Boosted Decision Trees for Separating Signal from Background Measurements in Search for Long Lived, Weakly Interacting Particles in pp Collisions at √s = 13 TeV

Ashley Batchelor, University of Washington Department of Physics

Abstract

I used boosted decision trees to develop models for separating signal and background data of measurements of pp collision events taken by the ATLAS detector of the Large Hadron Collider. Such models are necessary in the search for long lived, weakly interacting decay products of pp collisions. In order to train the weights of the models, signal data was provided by Monte Carlo simulations of scalar particle (e.g. Higgs Boson) decays, and background data was provided by ATLAS detector data with feature selection based on missing energy values of less than 40 GeV.

Introduction

A primary aim of the Large Hadron Collider is to search for "Beyond Standard Model" (BSM) physics phenomena. This includes, for example, Supersymmetry (SUSY) models, dark matter, and baryon asymmetry. A potentially useful area for observing BSM physics is long-lived particles (LLPs) which are the products of high energy proton-proton (pp) collisions. Three processes of particular interest are introduced below.

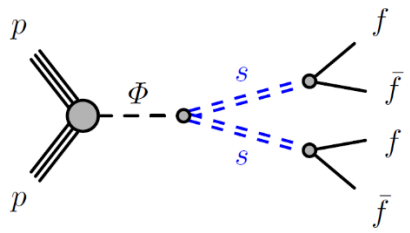


Figure 1 - Feynman diagram of a Higgs scalar portal model that gives rise to long lived particles. [1]

In an example shown in Figure 1, two protons collide and produce a scalar Φ (e.g. a Higgs Boson), which decays to a pair of long-lived particles s, which each decay to pairs of fermions. [5] This is sometimes referred to as a "Higgs scalar portal" when the scalar Φ is a Higgs Boson. [1]

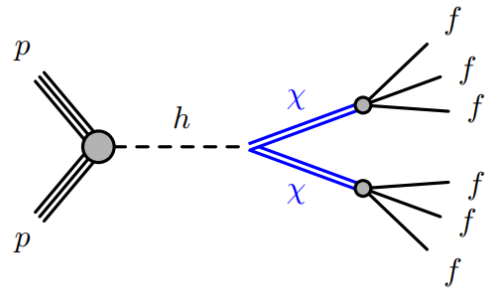


Figure 2 - Feynman diagram of a Higgs portal baryogenesis model that gives rise to long lived particles. [1]

In an example shown in Figure 2, two protons collide and produce a Higgs boson, which decays into a pair of long-lived Majorana fermions χ. The Majorana fermions χ decay into fermions. This process, referred to as "Higgs portal baryogenesis" violates baryon and/or lepton number conservation. [1]

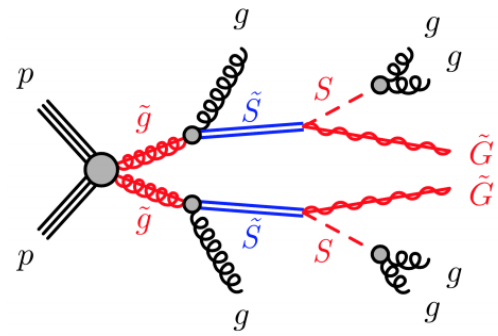


Figure 3 - Feynman diagram of a stealth SUSY model that gives rise to long lived particles. [1]

In an example shown in Figure 3, two protons collide and produce a pair of gluinos $\tilde{g}$, which each decay into a gluon and a long-lived singlino $\tilde{S}$. Each of the singlinos $\tilde{S}$ decays into a singlet S which decays into a gluon-jet g, and a light gravitino $\tilde{G}$. This process is referred to as "stealth SUSY." [1]

Theoretical Background

Motivation

The discovery of the Higgs Boson announced in 2012 [2] marked the final piece of the Standard Model (SM) to be observed in experiment. While the SM is a very successful theory which in most cases agrees very closely with experiment, sometimes to parts per trillion (e.g. the g-factor of the electron [29]), the Standard Model includes various insufficiencies for which theorists have proposed solutions. A major insufficiency of the SM is the hierarchy problem. The hierarchy problem occurs when a fundamental physical parameter differs greatly from its effective value in experiment. The most prominent hierarchy problem involves the mass of the Higgs Boson. The Higgs Boson has been observed to have a mass m_{obs} of 125 GeV. Accounting for this mass requires "fine-tuning." The first-order SM

correction to the observed Higgs mass (shown in the Feynman diagram below) is a quadratic divergence given by the equation:

$$m_{obs}^2 = m_0^2 - \frac{1}{16\pi^2}\lambda^2\Lambda^2 + \cdots$$

The term m_0^2 is a free parameter of the SM and λ is the Higgs-fermion coupling. If Λ is on the order of the Planck scale (10^{19} GeV), then the term m_0^2 must correct the observed Higgs mass m_{obs} to one part in 10^{34}. This massive correction is referred to as "fine-tuning," and remains a troublesome feature of the SM. [4]

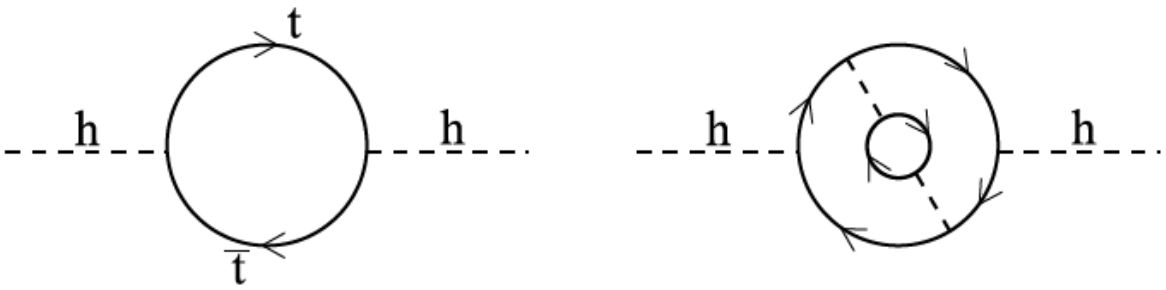


Figure 4 - One example of a Feynman diagram of a one-loop top quark correction to the Higgs mass (left), and a higher order correction (right) [3]

SUSY is one of the leading candidates for BSM physics and the subject of many studies with the LHC. SUSY introduces an addition symmetry to the Standard Model, where each fermion and boson has a superpartner which is a boson or fermion, respectively. For fermions, their superpartners are named with an s at the beginning of the name of the SM counterpart. For bosons, their superpartners are named with "ino" at the end of the name of the Standard Model counterpart. Thus, an electron has a superparter called a selectron, and a gluon has a superpartner called a gluino, each of which is indicated by the SM symbol topped with a ~. [3] Thus far, SUSY superpartner particles have never been observed at the LHC or elsewhere.

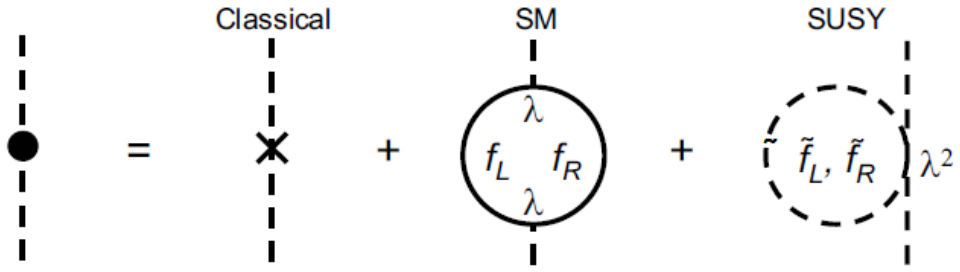


Figure 5 - Feynman diagrams for Higgs mass as a sum of classical behavior, a SM one-loop correction, and an additional SUSY correction [4]

SUSY modifies the SM Higgs mass correction such that:

$$m_{obs}^2 = m_0^2 - \frac{1}{16\pi^2}\lambda^2\Lambda^2 + \frac{1}{16\pi^2}\lambda^2\Lambda^2 \ldots$$

$$m_{obs}^2 \approx m_0^2 + (m_{\tilde{f}}^2 - m_f^2)\ln\frac{\Lambda}{m_f}$$

The two mass terms $m_{\tilde{f}}^2$ and m_f^2 are the masses of the two fermions in the rightmost Feynman diagram above. This equation assumes that the Higgs-fermion coupling constants λ are identical for fermions and their superpartners, and with this assumption, the one-loop quadratic divergence term cancels out for all fermions. The SUSY correction is logarithmic, and thus even a very large Λ will not require fine-tuning like the SM.

Thus far, superpartners have never been observed, which may be indicative of a breaking of supersymmetry. Several models for breaking of this symmetry either necessitate or include the possibility of particles which are long-lived and which decay into SM particles. [5]

Hidden Sector

Another potential solution to the hierarchy problem is "Hidden Sector" scalar fields. In various models, the Hidden Sector is coupled to the SM through a communicator particle. The communicator particle is a scalar boson (e.g. the Higgs Boson) which couples with a hidden sector scalar boson, which decays into SM particles. [6] Very small couplings to the Hidden Sector can result in decay modes with large branching ratios resulting in long-lived particles. [7] At present, the upper bound for the branching ratio into undetected particles is at 25%, which allows for very large branching ratios for BSM processes. [1] Searches thus far have focused on searching for communicator particles with a mass range of 100 GeV to 900 GeV, which includes the Higgs Boson as a potential communicator. [6] [8]

Dark Matter

Long-lived particles may also be indicative of dark matter. The lightest particle allowed by SUSY is sometimes referred to as the Lightest Supersymmetry Particle (LSP), which in many models is a neutralino. The neutralino may be a dark matter particle. Furthermore, if it is produced in pp collisions, a large amount of energy from the collision event will not be detected by the ATLAS detector. Therefore, the search for dark matter in colliders is frequently based on missing transverse momentum $E_{miss,T}$ from dark matter particles (e.g. a neutralino) recoiling against a SM particle X. This is referred to as an $X + E_{miss,T}$ signature. A Higgs Boson as particle X is a promising area of probing for dark matter. [9]

Processes that follow the Higgs portal model shown in Figure 1 have two potential signatures in signal events. First, a signal event may have two or more decay vertices. This signature is denoted 2MSVx. Second, a signal event may have precisely one decay vertex and a missing energy $E_{miss,T}$ of at least 30 GeV which is not detected by the ATLAS detector, in the case of a portal scalar Φ which is a Higgs Boson with a 125 GeV mass. This signature is denoted 1MSVx + $E_{miss,T}$. [1]

Baryon Asymmetry

An important goal for BSM physics is to explain baryon asymmetry, or more specifically, the relative abundance of baryon matter in the observable universe compared to antibaryonic matter. Various models for baryon asymmetry also address the Higgs mass hierarchy problem described above.

One such model, the Higgs portal baryogenesis previously shown and described in Figure 2, is expected to produce a high level of Majorana fermions at 13 TeV. The decay modes of the Majorana fermions in this model violate baryon asymmetry. [1]

Processes that follow the Higgs portal baryogenesis models have two potential signatures in signal events. First, a signal event may have two or more decay vertices. This signature is denoted 2MSVx. Second, a signal event may have precisely one decay vertex and a missing energy $E_{miss,T}$ of at least 30 GeV which is not detected by the ATLAS detector, in the case of a portal scalar Φ which is a Higgs Boson with a 125 GeV mass. This signature is denoted 1MSVx + $E_{miss,T}$. [1]

Stealth SUSY

One model of interest to ATLAS studies does not have a large missing energy $E_{miss,T}$ like that of the portal scalar or Higgs portal baryogenesis models and is thus called, Stealth SUSY, because it may have gone undetected at the LHC and Tevatron. Because of this lack of missing energy, superpartner particles may have been produced in abundance while going undetected. A hidden sector may be the cause of the lack of missing energy. [10]

Processes that follow the stealth SUSY models have two potential signatures in signal events. First, a signal event may have two or more decay vertices. This signature is denoted 2MSVx. Second, a signal event may have precisely one decay vertex and at least two jets with energy E_T (associated with transverse momentum) of at least 150 GeV. This signature is denoted 1MSVx + Jets. [1]

Methods

ATLAS Detector

The ATLAS Detector uses cylindrical coordinates (z, r, ϕ), where the origin is defined as the collision interaction point, z is in the direction of the colliding beams, r is the radial distance transverse to the colliding beams, and ϕ is the azimuthal angle. A polar angle θ is defined as an angle with respect to the z direction. The pseudorapidity η associated with the polar angle θ is defined as [11]:

$$\eta = -\ln \tan \frac{\theta}{2}$$

An angular distance ΔR in the η and ϕ directions is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

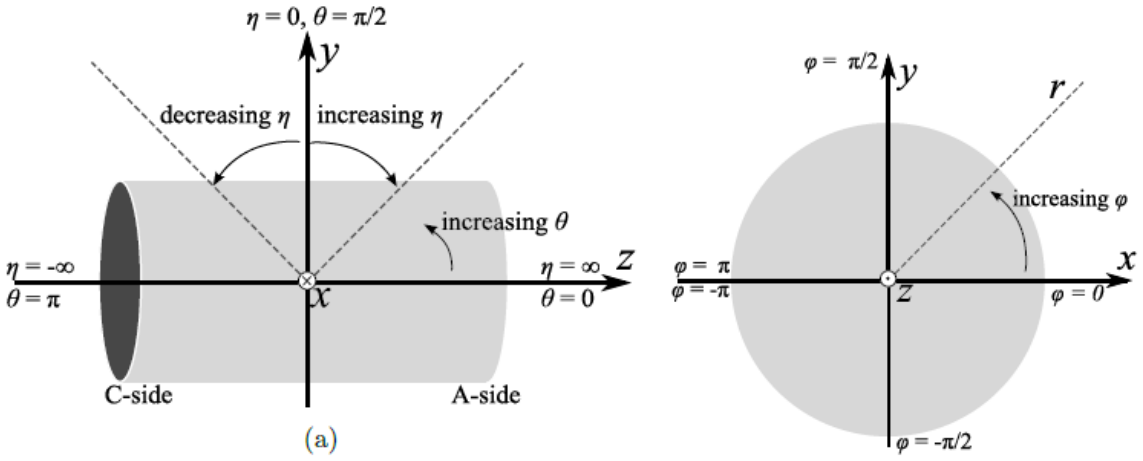


Figure 6 - Coordinate system of the ATLAS Detector [3]

Momentum is often expressed in terms of x, y, and z coordinates, where x is horizontal, and y is vertical. A transverse momentum component p_T represents momentum in the direction transverse to the beams, i.e. in the r direction (x and/or y).

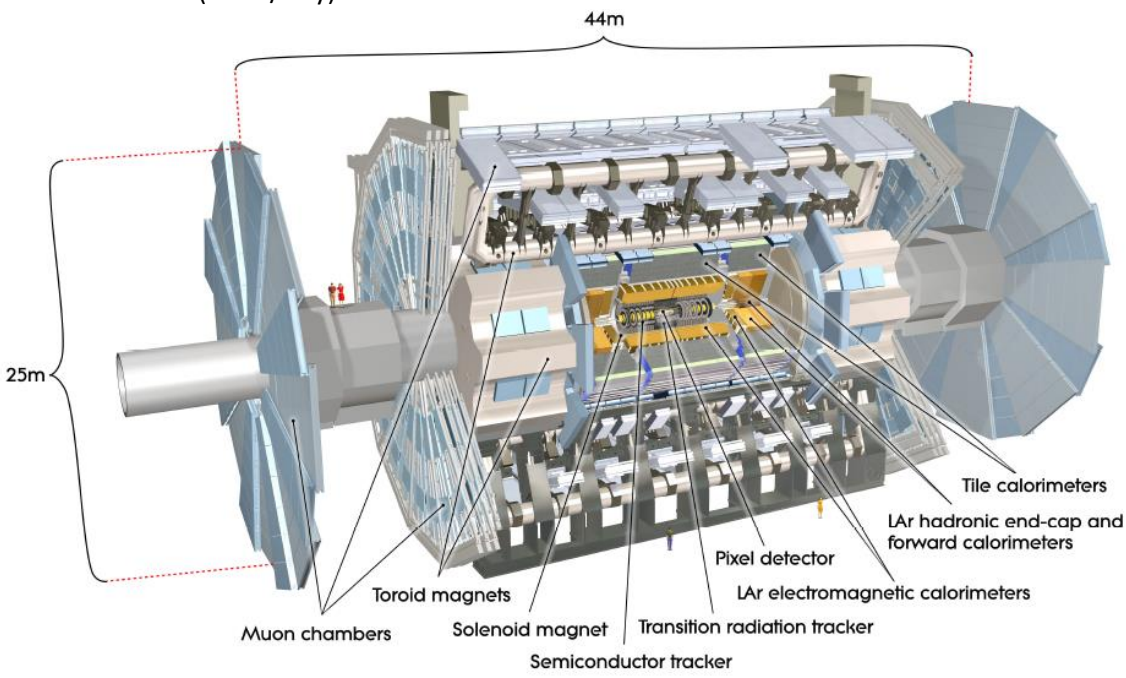


Figure 7 - Schematic of the ATLAS Detector [12]

Measurements associated with long lived particles generally take place in the outermost portion of the ATLAS detector in the muon chamber. As the name suggests, the muon chamber provides triggering and tracking of muon locations and momentum. The muon chamber includes a barrel and two endcaps.

The barrel includes three sections of resistive plate chambers (RPCs) and the endcaps include three sections of thin gap chambers (TGCs) for triggering. The barrel and endcaps also include monitored drift tubes (MDTs) and cathode strip chambers (CSCs) for tracking. The MDTs can track over most of the range of η while the CSCs track over the range the range $2.0 < |\eta| < 2.7$. The MDTs are arranged in chambers which have multilayers with three or four layers of tubes. [13]

A toroid magnet is configured to bend trajectories of particles in the range $|\eta| < 1.4$ in the barrel region. Two smaller toroid magnets are configured to bend trajectories of particles in the range $1.6 < |\eta| < 2.7$ in the end cap region. The combined fields of the barrel region magnet and the end cap region magnets are configured to bend trajectories of particles in the range $1.4 < |\eta| < 1.6$. [3]

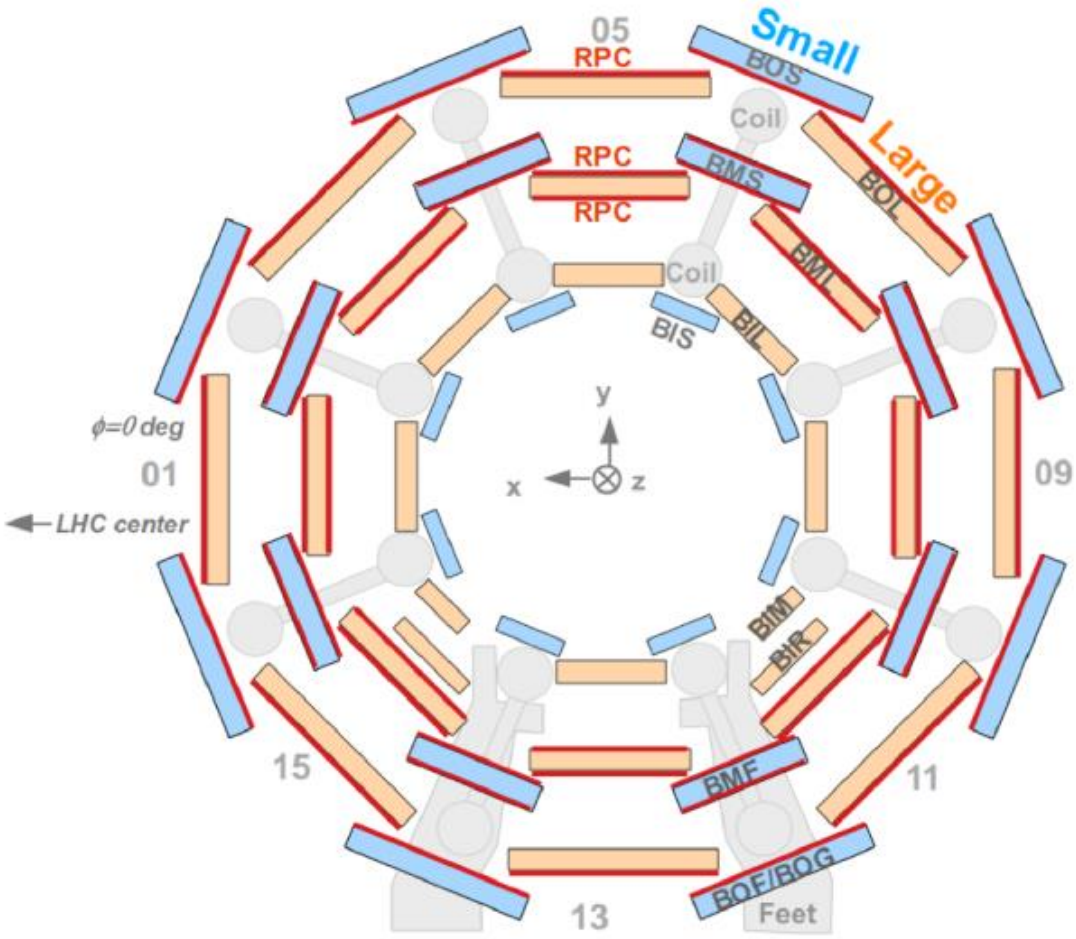


Figure 8 - Schematic cross-section of the barrel portion of the MS showing the arrangement of the MDT chambers and the RPCs with respect to the particle coordinate system. The MDT chambers are light blue and orange. The RPCs are red. [13]

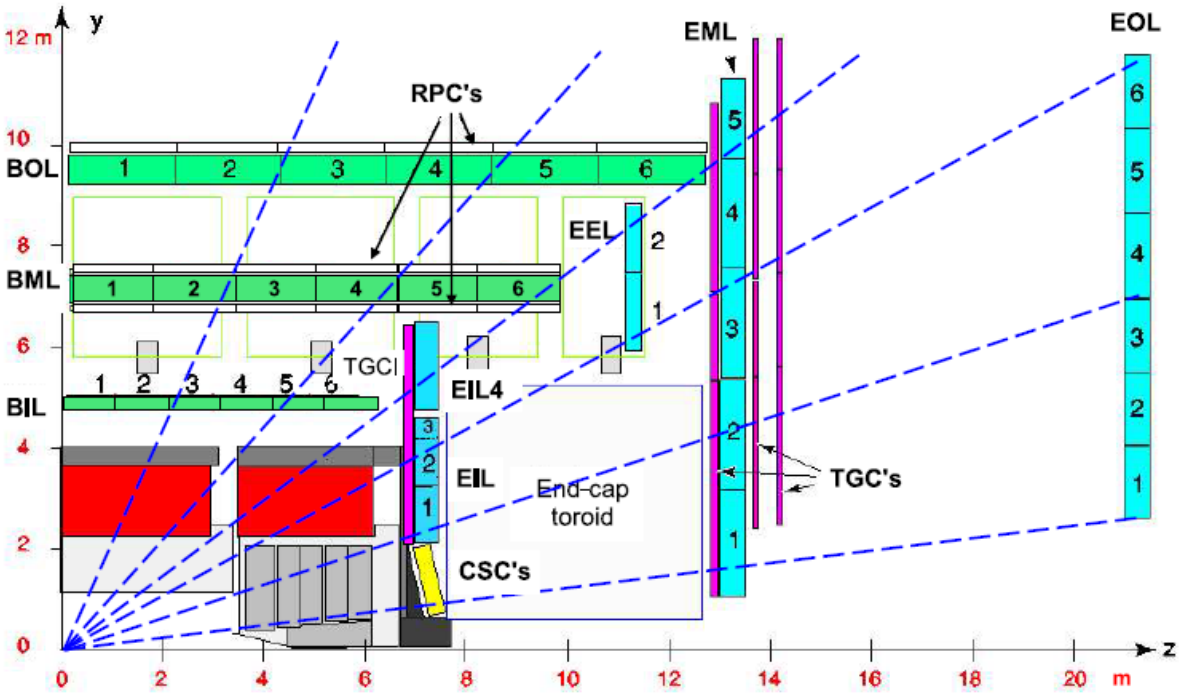


Figure 9 - Schematic cross-section of the ATLAS Detector in the y-z plane. The MDT chambers in the barrel are green, while the MDT chambers in the endcap are blue. The TGCs in the endcap are purple. [12]

The MS is the outermost detector at ATLAS. As such, it is the best detector for long lived particles that are of interest. Its detection ability is limited to detecting decay products of neutral long-lived particles with decay distances within the size of the MS. The MATHUSLA Detector (MAssive Timing Hodoscope for Ultra Stable neutraL pArticles) has been proposed as a means to detect neutral particles with very long lifetimes which are too long to be detected by the MS through their decay products. The MATHUSLA detector would be located above ground near the ATLAS detector. The MATHUSLA could detect particles carrying the missing energy $E_{miss,T}$ which would be a great benefit to dark matter and SUSY searches. [7]

Vertex Reconstruction

The barrel and endcap detectors use different algorithms for MS vertex reconstruction, although the general procedure is similar for each. The general procedure for each algorithm is defined in [13]:

1. MDT hits within a chamber multilayer are formed into 3- or 4-hit segments

2. Segments in the two multilayers of a chamber are joined to form tracklets

3. Clusters of tracklets are identified

4. Tracklets are back-extrapolated

5. Locations where multiple tracklets cross are identified as vertices

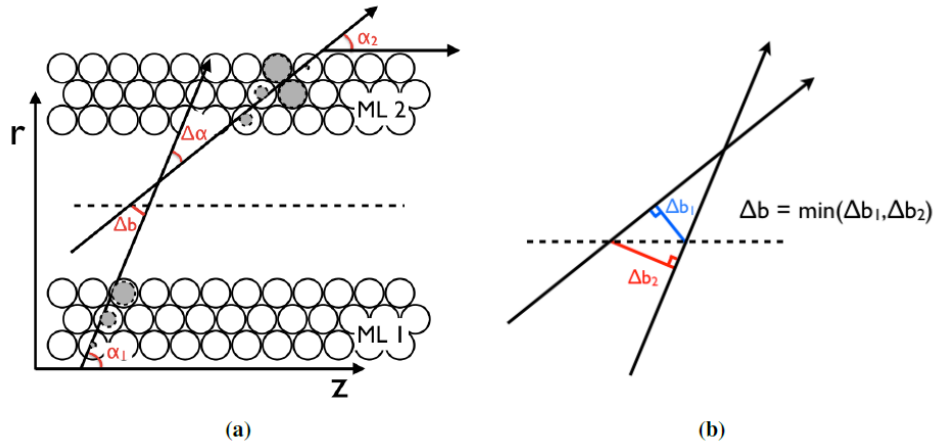


Figure 10 - Schematic diagrams of two reconstructed segments located respectively in MDT barrel multilayers denoted ML1 and ML2. [13]

In each MDT multilayer, measurements in adjacent tubes are used to reconstruct segment trajectories. After determining segments, the segments may be used for tracklet reconstruction. As shown in Figure 10, the segment in ML1 has an angle α_1 with respect to the z direction, and the segment in ML2 has an angle α_2 with respect to the z direction. An angle $\Delta\alpha$ is the difference between α_1 and α_2. A distance Δb is the minimum distance of the between the two segments at a plane halfway between the multilayer ML1 and the multilayer ML2. Each combination of segments which has parameters $\Delta\alpha$ and Δb below specified thresholds may be paired to define a tracklet. When a multilayer is located within the MS magnetic field, a value of Δb=0 indicates that segments belong to the same particle and may be combined to a tracklet. For a multilayer outside the MS magnetic field, the angle $\Delta\alpha$ is used to match segments. For multilayers inside the MS magnetic field, the angle $\Delta\alpha$ may be used to determine the tracklet momentum (for low p_{T}). The tracklet momentum is inversely proportional to the angle $\Delta\alpha$, by a proportionality constant k which depends on the magnetic field within each of a respective pair of multilayers. Finally, a clustering algorithm is applied to the tracklets to cluster the tracklets into one or more vertices with a calculated vertex center. [3] [13]

Background Estimation

In order to estimate a background rate, ABCD techniques are used [14]. To illustrate ABCD techniques, a variable space with two uncorrelated variables is divided into four regions A, B, C, and D. The region A contains most of the signal. The amount of background in A may be estimated from the events in B, C, and D by a simple linear relation. Each region contains a population of events N_A, N_B, N_C, and N_D. The population NA may then be estimated as:

$$N_A = N_B \frac{N_C}{N_D}$$

This method assumes that there is a sufficient number of events in B, C, and D to estimate a linear uncertainty in A; that there is no signal in B, C, and D; the background comes from a single source; and the two variables are uncorrelated. This equation may be used for a validation region and a signal

region, where the validation region is dominated by background, and where the signal region is dominated by the signal of interest. [1]

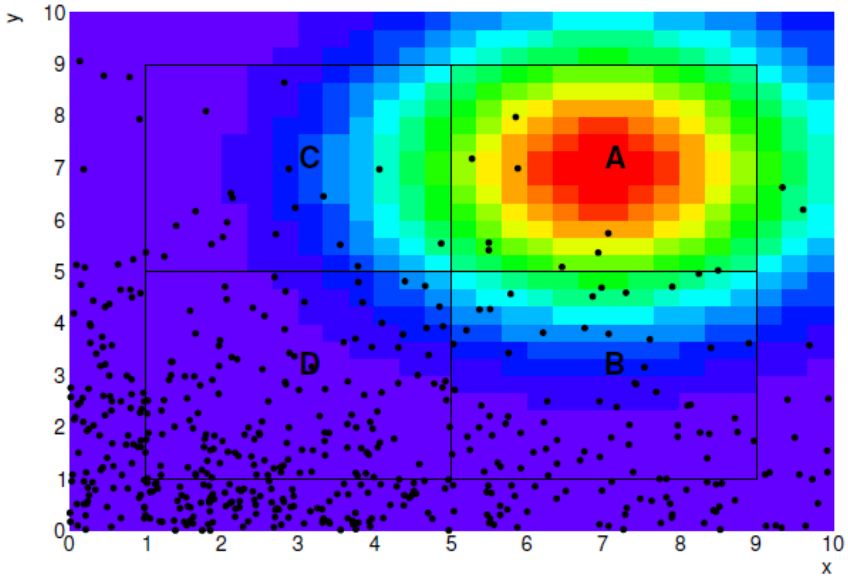


Figure 11 - A simple representation of an ABCD plane for uncorrelated variables x and y. The signal region is A and the background regions are B, C, and D. [14]

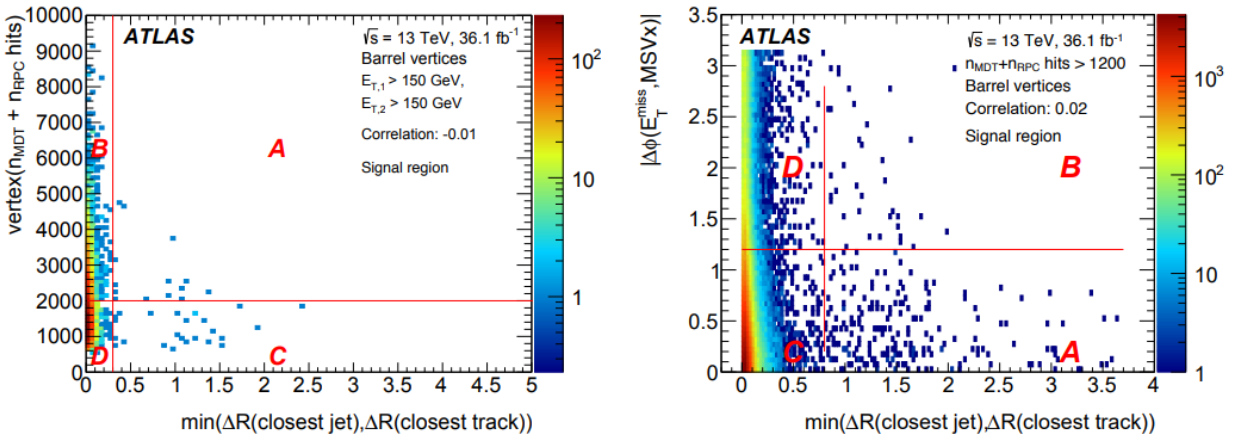


Figure 12 - Examples of ABCD planes for signal region for 1MSVx + jets (left) and 1MSVs + $E_{miss,T}$ (right) [1]

At present there are three variables used to define the ABCD plane (two at a time), which are shown in Figure 12:

$$n_{MDT} + n_{RPC}$$

$$\min(\Delta R(\text{closest jet}), \Delta R(\text{closest track}))$$

$$\Delta\phi(E_{miss,T}, MSVtx)$$

The value $n_{MDT} + n_{RPC}$ is the number of hits associated with a reconstructed vertex. The value $\min(\Delta R(\text{closest jet}), \Delta R(\text{closest track}))$ is a vertex isolation between tracks and jets based on an angular difference. The value $\Delta\phi(E_{miss,T}, MSVtx)$ is an angular difference between missing energy and an associated vertex direction.

In order to separate signal from background I developed models based on boosted decision trees, as described in detail below, with data labeled "signal" and "background." For the data labeled as "signal," I used Monte Carlo (MC) simulated data. For data labeled as "background," I used ATLAS detector data with $E_{miss,T}$ values of less than 40 GeV. For this work, I assumed that $E_{miss,T}$ below 40 GeV was a signal free region.

Signal Simulation

The signal data for fitting models was provided by MC simulations prepared with Pythia 8.210 software [15] to provide a parton shower model to the MadGraph5 MG5_aMC@NLO 2.2.3 software [16]. The simulations used an A14 set of parameters [17] with a NNPDF2.3LO parton distribution function (PDF) set [18]. Events were generated with EvtGen 1.2.0 software [19]. The events were processed with Geant4 software [20] to simulate the ATLAS geometry and the event response to the geometry [21]. Pileup was simulated with Pythia 8.210 software with an A2 parameter set [22] and a MSTW2008LO PDF set [23]. For the work described herein, I used MC simulations for a mass 125 GeV scalar decaying to 5, 35, and 55 GeV products. Other mass parameters relevant to this work were described in reference [1] as shown in Figure 13.

Model	m_ϕ [GeV]	m_s [GeV]
Scalar portal	100	8 , 25
	125	5 , 8 , 15 , 25 , 40
	200	8 , 25 , 50
	400	50 , 100
	600	50 , 150
	1000	50 , 150 , 400
	m_χ [GeV]	χ decay channel
Higgs portal baryogenesis	10	$\tau^+ \tau^- \nu_\ell$, $c\,b\,s$, $\ell^+ \bar{c}\,b$, $\nu\,b\bar{b}$
	30	
	50	
	100	
	$m_{\tilde{g}}$ [GeV]	$m_{\tilde{S}}$, m_S [GeV]
Stealth SUSY	250	100 , 90
	500	
	800	
	1200	
	1500	
	2000	

Figure 13 – Mass parameters for MC models used in reference [1]

Feature Selection

I used the variables below as features to train the boosted decision trees, which I calculated from the MC data. They are listed in descending order of importance in the model fits.

1. nHits – The total number of muon spectrometer hits
2. Segment_avg_dR – The average angular distance between segment trajectories and vertex trajectories
3. Tracklet_avg_dR – The average angular distance between tracklet trajectories and vertex trajectories
4. Segment_Sphericity – The sphericity metric of the segment trajectories
5. Segment_sumpT – The sum of the transverse momentum of the segment trajectories

For each event, I calculated nHits as a sum of the total number of MDT hits (nMDT), the total number of RPC hits (nRPC), and the total number of TGC hits (nTGC).

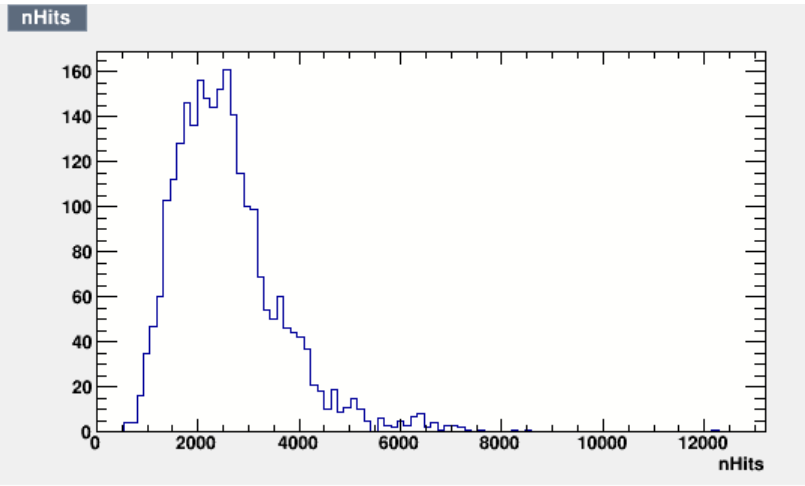


Figure 14 - Example of a histogram of nHits for 55 GeV decay products.

Each event was associated with m tracklets and n segments. For each event, I calculated the average ΔR between tracklet trajectories, and vertex trajectories as:

$$\overline{\Delta R}_{tracklet} = \frac{1}{m}\sum_{i}\sqrt{(\eta_{tracklet,i} - \eta_{vertex})^2 + (\phi_{tracklet,i} - \phi_{vertex})^2}$$

$$\overline{\Delta R}_{segment} = \frac{1}{n}\sum_{i}\sqrt{(\eta_{segment,i} - \eta_{vertex})^2 + (\phi_{segment,i} - \phi_{vertex})^2}$$

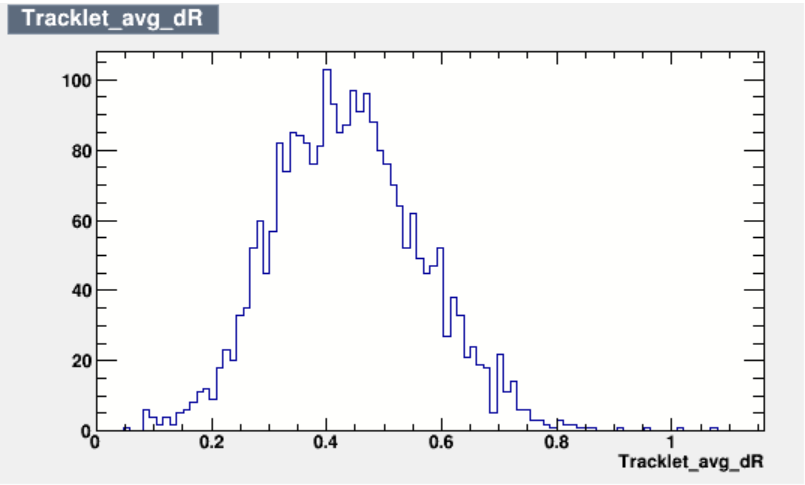


Figure 15 - Example of a histogram of Tracklet_avg_dR for 55 GeV decay products in the barrel region.

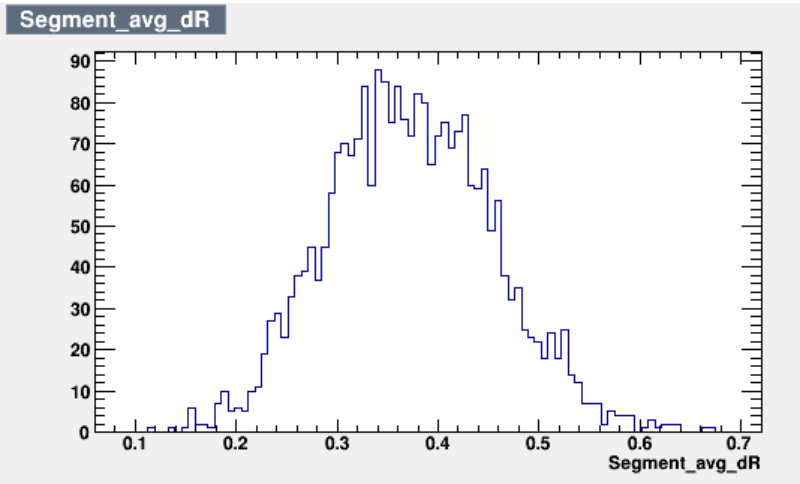


Figure 16 - Example of a histogram of Segment_avg_dR for 55 GeV decay products in the barrel region.

For the calculations of the averages above, I only counted segments and tracklets matched to vertices. I only included tracklet and segment measurements which satisfied the restrictions:

$$\left|\eta_{tracklet,i} - \eta_{vertex}\right| < 0.7$$

$$\left|\phi_{tracklet,i} - \phi_{vertex}\right| < 1.0$$

$$\left|\eta_{segment,i} - \eta_{vertex}\right| < 0.7$$

$$\left|\phi_{segment,i} - \phi_{vertex}\right| < 1.0$$

For each event, I calculated the segment sphericity value from the momentum tensor $M_{x,y,z}$ defined in reference [31] as:

$$M_{x,y,z} = \sum_i \begin{pmatrix} p_{xi}^2 & p_{xi}p_{yi} & p_{xi}p_{zi} \\ p_{yi}p_{xi} & p_{yi}^2 & p_{yi}p_{zi} \\ p_{zi}p_{xi} & p_{zi}p_{yi} & p_{zi}^2 \end{pmatrix}$$

Where p_{xi}, p_{yi}, and p_{zi} are the momentum components of segments of an event denoted by subscript i. The momentum tensor $M_{x,y,z}$ is summed over all segments of the event. The normalized eigenvalues of the tensor are defined in order of value $\lambda_1 > \lambda_2 > \lambda_3$, where $\sum_i \lambda_i = 1$. The sphericity S is defined as:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3)$$

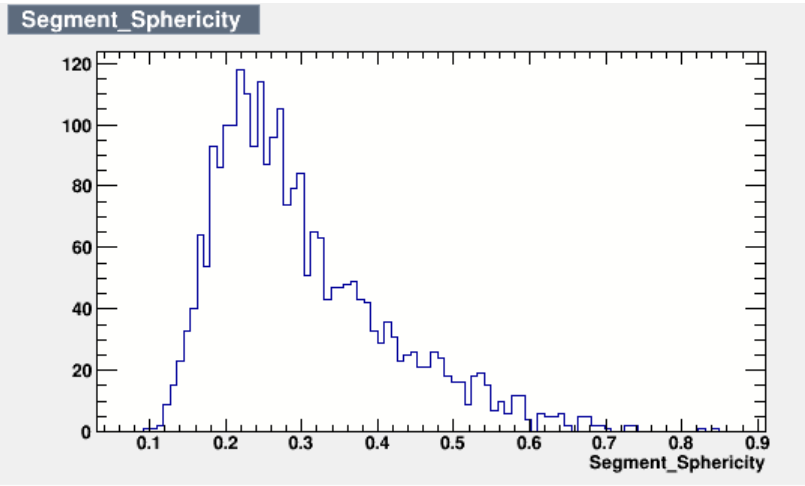


Figure 17 - Example of a histogram of Segment_Sphericity for 55 GeV decay products in the barrel region.

For each event, I calculated the sum of the segment transverse momentum $p_{T,segment,tot}$ as:

$$p_{T,segment,tot} = \sum_{i=0}^{n-1} p_{T,segment,i}$$

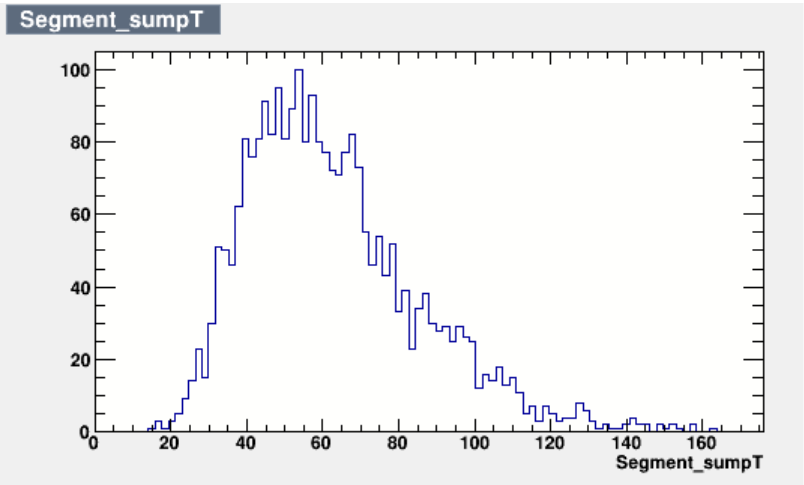


Figure 18 - Example of a histogram of Segment_sumpT for 55 GeV decay products in the barrel region.

In order to determine whether a vertex of an event is associated with a barrel or endcap detector region, or another region, I applied a set of criteria based on ranges of the pseudorapidity η, a transverse decay position L_{xy}, and a longitudinal decay position L_z. [1] For events with a vertex in the range $|\eta| < 0.7$ and $3{,}000\ \text{mm} < L_{xy} < 8{,}000\ \text{mm}$, I assigned the vertex to the barrel region. For events with a vertex in the range $1.3 < |\eta| < 2.5$, $L_{xy} < 10{,}000\ \text{mm}$, and $5{,}000\ \text{mm} < |L_z| < 15{,}000\ \text{mm}$, I assigned the vertex to the endcap region.

Decision Trees

A decision tree is a recursive series of cuts applied to at least two variables to separate data from at least two classes. In the scope of the present work, the two classes were signal and background. Figure 19 shows an example of a decision tree with 5 nodes defined illustrated by 5 cuts in the X-Y plane which separate either signal, background, or an additional node. The decision tree in Figure 19 uses rectangular cuts, where each node specifies a cut based on a single variable X or Y.

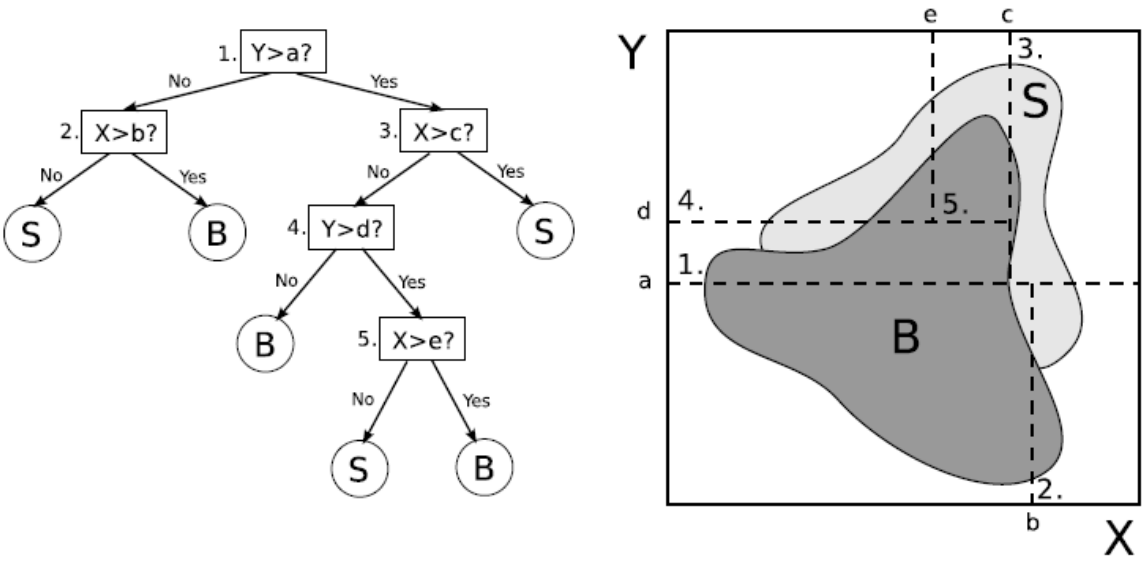


Figure 19 - Example of a basic decision tree (left) for variables X and Y, which uses single variable rectangular cuts (right) for separating signal S from background B. [24]

An improvement to the basic decision tree described above is to use boosting techniques. In order to make the tree stable to statistical fluctuations, a tree may be constructed by averaging multiple trees based on boosting techniques [25]. The most popular boosting technique is the AdaBoost algorithm (short for "adaptive boosting") which was first outlined by Freund and Schapire in 1996 [26]. The present work utilizes AdaBoost for constructing decision trees.

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in \mathcal{X}$, $y_i \in \{-1, +1\}$.
Initialize: $D_1(i) = 1/m$ for $i = 1, \ldots, m$.
For $t = 1, \ldots, T$:

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : \mathcal{X} \to \{-1, +1\}$.
- Aim: select h_t with low weighted error:

$$\varepsilon_t = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i].$$

- Choose $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$.
- Update, for $i = 1, \ldots, m$:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

Figure 20 - AdaBoost algorithm from Schapire, "Explaining AdaBoost." [26]

The AdaBoost algorithm is an iterative method of building a decision tree based on weighted observations of a binary classifier. A set of m observations is given by (x_1, y_1) though (x_m, y_m), where

x_i is a data point of a set X and y_i is a binary classifier equal to -1 or 1. An initial set of weights $D_i(i)$ is set uniformly to $1/m$ for each i.

Over T iterations ($t = 1\ to\ T$), during each iteration, a weak learner (i.e. a decision tree) is fit using the weights $D_t(i)$ by selecting a weak hypothesis h_t, which minimizes an error ε_t, which is a sum of the weights $D_t(i)$ of each of the misclassified points. A value α_t is set to:

$$\alpha_t = \frac{1}{2}\ln\frac{1-\varepsilon_t}{\varepsilon_t}$$

The next set of weights $D_{t+1}(i)$ is updated to:

$$D_{t+1}(i) = \frac{D_t(i)e^{-\alpha_t y_i h_t(x_i)}}{Z_t}$$

where Z_t is a normalization factor to normalize $D_{t+1}(i)$. Finally, the output hypothesis is given by:

$$H(x_i) = sign\left(\sum_{t=1}^{T}\alpha_t h_t(x_i)\right)$$

An improvement to a decision tree such as an AdaBoost tree is to use multivariate cuts. Multivariate cuts may be based on a Fisher discriminant, as we did with the present work. A Fisher discriminant is a basic technique of linear discriminant analysis (LDA), which seeks to find a vector $\vec{w}$ onto which two classes of data may be projected, which is the normal of the Fisher discriminant hyperplane that defines a cut. The two classes (class 1 and class 2) have respective means $\vec{w}\cdot\vec{\mu}_1$ and $\vec{w}\cdot\vec{\mu}_2$, and variances $\vec{w}^T\sigma_1\vec{w}$ and $\vec{w}^T\sigma_1\vec{w}$. The Fisher discriminant maximizes the separation S defined by [27]:

$$S = \frac{(\vec{w}\cdot\vec{\mu}_1 - \vec{w}\cdot\vec{\mu}_2)^2}{\vec{w}^T\sigma_1\vec{w} + \vec{w}^T\sigma_2\vec{w}}$$

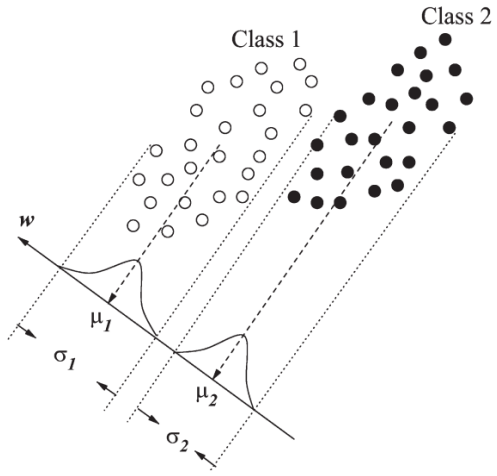


Figure 21 - Example of a Fisher Discriminant indicated by its normal vector $\vec{w}$ for two classes with means μ_1 and μ_2, and variances σ_1 and σ_2 as projected onto the direction $\vec{w}$. The goal is to determine a discriminant to maximize the difference between the means μ_1 and μ_2 and minimize the variances σ_1 and σ_2 in order to find the best separation of the classes 1 and 2. [27]

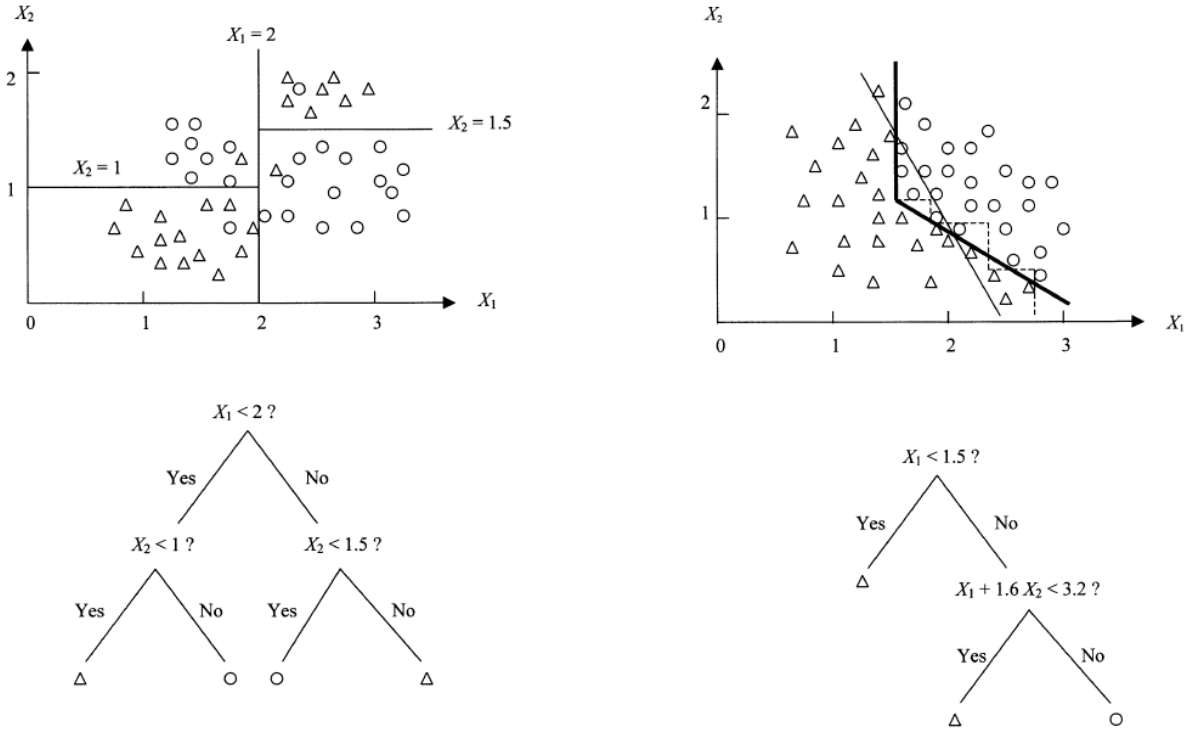


Figure 22 - Example of a decision tree for two classes of observations in variables X_1 and X_2, indicated by triangles and circles employing rectangular single variable cuts (left) and a decision tree employing multivariable Fisher cuts (right) [28]

A decision tree using multivariable Fisher cuts includes nodes which may have cuts defined by two or more variables at a time. As shown in Figure 22, a single cut based on multiple variables may provide a more optimal cut than a series of many rectangular cuts based on single variables.

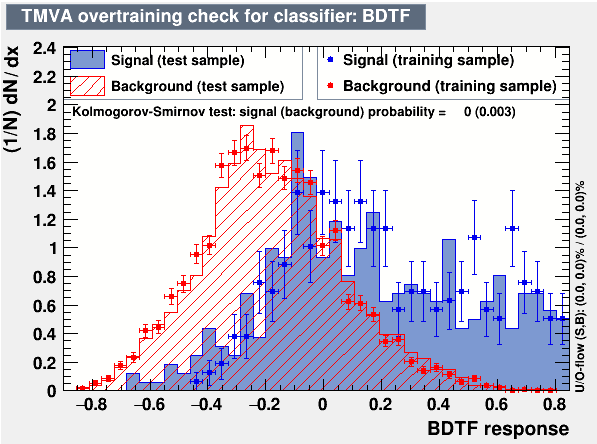
For each BDTF analysis, I used a set of event selection criteria corresponding to the Higgs Portal model illustrated in Figure 1. For the background, I selected events with $E_{miss,T}$ below 40 GeV, which was assumed to be a signal free region. For both background and signal I included several criteria. I selected half of the data for signal and background for machine learning, and half to use later for ABCD methods for background estimation using alternating entries from each data set (even and odd, respectively). I selected either barrel or endcap data (performing a separate BDTF for each). I selected events with one vertex which matched a muon Region of Interest (RoI) for that event with $\Delta R < 0.4$. A RoI is an angular region in (η,ϕ) which is used by ATLAS triggers for reducing unnecessary data. A RoI trigger is generally defined by a number of coinciding hits in the RPC or TGC. [1] [30]

I prepared the BDTF analysis using the Toolkit for Multivariate Data Analysis (TMVA) with ROOT software from CERN.

In order to improve the accuracy of predictive labelling of the models, I prepared an automated grid search of each set of Monte Carlo simulated signal data, varying the AdaBoostBeta parameter

(more specifically, the learning rate) and the number of trees used in the model fit. I looked at the integral of the ROC curve for each set of parameters as a metric for tuning improvement.

Computational Results



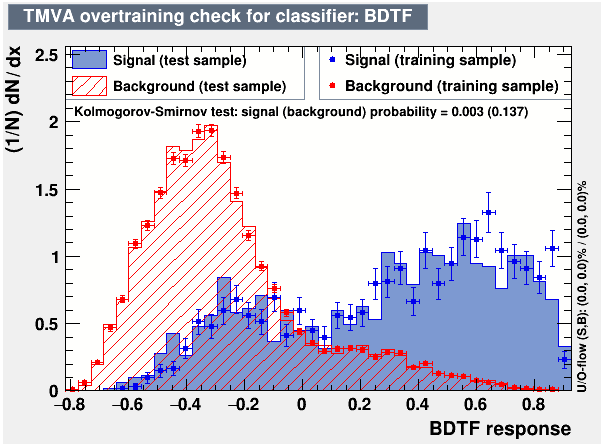
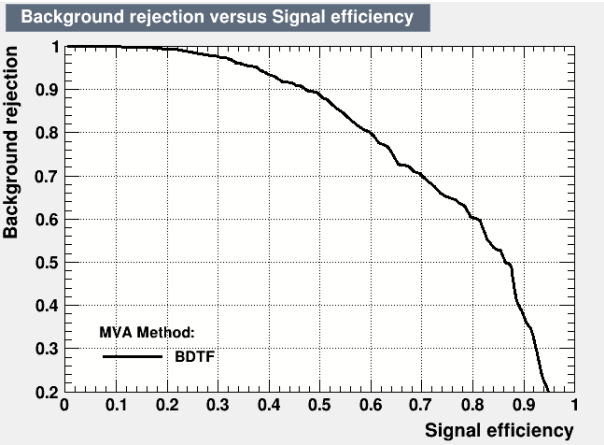


Figure 23, Figure 24 – BDTF Results for 5 GeV decay products, barrel and endcap.



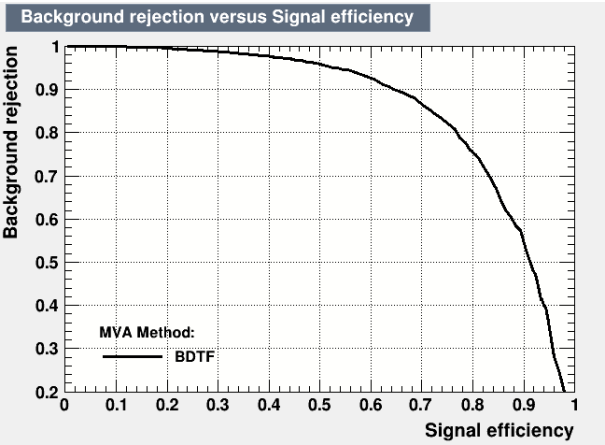


Figure 25, Figure 26 – ROC curves for 5 GeV decay products, barrel and endcap.

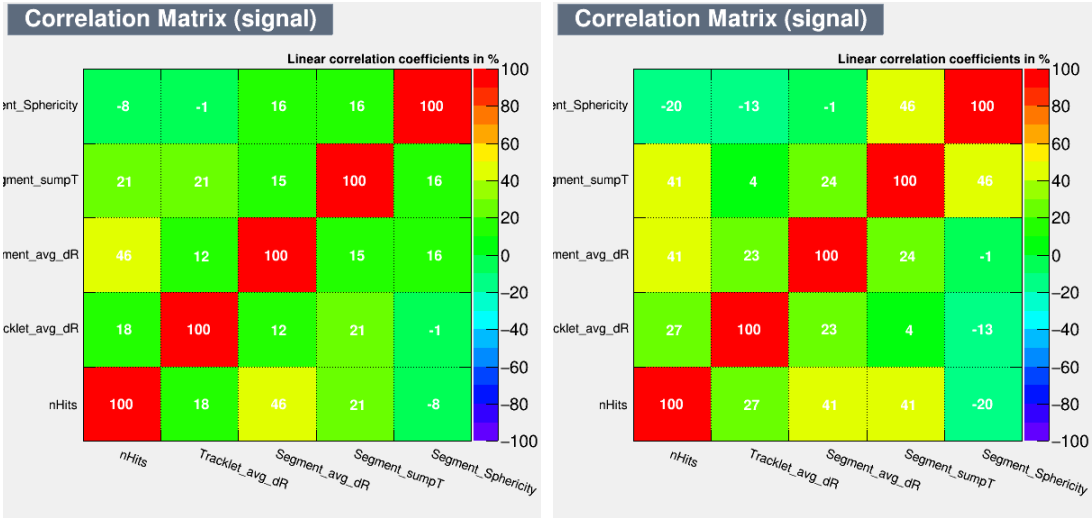


Figure 27, Figure 28 – Correlation matrices for 5 GeV decay products, barrel and endcap signal.

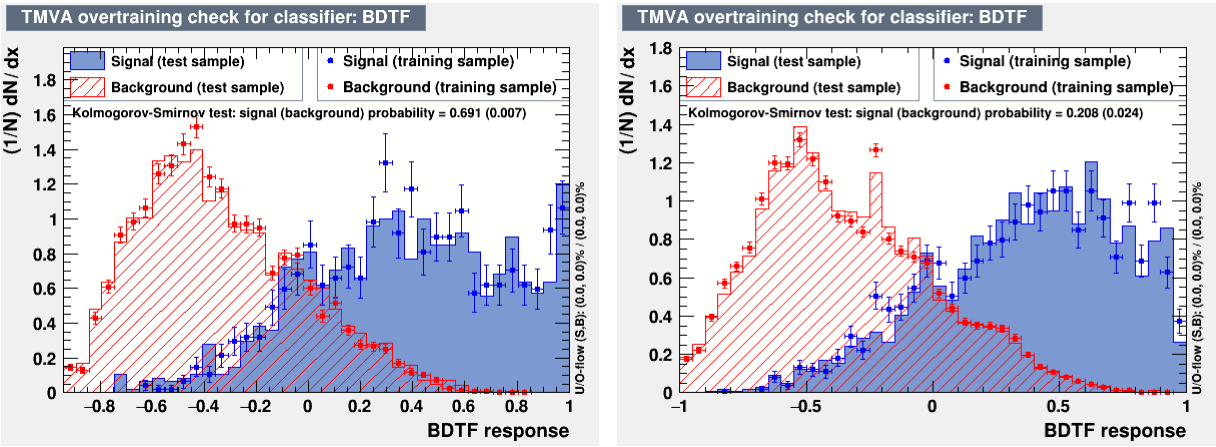


Figure 29, Figure 30 – BDTF Results for 35 GeV decay products, barrel and endcap.

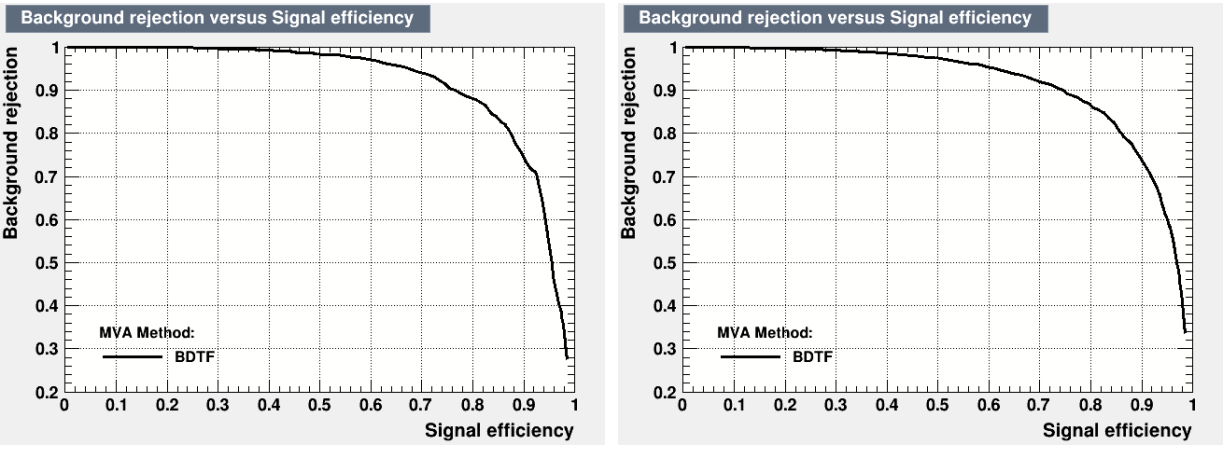
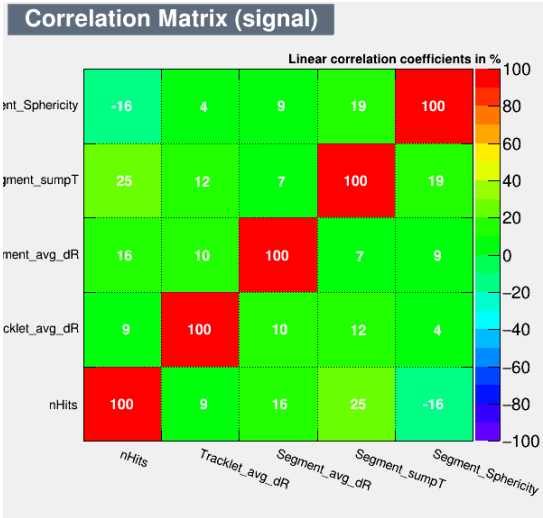


Figure 31, Figure 32 – ROC curves for 35 GeV decay products, barrel and endcap.



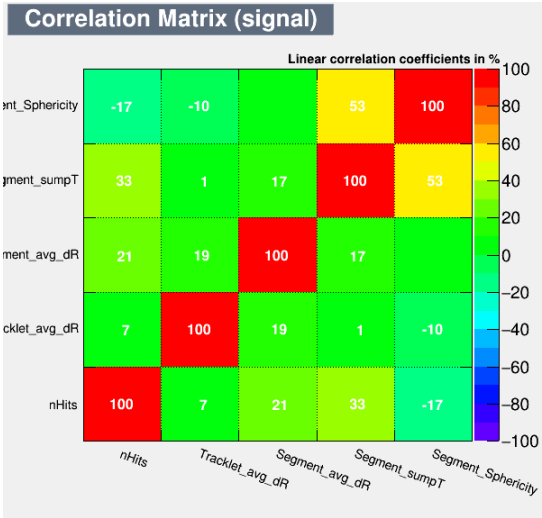
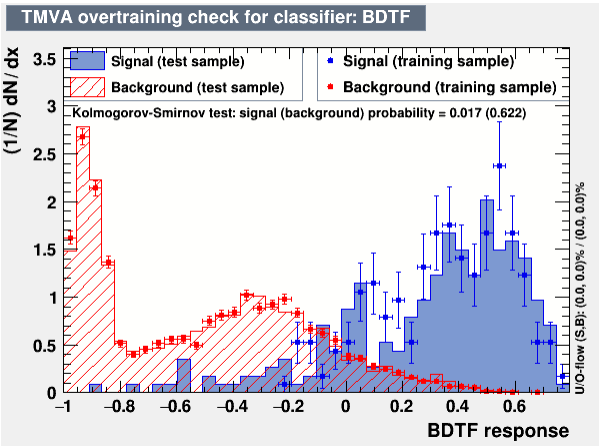


Figure 33, Figure 34 – Correlation matrices for 35 GeV decay products, barrel and endcap signal.



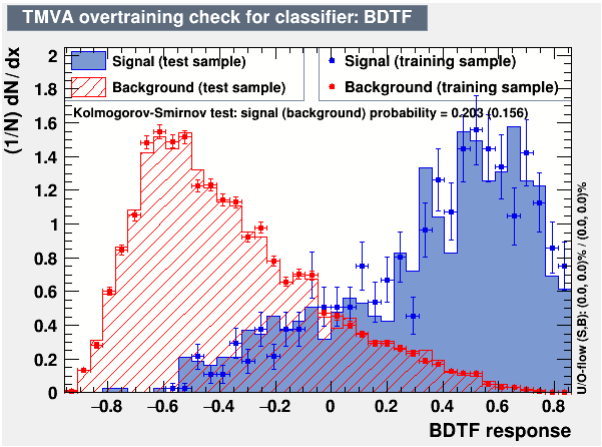
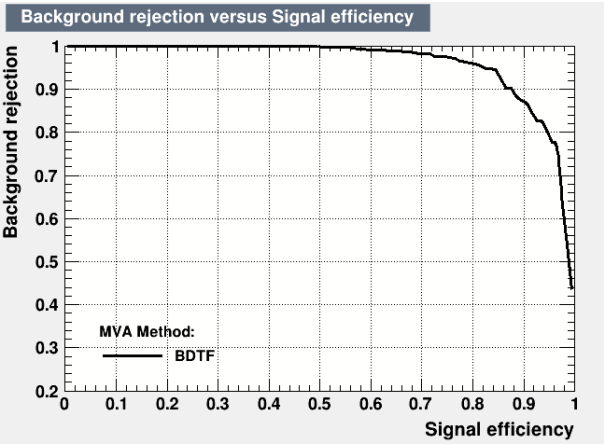


Figure 35, Figure 36 – BDTF Results for 55 GeV decay products, barrel and endcap.



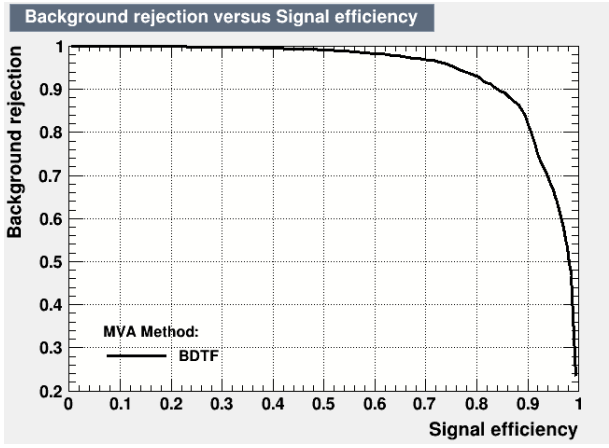
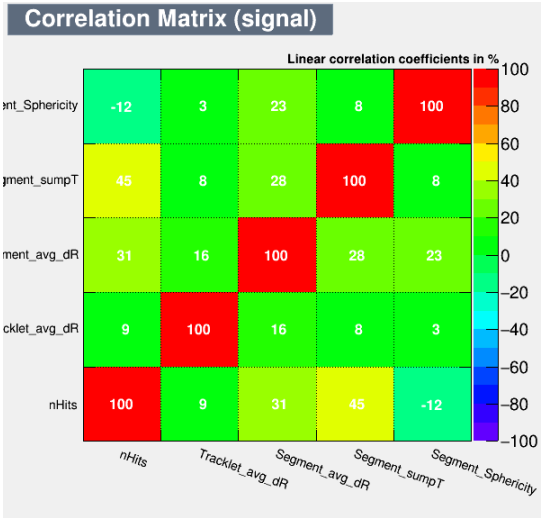


Figure 37, Figure 38 – ROC curves for 55 GeV decay products, barrel and endcap.



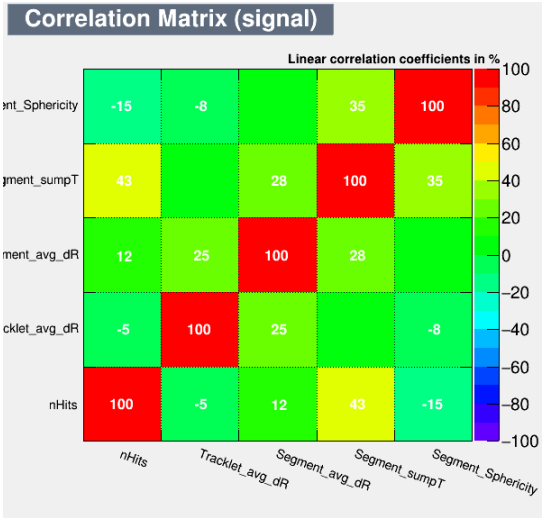


Figure 39, Figure 40 – Correlation matrices for 55 GeV decay products, barrel and endcap signal.

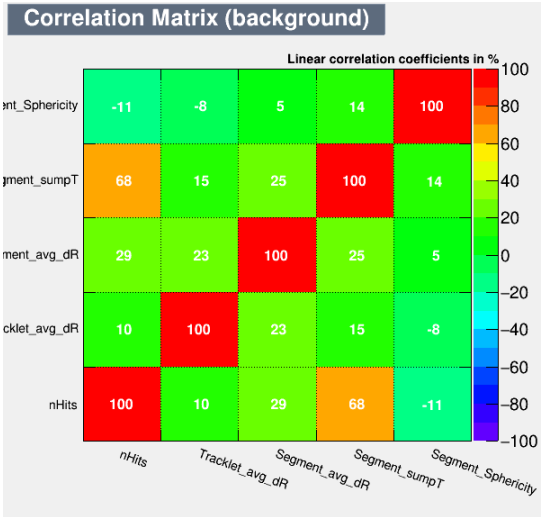


Figure 41 – Example of correlation matrix for background, from 55 Gev end cap.

The results of Figures 23-41 were from boosted decision trees with a learning rate of 0.5 and 50 trees, which are the default parameters used by TMVA for BDTF learning. Figures 42 and 43 show how the ROC integral was affected by varying these parameters.

Decay Product (GeV)	Learning Rate	Number of Trees	ROC Integral
5	0.2	30	0.782
		50	0.783
		100	0.78
	0.5	30	0.782
		50	0.78
		100	0.769
	1	30	0.759
		50	0.76
		100	0.749
35	0.2	30	0.917
		50	0.918
		100	0.919
	0.5	30	0.916
		50	0.916
		100	0.914
	1	30	0.91
		50	0.909
		100	0.91
55	0.2	30	0.963
		50	0.961
		100	0.96
	0.5	30	0.96
		50	0.96
		100	0.958
	1	30	0.953
		50	0.949
		100	0.95

Figure 42 – Table of tuning parameter results for barrel region (ROC integral)

Decay Product (GeV)	Learning Rate	Number of Trees	
5	0.2	30	0.857
		50	0.86
		100	0.861
	0.5	30	0.859
		50	0.859
		100	0.861
	1	30	0.856
		50	0.857
		100	0.857
35	0.2	30	0.914
		50	0.915
		100	0.915
	0.5	30	0.911
		50	0.912
		100	0.913
	1	30	0.906
		50	0.908
		100	0.911
55	0.2	30	0.943
		50	0.944
		100	0.945
	0.5	30	0.941
		50	0.941
		100	0.942
	1	30	0.935
		50	0.938
		100	0.938

Figure 43 – Table of tuning parameter results for end cap region (ROC integral)

Conclusions

Tuning the trees only provided a marginal improvement, at most a 0.5% improvement of the ROC integral. A higher number of events in the training sets would be the best way to improve the results. This is challenging because an initial MC simulation may include 300,000 or more events, but less than 1,000 of those may be suitable for training after event selection has been applied. In some cases, the

training set only included as many as 300 events. Thus, a very large set of MC data is required to improve the statistical capabilities of the training sets.

Acknowledgements

This work was prepared under the supervision of Henry Lubatti of the University of Washington ATLAS Group, with much help from Cristiano Alpigiani at CERN in Geneva.

References

[1] The ATLAS Collaboration. "Search for long-lived particles produced in pp collisions at √s = 13 TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer." Phys. Rev. D 99, 052005 (2019) arXiv: 1811.070730

[2] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." Phys.Lett. B716 (2012) 1-29. arXiv: 1207.7214

[3] Russell, Heather. "Search for long-lived particles decaying in the muon spectrometer of the ATLAS detector at the LHC." Phys. Rev. D 99, 052005 (2019).

[4] Feng, Jonathon, L. "Supersymmetry and Cosmology." SLAC Summer Institute, July 28 - August 8, 2003, Stanford, California.

[5] Rosten, Rachel. "A Search for Long Lived Neutral Particles Decaying to Hadronic States in Proton-Proton Collisions in the ATLAS Detector at √s = 8 TeV and √s = 13 TeV at the LHC." CERN-THESIS-2017-075.

[6] ATLAS Collaboration. "Search for pair-produced long-lived neutral particles decaying in the ATLAS hadronic calorimeter in pp collisions at √s = 8 TeV." Physics Letters B 743 (2015) 15-34. arXiv: 1501.04020

[7] Chou, John Paul; Curtin, David; Lubatti, H.J. "New detectors to explore the lifetime frontier." Physics Letters B. Volume 767, 10 April 2017, Pages 29-36

[8] Chang et al. "Nonstandard Higgs Boson Decays." Ann.Rev.Nucl.Part.Sci.58:75-98, 2008. arXiv: 0801.4554

[9] ATLAS Collaboration. "Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $b\bar{b}$ using 36 fb−1 of pp collisions at √s = 13 TeV with the ATLAS Detector." Phys. Rev. Lett. 119, 181804 (2017). arXiv: 1707.01302

[10] Fan, JiJi; Reece, Matthew; and Ruderman, Joshua T. "Stealth Supersymmetry." JHEP 1111 (2011) 012. arXiv: 1105.5135

[11] ATLAS Collaboration. "Search for long-lived, weakly interacting particles that decay to displaced hadronic jets in proton-proton collisions at √s = 8 TeV with the ATLAS detector." Phys. Rev. D 92 (2015) 012010. arXiv: 1504.03634

[12] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider." JINST 3 (2008)

S08003.

[13] ATLAS Collaboration. “Standalone vertex finding in the ATLAS muon spectrometer.” JINST 9 (2014) P02001. arXiv: 1311.7070

[14] ATLAS Statistics Forum. ABCD Method in Searches.

[15] T. Sjöstrand et al., An Introduction to PYTHIA 8.2, Comput. Phys. Commun. 191 (2014) 159, arXiv: 1410.3012.

[16] J. Alwall et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP 7 (2014) 79, arXiv: 1405.0301.

[17] ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021, 2014, url: https://cds.cern.ch/record/1966419.

[18] R. D. Ball et al., Parton distributions with LHC data, Nucl. Phys. B 867 (2013) 244, arXiv: 1207.1303 [hep-ph].

[19] D. J. Lange, The EvtGen particle decay simulation package, Nucl. Instr. Meth. A 462 (2001) 152.

[20] S. Agostinelli et al., Geant4 – a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250.

[21] ATLAS Collaboration, The ATLAS Simulation Infrastructure, Eur. Phys. J. C 70 (2010) 823, arXiv: 1005.4568.

[22] ATLAS Collaboration, Summary of ATLAS Pythia 8 tunes, ATL-PHYS-PUB-2012-003, 2012, url: https://cds.cern.ch/record/1474107.

[23] A. Martin, W. Stirling, R. Thorne, and G. Watt, Parton distributions for the LHC, Eur. Phys. J. C 63 (2009) 189, arXiv: 0901.0002

[24] EN Dawe, DC O'Neil, and S Protopopescu. “Using Boosted Decision Trees for Hadronic Tau Identification: On behalf of the Tau Performance Group.” Technical Report ATL-COM-PHYS-2010-603, CERN, Geneva, Aug 2010.

[25] Dale, Ørjan “Discovering SUSY with tau+tau- over tt_ Background - Evaluation of ATLAS Potential with 2010 Data.” https://wiki.uib.no/ift/images/4/49/Thesis_orjan_dale.pdf

[26] Schapire, R.E. “Explaining adaboost.” Empirical Inference: Springer (2013)

[27] Bandos, Tatyana V. et al. “Classification of Hyperspectral Images with Regularized Linear Discriminant Analysis.” IEEE Transactions on Geoscience and Remote Sensing, Vol. 47, No. 3, March 2009.

[28] Li, Xiao-Bai et al. “Multivariate Decision Trees Using Linear Discriminants and Tabu Search.” IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 33, No. 2, March 2003

[29] Hanneke, D., Fogwell, S., and Gabrielse, G. “New Measurement of the Electron Magnetic Moment and the Fine Structure Constant.” Phys. Rev. Lett. 100, 120801 (2008).

[30] The ATLAS Collaboration. “Triggers for displaced decays of long-lived neutral particles in the ATLAS detector.” JINST 8 (2013) P07015. arXiv: 1305.2284

[31] ATLAS Collaboration. “Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at √s = 7 TeV.” Eur. Phys. J. C (2012) 72: 2211. arXiv: 1206.2135