Universität Bonn

Physikalisches Institut

Improvement of the Higgs mass resolution in the $ZH \to l^+l^-b\bar{b}$ channel using a Kinematic Fit with constraints

Holger Ritter

The measurement of the Higgs mass in the $H \to b\bar{b}$ decay suffers from low accuracy of b-jet momentum reconstruction. In order to improve the resolution in $m_{b\bar{b}}$ a least squares kinematic fit was applied in the $ZH \to l^+l^-b\bar{b}$ channel. In the fit the b-jet momentum vectors are changed such that they are closer to those of the true b-quarks. Momentum conservation in the transverse plane was used as a constraint. Initial state radiation carries away momentum of the ZH system. To account for the loss in transverse momentum an additional momentum vector is constructed. An improvement of 8 % was achieved in the Higgs mass resolution. With a cut on the chi-square probability of the kinematic fit the expected significance could be improved by 17 %. A cross check on the $ZZ \to l^+l^-b\bar{b}$ decay channel showed an improved of 23 % in the expected significance.

Physikalisches Institut der
Universität Bonn
Nussallee 12
D-53115 Bonn

BONN-IB-2014-14
December 2014

Acknowledgements

I would like to thank first and foremost Dr. Eckhard von Toerne, who supervised me over the last year, helped me with any problems and questions, came up with new ideas and suggestions, and proof read my thesis. I would like to thank Prof. Dr. Norbert Wermes for giving me the opportunity to write my master thesis in his research group. I would like to thank Elisabeth Schopf, Nicklas Denis and Alexander Ermakov for helping with various problems and making the time in the office a lot more enjoyable. I would like to thank Dr. Goetz Gaycken and Stephan Hageboeck for assisting me in technical and physical questions. Finally I want to thank everyone else of my colleagues, friends and family who helped me completing my thesis over the last year in one way or another.

Contents

CHAPTER 1

Introduction

Since the discovery of the tau neutrino in 2000 the Higgs boson remained the last particle in the Standard Model of particle physics yet unconfirmed. To find this last missing piece of one of the most important theories in modern physics was among the main motivations for the construction of the Large Hadron Collider. In summer of 2012 a new boson decaying into two photons or two Z bosons has been discovered by the ATLAS and CMS experiment and in March 2013 it was confirmed by CERN that this is indeed most likely the Higgs boson. Since then it has been observed to decay into WW and $\tau\tau$ but not into two b-quarks even though this decay channel has the highest branching ratio. The $H \to b\bar{b}$ channel has an overwhelming background from multijet production which makes triggering extremely difficult, therefore one considers the associated production with a vector boson W/Z which decays into ll,$l\nu$ or $\nu\nu$. These events can be triggered using the lepton(s), or $E_{\mathrm{T}}^{\mathrm{miss}}$ in the case of neutrinos, on the downside the branching ratio is much lower. Quarks inevitably hadronise and form jets which are far less accurately measured than leptons and photons, leading to a much worse resolution in the Higgs mass which is determined by the invariant mass of the b-jets $m_{b\bar{b}}$.

"*Kinematic fitting* is a mathematical procedure in which one uses the physical laws governing a particle interaction or decay to improve the measurements of the process."[1] In this thesis one of the most fundamental laws of physics, momentum conservation, is used in a kinematic fit which aims to improve the Higgs mass resolution by correcting the b-jet momentum measurement. Momentum conservation can only be assumed in the plane transverse to the particle beam since events have an unknown boost in beam direction. For this the decay channel $ZH \to l^+l^-b\bar{b}$ is studied as there are no high energy neutrinos involved that would remove an unknown amount of momentum from the measurement. Additionally, since the Z boson is a known particle with a mass of 91.2 GeV, the invariant lepton mass can be constrained to m_Z. The main problem that needs to be dealt with is the occurrence of initial state radiation. Gluons are radiated off the interacting partons and carry away transverse momentum, thus changing the transverse momentum of the ZH system. This needs to be considered in the fit, but is often indistinguishable from the underlying event which is independent from the main scattering process. To a lesser degree final state radiation, gluons irradiated by a b-quark, can have the same effect though most of it becomes part of the b-jet.

The structure of this document is as follows: Chapter 2 contains an overview over the physical background, methods and tools relevant for this thesis. In chapter 3 the construction of the analysis is discussed. The results for the kinematic fit including its application on the background are presented in chapter 4. Finally a summary is given in chapter 5.

CHAPTER 2

Technical Introduction

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a 26.7 km long synchrotron buried 45 m − 170 m underground and located at CERN in Switzerland and France. Collisions between proton bunches take place in 4 intersection points, where the main experiments ATLAS, CMS, ALICE and LHCb are located. In 2011 the first run with proton-proton collisions started with a collision energy of $\sqrt{s} = 7\,\text{TeV}$. In 2012 the collision energy was increased to 8 TeV. In 2013 the LHC was shut down for a major upgrade and is planned to be reactivated with $\sqrt{s} = 13\,\text{TeV}$ in 2015. The integrated luminosity, that is directly proportional to the number of collision events, is $\mathcal{L} = 4.7\,\text{fb}^{-1}$ for 2011 and $\mathcal{L} = 20.3\,\text{fb}^{-1}$ for 2012.[2]

2.2 ATLAS detector

The ATLAS detector is the largest experiment found at the LHC. It has a cylindrical shape that layers different parts of the detector, as shown in figure 2.1. Going from the centre outwards, these are:[4]

Tracker Designed to reconstruct the tracks of charged particles in the centre of the detector, it features a very high granularity and thus spatial resolution. The tracker consists of three parts: the Pixel Detector and the Semi-Conductor Tracker, which are silicon detectors, and the Transition Radiation Tracker which uses drift tubes filled with gas to cover a larger area. The latter can separate electrons/positrons from other charged particles.

Calorimeters A calorimeter measures the energy of a particle by absorbing it via a particle shower. There are two calorimeters in the ATLAS detector: the electromagnetic measuring electrons and photons and the hadronic measuring hadrons.

Muon spectrometer Muons escape the calorimeters and are tracked in the muon spectrometer to measure their momentum and separate them from other particles.

The coordinate system of the ATLAS detector has its origin at the nominal collision point. The x-axis points towards the centre of the LHC ring, the y-axis towards the surface and the z-axis in beam direction. The parametrisation that is usually used for describing particle momenta are modified spherical coordinates:

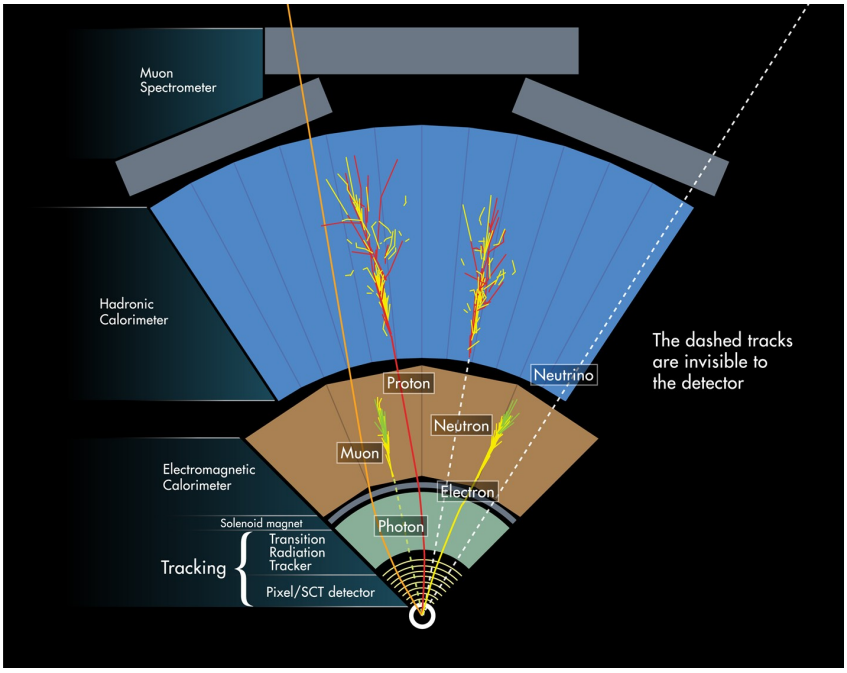


Figure 2.1: Cross section of the ATLAS detector.[3] This picture shows the different parts of the detector and the particles that are observed in them. It also shows how different particles can be distinguished from each other.

- $p_T = |\vec{p}| \sin\theta$, the transverse momentum
- $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$, the pseudorapidity which is invariant under boosts in z-direction
- ϕ, the angle in the x-y-plane

The angular difference between two particles a and b is defined as

$$\Delta R = \sqrt{(\eta^a - \eta^b)^2 + (\phi^a - \phi^b)^2} \tag{2.1}$$

With an expected collision frequency of 20 MHz in 2011 and 2012 the detector would generate more data than it is possible to save. Therefore a selection is made by so called triggers that filter out all events except ones which inhibit certain signatures. After a three layer triggering system the rate at which events can be recorded is reduced to 200 Hz.

2.3 Particle reconstruction

The information extracted from the detector is used to reconstruct momentum vectors and determine their identity. These can only be electrons, muons, photons, certain hadrons and jets. Neutrinos escape undetected and have to be reconstructed using the missing transverse energy, E_T^{miss}. This is a vector in the transverse plane, defined as the negative value of the sum of all calorimeter entries. All other particles decay before they reach the detector and can only be reconstructed using their decay products.

The energy-momentum vector of a decayed particle is the sum of the energy-momentum vectors of its decay products. The original mass of a particle H with decay products b_1, b_2 is called invariant mass:

$$m_H = m_{bb} = \sqrt{(p^{b_1} + p^{b_2})^2} \tag{2.2}$$

The invariant mass function is asymmetric with more pronounced tails towards lower masses. The Bukin fit function[5], designed for asymmetric distributions in particle physics, is used to determine the width and peak position of the invariant mass distribution. The resolution is defined as width over peak position.

2.3.1 Jets

Quarks and gluons carry colour charges of the strong interaction and cannot exist alone. To satisfy the condition of colour neutrality new particles are created. This process is called hadronisation and leads to the creation of so called jets consisting mainly of hadrons.[6] To reconstruct the momentum vector of the original particle, a jet finder algorithm is employed. The reconstruction uncertainties are high: Low energy particles have a low resolution in the calorimeters. The jet finder may miss particles belonging to the jet and assign others that are independent. b and c-quarks in a jet have a chance to decay into an electron or muon and their corresponding neutrino that cannot be detected. As a result jets have a much lower energy (and angular) resolution than other particles. The jets in this thesis were reconstructed with the anti-k_t jet clustering algorithm using a cone radius of 0.4.[7]

To determine what kind of particle originally caused the jet, further research is necessary. A flavour is assigned to a jet with a 'tagger'. A b-tagger for instance examines different properties of a jet to decide if the jet was caused by a b-quark or not. It utilises the decay signature of long-lived B-hadrons that typically fly several 100 µm and thus create a secondary vertex that can be reconstructed due to the vertexing capabilities of the ATLAS tracking system. The properties of the b-tagger are described by several quantities. The efficiency is the percentage of real b-jets that are identified as such. The purity is the percentage of all b-tagged jets, which are tagged correctly. One wants to have a high efficiency for greater statistics as well as a high purity to reduce the number of wrongly tagged jets. For this a trade off has to be found.

2.3.2 Underlying event and pile-up

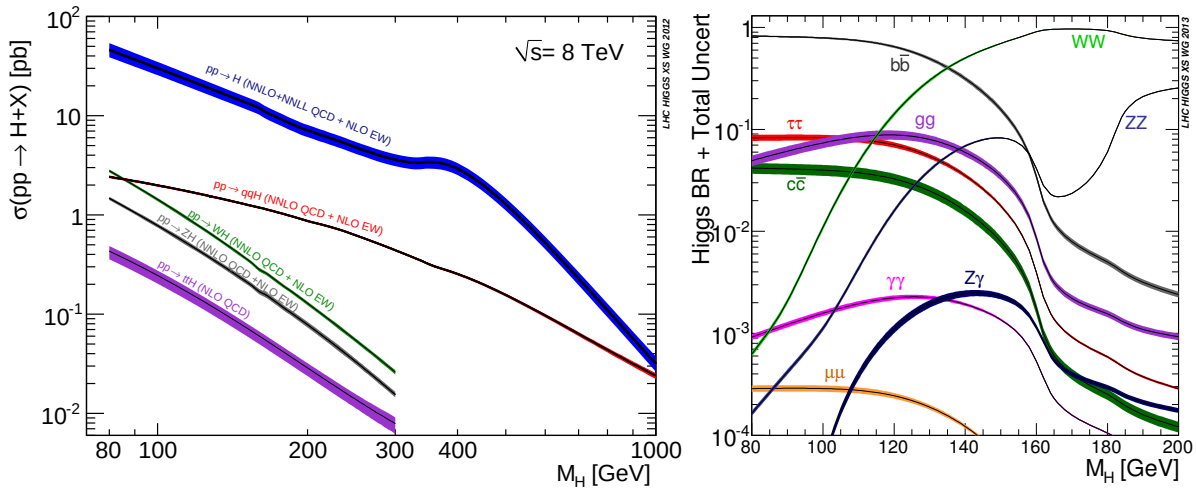
In a proton collision one is interested in the main event, called the 'hard event', the interaction with the signature causing the triggers to react. However a lot more is happening that cannot be ignored. This is called the underlying event and has two different causes: the beam remnant which is the non-interacting fragments of the two beam particles and different partons interacting from the same beam particles as the hard scattering process.[8] Additionally there are pile-up events which are soft scattering processes between other protons in the bunch crossing and spatially separated from the hard collision. In 2012 there was an average of about 20 pile-up interactions for every bunch crossing.

Most outgoing particles of these scattering processes have a very shallow angle towards the beam axis and are therefore detected only in the forward region or not at all. However not all tracks and calorimeter entries caused by these can be completely separated from the hard event which influences the jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ measurement.

2.4 Search for the Higgs Boson

The Higgs mechanism was theorised to explain how the W and Z boson, the gauge bosons of the weak force, could have mass. The Higgs field was introduced as an auxiliary field to accomplish that. A consequence of this theory is the existence of the Higgs boson. In the standard model of particles it is supposed to gives rise to the mass of every massive particle. As a result the Higgs boson couples to all of them which needs to be confirmed experimentally as far as that is possible.

In figure 2.2b are the branching ratios of the Higgs decay shown. So far its decay into $\gamma\gamma$, ZZ, WW and $\tau\tau$ has been confirmed. The decay into two b-quarks is with 53 % the most likely, but it has not yet been observed. There is an overwhelming background from multijet production, which drowns out the Higgs signal. To avoid this problem, the associated production with a vector boson, where a W or Z boson irradiates a Higgs boson and then decays leptonically, is studied. The trigger for these events are then one or two leptons, or $E_{\mathrm{T}}^{\mathrm{miss}}$ in the neutrino-only decay. The downside of the associated production is the low relative cross section as shown in figure 2.2a.



(a) Cross section

(b) Branching ratios

Figure 2.2: Predicted production cross section (a) and decay branching ratios (b) of the Higgs boson.[9]

2.4.1 The $ZH \rightarrow l^+l^-b\bar{b}$ decay channel

The decay channel discussed in this thesis is $ZH \rightarrow l^+l^-b\bar{b}$, the associated production of a Higgs with a Z boson where the Higgs decays into a b-quark pair and the Z into two charged leptons, i.e. electron/positron or muon pair. In this channel are no high energy neutrinos produced such that the momentum in the transverse plane should be conserved.

Additional gluons are irradiated by the strongly interacting particles of the event. Initial state radiation (ISR) are gluons emitted by the interacting partons and final state radiation (FSR) are gluons emitted by the b-quarks before hadronisation. These gluons hadronise and create jets of their own. Most ISR and FSR is soft, i.e. have low energy and their jets cannot be isolated, thus detection is quite difficult. As a consequence it is often not possible to distinguish ISR from the underlying event, while most of the FSR becomes part of the b-jets. Some gluons however have a higher energy and can form their own detectable jets. This is less often the case for FSR, as the angle to the b-quark is generally too small to form an individual jet. The corresponding Feynman graph including ISR and FSR is shown in figure 2.3.

There are two main backgrounds for this decay channel: $t\bar{t}$, two top quarks that each decay into a b-quark and a W boson which in turn decays leptonically, thus creating the signature of two b-jets and leptons. Z+bb is the creation of a Z which decays leptonically and additional b-jets from e.g. ISR gluon radiation. Z+bb is much more difficult to distinguish from $ZH \rightarrow l^+l^-b\bar{b}$ since $m_{b\bar{b}}$ is the only separating variable.

In this analysis the following notation is used: l_1 and l_2 denote the leptons from the Z decay. b_1 and

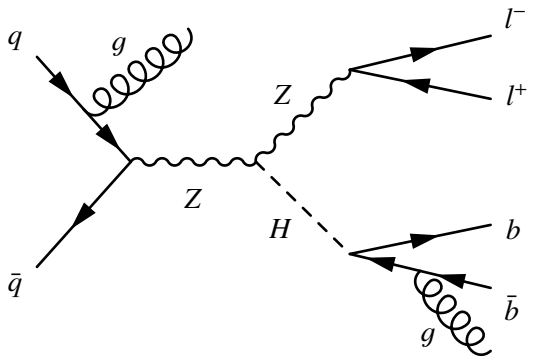


Figure 2.3: The Feynman graph of the $ZH \to l^+l^-b\bar{b}$ decay channel including gluons from initial (left gluon) and final (right gluon) state radiation.

b_2 denote the b-tagged jets, assuming to be tagged correctly and thus corresponding to the quarks from the H decay. In the case of true particles b_1 and b_2 denote the quarks before any final state radiation. j is used for third jets which are not b-tagged.

2.5 Monte Carlo simulated events

An analysis is typically performed on simulated events before it is applied to real measured data. Collections of simulated events are called Monte Carlo samples (MC samples) which try to recreate everything that is happening in the detector such that the analysis on data and MC samples is identical. It also allows to analyse only specific processes without having to deal with the complete background and to test different theoretical models.

MC samples consist of truth and reconstructed information. Truth information are the underlying physical processes which are created randomly according to probabilities calculated from theory. After this a detector simulation emulates the measurement process as close as possible to the actual detector. The simulations are not perfect. The theoretical calculations of the true event are only approximations (leading order or next-to-leading order of Feynman graphs) and have their own uncertainties. QCD processes often cannot be predicted due to their non-perturbative nature and have to be modelled on the basis of prior measurements. If the measured data and simulated Monte Carlo samples disagree outside of the signal region, a rescaling on the background samples can be applied to bring data and MC in accordance.[10]

Using Monte Carlo samples the expected significance in a given signal region of a certain distribution can be calculated, with S the number of signal and B the number of background events:[11]

$$\sqrt{2((S+B)\ln(1+S/B)-S)} \approx \frac{S}{\sqrt{B}} \tag{2.3}$$

2.6 Data

The Monte Carlo samples used in this analysis were created for the data accumulated by the ATLAS detector in 2012. The centre of mass collision energy is $\sqrt{s} = 8\,\text{TeV}$ and the integrated luminosity $20.3\,\text{fb}^{-1}$. The $q\bar{q} \to ZH$ signal samples, with which the main part of this analysis is conducted, are created with Pythia8 and a Higgs mass of 125 GeV. For simplicity $gg \to ZH$ events are not considered

for the signal-only analysis since they account for less than 10 % of the cross section and their topology is very similar to $q\bar{q} \to ZH$ events. Of course they are included when using the full signal+background samples.

2.7 Data processing

The data generated by the detector is processed in stages. The complete information gathered is too large to store and process for every physics analysis, therefore different data formats are used where each derivation takes considerably less storage space than the previous format. The raw data is processed to ESDs (Event Summary Data) containing the full output of reconstruction such as track and calorimeter information. Using ESD actual physical objects like electrons or jets are reconstructed and stored in AODs (Analysis Object Data). D3PDs (Derived Physics Data) are specialised selections from AODs containing only information relevant for the analysis at hand.[12]

Normally an ATLAS analysis is performed on the Worldwide LHC Computing Grid (WLCG). In the Bonn $H \to b\bar{b}$ analysis group custom made "BonnAnalysisD3PD" are derived from the AODs which use much less space than the official "SMD3PD".[13] These can then be used on the local computing cluster in Bonn without having to rely on the Grid. The derivation of the D3PDs requires several weeks of Grid processing. However the local analysis, which is executed far more often, is considerably faster than it would be otherwise.

2.7.1 Analysis frameworks

The analysis frameworks used in this thesis are, like most in high energy particle physics, based on ROOT. While the standard in ATLAS is Athena which uses AODs or D3PDs on the Grid, the Bonn D3PDs are processed differently. The analysis is performed with the following frameworks:

Private n-tuples Information from the Bonn D3PDs are written out into n-tuples which are then saved as ROOT trees. These trees can directly be used to perform an analysis. There are two main advantages to this: It is easy to use since the trees contain high-level information about the main physics objects that are easily processed with ROOT. The second advantage is the possibility to instantaneously create histograms for any given configuration which is very beneficial while developing an analysis. This framework is used in chapter 3.

Overkill The Overkill framework[13] has been developed in Bonn for the $H \to b\bar{b}$ analysis. It directly uses the D3PDs and implements the complete official $V, H \to b\bar{b}$ analysis. It is built modular and highly configurable with its own physics objects, making it very flexible. As a consequence it has a high learning curve and lacks the ability to instantly create histograms. This framework was used in the last sections of chapter 3 and to obtain the results in chapter 4.

2.8 Event selection and corrections

In the ATLAS $V, H \to b\bar{b}$ group two different analyses are performed.[14] The cut-based analysis focuses on improving the significance in $m_{b\bar{b}}$ with cuts, while the multivariate analysis uses BDT classifiers to discriminate between background and signal. Hence the MVA analysis uses a looser event selection than the cut-based.

2.8.1 Private n-tuples

The event selection for the first part of the analysis, implemented using the private n-tuples, is modelled as close as possible to the cut-based analysis. Additionally, truth matching is applied for the b-jets.

- $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ for the leading jet
- $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ for the second jet
- b-tagging working point of 80 %
- truth matching: $\Delta R < 0.3$ between b-jets and true b-quarks
- $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ for the leptons
- $83\,\mathrm{GeV} < m_{ll} < 99\,\mathrm{GeV}$
- $\Delta R > 0.3$ between jets and leptons
- ΔR between b-jets in dependence of p_{T}^{Z} according to the following table:

p_{T}^{Z}	$0-90$	$90-120$	$120-160$	$160-200$	>200
$\Delta R(b_1, b_2)$	$0.7-3.4$	$0.7-3.0$	$0.7-2.3$	$0.7-1.8$	<1.4

The b-tagging algorithm is JetFitterCOMBNN.

The b-jet measurement can be improved significantly when correcting for semileptonic decays and mismodeling in jet energy. If the angular difference between a muon and a jet is less than $\Delta R < 0.4$ the muon is considered part of the jet. The jet energy is reduced by 3 GeV to account for the energy that the muon left in the calorimeter and then the muon vector is added to the jet vector. After that the jet momentum is scaled in dependence of its transverse momentum according to the following formula:[15]

$$\vec{p}^{\,b}_{\mathrm{corrected}} = (0.98 + 0.6e^{-0.031 \cdot p_{\mathrm{T}}^{b}[\mathrm{GeV}]})\vec{p}^{\,b} \tag{2.4}$$

The effect of these improvements in the $m_{b\bar{b}}$ resolution is shown in figure 2.4.

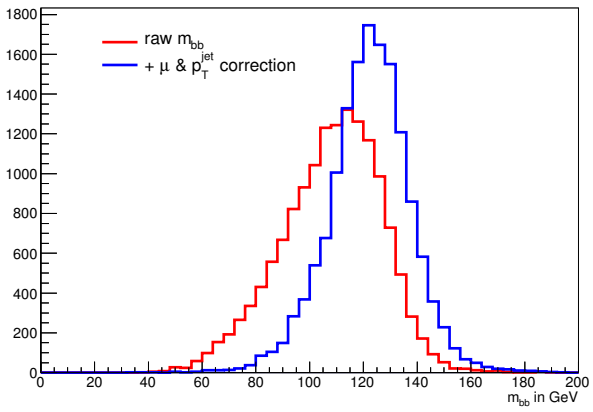


Figure 2.4: The $m_{b\bar{b}}$ distribution before and after applying the μ and p_{T}^{b} corrections on the private n-tuples, showing a good improvement in peak width and position.

2.8.2 Overkill framework

The event selection that was designed for the MVA analysis, implemented in the Overkill framework and thus used in chapter 4 is:

- exactly 2 leptons with identical flavor
  - $p_{\mathrm{T}}^{l_1} > 25\,\mathrm{GeV}$
  - $|\eta^{l_1}| < 2.5$
  - $p_{\mathrm{T}}^{l_2} > 7\,\mathrm{GeV}$
  - $71\,\mathrm{GeV} < m_{ll} < 121\,\mathrm{GeV}$
- 2 b-tagged jets with higher p_{T} than any third jet
  - $p_{\mathrm{T}}^{b} > 20\,\mathrm{GeV}$
  - $|\eta^{b}| < 2.5$
  - $\Delta R(b_1, b_2) > 0.7$ for $p_{\mathrm{T}}^{Z} < 200\,\mathrm{GeV}$
- Veto on forward jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| > 2.5$

Additionally a 70 % b-tagging efficiency with a purity of 93 %, which corresponds to 'medium' tagged jets in the ATLAS analysis, is chosen.[16] The b-tagging algorithm is MV1c which is a variation of MV1 with an improved rejection of c-jets. MV1 is an artificial neural network that uses the output weights of other b-tagging algorithms as its input.

Corrections on the b-jets are also already implemented in Overkill. Similarly to above this is a correction for muons in jets and a p_{T} dependent energy scaling, however it differs from the correction in equation 2.4. There are other calibrations applied to leptons and jets, furthermore events are assigned a certain weight. More information about this can be found in [17].

2.9 Kinematic Fit with constraints

This section about the kinematic fit, which is the main tool of this analysis, follows the description in [18]. It is a least square fit that uses kinematic variables such as momentum and energy of particles in a single event. These variables are modified in a kinematic fit such that they fulfil predefined constraints. The constraints are chosen such that the true (generally unknown) values fulfil them. Thus the fitted variables are expected to be closer to the real values. A kinematic fit can be performed if the event topology allows to formulate constraints that express physical laws such as momentum conservation and known quantities such as the mass of a particle.

Let y_i be n measured uncorrelated kinematic variables, then the constraints f_k should be fulfilled for the true values $\overline{y}_i$.

$$f_k(\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_n) = 0 \qquad \text{for} \quad k = 1, 2, \ldots, m \tag{2.5}$$

The corrected variables $y_i' = y_i + \Delta y_i$ are calculated in such a way that they minimise the likelihood L:

$$L(\vec{y}, \vec{\lambda}) = S(\vec{y}) + 2\sum_{k=1}^{m} \lambda_k f_k(\vec{y}) \qquad S(\vec{y}) = \sum_{i=1}^{n} \frac{(y_i' - y_i)^2}{\sigma_i^2} \tag{2.6}$$

Here λ_k are Langrange Multipliers and σ_i denotes the standard deviation of measured values y_i with regard to their true values. This is equivalent to minimising $\chi^2 = S(\vec{y})$ while the constraints f_k are fulfilled.

With non-linear constraints the solution can in general only be derived iteratively through linearisation.[19] This is done by approximating the constraints in every iteration step as

$$f_k(\vec{y}') \approx f_k(\vec{y}^*) + \sum_{i=1}^{n} \frac{\partial f_k}{\partial y_i} \cdot (y_i' - y_i^*) \approx 0 \tag{2.7}$$

Here $\vec{y}^*$ are the measured values after the last iteration and $\vec{y}'$ after the current. With this linear constraint the solution $\vec{y}'$ can be derived in every step. The fit is defined as converged if χ^2 does not change significantly anymore and the constraints are fulfilled up to an ϵ value. The number of degrees of freedom in this fit is equal to the number of constraints.

For this kinematic fit the implementation *KinFitter*[18, 20] is used. KinFitter is a software package for ROOT and as such is easily integrated into the analysis. Its behaviour has been verified in section A.1.

2.9.1 χ^2 probability test

The minimised χ^2 value is expected to be distributed according to the chi-square probability density function $f(\chi^2, \nu)$, where $\nu = m$ is the number of degrees of freedom. The χ^2 probability $P(\chi^2)$ is defined as[21]

$$P(\chi^2) = \int_{\chi^2}^{\infty} f(\chi'^2, \nu)\ \mathrm{d}\chi'^2 \tag{2.8}$$

which is the probability that another minimised χ'^2 is larger than χ^2. A large χ^2 leads to a small $P(\chi^2)$ which means that it is unlikely to occur and thus can be classified as a bad fit. Under ideal circumstances $P(\chi^2)$ should be uniformly distributed in its range $[0, 1]$. In reality there is often a peak at low probabilities corresponding to events where the constraints of the fit are not compatible with the data. If the distribution is skewed towards high (low) values, the errors are chosen too large (small).[22] The actual formula of $P(\chi^2)$ depends on the degrees of freedom in the fit.[1]

2.9.2 Pull histograms

Another tool to analyse the behaviour of a fit are so called stretch or pull functions. A pull is defined for every kinematic variable in the fit and shows how much it is changed. If a measured parameter has the value y with an uncertainty of σ and is distributed around the true value μ, the general definition of the pull is[23]

$$g = \frac{y - \mu}{\sigma} \tag{2.9}$$

which for Gaussian errors is per definition standard distributed.[2] For this kinematic fit two separate definitions will be used, both describing different aspects.

[1] For two degrees of freedom it is $P(t) = \frac{e^{-t/2}}{2}$ and for three $P(t) = \frac{e^{-t/2}\sqrt{t}}{\sqrt{2\pi}}$.

[2] The standard distribution is a Gaussian distribution with mean 0 and variance 1.

$$g_{\mathrm{m}} = \frac{y_{\mathrm{f}} - y_{\mathrm{m}}}{\sqrt{\sigma_{\mathrm{m}}^2 - \sigma_{\mathrm{f}}^2}} \tag{2.10}$$

$$g_{\mathrm{tr}} = \frac{y_{\mathrm{f}} - y_{\mathrm{tr}}}{\sigma_{\mathrm{f}}} \tag{2.11}$$

Here y_{m} is the originally measured value, y_{f} the fitted value and y_{tr} the true value. σ_{m} denotes the error value of the measurement, which is used as input to the fit, and σ_{f} the error after the fit, defined as the value the variable needs to be changed such that the minimised χ^2 value is increased by one.[23] In an ideal case both pulls would be standard distributed.

g_{m} can be calculated without having any truth information available. A broader (smaller) width means that the variable is changed more (less) than expected by the errors. This can indicate that the errors are chosen wrong. Another reason for broad widths is a badly chosen constraint that forces the variables to move more then they should. If the peak of the distribution is moved, the variable is mainly shifted in one direction by the fit. The cause can be a systematic measurement error or a wrong constraint. The same applies to asymmetric tails, where this is the case for some parts of the measurement.

g_{tr} measures how close the variables are to their true values after the fit. It is more suitable for recognising how well the errors are chosen. If the width is too broad (small) then the error is chosen too low (high).[22] Asymmetric distributions only occur if the fitted variable is systematically shifted from the true value. Asymmetry in g_{tr} does not necessitate asymmetry in g_{m} and vice versa.

2.10 Boosted Decision Trees

This sections is a short introduction to Boosted Decision Trees (BDTs), a kind of multivariate analysis (MVA), which will be used in section 3.4.1. A multivariate analysis is a machine learning algorithm which is able to use many input variables to predict an outcome for a measurement. Before the MVA can be used, it has to be trained on data where the true outcome is known. In the training process a model is built which predicts the true outcome for the given input. After that the prediction model can estimate the outcome for input data where the true outcome is unknown. The performance of the trained MVA is typically tested by splitting the data with known outcome into a train and a test sample. In particle physics the training data are MC samples and the application can be on other MC samples and measured data.

There are two kinds of multivariate analysis: classification and regression. The difference is the target: classification categorises the data, while regression returns a real valued quantity. In the following a regression analysis is used.

A decision tree starts at its root node with the full sample. This sample is split in two by choosing a cut value for an input variable. The events with a value higher than the cut form a next node, those with a lower value another. This is then repeated for these subsamples where the input variable may change. If a certain depth is reached the decision tree stops and returns the result for the remaining subsamples. Boosting is the combination of many trees after another while training. Here the performance of the previous tree is rated and this information is used to create a more accurate tree. After enough iterations the BDT will converge, at which point it can be tested and used.

CHAPTER 3

Kinematic fit for the $ZH \to l^+l^-b\bar{b}$ channel

3.1 Motivation

Events in the $ZH \to l^+l^-b\bar{b}$ channel have on truth level no $E_{\mathrm{T}}^{\mathrm{miss}}$, momentum in the transverse plane should thus be conserved. The only possible neutrinos with significant energies come from semileptonic b decays (or to a lesser degree c decays) and add to the already high measurement uncertainty of the jets. As a result of these uncertainties the momentum conservation is not fulfilled if one considers the reconstructed quantities. This can be exploited to improve the particle momentum vectors with a kinematic fit on event level. In this thesis a least squares fit with constraints is employed. The aim is to improve the invariant mass resolution of the b-tagged jets, and thus the Higgs boson, by bringing them closer to the momentum of the true b-quarks.

Besides the momentum conservation in p_{T} another constraint arises from the invariant mass of the leptons, since they decay from a Z boson which has a known mass of $m_Z = 91.2\,\mathrm{GeV}$ and decay width of $\Gamma = 2.5\,\mathrm{GeV}$.[24] Fitting the leptons does not influence the $m_{b\bar{b}}$ distribution directly, however if combined with the momentum conservation fit it will indeed affect the b-jets. It is not expected to have a large impact however since the leptons are already measured much more accurately than the jets in the first place.

The kinematic fit algorithm used in this chapter is discussed in section 2.9. The analysis is performed on the private n-tuples first since it has advantages for the initial implementation (see section 2.7.1). There are several cuts and corrections applied before the kinematic fit. These are not designed to replace a full analysis but to provide a realistic testbed for the fit while filtering out bad events that could poison the study with a truth matching. They can be found in section 2.8.1. Afterwards the analysis will be implemented into the Overkill framework with a different event selection (see section 3.6).

3.2 Constraints

Two assumptions are made in the $ZH \to l^+l^-b\bar{b}$ decay channel. The first is that the invariant mass of the leptons should be equal to the mass of the Z boson. The resulting constraint for the fit reads as

$$(p_{l_1} + p_{l_2})^2 - m_Z^2 = 0 \qquad m_Z = 91.2\,\mathrm{GeV} \tag{3.1}$$

This is not quite correct as the Z mass is expected to have a Breit-Wigner distribution in contrast to

the delta distribution imposed by this constraint. On the other hand the peak of the m_{ll} distribution is broader and lower positioned than the true m_{ll} distribution (this will be shown in figure 3.11) due to systematic measurement uncertainties. The effects of this simplification will be discussed in section 3.8.

The second and most important assumption is momentum conservation of observed particles. The lack of high energy neutrinos results in a very low $E_{\mathrm{T}}^{\mathrm{miss}}$, such that the sum of all observed momenta should vanish. This is only true for the x- and y-axis in the lab frame since the centre of mass of the collision event is generally boosted in beam direction. The resulting constraints are then

$$p_x^{l_1} + p_x^{l_2} + p_x^{b_1} + p_x^{b_2} + p_x^r = 0 \tag{3.2}$$

$$p_y^{l_1} + p_y^{l_2} + p_y^{b_1} + p_y^{b_2} + p_y^r = 0 \tag{3.3}$$

Here r denotes the *Remnant Vector*. It is a constructed momentum vector that is necessary due to initial and final state radiation which carry transverse momentum away. The Remnant Vector will be treated as a tool to improve the fit rather than a physical quantity. Its definition will be discussed in section 3.4. The chosen coordinate system for r is Cartesian since its p_{T} value is by design close to zero, making polar coordinates impractical.

Particles without mass have three independent variables parametrising the momentum vector. The mass of the b-jets, which is non-vanishing, is chosen constant. There are five momentum vectors that get fed into the fit. These are the two b-jets, the two Z leptons and the Remnant Vector. The constraints do not change every kinematic variable. The p_{T} and ϕ (or p_x and p_y respectively) values of all fit objects will be modified to satisfy equations 3.2 and 3.3. The whole lepton momentum needs to be adjusted to fulfil equation 3.1 which means that p_η^l will also be shifted. On the other hand η^b, and p_z^r are independent variables not affected by the constraints.

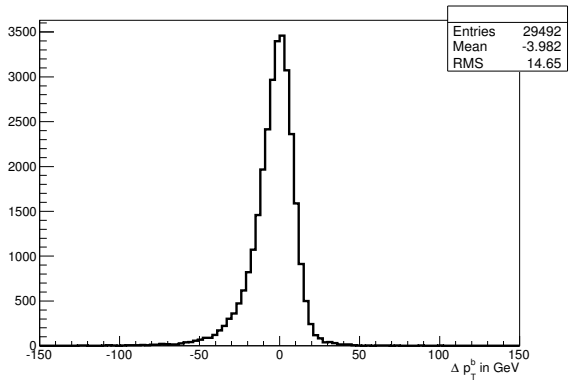
3.3 Error parametrisation

For the fit to work one needs to estimate the standard deviation values σ_{m} of the kinematic variables involved. The fit quality depends highly on the right choice of errors as examined in section A.1.

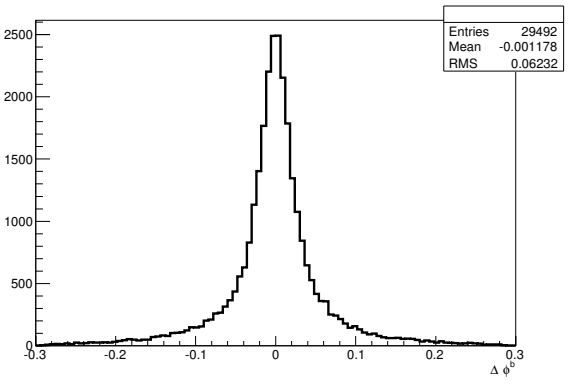
There are several publications which estimate the resolutions for leptons and jets.[25, 26] It is questionable how reliable those uncertainties are after applying corrections to the reconstructed particle vectors. The deviations for variables of the Remnant Vectors cannot be determined through known values of other objects, since the relation between r_{truth} and the different definitions discussed in section 3.4 is impossible to determine.

The deviation values can be obtained empirically using truth information contained in the Monte Carlo sample. For this each fit object needs to be associated to a truth object. The signal leptons are assigned to truth leptons decayed from the Z boson. Similarly, the b-jets are associated to the b-quarks directly decaying from the Higgs before any final state radiation. The truth object with the smallest angular difference ΔR to the measured object is chosen. The Remnant Vector is simply assigned r_{truth} as defined in equation 3.4.

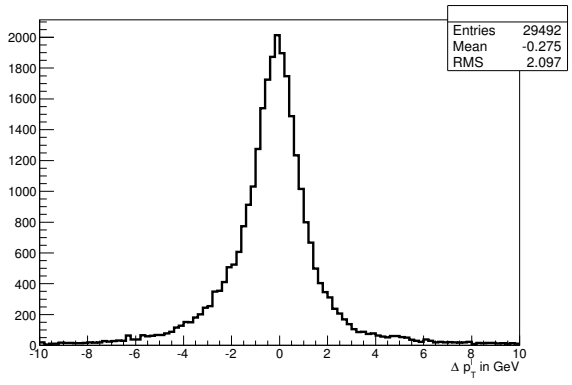
The differences between the kinematic variables of measured and associated truth objects are then filled into histograms which are shown in figure 3.1. The RMS value of these histograms is close to the standard deviation of the corresponding variable and can be used in the kinematic fit. This is not perfect as the deviation distributions are not Gaussian, however small changes on the σ_{m} values have only minimal impact on the fit result. In figure 3.1f the deviations of p_x and p_y are merged into one histogram due to their symmetry, the RMS value is used for both. The absolute value of the errors is



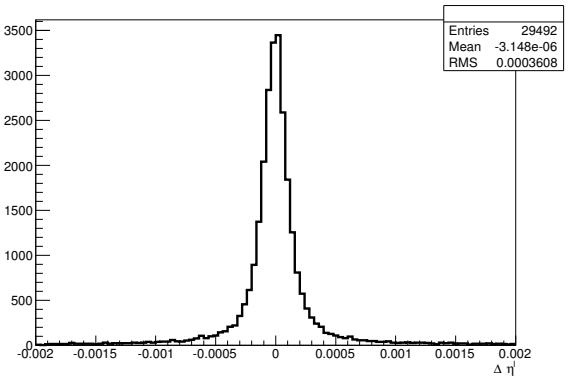
(a) $\Delta p_{\mathrm{T}}^{b}$



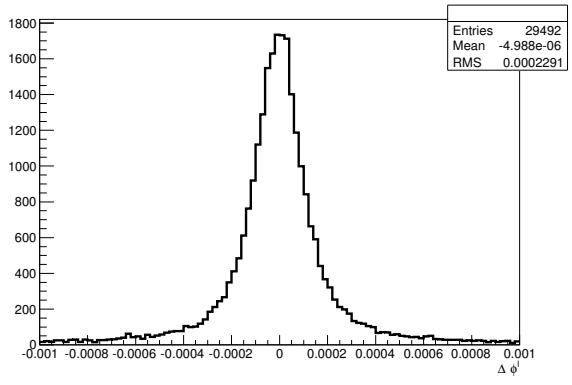
(b) $\Delta\phi^{b}$



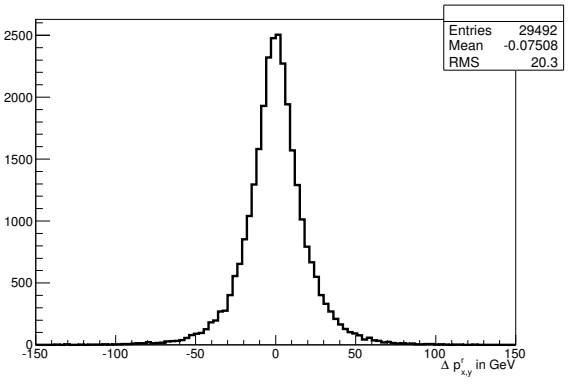
(c) $\Delta p_{\mathrm{T}}^{l}$



(d) $\Delta\eta^{l}$



(e) $\Delta\phi^{l}$



(f) $\Delta p_{x,y}^{r}$ with $r = r_{\mathrm{jets}}$

Figure 3.1: Deviation plots for the σ_{m} determination of the kinematic variables in the fit. The definition for a variable y is $\Delta y = y_{\mathrm{meas}} - y_{\mathrm{truth}}$. The shape of $\Delta p_{\mathrm{T}}^{b}$ shows the bias of reconstructed jets towards lower p_{T} even after corrections. This is less prominently also visible for the leptons.

not important - rescaling all errors equally does not change the fit result - but their relative value to the other errors. For that reason all uncertainties should be obtained using the same method i.e. using truth information rather than mixing it with externally obtained values.

3.4 Remnant Vector

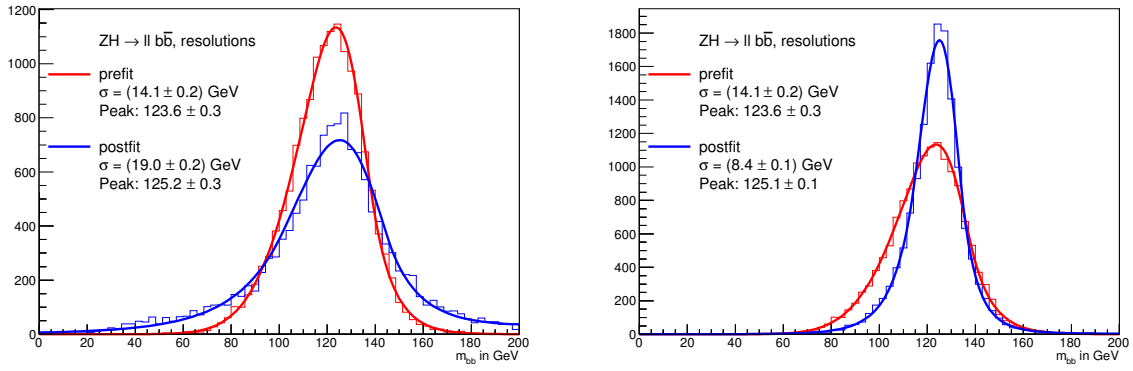
The correct determination of the Remnant Vector is vitally important for the performance of the fit as shown in section A.1. It should be noted that its value of the z-component is irrelevant since p_z^r (or η^r) does not influence the fit. The following definitions are tested and compared in fit performance:

No Remnant Vector in figure 3.2a is demonstrated how the fit behaves if one where to leave out the Remnant Vector completely. This is obviously not an option.

Truth To get a feeling what could be achieved by the right choice of the remnant vector, it is instructive to look at the optimal values using truth information. r_{truth} is defined as:

$$r_{\text{truth}} = -(b_1^{\text{truth}} + b_2^{\text{truth}} + l_1^{\text{truth}} + l_2^{\text{truth}}) \tag{3.4}$$

With this r_{truth} is equal to the ISR of the hard event. In figure 3.2b it is shown that the choice of r_{truth} in the fit can almost halve the width of the distribution. Clearly choosing r as good as possible is very important. r_{truth} is also used to determine the deviation $\Delta p_{x,y}^r$ and thus σ_{m} of the other definitions.



(a) without r (b) r_{truth}

Figure 3.2: Performing the kinematic fit without the Remnant Vector (a) leads to a deterioration of the mass resolution. Fitting with the true definition of r (b) shows what alone a perfectly chosen Remnant Vector could achieve.

Inverse If one were to define the Remnant Vector equivalent to r_{truth} with the reconstructed particle vectors, equations 3.2 and 3.3 would be fulfilled automatically and the fit would have no effect apart from minimal changes caused by the m_{ll} constraint. However, one can add the measured $E_{\text{T}}^{\text{miss}}$ to the equation assuming this is primarily due to b-jet mismeasurement. This results in the definition of r_{inv}:

$$r_{\text{inv}} = -(b_1 + b_2 + l_1 + l_2 + E_{\text{T}}^{\text{miss}}) \tag{3.5}$$

Note that this does not mean the fit achieves $E_{\mathrm{T}}^{\mathrm{miss}} = 0$. To let $E_{\mathrm{T}}^{\mathrm{miss}}$ vanish completely, one has to replace the right hand side of equations 3.2 and 3.3 with $E_{\mathrm{T}}^{\mathrm{miss}}$. The resulting kinematic fit however only broadens the peak width, regardless of chosen Remnant Vector, and should not be used.

Zero It is possible to simply set

$$r_{\mathrm{zero}} = 0 \tag{3.6}$$

This assumes that the ISR has a low impact on the measurement. This can be reasonable to assume for the 2 jet region and is equivalent to using a soft constraint on the momentum conservation.

Jets r_{jets} is defined as the sum over all non-tagged jets:

$$r_{\mathrm{jets}} = \sum_{j \neq b} j \tag{3.7}$$

This is motivated by the fact that the only observed non-signal objects with significant momentum are additional jets from initial and final state radiation. If a third jet is caused by final state radiation this definition is counter-productive since it is missing from the invariant mass and even pushes $m_{b\bar{b}}$ down, however most final state radiation is included in the b-jet cones. Only jets with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ are recognised. In events containing only the 2 b-jets r_{jets} is equivalent to r_{zero}.

In figure 3.3 are the fit results with the different definitions of the Remnant Vector shown. r_{jets} is superior to r_{zero} since it models events with more than 2 jets better. r_{inv} yields a better result for the peak position than r_{jets}, however the width for the latter is smaller which is more important. The r_{inv} definition suffers from a badly measured $E_{\mathrm{T}}^{\mathrm{miss}}$ which is caused by calorimeter uncertainties as well as soft scattering processes caused by the underlying event and pile-up. In the following sections r_{jets} is used as default.

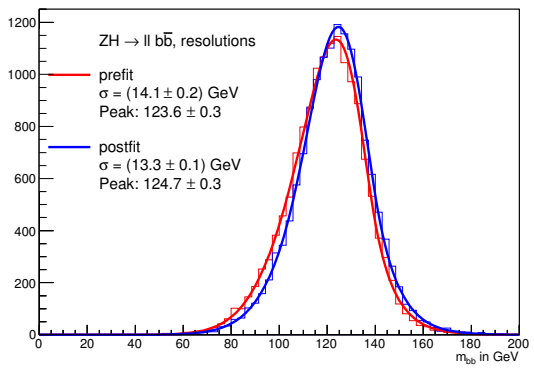
There are more definitions possible than those presented, like choosing $r = \pm E_{\mathrm{T}}^{\mathrm{miss}}$ and mimicking the r_{truth} statistical properties with random values, which fail for various reasons. Other possibilities include adding weights and statistical variations to combine different definitions, this has not been extensively tested.

3.4.1 Improving the Remnant Vector with multivariate regression

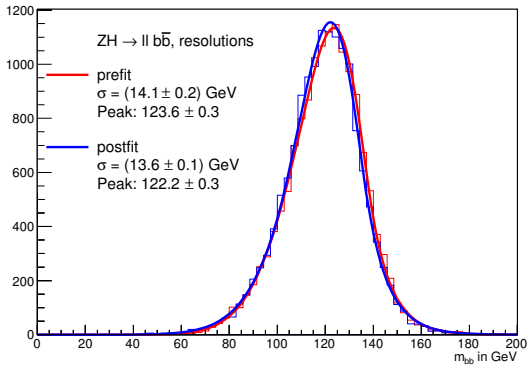
Multivariate regression has been successfully used to improve the b-jet energy measurement.[27] Here it is already corrected using a different method, but the regression can be used to improve the Remnant Vector instead. r needs to be as close to its true value as possible to improve the $m_{b\bar{b}}$ resolution and it is reasonable to assume that a sophisticated method such as regression might improve the rather simple choice of r_{jets}.

The software used for the regression is TMVA[28] which is included in ROOT. The multivariate method is Boosted Decision Trees (see section 2.10) with Gradient Boosting. To build upon an already reconstructed Remnant Vector r_{reco}, the deviation $\Delta r_{\mathrm{reco}} = r_{\mathrm{reco}} - r_{\mathrm{truth}}$ is chosen as a target rather than r_{truth} directly.

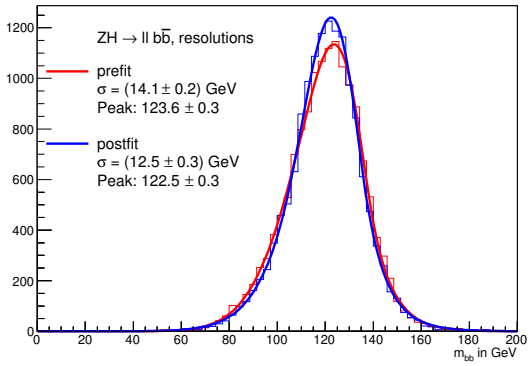
$$X_{\mathrm{target}} = p_x^{r_{\mathrm{reco}}} - p_x^{r_{\mathrm{truth}}} \tag{3.8}$$



(a) r_{inv}



(b) r_{zero}



(c) r_{jets}

Figure 3.3: The invariant b-jet mass after the fit using different definitions for r. r_{inv} yields the best result in peak position while r_{jets} yields the best improvement in peak width.

The regression on target X_{target} then yields the weights which are applied to create X_{trained}. With this a new Remnant Vector r_{BDT} and its deviation is defined:

$$p_x^{r_{\text{BDT}}} = p_x^{r_{\text{reco}}} - X_{\text{trained}} \tag{3.9}$$

$$\Delta p_x^{r_{\text{BDT}}} = p_x^{r_{\text{reco}}} - X_{\text{trained}} - p_x^{r_{\text{truth}}} \tag{3.10}$$

The p_y^r correction is derived by applying the same weights while interchanging $x \leftrightarrow y$ coordinates. This has been confirmed to be justified as training both variables individually leads to very similar results. The basis for the regression is chosen as $r_{\text{reco}} = r_{\text{jets}}$. For kinematic input variables Cartesian coordinates are used since the target is Cartesian as well. The input variables are determined the following way:

- Start with a basic set of variables: p_x^r, $p_{x,y}^{b_{1,2}}$
- Add more variables[1] and keep them if RMS of $\Delta p_{x,y}^r$ improves
- Try removing variables without deteriorating the regression result
- The final choice of variables is the following:
  - p_x of r_{reco}
  - p_x of b_1, b_2
  - p_x of l_1, l_2
  - b-tag weight of b_1, b_2
  - $E_{\text{T},x}^{\text{miss}}$

Figure 3.5a shows that the deviation of r_{BDT} is about 20 % smaller than r_{jets}, the p_{T} shape in figure 3.4 looks better as well. From this one would expect an improvement in the fit result, however in figure 3.5b the width after fitting is clearly broader. On the other hand the peak is shifted more towards the true value. In this it is more similar to the result with r_{inv}. If a variable set has less improvement in σ_{m} the fit result of r_{BDT} is closer to that of r_{jets}.

It is unclear why the multivariate regression has this negative effect. One possible reason could be the use of kinematic variables of the b-jets and leptons, creating a correlation between the Remnant Vector and the other fit objects. Repeating the regression with an adapted set containing no b-jet and lepton variables also yields no positive result; the improvement in $\Delta p_{x,y}^r$ is much smaller and so is consequently the negative effect on the fit. Another idea is to change the target and input variable coordinate system, thus training on Δp_{T}^r and $\Delta\phi^r$. The result here is quite similar to using Cartesian: the deviations of the variables gets smaller while the fit resolution gets worse. To assure that those two variables do not influence each other the following check is executed: one variable is chosen optimally as $\phi^r = \phi^{r_{\text{truth}}}$ while the fit result is compared using $p_{\text{T}}^{r_{\text{reco}}}$ and $p_{\text{T}}^{r_{\text{BDT}}}$. The same is repeated with p_{T} and ϕ interchanged. In both cases the resolution gets slightly worse. Other things that were tried, but did not lead to any gain:

- A different r_{reco} choice

[1] The following variables were considered: energy and momentum variables of leptons, b-jets, third jets and $E_{\text{T}}^{\text{miss}}$; the additional b-jet variables b-tag weights, vertex fraction, EM fraction, 0.6 cone momenta, number of tracks, in-cone muons and area size; $\Delta R(b_1, b_2)$; scalar sum of transverse momenta; number of jets; lepton ID (e or μ)

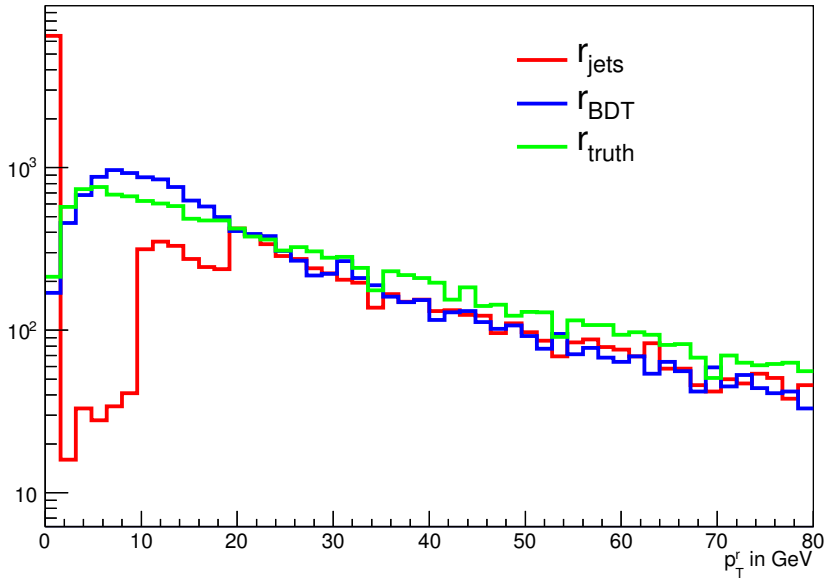
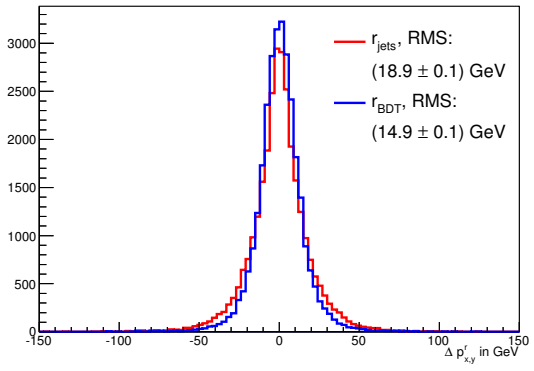
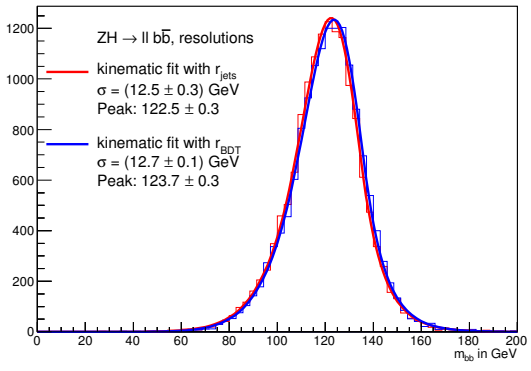


Figure 3.4: Comparison between p_T^r of r_{jets}, r_{BDT} and r_{truth}. The peak at zero for r_{jets} is the 2 jet region. The steps are a result of the p_T of the non-tagged jets; the jet finder algorithm requires at least $p_T > 10\,\text{GeV}$ and has looser criteria for $p_T > 20\,\text{GeV}$. After regression r_{BDT} resembles more closely the shape of r_{truth}.



(a) Deviation of $p^r_{x,y}$



(b) $m_{b\bar{b}}$ resolution

Figure 3.5: Results of the multivariate regression: The enhanced Remnant Vector (a) does not improve the kinematic fit (b).

- Different object selection (e.g. only 2-jet or 3-jet region, looser cuts for greater statistics)
- Using σ_m of r_reco also for r_BDT to exclude deterioration caused by too small errors

While the exact reasons remain unclear, the conclusion is that BDT regression does not seem to be applicable here.

3.5 Splitting the phase space for the error parametrisation

In section 3.3 was discussed how one error value per kinematic variable can be obtained. In reality the deviations change for different regions of the phase space; for example jets with higher p_T have greater absolute uncertainties in the energy measurement. In order to obtain higher precision the deviation histograms are split into different phase space regions of p_T and η for each momentum vector, i.e. σ_m for p_T and ϕ of a jet will depend on its p_T and η values. Since η is meaningless for the Remnant Vector, it is only segmented into p_T regions. More segments mean higher granularity which in turn leads to a higher accuracy of deviation values for each segment. On the other hand it is necessary that the statistics are sufficiently high in each segment, a low number of entries leads to an imprecise RMS value. So a trade off has to be found.

The binning for the p_T^r segments is determined dynamically: Starting with the highest p_T value, the lower boundary of the bin is chosen such that it has ~ 1000 entries.[2] This is repeated until $p_\mathrm{T} = 0$, if the lowest bin has too few entries it is merged with the previous. In figure 3.6 are the resulting bins and deviations for r_jets. A Remnant Vector with low p_T^r is generally more accurate. The lowest bin has a higher σ_m value than its neighbour since it includes all events in the 2 jet region for which r_jets is zero.

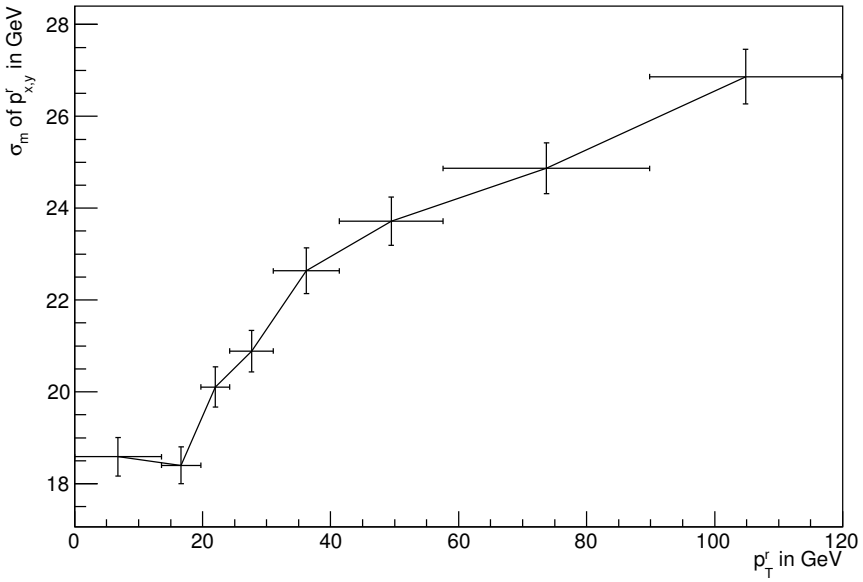
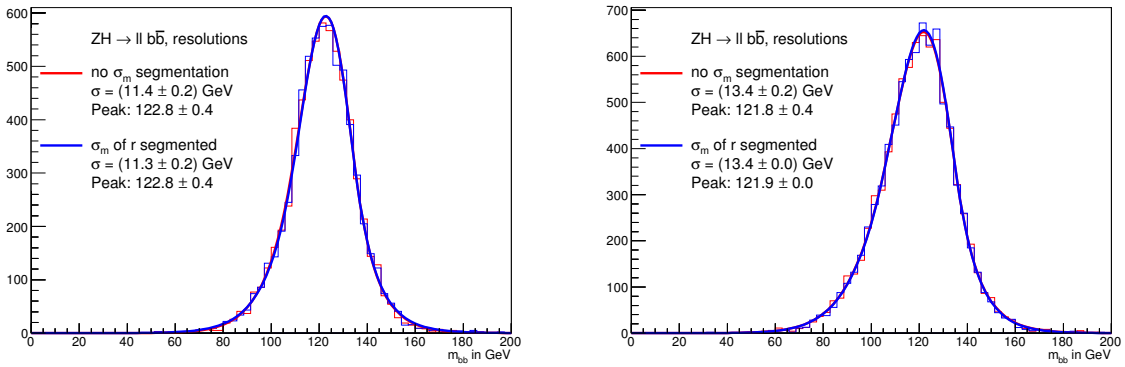


Figure 3.6: σ_m values for $p_{x,y}^r$ obtained through dynamic segment binning. The error on the x-axis indicates which p_T^r range it covers. In events with $p_\mathrm{T}^r > 120\,\mathrm{GeV}$ the highest bin is used.

[2] This is done using a binary search algorithm which is probably far from optimal, but sufficient here.

For the jets and leptons the segments are chosen manually since a dynamic σ_{m} determination would be much more complex. There are 3 p_{T} and 3 η regions combining into 9 segments such that the statistics are not too low in any segment. The resulting errors can be found in table 3.1.

In figure 3.7 is shown that the p_{T}^r error segmentation yields no significant improvement in the final fit result regardless of the region. Neither does the additional lepton and jet segmentation as shown in figure 3.8. This may be due to the fact that the statistics used are too low to get a noticeable effect. Since there is at least a minimal although not significant improvement and it does not harm the fit, the segmentation will be used in the next sections.



(a) Number of jets = 2

(b) Number of jets > 2

Figure 3.7: The invariant mass fitted without and with error segmentation in p_{T}^r in events with 2 jets (a) and events with more than 2 jets (a). There is no significant improvement.

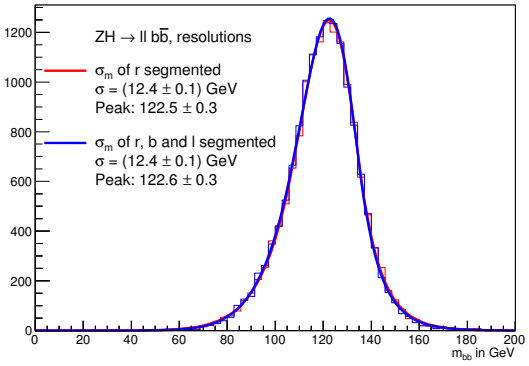


Figure 3.8: The invariant mass fitted without and with error segmentation for $b_{1,2}$ and $l_{1,2}$.

3.6 Porting the analysis to the *Overkill* framework

For the results the kinematic fit is implemented into the Overkill framework (see section 2.7.1) such that it is comparable to the official ATLAS $ZH \to l^+l^-b\bar{b}$ analysis. The kinematic fit is implemented as Overkill modules. The following sections on the χ^2 probability (section 3.7) and pulls (section 3.8) are also done with Overkill.

$p_T \setminus \eta$	> 0.7	$0.7 - 1.7$	> 1.7
< 40	11.1	10.9	10.7
$40 - 70$	12.7	12.9	11.9
> 70	17.8	18.1	15.3

(a) σ_m for p_T^b in GeV

$p_T \setminus \eta$	> 0.7	$0.7 - 1.7$	> 1.7
< 40	0.0894	0.0911	0.0849
$40 - 70$	0.0609	0.0635	0.0572
> 70	0.0416	0.0415	0.0403

(b) σ_m for ϕ^b

$p_T \setminus \eta$	> 0.7	$0.7 - 1.6$	> 1.6
< 45	1.31	1.44	1.40
$45 - 60$	1.42	1.75	1.82
> 60	2.25	2.80	2.89

(c) σ_m for p_T^l in GeV

$p_T \setminus \eta$	> 0.7	$0.7 - 1.6$	> 1.6
< 45	0.000664	0.000455	0.000525
$45 - 60$	0.000629	0.000417	0.000507
> 60	0.000622	0.000408	0.000451

(d) σ_m for η^l

$p_T \setminus \eta$	> 0.7	$0.7 - 1.6$	> 1.6
< 45	0.000222	0.000243	0.000304
$45 - 60$	0.000215	0.000220	0.000279
> 60	0.000195	0.000216	0.000253

(e) σ_m for ϕ^l

Table 3.1: Error values for lepton and jet variables segmented into p_T and η regions. p_T categories are in GeV.

The σ_m determination is implement similarly as in section 3.3. There are additional truth matching cuts applied for creating the deviation plots only, such that incorrectly reconstructed particles do not skew the distributions. These are $\Delta R < 0.3$ between true leptons/b-quarks and their reconstructed counterparts and $90\,\text{GeV} < m_{b\bar{b}} < 140\,\text{GeV}$. The dynamic segmentation for p_T^r has not been implemented for technical reasons, the binning is chosen by hand instead. The σ_m values used can be found in table A.1. For the Remnant Vector $r = r_{\text{jets}}$ is confirmed to be still the best choice.

3.7 χ^2 probability test

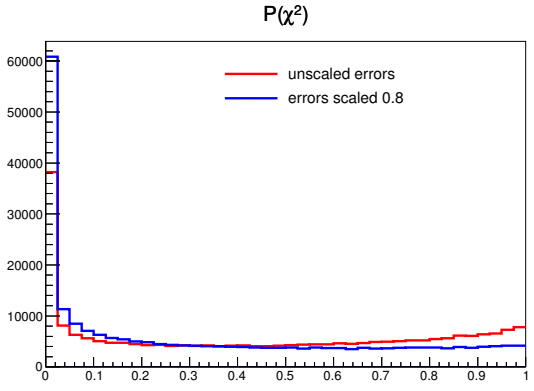


Figure 3.9: $P(\chi^2)$ before and after scaling all σ_m values with 0.8. There is no change in the actual fit result.

The probability distribution of the minimised χ^2 of the fit should be uniform for a model that describes the data perfectly (see section 2.9.1). As shown in figure 3.9 this is not the case here. There is a large

peak at a probability of zero which can be attributed to events that do not fit the model. This is expected to some degree. It can be events where the reconstructed leptons or b-jets do not correspond to the true signal particles or the mismeasurement is simply too high. Another possibility is third jets that do not come from initial or final state radiation but the underlying event such that they should not be included in the transverse momentum of the event.

There is also a rise of the distribution towards high probabilities indicating that the errors are chosen too high. Hence all errors are scaled down with a factor of 0.8, creating an even distribution apart from the peak at zero which grows in the process. The actual fit result does not change when scaling all errors down evenly, it is equivalent to changing the step size for the iterations in the fit algorithm.

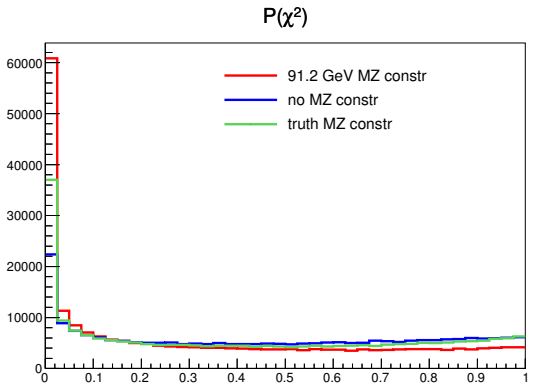


Figure 3.10: The χ^2 probability when fitting with m_Z constrained to 91.2 GeV, to m_{ll}^{truth} and without Z mass constraint. A high amount of events reject the fit hypothesis with a constrained invariant lepton mass.

$P(\chi^2)$ for the fit without the m_Z constraint (figure 3.10) has a much lower peak at low probability. The obvious explanation for that is the wrong assumption of a hard mass constraint of 91.2 GeV that does not account for the Breit-Wigner distribution of the mass peak as can be seen in figure 3.11. To study this the fit is repeated with the mass constraint in equation 3.1 set to $m_Z = m_{ll}^{\text{truth}}$. As shown in figure 3.10 the peak is smaller than with the 91.2 GeV constraint, but not as small as fitting without it. The reason for this can be found in the pull distributions discussed in the next section.

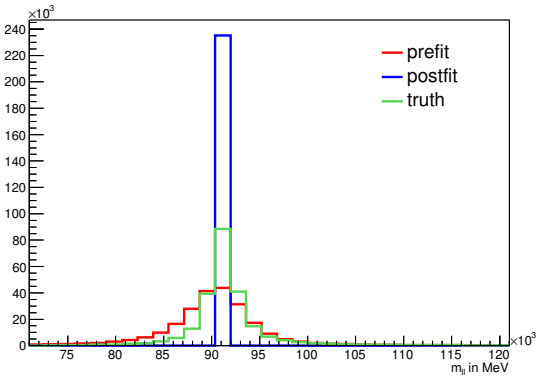


Figure 3.11: The invariant mass of the leptons before and after the kinematic fit, and of the true leptons. The delta peak in the postfit distribution is caused by the $m_{ll} = 91.2\,\text{GeV}$ constraint. The prefit m_{ll} distribution is systematically lower than the true distribution, this is caused by leptons with low reconstructed p_{T}.

3.8 Pull tests

In figure 3.12 are the pulls g_{m}, as defined in equation 2.10, shown fitted with and without the Z mass constraint. The pulls without the mass constraint are close to normal distributions as expected. This means that all kinematic variables are shifted evenly with regards to their error value. If the m_Z constraint is used however, the lepton variables have a broader distribution than the others. This is a consequence of the fact that the actual mass of the Z boson, which the leptons are spawned from, is indeed not exactly 91.2 GeV as the constraint assumes but Breit-Wigner distributed. Therefore some leptons have to be shifted more strongly than the other particles to satisfy the constraint. When fitting with m_Z constrained to m_{ll}^{truth} they are closer to standard distributions, but there is still some difference in the width for p_{T}^{l} and ϕ^{l}. The constraint leads also to a shifted p_{T}^{l} peak (figure 3.12c) which means that the fitted value is generally higher than before. This is not necessarily a bad thing as the reconstructed energy of the reconstructed leptons is systematically too low. On the other hand the fit without m_Z constraint shows a tail towards even lower p_{T}^{l} values. This is because the Z and H bosons - and thus leptons and b-jets - are likely to have opposite transverse momentum. If p_{T}^{b} is pushed to higher values because of p_{T} conservation then p_{T}^{l} is necessarily lower.

For the non-lepton variables the distributions are very similar since the mass constraint does not influence them directly. The indirect effect is small since the lepton uncertainties are significantly lower thus changing much less, leading to relatively little additional adjustment for the other variables. The transverse momentum of the b-jets is slightly shifted towards higher values (figure 3.12a) reflecting that the corrections made are not completely eliminating the bias towards lower jet energies.

3.8.1 Using truth information

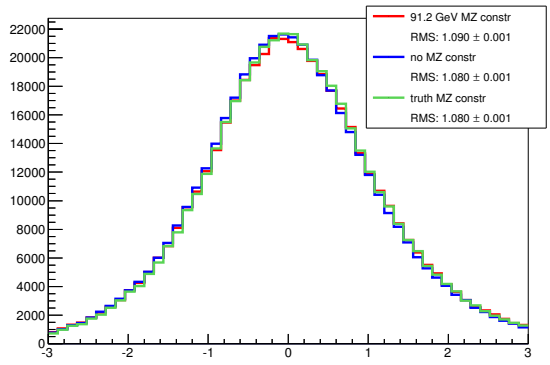
The pulls g_{tr} using truth information as in equation 2.11 give more specific details how the fit behaves. In an optimal situation the variables would be normal distributed. Here however the g_{tr} shape is not even Gaussian, it is most likely distorted for the same reasons that cause the spike at low $P(\chi^2)$. The width for the different variables varies quite a lot indicating a suboptimal error choice. Therefore all σ_{m} values are scaled such that an RMS value of one is reached in every pull distribution.[3] The result can be seen in figure 3.13, the scaling factors for the errors of every kinematic variable can be found in table 3.2. The probability $P(\chi^2)$ after rescaling is again skewed towards one as can be seen in figure A.3.

	p_{T}^{b}	ϕ^{b}	p_{T}^{l}	η^{l}	ϕ^{l}	$p_{x,y}^{r}$
with m_Z constr.	1.11	0.85	1.95	1.10	0.80	2.40
without m_Z constr.	1.11	0.85	0.96	-	0.80	2.40

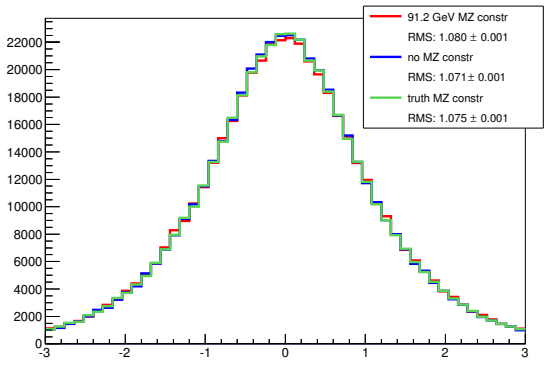
Table 3.2: The scale factors for the fit input errors σ_{m} determined by optimising the pulls g_{tr} in figure 3.13 and A.2. The scales of the lepton errors depend on whether the m_Z constraint is used. The errors for the Remnant Vector would have to be chosen very high to reach an RMS of 1.0 in the pull, the factor 2.4 is a compromise. The difference in the fit result is marginal if choosing e.g. 1.4.

Another thing to notice in the g_{tr} distributions is that p_{T}^{l} in figure 3.13c is not skewed like g_{m} is. This supports the claim that the systematically higher fitted value is a sensible result. On the other hand p_{T}^{b} in figure 3.13a is systematically too low after the fit. This explains that the mass peak in $m_{b\bar{b}}$ is shifted towards lower values than 125 GeV.

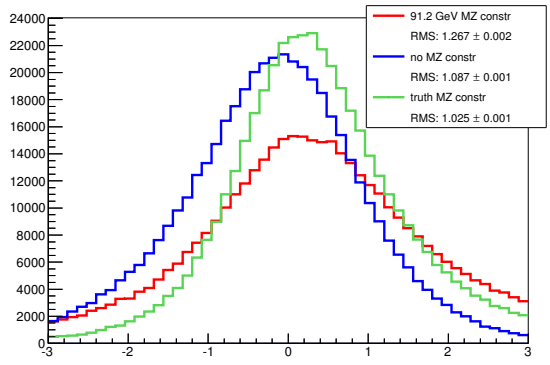
[3] Since the distributions contain badly measured events and are in general not Gaussian distributed, it is not necessarily optimal that the width is determined purely by histogram RMS values.



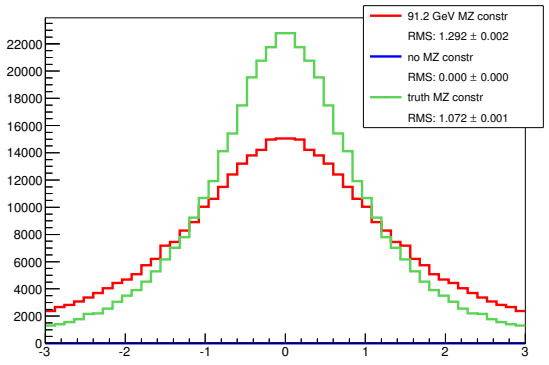
(a) p_{T}^{b}



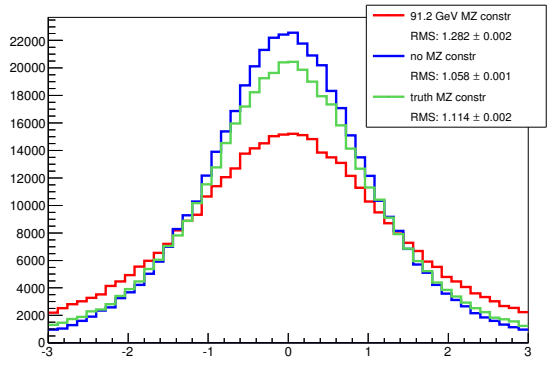
(b) ϕ^{b}



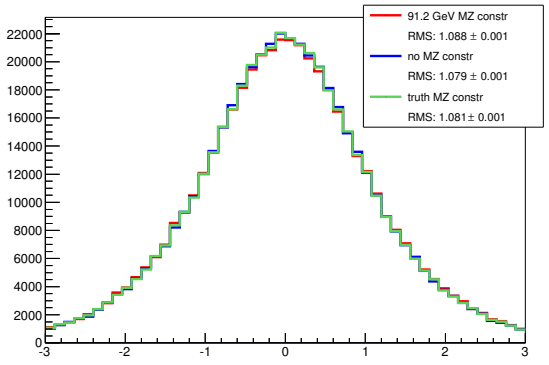
(c) p_{T}^{l}



(d) η^{l}

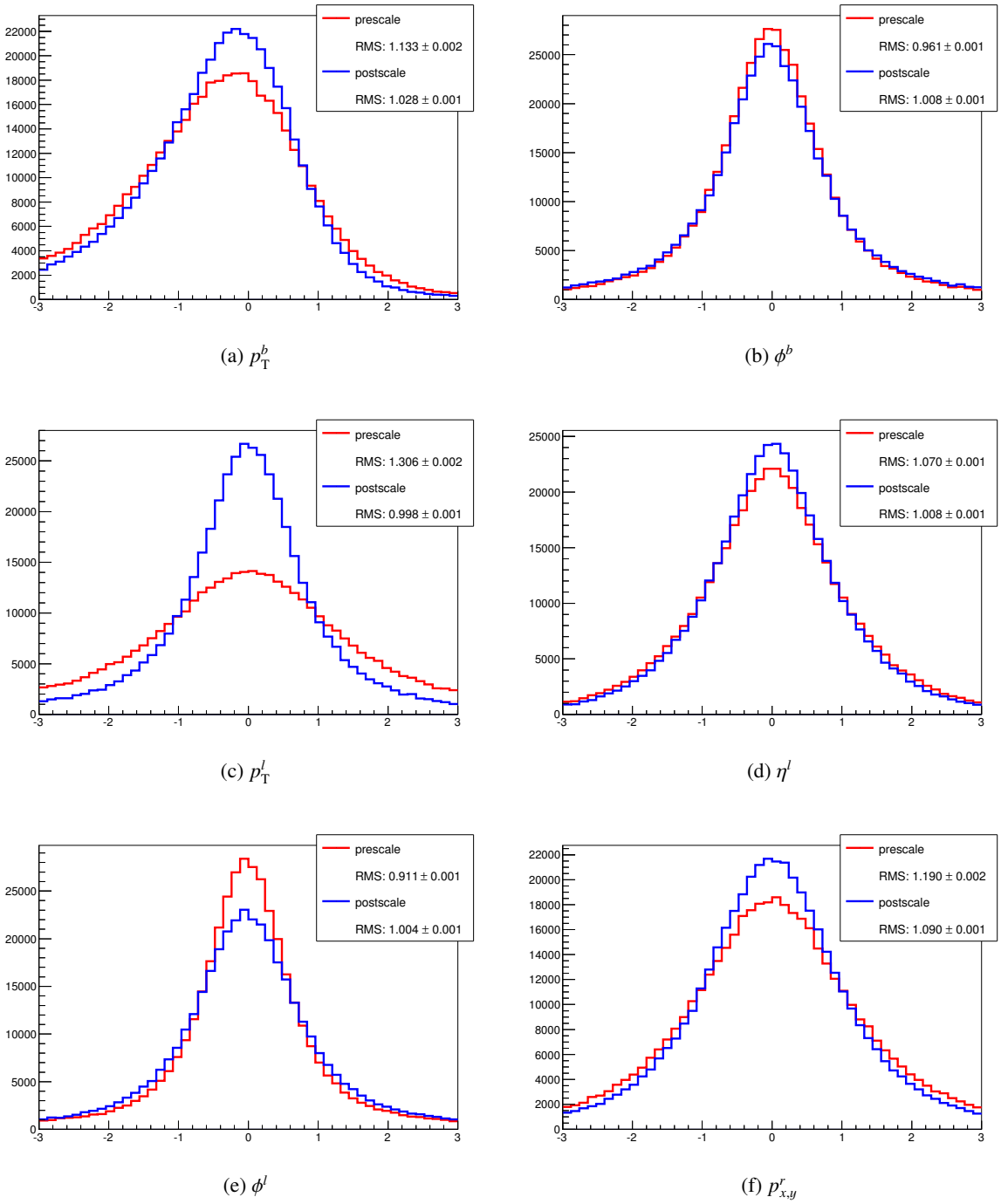


(e) ϕ^{l}



(f) $p_{x,y}^{r}$

Figure 3.12: Pulls g_{m} for the fit variables with the m_Z constraint fixed to 91.2 GeV, set to m_{ll}^{truth} and left out. The pulls are similar for non lepton variables as the mass constraint does not influence them directly and their resolution is worse anyway. The η^{l} variable (d) is not used when fitting without m_Z constraint.



(a) p_{T}^{b} (b) ϕ^{b}

(c) p_{T}^{l} (d) η^{l}

(e) ϕ^{l} (f) $p_{x,y}^{r}$

Figure 3.13: Pulls g_{tr} for the kinematic variables when fitting with m_Z constrained to 91.2 GeV. It was not possible to reach an RMS value of 1.0 for the $p_{x,y}^{r}$ variable.

The rescaling does not improve the fit result in $m_{b\bar{b}}$ as shown in figure A.4. The reason may be bad events that are distorting the pull distributions and are causing the peak at low $P(\chi^2)$. This could be mitigated by choosing the cut that is used for the error determination also for the pulls and by using a Bukin fit function for a better width estimation. Since the rescaling does not yield any positive effects, it will not be used.

CHAPTER 4

Results

All results in this chapter are derived within the Overkill framework (see section 3.6) such that the kinematic fit is implemented on the same data as the maximum likelihood fit which is used in the ATLAS analysis. After full implementation the result in figure 4.1 shows an improvement of 8 % for the Higgs mass resolution. If the Z mass constraint is used in the fit, the resolution gets worse by a minimal amount.

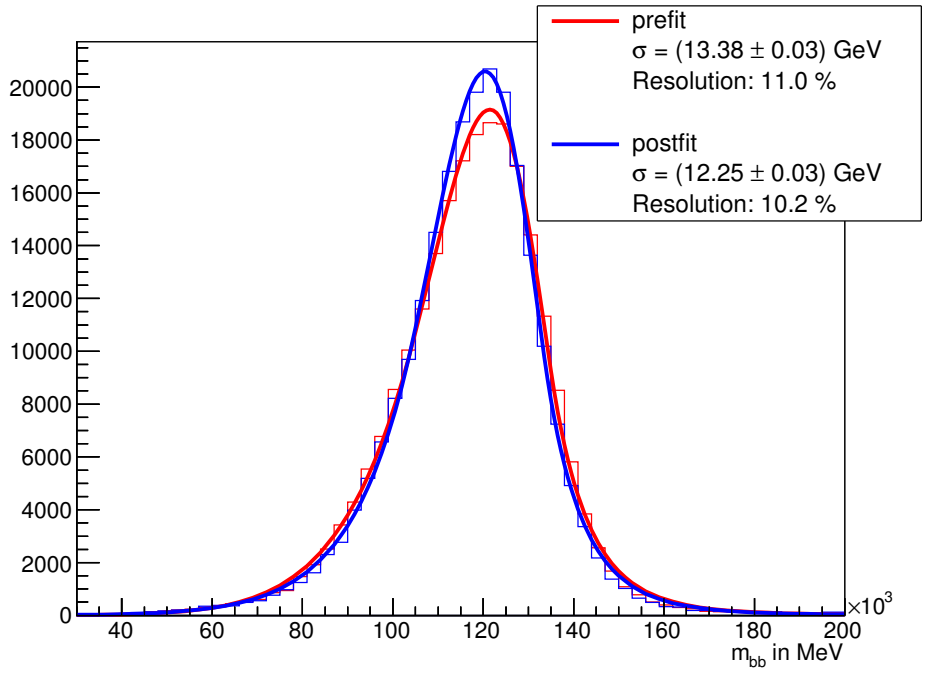


Figure 4.1: $m_{b\bar{b}}$ after implementing the kinematic fit in Overkill without m_{ll} constraint yields an 8 % improvement. If the $m_{ll} = 91.2\,\text{GeV}$ constraint is added, the width is $\sigma = (12.31 \pm 0.03)$ GeV.

4.1 Application on background

So far the kinematic fit has been optimised for the application on the $ZH \to l^+l^-b\bar{b}$ signal samples. In this section its application on the full Monte Carlo simulation with all relevant backgrounds will be discussed to determine the expected significance (see equation 2.3). The measured data will only be used for the rescaling of the background since the expected significance is below one and thus too low to get a meaningful observed significance. Here a very simple rescaling procedure is applied: take the two main backgrounds Z+bb and $t\bar{t}$ and fit their scale factors in the m_{ll} distribution where these backgrounds have a very different shape. Since the signal is so small compared to the statistical errors of the data, the signal region does not need to be excluded for the rescaling. The result can be found in figure 4.2. Only statistical errors are considered since determination of systematic errors is beyond the scope of this thesis.

In figure 4.3 is $m_{b\bar{b}}$ before and after the fit shown. The significances in the signal region $80\,\text{GeV} < m_{b\bar{b}} < 150\,\text{GeV}$ are 0.536 and 0.540 respectively which amounts to an improvement of only $\sim 0.7\,\%$. The reason for this small improvement compared to the much larger improvement in signal resolution is that the Z+bb background is very similar to the signal such that the fit does not help in distinguishing it. On the other hand $t\bar{t}$ is shifted towards lower values, but this also means that events with higher $m_{b\bar{b}}$ are moving into the signal region negating the positive effect.

To understand this effect one has to look at the $P(\chi^2)$ distributions in figure 4.4. The bulk of the $t\bar{t}$ background is gathered around zero and thus is not compatible with the fit constraints at all, explaining the different behaviour in $m_{b\bar{b}}$. This can be exploited with a cut on $P(\chi^2)$ such that the $t\bar{t}$ background will be filtered out. The cut value for the maximal significance is found by filling $(1 - P(\chi^2))$ in the signal region, integrating and then calculating the significance for each bin. The results can be found in table 4.1, showing that the best result of $\sim 17\,\%$ improvement in expected significance is reached when fitting without m_Z constraint.

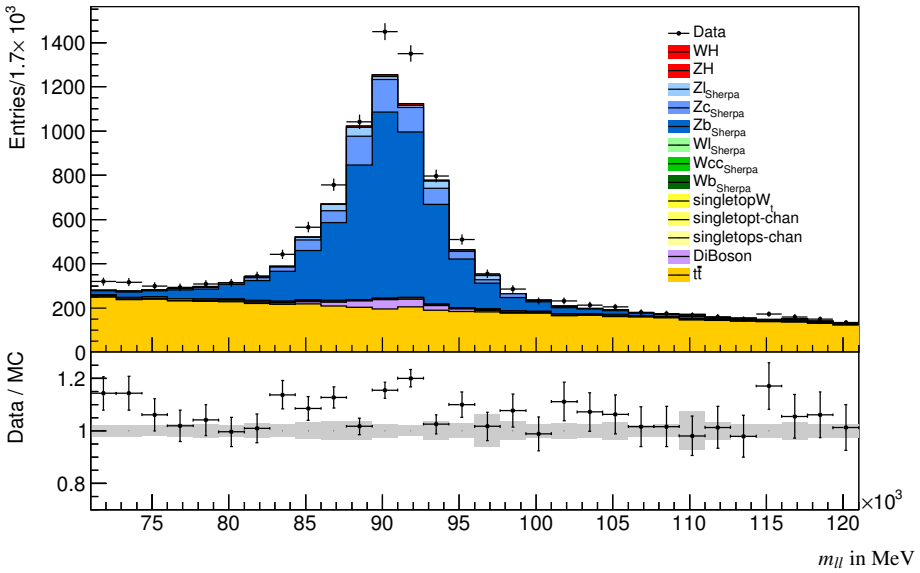
Configuration	exp. sign.	rel. impr.	$P(\chi^2)$ cut
prefit	0.536	-	-
with m_Z constr.	0.540	0.7 %	-
without m_Z constr.	0.540	0.7 %	-
with m_Z constr., with $P(\chi^2)$ cut	0.625	16.6 %	> 0.0005
without m_Z constr., with $P(\chi^2)$ cut	0.629	17.4 %	> 0.017

Table 4.1: Improvements in the expected significance for $ZH \to l^+l^-b\bar{b}$ in the $80\,\text{GeV} < m_{b\bar{b}} < 150\,\text{GeV}$ signal region after the fit. An additional cut on $P(\chi^2)$ is chosen such that the significance is maximised. The difference in the cut values is a consequence of the different distributions in figure 4.4.

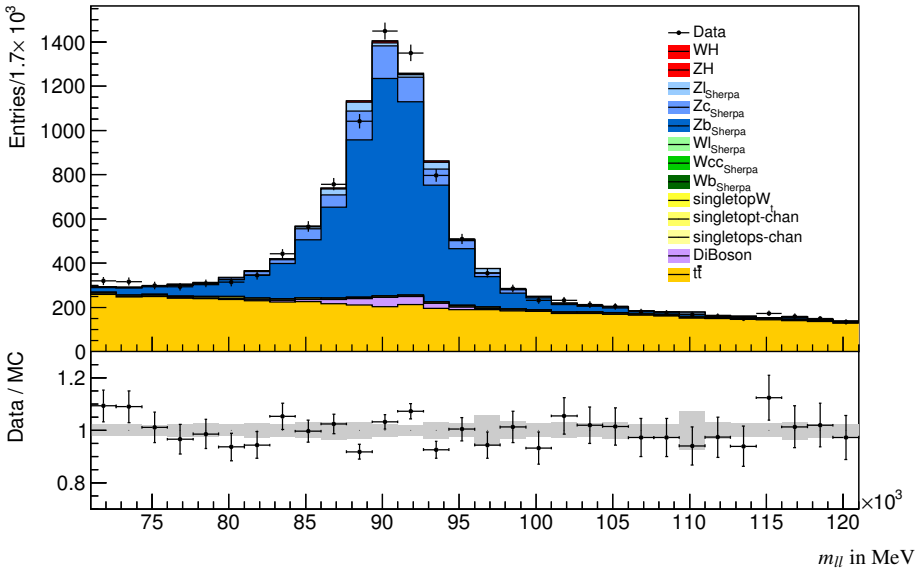
4.1.1 Cross check with diboson

It is helpful to cross check the fit results with the diboson channel - mostly $ZZ \to l^+l^-b\bar{b}$, though some WZ events are included - since it has the same signature as $ZH \to l^+l^-b\bar{b}$. The only difference is the smaller mass of 91.2 GeV which is reflected in a lower $m_{b\bar{b}}$ distribution and p_T values while the cross section is about 5 times bigger.[24] The fit can be used on $Z \to b\bar{b}$ without any modifications since $m_H = 125\,\text{GeV}$ is only used for optimising the error determination of the kinematic variables, which are very similar in this channel. The $m_{b\bar{b}}$ distributions of the ZZ signal sample can be found in figure A.5.

The significances are obtained using the same method as for ZH only that the signal region for the Z boson is defined as $60\,\text{GeV} < m_{b\bar{b}} < 120\,\text{GeV}$. When looking at this region in figure 4.3 the amount

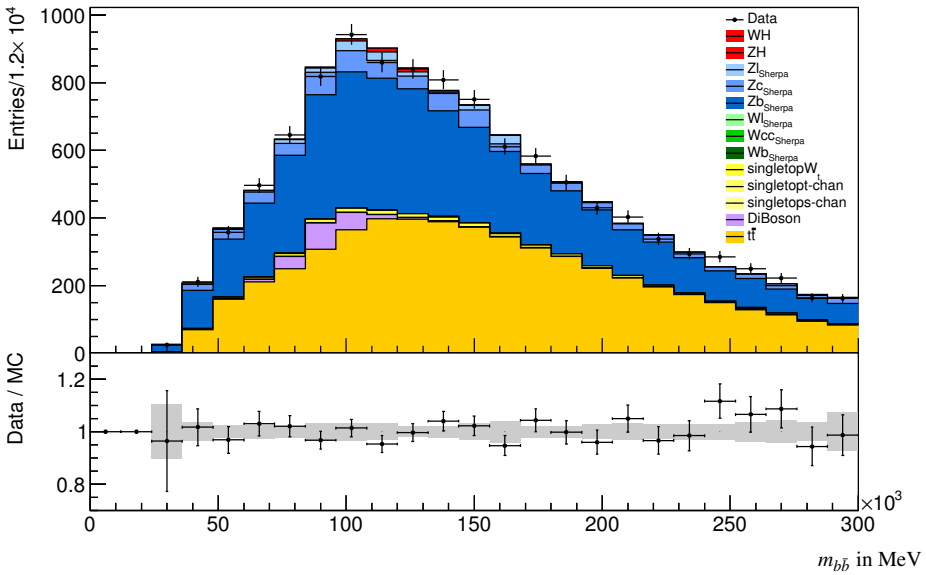


(a) Before rescaling

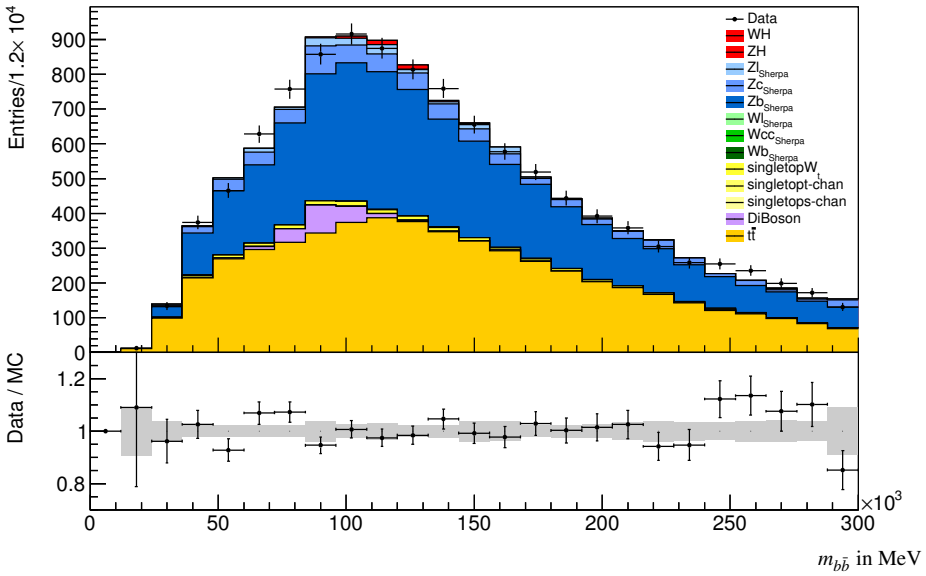


(b) After rescaling

Figure 4.2: The measured data points are used to rescale the main backgrounds Z+bb (Zb in the plot legend) and $t\bar{t}$ in m_{ll}. A fit is performed that finds the scale factors for the selected backgrounds such that the χ^2 between data and Monte Carlo is minimised. The calculated scale factors are: Z+bb: 1.169 ± 0.025 and $t\bar{t}$: 1.039 ± 0.019. If the next biggest background Z+cc would be included into the fit its scale would be 0.2 ± 0.3, which is unlikely to be correct. The fit is performed with MINUIT[29] which is part of ROOT.

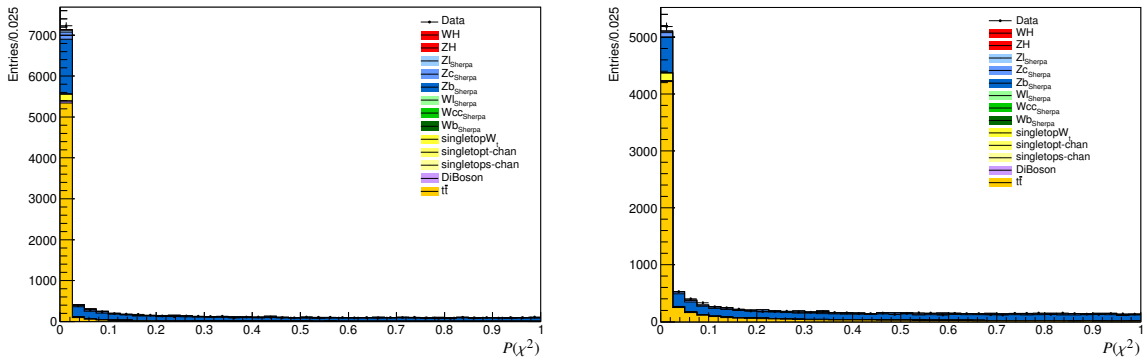


(a) prefit



(b) postfit

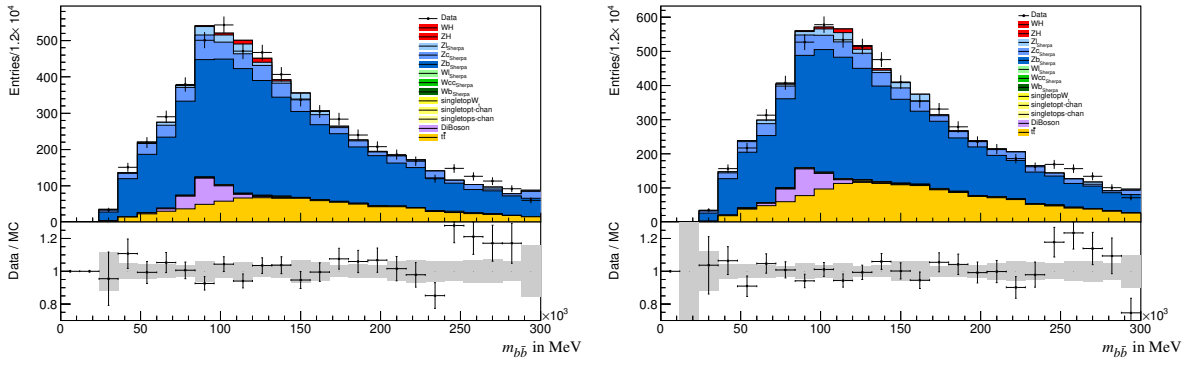
Figure 4.3: The $m_{b\bar{b}}$ distributions of the full Monte Carlo with data. The main difference of the fit is a shift of the $t\bar{t}$ background to lower values.



(a) with m_Z constraint

(b) without m_Z constraint

Figure 4.4: Probabilities of the minimised χ^2 for the kinematic fit on the full Monte Carlo sample and data.



(a) with m_Z constraint

(b) without m_Z constraint

Figure 4.5: The $m_{b\bar{b}}$ distributions after cutting on $P(\chi^2)$ according to the values found in table 4.1. Mainly $t\bar{t}$ is affected by the cut. If using the m_Z constraint, the loss is higher altogether.

of $t\bar{t}$ background is actually higher after the fit, leading to a deterioration of the expected significance. As before the significance can be improved considerably using a $P(\chi^2)$ cut. The results can be found in table 4.2.

The improvement in expected significance with a $P(\chi^2)$ cut is higher than for $ZH \to l^+l^-b\bar{b}$, as is the optimal cut value, since there is a smaller fraction of events gathered around $P(\chi^2) = 0$ in the $ZZ \to l^+l^-b\bar{b}$ channel. The cause of this may be the overall lower energies of the particles involved, leading to a smaller low p_T tail of the b-jets.

Configuration	exp. sign.	rel. impr.	$P(\chi^2)$ cut
prefit	3.09	-	-
with m_Z constr.	3.01	−2.6 %	-
without m_Z constr.	3.01	−2.6 %	-
with m_Z constr., with $P(\chi^2)$ cut	3.68	19.1 %	> 0.004
without m_Z constr., with $P(\chi^2)$ cut	3.79	22.7 %	> 0.07

Table 4.2: Improvements in the expected significance for $ZZ \to l^+l^-b\bar{b}$ in the 60 GeV $< m_{b\bar{b}} <$ 120 GeV signal region after the fit similarly to table 4.1.

4.2 Comparison with maximum likelihood fit

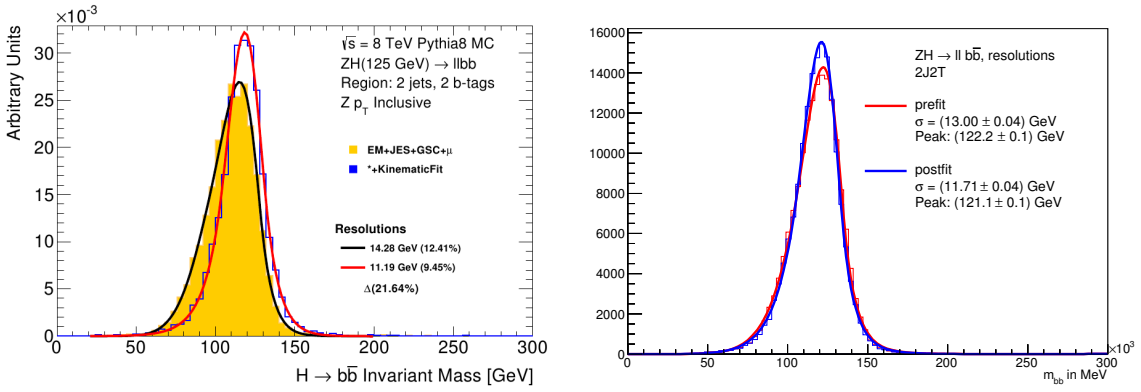
In the ATLAS analysis a maximum likelihood fit is performed in events with 2 jets which can be compared to the kinematic fit in this thesis. It works differently in that the kinematic variables and constraints are expressed as probability distributions which are multiplied to a likelihood function. The parameters for which the function is maximised are the fitted values. The inputs are[30]

- the angles and energies of the leptons and angles of the b-jets, Gaussian distributed probabilities
- the energies of the b-jets with derived asymmetric distributions
- the transverse momentum sum of leptons and jets constrained to zero with a width of 9 GeV
- m_{ll}, constrained to a Z boson Breit-Wigner distribution

The third constraint is equivalent to equations 3.2 and 3.3 with $r = 0$, which is the case for r_jets in 2 jet events. As such the main differences in assumptions are the jet energy distribution and the m_{ll} Breit-Wigner distribution. Although the possibility is mentioned, there is no fit conducted in the 3 jet region.

In figure 4.6 the comparison of the fitting techniques is shown. The improvement for the likelihood fit is about 22 % while the least squares fit yields 9 %. While the two plots are not completely comparable the likelihood fit appears to give a higher resolution improvement. The advantage of that model is primarily in the asymmetric p_T^b shapes and the Breit-Wigner constraint for m_{ll}.

The likelihood fit is conducted on the ZH signal and the Z+jets background to evaluate the change in significance in the signal region of 80 GeV $< m_{b\bar{b}} <$ 150 GeV. The improvement achieved is then $\approx$ 4 %. The same setup for the least squares fit results in an improvement of 2.3 % both for the fit with and without the m_Z constraint. This is roughly consistent with the mass resolution comparison. With the likelihood fit events tend to migrate to higher values in the $m_{b\bar{b}}$ spectrum while for the least squares fit the exact opposite is the case.



(a) Maximum likelihood fit[17]

(b) Kinematic least squares fit, no m_{ll} constraint

Figure 4.6: Comparison with the maximum likelihood fit used by the ATLAS analysis group in the 2 jet region. The used samples are obviously not identical, however there are no information that a different event selection was used. The least squares fit shifts the peak in the wrong direction but it is still closer to the true value than after the likelihood fit.

The likelihood fit yields the better result in direct comparison but it is unclear how it behaves in events with 3 or more jets. There is no exploitation of the likelihood function like the $P(\chi^2)$ cut, as such the full potential for a cut based analysis is hard to compare.

CHAPTER 5

Summary

In this thesis a least-squares kinematic fit using the transverse momentum conservation as a constraint was successfully applied in the $ZH \to l^+l^-b\bar{b}$ decay channel to improve the Higgs mass resolution of a Higgs with $m_H = 125\,\text{GeV}$. The problem of badly measurable initial state radiation was solved by introducing the Remnant Vector as a fifth fit vector, which was chosen as the sum of non-tagged jets as this definition proved to give the best fit result. A χ^2 probability test and a pull test revealed that an optional hard constraint on the invariant mass of the leptons to 91.2 GeV was inadequate, since the true m_{ll} is Breit-Wigner distributed, and lead to a slightly worse fit result than leaving it out completely. An improvement of 8 % was achieved in the Higgs mass resolution.

The application of the fit on the full Signal+Background samples showed only a minimal improvement of 0.7 % in expected significance since the background in the signal region does not get smaller. A single cut on the χ^2 probability $P(\chi^2)$, cutting away most of the $t\bar{t}$ background, was able to give an improvement of 17 % in expected significance. A cross check on the diboson background (mainly $ZZ \to l^+l^-b\bar{b}$), which is very similar to $ZH \to l^+l^-b\bar{b}$ apart from the lower Z mass, even showed a deterioration of the significance since more background events wandered in the signal region. Employing again a $P(\chi^2)$ cut however gave an improvement of 23 % in expected significance.

The least-squares fit was compared to a maximum likelihood fit used in the standard ATLAS analysis in $ZH \to l^+l^-b\bar{b}$. An exact comparison between the two methods was not possible as the exact event selection of the cited results are unknown. In comparison, as far as that was possible, the maximum likelihood fit seems to yield a somewhat higher improvement in the Higgs mass resolution as well as the expected significance without any additional cut. However it is only used in the 2 jet region whereas the least squares fit improves the resolution and significance for any number of additional jets. It is also unknown if the significance could be improved in a similar way as with the $P(\chi^2)$ cut.

APPENDIX A

Appendix

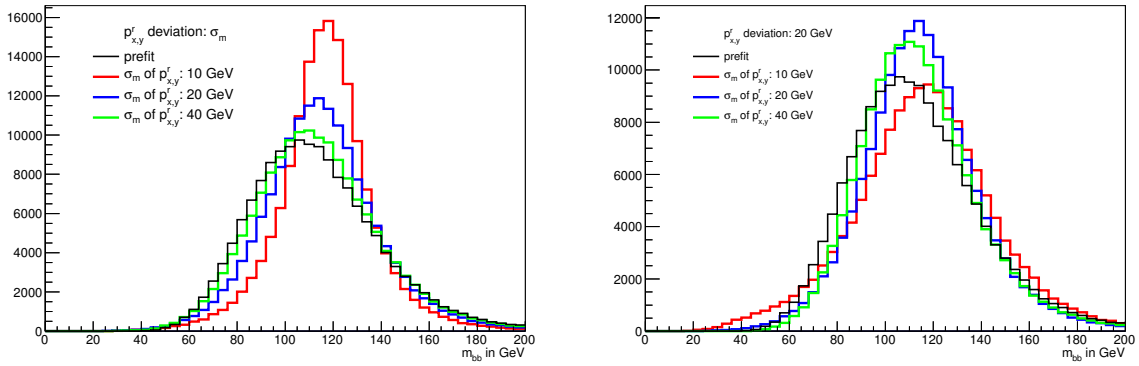
A.1 KinFitter validation

Before applying the kinematic fit on "real" Monte Carlo simulated data, it makes sense to test how the KinFitter package behaves for the $ZH \to l^+l^-b\bar{b}$ setup. Events are created in a very simplified model which is trying to mimic the basics of the decay channel. It is not comparable to the physics and detector inspired MC samples that are used to tune the fit for the real data, but rather uses similar kinematic properties while maintaining complete control over the connection between the true and measured values. This is done for two reasons: to test if the KinFitter works as intended under ideal circumstances, and to determine its sensitivity towards the Remnant Vector. The model consists of a pure $ZH \to l^+l^-b\bar{b}$ decay with an added random p_T to simulate ISR. The RMS values for the truth distribution of the Remnant Vector and deviation distributions of all measured kinematic variables from their truth value are chosen to roughly align with the framework used in the real analysis.

A Z boson with a Breit-Wigner distributed mass of 91.2 GeV and $\Gamma = 2.5$ GeV and a H boson with a delta distributed mass of 120 GeV are created in the centre of mass system. Then a random boost in z direction is applied to both. Now they are decayed into leptons and b-quarks respectively. p_x and p_y of the Remnant Vector are created Gaussian around 0 with $\sigma = 42$ GeV, thus the true $E_\mathrm{T}^\mathrm{miss}$ is non-zero.

With the truth information available, the measured data can now be derived. The true values p_T, η, ϕ of leptons and b-quarks as well as p_x, p_y of the Remnant Vector are smeared and taken as measured particles. Additionally, to mimic the b-jet distribution better, their p_T value is lowered by 10 GeV to account for generally lower measured values. While the shape is still wrong, it should be sufficient to study the effect of an asymmetric deviation of measured jets from their true values on the kinematic fit. A cut of 25 GeV on lepton p_T and 20 GeV on jet p_T is applied. After that the fit is performed. The errors used are equal to the deviations that smeared the variables, thus are the optimal error values without using event level truth information.

This is done for different deviations of the Remnant Vector to imitate good and bad approximations for its true value. In figure A.1a is shown that a better r reconstruction significantly improves the fit result. Given these results it becomes clear that the Remnant Vector has to be chosen very carefully to minimise its deviation and ensure a good fit result. Additionally, the fit is repeated for the same Remnant Vector but different σ_m values. In figure A.1b is shown that an optimal error value is very important.



(a) Fit with different accuracies of the Remnant Vector

(b) Fit with the same Remnant Vector and different error input

Figure A.1: $m_{b\bar{b}}$ of a simplified model to test the behaviour of the kinematic fit. In (a) is shown that a better reconstructed Remnant Vector significantly improves the fit result. In (b) is shown that the fit result is best with optimal error parametrisation.

A.2 Additional figures and tables

p_{T}	> 20	$20 - 30$	$30 - 40$	$40 - 50$	$50 - 60$	> 60
σ_{m} of $p_{x,y}$	10.8	14.3	15.2	16.1	16.2	17.5

(a) σ_{m} for $p^{r}_{x,y}$ in GeV

$p_{\mathrm{T}} \setminus \eta$	> 0.6	$0.6 - 1.3$	> 1.3
< 50	7.0	6.9	6.9
$50 - 70$	8.1	8.2	7.9
> 70	11.5	11.4	10.7

(b) σ_{m} for p^{b}_{T} in GeV

$p_{\mathrm{T}} \setminus \eta$	> 0.6	$0.6 - 1.3$	> 1.3
< 50	0.0643	0.0648	0.0604
$50 - 70$	0.0472	0.0475	0.0460
> 70	0.0313	0.0306	0.0304

(c) σ_{m} for ϕ^{b}

$p_{\mathrm{T}} \setminus \eta$	> 0.5	$0.5 - 1.2$	> 1.2
< 30	1.02	1.13	1.28
$30 - 60$	1.19	1.34	1.68
> 60	2.10	2.31	2.81

(d) σ_{m} for p^{l}_{T} in GeV

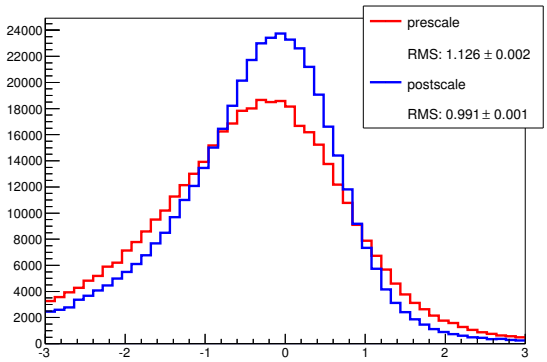
$p_{\mathrm{T}} \setminus \eta$	> 0.5	$0.5 - 1.2$	> 1.2
< 30	0.000581	0.000441	0.000448
$30 - 60$	0.000550	0.000406	0.000374
> 60	0.000516	0.000378	0.000345

(e) σ_{m} for η^{l}

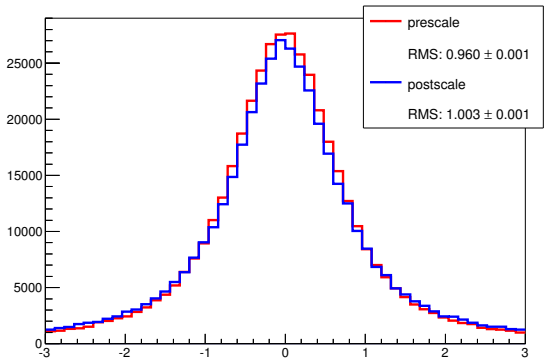
$p_{\mathrm{T}} \setminus \eta$	> 0.5	$0.5 - 1.2$	> 1.2
< 30	0.000217	0.000236	0.000294
$30 - 60$	0.000180	0.000189	0.000224
> 60	0.000156	0.000165	0.000191

(f) σ_{m} for ϕ^{l}

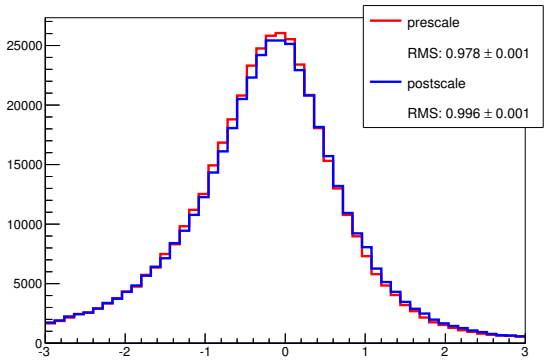
Table A.1: Error values determined in the Overkill framework for Remnant Vector, lepton and jet variables segmented into p_{T} and η regions. p_{T} categories are in GeV.



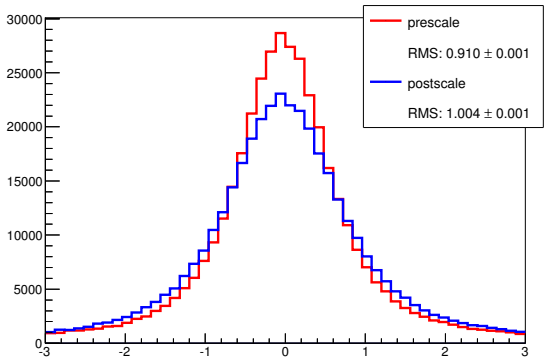
(a) p_{T}^{b}



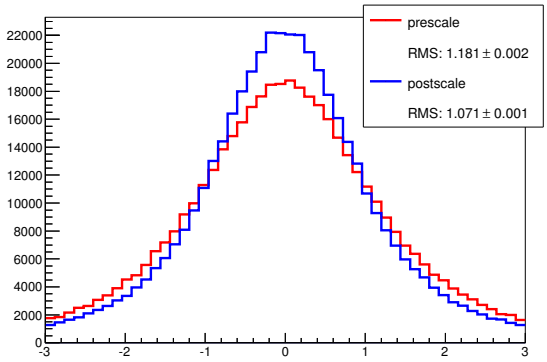
(b) ϕ^{b}



(c) p_{T}^{l}



(d) ϕ^{l}



(e) $p_{x,y}^{r}$

Figure A.2: Pulls g_{tr} for the kinematic variables when fitting without m_Z constraint.

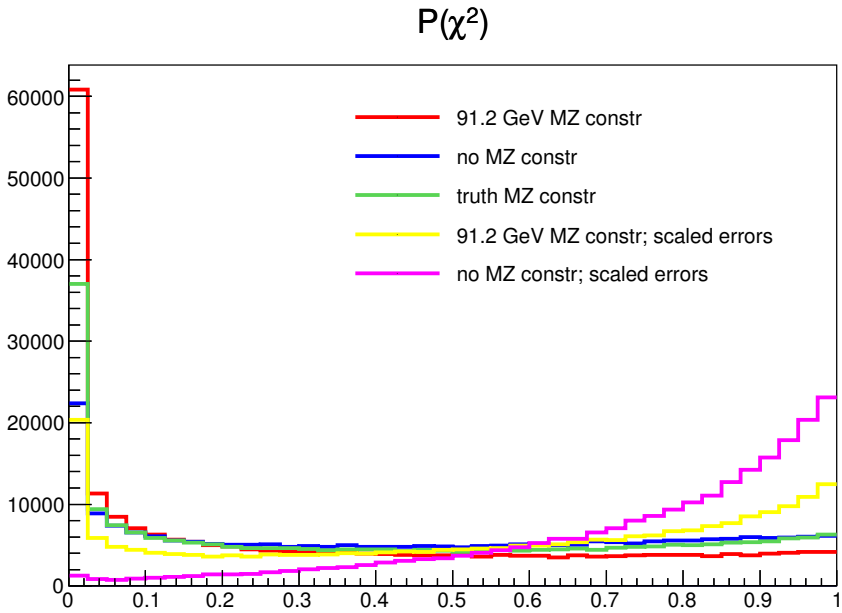


Figure A.3: $P(\chi^2)$ for the kinematic fit with different m_Z constraints and different scaled σ_m values. The scale factors can be found in table 3.2. A factor of 0.8 is applied to all errors.

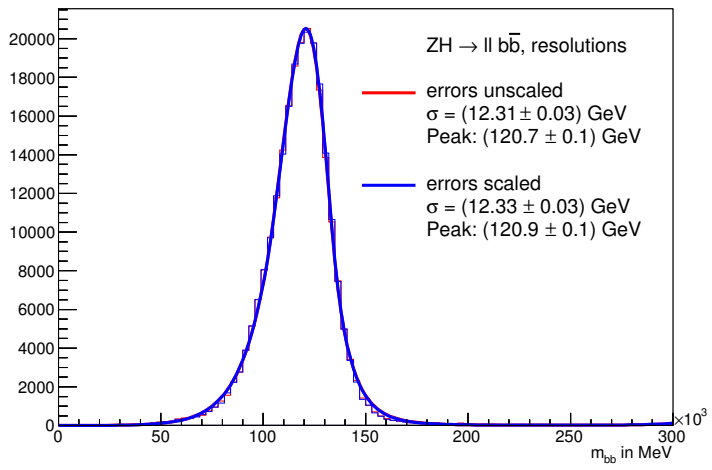


Figure A.4: The kinematic fit applied with and without rescaled errors from table 3.2. The width gets worse while the peak position improves. The resolution is approximately the same.

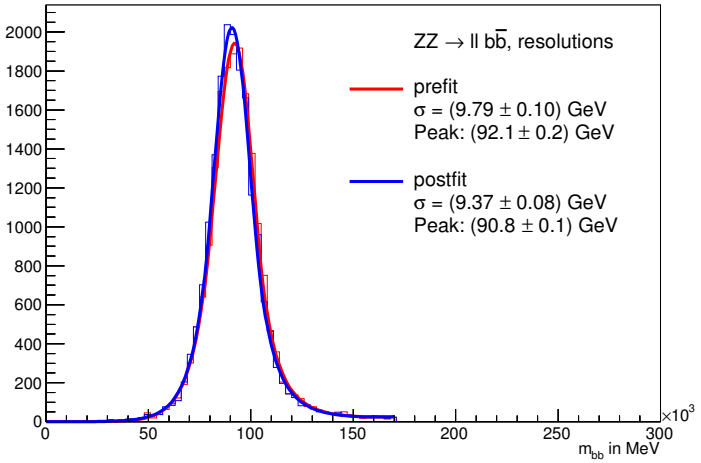


Figure A.5: The kinematic fit applied to $ZZ \to l^+l^-b\bar{b}$. The width is improved with the kinematic fit. Contrary to $ZH \to l^+l^-b\bar{b}$, the mass peak is higher than the true value $m_Z = 91.2\,\text{GeV}$ as the jet corrections are optimised for the higher Higgs mass, it is shifted closer through the kinematic fit.

Bibliography

[1] P. Avery, *Vertexing and Kinematic Fitting, Part I: Basic Theory*, Aug. 1998, URL: http://www.phys.ufl.edu/~avery/fitting/kinfit_talk1.pdf.

[2] L. Evans and P. Bryant, "The CERN Large Hadron Collider: Accelerator and Experiments", *Journal of instrumentation* 3.S08001 (2008).

[3] ATLAS Collaboration, *ATLAS Experiment - Photos*, URL: http://www.atlas.ch/photos/.

[4] G. Aad et al.,
"Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics" (2009), arXiv: 0901.0512 [hep-ex].

[5] A. Bukin, "Fitting function for asymmetric peaks", 2007, arXiv: 0711.4449v2.

[6] R. Cahn and G. Goldhaber, *The Experimental Foundations of Particle Physics*, 2nd ed., Cambridge University Press, 2009.

[7] M. Cacciari, G. P. Salam and G. Soyez, "The Anti-k_t jet clustering algorithm", *JHEP* 0804 (2008) 063, DOI: 10.1088/1126-6708/2008/04/063, arXiv: 0802.1189 [hep-ph].

[8] T. Aaltonen et al., "Studying the Underlying Event in Drell-Yan and High Transverse Momentum Jet Production at the Tevatron", *Phys.Rev.* D82 (2010) 034001, DOI: 10.1103/PhysRevD.82.034001, arXiv: 1003.3146 [hep-ex].

[9] ATLAS Collaboration, *LHC Higgs Cross Section Working Group*, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[10] T. Sjostrand, "Monte Carlo Generators" (2006) 51–74, arXiv: hep-ph/0611247 [hep-ph].

[11] G. Cowan and E. Gross, *Discussion and guidelines on significance*, 2009, URL: https://twiki.cern.ch/twiki/pub/AtlasProtected/ATLASStatisticsFAQ/SigRec_v1.pdf.

[12] ATLAS Collaboration, *The ATLAS Computing Model*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasComputing/WorkBookComputingModel.

[13] P. I. Bonn, *Overkill Overview - Overkill wiki page*, URL: https://wiki.physik.uni-bonn.de/atlas/private/index.php/Overkill_Overview.

[14] ATLAS Collaboration, "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector", CERN, 2014, arXiv: 1409.6212.

[15] V. Kostyukhin, *First bb-mass correction*, URL: https://indico.cern.ch/getFile.py/access?contribId=1&resId=1&materialId=slides&confId=213579.

[16] ATLAS Collaboration, *B-Tagging Benchmarks*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTaggingBenchmarks.

[17] F. Ahmadov et al., "Supporting Document for the Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector", tech. rep. ATL-COM-PHYS-2014-051, CERN, Jan. 2014.

[18] J. E. Sundermann and T. Göpfert, *KinFitter – A Kinematic Fit with Constraints*, URL: http://www.iktp.tu-dresden.de/~goepfert/KinFitter.pdf.

[19] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Online Edition, 2012, ISBN: 978-3-935702-66-9, URL: http://www-library.desy.de/preparch/books/BloLoBuch.pdf.

[20] J. E. Sundermann and T. Göpfert, *KinFitter source code*, URL: https://github.com/goepfert/KinFitter.

[21] R. J. Barlow, *Statistics, A Guide to the Use of Statistical Methods in the Physical Sciences*, 1st ed., John Wiley & Sons Ltd., 1997.

[22] A. G. Frodesen, O. Skjeggestad and H. Tøfte, *Probability and Statistics in Particle Physics*, Universitetsforlaget, 1979, ISBN: 82-00-01906-3.

[23] L. Demortier and L. Lyons, "Everything you always wanted to know about pulls", CDF/ANAL/PUBLIC/5776, Aug. 2002, URL: http://physics.rockefeller.edu/luc/technical_reports/cdf5776_pulls.pdf.

[24] Particle Data Group, K. A. Olive et al., "Review of Particle Physics", *Chinese Physics C* 38 (2014) 090001, URL: http://pdg.lbl.gov.

[25] T. A. collaboration, "Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data" (2013).

[26] G. Unal, "Electron and Photon energy resolutions and their uncertainties for run-1 calibration", tech. rep. ATL-COM-PHYS-2013-1651, CERN, Dec. 2013.

[27] E. Schopf, "Improvement of the Higgs Mass Resolution in the Decay Channel $H \to b\bar{b}$ in Associated Production with a W Boson at the ATLAS Detector Using Multivariate Regression", Universität Bonn, 2013.

[28] A. Hoecker et al., *TMVA 4.2*, URL: http://tmva.sourceforge.net.

[29] F. James, *MINUIT Reference Manual*, URL: http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[30] A. Buzatu et al., "Invariant Mass Studies for the $H \to b\bar{b}$ Measurements for LHCP", CERN, 2013.

List of Figures

List of Tables