Institut de Physique Nucléaire de Lyon
Internship carried out from 2018/03/12 to 2018/07/13

Master 2 internship report

Signal vs background discrimination in γ+jet events, recorded by the CMS experiment at LHC.

Author :
Maxime GIRAUD

Supervisor :
Viola SORDINI

Version 0.1 du
June 19, 2018

Introduction

First the CMS experiment at CERN will be described then the data that has been used will be described then the technics that has been used will be described then the exploitation will be described

γ+jet event classification in LHC collisions

1.1 CMS experiment at LHC

The Compact Muon Solenoid (CMS) is a particle physics detector built on the Large Hadron Collider (LHC) at CERN in switzerland and France. The goal of CMS experiment is to investigate the physics beyond the Standard Model. CMS is designed as a general-purpose detector, capable of studying many aspects of proton collisions at 0.9-13 TeV, the center-of-mass energy of the LHC particle accelerator.
It is made of multiple particle detectors designed to measure the energy and momentum of products of the collisions. The first layer called the "Tracker" reconstruct the paths of high-energy muons, electrons and hadrons as well as see tracks coming from the decay of very short-lived particles.
Next the "Electromagnetic Calorimeter" is designed to measure with high accuracy the energies of electrons and photons.
The Hadronic Calorimeter measures the energy of hadrons and provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos.
Theses layers all fit inside a large solenoid magnet of 3.8 Tesla, this allows the charge/mass ratio of particles to be determined from the curved track that they follow in the magnetic field. Finally the "Muon detectors and return yoke" are placed outside of the solenoid.

1.2 Hadronic jets in proton-proton collisions

In particle physics, jets are the experimental signatures of quarks and gluons produced in high-energy processes.
These particles having a net colour charge cannot exist freely due to couloour-confinement, thereby they are not directly observed in nature. Instead, they come together to form colour-neutral hadrons by a process called hadronisation that leads to a collimated spray of hadrons called a jet. The detailed understanding of both the jet energy scale and of the transverse momentum resolution is of crucial importance for many physics analyses.

Collision data

In this chapter will be described the various sources of data, and input variable that has been used during this work.

2.1 Monte-Carlo simulation

?

2.2 CMS data

Run 2 at $\sqrt{s} = 13TeV$ for an integrated luminosity of $36fb^{-1}$ number of events ? slide de hugues ?

2.3 MVA variables

In order to perform a multivariate analysis we used multiple variables representing various aspects of jets :

Isolation variables represent additional objects (photons,charged hadron and neutral hadron) reconstructed in a ΔR radius cone around the processed photon. These variables permit to discriminate between isolated prompt photons and neutral pions within a jet.

Charged Hadron isolation (CHiso) : $I_{cha} = \sum_{cha_i}^{\Delta R} p_{T,cha_i}$
cha_i corresponds to reconstructed charged hadron.

Neutral Hadron isolation (NHiso) : $I_{neu} = \sum_{neu_i}^{\Delta R} p_{T,neu_i}$
neu_i corresponds to reconstructed neutral hadron.

Photon isolation (Photoniso) : $I_{\gamma} = \sum_{\gamma_i}^{\Delta R} p_{T,\gamma_i}$
γ_i corresponds to reconstructed photons, the sum doesn't account for the p_T of the processed photon. (parler du pile-up avec ρ ?)

$\sigma_{i\eta i\eta}$: Energy weighted spread within the 5x5 crystal matrix centred on the crystal with the largest energy deposit in the supercluster. Obtained by measuring position by countinig crystals.

$\sigma_{i\eta i\varphi}$: Energy weighted spread within the 5x5 crystal matrix centred on the crystal with the largest energy deposit in the supercluster. Obtained by measuring position by countinig crystals.

η_{width} γ: Shower width in η

φ_{width} γ: Shower width in φ

R_9 γ: Energy sum of the 3x3 crystals centred on the most energetic crystal in the supercluster divided by the supercluster's energy. Lower values of R_9 for converted photons than those of unconverted photons.

Had/Em : Hadronic calorimeter energy deposit over Electromagnetic calorimeter energy deposit

E_{nxm}/E_{5x5} : Energy of most energetic nxm crystal set over energy of 5x5 crystal set

ρ: Pile-up energy, median of the transverse energy density per unit area.

Input variable analysis

A large set of variables is available from CMS data. MVA training can be time consumming and the "curse of dimensionality"(1) forces us to select only a few of them based on two main criteria :

Background vs Signal discrimination : Variables with most differences of shape for background and signal will be picked.

Low correlation between variables : Needed in order to reduce redundancy of input data and thus will permit to reduce MVA complexity (for example number of hidden neurons in ANN).

reference [Collaboration 2015].

3.1 Background vs Signal discrimination

It is necessary to pick the smallest set of input variable for the MVA. This selection is done by looking at variable shape for background and signal data from MC simulation. (fig. 3.1)

3.2 Variable correlations

Training data needed-quantity increases with network complexity. So correlation between variables must be avoided in order to get the minimum redundancy. By looking at the correlation matrix (fig. 3.2) we can see that CHiso is a good candidate and so will be used next for the background estimation.

3.3 Data driven background estimation

MVA will be performed with real data for the background, thereby a sideband (fig. 3.3) has to be defined on a low-correlated variable that won't be used in the MVA. MC simulations allow us to estimate the signal purity after applying a cut on CHiso.

Then we compare variables shape for background MC and DATA in the sideband region (fig. 3.4)

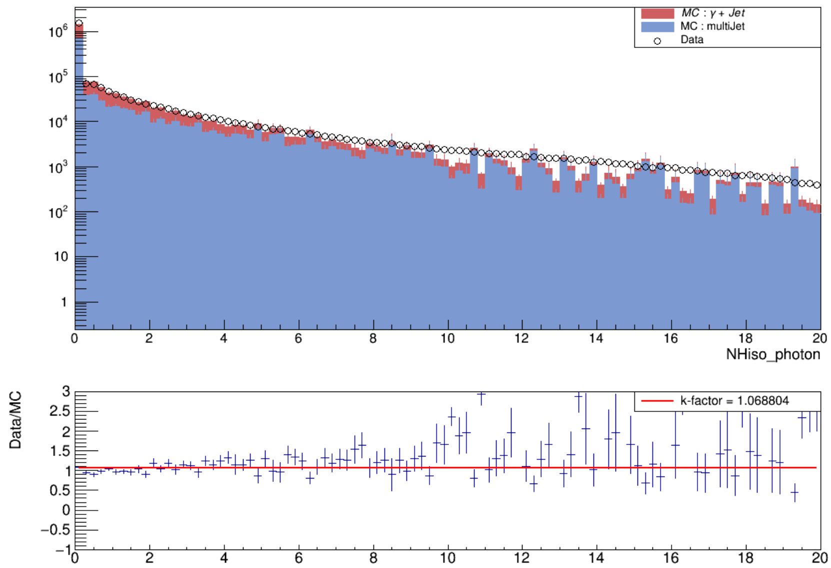


Figure 3.1: Neutral hadron isolation for background and signal MC

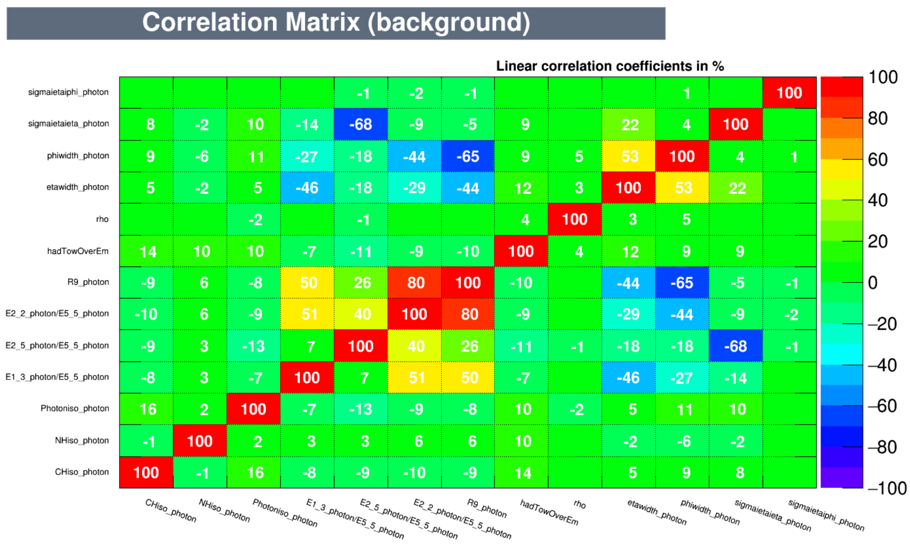


Figure 3.2: Correlation matrix for background MC

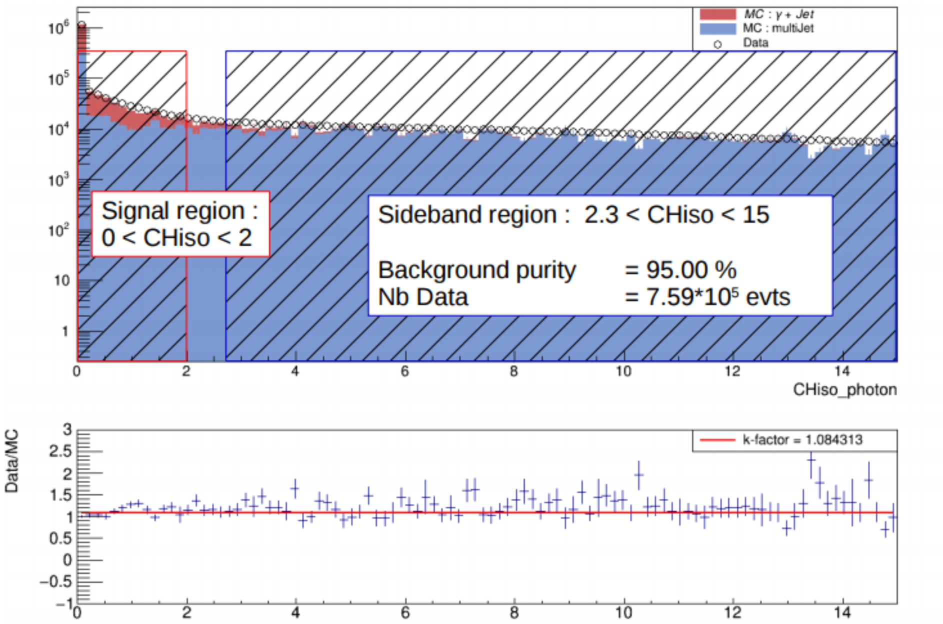


Figure 3.3: Charged hadron isolation for background and signal MC (histograms), signal region (red shaded area) and sideband (blue shaded area).

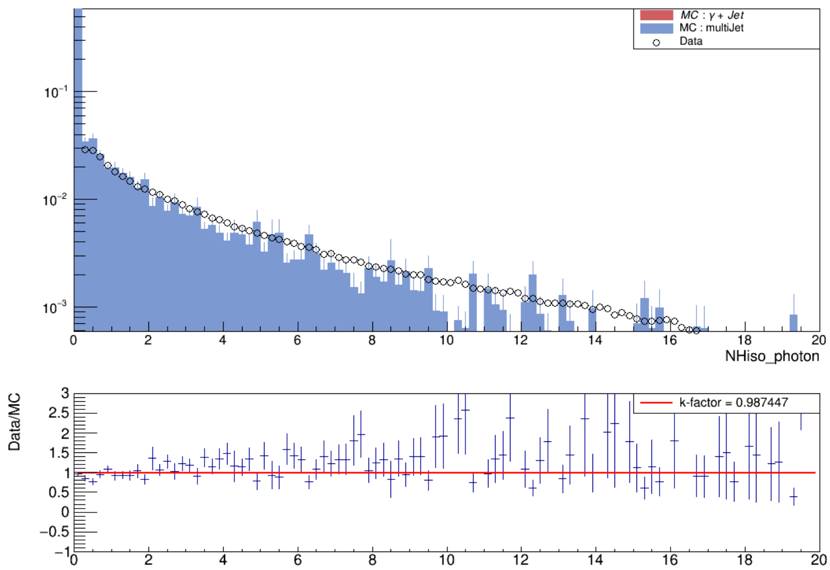
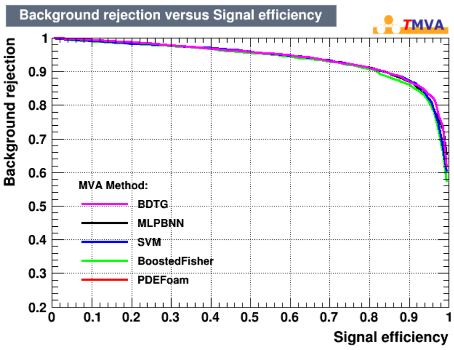


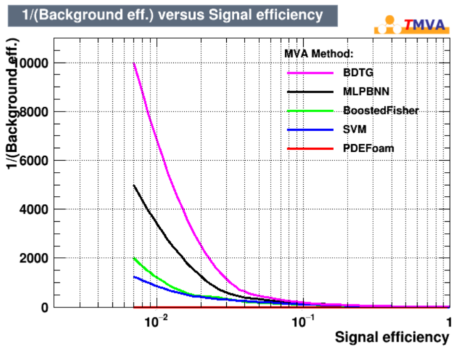
Figure 3.4: Neutral hadron isolation for background MC and DATA in the sideband region

MultiVariate Analysis

Now that we get background and signal samples we can perform the MVA for classification. For this work the TMVA framework from ROOT was used. Multiple MVA were tested (fig. 4.5a) with default configuration then the 2 bests were selected for the tuning of their parameters.



(a) ROC curve for the 5 bests MVA that has been tested.



(b) Inverse ROC curve for the 5 bests MVA that has been tested.

4.1 Artificial Neural Network

Input variable optimization

Signal extracted on DATA

5.1 PDF parametrization

5.2 Fit on Data

Conclusion and future outlook

Bibliography

[Collaboration 2015] CMS Collaboration. *Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV.* In JINST 10, 2015.

Résumé — Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Cras elementum ultrices diam. Maecenas ligula massa, varius a, semper congue, euismod non, mi. Proin porttitor, orci nec nonummy molestie, enim est eleifend mi, non fermentum diam nisl sit amet erat. Duis semper. Duis arcu massa, scelerisque vitae, consequat in, pretium a, enim. Pellentesque congue. Ut in risus volutpat libero pharetra tempor. Cras vestibulum bibendum augue. Praesent egestas leo in pede. Praesent blandit odio eu enim. Pellentesque sed dui ut augue blandit sodales. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam nibh. Mauris ac mauris sed pede pellentesque fermentum. Maecenas adipiscing ante non diam sodales hendrerit. Ut velit mauris, egestas sed, gravida nec, ornare ut, mi. Aenean ut orci vel massa suscipit pulvinar. Nulla sollicitudin. Fusce varius, ligula non tempus aliquam, nunc turpis ullamcorper nibh, in tempus sapien eros vitae ligula. Pellentesque rhoncus nunc et augue. Integer id felis.

Mots clés : Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor.