The search for the Higgs Boson at the ATLAS Experiment using Multivariate Techniques

Diplomarbeit

vorgelegt der

Fakultät für Physik

der

Ludwig-Maximilians-Universität München

von

Johannes Ebke

München, den 25. November 2008

Erstgutachterin: Prof. Dr. Dorothee Schaile
Zweitgutachter: Prof. Dr. Christian Kiesling

Abstract

This thesis describes a search for the Standard Model Higgs boson in the $H \to W^+W^- \to \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$ decay channel. A traditional cut analysis is compared with multivariate classification algorithms: neural networks, boosted decision trees, bagged randomized trees and the Fisher classifier are studied. The analysis uses simulated data of the ATLAS detector at the Large Hadron Collider (LHC), where protons will be collided at a center of mass energy of 14 TeV starting from 2009.

The goal of this thesis is to determine the performance and usability of the various multivariate classifiers in a realistic study. All relevant background processes are considered, and the expected significance of the Higgs boson decay signals for Higgs boson masses of 150 GeV, 170 GeV and 190 GeV is calculated for each technique. The dependency of classifier performance on training parameters is examined, and a consistent system for evaluating and estimating systematic uncertainties is proposed and demonstrated.

Although slower in analysis, boosted decision trees can be easily trained to outperform an optimized cut analysis on the same variables. Neural networks are also capable of improving upon a classical analysis in many cases, although more care and manual tuning is necessary. However, systematic uncertainties affect multivariate classifiers almost always much stronger than a cut analysis. Additionally, accurate estimates of systematic uncertainties for multivariate classifiers can only be obtained if uncertainties are simultaneously evaluated: considering each uncertainty individually and then combining them works well for the cut analysis but seriously underestimates the total uncertainty in the case of multivariate classifiers.

Zusammenfassung

Diese Diplomarbeit beschreibt eine Suche nach dem Higgs Boson im Zerfallskanal $H \to W^+W^- \to \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$. Eine übliche schnittbasierte Analyse wird mit folgenden multivariaten Techniken verglichen: Neuronale Netze, Boosted Decision Trees, Bagged Randomized Trees und die Fisher-Methode werden untersucht. Die Analyse verwendet simulierte Daten des ATLAS Detektors am Large Hadron Collider (LHC), an dem von 2009 an Protonen mit einer Schwerpunktsenergie von 14 GeV kollidiert werden sollen.

Das Ziel dieser Arbeit ist die Effizienz und Benutzbarkeit der verschiedenen multivariaten Techniken in einer realistischen Studie zu vergleichen. Alle relevanten Untergrundprozesse werden berücksichtigt, und die erwartete Signifikanz der Higgs-Boson-Zerfallssignale für die Higgs-Boson-Massen 150 GeV, 170 GeV und 190 GeV für jede Methode berechnet. Die Abhängigkeit der Effizienz der verwendeten Techniken von Trainingsgegebenheiten und -parametern wird untersucht, und ein konsistentes System zur Auswertung und Abschätzung systematischer Unsicherheiten wird vorgestellt und angewendet.

Obwohl Boosted Decision Trees in der Anwendung am langsamsten sind, können diese leicht so traininert werden, dass sie eine höhere Effizienz als eine Schnittanalyse aufweisen. Neuronale Netze können ähnliche Effizienz erreichen, es ist jedoch mehr Sorgfalt und manuelles Tuning nötig. Systematische Unsicherheiten beeinflussen multivariate Techniken jedoch oft deutlich mehr als eine Schnittanalyse. Auch können systematische Unsicherheiten für multivariate Techniken nur gut abgeschätzt werden wenn diese gleichzeitig betrachtet werden: Wenn sie einzeln ausgewertet und dann kombiniert werden wird die gesamte systematische Unsicherheit der multivariaten Techniken stark unterschätzt.

Contents

1. Introduction

The proposal of the electron by J.J. Thomson in 1897 opened up a new field of physics: the search for sub-atomic particles and the laws that govern them. During the 20th century, a veritable zoo of particles was discovered in cosmic rays and at particle accelerators. The quark model proposed and developed by Gell-Mann, Nishijima, Ne'eman and Zweig [1, 2] explains many of these particles as composite objects, consisting of quarks of three colours. On the close of the 20th century, the Standard Model of particle physics with its three families of quarks and leptons was nearly unanimously accepted as an excellent description of our current state of knowledge. Its predictions match experiments extremely well, in some cases exceeding a precision of nine significant digits [3]. Yet the Standard Model is not completely confirmed: the question how the electroweak bosons W^+, W^- and Z acquire mass is still open. The most prominent explanation consistent with experiment is the existence of a Higgs field, and it predicts the existence of a spin-0 boson, remnant of a process of spontaneous symmetry breaking. Since this boson has not been discovered at previous experiments, the decision what model best fits reality can only be decided by new high-energy colliders.

The ATLAS detector at the Large Hadron Collider, where protons will be collided at a center of mass energy of 14 TeV is a window to this new region of high energy physics. It is expected that results either confirm the extremely successful Standard Model of particle physics by finding the Higgs boson, or see signs of new physics beyond it. However, the effects of a Higgs boson are expected to be tiny compared to several Standard Model processes. In this situation, advanced analysis techniques as used in this thesis can significantly improve sensitivity to new physics [4].

In the first part of this thesis, an introduction to the matter is given: The Standard Model of particle physics is shortly described, as well as the theoretical expectations for the Higgs boson production and the background processes at LHC. Multivariate classifiers are presented and explained. The LHC and the ATLAS Experiment are introduced, as well as the methods to simulate the detector and to analyze the data.

The main part deals with the analysis of the data obtained by simulation. Firstly, a coherent statistical treatment of occurring uncertainties, statistical as well as systematical, is developed. Preparations and procedures for the analysis are described. Some features of the relevant processes and of the data used are shown, and experiences in training, application and performance of multivariate classifiers are presented. Finally, the results of the optimized cut and mulltivariate analyses are given in detail, the expected discovery luminosity for each method stated, and the results discussed in view of practicality, reliability and performance.

The Appendix contains technical details that are important for a practical implementation of the analysis.

2. Theory and Methods

2.1 The Standard Model of Particle Physics

The Standard Model of fundamental particles and interactions describes almost all readily observable effects in nature - apart from gravity - exceedingly well. In total 24 fermion fields - six quarks in three colours each, and six leptons - are part of it, manifesting as particles. Twelve boson fields convey the different interactions: The massless photons carry the electromagnetic, the massive W^+, W^- and Z the weak interaction, and the eight multicoloured gluons the strong interaction dominant in atomic nuclei [2].

Since the discovery of the last quark fitting into the apparent periodic structure of the Standard Model at TeVatron in 1995 [5], this remarkable Theory of particle physics seems to be nearing completion.[1]. The last missing piece seems to be the Standard Model Higgs particle, a boson, its predicted mass being around 91^{+45}_{-32} GeV [3], with a mass of 170 GeV already excluded by TeVatron [7] at 95% confidence. Why now is the Higgs boson necessary, and how can we detect it? In the following a short overview over the Standard Model tries to explain the last two points.

As far as we know, particle physics excluding gravity is described by a *quantized least action principle*[2].

Least Action Principle: In classical physics, a least action principle can be described as the following prescription for the field ϕ:

$$\min_{\phi, \partial_\mu \phi} \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x \tag{2.1}$$

From this prescription a set of equations for the Lagrangian $\mathcal{L}$, the Lagrange equations, can be derived. These determine exactly the evolution of the field ϕ, given initial values and first derivatives[3].

[1] Yet in the same decade, the discovery of neutrino oscillations [6] shook long-held beliefs, and makes modifications to the Standard Model necessary

[2] Gravity is currently described best by the general theory of relativity, also defined by a least action principle. However, no successful quantized version of this theory is yet known.

[3] Actual calculation can of course still be very difficult

Quantum Field Theory: A full introduction to Quantum Field Theory is not in the scope of this thesis, a good introductory textbook is for example [8]. Still, some points are important for the understanding of accelerator experiments.

In Quantum Field Theory, the fields themselves are replaced by operators. Using a quantization procedure one can extract serveral interesting features and predictions of the theory[4]. The usual assumption in high energy physics is that interactions are *weak* - an assumption that, except for bound states of quarks, is very reasonable. This allows us to expand the theory in a power series of the (dimensionless) interaction strength, and for each power and each process we obtain several *Feynman diagrams*, which are diagrammatical expressions of integrals. These diagrams can be evaluated to give cross sections depending on the initial and final conditions, and the sum over all powers gives the actual prediction of the Theory.[5] Since calculating corrections at higher orders is very difficult, most of the time only predictions to the next to leading and perhaps second to leading order are available for a given process, and many Monte Carlo generators work only to the leading order of a process.

The Standard Model: Central to the Standard Model[6] is the gauge symmetry group $SU(3) \otimes SU(2) \otimes U(1)$ that "generates" the bosonic fields. For each symmetry, local gauge invariance demands a gauge field to exist.

We have therefore for the Standard Model in total $8 + 3 + 1$ gauge fields, one for each generator of each of the groups: The eight gluon gauge fields G^a_μ, and then the electroweak fields, a weak isospin triplet field W^i and a weak isosinglet field B. Additionally to the gauge fields, we introduce two kinds of fermionic fields: Six quark fields q_j which couple to electroweak and strong gauge fields, and six lepton fields that couple only to the electroweak fields.

At this point in theoretical development it was proven that unfortunately there is no way to directly give gauge bosons a mass without making the theory nonrenormalizable. Fortunately, it is possible to introduce mass indirectly using a two-component complex scalar gauge field ϕ - the Higgs field. The potential of this field is symmetric, but with an instability around the symmetry axis. This enables a spontaneous breaking of the symmetry at the beginning of (theory) time - this can give gauge bosons a mass without

[4]It is also possible to discretize the problem and apply evolution equations to a field on a lattice, especially in the study of bound states in Quantum Chromodynamics this is the only way to make predictions

[5]This description ignores the fact that to get to the point where actual, second order predictions can be made, the theory has to be regularized, renormalized and the gauge fixed, and that calculating even a one one-loop correction is a serious undertaking.

[6]For a historical overview and a bibliography of original papers, see [9]

breaking gauge invariance. This mechanism will be explained in more detail later.

The complete Lagrangian $\mathcal{L}$ of the Standard Model is therefore given by:

$$\mathcal{L} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{Higgs}} \tag{2.2}$$

The gauge fields of QCD are the G_μ^α, where α is the group index. The strong field and field strength are then defined as:

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - ig_s \left[G_\mu, G_\nu\right] \tag{2.3}$$

$$G_\mu = \sum_{\alpha=1}^{8} G_\mu^\alpha \frac{\lambda^\alpha}{2} \tag{2.4}$$

Using the fields $\boldsymbol{W}$ and B the electroweak partial derivative D_μ and the electroweak field strengths $F_{\mu\nu}^j$ and $B_{\mu\nu}$ are defined as follows:

$$D_\mu = \partial_\mu - ig\frac{\boldsymbol{\sigma}}{2}\boldsymbol{W}_\mu - ig'\frac{Y}{2}B_\mu \tag{2.5}$$

$$F_{\mu\nu}^j = \partial_\mu W_\nu^j - \partial_\nu W_\mu^j + g\epsilon^{jkl}W_\mu^k W_\nu^l \tag{2.6}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.7}$$

$$\tag{2.8}$$

Using this definitions the parts of the Lagrangian can be written as:

$$\mathcal{L}_{\text{QCD}} = \sum_j i\bar{q}_j\gamma_\mu \left(D^\mu - ig_s G^\mu\right) q_j - \frac{1}{2}\text{tr}G_{\mu\nu}G^{\mu\nu} \tag{2.9}$$

$$\mathcal{L}_{\text{EW}} = \sum_k i\bar{\psi}_k\gamma_\mu D^\mu\psi_k - \frac{1}{4}F^{j\mu\nu}F_{\mu\nu}^j - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{2.10}$$

$$\mathcal{L}_{\text{Higgs}} = |D_\mu\phi|^2 - V(\phi) - \sum_j c_j\bar{q}_j\phi q_j - \sum_k f_k\bar{\psi}_k\phi\psi_k \tag{2.11}$$

Each of these parts describes various particles and their interactions, and will be explained now in more detail.

2.1.1 $\sum_j i\bar{q}_j\gamma_\mu \left(D^\mu - ig_s G^\mu\right) q_j - \frac{1}{2}\mathbf{tr}G_{\mu\nu}G^{\mu\nu}$

Kinematic term for quarks and their electroweak interaction The first term in this expression is a kinematic term that describes the propagation of quark fields - $\bar{q}_j\gamma_\mu\partial^\mu q_j$. It also includes various interaction terms. The covariant electroweak derivative D^μ includes a full set of electroweak interaction. The strength of this electroweak interaction depends on the isospin $\boldsymbol{\sigma}$ and weak hypercharge Y of the participating fields.

Strong interaction: The term $\bar{q}_j \gamma_\mu i g_s G^\mu q_j$ describes the strong interaction of quarks and gluons, coupled with the strong coupling constant g_s. The peculiarity of the strong interaction now is not only that the coupling constant g_s is large, but also that the next term - $-\frac{1}{2}\mathrm{tr} G_{\mu\nu} G^{\mu\nu}$ - describes a gluon-gluon self-interaction. At higher energies, this leads to a self-similarity of quarks and gluons: gluons can be observed to split into virtual quark-antiquark pairs, and these quarks again contain gluon loops. In this theory, this self-similarity continues up to infinite energies -infinitely small distances - but it is expected that at the Planck scale Λ, when gravity becomes as strong as the other interactions, this SU(3) symmetry turns out to be only a fragment of a spontaneously broken greater group[7].

Running coupling constant: One other interesting feature of this part of the Lagrangian is that - after the process of renormalization, or adapting to the energy scale of a process - the coupling constant seems to decrease with increasing energy. This means that the hard part of collisions in particle accellerators can often be described without taking into account the full non-perturbative strong interaction and assuming that the coupling is small. This effect is called "asymptotic freedom"[8]. This of course mostly affects the calculation of cross-sections, since the subsequent "hadronization" of the colour-charged remnants of the hard interaction can as yet only be described by empirical models and numerical approximations.

Colour and Confinement: In the expression of the gluon fields G^μ we find the λ^a, the generators of the $SU(3)$ group. SU(3) is the group of rotations of a three-dimensional sphere, and so the λ terms do not commute. According to Noethers Theorem, this leads to three separate conserved gauge currents that can be derived from the Lagrangian - called colloquially the red, green and blue current. It was discovered that because the coloured gluons interact with themselves, the above Lagrangian does not allow "naked" color to exist: If a red and an antired quark are pulled apart, the field strength increases with distance until from the built-up energy an colour-anticolour quark pair is formed and two colourless parts remain.

2.1.2 $\sum_k i\bar{\psi}_k \gamma_\mu D^\mu \psi_k - \frac{1}{4} F^{j\mu\nu} F^j_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}$

Transformation into charge eigenstates On further examination of this Lagrangian, it turns out that the fields W and B are not charge eigenstates and therefore can not really be identified with particles. One can, however, choose four linear combinations of these fields that each have distinct properties and can be identified with particles: The W^+ and W^- fields correspond

[7] This is the domain of Great Unified Theories, or GUTs

[8] Nobel Prize 2004, F. A. Wilczek, H. D. Politzer and D. Gross

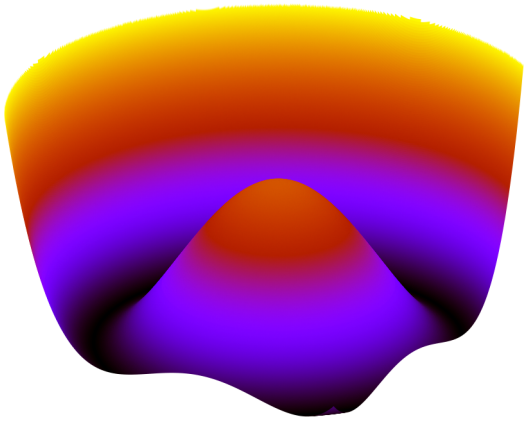


Figure 2.1: The usual "Mexican Hat" potential of the Higgs field

to charged particles, the neutral Z couples to a combination of weak and electric charge while the well-known photon field A couples only to electric charge.

History and Experiment The electromagnetic part A of the electroweak force was the first part of the Standard Model to be transformed into a quantum field theory. James Clerk Maxwell discovered empirically the equations that govern the electromagnetic field, and later Feynman and Schwinger [10] were the first to quantize a classical field theory and apply it to reality as Quantum Electrodynamics, identifying the quantized electric field A as the photon. In the 70s, Glashow, Salam, Weinberg and others developed the unification of the electromagnetic and the weak interaction and proposed the W and Z particles [9]. The discovery of these particles at the UA1/UA2 experiments at CERN in 1983 was one of the great successes of this theory, and the success of the Standard Model is firmly founded on these experiments.

2.1.3 The Higgs Mechanism

In the last two sections, the question of mass was ignored. Static mass terms in a Lagrangian have the form $-\bar{\psi}m\psi$ or $\mu^2 A_\mu A^\mu$ for spin-1/2 and spin-1 particles, respectively. If we now would just add such a mass term for the massive boson fields W and Z, gauge invariance would be broken explicitely, and the theory would not be renormalizable. This means we cannot introduce mass statically.

The prime method for *dynamical* mass generation was proposed by Peter Higgs in 1964 [11]. He introduced an additional two-component complex scalar quantum field that has a so-called "mexican hat" potential (see figure 2.1)

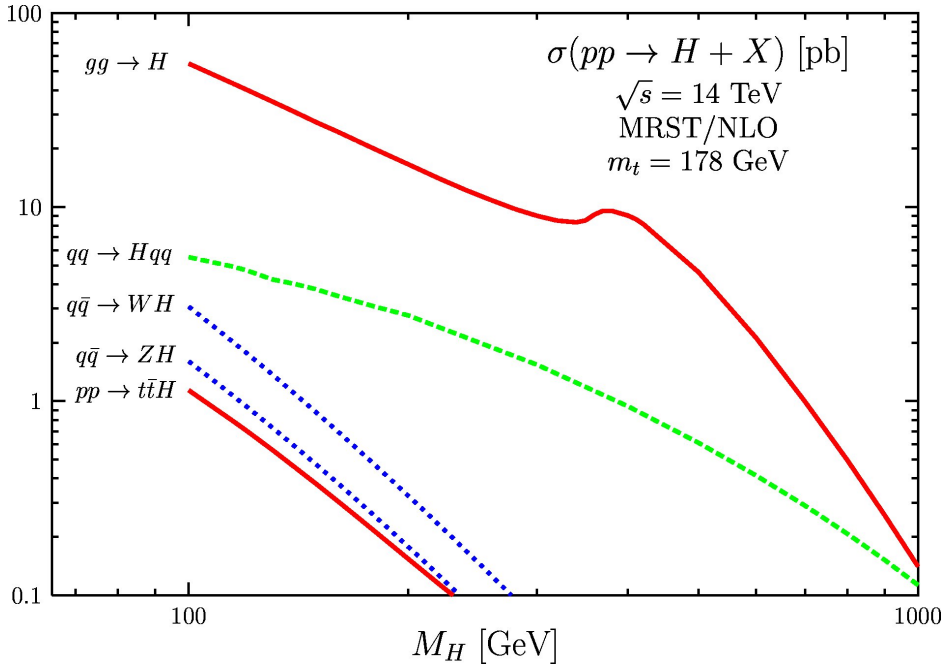


Figure 2.2: The Higgs boson production cross-sections for different Higgs boson masses at the LHC [12]

In the usual model, shortly after the big bang, the temperature fell below the energy at the origin of the potential. Small fluctuations lead then to the breaking of the initial symmetry, and the field settles after further cooling at the bottom of the valley. This leads to a scalar field that is nonzero everywhere in space.

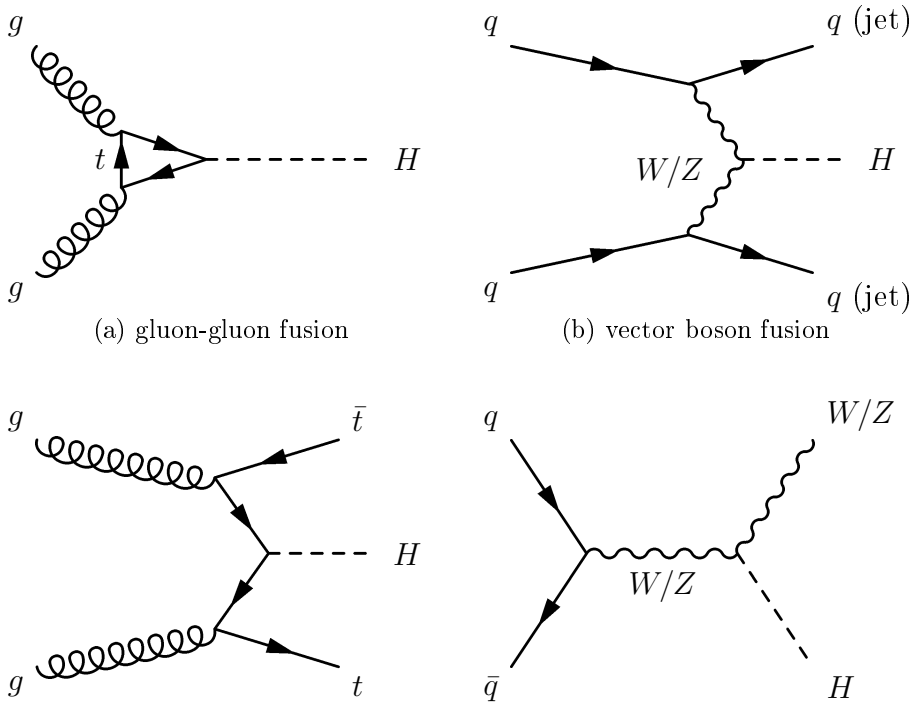
Now, a complex scalar field with two components has four degrees of freedom. Three are transformed by a gauge transformation to give the spin-1 particles W^-, W^+ and Z a third spin component [9]. So there is one component left, which describes a real scalar field, and couples to each particle with a strength proportional to its mass. The excitation of this Higgs field would manifest itself as a spin-0 particle. This particle, the Higgs boson, is essential to the theory, and is the only particle of the Standard Model not yet observed.

2.2 Standard Model Processes

2.2.1 Higgs Boson Production

A Standard Model Higgs boson can be produced at LHC in four different channels (figures 2.2 and 2.3), most prominently in the gluon-gluon-fusion

[9] massless particles can only have spin in forward or backward directions



(a) gluon-gluon fusion

(b) vector boson fusion

(c) associated t-Higgs boson production

(d) associated W-Higgs boson production

Figure 2.3: Production channels of the Higgs boson

and the vector boson fusion (VBF) channels. Strongest is gluon fusion via the top quark loop. In this channel, the decay products of the Higgs boson are the only objects in the event with high transverse momentum. In contrast, vector boson fusion is accompanied by the two participating quarks as two jets with low transverse momentum. This enables us to separately analyze the VBF channel. In addition to these channels, the Higgs boson can also be produced in association with quarks and in association to a W boson. Since these processes have a very small cross-section and need quite different analysis methods they are ignored in the following.

2.2.2 Higgs Boson Decays

The Higgs boson is an uncharged spin-0 particle. It can decay to vector bosons, to quarks and - via a top quark loop - to photons. The branching ratios are strongly dependent on the Higgs boson mass (see figure 2.4). If the Higgs boson is light, up to 140 GeV, the decay to $b\bar{b}$ is strongest. Since hadronic decays are difficult to separate from the huge backgrounds with similar topologies, the decay to two photons is also important in this region. For higher Higgs boson masses, the decay to vector bosons is dominant, above 200 GeV the decay of ZZ to four leptons provides a very clear signal. Yet in

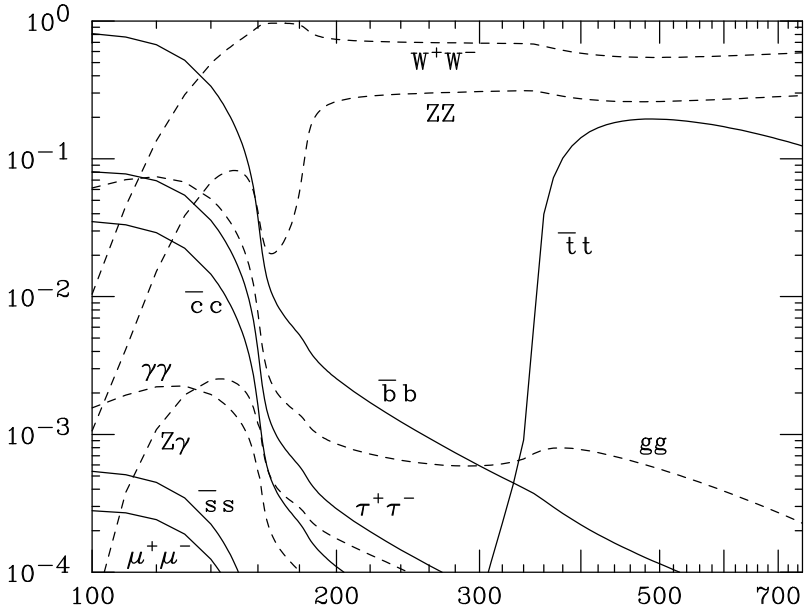


Figure 2.4: The Higgs boson decay branching ratios for different Higgs boson masses Particle Data Group, 2007 Web Update of Review of Particle Physics

the region of 140 GeV to 200 GeV the decay to W^+W^- and the subsequent decay of the Ws into muons (see figure 2.5) is a central channel for a Higgs boson discovery. Additionally, in some theories the Higgs-W coupling is enhanced, and the W^+W^- decays are also relevant at lower Higgs boson masses [13].

2.2.3 Background Processes

In this section, the four most important Standard Model processes that can cause similar detector signatures as the $H \rightarrow W^+W^- \rightarrow \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$ decay are presented. In figure 2.6 the relevant Feynman diagrams are shown.

Drell-Yan process The di-muon (and di-tau) production via Z/γ^* as shown in figure 2.6(a) is the dominant background of this analysis channel, even though the final state lacks the neutrinos of the Higgs boson decay. Since the expected total initial cross-section in this channel with an invariant mass above 15 GeV is 3.35 nb a good understanding of this background is crucial to detect Higgs boson production with a cross-section of approximately 1 pb. Luckily, good separation of signal and backround is possible, since the kinematic properties of Drell-Yan as a direct $\mu\mu$ decay differ markedly from a decay where each muon originates from a W boson.

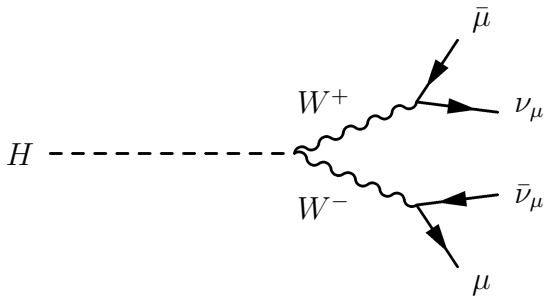
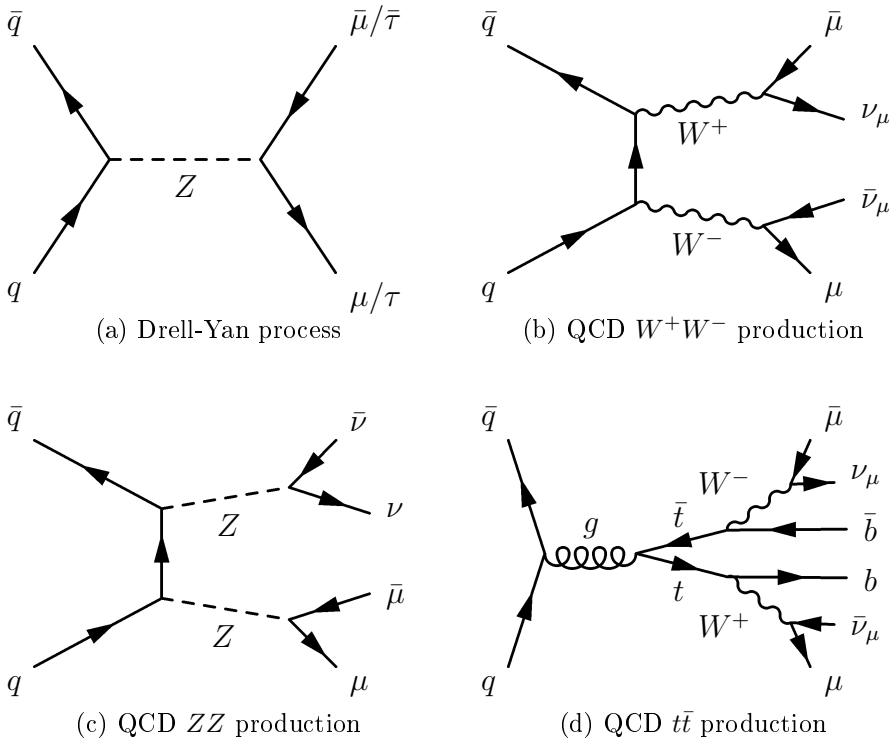


Figure 2.5: Feynman diagram for the $H \rightarrow W^+W^- \rightarrow \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$ decay examined in this thesis

QCD W^+W^-: The direct QCD W^+W^- production is the most challenging background to this analysis: The kinematic properties as seen in the Feynman diagram 2.6(b) are very similar to the Higgs boson decay. Yet there is one separation criterion: Since the two W bosons of the Higgs boson decay are produced at one vertex with very little transverse momentum and the Higgs boson has spin 0, their spins are preferentially antiparallel. Since the neutrino emitted by the W^+ must be left-handed the muon from the W^+ is emitted preferentially in the direction of its spin, and since the antineutrino emitted by W^- must be right-handed the muon from the W^- is emitted preferentially in the opposite direction of its spin. This means that the opening angle between the two muons is tending to smaller angles. This will be exploited in the analysis [14].

QCD ZZ: Direct di-Z production as shown in figure 2.6(c) has much less influence on the analysis than W^+W^- production. This is the case since the two muons originate from one Z, and their distribution is therefore similar to the Drell-Yan background.

QCD $t\bar{t}$ production Top quark decays pose a problem mostly for the vector boson fusion channel. According to the Standard Model, top quarks decay almost exclusively to Wb (see figure 2.6(d)). If then both W decay to muons, the signature is quite similar to the VBF Higgs boson production channel: The two b-jets from the top decays mimic the two jets seen in the Higgs boson production process. Also, the NLO prediction for the total $t\bar{t}$ cross-section at the LHC is 833 pb, and the cross-section for decays which include at least one lepton is 450 pb, still much more than the predicted Higgs boson cross section of approximately 1 pb. One possibility is the use of the $b-tagging$ technique, where secondary vertices arising from the "long" lifetimes of the b-mesons are identified. While this could reduce background



(a) Drell-Yan process (b) QCD W^+W^- production

(c) QCD ZZ production (d) QCD $t\bar{t}$ production

Figure 2.6: Feynman diagrams for main background processes

significantly, the efficiency of the technique is very difficult to estimate by simulations, and will have to be determined by actual data. Therefore, the technique is ignored in the following.

2.2.4 Monte Carlo Event Generators

The exact Standard Model differential cross sections including hadronization and radiative corrections can not be calculated analytically. The same holds for the exact response of the detector to each possible event. Therefore, a study must use Monte-Carlo event generators to simulate virtual collisions. This is not a simple task: several approximations, parametrisations and auxiliary models have to be introduced [15].

In table 2.1 an overview to the datasets used for this analysis is given: the Monte Carlo generator used for each process, the corresponding cross section times branching ratio $\sigma \times BR$ after initial filters and the numbers of events simulated are given. The event numbers in the columns "full events" and "fast events" refer to the degree of detail of the detector simulation (see 3.3). The dataset number in the table refers to unique number given to the corresponding ATLAS configuration file.

One of the most versatile generators is named Pythia [16]. Its predecessors

are in development since 1978, and it can generate almost all processes out of the box. Processes are simulated to tree level and corrections for initial and final state radiation as well as for renormalization are applied. Also, models for hadronization of final-state quarks and gluons are included. To improve the modeling of more complicated, non-tree level contributions of certain processes it is possible to use AcerMC [17] to simulate the hard process before using Pythia for further decays.

Another general-purpose tree-level generator is Herwig [18]. It is used for hadronization for the next-to-leading order generator MC@NLO [19, 20] and for the generator Alpgen [21] specialized on multi-parton processes in hadronic interactions.

process	dataset nr.	generator	full events	fast events	sample cross-section $\sigma \times BR$ [pb]
$Z/\gamma^* \to \mu^+\mu^-$ ($m_{\mu\mu} > 60$ GeV)	5145	Pythia	425900	400000	1793
$Z/\gamma^* \to \mu^+\mu^-$ (15 GeV $< m_{\mu\mu} <$ 60 GeV)	-	Pythia	8850	521263	1192
$Z/\gamma^* \to \tau^+\tau^-$	5146	Pythia	149600	0	77
$t\bar{t}$	5200	MC@NLO	403580	0	450
$H \to W^+W^-$(170 GeV)	5320	Pythia	25700	0	1.24
$H \to W^+W^-$(150 GeV)	-	Pythia	19000	0	1.16
$H \to W^+W^-$(190 GeV)	-	Pythia	20000	0	0.88
Wt	5500	AcerMC	13950	0	26.7
$W^+W^- \to e^+e^-$	5921	MC@NLO	13624	0	1.4
$W^+W^- \to e^+\mu^-$	5922	MC@NLO	13626	0	1.4
$W^+W^- \to e^+\tau^-$	5923	MC@NLO	12808	0	1.4
$W^+W^- \to \mu^+\mu^-$	5924	MC@NLO	10220	0	1.4
$W^+W^- \to \mu^+e^-$	5925	MC@NLO	13868	0	1.4
$W^+W^- \to \mu^+\tau^-$	5926	MC@NLO	10390	0	1.4
$W^+W^- \to \tau^+\tau^-$	5927	MC@NLO	13534	0	1.4
$W^+W^- \to \tau^+e^-$	5928	MC@NLO	12262	0	1.4
$W^+W^- \to \tau^+\mu^-$	5929	MC@NLO	13740	0	1.4
ZZ	5986	Herwig	48000	0	2.81
WZ	5987	Herwig	4450	0	13.86
$W(\mu) + jets$, 0 partons	6107	Alpgen	16500	99500	819
$W(\mu) + jets$, 1 partons	6108	Alpgen	10000	300000	1630
$W(\mu) + jets$, 2 partons	6109	Alpgen	9950	62000	588
$W(\mu) + jets$, 3 partons	6110	Alpgen	10450	100000	174
$W(\mu) + jets$, 4 partons	6111	Alpgen	9500	22000	50.4
$Wbb \to l\nu b\bar{b}$	6366	Alpgen	19500	0	29

Table 2.1: Monte Carlo sample information. All $\sigma \times BR$ are in NLO, NNLO or LO + k-factor correctrions to NLO

2.3 Multivariate Classifiers

In general, multivariate classifiers are mathematical algorithms that sort data points into different categories. In the case of "supervised learning" used in this analysis a pre-classified set of data "events" is provided, and the classifier trained to perform well on the training set. The efficiency of the classifiers is then tested on an independent testing sample. In the high energy physics case, only two categories are used: signal and background. In this case, a multivariate classifier is simply a scalar function f of the input variables. Higher values correspond to "signal-like" events, whereas lower values indicate background. A clearly defined "signal region" can then be defined as the region in input variable space where for which $f(\boldsymbol{x}) > c$ for the input variable vector $\boldsymbol{x}$, with c as a constant.

2.3.1 Training and Testing samples

The available Monte Carlo data is first split into training and testing samples: The classifiers as well as the analysis cuts are trained and optimized, respectively, on the training sample, whereas all results are only derived from the testing sample. This increases the uncertainty due to Monte Carlo statistics, but it is essential to get correct results: multivariate classifiers tend to perform much better on the training sample than on the testing sample. This "overtraining" is especially visible in classifiers with many parameters. However, as long as the testing sample is statistically independent from the training sample the derived results are accurate, even if the classifier is overtrained.

2.3.2 The Toolkit for Multivariate Analysis: TMVA

The correct implementation of a multivariate classifier and its training algorithm is a non-trivial task. To compare the results from different classifiers consistently is often also very difficult [22]. To simplify the procedure, the Toolkit for Multivariate Analysis (TMVA, [22], version 3.9.4) was used in this thesis for all multivariate methods. In the following, a short description of the classifiers and their parameters is given.

2.3.3 Boosted Decision Trees

Boosted decision trees (BDT) are a popular tool for event classification in high energy physics [23]. Firstly, a decision tree is simply a binary tree of cuts on one variable with "signal" and "background" baskets at the end as illustrated in figure 2.7. In training or growing a decision tree, one starts with a root node. Then a variable and cut is searched that maximizes the sum of

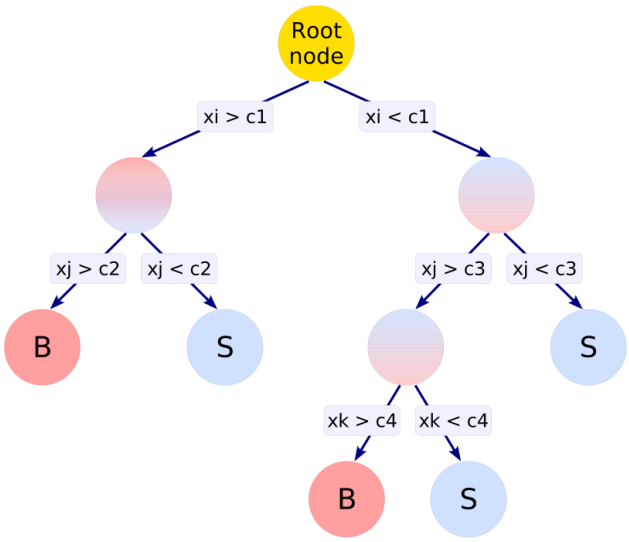


Figure 2.7: One Decision Tree: Binary Tree of Cuts

the significances $S/\sqrt{S+B}$ of each new subsample weighted by the number of events in it; for each variable 20 possible cuts between the minimum and the maximum are tried out. Two leaf nodes, one for each subsample are added and the process repeated for each node. This leads quickly to a tree where each training event is classified perfectly, so a "pruning" method is used to cut the most statistically insignificant nodes. Such a tree is very easy to interpret, yet is extremely sensitive to slight changes in the training sample - if at one node a different variable is chosen, the whole tree below it has a different structure.

AdaBoost: A boosting procedure promotes one classifier optimized to one training sample to many classifiers derived from reweighted training samples. The result of the now "boosted" classifier is the number of votes of the single classifiers. In the used "AdaBoost" algorithm, the weight of previously misclassified events is multiplied by the common "boost weight"

$$\alpha = \frac{1-\epsilon}{\epsilon}$$

Here ϵ is the misclassification rate of the previous tree. The output of the boosting algorithm is the weighted sum over all trees: $\sum_{i \in trees} \ln \alpha_i h_i(\boldsymbol{x})$, where $h(\boldsymbol{x})$ is $+1$ if $\boldsymbol{x}$ is classified as signal by the ith tree and -1 otherwise.

The boosting reduces the dependence on the choice of the training sample, yet the easy interpretation of decision trees is lost. Additionally, many additional degrees of freedom are introduced, and most boosted decision trees are overtrained. This makes it imperative that the selection efficiency is deter-

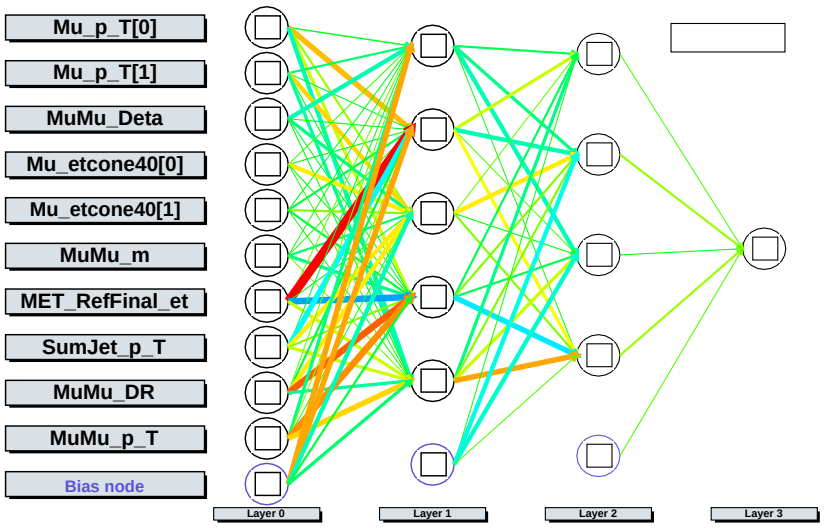


Figure 2.8: Example for a neural network: this network is a two-layer perceptron trained to identify Higgs decays with a Higgs boson mass of 170 GeV

mined on an independent control sample. On the other hand, the boosting method makes the decision tree a classifier that performs well "out-of-the-box", with little tuning required.

Bagging: An alternative way to increase the power of decision trees is "bagging": Reweighted training samples are constructed by randomly picking events from the training sample according to their weight, and also allowing an event to be picked more than once. The constructed samples then have the same probability distribution as the parent sample, and the combination of the different trees then gives a more statistically stable classifier. An additional possibility is to only use a random set of variables to choose at each split. This can in some cases increase the strength of splits further on. The decision tree classifier using bagging and a random choice of variables at each node is named "bagged randomized trees" (BRT) in this thesis.

2.3.4 Multilayer Perceptrons

The field of neural networks is wide, yet the multilayer perceptron (MLP) commonly used in high energy physics is one of the simplest examples: Each neuron of each layer is wired with each neuron of the preceding layer: in figure 2.8 one such wiring is depicted. Each neuron applies a sigmoid ($tanh$) function on the sum of its weighted input signals, the result is passed on to the next layer.

A neural network with N layers can therefore be described by $N + 1$ matrices: Let $\boldsymbol{x}$ be the input variables, and $\mathbf{W}_i$ the matrix of the ith layer,

then the output of a neural net is

$$\mathbf{W}_2 \cdot \tanh\left(\mathbf{W}_1 \cdot \tanh\left(\mathbf{W}_0 \cdot \boldsymbol{x}\right)\right)$$

Here the tanh function is understood to apply elementwise to each vector.

Training a neural network is achieved by setting the inputs to known (classified) events and then adjusting the strength of the network connections. This is done using the back-propagation algorithm [24]. A properly trained neural network then maps a score to each point in phase space that is high in signal-like regions.

The advantage of neural networks is that they are less susceptible to overtraining than boosted decision trees. Also, their output is more continuous with changes in the input variables. However, neural networks need large training sets to obtain good performance, a major problem for many applications.

2.3.5 The Fisher Classifier

The Fisher Classifier or Fisher method is simple: its output is the scalar product between the vector of variables - minus its training mean - and a coefficient vector $\boldsymbol{w}$ [25]. It can be visualized as a line in variable space onto which each event is projected. One end of the line corresponds to signal, the other to background. The method is trained by maximizing the Fisher criterion $J(\boldsymbol{w})$ over different coefficient vectors $\boldsymbol{w}$, where $J(\boldsymbol{w})$ is

$$J(\boldsymbol{w}) = \frac{|m_1 - m_2|^2}{s_1^2 + s_2^2}$$

Here m_1 and m_2 are the means of the signal and background samples, and s_1, s_2 their variances. This optimization problem can be solved in closed form, this yields a fast training algorithm.

Since the number of free parameters is rather small, it is impossible to overtrain the Fisher classifier. And despite its obvious shortcomings, it still gives surprisingly good results, expecially when other classifiers fail because of small training samples. It is therefore a good indicator of such problems - if a more sophisticated classifier yields worse results than Fisher on the test sample and a better performance would be expected, some aspect of training probably has to be improved.

2.3.6 Other Classifiers

Other classifiers tested in this thesis have been Support Vector Machines, gaussian maximum Likelihood estimators and probability density estima-

tors [22]. Since these classifiers yielded worse results than either BDTs or MLPs in several tries, they are not listed in the analysis and the results. [10]

[10] This of course does not exclude that for data unlike the examined samples, these classifiers may be better.

3. Experimental Setup

The ATLAS experiment is situated at the Point 1 of the Large Hadron Collider Ring. In the center of the detector protons are collided with protons at high luminosity every 25 ns. The resulting fragments and particles are then measured by the ATLAS Detector, the data written to tape, processed, and analyzed.

3.1 The Large Hadron Collider

The Large Hadron collider is a synchrotron 27 km in circumference that accelerates protons from injection energies of 450 GeV to 7 TeV in both directions. It is the last stage of a chain of pre-accelerators (figure 3.1). The protons are kept on course and focussed by superconducting magnets. At the experiment site of ATLAS, the two beams are brought to collision with a center of mass energy of 14 TeV and an instantaneous luminosity of $10^{34} cm^{-2} s^{-1}$, approximately a factor of 100 more than the $TeVatron$ accelerator [27, 28].

The protons in LHC are kept in 2808 "bunches" of about 10^{11} particles. This leads to one bunch crossing each 25 ns at the experiment, each of which leads to several interactions. The LHC finished its cooldown phase in August 2008, and injected the first beam on September 10. First physics data is expected in Spring 2009.

3.2 The ATLAS Experiment

3.2.1 Design

The ATLAS experiment is layered around the LHC beam interaction point. Each layer measures different properties of the created particles, and by combination of the layers particle types and momenta can be reconstructed. For detailed technical references, see [27].

The layers of ATLAS follow the traditional outline (see figure 3.2):

- Inner tracker
- Electromagnetic calorimeter

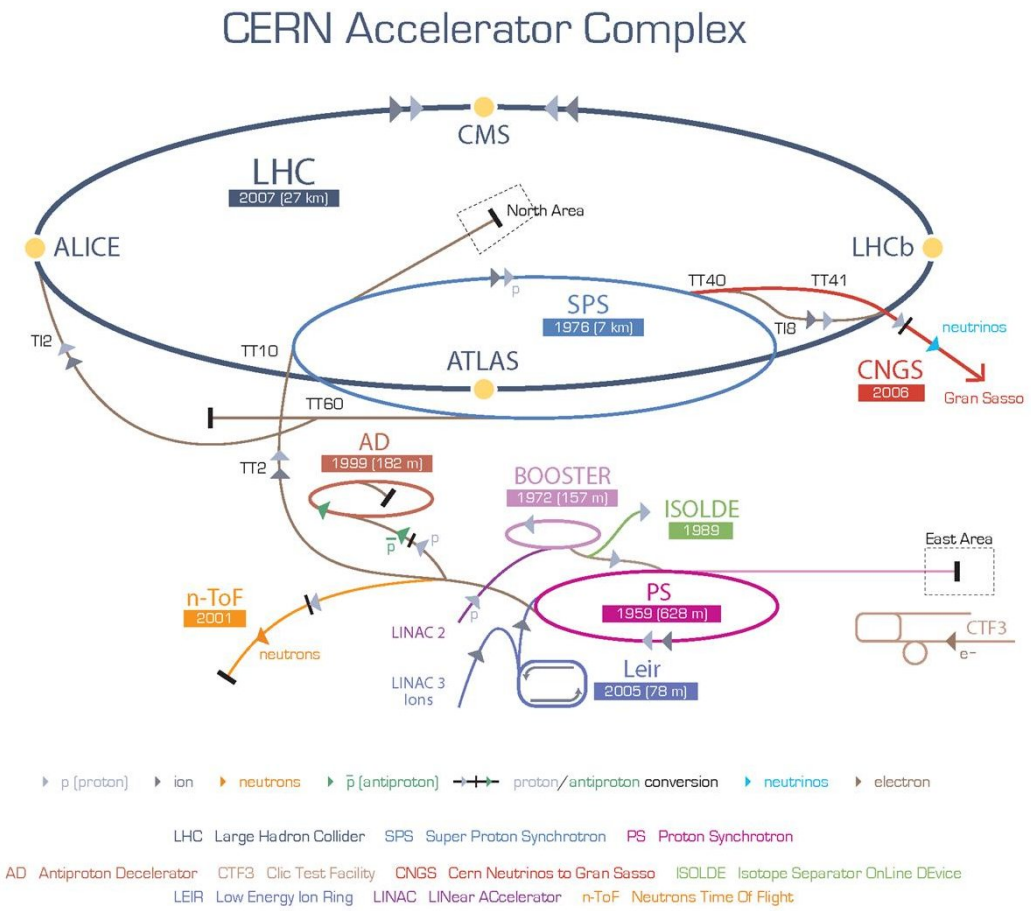


Figure 3.1: Overview over the CERN Accelerator Complex [26].

- Hadronic calorimeter
- Muon spectrometer

The inner detector with a radius of 1.15m is enclosed by a solenoid magnet, creating a homogenous field of 2T. This curves the tracks of charged particles, the radius of curvature then being proportional to the transverse momentum of the particle. The inner detector itself consists of three layers: First silicon pixels 50μm$\times300\mu$m large for good vertex resolution, then longer silicon strip detectors 80μm$\times$12cm for further tracking, and finally straw-tubes with a resolution of approx. 200μm for tracking and transition radiation measurement. The inner detector can resolve the transverse momentum of a 20 GeV muon flying ortogonal to the beam in good conditions to 1.4%.

The electromagnetic calorimeter is built using liquid argon technology, the hadronic calorimeter is a scintillator-tile structure. In the endcaps, the

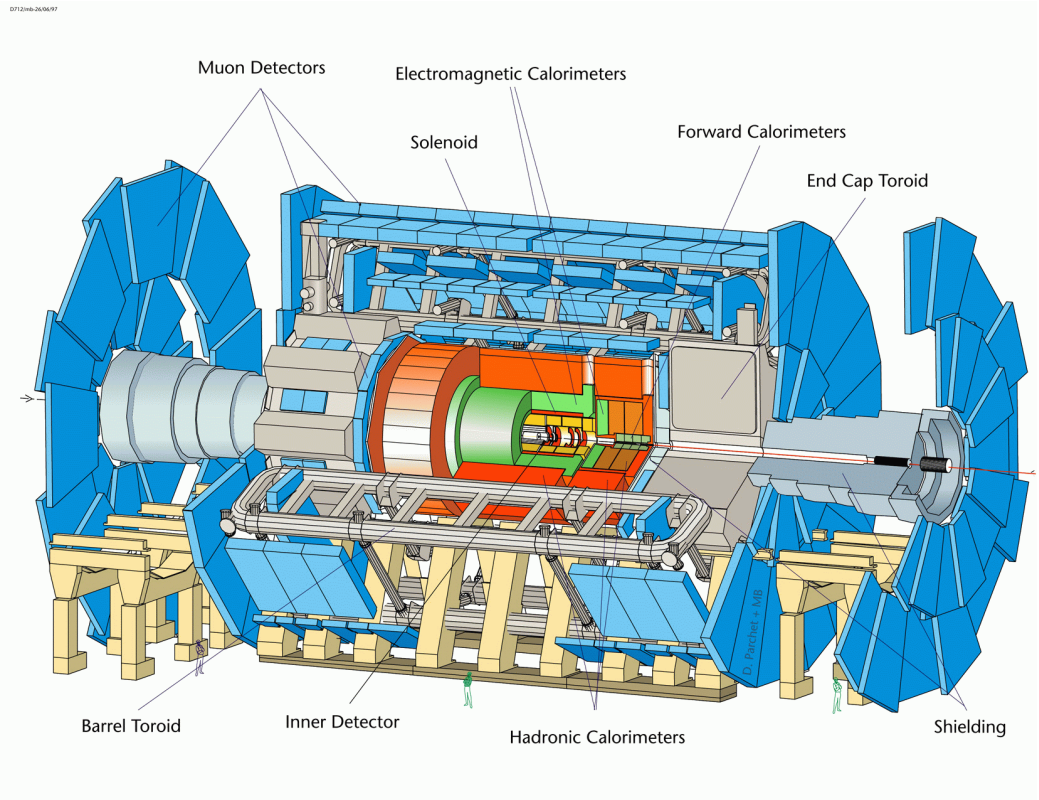


Figure 3.2: Overview of the ATLAS Detector.

hadronic calorimetry is as well taken on by liquid argon. Only muons can transverse the calorimeter mostly unhindered, all other particles are stopped and their energy is measured. The combined uncertainty on jet energy measurements is difficult to estimate without collision data, and is conservatively given [29] at $67\%/\sqrt{E_{jet}[\mathrm{GeV}]}$ for resolution uncertainty and 10% for systematic scale uncertainty.

The whole calorimeter structure, with a radius of 4.25m, is surrounded by the muon spectrometer. The muon spectrometer at ATLAS consists of an air-filled toroid magnet system, in which there are three layers of tracking stations in every direction. Muons are tracked in the barrel and endcap using monitored drift tubes. Triggering is achieved by fast-acting resistive plate chambers in the barrel region and thin gap chambers in the endcaps. The open structure of the muon spectrometer reduces multiple scattering, and helps improving muon momentum resolution. The total radius of the muon spectrometer and therefore of ATLAS is about 11m. The combined muon resolution of the inner detector and the muon system is estimated [30] to be about 3% in the region of interest to this analysis ($10 - 100$ GeV).

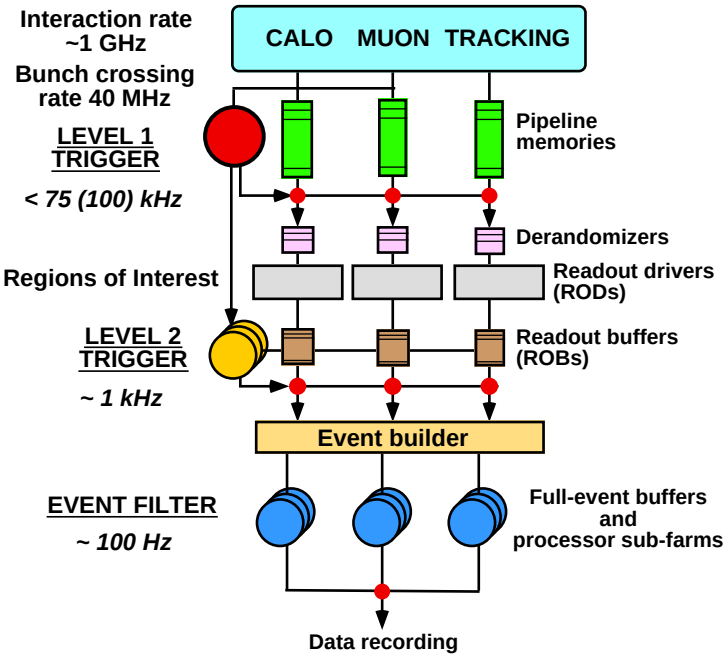


Figure 3.3: The ATLAS Trigger Levels and approximate rates.

3.2.2 Trigger

Since there are approximately 40 million bunch crossings per second, and an electronic record of one event takes several megabytes, recording every bunch crossing is far beyond current technology. Therefore, three levels of event triggers are used (figure 3.3). The first level is built into the calorimeter and muon system hard- and firmware. If an event seems sufficiently interesting, this system defines "regions of interest". The level two trigger system - still underground, close to the detector - fetches the readout buffers from these regions and checks the event for a second time. Then, it is passed topside to a computing farm where a preliminary reconstruction of particles is performed. After the last trigger, the event rate is reduced to approximately 100 events per second, translating to about 500 MB/s.

3.2.3 Event reconstruction

To reconstruct the track of a charged particle, detector hits must be connected. For muons, tracks from the inner detector and the muon spectrometer must be matched, and the output of several reconstruction algorithms are available. For this analysis, a recommended algorithm, the "STACO" algorithm (for an evaluation see [31]) has been used.

Still, there are two other possibilities to consider: A stable hadron, car-

rying most of the momentum of a hadronic jet, can in rare cases "sneak" through the hadronic calorimeter and register as a muon. Also, if a jet has high enough energy, it can "punch through" the hadronic calorimeter and cause enough hits in the muon chambers to fake a muon. Since especially the first case is difficult to detect, in an analysis on real data an additional matching with QCD data must be done to model these backgrounds.

Missing Transverse Energy During event reconstruction, the missing transverse energy, the negative vector sum of all measured (and estimated) transverse momenta in the calorimeter is calculated and corrected for known effects. The transverse momenta in the calorimeter are calculated using the location and amount of the deposited energy. The missing transverse energy is is a quantity extremely sensitive to detector effects. The current estimate on the accuracy in very good conditions is about 5 GeV [32].

3.3 Detector Simulation

The complex nature of the physics events and the multitude of subdetectors in ATLAS makes an accurate simulation of the detector indispensable. The simulation for ATLAS is built using GEANT4 [33] and an accurate geometric model of the ATLAS detector. This simulation tracks particles through matter, incorporates a wide range of interactions and simulates the detector response. The full simulation of the detector with GEANT4 also takes a large amount of CPU time: simulating an average event on a current CPU (AMD Opteron 275) takes about half an hour. Since millions of simulated events are necessary for a full study, the CPU time necessary is in the order of decades. The solution is the use of the resources of the LHC Community Grid (LCG), a large network of computing centres dedicated to the LHC experiments. On the Grid many physics samples for ATLAS are already centrally produced. For this study, approximately 330000 events were simulated in full simulation in addition to the officially produced samples, this is the equivalent of about 19 years of CPU time. Without the resources of the LCG this would not have been possible.

Fast Simulation: Even with the resources of the LCG it is impractical to obtain samples of a reasonable size for some processes. One option to speed up the simulation is to use ATLFAST II [34], where the exact simulation of the calorimetetry with GEANT4 is replaced by a fast parametrized version. This reduced processing time considerably, and allowed an additional 1.5 million events to be produced for this analysis.

4. Statistical Methods for Uncertainties

In experimental statistics there are two different approaches to statistical analysis, the *Frequentist* and the *Bayesian*[35]. For more than a hundred years the question which method to use under which circumstances has been controversial. Since both are used in particle physics analyses, a careful examination of the differences of the methodologies is in order.

Firstly, the two approaches are presented. Then the actual statistical questions relevant to the analysis are clearly stated, and the answers proposed by the two schools of thought are examined. Finally the systematic uncertainties arising from imperfect knowledge of the detector are integrated into the statistical analysis method, and a concise prescription for evaluating an analysis is given.

4.1 Frequentists and Bayesian Approaches to Probability

4.1.1 Frequentist Probability

The frequentist probability of an event e is defined as the frequency of e in an ensemble of identical random experiments. This means one can only assign probabilites to outcomes of random experiments - assigning a probability to the event "it will rain tomorrow" is not possible. This means that one also can not assign probabilities to possible values of theory parameters.

To interpret an experiment measuring the (parameter) X, a frequentist can now give a 95% confidence level interval with the property: "No matter what the true value of X is, the probability that the confidence interval that will be obtained by this experiment covers it should be 95%". This means that if the experiment is repeated, in the limit 95% of the confidence intervals will cover the true value.[36]

There are several possibilities to construct such a confidence interval, and for the frequentist method to be accurate it is absolutely imperative that the method is chosen before the experiment. If the method can be chosen after

looking at the data, the probability of covering the true value, the "coverage", can be smaller than intended: the method can "undercover". This is especially important if it is not known beforehand if only an upper or also a lower limit will be given. In this case, a "unified method" which avoids this problem, proposed by Feldman and Cousins [37] is recommended [3].

The most common conceptual problem arising from a frequentist confidence interval concerns the interpretation of a single measurement. A 95% confidence interval does not mean that the true value lies in this particular interval at 95% probability. Connected with this is the fact that the choice of method can also lead to very different confidence intervals for the same confidence level, or confidence levels for the same intervals. One other problem also inherent in the "unified method", is that in experiments with known background there are cases where one can know that the obtained 95% confidence interval is probably in the 5% that do not cover the true value, and too optimistic limits are set. [36]

The frequentist construction does have the advantage of relative simplicity, at least applied to one-dimensional problems. The construction of a confidence belt can be performed in advance and the results tabulated.

4.1.2 Bayesian Probability

In strong conceptual contrast, Bayesian probability is a mathematical description of subjective beliefs. Its origins can be traced to questions about gambling, and the odds that are "fair" in a given betting situation.

The central tool of Bayesian Statistics is Bayes Theorem

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A) \tag{4.1}$$

Here $P(A|B)$ denotes the conditional probability of A subject to B, the probability of A given that B is true.

This formula allows a Bayesian observer to adjust its *prior* belief about A, specified by $P(A)$, according to the observation B. The multiplicative factor $\frac{P(B|A)}{P(B)}$ is called the "belief updating ratio". This is an extremely powerful tool, yet since every physicist can have a different prior belief about, for example, the existence of the Higgs boson and limits on its mass, it is difficult to give a consensus number for $P(A|B)$. On the other hand, Bayesian probability is not bound to the limits of frequentist statistics: Probabilities can be given for one-off events (like if it will rain tomorrow), decisions can be subjected to mathematical comparison, and uncertainties in measurement can easily be incorporated into the result. To summarize an experiment, a Bayesian observer can give a subjective "95% credibility interval": He would think a bet with the odds 20:1 for the value to be in this interval is fair. Also, in cases where the prior belief of different physicists differs greatly, one could

additionally publish the belief updating ratio, and enable everyone to draw their own conclusions. [36]

To summarize, in cases where we have clearly defined random experiments it is often advantageous to use frequentist statistics - we do not need any more prior information[1]. Several reasonable schemes exist to remove some of the problems of the unified method, for example the CLs method used by the LEP experiment [38]. Also, some mixed approaches exist to incorporate systematic uncertainties into frequentist statistics [3], yet most of them rely on a Bayesian understanding of the problem.

4.2 Statistical Questions

A search at a particle accelerator in channels where no mass can be directly reconstructed relies heavily on Monte Carlo simulation of events and their detector response. Simulated Standard Model processes are used as model background, and several signal samples are considered. Then signal regions with a good signal significance are specified, and real events falling into these regions are counted and compared to the Monte Carlo background prediction.

To correctly estimate the background in a signal region, each Monte Carlo sample has to be split in half between training and testing data. The training data is used for training classifiers and optimizing cuts, and the testing data - statistically independent of the training data - is used for the estimation. This is absolutely necessary, since the choice of signal region can depend on statistical fluctuations, and the significance can be considerably overestimated, especially in cases where background statistics are low.

The first statistical question concerns this estimation of the "Monte Carlo background". Since each background sample is finite, we do not know the probability distribution of the event generator exactly. Disregarding other uncertainties, we have to answer: how can we set a credible limit on background events in the signal region?

In addition to using the test data directly to estimate the background, several detector uncertainties are parametrized and the estimation repeated with accordingly distorted test data. The second question is twofold: how and how often do we distort the test data, and how can we credibly estimate the influence of the detector uncertainties on the analysis from the results?

In this analysis it is additionally assumed that the absolute normalization of the various backgrounds is kept under control by careful evaluation of sidebands, and these two sources of uncertainty are the only relevant ones to the estimation, then the third question is: how can I determine if I can claim

[1] Of course we have to know with "absolute certainty" that our description of the random experiment is correct

evidence, discovery or exclusion, and at what luminosity do I expect to do so if a signal is present?

To summarize:

- *Monte Carlo Uncertainty:* How can we set a credible limit on the background events?
- *Systematic Uncertainty:* How can we get from a parametrization of e.g. detector uncertainties to an uncertainty of the analysis result?
- *Measurement Significance:* When can we say to have found evidence, discovered, or excluded a signal?

It is interesting to note that the first two uncertainties do not intrinsically depend on the integrated luminosity if the analysis remains unchanged, although of course the understanding of sytematic uncertainties will increase with longer run time of the detector. Yet, this means that with a specific set of Monte Carlo samples and a specific uncertainty on detector parameters one can only reach a *limit significance* that is independent on run time of the detector. At the end of the analysis, this limit significance will be given for statistical and systematical uncertainties seperately as well as combined. For an extended analysis with real data, the first uncertainty will hopefully be much smaller than in this analysis

4.2.1 Estimating Statistical Uncertainty on Background

Setting limits on the expected background is particularly important for discoveries. If we want to estimate the probability density in a region R of phase space, the straightforward estimate of the integral of the probability density is $\int_R \rho = \frac{N_R}{N}$, where N_R is the number of Monte Carlo events in R and N is the total number of events in the sample[2]. This leads directly to the uncertainty statistics: In case of large N_R one can easily use Gaussian uncertainties, while for small N_R Poisson statistics must be used.

Setting limits to Poisson processes is traditionally a frequentist domain. The basis for all frequentist methods is the Neyman construction of a confidence belt [39].

The Neyman Construction

Firstly, for each possible "true" value one constructs a range of possible measured values that include the actually measured values with $N\%$ probability, with $N\%$ as the confidence level. This region is called the "acceptance region". Then one can obtain a confidence interval by drawing a line through

[2]This estimate is based on a discrete probablity density: One δ-function at each Monte Carlo event

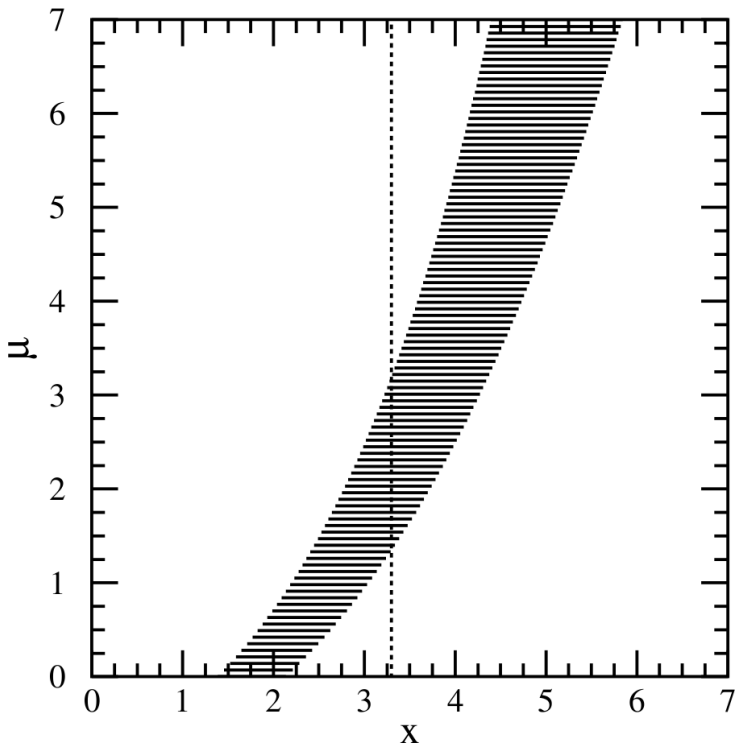


Figure 4.1: Example for construction and use of a frequentist confidence belt: the measurement of a theory parameter μ given the experimental (uncertain) result x for the value of μ. For each possibly true value of μ an acceptance region is constructed. This region must cover a fraction of possible experimental outcomes at least equal the confidence level in probability. On an experimental outcome one then draws a vertical line, the region of intersection then being the confidence interval

the actually measured value: Each true value for which the line lies in the acceptance region is not excluded at $N\%$ confidence.

This construction must be done in advance of seeing the data, and in this form is not fully specified: An acceptance region can be constructed to find an upper limit, a lower limit or a two-sided limit. Yet, also this decision must be made before seeing the data, since changing your mind and choosing between several confidence belts after looking at the data destroys the frequentist properties of the published confidence interval. [37]

The most popular form of actual implementation of the Neyman construction, the Feldman-Cousins method, avoids this, as it constructs only one confidence belt that can give one- or two sided limits, depending on the data. This method is also recommended by the Particle Data Group for setting limits [3].

The unified approach of Feldman and Cousins

The Neyman construction leaves out the construction of the acceptance region. Feldman and Cousins now propose [37] to order possible outcomes $N_{observed}$ by the fraction $P(\mu|N_{observed})/P(\mu_{best}|N_{observed})$ for a given "truth" μ, where μ_{best} is the truth that best fits $N_{observed}$. Then the acceptance region is determined by the set of top ranked $N_{observed}$, with a cut-off determined by the fact that the sum of $P(\mu|N_{observed})$ should be at least the desired confidence level.

Yet one extension is missing: since we have several background samples, each with events representing different cross-sections, we have to put limits not on the individual backgrounds but on the sum of all background samples.

Multidimensional Extensions of the Frequentist Methods

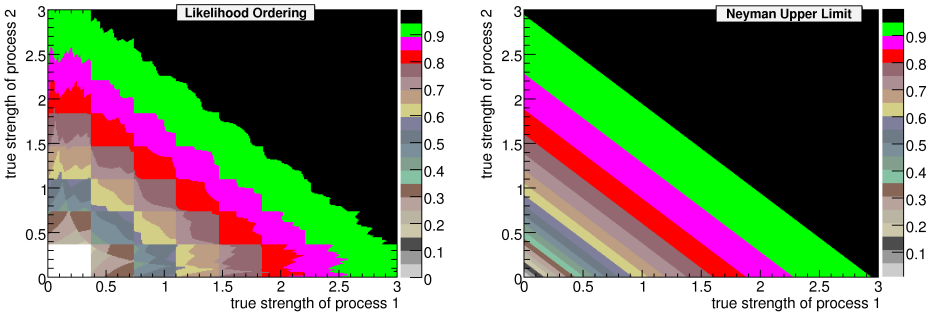
In the case of several background samples - especially if no or only a few Monte Carlo events remain after selection - it would be interesting to set a combined limit, a limit on the sum of all backgrounds. It is not really interesting for us if a certain event is from one background or another.

Multidimensional Neyman construction Firstly, one must exted the framework of the Neyman construction to n dimensions, so that we can handle n background samples. Now we have to construct an acceptance region for each n-tuple of true background strengths, given the observed number of Monte Carlo events, the n-tuple $N_{observed}$. We can calculate the sum of backgrounds v by scalar multiplication of the true background n-tuple with the n-dimensional weight vector. Now we are done: the confidence interval is the set of all v for which there exists an n-tuple of true values that sum up (by weight) to v and are in the acceptance region for the $N_{observed}$. This construction is completely analogous to the one-dimensional Neyman construction. It evidently gives a frequentist confidence interval, yet is as the original Neyman construction not fully specified.

Multidimensional Feldman-Cousins To extend the approach of Feldman and Cousins to more than one dimension is straightforward: The ranking of possible $N_{observed}$ given the true μ is still done as before, only with multidimensional N and μ. This has been implemented using the one-dimensional implementation of ROOT [40] as a starting point.[3].

Yet if we look at the example (a) in figure 4.2, there are some inconsistencies. In the figure, each point represents a "true" value, its colour indicating if it would be included in an acceptance region with a given confidence interval in the case where twice zero events were selected ($N_{observed} = (0, 0)$).

[3]Source code available on request



(a) multidimensional likelihood ordering (b) multidimensional Neyman construction - only upper limit

Figure 4.2: Comparison in two dimensions: The confidence belts for different confidence levels for two Poisson processes without background, with no event measured for each process.

For example, one notices that in the case of an 90% CL (the regions below the pink region being the acceptance region), the tuple $(2.5, 0.2)$, lying in the pink region, would be considered "accepted", yet the tuple $(2.45, 0.1)$, lying in the green region, would be excluded! Although this is not critical in this case, it raises doubts about the ordering principle adopted, and questions its application in more than one dimensions.

Multidimensional Poisson Upper Limit The problem of combining background limits can also be solved by an easier frequentist construction: Since we are exclusively interested in upper limits, the complex construction required by the ranking of Feldman and Cousins can be replaced by a simple integration: Instead of choosing the top ranked "observed" values until we reach the confidence limit, we sum up from the minimum weighted sum of backgrounds. This can further be simplified by not requiring the full n-tuple as "truth", but only the weighted sum v, and calculating the probability of v occurring directly via Poisson statistics. This method was implemented as well, the resulting confidence belts can be seen in 4.2 (b). It seems that this method handles the extension to more than one dimension much better, although it is of course less general.

For the simple case of the combination of two Poisson processes with equal weight and no observed events, the limit behaves as expected: the limit for the combined process is the same as for the single process. This is the case, since by superimposing two poisson processes with intensity x and y we get another poisson process of intensity $x+y$, and we can refer to the one-dimensional case. If we extend this case to poisson processes with different weights, still for simplicity assuming no events were found, the result is that the upper limit is determined solely by the background with the greatest

weight. This is mathematically correct, the method does never "undercover". Although, one would expect the limit to increase if additional backgrounds are added.

Bayesian Upper Limits

An additional approach would be to use a Bayesian technique for estimating the background from the background Monte Carlo.

Choice of prior: To obtain a posterior probability distribution we need a prior for the background estimate. One sensible option would be a uniform prior in the number N of expected events, cut of at the number of events in the sample. One other prior, recommended for its objectivity, is the Jeffreys prior; in this case it is $P(N) \sim 1/\sqrt{N}$ [41]. This prior is unfortunately not applicable in cases where no events are found, since the posterior found by application of Bayes theorem can not be normalized. Since in addition the Jeffreys prior would always be less conservative than a uniform prior in N, the uniform prior was chosen.

Convolution: We now have for each background i found n_i events in the signal region. Bayes law together with a uniform prior gives us then the posterior distributions

$$P(\lambda_i|n_i) \sim \frac{1}{n_i!} e^{-\lambda_i} \lambda_i^{n_i}$$

We now have to get the distribution of the sum of the backgrounds. Let now λ be the sum, w_1, w_2 the weights and λ_1, λ_2 the strengths of two backgrounds.

$$\begin{aligned} & P(\lambda = w_1\lambda_1 + w_2\lambda_2|n_1, n_2) = \\ = & \int\int \delta(\lambda - w_1\lambda_1 - w_2\lambda_2) P_1(\lambda_1|n_1) P_2(\lambda_2|n_2) \, \mathrm{d}\lambda_1 \mathrm{d}\lambda_2 \\ = & \frac{1}{w_1} \int P_1\left(\frac{\lambda - w_2\lambda_2}{w_1}|n1\right) P_2(\lambda_2|n_2) \, \mathrm{d}\lambda_2 \\ = & \frac{1}{w_1 w_2} \int P_1\left(\frac{\lambda - x}{w_1}|n1\right) P_2\left(\frac{x}{w_2}|n_2\right) \mathrm{d}x \\ = & \frac{1}{w_1 w_2} \left(P_1\left(\frac{\cdot}{w_1}|n1\right) * P_2\left(\frac{\cdot}{w_2}|n2\right)\right)(\lambda) \end{aligned}$$

In the last line the convolution operator $*$ is used, the two places where the corresponding convolution variable has to be put in is denoted by a $\cdot$ dot. It follows that we can simply convolve the individual distributions together, normalize and obtain the final distribution. This seems computationally difficult, but the convolution operation can be sped up considerably by fast

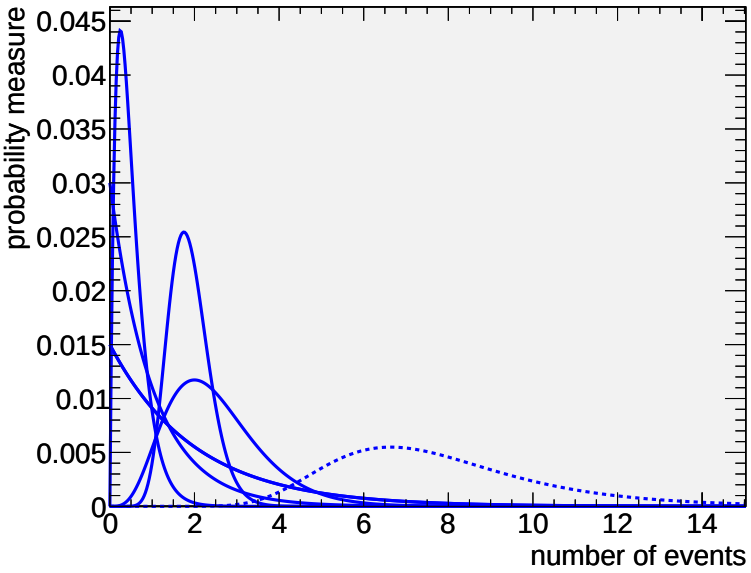


Figure 4.3: Example for combining five backgrounds with different weights. The weights of the backgrounds- the ratio of Monte Carlo to predicted events - were $\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1$ and 2, and the respective numbers of Monte Carlo events were $14, 1, 4$ and twice 0. The posterior probability distributions from the individual backgrounds using a uniform prior are plotted in solid blue, and the resulting probability distribution for the predicted events is dotted.

fourier transformation, and a package for this purpose (RooFit) is already available in ROOT [40].

This method is guaranteed to give consistent results in the framework of Bayesian reasoning, and the resulting posterior distribution can easily be used for combination with systematic uncertainties. An example is given in figure 4.3.

Choice of Method: Both the newly presented frequentist multidimensional Poisson upper limit and the Bayesian method would be suitable for the analysis. Since, however, the Bayesian method allows direct combination of the systematic with the statistical uncertainties and also provides a simple method of combining two analysis channels the Bayesian method is chosen for this thesis. Additionally, it is slightly more conservative, and is therefore a sensible choice if avoiding false positives is a priority.

4.2.2 Estimating Systematic Uncertainties

To accurately determine the sensitivity of an analysis, many classes of systematical uncertainties have to be considered. They can be separated in

detector-bound systematic uncertainties and theoretical systematic uncertainties:

Theoretical Systematic Uncertainties

- uncertainty on process cross-sections
- uncertainty in parton distribution functions
- correctness and technical accuracy of the Monte Carlo generator

In this thesis the theoretical uncertainty on cross-sections is for simplicity assumed to be under control by auxiliary measurements in sidebands. In case of an analysis with real data, the uncertainties of the auxiliary measurements can be combined with the other uncertainties following the process for the detector systematic uncertainties.

Detector Systematic Uncertainties The systematic uncertainties most relevant to this analysis are:

- measurement uncertainty on total luminosity
- resolution of particle momentum
- calibration of momentum scale
- calorimeter resolution and/or noise
- missing transverse energy calculation

For the total luminosity of the data good estimates will be available. Also, the Z peak can be used as a standard candle of normalization, although this introduces dependencies on theoretical extrapolations and on the muon resolution and reconstruction efficiency. Still, it is expected that these uncertainties are either small or easily detectable. For further discussion on Z peak normalization, see [42].

The systematic uncertainties on the momentum resolutions will also decrease once the resolutions are measured, however, the analysis as it stands has still to incorporate these uncertainties. When improved estimates of the resolution parameters are available, the analysis will have to be repeated using the updated resolution, and the systematic error will decrease accordingly.

Calculating the Systematic Uncertainty: The usual approach would now be to consider each systematic uncertainty individually, "smear" the data by $\pm\sigma$ and then take as the total uncertainty the root of the sum of squares of individual cross-section shifts. But since uncertainties are not independent - the muon scale, for example, affects both muon momentum and missing transverse energy - this approach does in general not give correct estimates [43]. In order to obtain a better estimate, additionally the following combined approach was carried out:

Combined Systematic Uncertainties: Firstly, each systematic uncertainty was parametrized using available $\pm\sigma$ values (see section 3.2 and table 4.1). The magnitudes of the d systematic uncertainties then span a d-dimensional "uncertainty space". Secondly, the analysis was repeated for 400 points in this uncertainty space randomly chosen according to a normal distribution around zero, with the events accordingly distorted. For each such point and each cut flow entry this gives a distribution of event numbers. This procedure corresponds to Bayesian numerical "marginalization", a numerical evaluation of

$$P(x) = \int_{\theta} P(x|\theta)P(\theta)$$

Here θ represents the parameters for the systematic uncertainty. This means, the resulting distribution approximates the probability distribution of the event count, if we assume that the detector systematic uncertainties are distributed as described.

Using this distribution, we have two possibilities: we can either directly derive confidence intervals from this distribution, which can only be used to a confidence level of about 90% since we have only 400 sample points, or we can assume for simplicity the total systematic uncertainty can be described by a Gaussian. In the latter case, we can take the standard deviation of the distribution to be the estimate of the systematic uncertainty. To achieve approximately 5% accuracy in the resulting standard deviation 400 points were chosen, since a Monte Carlo integration with N points has the asymptotic error of $1/\sqrt{N}$ in any number of dimensions, in this case this yields $1/\sqrt{400} = 5\%$. This uncertainty is small enough to be disregarded in this case.

In the results section, these two methods for estimating systematic uncertainties will be compared.

4.2.3 Significance

An experiment at a collider takes data, the relative statistical uncertainty of a finite data sample decreases. Thus, if enough data statistics are available,

systematic uncertainty	1σ value	reference
muon resolution	3%	[30]
muon scale	3%	[30]
jet resolution	$67\%/\sqrt{E[\text{GeV}]}$	[29]
jet scale	10%	[29]
missing transverse energy resolution	5 GeV	[32]

Table 4.1: 1σ values of systematic uncertainties used for "smearing"

systematic uncertainties can as well be a limiting factor. For a particular analysis with a specific set of Monte Carlo samples we therefore have an "expected maximal signal significance" for all types of uncertainties, which is the expected significance of the signal in case of infinite detector statistics. These expected significances can be read off from the background probability distribution function: the integral from the expected signal+background value to positive infinity.

To obtain a combined probability density function (p.d.f.) for the background we convolute the statistical and systematical uncertainties together, and then evaluate the integral. Also, for combining the two channels we can convolute the background estimates and add the signal estimates - it is actually not possible to reliably combine experiments or uncertainties if a p.d.f. or a likelihood function is not available. Having followed a Bayesian approach, this integral can now easily be interpreted: it is the probability of the observed excess being caused by background, in light of the known uncertainties included in the analysis.

For the estimate of the final expected significance a multitude of good tools and approaches is available. In this thesis the S_{cP} significance [44] has been used, which calculates the significance statement using Poisson statistics and takes an uncertainty term on the background. This method is widely used, and yields sensible results. However, in cases with high systematic uncertainty on the background, the value returned is negative.

5. Search for the Higgs Boson

In this section the search for a standard model Higgs Boson in the Higgs $H \to W^+W^-$ channel is described. Only the events where both W bosons decay into muons are considered. Both a classical cut analysis and multivariate techniques are used and compared, and several systematical uncertainties are considered.

5.1 Procedure

To find a weak signal in the presence of strong background is challenging, and care is needed not to overestimate the power of the analysis. To determine the sensitivity correctly all Monte Carlo samples have been split into training and testing parts. For the training of multivariate classifiers and even for the optimization of the analysis cuts, only the training samples have been used: first tests have indicated that not only multivariate classifiers but also normal analysis cuts can be "overtrained".

In the following, preparations for the analysis are given: the event model is described, the relevant topologies and features of the Monte Carlo data are shown. The preselection for both cuts and multivariate classifiers is presented, as well as the splitting into the gluon-gluon and VBF channels.

Both multivariate classifiers and cuts are optimized to select a "signal region" with maximal discovery significance

$$\frac{\#\text{ of signal events}}{\sqrt{\#\text{ of background events}}} = \frac{S}{\sqrt{B}} \tag{5.1}$$

As an additional tool for optimization of the cuts the ScP significance [44] has been used, which corrects the significance statement using Poisson statistics. In cases where only low Monte Carlo statistics for B is available this method reduces fake "statistical peaks" in the significance if B falls close to zero and S is small as well. The training, visualisation and initial performance evaluation of the multivariate classifiers is shown, and the dependence of the performance on parameters tested.

After the training, the multivariate classifiers as well as the cuts are evaluated using the test data. The numerical results of all analyses are given,

together with the statistical, systematical and combined uncertainties, the expected maximal significance and, if applicable, the expected discovery luminosity.

5.2 Event description

In the case an event is triggered in the ATLAS Detector, the full detector is read out and processed in many steps to analysis data. This analysis data is then converted to ROOT [40] trees, and for each event several variables are defined to describe the event.

- four-momentum of the dimuon system $p_{\mu\mu} = p_{\mu,1} + p_{\mu,2}$
- invariant mass of the dimuon system $m_{\mu\mu} = \sqrt{p_{\mu\mu}^2}$
- opening angle of the muons in the transverse plane $\Delta\phi_{\mu\mu}$, the pseudorapidity $\Delta\eta_{\mu\mu}$ and the distance of the muons in the $\phi_{\mu\mu}$-$\eta_{\mu\mu}$ plane $\Delta R_{\mu\mu} = \sqrt{\Delta\phi_{\mu\mu}^2 + \Delta\eta_{\mu\mu}^2}$
- angle between the dimuon system and the missing transverse energy $\Delta\phi_{\mu\mu,/E_T}$ in the transverse plane
- transverse mass of the dimuon/missing transverse energy system $m_{T,\not{E}_T/\mu\mu} = \sqrt{2/E_T p_{T,\mu\mu}(1 - cos(\Delta\phi_{\mu\mu,/E_T}))}$
- $\sum_{jet} p_T = \sum_{jet,p_{T,jet}>15\ \text{GeV}} p_{T,jet}$; the transverse momentum of a jet is the summed transverse momenta of the particle tracks associated with the jet. This definition was chosen to match the usage in [45]
- the number of jets with transverse energy $E_T > 30$ GeV. This is also chosen so the 0 jet analysis is equivalent to [45].
- the number of electrons with transverse momentum $p_T > 30$ GeV

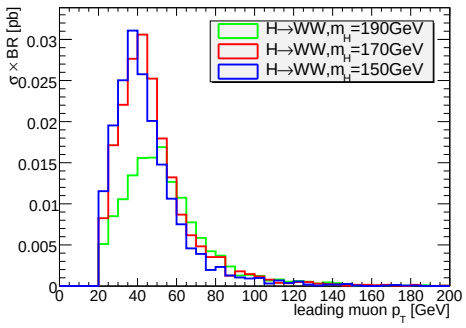
5.3 Process Topologies

In the following, the most important features of the signal and background samples are presented.

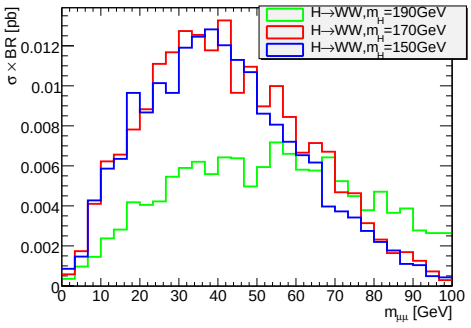
5.3.1 Signal Structure

In $H \to W^+W^-$ decays the most interesting kinematic feature is the spin correlation between the W^+ and the W^- as mentioned in section 2.2.3: the opening angle of the two muons from the Higgs boson decay tend to small

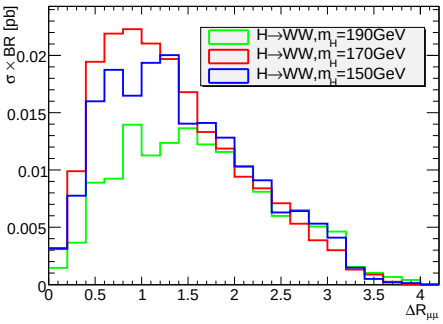
$$H \rightarrow W^+W^- \rightarrow \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$$



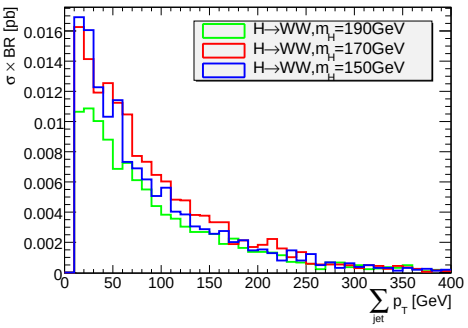
(a) transverse momentum p_T of the leading muon



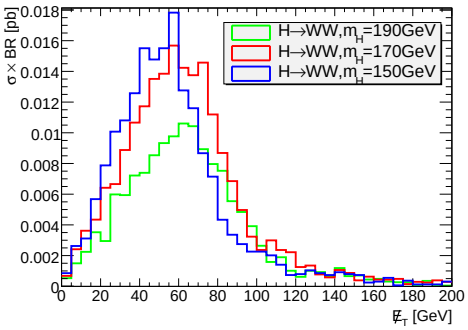
(b) invariant mass of the two leading muons



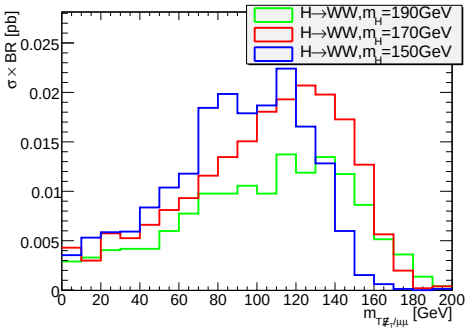
(c) distance of the two leading muons in the η-ϕ plane



(d) sum of the jet transverse momenta



(e) missing transverse energy



(f) transverse mass of the missing transverse energy and the transverse momentum of the two leading muons

Figure 5.1: Histograms of $H \rightarrow W^+W^- \rightarrow \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$ decays with $m_H = 150, 170, 190$ GeV after preselection.

values. In figure 5.1c one can see that the correlation is strongest for a Higgs boson mass of 170 GeV and is reduced as the Ws get virtual, or respectively have higher momenta.

5.3.2 Background Processes

In the following the most important background processes and their most relevant features are described.

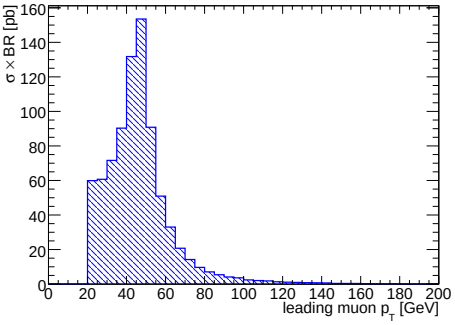
$Z/\gamma^* \rightarrow \mu^+\mu^-$: The dimuon production via Z/γ^* is the dominant background of this analysis channel, with a very high cross-section of 3.35 nb after selecting two reconstructed muons and restricting the invariant mass to $m_{\mu\mu} > 15$ GeV. Luckily, good separation of signal and backround seems possible, since the kinematic properties of Drell-Yan decays as direct $\mu^+\mu^-$ decays differ markedly from W^+W^- decays. The most significant distinction is the missing transverse energy (figure 5.2e). As Higgs boson decays produce two neutrinos there should be a significant amount of missing transverse energy, whereas in the Drell-Yan process there is only a small value of missing transverse energy expected. Secondly, the Drell-Yan spectrum of the dimuon mass is quite distinctive with the Z peak as seen in figure 5.2b, yet there is still a significant overlap, since the cross-section is so much larger. Thirdly, the muons from the Z tend to be emitted back-to-back, so the same separation criterion as in direct W^+W^- production can be used (see figure 5.2c).

QCD $gg/qq \rightarrow W^+W^-$: The direct QCD W^+W^- production is the most challenging background to this analysis: The kinematic properties of W^+W^- decays are very similar to the Higgs boson decay. Yet, there are still some detectable differences. For one, if the Higgs boson mass is close to twice the W mass, the Ws tend to be closer to the mass shell than the Ws produced directly. This leads to a peak of the muon transverse momenta at approximately 40 GeV (see figure 5.1a). In the case of the VBF channel the W^+W^- background is less important, since the two characteristic high η jets are absent. Finally, strongest separation is given by the opening angle of the muons, since the spin correlation is absent in direct W^+W^- production (figures 5.1c, 5.3c).

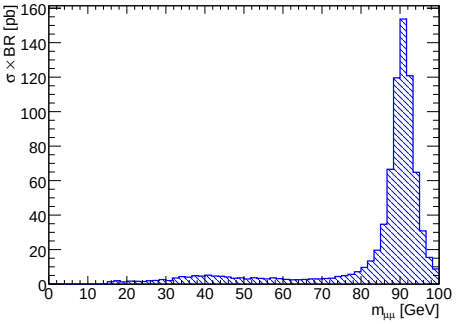
Other diboson productions like WZ and ZZ are also considered, yet their contribution is small compared to the major backgrounds.

QCD $t\bar{t}$ production: Top quark decays pose a significant problem to this channel. The NLO prediction for the total $t\bar{t}$ cross-section at the LHC is 833 pb, and the cross-section for decays which include at least one lepton is 450 pb. This is much less than the Drell-Yan background, yet it is much

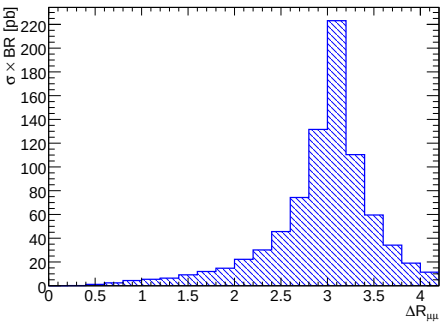
$$Z/\gamma^* \rightarrow \mu^+\mu^-$$



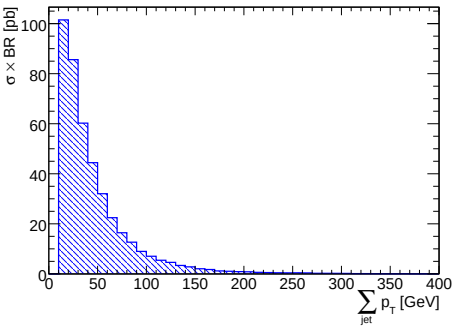
(a) transverse momentum p_T of the leading muon



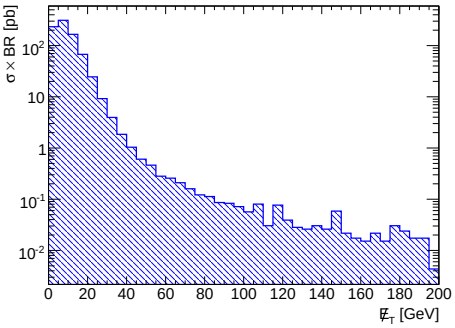
(b) invariant mass of the two leading muons



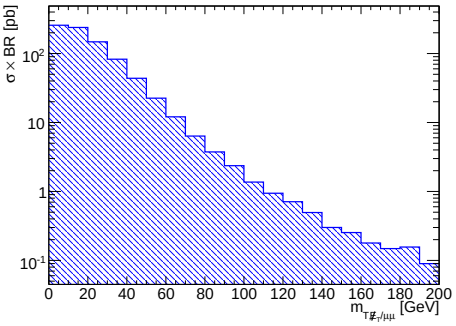
(c) distance of the two leading muons in the η-ϕ plane



(d) sum of the jet transverse momenta



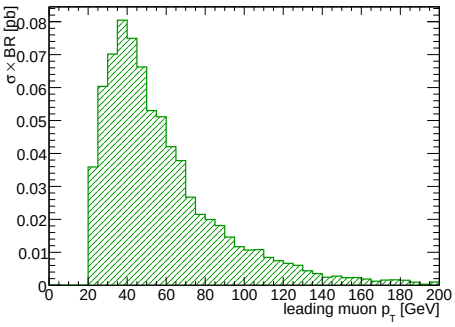
(e) missing transverse energy



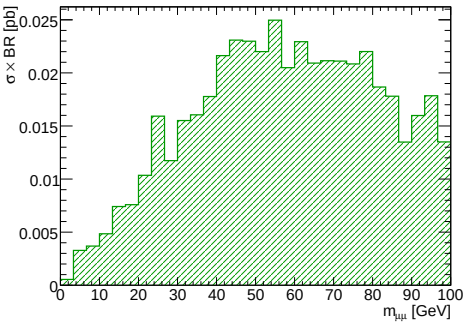
(f) transverse mass of the missing transverse energy and the transverse momentum of the two leading muons

Figure 5.2: Histograms of $Z/\gamma^* \rightarrow \mu^+\mu^-$ decays after preselection.

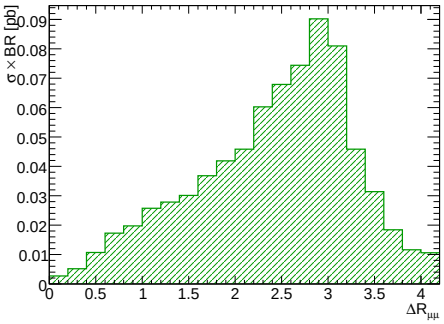
QCD W^+W^-



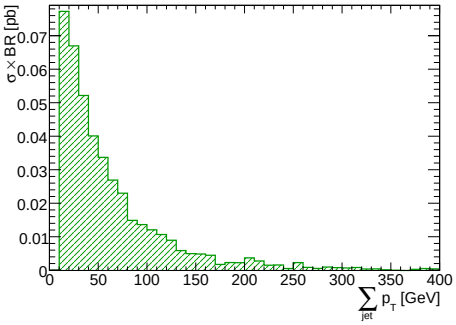
(a) transverse momentum p_T of the leading muon



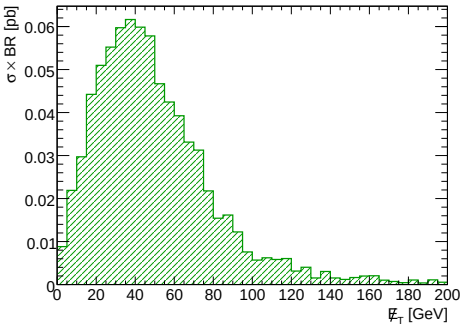
(b) invariant mass of the two leading muons



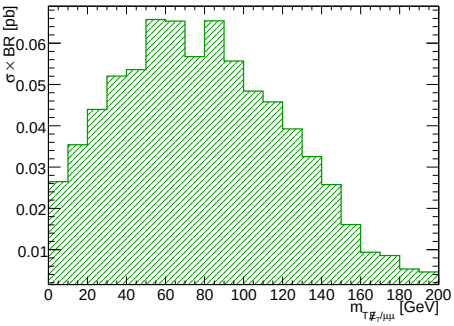
(c) distance of the two leading muons in the η-ϕ plane



(d) sum of the jet transverse momenta



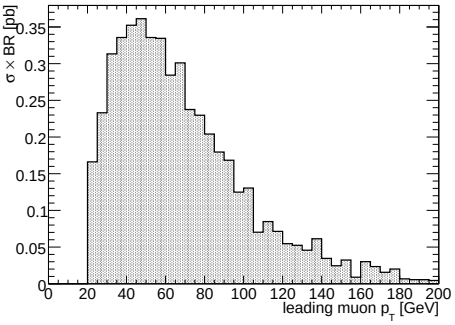
(e) missing transverse energy



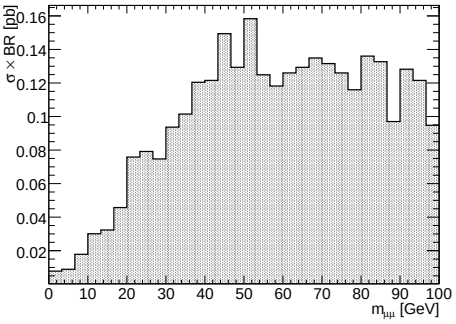
(f) transverse mass of the missing transverse energy and the transverse momentum of the two leading muons

Figure 5.3: Histograms of QCD direct W^+W^- production after preselection.

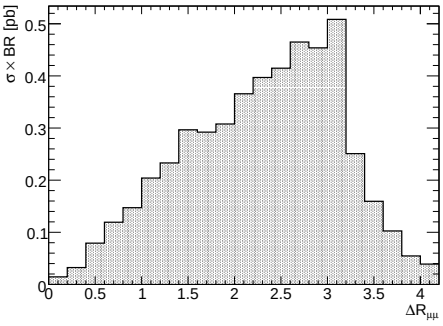
QCD $t\bar{t}$



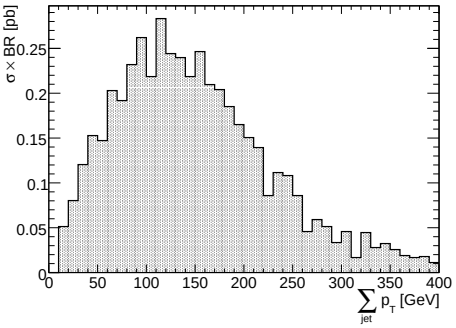
(a) transverse momentum p_T of the leading muon



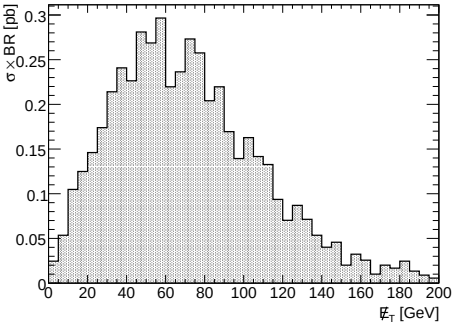
(b) invariant mass of the two leading muons



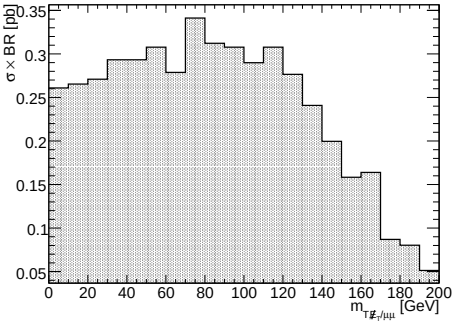
(c) distance of the two leading muons in the η-ϕ plane



(d) sum of the jet transverse momenta



(e) missing transverse energy



(f) transverse mass of the missing transverse energy and the transverse momentum of the two leading muons

Figure 5.4: Histograms of QCD $t\bar{t}$ production after preselection.

more difficult to separate, since the signature can be quite similar to the VBF Higgs boson production channel. As already mentioned, the b-jets resulting from t decays could still be detected using b-tagging, yet this will probably remain difficult until much more experience with actual data has been made.

W+jets There is a multitude of QCD processes that also produce a W boson. Most important for this analysis are those also producing a muon, namely single top $W + t$ and $Wb\bar{b}$ production. However, the cross-sections of these processes are rather small. More significant is the probability of fake muons from QCD multijet events, as mentioned in section 3.2.3. In this analysis, several samples from Monte Carlo were used as a baseline, but the number of Monte Carlo events passing the preselection cut is still very small, therefore no plots of this background are shown. For an analysis with real data it would be essential to use data to estimate the shape of this background, and then use a fitting procedure to estimate the strength of this background in the signal region.

5.4 Preselection

To reduce the amount of data processed and to limit the influence of systematic uncertainties that only affect events far from the signal region a preselection is applied to the data as follows:

- We require two well-reconstructed oppositely charged muons tracked both in the inner detector and the muon system, reconstructed by the STACO algorithm.
- The muons have to have a minimal transverse momentum of $p_T > 20$ GeV for the leading and $p_T > 15$ GeV for the trailing muon.
- The muons have to be measured in the muon detector acceptance region: $|\eta| < 2.5$.
- The muons have to be isolated: the amount of energy measured in the calorimeter in a cone of 0.4 radius around the muon in the η-ϕ plane must not exceed 5 GeV
- Since the studied Higgs processes have only two muons in the final state, a muon veto for additional muons above 10 GeV is introduced. This cut has an 98% efficiency for the signal, yet rejects 21% of the $t\bar{t}$ background.

Since a bug in the used software version made it impossible to evaluate trigger decisions, these limits ensure that the events are triggered with high

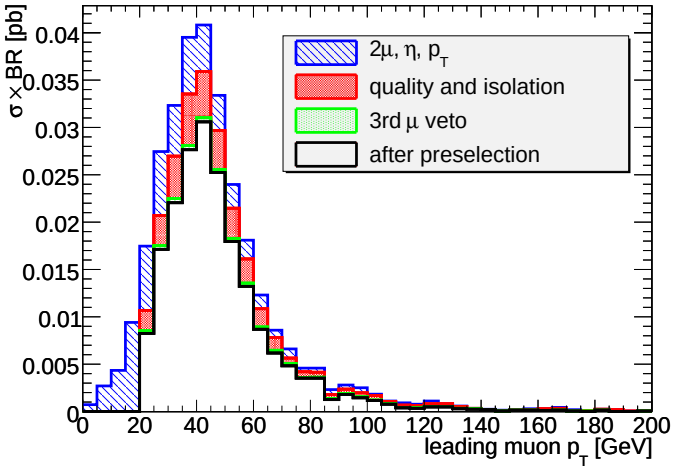


Figure 5.5: Histogram of the leading muon transverse momentum for $H \to W^+W^- \to \mu^+\nu_\mu\mu^-\bar{\nu}_\mu$ decays with $m_H = 170$ GeV: blue events are removed by the η and p_T cuts, red events are removed by quality and isolation cuts, and green events are removed by a veto of a third muon. The white events pass the preselection cuts.

probability, since we can combine the trigger for a muon higher than 20 GeV with the trigger on two muons greater than 10 GeV. Therefore, the trigger is ignored in the following.

In figure 5.5 the effects of the preselection cuts on the momentum of the leading muon can be seen. The signal is not significantly reduced, yet the confidence in the reconstructed particles is much increased. For comparison, the effect of preselection on the Drell-Yan background is given in figure 5.6.

Limited Monte Carlo events: At this point it is necessary to consider the finite amount of Monte Carlo statistics available. The number of events seems sufficient to make well-founded statements, yet in the later stages of the analysis small numbers of Monte Carlo events are often encountered. In table 5.1 two numbers are given: first the total $\sigma \times BR$ for the used sample after preselection, and second the fraction $f = \frac{\sigma \times BR}{\#\text{of MC events}}$. This fraction is evidently constant throughout the analysis, and can be seen as the cross-section represented by one Monte Carlo event. An estimate for the statistical error for a cross-section is therefore $f \cdot \sqrt{N_{MC}}$, and will be given in the following. A more accurate estimate of the statistical error is done for the results as described in section 4.2.1. Uncertainties on the total sample cross-section are disregarded for this analysis, since they affect all compared classifiers and cuts equally.

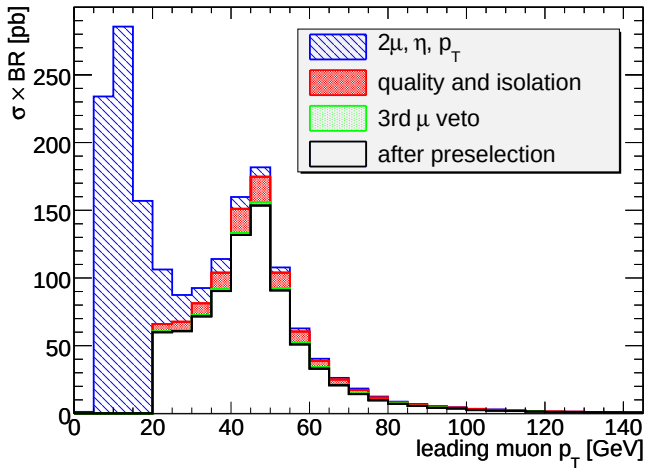


Figure 5.6: Histogram of the cut effects on the leading muon transverse momentum for the Drell-Yan background. It is clearly visible that the low-momentum part of the spectrum is mostly removed by the cuts on the muon transverse momenta.

process	$\sigma \times BR$[pb]	f[fb/event]
$H \rightarrow W^+W^-(150\ \text{GeV})$	186.3	0.061
$H \rightarrow W^+W^-(170\ \text{GeV})$	198.8	0.048
$H \rightarrow W^+W^-(190\ \text{GeV})$	142.4	0.044
$Z/\gamma^* \rightarrow \mu^+\mu^-$	$8.227 \cdot 10^5$	2.2 - 2.9
$Z/\gamma^* \rightarrow \tau^+\tau^-$	4984	0.52
$t\bar{t}$	4996	1.1
W^+W^-	773.1	0.10 - 0.14
ZZ	430.4	0.059
WZ	831.6	3.1
$W + jets$	252.8	1.5 - 8.2

Table 5.1: Cross section times branching ratio of the different processes after preselection, estimated from Monte Carlo data

$\sigma\times$ BR [fb]	preselection	0 jet	2 jet
$Z/\gamma^* \to \mu^+\mu^-$	$(8.227 \pm 0.014) \cdot 10^5$	$(6.588 \pm 0.012) \cdot 10^5$	$(1.414 \pm 0.006) \cdot 10^5$
$Z/\gamma^* \to \tau^+\tau^-$	4984 ± 50.7	3880 ± 44.7	951.7 ± 22.1
$t\bar{t}$	4996 ± 74.6	402.5 ± 21.2	4552 ± 71.2
W^+W^-	773.1 ± 10.2	549.4 ± 8.56	203.8 ± 5.23
WZ	831.6 ± 50.9	230.5 ± 26.8	579.3 ± 42.5
ZZ	430.4 ± 5.02	148.3 ± 2.95	275.7 ± 4.02
W+QCD	252.8 ± 38.8	130.3 ± 27.6	122.4 ± 28.8
All Background	$(8.350 \pm 0.014) \cdot 10^5$	$(6.642 \pm 0.012) \cdot 10^5$	$(1.481 \pm 0.006) \cdot 10^5$
$S, m_H = 150$ GeV	186.3 ± 3.37	94.51 ± 2.40	83.95 ± 2.26
$S, m_H = 170$ GeV	198.8 ± 3.10	94.33 ± 2.13	96.74 ± 2.16
$S, m_H = 190$ GeV	142.4 ± 2.50	66.75 ± 1.71	69.83 ± 1.75

Table 5.2: Cross section times branching ratio of the different processes after preselection, estimated from Monte Carlo data. Only statistical errors are given.

5.5 Signal Decay Channels

In this analysis two distinct "regions" or "channels" corresponding to the Higgs production processes of vector boson and gluon fusion are defined:

- The "0 jet" region, where no jet may have a larger transverse energy than 30 GeV,
- The "2 jet" region, where there must be at least one jet with transverse energy greater than 30 GeV and at least two reconstructed jets.

This improves the individual selection: if we would look at the data directly after preselection, a good selection cut would be to remove all events with high jet activity. This would remove most of the $t\bar{t}$ background, yet also discard almost all VBF events. In the 2 jet region we can deal with the $t\bar{t}$ background separately using different variables.

The overlap between the regions so defined is zero, and the loss of signal due to events with one high-E_T jet and no additional jet is with about 4% very small (see table 5.2). Now the 0 jet region or "channel" can be optimized for the search for Higgs bosons from quark/gluon production and the 2 jet "channel" can be used to search for Higgs bosons from Vector Boson Fusion. In later stages, this split approach could be used to measure the Higgs boson couplings of quarks (in qq/gg fusion) separately from the the coupling of the Higgs boson to Vector bosons.

It can be noted that the number of events in the 2 jet channel is significantly greater than expected just from vector boson fusion. This means that also a substantial fraction of gluon-gluon fusion events radiate hard gluons and produce jets. Further study, perhaps on generator level, would be needed for a better separation of the two production channels.

5.6 Selection Cuts

To evaluate sophisticated multivariate classifiers, a baseline for comparison is needed. For this reason, a traditional analysis that determines a signal region as a hypercube in variable space using cuts was set up. The cuts were optimized for maximum signal significance $S/\sqrt{B}$ for both channels and for all three examined Higgs boson masses.

In both channels the following properties of the two leading muons are used: their invariant mass, their transverse momentum, their distance in the η-ϕ plane and the transverse mass of the two leading muons combined with the missing transverse energy $m_{T,\not{E}_{T/\mu\mu}}$. Additionally, the missing transverse energy $\not{E}_T$, the sum of the jet transverse momenta $\sum_{Jet} p_T$ and the number of electrons with more than 30 GeV transverse momentum are used to suppress various backgrounds.

5.6.1 0-Jet Channel

In the 0 jet channel - the quark/gluon fusion channel - no additional variables are used for cuts, since all other jet activity should be background.

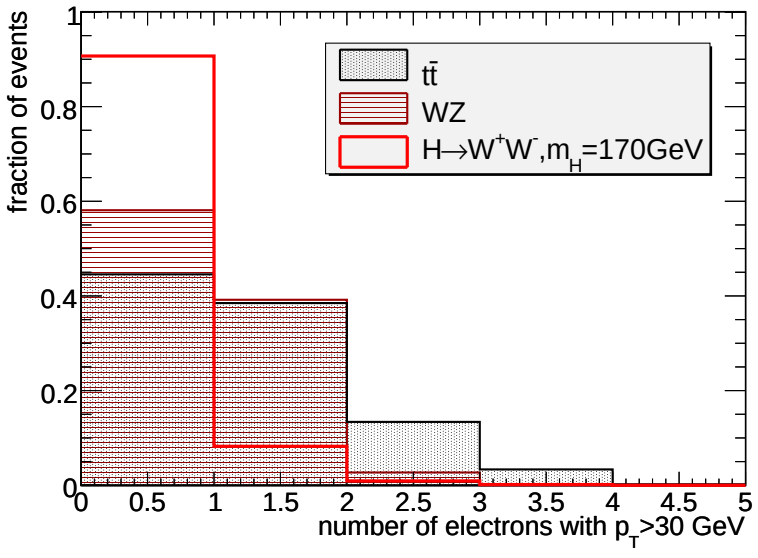


Figure 5.7: Histogram of the number of electrons with a transverse momentum $p_T > 30$ GeV. The plot is shown as overlay of three processes in the 0 jet channel.

- The first cut removes all events with a reconstructed electron with a transverse momentum $p_T > 30$ GeV. In figure 5.7 the number of

hard electrons in the 170 GeV signal is compared to the $t\bar{t}$ and WZ backgrounds.

- Secondly, a window in the invariant mass distribution of the two leading muons is chosen. Since the signal is spread from 10 to 80 GeV the borders will be determined by the background - Drell-Yan decays still absolutely dominating. A minimal lower border of 15 GeV is required by the influence of the strong J/ψ and Υ resonances. The optimized upper border depends on the Higgs boson mass, and lies between 51 and 73 GeV. Figure 5.8 shows this invariant mass.

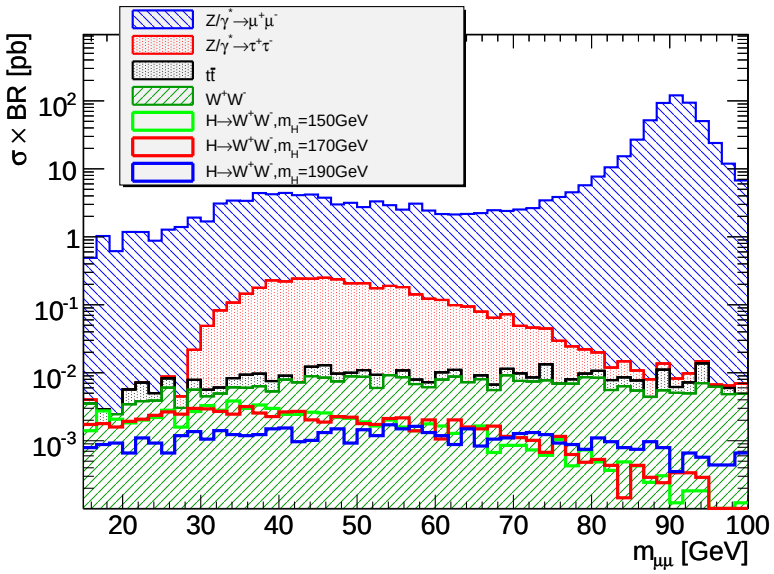


Figure 5.8: Histogram of the invariant mass of the two leading muons $m_{\mu\mu}$ on a logarithmic scale with backgrounds after preselection and hard electron veto. $Z/\gamma^* \to \mu^+\mu^-$ events have been simulated only with $m_{\mu\mu} > 15$ GeV.

- Another cut that works well against most of the backgrounds is a cut on the transverse momentum of the second muon. Since the W bosons are approximately on the mass shell and ideally at rest at $m_H = 170$ GeV the second muon momentum is approximately evenly distributed between 0 and 35 GeV, then falls off quickly. In figure 5.9 this is shown for all three Higgs signals, compared to background. The optimization of this cut for significance, in context of the other cuts, gives 14 GeV for $m_H = 150$ GeV, 26.5 GeV for $m_H = 170$ GeV and no cut for $m_H = 190$ GeV.

- The next cut exploits the spin correlation of the W bosons from the Higgs boson decay. The opening angle of the two muons, corresponding

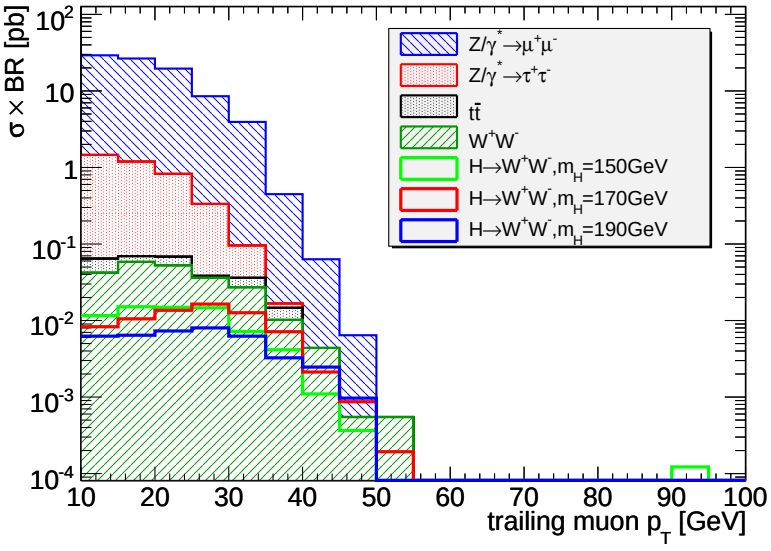


Figure 5.9: Histogram of the transverse momentum of the trailing muon in the 0 jet channel after previous cuts for $m_H = 170$ GeV

to $\Delta R_{\mu\mu}$, their distance in the ϕ-η plane, is a very efficient separator suppressing the Drell-Yan background to 3% at a signal efficiency of 80%. In figure 5.10 one can see that also the W^+W^- background is also slightly reduced. Additionally cutting on the angle in the transverse plane $\Delta\phi$ is also advantageous in most cases.

At this point, considering $m_H = 170$ GeV, using only cuts on muon momentum and direction measurements, all but two backgrounds are strongly suppressed: Drell-Yan decays with approximately 310 fb, and QCD W^+W^- decays with 36.5 fb. The Higgs signal for $m_H = 170$ GeV is at this point 31.2 fb. If there was no uncertainty on these numbers[1] they would correspond to an expected 1.6σ signal at 1 fb^{-1}. Still, we can do much better if we now add the missing transverse energy and derived variables like the transverse mass of the missing transverse energy and the dimuon system $m_{T,\not{E}_{T/\mu\mu}}$. This of course assumes that the detector is well-understood and well calibrated.

- Requiring the value for the missing transverse energy $\not{E}_T \gtrsim 40$ GeV reduces the Drell-Yan background to almost zero - There is a 1.9% efficiency at 86% signal efficiency for $m_H = 170$ GeV (see figure 5.11)

[1] Needless to say, this is not the case: the "significances" given in this section are only for the orientation of the reader and comparison. Accurate estimates of the resulting significance will be given in later sections.

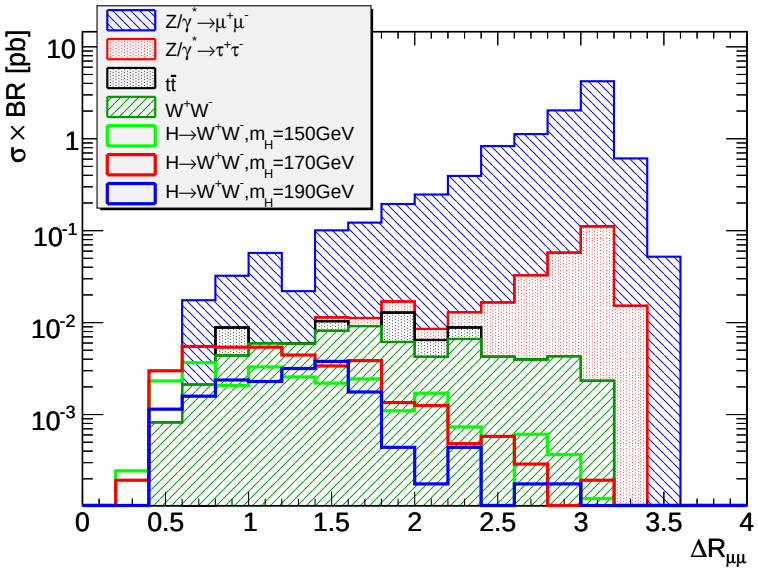


Figure 5.10: Histogram of the opening angle of the muons in the η-ϕ plane in the 0 jet channel after previous cuts for $m_H = 170$ GeV

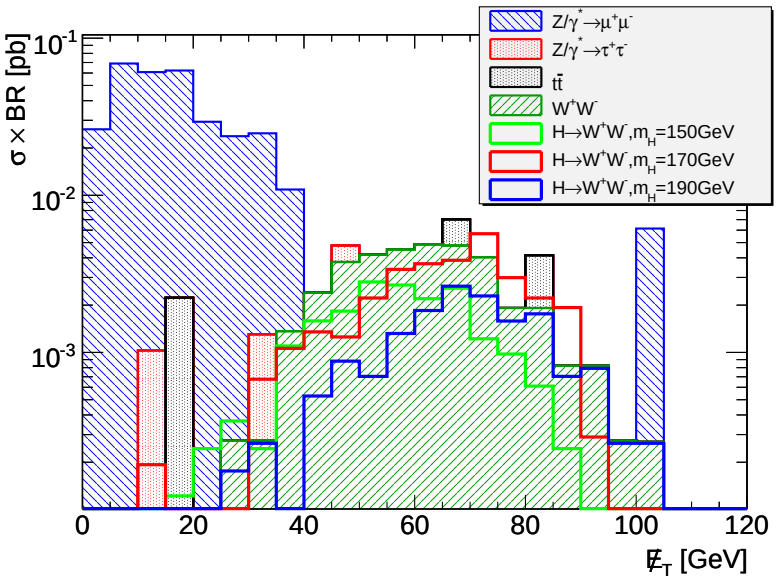


Figure 5.11: Histogram of the missing transverse energy in the 0 jet channel after previous cuts for $m_H = 170$ GeV

- The other significant background, QCD W^+W^- decays, is very similar to the Higgs boson decays in both the missing transverse energy $\not{E}_T$ and the transverse mass $m_{T,\not{E}_{T/\mu\mu}}$ (see figure 5.12). Not much can be gained there. But still, the final cut on the transverse mass $m_{T,\not{E}_{T/\mu\mu}}$ reduces some remnants of other backgrounds, and we get - for $m_H = 170$ GeV - a signal that is expected to be 3.8σ over the expected background at 1 fb^{-1}, again without considering uncertainties.

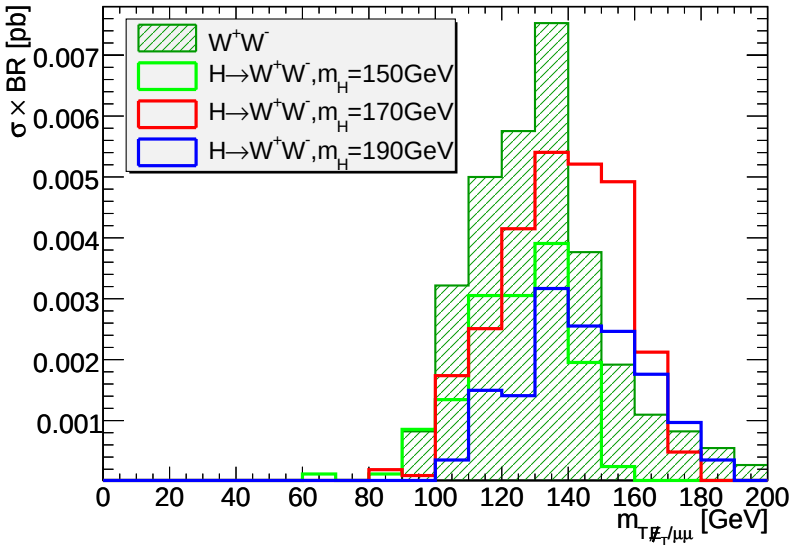


Figure 5.12: Histogram of the transverse mass of the two muon system and the missing transverse energy $m_{T,\not{E}_{T/\mu\mu}}$ in the 0 jet channel after previous cuts for $m_H = 170$ GeV.

In table 5.6 the optimized cut parameters for the 0 jet channel are summarized. Shortened cut flows of the respective analyses for each Higgs boson mass containing only the backgrounds most important to this channel, the QCD W^+W^- and Drell-Yan $\mu^+\mu^-$ decays, are given in tables 5.3, 5.4 and 5.5. The cut flows with all background processes included can be found in Appendix B.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
0j	$(6.59 \pm 0.01) \cdot 10^5$	549 ± 8.6	$(6.64 \pm 0.01) \cdot 10^5$	94.5 ± 2.4	0.12	0.12
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	517 ± 8.3	$(6.45 \pm 0.01) \cdot 10^5$	86.2 ± 2.3	0.11	0.11
$m_{\mu\mu}$ range	$(5.25 \pm 0.06) \cdot 10^4$	121 ± 5.7	$(5.50 \pm 0.06) \cdot 10^4$	52.4 ± 2.5	0.22	0.22
$p_{T,trailing\mu}$	$(3.18 \pm 0.04) \cdot 10^4$	93.9 ± 5.0	$(3.32 \pm 0.04) \cdot 10^4$	44.6 ± 2.3	0.24	0.24
max $\Delta R_{\mu\mu}$	$(6.42 \pm 0.19) \cdot 10^3$	64.4 ± 4.2	$(6.52 \pm 0.20) \cdot 10^3$	39.3 ± 2.2	0.49	0.48
min $\not{E}_T$	5.87 ± 7.3	43.1 ± 3.4	60.4 ± 29	30.3 ± 1.9	3.90	3.59
$m_{T,\mu\mu}$ range	5.87 ± 7.3	25.6 ± 2.7	33.8 ± 29	23.1 ± 1.7	3.97	3.57

Table 5.3: Cut flow table for $m_H = 150$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
0j	$(6.59 \pm 0.01) \cdot 10^5$	549 ± 8.6	$(6.64 \pm 0.01) \cdot 10^5$	94.3 ± 2.1	0.12	0.12
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	517 ± 8.3	$(6.45 \pm 0.01) \cdot 10^5$	85.5 ± 2.0	0.11	0.11
$m_{\mu\mu}$ range	$(8.44 \pm 0.07) \cdot 10^4$	232 ± 7.9	$(8.84 \pm 0.07) \cdot 10^4$	71.8 ± 2.6	0.24	0.24
$p_{T,trailing\mu}$	$(9.74 \pm 0.21) \cdot 10^3$	68.1 ± 4.3	$(1.01 \pm 0.02) \cdot 10^4$	35.3 ± 1.9	0.35	0.35
max $\Delta R_{\mu\mu}$	309 ± 42	36.5 ± 3.2	364 ± 50	31.2 ± 1.7	1.63	1.59
min $\not{E}_T$	5.87 ± 7.3	30.5 ± 2.9	52 ± 29	26.8 ± 1.6	3.72	3.42
$m_{T,\mu\mu}$ range	0 ± 7.3	28.3 ± 2.8	45.8 ± 29	26.1 ± 1.6	3.85	3.51

Table 5.4: Cut flow table for $m_H = 170$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
0j	$(6.59 \pm 0.01) \cdot 10^5$	549 ± 8.6	$(6.64 \pm 0.01) \cdot 10^5$	66.8 ± 1.7	0.08	0.08
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	517 ± 8.3	$(6.45 \pm 0.01) \cdot 10^5$	60.5 ± 1.6	0.08	0.07
$m_{\mu\mu}$ range	$(7.76 \pm 0.07) \cdot 10^4$	206 ± 7.4	$(8.15 \pm 0.07) \cdot 10^4$	38.2 ± 1.8	0.13	0.13
$p_{T,trailing\mu}$	$(7.76 \pm 0.07) \cdot 10^4$	206 ± 7.4	$(8.15 \pm 0.07) \cdot 10^4$	38.2 ± 1.8	0.13	0.13
max $\Delta R_{\mu\mu}$	$(1.00 \pm 0.02) \cdot 10^4$	91.7 ± 5	$(1.02 \pm 0.02) \cdot 10^4$	29.2 ± 1.6	0.29	0.29
min $\not{E}_T$	29.4 ± 14	74.7 ± 4.5	139 ± 32	26.7 ± 1.5	2.26	2.17
$m_{T,\mu\mu}$ range	0 ± 7.3	8.21 ± 1.6	13.7 ± 29	9.33 ± 0.91	2.52	2.22

Table 5.5: Cut flow table for $m_H = 190$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

cut	150 GeV	170 GeV	190 GeV
min $m_{\mu\mu}$	15 GeV $< m_{\mu\mu}$	15 GeV $< m_{\mu\mu}$	15 GeV $< m_{\mu\mu}$
max $m_{\mu\mu}$	$m_{\mu\mu} < 51.3$ GeV	$m_{\mu\mu} < 73.5$ GeV	$m_{\mu\mu} < 69$ GeV
trailing muon p_T	$p_T > 15$ GeV	$p_T > 26.5$ GeV	$p_T > 15$ GeV
$\Delta R_{\mu\mu}$	$\Delta R_{\mu\mu} < 1.8$	$\Delta R_{\mu\mu} < 1.8$	$\Delta R_{\mu\mu} < 1.8$
$\Delta\phi_{\mu\mu}$	$\Delta\phi_{\mu\mu} < 1.8$	-	$\Delta\phi_{\mu\mu} < 1.65$
$\not{E}_T$	$\not{E}_T < 44$ GeV	$\not{E}_T < 48$ GeV	$\not{E}_T < 40$ GeV
min $m_{T,\mu\mu}$	100 GeV $< m_{T,\mu\mu}$	100 GeV $< m_{T,\mu\mu}$	148 GeV $< m_{T,\mu\mu}$
max $m_{T,\mu\mu}$	$m_{T,\mu\mu} < 140$ GeV	$m_{T,\mu\mu} < 170$ GeV	$m_{T,\mu\mu} < 210$ GeV

Table 5.6: Cut parameters for the 0 jet channel. The parameters have been individually optimized for each Higgs boson mass. The strong correlation between the $\Delta\phi_{\mu\mu}$ and $\Delta R_{\mu\mu}$ cuts lead to the situation that the $\Delta\phi_{\mu\mu}$ cut is sometimes not necessary (denoted by -).

5.6.2 2-Jet Channel

In the 2 jet channel additional variables are necessary: the sum of the jet transverse momenta $\sum_{jet} p_T$, and the η directions of the leading jet. At the beginning, however, we start by noting that in this channel the cut on hard electrons removes 12% of the signal compared to 7% in the 0 jet channel, presumably because of electrons occurring in the jets. This makes this cut inefficient, and it is dropped.

- The next cut, as in the 0 jet channel, a window in the dimuon mass distribution is chosen (see figure 5.13), the lower limit being 15 GeV for $m_H = 150$ GeV and 170 GeV and 22 GeV for $m_H = 190$ GeV and the upper limit between 56.5 GeV and 70.5 GeV.

- Now the minimum transverse momentum of the trailing muon is raised: in figure 5.14 one can see that a cut from 20 to 30 GeV can remove a great amount of background, especially in the case of $m_H = 170$ GeV.

- The ΔR and $\Delta\phi$ cuts are in this channel not very effective, since the W^+W^- background is not as relevant in this channel (see figure 5.15). During optimization only the $m_H = 150$ GeV analysis showed an increase in significance.

- The next cut limits the sum of the jet transverse momenta (see figure 5.16). This cut removes most of the Drell-Yan $\tau\tau$ decays as well as some of the $\mu\mu$ decays and $t\bar{t}$ background.

- Subsequently a central jet veto is applied: the leading jet must not lie in an area that is transverse to the beam axis, the value of $|\eta|$ must be greater than a certain value (see table 5.7 for a summary of

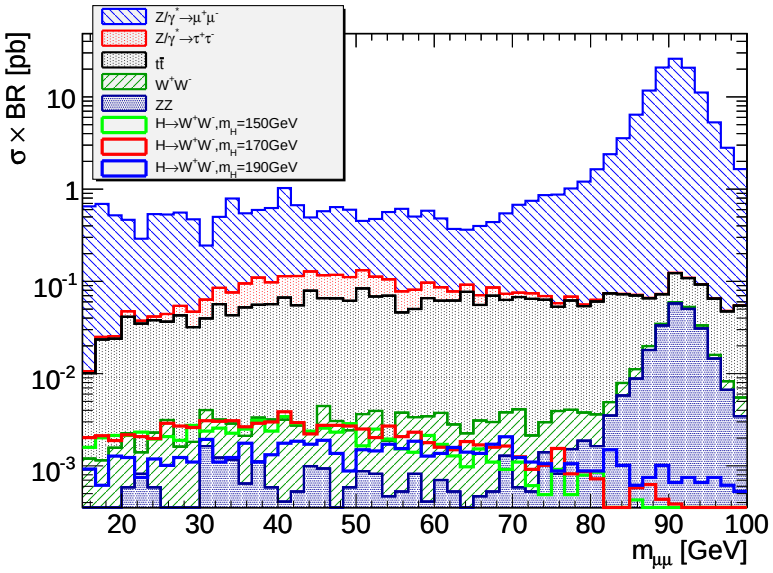


Figure 5.13: Histogram of the invariant mass of the two leading muons $m_{\mu\mu}$ on a logarithmic scale with backgrounds after preselection in the 2 jet channel.

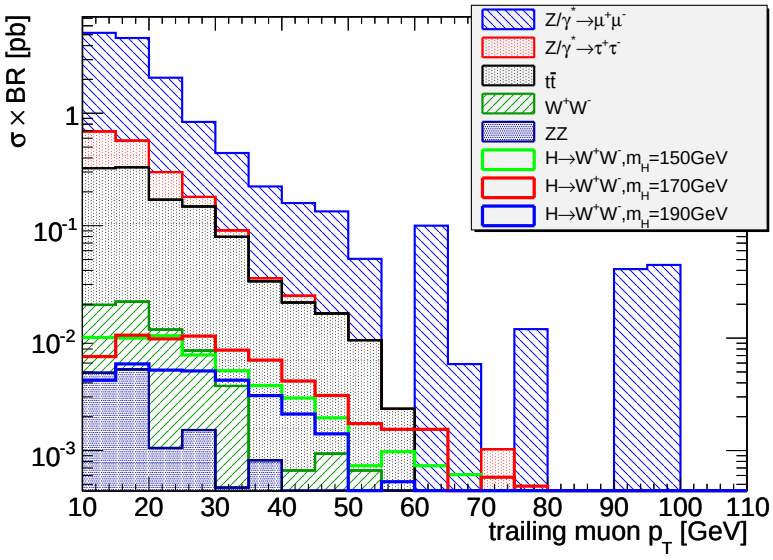


Figure 5.14: Histogram of the trailing muon transverse momentum p_T after invariant mass cuts in the 2 jet channel.

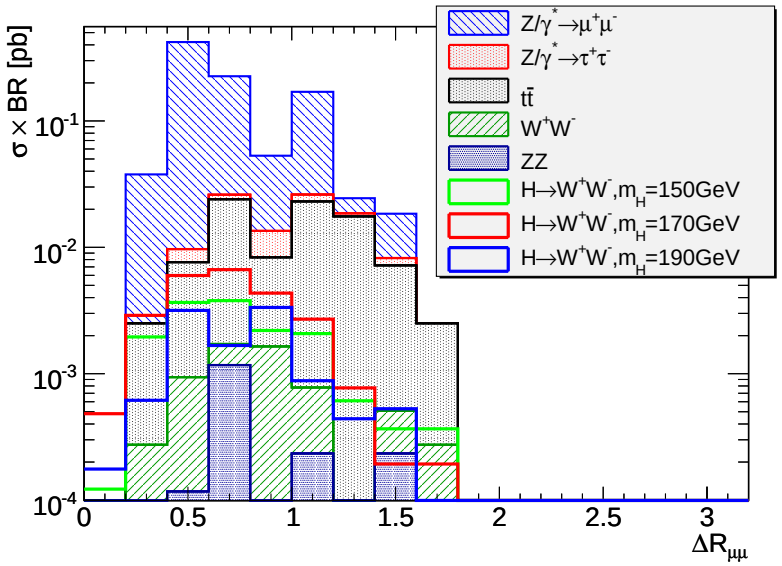


Figure 5.15: Histogram of the distance between the two leading muons in the η-ϕ plane $\Delta R_{\mu\mu}$ after invariant mass and trailing muon p_T cuts in the 2 jet channel.

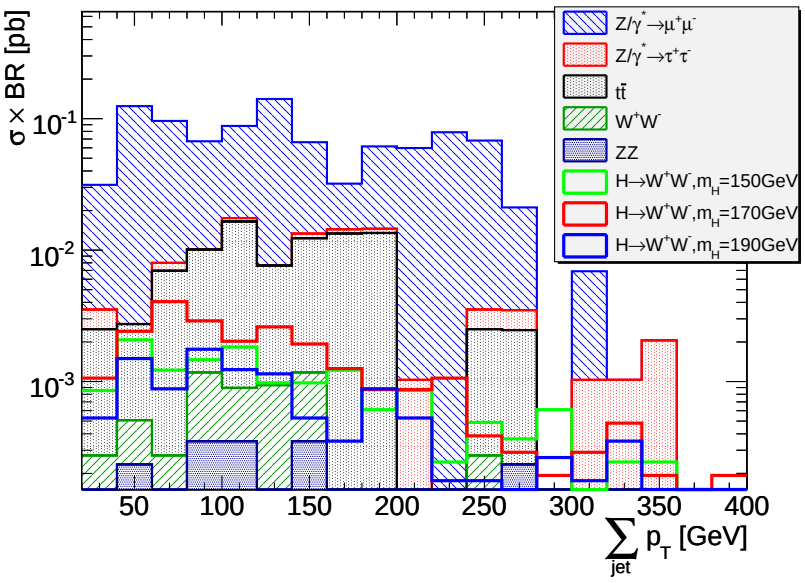


Figure 5.16: Histogram of the sum of the jet transverse momentum $\sum_{jet} p_T$ after previous cuts in the 2 jet channel.

cut parameters). In figure 5.17 this quantity is histogrammed directly after preselection, since the lack of good Monte Carlo statistic after the previous cuts makes it hard to see the structure. Here we can clearly see that this cut suppresses $t\bar{t}$ more than Higgs boson decays.

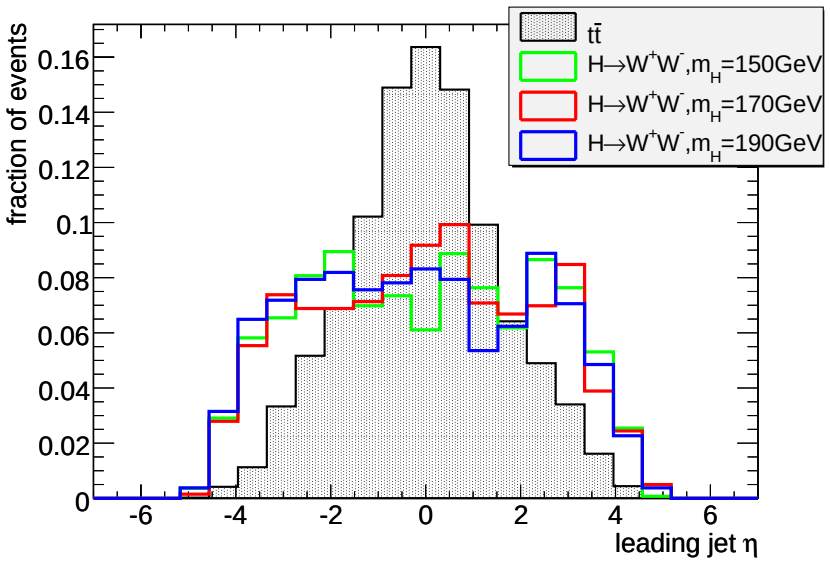


Figure 5.17: Histogram of η of the leading jet in the 2 jet channel. For better visibility only preselection cuts are applied.

- After these cuts specific to the 2 jet channel we continue to apply the cuts already used in the 0 jet channel. The missing transverse energy $\not{E}_T$ and the transverse mass of the missing transverse energy combined with the two leading muons $m_{T,\not{E}_{T/\mu\mu}}$ are shown (after preselection) in the figures 5.18 and 5.19. The corresponding cuts are very similar to the 0 jet channel.

In table 5.7 the cuts for the 2 jet channel for each Higgs boson mass are summarized. Again, shortened cut flows of the respective analyses for each Higgs boson mass containing only the backgrounds most important to this channel, the QCD $t\bar{t}$ and Drell-Yan $\mu^+\mu^-$ decays, are given in tables 5.8, 5.9 and 5.10. The full cut flows are given in the appendix.

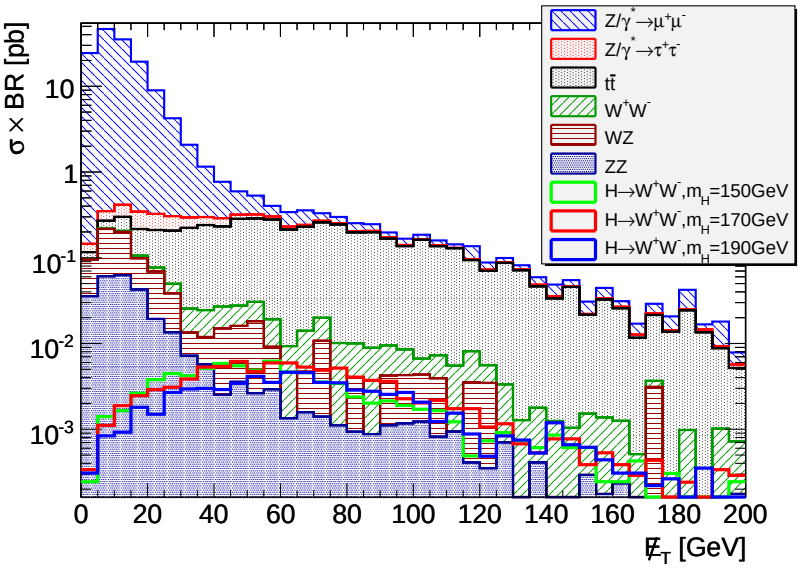


Figure 5.18: Histogram of the missing transverse energy $\not{E}_T$ in the 2 jet channel. For better visibility only preselection cuts are applied.

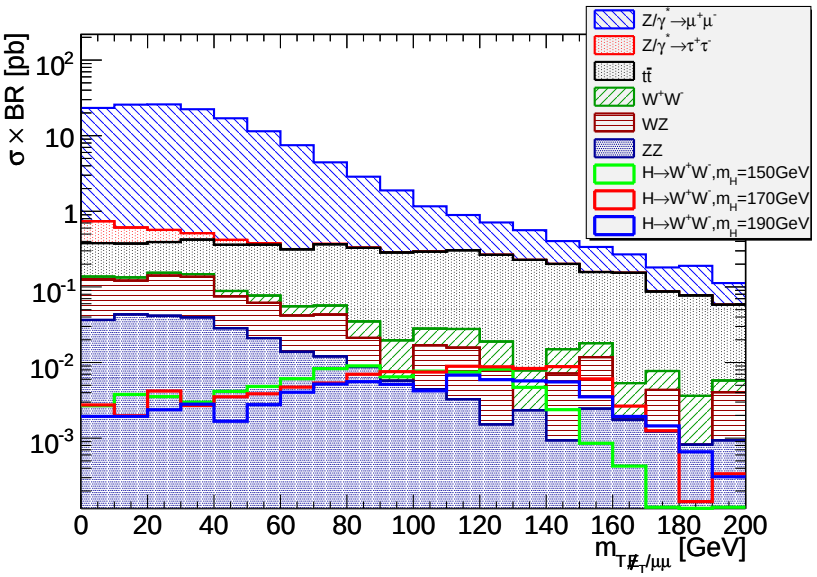


Figure 5.19: Histogram of the transverse mass of the missing transverse energy and the two leading muons $m_{T,\not{E}_T/\mu\mu}$ in the 2 jet channel. For better visibility only preselection cuts are applied.

cut	150 GeV	170 GeV	190 GeV
min $m_{\mu\mu}$	15 GeV $< m_{\mu\mu}$	15 GeV $< m_{\mu\mu}$	22 GeV $< m_{\mu\mu}$
max $m_{\mu\mu}$	$m_{\mu\mu} < 68$ GeV	$m_{\mu\mu} < 56.5$ GeV	$m_{\mu\mu} < 70.5$ GeV
trailing muon p_T	$p_T > 20$ GeV	$p_T > 32.5$ GeV	$p_T > 20.5$ GeV
$\Delta R_{\mu\mu}$	-	-	-
$\Delta\phi_{\mu\mu}$	$\Delta\phi_{\mu\mu} < 2$	-	-
$\sum_{jet} p_T$	< 230 GeV	< 230 GeV	< 190 GeV
leading jet η	$\lvert\eta\rvert > 1.55$ GeV	$\lvert\eta\rvert > 0.65$ GeV	$\lvert\eta\rvert > 1.45$ GeV
$\not{E}_T$	$\not{E}_T < 42$ GeV	$\not{E}_T < 45$ GeV	$\not{E}_T < 42$ GeV
min $m_{T,\mu\mu}$	100 GeV $< m_{T,\mu\mu}$	100 GeV $< m_{T,\mu\mu}$	95 GeV $< m_{T,\mu\mu}$
max $m_{T,\mu\mu}$	$m_{T,\mu\mu} < 150$ GeV	$m_{T,\mu\mu} < 190$ GeV	$m_{T,\mu\mu} < 185$ GeV

Table 5.7: Cut parameters for the 2 jet channel. The parameters have been individually optimized for each Higgs boson mass.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
2j	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	$(1.48 \pm 0.01) \cdot 10^5$	84 ± 2.3	0.22	0.22
$m_{\mu\mu}$ range	$(1.48 \pm 0.06) \cdot 10^3$	72 ± 4.4	$(1.72 \pm 0.03) \cdot 10^4$	65.2 ± 2.8	0.50	0.49
$p_{T,trailing\mu}$	685 ± 39	34.1 ± 3.1	$(6.06 \pm 0.17) \cdot 10^3$	43.1 ± 2.3	0.55	0.55
max $\Delta R_{\mu\mu}$	584 ± 36	27.4 ± 2.8	$(4.84 \pm 0.16) \cdot 10^3$	39.7 ± 2.2	0.57	0.56
$\sum_{jet} p_T$	504 ± 34	26.3 ± 2.7	$(4.57 \pm 0.16) \cdot 10^3$	33.8 ± 2.0	0.50	0.49
cjv	167 ± 19	10.0 ± 1.8	$(1.60 \pm 0.09) \cdot 10^3$	21.7 ± 1.6	0.54	0.53
min $\not{E}_T$	114 ± 16	8.42 ± 1.6	154 ± 34	13.9 ± 1.3	1.12	1.08
$m_{T,\mu\mu}$ range	58 ± 11	4.11 ± 1.2	62.3 ± 31	9.16 ± 1.1	1.16	1.09

Table 5.8: Cut flow table for $m_H = 150$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
2j	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	$(1.48 \pm 0.01) \cdot 10^5$	96.7 ± 2.2	0.25	0.25
$m_{\mu\mu}$ range	$(1.06 \pm 0.05) \cdot 10^3$	53.2 ± 3.8	$(1.41 \pm 0.03) \cdot 10^4$	66.5 ± 2.5	0.56	0.56
$p_{T,trailing\mu}$	84.7 ± 14	4.11 ± 1.2	951 ± 77	24.2 ± 1.5	0.79	0.77
max $\Delta R_{\mu\mu}$	84.7 ± 14	4.11 ± 1.2	951 ± 77	24.2 ± 1.5	0.79	0.77
$\sum_{jet} p_T$	80.3 ± 13	3.83 ± 1.2	809 ± 72	20.6 ± 1.4	0.72	0.71
cjv	51.3 ± 11	2.47 ± 1.0	481 ± 58	16.6 ± 1.3	0.76	0.74
min $\not{E}_T$	31.2 ± 8.3	1.64 ± 0.91	43.1 ± 30	11.6 ± 1.1	1.76	1.65
$m_{T,\mu\mu}$ range	22.3 ± 7.0	1.10 ± 0.83	23.6 ± 30	9.94 ± 0.98	2.04	1.86

Table 5.9: Cut flow table for $m_H = 170$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
2j	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	$(1.48 \pm 0.01) \cdot 10^5$	69.8 ± 1.8	0.18	0.18
$m_{\mu\mu}$ range	$(1.46 \pm 0.06) \cdot 10^3$	74.1 ± 4.5	$(1.56 \pm 0.03) \cdot 10^4$	42.4 ± 1.9	0.34	0.34
$p_{T,trailing\mu}$	676 ± 39	35 ± 3.1	$(5.75 \pm 0.17) \cdot 10^3$	31.3 ± 1.7	0.41	0.41
max $\Delta R_{\mu\mu}$	676 ± 39	35 ± 3.1	$(5.75 \pm 0.17) \cdot 10^3$	31.3 ± 1.7	0.41	0.41
$\sum_{jet} p_T$	406 ± 30	30.4 ± 2.9	$(4.93 \pm 0.16) \cdot 10^3$	23.1 ± 1.4	0.33	0.32
cjv	125 ± 17	13.1 ± 1.9	$(1.95 \pm 0.10) \cdot 10^3$	13.8 ± 1.1	0.31	0.30
min $\not{E}_T$	80.3 ± 13	8.90 ± 1.6	130 ± 33	10.7 ± 0.97	0.93	0.89
$m_{T,\mu\mu}$ range	49.1 ± 10	6.50 ± 1.4	61.7 ± 31	8.01 ± 0.84	1.02	0.96

Table 5.10: Cut flow table for $m_H = 190$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

5.7 Multivariate Classifiers

This thesis examines the following multivariate methods:

- boosted decision trees
- boosted randomised trees
- multilayer perceptrons (neural networks)
- the Fisher classifier

Two different approaches have been tried out:

- each classifier is trained against all background processes together
- the background is split into three groups: background containing Z, QCD W^+W^- production and other QCD events. For each group, a neural network and a fisher classifier is trained. Then, cuts on the classifier output are optimized for maximal significance as in the cut analysis.

The last approach is tried to see if improvement can be made if each neural network has only to adapt to a few background topologies.

5.7.1 Training

Training settings: The multivariate classifiers are trained using TMVA with the following settings:

- Boosted Decision Tree: Number of Trees: 1000,
  Strength of Pruning: 8
- Boosted Randomised Tree: Use Randomized Bagged Trees with 3 variables per node, Number of Trees: 1000
- Multilayer Perceptron: Default number of Hidden Layers; for N input variables there are N neurons in the first layer and $N-1$ neurons in the second layer.
- Fisher Classifier: Default settings

The pruning strength of the boosted decision tree was increased from the default 4.5 to reduce the amount of overtraining. The number of trees was increased from 400 since evaluation speed was not a critical factor in this analysis.

Additional preselection: To reduce the number of training events to reasonable numbers the following loose preselection cuts were applied in addition to the full preselection:

- no hard electrons ($p_T > 30$ GeV) in the 0 jet channel, the same as the first cut in the cut analysis
- an invariant mass window of 15 GeV $< m_{\mu\mu} <$ 70 GeV

Input variables: The following variables have been used for training in both the 0 jet and the 2 jet channel:

- invariant mass of the dimuon system $m_{\mu\mu}$
- missing transverse energy $\not{E}_T$
- transverse momentum of the system of the two leading muons $p_{T,\mu\mu}$
- transverse momentum of the trailing muon
- opening angle of the muons in the transverse plane $\Delta\phi_{\mu\mu}$
- distance of the muons in the $\phi_{\mu\mu}$-$\eta_{\mu\mu}$ plane $\Delta R_{\mu\mu}$
- transverse mass of the missing transverse energy and the system of the two leading muons $m_{T,\not{E}_T/\mu\mu}$

In the 2 jet channel these variables were added to characterize the increased jet activity:

- sum of transverse momenta of the jets $\sum_{jet} p_T$
- η of the leading jet
- η of the trailing jet

5.7.2 Neural Net Architecture

The number of nodes and layers of a neural network has a big influence on the performance of the neural net. The default architecture in TMVA is N neurons in the first, and $N-1$ neurons in the second layer, N being the number of input variables. This is a good choice for many applications, yet the number of nodes and especially the number of weights to be trained increases with N^2. This leads to serious limitations in cases where only limited training statistics are available: for the QCD background in the 0 jet channel and the W^+W^- background in the 2 jet channel the number of Monte Carlo events was apparently not enough to effectively train the neural network: the response detoriated to a random response. In these cases the Fisher classifier was used for the respective cut. A more detailed look at the influence of the architecture of the neural network is given in appendix A.3.

5.7.3 Visualization

A multivariate classifer maps a score to each point of the multidimensional variable space. Since more than two input variables are used here, a direct visual representation is not possible. Therefore, several methods are used to visualize the classifiers, thus gaining a better understanding of their performance.

Variation of the "Mean Signal Event"

This method is used to examine the effect of one input variable on the classifier response. For both channels and all examined Higgs boson masses the "mean Higgs boson decay" is determined by taking the mean of the distribution of the Higgs boson decay sample after the channel cuts. These mean events are listed in table 5.11. Since η of the leading jets is approximately symmetric around 0, it is set to -3 for the leading to $+3$ for the trailing jet to mimic a "typical" vector boson fusion event.

These mean events are taken as a baseline, and each input variable is varied and the respective classifier responses are plotted (figures 5.20-5.27). The most important information that can be seen from these plots is which variable values the classifier deems to be most "Signal-like", by looking at each curve indivudually. These plots can therefore show which features of each variable distribution is picked up or ignored by the classifier, and perhaps what prevents a certain classifier from doing better. The absolute offset on the y axis is not really important in these plots, since the distribution of the classifier response to background can be different for each classifier, even one of the same type[2].

This method does not completely characterize a classifier, but it gives impressive insights to its nature. In figures 5.20 and 5.21 one can see the jagged edges resulting from the binary nature of the decision trees, while figure 5.22 shows the continuous nature of neural networks. In figure 5.23 the linearity of the Fisher classifier is evident.

Boosted Decision Trees Sharp edges and spikes characterize the plots for the boosted decision trees (figures 5.20 and 5.24). The shapes suggest that the method can pick out several regions with high signal significance, and reject small sections with high background probability. This is evidenced by sudden dips in classifier response, for example in the invariant mass $m_{\mu\mu}$. The curves for the 0 jet channel seem to indicate that the training statistics were not completely sufficient to provide enough information to the method, and that statistical fluctuations have influenced the result. In the 2 jet channel,

[2]However, one can compare the values on the absolute scale with the classifier output distributions in section 5.7.3.

the classifier seems more regular. Yet in both cases many of the distinctive features of the signal are captured: a preference for smaller $\Delta R_{\mu\mu}$, the distributions of the transverse mass $m_{T,\not{E}_{T/\mu\mu}}$ for the different Higgs boson masses and some missing transverse energy $\not{E}_T$. Also, in the plot for the leading jet η one can see a preference for jets in the positive η direction, since the trailing jet is fixed at $\eta = -3$.

Bagged Randomized Trees The randomized trees show a similar behaviour to the boosted decision trees, yet in the 0 jet channel seem to have a "plateau", where all trees classifiy the event as signal. This is most probably a consequence of the training process: since each tree is trained independently of the other, the background events that are very close to the signal and are "costly" to remove are never actually separated.

Neural Networks The neural network response is easily distinguished from the jagged boosted decision trees. Since we assume that the "real" distributions of signal and background probabilities should not have sudden jumps, the smooth response of a neural network seems to approximate reality much better. Yet this smooth nature makes also two things very difficult: firstly, in contrast to the boosted decision trees, small features of the data can be smoothed out, and the performance diminished. Secondly the neural net *extrapolates* its response into areas where no or little training statistic is available, which is visible for example in the trailing muon p_T in figures 5.20 and 5.22. In contrast the boosted decision tree starts to return a constant value if one leaves the trained area. This can be especially unfortunate if unknown systematic uncertainties push events into regions where little training statistics are available, and where the neural network still classifies the events as signal.

Fisher classifier The responses of the Fisher classifier are linear, as expected. Still, many important features of the variables can be captured: $m_{T,\not{E}_{T/\mu\mu}}$, $\not{E}_T$, $\Delta R_{\mu\mu}$. Yet for example in figure 5.27 we can see that the opposite VBF jets can not be well used by this method, in contrast to the other, nonlinear methods: most of the lines in the plots for leading and trailing jet η indicate no preference, and therefore no separation power.

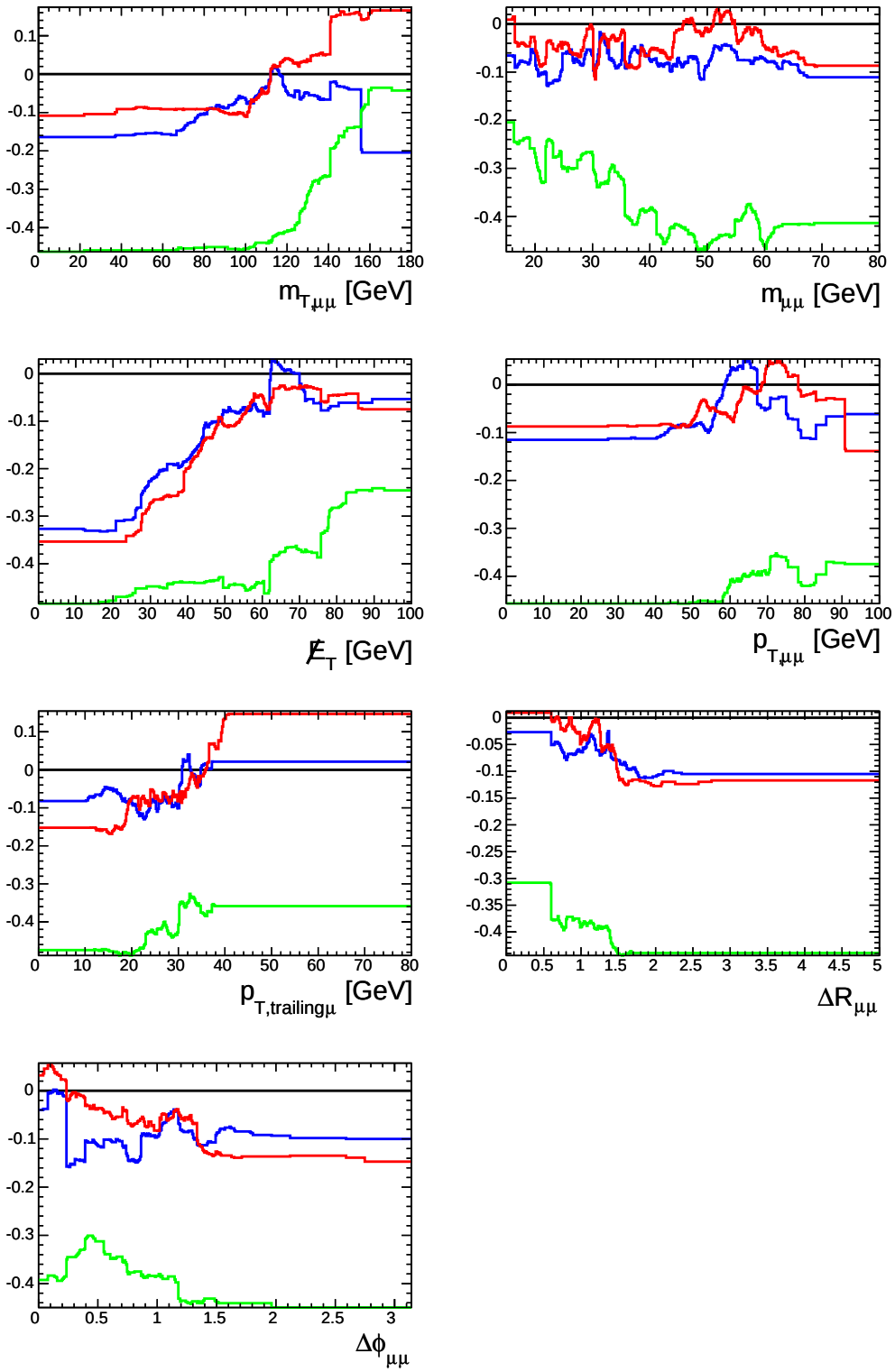


Figure 5.20: Visualization of the boosted decision trees trained in the 0 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

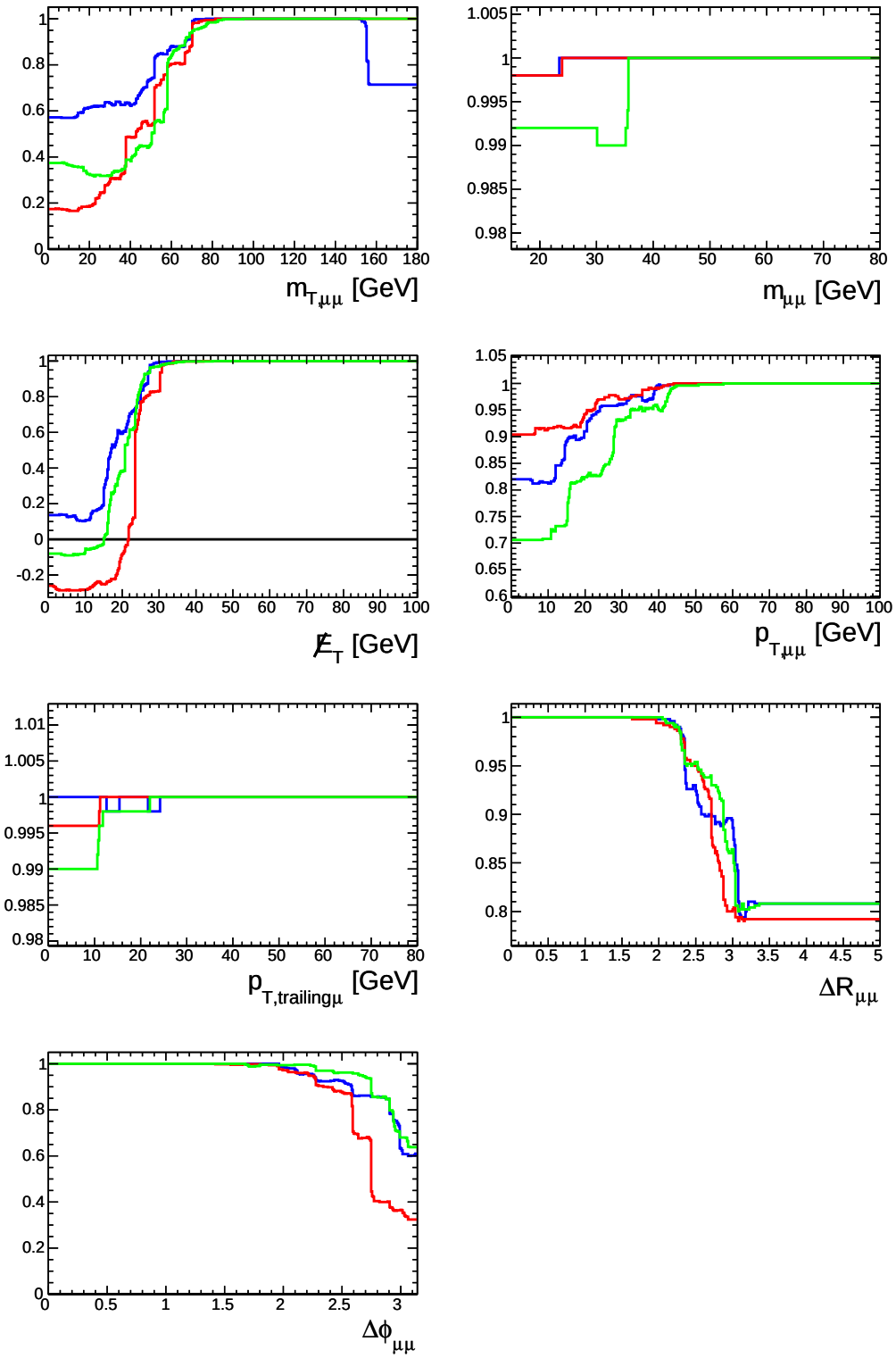


Figure 5.21: Visualization of the bagged randomized trees trained in the 0 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

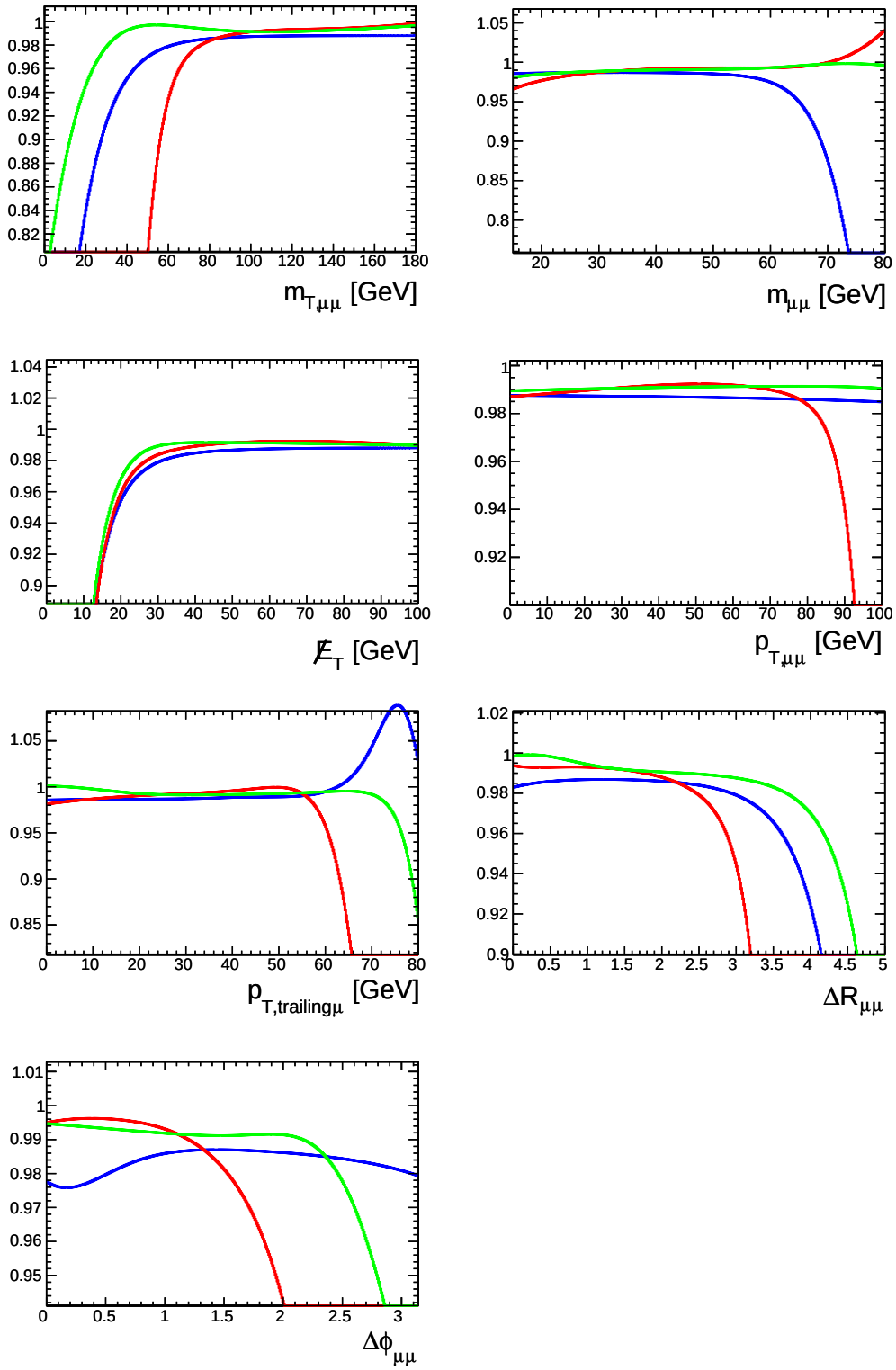


Figure 5.22: Visualization of the neural networks trained in the 0 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

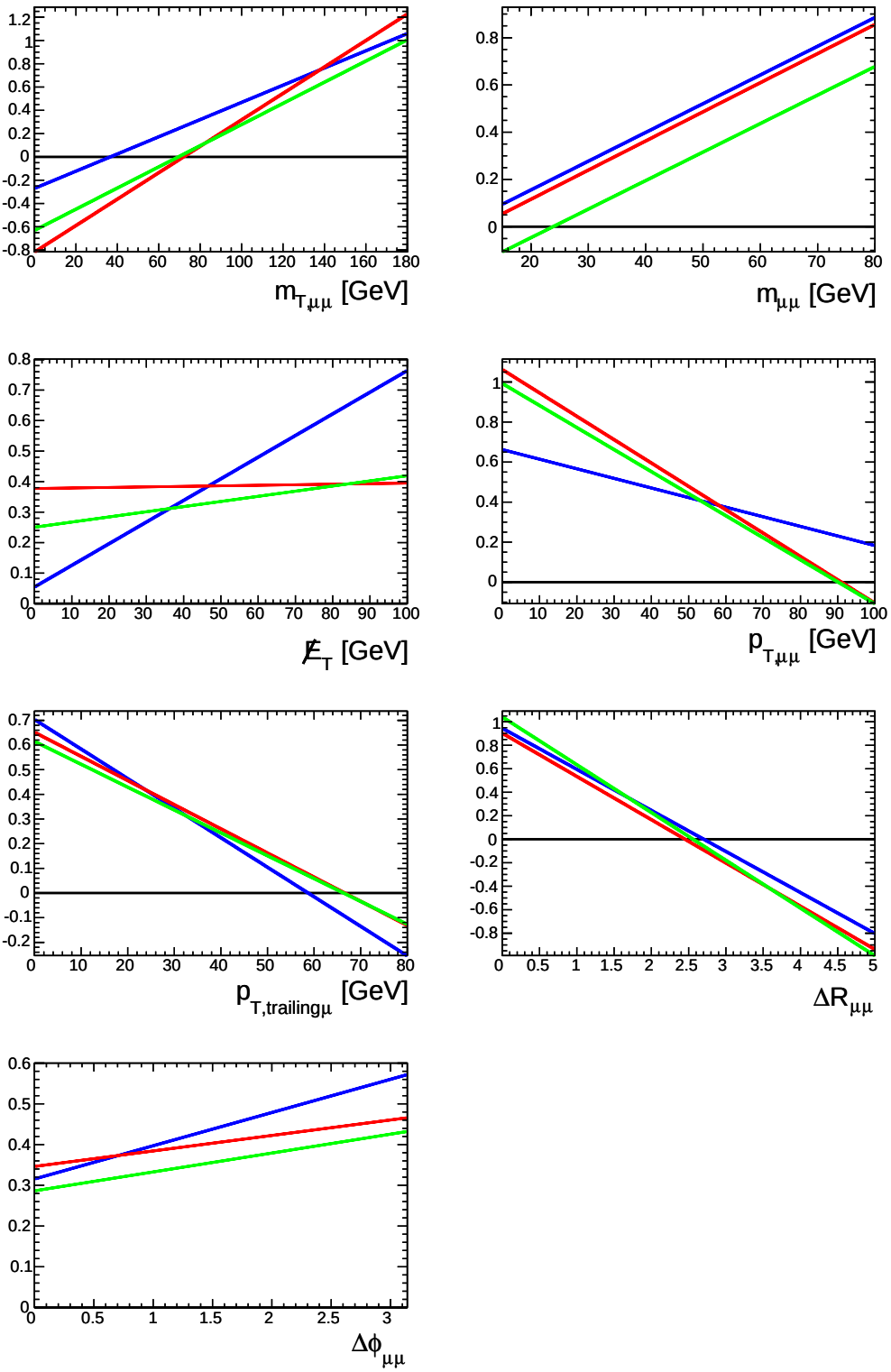


Figure 5.23: Visualization of the Fisher classifiers trained in the 0 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

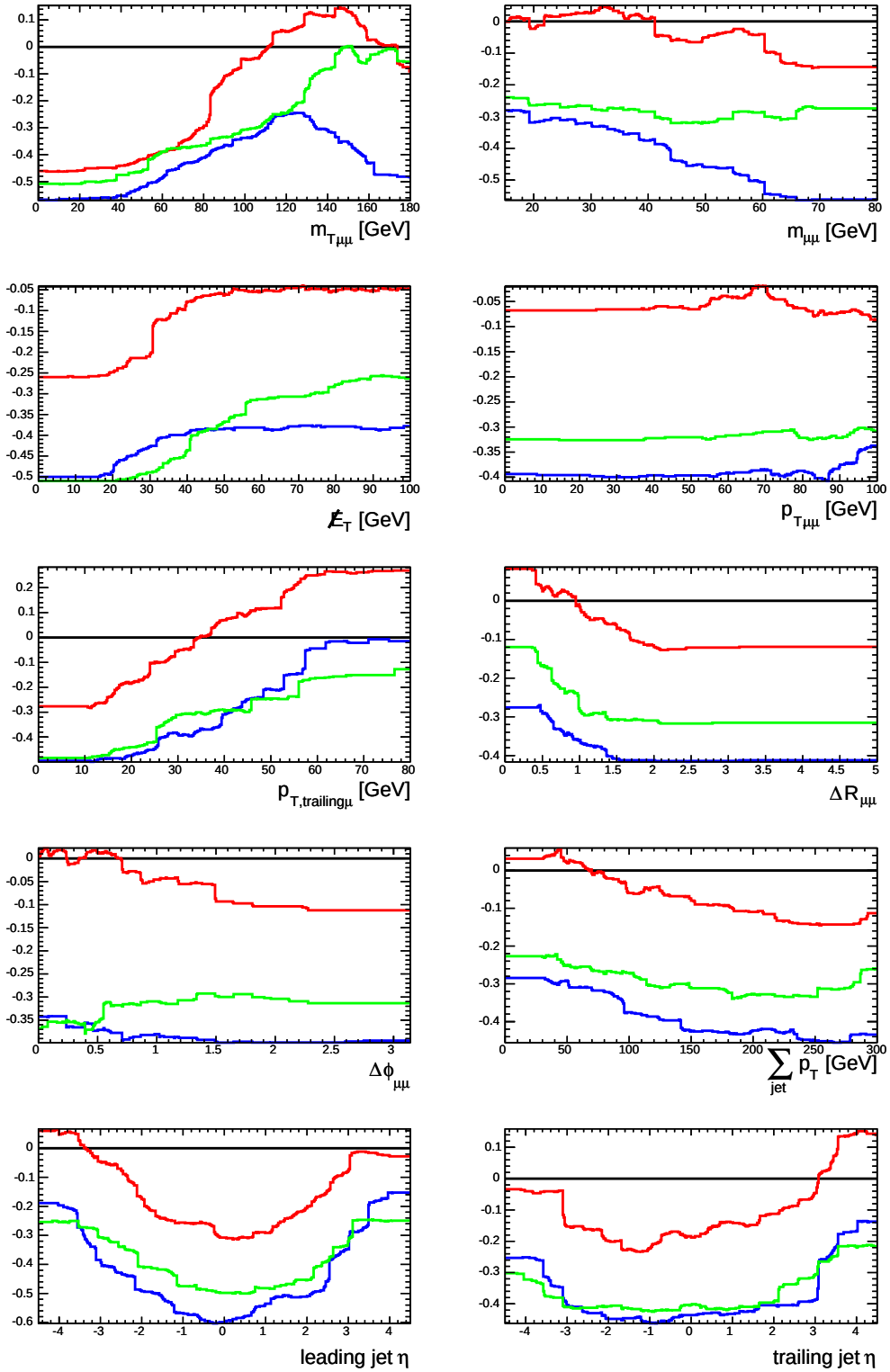


Figure 5.24: Visualization of the boosted decision trees trained in the 2 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

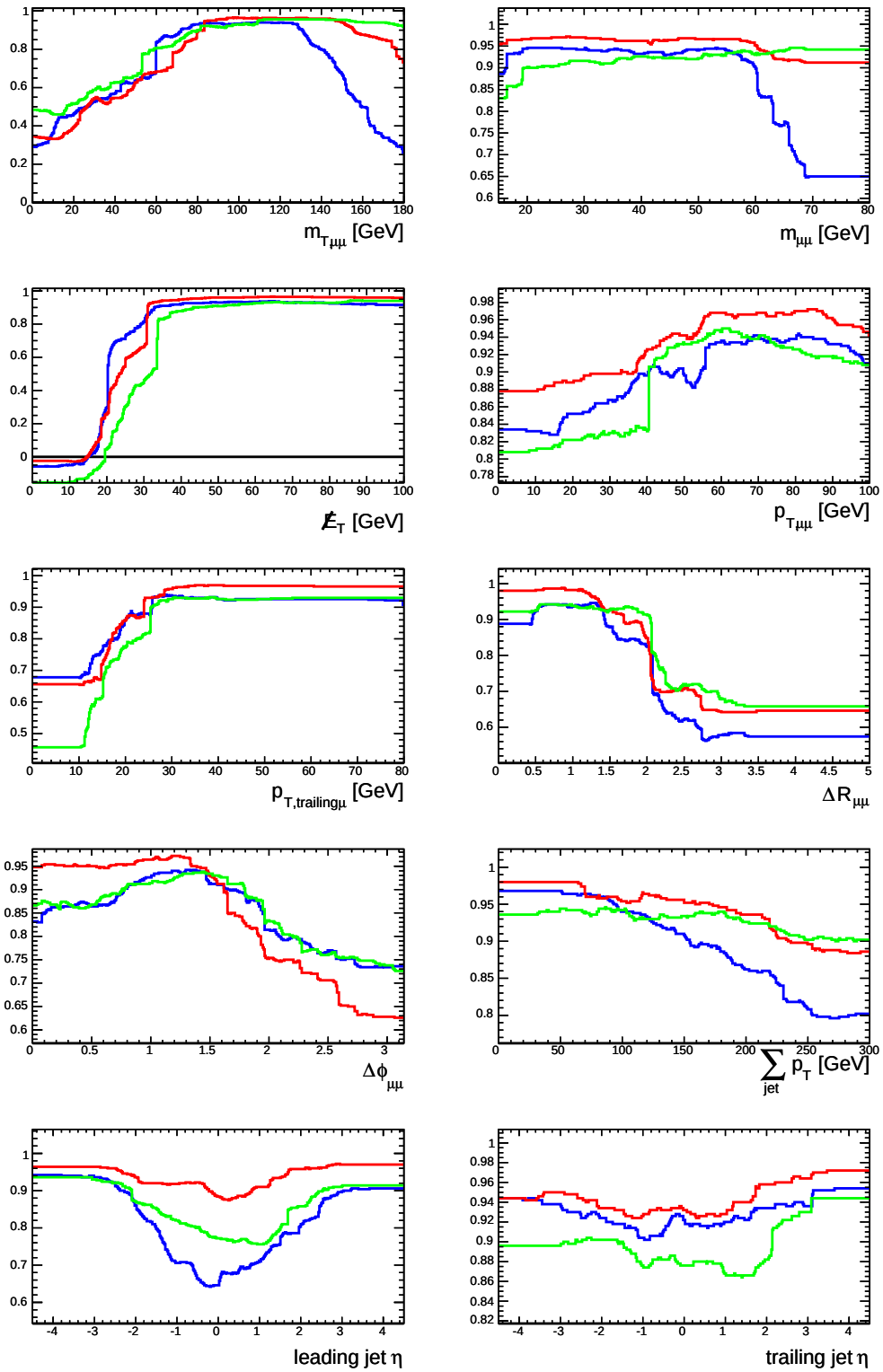


Figure 5.25: Visualization of the bagged randomized trees trained in the 2 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

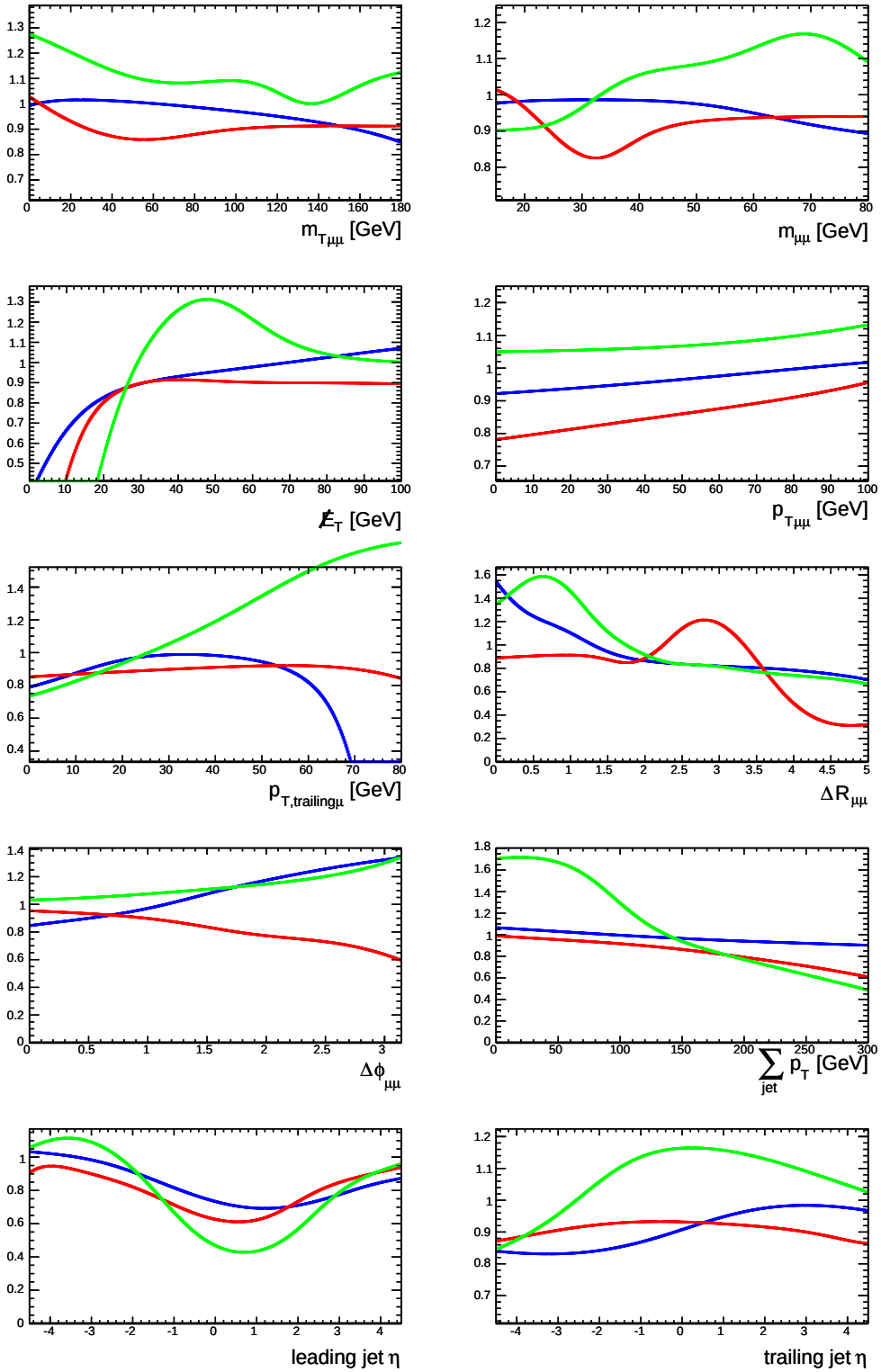


Figure 5.26: Visualization of the neural networks trained in the 2 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

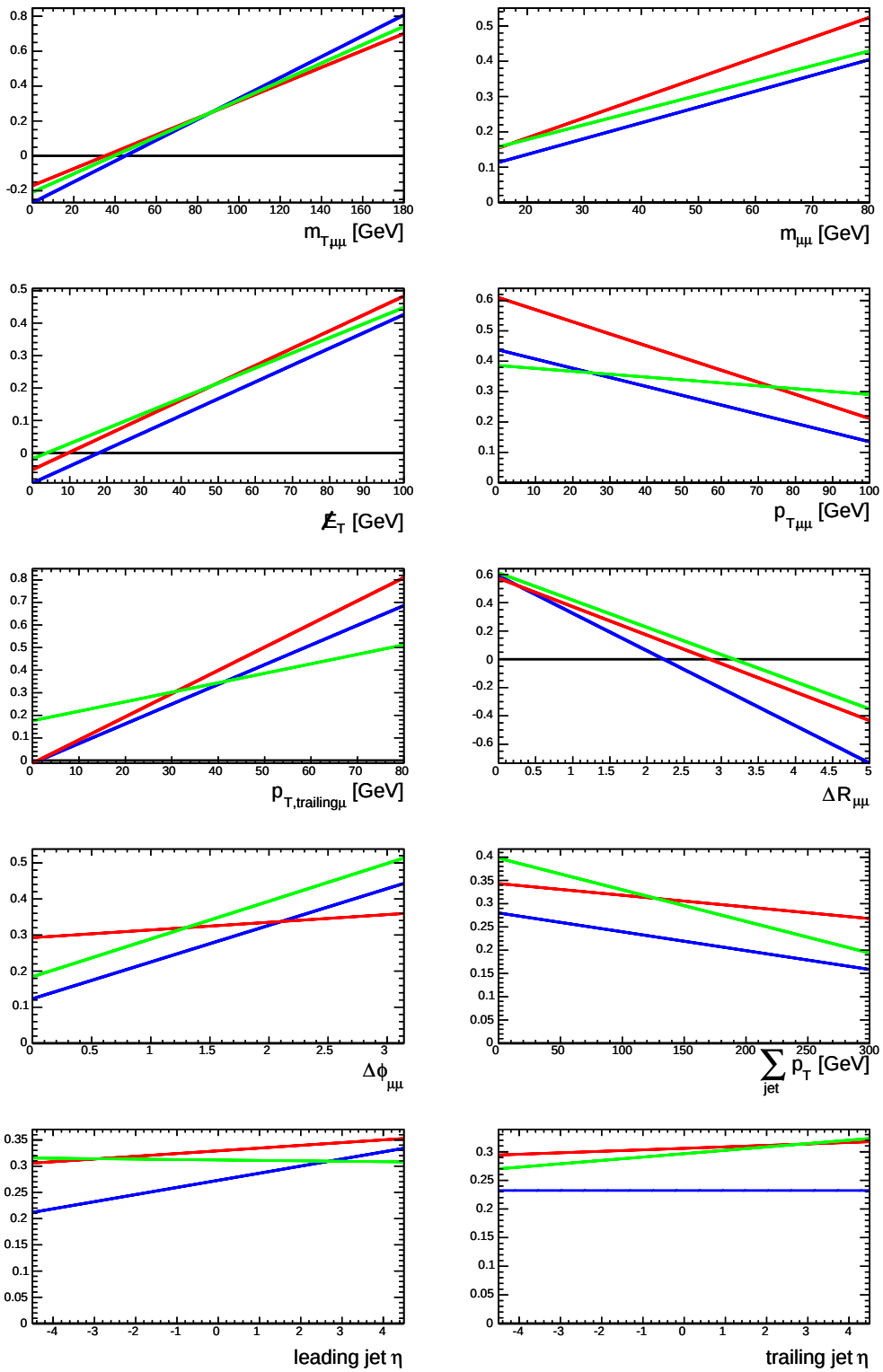


Figure 5.27: Visualization of the Fisher classifiers trained in the 2 jet channel. The blue line represents the classifier trained for $m_H = 150$ GeV, red is for $m_H = 170$ GeV and green for $m_H = 190$ GeV

channel	0 jet			2 jet		
m_H [GeV]	150	170	190	150	170	190
$m_{T,\not{E}_{T/\mu\mu}}$ [GeV]	92.6	106	108	83.8	100	98.9
$m_{\mu\mu}$ [GeV]	41.2	42.0	52.8	41.5	43.0	52.5
$\not{E}_T$ [GeV]	50.6	57.0	58.7	62.6	68.3	71.3
$p_{T,\mu\mu}$ [GeV]	51.9	57.8	58.4	67.8	74.3	75.1
$p_{T,\mu\mathrm{trail}}$ [GeV]	24.2	27.0	26.6	28.0	31.7	32.9
$\Delta R_{\mu\mu}$	1.52	1.40	1.70	1.35	1.29	1.54
$\Delta\phi_{\mu\mu}$	1.20	1.08	1.36	1.07	0.98	1.24
$\sum_{jet} p_T$ [GeV]	-	-	-	118	118	123

Table 5.11: Means of the Higgs boson decay sample distributions in the respective channels

ROC Curve

The "Receiver Operation Characteristics" or ROC curve is a tool for visualizing and comparing the performance of several methods. In it the signal efficiency and the background rejection are plotted against each other. Points with high signal efficiency and high background rejection are desirable, the optimal point depends on the definition of the desired significance. Figure 5.28 shows the ROC curve for $m_H = 170$ GeV in the 0 jet channel. This particular example shows that the classifier performance depends on the operating region: above a signal efficiency value of 0.6 the Fisher and bagged randomized trees classifiers are best, in the lower region the boosted decision trees outperform the other classifiers. For this Higgs boson mass and channel, the BDT working point is at 51% signal efficiency, and the MLP working point at 82%, as can be calculated using table 5.13.

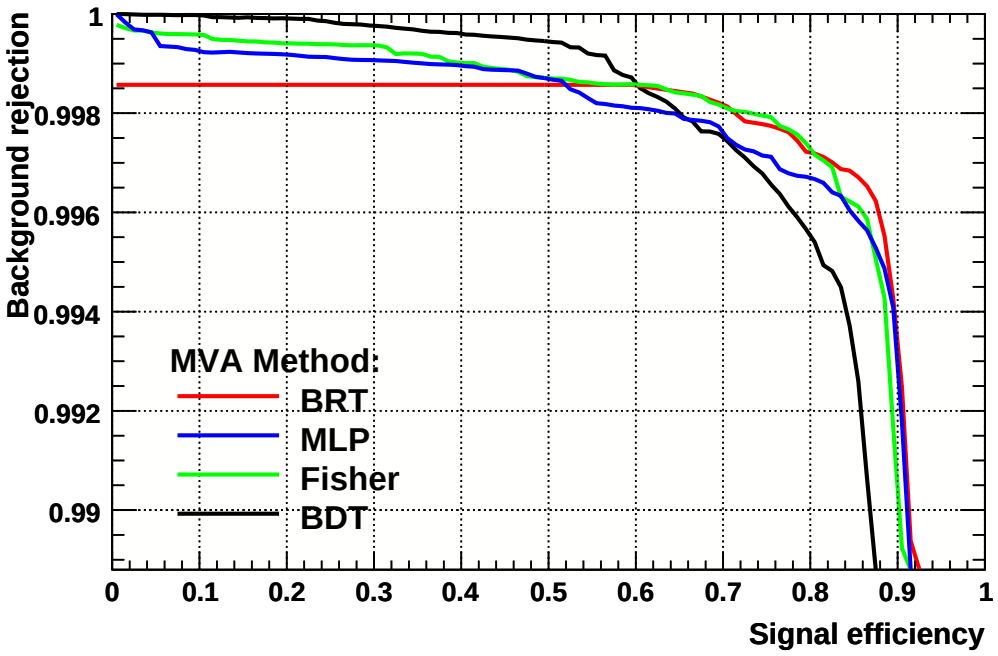


Figure 5.28: ROC Curves for classifiers trained for $m_H = 170$ GeV, 0 jet channel

Output Distributions

To estimate the expected data distribution, the distributions of the classifier response for signal and background test data are histogrammed. In figures 5.31-5.54 the response of all trained classifiers is given: all background samples are scaled to their expected cross section and stacked. The signal

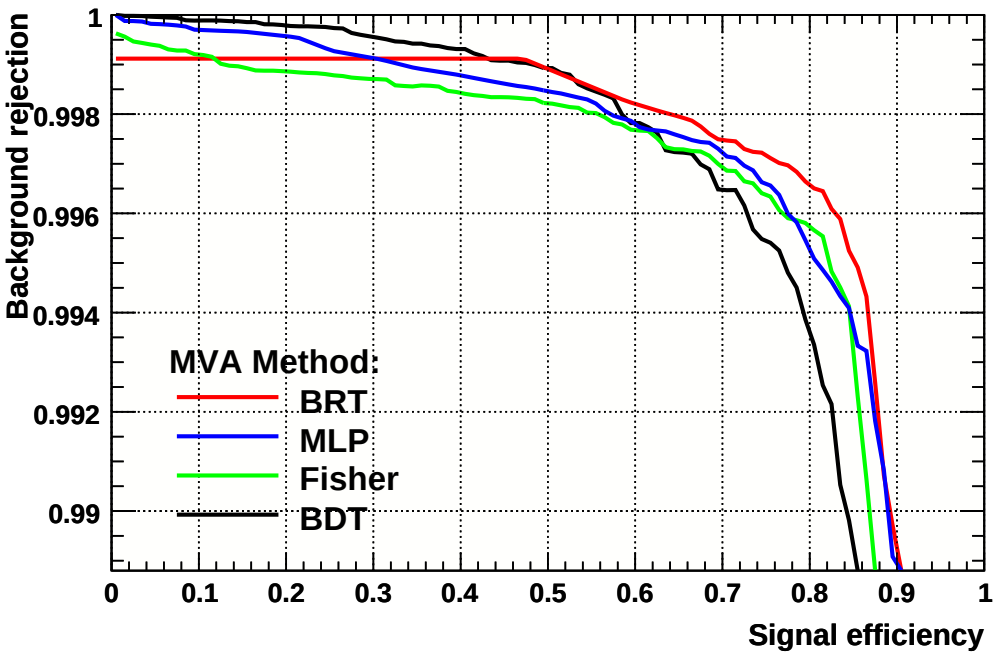


Figure 5.29: ROC Curves for classifiers trained for $m_H = 150$ GeV, 0 jet channel

distribution is then superimposed. These distributions are the most important indicators of classifier performance: ideally, the signal and background should be well separated.

Boosted Decision Trees The most distinctive feature of the boosted decision tree histograms is the almost constant distribution of the signal (see figure 5.31). This is a consequence of the boosting procedure: since every new tree is trained with events reweighted according to the misclassification rate, the new tree does not have to include previously well-classified events. This leads to the situation where only events in an extremly "signal-like" region without much background are classified as signal by virtually every tree, and are thus on the right hand side of the histogram.

Bagged Randomized Trees The BRT classifier uses another boosting procedure, described in section 2.3.3. Each decision tree is trained individually. This leads to the situation where a signal event is classified as a signal by most trees, and therefore the signal distribution peaks strongly at 1 (figure 5.32). This obviously also leads to increased background at high classifier output values, since cuts that remove weak background are not "locally good" for any decision tree trained on a reweighted sample. This is most evident at

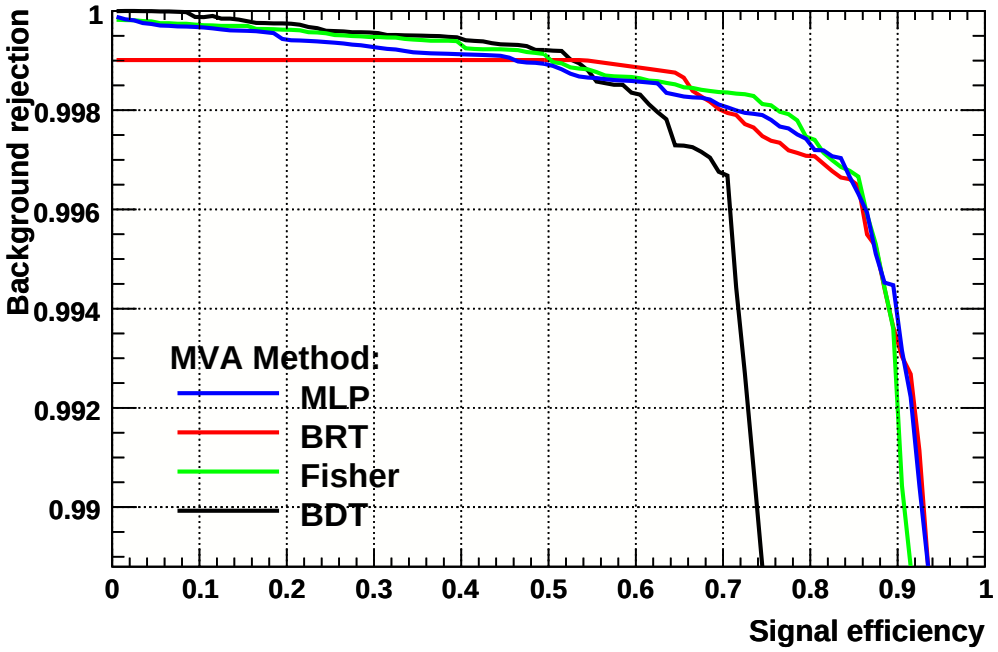


Figure 5.30: ROC Curves for classifiers trained for $m_H = 190$ GeV, 0 jet channel

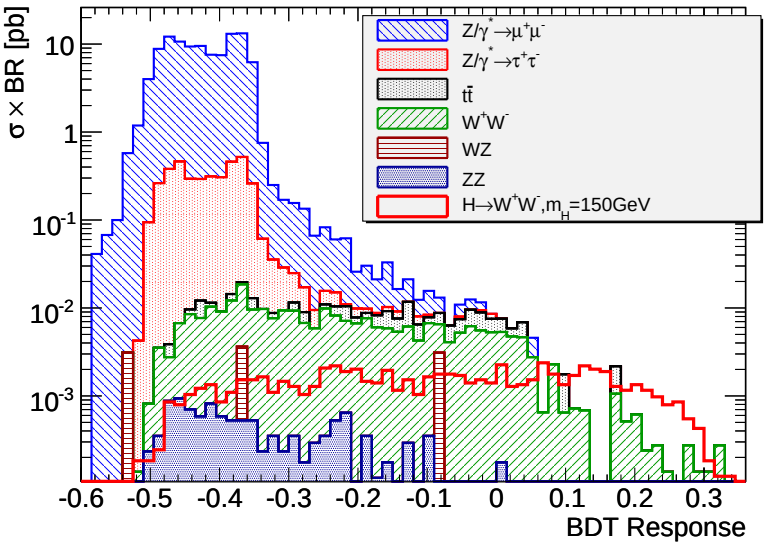


Figure 5.31: Boosted decision tree response in the 0 jet channel, $m_H = 150$ GeV. Most background processes peak at -0.4 and do not extend over 0, QCD W^+W^- decays being the exception. However, the signal is more widely spread as in the other classifiers. The optimal cut value is close to 0.

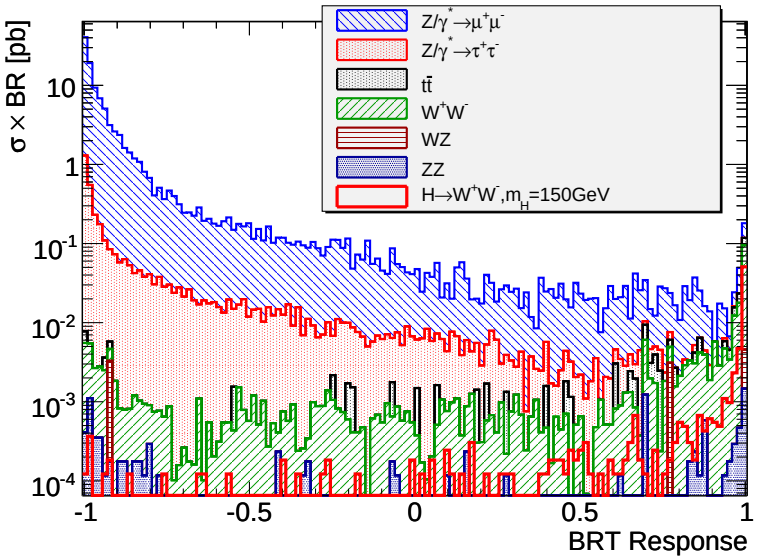


Figure 5.32: Bagged randomized trees response in the 0 jet channel, $m_H = 150$ GeV. Only the well-separable Drell-Yan decays peak at -1, the diboson and Higgs distributions peak at $+1$. The optimal cut value is at 0.995.

the example of the diboson ZZ background in figure 5.32 compared to figure 5.31.

Neural Networks The smooth nature of the response of neural networks is mirrored in the output distribution. Background-like events peak at -1, while signal-like events peak at 1. The extrapolation seen in the previous visualization causes some background events to be classified above $+1$, a behaviour that indicates that these events were not encountered in training.

Fisher classifier The Fisher classifier responses for each background closely resemble a skewed Gaussian. This is not surprising, since the classifier sums up several seemingly "random" variables to arrive at the score, and the central limit theorem then suggests a gaussian output distribution. In the 0 jet channel this separation works quite well, yet in the 2 jet channel the $t\bar{t}$ background is obviously too similar to the signal in the variables used by this linear classifier (see figure 5.50)

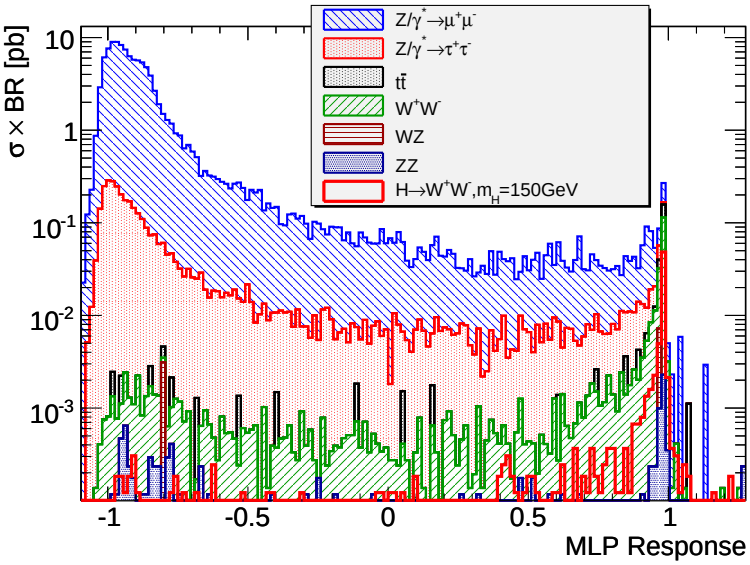


Figure 5.33: Neural network response in the 0 jet channel, $m_H = 150$ GeV. Most backgrounds show a primary peak at -1 and a weaker, secondary peak at $+1$. Some background events are classified higher than $+1$, as discussed in the text. The optimal cut value is at 0.97.

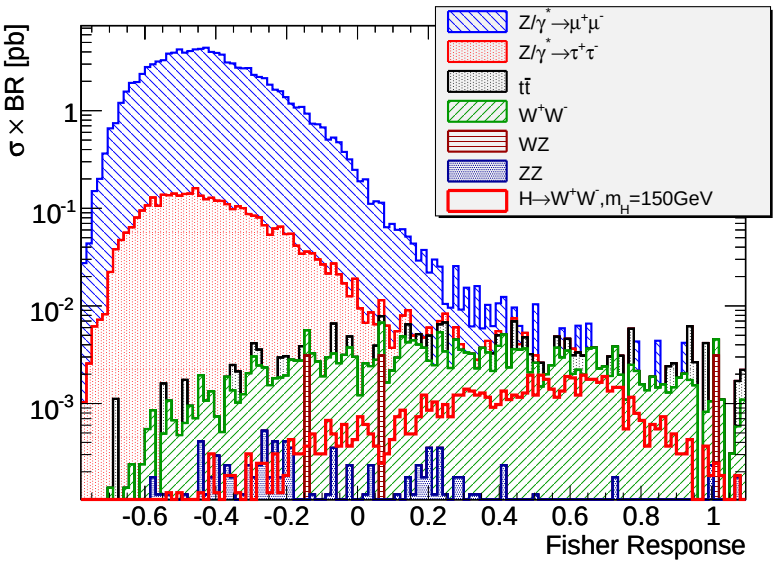


Figure 5.34: Fisher classifier response in the 0 jet channel, $m_H = 150$ GeV. Most backgrounds could be described by a skewed normal distribution. Much of the W^+W^- background can not effectively be separated. The optimal cut value is at 0.31.

$m_H = 170$ GeV, 0 jet channel

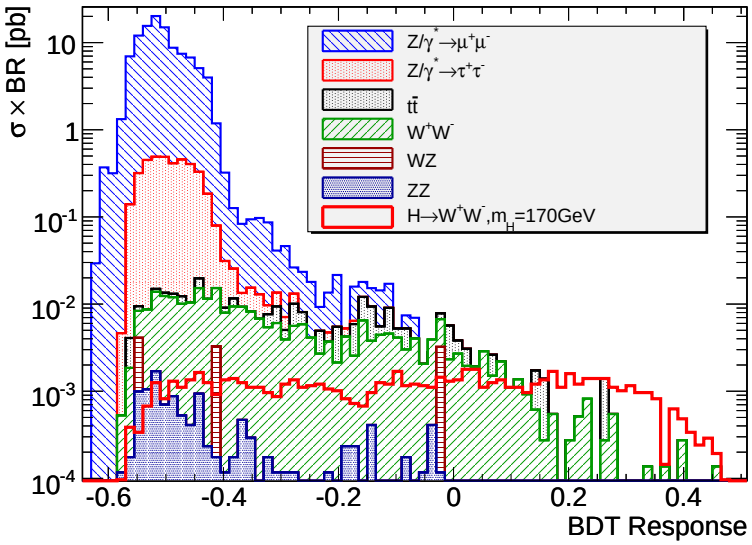


Figure 5.35: Boosted decision tree response in the 0 jet channel, $m_H = 170$ GeV. Very similar to the $m_H = 150$ GeV distribution. The optimal cut value is at -0.08.

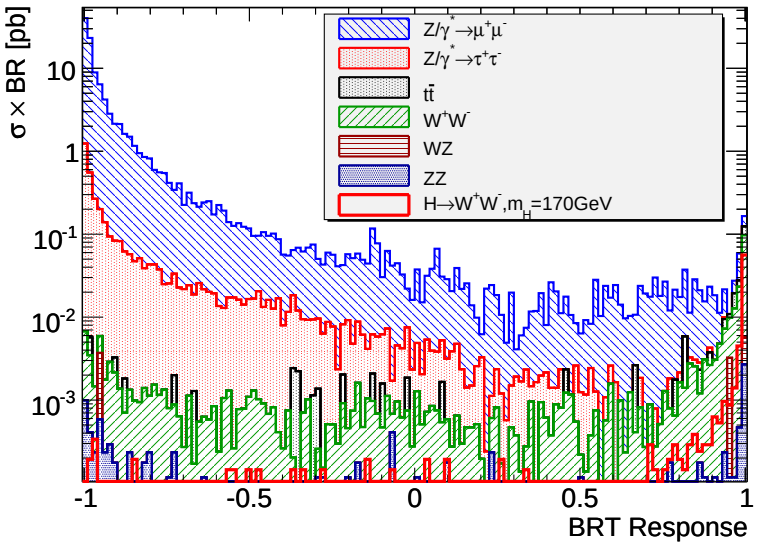


Figure 5.36: Bagged randomized trees response in the 0 jet channel, $m_H = 170$ GeV. The optimal cut value is at 0.995.

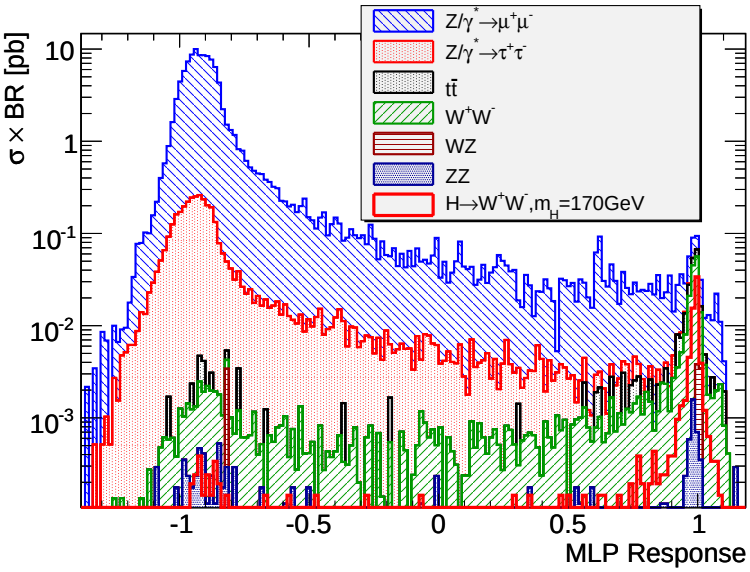


Figure 5.37: Neural network response in the 0 jet channel, $m_H = 170$ GeV. In this case, even more events are classified above +1 as in figure 5.33. The optimal cut value is at 0.97.

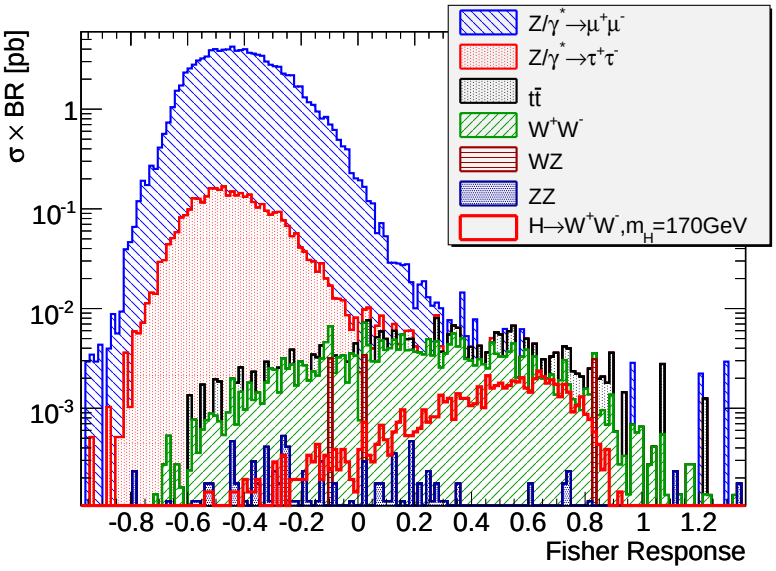


Figure 5.38: Fisher classifier response in the 0 jet channel, $m_H = 170$ GeV. The classifier here performs much better than at $m_H = 150$ GeV. The optimal cut value is at 0.41.

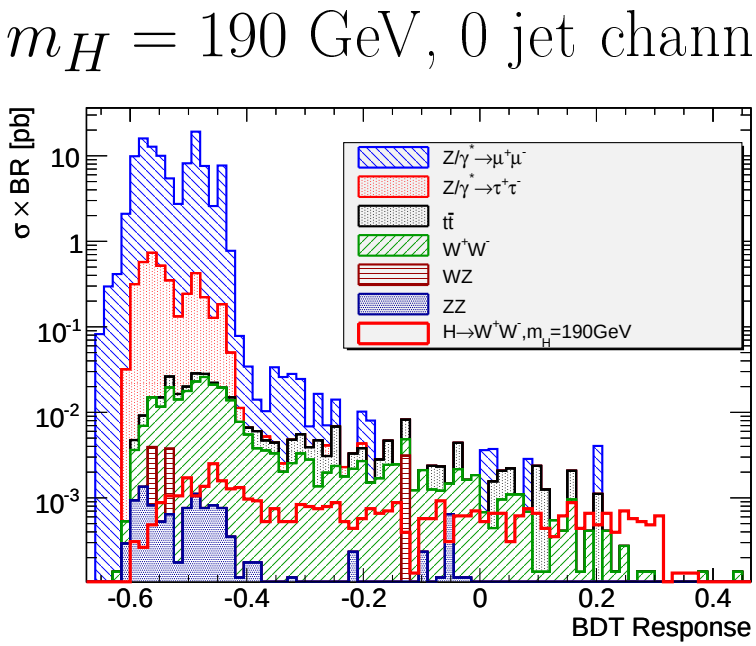


Figure 5.39: Boosted decision tree response in the 0 jet channel, $m_H = 190$ GeV. The optimal cut value is at -0.25.

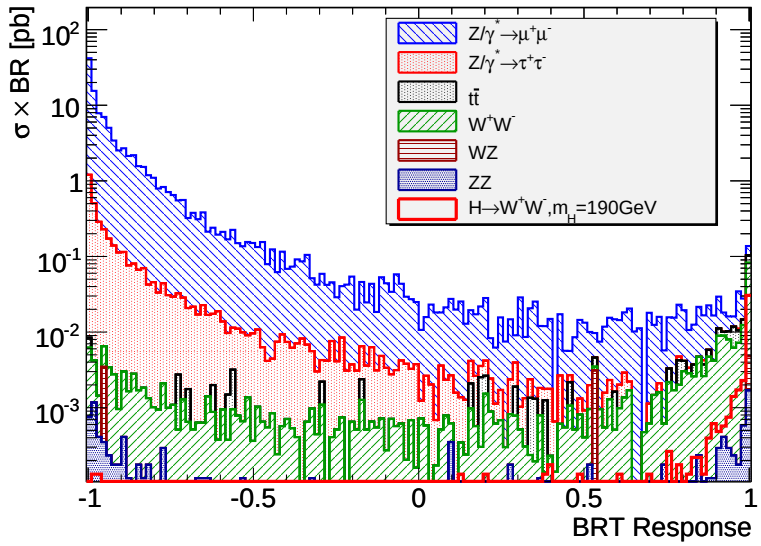


Figure 5.40: Bagged randomized trees response in the 0 jet channel, $m_H = 190$ GeV. The optimal cut value is at 0.995.

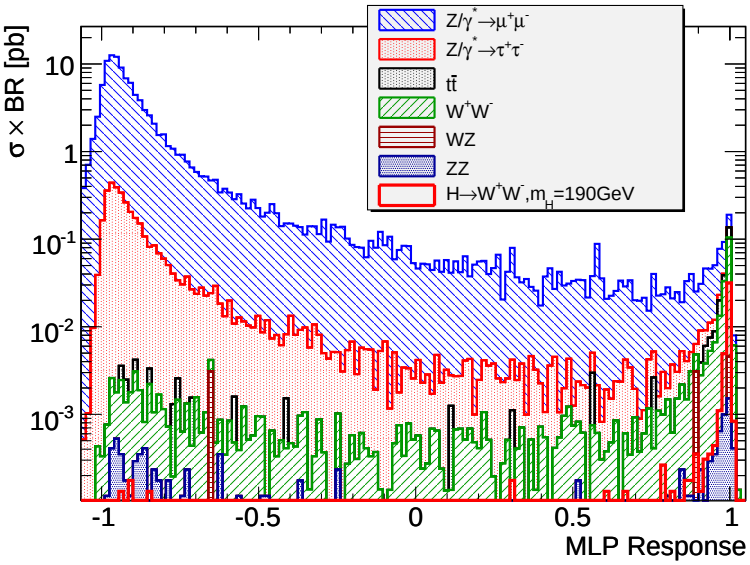


Figure 5.41: Neural network response in the 0 jet channel, $m_H = 190$ GeV. At this Higgs boson mass, almost no background events are classified higher than +1. The optimal cut value is at 0.97.

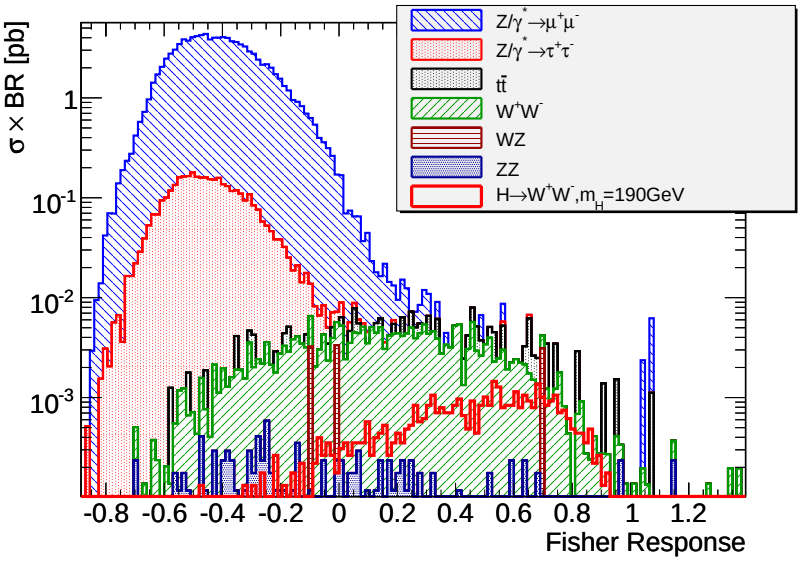


Figure 5.42: Fisher classifier response in the 0 jet channel, $m_H = 190$ GeV. This classifier is particularly efficient at this Higgs boson mass, scoring just below bagged randomized trees and better than boosted decision trees (see table 5.14). The optimal cut value is at 0.29.

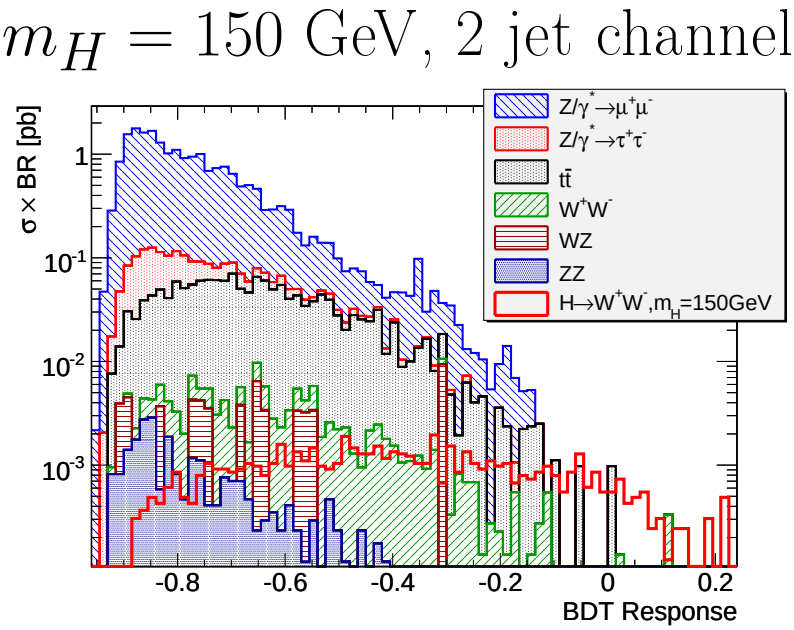


Figure 5.43: Boosted decision tree response in the 2 jet channel, $m_H = 150$ GeV. The different conditions of the 2 jet channel are visible: $t\bar{t}$ decays are much more dominant. In comparison to the cut analysis however this classifier performs exceedingly well (see table 5.15). The optimal cut value is at -0.11.

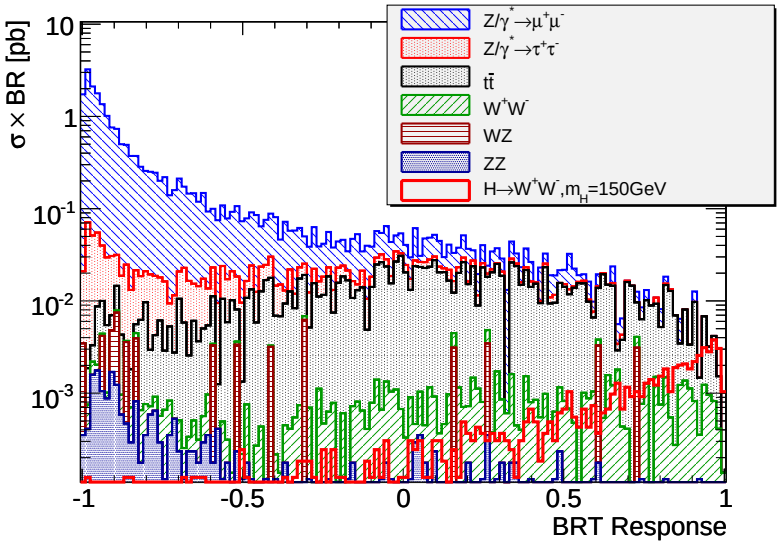


Figure 5.44: Bagged randomized trees response in the 2 jet channel, $m_H = 150$ GeV. Only the signal peaks around $+1$: the dominant $t\bar{t}$ is widely spread with the tail above the optimal cut value of 0.84 approximately the same strength as the signal.

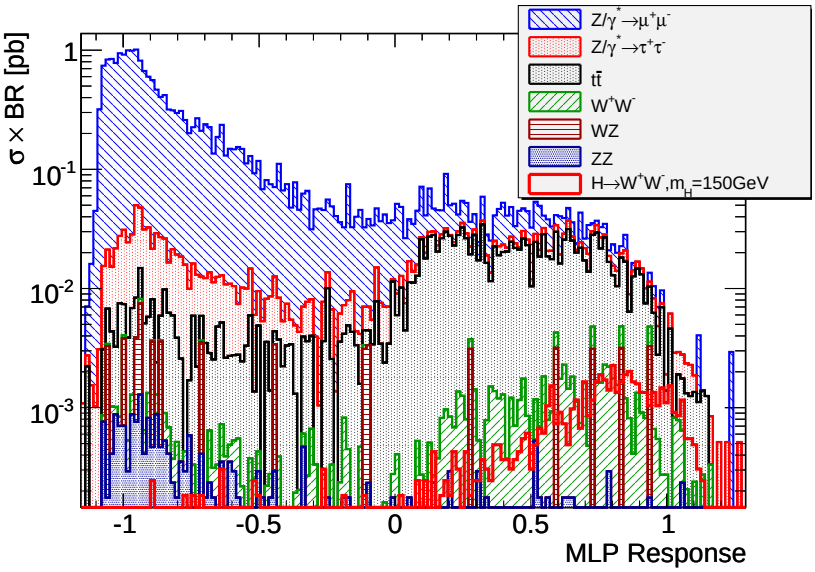


Figure 5.45: Neural network response in the 2 jet channel, $m_H = 150$ GeV. The neural network still has background events classified above +1 in this channel, and the $t\bar{t}$ is approximately six times stronger than the signal above 0.8, the position of the optimal cut value.

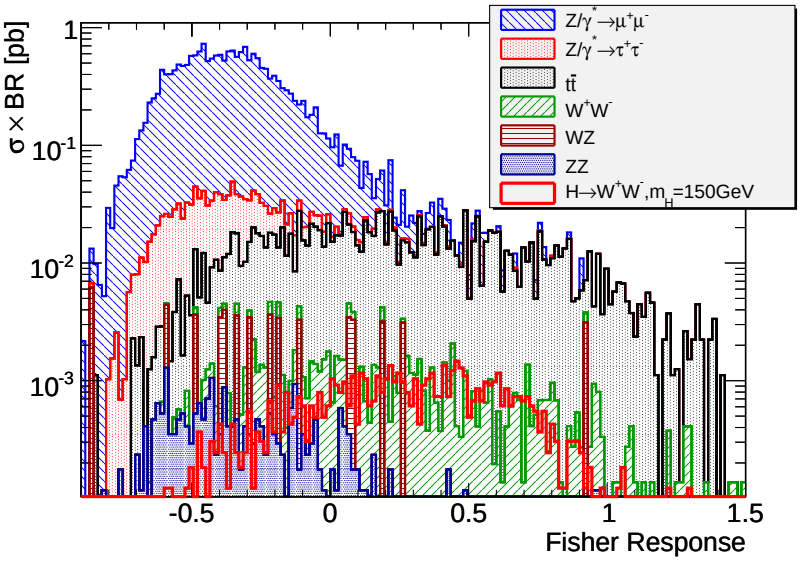


Figure 5.46: Fisher classifier response in the 2 jet channel, $m_H = 150$ GeV. This channel shows the disadvantage of a linear classifier: the $t\bar{t}$ background is virtually not suppressed. The optimal cut value is at 0.61.

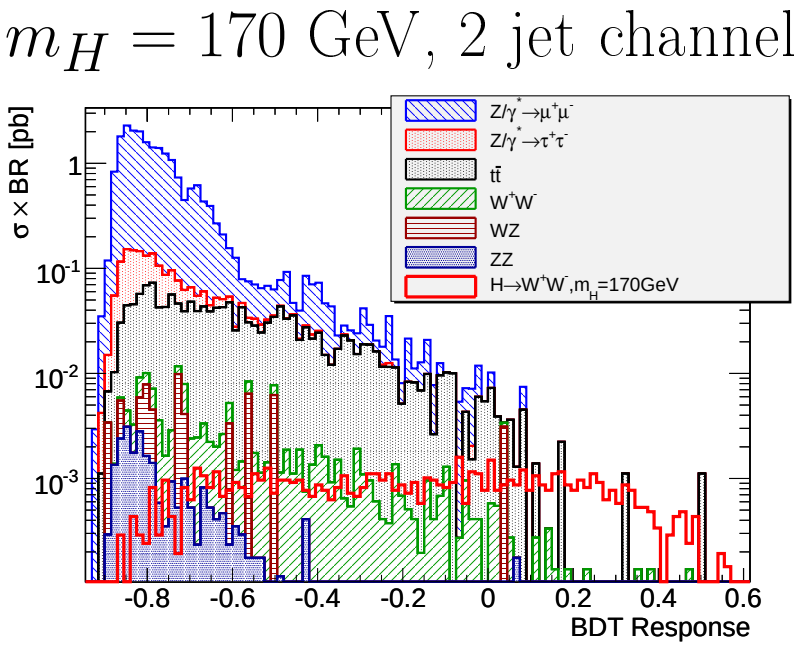


Figure 5.47: Boosted decision tree response in the 2 jet channel, $m_H = 170$ GeV. Also at this Higgs boson mass the boosted decision trees separate signal and background very well: capturing the more complex dynamics of $t\bar{t}$ using a cut analysis seems more difficult (see table 5.16). The optimal cut value is at 0.12.

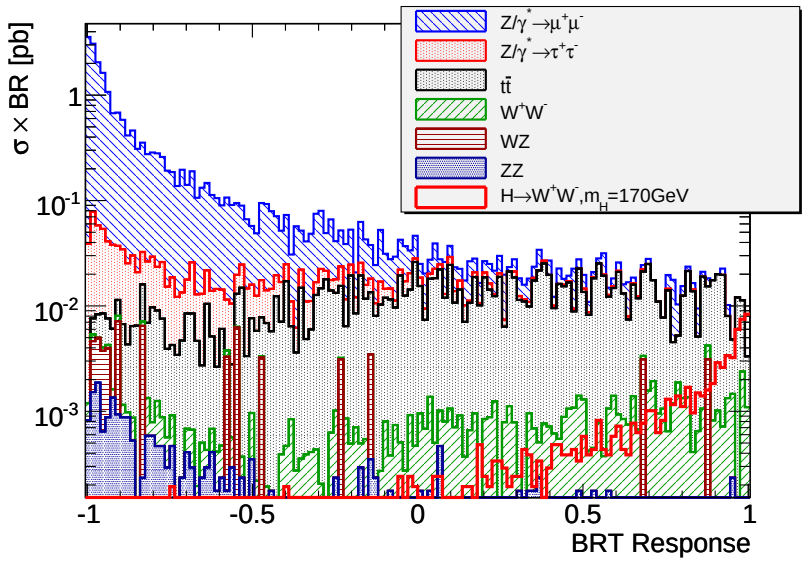


Figure 5.48: Bagged randomized trees response in the 2 jet channel, $m_H = 170$ GeV. The optimal cut value is at 0.952.

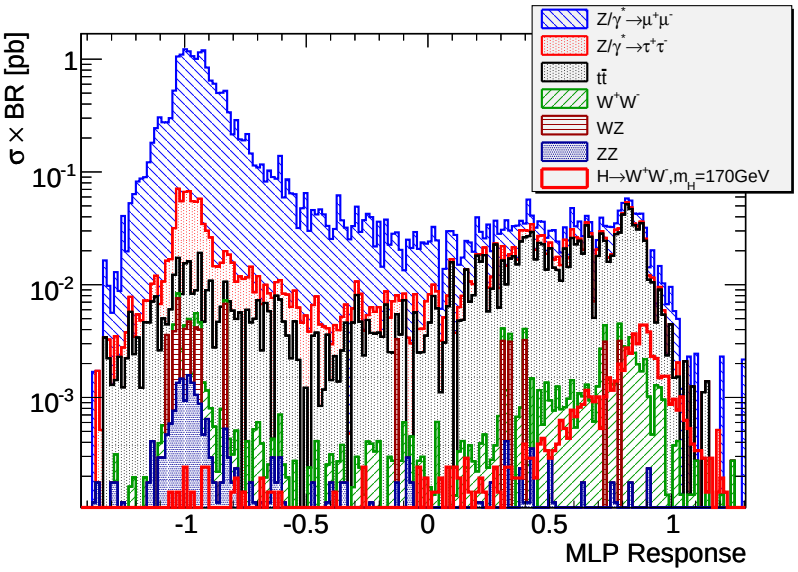


Figure 5.49: Neural network response in the 2 jet channel, $m_H = 170$ GeV. The optimal cut value is at 0.84.

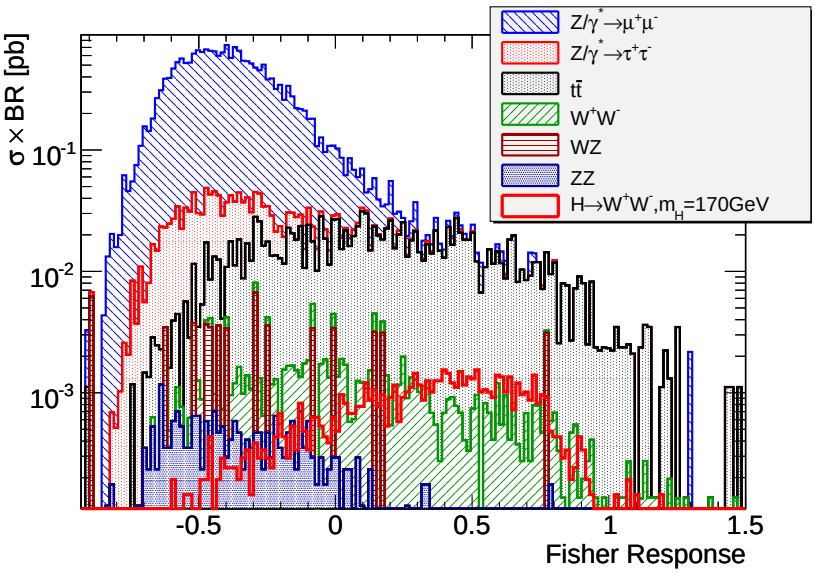


Figure 5.50: Fisher classifier response in the 2 jet channel, $m_H = 170$ GeV. The optimal cut value is at 0.50.

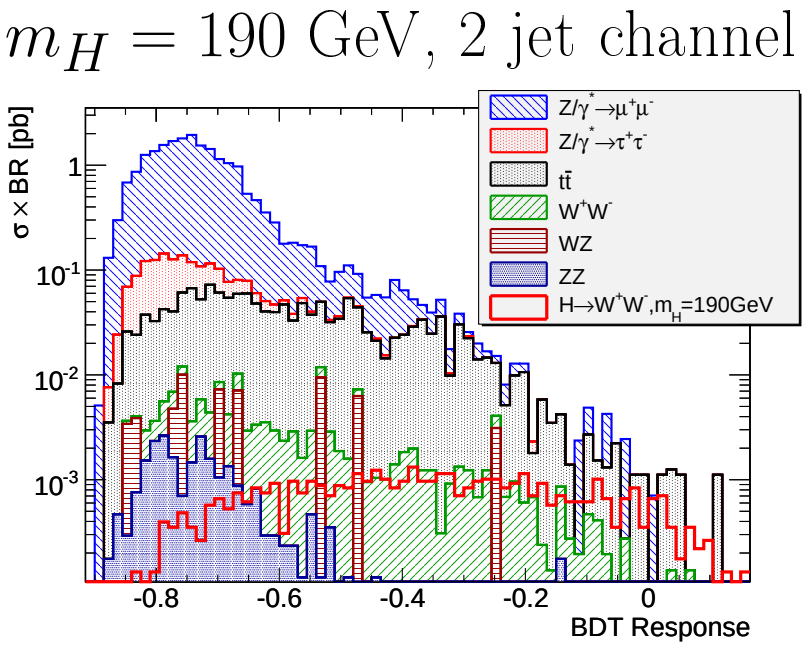


Figure 5.51: Boosted decision tree response in the 2 jet channel, $m_H = 190$ GeV. This is the most difficult separation problem in this thesis. It is problematic for every classifier. Additionally, the signal is weaker than at the other Higgs boson mass points. The optimal cut value is at -0.16.

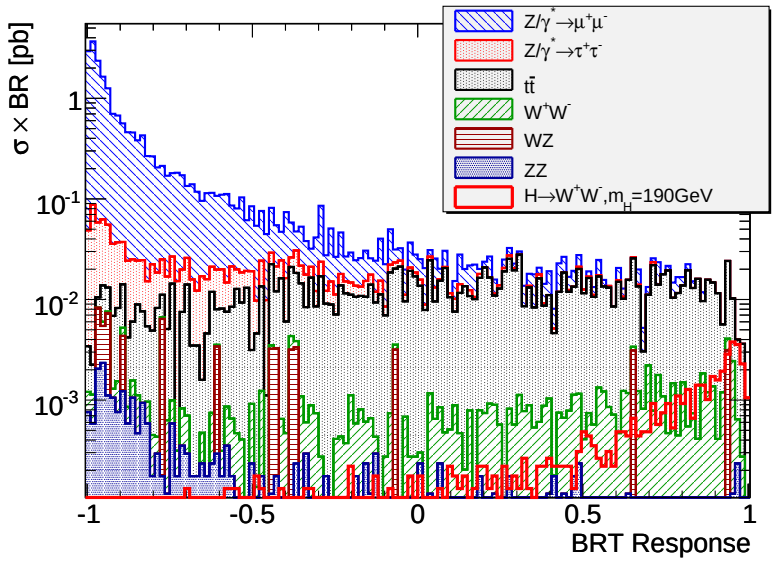


Figure 5.52: Bagged randomized trees response in the 2 jet channel, $m_H = 190$ GeV. The optimal cut value is at 0.898.

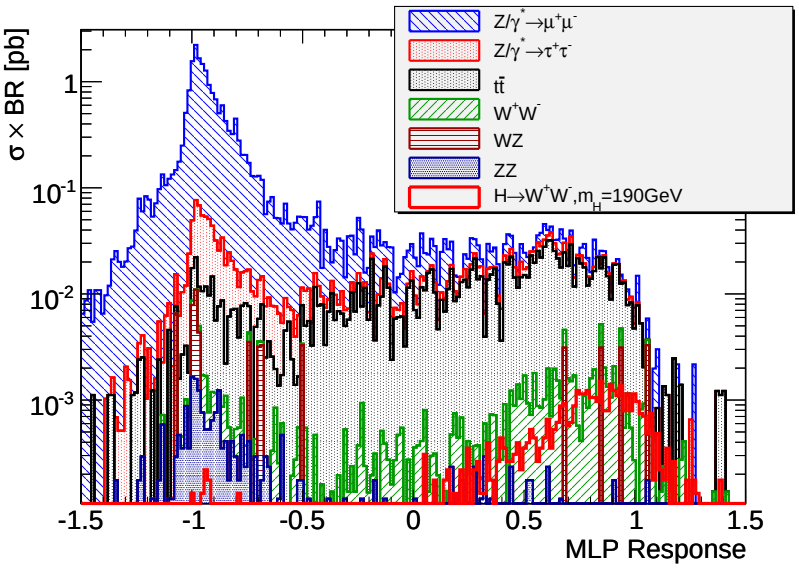


Figure 5.53: Neural network response in the 2 jet channel, $m_H = 190$ GeV. The optimal cut value is at 0.68.

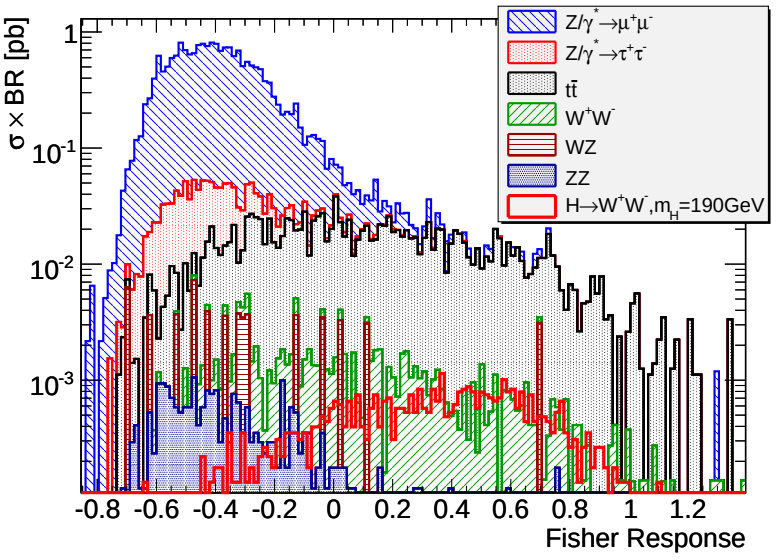


Figure 5.54: Fisher classifier response in the 2 jet channel, $m_H = 190$ GeV. The optimal cut value is at 0.34.

5.7.4 Single Classifier Results

In tables 5.12 - 5.17 the results of an optimized cut on the respective classifier output are given. As in the cut analysis, only the numbers for the most important backgrounds are given: Drell-Yan and W^+W^- decays in the 0 jet and $t\bar{t}$ and W^+W^- decays in the 2 jet channel. The full results are given in Appendix B.

0 jet channel: The first observation to make is that for $m_H = 150$ GeV the cut analysis has actually the best S_{cP} significance (see tables 5.3 and 5.12). This is somewhat surprising since multivariate classifiers should be able to improve on simple cut analysis by taking advantage of correlations. One simple explanation is that for this case the exploitable correlations are more than compensated by the limited training statistics[3].

Another interesting feature seen not only to the 0 jet channel is that using the optimal working point the boosted decision trees select less signal events than the other classifiers: the bagged randomized trees at $m_H = 170$ GeV select more than twice, the other two classifiers approximately 2.5 times the number of signal events. However, this increase in signal events is counterbalanced by the increase in background.

For $m_H = 170$ GeV apparently more useable information could be extracted by most multivariate classifiers: Compared to an $S_{cP} = 3.5$ for the cut analysis the boosted decision trees reached $S_{cP} = 4.43$, a 26% increased significance. Also, the bagged randomized trees and the Fisher classifier obtain a better significance than the simple cut analysis. The neural network is lagging behind, presumably because of insufficient training data - neural networks as more general classifiers need more information to accurately fit to the background distribution.

Only bagged randomized trees and the Fisher classifier showed a better performance than the cut analysis for $m_H = 190$ GeV (see table 5.5 and 5.14). This again points at a situation similar to $m_H = 150$ GeV: the Fisher classifier, for one, can easily be well-trained even with very little training data, in great contrast to neural networks. If it can fit the problem well the performance even exceeds that of complicated multivariate classifiers like boosted decision trees.

2 jet channel: In this channel the advantage of nonlinear combinations of input data becomes evident: the nonlinear classifiers boosted decision trees, bagged randomized trees and neural networks improve upon the cut analysis significance by 272%, 188% and 80%, respectively, for $m_H = 150$ GeV (see

[3]A typical multivariate classifier has many more parameters to train than a cut analysis. Therefore the number of training Monte Carlo events should be greater than necessary for a cut analysis, especially in the case of neural networks

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	253 ± 5.8	$(1.06 \pm 0.01) \cdot 10^5$	75.4 ± 2.1	0.23	0.23
BDT	5.87 ± 7.3	27.6 ± 2.8	38.1 ± 29	20.6 ± 1.6	3.34	3.05
BRT	58.7 ± 19	82.4 ± 4.7	171 ± 35	43.8 ± 2.3	3.35	3.20
MLP	133 ± 27	134 ± 6.0	351 ± 41	52.9 ± 2.5	2.82	2.74
Fisher	92.2 ± 22	130 ± 5.9	298 ± 39	49.2 ± 2.5	2.85	2.76

Table 5.12: Cut flow table for $m_H = 150$ GeV multivariate classifiers, 0 jet channel.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	253 ± 5.8	$(1.06 \pm 0.01) \cdot 10^5$	74.7 ± 1.9	0.23	0.23
BDT	5.87 ± 7.3	46.8 ± 3.6	60.4 ± 29	37.9 ± 1.9	4.88	4.43
BRT	41.1 ± 16	87.0 ± 4.9	165 ± 33	51.6 ± 2.2	4.02	3.81
MLP	150 ± 27	150 ± 6.4	366 ± 42	61.2 ± 2.4	3.20	3.10
Fisher	22 ± 11	83.3 ± 4.8	150 ± 31	47.9 ± 2.1	3.91	3.70

Table 5.13: Cut flow table for $m_H = 170$ GeV multivariate classifiers, 0 jet channel.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	253 ± 5.8	$(1.06 \pm 0.01) \cdot 10^5$	45.1 ± 1.4	0.14	0.14
BDT	26.3 ± 12	44 ± 3.5	95.1 ± 31	22.5 ± 1.4	2.31	2.19
BRT	17.6 ± 11	69.2 ± 4.3	110 ± 31	28.4 ± 1.6	2.71	2.57
MLP	118 ± 25	150 ± 6.4	328 ± 39	38.5 ± 1.8	2.12	2.07
Fisher	26.3 ± 12	100 ± 5.2	176 ± 32	33.8 ± 1.7	2.55	2.45

Table 5.14: Cut flow table for $m_H = 190$ GeV multivariate classifiers, 0 jet channel.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	$(2.14 \pm 0.02) \cdot 10^4$	63.3 ± 2	0.43	0.43
BDT	0± 2.2	1.10 ± 0.83	1.33 ± 29	8.06 ± 0.99	6.99	4.41
BRT	53.5 ± 11	8.61 ± 1.6	68.2 ± 31	18.3 ± 1.5	2.22	2.09
MLP	154 ± 18	15.9 ± 2.1	235 ± 36	23.1 ± 1.7	1.51	1.46
Fisher	355 ± 28	19.0 ± 2.3	410 ± 42	11.7 ± 1.2	0.58	0.56

Table 5.15: Cut flow table for $m_H = 150$ GeV multivariate classifiers, 2 jet channel.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	$(2.14 \pm 0.02) \cdot 10^4$	73.0 ± 1.9	0.50	0.50
BDT	8.92 ± 4.5	1.37 ± 0.87	10.5 ± 29	16.0 ± 1.2	4.94	4.06
BRT	29 ± 8.0	4.38 ± 1.3	33.8 ± 30	20.2 ± 1.4	3.47	3.14
MLP	172 ± 20	15.3 ± 2.1	280 ± 39	33.9 ± 1.8	2.02	1.96
Fisher	355 ± 28	18.5 ± 2.3	410 ± 42	27.0 ± 1.6	1.33	1.30

Table 5.16: Cut flow table for $m_H = 170$ GeV multivariate classifiers, 2 jet channel.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	$(2.14 \pm 0.02) \cdot 10^4$	45.2 ± 1.4	0.31	0.31
BDT	20.1 ± 6.7	4.38 ± 1.3	36.7 ± 30	8.80 ± 0.88	1.45	1.35
BRT	66.9 ± 12	5.48 ± 1.4	79.1 ± 31	13.4 ± 1.1	1.50	1.43
MLP	330 ± 27	32.6 ± 3.0	480 ± 44	27.4 ± 1.6	1.25	1.22
Fisher	466 ± 32	24.6 ± 2.6	554 ± 46	23.9 ± 1.4	1.01	0.99

Table 5.17: Cut flow table for $m_H = 190$ GeV multivariate classifiers, 2 jet channel.

tables 5.8 and 5.15). This becomes especially visible in the $t\bar{t}$ suppression: boosted decision trees almost completely suppress this background with a signal slightly smaller than in the cut analysis, the bagged randomized trees manage to double the signal while maintaining approximately the same level of background. Neural networks increase the signal even further, but pay with a much increased background. The linear Fisher classifier can not compete in this situation.

Much the same situation is present at $m_H = 170$ GeV, at $m_H = 190$ GeV the significance of all classifiers drops, the separation problem becomes apparently more difficult. Yet still, the nonlinear classifiers provide a $22\% - 43\%$ improvement on the cut analysis (see table 5.17).

5.7.5 Combined Neural Network Results

In this section the results for the approach of combined neural networks is presented. This approach separates the backgrounds into three groups, Drell-Yan, QCD and W^+W^-, as described in section 5.7. Against each background group a neural network is trained, and then the cuts on the outputs of the three classifiers are jointly optimized. The classifier for the QCD background including $t\bar{t}$ was replaced by a Fisher classifier in the 0 jet channel, since the number of training events were not sufficient to train an efficient neural network. In the 2 jet channel the training statistics for the QCD background were sufficient, but for the W^+W^- decays a Fisher classifier was used for the same reason. In the cut flows first the cut on Drell-Yan, then the cut on QCD and finally the cut on W^+W^- is performed.

0 jet channel: In the 0 jet channel this method gives excellent performance, except for $m_H = 190$ GeV, where a statistical fluctuation in the training data probably indicated that a very strong cut on the output of the W^+W^- classifier has been advantageous, in contrast to the test data. However, for $m_H = 150$ GeV and $m_H = 170$ GeV the combination of multiple classifiers improved the significance respective to the neural nets by 25% respective 34%.

2 jet channel: In the two-jet channel this combined classifier actually has a worse performance than the directly trained neural network. It seems therefore that the signal region in this channel can better be selected by the training algorithms if all backgrounds are used for training together. This can also be seen as positive: a neural network can be trained against the total background, and seperate training is not necessary to improve the classifier in this situation.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	213 ± 7.5	$(8.30 \pm 0.07) \cdot 10^4$	68.1 ± 2.9	0.24	0.24
Drell-Yan Cut	86.6 ± 22	118 ± 5.7	273 ± 38	52.3 ± 2.5	3.16	3.05
QCD Cut	74.8 ± 21	115 ± 5.6	234 ± 36	50.2 ± 2.5	3.28	3.15
W^+W^- Cut	23.5 ± 12	57.8 ± 4	94.3 ± 31	35.8 ± 2.1	3.69	3.45

Table 5.18: Cut flow table for $m_H = 150$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	213 ± 7.5	$(8.30 \pm 0.07) \cdot 10^4$	70.3 ± 2.6	0.24	0.24
Drell-Yan Cut	216 ± 34	152 ± 6.4	461 ± 47	62.5 ± 2.5	2.91	2.84
QCD Cut	73.0 ± 20	72.7 ± 4.5	173 ± 35	42.6 ± 2.0	3.24	3.10
W^+W^- Cut	5.87 ± 7.3	35.9 ± 3.2	49.3 ± 29	32.2 ± 1.8	4.59	4.15

Table 5.19: Cut flow table for $m_H = 170$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	213 ± 7.5	$(8.30 \pm 0.07) \cdot 10^4$	38.9 ± 1.9	0.14	0.13
Drell-Yan Cut	178 ± 31	146 ± 6.3	415 ± 45	34.3 ± 1.7	1.69	1.65
QCD Cut	98.1 ± 23	95.7 ± 5.1	229 ± 37	27.2 ± 1.6	1.80	1.74
W^+W^- Cut	0± 7.3	9.52 ± 1.7	12 ± 29	7.66 ± 0.82	2.21	1.94

Table 5.20: Cut flow table for $m_H = 190$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	$(1.78 \pm 0.03) \cdot 10^4$	66.7 ± 2.9	0.50	0.50
Drell-Yan Cut	430 ± 31	25.8 ± 2.7	525 ± 43	26.4 ± 1.8	1.15	1.13
QCD Cut	20.1 ± 6.7	3.69 ± 1.1	25.1 ± 30	7.94 ± 0.98	1.58	1.45
W^+W^- Cut	15.6 ± 5.9	2.19 ± 0.95	19.2 ± 29	7.45 ± 0.95	1.70	1.54

Table 5.21: Cut flow table for $m_H = 150$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	$(1.78 \pm 0.03) \cdot 10^4$	79.7 ± 2.8	0.60	0.59
Drell-Yan Cut	321 ± 27	18.3 ± 2.3	368 ± 40	26.2 ± 1.6	1.37	1.34
QCD Cut	8.92 ± 4.5	3.01 ± 1.1	17.9 ± 29	8.30 ± 0.90	1.96	1.76
W^+W^- Cut	8.92 ± 4.5	1.10 ± 0.83	16.0 ± 29	7.33 ± 0.84	1.83	1.64

Table 5.22: Cut flow table for $m_H = 170$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$t\bar{t}$	WW	All Backgrounds	Signal	$S/\sqrt{B}$	S_{cP}
Preselection	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	$(1.78 \pm 0.03) \cdot 10^4$	46.6 ± 2.0	0.35	0.35
Drell-Yan Cut	$(1.03 \pm 0.05) \cdot 10^3$	48.4 ± 3.6	$(1.26 \pm 0.06) \cdot 10^3$	36.1 ± 1.8	1.02	1.00
QCD Cut	51.3 ± 11	4.65 ± 1.3	68.6 ± 31	8.27 ± 0.85	1.00	0.94
W^+W^- Cut	51.3 ± 11	4.65 ± 1.3	68.6 ± 31	8.27 ± 0.85	1.00	0.94

Table 5.23: Cut flow table for $m_H = 190$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

5.8 Monte Carlo Statistical Uncertainty

The statistical uncertainty of the Monte Carlo data is relatively high in this study, making an accurate estimate important. However, for an extended study with real data this uncertainty would very probably be smaller, because more Monte Carlo events will be available. For this reason this uncertainty is discussed separately from the other systematic uncertainties, where the treatment would be much the same in an analysis with more Monte Carlo data.

Doing a counting experiment using Monte Carlo background subtraction is a daunting and difficult task. Using the procedure described in section 4.2.1 one can derive a posterior probability distribution function (p.d.f.) from the number of background Monte Carlo events in the signal region. This was done for each method. The main characteristics of the resulting p.d.f.s are summarized in tables 5.24 to 5.29 by using several values:

- the mean of the p.d.f. $\hat{B}$ (this is the expected background $\sigma \times BR$)
- the root mean square width of the p.d.f., σ_{stat}.

For a gaussian p.d.f. this information would be sufficient, and the maximal signal significance would be $\hat{S}/\sigma_{stat}$. The tails of the actual distribution are, however, much heavier than those of a gaussian. Therefore additional information is given:

- the maximal significance of the expected signal $\hat{S}$, given by the integral $\int_{\hat{B}+\hat{S}}^{\infty} p(x)\mathrm{d}x$ of the background p.d.f. $p(x)$
- the 3σ and 5σ upper limits for the background from the p.d.f. in fb
- and for comparison: the S_{cP} of the expected signal at an integrated luminosity of 1 fb^{-1} for the given background uncertainty σ_{stat}[4]

[4]The S_{cP} method can at the moment only treat gaussian systematic errors

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	110 ± 30	23.1 ± 1.7	1.29	0.728	221	366
BDT	114 ± 30	20.6 ± 1.6	1.23	0.648	225	370
BRT	244 ± 37	43.8 ± 2.3	1.56	1.08	371	520
MLP	422 ± 43	52.9 ± 2.5	1.58	1.10	563	716
Comb MLP	162 ± 32	35.8 ± 2.1	1.51	1.03	277	423
Fisher	370 ± 41	49.2 ± 2.5	1.56	1.07	508	661

Table 5.24: Monte Carlo statistical uncertainty in the 0 jet channel for m_H = 150 GeV. Significance is given in equivalent standard deviations (σ).

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	122 ± 31	26.1 ± 1.6	1.32	0.783	236	382
BDT	136 ± 30	37.9 ± 1.9	1.61	1.17	247	393
BRT	239 ± 36	51.6 ± 2.2	1.74	1.32	362	510
MLP	440 ± 44	61.2 ± 2.4	1.70	1.24	584	739
Comb MLP	119 ± 30	32.2 ± 1.8	1.49	1.00	230	375
Fisher	224 ± 34	47.9 ± 2.1	1.71	1.27	344	491

Table 5.25: Monte Carlo statistical uncertainty in the 0 jet channel for m_H = 170 GeV. Significance is given in equivalent standard deviations (σ).

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	90.6 ± 29	9.33 ± 0.91	0.982	0.303	200	345
BDT	171 ± 32	22.5 ± 1.4	1.22	0.644	286	432
BRT	184 ± 33	28.4 ± 1.6	1.33	0.789	303	450
MLP	400 ± 41	38.5 ± 1.8	1.37	0.831	537	689
Comb MLP	83.5 ± 29	7.66 ± 0.82	0.943	0.250	193	338
Fisher	250 ± 35	33.8 ± 1.7	1.41	0.878	371	519

Table 5.26: Monte Carlo statistical uncertainty in the 0 jet channel for m_H = 190 GeV. Significance is given in equivalent standard deviations (σ).

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	139 ± 31	9.16 ± 1.1	0.951	0.275	251	397
BDT	78.6 ± 29	8.06 ± 0.99	0.954	0.260	188	333
BRT	144 ± 31	18.3 ± 1.5	1.15	0.541	258	403
MLP	308 ± 37	23.1 ± 1.7	1.15	0.555	434	582
Comb MLP	90.1 ± 29	7.45 ± 0.95	0.933	0.241	200	345
Fisher	485 ± 43	11.7 ± 1.2	0.895	0.242	622	773

Table 5.27: Monte Carlo statistical uncertainty in the 2 jet channel for m_H = 150 GeV. Significance is given in equivalent standard deviations (σ).

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	100 ± 30	9.94 ± 0.98	0.986	0.315	211	356
BDT	87.6 ± 29	16.0 ± 1.2	1.13	0.519	197	342
BRT	110 ± 30	20.2 ± 1.4	1.21	0.631	221	367
MLP	354 ± 40	33.9 ± 1.8	1.31	0.758	486	635
Comb MLP	86.8 ± 30	7.33 ± 0.84	0.926	0.233	198	343
Fisher	485 ± 43	27.0 ± 1.6	1.15	0.559	622	772

Table 5.28: Monte Carlo statistical uncertainty in the 2 jet channel for m_H = 170 GeV. Significance is given in equivalent standard deviations (σ).

method	$\hat{B} \pm \sigma_{stat}$ [fb]	$\hat{S} \pm \sigma_{stat}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}	B upper limit 3 σ	B upper limit 5 σ
Cuts	138 ± 31	8.01 ± 0.84	0.922	0.239	251	397
BDT	113 ± 31	8.80 ± 0.88	0.946	0.269	226	371
BRT	155 ± 32	13.4 ± 1.1	1.03	0.389	269	415
MLP	554 ± 45	27.4 ± 1.6	1.13	0.536	697	850
Comb MLP	137 ± 32	8.27 ± 0.85	0.922	0.242	252	398
Fisher	628 ± 47	23.9 ± 1.4	1.05	0.447	775	928

Table 5.29: Monte Carlo statistical uncertainty in the 2 jet channel for m_H = 190 GeV. Significance is given in equivalent standard deviations (σ).

These numbers, especially the upper limits, clearly show that counting events without using actual data to constrain certain backgrounds is absolutely futile even when one only looks at the uncertainty of Monte Carlo statistics. However, some information on the classifiers can still be extracted from these numbers.

One observation is that the significance numbers increase with increasing signal cross-section: methods with very hard cuts are penalized. This is actually a valid concern: hard cuts lead to less background statistics, which means the background in the respective area is statistically less well understood. This favors especially the neural networks, and shows their strong point.

Another interesting point can be seen for example in table 5.27 comparing the bagged randomized trees and the neural networks. Both have the same limit significance, although the background for the neural network is almost doubled compared to the BRT, and the signal is not very much stronger for the neural networks. The reason is that the $t\bar{t}$ background that contributes to the bulk of the increase has comparatively good Monte Carlo statistics, and the increase in signal for the neural network compensates for the increased statistical uncertainty.

Finally one can note that the cut analysis is affected strongest by this uncertainty, since the optimal cuts tend to be relatively strict. The same applies to a lesser degree to the combined neural networks, and even less to the boosted decision trees. Bagged randomized trees and especially neural networks show the best performances, if only this uncertainty was important.

5.9 Systematic Uncertainties

Now leaving out the uncertainty by limited Monte Carlo statistics we can examine systematic uncertainties from detector resolution and scale uncertainties: on muon resolution and scale, jet resolution and scale, and the resolution of the missing transverse energy. Using the method described in 4.2.2 and the values in table 4.1 the systematical shifts have been calculated for every method. The results are summarized in tables 5.30 to 5.35. The tables contain the expected background and signal values together with two sets of uncertainties. $\sigma_{S,ssq}$ and $\sigma_{B,ssq}$ are calculated using the traditional method (sum of squares): each individual uncertainty is varied by $\pm 1\sigma$ and the squares of the shifts are summed up and the square root is taken. This would be accurate for systematic uncertainties that are linearly increasing and not correlated in the directions of variation. $\sigma_{S,mc}$ and $\sigma_{B,mc}$ however are calculated by taking 400 random "distortion" points and evaluating the analysis with the chosen systematic uncertainties. The mean square of the distances to the undistorted result is then taken as the the uncertainty on the

result (see section 4.2.2). The limit of the expected signal $\hat{S}$ significance has the same meaning as in the tables for the Monte Carlo statistical uncertainties: the probability in equivalent standard deviations that the background B with uncertainty $\sigma_{B,mc}$ is actually as large as $\hat{S}+\hat{B}$. The S_{cP} for 1 fb^{-1} is calculated with $\sigma_{B,mc}$ as systematical uncertainty on the background.

Already in table 5.30 one can see that these two methods can strongly differ - the uncertainty on the background by Monte Carlo is double that given by $\pm 1\sigma$ variation for the cut analysis, bagged randomized trees, and combined neural networks, even five times higher for neural networks. If one also looks at the other tables, one can see that the two methods roughly agree for the cut analysis. Since the traditional method is used widely to estimate systematic uncertainties in cut analyses, this is a reassuring result. However, looking especially at the neural network results, one can see that the traditional method is not favorable to estimate systematic uncertainties for nonlinear classifiers: In the 0 jet channel, the uncertainty on the neural network response is underestimated by a factor of five.

Looking at the systematic values producing the greatest shift in neural net output reveals that a decrease in missing transverse energy resolution coupled with other systematic uncertainties causes a large number of background events to be classified as signal. This is probably due to the extrapolating nature of the neural network, where events can be assigned high scores even though they are not in the region where the network has been well trained. A similar effect can be seen in figure 5.37, where several test events are assigned a score higher than one.

If only these systematic uncertainties from the detector would be considered, and no constraints by side-bands imposed, the traditional cut analysis gives consistently the best results. This result confirms a common expectation: by using sophisticated classifiers, high significance under good conditions is traded against robustness. However, the boosted trees also show a very robust behaviour. The combined neural networks also show a more robust behaviour than the network trained against all backgrounds. The Fisher classifier and the neural network however show very great sensitivity to these systematic uncertainties in both channels at all Higgs boson masses.

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	S_{cP}@1 fb^{-1}
Cuts	33.8	7.05	14.9	23.1	1.96	2.0	1.53	1.38
BDT	38.1	11.5	9.34	20.6	4.52	2.51	2.14	1.70
BRT	171	50.9	106	43.8	4.14	3.13	0.415	0.364
MLP	351	93.5	497	52.9	5.58	3.46	0.106	-0.246
Comb MLP	94.3	41.4	97.8	35.8	3.87	2.53	0.366	0.184
Fisher	298	77.6	277	49.2	5.16	3.59	0.178	0.00858

Table 5.30: Uncertainty due to systematics in the 0 jet channel for $m_H = 150$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}
Cuts	45.8	10.7	10.0	26.1	3.99	2.76	2.50	1.99
BDT	60.4	25.4	27.7	37.9	6.0	3.35	1.36	1.29
BRT	165	39.4	65.5	51.6	6.37	3.97	0.787	0.767
MLP	366	167	786	61.2	7.37	4.57	0.0779	-0.477
Comb MLP	49.3	9.45	13.3	32.2	5.69	3.64	2.34	1.99
Fisher	150	35.5	49.2	47.9	5.58	3.90	0.970	0.939

Table 5.31: Uncertainty due to systematics in the 0 jet channel for $m_H = 170$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}
Cuts	13.7	4.09	4.93	9.33	2.65	1.63	1.80	1.34
BDT	95.1	17.8	23.7	22.5	2.43	1.71	0.949	0.866
BRT	110	30.4	40.8	28.4	2.68	1.89	0.695	0.668
MLP	328	128	703	38.5	4.30	2.70	0.0547	-0.515
Comb MLP	12	3.54	3.03	7.66	1.99	1.18	2.36	1.04
Fisher	176	27.7	86.1	33.8	2.82	2.10	0.392	0.370

Table 5.32: Uncertainty due to systematics in the 0 jet channel for $m_H = 190$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}
Cuts	62.3	14.9	9.01	9.16	1.39	0.979	1.01	0.749
BDT	1.33	6.19	11.0	8.06	2.01	0.950	0.729	0.143
BRT	68.2	34.1	43.3	18.3	1.40	1.06	0.423	0.362
MLP	235	108	114	23.1	1.40	1.15	0.202	0.182
Comb MLP	19.2	9.40	12.9	7.45	1.09	0.704	0.576	0.481
Fisher	410	93.3	81.8	11.7	3.62	2.60	0.143	0.139

Table 5.33: Uncertainty due to systematics in the 2 jet channel for $m_H = 150$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}
Cuts	23.6	11.6	9.37	9.94	1.61	0.979	1.06	0.898
BDT	10.5	13.9	23	16.0	2.23	1.28	0.697	0.353
BRT	33.8	16.1	19.3	20.2	2.20	1.44	1.04	0.955
MLP	280	135	158	33.9	1.96	1.86	0.215	0.177
Comb MLP	16.0	12.7	11.7	7.33	1.59	0.943	0.623	0.505
Fisher	410	72.2	69.9	27.0	4.80	3.59	0.386	0.370

Table 5.34: Uncertainty due to systematics in the 2 jet channel for $m_H = 170$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

method	$\hat{B}$ [fb]	$\sigma_{B,ssq}$ [fb]	$\sigma_{B,mc}$ [fb]	$\hat{S}$ [fb]	$\sigma_{S,ssq}$ [fb]	$\sigma_{S,mc}$ [fb]	$\hat{S}$ significance limit [σ]	$\hat{S}$ significance S_{cP}@1 fb^{-1}
Cuts	61.7	13.8	11.3	8.01	0.821	0.434	0.706	0.567
BDT	36.7	29.7	49.2	8.80	1.59	1.26	0.179	-0.131
BRT	79.1	18.1	23.5	13.4	1.31	1.01	0.568	0.525
MLP	480	154	214	27.4	2.35	1.33	0.128	0.115
Comb MLP	68.6	9.68	20.6	8.27	0.835	0.523	0.402	0.372
Fisher	554	131	105	23.9	3.31	2.16	0.227	0.222

Table 5.35: Uncertainty due to systematics in the 2 jet channel for $m_H = 190$ GeV. Negative S_{cP} values indicate high systematic uncertainties on the background

5.10 Result Overview

Finally, we can include all uncertainties into one total uncertainty and then calculate the significance of the expected signal at 1 fb^{-1}. This is done by using the S_{cP} formalism with the square root of the sum of squares of the considered uncertainties, detector systematics and Monte Carlo statistics, as the background uncertainty. The results are given in given in tables 5.36 to 5.39, where also the S_{cP} results considering no uncertainties (naive), only considering Monte Carlo statistical uncertainty (stat) and only considering the detector systematical uncertainty (syst) are repeated.

The resulting numbers do not look very promising, but one has to take into account that in the absence of real data neither background nor systematic uncertainties can be constrained.

In the 0 jet channel, boosted decision trees and bagged randomized trees are consistently among the best classifiers. The combined neural networks give best performance for $m_H = 170$ GeV, yet perform poorly at the other Higgs boson masses. Still, tuning the number of neurons in the network would probably increase the performance in these channels. For neural networks trained against the whole background the bad performance in the systematic study causes an overall bad result - improvements in this region could for example include training the network also with events distorted using models of the systematic uncertainties.

For the 2 jet channel the picture changes: the boosted decision trees are strongly affected by the systematical uncertainties, and trail all other classifiers. One reason is the extremely small signal region and the small cross-section of the selected signal that is then easily swamped by the combination of statistical and systematic uncertainties. The randomized version, however, even increases its relative performance, and performs much better than most other classifiers: only the cut analysis at $m_H = 150$ GeV, relatively untouched by the systematic uncertainties, shows better results. The Fisher classifier, despite trailing in the performace without uncertainties performs well in the total significance. The neural net, yielding always the largest signal cross-section, also does well in the final comparison.

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	23.1	33.8	3.57	0.728	1.38	0.538
BDT	20.6	38.1	3.04	0.648	1.70	0.541
BRT	43.8	171	3.20	1.08	0.364	0.328
MLP	52.9	351	2.74	1.10	-0.246	-0.248
Comb MLP	35.8	94.3	3.45	1.03	0.184	0.148
Fisher	49.2	298	2.76	1.07	0.00858	0.00194

Table 5.36: Overview over S_{cP} with different kinds of uncertainties included, in the 0 jet channel for $m_H = 150$ GeV

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	26.1	45.8	3.52	0.783	1.99	0.714
BDT	37.9	60.4	4.43	1.17	1.29	0.854
BRT	51.6	165	3.81	1.32	0.767	0.669
MLP	61.2	366	3.10	1.24	-0.477	-0.479
Comb MLP	32.2	49.3	4.15	1.00	1.99	0.898
Fisher	47.9	150	3.70	1.27	0.939	0.777

Table 5.37: Overview over S_{cP} with different kinds of uncertainties included, in the 0 jet channel for $m_H = 170$ GeV

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	9.33	13.7	2.22	0.303	1.34	-0.126
BDT	22.5	95.1	2.19	0.644	0.866	0.540
BRT	28.4	110	2.57	0.789	0.668	0.515
MLP	38.5	328	2.07	0.831	-0.515	-0.514
Comb MLP	7.66	12.0	1.94	0.250	1.04	-0.251
Fisher	33.8	176	2.45	0.878	0.370	0.334

Table 5.38: Overview over S_{cP} with different kinds of uncertainties included, in the 0 jet channel for $m_H = 190$ GeV

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	9.16	62.3	1.09	0.275	0.749	0.250
BDT	8.06	1.33	4.41	0.260	0.143	-0.707
BRT	18.3	68.2	2.09	0.541	0.362	0.238
MLP	23.1	235	1.46	0.555	0.182	0.167
Comb MLP	7.45	19.2	1.54	0.241	0.481	-0.141
Fisher	11.7	410	0.558	0.242	0.139	0.125

Table 5.39: Overview over S_{cP} with different kinds of uncertainties included, in the 2 jet channel for $m_H = 150$ GeV

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	9.94	23.6	1.86	0.315	0.898	0.0504
BDT	16.0	10.5	4.06	0.519	0.353	-0.102
BRT	20.2	33.8	3.15	0.631	0.955	0.393
MLP	33.9	280	1.97	0.758	0.177	0.164
Comb MLP	7.33	16.0	1.64	0.233	0.505	-0.227
Fisher	27.0	410	1.30	0.559	0.370	0.319

Table 5.40: Overview over S_{cP} with different kinds of uncertainties included, in the 2 jet channel for $m_H = 170$ GeV

method	$\hat{S}$ [fb]	$\hat{B}$ [fb]	naive S_{cP}	stat S_{cP}	syst S_{cP}	total S_{cP}
Cuts	8.01	61.7	0.957	0.239	0.567	0.207
BDT	8.80	36.7	1.35	0.269	-0.131	-0.235
BRT	13.4	79.1	1.43	0.389	0.525	0.307
MLP	27.4	480	1.22	0.536	0.115	0.111
Comb MLP	8.27	68.6	0.940	0.242	0.372	0.176
Fisher	23.9	554	0.994	0.447	0.222	0.204

Table 5.41: Overview over S_{cP} with different kinds of uncertainties included, in the 2 jet channel for $m_H = 190$ GeV

6. Summary and Outlook

In table 6.1 the major results of this thesis are summarized in a ranking of classifiers. From the results presented and the ranking, the following conclusions can be made:

- Any discovery via event counting needs strong constraints on all systematic uncertainties. Methods for using real data to constrain background contributions are therefore essential and a topic for further studies.

- Multivariate classifiers can improve significance (without considering uncertainties) especially in situations where nonlinearities can be exploited, which here is the case for the 2 jet channel. Boosted and randomized decision trees are particularly efficient.

- Some multivariate classifiers select a greater signal cross-section at the optimal working point, especially neural networks and to a lesser extent bagged randomized trees. This leads to reduced uncertainty from Monte Carlo statistics.

- Multivariate classifiers increase the uncertainty resulting from systematic uncertainties. This negates their initial advantage if the systematic uncertainties can not be otherwise constrained. Neural networks are particularly susceptible: event types that did not occur in the training sample could be classified as signal-like resulting from the extrapolating nature of the neural network.

- In total, the bagged randomized trees classifier is the most robust and best-performing classifer, not always outperforming the classical cut analysis. It increases the selected signal cross-section on average by a factor of two, and therefore is more resistant to statistical uncertainties.

The different classifiers have shown very different strengths and weaknesses during the course of this analysis:

- The traditional cut analysis has shown that it is still a viable and robust method compared to multivariate classifiers, especially if the systematic

method	performance		performance considering uncertainties					
			statistical		systematical		all uncertainties	
	0 jet	2 jet	0 jet	2 jet	0 jet	2 jet	0 jet	2 jet
Cuts	o	−	−−	−	+	++	o	o
BDT	+	++	o	o	+	−	+	−−
BRT	+	+	+	+	o	+	+	+
MLP	−−	o	+	++	−−	−	−−	o
Comb. MLP	+	−	−	−	o	o	o	−
Fisher	o	−	+	o	o	−	o	+

Table 6.1: Ranking of the examined classifiers by performance considering different uncertainties. The rank is given relative to the mean m and root mean square σ of the mean over the Higgs boson masses of the S_{cP} significances. A result more than 1.5σ from the mean is given a ++ or −− respectively, a result more than 0.5σ from the mean a + or − respectively.

uncertainty is high. Usually the optimization results in only a small fraction of the signal events being selected.

- Boosted decision trees obtain the best performance if uncertainties are not considered, but their performance in difficult situations decreases rapidly in the presence of systematic uncertainties. The fraction of signal events selected is slightly better than in the cut analysis, but often comparable.

- Bagged randomized trees are not as performant as boosted decision trees, but much more robust. The examined classifiers had on average double the selected signal cross-section than the simple cut analysis.

- Neural networks need high training statistics. The neural network classifier consistently selected two to three times the signal cross-section compared to the cut analysis, and especially in the nonlinear 2 jet channel outperformed it as well. The neural networks used were extremely sensitive to systematic uncertainties - additional study to reduce this effect is necessary.

- The combination of multiple neural networks improved upon the neural network performance in the 0 jet channel, the resulting combined classifier was one of the best. However, in the 2 jet channel the performance was not nearly as good as the single neural network.

- In situations where it can be applied, the Fisher classifier performs well. It can be trained very easily, and selects almost as much signal cross-section as neural networks. To improve its performance also squares

or other nonlinear forms of variables could be used as inputs, this, however, would have to be separately investigated.

It is of course possible to improve the relative performance of multivariate classifiers by providing more variables. However, since this thesis was intended to be a first comparison, only the variables used in the cut analysis were made available to the classifiers.

A. Datasets and Methods

A.1 Datasets

Table 2.1 lists the datasets used for this study. When available, the official ATLAS dataset number is given. The low-mass Drell-Yan sample has been generated using a modified high-mass Pythia configuration file, setting the invariant mass to a range of 15 GeV $< m_{\mu\mu} <$ 60 GeV. The Higgs signal samples ahve been generated using modified 5320 sample configuration files. All samples have been simulated and reconstructed using the Athena software version 13.0.30. Except for the Drell-Yan and Higgs samples, all generated events were from official production.

The fast simulation was performed using ATLFAST II [34], using Athena version 13.0.40. These samples have been compared to fully simulated samples, and up to statistical accuracy no significant difference was observed in the relevant variables.

A.2 Analysis Methods

In this section, the software and methods used for the analysis are described.

A.2.1 Deriving Physics Data from AOD

To obtain plain ROOT NTuples from Analysis Object Data the HighPtView tool from the Athena 13.0.40 release has been used. The default settings were changed to not require a minimum muon transverse momentum p_T to include muons with $p_T < 20$ GeV. Apart from this, the default settings were not modified.

A.2.2 Preprocessing

From these NTuples an additional physics ntuple has been constructed by only including events with at least two reconstructed muons, removing unused variables and adding additional variables as described in section 5.2. These ntuples are the basis for further MVA and cut analysis.

A.2.3 Analysis Framework

To quickly set up, extend and execute an analysis a set of ROOT classes has been written: the first part (selectors) makes it possible to easily specify an analysis including cuts, multivariate classifiers, splits into training and test data and systematic shifts. An analysis so defined can be applied to the each NTuple individually, making easy local and distributed parallelization possible. The output of this analysis is then summarized, inspected and visualized by another set of ROOT classes (analysis).

Selectors

A helper class "MySelector" has been derived from a modified TSelector created by the ROOT MakeClass feature. This helper class can process each event several times in a "smearing loop" where events can be modified according to models of systematic shifts. Histograms for all cuts and smearing loops can easily be created. Cuts are not defined in this class, they can be defined in derived classes which can be very small. Also, the automated filling of optimization histograms and the creation of a tree of boolean values that holds the results of the individual cuts is supported.

From this class a "SmearSelector" has been derived that defines the number of smearing loops, which modifies each event for each loop, and stores the corresponding nuisance parameter values for each set of loops in the output ROOT file. For the initial optimization of cuts smearing can be easily deactivated in the derived classes as needed.

The preselection classes for both channels are then derived from "SmearSelector". The preselection and channel cuts are defined there, together with many common histograms. These are then automatically filled for each cut defined, even for cuts defined only in further derived classes.

Therefore, the HiggsCuts class is rather short: it only has to define the parametrized cuts. In the specific HiggsCuts170 class that defines the corresponding analysis for the Higgs boson mass of 170 GeV only these parameters have to be set, so it only consists of 15 lines, 10 of which are parameter settings.

This construction allows the user to quickly define and change a comprehensive analysis. The output ROOT files of these classes can now be used as input for the analysis classes.

Analysis

The analysis classes have the following structure:

- Sample: this class encapsulates one sample that can be spread over multiple files, but shares one physical cross-section. All classes that

build on this class do not have to interact with the ROOT files on disk anymore.

- Process: this class represents a physical process as a collection of several samples. The Drell-Yan decays, for example, are combined from the sample for $m_{\mu\mu} > 60$ GeV and $m_{\mu\mu} < 60$ GeV. This class can be used to get combined histograms and cross-sections of the samples, error propagation is automatically handled. Processes can be signal, background or data processes.

- Channel: here the processes are combined to one channel, and many presentation and analysis functions are defined. It can also create necessary processes from specifications in a configuration file.

Additionally, a graphical user interface to the analysis class has been implemented. There many different types of histograms can be plotted with different processes after different cuts. All histograms in this thesis were created by this tool with only slight modifications.

A.3 Neural Net Architecture Optimization

To visualize the dependence of the neural net performance on the architecture, a series of networks with different architectures were trained against the W^+W^- background in the 0 jet channel, where good training statistics are available. In figure A.1 one can see the results for a one-layer neural network, scaled to the performance of a boosted decision tree trained with the same data. In figure A.2 several two-layer configurations starting with 1 neuron in the first and 2 neurons in the second layer to 20 neurons in the first and 20 neurons in the second layer are tested. As one can see, the dependency of performance versus number of neurons is not necessarily increasing, and the optimal point is difficult to determine.

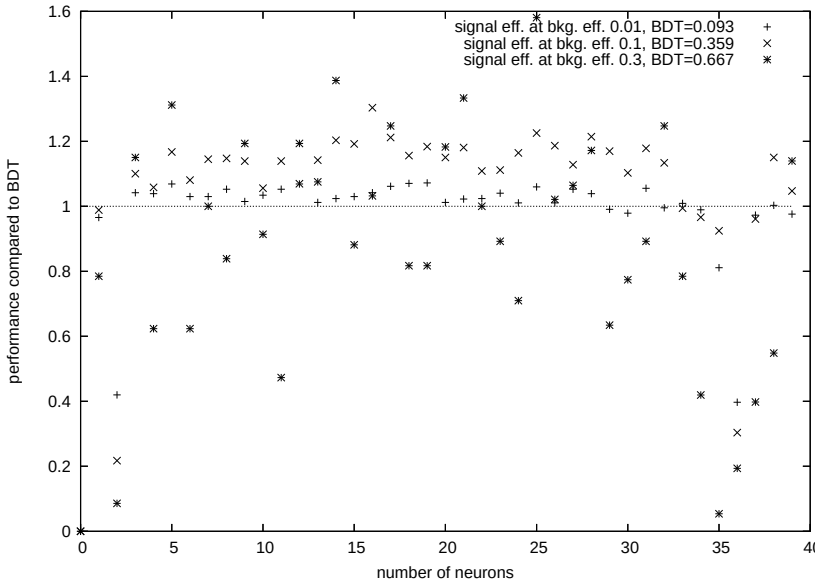


Figure A.1: Plot of the quality of a one-layer neural network trained versus QCD W^+W^- background versus number of nodes. As quality the signal efficiencies at background efficiencies of 1%, 10% and 30% are given, normalized to the corresponding efficiency for a boosted decision tree trained on the same data: $\frac{\text{signal eff. of MLP}}{\text{signal eff. of BDT}}$. The plus signs can be understood as the "tail performance" for high background rejection, the cross for intermediate and the star for high signal efficiency. Since the shape of the signal region can be more complex in the last case, more variation in the performance depending on the number of neurons is observed.

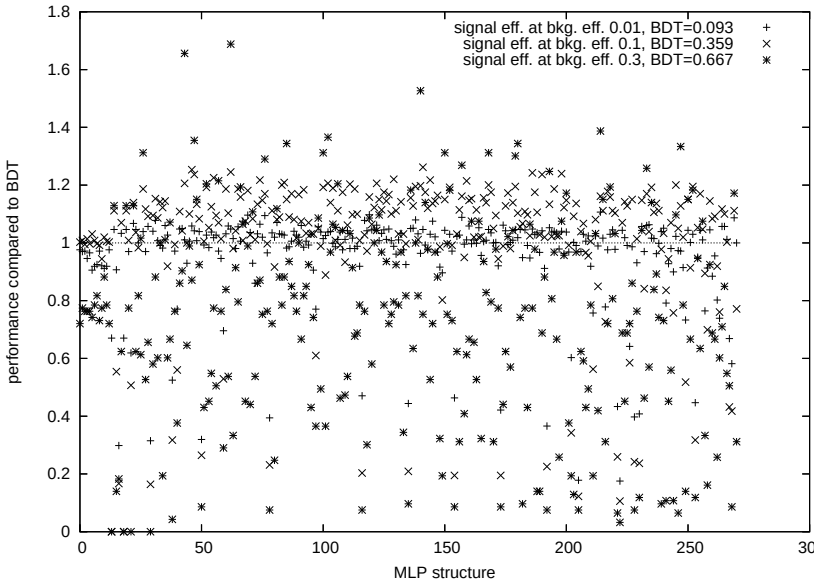


Figure A.2: Plot of the quality of a two-layer neural network trained versus QCD W^+W^- background, scanned over many different numbers of nodes in the two layers. The symbols are the same as in figure A.1.

B. Detailed Cut Flow Tables

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
0j	$(6.59 \pm 0.01) \cdot 10^5$	$(3.88 \pm 0.04) \cdot 10^3$	402 ± 21	549 ± 8.6	230 ± 27	148 ± 3	130 ± 28	$(6.64 \pm 0.01) \cdot 10^5$	94.5 ± 2.4
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	$(3.75 \pm 0.04) \cdot 10^3$	173 ± 14	517 ± 8.3	134 ± 20	113 ± 2.6	127 ± 28	$(6.45 \pm 0.01) \cdot 10^5$	86.2 ± 2.3
$m_{\mu\mu}$ range	$(5.25 \pm 0.06) \cdot 10^4$	$(2.29 \pm 0.05) \cdot 10^3$	42.4 ± 9.7	121 ± 5.7	6.23 ± 6.2	7.26 ± 0.92	50 ± 30	$(5.50 \pm 0.06) \cdot 10^4$	52.4 ± 2.5
$p_{T,trailing\mu}$	$(3.18 \pm 0.04) \cdot 10^4$	$(1.24 \pm 0.04) \cdot 10^3$	29 ± 8.0	93.9 ± 5.0	6.23 ± 6.2	4.45 ± 0.72	16.8 ± 27	$(3.32 \pm 0.04) \cdot 10^4$	44.6 ± 2.3
max $\Delta R_{\mu\mu}$	$(6.42 \pm 0.19) \cdot 10^3$	2.06 ± 1.5	15.6 ± 5.9	64.4 ± 4.2	0 ± 6.2	3.04 ± 0.60	16.8 ± 27	$(6.52 \pm 0.20) \cdot 10^3$	39.3 ± 2.2
min $\not{E}_T$	5.87 ± 7.3	1.03 ± 1.0	8.92 ± 4.5	43.1 ± 3.4	0 ± 6.2	1.52 ± 0.42	0 ± 27	60.4 ± 29	30.3 ± 1.9
$m_{T,\mu\mu}$ range	5.87 ± 7.3	0 ± 1.0	2.23 ± 2.2	25.6 ± 2.7	0 ± 6.2	0.117 ± 0.12	0 ± 27	33.8 ± 29	23.1 ± 1.7

Table B.1: Cut flow table for $m_H = 150$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	$(3.70 \pm 0.04) \cdot 10^3$	66.9 ± 8.6	253 ± 5.8	9.34 ± 5.4	11.6 ± 0.82	92.5 ± 25	$(1.06 \pm 0.01) \cdot 10^5$	75.4 ± 2.1
BDT	5.87 ± 7.3	0 ± 1.0	4.46 ± 3.1	27.6 ± 2.8	0 ± 6.2	0.117 ± 0.12	0 ± 27	38.1 ± 29	20.6 ± 1.6
BRT	58.7 ± 19	0 ± 1.0	15.6 ± 5.9	82.4 ± 4.7	0 ± 6.2	0.937 ± 0.33	13.7 ± 27	171 ± 35	43.8 ± 2.3
MLP	133 ± 27	14.4 ± 3.9	49.1 ± 10	134 ± 6.0	0 ± 6.2	3.51 ± 0.64	16.8 ± 27	351 ± 41	52.9 ± 2.5
Fisher	92.2 ± 22	4.12 ± 2.1	42.4 ± 9.7	130 ± 5.9	0 ± 6.2	2.69 ± 0.56	27.3 ± 29	298 ± 39	49.2 ± 2.5

Table B.2: Cut flow table for $m_H = 150$ GeV multivariate classifiers, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	$(3.55 \pm 0.06) \cdot 10^3$	60.2 ± 12	213 ± 7.5	6.23 ± 6.2	9.60 ± 1.1	101 ± 38	$(8.30 \pm 0.07) \cdot 10^4$	68.1 ± 2.9
Drell-Yan Cut	86.6 ± 22	10.3 ± 3.3	37.9 ± 9.2	118 ± 5.7	0 ± 6.2	3.16 ± 0.61	16.8 ± 27	273 ± 38	52.3 ± 2.5
QCD Cut	74.8 ± 21	0 ± 1.0	24.5 ± 7.4	115 ± 5.6	0 ± 6.2	3.16 ± 0.61	16.8 ± 27	234 ± 36	50.2 ± 2.5
W^+W^- Cut	23.5 ± 12	0 ± 1.0	8.92 ± 4.5	57.8 ± 4	0 ± 6.2	0.937 ± 0.33	3.15 ± 27	94.3 ± 31	35.8 ± 2.1

Table B.3: Cut flow table for $m_H = 150$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
0j	$(6.59 \pm 0.01) \cdot 10^5$	$(3.88 \pm 0.04) \cdot 10^3$	402 ± 21	549 ± 8.6	230 ± 27	148 ± 3	130 ± 28	$(6.64 \pm 0.01) \cdot 10^5$	94.3 ± 2.1
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	$(3.75 \pm 0.04) \cdot 10^3$	173 ± 14	517 ± 8.3	134 ± 20	113 ± 2.6	127 ± 28	$(6.45 \pm 0.01) \cdot 10^5$	85.5 ± 2.0
$m_{\mu\mu}$ range	$(8.44 \pm 0.07) \cdot 10^4$	$(3.64 \pm 0.06) \cdot 10^3$	62.4 ± 12	232 ± 7.9	6.23 ± 6.2	9.72 ± 1.1	101 ± 38	$(8.84 \pm 0.07) \cdot 10^4$	71.8 ± 2.6
$p_{T,trailing\mu}$	$(9.74 \pm 0.21) \cdot 10^3$	235 ± 16	13.4 ± 5.5	68.1 ± 4.3	0 ± 6.2	0.937 ± 0.33	10.5 ± 27	$(1.01 \pm 0.02) \cdot 10^4$	35.3 ± 1.9
max $\Delta R_{\mu\mu}$	309 ± 42	3.09 ± 1.8	4.46 ± 3.1	36.5 ± 3.2	0 ± 6.2	0.703 ± 0.29	10.5 ± 27	364 ± 50	31.2 ± 1.7
min $\not{E}_T$	5.87 ± 7.3	0 ± 1.0	4.46 ± 3.1	30.5 ± 2.9	0 ± 6.2	0.585 ± 0.26	10.5 ± 27	52 ± 29	26.8 ± 1.6
$m_{T,\mu\mu}$ range	0 ± 7.3	0 ± 1.0	6.69 ± 3.9	28.3 ± 2.8	0 ± 6.2	0.351 ± 0.20	10.5 ± 27	45.8 ± 29	26.1 ± 1.6

Table B.4: Cut flow table for $m_H = 170$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	$(3.70 \pm 0.04) \cdot 10^3$	66.9 ± 8.6	253 ± 5.8	9.34 ± 5.4	11.6 ± 0.82	92.5 ± 25	$(1.06 \pm 0.01) \cdot 10^5$	74.7 ± 1.9
BDT	5.87 ± 7.3	0± 1.0	6.69 ± 3.9	46.8 ± 3.6	0± 6.2	1.05 ± 0.35	0± 27	60.4 ± 29	37.9 ± 1.9
BRT	41.1 ± 16	0± 1.0	24.5 ± 7.4	87.0 ± 4.9	0± 6.2	1.64 ± 0.44	10.5 ± 27	165 ± 33	51.6 ± 2.2
MLP	150 ± 27	2.06 ± 1.5	26.8 ± 7.7	150 ± 6.4	6.23 ± 6.2	3.28 ± 0.62	27.3 ± 29	366 ± 42	61.2 ± 2.4
Fisher	22 ± 11	1.03 ± 1.0	31.2 ± 8.3	83.3 ± 4.8	0± 6.2	1.64 ± 0.44	10.5 ± 27	150 ± 31	47.9 ± 2.1

Table B.5: Cut flow table for $m_H = 170$ GeV multivariate classifiers, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	$(3.55 \pm 0.06) \cdot 10^3$	60.2 ± 12	213 ± 7.5	6.23 ± 6.2	9.60 ± 1.1	101 ± 38	$(8.30 \pm 0.07) \cdot 10^4$	70.3 ± 2.6
Drell-Yan Cut	216 ± 34	12.3 ± 3.6	42.4 ± 9.7	152 ± 6.4	6.23 ± 6.2	3.86 ± 0.67	27.3 ± 29	461 ± 47	62.5 ± 2.5
QCD Cut	73.0 ± 20	0± 1.0	15.6 ± 5.9	72.7 ± 4.5	0± 6.2	0.937 ± 0.33	10.5 ± 27	173 ± 35	42.6 ± 2.0
W^+W^- Cut	5.87 ± 7.3	0± 1.0	6.69 ± 3.9	35.9 ± 3.2	0± 6.2	0.820 ± 0.31	0± 27	49.3 ± 29	32.2 ± 1.8

Table B.6: Cut flow table for $m_H = 170$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
0j	$(6.59 \pm 0.01) \cdot 10^5$	$(3.88 \pm 0.04) \cdot 10^3$	402 ± 21	549 ± 8.6	230 ± 27	148 ± 3	130 ± 28	$(6.64 \pm 0.01) \cdot 10^5$	66.8 ± 1.7
no hard e	$(6.40 \pm 0.01) \cdot 10^5$	$(3.75 \pm 0.04) \cdot 10^3$	173 ± 14	517 ± 8.3	134 ± 20	113 ± 2.6	127 ± 28	$(6.45 \pm 0.01) \cdot 10^5$	60.5 ± 1.6
$m_{\mu\mu}$ range	$(7.76 \pm 0.07) \cdot 10^4$	$(3.52 \pm 0.06) \cdot 10^3$	60.2 ± 12	206 ± 7.4	6.23 ± 6.2	9.48 ± 1.1	101 ± 38	$(8.15 \pm 0.07) \cdot 10^4$	38.2 ± 1.8
$p_{T,trailing\mu}$	$(7.76 \pm 0.07) \cdot 10^4$	$(3.52 \pm 0.06) \cdot 10^3$	60.2 ± 12	206 ± 7.4	6.23 ± 6.2	9.48 ± 1.1	101 ± 38	$(8.15 \pm 0.07) \cdot 10^4$	38.2 ± 1.8
max $\Delta R_{\mu\mu}$	$(1.00 \pm 0.02) \cdot 10^4$	6.18 ± 2.5	24.5 ± 7.4	91.7 ± 5	0± 6.2	3.75 ± 0.66	27.3 ± 29	$(1.02 \pm 0.02) \cdot 10^4$	29.2 ± 1.6
min $\not{E}_T$	29.4 ± 14	4.12 ± 2.1	17.8 ± 6.3	74.7 ± 4.5	0± 6.2	2.11 ± 0.50	10.5 ± 27	139 ± 32	26.7 ± 1.5
$m_{T,\mu\mu}$ range	0± 7.3	0± 1.0	4.46 ± 3.1	8.21 ± 1.6	0± 6.2	1.05 ± 0.35	0± 27	13.7 ± 29	9.33 ± 0.91

Table B.7: Cut flow table for $m_H = 190$ GeV cut analysis, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.02 \pm 0.01) \cdot 10^5$	$(3.70 \pm 0.04) \cdot 10^3$	66.9 ± 8.6	253 ± 5.8	9.34 ± 5.4	11.6 ± 0.82	92.5 ± 25	$(1.06 \pm 0.01) \cdot 10^5$	45.1 ± 1.4
BDT	26.3 ± 12	1.03 ± 1.0	22.3 ± 7.0	44 ± 3.5	0± 6.2	1.52 ± 0.42	0± 27	95.1 ± 31	22.5 ± 1.4
BRT	17.6 ± 11	0± 1.0	11.2 ± 5	69.2 ± 4.3	0± 6.2	1.64 ± 0.44	10.5 ± 27	110 ± 31	28.4 ± 1.6
MLP	118 ± 25	0± 1.0	40.1 ± 9.5	150 ± 6.4	0± 6.2	3.04 ± 0.60	16.8 ± 27	328 ± 39	38.5 ± 1.8
Fisher	26.3 ± 12	1.03 ± 1.0	35.7 ± 8.9	100 ± 5.2	0± 6.2	1.87 ± 0.47	10.5 ± 27	176 ± 32	33.8 ± 1.7

Table B.8: Cut flow table for $m_H = 190$ GeV multivariate classifiers, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(7.90 \pm 0.07) \cdot 10^4$	$(3.55 \pm 0.06) \cdot 10^3$	60.2 ± 12	213 ± 7.5	6.23 ± 6.2	9.60 ± 1.1	101 ± 38	$(8.30 \pm 0.07) \cdot 10^4$	38.9 ± 1.9
Drell-Yan Cut	178 ± 31	16.5 ± 4.1	42.4 ± 9.7	146 ± 6.3	0 ± 6.2	3.86 ± 0.67	27.3 ± 29	415 ± 45	34.3 ± 1.7
QCD Cut	98.1 ± 23	5.15 ± 2.3	17.8 ± 6.3	95.7 ± 5.1	0 ± 6.2	1.76 ± 0.45	10.5 ± 27	229 ± 37	27.2 ± 1.6
W^+W^- Cut	0 ± 7.3	0 ± 1.0	2.23 ± 2.2	9.52 ± 1.7	0 ± 6.2	0.234 ± 0.17	0 ± 27	12 ± 29	7.66 ± 0.82

Table B.9: Cut flow table for $m_H = 190$ GeV combined neural networks, 0 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
2j	$(1.41 \pm 0.01) \cdot 10^5$	952 ± 22	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	579 ± 42	276 ± 4.0	122 ± 29	$(1.48 \pm 0.01) \cdot 10^5$	84 ± 2.3
$m_{\mu\mu}$ range	$(1.46 \pm 0.03) \cdot 10^4$	942 ± 31	$(1.48 \pm 0.06) \cdot 10^3$	72 ± 4.4	56.1 ± 19	17.9 ± 1.4	82.2 ± 39	$(1.72 \pm 0.03) \cdot 10^4$	65.2 ± 2.8
$p_{T,trailing\mu}$	$(5.03 \pm 0.17) \cdot 10^3$	267 ± 17	685 ± 39	34.1 ± 3.1	24.9 ± 12	6.32 ± 0.86	16.3 ± 27	$(6.06 \pm 0.17) \cdot 10^3$	43.1 ± 2.3
max $\Delta R_{\mu\mu}$	$(4.06 \pm 0.15) \cdot 10^3$	122 ± 11	584 ± 36	27.4 ± 2.8	18.7 ± 11	4.92 ± 0.76	16.3 ± 27	$(4.84 \pm 0.16) \cdot 10^3$	39.7 ± 2.2
$\sum_{jet} p_T$	$(3.90 \pm 0.15) \cdot 10^3$	106 ± 10	504 ± 34	26.3 ± 2.7	18.7 ± 11	4.68 ± 0.74	16.3 ± 27	$(4.57 \pm 0.16) \cdot 10^3$	33.8 ± 2.0
cjv	$(1.37 \pm 0.09) \cdot 10^3$	38.1 ± 6.3	167 ± 19	10.0 ± 1.8	12.5 ± 8.8	1.76 ± 0.45	0 ± 27	$(1.60 \pm 0.09) \cdot 10^3$	21.7 ± 1.6
min $\not{E}_T$	11.8 ± 9.4	13.4 ± 3.7	114 ± 16	8.42 ± 1.6	6.23 ± 6.2	0.234 ± 0.17	0 ± 27	154 ± 34	13.9 ± 1.3
$m_{T,\mu\mu}$ range	0 ± 7.3	0 ± 1.0	58 ± 11	4.11 ± 1.2	0 ± 6.2	0.234 ± 0.17	0 ± 27	62.3 ± 31	9.16 ± 1.1

Table B.10: Cut flow table for $m_H = 150$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.88 \pm 0.02) \cdot 10^4$	883 ± 21	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	49.8 ± 12	23.8 ± 1.2	81.3 ± 24	$(2.14 \pm 0.02) \cdot 10^4$	$63.3 \pm 2.$
BDT	0 ± 7.3	0 ± 1.0	0 ± 2.2	1.10 ± 0.83	0 ± 6.2	0.234 ± 0.17	0 ± 27	1.33 ± 29	8.06 ± 0.99
BRT	5.87 ± 7.3	0 ± 1.0	53.5 ± 11	8.61 ± 1.6	0 ± 6.2	0.234 ± 0.17	0 ± 27	68.2 ± 31	18.3 ± 1.5
MLP	26.3 ± 12	28.8 ± 5.5	154 ± 18	15.9 ± 2.1	6.23 ± 6.2	0.703 ± 0.29	3.15 ± 27	235 ± 36	23.1 ± 1.7
Fisher	27.8 ± 12	2.06 ± 1.5	355 ± 28	19.0 ± 2.3	6.23 ± 6.2	0.703 ± 0.29	0 ± 27	410 ± 42	11.7 ± 1.2

Table B.11: Cut flow table for $m_H = 150$ GeV multivariate classifiers, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.51 \pm 0.03) \cdot 10^4$	956 ± 31	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	56.1 ± 19	18.9 ± 1.5	82.2 ± 39	$(1.78 \pm 0.03) \cdot 10^4$	66.7 ± 2.9
Drell-Yan Cut	22 ± 11	25.7 ± 5.2	430 ± 31	25.8 ± 2.7	6.23 ± 6.2	0.820 ± 0.31	13.7 ± 27	525 ± 43	26.4 ± 1.8
QCD Cut	0 ± 7.3	1.03 ± 1.0	20.1 ± 6.7	3.69 ± 1.1	0 ± 6.2	0.351 ± 0.20	0 ± 27	25.1 ± 30	7.94 ± 0.98
W^+W^- Cut	0 ± 7.3	1.03 ± 1.0	15.6 ± 5.9	2.19 ± 0.95	0 ± 6.2	0.351 ± 0.20	0 ± 27	19.2 ± 29	7.45 ± 0.95

Table B.12: Cut flow table for $m_H = 150$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
2j	$(1.41 \pm 0.01) \cdot 10^5$	952 ± 22	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	579 ± 42	276 ± 4.0	122 ± 29	$(1.48 \pm 0.01) \cdot 10^5$	96.7 ± 2.2
$m_{\mu\mu}$ range	$(1.21 \pm 0.03) \cdot 10^4$	785 ± 28	$(1.06 \pm 0.05) \cdot 10^3$	53.2 ± 3.8	49.8 ± 18	14.9 ± 1.3	65.8 ± 36	$(1.41 \pm 0.03) \cdot 10^4$	66.5 ± 2.5
$p_{T,trailing\mu}$	846 ± 70	14.4 ± 3.9	84.7 ± 14	4.11 ± 1.2	0 ± 6.2	1.76 ± 0.45	0 ± 27	951 ± 77	24.2 ± 1.5
max $\Delta R_{\mu\mu}$	846 ± 70	14.4 ± 3.9	84.7 ± 14	4.11 ± 1.2	0 ± 6.2	1.76 ± 0.45	0 ± 27	951 ± 77	24.2 ± 1.5
$\sum_{jet} p_T$	716 ± 65	7.21 ± 2.7	80.3 ± 13	3.83 ± 1.2	0 ± 6.2	1.52 ± 0.42	0 ± 27	809 ± 72	20.6 ± 1.4
cjv	423 ± 50	4.12 ± 2.1	51.3 ± 11	2.47 ± 1.0	0 ± 6.2	0.820 ± 0.31	0 ± 27	481 ± 58	16.6 ± 1.3
min $\not{E}_T$	5.87 ± 7.3	4.12 ± 2.1	31.2 ± 8.3	1.64 ± 0.91	0 ± 6.2	0.234 ± 0.17	0 ± 27	43.1 ± 30	11.6 ± 1.1
$m_{T,\mu\mu}$ range	0 ± 7.3	0 ± 1.0	22.3 ± 7.0	1.10 ± 0.83	0 ± 6.2	0.234 ± 0.17	0 ± 27	23.6 ± 30	9.94 ± 0.98

Table B.13: Cut flow table for $m_H = 170$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.88 \pm 0.02) \cdot 10^4$	883 ± 21	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	49.8 ± 12	23.8 ± 1.2	81.3 ± 24	$(2.14 \pm 0.02) \cdot 10^4$	73.0 ± 1.9
BDT	0 ± 7.3	0 ± 1.0	8.92 ± 4.5	1.37 ± 0.87	0 ± 6.2	0.234 ± 0.17	0 ± 27	10.5 ± 29	16.0 ± 1.2
BRT	0 ± 7.3	0 ± 1.0	29 ± 8.0	4.38 ± 1.3	0 ± 6.2	0.468 ± 0.23	0 ± 27	33.8 ± 30	20.2 ± 1.4
MLP	82.8 ± 20	10.3 ± 3.3	172 ± 20	15.3 ± 2.1	0 ± 6.2	0.468 ± 0.23	0 ± 27	280 ± 39	33.9 ± 1.8
Fisher	27.8 ± 12	2.06 ± 1.5	355 ± 28	18.5 ± 2.3	6.23 ± 6.2	0.703 ± 0.29	0 ± 27	410 ± 42	27.0 ± 1.6

Table B.14: Cut flow table for $m_H = 170$ GeV multivariate classifiers, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.51 \pm 0.03) \cdot 10^4$	956 ± 31	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	56.1 ± 19	18.9 ± 1.5	82.2 ± 39	$(1.78 \pm 0.03) \cdot 10^4$	79.7 ± 2.8
Drell-Yan Cut	20.4 ± 10	7.21 ± 2.7	321 ± 27	18.3 ± 2.3	0 ± 6.2	0.703 ± 0.29	0 ± 27	368 ± 40	26.2 ± 1.6
QCD Cut	5.87 ± 7.3	0 ± 1.0	8.92 ± 4.5	3.01 ± 1.1	0 ± 6.2	0.117 ± 0.12	0 ± 27	17.9 ± 29	8.30 ± 0.90
W^+W^- Cut	5.87 ± 7.3	0 ± 1.0	8.92 ± 4.5	1.10 ± 0.83	0 ± 6.2	0.117 ± 0.12	0 ± 27	16.0 ± 29	7.33 ± 0.84

Table B.15: Cut flow table for $m_H = 170$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
2j	$(1.41 \pm 0.01) \cdot 10^5$	952 ± 22	$(4.55 \pm 0.07) \cdot 10^3$	204 ± 5.2	579 ± 42	276 ± 4.0	122 ± 29	$(1.48 \pm 0.01) \cdot 10^5$	69.8 ± 1.8
$m_{\mu\mu}$ range	$(1.29 \pm 0.03) \cdot 10^4$	952 ± 31	$(1.46 \pm 0.06) \cdot 10^3$	74.1 ± 4.5	49.8 ± 18	18.3 ± 1.5	82.2 ± 39	$(1.56 \pm 0.03) \cdot 10^4$	42.4 ± 1.9
$p_{T,trailing\mu}$	$(4.74 \pm 0.16) \cdot 10^3$	262 ± 16	676 ± 39	35 ± 3.1	12.5 ± 8.8	6.79 ± 0.89	16.3 ± 27	$(5.75 \pm 0.17) \cdot 10^3$	31.3 ± 1.7
max $\Delta R_{\mu\mu}$	$(4.74 \pm 0.16) \cdot 10^3$	262 ± 16	676 ± 39	35 ± 3.1	12.5 ± 8.8	6.79 ± 0.89	16.3 ± 27	$(5.75 \pm 0.17) \cdot 10^3$	31.3 ± 1.7
$\sum_{jet} p_T$	$(4.24 \pm 0.15) \cdot 10^3$	224 ± 15	406 ± 30	30.4 ± 2.9	12.5 ± 8.8	5.85 ± 0.83	16.3 ± 27	$(4.93 \pm 0.16) \cdot 10^3$	23.1 ± 1.4
cjv	$(1.71 \pm 0.10) \cdot 10^3$	93.7 ± 9.8	125 ± 17	13.1 ± 1.9	6.23 ± 6.2	2.58 ± 0.55	0 ± 27	$(1.95 \pm 0.10) \cdot 10^3$	13.8 ± 1.1
min $\not{E}_T$	16.1 ± 9.4	18.5 ± 4.4	80.3 ± 13	8.90 ± 1.6	6.23 ± 6.2	0.351 ± 0.20	0 ± 27	130 ± 33	10.7 ± 0.97
$m_{T,\mu\mu}$ range	5.87 ± 7.3	0 ± 1.0	49.1 ± 10	6.50 ± 1.4	0 ± 6.2	0.234 ± 0.17	0 ± 27	61.7 ± 31	8.01 ± 0.84

Table B.16: Cut flow table for $m_H = 190$ GeV cut analysis, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.88 \pm 0.02) \cdot 10^4$	883 ± 21	$(1.45 \pm 0.04) \cdot 10^3$	77.5 ± 3.2	49.8 ± 12	23.8 ± 1.2	81.3 ± 24	$(2.14 \pm 0.02) \cdot 10^4$	45.2 ± 1.4
BDT	11.8 ± 9.4	0 ± 1.0	20.1 ± 6.7	4.38 ± 1.3	0 ± 6.2	0.468 ± 0.23	0 ± 27	36.7 ± 30	8.80 ± 0.88
BRT	0 ± 7.3	0 ± 1.0	66.9 ± 12	5.48 ± 1.4	6.23 ± 6.2	0.468 ± 0.23	0 ± 27	79.1 ± 31	13.4 ± 1.1
MLP	65.4 ± 18	32.9 ± 5.8	330 ± 27	32.6 ± 3.0	18.7 ± 11	0.820 ± 0.31	0 ± 27	480 ± 44	27.4 ± 1.6
Fisher	55.4 ± 17	1.03 ± 1.0	466 ± 32	24.6 ± 2.6	6.23 ± 6.2	0.937 ± 0.33	0 ± 27	554 ± 46	23.9 ± 1.4

Table B.17: Cut flow table for $m_H = 190$ GeV multivariate classifiers, 2 jet channel. Only statistical errors are given.

$\sigma\times$ BR [fb]	$Z/\gamma^* \to \mu^+\mu^-$	$Z/\gamma^* \to \tau^+\tau^-$	$t\bar{t}$	W^+W^-	WZ	ZZ	W+QCD	All Backgrounds	Signal
Preselection	$(1.51 \pm 0.03) \cdot 10^4$	956 ± 31	$(1.53 \pm 0.06) \cdot 10^3$	76.0 ± 4.5	56.1 ± 19	18.9 ± 1.5	82.2 ± 39	$(1.78 \pm 0.03) \cdot 10^4$	46.6 ± 2.0
Drell-Yan Cut	74.6 ± 20	72.1 ± 8.6	$(1.03 \pm 0.05) \cdot 10^3$	48.4 ± 3.6	12.5 ± 8.8	1.76 ± 0.45	22.6 ± 27	$(1.26 \pm 0.06) \cdot 10^3$	36.1 ± 1.8
QCD Cut	5.87 ± 7.3	0 ± 1.0	51.3 ± 11	4.65 ± 1.3	6.23 ± 6.2	0.585 ± 0.26	0 ± 27	68.6 ± 31	8.27 ± 0.85
W^+W^- Cut	5.87 ± 7.3	0 ± 1.0	51.3 ± 11	4.65 ± 1.3	6.23 ± 6.2	0.585 ± 0.26	0 ± 27	68.6 ± 31	8.27 ± 0.85

Table B.18: Cut flow table for $m_H = 190$ GeV combined neural networks, 2 jet channel. Only statistical errors are given.

Bibliography

[1] M. Gellmann. A schematic model of baryons and mesons. *Physics Letters*, 8:214–215, February 1964.

[2] D. Griffiths. *Introduction to Elementary Particles.* Wiley-VCH, March 1987.

[3] Amsler, C. et al. (Particle Data Group). 2008 Review of Particle Physics. *Phys. Lett. B*, 667(1), 2008.

[4] Pushpalatha C. Bhat, Russell Gilmartin, and Harrison B. Prosper. Strategy for discovering a low-mass Higgs boson at the Fermilab Tevatron. *Phys. Rev. D*, 62(7), Sep 2000.

[5] Abe, F. et al. (CDF Collaboration). Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74(14):2626–2631, Apr 1995.

[6] Ahmad, Q. R. et al. (SNO Collaboration). Measurement of the Rate of $\nu e + d \rightarrow p + p + e-$ Interactions Produced by $B8$ Solar Neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87(7), Jul 2001.

[7] Bernardi, G. et al. (Tevatron New Phenomena Higgs Working Group). Combined CDF and Dzero Upper Limits on Standard Model Higgs Boson Production at High Mass (155-200 GeV/c^2) with 3 fb^{-1} of data. *arXiv*, 2008.

[8] M.E. Peskin and D.V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, Boulder, U.S.A., 1995.

[9] Steven Weinberg. The making of the standard model. *Eur. Phys. J.*, C34:5–13, 2004.

[10] R. P. Feynman. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Rev. Mod. Phys.*, 20(2):367–387, Apr 1948.

[11] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12:132–133, September 1964.

[12] Abdelhak Djouadi. The anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008.

[13] Gian F. Giudice and Oleg Lebedev. Higgs-dependent Yukawa couplings. *Phys. Lett.*, B665:79–85, 2008.

[14] M. Dittmar and Herbert K. Dreiner. How to find a Higgs boson with a mass between 155-GeV to 180-GeV at the LHC. *Phys. Rev.*, D55:167–172, 1997.

[15] M. A. Dobbs et al. Les Houches guidebook to Monte Carlo generators for hadron collider physics. In *Les Houches 2003, Physics at TeV colliders*, pages 411–459, 2004.

[16] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[17] Borut Paul Kersevan and Elzbieta Richter-Was. The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5, 2004.

[18] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.

[19] S. Frixione and B. R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 0206:029, 2002.

[20] S. Frixione, P. Nason, and B. R. Webber. Matching NLO QCD and parton showers in heavy flavour production. *JHEP*, 0308:007, 2003.

[21] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa. ALPGEN, A Generator for Hard Multiparton Processes in Hadronic Collisions. *JHEP*, 0307:001, 2003.

[22] A. Hocker et al. TMVA-Toolkit for Multivariate Data Analysis. *arXiv physics/0703039*, 2007.

[23] Abazov, V. M. et al. (D0 Collaboration). Evidence for Production of Single Top Quarks and First Direct Measurement of $|V_{tb}|$. *Physical Review Letters*, 98(18):181802, 2007.

[24] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR Upper Saddle River, NJ, USA, 1998.

[25] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

[26] European Organization for Nuclear Research (Geneva). CERN Home Page. www.cern.ch, October 2008.

[27] European Organization for Nuclear Research (Geneva). *ATLAS Detector and Physics Performance: Technical Design Report*. CERN, 1999.

[28] R.S. Moore. Performance and Future of the Tevatron. *Nuclear Physics B (Proceedings Supplements)*, 177:5–7, 2008.

[29] Karl Jakobs and Dave Charlton. CSC Notes. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CSCNotesList#Recommendations_for_the_Notes, October 2008.

[30] D. L. Adams, A. Ouraou, A. Poppleton, and S. Willocq. Muon Reconstruction and Identification Performance in ATLAS: Studies with Simulated Monte Carlo Samples. Draft version 3.02, April 2008.

[31] Benjamin Ruckert. Muon Reconstruction and the Search for Leptoquarks at LHC. Diploma thesis, Ludwig-Maximilians-Universität München. Master's thesis, July 2006.

[32] Wolfgang Mader and Jimmy Proudfoot. Jet, Missing Et and Tau Combined Performance. May 2008.

[33] S. Agostinelli et al. Geant4 - A Simulation Toolkit. *Nuclear Inst. and Methods in Physics Research, A*, 506(3):250–303, 2003.

[34] M Dührssen. The fast calorimeter simulation FastCaloSim. Technical Report ATL-PHYS-INT-2008-043, ATL-COM-PHYS-2008-093, CERN, Geneva, Jul 2008.

[35] B. Efron. Controversies in the foundations of statistics. *Amer. Math. Monthly*, 85:231–246, 1978.

[36] G. Zech. Frequentist and Bayesian confidence intervals. *EPJ direct*, 4(1):1–81, 2002.

[37] Gary J. Feldman and Robert D. Cousins. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57(7):3873–3889, Apr 1998.

[38] AL Read. Modified Frequentist Analysis of Search Results (The CLs Method) in proceedings of the 2000 CERN Workshop on Confidence Intervals. *CERN report*, 5(2000):81, 2000.

[39] J. Neyman. On the problem of confidence intervals. *Ann. Math. Stat*, 6(1):1, 1935.

[40] R. Brun, F. Rademakers, et al. ROOT-An Object Oriented Data Analysis Framework. *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Inst. & Meth. in Phys. Res. A*, 389:81–86, 1997.

[41] P Clifford. Interval Estimations as Viewed from the World of Mathematical Statistics. *CERN report*, 5(2000):157, 2000.

[42] Gabriele Reiter. Studien zur Higgs-Produktion im Kanal $H \rightarrow W^+W^- \rightarrow \mu^+\mu^-\nu_\mu\bar{\mu}_\nu$ mit dem ATLAS Detektor am LHC. Diploma thesis, Ludwig-Maximilians-Universität München, July 2007.

[43] J. T. Linnemann. A Pitfall in Evaluating Systematic Errors. In *PHYSTAT LHC Workshop on Statistical Issues for LHC Physics*, volume CERN-2008-001, 2008.

[44] S. I. Bityukov, S. E. Erofeeva, N. V. Krasnikov, and A. N. Nikitenko. Program for Evaluation of Significance, Confidence Intervals and Limits by Direct Calculation of Probabilities. In L. Lyons and M. Karagöz Ünel, editors, *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pages 106–+, 2006.

[45] Hai-Jun Yang, Tiesheng Dai, Dan Levin, Xuefei Li, Alan Wilson, Zhengguo Zhao, and Bing Zhou. ATLAS $H \rightarrow WW$ Detection Sensitivity with Boosted Decision Trees. *ATLAS Note COM-PHYS-2008-023*, 2008.

Acknowledgements

Zum Schluss möchte ich mich bei den Personen bedanken, die mich bei der Anfertigung der Diplomarbeit unterstützt haben. Herzlichen Dank!

- Prof. Dr. Dorothee Schaile für die Erstkorrektur, das Thema, die vielen Anregungen sowie die Möglichkeit auf Konferenzen und Treffen mit vielen Leuten über diese Themen zu sprechen
- Prof. Dr. Christian Kiesling für die Erstellung des Zweitgutachtens
- Dr. Johannes Elmsheuser für die vielen Ratschläge, Tips und allgemein die ausgezeichnete Betreuung
- Dr. John Kennedy für die Einführung in das Grid, und die Beantwortung vieler weiterer Fragen
- Dr. Günther Duckeck und Dr. Johannes Elmsheuser für das Korrekturlesen dieser Arbeit
- Dr. Gernot Krobath und Benjamin Ruckert für die gute Stimmung und die interessanten Diskussionen im Büro
- Frau Hertha Franz, die mir bei meinen administrativen Problemen immer freundlich mit Rat und Tat beistand
- allen Leuten aus dem Lehrstuhl für die interessanten Gespräche und die schöne Arbeitsatmosphäre
- auch besonders meiner Freundin Amelie, die immer für mich da ist und mich immer wieder aufbaut
- und schließlich meiner Familie, insbesondere meinen Eltern, die mich mein ganzes Leben lang unermüdlich unterstützt und gefördert haben.

Erklärung

Ich erkläre hiermit, dass ich meine Diplomarbeit mit dem Titel

The search for the Higgs Boson at the ATLAS Experiment using Multivariate Techniques

selbständig verfasst sowie keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 25. November 2008

_____

(Johannes Ebke)