Machine Learning – Multivariate Analysis with Boosted Decision Trees

$H \longrightarrow W^+W^- \longrightarrow e\,\mu\,\nu_e\,\nu_\mu$

Jessica Strickland

Supervisors:

Prof. Michel Lefebvre

Dr. Manuela Venturi

INSTITUTE OF PARTICLE PHYSICS

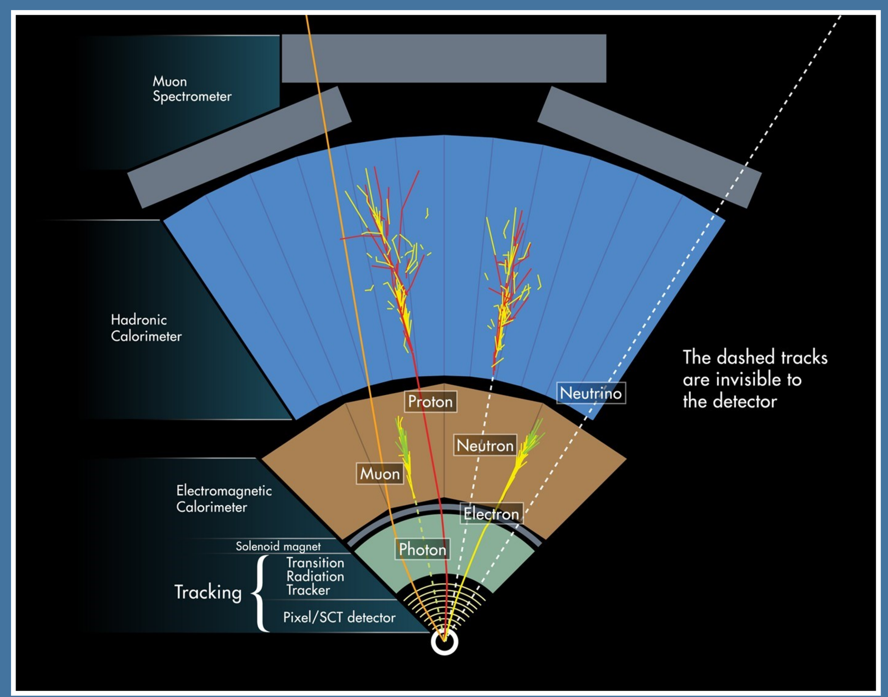
MEMORIAL UNIVERSITY

University of Victoria

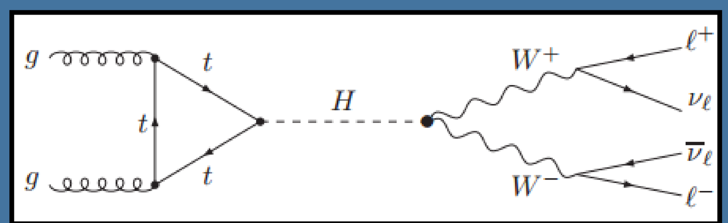
ATLAS

The Large Hadron Collider (LHC) is the world's largest and highest energy particle accelerator. It operated at $\sqrt{s}$ = 8 TeV in 2012, and is aspiring to 14 TeV during the second run starting in June, 2015. The LHC has 4 main experiments, one of which is the ATLAS (A Toroidal LHC Apparatus) detector, a general purpose one. Proton beams are accelerated by the LHC and subsequently collided at the ATLAS interaction point. Given the overwhelming amount of data produced in the collision, two levels of pre-selection triggers are applied, followed by an event filter stage which is able to reduce the rate to 200 events/s. To detect a wide variety of phenomenon, good resolution for leptons, photons, and jets is required. The detector also includes a very accurate muon sub-detector. Such an excellent resolution is crucial to account for the missing energy of neutrinos which is also used in this project. ATLAS has multiple layered components to detect different particles:

- Inner Detector: it uses a 2 T solenoid magnetic field to bend the particles in the plane perpendicular to the beam axis (transverse). Here, the tracker is used to accurately measure the trajectory of charged particles.
- Calorimetry: there is an electromagnetic calorimeter and a hadronic one. The information they provide is crucial to measure the energy of electrons, photons, jets, and to calculate the missing energy in the transverse plane.
- Muon Spectrometer: located in the outer part of the ATLAS detector to absorb and measure energy and momentum of muons.

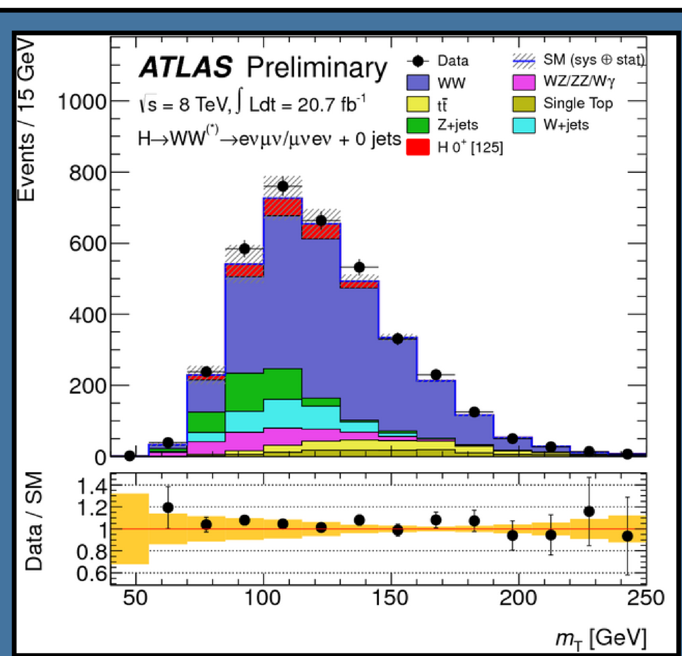


Final State Topology



The $H \longrightarrow W W$ channel signature is characterized by two oppositely charged leptons and their corresponding neutrinos. Only the eμ final state is considered here, with 0 jets, since it is the most sensitive.

As shown in the plot to the right, the main background is the continuous WW production [1]. This project focuses on the main background in the qq and gg channels. Other backgrounds of this final state include the top and anti-top quark, W+ jets, and diboson backgrounds.



Analysis

The aim of this project is to follow the ATLAS HWW BDT analysis and try to optimize training variables, pre-selection cuts and training parameters [1, 2]. While the standard analysis focuses on all backgrounds at once, this project aims to separate the signal ($gg \longrightarrow H \longrightarrow W W$) from the continuous WW background only, so differences can arise and be exploited.

• TMVA/BDTs

The Toolkit for Multivariate Analysis (TMVA) provides a ROOT-integrated environment for the processing, parallel evaluation and application of multivariate classification and regression techniques [3]. This project employs in particular, the decision tree, which is boosted in order to reduce the impact of statistical fluctuations. A boosted decision tree (BDT) is a progression of binary decisions for one variable at a time until a stop criterion is satisfied. The trees are finally combined into a single classifier which is given by a (weighted) average of the individual decision trees (training) [3].

• Cuts and Depth

Starting from the standard pre-selection cuts, made to optimize BDT efficiency before the training (GeV):

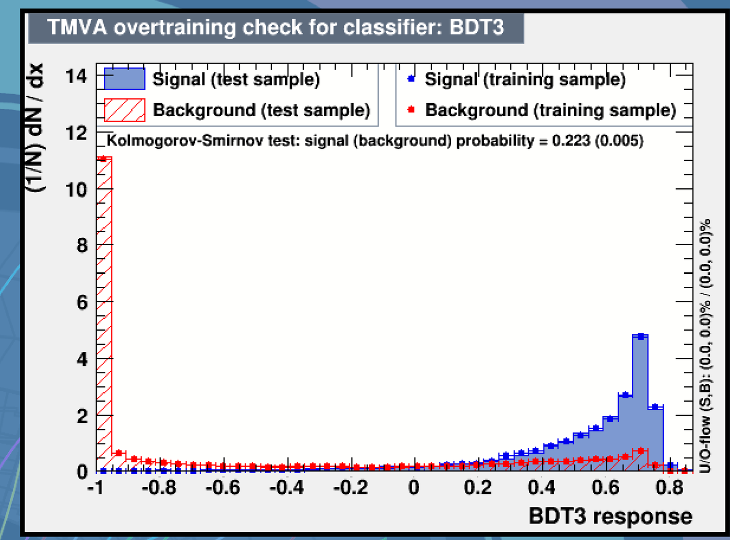
$P_{T,\ell\ell} > 20$, $m_{\ell\ell} > 10$, $P_{T,\ell 0} > 22$, $P_{T,\ell 1} > 15$, MET > 20

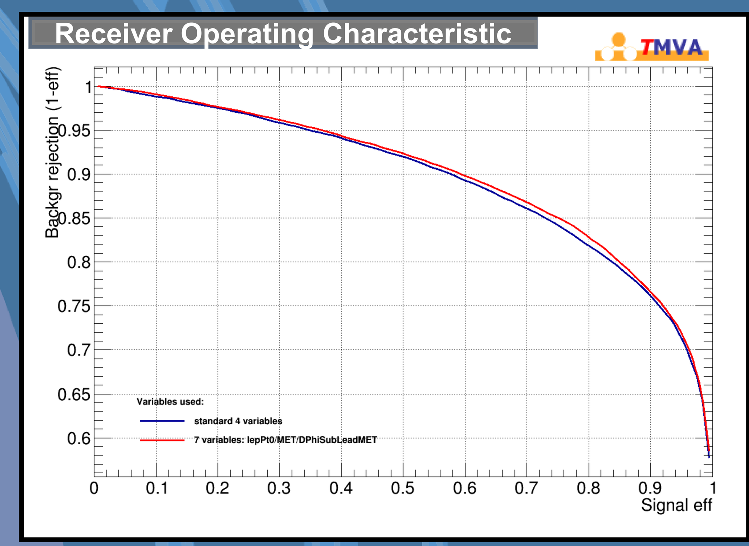
we tried to change all the thresholds in order to focus better on the signal and WW topologies [1, 2]. Multiple different cuts were attempted, but the standard pre-selection cuts produced the best results.

The depth is the number of times that there is splitting at a node during the training. A depth of 3 was used to prevent overtraining of the machine. Two different samples of signal and background were used and compared to ensure that the machine was not over trained. Essentially, if the machine is over trained it is too specific to that set and will be trained according to their statistical fluctuations. It was seen through trial that increasing the depth up to 6 did not lead to significant improvement relative to the risk of overtraining.

• BDT Output and Corresponding ROC Curves

Monte Carlo samples are used to simulate physics phenomena. Boosted decision tree training is a machine learning algorithm which learns how to discriminate between *signal* and *background* events using the variables provided. Different combinations of variables have been tested, ranging from 4 to 7. The weights produced during training can be used on another sample which is not already classified. The metric for optimization is the receiver operating characteristic (ROC curve), which represents the signal efficiency at a given background rejection. It is representative of all the possible value selections of BDT variable. By maximizing the area under the ROC Curve, we maximize the efficiency, which implies a higher separation power of the BDT.





Variables of Interest

The standard training is executed using 4 variables: m_T, $m_{\ell\ell}$, $P_{T,\ell\ell}$ and $\Delta\phi_{\ell\ell}$ [1, 2]. To this basic set, more variables were added and compared, in different combinations. If the number of variables used during training is reduced below 4, drastic decreases in separation power are observed. At most, 7 variables were used. Increasing the number of input variables further does not result in an increase in the separation power as it ignores non-discriminating variables. The following variables were found to perform the best:

⇒ **MET** - Missing Transverse Energy: the negative sum of the transverse momentum of all detected particles is equivalent to the sum of the transverse momentum of the neutrinos.

⇒ m_T - Transverse Mass: the final state mass can not be fully reconstructed due to the presence of the two neutrinos, so an approximation is used.

$$m_T = \sqrt{E^2_{T,\ell\ell+\nu\nu} - |\vec{p}_{T,\ell\ell+\nu\nu}|^2} = \sqrt{\left(E_{T,\ell\ell+\nu\nu} - E_T^{miss}\right)^2 - |\vec{p}_{T,\ell\ell+\nu\nu} + \vec{p}_T^{\,miss}|^2}$$

⇒ $P_{T,\ell\ell}$ – Transverse Momentum of the Two Leptons

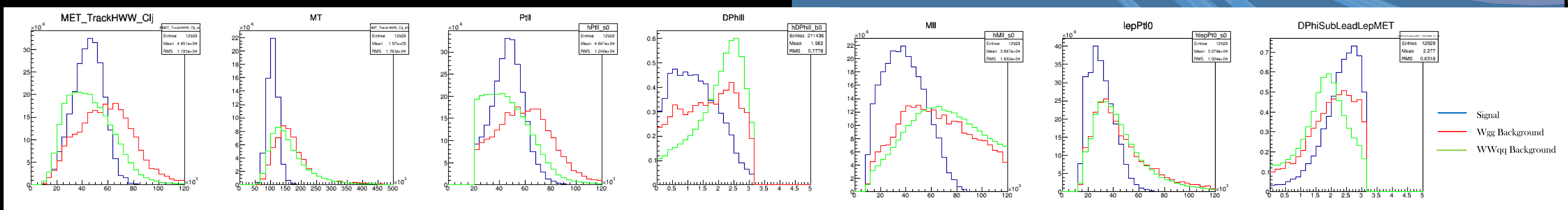
$$P_{T,\ell\ell} = \sqrt{(P_{x,e} + P_{x,\mu})^2 + (P_{y,e} + P_{y,\mu})^2}$$

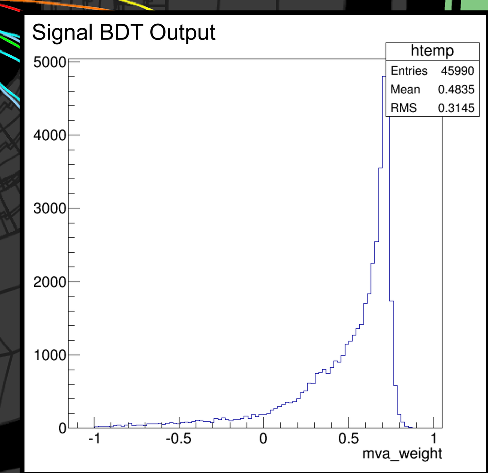
⇒ $\Delta\phi_{\ell\ell}$ - Azimuthal Angle Between the Two Leptons

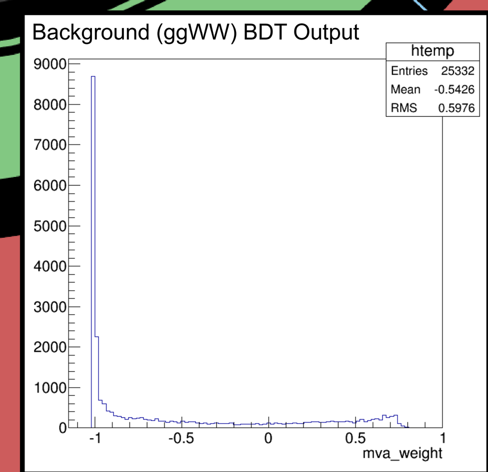
⇒ $m_{\ell\ell}$ - Invariant Mass of the Two Leptons

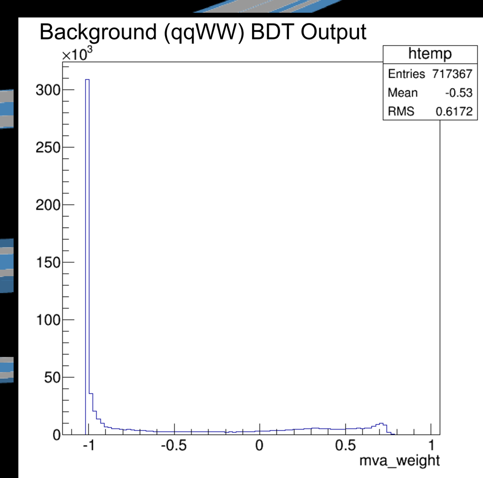
⇒ $P_{T,\ell 0}$ – Transverse Momentum of the leading Lepton

⇒ $\Delta\phi_{\ell 1,MET}$ - Angle Between the sub-leading Lepton and the Missing Transverse Energy









Results

During the training, the variables were ranked by variable importance. The order shown starts with largest separation power:

m_T, $\Delta\phi_{\ell\ell}$, $m_{\ell\ell}$, $\Delta\phi_{\ell 1,MET}$, MET, $P_{T,\ell\ell}$, $P_{T,\ell 0}$

Concerning pre-selection cuts, no change in thresholds has been found (so far) to perform better than the standard configuration.

The same can be said about training settings, showing that overall the standard ATLAS HWW BDT analysis is close to optimal [1, 2].

Once the machine was trained, the weights were used to classify a new sample. The BDT variable that was generated was then plotted (left) for the signal and both background samples without any region cuts. It is seen that the signal is concentrated close to 1, while the background is concentrated around -1. Sharp peaks show that the machine has done a good job at separating the two. It is observed that most of the signal sample was classified as signal, and most of the background was interpreted as background.

References

[1] The ATLAS Collaboration. Study of the spin properties of the Higgs-like particle in the $H \to WW^{(*)} \to e\nu\mu\nu$ channel with 21 fb^{-1} of $\sqrt{s}$ = 8 TeV data collected with the ATLAS detector. ATLAS-CONF-2013-031, 2013.

[2] A. Hoeker and others. TMVA 4: Toolkit for Multivariate Data Analysis with ROOT. arXiv:physics/0703039v5, 2009.

[3] The ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. Phys.Lett., B726:120–144, 2013.