



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Física Armando Dias Tavares

Miquéias Melo de Almeida

**Measurement of Higgs Production Cross Section via Vector
Boson Fusion in $H \rightarrow ZZ \rightarrow 4l$ final state at 13 TeV using
Artificial Neural Networks**

Rio de Janeiro

2019

Miquéias Melo de Almeida

**Measurement of Higgs Production Cross Section via Vector Boson Fusion in
 $H \rightarrow ZZ \rightarrow 4l$ final state at 13 TeV using Artificial Neural Networks**



Orientador: Prof. Dr. Andre Sznajder
Coorientador: Prof. Dr. Nicola De Filippis

Rio de Janeiro
2019

Miquéias Melo de Almeida

**Measurement of Higgs Production Cross Section via Vector Boson Fusion in
 $H \rightarrow ZZ \rightarrow 4l$ final state at 13 TeV using Artificial Neural Networks**

Tese apresentada, como requisito parcial para obtenção do título de Doutor, ao Programa de Pós-Graduação em Física, da Universidade do Estado do Rio de Janeiro.

Aprovada em 11 de Abril de 2019.

Banca Examinadora:

Prof. Dr. Andre Sznajder (Orientador)
Universidade do Estado do Rio de Janeiro - UERJ

Prof. Dr. Nicola De Filippis (Coorientador)
Pilitecnico and Institute Nazionale di Fisica Nucleare sezione di Bari

Prof. Dr. Dilson de Jesus Damião
Universidade do Estado do Rio de Janeiro - UERJ

Prof^a. Dr^a. Helena Brandão Malbouisson
Universidade do Estado do Rio de Janeiro - UERJ

Prof. Dr. Luiz Martins Mundim Filho
Universidade do Estado do Rio de Janeiro - UERJ

Prof. Dr. Carsten Hensel
Centro Brasileiro de Pesquisas Físicas - CBPF

Prof. Dr. Leandro Salazar de Paula
Universidade Federal do Rio de Janeiro - UFRJ

Prof. Dr. Arthur Marques Moraes
Centro Brasileiro de Pesquisas Físicas - CBPF

Rio de Janeiro

2019

DEDICATION

I dedicate this work to my family, which is the foundation of my life.

ACKNOWLEDGMENTS

I would like to say thank you:

- to my family for being the foundation of my life. In special to my parents Valdir and Lucia for their dedicated love and unending example of life to me. If up to here I came it is because of their great support in several moments. In special I also thank you to my sister Queila, which more than a big sister, is the closest friend I have in life. And the last but not the least, I thank you my closest grandmother Carmelita that always teaches me about perseverance and goodness;
- to prof.(s) Andre Sznajder and Nicola De Filippis for their friendship and advising since the pledging of my master degree. Both of them have proportioned to me great experiences, learnings and growing which contributed significantly for this present work;
- to prof.(s) Sergo Jindariani and Luciano Ristori which I had the opportunity to met at Fermilab and learn many things during a very productive collaboration;
- to the Universidade do Estado do Rio de Janeiro (UERJ) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the support during this PhD which has allowed me to enhance my knowledge about Particle Physics;
- to the Fermi National Laboratory (Fermilab) and the US Department of Energy for the support given to me when I was working into the CMS L1TT project guided by the AM+FPGA Working Group at Fermilab;
- to the Università degli Studi di Bari Aldo Moro (UNIBA) and the Global Doc project via the Puglia region, which allowed me to have an interchange period closer to my co-advisor in Italy;
- to the workers in the several sections of the UERJ, from the cleaners to the rector, which are all essential for the proper functionality of the university;
- to everyone that in same level has contributed for the accomplishment of the present thesis or for moments of happiness, including friends and colleagues from the places I have gone. I don't quote any name in order to avoid the injustice of forgetting someone.

No one can balance the world
without balancing oneself.

Unknown author

ABSTRACT

This work presents an isolated measurement of the Higgs boson production cross section via Vector Boson Fusion (VBF) production mode, with the Higgs decaying through the $H \rightarrow ZZ \rightarrow 4l(l = e, \mu)$ channel. The study is performed using data samples corresponding to an integrated luminosity of 35.9 fb^{-1} from pp collisions at $\sqrt{s} = 13 \text{ TeV}$, which has been collected by the CMS experiment during 2016 at the LHC. A multivariate analysis is performed through the usage of Artificial Neural Networks (ANNs). Statistical shape analysis of ANNs developed for two orthogonal jet-based categories is done by combining the discriminants distribution from each category. The Higgs VBF signal strength modifier is measured to be $\mu_{qqH} = 1.28^{+1.24}_{-0.84}$ for an expected Higgs boson of $m_H = 125 \text{ GeV}$. This result is compatible with the SM expectation. The observed significance of the present analysis is $Z_{qqH}^{obs} = 1.9$, while the expected one is $Z_{qqH}^{exp} = 1.8$. The observed and expected 95%CL limits are estimated as $\mu_{qqH}^{obs} < 3.79$ and $\mu_{qqH}^{exp} < 1.66$, respectively. A projection for future luminosities is also presented and it is expected that the present analysis will have enough significance for the VBF Higgs production evidence (3.4σ) at 150 and the observance (5.1σ) at 359 fb^{-1} , respectively. Additionally, the present work brings on its appendixes full results obtained by the author during his collaboration in one of the CMS L1 Tacking Trigger approaches in 2015 and 2016, which has been leaded by Fermilab. The appendixes also contains a summary of results obtained by the author when working on an event tagging procedure called *FastME*, which is based on Monte Carlo events topology.

Keywords: Higgs. Vector Boson Fusion. Artificial Neural Network. CMS. L1 Tracking Trigger. Fast Matrix Element.

RESUMO

Este trabalho apresenta uma medida isolada da seção de choque de produção do boson de Higgs via Fusão de Bosons Vetoriais (VBF), com o Higgs decaindo pelo canal $H \rightarrow ZZ \rightarrow 4l(l = e, \mu)$. O estudo foi feito usando-se amostras de dados coletadas através do detetor CMS em colisões proton-proton com $\sqrt{s} = 13$ TeV realizadas no LHC em 2016 durante o RunII. Essas amostras correspondem à uma luminosidade integrada de 35.9 fb^{-1} . Uma análise de várias variáveis (*Multivariate Analysis - MVA*) é realizada através do uso de Redes Neurais Artificiais (ANNs) desenvolvidas para dois sub-conjuntos de dados categorizados de acordo ao número de jatos por evento selecionados na análise. Resultados finais são obtidos por meio de uma análise estatística da combinação das ANNs associadas à cada sub-conjunto. A intensidade do sinal de Higgs produzido via VBF é medida como $\mu_{qqH} = 1.28^{+1.24}_{-0.84}$ para um boson de Higgs de massa $m_H = 125\text{GeV}$, resultado este, que encontra-se em compatibilidade com o Modelo Padrão (SM) da Física de Partículas. A significância observada nesta análise é de $Z_{qqH}^{obs} = 1.9$, enquanto que, a esperada é de $Z_{qqH}^{exp} = 1.9$. Limites observado e esperado com 95% de confidência são obtidos como $\mu_{qqH}^{obs} < 3.79$ e $\mu_{qqH}^{exp} < 1.66$, respectivamente. Projeções para futuros cenários de luminosidade no LHC são também apresentados e indicam esta análise terá significância suficiente para evidenciar (3.4σ) e observar (5.1σ) a produção de Higgs via VBF com 150 e 359 fb^{-1} , respectivamente. Adicionalmente, este trabalho apresenta em seus apêndices resultados completos obtidos pelo autor durante sua colaboração em um dos projetos para criação do trigger de nível 1 para o detetor de trajetografia interna do CMS, o qual foi liderado pelo Fermilab. Os apêndices também contêm um sumário de resultados obtidos pelo autor sobre um procedimento de discriminação de eventos observados baseado na topologia de eventos de Monte Carlo.

Palavras-chave: Higgs. Fusão de Bosons Vetoriais. Redes Neurais Artificiais. CMS. L1 Tracking Trigger. Fast Matrix Element.

LIST OF FIGURES

Figure	1 - The $V(\phi)$ potential for different choices of μ^2	26
Figure	2 - The main Higgs (represented by the dotted line) production modes at the LHC.	32
Figure	3 - Higgs production XS and branching ratios.	33
Figure	4 - Feynman diagram for the Higgs production via vector boson fusion. The initial state quarks (or anti-quarks) are p_1 and p_2 , while the momentum at final state are p'_1 and p'_2 , respectively.	34
Figure	5 - Sketch of the Higgs VBF production description through the structure function approach.	36
Figure	6 - Main processes in this analysis: Higgs production modes (a) vector boson fusion (VBF), (b) gluon-fusion (ggH), (c) Higgsstrahlung (VH), (d) production associated to top quark, and the main SM backgrounds (e) and (f) ZZ. Additionally there is the $Z + X$ background which is derived via the Fake Rate <i>data-driven</i> method and will be presented later.	38
Figure	7 - Simplified transverse section of the LHC.	41
Figure	8 - Accelerators composing the LHC at CERN.	42
Figure	9 - General view of CMS layers.	44
Figure	10 - Transverse section view showing the detection that each sub-detector does.	45
Figure	11 - Scheme of the pixel detector.	46
Figure	12 - Illustration of a super-module and the entire ECAL.	48
Figure	13 - Scheme of the positions of the HB towers.	49
Figure	14 - Scheme of the hadronic calorimeter in the endcaps, annexed to the muon detection system.	50
Figure	15 - Scheme of the positions of HO in the muon detection system and its effect on the measurement of the energy from hadronic showers.	51
Figure	16 - An HF quadrant showing the modules of steel plates with grooves where optical fibers sit.	52
Figure	17 - The CMS magnetic system.	54
Figure	18 - Comparison of the resolution on the muon transverse momentum measurement obtained by using the muon chambers and the tracker separated and combined, left. Scheme of the muon chambers position in the CMS, right	55
Figure	19 - Types of detectors used in the muon detection system.	57

Figure 20 - Electron selection efficiencies measured using the tag-and-probe technique for: (a) non-gap electrons and (b) gap electrons.	65
Figure 21 - Muon reconstruction and identification efficiency for $p_T < 20\text{GeV}$ as a function of p_T in the barrel (a) and endcaps (b), and as a function of η for $p_T > 7\text{GeV}$ (c). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot of the two efficiencies, the black error bars represent the statistical uncertainty, the orange ones the systematical uncertainty, and the violet ones the total uncertainty.	68
Figure 22 - Efficiency of the muon impact parameter requirements as a function of p_T in the barrel (a) and endcaps (b), and as a function of η for $p_T > 20\text{GeV}$ (c). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot the black error bars are for the statistical uncertainty, the orange ones for the systematical uncertainty and the violet ones include both uncertainties.	68
Figure 23 - Efficiency of the muon isolation requirement as a function of p_T in the barrel (a) and endcaps (b). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot the black error bars are for the statistical uncertainty, the orange ones for the systematical uncertainty and the violet ones for the total uncertainty.	69
Figure 24 - Tracking efficiency in data and simulation as a function of η for muon $p_T < 10\text{GeV}$ (a) and $p_T > 10\text{GeV}$ (b) with ReReco data.	69
Figure 25 - Control plots of four-lepton invariant mass for: (a) events passing the SM Higgs selection, (b) events in the $m_{4l} = [118, 130]\text{ GeV}$ range and (c) events in the VBF signal region.	76
Figure 26 - (a) Graphical representation of a neuron. Each input x_i (from user or from the hidden neurons) is weighted by w_i and summed to a bias b . The sum is passed to a function (the activation function - neuron type) which outputs a vector of values (one for each input). (b) Graphical representation of a NN with three hidden layers. In this type of ANN the outputs of all neurons from a previous layer are fed into the neurons of a next layer.	78
Figure 27 - The flow of signal through the neuron j up to its instantaneous error e_j	80

Figure 28 - The flow of signal through a hidden neuron j up to its <i>signal function</i> $y_j(n)$ which is forwarded as an input to an neuron k in an output layer. Note that, the <i>activation function</i> can now be different between the layers and they are represented as f_j and f_k for the neurons j and k , respectively.	82
Figure 29 - Monte Carlo and Data distribution of the variables used as inputs to train the ANNs in this analysis.	84
Figure 28 - ROC curves comparing the performance of a trained NN, the VBF MELA discriminant and D_{jet} : (a) the performance of NN without scaling the <i>loss</i> and (b) performance after scaling the <i>loss</i>	87
Figure 29 - Example of checking plots produced during ANNs training. (a) $D_{qqH,2j}^{MELA}$ (for reference) and (b) ANN distributions, (c) $\epsilon_s.\pi$ vs ϵ , (d) signal significance and (e) ROC curves computed using all backgrounds together for training and testing sets. (f) ROC areas computed separately for each background using the testing set (the filled lines are for the ANN and the dotted ones are for $D_{qqH,2j}^{MELA}$).	90
Figure 30 - Summary of the results obtained with the parameters scanning. The plot shows the difference between trained ANNs and $D_{qqH,2j}^{MELA}$ in three metrics used at that moment: $\max(\epsilon_s.\pi)$, ROC area and significance . . .	91
Figure 31 - Performance of VBF discrimination against the backgrounds for the ANN, MELA and Djet discriminants. On (a) is the performance when considering only 2 jets per event and on (b) when considering up to 3 jets per event.	93
Figure 32 - Fake rates in terms of the lepton probe p_T and η . The p_T distributions (barrel and endcap) corresponds to the regions defined by $ \eta < 1.479$ (1.2) for electrons (muons). The background WZ from MC has been subtracted.	96
Figure 33 - Fake rates as a function of the probe lepton p_T and η . The background WZ from MC has been subtracted. This is the mapping for fake rates applied into the control regions.	97
Figure 34 - Four-lepton invariant mass distribution of the events selected in the 2P2F control region in the 13TeV dataset: (a) 4μ , (b) $4e$, (c) $2\mu2e$ and (d) $2e2\mu$ channels.	99
Figure 35 - Four-lepton invariant mass distribution of the events selected in the 3P1F control region in the 13TeV dataset: (a) 4μ , (b) $4e$, (c) $2\mu2e$ and (d) $2e2\mu$ channels.	100
Figure 36 - Four-lepton mass distribution of $Z+X$ for events after the (a) SM Higgs and (b) VBF-SR selections. The shapes are shown for each separate channel and for their combination into $4l$ final state.	101

Figure 37 - Fake rate, for each reducible background (DY , $t\bar{t}$, WZ, ZZ, $Z\gamma$), the average and the re-weighted distributions for electron (a) p_T barrel, (b) p_T endcap and (c) η and for muons (d) p_T barrel, (e) p_T endcap and (f) η	103
Figure 38 - Fake rate versus p_T and η for the average and the re-weighted distributions for electron (a) and (c) and, for muon (b) and (d), respectively.	105
Figure 39 - Effect of systematic uncertainties in the shape of one of the trained NNs for each of the processes and lepton channels. In this plot the histograms show the nominal distribution of the NN and the error bars indicate the maximum yield oscillation (up/down) per bin due to the systematic uncertainties of the inputs.	110
Figure 40 - NN distributions and their ratio in the Njets3 category from VBF-H2J and VBF-H3J for (a) 4μ , (b) $4e$ and (c) $2e2\mu$	112
Figure 41 - Final four-lepton mass distributions for the SM Higgs separated in the channels (a) 4μ , (b) $4e$ and (c) $2e2\mu$ and, combined into the (d) $4l$ final state. The distributions are show for $70 < m_{4l} < 400$ GeV and include the $Z + X$ background estimated by the data-driven method explained in Sec. 8.4.	115
Figure 42 - Final four-lepton mass distributions for the events passing the (a) SM Higgs and the (b) VBF-SR selections combined into the $4l$ final state. The distributions are show for $118 < m_{4l} < 130$ GeV (SM Higgs signal region defined by CMS Collaboration) and include the $Z + X$ background estimated by the data-driven method explained in Sec. 8.4.	116
Figure 43 - Pre-fit distributions of the neural network developed for (a) Njets2 and (b) Njets3 categories. In (c) is the distribution resulting from the combination of the two categories.	117
Figure 44 - The best fit value found, given the observed data, for μ_{qqH} . In (a) is shown the compatibility between the channels from each category of number of jets and in (b) is shown the likelihood ratios profile of the two categories and their combination.	120
Figure 45 - Post-fit NN distribution. The fit is done with the assumption of S+B hypothesis ($\mu_{VBF} = 1$ a priori).	122
Figure 46 - Post-fit nuisances impact plot obtained with the combination of the two categories of number of jets. This plot shows the effect of each systematic uncertainty on the estimation of the VBF signal strength (μ_{qqH}).	123

Figure 47 - (a) Observed and expected 95% CL upper limit on the ratio of the Higgs VBF production cross section to the SM expectation, for the two jet-based categories and their combination. The expected 1σ (2σ) ranges of the expectation for the background-only hypothesis are also shown by the green (yellow) bands. (b) The observed significance of the local excess with respect to the SM background expectation for the two categories and their combination.	126
Figure 48 - Comparison of the expected VBF Higgs signal strength best fits, and their uncertainties, between the present luminosity and the future 150fb^{-1}	127
Figure 49 - Event display of the most VBF-like event ($H \rightarrow ZZ \rightarrow 2e2\mu$) according to the ANN in the Njets2 category (ANN score = 0.93). As expected, the two jets present in the event have a very large η separation and no particle activity between them.	138
Figure 50 - Event display of the most background-like event ($H \rightarrow ZZ \rightarrow 4e$) according to the ANN in the Njets2 category (ANN score = 0.04). Differing from the most VBF-like event, this one has the two jets very close in the $\eta - \phi$ space.	139
Figure 51 - Event display of the most VBF-like event ($H \rightarrow ZZ \rightarrow 4\mu$) according to the ANN in the Njets3 category (ANN score = 0.77). By comparison, the η gap between the jets and the particle activity is more evident than what is observed in the Njets3 category for background (see next figure).	140
Figure 52 - Event display of the most background-like event ($H \rightarrow ZZ \rightarrow 2e2\mu$) according to the ANN in the Njets3 category (ANN score = 0.15).	141
Figure 53 - The architecture of the neural network created in this analysis for Njets2 category. The black dots represent the inputs and the neurons, while the lines represent the size of the parameters found after training those NNs. Wider and brighter lines means that the parameter (weight or bias) associated to a given input for a neuron is larger (in other words its contribution is more relevant). This ANN has 21:13:8 hidden neurons of SeLU type.	142
Figure 54 - The architecture of the neural network created in this analysis for Njets3 category. This ANN has 11:9:7 hidden neurons of SeLU type.	143
Figure 55 - The significance (computed as S/\sqrt{B}) (a) and the efficiency times purity ($\pi.\epsilon$) (b) achieved when cutting at the NN for each jet-based category. The cut applied corresponds to the low edge value of a given bin in the histogram. Note that the significance computed here does not take into account the uncertainties.	144

Figure 56 - The behavior of the NN neuron output for the (a) first hidden layer, (b) second hidden layer and (c) output layer. The subplots in (a), (b) and (c) show, from left to right and sorted, the p_T 's (top row), the η 's (middle row) and the ϕ 's (bottom row) of the four leptons and the three jets.	146
Figure 57 - The calendar for the upgrades of the LHC and its detectors.	148
Figure 58 - Scheme of the sensors in a layer of the CMS detector after the phase II upgrade. An object called <i>stub</i> is defined in terms of the distance between the centroids of back-to-back sensors (or cluster of them) activated when a particle cross the layer.	149
Figure 59 - Scheme of the partition adopted in the CMSL1TT AM+FPGA for the cluster of strips activated in the silicon sensors in the detector layers. .	150
Figure 60 - Road sorting effect on efficiency without truncation and with truncation in 200 roads and 500 combinations for $t\bar{t} + PU200$	152
Figure 61 - How the ΔS mitigate AM patterns fired by random tracks: two random tracks trigger a valid pattern and the stubs are stored for further analysis but, the stubs combination doesn't come from a real track. . .	153
Figure 62 - Stub bending distribution per detector layer for $t\bar{t} + PU200$ and the ΔS ranges for the configuration ASYM115577. In the ASYM mode only the central range can have width defined by user and ΔS always is divided only into 3 ranges.	155
Figure 63 - Stub bending distribution per detector layer for $t\bar{t} + PU200$ and the ΔS ranges for the configuration SYM115577. In the SYM mode ranges have the same width, except the outer ranges (those ones comprehending $ \Delta S = 6.5$).	156
Figure 64 - Stub bending distribution for $jet(p_T = 250GeV) + PU200$. Different events seems to have the same limits but, as one can see the occupancy is no equal (comparing with the pictures for $t\bar{t} + PU200$).	157
Figure 65 - Roads and combinations average and 95 percentile per event versus different ΔS configurations at $sf1_nz1$ for $t\bar{t} + PU200$ (left) and $jet(p_T = 250GeV) + PU200$ (right).	158
Figure 66 - Road efficiency versus different ΔS configurations at $sf1_nz1$ for $t\bar{t} + PU200$ (left) and $jet(p_T = 250GeV) + PU200$ (right).	159
Figure 67 - Stub fountain scaler factor scan for $t\bar{t} + PU200$	159
Figure 68 - Muon road efficiency versus pile up. The plot (a) is the result without ΔS approach (done by another person in the CMS L1TT working group) and the plot (b) with it (done by the author).	160
Figure 69 - Track fitter χ^2 from all muon tracks in $t\bar{t} + PU200$ versus track p_T . .	162

Figure 70 - TF χ^2 distribution for 6/6 muon tracks in each track p_T range from $t\bar{t} + PU200$	163
Figure 71 - TF χ^2 distribution for 5/6 muon tracks in each track p_T range from $t\bar{t} + PU200$. Vertical bars point the cuts defined.	164
Figure 72 - Track parameters resolution as function of q/p_T for single pion no PU (first column), single muon no PU (second column) and single electron no PU (last column). From top to bottom the track parameters are q/p_T , ϕ_0 , z_0 and $\cot \theta$	167
Figure 73 - χ^2_{match} distribution from single pion and single muon event without and with PU.	168
Figure 74 - AM track characterization and synthetic efficiency from single muon no PU events.	169
Figure 75 - AM tracks categories fraction and truth stub fraction versus χ^2_{match} cut for $\mu + PU200$ on left column and for $jet(p_T = 250GeV) + PU200$ on right column.	171
Figure 76 - Track reconstruction efficiency for $t\bar{t} + PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplicate removal was applied by requiring DR=0.	175
Figure 77 - Track reconstruction efficiency for $jets(p_T = 250GeV) + PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.	176
Figure 78 - Track reconstruction efficiency for $\mu+PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.	177
Figure 79 - Track reconstruction efficiency for $\mu+PU300$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.	178
Figure 80 - Track reconstruction efficiency for $\mu + PU400$ sample. The pattern bank used had 64k and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.	179
Figure 81 - Comparison between the pattern bank $sf1_nz1$ and its variants due to the stub bending approach.	180
Figure 82 - Truncation is not a linear behavior for ΔS . Increasing combination is better than increasing road truncation limit. Increasing combination truncation limit to 1k increases 3% almost equally for 200/400/800 roads.	182

Figure 83 - Overview of the FNAL AM CMSL1TT hardware: (a) Pulsar-IIB board, (b) inspection of Pulsar-IIB components, (c) and (d) the setup for measure data transmission quality from the PRMs via optical cables, (e) crate containing Pulsar-IIB boards and PRMs (not visible since Pulsar boards sit right besides each other) with full setup ready for demonstration of the approach.	184
Figure 84 - Illustration of the analysis process done by the <i>FastME</i> . The MC events present a topology associated to the EM of a given physical process, such as, each particle has a correlation with the others particles in the event. An data event (black points) receives a probability of being from a kind or other (blue and red point) via the correlation of the distances (represented by the blue and red circles) between it and the MC events.	188
Figure 85 - Comparison between the <i>FastME</i> discriminants distributions as defined by Eq. 80 (a) and 81 (b), obtained samples of $gg \rightarrow H \rightarrow 2e2\mu$ and $qq \rightarrow 2e2\mu$ (ZZ) generated with Powheg.	190
Figure 86 - Classification dependency for signal and background events accordingly to the size of the pattern banks used. On graphs (a) and (b) the size of the background pattern bank has a fixed size while the signal one is varied. On graphs (c) and (d) the opposite case is shown. Note, here "purity" is computed using the absolute number of events (without normalization).	191
Figure 87 - Variação da fração de signal e <i>background</i> identificados na amostra com 50% de cada classe.	192
Figure 88 - (a) Scan of the ideal cut (to correct classify signal and background events) in function of the ratio between the size of the signal and back- ground pattern banks. (b) The effect on the fraction of signal and background events being classified as so, before and after the correc- tion on the discriminant.	193
Figure 89 - Impact on the D (Eq. 79) distribution for signal and background due to use or not of ϕ in the computation of the distances between particles and events.	194
Figure 90 - Dependency on the variation of the fraction of signal events being mis- classified as background in function of the δp_T scale factor (the σ_{v_k} in Eq. 78). A fixed value of 5.0 was used as the $\delta\eta$ scale factor.	195

Figure 91 - (a) Observed events classified as signal and background by the <i>FastME</i> , (b) the discriminant distribution (from observed data), (c) and (d) distribution of <i>FastME</i> and <i>MELA</i> discriminants as obtained from observed data and MC events, (e) and (f) the distribution of those discriminants versus the m_{4l} mass value of the analyzed events. The distributions are not normalized by the events weight. The distribution of <i>FastME</i> discriminant in (b) looks different from what is shown in (c) and in Fig. 89 because that was the original shape. Later it was decided to re-define the discriminant as expressed by Eq. 80 in order to keep it in the range [0, 1].	197
Figure 92 - View of the <i>FastME</i> repository on <i>Github</i> created by the author.	198
Figure 93 - <i>FastME</i> input card. This card is parsed to the <i>FastME</i> executable in order to set up needed parameters in order to run an analysis.	200
Figure 94 - Comparison between the two methods developed to define the classifi- cation of a event probe. On (a) the discriminant distributions for the minimum distance matching and on (b) its performance.	201
Figure 95 - Variation on the <i>FastME</i> performance due the introduction of extra objects with random properties in the signal and background testing samples. In (a) the variation on the signal and background distributions for the discriminant (based on distance) and in (b) the variation on the ROC curve.	203
Figure 96 - <i>FastME</i> performance on the discrimination of qqH against ggH using the discriminant based on the distance (left). The area under the ROC curve (right) corresponds to $\sim 66\%$	204
Figure 97 - Some pictures of the seminar given by the author at UNIBA during his participation in the Global Doc project. The organizers, the local tutor prof. Giovanna Selvagi and the co-advisor prof. Nicola were pre- sent in the seminar, which was also opened to the university academic community.	206
Figure 98 - The RECAS computing farm at the UNIBA campus where the Physics department is located.	207
Figure 99 - Briefly of the Dark Matter analysis: (a) the LO diagrams of the sear- ched processes and (b) the MET distribution (stacked) for SM processes and (superimposed) Dark Matter candidates.	208

LISTA OF TABLES

Table	1 - Elementary particles and their interactions in the SM.	24
Table	2 - Some parameters of the LHC.	40
Table	3 - Comparison between the energy and velocity achieved by a proton in the different accelerators located at CERN.	41
Table	4 - Datasets used in the analysis.	59
Table	5 - Trigger paths used in 2016 collision data.	60
Table	6 - MC simulated samples and their respective cross section.	61
Table	7 - Overview of input variables to the electron identification classifier. Va- riables not used in the RunI MVA are marked with *.	63
Table	8 - Minimum BDT score required for passing the electron identification. .	63
Table	9 - The requirements for a muon to pass the tracker high- p_T ID. Note that these are equivalent to the Muon POG high- p_T ID with the global track requirements removed.	66
Table	10 - Filters from JETMET POG used to improve signal-to-noise ratio. . .	72
Table	11 - Number of expected events for background and signal, with statisti- cal uncertainty reported, and number of observed events after the SM Higgs selections.	75
Table	12 - Number of expected events for background and signal and number of observed events after the VBF-SR selections.	75
Table	13 - Summary of the scanned parameters during ANN trainings.	89
Table	14 - Fraction of events, after the VBF-SR selections, containing the jet_i (i-th jet).	92
Table	15 - Yields estimated for Z+X in the signal region from the measurements on data using the fake rate method with opposite-sign (OS) leptons. Estimates reported with statistical uncertainties comparing Higgs sig- nal SR and after applying the cuts defining the VBF-SR.	101
Table	16 - The fake ratios for individual background processes, the average fake ratio and the fake ratio reweighed according to the composition of back- grounds in 2P2F control region.	102

Table 17 - Final yields estimated for Z+X in the signal region from the measurements on data using the fake rate methods of opposite-sign (OS-OS) and same-sign (OS-SS) leptons. The estimates are reported with the total uncertainty for each final state after the SM Higgs and VBF-SR selections. The m_{4l}^{OS}/m_{4l}^{SS} is the ratio between the m_{4l} distribution for each final state obtained for Z+X in the OS and SS procedures (it is not computed for VBF-SR due to the very few Z+X events estimated there).	104
Table 18 - Summary of experimental systematic uncertainties accounted into this analysis.	107
Table 19 - Summary of theoretical systematic uncertainties accounted into this analysis.	108
Table 20 - Number of expected events from background and signal, with total (statistical+systematic) uncertainty reported, and number of observed events after the SM Higgs selections in the mass range $m_{4l} > 70$ GeV.	114
Table 21 - Number of expected events from background and signal, with total uncertainty (statistical+systematic), and the number of observed events after the VBF-SR selections (see Sec. 6.3).	114
Table 22 - Expected and observed signal strength modifiers from each category and four-lepton channel for 35.9fb^{-1} of observed data at $\sqrt{s} = 13\text{TeV}$	121
Table 23 - Background and signal estimations, with total uncertainty (statistical+systematic), derived from fitting the s+b model to the observed data, accounting 35.9fb^{-1} at $\sqrt{s} = 13\text{TeV}$	121
Table 24 - Comparison of the expected VBF significance between the present luminosity (35.9fb^{-1}) and future scenarios at the LHC.	127
Table 25 - Stub bending possible divisions. The SYM method uses such table to decide how to split the ΔS values based on the number of ranges requested by user. The negative related of the ranges were omitted in the table.	157
Table 26 - Track resolution function parameters from single pion no PU.	166
Table 27 - AM track categories per event versus DR configuration for $\mu + PU200$, $t\bar{t} + PU200$ and $jet(p_T = 250\text{GeV}) + PU200$	173
Table 28 - AM pattern banks size for different configurations and coverages. The configurations stands for the way the CMS detector is partitioned in ϕ and z directions. The coverage stand for the fraction of track reconstructed over the total number of track patterns in a given configuration.	180

Table 29 - AM pattern banks size for different configurations and coverages. The configurations stands for the way the CMS detector is partitioned in ϕ and \mathbf{z} directions. The coverage stand for the fraction of track reconstructed over the total number of track patterns in a given configuration. The meaning of ASYM and SYM nomenclature have been introduced in Sec.E.2.1.3.	181
Table 30 - Time to compute the weight of one event using <i>MadWeight5</i> . For an usual analysis these numbers multiplies by thousand.	186

SUMMARY

1	INTRODUCTION	22
2	THE STANDARD MODEL AND THE HIGGS BOSON	23
2.1	The Standard Model Structure	23
2.2	Spontaneous Symmetry Breaking	25
2.2.1	<u>The Brout-Englert-Higgs Mechanism</u>	27
2.2.2	<u>The Higgs Mechanism and Electroweak Symmetry Breaking</u>	29
2.3	The Higgs Boson at the LHC	31
2.4	The Higgs VBF Production Mode	33
2.4.1	<u>The Higgs VBF LO Cross Section and its Kinematics</u>	34
2.4.1.1	Corrections of Higher Orders	36
2.5	The Physics Scenario of this Analysis	37
3	THE LHC AND THE CMS EXPERIMENT	39
3.1	The Accelerators at CERN	39
3.2	The CMS Experiment at the LHC	42
3.2.1	<u>The Tracker System</u>	43
3.2.1.1	The Silicon Pixels Detector	43
3.2.1.2	The Silicon Strips Detector	46
3.2.2	<u>The Calorimeters</u>	47
3.2.2.1	The Electromagnetic Calorimeter (ECAL)	47
3.2.2.2	The Hadronic Calorimeter (HCAL)	48
3.2.3	<u>The Magnetic System</u>	52
3.2.4	<u>The Muon Chambers</u>	53
3.2.5	<u>The Trigger and Data Acquisition System</u>	56
4	DATA AND MONTE CARLO SAMPLES	58
4.1	Data	58
4.2	Simulated Samples	58
5	PHYSICS OBJECTS SELECTIONS	62
5.1	Electrons	62
5.1.1	<u>Electron Efficiency Measurements</u>	64
5.2	Muons	65
5.2.1	<u>Muon Efficiency Measurements</u>	67
5.3	Photons and FSR Recovery	70
5.4	Jets	70
5.5	Missing Transverse Energy	71
6	EVENT SELECTION	73
6.1	Preselection	73

6.2	Selection of ZZ System	73
6.3	VBF Signal Region (VBF-SR)	74
7	NEURAL NETWORKS	77
7.1	The Learning Process	79
7.2	Developing and Training the Neural Networks	83
7.3	Preparation of the Datasets	83
7.4	Scaling Events Contribution in the Training	86
7.5	The Training Procedure	87
7.6	NNs Training Performances	89
7.6.1	<u>The Impact of the VBF 3rd Jet</u>	92
8	BACKGROUND ESTIMATION	94
8.1	Estimation of the SM Higgs Production Modes	94
8.2	Modeling of $q\bar{q} \rightarrow ZZ$	94
8.3	Modeling of $gg \rightarrow ZZ$	94
8.4	Estimation of Z+X Background	95
8.4.1	<u>Measuring the Fake Rates</u>	95
8.4.2	<u>Building Control Regions</u>	97
8.4.3	<u>Uncertainties on the Estimation of Z+X</u>	102
8.4.4	<u>Same-Sign Cross-checking and Final Z+X Estimation</u>	104
8.4.5	<u>Z+X Neural Network Output Shape</u>	106
9	SYSTEMATIC UNCERTAINTIES	107
9.1	Experimental Uncertainties	107
9.2	Theoretical Uncertainties	108
9.3	Discriminants Shape Uncertainties	108
9.3.1	<u>Systematic Uncertainty on the 3rd Jet</u>	111
10	RESULTS	113
10.1	Yields and Distributions	113
10.2	Statistical Analysis	113
10.3	Projection for Future Luminosity Scenarios	125
11	CONCLUSIONS	128
	REFERÊNCIAS	129
A	DISPLAY OF VBF AND BACKGROUND-LIKE EVENTS	138
B	FINAL ANN ARCHITECTURES	142
C	ANNS EXPECTED S/\sqrt{B} AND $\pi \cdot S/(S+B)$	144
D	THE ANN INTERNAL REPRESENTATION SPACE	145
E	THE CMS L1 TRACKING TRIGGER (CMSL1TT) PROJECT	147
F	THE FAST MATRIX ELEMENT (<i>FASTME</i>)	185
G	PARTICIPATION AT THE GLOBAL DOC PROJECT	205

1 INTRODUCTION

The Standard Model (SM) of Particle Physics is an effective theory supported by several experimental observations. The SM gathers quantum field theories which describes the elementary particles building the ordinary matter and how they interact via three mechanisms: the strong nuclear, the weak nuclear and the electromagnetic interactions. The strong and weak interactions are constrained to the atomic nuclei radius (10^{-14} m) and are the responsible for keeping protons together and nuclei decaying processes, for instance. The electromagnetic interaction is the reason why opposite charged particles attract each other, for instance. Each of these interactions is mediated by particles called bosons. The strong nuclear interaction is mediated by gluons (g) which are particles without mass and that can interact with themselves. The weak nuclear interaction is mediated by massive bosons Z and W^\pm . And the electromagnetic interaction is mediated by the photon (γ) which does not have electric charge and neither mass.

One of the biggest mysteries of the SM is the mass of the elementary particles. According to the theory, for keeping certain symmetry conditions (which are needed since Physics must be invariant under frame variation) the elementary particles has been initially treated as non-massive objects. Around 1964, though, an idea for spontaneously generate the particles mass in the theory without breaking its symmetry was conceived. Nowadays it is known as *Higgs mechanism*. This mechanism is associated to a new boson (Higgs), which was the last missing piece on the SM basis for many years. The search for it was one of the motivations for the building of the LHC (*Large Hadron Collider*). Finally, in 2012 a SM Higgs boson-like was observed by ATLAS and CMS collaborations, opening a new chapter to elucidate the origin of particles' mass.

After the Higgs discover several measurements of its properties have been done and so far they shown good compatibility with SM expectations. However, there is still room for beyond SM effects in the Higgs sector. An interesting channel is the Higgs production trough Vector Boson Fusion (VBF). The VBF is the second most important Higgs production mode and presents a physic process clean of beyond SM effects. Hence it is a interesting channel to check the expectation from the SM. The Higgs produced via VBF followed by the presence of two energetic jets separated by a η gap, which is often used as a tagging input. In this context, this analysis presents the results of combining the power of Artificial Neural Networks (ANNs) and the extra information in the events given by a 3^{rd} jet passing some required selections. An isolated measurement on the Higgs VBF signal strength is then made on dedicated subsets of events, selected in a defined signal region and orthogonally divided into two jet-based categories.

2 THE STANDARD MODEL AND THE HIGGS BOSON

2.1 The Standard Model Structure

The Standard Model (SM) is a set of quantum field theories, locally gauge invariants, that describes the interactions¹ between the fundamental particles building the known matter. Such interactions arises by requiring the gauge symmetry on the lagrangians of the theories contained into the SM. These theories are the Quantum Electrodynamic (QED), the Glashow-Weinberg-Salam (GWS - electroweak interaction) and the Quantum Chromodynamic (QCD). The QED is based on the symmetry group $U(1)$ and describes the electromagnetic interactions between charged particles. The GWS is the theory describing the electroweak processes, such as the particles decay, and is based on the symmetry group $SU(2)_L \times U(1)_Y$. The QCD is structured on the symmetry group $SU(3)_C$ and describes the strong interactions, which is responsible, for instance, to keep the atomic nucleon stable [1, 2].

Based on the fact that all the known matter is constituted of elementary particles which interacts in different ways between them, the SM organizes such particles in several classes, with the most elementary ones being: quarks, leptons and bosons. The quarks do not appear alone in nature as far it is known. These particles are always observed in pairs or triplets forming bigger particles classified as mesons (pair quark-antiquark) and barions (triplets of quarks and antiquarks). The barions and mesons are grouped into another class called hadrons. Examples of mesons are the pions and kaons produced in collision experiments and in cosmic rays and, examples of barions are the protons and neutrons.

The leptons and quarks are together classified as fermions because they have fractioned quantum number of spin and are divided into six flavors organized in doublets in the SM. Such doublets are often identified in three generations. The ordinary matter is composed by the quarks of first generation while the other ones appear in particles produced in accelerators or cosmic rays, for instance. The leptons can interact weak and electromagnetically (excepting the neutrinos which only interact weakly), while the quarks can interact through any of the three SM interactions (since they have electric and color charge). The color charge is another quantum number that has been introduced originally in order to fit the existence of particles composed by three quarks of same flavor (the case of Ω^-). Without this consideration the Pauli exclusion principle avoids the existence of such particle. There are three color charges for each quark and often they are referred as

¹ Except the gravity which is not described by the SM currently.

Table 1 - Elementary particles and their interactions in the SM.

Generation	Flavor	Charge [e]	Mass [MeV]	Lifetime [s]
Leptons				
1 ^a	e (electron)	-1	0.51	1.4×10^{34}
	ν_e (neutrino do e)	0	≈ 0	-
2 ^a	μ (muon)	-1	105.66	2.2×10^{-6}
	ν_μ (μ neutrino)	0	≈ 0	-
3 ^a	τ (tau)	-1	1776.82	2.9×10^{-13}
	ν_τ (τ neutrino)	0	≈ 0	-
Quarks				
1 ^a	u (up)	2/3	2.30	-
	d (down)	-1/3	4.80	-
2 ^a	c (charm)	2/3	1275.00	-
	s (strange)	-1/3	95.00	-
3 ^a	t (top)	2/3	173500.00	-
	b (bottom)	-1/3	4180.00	-
Bosons				
Interaction	Associated Boson	Charge [e]	Mass [GeV]	Coupling Strength
Nuclear Strong	g (8 gluons)	0	0	1
Nuclear Weak	W^\pm	± 1	80.38	10^{-6}
	Z	0	91.19	
Electromagnetic	γ (photon)	0	0	10^{-2}

Source: Adapted from [5], [6], p. XII and [7], p. 26.

red, yellow and blue, in analogy to the primary colors which combined produce the white (no color), similar to the hadrons that do not have color charge [6, 7].

The bosons are the interactions mediators between the elementary particles and present integer quantum numbers of spin. The mediator of the strong interaction are the non-massive gluons (g), which interacts with color charged particles and themselves, exiting in 8 types. The electromagnetic interaction is mediated by the non-massive photon (γ), which does not have electric charge and thus do not self interact. Associated to the weak interaction there are three massive bosons, two charged (W^\pm) and one neutral (Z). The Tab. 1 resumes the properties of the elementary particles and their interactions described by the SM.

Although the SM has shown great success in the description of the elementary particles and their interactions, the method of introducing such interaction by requiring the gauge symmetries created a divergence with the observations. The gauge symmetry does not allow mass terms in the SM theory. However, since Glashow, proposing the $SU(2)$ structure for the electroweak interactions, it was known that the boson W has mass. This inconsistency was solved by so called Higgs mechanism, which states the existence of a

scalar boson (spin 0) whose interaction with the elementary particles generates their mass through the process known as Spontaneous Symmetry Breaking (SSB) [1, 6].

2.2 Spontaneous Symmetry Breaking

The spontaneous symmetry breaking is a phenomenon not restricted to the Particle Physics. In such context, though, it can be used as a mechanism to introduce mass terms in the SM theory without violating the gauge symmetries. It is important to highlight that different aspects can be observed for different situations when applying the SSB. As a general illustration, consider the following lagrangian for two scalar fields ϕ_1 and ϕ_2 (Goldstone Model) given by

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu \phi_1)^2 + (\partial_\mu \phi_2)^2] - V(\phi_1^2 + \phi_2^2) \quad (1)$$

which is invariant under transformations of the $SO(2)$ rotation group

$$\phi \equiv \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}. \quad (2)$$

Now, suppose a potential for the lagrangian in the following form

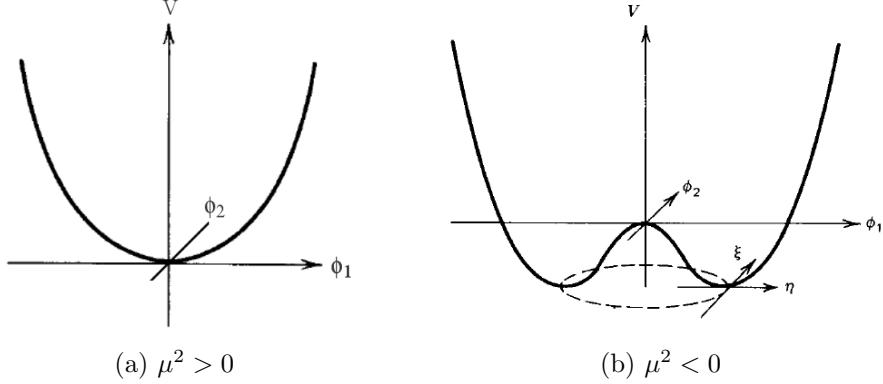
$$V(\phi^2) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}|\lambda|(\phi^2)^2 \quad (3)$$

where, $\phi^2 = \phi_1^2 + \phi_2^2$. The choice of the parameter μ defines two distinguishable cases for the potential. The case $\mu^2 > 0$ corresponds to the exact symmetry, which the vacuum state occurring for

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (4)$$

and the potential in this case has the profile represented in Fig. 1(a). The lagrangian, in such condition, describes just a pair of scalar particles with common mass μ . The case $\mu^2 < 0$ leads to the spontaneous breaking of the symmetry $SO(2)$, in which the absolute minimum of the potential, Fig. 1(b), corresponds to different vacuum states (all

Figure 1 - The $V(\phi)$ potential for different choices of μ^2 .



Source: Adapted from [7], p. 322 and 325.

degenerated in energy) given by

$$\langle \phi_0 \rangle = -\frac{\mu^2}{|\lambda|} = v^2 \quad (5)$$

Considering now one of those vacuum states (in terms of ϕ_1 and ϕ_2) and making the expansion around the chosen state, one has

$$\langle \phi \rangle = \begin{pmatrix} v \\ 0 \end{pmatrix} \quad (6)$$

$$\phi' \equiv \phi - \langle \phi_0 \rangle \rightarrow \langle \phi \rangle = \begin{pmatrix} v + \eta \\ \xi \end{pmatrix}. \quad (7)$$

Inserting this new displaced field in the lagrangian, it becomes

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu \eta)^2 + 2\mu^2 \eta^2] + \frac{1}{2}(\partial_\mu \xi)^2 + \mathcal{O}_{sup} + const. \quad (8)$$

where, \mathcal{O}_{sup} , representing terms of interaction and self interaction of the fields, and the constant are not important in this discussion.

The lagrangian obtained contains two particles: one (η) with mass $m_\eta = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2}$ and one massless particle (ξ). The massless particle is a consequence of the lagrangian SO(2) invariance and its origins is stated by the Goldstone Theorem. According to this theorem, scalar massless will arises every time a continuous symmetry is broken, or more specifically, for each generator of the original symmetry group broken there will

arise a massless and spinless particle [6–8]. The Higgs mechanism is an extension of the process showed here: an application of the spontaneous symmetry breaking in locally gauge invariant theories.

2.2.1 The Brout-Englert-Higgs Mechanism

In 1963, Phillip Andresson proposed that symmetries broken spontaneously would provide a base to explain the massive gauge bosons in non-relativistic systems and later, these ideas were studied in the context of quantum field theories. It was shown that a complex scalar field which potential is particularly chosen can spontaneously break the gauge symmetry. Peter Higgs suggested that this indicated the existence of a new massive scalar particle. Glashow, Weinberg and Salam showed that the Higgs mechanism could be used to break the gauge symmetry of the electroweak theory and produces the known electroweak interactions. They were even able to predict accurately the mass of the Z boson which has been observed directly in 1983 [9, 10].

The Higgs boson is the quantum of a scalar field which breaks the electroweak symmetry (EWSSB - *Electroweak Spontaneous Symmetry Breaking*) generating the mass of the W^\pm and Z vectorial bosons, providing yet a reasonable base to explain the origin of the leptons and quarks mass ² [11]. Understand the mechanism of the electroweak symmetry breaking is one of the most fundamental problems in the Particle Physics and the Higgs discovery is one of the bases for it.

Exemplifying the Higgs mechanism in the SM be the following lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial^\mu \phi) - \mu^2 \phi^* \phi - \lambda (\phi^* \phi)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (9)$$

where, ϕ is the complex field

$$\phi = \frac{\phi_1 \pm i\phi_2}{\sqrt{2}}. \quad (10)$$

In order that this lagrangian be invariant under local gauge transformation, that is, $\phi \rightarrow e^{iq\alpha(x)} \phi$, the ordinary partial derivative, ∂_μ , needs to be replaced by the so called *covariant* derivative $D_\mu = \partial_\mu + iqA_\mu$ (this is the process in which the interactions

² In the case of neutrinos, in modern theories (since the SM does not include mass for them), it is believed that the Higgs mechanism is not entirely responsible for the neutrinos mass [5].

are introduced). The gauge field A_μ must also transform as $A_\mu \rightarrow A_\mu - \partial_\mu \theta$ to complete the invariant lagrangian. Considering $\mu^2 < 0$ (condition for the symmetry breaking) the potential has a continuum of degenerated vacuum states at $\langle |\phi_0|^2 \rangle = -\mu^2/2|\lambda| = v^2/2$. Again, by the formalism of the theory, the field ϕ must be expanded in terms of a particular vacuum state. Choosing, for instance, $\langle \phi_0 \rangle = v/\sqrt{2}$, the field ϕ is conveniently parameterized as

$$\phi = \frac{e^{i\xi/v}(v + \eta)}{\sqrt{2}} \simeq \frac{(v + \eta + i\xi)}{\sqrt{2}} \quad (11)$$

which inserted in Eq. 9 and considering the covariant derivative and the field A_μ transformation, gives

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu \eta)^2 + 2\mu^2 \eta^2] + \frac{1}{2}(\partial_\mu \xi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + qvA_\mu \partial^\mu \xi + \frac{q^2 v^2}{2}A_\mu A^\mu + \mathcal{O}_{sup} + const. \quad (12)$$

As expected from the Goldstone theorem in the lagrangian of Eq. 12 there is a massless field, ξ . The term $qvA_\mu \partial^\mu \xi$ does not have a clear interpretation. Terms like that, bilinear in two different fields, indicate that the fundamental particles of the theory were not correctly identified. Also, note that the lagrangian of Eq. 9 has four degrees of freedom: two coming from the scalar field ϕ and two from the vectorial field A_μ . The lagrangian of Eq. 12, however, presents five degrees of freedom: two from the scalar field and three from the vectorial field due to its mass acquisition. Such non-conservation of the degrees of freedom means that exist non-physical fields in the lagrangian. This can be handled by applying the gauge symmetry. Note that the terms involving the fields ξ and A_μ can be written as the combination

$$\frac{q^2 v^2}{2} \left(A_\mu + \frac{1}{qv} \partial_\mu \xi \right) \left(A^\mu + \frac{1}{qv} \partial^\mu \xi \right) \quad (13)$$

which clearly suggests the gauge transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{qv} \partial_\mu \xi \quad (14)$$

such that,

$$\partial_\mu A'_\nu - \partial_\nu A'_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (15)$$

$$A'_\mu A'^\mu = A_\mu A^\mu + \frac{2}{ev} A^\mu \partial_\mu \xi + \frac{1}{e^2 v^2} (\partial_\mu \xi)^2. \quad (16)$$

Note that, this also corresponds to a phase rotation over the field ϕ parameterized from Eq. 11 (called unitary gauge)

$$\phi \rightarrow \phi' = e^{-i\xi(x)/v} \phi(x) = (v + \eta)/\sqrt{2} \quad (17)$$

and, as the lagrangian is invariant under local gauge transformations, it is possible to rewrite it as following

$$\mathcal{L} = \frac{1}{2} [(\partial_\mu \eta)^2 + 2\mu^2 \eta^2] - \frac{q^2 v^2}{2} A'_\mu A'^\mu - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \mathcal{O}_{sup} + const. \quad (18)$$

The Eq. 18 summarizes the Higgs mechanism. The terms between the brackets form a Klein-Gordon field associated to a massive particle while the other two terms represent a vectorial field associated to a massive gauge boson. It is interesting to observe as the gauge boson appears in the lagrangian. The original field A_μ has just two degrees of freedom (as the photon has just two polarizations). However, the Goldstone boson, also without mass and which appears from the spontaneous continuous symmetry breaking, combines with the gauge field in order to give it a longitudinal degree of freedom: the mass. Note that, the Goldstone does not disappear from the theory, it just is not explicitly apparent anymore due to the new gauge (Eq. 17) used in the lagrangian. The remaining component of the parameterized field ϕ , that is, the scalar field η , is the so called Higgs boson. This phenomenon of a gauge boson acquiring mass by "absorbing" a Goldstone boson is the Higgs mechanism: a combination of local gauge invariance and the spontaneous symmetry breaking [6–8, 19].

2.2.2 The Higgs Mechanism and Electroweak Symmetry Breaking

The elementary particles interacting weakly are described by multiplets belonging to the weak isospin group $SU(2)$. Furthermore, only the components left-handed of the fermionic fields participates in the electroweak interactions, so that, the isospin group is often identified as $SU_L(2)$. This means that only such components are modified on symmetry transformations. Experimentally is observed that the weak interactions vio-

late the weak isospin and hipercharge symmetries. In other words, the particles do not conserve the quantum numbers associated to these symmetries in process guided by weak interactions. The hipercharge is a quantum number associated to the symmetry group $U(1)$, which combined to the third component of the isospin, as $Q = I_3 + Y/2$, comports the values of electric charge of each particle weakly interacting [19, 20].

As already mentioned, the theories involved in the SM are theories locally gauge invariants, which contains massless fields. However, the weak processes are marked by interactions of short range, what is translated in the existence of an associated massive boson as the mediator of the interaction. In such way, the Higgs mechanism is associated to the breaking of the isospin and hipercharge symmetries, and the theory is constructed over the symmetry group $SU_L(2) \times U(1)_Y$. Additionally, the breaking of the symmetry must occur in such way that the electroweak symmetry survives, keeping a massless gauge field (the photon). Schematically the electroweak symmetry breaking is, then, represented as $SU_L(2) \times U_Y(1) \rightarrow U_{em}(1)$. The gauge fields, that are expected to result in mediators, are determined by the structure of the local symmetry of the theory. Thus, three vectorial fields are introduced for the $SU_L(2)$ symmetry and a vectorial field for the $U_Y(1)$ symmetry (for instance, W_μ^a ($a = 1, 2, 3$) and Y_μ , respectively) [21]. The electroweak lagrangian can, then, be written as

$$\mathcal{L} = \left(\partial^\mu \phi^\dagger + ig W^\mu \cdot T \phi^\dagger + \frac{1}{2} ig' Y^\mu \phi^\dagger \right) \left(\partial_\mu \phi - ig W_\mu \cdot T \phi - \frac{1}{2} ig' Y_\mu \phi \right) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (19)$$

where, g and g' correspond to the coupling constants for the gauge interactions associated to the symmetry groups $SU_L(2)$ and $U_Y(1)$, respectively. These constants has the following relation with the Weinberg angle (electroweak mixture angle)

$$\text{sen}^2 \theta_W = \frac{g'^2}{g^2 + g'^2} \approx 0.23 \quad (20)$$

The Higgs field is also described by a multiplet of $SU(2)$, which minimum configuration capable to generate the expected symmetry breaking is given by the doublet of complex fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (21)$$

where, the field ϕ^+ has charge $Q = 1$ and the field ϕ^0 has charge $Q = 0$. As

exemplified in the previous sections, the mass of the gauge bosons is identified after the symmetry breaking done by the choice of a particular vacuum state to expand the field ϕ . In this procedure, the gauge freedom is used again in order to write the lagrangian containing only physical fields, which is often achieved by using the gauge

$$\phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (22)$$

In the end, the massive gauge bosons arise through the combination of the fields $W_\mu^{1,2}$ (for the bosons W^\pm) and of the fields Y_μ and W_μ^3 (for the bosons Z and γ - the photon) through the expressions

$$\left. \begin{aligned} W_\mu^\pm &= \frac{(W^1 \mp iW^2)}{\sqrt{2}}, & M_W &= \frac{vg}{2} \\ Z_\mu &= \frac{gW_\mu^3 - g'Y_\mu}{\sqrt{g^2 + g'^2}}, & M_Z &= \frac{v\sqrt{g^2 + g'^2}}{2} \end{aligned} \right\} \text{(bósons vetoriais eletrofracos)} \quad (23)$$

$$A_\mu = \frac{g'W_\mu^3 + gY_\mu}{\sqrt{g^2 + g'^2}}, \quad M_\gamma = 0 \quad (\text{fóton}) \quad (24)$$

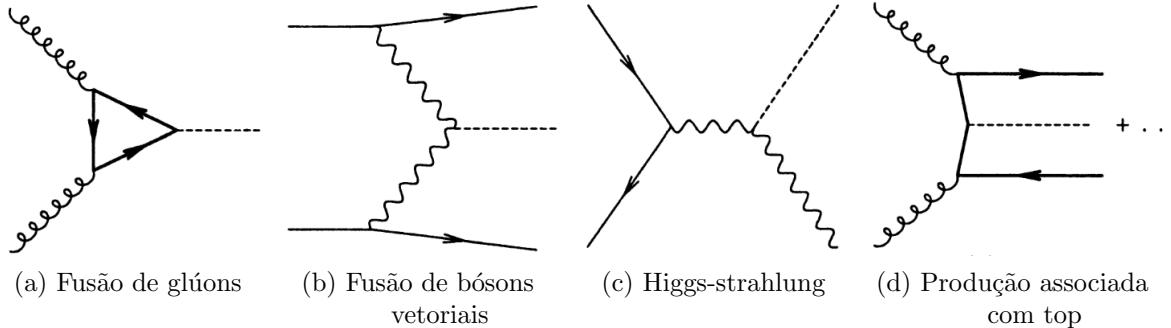
where, the parameter $v \approx 246$ GeV is the expected vacuum value of the Higgs field. It is interesting to note that, in order to generate the leptons and quarks mass (Yukawa sector) the same Higgs doublet (Eq. 21) can be used. For the leptons the application is immediate, while for the quarks, which have the fields in the up part of the doublets also massive³, it is needed a new doublet obtained from the Eq. 21 [7].

2.3 The Higgs Boson at the LHC

The standard model (SM) of electroweak interactions relies on the existence of the Higgs boson, the scalar particle associated with the field responsible for spontaneously breaking electroweak symmetry [24, 25]. The ATLAS and CMS collaborations reported in 2012 for the first time the discovery of a new boson consistent with the SM Higgs [3-8] by observing proton-proton (pp) collisions from the LHC at $\sqrt{s} = 7$ and 8 TeV.

³ The lepton doublets are composed by a lepton and its respective neutrino, which in the SM is considered massless.

Figure 2 - The main Higgs (represented by the dotted line) production modes at the LHC.



Source: Reference [21], p. 398.

Several studies carried out by CMS and ATLAS, and the combination of these results, have shown so far that the properties of the observed scalar boson is consistent with the SM expectations [9-12].

At the LHC, the Higgs boson can be produced in four main modes (the most significant in terms of their cross section). These modes are [21]:

- a) Gluon fusion: $gg \rightarrow H$;
- b) Vector boson fusion (VBF): $qq \rightarrow Hqq$;
- c) Vectorial boson association (Higgstrahlung): $q\bar{q} \rightarrow WH/ZH$;
- d) Top quark association: $gg/q\bar{q} \rightarrow t\bar{t}H$.

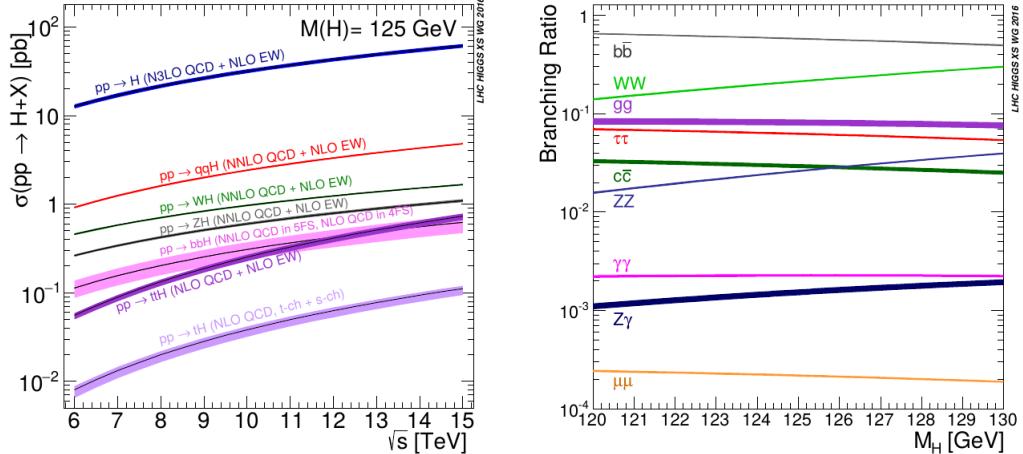
The Feynman diagrams at LO (*leading-order*) for each of this processes are shown in Fig. 2. The gluon fusion is the production mode in which the Higgs is produced through a loop of quarks top induced by the interaction of gluons in the initial state. This is the production mode with the highest probability to happen, presenting a XS of about 43.9 pb for $m_H = 125\text{GeV}$ at $\sqrt{s} = 13\text{TeV}$ [91].

The second most important production mode if vector boson fusion. In this production mode the Higgs is created by the fusion of weak vectorial bosons (Z , W^\pm) which arises during the scattering of quarks in the colliding protons. The main characteristic of this mode is the presence of two energetic jets in the final state and separated by a pseudorapidity gap. Its cross section at 13TeV for $m_H = 125\text{GeV}$ is about 3.7 pb [22, 91].

The XS of the Higgs production associated to vectorial bosons is even smaller, but they are useful for observing the Higgs decaying at $b\bar{b}$ ou $\gamma\gamma$ since the decay can be tagged by the presence of the bosons W/Z . The production of Higgs up to $M_H \sim 400\text{ GeV}$ associated to quark top is the smallest one between the main production modes. This process is similar to the Higgstrahlung, but the Higgs in this case is emitted from top quarks through gg or $q\bar{q}$ interactions [21].

The Higgs can decay in several final states and each of them presents advantages

Figure 3 - Higgs production XS and branching ratios.



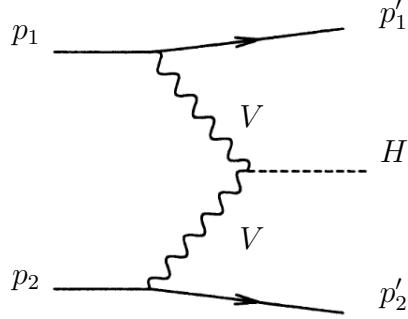
Source: Reference [16], p. 184.

and disadvantages. The decaying channel $ZZ \rightarrow 4l$, handled in this analysis, is one of the best ones. This channel is actually called "the golden channel" because all the decay products can be fully reconstructed in the CMS detector and it also has relatively small backgrounds at the signal region. However, the branching ratio (BR) into $ZZ \rightarrow 4l$ is small, about $3.3e^{-6}$ for $m_H = 125\text{GeV}$ at 13TeV [91]. The Fig. 3 shows the Higgs cross sections and branching ratios at 14TeV .

2.4 The Higgs VBF Production Mode

Many measurements of the properties of the Higgs boson are still statistics limited, and there is still significant room for beyond standard model effects in the Higgs sector. A particularly sensitive channel for observing non-standard model effects is the cross section for Higgs boson production through the Vector Boson Fusion (VBF) mode. There are some special features about this production mode. It has the largest cross section between the processes that involves tree-level production of the Higgs boson. It is also the second largest among all the processes. Its signature of two energetic forward jets makes possible to tag events and identify Higgs decays that normally have large backgrounds. Another feature is that the Higgs p_T is nonzero, even at lowest order, facilitating searches of invisible decay modes. The VBF mode also has a particular sensitivity to the charge-parity properties of the Higgs boson and non-standard Higgs interactions via the angular correlations between the jets [47–52].

Figure 4 - Feynman diagram for the Higgs production via vector boson fusion. The initial state quarks (or anti-quarks) are p_1 and p_2 , while the momentum at final state are p'_1 and p'_2 , respectively.



Source: Adapted from [13], p. 196.

2.4.1 The Higgs VBF LO Cross Section and its Kinematics

Assuming the Higgs production via the VBF mode as shown in Fig. 4, the matrix element for this process, in terms of the momentum of the involved particles, can be expressed as [13]

$$\mathcal{M} = g_{VWH} \frac{\bar{u}(p'_1)\gamma^\lambda(g_V + g_A\gamma_5)\bar{u}(p_1) \bar{u}(p'_2)\gamma_\lambda(g'_V + g'_A\gamma_5)\bar{u}(p_2)}{(q_1^2 - M_V^2)(q_2^2 - M_V^2)}, \quad (25)$$

where for,

$$V = W:$$

$$g_{VWH} \rightarrow g_{WWH} = gM_W \quad (26)$$

$$g_V = -g_A = \frac{e}{2\sqrt{2} \sin \theta_W} \quad (27)$$

$$V = Z:$$

$$g_{VWH} \rightarrow g_{ZZH} = \frac{gM_W}{\cos^2 \theta_W} \quad (28)$$

$$g_V = \frac{g(T_{3L}/2 - Q \sin^2 \theta_W)}{\cos \theta_W} \quad (29)$$

$$g_A = -\frac{gT_{3L}}{2 \cos \theta_W}. \quad (30)$$

The term $e/\sin \theta_W$ is the weak SU(2) gauge coupling and the terms T_{3L} and Q stand for the third component of the weak isospin and the quark charge⁴, respectively. Now, using $g_L = (g_V - g_A)/2$ and $g_R = (g_V + g_A)/2$ the average squared matrix element

⁴ For quarks and leptons $T_{3L} \pm 1/2$.

becomes [13–15]

$$|\mathcal{M}|^2 = 64 g_{VWH}^2 \frac{C_1(p_1 \cdot p_2)(p'_1 \cdot p'_2) + C_2(p_1 \cdot p'_2)(p'_1 \cdot p_2)}{(q_1^2 - M_V^2)^2 (q_2^2 - M_V^2)^2} \quad (31)$$

where $C_1 = g_L^2 g_L'^2 + g_R^2 g_R'^2$ and $C_2 = g_L^2 g_R'^2 + g_R^2 g_L'^2$. The differential cross section at LO is, then, given by [14]

$$d\hat{\sigma}_{LO} = \frac{1}{72} \frac{1}{(2\pi)^5} \frac{1}{\hat{s}} |\mathcal{M}|^2 \frac{d^3 p_H}{2dE_H} \frac{d^3 p'_1}{2dE'_1} \frac{d^3 p'_2}{2dE'_2} \delta^4(p_1 + p_2 - p'_1 - p'_2 - p_H) \quad (32)$$

which integration over the variables p'_1 and p'_2 is conveniently performed in the rest frame of the final-state quarks ($\vec{p}'_1 + \vec{p}'_2 = 0$), leading to [14]

$$\frac{d\hat{\sigma}_{LO}}{dE_H d\cos \theta} = \frac{G_F^3 M_V^8}{9\sqrt{2}\pi^3 \hat{s}} \frac{p_H}{32s_1 s_2} (C_1 \mathcal{H}_1 + C_2 \mathcal{H}_2) \quad (33)$$

where $\hat{s} = (p_1 + p_2)^2$ is the energy at the CM (*center of mass*), G_F is the Fermi constant, $s_{1,2} = \sqrt{\hat{s}}[(\sqrt{\hat{s}} - E_H) \pm p_H \cos \theta]$, with p_H being the Higgs momentum and θ the final-state parton scattering angle⁵. The $\mathcal{H}_{1,2}$ are terms depending on $s_{1,2}$, \hat{s} , p_H , θ and M_V . For the interested reader, please see [14].

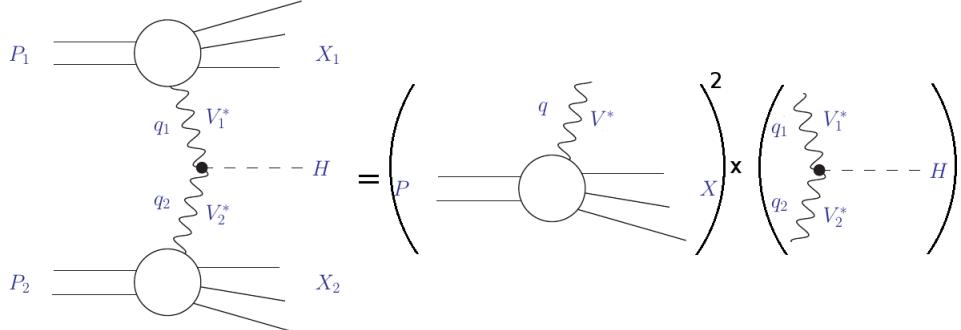
The total partonic cross section of the Higgs produced via VBF is obtained by integrating the Eq. 33, which can properly done in terms of the Higgs transverse momentum and rapidity as

$$\hat{\sigma}_{LO}(qq \rightarrow Hqq) = \int_{y_-}^{y_+} dy \int_0^{p_T^{max}} dp_T (2\pi p_T) \frac{d^2 \hat{\sigma}_{LO}}{dy dp_T}. \quad (34)$$

The integration on y and p_T on Eq. 34 can not be done analytically and numerical integration is used to get physical results [14,15]. As reported on this previous references, the results show that the quarks at the final state present large energies while their transverse momentum is usually $p_T \sim M_V$ but, varying with m_H (see Fig. 3.14 of [14] and Fig. 6 of [15], for instance). The relatively small transverse momentum of the final-state partons can be translated as a small scattering angles, which in terms of pseudo-rapidity (η), means that, the rising jets are forward (a well known signature of VBF).

⁵ Do not confuse that with the θ_W , which is the weak-mixing angle (Weinberg angle).

Figure 5 - Sketch of the Higgs VBF production description through the structure function approach.



Source: Adapted from [56], p. 3.

2.4.1.1 Corrections of Higher Orders

Through the past years considerable enhancements have been achieved in the phenomenological interpretation of the Higgs VBF production mode. Computations for the total cross section and differential distributions to Higgs produced via VBF and including NLO QCD and EW corrections were presented by [53, 54] and are available in flexible parton-level MC generators. Partial NNLO QCD corrections have also been presented by [55, 56], which reduces the uncertainties on the inclusive VBF XS from 5-10% to 1-2%, making VBF the theoretically most accurate production mode at hadron colliders [12].

The most recent developments done on the VBF theory formalism comes from the application of the so called *structure function approach*, which is based on the description, in very good approximation, of the VBF process as a double deep-inelastic scattering process (DIS), see Fig. 5. In such approximation the two virtual vector bosons \$V_{1,2}^*\$ are independently emitted from the hadronic initial states and fuse into the Higgs boson.

This approximation builds on the absence (or smallness) of QCD interference between the two inclusive final states \$X_1\$ and \$X_2\$. The total cross section is given as a product of the matrix element \$\mathcal{M}^{\mu\nu}\$ for VBF (\$V_1^\mu V_2^\nu \rightarrow H\$) and the DIS hadronic tensor \$W_{\mu\nu}\$, giving a VBF differential cross section of the form [54, 56],

$$d\sigma = \frac{1}{2S} 2G_F^2 M_{V_1}^2 M_{V_2}^2 \frac{1}{(Q_1^2 + M_{V_1}^2)^2} \frac{1}{(Q_2^2 + M_{V_2}^2)^2} \times W_{\mu\nu}(x_1, Q_1^2) \mathcal{M}^{\mu\rho} \mathcal{M}^{*\nu\sigma} W_{\rho\sigma}(x_2, Q_2^2) \times \frac{d^3 P_{X_1}}{(2\pi)^3 2E_{X_1}} \frac{d^3 P_{X_2}}{(2\pi)^3 2E_{X_2}} ds_1 ds_2 \frac{d^3 P_H}{(2\pi)^3 2E_H} (2\pi)^4 \delta^4(P_1 + P_2 - P_{X_1} - P_{X_2} - P_H), \quad (35)$$

where, \$S = (\sqrt{S})^2\$ stands for the center-of-mass energy of the collider, \$G_F\$ is the

Fermi's constant, $Q_1^2 = -q_i^2$ and $x_i = Q_i^2/(2P_i \cdot q_i)$ are the usual DIS variables, M_{V_i} stands for the vector boson mass. The hadronic tensor $W_{\mu,\nu}$ represents the unknown physics at the $V_{\mu,\nu}^{1,2}$ -parton vertexes. This function depends on the four-momentum of both vector boson and the parton.

The theoretical discussion about the fully computation of the total cross section exceeds the scope of this analysis and it will not be presented here. For the reader interested on more details, please have a look on [14], [56] and [54].

2.5 The Physics Scenario of this Analysis

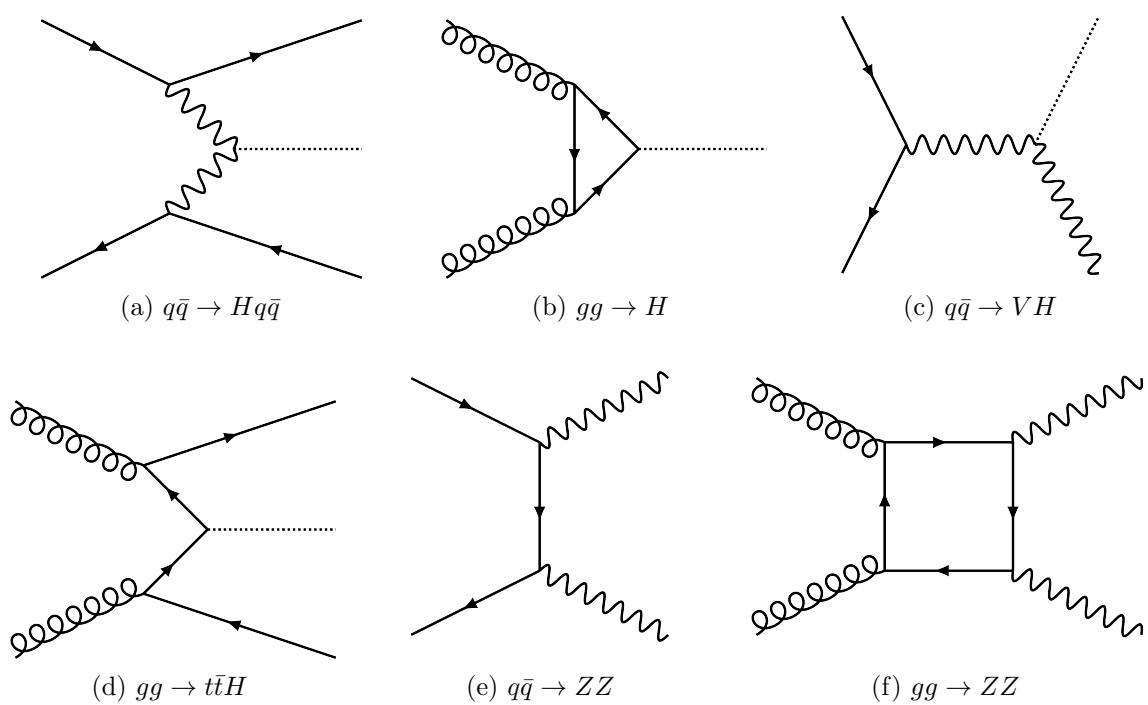
The VBF production mode, as presented on previous sections, is a key process for precision measurements of Higgs properties since it is a clean channel with very distinctive kinematics. Such features provide good scenario for measurements of Higgs couplings [57]. The VBF channel is a direct probe of the coupling between vector bosons and the Higgs boson, and hence directly probes the electroweak sector of the standard model.

Additionally, the $H \rightarrow ZZ \rightarrow 4l$ ($l = e, \mu$) decay channel has a large signal-to-background ratio because it is possible to completely reconstruct the final state leptons, which present excellent momentum resolution. This features makes this decay channel one of the most important for the studies of the Higgs boson properties. Several different measurements have been performed with the data collected during the LHC RunI using this channel [27–30, 46]. This decay channel, thus, presents a good frame to measure Higgs VBF production.

This analysis presents a measurement of the Higgs boson VBF production cross-section obtained in the $H \rightarrow ZZ \rightarrow 4l$ decay channel using data collected at $\sqrt{s} = 13\text{TeV}$. The data accounts for 35.9fb^{-1} of pp collision collected with the CMS detector at the LHC RunII in 2016. This analysis uses similar requirements as the CMS *HZZ4L* 2016 analysis [72] and [77], relying essentially on the reconstruction, identification and isolation of leptons and requiring the presence of jets and Missing Energy Transverse (MET). The specific selection and analysis requirements will be discussed in more detail in the following sections.

As it will also be shown, the main processes in this analysis, besides the VBF itself, are the $gg \rightarrow H$ Higgs production mode and the reducible SM background $q\bar{q} \rightarrow ZZ$ (which in the context of the present analysis become a non-reducible background). In a smaller, but still significant, contribution appears the remaining Higgs production modes, Higgs-strahlung (VH) and associated with top quark ($t\bar{t}H$), and the SM background $gg \rightarrow ZZ$. The tree-level Feynman diagrams for each of these processes are shown in Fig. 6.

Figure 6 - Main processes in this analysis: Higgs production modes (a) vector boson fusion (VBF), (b) gluon-fusion (ggH), (c) Higgsstrahlung (VH), (d) production associated to top quark, and the main SM backgrounds (e) and (f) ZZ . Additionally there is the $Z + X$ background which is derived via the Fake Rate *data-driven* method and will be presented later.



Source: The AUTHOR, 2018.

3 THE LHC AND THE CMS EXPERIMENT

3.1 The Accelerators at CERN

The CERN (*European Organization for Nuclear Research*) it is a scientific complex created in 1954 and designed mainly for Particle Physics research. Its name derives from the original French acronym *Conseil Européen pour la Recherche Nucléaire*, which was proposed by the funding group at 1952, aiming to create an international organization for research in fundamental Physics⁶. The research lines developed nowadays at CERN goes far beyond the atomic nucleon and the group of scientists associated to it, is very diversified. The main instrumentation used at CERN are the particle accelerators and their associated detectors, which allow the scientists to access informations about the Physics in the very small world. The biggest of those accelerators is the LHC (*Large Hadron Collider*) which started to be build in 1994 and was activated for the first time in September of 2008. On its first official operation phase for data collection, period from 2011 to 2013, the LHC achieved 7 and 8TeV of energy, resulting in a integrated luminosity of $\mathcal{L} = 5.051$ and 19.712 fb^{-1} , respectivamente [31].

The LHC it is a hadronic collider - collides protons and lead ions (Pb) - projected by CERN to operate in a energy up to $\sqrt{s} = 14\text{TeV}$ in the center of mass (frame in which the total momentum of the colliding particles is null). The main motivation for its construction is the elucidation of the nature of the Electroweak Symmetry breaking (EWSB). Also, though the SM has been consistent so far with the observations, it is an effective theory. The LHC allows the test of the present theories in such way never done before. The conditions (energy density and temperature, for instance) created during the collisions are estimated to be similar to the ones existing few moments after the so called Big Bang [32]. Added to this, there is the possibility to observe process related to Dark Matter, including mechanisms involving Higgs decays, for instance.

The hadronic colliders have an increased potential to discover new particles, since they make the collisions in a range of energy (since the partons inside the protons can carry different fractions of the proton total energy). Also, the energy in the collision is higher than in a linear collider since both colliding objects are moving against each other ($E = E_{beam1} + E_{beam2}$). Another advantage is that, the protons are relatively heaviest and loss less energy while being accelerated in an intense magnetic field [33].

The LHC was constructed in a tunnel of ~ 27 km. Such tunnel is located at 100 m underground and is located in the frontiers between France and Switzerland. The

⁶ Back to that time, only physicists from Nuclear Physics integrated the group, what justifies its name.

Table 2 - Some parameters of the LHC.

Parameter	Value
Exact circumference	26 659 m
Dipoles operation temperature	1.9 K (-271.3°C)
Number of magnets	9593
Number of main dipoles	1232
Number of main quadrupoles	392
Number of RF cavities (per beam)	8
Dipoles maximum magnetic field	8.33 T
Minimum distance between the bunches	~7 m
Projected Luminosity	$10^{34} \text{ cm}^{-2}\text{s}^{-1}$
Bunches per beam	2808
Protons per bunch (at the beginning)	$1.1 * 10^{11}$
Cycles in the ring by a beam	11 245/s
Number of collisions	$6.e^8 /s$
Energetic consuming	~120 MW*
Pressure in the beam pipes	10^{-13} atm
Beam energy (at $\sqrt{s} = 14 \text{ TeV}$)	350 MJ**

*230 MW needed to supply all CERN.

**Equivalent to a 500 ton train at 150km/h.

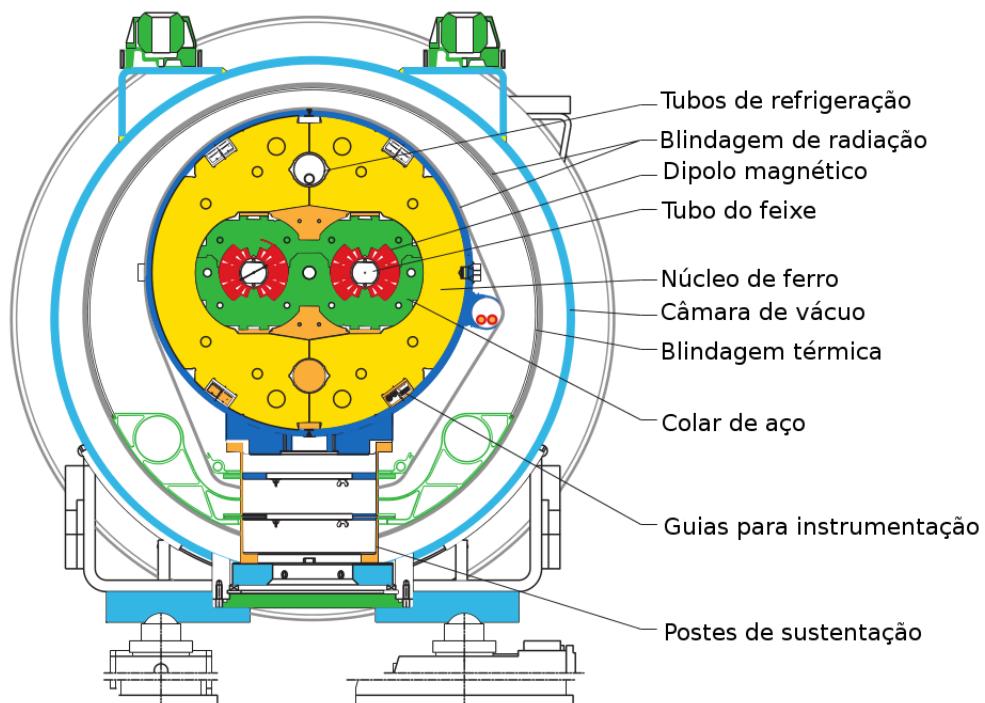
Source: Adapted from [32], p. 30.

accelerator is approximately circular, being composed by eight main arcs in which there are 154 magnetic dipoles used to keep the particle beams in the curve trajectory. Each arc has an internal structure (Fig. 7) divided into layers: the inner layer, which supports the beam pipes and magnetic dipoles, is constituted of iron and the outer layers are used to create a high vacuum region, being coated by a radiation and thermic shielding. Such configuration enhances the magnetic field and the system cooling made with liquid He inside the tube. The particle acceleration is done by eight (for each beam) super-conductive radio-frequency cavities disposed along the arcs, which outputs up to 2 MV at 400 MHz. Tab. 2 shows some parameters of the LHC operation.

The LHC operation was structured on the already existent accelerators at CERN. These accelerators are nowadays used to make the initial particle acceleration at smaller energies. The Tab. 3 summarizes the energies and velocities achieved by a proton in each accelerator up to the LHC and Fig. 8 shows an scheme of the accelerators.

At the LHC there are four main experiments (big detectors): ALICE (*A Large Ion Collider Experiment*), ATLAS (*A Toroidal LHC Apparatus*), CMS (*Compact Muon Solenoid*), LCHb (*Large Hadron Collider beauty*). These experiments are located on the beams collision points and each of them have a complex computing system, which the main frame is called Trigger system. The Trigger is responsible to combine the information coming from the sub-detectors in each experiment and identify the particles in a fast way in

Figure 7 - Simplified transverse section of the LHC.



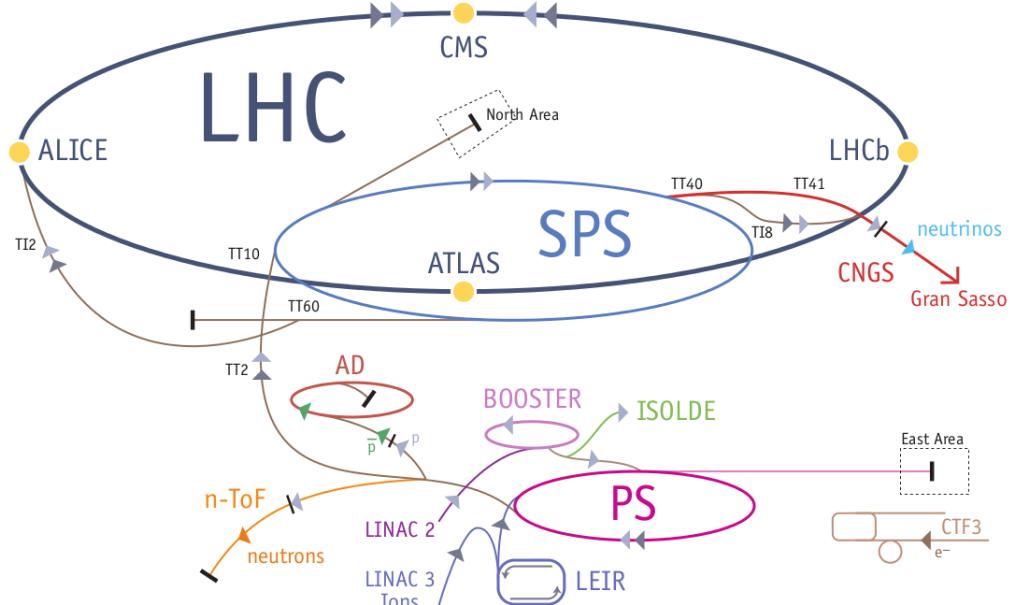
Source: Adapted from bib:Nature-448-2017, p. 286.

Table 3 - Comparison between the energy and velocity achieved by a proton in the different accelerators located at CERN.

Accelerator	Kinetic energy (K)	Velocity (%c)
Linac 2	50 MeV	31.4
PS Booster	1.4 GeV	91.6
PS	25 GeV	99.93
SPS	450 GeV	99.9998
LHC	7 TeV	99.999991

Source: Reference [32], p. 5.

Figure 8 - Accelerators composing the LHC at CERN.



Source: Reference [32], p. 13.

order to decide if a given event is interesting to be stored and further processed (in a more refined way). Such system is very important since, at 14TeV μ the bunches composing the beams are spaced by 25 ns and generates $\sim 4 \cdot 10^7$ crossings every second, what can lead up to ~ 1 G events/s [32, 34]. In fact, there will be upgrades on this system for future collision scenarios as it is discussed at Appendix E.

3.2 The CMS Experiment at the LHC

The CMS is a general purpose experiment, which has similar physics goals as the ATLAS detector and both are used to look for different Physics topics. Different from ATLAS, the CMS has different detection techniques and a different magnetic system profile. Its construction is based on the traditional barrel model with two endcaps. As the name suggests, it was constructed to provide a good detection and resolution of muons. It has a giant cylindric coil made of super-conducting material, being able to generate a 3.8 T field (which is about 80 thousand times larger than the earth magnetic field). The CMS has 21 m of length, 15 m of diameter, weigh about 12500 tons and its construction cost is about R\$ 1,317 billions [32, 35].

The construction of the CMS follows some requirements needed to the Physics goal in the TeV energy scale. Between them, is the good identification and resolution in the muons transverse momentum and charge, good reconstruction efficiency of the tracks

from the tracker systems, good resolution in the energy deposited on the calorimeters and good resolution on the detection of missing energy transverse (MET). The coordinates system adopted in CMS has the origin in the beams collision point, while the **y** axis is in the vertical direction (positive to up) and the **x** axis points radially to the center of LHC circumference. Thus, the **z** axis is in the beams moving direction. The azimuthal angle ϕ is on the plane **xy**, being measured from the **x** axis and the polar angle θ is measured from the **z** axis. Usually, due to Lorentz invariance, the angle θ is replaced by the pseudorapidity which is defined as $\eta = -\ln[\tan(\theta/2)]$ [32, 35].

The CMS experiment is composed by several sub-detectors distributed in radial layers (onion-like) around the beam pipe and each sub-detector is specialized in the detection of specific types of particles. The Fig. 9 show a sectioned view of the experiment and Fig. 10 how each type of particle interacts with each sub-detector. The main parts of the CMS are the tracker, the superconducting solenoid, the calorimeters, the muon chambers and the frontal detectors.

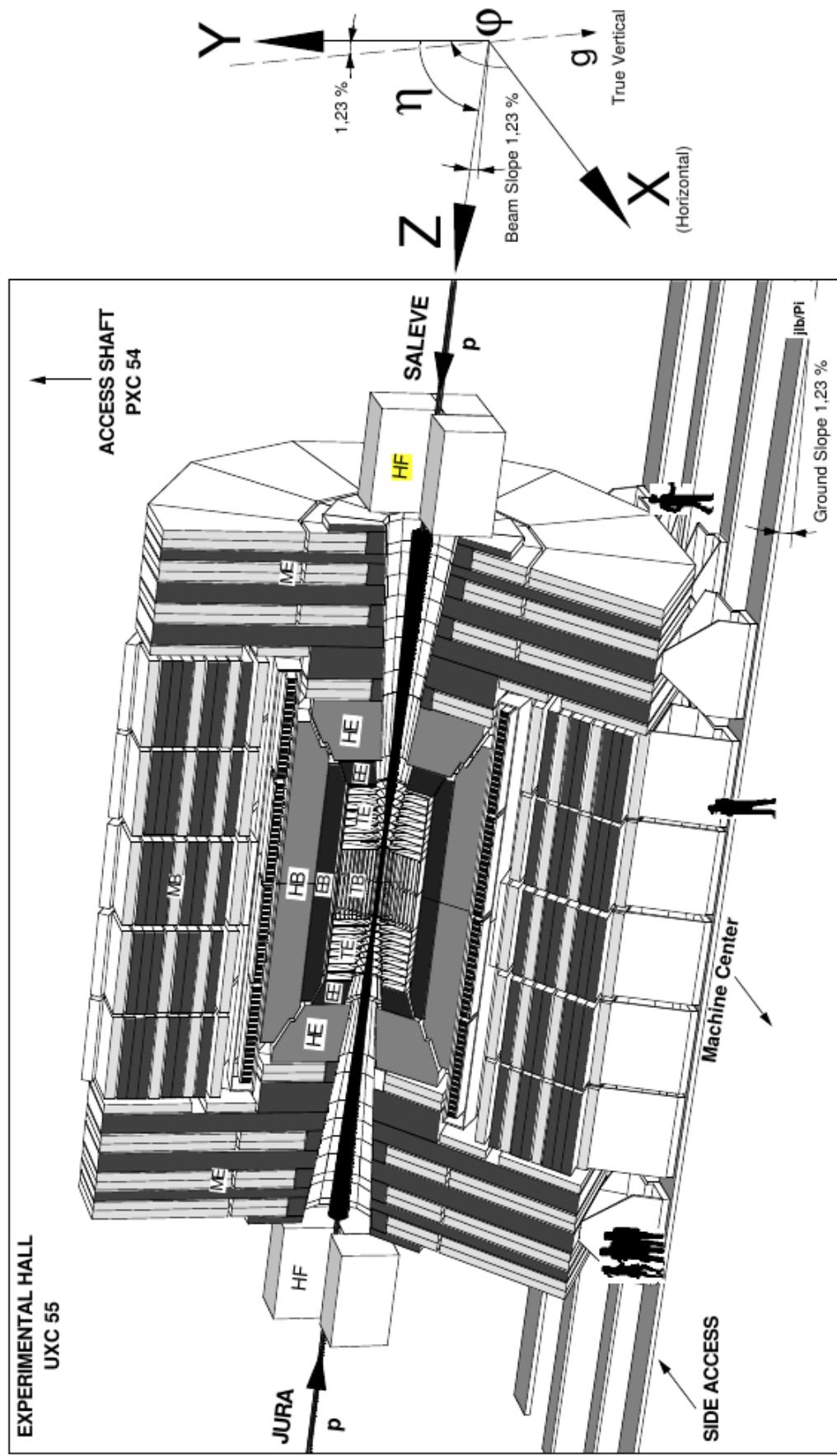
3.2.1 The Tracker System

The tracker system is composed by two sets of detectors (Pixels and Strips) disposed radially around the beam pipe and has a total length of ~ 5.40 m and total radius of ~ 1.10 m. Each set presents a specific resolution but both of them detect the charged particles trajectory. The detectors used in the tracker are projected to support high radiation doses, since they are the ones closest to the collision point. Summing both sets of detectors, the CMS tracker has ~ 200 m² of active silicon (SiO_2). Its working principle is based on the ionization of silicon solid state sensors and in the collection of the generated charges by an electric field. These detectors stand out of the gas chambers due to their higher density, which promotes absorption of higher energy particles [34, 41, 42]. Note, however, that this detector does not have the purpose to stop the particles, this is done by the calorimeters.

3.2.1.1 The Silicon Pixels Detector

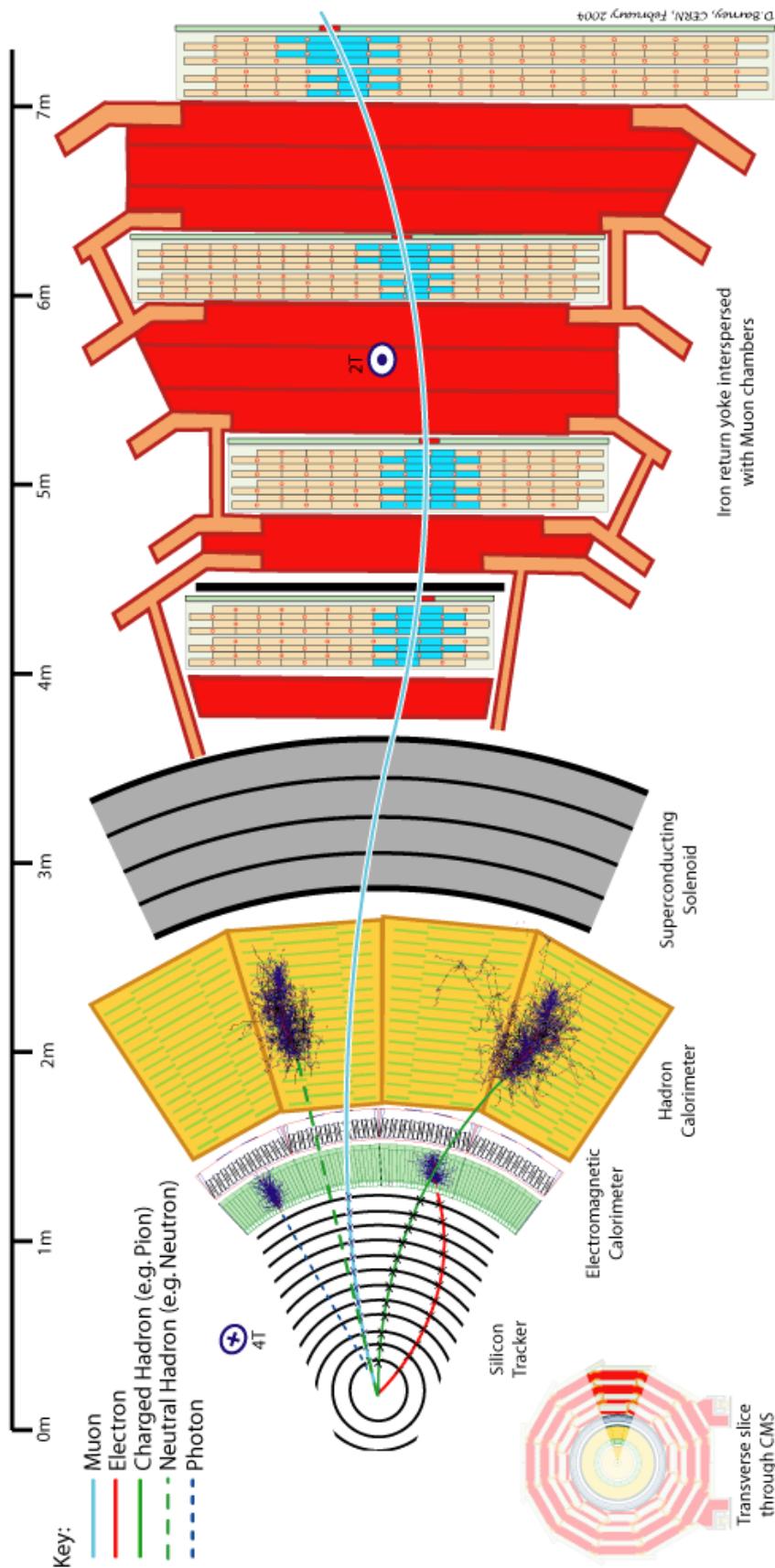
The pixels detector comprehend the inner section of the tracker system and is projected to identify with high precision the particles trajectories in that region. It has great importance in the reconstruction of secondary vertices and in the measurement of small impact parameters, which are relevant for the identification of particles associated to heavy flavor quarks. This detector is composed by three cylindric layers, located around the beam pipe, and two discs (endcaps) located in the beam pipe transverse plane (Fig. 11). The complete detector has 1 m² of silicon with 66 millions of pixels. The layers

Figure 9 - General view of CMS layers.



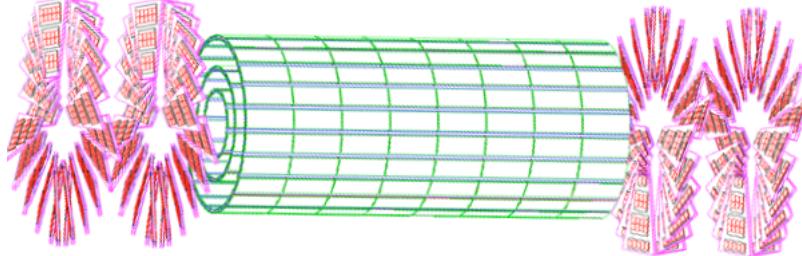
Source: Reference [41], p. 5.

Figure 10 - Transverse section view showing the detection that each sub-detector does.



Source: Reference [35].

Figure 11 - Scheme of the pixel detector.



Source: bib:CMS-PTDR-2006, p. 20.

in the barrel extend for 53 cm and are disposed radially at 4.4, 7.3, 10.2 cm from the **z** axis. In the barrel there are 768 pixels with area $100 \times 150 \mu\text{m}^2$ and disposed in a half-stair fashion way. In the discs exist 672 pixels disposed in helice and rotated by 20° around the radial direction. The pixel detector has a coverage of $|\eta| < 2.5$ and presents a resolution of about $10 \mu\text{m}$ in the **r**- ϕ and $20 \mu\text{m}$ in the **z** direction.

3.2.1.2 The Silicon Strips Detector

The silicon strips detector is the tracking system positioned around the pixels detector, being divide into two parts identified as: the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). The TIB is composed by four layers of silicon sensors with thickness of $320 \mu\text{m}$ and cover the region for $|z| < 65 \text{ cm}$. The TOB has 6 layers, also of SiO_2 , with average length of $|z| < 110 \text{ cm}$. The silicon strips detector also has endcaps, which are divided into two categories: Tracker EndCaps (TEC) and Tracker Inner Discs (TID). Each TEC is composed by 9 discs which are disposed in the region $120 < |z| < 280 \text{ cm}$ and each TID is composed by 3 small discs which fill the space between the TIBs and the TECs. Both TID and TEC are organized in rings around the beam pipes and the sensors that constitute them have thickness of $320 \mu\text{m}$ (the entire TID and the 3 inner layers of the TEC) and $500 \mu\text{m}$ (remaining TEC layers). This entire detector has the length of 5.8 m and diameter of 2.6 m, containing about 10 million SiO_2 fibers. Its main designation is the track reconstruction from muons, isolated electrons and hadrons with high transverse momentum, keeping an efficiency higher than 98% in the interval $|\eta| < 2.5$.

3.2.2 The Calorimeters

A calorimeter is a system in which the particles are totally absorbed in order to measure their energy, that is, the particle is stopped. There are two types of calorimeters in the CMS: the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL). The ECAL is designed to measure the energy of particles which interacts mainly electromagnetically with its material, such as electrons and photons. At high energy, the electrons loss their energy almost exclusively by the *bremsstrahlung* process. The photoelectric effect and Compton also happen. The HCAL is built to measure the energy of particles which interacts through the strong and weak interactions, the case of hadrons such as pions, kaons and so on. The particles interacting with the calorimeters usually produces shower of other particles which appears from the decay of the first ones. The precision on the measurement of the energy deposited in a calorimeter is given by $\Delta E/E \propto 1/\sqrt{E}$, and thus, the LHC can measure the particles energy with higher precision than the previous colliders.

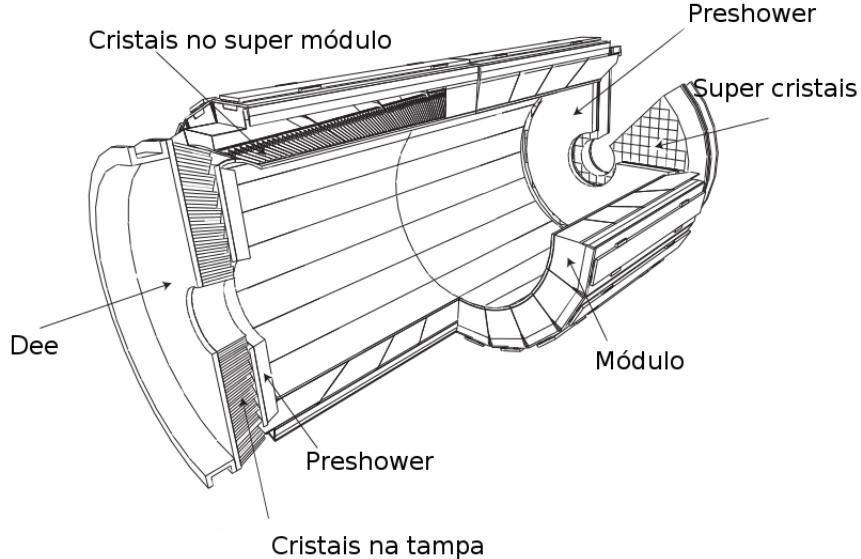
3.2.2.1 The Electromagnetic Calorimeter (ECAL)

The ECAL is a hermetic⁷ and homogeneous calorimeter, composed by 61200 $PbWO_4$ crystals disposed over the central barrel (EB) and by 7324 crystals in each endcap (EE). Between the properties which lead to the choice of the $PbWO_4$ as the material for the calorimeter scintillators, are the short radiation wave length ($\chi = 0.89$ cm) and the short Molière radius (2.2 cm). This allowed the construction of a compact calorimeter. The Molière is a radial measurement of a cylinder, around the shower axis, which contains about 95% of its total energy, being almost independent of the energy of the incident particle. Additionally, the $PbWO_4$ scintillators are fast such that about 80% of the produced light is emitted in the interval of 25 ns [34, 42].

In the barrel, the ECAL-EB covers a region of $|\eta| < 1.479$, with granularity of 1 crystal for each 1° in ϕ . These crystals measure 23 cm in length and have faces with 22×22 mm² (front) e 26×26 mm² (back, where the photo-multipliers are coupled), forming a total volume of 8.14 m³ (and weighting 67.4 tons). The $PbWO_4$ crystals are supported by a aluminum structure that keeps them spaced from each other by 0.35 mm and grouped in sub-modules. Such sub-modules are grouped in modules containing from 400 to 500 crystals. Finally, a trapezoidal structure groups four modules in so called super-modules,

⁷ Capable to detect all the products of a decay by covering a big area around it (in other words, maximum coverage in the solid angle).

Figure 12 - Illustration of a super-module and the entire ECAL.



Source: Adapted from [34], p. 93 e 95.

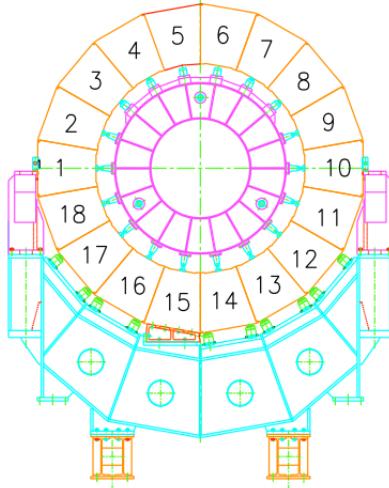
which form a half angular section of the barrel containing 1700 crystals (Fig. 12). In order to form half arc of the barrel, 18 super-modules are needed, each covering 20° . On each sub-module the crystals are positioned forming a small angle with respect to the collision point in order to have coincidence between the crystal axis and the radial axes from the coordinates system.

The ECAL endcaps cover a region with $1.479 < |\eta| < 3.0$ and are installed 314 cm away from the CMS coordinates system. In that region the endcap is composed by crystals grouped in 5×5 unities (called super-crystals or SCs) by a structure made of carbon fiber. Each crystal has 22 cm in length and present faces of 8.4 cm^2 (front) and 9 cm^2 (back). On each endcap, the ECAL is divided in two halves (called *Dees*) which contain 3662 crystals each one. The crystals and the SCs are organized in a rectangular grid (**xy**) pointing to the origin of the CMS coordinates system. The two endcaps together comprehend a volume of 290 cm^3 and weigh 24 tons. In front the scintillators, on each endcap, there is also a shower detector (called *preshower*). This detector is composed of two planes of silicon trips. The main goal of the preshower is to identify neutral pions in the region $1.653 < |\eta| < 2.6$, which allows the suppression of significant backgrounds in the $H \rightarrow \gamma\gamma$ channel [34, 41].

3.2.2.2 The Hadronic Calorimeter (HCAL)

Hadronic calorimeters are particularly important to the detection of jets and particles that leads to unbalance of the total energy (MET). The HCAL is installed around

Figure 13 - Scheme of the positions of the HB towers.



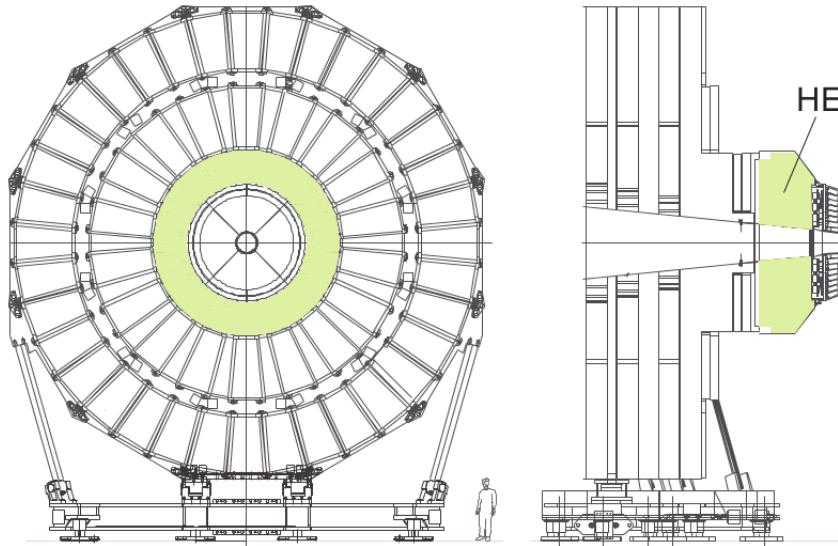
Source: Reference [34], p. 125.

the ECAL in the CMS and its profile was strongly guided by the behavior of the magnetic field (since the biggest part of the HCAL is inside the solenoid). An important requirement for this detector is the enhancement on the resolution of the energy deposited on it and provide good hermeticity for correct measurement of MET. This is achieved by the addition of an extra layer positioned after the solenoid, such that, the CMS HCAL can be divided in three parts: the barrels inside (HB - *Hadron Barrel*) and outside (HO - *Hadron Outer*) of the solenoid and, the endcaps (HE - *Hadron Endcap*). In the barrel the HCAL goes radially from the ECAL ($R = 1.77$ m) up to the solenoid ($R = 2.95$ m).

The HB is composed by 36 towers divided in two halves with respect to the ***z*** axis (Fig. 13) and covers the region with $|\eta| < 1.3$. The towers are made of absorbent brass, constituted of 70% of Cu and 30% of Zn (density of 8.53 g/cm^3). There are inner and outer plates made of stainless steel that holds the towers. The complete HB towers have, then, the following structure: a steel plate with thickness of 4 cm, followed by eight brass plates with thickness of 5.1 cm and six brass plates of 5.7 cm, finishing with another steel plate with 7.5 cm. The effective HB thickness increases with the polar angle (since the particle trajectory increases) and the presence of the ECAL before the HCAL adds about $1.1 \lambda_I$ de material [34, 41].

The HE covers a region with $1.3 < |\eta| < 3$, which contains in general about 34% of the particles produced as final states. The material used on its construction was constrained by the conditions of high luminosity, the capacity of high rates counting and the high magnetic field. The chosen material for this detector was the so called C26000, which is a compound also based in brass. Each HE is coupled to the support of the muon detection system in the CMS endcaps and constitutes 36 towers positioned behind the EE (Fig. 14). The geometry used in the HE construction has the goal to reduce

Figure 14 - Scheme of the hadronic calorimeter in the endcaps, annexed to the muon detection system.



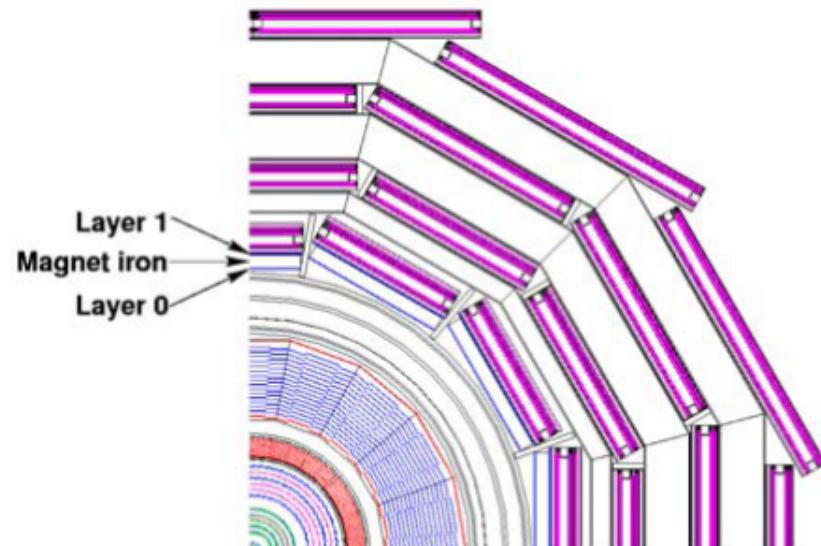
Source: Reference [34], p. 131.

the discontinuities between it and the HB, besides aiming a better energy resolution of a single particle. The used brass plates have thickness of 7.9 cm and present between them plastic scintillators with ~ 9 mm.

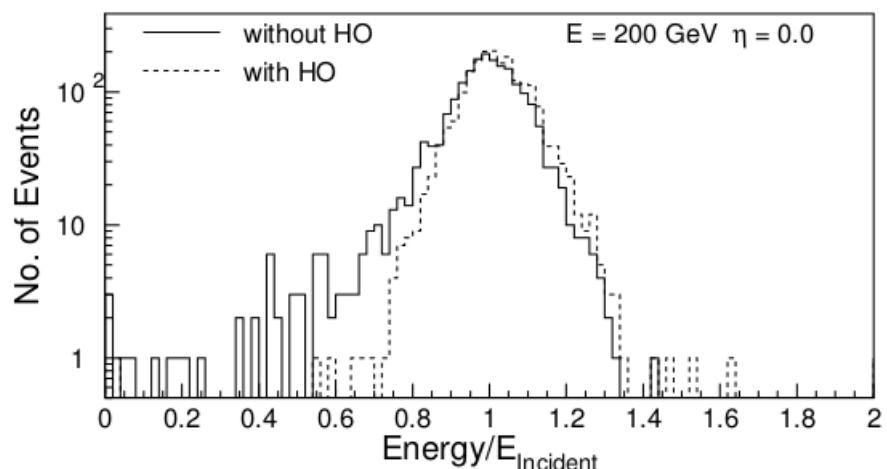
Because of the radial limitation associated to the solenoid there is an additional HCAL layer positioned behind it. This is needed since the combined absorption power of the EB and the HB is not enough to fully handle the some hadronic showers. The HO covers a region of $|\eta| < 1.3$ and uses the solenoid as additional material, being able to detect showers with retarded start and thus, the reminiscent energy after the HB. The physical impact of the HO has been studied in CMS simulation. Without it one observes an excess in the ratio $E_{medida}/E_{incidente} < 1$ (Fig. 15), which means that, part of the total energy is lost in the reconstruction (showers "leakage- the showers are not completely reconstructed). Such effect has a direct impact on the measure of MET and thus, the HO is an important detector for the study of processes involving strong interactions [34, 40].

In the forward CMS directions there are also the HF (Hadron Forward) detectors. The HF are calorimeters which receives very high fluxes of particles (on average 760 GeV per pp interaction is deposited into the two forward calorimeter - very small compared to the 100 GeV for the rest of CMS). Also, such energy is not uniformly distributed but increases with higher pseudorapidities. The HF are essentially a cylindrical steel absorber structure composed by 5 mm thick grooved plates. Quartz Optical fibers are inserted in such grooves. The detector are functionally divided into two longitudinal segments. Half of the fibers run over the full depth of the absorber (165 cm) while the other half starts at a depth of 22 cm from the front of the detector. These two set of

Figure 15 - Scheme of the positions of HO in the muon detection system and its effect on the measurement of the energy from hadronic showers.



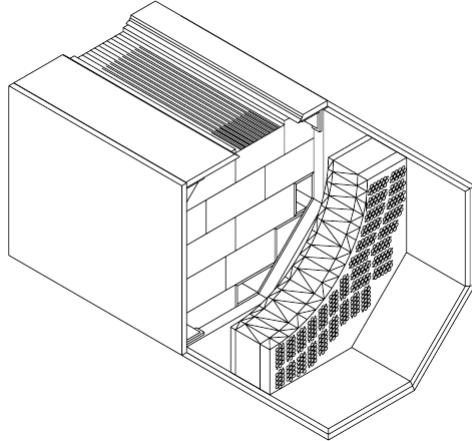
(a) HO transverse section.



(b) HO effect on the balance of energy measured in the CMS.

Source: Reference [34], p. 138 e 140.

Figure 16 - An HF quadrant showing the modules of steel plates with grooves where optical fibers sit.



Source: Reference [40], p. 138 e 140.

fiver have separated read out systems. Such arrangement makes it possible to distinguish between showers generated by electrons and photons (which deposit a large fraction of their energy in the first 22 cm), from the showers generated by hadrons (which produce nearly equal signals in both calorimeter segments). The HFs are located at about 11.2 m from the collision point and their inner radius is at 12.5 cm from the beam line center. The outer radius goes up to 130.0 cm. The total coverage of the HFs comprehend the region $3.0 < |\eta| < 5.0$ [34, 40]. Fig. 16 shows a scheme of a quadrant of the HF detector and the grooves where the optical fibers sit.

3.2.3 The Magnetic System

The CMS magnetic system has the main goal the efficiency in the detection muon system and thus, also the power to bend such particles in the momentum range of ~ 1 TeV in order to distinguish their charge. That requires a resolution of $\Delta p/p \sim 10\%$ at this given momentum. For this reason, the CMS Collaboration choice was a big superconducting solenoid with 12.9 m of length, consisting of 5 rings of 5.9 m of inner radius (Fig. 17). Because of the huge number of wire turns need to achieve the required magnetic field, the solenoid coil is composed by 4 layers, differing from the usual one and two layers, as for Aleph and Delphi and, ZEUS and BaBar experiments, respectively. In total CMS has 2168 wire turns. The solenoid and the wires are made of NbTi and aluminum of high purity. Due to the huge magnetic field, a big part of the solenoid has structural functionality.

Because of physical reasons the radial extension of the solenoid is relatively small

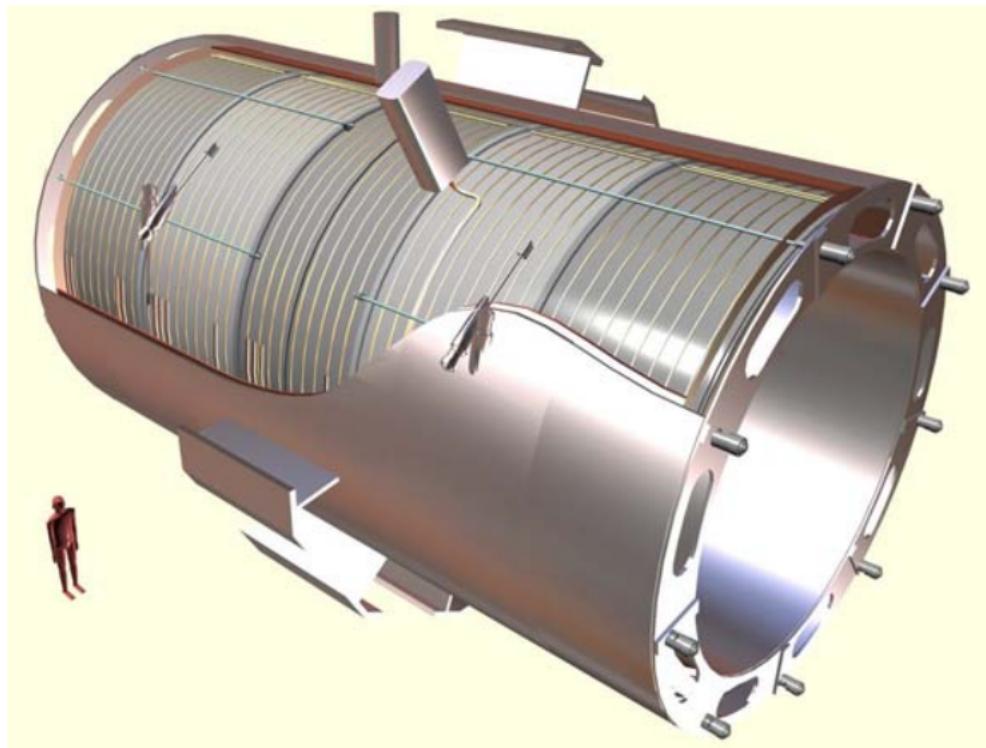
$(\Delta R/R \sim 0.1)$ but even so, its mass is about 220 tons and it can holds up to 2.6 GJ of energy. In order to keep the solenoid superconducting a huge cryogenic system is needed. This system was constructed with a cooling capacity of 800 W at 4.45 K and 4500 W between 60-80 K and, simultaneously for a liquefaction capacity of 4 g/s. Associated to it, there is also a vacuum system capable of produce good isolation for a volume of 40 m³. In order to assure the return of the magnetic field generated by the solenoid, a big structure made of iron (*yoke*) is used in the external region around the solenoid. Such structure is composed of 11 big sections: 6 discs (endcaps) and 5 cylinders (barrel region), with masses that vary between 400 and 1920 tons. For a precise moving of such structures (during the CMS maintenances, for instance), they sit over platforms that float over air or oil (*heavy-duty air pads* and *grease pads*). Because of this system the alignment of the parts composing the *yoke* was done in such a good way that the final accuracy was of 2 mm of difference from the ideal center (the center of the coordinates system) [34].

3.2.4 The Muon Chambers

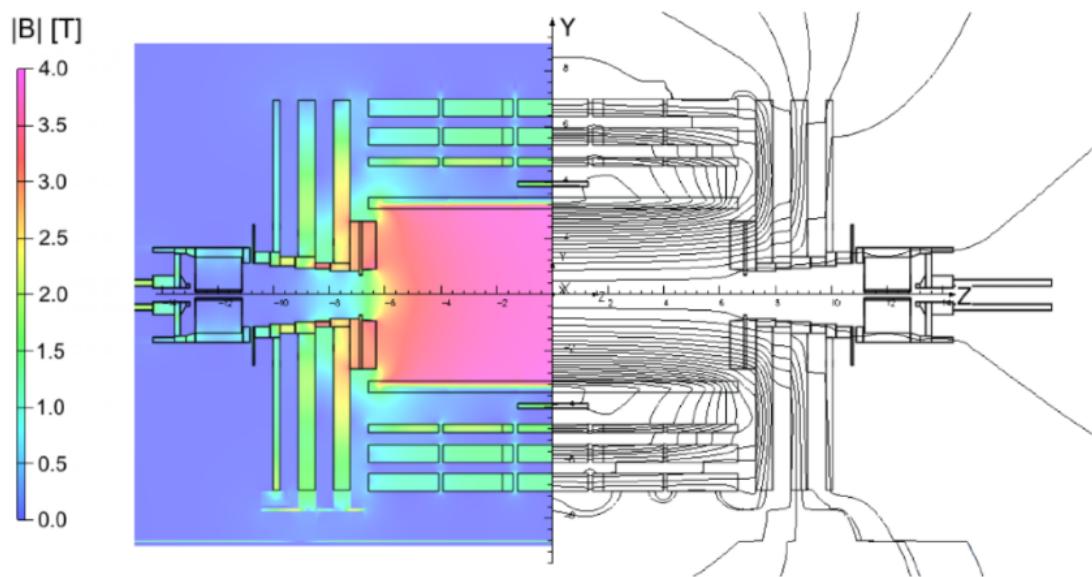
This is the last and outermost set of detectors in the CMS experiment. It is constituted of 4 radial stations inserted in cavities of the *yoke* and is responsible for identify and measure the transverse momentum of muons. The muon chambers are divided in three categories: the drift tube chambers (DTCs), the Cathode Strip Chambers (CSCs) and the Resistive Plate Chambers (RPCs). The DTCs are located in the barrel region, while the CSCs in the endcaps and the RPCs are installed in both barrel and endcaps. Such detectors covers the region of $0 < |\eta| < 2.4$. In the barrel the DTs cover $0 < |\eta| < 1.3$ while in the endcaps the CSCs cover $0.9 < |\eta| < 2.4$. The measurement of the muons transverse momentum using only the muon chambers is determined essentially by the curvature angle of the muons coming out of the solenoid and is not efficient as the tracker system. However, the combination of the tracker with the muon chambers can increases the resolution on the transverse momentum measurement by a factor of up to 10x (Fig. 18) [41, 43].

In the CMS barrel region, the muon chambers are composed by 4 stations that form concentric cylinders positioned radially at 4.0, 4.9, 5.9 and 7.0 m of the beam line center (totalizing 250 chambers). The geometry of the *yoke* presents 12 sectors covering 30° in ϕ and the chambers are disposed in such a way that a high momentum muon produced close to the borders of adjacent sectors crosses at least 3 or 4 chambers. There are 12 chambers in each of the 3 inner stations, while in the fourth station, the sectors above and below (hemispherically) hold 2 chambers. Each station was designed to produce a spacial vector associated to the muons with a precision em ϕ better than 100 μ m. The muon chambers in the endcaps is constituted of 468 CSCs in trapezoidal shape and are superposed between each other, covering the full azimuthal region (orthogonal to the beam

Figure 17 - The CMS magnetic system.



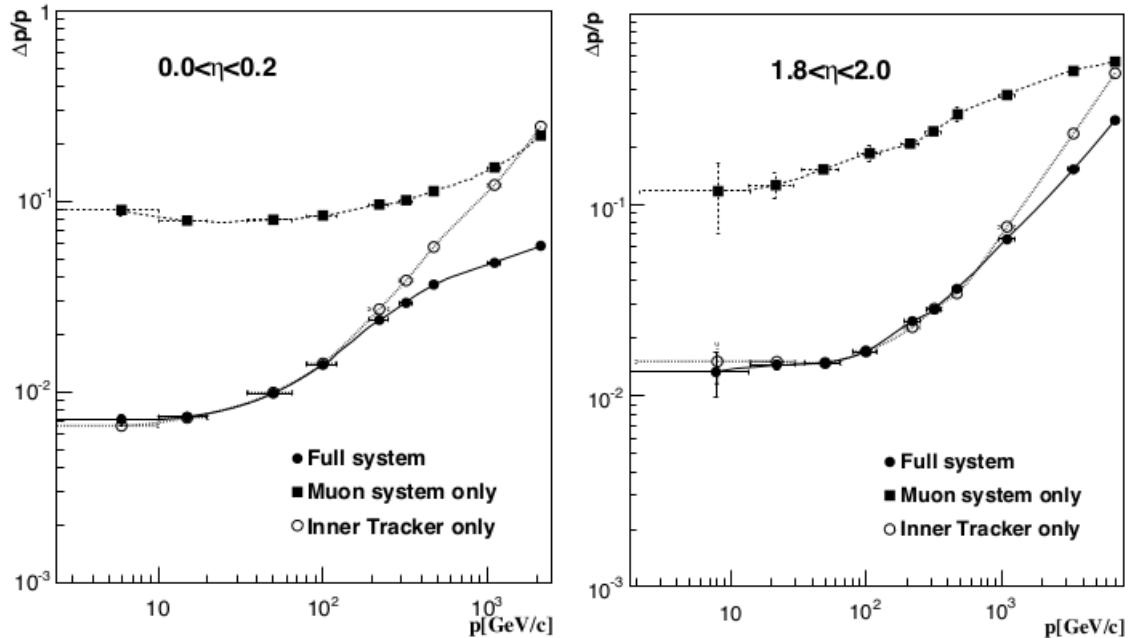
(a) Scheme of the CMS solenoid.



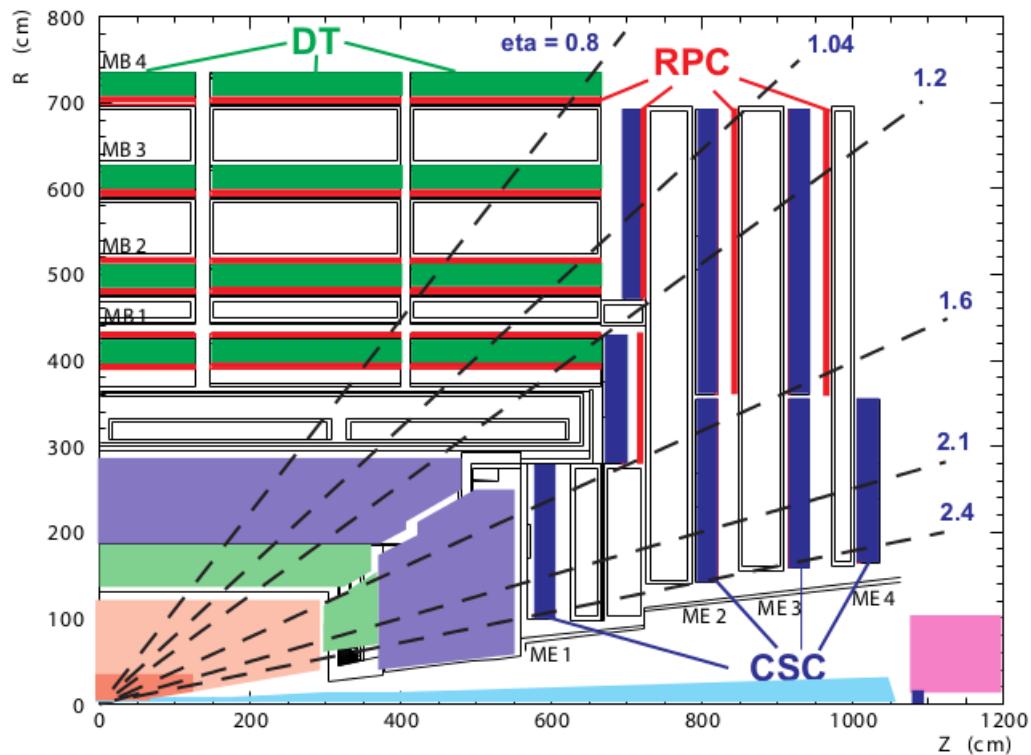
(b) Mapping of the magnetic field created by the solenoid.

Source: Reference [34] p. 7 and [] p. 30.

Figure 18 - Comparison of the resolution on the muon transverse momentum measurement obtained by using the muon chambers and the tracker separated and combined, left. Scheme of the muon chambers position in the CMS, right



(a) Resolutions on the measurement of the muons transverse momentum.



(b) Location of the muon chambers in a section of the CMS detector.

Source: Reference [41], p. 11 and 12.

axis). The chambers in that region has thickness of 60-10 cm, and each one covers about 10° (radially outer chambers) and 20° (radially inner chambers). All RPCs installed in the endcaps cover 20° .

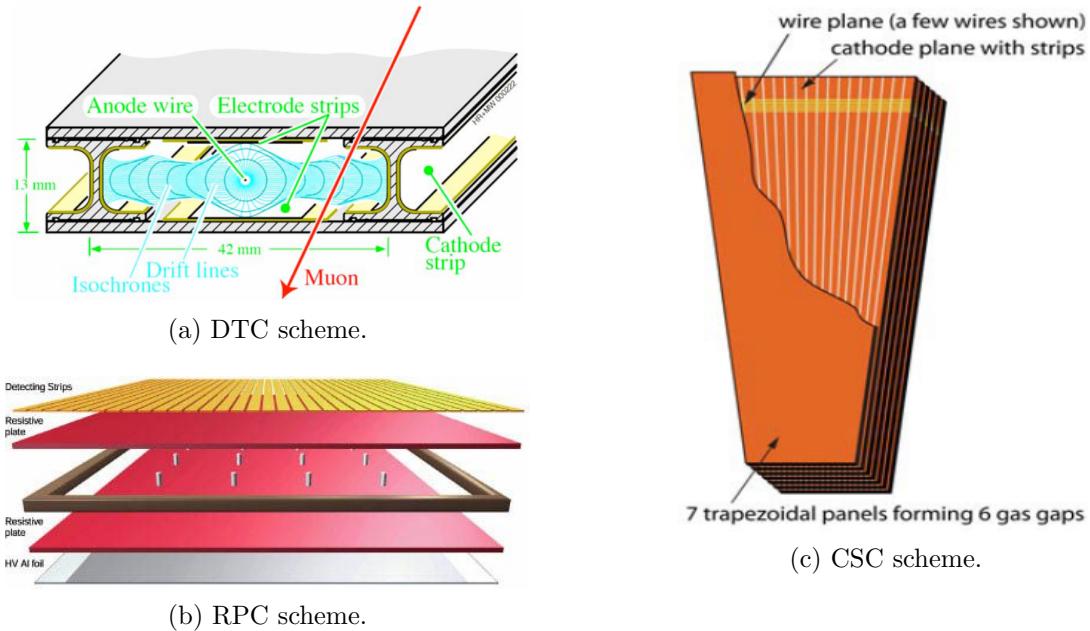
Although there are structural differences between the technologies used in each type of chamber, all of them work in similar way (cathode-anode principle). The DTCs are made of tubes containing gas ($Ar+CO_2$) in which there is a conducting wire Fig. 19(a). A potential difference created between the tube and the wire creates electric field such that, any charged particle crossing the tube ionizes the gas and produces a bunch of electrons (which can produce even more electrons - avalanche process) that are drifted to the wire and produces an excess of current in the detector (the signal). The distance between the particle and the wire is a function of the velocity and the time associated to the electrons in order to achieve the wire. Each DTC is composed of 12 layers of aluminum organized in 3 groups of 4, containing up to 60 tubes each one. Each layer measures in average $2 \times 2.5 \text{ m}^2$ and each tube is 4 cm wide. The CSCs consist of wires perpendicularly positioned to cooper fibers inserted in a gas Fig. 19(c). The difference from the DTCs are basically the higher resolution due to the use of cathodes in strip shape, which allows higher precision on the estimation of the electrons avalanche position. The RPCs are detectors composed by two parallel plates made of a plastic material with high electric resistivity and separated by gas Fig. 19(b). The plates are transparent to the electrons generated in the ionization process and those are collected by metallic fibers located after the plates. RPCs are detectors that combine a good spatial and temporal (1 ns) resolution and, because of that they are used to trigger the presence of particles in the muon detection system [41–43].

3.2.5 The Trigger and Data Acquisition System

The trigger system and data acquisition (TriDAS - *Trigger and Data Acquisition System*) in a hadronic collider has a very important function, since, both the collision rates and the amount of data produced are much larger than the speed and size which can be handled. At the LHC, about 1 MB of data can be produced every crossing of the proton bunches. In such way a system that reduces the amount of data at the moment of the collision is needed. Much of what is produced in the LHC is already known Physics and very efficient filters need to be applied in order to store the Physics potentially new. Since the event rate is very high ($\mathcal{O}(10^6)$), the filtering is divided into two stages: the level-1 trigger (L1T) and the high level trigger (HLT) [34, 44].

The L1T is composed by custom programmed hardware in (or very close to) the sub-detectors and is designed to reduce the event rate to 100 kHz. This trigger can be divide into local, regional and global components, which use informations coming from

Figure 19 - Types of detectors used in the muon detection system.



Source: Reference [34], p. 169 and [41], p. 94.

calorimeters and muon chambers to identify and organize specific objects (*trigger object* - such as the EG candidates) according to their energy/momenta and the quality of such informations (which constitutes the confidence level). In 2016 there was a list (L1 trigger menu) of about 200 requirements (L1 seeds), which constitute specifications that an event should satisfy. In order to be forward for further analysis, an event has to fire the L1 seeds. If that happens then the event is sent to the next level, the HLT [34, 44].

The HLT is a more complex filter based on pure software. It filters the events using a set of algorithms running on a computer farm. These algorithms are a simplified version of the algorithms used for event offline reconstruction. The HLT is designed to reduce the event rate to 100 Hz. In order to be stored an event needs to satisfy at least one of the HLT requirements (trigger paths). When that happens the event is marked for permanent storage and it is transferred to the CERN T0. Once an event is triggered by the HLT and stored in the CERN T0 it follows to the full event reconstruction. This processing is done offline (no time constraints) using the complete resolution and full information of the CMS sub-detectors and, takes at least 48h to be complete. After that, the event is ready for Physics analysis [34, 44].

4 DATA AND MONTE CARLO SAMPLES

4.1 Data

The data samples used in this analysis were recorded during 2016 and correspond to 35.9fb^{-1} of certified data (JSON⁸: Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt). The list of datasets is in Tab. 4 along with the integrated luminosity. The analysis relies on five different primary datasets (PDs), DoubleEG, DoubleMuon, MuEG, SingleElectron, and SingleMuon, each of them combines a certain collections of HLT paths. In order to avoid duplicate events from different PDs, the events are taken:

- from DoubleEG if they pass the diElectron or triElectron triggers;
- from DoubleMuon if they pass the diMuon or triMuon triggers and fail the diEle and triEle triggers;
- from MuEG if they pass the MuEle or MuDiEle or DiMuEle triggers and fail the diEle, triEle, diMuon and triMuon triggers;
- from SingleElectron if they pass the singleElectron trigger and fail all the triggers above;
- from SingleMuon if they pass the singleMuon trigger and fail all the triggers above.

The High Level Trigger (HLT) paths used are listed in Tab. 5 together with their L1 seed, prescale value and the associated PD. The trigger efficiency measured using $4l$ events is found to be larger than 99% for each of the three final states [77].

4.2 Simulated Samples

In this analysis Higgs bosons produced via (VBF) are the signal, while the other production modes and SM backgrounds constitute the total background. The simulation of the SM Higgs is obtained via POWHEG V2 [58, 60] generator for the five main production modes: gluon fusion (ggH) including quark mass effects [61], vector boson fusion (VBF) [62], and associated production (WH, ZH, and ttH [63]). For ggH its MiNLO HJJ extension is used, while in the case of WH and ZH the MiNLO HVJ is used [64]. Higgs

⁸ The JSON file contains the official runs and luminosity section which should be consider when processing the datasets for a physics analysis.

Table 4 - Datasets used in the analysis.

Run range	Datasets	Integ. luminosity
	Global Tag: 80X_dataRun2_2016SeptRepro_v7	
273150-275376	/DoubleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD /DoubleEG/Run2016B-03Feb2017_ver2-v2/MINIAOD /MuonEG/Run2016B-03Feb2017_ver2-v2/MINIAOD /SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD /SingleElectron/Run2016B-03Feb2017_ver2-v2/MINIAOD	5.892 fb ⁻¹
275656-276283	/DoubleMuon/Run2016C-03Feb2017-v1/MINIAOD /DoubleEG/Run2016C-03Feb2017-v1/MINIAOD /MuonEG/Run2016C-03Feb2017-v1/MINIAOD /SingleMuon/Run2016C-03Feb2017-v1/MINIAOD /SingleElectron/Run2016C-03Feb2017-v1/MINIAOD	2.646 fb ⁻¹
276315-276811	/DoubleMuon/Run2016D-03Feb2017-v1/MINIAOD /DoubleEG/Run2016D-03Feb2017-v1/MINIAOD /MuonEG/Run2016D-03Feb2017-v1/MINIAOD /SingleMuon/Run2016D-03Feb2017-v1/MINIAOD /SingleElectron/Run2016D-03Feb2017-v1/MINIAOD	4.353 fb ⁻¹
276831-277420	/DoubleMuon/Run2016E-03Feb2017-v1/MINIAOD /DoubleEG/Run2016E-03Feb2017-v1/MINIAOD /MuonEG/Run2016E-03Feb2017-v1/MINIAOD /SingleMuon/Run2016E-03Feb2017-v1/MINIAOD /SingleElectron/Run2016E-03Feb2017-v1/MINIAOD	4.117 fb ⁻¹
277932-278808	/DoubleMuon/Run2016F-03Feb2017-v1/MINIAOD /DoubleEG/Run2016F-03Feb2017-v1/MINIAOD /MuonEG/Run2016F-03Feb2017-v1/MINIAOD /SingleMuon/Run2016F-03Feb2017-v1/MINIAOD /SingleElectron/Run2016F-03Feb2017-v1/MINIAOD	3.186 fb ⁻¹
278820-280385	/DoubleMuon/Run2016G-03Feb2017-v1/MINIAOD /DoubleEG/Run2016G-03Feb2017-v1/MINIAOD /MuonEG/Run2016G-03Feb2017-v1/MINIAOD /SingleMuon/Run2016G-03Feb2017-v1/MINIAOD /SingleElectron/Run2016G-03Feb2017-v1/MINIAOD	7.721 fb ⁻¹
	Global Tag: 80X_dataRun2_Prompt_v16	
281207-284068	/DoubleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD /DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD /MuonEG/Run2016H-03Feb2017_ver2-v1/MINIAOD /SingleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD /SingleElectron/Run2016H-03Feb2017_ver2-v1/MINIAOD /DoubleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD /DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD /MuonEG/Run2016H-03Feb2017_ver3-v1/MINIAOD /SingleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD /SingleElectron/Run2016H-03Feb2017_ver3-v1/MINIAOD	8.857 fb ⁻¹

Source: Adapted from [77], p. 12.

Table 5 - Trigger paths used in 2016 collision data.

HLT path	L1 seed	Prescale	Primary dataset
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_15_10	1	DoubleEG
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_22_10	1	DoubleEG
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL_(Multiple)	(Multiple)	1	DoubleEG
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	L1_TripleEG_14_10_8	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_TripleMu_12_10_5	L1_TripleMu_5_5_3	1	DoubleMuon
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG15	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG20	1	MuonEG
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu12_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu20_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	L1_SingleMu*	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	L1_Mu6_DoubleEG10	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	L1_DoubleMu7_EG7	1	MuonEG
HLT_Ele25_eta2p1_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_eta2p1_WPLoose_Gsf	L1_SingleEG*	1	SingleElectron
HLT_IsoMu20 OR HLT_IsoTkMu20	L1_SingleMu*	1	SingleMuon
HLT_IsoMu22 OR HLT_IsoTkMu22	L1_SingleMu*	1	SingleMuon

Source: Reference [77], p. 13.

boson decay into ZZ and subsequently into four leptons is obtained via the JHUGEN generator [65]. In the case of WH , ZH , and $t\bar{t}H$, the Higgs boson is allowed to decay as $H \rightarrow ZZ \rightarrow 2l2X$ such that 4-lepton events where two leptons originate from the decay of associated Z or W bosons (or yet top quarks) are also taken into account in the simulation. The SM Higgs boson is also simulated when produced through ggH, VBF and associated production (WH , ZH , $b\bar{b}H$), while decaying as $H \rightarrow WW \rightarrow 2l2\nu$ (in the case of $b\bar{b}H$ the MadGraph5 (aMC@NLO) generator is used). The parton showering and the hadronization are carried out by PYTHIA8.209. All samples are generated with the NNPDF 3.0 NLO parton distribution functions (PDFs) [66]. The list of SM Higgs samples and their cross sections is shown in Tab. 6.

The ZZ production via $q\bar{q}$ annihilation is generated at NLO using POWHEG V2 [67] and PYTHIA8, with the same settings as for the Higgs signal. As this simulation covers a large range of ZZ invariant masses, dynamical QCD factorization and renormalization scales have been chosen to be equal to m_{ZZ} .

The $gg \rightarrow ZZ$ process is simulated at LO with MCFM [68, 69]. In order to match the $gg \rightarrow H \rightarrow ZZ$ transverse momentum spectra predicted by POWHEG at NLO, the showering for MCFM samples is performed with different PYTHIA8 settings, allowing only emissions up to the parton-level scale ("wimpy" shower).

Additional MC samples of WZ , Drell-Yan+jets, $t\bar{t}$, $t\bar{t}V$, and VVV are generated using MadGraph5 either inclusively or merging several jet multiplicities. Tab. 6 summarizes the MC simulated samples used in this analysis. All samples are processed through GEANT4 [70, 71] simulating the CMS detector and then, reconstructed through the official production chain. The Monte Carlo samples are also re-weighted to match the pileup distribution observed in 2016 data.

Table 6 - MC simulated samples and their respective cross section.

Process	Dataset Name	SM Higgs MC Samples	$\sigma \cdot BR (fb)$
$gg \rightarrow H \rightarrow ZZ \rightarrow 4l$	[1] GluGluHToZZTo4L_M125_13TeV-powheg2_minloHJJ_JHUgenV6_pythia8	12.180	
$q\bar{q} \rightarrow H q\bar{q} \rightarrow ZZ q\bar{q} \rightarrow 4l q\bar{q}$	[1] VBF_HToZZTo4L_M125_13TeV-powheg2_JHUgenV6_pythia8	1.044	
$q\bar{q} \rightarrow W^+ H \rightarrow W^+ ZZ \rightarrow 4l + X$	[1] WplusH_HToZZTo4L_M125_13TeV-powheg2_minlo-HWJ_JHUgenV6_pythia8	0.232	
$q\bar{q} \rightarrow W^- H \rightarrow W^- ZZ \rightarrow 4l + X$	[1] WminusH_HToZZTo4L_M125_13TeV-powheg2_minlo-HWJ_JHUgenV6_pythia8	0.147	
$q\bar{q} \rightarrow ZH \rightarrow ZZZ \rightarrow 4l + X$	[1] ZH_HToZZ_4LFILTER_M125_13TeV_V_powheg2_minlo-HZJ_JHUgenV6_pythia8	0.668	
$q\bar{q} \rightarrow t\bar{t}H \rightarrow t\bar{t}ZZ \rightarrow 4l + X$	[1] ttH_HToZZ_4LFILTER_M125_13TeV_V_powheg2_minlo-HZJ_JHUgenV6_pythia8	0.393	
$gg \rightarrow H \rightarrow WW \rightarrow 2l2\nu$	[1] GluGluHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8	1101.790	
$q\bar{q} \rightarrow H q\bar{q} \rightarrow WW q\bar{q} \rightarrow 2l2\nu$	[1] VBFHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8	85.776	
$q\bar{q} \rightarrow W^+ H \rightarrow W^+ WW \rightarrow l\nu 2l2\nu$	[1] HWplusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8	2.138	
$q\bar{q} \rightarrow W^- H \rightarrow W^- WW \rightarrow l\nu 2l2\nu$	[1] HWminusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8	1.357	
$q\bar{q} \rightarrow ZH \rightarrow ZWW \rightarrow 2l2l2\nu$	[1] HZJ_HToWWTo2L2Nu_ZTo2L_M125_13TeV_powheg_pythia8	2.029	
$gg \rightarrow b\bar{b}H \rightarrow b\bar{b}WW \rightarrow 2l2\nu + X$	[1] bbHToWWTo2L2Nu_M-125_4FS_yb2_13TeV_amcatnlo	11.068	
SM Backgrounds MC Samples			
$q\bar{q} \rightarrow ZZ \rightarrow 4l$	[2] ZZTo4L_13TeV_powheg_pythia8	1256.000	
$q\bar{q} \rightarrow ZZ \rightarrow 4l$	[2] ZZTo4L_13TeV-amcatnloFXFX_pythia8	1212.000	
$q\bar{q} \rightarrow ZZ \rightarrow 4l + jets(EWK)$	[2] ZZJHTo4L_EWK_13TeV-madgraph_pythia8/	4.404	
$gg \rightarrow ZZ \rightarrow 4e$	[2] GluGluToContinToZZTo4e_13TeV_MCFM701	1.590	
$gg \rightarrow ZZ \rightarrow 4\mu$	[2] GluGluToContinToZZTo4mu_13TeV_MCFM701	1.590	
$gg \rightarrow ZZ \rightarrow 4\tau$	[2] GluGluToContinToZZTo4tau_13TeV_MCFM701	1.590	
$gg \rightarrow ZZ \rightarrow 2e2\mu$	[2] GluGluToContinToZZTo2e2mu_13TeV_MCFM701	3.190	
$gg \rightarrow ZZ \rightarrow 2e2\tau$	[2] GluGluToContinToZZTo2e2tau_13TeV_MCFM701	3.190	
$gg \rightarrow ZZ \rightarrow 2\mu2\tau$	[2] GluGluToContinToZZTo2mu2tau_13TeV_MCFM701	3.190	
$Z \rightarrow ll + jets$	[2] DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8	6.104e ⁶	
	[2] DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8	1.861e ⁷	
$t\bar{t}$	[2] TTJets_TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8	815.96e ⁵	
$t\bar{t} \rightarrow 2l2\nu 2b$	[2] TTTTo2L2Nu_13TeV-powheg	8.731e ⁴	
$WZ \rightarrow 3l\nu$	[2] WZTo3LNu_TuneCUETP8M1_13TeV-powheg_pythia8	4430	
ZZZ	[2] ZZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	13.980	
WWZ	[2] WWZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	165.100	
WZZ	[2] WZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	55.650	
$t\bar{t}W$	[2] TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin_pythia8	204.300	
$t\bar{t}Z$	[2] TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo_pythia8	252.900	

[1] RunIIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheV_V_v6-v1/
[2] RunIIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheV_V_v6-v1/

Source: Adapted from [77] p(s). 14 and 15 and, [78] p. 3.

5 PHYSICS OBJECTS SELECTIONS

The present analysis is based on the following physics objects: electrons, muons, missing transverse energy (E_T^{miss}), and jets. In the following subsections the main selection criteria for such object is presented while a detailed and complete description can be found in [72, 73]. The selection (reco, ID and isolation) of objects is the same as in the SM $H \rightarrow ZZ \rightarrow 4l$ analysis.

5.1 Electrons

Electrons are required to have transverse momentum $p_T > 7\text{GeV}$, $|\eta| < 2.5$, and to satisfy a loose primary vertex (PV) constraint defined as $|d_{xy}| < 0.5\text{cm}$ and $|d_z| < 1\text{cm}$. Such electrons are referred to as *loose electrons*. The early runs in 2016 data-taking exhibit a tracking inefficiency originating from a reduced hit reconstruction efficiency in the strip detector ("HIP" effect). The resulting data-MC discrepancy is corrected using scale factors as it is done for the electron selection with efficiencies measured in data using the same tag-and-probe technique outlined later. These studies are carried out by the EGM POG and electron reconstruction (tracking) scale factors as a function of the super cluster η are derived and used for this analysis (it was shown that the p_T dependence of the scale factor is negligible). More details on electron reconstruction can be found in [74].

Reconstructed electrons are identified by means of a Gradient Boosted Decision Tree (GBDT) multivariate classifier algorithm, which exploits observables from the electromagnetic cluster, the matching between the cluster and the electron track as well as observables based exclusively on tracking measurements. The BDT has been retrained using *CMSSW_8_0_X* samples. The classifier is trained on a $DY + \text{jets}$ MC sample for both signal and background. Tab. 7 summarizes the full list of observables used as input to the classifier and Tab. 8 lists the cut values applied to the BDT score for the chosen working point. For the analysis, we define tight electrons as the loose electrons that pass this MVA identification working point.

Electrons are required to be isolated and the relative isolation is defined as

$$RelPFiso = \frac{\left(\sum_{\text{charged}} p_T + \sum_{\text{neutral}}^{\text{corr}} p_T \right)}{p_T^{\text{lepton}}} \quad (36)$$

Table 7 - Overview of input variables to the electron identification classifier. Variables not used in the RunI MVA are marked with *.

Observable	Observable name
Cluster shape	RMS of the energy-crystal number spectrum along η and ϕ ; $\sigma_{i\eta i\eta}$, $\sigma_{i\sigma i\sigma}$ super cluster width along η and ϕ ratio of the hadronic energy divided by the supercluster energy, H/E circularity $(E_{5\times 5} - E_{5\times 1})/E_{5\times 5}$ sum of the seed and the 9 adjacent crystals divided by the supercluster energy, R_9 for Endcap training bins: energy fraction in pre-shower, E_{PS}/E_{raw}
Track-cluster matching	energy-momentum agreement E_{tot}/p_{in} , E_{ele}/p_{out} , $1/E_{tot} - 1/p_{in}$ position matching $\Delta\eta_{in}$, $\Delta\phi_{in}$, $\Delta\eta_{seed}$
Tracking	fractional momentum loss $f_{brem} = 1 - p_{out}/p_{in}$ number of hits of the KF and GSF track N_{KF} , N_{GSF} * reduced χ^2 of the KF and GSF track χ^2_{KF} , χ^2_{GSF} number of expected but missing inner hits * probability transform of conversion vertex fit χ^2 *

Source: Reference [77], p. 16.

Table 8 - Minimum BDT score required for passing the electron identification.

Minimum BDT score	$ \eta < 0.8$	$0.8 < \eta < 1.479$	$ \eta > 1.479$
$5 < p_T < 10 GeV$	-0.211	-0.396	-0.215
$p_T > 10 GeV$	-0.870	-0.838	-0.763

Source: Reference [77], p. 16.

where the corrected neutral component of isolation is computed using the formula

$$\sum_{neutral}^{corr} p_T = \max\left(\sum_{neutral}^{uncorr} p_T - \rho \times A_{eff}, 0\right) GeV \quad (37)$$

and the mean pile-up contribution to the isolation cone is obtained by

$$PU = \rho \times A_{eff} \quad (38)$$

where ρ is the mean energy density in the event and A_{eff} the effective area which is defined as the ratio between the slope of the average isolation and the slope of ρ as a function of the number of vertices. The electron isolation working point was optimized in [76] and was chosen as $RelPFiso(\Delta R = 0.3) < 0.35$.

Electrons in data are also corrected for features in ECAL energy scale in bins of p_T and $|\eta|$. Corrections are calculated on a $Z \rightarrow ee$ sample to align the di-electron mass spectrum in the data to the one in simulation and to minimize the width of the distribution.

The $Z \rightarrow ee$ mass resolution in simulation is made to match data by applying a pseudo-random Gaussian smearing to electron energies, with Gaussian parameters varying in bins of p_T and $|\eta|$. This has the effect of convolving the electron energy spectrum with a Gaussian. The electron energy scale is measured in data by fitting a Crystal-Ball function to the di-electron mass spectrum around the Z peak in the Z+jets control region.

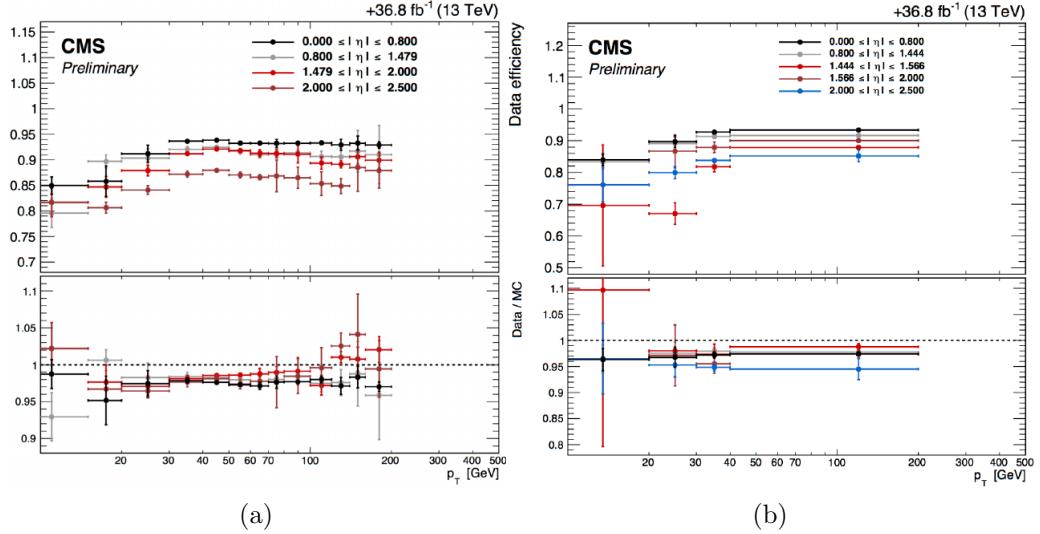
5.1.1 Electron Efficiency Measurements

The tag-and-probe (T&P) study was performed on the single electron primary datasets listed in Tab. 4 using a JSON corresponding to $36.8 fb^{-1}$. More details on the tag-and-probe method can be found in [76]. Tagged electrons need to satisfy the following quality requirements:

- trigger matched to HLT_Ele27_eta2p1_WPTight_Gsf_v*;
- $p_T > 30 GeV$, super cluster (SC) $|\eta| < 2.1$ but on in EB-EE gap ($1.4442 < |\eta| < 1.566$);
- "tight" working point of the Spring16 cut-based electron ID.

Probe electrons only need to be reconstructed as GsfElectron. The Final State Radiation (FSR) recovery algorithm used in the main analysis is used consistently throughout

Figure 20 - Electron selection efficiencies measured using the tag-and-probe technique for: (a) non-gap electrons and (b) gap electrons.



Source: Reference [77], p. 18.

the efficiency measurement: the isolation is calculated after removing any FSR photon matched to the electron and the electron p_T itself, and the di-electron invariant mass is estimated by including the FSR photons (if any). The nominal MC efficiencies are evaluated from the LO MadGraph DY sample, while the NLO systematics use the 0 and 1 jet MadGraph_AMCatNLO sample listed in Tab. 6.

In contrast to previous efficiency measurements, a template fit is used here. The m_{ee} signal shape of the passing and failing probes is taken from MC and convoluted with a Gaussian. The data are then fitted with the convoluted MC template and a CMSShape (an error-function with a one-sided exponential tail). This change follows from the usage of the new T&P tool developed by the EGM POG. The electron selection efficiency is measured as a function of the probe electron p_T and η , and separately for electrons falling in the ECAL gaps. Fig. 20 shows the p_T turn-on curves measured in data; final scale factors for the efficiencies are derived and used within the analysis. The EGM recommendations on the evaluation of tag-and-probe uncertainties for efficiency measurements are followed.

5.2 Muons

Further details on muon reconstruction can be found in [58]. Two types of selection are performed. We define **loose muons** as the muons satisfying $p_T > 5\text{GeV}$, $|\eta| < 2.4$, $d_{xy} < 0.5\text{cm}$ and $d_z < 1\text{cm}$, where d_{xy} and d_z are defined with respect to the primary

Table 9 - The requirements for a muon to pass the tracker high- p_T ID. Note that these are equivalent to the Muon POG high- p_T ID with the global track requirements removed.

Plain-text description	Technical description
Muon station matching	Muon is matched to segments in at least two muon stations
Good p_T measurement	$p_T/\sigma_{p_T} < 0.3$
Vertex compatibility (x-y)	$d_{xy} < 2\text{mm}$
Vertex compatibility (x)	$d_z < 5\text{mm}$
Pixel hits	At least one pixel hit
Tracker hits	Hits in at least six tracker layers

Source: Reference [77], p. 18.

vertex (PV) and using the *muonBestTrack*. Muons have to be reconstructed by either the *GlobalMuon* or *TrackerMuon* algorithm. Standalone muon tracks reconstructed only in the muon system are rejected. Muons with *muonBestTrackType == 2* (standalone) are also discarded, even if they are marked as global or tracker muons. Loose muons with $p_T < 200\text{GeV}$ are considered *tight muons* if they also pass the PF muon ID (note that the naming convention used for these IDs differs from the muon POG naming scheme, in which the "tight ID" used here is called the "loose ID"). Loose muons with $p_T > 200\text{GeV}$ are considered tight muons if they pass the PF ID or the Tracker High- p_T ID, the definition of which is shown in Tab. 9.

An additional "ghost-cleaning" step is performed to deal with situations when a single muon can be incorrectly reconstructed as two or more muons:

- Tracker Muons that are not Global Muons are required to be "arbitrated", i.e. associated to the segments in the outer muon detectors;
- If two muons are sharing 50% or more of their segments, the muon with lower quality is removed.

Muons are required to be isolated similarly as described for electrons. The pileup contribution subtraction is performed in a different way for muons. A correction $\Delta\beta = \frac{1}{2}(\sum_{PU}^{charged had.} p_T)$, that gives an estimate of the energy deposit from neutral particles coming from pileup vertices, is applied. The relative muon isolation is then defined as

$$RelPFiso = \frac{\sum_{lepton}^{charged had.} p_T + \max(\sum_{lepton}^{neutral had.} E_T + \sum_{lepton}^{photon} E_T - \Delta\beta, 0)}{p_T} \quad (39)$$

The muon isolation was optimized in [76] and the working point is chosen to be equal to the one for electrons, that is, $\text{RelPFiso}(\Delta R = 0.3) < 0.35$.

5.2.1 Muon Efficiency Measurements

Muon efficiencies are also measured with the T&P method, which is then performed on $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ events in bins of p_T and η . More details on the methodology can be found in [76]. The Z sample is used to measure the muon reconstruction and identification efficiency at high p_T , and the efficiency of isolation and impact parameter at p_T . The J/ψ sample is used to measure the reconstruction efficiency at low p_T as it benefits from a better purity in such kinematic regime. Events are collected using HLT_Mu7p5_Track2_Jpsi_v* when probing the reconstruction and identification efficiency in the muon system. For probing the muon tracking efficiency HLT_Mu7p5_L2Mu2_Jpsi_v* is used.

Muon reconstruction and identification efficiency for $p_T > 20\text{GeV}$ have been derived by the Muon POG. The probe in this measurement are tracks reconstructed in the inner tracker, and the passing probes are those that are also reconstructed as a global or tracker muon and passing the Muon POG Loose muon identification. For low p_T muons J/ψ events have been used, with the same definitions of "probe" and "passing probe". The systematic uncertainties are estimated by varying the analytical signal and background shape models used to fit the di-muon invariant mass. Details on the procedure can be found in [53]. The efficiency and scale factors used for low p_T muons are the ones derived using single muon prompt-Reco dataset. The efficiencies in data and in simulation are shown in Fig. 21.

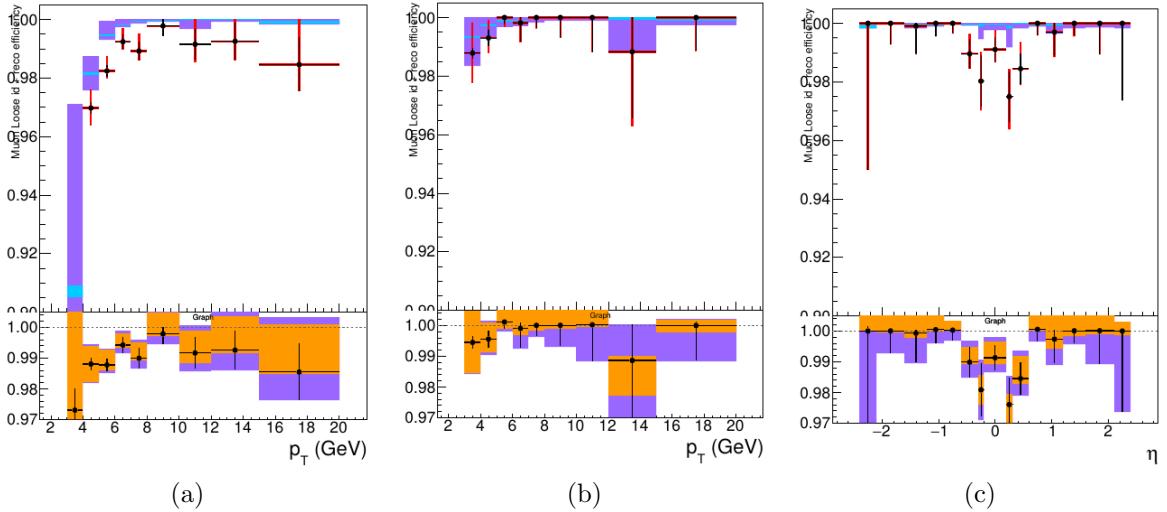
The requirements for the impact parameter is studied using Z events selected with the trigger HLT_IsoMu20_v* or HLT_IsoMu22_v*. For this measurement, the probe is a muon passing the POG "loose" identification criteria, and it is considered a "passing" probe if it satisfies the SIP_{3D} , d_{xy} and d_z cuts of this analysis. The results are shown in Fig. 22.

The muon isolation efficiency is measured using events from the Z decay at any p_T . The events are selected with the same triggers as for the SIP study. The isolation of the muons are calculated after recovery of the FSR photons and subtracting their contribution from the isolation cone of the muons. More detailed description of the method can be found in [54]. Results are shown in Fig. 23.

The efficiency to reconstruct a muon track in the inner detector is measured using tracks reconstructed only in the muon system. The method for measuring the tracking efficiency is the same as in [75] and the results on 2016 data are briefly discussed here. The efficiency and data-to-simulation scale factors are measured from Z events as a function of η . The values of data-to-simulation scale factors used are from the ReReco version of the full dataset collected in 2016. The tracking efficiency in data and in simulation as a function of η is shown in Fig. 24.

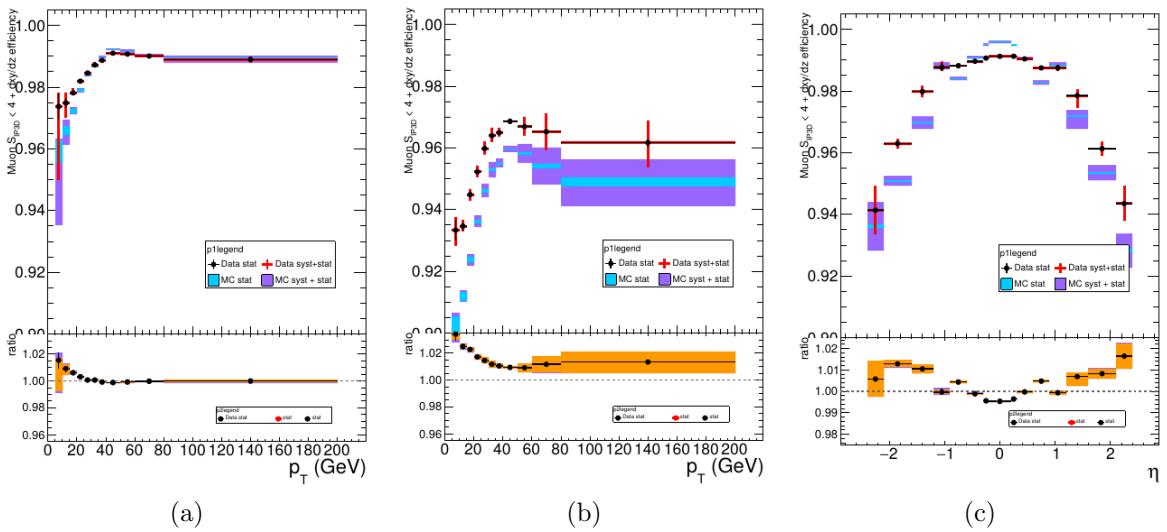
The product of all data-to-simulation scale factors for muon tracking, reconstruction, identification, impact parameter and isolation requirements is used in the analysis.

Figure 21 - Muon reconstruction and identification efficiency for $p_T < 20\text{GeV}$ as a function of p_T in the barrel (a) and endcaps (b), and as a function of η for $p_T > 7\text{GeV}$ (c). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot of the two efficiencies, the black error bars represent the statistical uncertainty, the orange ones the systematical uncertainty, and the violet ones the total uncertainty.



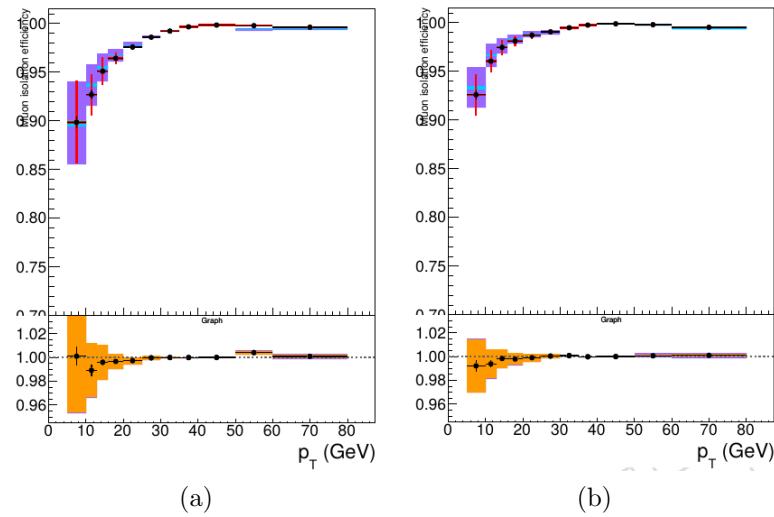
Source: Reference [77], 20.

Figure 22 - Efficiency of the muon impact parameter requirements as a function of p_T in the barrel (a) and endcaps (b), and as a function of η for $p_T > 20\text{GeV}$ (c). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot the black error bars are for the statistical uncertainty, the orange ones for the systematical uncertainty and the violet ones include both uncertainties.



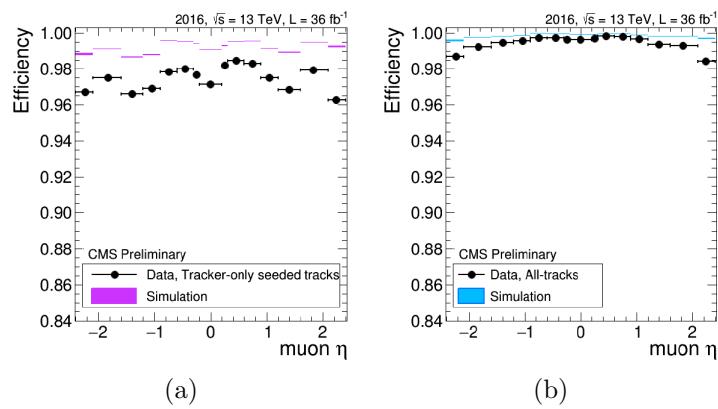
Source: Reference [77], p. 20.

Figure 23 - Efficiency of the muon isolation requirement as a function of p_T in the barrel (a) and endcaps (b). In the upper panel the cyan (black) error bars show only the statistical uncertainty and the red (violet) ones represent the total uncertainty from MC (data). In the ratio plot the black error bars are for the statistical uncertainty, the orange ones for the systematical uncertainty and the violet ones for the total uncertainty.



Source: Reference [77], p. 21.

Figure 24 - Tracking efficiency in data and simulation as a function of η for muon $p_T < 10\text{GeV}$ (a) and $p_T > 10\text{GeV}$ (b) with ReReco data.



Source: Reference [77], p. 22.

5.3 Photons and FSR Recovery

The FSR recovery algorithm was considerably simplified with respect to what was done in RunI, while maintaining a similar performance. The selection of FSR photons is now only done per-lepton and no longer depends on any Z mass criteria, which significantly simplifies the building and selection of ZZ candidates. In the association of photons and leptons the rectangular cuts on $\Delta R(\gamma, l)$ and $E_{T,\gamma}$ have been replaced by a cut on $\Delta R(\gamma, l)/E_{T,\gamma}^2$.

Starting from the collection of photons provided by the PF algorithm (PF photons), the selection of photons and their association to a lepton proceeds as follows:

- 1 The pre-selection of PF photons is done by requiring $p_T^\gamma > 2\text{GeV}$, $|\eta^\gamma| < 2.4$, and a $\text{RelPFiso}(\Delta R = 0.3) < 1.8$. The RelPFiso computation uses a threshold of 0.2GeV on charged hadrons, with a veto cone of 0.0001, and 0.5GeV on neutral hadrons and photons, with a veto cone of 0.01, also including the contribution from pileup vertices (with the same radius and threshold as per-charged isolation);
- 2 Supercluster veto: we remove all PF photons that match with any electron passing both the loose ID and SIP cuts. The matching is performed by directly associating the two PF candidates;
- 3 Photons are associated to the closest lepton in the event among all those that pass both the loose ID and SIP cuts;
- 4 Photons that do not satisfy $\Delta R(\gamma, l)/E_{T,\gamma}^2 < 0.012$ and $\Delta R(\gamma, l) < 0.5$ are discarded;
- 5 If more than one photon is associated to the same lepton, we select the one with lowest $\Delta R(\gamma, l)/E_{T,\gamma}^2$;
- 6 Each selected FSR photon is removed from the isolation sum of all the leptons in the event that pass both the loose ID and SIP cuts. This concerns the photons that are in the isolation cone and outside the isolation veto of said leptons ($\Delta R < 0.4$ AND $\Delta R > 0.01$ for muons, and $\Delta R < 0.4$ AND ($\eta^{SC} < 1.479$ OR $\Delta R > 0.08$) for electrons).

More details about the photon FSR optimization can be found in [72, 76].

5.4 Jets

Jets are reconstructed through the anti- k_T clustering algorithm using PF candidates after rejecting the charged hadrons that are associated to a pileup primary vertex. We use a distance parameter $R = 0.4$.

To reduce instrumental background, the loose working point for the jet identification suggested by the JetMET POG is applied. Jets are required to have $p_T > 30\text{GeV}$ and $|\eta| < 4.7$ and are cleaned from any tight leptons and FSR photons by a separation criterion of $\Delta R(jet, l/\gamma) > 0.4$.

Since the calorimeter response to particles is not linear, Jet Energy Corrections (JEC) are needed to translate the measured jet energy to the true particle/parton energy. Standard JEC are applied on reconstructed jets, which consist of L1 Pileup, L2 Relative Jet Correction, L3 Absolute Jet Correction for both MC samples and data, and also residual calibration for data.

For the purpose of reducing the background coming from $t\bar{t}$ events, the *Combined Secondary Vertex* (CSV) algorithm is used as b-tagging algorithm. It combines information about the SIP, the secondary vertex (SV) and jet kinematics. These variables are combined through a likelihood ratio technique to compute the b-tag discriminator. In this analysis, a jet is considered to be b-tagged if the discriminator $pfCombinedInclusiveSecondaryVertexV2BJetTags > 0.8484$ (i.e. the jet passes the *CSVv2M* "medium" working point) [72, 76]. Data-to-simulation scale factors for b-tagging efficiency are provided for this working point for the full dataset as a function of jet p_T , η and flavour. Such scale factors are applied to simulated jets by downgrading (upgrading) the b-tagging status of a fraction of the b-tagged (untagged) jets that have a scale factor smaller (larger) than one.

5.5 Missing Transverse Energy

The missing transverse energy (MET), E_{pT}^{miss} , of an event consists of the imbalance in the transverse plane with respect to the beam axis. Since the sum of momentum in the transverse plan must be zero, any imbalance is attributed to undetected particles (such as neutrinos) escaping the CMS detector. Raw E_T^{miss} (PFMET) is defined as the magnitude of the negative vectorial sum of all reconstructed PF particles p_T ,

$$\vec{E}_T^{miss} = - \sum_{jetall} \vec{p}_{T,i} \quad (40)$$

An alternative definition of MET, which is called Type-I corrected E_{pT}^{miss} , takes into account the JEC, correcting the MET for detector inefficiencies and non-linear responses

Table 10 - Filters from JETMET POG used to improve signal-to-noise ratio.

Filter	Description
HBHENoiseFilter HBHENoiseIsoFilter	remove noisy events from the HCAL, where the HBHE scintillator produces anomalous signals with pulse shapes and pixel multiplicities discrepant from those from a clean signal
EcalDeadCellTriggerPrimitiveFilter	removes events with non-functioning ECAL data links, comparing the sum of energy deposited in each supercluster cell to the energy saturation of the trigger primitive
goodVertices	filter events with noisy vertex reconstruction (due to pileup effects) by requiring the reconstruction of at least one good vertex full filling the following criteria: high number of degree of freedom ($NPV > 4$), collisions restricted along the z-axis ($zPV < 24\text{cm}$) and small radius of the PV ($rPV < 2\text{cm}$)
eeBadScFilter	removes events with noisy ECAL endcap superclusters
globalTightHalo2016Filter	removes events with enhanced MET from beam-halo particles which are in time with the beam
BadPFMuonFilter BadChargedCandidateFilter	remove events with mis-reconstructed muon and charged hadron PF candidates

Source: Adapted from [77], p. 24.

in the calorimeters. The Type-I corrected MET is given by

$$\vec{E}_T^{miss} = - \left(\sum_{jets} \vec{p}_{T,jet}^{JEC} + \sum_{i \in uncl.} \vec{p}_{T,i} \right) \quad (41)$$

where the contribution of jets ($\vec{p}_{T,jet}^{JEC}$) and the remaining unclustered objects ($\vec{a}_{T,i}$) are accounted separately.

In order to improve signal-to-noise ratio several filters developed in the JETMET POG [72] are used to select the events in this analysis. Such filters are presented in Tab. 10.

6 EVENT SELECTION

The selections performed in this analysis are designed to reconstruct a final state with four charged leptons (4μ , $4e$ or $2e2\mu$) and MET. The four-lepton system must satisfy the SM Higgs selections described in [72].

6.1 Preselection

The events must have fired at least one of the HLT paths presented in Sec. 4. They are also required to pass the MET filters described in Sec. 5.

6.2 Selection of ZZ System

The events with four lepton candidates are selected from what is called *selected leptons*, which are the tight leptons defined in Sec. 5. Such leptons have $SIP_{3D} < 4$ as a vertex constraint and isolation cuts, where the FSR photons are removed from the isolation cone. A lepton cross cleaning, which discards leptons with $\Delta R \geq 0.05$ from each other, is applied.

The building of a ZZ system from four candidate leptons proceeds according to the following sequential steps:

- 1 **Z candidates:** are built from a pair of selected leptons of opposite charge and same flavor (e^+e^- , $\mu^+\mu^-$), having invariant mass satisfying $12 < m_{ll(\gamma)} < 120\text{GeV}$, where the Z candidate mass takes into account (if any) a selected FSR photon;
- 2 **ZZ candidates:** are built from a pair of Z candidates which do not have common leptons (non-overlapping). The Z with closest $m_{ll(\gamma)}$ to the nominal Z boson mass is denoted as Z_1 and the second one is the Z_2 . The built ZZ system must satisfy the following requirements:
 - **Ghost removal:** any two leptons must have $\Delta R(\eta, \phi) < 0.02$;
 - **Lepton p_T :** at least two out of the four leptons must have $p_T^i > 10$ and $p_T^i > 20$ GeV;
 - **QCD suppression:** all opposite-sign lepton pair that can be built out of the four leptons (regardless of lepton flavor) must satisfy $m_{ll} > 4$ GeV. Here, the selected FSR photons are not included in the mass computation, since QCD-induced low mass dilepton (e.g. J/Ψ) may have photons nearby (e.g. from

- $\pi^0)$;
- **Z_1 mass:** $m_{Z_1} > 40$ GeV;
 - **Smart cut:** defining Z_a and Z_b as the mass-sorted alternative pairing Z candidate (Z_a being the closest one to the nominal Z mass), require $\text{NOT}(|m_{Z_a} - m_Z| < |m_{Z_1} - m_Z| \text{ AND } m_{Z_b} < 12)$. Selected FSR photons are included in the m_Z 's computation. This cut discards 4μ and $4e$ candidates which have the alternative pairing similar to an on-shell $Z +$ a low-mass pair l^+l^- ;
 - **Four-lepton invariant mass:** $m_{4l} > 70$ GeV;
 - **Choice of the best ZZ candidate:** if more than one ZZ candidate survives the previous selections, the one with the highest four leptons p_T scalar sum is chosen.
- 3 **SM Higgs selection:** events containing at least one ZZ system satisfying all the previous selections are then used to tag the Higgs decaying into four lepton final state, as it is done for the SM Higgs analysis [72];
- 4 **Signal region (SR):** in order to enhance the presence of VBF events over the other SM Higgs production modes and the SM backgrounds, additional cuts are applied as discussed in the next section.

6.3 VBF Signal Region (VBF-SR)

In order to optimize the analysis for the VBF production mode few additional requirements are imposed to the events selected after the steps described above (until step 3). These additional selections are:

- **Number of jets:** events must have,
 - EITHER, two or three jets from which at most one b-tagged jet;
 - OR, more than three jets with no b-tagged jet;
- **Four-lepton invariant mass:** events must have $118 \leq m_{4l} \leq 130$ GeV since the significant fraction of VBF yields its contained within that range. This m_{4l} range is equal to the SM Higgs signal region adopted by CMS collaboration [86].

Note that these VBF-SR requirements are similar to the standard ones used by CMS Collaboration defining the VBF category. The difference, since the NNs are the discriminants of interest in this analysis, is that the CMS MELA discriminant for VBF 2-jets category (denoted by $D_{qqH,2j}^{\text{MELA}}$ from now on in the text) is not used to categorize events. The Tab. 11 and Tab. 12 show the yields for each process after the SM Higgs and VBF-SR

Table 11 - Number of expected events for background and signal, with statistical uncertainty reported, and number of observed events after the SM Higgs selections.

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	19.34 ± 0.14	11.02 ± 0.11	25.99 ± 0.16	56.35 ± 0.24
VH	1.45 ± 0.01	0.92 ± 0.01	2.14 ± 0.01	4.51 ± 0.01
ttH	0.36 ± 0.00	0.23 ± 0.00	0.48 ± 0.00	1.07 ± 0.01
qqZZ+ZZJJ	387.01 ± 1.67	234.64 ± 1.31	538.35 ± 1.97	1160.00 ± 2.90
ggZZ	65.81 ± 0.09	43.85 ± 0.08	102.32 ± 0.13	211.98 ± 0.18
\sum backgrounds	473.97 ± 1.68	290.66 ± 1.31	669.29 ± 1.98	1433.91 ± 2.91
qqH (signal $m_H = 125\text{GeV}$)	1.86 ± 0.01	1.10 ± 0.01	2.53 ± 0.01	5.49 ± 0.02
Total expected	475.83 ± 1.68	291.76 ± 1.74	671.81 ± 1.98	1433.91 ± 2.92
Observed	503	287	669	1459

Source: The AUTHOR, 2018.

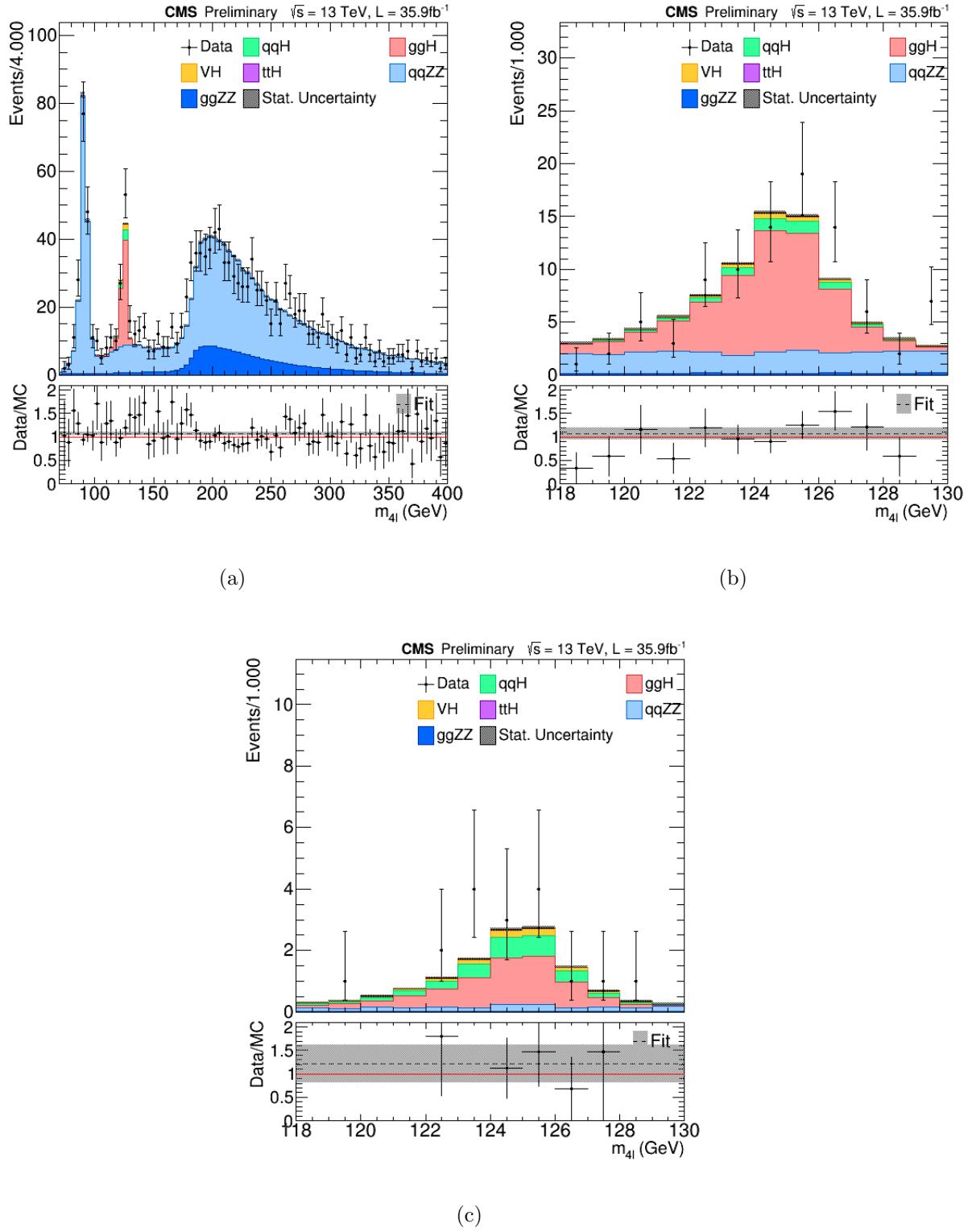
Table 12 - Number of expected events for background and signal and number of observed events after the VBF-SR selections.

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	2.46 ± 0.05	1.28 ± 0.04	3.10 ± 0.06	6.84 ± 0.08
VH	0.34 ± 0.00	0.20 ± 0.00	0.46 ± 0.00	1.00 ± 0.01
ttH	0.05 ± 0.00	0.03 ± 0.00	0.06 ± 0.00	0.14 ± 0.00
qqZZ+ZZJJ	0.67 ± 0.07	0.36 ± 0.05	0.74 ± 0.07	1.77 ± 0.10
ggZZ	0.05 ± 0.00	0.03 ± 0.00	0.06 ± 0.00	0.14 ± 0.01
\sum backgrounds	3.57 ± 0.08	1.90 ± 0.06	4.42 ± 0.09	9.89 ± 0.13
qqH (signal $m_H = 125\text{GeV}$)	1.05 ± 0.01	0.58 ± 0.01	1.39 ± 0.01	3.02 ± 0.02
Total expected	4.62 ± 0.08	2.48 ± 0.06	5.81 ± 0.09	12.91 ± 0.13
Observed	5	2	10	17

Source: The AUTHOR, 2018.

selections, respectively. The Fig. 25 shows m_{4l} distributions for such events. Note that those tables don't include yet the contribution coming from the reducible background (Z+X) that will be discussed in detail in Sec. 8.4.

Figure 25 - Control plots of four-lepton invariant mass for: (a) events passing the SM Higgs selection, (b) events in the $m_{4l} = [118, 130]$ GeV range and (c) events in the VBF signal region.



Source: The AUTHOR, 2018.

7 NEURAL NETWORKS

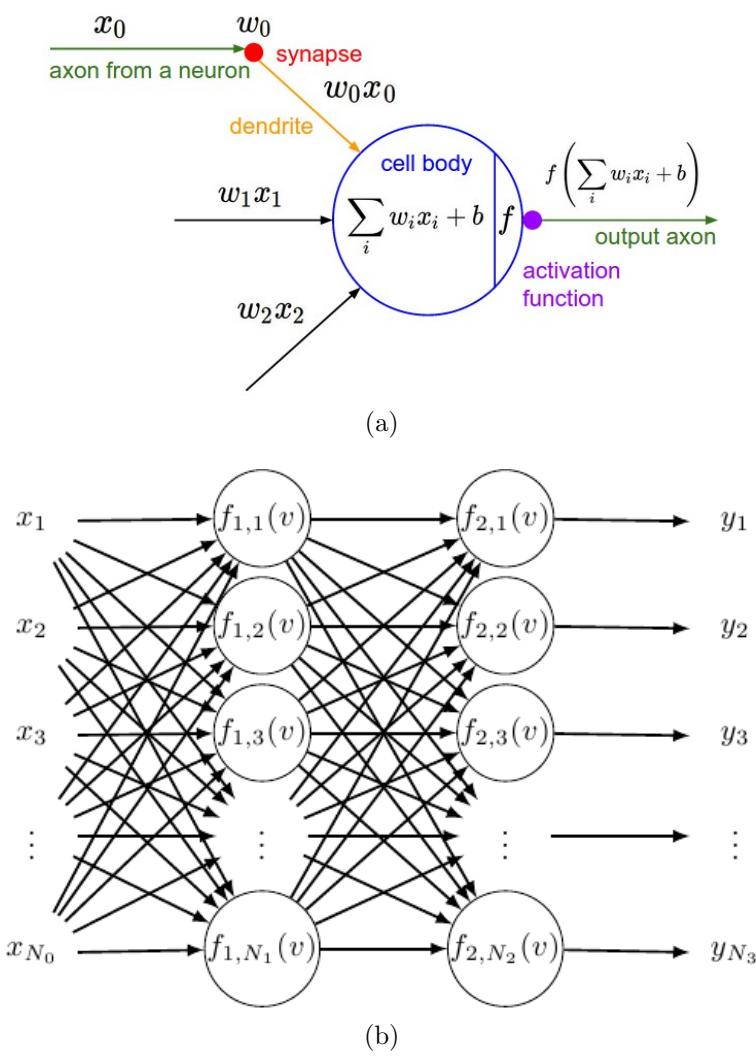
As shown on previous sections the Higgs VBF signal is quite small compared to its main background (the main Higgs production mode ggH). In order to enhance the signal efficiency this analysis was meant to be based on Artificial Neural Networks (ANNs) instead of applying rectangular cuts as usual.

An ANN (frequently referred just as neural network - NN) is an interconnected assembly of small processing units usually called neurons (also activations or nodes). Each of these units performs individual operations and communicates between themselves their results. Such a design was motivated by analogy with the brain, which can be thought as a highly complex, nonlinear and parallel computer. It is estimated that the human brain has 100 billion neurons and each one of them typically receives thousands of connections from other neurons. The inter-neuron connections are mediated by electrochemical junctions, the so called synapses, which happens on branches of the cell called dendrites. These thousand of signals are then combined and depending of the result of this combination the neuron outputs a signal to its neighborhood [79, 80].

Similar to the brain, NNs have the capability to learn from patterns and generalize the modeling of such patterns. This generalization means the possibility to correct identification of a pattern that was not seen before by the NN during the learning process. The neurons contain a function which can receives also several arguments (inputs) and combine them into a single output value (Fig. 26a). There are several types of functions that can be used in the neurons and the choice of the best one might depend on each type of problem. The way the neurons are assembled also can be done in several ways, such that there are different architecture for the connections of the neurons. But a typical and well representative NN architecture can be seen in Fig. 26. This kind of NN is usually called Deep Neural Network (DNN) due to its many layers. Such NN has shown capability to learn very complex tasks and its success comes from relatively recent improvements in the Machine Learning (ML) field [81–83].

Nowadays there are many different flavors of NNs: Fully Connected Neural Networks (FC-NNs), Convolutional Neural Networks (CNNs), Long Short Term Memory Neural Networks (LSTM-NNs), etc. In principle any of these types of NNs can be used in any application but each one of them has some particular structure and workflow that benefits differently each application. In the present analysis the FC-NN was chosen. In a fully connected neural network the outputs from all the neurons in a previous layer are passed to all neurons in the next layer and this process happens across the entire NN in a sequential way (Fig. 26b).

Figure 26 - (a) Graphical representation of a neuron. Each input x_i (from user or from the hidden neurons) is weighted by w_i and summed to a bias b . The sum is passed to a function (the activation function - neuron type) which outputs a vector of values (one for each input). (b) Graphical representation of a NN with three hidden layers. In this type of ANN the outputs of all neurons from a previous layer are fed into the neurons of a next layer.



Source: <https://ifun01.com/8RDXSFC.html>.

7.1 The Learning Process

The way NNs can be trained allows one to classify the learning process basically in two types: unsupervised and supervised. The former stands for the case when one doesn't know if a given dataset contains some pattern, in other words, if it is not possible to label the objects (read signal/background events) in the dataset. The NN can be still trained to find out possible patterns existing in the dataset. The supervised procedure is the opposite situation in which one knows that the dataset contains patterns and such patterns are known. The NN is then trained to correctly assign a given label to a given example. The labels can arbitrarily be chosen and usually for classification task of two classes one uses 0 or 1 for labeling the examples. In order to verify the performance of labels assignment there is a function, usually called *loss* (or cost/risk), which measures the error between the NN response and what is expected [79]. The output of the *loss* function is the main input used to modify the NN parameters in order to evolve the modeling of the data. The modification in the NN parameters (Fig. 26a) is done by an algorithm called *back-propagation* which consists basically of an application of the chain rule for partial derivatives.

Exemplifying such process, let's assume a neuron j in an intermediate layer of a neural network. Calling w_{ji} as the weight assigned by that neuron j to an input $y_i(n)$ at a given iteration n , then such neuron produces a *local field* given by

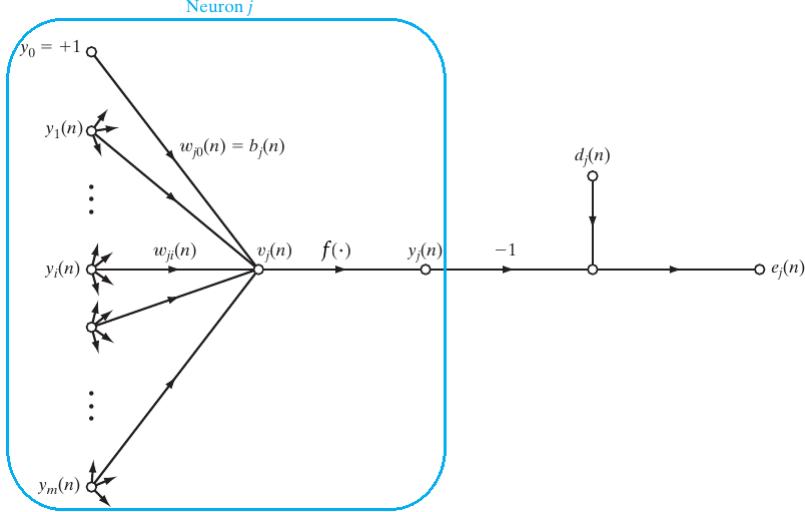
$$v_j(n) = \sum_{i=1}^m w_{ji}(n)y_i(n) + b_j(n) \quad (42)$$

in which m stands for the total number of inputs received by the neuron and $b_j(n)$ is the so called bias (which is just one value per neuron). This *local field* is then given as an argument to an activation function (represented by f in Fig. 26) which produces the *signal function* $y_j(n)$ of the neuron j at iteration n ,

$$y_j(n) = f[v_j(n)] \quad (43)$$

It is possible to compute an instantaneous error for the neuron j (Eq. 44) using its *signal error* $e_j(n)$ which is defined as the difference between the neuron *signal function* $y_j(n)$ (its output after the activation function) and the expected *response* $d_j(n)$. Note that, the expected response $d_j(n)$ can be the label identifying a training example - for neurons in the output layer - or an actual response that a hidden neuron - in the intermediate layers - should have based on the example label. The Eq. 44 is the core of the so called *loss* or *empirical risk* and it is an example of *loss* function used here in order to exemplify

Figure 27 - The flow of signal through the neuron j up to its instantaneous error e_j .



Source: Adapted from [79], p. 129.

$$\frac{\partial \mathcal{E}(n)}{\partial e_j(n)} = e_j(n) \quad (46) \quad \frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (47)$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = f'[v_j(n)] \quad (48) \quad \frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (49)$$

the NN learning process but, be aware that there are other types. Summarizing, the processes described so far can be graphically represented as shown in Fig. 27.

$$\mathcal{E}(n) = \frac{1}{2}e_j^2(n), \quad \text{with } e_j(n) = d_j(n) - y_j(n) \quad (44)$$

Now, taking the partial derivative of the *loss* by application of the chain rule one gets,

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}} = \frac{\partial \mathcal{E}(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)}. \quad (45)$$

and replacing the derivatives accordingly like so,

one reaches to

$$\frac{\partial \mathcal{E}(n)}{\partial w_{ji}} = -e_j(n) f'[v_j(n)] y_i(n). \quad (50)$$

Finally, the correction applied to the weights w_{ji} is given by the *delta rule* as

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial w_{ji}} \quad (51)$$

where the signal $-\cdot$ accounts for the descending gradient in the weights space (that is, it seeks for a direction in which the weight change leads to decreasing of $\mathcal{E}(n)$). The η term is the so called *learning rate* of the back-propagation algorithm. The Eq. 51 is the core of the synaptics (w_{ji}) update within the neural network. Replacing the Eq. 50 into Eq. 51 yields

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n), \quad \text{with } \delta_j(n) = \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} \quad (52)$$

The term $\delta_j(n)$ is the *local gradient*. From this point there are two situations now:

- 1 When the neuron is located in the output layer: in such case that neuron is supplied by the labels classifying the examples (which is its *expected response* v_j) and thus, by Eq. 44, one has the error and can compute the *local gradient* $\delta_j(n)$;
- 2 When the neuron is in a hidden layer some complications arise but still Eq. 51 holds. The Fig. 28 shows a scheme of the signal flow in this case. According to Eq. 52 and Eq. 47 it is possible to redefine the *local gradient* for a hidden neuron j as

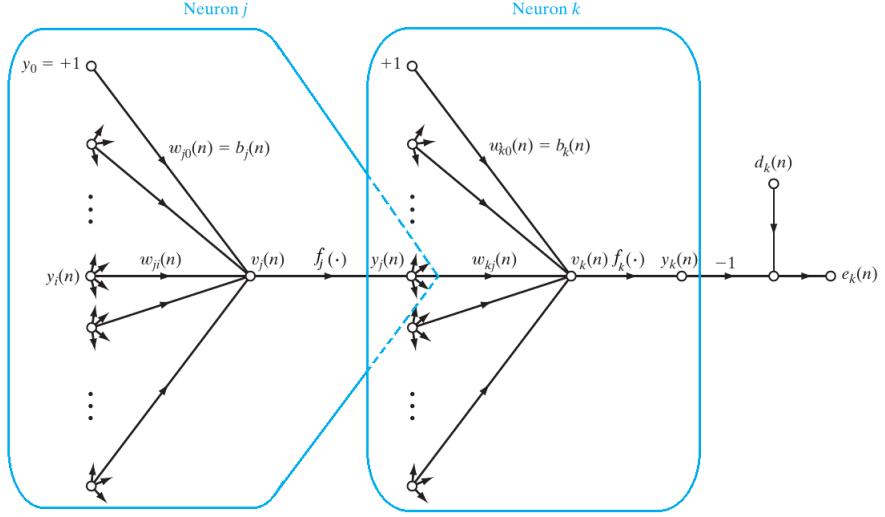
$$\delta_j(n) = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = -\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} f'[v_j(n)]. \quad (53)$$

From Eq. 44 and Fig. 28 it is clear that $\mathcal{E}(n) = (1/2) \sum_{k \in C} e_k^2(n)$. Differentiating this expression with respect to $y_j(n)$ one has

$$\frac{\partial \mathcal{E}(n)}{\partial y_j(n)} = \sum_{k \in C} e_k(n) \frac{\partial e_k(n)}{\partial y_j(n)} \equiv \sum_{k \in C} e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \quad (54)$$

Now, from Fig. 28 and the previous discussions one sees that,

Figure 28 - The flow of signal through a hidden neuron j up to its *signal function* $y_j(n)$ which is forwarded as an input to an neuron k in an output layer. Note that, the *activation function* can now be different between the layers and they are represented as f_j and f_k for the neurons j and k , respectively.



Source: Adapted from [79], p. 132.

$$e_k(n) = d_k(n) - y_k(n) \equiv d_k(n) - f'_k[v_k(n)] \quad (55)$$

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -f'_k[v_k(n)] \quad (56)$$

$$v_k(n) = \sum_{j=1}^m w_{kj}(n) y_j(n) + b_k(n) \quad (57)$$

$$\frac{\partial v_k(n)}{\partial y_i(n)} = w_{kj}(n) \quad (58)$$

and, using Eq(s). 56 and 58 into 54 one has

$$\begin{aligned} \frac{\partial \mathcal{E}(n)}{\partial y_j(n)} &= - \sum_{k \in C} e_k(n) f'_k[v_k(n)] w_{kj}(n) \\ &= - \sum_{k \in C} \delta_k(n) w_{kj}(n) \end{aligned} \quad (59)$$

where the definition of *local gradient* has been used. Finally, using Eq. 59 in 53

one finds that the *back-propagation* formula for the hidden neuron j is

$$\delta_j(n) = f'_j[v_j(n)] \sum_{k \in C} \delta_k(n) w_{kj}, \quad \text{neuron } j \text{ is hidden.} \quad (60)$$

Note that, the sum in Eq. 60 is over all the neurons in the output layer. Also note that, $e_k(n)$ can be extend to internal layers, such that, it will corresponds to the signal errors coming from neurons in the forward hidden layers.

7.2 Developing and Training the Neural Networks

In this analysis the base software used to build up and train the NNs is Keras [90]. Keras is a widely used and powerful set of python libraries that allows one to quickly create any type of known NN. It contains sets of modules that simplifies much of the coding needed by the user. Keras has recently been integrated into TMVA [92] and the user has the flexibility to either prepare a C++ code version or to transfer a python version to TMVA which plugs it into Keras. Note however that in the beginning of this analysis that wasn't available and an entire Keras-python framework was developed to make the studies. Using recent version of TMVA cross-checks were made comparing the results got by Keras-python and Keras-TMVA, and the results are in agreement.

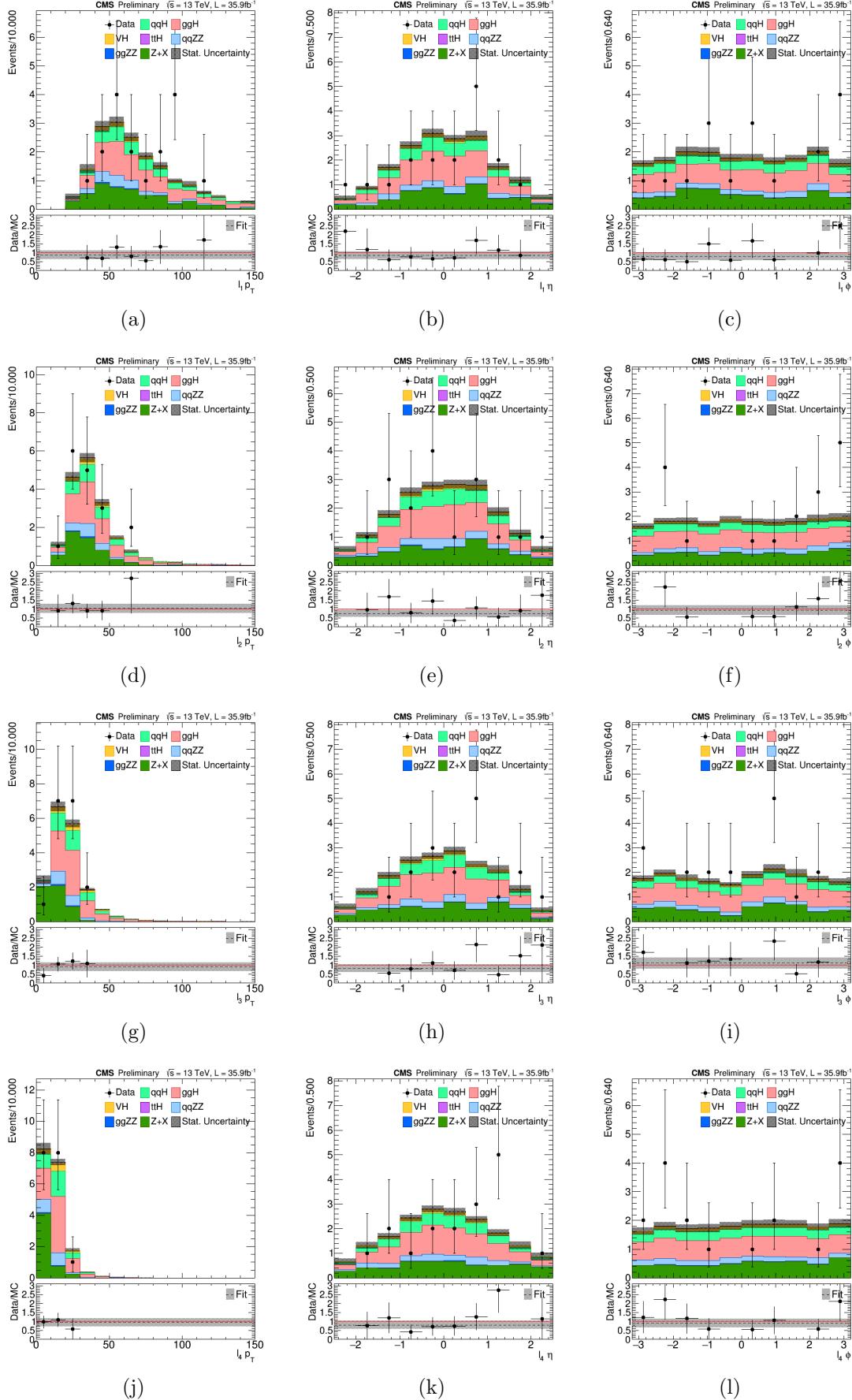
7.3 Preparation of the Datasets

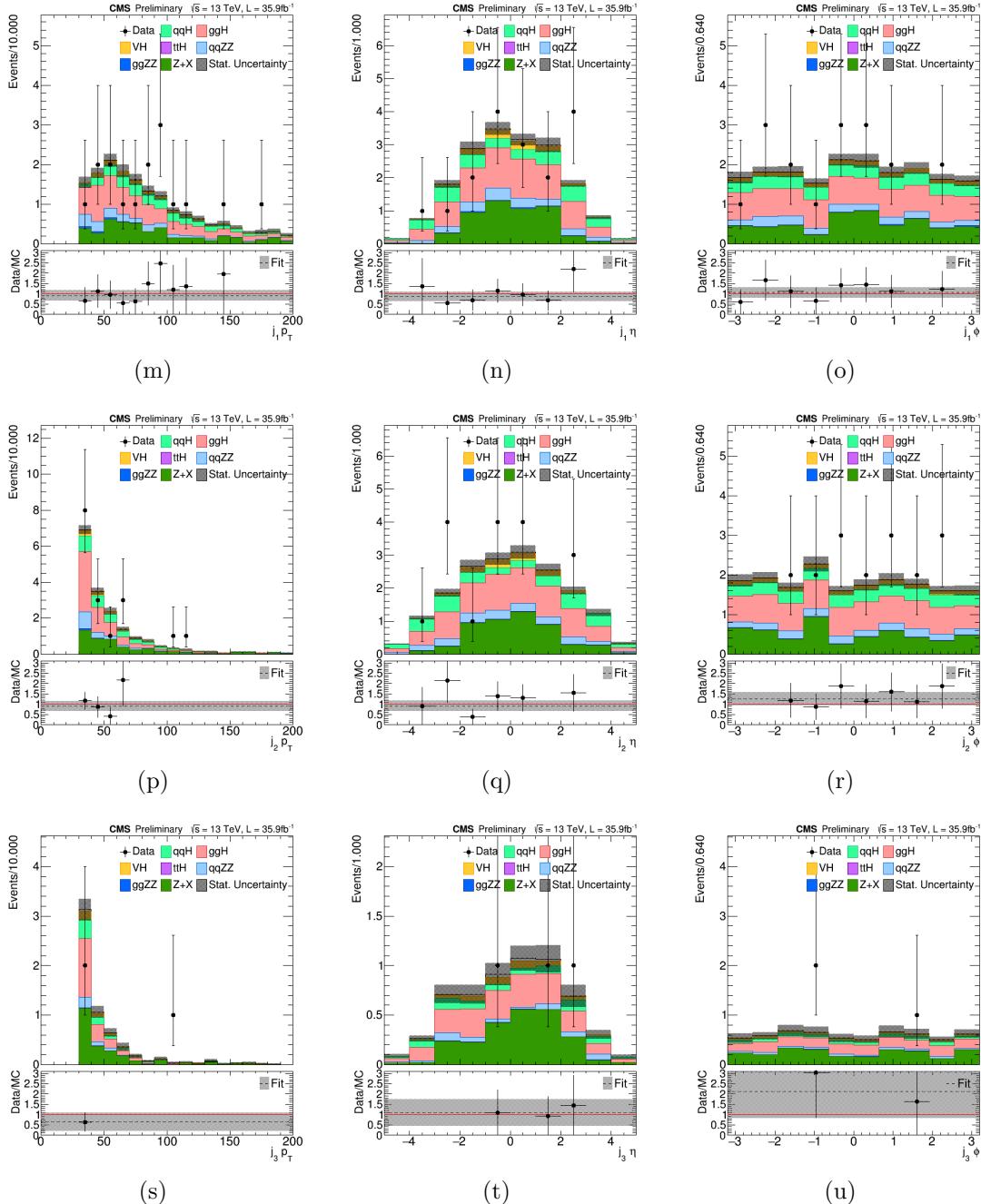
The Neural Networks have been trained using the simulated samples described in Section 4 and only with the events passing the selections defining the VBF-SR. As it will be discussed later the variables used into the final ANNs are the leptons and jets three-momentum (that is, p_T , η and ϕ). The distribution of such variables, for both MC and Data, is shown in Fig. 27 as a cross-check of the variables shape and Data-MC agreement (the low binning has been chosen due to the low statistic available for Data at the VBF-SR). The Z+X background on those pictures is derived via a established CMS data-driven method, which is described in Sec. 8.4.

In order to optimize the analysis and properly handle the usage of the third jet in the events, two jet-based categories have been defined and are labeled as **Njet2** and **Njet3** from now on. The **Njet2** category comprehend the events with exactly two jets while the **Njet3** category are composed by the events with at least three jets. For each of this category a NN was developed.

Assuring an adequate distribution of the events when splitting them, forming the

Figure 29 - Monte Carlo and Data distribution of the variables used as inputs to train the ANNs in this analysis.





Source: The AUTHOR, 2018.

exclusive training and testing sets, the following procedure was adopted for the VBF-SR events in each jet-based category:

- 1 Each simulated process has three samples one for each final state (4μ , $4e$ and $2e2\mu$). Those three samples were merged into just a single sample with the events from each final state being randomized, such that the final states randomly populated the merged sample;
- 2 The merged and randomized samples for each process were divided into two parts: one containing 80% of the total events and one containing the remaining 20%;
- 3 The first and second parts from each process were merged and once more the events were randomized in order to randomly populate each process inside the two parts;

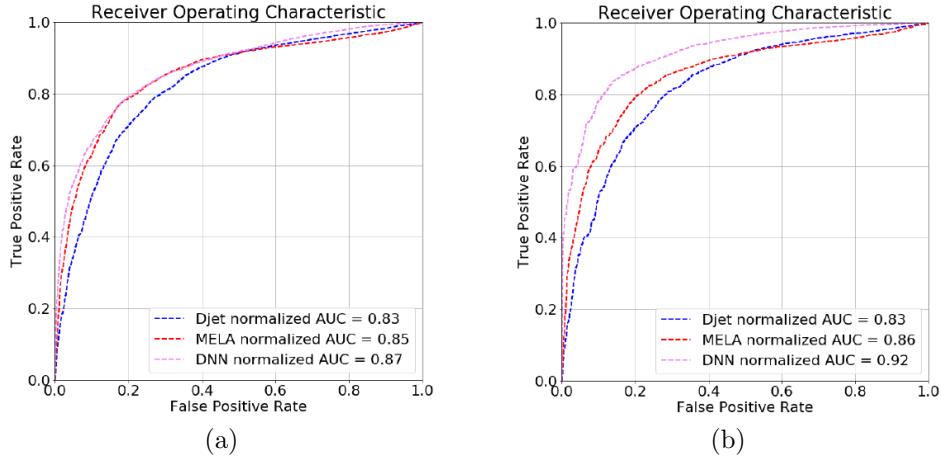
The merged first parts constitutes the training set, that is the set of events used to train the NN. The merged second parts constitute the testing set which is used to test the trained NNs. Note that the two sets are completely independent such that the NN was tested with completely new (and thus unseen by the NN) events. Also, processes which has too few events, such as TT4L and VVV were not included in the training set or in the testing set. As explained on 7 the NN training is very similar to a fit procedure and thus few examples (points) produces fits with large errors.

7.4 Scaling Events Contribution in the Training

In the beginning of this analysis the NN trainings were carried out without taking into account any kind of event weight. In such way all events are seen by the NN in an equal basis. Keras has some features that allows one to include scale factor (example weight) which are in general used to balance the training when the number of events from different classes are very different. In the physics scenario one could use the individual MC event weights ($\sigma.\epsilon.BR$) or yet the sum of those weights (which constitutes the expected yields). The advantage of the first approach is that some events, from signal or background, crossing the classes zones defined by the NN could have small weights and would be worthy to allow them to come in with the benefit to possibly improve the final discrimination. The disadvantage is that the individual weights might not keep the hierarchical contribution of each process since the expected cross-section is divided by the number of events (such that if one have a large number of events the weights might become smaller for important processes than the ones for other small processes).

Both approaches were tested. The weight is used by Keras as a scale factor for the *loss* function such that during the training each example is seen by the NN with a different importance. That affects the direction in which the *minimizer* computes the

Figure 28 - ROC curves comparing the performance of a trained NN, the VBF MELA discriminant and D_{jet} : (a) the performance of NN without scaling the *loss* and (b) performance after scaling the *loss*.



Source: The AUTHOR, 2018.

gradients of the *loss* function. The advantage of doing so is that even if a process has few events, which is very frequent in many analysis, it will be properly taken into account with respect to the other process having more events. The Fig. 28 shows the impact of scaling the *loss*. In the beginning of the analysis the sample ggH-minlo was produced slightly different from the other samples: the leptons were sorted by p_T in the 4μ final state. Without weighting the events during the training the typical ROC curves observed were like the one showed in Fig. 28a. Once the weights were introduced the discrimination of VBF and ggH processes increased significantly as showed in Fig. 28b. Although this is not a physical property and thus had to be fixed, this shown the big importance of scaling the events contribution. It shows that the NN may not learn some features from a process if a scaling is not applied. As seen in Tab. 12 ggH has the biggest yield and has a relatively good number of events but still without weighting it the small difference was not taken into account by the NN. Since there all the further developed studies were done with the events weighted.

7.5 The Training Procedure

Training MVA methods is a procedure that always need to be done in multiple fronts. It is hard to guarantee that for a set of inputs a set of MVA training parameters are the optimal choice, since there is an infinity of combinations that one can build. Those configurations (including the training dataset) can strongly affect the evolution of the MVA training. In order to optimize the study the framework developed in python has

the feature to build up through Keras different NN architectures and perform multiple parallel scans. These scans can run over different sets of inputs and the several parameters that needs to be configured for the training. The result of each NN training can be retrieved to produce plots which are used to validate and classify the quality of each training (one can check if over-fitting happened, for instance). The set of tested inputs includes combinations of the leptons and jets kinematic variables (p_T , η , ϕ and E), MET, Njets and Nbjets. The scanned NN parameters are summarized in the Tab. 13. These parameters are:

- **Pre-processing:** it is a usual practice in Machine Learning field to apply some operation in the NN inputs in order to standardize them (keep them in the same range of values for instance). In the scans done the pre-processing showed negligible effects and it was decided to not adopt such procedure for further studies. Note that, when pre-processing is applied in the training it will be need to also apply the same procedure when using the trained NN (otherwise the results will be mistaken, since the NN are expecting preprocessed inputs). One also should note that preprocessing changes the input variables what may destroy the possibility of the NN reconstruct some important property(e.g. invariant masses);
- **Topology:** the architecture of the NN, that is, the number of hidden layers, neurons and the neuron type. In Keras there are several types of neuron which even includes possible learning parameters (during training). The *Rectified Linear Unity* (ReLU) is the most recommended due to its property of non-vanishing gradient;
- **Batch size:** the number of events in a subset from the training set used to compute the gradients and update the w_i 's and b in each neuron;
- **Epochs:** the number of iterations over the full training set. The total number of iterations is a combination of the size of the training set, the batch size and the number of epochs. For instance, setting a training of 10 epochs and a batch size of 10 for a training set size of 100 means that Keras performs 10^2 updates on the w_i 's and b 's;
- **Early stop:** a parameter to set the number of epochs which Keras should wait if not improvement in the loss function is observed. If still not improvement is seen after that number of epochs the training is stopped;
- **Minimizer:** is the method to compute the gradients. There are several options in Keras (SGD, Adam, RMSprop, etc.) and after testing most of them it was decided to keep Adam. Adam stands for Adaptive Momentum and has the property of fast convergence due to a NN updating directed towards the loss function minimum. This minimizer is widely applied in ML studies;

Table 13 - Summary of the scanned parameters during ANN trainings.

Parameter	Tested options
Inputs	leptons/jets(p_T, η, ϕ), MET
Pre-processing	none, normalization, standardization
Topologies	7:5:3, 9:7:5, 11:9:7, 15:10:5, 21:13:8, 10:10:10:10, 30, 100, 100:100:50, 10 ³
Early stop	100, 600, 3000
Minimizer	SGD, Adam, Adagrad, Adadelta, RMSprop
Batch size	1, 5, 32, 64, 128, 786
Neuron	ReLU, SeLU
Loss Scaling	XS (process total XS), $\sigma \cdot \epsilon \cdot BR$ (event weight)
Dropout	none, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99, 0.3:0.4:0.2, 0.5:0.25:0.1

Source: The AUTHOR, 2018.

- **Scaling:** Keras allows one to scale the loss function by some weight that can be independent for each training example or the same for a entire class (signal or background, for instance). It was tested the impact of using the expected yield (XS) of each process and the individual weights of each event ($\sigma \cdot \epsilon \cdot BR$);
- **Dropout:** it is a procedure in which keras randomly sets a fraction of the input units (literally the ANN input variables or then neurons outputs - which are inputs for forward neurons in the net) to zero at each update during the ANN training. This mechanism helps to prevent the over-fitting issue and also can build up more complex models. Keras allows one to apply this procedure in each layer independently (so one can apply it just on the first layer and not in the other ones);

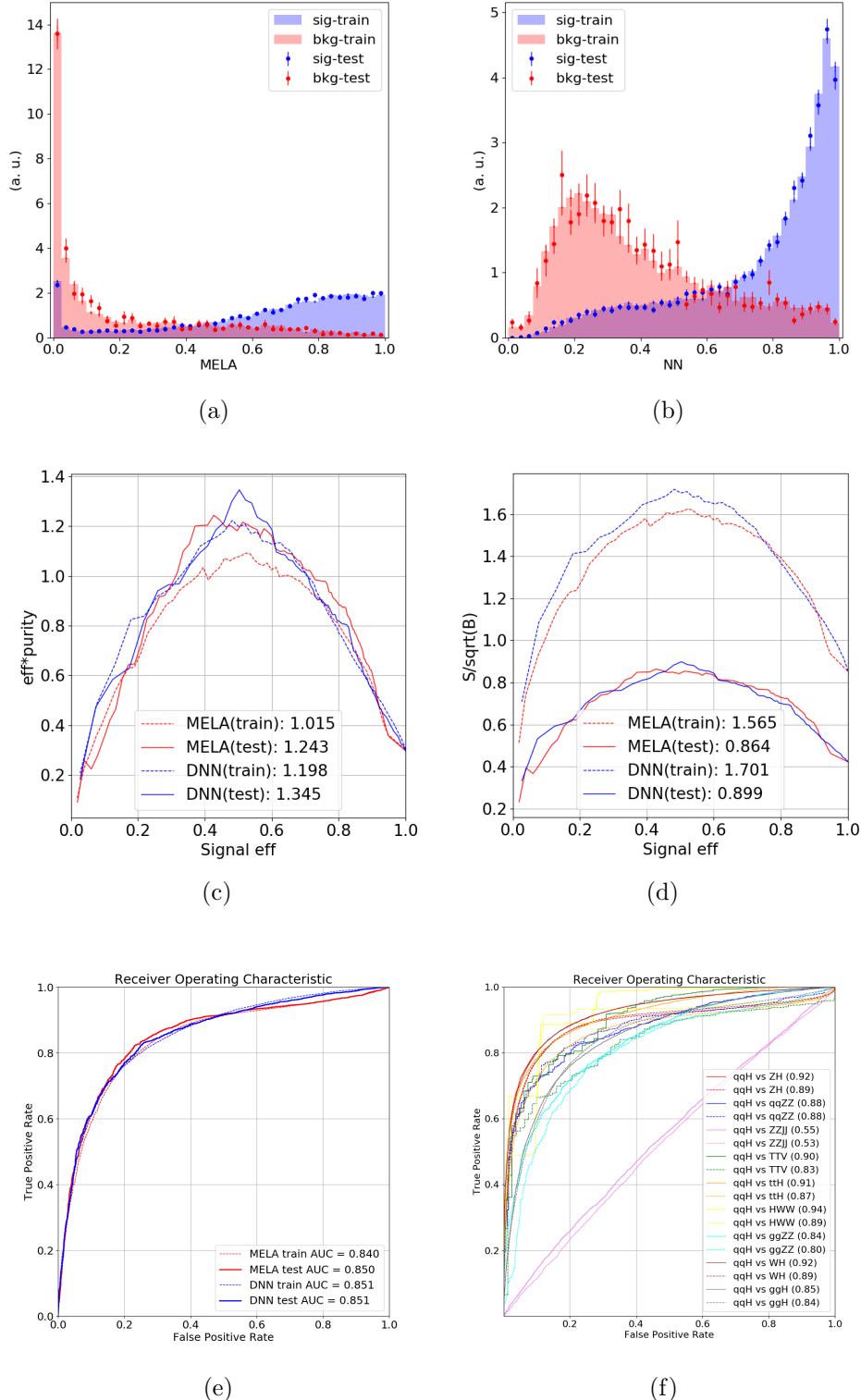
The ANNs achieved through these trainings are produced together with some plots which are used to validate the choice of the best ANN architecture. These plots comprehend the metrics adopted (ROCs, $\epsilon_s \cdot \pi$, σ) and the ANN distributions for the training and test sets, which helps to check if there was over-fitting for a given configuration. Fig. 29 shows these plots. The ROCs curve from training and test sets are also used to check if there was over-fitting. The plots in Fig. 29 show a example in which the ANN is considered good.

7.6 NNs Training Performances

In order to better visualize the results from the several scannings and choose the best ANN configuration, the metric $\epsilon_s \cdot \pi$ were chosen to be the one representing the ANNs performance score. The Fig. 30 shows a summary of those scannings. It is the difference in terms of $\max(\epsilon_s \cdot \pi)$ between each ANN and the $D_{qqH,2j}^{MELA}$ applied to the same testing set.

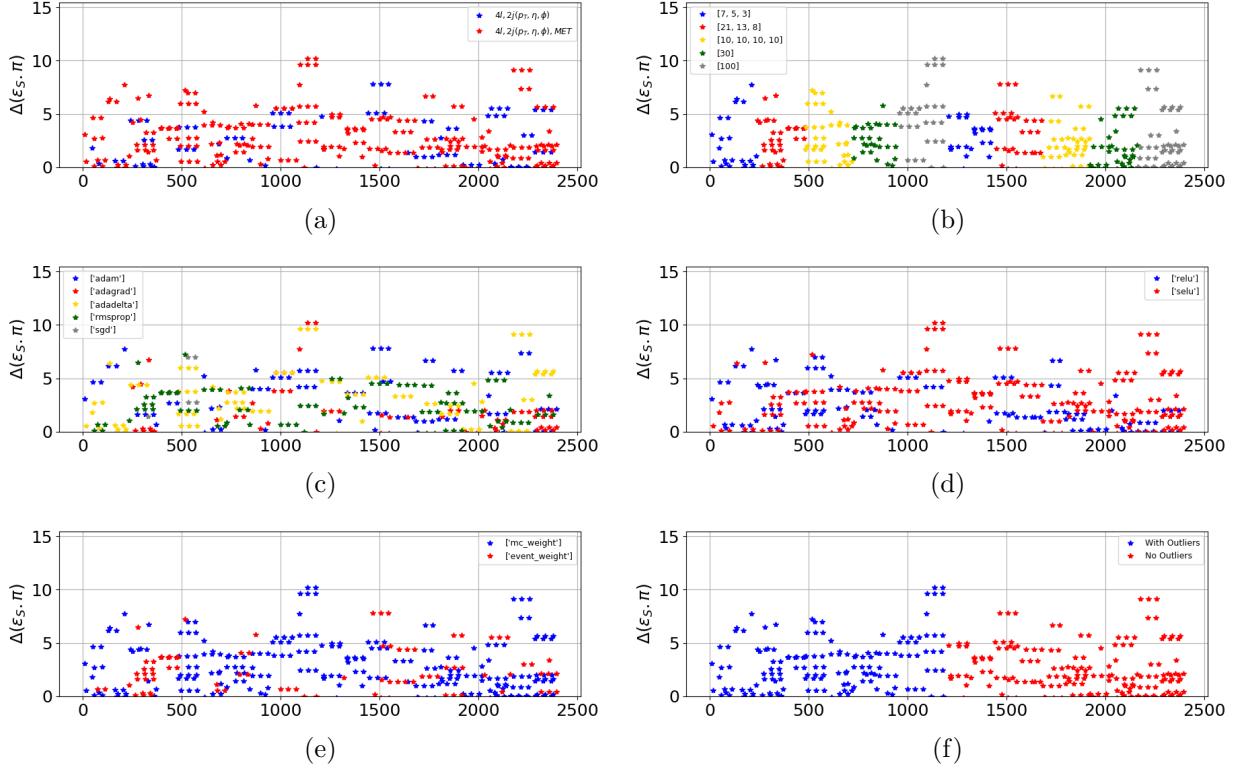
From those several ANN trainings, some observations were drawn:

Figure 29 - Example of checking plots produced during ANNs training. (a) $D_{qqH,2j}^{MELA}$ (for reference) and (b) ANN distributions, (c) $\epsilon_s \cdot \pi$ vs ϵ , (d) signal significance and (e) ROC curves computed using all backgrounds together for training and testing sets. (f) ROC areas computed separately for each background using the testing set (the filled lines are for the ANN and the dotted ones are for $D_{qqH,2j}^{MELA}$).



Source: The AUTHOR, 2018.

Figure 30 - Summary of the results obtained with the parameters scanning. The plot shows the difference between trained ANNs and $D_{qqH,2j}^{MELA}$ in three metrics used at that moment: $\max(\epsilon_s, \pi)$, ROC area and significance



Source: The AUTHOR, 2018.

- 1 As expected the jets are the objects which have almost all the discrimination power for VBF against the backgrounds. This is clear by noticing that when only the leptons are used the ANNs performance are very low (for any configuration). And also, noticing that with only the jets there were ANNs with performance similar to $D_{qqH,2j}^{MELA}$;
- 2 p_T , η and ϕ seems to be enough for the discrimination. As can be noticed adding jet energy caused some trouble and none of the attempted configuration got better performance than when not using the jet energy;
- 3 Additional jets lead to more discrimination power. Different from $D_{qqH,2j}^{MELA}$, the ANNs can use information from extra jets ($n_{jets} > 2$) when those are present in the event;
- 4 The usage of MET, Njets and Nbjets produced an increasing in performance similar to the ones observed when extra jets are used. For the final ANNs, Njets and Nbjets were not used since it is not straightforward how to estimate their systematic uncertainty.

Table 14 - Fraction of events, after the VBF-SR selections, containing the jet_i (i-th jet).

	$j_3(\%)$	$j_4(\%)$
qqH	21.3	5.0
ggH	28.8	7.1
$qqZZ$	19.5	2.9
<i>Data</i>	18.0	0.0

Source: The AUTHOR, 2017.

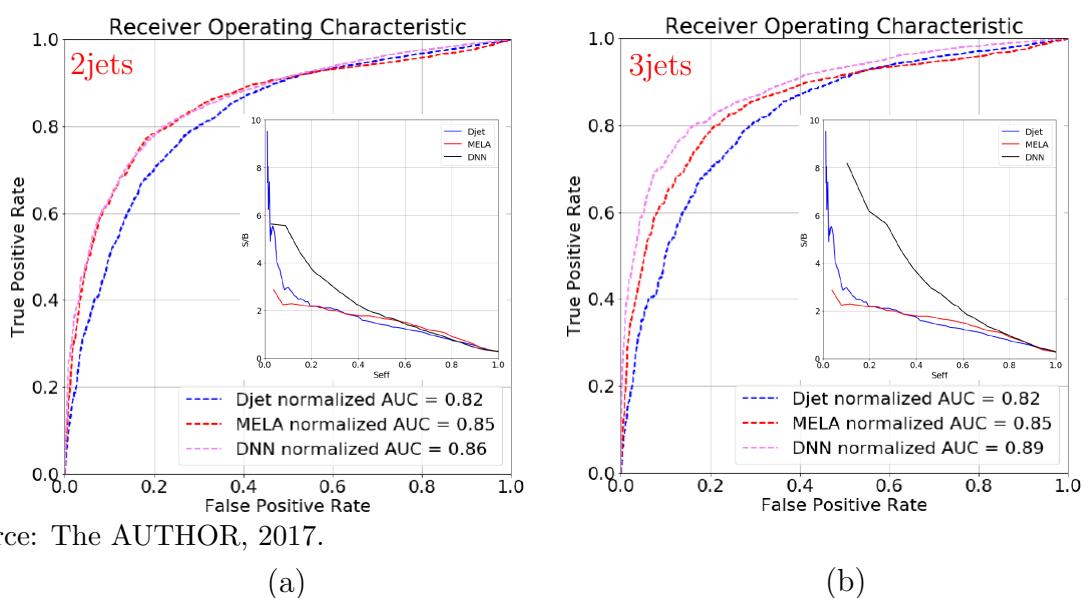
The final ANN for each of the jet-based categories were chosen to be the one with highest $\max(\epsilon_s, \pi)$ and are presented in Sec. 10.2. For the seek of the reader in Appendix B the architecture of the two final ANNs can be seen. They give an idea of the correlation between the input variables.

7.6.1 The Impact of the VBF 3rd Jet

The usual VBF process is characterized by the presence of two high energetic jets with a big gap between them on η , where the particle activity is rather small or even nonexistent around it. This is the LO process. However, is possible that more jets appear in a VBF process and such jets are radiative corrections (higher order diagrams - NLO) which contributes for the total VBF cross section. These extra jets can contribute for discriminate the VBF process against the backgrounds.

Looking into the number of events which has extra jets one sees that there is a significant fraction for VBF , ggH , $qqZZ$ and even the observed data events with at least 3 jets. These fractions are place on Tab. 14. The usage of the ANNs including a 3rd jet in the events indicated at the beginning of this analysis that one could improve the discrimination between VBF and the background. Fig. 31 shows that the ANNs present similar behavior as the MELA for a case with only two jets, while they show a improvement when the 3rd jet is present as an input for the ANN. This is where the motivation to use the 3rd jet comes from in the present analysis.

Figure 31 - Performance of VBF discrimination against the backgrounds for the ANN, MELA and Djet discriminants. On (a) is the performance when considering only 2 jets per event and on (b) when considering up to 3 jets per event.



Source: The AUTHOR, 2017.

8 BACKGROUND ESTIMATION

In this analysis the backgrounds are composed by the three SM Higgs production modes, gluon fusion (ggH) and associated vector (WH , ZH and $t\bar{t}H$), and the SM backgrounds $q\bar{q} \rightarrow ZZ$, $gg \rightarrow ZZ$ and $Z + X$.

8.1 Estimation of the SM Higgs Production Modes

The SM Higgs production modes are estimated from MC by normalizing the yields according to the cross sections times the branching ratio fractions ($\sigma \cdot BR$) as computed by the LHC Higgs Cross Section Working Group [77].

8.2 Modeling of $q\bar{q} \rightarrow ZZ$

The $q\bar{q} \rightarrow ZZ$ background events are generated at NLO. The fully differential XS, which is not yet available in a partonic level event generator, is computed at NNLO. NNLO/NLO k -factors are then applied to the $q\bar{q} \rightarrow ZZ$ NLO POWHEG sample. The inclusive cross sections obtained using the same PDF, as well the renormalization and factorization scales, as the POWHEG LO, NLO and NNLO are used. The NNLO/NLO k -factors are applied as function of m_{ZZ} [72, 77].

Additional NLO electroweak corrections, which depend on the initial state quark flavor and kinematics, are also applied in the mass range of $m_{ZZ} > 2m_Z$, where the corrections have been computed [72].

8.3 Modeling of $gg \rightarrow ZZ$

The $gg \rightarrow ZZ$ background is generated at LO with the generator MCFM 7.0 [93]. This background does not have exact calculation beyond LO. However, it has shown that the soft collinear approximation is able to describe such background cross section and the interference term at NNLO [94]. Also, the k -factors are very similar at NLO for signal (Higgs) and background [95] and, at NNLO for signal and the interference terms [96]. Because of that, the same k -factor is used for signal and background. The k -factor at NNLO for the signal is obtained as a function of m_{4l} using the HNNLO v2 MC program by calculating the NNLO and LO $gg \rightarrow H \rightarrow 2l2l$ cross sections at the small H boson decay width of 4.07 MeV and taking their ratios [97–99]. The NNLO, as well the NLO,

k-factors and the cross sections are reported in [72] along with the NNLO, NLO and LO cross sections at the SM Higgs boson decay width.

8.4 Estimation of Z+X Background

The commonly called Z+X background is a reducible background in the $H \rightarrow ZZ \rightarrow 4l$ analysis and originates from processes which contain one or more non-prompt leptons in the four-lepton final state. Non-prompt leptons are mainly non-isolated leptons which can originate from the heavy-flavor mesons decay, mis-reconstructed jets (usually coming from light-flavor quarks) and electrons from γ conversions. Such non-prompt leptons are usually called "fake" leptons.

The estimation of such background in the $H \rightarrow ZZ \rightarrow 4l$ analysis is done by measuring the f_e (f_μ) probability of fake electrons (fake muons) passing the loose selection criteria (described in Sec. 6) to also pass the final selection criteria (the *tight* selection described in Sec. 6). These probabilities are called as fake ratios or fake rates and are applied into two defined control regions (CRs) in order to extract the expected Z+X yield in the signal region.

In the following sections the steps needed for the estimation of Z+X yields and shape in this analysis are described.

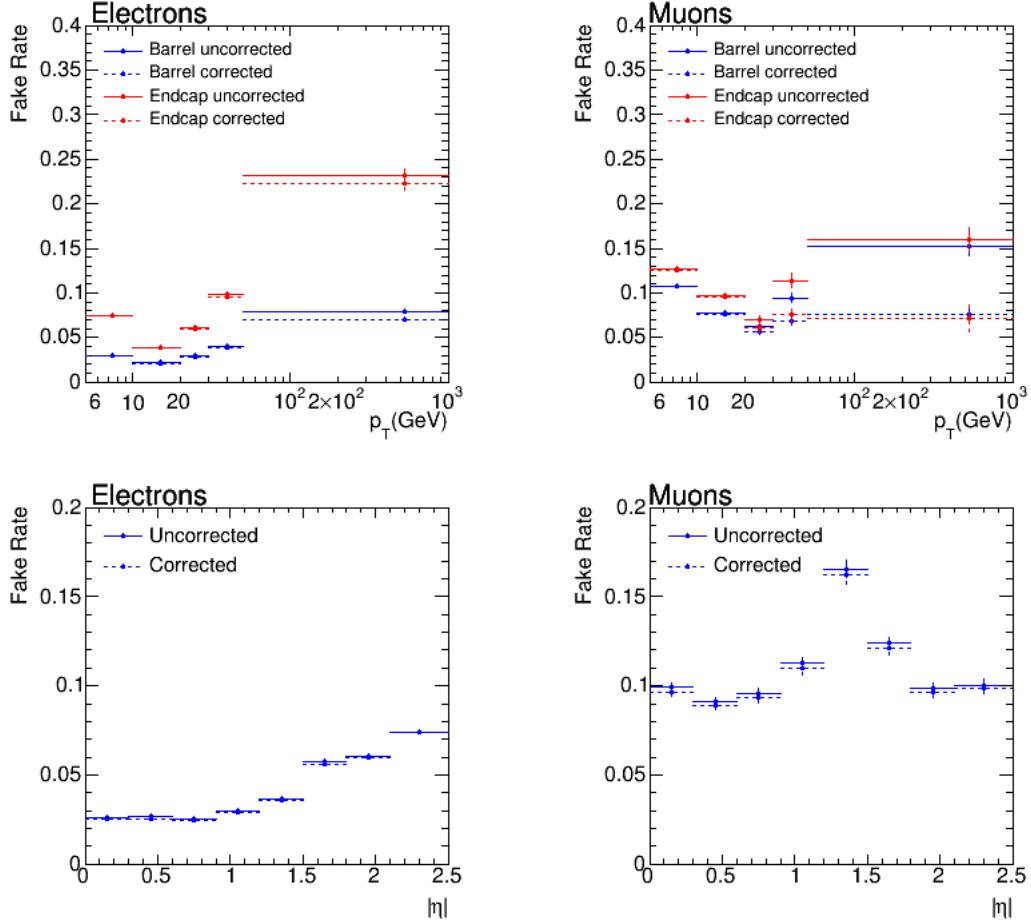
8.4.1 Measuring the Fake Rates

The measurement of the lepton fake rates requires a selection of samples of $Z_{ll} + e$ and $Z_{ll} + \mu$ events. Such events are expected to be dominated by a Z boson and a fake lepton. The leptons forming the Z must be opposite sign, same flavor and have $p_T > 20(10)$ GeV. The third lepton (e/μ) has to pass the loose selection and is used as the probe to compute the f_e and f_μ probabilities. The invariant mass given by the fake lepton and the opposite-sign tight lepton composing the Z is required to be $m_{ll} > 4$ GeV to reduce QCD contamination. Also, in order to suppress contamination from γ conversions to electrons the invariant mass of the two leptons composing the Z must satisfy $|m_{Z_{ll}} - m_Z| < 7$ GeV. The remaining events are expected to also contain contribution from WZ and $t\bar{t}$, which are reduced by requiring $E_T^{miss} < 25$ GeV.

The fake rate is parameterized in terms of the probe lepton p_T and η . It has been shown in [77] that there is no dependence between the fake rate and the probe lepton charge. Fig. 32 shows the fake rate estimation separately for electrons and muons in terms of p_T and η .

There are two FR distributions in Fig. 32. The points with solid error bars are the

Figure 32 - Fake rates in terms of the lepton probe p_T and η . The p_T distributions (barrel and endcap) corresponds to the regions defined by $|\eta| < 1.479$ (1.2) for electrons (muons). The background WZ from MC has been subtracted.

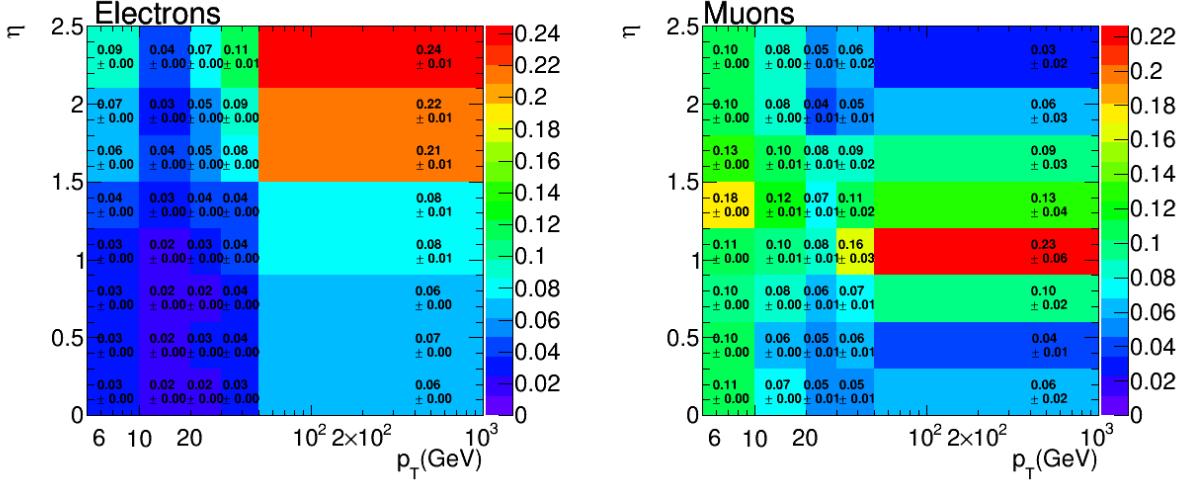


Source: The AUTHOR, 2018.

fake rates computed directly from data through the procedure described above. But, since WZ background potentially contributes with three real leptons and such background is already included in the analysis via MC, it is then subtracted from data (separately in the numerator and in the denominator used to compute the FR's). The resulting FR after this subtraction is shown by the points with dotted error bars in Fig. 32.

Now, as one can see, there is a significant dependence between the FR's and the (p_T, η) of the loose leptons. For this reason the final parameterization of the FR's is made as a function of these two variables and it is shown in the Fig. 33. The binning used for such distributions have been chosen in order to control the statistical uncertainty across the bins and avoid large uncertainties on the FR's. The muon case is worse than electron due to low statistics in data.

Figure 33 - Fake rates as a function of the probe lepton p_T and η . The background WZ from MC has been subtracted. This is the mapping for fake rates applied into the control regions.



Source: The AUTHOR, 2018.

8.4.2 Building Control Regions

In order to apply the fake rates described previously, two control regions are defined requiring events with two leptons passing the selection of the first Z (tight leptons) and additionally a pair of loose leptons of same flavor, opposite charge and passing the SIP_{3D} cut. These events also must satisfy all the kinematic cuts applied for the Higgs phase space selection (See section 6).

The first control region is then built by requiring that the two loose leptons do not pass the final identification and isolation criteria. This control sample is nominated as "2Prompt + 2Fail" (being referred as 2P2F from now on) and it is expected to be populated by events that have only two prompt leptons: mostly DY (Drell Yan) with a small fraction of $t\bar{t}$ and $Z\gamma$.

The second control region is obtained by requiring that just one of the four leptons does not pass the final identification and isolation criteria. The remaining three leptons in the event should pass those selections. This control region is then nominated "3Prompt + 1Fail" (being referred as 3P1F from now on). The events appearing in this control region are the same present in the 2P2F, with different relative proportions though, plus WZ events that are expected to have three prompt leptons.

This two control regions, which are enriched by fake leptons, are orthogonal to the Higgs signal region and are used to estimate the yields and shape of the $Z+X$ background in the signal region. The four lepton invariant mass for the 2P2F and 3P1F control

regions, as obtained for data and simulation separated by final state, can be seen in the Fig. 34 and Fig. 35.

The expected number of reducible background events in the 3P1F control region can be computed by weighting each event observed in the 2P2F control region with the factor $(\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j})$, in which the f 's correspond to fake rates as derived from the map shown in Fig. 33 using (p_T, η) of each loose lepton. This estimation of 3P1F events from the 2P1F control region is shown in Fig. 35.

The discrepancy observed in the 2P2F control region for $Z_1 + \mu\mu$ is related specially to the limited statistics of $Z + b\bar{b}$ and $Z + c\bar{c}$ events in the inclusive $Z + \text{jets}$ sample (Drell Yan) as has been pointed previously in [84, 85]. The $Z_1 + ee$ channels are better described by the available MC simulation. The discrepancies also arise because the fake rates do not properly take into account the background composition of 2P2F as detailed in Sec. 7.2 of [72].

In order to correct those discrepancies and correctly estimate the $Z + X$ background, one needs two components:

- a component directly from 2P2F, computed by weighting each observed event in the 2P2F control region using the factor $(\frac{f_i}{1-f_i} \frac{f_j}{1-f_j})$, in which the f 's are the fake rates of loose leptons as explained before;
- a component from 3P1F, obtained by the difference between the observed events in 3P1F (N_{3P1F}), the expected contribution from 2P2F and from ZZ in the signal region ($N_{3P1F}^{ZZ} + N_{3P1F}^{bkg}$). The ZZ contribution is taken from the MC simulated events selected in the 3P1F control region. The N_{3P1F}^{bkg} is the 3P1F extrapolation from 2P2F as discussed above.

the final expression for the estimation of $Z + X$ can be then symbolically represented as,

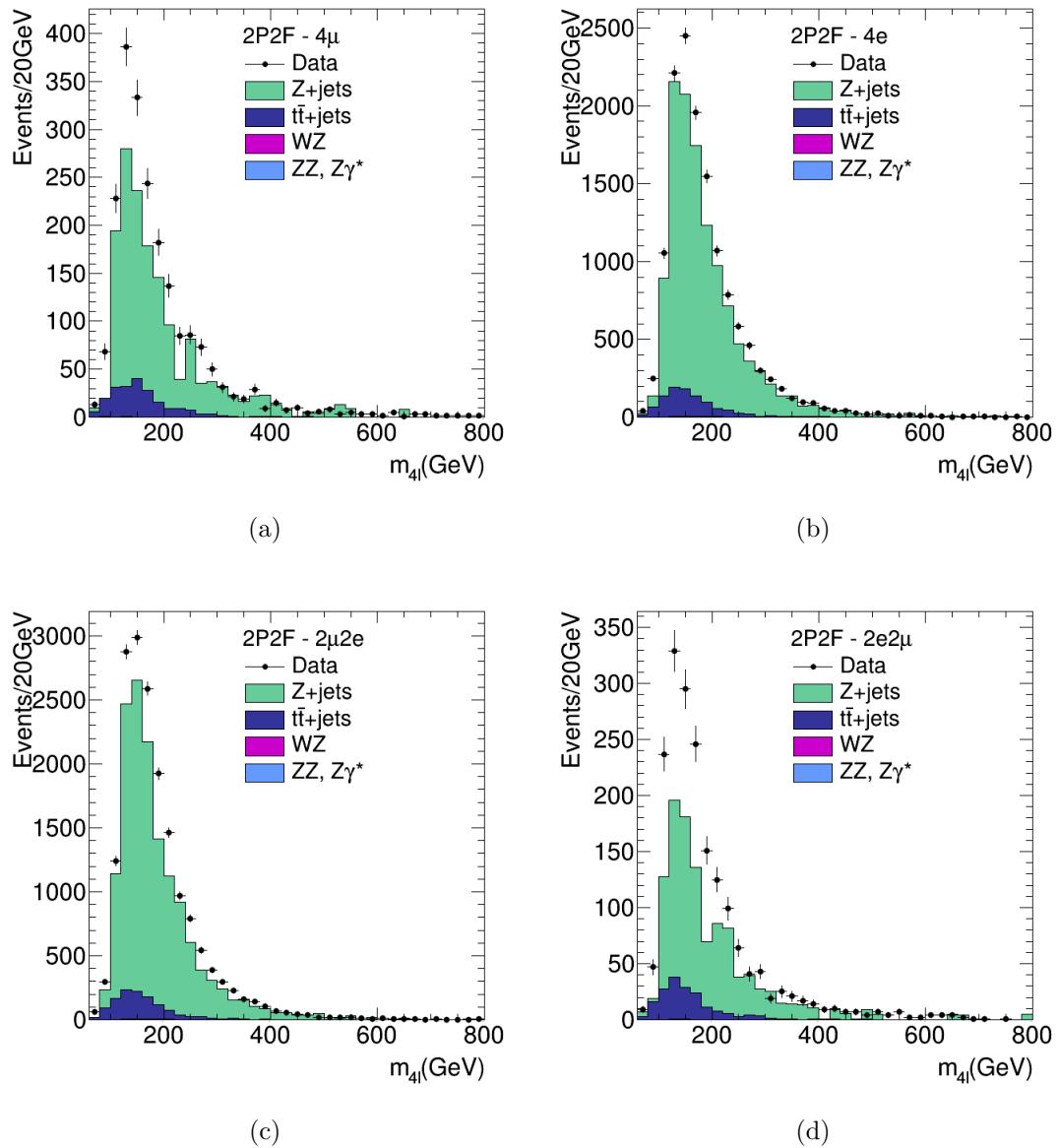
$$N_{SR}^{bkg} = \sum \frac{f_a^i}{(1-f_a^i)} (N_{3P1F} - N_{3P1F}^{bkg} - N_{3P1F}^{ZZ}) + \sum \frac{f_b^i}{(1-f_b^i)} \frac{f_c^j}{(1-f_c^j)} N_{2P2F} \quad (61)$$

which is more conveniently expressed as,

$$N_{SR}^{bkg} = (1 - \frac{N_{3P1F}^{ZZ}}{N_{3P1F}}) \sum_i \frac{f_a^i}{(1-f_a^i)} - \sum_j \frac{f_b^j}{(1-f_b^j)} \frac{f_c^j}{(1-f_c^j)}. \quad (62)$$

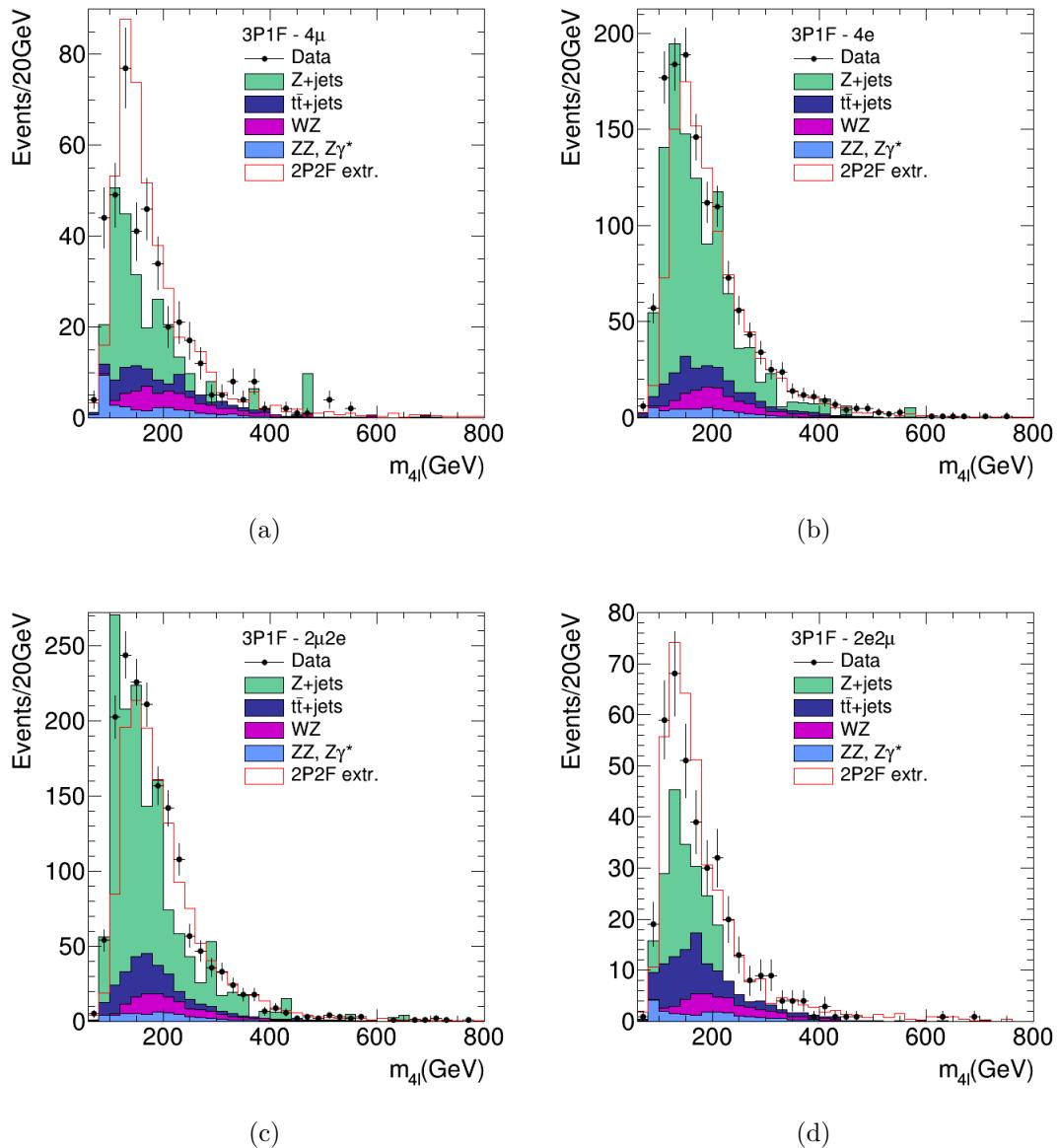
The Eq. 62 represents what has been done to estimate the $Z + X$ background contribution in this analysis. In order to get the appropriate yields and their statistical uncertainties the Eq. 62 is computed using histograms. The two sums are represented

Figure 34 - Four-lepton invariant mass distribution of the events selected in the 2P2F control region in the 13TeV dataset: (a) 4μ , (b) $4e$, (c) $2\mu 2e$ and (d) $2e 2\mu$ channels.



Source: The AUTHOR, 2018.

Figure 35 - Four-lepton invariant mass distribution of the events selected in the 3P1F control region in the 13TeV dataset: (a) 4μ , (b) $4e$, (c) $2\mu 2e$ and (d) $2e 2\mu$ channels.



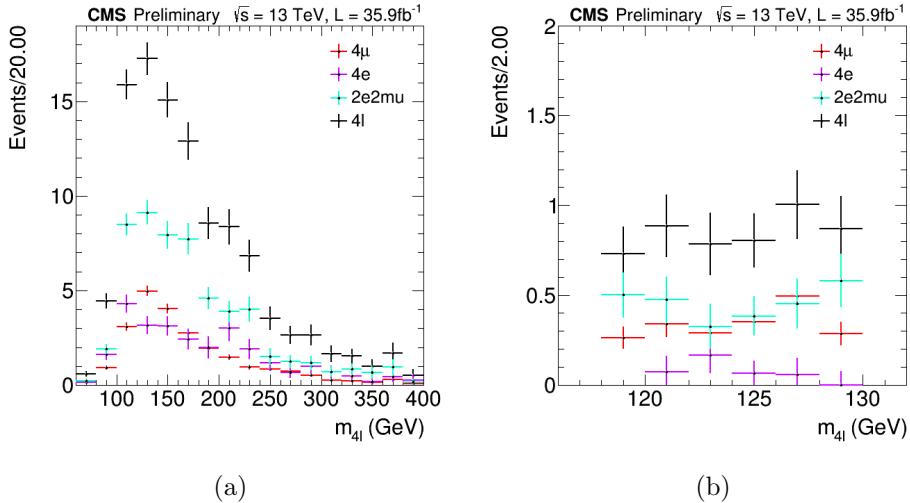
Source: The AUTHOR, 2018.

Table 15 - Yields estimated for Z+X in the signal region from the measurements on data using the fake rate method with opposite-sign (OS) leptons. Estimates reported with statistical uncertainties comparing Higgs signal SR and after applying the cuts defining the VBF-SR.

Selection	4μ	$4e$	$2e2\mu$	$4l$
SM Higgs	24.28 ± 0.60	27.80 ± 1.81	56.90 ± 2.11	108.99 ± 2.84
VBF-SR	2.03 ± 0.18	0.33 ± 0.20	2.72 ± 0.17	5.08 ± 0.32

Source: The AUTHOR, 2018.

Figure 36 - Four-lepton mass distribution of Z+X for events after the (a) SM Higgs and (b) VBF-SR selections. The shapes are shown for each separate channel and for their combination into $4l$ final state.



Source: The AUTHOR, 2018.

by two separated histograms with weights given by the fake rate fractions from the loose leptons. Then, the histogram representing the first sum (left one) is scaled by the multiplying term seen in Eq. 62. In that term, N_{3P1F}^{ZZ} and N_{3P1F} are, respectively, the number of expected ZZ events (from MC) and the number of observed events selected in the 3P1F control region. Finally, the histogram representing the second sum (right one) is subtracted from the scaled histogram. This procedure is convenient to get not only the yields (which corresponds to the integral of the resulting histogram) and the statistical uncertainties correctly but, also the shape of Z+X. The yields computed in this way are shown in Tab. 15 and the derived m_{4l} shapes are shown in Fig. 36 for events after (a) SM Higgs and (b) VBF-SR selections.

Table 16 - The fake ratios for individual background processes, the average fake ratio and the fake ratio reweighed according to the composition of backgrounds in 2P2F control region.

FR	$ZZ, Z\gamma^*$	WZ	$t\bar{t} + jets$	$Z + jets$	Average	Reweighted	Uncertainty (%)
e	0.750 ± 0.005	0.848 ± 0.014	0.077 ± 0.008	0.039 ± 0.001	0.040 ± 0.001	0.051 ± 0.001	27.5
μ	0.899 ± 0.007	0.963 ± 0.017	0.165 ± 0.020	0.114 ± 0.002	0.118 ± 0.002	0.134 ± 0.003	13.3

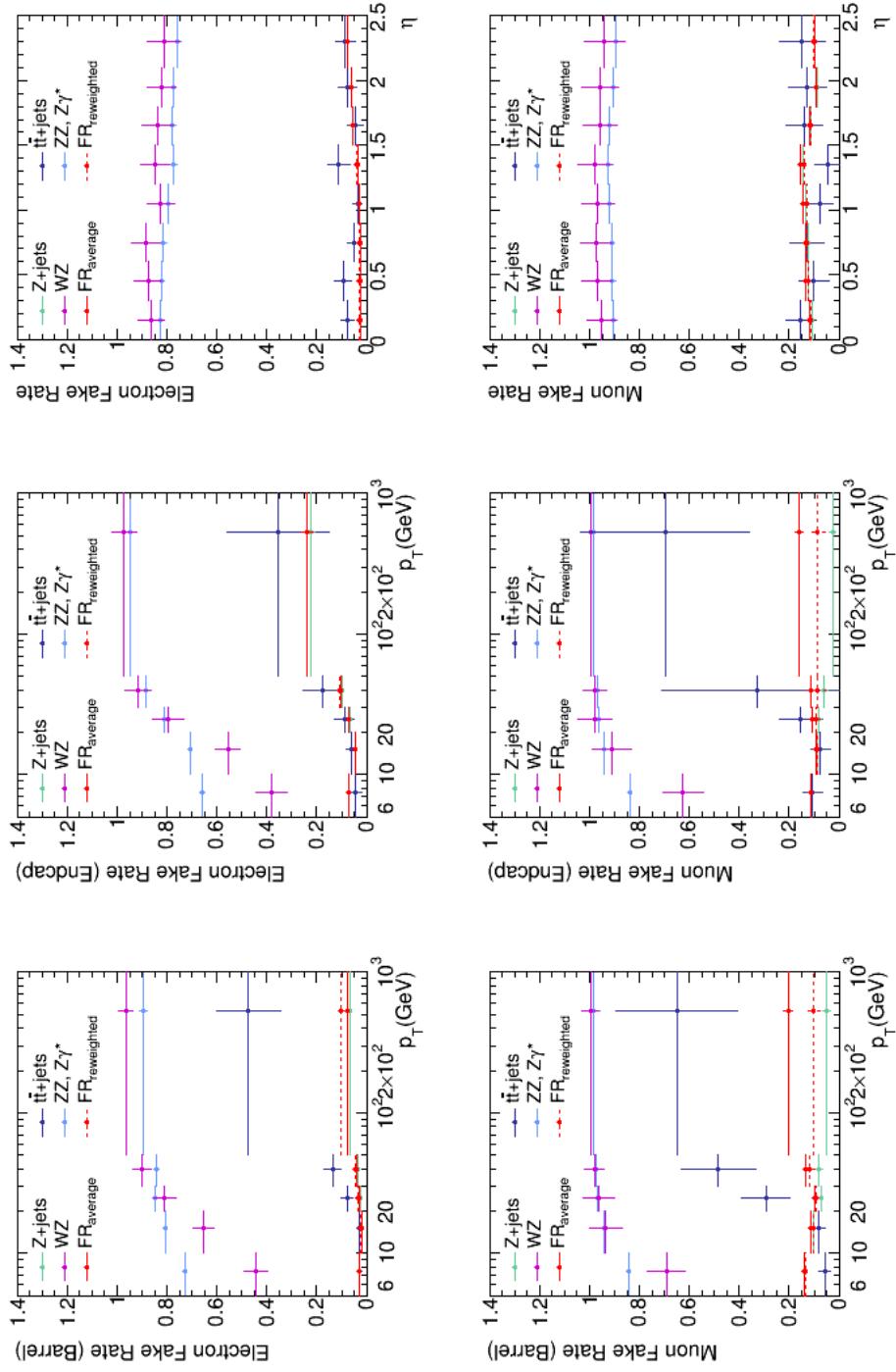
Source: The AUTHOR, 2018.

8.4.3 Uncertainties on the Estimation of Z+X

The statistical uncertainty on the $Z+X$ estimation comes from the limited number of events selected in the control regions where one measures and applies the fake ratio methods. Such uncertainty is typically in the range of 2-7% (SM Higgs) and 8-61% (VBF-SR). The systematic uncertainty arises because the composition of reducible backgrounds in the control regions where one measures and applies the fake ratios are usually not the same. This is the main source of systematic uncertainty of the fake ratio method. This uncertainty is estimated in three steps. First, one measures the fake rates for each background process ($DY, t\bar{t}, WZ, ZZ, Z\gamma$) using MC and applying the selections explained in the Sec. 8.4.1 to obtain samples of $Z + l$. Then, one defines the fake rates for MC (simulation) by computing the weighted average of those individual fake rates. Second, one reweights the individual fake rates using the expected yields of the reducible backgrounds selected in the 2P2F control region. Then, one applies the average and the re-weighted fake rates in order to get the final yields of $Z+X$ and finally, the difference between them is used as an estimation of the uncertainty on the measurement of fake rates. The average and the re-weighted fake rates are shown in Tab. 16.

The fake rates versus p_T and η computed for each reducible background are shown in Fig. 37 and the 2D maps used to apply the fake rates can be seen in Fig. 37. The final yields for $Z+X$ along with the statistical, systematic and combined uncertainties are shown in Tab. 17.

Figure 37 - Fake rate, for each reducible background ($D\gamma$, $t\bar{t}$, WZ, ZZ, $Z\gamma$), the average and the re-weighted distributions for electron (a) p_T barrel, (b) p_T endcap and (c) η and for muons (d) p_T barrel, (e) p_T endcap and (f) η .



Source: The AUTHOR, 2018.

Table 17 - Final yields estimated for Z+X in the signal region from the measurements on data using the fake rate methods of opposite-sign (OS-OS) and same-sign (OS-SS) leptons. The estimates are reported with the total uncertainty for each final state after the SM Higgs and VBF-SR selections. The m_{4l}^{OS}/m_{4l}^{SS} is the ratio between the m_{4l} distribution for each final state obtained for Z+X in the OS and SS procedures (it is not computed for VBF-SR due to the very few Z+X events estimated there).

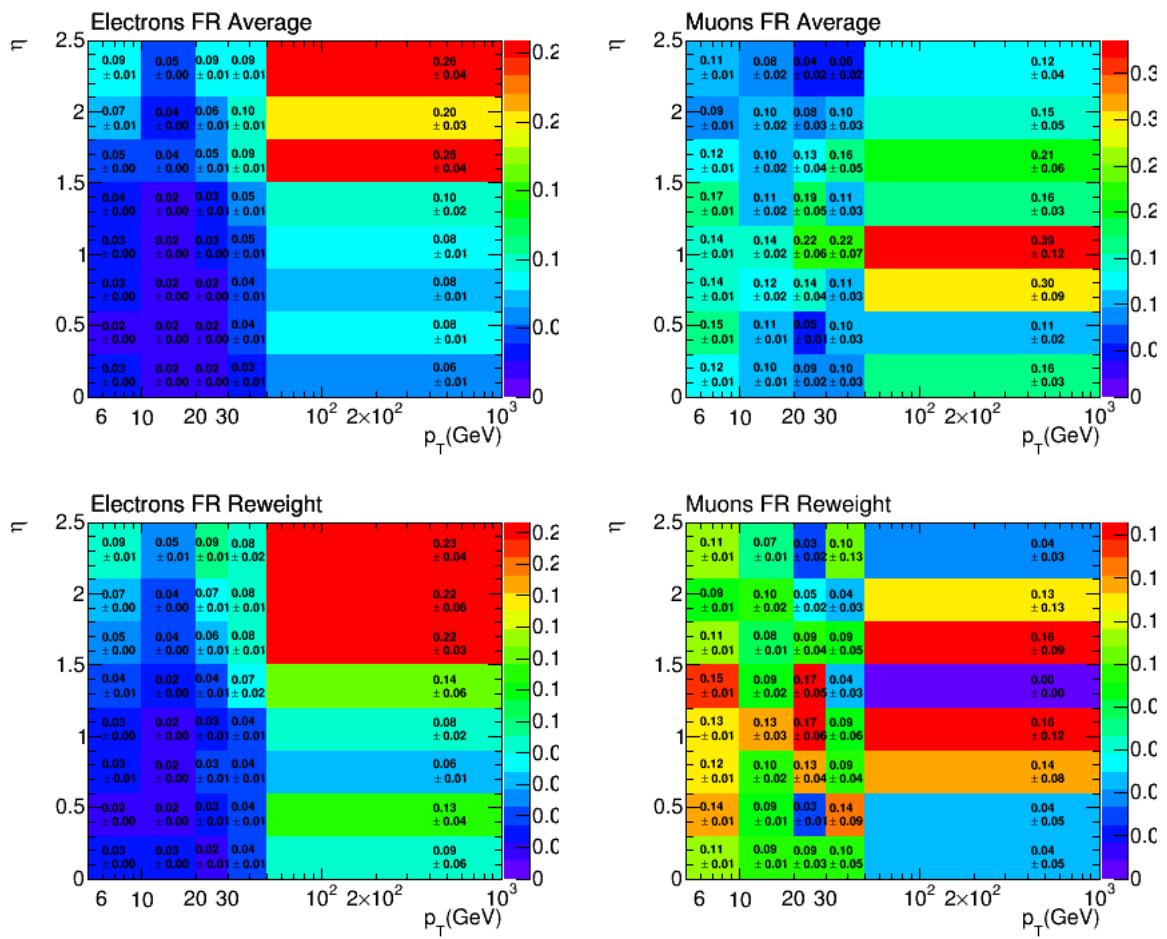
Z+X	4 μ	4e	2e2 μ	4l
		SM Higgs		
OS-OS	24.28 ± 7.79	27.80 ± 1.84	56.90 ± 6.71	108.99 ± 10.33
OS-SS	24.00 ± 4.90	36.00 ± 6.00	64.00 ± 8.00	124.00 ± 11.14
m_{4l}^{OS}/m_{4l}^{SS}	0.75 ± 0.31	0.88 ± 0.33	0.98 ± 0.35	0.85 ± 0.33
		VBF-SR		
OS-OS	2.03 ± 0.49	0.33 ± 0.20	2.72 ± 0.42	5.08 ± 0.68
OS-SS	1.00 ± 1.00	1.00 ± 1.00	0.00 ± 0.00	2.00 ± 1.41

Source: The AUTHOR, 2018.

8.4.4 Same-Sign Cross-checking and Final Z+X Estimation

A cross-checking procedure (see [18, 77]) is applied in order to verify the Z+X estimation from OS method by applying the selections used in the Same-Sign (SS) method. The SS method requires similar selections to get the 2P2F and 3P1F control regions explained in Sec. 8.4.2. The events are obtained as a subset of the events satisfying the first step of the selection, which correspond to the first Z selection composed by two opposite-sign and same-flavor (SF) leptons. Additionally, one requires a pair of loose leptons same-sign (avoiding signal contamination) and same-flavor. The SS-SF leptons are requested to pass the SIP_{3D} requirement (as previously in the OS method) and, now, also to pass both the identification and isolation criteria, which are imposed to the signal events. As for the SM Higgs selections one requires that $m_{ll} > 4\text{GeV}$ (the QCD suppression cut between any two leptons), $12 < m_{ll} < 120\text{GeV}$ and the $m_{4l} > 70\text{GeV}$. The observed events passing this selections provide a good estimate of the reducible background. The estimates of Z+X applying the OS method (OS-OS) and the SS method (OS-SS) are shown in Tab. 17 for each final state after the SM Higgs and VBF-SR selections and, they are compatible within the uncertainties. Additionally, it has been computed the ratio between the m_{4l} distribution in each final estate for the $Z + X$ estimated by the OS and SS procedures. That ratio (m_{4l}^{OS}/m_{4l}^{SS}) is also shown in Tab. 17 for the events passing the SM Higgs selections, showing the agreement between the two procedures within the uncertainties. Additionally, one also can see that the estimation of Z+X is in agreement with the central HZZ4L analysis by looking at Tab. 20 (SM Higgs) and Tab(s). 15-19 (VBF) of [72].

Figure 38 - Fake rate versus p_T and η for the average and the re-weighted distributions for electron (a) and (c) and, for muon (b) and (d), respectively.



Source: The AUTHOR, 2018.

8.4.5 Z+X Neural Network Output Shape

The samples of events created in the Z+X analysis for the control regions have four leptons, all possible jets and the MET passing the requirements specified in Sec. 5 and Sec. 6. In this way, the events composing the 2P2F and 3P1F control regions can be used as input to a NN and produce the correct shape of the output discriminant. Then, in order to get the final yields and shapes one follows the procedure described in the previous subsections. The yields of course will not change as far as a cut is not applied to the NN discriminant. So, the full shape of the NN gives the same yields reported in Tab. 17 for VBF-SR. The Z+X shape derived in such way is the one used for the final statistical analysis. The statistical and systematic uncertainty are automatically included by the procedure described above and additionally by the systematic uncertainty from each object used as input to the NN discriminant as it will be discussed in Sec. 9.3.

9 SYSTEMATIC UNCERTAINTIES

The systematic uncertainties and their treatment in this analysis is discussed in this section. The general strategy is first to rely on the studies done for the SM HZZ4L search, since the selections applied in the present analysis are based on it. Then, the uncertainties associated directly to the signal and background modeling by the Neural Network, due to the uncertainties associated to the objects used as inputs, is discussed and follows similar strategy to what has been done for the MELA discriminants by the CMS Collaboration [76].

9.1 Experimental Uncertainties

The first source of experimental uncertainties, which affect both signal and background, are the uncertainty on the integrated luminosity, equals to 2.6%, and the uncertainty on lepton identification and reconstruction efficiency, which variates from 2.5 to 9% on overall event yield for the 4μ and $4e$ channels, respectively. The uncertainty on the lepton energy scale is estimated by measuring the difference between the position of $Z \rightarrow ll$ peak reconstructed from data and simulation, as described in Sec. 9 of [72], and has been determined to be 0.04% (0.3%) for the 4μ ($4e$) channel. The uncertainty on the $4l$ mass resolution, coming from the uncertainty on each lepton energy resolution, is estimated to be 20% as described in Sec. 5 of [72]. Experimental uncertainties in the fake rate method are the ones used to estimate the $Z + X$ background uncertainty, as described in Sec. 8.4. This uncertainty is introduced in the analysis by using the $Z + X$ shapes from the average and reweight procedures and it amounts from 6% ($4e$) to 23% (4μ).

Table 18 - Summary of experimental systematic uncertainties accounted into this analysis.

Source	Magnitude (%)
Luminosity	2.6
Lepton $\epsilon_{ID/Reco}$	2.5-9
Lepton energy scale	0.04-0.30
m_{4l} resolution	20
Jet energy scale	3.3
E_T^{miss}	7-26
b-tagging	1
$Z + X$	6-23

Source: The AUTHOR, 2018.

Table 19 - Summary of theoretical systematic uncertainties accounted into this analysis.

Source	Magnitude (%)
QCD scale (VBF)	+0.4/-0.3
PDF set (VBF)	± 2.1
QCD scale (gg)	± 3.9
PDF set (gg)	± 3.2
Bkg K factor (gg)	± 10.0
QCD scale (WH)	+0.5/-0.7
PDF set (WH)	± 1.9
QCD scale (ZH)	+3.8/-3.1
PDF set (ZH)	± 1.6
QCD scale ($t\bar{t}H$)	+5.8/-9.2
PDF set ($t\bar{t}H$)	± 3.6
QCD scale ($q\bar{q} \rightarrow ZZ$)	+3.2/4.2
PDF set ($q\bar{q} \rightarrow ZZ$)	+3.1/-3.4
Electroweak corrections ($q\bar{q} \rightarrow ZZ$)	± 0.1
$BR(H \rightarrow ZZ \rightarrow 4l)$	2.0

Source: Adapted from [72, 77].

9.2 Theoretical Uncertainties

The theoretical uncertainties, affecting both signal and background estimation, include the renormalization and factorization scales and the choice of the Parton Density Function (PDF) set. The variation of these scales between 0.5 and 2 times their nominal value, while keeping their ratio between 0.5 and 2, provides the uncertainty on the normalization and factorization scales. The uncertainty from the PDF set is determined by taking the root mean square of the variation when using different replicas of the default NNPDF set. An additional uncertainty of 10% on the k-factor used for the prediction of $gg \rightarrow ZZ$ is applied as described in Sec. 8.3. Associated to the $H \rightarrow ZZ \rightarrow 4l$ BR there is 2% of systematic uncertainty, which affects only the signal yields (VBF and the others).

9.3 Discriminants Shape Uncertainties

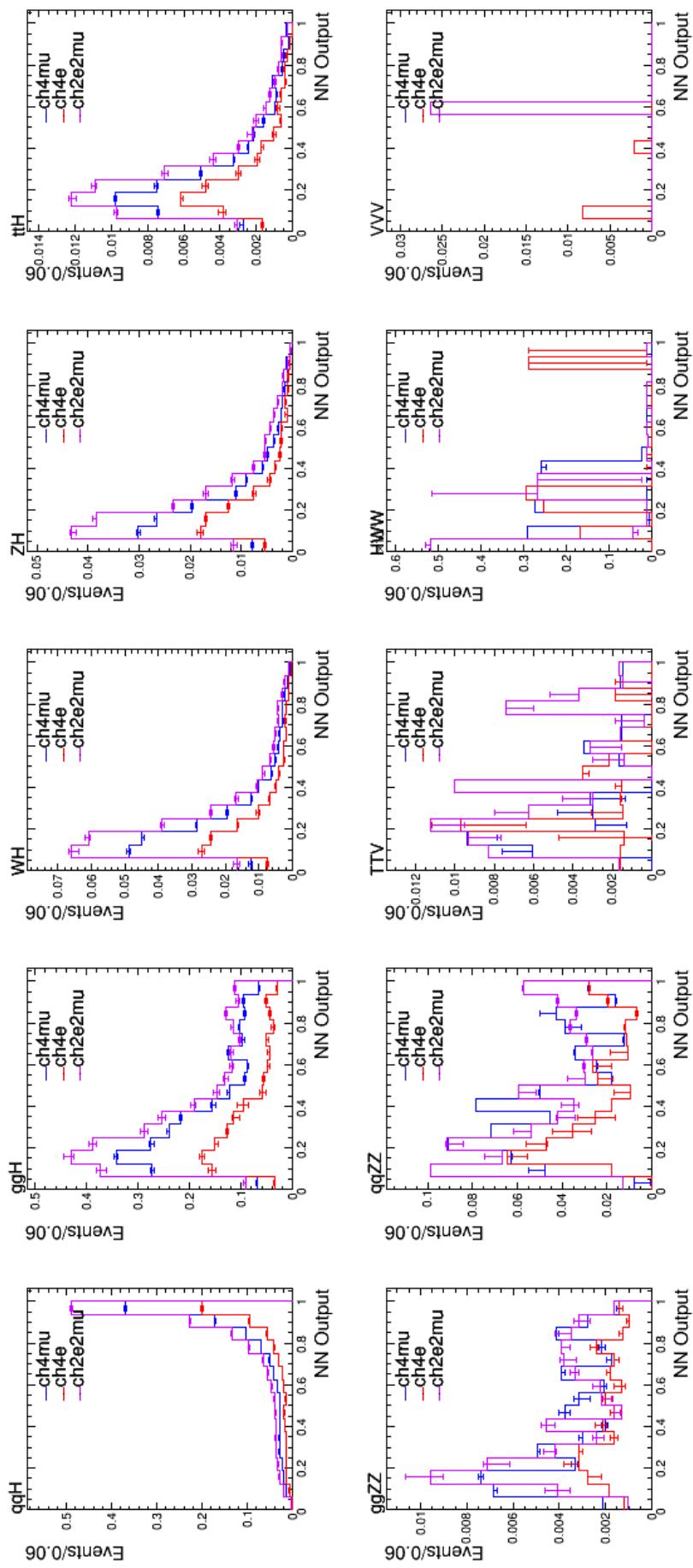
The systematic uncertainties in data classification using the NNs were estimated by checking the impact of systematic uncertainties in the input variables on the shape of the NN output. The main types of systematic uncertainties are the ones that changes the values of the NN output, which can cause event migration across the signal and background regions defined by the discriminant. That is an issue one should take into account when applying some specific cut in the discriminant. For shape analysis the

effect is amplified since, for binned analysis, the bins will shift up and down due to the uncertainties. Also, the bin width has some impact: a smaller bin width increase the probability of event migration across the bins and one would expect larger fluctuations than in the case with larger bin width.

The procedure to estimate the uncertainties propagated through the NNs was based on similar procedure adopted by [87–89]. The systematic uncertainties of the objects used as inputs for the NNs were applied to the nominal input value. Those are the uncertainty in the leptons and jet energy and the uncertainty in the MET measurement (for the cases where it is used). The leptons energy uncertainty comes from Particle Flow (muon calibrator and calorimeter-based - for electrons). The uncertainty on jet energy comes from the jet energy correction procedure. The MET uncertainties come from the 1σ shift up and down of the energy/resolution of different PF objects used to compute the MET.

The NNs were fed with the nominal inputs (no systematic uncertainty shifts applied) and with $\pm 1\sigma$ shifts from the nominal inputs. These shifts are done one variable at a time such that after all shifts have been done there are $N_{(Inputs)} \times [1 + 2 \times N_{(InputUncertainties)}]$ output values for each event (and thus the same amount of NN distributions). Fig. 39 shows the impact of the systematic uncertainties on the shape of a NN. The error bars in each bin shows the maximum oscillation (up/down) of the yield in that bin due to the shifting of the inputs $\pm 1\sigma$ from their nominal values.

Figure 39 - Effect of systematic uncertainties in the shape of one of the trained NNs for each of the processes and lepton channels. In this plot the histograms show the nominal distribution of the NN and the error bars indicate the maximum yield oscillation (up/down) per bin due to the systematic uncertainties of the inputs.



Source: The AUTHOR, 2018.

9.3.1 Systematic Uncertainty on the 3rd Jet

As mentioned in Sec. 7.3, two jet-based categories have been defined in this analysis. One of the reasons for it is the possibility to properly handle the systematic uncertainty associated to the usage of the third jet. This uncertainty arises because the third jet in the VBF sample (see Sec. 4) is generated at LO. Currently there is a better description of that jet at NLO (VBF-H3J), implemented in the Powheg-Box V2 MC generator. However, its available version is not a MiNLO as it is for the gluon-fusion process (which, by the way we are using in this analysis in replacement of the standard gluon-fusion sample used in the HZZ4L central analysis). Due to that, the cross-section of the VBF-H3J as computed by the generator is not inclusive for the other categories and thus, it can not replace the present VBF sample (VBF-H2J) used in this analysis.

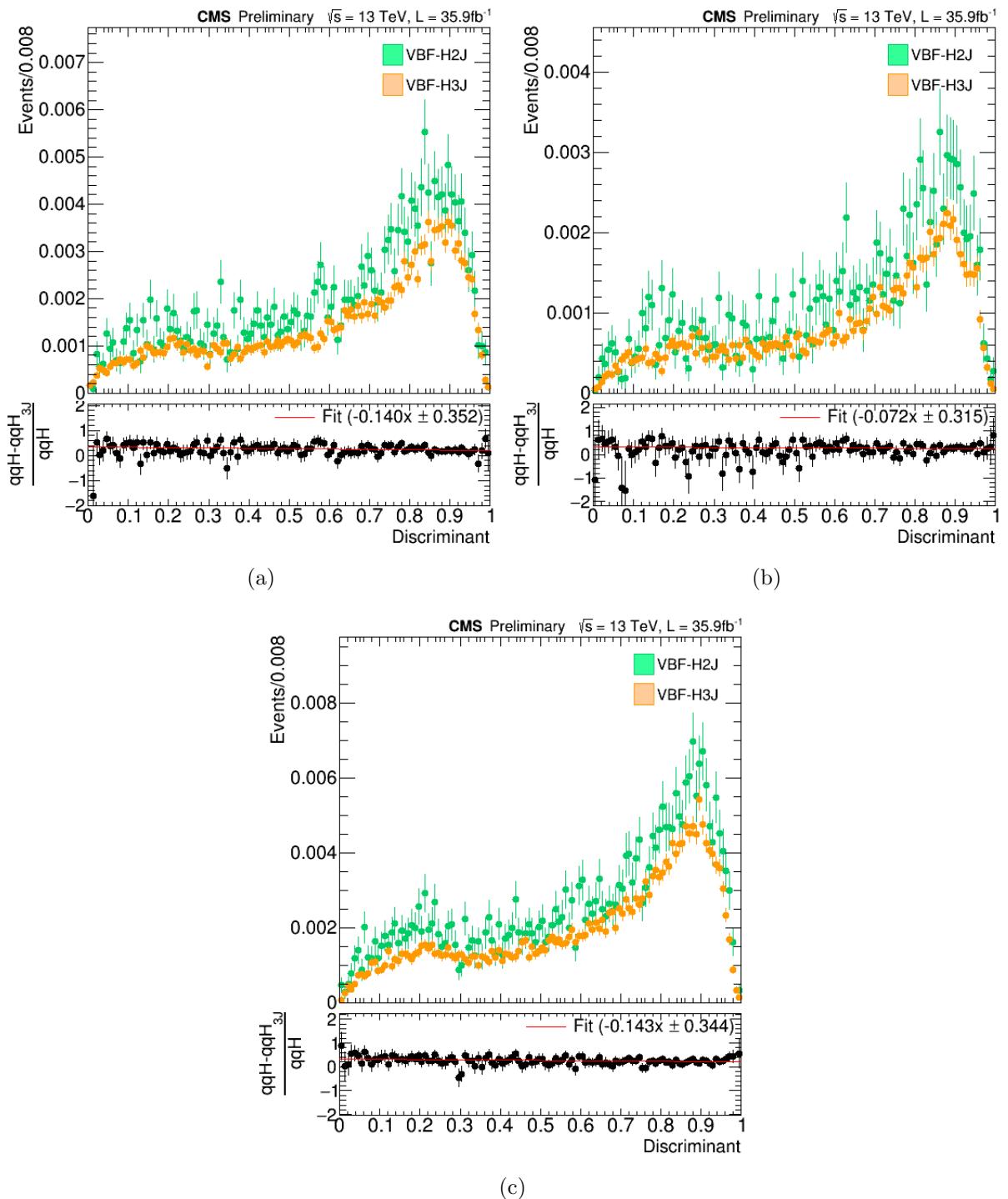
For that reason, the third jet of the VBF-H3J is used and a procedure to estimate the uncertainty on it, affecting the NN, were made using a private VBF-H3J sample⁹. This sample was produced using the Powheg-Box V2 MC generator and the events were processed through the complete CMS simulation chain adopted for RunII 2016.

The events in the **Njets3** category from each sample were fed into the NN and the ratio between their shapes gives an uncertainty in both normalization and shape variation between the two processes. This was done for each four-lepton final state separately. Fig. 40 shows the NN distribution from the two samples and their ratio.

The fit of the ratio smooth down the statistical variation bin by bin, since the VBF-H2J has less events. Using the linear function derived from that ratio, and uncertainty (up and down) is introduced around the distribution given by the VBF-H2J sample, which is the nominal one. Such uncertainty is introduced in the statistical analysis following the procedure for shape analysis as it will be explained in Sec. 10.2.

⁹ This procedure was adopted since it was concluded, after some discussion with MC generators experts, that a merging procedure between the two samples is not straightforward.

Figure 40 - NN distributions and their ratio in the **Njets3** category from VBF-H2J and VBF-H3J for (a) 4μ , (b) $4e$ and (c) $2e2\mu$.



Source: The AUTHOR, 2018.

10 RESULTS

In this section are summarized the yields and distributions obtained after the analysis selection showing the inputs that are used into the statistical analysis for the estimation of the limit on the Higgs VBF production mode.

10.1 Yields and Distributions

The number of estimated events from signal and background, as well the number of observed events after the full selection for SM Higgs, as explained in Sec. 6, are reported in Tab. 20, for $m_{4l} > 70$ GeV. Tab. 12 shows the same for the events selected through the VBF-SR selections, which combines the SM Higgs selections and the extra requirements explained in Sec. 6.3. The quoted uncertainties include both statistical and systematic sources and, have been derived (as pre-fit statistics) through the statistical tool explained in Sec. 10.2.

Since it is proposed in this analysis the usage of up to three jets, differing of the standard VBF category in which one takes only two jets, an extra categorization have been implemented as mentioned previously. The events selected in the VBF-SR (Sec. 6.3) were split into two categories accordingly to the number of jets passing the selections. Such categories, labeled as **Njets2** and **Njets3**, are orthogonal since they are mutually exclusive. The discriminant from each category will be from now on referred as $D_{qqH,2J}^{NN}$ and $D_{qqH,3J}^{NN}$ for simplicity.

The Fig(s). 43(a), (b) show the distribution of the final neural networks chosen to be the discriminant in the **Njets2** and **Njets3** categories, while the Fig. 43(c) shows their combination. This distributions include the three four-lepton final states. The $Z + X$ estimation follows the same procedure explained in Sec. 8.4 and the shape has been derived by feeding the NNs with the events selected in the two CRs that are used to make the estimation. The statistical analysis is performed for each separated category and also for the combined case.

10.2 Statistical Analysis

The discriminants $D_{qqH,2J}^{NN}$ and $D_{qqH,3J}^{NN}$ (Fig. 43), as well their statistical and systematic uncertainties, derived for signal, background and observed data as explained in previous sections, constitute the inputs (in form of histograms) to perform the statistical analysis. Such analysis is carried out by using the HiggsAnalysis-CombinedLimit package,

Table 20 - Number of expected events from background and signal, with total (statistical+systematic) uncertainty reported, and number of observed events after the SM Higgs selections in the mass range $m_{4l} > 70$ GeV.

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	19.34 ± 3.73	11.02 ± 2.34	25.99 ± 5.20	56.35 ± 6.81
VH	1.45 ± 0.21	0.92 ± 0.14	2.14 ± 0.32	4.51 ± 0.41
ttH	0.36 ± 0.08	0.23 ± 0.05	0.48 ± 0.11	1.07 ± 0.14
qqZZ+ZZJJ	387.01 ± 24.48	234.64 ± 24.81	538.35 ± 43.61	1160.00 ± 55.83
ggZZ	65.81 ± 7.43	43.85 ± 6.14	102.32 ± 12.60	211.98 ± 15.86
Z+X	24.28 ± 7.71	27.80 ± 8.09	56.90 ± 17.09	108.99 ± 20.42
\sum backgrounds	498.25 ± 26.98	318.46 ± 26.91	726.18 ± 48.78	1542.89 ± 61.90
qqH (signal)	1.86 ± 0.36	1.10 ± 0.24	2.53 ± 0.52	5.49 ± 0.68
Total expected	500.11 ± 26.98	319.56 ± 26.91	728.71 ± 48.78	1548.38 ± 61.90
Observed	503	287	669	1459

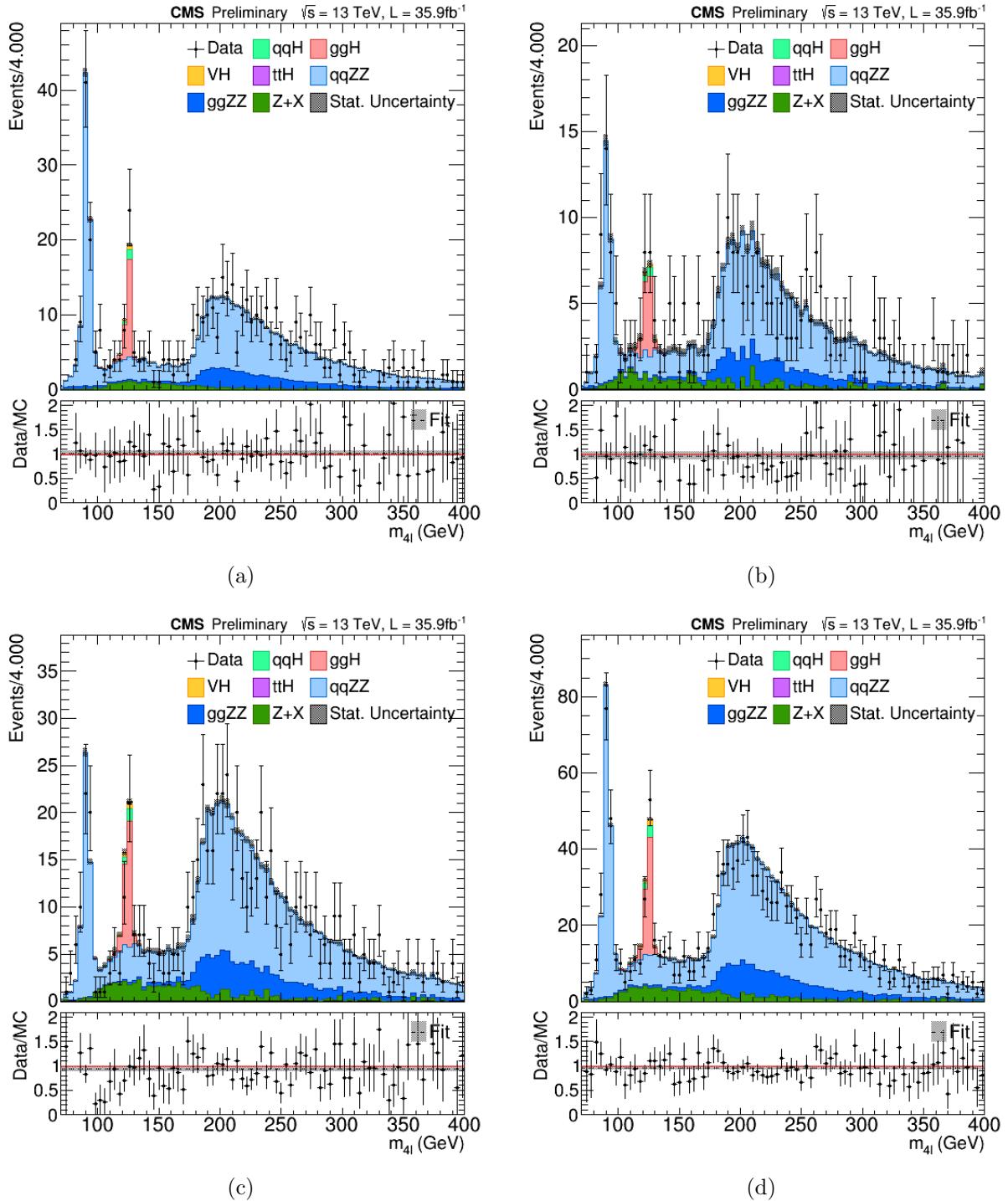
Source: The AUTHOR, 2018.

Table 21 - Number of expected events from background and signal, with total uncertainty (statistical+systematic), and the number of observed events after the VBF-SR selections (see Sec. 6.3).

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	2.46 ± 0.53	1.28 ± 0.30	3.10 ± 0.69	6.84 ± 0.92
VH	0.34 ± 0.05	0.20 ± 0.03	0.46 ± 0.07	1.00 ± 0.09
ttH	0.05 ± 0.01	0.03 ± 0.01	0.06 ± 0.01	0.14 ± 0.02
qqZZ+ZZJJ	0.67 ± 0.04	0.36 ± 0.04	0.74 ± 0.06	1.77 ± 0.08
ggZZ	0.05 ± 0.01	0.03 ± 0.01	0.06 ± 0.01	0.14 ± 0.02
Z+X	2.03 ± 0.83	0.33 ± 0.04	2.72 ± 0.76	5.08 ± 1.12
\sum backgrounds	5.60 ± 0.99	2.23 ± 0.31	7.14 ± 1.03	14.97 ± 1.46
qqH (signal)	1.05 ± 0.22	0.58 ± 0.13	1.39 ± 0.30	3.02 ± 0.39
Total expected	6.65 ± 1.01	2.81 ± 0.34	8.53 ± 1.07	17.99 ± 1.51
Observed	5	2	10	17

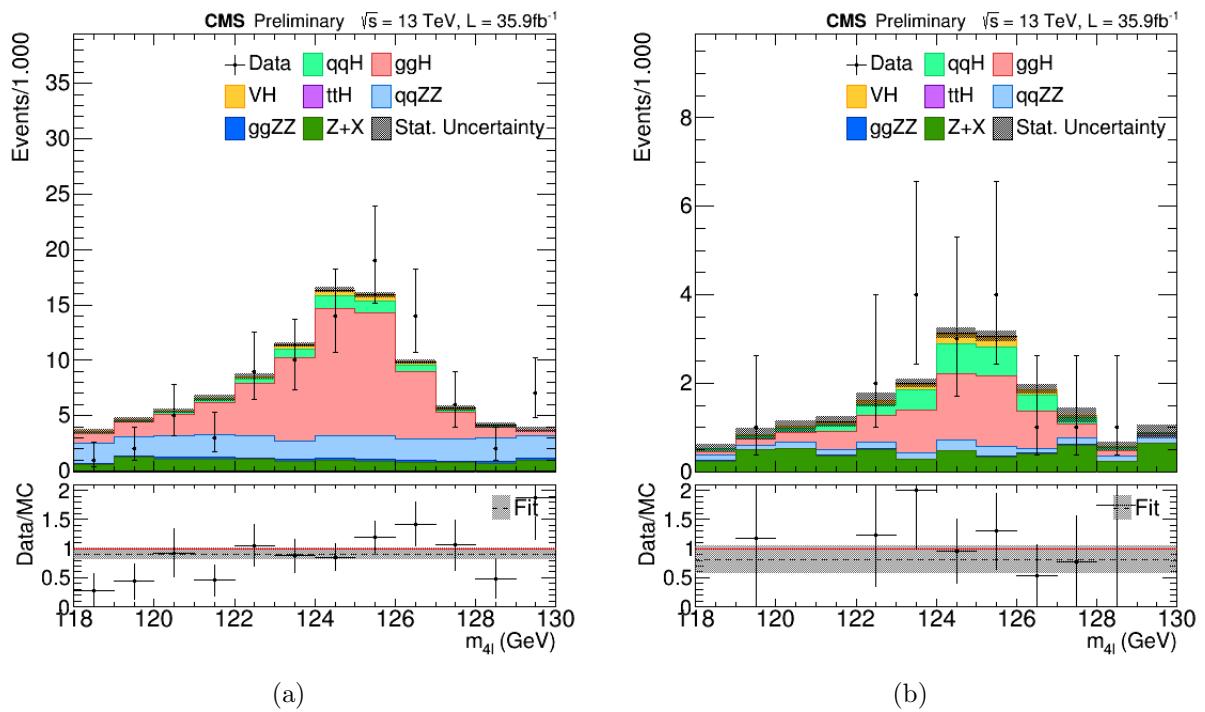
Source: The AUTHOR, 2018.

Figure 41 - Final four-lepton mass distributions for the SM Higgs separated in the channels (a) 4μ , (b) $4e$ and (c) $2e2\mu$ and, combined into the (d) $4l$ final state. The distributions are shown for $70 < m_{4l} < 400$ GeV and include the $Z + X$ background estimated by the data-driven method explained in Sec. 8.4.



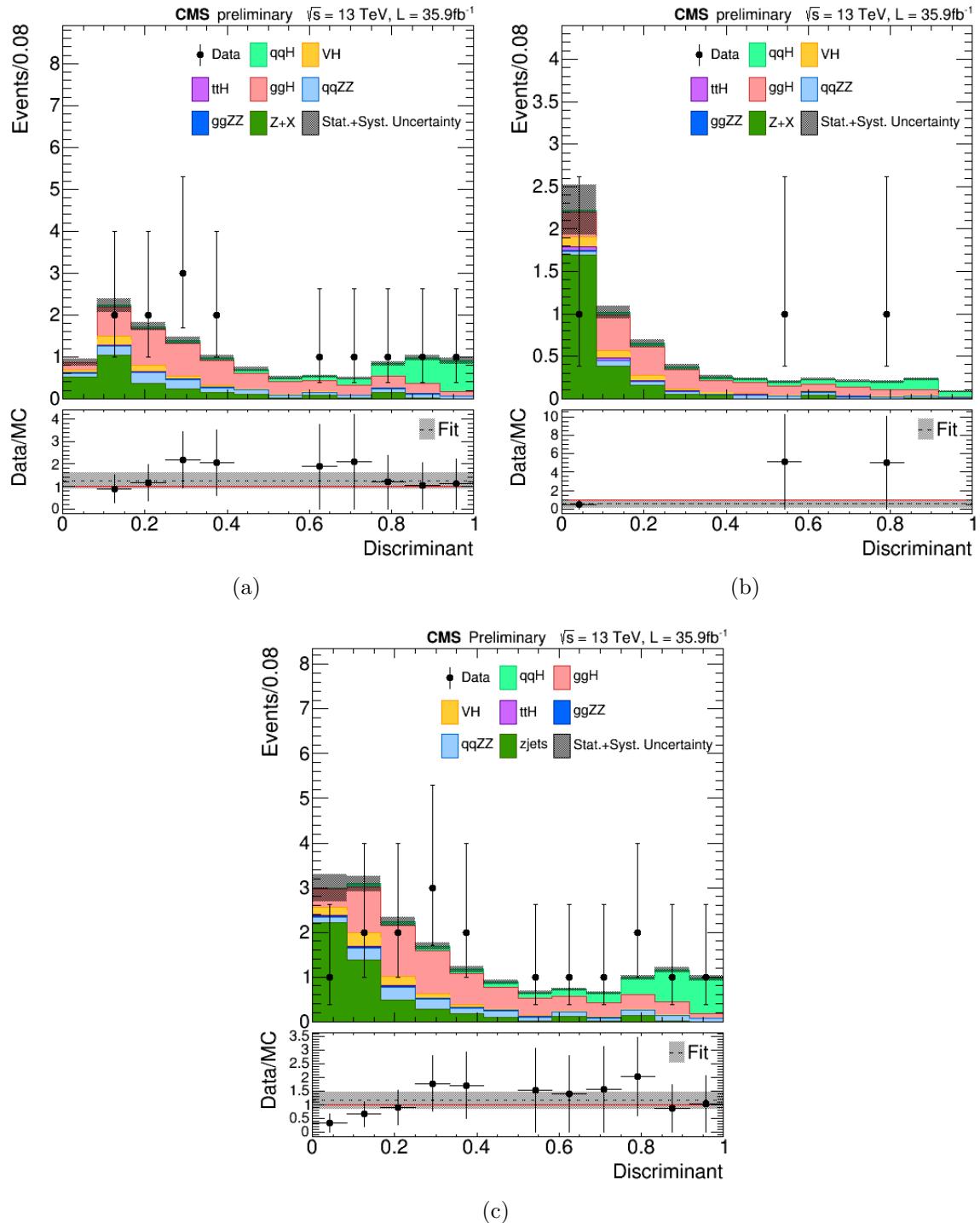
Source: The AUTHOR, 2018.

Figure 42 - Final four-lepton mass distributions for the events passing the (a) SM Higgs and the (b) VBF-SR selections combined into the $4l$ final state. The distributions are shown for $118 < m_{4l} < 130$ GeV (SM Higgs signal region defined by CMS Collaboration) and include the $Z + X$ background estimated by the data-driven method explained in Sec. 8.4.



Source: The AUTHOR, 2018.

Figure 43 - Pre-fit distributions of the neural network developed for (a) **Njets2** and (b) **Njets3** categories. In (c) is the distribution resulting from the combination of the two categories.



Source: The AUTHOR, 2018.

which assembles a collection of RooStats-based software [100, 101]. The HiggsAnalysis-CrossedLimit package returns the expected and observed upper limits on the cross section of the VBF process times the branching ratio at 95% confidence level (CL). The one and two sigma deviations from the expected limits are also calculated. The VBF signal strength (μ_{qqH}), which is the ratio between the observed and the theoretical cross sections are also computed.

As recommended by the LHC Higgs Combination Group [100] the so called HybridNew method, which computes fully frequentist limits, is the one used in this analysis. Also, because the number of background and signal events in certain categories and/or four-lepton final state don't satisfy the requirements needed to apply the AsymptoticLimits method [102]. The statistical and systematic uncertainties are properly taken into account by defining log-normal and shape uncertainties that affects the shape analysis of the discriminants.

In order to extract the VBF signal strength μ_{qqH} a multi-dimensional likelihood fit is performed on the discriminants $D_{qqH,2J}^{NN}$ and $D_{qqH,3J}^{NN}$. The systematic uncertainties enter in such likelihood as nuisance parameters, which are left to float during the fit. The likelihood implemented within the Higgs Combine tool is a function of the probabilities related to the parameters of interest (POIs) and the systematic uncertainties in the measurements (the nuisance parameters). It can, thus, be expressed as [102],

$$\mathcal{L}(\vec{\mu}, \vec{\theta}) \sim p(Data|\vec{\mu}, \vec{\theta}) \cdot \pi(\vec{\theta}) \quad (63)$$

where the $\vec{\mu}$ stands for the POIs and $\vec{\theta}$ for the values of the nuisance parameters. The POI in the present case is the μ_{qqH} . The essence of this likelihood analysis is based on the number of observed events (n) such that the probability term is just a *Poisson* probability which has the simplest form of

$$p(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \quad (64)$$

where the $\lambda = \mu \cdot s + b$, being s and b the representation of the expected yields for signal and background, respectively. In the case of a binned shape analysis (the present one), the probability $p(n|\lambda)$ becomes the product of the individual bin probabilities (product rule of probability), and thus,

$$p \rightarrow \prod_i p_i = \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \quad (65)$$

For the uncertainties present in this analysis the probability function is

$$\pi(\vec{\theta}) = e^{-\frac{1}{2} \theta^2} \quad (66)$$

which is known as a *log-normal* nuisance parameter distribution. So, for instance, taking L as the luminosity and supposing it is known to 10% then, this systematic uncertainty affects L by $L.(1+0.1)^\theta$. The nominal case (no systematic uncertainty) corresponds to $\theta = 0$ and, $\theta = \pm 1$ is identified as the $\pm 1\sigma$ uncertainty on a given quantity that affects the analysis.

Now, the value of a likelihood itself does not have any meaning (it can be larger than 1, so it clearly can not be a probability). However, the relative values of likelihoods are useful and one usually computes the so called likelihood ratio. This ratio is computed between the likelihood maximum value and its variation as one goes over different values of the POIs. In order to avoid very small or very large values of the likelihoods one takes their negative logs and thus, the maximum likelihood value means the minimum negative log likelihood. Mathematically,

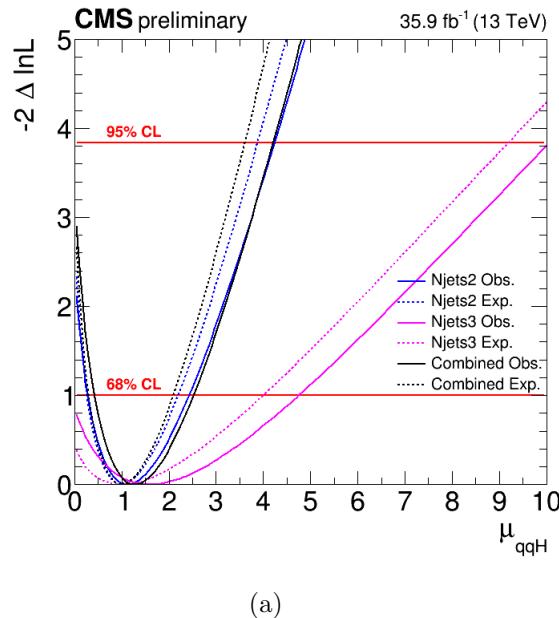
$$-\ln \mathcal{L}(\mu) - [-\ln \mathcal{L}(\hat{\mu})] = -\frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})} \quad (67)$$

where $\hat{\mu}$ is the POI value which maximizes the likelihood. Eq. 67 is usually represented as $-2\Delta \ln \mathcal{L}(\mu)$ in likelihood scan plots. The interesting point of the likelihood ratios comes from Wilk's theorem [103, 104], which states that a 68% CL interval can be found for the POI by looking at the region for which $-\Delta \ln \mathcal{L}(\mu) < 0.5$ (or $-2\Delta \ln \mathcal{L}(\mu) < 1.0$). The minimum value of the likelihood ratio gives the best estimation of a given POI. Fig. 44 shows (a) the likelihood ratio profiles and (b) the summary of the minimum likelihood ratio values and the 68% CL obtained in this analysis for each category and their combination.

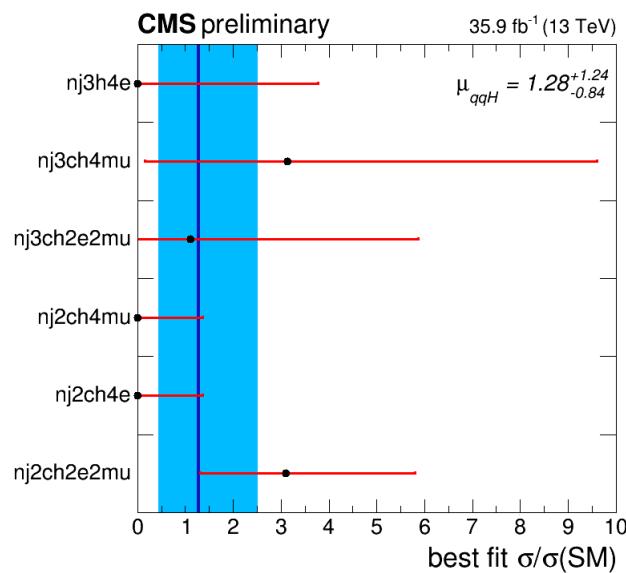
In this case, for the signal strength estimation, the likelihood ratio is performed with $\mu = 1$ (VBF signal expected from SM) and $\hat{\mu}$ is left to float during the fit on the observed events (VBF signal estimated from observed data). In this analysis the Higgs VBF signal strength have been measured to be $\mu_{qqH} = \sigma_{Obs}^{qqH}/\sigma_{SM}^{qqH} = 1.28_{-0.84}^{+1.24}$, based on the SM Higgs boson with mass 125GeV and the collected 2016 data which accounts 35.9fb^{-1} . Tab. 22 summarizes the signal strengths measured in each category and channel and their combination.

The Fig. 45(a), (b) and (c) show the distributions of the $D_{qqH,2J}^{NN}$ and $D_{qqH,3J}^{NN}$ discriminants and their combination, respectively, after fitting the model **s+b** to the observed data. The histograms shown the estimated yields and the associated uncertainty

Figure 44 - The best fit value found, given the observed data, for μ_{qqH} . In (a) is shown the compatibility between the channels from each category of number of jets and in (b) is shown the likelihood ratios profile of the two categories and their combination.



(a)



(b)

Source: The AUTHOR, 2018.

Table 22 - Expected and observed signal strength modifiers from each category and four-lepton channel for 35.9fb^{-1} of observed data at $\sqrt{s} = 13\text{TeV}$.

	$\mu_{qgH}^{4\mu,2J}$	$\mu_{qgH}^{4e,2J}$	$\mu_{qgH}^{2e2\mu,2J}$	$\mu_{qgH}^{4\mu,3J}$	$\mu_{qgH}^{4e,3J}$	$\mu_{qgH}^{2e2\mu,3J}$	$\mu_{qgH}^{4l,2J+3J}$
Expected	$1.00^{+2.13}_{-1.49}$	$1.00^{+3.11}_{-1.00}$	$1.00^{+1.83}_{-0.96}$	$3.13^{+5.79}_{-1.00}$	$1.00^{+8.24}_{-1.00}$	$1.00^{+4.67}_{-1.00}$	$1.00^{+1.08}_{-0.70}$
Observed	$0.00^{+1.36}_{-0.00}$	$0.00^{+1.36}_{-0.00}$	$3.10^{+2.69}_{-1.79}$	$3.13^{+6.47}_{-2.98}$	$0.00^{+3.78}_{-0.00}$	$1.10^{+4.77}_{-1.14}$	$1.28^{+1.24}_{-0.84}$

Source: The AUTHOR, 2018.

Table 23 - Background and signal estimations, with total uncertainty (statistical+systematic), derived from fitting the **s+b** model to the observed data, accounting 35.9fb^{-1} at $\sqrt{s} = 13\text{TeV}$.

Process	4μ	4e	2e2 μ	4l
ggH	2.48 ± 0.34	1.29 ± 0.19	3.14 ± 0.44	6.91 ± 0.59
VH	0.34 ± 0.03	0.20 ± 0.02	0.46 ± 0.04	1.00 ± 0.06
ttH	0.05 ± 0.01	0.03 ± 0.00	0.06 ± 0.01	0.14 ± 0.01
qqZZ+ZZJJ	0.67 ± 0.04	0.36 ± 0.03	0.75 ± 0.05	1.77 ± 0.07
ggZZ	0.05 ± 0.01	0.03 ± 0.00	0.06 ± 0.01	0.14 ± 0.01
Z+X	1.74 ± 0.34	0.29 ± 0.06	2.35 ± 0.44	4.37 ± 1.34
\sum backgrounds	5.34 ± 0.49	2.19 ± 0.20	6.81 ± 0.63	14.34 ± 0.82
qqH (signal $m_H = 125\text{GeV}$)	1.35 ± 0.77	0.76 ± 0.42	1.79 ± 1.01	3.90 ± 1.34
Total estimated	6.69 ± 0.91	2.95 ± 0.47	8.60 ± 1.19	18.24 ± 1.57

Source: The AUTHOR, 2018.

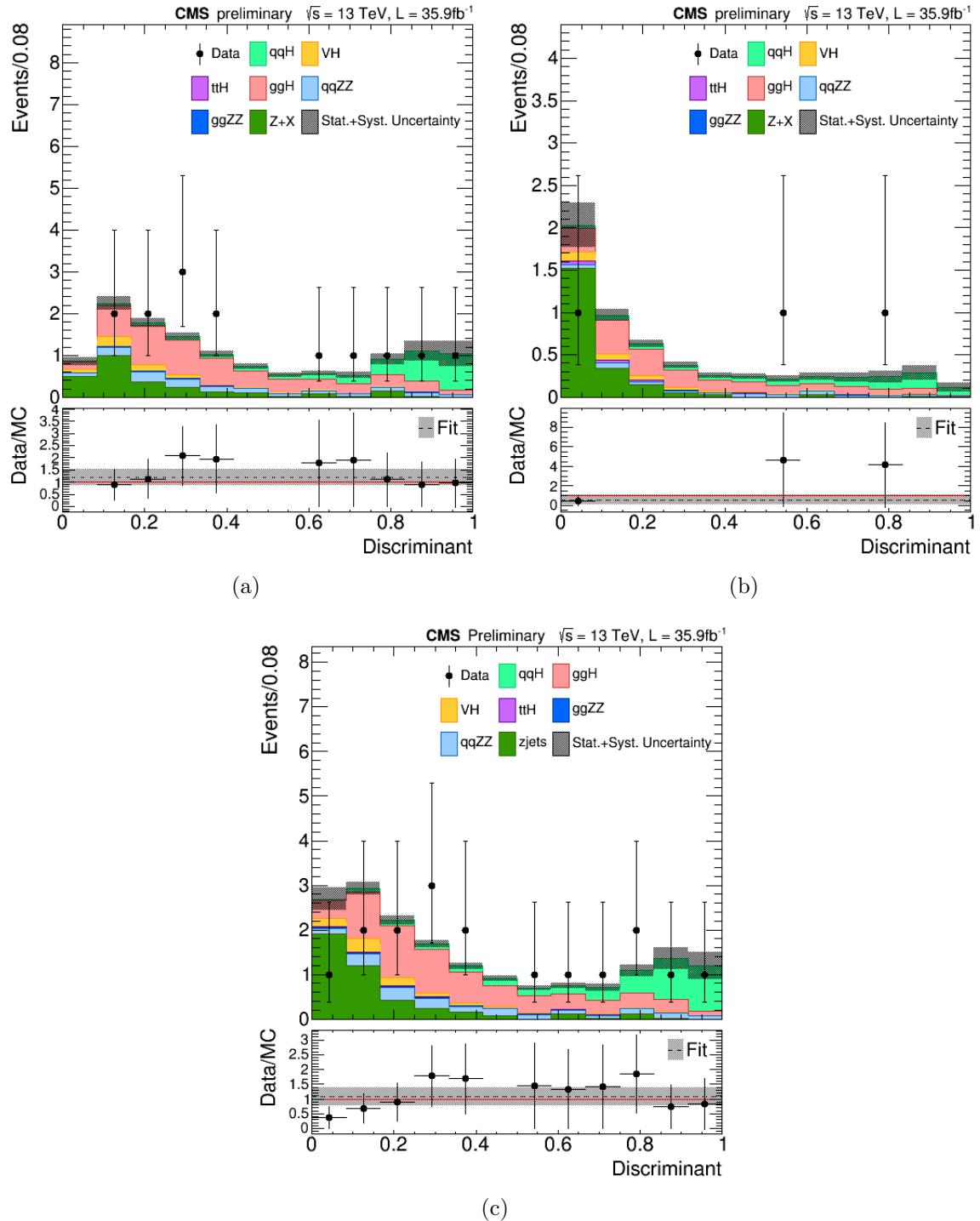
in each bin. Tab. 23 summarizes the total post-fit yields and total uncertainties for each process.

The Higgs Combine toolkit allows one to also compute the impact of each systematic uncertainty on the signal strength estimation. The procedure leaves just one systematic uncertainty floating, while the other ones are kept fixed at their expected values (given by the user in the datacards), during the likelihood scan and retrieves how the estimated signal strength changes due to the floating systematic uncertainty. The impact of the systematic uncertainties present in this analysis over the VBF signal strength, combining the **Njets2** and **Njets3** categories, is shown in Fig. 46. It shows that the resolution in the m_{4l} is the dominant uncertainty, which was also observed in the analysis reported in [77] (which used the same selections as the present analysis until the SM Higgs step - differing only by the extra requirements defining the VBF-SR).

Additionally to the estimation of the Higgs VBF signal strength, 95%CL limits have been also calculated as well the significance of the analysis for each category and four-lepton channel. The limits and significance calculation follows the HybridNew method within Combine tool, which tosses MC toys to compute the distribution of a test statistic.

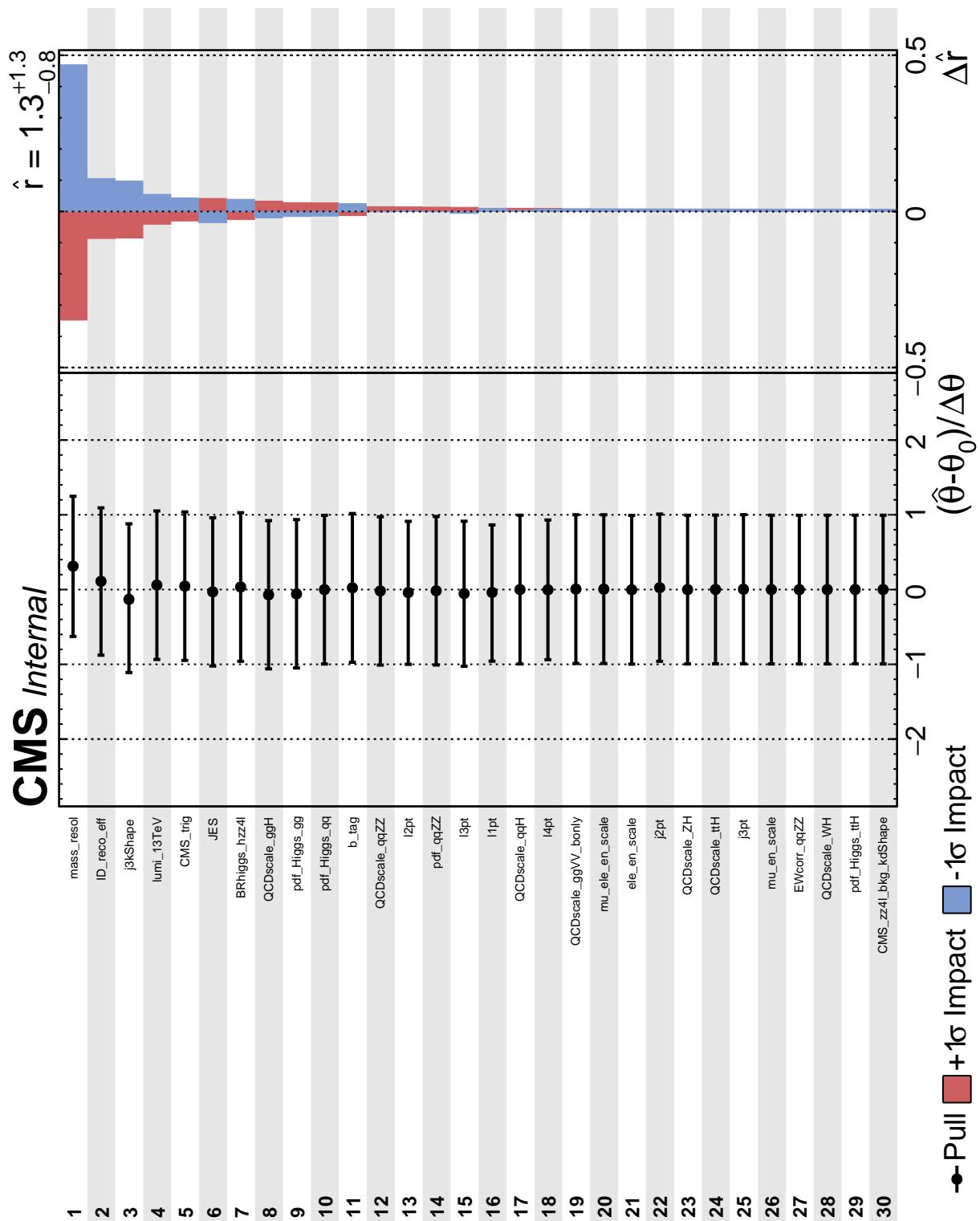
As usual, the limits are computed through the *Frequentist* CLs criterion (Neyman-

Figure 45 - Post-fit NN distribution. The fit is done with the assumption of S+B hypothesis ($\mu_{VBF} = 1$ a priori).



Source: The AUTHOR, 2018.

Figure 46 - Post-fit nuisances impact plot obtained with the combination of the two categories of number of jets. This plot shows the effect of each systematic uncertainty on the estimation of the VBF signal strength (μ_{qqH}).



Source: The AUTHOR, 2018.

Pearson's intervals construction [105–107]), which in the **s+b** exclusion case, when $CL_s \equiv CL_{s+b}/CL_b \leq 0.05$, gives the 95%CL exclusion of a given signal hypothesis. The CL_{s+b} , probability of **s+b** given the value of a test statistic q_μ , is computed as

$$CL_{s+b} \equiv p_\mu \equiv P(q_\mu \geq q_\mu | s + b) = \int_{q_\mu^{obs}}^{+\infty} f(q_\mu^{obs} | \mu, \hat{\theta}_\mu^{obs}) dq_\mu \quad (68)$$

and CL_b as

$$CL_b \equiv 1 - p_b \equiv P(q_\mu \geq q_\mu^{obs} | b - only) = \int_{q_\mu=0}^{+\infty} f(q_\mu | \mu = 0, \hat{\theta}_{\mu=0}^{obs}) dq_\mu \quad (69)$$

where, μ can be set as the unity, case in which one computes the expected limit, or left to float when fitting observed data, giving the observed limit. As one can notice from Eq(s). 68 and 69, the expected limit is computed trying to exclude with 95%CL the expected signal hypothesis ($\mu = 1$) against the background-only hypothesis. The observed limit follows the same idea but μ is derived from the likelihood calculation on the observed data. The test-statistic q_μ adopted by the LHC community (and implemented within Combine tool [102, 106, 107]) is defined in the following way

$$q_\mu = -2 \ln \left[\frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})} \right] \quad (70)$$

where, the variables with "̂" stand for maximum likelihood estimators. Therefore, this test-statistic demands two fits: one for a fixed value of signal strength ($\mu = \hat{\mu}$) and one floating the signal strength. Also, the systematic uncertainties are handled accordingly: in the numerator the uncertainties are the ones which maximize the likelihood for a given μ (which is floating during the fit), while in the denominator the uncertainties are the ones which maximizes the likelihood for the given μ estimator.

Fig. 47(a) shows the expected and observed limits obtained for the categories defined in this analysis as well their combination. In a limit plot, the solid black line (observed limit) indicates the ratio between the observed and expected (from SM) cross sections. The dotted line indicates the expected limit without the presence of given signal, that is, the background-only case. The green and yellow bands indicate the ± 1 and $\pm 2\sigma$ uncertainty bands on the expected limit, respectively. As it can be seen in Fig. 47(a), the solid black line goes above the 2σ background-only expectation. That can be a hint for the presence of Higgs VBF production, or it could be a sign of background processes or even of systematic uncertainties that were not well understood. The conclusion drawn from the limit is that the VBF Higgs expected SM cross section can not be excluded

given the observed data in this analysis. Additionally, this analysis set $\mu_{qqH}^{obs} < 3.79$ and $\mu_{qqH}^{exp} < 1.66$.

In order to quantify the observation in each category and from their combination, the significance have been computed. In this case, one wants to exclude the background-only hypothesis. Similarly to the limits, a test-statistic is implemented within Combine tool and is defined as

$$q_0 = -2\ln \left[\frac{\mathcal{L}(\mu = 0, \hat{\theta}_{\mu=0})}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right]. \quad (71)$$

The p-value, probability of the background fluctuate in such way to give a certain excess over the expected, is computed as the cumulative region of the background-only test-statistic distribution as follow

$$P_0 = \int_{q_0^{obs}}^{+\infty} f(q_0 | \mu = 0) dq_0. \quad (72)$$

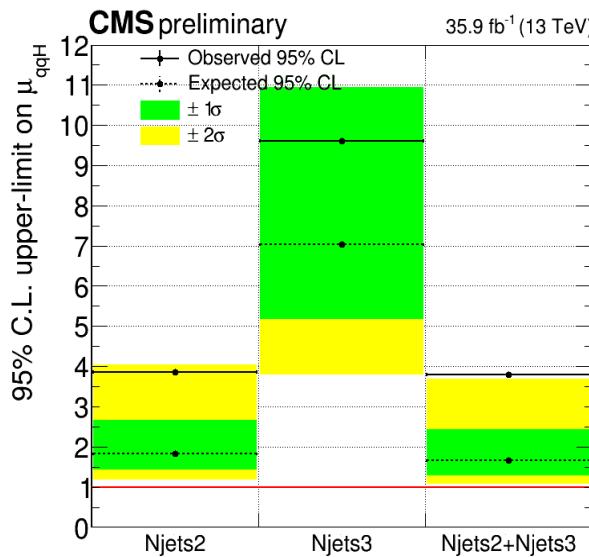
Fig. 47(b) shows the expected and observed significances for the two jet-based categories and their combination. As expected, based on the limits, the significance observed in the two jet-based categories is not enough to state the presence of VBF Higgs events¹⁰. As noticed from the limits, the combination of the two jet-based categories produces an excess over the background-only expectation and thus the significance is higher than the ones observed for each category alone. Even though, that significance ($Z_{qqH}^{obs} = 1.9$, $Z_{qqH}^{exp} = 1.8$) is still not sufficient to state an observation. Even though, this significance is about 2.6x higher than the significance that can be obtained by looking at the m_{4l} distribution. In the next section, a projection of the significance for this analysis in future luminosity scenarios is presented.

10.3 Projection for Future Luminosity Scenarios

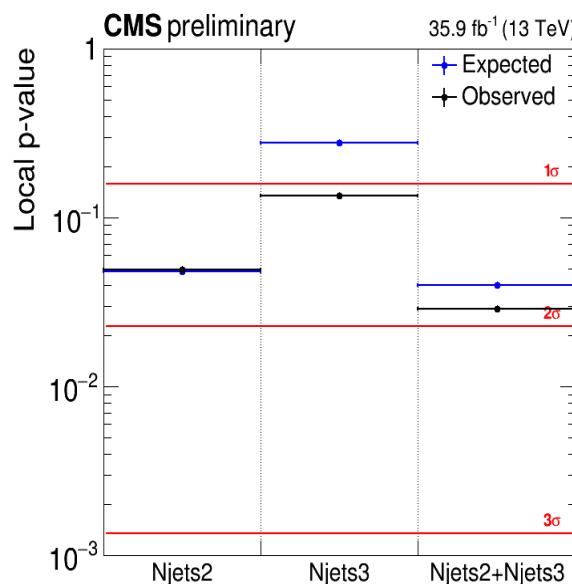
In this section it is presented a simple projection of some results for the future planned luminosities to be achieved by the LHC. The systematic uncertainties have not been fixed, however, none approach to scale them was applied since that demands a more dedicated study of the systematic uncertainties dependences (which goes beyond

¹⁰ For an statement of VBF Higgs observation it is needed $p_0 \leq 2.87e^{-6}$, which corresponds to 5σ .

Figure 47 - (a) Observed and expected 95% CL upper limit on the ratio of the Higgs VBF production cross section to the SM expectation, for the two jet-based categories and their combination. The expected 1σ (2σ) ranges of the expectation for the background-only hypothesis are also shown by the green (yellow) bands. (b) The observed significance of the local excess with respect to the SM background expectation for the two categories and their combination.



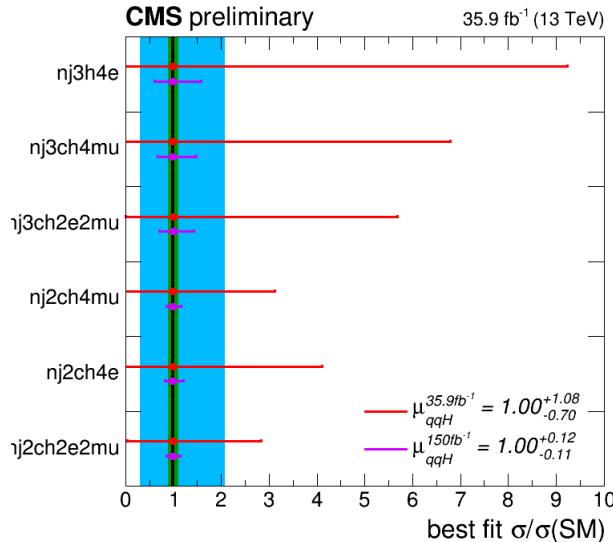
(a)



(b)

Source: The AUTHOR, 2018.

Figure 48 - Comparison of the expected VBF Higgs signal strength best fits, and their uncertainties, between the present luminosity and the future 150fb^{-1} .



Source: The AUTHOR, 2018.

Table 24 - Comparison of the expected VBF significance between the present luminosity (35.9fb^{-1}) and future scenarios at the LHC.

Luminosity (fb^{-1})	35.9	150.0	300.0	359.0	1077.0	1795.0	3000.0
Factor	1.00	4.18	8.36	10.00	30.00	50.00	83.57
$Z_{exp.}$	1.8	3.4	4.7	5.1	8.6	10.9	14.0

Source: The AUTHOR, 2018.

the scope of this analysis). The Fig. 48 shows a comparison between the expected signal strength modifiers for the present luminosity and the projection for 150fb^{-1} scenario. The reduction in the uncertainty of the combined signal strength measurement is expected to be about 87%.

The significance evolution of the present analysis with the future luminosities planned to be achieved at the LHC (and also some intermediary ones) have also been computed. The same procedure adopted for the systematic uncertainties on the expected signal strength estimation above have been used. According to this projection, it is expected that with a luminosity 10x larger than the present one (35.9fb^{-1}) the VBF Higgs signal will have enough significance through using the NN-based discriminants constructed in this analysis.

11 CONCLUSIONS

Write the conclusions!!!

REFERÊNCIAS

- 1 WHITBECK, A. J. Discovery and Characterization of a Higgs-like Resonance using the Matrix Element Likelihood Approach. Thesis (Doctor of Philosophy) - Johns Hopkins University, Baltimore, 162 p., 2013. Available at: <<http://inspirehep.net/record/1286374/files/>>. Accessed on 19/08/2017.
- 2 BRACHEM, C. Studies for a top quark mass measurement using the matrix element method in the semileptonic channel. Dissertation (Master of Particle Physics) - Georg-August-Universität, Göttingen, 94 p., 2012. Available at: <<http://physik2.uni-goettingen.de/research/high-energy/publications/MastersBrachem.pdf>>. Accessed on: 21/08/2017.
- 3 CHOL, S. K. *et al.* Observation of a Resonance like Structure in the $\pi^{+-}\psi'$ Mass Distribution in Exclusive $B \rightarrow K\pi^{+-}\psi'$ Decays. Physics Review Letters, USA, v. 100, n. 14, p. 142001, 2008.
- 4 AALJ, R. *et al.* Observation of the Resonant Character of the Z(4430) - State.. Physics Review Letters, USA, v. 112, n. 22, p. 222002, 2014.
- 5 BERNINGER, J. *et al.* (Particle Data Group). Review of Particle Physics, Physics Review D, USA, v. 86, n. 1, p. 1528, 2012.
- 6 GRIFFITHS, D. Introduction to Elementary Particles. Wiley, 2008.
- 7 HALZEN, F.; MARTIN, A. D. Quarks and Leptons: an Introductory Course in Modern Particle Physics. Wiley, 1984.
- 8 QUIGG, C. Gauge Theories of the Strong, Weak and Electromagnetic Interactions. Menlo Park: Westview Press, 1983.
- 9 ARNISON, G. *et al.* Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c² at the CERN SPS Collider. Physics Letters B, USA, v. 126, n. 18, p.398, 1983.
- 10 BRIANTI, G.; GABATHULER, E. Intermediate vector bosons. Production and identification at the CERN proton anti-proton collider. Europhysics News, France, v. 14, n. 10, p. 1, 1983.
- 11 VARELA, J. Measurement of the properties of the 125 GeV Higgs boson with the CMS detector. Journal of Physics, USA, v. 447, n. 1, p. 012010, 2013.
- 12 OLIVE, K.A. *et al.* (Particle Data Group). Review of Particle Physics. Chinese Physics Courier, China, v. 38, n. 9, p. 1676, 2014.

- 13 CAHN, R.N. and DAWSON, S. Production of Very Massive Higgs Bosons. Physics Letters B, v. 136, n. 3, p. 196, 1984.
- 14 DIJOURDI, A. The Anatomy of Electro–Weak Symmetry Breaking. Tome I: The Higgs boson in the Standard Model. Physics Reports, v. 457, n. 1, p. 1, 2005.
- 15 ALTARELLI, G. and MELE, B. and PITOLLI, F. Heavy Higgs Bosons Production at Future Colliders. Nuclear Physics B, v. 287, n. 1, p. 205, 1987.
- 16 Tanabashi, M *et al.* (Particle Data Group) Review of Particle Physics. Physics Review D, v. 98, n. 9, p. 030001, 2018.
- 17 FILIPPIS, N. Search for the Standard Model Higgs in $H \rightarrow ZZ \rightarrow 4l$ channel with the CMS experiment. Journal of Physics, USA, v. 455, n. 1, p. 012027, 2013.
- 18 CMS COLLABORATION. Properties of the Higgs-like boson in the decay $H \rightarrow ZZ \rightarrow 4l$ in pp colisions at $\sqrt{s} = 7$ e 8 TeV. CMS Physics Analysis Summary, n. 002, CERN, 2013.
- 19 DAS, A. Lectures on Quantum Field Theory. Singapore: World Scientific, 2008.
- 20 SEIDEN, A. Particle Physics: A Comprehensive Introduction. San Francisco: Addison Wesley, 2005.
- 21 ELLIS, R.K.; STIRLING W.J.; WEBBER, B.R. QCD and Collider Physics. Cambridge: Cambridge University Press, 2003.
- 22 BOZZI, G. *et al.* Vector Boson Pair Production via Vector Boson Fusion at NLO QCD. XVI Incontri di Fisica delle Alte Energie, Italy, v. 14, p. 69, 2008.
- 23 LANGACKER, P. The Standard Model and Beyond, New York: CRC Press, 2009.
- 24 GLASHOW, S. L. Partial Symmetries of Weak Interactions. Nuclear Physics, v. 22, p. 579, 1961.
- 25 WEINBERG, S. A Model of Leptons. Physics Review Letters, v. 19, p. 1264, 1967.
- 26 ENGLERT, F. and BROUT, R. Broken Symmetry and the Mass of Gauge Vector Mesons. Physics Review Letters, v. 13, p. 321, 1964.
- 27 CMS COLLABORATION. Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs. Physics Review Letters, v. 110, p. 081803, 2013.
- 28 CMS COLLABORATION. Measurement of the properties of a Higgs boson in the four-lepton final state. Physics Review D, v. 89, p. 092007, 2014.

- 29 CMS COLLABORATION. Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs. Physics Letters B, v. 736, p. 64, 2014.
- 30 CMS COLLABORATION. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. Physics Review D, v. 92, p. 012004, 2015.
- 31 CERN WEB SITE. Available at: <<http://home.web.cern.ch/about>>. Accessed on: 24/11/2017.
- 32 CERN. LHC Guide. CERN-Brochure-2008-001-Eng. Available at: <<http://cds.cern.ch/record/1092437/files/CERN-Brochure-2008-001-Eng.pdf>>. Accessed on: 24/11/2017.
- 33 BRUNING, O; COLLIER, P. Building a behemoth. Nature, United Kingdom, v. 448, p. 258, 2007.
- 34 CMS COLLABORATION. The CMS experiment at the CERN LHC. Journal of Instrumentation, Italy, v. 3, p. 362, 2008.
- 35 CMS. CMS Webpage. Available at: <<http://cms.web.cern.ch/>>. Accessed on: 20/11/2017.
- 36 ATLAS. ATLAS Webpage. Available at: <<http://atlas.ch/>>. Accessed on: 20/11/2017.
- 37 LHCb. LHCb Webpage. Available at: <<http://lhcb.web.cern.ch/lhcb/>>. Acessed in: 20/11/2017.
- 38 LHCf. LHCf Webpage. Available at: <<http://home.web.cern.ch/about/experiments/lhc>>. Accessed on: 20/11/2017.
- 39 TOTEM. TOTEM Webpage. Available at: <<http://totem.web.cern.ch/Totem/>>. Accessed on: 20/11/2017.
- 40 CMS COLLABORATION. CMS Technical Design Report: Hadron Calorimeter Technical Design Report. Available at: <[http://cds.cern.ch/record/357153/files/HCAL\\$_\\$TDR1997.pdf?version=1](http://cds.cern.ch/record/357153/files/HCAL$_$TDR1997.pdf?version=1)>. Accessed on: 20/11/2017.
- 41 CMS COLLABORATION. CMS Physics: Technical Design Report 1. Available at: <<http://cds.cern.ch/record/922757/files/lhcc-2006-001.pdf>>. Accessed on: 20/11/2017.
- 42 GRUPEN, C.; SHWARTZ, B. Particle Detectors. New York: Cambridge University Press, 2008.

- 43 CMS COLLABORATION. CMS Technical Design Report: The Muon Project. 1997. Available at: <<http://cds.cern.ch/record/343814/files/LHC-97-032.pdf>>. Accessed on: 20/11/2017.
- 44 CMS COLLABORATION. CMS The TriDAS Project Technical Design Report 2: Data Acquisition and High-Level Trigger. 2002. Available at: <<https://cdsweb.cern.ch/record/578006/files/cer-2336481.pdf>>. Accessed on: 20/11/2017.
- 45 PIGARD, P. Electron studies and search for vector boson scattering in events with four leptons and two jets with the CMS detector at the LHC. Thesis (Doctor of Philosophy) - Universite Paris-Saclay, France, 181 p., 2017. Available at: <<https://cds.cern.ch/record/2290139>>. Accessed on 19/08/2018.
- 46 CMS COLLABORATION. Limits on the Higgs boson lifetime and width from its decay to four charged leptons. Physics Review D, v. 92, p. 072010, 2015.
- 47 DITTMAIER, S. *et al.* LHC Higgs Cross Sections. CERN, 2011.
- 48 DITTMAIER, S. *et al.* LHC Higgs Cross Sections. CERN, 2012.
- 49 DITTMAIER, S. *et al.* LHC Higgs Cross Sections. CERN, 2013.
- 50 CMS COLLABORATION. Search for Invisible Decays of Higgs Bosons in the Vector Boson Fusion Production Mode. CMS-PAS-HIG-14-038. CERN, 2014.
- 51 ATLAS COLLABORATION. Search for an Invisibly Decaying Higgs Boson Produced via Vector Boson Fusion in pp Collisions at $\sqrt{s}=8\text{TeV}$ Using ATLAS Detector at the LHC. ATLAS-CONF-2015-004. CERN, 2015.
- 52 PLEHN, T. and RAINWATER, D. L. and ZEPPENFELD, D. Determining the Structure of Higgs Couplings at the CERN Large Hadron Collider. Physics Review Letters, v. 88, p. 051801, 2002.
- 53 SPIRA, M. QCD Effects in Higgs Physics. Fortsch Physics, v. 46, p. 203, 1998.
- 54 HAN, T. and VALENCIA, G. and WILLENCROCK, S. Structure-Function Approach to Vector-Boson Scattering in pp Collisions. Physics Review Letters, v. 69, p. 3274, 1992.
- 55 HARLANDER, R. V. and VOLLINGA, J. and WEBER, M. M. Gluon-induced weak boson fusion. Physics Review D, v. 77, p. 053010, 2008.
- 56 BOLZONI, P. *et al.* Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD. Physics Review Letters, v. 105, p. 011801, 2010.

- 57 ZEPPENFELD, D. and *et al.* Measuring Higgs boson couplings at the LHC. Physics Review D, v. 62, p. 013009, 2000.
- 58 ALIOLI, S. *et al.* NLO vector-boson production matched with showed in POWHEG. Journal of High Energy Physics, v. 07, p. 060, 2008.
- 59 NASON, P. A new method for combining NLO QCD with shower Monte Carlo algorithms. Journal of High Energy Physics, v. 11, p. 040, 2004.
- 60 FRIXIONE, S. and NASON, P. and OLEARI, C. Matching NLO QCD computations with parton shower simulations: the POWHEG method. Journal of High Energy Physics, v. 11, p. 070, 2007.
- 61 BAGNASHI, E. *et al.*, Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. Journal of High Energy Physics, v. 02, p. 088, 2012.
- 62 NASON, P. and OLEARI, C. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. Journal of High Energy Physics, v. 02, p. 037, 2010.
- 63 HARTANTO, H. B. *et al.*, Higgs boson production in association with top quarks in the POWHEG BOX. Physics Review D, v. 91, p. 094003, 2015.
- 64 LUISONI, G. *et al.*, $HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. Journal of High Energy Physics, v. 10, p. 083, 2013.
- 65 GAO, Y. *et al.* Spin determination of single-produced resonances at hadron colliders. Physics Review D, v. 81, p. 075022, 2010.
- 66 NNPDF COLLABORATION. Parton distributions for the LHC Run II. Journal of High Energy Physics, v. 04, p. 040, 2015.
- 67 NASON, P. and ZANDERIGHI, G. W^+W^- , WZ , ZZ production in the POWHEG-BOX-V2. European Physics Journal C, v. 74, p. 2702, 2014.
- 68 CAMPBELL, J. M. and ELLIS, R. K. MCFM for the Tevatron and the LHC. Nuclear Physics Proceedings Supplements, v. 205, p. 10, 2010.
- 69 CAMPBELL, J. M. and ELLIS, R. K. and WILLIAMS, C. Bounding the Higgs width at the LHC using full analytic results for $gg \rightarrow e^-e^+\mu^-\mu^+$. Journal of High Energy Physics, v. 04, p. 060, 2014.
- 70 GEANT4 COLLABORATION. GEANT4 a simulation toolkit. Nuclear Instruments and Methods A, v. 506, p. 250, 2003.

- 71 GEANT4 COLLABORATION. GEANT4 developments and applications. IEEE Transactions on Nuclear Science, v. 53, p. 270, 2006.
- 72 CMS COLLABORATION. Measurements of properties of the Higgs boson in the four-lepton final state at $\sqrt{s} = 13\text{TeV}$, CMS Analysis Notes, n. 442, CERN, 2016.
- 73 CMS COLLABORATION. New results on the study of Higgs boson production in the four-lepton final state at $\sqrt{s} = 13\text{TeV}$. CMS Analysis Notes, n. 217, CERN, 2016.
- 74 CMS COLLABORATION. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\text{ TeV}$. Journal of Instrumentation, v. 10, p. 06005, 2015.
- 75 CMS COLLABORATION. Measurement of the muon-tracking efficiency using the tag and probe method on 2015 RunII data. CMS Analysis Notes, n. 215, CERN, 2015.
- 76 CMS COLLABORATION. Measurement of the properties of the Higgs boson in the four-lepton final state at $\sqrt{s} = 13\text{ TeV}$. CMS Analysis Notes, n. 277, CERN, 2016.
- 77 CMS COLLABORATION. Search for Dark Matter Produced in Association with a Higgs Boson in the four lepton final state at 13 TeV. CMS Analysis Notes, n. 328, CERN, 2016.
- 78 CMS COLLABORATION. Measurement of the total and differential ZZ production cross sections, $Z \rightarrow ll'l'l'$ branching fraction and a TGCs in pp collisions at $\sqrt{s} = 13\text{ TeV}$. CMS Analysis Notes, n. 359, CERN, 2016.
- 79 HAYKIN, S. Neural Networks and Learning Machines. Pearson, 2009, ed. 3rd, isbn: 0-13-147139-2.
- 80 GURNEY, K. An Introduction to Neural Networks. UCL Press, 2004, isbn: 1-85728-503-4.
- 81 GLOROT, X. and BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. International Conference on Artificial Intelligence and Statistics, p. 249, 2010.
- 82 NAIR, V. and HINTON, G. E. Rectified linear units improve restricted boltzmann machines. International Conference on Machine Learning, p. 807, 2010.
- 83 ZEILER, M. D. *et al.* On rectified linear units for speech processing. IEEE International Conference on Acoustics, Speech and Signal Processing, p. 235, 2013.
- 84 CMS COLLABORATION. Updated results on the new boson discovered in the search for the standard model Higgs boson in the $H \rightarrow ZZ \rightarrow 4l$ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV . CMS Analysis Notes, n. 367, CERN, 2012.

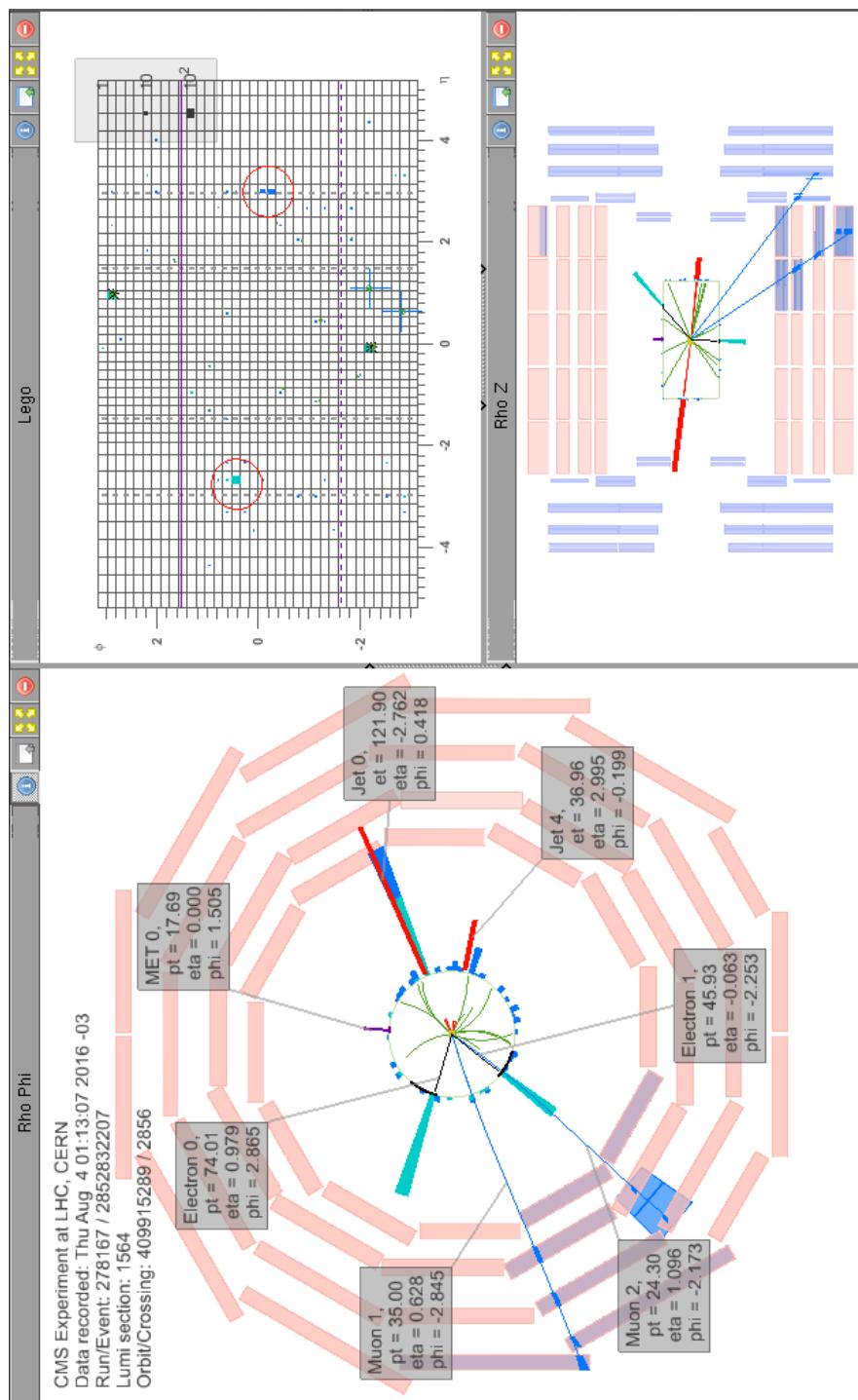
- 85 CMS COLLABORATION. Measurement of the production and decay of a Higgs boson in the four-lepton final state. CMS Analysis Notes, n. 108, CERN, 2013.
- 86 CMS COLLABORATION. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13\text{TeV}$. Journal of High Energy Physics, v. 11, p. 047, 2017.
- 87 CMS COLLABORATION. Search for the standard model Higgs boson in the decay channel $H \rightarrow ZZ \rightarrow 4l$ in pp collisions. CMS Analysis Note, n. 141, CERN, 2012.
- 88 CMS COLLABORATION. Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s} = 13\text{ TeV}$. CMS Physics Analysis Summary, n. 022, CERN, 2016.
- 89 ATLAS COLLABORATION. Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector. ATLAS Conference Notes, n. 041, CERN, 2017.
- 90 KERAS WEB PAGE. Available at: <https://keras.io/> Accessed on: 12/01/2016.
- 91 LHC HIGGS CROSS SECTION WORKING GROUP. Available at: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV2014#s_13_0_TeV. Accessed on: 29/11/2018.
- 92 BAGOLY, A. *et al.* Machine Learning Developments in ROOT. Journal of Physics Conference Series, v. 898, p. 072046. 2017.
- 93 CAMPBELL, J. M. and ELLIS, R. K. MCFM for the Tevatron and the LHC. Nuclear Physics Proceeding Supplements, v. 10, p. 205, 2010.
- 94 BONVINI, M. *et al.* Signal-background interference effects in $gg \rightarrow H \rightarrow WW$ beyond leading order. Physical Review D, v. 88, p. 034032, 2013.
- 95 MELNIKOV, K. and DOWLING, M. Production of two Z-bosons in gluon fusion in the heavy top quark approximation. Physical Letters B, v. 744, p. 43, 2015.
- 96 LI, C. S. *et al.* Soft gluon resummation in the signal-background interference process of $gg(\rightarrow h*) \rightarrow ZZ$. Journal of High Energy Physics, v. 1508, p. 065, 2015.
- 97 CATANI, S. and GRAZZINI, M. An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. Physics Review Letters, v. 98, p. 222002, 2007.
- 98 GRAZZINI, M. NNLO predictions for the Higgs boson signal in the $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ decay channels. Journal of High Energy Physics, v. 02, p. 043, 2008.

- 99 GRAZZINI, M. and SARGSYAN, H. Heavy-quark mass effects in Higgs boson production at the LHC. *Journal of High Energy Physics*, v. 09, p. 129, 2013.
- 100 Higgs Combine Tool Web Page. Available at: [@cms-hcomb](https://legacy.gitbook.com/@cms-hcomb). Accessed on: 08/06/2016.
- 101 MONETA, L. *et al.* The RooStats Project. *Proceeding of Science*, v. ACAT2010, p. 057, 2010.
- 102 HIGGS COMBINE TOOL WORKSHOP. Fermilab (FNAL). 2018. Available at: <https://indico.cern.ch/event/747340/timetable/>. Accessed on: 27/11/2018.
- 103 Wilk's Theorem; A Likelihood Ratio Test for Nested Composite Hypothesis. Massachusetts Institute of Technology (MIT), USA. 2018. Available at: <https://math.mit.edu/~rmd/650/wilks.pdf>. Accessed on: 01/11/2018.
- 104 BARTLETT, P. Theoretical Statistics: Lecture 26. Berkeley, University of California, USA. 2018. Available at: <https://www.stat.berkeley.edu/~bartlett/courses/2013spring-stat210b/notes/26notes.pdf>. Accessed on 01/11/2018.
- 105 GLEN, C. Statistical Data Analysis. Clarendon Press, 1998, isbn: 0198501552.
- 106 EILAM, G. *et al.* Search and Discovery Statistics in HEP: Lecture 1. Electroweak Symmetry Break Spring School (EWSB), Italy. 2018. Available at: https://indico.cern.ch/event/673580/contributions/2790011/attachments/1638839/2615856/Statistics_CERN_2018_part1.pdf. Accessed on: 05/11/2018.
- 107 EILAM, G. *et al.* Search and Discovery Statistics in HEP: Lecture 2. Electroweak Symmetry Break Spring School (EWSB), Italy. 2018. Available at: https://indico.cern.ch/event/673580/contributions/2790011/attachments/1638839/2615856/Statistics_CERN_2018_part1.pdf. Accessed on: 05/11/2018.
- 108 HAHN, K. *et al.* L1 Track Finding for the CMS HL-LHC Upgrade. 2017. Available at: <https://indico.cern.ch/event/627245/contributions/2677000/attachments/1524022/2382144/VERTEX2017-CMS-L1Tk-KristianHahn.pdf>, Accessed on: 01/11/2018.
- 109 FNAL-FSU-TAMU CMSL1TT WORKING GROUPS. SLHCL1TrackTriggerSimulations. 2016. Available at: <https://github.com/jiafulow/SLHCL1TrackTriggerSimulations/wiki>. Accessed on: 27/11/2017.
- 110 CMS COLLABORATION. The Phase-2 Upgrade of the CMS Tracker. CMS Technical Design Report, CERN, 2017.
- 111 BAULIEU, G. *et al.* Emulation of a track reconstruction system based on associative memories. Detector Notes, CERN, 2015.

- 112 XILINIX COMPANY. Field Programmable Gate Array (FPGA). Available at: <<https://www.xilinx.com/products/silicon-devices/fpga/what-is-an-fpga.html>>. Accessed on: 27/01/2017.
- 113 ALMEIDA, M. M.; JINDARIANI, S.; RISTORI, L. Synthetic efficiency study. Technical Simulation Report. Available at: <indico.cern.ch/event/485680/contributions/2004800/>. Accessed on: 01/11/2018.
- 114 ALMEIDA, M. M.; JINDARIANI, S.; RISTORI, L. Performance of simple duplicate removal strategy. Technical Simulation Report. Available at: indico.cern.ch/event/520985/contributions/2136443/. Accessed on: 01/11/2018.
- 115 ALMEIDA, M. M.; JINDARIANI, S.; RISTORI, L. Update on synthetic efficiency study. Technical Simulation Report. Available at: <indico.cern.ch/event/522593/contributions/2140692/>. Accessed on: 01/11/2018.
- 116 ALMEIDA, M. M.; JINDARIANI, S.; RISTORI, L. Synthetic Efficiency and Duplicate Removal. Technical Simulation Report. Available at: <indico.cern.ch/event/524621/contributions/2148537/>. Accessed on: 01/11/2018.
- 117 ALMEIDA, M. M.; JINDARIANI, S.; RISTORI, L. Using Bend Information in the AM - Performance Studies. Technical Simulation Report. Available at: <indico.cern.ch/event/660323/contributions/2693623/>. Accessed on: 01/11/2018.
- 118 SZNAJDER, A.; ALMEIDA, M. M. A Fast Matrix Element Method (FastME). XXVII International Symposium on Lepton Photon Interactions at High Energies. 2015. Available at: <<https://pos.sissa.it/245/057/pdf>> Accessed on: 10/08/2016.
- 119 HOECKER, A. *et. al.* TMVA4 Toolkit for Multivariate Data Analysis with ROOT (Users Guide). 2017. Available at: <<http://tmva.sourceforge.net/docu/TMVAUUsersGuide.pdf>>. Accessed on: 30/03/2017.
- 120 ALMEIDA, M. M.; SZNAJDER, A. Fast Matrix Element Package (FastME). 2016. Available at: <<https://github.com/mmelodea/FastMatrixElement>>. Accessed on: 27/11/2017.
- 121 BELUFFI, C. The Matrix Element Method within CMS. 2014. Available at: <<https://indico.cern.ch/event/258092/contributions/1588574/attachments/454217/629628/ACAT14.pdf>>. Accessed on: 27/11/2017.

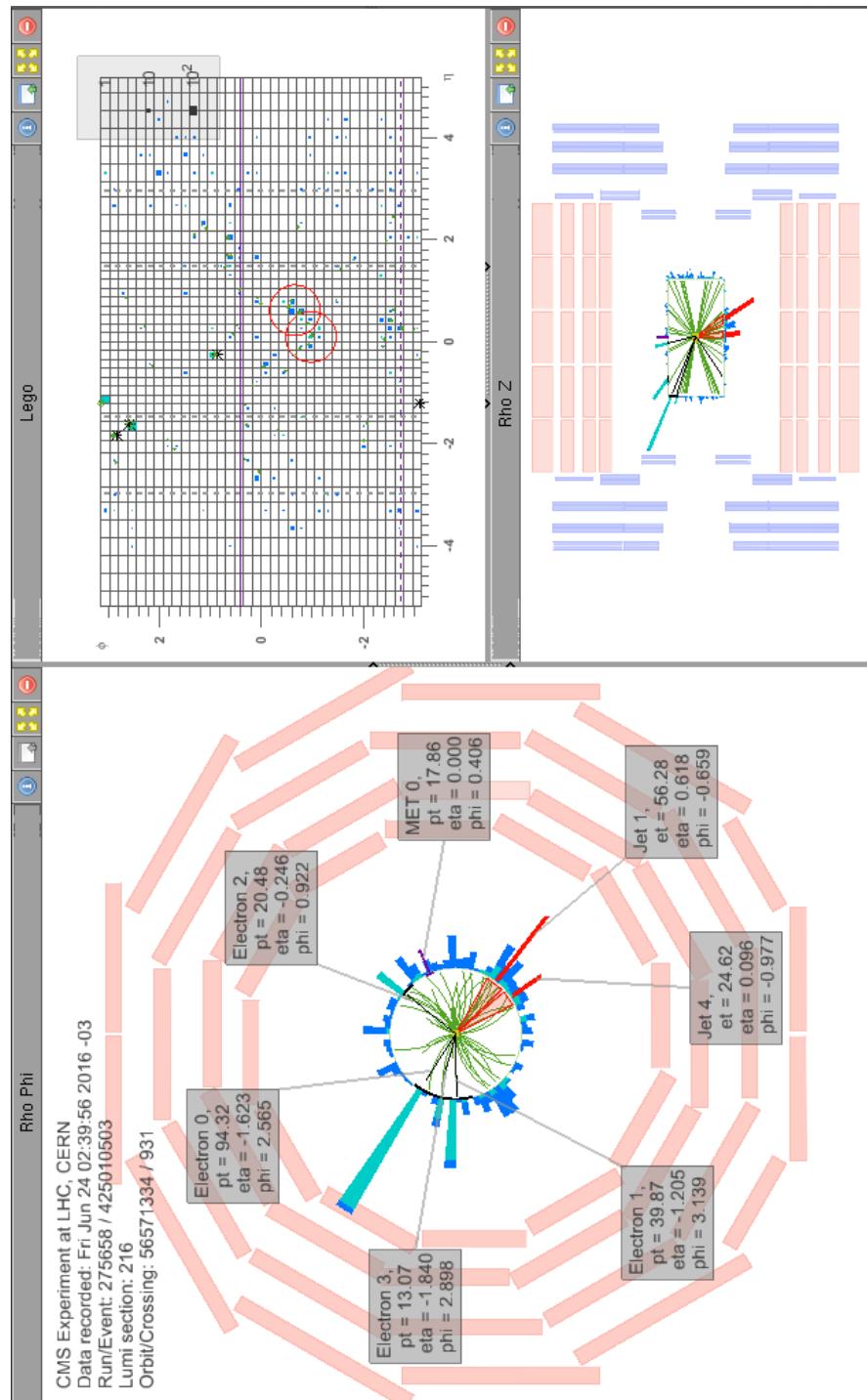
A DISPLAY OF VBF AND BACKGROUND-LIKE EVENTS

Figure 49 - Event display of the most VBF-like event ($H \rightarrow ZZ \rightarrow 2e2\mu$) according to the ANN in the **Njets2** category (ANN score = 0.93). As expected, the two jets present in the event have a very large η separation and no particle activity between them.



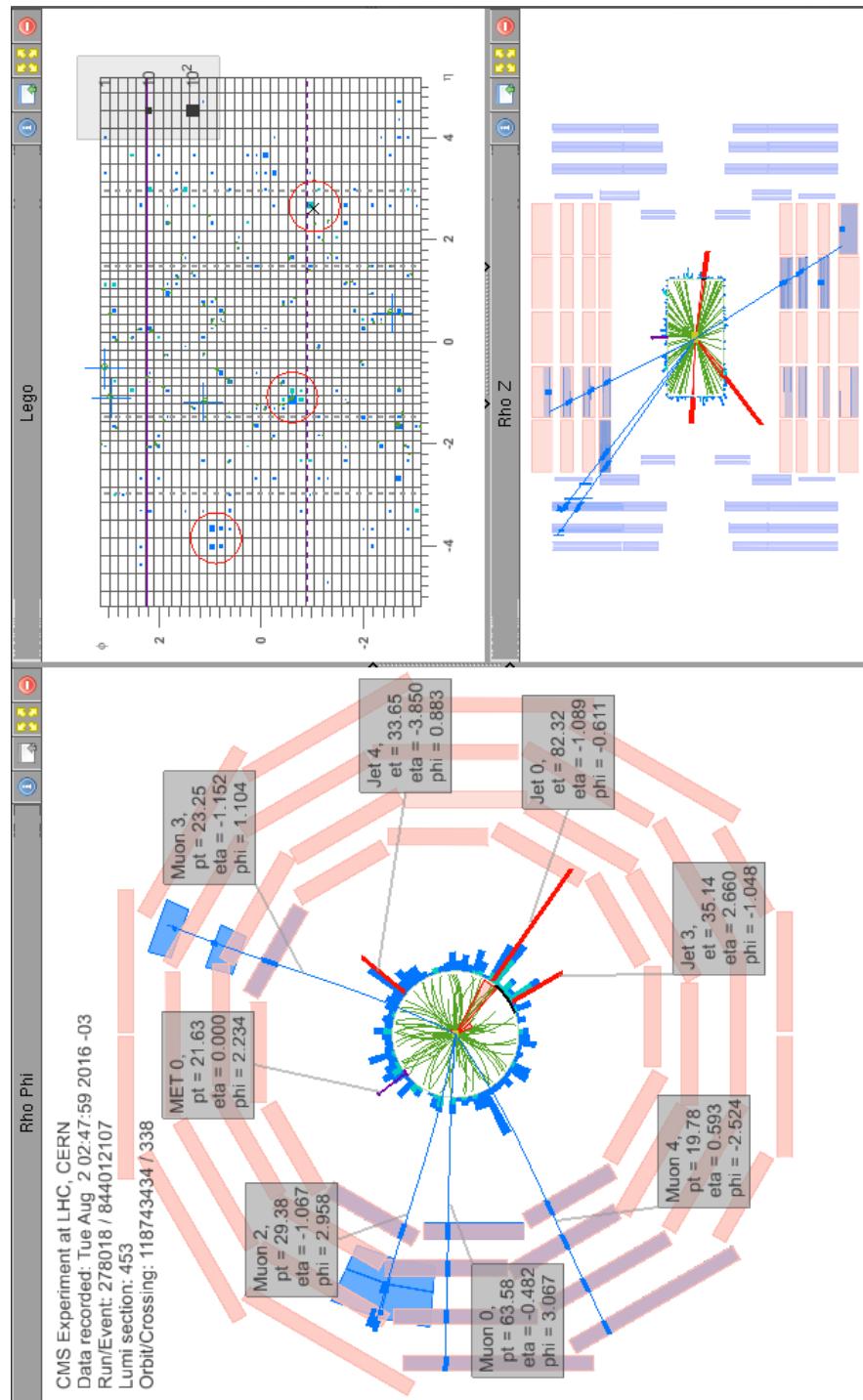
Source: The AUTHOR, 2018.

Figure 50 - Event display of the most background-like event ($H \rightarrow ZZ \rightarrow 4e$) according to the ANN in the **Njets2** category (ANN score = 0.04). Differing from the most VBF-like event, this one has the two jets very close in the $\eta - \phi$ space.



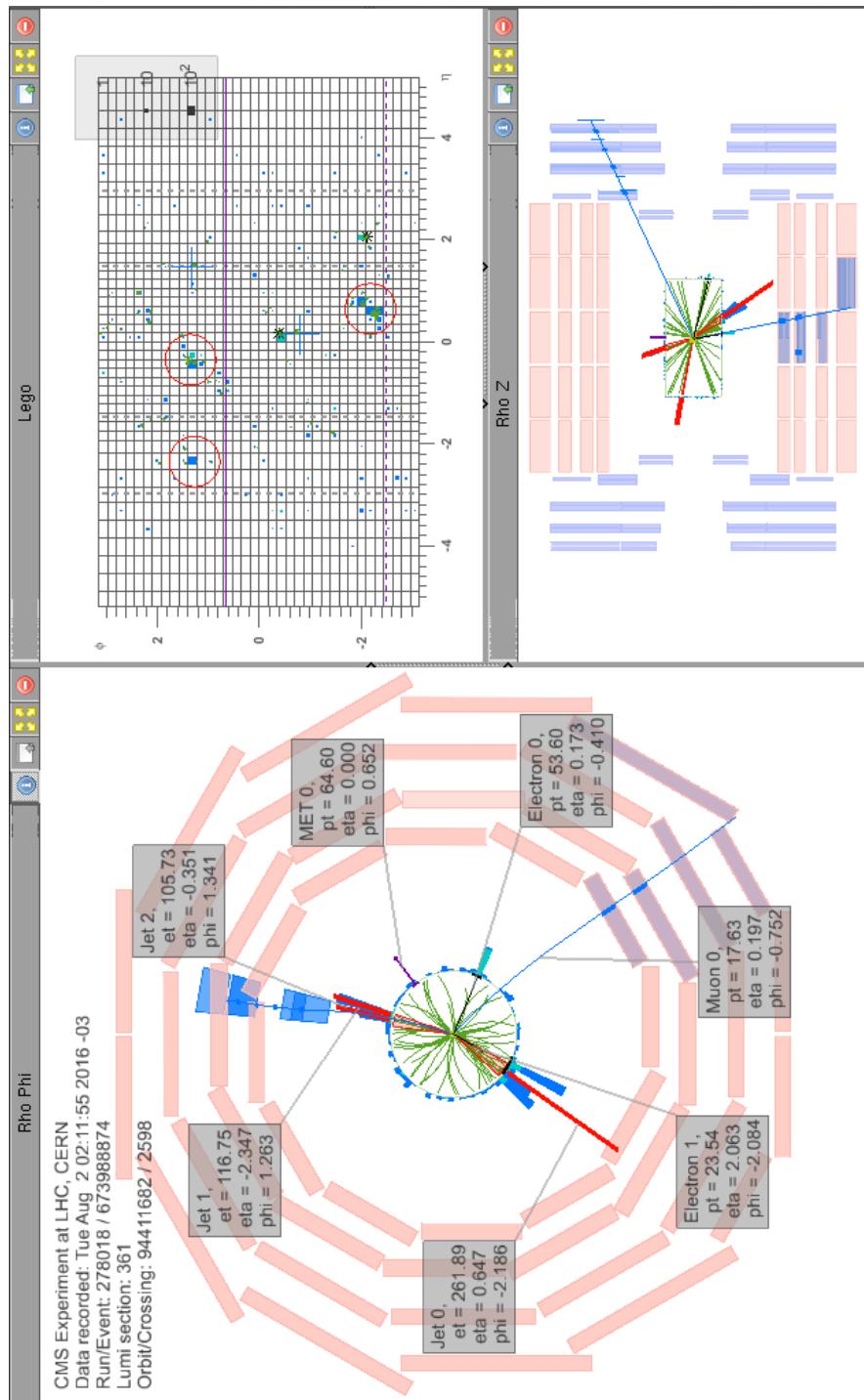
Source: The AUTHOR, 2018.

Figure 51 - Event display of the most VBF-like event ($H \rightarrow ZZ \rightarrow 4\mu$) according to the ANN in the **Njets3** category (ANN score = 0.77). By comparison, the η gap between the jets and the particle activity is more evident than what is observed in the **Njets3** category for background (see next figure).



Source: The AUTHOR, 2018.

Figure 52 - Event display of the most background-like event ($H \rightarrow ZZ \rightarrow 2e2\mu$) according to the ANN in the Njets3 category (ANN score = 0.15).

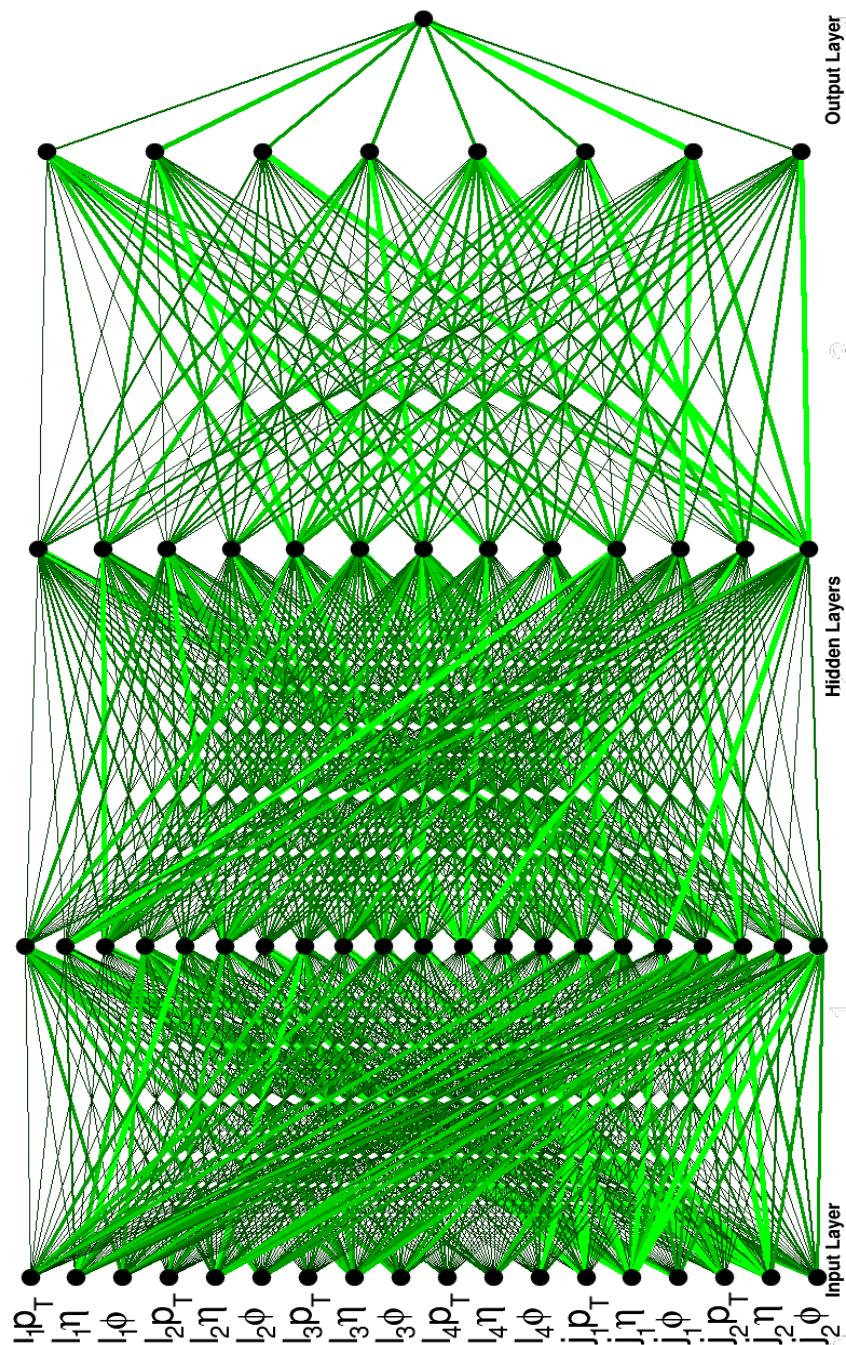


Source: The AUTHOR, 2018.

B FINAL ANN ARCHITECTURES

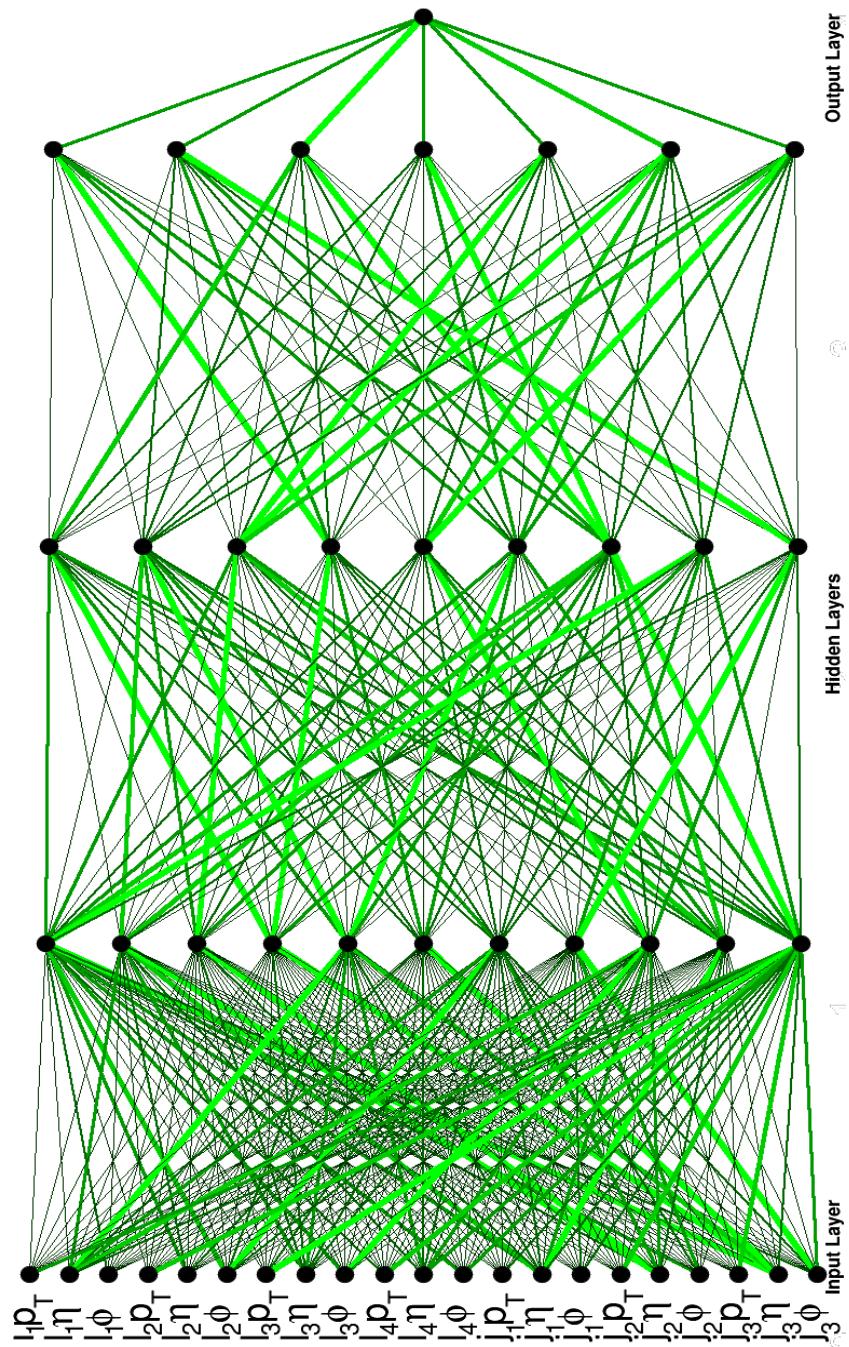
Figure 53 - The architecture of the neural network created in this analysis for **Njets2** category.

The black dots represent the inputs and the neurons, while the lines represent the size of the parameters found after training those NNs. Wider and brighter lines means that the parameter (weight or bias) associated to a given input for a neuron is larger (in other words its contribution is more relevant). This ANN has 21:13:8 hidden neurons of SeLU type.



Source: The AUTHOR, 2018.

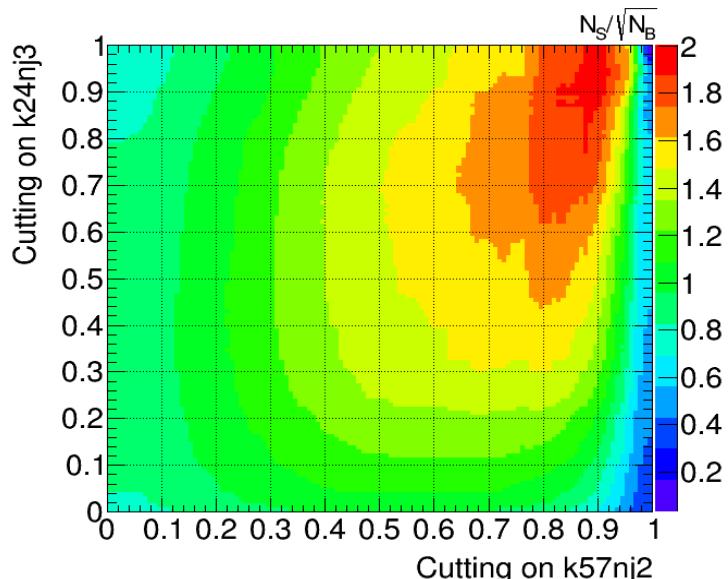
Figure 54 - The architecture of the neural network created in this analysis for **Njets3** category.
This ANN has 11:9:7 hidden neurons of SeLU type.



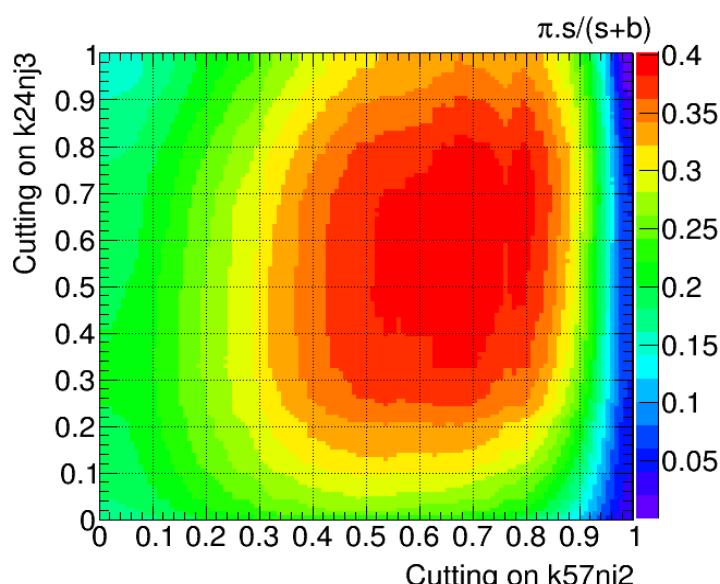
Source: The AUTHOR, 2018.

C ANNS EXPECTED S/\sqrt{B} AND $\pi \cdot S/(S+B)$

Figure 55 - The significance (computed as S/\sqrt{B}) (a) and the efficiency times purity ($\pi \cdot \epsilon$) (b) achieved when cutting at the NN for each jet-based category. The cut applied corresponds to the low edge value of a given bin in the histogram. Note that the significance computed here does not take into account the uncertainties.



(a)



(b)

Source: The AUTHOR, 2018.

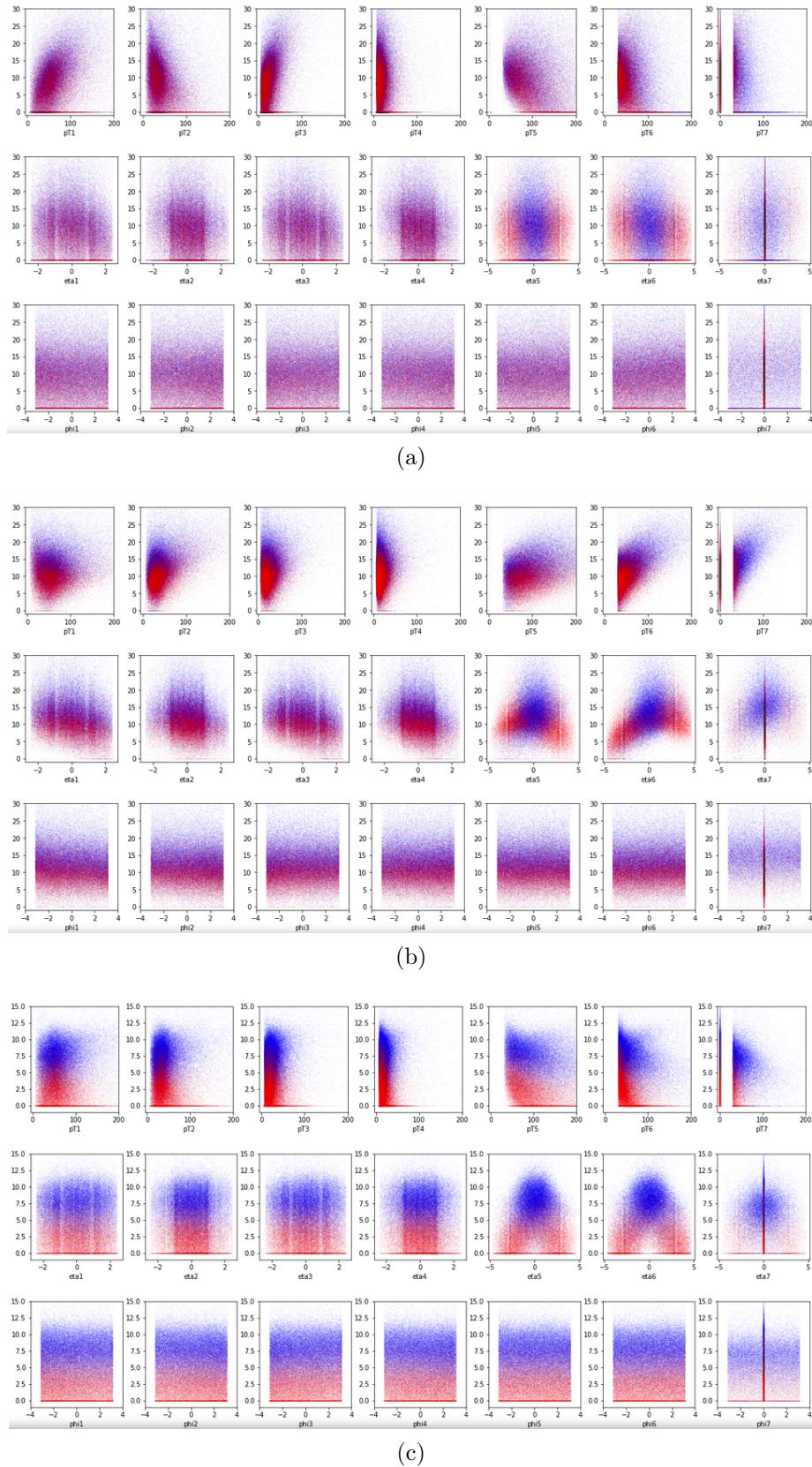
D THE ANN INTERNAL REPRESENTATION SPACE

In this appendix, the author would like to call the attention for the interesting internal representation created by a NN. The internal representation here could be replaced by the question: "what/how do the NN neurons are firing for". The inspiration for the discussion presented here comes from the well known behavior of the Convolutional Neural Networks (CNNs). The CNNs are special because the filters created by the convolution procedure during the network training can describe objects which humans are familiar with. For instance, in face recognition the filters can (and usually do) learn properties of human face, as the nose, the eyes, the mouth and so on.

In the case of the Fully Connected NNs used in this analysis, it is not so obvious how one could extract similar features from them and in fact, there seems not to exist a work about such topic in the literature. But, motivated by how the features are extracted from the filters of a CNN, the author looked to the response coming from the neurons in different layers when the inputs are given to the NN. Then, superimposing the behavior of such neurons when responding to signal and background inputs returned some pictures like the ones showed in Fig. 56. Note that, those pictures are an example since the NNs used have three layers and several neurons. Taking the NN for the **Njets2** category, for instance, it has the topology 21:15:8 and thus, (for six particles and with p_T , η , ϕ as inputs) it generates more than 500 pictures. Even so, it is possible to draw some observations from them and Fig. 56 represents well what was seen for the NN in the other neurons in each layer.

The first observation that can be quickly derived by looking through the plots in Fig. 56(a), (b) and (c) is that the separation between signal (in red) and background (in blue) increases gradually across the hidden layers. This is more evident for the leptons since they do not produce discrimination between the VBF process and the other SM Higgs production modes. While that, the jets (the two highest p_T at least) have a good separation in η already in the first hidden layer. This is not surprising since we expect such behavior in η for the VBF jets. It's also possible to notice that the patterns of discrimination across the hidden layers also gets more complex as the pictures changes their shapes. The most surprising observation (at the author point of view) is the separation in ϕ , which has a exactly same distribution (roughly a uniform distribution between $-\pi$ and $+\pi$) for signal and backgrounds. Note that the plots in Fig. 56(c) are not coming from the output layer, those are the responses of one of the neurons in the last hidden layer before the output neuron, which only give responses in the range [0, 1] (which are the labels for the two classes, background and VBF, respectively).

Figure 56 - The behavior of the NN neuron output for the (a) first hidden layer, (b) second hidden layer and (c) output layer. The subplots in (a), (b) and (c) show, from left to right and sorted, the p_T 's (top row), the η 's (middle row) and the ϕ 's (bottom row) of the four leptons and the three jets.



E THE CMS L1 TRACKING TRIGGER (CMSL1TT) PROJECT

The CMS Level 1 tracking trigger is a project proposed for the upgrade of the CMS detector during the third stop period of the LHC operation, which is planned to occur around 2020 and goes until around 2025, when should be started the era of the High Luminosity LHC (HL-LHC) Fig. 57.

In the scenario of the HL-LHC the expected instantaneous luminosity to be provided will be $\sim 10^{34} \text{cm}^2 \text{s}^{-1}$ (integrated luminosity of 3000fb^{-1}) with collisions at the nominal energy at which the LHC was projected, 14TeV. The occurrence of PU has been estimated to be of about 140/200 in average per event [108]. Which is a critic condition for the detector operation and for selection of the primary vertices. Since there were some larger PU occurrence (≥ 50 PUs) than the expected one (≥ 30 PUs), the simulation studies also take into account possible extreme scenarios of 250, 300 and 400 PUs.

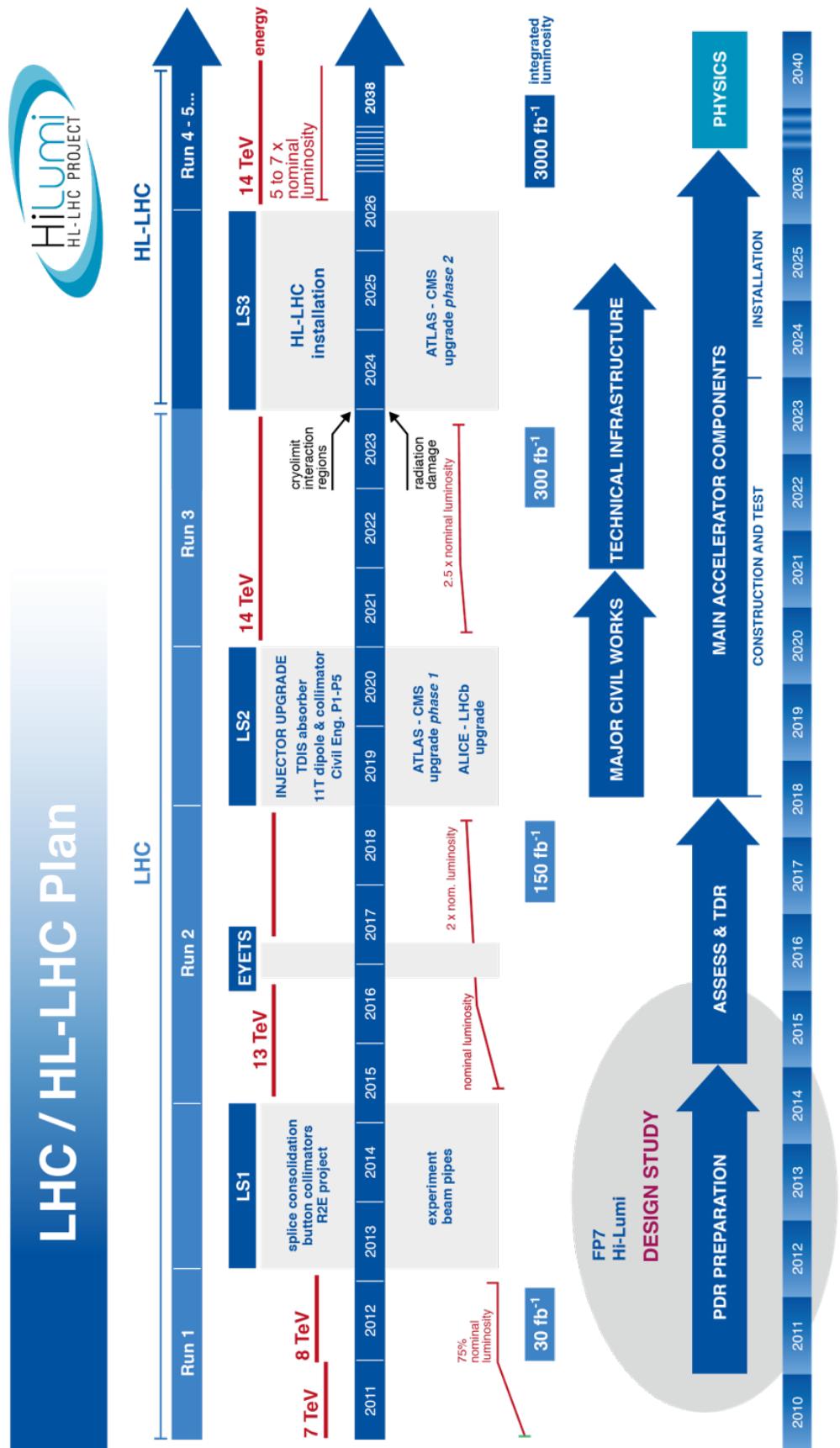
The main idea behind the CMSL1TT project is based on the hardware configuration for the phase II upgrade (LS3 in Fig. 57). The idea is to include the tracker the information coming from the CMS tracker into the L1 trigger. However, there some significant challenges for such approach. A reasonable estimation of the amount of data given by the tracker is about 600 Gb/s [108]. Additionally, the decision returned by the L1 trigger must still be fast ($5\mu\text{s}$).

In order to enhance the filtering of PU contribution the hardware of the tracker after phase II will be enhanced. Each tracker layer will be composed of two silicon sensors, back-to-back and interconnected, such that, every time a particle crosses the layer it will produce an object called as *stub*. The stub is a kind of vector which links the centroids of sensor (or cluster of them) in each of the detector layer. Additionally, the distance between those centroids (SW in Eq. 73 and ΔS in Fig. 58), taken orthogonally to the sensors, can be correlated to the p_T o the crossing particle. The relation is usually expressed as

$$p_T = 0.57 \frac{q.r.\epsilon}{SW.p} \sqrt{1 + \frac{\sin\theta_0}{\cos(\theta_0 - \alpha)}} \quad (73)$$

where, q stands for the charge of the particle, r stands for the average distance of the sensors to the detector coordinates origin (\mathbf{xy}), ϵ stands for the distance between the sensors back-to-back, θ_0 is the angle between the particle trajectory and the beam axis and, p is the strip pitch in millimeters [110]. The Fig. 58 shows schematically the idea of the stub. With this approach, it is easy to apply threshold in the p_T of the detected particles and remove in a fast way PU particles, which have in average $p_T < 2\text{GeV}$. This is the core working principle of the CMSL1TT project.

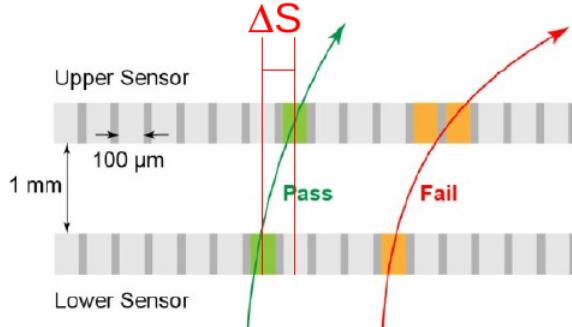
Figure 57 - The calendar for the upgrades of the LHC and its detectors.



Source: (<https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>).

Figure 58 - Scheme of the sensors in a layer of the CMS detector after the phase II upgrade.

An object called *stub* is defined in terms of the distance between the centroids of back-to-back sensors (or cluster of them) activated when a particle cross the layer.



Source: Adapted from [111].

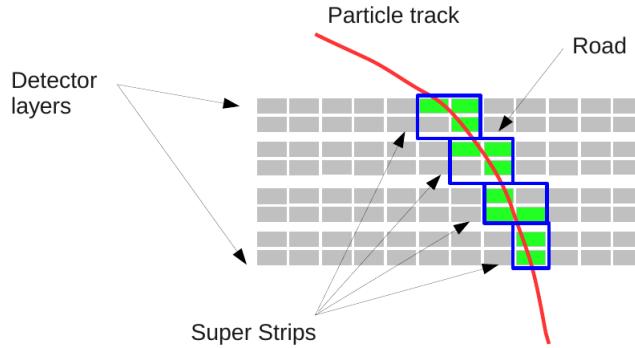
E.1 The Three CMSL1TT Approaches

At the time the author was cooperating with Fermilab, there were three groups carrying different approaches on how to deal with the track information extraction within the CMSL1TT. These three approaches were: the Associative Memory (AM) + FPGA, the FPGA-based Hough Transform and the FPGA-based Tracklet. As it is clear from their names, the three of the approaches uses FPGAs¹¹.

The three approaches treat the CMS detector in similar way. In order to properly process the signals coming from the CMS, the detector is divided into small regions called *trigger towers*. Each of these trigger towers are handled separately but using the same idea of the give approach. This first step is usually called *data formatting*. Then, the second step is to find coarse patterns across the hits left by particles in the detector. This step is usually called *pattern recognition* and is here where the three approaches start to go over different paths. The name of the approaches also explains this step for each of them. The AM+FPGA is the one developed by the Fermilab group and the one the author got involved. So, it will be discussed in more details in the next sections. More details about all these approaches can be found in [110].

¹¹ Field Programmable Gate Array (FPGA) is a semiconductor device (a special chip), which can be programmable with routines at hardware level (and not software level) and allows one to create several logic circuits inside just one chip. Different from usual chips, usually called ASIC (Application Specific Integrated Circuits) which are fabricated for specific purposes, the FPGAs can be used for any application. More details see [112].

Figure 59 - Scheme of the partition adopted in the CMSL1TT AM+FPGA for the cluster of strips activated in the silicon sensors in the detector layers.



Source: The AUTHOR, 2016.

E.2 The AM+FPGA CMSL1TT Approach

The AM+FPGA approach makes the pattern recognition by using the so called Associative Memory. This method runs a parallel matching of the hits found in the detector and the hits found in simulated patterns. These patterns are the so called *roads*, which are coarse track patterns. These roads are composed by what have been called in this approach as *super-strips*, which are the cluster of activated silicon strips in detector layers. Fig. 59 illustrates these nomenclatures. Once the number of fired super-strips is achieved for a given pattern, that pattern is triggered and the hits observed in the detector are saved for a most robust fit in order to extract the adequate track properties.

After the pattern recognition step, the triggered hits coming from the detector are given to a module called *combination builder*, which builds up all the possible tracks by combining those hits. The combinations are then given to another module, called *track fitter*, which performs a fit based on a linearized χ^2 . In this step one gets the tracks with good resolution parameters. To enhance the selection of tracks and random combination that still looks like a good track, it was developed the procedure called *duplicate removal* (which will be discussed in the following sections). Finally, the tracks that pass the duplicate removal steps are the ones that should be give ot the L1 trigger.

In the following sections the work developed by the author inside the CMSL1TT AM+FPGA approach is presented. For full details, refer to the author presentations ([113–117]) about these studies.

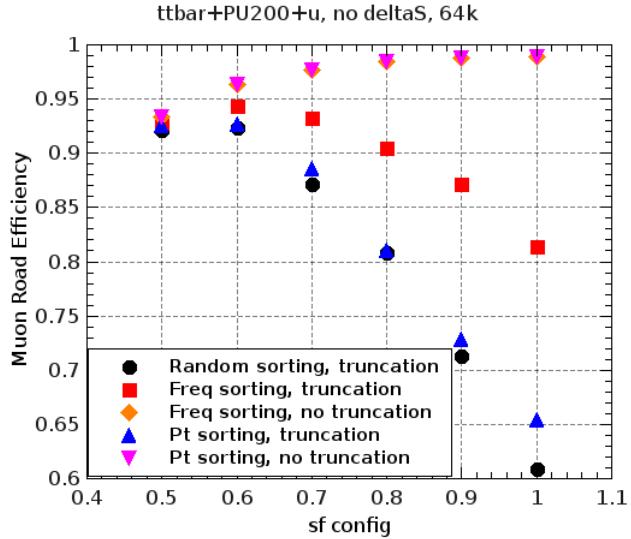
E.2.1 The Simulation Studies for the CMSL1TT AM+FPGA

In the simulation side, a couple of studies and implementations have been done through the AM simulation package maintained by Fermilab, Florida and TAMU working groups [109]. In the next sections, each study will be described in separated sections and full relevant details about them will be shown. Such studies comprehend synthetic efficiency, duplicate removal, stub bending, roads and combinations truncation, road sorting effect under truncation and track fitter χ^2 . Synthetic efficiency is designed to quantify track reconstruction efficiency based on track parameters, which can be linked to the analytic efficiency (based on stubs) as will be presented here. Once the AM + FPGA approach produces replicated tracks a duplicate removal (DR) method is needed in order to mitigate those redundant tracks. The stub bending information (ΔS) was used to reduce the number of roads and combinations coming out from AM in order to reduce processing time latency and reduce truncation effects (being very important for busy environment cases such $\mu + PU200$). Good results were achieved using DR and ΔS , keeping track reconstruction efficiency in an acceptable level. Also, when truncation is applied the road ordering has significant impact on the pattern match efficiency. The results presented here were obtained over different event samples which were: $(\mu/\pi/e) + PU(140,200,300,400)$, $\nu + PU(140, 200, 250)$, $t\bar{t} + PU200$ and jet($p_T = 250 GeV$) + PU200. Jet sample had 2k events and all the others had 10k.

E.2.1.1 Roads and Combinations Truncation

Some of the results presented in these notes concern to simulations in which truncation on roads and combinations was applied. The standard truncation limits were 200 roads and 500 combinations. In order to make the truncation easily modifiable in case of need, one extra flag was implemented in the AM package. The flag for road truncation is `--maxRoads` (implemented on **PatternMatcher.cc** and **TrackFitter.cc** modules) and it can act either during the pattern recognition (pattern match) or tracking fitting stage. The combination truncation were previously implemented to truncate combinations per road but, the adequate one is a truncation in combinations per event. For that, the flag `--maxCombsPreCB` was implemented. It works only in the track fitting stage and it essentially counts how many simple combinations are being generated in a given event. Simple combination means that, for each road the number of combinations is the product of the number of stubs per layer. Note that, the advanced combination builder builds extra combinations from 6/6 and it was decided to not account them for the truncation. By default, no road or combination truncation is active in the package (user must specify the truncation limits in the command line).

Figure 60 - Road sorting effect on efficiency without truncation and with truncation in 200 roads and 500 combinations for $t\bar{t} + PU200$.

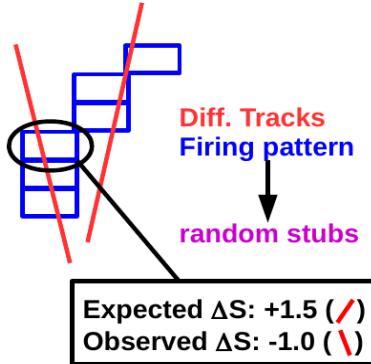


Source: The AUTHOR, 2016.

E.2.1.2 Road Sorting Effects under Truncation

Truncation is a latency holder that needs to be carried out carefully. When truncation is applied it might throw away things it should not. One thing that can affect that is how the list of roads and combinations are handled in AM processing stages. A study was made to verify how the road sorting affect the road efficiency when truncation is applied. Fig. 60 shows that for three road sorting ways in $t\bar{t} + PU200$ events: based on pattern p_T , based on pattern frequency and randomly sorted. The random sort is not straightforward to interpret since it depends on how the randomization sits the roads in the list. In order to get a more precise result it was done a sampling for each sf config. The points in Fig. 60 for random sorting are the average from those runs. No big variation was seen (max of 0.003 for sf0.8). It was pointed that sorting the roads by frequency is the better way and that is confirmed by the plot. In the AM package the roads come out from the pattern match stage naturally already sorted by frequency, so no need to implement something new for that. It was also decided to remove any sorting after the pattern recognition stage (since that is not straightforward to have on hardware). Then, no track sorting by decreasing p_T after the tracking fit stage was used. Also, in DR the track sorting based on increasing track fitter χ^2 was removed.

Figure 61 - How the ΔS mitigate AM patterns fired by random tracks: two random tracks trigger a valid pattern and the stubs are stored for further analysis but, the stubs combination doesn't come from a real track.



Source: The AUTHOR, 2016.

E.2.1.3 Stub Bending (ΔS)

The stub bending is one of the core concepts of CMS L1TT being the main variable projected to allow fast cut on low p_T tracks such as the ones coming from pile up. Stub bending is a measurement of the distance between cluster centroids of correlated silicon sensors in each of the detector layers. That distance is measured in terms of half silicon strip. A study was developed in order to investigate the usage of stub bending information as a way to avoid random pattern firing. That is, without ΔS an AM pattern is fired by random tracks that can cross the layers in very different angles. Looking at the ΔS 's of combinations produced by such random tracks one can see that they are incompatible as sketched in Fig. 61. Now, if the stub bending can be used to build up AM pattern banks (such as with ϕ or z segmentation) than those random tracks will not fire a pattern anymore.

The ΔS encoding to create AM pattern banks was done by dividing the stub bending distribution in ranges. Each range then is identified by an integer index which one is used to define the superstrips ID (ss ID). Once the standard configuration adopted at that time was *sf1_nz1*, which means no z segmentation, the same formula to build the ss ID on that case was used, with the change of the z segment index by the stub bending index. The formula, present on module **SuperstripArbiter.cc**, is the following

$$ss = i_{\Delta S} * N_{\phi} + i_{\phi} \quad (74)$$

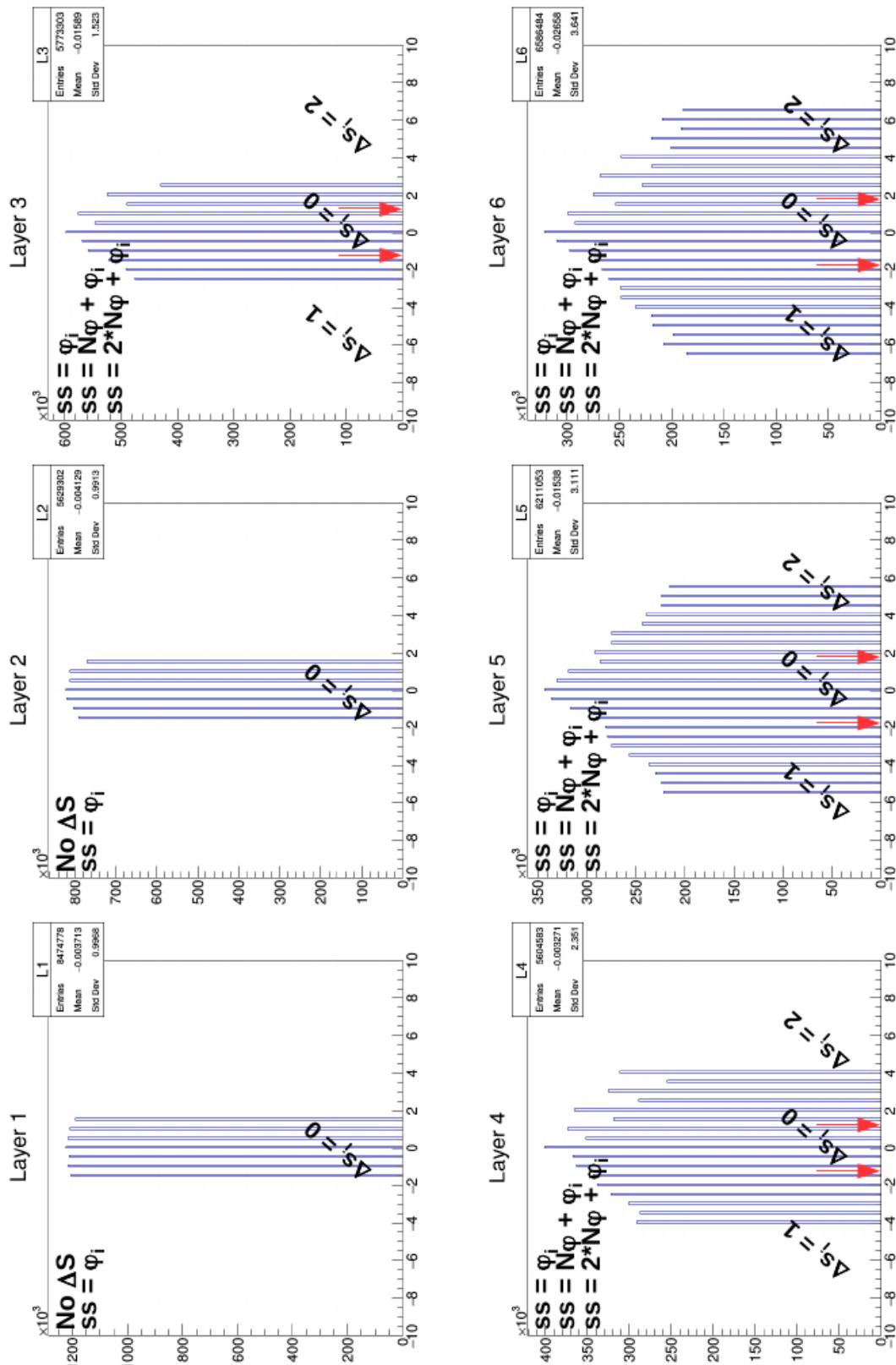
in which $i_{\Delta S}$ is the stub bending index associated to the range from which a stub belongs, N_{ϕ} is the number of ϕ segmentations (which depends on the stub fountain scale

factor configuration - sf) and i_ϕ is the ϕ segment index associated to the stub. As one can see in Fig. 62 the ΔS distributions in all layers are centered in zero. The stub bending index is defined from there, that is, the region close to $\Delta S = 0$ is the central range. Then, one can divide the stub bending distribution in many different range sizes. Also, one can try to use ranges with different widths.

On those studies two approaches were developed: one called asymmetric (ASYM) and one called symmetric (SYM). They have completely different structures but show similar performance. In the ASYM the ΔS distribution is divided always in three ranges and only the central range can have the width adjusted. The SYM was meant to produce all ranges with equal width, being possible to specify how many ranges should be produced. In the AM package that is controlled by two flags: `--deltaSM` that receives a string identifying the approach to be used (ASYM or SYM) and `--deltaS` that receives an integer number of 6 digits that specifies the configuration in each layer. So, for instance, ASYM115577 means that in the first two inner detector layers ΔS is not divided (that is, the standard way, no stub bending usage), in the two middle layers a central range with 5 ΔS values is created and in the two outer layers a central range with 7 values. For a configuration like SYM115577, the two inner layer again will not receive any operation, then the two middle layer will contain 5 ranges and the two outer layers will have 7 ranges. In order to calculate how to divide those distributions it was considered the maximum ΔS absolute value observed. It was found to be the same for $t\bar{t} + PU200$ and $jet(p_T = 250GeV) + PU200$ (distribution on Fig. 64), and is equal to 6.5. The division then was previously computed, producing the Tab. 25 which was encoded on **SuperstripArbiter.cc** module. An important note here: the flags specifying ΔS configuration must be used when generating pattern bank and when doing the pattern recognition step. In tracking fit stage it doesn't play any function since the ss are not used but instead the stubs properties. Also, the maximum value allowed to specify the range width or the number of ranges is 9. Additionally, the numbers must be always odd (since there's the $\Delta S = 0$ value). Any other attempt will produce an error message showing the allowed configurations.

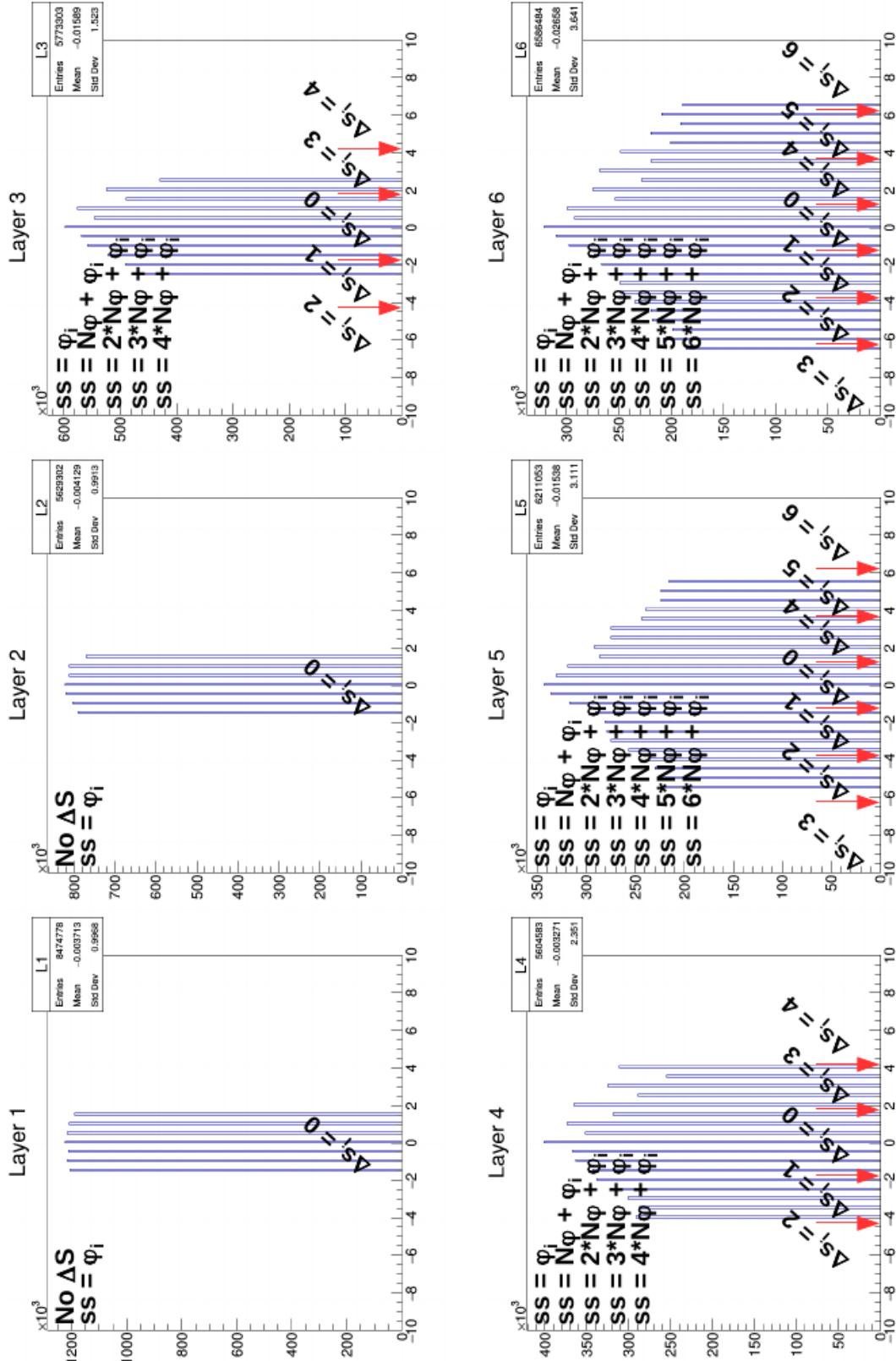
Using such implementation many configurations were studied for *sf1_nz1*. Fig. 65 shows all tested configurations in comparison with the standard case when ΔS is not used for $t\bar{t} + PU200$ and for $jet(p_T = 250GeV) + PU200$. There are three noticeable points on those plots. First, the stub bending is a strong approach to mitigate the proliferation of roads and combinations. In average the reduction factor is about up to 10x for roads and 25x for combinations. Second, there's no big difference on the performances from ASYM and SYM approaches (except for 335577 configuration). Third, ΔS segmentation in detector inner layers has less effect than in the outer layers. That can be noticed by comparing the black circles, for instance. They represent the combinations @95%. The reduction from standard *sf1_nz1* to ASYM115577 is about 90%. Then for ASYM335577

Figure 62 - Stub bending distribution per detector layer for $t\bar{t} + PU200$ and the ΔS ranges for the configuration ASYM115577. In the ASYM mode only the central range can have width defined by user and ΔS always is divided only into 3 ranges.



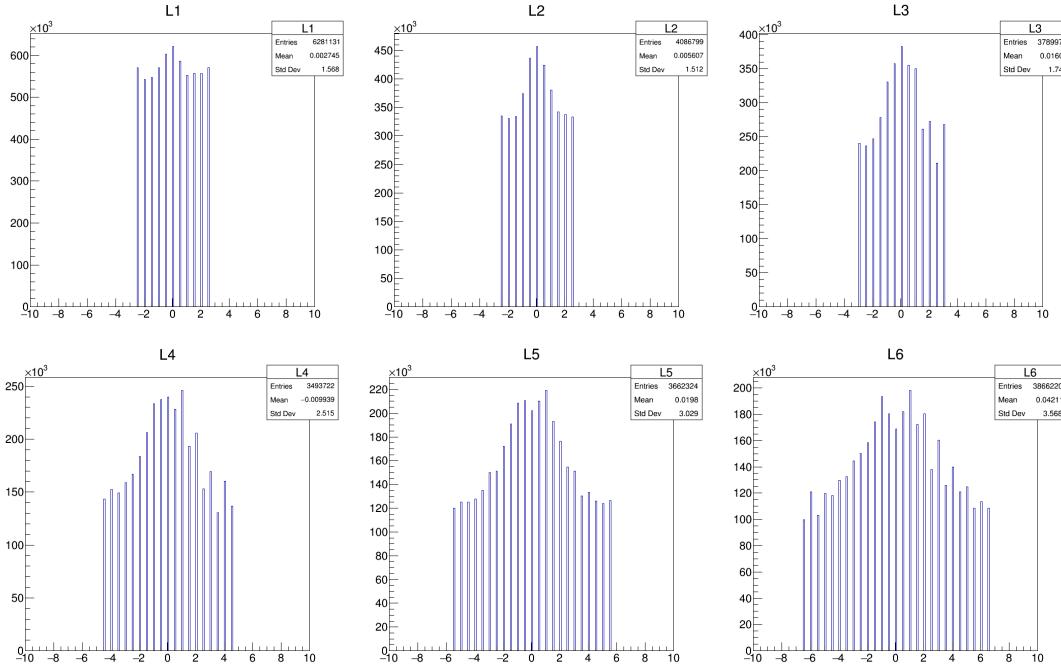
Source: The AUTHOR, 2016.

Figure 63 - Stub bending distribution per detector layer for $t\bar{t} + PU200$ and the ΔS ranges for the configuration SYM115577. In the SYM mode ranges have the same width, except the outer ranges (those ones comprehending $|\Delta S| = 6.5$).



Source: The AUTHOR, 2016.

Figure 64 - Stub bending distribution for $jet(p_T = 250GeV) + PU200$. Different events seems to have the same limits but, as one can see the occupancy is no equal (comparing with the pictures for $t\bar{t} + PU200$).



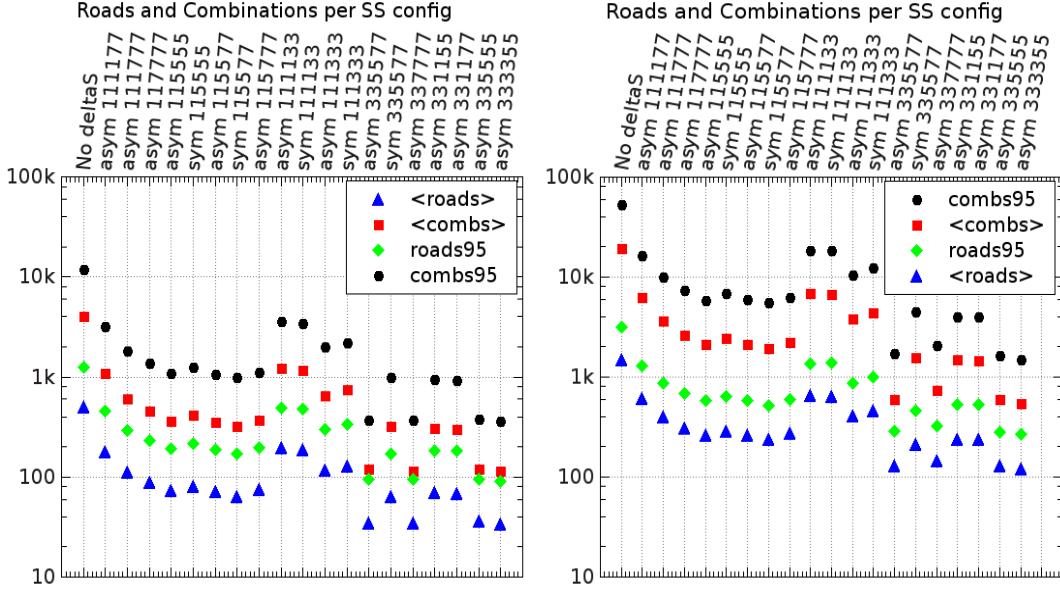
Source: The AUTHOR, 2016.

Table 25 - Stub bending possible divisions. The SYM method uses such table to decide how to split the ΔS values based on the number of ranges requested by user. The negative related of the ranges were omitted in the table.

#ranges	range width	ΔS values ([] central ranges)
3	9	[-2.0, 2.0], [2.5, ...]
5	7	[-1.5, 1.5], [2.0, 5.5], [6.0, ...]
7	5	[-1.0, 1.0], [1.5, 3.5], [4.0, 6.0], [6.5, ...]
9	3	[-0.5, 0.5], [1.0, 2.0], [2.5, 3.5], [4.0, 5.0], [5.5, ...]

Source: The AUTHOR, 2016.

Figure 65 - Roads and combinations average and 95 percentile per event versus different ΔS configurations at $sf1_nz1$ for $t\bar{t} + PU200$ (left) and $jet(p_T = 250GeV) + PU200$ (right).



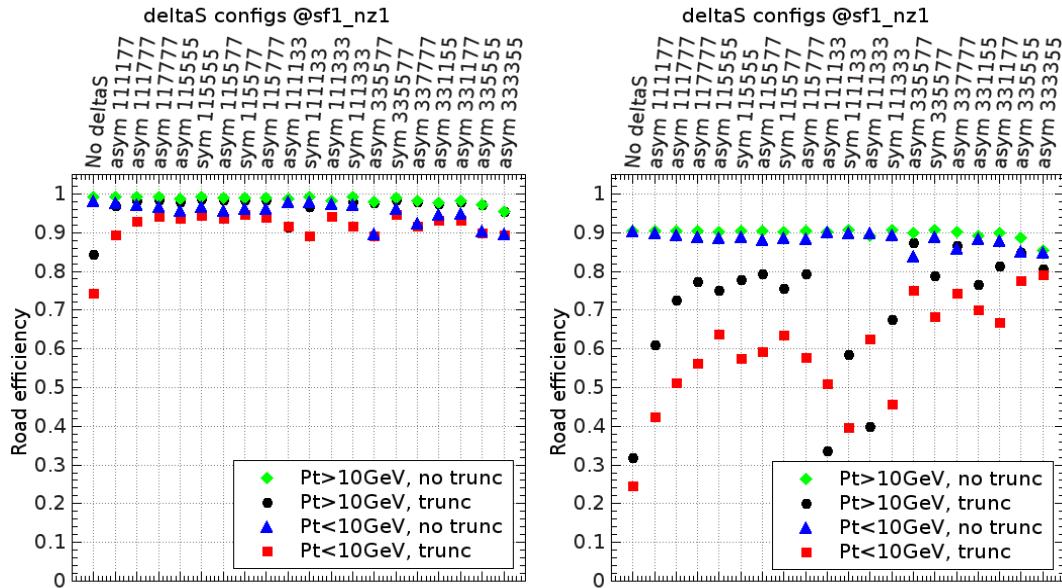
Source: The AUTHOR, 2016.

the extra reduction is just about 6%.

The next quantity to look on was the efficiency. In order to avoid the effects from other AM processing stages as tracking fit and DR, it was used the analytic efficiency from roads. Also, it was applied the standard truncation of 200 roads and 500 combinations. As showed in Fig. 66 the stub bending has an important role when truncation is applied and the effect is even more significant in jet case. In the better situation it's possible to recover 14% and 50% of efficiency for $t\bar{t} + PU200$ and $jet(p_T = 250GeV) + PU200$, respectively. It's noticeable that for $t\bar{t} + PU200$ the efficiency without and with truncation match for the two last stub bending configurations. Also, one can see that the difference between the ASYM and SYM approaches are more evident in terms of the efficiency. Based on Fig. 65 and Fig. 66, it was decided that the configuration ASYM335577 is, simultaneously for $t\bar{t}$ and jets, the best one.

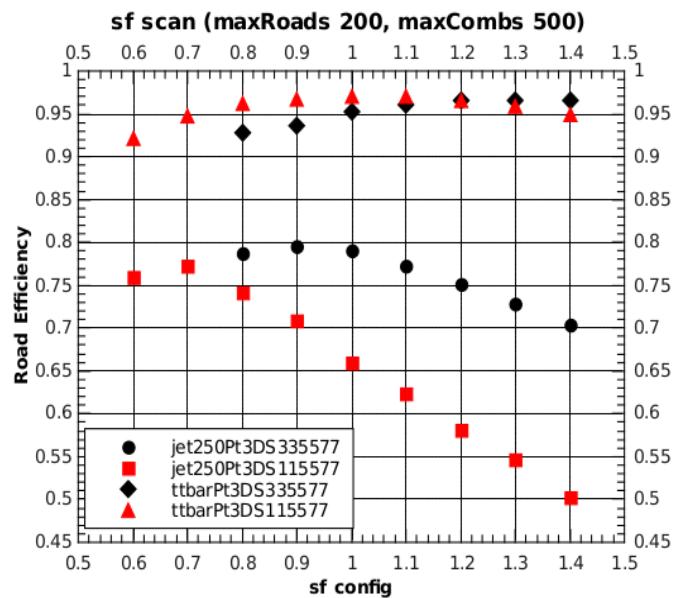
In order to make sure that the $sf1_nz1$ configuration is the best a scan was made for the stub fountain scale factor. It was also included for such scan the ASYM115577 configuration since it looks a better choice for $t\bar{t} + PU200$. The sf scan can be seen in the Fig. 67. That plot shows that in fact the ASYM335577 @ $sf1_nz1$ is an appropriate configuration.

Figure 66 - Road efficiency versus different ΔS configurations at $sf1_nz1$ for $t\bar{t} + PU200$ (left) and $jet(p_T = 250GeV) + PU200$ (right).



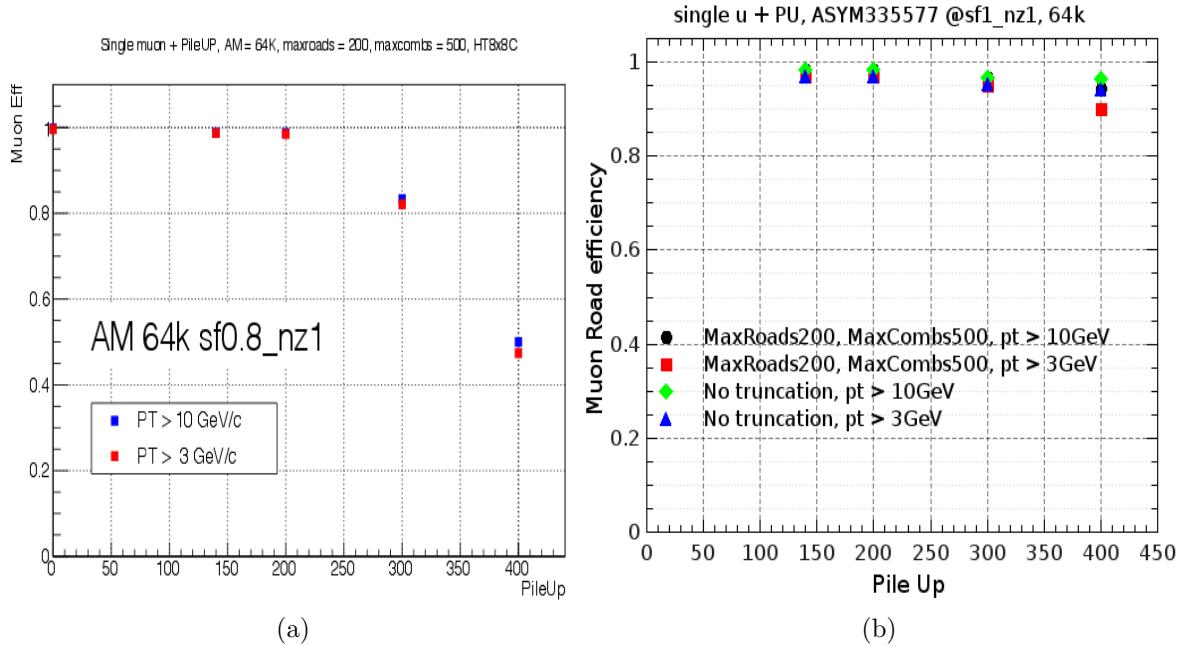
Source: The AUTHOR, 2016.

Figure 67 - Stub fountain scaler factor scan for $t\bar{t} + PU200$.



Source: The AUTHOR, 2016.

Figure 68 - Muon road efficiency versus pile up. The plot (a) is the result without ΔS approach (done by another person in the CMS L1TT working group) and the plot (b) with it (done by the author).



Source: The AUTHOR, 2016.

E.2.1.4 ΔS Approach on $\mu + \text{PU}$'s

Studies have been developed about the limits of AM+FPGA approach. It was showed that AM start to lose significant efficiency when PU300/400 is present¹² (even with Hough transform - worst without it). It was worth then to see what ΔS could do about it. Using the same samples ($\mu + \text{PU}140, 200, 300, 400$) a similar plot was produced to see how the analytical road efficiency is affected by pile up presence when ΔS is used. The result can be seen in Fig.68 where an original plot (a), without ΔS approach, is showed along with the results with stub bending approach (b). As one can see, with *sf0.8_nz1* and 64k the efficiency loss is big at PU400. Using ΔS approach the efficiency is ~ 40 higher than for *sf0.8-64k*.

¹² <https://indico.cern.ch/event/653731>

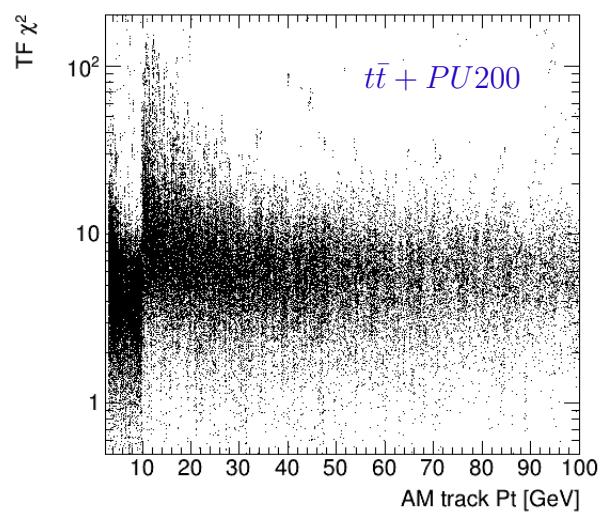
E.2.1.5 Track Fitter χ^2 Revision

Before proceeding to get final results using ΔS approach, some stages in the AM package were revisited. The first of them was the track fitter. Currently the tracks χ^2 have only a single cut equal to 14.6. However, there are two strong features in the TF χ^2 : it has two set of constants and its value has a strong dependence on track p_T (Fig. 69). Then it was decided to redefine the TF χ^2 cut. First, the cut should be defined separately for each logic (6/6, 5/6), keeping the cut for 6/6 tight. Also, the cut should be parameterized in terms of track p_T for 5/6. For that 8 p_T ranges were chosen: [3, 5], [5, 10], [10, 13], [13, 15], [15, 20], [20, 25], [25, 30], [30, ∞] GeV. Then for 6/6 it was decided to take the 99% of the theoretical $\chi^2_{ndof=8}$ curve (Eq. 75) being equal to 20.2. Fig. 70 shows how the all 6/6 reconstructed muon tracks (from $t\bar{t} + PU200$) behave compared to the expected distribution. Note that here no DR was used.

$$\chi^2 = \frac{x^{k/2-1}}{2^{k/2} \cdot \Gamma(k/2) \cdot e^{x/2}}, \quad k = \text{degrees of freedom} \quad (75)$$

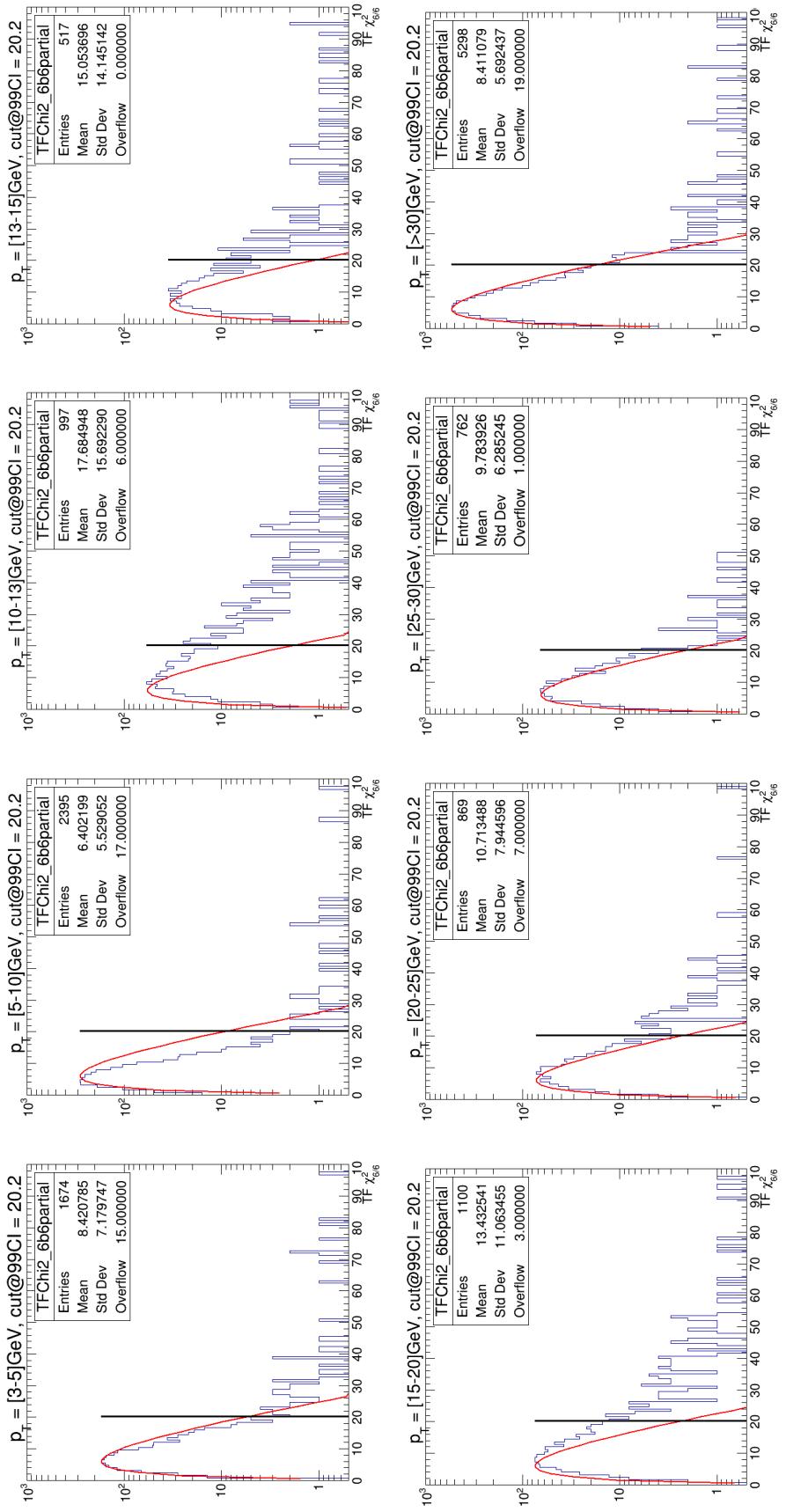
The normalization of the expected distribution was made by applying a scale factor to the Eq. 75 such that its maximum was equal to the maximum observed (the maximum in each histogram). As it can be seen in the plots, the distributions from some p_T ranges are wider or tinner than the expected curve. Such behavior matches with Fig. 69. The TF χ^2 has higher values for low p_T tracks (< 5 GeV) and in the transition region (> 10 GeV). Also, the tail in the 6/6 χ^2 distribution in disagreement with the expected curve shows that many bad stubs are being accepted with a cut of 14.6. The cut for 5/6 tracks were defined finding where the cut should be in order to get $\epsilon_{tracks} = 0.99 * \epsilon_{roads}$ for each p_T range. The distributions for $t\bar{t} + PU200$ and the cuts found for it are in Fig. 71. Those cuts were implemented in the package and used to fit tracks in the further studies.

Figure 69 - Track fitter χ^2 from all muon tracks in $t\bar{t} + PU200$ versus track p_T .



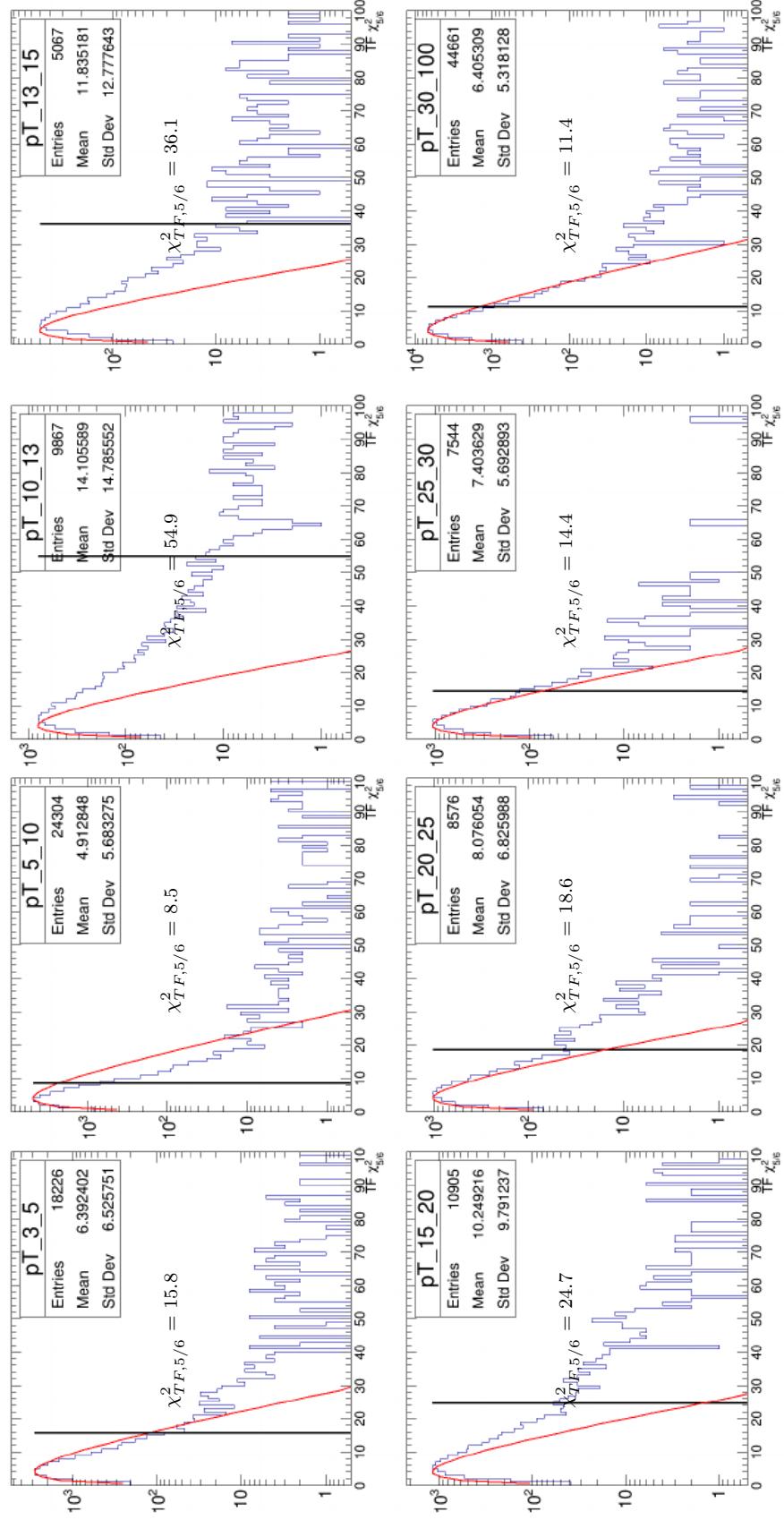
Source: The AUTHOR, 2016.

Figure 70 - TF χ^2 distribution for 6/6 muon tracks in each track p_T range from $t\bar{t} + PU200$.



Source: The AUTHOR, 2016.

Figure 71 - TF χ^2 distribution for 5/6 muon tracks in each track p_T range from $t\bar{t} + PU200$. Vertical bars point the cuts defined.



Source: The AUTHOR, 2016.

E.2.1.6 Synthetic Efficiency

Currently in the AM simulation package there are two methods to compute track reconstruction efficiency: analytic and synthetic. The analytic approach is based only on stubs produced by the tracking particles. Its threshold is 5 stubs, that is, a tracking particle (MC truth track) only can be matched to reconstructed tracks (AM track) that have at least 5 stubs belonging to the tracking particle. The synthetic efficiency approach uses only the four track parameters (q/p_T , ϕ_0 , z_0 and $\cot \theta$) to determine the AM track compatibility with a MC track. That is an important way to measure the efficiency because it happens that a stub belonging to a different MC tracking particle can build up a good combination with the other stubs, producing then a track that has compatible parameters (within track resolutions) to the desired MC tracking particle. In the AM package the synthetic efficiency is encoded in the module **MCTruthAssociator.cc** and it's called during tracking fit stage after the fit.

The base of synthetic efficiency is described by an algorithm that defines how to pair the collections of MC and AM tracks. Its current structure implemented in the AM package is the following:

- 1- Sort AM tracks with 6/6 first and 5/6 last, and in each group sort by decreasing p_T ;
- 2- Sort MC tracks by decreasing p_T ;
- 3- For each MC track, scan the collection of AM tracks to find those ones that satisfy certain cuts on all four track parameters:
 - Flag the AM track that gives the best match as **good**;
 - Flag all remaining AM tracks matching to a same MC track as **duplicate**;
- 4- Once all MC tracks have been scanned, flag as **false** all remaining AM track that is not marked as **good** or **duplicate**.

From such algorithm the definition of synthetic efficiency is eminent. The denominator is the number of MC tracks. The numerator is the number of AM tracks marked as **good**. In order to measure the track compatibility and decide which one gives the best match a metric is needed. Such metric was created via a χ^2 -like formula. Then each AM track paired to a MC track receives a rank that is computed based on the differences in the four track parameters (δp_i) normalized by their respective resolutions. The resolutions were determined by parameterizing the δp_i in function of q/p_T . For such task only single particle (no PU) events and only 6/6 tracks were used since they are expected to reproduce the original track parameters with better precision. The δp_i 's in each track parameter were fitted with the same function (constant+gaussian) for different p_T ranges,

Table 26 - Track resolution function parameters from single pion no PU.

Resolution function: $\Omega(q/p_T) = c1 + c2 * Gaus(q/p_T, \mu, \sigma)$				
Parameter	q/p_T	ϕ_0	z_0	$\cot \theta$
c_1	0.001702	0.000663	0.178339	0.003157
c_2	-0.001517	-0.000559	-0.099997	-0.001072
μ	0.000256	-0.000678	0.009426	0.005015
σ	0.148572	0.155241	0.838881	0.268481

Source: The AUTHOR, 2016.

producing the points showed in Fig. 72. The dependence on q/p_T is stronger for q/p_T and ϕ parameters. For z_0 and $\cot \theta$ it is small but the points can still be described with the same function. As one can see the function parameters are not too different between different event types and its shape is approximately the same, even when PU is added. Since pion tracks are the majority of tracks reaching the detector in particle collisions and a common function for all event types is appropriate to be defined, the function parameters from pions (Tab. 26) were taken. After that a pseudo- χ^2 (usually called χ^2_{match}) was defined allowing one to control the tracking matching window.

The χ^2_{match} correlates the four track parameters through the following formula,

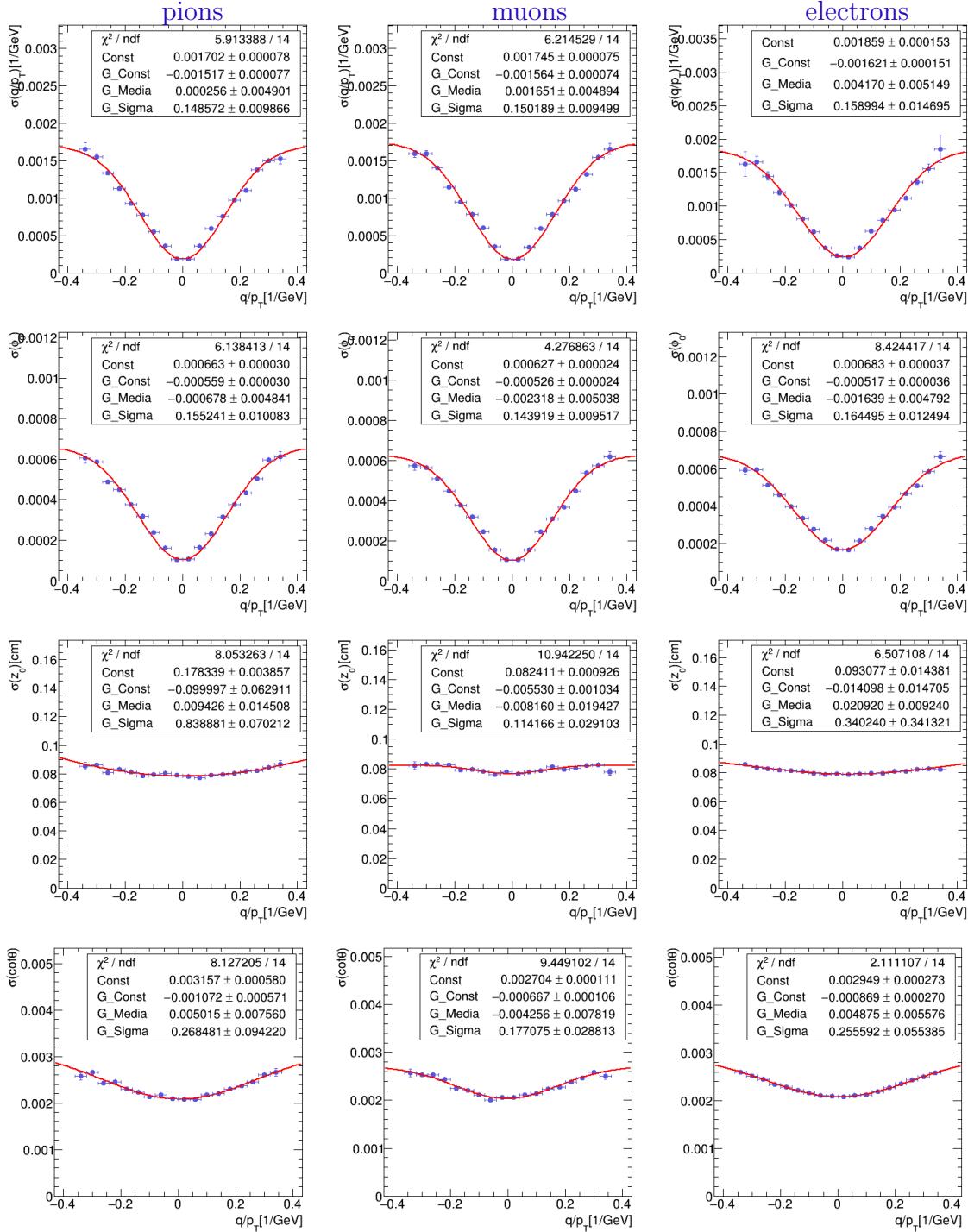
$$\chi^2_{match} = \sum_{i=0}^4 \frac{\delta^2 p_i}{\Omega^2(q/p_T)_i}, \quad \delta p_i = (p_i^{MC} - p_i^{Reco}) \quad (76)$$

where p_i^{MC} (p_i^{Reco}) stands for the four track parameters from MC (Reco) tracks. Just for completeness the χ^2_{match} is divided by the number of track parameters, that is, four. This reduced χ^2_{match} then gives a ruler for the synthetic efficiency. Once a cut is chosen it defines where the good, duplicate and fake tracks sit. Tracks under such threshold will be good or duplicate and tracks above it will be fake tracks.

First time the cut was chosen based on the capability of finding 99% of the pion tracks, being at $\chi^2_{match}/ndof = 12.8$. As one can see in Fig. 73 it's a good cut even when PU140 is present. Also, it's good for muon case as expected since it's not a busier environment. Using such cut, a group of 4 different event types (muon, pion, electron and pure PU) with different PU levels could be characterized in terms of tracks category (good/duplicate/fake) and track reconstruction efficiency. In Fig. 74 one can see an example. The turn-on at low p_T range (3 to 6GeV) in the efficiencies is associated to another effect and not due to tightness of the synthetic match cut. It has to do with trigger tower definition and will not be discussed here.

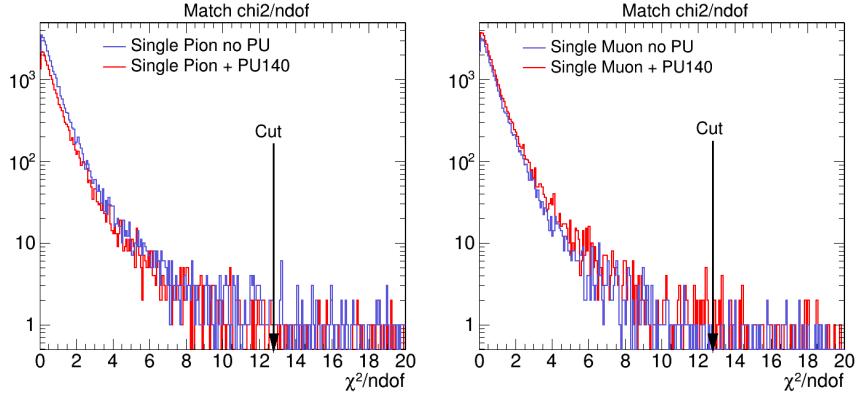
In the track files created by AM package there are three branches that store important information from synthetic approach. Currently, the branch **AMTTTracks_synMatchCat** stores an integer that tags AM track category: 1 for good tracks, -1 for duplicates and -2

Figure 72 - Track parameters resolution as function of q/p_T for single pion no PU (first column), single muon no PU (second column) and single electron no PU (last column). From top to bottom the track parameters are q/p_T , ϕ_0 , z_0 and $\cot \theta$.



Source: The AUTHOR, 2016.

Figure 73 - χ^2_{match} distribution from single pion and single muon event without and with PU.



Source: The AUTHOR, 2016.

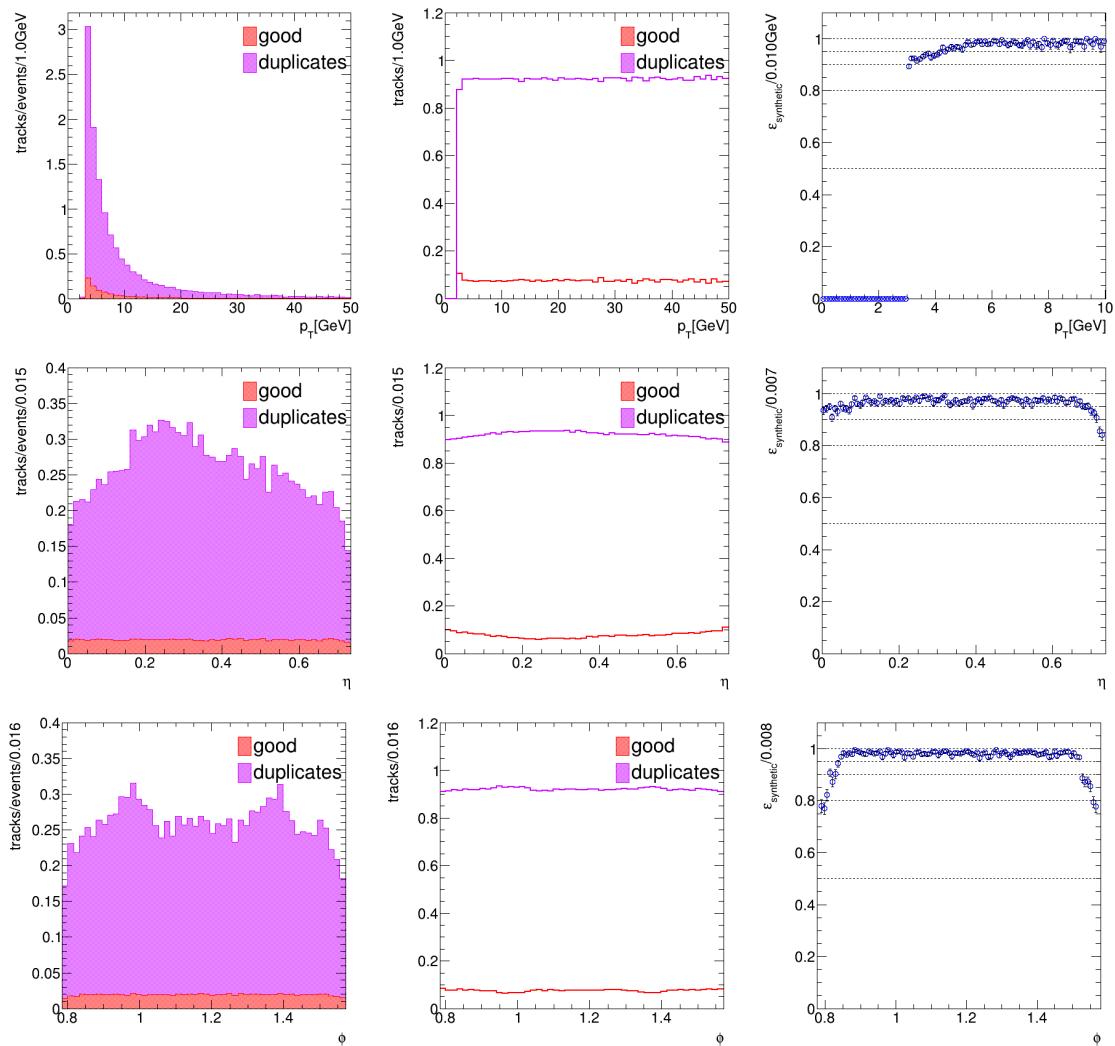
for fakes. The branch **AMTTTracks_synTpId** stores the tracking particle index in a event to which the AM track was matched (good or duplicate). Index equal to -1 means that the AM track is fake, that is, it could not match to any MC tracking particle then it can't have an index. The branch **AMTTTracks_matchChi2** stores the χ^2_{match} value for AM tracks flagged as a good or duplicate track. For fake tracks the value is 999999999.

E.2.1.7 χ^2_{match} Optimization and Synthetic-Analytical Methods Connection

In order to better decide the cut to be used on the χ^2_{match} , avoiding then random combinations of stubs but without loosing efficiency in busier event a new strategy was designed. It also allowed an unification between analytical and synthetic approaches. Such strategy consisted on scanning through different cuts on χ^2_{match} and observing how the synthetic efficiency and the fraction of good/duplicate/fake tracks change. It's intuitive that a very restrictive cut will cause loss of efficiency making good tracks being classified as fake. On the other direction a very loose cut will make tracks with quite different parameters to be matched, so real fake tracks will be classified as good or duplicate and duplicate tracks can be classified as good ones.

Other quantity that was monitored through those scannings was the fraction of truth stubs in the AM tracks being classified as good or duplicate. That fraction was computed checking how many stubs the AM track had in common with the matching MC track. Both of those informations can be seen in Fig. 75. The notation used in those plots are track category and number of stubs. For instance, **GOOD 6S** stands for AM tracks flagged as good and have 6 stubs that come from the matching tracking particle (the same for the duplicates). The plots on the top show that a very restrictive cut on χ^2_{match}

Figure 74 - AM track characterization and synthetic efficiency from single muon no PU events.



Source: The AUTHOR, 2016.

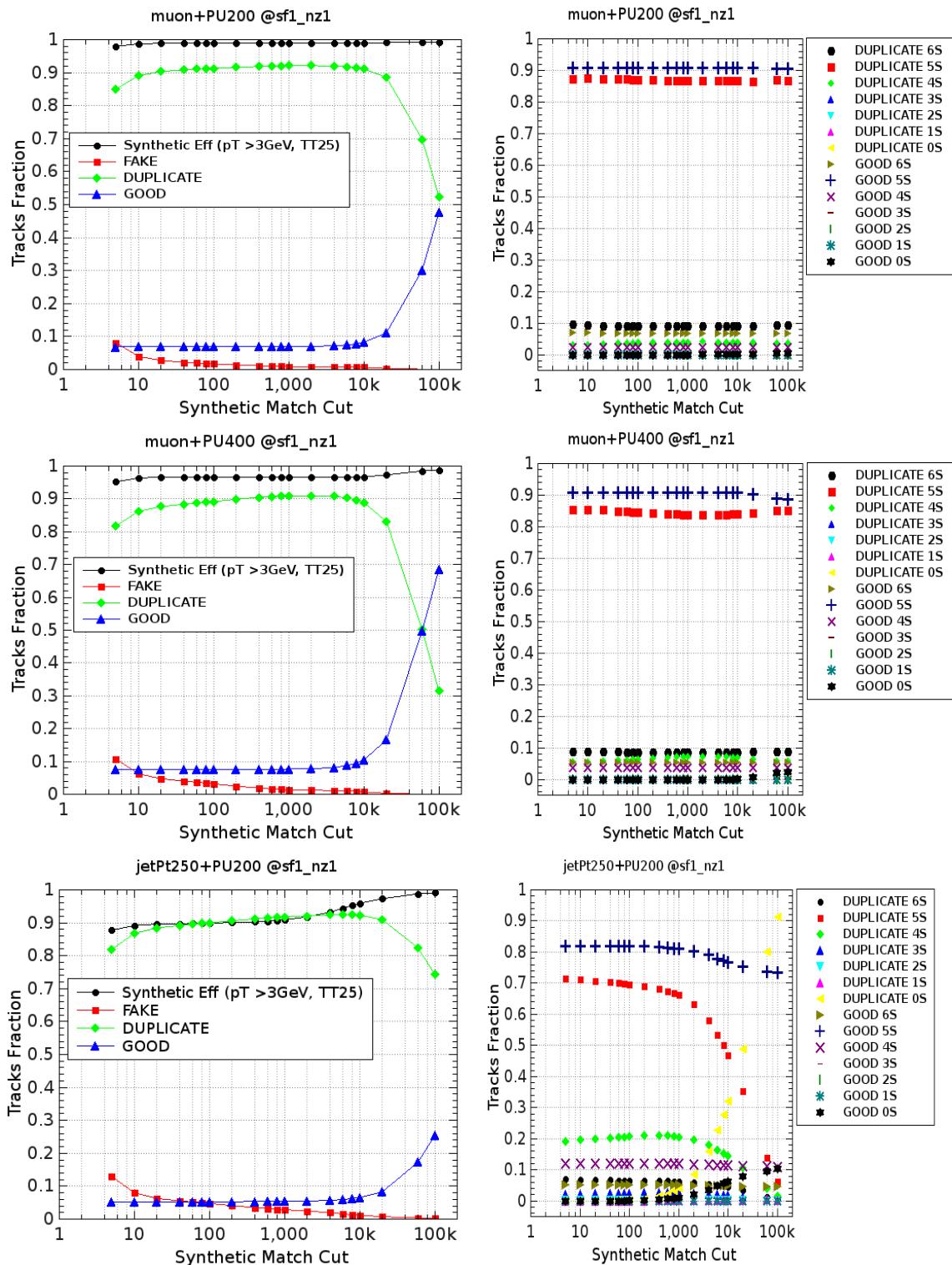
produces many fake tracks as it's too tight to compensate the track parameters resolution. Once the cut starts to increase the fraction of fake tracks quickly gets reduced while it's possible to see the fraction of duplicates and the synthetic efficiency increasing (as the fraction of good tracks also slowly increases). After that the three curves reduces their slope - the synthetic efficiency becomes actually flat. Then, on $\chi^2_{match} \sim 1k$ the curves start to change faster again and after 10k the fraction of good tracks increases abruptly and duplicates decrease abruptly. A last thing to notice is the flat region for synthetic efficiency in the plots for $\mu +PU200$ and $+PU400$. It looks quite similar, showing no strong dependence on PU.

Looking into the bottom plots one see what happens in the previous stages through the analytical view. First thing to make clear is that, in that plot the number of stubs stated in the legends means the number of stubs belonging to the tracking particle that the AM track was matched - and not the actual number of stubs the track has. Then, the second thing to notice is that the majority of AM tracks have 5 stubs coming from the tracking particle (either good or duplicate tracks) - if the AM track is a 6/6, it means that the track has one stub from other tracking particle. In such case, that majority comes from three sources: one is that naturally there are more 5/6 than 6/6 tracks, two is that the advanced combination builder produces 5/6 combinations from all 6/6 combinations and three is that a 6/6 AM track may not have 6 stubs from the same tracking particle. In the frame of synthetic efficiency the most significant result is the fraction of tracks that don't have any stub coming from the tracking particle to which it was matched (GOOD 0S and DUPLICATE 0S). In other words, the χ^2_{match} cut is too loose that random combinations of stubs are being flagged as a track. So, for sure one needs to avoid that region which imposes an upper limit on $\chi^2_{match} = 100$ based on the plot for jets. For μ is not so evident maybe because of the muon track is more un-like a pion track (that essentially composes the pile up). While, for jets, is more susceptible to occur that a pion track ends up reconstructed with stubs coming from other pion. Currently, an flag was developed to control the χ^2_{match} cut. Its default value is 40 that was agreed to be the adequate cut based on Fig 75. An user can change it during simulation actioning the flag **--maxChi2Match** followed by the cut value.

E.2.1.8 Duplicate Removal (DR) Based on Stubs

The duplicate removal was designed to mitigate tracks replicas. Those are tracks that are very similar between them, sharing equal stubs and usually having equal/similar parameters. In the level of AM work structure, those tracks arise due to stubs combinatorics (from the same road) generated (in tracking fit stage) in order to find out the appropriate group of stubs coming from a real track. A road can also have extra stubs

Figure 75 - AM tracks categories fraction and truth stub fraction versus χ^2_{match} cut for $\mu + PU200$ on left column and for $jet(p_T = 250GeV) + PU200$ on right column.



Source: The AUTHOR, 2016.

due to secondary radiation from original tracking particle or due to different particles crossing the same detector region that comprehends an expected pattern. As mentioned, the advanced combination builder generates 5/6 combinations from 6/6. Those 5/6 combinations are also duplicates.

The DR based on stubs is an analytical approach since it uses the stubs shared between different tracks. It's current algorithm is:

- 1- Take the first AM track and store it into a new list (unique tracks list);
- 2- Then compare each remaining AM track to the tracks stored in the unique tracks list:
 - If an AM track and an unique track have more equal stubs than allowed remove the AM track;
 - Else, add the AM track into the unique tracks list;
- 3- Repeat the process until all AM tracks have been scanned.

A scan on the DR option was made monitoring the fraction of AM track categories per event and the synthetic efficiency for three samples: μ , $t\bar{t}$ and jet +PU200. The synthetic efficiency was computed using only the muons from μ and $t\bar{t}$ +PU200 samples, while for jet sample all tracks are included in the computation. The results are showed in Tab. 27. From them, first important thing to notice is the ratio of synthetic efficiency and the track efficiency. The track efficiency is the analytical efficiency computed using the tracks, that is, a tracking particle is considered found if at least a combination has at least 5 stubs coming from it. As expected the synthetic efficiency is slightly higher than the analytical efficiency due to extra stubs.

The numbers also show that the majority of the duplicate tracks have 4 stubs in common with the good tracks since DR options 5 and 4 are not effective on duplicate removal. As it can be seen $DR \leq 3$ significantly affects the analytical track efficiency although synthetic efficiency reduces very few. That suggests a DR option even more restrictive like $DR = 0$ (the only one actually capable to completely remove all duplicate tracks). So, it was decided to reduce the DR option previously adopted of 2 to 0. It was also noticed that DR is also removing fake tracks. That is a new and important result since those fake tracks mean hardware latency after all. So, DR will also helps to reduce the AM processing latency. A study from where the extra stubs (that makes synthetic efficiency higher than the analytic one) come from was done. Through it, however, it was discovered that the majority of those stubs come from particle not identified by the current software (**cmssw**). And a very small fraction come from pions (which ones would be called random stubs since are not expected to be correlated to the μ - as an irradiated γ , for instance).

Table 27 - AM track categories per event versus DR configuration for $\mu + PU200$, $t\bar{t} + PU200$ and $jet(p_T = 250GeV) + PU200$.

$\mu + PU200$					
DR option	Goods	Duplicates	Fakes	Track eff	Synthetic eff
None	1.976	25.785	0.614	0.985	0.989
5	1.976	25.785	0.614	0.985	0.989
4	1.976	8.898	0.275	0.98	0.989
3	1.973	0.604	0.095	0.964	0.989
2	1.969	0.065	0.047	0.953	0.989
1	1.967	0.007	0.039	0.951	0.988
0	1.966	0.000	0.038	0.951	0.988

$t\bar{t} + PU200$					
DR option	Goods	Duplicates	Fakes	Track eff	Synthetic eff
None	3.206	43.205	3.252	0.981	0.983
5	3.206	43.205	3.252	0.981	0.983
4	3.205	14.993	1.384	0.976	0.983
3	3.197	1.023	0.400	0.957	0.98
2	3.183	0.111	0.253	0.948	0.977
1	3.179	0.008	0.230	0.945	0.976
0	3.178	0.000	0.226	0.945	0.976

$jet(p_T = 250GeV) + PU200$					
DR option	Goods	Duplicates	Fakes	Track eff	Synthetic eff
None	8.506	143.735	8.924	0.89	0.897
5	8.506	143.735	8.924	0.89	0.897
4	8.506	52.935	4.109	0.883	0.897
3	8.481	4.746	1.167	0.823	0.895
2	8.431	0.642	0.597	0.754	0.889
1	8.412	0.067	0.506	0.738	0.887
0	8.406	0.003	0.482	0.74	0.886

Source: The AUTHOR, 2016.

The DR code is wrote in the **DuplicateRemoval.cc** module and it has the flag **--rmDuplicate** that can be used to set the maximum number of stubs to be shared between unique tracks. The values must be integers from 0 to 6. The default option is 0 and to set off the DR the user must pass the argument -1 to the flag. Since DR is an analytical approach it acts right after tracking fit stage and before the synthetic match stage being actioned.

E.2.1.9 Final FOMs

In the following pages will be show the common FOMs (figure of merits) usually used to characterize the final performance of AM simulation. Such plots (Fig(s). 76, 77, 78, 80, 80) show the synthetic efficiency for the track reconstruction and the track categories ratios in terms of the track p_T , η and ϕ . The plots also show the comparison between the standard *sf1_nz1* configuration and two chosen ΔS schemes: ASYM335577 and SYM335577. Also, truncation at 200 roads and 500 combinations is applied.

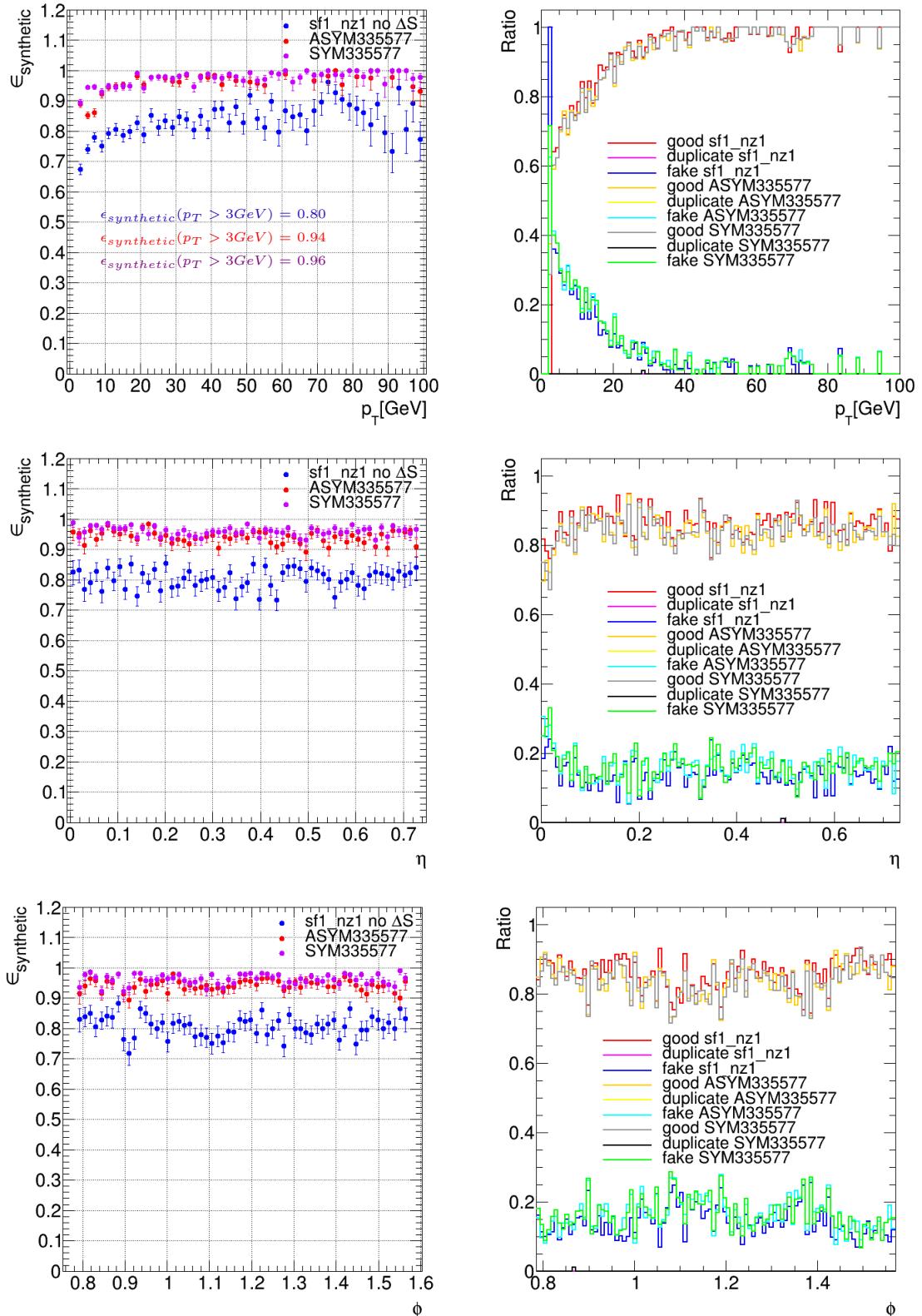
E.2.1.10 Generated AM Pattern Banks

During the studies presented here several pattern banks have been generated and for the information of the reader and even possible reference, the properties of such banks are summarized in this section through the Tab(s). 28, 29 and the Fig. 81. As presented along the text it was decided to work around the configuration *sf1_nz1*, which has a good road efficiency (as shown in Fig. 67) to develop the stub bending approach and that is why, in the tables, the notation for the stub bending approach only appears for that particular configuration. Fig. 81 compares the size of the banks for the standard *sf1_nz1* and its variants with the application of the stub bending approach. As it is expected such approach increases the number of pattern due to the extra parameter assigned to each stub (remember, that, though, does not increase the partition of the detector).

E.2.1.11 Road Efficiency vs Truncation in $\mu + PU400$

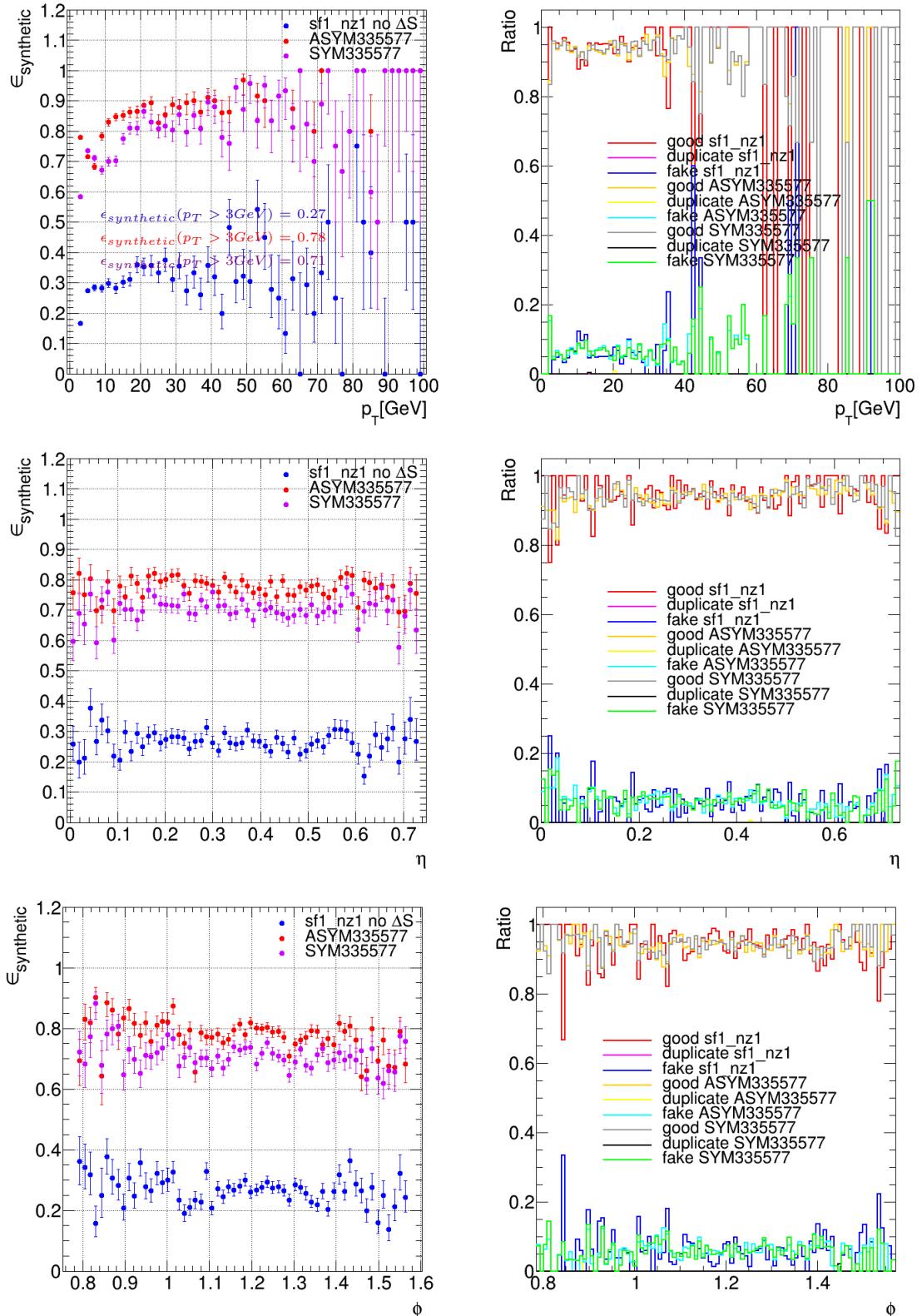
This section is just a extra piece of information for the reader, even more in a case of future reference in case the CMSL1TT approach starts to move forward again. After the implementation of the stub bending approach, the author had decided to have a close look on the effect of the truncation over the road find efficiency. The study was performed using a sample of single muon in a environment with 400 PU. The pattern bank used had

Figure 76 - Track reconstruction efficiency for $t\bar{t} + PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplicate removal was applied by requiring DR=0.



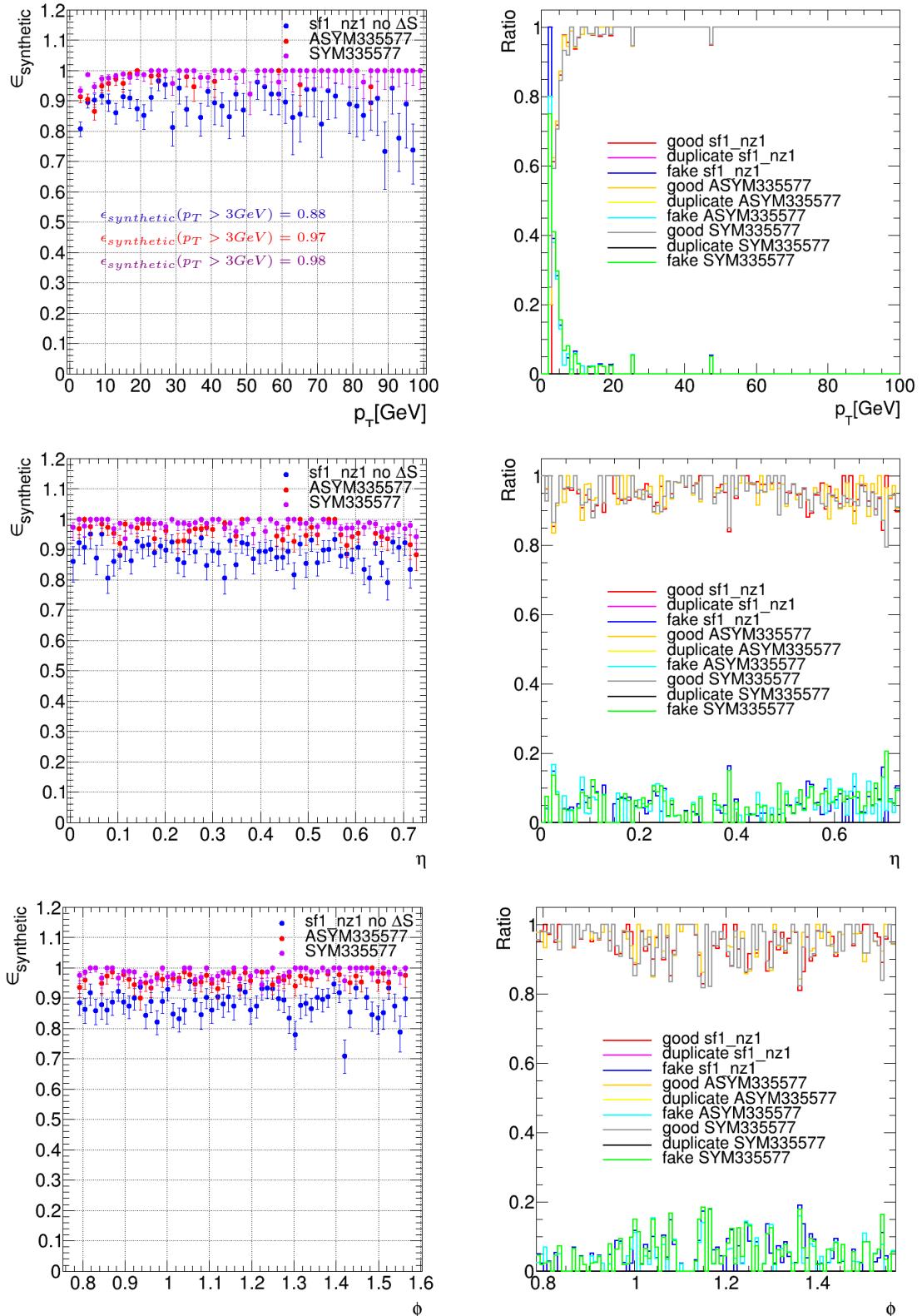
Source: The AUTHOR, 2016.

Figure 77 - Track reconstruction efficiency for $jets(p_T = 250GeV) + PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.



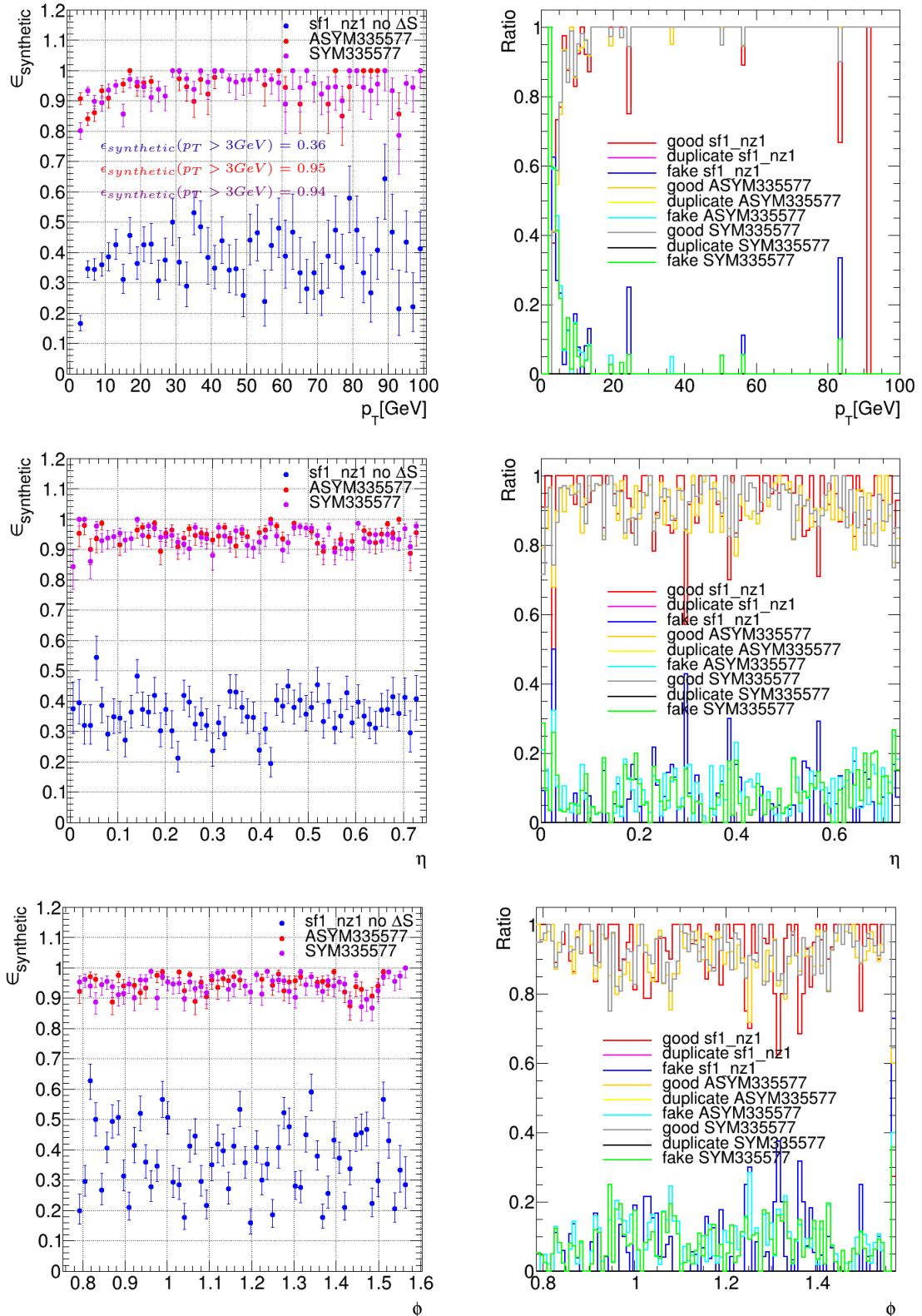
Source: The AUTHOR, 2016.

Figure 78 - Track reconstruction efficiency for $\mu + PU200$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.



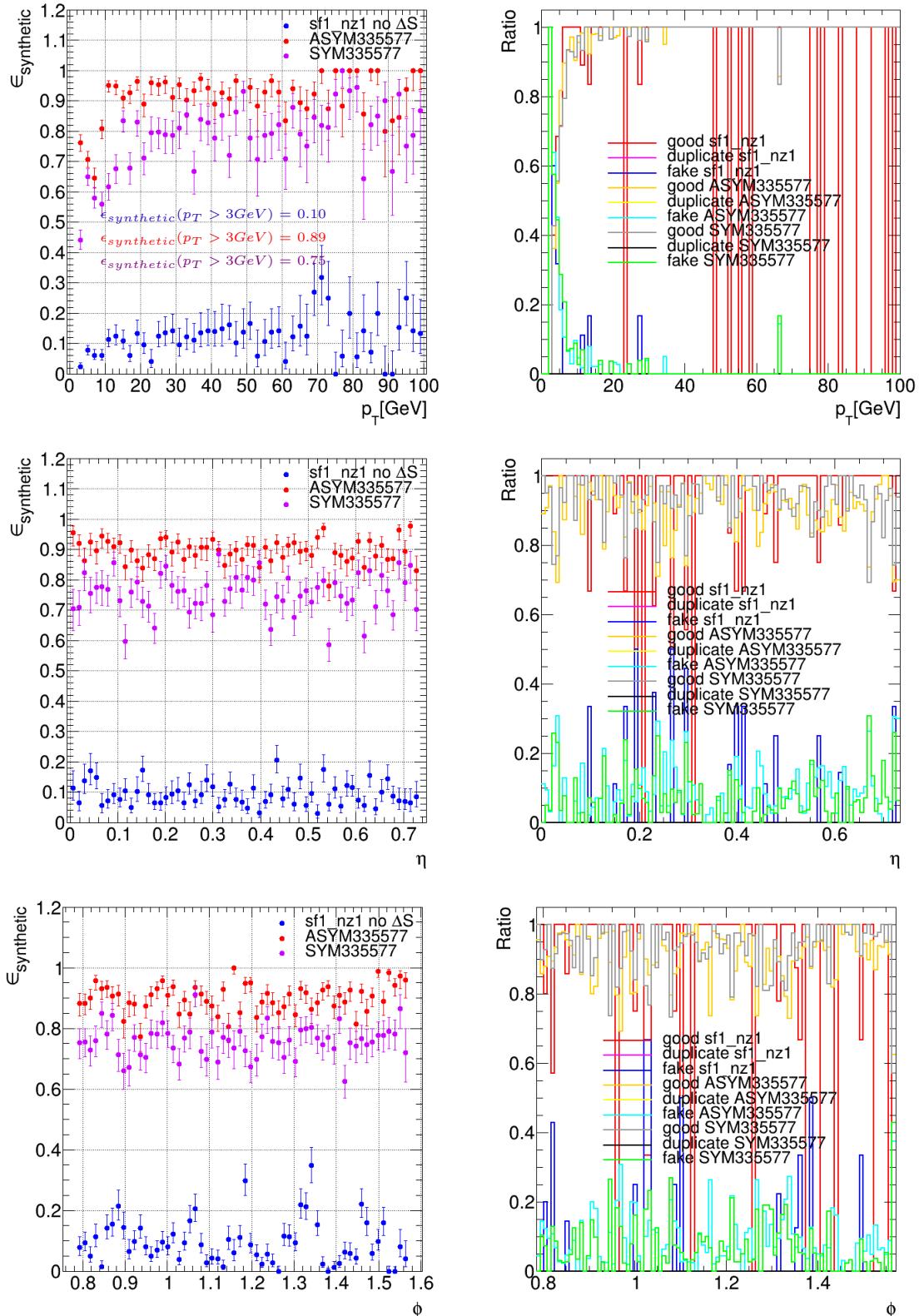
Source: The AUTHOR, 2016.

Figure 79 - Track reconstruction efficiency for $\mu + PU300$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.



Source: The AUTHOR, 2016.

Figure 80 - Track reconstruction efficiency for $\mu + PU400$ sample. The pattern bank used had 64k and truncation at 200 roads and 500 combinations has been applied. Duplication removal was applied by requiring DR=0.



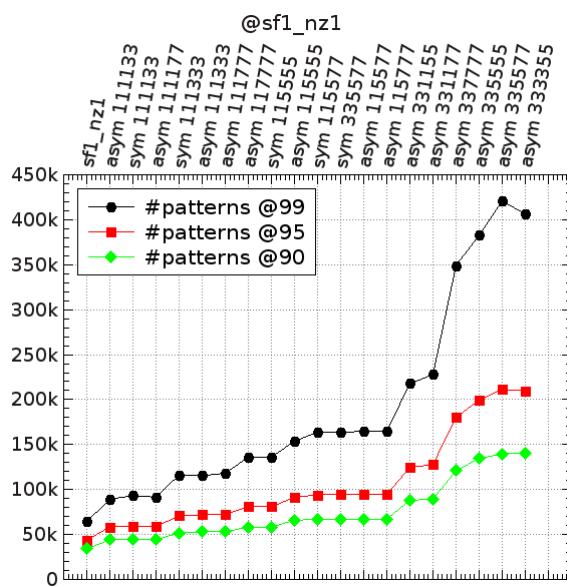
Source: The AUTHOR, 2016.

Table 28 - AM pattern banks size for different configurations and coverages. The configurations stands for the way the CMS detector is partitioned in ϕ and \mathbf{z} directions. The coverage stand for the fraction of track reconstructed over the total number of track patterns in a given configuration.

Bank	@90	@95	@99
<i>sf0.3_nz1</i>	1011898	1408735	2393133
<i>sf0.4_nz1</i>	385610	527271	857645
<i>sf0.5_nz1</i>	197421	265719	419641
<i>sf0.6_nz1</i>	119960	159487	246058
<i>sf0.7_nz1</i>	80199	105252	159175
<i>sf0.8_nz1</i>	57483	75332	112215
<i>sf0.9_nz1</i>	43718	56422	82767
<i>sf1.0_nz1</i>	34081	43874	64253
<i>sf1.1_nz1</i>	27437	35110	50834
<i>sf1.2_nz1</i>	22720	28907	41472
<i>sf1.3_nz1</i>	18930	24144	34451
<i>sf1.4_nz1</i>	16205	20560	29058

Source: The AUTHOR, 2016.

Figure 81 - Comparison between the pattern bank *sf1_nz1* and its variants due to the stub bending approach.



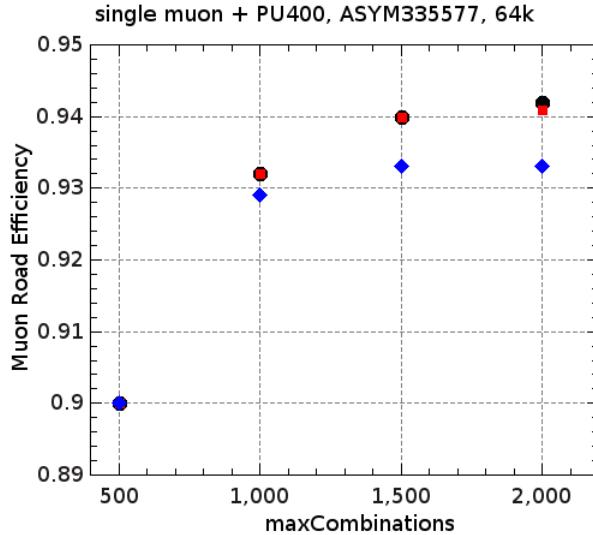
Source: The AUTHOR, 2016.

Table 29 - AM pattern banks size for different configurations and coverages. The configurations stands for the way the CMS detector is partitioned in ϕ and \mathbf{z} directions. The coverage stand for the fraction of track reconstructed over the total number of track patterns in a given configuration. The meaning of ASYM and SYM nomenclature have been introduced in Sec.E.2.1.3.

Pattern Bank	@90	@95	@99
ASYM 111133 @ $sf1_nz1$	44043	58323	89271
SYM 111133 @ $sf1_nz1$	44410	59344	93151
ASYM 111177 @ $sf1_nz1$	44141	58813	91459
ASYM 111333 @ $sf1_nz1$	52926	72009	115405
SYM 111333 @ $sf1_nz1$	51744	70820	116106
ASYM 117777 @ $sf1_nz1$	58407	80810	136003
ASYM 115555 @ $sf1_nz1$	65816	91351	153837
SYM 115555 @ $sf1_nz1$	58408	80813	136022
ASYM 115777 @ $sf1_nz1$	67069	94598	164988
ASYM 331177 @ $sf1_nz1$	89307	127762	227835
ASYM 335555 @ $sf1_nz1$	134630	199502	382805
ASYM 333355 @ $sf1_nz1$	140635	209775	406399
ASYM 111777 @ $sf1_nz1$	52961	72335	118083
ASYM 115577 @ $sf1_nz1$	67057	94756	164687
SYM 115577 @ $sf1_nz1$	67015	94120	163642
ASYM 331155 @ $sf1_nz1$	88075	124587	217815
ASYM 337777 @ $sf1_nz1$	120946	180225	349130
ASYM 335577 @ $sf1_nz1$	139533	211445	420652
SYM 335577 @ $sf1_nz1$	67018	94123	163645
ASYM 115577 @ $sf0.3_nz1$	2062636	3123107	6826943
ASYM 115577 @ $sf0.4_nz1$	771240	1143988	2246040
ASYM 115577 @ $sf0.5_nz1$	391520	573532	1078009
ASYM 115577 @ $sf0.6_nz1$	236785	342370	623608
ASYM 115577 @ $sf0.7_nz1$	157642	225866	405288
ASYM 115577 @ $sf0.8_nz1$	113817	161966	286350
ASYM 335577 @ $sf0.8_nz1$	236767	361991	739049
ASYM 115577 @ $sf0.9_nz1$	85933	121500	212050
ASYM 335577 @ $sf0.9_nz1$	178506	271420	543726
ASYM 115577 @ $sf1.1_nz1$	54038	75917	130618
ASYM 335577 @ $sf1.1_nz1$	112462	169536	333569
ASYM 115577 @ $sf1.2_nz1$	44862	62813	107550
ASYM 335577 @ $sf1.2_nz1$	93467	140240	274192
ASYM 115577 @ $sf1.3_nz1$	37557	52558	89299
ASYM 335577 @ $sf1.3_nz1$	78456	117559	227121
ASYM 115577 @ $sf1.4_nz1$	32166	44774	75484
ASYM 335577 @ $sf1.4_nz1$	66909	99746	192147

Source: The AUTHOR, 2016.

Figure 82 - Truncation is not a linear behavior for ΔS . Increasing combination is better than increasing road truncation limit. Increasing combination truncation limit to 1k increases 3% almost equally for 200/400/800 roads.



Source: The AUTHOR, 2016.

64k patterns, which was a standard size at that time in the CMSL1TT studies. Also, stub bending approach have been used by the requirement of configuration ASYM335577.

The conclusion drawn by the author is that the truncation and the loss of efficiency is not a linear relation. Fig. 82 shows the variation in the muon road finding efficiency versus the truncation in the number of combinations and roads. The blue lozenges, the red squares and the black dots represent the efficiency variation for truncation at 200, 400 and 800 roads, respectively. As it can be seen, it is better to loose the truncation in the combinations than in the roads. The efficiency increases $\sim 3\%$ doubling the threshold in the combinations, while it does not increase significantly when changing the truncation in the roads (just $\sim 0.2\%$ by increasing both the number of roads and combinations). The author calls the attention, however, to the fact that here the stub bending approach has been applied and there is not knowledge about other circumstances, so that, this behavior can be not always true.

E.2.2 The Hardware work in the CMSL1TT

With less involvement (due to lack of experience) but yet crucial task, in the hardware side, the author was part of a group with three students (guided by another experienced student in the hardware) which was responsible for checking the status of the electronic boards used in the project. Also, the group was responsible to check the

quality of the communication channels of such boards and the optical cables used to interconnect them. For that some (very) few manipulation with code HVDL was carried out to program the FPGAs in the boards and make few data flow across the boards.

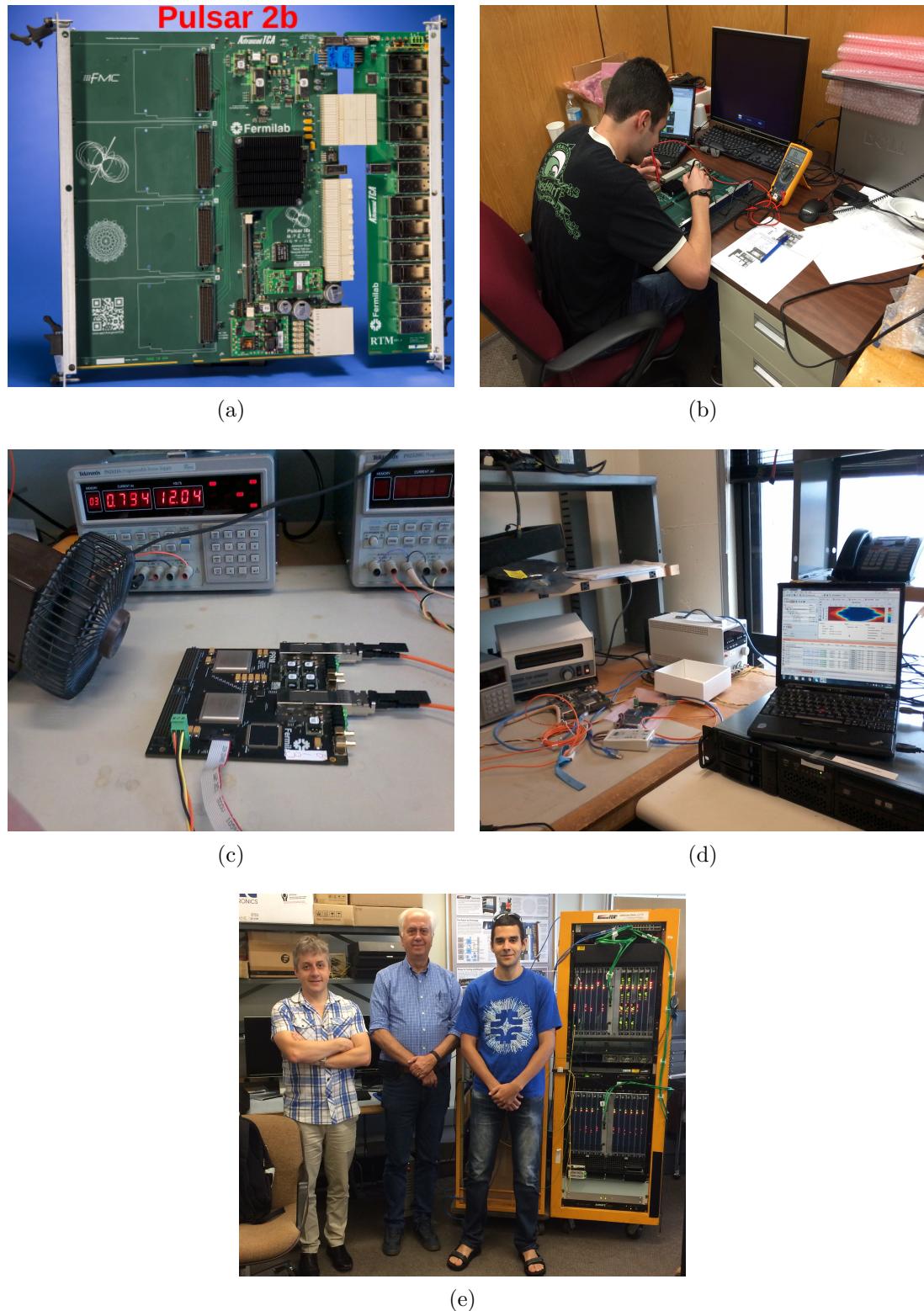
The hardware used in the CMSL1TT AM+FPGA is summarized in Fig. 83. The main component is the Pulsar IIB board, which is shown in Fig. 83(a). The Pulsar IIB is the responsible for organizing the data coming in and out of the tracker hits filtering process. It has a big FPGA located close to its center (under the heat sink), which holds the frame of the CMSL1TT AM+FPGA approach, such that, the boards can communicate between them (since one does not want to loose tracks happening in the boundary of the trigger towers). That FPGA takes care of organizing the hits coming from the detector, and send them to the mezzanine cards where the pattern match and further steps happen. Another feature is that, the Pulsar IIB process 8 bunch crosses per time, that is, the board receives a train with information of 8 bunch crosses and then organizes them for the full processing.

The mezzanine boards, usually called in the group as PRMs (Patter Recognition Mezzanines), shown in Fig. 83(c), holds two FPGAs with lower processing power. Those FPGAs, during the development and demonstration of the AM+FPGA approach were used to store the pattern banks and to hold the framework for the road matcher, the combination builder, the track fitter and the duplicate removal procedure discussed in previous sections. The track hits can flow from the Pulsar IIB board to the PRM by two ways: the LVDs or the optical connections. The ones used were the optical ones, but the LVDs were thought as a backup connection.

The author got involved in the inspection of both boards and also the cables in order to assure good connections between the boards and mezzanines and also with the crate, where there is an extra serial connection to which all the Pulsar IIB boards connect (receiving power supply and having connection between them). The inspection of the Pulsar IIB boards consisted first of checking its electronic components, such as capacitors, resistors, diodes and so on (Fig. 83(b)). The procedure has de goal to verify that there was no important component missing or with bad weld. In fact, about 2 boards were found to be missing capacitors and with bad weld, which was identified during the inspection of the board backend connections. Once the components were found without problems then the board was connected to a power supply and small routines were programmed within the FPGA, those which show the status of the board (if connected or not, if there was a mezzanine installed and so on).

The mezzanine inspections were similar. The main components were first visually checked and then the mezzanines were powered in order to verify the FPGAs and check the optical connections (Fig. 83(c) and (d)). The picture displayed in the laptop in Fig. 83(d) is the typical performance graph verified during the inspections and it shows the quality of data transmitted, being better as the blue area increases.

Figure 83 - Overview of the FNAL AM CMSL1TT hardware: (a) Pulsar-IIB board, (b) inspection of Pulsar-IIB components, (c) and (d) the setup for measure data transmission quality from the PRMs via optical cables, (e) crate containing Pulsar-IIB boards and PRMs (not visible since Pulsar boards sit right besides each other) with full setup ready for demonstration of the approach.



Source: The AUTHOR, 2016.

F THE FAST MATRIX ELEMENT (*FASTME*)

F.1 Introduction

In this chapter it will be described a method studied during the first year of this present PhD. This method has been called *Fast Matrix Element (FastME)*. The first studies for this method were developed using samples of simulated events of ggH and qqZZ (the SM Higgs main background), both decaying into four leptons in the final state. These events were generated using MadGraph (MC generator which produces unweighted events) and Sherpa (which produces weighted events). Later on, once the parameters associated to the method was optimized, the *FastME* was applied to the real CMS data from LHC RunI coming from the 2015 HZZ4L analysis (the same datasets used by the author during his master). Differing from the MC generated by the author, such CMS data includes the whole CMS detector simulation and effects of PU. As it will be shown, the method didn't have significant loss of performance in that data.

The results achieved through the *FastME* applied to the generated MC events were presented in the XXVII *International Symposium on Lepton Photon Interactions at High Energies* [118]. The development of the method has stopped in a stage where the codes created by the author were organized into a package that can be downloaded from *Github* [120] and compiled inside a proper **cmssw** release (similar to any of the packages developed and used by the CMS collaboration). Unfortunately, the development of *FastME* was abandoned. First, because the method did not show the expected discrimination power when applied to separate VBF and ggH (that was the focus of this PhD) and second (the actual definitive reason) was the publication of a similar method by the TMVA team [119]. Although that happened, the amount of developed work produced interesting results that without any doubt are worthy to be included in this thesis.

F.2 Description of the *FastME*

The goal of the *FastME* was to be a fast and efficient discriminating method of real events by using MC events topology. Its implementation is based on an idea started originally by Dr. Andre Sznajder and Dr. Stephen Mrenna (Computing and Simulation Division, FNAL). The theoretic motivation behind the *FastME* was verify the capacity of retrieving the Matrix Element (ME) from MC events, which are produced through the MC generators by computing ME of the physical processes. The advantage of such approach, instead of using traditional methods as the *Matrix Element Likelihood Approach*, comes from the possibility of skipping the computing latency and the possibility

Table 30 - Time to compute the weight of one event using *MadWeight5*. For an usual analysis these numbers multiplies by thousand.

Process	Time/Event (s)
ZH	<5
$t\bar{t}$ fully-leptonic	10
Zbb	18
$t\bar{t}$ semi-leptonic	41
$t\bar{t}H$ fully-leptonic	60

Source: Reference [121], p. 7.

of using informations not implemented on them yet, as higher order corrections (NLO, NNLO, ...), for instance.

Nowadays several discriminative methods are based on the EM. Such methods assign to events their probability to be compatible with a certain theoretical model via the Eq. 77, where

$$\mathcal{P}(x|\alpha) = \frac{1}{\sigma_\alpha} \int d\omega_1 d\omega_2 f(\omega_1) f(\omega_2) \int d\Phi(y) |\mathcal{M}_\alpha(y)|^2 W(x, y) \quad (77)$$

- σ_α : process cross-section;
- $d\Phi$: kinematic acceptance of detector/analysis;
- $f(\omega_i)$: protons probability density function (PDF), where ω_i is a momentum fraction of incident proton;
- $\mathcal{M}_\alpha(y)$: matrix element, where y is the representation of all kinematic variables describing the process;
- $W(x, y)$: transference function, which describes the probability of an x state be measured as y (accounts for the expected detector effects - extracted from simulation);

The integration of the EM is not trivial due to the presence of singular functions (Breit-Wigner's and transfer functions), which forces specific parameterizations of the phase space (which is not possible for all physical processes). Furthermore, the method of EM is a technique that depends on the theoretical model adopted. Another issue is the presence of MET which forces the integration of the EM over unknown degrees of freedom. Also, although the method of EM maximizes the quantity of theoretical information extracted from the observed data and not relies on pre-training, the computation of Eq.77 can take much time (Tab. 30). And a last note is that the current EM methods can only make EM's at LO [121].

The original scheme for the *FastME* was similar to the EM method: one wanted to estimate the weight for a data event by association to MC events. However, this scheme has shown low performance and usually the discriminant distributions for signal and background had few separation. Because of that, the development of the studies was focused into a scheme in which the association between real data and MC events is given by a distance measurement. The metric for such distance was developed empirically through the studies and is given by Eq. 78 (a χ^2 -type),

$$R_{(i,j)}^2 = \sum_{k=1}^n \left(\frac{\delta v_k}{\sigma_{v_k}} \right)^2, \quad \text{with } \delta v_k = v_k^{(i,MC)} - v_k^{(j,Data)} \quad (\text{FastME metric}) \quad (78)$$

where, $R_{(i,j)}$ returns the distance between the particles i (from MC event) and j (from Data event) computing the difference between the variables v_k , which were in the beginning p_T , η and ϕ .

The Eq. 79 gives the distance between events (data-MC) by summing the smallest $R_{(i,j)}$ found for each particle i , with the constraint that j for a given $R_{(i,j)}$ can not be used for other pairs¹³. The σ_v^2 factor was thought to be a parameter to include the uncertainty associated to the corresponding variable (its resolution). In the carried studies, however, σ was initially kept equal to one and later on received a value to scale the contribution of each term in the sum (it will be discussed later). Until the last studies with *FastME* the uncertainties were not included.

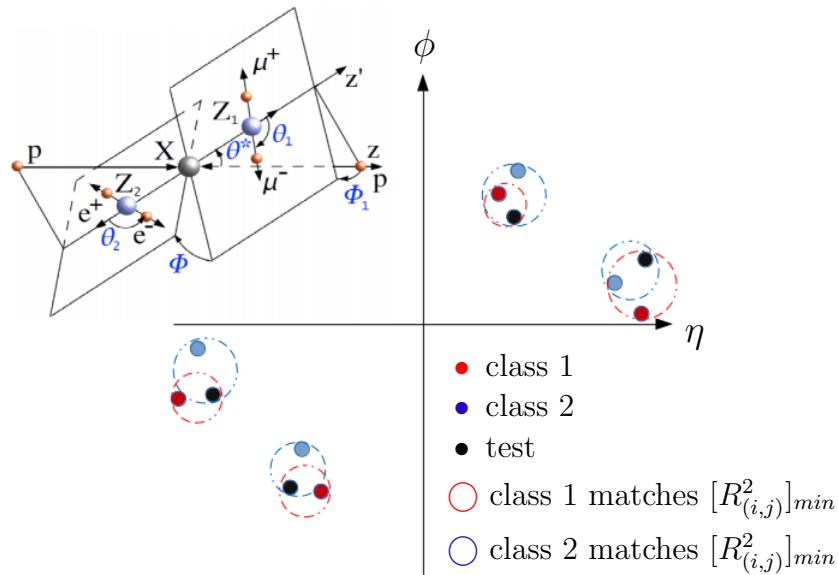
The reason why the distances between the particles are computed towards the MC particles (and not the data) is due to the possibility of find proper final particles in a data sample where one has more particles than it is required. That is, one can use the MC (which has been previously filtered through an analysis selections) to select the proper data particles.

$$D^2 = \sum_{i=1}^m [R_{(i,j)}^2]_{min}, \quad \text{with } j(i+1) \neq j(i) \quad (79)$$

In such scheme the sample of MC events serves as a pattern bank and each data event can be compared to the patterns in order to compute a compatibility between the

¹³ The author calls the attention here for a problem in this algorithm that have not been noticed during the development of the method. In the $R_{(i,j)}$ calculation the search for the smallest values is done towards the particle i and that makes impossible that the particle j can be used as pair for the remaining particles ($i+1, i+2, \dots$) in the event. In order to correct that, the correct algorithm would be to build a matrix with all the $R_{(i,j)}$ values and then find the global minimum (in such way, the final pairing order (i,j) might not follow the particles i).

Figure 84 - Illustration of the analysis process done by the *FastME*. The MC events present a topology associated to the EM of a given physical process, such as, each particle has a correlation with the others particles in the event. A data event (black points) receives a probability of being from a kind or other (blue and red point) via the correlation of the distances (represented by the blue and red circles) between it and the MC events.



Source: The AUTHOR, 2018

observed event and the simulated pattern. The compatibility is computed by a discriminant given by Eq. 80 where D_{Min}^{Bkg} (D_{Min}^{Sig}) is the smallest event distance (given by Eq. 79) between a data event and a signal (background) MC event (here signal and background can be a group of processes together constituting two classes one wishes to separate).

$$P_{SB}^D = \frac{D_{Min}^{Bkg}}{D_{Min}^{Bkg} + D_{Min}^{Sig}} \quad (80)$$

The working structure of the *FastME* is sketched in Fig. 84 and can be summarized in the following way:

- 1 Pair each data particle to a MC particle of a given MC sample using Eq. 78;
- 2 Compute the event distance using those particle distances as given by Eq. 79;
- 3 Steps 1 and 2 are made for different MC processes. Then compute the probability of the data event belongs to class *sig* or *bkg* accordingly to Eq. 80.

By definition, P_{SB} has values in the range $[0, 1]$ and one could intuitively think that an ideal threshold is $P_{SB} = 0.5$. However, as it will be shown in the simulation studies,

the distribution of that variable usually presents asymmetry and depends on the MC's being used. This can be an effect of the phase space population since this is one strong point that introduces bias to this method based on distance. This effect was studied and it will be discussed in the next sections.

Additionally to the discriminant defined by Eq. 80, which is based on distance between the events, there was a discussion also about another ways during the development os this project. As mentioned before, the original motivation of the *FastME* idea was the possibility to retrieve the weight of events by association with MC samples. So, one of the first discriminants thought, after the success of method based on distance, was the usage of the weight of the MC event in place of the distance (Eq. 79). That is, the form of the discriminant defined by Eq. 80 does not change, but the terms D are replaced by the weight of the associated MC events. So, the form of the second discriminant is

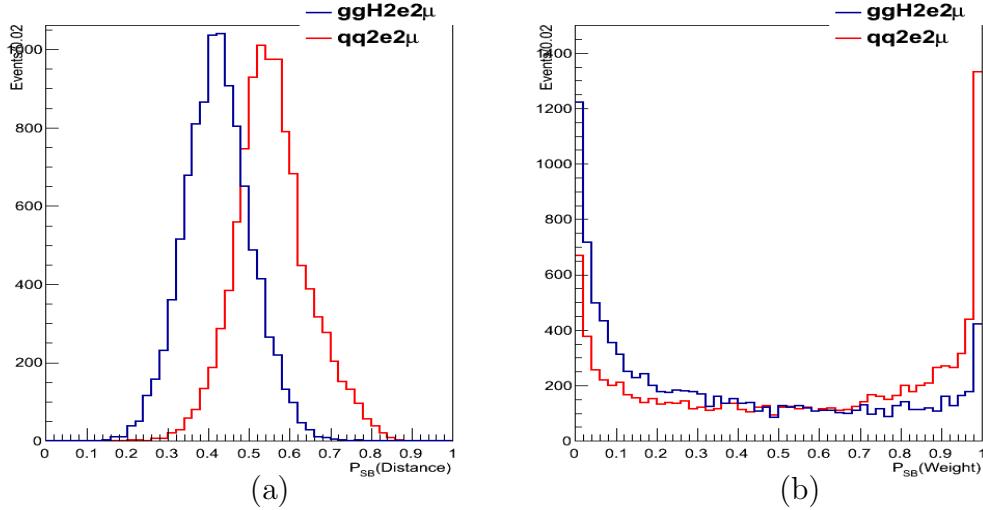
$$P_{SB}^W = \frac{W_{D_{min}}^{Bkg}}{W_{D_{min}}^{Bkg} + W_{D_{min}}^{Sig}} \quad (81)$$

where $W_{D_{min}}^{Sig}$ ($W_{D_{min}}^{Bkg}$) stands for the weight associated to the closest signal (background) MC event matched to an probe event. The shape of the discriminants P_{SB}^D and P_{SB}^W are quite differentiable and their performance have shown to be significantly different. The method based on weight had the main issue of cost too much the signal efficiency [118], for instance. Fig. 85 shows the shape of such discriminants as obtained with private samples of $gg \rightarrow H \rightarrow 2e2\mu$ and $qq \rightarrow ZZ \rightarrow 2e2\mu$.

F.3 Studies and Results

Several studies for comprehension and optimization of *FastME* were made. As mentioned before, the base of the method are MC events to work as pattern bank. In order to make studies each MC sample was divided into two sub-samples: one for work as pattern bank and another to work as testing events. The first studies investigated the importance of the variables to be used in Eq. 78. Originally, the studies started with the four-momentum of the particles. Then, it was noticed that the particle energy could be omitted without loss of performance. So, it was kept only p_T , η and ϕ , which performances will be presented in the sequence.

Figure 85 - Comparison between the *FastME* discriminants distributions as defined by Eq. 80 (a) and 81 (b), obtained samples of $gg \rightarrow H \rightarrow 2e2\mu$ and $qq \rightarrow 2e2\mu$ (ZZ) generated with Powheg.



Source: The AUTHOR, 2018.

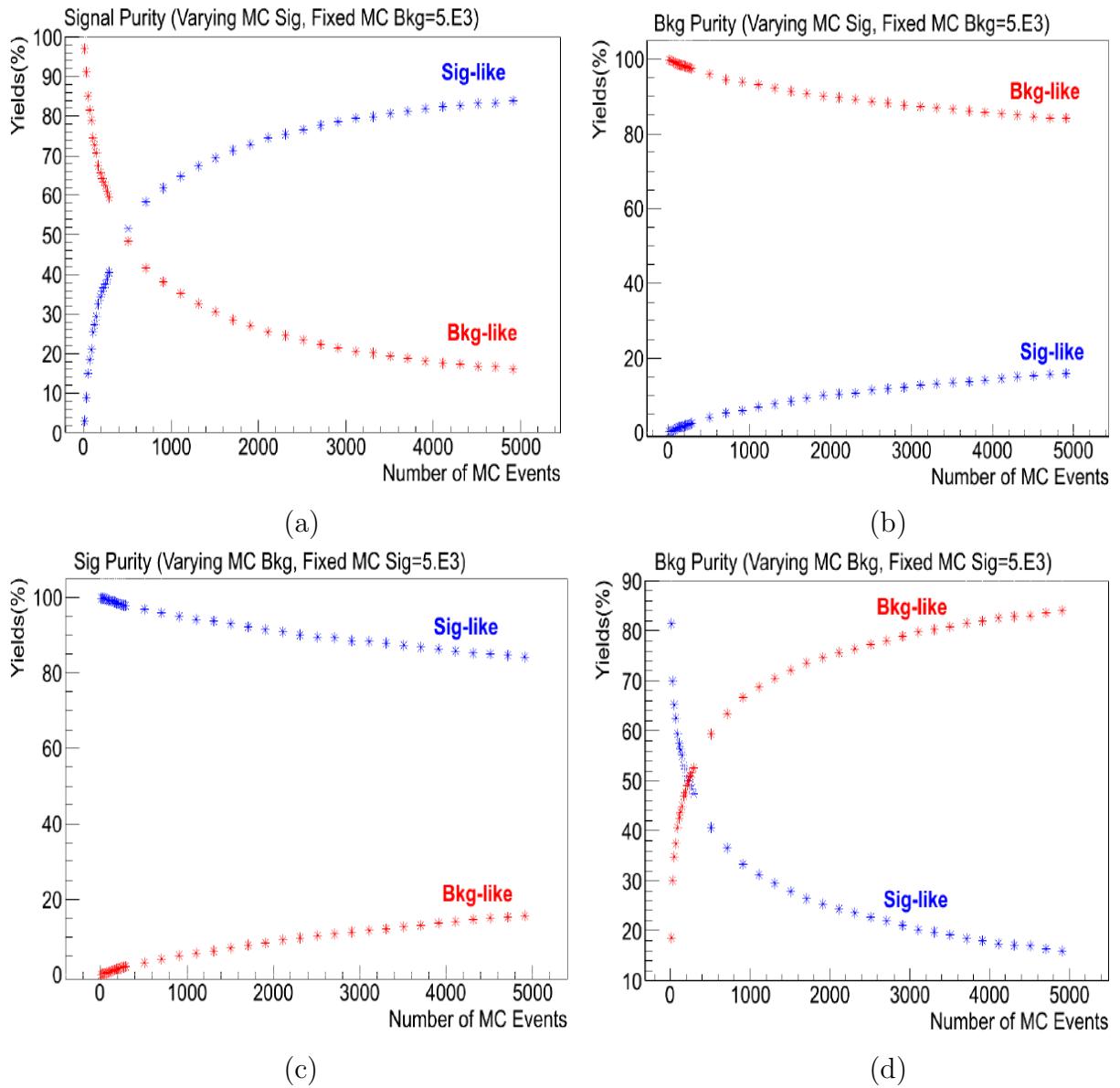
F.3.1 Bias due to Pattern Bank Size

Once set those three variables as the ones for the method, the next step was to study the dependence between the discriminant and the size of the pattern banks. As one can intuitively think, the probability of an event be classified (by the *FastME* method) as type A or B is associated to the size of the MC samples, which determines the quality of phase-space coverage and resolution.

The strategy adopted to verify the impact of the size of the pattern banks in the discrimination given by the method was to map the purity variation for signal and background, as shown in Fig. 86. The signal here is the ggHZZ4L and the background is the qqZZ4L processes. Note that, what is called "purity" in that plot is actually the efficiency, that is, the absolute number os events os signal (background) being correctly selected as signal (background).

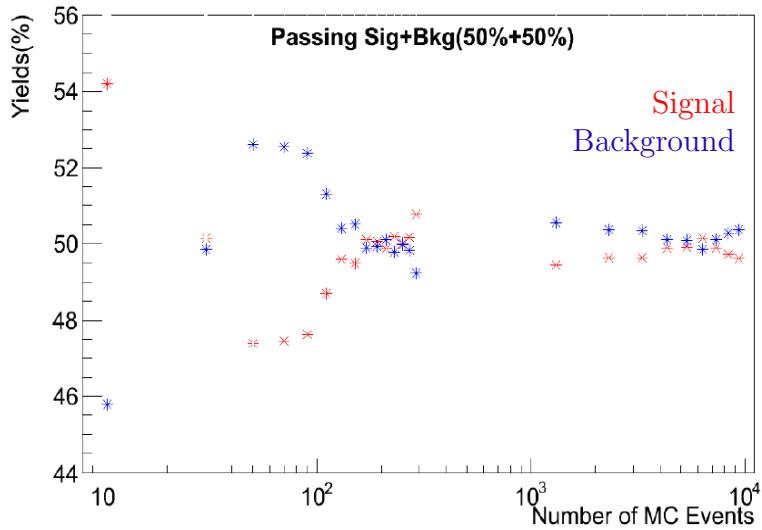
Focusing on Fig. 86(a) one sees that a small pattern bank for the signal leads to a high classification error rate, such that, almost all events are classified as background. As the size of the pattern bank starts to increase also the fraction of signal events being correctly classified increases. Looking at graph on Fig. 86(b) one sees what happens to the background events. As the size of the signal pattern bank increases the fraction of background events being mis-classified as signal also increases but much more slowly. In Fig. 86(c) and (d) is shown what happens when the size of the signal pattern bank is kept fixed and the background one is varied. A similar behavior as described before for the signal is seen now for the background (what of course is expected).

Figure 86 - Classification dependency for signal and background events accordingly to the size of the pattern banks used. On graphs (a) and (b) the size of the background pattern bank has a fixed size while the signal one is varied. On graphs (c) and (d) the opposite case is shown. Note, here "purity" is computed using the absolute number of events (without normalization).



Source: The AUTHOR, 2018

Figure 87 - Variação da fração de signal e *background* identificados na amostra com 50% de cada classe.

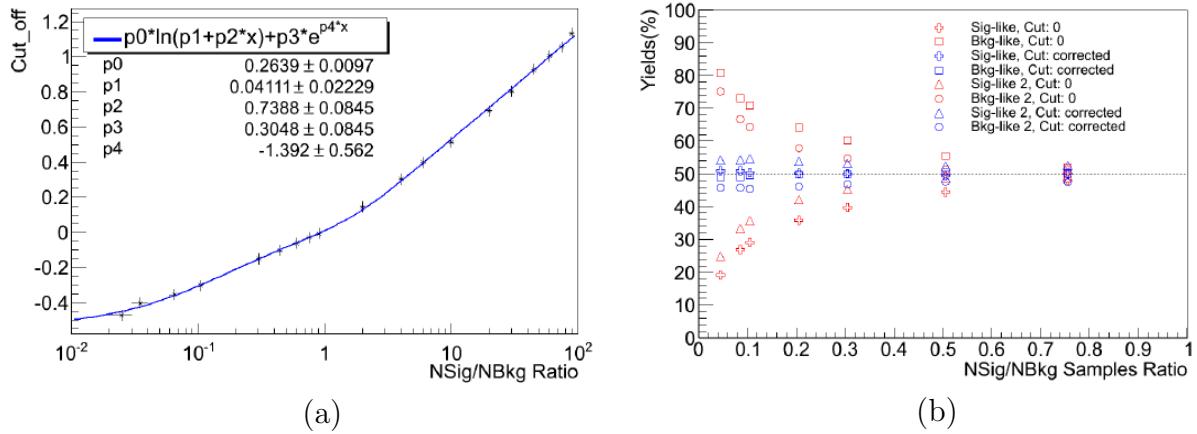


Source: The AUTHOR, 2018.

Completing the study on the pattern bank size dependency a test sample composed of 50% signal and background was prepared ($1.e^4$ events from each). Then the same study was repeated to check the impact of the bank size on the classification. As it is shown in Fig. 87 the dependency is not trivial, such that is not possible to point the ideal size, since the classification error increased again for the bank size of 400. Fig. 87 also points a trending: the classification error gets smaller as the bank size increase. Note that the increasing observed after the decreasing is sequentially smaller, that is, although it can increase again the sub-sequent increasing is not bigger than the previous one. From this behavior it was decided that an "ideal" minimum bank size should be $1.e^3$ events, which generates for those samples an average classification error $< 2\%$ (note that in that study the signal and backgrounds had the same size). It's obvious, though, that this conclusion can not be generalized to other processes, since the population of the phase space by the events can easily be different.

Based on that dependency between the classification error and the bank size a correction function was studied just as an extra complementation of the study. The correction was derived by finding the dependency between the ideal cut on the discriminant (which gives the correct classification of each class) and the ratio between the banks (signal/background) size. The ideal cuts and the found correction function is shown on Fig. 88(a) and the effect of this correction on the classification of the events are shown in Fig. 88(b). Note (again), the derived function would be need to found for each particular case. Also, it is interesting to notice that there are essentially two regimes for the correction ($sig < bkg$ and $sig > bkg$) and they approximately present a linear dependency.

Figure 88 - (a) Scan of the ideal cut (to correct classify signal and background events) in function of the ratio between the size of the signal and background pattern banks. (b) The effect on the fraction of signal and background events being classified as so, before and after the correction on the discriminant.



Source: The AUTHOR, 2018.

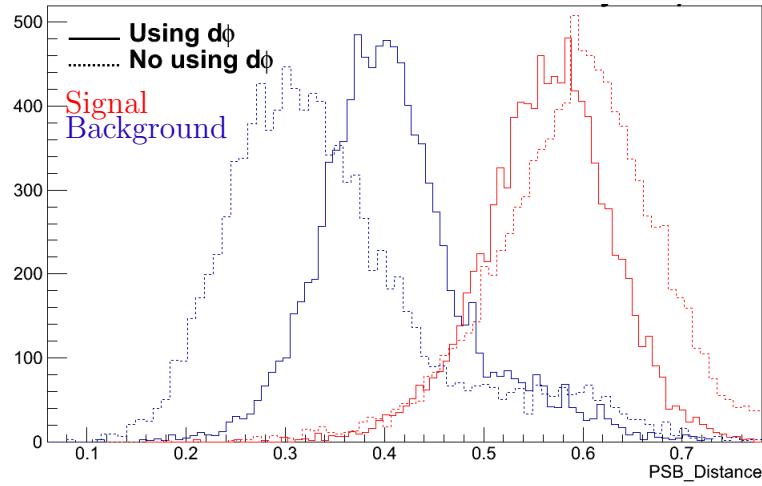
F.3.2 Impact of ϕ on the Discriminant

Between the developed studies it was noticed that the contribution coming from the variable ϕ is not significant to the discriminant or it can even degrade its performance, which can be seen in Fig. 89. This result indicated that just the variables p_T and η are enough to provide good discrimination power through the method present here. Because of that, the posterior developments and studies were drove toward this principle and the variables p_T and η became the standard ones to be used, while ϕ has been kept as an optional parameter (which can be used by calling it in the *FastME* framework configuration card). Similar to ϕ , the energy (E) is a variable that did not present a substantial contribution to the discriminant performance and that is why it is not used too. Although plots have not been saved, it is mentioned here for information of the reader.

F.3.3 Scaling of the Metric Terms

The metric defined by Eq. 78 allows the usage of variables with big scale variations. While η and ϕ are usually limited by the analysis acceptance and geometrical detector definitions ($|\eta| < 4.7$ and $|\phi| \leq \pi$), the same does not happen to p_T (which can be up to two orders of magnitude higher). So, it is evident the needing of a scale factor to weight the contribution of each term in Eq. 78. Note that, this factor would be combined with the uncertainty σ on the measure property. The study of how that combination should

Figure 89 - Impact on the D (Eq. 79) distribution for signal and background due to use or not of ϕ in the computation of the distances between particles and events.



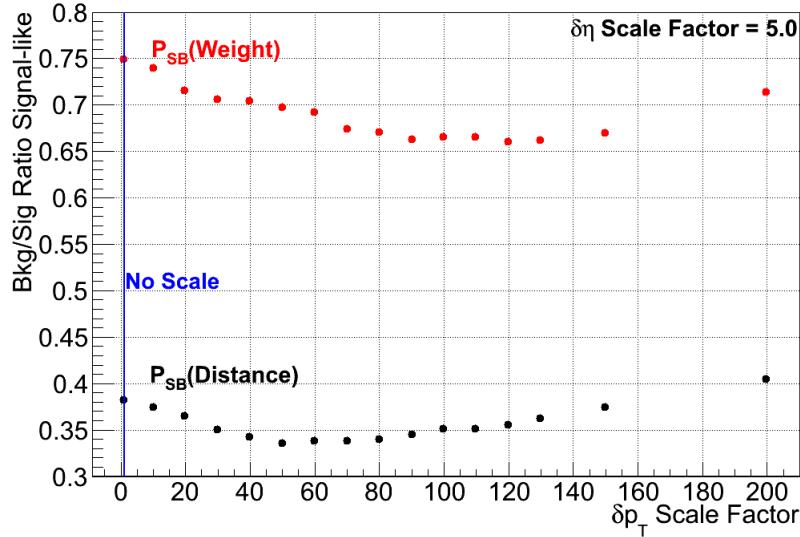
Source: The AUTHOR, 2018.

be done was not made, since the project was stopped before such step. The study of a scaling factor for p_T , though, was carried out and it is discussed here.

In order to determine an adequate value to scale the δp_T in Eq. 78 a fixed value of 5.0 (because it is $\sim 2_{leptons}$) was chosen to scale the $\delta\eta$. Then, using the already mentioned MC samples, the discriminant was computed for several values of scale factors for p_T . The effect on its performance in the classification of a signal MC sample is shown in Fig. 90. That plot shows the ratio between the amount of background being (wrongly) classified as signal and the amount of events correctly assigned as signal. As one can see, the best value in that case was around 50. Fig. 90 also shows the performance of the two methods developed: the one based on the distance - $P_{SB}(Distance)$ - and the one based on the event weight - $P_{SB}(Weight)$ - (which has much lower performance as already said).

Later it was noticed that the cumulative p_T distribution of the MC particles has the peak around 50GeV. With the idea of automatize the way to find the scale factor, that was seen as a correlation and it was introduced in the *FastME* codes a routine to compute the scale factor from the cumulative p_T distribution of the particles in the MC samples. Two procedures were implemented: the arithmetic average between the particles p_T and the value of p_T at the distribution peak. Similarly, a procedure was introduced for η which takes either the maximum value among all particles. Maybe instead of a fixed scale factor, a parameterized scale factor (as function of p_T and/or η) could work better but such studies were not produced before the stop of the project.

Figure 90 - Dependency on the variation of the fraction of signal events being mis-classified as background in function of the δp_T scale factor (the σ_{v_k} in Eq. 78). A fixed value of 5.0 was used as the $\delta\eta$ scale factor.



Source: Source: The AUTHOR, 2018.

F.4 *FastME application on the HZZ4L Data (CMS RunI, 2015)*

In this chapter the results obtained by applying the *FastME* on the official HZZ4L 2015 (RunI) CMS data will be discussed. These are the most expressive results of the method and the ones that gave the motivation to format the standalone codes developed into a package. The results presented here shows the feasibility of the method, since so far, it had not been applied to MC samples with full simulation (which contains PU and detector effects).

The events from 2015 account for $24.8 fb^{-1}$ coming from collisions at $\sqrt{s} = 7$ and 8TeV. Just as a remind, the biggest background to the SM Higgs is the ZZ . The background $Z + X$, as already discussed previously, is estimated by the a *data-driven* method and was the second biggest background in RunI. For simplification, were taken only the Higgs signal and the ZZ background MC samples (since, also, that the $Z + X$ contributed with just 2 events at the Higgs signal region). The observed data was also used in order to validate the *FastME* distributions from MC. The MC samples of signal and background were split in order to have two sets of each source, which were used as the pattern banks and the testing events.

The Fig. 91 shows the performance of the *FastME* compared to the *MELA* CMS discriminant. Note, that is another discriminant (not the *VBF* one) from the *MELA* method and it is dedicated to separate SM Higgs from the ZZ background. As one can see, the *FastME* method works well for this case and even presents a better performance

the *MELA* discriminant at that time. On Fig. 91(a) is shown the m_{4l} distribution for observed events as classified by the *FastME* like signal (red) and background (blue) - the stars just show the total distribution. On Fig. 91(c) and (d) the distributions for MC and the observed events from *FastME* and *MELA* are shown. The MC normalization is arbitrary there because those were preliminary studies and there were missing some proper statistical treatment of the events (note that is also missing the error bars from the observed data points). Fig. 91(e) and (f) shows the 2D distribution of each discriminant versus the m_{4l} mass. Again it is clear a better separation between the signal and background MC events through the *FastME*.

F.5 The *FastME* Package

In the previous sections the discussion was restricted to the bases of the *FastME* idea, which are the core of the algorithms developed. In this section a more technical description of the package that has been created to hold the project will be discussed.

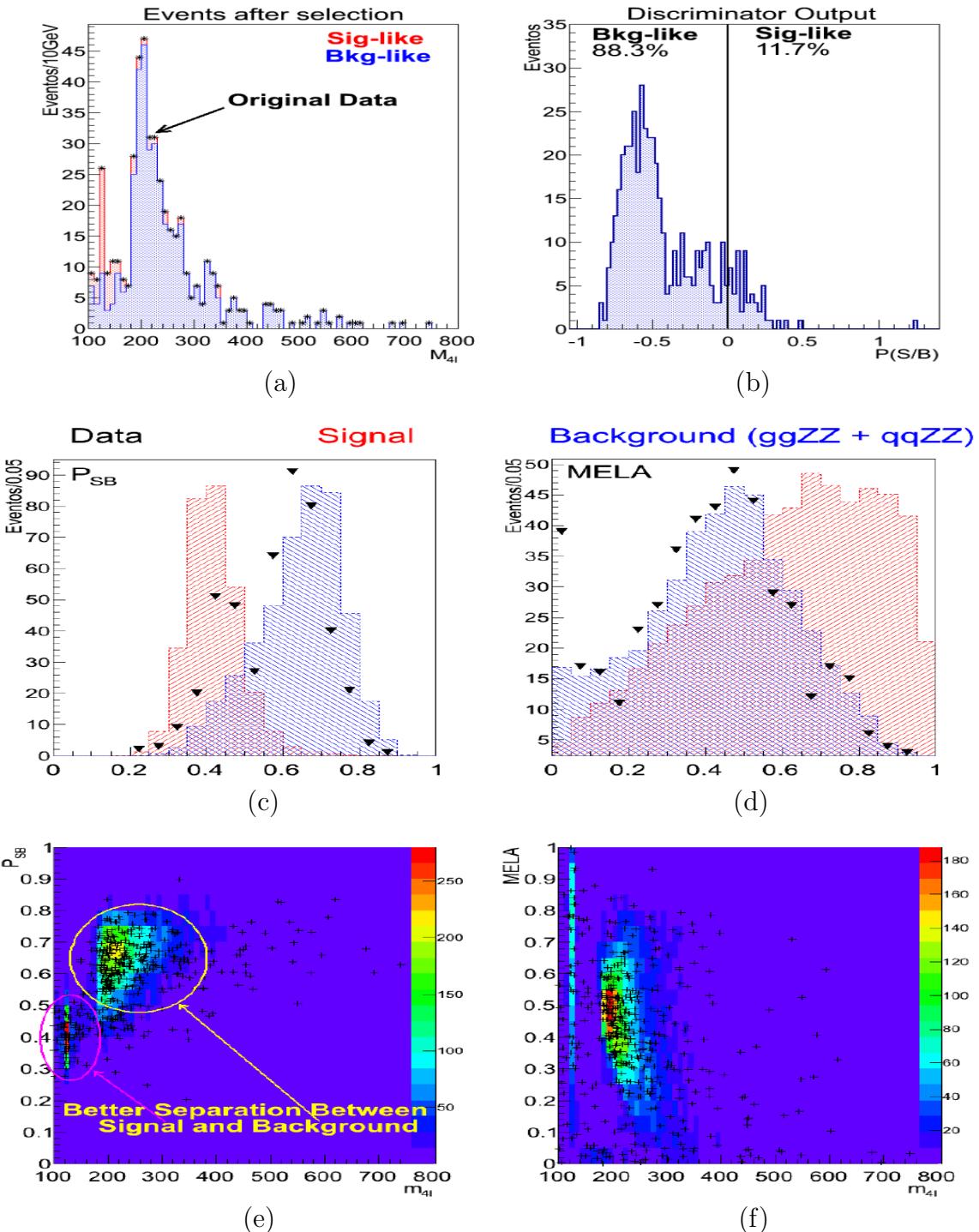
F.5.1 Features of the Package

After the performance obtained with the official HZZ4L events from CMS RunI there was a motivation to create a package that could make more flexible the usage of the *FastME* idea and allow further improvements. The goal was the possibility to use it inside a **cmssw** release as it is for several packages used within the CMS collaboration. The main part of development of such package was carried out during the first stay of the author at Fermilab as a visiting-scientist, which was mainly supported by another researcher from the FNAL-CMS-L1TT working group. In that time the author also had some few discussions with Dr. Stephen Mrenna.

The standalone codes were organized in such a way that each processing part has its own separated routine and all of them are actioned by a central executable that drives the flow of an analysis with the *FastME* method. Several enhancement have been done after that and the most important was the migration of the codes to a version capable to parallelize the step of comparison between data and MC events (since one of the proposals were a short time analysis). That was done by using the *c++* class *TProcPool* available in ROOT (6.0 or newer versions), which allows one to easily manipulate the CPU cores in a computer. Additionally to that, a command to find out the number of available CPU cores in a computer was implemented in the package, so that, the user can optimize its analysis timing.

The *FastME* package in its latest version can be found at the *GitHub* repository

Figure 91 - (a) Observed events classified as signal and background by the *FastME*, (b) the discriminant distribution (from observed data), (c) and (d) distribution of *FastME* and *MELA* discriminants as obtained from observed data and MC events, (e) and (f) the distribution of those discriminants versus the m_{4l} mass value of the analyzed events. The distributions are not normalized by the events weight. The distribution of *FastME* discriminant in (b) looks different from what is shown in (c) and in Fig. 89 because that was the original shape. Later it was decided to re-define the discriminant as expressed by Eq. 80 in order to keep it in the range [0, 1].



Source: The AUTHOR, 2015.

Figure 92 - View of the *FastME* repository on *GitHub* created by the author.

mamelodea	Update README	Latest commit d31486c on Sep 22
<code>FastMatrixElement</code>	Updating composer structure	9 months ago
<code>tests</code>	Example of root file from event generation feature	10 months ago
<code>README.md</code>	Update README	2 months ago
<code>SetUpEnvironment.sh</code>	Automatic setup	a year ago

`README.md`

FastME

Probing and Discriminating Data by MC Topology (stable & updated version of the project)

Fast Matrix Element main goal is to achieve similar capability and full physic information as the many types of Matrix Element method available today. However, FastME try to reduce the computing time needed to get such info. The Monte Carlo events already contain the full physic information (that can be in higher orders of parameters correction) about the related event and FastME uses these events as inputs to compute in a fast way the nature (signal/background) of a specific event (a data event, for instance). The process relies into compare de available parameters (pT and η) between the event and the MC. Such comparison allows one to compute the distance between an event and each MC in that phase space.

Source: The AUTHOR, 2018.

of the author [120] (Fig. 92). Some few modifications and studies were done after the stop of the project and they are not completely synchronized. On the main page of the repository one finds informations about the method and the main commands needed to run the package, including an image of it running a standard analysis flow. A *shell-script* is provided inside the package to the user as a way to quickly set it up within a **cmssw** release. The code essentially creates the adequate release for a **gcc4.9.3** compiler and then, compiles the package. After concluded the installation, the used will have available the executable **fastme**, which usage will be further explained in the next sections.

F.5.2 The Package Workflow

The *FastME* package is divided essentially in six modules that execute different tasks, which might or might be not sequentially dependent. Such modules and their properties are:

- **Cortana:** drives the interaction between the user and the package. Produces the object **FmeSetup** which contains all the configuration parameters specified by the user in a standard format for the package. It also gives to the user the available

options in the package (as the number of available CPU cores, for instance);

- **Librarian:** identifies and replicates the input data to be analyzed (observed data or MC), producing another files that are used by another modules (this is a security procedure, the *FastME* does not run the analysis over the original user files);
- **Cartographer:** it is the main module, responsible for the algorithms of particle/event pairing (that is, it computes the Eq(s). 78 and 79). It produces a ROOT file containing the information of the closest events from each class (signal/background). The parallelization happens in this module (if required by the user);
- **Scaler:** computes the scale factor, as described in Sec. F.3.3, that normalizes the δv_k 's (Eq. 78). It is used by the **Cartographer**;
- **Arbiter:** it is the module where the discriminants are computed (Eq. 80 and 81). It receives as input the ROOT file produced by the *Cartographer*;
- **Composer:** this was the latest module to be developed and its implementation have not been concluded. Its purpose was the generation of events based on the topology give be the MC events in the pattern banks. The goal was to find an alternative way to increase the discrimination power of the method. The algorithm for it consists of generating events with randomly v_k 's and filter such events using the discriminant. However, until the stop of the project it was not found expressive results.

The modules described above are centralized and controlled by the executable **fastme**. This executable is responsible for receive, and transmit to the modules, the commands and parameters given by the user for the analysis. The commands are passed via command-line and the parameters are given via a configuration file called **fme_config.dat**, which is present in the package. The details about the parameters presents in that file can be seen in Sec F.5.3. For a complete analysis in the *FastME* the user needs to run the executable **fastme** twice. The first run to launch the **Cartographer** receiving the input data and producing the output file that will be used by the **Arbiter**, which will finally produce the analysis results.

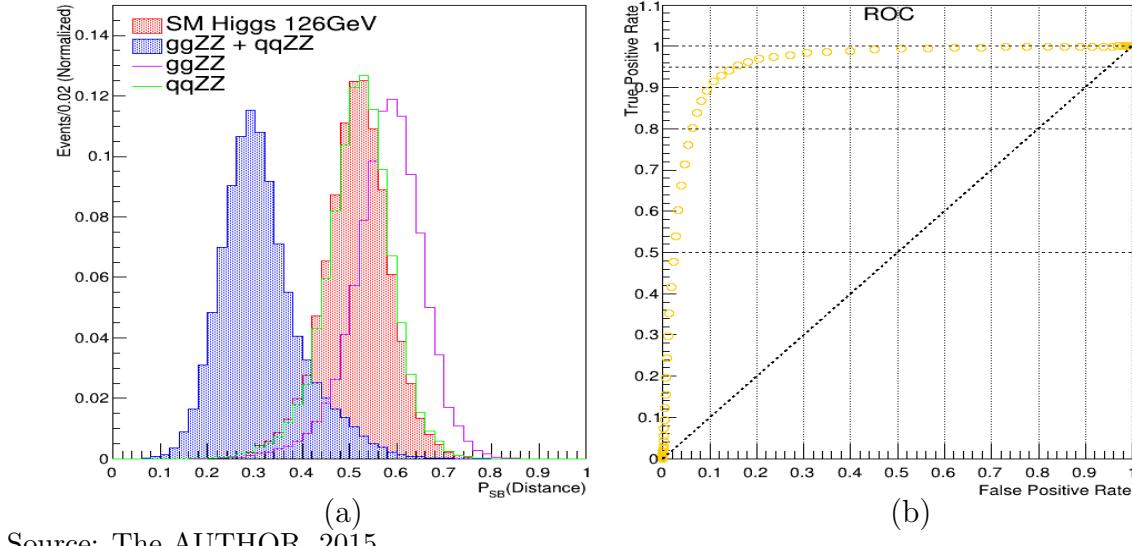
F.5.3 The *FastME* Configuration Card

Fig. 93 view of the configuration card file that an user needs to give to *FastME* in order to run an analysis. For more details on the parameters specified in the card file, see [120].

Figure 93 - *FastME* input card. This card is parsed to the *FastME* executable in order to set up needed parameters in order to run an analysis.

Source: The AUTHOR, 2015.

Figure 94 - Comparison between the two methods developed to define the classification of a event probe. On (a) the discriminant distributions for the minimum distance matching and on (b) its performance.



Source: The AUTHOR, 2015.

F.5.4 The results provided by the *FastME* package

Once the full analysis in the *FastME* has been ran, the user will have some plots which gives the information about the performance of the method on the given dataset. Between these plots the most important are the discriminant distribution for signal and background (it was not developed a feature for superimposing the observed data distribution to the MC). Here is an example of the plots (Fig. 94) that are produced by the package. On (a) one sees the discriminant distribution coming from the *FastME* method based on the minimum events distance and on (b) it is shown its performance by usage of a ROC curve.

F.5.5 Data Filtering by *FastME*

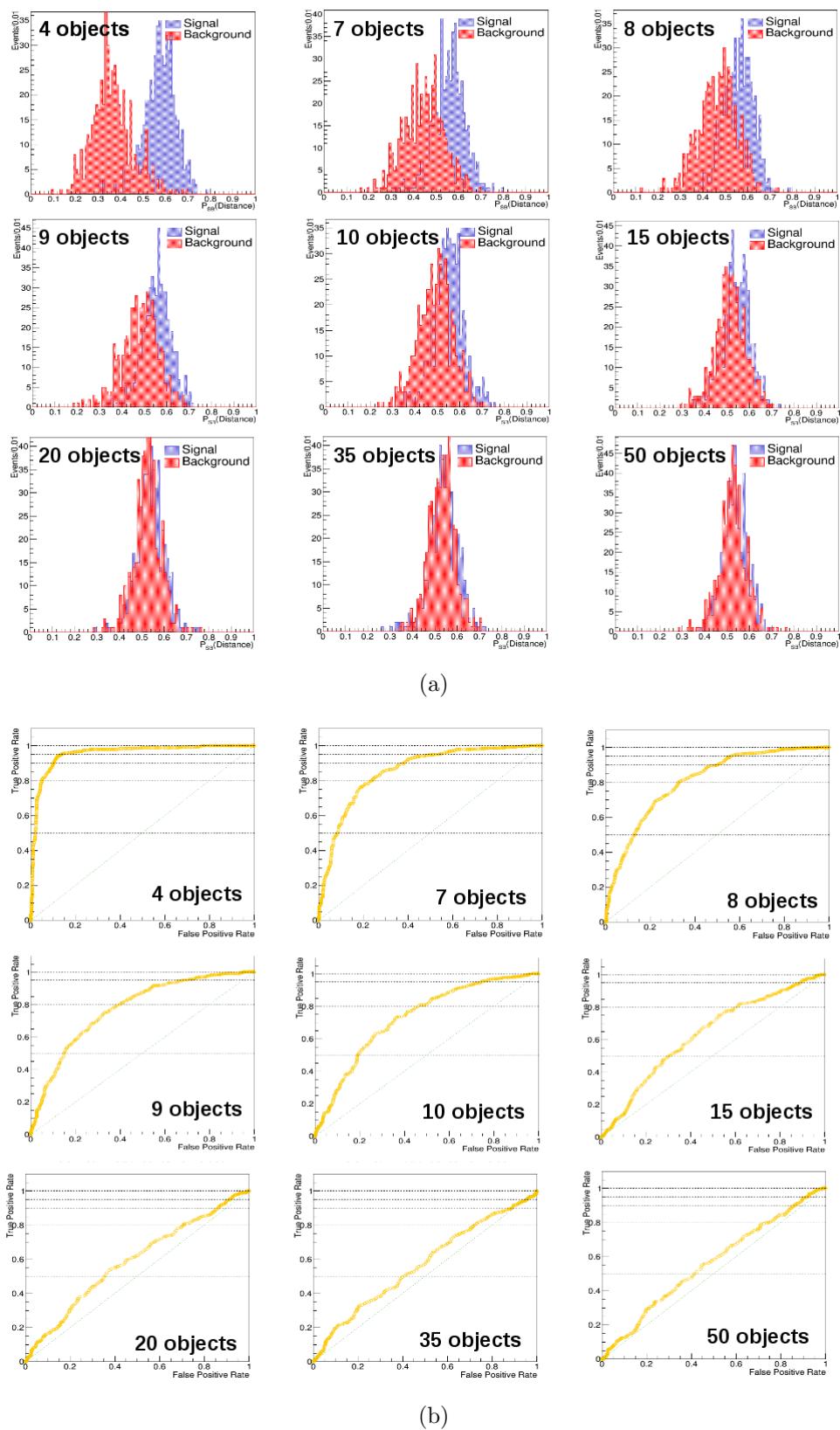
Another idea that came later for the usage of the *FastME* was the particle selection based on the MC. This selection is different from what have been presented so far in the sense that, originally, the *FastME* was planed to be used on final data, that is, after a full-selection analysis. However, as the core of the method is based on the idea of matching a given particle from a data event probe to a given particle from a MC event, the straight derivative idea is the possibility of filter out such data probe particles through the same principles. So, the idea here was to test and see how the method works when one has more particles in data than it is expected from the MC.

In order to test this situation, random values of p_T , η and ϕ (constrained within the expected range of such variable) were introduced in the sample of observed events. Here, the tests were done for the case without jets and only the four leptons coming from the Higgs decay have been used (the same samples presented before). The Fig. 95 shows how the discriminant behaves under such situation, where it was tested the presence of up to 46 extra objects in the data event (there were always the selected four leptons of the probing event). As one can see, the method is sensitive to the presence of extra particles and its performance decreases fast the number of those particles. But, it seems to work reasonably up to 4 extra particles. Note that in those tests the size of the bank was constrained to 5k. With a larger pattern bank these performances could increase.

F.5.6 The *FastME* Applied to qqH vs. ggH

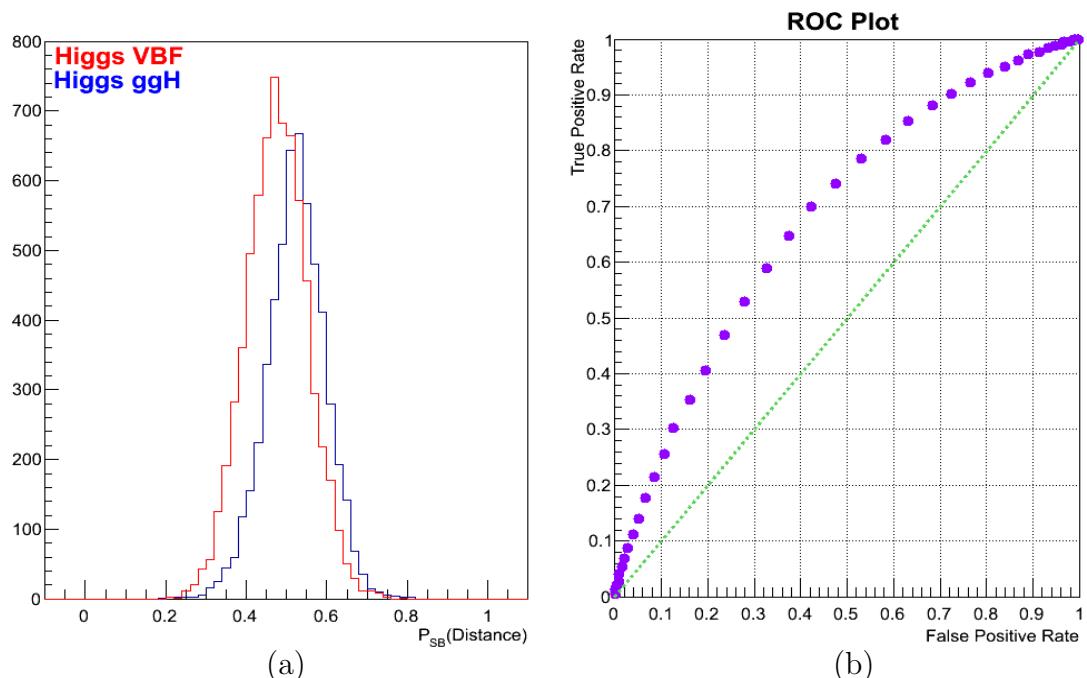
The final topic of study involving the *FastME* was the attempt of VBF discrimination against gluon-fusion. The main subject of the present thesis is the measurement of Higgs VBF production XS. So, it was natural to think about applying the *FastME* to the VBF selection. In this case, the inputs were the four leptons and two jets p_T , η and ϕ . The Fig. 96 shows the outcome found when trying to discriminate the VBF process against the gluon-fusion. The overlap between the discriminant distributions from each process is visibly large and its performance, as it is shown by the ROC curve, is very low (AUC $\sim 66\%$). Here it is shown only the performance from the method based on the distance, since, as already said on previous sections, this method was the one presenting highest performance.

Figure 95 - Variation on the *FastME* performance due the introduction of extra objects with random properties in the signal and background testing samples. In (a) the variation on the signal and background distributions for the discriminant (based on distance) and in (b) the variation on the ROC curve.



Source: The AUTHOR, 2016.

Figure 96 - *FastME* performance on the discrimination of qqH against ggH using the discriminant based on the distance (left). The area under the ROC curve (right) corresponds to $\sim 66\%$.



Source: The AUTHOR, 2016.

G PARTICIPATION AT THE GLOBAL DOC PROJECT

During his PhD the author of the present work spent a five month interchange period in Italy, at the Università degli Studi di Bari Aldo Moro (UNIBA) located in Bari (Puglia region). The interchange was funded by the UNIBA through the Global Doc project, which was developed to promote the local interaction of PhD students, from several countries around the world and different research fields, with the UNIBA local community. This interchange allowed the author to be very close to his co-advisor since the Physics department of UNIBA (Dipartimento Interateneo di Fisica Michelangelo Merlin) sits at the same campus where is located the INFN and Politecnico di Bari. This period abroad also allowed the author to personally participate of two very good events: the Electroweak Symmetry Breaking Spring school (EWSB)¹⁴ and the Monte Carlo net school (MCnet)¹⁵.

The period spent at UNIBA was also very productive for the author. Differently from the period spent at Fermilab as visiting scientist, the Global Doc project promoted the interaction of the author with several (PhD, Master and Grad) students (even another Brazilians) from different research fields. This is a good scenario for interchange of ideas, which finally achieved his apex during the Global Doc Seminars section. In this seminar each student enrolled in the project had to give a presentation about his (her) research field, allowing the students to expose their work and receive questions from very different view angles. Fig. 97 shows some picture from the seminar given by the author at UNIBA.

In terms of the Physics, the period spent at UNIBA brought good experiences for the author. One of the main impacts was the development of the strategy to estimate the systematic uncertainties associated to the VBF ANN discriminants due to their inputs. That demanded the re-processing of MC samples used in this present analysis which was possible by using the computer farm RECAS¹⁶ at UNIBA Fig. 98. This procedure has been also important for another analysis about Dark Matter leaded by the author's co-advisor. Because of that, the author spent also some time in that analysis contributing on the final systematic uncertainties estimation and on the statistical analysis through the advanced features contained into Higgs Combine tool. This was a good experience for the author which applied similar studies to work presented in this thesis.

This Dark Matter analysis is essentially based on the search for Dark Matter candidates associated to Higgs production mechanisms (Higgsstrahlung) as shown in Fig. 99(a).

¹⁴ EWSB school: [\(https://indico.cern.ch/event/673580/\)](https://indico.cern.ch/event/673580/).

¹⁵ MCnet school: [\(https://indico.cern.ch/event/669309/\)](https://indico.cern.ch/event/669309/).

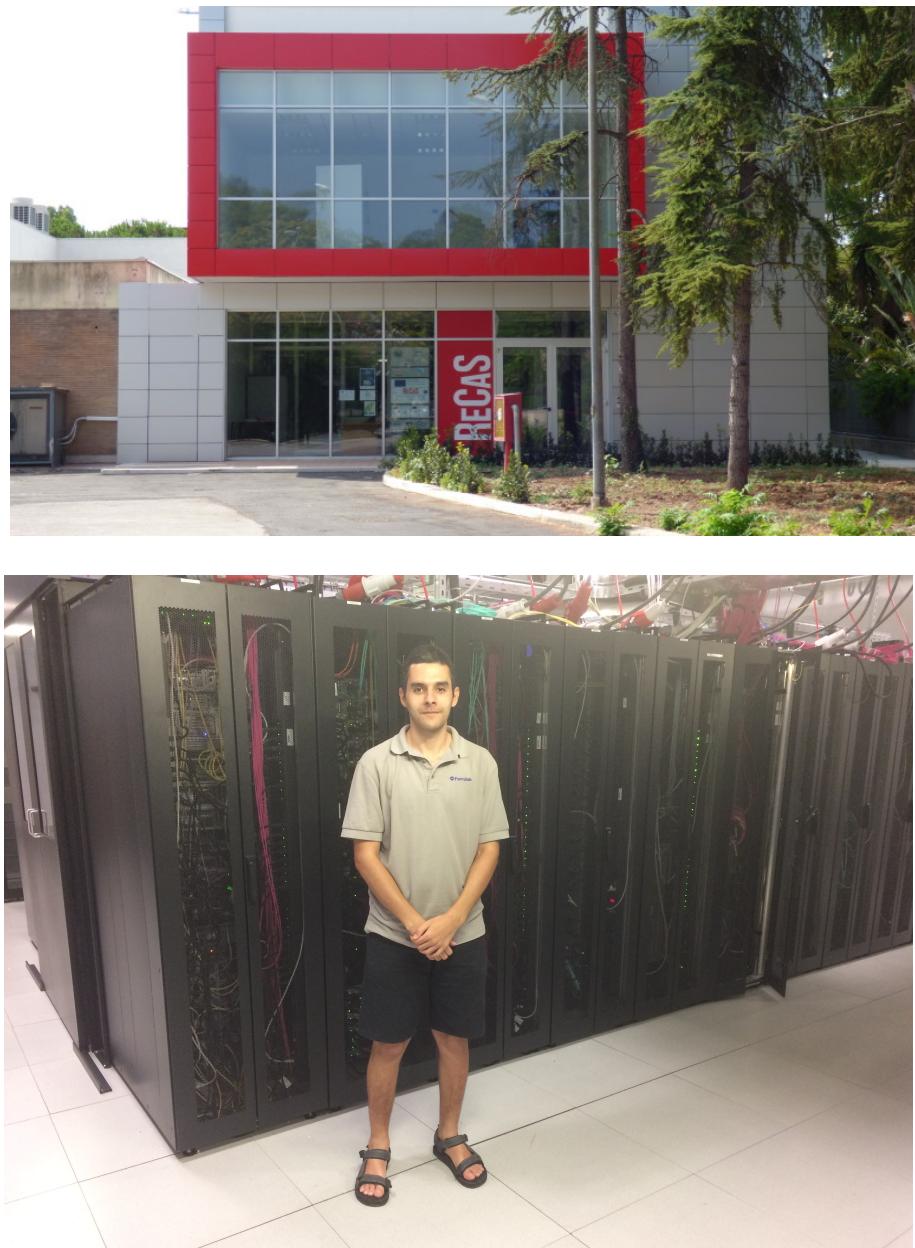
¹⁶ Details here: [\(https://www.recas-bari.it/index.php/it/\)](https://www.recas-bari.it/index.php/it/).

Figure 97 - Some pictures of the seminar given by the author at UNIBA during his participation in the Global Doc project. The organizers, the local tutor prof. Giovanna Selvagi and the co-advisor prof. Nicola were present in the seminar, which was also opened to the university academic community.



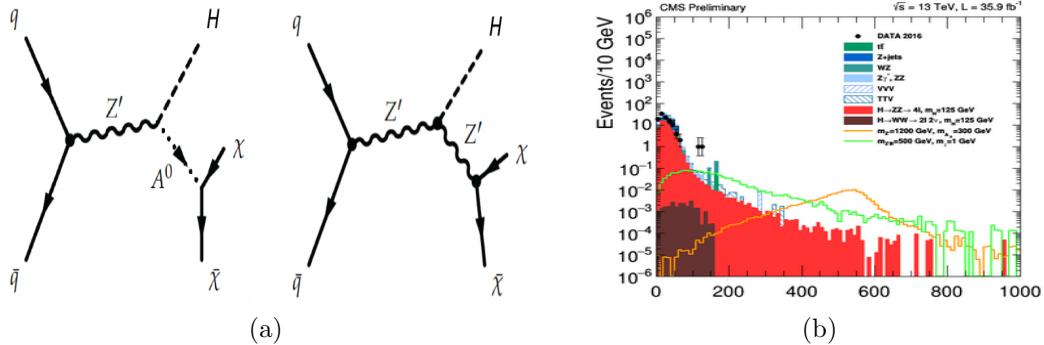
Source: The AUTHOR, 2018.

Figure 98 - The RECAS computing farm at the UNIBA campus where the Physics department is located.



Source: The AUTHOR, 2018.

Figure 99 - Briefly of the Dark Matter analysis: (a) the LO diagrams of the searched processes and (b) the MET distribution (stacked) for SM processes and (superimposed) Dark Matter candidates.



Source: Reference [77] p. 10 and adapted from [77] p. 49.

The Dark Matter candidate in this case is the Z' boson from two models. This particle appears in simplified models extending the SM gauge group with new symmetries, which in the referring analysis are the 2HDM (Two Higgs Doublet Model) and the Baryonic. In the 2HDM model the Z' decays into a Higgs (SM or not) and a neutral pseudoscalar particle A^0 . In the Baryonic model the Z' irradiates a Higgs (not SM). In both models there is MET in the final state due to the decay of the A^0 or Z' into a pair $\chi\bar{\chi}$ (neutralinos - supersymmetrical partners of neutrinos). Hence, the analysis is ultimately done over the MET distribution which is very sensitive for discriminate SM processes from these Dark Matter candidate process, see Fig. 99(b). For more details about this analysis, please see [77].

Another experience at UNIBA for the author, though shortly, has been the guiding of another student of the author's co-advisor through some basics on the MadGraph MC generator. This student was developing his PhD research project on the double and triple Higgs production processes, aiming the study of the associated couplings.

Finally, a very briefly documentation of each student participating in the Global Doc has been prepared and presented by the organizers of the project at the **Verso Horizon 2012-2017. La ricerca in rete in Europa**¹⁷ congress, which happened on October 5th at the UNIBA. For the interested reader, please see <https://www.uniba.it/elenco-siti-tematici/altri-siti-tematici/globaldoc>, which is the Global Doc web page and where one can find detailed information on this project and about future selective processes. According to the organizers, the documentation about the students participation on the project will be attached to that web page in the future.

¹⁷ See: <https://www.uniba.it/eventi-alluniversita/2018/verso-horizon-europe-2021-2027>.