher p_{T} than any acceptor jet or any donor jet overlaps within $\Delta\mathcal{R} < 0.5$ with any acceptor jet) are vetoed. For matching donor–acceptor event pairs a background event is generated by inserting the softest jets (in p_{T}) from the donor event into the acceptor event. (So the fifth and sixth jets from six-jet events are added to matched four-jet events, and only the sixth is added to a matched five-jet event.)

Two versions of the background sample are generated: one starting from the four-jet acceptor sample (`4+2`) and one starting from the five-jet acceptor sample (`5+1`). The central background sample used in the analysis consists of equal contributions from each of the `4+2` and `5+1` samples. (Due to the sizes of the acceptor samples the `4+2` sample is significantly larger than the `5+1` sample. For simplicity in estimating systematic uncertainties a $50 - 50$ mix was chosen for the combined background sample.)

The second step in the background generation applies two phase space shaping cuts to improve the agreement of the generated background sample in the validation sample (see section 3.3):

$\not{E}_{\mathbf{T}} > 5$ **GeV**
: The premise of this cut is that, since QCD multi-jet events are not expected to contain significant missing transverse energy, the presence of missing E_{T} implies the presence of a missed jet. In this case the acceptor event really has a five-jet-like phase space.

$\not{E}_{\mathbf{T}}/H_{\mathbf{T4}} < 0.1$
: This cut on the relative size of the missing E_{T} with respect to the p_{T}-sum of the leading four jets H_{T4} removes badly reconstructed events as well as events with leptons or neutrinos.

It should be noted that the distributions of most of the event shape variables are fairly well modeled by the first step of the background model, without shaping cuts. The p_{T} and H_{T} (the p_{T}-sum of all jets) spectra, however, in that case tend to be slightly too soft.

3.3 Background validation

The performance of the background model is studied in a sample with negligible signal contamination: the preselected five-jet sample. To verify the background modeling a five-jet background validation sample (`4+1`) was generated starting from a four-jet acceptor sample and matching to a five-jet donor sample. Comparisons between these samples for many kinematic and topological variables used in **??** are shown in Figure 3.2. The agreement between the generated five-jet background sample and the original five-jet donor

[1] An alternative was tried in matching the fourth acceptor- to the fourth donor jet. The agreement between the generated background and the data was significantly worse than for the third-to-fourth jet scheme. This was understood as the fourth-to-fourth jet scheme too strongly matching events with similar leading four jets, leaving too little room for the addition of two extra jets 'below' the fourth jet.

[2] This is implemented as follows. Randomly select an acceptor event. Select all donor events with similar leading- and match jet p_{T}'s and randomly try each of these. If none is found, increase the leading jet p_{T} cut and repeat.

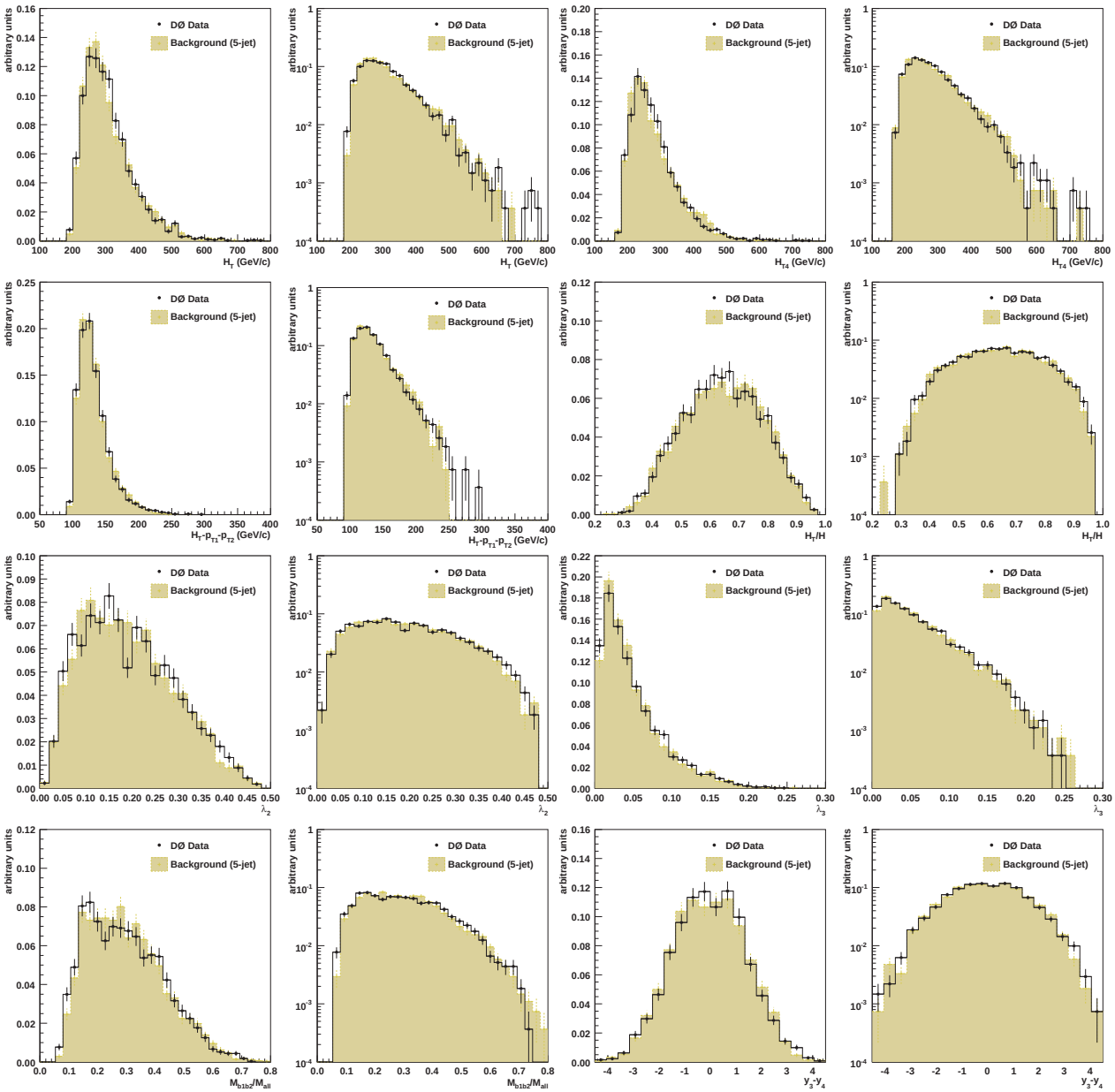


Figure 3.2: Comparison between the 4+1 background validation sample and the original five-jet donor sample. Left-hand side: linear scale, right-hand side: logarithmic scale.

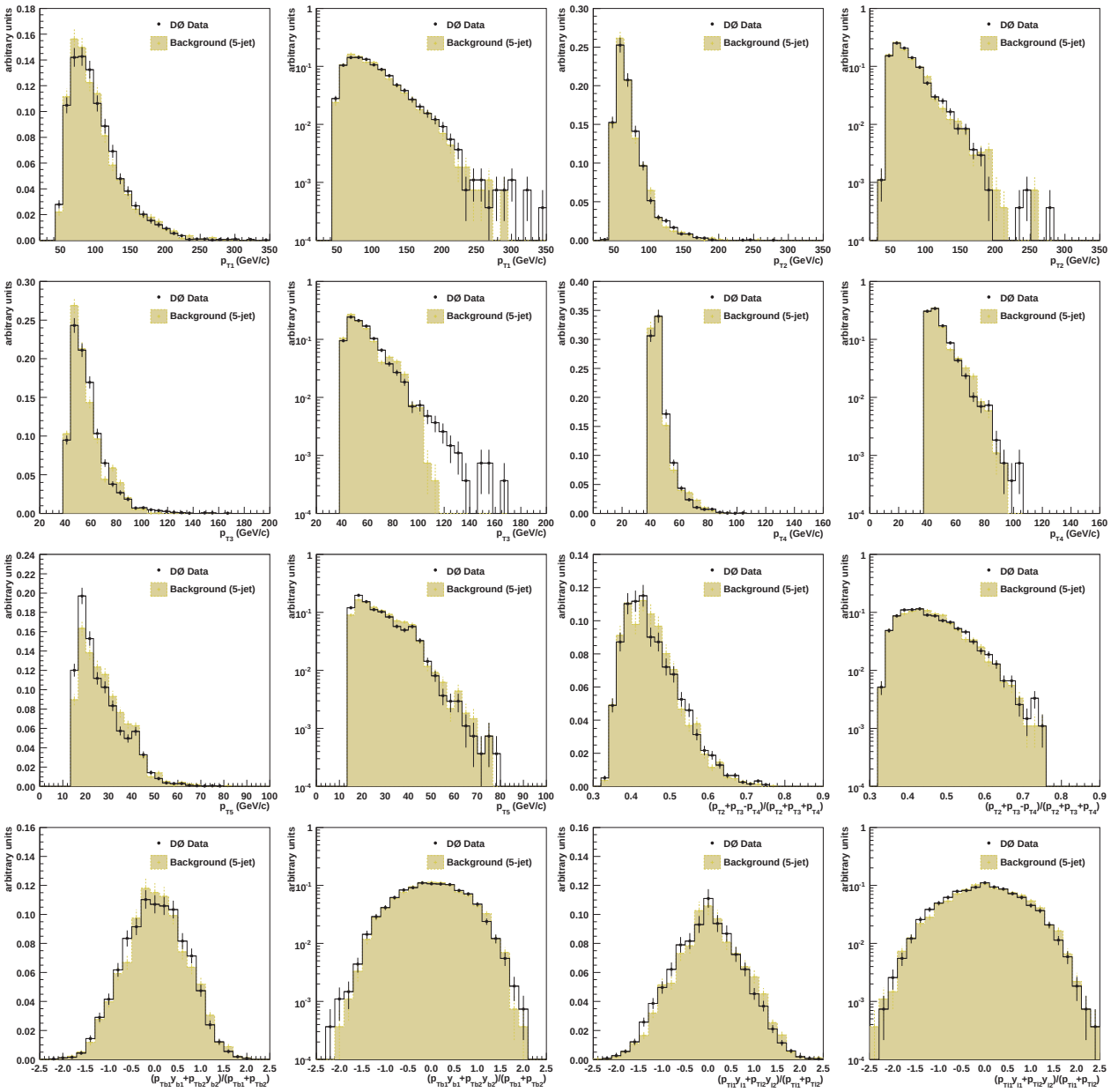


Figure 3.2: Comparison between the 4+1 background validation sample and the original five-jet donor sample. Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

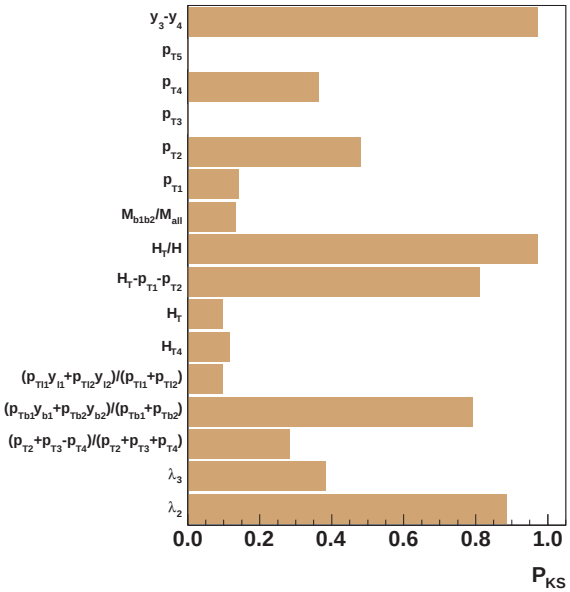


Figure 3.3: Kolmogorov-Smirnov test results corresponding to figure 3.2 for each of the test variables.

sample was quantified using unbinned Kolmogorov-Smirnov tests for a series of representative variables. The Kolmogorov-Smirnov probabilities are shown for each of the test variables in figure 3.3. Clearly the 4+1 background sample describes the original five-jet sample well. Overall, the event shape variables are modeled better than the single jet variables like p_{T}. This is not surprising since those variables are more strongly correlated with the phase space configuration of the event. It should be noted that part of the discrepancy in the distributions of, for example, the jet transverse momenta is enhanced by low statistics at high p_{T}.

3.4 Systematic studies of the background modeling

The crucial piece of the background model is the matching and shaping of the generated phase space. Two individual systematic uncertainties, both related to the phase space occupied by the generated background sample, are considered. The effects of these systematic uncertainties on the generated background samples are discussed in the following sections, the influence of these systematic background variations on the measured cross section are summarised together with all other systematic uncertainties in section 5.2.

3.4.1 Phase space of the acceptor sample

The difference between the 4+2 and 5+1 background samples shows the sensitivity of the background model to the intrinsic phase space of the acceptor sample. Figures 3.4 and 3.5 show a comparison between these two different background samples and the central (mixed) background sample. The two samples differ significantly in distribution shape for some of the test variables. Most of these differences are directly related to the intrinsic phase space differences between the four-jet and five-jet donor samples. This is also reflected by the fact that the largest differences occur for the jet transverse momenta: the individual jet momenta are more sensitive to phase space effects than the global event shape variables.

When considering the jet transverse momenta it's important to keep in mind the origin of the jets. For both the 4+2 and 5+1 samples the leading four jets originate from the acceptor event, but from four-jet and

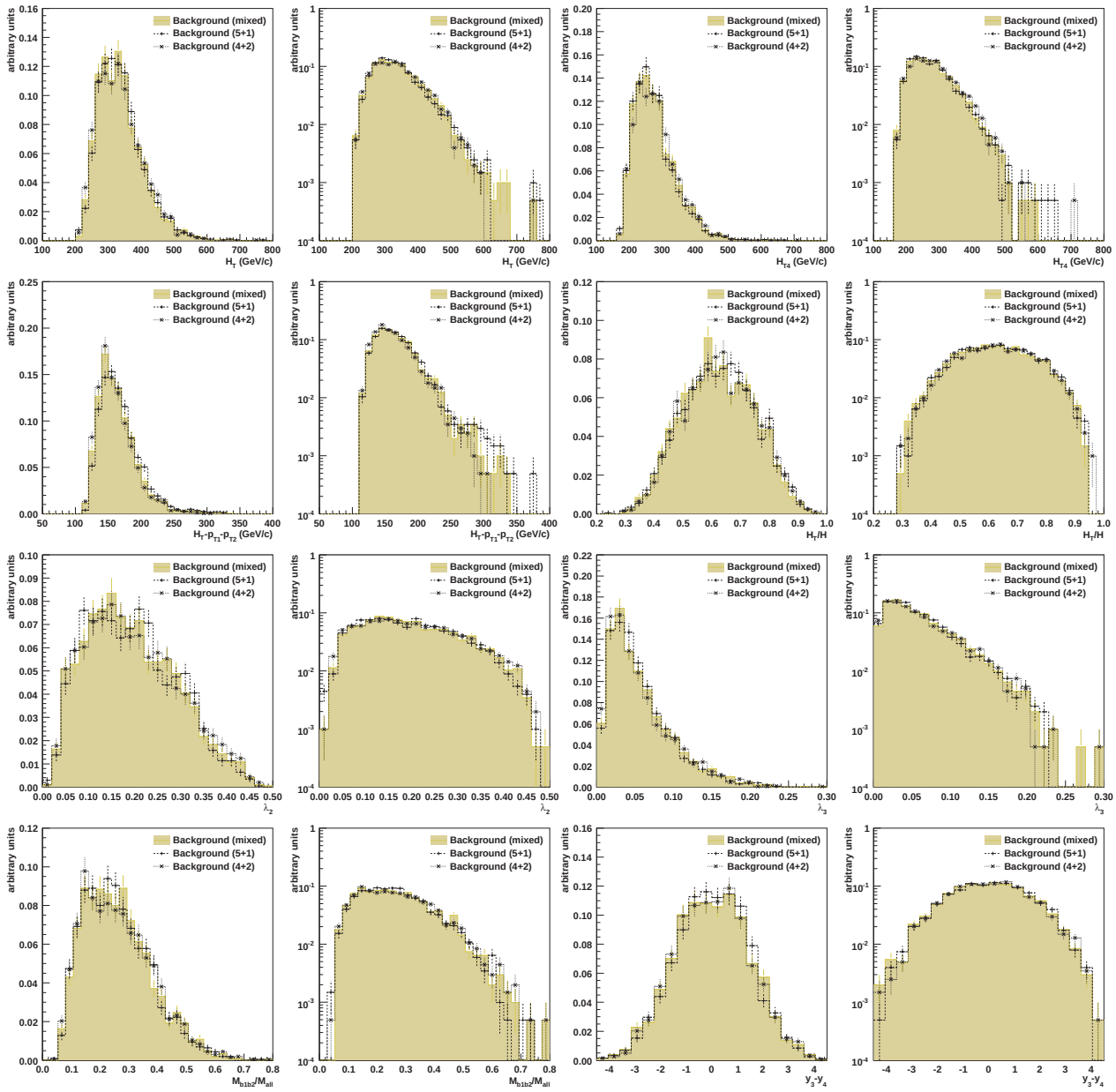


Figure 3.4: Comparison between the 4+2 (× markers) and 5+1 (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.

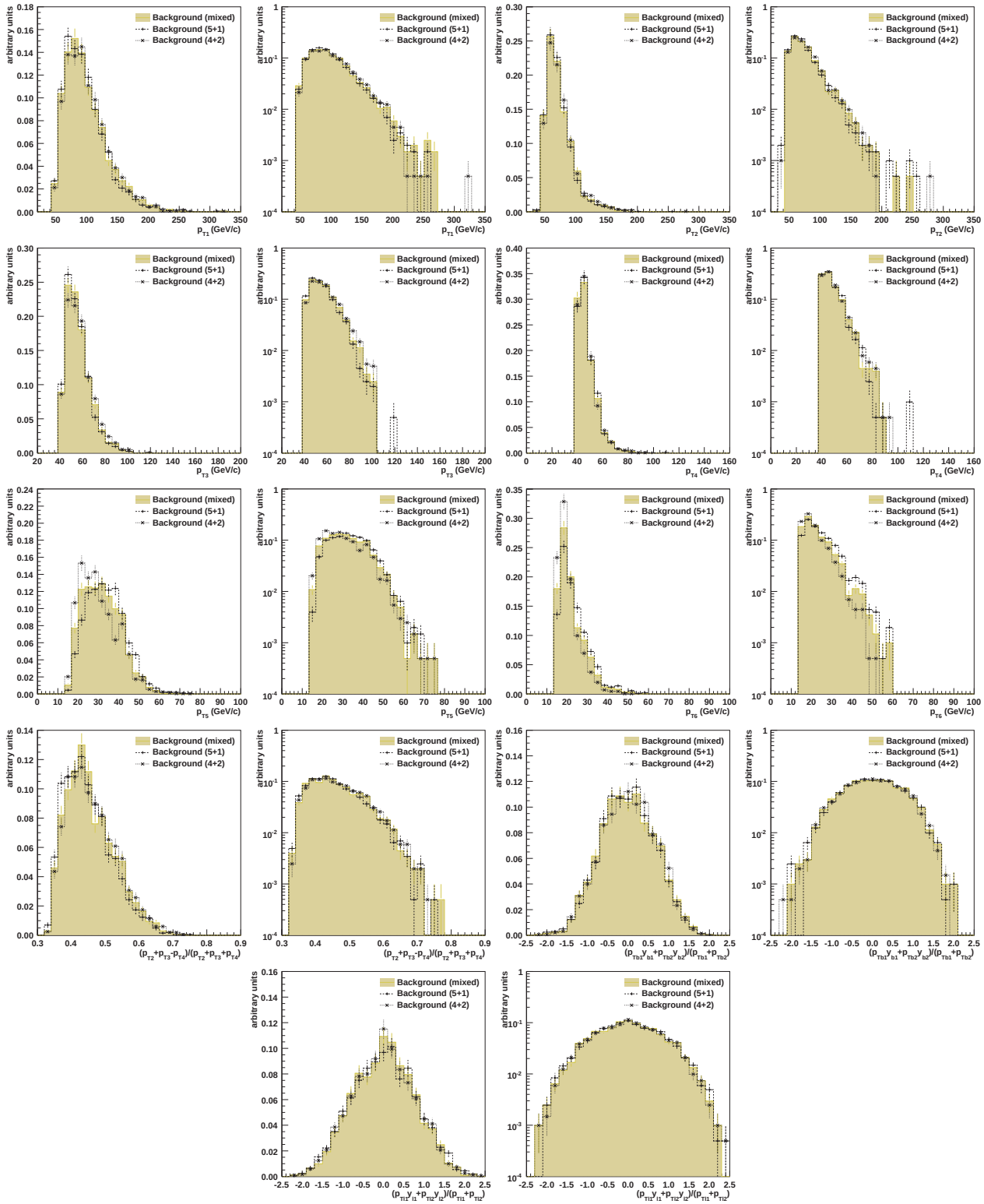


Figure 3.4: Comparison between the 4+2 (× markers) and 5+1 (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.

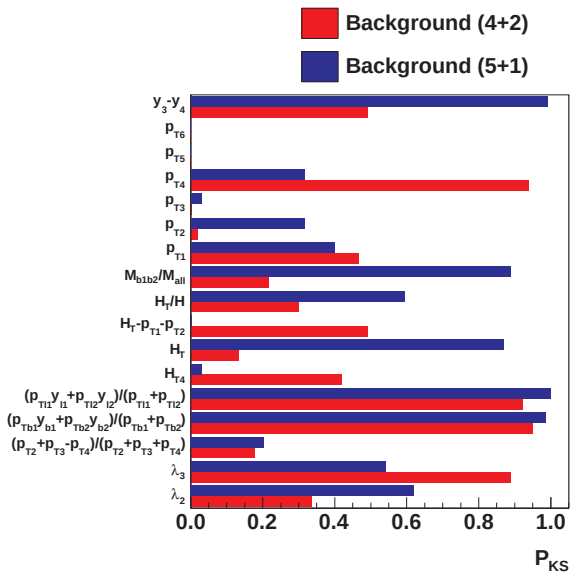


Figure 3.5: Kolmogorov-Smirnov test results from comparing the 4+2 and 5+1 background samples to the central background sample. The disagreement for the jet p_Ts is made worse by low statistics at the high p_T end. This explains why the agreement for H_T is still reasonable.

five-jet acceptor events respectively. The fifth jet in the 4+2 sample is the first donor jet, coming from a six-jet event, for 5+1 event the fifth jet is the softest acceptor jet. For both background samples the sixth and softer jets were part of the six-jet donor event.

The p_{T} distributions of the leading two jets are very similar (i.e. the same within statistics) in the 4+2 and 5+1 samples, demonstrating the dominance of $2 \to 2$ processes. The difference between the p_{T} distributions of the third jet is partly due to the Q^2 matching: four-jet events at a given Q^2 tend to have harder third jets than five-jet events with similar Q^2. A similar feature can be seen in the distribution of the fifth jet p_{T}: the fifth jet in five-jet events (the 5+1 histogram) tends to be harder than in six-jet events (the 4+2 histogram). The sixth jet in both background samples always originates from the donor sample; the differences between the sixth jet p_{T} spectra in the 4+2 and 5+1 samples directly reflect the phase space differences between six-jet events matched to four-jet and to five-jet events.

It is interesting to note the difference in shape of the $(p_{\mathrm{T}2} + p_{\mathrm{T}3} - p_{\mathrm{T}4})/(p_{\mathrm{T}2} + p_{\mathrm{T}3} + p_{\mathrm{T}4})$ variable distributions in figure 3.4. This variable is related to the asymmetry discussed in section 1.1. The observed difference in this case, however, is just a result of the differences in third jet p_{T} spectra. The second and fourth jets have very similar p_{T} distributions, but the softer third jet in the 5+1 sample moves this asymmetry variable towards lower values.

3.4.2 Variations of the shaping cuts

The shaping cuts were optimised on the agreement between the 4+1 validation sample and the original five-jet donor sample. The lower bound on $\not{E}_{\mathrm{T}}$ opens up available phase space for the addition of jets and is an essential part of the background model. Varying the $\not{E}_{\mathrm{T}}/H_{\mathrm{T}4}$ cut allows for more or less phase space for additional jets. The effects of variations in each of these two cuts were studied, showing that the performance of the background model is stable under variations of the $\not{E}_{\mathrm{T}}$ cut, but does depend on the value

of the $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}}$ cut. Therefore, it was decided to keep the $\not{E}_{\mathrm{T}}$ cut fixed and only vary the $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}}$ cut to reflect the effects of the choice of the shaping cuts. 'Up' and 'down' background samples were generated using $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}} < 0.12$ and $\not{E}_{\mathrm{T}}/H_{\mathrm{T4}} < 0.08$ respectively. The effects are shown in figure 3.6. Clearly the upper shaping cut is not able to change the agreement between the 'up' and 'down' backgrounds and the central value significantly.

The results of the background systematics studies are summarized together with the other systematic uncertainties in section 5.2.

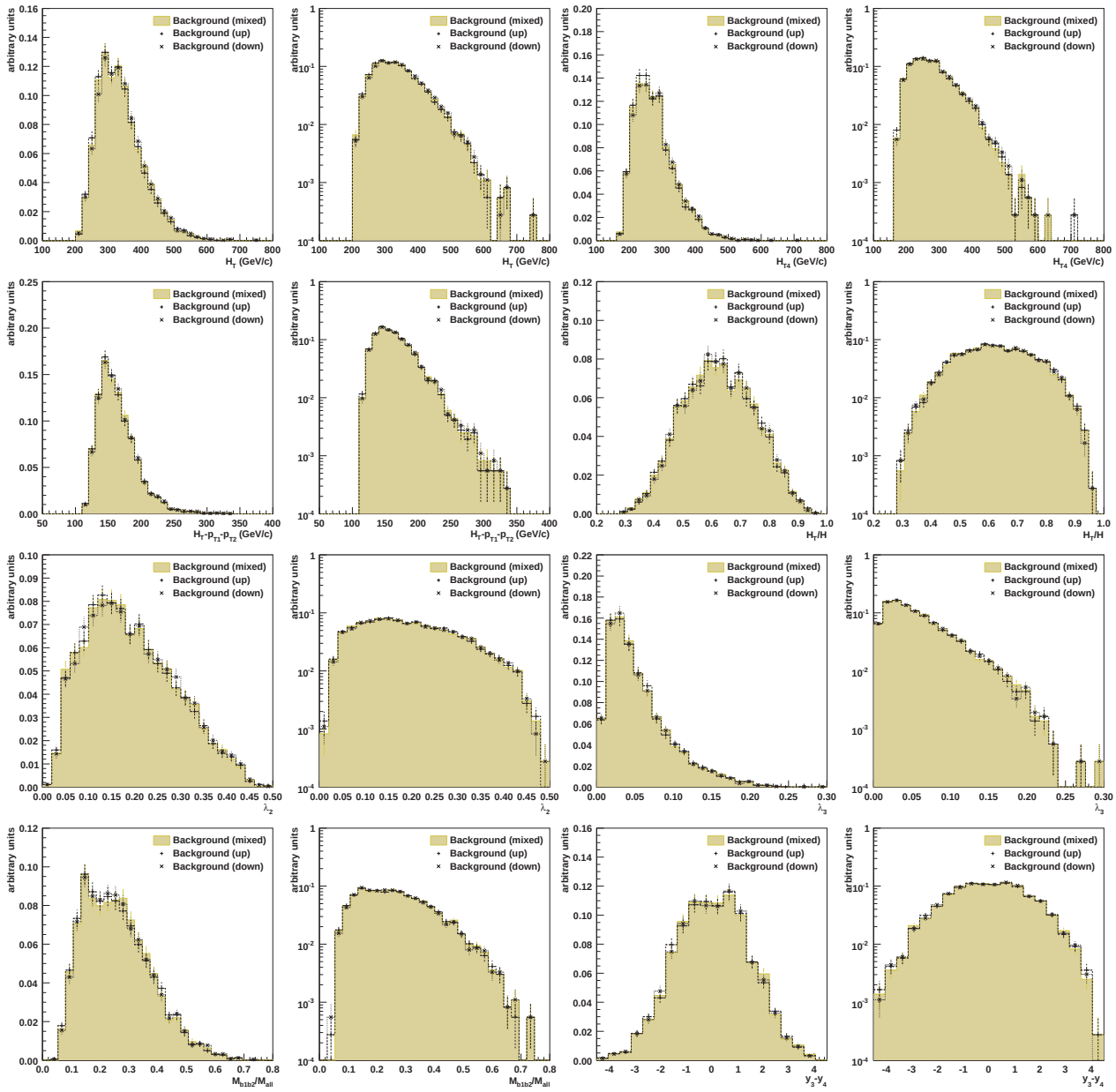


Figure 3.6: Comparison between the 'up' (× markers) and 'down' (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.

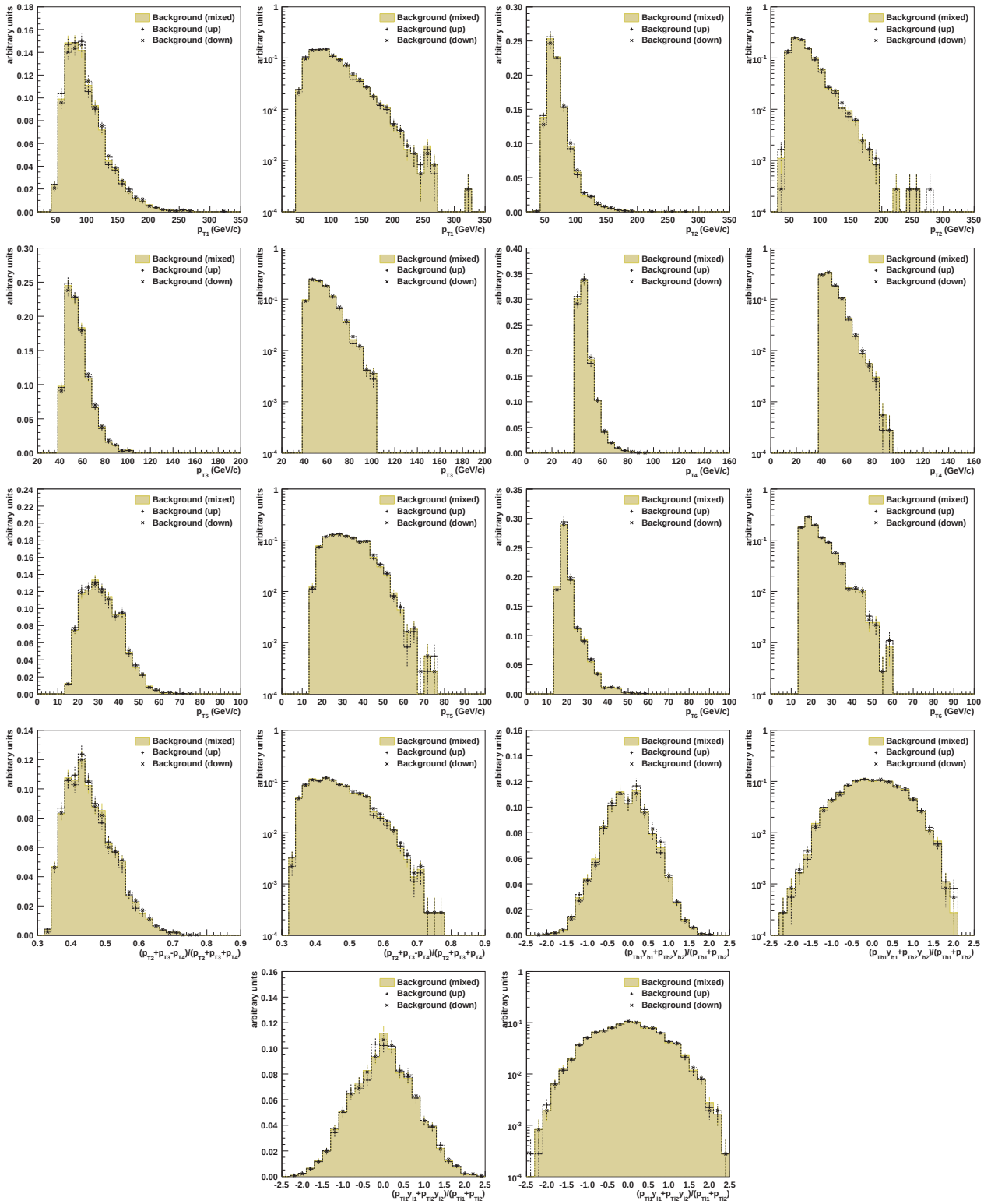


Figure 3.6: Comparison between the 'up' (× markers) and 'down' (+ markers) background samples and the central background sample (solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

Chapter 4

Separating signal from background

Signal- and background contributions in the preselected data sample are estimated using a template based likelihood method. Technically this is implemented using the TMVA [24] package version v3.8.11.

Notation Jets are numbered in order of decreasing p_{T}, starting at one for the leading jet. This jet index is used as subscript for jet properties, e.g. $p_{\mathrm{T}1}$ for the leading jet p_{T}. b-Jets and light jets are marked with subscripts b and l respectively.

4.1 Likelihood input variables

All input variables were chosen based on the following criteria:

- they should provide good separation between signal and background,
- they should be well modeled in the toy background sample,
- they should introduce as little as possible mass dependence into, and the likelihood,
- they shouldn't have significant correlations with each other (Figure 4.1).

The following nine variables were used in the likelihood:

$\boldsymbol{H_{\mathrm{T}}/H}$

The centrality is the ratio of summed jet transverse momenta to the summed jet energies (where the sum runs over all jets in the event):

$$\frac{H_{\mathrm{T}}}{H} \equiv \frac{\sum_{\mathrm{jets}} p_{\mathrm{T}\,\mathrm{jet}}}{\sum_{\mathrm{jets}} E_{\mathrm{jet}}}.$$

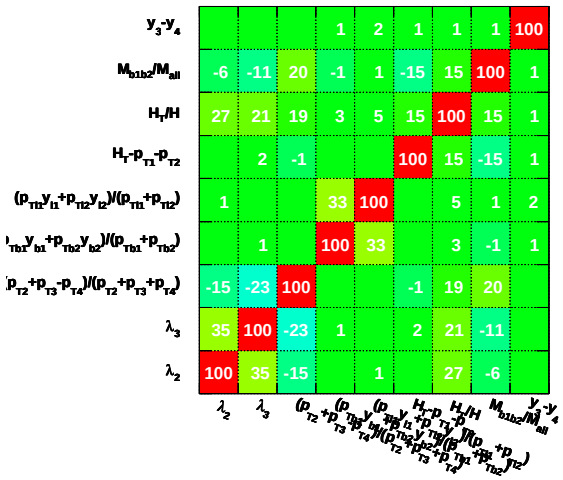
It provides a handle on the mass of the particles produced in the hard scatter. High mass multi-jet events, like the $\mathrm{t\bar{t}}$ signal, are expected to be more central than QCD multi-jet events.

$\boldsymbol{H_{\mathrm{T}} - p_{\mathrm{T}1} - p_{\mathrm{T}2}}$

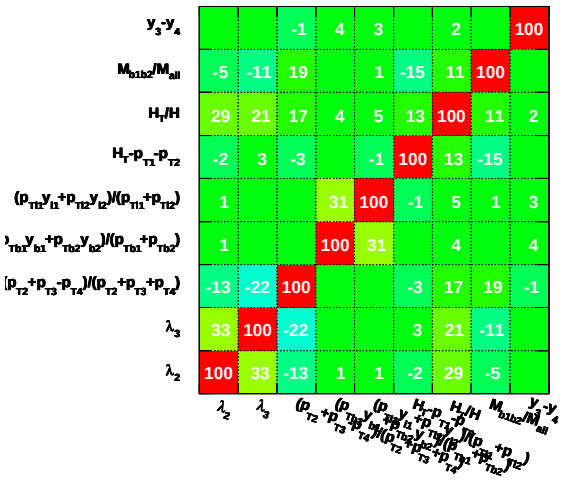
The 'reduced H_{T}' is used as a compromise between H_{T}, which provides strong separation between signal and background, and introducing a mass dependence. H_{T} itself is strongly correlated with the top quark mass, subtracting the two leading jet p_{T}s almost completely removes this mass dependence while retaining most of the separation.

$\boldsymbol{M_{\mathrm{b1b2}}/M_{\mathrm{all}}}$

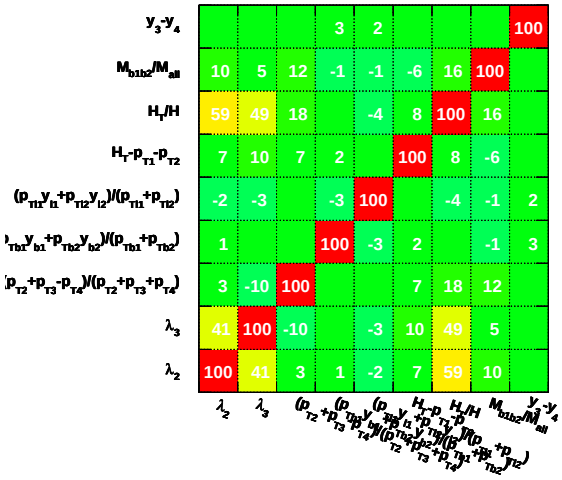
Unlike in $\mathrm{t\bar{t}}$ decays, the QCD background contains a significant amount of $\mathrm{b\bar{b}}$ pairs from gluon splitting. Those $\mathrm{b\bar{b}}$ pairs are expected to be nearly collinear and have small invariant mass. This contribution is visible in the $M_{\mathrm{b1b2}}/M_{\mathrm{all}}$ distribution as the additional peak to the left of the bulk of the distribution.



(a) Alpgen $t\bar{t}$ signal with $m_t = 170$ GeV/c^2



(b) Alpgen $t\bar{t}$ signal with $m_t = 175$ GeV/c^2



(c) Background

Figure 4.1: Correlations between all likelihood input variables for the $t\bar{t}$ signal (b) and for the QCD background (c).

$\boldsymbol{\lambda_2}$, $\boldsymbol{\lambda_3}$

Both are Eigenvalues of the squared momentum tensor [25]:

$$M^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}$$

where the index i runs over all jets in the event and $\alpha, \beta = 1, 2, 3$ denote the three spatial components of the jet momenta. This tensor has three Eigenvalues: λ_1, λ_2 and λ_3 ordered in order of decreasing value. The sum of all three Eigenvalues is one. The sum of λ_2 and λ_3 is a measure of the spherical symmetry of the event. Compared to the QCD background hadronic $\mathrm{t\bar{t}}$ events generally are more spherically symmetric and have correspondingly larger value of λ_2 and λ_3. By separating the two, correlations between the input variables can be avoided. The third and smallest Eigenvalue of the squared momentum tensor is equal to the smallest sum of squares of jet transverse momenta. λ_3 Determines the aplanarity of an event: a measure of how well all jets lie in a single plane. $\mathrm{t\bar{t}}$ Multi-jet events tend to be more spherical, and thus have larger values for λ_3, than QCD multi-jet events.

$\boldsymbol{y_3 - y_4}$

In the case of hadronic $\mathrm{t\bar{t}}$ decays, often the two leading jets are the b-jets from the top quark decays. The third and fourth jets are often a quark-antiquark pair from the decay of the same W boson and will predominantly be close together. This rapidity difference provides an estimate of the average separation between the two jets from the leading W boson in the $\mathrm{t\bar{t}}$ signal. For the case of QCD multi-jet events these jets will in general be spread more widely apart. (The QCD background is dominated by di-jet events with additional radiative jets. For those events the third and fourth jets will generally be in opposite hemispheres.)

$\boldsymbol{(p_{\mathrm{T}2} + p_{\mathrm{T}3} - p_{\mathrm{T}4})/(p_{\mathrm{T}2} + p_{\mathrm{T}3} + p_{\mathrm{T}4})}$

This variable is related to the energy asymmetry between the quark-antiquark pair from a W decay originating from a top quark discussed in Section 1.1. It is unknown which quarks originate from which W boson. In general the two b-jets will be the leading jets. The third and fourth jets would then be the jets from the highest p_{T} W. To account for the fact that sometimes the b-jets may not be the leading jets the calculated asymmetry treats the second and thirds jet equally. Even though the effect is diluted by these assumptions, as well as by the use of the laboratory frame instead of the (anti)top quark rest frame, the expected tendency of the signal towards smaller asymmetries is still clearly present.

$\boldsymbol{(p_{\mathrm{Tb1}} y_{\mathrm{b1}} + p_{\mathrm{Tb2}} y_{\mathrm{b2}})/(p_{\mathrm{Tb1}} + p_{\mathrm{Tb2}})}$

The p_{T}-weighted rapidity sum of the leading two b-jets is related to the mass of the all-jet system. The large top quark mass prevents large boosts along the z-direction. This maintains relatively small jet rapidities while at the same time keeping the b-jets more back-to-back than is the case for QCD $\mathrm{b\bar{b}}$ production. Both effects lead to a narrower distribution for the $\mathrm{t\bar{t}}$ signal than for the QCD background.

$\boldsymbol{(p_{\mathrm{Tl1}} y_{\mathrm{l1}} + p_{\mathrm{Tl2}} y_{\mathrm{l2}})/(p_{\mathrm{Tl1}} + p_{\mathrm{Tl2}})}$

This variable represents the light-jet version of the previous parameter. The higher light-jet multiplicity broadens the distribution.

The distributions of the above variables for both $\mathrm{t\bar{t}}$ signal and background samples are shown in Figures 4.2 and 4.3. Clearly none of these variables possesses enough separating power by themselves to distinguish signal from background.

4.2 Likelihood performance

The likelihood templates were determined separately for the $m_{\mathrm{t}} = 170\,\mathrm{GeV}/c^2$ and $m_{\mathrm{t}} = 175\,\mathrm{GeV}/c^2$ ALPGEN all-hadronic $\mathrm{t\bar{t}}$ signal samples. The signal- and background samples were randomly split into two halves each: one used to determine the templates for the likelihood analysis and one for use as a test sample to verify the likelihood performance. The likelihood output distributions for data and signal are shown in both linear-

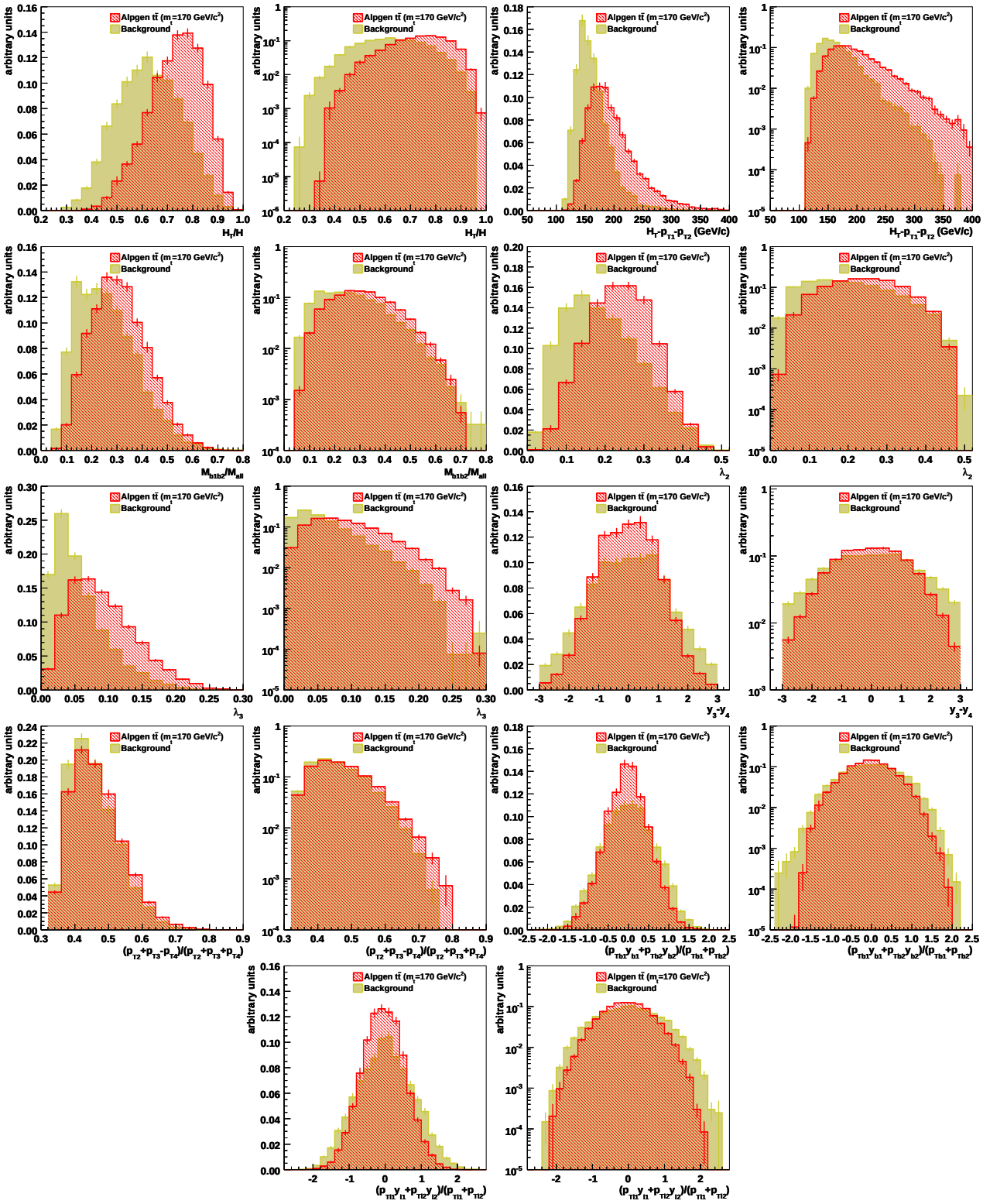


Figure 4.2: Distributions of the likelihood input variables for the 170 GeV/c^2 $t\bar{t}$ signal (orange, hashed fill) and background (brown, solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.

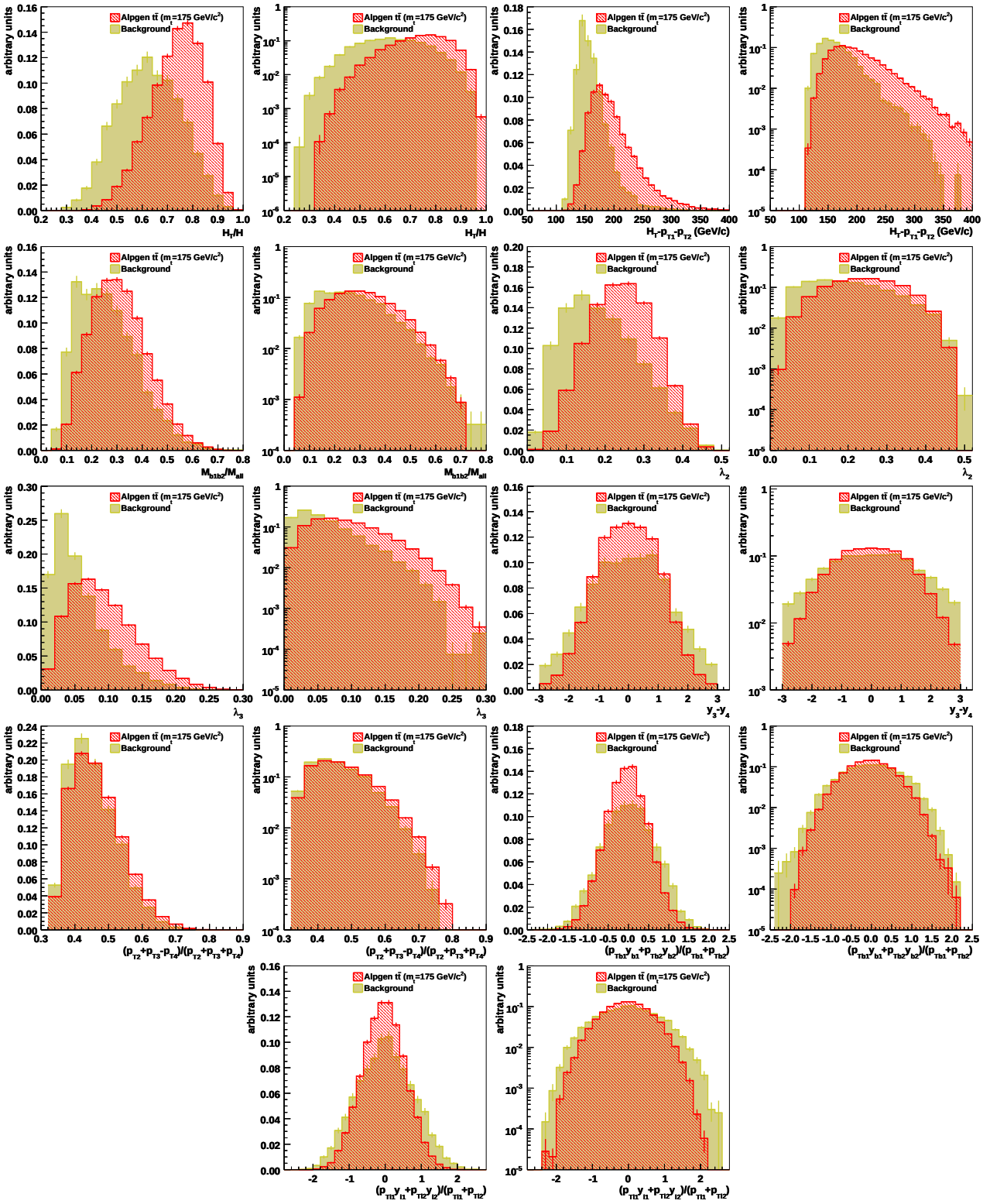
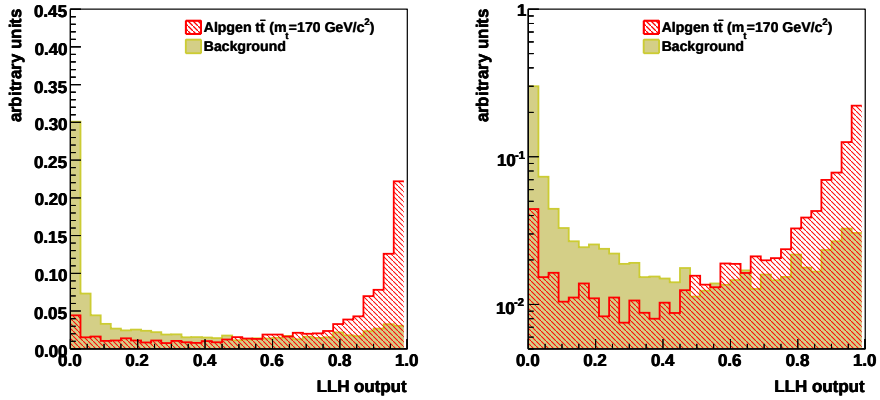
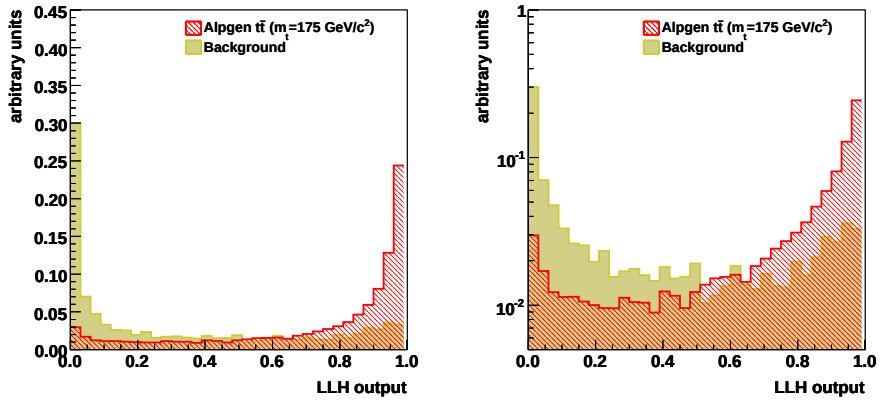


Figure 4.3: Distributions of the likelihood input variables for the 175 GeV/c^2 $t\bar{t}$ signal (orange, hashed fill) and background (brown, solid fill). Left-hand side: linear scale, right-hand side: logarithmic scale.



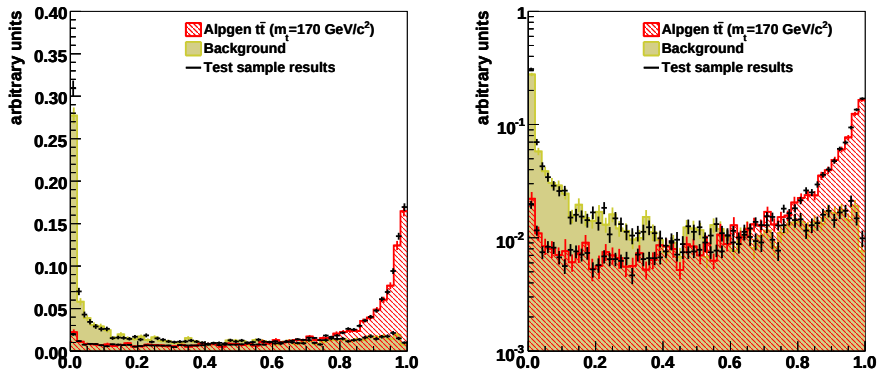
(a) 170 GeV/c^2



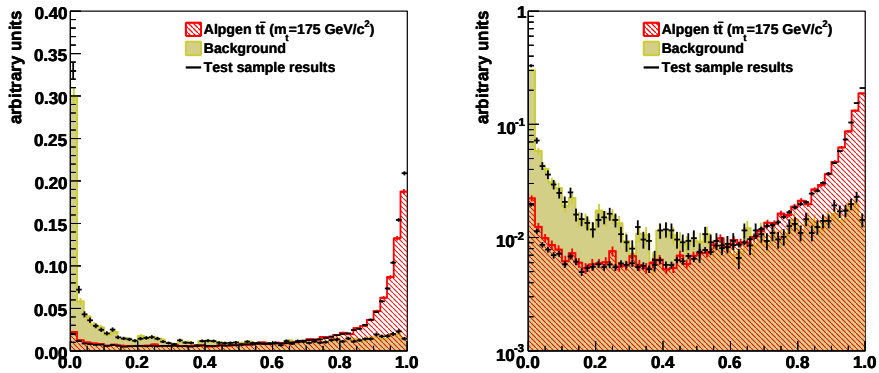
(b) 175 GeV/c^2

Figure 4.4: Likelihood output distribution. Left-hand side: linear scale, right-hand side: logarithmic scale.

and log scales in Figure 4.4. Figure 4.5 shows an overtraining check performed on the used likelihood. This verifies that the 'training' samples used to determine the signal- and background templates were sufficiently large to span the full sample variations. The good agreement between the results from the test sample and the templates themselves shows overtraining is no issue in this case.



(a) 170 GeV/c^2



(b) 175 GeV/c^2

Figure 4.5: Shown are the signal- and background templates (hashed orange- and solid black fill respectively) and the likelihood output from the test samples (the second halves of the 'training' samples). The good agreement between the templates and the test results shows there is no sign of overtraining. Left-hand side: linear scale, right-hand side: logarithmic scale.

Chapter 5

Results

5.1 Signal Purity

Figure 5.1 shows the direct result of a fit of the signal and background likelihood templates to the data likelihood distribution. Distributions for several observables not included in the likelihood, using the signal and background fractions from the fits, are shown in Figures 5.2 and 5.3 for $m_{\mathrm{t}} = 170$ and 175 GeV/c^2, respectively. Comparisons with other variables are shown in Appendices B and C. The fit results show good agreement with the data, as well as a good Kolmogorov-Smirnov probability. The signal fractions are $(12.9 \pm 2.4)\%$ for m_{t} =170 GeV/$c^2$and $(12.5 \pm 2.3)\%$ for m_{t} =175 GeV/c^2. Given 1051 data candidate events, this results in 136 and 131 $\mathrm{t\bar{t}}$ events, respectively.

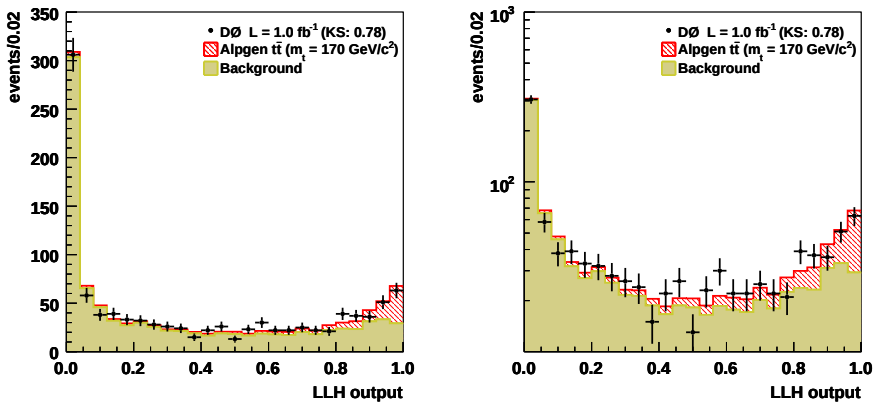
5.2 Systematic uncertainties

The effects of systematic uncertainties and variations in input variables are studied using ensemble tests. With the exception of the background systematics, all of the systematic uncertainties are associated with the signal Monte Carlo. All systematic uncertainties on the measured hadronic $\mathrm{t\bar{t}}$ production cross section for both $m_{\mathrm{t}} = 170$ GeV/c^2 and $m_{\mathrm{t}} = 175$ GeV/c^2 mass points are summarised in Table 5.1. Many of these are self-explanatory or, like the background uncertainty (Section 3.4) have been already discussed. The others are detailed below.

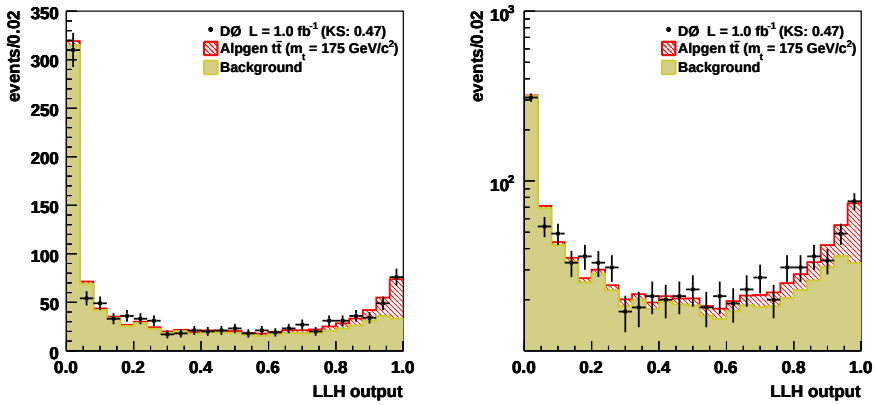
5.2.1 Signal model

This analysis relies on the ALPGEN +PYTHIA Monte Carlo for the $\mathrm{t\bar{t}}$ signal estimation. This is key for both the measurement of the preselection efficiency (Table 2.3 and for the shapes used in the likelihood determination (Chapter 4). It is possible that the signal MC does not properly reproduce the properties of the $\mathrm{t\bar{t}}$ system. Many other $\mathrm{t\bar{t}}$ analyses published by the DØ collaboration in the lepton+jets and dilepton decay channels have found excellent agreement between the ALPGEN Monte Carlo and the reconstructed data. Nevertheless, the MC might mis-estimate the jet multiplicity through differences in the QCD radiation or the underlying event. Given a lack of alternate choices for the signal model and a lack of MC samples with varied parameters (e.g., more or less initial state radiation), we must extract the systematic uncertainty associated with the signal model from the data.

Changes in the signal model are reflected in the reconstructed jet multiplicity. We therefore took the fraction of reconstructed $\mathrm{t\bar{t}}$ events (using the measured signal purities) with seven or more jets and weighted the signal MC events up and down by one sigma in the statistical uncertainty on this ratio. The entire analysis was repeated and the resulting difference in the mean cross section applied as a systematic uncertainty. The fraction of events with seven or more jets was 0.29 ± 0.04 for both top quark masses. We therefore weighted



(a) $m_{\mathrm{t}} = 170$ GeV/c^2



(b) $m_{\mathrm{t}} = 175$ GeV/c^2

Figure 5.1: Likelihood output distribution for the selected data sample (black markers), together with the signal and background contributions (hashed orange and solid brown fill respectively) obtained from the fraction-fit described in Chapter 4. Left: linear scale, right: logarithmic scale.

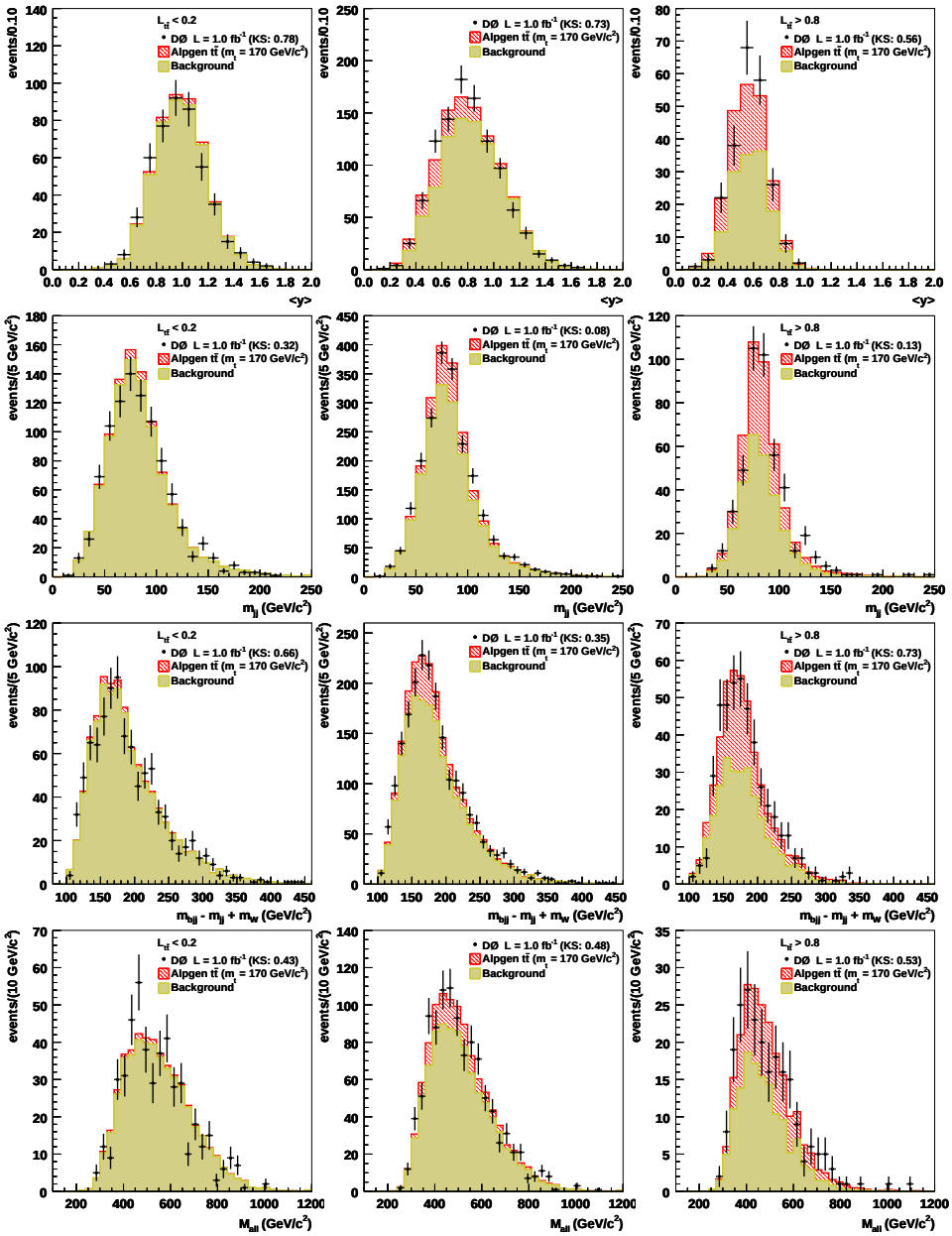


Figure 5.2: Center: Distributions for selected observables (black markers), together with the signal and background contributions (hashed orange and solid brown fill respectively) obtained from the fraction-fit described in Chapter 4. Left: same, but requiring likelihood< 0.2. Right: same, but requiring likelihood> 0.8. All figures are for $m_t = 170\ GeV/c^2$

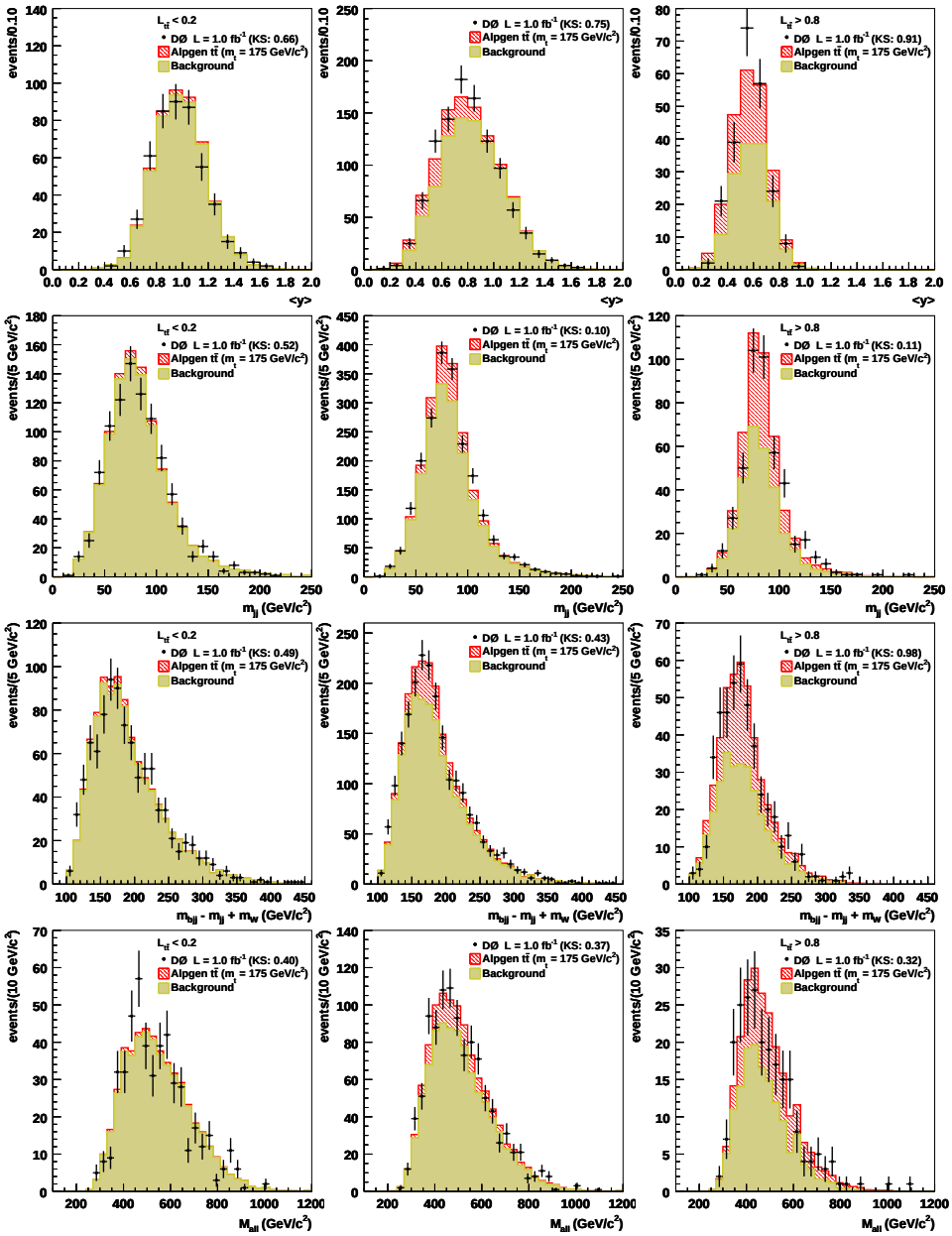


Figure 5.3: Center: Distributions for selected observables (black markers), together with the signal and background contributions (hashed orange and solid brown fill respectively) obtained from the fraction-fit described in Chapter 4. Left: same, but requiring likelihood< 0.2. Right: same, but requiring likelihood> 0.8. All figures are for $m_t = 175\ GeV/c^2$

Table 5.1: Systematic uncertainties on the $t\bar{t}$ cross section.

Source	Uncertainty pb		Uncertainty %	
data statistics	±1.45		±18.3	
background statistics	±0.30		±3.8	
signal statistics	±0.09		±1.1	
background multiplicity	±0.80		±10.1	
background phase space	±0.28		±3.5	
PDF	−0.29	+0.11	−3.7	+1.4
signal jet multiplicity	±0.03		±0.4	
jet energy resolution	−0.27	+0.12	−3.4	+1.6
jet ID	−0.19	+0.28	−2.4	+3.6
jet energy scale	−0.19	+0.12	−2.4	+1.5
jet energy scale shifting	±0.82		±10.4	
b fragmentation	±0.02		±0.2	
b/light jet response	−0.17	+0.09	−2.2	+1.2
TRF	−0.65	+0.72	−8.2	+9.1
taggability	±0.71		±8.9	
trigger	−0.16	+0.32	−2.0	+4.1
lepton veto	±0.02		±0.3	
luminosity profile	±0.08		±1.0	
vertex reweighting	±0.01		±0.1	
systematic uncertainty	±1.63		±20.6	

(a) $m_t = 170$ GeV/c^2

Source	Uncertainty pb		Uncertainty %	
data statistics	±1.28		±18.5	
background statistics	±0.26		±3.8	
signal statistics	±0.04		±0.5	
background multiplicity	±0.73		±10.5	
background phase space	±0.12		±1.8	
PDF	−0.22	+0.08	−3.2	+1.1
signal jet multiplicity	±0.02		±0.3	
jet energy resolution	−0.22	+0.15	−3.1	+2.2
jet ID	−0.18	+0.20	−2.5	+3.0
jet energy scale	−0.16	+0.11	−2.3	+1.6
jet energy scale shifting	±0.73		±10.6	
b fragmentation	±0.03		±0.4	
b/light jet response	−0.10	+0.07	−1.5	+1.1
TRF	−0.59	+0.63	−8.5	+9.1
taggability	±0.61		±8.8	
trigger	−0.14	+0.27	−2.0	+3.9
lepton veto	±0.01		±0.2	
luminosity profile	±0.01		±0.1	
vertex reweighting	±0.01		±0.1	
systematic uncertainty	±1.42		±20.5	

(b) $m_t = 175$ GeV/c^2

up and down by 15%. This resulted in a very small change in the cross section; this is listed in Table 5.1 as 'signal jet multiplicity'.

The Monte Carlo was also reweighted to correspond to the central PDF of the CTEQ6.5M PDF set. This set also includes forty alternative PDFs representing up/down variations along the twenty eigenvectors of the parameter space. The modified tolerance approach [26] is used to estimate the effects of the individual PDF uncertainties on the measured cross section. This method properly estimates the uncertainties for physical observables X also in the case in which both up and down variations of a given PDF influence X in the same direction. Considering all N (twenty for CTEQ6.5M) up and down variations the uncertainty on the observable X is given by:

$$\Delta X^{+}_{\mathrm{max}} = \sqrt{\sum_{i=1}^{N} \left[max\left(X_i^{+} - X_0, X_i^{-} - X_0, 0\right)\right]^2}$$

$$\Delta X^{-}_{\mathrm{max}} = \sqrt{\sum_{i=1}^{N} \left[max\left(X_0 - X_i^{+}, X_0 - X_i^{-}, 0\right)\right]^2}$$

where the index i runs over the different eigenvectors and X_i^{+} (X_i^{-}) represent the values of X using the up (down) variation of the i-th eigenvector.

5.2.2 Jet energy, resolution and ID

The standard uncertainties on the jet energy scale [27] are propagated through the analysis by shifting the JES corrections up and down by one sigma. Jet resolution effects are estimated by shifting all Monte Carlo jets (in p_{T}) by the relative data-to-MC jet energy scale [21]. This assumes that the full jet energy scale difference is due to ununderstood resolution and efficiency effects in the Monte Carlo. The jet identification efficiency is almost 100%. Statistics-based upward variations would exceed unity, so the systematics effects due to the jet ID efficiency use the statistical uncertainty on the efficiency for the downward variation, while for the upward variation the efficiency is set to 100%.

5.2.3 b-Jet related uncertainties

The b-jet related uncertainties contain contributions from:

- b-fragmentation,
- jet taggability,
- b-jet identification, and
- b-jet energy calibration.

The b-fragmentation reweighting discussed in Section 2.1 contains a single alternative parameterisation for use in systematic studies. This alternative reweighting is applied symmetrically by defining $w_{\mathrm{down}} = w_{\mathrm{central}} - (w_{\mathrm{up}} - w_{\mathrm{central}})$. As discussed in Section 2.8, the taggability was varied down by 2.0% (absolute); the same result used for the upwards systematic. The effects of the b-tagging efficiency were estimated by varying the TRFs [23] in both data and Monte Carlo by their respective one-sigma uncertainties. The jet energy scale was determined on a sample dominated by light-jets. The systematic uncertainty due to differences in both kinematics and particle composition between b-, and light-jets, the relative b-to-light jet response is estimated by scaling the relative response for b-jets in Monte Carlo down to 0.982 and up to $1/0.982 = 1.018$, following the prescription in ref. [28].

5.2.4 Luminosity

The uncertainty on the total, integrated luminosity recorded by D0 during Run IIa is 6.1% [29]. This contribution is not combined within the systematic uncertainty but quoted separately. The systematic uncertainty on the reweighting of the overlaid zero-bias luminosity in Monte Carlo was estimated as the (symmetrised) full difference between performing the analysis using Monte Carlo samples with and without this luminosity reweighting.

5.2.5 Primary vertex position and reweighting

To estimate the effect of a possibly imperfect primary vertex distribution in the Monte Carlo ('Beam spot profile' in Table 5.2a) the analysis was repeated without applying the primary vertex position reweighting discussed in Section 2.1. The full difference with the measured cross section was applied as a symmetric uncertainty.

5.3 Result

Combined with the preselection efficiencies (Table 2.3) and luminosity (Table 2.2), this results in the following cross sections:

$$\sigma_{t\bar{t}}(m_t = 170\,\mathrm{GeV}/c^2) = 7.9\ \pm 1.5\ (\mathrm{stat})\ \pm 1.6\ (\mathrm{sys})\ \pm 0.5\ (\mathrm{lum})\ \mathrm{pb}$$
$$\sigma_{t\bar{t}}(m_t = 175\,\mathrm{GeV}/c^2) = 6.9\ \pm 1.3\ (\mathrm{stat})\ \pm 1.4\ (\mathrm{sys})\ \pm 0.4\ (\mathrm{lum})\ \mathrm{pb}$$

The statistical uncertainty includes the statistical uncertainties associated with the signal and background templates. This was determined by re-fitting the data 100k times but allowing the signal and background templates to vary according to their bin-to-bin statistical uncertainties. The resulting uncertainties are summarized in Table 5.1.

Figure 5.4 shows the standard model prediction together with the measured cross section from this analysis. The measured cross section is in perfect agreement with the expectation.

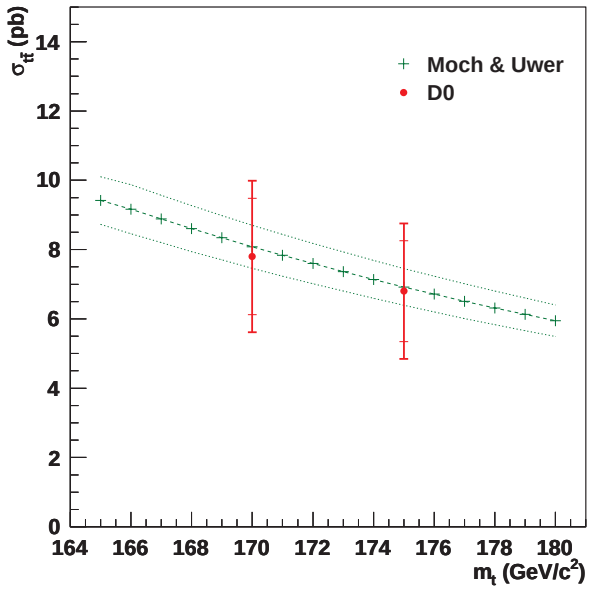


Figure 5.4: The $t\bar{t}$ production cross section measured in this analysis (• markers) together with the standard model prediction [2]. Shown with the measured cross section are the corresponding systematic uncertainty (including luminosity contributions, inner error bars) and the total uncertainty (outer error bars).

Appendix A

Trigger Details

Run range	Trigger version [from	to)	Run numbers [from-to)
v8	v8.00	v9.00	160582-167016
v9 - v10	v9.00	v11.00	167019-174803
v11	v11.00	v12.00	174845-178018 178080-178097 178104-178560 178645-178722
v12	v12.00	v13.03	178019-178071 178097-178104 178618-178620 178722-194567 194595-194598
v13a	v13.03	v13.30	194567-194574 194662-202025
v13b	v13.30	v14.00	202025-207217 207229-207258 207288-207343 207350-207681 207749-208146 208246-208247 208444-208501 209102-209105
v14a	v14.00	v14.80	207217-207223 207279-207280 207343-207346 207719-207741 208168-208207 208276-208435 208512-209100 209147-211483
v14b	v14.80	v15.00	211513-215671

Table A.1: The different run ranges used defined in terms of run numbers

Run range	Trigger	Level	Terms
v8	4JT10	Level 1 Level 2 Level 3	CJT(4,5) 3JET(0,8.,JET(0,7.)) L3FJet(SCJET_9,4,10.,0.,3.,0.,1.)
v9 - v10	4JT10	Level 1 Level 2 Level 3	CJT(4,5) 3jet8_ht90 JET(SCJET_9,4,10.)_JET(SCJET_9,2,20.)
v11	4JT10	Level 1 Level 2 Level 3	CJT(4,5)_ncu 3jet8_ht90 JET(SCJET_9,4,10.)_JET(SCJET_9,2,20.)
v12	4JT12	Level 1 Level 2 Level 3	CJT(3,5)_ncu 3jet8_ht50 L3FJet(SC5JET_9_PV1,4,12.,0.,3.6)_L3FJet(SC5JET_9_PV1,3,15.,0.,3.6)_ L3FJet(SC5JET_9_PV1,2,25.,0.,3.6)
v13a	JT2_4JT12L_HT	Level 1 Level 2 Level 3	CJT(3,4eta2.4)CJT(2,5)_ncu 3JET6_HT70 Jet(JET9_PV1,4,12.,3.6)_Jet(JET9_PV1,3,15.,3.6)_Jet(JET9_PV1,2,25.,3.6)_ Ht(2Jet(JET9_PV1,1,9.,3.6),240.)
v13b	JT2_4JT12L_HT	Level 1 Level 2 Level 3	CJT(3,5)_ncu 3JET6_HT70 Jet(JET9_PV1,4,12.,3.6)_Jet(JET9_PV1,3,15.,3.6)_Jet(JET9_PV1,2,25.,3.6)_ Ht(2Jet(JET9_PV1,1,9.,3.6),240.)
v14a	JT2_4JT20	Level 1 Level 2 Level 3	CJT(3,5)CJT(3,4eta2.6)_ncu 3JET6_ETA26HT75 Jet(JET9_PV3,4,20.,3.6)_Jet(JET9_PV3,2,25.,3.6)
v14b	JT2_4JT20	Level 1 Level 2 Level 3	CJT(3,5)CJT(3,4eta2.6)CJT(1,7eta1.8)_ncu 3JET6_ETA26HT75 Jet(JET9_PV3,4,20.,3.6)_Jet(JET9_PV3,2,25.,3.6)

Table A.2: The 4JT trigger definitions. Minor differences between trigger terms have been ignored, as have version numbers and mark-and-pass terms.

Appendix B

Likelihood control plots

This appendix combines all the control plots for the likelihood analysis used to separate signal from background. Distributions are shown for a series of 59 relevant variables (figure C.4) including all likelihood input variables. The signal and background normalizations were taken from the cross section measurement given in Section 5.3. Agreement was quantified using Kolmogorov-Smirnov tests on the histograms. [1] Performing the tests on the histograms instead of on the unbinned data points biases the results towards slightly higher probabilities. Since the real criterion is for the distribution of the test results to be uniform, this effect can safely be ignored here. Figure C.1 shows the distribution of the resulting Kolmogorov-Smirnov test results.

Apart from the variables already mentioned in the previous chapter, some additional variables are used on the control plots:

Fox-Wolfram moments

In 1979 Fox and Wolfram devised a complete set of spherically symmetric 'event shape' variables to describe the final states of electron-positron collisions [30]. The l-th Fox-Wolfram moment H_l is defined as

$$H_l = \left(\frac{4\pi}{2l+1}\right) \sum_{m=-l}^{+l} \left| \sum_{i}^{\text{particles}} Y_l^m(\Omega_i) \frac{|\vec{p}_i|}{E_{\text{tot}}} \right|^2$$

where $\vec{p}_i$ is the three-momentum of the i-th particle in the final state (which are jets in our case). $\Omega_i = (\theta_i, \phi_i)$ shows the direction of $\vec{p}_i$, E_{tot} is the total energy of all final state particles and Y_l^m are the spherical harmonics. All H_l fall into the range $[0, 1]$. Conservation of momentum in the final state ($\sum_i \vec{p}_i = \vec{0}$) gives $H_1 = 0$.

The combination of the spherical harmonics Y_l^m with the directions of the final state particles $\vec{p}_i$ gives a unique relation between the spatial shape of the final state and the values for H_l..

For perfect back-to-back di-jets events $H_l = 0$ (1) for all even (odd) l. Perfectly spherical events will lead to all H_l being zero for all $l > 0$.

$\cos\theta^*$

The absolute value of the cosine of the angle θ^* between the three-momentum of the leading jet and the beamline in the restframe of the multi-jet (all jets in the event) system.

$\mathbf{\Delta Y_{max}}$

The maximum rapidity difference between any two jets in the event.

H

The sum of the energies of all jet in the event.

[1] Most of the histograms shown have been rebinned for visual reasons. All Kolmogorov-Smirnov tests were performed on the histograms before rebinning.

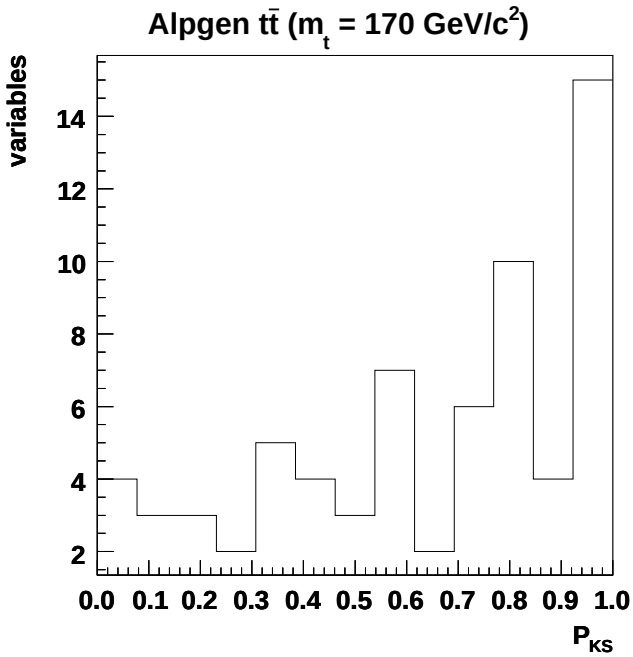


Figure B.1: Kolmogorov-Smirnov test results distribution corresponding to the variables in figure C.4

$\mathbf{M_{b_0 b_1}}$
The invariant mass of the four-vector-sum of the two leading b-jets.

$\mathbf{M_{all}}$
The invariant mass of the four-vector-sum of all jets in the event.

$\mathbf{m_{jj}}$
The dijet invariant mass of the two jets assigned to a PW-boson according to the best χ^2 jet-parton combination.

$\mathbf{m_{bjj}}$
The trijet invariant mass of the two jets assigned to a PW-boson together with the corresponding b-quark according to the best χ^2 jet-parton combination.

$\mathbf{m_{bjj} - m_{jj} + M_W}$
The estimate of the top quark mass from the above two multi-jet masses.

$\log(\mathbf{best}\ \chi^{\mathbf{2}})$
The $\log(\chi^2)$ of the best parton-jet assignment.

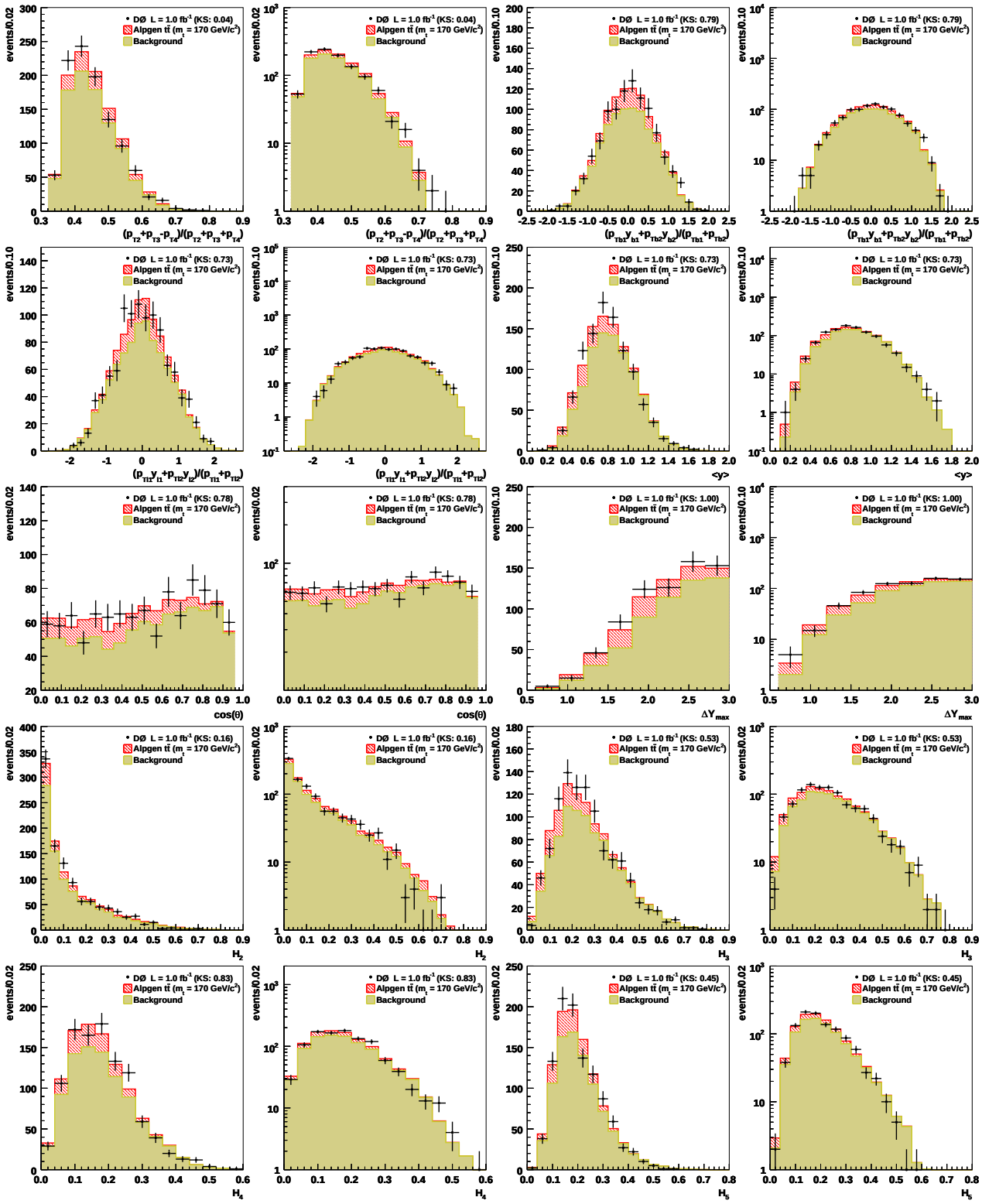


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

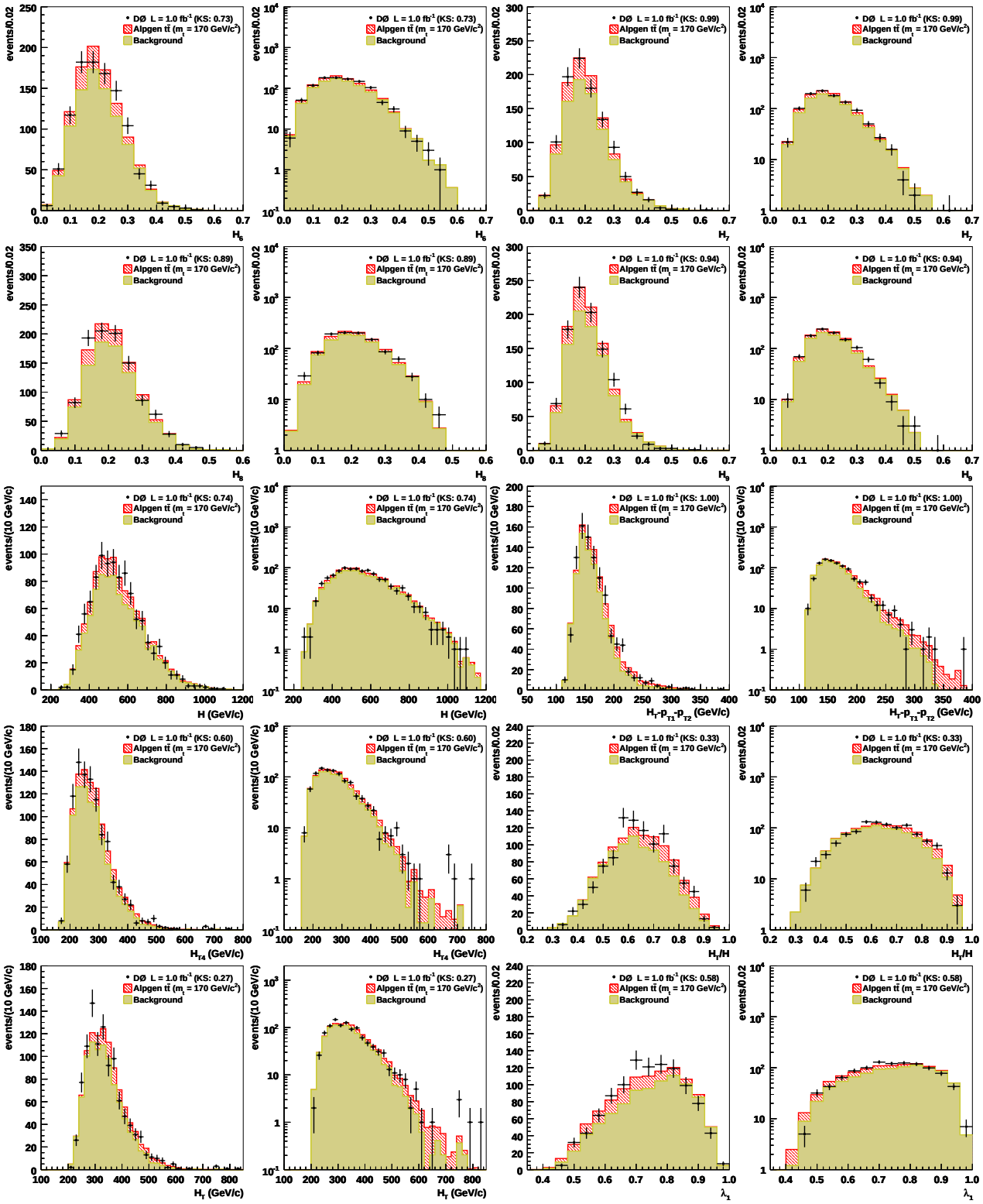


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

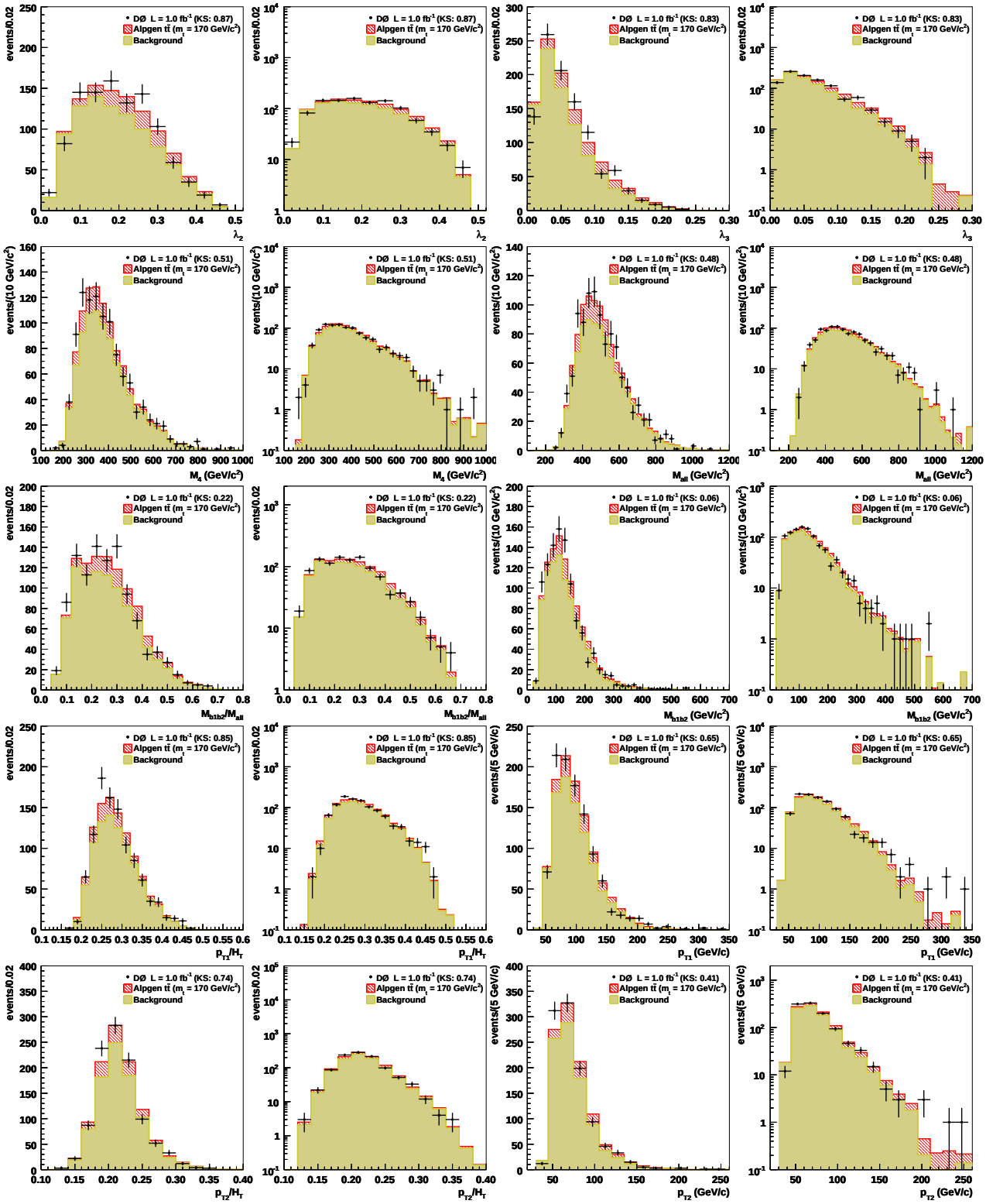


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

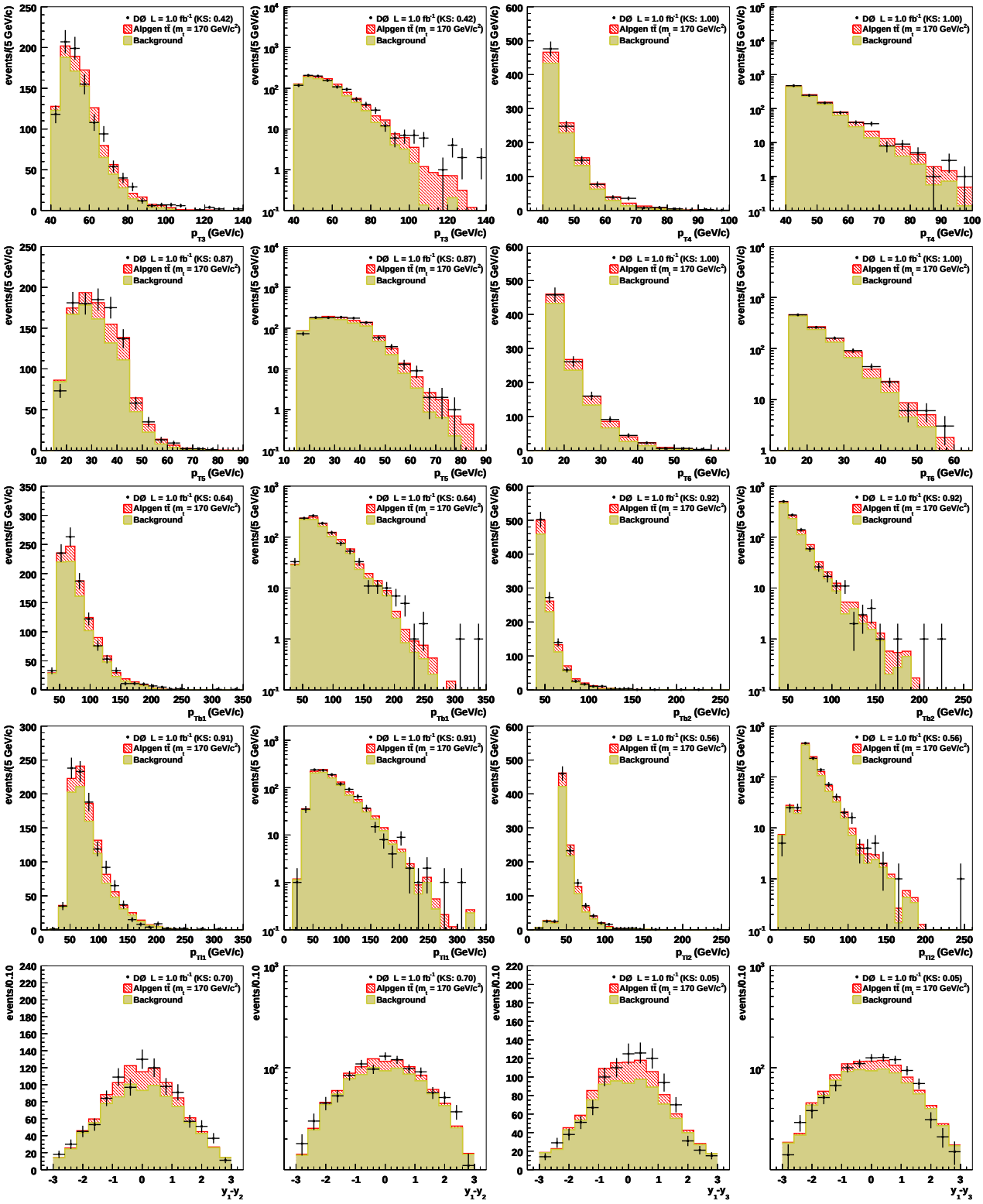


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

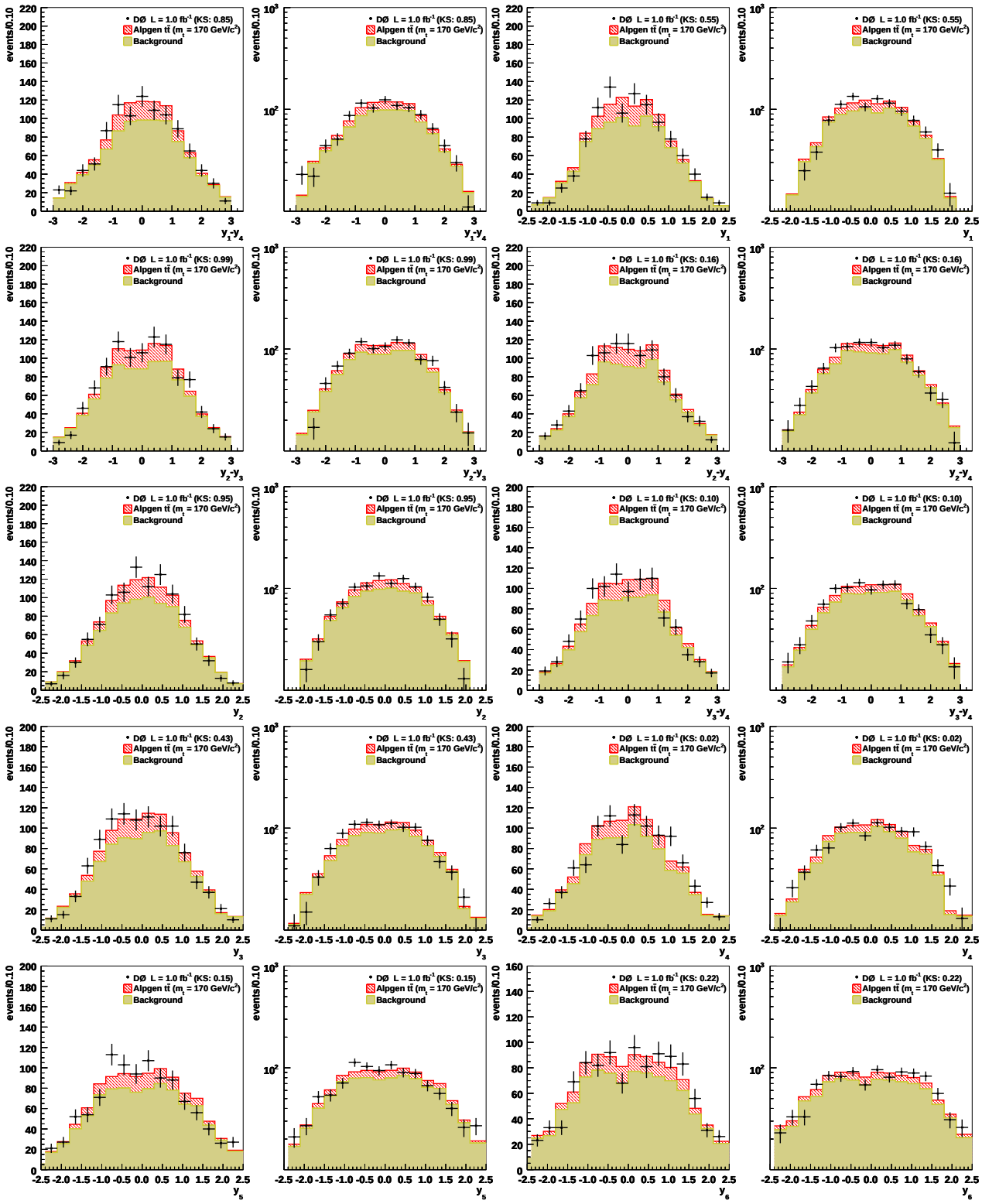


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

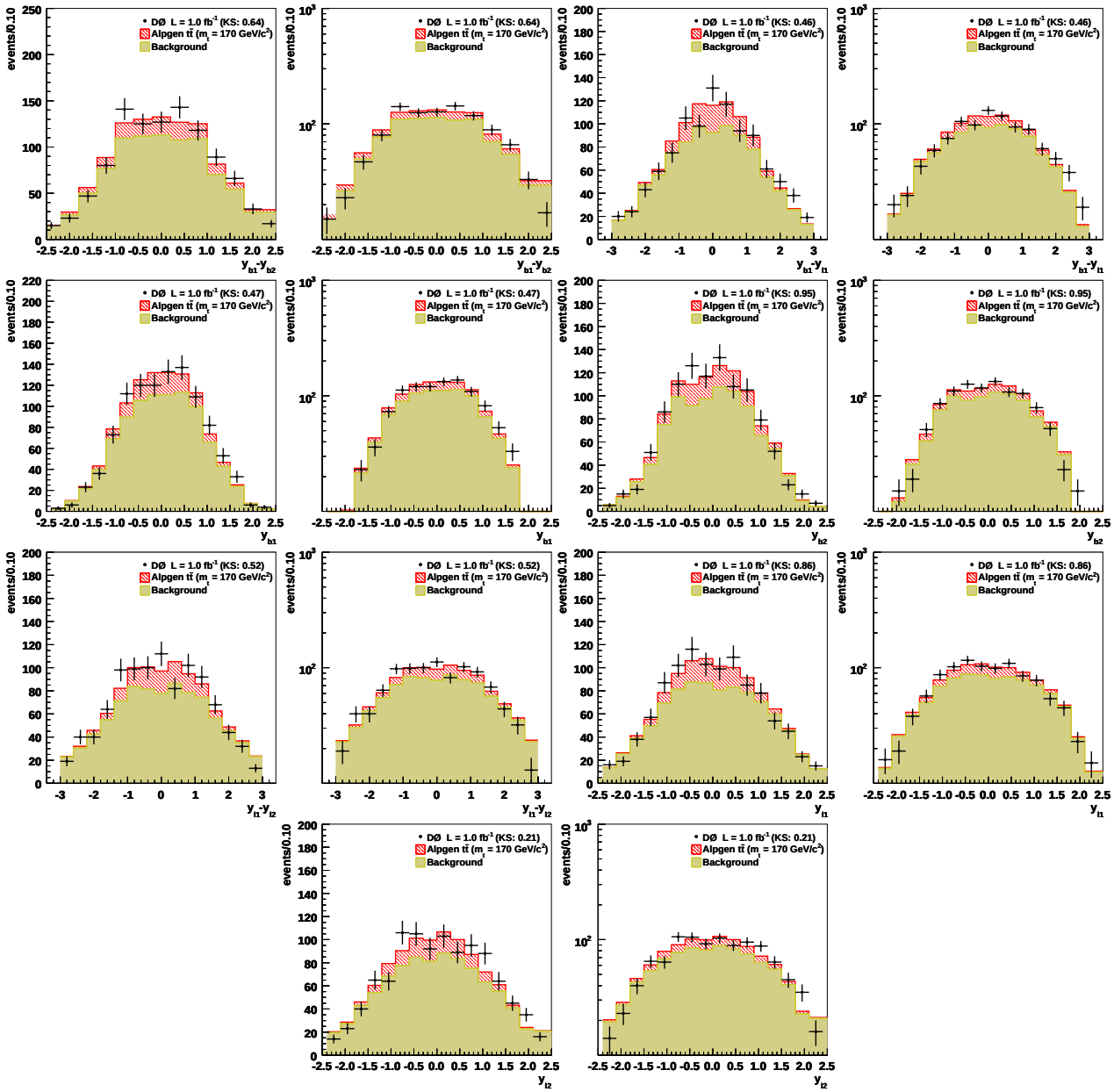


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

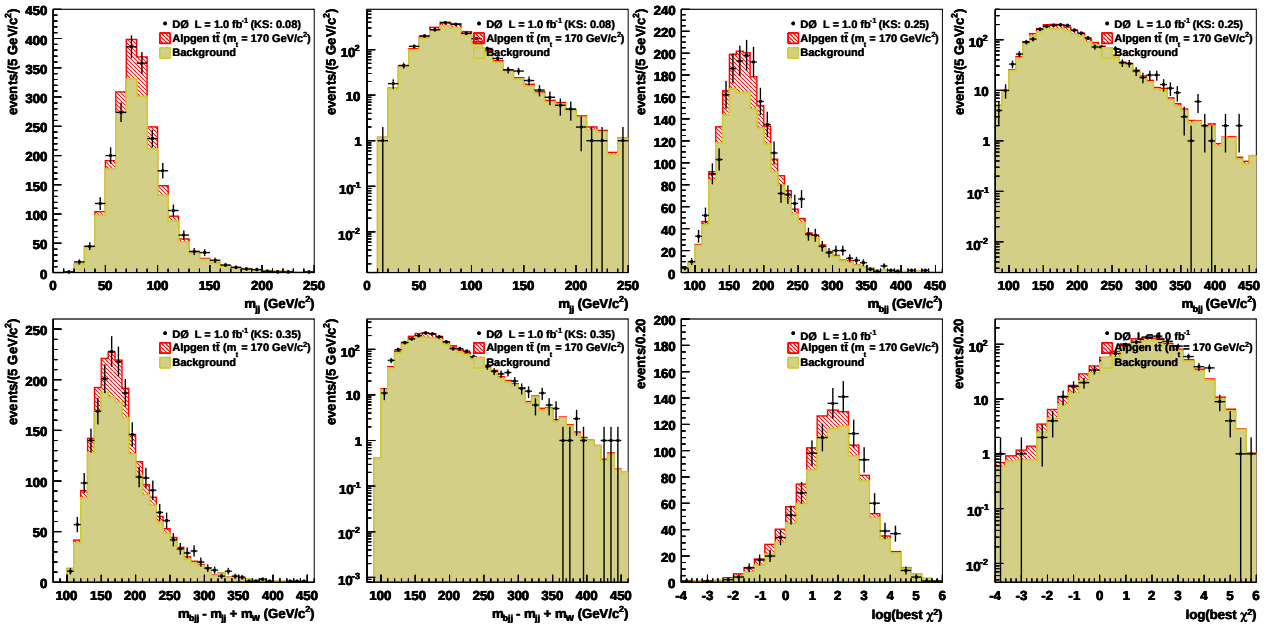


Figure B.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

B.1 Control plots after cutting on likelihood output

By cutting on the likelihood output value the relative signal-background contributions can be varied over a wide range. Additional control plots were made for two different likelihood cuts:

- selecting only events with LLH < 0.2 to obtain a background dominated sample, and
- selectin only events with LLH > 0.8 to obtain a signal-pure sample.

B.1.1 Likelihood output less than 0.2

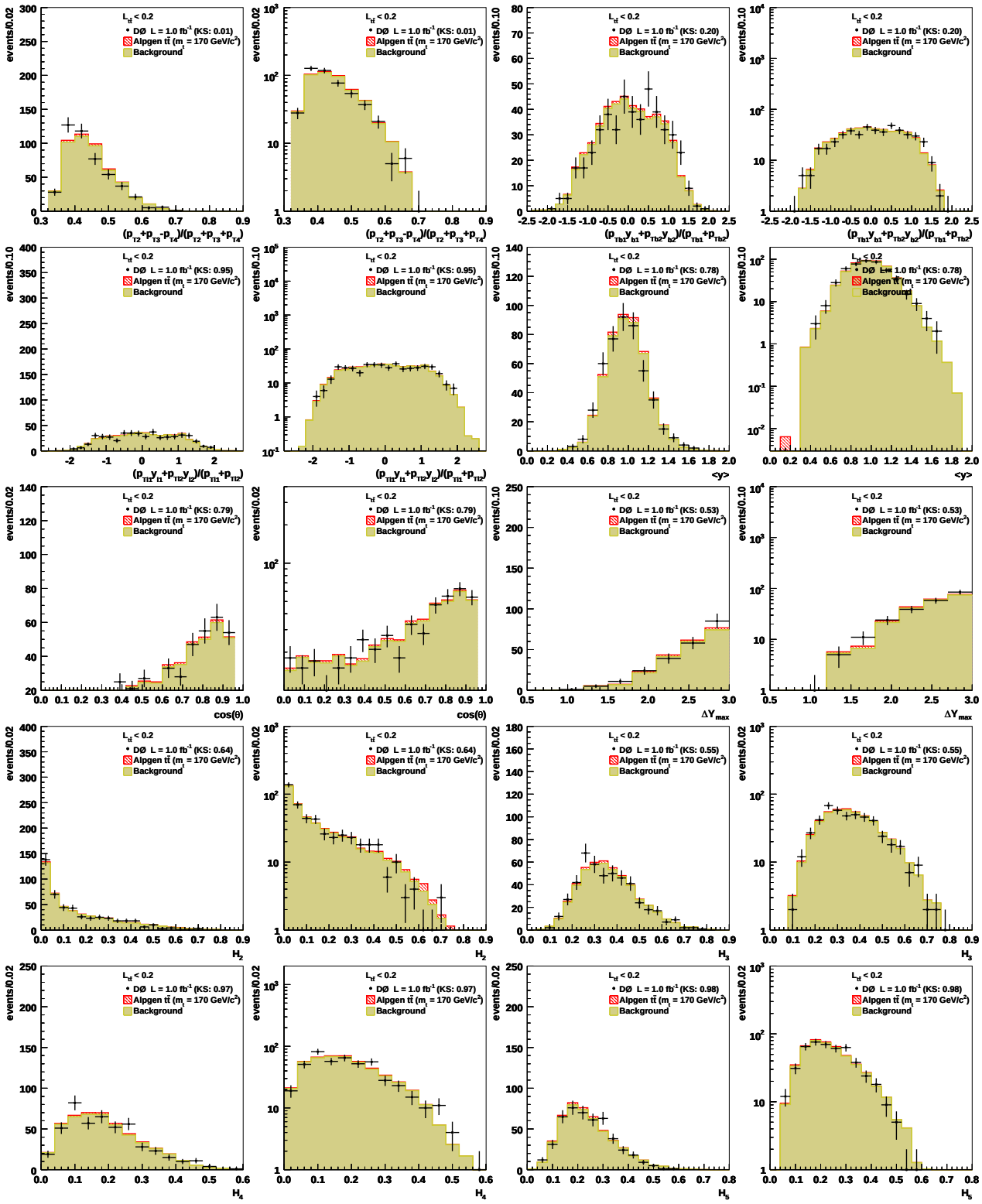


Figure B.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

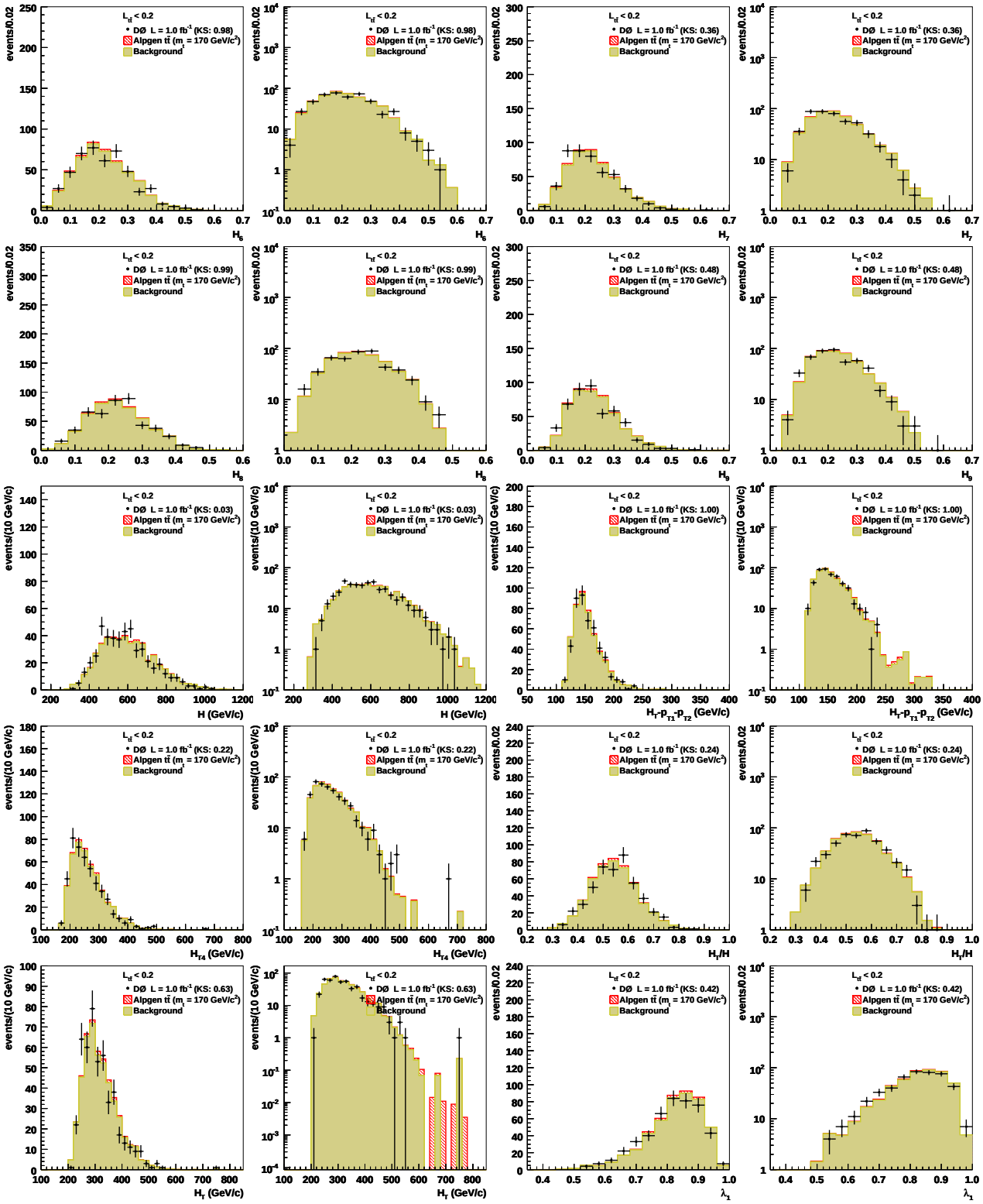


Figure B.3: Results of the likelihood analysis compared to data after cutting on $LLH < 0.2$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

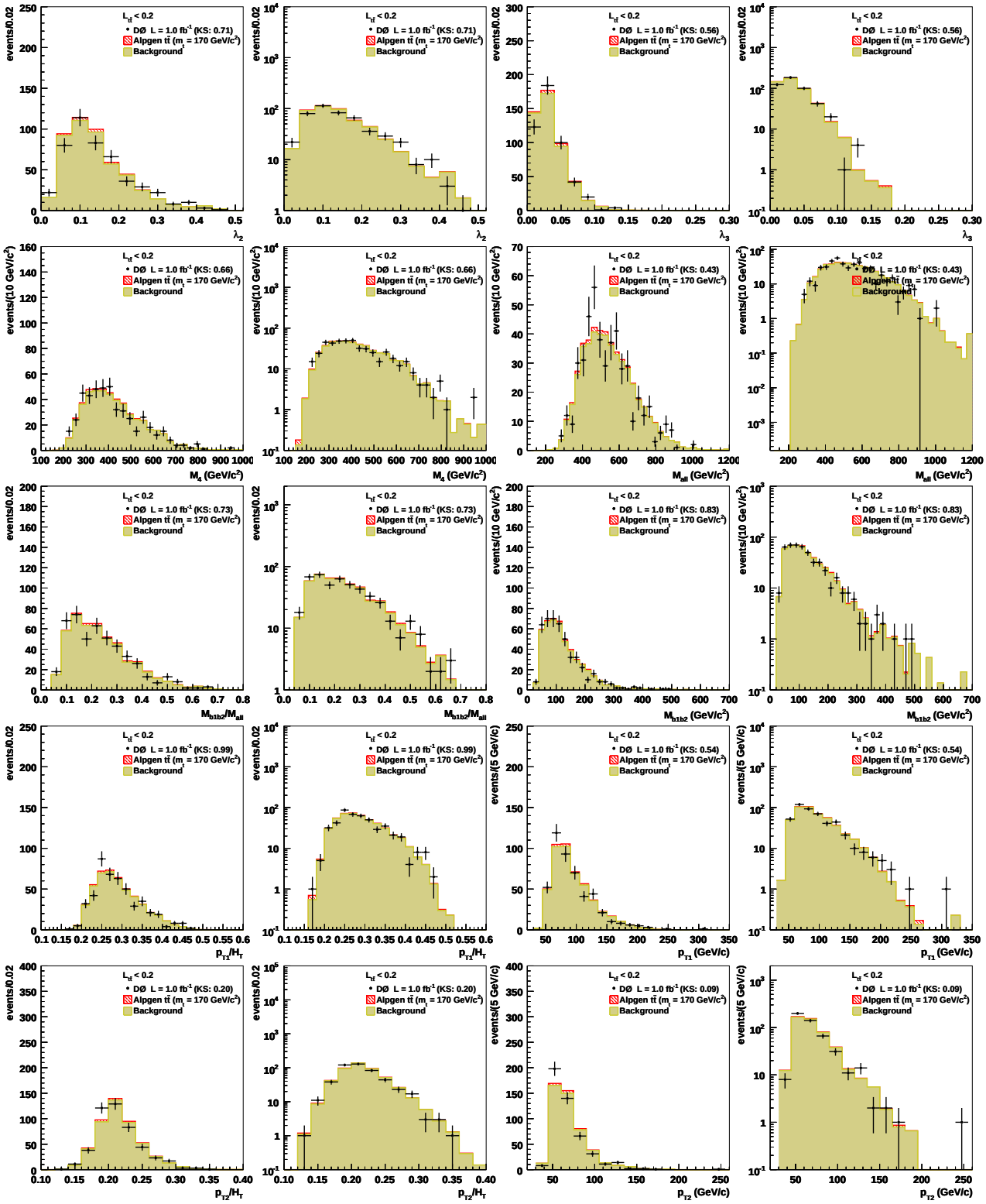


Figure B.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

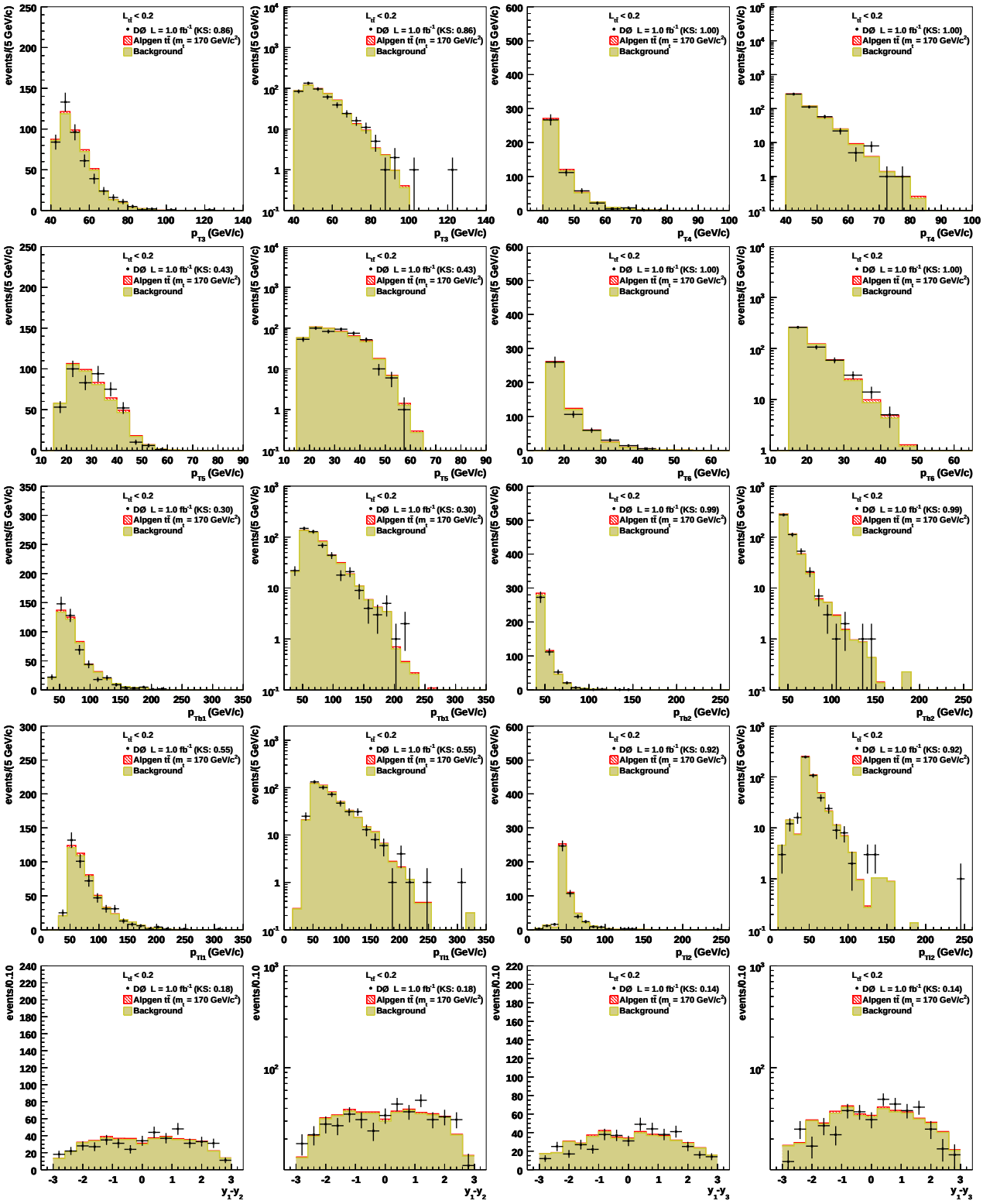


Figure B.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

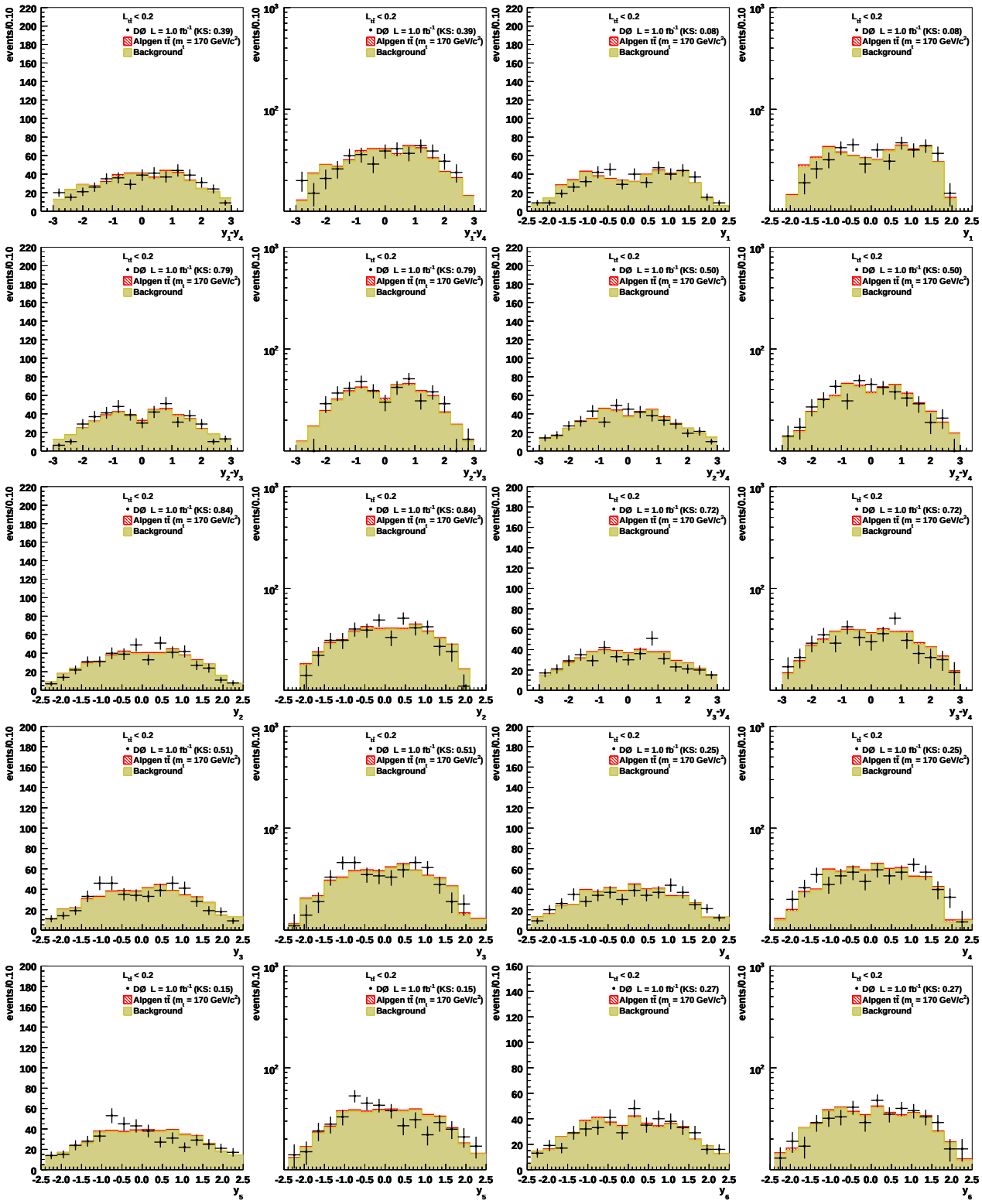


Figure B.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

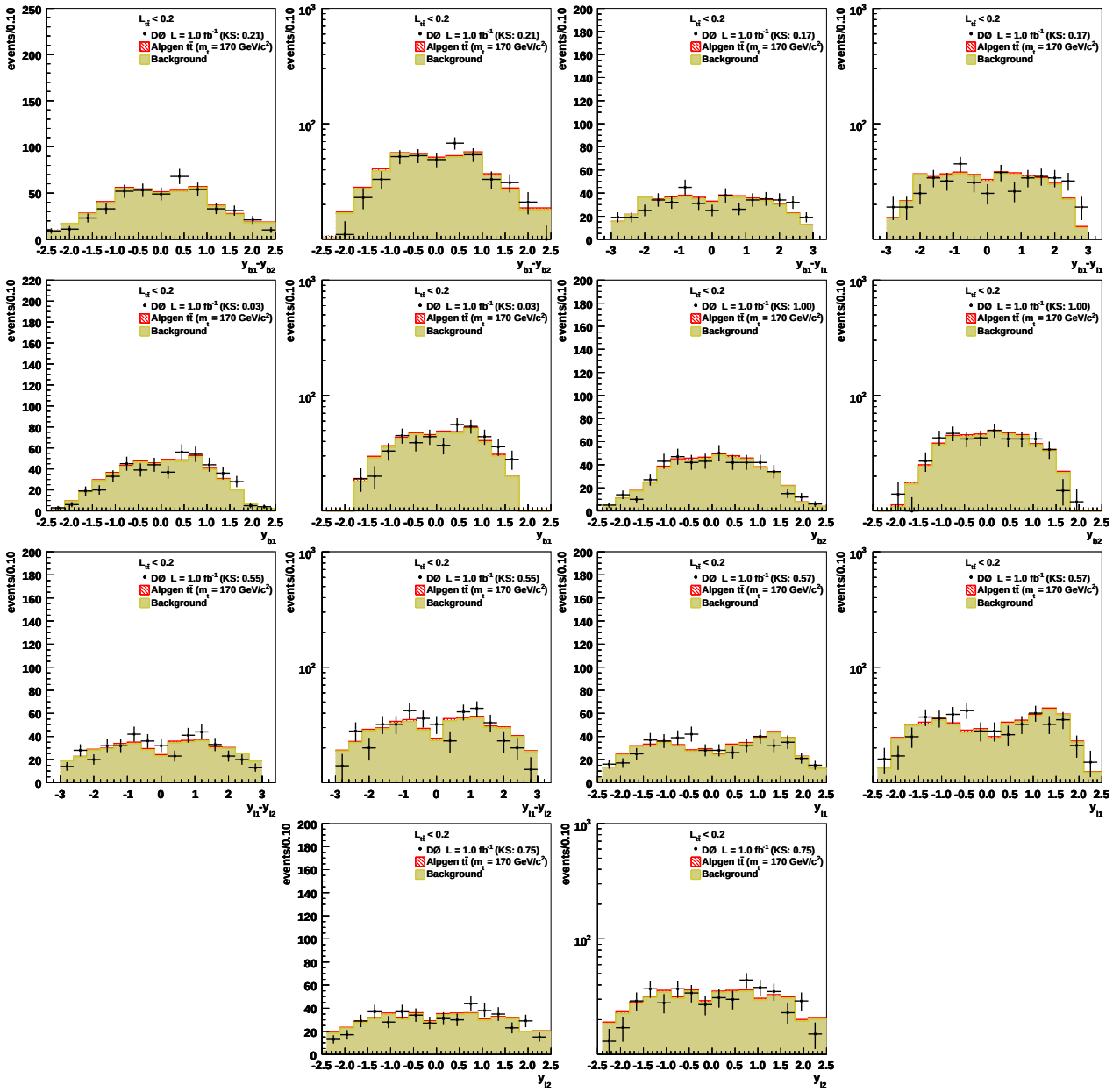


Figure B.3: Results of the likelihood analysis compared to data after cutting on $LLH < 0.2$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

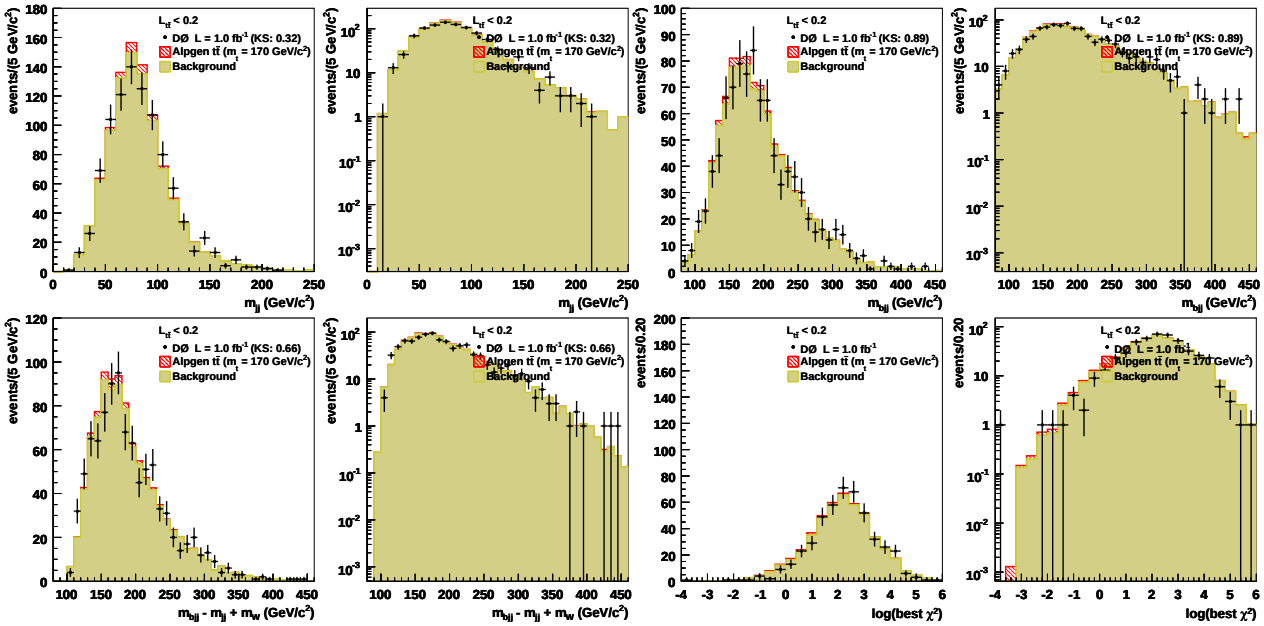


Figure B.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

B.1.2 Likelihood output more than 0.8

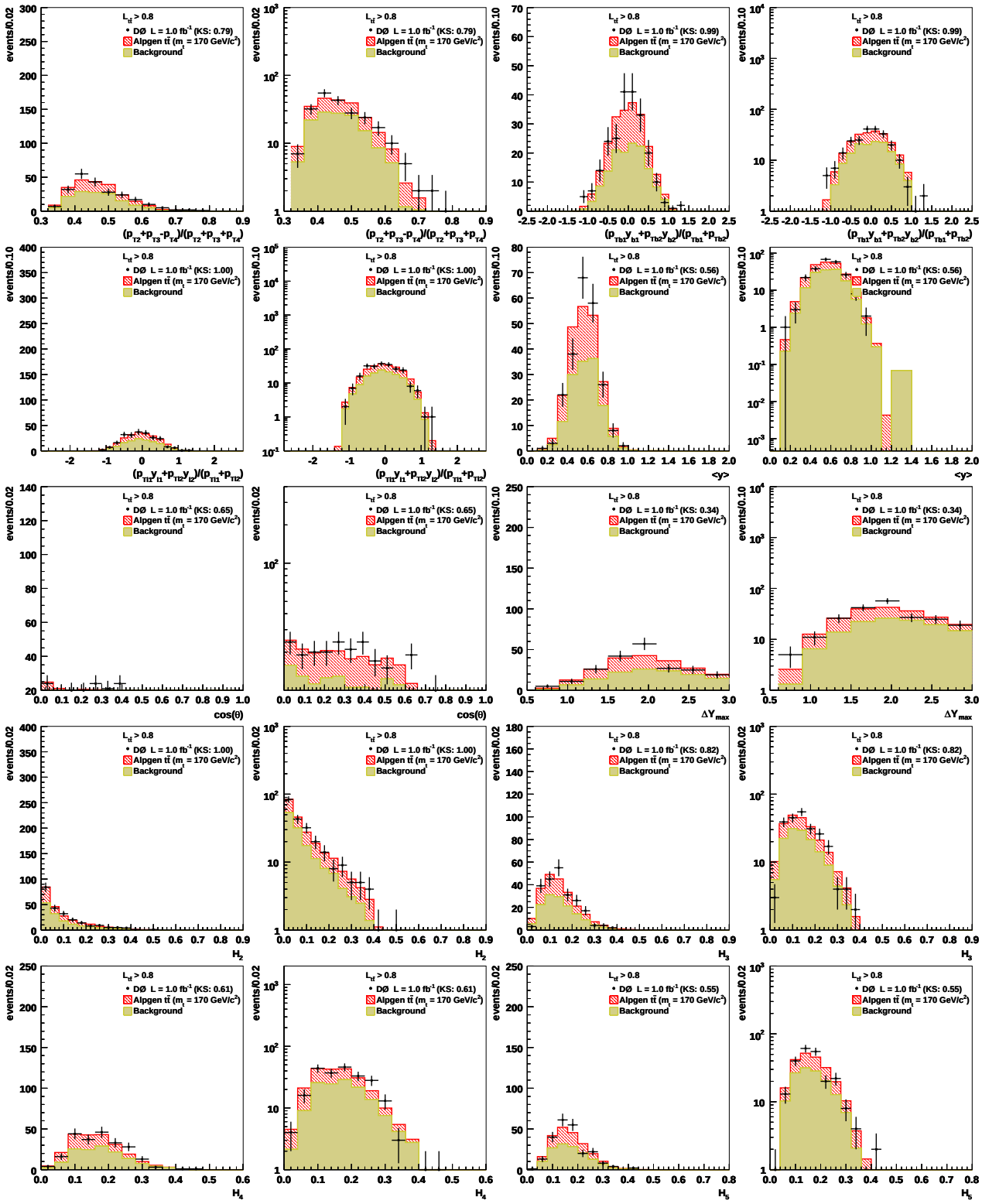


Figure B.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

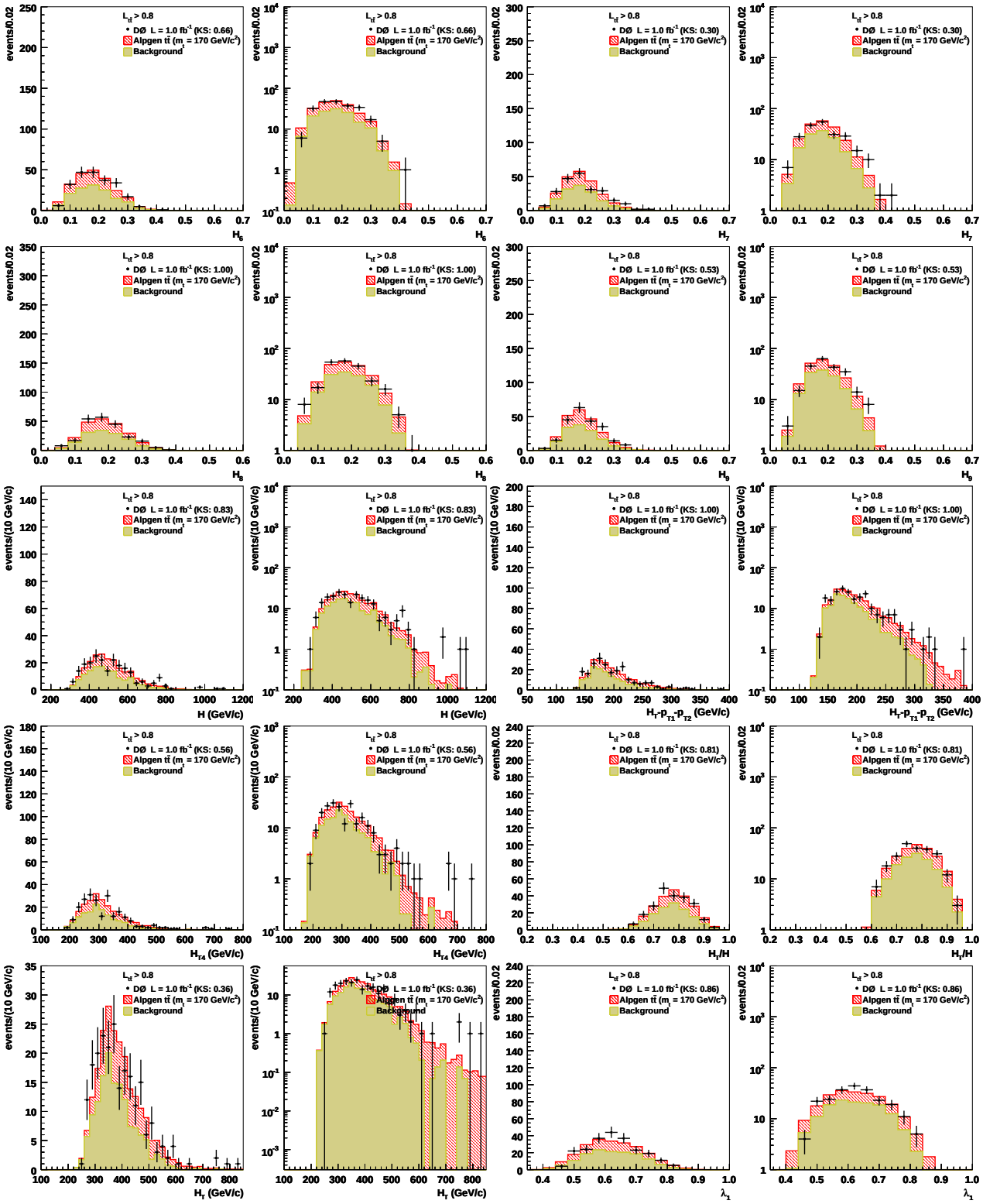


Figure B.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

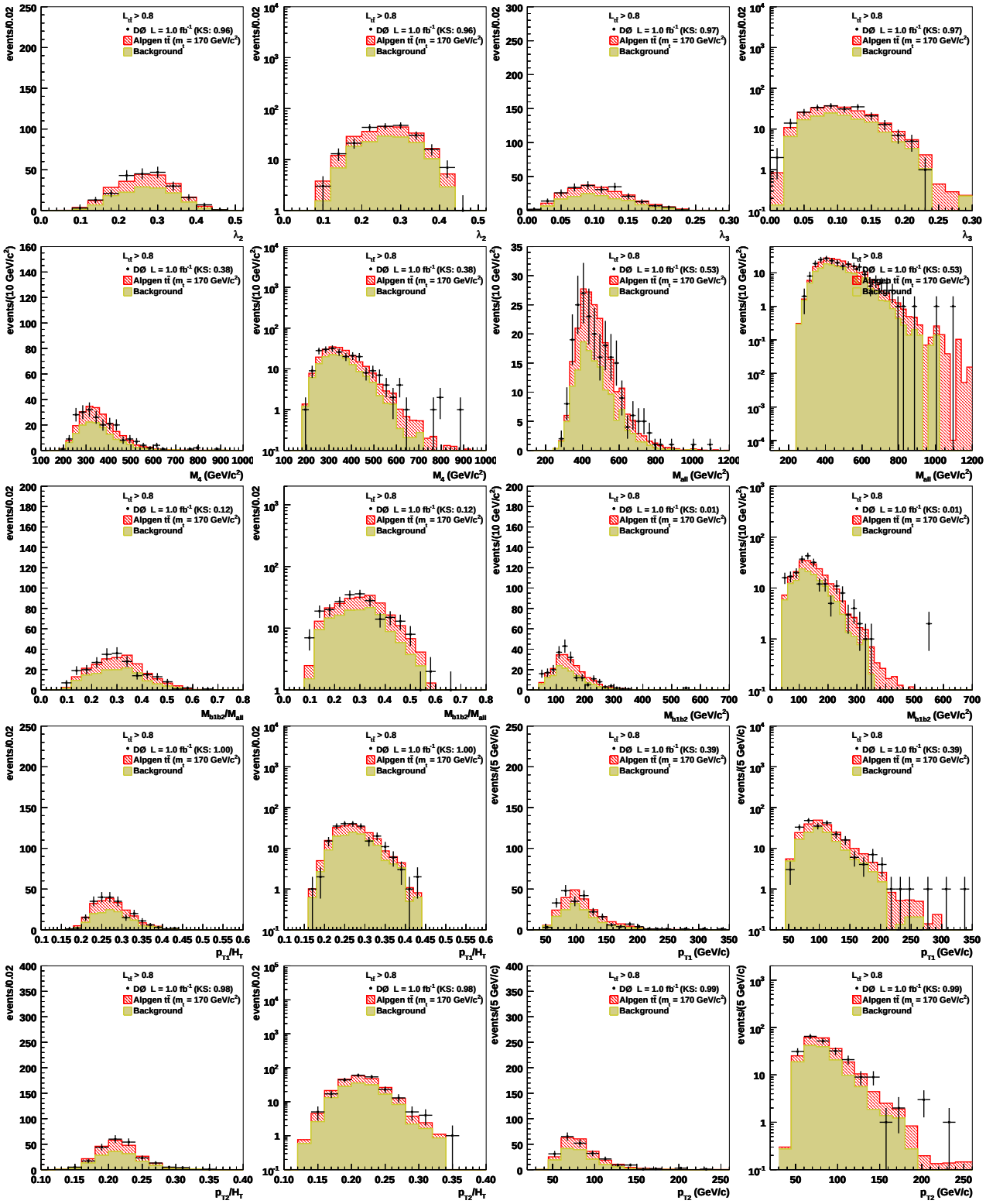


Figure B.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

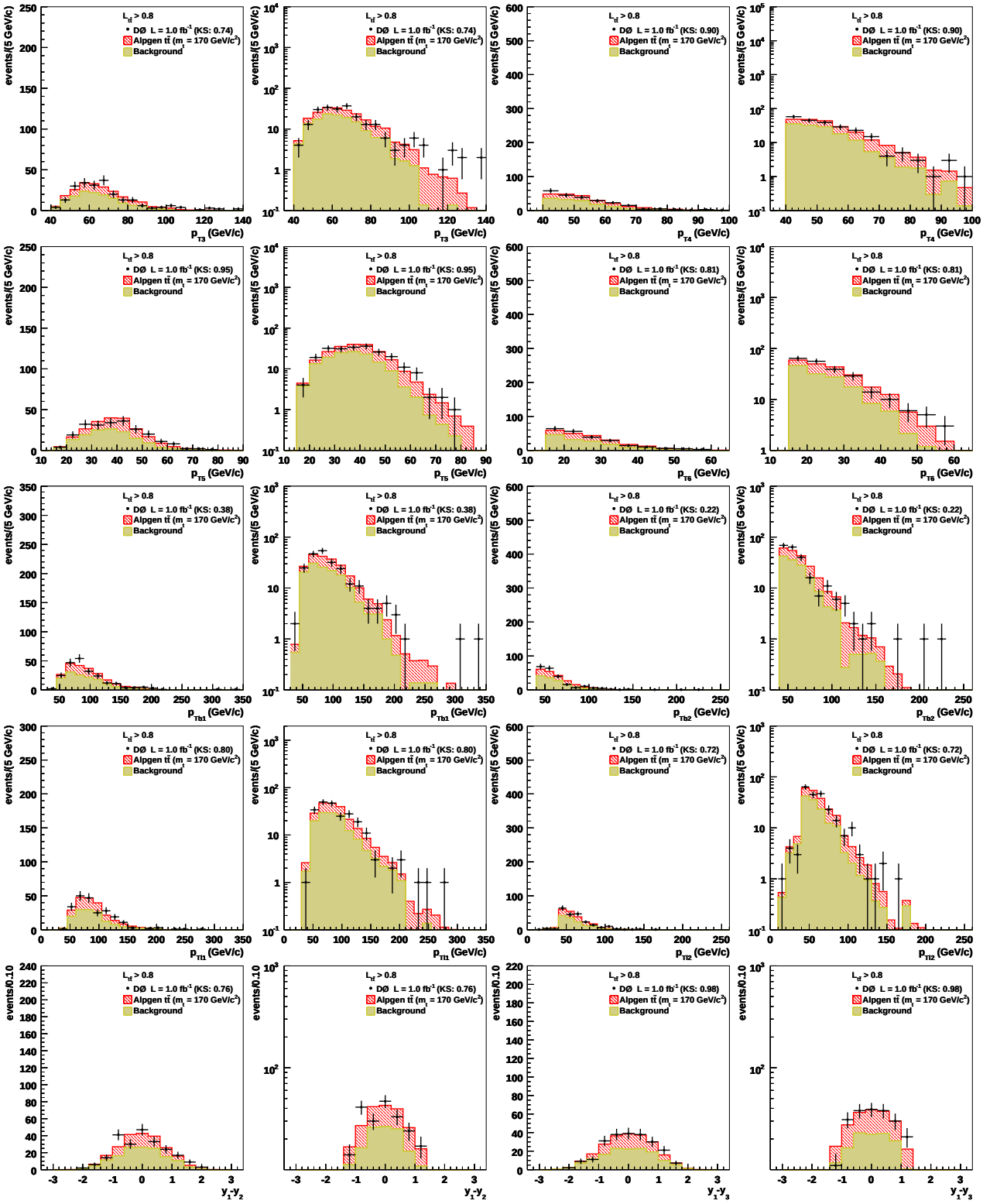


Figure B.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

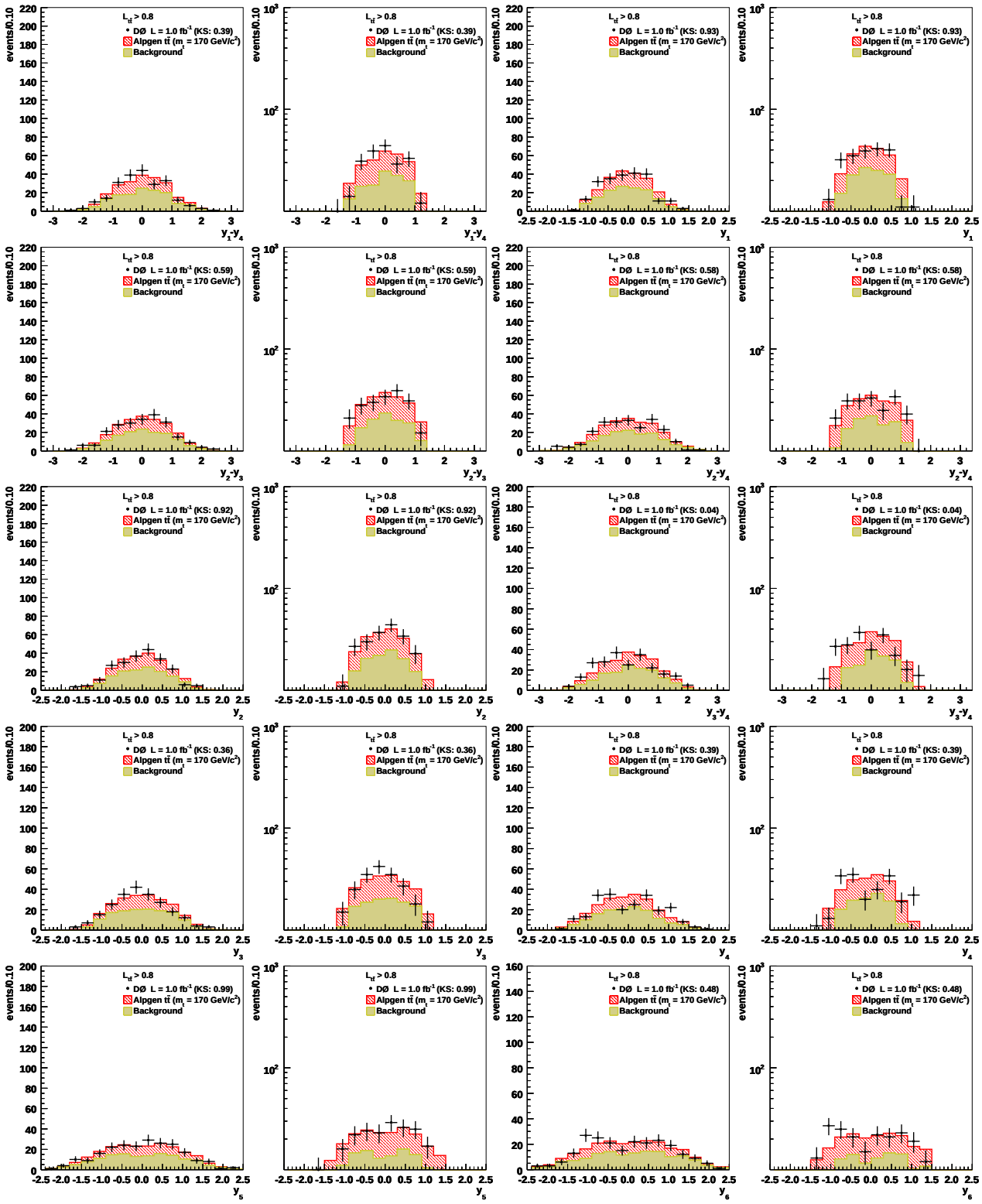


Figure B.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

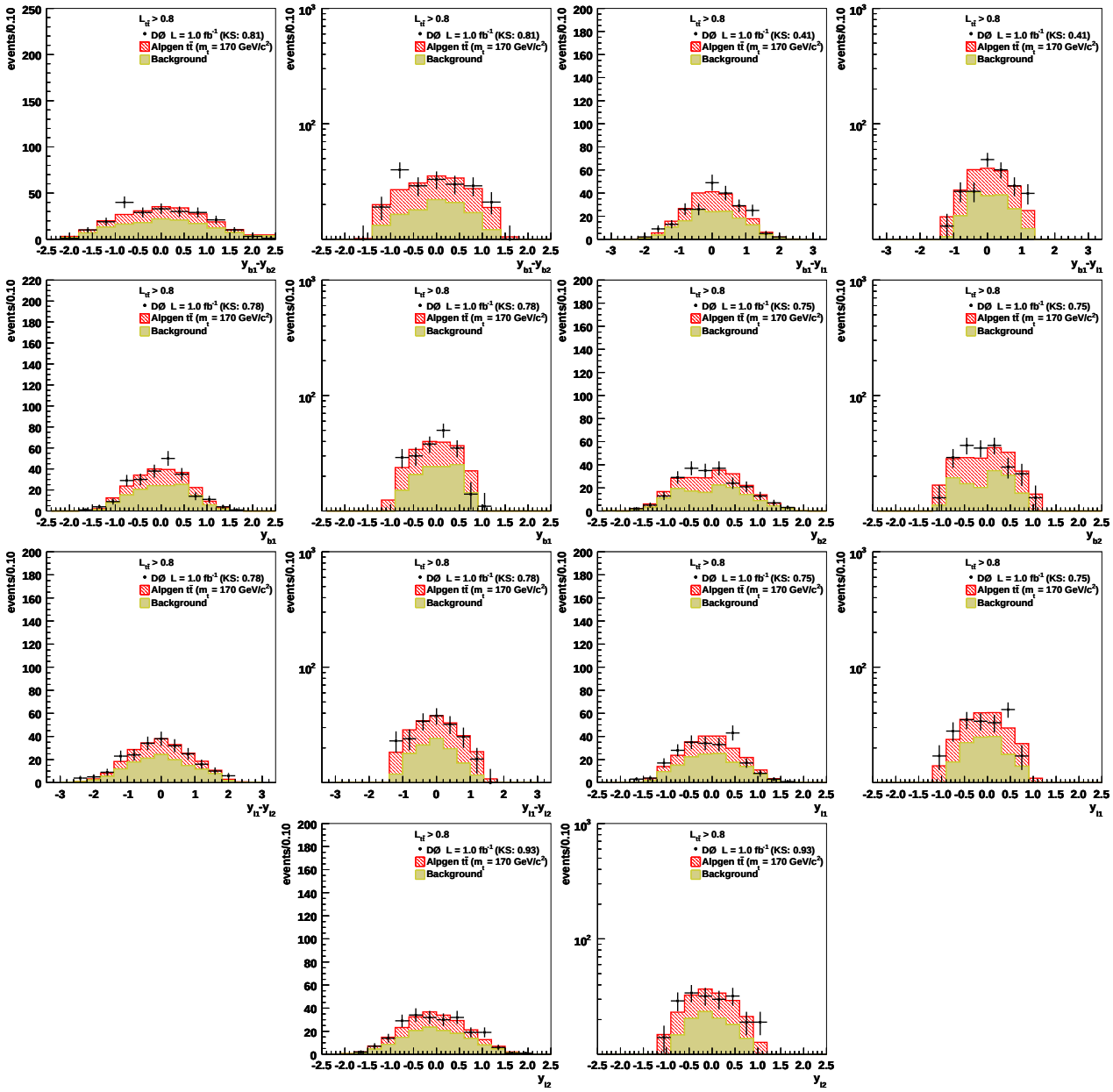


Figure B.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

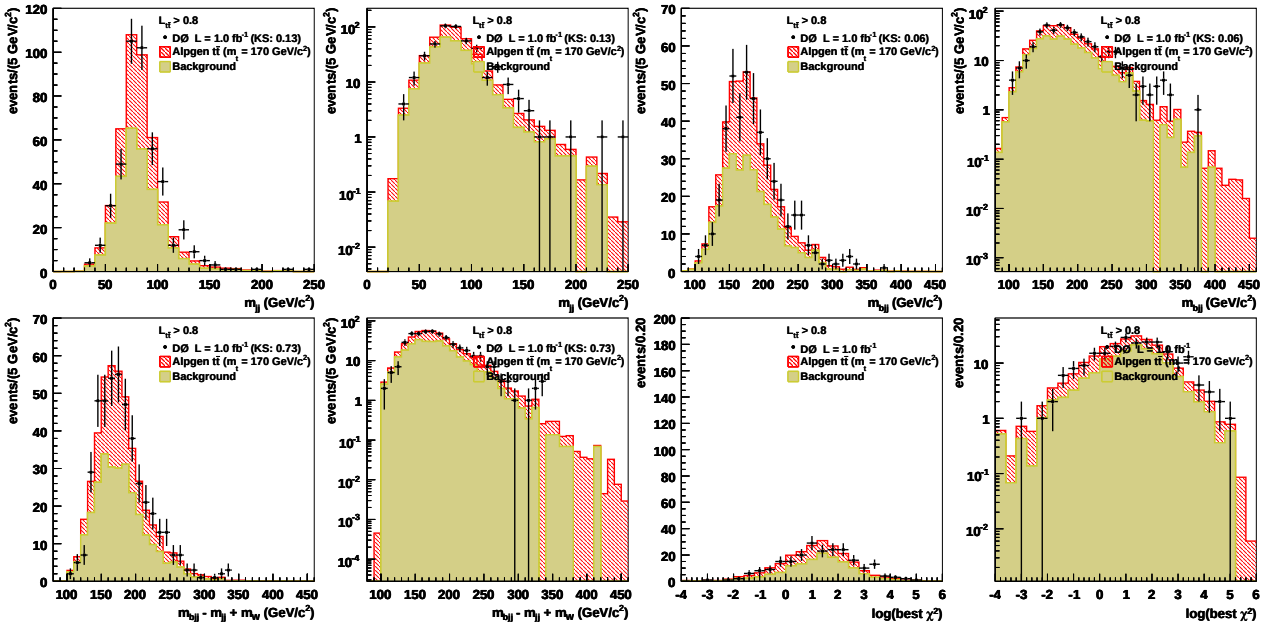


Figure B.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 170 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

Appendix C

Figures for $m_{\mathbf{t}} = 175$ GeV/c^2

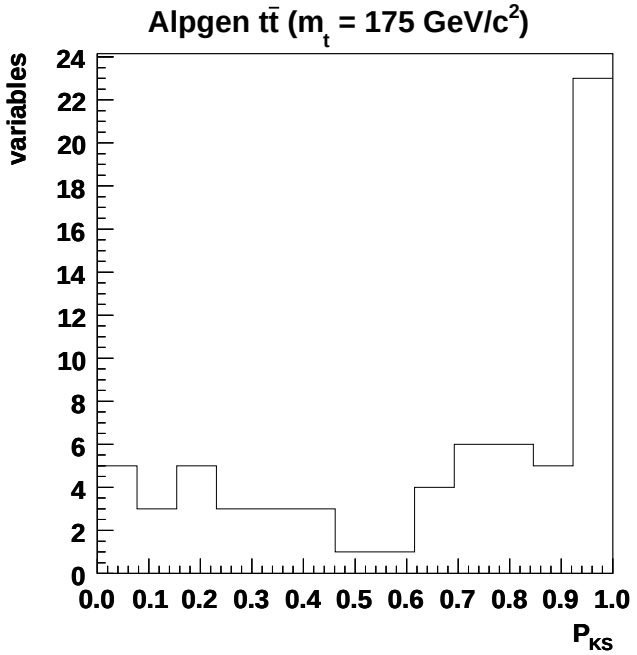


Figure C.1: Kolmogorov-Smirnov test results distribution corresponding to the variables in figure C.4

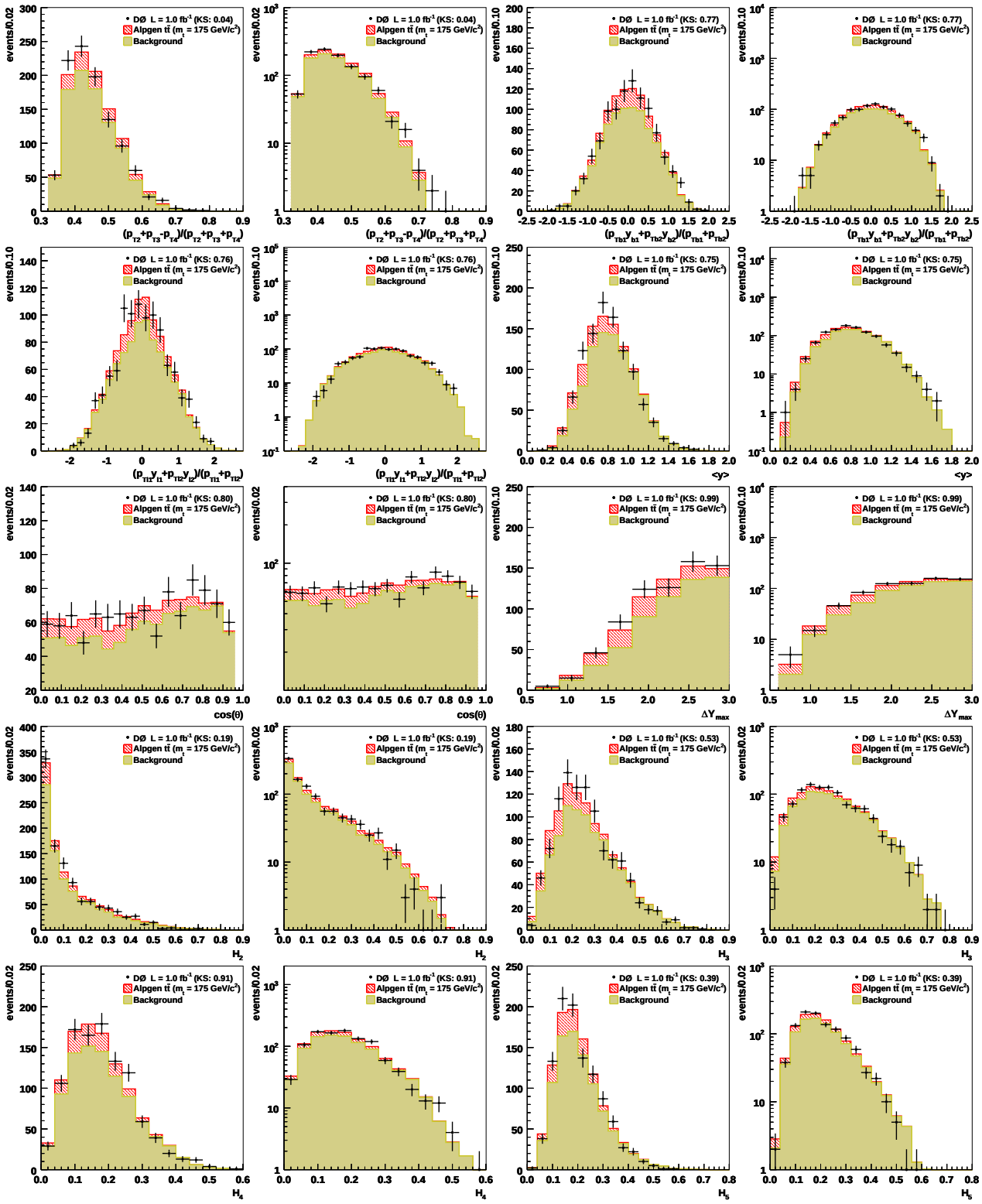


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

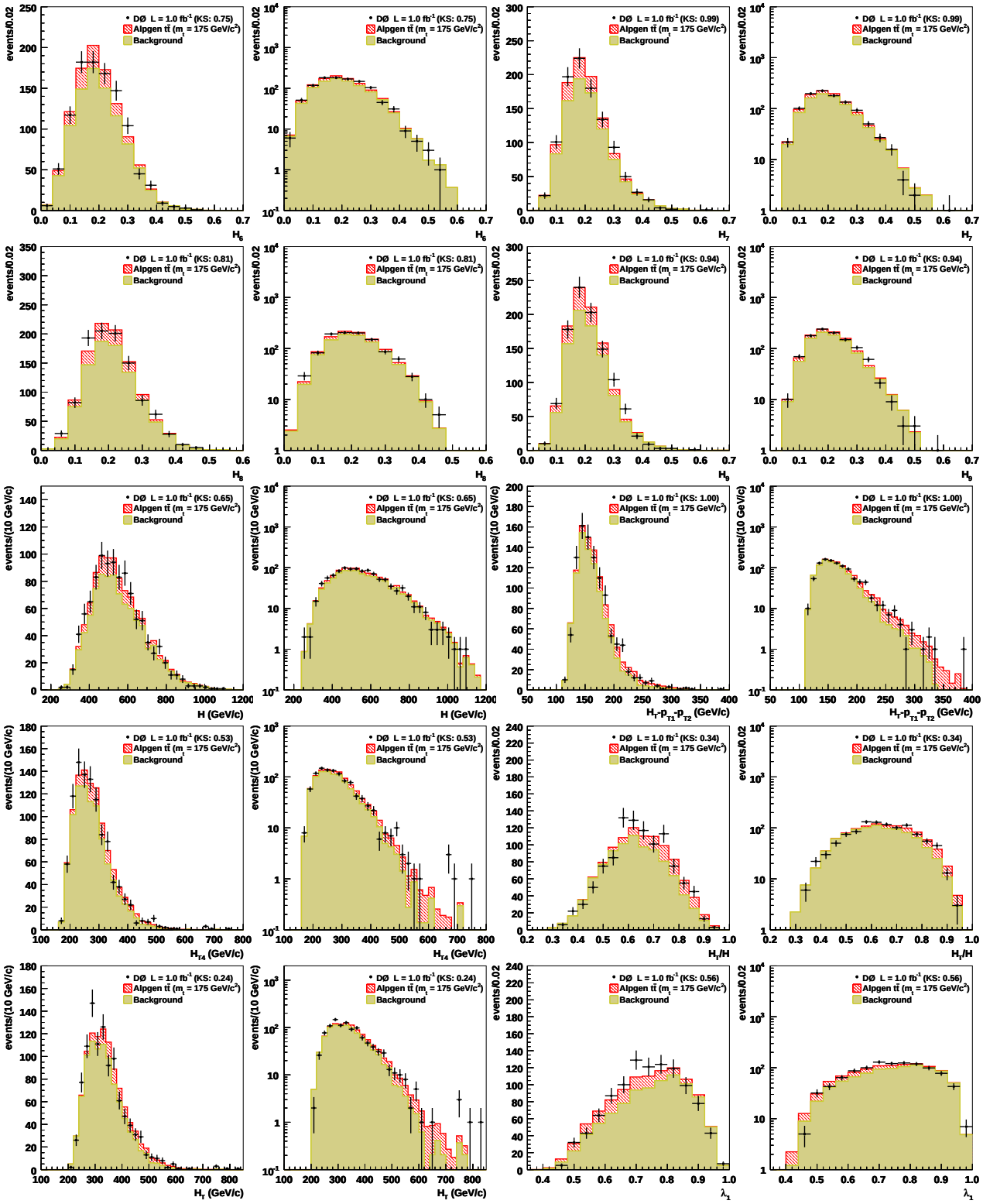


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

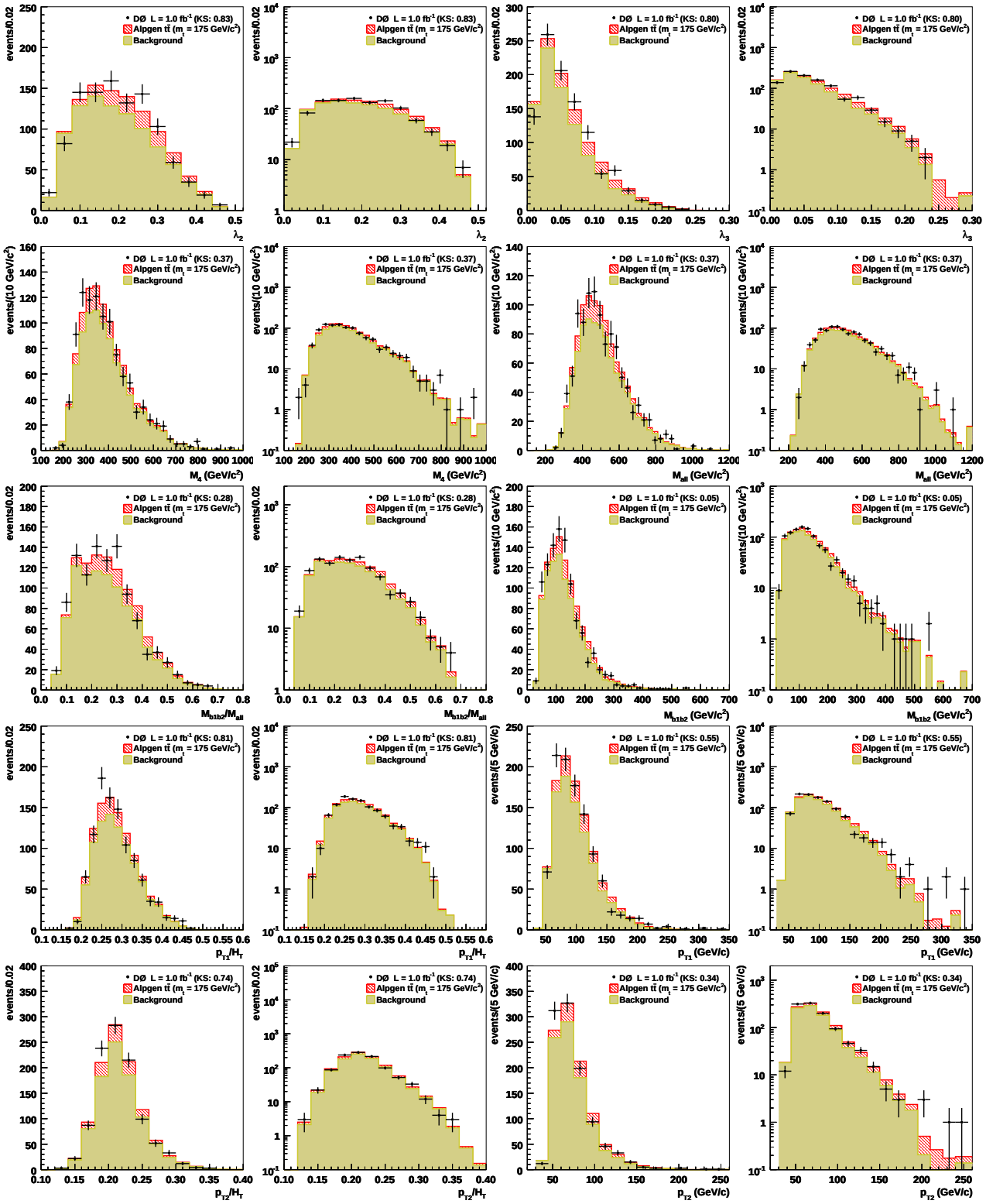


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

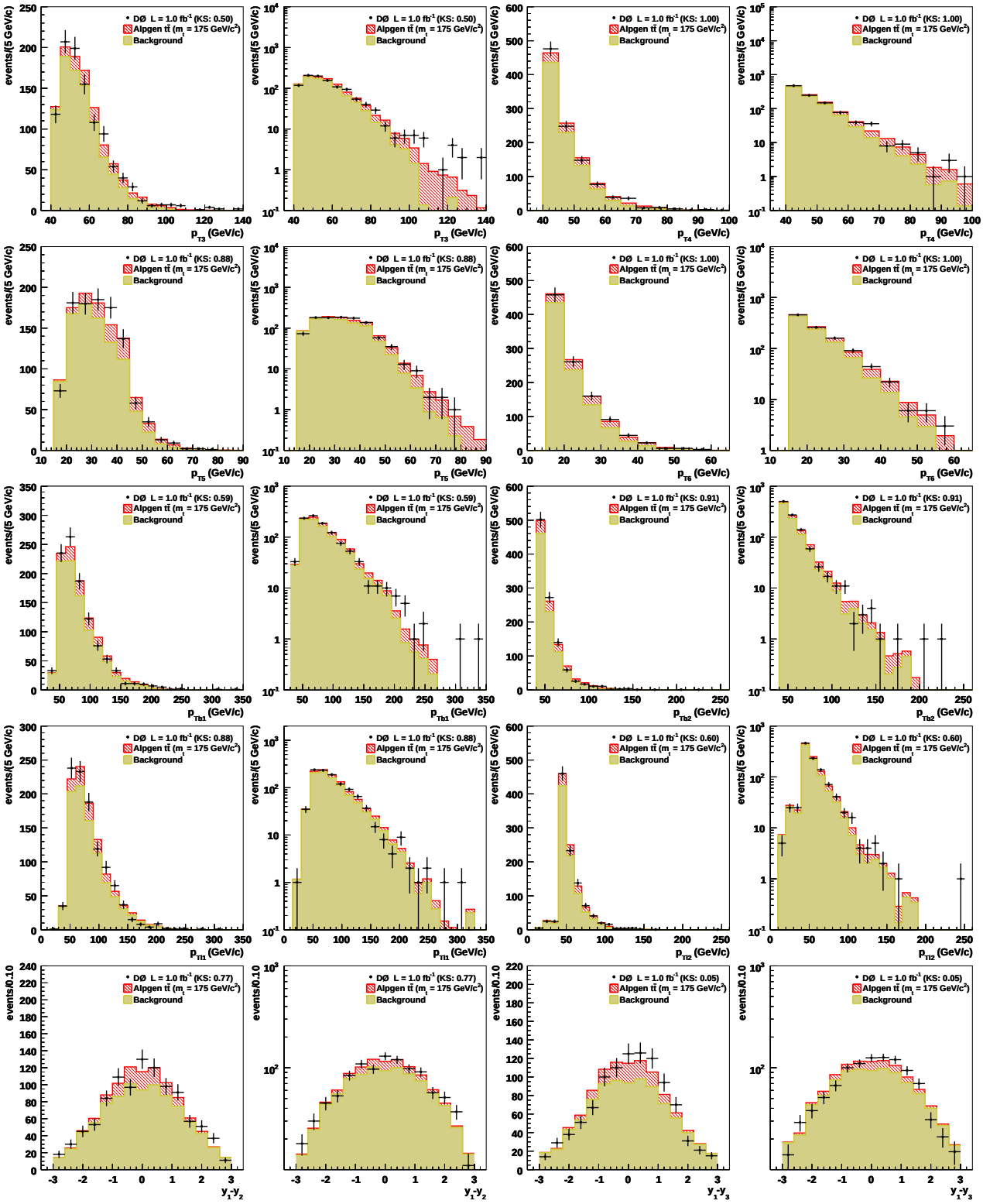


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

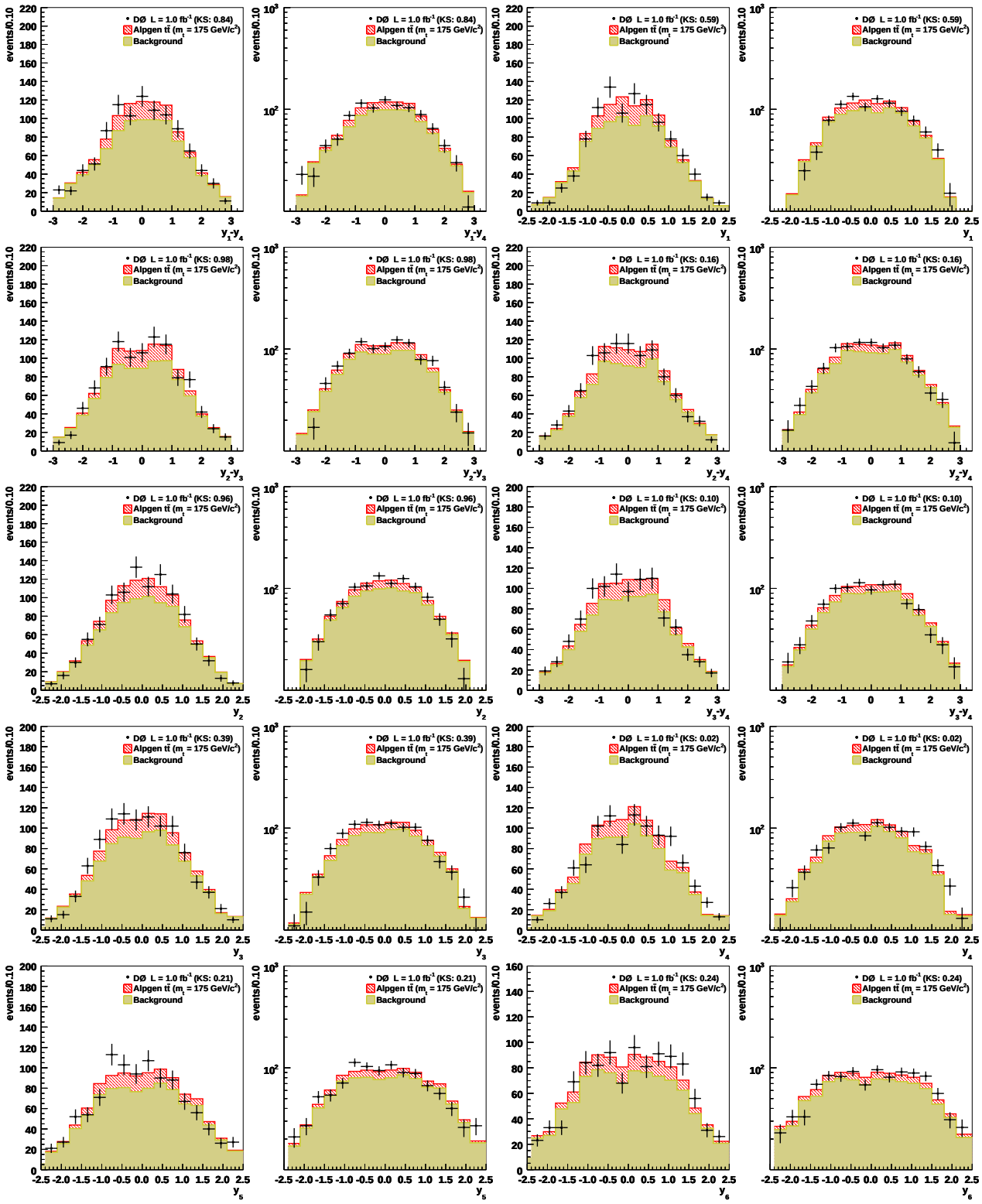


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

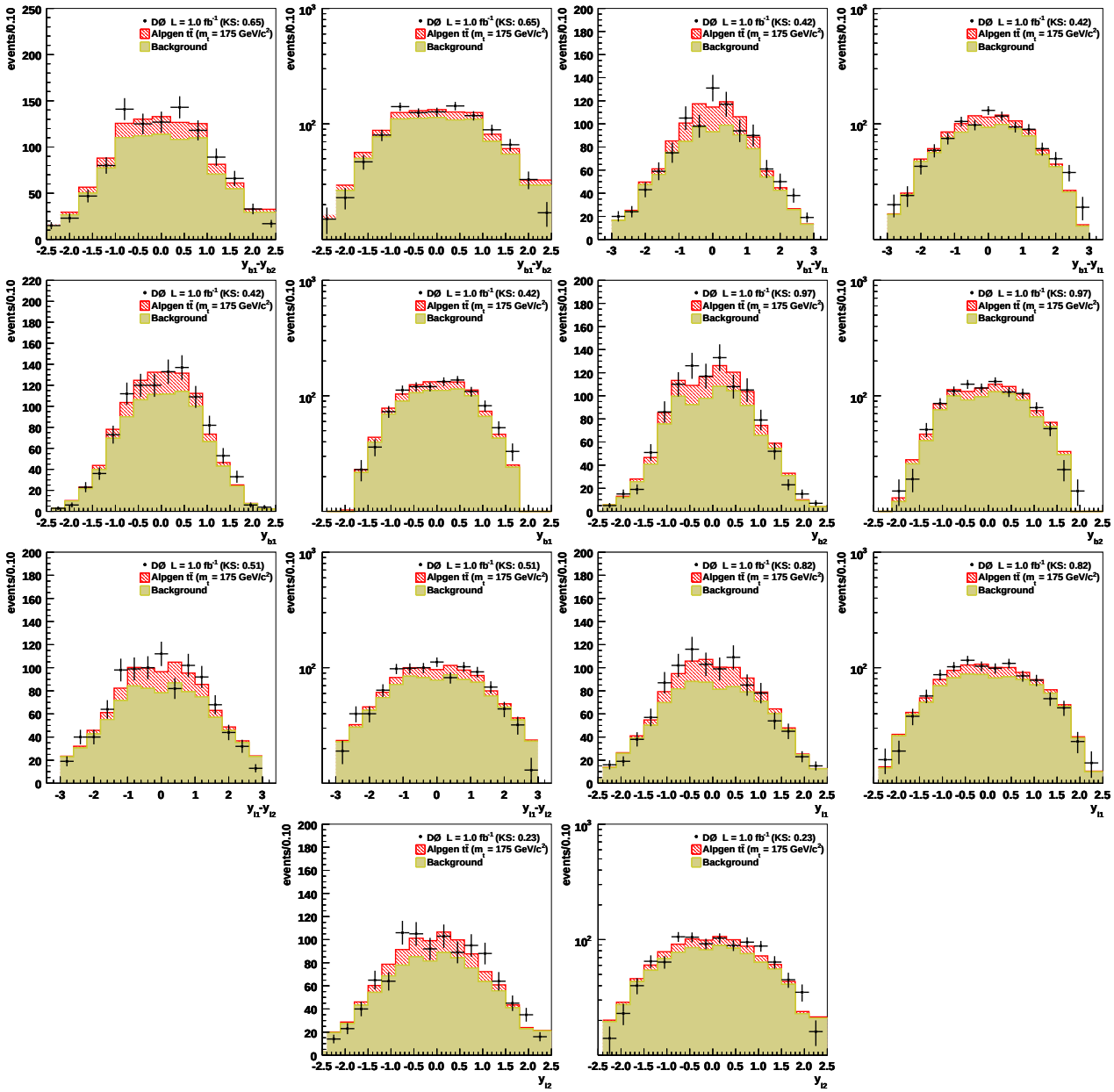


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

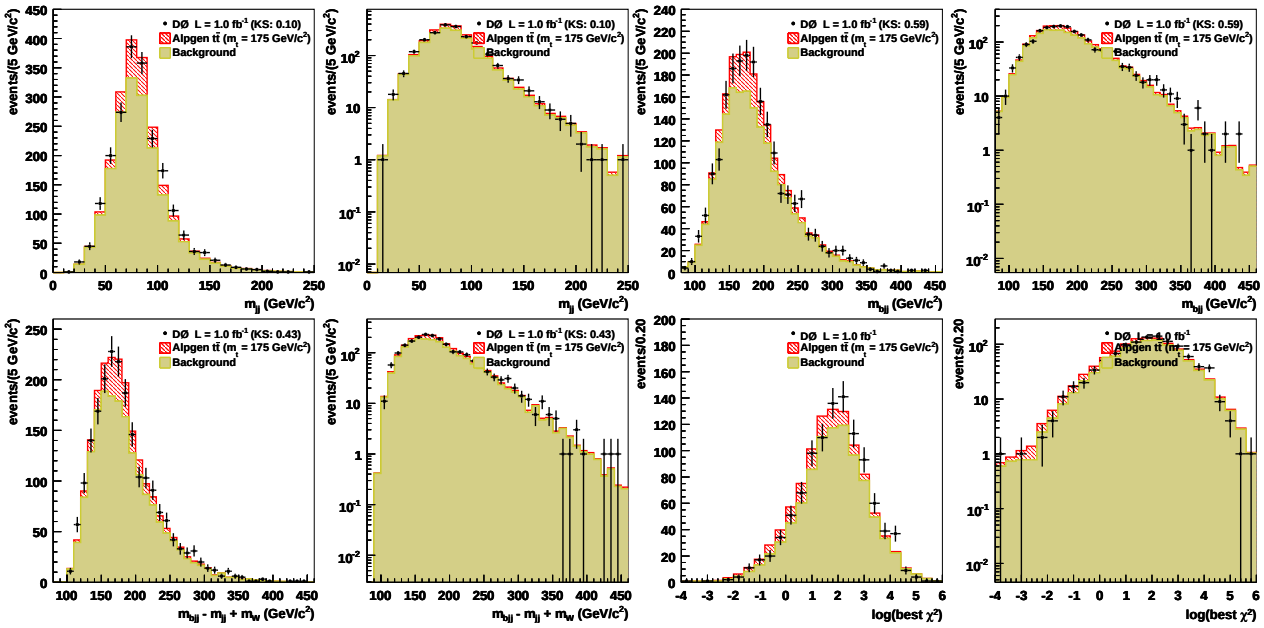


Figure C.2: Results of the likelihood analysis compared to data. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

C.1 Control plots after cutting on likelihood output

C.1.1 Likelihood output less than 0.2

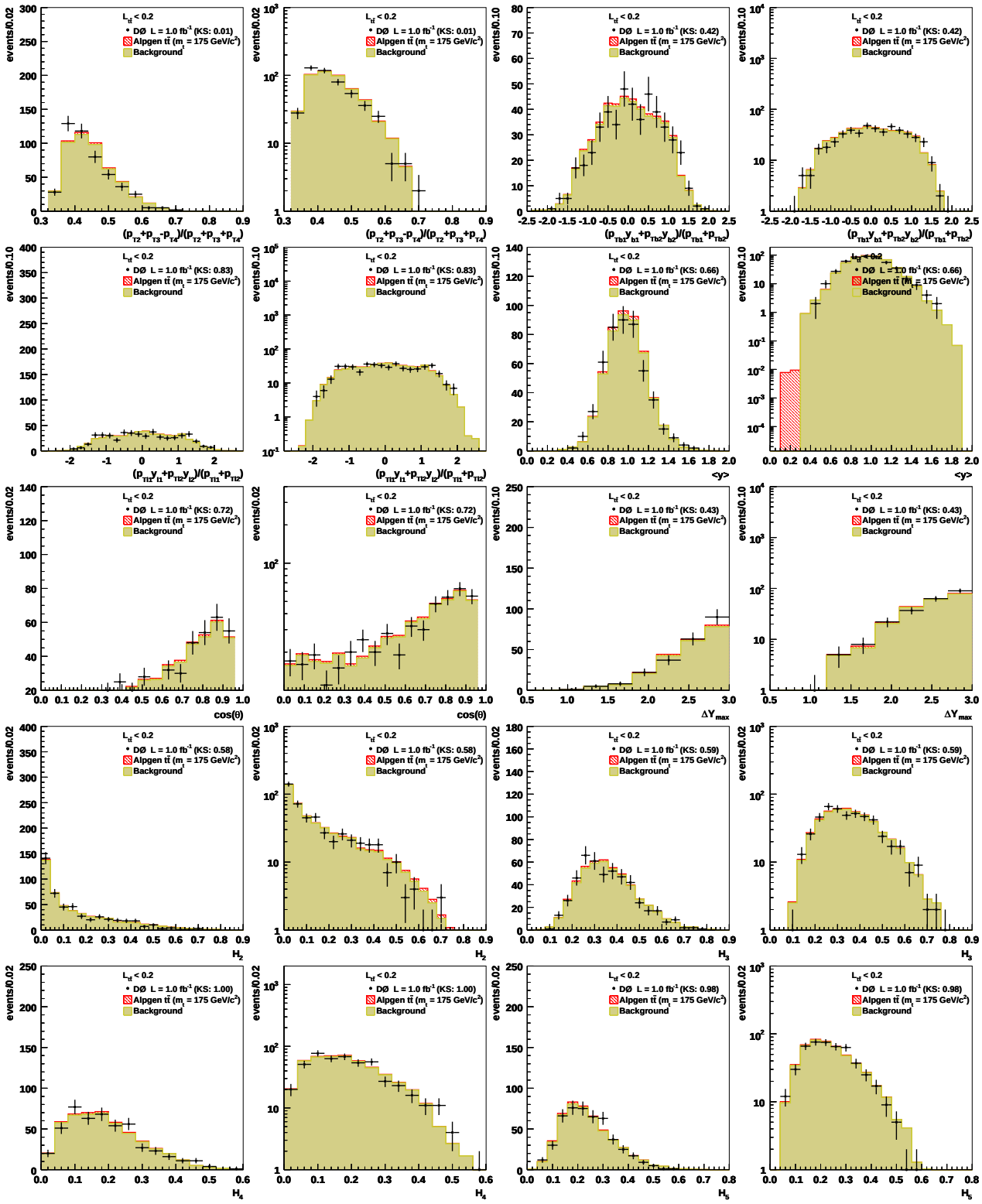


Figure C.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

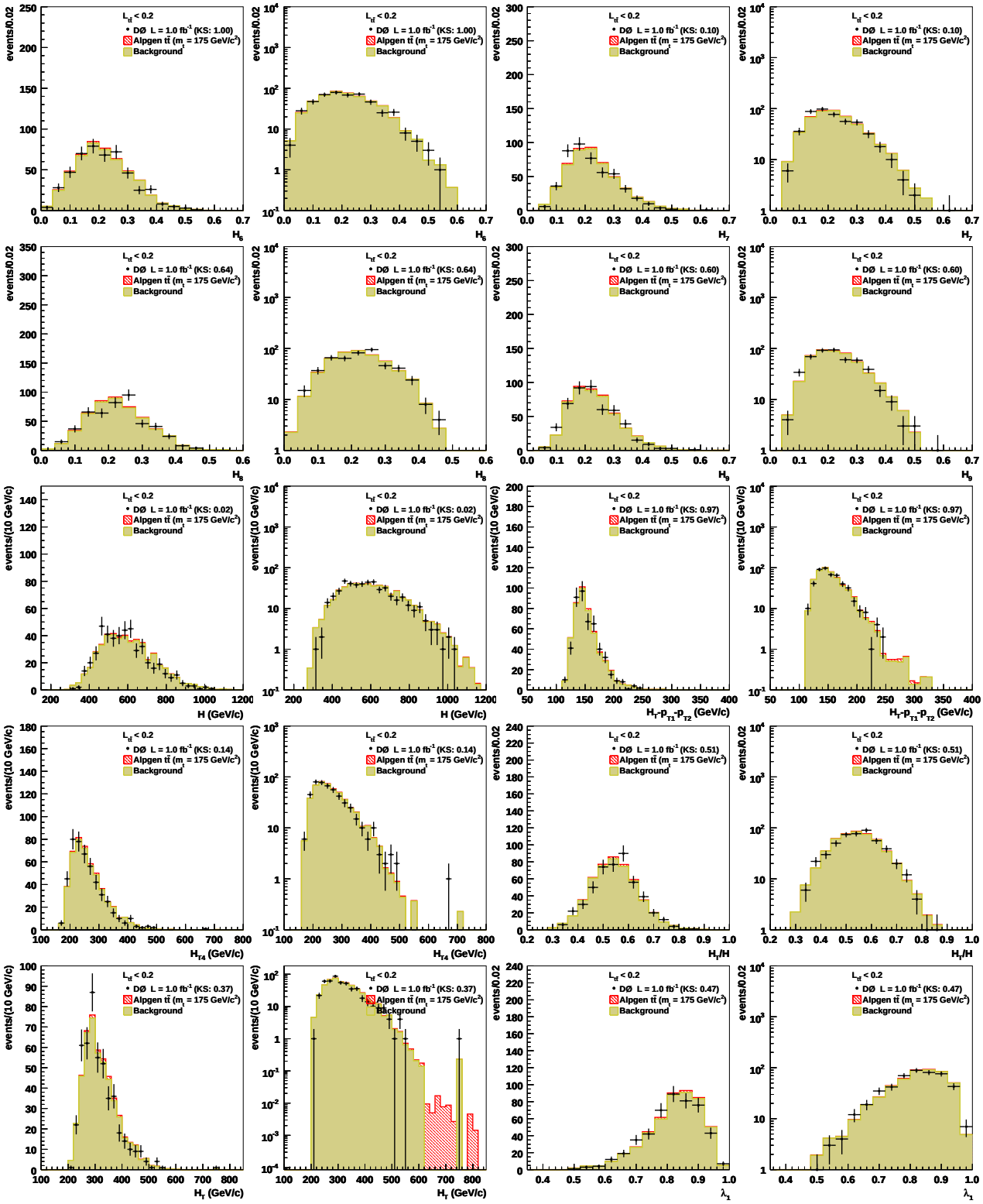


Figure C.3: Results of the likelihood analysis compared to data after cutting on $LLH < 0.2$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

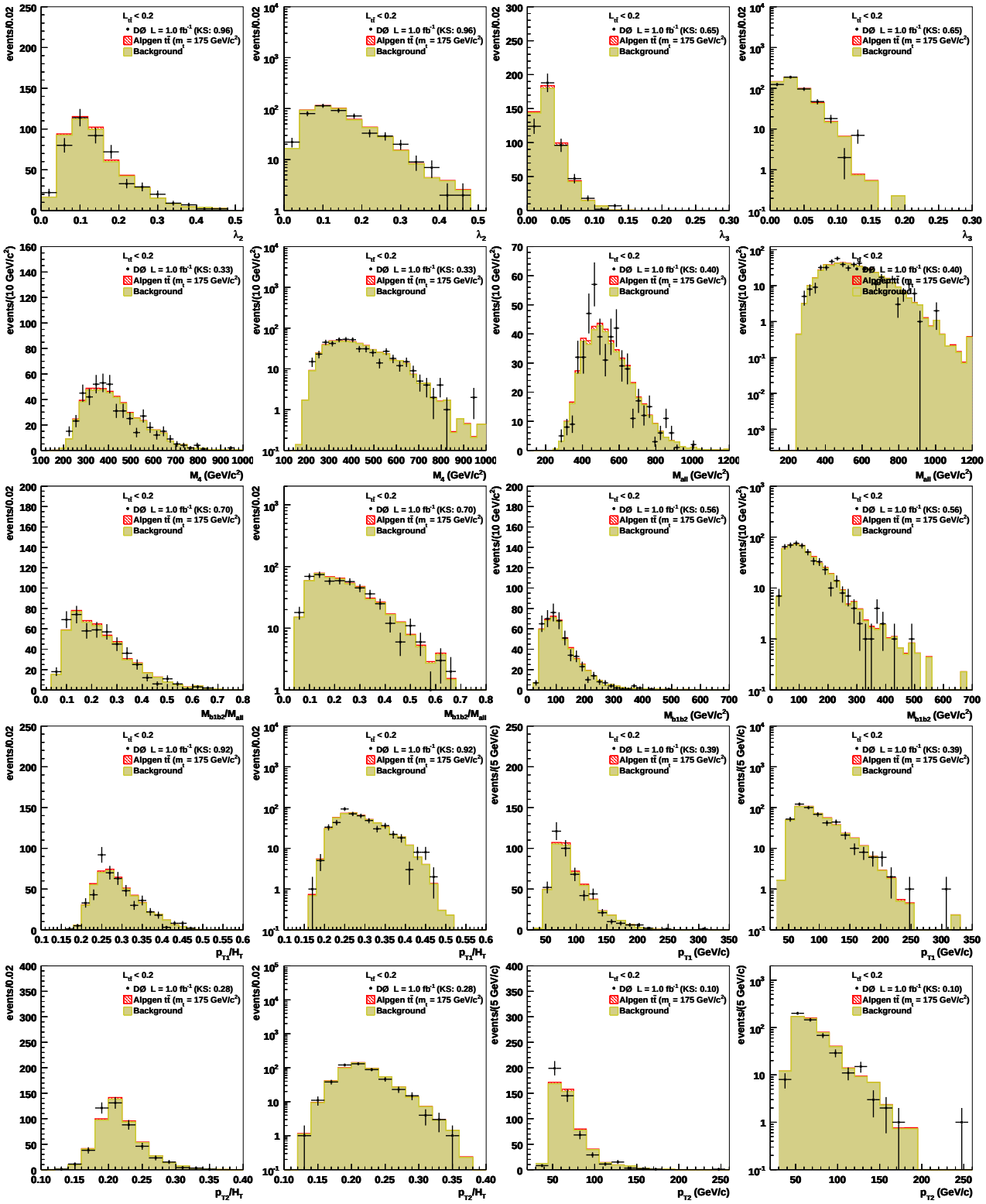


Figure C.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

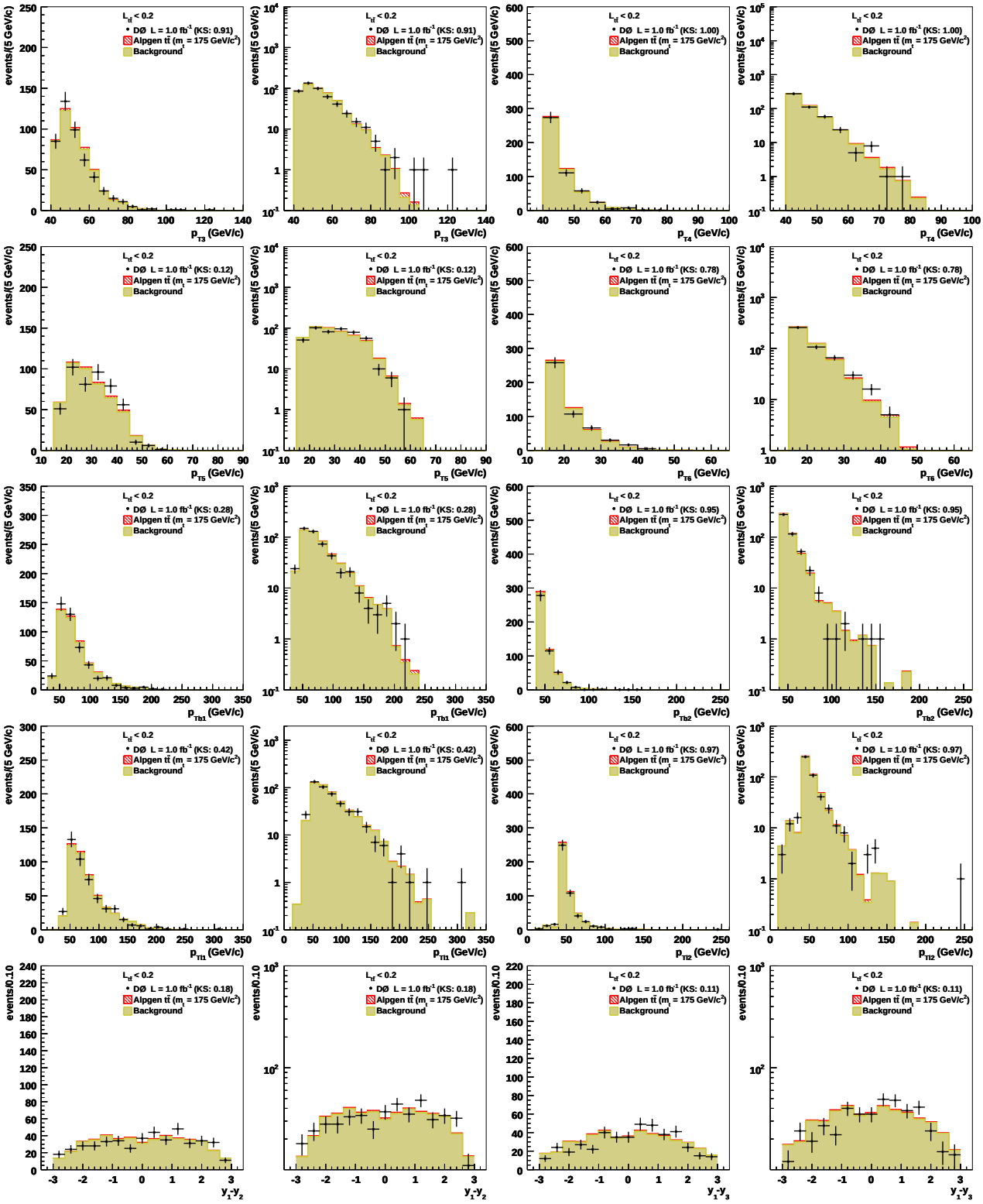


Figure C.3: Results of the likelihood analysis compared to data after cutting on $LLH < 0.2$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

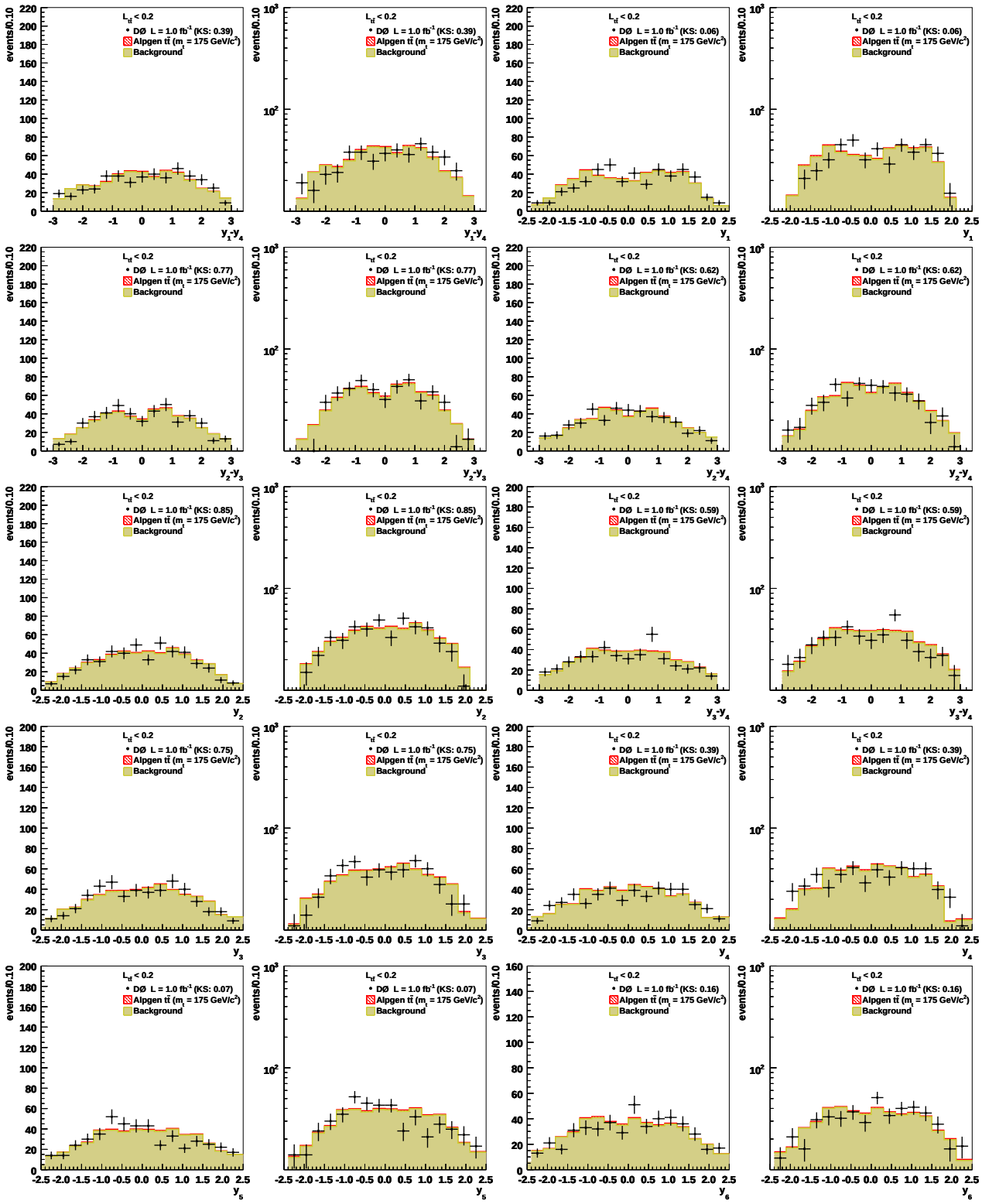


Figure C.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

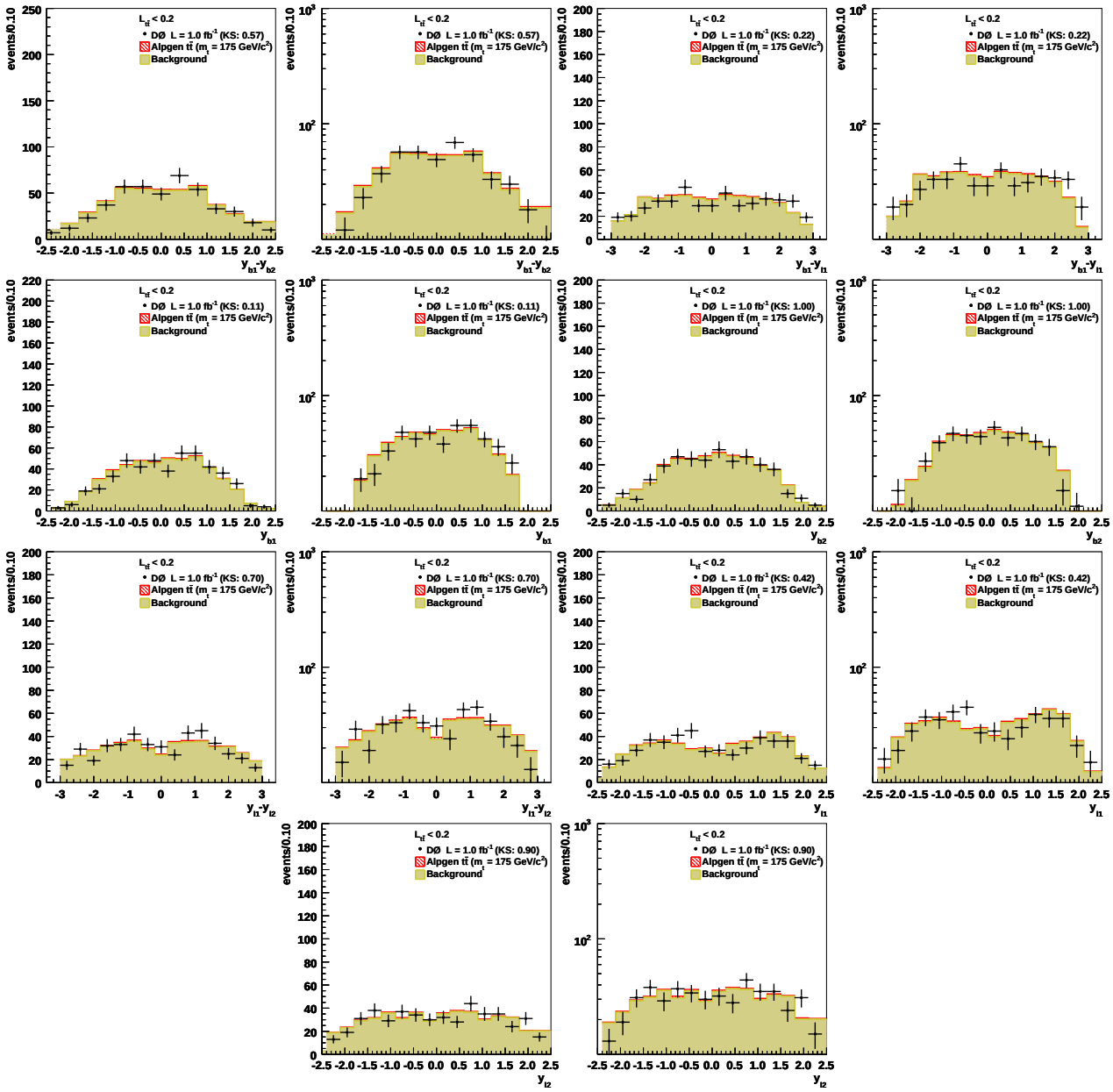


Figure C.3: Results of the likelihood analysis compared to data after cutting on $LLH < 0.2$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

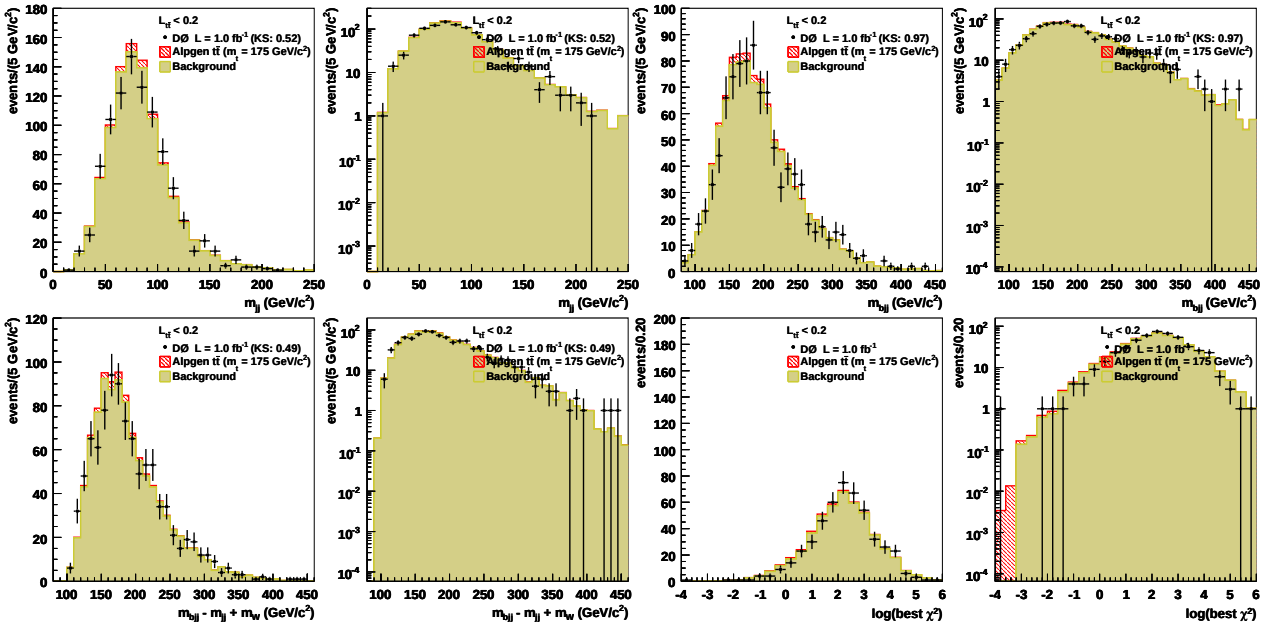


Figure C.3: Results of the likelihood analysis compared to data after cutting on LLH < 0.2. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

C.1.2 Likelihood output more than 0.8

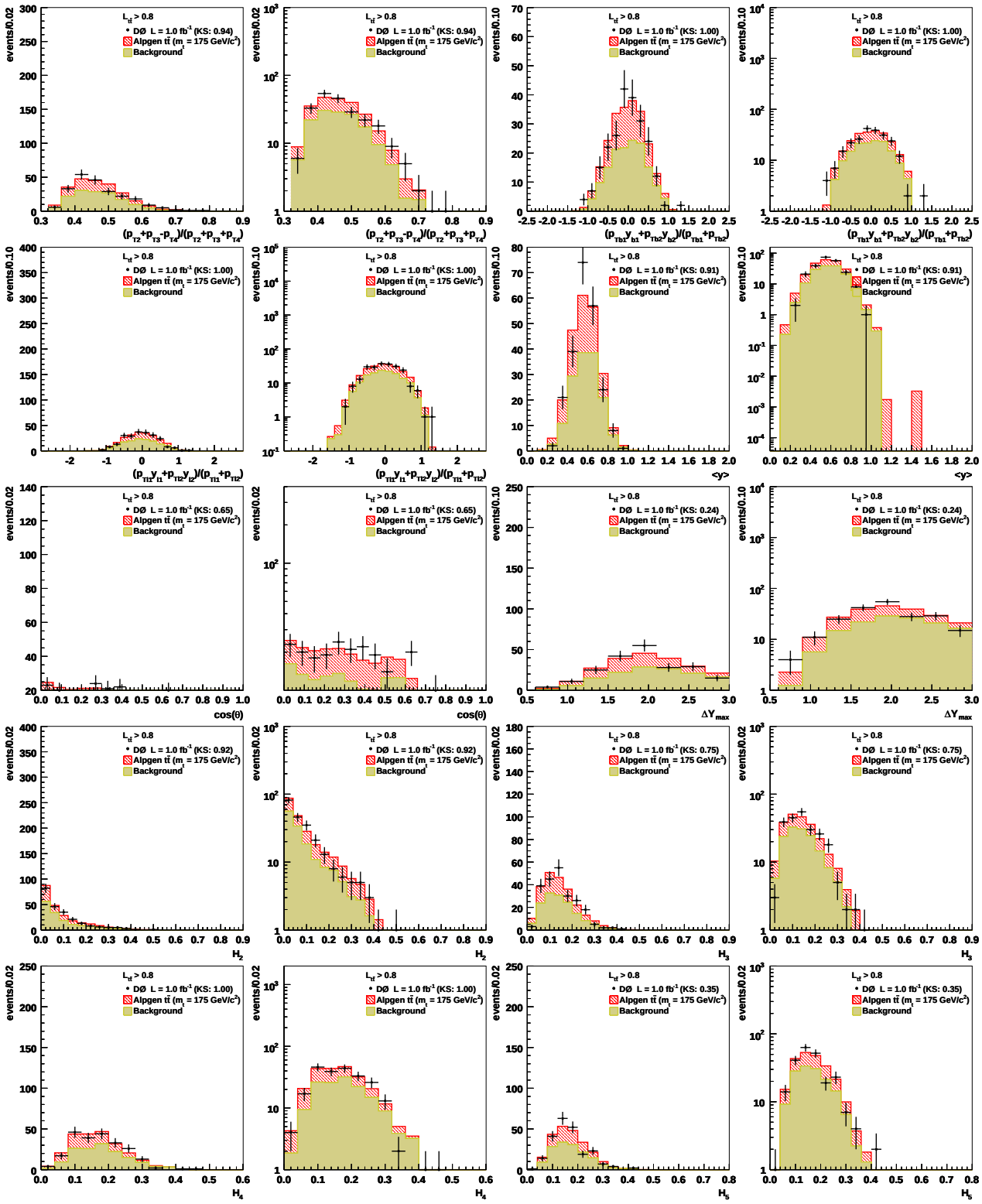


Figure C.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale.

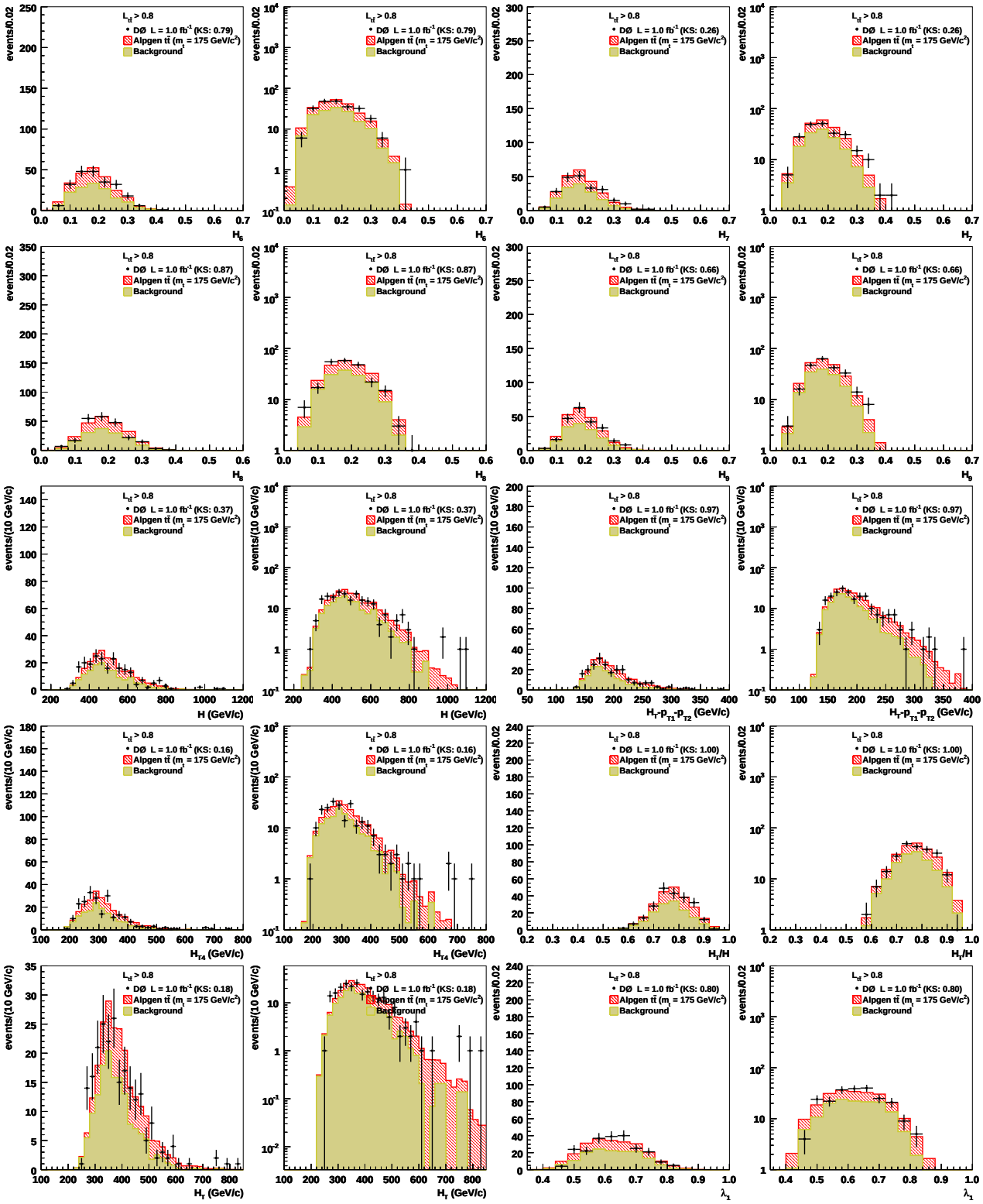


Figure C.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

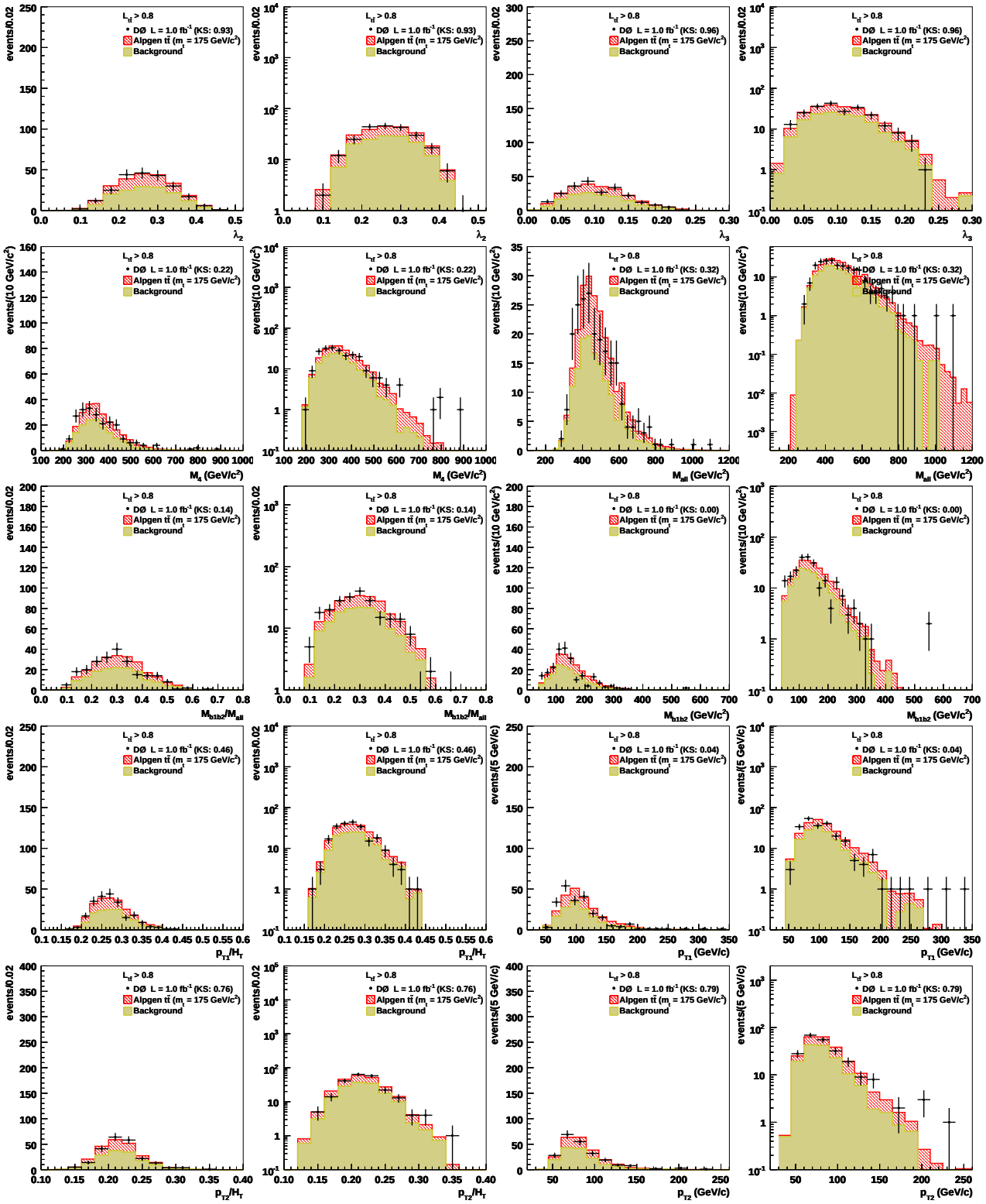


Figure C.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

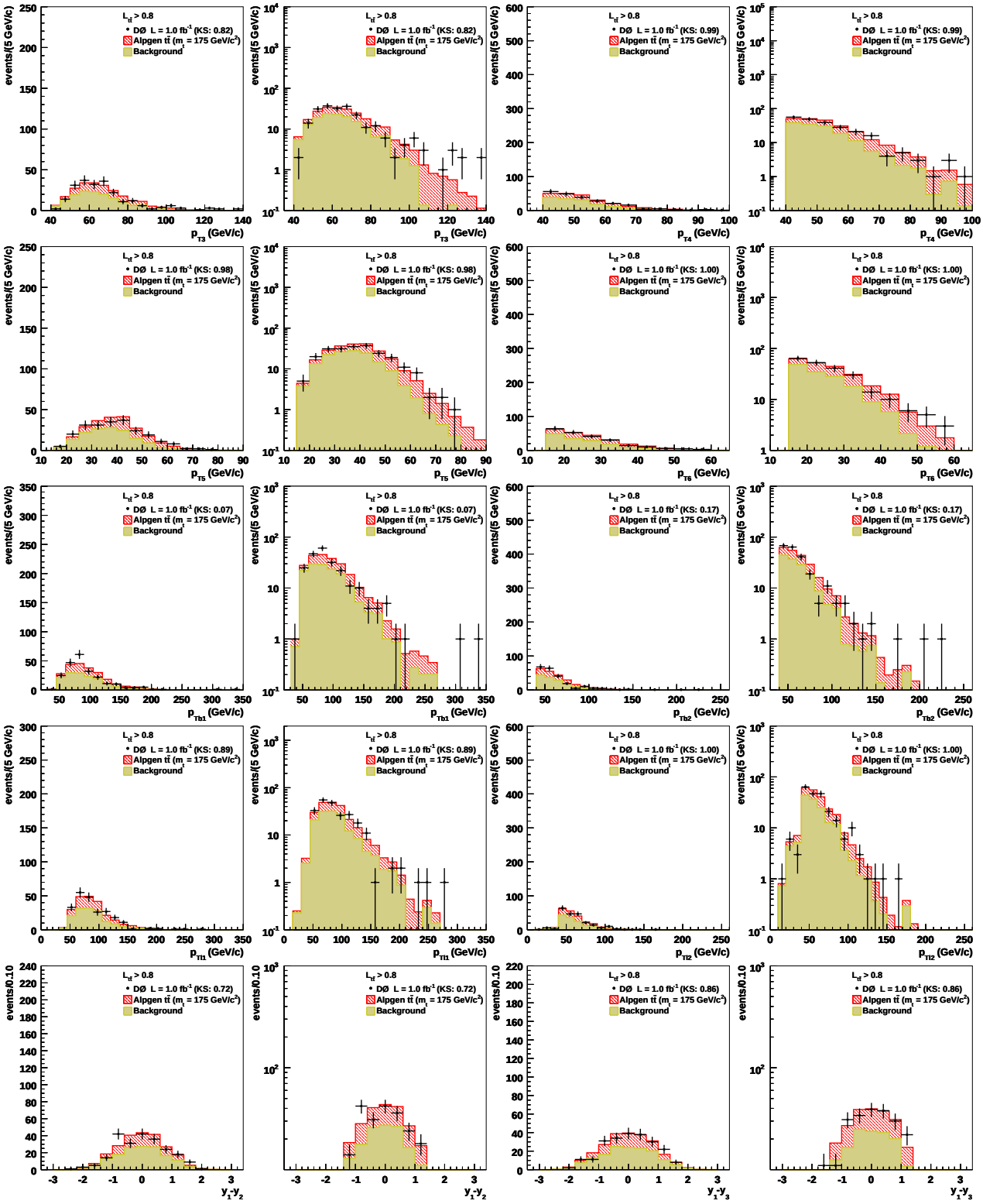


Figure C.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

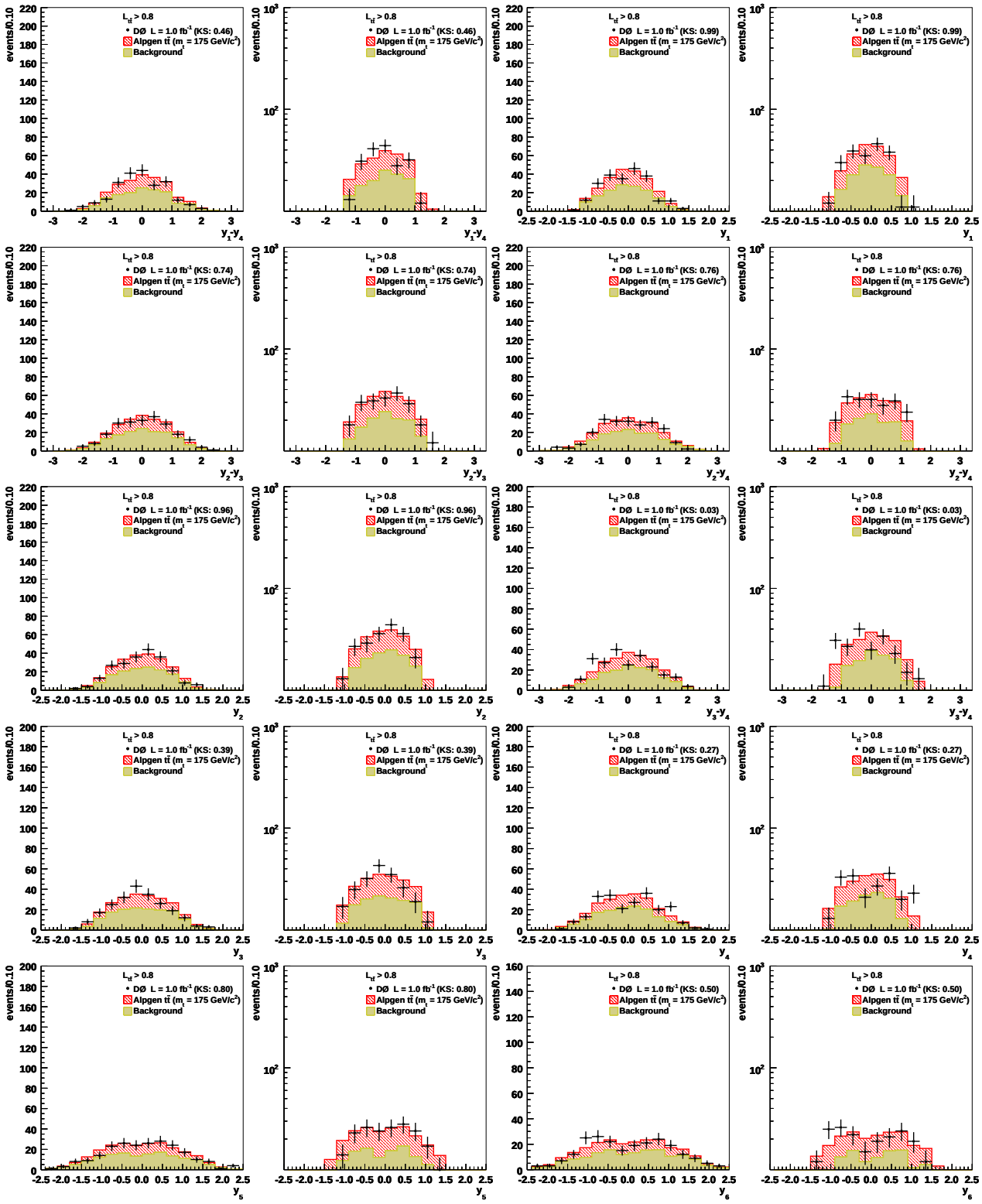


Figure C.4: Results of the likelihood analysis compared to data after cutting on LLH > 0.8. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

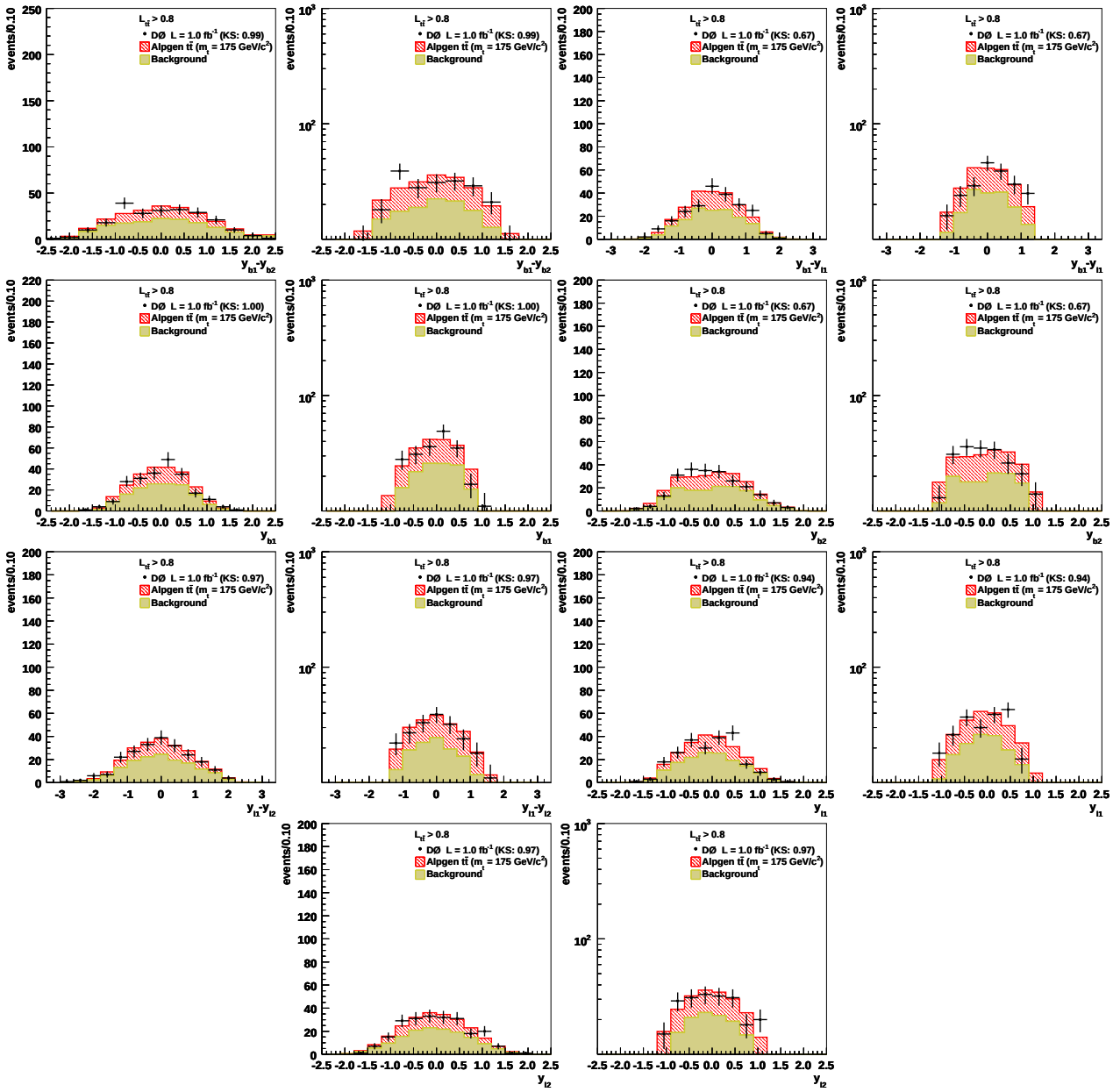


Figure C.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

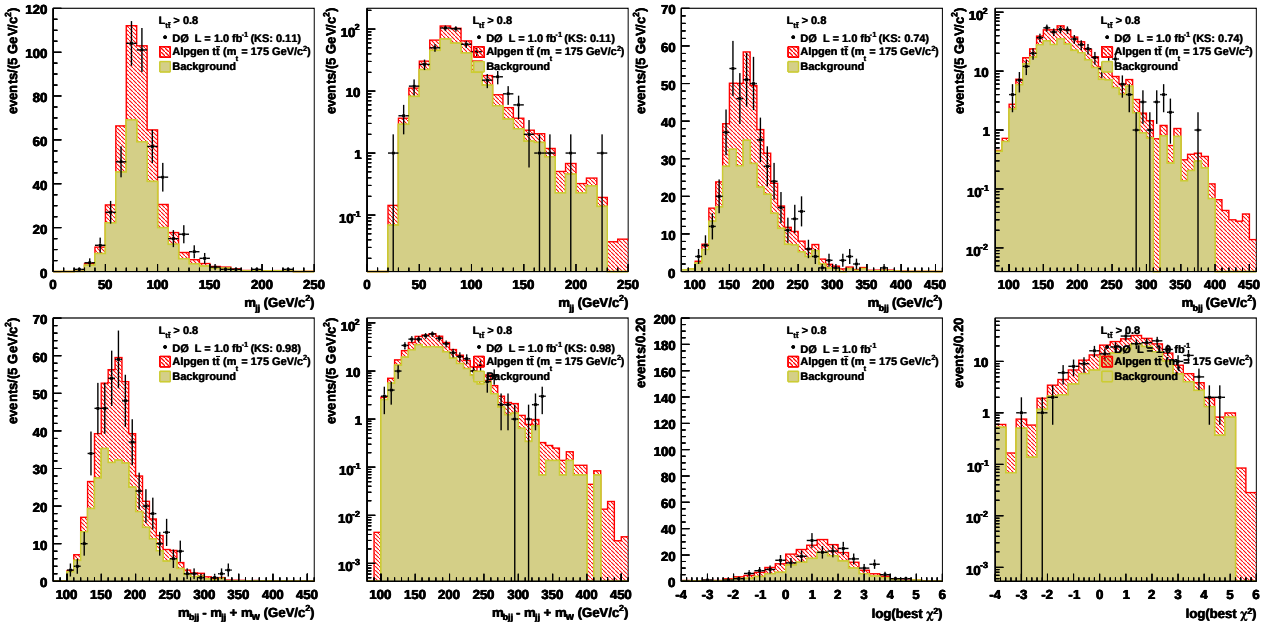


Figure C.4: Results of the likelihood analysis compared to data after cutting on $LLH > 0.8$. Shown are the background contribution (solid fill), the 175 GeV/c^2 $t\bar{t}$ signal contribution (hashed fill) and the data (black markers). Left-hand side: linear scale, right-hand side: logarithmic scale. (continued)

Bibliography

[1] M. Cacciari, S. Frixione, M. L. Mangano, P. Nason, and G. Ridolfi, The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A study of the systematics due to parton densities and scale dependence, JHEP **04**, 068 (2004).

[2] S. Moch and P. Uwer, Theoretical status and prospects for top-quark pair production at hadron colliders, Phys. Rev. **D78**, 034003 (2008), `arXiv:0804.1476 [hep-ph]`.

[3] C. A. Nelson, B. T. Kress, M. Lopes, and T. P. McCauley, General tests for t –> W+ b couplings at hadron colliders, Phys. Rev. **D56**, 5928 (1997), `arXiv:hep-ph/9707211`.

[4] R. H. Dalitz and G. R. Goldstein, The Decay and polarization properties of the top quark, Phys. Rev. **D45**, 1531 (1992).

[5] G. L. Kane, G. A. Ladinsky, and C. P. Yuan, Using the top quark for testing standard model polarization and CP predictions, Phys. Rev. **D45**, 124 (1992).

[6] J. A. Aguilar-Saavedra, J. Carvalho, N. Castro, F. Veloso, and A. Onofre, Probing anomalous W t b couplings in top pair decays, Eur. Phys. J. **C50**, 519 (2007), `arXiv:hep-ph/0605190`.

[7] cited 2008-06-17, `http://mlm.home.cern.ch/mlm/alpgen/`.

[8] T. Sjostrand, L. Lonnblad, S. Mrenna, and P. Skands, PYTHIA 6.3: Physics and manual, (2003), `arXiv:hep-ph/0308153`.

[9] P. Renkel, Correction in inclusive bin cross-sections for Alpgen MC samples, D0 note 5330, 2007.

[10] W. K. Tung *et al.*, Heavy quark mass effects in deep inelastic scattering and global QCD analysis, JHEP **02**, 053 (2007).

[11] D. Stump *et al.*, Inclusive jet production, parton distributions, and the search for new physics, JHEP **10**, 046 (2003).

[12] M. R. Whalley, D. Bourilkov, and R. C. Group, The Les Houches accord PDFs (LHAPDF) and LHAGLUE, (2005).

[13] Y. Peters, M. Begel, K. Hamacher, and D. Wicke, Reweighting of the fragmentation function for the D0 Monte Carlo, D0 note 5325, 2007.

[14] L. Duflot, V. Shary, I. Torchiani, and R. Zitoun, cal_event_quality package, D0 note 4614, 2004.

[15] The D0 data quality group, The D0 data quality group web page, cited 20071220, `http://www-d0.fnal.gov/computing/data_quality/`.

[16] The D0 data quality group, The dq_defs package v2007-10-01, `http://cdcvs0.fnal.gov/cgi-bin/public-cvs/cvsweb-public.cgi/dq_defs/?cvsroot=d0cvs&only_with_tag=v2`

[17] J. Hegeman, Event data quality efficiency for the p17 hadronic top cross section analyses, D0 note 5562, 2008.

[18] J. Hegeman, Luminosity determination and reweighting of Monte Carlo overlay luminosity for p17 hadronic top analyses, D0 note 5561, 2008.

[19] L. Christofek, R. Schwienhorst, M. Begel, and M. Kopal, Probability calculation for multi-object trigger configurations, D0 note 4882, 2005.

[20] A. Harel, Capping the jes muon corrections, D0 note 5563, 2008.

[21] C. Ochando and J.-F. Grivaz, SSR for p17, D0 note 5609, 2008.

[22] T. Scanlon, A neural network b-tagging tool, D0 note 4889, 2005.

[23] M. Anastasoaie, S. Robinson, and T. Scanlon, Performance of the NN b-tagging tool on p17 data, D0 note 5213, 2006.

[24] A. Hocker *et al.*, Tmva: Toolkit for multivariate data analysis, (2007).

[25] G. Hanson *et al.*, Evidence for Jet Structure in Hadron Production by e+ e- Annihilation, Phys. Rev. Lett. **35**, 1609 (1975).

[26] D. Bourilkov, R. C. Group, and M. R. Whalley, LHAPDF: PDF use from the Tevatron to the LHC, (2006).

[27] J. Group, Jet energy scale determination at D0 Run II (final p17 version), D0 note 5382, 2007.

[28] A. Harel, An estimate of the data over mc, b over light jet response uncertainty for runiia jes, D0 note 5654, 2008.

[29] T. Andeen *et al.*, The D0 experiment's integrated luminosity for Tevatron Run IIa, D0 note 5398, 2007.

[30] G. C. Fox and S. Wolfram, Event shapes in e+ e- annihilation, Nucl. Phys. **B149**, 413 (1979).