

Observation of a new particle in the search for the Standard Model Higgs boson at the CMS detector

Nicholas Wardle

December 4, 2012

0.1 Introduction

Preamble, declaration of work, description etc. . .
≈ 10 pages (including list of figs/tables)

Chapter 1

Theory and Motivations

1.1 The Standard Model

Local Gauge theory + SM Lagrangian ≈ 3 pages

1.2 The SM Higgs

1.2.1 The Higgs mechanism

$\approx 2 - 3$ pages

1.2.2 Constraints and previous searches

$\approx 2 - 3$ pages Results from LEP and electroweak fits (most recent) Include
Tevatron searches ...?

1.2.3 Higgs production at the LHC

≈ 2 pages

Chapter 2

The CMS Detector

2.1 Detector components

very general description about detector geometry, components ≈ 2 pages

2.1.1 Electromagnetic Calorimeter

≈ 2 pages

2.1.2 L1 Trigger

$\approx 3 - 4$ pages if include L1 JEC work (c+p from Internal Note) otherwise remove as a subsection

Chapter 3

Higgs decaying to two photons

3.1 Object Reconstruction and Identification

3.1.1 Photon Energy Reconstruction

The natural width of the Higgs for at low mass is negligible when compared to the experimental resolution of the calorimeter.

≈ 6-8 pages in total just using information from PAS + Notes subsections on photon selection and vertex selection.

MC / data comparison plots

Event selection After applying the selection, the majority of the remaining background comprises of real

3.2 Signal Extraction

The signature for the decay $H \rightarrow \gamma\gamma$ is the presence of a narrow peak on a smoothly falling background in the invariant mass spectrum. The signal to background ratio can be dramatically increased by focusing on events falling in a window around the mass of the Higgs boson, m_H . Since this mass is unconstrained in the Standard Model, the search is performed for a range of mass hypotheses effectively sliding the signal window across the diphoton mass spectrum.

As the signal yield for a SM Higgs decaying to two photons is expected to be small, additional event information from the detector and the kinematics of the diphoton system can be used to increase the sensitivity of the search.

This section describes a MVA based approach to extracting the signal, categorizing events within a sliding signal region window based on a single event discriminator. The approach allows for use of data in sidebands to

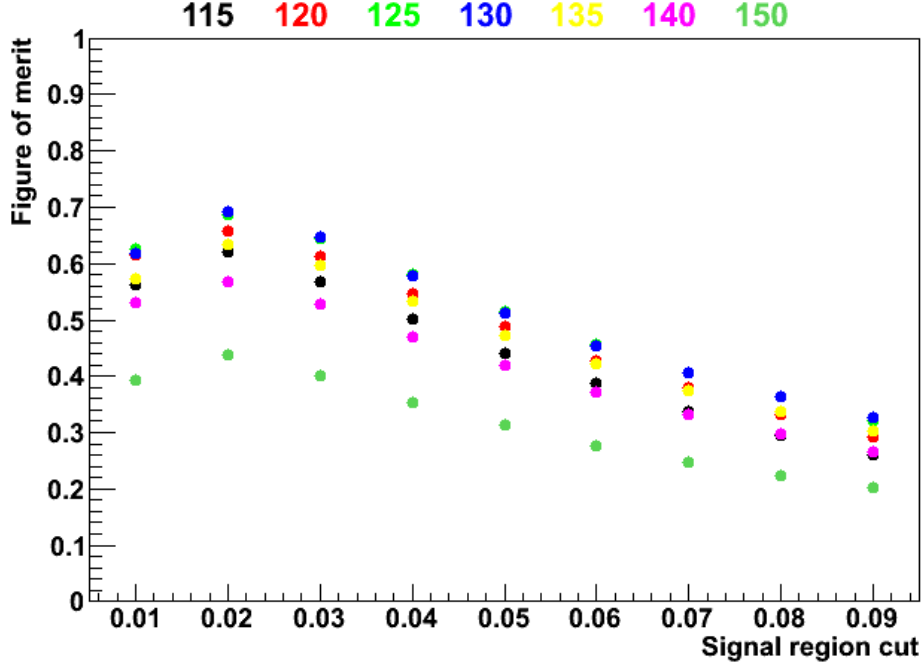


Figure 3.1: Figure of merit for selection of the signal region cut value, w . Each color shows the evaluation under different Higgs mass hypotheses.

determine expected event yields within the signal region, making little assumption about the specific composition and kinematics of the background.

3.2.1 Definition of the Signal region

Once the expected resolution of the Higgs peak is determined, the choice of signal window can be optimized to reduce the uncertainty on the background while selecting as many signal events as possible. The size of the signal window is chosen using a simplified analysis in which the number of signal events from a SM Higgs with hypothesised mass m_H expected within the range $|\Delta M/M_H| = |(m_{\gamma\gamma} - m_H)/m_H| < w$ is compared to the uncertainty on the total number of events (from background and signal) in that range. The figure of merit, $N_S/\sigma = N_S/\sqrt{\sigma_S^2 + \sigma_B^2}$, is calculated as a function of signal region cut value, w , for a range of mass hypotheses as shown in Figure 3.1. The error on the background is calculated using the procedure defined in [1] whereas the error on the signal is purely statistical.

For this analysis, $w = 0.02$ was chosen as the optimal signal region cut

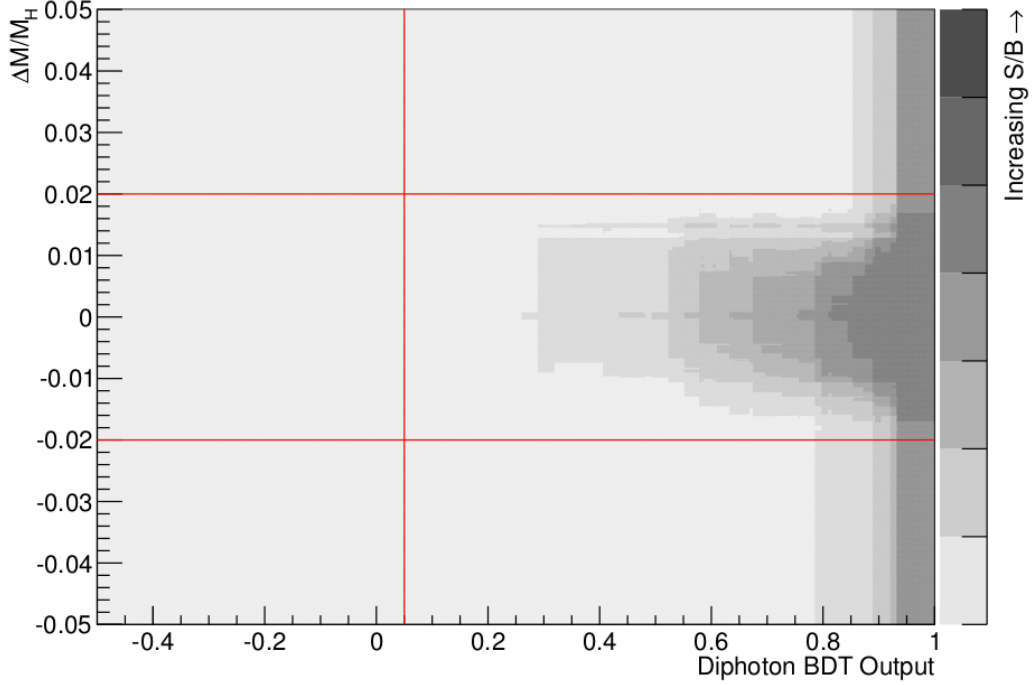


Figure 3.2: Signal to background ratio as a function of diphoton BDT output and $\Delta M/M_H$. The red lines indicate the cuts applied before the training and for applying the event selection.

value.

3.2.2 Event BDT Descriminator

The inputs to the diphoton BDT contain information from the event kinematics and the quality of the photons and vertex location in the form of the photon ID MVA output and event resolution estimators. The output of the diphoton BDT combined with the invariant mass of the diphoton system therefore provides the necessary information to separate signal from background.

Figure shows the variation in the signal to background ratio across different regions in the two-dimensional plane defined by the output of the diphoton MVA and $\Delta M/M_H$.

The two variables are combined to produce a single event discriminator by training a BDT with TMVA using the diphoton BDT output and $\Delta M/M_H$

as inputs. The BDT is trained with Higgs signal MC with $m_H = 123$ GeV including all four production processes and background MC including prompt-prompt, prompt-fake and fake-fake events.

The performance of several different training methodologies were compared to find which gave the optimum separation of signal and background. Two different choices of boosting were studied, one in which BLAH BLAH (adaptive boosting) and the other BLAH BLAH (gradient boosting). In addition, these were compared to a simple likelihood and decorrelated likelihood (which considers at most linear correlations between the variables) as shown in Figure The gradient boosting method was found to give the best performance although the variation between methodologies was small.

In this analysis, the background is estimated entirely from data. This means that any disagreement between data and MC will only effect the performance of the BDT and not the validity of the final results. The agreement between the data and MC is shown in Figure 3.3 for a mass hypothesis, $m_H = 145$ GeV. The level of agreement is sufficient so as not to require indepth study of the BDT output distribution of the background MC.

3.2.3 Binning of the BDT Output Distribution

The BDT provides a single variable with which to classify events based on their signal to back- ground ratio which will have a discrete number of response values based on the number of trees used. The boosting procedure provides a pseudo-continuous distribution which is used to model the signal and background. However, the resulting distribution will still be only pseudo-continuous. In addition, the BDT response does not directly correspond to a physical distribution and it is therefore difficult to motivate any parameterisation of either the signal or background distributions. To overcome these issues, a binning procedure is defined to construct templates which are used as models for the signal and background expectation as a function of BDT response range (BDT bin). This procedure is designed firstly to ensure that no bin in the data-driven background model is empty and secondly that as few bins as possible are used without reducing too much the sensitivity of the BDT to improve performance of the limit calculation and allow for better modelling of systematic variations.

The binning procedure is defined as follows:

1. The distribution of background MC is binned very finely to provide an almost discrete dataset (5000 equally spaced bins are used). The background is re-binned such that there are 20 expected events per bin at a luminosity of 5.1 fb^{-1} . The procedure starts from the highest BDT

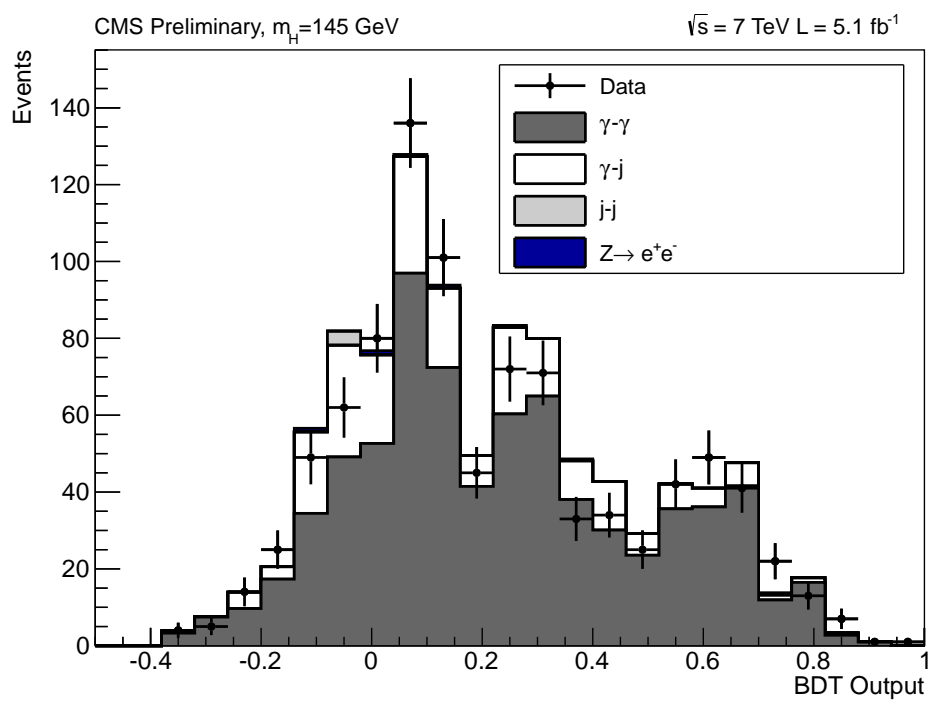


Figure 3.3: Distribution of data and MC in BDT output for $m_H = 145$ GeV, binned in constant bin width.

value bin since the final step bin may have less than 20 events. If that is the case, the last and penultimate bins are combined.

2. Smoothed versions of the signal (at each 5 GeV step mass) and background MC templates are produced in order to obtain a stable model of the signal to background ratio S/B as a function of BDT bin. The smoothing procedure is done via binning a fit (of a 9th order polynomial) to the signal distribution. Other smoothing techniques were found to give less stable performance.
3. N bin edges (boundaries), b_i , are defined on the remaining bins such that $N + 1$ bin-ranges (categories) are formed with $b_1 < b_2 < \dots < b_N$. A full scan is performed varying the bin edges to find the maximum expected significance in the presence of a SM Higgs signal.
4. An extra boundary is added and the scan is repeated and the maximum expected significance is found for $N + 1$ boundaries. If the maximum expected significance is increased by more than 0.1% compared to that of step 3, the new boundary is kept and step 4 is repeated, if not, the procedure terminates.

The scan in step 3 is split into two parts, first using a large step size to find the region where the maximum lies and then small steps within that region. The ratio of small to large step size is chosen to be that which minimizes the total number of iterations in the scan reducing the time taken for the procedure. An example of the binning procedure is shown in Figure 3.4. The red histogram is the S/B distribution after step 1, the blue after step 2 and the black vertical lines show the final set of 7 bins chosen for this analysis.

3.2.4 Background modelling

\approx 4-5 pages. Datafits to sidebands, systematic studies on background normalization and shape (linear, 2nd order pol ...)

3.2.5 Signal modelling

\approx 4-5 pages describing systematics and how they are included into the model. Describe relevant measurements from Z's.

3.2.6 7TeV results

\approx 1 pages limits, pvalues ... Since the mass of the Higgs is unknown, the data are tested against a range of possible mass hypotheses

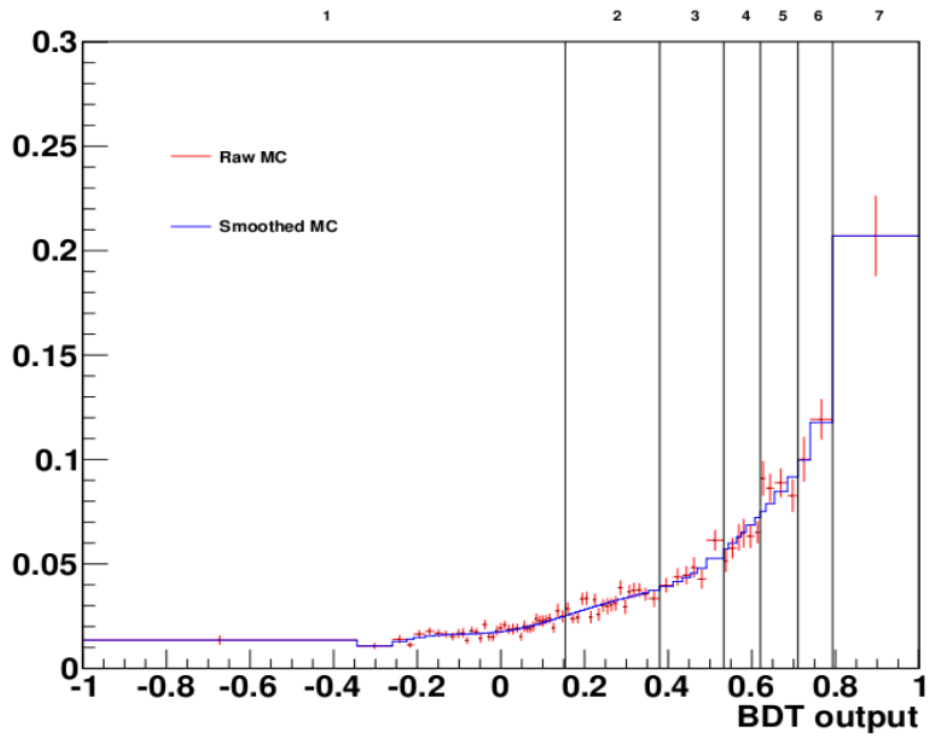


Figure 3.4: Signal to background ratio as a function of BDT output bin. The red and blue histograms show the distribution after applying step 1 of the binning procedure before and after smoothing respectively. The black vertical lines indicate the boundaries of the final binning choice from the full procedure.

3.3 8TeV results

\approx 1-2 pages, describing changes to analysis in 2012, results from 2012 + combined with 2011

Chapter 4

Higgs Combinations and Properties

4.1 Statistical interpretation of data

\approx 2-3 pages explaining frequentist confidence intervals, CLs, p-values etc. . .

4.1.1 Diagnostics

\approx 2-5 pages depending on how much just gets lifted from the note.

4.1.2 Combined Higgs search results

\approx 3 pages results from sub-combinations, full combination, discovery

4.2 Properties

\approx 10 pages . . . unknown yet!

Chapter 5

Conclusions

≈ 1 page