Search for new physics in multilepton final-states using multivariate techniques

- Ask Emil Løvschall-Jensen

Academic advisor - Associate Professor Jørgen Beck Hansen

This thesis has been submitted to the PhD School of The faculty of Science, University of Copenhagen

Ask Emil Løvschall-Jensen
ask.emil.jensen@gmail.com
askaboutphysics.com

Supervised by Ass. Prof. Jørgen Beck Hansen

Abstract

A novel use of multivariate methods has been developed for searches for new physics in high-energy particle physics. Through the use of Principal Component Analysis a range of input kinematic and event-observables are transformed into linear uncorrelated *principal components*. The resulting *principal components* are ranked by variance associated to the degree of information they contain. The potential usage of this transformation on both data and model expectations is presented and it is shown how the highest ranked *principal component* can be used in a composite likelihood fit of data to model expectation, combining much more of the available information than any one original observable could provide.

The independence of the *principal components* is tested using the *Mutual Information* measure. *Principal components* found to be independent are combined to add an even larger amount of information to the fit and improve separation power to new physics. This is applied in both tests of the Standard Model and a search for new physics. As a benchmark model upper limits are set on the Type III seesaw mechanism.

The method is shown to hold significant potential for use in both searches for new physics and in testing the Standard Model.

Resumé

En ny anvendelse af multivariate metoder er blevet udviklet til at søge efter nye fysik fænomener i højenergi partikelfysik. Ved at anvende Principal Component Analysis kan en række kinematiske og begivenhedsbaserede observabler transformeres med lineært ukorrelerede *principielle bestanddele* til følge. Disse *principielle bestanddele* bliver rangeret efter varians forbundet med mængden af information de indeholder. Anvendelsesmulighederne på data og model-forventninger bliver præsenteret og det vises hvordan den højest rangerende *principielle bestanddel* kan bruges i et *likelihood-fit* af model-forventning til data og på den måde inkludere en langt større del af den oprindelige information end det kunne opnås med en enkelt af de oprindelige observabler.

Uafhængigheden de *principielle bestanddele* imellem udregnes vha. *Mutual Information* målet. De bestanddele, der er uafhængige, kombineres så en endnu større grad af information opnås i likelihood-fittet. Dette giver ydermere mulighed for bedre at diskriminere mellem forventning og bidrag fra ny fysik. Metoden anvendes i både test af partikelfysikkens *Standardmodel* og til at lede efter ny fysik. *Type III seesaw* mekanisme bruges som benchmarkmodel, hvor der sættes øvre grænser på produktionstværsnittet.
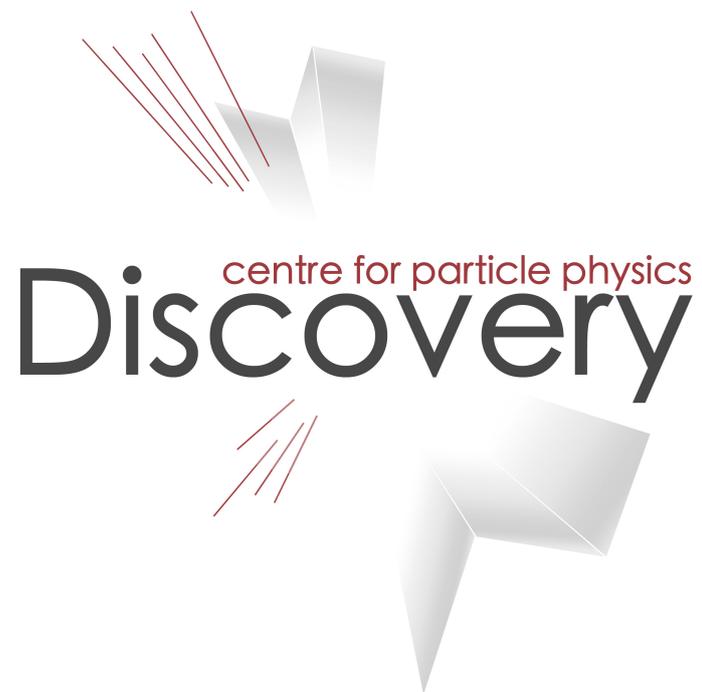
Acknowledgement

Throughout the work on this thesis and my work at the Niels Bohr Institute (NBI) I have been given the opportunity to develop a novel method for fitting Standard Model expectations to data. I owe a huge thanks to my supervisor and sparring partner over the last years, Associated Professor Jørgen Beck Hansen for being shown that confidence.

The work presented is to a very large extent my own but the ideas have sparked and evolved through discussions with both Jørgen Beck Hansen and my office mates Kristian Grekersen and Peter Rosendahl and I would like to thank all three of you for this.

It has been a pleasure to be part of the HEP group at NBI and to work with such competent and dedicated people.

Disclaimer

Some of the plots and tables in this thesis are based on data recorded at the ATLAS experiment during 2012. The plots represent the writer's interpretation of the data and are not officially approved by ATLAS.

centre for particle physics

Discovery

Contents

Foreword

Through my PhD I have had the change to get involved in several exciting projects. In the initial phase Kristian Grekersen, Jørgen Beck Hansen and I developed an extension to the analysis framework SFrame [1] that allowed for sequential addition of object selectors and event tools. The code including documentation is openly available at:
http://sourceforge.net/projects/cyclesequencer/reviews.

During all three years I have been involved in the work on the *Transition Radiation Tracker* in ATLAS; with performance studies and lately with the development of a common data format to help improve and speed up the many performance studies. The work has been documented in the relevant groups but as it is not a direct part of what I consider my main work on this thesis I have not included it here. The following will present my work on the usage of multivariate methods in searches for new physics in three-lepton final states using data from 2012 recorded at the ATLAS experiment.

Challenging the current views

The Standard Model of particle physics is one of the most successful models in physics if not all of science. With the discovery in 2012 [2, 3] of a particle consistent with the long sought Higgs, the Standard Model became consistent and even with the degree of precision it has been tested to, no significant signs of anything beyond it have been observed. It might seem futile to further test it but this is by no means the case. The Standard Model is an effective model and not a fundamental theory and it has several deficiencies. Furthermore there is theoretical motivation for physics beyond the Standard Model to occur in data from the LHC.

To search for new physics, data is compared to expectations based on simulation of the Standard Model processes. To be able to conclude anything selections are performed on the data to get phase-spaces with predictable background contribution. The phase-spaces are typically tested in one or a few select observables. The approach is often to look for new physics with selections specific to new models but a model-independent approach can also be taken. The result has so far been strict limits on many types of new physics in the available energy ranges.

To really test the Standard Model it would be ideal to compare the measured data to the actual theoretical predictions it gives. This would mean comparing the event kinematics directly to the matrix elements, that constitute the core of the simulation. This however does not give a meaningful result. Collision events are measured through detectors, that per construction have limited resolution and all event-kinematics have to be reconstructed and particles identified using sophisticated tracking and energy reconstruction software. The reconstruction leads to overall smearing and the Standard Model expectation should be simulated accordingly.

One way to fully compare an event to expectation is by using a large set of observables to simultaneously look for discrepancies. This has not been done to large extent before, as fitting data to expectation in many observables demands that either all observables are independent or that their correlations are known.

This thesis introduces a novel method where Principal Component Analysis is used to map a list of observables into linearly uncorrelated variables by maximising the signal-to-noise ratio. Two approaches are shown to be possible using Principal Component Analysis that

allow for a significant addition of information to the fits.

Theory

1.1 Our current view - the Standard Model of Particle physics

The discovery of the Higgs particle was the last brick needed for *the Standard Model* of particle physics to be consistent. It is perhaps one of the most successful scientific models or effective theories to this date and it explains the dynamics of our universe at the most fundamental scale with extreme precision. It describes the existence of a set of fundamental matter particles; the fermions and three of the four fundamental forces of nature with associated force carrying particles - the gauge bosons. In the following a brief overview of the central points are presented. It is an attempt at making the absolute minimal introduction for the reader already familiar with quantum mechanics and field theory as the theoretical foundation of the Standard Model is not part of this thesis. The interested reader is referred to the excellent textbook descriptions in [4], [5] or [6].

1.1.1 Quantum field description

The Standard Model is based on a quantum field theory subject to $SU(3)_C$ x $SU(2)_L$ x $U(1)_Y$ gauge symmetries. Central to it is the electroweak theory and the theory of strong interactions, the quantum chromo dynamics. The electroweak theory came to be through the work of Sheldon Glashow, Steven Weinberg and Abdus Salam in the '60's. All three were rewarded with a Nobel prize for the work in '79. The theory combines the phenomena of electromagnetism and the weak interactions under one description. Quantum chromo dynamics is the description of the strong force governing the structure and constituents of the atomic nuclei.

1.1.2 Gauge invariance

In the framework of quantum field theory, particles can, contrary to the point-particles of classical mechanics, be described as the quantised excitations of fields. The description of fields is most commonly recognised from the description of light: the photon exhibiting both wave- and particle-properties.

In classical mechanics the fundamental quantity is the action, S. It contains all the information needed to determine the dynamics and kinematics of a system, and is found by time integrating the Lagrangian. The Lagrangian is here the spatial integral of the Lagrangian

density, $\mathcal{L}$.

$$S = \int L \, dt = \int \mathcal{L} \, d^4 x = \int (T - V) \, dt, \tag{1.1}$$

where T is kinetic energy and V is potential energy of the system. The term *gauge* refers to an excess degree of freedom in the Lagrangian. Transformation between different *gauges* form a symmetry group called the gauge group of the theory. The transformations are called *gauge transformations*. Basically gauge invariance means that if the physical predictions of a theory remain unaltered by a local or global transformation, then the theory is gauge invariant. A gauge invariant Lagrangian is thus invariant under

$$\psi(x) \to e^{i\alpha} \psi(x), \tag{1.2}$$

where ψ is an arbitrary field and α its phase. α is unmeasurable and can be chosen arbitrarily but as soon as it is fixed, it is specified for all points in space-time and it forms a global gauge transformation. If α is dependant on space time e.g. $\alpha(\mathbf{x})$, it forms a local gauge symmetry.

Principle of least action

In classical mechanics a system making a transition from one state to another does so along the path in configuration space, where the action is a minimum. This principle is called the principle of least action. In other words, the classical system will always take the shortest *path* in space-time. The extremum is found by varying the action with regard to a field like the above $\psi(x)$, with the demand, that these variations vanish for a given set of boundary conditions, that correspond to the inherent physics. This gives the Euler-Lagrange version of the equation of motions for that field $\psi(x)$. This principle is not necessarily true when moving to quantum mechanics. The classical path is only one of the paths, and in principle all other paths are allowed. The classical path is however often the dominant path but those close to it, e.g. *quantum fluctuations* can influence the results significantly. We can probe these small fluctuations and get a new form of our Lagrangian.

The dynamics of a fermion field like the electron, $\psi(\mathbf{x})$ are expressed by the Dirac Lagrangian[1] :

$$\begin{aligned}
\mathcal{L}_{Dirac} &= \bar{\psi}(\mathbf{x})(i\gamma^{\mu}\partial_{\mu} - m)\psi(\mathbf{x}) \\
&= \underbrace{i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi}_{\text{kinetic part}} - \underbrace{m\bar{\psi}\psi}_{\text{mass term}}
\end{aligned} \tag{1.3}$$

It is interesting to try to make a gauge transformation on this equation. The partial derivative will transform as

$$\partial_{\mu} \to D_{\mu} = \partial_{\mu} - iqA_{\mu}, \tag{1.4}$$

where A_{μ} is the gauge field. D_{μ} is the referred to as the gauge covariant derivative. This must, due to gauge invariance, transform along with $\psi(\mathbf{x})$ by a phase transformation (see e.g. [4] page 482)

$$\psi(\mathbf{x}) \to e^{i\alpha(\mathbf{x})}\psi(\mathbf{x}) \qquad \text{and} \qquad A_{\mu}(\mathbf{x}) \to A_{\mu}(\mathbf{x}) + \frac{1}{e}\partial_{\mu}\alpha(\mathbf{x}), \tag{1.5}$$

---

[1]This is the Lagrangian density, but as is customary in the field's notation, it is henceforth referred to as the Lagrangian.

Here e is the elementary charge. When this is done, the Dirac Lagrangian changes form and can be written as

$$\mathcal{L} = \bar{\psi}(\mathbf{x})(i\gamma^\mu \partial_\mu - m)\psi(\mathbf{x}) - \mathcal{L}_{int}, \tag{1.6}$$

where $\mathcal{L}_{int}$ is the interaction part of the Lagrangian given by

$$\mathcal{L}_{int} = -e\bar{\psi}(\mathbf{x})\gamma^\mu \psi(\mathbf{x})A_\mu \tag{1.7}$$

The field A_μ can be interpreted as the photon field, but it cannot propagate in its current form, as it has no kinematic degree of freedom. To give the field the ability to propagate as desired, we impose terms containing first order derivatives in time to our Lagrangian. By requiring only gauge invariance this gives a series of terms which is unsatisfactory, but by requiring also P parity symmetry, we are left with one allowed term [4]

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.8}$$

for the electromagnetic field strength tensor

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu \tag{1.9}$$

The addition of a mass term is not allowed by the demand of gauge invariance so this is the final form of the quantum electrodynamic Lagrangian given in full by

$$\mathcal{L}_{QED} = \bar{\psi}(\mathbf{x})(i\gamma^\mu D_\mu - m)\psi(\mathbf{x}) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.10}$$

QED embodies the imposition of a U(1) gauge symmetry on a field theory of fermionic fields. The gauge field is the electromagnetic field and the symmetry of the group demands electric charge conservation and, as it is part of the Standard Model, it means charge must always be conserved in particle interactions.

The mass term here is the fermion mass term introduced by the Dirac Lagrangian. It has not been explained and is indeed put in by hand without explanation of its origin.

1.1.3   The electroweak sector

The work presented in this thesis will concern three-lepton final states to which the electroweak sector is of special relevance.

The electroweak part of the Lagrangian is formed by interactions that have $SU(2)_L \times U(1)_Y$ structure. This is broken into the $U(1)_{EM}$ symmetry by electroweak symmetry breaking as will be explained. Y is weak hypercharge, I_W is weak isospin and I_W^3 its third component. They are related to the electric charge by

$$Q = \frac{Y}{2} + I_W^3 \tag{1.11}$$

With respect to the weak isospin, fermions can be arranged in left-handed doublets like leptons and neutrinos or up-type and down-type quarks. Right-handed fermions are invariant under

rotations in weak isospin space, which means they do not possess weak-isospin, and are therefore singlets under $SU(2)_L$.[2] They both, however, possess weak hypercharge. This equation is central for the understanding of the Standard Model interactions. Weak hypercharge is defined as follows

$$Y \equiv B + X, \tag{1.12}$$

where B is the quantum number called baryon number and X represents a set of quantum numbers for each type of quark. There is a similar quantum number for the leptons demanding conservation of electron, muon and tauon numbers e.g. $L_e \equiv N(e^-) - N(e^+) + N(\nu_e) - N(\bar{\nu}_e)$. Along with the lepton number [7], these numbers govern the possible decays and production mechanisms of all particles.

The bosons of the electroweak interactions stems from the description of the $SU(2)_L \times U(1)_Y$ by a massless isotriplet $W_\mu^{1,2,3}$, and a massless isosinglet, B_μ. Special unitary groups like the SU(2) are described by the so called Lie algebra [8]. The group has generators, T_a that are proportional to the Pauli matrices σ_a:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.13}$$

Because the Pauli matrices do not commute with each other the W_μ field tensors have, in addition to their kinetic energy, a contribution from their self-interaction [9]

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon^{abc}W_\mu^b W_\nu^c, \tag{1.14}$$

and ϵ^{abc} are the structure constants of SU(2) and g is the coupling constant of the left handed fermions weak isospin to $W_\mu^{1,2,3}$. It is related to g' - the coupling of the weak hypercharge to B_μ via the weak mixing angle θ_W and the elementary charge, e, by

$$g \cdot \sin\theta_W = g' \cos\theta_W = e. \tag{1.15}$$

For the B_μ field there exist no self-interactions and the field tensor is described by

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.16}$$

Electroweak Lagrangian

Using the U(1) gauge field B_μ and SU(2) gauge fields W_μ^a (a=1,2,3) the electroweak Lagrangian is defined as:

$$\mathcal{L}_{gauge}^{EW} = \underbrace{-\frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}}_{W^\pm, Z, \gamma \text{ field part}}. \tag{1.17}$$

---

[2]The term handedness is used due to transformation properties of the left and right-handed fermions. Left-handed massless fermions will have spin opposite the direction of momentum whereas right-handed massless fermions will have spin in the same direction. Massive fermions will transform under rotation θ and boost β as (see e.g. [4] page 44): $\psi_L \rightarrow (\mathbf{1} - i\theta \cdot 0.5\sigma - \beta \cdot 0.5\sigma)\psi_L$ and for right-handed the sign in front of the boost will change.

The $W^\pm, Z^0$ and γ bosons are related through the transformations:

$$A_\mu = cos\theta_W B_\mu + sin\theta_W W_\mu^3 \tag{1.18}$$

$$Z_\mu = sin\theta_W B_\mu - cos\theta_W W_\mu^3$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$$

$$\tag{1.19}$$

The W and Z field can achieve mass by including the terms:

$$m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu, \tag{1.20}$$

but this would explicitly break the invariance under the U(1) gauge transformation in equation 1.5 resulting in a non-normalisable model.

The interaction between the gauge boson and fermions can be generally described as:

$$\mathcal{L}_f = i\bar{\psi}_{L,i}\slashed{D}\psi_{L,i} + \bar{\psi}_{R,i}\slashed{D}\psi_{R,i},$$

where the Dirac notation $\slashed{D} \equiv \gamma^\mu \mathcal{D}_\mu$ has been introduced. The Lagrangian will contain terms from both the quarks and leptons and the subscript i will run over the three generations of fermions. For the leptons this will be the left-handed doublet and right handed singlet electron fields.

1.1.4 Spontaneous symmetry breaking

To generate masses a scalar field ϕ is added to the Standard Model. This is the Higgs field, introduced by Peter Higgs[3] in 1964 [10]. The fundamental understanding is that the universe was initially in a higher vacuum state and the symmetry was broken by the transition to the (local) lowest asymmetric state.

In the Lagrangian, it contributes with a term of the form

$$\mathcal{L}_\psi = |\mathcal{D}_\mu \phi|^2 - V(\phi) \tag{1.21}$$

$$= \left|\left(i\partial_\mu - g\vec{T}\cdot\vec{W}_\mu - g'\frac{Y}{2}B_\mu\right)\phi\right|^2 - V(\phi),$$

describing the kinetic energy and interaction term of the scalar field ϕ with $W^{1,2,3}$ and B_μ fields. $\vec{T}$ are the SU(2) generators related to the Pauli matrices by $T_i = \frac{1}{2}\sigma_i$ see equation 1.13. The potential is given by

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{1.22}$$

---

[3]the Higgs mechanism was in general developed by different people independently and could also be referred to as the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism. The application of the Higgs mechanism was in fact done by Steven Weinberg and Abdus Salam. Furthermore, Gerard 't Hooft showed that the Standard Model was renormalisable with the Higgs-mechanism.

The Lagrangian still has to be gauge invariant and it has to break the $SU(2)_L \times U(1)_Y$ symmetry so masses are generated. This means, that the ground state must have non-vanishing values for hypercharge and weak isospin but cannot be electrically charged e.g. $0 = \frac{Y}{2} + I_W^3$. The simplest choice is a scalar with weak hypercharge $Y = 1$ and weak isospin $I_W^3 = -\frac{1}{2}$

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.23}$$

Considering equation 1.22, λ has to be positive as the energy of the ground state should be finite. Choosing $\mu^2 > 0$ the potential will have a minimum at $\phi = 0$ which will not allow for the generation of mass terms, so μ^2 has to be negative.

This will give an infinite number of equally likely states at lowest energy. These states will have non-vanishing expectation value, ν, with $\nu^2 = -\mu^2/\lambda$. The final choice of ground state is assumed to be randomly selected by Nature to be one of the minima of the potential. The idea is represented in figure 1.1. After the selection of ground state, the symmetry is broken. The ground state has to yield the correct mass relations and break symmetry as well as be
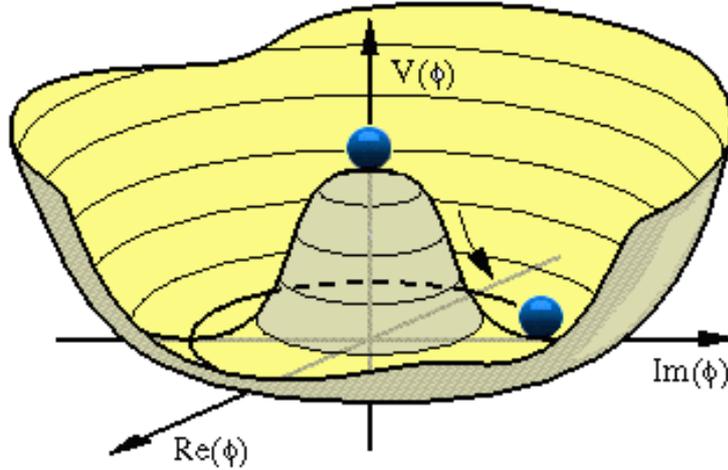
Figure 1.1: The Higgs potential depicted in two degrees of freedom. The $\phi = 0$ state of the field is not stable and a ground state is only found after Nature "rolls the ball to side" e.g. selects one of the minima of the potential as physical ground state, thus breaking the symmetry.

invariant under the $U(1)_{EM}$ symmetry. One choice that fulfils this is

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{1.24}$$

Transforming this under $SU(2)_L$, $\phi \to e^{i\alpha_i \sigma_i/2}\phi$ results in:

$$\phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} e^{i\vec{\sigma}\cdot\vec{\theta}(x)/\nu}, \tag{1.25}$$

where $h(x)$ is the Higgs field and θ are Goldstone bosons - excess degrees of freedom. They are absorbed by the gauge bosons, $W_\mu^{1,2,3}$ through what is known as the Higgs mechanism.

This is described in [4] chapter 20. The result is the relations of equation 1.18 and the gauge bosons receive the mass terms:

$$m_W = \frac{1}{2}\nu g \qquad (1.26)$$

$$A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2 + g'^2}} \qquad \text{with mass} \qquad m_A = 0$$

$$Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}} \qquad \text{with mass} \qquad m_Z = \frac{1}{2}\nu\sqrt{g^2 + g'^2}$$

After the Higgs field acquires a vacuum expectation value the fermion masses are generated by *Yukawa interactions* of the form $g_f\bar{\psi}\phi\psi$, with g_f being proportional to the fermion mass.

From the Lagrangian to the three-lepton final state.

The final-state probed in this thesis includes four leptons with three of them being any combination of electrons and muons. The fourth is a neutrino but as the neutrino has such a low interaction probability with other matter it will not be detected and the final-state is called a three-lepton final state. The cross-section for a process is a measure used to estimate how often we will see an event of that type in a particle collision. If a quark and an anti-quark collide the rate at which they result in e.g. three electrons or muons and a neutrino would be denoted $\sigma(q\bar{q} \to lll + \nu)$, where l is e or μ. Here the charge of the leptons have been left out for simplicity but the process must conserve lepton numbers.

The amplitude of the process can be calculated using quantum field theory (using the $|\mathcal{M}^2|$ method described in e.g. [4]). This is related to the cross-section through

$$\left(\frac{d\sigma}{d\Omega}\right)_{CM} = \frac{1}{2E_a2E_b|\nu_a - \nu_b|}\frac{|\mathbf{p_1}|}{(2\pi)^24E_{cm}}|\mathcal{M}(p_a, p_b \to p_1, p_2)|^2, \qquad (1.27)$$

where $d\Omega$ is $\sin\theta d\theta d\phi$ with $\int d\Omega = 4\pi$, $|\nu_a - \nu_b|$ is the relative velocity of the incoming particles in the laboratory frame, the subscripts 1 and 2 the outgoing particles and E and p, their energy and momentum, with E_{CM} being the total initial energy. The equation has the assumption that the incoming particles can be seen as coming from an infinite past, while the resulting particles are in the infinite future i.e. that anything happening before or after can be factorised away.

These calculations can be illustrated by *Feynman diagrams* where a type of line is drawn representing each kind of particle (the propagator) and each vertex represents the coupling between the particles meeting in that vertex. An example of a Feynman diagram can be seen in figure 1.2. Here the t-axis represents displacement in time while the s-axis represents displacement in space. The figure shows the scattering of two fermions described by the interaction part of the Lagrangian in equation 1.10 when interpreting $\psi(\bar{\psi})$ as the representation of the electron (positron) and A_μ as the photon.

In the Standard Model the only process contributing to a final-state of exactly three leptons and a neutrino is the decay of a W and a Z boson produced through $q\bar{q}$ (proton-proton) collisions. By Z we actually refer to the quantum mechanical mixture of the two fields as seen from eq. 1.26. The label Z is used for simplicity but refers to Z/γ^*. The Feynman diagrams for this process is presented in figure 1.3.
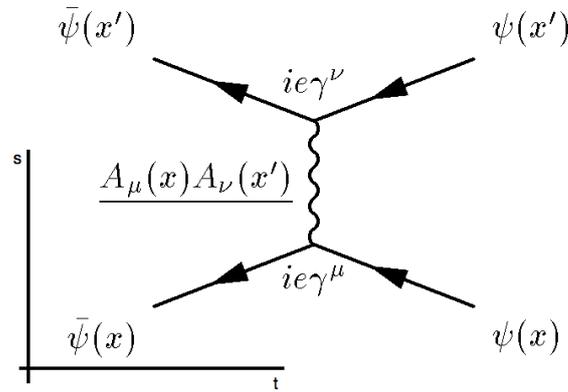
Figure 1.2: A Feynman diagram of the scattering of two fermions each described by a Lagrangian like in Eq. 1.10.
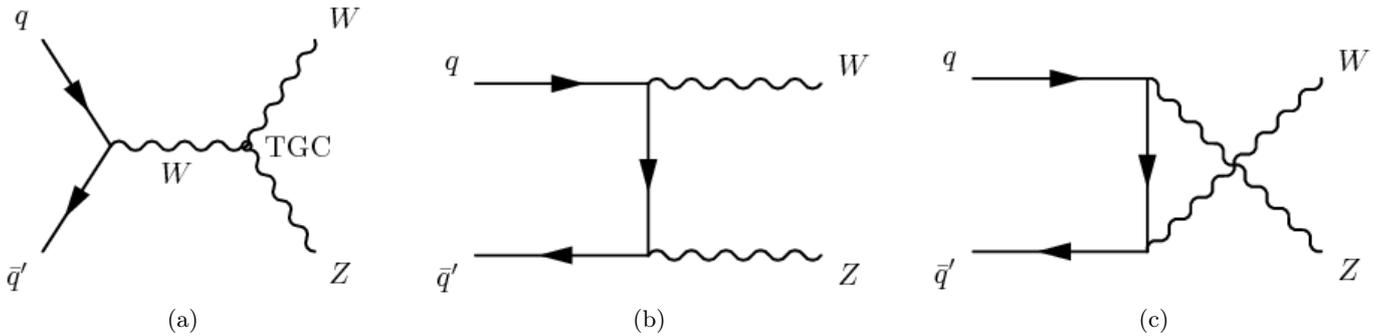

Figure 1.3: The production mechanism for WZ in the Standard Model. Equivalent diagrams exist for γ^*.

The Standard Model botanics

The Standard Model assumes that we know all existing particles at fundamental scale and that they are elementary particles meaning they at least from the view of the Standard Model contain no substructure. Furthermore all interactions are described by the Lagrangian and the symmetries involved. The *botanics* of the Standard Model are shown in figure 1.4.

1.2 Beyond the Standard Model

The minimal description presented above gives the most basic overview of the Standard Model. It is a very successful model but it is not a fundamental theory. The Standard Model assumes all particles to be point-like particles in the effective model. This might be true but there might be a more fundamental understanding like proposed in for instance String Theory where particles are thought of as vibrating strings. The Standard Model combines the description of three of the four fundamental forces, but does not include gravity. It might be that gravity is not relevant at these scales, but if the four forces originate from one fundamental force at
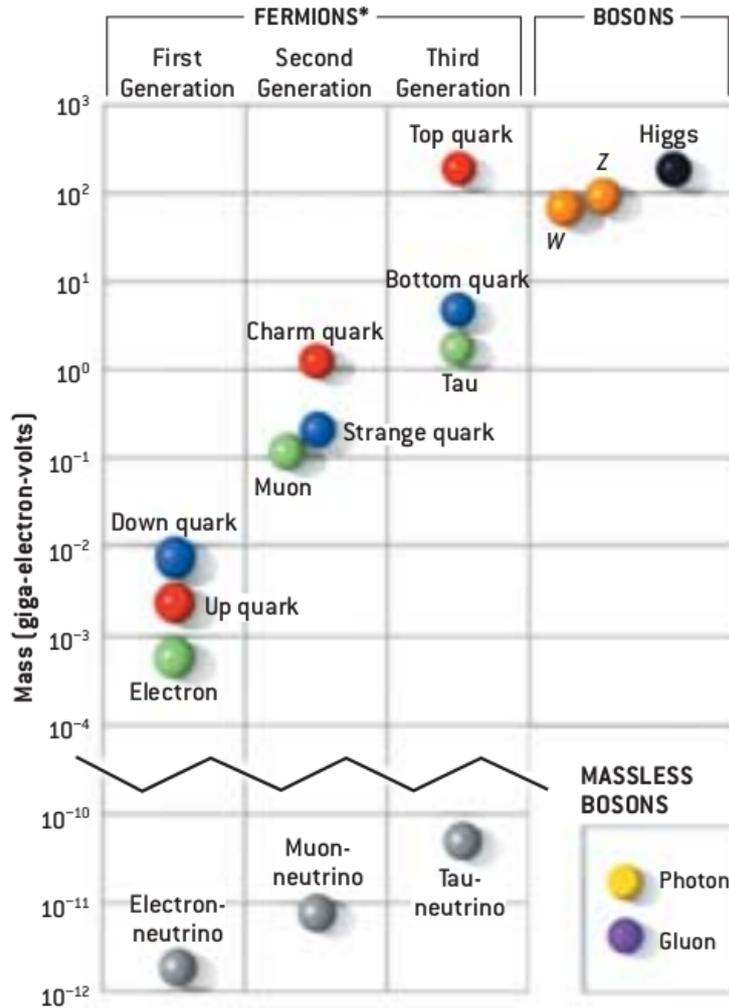
Figure 1.4: The particle botanics of the Standard Model - Image credit: Gordon Kane, Scientific American, May 2003.

the birth of the universe, gravity must be part of the fundamental theory describing these dynamics.

The universe as astronomers observe it consists almost entirely of matter and very little anti-matter is observed. Way to little to be consistent with the proposed model for the Big Bang and the evolution of the universe. Basically matter and anti-matter must always be created in nearly equal amounts following the Standard Model. Several effects can explain this anomaly but they all lie outside the Standard Model - some even requiring a new fundamental theory.

Another open question in the Standard Model is the origin of the mass of the neutrinos and the ultra-light scale of it. If all Standard Model particles get their mass from the Higgs mechanism why would the electron neutrino be 10^5 times lighter than the electron? In other words why does the neutrinos couple so lightly to the Higgs field. As there exist no right-handed neutrinos in the Standard Model they do not gain mass from the Yukawa coupling.

With the discovery of neutrino oscillation (see e.g. [11]) it became clear that at least two of the neutrinos had a mass. This is not a problem for the Standard Model as the neutrino mass can be added within gauge invariance. It is however a widespread assumption among physicists that it is unnatural with the large difference in mass scale without a underlying explanation. A possible explanation will be discussed in section 1.4.

1.3 New Physics

As it stands today no clear hints of new physics exist. Several experiments have made 3σ *observations* of new physics including the proposed signal from Dark Matter in the form of sterile neutrinos [12] and the much debated BICEP2 claim [13] of discovery of primordial B-waves that in short would be an indication or even proof of quantised gravity. Common to them all is that none have stood their ground and made it into a discovery (of which the statistical requirements are set higher than to claim an observation) and at the moment the physics community does not know at what energies new physics will occur. We do know, however, that it will - if not before at least at the Planck scale where quantum field theory breaks down in its current form. The metastability of the Higgs, due to the measured mass just above 125 GeV, could mean that there exists a state with lower vacuum expectation value. It is not yet clear at what energy scale this physics will be present. In principle the current universe could be in a local minimum and could go through another phase transition e.g. tunnel to an even lower state at much higher energies. We would not be there to experience the new state but fortunately this is not something that should keep you awake at night as the metastability is at a level where this is not expected to happen [14].

1.3.1 Model Independence VS. specific searches

The search for new physics is commonly done by optimising the analysis for a given model but one can also adopt a model-independent approach where the search looks for general signatures of new physics. The model independent search has several advantages one of them being it tries to be as unbiased towards a specific type of new physics as possible. This is a logical way to attack the problem when it is unclear where the new physics might occur.

The practical implementation of this is not always as easy though. Complete model independence can easily result in a search with little sensitivity to any kind of new physics. Assumptions must be made to the final state to be probed and the expected decay-modes, the resulting particles originate from.

1.3.2 Exotic models resulting in multilepton final states

Several types of *exotic* models exist that try to encompass one or more of the problems or inadequacies of the Standard Model. Common to most of them is that their contribution to the Standard Model Lagrangian must still leave it gauge invariant although some of the symmetries of the Standard Model are allowed to be broken. This might even be the purpose of the model in the case of e.g. models trying to explain the matter-anti-matter asymmetry. It is of interest to this thesis that multilepton final states are a common denominator for many of the widely accepted possible extensions to the Standard Model.

SUSY

Supersymmetry has been and still is one of the major types of exotic models being probed, partly because it has candidates for Dark Matter particles. Numerous distinct variations exist excellently described in [15]. They all add fields to the Standard Model resulting in more than one Higgs particle but equally importantly in a supersymmetric spin-half partner to all of the existing bosons. An example from SUSY resulting in lepton final states, which are of special interest to the work presented here, is the decay of sleptons - the supersymmetric partner of the leptons - which will happen through the decay to leptons and lighter SUSY particles reconstructed as missing energy.

Gravity and extra spatial dimensions

Extra spatial dimensions (ED) is not as such a theory but a phenomenon needed for several types of exotic models including fundamental *String Theory*. One of the features of extra dimensions is that it allows for more than one graviton - the hypothesised information carrying boson associated with quantised gravity - to exist effectively allowing gravity to grow in strength at small distances depending on the specific model for extra dimensions see e.g [16], [17] and the more recent [18]. Some of the main final states for ED are heavy Standard Model particle partners (Kaluza-Klein particles) and microscopic black holes both of which can decay to multilepton final states.

More generally a range of models lead to heavy (excited) particles that will decay to known particles and as many of them will couple to the electroweak force they will have final states consisting of leptons.

1.4 Seesaw mechanism as a benchmark model

The work presented in this thesis will include a test for physics beyond the Standard Model. As this will serve as a benchmark test it could in principle be any new physics model. The choice fell on a model that explains the generation of masses for neutrinos by introducing a type of heavy right-handed partners. The model was chosen as it has final-states involving Z, W and leptons.

1.4.1 Seesaw mechanism

The principle behind seesaw mechanisms can be illustrated with a 2x2 matrix:

$$\mathcal{X} = \begin{pmatrix} 0 & M \\ M & Y \end{pmatrix}, \tag{1.28}$$

where Y is much larger than M. The eigenvalues of it will be:

$$\lambda_{\pm} = \frac{Y \pm \sqrt{Y^2 + 4M^2}}{2} \tag{1.29}$$

λ_{+} is approximately Y but λ_{-} will be approximately M^2/Y by taylor expansion. The two eigenvalues are inversely proportional so if one goes up the other goes down. This feature has

let to the name "seesaw". By imagining there exist a right handed neutrino which is a singlet under weak isospin, $\psi_R = \nu_R$, mass terms can be added to the Lagrangian as:

$$\mathcal{L}_{seesaw,mass} = -\frac{1}{2}(\bar{\nu}_L \bar{\nu}_R^c)\mathcal{M}\begin{pmatrix} \nu_L^c \\ \nu_R \end{pmatrix} - \frac{1}{2}(\bar{\nu}_L^c \bar{\nu}_R)\mathcal{M}\begin{pmatrix} \nu_L \\ \nu_R^c \end{pmatrix} \tag{1.30}$$

where $\mathcal{M}$ is a mass-mixing matrix for the neutrinos much like eq. 1.28:

$$\mathcal{M} = \begin{pmatrix} m_M^L & m_D \\ m_D & m_M^R \end{pmatrix}. \tag{1.31}$$

The off-diagonal elements m_D are Dirac mass terms that are only allowed after spontaneous symmetry breaking. After this m_D terms are generated via the Yukawa interactions with the Higgs field:

$$\mathcal{L} = g_\nu \psi_R \psi_L H * +... \tag{1.32}$$

where ψ_L is the lepton isospin doublet. The diagonal terms are the *Majorana* mass terms and the neutrinos will get their mass according to the interaction illustrated in figure 1.5.
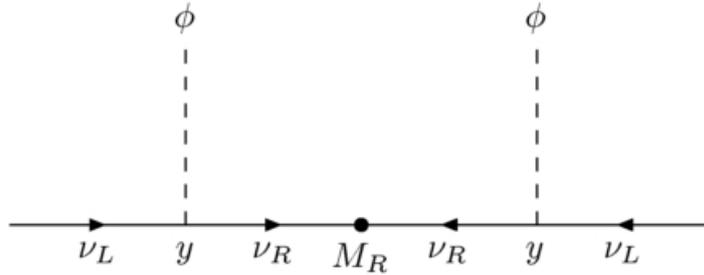

Figure 1.5: The (type I) seesaw mechanism for generating mass to the neutrinos.

Type III seesaw mechanism

Several models exist for seesaw mechanism. Common to them is the existence of heavier right-handed partners that interact with the left-handed neutrinos to give them mass. The model used in this thesis is the so called Type III seesaw mechanism. It was chosen as it is the most promising for searches in high-energy particle collisions. It is unfortunately also the most complex of the models so the following description of it will only serve to give an overview.

As described in [19] the Standard Model can be extended by the addition of a fermionic *triplet*. The triplet is introduced under the form of a heavy charged lepton $N^\pm$ and a heavy neutral lepton N^0 being its own antiparticle. They all have zero weak hypercharge.

The interest to the work presented here is not the underlying theory but the phenomenology the model represents. The fermions N^a which has the three components, $N^0 := N^3$ with zero charge and $N^\pm := (N^1 \mp iN^2)/\sqrt{(2)}$ with charge ± 1, decay to observable particles. As explained in detail in [19] the lightest triplet including both $N\pm$ and N^0 will couple to the Standard Model leptons and Z and W.

At the LHC (see description later) the dominant processes are through pair production of the triplets in $q\bar{q}' \to W^{\pm}l^{\mp}W^{\pm}\nu$ [20] illustrated in figure 1.6. The final states will have large contributions from either 3 leptons + missing energy (and jets) or 2 same-sign leptons + missing energy and jets [20]. The processes have been included in table 1.1. The first of these two constitutes an excellent final state for a benchmark model to be used in the work presented here. It should be noted that for each process involving a decay to a Z there exist an identical process but with the Z exchanged by a Higgs-particle at a lower branching ratio [20].
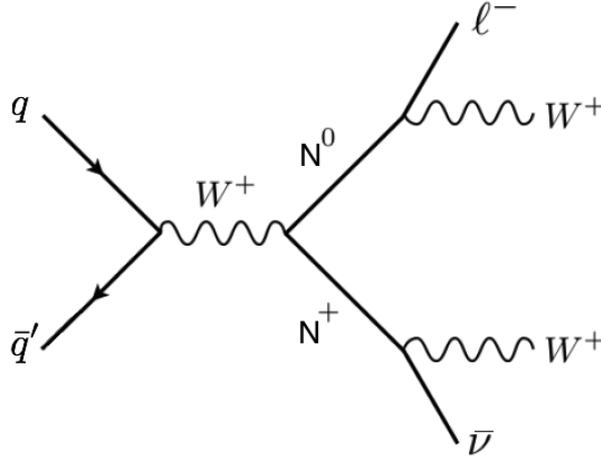

Figure 1.6: The dominant production and decay of type III seesaw triplets at LHC.

Table 1.1: Seesaw triplet production and decay
$$pp \to N_0 + N_{\pm} \to l^{\pm}W^{\mp} + l^{\pm}Z$$
$$pp \to N_+ + N_- \to l^-Z + l^+Z$$
$$pp \to N_+ + N_- \to l^+Z + \nu_l W^-$$
$$pp \to N_+ + N_- \to \nu_l W^+ + l^-Z$$
$$pp \to N_0 + N_{\pm} \to \nu_l Z + \nu_l W^{\pm}$$
$$pp \to N_0 + N_{\pm} \to \nu_l W^{\pm} + l^{\pm}W^{\mp}$$

The accurate simulation of the expectations of the Standard Model in a given final-state is absolutely critical to any particle physics analysis. If deviations from the Standard Model are to be measured, it is imperative to have very precise predictions for the signature of the Standard Model expectations. This is not a simple task. The cross-section for WZ production can, as described in the theory section, be calculated. The step from the theoretical cross-section to the full detector simulation will be described in the following chapter with emphasis on generators for the processes relevant for this study.

2.1 The Monte Carlo generation method

Monte Carlo methods rely on computer algorithms that use repeated random sampling to compute particle interaction probabilities in particle physics. Monte Carlo *generators* are written to perform the task of simulating realistic events of entire collisions. It is an absolutely central assumption that the calculations can be factorised. This means every step can be calculated independently and later combined.

The simulations of events can be split up into four different parts: 1. The hard process, 2. Parton Showering 3. Hadronisation and final decays and 4. the Underlying Event. They have been illustrated pictorially for a hadron collision in figure 2.2 and are described below.

2.1.1 The hard process

The cross-section calculated from theory is $\sigma(q\bar{q} \to WZ)$ with the possibility of adding the decay of the W and Z to leptons through multiplication of their branching ratios. But colliders accelerate hadrons or leptons and not individual partons. This means the first step in a simulation of a WZ event is to know the probability of obtaining a quark and an anti-quark from the initial colliding hadrons at given energies. The functions describing this probability is known as parton distribution functions and several groups are developing these in parallel. Two of the major groups are CTEQ [21] and MRST (MSTW) [22]. These groups extract specific *tunes* of the parton distribution functions from fits to large datasets and tunes are updated as new data becomes available. The contribution of anti-quarks in the structure function can be obtained from the collisions of e.g. two protons with large energy as some of the energy will result in the creation of quark-anti-quark pairs; the sea-quarks.

The total cross-section for WZ will have the form:

$$\sigma(p(P_i)p(P_j) \to WZ \to lll + \nu) = \int_0^1 dx_i \int_0^1 dx_j \sum_q f_q(x_i)f_{\bar{q}}(x_j) \cdot \sigma(q\bar{q} \to WZ \to lll + \nu) \quad (2.1)$$

where x_i is the momentum fraction the i'th parton takes from the hadron it is produced in (P_i) and $f_{q(\bar{q})}(x_i)$ is the parton distribution function describing the probability to draw a quark (anti-quark) from the distribution for a given energy. The integral over all partons sum to one.
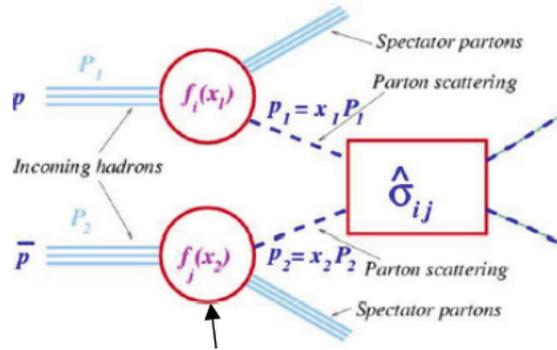

Figure 2.1: The *Parton Distribution Functions*, $f(x)$ are a description of the proton constituents behaviour governing the collisions. The functions $f(x)$ describe the probability of a given parton to carry a momentum fraction, x, of the proton momentum. The cross-section for a given interaction, $\hat{\sigma}$, is dependant on these functions.

The initial particles of the process have momentum transfer, Q^2 between them, determined by the hard process. For collisions involving high Q^2 all processes will happen within short distances, and the calculations can be done perturbatively using matrix elements. If the calculations are done at tree-level it is called leading order, LO. Corrections can be made to these calculations but if loop diagrams and virtual particles are taken into account, the calculations are next to leading order, NLO.

2.1.2 Parton Showers

The large momentum transfer in the hard process means the partons are accelerated to a degree where they emit gluons just like accelerated electrons emit photons. As the gluons have colour charge they can radiate further resulting in what is known as parton showers of quarks and gluons. The partons can in principle be thought of as higher order corrections to the hard process but instead of calculating these exactly (which is not feasible) an approach is taken where splittings are assumed to be collinear and/or gluon emissions soft in energy.

The terms initial- and final state radiation are often used to describe different origins of the parton showers. The initial state radiation consists of the gluons and quarks radiated off before the hard scattering and final state radiation the gluons and quarks emitted after the hard scattering. If there are no outgoing partons from the hard process no final state radiation will be present so for the production of WZ only initial state radiation is present.

After the initial hard process the parton showers are evolved until the partons reach the energy threshold of the pertubative description. This is around 1 GeV where the strong
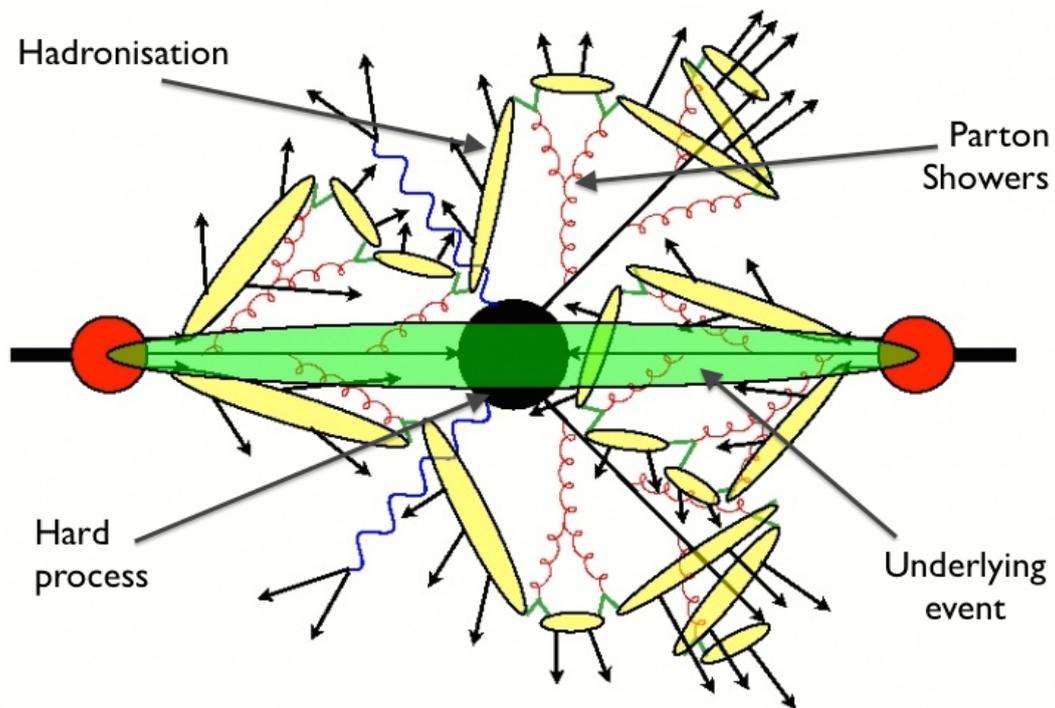
Figure 2.2: The simulation of a hadron collision is factorised into different parts. This is a necessity as the different parts cannot all be calculated analytically. The figure illustrates the complexity of event simulation. The figure is courtesy of [23] - the labels are added by the author.

coupling constant becomes large. [24, 25] There is a subtlety that need to be described. If for instance the process containing a Z and 1 parton in the final state is calculated the following parton showering might result in partons almost identical to the parton from the NLO diagram. Figure 2.3 illustrates the subtlety. The first row represents the calculation of the matrix element to leading order and with all radiation coming from parton showering illustrated by the diagrams going towards the right. Alternatively the matrix elements, ME, can also be calculated to NLO and then both LO and NLO diagrams can be evolved via the Parton Showers. This is illustrated by the following rows. There are overlaps between the diagrams with gluons from the initial matrix elements and the gluons from parton showering. The difference between the two will lie in the momentum or the angle of the emitted gluon. For the matrix elements, the gluon will have energy of the order of the transferred momentum. The radiation from parton showering will have to be either collinear or soft, so for all transverse showers, gluons will be soft. The difference between the calculations is however small. As a consequence a careful matching and merging of the diagrams must be performed which is done by dedicated matching algorithms.
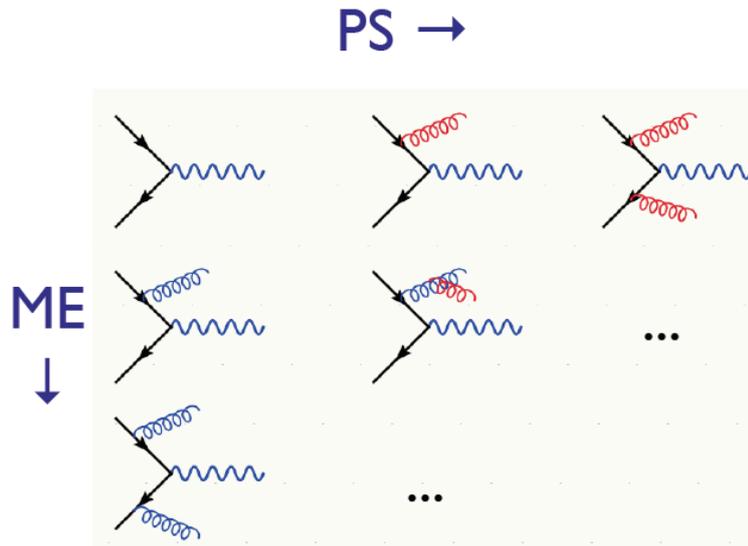

Figure 2.3: The evolution of parton showers can result in overlap between diagrams produced in the hard process and diagrams resulting from parton showering. The matching and selection of which diagrams to use, is performed by matching algorithms. The colour codes are: blue for matrix elements and red for parton showers.

Calculations to NLO has the strong advantage over LO, that matching and merging has only to be done between the NLO matrix element diagram and corresponding one parton, Parton Shower diagram. Calculation of NLO diagrams is an ongoing theoretical discipline and many processes have not yet been calculated to NLO and very few to higher orders.

2.1.3 Hadronisation and decays

As the Parton showers reach the pertubative threshold the particles can no longer be described as asymptotically free and theoretical calculations are no longer possible. At this point the

showers will *hadronise*. The colour confinement of the quarks dictates that no free quarks can exist after an interaction. The quarks and gluons emerging from a collision therefore have to be merged into colour neutral objects by the algorithm performing the hadronisation. Hadronisation is performed differently for the different generators but can be split into three different generic types [26]: Independent fragmentation, string fragmentation, and cluster hadronisation.

Independent fragmentation does not take into account the colour connections and has not been used for any of the generators used in this study.

String fragmentation give better agreement with data so far but also contain a lot of parameters to describe the flavour composition.

Cluster hadronisation is the main alternative to string fragmentation. It has fewer parameters but also has more difficulties in simulating some parts of data correctly.

Lastly, unstable particles created during previous steps, like heavy hadrons or taus, will have to be decayed after hadronisation. The decays are governed by the branching ratios and can be calculated theoretically, making it a relatively well understood part of the event generation process.

2.1.4 Underlying event

The underlying event consists of the rest of the partons originating from the protons colliding. They must of course also become colour neutral but distances to the hard process are large enough that there is generally no colour exchange between these beam remnants and the hard process. The underlying event however depends on the hard process and the evolution is done by following the steps of parton showering and hadronisation.

It may happen in colliders that more than one pair of protons collide in an event. This is referred to as pile-up. These interactions are far enough away from each other in terms of time scales involved that pile-up is simulated by adding events on top of each other. It is also possible to have multi-parton interactions in a given event where several pairs of partons collide. Apart from the effect of other collision in the event, out-of-time pile-up exist. This is the effects of remnants from earlier collisions in the detector. All pile-up simulation is handled by dedicated Pile-up tools as described in section 6.2.5.

2.2 The generators

The combination of these four parts into a Monte Carlo generator is done in many different ways. Often different programs will be used to perform each specific part of the tasks. These can be combined via the *Les Houches Accord*. Programs exists that can calculate all four parts themselves, e.g. that can be used as standalone Monte Carlo generators. These constitute the backbone for generating events, but will often be interfaced to other programs to perform one or more of the four parts of the calculation. Pythia [27], HERWIG [28] and Sherpa [29] are examples of programs that generate full events that can be used directly in physics analysis or interfaced to detector simulation. The work in this thesis uses Monte Carlo samples generated using both Pythia and Sherpa and HERWIG to generate parton showers and hadronisation for some processes.
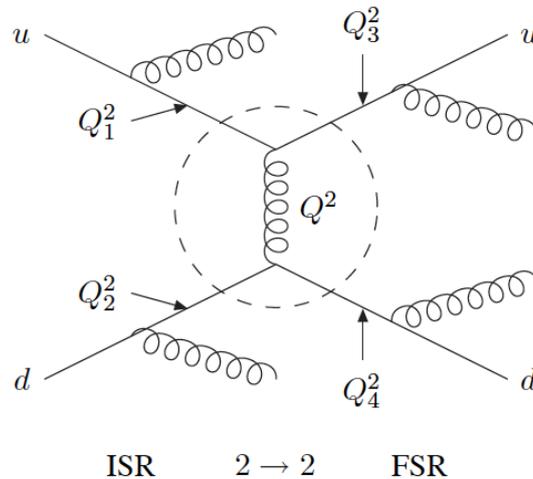
Figure 2.4: The calculation of $2 \to n$ processes would not be possible in generators for n much higher than 2 if not for the method of factorising the event. In this way different parts of the process can be calculated independently and merged afterwards, to a complete numerical calculation. The figure shows the factorisation of a $2 \to 2$ process with both ISR and FSR. From [25].

2.2.1 PYTHIA

Pythia [27] is one of the most widely used generators, as it is a standalone generator for everything from few-body hard processes to complex multihadronic final states through parton showering and hadronisation. All main aspects of the events are simulated, such as hard-process selection, initial- and final-state radiation, beam remnants, fragmentation, decays, etc. It uses the CTEQ 5L [21] parton distribution function as default but can be used with others.

Pythia has the limitation that it only calculates $2 \to 2$ or $2 \to 1$ interactions of the hard process, which means that all further particles will have to be created by parton showering. $2 \to n$ final states can, however, be created this way. The method is based on the already described factorisation and is shown in figure 2.4. If other programs are used for the calculations of matrix elements, a matching and merging of diagrams arising from those and the Pythia parton showering is performed.

The hadronisation is done by string fragmentation, followed by the decays of unstable particles. The method used, is the Lund string model that relies on a description of the QCD field lines as compressed tubelike regions for large charge separation. These field lines are represented as gluon strings that, when stretched long enough, break up into quark-antiquark pairs, which form up with the existing quarks forming colour neutral mesons. This part is almost completely non-perturbative, and so requires extensive modelling and parametrisation from existing data. Pythia has also the ability to simulate effects of the underlying event.

2.2.2   HERWIG

Herwig [28] works in many ways like Pythia. There is, however, a wider range of programs that can be interfaced to Herwig for calculating e.g. matrix elements for $2 \to n$ processes, with n greater than 2. Calculation with Herwig alone will like in Pythia, only be to lowest nontrivial order without e.g. loop corrections.

For hadronisation Herwig does cluster hadronisation. Colourless clusters are formed from colour connected quarks. They consist of quark-antiquark (meson-like clusters), quark-diquark (baryon-like), or antiquark-antidiquark (antibaryon-like) pairs. The basic idea of the model is that the clusters decay according to the phase space available to the decay products. In other words, the initial partons are merged into colourless clusters, that are then decayed to hadrons.

The newer Herwig++ handles simulation of the underlying event whereas the older version needs to be interface with a program called JIMMY [30]. JIMMY focuses on the impact of multiple parton interactions (from e.g. beam remnants) on the event.

2.2.3   SHERPA

Sherpa separates itself from the previous two by being able to perform hard processes going beyond $2 \to 2$. Sherpa uses a fully automated matching and merging of multi-leg tree-level matrix elements $2 \to n$ with the parton shower. The underlying idea for Sherpa is to split the kinematical range of parton emission by a $k_\perp$ algorithm [31] into either a regime of jet production covered by the appropriate matrix elements or a regime of jet evolution covered by cluster hadronisation like for Herwig. The matrix elements are reweighted through Sudakov form factors and any hard emissions in the parton shower are vetoed if they would lead to a jet in the phase space covered by the matrix element calculation. In this way Sherpa resembles full leading order calculations quite well and can be used for central processes with better results than Herwig or Pythia standalone.

The diboson hard processes, WZ, ZZ and WW have been generated using Sherpa with the CT10 parton distribution functions [32] optimised for the energy and luminosity of the LHC 2012 data. These are benchmark samples used in many of the Standard Model and exotic searches involving multilepton final states in ATLAS.

2.2.4   Generators for the hard process

For several of the samples used in this thesis, the matrix elements have been calculated by interfacing Pythia or Herwig to a dedicated generator of the hard process. The programs are optimised for a specific process, either with speed or precision in mind. As the event is stored in the Les Houches accord format, the task of interfacing is made much simpler. The calculation of the hard process is still done with different approaches each having their own advantages.

The program MC@NLO does multi-leg matrix elements with next-to-leading-order QCD matrix elements [33]. It is a package that allows linkage between HERWIG and next to leading order calculations of rates for QCD processes. Although matching between matrix elements and Parton Shower diagrams still has to be done, only NLO and not LO diagrams have to be evolved with parton showering, so fewer diagrams have to be matched.

MC@NLO includes a range of production mechanisms spanning Higgs boson, single vector boson, vector boson pair, heavy quark pair, single top (with and without associated W or

charged Higgs), lepton pair and associated Higgs+W/Z production in hadron collisions. The generator is used in this thesis to simulate events where proton collisions produce leptons through single- and pair-production of top quarks.

The calculation of Z and W plus partons, have been done with ALPGEN [34]. ALPGEN is specialised in multi parton processes. It is based on calculations using tree-level matrix elements, but has been developed especially with multi-jet events in mind. It calculates matrix elements to LO, with a fixed number of additional partons in a process. This is a better approach for events with high jet multiplicities with large p_T than the Parton Shower method, where additional partons (with respect to the initial $2 \rightarrow 2$ process) are generated only during the shower evaluation. The Parton Showers and hadronisation and the following matching will be done by Pythia with the distinct P2011C tune [35].

Triboson events, ZZZ, ZWW, WWW and $t\bar{t} + W/Z/WW$ have all been generated using MadGraph [36]. MadGraph performs NLO hard process calculations and Pythia has been used for the rest of the event generation. For reference the parton distribution function used is AUET2BCTEQ6L. CTEQ6 also used in the P2011C tune of Pythia is not optimal for the 8 TeV 2012 data but the sample constituted the best available at the time of writing. As the processes using the slightly older parton distribution functions are background processes with relatively small expected events in three lepton final states as will be shown later, the impact on the final results is excepted to be small.

<div style="text-align: right;">*3*</div>

Measuring physics at the LHC

3.1 The Large Hadron Collider

The Large Hadron Collider, LHC is the largest particle accelerator in the world. It is located at the border of Switzerland and France. It has been built by the international research organisation CERN (Conseil Européenne pour la Recherche Nucléaire - The European Organisation for Nuclear Research) which include 21 member states and several thousand scientists associated.

The LHC has been built to accelerate protons to unpredecented centre-of-mass energies and has peaked in 2012 with mean energies of 4 TeV per proton resulting in a centre-of-mass energy of 8 TeV. This is the highest energy ever recorded for a man-made proton collider. LHC is built underground with depths ranging from 50 to 175 metres and the circumference of the tunnel is approximately 27 kilometres.

It is designed to accelerate both protons and lead ions depending on the physics of interest. For this study proton collisions are considered. In the accelerator two beams of protons are accelerated inside beam pipes in the tunnel in opposite directions. The protons in the beams are held in position, focused and squeezed tighter together by appropriate magnets along the beam pipes. Particles are initially accelerated by various smaller accelerators before entering LHC and from that point on electric fields will accelerate the protons up to a speed only 3 metres per second slower than the speed of light, $0.999999991c$. Superconducting dipole magnets are used to hold them in place.

At four points on the LHC ring, the beam pipes cross to allow for collision of protons from the oppositely moving beams. At these points the four major experiments at the LHC are situated: ALICE, ATLAS, CMS and LHC-b.

The rate of events produced at ATLAS is given by $N = L\sigma$, where σ is the total cross-section and L is the instantaneous luminosity given in $cm^{-2}s^{-1}$. The luminosity is a measure of the number of collisions per unit of area and time. For data recorded over a period the number of events is given by the time integral of the luminosity, denoted $\mathcal{L}$. The unit used for cross-section is barns and it is customary to present the integrated luminosity in inverse femtobarns, fb^{-1}.

The total integrated luminosity collected by the ATLAS experiment has been increasing drastically each year of running as the beam intensity has been ramped up. The result is that over 20 fb^{-1} was recorded in 2012 alone at 8 TeV. The amount of data delivered in 2012 can

be seen in figure 3.1 and it should be noted that ATLAS has recorded only slightly less data. The data delivered is the total luminosity delivered by LHC whereas the recorded data is the data as seen by the detector (when the detector was fully online). The high-luminosity will result in several protons colliding on top of each other and to illustrate how significant this is the average interactions per event is shown in figure 3.2.
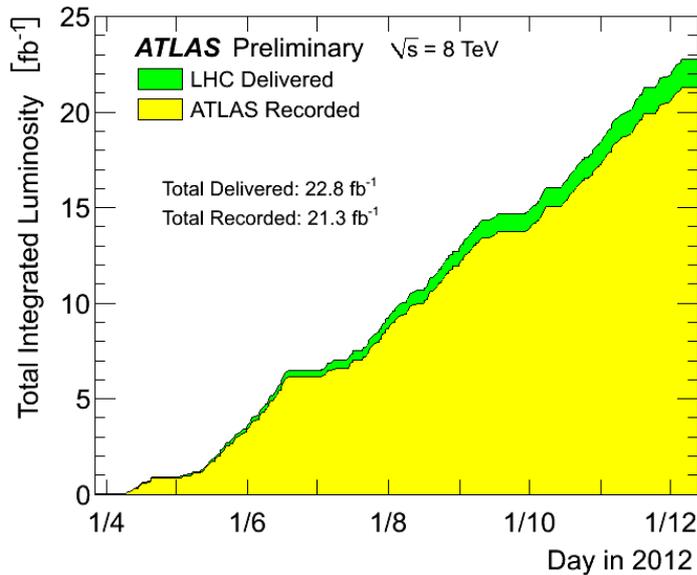

Figure 3.1: Total Integrated Luminosity for the LHC and the amount recorded by the ATLAS experiment during the periods of running in 2012.

3.1.1 A collider designed to look for new physics

The LHC is no ordinary experiment. It is the absolute state-of-art and it needs to be as its design purpose is to push the boundaries for our understanding of the most fundamental dynamics of our universe.

With the discovery of the Higgs, the first goal of the LHC has been achieved, but LHC was all along designed for more. It is currently in shut-down while upgrade work is being performed to increase energy and luminosity of collisions. The cross-section for many new types of physics are quite low, so they can only be fully probed with this higher energy and luminosity, but many of these can already be searched for in the current data.
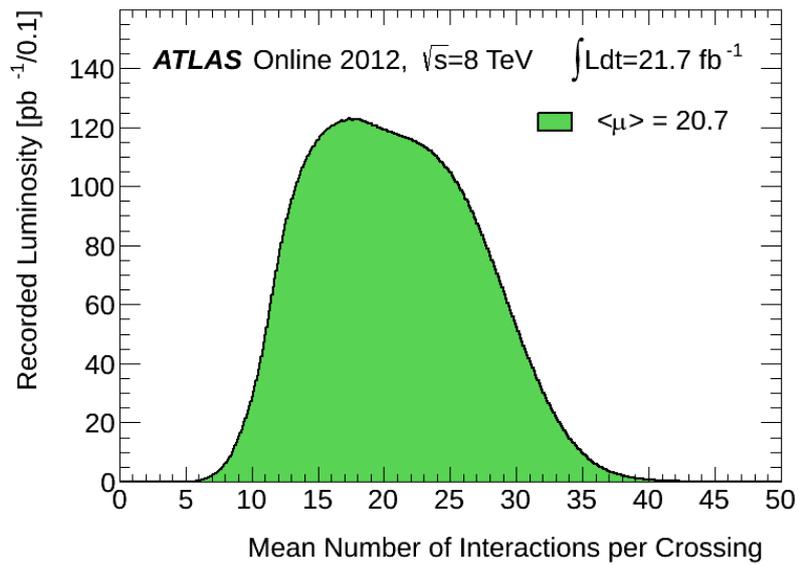
Figure 3.2: The average number of interactions per crossing of the proton-beams for ATLAS in 2012 data.
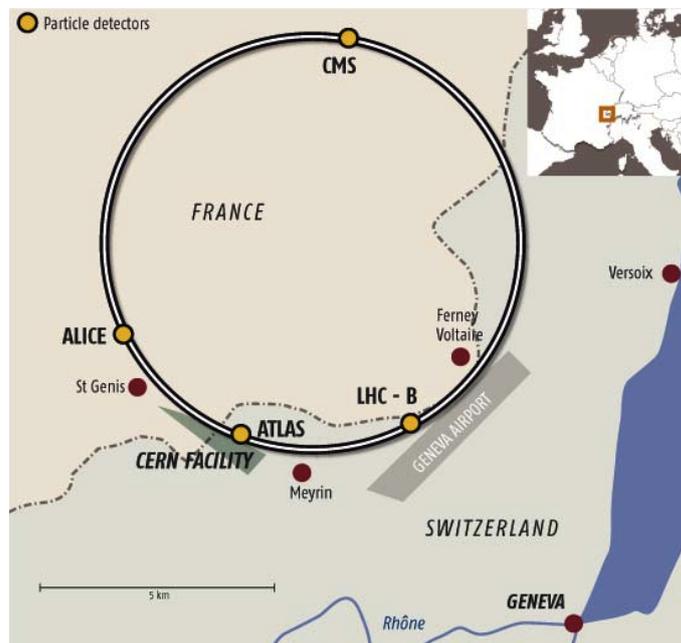

Figure 3.3: The LHC is buried many metres underground to screen the experiments from cosmic radiation and other possible external disturbances. The idea for the ATLAS detector stems from the 1980's and to run and analyse the data collected by ATLAS around 3800 scientist are participating from over 30 countries.

3.2   The ATLAS detector

The ATLAS [37] (A Toroidal LHC Apparatus) is a multipurpose detector designed to cover a wide range of physics including tests of the Standard Model and physics beyond the Standard Model. It consists of several parts, referred to as sub-detectors, each designed and optimised for special tasks. The sub-detectors allow for the sophisticated reconstruction of objects that are essential for the work presented in this thesis. The following will describe the sub-detectors used for the reconstruction of objects in the data-events presented in this thesis.
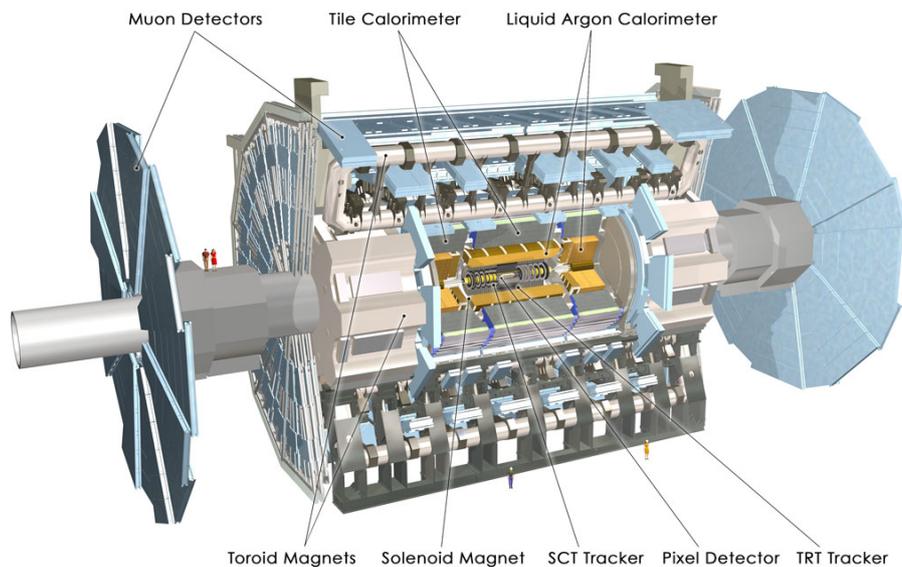
Figure 3.4: The ATLAS detector with name tags on the different sub-detectors.

3.2.1   Inner Detector

The inner detector consist of three different detectors. The pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). All three are arranged into central barrel parts and two end-cap parts composing the forward detector.

Pixel detector

The pixel detector is a silicon pixel detector with a very high granularity which provide high precision measurements of tracks as close to the interaction point as possible. The innermost part is called the b-layer and this is used to provide the first couple of points to extrapolate the particle track from. The innermost part of the tracking is important as it helps establish if particles originate from the primary vertex or from particles that travel a distance before decaying in the detector.
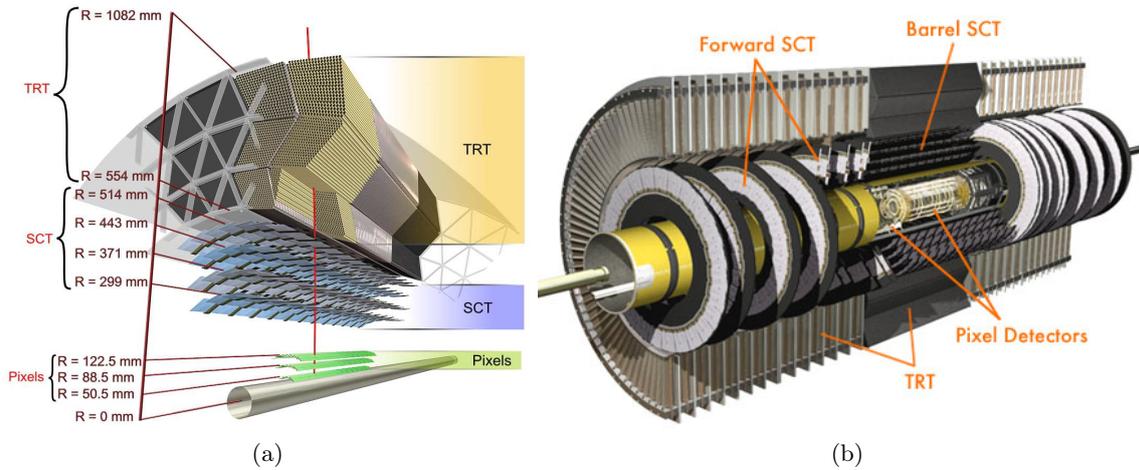
Figure 3.5: The ATLAS inner detector.

SCT

The SCT is similar in construction and function to the pixel detector but differs by having long narrow strips rather than small pixels. This enables it to cover a larger area giving more sampled points and with this roughly the same accuracy as the pixel detector. The strips are read out one-dimensionally opposed to the pixel detector, where each pixel gives a signal (readout) with two-dimensional information. The SCT is designed to provide eight precision measurements per track in the intermediate radial range. Together with the silicon detector this enables exact determination of where the interaction occurred (the vertex position), how much momentum a particle has (due to its curvature in the magnetic field) and secondary vertices and impact parameter. This helps understand the process and is central in the tagging of jets stemming from bottom quarks.

TRT

The transition radiation tracker is the outermost of the three inner detectors. It is a detector made of straws containing gas with a wire at the centre to detect electron avalanches. When a particle crosses a straw, the atoms of the (Xenon) gas are ionised resulting in free electrons. As the central wire carries an electric potential, the electrons will drift towards the centre. The drift of electrons creates secondary ionisation which results in an avalanche of electrons reaching the wire. This allows for a readout of the signal giving a drift time measurement, that gives a spatial resolution of $\mathcal{O}(100)$ μm per straw.

Additionally the material between the straws, called the radiator, is composed of materials with different dielectric constants. This causes ultra-relativistic particles to radiate off photons in the X-ray region. Xenon is chosen as it is particularly sensitive to absorbing those photons resulting in massive ionisation and a much larger signal readout. This type of signal is called a high threshold signal while signal from only ionising charged particles is called low threshold. This is used to separate electrons that cause transition radiation from particles like the pion.

3.2.2 The Magnet system

ATLAS has two different magnet systems, the Solenoidal magnet and the Toroidal magnets. The solenoidal magnet is a superconducting magnet made from a composite that consists of a flat superconducting cable located in the centre of a rectangular aluminium stabiliser. The magnet is designed to provide a 2 T magnetic field parallel to the beam axis. It is placed so the magnetic field surrounds the inner detector while the radial thickness and field in the calorimeters are minimal. The magnetic field causes particles to bend according to their charge and momentum, so these quantities can be determined in offline analysis. This bending also means that particles below roughly 400 MeV are curved to a degree where they will loop repeatedly in the field and are less likely to be measured. This helps to reduce the noise of the irrelevant low p_T particles.

The Toroidal magnet system is made of eight very large air-core superconducting coils forming a barrel, symmetric around the beam axis, and two end-cap parts rotated with respect to the barrel so the coils interleave. It is situated outside the calorimeter and within the muon systems. For the barrel, each coil has its own cryostat, with the coils connected together to form a rigid cold mass which contains the large magnetic forces acting radially inwards.

The toroid system creates a magnetic field around 4 T which is strong enough to bend particles (muons) with energy up to 1 TeV, poorly measured by the inner part of the detector alone, so their momentum and charge can be determined.

3.2.3 The Calorimeters

There are two different calorimeter systems in the ATLAS detector. The electromagnetic- (EM) and hadronic- calorimeter. They are both *sampling calorimeters* which means they alternate layers of high density absorbing materials and active sampling layers, which collect the signal from the resulting particle shower. The energy of the passing particles can be inferred from these showers. The physics governing the two calorimeters are however not the same.

EM calorimeter

The electromagnetic calorimeter is a Liquid Argon calorimeter with interlacing layers of lead and stainless steel. To keep the argon in its liquid state a cryostat surrounds the entire EM calorimeter. Lead is chosen for the interlacing plates, as it has a short radiation length which means electrons or photons moving through the calorimeter will shower and create a cascade of photons within short distances. The secondary electrons will ionise the argon in the narrow gaps. An electric field results in the electrons drifting in the gas-gaps and being readout by copper electrodes. The size of an electromagnetic shower depends linearly in units of radiation length X_0 of the calorimeter material. Figure 3.6a illustrates the electromagnetic calorimeter. The calorimeter can be split into four layers.

- **The presampler** is a single layer of argon without any lead in front. The sole purpose of this layer is to correct for the energy loss in the inner detector, the solenoid magnet and the cryostat wall.

- **The 1st sampling** has a layer of 4.3 radiation lengths in depth. The readout is done from thin strips positioned in the η direction (see figure 3.6a) which provide good reso-

lution in this coordinate for photon/π^0 separation. The magnetic field causes photons to produce showers similar to the π^0 in the ϕ direction.

- **The 2nd sampling** is with its 16 radiation lengths of material the largest layer in the EM calorimeter and it is here the largest part of the energy is deposited. All clusters with energy below 50 GeV are contained within the second sampling [38].

- **The 3rd sampling** is a layer that will only be reached by the most energetic particles and the cell sizes have been doubled in η without loss of resolution, as the energetic particles reaching this layer will give a much wider shower.

The end-cap regions are split into two. Out until $|\eta| = 3.2$ the structure is the same as for the barrel but without the presampler and with less material. From there out to $|\eta| = 4.9$ the calorimeter is made from copper and tungsten. This choice was made to limit the width and depth of showers from high energy jets close to the beampipe and to contain particles from the forward region.

The overlapping region between the barrel and the end-cap calorimeters result in a "dead area" with poor energy resolution but this area has been made as small as possible with room still for the cables and cooling pipes for the inner detector. Electrons from this part of the detector are vetoed in the selection.

Hadronic calorimeter

The hadronic part of the calorimeter in ATLAS is situated outside the cryostat of the EM calorimeter. It is a tile calorimeter build of a steel frame with plastic scintillators inserted as tiles. An outgoing slice of the calorimeter can be seen in figure 3.6b. The scintillator material emits blue light from ionising particles in the hadronic showers. The blue light is sent via wavelength-shifting fibres to the outside of the calorimeter where photomultipliers read out the now longer wavelengths from the fibres.

Unlike the electromagnetic showers, that have a rather constant shower-energy to particle-energy ratio, the energy deposited by hadrons in the hadronic calorimeter varies much. Neutral pions decay to photons like in the EM calorimeter while secondary $\pi^\pm$, neutrons etc. from the nuclear processes, caused by incoming particles interacting with the material in the hadronic calorimeter, give large variations in the estimate of their energy. The size of hadronic showers depends linearly on the interaction length λ of the material which is always longer than the radiation length. [38]

In order to compensate for the variations of the hadronic showers, the ratio of the EM- and the hadronic- calorimeter, e/h, is measured. For a good energy resolution this value should be as close as possible to one.

3.2.4 The Muon detector

The last and outermost detector is the muon detector designed to catch the muons that have otherwise deposited little energy in the inner parts of the detector. The principle for the detector is like for the inner detector. The magnetic field from the toroid bend the muons, so their momentum and charge can be identified and most tracks in the muon detectors can be considered muons, as few other particles make it through the calorimeters to the muon detector.
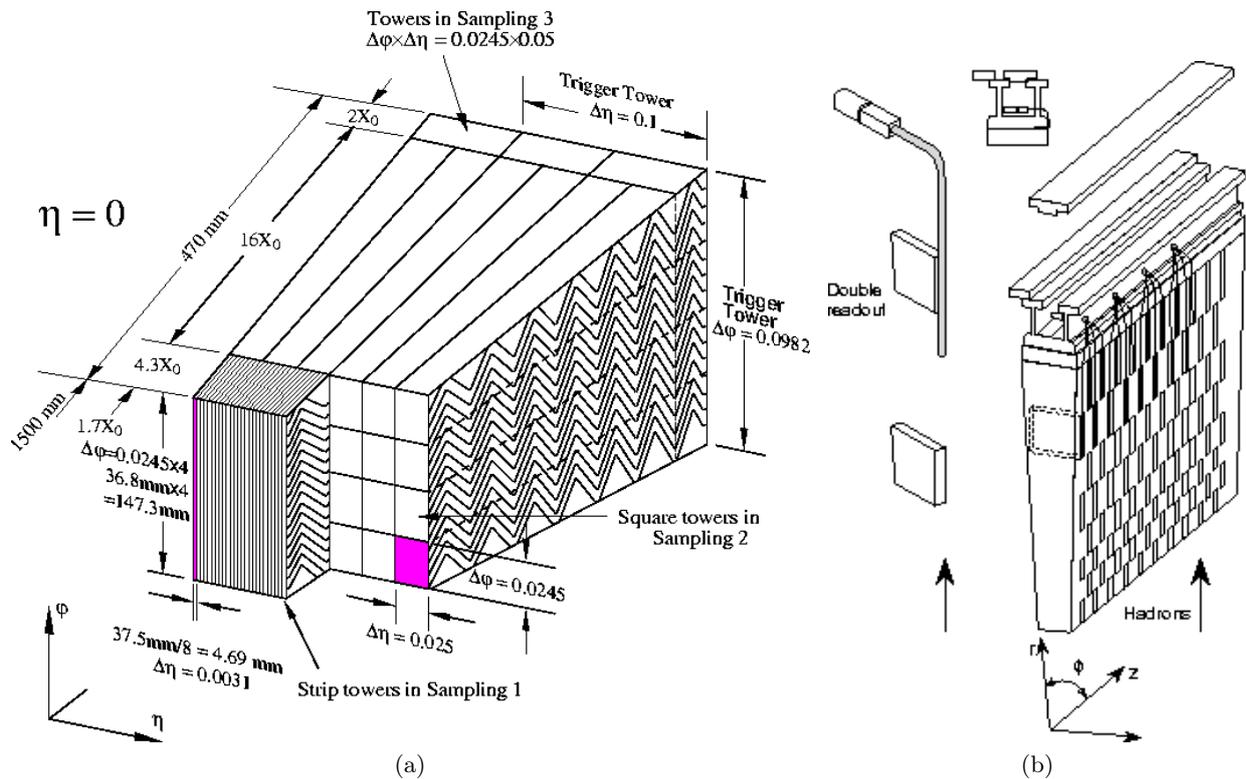
Figure 3.6: The ATLAS Calorimeters. Figures from [38].

The typical signals in the ATLAS detector for different particles is drawn in figure 3.7. For multilepton final states electrons and muons are important, but as is clear from the above no part of the detector could be left out. The force of the many sub-detectors lie both in their individual specialisations but certainly also in their combinatorial possibilities.

3.3 The triggering and data acquisition systems

To round off the experiment section, a few words about the triggering and data distribution- and analysis- systems are needed.

3.3.1 The trigger-system

When LHC is running at peak design luminosity, the interaction rate will be of the order of 1 GHz with a bunch crossing rate of around 40 MHz. This means that approximately 40 million proton bunches collide every second. Each collision results in an event with raw data. If all these events were to be stored, the total amount of data from the ATLAS experiment alone would be in sizes of terabytes or even Petabytes each second. It is impossible to store such an amount and luckily not all events are equally interesting. The cross-sections for production of quarks and gluons in the initial collisions are much higher than for creating Z bosons, and for many searches, the quark events are not as interesting and can be filtered away. To do this a set of *triggers* have been developed to select only events with interesting physics. The triggers
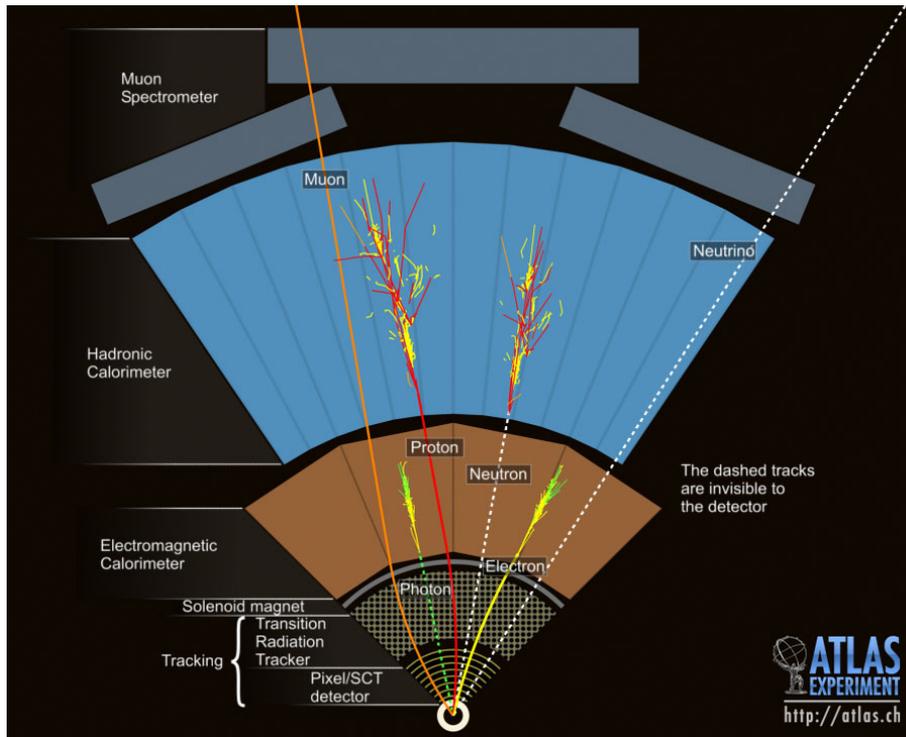
Figure 3.7: The signal from various particles as seen in the ATLAS detector. Different particles leave signals in different parts of the detector and can be identified from this.

can be split into two major groups. Level 1 (L1) triggers, and high level triggers (HLT).

The L1 triggers are based on hardware implemented logic decisions in the detector and get their information from the various parts of the ATLAS detector that have fast readout rates. They use this information to select regions in the detector, that might contain interesting physics and does so within microseconds. Events with no interesting regions are immediately thrown away and anaylsis is performed on the next event. This step reduces the rate to around 75 KHz.

The regions of interest, ROI, are then passed on to the HLT. The HLT consist of two steps. The Level 2, L2, trigger and the Event Filter, EF. The L2 trigger runs more thorough algorithms on the ROI's and reconstructs data in those regions to see whether it is interesting. As the rate of events from the L1 trigger is lower, there is more time to decide whether the event in question is interesting and the algorithms can be more complex. If interesting physics is found, the event is passed on with a rate around 1 KHz.

Finally the event is sent to the EF where reconstruction algorithms are run to find particles in the entire detector or just the regions of interest. If the event is found to be of interest in the end, it is sent to be recorded and is now ready for further analysis.

There are many different triggers spanning from triggers that demand a reconstructed electron of a specific p_T to triggers demanding jets and various other processes. For triggers like the electron trigger, there are often an entire range of triggers demanding electrons with different minimum p_T or with different degree of certainty in the identification, called tightness or even isolation requirements.
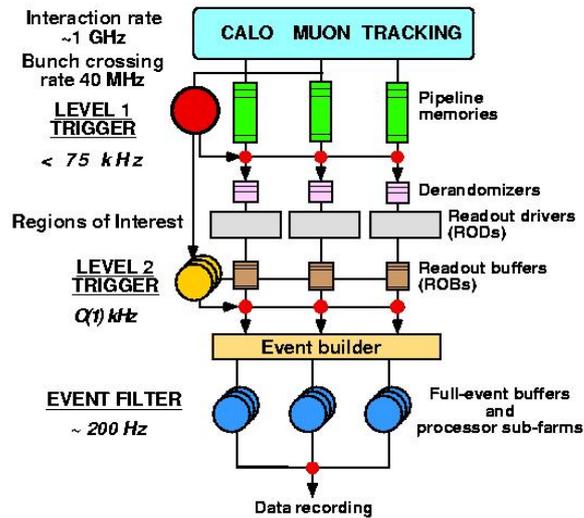
Figure 3.8: The triggering system can be split into the low-level Level 1 trigger and the high-level Level 2 and Event Filter triggers. Level 1 triggers are hardware based and use only very fast algorithms while the high-level triggers use more advanced algorithms and incorporate larger part of the detector.

The triggers have been systematically deployed for the conditions present in different periods of data-taking.

3.4   Reconstruction

To reconstruct a collision in the ATLAS detector all sub detectors are used to combine their respective signals into objects. A series of sophisticated algorithms, much like those of the event filter triggers, help determine what particles where present in the event. An energy deposit in the calorimeters would for instance have to be paired to a track in the inner detector and the muon detector to be considered a muon. The same is true for simulated events. Their response to the detector system is simulated and the reconstruction is run on the events just like on data including all trigger systems. The following section describes the details.

3.4.1   Detector simulation

Before an event generated by the Monte Carlo generators can be compared to experimental data, a simulation of the detector-system must be performed. In ATLAS this can be done with two official programs, ATLFAST II [39] and Geant4 [40]. ATLFAST II combines full simulation of some parts of the detector with faster simulations and retains the storage format and naming convention of real reconstructed data. It is much faster and retains a high degree of detail compared to full simulations of the detector. If the full detail is however desired GEANT4 can be used. Geant4 is a program for simulating particle movement through all kinds of material and is used in ATLAS to simulate the entire detector. It contains information on the weight of each screw and bolt in the detector down to a precision in grams. For a detector of over

1000 tons, this is astonishing. The datasets used in this thesis have been generated with the
generators described above and have all gone through **full** detector simulation.

3.4.2 Electron reconstruction

The electron is reconstructed from the trackers and the calorimeters. Different degrees of
tightness are being used as default: Loose++, Medium++ and Tight++ [1]. They denote
different degrees of discrimination of the electron against objects that could resemble it in the
detector. The amount of parameters and parts of the detector used to identify the electrons
vary with each degree of tightness. An overview of the different variables can be seen in figure
3.10 from [41]. The loose++ identification is based on calorimeter information and inner
detector and takes advantage of the most basic assumptions to reconstruct electrons. The
medium++ electron identification tightens hit requirements in the inner detector including
the TRT and adds a cut on the transverse impact parameter that describes the distance from
the track origin to the primary vertex. The TRT ratio of high-threshold hits to total hits is
especially useful for separating electrons from pions. A tight++ electron adds track-cluster
matching and tightens cuts on track quality and photon conversion rejection. The efficiency
of the electron selection depends on both the η of the electron, p_T and the number of tracks in
the detector and number of reconstructed primary vertices. The last two are highly correlated.
Figure 3.9 shows the dependence of the electron reconstruction efficiency on number of primary
vertices.
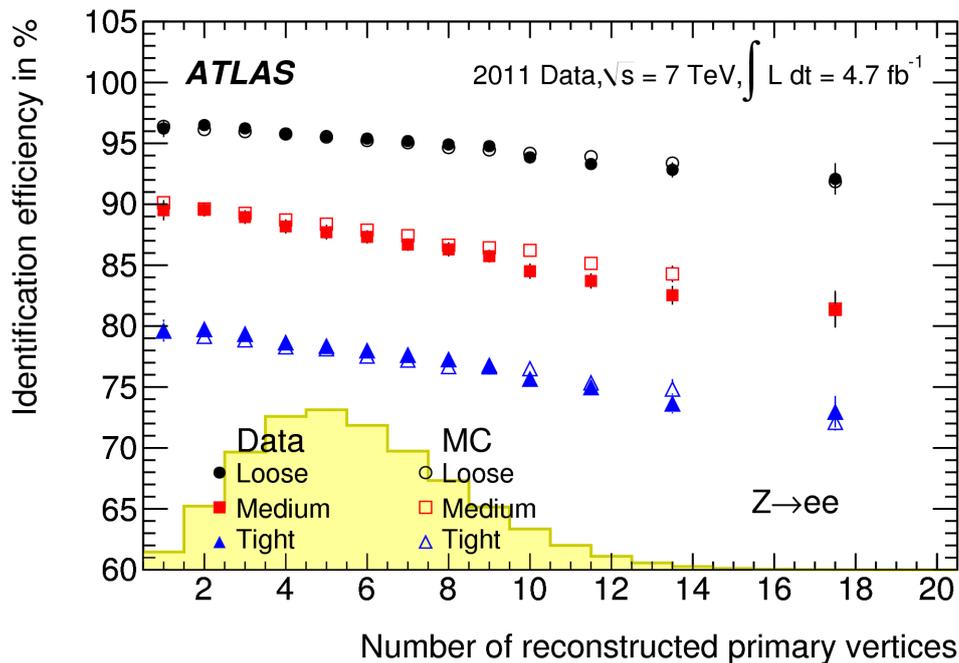
Figure 3.9: The electron reconstruction efficiency as a function of number of reconstructed primary
collisions (corresponding to the number of proton-collisions per event). The same plot for 2012
data is not yet public but the dependence on number of collisions is comparable [41].

---

[1]The ++ is to separate the definitions from the 2011 tightness definitions that resemble that of 2012 but
have looser cuts due to the different pile-up conditions.

Type	Description	Name
LoosePP selection		
Acceptance	$|\eta| < 2.47$	
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $|\eta| < 0.8$ and $|\eta| > 1.37$)	R_{had1}
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $|\eta| > 0.8$ and $|\eta| < 1.37$)	R_{had}
Middle layer of EM calorimeter	Ratio of the energy in 3x7 cells over the energy in 7x7 cells centred at the electron cluster position	R_η
	Lateral width of the shower	$w_{\eta 2}$
Track quality	Number of hits in the pixel detector (≥ 1)	n_{pixel}
	Number of total hits in the pixel and SCT detectors (≥ 7)	n_{Si}
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track ($|\Delta\eta| < 0.015$)	$\Delta\eta$
Strip layer of EM calorimeter	Total shower width	w_{stot}
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	E_{ratio}
MediumPP selection (includes LoosePP)		
Shower-shapes	All LoosePP shower shape cuts but at tighter values	
Track-cluster matching	Tighter $\Delta\eta$ track-cluster matching ($|\Delta\eta| < 0.005$)	$\Delta\eta$
Track quality	Transverse impact parameter ($|d_0| < 5$ mm)	d_0
	Number of b-layer hits (≥ 1) for $|\eta| < 2.01$	n_{BL}
	Stricter requirements for pixel hits (> 1) above $|\eta| > 2.01$	n_{pixel}
TRT	Ratio of the number of high-threshold hits to the total number of hits in the TRT	f_{HT}
TightPP selection (includes MediumPP)		
Track-cluster matching	$\Delta\phi$ between the cluster position in the middle layer and the extrapolated track ($|\Delta\phi| < 0.02$)	$\Delta\phi$
	Ratio of the cluster energy to the track momentum	E/p
	Tighter $\Delta\eta$ requirement ($|\Delta\eta| < 0.005$)	$\Delta\eta$
Track quality	Tighter transverse impact parameter requirement ($|d_0| < 1$ mm)	d_0
	Number of b-layer hits (≥ 1) over the full η range	n_{BL}
TRT	Total number of hits in the TRT	n_{TRT}
Conversions	Number of hits in the b-layer (≥ 1)	n_{BL}
	Veto electron candidates matched to reconstructed photon conversions	

Figure 3.10: The loose, medium and tight electron identification cuts [41].

3.4.3   Muon reconstruction

Several groups exist that develop the muon reconstruction algorithms. They are divided into three main groups, STACO, MuTag [42] and Muid [43] . During 2013/2014 these have been merged into one common container, but the data samples used in this study did not yet have this container and therefore the STACO algorithms have been used. The STACO algorithm combines an inner detector track with a muon spectrometer track using a statistical method. On top of that two other algorithms are used: Muonboy and Mutag [42]. Muonboy starts from hit information in the muon spectrometer and produces standalone segments and tracks that are extrapolated to the interaction vertex. Mutag associates inner detector tracks with Muonboy segments. To clarify only inner detector tracks not combined in STACO will be used. The same counts for muon spectrometer only tracks that have not been combined with an inner detector track in STACO. More specifically the principle of the STACO method is the statistical combination of two independent measurements by means of their covariance matrices. A χ^2 test is made to determine how good the matching is [42]. The combination of inner detector tracks with the outer parts of the detector has lead to the "isCombined" tag used in selection. The efficiency of the muon reconstruction for both the STACO, Muonboy and Mutag are shown in figure 3.11.
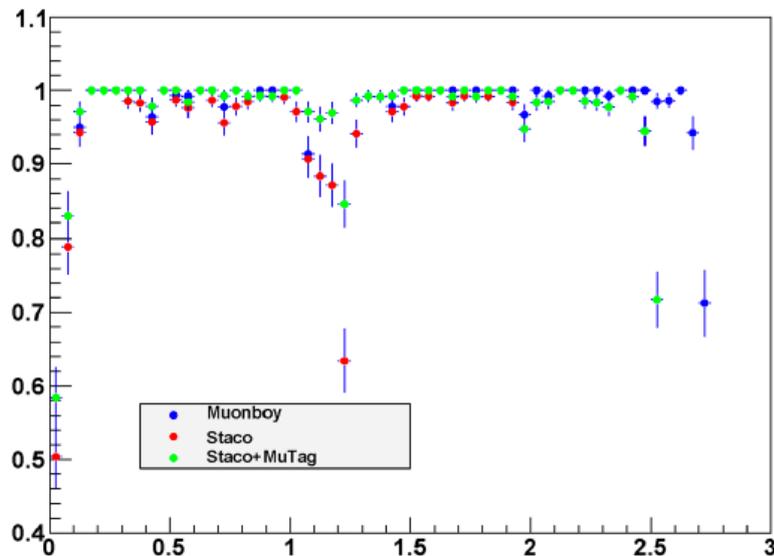
Figure 3.11: Efficiencies as function of $|\eta|$ from standalone and combined μ reconstruction algorithms, obtained on a single muon simulated sample of $p_T = 100$ GeV / c. [44]. The drops in efficiency is from regions where the Muon Spectrometer coverage is thin.

3.4.4   Missing transverse energy

Besides leptons the final state presented in this work contain neutrinos that will not be measured in the detector. Missing transverse energy, $\not{E}_T$ is calculated from the energy deposited in the calorimeters as well as from the muon detectors in ATLAS. The ATLAS cryostat - the container that hold the liquid argon needed for the EM calorimeters - causes a loss of energy

that is corrected for when reconstructing $\not{E}_T$. There are a set of different reconstruction methods depending on how the energy is read out from the calorimeter cells and what types of noise suppression methods are used [45].

When constructing the $\not{E}_T$, calorimeter cell readouts are associated with a parent object - a reconstructed, identified high-p_T object - in a chosen order: electrons, photons, muons, hadronically decaying tau-leptons, b-jets and light jets. A refined calibration of these objects are then used instead of the initial global calibration of cells allowing for a more accurate calibration and thus a better $\not{E}_T$ reconstruction.

This refined $\not{E}_T$ measure thus contains the sum of $\not{E}_T$ calculated from both interacting particles as well as muons that will generally not interact much in the calorimeters and from known detector effects.

Statistical procedure

4.1 Introduction

The statistical approach is important when finer details have to be studied. By finer details we could for instance refer to the small deviation from the Standard Model due to a new signal. It is very feasible that new physics will be at very low cross-sections which means statistical methods must be optimised for this, either by cutting away events characterised as background to the signal or by being able to separate the signal from background in one or more observables. The aim of this thesis is to fit the prediction of a range of Standard Model processes in several observables to data simultaneously in order to look for small deviations from the expected. The following chapter describes the statistical methodology used in the thesis.

4.2 Modelling of simulated processes

In any quantum theory, the observables characterising a particle physics reaction (e.g. A + B → X) will be stochastic variables. In order to predict their distribution, Monte Carlo methods are used for integrating elementary differential cross-sections over the relevant phase-space, but also for modelling the random perturbations from detector smearing. In this way, any physics process contributing to an observable will get an associated *probability density functions* (p.d.f.) of that observable, at least to the extent that quantum interference effects can be ignored.

4.2.1 Functional description

For some processes a simple analytical expression is available. The dilepton mass spectrum from Drell-Yan [46] is one such example well approximated by an exponential function (It, however, needs modifications to take into account detector effects [2]). In most cases a functional description of the process in a given observable is not available or is too time-consuming to calculate and other methods are used to model the data.

4.2.2 Histograms

One of the most widely used methods of creating a p.d.f for a process is to bin it as a histogram in one or more observables. Depending on the amount of data, the bin size of the histogram can be decreased to the limit where, given enough statistics, the histogram will resemble the actual functional description of the process if such a description exist. The fluctuations of each bin is expected to be Poissonian and well approximated by Gaussian fluctuations for more than 10 counts in the bin. In cases with low statistics or with large variations from bin to bin, histograms will not perform particularly well at modelling the process.

4.2.3 Kernel estimation - keys p.d.f's

Modelling of samples can in some cases be improved with the use of kernel estimation. This study utilises the ROOT *RooKeysPdf* [47] class to perform this modelling. The principle behind, is that a number of convoluted Gauss functions are used to model the process. The width and height of each Gauss function is proportional to the number of events it describes for a given volume step - in other words, they describe the density of the process. The number of Gauss functions needed to model the process is decided dynamically from a definition of what the minimal number of events per Gauss function is. The result is that volume steps with many events, e.g. a large density will have high and narrow Gauss functions that have a high impact on the overall functional description, whereas volumes with few events will have small and wide Gauss functions with a lower impact on the overall functional description. The *RooKeysPdf* has the option of using adaptive bandwidth i.e. width of the Gaussians to ensure optimal modelling and minimal bias.

This method is particularly interesting for the study of low statistics samples where improvements over binned histograms can be achieved.

4.3 Hypothesis Testing

In the end all analyses aim to test a given hypothesis. This can be of a specific model with a well modelled signal that can be searched for or it can be with a more general approach to look for deviations from a reference model. The use of well-understood and robust statistical methods are central to all of these. For the discovery of a new particle one would test the contributions from the reference model only, H_0 or null-hypothesis, against an alternate H_1 hypothesis involving new particles[1]. After a measurement one has to decide whether to reject or accept H_0. For model-independent searches the test is of the null-hypothesis alone.

4.3.1 Type I and II errors

To define the goals of a hypothesis test the Type I and Type II errors are introduced with the definition in figure 4.1a. It is meaningful to treat the two hypotheses asymmetrically as the null hypothesis is the "current best model" so the approach is normally to fix the allowed rate of Type II error, α. The well defined goal is then to maximise the *power*, (1-β), where β is defined as the rate of Type I error. The meaning of a "5σ" discovery would correspond to $\alpha = 2.87 \cdot 10^{-7}$. In other words there is a very small chance that the H_0 (i.e. the Standard

---

[1]The alternative model could equally well describe changes in amplitudes of known processes or changes to kinematic distributions due to e.g. the spin deviating from the expected.

Model) is rejected by mistake. The sensitivity of a measurement to new physics depends on the expected signal to background ratio in the given region and the ability to separate the two. Using counting experiments in one observable (see figure 4.1b) it is easy to determine which regions are sensitive but this is not always the case in higher dimensions. Figure 4.2 shows a few examples of distributions of two hypotheses in higher dimensions and the attempts to separate them by finding suitable discrimination criteria.

The Neyman-Pearson Lemma

There exist several approaches to determining whether a given hypothesis should be kept or rejected given observed data. One such hypothesis test is the Neyman-Pearson Lemma [48], states that the most *powerfull* test, the one maximising $1 - \beta$ among other alternative hypothesis tests, is based on the *likelihood ratio*:

$$
\begin{aligned}
\lambda(x) & = \frac{P(x|H_0)}{P(x|H_1)} \leq k_\alpha \\
& = \frac{\mathcal{L}(\theta_0|x)}{\mathcal{L}(\theta_1|x)} \leq k_\alpha,
\end{aligned}
\tag{4.1}
$$

where x represents the data and θ an estimator taking the value θ_0 under the null hypothesis and θ_1 under H_1. The test which rejects H_0 in favour of H_1 is:

$$
\lambda(x) \leq \eta,
\tag{4.2}
$$

where η is chosen such that:

$$
P(\lambda(x) \leq \eta|H_0) = \alpha.
\tag{4.3}
$$

The region Ω that minimises the probability of wrongly accepting H_0 is just a contour of this likelihood ratio and any other region of the same size will have less power. The intuitive understanding of the Neyman-Pearson lemma can be understood from looking again at figure 4.1b. Imagine the Gaussian distribution to the left is the H_0 and the Gaussian to the right is H_1. If we observe x=160 in the given observable, it makes sense to reject H_0. The critical region Ω in this one-observable case is simply the interval of x required to give the rate of rejection α so $P(x \notin \Omega|H_0) = \alpha$. Several other test-statistics exist, that summarise data in real-valued functions and the choice of test depends on the goal of the analysis.

4.4 Maximum likelihood estimation with multiple observables

This thesis will compare the modelled distributions of a series of processes resulting in lepton final states to data. This will be done for all processes and in more than one observable simultaneously. The rate of each process will be a parameter in the fit, that can be allowed to float and each kinematic distribution an observable in the probabilistic description of the process. To perform this fit the maximum likelihood estimate is used.

The probability of n events to appear in e.g. a given bin of a histogram can be described by the Poisson function:

$$
Pois(n|\nu) = \nu^n \frac{e^{-\nu}}{n!}.
\tag{4.4}
$$

The likelihood of ν given n is the same equation evaluated as a function of ν:

$$\mathcal{L}(\nu) = Pois(n|\nu). \tag{4.5}$$

It is a continuous function but not a probability distribution function as the name might otherwise hint. It is commonly plotted as $-2ln(\mathcal{L})$ as can be seen in figure 4.3. The resulting function is related to the χ^2 distribution and the global minimum of $-2ln(\mathcal{L})$ corresponds to the best fit of expectation to observation. In this example the only parameter allowed to float in the fit, would be the total number of excepted events, ν.

An important feature of the log-likelihood is that it is additive for independent measurements. Specifically the probability of a measurement in two independent observables is treated as two completely independent datasets so $P_{tot}(x) = P(x_{obs1}) \cdot P(x_{obs2})$ and the likelihood becomes

$$ln(\mathcal{L}_{tot}) = ln(\mathcal{L}_{obs1}) + ln(\mathcal{L}_{obs2}). \tag{4.6}$$

4.4.1 Multiple processes

Imagine the case where several processes contribute to the number of events in a given observable. Each process will be described by a p.d.f in that observable henceforth referred to as the *template* of that process in a given observable. The total number of predicted events in a given observable, is the sum of the total number of expected events for each template. The probability will be given by:

$$P(n|\nu_0 + \nu_1 + ... + \nu_p), \tag{4.7}$$

where ν_i is the expected number of events from the i'th template. When this is expanded to observables, $\{o_1, ..., o_M\}$ it can be described by what is normally referred to as the marked Poisson [50]. The term marked refers to the fact, that at each step (bin/event) the probability is build from a sum of distinct sources in a discriminating variable. The probably takes the form:

$$P(n|\nu_0 + \nu_1 + ... + \nu_p) = Pois(n|\sum_{i=0}^{p} \nu_i) \cdot \prod_{j=0}^{M} \frac{\nu_0 \cdot f_0(o_j) + ... + \nu_p \cdot f_p(o_j)}{\nu_0 + ... + \nu_p}, \tag{4.8}$$

where $f_i(o_j)$ is the probability density function of process ν_i in observable o_j. Taking the logarithm and rewriting, the likelihood function becomes:

$$ln\mathcal{L} = ln\left(Pois(n|\sum_{i=0}^{p} \nu_i)\right) + \sum_{j=0}^{M} ln\left(\frac{\sum \nu_i \cdot f_i(o_j)}{\sum_{i=0}^{p} \nu_i}\right), \tag{4.9}$$

The first term is what is referred to as the extended part. The actual implementation of the negative log-likelihood has been done in two different ways: A version, where both data and background templates are binned and an unbinned method where the probability of each event in the data is calculated from finely binned templates.

Introducing the fit parameters

Each process will be associated with a fit parameter that will determine the contribution from the process in all observables. The parameters will be defined as:

$$\frac{\alpha_i}{\sum_{i=0}^{p} \alpha_i} \cdot N_{exp,tot} = \nu_i, \tag{4.10}$$

such that α_i represent the fraction of total expected events ($N_{exp,tot}$) process i is contributing with. The sum over parameters should be one when the number of observed events is equal to the number of expected. It results in a unit probability when introduced in the last term of eq. 4.9:

$$\sum_{j=0}^{M} ln \left(\frac{\sum \nu_i \cdot f_i(o_j)}{\sum_{i=0}^{p} \nu_i} \right) \tag{4.11}$$

$$\rightarrow \sum_{j=0}^{M} ln \left(N_{obs} \sum_{i=0}^{p} \frac{\alpha_i \cdot f_i(o_j)}{\sum_{i=0}^{p} \alpha_i} \right),$$

where $f_i(o_j)$ are the templates normalised to unity. The probability in each observable is calculated per bin or event as described in the following:

Binned likelihood implementation

The binned likelihood is calculated using a Poisson for each bin. It has to take into account the different templates in each observable. As these are independent measurements they can be added in the log-likelihood, which takes the form:

$$
\begin{aligned}
ln\mathcal{L} \;=\; & ln \left(Pois(n| \sum_{i=0}^{p} \nu_i) \right) + \dots \\
& \dots + \sum_{j=0}^{M} ln \left[\sum^{bins} Pois \left(N_{obs,bin}, N_{obs,bin} \cdot \sum_{i=0}^{p} \frac{\alpha_i \cdot \nu_{o,i,bin}}{\sum_{i=0}^{p} \alpha_i} \right) \right],
\end{aligned}
\tag{4.12}
$$

where $\nu_{o,i,bin}$ is the number of expected events in the given bin in the template of the i'th process in the o'th observable. $N_{obs,bin}$ is introduced to remove any dependence on overall normalisation and simply fit the shape of templates to the observed. The binning of the template will affect the result quite significantly and a large number of bins is desirable to the extent that the bins are still adequately filled. Any bin with zero expectation will be excluded from the fit so any data points ending up in a bin with zero expectation will be removed in the fit, resulting in a loss of information, which leads to larger fit errors.

Unbinned likelihood implementation

In the unbinned likelihood the probability is calculated per event instead of per bin. The probability of each data-event to originate from either of the processes in a given observable is simply given by:

$$P(event|o_j) = \sum^{bins} \sum_{i=0}^{p} \frac{\alpha_i \cdot f_i(o_j)_{bin}}{\sum_{i=0}^{p} \alpha_i},$$

and the likelihood takes the form:

$$ln\mathcal{L} \;=\; ln\left(Pois(n|\sum_{i=0}^{p}\nu_i)\right) + \dots \tag{4.13}$$

$$\dots + \sum_{j=0}^{M} ln\left[\sum^{events} N_{obs,bin}\cdot\sum_{i=0}^{p}\frac{\alpha_i\cdot f_i(o_j)_{bin}}{\sum_{i=0}^{p}\alpha_i}\right],$$

where $f_i(o_j)_{bin}$ is the template p.d.f representing the probability for the data-event to originate from the i'th process in the bin it ends up in in the o'th observable.

4.4.2 Nuisance parameters

All processes will be affected by a range of systematic uncertainties in collider experiments. These uncertainties can be represented in the fit by what is generally referred to as nuisance parameters - additional parameters in the fit, that are allowed to vary to simulate a given uncertainty on the construction of a template. To put it simple, the likelihood becomes a function of both the parameters associated with the process in question, α and the nuisance parameters θ. The nuisance parameters must be constrained in order not to completely dominate the fit. It is assumed that the measured value of a given nuisance parameter, m, known to within $\pm 1\sigma$, is expected to fluctuate about θ, the estimated value of the nuisance parameter. The implementation is therefore done by adding Gaussian constrained nuisance parameters to the likelihood function:

$$P(x|\bar{\alpha}) \to P(x|\bar{\alpha}(\bar{\theta}))\cdot G(\bar{v}|\bar{\theta},\bar{\sigma}) \tag{4.14}$$

When performing fits that contain several processes, sometimes referred to as composition fits or composite fits, the fit results can be arbitrary if there is little or no separation power between the different processes in a given observable. In this case it is customary to constrain some of the process strengths, α, to be within experimental uncertainties of external measurements of these quantities.

4.5 p-value test of data to model expectations

One of the most simple, yet very effective, tests of the observed data to model expectations is the p-value test. It is a test of the data against a null hypothesis. In its most simple form in particle physics it can be used to test the number of observed events in each bin of a binned model p.d.f. The probability of n being equal to or greater than the number of observed events, n_{obs} is then:

$$p(n \geq n_{obs}) = \sum_{n=n_{obs}}^{\infty} f(n;\nu_{N.P.}=0,\nu_{sm}) \;=\; 1 - \sum_{n=0}^{n_{obs}-1} f(n;\nu_{N.P.}=0,\nu_{sm}) \tag{4.15}$$

$$= \; 1 - \sum_{n=0}^{n_{obs}-1}\frac{\nu_{sm}^{n}}{n!}e^{-\nu_{sm}},$$

where $\nu_{N.P./sm}$ is the excepted number of events from new physics phenomena divided by the contribution from the Standard Model. The last equal assumes each bin is Poisson distributed.

Notice that no knowledge of a signal expectation is needed as the test is of the likeliness of the null (Standard Model) to fluctuate to the observed value or above. This very simple test is a powerful tool for discovering deviations from the Standard Model expectations in data although the interpretation of a deviation might not be straightforward. It is also closely related to the approach for setting limits on a given model.

4.6 Claiming discovery and setting limits with likelihood ratios

A test statistic that maximises the power was in section 4.3.1 claimed to be the likelihood ratio of two hypotheses. The goal of all searches is to quantify the significance of any observed excess or set limits on new physics. If the signal strength of a hypothesised new signal is denoted μ, with $\mu = 0$ being the Standard Model and $\mu = 1$ the nominal signal strength of the new physics process, the *profile likelihood ratio* can be constructed as:

$$q_\mu = -2 \cdot ln(\lambda(\mu)) = -2 \cdot ln\left(\frac{\mathcal{L}(\mu, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right), \tag{4.16}$$

where $\mathcal{L}(\hat{\mu}, \hat{\theta})$ corresponds to the overall best fit (i.e. that maximises the likelihood) of signal, background and nuisance parameters and $\mathcal{L}(\mu, \hat{\hat{\theta}})$ the best fit for a given, fixed μ. This follows the ATLAS recommendations as elaborated in [51]. The use of q as the likelihood ratio test-static will be kept for the rest of the thesis. $\lambda(\mu)$ will take on values between zero and one with values near one implying good agreement between data and hypothesised value of μ. This means the test-statistic q_μ will grow with increased discrepancy between data and model.

In statistics, a test statistic independent of nuisance parameters is referred to as a pivot. The profile likelihood ratio satisfies this in the asymptotic limit which in the case of this thesis means having enough data to probably fill out the phase space probed. *Willks' theorem* [52] states that in the asymptotic limit the distribution of the test-statistic will follow a χ^2-distribution with n degrees of freedom, equal to the dimensionality of the tested parameter. For the signal strength μ this is one:

$$f(q_\mu) = f(-2 \cdot ln(\lambda(\mu))) = \chi^2_{Dim(\mu)}, \tag{4.17}$$

Wilks' theorem only holds for $\mu = \dot{\mu}$ i.e μ equal to the true value, $\dot{\mu}$. It can be shown (see [51] for the intermediate steps) that for the special case of $\mu = \dot{\mu}$ in the asymptotic limit, the cumulative distribution of q_μ is:

$$F(q_\mu|\mu) = 2\Phi(\sqrt{q_\mu}) - 1, \tag{4.18}$$

where Φ is the cumulative standard normal distribution (zero mean, unit variance). It is implicitly understood that $\Phi(x)$ is the integral from minus infinity to x. The p-value of μ for the observed q_μ follows to be:

$$p_\mu = 1 - F(q_\mu|\mu) = 2(1 - \Phi(\sqrt{q_\mu})), \tag{4.19}$$

and the significance of the observed signal strength μ:

$$S_\mu = \Phi^{-1}(1 - p_\mu) = \Phi^{-1}(2\Phi(\sqrt{q_\mu}) - 1). \tag{4.20}$$

The requirements for a discovery to be claimed in particle physics demands a significance of $S = 5$ corresponding to a p-value of the null-hypothesis only of $p = 2.87 \cdot 10^{-7}$ whereas exclusion of a possible new signal is done at 95% confidence level corresponding to a p-value threshold of the signal of $p = 0.05$ and $S = 1.64$. With this definition a discovery is based on the improbability of the null-only hypothesis and the degree of belief in the discovery should be dependent on the feasibility and impact of the hypothesised new physics. If the new model/theory drastically alters our view of the world a more stringent requirement should be set on the significance.

4.6.1 Testing the null-hypothesis

To test for a discovery the special case of $\mu = 0$ is important. Basically the test of disagreement with the null-hypothesis can be used as a test for a new signal, so testing if a given signal is likely to have zero signal strength can determine if it can be excluded. With a test-statistic q_0 defined for $\mu = 0$ as:

$$q_0 = -2 \cdot ln(\lambda(0)) = -2 \cdot \left(\frac{\mathcal{L}(0, \hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right), \tag{4.21}$$

with $q_0 = 0$ for $\hat{\mu} < 0$, as this test-statistic only concerns deviations due to an upward fluctuation in events. If a χ^2 distribution is assumed for the likelihood then (using eq. 4.19 and 4.20) the p-value is defined as (see [51] again):

$$p_0 = 1 - F(q_0|0) = 1 - \Phi(\sqrt{q_0}), \tag{4.22}$$

with corresponding significance:

$$S_0 = \Phi^{-1}(1 - p_0) = \sqrt{q_0}. \tag{4.23}$$

These can both be calculated analytically in the asymptotic limit, i.e.. if the test-statistic is χ^2 distributed. If not the distribution of the test-statistic must be generated through pseudo experiments, sampling the expected values and calculating the likelihood ratio. The p-value can then be calculated from the more general definition:

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu)dq_\mu, \tag{4.24}$$

where q_μ is the value of the test-statistic for the observed μ and $f(q_\mu|\mu)$ the probability density function of q_μ for the given μ.

4.6.2 Upper limits on μ

The limit on μ can be defined in several ways. It is meaningful to find limits within a given probability of the null to be true. The standard is to find the 95% upper limit - the limit where the p-value of the tested model is still above 0.05 for the given μ. The definition of the test-statistic will be the same as for q_0 with the exception that it is now set to 0 for μ below the best fit value as only upper limits are tested. With that definition the 95% upper limit on μ is found by finding the $\mu_{95\%}$ where

$$p_{\mu_{95\%}} = \int_{q_\mu,obs}^{\infty} f(q_{\mu_{95\%}}|\mu_{95\%})dq_{\mu_{95\%}} = 0.05. \tag{4.25}$$

The distribution of q_μ should in principle be sampled (generated using pseudo-experiments) for each value of μ to calculate the p-value. If the p.d.f of the test-statistic can be shown to follow a χ^2 distribution for one value of μ it is customary to assume it will for other values of μ. If this is the case eq. 4.19 can be used to calculate the p-value of q_μ for the tested μ and the 95% upper limit found with significant less computation. The actual implementation will be to iterate to find the expected $\mu_{95\%}$. A qualified guess for the starting value is the observed $\mu_{95\%}$.

4.7 Principal Component Analysis

The most widespread approach is to base fits on distributions in one observable. If a fit is done multidimensionally or in several event-observables simultaneously, however, it will potentially give a much more stringent comparison. The following describes a procedure for finding independent variables and introduces the Principal Component Analysis (PCA) method to generate linearly uncorrelated observables.

4.7.1 Correlation of observables

The correlation between two variables is a measure of the dependence of one on the other. Observables like the energy and momentum of a particle are sharing a large amount of information especially if the mass of the particle is low compared to the total energy. Treating the two as independent observables will likely result in erroneous fits as the same information enters twice in the fit and any differences are thus artificially enhanced. To combine several observables in a fit the observables will have to be independent or the degree of dependence known.

Variance

The variance is a measure for the spread of a set of discrete measurements. It is defined as the sum of squared deviations from the mean:

$$Var(x) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma^2, \tag{4.26}$$

where $\bar{x}$ is the mean of the sample and σ is the standard deviation. The *co*variance between two discrete measurements x and y is given as[2]:

$$cov[x, y] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}), \tag{4.27}$$

Linear correlation

The linear dependence between two variables x and y is commonly described by the Pearson product-moment correlation [54] coefficient or simply correlation coefficient defined for a series

---

[2]The factor n in the denominator should actually be n-1 if the data in question is a subsample of a larger population but it is left as n here for simplicity.

of measurements of x and y as:

$$\rho_{xy} = \frac{cov[x,y]}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}. \tag{4.28}$$

$\rho = 1$ means that the observables are completely correlated whereas $\rho = 0$ means they are linearly independent. Figure 4.4 show the correlation of a range of distributions.

4.7.2 Variance as a measure of information

The variance plays a central role in understanding the degree of information that can be obtained from an observable. The assumption valid for many experiments is that the axis in data, that has the largest variance, contain the largest amount of information. Imagine a two-dimensional sample as the one in figure 4.5. The variance calculated for the projection to the x axis will be roughly equal to the variance for the projection on the y axis. In the given distribution it is possible to define a new axis with much larger variance. The assumption valid for many experiments is that this axis will contain most of the signal, whereas the orthogonal axis after the new definition can be thought of as noise. With this definition of signal variance, σ_{signal}^2 and noise variance, σ_{noise}^2 the signal-to-noise-ratio (SNR) for a given observable, defined as

$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2}, \tag{4.29}$$

is high for the axes that most precisely describe the measured space. In other words the directions in data with largest variance should correspond to fundamental axes of interest with a large degree of information and low noise. Figure 4.5 is an example of data where an axis exists, that clearly describes the data much more precisely than one of the x and y axes.

Redundancy

Two measurements of the thickness of all chips (crisps) in a bag done with one ruler with distances in inches and the other in centimetres will in principle be independent. The distribution of one plotted against the other should, if noise is low enough, result in a very narrow straight line. In other words the two variables are completely correlated and no or very little new information is added by adding the second measurement. This is referred to as measurements with a high degree of redundancy equivalent to a high ρ. Measuring the thickness and diameter of chips could likely yield results similar to the distribution in the top centre of figure 4.4. Here there is a low degree of correlation and a low degree of redundancy as both the thickness and diameter of a chip carries information highly independent of the other.

Covariance matrix and Principal Component Analysis

The covariance can be generalised to n measurements of N variables. For each variable i the n measurements are described by the vector X_i. Let's define the matrix:

$$A = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \tag{4.30}$$

where each row contains the n-measurements of the deviation of variable N from its mean value, or in other words, each column is a set of measurements from one trial. The i,j'th entry of the covariance matrix will be:

$$C_{A,i,j} := cov[X_i, X_j] = \frac{1}{n} \sum_{event=1}^{n} (X_{i,event} - \bar{X}_i)(X_{j,event} - \bar{X}_j) \tag{4.31}$$

redefining X to be the relative distance:

$$X = \frac{X - \bar{X}}{\bar{X}}, \tag{4.32}$$

the short hand notation is

$$C_A \equiv \frac{1}{n} A A^T. \tag{4.33}$$

C_A will be a symmetric N x N matrix with the diagonal being the variance of particular types and the off-diagonal terms the covariance between measurements which will be zero for independent variables. The aim of a Principal Component Analysis is to find a matrix of the form in equation 4.33 where the off-diagonal elements are zero.

4.7.3 Principal Component Analysis

If the covariance matrix was calculated based on figure 4.5 the eigenvectors of the covariance matrix the off-diagonal entries would be positive, showing that the two variables increase together. The eigenvalues would hold information of the variance in the direction of the eigenvectors. This alone is actually enough to explain the principle of Principal Component Analysis - to find the eigenvectors of the covariance matrix with highest variance and use these to transform data from the original *pattern space* to a resulting *feature space*.

The covariance matrix C_A can be diagonalised by a matrix U_A such that:

$$U_A^T C_A U_A = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_N \end{pmatrix} \tag{4.34}$$

with

$$U_A = (\vec{v}_{e,1}, \vec{v}_{e,2}, \cdots, \vec{v}_{e,N}) \tag{4.35}$$

where $\vec{v}_{e,i}$ are the eigenvectors corresponding to the eigenvalues λ_i. The final step in Principal Component Analysis is to order these after the values of the eigenvalues so the first principal component is the eigenvector with largest eigenvalue. The use of the relative distance defined in equation 4.32 means that the Principal Component Analysis is independent of dynamic scales so it can compare observables in e.g. MeV and GeV without problems.

4.7.4   Test for non-linear correlations

Figure 4.4 illustrates how linearly uncorrelated observables can still be highly dependent. All distributions in the bottom row have clear symmetries and for several the actual functional dependence of one variable on the other can be guessed. Imagine the measurement of the height, y of a wave at a certain point for equal waves moving along the x-axis. If this was plotted against the time of the measurement a distribution similar to that bottom left in figure 4.4 could be achieved e.g. a sine or cosine distribution. There is no obvious linear correlation between the variables but it is not easy to decide the degree of redundancy either. The reason is of course that the passing wave's heights at given times are best described by a function, that cycles in time. The simple linear description is not adequate and PCA will most likely not give anything that better describes the data. There exist many types of non-linear correlations that can potentially spoil the positive effects of a PCA transformation. The following will describe a method to test for non-linear correlations that is used in this thesis.

Mutual Information

A more general notion of correlation comes from *Mutual Information* (MI) defined as (see e.g. [56]:

$$I_{X,Y} = \sum_{i,j} p_{i,j} \log \left(\frac{p_{i,j}}{p_i(X)\, p_j(Y)} \right). \tag{4.36}$$

Here $p_i(X)$ is the probability of $X = x_i$, the probability density function of X. $p_j(Y)$ is the probability density function of Y and $p_{i,j}$ that of the combined distribution of X and Y. This could for instance be the binned one-dimensional histograms of X and Y and the two-dimensional histogram of (X,Y). The MI-measure will be described in more detail later where results are generated for a series of non-linearly correlated variables. For now it should be noted, that MI can be defined both for discrete and continuos distributions and could be extended to higher orders, referred to as generalised redundancy, although this is computationally more demanding. The definition is on a p.d.f basis and the value of the MI measure is not directly comparable to linear correlation defined on an event-basis. It will, however, serve fine as an independent test of correlations.

		Actual condition	
		Accept	Reject
Decision	**Verdict of accept**	True positive	False positive (i.e. H0 accepted while false) **Type I error**
	Verdict of reject	False negative (i.e. reject H0 while true) **Type II errors**	True Negative

(a) The definition of error types in hypothesis testing.
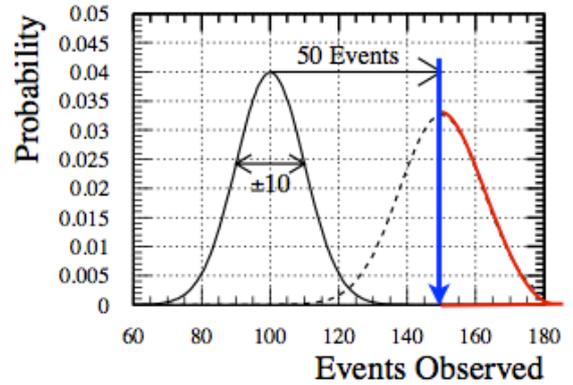
(b) A number-counting experiment. Figure from talk by G. Cowan at CERN statistics forum.

Figure 4.1: The Type I and II error definitions and a simple counting experiment.
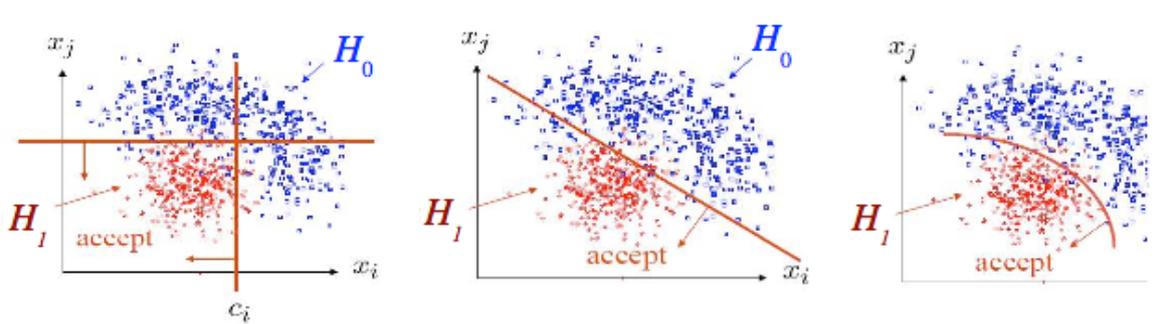
Figure 4.2: Hypothesis testing in higher dimensions challenges the method used for separating background and signal. Figure from talk by G. Cowan at CERN statistics forum.
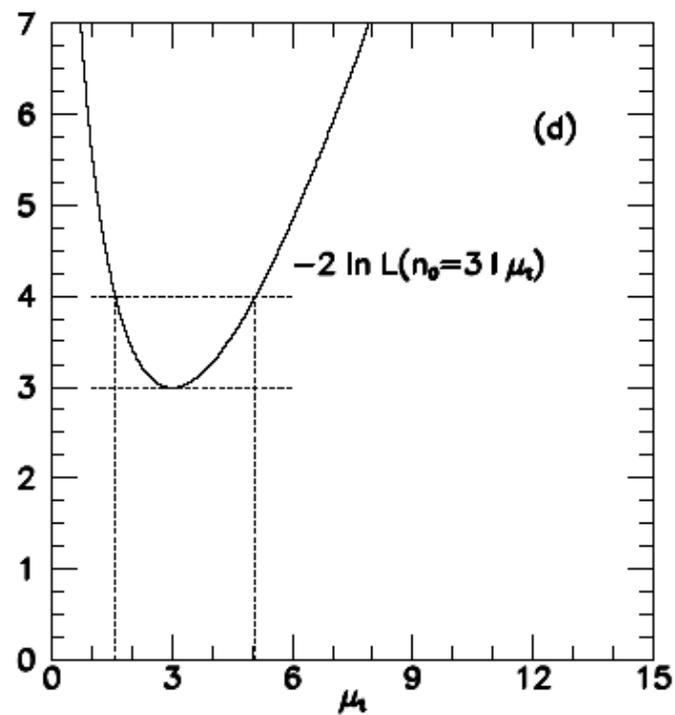
Figure 4.3: The likelihood of ν (μ) plotted as $-2ln(\mathcal{L})$ from [49].
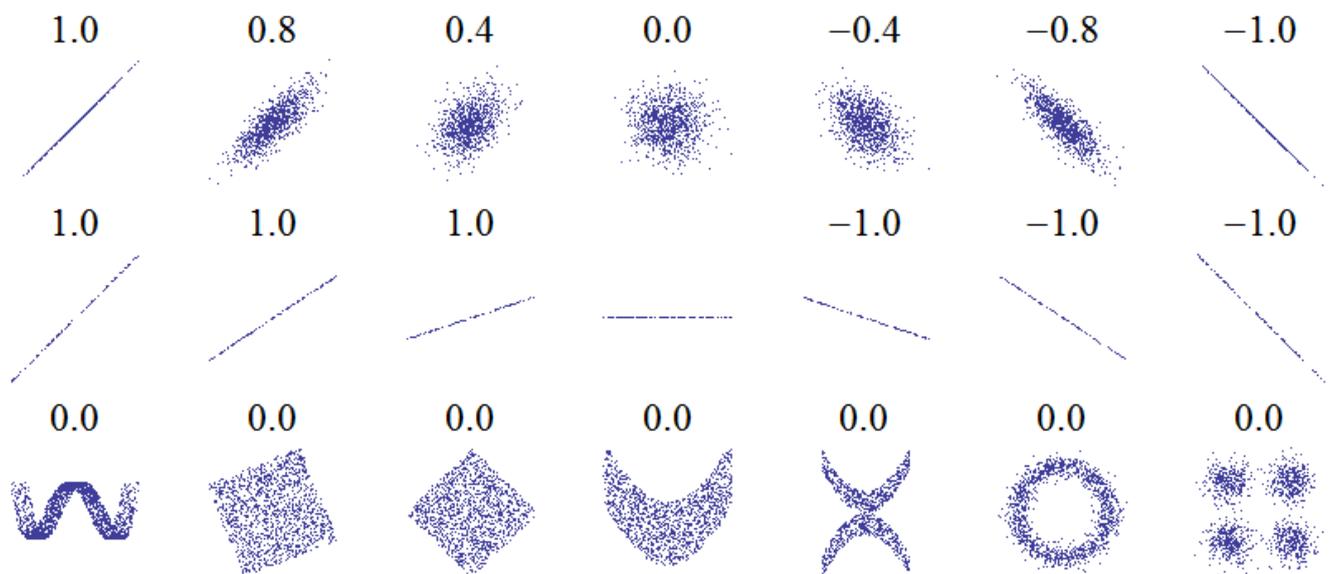
Figure 4.4: The figure shows the correlation between variables with varying degrees of linear correlation. A set of variables that are clearly not independent but with a correlation coefficient of 0 is shown. These all have non-linear correlations that are not described by Pearson's correlation coefficient. From [53].
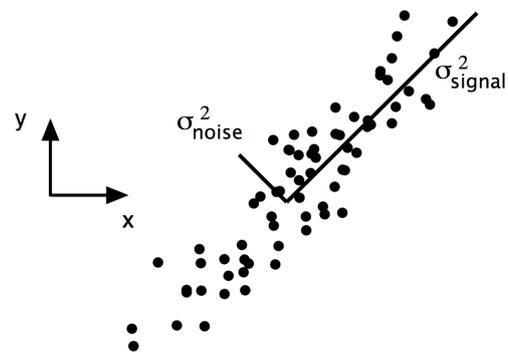
Figure 4.5: Simulated data of a camera measuring the movement of a spring. The camera is rotated, so neither the x nor y axis corresponds to the movement of the spring. For this case the axis describing the signal best, is the axis with greatest variance and largest signal to noise ratio. From [55].

4.7.5 Usage of Principal Component Analysis for Standard Model processes

The central approach for this thesis will be the transformation of simulated and real data using Principal Component Analysis. The transformation matrix will be generated once and applied to all processes including data. It can in principle be generated from any number of simulated processes and even from data but the choice affects the results significantly as will be discussed later in detail. In short the principal components' independence depends on the input to the Principal Component Analysis. The hypothesis is that data is transformed into a feature space that optimally characterises the given process. For the Principal Component Analysis transformation based on WZ, the same transformation on data will result in observable distributions that are optimised for fitting the WZ "part" of data or more generally for fitting any signal with three real, isolated leptons.

A note on likelihood implementation for Principal Component Analysis

Assuming the variables resulting from the transformation are independent they can be treated as independent measurements and all resulting variables can be fitted simultaneously. The likelihood functions described for both the binned and unbinned case remain as described earlier and both will work with Principal Component Analysis transformed events. It was however found during the studies that the unbinned likelihood is significantly slower to calculate. This is not surprising as all calculations are performed on a per-event basis opposed to the binned likelihood, which only calculates the likelihood contribution in each bin. For e.g. 20 bins compared to 2000 data-events it is easily more than a factor 100 in difference in the calculation time. As the unbinned approach was not found to improve results the binned likelihood will be used throughout this study.

Testing the fit procedure

5.1 A proof of concept with simulated data

Although Principal Component Analysis is not a new method, fitting Monte Carlo to data through the use of *principal components*, is a novel approach. It is therefore essential to test the robustness of the method. The following will describe a case study using simulated Standard Model processes. The first subsection will present the distribution of the *principal components* based on well understood Monte Carlo samples along with some measures to help illuminate the composition and information in the *principal components*. As the final goal is to fit model-expectation to data using *principal components*, a section is dedicated to the behaviour of fits in *principal components*.

5.1.1 Comparing original observables to *principal components*

The first benchmark test is made using two fully simulated ATLAS samples generated with Sherpa, one with WZ and the other with ZZ in the final state, normalised to 2600 events corresponding to the total number of data-events expected for an integrated luminosity of $20fb^{-1}$. The available Monte Carlo statistics are a factor 10 higher. A selection of three prompt leptons (electrons or muons) with a high degree of isolation has been performed leaving 1611 WZ and 989 ZZ events. The selection details will be presented in section 6.2 describing the signal region. The selection criteria are optimised for selecting leptons produced through electroweak decays or very similar processes. WZ is expected to be the largest Standard Model contribution to such a three-lepton final state and ZZ the second largest. As ZZ events also have three prompt, isolated leptons the ability to separate it from WZ is important if the contribution from each or either of the processes is to be correctly estimated through a fit to data.

Figure 5.1 shows the distribution of the missing transverse energy ($\not{E}_T$) and p_T of leading lepton [1] for the the WZ and ZZ samples. The bottom panel shows the ratio of the two samples both normalised to unity. The two observables are chosen to have one ($\not{E}_T$) with strong separation power and the other with weaker separation power[2]. The linear correlation

---

[1] Ranked by p_T

[2] The separation power referred to here is the linear separation power found from a linear (Fisher) discriminant.

between the two observables is 29% so a Principal Component Analysis will result in a first *principal component*, that combines information from both. The actual implementation is done through the use of ROOTs TPrincipal class [57]. The eigenvalues of the resulting *principal components* are plotted in figure 5.2. It shows the first *principal component* holds roughly 67% of the variance in the two input observables and the second the remaining 33%. As described in section 4.7.2 the variance is closely related to the information in a given observable and the word *information* will be used in this context henceforth.
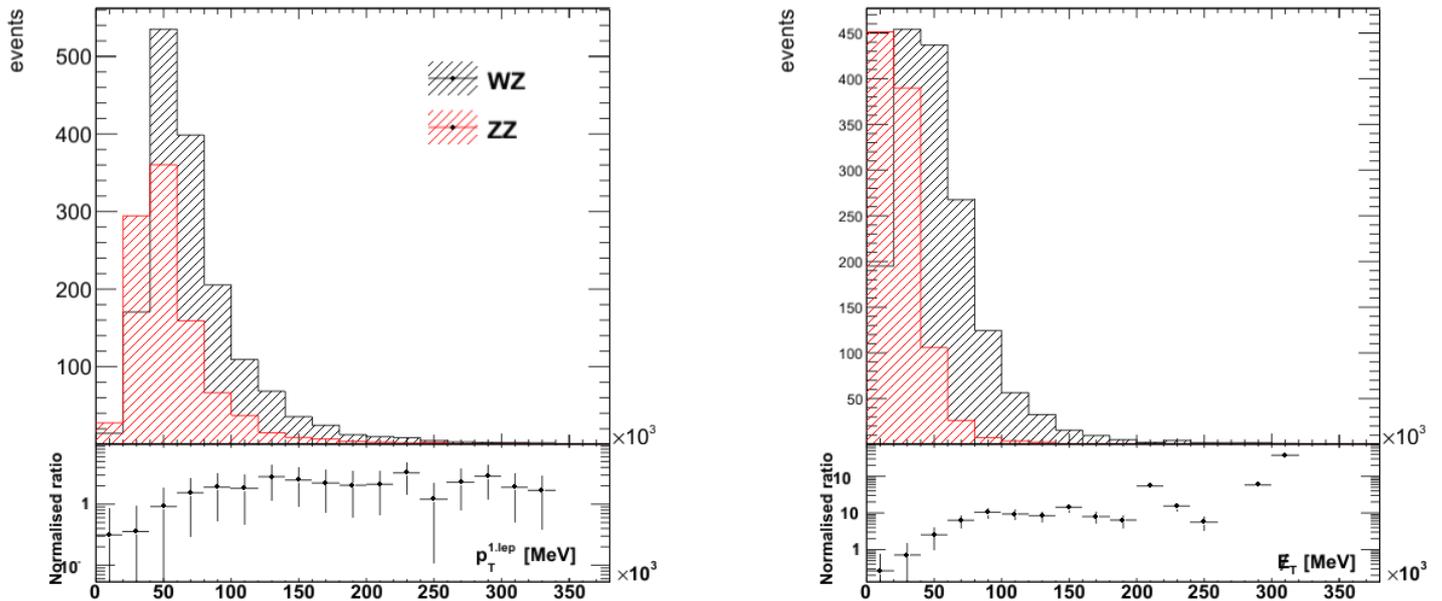
(a) The transverse momentum of the lepton with largest transverse momentum plotted for the two samples.

(b) The missing transverse energy in the event, based on energy reconstructed in calorimeters and corrected for energy in reconstructed objects.

Figure 5.1: The distributions of two observables in the WZ and ZZ samples. The bottom panel shows the ratio of the two samples both normalised to unity.

The transformation matrix to go from the original pattern space to the feature space, resulting from the Principal Component Analysis, consists of the eigenvectors and is given by:

$$
\begin{matrix} v_{e,1} & v_{e,2} \end{matrix} \qquad (5.1)
$$
$$
\begin{bmatrix} 0.873 & 0.487 \\ 0.487 & -0.873 \end{bmatrix}
$$

with eigenvalues 0.670 and 0.330 respectively. The original data has the values of the leading lepton p_T as the first row. The distributions of the resulting *principal components* are plotted in figure 5.3.

5.1.2   Fitting with PCA

In order to understand the behaviour of *principal components*, pseudo-data was sampled from the combined WZ and ZZ sample. A model with two processes, WZ and ZZ, is fitted to the
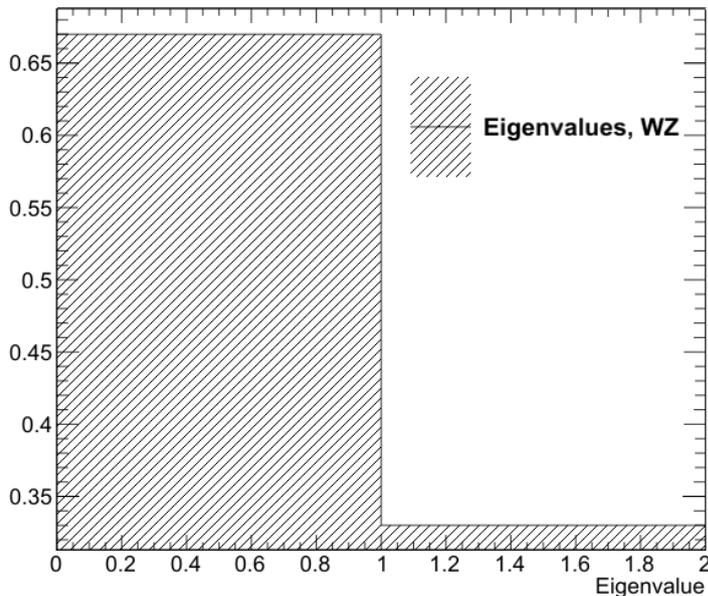
Figure 5.2: The eigenvalues resulting from performing a Principal Component Analysis on a simulated WZ sample. As described in section 4.7.3 the size of the eigenvalues correspond to the amount of variance associated to information from the original pattern space, the corresponding *principal component* spans.

pseudo-data with the fraction of WZ events as a free parameter. The full Monte Carlo statistics were used to determine the binned templates of the two processes. Two fits were performed, one minimising the total negative log-likelihood of eq. 4.12 and another minimising only the shape part. The results are in table 5.1. Of the two original observables, $\not{E}_T$ has the larger separation power and the fit to this observable using only the shape-part of the likelihood leads to a smaller statistical error on α_{WZ}, than the corresponding fit to the leading lepton p_T (The numbers in parentheses indicate the insignificant digit to be rounded). A fit to the highest ranking *principal component* yields a statistical uncertainty slightly lower, but comparable with the uncertainty from the $\not{E}_T$ fit. The smallest uncertainty, still using only the shape-part of the likelihood, is achieved by a combined fit to p_1 and p_2. It is worth noticing that the fit parameters, α, are defined such that their sum can deviate from one hence α_{WZ} is not given by fixing the expected contribution from ZZ. If only the extended part of the negative log-likelihood is minimised while keeping ZZ fixed at its expected (true) contribution, the uncertainty becomes $0.62 \cdot \sqrt{2600}/1611 = 0.02$ as returned by the fit. The fit including the overall normalisation has uncertainty dominated by this extended uncertainty and the reduction in statistical uncertainty from the shape-part does not affect the significant digits. In order to test the parameter estimate and assignment of uncertainties the pseudo-data bin contents are Poisson fluctuated around their expectation value and the fit repeated as described in the following section.

(a) The highest ranked principal component.

(b) The second principal component.
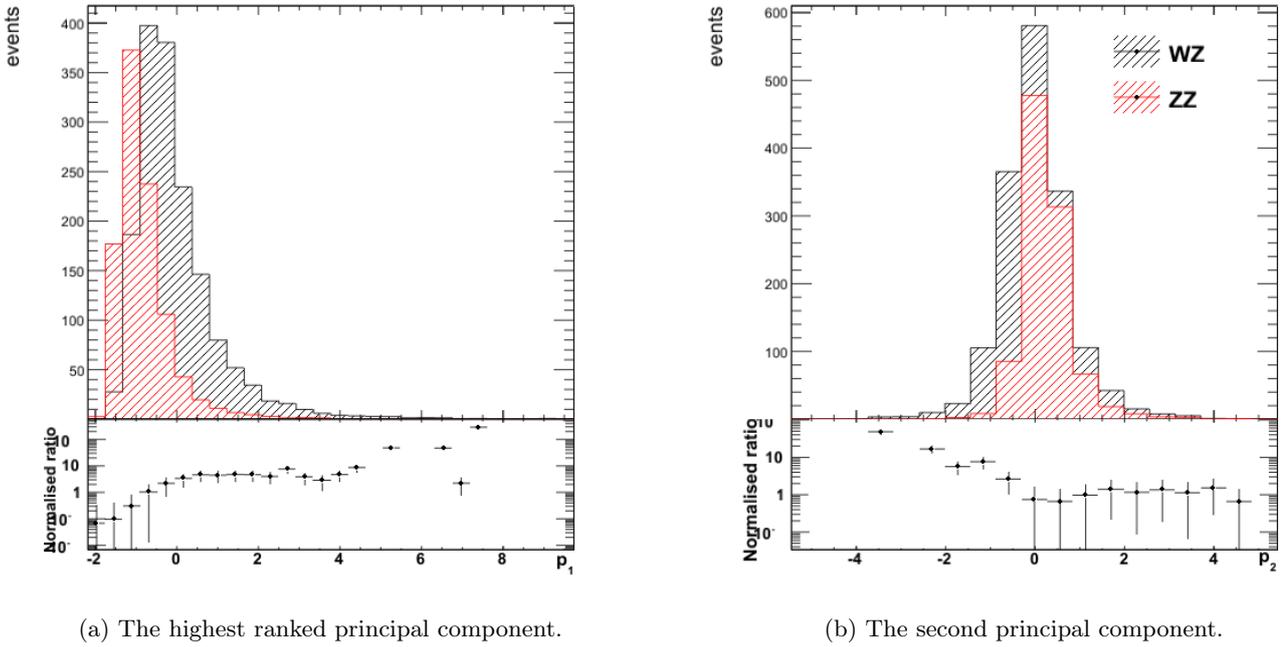
Figure 5.3: The *principal components* of the WZ sample with the same transformation performed on the ZZ sample plotted on top. The first and highest ranked principal component representing the linear component with the largest variance is p_1.

Binned shape-fit in 50 bins		
observable(s)	α_{WZ}	$\sigma_{\alpha_{WZ}}$
leading lepton p_T	0.62	$\pm\ 0.08(4)$
$\not{E}_T$	0.62	$\pm\ 0.05(2)$
p_1	0.62	$\pm\ 0.05(0)$
p_2	0.62	$\pm\ 0.09(3)$
p_1 & p_2	0.62	$\pm\ 0.04(4)$

Fit of overall normalisation alone		
any	0.62	$\pm\ 0.02(0)$

Extended, binned fit (including overall normalisation).		
leading lepton p_T	0.62	$\pm\ 0.01(9)$
$\not{E}_T$	0.62	$\pm\ 0.01(8)$
p_1	0.62	$\pm\ 0.01(8)$
p_2	0.62	$\pm\ 0.01(9)$
p_1 & p_2	0.62	$\pm\ 0.01(8)$

Table 5.1: The fit values and associated uncertainties for WZ fitted to pseudo-data in a WZ + ZZ sample. α_{WZ} is the fraction of WZ events in the sample and $\sigma_{\alpha_{WZ}}$ the associated uncertainty in the fit. The expected α_{WZ} is 0.62.

5.1.3 Pull distributions

The pull distribution is a useful measure to test the behaviour of a fit. It is defined as:

$$\frac{\alpha_{fit} - \alpha_{true}}{\sigma_{fit}}, \tag{5.2}$$

and should result in a unit Gaussian if the parameter and error estimation is sensible. When using ROOTs TMinuit [58] package to calculate fit uncertainties, the parabolic error estimate from the likelihood is returned per default. As described in [59, 60] the asymmetric errors calculated through MINOS in TMinuit represent a more correct estimate and they have been used for the following estimates.

To plot the pull distributions and test the fit 10000 pseudo-datasets were generated. Each time the overall normalisation was determined from a Poisson around the total 2600 expected events. The distribution of data in each observable was then sampled, assuming Poisson-statistics in each bin, with the normalisation fixed to the total number of events determined for each pseudo-dataset. This means that the *principal components* are fully uncorrelated per construction and the results will quantify the behaviour under that condition.

The resulting pull distributions are plotted in figure 5.4. The fits are made without the extended term concerning the overall normalisation. If the fit behaves correctly the distributions should be Gauss distributed around zero, with a width of 1. The pull-distributions for $\not{E}_T$, shown in figure 5.4a, is a unit Gaussian within uncertainties but the fit in leading lepton p_T, seen in figure 5.4b, is slightly skewed. The pull-distributions for the fits using p_1 and p_1 and p_2 combined also agree with expectations but p_2 in figure 5.4d is slightly off with a mean of 0.04 ± 0.01 and width 1.017 ± 0.007. The ratio of WZ to ZZ is flat in p_2 except for the tail to the left of the peak, where ZZ has limited statistics. This means the two distributions will be very difficult to separate in the fit.

Figure 5.5 show the distribution of fitted α_{WZ} from the pseudo-data samples. The width represents the uncertainty on the parameter estimate in the given observable. Figure 5.5b and 5.5d show that the parameter estimate in leading lepton p_T and p_2 do not follow a Gaussian distribution very well. The deviation in the pull distribution is $\sim 1\%$. The behaviour of the fit using both *principal components* is very sensible.

The extended part of the fit, which only concerns overall normalisation, is also fitted for the pseudo-data resulting in a mean of $\alpha_{WZ} = 0.619 \pm 0.020$.

5.1.4 Input observables

The two observables used for the study of fit results using the WZ and ZZ sample were somewhat arbitrarily chosen and a vast number of possible observables can be constructed in ATLAS data. The relevant observables, however, depend on the final state in question and the requirement that they must be suited as input for a Principal Component Analysis. As this thesis will concern final states with exactly three prompt and isolated leptons, the observables must characterise this type of events including possible jets and missing energy in the event. A series of observables have been tested in the multilepton group and found to be sensitive to a range of new physics phenomena (see e.g. [61, 62]). The following is a combination of these along with several other kinematic variables used to characterise events. Common to all of them is that, if they are to be used as input to a Principal Component Analysis, they must be continuous and defined everywhere.
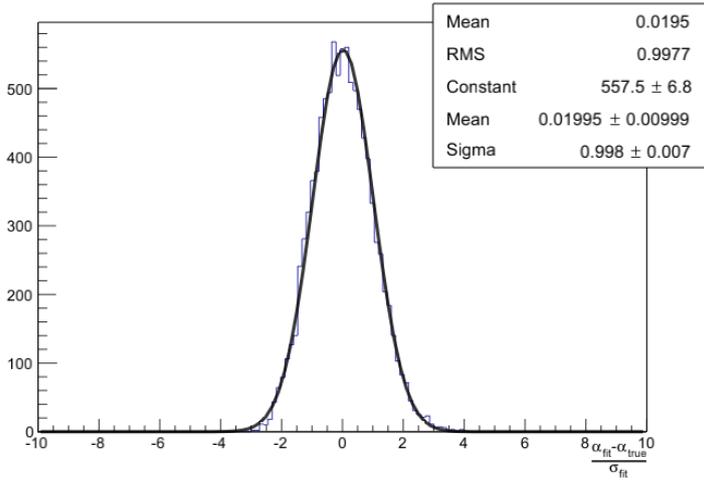
(a) The pull-distribution of α_{WZ} in $\not{E}_T$
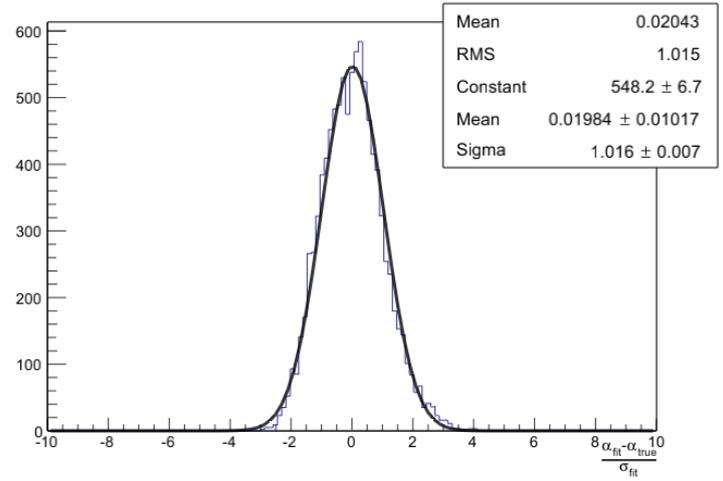
(b) The pull-distribution of α_{WZ} in p_T of leading lepton
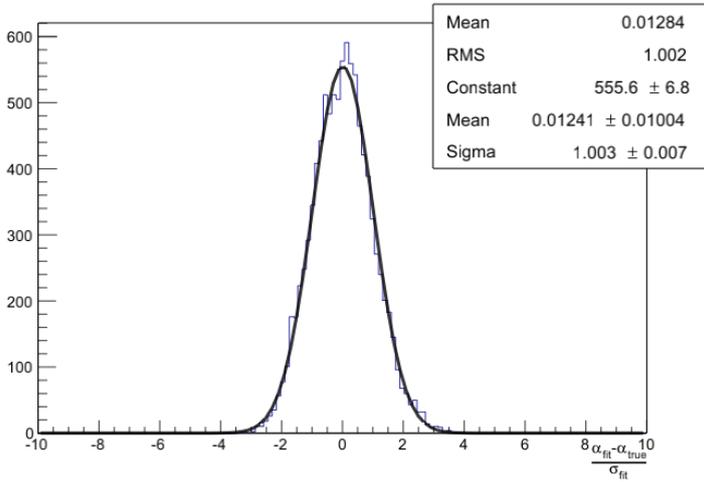
(c) The pull-distribution of α_{WZ} in p_1
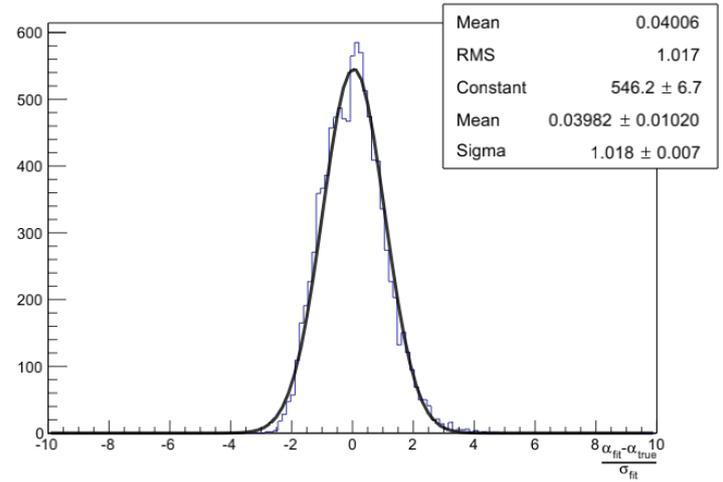
(d) The pull-distribution of α_{WZ} in p_2
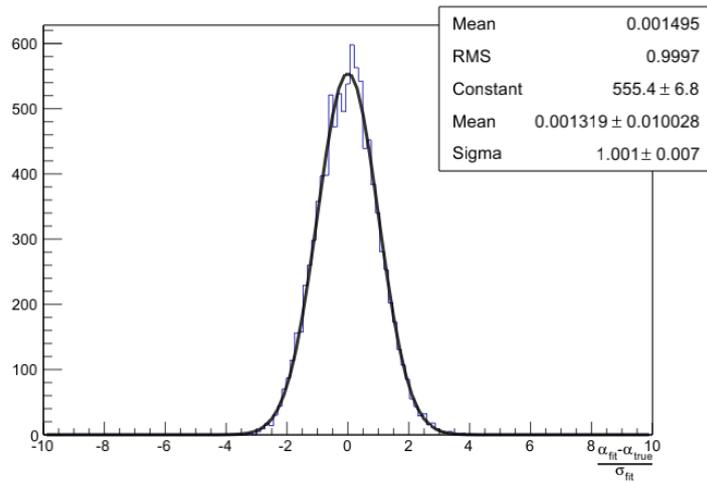
(e) The pull-distribution of α_{WZ} in p_1 and p_2

Figure 5.4: Pull distributions for binned shape-fits.
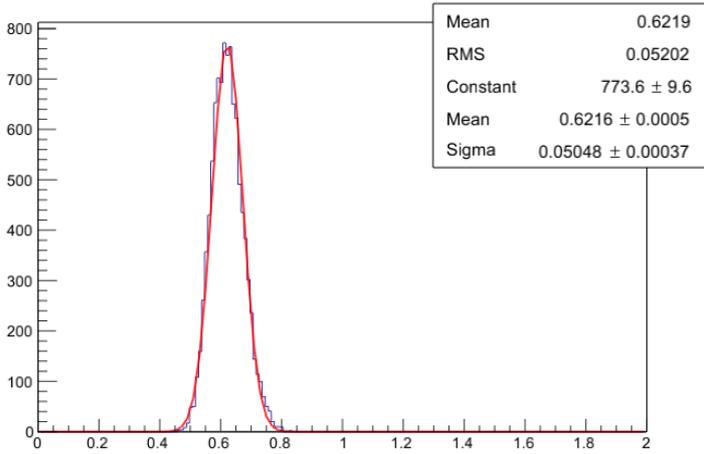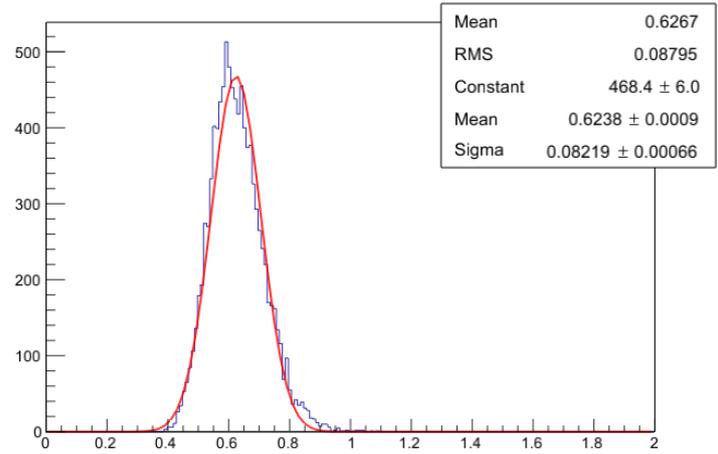
Mean	0.6219
RMS	0.05202
Constant	773.6 ± 9.6
Mean	0.6216 ± 0.0005
Sigma	0.05048 ± 0.00037

(a) The distribution of α_{WZ} in $\not{E}_T$

Mean	0.6267
RMS	0.08795
Constant	468.4 ± 6.0
Mean	0.6238 ± 0.0009
Sigma	0.08219 ± 0.00066

(b) The distribution of α_{WZ} in p_T of leading lepton

Mean	0.622
RMS	0.05073
Constant	793 ± 10.0
Mean	0.6214 ± 0.0005
Sigma	0.04931 ± 0.00037

(c) The distribution of α_{WZ} in p_1

Mean	0.6275
RMS	0.09812
Constant	429.2 ± 5.6
Mean	0.6237 ± 0.0011
Sigma	0.08771 ± 0.00071

(d) The distribution of α_{WZ} in p_2

Mean	0.6219
RMS	0.04461
Constant	908.8 ± 11.3
Mean	0.6216 ± 0.0005
Sigma	0.04284 ± 0.00031
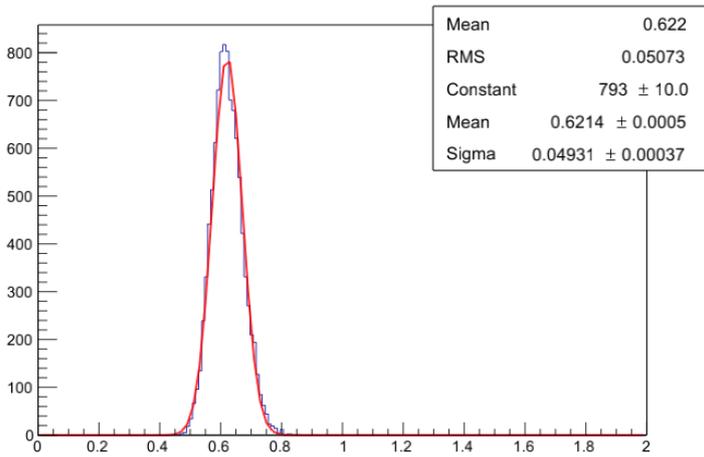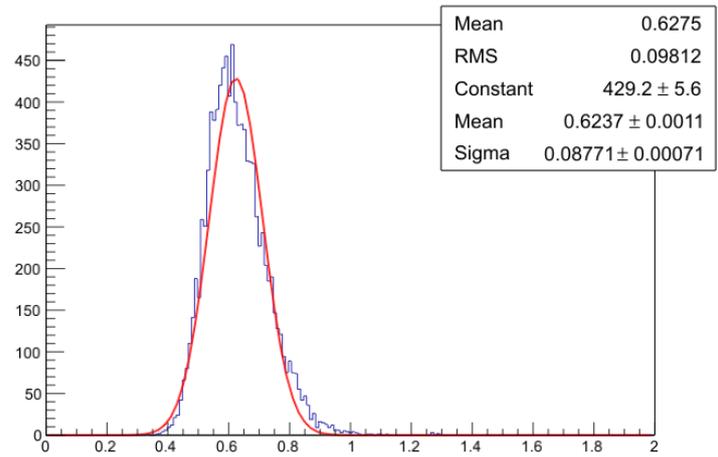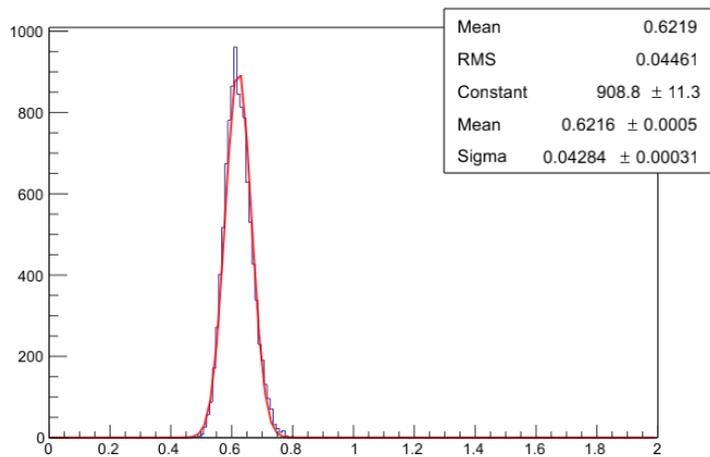
(e) The distribution of α_{WZ} in p_1 and p_2

Figure 5.5: Fitted α_{WZ} distributions for the different observables and the associated uncertainty assuming a Gaussian distribution.

Leptons

The first obvious choice is the kinematic information of the leptons. The p_T of each lepton is used and defined by the leptons ranked by p_T as with the leading lepton p_T observable already presented. To have an observable sensitive to resonances from new physics the transverse mass of the three leptons is included. Through the work in the multilepton group leading to [61] it was found that the invariant mass of two same-flavour leptons with opposite sign (charge) (SFOS) closest to the Z-mass is also sensitive to a range of new physics. It effectively separates events with a Z-like particle from events without. To ensure continuity the definition is the invariant mass of either 2 SFOS leptons or all three leptons closest to the Z-mass. This ensures continuity in both the case where a SFOS pair is not present and in the extreme case where three electrons or muons with at least one opposite charge are present in the event.

The angular information regarding the leptons constitute numerous observables, if the information from the objects are combined, but not all information is equally interesting. The tradition in ATLAS is to define the angular separation between objects, ΔR as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{5.3}$$

where $\Delta\eta$ is the difference in η between the two objects and $\Delta\phi$ the ϕ difference. This holds much more information than the individual angles. The smallest angular separation between leptons is especially interesting as it holds information that characterise the two leptons from the Z and clearly separates these from a range of other possible production mechanisms. For this reason the smallest ΔR between leptons has been used as angular observable in this study.

Jets

Jets can be present in the events either from initial- or final state radiation in WZ events or from other processes. There will, however, not be a requirement on the number of jets, so the kinematic observable concerning individual jets cannot be used. The total number of jets and summation of energy of jets can be used as these will be continuous distributions and the sum of jet transverse momenta (H_T^{jets}) was chosen as this contains information on both the momentum and the number of jets.

Furthermore an observable describing the probability for a b-jet to be present is used. This is the maximum value of the ATLAS MV1 variable. The MV1 variable is a combined probability constructed from several individual jet-observables using boosted decision trees and has a value between 0 and 1 for each jet. It will be continuous down to 0 which is also the value events without jets will be given.

Missing energy

The final object-type in the event, is the calculated missing energy. The observable used for this is simply the transverse component of the missing momentum vector.

Event observables

Besides the kinematic observables concerning the objects in the event, the total number of tracks from the primary vertex is included. This was chosen for its possible separation power between diboson events and the boson-jet background as well as its possible sensitivity to high-multiplicity new physics like black holes in extra spatial dimensions.

5.1.5 PCA with all observables

The correlation between input observables is shown in figure 5.6a. The clear correlations seen indicates that a linear combination should be able to include much of the information in a fewer number of *principal components*. Performing a PCA based on the WZ sample results in *principal components* with the eigenvalues in figure 5.6b. The first *principal component*, p_1 has 36% of the total information of the 10 input observables. The following *principal components* contain from 13% down to a few % of the total information. The eigenvectors are:

$$
\begin{array}{cccccccccc}
v_{e,1} & v_{e,2} & v_{e,3} & v_{e,4} & v_{e,5} & v_{e,6} & v_{e,7} & v_{e,8} & v_{e,9} & v_{e,10}
\end{array}
\tag{5.4}
$$

$$
\begin{bmatrix}
0.43 & 0.12 & -0.16 & 0.11 & 0.08 & -0.13 & 0.35 & -0.51 & -0.09 & 0.59 \\
0.39 & 0.21 & -0.05 & 0.04 & 0.05 & -0.35 & -0.48 & 0.57 & 0.02 & 0.35 \\
0.26 & 0.02 & 0.47 & -0.50 & -0.18 & 0.43 & 0.30 & 0.28 & 0.12 & 0.23 \\
0.31 & -0.38 & 0.24 & -0.13 & -0.08 & 0.19 & -0.66 & -0.45 & -0.07 & -0.03 \\
0.11 & -0.49 & 0.29 & 0.24 & 0.74 & -0.02 & 0.18 & 0.18 & 0.02 & 0.00 \\
0.19 & -0.19 & -0.58 & 0.26 & -0.05 & 0.67 & -0.02 & 0.24 & -0.11 & 0.05 \\
0.14 & 0.42 & 0.43 & 0.63 & -0.01 & 0.34 & -0.15 & -0.08 & 0.22 & 0.15 \\
0.13 & -0.48 & 0.18 & 0.43 & -0.63 & -0.28 & 0.22 & 0.14 & 0.03 & -0.01 \\
0.47 & 0.12 & -0.15 & 0.02 & 0.05 & -0.04 & 0.09 & -0.09 & 0.72 & -0.45 \\
0.43 & 0.32 & 0.19 & 0.13 & 0.01 & 0.01 & 0.12 & 0.04 & -0.62 & -0.50
\end{bmatrix}
$$

for the input observables:

$$
\begin{bmatrix}
p_{T,lead}^{lep} & p_{T,second}^{lep} & p_{T,third}^{lep} & H_T^{jet} & MV1_{max} & \not{E}_T & \Delta R & nTracks_{vertex} & M_{inv}^{SFOS} & M_T^{3lep}
\end{bmatrix}
$$

The distribution of WZ and ZZ in the resulting *principal components* is shown in figure 5.7 and 5.8.
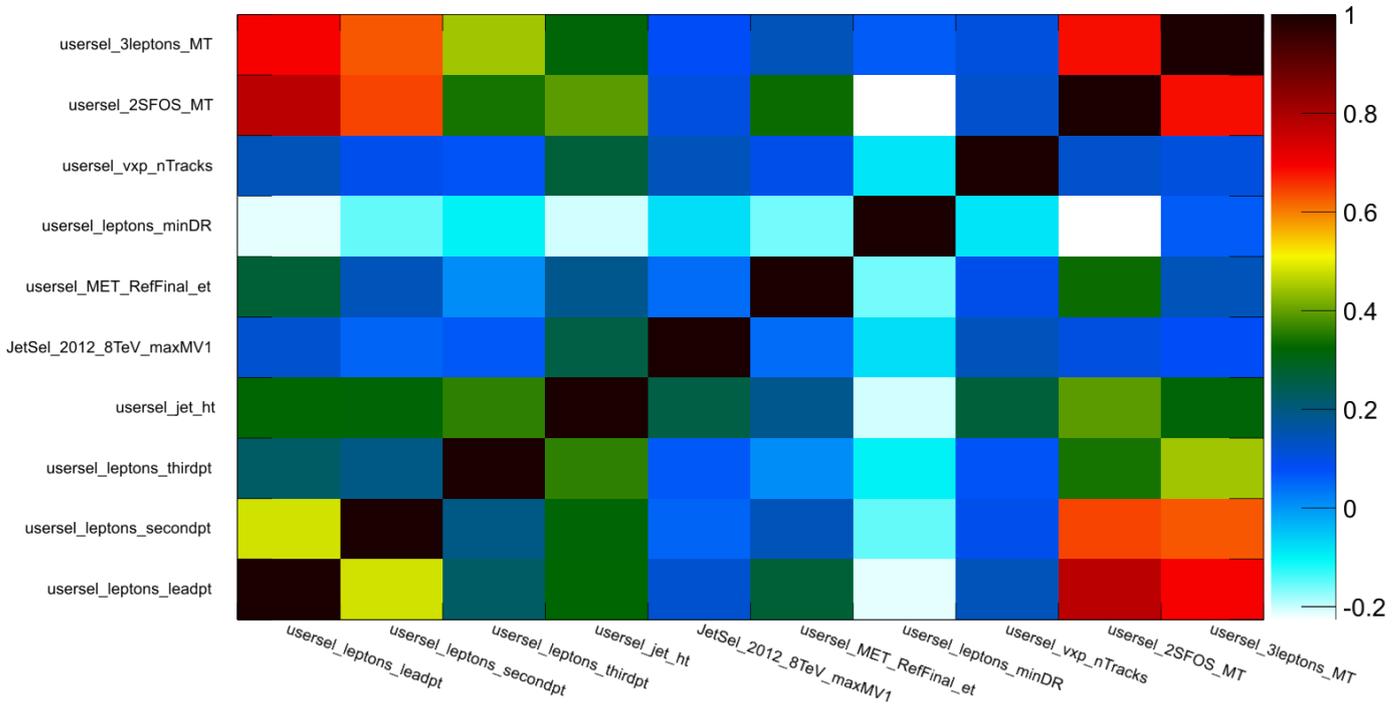
(a) The correlation coefficient between calculated for each pair of input observables.
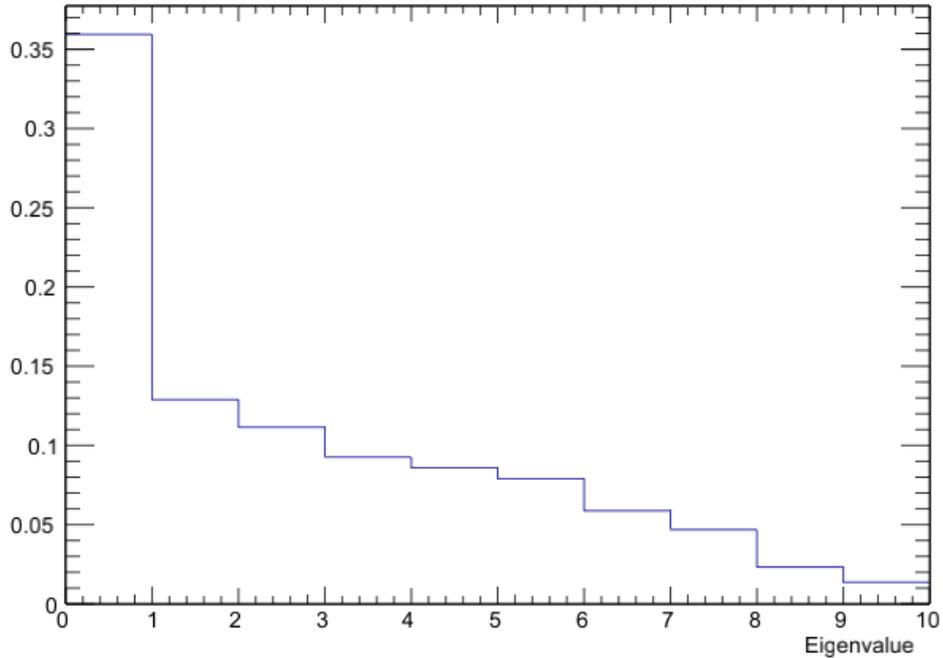
(b) The eigenvalues resulting from performing a Principal Component Analysis on a simulated WZ sample with 10 input observables. As described in section 4.7.3 the size of the eigenvalues correspond to the amount of information from the original pattern space, the corresponding *principal components* spans.

Figure 5.6: Correlation of input observables in the WZ sample and eigenvalues of resulting *principal components*.
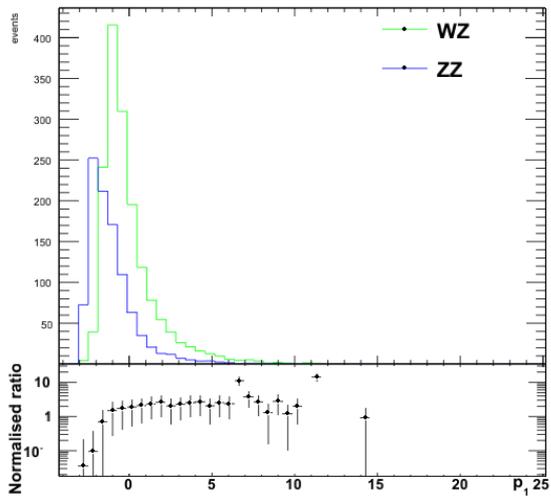
(a) PC 1
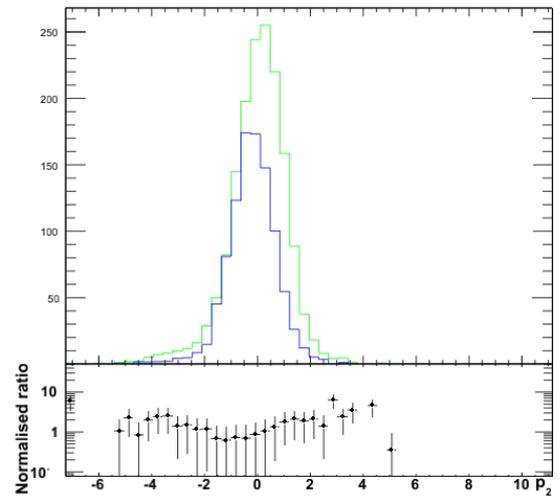
(b) PC 2
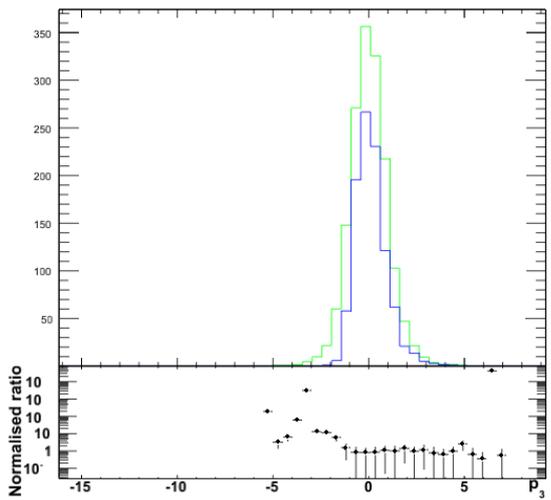
(c) PC 3
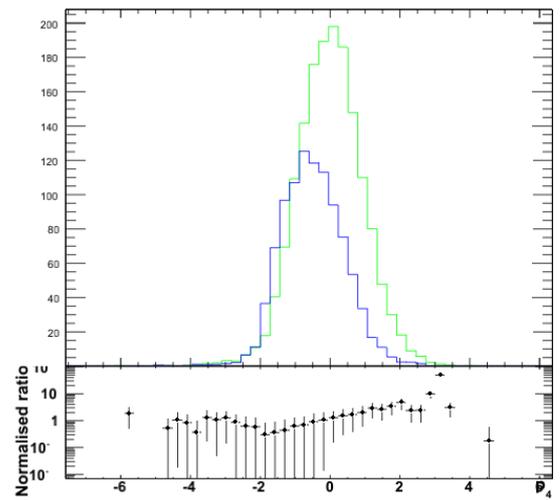
(d) PC 4
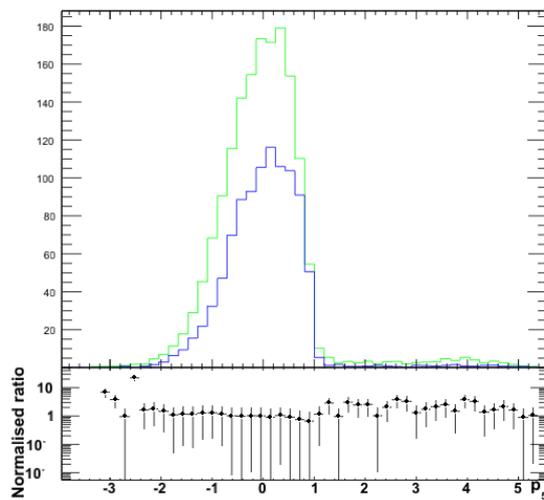
(e) PC 5

Figure 5.7: The first 5 *principal components* of the WZ sample constructed from the 10 input observables. The same transformation is performed on the ZZ sample which is plotted on top. The first and highest ranked principal component is p_1.

(a) PC 6

(b) PC 7

(c) PC 8

(d) PC 9

(e) PC 10
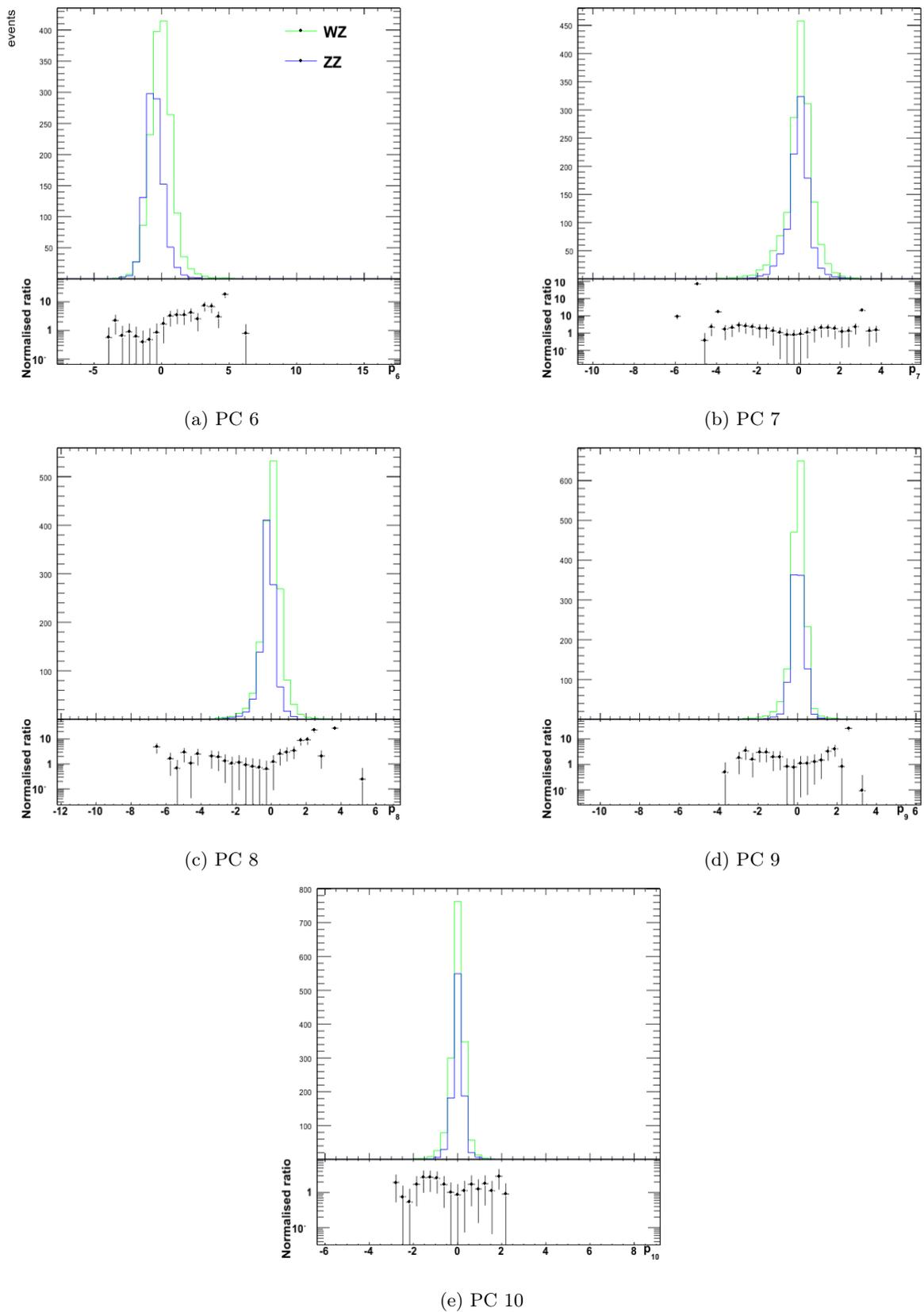
Figure 5.8: The last 5 *principal components* of the WZ sample constructed from the 10 input observables. The same transformation is performed on the ZZ sample which is plotted on top.

5.1.6   Fitting with all *principal components*

Pseudo-data was again generated 10000 times and the fraction of WZ events fitted for each pseudo-sample, both for the individual *principal components* and the combination of them. The pull-distributions were generated, as was the fitted parameter values and uncertainty. The pull distributions are included in appendix A and summarised here. The fit of one *principal component* at a time is the most stringent test of the behaviour of the fit. If separation between the two processes is not possible in the given variable, the returned fit uncertainty should be large. For p_5,p_7,p_9 and p_{10} this is to some extent the case. The pull-distributions are asymmetric with a tail towards larger values. The overall fit results in widths close to 1 and mean 0 but when the Gaussian peak is fitted alone the width is smaller than one by 14%,10%,6% and 4.5% respectively. The uncertainty returned is quite large for all these four *principal components*. Looking at figure 5.7 and 5.8, the ratio of WZ to ZZ in p_5,p_7,p_9 and p_{10}, are all close to constant for the bulk of the distributions. This means that any separation between the two processes must be found in the limited statistics of the tails which are sensitive to fluctuations causing the pull-distributions to deviate from a unit Gaussian.

For each *principal component* the mean value and 1-sigma width of the distribution of α_{WZ} returned by the fit was calculated and the numbers are given in table 5.2. This was done both for the one *principal component* at a time and for the combination of several *principal components*. Furthermore the normalisation was first left out, to study the effects of fitting the shape of the two samples in several distributions. The last part of the table is of the combined fit of *principal components* including information of both the shape and the overall normalisation.

The fit uncertainty from the 10000 fitted pseudo-data samples is plotted in figure 5.9 for fits without overall normalisation. First it should be noted how the uncertainty, as expected, is reduced with the addition of each new *principal component*. As each *principal component* contains less new information the reduction of uncertainty should be smaller with each step. This is indeed the case. The fit-uncertainties are of course affected by the separation power between the two samples in each variable and not only by the eigenvalue of the given *principal component*. The variance of a weighted sample mean, $\bar{X}$ is given as:

$$\sigma_{\bar{X}}^2 = \sum_{i=1}^{n} \omega_i^2 \sigma_i^2. \tag{5.5}$$

In the case of complete independence and 100% new information (i.e. identical variance and eigenvalue) in each observable, the weights must be equal and it follows that

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}. \tag{5.6}$$

The red line in figure 5.9 is the expected uncertainty assuming equal weights and complete independence. The actual fit uncertainties are larger as the first *principal component* has the majority of information in the fit.

Sum of square residuals - a distance measure

Another way to quantify the effects of the Principal Component Analysis is to look at a distance measure between the original data and the space spanned by the principal components, *the sum of square residuals*, SQR. The value is calculated as the 'distance' from a space

Shape-fit of one *principal component* at a time

PC	1	2	3	4	5
α_{WZ}	0.622	0.641	0.649	0.628	0.689
σ_{WZ}	± 0.056	± 0.159	± 0.177	± 0.097	± 0.321
	6	7	8	9	10
α_{WZ}	0.626	0.668	0.630	0.676	0.674
σ_{WZ}	± 0.0808	± 0.242	± 0.110	± 0.283	± 0.297

Shape-fit of *principal components* combined

# PCs	1	2	3	4	5
α_{WZ}	0.622	0.622	0.622	0.622	0.622
σ_{WZ}	± 0.0557	± 0.052	± 0.049	± 0.044	± 0.043
	6	7	8	9	All 10
α_{WZ}	0.622	0.621	0.621	0.622	0.621
σ_{WZ}	± 0.038	± 0.037	± 0.035	± 0.035	± 0.034

Fit of shapes and overall normalisation for *principal components* combined

# PCs	1	2	3	4	5
α_{WZ}	0.619	0.619	0.620	0.620	0.620
σ_{WZ}	± 0.0188	± 0.0182	± 0.0182	± 0.0180	± 0.0180
	6	7	8	9	All 10
α_{WZ}	0.620	0.620	0.620	0.620	0.620
σ_{WZ}	± 0.0175	± 0.0174	± 0.0172	± 0.0169	± 0.0170

Table 5.2: Fit values for a shape-fit without overall normalisation to first one *principal component* at a time and secondly to the combined *principal components*. The values are obtained from the mean and width of the parameter distribution from 10000 fits to pseudo-data. An extra digit has been included for the combined fit to see changes with each new *principal component*.
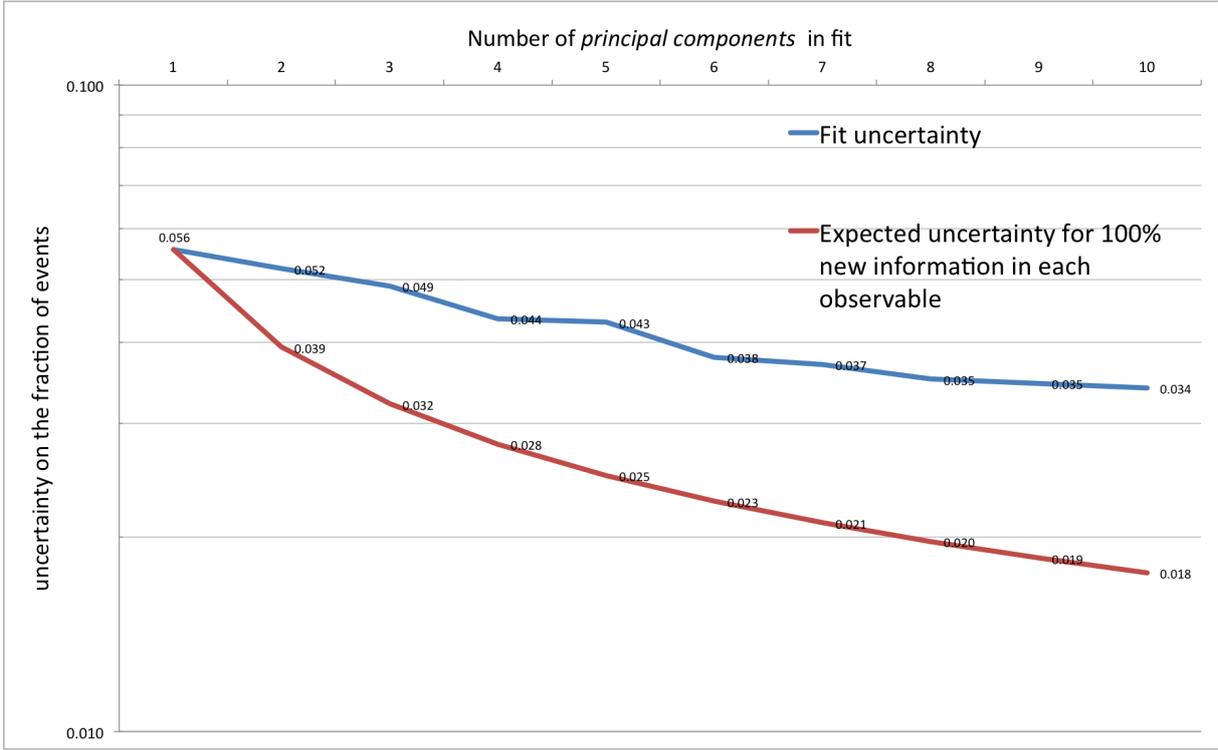
Figure 5.9: The uncertainty found through fits to 10000 pseudo-data samples for the combination of several *principal components* using the shape information in each variable but not overall normalisation.

spanned by a subset of *principal components* to the original data and it effectively describes how much phase-space is not described by the *principal components*. The definition is as follows for each event. For a total of P parameters and $N = P$ principal components the distance between the space spanned by the $m \leq N$ principal components is defined for the data vector X as:

$$E_m = \sum_{i=1}^{P} \left(X_i - X_i' \right)^2, \tag{5.7}$$

where X' is the value in original patternspace obtained from performing the inverse transformation from feature space, based only on m principal components.

$$X_i' = \sum_{j=i}^{m} p_i e_{i,j}, \tag{5.8}$$

E_m is calculated for each event and the sample average for the WZ sample is shown in figure 5.10. The X-axis denotes the number of *principal components* included in the calculation of the SQR and ranges from 1 to 9 of the 10 *principal components*. A large distance measure indicates, that there is still a significant amount of the data not described by the space spanned by the m principal components. The Y-axis is the E_m sum of all events in the sample normalised to the
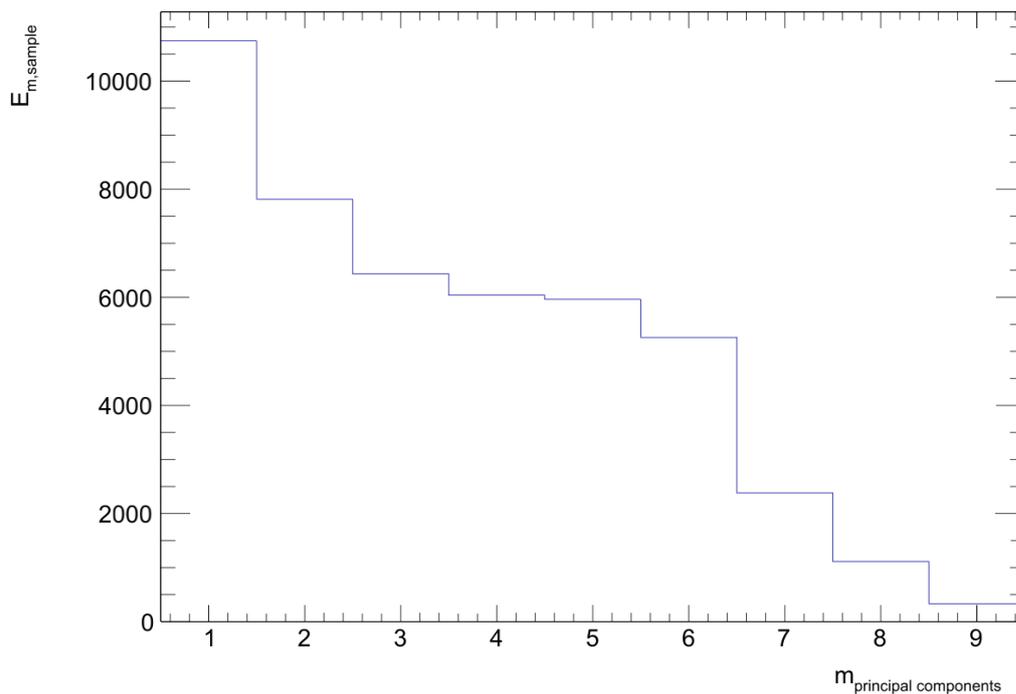
Figure 5.10: The sum of squared residuals - a measure of the distance from the original data to the space spanned by the $m \leq N$ principal components.

number of entries. The SQR is useful for reduction of dimensionality and figure 5.10 shows that the inclusion of p_2 and p_3 both results in a significantly smaller SQR, while the next large reduction is with the inclusion of p_6 and especially p_7. The last *principal components*, however, have small eigenvalues and will likely not improve the fit significantly as supported by the results in table 5.2.

5.2 A study of the method in real data

The pseudo-data generated from Monte Carlo samples will resemble the simulated data quite closely even after sampling. It is, however, interesting to see how an analysis using Principal Component Analysis behaves when detector smearing and model inaccuracies enter the equation. To test this a region of data, not overlapping with the final signal region, has been used. The region must be one, where new physics overlapping with the signal region are not believed to occur. For a signal region of three prompt leptons, highly isolated and with high p_T, one orthogonal region is the region with exactly two leptons that satisfy all the same selection cuts on the leptons with the additional requirement that the leptons must have same flavour and charge to suppress processes that can produce three lepton final states [3]. To suppress potential new physics (resonances as e.g. doubly-charged Higgs [63] or right-handed leptons from the seesaw [19]) the sum of jet p_T is required to be below 500 GeV and the transverse mass of the two leptons to be below 150 GeV. Furthermore the transverse mass of the two leptons is required to be above 60 GeV to avoid low-mass resonances and Drell-Yan background. This will also reduce the amount of events with leptonic τ-decays as the mass peak for e.g. $Z \to \tau\tau$ lies below 60 GeV resulting in a more pure sample of prompt leptons.

The following section is intended as a study of the use of Principal Component Analysis on real data to create independent observables and to show that these will give sensible fit results. It is not intended as a test of the Standard Model in the control region. Some Standard Model process contributions should preferably be calculated using data-driven methods and the final analysis presented in this thesis will use data-driven methods to properly model fakes. For this initial study only Monte Carlo samples have been used to model the expected number of events. A range of corrections for known reconstruction- and detector problems will be applied in the signal region but these are not included here.

5.2.1 Selecting $e^\pm e^\pm$ and $\mu^\pm \mu^\pm$

As charge must be conserved in any Standard Model hard process, the creation of a lepton pair with same flavour, e.g., two electrons or two muons and with same electric charge is in principle ruled out. But since the processes have to be measured, some processes can create such a signal, by having e.g. a jet being (mis)reconstructed as a lepton. The Monte Carlo samples in table 5.3 have been used to model the Standard Model contribution to the SFSS control region. The k-factor is the correction to the dataset cross-section known from higher order theoretical calculations where such exist. Reference to k-factors for the given samples are found in [64].

To select these same-flavour, same-sign (SFSS) events the following selection has been performed on the Monte Carlo samples and for real data.

Event Selection

The event-selection is developed in the multilepton group [61] and identical to the lepton selection in the signal region while the event requirements are for exactly two SFSS leptons. The data is obtained from two data streams. The 'Muons' stream and the 'Egamma' streams. These include all events that have triggered at least one muon and/or one electron. All events

---

[3]There will be a small overlap between the regions as e.g. $t\bar{t}$ and Z/W + jets can produce same-sign dilepton and three-lepton events.

Table 5.3: Monte Carlo datasets for 8TeV 2012 SFSS data-modelling

Process type	dataset id number	k-factor
Single boson processes		
$Z \to ll +$ jets	117650-117674	1.18
$W \to l +$ jets	147774-147776	1.019
Diboson processes		
WZ	126893	1.05
ZZ	126894	1
WW	126892	1.05
top related processes		
$t\bar{t}$	110001	1.2177
top + bosons		
$t\bar{t} + W(+jet)$	119353-119354	1.17
$t\bar{t} + Z(+jet)$	119355-119356	1.35

selected must pass a trigger match, e.g., one electron or muon in the event must match a triggered object. The triggers used for both data and Monte Carlo are:

- **Muon stream**

  – pass EF_e24vhi_medium1 || EF_e60_medium1 || EF_mu24i_tight || EF_mu36_tight

- **Egamma stream**

  – pass EF_e24vhi_medium1 || EF_e60_medium1

  – fail EF_mu24i_tight **and** EF_mu36_tight,

where e/muXX refers to the minimum momentum threshold in GeV, i is an isolation requirement and medium or tight are reconstruction algorithms deciding the degree of tightness required of the object to be reconstructed.

The reasoning behind the choice of two triggers is as follows. The lowest unprescaled p_T threshold triggers for all 2012 data has been selected as the primary trigger. These triggers have requirements on the isolation of the leptons that leads to inefficiencies for high p_T leptons. To compliment the low threshold triggers, the lowest unprescaled triggers without isolation requirements are also used.

- the event must pass the ATLAS standard jet cleaning requirements [65].

- Real data must pass the ATLAS GoodRunsList, guaranteeing that all relevant detectors of the ATLAS experiment were operational and data has been correctly stored.

- The sum of transverse momenta for all reconstructed jets in the event must not exceed 500 GeV.

Finally, exactly two leptons with same flavour and same charge must be found satisfying:

Electron pair selection

The electrons are required to both fulfil the following requirements:

- **ID class:** Tight++

  Forward and soft electrons are vetoed (author 1 and 3).

- **Leading/Triggered lepton:** $E_T \geq 26 GeV$

- **Subleading lepton:** $E_T \geq 15 GeV$

- **Acceptance:** $(\mid \eta \mid < 1.37) \mid\mid (1.52 \leq \mid \eta \mid < 2)$

  Requiring $\eta \leq 2$ ensures TRT coverage ensuring better identification and pion and conversion rejection.

- **Calorimeter Isolation:** The E_T sum based on calorimeter readout in a cone of $\Delta R \leq 0.3$ around the electron divided by electron transverse energy - $\frac{TopoEtCone30}{E_T} < 0.10$

- **Track Isolation:** The p_T sum of tracks in a cone of $\Delta R \leq 0.3$ around the electron divided by electron transverse energy - $\frac{ptcone30}{E_T} < 0.10$

- **Vertex:** $\mid Z_0 sin(\theta) \mid < 0.5 mm$

- **Impact parameter significance:** $\mid \frac{d_0}{\sigma(d_0)} \mid < 3$

Muon pair selection

The muons are required to both fulfil the following requirements:

- **ID class:** Identification Combined tight with:

  1. a B-layer hit (if expected)
  2. ≥ 1 pixel hits (including dead sensors)
  3. ≥ 5 SCT hits (including dead sensors)
  4. < 3 holes in the SCT and Pixels combined
  5. if $0.1 < \mid \eta \mid < 1.9$, then TRT hits + outliers > 5 require $\frac{\text{outliers}}{\text{all hits}} < 0.9$
  6. if $\mid \eta \mid \leq 0.1$ or $\mid \eta \mid \geq 1.9$ and TRT hits + outliers > 5 require $\frac{\text{outliers}}{\text{all hits}} < 0.9$

- **Leading/Triggered lepton:** $E_T \geq 26 GeV$

- **Subleading lepton:** $E_T \geq 15 GeV$

- **Trigger acceptance:** $(\mid \eta \mid < 2.4)$

- **Acceptance:** $(\mid \eta \mid < 2.5)$

- **Author:** The muons are required to be combined muons with track segments both in the inner tracker and muon detectors.

- **Calorimeter Isolation:** The E_T sum based on calorimeter readout in a cone of $\Delta R \leq 0.3$ around the muon divided by muon track transverse momentum - $\frac{EtCone30}{p_T} < 0.10$

- **Track Isolation:** The p_T sum of tracks in a cone of $\Delta R \leq 0.3$ around the muon divided by muon transverse energy - $\frac{ptcone30}{E_T} < 0.10$

- **Vertex:** $\mid Z_0 sin(\theta) \mid < 0.5mm$

- **Impact parameter significance:** $\mid \frac{d_0}{\sigma(d_0)} \mid < 3$

Exactly two same-charge electrons or muons passing the object selection are required to be present in the event and to have transverse mass between 60 and 150 GeV. These are both required not to overlap with other reconstructed objects. The forward region of the ATLAS detector has more non-detector material and a weaker solenoid magnetic field. This means that more conversions are present and that both charge and momentum reconstruction is poorer especially for electrons. This is especially true for the region outside the TRT η range, so to reduce the impact of these objects in the event reconstruction an additional veto on electrons with $\eta \geq 2.0$ is imposed. The result is 4897 data events passing the selection, with a total of 4995 ± 82 (stat) expected of which 4945 events are V + jets. The modelled background does not contain dedicated $b\bar{b}$ and $c\bar{c}$ samples which is likely to cause a slight deviation. Even more importantly no data-driven background estimate of light-flavour jet contribution, conversions and charge mis-reconstruction has been performed.

5.2.2 Distribution of processes

All other relevant processes have been included and the distributions of data and Monte Carlo are shown in figure 5.11 for a selection of observables. The general agreement is good although smaller deviations are seen. These are likely to originate from incomplete modelling. Performing the two standard tests Kolmogorov-Smirnov and a χ^2, on data compared to total expectation, shows that the agreement between model expectations and data is not good everywhere. The normalisation difference before fitting is likely to affect the χ^2 probability as is the non-negligible number of bins with less than 10 events. The shape is seen to agree reasonable well in most observables with reasonable KS values. Besides this, it is evident that V + jets completely dominates this region. For W + jets this is expected as W can decay leptonically to a lepton and a neutrino and the jet can fake a lepton with same charge. Z + jets contributes if the Z decays to leptons where one is not reconstructed. Again, the other lepton must come from a jet. In principle the charge of a lepton could be misidentified also, but the contribution from this is small for the required η range. Conversion of photons can also add reconstructed leptons to the final state, but as the photon has no track in the inner detector the rejection is good.

$\mathbf{PCA}_{V+jets}$ to fit expectation to data

As V+jets so clearly dominates the region a Principal Component Analysis based on V+jets has been chosen. The eigenvalues and sum of square residuals is plotted in figure 5.12 for the PCA-transformation based on V + jets. As a reference the eigenvectors are included in appendix B along with the list of input observables. To understand the eigenvalues we need to know the correlation between observables in the pattern space. They are plotted for V + jets in figure 5.13. Correlation up to above 60% exists and a Principal Component Analysis should be able to combine much of this in the highest ranked *principal component*. As expected, the resulting eigenvalues of the *principal components* for V+jets, fig. 5.12a show that p_1 combines
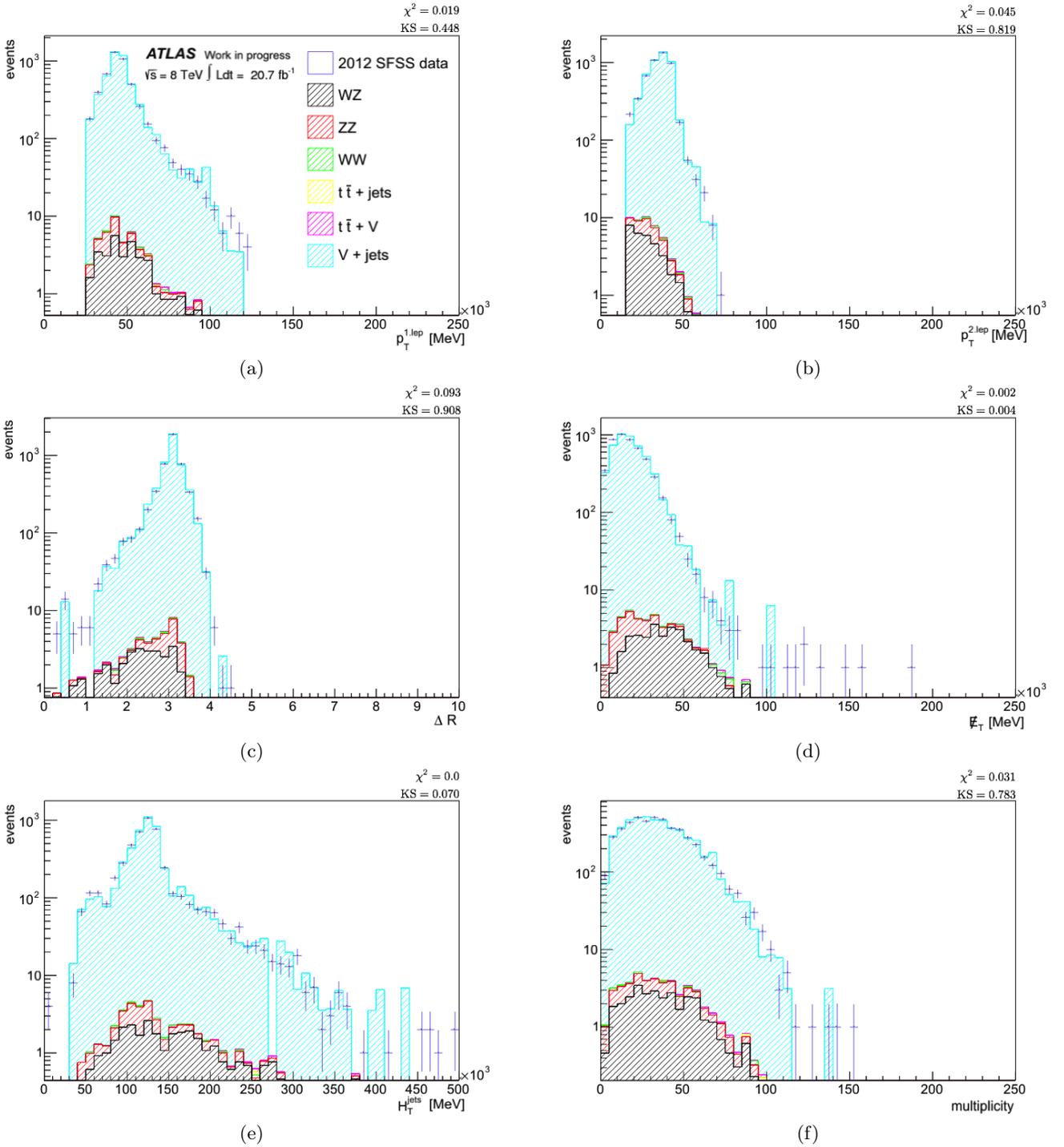
Figure 5.11: The distribution of processes contributing to the SFSS dilepton final state. It is clear that the W/Z + jets completely dominates this region, contributing to the vast majority of the observed events.

roughly 45% of the information from the correlated observables into one parameter. The second and third are at roughly 17% and 13%. Figure 5.12b shows the distance measure for the *principal components*. It shows that the space is not fully spanned until the inclusion of p_6 after which a very small distance remains. The measure is here again for the entire sample and normalised to number of observed events.
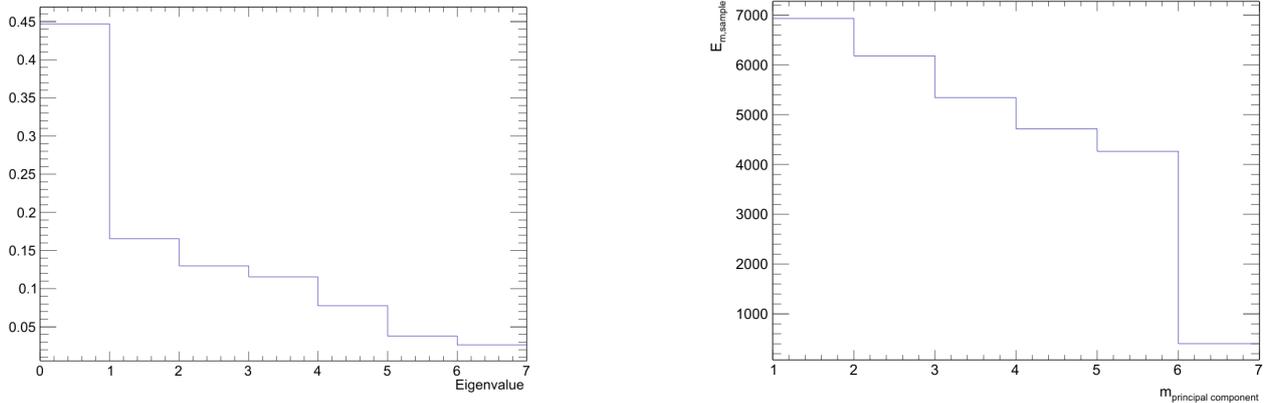
(a) The eigenvalues of the eigenvectors of the covariance matrix of V + jets.

(b) The distance measure, E_m, for *principal components* for the entire V + jets sample.

Figure 5.12: The eigenvalues and sum of squared residuals resulting from the transformation of the four observables in fig. 5.11 using Principal Component Analysis for V + jets and 2012 SFSS data.

The final comparison of distributions is shown in fig. 5.14 and 5.15. Here all *principal component* distributions are plotted for data and Standard Model expectation. The errors here are purely statistical. Before the fit of the Standard Model expectation to data in the SFSS region is presented we return to the question of independence of the *principal components* .
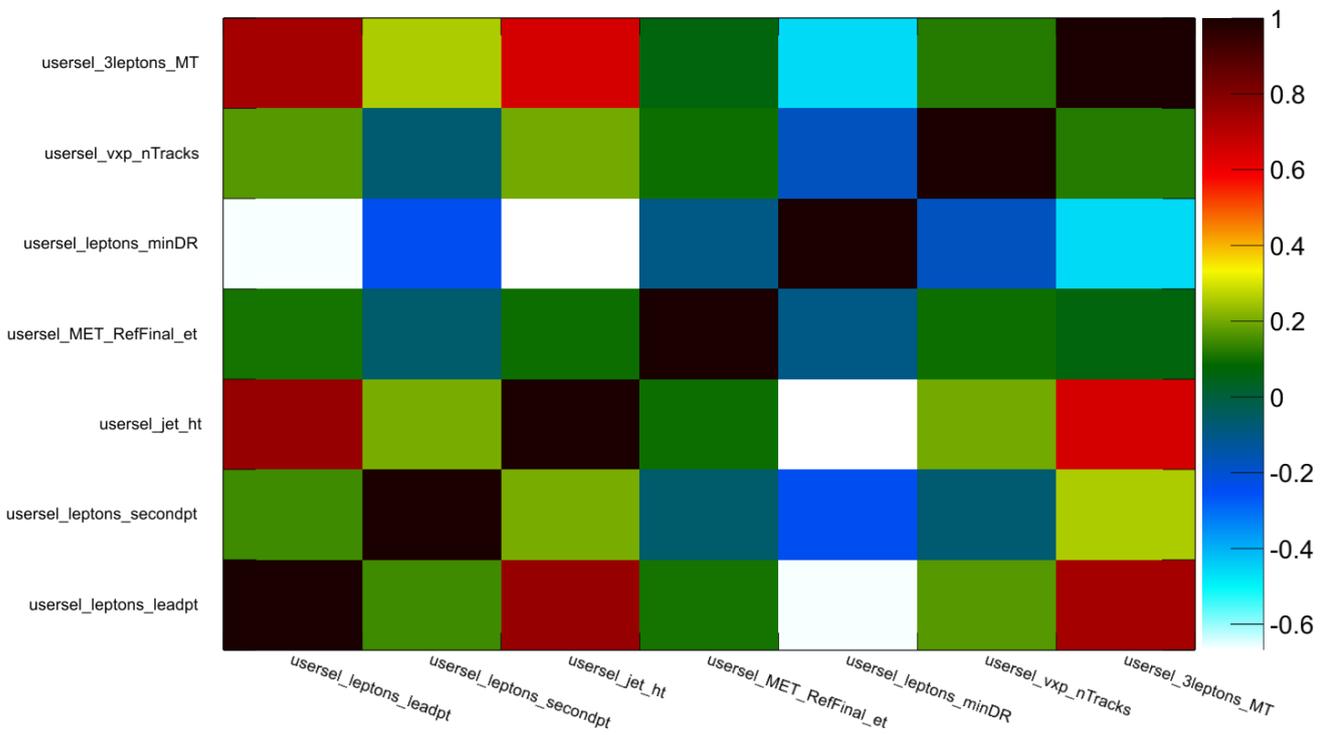
Figure 5.13: The linear correlation between input observables in V + jets for the SFSS selection.
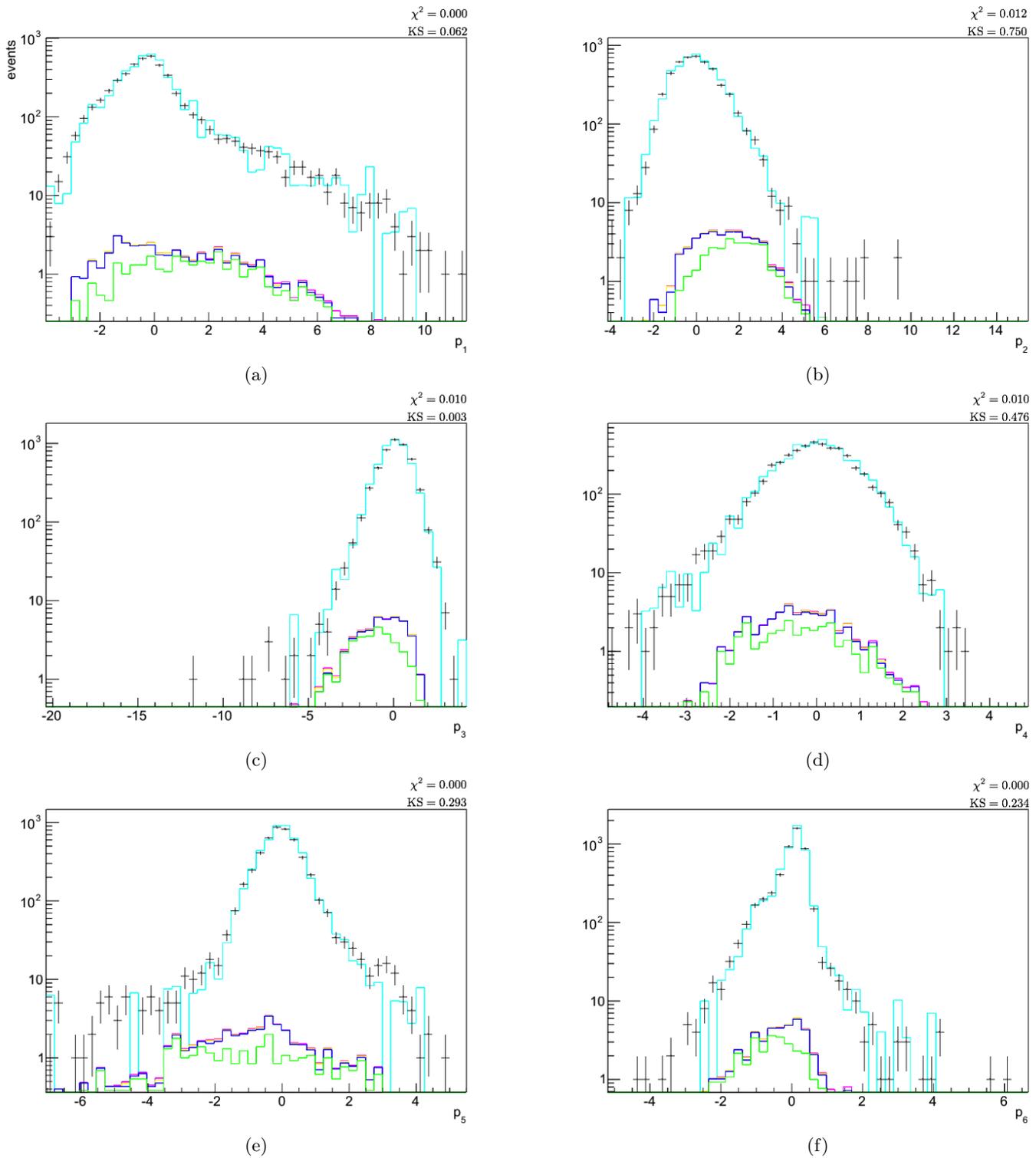
Figure 5.14: The *principal components* of the SFSS expectations and data. The first and highest ranked *principal component* representing the linear component with the largest variance is P1 and lowest rank is P6.
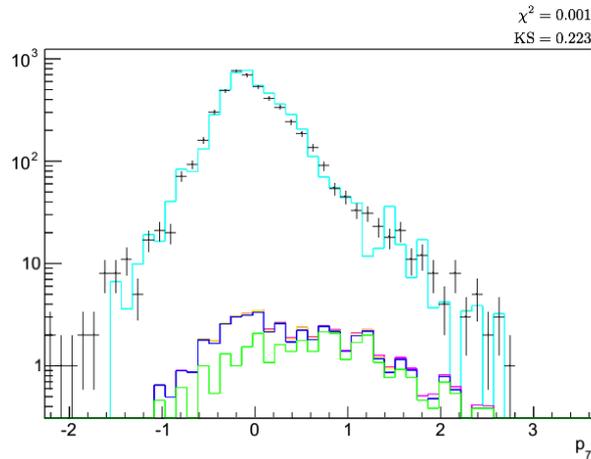
Figure 5.15: The last *principal component* of the SFSS sample.

5.3   Mutual information and fitting real data with PCA

If templates in several *principal components* are to be fitted simultaneously without handling correlations, the *principal components* must be independent to provide sensible errors. Two questions arise. Principal Component Analysis ensures linear independence but does not guarantee non-linear independence. Secondly the transformation matrix only guarantees independence for the process it is generated from, so the resulting *principal components* for other Standard Model processes might not be independent. To test whether *principal components* have non-linear correlations the Mutual Information (MI) method as defined in eq. 4.36 is used. The implementation can be done in several ways. The most straightforward is through binned distributions but as described in [66] and [67] the usage of kernel density estimation to model distributions removes the dependence on origin and binning, giving a better estimate. The kernel estimate of ROOT's RooKeysPDF has been used to model the distributions for the mutual information measure here. RooKeysPDF uses kernel estimation to get a functional description of a given distribution (by convoluting Gaussians). The widths of the Gaussians are through an 'adaptive mode' to better model data and minimise bias. For multi-dimensional kernel estimates the associated *RooNDKeys* class ensures that the kernels are constructed to reflect the correlation coefficient between the observables in the input dataset.

MI for varying degree of dependences

To study what the benchmark value should be for linearly uncorrelated variables to be considered independent a series of correlated functions have been studied. Common to them are that they have non-linear correlation between variables. A completely independent sample was generated using a random number generator [68] to get random number between 0 and 1 in variables x and y. The seed for the generator made the two series completely independent.

For studying non-linear correlations three functions with non-linear correlations were used:

$$\begin{aligned} y &= cos(x), & (5.9)\\ y &= sin(x)^2 \\ y &= \sqrt{(x - 0.11) \cdot (x - 0.11) - 0.1} + 0.1. \end{aligned}$$

To test the effect of partial correlations on the mutual information measure the fraction of correlated and independent data points in a final dataset was varied from 0 to 100%. The result is plotted in figure 5.16. It shows the plot of the MI value as a function of degree of correlation of the two variables x and y. The plotted points have approximately 50% errors.

For illustrative reasons a second degree polynomial has been fitted to each of the three distributions. The conclusions from the figure is that the value of mutual information is non-zero even for independent series, it rises slowly for low percentage of correlated datapoint and faster for higher percentage of dependent events. The rise seems to follow a second degree polynomial but this feature is not probed further here. MI is not expected to give exactly 0 as statistics influence the value [56] [66], so a non-zero start value is acceptable. This should be considered when considering whether a given set of *principal components* are independent. From fig. 5.16 it seems that for the given functions, an MI measure of 0.1 will correspond to samples with less than 30% correlated data points, despite the type of correlation.
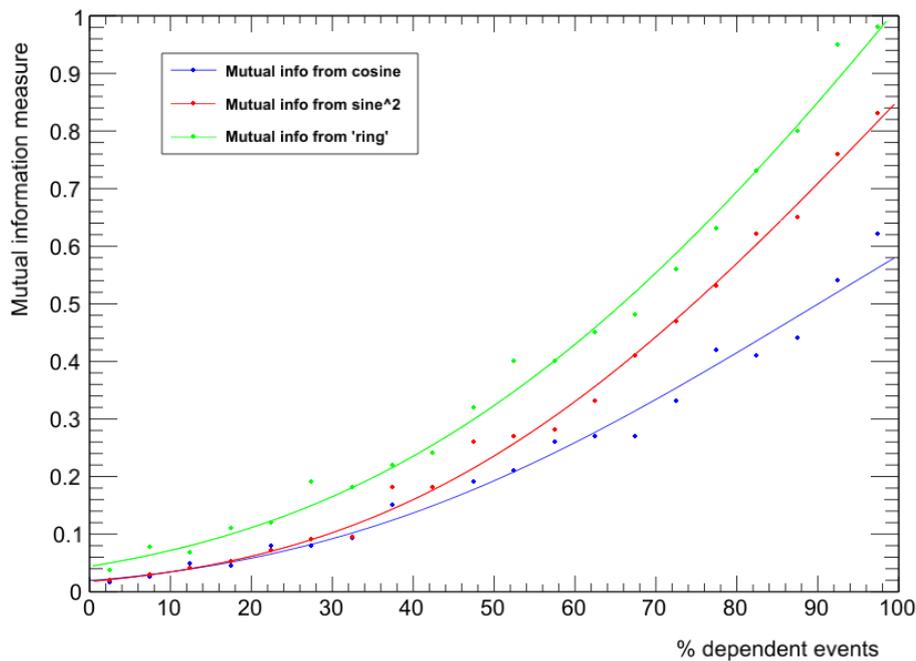
Figure 5.16: The measure of mutual information as a function of percentage of correlated generated data for three different non-linear correlations. Even for two completely independent generated datasets the MI measure still has a small value. High degrees of non-linear correlation between observables quickly gives rise to higher values of MI.
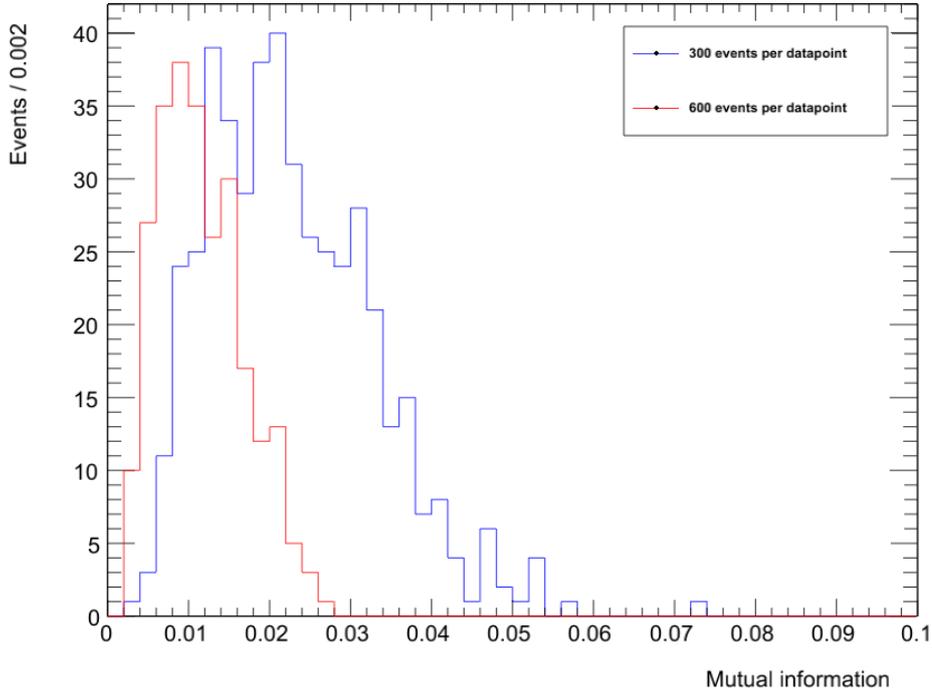
Figure 5.17: Mutual information between two independent observables for samples with 300 and 600 entries per datapoint in the mutual information calculation.

5.3.1 MI dependence on number of entries

To probe the spread as a function of number of data points in the calculation of each MI measure, a series of events with 300 entries per datapoint and 600 entries per datapoint have been generated. The two variables are have been generated to be independent per construction. The MI has been calculated for each datapoint and plotted in fig. 5.17. The number of entries clearly influence the spread significantly. Both distributions have tails of events with larger than average MI. It falls off fast but the sample with 600 events per datapoint reaches values up to 0.03. The sample with 300 entries per datapoint have values as high as 0.07 even though the two observable are completely independent by construction.

The MI for *principal components* from V + jets in SFSS region

The mutual information was calculated for all combinations of *principal components* in 600 events from the V + jets sample. No combinations had an MI value above 0.05 and the highest value found was $MI_{V+jets} = 0.027$. The threshold was chosen from the above result on MI as a function of events and the MI measure is not sensitive to anything below 0.05 as this corresponds to the upper fluctuation expected from 600 events in independent variables. The resulting number is so low that it is more than fair to assume non-linear correlations does not exist and the *principal components* in V + jets are indeed completely independent.

The other question regarding independence of the *principal components* was associated

to processes other than V + jets used to generate the transformation matrix. The MI was also calculated for data and for the other Monte Carlo processes as well. No combinations of *principal components* were above 0.05 in data. WZ is the largest background but still only has 33 events expected so the MI measure is likely to be above the threshold for any combination of *principal components*. It was found to be especially large for $MI_{2,3}^{WZ} = 0.12$, $MI_{2,4}^{WZ} = 0.08$, $MI_{2,6}^{WZ} = 0.09$ and $MI_{5,7}^{WZ} = 0.11$ with several other above 0.06. The transformation matrix only guarantees linear independence in data-points that resemble[4] the sample used to generate the transformation matrix. The two SFSS leptons from WZ likely deviate so much from the SFSS leptons in V + jets that the transformation matrix does not generate independent variables.

The dependence on the number of entries actually means that MI does not give meaningful values for processes with very few events in the signal region. This will be discussed later. The conservative approach for the final result will be to look at the two-dimensional distributions of all *principal component* pairs that are above threshold of 0.05 per 600 events in the sample under study.

Correlations between *principal components* in data and WZ

As the MI measure clearly indicated correlations, between several of the *principal components* in WZ, it is interesting to take another look at the linear correlations in data and WZ. The linear correlation is zero between all *principal components* in V + jets per construction but fig. 5.18a shows the linear correlation of the same *principal components* in the 2012 SFSS data sample. The correlation is below 20% for all *principal components* in data and several are close to 0. In WZ several combinations exist with low correlation, however, the correlation coefficient between p_2 and p_3 is below -60% and not in any way independent. The *principal components* p_2 and p_3 where also found to have a high MI value in WZ. The electroweak Monte Carlo processes represented by WZ in fig. 5.18b all have high correlation between p_2 and p_3 but due to their small contribution to total number of expected events the correlation is not found in data.

The dependence of the observables means that they cannot be combined in the fit without taking into account correlations. This is strictly true for electroweak Monte Carlo but the V + jets component of data could still be fitted if the other Standard Model contributions are fixed as the *principal components* where seen to be independent in data too. The next subsection will fit V + jets using Principal Component Analysis and present the results for all *principal components* whereas the final results for this thesis will discuss dependence of processes in the signal region further. All other processes will be fixed to their expected value.

5.3.2 PCA fit results in SFSS signal validation region

Like for the purely Monte Carlo based test of the fit method in section 5.1 a series of fits have been performed in the SFSS validation region. The result of the fits are shown in table 5.4. All quoted errors are the MINOS errors returned by TMinuit in the fitting routine. The first results are for the fit of each individual *principal component* alone and the second part is of the combination of *principal components*. To compare the results, a fit has been performed in $\not{E}_T$, chosen as it was found to have the largest discrimination power between V + jets and

---

[4]In principle the independence is only strictly guaranteed for the sample used to generate the transformation matrix
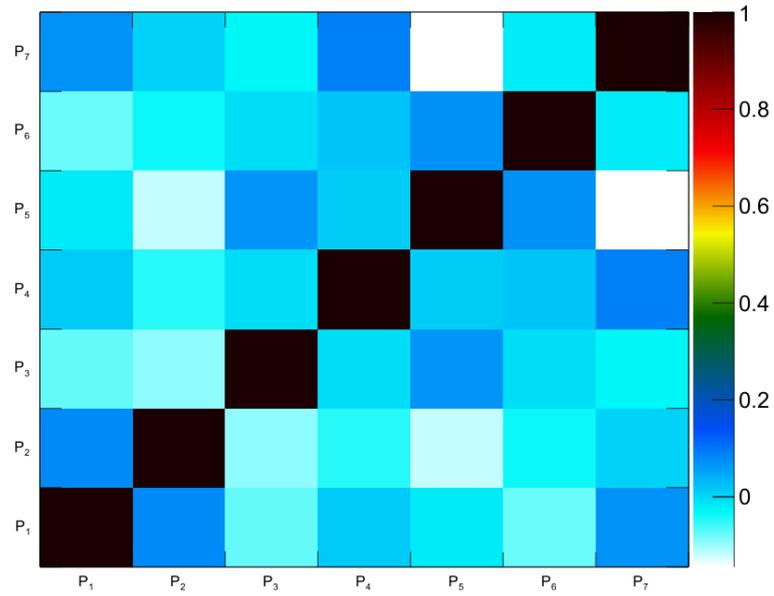
(a) Linear correlation in data.
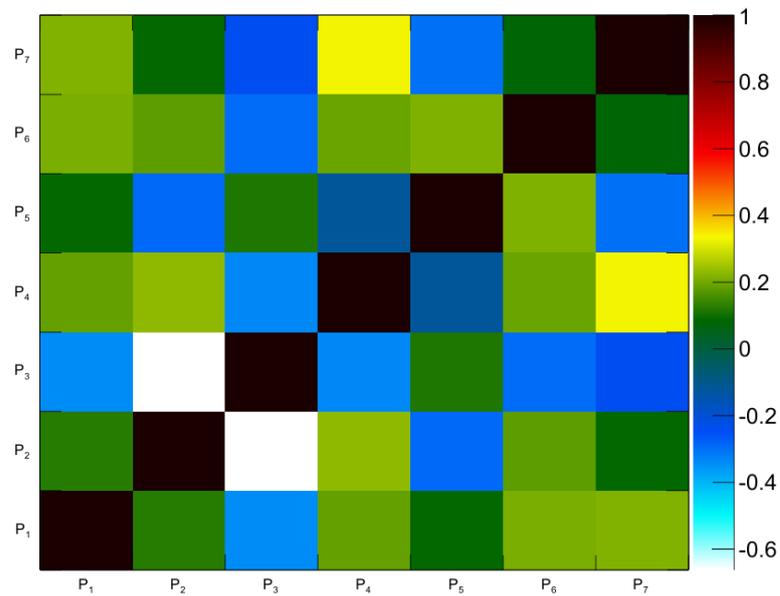
(b) Linear correlation in WZ.

Figure 5.18: The linear correlation between *principal components* in data and in WZ. The transformation matrix was generated from V + jets Monte Carlo.

data using a Fisher discriminant in TMVA [69]. The result was $\alpha_{V+jets} = 0.979 \pm 0.014(14)$. The fit-uncertainties have been presented with extra digits in the table to hint at the effect of the shape on the fit, but as was the case with WZ and ZZ Monte Carlo the uncertainty associated with the extended part dominates.

Extended and shape-part fit - individual *principal components* fitted

principal component	1	2	3	4
α_{V+jets}	0.97	0.97	0.97	0.97
σ_{fit}	$\pm$ 0.01(4)	$\pm$ 0.01(4)	$\pm$ 0.01(4)	$\pm$ 0.01(4)
principal component	5	6	7	
α_{V+jets}	0.96	0.96	0.97	
σ_{fit}	$\pm$ 0.01(4)	$\pm$ 0.01(4)	$\pm$ 0.01(4)	

Extended and shape-part fit - combining *principal components* in fit

# *principal components*	1	2	3	4
α_{V+jets}	0.967	0.966	0.963	0.961
σ_{fit}	$\pm$ 0.013(99)	$\pm$ 0.013(98)	$\pm$ 0.013(96)	$\pm$ 0.013(96)
principal component	5	6	7	
α_{V+jets}	0.948	0.940	0.934	
σ_{fit}	$\pm$ 0.013(85)	$\pm$ 0.013(79)	$\pm$ 0.013(75)	

Table 5.4: Fit values in the SFSS region.

The fit using only overall normalisation resulted in $\alpha_{V+jets} = 0.97 \pm 0.014$. Fitting in both the highest ranked *principal component* and the original observable, results in fit uncertainty and value very close to the extended alone. This indicates that the fit is driven by the normalisation and very little information is gained from the shape. The fit of individual *principal components* supports this claim as very little variation is seen. The fit in combined *principal components* shows a slight decrease in the fit-uncertainties hinting the same behaviour as seen in the Monte Carlo study. The effect is not significant for the given region.

5.3.3 Conclusions for the SFSS region study

One of the most noteworthy points of this chapter has been to clarify the challenge of dependent *principal components*. The transformation matrix only guarantees linear independence for the V + jets sample and even in data, where this process constituted the vast majority of events, the *principal components* where not all linearly independent. The choice of a not so well modelled region (e.g., without dedicated b-samples and data-driven estimates) is a worst case scenario but as several processes are significant in the final signal region the study of dependence will be even more essential. The MI measure was shown to find the *principal components* with linear correlation in WZ but also showed that no non-linear correlations existed between *principal components* in V + jets and data. This validated that the *principal components* in V + jets are indeed completely independent and that the independence generalises to data in this very pure V + jets region.

The fit of V+jets in the region showed that the method worked and resulted in fits comparable to the original observables. The fit-uncertainty decreased with the addition of *principal components* although the effect is in the insignificant digits and does not fully validate the behaviour. It is important to stress that the fit does not exclude the Standard Model. The

modelling is known to deviate and the region simply cannot be described probably without data-driven estimates. The intent of the SFSS study was not to test the Standard Model but to illuminate the challenges and features of the method concerning dependence and sensibility to deviations.

6

Standard Model expectation in three-lepton final states

6.1 Standard Model processes

In this thesis we have chosen the final state consisting of three leptons selected according to the characteristics of a prompt electroweak lepton to look for new physics. By prompt electroweak we refer to muons or electrons produced in the prompt decay of a W or a Z boson or of similar characteristics, that is real leptons with little hadronic activity in their vicinity. In the simplest form these leptons will be high-p_T, well isolated and well characterised by the sub-detectors of ATLAS i.e. all electrons for instance passing the tight++ selection criteria. This final-state is chosen, as it is one of the cleanest, so any deviation caused by new physics are likely to result in a clear signal.

In the Standard Model the only hard interaction contributing to a final state with exactly three electroweak, prompt leptons is the production of a W and a Z boson and their decay to three leptons. In reality three-lepton events will be the events we reconstruct as having three *electroweak and prompt*-like leptons and to do this properly a series of cuts have to be imposed including minimum p_T of the leptons, isolation requirements and matching of hits or tracks to energy deposits in the calorimeters. The result is that hard interactions with more than three leptons in the final-state will contribute if one of the leptons fail the selection.

Other hard interactions with three lepton final states will contribute even though the leptons might not be prompt electroweak in origin. The decay of a heavy quarks through a W to an electron or muon is one of the major contributors to this type of real non-promt leptons. Specifically hard interactions involving top-quarks and bosons result in non-negligible contribution to three-lepton final-states. The conversion of a photon into two electrons is another source of real leptons and will contribute noticeably when produced along with a Z-boson. Furthermore non-leptonic objects can be mis-reconstructed as leptons and several processes will contribute as so-called fakes.

The following chapter will describe the selection developed in the multilepton group at CERN for three-lepton final-states and present the expected number of events from Standard Model processes.

6.2   Signal event selection

6.2.1   Event based selection

To select events with three prompt, electroweak leptons both event-level and object based selections have been imposed. To ensure the event is a hard process and to reduce fake and pile-up related effects, a cut on the number of tracks from the primary vertex with highest summed transverse momentum of tracks has been imposed:

- 3 or more tracks must be associated to the vertex with largest $\sum p_{T,track}^2$

As badly reconstructed jets can mimic a range of objects and have a significant impact on the reconstruction and understanding of an event, all events have undergone the ATLAS jet-cleaning requirements.

- Event must pass ATLAS jet-cleaning requirements [65] or the entire event is discarded.

As all data is recorded using triggers to store only interesting physics events, a requirement that the relevant triggers have fired is imposed. The choice of triggers are the unprescaled triggers with lowest p_T requirements for electrons and muons. The triggered events are recorded in two streams that will have overlapping events. To avoid overlap all events with a triggered muon- or electron-object was used from the muon stream while events were only used from the electron (Egamma) stream if they contained a triggered electron- and no triggered muon-objects. The exact definition used is:

- **Muon stream**

  – pass EF_e24vhi_medium1 || EF_e60_medium1 || EF_mu24i_tight || EF_mu36_tight

- **Egamma stream**

  – pass EF_e24vhi_medium1 || EF_e60_medium1
  – fail EF_mu24i_tight **and** EF_mu36_tight

As already mentioned in section 5.2.1 the reason for having an OR between two triggers, is that the isolation requirement (i for isolated) on the lowest unprescaled trigger leads to inefficiencies for leptons with higher p_T [61]. To ensure efficiencies are correctly modelled at least one of the leptons in the event is required to match the triggered object. The efficiency of the lepton triggers depends on the p_T the lepton. The trigger matched lepton is required to have $p_T \geq 26$ GeV to ensure they are selected in the efficiency plateau of the trigger, i.e. where trigger efficiencies are close to 100%. Muons matched to the trigger object are further required to have $|\eta| \leq 2.4$ to avoid uninstrumented regions.

Finally the ATLAS detector is required to be fully functional and have all relevant detector subsystems active for the given event, so the information can be trusted. This is taken care of by the ATLAS GoodRunsLists that contain all relevant status-flags.

- Real data must pass the ATLAS standard *StandardGRL_ All_ Good GoodRunsList*, guaranteeing that all subdetectors of the ATLAS experiment were operat ional and data has been correctly recorded.

6.2.2 Electron Selection

As the electron selection should favour prompt, electroweakly produced electrons the object requirements can be very strict regarding object reconstruction, centrality, origin and isolation. The requirements for an object to be accepted as an electron are:

- **ID class:** Tight++

  Including a requirement of the electrons to originate from the barrel or end-cap and to be above a p_T threshold (*author* 1 or 3)

- **Leading/Trigger:** $E_T \geq 26 GeV$

- **Subleading:** $E_T \geq 15 GeV$

- **Acceptance:** $(\mid \eta \mid < 1.37) \mid\mid (1.52 \leq \mid \eta \mid < 2.47)$

- **Calorimeter Isolation:** $\frac{TopoEtCone30}{E_T} < 0.10$

- **Track Isolation:** $\frac{ptcone30}{E_T} < 0.10$

- **Vertex:** $\mid Z_0 sin(\theta) \mid < 0.5 mm$

- **Impact parameter significance:** $\mid \frac{d_0}{\sigma(d_0)} \mid < 3$

The identification of the electron is required to agree very well with the expected signature of a real electron following the ATLAS electron tight++ identification. This alone results in a significant rejection of especially light hadrons faking electrons. As the search in this thesis will be sensitive predominantly to high p_T physics, the electron p_T is required to be above 15 GeV for all electrons and at least 26 GeV for the triggered electron. Note that the trigger matched electron does not have to be the leading but the trigger matched electron must also have p_T of at least 26 GeV. The difference is very subtle - the trigger matching is performed before energy and momentum corrections [1] so the matched lepton could in principle be below the cut after corrections. The leading lepton cut ensures this does not happen.

The author requirement rejects forward and *soft* (low-p_T) electrons and the vertex and impact parameter requirements rejects photon conversions and helps ensure that the reconstructed object originates from the hard process.

The η requirements on the electron is to avoid detector regions with problematic reconstruction and to be within coverage of the relevant detectors. As this search aims to be sensitive to prompt isolated electrons, like electrons from W and Z leptonic decays, strict requirements are imposed on the electron isolation. The isolation criteria significantly rejects electrons originating from in-flight decays of hadrons and heavy quark decays to some extent. As electrons will naturally radiate *bremsstrahlung* with increased momentum, the requirements are taken as the ratio of transverse energy or momentum in a cone ($\Delta R = 0.3$) around the object to the total transverse energy deposited by the object.

---

[1] It has to be as the trigger algorithm uses the uncorrected values to identify the trigger object.

6.2.3 Muon selection

Much of the reasoning behind the muon selection is equivalent to the reasoning for the electron selection cuts. Prompt high p_T muons are expected to deposit little energy in the inner parts of the detector and must reach the muon chambers. They are expected to be well isolated and are as such one of the objects that are not easily faked by other objects if selected correctly. To ensure this the following selection has been performed:

- **ID class:** Combined tight (*author* 6)

  1. a B-layer hit (if expected)
  2. ≥ 1 pixel hits (including dead sensors)
  3. ≥ 5 SCT hits (including dead sensors)
  4. < 3 holes in the SCT and Pixels combined
  5. if $0.1 < \mid \eta \mid < 1.9$, then TRT hits + outliers > 5 and $\frac{\text{outliers}}{\text{all hits}} < 0.9$
  6. if $\mid \eta \mid \leq 0.1$ or $\mid \eta \mid \geq 1.9$ and TRT hits + outliers > 5, require $\frac{\text{outliers}}{\text{all hits}} < 0.9$

- **Leading/Trigger:** $E_T \geq 26 GeV$

- **Subleading:** $E_T \geq 15 GeV$

- **Trigger acceptance:** ($\mid \eta \mid < 2.4$)

- **Acceptance:** ($\mid \eta \mid < 2.5$)

- **Calorimeter Isolation:**

  - $p_T^\mu < 100 GeV$: $\frac{EtCone30}{p_T} < 0.10$
  - $p_T^\mu \geq 100 GeV$: $EtCone30 < 10 GeV + 0.01 \cdot p_T^\mu$

- **Track Isolation:**

  - $p_T^\mu < 100 GeV$: $\frac{PtCone30}{p_T} < 0.10$
  - $p_T^\mu \geq 100 GeV$: $PtCone30 < 10 GeV + 0.01 \cdot p_T^\mu$

- **Vertex:** $\mid Z_0 sin(\theta) \mid < 0.5 mm$

- **Impact parameter significance:** $\mid \frac{d_0}{\sigma(d_0)} \mid < 3$

Apart from the cuts already described for leptons a set of inner detector requirements on the muon-track has been imposed to help with identification. The track is required to have adequate hits matching expectations from simulated muons in both pixel detectors and in the TRT. Furthermore a hard limit is set on the number of holes allowed on the track and on outliers for the TRT. Outliers are hits used in the track reconstruction where either the central wire or edge of the TRT straw was hit. The number of outliers and holes on a track significantly impacts the reconstructed track, so imposing these requirements ensures tracks are well reconstructed and increases probability of reconstructing momentum and charge correctly.

The muon isolation efficiency was found to be dependent on p_T and the cuts have been optimised by redefining the cut for muons with p_T above 100 GeV.

Jet reconstruction

Jets have been selected using the anti-k$_T$ algorithm for a cone ΔR=0.4 based on topological clusters in the calorimeter with corrections from both the electromagnetic calorimeter and jet energy-scale. The requirement for a jet to pass object-selection is p_T larger than 25 GeV and $\eta \leq 4.5$. Furthermore a jet-cleaning has been performed where jets ending up in areas in the Liquid Argon calorimeter known to have holes are discarded. Finally, the jet is required to originate from the primary vertex with a given probability ($|\mathrm{JVF}| > 0.75$).

6.2.4 Overlap removal

A series of objects will be reconstructed on top of each other. Electrons will for instance very often be reconstructed as jets. These objects do not constitute true individual objects but are artefacts of the reconstruction procedure and overlapping objects need to be handled by storing only the objects believed to be the true object. A set of standard overlap removal cuts are imposed to ensure this:

If there are two electrons within $\Delta R < 0.1$, remove the electron with lower p_T

If there is an electron and jet within $\Delta R < 0.2$, remove the jet

If there is still a jet within $\Delta R < 0.4$ of the electron, remove the electron

If there is a muon within $\Delta R < 0.1$ of an electron, remove the electron

If there is a muon within $\Delta R < 0.4$ of a jet, remove the muon

Finally exactly three reconstructed leptons passing all cuts are required for an event to pass the signal region selection. If any of these are within $\Delta R = 0.2$ of each other the event is discarded.

6.2.5 Corrections applied to Monte Carlo samples

Several corrections to Monte Carlo distributions are recommended by ATLAS. These have to do with discrepancies between data and Monte Carlo that originate from simulation flaws or incomplete modelling/understanding of the detectors.

A scale shift has been observed in $Z \to l^+l^-$ events with regards to simulated data [70,71]. To correct for this the magnitude of transverse energy of electrons and transverse momentum of muons has been smeared in Monte Carlo samples according to the recommendations from ATLAS. The corrections are typically below 1%. A calorimeter response disagreement is corrected by scaling electron energies with corrections typically below 0.5% Muon and electron identification and reconstruction is also known to need corrections and the ATLAS recommended scale factors have been applied here. Scale factors have also been applied to account for a disagreement between the single lepton trigger efficiencies in data and Monte Carlo and for a mis-modelling of the track impact parameters d_0 and z_0 and track isolation. The lepton efficiency, reconstruction and identification scale factor systematics sum to an averaged 1.19 % for all processes and the track isolation and impact parameter corrections another 1.55 %.

For nearly all events recoded in 2012 several proton pairs collided simultaneously. The collisions are separated in space, so the impact on the hard process is negligible but it affects a range of parameters in the event including number of tracks in the event. To model this correctly Monte Carlo samples have been generated with pile-up (multiple events on top of the hard process) but to properly resemble the actual conditions a dedicated pile-up tool is used to reweigh Monte Carlo events [72].

6.3   Irreducible processes

The majority of events passing the selection is expected to originate from the Standard Model processes, that naturally decay to three-lepton final-states with WZ being by far the dominant process. To include all possible irreducible processes, the selection was run for all Monte Carlo samples listed in table 6.1. Furthermore contributions from triboson and $t\bar{t}$ + WW and singletop samples where tested and found to be negligible. W/Z + jets are included in the table as the samples have been used in the thesis but for the final results these will be estimated as a reducible (fake) background. The k-factors represent corrections to the process cross-section when calculations to higher-order than those used in the specific generator matrix elements exists. The corrections are from various sources all referenced in [64] - for the central diboson samples the correction is found through the use of MCFM [73] that performs next-to-leading order calculations of cross-sections.

Selection implementation

The selection described in 6.2 was implemented in several steps. A preselection with the jet cleaning and number of tracks from the primary vertex was the first cut followed by the demand that one of the chosen triggers had fired for the given event. After the object selection was run and after all overlap removal and corrections were applied, a demand of three leptons was imposed. Finally the cut on ΔR between the final leptons was applied. The actual efficiency of the object selection cuts varies from process to process. The selection optimisation has not been done by the author but in parallel with the CERN multilepton group. The cutflow histograms for object selectors have been included though, for some of the main processes, in appendix C and table 6.2 lists the cutflow values for the Monte Carlo processes used. The total integrated luminosity of the recorded 2012 data-stream used in this thesis is $20.28 \pm 0.56 fb^{-1}$ and the last column lists the expected number of events scaled to this luminosity.

Table 6.1: Monte Carlo datasets for 8TeV 2012 data-modelling

Process type	dataset id number	$\sigma_{gen} \cdot \epsilon_{filter}$ (pb)	ϵ_{filter}	K Factor	Gen. Int. lumi. (fb^{-1})
Diboson processes - Sherpa					
WZ	126893	9.75	1	1.05	263.7
ZZ	126894	8.74	1	1.00	467.6
WW	126892	5.5	1	1.05	434.9
$Z \rightarrow ee + \gamma$	145161	32.3	1	1.23	269.5
$Z \rightarrow \mu\mu + \gamma$	145162	32.3	1	1.23	269.3
Single boson processes - Alpgen + Pythia P2011C tune					
$Z \rightarrow ee$ no partons	117650	711.8	1	1.23	9.3
$Z \rightarrow ee + 1$ parton	117651	155.2	1	1.23	8.6
$Z \rightarrow ee + 2$ partons	117652	48.75	1	1.23	8.3
$Z \rightarrow ee + 3$ partons	117653	14.23	1	1.23	7.7
$Z \rightarrow ee + 4$ partons	117654	3.76	1	1.23	8.0
$Z \rightarrow ee + 5$ or more partons	117655	1.095	1	1.23	9.1
$Z \rightarrow \mu\mu$ no partons	117660	712.1	1	1.23	9.3
$Z \rightarrow \mu\mu + 1$ parton	117661	154.8	1	1.23	8.6
$Z \rightarrow \mu\mu + 2$ partons	117662	48.91	1	1.23	8.3
$Z \rightarrow \mu\mu + 3$ partons	117663	14.25	1	1.23	7.7
$Z \rightarrow \mu\mu + 4$ partons	117664	3.784	1	1.23	8.0
$Z \rightarrow \mu\mu + 5$ or more partons	117665	1.115	1	1.23	9.1
$Z \rightarrow \tau\tau$ no partons	117670	711.8	1	1.23	9
$Z \rightarrow \tau\tau + 1$ parton	117671	155.1	1	1.23	9
$Z \rightarrow \tau\tau + 2$ partons	117672	48.80	1	1.23	8
$Z \rightarrow \tau\tau + 3$ partons	117673	14.16	1	1.23	8
$Z \rightarrow \tau\tau + 4$ partons	117674	3.774	1	1.23	8
$Z \rightarrow \tau\tau + 5$ or more partons	117675	1.116	1	1.23	9
top related processes - McAtNlo + Jimmy					
$t\bar{t} +$ jets	110001	207.78	1	1.2177	39.5
t (Wt-channel)	108346	0.207	1	1.0825	44692
top + bosons - MadGraph + Pythia					
$t\bar{t} + W$	119353	0.1041	1	1.17	3284
$t\bar{t} + Wj$	119354	0.0534	1	1.17	6406
$t\bar{t} + Z$	119355	0.0677	1	1.35	4377
$t\bar{t} + Zj$	119356	0.0454	1	1.35	6532

Table 6.2: Cut-flow values for the signal region
Monte Carlo data

process	#events	Preselection	Trigger requirement	3-lep selection	Scaled to $20\,fb^{-1}$
WZ	2699893	2686091	599216	16712	1239
%	100.000	99.489	22.194	0.619	
ZZ	3799491	3779408	1014732	10206	461.1
%	100.000	99.471	26.707	0.269	
WW	1390000	1382085	820036	18	0.8481
%	100.000	99.431	58.995	0.001	
$t\bar{t}$ + V	799993	799862	269185	2912	14.03
%	100.000	99.984	33.648	0.364	
t (Wt-chan)	999799	999590	263873	4	0.0025
%	100.000	99.979	26.393	0.0004	
$t\bar{t}$+jets	200000	199954	131062	21	8.115
%	100.000	99.977	65.531	0.011	
$Z \rightarrow ll+\gamma$	18033252	17844170	10676526	452	31.19
%	100.000	98.951	59.205	0.003	
Total number of expected events from Monte Carlo:					1755 ± 42 (stat)

6.4   Reducible processes

As already mentioned the reconstruction of event kinematics means that three-lepton final-state events can originate from a range of hard-processes which do not involve prompt electroweak leptons. The leptons are called fake leptons as a general term but the characteristics of these differ. Prompt electroweak electrons can be faked by either semi-leptonic heavy flavour decays, like the decay of a top quark through a W to a lepton, neutrino and a b-quark, or from misidentified light hadrons and conversions. Pions will for instance resemble an electron quite a lot and even with the proposed isolation and identification criteria some light hadrons will be identified as electrons. Photon conversions will typically create an electron pair in the beam-pipe and either one or both the electrons can be reconstructed as prompt electrons. Muon fakes originate from decays of b- and c-hadrons [74] and to some extent also from Pion and Kaon decays or punch-through, where particles reach the muon detectors and deposit energy like a muon would. The amount of fakes to be expected in the signal region cannot be probed using Monte Carlo samples, as the fake dynamics are not well enough modelled. Instead a data-driven method, where scaling factors are calculated for the different fake types in a region orthogonal to the final signal region, is used.

6.4.1   Data-driven methods

Defining a numerator object, N, with selection identical to the signal region object of the type (e.g. electron/muon) and a denominator object that pass most selection criteria for the object but fail a selected few, D, the fake-factor is simply the ratio:

$$f = \frac{N}{D}. \tag{6.1}$$

The principle of the fake-factor method is to estimate this scaling factor in a control region orthogonal to the signal region and apply it to events in the signal region, where one or more of the three selected leptons fail nominal cuts but pass the denominator cuts, to get the total number of nominator objects expected in the signal region. In the perfect denominator selection of the fake control region no nominator objects are present. For this reason the contribution from simulated processes with real leptons have been subtracted from the fake-factor control region. The nominator and denominator selection criteria are described in the following sections and the combinatorics leading to the final estimate in 6.4.2.

Electron denominator definition

The fake electron selection must allow semi-leptonic heavy flavour decays and misidentified light hadrons to pass. The impact parameter significance is an excellent observable for enhancing semi-leptonic heavy flavour decays as the heavy (b-)quarks will travel a measurable distance before decaying. The electrons from a semi-leptonic heavy flavour decay and from conversions are real electrons, but they are not prompt electroweak electrons and are therefore both labelled fakes. Another source of electron fakes is misidentified light hadrons. These objects are not electrons but hadrons that have been identified as an electron. The electron identification algorithm Tight++ is specifically designed to reject these misidentified hadron, so loosening the tightness of this algorithm greatly enhances number of misidentified hadrons in the sample. The same applies to electrons from conversions which are also highly suppressed

by the identification criteria. The inverted cuts are shown in table 6.3. These have been developed in the multilepton group and optimised with respect to fakes in the three-lepton signal region. The electron denominator objects must pass inverted cuts but candidates that fail both nominator criteria are rejected as they are too far from the nominator objects.

Electron fake selection		
Cut	isEM ID tightness	Impact parameter significance
Numerator	Tight++	$\frac{|d_0|}{\sigma_{d_0}} < 3$
Denominator	Loose++	$3 \leq \frac{|d_0|}{\sigma_{d_0}} < 10$

Table 6.3: The nominal and inverse selection of electrons used to measure fake-factors.

Muon denominator definition

The primary source of non-prompt muons are semi-leptonic heavy flavour decays [74]. These will have larger impact significance parameters just like the electrons from semi-leptonic heavy flavour decay. Muons from hadrons like Pions and Kaons and misidentification from punch-through will all have significantly poorer isolation than prompt muons and these can be selected by loosening the isolation requirements for the object. The inverted cuts for muon fakes can be seen in table 6.4. The muon denominator objects must fail at least one nominator selection cut and pass both of the inverted cuts.

Muon fake selection		
Cut	track isolation	calorimeter isolation
	for $p_T^\mu < 100 GeV$:	
Numerator	$\frac{EtCone30}{p_T^{mu}} < 0.10$	$\frac{PtCone30}{p_T^{mu}} < 0.10$
Denominator	$\frac{EtCone30}{p_T^{mu}} \geq 0.10$	$\frac{PtCone30}{p_T^{mu}} \geq 0.10$
	for $p_T^\mu \geq 100 GeV$:	
Numerator	$EtCone30 < 10GeV + 0.01 \cdot p_T^\mu$	$PtCone30 < 10GeV + 0.01 \cdot p_T^\mu$
Denominator	$\frac{EtCone30}{p_T^{mu}} < 2.0$	$\frac{PtCone30}{p_T^{mu}} < 2.0$

Cut	Impact parameter significance
Numerator	$\frac{|d_0|}{\sigma_{d_0}} < 3$
Denominator	$3 \leq \frac{|d_0|}{\sigma_{d_0}} < 10$

Table 6.4: The nominal and inverse selection of muons used to measure fake-factors.

During the studies it was found that the amount of high-p_T muon fakes was too low to give proper fake-factor estimates. One of the cuts reducing the number of muons was the ΔR requirement between muons and jets. As muon fakes from semi-leptonic heavy flavour decays are expected to be close to or within jets this requirement was removed for muons with p_T above 40 GeV. This has been studied in detail in the multilepton group and is described in [62]. The requirement is loosened in both control regions for the fake factor estimate to give the correct final estimate. The muons are set to fail isolation if either or both of the isolation cuts fail.

6.4.2 Fakes in three-lepton finale states

For an event with three reconstructed leptons either one, two or three leptons could potentially be fake leptons. The three nominator leptons are however not identical, as the triggered lepton is required to have p_T above 26 GeV and is affected by the trigger algorithms. For both the lowest unprescaled electron and muon trigger, isolation criteria are applied significantly enhancing the probability of the lepton to be real. The combinatorics are:

$$\begin{aligned}
l_{trig}^N l^N l^N &= l_{trig}^R l^R l^R + l_{trig}^F l^R l^R + l_{trig}^R l^F l^R + l_{trig}^R l^R l^F + l_{trig}^F l^F l^R \\
&+ l_{trig}^F l^R l^F + l_{trig}^R l^F l^F + l_{trig}^F l^F l^F
\end{aligned} \tag{6.2}$$

where l^R are real leptons and l^F fake. Including leptons from the denominator selection and using the fake-factor definition:

$$\begin{aligned}
f_1 \cdot l_{trig}^D l^N l^N &= l_{trig}^F l^R l^R + l_{trig}^F l^F l^R + l_{trig}^F l^R l^F + l_{trig}^F l^F l^F \\
f_2 \cdot l_{trig}^N l^D l^N &= l_{trig}^R l^F l^R + l_{trig}^R l^R l^F + l_{trig}^F l^R l^F + l_{trig}^R l^F l^F + l_{trig}^F l^F l^R + l_{trig}^F l^F l^F \\
f_1 \cdot f_2 \cdot l_{trig}^D l^D l^N &= l_{trig}^F l^F l^R + l_{trig}^F l^R l^F + l_{trig}^F l^F l^F
\end{aligned} \tag{6.3}$$

Inserting equations 6.3 into eq. 6.2 results in:

$$\begin{aligned}
l_{trig}^N l^N l^N &= l_{trig}^R l^R l^R + f_1 \cdot l_{trig}^D l^N l^N + l_{trig}^R l^F l^R + l_{trig}^R l^R l^F + l_{trig}^R l^F l^F \\
&= l_{trig}^R l^R l^R + f_1 \cdot l_{trig}^D l^N l^N + f_2 \cdot l_{trig}^N l^D l^N - l_{trig}^F l^F l^R - l_{trig}^F l^R l^F - l_{trig}^F l^F l^F \\
&= l_{trig}^R l^R l^R + f_1 \cdot l_{trig}^D l^N l^N + f_2 \cdot l_{trig}^N l^D l^N - f_1 \cdot f_2 \cdot l_{trig}^D l^D l^N
\end{aligned} \tag{6.4}$$

The last line can better be understood by defining the regions: DNN where the triggered lepton fail the nominal selection object, NDN where one of the non-triggered leptons fail nominal selection and DDN where both the triggered and one of the non-triggered leptons fail nominal selection requirements. With this definition, the total number of fakes in the three-nominal-lepton final-state is:

$$N_{fakes}^{3l} = f_1 \cdot N_{DNN} + f_2 \cdot N_{NDN} - f_1 \cdot f_2 \cdot N_{DDN}, \tag{6.5}$$

where N_X is the number of events in the X'th region. In practice the f_1 and f_2 fake-factors will be estimated separately for electrons and muons and applied as appropriate to the flavour of the denominator object.

6.4.3 Fake-factor control region

To estimate $f_{1,e/\mu}$ and $f_{2,e/\mu}$ in eq. 6.5 a region orthogonal to the signal region must as mentioned be used. It should preferably have the same ratio of heavy- to light flavour fakes as the signal region. One region shown to work especially for muons [62, 75, 76], is the region with two same-flavour same-sign leptons. The same sign requirement helps to greatly enhance the number of fake candidates and suppress prompt leptons especially from W/Z + jets events. The region was found to work well for electron fakes as well and has been used for both electron and muon fake factor estimates. The regions allows for both f_1 and f_2 to be determined as

there is a triggered and a non-triggered lepton in the event. To reduce the amount of prompt leptons in the sample the Monte Carlo contribution from WZ, ZZ, WW, $t\bar{t} + jets$, singletop and $t\bar{t} + V$ has been subtracted from both nominator and denominator objects using the datasets shown in table 6.1. W and Z + jets have not been subtracted. The contribution from both will be estimated as part of the fake estimate. The inadequate modelling of V+jets seen in section 5.2 could be improved by including dedicated heavy flavour samples but as the third lepton in Z + jets in the signal region will be a fake, it is appropriate to included the contribution from V + jets in the fake estimate.

If fake-factors were independent of energy/momentum and geometry an average fake-factor could be used but during this study the fake-factors were found to be dependent on p_T. The electron fake-factors as a function of p_T can be seen for the triggered denominator objects in figure 6.1a and for the non-triggered objets in 6.1b and for the muons in figure 6.2a and 6.2b. The bins size is variable defined as [15-35,35-55,55-100,100-250,250+] in GeV to compensate for the falling statistics with higher p_T. The last bin has a large uncertainty but the contribution from fake objects at p_T above 250 GeV in the signal region is very small.
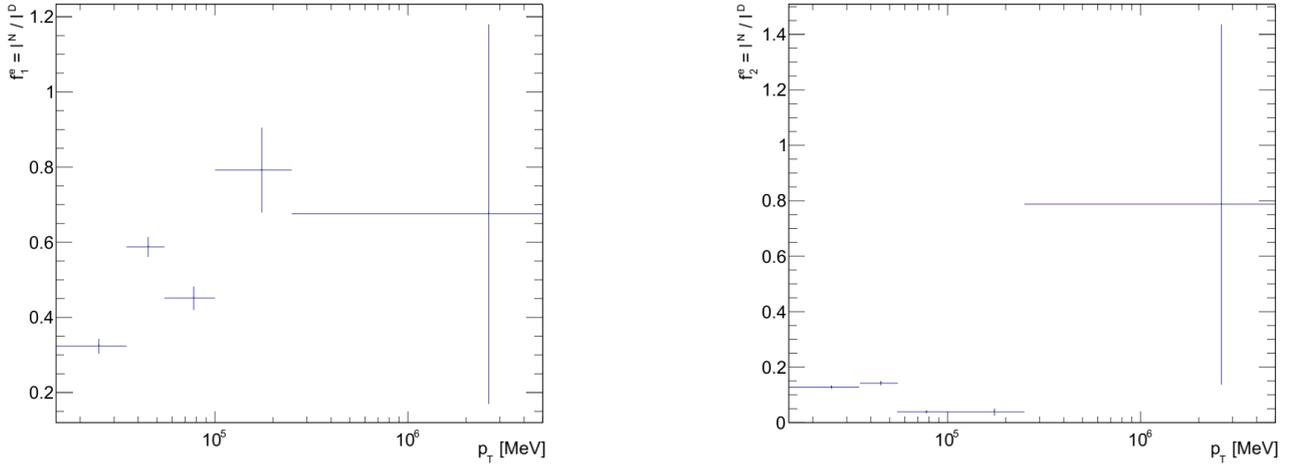
(a) Estimated electron fake-factors as a function of the p_T of the trigger-matched electrons.

(b) Estimated electron fake-factors as a function of the p_T for non-triggered electrons.

Figure 6.1: Bins defined as [15-35,35-55,55-100,100-250,250+] in GeV. It is worth recalling that the lowest unprescaled trigger for electrons has isolation criteria up until 60 GeV.
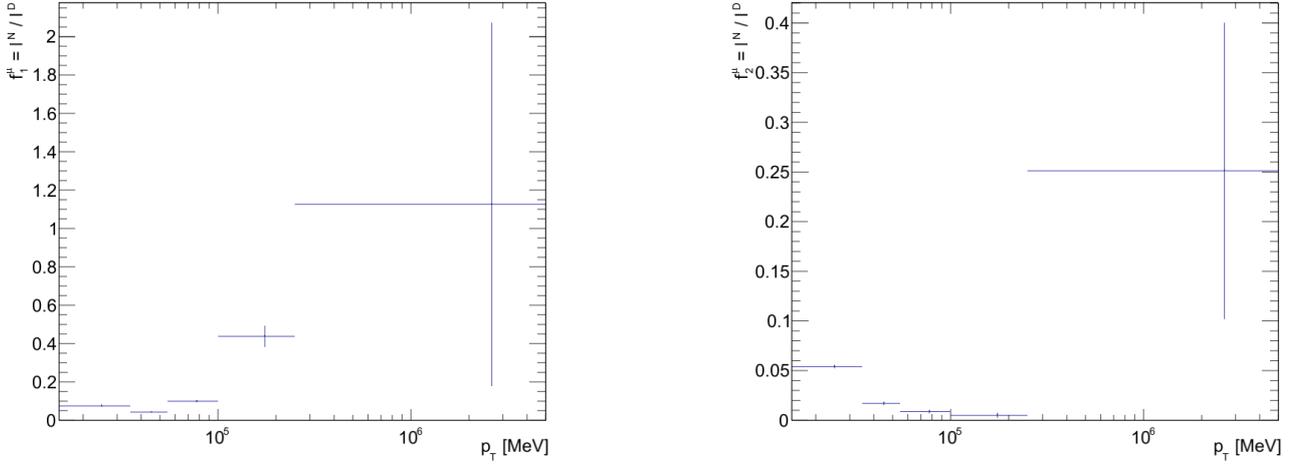
(a) Estimated muon fake-factors as a function of the p_T of the trigger-matched muons.

(b) Estimated muon fake-factors as a function of the p_T for non-triggered muons. It is worth recalling that the lowest unprescaled trigger for muons has isolation criteria up until 36 GeV.

Figure 6.2: Bins defined as [15-35,35-55,55-100,100-250,250+] in GeV.

Fakes in the signal region

The fake contribution to the signal region is estimated from a region identical to the signal region but with the requirement that at least one of the three leptons is a denominator lepton. Again the MC contribution is subtracted for all processes except V + jets. The resulting number of events in the triggered/non-triggered regions are shown in table 6.5.

	nDNN	nNND	nDDN
data	378	11988	705
Monte Carlo	92.4	4103.6	148.6

Total after subtraction		
286	7884	556

Table 6.5: The number of events with either- or both a triggered and non-triggered denominator object for the fake estimate.

To get the fake contribution in the signal region, 286 DNN events, 7884 NND events and 556 DDN events are stored with the appropriate weights following equation 6.5. The weight of the event is determined by the associated fake-factor dependent on flavour, region and the p_T bin the objects falls in. This results in a total of 857 ± 45 (stat) fakes in the signal region. The uncertainty is calculated as the quadratic sum of the errors defined by the error on the fake-factor. This includes the contribution from Z (and W) + jets.

Improvements to the method

The method used here gives an estimate of the number of fakes, but as it has not been the main focus of this study, it can be improved. A thorough study of the composition of heavy and light flavour fakes in the control regions and signal region would reveal whether the ratio of fake types is correct. Furthermore, the estimate might gain from increased statistics for the triggered fake-electrons by the inclusions of pre-scaled triggers with a lower p_T threshold or the use of another control region. The extensive study of an improved fake-estimate is worthy of a PhD. in itself and has not been investigated further here. The fake-factor estimate is of course associated with systematic uncertainties that will be summarised in section 6.5.

6.5   Systematic uncertainties

When analysing ATLAS data and fitting Monte Carlo distributions to it, the result will be influenced by a range of systematic uncertainties. Some constitute overall normalisation uncertainties whereas others are for objects like electrons and muons and other again are specific to a given process. Systematic uncertainties that have been calculated by other studies, have been used in this study, if they can be directly applied. Other uncertainties have been calculated for this analysis specifically. The source has been clearly cited if the uncertainties are taken from another source.

How systematic uncertainties are summed depends on their correlation. Uncorrelated systematic uncertainties are added in quadrature whereas fully correlated uncertainties are added linearly. The determination of correlation between systematics is a huge study in itself. For the work presented in this thesis all systematic uncertainties are assumed to be either fully correlated or completely independent, as other ATLAS analyses do.

6.5.1   Global systematics

The global systematics refer to all systematic uncertainties that influence the total number of events. This involves uncertainty on integrated luminosity as well as cross-section of simulated processes.

- The ATLAS official uncertainty on the integrated luminosity is $\pm$ 2.8%. It is derived, following the same methodology as that detailed in [77], from a preliminary calibration of the luminosity scale derived from beam-separation scans performed in November 2012. This affects all data distributions and is reflected in an uncertainty in the total number of events, but not on the shape of the distribution.

- The uncertainty on the cross-section of Monte Carlo samples comes from two sources: the *parton distribution function*, PDF uncertainty and the renormalisation scale uncertainty. The uncertainties have been summarised in table 6.6.

  - Z+jets and W + jets uncertainties have both been taken from [78]. The total uncertainty follows the PDF4LHC [79] standards, as does the following uncertainties quoted. This means combined uncertainty is the scale and PDF uncertainties added in quadrature.

  - For $t\bar{t} + V$ the uncertainties have been calculated by the top working group. The scale and PDF uncertainty is taken from [80] as the final quoted uncertainties

for both $t\bar{t} + W(j)$ and $t\bar{t} + Z(j)$. This choice is driven by the larger, and more conservative, scale variation range and also the consistent PDF variation which is not quoted in [81]. Although [80] does not quote these uncertainties for ttbar+Z the top group concluded from [81] that the scale uncertainties are of the same size for ttbar+W and ttbar+Z and therefore quote the same uncertainty for both processes. The largest uncertainty is taken per variation and symmetrised, resulting in a 21% scale uncertainty and 8% PDF uncertainty. Adding these in quadrature gives the total uncertainty on the cross-section of 22%.

– For ZZ and WZ (and WW) the cross-section uncertainties are also given in [78] for the MCFM generator but have been specifically studied for the sherpa generator and with jets in [61] appendix A.5. The uncertainties from the dedicated study are close to the uncertainties in [78] but slightly lower. The uncertainty of [61] quoted for sherpa has been used for ZZ and WZ and the total is the scale and PDF added in quadrature.

– $t\bar{t}$ and single-top have rather discrepant uncertainties listed in various studies with the result in [61] being the largest known to the author. The quoted total 10% is therefore used for both process types.

<div align="center">Systematic uncertainties on cross-section of Monte Carlo samples</div>

Sample	Scale uncertainty (%)	PDF uncertainty (%)	Total uncertainty (%)	Reference
Z + jets	(+4.5, -3.9)	(+3.3, -3.8)	(+5.6,-5.4)	[78]
W + jets	(+4.5, -3.9)	(+3.3, -3.8)	(+5.6,-5.4)	[78]
$t\bar{t} + V$	± 21	± 8	± 22	[80] and [81]
WZ	± 6.6	± 3.8	± 7.6	[61]
ZZ	-	-	± 4.3	[61]
WW	-	-	± 5	[61]
$t\bar{t}$	-	-	± 10	[61]
single top	-	-	± 10	[61]
$Z \to ll + \gamma$	-	-	± 17	[82]

Table 6.6: The systematic uncertainties associated to the Monte Carlo samples used for Standard Model expectation modelling. The PDF and scale constituents of the total uncertainty are given where available.

All global systematics are assumed to be uncorrelated from process to process, with the exception of luminosity that is 100% correlated between processes. The uncertainties are implemented by allowing the total number of events expected to float in the fit within the Gaussian constraint described in section 4.4.2. As table 6.6 shows some uncertainties are asymmetric. This has been implemented by using a two-sided Gaussian with sigma on each side set to the quoted value.

6.5.2   Fake systematics uncertainties

The fake or reducible background estimate has uncertainties associated to both the estimate of fake events and the fake-factors. The MC subtraction of nominator and denominator objects is affected by the uncertainty from the MC samples used. A 10% uncertainty is used consistent with the cross-section uncertainty on $t\bar{t}$. The MC subtraction varies from 20 to 35%

on average between the NDN, DNN and DDN regions. As the final number of estimated fakes is, however, dependent on the weights associated with each region, an uncertainty associated to the migration between regions is needed. As a test of the effect the number of events in the DNN and NDN regions was varied up and down by 10% respectively. The result was a difference between bins ranging from 0 to around 10% and an overall variation in events of 10%. This indicates a linear dependence so a 10% uncertainty on the MC subtraction of 30% results in at least up to 3% variations in both the final total number of estimated fakes and bin-to-bin variations. The statistical uncertainty on the number of events falling in the triggered/non-triggered regions will result in a similar contribution. The largest value is obtained from the low stat DNN region which results in 5% uncertainty on the final estimate. These contributions are all added as one total uncertainty of 10% assumed to be fully correlated across bins.

The fake-factors were seen to vary with p_T for both electrons and muons. The errors have been calculated by propagating the fractional uncertainty from the nominator and denominator objects to an uncertainty on the weight for each event. The effect is generally below 10% and resulted in a total uncertainty of 45 events. The contribution to each bin in the final fake distributions is the quadratic sum of uncertainty on weights for events falling in that bin. These are assumed to be fully correlated across bins. Charge flips and conversions are not included as a systematic uncertainty in the fake studies as there are no requirements on the charge composition in the signal region while photons from bremsstrahlung are included in the fake-factor estimate.

Finally the fraction of fakes from heavy- and light flavour sources was assumed to be identical in both control regions and the signal region. The normalised distributions of the MV1 b-tagging variable is plotted in figure 6.3a for the $\mu^{\pm}\mu^{\pm}$ control region and for the $\mu\mu\mu$ subsample of the three lepton control region. The first bin is completely dominated by c and lighter flavour jets and the last bin is dominated entirely by b-jets. As a reference the composition of b- and c-jets in D-meson decays is included in figure 6.3b. The difference in the first bin is 16% and 32% in the last bin. Clearly the composition of flavours in the jets are not identical for the two regions and a systematic uncertainty should be introduced to describe this. Ideally this should be done by a detailed Monte Carlo study but this has not been within the scope of this thesis. The distributions agree much better for electrons where variations are below 1% for all high-statistics bins. As a consequence a systematic uncertainty of 30% has been added on the fake events with a MV1 value above 0.7 if the event contains at least one muon and a 16% uncertainty if the MV1 tag is below 0.7.

6.5.3   Event, object identification and reconstruction systematics

There is a set of systematic uncertainties resulting from corrections of object- or event-reconstruction. All scale factors mentioned in section 6.2.5 could affect the shape of the distributions. The lepton ID and reconstruction correction scale factors affect the total number of reconstructed leptons passing the selection requirements and the uncertainty on the correction therefore constitute a central systematic uncertainty. Energy measurement corrections also affect the efficiency for leptons close to the threshold and as such the uncertainty on the correction constitutes a systematic uncertainty on the total selection efficiency. Likewise the energy scale correction and trigger efficiency have systematic uncertainties associated. As the aim of this study is to fit the distributions resulting from the Principal Component Analysis any systematics distorting the shape of these distributions, could directly influence the
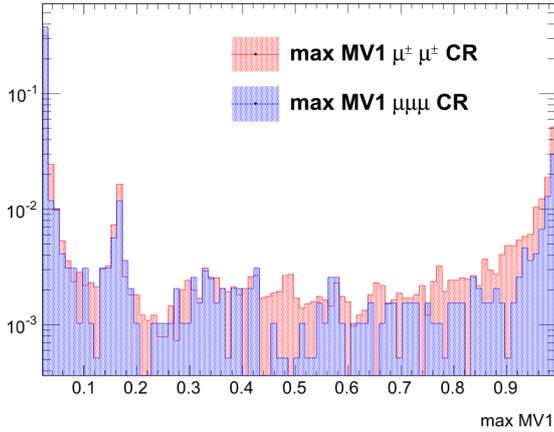
(a) The maximum value of the b-tagging variable MV1 for muon events in the two control regions.
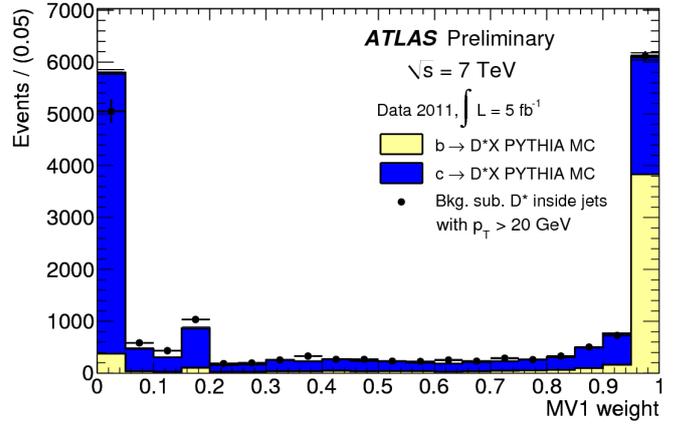
(b) D-meson decays in ATLAS as reference for the MV1 composition of b- and c-jets [83].

Figure 6.3: MV1 b-tagging values.

final result. The Principal Component Analysis is a linear composition of the input observables so any *principal component* used in the final fit should conservatively be assumed to be correlated to all input variables.

Event weight uncertainties

Some of the uncertainties are on the scale-factor or weight applied to each event. This is the case for trigger efficiency, lepton isolation and impact parameter corrections and the lepton reconstruction and identification scale-factors. The scale-factors are applied as an event weight and the tools provide the error associated with each weight. The uncertainties provided by the official ATLAS tools are based on underlying distributions typically in p_T and η resulting in bin-to-bin correlations. The uncertainties are therefore assumed to be fully correlated across bins. They are furthermore assumed to be correlated to each other and are implemented using one common nuisance parameter. The effect on overall number of events passing selection is seen in table 6.8.

A test of shape dependence of remaining systematics

For energy smearing, momentum corrections and energy scale corrections it is customary to quantise the effect as an overall up- or downwards fluctuation in number of selected events, but the shape of distributions of *principal components* could be influenced by these.

The full selection has been run twice for each of the energy and momentum corrections. Each time the systematic effect is varied up or down one sigma using the errors on the correction provided by the official tools, as recommended by ATLAS. Figure 6.4 shows the distributions of these variations in the highest ranked *principal component* for a range of processes

and uncertainties[2]. Below each plot is the *variation percentage* from the nominal defined as:

$$\frac{\Delta_{up(down)}}{N_{nominal}} \cdot 100\%, \tag{6.6}$$

where $\Delta_{up(down)}$ is the difference between the nominal selection and the selection with the systematic varied up or down one standard deviation and, $N_{nominal}$, the number of events passing the selection with the given smearing applied. The *variation %* has only been calculated for bins with at least 10 events and the following results are based on these. If the bin-uncertainties, for the given variation, are correlated the behaviour of the bins when varying up and down are expected to be coherent e.g. to appear synchronised. In other words the *variation %* of a bin varied up is expected to be approximately equal to the *variation %* of the bin varied down. This is to some extent the case for the distributions in Figure 6.4 but to further test if an actual shape shift occurs the *variation %* up has been fitted with a first degree polynomial. A shift of the shape would likely result in a slope on the line, but this is not observed. $Z \rightarrow ee + \gamma$ and $t\bar{t} + W$ does not fit a straight line very well, but the combined samples, $Z \rightarrow ll + \gamma$ and $t\bar{t} + V$ agree within uncertainties. The *variation %* has furthermore been fitted with a zero degree polynomial and the results are in table 6.7 for the highest ranked *principal component*. The probabilities are in general very close to one and the systematic uncertainties can be treated as an overall uncertainty on the normalisation. WW and $t\bar{t}$ + jets have so few events passing selection even before scaling that the distributions are completely dominated by statistical fluctuations and they have not been presented in table 6.7.

Obtaining the overall process normalisation

The overall normalisation to be applied to each process has been obtained from the performed fits to a flat line. The fitted constant gives the value of the deviation in % of the 1 σ variation up or down. As the normalisation should affect the distributions in each *principal component* equally the mean has been taken to avoid vastly overestimating the uncertainty due to statistical fluctuations. The results are presented in table 6.8.[3]

---

[2]MS is an abbreviation for Muon Spectrometer and ID for Inner Detector and the p_T smearing is performed on each separately.

[3]The energy scale correction was found not to alter the number of events passing selection and hence no uncertainty is added from it. WW and $t\bar{t}$ have not been assigned systematic uncertainties due to smearing as the distributions are completely dominated by statistical fluctuations.
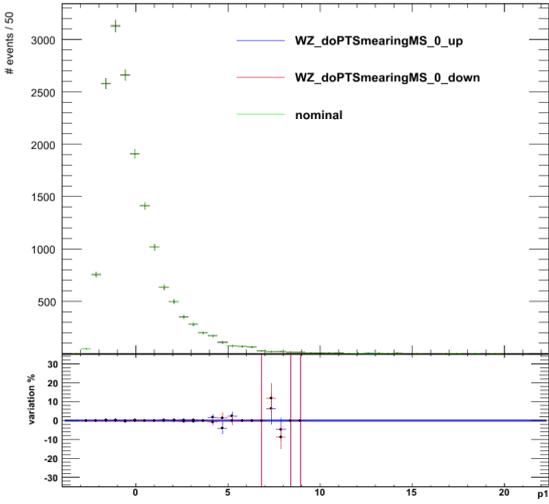
(a) 1. polynomial fit to first *principal component* for WZ with p_T smearing MS.
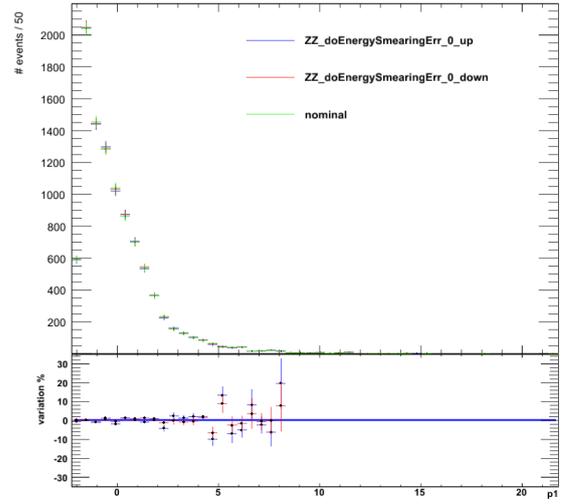
(b) 1. polynomial fit to first *principal component* for ZZ with energy smearing.
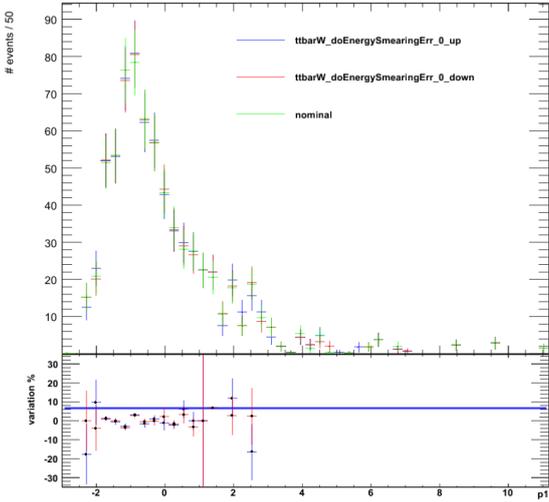
(c) 1. polynomial fit to first *principal component* for $t\bar{t} + W$ with energy smearing.
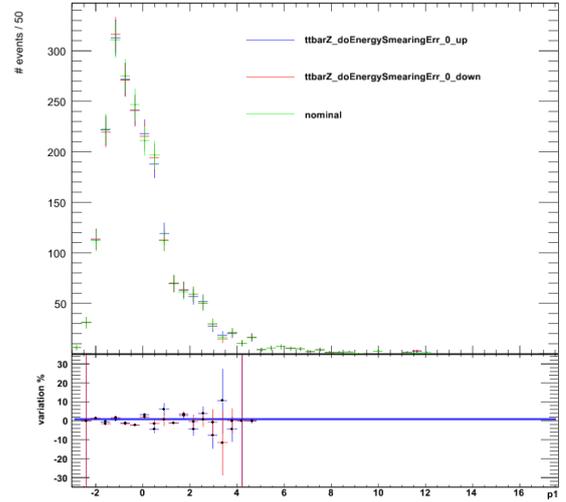
(d) 1. polynomial fit to first *principal component* for $t\bar{t} + Z$ with energy smearing.
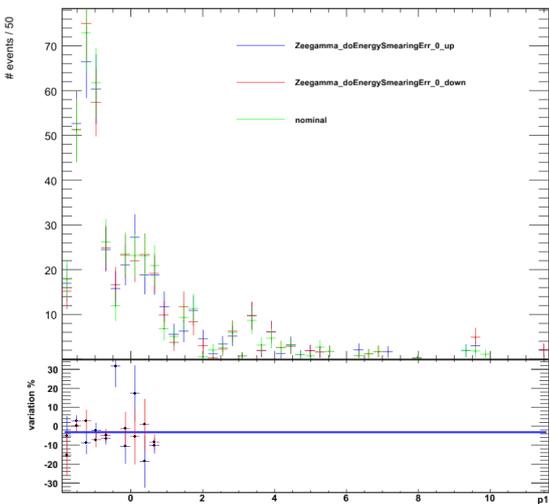
(e) 1. polynomial fit to first *principal component* for $Z \to ee + \gamma$ with energy smearing.
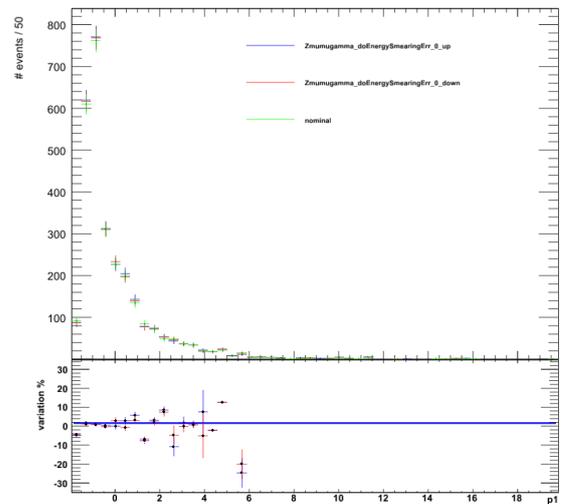
(f) 1. polynomial fit to first *principal component* for $Z \to \mu\mu + \gamma$ with energy smearing.

Figure 6.4: Distributions for energy and p_T smearing varied up and down one standard deviation for the central processes. In general small effects are observed. The blue line in the ratio plots is the fit to the *variation %* for upwards variations.

Process	principal component	probability of χ^2	$\frac{\chi^2}{NDF}$	p_0	error
energy smearing					
ZZ	1	0.59	0.90	-0.03	± 0.05
$t\bar{t} + W$	1	1.00	0.26	0.00	± 0.08
Z γ	1	0.97	0.44	0.14	± 0.12
$t\bar{t} + Z$	1	0.97	0.46	-0.00	± 0.06
WZ	1	0.93	0.60	0.01	± 0.02
p_T smearing ID					
ZZ	1	0.99	0.40	-0.00	± 0.00
$t\bar{t} + W$	1	1.00	0.16	-0.00	± 0.05
Z γ	1	1.00	0.19	0.00	± 0.00
$t\bar{t} + Z$	1	0.98	0.40	0.00	± 0.00
WZ	1	0.99	0.40	0.00	± 0.01
p_T smearing MS					
ZZ	1	1.00	0.38	-0.00	± 0.04
$t\bar{t} + W$	1	1.00	0.19	0.21	± 0.70
Z γ	1	1.00	0.26	0.00	± 0.00
$t\bar{t} + Z$	1	0.99	0.35	0.00	± 0.01
WZ	1	0.98	0.49	-0.00	± 0.00

Table 6.7: The results of a straight line (zero'th polynomial) fit to the *variation %* upwards for each varied systematic effect.

Effect/correction	σ_{up} (%)	σ_{down} (%)
Lepton ID, reconstruction, isolation		
and impact parameter scale-factors	2.6	2.6
Trigger efficiency	0.4	0.3
Energy scale	0	0
Total	2.63	2.62

Effect/correction	$\sigma_{up,mean}$ (%)	$\sigma_{down,mean}$ (%)
WZ		
Energy smearing	0.2	0.2
Momentum smearing ID	0.1	0.1
Momentum smearing MS	0.05	0.1
σ_{tot} (including theoretical)	8.0	8.0
ZZ		
Energy smearing	0.42	0.46
Momentum smearing ID	0.01	0.01
Momentum smearing MS	0.04	0.05
σ_{tot} (including theoretical)	4.7	4.7
WW		
$t\bar{t}$+ jets		
Energy smearing	<0.01	<0.01
Momentum smearing ID	<0.01	<0.01
Momentum smearing MS	<0.01	<0.01
σ_{tot} (including theoretical)	5.0	5.0
$t\bar{t}$+ V		
Energy smearing	0.09	0.05
Momentum smearing ID	0.01	<0.01
Momentum smearing MS	0.03	0.03
σ_{tot} (including theoretical)	22.1	22
$t\bar{t}$+ jets		
Energy smearing	<0.01	<0.01
Momentum smearing ID	<0.01	<0.01
Momentum smearing MS	<0.01	<0.01
σ_{tot} (including theoretical)	10	10
$Z \to ll + \gamma$		
Energy smearing	1.34	1.29
Momentum smearing ID	0.01	0.01
Momentum smearing MS	0.12	0.13
σ_{tot} (including theoretical)	18.5	18.4

Table 6.8: The object, reconstruction and calibration systematic uncertainties. The values presented are the mean effect over all observables for each distribution.

6.6   The use of Principal Component Analysis for final fits

So far the Principal Component Analysis transformation has been described as "a near perfect" method to create independent variables from any input observable but the test of the method in the two-lepton SFSS control region revealed that this is not always the case. Principal Component Analysis correctly transforms a dataset with N observables into a maximum of N new linearly uncorrelated observables. Now the first issue has to do with the generalisation of the transformation matrix. In principle the transformation matrix only guarantees linear independence for the dataset used to generate the matrix. In reality the transformation works on a more general basis and it is possible to transform other similar datasets resulting in still nearly uncorrelated variables.

6.6.1   Multiple principal components for final fits.

If the resulting *principal components* are to be used in a fit they must be independent or at the least with negligible correlations for all fitted processes. This might not be true for the rather complicated situation where several processes exist that all contribute to the three-lepton signal region, each with slightly different characteristics. Furthermore the result of any such test depends on the data used to generate the transformation matrix. Even if all simulated data is used to construct the matrix, it will be largely determined by the dominant processes and linear dependence might still exist for processes with low contribution.

One can to some extent get around this by fixing parameters for processes where the given PCA-matrix does not yield independent eigenvectors. That is, if the parameters for WZ, ZZ and fakes are the only free parameters in a fit, the *principal components* generated for e.g Z + jets are not required to be independent. This can be thought of as a noise reduction where the expected number of events from a set of fixed background processes are removed before performing the fit with the remaining floating parameters and processes.

To test the use of multiple *principal components*, a transformation matrix was generated using only WZ Monte Carlo and the covariance matrices were constructed for all resulting *principal components* in all processes. WZ was chosen following the idea that any transformation used should model well the three-lepton signal region and should be able to catch real prompt leptons from new physics which resembles WZ. The observables used as input to the generation of the PCA_{WZ} matrix affect how independent the resulting *principal components* are for other processes. By trial and error the input observables resulting in lowest correlation between *principal component* in processes other than WZ where found to be:

$$\begin{bmatrix} H_T^{lep} & H_T^{jets} & \not{E}_T & \Delta R & nTracks_{vertex} & M_T^{3lep} \end{bmatrix}$$

The correlation matrix for WZ is of course zero for all entries outside the diagonal per construction. The resulting correlation matrices for ZZ, fakes and real data are shown in figure 6.5. The scales are different but it is clear that neither have exact zero correlation between all *principal components*. The correlation is, however, very low for most of the *principal components*. The resulting *principal components* in data have correlations close to zero for the first three components. The highest correlation is seen for p_2 and p_4 which are roughly 20% anti-correlated and p_1 and p_5 close to 20% correlated. For fakes the first two components are highly uncorrelated and the correlation between the first and third is below 20%. Higher correlations are present for *principal components* after the third. For ZZ p_1 and p_2 and p_1 and p_3 are approximately 15% anti-correlated. Effects like random fluctuations and statistics

of the sample influence the correlation value. The first three *principal components* have been assumed to be independent for now and will go through the *mutual information* test to test this. For all $t\bar{t} + V$ and $Z\gamma$ the first three *principal components* also have correlations below 20%. WW and $t\bar{t}$ + jets have too low statistics for proper values of correlation coefficient to be found. They will be kept fixed in final fits.

A similar test was done using data, ZZ and $Z\gamma$ to generate the Principal Component Analysis matrix and perform the transformation. The results were very comparable for data but for other processes the linear dependence increased.

6.6.2 Test of mutual information in the signal region

The test for mutual information (MI) was performed on all resulting *principal components* for each process, one process at a time. The final transformation used in this thesis is the Principal Component Analysis based on the WZ Monte Carlo that should maximise the ability to describe final states with three real, prompt leptons. Using WZ to generate the PCA-matrix and a threshold of 0.05 (see description of statistical influence on the MI measure in section 5.3) resulted in just two distributions above the threshold: fakes p_2 vs. p_5 and $Z\gamma$ also p_2 vs. p_5. To be even more conservative, distributions in p_1 through p_3 with linear correlation above 0.2 was also found if the MI was above 0.01 and the results are shown in figure 6.6. The value in the upper right corner is the linear correlation. In data no distributions were found with mutual information above 0.01. It should be noted that for data and MC processes with more than 600 events the MI has only been calculated for 600 events. For this sample size even completely independent distributions could result in MI above 0.01 so the threshold should catch distributions even with very low correlation of any type.

The distributions do not show any clear correlations, neither linear nor non-linear (i.e. donut shapes or other clear non-linear structures). p_2 and p_5 have some linear structure resembling anti-correlation for both fakes and ZZ. Both plots have correlations factors of below -20% and some dependence is likely to exist. Focusing on just the first three *principal components*, p_1 and p_3 has a linear correlation of ≈ 20 % for fakes and $t\bar{t}$ + V but this is not clear from the distribution in figure 6.6. p_1 and p_2 for ZZ has some structure but no clear dependence. As already mentioned more clear (linear) correlations existed for p_4 and up (lower ranked) and these will not be included in fits for final results. It is worth noticing that no correlations were found in data even using a lower threshold of 0.01 and studying the first three *principal components*.

For the final results in this thesis p_1, p_2 and p_3 are assumed to be completely independent for all processes.
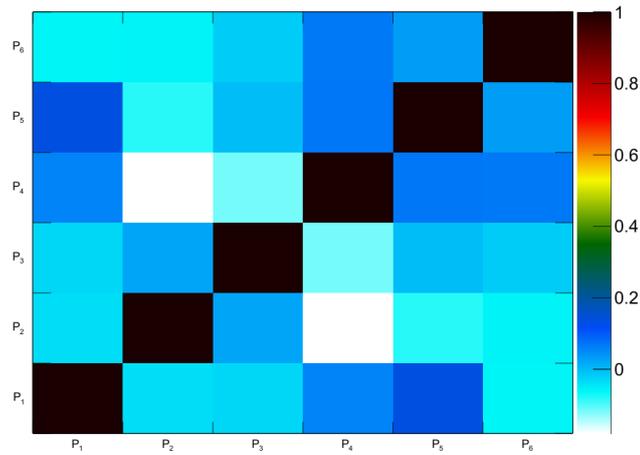
(a) Correlation matrix for 2012 8TeV ATLAS data
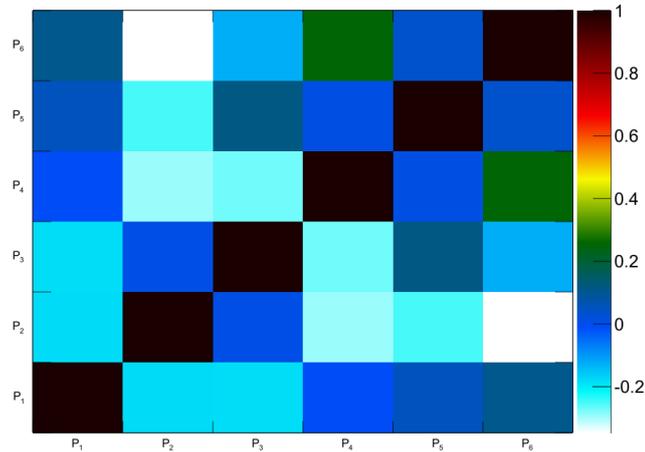
(b) Correlation matrix for ZZ
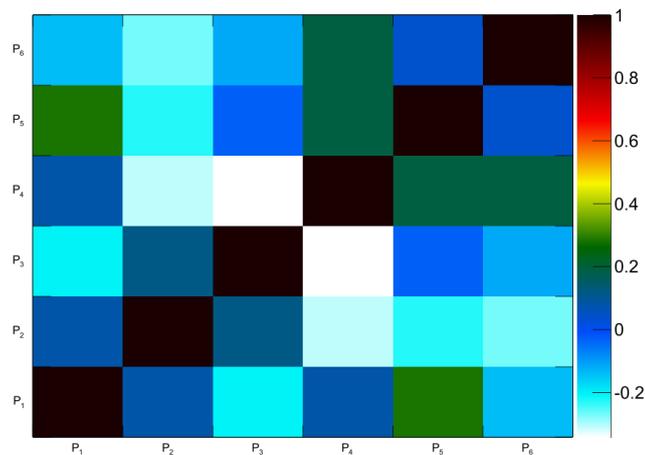
(c) Correlation matrix of fakes

Figure 6.5: The linear dependence of *principal components* in the other major processes after applying a transformation based on a matrix generated using WZ Monte Carlo. Notice the different scales.
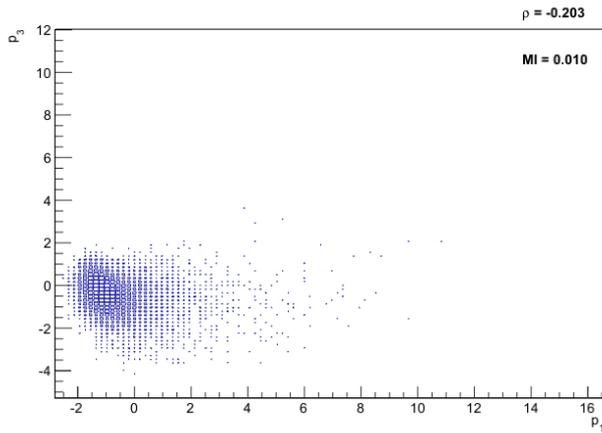
(a) The mutual information of fakes p_1 vs. p_3.
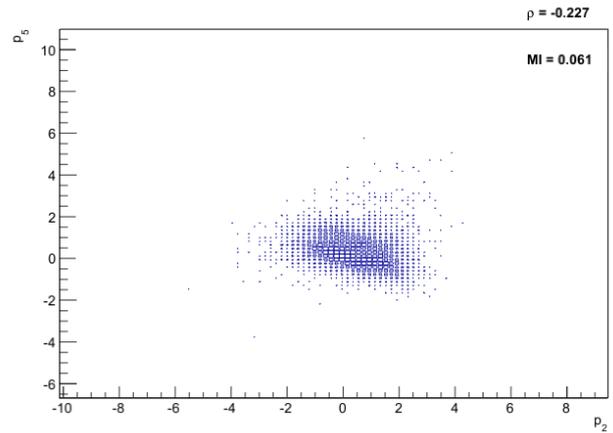
(b) The mutual information of fakes p_2 vs. p_5.
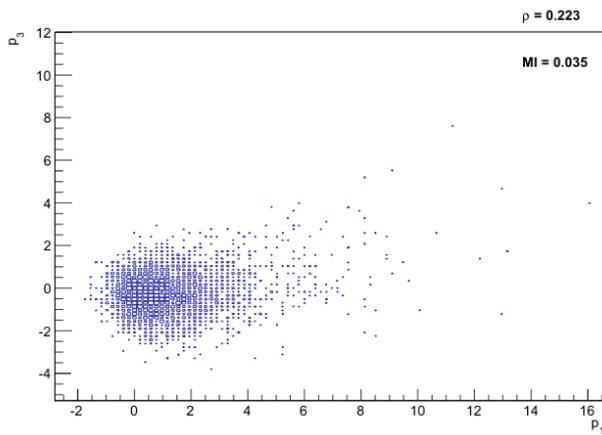
(c) The mutual information of $t\bar{t} + $ V p_1 vs. p_3.
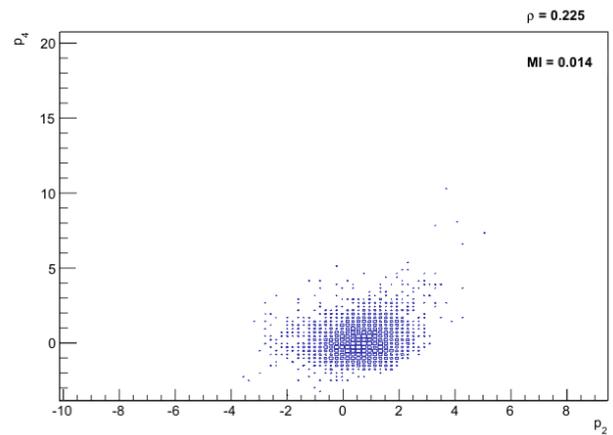
(d) The mutual information of $t\bar{t} + $ V p_2 vs. p_4.
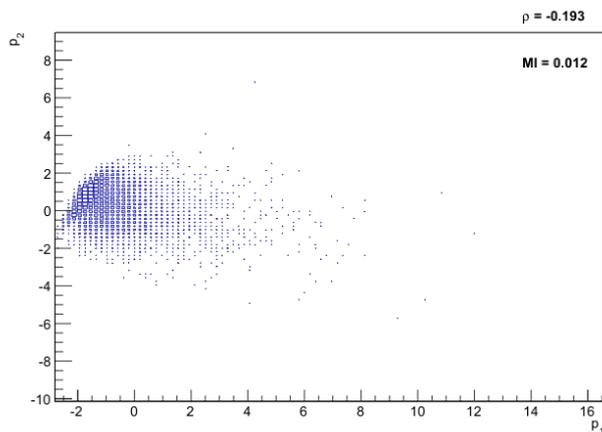
(e) The mutual information of ZZ p_1 vs. p_2.
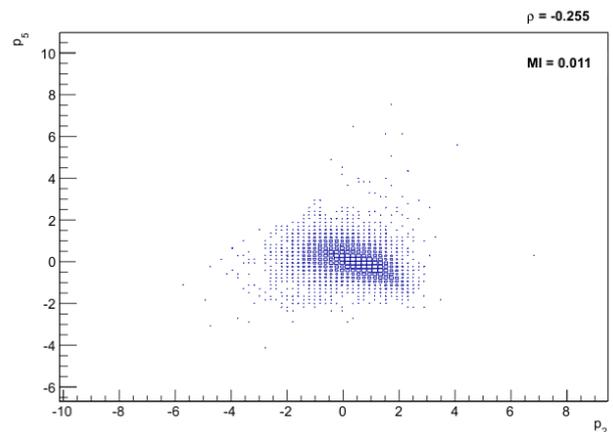
(f) The mutual information of ZZ p_2 vs. p_5.

Figure 6.6: The *principal component* pairs found by the mutual information test as defined in section 4.7.4 to be above the threshold of 0.05 or 0.01 if linear correlation around 20% or more are known to exist. Further correlation exist for p_4 and above. As a reference the linear correlation between the two components, ρ, has been plotted.

6.6.3   Separability of processes in principal components

It is not straightforward to illustrate separation in a multidimensional space, but figure 6.7 attempts to do it by presenting the 2-dimensional distributions of the three most significant processes and data against each other for some of the *principal components* resulting from the PCA$_{WZ}$ transformation. All distributions have been normalised to unity to focus on the shape and not their overall contribution. The processes have been plotted two and two to show their distinct separation. Figure 6.7a through 6.7c show that separation between Standard Model processes is indeed possible to some extent. Figure 6.7d is included to show that several *principal components* contain separation power between data and WZ. This is essential if anything but WZ must be measured although it is no guarantee that a new signal can be clearly separated. Figure 6.8a shows the 3-dimensional distribution of WZ, ZZ and fakes - the three most dominant processes - here in three *principal components*. Figure 6.8b shows the projection in the p_1 vs. p_2 plane. The plots illustrate that there is clear separation potential in adding additional *principal components* even though all processes undergo the same WZ PCA-matrix transformation.

6.7 p_1 as an optimal (multidimensional) observable

Even without the use of multiple *principal components* in the fit, Principal Component Analysis will be able to deliver a much larger portion of information to the fit than any one-observable fit. The transformation maximises the variance as described in section 4.7.3. The highest ranked *principal components* is therefore the component with the largest information from the original pattern space and it combines information from numerous observables. It will be multidimensional in the sense that the one dimensional *principal component*, will contain information from several of the dimensions of the original pattern space. If correlated observables are added to the pattern space, the eigenvalue and information contained in p_1 will increase.[4]

Including all 10 observables described in section 5.1.4 results in *principal components* with the eigenvalues shown in figure 6.9a. Close to 40% of all information in the eigenvectors are described by the first *principal component* alone. This really is one of the strongest points of the method. A single variable fit using the highest ranked *principal component* will contain much more information than any single observable in the original pattern space. The highest ranked *principal component* has been plotted in figure 6.10. The two plots in normal and logarithmic scale, respectively, show that the fit will be dominated by WZ, ZZ and fakes. $t\bar{t} +$ V, WW and Zγ have very small contributions but will affect the fit in the second, third and fourth bin.

---

[4]To the extent that adding 100% correlated observables will of course not add new information.
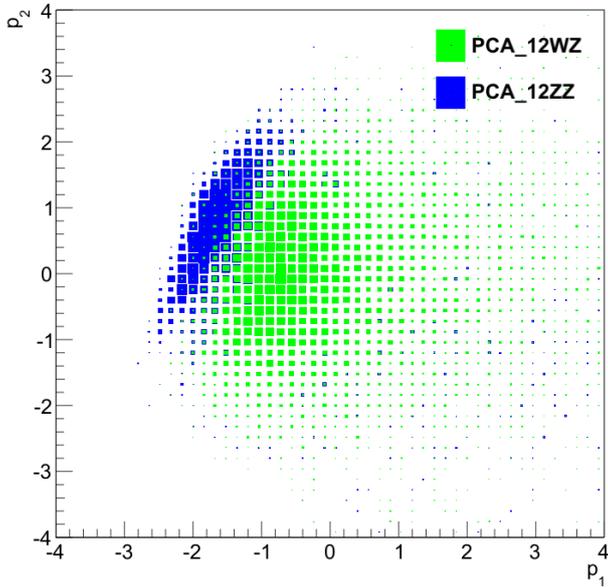
(a) Separation in p_1 and p_2 between WZ and ZZ.
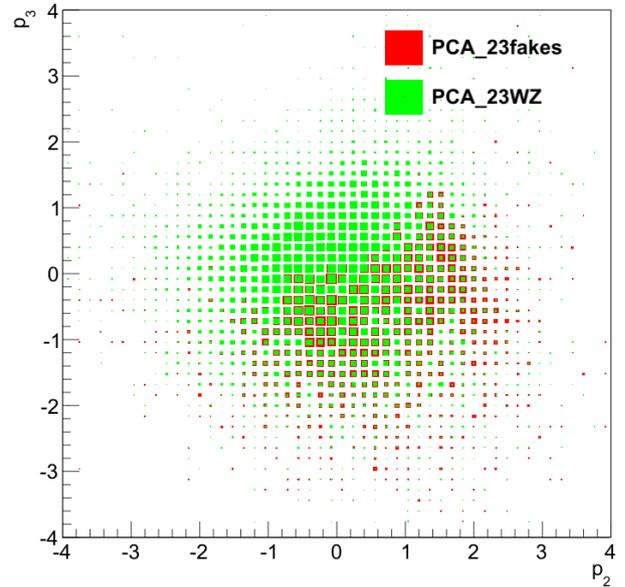
(b) Separation in p_2 and p_3 between WZ and fakes.
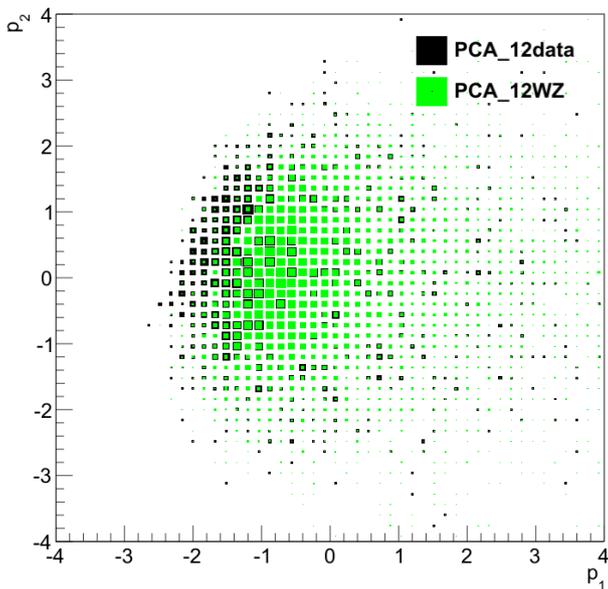
(c) Separation in p_1 and p_2 between data and WZ.
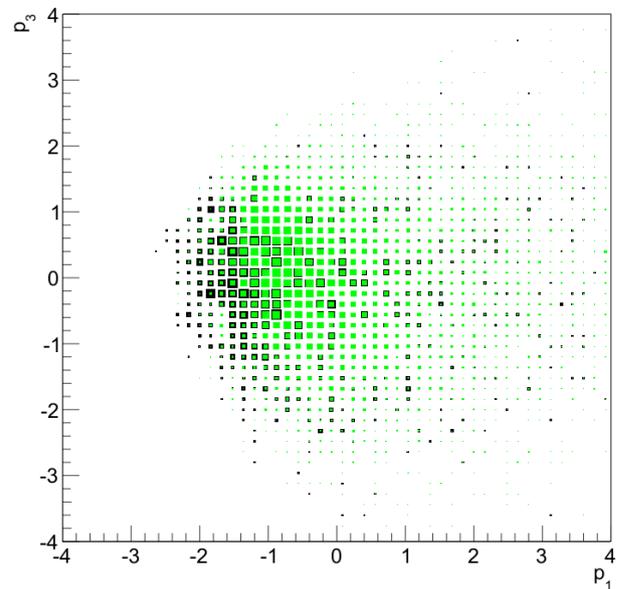
(d) Separation in p_1 and p_3 between data and WZ.

Figure 6.7: The distribution of highest ranked *principal component* made from WZ Monte Carlo for the different Standard Model processes contributing to the signal region. Each template is normalised to unity.
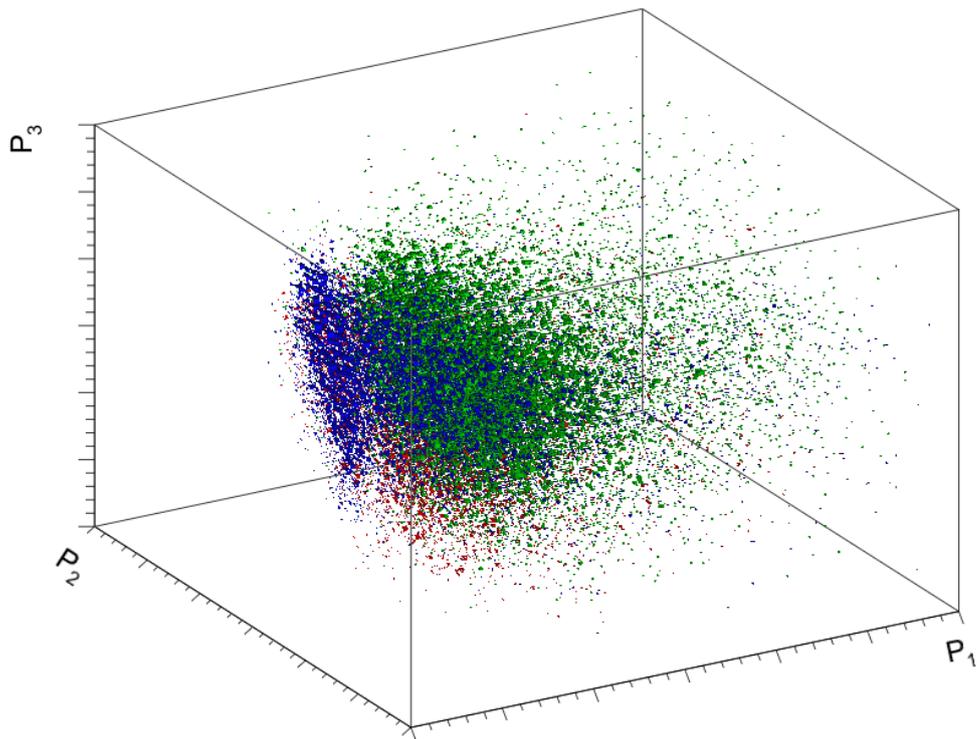
(a) The 3D distribution of p_1, p_2 and p_3 for WZ, fakes and ZZ.
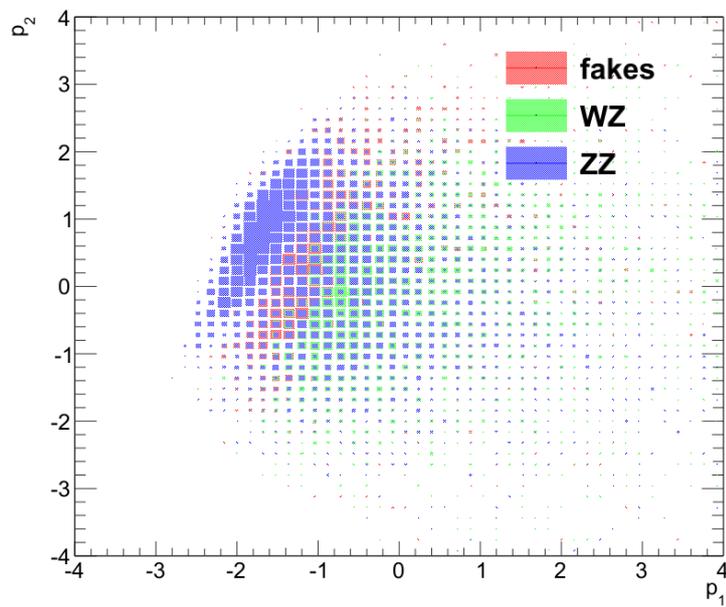
(b) The projection to the p_1,p_2 plane.

Figure 6.8: Separation potential in higher dimension.
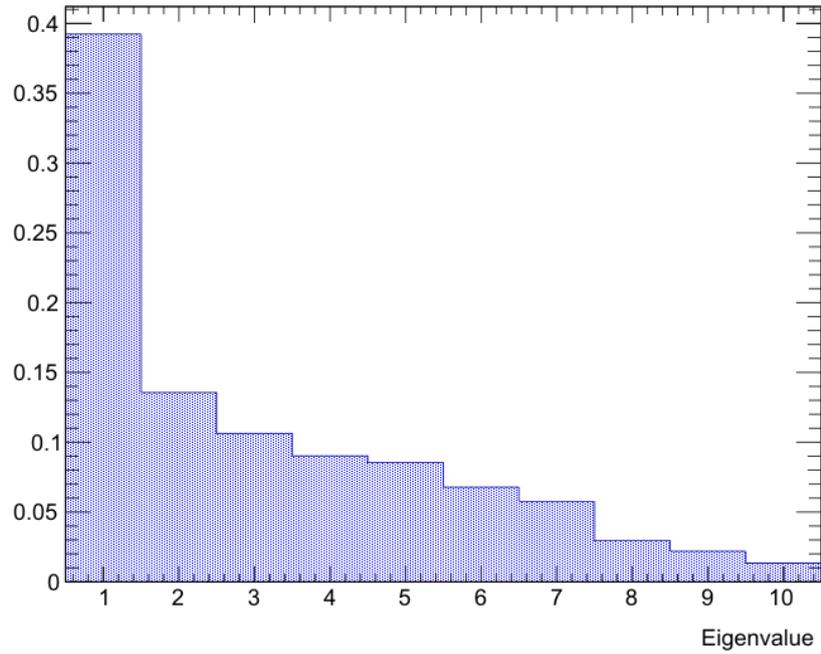
(a) The eigenvalues of the *principal components* of a PCA$_{WZ}$ transformation based on 10 input observables maximising variance in one *principal component*.
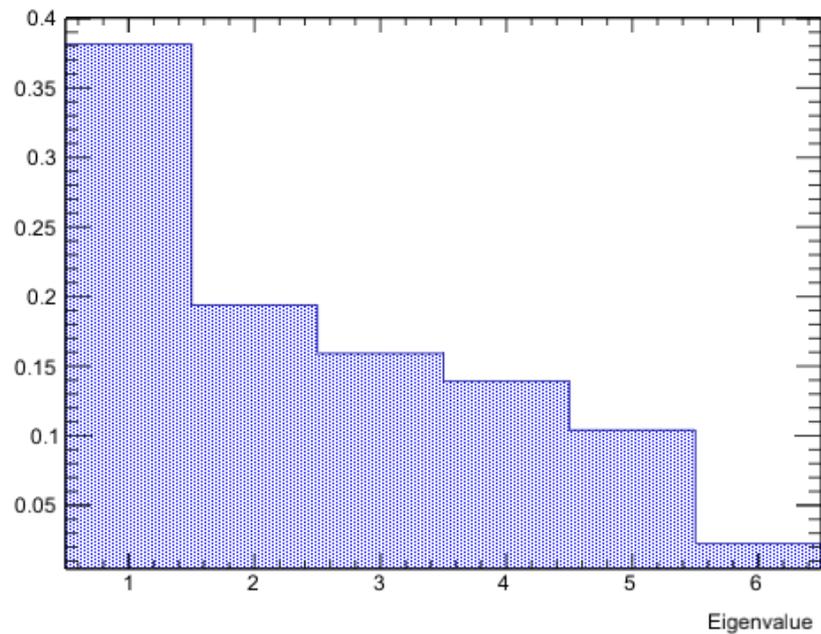
(b) The eigenvalues of the *principal components* of a PCA$_{WZ}$ transformation based on 6 input observables used for fits combining *principal components*. This combination maximises independence between *principal components* in all subprocesses.

Figure 6.9: The eigenvalues of the PCA$_{WZ}$ based on 10 and 6 observables respectively.
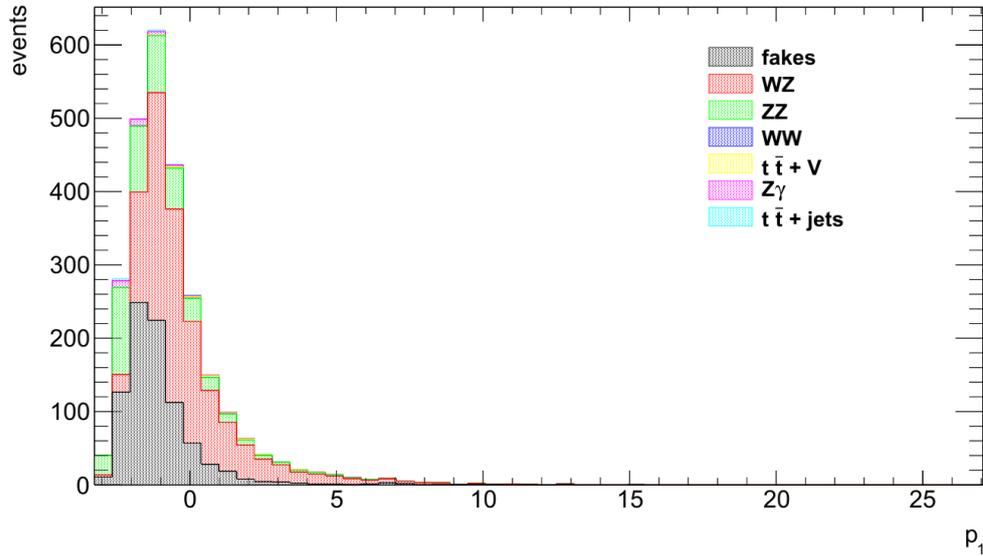
(a) The distributions in 50 bins allow for some separation while keeping the number of empty or low-content bins low.
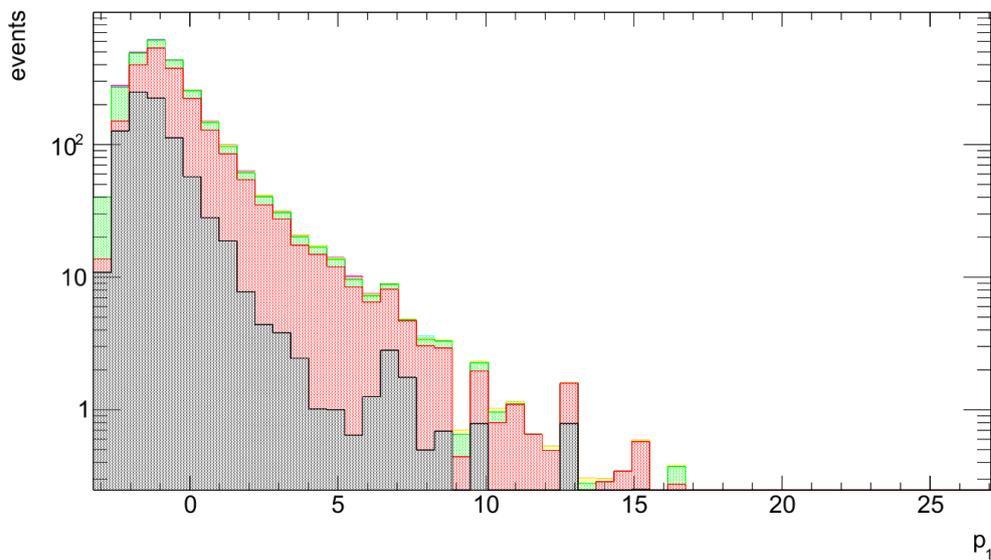
(b) Plotting on logarithmic scale help illustrate the difference in shapes of the processes.

Figure 6.10: The distribution of highest ranked *principal component* of the PCA_{WZ} transformation, constructed from the 10 input observables. All Standard Model processes contributing significantly to the signal region are included.

6.7.1   Seesaw neutrinos as a benchmark model.

To test the fit method on a search for new physics, the signal originating from the possible existence of type III seesaw heavy leptons is used as a benchmark model. The signal has been generated using madgraph for production and decay of the new signal and initial- and final state radiation is simulated using Pythia, which also handles the decay of the resulting W and Z bosons. The sample is generated with a filter for at least three leptons. The datasets used for the Type III seesaw mechanism heavy leptons are:

Table 6.9: Datasets used for Type III seesaw

dataset (MC) ID	mass of new particle [GeV]	Generated Luminosity $[fb^{-1}]$
158693	95	3020
158694	100	476
158695	120	288
158696	160	749
158697	200	5331
158698	300	17590
158699	500	97559

The simulation was done using the AU2CTEQ6L1 parton density function, which is not as recent as CTEQ10 but for the purpose of benchmark testing it is adequate.

Characteristics of the type III seesaw in PCA

Figure 6.11 presents the distribution of the Standard Model processes along with the distribution for type III seesaw triplets with mass 95,160 and 500 GeV. The observable is the highest ranked *principal component*, p_1 constructed from WZ Monte Carlo using 10 input observables. It is very interesting to see how well the first *principal component* alone, separates the signal from the Standard Model processes for the 95 GeV mass-point.
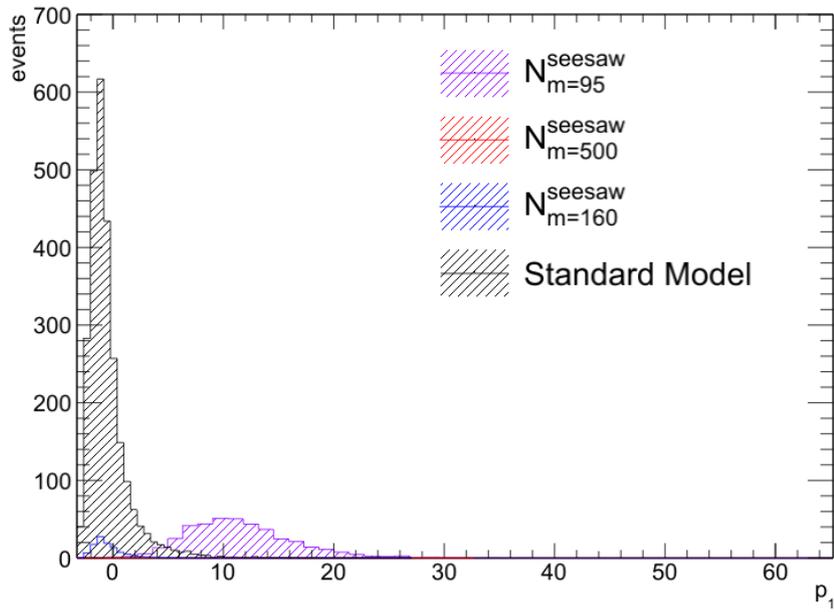
(a)
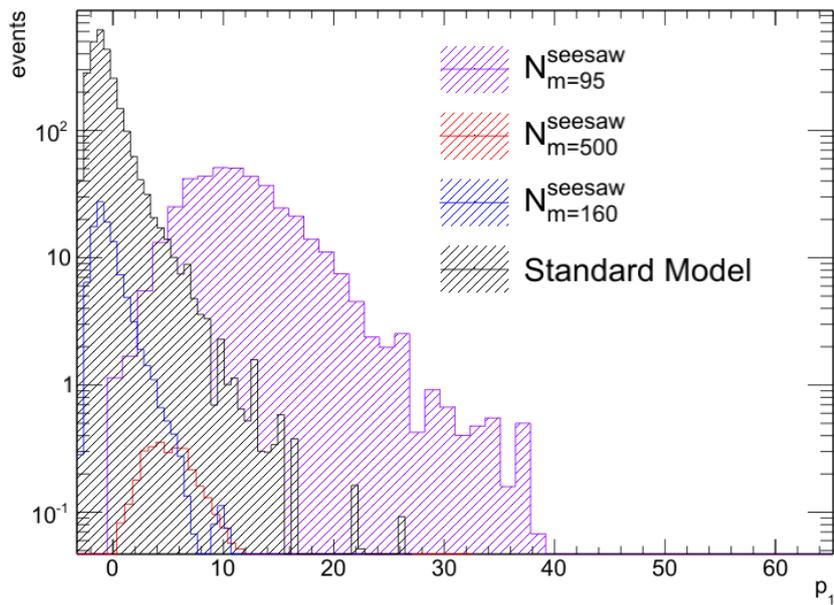
(b)

Figure 6.11: The expected signal in p_1 of the 10 observable PCA_{WZ} for the Type III seesaw triplet with mass 95,160 and 500 GeV.

7

Results using 20 fb^{-1} $\sqrt{s} = 8$ TeV ATLAS data from 2012

The final chapter will present the results made using 20.28 fb^{-1} of ATLAS data recorded in 2012. The methodology applied is central to the thesis, so to summarise the procedure:

- The full selection has been run to select events with exactly three leptons, reconstructed with a large probability to be real and highly isolated as described in section 6.2. In this final state the contribution of each process is described by an associated fit parameter, α. When several *principal components* are used in the fit, α is determined from the distribution of the process in each *principal component* simultaneously, as described in section 4.4.

- The implementation of systematic uncertainties is described in detail in section 6.5. Briefly summarised each processes has an associated uncertainty calculated as the quadratic sum of the lepton identification and reconstruction uncertainties including the uncertainties associated with reweigthing, as well as the theoretical PDF, renormalisation and factorisation scale uncertainties. If *constrained*, the associated fit parameter is constrained to be within the Standard Model expectation by a unit Gaussian with width σ_{sys}. Nuisance parameters also enters for luminosity and energy and momentum correction. The last is incorporated in the individual processes when systematic uncertainties are quoted. Signal parameters are not constrained. The total uncertainty quoted in distributions in this section, is the quadratic sum of the systematic uncertainties and the Monte Carlo statistical uncertainty.

- The Principal Component Analysis will be constructed from a transformation of WZ Monte Carlo and the resulting transformation will be applied to all other processes, to construct their *principal components*. One of two sets of input observables will be used for the construction of the *principal components*. 1) for the use of p_1 as one multidimensional observable, 10 input observables will be used as described in section 6.7. 2) When several *principal component* are fitted the 6 input observables described in section 6.6.1 are used to construct the *principal components* that minimises correlations in subprocesses.

119

7.1 Best fit to the Standard Model using PCA$_{WZ}$

To see how well the Standard Model agrees with data, a fit in the highest ranked *principal component* was performed. All processes where constrained to their Standard Model expectation. A total of 2563 events are observed. After the fit the total number of expected events is 2569 ± 108 (total). The signal strength, μ, calculated from the resulting parameter values, are shown in table 7.1. The total uncertainty includes statistical and systematics uncertainties profiled through the inclusion of nuisance parameters and against other processes. The systematic uncertainty used to constrain the parameter is given for reference. All parameters agree with Standard Model expectations within their uncertainty. The last four processes are dominated by low statistics as reflected by the uncertainty. For both WW, $t\bar{t}$ + jets and $t\bar{t}$ + V the fit did not converge correctly (parameter hit lower limit) and the quoted statistical uncertainty is overestimated. The distribution in p_1 resulting from the fit is shown in figure 7.1. A Kolmogorov-Smirnov test (KS) between data and fitted expectations results in a value of 1.00 indicating a good agreement.

Table 7.1: Best fit values

$\mu = \frac{N_{fit,process}}{N_{exp,process}}$			σ_{stat}	σ_{sys}	$N_{fit,process}$
μ_{WZ}	=	1.02	±0.04	±0.08	1259
μ_{ZZ}	=	0.99	±0.09	±0.05	455
μ_{fakes}	=	1.00	±0.05	±0.10	854
$\mu_{Z\gamma}$	=	1.00	±0.67	±0.19	30
μ_{WW}	=	1.00	-	-	1
$\mu_{t\bar{t}+V}$	=	1.00	-	-	14
$\mu_{t\bar{t}+jets}$	=	1.00	-	-	8
α_{lumi}	=	0.98	±0.02	±0.03	

The correlation between fit parameters is listed in table D.1 in appendix D. Correlation between parameters are in general 5% or less with the exception that the luminosity is correlated to the three major processes. To test the effect of the four low-stat processes an additional fit was performed with the contribution for the four processes fixed to their expectation. This had no influence on the results for the major processes and the result was exactly as quoted in table 7.1.

7.1.1 Goodness-of-fit

To test the goodness of fit both a χ^2 test and a Kolmogorov-Smirnov test have been calculated. The KS value of 1.00 indicates a good agreement between data and Standard Model predictions and the χ^2 test supports this with a probability of 0.992 indicating a very good agreement. To look for local deviations the p-value was calculated in each bin and all deviations were found to be below 1σ.

Looking at table 7.1 again, it is interesting to note that all electroweak processes are well within their uncertainty of expectation. The luminosity is correlated to WZ, ZZ and fakes (all dominant processes) so fixing it, to the nominal value in the fit, as one would do in a cross-section measurement, would predominantly decrease these three processes. To fully test the
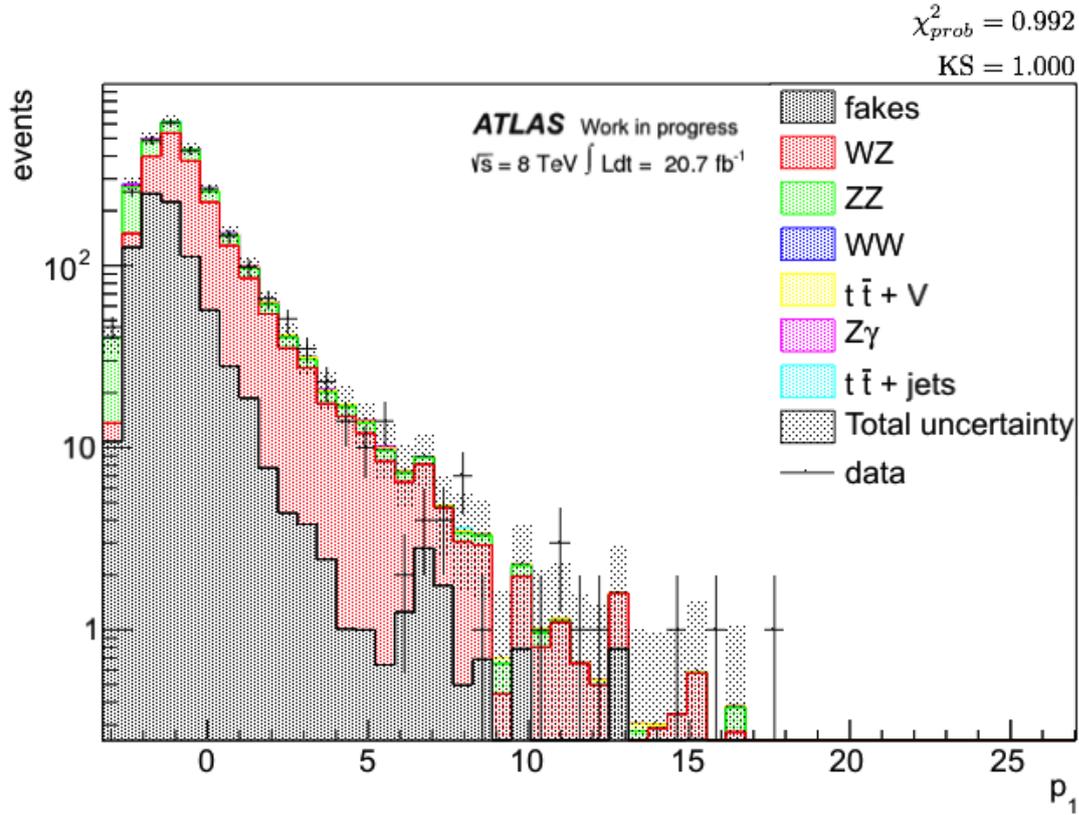
Figure 7.1: The distributions in the highest ranked *principal component* p_1 for the Standard Model expectation fitted to 2012 data from ATLAS. The *principal components* have been generated from the 10 input observables to create one optimal observable.

agreement of the dominant Standard Model contribution, WZ, to data, a fiducial cross-section measurement of WZ will be presented as the next result.

7.2 WZ three-lepton fiducial cross-section

The Standard Model has been so well tested, that most agree that a new fundamental theory should encompass it one way or another. In reality this means that many of the new types of physics we expect to see are likely first to be observed as deviations from the Standard Model. A way to find such deviations is by testing the known extensively and the use of one or more *principal components* could prove very useful for this. The first most general deviation (albeit small) will be seen in the cross-section of a process, and so a measurement of $\sigma_{WZ}(pp \to lll + X)$ for the given acceptance of this study was performed, with X being anything but leptons. The cross-section measurement of WZ is special in the sense that the requirement of independence between *principal components*, is only relevant for WZ and data and not the other Standard Model processes as these enters as a fixed background. Figure 6.5a showed that linear correlation was below 20% between any of the 6 *principal components* in data and no non-linear correlations where found for data nor WZ. For the benefit of testing the method the WZ cross-section has been found in fits including from one to all six of the *principal components* described in section 6.6.1. The only parameter allowed to float in the fit was α_{WZ}. A fit in the highest ranked *principal component* of the 10 observable PCA$_{WZ}$ has also been performed as well as a fit in the original observable H_T^{lep}. The results are shown in table 7.2 and the distributions for the fit in *principal components* is shown in figure 7.2. It should be noted that the 6-*principal components* results are only correct under the assumption that the correlations are negligible.

Table 7.2: $\sigma_{WZ}(pp \to lll + X)$

n PCs	$\alpha_{obs} = (\frac{\sigma_{obs}}{\sigma_{SM}})$	σ_{obs} [pb]
Fit in *principal components* from 6 observable PCA$_{WZ}$		
1	$0.92(7) \pm 0.03(8)$	9.5 ± 0.4 (stat.) ± 1.4 (syst.)
2	$0.95(7) \pm 0.03(8)$	9.8 ± 0.4 (stat.) ± 1.4 (syst.)
3	$0.97(8) \pm 0.03(7)$	10.0 ± 0.4 (stat.) ± 1.4 (syst.)
4	$0.99(6) \pm 0.03(6)$	10.2 ± 0.4 (stat.) ± 1.4 (syst.)
5	$0.99(8) \pm 0.03(5)$	10.2 ± 0.4 (stat.) ± 1.4 (syst.)
6	$1.00(5) \pm 0.03(5)$	10.3 ± 0.4 (stat.) ± 1.4 (syst.)
Fit in p_1 of the 10 observable PCA$_{WZ}$.		
1	0.987 ± 0.038	10.1 ± 0.4 (stat.) ± 1.4 (syst.)
Fit in original H$_T^{lep}$		
-	1.014 ± 0.040	10.4 ± 0.4 (stat.) ± 1.4 (syst.)

    The WZ theoretical NLO cross-section has been calculated for WZ using MCFM with PDF MSTW2008NLO to be 10.2 pb for the $l\nu ll$ final state [84]. The reference is for ATLAS members only but MCFM is publicly available. Both the fits in pattern space and feature space agree with the theoretical prediction. From table 7.1 it is evident that the statistical uncertainty can be reduced with the addition of *principal components* while it is dominated by the overall normalisation just as it was the case in the Monte Carlo study of WZ + ZZ alone in section 5.1.

7.3 Goodness-of-fit

The goodness-of-fit is not easily determined in multiple dimensions as the multidimensional bin content will often be too low for a χ^2 test and alternative methods can be extremely time-demanding. The goodness-of-fit in distributions projected to one dimension can, however, give an indication of the overall agreement, as a deviation will signify a disagreement in the multidimensional space. This is not perfect for p_2 and especially poor for p_3 in the 6 observable PCA$_{WZ}$ seen in figure 7.2 as can be inferred from the KS and χ^2 probabilities shown with the plots. The original distributions and the eigenvector of the PCA$_{WZ}$ are included in appendix D table D.2 and figure D.1. These show that the number of tracks from the primary vertex and minΔR between leptons are significant constituents of the second and third *principal component*. The distribution of these in data clearly deviate from the model expectations. The cause of this is either new physics or mis-modelling. In the case of number of tracks this is very likely due to mis-modelling and not new physics. The number of tracks is extremely dependent on pile-up and although some corrections are performed to account for mis-modelling these are known not to be perfect. It should be noted that the deviations are all within the uncertainties with the exception of a few low-statistics bins. The KS-probability is actually lower for p_1 in the fit of both the first two (KS = 0.25) and first three *principal components* (KS = 0.28) than in the fit in p_1 alone (KS = 0.423). The p-value, calculated for each bin of the distributions, are included in appendix D figure D.2. A KS-test has been performed on histograms with the fraction of events in p_1 falling in each bin in p_2. A total of 50 such slices resulted in KS-test values ranging from 0.001 to 0.995 with a mean of 0.30 confirming that the goodness-of-fit is not perfect in p_1 and p_2. The choice of transformation could have been reconsidered to optimise Standard Model modelling. This was not done as the choice of observables illustrates both one of the great strengths and weaknesses of using *principal components* to fit model expectation to data. The high-ranked *principal components* can be, and are in this case, very sensitive to deviations in any of the constituent observables. This is highly desirable in searches for new physics but obviously constitutes a challenge with regard to modelling of input-observables.
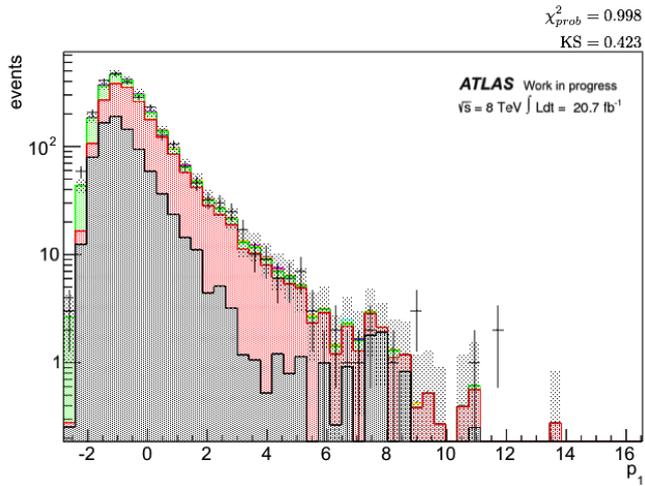
(a) Highest ranked *principal component*.
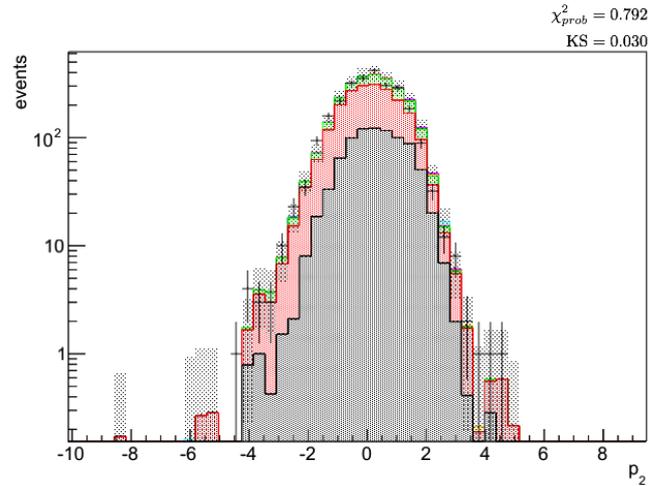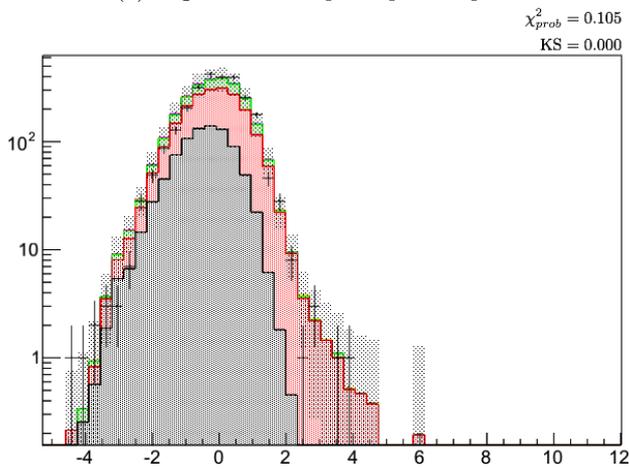
(b) Second *principal component*.

(c) Third *principal component*.
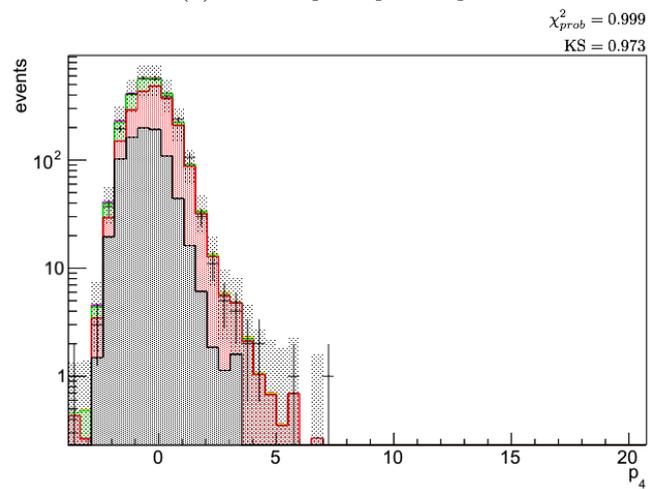
(d) Fourth *principal component*.
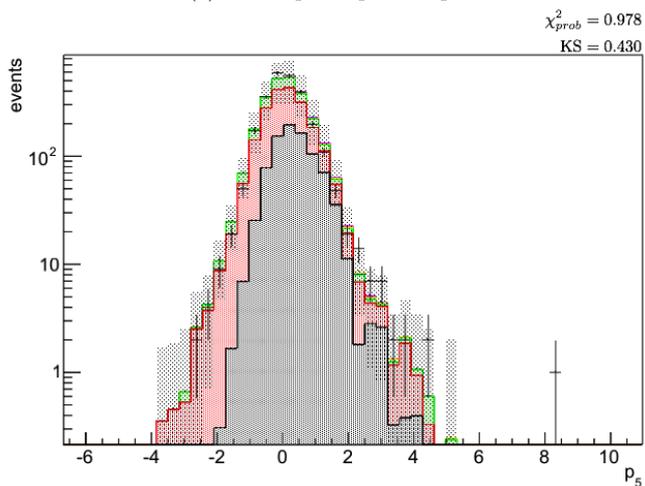
(e) Fifth *principal component*.
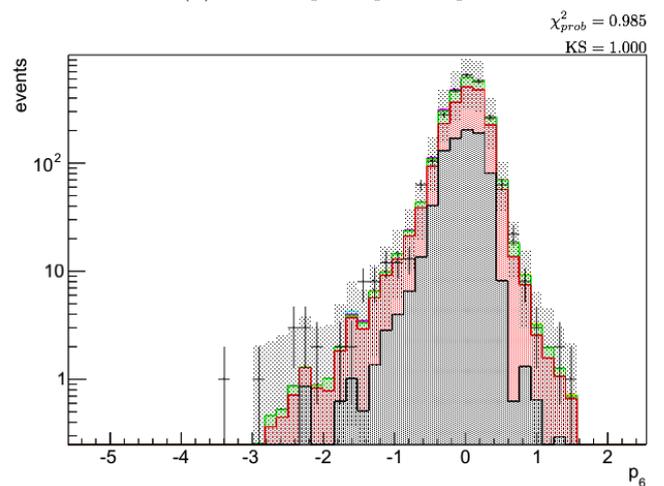
(f) Sixth *principal component*.

Figure 7.2: The distributions resulting from the fit of σ_{WZ} in 6 *principal components*. The KS-probability is quoted for fits performed only in the given *principal component*.

7.4 Type III seesaw mechanism

The final result of the fit-method using Principal Component Analysis is a search for the type III seesaw mechanism described in section 1.4. Seven different masses of the resulting heavy lepton are tested. For each hypothesised mass-point the distribution of observed and expected 95% upper limits on the signal strength are found through the construction of the likelihood ratio as described in section 4.6. For the expected limits 1000 pseudo-experiments were generated in each mass point and the power constrained limits (PCL) [85,86] were found. The test statistic was found to be χ^2 distributed for one μ and assumed to be for all other.

Figure 7.3 show two things: The resulting observed upper limit on the signal strength of the model translated into a cross-section and the median expected upper limit on the signal strength with 1σ and 2σ confidence bands assuming the null-hypothesis (SM). The limits have been set in both the p_1 resulting from 10 input observables and the sum of lepton momenta, H_T^{lep}. H_T^{lep} is chosen as it possesses the best linear separation power between the Standard Model and the seesaw signal in most mass-points. The cross-section presented is the fiducial cross-section of the three-lepton final state with the given selection. A zoom-in of the 95 GeV mass point is presented in figure 7.4.

The observed limit is above for a few mass-points in H_T^{lep} although within the 2-σ band, while it agrees better with the expected limits in p_1. The expected limits are close to identical for the two. With the poor goodness-of-fit found for especially p_3 in the 6 observable PCA$_{WZ}$ an observed limit is not very trustworthy. The expected limits, however, has been calculated and it is shown along with the other two in figure 7.5. While p$_1$ and the H_T^{leps} lie close, the combined fit of p_1, p_2 and p_3 results in a much lower expected limit for all mass points. This clearly illustrates the potential of combining *principal components* in a multidimensional fit. It is especially clear for the mass points from 100 to 200 GeV where the difference exceeds 200 fb. The observed limit for the fit combining 3 *principal components* with poor Standard Model goodness-of-fit has been included in figure D.4 in appendix D.3.

7.5 Discussion

The results presented in this section has illustrated the potential of the method for fitting in a much larger phase-space resulting in much more stringent tests of the Standard Model. When searching for new physics with very low cross-section, very tight cuts in phase-space are needed resulting in regions with few events. In these final-states a method that reduces statistical uncertainty is essential and the use of Principal Component Analysis can fulfil this, especially when the statistical uncertainty of the overall normalisation is not dominant.

It is clear that the approach is highly sensitive to any deviations from the Standard Model. In the light of this the approach, however, also comes with a word of warning. The increased sensitivity will catch any deviations and the higher precision must come with a demand for even more precise understanding and modelling of the known physics.
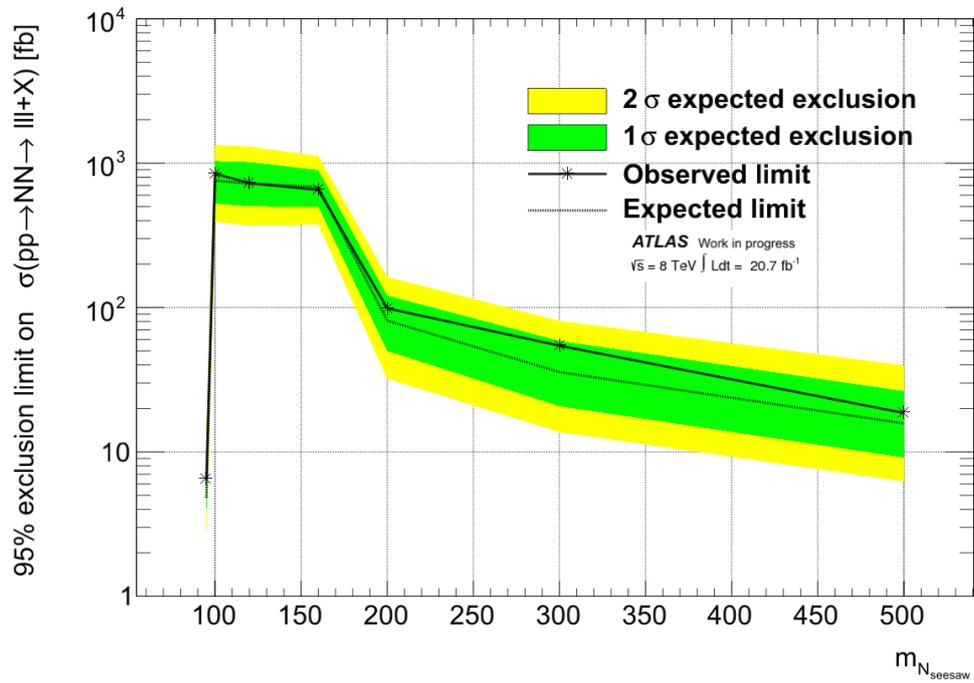
(a) The highest ranked *principal component* generated using 10 input observables.
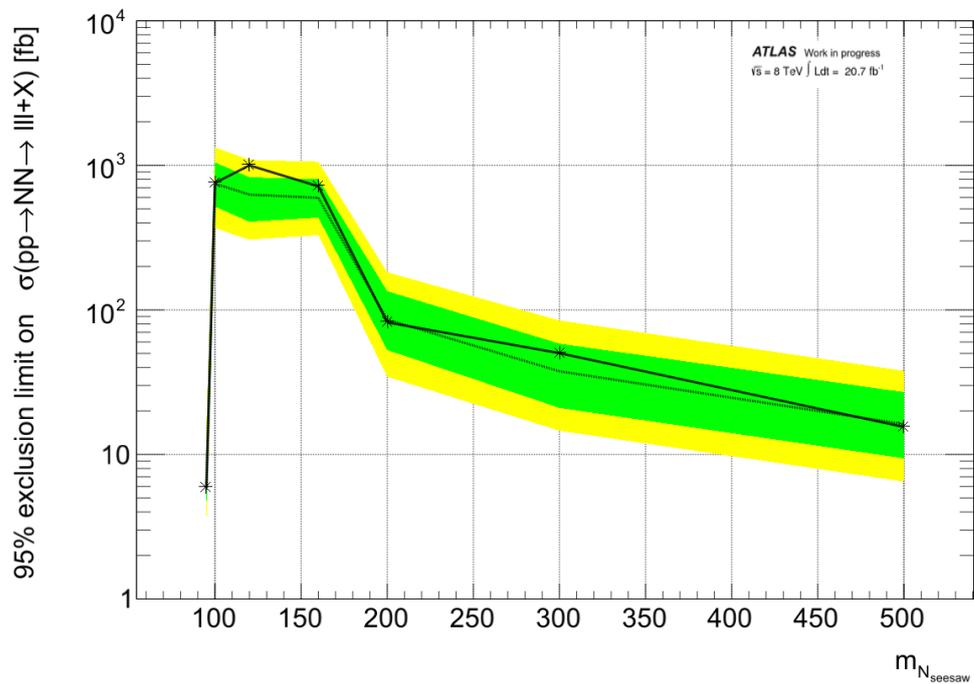
(b) Fitted in the sum of lepton momenta observable of the original pattern space.

Figure 7.3: The 95% confidence intervals set on the cross-section of the excited seesaw lepton at various masses.
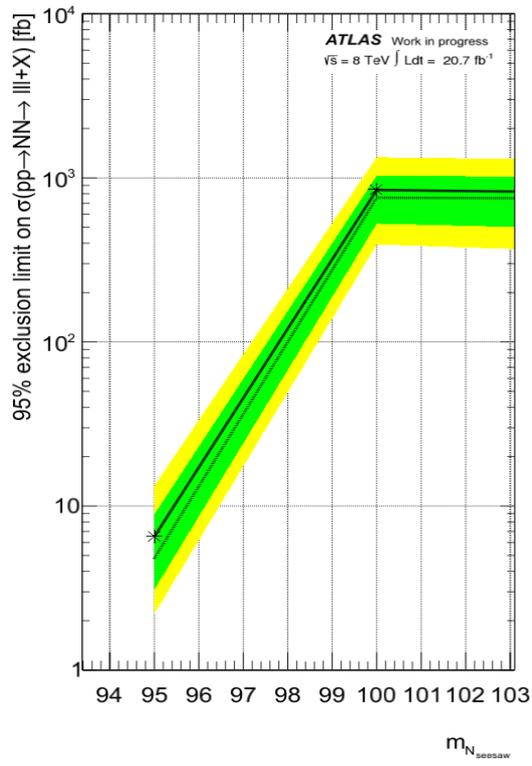
(a) p_1 from 10 observables.
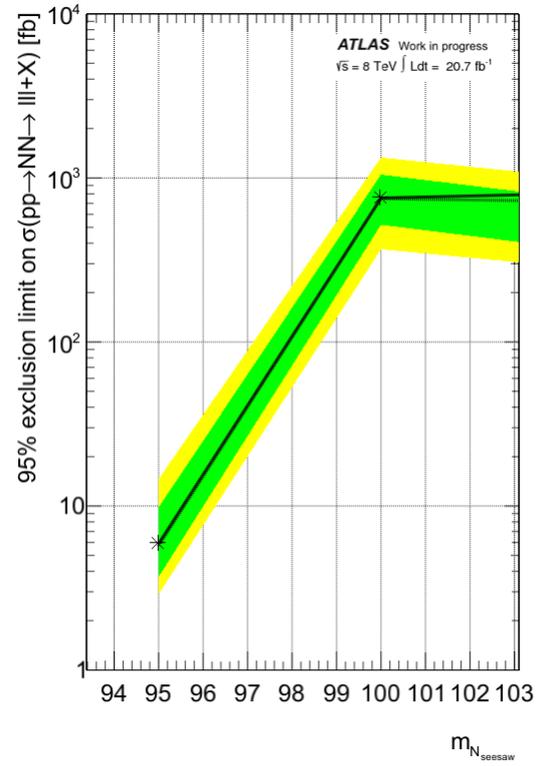
(b) H_T^{lep}.

Figure 7.4: A zoom-in on the 95 GeV mass point limits.
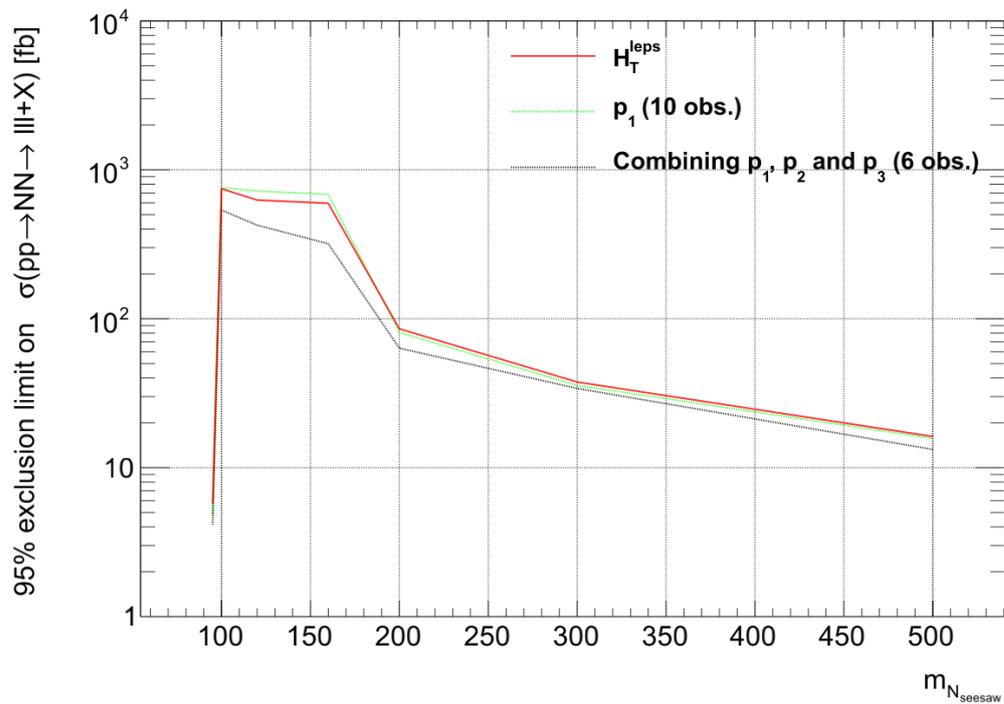
(a) The independent p_1, p_2 and p_3 from the 6 input observables.
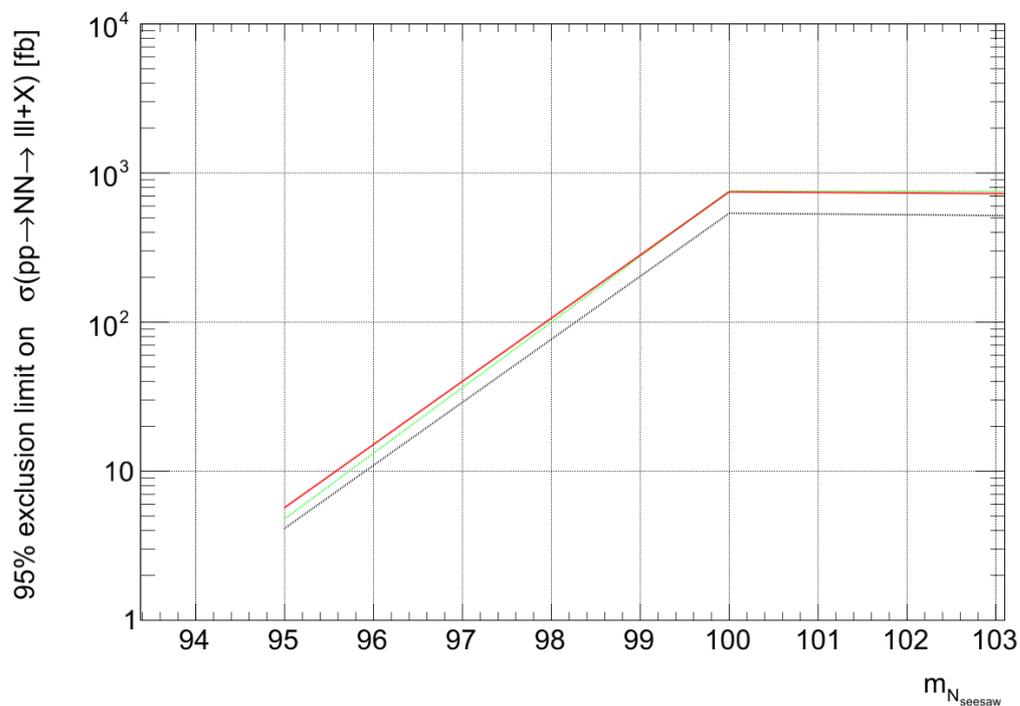
(b) A zoom of the 95 GeV region.

Figure 7.5: Comparing the expected limits of H_T^{leps} and p_1 from 10 observables to the expected limits from combining the three highest ranked *principal components* from the 6 observable matrix minimising correlation in subprocesses. The combination of *principal components* is seen to hold significant potential for setting tighter limits on new physics.

8

Summary and outlook

The work performed in this PhD. thesis has led to a novel fit procedure based on Principal Component Analysis. By diagonalising the covariance matrix of the original list of observables, linearly uncorrelated *principal components* are constructed. The Principal Component Analysis transforms this list of original observables into a new set of variables - the *principal components*. This has several interesting features:

- It has been shown that independent *principal components* can be found, such that several can be combined simultaneously in a likelihood fit.

- The use of the highest order *principal component* - the eigenvector with the largest variance in the original pattern space - can be used as a multidimensional observable. It combines information from several of the observables from the original pattern space.

- Both approaches can potentially reduce the total statistical error on the fit.

- The *principal components* have been shown to have the potential of separating Standard Model processes individually and in general finding deviations from expectation which could be new physics.

A measurement of the Standard Model expectations in the signal region resulted in 2563 observed events and a total of 2569 ± 108 (total) expected events. No significant deviations were found. Following this a measurement of the fiducial cross-section of WZ was performed. Using 6 *principal components* in the fit the result was:

$$\sigma_{WZ}(pp \to lll + X) = 10.3 \pm 0.4(stat.) \pm 1.4(syst.) \text{ pb,} \tag{8.1}$$

in agreement with the theoretical prediction of 10.2 pb. The statistical uncertainty is 3.4%. The current latest WZ cross-section measurement is from [87]. The result is for the three-lepton final state but with different phase-space cuts using 13 fb^{-1} 2012 data recorded with ATLAS. The absolute value of the results are not comparable but the fractional uncertainties can be compared. The result is:

$$\sigma_{WZ}(pp \to lll + \nu)_{reference} = 20.3^{0.8}_{-0.7}(stat.)^{+1.2}_{-1.1}(syst.)^{+0.7}_{-0.6}(lumi.) \text{ pb,} \tag{8.2}$$

with a 3.9% statistical uncertainty based on 1094 candidate events and 277 background expectation.

The statistical error was seen in general to be smaller for the highest ranked *principal component* compared to the best discriminating original observable. It was also shown to fall with the inclusion of additional *principal components* but the total uncertainty was dominated by overall normalisation for the scenarios tested in this thesis. As a final result the use of the procedure on a benchmark model for new physics involving right-handed massive neutrinos was performed. The observed and expected upper limit for the cross-section was calculated in 7 mass-points. The expected limits were seen to be almost identical for p_1 generated from 10 input observables and H_T^{lep} effectively giving a much more stringent test of the Standard Model. The use of several *principal components* was seen to significantly decrease the expected limits proving the potential of the approach for not only a more stringent test of the Standard Model but also better limits on new physics for a given amount of data. A slight excess within the 2-σ band was observed in H_T^{lep} but was not found in p_1.

The method as an alternative to cutting on background processes.

It is important to stress that a main motivation for the development of the procedure is the expected ability to discriminate between processes by using several *principal components*. This is fundamentally different from any cut-based separation, as there is no need to remove events, resulting in larger statistics in the signal region.

The final results revealed some deviations from the Standard Model expectations although none where significant. The use of the much larger phase-space, that is the result of fitting in the highest ranked or combining several *principal components*, will constitute a much more stringent test of the Standard Model.

The near future

By 2015 upgrade work on the LHC will have concluded and ATLAS will be able to record collisions at even higher energy and integrated luminosity. This will be an excellent opportunity to apply the fit procedure. With the ability to combine several *principal components* the procedure will be able to test Standard Model expectations to great detail. For dedicated searches it might hold potential for separation of the new physics and could be used simply to generate optimal observables. For general searches a much larger phase-space can be probed enhancing the chances of finding deviations without prior knowledge (or bias) of their characteristics. Time will tell what we find but there is no one in the physics community, who believe we have seen everything there is to see.

List of Figures

133

List of Tables

A.1   Pull distributions for the 10-observable input set

(a) Pull distribution for PC 1
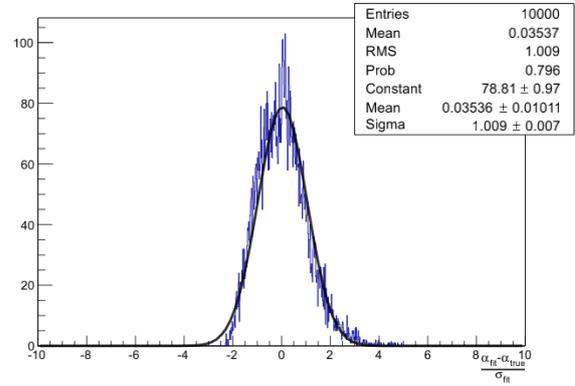
(b) Pull distribution for PC 2
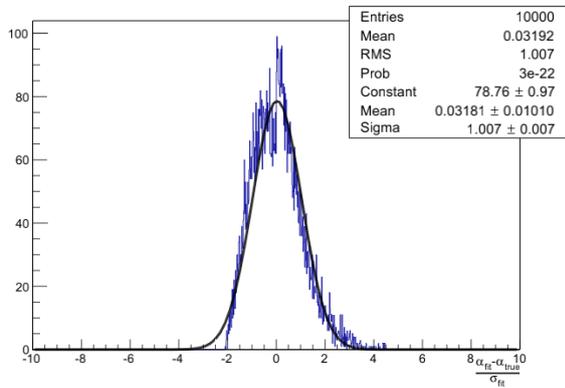
(c) Pull distribution for PC 3
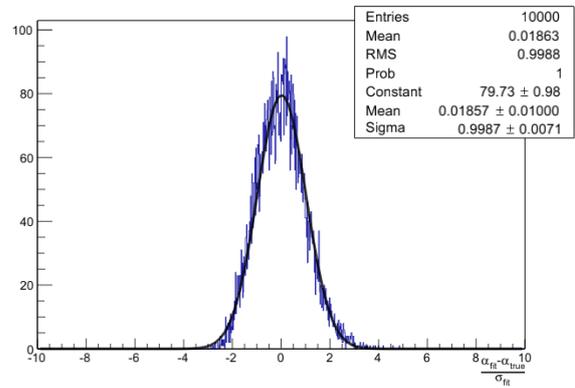
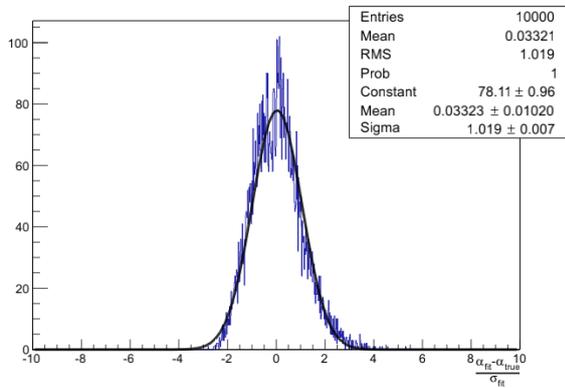(d) Pull distribution for PC 4
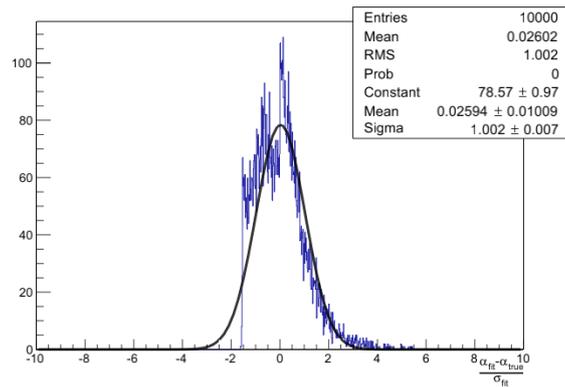
(e) Pull distribution for PC 5

Figure A.1: The pull distribution based on 10000 pseudo-samples for the first 5 *principal components* of the WZ sample constructed from the 10 input observables.
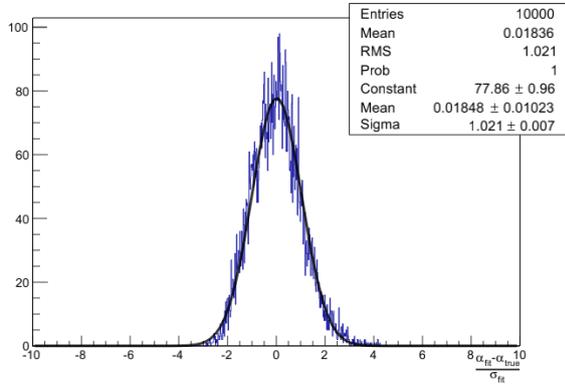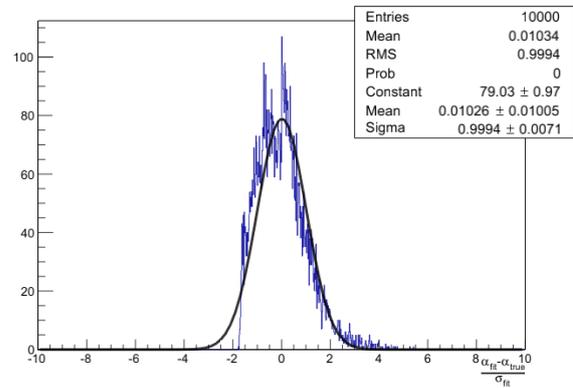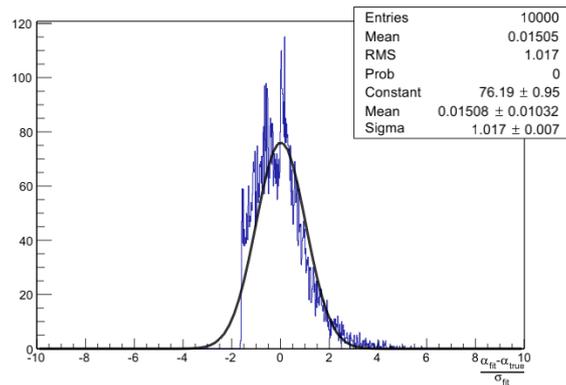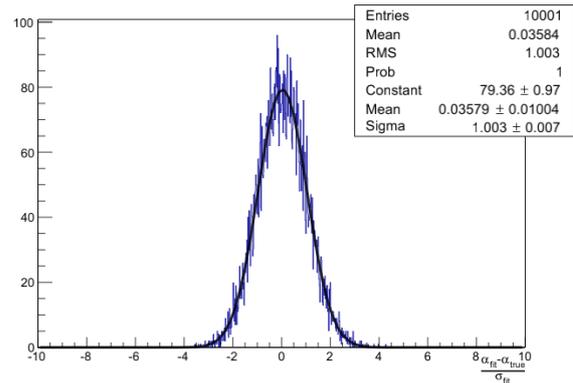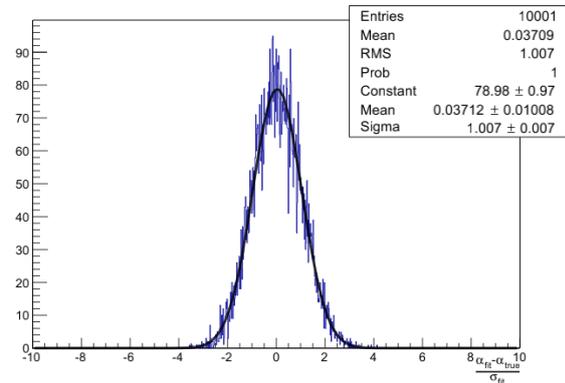
(a) Pull distribution for PC 6
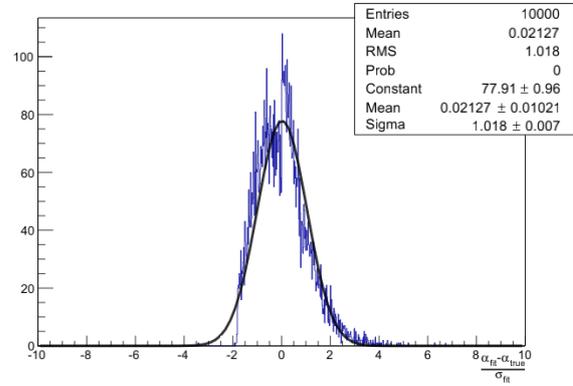
(b) Pull distribution for PC 7

(c) Pull distribution for PC 8

(d) Pull distribution for PC 9

(e) Pull distribution for PC 10

Figure A.2: The pull distributions for the last 5 *principal components* of the WZ sample constructed from the 10 input observables.
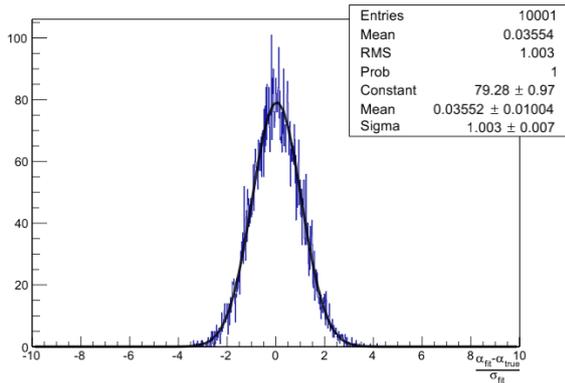
(a) Pull distribution for PC 5 including overall normalisation.

(b) Pull distribution for PC 7 including overall normalisation.

(c) Pull distribution for PC 9 including overall normalisation.
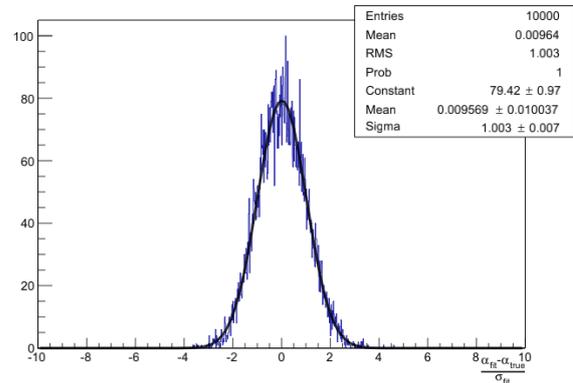
(d) Pull distribution for PC 10 including overall normalisation.

Figure A.3: The pull distributions for p_5, p_7, p_9 and p_{10} when normalisation is included in the fit along with shape information. The PCA is constructed from WZ using 10 input observables. The fit is to WZ while ZZ is kept fixed. The agreement with unit Gaussians is good.

B.1 SFSS eigenvector and eigenvalues for PCA$_{V+jets}$

PCA$_{V+jets}$ eigenvector matrix :

$$
\begin{array}{ccccccc}
v_{e,1} & v_{e,2} & v_{e,3} & v_{e,4} & v_{e,5} & v_{e,6} & v_{e,7}
\end{array}
\begin{bmatrix}
0.512 & 0.030 & 0.027 & -0.240 & 0.092 & -0.146 & -0.805 \\
0.183 & -0.589 & -0.233 & 0.734 & 0.054 & 0.054 & -0.136 \\
0.503 & 0.017 & 0.044 & -0.115 & -0.093 & 0.826 & 0.195 \\
0.084 & 0.561 & -0.804 & 0.169 & 0.044 & 0.009 & 0.000 \\
-0.454 & 0.006 & -0.001 & -0.019 & 0.734 & 0.422 & -0.276 \\
0.149 & 0.571 & 0.543 & 0.583 & 0.115 & -0.032 & -0.019 \\
0.463 & -0.098 & -0.017 & -0.137 & 0.652 & -0.335 & 0.467
\end{bmatrix}
\tag{B.1}
$$

for the input observables:

$$
\begin{bmatrix} p_{T,lead}^{lep} & p_{T,second}^{lep} & H_T^{jets} & \not{E}_T & \Delta R & nTracks_{vertex} & M_T^{3lep} \end{bmatrix}
$$

C

Object selection

C.1 Cutflow histograms for object selection

The selection of prompt electroweak electrons and muons has been optimised by the multi-lepton group at CERN for the 2012 8TeV ATLAS data. The efficiency of the various cuts is expected to vary for different Standard Model processes and the outflow histograms for the object selectors are included in this appendix for reference to other studies. Besides the electron and muon objects, jets have to be identified to handle overlap removal and for use in input observables for the final fitting.
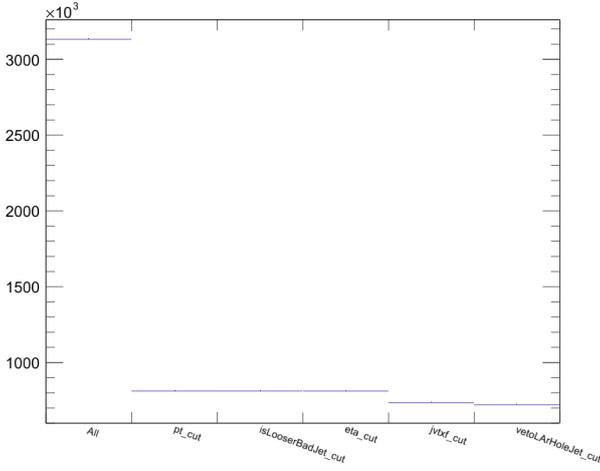
(a) The cutflow for the jet-object selector for WZ
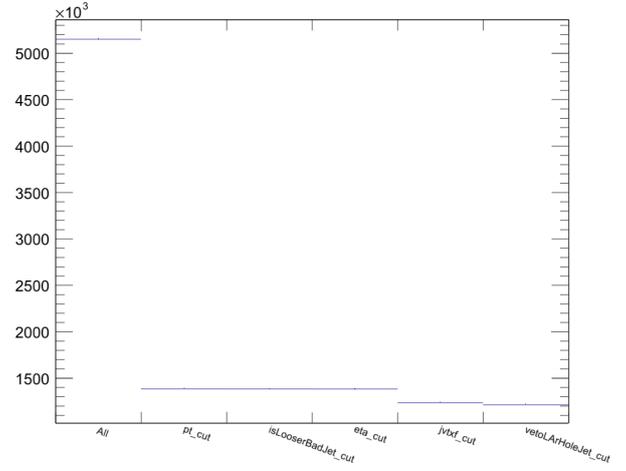
(b) The cutflow for the jet-object selector for ZZ
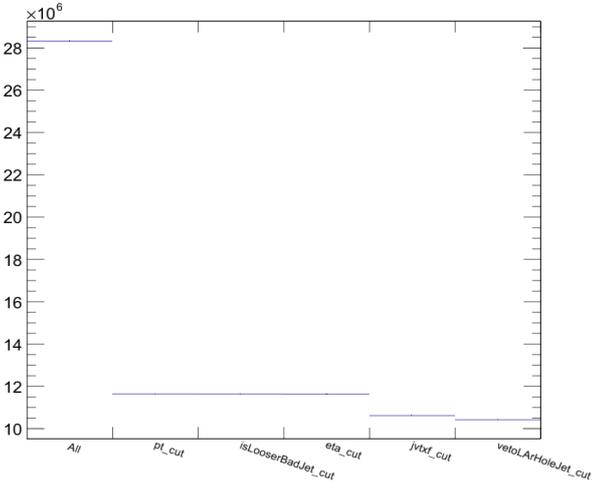
(c) The cutflow for the jet-object selector for $Z \to ee + \text{jets}$
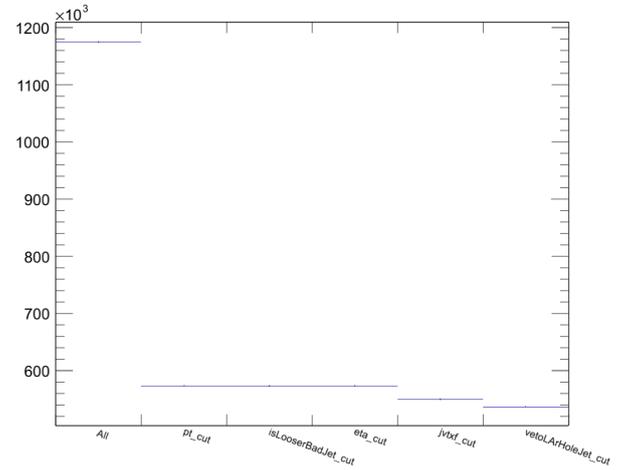
(d) The cutflow for the jet-object selector for $t\bar{t} + \text{jets}$
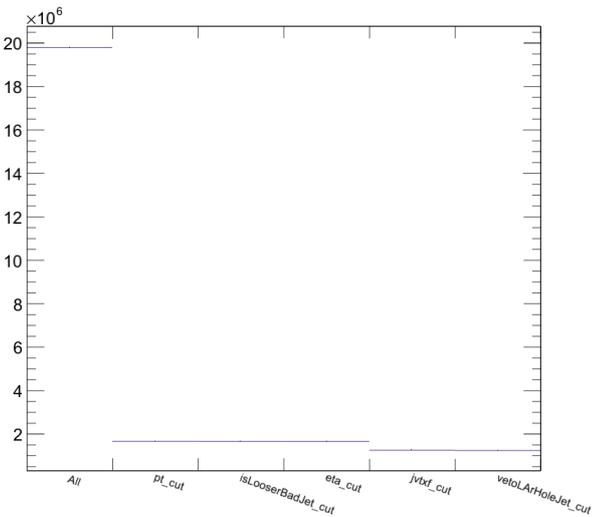
(e) The cutflow for the jet-object selector for $Z \to \mu\mu + \text{jets}$
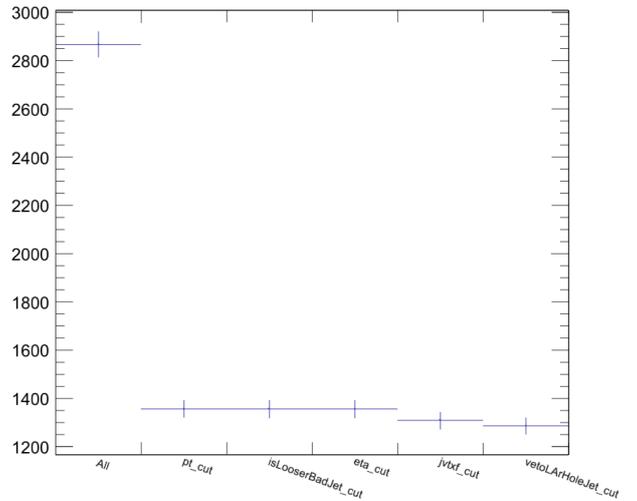
(f) The cutflow for the jet-object selector for $Z \to ll + \gamma$

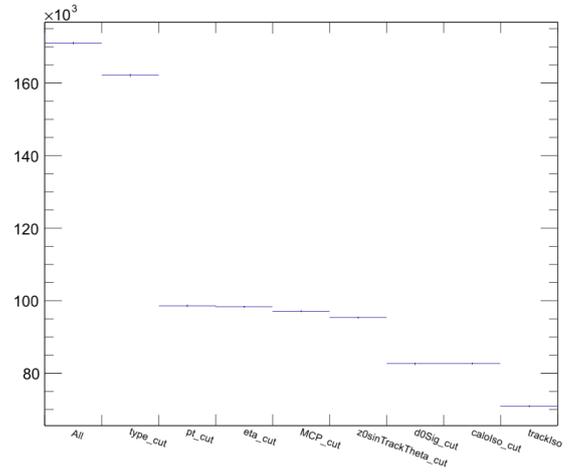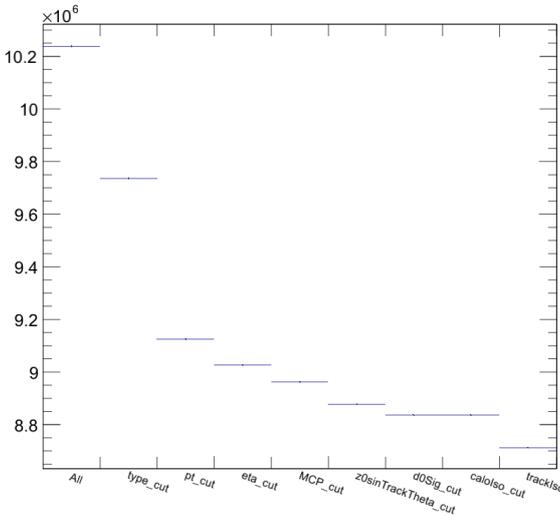Figure C.1: The jet object selector cutflow diagrams for various processes.
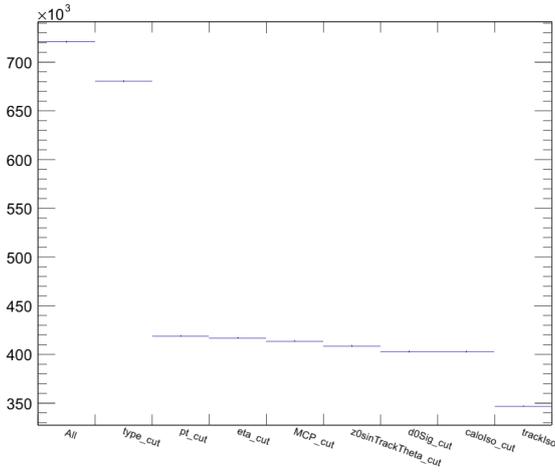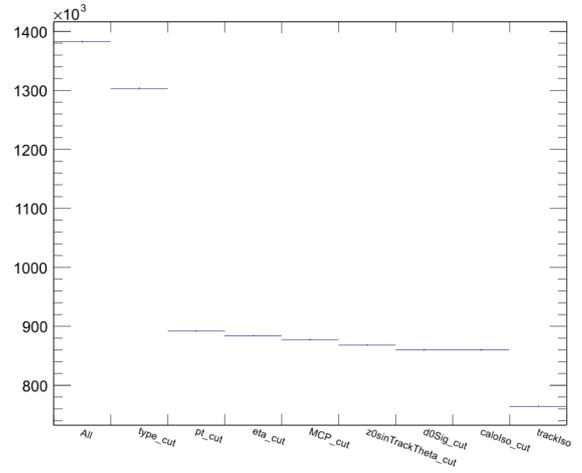
(a) The cutflow for the muon-object selector for WZ

(b) The cutflow for the muon-object selector for ZZ
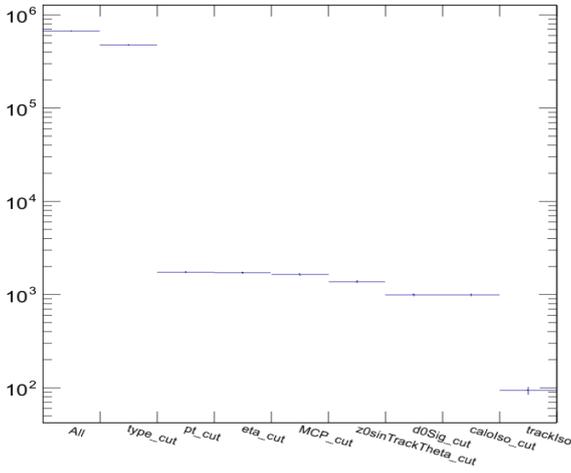
(c) The cutflow for the muon-object selector for $Z \to ee +$ jets. Log-scale is used here due very low efficiency on the p_T cut.

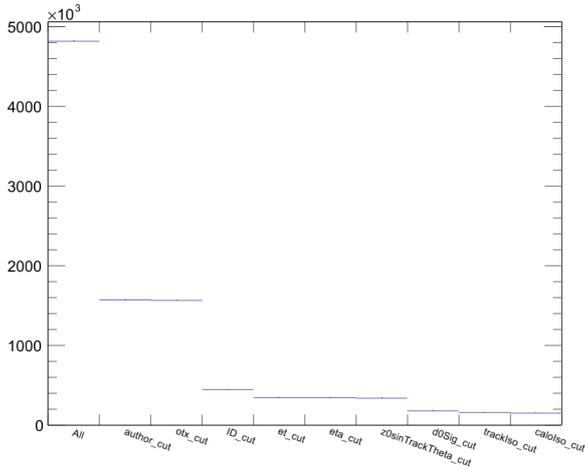(d) The cutflow for the muon-object selector for $t\bar{t} +$ jets

(e) The cutflow for the muon-object selector for $Z \to \mu\mu +$ jets
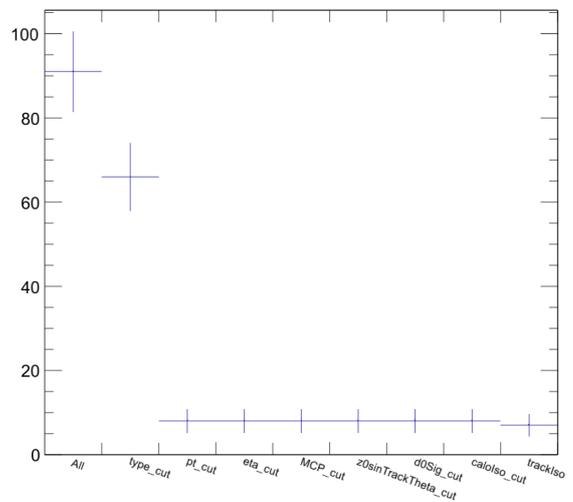
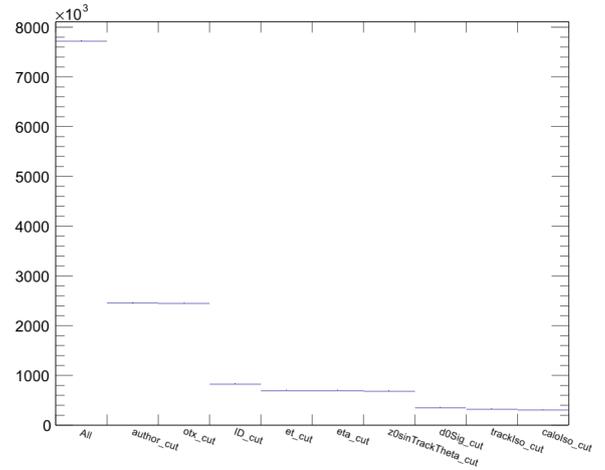(f) The cutflow for the muon-object selector for $Z \to ll + \gamma$

Figure C.2: The muon object selector cutlow diagrams for various processes.

(a) The cutflow for the electron-object selector for WZ
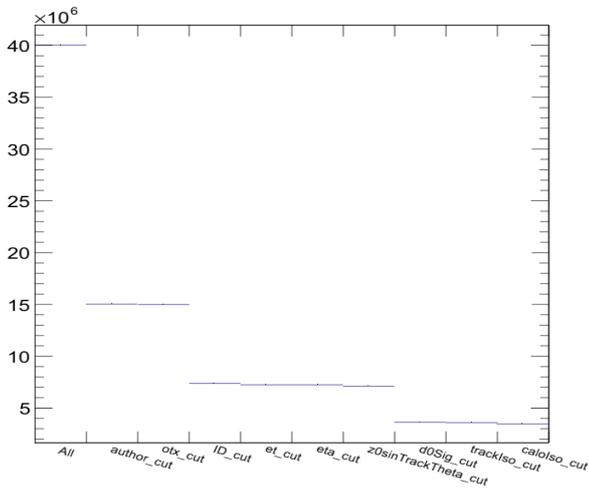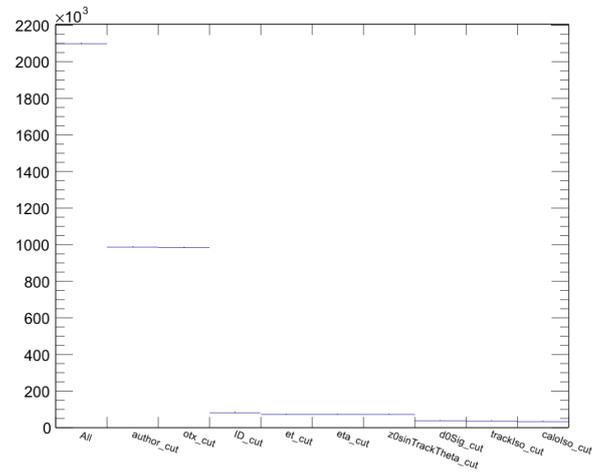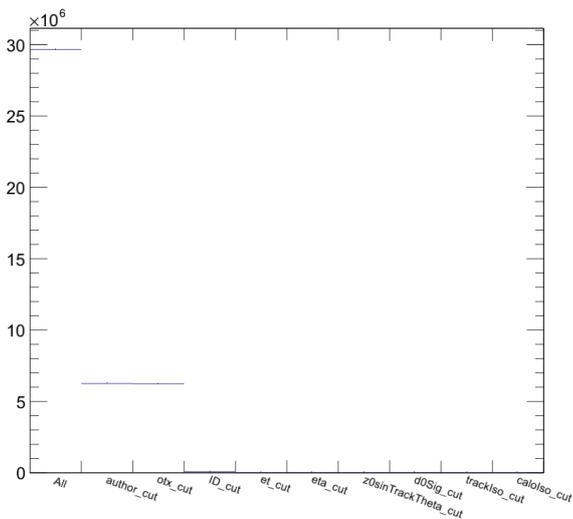
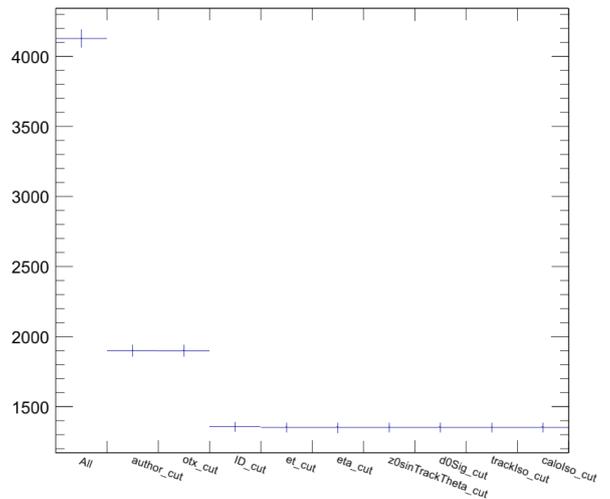(b) The cutflow for the electron-object selector for ZZ

(c) The cutflow for the electron-object selector for $Z \to ee + \text{jets}$

(d) The cutflow for the electron-object selector for $t\bar{t} + \text{jets}$

(e) The cutflow for the electron-object selector for $Z \to \mu\mu + \text{jets}$

(f) The cutflow for the electron-object selector for $Z \to ll + \gamma$

Figure C.3: The electron object selector cutlow diagrams for various processes.

D.1   6 obs PCA$_{WZ}$ for fitting mulitple *principal components*

PCA$_{WZ}$ eigenvector matrix :

$$
\begin{array}{cccccc}
v_{e,1} & v_{e,2} & v_{e,3} & v_{e,4} & v_{e,5} & v_{e,6}
\end{array} \tag{D.1}
$$
$$
\begin{bmatrix}
0.59 & -0.23 & 0.15 & -0.15 & -0.20 & 0.72 \\
0.45 & 0.20 & -0.22 & -0.13 & 0.83 & -0.07 \\
0.27 & 0.34 & 0.42 & 0.80 & 0.01 & -0.03 \\
-0.20 & -0.66 & -0.35 & 0.54 & 0.27 & 0.21 \\
0.24 & 0.34 & -0.79 & 0.19 & -0.40 & 0.01 \\
0.53 & -0.49 & 0.05 & -0.01 & -0.21 & -0.66
\end{bmatrix}
$$

for the input observables:

$$
\begin{bmatrix} H_T^{lep} & H_T^{jets} & \not{E}_T & \Delta R & nTracks_{vertex} & M_T^{3lep} \end{bmatrix}
$$
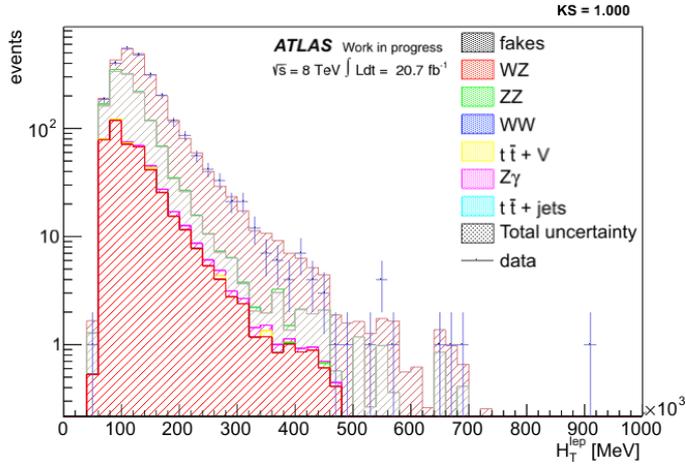
(a) Plot of H_T^{lep} before fitting MC to data.
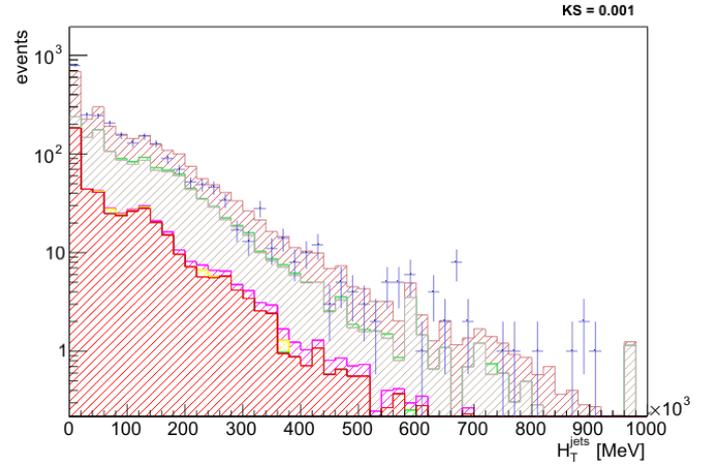
(b) Plot of H_T^{jets} before fitting MC to data.
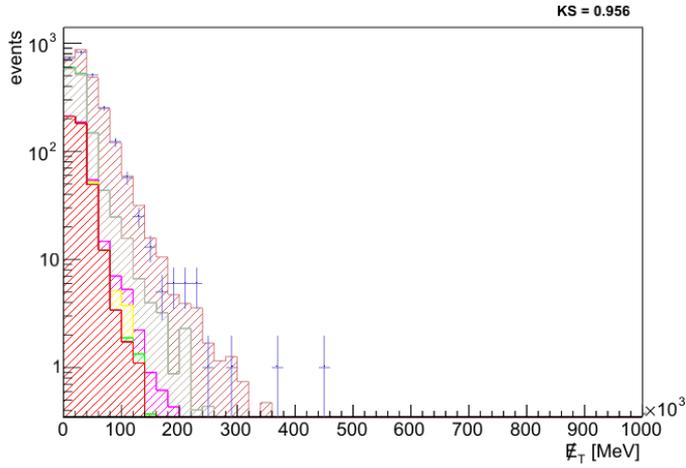
(c) Plot of $\not{E}_T$ before fitting MC to data.
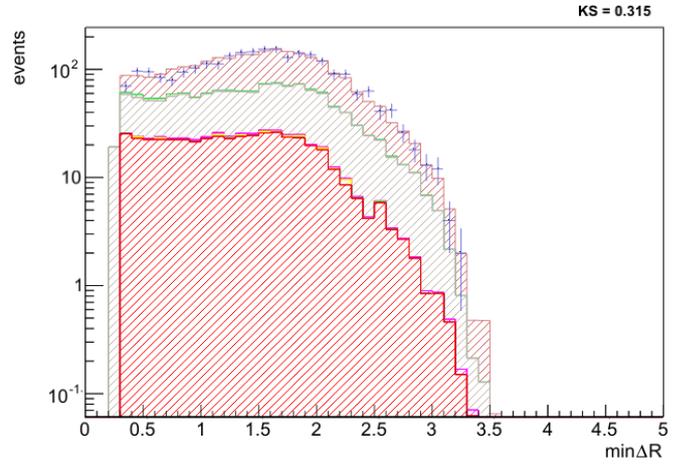
(d) lot of $\min\Delta R$ before fitting MC to data.
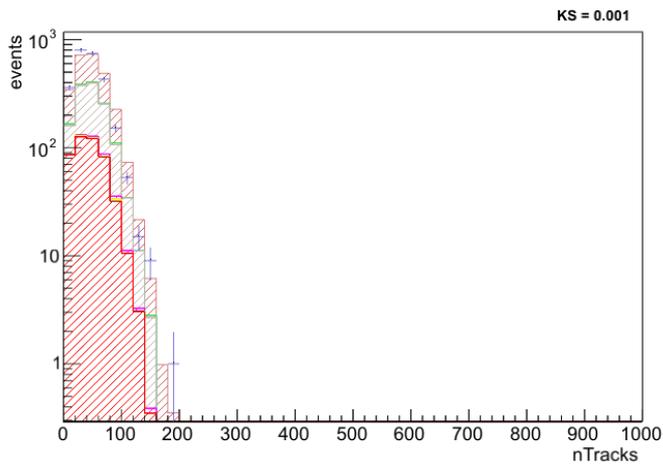
(e) Plot of nTracks$_{vertex}$ before fitting MC to data.
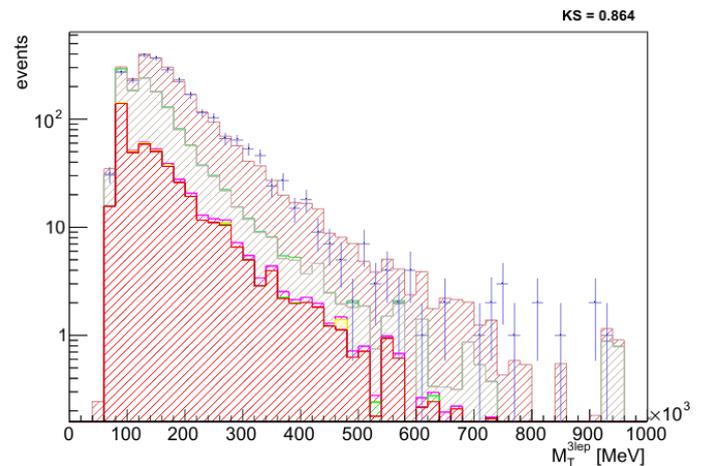
(f) Plot of M_T^{3lep} before fitting MC to data.

Figure D.1: The observables used as input for the 6 observable PCA$_{WZ}$. The distributions for Standard Model expectations are the initial distributions before fitting to real data.
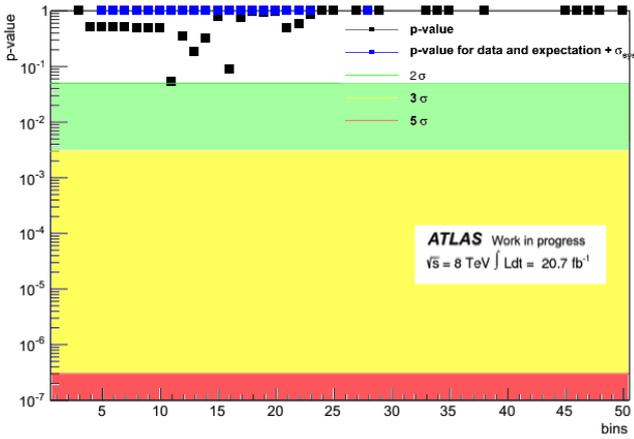
(a) Highest ranked *principal component*.
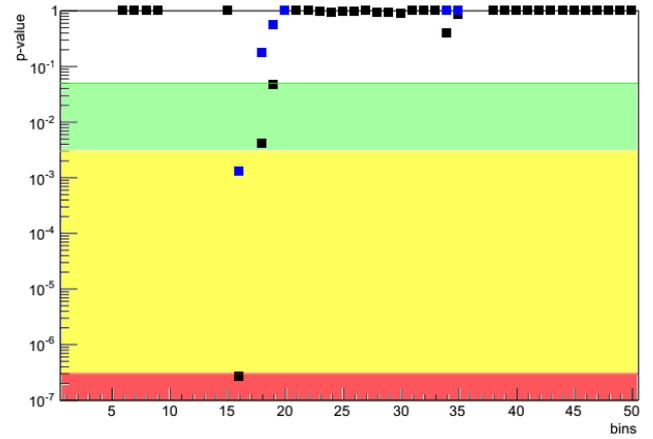
(b) Second *principal component*. The large deviation is in a low-statistics bin.
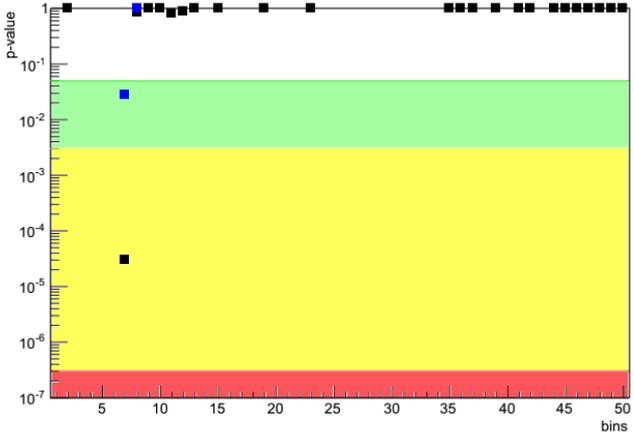
(c) Third *principal component*.
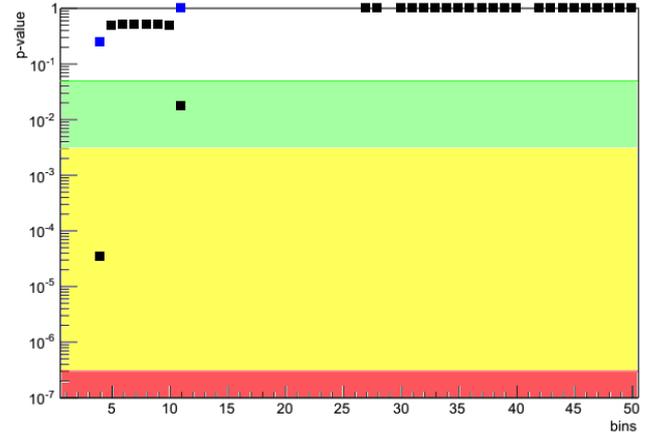
(d) Fourth *principal component*.
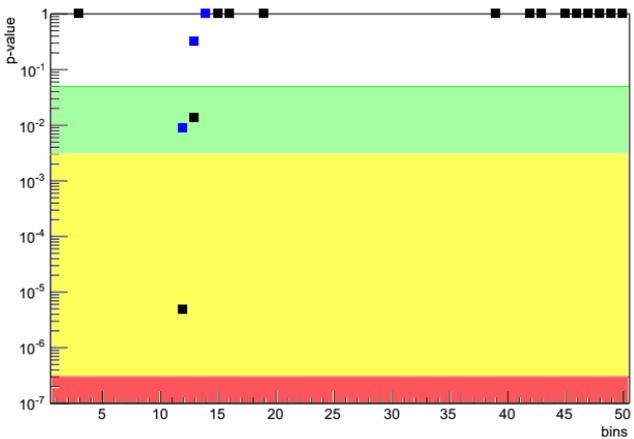
(e) Fifth *principal component*.
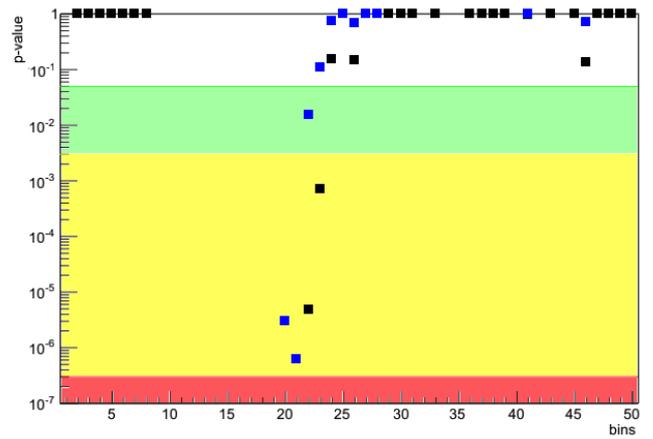
(f) Sixth *principal component*. The two large deviations are in low-statistics bins.

Figure D.2: The p-value calculated per bin in each of the *principal components*. Several deviations are present although none are above 5 σ after accounting for systematic uncertainties as indicated by the blue dots as opposed to the black indicating the deviation without accounting for the systematic uncertainty.

Figure D.3: The profile likelihood ratio for the expected μ for WZ using p_1 and p_2 combined in the fit, generated using 10.000 pseudo-experiments.

D.2   10 observable PCA$_{WZ}$ for fitting in highest ranked *principal component*

The eigenvector matrix as also presented in section 5.1.5 is:

$$
\begin{array}{cccccccccc}
v_{e,1} & v_{e,2} & v_{e,3} & v_{e,4} & v_{e,5} & v_{e,6} & v_{e,7} & v_{e,8} & v_{e,9} & v_{e,10}
\end{array} \tag{D.2}
$$

$$
\begin{bmatrix}
0.43 & 0.12 & -0.16 & 0.11 & 0.08 & -0.13 & 0.35 & -0.51 & -0.09 & 0.59 \\
0.39 & 0.21 & -0.05 & 0.04 & 0.05 & -0.35 & -0.48 & 0.57 & 0.02 & 0.35 \\
0.26 & 0.02 & 0.47 & -0.50 & -0.18 & 0.43 & 0.30 & 0.28 & 0.12 & 0.23 \\
0.31 & -0.38 & 0.24 & -0.13 & -0.08 & 0.19 & -0.66 & -0.45 & -0.07 & -0.03 \\
0.11 & -0.49 & 0.29 & 0.24 & 0.74 & -0.02 & 0.18 & 0.18 & 0.02 & 0.00 \\
0.19 & -0.19 & -0.58 & 0.26 & -0.05 & 0.67 & -0.02 & 0.24 & -0.11 & 0.05 \\
0.14 & 0.42 & 0.43 & 0.63 & -0.01 & 0.34 & -0.15 & -0.08 & 0.22 & 0.15 \\
0.13 & -0.48 & 0.18 & 0.43 & -0.63 & -0.28 & 0.22 & 0.14 & 0.03 & -0.01 \\
0.47 & 0.12 & -0.15 & 0.02 & 0.05 & -0.04 & 0.09 & -0.09 & 0.72 & -0.45 \\
0.43 & 0.32 & 0.19 & 0.13 & 0.01 & 0.01 & 0.12 & 0.04 & -0.62 & -0.50
\end{bmatrix}
$$

for the input observables:

$$
\begin{bmatrix}
p^{lep}_{T,lead} & p^{lep}_{T,second} & p^{lep}_{T,third} & H^{jet}_{T} & MV1_{max} & \not{E}_{T} & \Delta R & nTracks_{vertex} & M^{SFOS}_{inv} & M^{3lep}_{T}
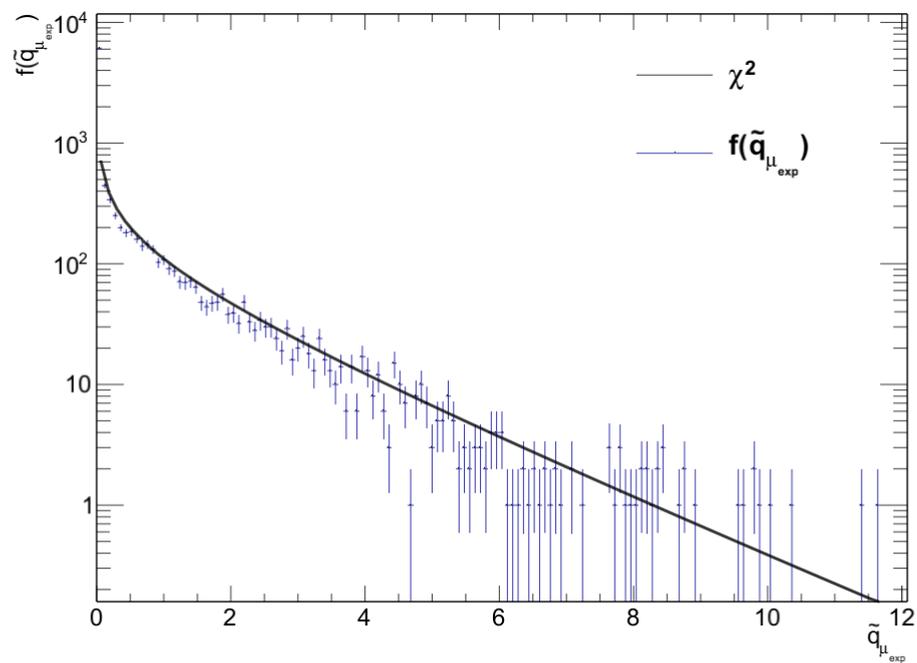\end{bmatrix}
$$

Table D.1: Correlation between parameters in fit.

process	GLOBAL	fakes	WZ	ZZ	WW	$t\bar{t} + $ V	Zγ	$t\bar{t} + $ jets	lumi
fakes	0.21	1.00	-0.05	-0.037	0.00	0.00	-0.01	0.00	-0.11
WZ	0.69	-0.05	1.00	-0.04	0.00	-0.05	0.00	0.00	-0.65
ZZ	0.32	-0.04	-0.04	1.00	0.00	0.00	-0.03	0.00	-0.20
WW	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
$t\bar{t} + $ V	0.07	0.00	-0.05	0.00	0.00	1.00	0.00	0.00	0.00
Zγ	0.10	-0.01	0.00	-0.03	0.00	0.00	1.00	0.00	-0.06
$t\bar{t} + $ jets	0.014	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-0.01
Luminosity	0.70	-0.11	-0.65	-0.20	0.00	0.00	-0.06	-0.01	1.00

D.3   Observed limits on seesaw benchmark model combining 3 *principal components*

The observed and expected upper limits (PCL) on the seesaw signal, based on the the likelihood ratio combining 3 *principal components*. The goodness-of-fit to the Standard Model was shown to be poor for p_3 while better for p_2. The downwards fluctuations in mass-points 100-160 GeV is due to a downwards fluctuation in data compared to Standard Model expectations in two bins central to the mass peak of all three seesaw mass-points in p_1. This originates from two bins in H^{leps}_T and M^{3lep}_T, the main contributors to p_1. This is seen in figure D.5.

Table D.2: Selection efficiency for the seesaw mass points

Mass	Events in signal region	sample size	efficiency
95	1250	100000	1.25 %
100	1324	100000	1.32 %
120	4103	100000	4.10 %
160	5760	100000	5.76 %
200	3159	50000	6.32 %
300	3420	50000	6.84 %
500	2798	39000	7.17 %

(a) The three highest ranked *principal components* generated using 6 input observables minimising correlation between *principal components* in subprocesses. The dotted line indicates the observed limit before imposing the power constraint as recommended by ATLAS [85, 86]
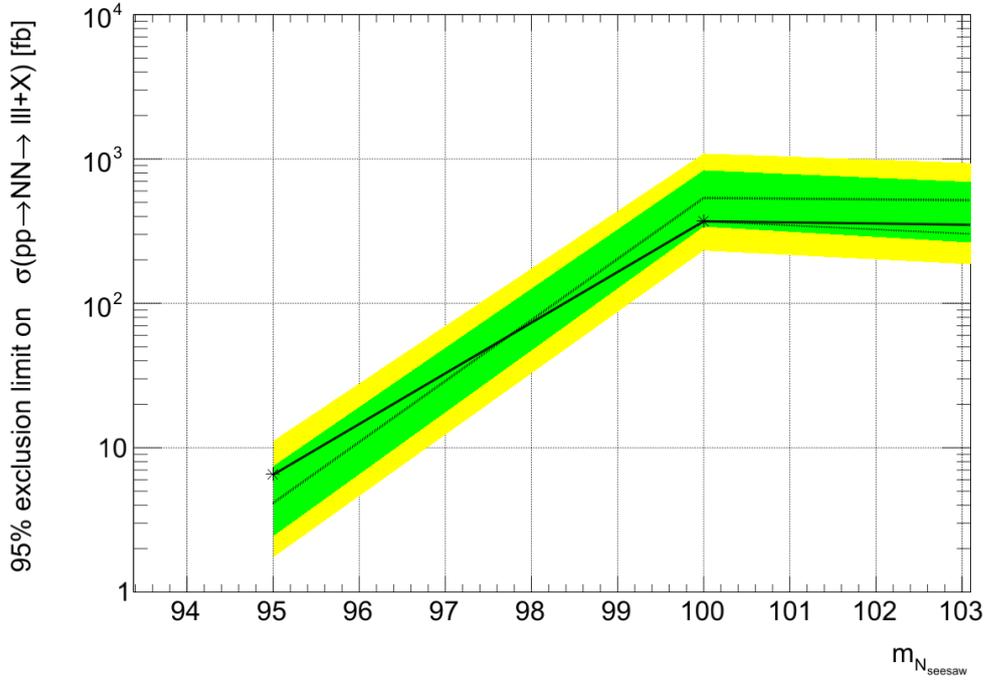
(b) p_1, p_2 and p_3 combined for 95 and 100 GeV seesaw mass points.

Figure D.4: The 95% confidence intervals (PCL) set on the cross-section of the excited seesaw lepton at various masses.

(a) $\sum_{i=0}^{leptons} p_T^i$ with seesaw signal at 100 GeV mass-point.

(b) $\sum_{i=0}^{leptons} p_T^i$ with seesaw signal at 160 GeV mass-point.

Figure D.5: The sum of lepton p_T and seesaw signal overlaid for two of the mass points with downward fluctuations. As seen, data has a downwards fluctuation compared to Standard Model expectation for two of the bins central to the seesaw signal in both distributions. The same is true for the 120 GeV mass-point and for the distributions in transverse mass.

Bibliography

[1] Berge Ask, Eifert Berger, Krasznahorkay Hoecker, and Haller. Sframe analysis tool. URL `http://sourceforge.net/p/sframe/wiki/Home/#developers`.
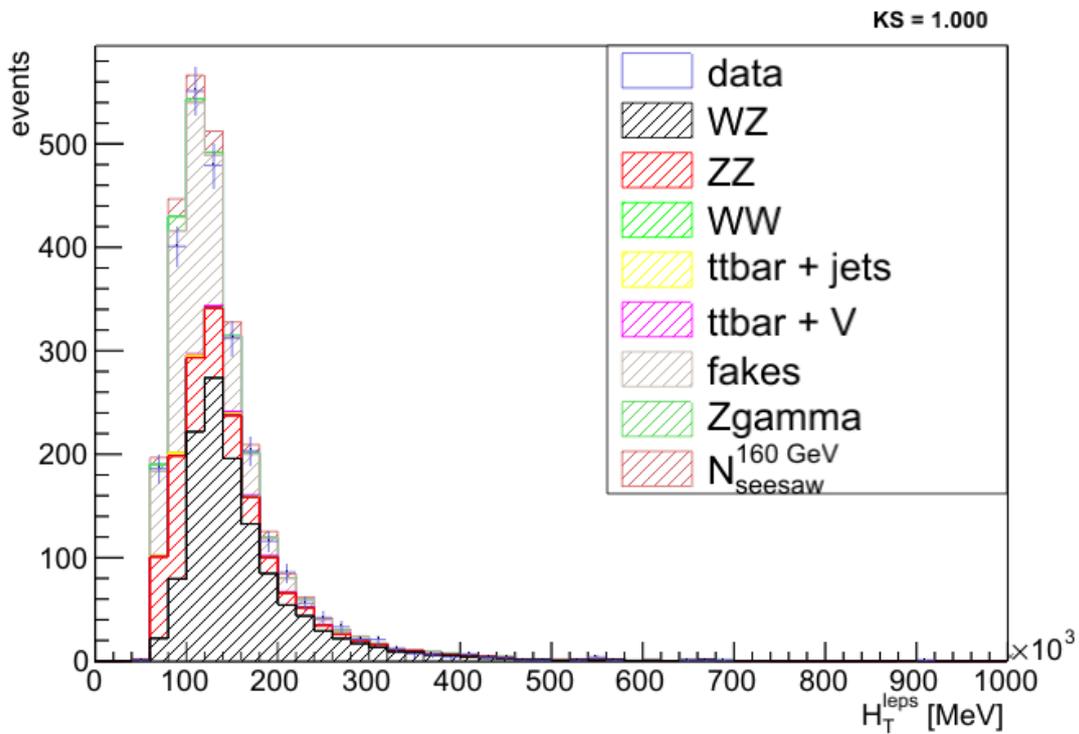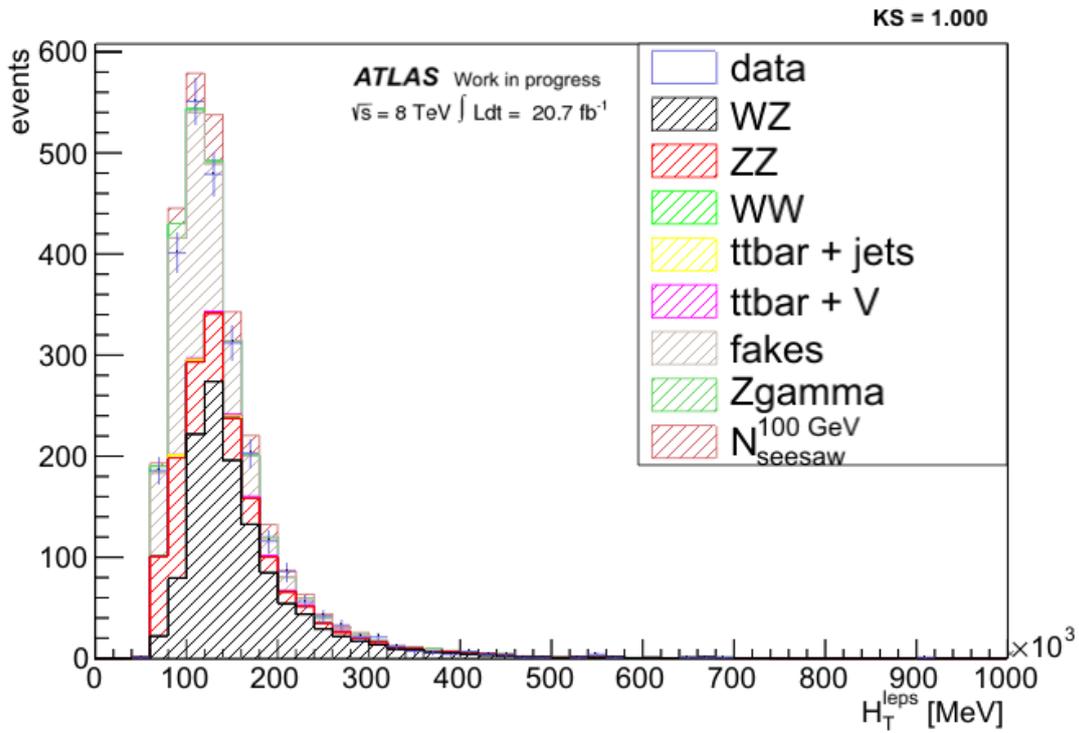
[2] Georges Aad et al. Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector. 2014.

[3] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012.

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995.

[5] A. Zee. *Quantum Field Theory in a nutshell.* Princeton University Press, 2003.

[6] F. Mandl and G. Shaw. *Quantum Field Theory.* John Wiley & Sons, Ltd, 1984.

[7] B.R.Martin and G. Shaw. *Particle Physics.* John Wiley and Sons, 2007. 22 - 28 pp.

[8] Shlomo Sternberg. Lie algebras, 1999. preprint November.

[9] Iris Rottländer. *Development of a Benchmark Parameter Scan for Higgs Bosons in the NMSSM model and a Study of the Sensitivity for $H \to AA \to 4\tau$ in Vector Boson Fusion with the ATLAS Detector.* PhD thesis, Bonn University, Physics Institute, 2008.

[10] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13(16):508–509, Oct 1964.

[11] Q.R. Ahmad et al. Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory. *Phys.Rev.Lett.*, 87:071301, 2001.

[12] Esra Bulbul, Maxim Markevitch, Adam Foster, Randall K. Smith, Michael Loewenstein, et al. Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters. *Astrophys.J.*, 789:13, 2014.

[13] P.A.R. Ade et al. Detection of B-Mode Polarization at Degree Angular Scales by BICEP2. *Phys.Rev.Lett.*, 112:241101, 2014.

[14] Giuseppe Degrassi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, et al. Higgs mass and vacuum stability in the Standard Model at NNLO. *JHEP*, 1208: 098, 2012.

[15] Stephen P. Martin. A Supersymmetry primer. *Adv.Ser.Direct.High Energy Phys.*, 21: 1–153, 2010.

[16] Theodor Kaluza. Zum Unitätsproblem in der Physik. *Sitzungsber. Preuss. Akad. Wiss.*, 1921.

[17] Oskar Klein. Quantentheorie und fünfdimensionale Relativitätstheorie. *Zeitschrift für Physik A*, 37, 1926.

[18] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys.Rev.Lett.*, 83:3370–3373, 1999.

[19] Roberto Franceschini, Thomas Hambye, and Alessandro Strumia. Type-III see-saw at LHC. *Phys.Rev.*, D78:033002, 2008.

[20] C. Biggio and F. Bonnet. Implementation of the type III seesaw model in Feyn-Rules/MadGraph and prospects for discovery with early LHC data. *UAB-FT-694*, Feb 2012.

[21] H. L. Lai et al. Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.*, C12:375–392, 2000.

[22] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. MRST partons and uncertainties. *hep-ph/0307262*, 2003.

[23] B. Webber. Parton shower Monte Carlo event generators. *Scholarpedia*, 6(12):10662, 2011.

[24] T Sjostrand. Monte carlo generators and soft qcd, August 2013.

[25] T Sjostrand. Monte carlo generators. oai:cds.cern.ch:999717. *hep-ph/0611247. CERN-LCGAPP-2006-06*, page 23 p, Nov 2006.

[26] A Kupco. Cluster hadronization in herwig 5.9. *hep-ph/9906412*, 1999.

[27] S. Mrenna T. Sjöstrand and P. Skands. Pythia 8.135. *Comput. Phys. Comm. 178*, 852, 2008. [arXiv:0710.3820].

[28] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and B.R. Webber. Herwig 6.5. *010 [hep-ph/0011363]; hep-ph/0210213*, JHEP 0101, 2001.

[29] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009.

[30] J. M. Butterworth, Jeffrey R. Forshaw, and M.H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637–646, 1996.

[31] For the full list of references for sherpa see:, 2014. URL `https://sherpa.hepforge.org/doc/SHERPA-MC-2.1.0.html`. [Online; accessed 7June 2014].

[32] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010.

[33] S. Frixione and B.R. Webber. Matching nlo qcd computations and parton shower simulations. *029 [hep-ph/0204244]*, JHEP 0206, 2002.

[34] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. Polosa. a generator for hard multiparton processes in hadronic collisions. *hep-ph/0206293*, JHEP 0307:001, 2003.

[35] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys.Rev.*, D82:074018, 2010.

[36] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. 2014.

[37] ATLAS collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3, August 2008.

[38] Ulrik Egede. *The search for a standard model Higgs at the LHC and electron identification using transition radiation in the ATLAS tracker.* PhD thesis, Lund University, Faculty of mathematics and natural sciences, 1998.

[39] Elzbieta Richter-Was, Daniel Froidevaux, and Luc Poggioli. ATLFAST 2.0 a fast simulation package for ATLAS atlas note atl-phys-98-131.

[40] J. Allison et al. Geant4 developments and applications. *Nuclear Science, IEEE Transactions on*, 53(1):270–278, Feb 2006. ISSN 0018-9499.

[41] Georges Aad et al. Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data. 2014.

[42] R. Nicolaidou, L. Chevalier, S. Hassani, J.F. Laporte, E. Le Menedeu, and A. Ouraou. Muon identification procedure for the atlas detector at the lhc using muonboy reconstruction package and tests of its performance using cosmic rays and single beam data. *Journal of Physics: Conference Series*, 219, 2010.

[43] *Performance of E_T miss reconstruction in first ATLAS data*, volume Phenomenology Symposium 10 - 12 May, 2010.

[44] R Nicolaidou, L Chevalier, S Hassani, J F Laporte, E Le Menedeu, and A Ouraou. Muon identification procedure for the atlas detector at the lhc using muonboy reconstruction package and tests of its performance using cosmic rays and single beam data. *Journal of Physics: Conference Series*, 219(3):032052, 2010. URL `http://stacks.iop.org/1742-6596/219/i=3/a=032052`.

[45] Silvia Resconi Donatella Cavalli. Missing energy twiki: https://twiki.cern.ch/twiki/bin/view/atlasprotected/etmiss. URL `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EtMiss`.

[46] J. H. Christenson, G. S. Hicks, L. M. Lederman, P. J. Limon, B. G. Pope, and E. Zavattini. Observation of massive muon pairs in hadron collisions. *Phys. Rev. Lett.*, 25:1523–1526, Nov 1970. URL `http://link.aps.org/doi/10.1103/PhysRevLett.25.1523`.

[47] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput.Phys.Commun.*, 136:198–207, 2001.

[48] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:pp. 289–337, 1933. ISSN 02643952. URL `http://www.jstor.org/stable/91247`.

[49] R. Cousins. Why isn't every physicist a bayesian. *Am. J. Ph ys*, 63:398, 1995.

[50] Harry Nelson and Tim Nelson. A Guide to Unbinned Maximum Likelihood Fits. October 1998.

[51] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur.Phys.J.*, C71:1554, 2011.

[52] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62, 03 1938. URL `http://dx.doi.org/10.1214/aoms/1177732360`.

[53] Principal component analysis - wikipedia, 2013. URL `http://en.wikipedia.org/wiki/Principal_component_analysis`. [Online; accessed 7-November-2013].

[54] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[55] Jonathon Shlens. A tutorial on principal component analysis. In *A tutorial on Principal Component Analysis*. Systems Neurobiology Laboratory, Salk Institute for Biological Studies, 2005.

[56] A. Kraskov, H. Stögbauer, R. G. Andrzejak, and P. Grassberger. Hierarchical Clustering Based on Mutual Information. *eprint arXiv:q-bio/0311039*, November 2003.

[57] Root tprincipal class, 2000. URL `http://root.cern.ch/root/html/TPrincipal.html`. [Online; accessed 21-November-2013].

[58] Root tminuit class, 2000. URL `http://root.cern.ch/root/html/TMinuit.html`. [Online; accessed 25-November-2014].

[59] L. Demortier and Lyons L. Everything you always wanted to know about pulls. February 2002. URL `http://physics.rockefeller.edu/luc/technical_reports/cdf5776_pulls.pdf.`.

[60] Till Moritz Karbach and Maximilian Schlupp. Constraints on Yield Parameters in Extended Maximum Likelihood Fits. 2012.

[61] Search for New Physics in Events with Three Charged Leptons with the ATLAS detector. Technical Report ATLAS-CONF-2013-070, CERN, Geneva, Jul 2013. the full description is in the ATLAS-only supporting document [62].

[62] P-O DeViveiros, S Dube, M Hance, B Heinemann, I Hinchliffe, M Hurwitz, O Igonkina, J Mahlstedt, and D Yu. Search for New Physics in Events with Three Charged Leptons. Technical Report ATL-COM-PHYS-2014-236, CERN, Geneva, Mar 2014.

[63] Georges Aad et al. Search for anomalous production of prompt same-sign lepton pairs and pair-produced doubly charged Higgs bosons with $sqrts = 8$ TeV pp collisions using the ATLAS detector. 2014.

[64] Atlas top group: topmc12:, 2014. URL `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TopMC12`. [Online; accessed 27-Jan-2015].

[65] How to clean jets:, 2011. URL `Https://twiki.cern.ch/twiki/bin/view/atlasprotected/howtocleanjets2011`. [Online; accessed 05-May-2014].

[66] R Steuer, J Kurths, C. O. Daub, J. Weise, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *BIOINFORMATICS*, 18:S231–S240, 2002.

[67] Y. Moon, B. Rajagopalan, and U. Lall. Estimation of mutual information using kernel density estimators. *Phys. Rev. E*, 52:2318–2321, 1995.

[68] Makoto Matsumoto and Takuji Nishimura. Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator. *ACM Trans. Model. Comput. Simul.*, 8(1):3–30, January 1998. ISSN 1049-3301. URL `http://scholar.google.it/scholar?cluster=5600191046461070355`.

[69] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[70] Georges Aad et al. Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data. *Eur.Phys.J.*, C72:1909, 2012.

[71] Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010. 2011.

[72] Atlas internal: Pile-up reweighing tool:, 2014. URL `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/PileupReweighting`. [Online; accessed 12 June 2014].

[73] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. *Nucl.Phys.Proc.Suppl.*, 205-206:10–15, 2010.

[74] Georges Aad et al. Search for anomalous production of prompt like-sign lepton pairs at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1212:007, 2012.

[75] J-F Arguin, K Hamano, A Hawkins, B Heinemann, M Hurwitz, E Lytken, D Olivito, N Rodd, L Skinnari, and E Thomson. Search for anomalous production of prompt like-sign lepton pairs and constraints on physics beyond the Standard Model. Technical Report ATL-PHYS-INT-2012-086, CERN, Geneva, Nov 2012.

[76] Louise Anastasia Skinnari and Beate Heinemann. *A Search for Physics Beyond the Standard Model using Like-Sign Muon Pairs in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector.* PhD thesis, UC, Berkeley, Berkeley, Dec 2012. Presented 2012.

[77] Georges Aad et al. Improved luminosity determination in *pp* collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC. *Eur.Phys.J.*, C73:2518, 2013.

[78] J Butterworth, E Dobson, U Klein, B Mellado Garcia, T Nunnemann, J Qian, D Rebuzzi, and R Tanaka. Single Boson and Diboson Production Cross Sections in pp Collisions at sqrts=7 TeV. Technical Report ATL-COM-PHYS-2010-695, CERN, Geneva, Aug 2010.

[79] Michiel Botje, Jon Butterworth, Amanda Cooper-Sarkar, Albert de Roeck, Joel Feltesse, et al. The PDF4LHC Working Group Interim Recommendations. 2011.

[80] John M. Campbell and R. Keith Ellis. $t\bar{t}W^{+-}$ production and decay at NLO. *JHEP*, 1207:052, 2012.

[81] M.V. Garzelli, A. Kardos, C.G. Papadopoulos, and Z. Trocsanyi. t $\bar{t}$ W^{+-} and t $\bar{t}$ Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects. *JHEP*, 1211:056, 2012.

[82] S Stern. Searches for Rare Decays of the Higgs Boson with the ATLAS Detector. Dec 2013.

[83] *b*-jet tagging calibration on *c*-jets containing D^{*+} mesons. Technical Report ATLAS-CONF-2012-039, CERN, Geneva, Mar 2012.

[84] Reference cross-sections calculated using mcfm by the top group - atlas only, 2014. URL `https://twiki.cern.ch/twiki/pub/AtlasProtected/SusyMCSamples/sherpanorm.txt`. [Online; accessed 2 September 2014].

[85] ATLAS statistics forum. Frequentist limit recommendation, 2011.

[86] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), 2011. ISSN 1434-6044. URL `http://dx.doi.org/10.1140/epjc/s10052-011-1554-0`.

[87] A Measurement of WZ Production in Proton-Proton Collisions at sqrt(s)=8TeV with the ATLAS Detector. Technical Report ATLAS-CONF-2013-021, CERN, Geneva, Mar 2013.