

Searching for SUSY in events with Jets and Missing Transverse Energy using α_T with the CMS Detector at the LHC

Zoe Hatherell

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to Imperial College London
December 2011

Contents

Acknowledgements	2
Contents	2
1 Searching for Supersymmetry with α_T in all-hadronic events	3
1.1 Samples	4
1.1.1 Monte Carlo Simulation	4
1.1.2 Data Sample	5
1.2 Analysis Framework	6
1.3 Trigger	6
1.4 Object Definitions	9
1.4.1 Good Event Definition	9
1.4.2 Jets	9
1.4.3 H_T and \cancel{H}_T	9
1.4.4 Muons	10
1.4.5 Electrons	10
1.4.6 Photons	11
1.5 Pre-Selection	12
1.6 Final Signal Selection	13
1.7 An H_T Shape Analysis	13
1.8 Hadronic Signal Region Results	14
1.8.1 Data to Monte-Carlo Comparisons	14
1.8.2 R_{α_T} on H_T	16
1.8.3 Composition of Selected $t\bar{t}$ + jets and W + jets Background Events	19
1.9 Estimation of $t\bar{t}$ and W + Jets Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.	20
1.9.1 μ Control Sample Selection	20
1.9.2 Prediction Calculation	22
1.9.3 μ Control Sample Systematic Uncertainty	26
1.9.4 Full electroweak background prediction cross check	29
1.10 Estimation of $Z \rightarrow \nu\bar{\nu}$ + jets background using photon + jets events	30
1.10.1 Z Background Prediction Calculation	30

1.10.2	γ Control Sample Systematic Uncertainty	32
1.10.3	Cross-Prediction between Control Samples	33
1.11	Signal Region Systematic Uncertainties	34
1.12	Statistical Interpretation	35
1.12.1	Hadronic Signal Selection Likelihood	35
1.12.2	Expression of b^i using R_{α_T} evolution in H_T	36
1.12.3	Electroweak Control Sample Likelihoods	37
1.12.4	Presence of Signal	38
1.12.5	Total Likelihood	38
1.13	Testing the SM-only hypothesis	39
1.14	Excluding Signal in the CMSSM Plane	41
1.15	Conclusion	41
Bibliography		44

Chapter 1

Searching for Supersymmetry with α_T in all-hadronic events

The analysis presented here represents a model-independent search for new physics in the all-hadronic channel, where the final state is defined by the presence of jets and missing energy. Designed to search for signs of supersymmetry whilst remaining sensitive to other new physics models, an inclusive strategy is used imposing restrictions only on the final state. Events are chosen based on their compatibility with a topology of heavy new particles pair-produced in p-p collisions, which decay through a chain with an end product which is stable and undetectable.

Isolating these new physics events from Standard Model background processes is essential in order to identify an excess. Controlling the dominant background from QCD Multijet processes is the central feature of the strategy, implementing use of the powerful discriminant, the α_T variable described in Chapter ???. The remaining backgrounds from electroweak processes may then be accounted for using data-driven estimation techniques in muon and photon control samples. The analysis presented here was done in 2011 and uses 1.1fb^{-1} data, representing an update on the previous iteration of this analysis using 35pb^{-1} 2010 data, which will be periodically referred to and is fully documented at [1].

1.1 Samples

This analysis uses datasets both from Monte Carlo simulation (MC) and of data recorded by the CMS detector in 2011.

1.1.1 Monte Carlo Simulation

Datasets of simulated events with calculated cross-sections are required for any analysis at the LHC. The following samples are used, full details of which can be found in Appendix A.

Standard Model Background

- QCD Multijet
- W + jets
- $Z \rightarrow \nu\bar{\nu} + \text{jets}$
- $t\bar{t} + \text{jets}$
- $\gamma + \text{jets}$

CMSSM SUSY Signal

For the purpose of understanding the possible yields from CMSSM SUSY, two mSUGRA parameter points are used. CMS has a dedicated set of 10 Low Mass(LM) points designed for initial data-taking from which we have chosen LM4 and LM6, the values of which are found in Table 1.1.

mSUGRA Point	m_0	$m_{1/2}$	A_0	$\tan \beta$	$\text{sign}(\mu)$
LM4	210 GeV	285 GeV	0	10	+
LM6	85 GeV	400 GeV	0	10	+

Table 1.1: The two CMSSM SUSY signal points used and their corresponding mSUGRA parameter values.

These points are chosen for their existence above the exclusion limit set previously, shown in Figure 1.1 on the exclusion plot from the 2010 iteration of this analysis[1].

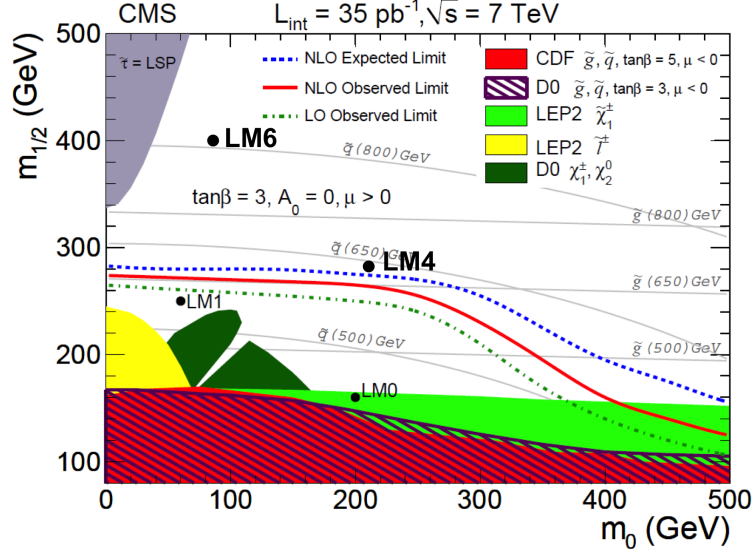


Figure 1.1: The exclusion limit set in the previous incarnation of this analysis with the 35pb^{-1} 2010 CMS dataset shown in the $m_0 - m_{1/2}$ plane with $\tan\beta = 10$, $A_0 = 0$ and $\text{sign}(\mu) = +$. Reference points LM4 and LM6 to be used in the 2011 analysis are illustrated, in the region yet to be excluded. Previously used reference points LM0 and LM1 are shown below the limit already excluded [1].

1.1.2 Data Sample

This analysis considers data collected by CMS at $\sqrt{s} = 7$ TeV in 2011 between March and June, during the data-taking period known locally as Run2011A. Analyses use only the data taken whilst CMS was fully operational, and thus the data used were specified by the certified list of “good runs” that correspond to $1.1 \pm 0.05 \text{fb}^{-1}$ of integrated luminosity [2].

As described previously, loose requirements on the types of triggers passed allow the sorting of the data into each Primary Dataset (PD). For the hadronic analysis the HT dataset is used, with basic low-threshold H_T triggers required. Higher threshold triggers are applied subsequently as part of the analysis selection, detailed in Section 1.3.

For the data-driven estimation techniques muon and photon control samples are defined. For the muon control sample, the HT PD is also used, as a low muon p_T requirement makes this better suited than the dedicated Muon PD. However the photon control sample uses the dedicated Photon PD that requires some basic threshold photon triggers to be passed. The photon control sample had slightly

lower statistics available (1.06 fb^{-1}) and henceforth a correction factor is applied to yields from this control sample to normalise to the signal region.

1.2 Analysis Framework

For the purpose of the analysis private ntuples are generated from the RECO samples using the CMSSW framework and CMS's Physics Analysis Toolkit (PAT), retaining all variables from the intermediate step of PAT-tuples recommended by the CMS SUSY group. Following this private code is used on the ntuples to perform the analysis.

1.3 Trigger

In order to select the signal events and minimise the contamination from backgrounds, a set of selection criteria is applied. As described previously in Section ??, data collected by the CMS detector is stored and organised according to the L1 and HLT trigger paths passed. While the datasets chosen in Section 1.1.2 have some basic level trigger requirements, we also require a more stringent set of requirements.

In the previous iteration of this analysis for the 2010 dataset, a set of pure H_T triggers was used. However these are unsuitable for the 2011 analysis as due to the increase in instantaneous luminosity the rate of triggers with desirable thresholds have become too high and become prescaled, rendering them unsuitable for the signal selection, although they will play a part in the control region. Moving to higher H_T thresholds is also undesirable as it would reduce a significant portion of the search region.

The use of cross-object triggers is now employed, requiring events that pass thresholds in both H_T and \cancel{H}_T for the signal region. As the era of data-taking progressed there were several menu-changes, during which time we required the lowest un-prescaled thresholds available at a given time to ensure signal yields are accurate, some of which are implemented in the menu under different versions, the relevant paths appended with `_v*`.

The first set of runs in the 2011 dataset correspond to a trigger used with thresholds in $[H_T, \cancel{H}_T] = [260, 60] \text{ GeV}$. After this the CMS standard thresholds

were shifted down by 10 GeV, relevant for the major portion of data taking. For all runs henceforth the lowest unrescaled cross-trigger had an H_T threshold of 250 GeV, during which time the \cancel{H}_T thresholds of the lowest unrescaled evolve through 60, 70 and 90 GeV.

HLT Trigger Path
HLT_HT260_MHT60_v2
HLT_HT250_MHT60_v2
HLT_HT250_MHT60_v3
HLT_HT250_MHT60_v4
HLT_HT250_MHT70_v1
HLT_HT250_MHT70_v3
HLT_HT250_MHT70_v4
HLT_HT250_MHT90_v1

Table 1.2: List of HLT trigger paths for the signal selection from which the lowest unrescaled was used or any given run.

The quantities H_T and \cancel{H}_T used in the trigger requirements differ from those used in the analysis. The trigger uses jets built in online reconstruction with uncorrected energy to form these quantities, whereas those quantities in our analysis use only jets passing the object requirements and with corrected energy. Thus it is not the case that an H_T trigger of threshold X is efficient for events with $H_T \geq X$ in the analysis. It is necessary to ensure the trigger is efficient with regard to analysis cuts in both H_T and \cancel{H}_T .

In this two step process, the efficiencies of pure H_T triggers of thresholds 250 and 260 are identified through comparison to an orthogonal sample, using a Muon H_T cross trigger. These triggers can be considered 100% efficient by an offline H_T cut of 275 GeV, and thus this is selected for the analysis. Having made this cut the efficiency of the \cancel{H}_T part may be tested, with reference to α_T (which is analogous to a cut on MHT). In the lowest bin of the analysis, $275 < H_T < 325$ GeV a small inefficiency was measured of $0.99^{+0.01}_{-0.02}$, and in all other bins $H_T > 325$ the analysis is measured as fully efficient - $1.0^{+0.00}_{-0.03}$.

This analysis makes use of those events which fail the α_T selection criteria also, as the hadronic bulk control sample. In this region the chosen signal triggers would not be efficient as we wish to use the events with low \cancel{H}_T which would not

pass that element of the trigger requirement. Here the pre-scaled H_T triggers are suitable for use, taking into account the prescale factors to gain yields in this bulk sample, an approach which is suitable due to the high statistics from QCD events. The lowest prescale of the trigger thresholds chosen for each H_T bin, shown in Table 1.3, are used at each point in the evolution of the trigger menu.

Analysis H_T Region	HLT Trigger Paths
$275 < H_T < 325$	HLT_HT250_v*, HLT_HT260_v*
$325 < H_T < 375$	HLT_HT300_v*
$375 < H_T < 475$	HLT_HT300_v*, HLT_HT350_v*
$475 < H_T < 575$	HLT_HT440_v*, HLT_HT450_v*
$H_T > 575$	HLT_HT520_v*, HLT_HT550_v*

Table 1.3: The prescaled HLT trigger paths used for each H_T region of the hadronic control sample, where $\alpha_T < 0.55$. N.B. The H_T that defines the region is built from jets with corrected energy that pass the requirements of the analysis, while the H_T quoted in the trigger definition is uncorrected and built using online reconstruction jets available to the trigger.

In the muon control sample, due to the low $p_T^\mu = 10$ GeV threshold we use the same triggers as for the hadronic signal sample. The photon sample makes use of the single photon trigger paths shown in Table 1.4, using the lowest unscaled threshold available for each given run in the data.

HLT Trigger Paths
HLT_Photon75_CaloIdVL
HLT_Photon75_CaloIdVL_IsoL
HLT_Photon90_CaloIdVL
HLT_Photon90_CaloIdVL_IsoL

Table 1.4: The list of HLT trigger paths available used to select the events for the Photon Control sample from which the lowest unscaled photon threshold is selected in any given run.

Events passing the relevant triggers proceed into the analysis selection.

1.4 Object Definitions

1.4.1 Good Event Definition

In order for an event to be considered suitable for use in physics analyses, it must be defined as a “Good Event”. Such an event is required to have at least one non-fake good primary vertex with $N_{dof} > 4$. Constraints on the vertex position along the beam axis $|z_{vtx}| < 24$ cm and perpendicular to the axis of $\rho < 2$ cm must be satisfied. Events that have many fake tracks are identified as monster events and removed, by requiring that the ratio of High Purity tracks to the total number be greater than 25% in events with more than 9 tracks.

1.4.2 Jets

The jets used in this analysis are Calo Jets, reconstructed as described in Section ?? using the anti- k_T jet clustering algorithm. In addition, a reconstructed jet must pass an additional selection in order to be considered for the analysis:

- Corrected jet transverse momentum requirement of $p_T > 50$ GeV
- Jet pseudo-rapidity $|\eta| < 3$ required to ensure within the fiducial range of the calorimeter systems.
- Passes “loose” jet identification criteria to reject jets resulting from unphysical energy using cuts in Table 1.5.

Any jet which passes the E_T and η requirements but fails the “loose” identification criteria is noted, and the event is marked as containing an “odd” jet, as the presence of such a particle reflects an event whose kinematics are poorly understood and may therefore lead to a misleading \cancel{H}_T .

1.4.3 H_T and \cancel{H}_T

The calculation of both H_T and \cancel{H}_T is performed using only the jets selected by the selection above in Section 1.4.2.

Definition	Variable	Cut
Fraction of jet energy contributed by the “hottest” hybrid photo-diode	f_{HPD}	< 0.98
Minimum number of cells required to contribute 90% of the jet energy	N_{cells}^{90}	≤ 2
Fraction of jet energy contributed by deposits in ECAL	f_{EM}	> 0.01
Balance of the energy measured in the short(E_S) and long(E_L) HF fibers.	$R_{HF} = \frac{(E_S - E_L)}{(E_S + E_L)}$ (if $p_T^{jet} > 80$ GeV)	$R_{HF} > -0.9$ ($-0.9 < R_{HF} < 1$)

Table 1.5: Set of cuts applied in “loose” CaloJet ID used to reject jets resulting from fake calorimeter deposits representing unphysical energy. DeVised using cosmic run data as a pure sample of non-collider “fake” jets, full details of which can be found in [3]

1.4.4 Muons

Although muons are not required by the analysis, a veto on them must be employed, based on muons that satisfy the following set of criteria:

- $p_T^\mu > 10$ GeV
- $|\eta| < 2.5$
- Relative Combined Isolation = $(Iso_{tracker} + Iso_{ECAL} + Iso_{HCAL})/p_T^\mu < 0.15^1$
- Passes “tight” muon identification, using cuts shown in Table 1.6.

1.4.5 Electrons

Similarly electrons are also defined for veto purposes, with the definition of an electron in the analysis as that passing the following cuts:

- $E_T^e > 10\text{GeV}$
- $|\eta| < 2.5$

¹The components $Iso_{tracker}(Iso_{ECAL}, Iso_{HCAL})$ represent the sum of $p_T(E_T)$ in the relevant detector component, calculated in a cone of $R = 0.3$ in $\eta - \phi$ around there muon trajectory. The track hits used to reconstruct the muon are not used and any muon deposit in the calorimeters is removed via a smaller veto cone.

Definition	Variable	Cut
Reconstructed with outside-in algorithm	Global Muon	Required
Reconstructed with inside-out algorithm	Tracker Muon	Required
Global muon track fit quality	χ^2	< 10
Number of hits in the silicon tracker included in track	N_{trk}^{hits}	> 10
Number of pixel hits in N_{trk}^{hits}	N_{pixel}^{hits}	> 0
Number of hits in muon system included in Global Muon	N_{muon}^{hits}	≥ 1
Transverse impact parameter with respect to vertex	d_{xy}	$< 2 \text{ mm}$

Table 1.6: Set of cuts applied in “tight” Muon ID, taken from [4]

- Combined Isolation = $(\sum_{tracker} p_T + \sum_{ECAL} E_T + \sum_{HCAL} E_T)/p_T^e < 0.15$
- Pass WP95 electron identification implemented using cuts in Table 1.7.

Definition	Variable	Barrel Cut	End-Cap Cut
RMS of the width in η of the crystals about the most energetic crystal in the seed	$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
Difference in ϕ between track and supercluster	$\Delta\phi_{vtx}$	< 0.8	< 0.7
Difference in η between track and supercluster	$\Delta\eta_{vtx}$	< 0.007	< 0.01
Ratio of HCAL energy in $\Delta R = 0.15$ to ECAL seed energy	H / E	< 0.15	< 0.07

Table 1.7: Set of cuts applied in “WP95” Electron ID, taken from [5] corresponding to an intended 95% efficiency for signal electrons in W events for electrons with $p_T > 20 \text{ GeV}$.

1.4.6 Photons

Photons in the analysis are defined by the following set of requirements:

¹where electron $p_T > 20 \text{ GeV}$, measured with a sample of $W \rightarrow e$ events.

- $p_T^\gamma > 25 \text{ GeV}$
- $|\eta| < 2.5$
- Passes “tight” photon cut-based identification (including isolation) using cuts shown in Table 1.8.

Definition	Variable	Barrel Cut	End-Cap Cut
Tracker Isolation in a cone of $R=0.4$	Iso_{trk}	$< (2.0\text{GeV} + 0.001E_T^\gamma)$	
ECAL Isolation in an outer cone of $R=0.4$ (inner cone $R=0.06$ removed).	Iso_{ECAL}	$< (4.2\text{GeV} + 0.006E_T^\gamma)$	
HCAL Isolation in an outer cone of $R=0.4$ (inner cone $R=0.15$ removed).	Iso_{HCAL}	$< (2.2\text{GeV} + 0.0025E_T^\gamma)$	
RMS of the width in η of the crystals about the most energetic crystal in the seed	$\sigma_{in\eta}$	< 0.013	< 0.030
Ratio of HCAL energy in $\Delta R = 0.15$ to ECAL seed energy	H / E	< 0.05	< 0.05

Table 1.8: Set of cuts applied in “tight” Photon ID, taken from [6]

1.5 Pre-Selection

A basic selection of events used for comparison of distributions between data and monte-carlo events shall be known as “pre-selection”, following the details set out in this section.

HCAL Barrel and End-cap (HBHE) Noise Fiter Events where noise has been identified in the HCAL are removed also, using an algorithm which checks for Photodetectors which have at east 17 out of 18 channels with an $E > 1.5 \text{ GeV}$.

Following these conditions events are then selected by the following criteria, using the definitions of physics objects according to the criteria stated previously:

- Pass triggers as detailed in Section 1.3.
- Pass Good Event selection as detailed in Section 1.4.1.

- Require events with $N_{jet} \geq 2$
- $N_{muon} = N_{electron} = 0$ to reduce the effects of missing energy from neutrinos.
- $N_{photon} = 0$ to ensure a pure hadronic set of events.
- Events with > 0 “odd” jets are vetoed.
- Additional constraint on the transverse momentum of the two leading jets $p_T^{j1}, p_T^{j2} > 100$ GeV.
- Additional constraint on the leading jet pseudorapidity $|\eta^{j1}| < 2.5$.
- $H_T \geq 275$ GeV

1.6 Final Signal Selection

After the preselection a final set of cuts is applied, including the α_T cut that defines the signal region alongside cleaning cuts which remove events that may lead to inaccurate results.

- $\alpha_T > 0.55$
- If the ratio $R_{miss} = \cancel{E}_T / \cancel{E}_T > 1.25$, the event is rejected. This protect the quantity α_T from the scenario where many jets fail the $p_T = 50$ GeV threshold thus resulting in fake \cancel{E}_T and thus misleading values of α_T .
- To prevent against fake missing energy resulting from dead or masked cells in the ECAL or the gap between the barrel and end-caps the following procedure is used: The jet most likely to be responsible for the \cancel{E}_T is found. If the angle ϕ between this jet and the vector $\vec{\cancel{E}}_T$ (known as $\Delta\phi^*$) less than 0.5 then the $\eta - \phi$ distance between the jet and the nearest masked ECAL cell is computed, along with the distance from the detector gap. If either distance is smaller than 0.3 then the event is rejected.

1.7 An H_T Shape Analysis

Previous iterations of this analysis strategy with the 35pb^{-1} 2010 LHC dataset [1] used a cut-and-count strategy for all events passing the selection, defining the

signal region by an $H_T > 375$ and using lower regions in H_T as control regions. The 2011 analysis follows the same selection but motivated by the increasing luminosity is undertaken as a Shape Analysis in bins of H_T , using the whole range $H_T > 275$ GeV as a signal region. This allows greater sensitivity to states of higher mass.

The set of lower bin edges are as follows: [275,325,375,475,574,675,775,875], where each bin is exclusive with an upper limit corresponding to the lower edge of the next bin, except in the case of the final bin which is inclusive $H_T > 875$. The background estimation techniques employed from data therefore are designed to identify the contribution in each distinct bin.

In order to include the two lowest bins in H_T it is necessary to scale the jet thresholds stated in Section 1.5 in order to maintain even event kinematics allowing a shape analysis approach. The background from $t\bar{t}$ + jets carries a bias to higher jet multiplicities compared to the other EWK components, and thus with identical jet definitions exhibits a turn on behaviour in H_T . In order to remedy this, the lowest two bins have both the p_T threshold required by definition and the additional second jet p_T requirement scaled. The scale factor is $l/375$, where l represents the bin lower edge in question, leading to the thresholds shown in Table 1.9.

H_T Region	Jet Definition	Second-Leading Jet Cut
$275 < H_T < 325$	$p_T > 36.7$ GeV	$p_T^{j1}, p_T^{j2} > 73.3$ GeV
$325 < H_T < 375$	$p_T > 43.3$ GeV	$p_T^{j1}, p_T^{j2} > 86.7$ GeV
$H_T > 375$	$p_T > 50$ GeV	$p_T^{j1}, p_T^{j2} > 100$ GeV

Table 1.9: The three different regions of jet scaling, with values indicated both for the basic definition of a jet used in the analysis, and the second-to-leading jet energy cut. The former is especially important as this alters the value of H_T as this is calculated using the jets in the event.

1.8 Hadronic Signal Region Results

1.8.1 Data to Monte-Carlo Comparisons

Distributions of the 2011 data with MC samples alongside are shown in this section. The MC samples are normalised to 1.1fb^{-1} for shape comparison and

to illustrate the accuracy of modelling provided, although these are not used in background estimation as data control samples are used later.

In Figure 1.2 distributions of H_T and the jet multiplicity (N_{Jet}) are shown for events that pass the pre-selection with an additional cut of $\cancel{E}_T > 100$ GeV to ensure trigger efficiency. For simplicity only bins with $H_T > 375$ GeV have been included in the plots, so as to maintain one set of jet thresholds. There is good agreement in both cases variables, with no noticeable shape disagreement. Using events with the same selection, Figure 1.3(a) shows the high discriminatory power of the α_T variable between the QCD “fake” \cancel{E}_T background and signal events with real \cancel{E}_T . The region $0.46 < \alpha_T < 0.6$ is expended in Figure 1.3(b), illustrating the rapid QCD fall-off to zero that motivates the chosen cut value of 0.55.

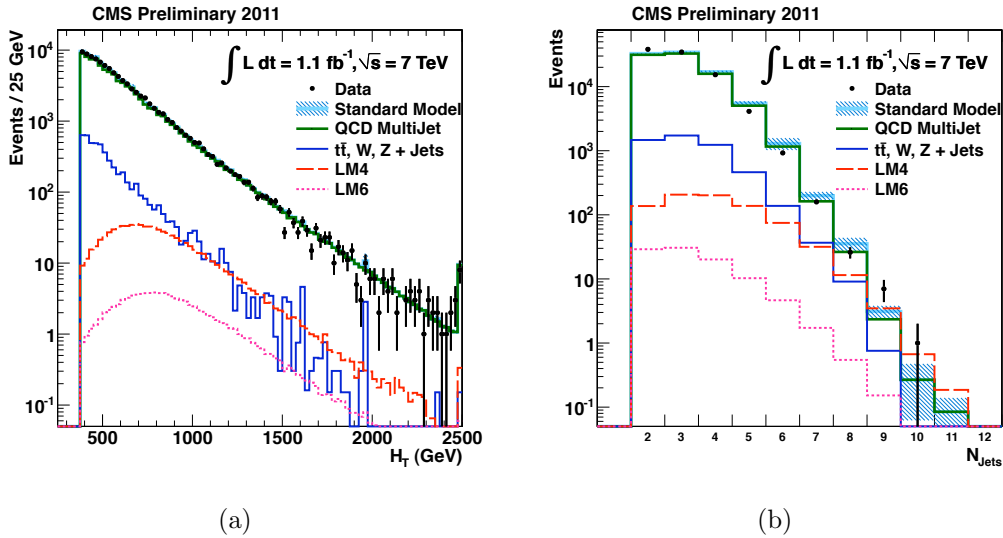


Figure 1.2: Distributions of (a) H_T , (b) N_{Jet} , showing comparisons of 1.1 fb^{-1} 2011 7TeV CMS Data and equivalently weighted Monte-Carlo prior to the α_T selection cut, for $H_T \geq 375$ GeV and $\cancel{E}_T > 100$ GeV. SUSY Signal reference points LM4 & LM6 shown for illustration of potential yields.

The distributions of jet multiplicity, $\Delta\phi^*$ and M_{eff} after the final selection cuts are applied can be seen respectively in Figure 1.4. Data shows a good overall comparison to the Standard Model MC, although here statistics are more limited accounting for fluctuations. The $\Delta\phi^*$ distribution in Figure 1.4(c) is consistent with the expectation that the α_T cut completely eradicate contamination from QCD events, as any evidence of such would lead to a peak at low values, instead of the flat behaviour seen. In addition no notable excess can be seen of data

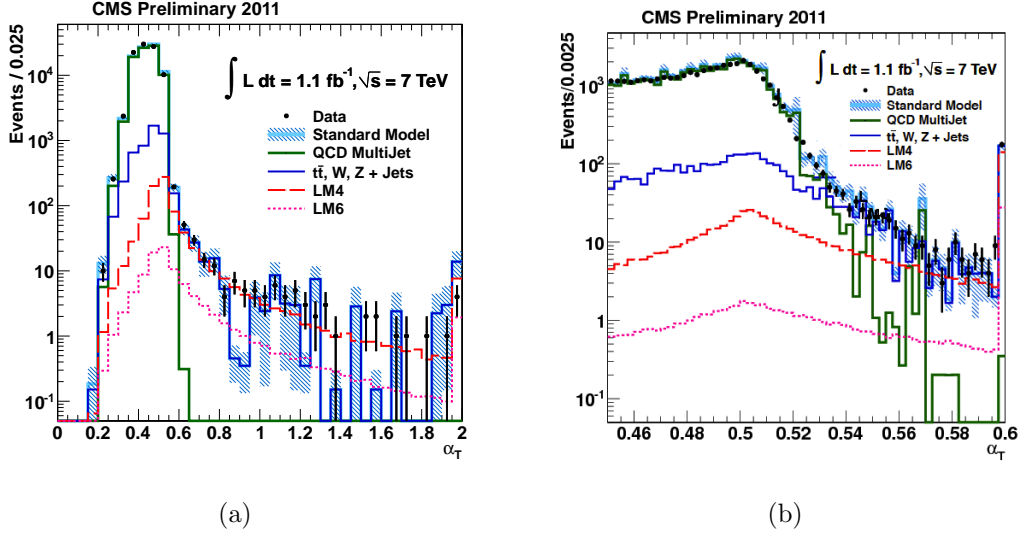


Figure 1.3: Distributions of α_T showing comparisons of 1.1 fb^{-1} 2011 7TeV CMS Data and equivalently weighted Monte-Carlo prior to the α_T selection cut, for $H_T \geq 375 \text{ GeV}$ and $\cancel{H}_T > 100 \text{ GeV}$. The α_T distribution is shown fully (left) and also shown zoomed (b) in the region $0.46 < \alpha_T < 0.6$. SUSY Signal reference points LM4 & LM6 shown for illustration of potential yields.

over MC although this observation is merely an aside, as a more detailed shape analysis will be used to evaluate this quantitatively in later sections.

1.8.2 R_{α_T} on H_T

Rather than a simple cut-and-count experiment a shape analysis in H_T bins as defined in Section sec:shape is desirable to look simultaneously in a large signal region in H_T . This is performed using the properties of the R_{α_T} variable, as discussed in Section ??, due to its unique properties separating the three possible components in its numerator: contamination from QCD, EWK backgrounds and SUSY signal events.

Hadronic Bulk Control Selection

While the numerator is defined by the final selection defined previously the denominator, which shall be known as the hadronic bulk control region, is defined by the pre-selection only, with the additional change of triggers essential as the cross triggers would put an inadvertent \cancel{H}_T cut, biasing the α_T distribution. Here, as it is a control region we use a suite of pre scaled H_T triggers described previously.

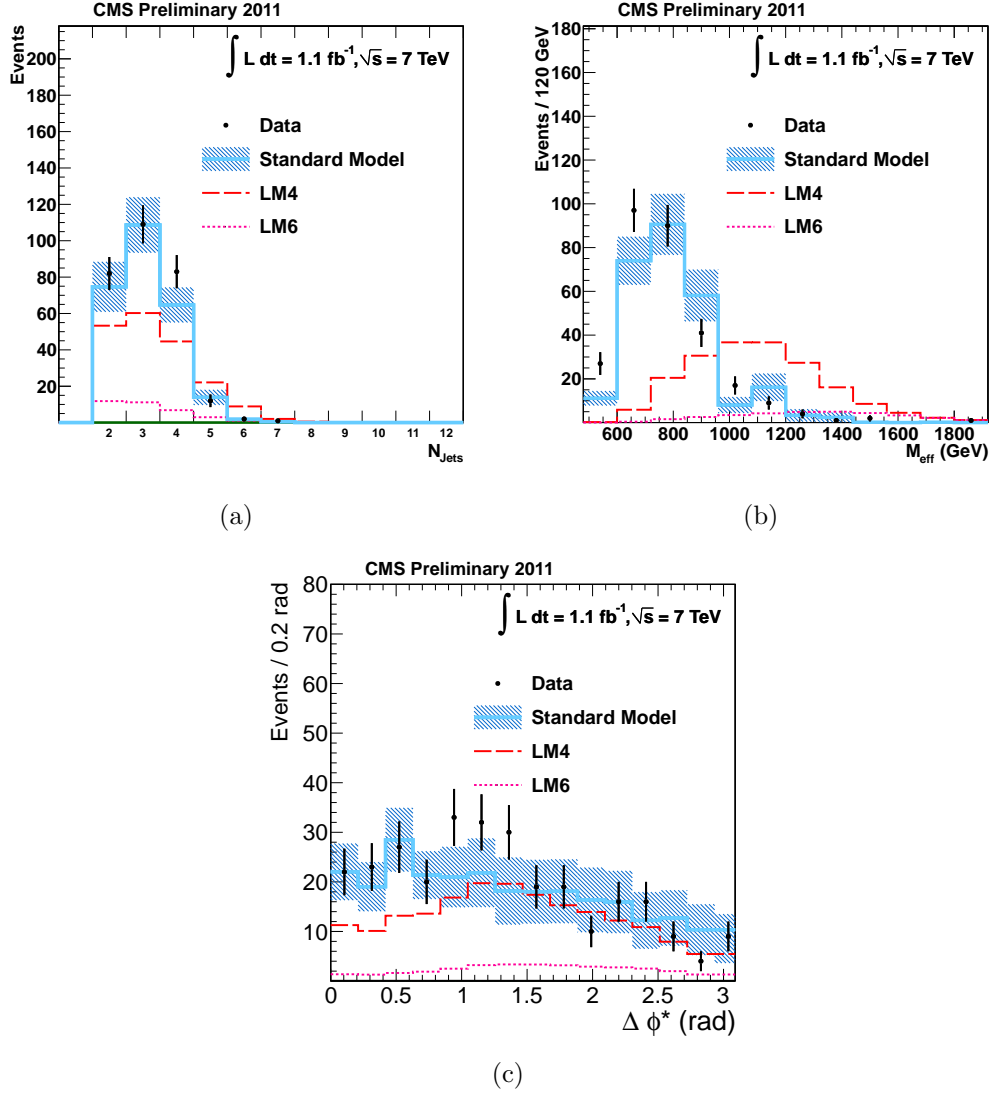


Figure 1.4: Distributions of (a) Jet Multiplicity, (b) $M_{\text{eff}} (= H_T + \cancel{H}_T)$ and (c) $\Delta\phi^*$ showing comparisons of 1.1 fb^{-1} 2011 7TeV CMS Data and equivalently weighted Standard Model Monte-Carlo in basic kinematic quantities after the full α_T selection. SUSY Signal reference points LM4 & LM6 shown for illustration of potential yields.

It is important to remove not just the α_T cut but all cuts of the final level selection as the cleaning cuts may introduce a bias to high missing energy. The bin-by-bin 1.1fb^{-1} yields for the hadronic signal and hadronic bulk selections are found in Table 1.10, along with each corresponding value for R_{α_T} .

H_T Bin (GeV)	275–325	325–375	375–475	475–575
$\alpha_T > 0.55$	782	321	196	62
$\alpha_T < 0.55$	$5.73 \cdot 10^7$	$2.36 \cdot 10^7$	$1.62 \cdot 10^7$	$5.12 \cdot 10^6$
$R_{\alpha_T}(10^{-5})$	$1.36 \pm 0.05_{\text{stat}}$	$1.36 \pm 0.08_{\text{stat}}$	$1.21 \pm 0.09_{\text{stat}}$	$1.21 \pm 0.15_{\text{stat}}$

H_T Bin (GeV)	575–675	675–775	775–875	875– ∞
$\alpha_T > 0.55$	21	6	3	1
$\alpha_T < 0.55$	$1.78 \cdot 10^6$	$6.89 \cdot 10^5$	$2.90 \cdot 10^5$	$2.60 \cdot 10^5$
$R_{\alpha_T}(10^{-5})$	$1.18 \pm 0.26_{\text{stat}}$	$0.87 \pm 0.36_{\text{stat}}$	$1.03 \pm 0.60_{\text{stat}}$	$0.39 \pm 0.52_{\text{stat}}$

Table 1.10: The number of events passing and failing the α_T cut and the resulting R_{α_T} value divided into H_T bins, for 1.1fb^{-1} of data collected in 2011.

The behaviour of R_{α_T} over the range of H_T bins is further explored in Figure 1.5(a), showing the values measured from data (black) alongside those derived from SM MC simulation events in three cases: without SUSY signal (red), with LM4 signal (blue) and with LM6 signal (green). The data is consistent with a hypothesis of a flat line with a p-value of 0.29, as is the MC SM only with a p-value of 0.50. The inclusion of the LM4(LM6) signal MC events renders the distribution non-consistent with a flat hypothesis, as expected. As it is known that QCD contamination in the numerator leads to an exponentially falling R_{α_T} as H_T increases we find the data consistent with our hypothesis that the signal sample is free of QCD contamination.

To test the sensitivity of the shape of R_{α_T} to the cross-sections used in SM MC, Figure 1.5(b) shows the H_T dependence with the effective cross-sections of the major EWK backgrounds varied individually by $\pm 15\%$, to encompass the level of our current certainty of their values. In all cases the behaviour continues to be consistent with a flat hypothesis, the worst case p-value of which is 0.47 confirming the validity of the conclusion of flat behaviour. As results support the theory that α_T has removed the QCD background, no dedicated background estimation technique is needed, although a small QCD contribution will be allowed in the eventual fit to ensure that there is no bias and to account for any small remaining

contribution.

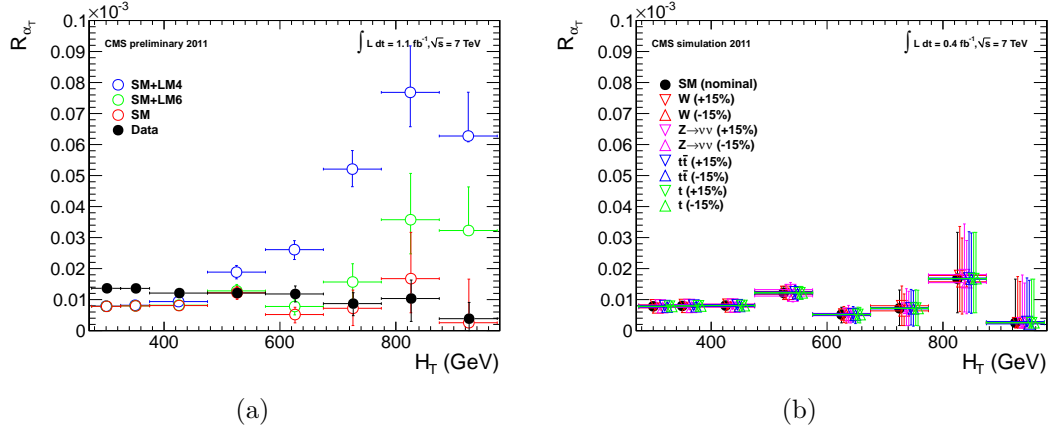


Figure 1.5: (Left) The dependence of R_{α_T} on H_T for events with $N_{\text{jet}} \geq 2$. (Right) Dependence of R_{α_T} on H_T when varying the effective cross-section of the four major EWK background components individually by $\pm 15\%$. (Markers are artificially offset for clarity.)

Having eradicated the majority of the QCD background the selection is left with a remaining dominating background stemming from electroweak processes where real missing energy is created in the form of neutrinos. There are three major backgrounds relevant, $Z + \text{jets}$, $W + \text{jets}$ and $t\bar{t}$. Events with $Z + \text{jets}$ form a true irreducible background with decay to $\nu\bar{\nu}$ producing a true event with jets and missing energy, and hence requires a method of estimation to quantify its contribution to the yield. A data-driven method is used with the construction of a $\gamma + \text{jets}$ control sample.

It can be seen that there is also an irreducible background from $W + \text{jets}$ and $t\bar{t}$ events, both of which in the kinematic space of the final selection concern the decays of boosted W 's. It is interesting to understand the composition of the decays responsible, as these should be leptonic and therefore easy to identify and eliminate with our vetoes.

1.8.3 Composition of Selected $t\bar{t} + \text{jets}$ and $W + \text{jets}$ Background Events

Figure 1.6 shows the breakdown of the decays for the $t\bar{t}$ component (a), the W component (b), and for the combined $t\bar{t}$ - W background (c) after all cuts in the

final selection. The category responsible for the greatest number of events involves the decay $W \rightarrow \tau\nu$, where the tau lepton decays hadronically. In this case the tau is identified as a jet and as such fulfils the selection criteria, accounting for 41.5% of the events. The contribution from W decays to $e\nu(\mu\nu)$ where the $e(\mu)$ is outside the p_T and $|\eta|$ acceptance of the analysis account for 19.6%(21.8%), a total of 41.4% divided evenly between the flavours. Another 15.2% of events represent the veto inefficiency where the W decays to $e, \mu\nu$ within this acceptance but failing the quality criteria required (isolation, ID), the larger proportion of which is from the electron veto. There is an additional 1.8% of fully leptonic decays where more than one lepton is missed, coming from the $t\bar{t}$ source only in the case that both W bosons decay leptonically. The remaining 0.1%, representing a negligible effect (< 1 event for 1.1fb^{-1}) consist of $t\bar{t}$ decays in which non-isolated leptons are produced within jet fragmentation and meson decay, therefore regarded as a fully hadronic decay despite the existence of missing energy from neutrinos.

1.9 Estimation of $t\bar{t}$ and $W + \text{Jets}$ Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.

In order to estimate the background contribution resulting from these boosted W decays from $W + \text{jets}$ and $t\bar{t} + \text{jets}$ events, a control sample is used. Here energetic W bosons that decay through a muon-neutrino pair are selected in the kinematic region of the search in order to extrapolate the expected yield in the hadronic selection.

1.9.1 μ Control Sample Selection

The aim of the μ control selection is to select events kinematically similar to those serving as background to the hadronic signal sample, but in the case of a well-identified muon, ensuring orthogonality with the signal selection. The muon veto mentioned earlier is replaced with a requirement for one and only one μ in the event, and the isolation requirement in the definition of a μ is tightened to 0.1 (compared with the hadronic analysis value of 0.15) to ensure a high purity sample of well reconstructed isolated μ 's. The final level cuts are also included

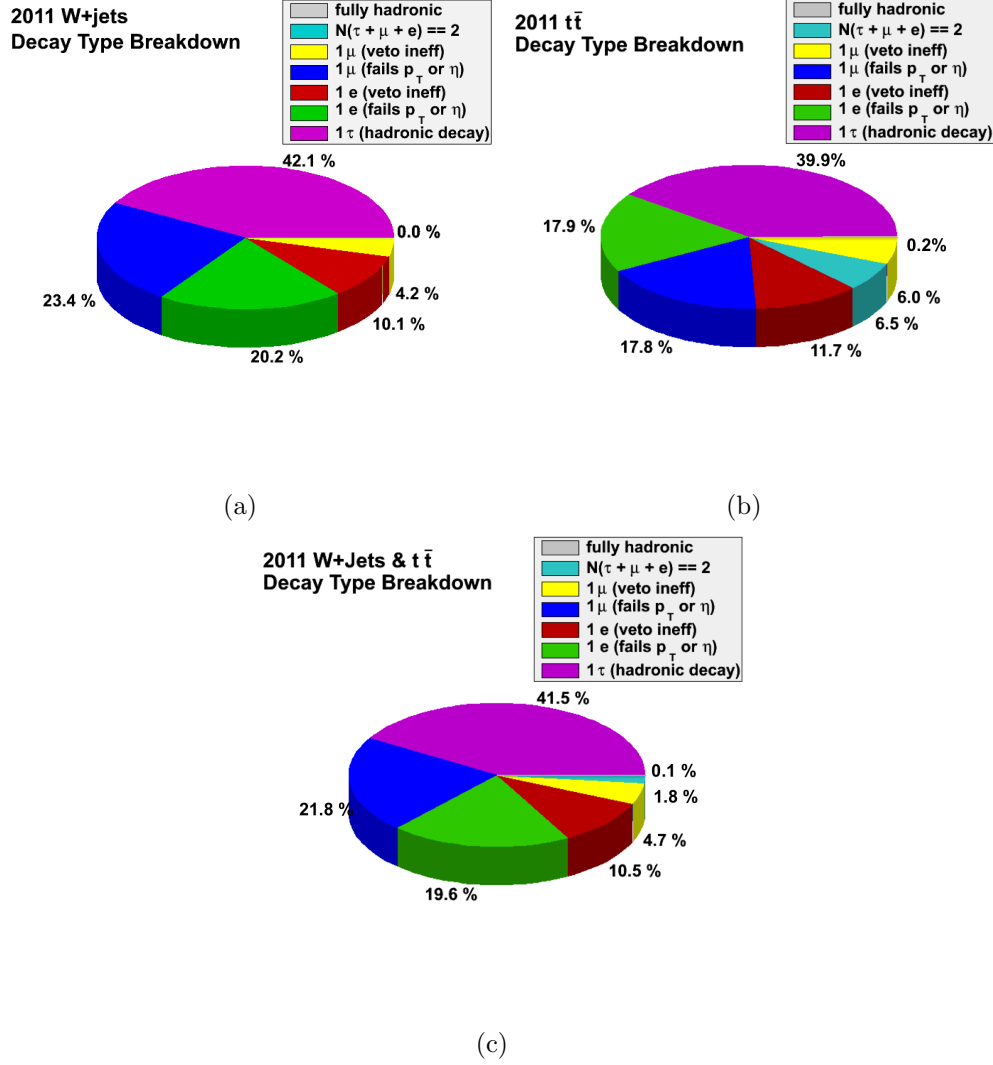


Figure 1.6: Type breakdown of decays resulting in a $W + \text{jets}$ and $t\bar{t} + \text{jets}$ event selected by the hadronic signal selection. Shown using Monte-Carlo truth information separately for $W + \text{jets}$ (a) and $t\bar{t} + \text{jets}$ (b) events, and both combined in the full $t\bar{t}$ - W background (c).

in this sample.

An additional set of requirements is also included with the muon requirement to select only events with kinematics that fit the decay $W \rightarrow \mu\nu$:

- $M_T > 30$ GeV where M_T is synonymous with the transverse mass of the W candidate.
- $\Delta R(\text{jet}, \text{muon}) > 0.5$.
- $\cancel{H}_T/H_T > 0.4$ placing an effective cut on p_T^W as this is synonymous with \cancel{H}_T .
- No second isolated muon outside of acceptance, reducing the contamination from $Z \rightarrow \mu\mu$.

Kinematic distributions of the events selected for the μ control sample are shown in Figure 1.7 prior to application of the α_T cut to demonstrate the agreement between data and Monte Carlo with high statistics. As the muon selection implicitly requires a cut on \cancel{H}_T/H_T that at its lowest ensures $\cancel{H}_T > 110$ GeV, no additional requirement is needed to ensure the trigger is efficient at the pre- α_T stage.

Agreement of data to $t\bar{t}$ - W MC in all distributions is good and no significant excess nor shape-disagreement are seen. In addition examination of the transverse mass distribution in Figure 1.7(f) confirms a peak at 80 GeV ($m^W \sim 80.385$) confirming the sample is dominated by W bosons. These observations lead to a conclusion that the sample is well modelled by Monte-Carlo and clean from contamination, confirming the validity of the selection and motivating the use of a Monte Carlo ratio in the prediction calculation. The QCD contamination is found to be negligible and therefore not seen in these distributions.

In Figure 1.8 the distributions are shown after the α_T cut, showing the same conclusions with lower statistics. This corresponds to the selection used for prediction.

1.9.2 Prediction Calculation

The prediction of the W boson decay contribution to the hadronic signal yield in the data, W_{data}^{had} , can be made from the analogous muon control yield W_{data}^{μ} providing the ratio between the hadronic and μ selections R_{μ}^{had} is known. This is

1.9. Estimation of $t\bar{t}$ and $W + \text{Jets}$ Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.

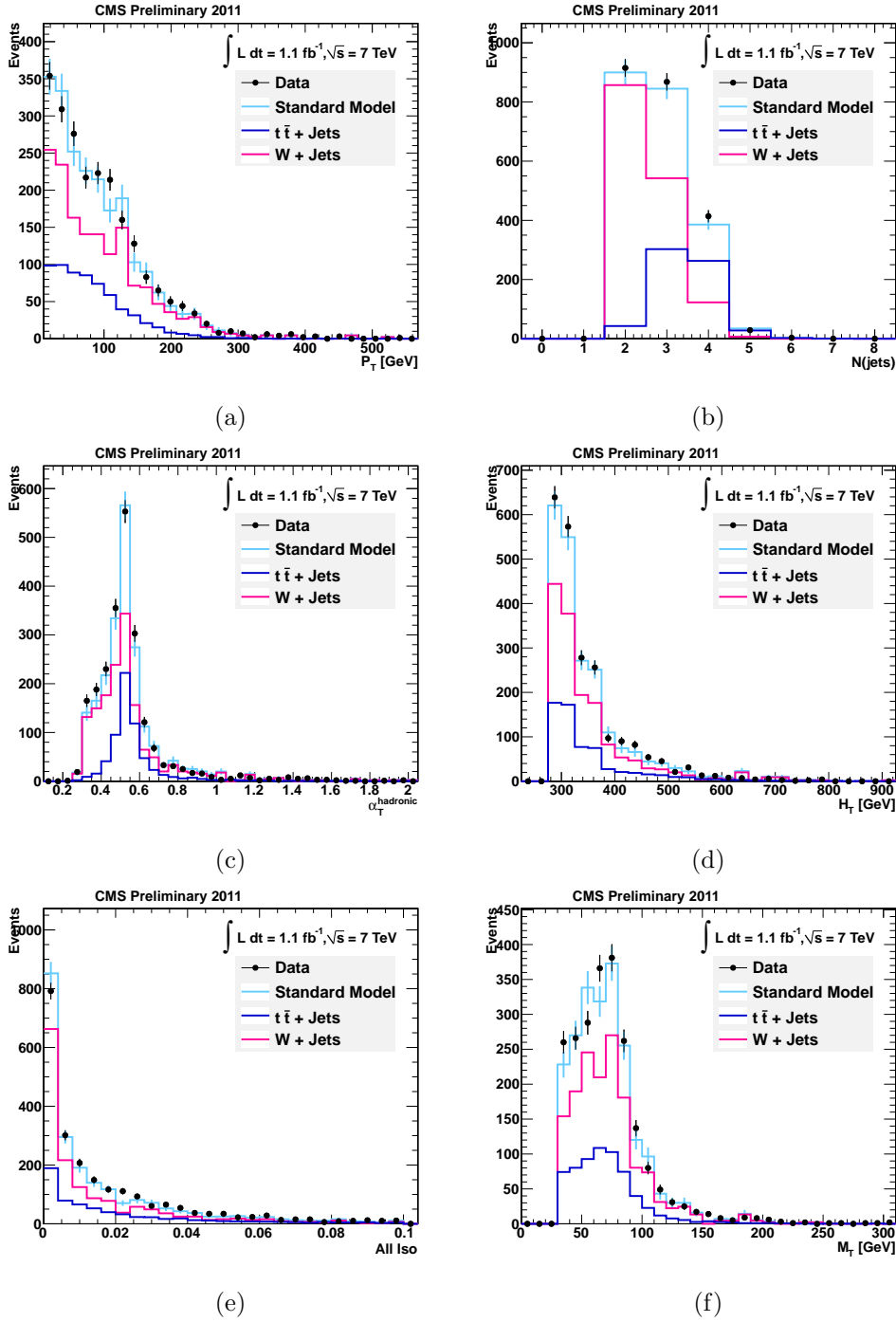


Figure 1.7: Distributions of (a) p_T^μ , (b) Jet Multiplicity ($N(\text{jet})$), (c) α_T , (d) H_T , (e) Muon Combined Isolation and (f) M_T for the μ control selection **before** the $\alpha_T > 0.55$ cut is applied. Shows comparisons of 1.1 fb^{-1} 2011 7TeV CMS Data and equivalently weighted Monte-Carlo.

1.9. Estimation of $t\bar{t}$ and $W + \text{Jets}$ Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.

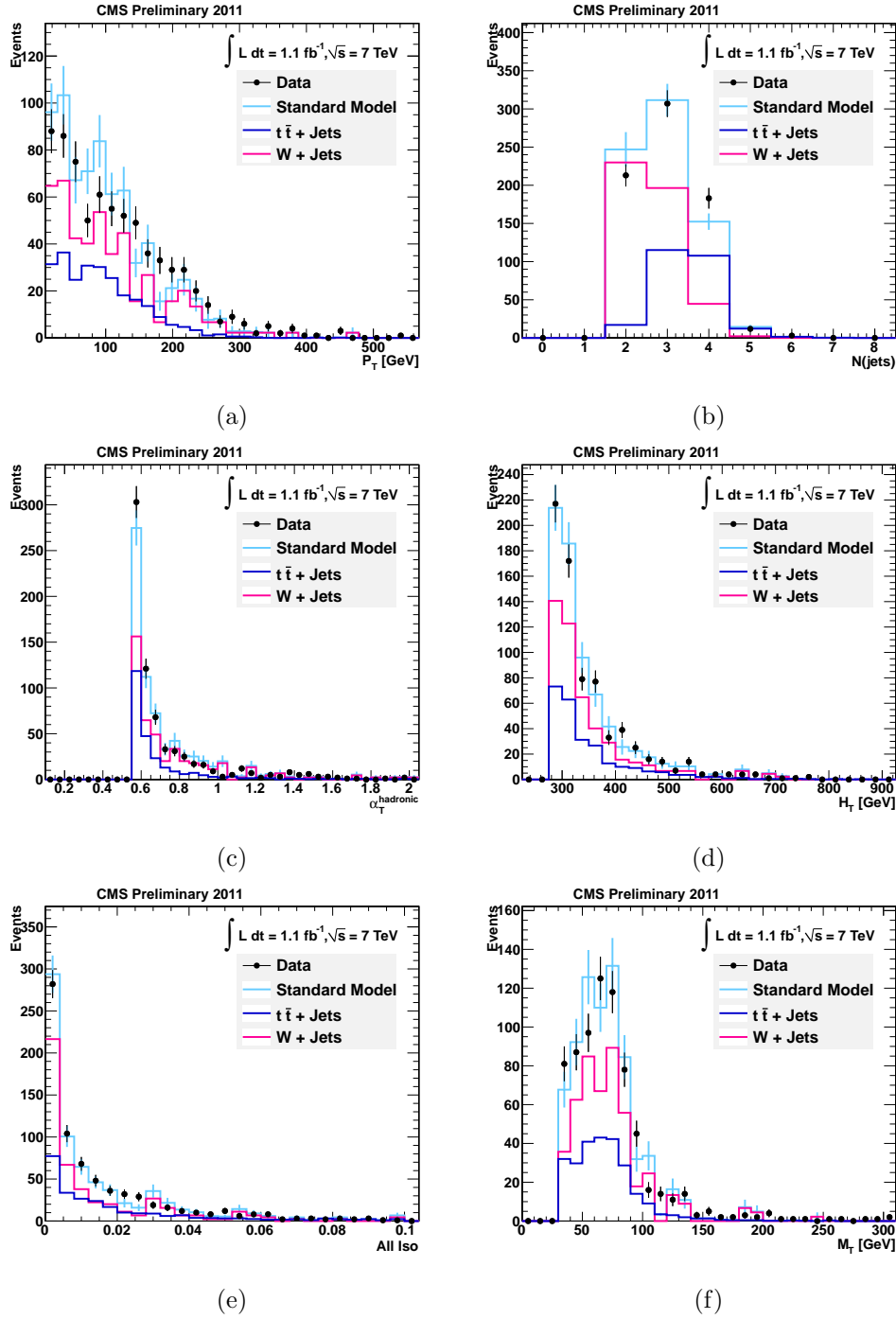


Figure 1.8: Distributions of (a) p_T^μ , (b) Jet Multiplicity ($N(\text{jet})$), (c) α_T , (d) H_T , (e) Muon Combined Isolation and (f) M_T for the μ control selection **after** the $\alpha_T > 0.55$ cut is applied. Shows comparisons of 1.1 fb⁻¹ 2011 7TeV CMS Data and equivalently weighted Monte-Carlo.

taken from the events passing each selection in the $t\bar{t} + \text{jets}$ and $W + \text{jets}$ Monte Carlo simulations W_{MC}^{had} and W_{MC}^μ , which corrects for the selection efficiencies and acceptance. The full estimation is made using Equation 1.1.

$$W_{data}^{had} = W_{data}^\mu \times R_\mu^{had} = W_{data}^\mu \times \left(\frac{W_{MC}^{had}}{W_{MC}^\mu} \right) \quad (1.1)$$

Calculating the contribution separately for each hadronic bin in this way results in the values for the ratio R_μ^{had} as shown in Table 1.11. As there are low MC statistics in the highest bins the errors become large, which affects the error of a prediction made. In addition, the values seem to have no trend, as expected as the behaviour of the backgrounds is not expected to change in H_T . Thus in order to improve results, one ratio R_μ^{had} is calculated for $H_T > 375$ GeV and used in the six highest bins, to provide six individual bin estimates. The two lowest bin estimates are calculated using exclusive ratios, as the MC statistics are sufficient.

H_T Bin (GeV)	R_μ^{had}
275–325	$1.14 \pm 0.06_{\text{stat}}$
325–375	$0.96 \pm 0.07_{\text{stat}}$
375–475	$0.88 \pm 0.09_{\text{stat}}$
475–575	$0.90 \pm 0.15_{\text{stat}}$
575–675	$1.31 \pm 0.37_{\text{stat}}$
675–775	$0.64 \pm 0.29_{\text{stat}}$
775–875	$0.34 \pm 0.27_{\text{stat}}$
875– ∞	$2.00 \pm 1.97_{\text{stat}}$
375–∞	$0.90 \pm 0.07_{\text{stat}}$

Table 1.11: The bin-by-bin Monte Carlo ratios R_μ^{had} for the 8 H_T bins of the signal region, along with the ratio calculated for the inclusive region encompassing the 6 highest bins. The ratios shown in **bold** indicate those chosen for the prediction calculation to minimise additional error propagation to the prediction introduced by low MC statistics.

The bin-by-bin results including prediction are shown in Table 1.12, with the normalised MC event yields that contribute to R_μ^{had} per bin (although the ratio quoted for the 6 higher bins is the overall ratio as described earlier) and the yields in data for the μ selection at 1.1fb^{-1} . Both statistical errors and systematic errors are quoted, the latter corresponding to a 30% uncertainty taken directly from the 2010 analysis, the calculation and validity of which are described below.

1.9. Estimation of $t\bar{t}$ and $W + \text{Jets}$ Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.

H_T Bin (GeV)	275–325	325–375	375–475	475–575
MC $W + t\bar{t}$	$463.0 \pm 16.0_{\text{stat}}$	$171.2 \pm 9.5_{\text{stat}}$	$116.3 \pm 8.3_{\text{stat}}$	$43.7 \pm 5.1_{\text{stat}}$
MC $\mu + \text{jets}$	$407.5 \pm 14.5_{\text{stat}}$	$179.1 \pm 9.6_{\text{stat}}$	$131.6 \pm 8.8_{\text{stat}}$	$48.7 \pm 5.5_{\text{stat}}$
MC Ratio	1.14	0.96	0.90	0.90
Data $\mu + \text{jets}$	389	156	113	39
$W + t\bar{t}$ Prediction	$442.0 \pm 22.4_{\text{stat}}$ $\pm 132.6_{\text{syst}}$	$149.1 \pm 11.9_{\text{stat}}$ $\pm 44.7_{\text{syst}}$	$101.9 \pm 9.6_{\text{stat}}$ $\pm 30.6_{\text{syst}}$	$35.2 \pm 5.6_{\text{stat}}$ $\pm 10.6_{\text{syst}}$

H_T Bin (GeV)	575–675	675–775	775–875	875– ∞
MC $W + t\bar{t}$	$17.5 \pm 3.2_{\text{stat}}$	$5.1 \pm 1.8_{\text{stat}}$	$1.1 \pm 0.7_{\text{stat}}$	$1.8 \pm 1.0_{\text{stat}}$
MC $\mu + \text{jets}$	$13.3 \pm 2.9_{\text{stat}}$	$8.0 \pm 2.3_{\text{stat}}$	$3.2 \pm 1.4_{\text{stat}}$	$0.9 \pm 0.7_{\text{stat}}$
MC Ratio	0.90	0.90	0.90	0.90
Data $\mu + \text{jets}$	17	5	0	0
$W + t\bar{t}$ Prediction	$15.3 \pm 3.7_{\text{stat}}$ $\pm 4.6_{\text{syst}}$	$4.5 \pm 2.0_{\text{stat}}$ $\pm 1.4_{\text{syst}}$	$0.0 \pm 1.0_{\text{stat}}$	$0.0 \pm 1.0_{\text{stat}}$

Table 1.12: Muon sample predictions with 1.1fb^{-1} . Errors quoted on predictions correspond to statistical errors and an additional conservative systematic uncertainty of 30%, as used in the 2010 analysis.

1.9.3 μ Control Sample Systematic Uncertainty

Although the prediction is data-driven the reliance on the ratio R_{MC}^{had} which is taken from Monte Carlo introduces sources of uncertainty based not the accuracy of modelling, and thus we apply conservative

Dependence of the Prediction Calculation

The number of events measured in data $W_{data,meas}^\mu$ is related to the actual total number of $W \rightarrow \mu\nu$ events $W_{data,actual}^\mu$ by the relation in Equation 1.2 where the purity p represents the fraction of events in the control sample that originate from $W + \text{Jets}$ and $t\bar{t}$ processes, and f_X, a_X and ϵ_X are the fraction of events, acceptance and efficiency of each process $X = W, t\bar{t}$.

$$W_{data,actual}^\mu = W_{data,meas}^\mu \times \left(\frac{f_W}{a_W \times \epsilon_W} + \frac{f_{t\bar{t}}}{a_{t\bar{t}} \times \epsilon_{t\bar{t}}} \right) \times p \quad (1.2)$$

The purity is assumed to be 1 due to the lack of QCD contamination demonstrated in MC.

In addition, the prediction of W_{data}^{had} in Equation 1.1 can be rewritten in

Equation 1.3 in terms of $W_{data,actual}^\mu$ and the probabilities P_X^{had} of an $W \rightarrow l\nu$ event from process X passing the hadronic signal selection as the charged lepton was not identified by the lepton vetoes.

$$W_{data}^{had} = W_{data,actual}^\mu \times \left(\frac{W_{MC}^{had}}{W_{MC,actual}^\mu} \right) = W_{data,actual}^\mu \times (f_W P_W^{had} + f_{t\bar{t}} P_{t\bar{t}}^{had}) \quad (1.3)$$

Using these two equations together yields the full dependence of the ratio R_W^{had} in Equation 1.4, analogous with Equation —refeq:mupred where we assume $p = 1$ and the other factors of which are taken directly from the MC yields.

$$W_{data}^{had} = W_{data,meas}^\mu \times \left(\frac{f_W}{a_W \times \epsilon_W} + \frac{f_{t\bar{t}}}{a_{t\bar{t}} \times \epsilon_{t\bar{t}}} \right) \times (f_W P_W^{had} + f_{t\bar{t}} P_{t\bar{t}}^{had}) \times p \quad (1.4)$$

Multiplying out the full dependence introduces quadratic terms in the fractions pertaining to each component, f_W , $f_{t\bar{t}}$ which multiply the factors $P_W^{had}/(a_W \times \epsilon_W)$ and $P_{t\bar{t}}^{had}/(a_{t\bar{t}} \times \epsilon_{t\bar{t}})$. This can be calculated in MC by dividing the yield in the MC of each component of the full hadronic signal selection without the lepton veto with that once the lepton veto has been applied. Yields normalised to 1.1fb^{-1} are shown in Table 1.13 without and with the lepton veto, and the corresponding values of $P^{had}/(a \times \epsilon)$, indicating these are similar. Using the full H_T range (left) there is a small difference not seen in the 2010 analysis which is reduced by using the highest 6 bins only. The similarity between these two factors indicates that fluctuations in the fractions of W and $t\bar{t}$ events have little effect despite their quadratic nature, allowing a linear treatment of the uncertainties on the remaining factors.

Components of the Overall Uncertainty

Although the estimation technique is data-driven the use of the ratio R_{MC}^{had} places a reliance on Monte Carlo, and so we treat all factors that affect this ratio conservatively when assigning uncertainties. The values chosen are in accord with that developed for the previous iteration of this analysis, with the following contributory factors:

- The largest contribution to the overall uncertainty is from the uncertainty

1.9. Estimation of $t\bar{t}$ and $W + \text{Jets}$ Backgrounds with a high p_T control sample using $W \rightarrow \mu\nu$ events.

All Bins, $H_T > 275$ GeV			6 High Bins, $H_T > 375$ GeV		
X	W + Jets	$t\bar{t} + \text{jets}$	X	W + Jets	$t\bar{t} + \text{jets}$
Before e, μ vetoes	2161.86	1035.50	Before e, μ vetoes	568.34	283.92
After e, μ vetoes	594.24	227.53	After e, μ vetoes	126.62	58.91
$P_X^{\text{had}}/(a_X \times \epsilon_X)$	0.27	0.22	$P_X^{\text{had}}/(a_X \times \epsilon_X)$	0.24	0.23

Table 1.13: $t\bar{t}$ - W MC yields having passed all the hadronic signal final selection cuts without and with the lepton vetoes, normalised for 1.1fb^{-1} . The factor $P_X^{\text{had}}/(a_X \times \epsilon_X)$ for each process X is calculated from the division of the full selection yield by the yield before the vetoes are applied.

of the probability p for a W decay to pass the lepton veto, divided into individual uncertainties on each type of decay averaged whilst weighted by the frequency of that decay as described earlier in Section 1.8.3. Overall this contributes 23%, consisting of the following components assigned in conduction with 2010 W and Z cross-section measurements [7]:

- Events surviving due to hadronic tau decays are governed by the tau-jet response. In order to obtain the uncertainty the response was varied by $\pm 10\%$, a conservative amount given the JES it at its greatest 6%. The MC yield from such decays changed by 7% under this variation, and so we choose this for our uncertainty.
- Events out of acceptance rely on accurate modelling of a . The properties of W are found to be well modelled but the measurement is performed for higher p_T than we use in this analysis, therefore we apply a conservative 10%.
- The inefficiency of lepton ID requirements measured in electrons has a 30% under-estimation from MC in data. In muons the under-estimate is slight, but the uncertainty from Isolation requirements is not well understood, as unknown pile-up effects could alter this drastically. Therefore a conservative 100% estimate is applied to remain above measured discrepancies, given the small proportion of events this effects.
- The small number of fully leptonic $t\bar{t}$ decays is dominated by di- τ events. The expected uncertainty should take into account all the above effects, as all are relevant and responsible for some of these

events remaining. In this case we assign a 50% uncertainty .

- The effect of acceptance a times efficiency ϵ $1/a \times \epsilon$ is predicted using the MC, but there is an uncertainty resulting from the difference in this between MC and data, which when measured for W and Z cross sections contributes 1.6%, and for $t\bar{t}$ it is 6.1%. Conservatively we chose 6.1% which represents an overestimate as the proportion of W events in reality reduces this component.
- There is an uncertainty on the assumption that the control sample is pure, i.e. that the fraction of events f in the sample that come from W and $t\bar{t}$ decays is 1. From Monte Carlo we have observed the contamination from QCD to be negligible, but to allow for poor QCD MC modelling assign a conservative uncertainty of 200% on the expected QCD yield, still only contributing a 3% uncertainty on f .

The overall uncertainty is then achieved by adding these components in quadrature, which yields a value of 24%. Whilst lower than the value taken from the 2010 analysis, 30%, it is consistent and shows there is not an under-estimate. The error is reduced due to the inclusion of the two low H_T bins which reduces the overall percentage contributed from the lepton veto inefficiency, which propagates the highest error. Excluding the lower two bins brings this calculation in line with 30% and as the overall sensitivity is lowest in these two bins the outcome of introducing an over-estimate of the uncertainty will have little effect on the limit.

1.9.4 Full electroweak background prediction cross check

In addition to making an estimate of the $t\bar{t} + \text{jets}$ and $W + \text{jets}$ backgrounds, it is also possible to extend the use of the μ control sample to make a prediction of the full electroweak background by including $Z \rightarrow \nu\bar{\nu}$ into the hadronic MC yield W_{μ}^{had} in order to create the ratio $R_{\mu}^{had,allEWK}$. This allows a useful cross-check with the estimation of this component from the γ control sample.

1.10 Estimation of $Z \rightarrow \nu\bar{\nu} + \text{jets}$ background using photon + jets events

The irreducible background from the Z boson decay to a $\nu\bar{\nu}$ pair is estimated with the use of an energetic γ control sample, in a similar manner to that described for the μ control sample. The similarity between the kinematics of $Z \rightarrow \nu\bar{\nu} + \text{jets}$ events and $\gamma + \text{jets}$ can be exploited, in the scenario where the photon is disregarded from calculations of quantities H_T , \cancel{H}_T and α_T . The events therefore appear to have missing energy with a similar spectrum to that of $Z \rightarrow \nu\bar{\nu}$ events, it being produced at a larger cross-section [8]. This method has been designed and documented in [9].

The γ control sample is defined similarly to that of the μ control sample, retaining the hadronic signal region final selection with the removal of the photon veto and the following photon requirements:

- $p_T^\gamma > 100 \text{ GeV}$ putting the photon momentum above the mass of the Z (91.2 GeV) to enhance the similarity of kinematics.
- $|\eta^\gamma| < 1.45$
- $\Delta R(\gamma, \text{jet}) > 1$

The α_T and Jet Multiplicity distributions of the γ control sample selection events prior to the α_T cut are shown in Figure 1.9 for the bins in the region $H_T > 375 \text{ GeV}$ for data and MC from QCD and $\gamma + \text{jets}$ events, taken from MADGRAPH. The distributions show good shape agreement although the total yield in data is higher than in MC. As the method will only use MC to make a comparison ratio between the two selections any inaccuracies of the cross-section have no relevance to the estimation and therefore the method is still valid.

After the α_T cut the γ control sample is mainly free from QCD although there is a small number of events remaining, resulting in a purity factor of 0.92 and 0.97 in the two lowest bins and 0.99 in all other bins.

1.10.1 Z Background Prediction Calculation

In a similar way to the $t\bar{t}$ -W prediction, it is possible to extrapolate the prediction for the Z component of the hadronic signal yield, Z_{data}^{pred} from the measured yield

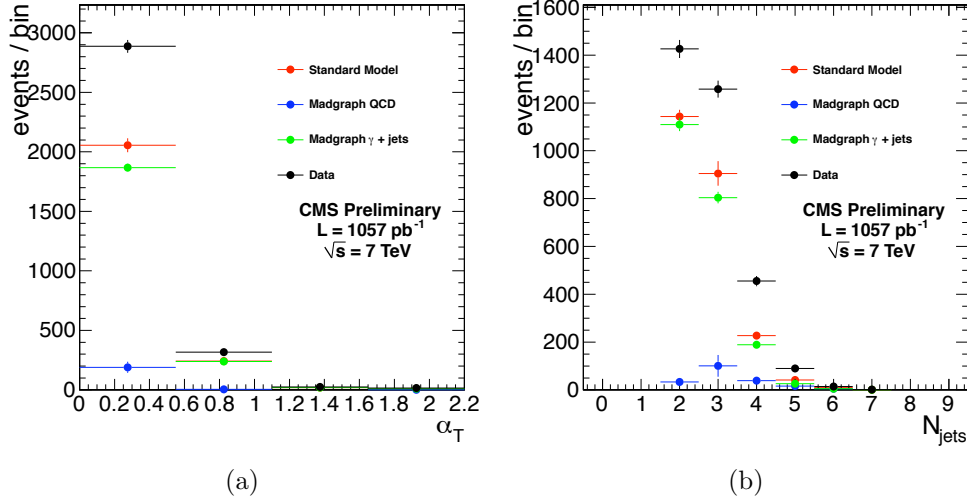


Figure 1.9: Data-MC comparisons for the photon control sample. $H_T > 375 \text{ GeV}$ and $\cancel{E}_T/H_T > 0.4$ are required. Left: the distribution of α_T . Right: the distribution of the number of jets.

of events in the $\gamma + \text{jets}$ control sample γ_{data}^{mess} , using as a translation factor the ratio $R_{Z/\gamma}^{MC}$. This is described in Equation 1.5, where the ratio $R_{Z/\gamma}^{MC}$ is created using the hadronic signal yield from $Z + \text{jets}$ MC events, Z_{MC} and the γ control selection yield in $\gamma + \text{jets}$ MC, γ_{MC} . The purity of the γ selection p is included as a multiplicative factor to correct for the QCD contamination in γ_{data}^{meas} , i.e. $\gamma_{data}^{corr} = p\gamma_{data}^{meas}$.

$$Z_{data}^{pred} = p \times \gamma_{data}^{meas} \times R_{Z/\gamma}^{MC} = \gamma_{data}^{corr} \times \left(\frac{Z_{MC}}{\gamma_{MC}} \right) \quad (1.5)$$

The calculation for the ratio $R_{Z/\gamma}^{MC}$ is performed once in each of the two lowest signal bins, whilst the six higher bins use one shared ratio as in the μ control method. The variation of the shapes with H_T are minimal and this approach minimises the error from MC statistics that will be passed on to the prediction in the highest bins.

The bin-by-bin results including prediction are shown in Table 1.14, with the normalised MC event yields that contribute to R_{μ}^{had} per bin and the yields in data for the γ selection at 1.1fb^{-1} . On the prediction both statistical errors and systematic errors are quoted, the latter corresponding to a 40% uncertainty used in the previous analysis, the components of which are listed below.

H_T Bin (GeV)	275–325	325–375	375–475	475–575
MC $Z \rightarrow \nu\bar{\nu}$	$212.6 \pm 20.4_{\text{stat}}$	$92.0 \pm 20.4_{\text{stat}}$	$61.3 \pm 20.4_{\text{stat}}$	$42.9 \pm 10.2_{\text{stat}}$
MC $\gamma + \text{jets}$	$613.1 \pm 20.4_{\text{stat}}$	$265.7 \pm 10.2_{\text{stat}}$	$168.6 \pm 10.2_{\text{stat}}$	$55.2 \pm 8.2_{\text{stat}}$
MC Ratio	0.35	0.35	0.44	0.44
Data $\gamma + \text{jets}$	867.5	313.7	214.6	68.5
Sample Purity	0.92	0.97	0.99	0.99
$Z \rightarrow \nu\bar{\nu}$ Prediction	$276.8 \pm 9.5_{\text{stat}}$ $\pm 110.7_{\text{syst}}$	$105.3 \pm 6.0_{\text{stat}}$ $\pm 42.1_{\text{syst}}$	$93.5 \pm 6.4_{\text{stat}}$ $\pm 37.4_{\text{syst}}$	$29.8 \pm 3.6_{\text{stat}}$ $\pm 11.9_{\text{syst}}$

H_T Bin (GeV)	575–675	675–775	775–875	875– ∞
MC $Z \rightarrow \nu\bar{\nu}$	$5.1 \pm 5.1_{\text{stat}}$	$0.0 \pm 3.1_{\text{stat}}$	$3.1 \pm 3.1_{\text{stat}}$	$0.0 \pm 3.1_{\text{stat}}$
MC $\gamma + \text{jets}$	$23.5 \pm 5.1_{\text{stat}}$	$3.1 \pm 2.0_{\text{stat}}$	$3.1 \pm 2.0_{\text{stat}}$	$2.0 \pm 1.0_{\text{stat}}$
MC Ratio	0.44	0.44	0.44	0.44
Data $\gamma + \text{jets}$	24.5	12.3	4.1	4.1
Sample Purity	0.99	0.99	0.99	0.99
$Z \rightarrow \nu\bar{\nu}$ Prediction	$10.7 \pm 2.2_{\text{stat}}$ $\pm 4.3_{\text{syst}}$	$5.3 \pm 1.5_{\text{stat}}$ $\pm 2.1_{\text{syst}}$	$1.8 \pm 0.9_{\text{stat}}$ $\pm 0.7_{\text{syst}}$	$1.8 \pm 0.9_{\text{stat}}$ $\pm 0.7_{\text{syst}}$

Table 1.14: Bin-by-bin prediction of $Z \rightarrow \nu\bar{\nu}$ irreducible background after final selection, using $\gamma + \text{jets}$ control sample with 1.1fb^{-1} 2011 data.

1.10.2 γ Control Sample Systematic Uncertainty

As in the μ control sample, the data-driven estimation techniques rely on a Monte Carlo ratio, and so we treat all factors that affect this ratio conservatively when assigning uncertainties. The values chosen are in accord with that developed for the previous iteration of this analysis, with the following contributory factors:

- As the Madgraph MC samples used are different in the numerator and denominator a factor of theoretical uncertainty exists on the relative cross-sections, taken conservatively as 30% [8] from MADGRAPH [10].
- The acceptance a is assigned an uncertainty of 5%, given the understanding of $\gamma + \text{jets}$ processes is good. The efficiency ϵ is assigned a conservative 20% as although the ID variables are validated with tag-and-probe [] the tests are not reformed in the high H_T
- The purity P is assigned an 20% uncertainty to take into account the uncertainty in the modelling of QCD MC which is used to estimate contamination.

Combining these events in quadrature and rounding to the nearest 10% yields an overall uncertainty of 40%

1.10.3 Cross-Prediction between Control Samples

A cross-prediction can be made between the two control samples, providing a cross-check in order to validate the methods and assigned systematics. The number of $W + \text{jets}$ events with μ decays $N_{data,pred}^W$ can be predicted from the γ control sample, exploiting the similarities between the kinematics of $W \rightarrow \mu\nu + \text{jets}$ and $Z \rightarrow \nu\bar{\nu} + \text{jets}$. In order to ensure the sample is not contaminated by events from $t\bar{t}$ production, an additional requirement that the jet multiplicity is constrained to 2 is made in both selections. The prediction proceeds using the same MC ratio strategy, as in Equation 1.6.

$$N_{pred}^W = N_{data}^\gamma \times \left(\frac{N_{MC}^W}{N_{MC}^\gamma} \right) \quad (1.6)$$

The ratio N_{MC}^W/N_{MC}^γ is found to be independent of H_T and therefore one factor 0.42 ± 0.04 is extracted for the whole set of bins. The results of the prediction are shown in Table 1.15 alongside the number of $W + \text{jets}$ events $N_{data,meas}^W$ found in data from the selection with MC statistical errors and systematic uncertainties in line with those described for the two control samples.

H_T	N_{data}^γ	$N_{MC}^W/N_{MC}^{\text{phot}}$	N_{pred}^W	N_{obs}^W
275	336	$0.42 \pm 0.04_{\text{MCstat}}$	$141.8 \pm 7.7_{\text{stat}} \pm 14.6_{\text{MCstat}} \pm 56.7_{\text{syst}}$	128
325	127	$0.42 \pm 0.04_{\text{MCstat}}$	$53.6 \pm 4.8_{\text{stat}} \pm 5.5_{\text{MCstat}} \pm 21.4_{\text{syst}}$	37
375	96	$0.42 \pm 0.04_{\text{MCstat}}$	$40.5 \pm 4.1_{\text{stat}} \pm 4.2_{\text{MCstat}} \pm 16.2_{\text{syst}}$	36
475	27	$0.42 \pm 0.04_{\text{MCstat}}$	$11.4 \pm 2.2_{\text{stat}} \pm 1.2_{\text{MCstat}} \pm 4.6_{\text{syst}}$	12
575	13	$0.42 \pm 0.04_{\text{MCstat}}$	$5.5 \pm 1.5_{\text{stat}} \pm 0.6_{\text{MCstat}} \pm 2.2_{\text{syst}}$	2

Table 1.15: Predictions of $W \rightarrow \mu\nu + 2 \text{ jets}$ events using the $\gamma + \text{jets}$ sample at 1.1fb^{-1} , including statistical errors and a systematic uncertainty of 40% as used in the $Z + \text{Jets}$ prediction.

The number of events predicted are compatible with those measured in data within the uncertainties estimated by the techniques.

1.11 Signal Region Systematic Uncertainties

A number of experimental systematic uncertainties have an effect on the efficiencies of potential signal, which are detailed here.

Luminosity

The measurement of luminosity taken propagates through to an uncertainty on the signal event yield when considering any new physics model, which is currently 6% [2]

Effect of dead ECAL cut

The cut that removes events where a jet points towards a region with masked ECAL towers has varying efficiencies for different signal models. This introduces an uncertainty based on the distribution about the mean. A study using several points in the CMSSM yields a standard deviation $\sim 2\%$. In addition there is a contribution to the uncertainty due to the resolution of ΔR , 0.05 for jets with $p_T > 100$ GeV. Such a variation corresponds to 2.2%, and the overall uncertainty for this cut is therefore 3%

Effect of $e/\mu/\gamma$ Vetoes

The rejection of leptons and photons have efficiencies that agree in data very well with QCD Monte Carlo using three different generators (PYTHIA6, PYTHIA8, MADGRAPH), the variation of which is at maximum 0.8% for the total effect which rejects $\sim 5\%$ of events. We choose to assign half this value, 2.5% representing 50% of the total veto inefficiency.

Jet Energy Scale and Resolution

Fluctuations of the JES affect the jets that pass the p_T and H_T requirement, which has an effect on the overall result. Varying this in accordance with results described in Section ?? show variations on yields of CMSSM points within +1.9% and -2.2%, of which a systematic of 2% is chosen as the latter would not artificially improve the yield. The resolution from MC is found to be 10-15% better than in data, and our correction for that yields a 1% uncertainty on signal yield. Overall 2.5% is contributed from JES and resolution.

Theoretical Uncertainty

In addition to the experimental uncertainties listed above, there is a theoretical contribution stemming from the choice of the renormalisation and factorisation scales used to calculate NLO cross sections and the PDF's used in signal Monte Carlo contribute an 10% effect.

Adding all these sources of uncertainty in quadrature gives an overall uncertainty of 12.5%, which is carried onwards to the limit calculation.

1.12 Statistical Interpretation

Having obtained the yields and predictions as detailed in previous sections, it is desirable to quantify and interpret the results of the three selections simultaneously with respect to the SM only hypothesis and also possible CMSSM SUSY signal. The results in all 3 selections are used simultaneously together in order to draw conclusions. A likelihood model is used for each of the three samples describing the relevant results along with the uncertainties, with the number of observed events in each assumed to have a Poisson Distribution $\text{Pois}(n|\mu)$ where the number of expected events is μ as in Equation 1.7.

$$\text{Pois}(n|\mu) = \frac{\mu^n}{n!} e^{-\mu} \quad (1.7)$$

The likelihood of μ being the number of expected events given the outcome n can then be expressed in terms of this probability distribution. Where μ depends on a set of unknown parameters maximising L provides estimates for the models parameters. The $N=8$ measurements corresponding to each H_T bin enter the likelihood distinctly and simultaneously through a product of Poisson distributions. It does not distinguish between bins of differing width.

1.12.1 Hadronic Signal Selection Likelihood

Given a set of observed event yields for the hadronic signal selection n^i in $i=1, \dots, N$ H_T bins, the likelihood L_{had} is described by:

$$L_{had} = \prod_i \text{Pois}(n^i | b^i + s^i) \equiv \prod_i \text{Pois}(n^i | b_{ewk}^i + b_{qcd}^i + s^i) \quad (1.8)$$

where the expected yields are composed of s^i the expected number of signal events in the i th bin, and b^i the expected Standard Model background, assuming $b^i \equiv b_{ewk}^i + b_{qcd}^i$ where b_{ewk}^i is the expected yield from electroweak processes, and b_{qcd}^i the expected yield from QCD.

1.12.2 Expression of b^i using R_{α_T} evolution in H_T

The separation of b^i into electroweak and QCD components allows the expression of each in terms of its characteristic behaviour of R_{α_T} as H_T evolves.

The hypothesis of R_{α_T} exponentially falling with increasing H_T can be expressed as a function of two parameters A and k in the following way:

$$R_{\alpha_T}(H_T) = Ae^{-kH_T}, \quad (1.9)$$

This can be used to express a hypothesis of flat behaviour also, by setting $k = 0$. A further observation is introduced, m^i which represents the observation of the hadronic bulk selection (where $\alpha_T < 0.55$). It is then possible to express the expected background b_p^i from a process p in terms of the H_T distribution of these bulk events, dN/dH_T :

$$b_p^i = \int_{x_i}^{x_{i+1}} \frac{dN}{dH_T} Ae^{-kH_T} dH_T, \quad (1.10)$$

where x_i is the lower bin edge in H_T , and x_{i+1} represents the upper edge (∞ in the case of the final bin). To simplify this continuous distribution the assumption is made that the full distribution of a given bin occurs at the mid-point of the bin, the mean $\langle H_T \rangle^i$, allowing the expression in terms of m^i :

$$\frac{dN}{dH_T} = \sum_i m^i \delta(x - \langle H_T \rangle^i), \quad (1.11)$$

yielding the full dependence of b_p^i in Equation 1.12.

$$b_p^i = \int_{x_i}^{x_{i+1}} m^i \delta(x - \langle H_T \rangle^i) Ae^{-kH_T} dH_T \quad (1.12)$$

This allows the expression in Equation ?? of the two components of b^i , in which the knowledge that R_{α_T} is flat in electroweak processes, and therefore k_{ewk}

$= 0$ is used.

$$b_{ewk}^i = m^i A_{ewk} \quad b_{qcd}^i = m^i A_{qcd} e^{-k_{qcd} \langle H_T \rangle^i}. \quad (1.13)$$

The likelihood of the full set of the background expectations b^i now depends on three nuisance parameters, A_{ewk} , A_{qcd} and k_{qcd} .

1.12.3 Electroweak Control Sample Likelihoods

The electroweak component of the background b_{ewk}^i can additionally be decomposed into terms of the expected number of Z and $t\bar{t}W$ components b_Z^i and $b_{t\bar{t}W}^i$. Reversing $b_{ewk}^i = b_Z^i + b_{t\bar{t}W}^i$ by introducing a set of fit parameters f_Z^i , the fraction of electroweak events that are Z in the i th bin, each component can be expressed as follows:

$$b_Z^i = f_Z^i \times b_{ewk}^i \quad b_{t\bar{t}W}^i = (1 - f_Z^i) \times b_{ewk}^i \quad (1.14)$$

Each of these components has an observational measurement n_γ and n_μ taken from the event yield in the photon and muon control samples respectively. Corresponding yields in simulation MC_γ^i and MC_μ^i are also known, along with the value in simulation of the expected amounts of Z and $t\bar{t}W$ in the hadronic signal region MC_Z^i and $MC_{t\bar{t}W}^i$, which combine to define the ratios r_γ^i and r_μ^i as follows:

$$r_\gamma^i = \frac{MC_\gamma^i}{MC_Z^i} \quad r_\mu^i = \frac{MC_\mu^i}{MC_{t\bar{t}W}^i} \quad (1.15)$$

The likelihoods regarding the two measured yields n_γ^i, n_μ^i can then be fully expressed as in Equation 1.16 and 1.17 with an additional parameter s_μ^i represents the expected signal events contaminating the muon control sample for a given signal model.

$$L_\gamma = \text{Gaus}(1.0 | \rho_\gamma^Z, \sigma_\gamma^Z) \prod_i \text{Pois}(n_\gamma^i | \rho_\gamma^Z r_\gamma^i b_Z^i) \quad (1.16)$$

$$L_\mu = \text{Gaus}(1.0 | \rho_\mu^{t\bar{t}W}, \sigma_\mu^{t\bar{t}W}) \prod_i \text{Pois}(n_\mu^i | \rho_\mu^{t\bar{t}W} r_\mu^i b_{t\bar{t}W}^i + s_\mu^i) \quad (1.17)$$

In addition to the Poisson product an additional Gaussian has been incorporated to account for the systematic error, in which $\rho_\gamma^Z, \rho_\mu^{t\bar{t}W}$ are the correction

factors that account for the systematic uncertainties $\sigma_\gamma^Z, \sigma_\mu^{t\bar{t}W}$ associated with the respective prediction constraints. We assume the uncertainties are fully correlated across the range of bins.

These two electroweak likelihoods add an additional dependance on the N parameters f_Z^i as well as two uncertainty correction factors $\rho_\gamma^Z, \rho_\mu^{t\bar{t}W}$.

1.12.4 Presence of Signal

Where the hypothesis includes presence of signal events the likelihood requires an additional component, representing the effect of systematic uncertainties associated with signal efficiency, σ^{sig} . This is treated with a Gaussian as in the electroweak cases, and introduces the correction factor ρ^{sig} :

$$L_{signal} = Gaus(1.0|\rho^{sig}, \sigma^{sig}) \quad (1.18)$$

With the inclusion of this uncertainty correction in the full likelihood the signal contribution s^i in Equation 1.8 can be rewritten in terms of the cross-section x of the model and the measured luminosity l , using the efficiency of the analysis in bin i , ϵ_{had}^i .

$$s^i = f\rho_{sig}xl\epsilon_{had}^i \quad (1.19)$$

This introduces the multiplicative factor f^x to be applied to the cross section. This represents the parameter of interest for which we shall determine an allowed interval. Analogously the signal contamination in the muon control region s_μ^i in Equation ?? is also written using these factors and the muon selection efficiency for signal ϵ_μ^i :

$$s_\mu^i = f\rho_{sig}xl\epsilon_\mu^i \quad (1.20)$$

1.12.5 Total Likelihood

The total likelihood is then expressed in Equation 1.21, as a product of the individual likelihood functions described previously.

$$L_{tot} = L_{had} \times L_\gamma \times L_\mu \times L_{signal} \quad (1.21)$$

The parameters of the full likelihood are then the R_{α_T} components: A_{ewk} , A_{qcd} and k_{qcd} , the N factors $\{f_Z^i\}$, the three uncertainty corrections ρ^{sig} , ρ_γ^Z and $\rho_\mu^{t\bar{t}W}$ and the strength of the signal f_x . These number 7 + N in total with f_x the primary parameter of interest whilst the other 6 + N are regarded as nuisance parameters.

1.13 Testing the SM-only hypothesis

Dropping contributions from possible signal to the likelihood (setting $s^i = s_\mu^i = 0$) allows consideration of the SM background-only hypothesis. Given the observations in data the likelihood is maximised over all of the parameters with the tools Minuit [?] and RooFit, and the maximum is recorded (L_{max}^{data}). The values of the parameters at this maximum value are known as maximum likelihood estimates (MLEs).

The likelihood function with the MLEs plugged in is then used as a p.d.f for observations, and many pseudo-experiments are generated from it. Starting with the observations in each pseudo-experiment, the likelihood is then re-maximised over all parameters in order to generate a distribution of the valleys of L at maximum, L_{max} . The p-value of the actual observation is found to be 0.56, obtained by examining the quantile of L_{max}^{data} in this distribution.

The results of the fit are shown in Figure 1.10 separately for the three samples, each with the actual yields from data (black) compared to the fit results (blue). The hadronic signal results (top) are shown alongside the predictions from the μ and γ control samples. For the hadronic and muon (middle) samples where signal (when present) would contribute, the distribution of SM with LM6 signal present is shown also for illustration purposes, although this does not enter the fit or affect the results.

All three samples show good agreement between the data and fit results, indicating the measured yields are compatible with the SM-only hypothesis. In Figure 1.11 the R_{α_T} distribution is shown by dividing the hadronic sample data and fit results by the measured bulk yields m^i . The electroweak component is flat as was required for the fit. The QCD component exhibits falling behaviour with MLEs $k = (5.2 \pm 5.6) \times 10^{-3}$ and $A_{qcd} = (1.4 \pm 1.9) \times 10^{-5}$, consistent with the concept of negligible QCD contamination. Although the presence of QCD

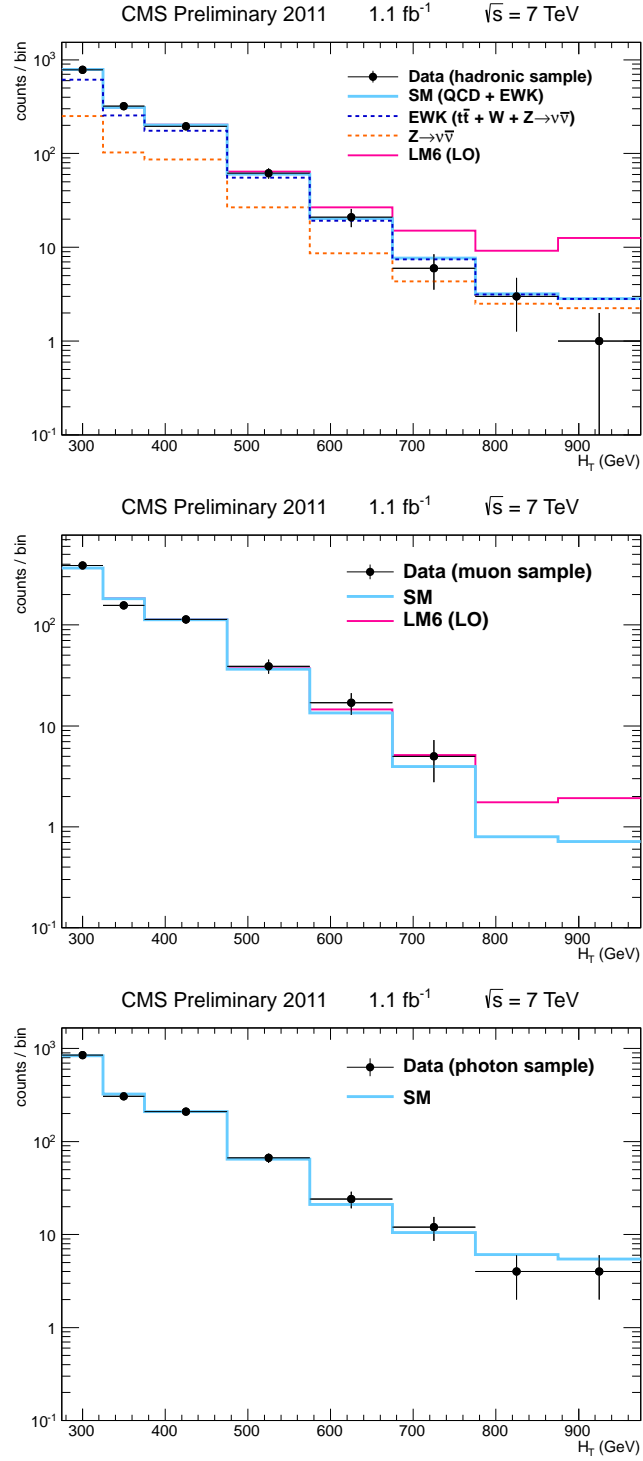


Figure 1.10: SM-only goodness of fit results for events in the hadronic (top), muon (middle) and photon (bottom) samples. Each H_T bin shows data observation (black points) and the outcome of the fit (blue line). For hadronic selection the breakdown of the individual background contributions as predicted by the control samples is shown also. Signal contribution from benchmark point LM6 is shown stacked with SM for illustration (pink line) but has no role in the fit. 41

was specified in the fit, a cross-check setting the QCD component to zero returns a p-value from the fit of 0.41 confirming the approach.

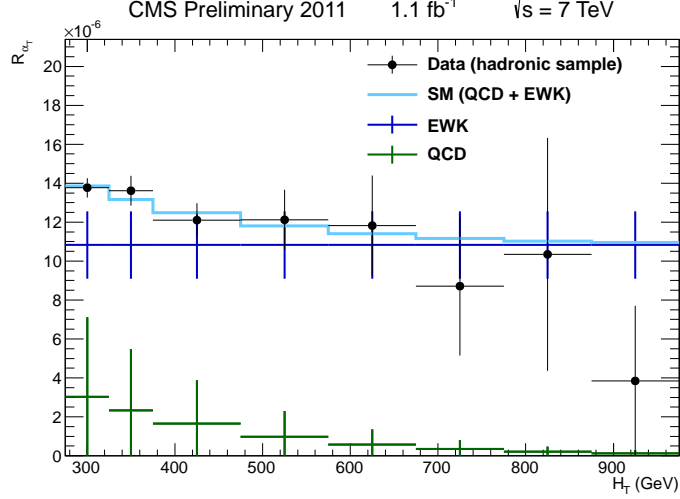


Figure 1.11: R_{α_T} as a function of H_T as observed in data (black points) and the results of the fit (blue) both divided through by the bulk observations. The components of the background are shown also, where the electroweak was fixed as flat during the fit whilst qcd was allowed to fall exponentially.

The good agreement between data and the Standard Model indicates a lack of signal, in which case the inclusion of signal in the likelihood is used to interpret the results in the plane of the CMSSM. The method of this is described in the following section.

1.14 Excluding Signal in the CMSSM Plane

The profile likelihood (PL) ratio of a given likelihood

$$\lambda(\mu) = \frac{L(\mu, \theta_\mu)}{L(\hat{\mu}, \hat{\theta})} \quad (1.22)$$

1.15 Conclusion

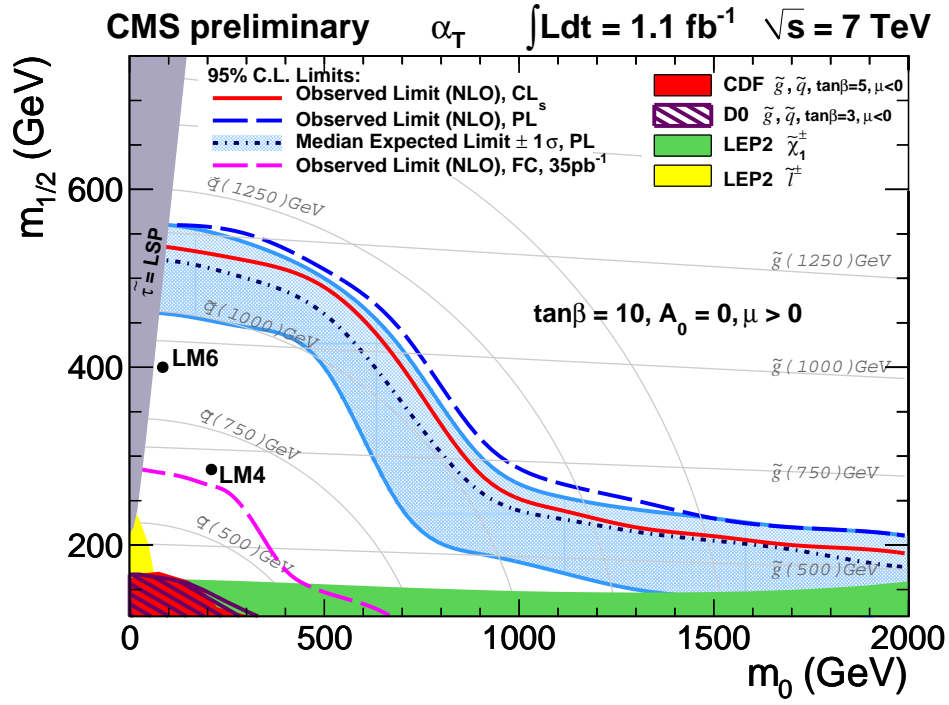


Figure 1.12: Observed and expected 95% CL exclusion contours in the CMSSM $(m_0, m_{1/2})$ plane ($\tan\beta = 10, A_0 = 0, \mu > 0$) using NLO signal cross sections using the Profile Likelihood (PL) method. The expected limit is shown with its 68% CL range. The observed limit using the CL_s method is shown as well.

H_T Bin (GeV)	275–325	325–375	375–475	475–575
W + $t\bar{t}$ background	363.7	152.2	88.9	28.8
Z $\rightarrow \nu\bar{\nu}$ background	251.4	103.1	86.4	26.6
QCD background	172.4	55.1	26.9	5.0
Total Background	787.4^{+32}_{-22}	310.4^{+8}_{-12}	202.1^{+9}_{-9}	60.4
Data	782	321	196	62

H_T Bin (GeV)	575–675	675–775	775–875	875– ∞
W + $t\bar{t}$ background	10.6	3.1	0.6	0.6
Z $\rightarrow \nu\bar{\nu}$ background	8.7	4.3	2.5	2.2
QCD background	1.0	0.2	0.1	0.0
Total Background	20.3	7.7	3.2	2.9
Data	21	6	3	1

Table 1.16: Fit results for 1.1fb^{-1} with data observations. Since the QCD fit parameters are compatible with zero (see text), the listed QCD contributions in this table are also compatible with zero.

Bibliography

- [1] Vardan Khachatryan et al. “Search for Supersymmetry in pp Collisions at 7 TeV in Events with Jets and Missing Transverse Energy”. *Phys.Lett.*, **B698** (2011) 196–218, 1101.1628.
- [2] The CMS Collaboration. “Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update”. **CMS PAS EWK-11-001**.
- [3] The CMS Collaboration. “Calorimeter Jet Quality Criteria for the First CMS Collision Data”. **CMS-PAS-JME-09-008**.
- [4] The CMS Collaboration. “Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV”. **CMS-PAS-MUO-10-002**.
- [5] N. Adam et al. “Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”. **CMS AN 2010/116**.
- [6] The CMS Collaboration. “Isolated Photon Reconstruction and Identification at $\sqrt{s} = 7$ TeV”. **CMS-PAS-EGM-10-006**.
- [7] The CMS Collaboration. “Measurement of the W and Z inclusive production cross sections at $\sqrt{s} = 7$ TeV with the CMS experiment at the LHC”. **CMS PAS EWK-10-002**.
- [8] Z. Bern, G. Diana, L.J. Dixon, F. Febres Cordero, S. Hoche et al. “Driving Missing Data at Next-to-Leading Order”. *Phys.Rev.*, **D84** (2011) 114002, 1106.1423.
- [9] The CMS Collaboration. “Performance of Methods for Data-Driven Background Estimation in SUSY Searches”. **CMS PAS SUS-10-001**.
- [10] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet et al. “MadGraph/MadEvent v4: The New Web Generation”. *JHEP*, **0709** (2007) 028, 0706.2334.