SEARCH FOR PRODUCTION OF A HIGGS BOSON AND A SINGLE TOP QUARK IN MULTILEPTON FINAL STATES IN pp COLLISIONS AT $\sqrt{s} = 13$ TeV.

by

Jose Andres Monroy Montañez

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfilment of Requirements

For the Degree of Doctor of Philosophy

Major: Physics and Astronomy

Under the Supervision of Kenneth Bloom and Aaron Dominguez

Lincoln, Nebraska

July, 2018

SEARCH FOR PRODUCTION OF A HIGGS BOSON AND A SINGLE TOP QUARK IN MULTILEPTON FINAL STATES IN pp COLLISIONS AT $\sqrt{s} = 13$ TeV.

Jose Andres Monroy Montañez, Ph.D.

University of Nebraska, 2018

Adviser: Kenneth Bloom and Aaron Dominguez

Table of Contents

List of Figures

List of Tables

Chapter 5

Statistical methods

In the course of analyzing the data sets provided by the CMS experiment and used in this thesis, several statistical tools have been employed; in this chapter, a description of these tools will be presented, starting with the general statement of the multivariate analysis methods, followed by the particularities of the Boosted Decision Trees (BDT) method and its application to the classification problem. Statistical inference methods used will also be presented. This chapter is based mainly on References [126–128].

5.1 Multivariate analysis

Multivariate data analysis (MVA) makes use of the statistical techniques developed to analyze more than one variable at once, taking into account all the correlations among variables. MVA is employed in a variety of fields like consumer and market research, quality control and process optimization. Using MVA it is possible to identify the dominant patterns in a data sample, like groups, outliers and trends, and determine to which group a set of values belong; in the particle physics context, MVA methods are used to perform the selection of certain type of events from a large data set.

Processes with small cross section, such as the tHq process ($\sigma_{SM}(\sqrt{s}=13\text{TeV}) = 70.96$ fb), are hard to detect in the presence of the processes with larger cross sections, $\sigma^{t\bar{t}}_{SM}(\sqrt{s}=13\text{TeV}) = 823.44$ fb for instance; therefore, only a small fraction of the data contains events of interest (signal), the major part is signal-like events, which mimic signal characteristics but belong to different processes, so they are a background to the process of interest. This implies that it is not possible to say with certainty that a given event is a signal or a background and statistical methods should be involved. In that sense, the challenge can be formulated as one where a set of events have to be classified according to certain special features; these features correspond to the measurements of several parameters like energy or momentum, organized in a set of *input variables*. The measurements for each event can be written in a vector $\mathbf{x} = (x_1, ..., x_n)$ for which

- $f(\mathbf{x}|s)$ is the probability density (*likelihood function*) that $\mathbf{x}$ is the set of measured values given that the event is a signal event (signal hypothesis).
- $f(\mathbf{x}|b)$ is the probability density (*likelihood function*) that $\mathbf{x}$ is the set of measured values given that the event is a background event (background hypothesis).

Figure 5.1 shows three ways to perform a classification of events for which measurements of two properties, i.e., two input variables x_1 and x_2, have been performed; blue circles represent signal events while red triangles represent background events. The classification on the left is *cut-based* requiring $x_1 < c_1$ and $x_2 < c_2$; usually the cut values (c_1 and c_2) are chosen according to some knowledge about the event process. In the middle plot, the classification is performed using a linear function of the input variables, hence the boundary is a straight line, while in the right plot the

the relationship between input variables is not linear thus the boundary is not linear either.

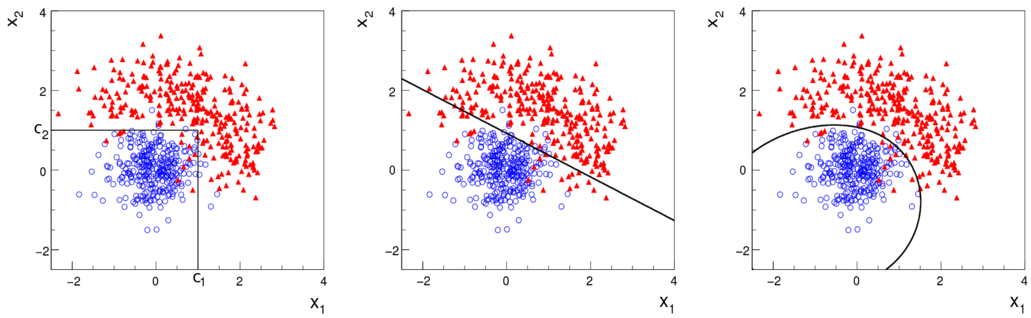

Figure 5.1: Scatter plots-MVA event classification. Distribution of two input variables x_1 and x_2 measured for a set of events; blue circles represent signal events and red triangles represent background events. The classification is based on cuts (left), linear boundary (center), and nonlinear boundary (right) [126]

.

In general, the boundary can be parametrized in terms of the input variables such that the cut is set on the parametrization instead of on the variables, i.e., $y(\mathbf{x}) = y_{cut}$ with y_{cut} being a constant; thus, the acceptance or rejection of an event is based on which side of the boundary the event is located. If $y(\mathbf{x})$, usually called *test statistic*, has functional form, it can be used to determine the probability distribution functions $p(y|s)$ and $p(y|b)$ and then perform a test statistic with a single cut on the scalar variable y.

Figure 5.2 shows an example of what would be the probability distribution functions under the signal and background hypotheses for a scalar test statistic with a cut on the classifier y. Note that the tails of the distributions indicate that some signal events fall in the rejection region and some background events fall on the acceptance region; therefore, it is convenient to define the *efficiency* with which events of a given type are accepted. The signal and background efficiencies are given by

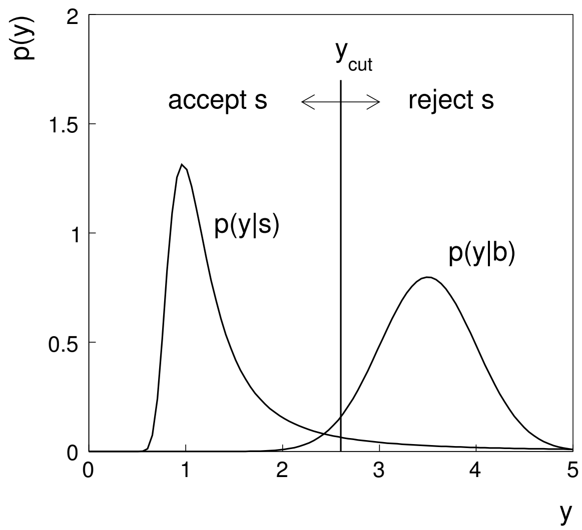

Figure 5.2: Distributions of the scalar test statistic $y(\mathbf{x})$ under the signal and background hypotheses. [126]

$$\varepsilon_{\mathrm{s}} = P(\text{accept event}|\mathrm{s}) = \int_{\mathrm{A}} f(\mathbf{x}|\mathrm{s})\, d\mathbf{x} = \int_{-\infty}^{y_{\mathrm{cut}}} p(y|\mathrm{s})\, dy \,, \tag{5.1}$$

$$\varepsilon_{\mathrm{b}} = P(\text{accept event}|\mathrm{b}) = \int_{\mathrm{A}} f(\mathbf{x}|\mathrm{b})\, d\mathbf{x} = \int_{-\infty}^{y_{\mathrm{cut}}} p(y|\mathrm{b})\, dy \,, \tag{5.2}$$

where A is the acceptance region. If the background hypothesis is the *null hypothesis (H_0)*, the signal hypothesis would be *alternative hypothesis (H_1)*; in this context, the background efficiency corresponds to the significance level of the test (α) and describes the misidentification probability, while the signal efficiency corresponds to the power of the test (1-β)[1] and describes the probability of rejecting the background hypothesis if the signal hypothesis is true. What is sought in an analysis is to maximize the power of the test relative to the significance level, i.e., set a selection with the largest possible selection efficiency and the smallest possible misidentification probability.

[1] β is the fraction of signal events that fall out of the acceptance region

5.1.1 Decision trees

For this thesis, the implementation of the MVA strategy, described above, is performed through decision trees by using the TMVA software package [127] included in the ROOT analysis framework [129]. In a simple picture, a decision tree classifies events according to their input variables values by setting a cut on each input variable and checking which events are on which side of the cut, just as proposed in the MVA strategy, but in addition, as a machine learning algorithm, decision trees offer the possibility to be trained and then perform the classification efficiently.

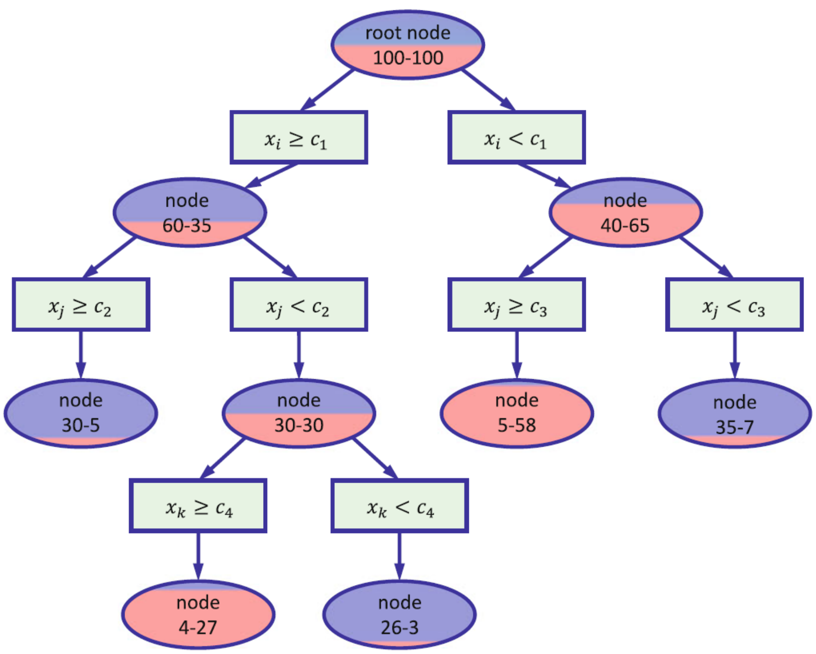

Figure 5.3: Example of a decision tree. Each node is fed with a MC sample mixing signal and background events (left-right numbers); nodes colors represent the relative number of signal/background events [128].

The training or growing of a decision tree is the process where the rules for classifying events are defined; this process is represented in Figure 5.3 and consists of several steps:

- take MC samples of signal and background events and split them into two parts each; the first parts will be used in the decision tree training, while the second parts will be used for testing the final classifier obtained from the training. Each event has associated a set of input variables $\mathbf{x} = (x_1, ..., x_n)$ which serve to distinguish between signal and background events. The training sample is taken in at the *root node*.

- Pick one variable, say x_i.
- Pick one value of x_i, each event has its own value of x_i, and split the training sample into two subsamples B_1 and B_2; B_1 contains events for which $x_i < c_1$ while B_2 contains the rest of the training events;
- scan all possible values of x_i and find the splitting value that provides the *best* classification[2], i.e., B_1 is mostly made of signal events while B_2 is mostly made of background events.
- It is possible that variables other than the picked one produce a better classification, hence, all the variables have to be evaluated. Pick the next variable, say x_j, and repeat the scan over its possible values.
- At the end, all the variables and their values will have been scanned, the *best* variable and splitting value will have been identified, say x_1, c_1, and there will be two nodes fed with the subsamples B_1 and B_2.

Nodes are further split by repeating the decision process until a given number of final nodes is obtained, nodes are largely dominated by either signal or background events, or nodes have too few events to continue. Final nodes are called *leaves* and they are classified as signal or background leaves according to the class of the majority of events in them. Each *branch* in the tree corresponds to a sequence of cuts.

[2] Quality of the classification will be treated in the next paragraph.

The quality of the classification at each node is evaluated through a separation criteria; there are several of them but the *Gini Index (G)* is the one used in the decision trees trained for the analysis in this thesis. G is written in terms of the purity (P), i.e., the fraction of signal events in the samples after the separation is made; it is given by

$$G = P(1-P) \tag{5.3}$$

note that P=0.5 at the root node while G=0 for pure leaves. For a node A split into two nodes B_1 and B_2 the G gain is

$$\Delta G = G(A) - G(B_1) - G(B_2). \tag{5.4}$$

The *best* classification corresponds to that for which the gain of G is maximized; hence, the scanning over all the variables in an event and their values is of great importance.

In order to provide a numerical output for the classification, events in a signal(background) leaf are assigned an score of 1(-1) each, defining in this way the decision tree *classifier/weak learner* as

$$f(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \;\; \text{in signal region,} \\ -1 & \mathbf{x} \;\; \text{in background region.} \end{cases}$$

Figure 5.4 shows an example of the classification of a sample of events, containing two variables, performed by a decision tree.

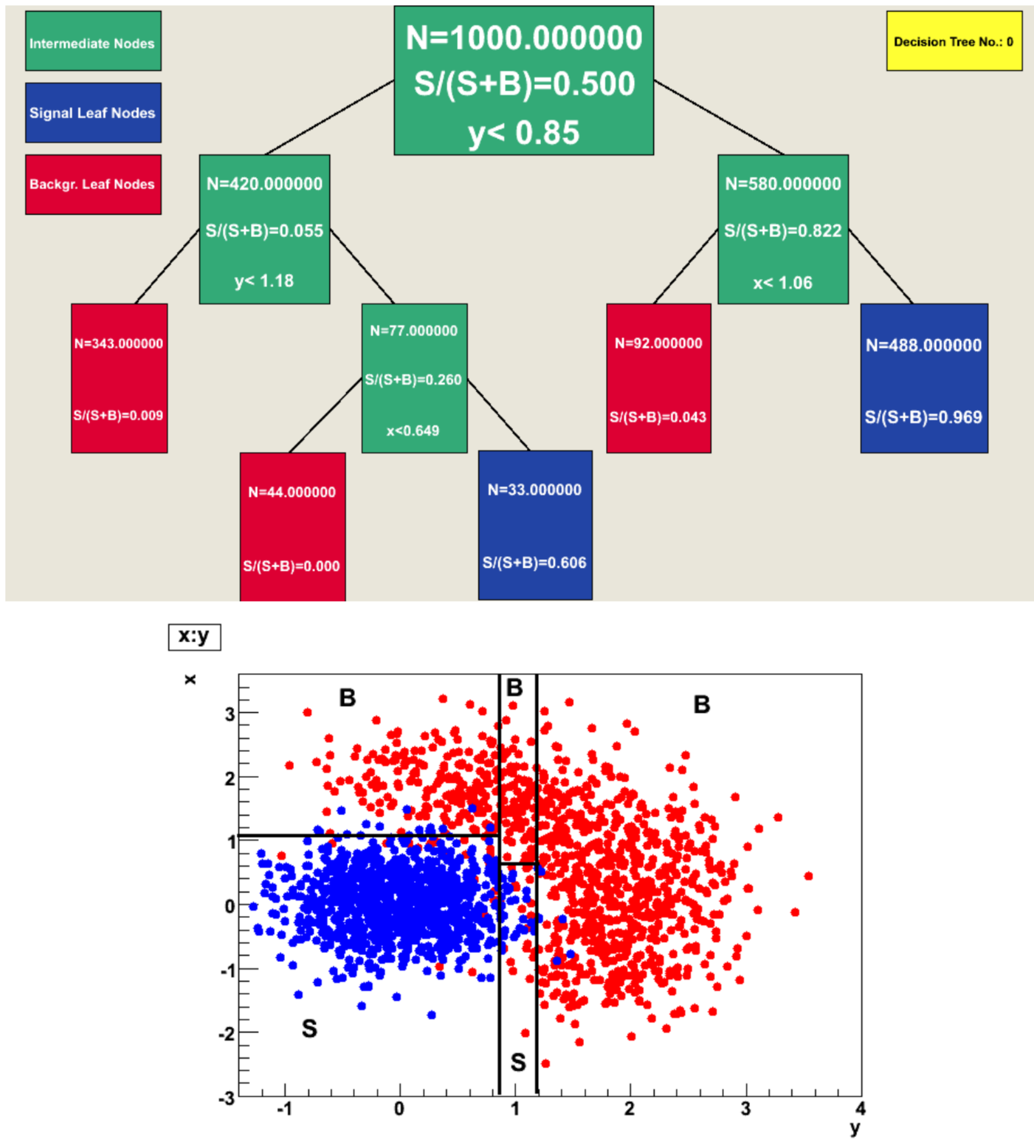

Figure 5.4: Example of a decision tree output. Each leaf, blue for signal events and red for background events, is represented by a region in the variables phase space [130].

5.1.2 Boosted decision trees (BDT).

Event misclassification occurs when a training event ends up in the wrong leaf, i.e., a signal event ends up in a background leaf or a background event ends up in a signal leaf. A way to correct it is to assign a weight to the misclassified events and train a second tree using the reweighted events; the event reweighting is performed by a boosting algorithm in such a way that when used in the training of a new decision tree the *boosted events* get correctly classified. The process is repeated iteratively adding a new tree to the forest and creating a set of classifiers, which are combined to create the next classifier; the final classifier offers more stability[3] and has a smaller misclassification rate than any individual ones. The resulting tree collection is known as a *boosted decision tree (BDT)*.

Thus, purity of the sample is generalized to

$$P = \frac{\sum_s w_s}{\sum_s w_s + \sum_b w_b} \tag{5.5}$$

where w_s and w_b are the weights of the signal and background events respectively; the Gini index is also generalized

$$G = \left(\sum_i^n w_i\right) P(1-P) \tag{5.6}$$

with n the number of events in the node. The final score of an event, after passing through the forest, is calculated as the renormalized sum of all the individual (possibly weighted) scores; thus, high(low) score implies that the event is most likely signal(background).

The boosting procedure, implemented in the *Gradient boosting* algorithm used in this thesis, produces a classifier $F(\mathbf{x})$ which is the weighted sum of the individual classifiers obtained after each iteration, i.e.,

$$F(\mathbf{x}) = \sum_{m=1}^{M} \beta_m f(\mathbf{x}; a_m) \tag{5.7}$$

where M is the number of trees in the forest. The *loss function* $L(F,y)$ represents the

[3] Decision trees suffer from sensitivity to statistical fluctuations in the training sample which may lead to very different results with an small change in the training samples.

deviation between the classifier $F(\mathbf{x})$ response and the true value y obtained from the training sample (1 for signal events and -1 for background event), according to

$$L(F,y) = \ln(1 + e^{-2F(\mathbf{x})y}) \tag{5.8}$$

thus, the reweighting is employed to ensure the minimization of the loss function; a more detailed description of the minimization procedure can be found in Reference [131]. The final classifier output is later used as a final discrimination variable, labeled as *BDT output/response*.

5.1.3 Overtraining

Decision trees offer the possibility to have as many nodes as desired in order to reduce the misclassification to zero (in theory); however, when a classifier is too much adjusted to a particular training sample, the classifier's response to a slightly different sample may leads to a completely different classification results; this effect is know as *overtraining*.

An alternative to reduce the overtraining in BDTs consists in prunning the tree by removing statistically insignificant nodes after the tree growing is completed but this option is not available for BDTs with gradient boosting in the TMVA-toolkit, therefore, the overtraining has to be reduced by tuning the algorithm, number of nodes, minimum number of events in the leaves, etc. The overtraining can be evaluated by comparing the responses of the classifier when running over the training and test samples.

5.1.4 Variable ranking

BDTs have a couple of particular advantages related to the input variables; they are relatively insensitive to the number of input variables used in the vector $\mathbf{x}$. The ranking of the BDT input variables is determined by counting the number of times a variable is used to split decision tree nodes; in addition, the separation gain-squared achieved in the splitting and the number of events in the node are accounted by applying a weighting to that number. Thus, those variables with small or no power to separate signal and background events are rarely chosen to split the nodes, i.e., are effectively ignored.

In addition, variables correlations play an important role for some MVA methods like the Fisher discriminant algorithm in which the first step consist of performing a linear transformation to a phase space where the correlations between variables are removed; in the case of BDT algorithm, correlations do not affect the performance.

5.1.5 BDT output example

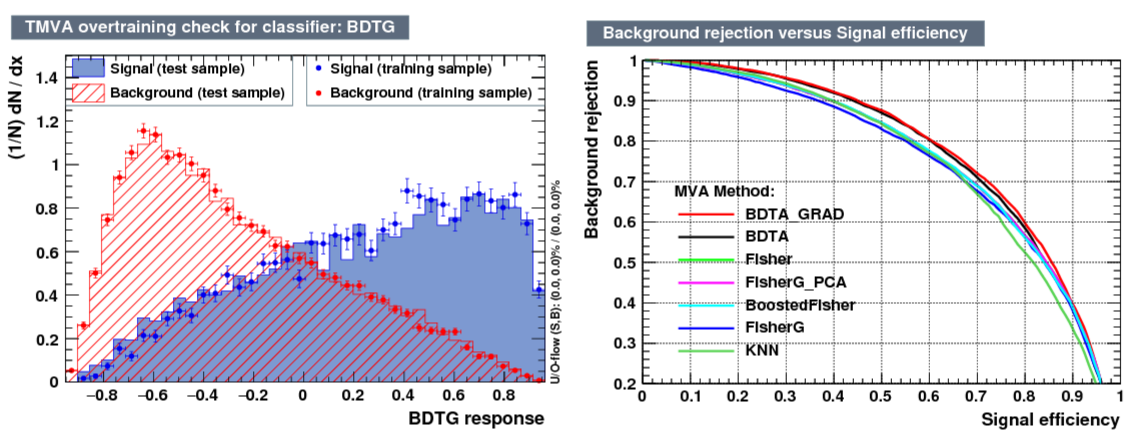

Figure 5.5: Left: Output distributions for the gradient boosted decision tree (BDTG) classifier using a sample of signal ($pp \to tHq$) and background ($pp \to t\bar{t}$) events. Right: Background rejection vs signal efficiency (ROC curves) for various MVA classifiers running over the same sample used to produce the plot on the left.

The left side of figure 5.5 shows the BDT output distributions for signal ($pp \to tHq$) and background ($pp \to t\bar{t}$) events; this plot is the equivalent to the one showed in Figure 5.2. A forest with 800 trees, maximum depth per tree = 3, and gradient boosting have been used as training parameters. The BDTG classifier offers a good separation power. There is a small overtraining in the signal distribution, while the background distribution is very well predicted which might indicate that the sample is composed of more background than signal events.

The right side of figure 5.5 shows the background rejection vs signal efficiency curves for several combinations of MVA classifiers-boosting algorithms running over the same MC sample; these curves are known as ROC curves and give an indication of the performance of the classifier. In this particular example, the best performance is achieved with the BDTG classifier (BDTA_GRAD), which motivate its use in this thesis.

5.2 Statistical inference

Once events are classified, the next step consists of finding the parameters that define the likelihood functions $f(\mathbf{x}|s), f(\mathbf{x}|b)$ for signal and background events respectively. In general, likelihood functions depend not only on the measurements but also on parameters (θ_m) that define their shapes; the process of estimating these *unknown parameters* and their uncertainties from the experimental data is called *inference*.

The statistical inference tools used in this analysis are implemented in the RooFit toolkit [132] and COMBINE package [133] included in the CMSSW software framework.

5.2.1 Nuisance parameters

The unknown parameter vector $\boldsymbol{\theta}$ is made of two types of parameters: those parameters that provide information about the physical observables of interest for the experiment or *parameters of interest*, and the *nuisance parameters* that are not of direct interest for the experiment but that need to be included in the analysis in order to achieve a satisfactory description of the data; they represent effects of the detector response like the finite resolutions of the detection systems, miscalibrations, and in general any source of uncertainty introduced in the analysis.

Nuisance parameters can be estimated from experimental data; for instance, data samples from a test beam are usually employed for calibration purposes. In cases where experimental samples are not availables, the estimation of nuisance parameters makes use of dedicated simulation programs to provide the required samples.

The estimation of the unknown parameters involves certain deviations from their true values, hence, the measurement of the nuisance parameter is written in terms of an estimated value, also called central value, $\hat{\theta}$ and its uncertainty $\delta\theta$ using the notation

$$\theta = \hat{\theta} \pm \delta\theta \tag{5.9}$$

where the interval $[\hat{\theta} - \delta\theta, \hat{\theta} + \delta\theta]$ is called *confidence interval*; it is usually interpreted, in the limit of infinite number of experiments, as the interval where the true value of the unknown parameter θ is contained with a probability of 0.6827 (if no other convention is stated); this interval represents the area under a Gaussian distribution in the interval $\pm 1\sigma$.

The uncertainties associated with nuisance parameters produce *systematic uncertainties* in the final measurement, while the uncertainties related only to fluctuations in data and that affect the determination of parameters of interest produce *statistical uncertainties*.

5.2.2 Maximum likelihood estimation method

The estimation of the unknown parameters that are in best agreement with the observed data is performed through a function of the data sample that returns the estimate of those parameters; that function is called an *estimator*. Estimators are usually constructed using mathematical expressions encoded in algorithms.

In this thesis, the estimator used is the likelihood function $f(\mathbf{x}|\boldsymbol{\theta})$[4] which depends on a set of measured variables $\mathbf{x}$ and a set of unknown parameters $\boldsymbol{\theta}$. The likelihood function for N events in a sample is the combination of all the individual likelihood functions, i.e.,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\mathbf{x}^i|\boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_1^i, ..., x_n^i; \theta_1, ..., \theta_m) \tag{5.10}$$

and the estimation method used is the *Maximum Likelihood Estimation* method (MLE); it is based on the combined likelihood function defined by eqn. 5.10 and the procedure seeks for the parameter set that corresponds to the maximum value of the combined likelihood function, i.e., the *maximum likelihood estimator* of the unknown parameter vector $\boldsymbol{\theta}$ is the function that produces the vector of *best estimators* $\hat{\boldsymbol{\theta}}$ for which the likelihood function $L(\boldsymbol{\theta})$ evaluated at the measured $\mathbf{x}$ is maximum.

Usually, the logarithm of the likelihood function is used in numerical algorithm implementations in order to avoid underflow the numerical precision of the computers due to the product of low likelihoods. In addition, it is common to minimize the negative logarithm of the likelihood function, therefore, the negative log-likelihood

[4] analogue to the likelihood functions described in previous sections

function is

$$F(\boldsymbol{\theta}) = -\ln L(\boldsymbol{\theta}) = -\sum_{i=1}^{N} f(\mathbf{x}^i|\boldsymbol{\theta}). \tag{5.11}$$

The minimization process is performed by the software MINUIT [134] implemented in the ROOT analysis framework. In case of data samples with large number of measurements, the computational resources necessary to calculate the likelihood function are too big; therefore, the parameter estimation is performed using binned distributions of the variables of interest for which the *binned likelihood function* is given by

$$L(\mathbf{x}|r,\boldsymbol{\theta}) = \prod_{i=1} \frac{(r \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-r \cdot s_i(\boldsymbol{\theta}) - b_i(\boldsymbol{\theta})} \prod_{j=1} \frac{1}{\sqrt{2\pi\sigma_{\theta_j}^2}} e^{-(\theta_j - \theta_{0,j})^2/2\sigma_{\theta_j}^2}, \tag{5.12}$$

with s_i and b_i the expected number of signal and background yields for the bin i, n_i is the observed number of events in the bin i and $r = \sigma/\sigma_{SM}$ is the signal strength. Note that the number of entries per bin follows a Poisson distribution. The effect of the nuisance parameters have been included in the likelihood function through the multiplication by a Gaussian distribution that models the nuisance. The three parameters, r, s_i and b_i are jointly fitted to estimate the value of r.

5.3 Upper limits

In this analysis, two hypotheses are considered; the background only hypothesis ($H_0(b)$) and the signal plus background hypothesis ($H_1(s+b)$), i.e., the sample of events is composed of background only events (r=0) or it is a mixture of signal plus background events (r=1). The exclusion of one hypothesis against the other means that the observed data sample better agrees with H_0 or rather with H_1. In order to discriminate these hypotheses, a test statistic is constructed on the basis of the likelihood function evaluated for each of the hypothesis.

The *Neyman-Pearson* lemma [135] states that the test statistic that provides the maximum power for H_1 for a given significance level (background misidentification probability α), is given by the ratio of the likelihood functions $L(\mathbf{x}|H_1)$ and $L(\mathbf{x}|H_0)$; however, in order to use that definition it is necessary to know the true likelihood functions, which in practice is not always possible. Approximate functions obtained by numerical methods, like the BDT method described above, have to be used, so that the *profile likelihood* test statistic is defined by

$$\lambda(\mathbf{r}) = \frac{L(\mathbf{x}|r,\hat{\hat{\boldsymbol{\theta}}}(r))}{L(\mathbf{x}|\hat{r},\hat{\boldsymbol{\theta}})}, \tag{5.13}$$

where, $\hat{r}$ and $\hat{\boldsymbol{\theta}}$ maximize the likelihood function, and $\hat{\hat{\boldsymbol{\theta}}}$ maximizes the likelihood function for a given value of the signal strength modifier r. In practice, the test statistic t_r

$$t_r = -2\ln\lambda(r) \tag{5.14}$$

is used to evaluate the presence of signal in the sample, since the minimum of t_r at $r = \hat{r}$ suggests the presence of signal with signal strength $\hat{r}$. The uncertainty interval for r is determined by the values of r for which $t_r = +1$.

The expected probability density function (p.d.f) $f(t_r|r,\boldsymbol{\theta})$ of the test statistic t_r can be obtained numerically by generating MC samples where one hypothesis, $H_0(b)$ or $H_1(s+b)$, is assumed; thus, MC samples contain the possible values of t_r obtained from *pseudo-experiments* as shown in Figure 5.6. The probability that t_r takes a value equal or greater than the observed value ($t_{r,obs}$) when a signal with a signal modifier r is present in the data sample, is called the *p-value* of the observation; it can be calculated using

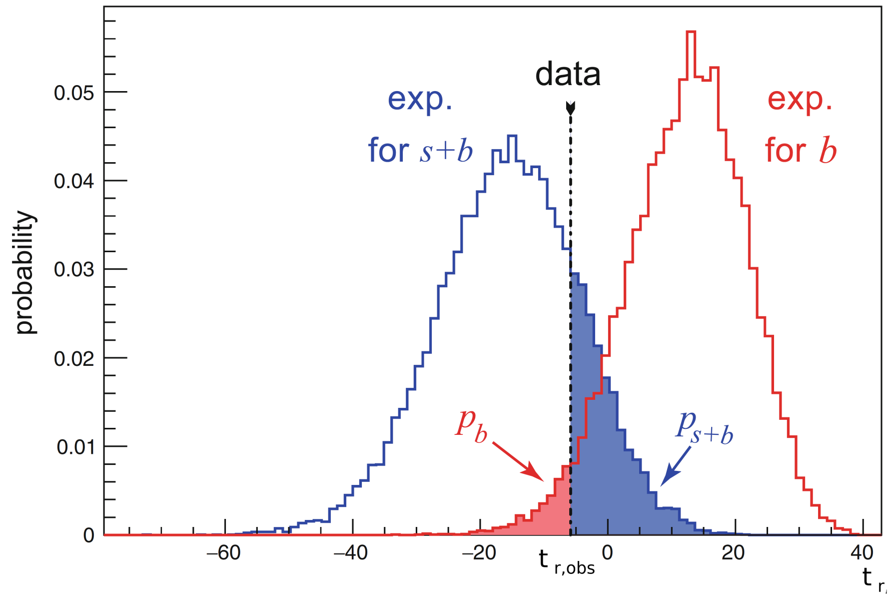

Figure 5.6: t_r p.d.f. from MC pseudo experiments assuming H_0 (red) and H_1 (blue). The black dashed line shows the value of the test statistic as measured from data. Adapted from Reference [128].

$$p_r = \int_{t_{r,obs}}^{\infty} f(t_r'|r,\boldsymbol{\theta})dt_r', \tag{5.15}$$

thus, $p_r < 0.05$ means that, for that particular value of r, H_1 could be excluded at 95% Confidence Level (CL). The corresponding background-only p-value is given by

$$1 - p_b = \int_{t_{r,obs}}^{\infty} f(t_r'|0,\boldsymbol{\theta})dt_r', \tag{5.16}$$

If the t_r p.d.f.s for both hypotheses are well separated, as shown in the top side of Figure 5.7, the experiment is sensitive to the presence of signal in the sample. If the signal presence is small, both p.d.f.s will be largely overlapped (bottom of Figure **??**) and either the signal hypothesis could be rejected with not enough justification because the experiment is not sensitive to the signal or a fluctuation of the background could be misinterpreted as presence of signal with the corresponding rejection of the

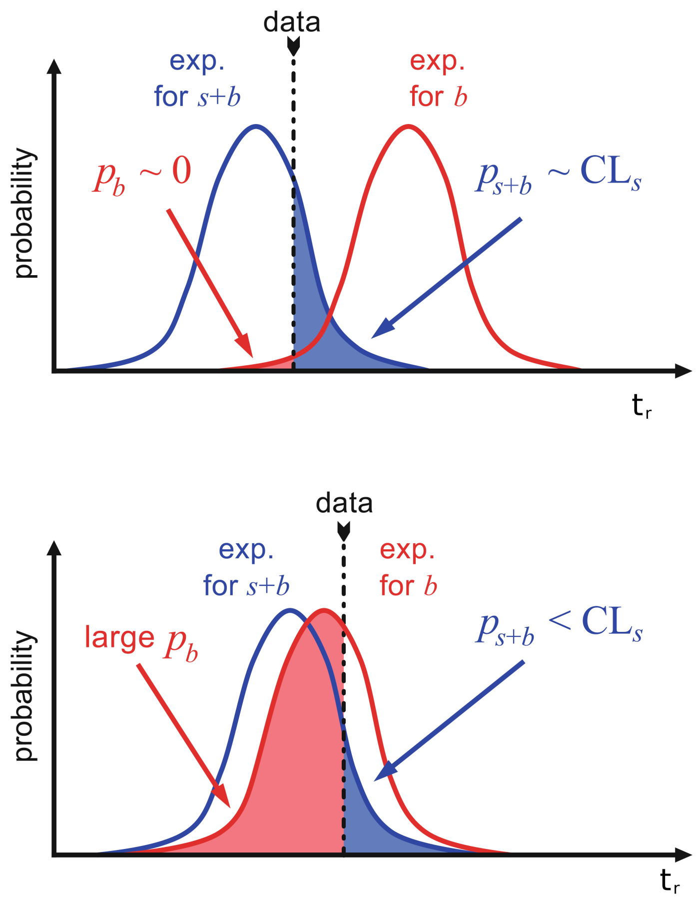

Figure 5.7: CL_s limit illustration. When the test statistic p.d.f. for the two hypotheses H_0 and H_1 are well separated (top) and when they are largely overlapped (bottom). Adapted from Reference [128].

background-only hypothesis. These issues are corrected by using the modified p-value [136]

$$p'_r = \frac{p_r}{1 - p_b} \equiv CL_s. \tag{5.17}$$

If H_1 is true, then p_b is small, $CL_s \simeq p_r$ and H_0 is rejected; if there is large overlap and an statistical fluctuation cause that p_b is large, then both numerator and denominator in Eqn. 5.17 become small but CL_s would allow the rejection of H_1 even if there is poor sensitivity to signal.

The upper limit of the parameter of interest r^{up} is determined by excluding the range of values of r for which $CL_s(r,\boldsymbol{\theta})$ is lower than the confidence level desired, normally 90% or 95%, e.g, scanning over r and finding the value for which $p_r^{'up} = 0.05$. The expected upper limit can be calculated using pseudo-experiments based on the background-only hypothesis and obtaining a distribution for r_{ps}^{up}; the median of that distribution corresponds to the expected upper limit, while the $\pm 1\sigma$ and $\pm 2\sigma$ deviations correspond to the values of the distribution that defines the 68% and 95% of the area under the distribution centered in the median. It is usual to present all the information about the expected and observed limits in the so-called *Brazilian-flag plot* as the one showed in Figure 5.8. The solid line represent the observed CL_s

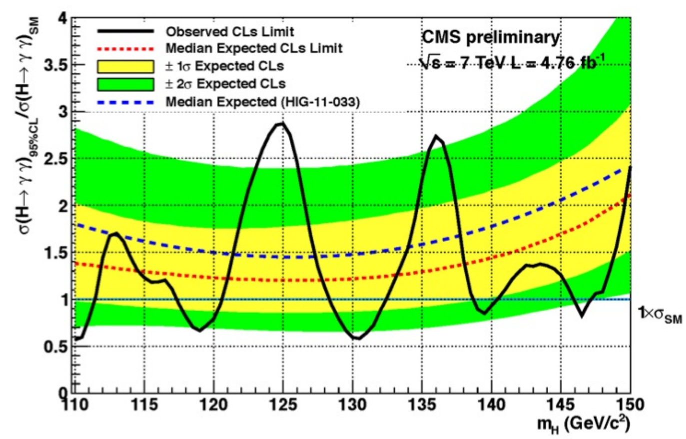

Figure 5.8: Brazilian flag plot of CMS experiment limits for Higgs boson decaying to photons [137].

5.4 Asymptotic limits

As said before, the complexity of the likelihood functions, the construction of test statistics, and the calculation of the limits and their uncertainties is not always manageable and requires extensive computational resources; in order to overcome those issues, asymptotic approximations for likelihood-based test statistics, like the ones described in previous sections, have been developed [138, 139] using Wilks' theorem. Asymptotic approximations replace the construction of the test statistics p.d.f.s using MC pseudo-experiments, with the approximate calculation of the test statistics p.d.f.s by employing the so-called *Asimov dataset*.

The Asimov dataset is defined as the dataset that produce the true values of the nuisance parameters when it is used to evaluate the estimators for all the parameters; it is obtained by setting the values of the variables in the dataset to their expected values [139].

Limits calculated by using the asymptotic approximation and the Asimov dataset are know as *asymptotic limits*.

References

[1] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". Physical Review. 74 (10): 1439-61, (1948). https://doi.org/10.1103/PhysRev.74.1439

[2] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". Physical Review. 76 (6): 769-89, (1949). https://doi.org/10.1103/PhysRev.76.769

[3] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields". Progress of Theoretical Physics. 1 (2): 27-42, (1946). https://doi.org/10.1143/PTP.1.27

[4] D.J. Griffiths, "Introduction to electrodynamics". 4th ed. Pearson, (2013).

[5] F. Mandl, G. Shaw. "Quantum field theory." Chichester, Wiley (2009).

[6] F. Halzen, and A.D. Martin, "Quarks and leptons: An introductory course in modern particle physics". New York: Wiley, (1984) .

[7] File: Standard_Model_of_Elementary_Particle_dark.svg. (2017, June 12) Wikimedia Commons, the free media repository. Retrieved November 27, 2017 from https://www.collegiate-advanced-electricity.com/single-post/2017/04/10/The-Standard-Model-of-Particle-Physics.

[8] E. Noether, “Invariante Variationsprobleme”, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, vol. 1918, pp. 235-257, (1918).

[9] C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016) and 2017 update.

[10] M. Goldhaber, L. Grodzins, A.W. Sunyar “Helicity of Neutrinos”, Phys. Rev. 109, 1015 (1958).

[11] Palanque-Delabrouille N et al. “Neutrino masses and cosmology with Lyman-alpha forest power spectrum”, JCAP 11 011 (2015).

[12] M. Gell-Mann. “A Schematic Model of Baryons and Mesons”. Physics Letters. 8 (3): 214-215 (1964).

[13] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking” (PDF). CERN Report No.8182/TH.401 (1964).

[14] G. Zweig. “An SU(3) Model for Strong Interaction Symmetry and its Breaking: II” (PDF). CERN Report No.8419/TH.412(1964).

[15] M. Gell-Mann. “The Interpretation of the New Particles as Displaced Charged Multiplets”. Il Nuovo Cimento 4: 848. (1956).

[16] T. Nakano, K, Nishijima. “Charge Independence for V-particles”. Progress of Theoretical Physics 10 (5): 581-582. (1953).

[17] N. Cabibbo, “Unitary symmetry and leptonic decays” Physical Review Letters, vol. 10, no. 12, p. 531, (1963).

[18] M.Kobayashi, T.Maskawa, "CP-violation in the renormalizable theory of weak interaction," Progress of Theoretical Physics, vol. 49, no. 2, pp. 652-657, (1973).

[19] File: Weak Decay (flipped).svg. (2017, June 12). Wikimedia Commons, the free media repository. Retrieved November 27, 2017 from https://commons.wikimedia.org/w/index.php?title=File:Weak_Decay_(flipped)\.svg&oldid=247498592.

[20] Georgia Tech University. Coupling Constants for the Fundamental Forces(2005). Retrieved January 10, 2018, from http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/couple.html#c2

[21] M. Strassler. (May 31, 2013).The Strengths of the Known Forces. Retrieved January 10, 2018, from https://profmattstrassler.com/articles-and-posts/particle-physics-basics/the-known-forces-of-nature/the-strength-of-the-known-forces/

[22] S.L. Glashow. "Partial symmetries of weak interactions", Nucl. Phys. 22 579-588, (1961).

[23] A. Salam, J.C. Ward. "Electromagnetic and weak interactions", Physics Letters 13 168-171, (1964).

[24] S. Weinberg, "A model of leptons", Physical Review Letters, vol. 19, no. 21, p. 1264, (1967).

[25] M. Peskin, D. Schroeder, "An introduction to quantum field theory". Perseus Books Publishing L.L.C., (1995).

[26] A. Pich. "The Standard Model of Electroweak Interactions" https://arxiv.org/abs/1201.0537

[27] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 122, 103 (1983).

[28] M. Banner et al. (UA2 Collaboration), Phys. Lett. B 122, 476 (1983).

[29] G. Arnison et al. (UA1 Collaboration), Phys. Lett. B 126, 398 (1983).

[30] P. Bagnaia et al. (UA2 Collaboration), Phys. Lett. B 129, 130 (1983).

[31] F.Bellaiche. (2012, 2 September). “What’s this Higgs boson anyway?”. Retrieved from: `https://www.quantum-bits.org/?p=233`

[32] M. Endres et al. Nature 487, 454-458 (2012) doi:10.1038/nature11255

[33] F. Englert, R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. Physical Review Letters. 13 (9): 321-23.(1964) doi:10.1103/PhysRevLett.13.321

[34] P.Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. Physical Review Letters. 13 (16): 508-509,(1964). doi:10.1103/PhysRevLett.13.508.

[35] G.Guralnik, C.R. Hagen and T.W.B. Kibble. “Global Conservation Laws and Massless Particles”. Physical Review Letters. 13 (20): 585-587, (1964). doi:10.1103/PhysRevLett.13.585.

[36] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. Physics Letters B. 716 (1): 30-61 (2012). arXiv:1207.7235. doi:10.1016/j.physletb.2012.08.021

[37] ATLAS collaboration. “Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC”. Physics Letters B. 716 (1): 1-29 (2012). arXiv:1207.7214. doi:10.1016/j.physletb.2012.08.020.

[38] ATLAS collaboration; CMS collaboration (26 March 2015). "Combined Measurement of the Higgs Boson Mass in pp Collisions at âĹŽs=7 and 8 TeV with the ATLAS and CMS Experiments". Physical Review Letters. 114 (19): 191803. arXiv:1503.07589. doi:10.1103/PhysRevLett.114.191803.

[39] LHC InternationalMasterclasses"When protons collide". Retrieved from `http://atlas.physicsmasterclasses.org/en/zpath_protoncollisions.htm`

[40] CMS Collaboration, "SM Higgs Branching Ratios and Total Decay Widths (update in CERN Report4 2016)". `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR` , last accessed on 17.12.2017.

[41] R.Grant V. "Determination of Higgs branching ratios in $H \to W^+W^- \to l\nu jj$ and $H \to ZZ \to l^+l^-jj$ channels". Physics Department, University of Tennessee (Dated: October 31, 2012). Retrieved from `http://aesop.phys.utk.edu/ph611/2012/projects/Riley.pdf`

[42] LHC Higgs Cross Section Working Group, Denner, A., Heinemeyer, S. et al. "Standard model Higgs-boson branching ratios with uncertainties". Eur. Phys. J. C (2011) 71: 1753. `https://doi.org/10.1140/epjc/s10052-011-1753-8`

[43] D. de Florian et al., LHC Higgs Cross Section Working Group, CERNâĂŞ2017âĂŞ002-M, arXiv:1610.07922[hep-ph] (2016).

[44] ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV," (2016). CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002.

[45] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Perez-Victoria, "Handbook of vector-like quarks: Mixing and single production", Phys. Rev. D 88 (2013) 094010, doi:10.1103/PhysRevD.88.094010, arXiv:1306.0572.

[46] A. Greljo, J. F. Kamenik, and J. Kopp, "Disentangling flavor violation in the top-Higgs sector at the LHC", JHEP 07 (2014) 046, doi:10.1007/JHEP07(2014)046, arXiv:1404.1278.

[47] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, "Higgs production in association with a single top quark at the LHC," European Physical Journal C, vol. 75, p. 267, (2015). doi:10.1140/epj c/s10052-015-3475-9, arXiv:1504.00611.

[48] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, "tWH associated production at the LHC", European Physical Journal C, vol. 77, p. 34, (2017). arXiv:1607.05862

[49] F. Maltoni, K. Paul, T. Stelzer, and S. Willenbrock, "Associated production of Higgs and single top at hadron colliders", Phys.Rev. D64 (2001) 094023, [hep-ph/0106293].

[50] S. Biswas, E. Gabrielli, F. Margaroli, and B. Mele, "Direct constraints on the top-Higgs coupling from the 8 TeV LHC data," Journal of High Energy Physics, vol. 07, p. 073, (2013).

[51] M. Farina, C. Grojean, F. Maltoni, E. Salvioni, and A. Thamm, "Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson," Journal of High Energy Physics, vol. 05, p. 022, (2013).

[52] T.M. Tait and C.-P. Yuan, "Single top quark production as a window to physics beyond the standard model", Phys. Rev. D 63 (2000) 014018 [hep-ph/0007298].

[53] CMS Collaboration, “Modelling of the single top-quark production in association with the Higgs boson at 13 TeV.” `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV`, last accessed on 16.01.2018.

[54] CMS Collaboration, “SM Higgs production cross sections at $\sqrt{s}$ = 13 TeV.” `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV`, last accessed on 16.01.2018.

[55] S. Dawson, The effective W approximation, Nucl. Phys. B 249 (1985) 42.

[56] S. Biswas, E. Gabrielli and B. Mele, JHEP 1301 (2013) 088 [arXiv:1211.0499 [hep-ph]].

[57] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 4.Deciphering the Nature of the Higgs Sector”, arXiv:1610.07922.

[58] J. Ellis, D. S. Hwang, K. Sakurai, and M. Takeuchi.“Disentangling Higgs-Top Couplings in Associated Production”, JHEP 1404 (2014) 004, [arXiv:1312.5736].

[59] CMS Collaboration, V. Khachatryan et al., “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV,” arXiv:1412.8662.

[60] ATLAS Collaboration, G. Aad et al., “Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-}1 of proton-proton collision data”, ATLAS-CONF-2014-009.

[61] File:Cern-accelerator-complex.svg. Wikimedia Commons, the free media repository. Retrieved January, 2018 from `https://commons.wikimedia.org/wiki/File:Cern-accelerator-complex.svg`

[62] J.L. Caron , “Layout of the LEP tunnel including future LHC infrastructures.”, (Nov, 1993). A C Collection. Legacy of AC. Pictures from 1992 to 2002. Retrieved from https://cds.cern.ch/record/841542

[63] M. Vretenar, “The radio-frequency quadrupole”. CERN Yellow Report CERN-2013-001, pp.207-223 DOI:10.5170/CERN-2013-001.207. arXiv:1303.6762

[64] L.Evans. P. Bryant (editors). “LHC Machine”. JINST 3 S08001 (2008).

[65] CERN Photographic Service.“Radio-frequency quadrupole, RFQ-1”, March 1983, CERN-AC-8303511. Retrieved from https://cds.cern.ch/record/615852.

[66] CERN Photographic Service “Animation of CERN’s accelerator network”, 14 October 2013. DOI: 10.17181/cds.1610170 Retrieved from https://videos.cern.ch/record/1610170

[67] C.Sutton. “Particle accelerator”.Encyclopedia Britannica. July 17, 2013. Retrieved from https://www.britannica.com/technology/particle-accelerator.

[68] L.Guiraud. “Installation of LHC cavity in vacuum tank.”. July 27 2000. CERN-AC-0007016. Retrieved from https://cds.cern.ch/record/41567.

[69] J.L. Caron, “Magnetic field induced by the LHC dipole’s superconducting coils”. March 1998. AC Collection. Legacy of AC. Pictures from 1992 to 2002. LHC-PHO-1998-325. Retrieved from https://cds.cern.ch/record/841511.

[70] AC Team. “Diagram of an LHC dipole magnet”. June 1999. CERN-DI-9906025 retrieved from https://cds.cern.ch/record/40524.

[71] CMS Collaboration "Public CMS Luminosity Information". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_proton_proton_13_TeV_collis, last accessed 24.01.2018

[72] J.L Caron. "LHC Layout" AC Collection. Legacy of AC. Pictures from 1992 to 2002. September 1997, LHC-PHO-1997-060. Retrieved from https://cds.cern.ch/record/841573.

[73] J.A. Coarasa. "The CMS Online Cluster:Setup, Operation and Maintenance of an Evolving Cluster". ISGC 2012, 26 February - 2 March 2012, Academia Sinica, Taipei, Taiwan.

[74] CMS Collaboration. "The CMS experiment at the CERN LHC" JINST 3 S08004 (2008).

[75] CMS Collaboration. "CMS detector drawings 2012" CMS-PHO-GEN-2012-002. Retrieved from http://cds.cern.ch/record/1433717.

[76] Davis, Siona Ruth. "Interactive Slice of the CMS detector", Aug. 2016, CMS-OUTREACH-2016-027, retrieved from https://cds.cern.ch/record/2205172

[77] R. Breedon. "View through the CMS detector during the cooldown of the solenoid on February 2006. CMS Collection", February 2006, CMS-PHO-OREACH-2005-004, Retrieved from https://cds.cern.ch/record/930094.

[78] Halyo, V. and LeGresley, P. and Lujan, P. "Massively Parallel Computing and the Search for Jets and Black Holes at the LHC", Nucl.Instrum.Meth. A744 (2014) 54-60, DOI: 10.1016/j.nima.2014.01.038"

[79] A. Dominguez et. al. “CMS Technical Design Report for the Pixel Detector Upgrade”, CERN-LHCC-2012-016. CMS-TDR-11.

[80] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” Journal of Instrumentation, vol. 9, no. 10, p. P10009,(2014).

[81] CMS Collaboration and M. Brice. “Images of the CMS Tracker Inner Barrel”, November 2008, CMS-PHO-TRACKER-2008-002. Retrieved from https://cds.cern.ch/record/1431467.

[82] M. Weber. “The CMS tracker”. 6th international conference on hyperons, charm and beauty hadrons Chicago, June 28-July 3 2004.

[83] CMS Collaboration. “Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process”. Jul 26, 2013. arXiv:1307.7135

[84] L. Veillet. “End assembly of HB with EB rails and rotation inside SX ”,January 2002. CMS-PHO-HCAL-2002-002. Retrieved from https://cds.cern.ch/record/42594.

[85] J. Puerta-Pelayo.“First DT+RPC chambers installation round in the UX5 cavern.”. January 2007, CMS-PHO-OREACH-2007-001. Retrieved from https://cds.cern.ch/record/1019185

[86] X. Cid Vidal and R. Cid Manzano. “CMS Global Muon Trigger” web site: Taking a closer look at LHC. Retrieved from https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_trigger

[87] WLCG Project Office, “Documents & Reference - Tiers - Structure,” (2014). http://wlcg.web.cern.ch/documents-reference , last accessed on 30.01.2018.

[88] CMS Collaboration. “CMSSW Application Framework”, https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework, last accesses 06.02.2018

[89] A. Buckleya, J. Butterworthb, S. Giesekec, et. al. “General-purpose event generators for LHC physics”. arXiv:1101.2599v1 [hep-ph] 13 Jan 2011

[90] A. Quadt. “Top Quark Physics at Hadron Colliders”. Advances in the Physics of Particles and Nuclei. Springer-Verlag Berlin Heidelberg. DOI: 10.1007/978-3-540-71060-8 (2007)

[91] DurhamHep Data Project, “The Durham HepData Project - PDF Plotter.” http://hepdata.cedar.ac.uk/pdf/pdf3.html , last accessed on 02.02.2018.

[92] G. Altarelli and G. Parisi. “ASYMPTOTIC FREEDOM IN PARTON LANGUAGE”, Nucl.Phys. B126:298 (1977).

[93] Yu.L. Dokshitzer. Sov.Phys. JETP 46:641 (1977)

[94] V.N. Gribov, L.N. Lipatov. “Deep inelastic e p scattering in perturbation theory”, Sov.J.Nucl.Phys. 15:438 (1972)

[95] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a reappraisal,” Journal of High Energy Physics, vol. 07, p. 022, (2012).

[96] B. Andersson, G. Gustafson, G.Ingelman and T. Sjostrand, “Parton fragmentation and string dynamics”, Physics Reports, Vol. 97, No. 2-3, pp. 31-145, 1983.

[97] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements;” European Physical Journal C, vol. 76, no. 3, p. 155, (2016).

[98] J. Alwall et. al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” Journal of High Energy Physics, vol. 07, p. 079, (2014).

[99] T. SjÃűstrand and P. Z. Skands, “Transverse-momentum-ordered showers and interleaved multiple interactions,” European Physical Journal C, vol. 39, pp. 129âĂŞ154, (2005).

[100] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method,” Journal of High Energy Physics, vol. 11, p. 070, (2007).

[101] S. Agostinelli et al., “GEANT4: A Simulation toolkit,” Nuclear Instruments and Methods in Physics, vol. A506, pp. 250âĂŞ303, (2003).

[102] J.Allison et.al.,“Recent developments in Geant4”, Nuclear Instruments and Methods in Physics Research A 835 (2016) 186-225.

[103] CMS Collaboration “Full Simulation Offline Guide”, `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideSimulation`, last accessed 04.02.2018

[104] A. Giammanco. “The Fast Simulation of the CMS Experiment” J. Phys.: Conf. Ser. 513 022012 (2014)

[105] A.M. Sirunyan et. al. “Particle-flow reconstruction and global event description with the CMS detector”, JINST 12 P10003 (2017) `https://doi.org/10.1088/1748-0221/12/10/P10003`.

[106] The CMS Collaboration. “ Description and performance of track and primary vertex reconstruction with the CMS tracker”. JINST 9 P10009 (2014). doi:10.1088/1748-0221/9/10/P10009

[107] J. Incandela. “Status of the CMS SM Higgs Search” July 4, 2012. Pdf slides. Retrieved from `https://indico.cern.ch/event/197461/contributions/1478917/attachments/290954/406673/CMS_4July2012_Final.pdf`

[108] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, Nucl. Instrum. Meth. A 294 219. (1990).

[109] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf C 0303241 (2003) TULT009 [physics/0306087].

[110] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression and related Optimisation Problems”, Proc. IEEE 86 (1998) 2210.

[111] R. Fruhwirth, W. Waltenberger and P. Vanlaer, “ Adaptive Vertex Fitting”, CMS Note 2007-008 (2007).

[112] CMS collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV ”, JINST 7 P10002 2012, [arXiv:1206.4071].

[113] Coco, Victor and Delsart, Pierre-Antoine and Rojo-Chacon, Juan and Soyez, Gregory and Sander, Christian, “Jets and jet algorithms”, Proceedings, HERA and the LHC Workshop Series on the implications of HERA for LHC physics: 2006-2008, pag. 182-204. `http://inspirehep.net/record/866539/files/access.pdf`, (2009), doi:10.3204/DESY-PROC-2009-02/54

[114] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," Journal of High Energy Physics, vol. 04, p. 063, (2008).

[115] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", Nuclear Physics B, vol. 406, pp. 187âĂŞ224, (1993).

[116] Y.L. Dokshitzer, G.D. Leder, S.Moretti, and B.R. Webber, "Better jet clustering algorithms," Journal of High Energy Physics, vol. 08, p. 001, (1997).

[117] B. Dorney. "Anatomy of a Jet in CMS". Quantum Diaries. June 1st, 2011. Retrieved from https://www.quantumdiaries.org/2011/06/01/anatomy-of-a-jet-in-cms/

[118] The CMS Collaboration."Event Displays from the high-energy collisions at 7 TeV", May 2010, CMS-PHO-EVENTS-2010-007, Retrieved from https://cds.cern.ch/record/1429614.

[119] The CMS collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". JINST 6 P11002 (2011). http://dx.doi.org/10.1088/1748-0221/6/11/P11002

[120] The CMS Collaboration, "Introduction to Jet Energy Corrections at CMS.". https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC, last accessed 10.02.2018.

[121] CMS Collaboration Collaboration. "Identification of b quark jets at the CMS Experiment in the LHC Run 2". Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, (2016). https://cds.cern.ch/record/2138504.

[122] CMS Collaboration Collaboration. "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector". Tech. rep. CMS-PAS-JME16-004. Geneva: CERN, 2016. https://cds.cern.ch/record/2205284.

[123] CMS Collaboration, "New CMS results at Moriond (Electroweak) 2013", Retrieved from http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/HIG13004_Event01_0.png?itok=LAwZzPHR

[124] CMS Collaboration, "New CMS results at Moriond (Electroweak) 2013", Retrieved from http://cms.web.cern.ch/sites/cms.web.cern.ch/files/styles/large/public/field/image/TOP12035_Event01.png?itok=uMdnSqzC

[125] K. Skovpen. "Event displays highlighting the main properties of heavy flavour jets in the CMS Experiment", Aug 2017, CMS-PHO-EVENTS-2017-006. Retrieved from https://cds.cern.ch/record/2280025.

[126] G. Cowan. "Topics in statistical data analysis for high-energy physics". arXiv:1012.3589v1

[127] A. Hoecker et al., "TMVA-Toolkit for multivariate data analysis" arXiv:physics/0703039v5 (2009)

[128] L. Lista. "Statistical Methods for Data Analysis in Particle Physics", 2nd ed. Springer International Publishing. (2017) https://dx.doi.org/10.1007/978-3-319-62840-0

[129] I. Antcheva et al., “ROOT-A C++ framework for petabyte data storage, statistical analysis and visualization ,” Computer Physics Communications, vol. 182, no. 6, pp. 1384âĂŞ1385, (2011).

[130] Y. Coadou. “Boosted decision trees”, ESIPAP, Archamps, 9 February 2016. Lecture. Retrieved from `https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208-BDT.pdf`

[131] J.H. Friedman. “Greedy function approximation: A gradient boosting machine”. Ann. Statist. Volume 29, Number 5 (2001), 1189-1232. `https://projecteuclid.org/download/pdf_1/euclid.aos/1013203451`.

[132] W. Verkerke and D. Kirkby, “The RooFit toolkit for data modeling,” arXiv preprint physics, (2003).

[133] CMS Collaboration, “Documentation of the RooStats-based statistics tools for Higgs PAG”. `https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit`, last accessed on 08.04.2018.

[134] F. James, M. Roos, “MINUIT: Function minimization and error analysis”. Cern Computer Centre Program Library, Geneve Long Write-up No. D506, 1989

[135] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of statistical hypotheses”. Springer-Verlag, (1992).

[136] A.L. Read. “Modified frequentist analysis of search results (the CL_s method),” (2000). CERN-OPEN-2000-205.

[137] C. Palmer. “Searches for a Light Higgs with CMS”, CMS-CR-2012-215. `https://cds.cern.ch/record/1560435`.

[138] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large”, Transactions of the American Mathematical society, vol. 54, no. 3, pp. 426âĂŞ482, (1943).

[139] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, European Physical Journal C, vol. 71, p. 1554, (2011).

[140] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, Annals of Mathematical Statistics, vol. 9, pp. 60-62, (03, 1938).

[141] B. Hespel, F. Maltoni, and E. Vryonidou, “Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM”, JHEP 06 (2015) 065, `https://dx.doi:10.1007/JHEP06(2015)065`, arXiv:1503.01656.

[142] ATLAS Collaboration, “Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC”, Phys. Lett. B726 (2013) 88-119, doi:10.1016/j.physletb.2014.05.011,10.1016/j.physletb.2013.08.010, arXiv:1307.1427. [Erratum: Phys. Lett.B734,406(2014)].

[143] CMS Collaboration, “Search for the associated production of a Higgs boson with a single top quark in proton-proton collisions at $\sqrt{s} = 8$ TeV”, JHEP 06 (2016) 177,doi:10.1007/JHEP06(2016)177, arXiv:1509.08159.

[144] B. Stieger, C. Jorda Lope et al., “Search for Associated Production of a Single Top Quark and a Higgs Boson in Leptonic Channels”, CMS Analysis Note CMS AN-14-140, 2014.

[145] M. Peruzzi, C. Mueller, B. Stieger et al., “Search for ttH in multilepton final states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-16-211, 2016.

[146] CMS Collaboration, “Search for H to bbar in association with a single top quark as a test of Higgs boson couplings at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-16-019, 2016.

[147] CMS Collaboration, “Search for production of a Higgs boson and a single top quark in multilepton final states in proton collisions at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-17-005, 2016.

[148] CMS Collaboration, “PdmV2016Analysis,” (2016). `https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmV2016Analysis#DATA`, last accessed 11.04.2016.

[149] M. Peruzzi, F. Romeo, B. Stieger et al., “Search for ttH in multilepton final1 states at $\sqrt{s} = 13$ TeV”, CMS Analysis Note CMS AN-17-029, 2017.

[150] B. Maier, “SingleTopHiggProduction13TeV”, February, 2016. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SingleTopHiggsGeneration13TeV`.

[151] B. WG, “BtagRecommendation80XReReco”, February, 2017. `https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco`.

[152] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC Run 2”, CMS Physics Analysis Summary CMS-PAS-BTV-15-001, 2016.

[153] CMS Collaboration, “Baseline muon selections for Run-II.” `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonIdRun2`, last accessed on 24.02.2018.

[154] G. Petrucciani and C. Botta, “Two step prompt muon identification”, January, 2015. https://indico.cern.ch/event/368007/contribution/2/material/slides/0.pdf.

[155] H. Brun and C. Ochando, “Updated Results on MVA eID with 13 TeV samples”, October, 2014. https://indico.cern.ch/event/298249/contribution/3/material/slides/0.pdf.

[156] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic Top Quarks at the LHC”, JHEP 03 (2011) 059, https://dx.doi:10.1007/JHEP03(2011)059, arXiv:1007.2221.

[157] CMS Collaboration. “Tag and Probe”, https://twiki.cern.ch/twiki/bin/view/CMSPublic/TagAndProbe, last accessed on 02.03.2018.