QCD Processes and Search for Supersymmetry at the LHC

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

Torben Schum

aus Hamburg

Hamburg
2012

Gutachterin/Gutachter der Dissertation:	Prof. Dr. Peter Schleper Prof. Dr. Johannes Haller
Gutachterin/Gutachter der Disputation:	Dr. Isabell Melzer-Pellmann Dr. Christian Sander
Datum der Disputation:	12.07.2012
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Peter Hauschild
Leiter des Departments Physik:	Prof. Dr. Daniela Pfannkuche
Dekan der MIN-Fakultät:	Prof. Dr. Heinrich Graener

Abstract

In this thesis, a data-driven method to estimate the number of QCD background events in a multijet search for supersymmetry at the LHC was developed. The method makes use of two models which predict the correlation of two key search variables, the missing transverse momentum and an angular variable, in order to extrapolate from a QCD dominated control region to the signal region. A good performance of the method was demonstrated by its application to $36\,\mathrm{pb}^{-1}$ data, taken by the CMS experiment in 2010, and by the comparison with an alternative method. Comparing the number of data events to a combined background expectation of QCD and data-driven estimates of the electroweak and top background, no statistically significant excess was observed for three pre-defined search regions. Limits were calculated for the $(m_0, m_{1/2})$ parameter space of the cMSSM, exceeding previous measurements. The expected sensitivity for further refined search regions was investigated.

Kurzfassung

Im Rahmen dieser Doktorarbeit wurde eine datengetriebene Methode zur Bestimmung des Anteils von QCD-Untergrund an Multijet-Ereignissen in einer Supersymmetrie-Suche am LHC entwickelt. Als zentraler Teil dieser Methode wird die Korrelation zweier Schlüsselvariablen (das fehlende transversale Momentum und eine Winkelvariable) mit Hilfe von zwei Modellen vorhergesagt, in denen jeweils aus einer QCD-dominierten Kontrollregion in die Signalregion extrapoliert wird. Die Methode wurde erfolgreich auf die $36\,\mathrm{pb}^{-1}$ Daten, die 2010 mit dem CMS-Experiment gesammelt wurden, angewendet. Das Ergebnis zeigte sich in Übereinstimmung mit einer alternativen Methode zur QCD-Bestimmung. Bei einer kombinierten Vorhersage aller beitragenden Untergrundprozesse, wozu neben der QCD-Bestimmung auch datengetriebene Methoden für die elektroschwachen und Topquark-Prozesse verwendet wurden, konnte für drei zuvor festgelegte Suchregionen keine statistisch signifikante Abweichung in den Daten gefunden werden. Die mit diesen Messungen verträgliche Region der $(m_0, m_{1/2})$-Parameterebene des cMSSM konnte, im Vergleich zu früheren Messungen, weiter eingeschränkt werden. Zudem wurde die zu erwartende Sensitivität weiterer Suchregionen untersucht.

"There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened."

— Douglas Adams, *The Restaurant at the End of the Universe*

Contents

1 Introduction **1**

2 Supersymmetric Extensions to the Standard Model **5**
2.1 The Standard Model of Particle Physics 5
2.2 Supersymmetry 10
2.2.1 Motivations for Supersymmetry 10
2.2.2 MSSM and Supergravity 11
2.2.3 Expected Signatures at the LHC 15

3 Experimental setup **19**
3.1 The Large Hadron Collider 19
3.2 The CMS Experiment 20
3.2.1 The Inner Tracker 22
3.2.2 The Calorimeters 23
3.2.3 The Muon System 24
3.2.4 The Trigger System 25
3.3 The Particle-Flow Algorithm 26

4 Search Design **27**
4.1 Search Strategy 27
4.2 Monte Carlo Samples 28
4.3 Event Selection, Trigger and Cleaning 28
4.4 Data-Simulation Comparison 33

5 Estimation of Electroweak and Top Background from Data **37**
5.1 W and Top Quark Background 37
5.1.1 The Lost Lepton Background Estimation 37
5.1.2 Hadronic τ Background Estimation 40
5.2 Invisible Z Background Estimation 42
5.2.1 Estimation of $Z \to \nu\bar{\nu}$ Background from γ + jets 42

6 Data-driven QCD Background Estimation **45**
6.1 The Factorization Method 45
6.1.1 A simple Idea how to estimate QCD from Data 45
6.1.2 Topology of QCD Events in hadronic SUSY Searches . . 47
6.1.3 The full Method and its technical Application 54
6.1.4 Closure and Robustness Check 56
6.1.5 Contamination from SUSY and SM Processes 68
6.2 Application of the Factorization Method in Data 69
6.2.1 Control Regions 69
6.2.2 Verification of the Model in Data 70
6.2.3 Systematic Uncertainties 70
6.2.4 Summary of Results 78
6.3 The Rebalancing and Smear Method 80
6.3.1 Basic Concept of the Method 80
6.3.2 Measuring the Jet Response 83
6.3.3 Results in Monte Carlo and Data 85
6.3.4 Comparision of the two Methods 88

7 Search Results **91**
7.1 Combination of Background Estimations 91
7.2 Limits on SUSY Signals 92
7.2.1 Signal Simulation and Uncertainty 94
7.2.2 The Hybrid CLs Method 94
7.2.3 Interpretation within cMSSM 99
7.3 Studying the Search Sensitivity 104
7.3.1 Variation of the inclusive Search Regions 104

8 Summary **113**

Bibliography **117**

1 Introduction

The Standard Model of particle physics (SM) is a very successful description of all known elementary particles and three of the four fundamental forces in nature. So far, it is in agreement with all experimental collider results and many discoveries of the last decades have been based on its predictions. While high-precision measurements of the parameters of the SM and the search for the last missing particle, the Higgs boson, are still on-going, many experimental tests aim at theories beyond the SM. These searches for new physics are driven by a wide range of theoretical and experimental unsolved problems, such as the composition of Dark Matter, which reveal the SM as an incomplete theory.

Since the development of the concept of supersymmetry in the early 1970s, many theories beyond the SM include this proposed space-time symmetry between fermions and bosons. Supersymmetric theories can not only provide candidates for Dark Matter but make it also possible to solve intrinsic shortcomings of the SM such as the hierarchy problem.

As no supersymmetric particle could be observed yet, supersymmetry can only be realized as a broken symmetry in nature. This leads to a wide range of supersymmetric models with different breaking scenarios which can to a large extent be tested in collider experiments.

The Large Hadron Collider (LHC) which started operating in 2009 provides unique possibilities for the search for new physics. The high center of mass energy of $\sqrt{s} = 7$ TeV in pp collisions, for the first time achieved in 2010, together with a high luminosity would allow discoveries in large fractions of parameter spaces of supersymmetric models. The prospects will be even improved once the design center of mass energy ($\sqrt{s} = 14$ TeV) and the design luminosity of the LHC will be reached.

In this thesis, a search for new physics using the signatures of large missing transverse momentum in multijet events in pp collisions at $\sqrt{s} = 7$ TeV is presented (published in [1]). The analysis makes use of $36\,\mathrm{pb}^{-1}$ of data collected with the CMS detector at the LHC from March until November of 2010. The results of the analysis are interpreted in the context of the constrained Minimal Supersymmetric extension of the Standard Model (cMSSM), and since no discovery could be claimed, limits on the main parameters of the

cMSSM are presented.

The two fundamental concepts of the analysis are, first, to keep the event selection generic in a way that allows to reach sensitivity in large parts of the signal model parameter spaces, and secondly to make use of data-driven methods for the estimation of SM backgrounds. By the second choice the analysis is independent of possible imperfections of the simulation of SM processes.

The first goal is achieved by introducing a baseline event selection with moderate cuts on the two key variables $\not{H}_T$ (missing transverse momentum constructed from jets) and H_T (sum of transverse momenta of jets as a measure of the total hadronic activity in the event) together with two evolved selections, one with an increased cut on $\not{H}_T$ ($> 250\,\text{GeV}$) and one with an increased cut on H_T ($> 500\,\text{GeV}$).

For this analysis, the most challenging to understand SM background contribution is multi-jet production due to QCD processes. To estimate the number of QCD events in the signal region, the so-called factorization method has been developed that makes use of the correlation between the missing transverse momentum $\not{H}_T$ and an angular variable between the $\not{H}_T$ vector and the leading jets. This correlation is used to predict the background in the tails of the $\not{H}_T$ distribution. The results of the factorization methods are compared to the results of an independent method.

The QCD prediction with these methods in combination with data-driven estimates for the $Z \to \nu\bar{\nu} + \text{jets}$, $t\bar{t}$ and $W + \text{jets}$ events yield a complete prescription of the H_T and $\not{H}_T$ distributions observed in data.

The analysis serves as a basis for further multijet searches for new physics with the CMS detector. These searches benefit from the increasing luminosity of the LHC and will be provided with a higher center of mass energy in future. This thesis presents a technique that allows to optimize the search cuts on the two key variables $\not{H}_T$ and H_T in terms of best expected sensitivity for future searches, based on the background estimations of this analysis.

The thesis is organized as follows. In chapter 2 short-comings of the SM are discussed and a brief introduction to supersymmetric models is given. The CMS experiment at the LHC is described in ch. 3. In ch. 4 the search criteria are defined and the data passing the event selection is compared to Monte Carlo simulation.

In the next two chapters the data-driven background estimations are discussed. While in ch. 5 the electroweak backgrounds are reviewed, ch. 6 is dedicated to the QCD background. The concepts and the application of the factorization method to data are described in detail. The R&S method is introduced and both methods are compared.

In ch. 7 the statistical interpretation of the observed data is presented and a search optimization technique for future searches is introduced. Finally, the thesis is summarized in ch. 8.

2 Supersymmetric Extensions to the Standard Model

The standard model of particle physics (SM) is the starting point of all searches for new physics. It has been proven to be extremely successful in the description of experimental collider results, but there remain unsolved problems in particle physics and cosmology.

In this chapter, the motivation for searches beyond the SM are reviewed, starting with the SM and a brief discussion of its general short-comings in sec. 2.1. Section 2.2 focusses on the promising concept of supersymmetry (SUSY), a theoretical elaboration demonstrating the possibility to solve numerous problems of the SM. The latter section also highlights the importance of the Large Hadron Collider (LHC) for experimental tests of SUSY models.

2.1 The Standard Model of Particle Physics

The SM is formulated as a relativistic quantum field theory that describes the elementary particles as well as their fundamental interactions. There are three generations of leptons and quarks which have in common that they are fermions (particles with half-integer spin). The first generation provides the components of ordinary matter, e.g. atoms. The quarks and leptons from the second and third generations are heavier copies of the first generation. The masses of the 12 elementary fermions enter the SM as free parameters and have to be determined by experiments. While leptons only interact with the fields of the electroweak forces, quarks also experience the *strong* force, or in other words they carry a color charge.

The force carrier particles, which belong to the gauge fields of the fundamental interactions, and the scalar Higgs boson, which is one quantum component of the Higgs field, are bosons (integer spin particles). The gauge bosons of the *electromagnetic* and the *strong* interaction (photons and gluons, respectively) are massless, whereas there exist gauge bosons of the *electroweak* interaction ($W^{\pm}$ and Z^0) which are massive, restricting these interactions to the very short range of $\sim 10^{-3}$ fm. The limited range of the *strong* interaction

of $\sim$1 fm is explained by the concept of color confinement which is observed in nature. Essential for the color confinement is that the gluons themselves are color charged and therefore strongly interact which each other. As a result, color charged particles cannot be separated and therefore are only observable in composites called hadrons. The strengths of the gauge couplings introduce another 3 free parameters to the SM.

The properties of the elementary particles of the SM are summarized in table 2.1. For each particle the associated antiparticle has the same mass and spin but opposite electric charge. The symmetry between particles and antiparticles (mediated through the CP operator) is violated by weak interactions, which can be expressed by a complex phase in the mixing matrices. Together with the three mixing angles which describe the violation of the quark flavor quantum number conservation by the charged weak interaction, another 4 free parameters are needed in the SM.

Originally, neutrinos were thought to be massless, leading to the assumption that mixing between the lepton generations does not exist even though the opposite had been proven for the quarks. The experimental observation of neutrino oscillations made the introduction of a lepton mixing matrix with at least 4 more free parameters necessary.[1]

The CP invariance of the strong interaction has been verified by measuring the electric dipole momentum of neutrons which leads to a very small upper limit. This symmetry is neither postulated nor predicted by the SM and the missing knowledge can be interpreted as one more free parameter. The CP symmetry violation in the SM is too small to explain the asymmetry between matter and antimatter in the universe.

The fundamental interactions in the SM can be described by local gauge field theories. The electroweak model is based on the gauge group $\boldsymbol{SU(2)_L \otimes U(1)_Y}$, whereas quantum chromodynamics (QCD) is based on the $\boldsymbol{SU(3)}$ symmetry. The electroweak symmetry is spontaneously[2] broken via the Higgs mechanism. By constructing a Higgs field that consists of two neutral and two charged component fields the masses of the heavy gauge bosons are incorporated in the SM. Furthermore, two free parameters are introduced one of which is the last not yet determined by experiments: The Higgs boson mass.

The 26 free parameters of the SM are summarized in tab 2.2.The fermion masses can also be expressed as Yukawa couplings to the Higgs field. Never-

[1] In the simple Dirac case, the lepton mixing is completely analogous to the quark mixing. However, since neutrinos could also be Majorana particles additional CP phases are possible.[5]

[2] Spontaneous symmetry breaking means the symmetry is broken by the non-zero vacuum expectation value and not by the Lagrangian. For details see e.g. [6].

Generation/ Interaction	Lepton flavor / Quark flavor / Gauge boson	Spin	Electric charge	Color charge	Mass
I	electron e^-	$\frac{1}{2}$	$-e$	-	0.51 MeV
	el. neutrino ν_e	$\frac{1}{2}$	0	-	< 2 eV
II	muon μ^-	$\frac{1}{2}$	$-e$	-	105.7 MeV
	muon neutrino ν_μ	$\frac{1}{2}$	0	-	< 2 eV
III	tau τ^-	$\frac{1}{2}$	$-e$	-	1.78 GeV
	tau neutrino ν_τ	$\frac{1}{2}$	0	-	< 2 eV
I	up u	$\frac{1}{2}$	$+\frac{2}{3}e$	r, g, b	1.7-3.1 MeV
	down d	$\frac{1}{2}$	$-\frac{1}{3}e$	r, g, b	4.1-5.7 MeV
II	charm c	$\frac{1}{2}$	$+\frac{2}{3}e$	r, g, b	1.3 GeV
	strange s	$\frac{1}{2}$	$-\frac{1}{3}e$	r, g, b	80-130 MeV
III	top t	$\frac{1}{2}$	$+\frac{2}{3}e$	r, g, b	172.9 GeV
	bottom b	$\frac{1}{2}$	$-\frac{1}{3}e$	r, g, b	4.2 GeV
Electroweak	photon γ	1	0	-	0
	W^-	1	$-e$	-	80.4 GeV
	Z^0	1	0	-	91.2 GeV
Strong	gluon g	1	0	$r\bar{r}, r\bar{g}, r\bar{b}$ $g\bar{g}, g\bar{r}, g\bar{b}$ $b\bar{b}, b\bar{r}, b\bar{g}$	0
Higgs boson	H	0	0	-	115-129 GeV

Table 2.1: The elementary particles of the SM. Three generations of leptons and quarks together with the gauge bosons of the fundamental interactions and the Higgs boson. For each charged particle there exists an antiparticle with same mass and opposite electric charge. Particle masses and limits taken from [2], apart from the SM Higgs boson limits which are taken from [3, 4].

Description	Free Parameters
Lepton masses	m_e, m_μ, m_τ m_{ν_e}, m_{ν_μ}, m_{ν_τ}
Quark masses	m_u, m_c, m_t m_d, m_s, m_b
Lepton mixing (PMNS matrix)	Θ_{ATM}, Θ_{reactor}, Θ_{solar}, δ_{Dirac}
Quark mixing (CKM matrix)	Θ_{12}, Θ_{13}, Θ_{23}, δ
Coupling constants	g_e, g_w, g_s
Higgs doublet	m_H, m_W
Strong CP	Θ_{CP}

Table 2.2: Free parameters of the SM. For the lepton mixing the simple Dirac case has been chosen.

theless, this does not reduce the high number of free parameters of the SM which can be seen as unsatisfactory for a fundamental theory.

As a consequence of the unified theoretical description of the electromagnetic and the weak interaction at energies of the electroweak scale $M_{EW} \sim 10^2$ GeV, one can also expect a unification with the strong interaction at even higher energies, known as the GUT (Grand Unified Theories) scale $M_{GUT} \sim 10^{16}$ GeV. The simplest GUT that contains the SM has a $\boldsymbol{SU(5)}$ gauge group[3] but also other have been proposed. The energy scale dependence of the three gauge couplings of the SM suggests that these are close to each other at the GUT scale. However, a much better matching of the running gauge couplings could be achieved by introducing supersymmetric models (fig.2.1).

At energies of the Planck scale $M_{Pl} \sim 10^{18}$ GeV, the gravitational interaction is of the same order in strength as the other fundamental interactions and has to be incorporated into a theory of particle physics. A serious problem for extensions of the SM to higher energies is the huge difference between the Planck scale and the electroweak scale up to which the parameters of the SM can be measured. Using renormalization to transform the fundamental quantities into observables it is found that extreme fine-tuning of the quantum loop corrections is necessary. This, addressed as the hierarchy problem, is a strong motivation for supersymmetry and is discussed in the following section.

[3]Experimentally $\boldsymbol{SU(5)}$ is ruled out by measurements of the proton lifetime [7].

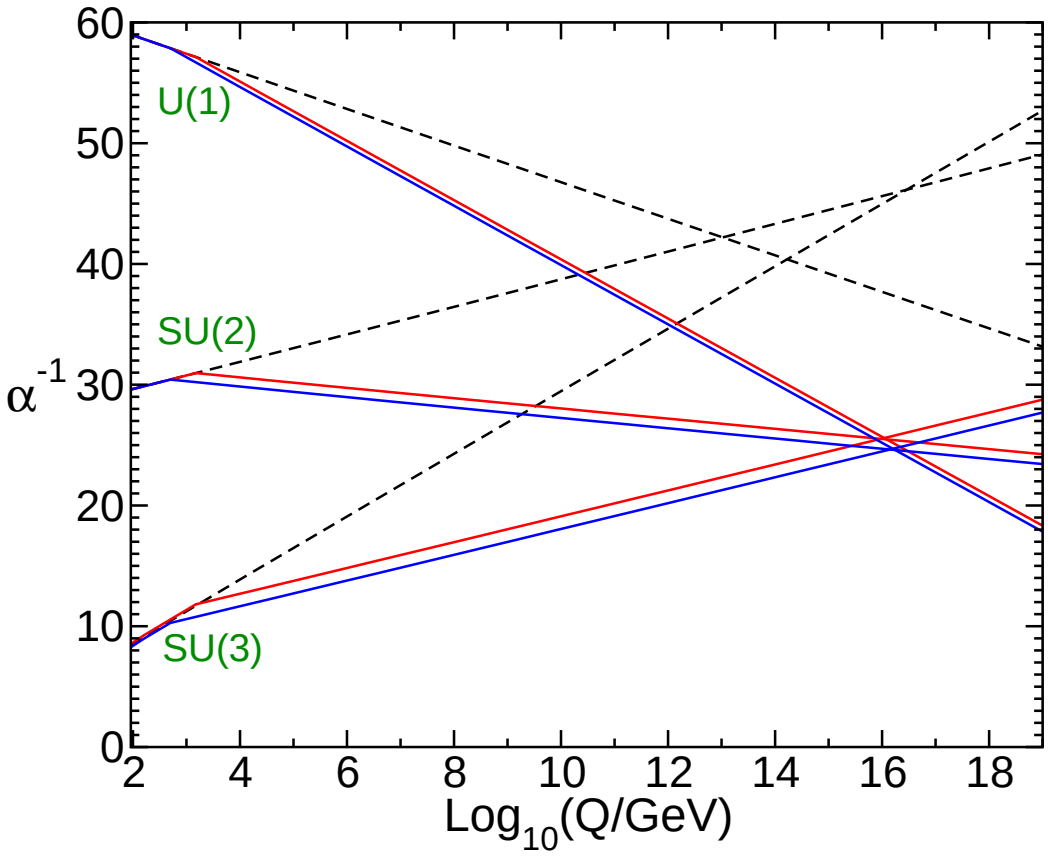


Figure 2.1: Evolution of the inverse gauge couplings of the electroweak interaction (α_1^{-1} for $\boldsymbol{U(1)_Y}$ and α_2^{-1} for $\boldsymbol{SU(2)_L}$) and the strong interaction (α_3^{-1}) in the SM (dashed lines) and the MSSM (solid lines). From [8].

2.2 Supersymmetry

The long-established theory of supersymmetry proposes a hypothetical symmetry between elementary fermions and bosons. Fermion states are transformed to boson states and vice versa via an operator Q which shifts the spin by $\frac{1}{2}$ and leaves the masses and the gauge charges unchanged. None of the *superpartners* defined by this procedure can be identified with a SM particle, hence new particles are introduced which significantly increases the number of elementary particles.

From the absence of observed supersymmetric particles, we know that supersymmetry must be broken. However, the breaking mechanism is unknown and there are several supersymmetry breaking scenarios proposed from theory which results a variety of models.

2.2.1 Motivations for Supersymmetry

The SM introduces particle masses through the Higgs mechanism, but at the same time it poses a theoretical problem with the Higgs boson mass, known as the hierarchy problem. In the Lagrangian, the coupling of the fermions to the Higgs field is described by a $-\lambda_f H\bar{f}f$ term, where the Yukawa coupling is largest for the heaviest SM fermion, the top quark with $\lambda_f \sim 1$. As a consequence, each fermion gives quantum corrections to the Higgs boson mass (from the Feynman diagram in fig. 2.2a). The dominant contribution is:

$$\Delta m_H^2|_f = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \dots \tag{2.1}$$

Here Λ_{UV} is the ultra-violet cutoff parameter which represents the energy scale at which new physics alters the high-energy behavior. A natural choice of Λ_{UV} would be the Planck scale, since we know that new physics has to appear here. However, the scale of the effective Higgs boson mass is far below Λ_{UV} which means that the Higgs boson mass would be extremely sensitive to a fine-tuning cancellation between the quadratic radiative corrections and the bare mass.

Choosing Λ_{UV} not too large, one still needs a new physics model that alters the propagators and cuts off the loop integral already at this scale. Furthermore, new heavy particles (with masses large compared to m_H) would induce similar problems via the second term in eq. 2.1 in the case of fermions and also in the case of heavy scalars as given by (corresponding Feynman diagram in fig. 2.2b):

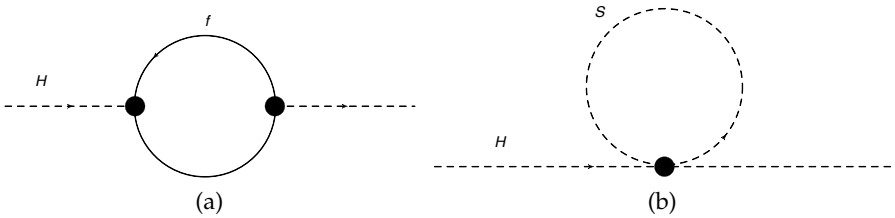


Figure 2.2: One-loop quantum corrections to the Higgs squared mass parameter m_H^2, due to (a) a Dirac fermion f, and (b) a scalar S.

$$\Delta m_H^2|_S \;=\; \frac{\lambda_S}{16\pi^2}\left(\Lambda_{\rm UV}^2 - 2m_S^2\,\ln\frac{\Lambda_{\rm UV}}{m_S}\right). \tag{2.2}$$

It is important to notice that the contributions to the Higgs mass correction given by eq. 2.1 and eq. 2.2 have opposite sign. The inclusion of supersymmetry gives an exact cancellation of the $\Lambda_{\rm UV}^2$ term, since each fermion is associated to two scalars (from the two real components of the Weyl spinor which describes fermions in the SM) with the same coupling to the Higgs field, which means $|\lambda_f|^2 = \lambda_S$.

The remaining contribution can be expressed in an approximation where the mass difference between the fermion and its superpartner scalar boson is small:

$$\Delta m_H^2|_{tot} \;\simeq\; \frac{\lambda_f^2}{4\pi^2}\left(m_S^2 - m_f^2\right)\,\ln\frac{\Lambda_{\rm UV}}{m_S}. \tag{2.3}$$

While an ideal supersymmetry would result in a vanishing correction to the Higgs boson mass, a symmetry breaking that produces mass differences of at most a few TeV would only lead to small quantum corrections [9].

The solution of the hierarchy problem makes supersymmetry an excellent candidate for new physics at the high-energy frontier. As such, supersymmetry is incorporated in super string theories which could be able give a fundamental theory by including quantum gravity.

2.2.2 MSSM and Supergravity

The scope of this section is to give a short description of the Minimal Supersymmetric Standard Model (MSSM) and to discuss the properties of the more constrained minimal Supergravity scenario (mSUGRA) which is one of

the most investigated supersymmetric models. To begin with, the MSSM is introduced in a phenomenological way by describing the particle content and the fundamental couplings.

The particles and their *superpartners* are arranged in supermultiplets which are constructed as irreducible representation of the supersymmetric algebra. Except for the spin, the members of the supermultiplets are identical in all quantum numbers.

The chiral left-handed and right-handed SM fermions are associated to different scalar (spin 0) supersymmetric particles: $\tilde{q}_L$ and $\tilde{q}_R$ (*squarks*), respectively $\tilde{l}_L$ and $\tilde{l}_R$ (*sleptons*). These form chiral supermultiplets (ψ, ϕ), where ψ is the fermion and ϕ is the complex scalar field.

Before electroweak symmetry breaking, the SM gauge bosons are massless, hence their two possible helicity states correspond to the two degrees of freedom of their spin $\frac{1}{2}$ fermion *superpartners* which are named *bino*, *winos* and *gluino*. Their left- and right-handed components must behave identical under gauge transformations and the multiplets are called vector or gauge supermultiplets (A, λ).

The MSSM contains two Higgs doublets which is the minimum number allowed in supersymmetry. Together with their fermion superpartners (*higgsinos*), the Higgs bosons also build chiral supermultiplets.

As a result of the electroweak symmetry breaking mass eigenstates are formed from particles with same quantum numbers. In the SM the γ and the Z boson are the eigenstates of a mass mixing matrix that is made of B^0 and W^0. In the MSSM two more types of mixtures occur with the corresponding mass eigenstates:

- *neutralinos* $\tilde{\chi}^0_{1-4}$ from mixing of $\tilde{B}^0$, $\tilde{W}^0$, $\tilde{h}^0$ and $\tilde{H}^0$,
- *charginos* $\tilde{\chi}^\pm_{1-2}$ from mixing of $\tilde{W}^+$, $\tilde{W}^-$, $\tilde{H}^+$ and $\tilde{H}^-$,

The mass eigenstates of the particles that have been introduced for the MSSM in addition to the known SM particles are summarized in tab. 2.3. This represents the minimal set of new particles which is extended in several supersymmetry scenarios.

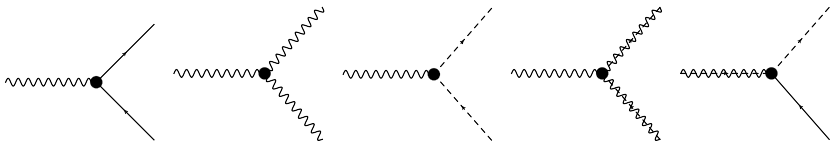
In order to distinguish the particles and their *superpartners*, a new multiplicative quantum number

$$P_R = (-1)^{3(B-L)+2S} \tag{2.4}$$

is introduced, where B (L) are the baryon (lepton) number and S is the spin. Therefore, P_R is +1 for all particles and -1 for all supersymmetric particles.

Name	Spin	P_R	mass eigenstates
Higgs Boson	0	+1	$h^0\ H^0\ A^0\ H^+\ H^-$
Squarks	0	-1	$\tilde{u}_L\ \tilde{u}_R\ \tilde{d}_L\ \tilde{d}_R$ $\tilde{c}_L\ \tilde{c}_R\ \tilde{s}_L\ \tilde{s}_R$ $\tilde{t}_L\ \tilde{t}_R\ \tilde{b}_L\ \tilde{b}_R$
Sleptons	0	-1	$\tilde{e}_L\ \tilde{e}_R\ \tilde{\nu}_{e,L}\ \tilde{\nu}_{e,R}$ $\tilde{\mu}_L\ \tilde{\mu}_R\ \tilde{\nu}_{\mu,L}\ \tilde{\nu}_{\mu,R}$ $\tilde{\tau}_L\ \tilde{\tau}_R\ \tilde{\nu}_{\tau,L}\ \tilde{\nu}_{\tau,R}$
Neutralinos	$\frac{1}{2}$	-1	$\tilde{\chi}_1^0\ \tilde{\chi}_2^0\ \tilde{\chi}_3^0\ \tilde{\chi}_4^0$
Charginos	$\frac{1}{2}$	-1	$\tilde{\chi}_1^\pm\ \tilde{\chi}_2^\pm$
Gluino	$\frac{1}{2}$	-1	$\tilde{g}$

Table 2.3: List of particles and *sparticles* that are incorporated in the MSSM, in addition to the Standard Model particles.



(a) $(A\psi\psi)$ (b) (AAA) (c) $(A\phi\phi)$ (d) $(A\lambda\lambda)$ (e) $(\lambda\phi\psi)$

Figure 2.3: Trilinear couplings in the MSSM. Full lines correspond to fermions, dashed lines to scalar bosons and wiggly lines to vector bosons. *Gauginos* are shown as a combination of dashed and wiggly lines.

There is a good motivation to take P_R as the conserved quantity of a new symmetry, the R-parity. Firstly, it is possible to explain the experimentally proven long life time of the proton even without conservation of the baryon number. Moreover, with R-parity the lightest supersymmetric particle (LSP) is stable. In the MSSM the only weakly interacting $\tilde{\chi}_1^0$ could be the LSP and as such is an excellent candidate for Dark Matter.

The full set of trilinear gauge couplings in the MSSM is shown in fig. 2.3a-2.3e. The SM coupling between fermions and gauge bosons $(A\psi\psi)$ and the three boson vertex (AAA) is extended by the couplings $(A\phi\phi)$, $(A\phi\psi)$ and $(A\lambda\lambda)$, where λ stands for the *gaugions*. Additional to the gauge couplings, there are also Yukawa couplings in the MSSM. It can be shown that the coupling strength of the Yukawa top coupling is not negligible compared to the gauge couplings.[9]

A soft symmetry breaking scenario is introduced, where the supersymmetry

Lagrangian is the sum of two parts:

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}} \tag{2.5}$$

and only the $\mathcal{L}_{\text{soft}}$ part violates supersymmetry. $\mathcal{L}_{\text{SUSY}}$ contains all the gauge and Yukawa couplings and its parameters are determined by the SM, whereas $\mathcal{L}_{\text{soft}}$ contains mass terms and couplings with positive mass dimension which breaks the symmetry. With this procedure, the stabilizing effect on the Higgs boson mass (see eq. 2.3) can be conserved. A derivation of the MSSM *superpotential* and the Lagrangian can be found in [8]. Here, we focus on the additional mass terms in the supersymmetry breaking for the MSSM:

$$\mathcal{L}_{\text{soft}} = -\sum_{\tilde{q},\tilde{l},H_{d,u}} m_{0,i}^2 |\Phi_i|^2 + (\frac{1}{2} m_{1/2,a} \lambda_a \lambda_a - A_{0,i} W_{3,i} - B_0 \mu H_u H_d) + h.c. \tag{2.6}$$

The matrices $m_{0,i}$ introduce mass to the scalar *superpartners* of fermions, $m_{1/2,a}$ introduces masses for the *gauginos* and $W_{3,i}$ represents the trilinear terms with their sign and Yukawa couplings (taken from [9]). The trilinear and bilinear soft breaking terms, arise from the *superpotential* multiplied by a parameter with mass dimension ($A_{0,i}$, B_0) in order to preserve the soft breaking.

Altogether 105 free parameters are additionally introduced in the MSSM. This large number can be reduced by considering that flavor changing neutral currents have not been found and the CP violation must be of the experimentally found level. In general, the number of free parameters in the MSSM is not seen as a fundamental problem because once the breaking mechanism is found it should explain the origin of the parameters.

It is assumed that the soft breaking terms arise indirectly or radiatively, since it seems to be very difficult to achieve a derivation from tree-level renormalizable couplings. For this purpose, a *hidden sector* is introduced where the supersymmetry breaking occurs and which has only very small couplings to the visible sector. The breaking is then mediated from the *hidden sector* to the visible sector via an unknown interaction. A popular idea is to claim that this interaction is gravitational which means that a local symmetry, called *supergravity*, exists which unifies the space-time symmetries with the local supersymmetry at energies above the Planck scale.

A gravitational field theory together with *supergravity* requires a graviton (spin 2 particle) and its *superpartner* the gravitino (spin $\frac{3}{2}$). While the graviton is massless the gravitino often is expected to be heavier than 100 GeV and is

therefore generally not the LSP. Furthermore, both particles are very weakly interacting which make them invisible for collider experiments.

Apart from the motivation with the gravitational interaction, the minimal supergravity scenario (mSUGRA) can effectively be described by a set of assumptions on the free parameters of the $\mathcal{L}_{\text{soft}}$ part of the MSSM Lagrangian (see eq. 2.6), which should hold at the GUT scale:

- $m^2_{0,i} = m^2_0$ (multiplying the identity matrix)
- $m^2_{1/2,a} = m^2_{1/2}$ (multiplying the identity matrix)
- $A_{0,i} = A_0$ (multiplying the Yukawa matrices)

This reduced the number of free parameters significantly. Furthermore, μ can be expressed in terms of the others and making use of the Z boson mass:

$$\mu^2 = \frac{m^2_{H_d} - m^2_{H_u} \tan^2 \beta}{\tan^2 \beta - 1} - \frac{M^2_Z}{2}, \tag{2.7}$$

where B_0 has been replaced by $\tan \beta = v_u / v_d$, the ratio of the vacuum expectation values of the Higgs fields H_u and H_d.

There remain 4 free GUT parameters and a sign:

$$m_0,\ m_{1/2},\ A_0,\ \tan \beta \text{ and } \text{sign}(\mu), \tag{2.8}$$

which make mSUGRA a highly predictive model.

An important contribution to the conditions of mSUGRA arise from the assumption that the mediating interaction between the *hidden sector* and the visible sector is flavor-blind. This is, however, not necessarily required by the gravitational mediation. Using only the above described conditions to reduce the number of free parameters, this model is also referred to as constrained Minimal Supersymmetric extension of the Standard Model (cMSSM).

Other scenarios of soft supersymmetry breaking describe the transition to the visible sector as gauge mediated (GMSB models) or anomaly mediated (AMSB models). In the analysis presented here, the cMSSM scenario is taken as a basis, since it provides a clear signature for the search at the LHC and is easily comparable to analyses from previous experiments.

2.2.3 Expected Signatures at the LHC

Supersymmetric models predict a wide variety of observable processes at high-energy colliders, such as the LHC. Especially the favored soft symmetry

breaking, also called weak-scale supersymmetry, offer the opportunity to explore the theoretically accessible parameter space, since it sets upper limits on the masses of the supersymmetric particles of a few TeV. [4]

Apart from the soft breaking mechanism, R-parity conservation has an important impact on the phenomenology of supersymmetric processes. As a first consequence of R-parity conservation supersymmetric particles are always pair-produced and each decay will produce another supersymmetric particle. Secondly, the LSP is stable since it can not decay into a lighter supersymmetric particle. In the final state, supersymmetric pair-production lead to cascade decays with many leptons, quarks (visible as jets) and two LSP's which manifest themselves as missing energy in most regions of the parameter space.

At the LHC, the highest production cross sections for supersymmetry comes from $\tilde{g}\tilde{g}$, $\tilde{g}\tilde{q}$ and $\tilde{q}\tilde{q}$. Since any further discussion is very model depended, we will focus here on the mSUGRA model (eq. 2.8).

Figure 2.4 spans the parameter space for the two parameters which mainly influence the mass spectrum of mSUGRA: m_0 and $m_{1/2}$. It can be seen, that the *gluino* mass is almost exclusively determined by $m_{1/2}$. The theoretically accessible region in the m_0, $m_{1/2}$ plane is constrained by two effects. Near the $m_{1/2}$-axis the $\tilde{\tau}_1$ would be the LSP which is ruled out because of its electric charge. Additionally, a region near the m_0-axis is forbidden for large m_0 since electroweak symmetry breaking would not be possible.

Indications of observable signatures can be derived from the dominant decays in the four subregions of the m_0, $m_{1/2}$ plane in fig. 2.4. In subregion **1**, leptonic searches are favored compared to other subregions but due to the neutrino production also searches using high missing energy have optimal conditions. The dominant decay into the lightest Higgs boson in subregion **2** makes it possible to search a $h \to b\bar{b}$ signal in this environment. Finally, in the subregions **3** and **4** the *gluino* is lighter than the *squark* and its decay will dominantly involve top quarks. These assumptions also hold qualitatively for a wide variation of the parameter $\tan\beta$.

The expected reaches of the various search channels have been thoroughly studies for the two multi purpose detectors ATLAS and CMS at the LHC. The fully hadronic channel using high missing transverse energy, multi jets and no lepton appears to be competitive for the whole parameter space in mSUGRA [13]. Thus, it is an excellent candidate for early discovery searches.

In fig. 2.5 the expected reach of an early fully hadronic search at LHC with

[4] This has been proposed for several years LHC running with $\sqrt(s) = 14\,\text{TeV}$ at high luminosity. See e.g. [10] and [11].

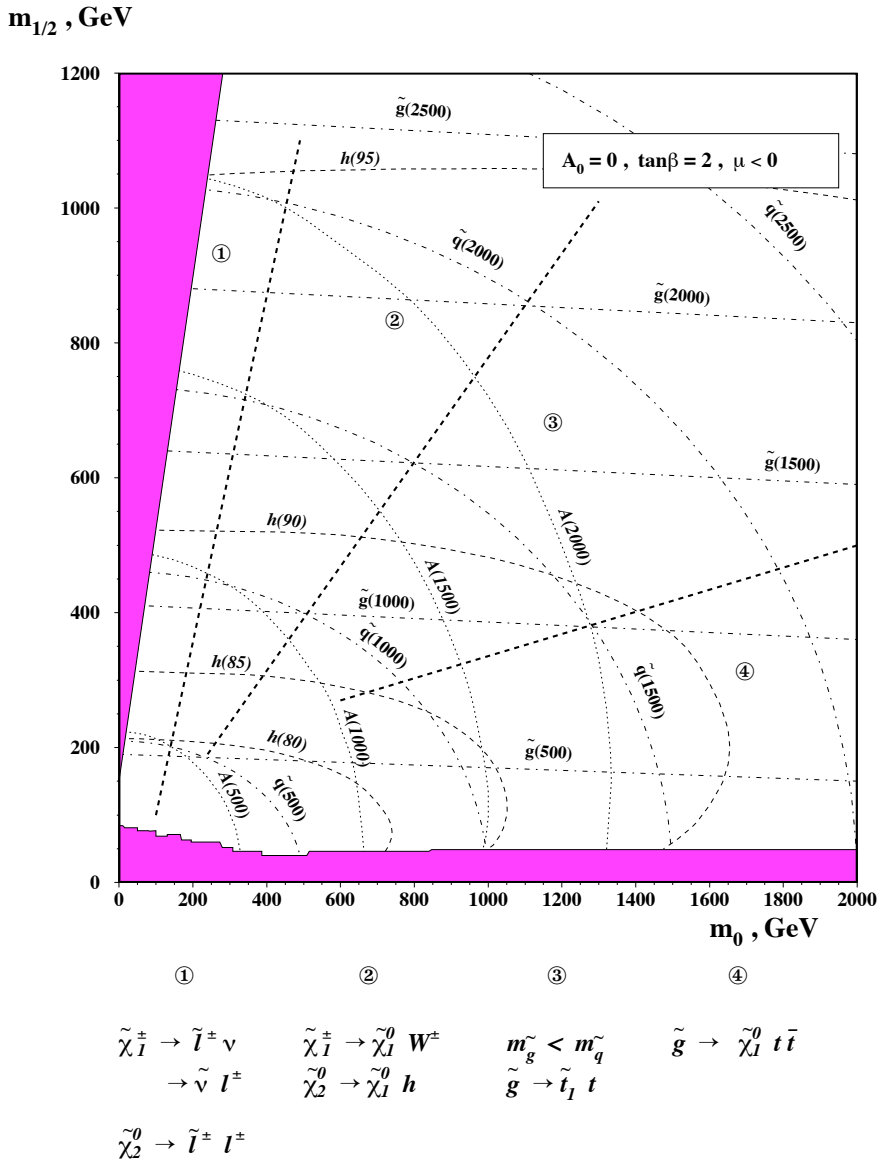


Figure 2.4: Domains of the $(m_0, m_{1/2})$ parameter space at $\tan\beta = 2$ with charachteristic predominant decay modes. Isomass contours for squarks, gluinos, light and pseudoscalar higgses are also shown as dashed lines. The shaded region near the $m_{1/2}$ axis shows the theoretically forbidden region of parameter space, and a similar region along the m_0 axis corresponds to both, theoretically and experimentally excluded portions of parameter space. Regions **1-4** are discussed in the text. From [12].

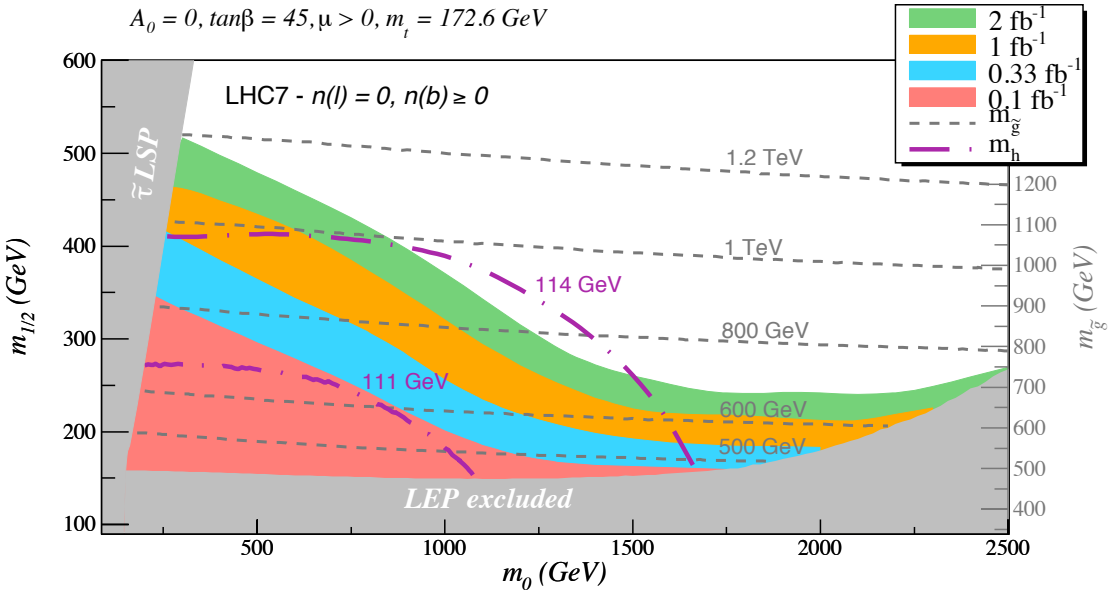


Figure 2.5: The optimized SUSY reach of LHC7 with different integrated luminosities for the $n(lepton) = 0$, $n(bquark) \geq 0$ channel. The fixed mSUGRA parameters are $A_0 = 0$, $\tan\beta = 45$ and $\mu > 0$. Gluino mass contours are shown by the dashed, dark grey curves. Higgs mass contours (dash-dotted purple) are also shown for $m_h = 111$ and 114 GeV. The shaded grey area is excluded due to stau LSPs or no electroweak symmetry breaking, while the shaded area marked "LEP excluded" is excluded by direct LEP bounds on sparticle masses. From [13].

$\sqrt{(s)} = 7$ TeV is presented. The different integrated luminosities shown there makes it possible to compare it to the fully hadronic supersymmetry searches connected to this thesis [1, 14].

3 Experimental setup

The Compact Muon Solenoid (CMS) [15] is one of the two multi purpose detectors at the Large Hadron Collider (LHC) [16–18], which produced the first proton-proton collisions at $\sqrt{s} = 7\,\text{TeV}$ in March 2010. The main physics motivations of the experiment are the detection of the Higgs boson and the search for indications of physics beyond the SM.

The LHC design is introduced in sec. 3.1, and the CMS experiment together with the basic features of its subdetectors are discussed in sec. 3.2.

The reconstruction of the physics objects, which makes use of a particle-flow algorithm, is described in sec. 3.3.

3.1 The Large Hadron Collider

The LHC has been built in the tunnel of the former e^+e^--storage ring LEP at the European Organization for Nuclear Research CERN near Geneva. It has a circumference of 27 km.

Only 10 days after the first successful circulation of proton beams on 10th September 2008, the operation had to be stopped due to a serious incident with a superconducting connection between magnets, leading to a break of the liquid helium containment.

The operation was restarted in mid-November 2009 with proton beams at an energy of 450 GeV. With 7 TeV, half the design center of mass energy, was reached on 30th March 2010. The LHC operated successfully with increasing instantaneous luminosity at $\sqrt{s} = 7\,\text{TeV}$ throughout 2010 (see fig. 3.1). The total good quality dataset of the 7 TeV run in 2010 recorded by CMS corresponds to an integrated luminosity of about $36\,\text{pb}^{-1}$. In 2011, about $5\,\text{fb}^{-1}$ of data has been collected. The design center of mass energy ($\sqrt{s} = 14$ TeV) will not be accessable before the foreseen upgrade in 2013.

With a peak luminosity of $4.67 \cdot 10^{32}\ \text{cm}^{-2}\text{s}^{-1}$ a new world record was established on 21st April 2011, and the since then the luminosity has been further increased. The design luminosity of the LHC is $\sim 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$. The large flux of particles from the proton-proton interactions results in high radiation levels, which require radiation-hard detectors and front-end

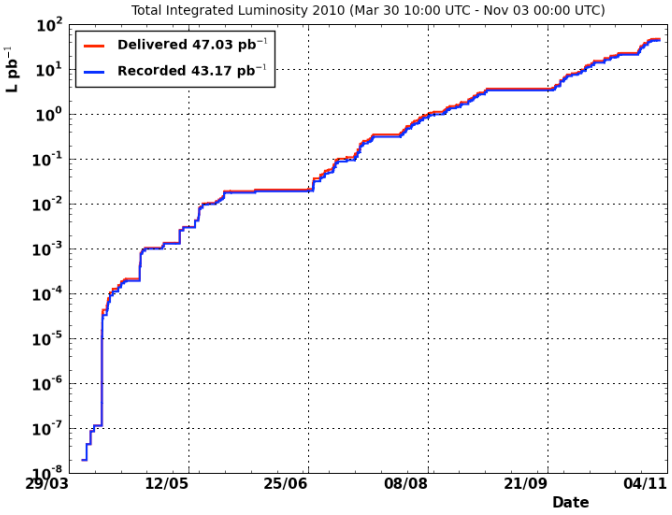


Figure 3.1: Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) during stable beams at $\sqrt{s} = 7\,\text{TeV}$ in 2010. From [19].

electronics.

At four intersection points of the beams large experiments are installed. The LHC-b experiment is dedicated to the study of CP violation in B meson decays, the heavy ion experiment ALICE will among other things investigate the properties of an exotic phase of matter, the quark gluon plasma, and the two multi purpose detectors ATLAS [20, 21] and CMS compete in the search for the Higgs boson and new physics and improve measurements of the parameters of the SM.

3.2 The CMS Experiment

The CMS detector at the LHC is designed for the discovery and investigation of a wide range of phenomena. Some of the corresponding signatures have been discussed in sec. 2.2.3, while many more have been included in the planning of CMS [22]. Basically a detector is required that is prepared for nearly everything, but several demands on the performance of the CMS detector can be made nevertheless:

- Since a promising signature of a light Higgs boson (less than 150 GeV)

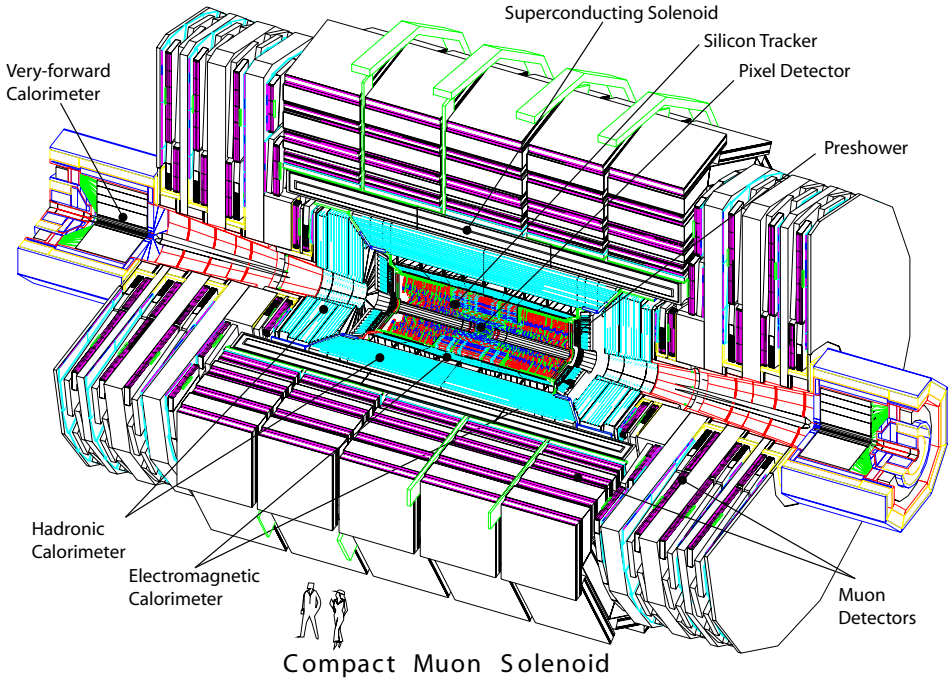


Figure 3.2: Three-dimensional view of the CMS detector [23].

is its decay to two photons, an excellent electromagnetic calorimeter is required.

- The high muon identification efficiency, as well as a good muon momentum resolution, are essentially important for other Higgs signatures.
- The best signature for supersymmetric particles will be missing transverse energy (or momentum), hence this search will benefit from a detector that is nearly hermetically closed and has also good electromagnetic and hadronic calorimeter resolutions.
- All searches require very good reconstruction of the momenta of charged particles and of all vertices. These tasks depend crucially on a high quality central tracking system.

The design of CMS is sketched in fig. 3.2. It has the layered structure that is typical for collider experiments. The overall shape is dictated by the choice of a solenoid magnet for bending particle tracks. A high magnetic

field was chosen, in order to achieve a good momentum resolution. The 4 T superconducting solenoid is large enough to accomodate the silicon pixel and strip trackers, the electromagnetic calorimeter and most of the hadronic calorimeter inside. The outer shell is made up of the return yoke that serves as the main support of the detector, and also allows four muon stations to be integrated. In total the CMS detector reaches a length of 21.5 m, a diameter of 15 m and a weight of 12,500 t.

The canonical coordinate system of the detector is a cylindrical one, with the positive z-axis counter clock-wise along the direction of the beam pipe. The y-axis points vertically upward and the x-axis points radially toward the center of the LHC. The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The polar angle θ can be replaced by the pseudorapidity that is defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad . \tag{3.1}$$

Differences in the pseudorapidity $\Delta\eta$ and the azimuthal angle $\Delta\phi$ are both invariant under Lorentz boost along the z-axis. Therefore, the difference ΔR in the (η, ϕ)-metric is a useful quantity.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.2}$$

The subdetectors of CMS are introduced in the following subsections, starting with the devices closest to the interaction point.

3.2.1 The Inner Tracker

The CMS inner tracker is subdivided into the barrel region ($|\eta| < 1.2$) and two endcaps, which provide a coverage up to $|\eta| < 2.4$. In the barrel part, close to the interaction vertex, are three layers of hybrid pixel detectors at radii r of 4, 7 and 11 cm. The size of the pixel is $100 \times 150\ \mu m^2$. In addition, ten layers of silicon microstrip detectors are placed at r between 20 and 115 cm. The tracker is operated at a temperature of $-10°C$ to increase the lifetime of the silicon modules in the high radiation environment near the interaction point.

In the two endcaps are two pixel and nine micro-strip layers each. Altogether, the tracking system consists of 66 million pixels and 9.6 million silicon strips.

The inner tracking system allows the precise measurement of charged particles which initiate signals (hits) within silicon sensors. The track reconstruction is done by fitting these hits to a helix. The tracks of the charged

particles are bend due to the strong magnetic field. Their transverse momenta can be computed by using:

$$p_T[\,\mathrm{GeV}] = 0.3B\rho \tag{3.3}$$

where ρ is the radius of the bent track in meters and B is the magnetic field in Tesla.

The resolution of the transverse momenta of high energetic charged particles ($p_T > 100\,\mathrm{GeV}$) is about 1-2% in the central part, decreasing towards higher $|\eta|$. The track reconstruction efficiency is also very high, e.g. $\approx 85\%$ for pions with $p_T = 2\,\mathrm{GeV}$ and over 90% for pions with $p_T = 10\,\mathrm{GeV}$.

With the high resolution of the pixel tracker also a precise vertex reconstruction is possible with ≈ 25 µm spatial resolution and ≈ 20 µm for the z measurement [24]. This allows the identification of secondary vertices that are produced by the relatively longer lifetime of B mesons.

3.2.2 The Calorimeters

The calorimetric system provides an important contribution for the event reconstruction. In the electromagnetic calorimeter the energies of electrons, positrons and photons are measured. Particles of hadronic showers, referred to as jets, deposit most of their energy within the hadronic calorimeter. The calorimeters need to be calibrated to account for their non-linear and non-compensating response.

The electromagnetic calorimeter (ECAL)

Lead tungstate ($PbWO_4$) crystals are used to measure the energy of electromagnetic particles. In the barrel section ($|\eta| < 1.479$) the crystals are arranged in an η-θ-grid. The crystals have a front face cross-section of $\approx 22 \times 22$ mm^2 and a length of 230 mm corresponding to $25.8X_0$ (radiation length $X_0 = 0.89$ cm). The two endcaps extend the coverage up to $|\eta| = 3.0$.

The performance of the energy resolution is parametrized with three term, firstly a stochastic term, secondly a noise term and thirdly a constant term. The measurement in a test beam resulted in:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \quad , \tag{3.4}$$

where E is in GeV [15].

A selective readout is used for the ECAL, which means that only a part of the calorimeter is read out without zero suppression (at about $3\sigma_{noise}$). In the barrel case, these are the trigger towers (5×5 crystals) above a threshold and those in the direct neighborhood of such a tower. During operation, some of these readout channels were found to be faulty, though the information from the trigger system was still available.

The hadronic calorimeter (HCAL)

The HCAL consists of the four parts, barrel region (HB), endcaps (HE), hadron outer detector (HO) and very forward calorimeter (HF). The absorber material is mostly brass, since it has a short absorption length and is non-magnetic.

Plastic scintillator tiles are used, which are read out with embedded wavelength-shifting fibers. While the HB and HE are located inside the magnet coil surrounding the ECAL, the HO is an additional layer of scintillators, lining the outside of the coil. However, in the this analysis the HO is not used in the jet reconstruction due to its high noise level.

The barrel region covers the pseudorapidity region of $|\eta| < 1.4$, and the endcaps cover the region of $1.3 < |\eta| < 3.0$. The two devices of HF ($3.0 < |\eta| < 5.0$) complete the good hermeticity which is essential for the $\not{H}_{\mathrm{T}}$ measurement. Th barrel consists of 2304 towers and has a segmentation of $\eta \times \Phi = 0.087 \times 0.087$. The granularity in the other parts is chosen such that the jet energy resolution, as a function of E_{T} is similar in the three parts (HB, HE and HF) of the HCAL.

As the HCAL has a worse energy resolution compared to other detector components, the jet energy resolution can be significantly improved by using information from several sub-detectors as described below in sec. 3.3.

3.2.3 The Muon System

The CMS muon system exploits three technologies, drift tubes (DT) in the barrel region ($|\eta| < 1.2$), cathode strip chambers (CSC) in the endcap region (up to $|\eta| < 2.4$) and resistive plate chambers (RPC) in both the barrel and the endcap. This choice has been made because of the different radiation environments and the large surface that is covered. Due to the presence of the return yoke, the magnetic field is of relatively low strength inside the barrel region and the muon rate is low, allowing a drift chamber tracking detector. Whereas in the two endcaps, the muon rate as well as the magnetic field penetration is high. The resistive plate chambers are added to provide a

fast response with good time resolution which is important for the first trigger level (L1).

The momentum measurement of the muon system is essentially determined from the muon bending angle at the exit of the 4 T coil. The tracking system and the muon system are used together to reconstruct the kinematics of the muons.

Muons have a very high reconstruction efficiency of 99% [25]. For low energetic muons the resolution is dominated by the inner tracker, but with increasing muon energy the energy loss in the material in front of the muon system becomes negligible and the longer lever arm for the measurement of the track curvature is needed. The p_{T} resolution of low energetic muons ($< 100\,\mathrm{GeV}$) is between 1% in the barrel and 6% in the endcap region, and still better than 10% for muons up to 1 TeV in the barrel.

3.2.4 The Trigger System

The CMS trigger system has to reduce the interaction rate by a factor of nearly 10^6 (at design luminosity), in order to achieve the rate of about 100 interactions/sec that can be written to archival media. The trigger and data acquisition system consists of 4 parts: the detector electronics, the Level-1 trigger (L1), the readout network, and an on-line event filter system that executes the software of the High-Level Triggers (HLT).

The information used for the data reduction in the Level-1 trigger is taken from coarse measurements in the calorimeters and the muon system. For the decision wheter an event is accepted, the thresholds on the transverse energy E_T or transverse momentum p_{T} of objects such as photons, electrons, muons and jets as well as E_T^{miss} (or $\not{H}_{\mathrm{T}}$) and sum of E_T (or H_{T}) are applied. For these quantities a quick preliminary reconstruction is done using the fastest detector components such as the RPCs. The decision has to be available after a limited time of 3.2 µs. During this time, the data is kept in the readout buffers.

If an event is accepted by the Level-1 trigger, it will be sent to the High-Level triggers. In the High-Level triggers, more time and more detailed information from the detectors are available to analyse an event. The processing takes place in a farm of about 1000 commercial CPUs and can take up to 1 s of processing time per event. In this step, the output rate of the Level-1 trigger of 50 kHz is reduced to an event rate of about 150 Hz with a size of 1 MB per event.

The trigger thresholds are constantly adjusted to match the increasing instant luminosity. To keep trigger thresholds low, cross triggers are used which include simultaneous cuts on several physics objects.

3.3 The Particle-Flow Algorithm

The physics objects used in this analysis are electrons ($e^{\pm}$), muons ($\mu^{\pm}$), photons and jets which originate from gluon radiation or quarks. These objects can be measured independently in the corresponding subdetectors, which is e.g. the ECAL for photons and electrons or both calorimeters for the jets. But, since CMS has a high resolution inner tracker, it is very promising to include its measurements in the reconstruction of all objects.

The idea of the CMS particle-flow algorithm [26, 27] is to individually identify and reconstruct all particles produced in the collision, namely charged hadrons, photons, neutral hadrons, muons, and electrons, by combining the information from the tracker, the calorimeters and the muon system.

With the particle-flow algorithm it is possible to better resolve electrons and photons that are inside a jet cone. Significant improvements in the jet energy resolution can be made using particle-flow jets, which are built of the individually measured hadrons, compared to calorimeter jets. This is due to the large contribution of charged hadrons to the jet energy. The improved jet energy resolution leads to a significant improvement in the resolution of missing transverse momentum ($\not{H}_{\mathrm{T}}$) which is important for the analysis presented here.

4 Search Design

The analysis presented in this thesis has been published in [1]. This chapter reviews the search criteria that have been defined for this analysis. The events selection presented here serves as a basis for the data-driven background estimation methods, discussed in the following chapters. Note that this thesis contributes to the analysis primarily with the development of a QCD background estimation method presented in ch. 6.

In sec. 4.1 the search strategy of the analysis is recapitulated. Details on the simulated samples used for the validation are given in sec. 4.2 and the event selection is introduced in sec. 4.3.

The chapter is concluded with a comparison between the 2010 dataset and Monte Carlo simulation of the backgrounds and two benchmark signals in sec. 4.4.

4.1 Search Strategy

The presented analysis was one of the first multijet searches for new physics with pp collisions at a center of mass energy of $\sqrt{s} = 7$ TeV using $36\,\mathrm{pb}^{-1}$ of data collected with the CMS detector at LHC in 2010. The first aim of the analysis was to detect visible signs of new physics already at this early stage of the experiment. The second aim was to establish reliable data-driven background estimation methods and check upon their usability for the succeeding analyses.

The analysis is focused on one central observable which is chosen to be the missing transverse momentum ($\not{H}_\mathrm{T}$). This allows the search to have sensitivity for the wide range of new physics models that yield a hadronic final state with missing momentum. The cuts of the baseline selection, discussed below, are chosen such that necessary background suppression results in a minimal kinematical bias of the expected signal. In the important case of the constrained Minimal Supersymmtreic extension of the Standard Model (cMSSM, see sec. 2.2.3), the selection efficiency of signal events is especially good for models where the sparticle masses are low enough to be produced with sizeable yield at limited integrated luminosities.

Regarding the lack of a visible signal event excess after the baseline selection, the analysis is extended with two evolved selections in order to gain in sensitivity for models with higher sparticle masses.

As part of a CMS collaboration wide strategy, this analysis is restricted to the signature of no leptons in the final state. Other searches with leptonic final states are accomplished in parallel [28–30]. Since the different search regions are statistically decoupled are combination of the results will be in principle possible.

4.2 Monte Carlo Samples

All contributing background processes for this search have been studied using Monte Carlo simulations. All these samples have been produced using PYTHIA [31, 32] and MADGRAPH [33] together with a detailed Geant-based [34, 35] CMS detector simulation. The important backgrounds, QCD multijet, $t\bar{t}$, $W + \text{jets}$ and $Z \to \nu\bar{\nu}$ are all generated with MADGRAPH, though for some Monte Carlo samples the choice depended on availability (especially for some tunes, see tab. 4.2).

The important QCD background is studied with both PYTHIA and MADGRAPH samples, which is especially important for the validation of the factorization method presented in sec. 6.1.

Further processes which have minor influence on the final selection but can be important for the control regions of the data-driven methods are γ+jets, dibosons and single top.

The properties of the simulated low mass benchmark cMSSM points LM0 and LM1 are given in tab. 4.1.

	cross section (NLO)	m_0	$m_{1/2}$	A_0	$\tan\beta$	μ
LM0	54.9 pb	200 GeV	160 GeV	-400	10	> 0
LM1	6.5 pb	60 GeV	250 GeV	0	10	> 0

Table 4.1: CMS low mass benchmark points for cMSSM. The NLO cross sections have been calculated with PROSPINO [36].

4.3 Event Selection, Trigger and Cleaning

The events used in this analysis are collected by trigger paths based on the quantity H_T^{trig}, defined as the scalar sum of transverse energy of reconstructed

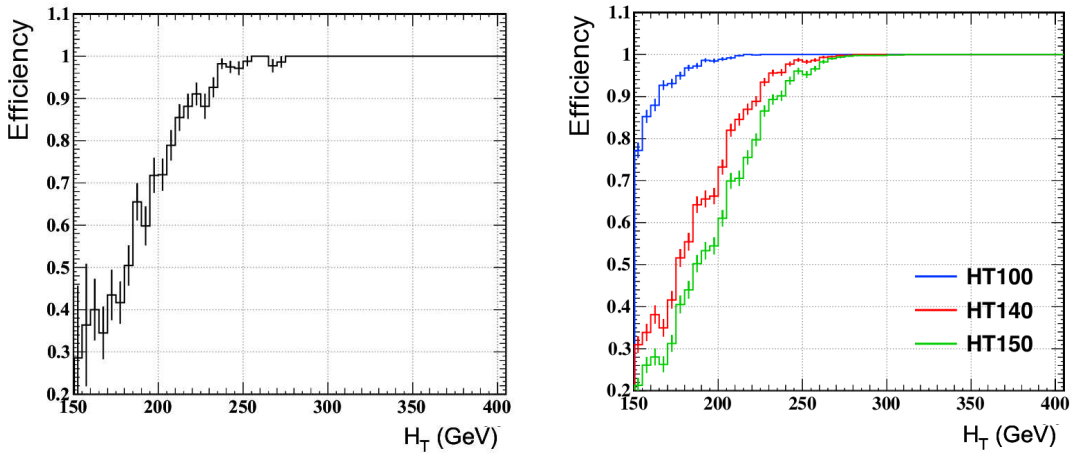


Figure 4.1: Trigger efficiency curves as a function of the particle-flow reconstructed offline H_T, measured in data using the `Jet15U` single-jet trigger (left) and in simulation for the LMo benchmark signal (right). For the baseline selection $H_T > 300\,\text{GeV}$ is required. From [37].

calorimeter jets (without response correction) with $p_T > 20\,\text{GeV}$ and $|\eta| < 5$. The threshold of the lowest unprescaled H_T trigger increased during 2010 data taking (due to the increase in luminosity) and reached finally 150 GeV.

The choice of the H_T trigger meets the requirements from the search strategy discussed above. It has a good acceptance for SUSY signals with low sparticle masses. Furthermore, the use of this trigger enables the simultaneous collection of a multijet control sample with low missing momentum which is used in the data-driven QCD estimation methods (see ch. 6).

The trigger efficiency as a function of the particle-flow based H_T (defined in sec. 4.3) have been constructed for both data and Monte Carlo signal samples, as shown in Figure 4.1. The data measurement makes use of the `Jet15U` single-jet trigger which triggers on a minimal uncorrected jet p_T of 15 GeV (a jet threshold below that used in the H_T^{trig} calculation). The data measurement shows full efficiency for the set of H_T triggers which are applied in this analysis. Also, the low mass LMo signal point is fully efficient for the simulated `HT150U` trigger at an offline H_T cut of 300 GeV.

The cuts for the baseline selection are listed below.

All physics objects used in the offline event selection, namely jets, electrons and muons, are reconstructed in a consistent way using the CMS particle-flow event description explained in sec. 3.3.

- Events are collected based on their H_T^{trig}.
- At least three central jets with $p_\mathrm{T} > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$ are required. Jets are clustered with the anti-kT ($D = 0.5$) cone algorithm [38]. Jets are corrected using Monte Carlo derived correction factors and, for data events, an additional parametrized residual jet energy correction derived from the data is applied [39].
- $H_\mathrm{T} > 300\,\mathrm{GeV}$, with H_T defined as the scalar sum of the p_T's of all the jets. Jets are required to fulfill the jet definition from above ($p_\mathrm{T} > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$). $H_\mathrm{T} = \sum_i |p_\mathrm{T}^{jet_i}|$.
- $\not{H}_\mathrm{T} > 150\,\mathrm{GeV}$, with $\not{H}_\mathrm{T}$ defined as the magnitude of the missing momentum vector, which is for simplicity denoted by the same symbol. The $\not{H}_\mathrm{T}$ vector is the negative *vectorial* sum of the p_T's of the jets in the events, where in this case jets are required to satisfy $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 5$, in order to suppress high $\not{H}_\mathrm{T}$ tails from QCD multijet events. $\not{H}_\mathrm{T} = -\sum_i p_\mathrm{T}^{jet_i}$.
- $|\Delta\phi(J_n, \not{H}_\mathrm{T})| > 0.5$, $n = 1, 2$ and $|\Delta\phi(J_3, \not{H}_\mathrm{T})| > 0.3$, vetoing alignment in the transverse plane between any one of the first three jets J_i and the $\not{H}_\mathrm{T}$ as defined above. The cut on $\Delta\phi$ at 0.5 was chosen to be equal to the jet cone size, while the looser cut at 0.3 was chosen to retain signal efficiency.
- No isolated muons and electrons in the event. Muon candidates are required to have $p_\mathrm{T} \geq 10\,\mathrm{GeV}$ and $|\eta| < 2.4$, to satisfy requirements for a global muon with good quality global and tracker tracks, to match to the primary vertex within $200\mu m$ transversely and 1 cm longitudinally with respect to the beam axis, and to be isolated, by having the value of the particle-flow-based relative isolation variable, defined as $\mu_{Iso} = \frac{\sum_{trk}^{\Delta R=0.3} p_T^{chargedhadron} + \sum_{ecal}^{\Delta R=0.3} E_T^{neutralhadron} + \sum_{hcal}^{\Delta R=0.3} E_T^{photons}}{p_\mathrm{T}}$, smaller than 20%. Electrons similarly should have $p_\mathrm{T} \geq 10\,\mathrm{GeV}$ and $|\eta| < 2.5$ (excluding the transition region $1.44 < |\eta| < 1.57$), be attached to a good-quality GSF track [40], and match to the primary vertex and be isolated as described above for muons.

While the cut on $\not{H}_T$ suppresses the vast majority of multijet QCD events, the requirements on the $\Delta\phi$ between $\not{H}_T$ and leading jets removes most events which have a single mismeasured jet that leads to high $\not{H}_T$. Inverting these cuts makes it possible to measure QCD dominated control regions which is used for the factorization method in sec. 6.1.

Although, it different cut values for the angular variables $\Delta\phi_1$, $\Delta\phi_2$ and $\Delta\phi_3$ have been chosen, the basic idea of this requirement can be put into one variable $\Delta\phi_{\text{min}}$ which is defined as the minimum value of the three. With this definition the variable will be used in the following. Similar to a higher cut on $\Delta\phi_3$, an inclusion of a forth leading jet in an event would significantly reduce the signal efficiency.

The leptonic final states of $t\bar{t}$ and V+jets processes are efficiently suppressed by using a loose lepton definition for the veto described above. Also here, an inversion of the cut is later on used to estimate such backgrounds (see sec. 5.1.1).

Two central search requirements are tightened individually, resulting in two additional search regions:

- a high-$\not{H}_T$ search region, with $\not{H}_T > 250\,\text{GeV}$;
- a high-H_T search region, with $H_T > 500\,\text{GeV}$.

In the first case, the high $\not{H}_T$ is motivated by the search for R-parity conserving supersymmetry (eq. 2.4), or more generally a dark matter candidate, and additionally because of the high background rejection. The second case benefits from cascade decays where higher object multiplicities are expected and more energy is transferred to visible energy rather then to dark matter candidates (e.g. LSP).

The analysis forgoes a preceding optimization of cuts on the two central search variables $\not{H}_T$ and H_T. However, this thesis exploits a possible search cut optimization procedure, which aims to maximize the sensitivity for important regions of a two parameter plane of the cMSSM (sec. 7.3).

Cleaning of Events

Since $\not{H}_T$ is an important variable for the analysis, inaccurate event reconstruction that leads to fake $\not{H}_T$ has to be investigated. Possible ways to remove fake $\not{H}_T$ were investigated using simulated multijet and signal samples, as well as the full 2010 data sample collected by the CMS experiment [41]. While some sources of fake $\not{H}_T$ can be traced back to the muon and electron reconstruction algorithms, and therefore only affect particle flow $\not{H}_T$, other are also

present for calorimeter-only $\not\!\!H_T$. Whenever possible, event filtering tools were designed and characterized using simulated data before being applied in the analysis.

For muons, two filters were introduced that were shown in simulation to suppress $\not\!\!H_T$ tails. Events with muons for which the tracker p_T and global track p_T deviate by more than 10% are rejected. Also, events are vetoed that contain particle-flow muons that absorb a calorimeter deposit with energy larger than the muon's momentum.

To reject fake energy deposits in the calorimeters, noise in the hadron calorimeter and beam halo backgrounds are rejected using CMS standard cleaning recipes [42, 43].

A new source of rare noise was identified that simultaneously affects the ECAL endcaps (EE) and the muon systems. Requiring the number of energy deposits in the EE to be smaller than 2500 was shown to suppress this noise adequately. Finally, $\not\!\!H_T$ also arises due to losses of energy for crystals in the ECAL that are not read out, mostly because of malfunctioning on-detector electronics. Two algorithms are used to identify and reject such events, as detailed in [44]. One uses the trigger-primitive information to identify the presence of an energy deposit above the saturation limit of 64 GeV in masked so-called towers of 5-by-5 crystals. The other filter puts a cut-off of 10 GeV on the amount of energy allowed in the crystals surrounding masked towers for which the trigger-primitive information is missing.

Also, tracking-related problems can produce events with a large fake $\not\!\!H_T$ that pass the event selection. Beam-background events can create such a large number of clusters in the pixels or silicon strips that the tracking algorithm of the standard reconstruction can not run completely. Also sattelite collisions were observed, displaced by 75 cm from the nominal interaction point, for which the standard CMS tracking algorithm parameters prevent reconstruction. A large apparent $\not\!\!H_T$ can be induced in such events by assuming jets to come from the nominal interaction point. To deal with these issues, a good primary vertex (ndof > 4) is required within the CMS luminous region of 24 cm in length and 2 cm in radius. Next, the standard beam-background veto is applied, requiring events with more than 10 tracks to have at least 25% of the tracks to be of good quality. Finally, an additional veto is applied to events for which the scalar sum of the p_T's of tracks associated to the primary vertex is less than 10% of the scalar sum of jet p_T's for all jets within the tracker acceptance.

In Figure 4.2 the integrated effect of all the cleanup cuts is shown for the Monte Carlo QCD multijet samples and for the data. Since some of the applied filters are sensitive to general reconstruction inaccuracies of the physics objects,

small deviations are also present in the simulation.

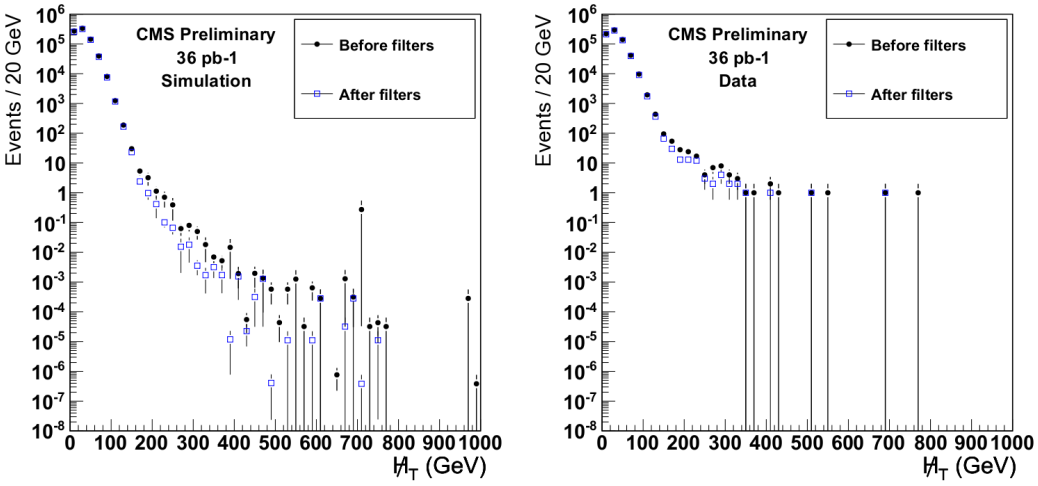


Figure 4.2: Distribution of $\not\!\!H_T$ for the full QCD sample (left) and the full data sample (right), before and after all event filters. The data sample being dominated by electroweak processes with real $\not\!\!H_T$, it cannot be directly compared to the QCD sample in the high $\not\!\!H_T$ region. From [37].

4.4 Data-Simulation Comparison

While in this analysis all the backgrounds are estimated from data (ch. 5, ch. 6), a direct comparison of data and simulation is an important starting point. However, final numbers will not be drawn from this comparison and systematic uncertainties are only needed for the simulated events of the signal scan used for the limit calculation in ch. 7.

The event yields in data and Monte Carlo simulated samples for the final steps of the event selection are summarized in tab. 4.2. The triggering, event cleaning and all other cuts have been applied before.

The distributions of data and simulation are compared for the observables $\not\!\!H_T$ and H_T in fig. 4.3 after the baseline selection. While in both distributions the sum of the Monte Carlo simulated processes is in agreement with the data, which is remarkable at this early stage of the experiment, some weaknesses of

	Baseline no $\Delta\phi$ cuts no e/μ veto	Baseline no e/μ veto	Baseline selection	high-$\not{H}_T$ selection	high-H_T selection
Data	482	180	111	15	40
Sum SM MC	418	155	94	14	32
LMo	391	303	231	84	126
LM1	71	60	45	31	34
$Z \to \nu\bar{\nu}$	27	21	21	6	6
$t\bar{t}_{semilep}(\mu\|\|\tau_\mu)$	21	15	5	1	1
$t\bar{t}_{semilep}(e\|\|\tau_e)$	22	15	6	1	2
$t\bar{t}_{semilep}(\tau_h)$ and $t\bar{t}(\tau_h\tau_h)$	15	10	10	1	2
$t\bar{t}_{other}$	13	9	2	0	1
$W(\mu)$ Z2-tune	29	18	4	0	1
$W(e)$ Z2-tune	33	21	6	1	2
$W(\tau_h)$ D6T-tune	17	9	9	3	2
$W(\tau_\mu)$ D6T-tune	7	4	2	1	1
$W(\tau_e)$ D6T-tune	8	4	2	0	1
WW+WZ+ZZ+Vγ+DY	4	2	1	0	0
QCD PYTHIA6	146	14	14	0	7
QCD PYTHIA + PU	222	21	20	0	13
QCD MADGRAPH	92	6	6	0	5

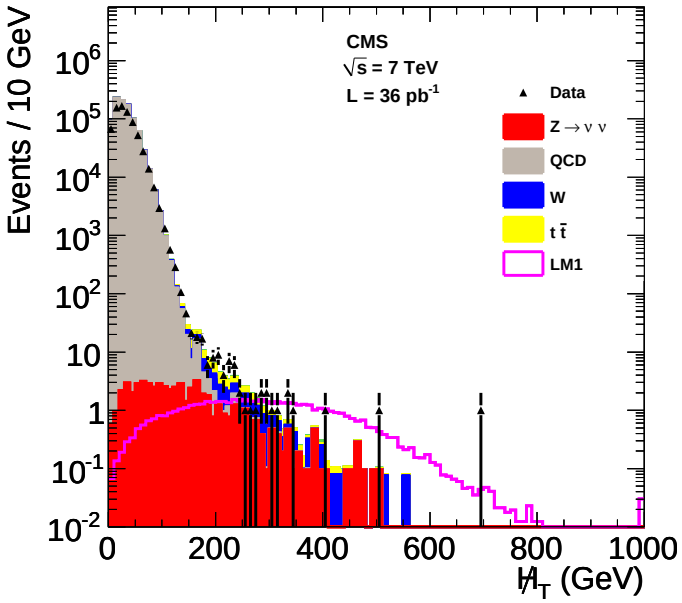
Table 4.2: Event yield in data and Monte Carlo simulation. The simulated samples are normalized to the integrated luminosity of the data: $36\,pb^{-1}$. For the sum of SM MC all given electroweak background processes and QCD PYTHIA plus pile-up are used.

the simulation can be spotted. Firstly, the accuracy of the QCD simulation can be judged from the region $\not{H}_T < 150\,GeV$ shown in the full $\not{H}_T$ distribution in fig. 4.3a. Here, it can be seen that the shapes of the distributions of data and simulation do not agree. The simulation is overestimating the data for $\not{H}_T < 80\,GeV$, while a data excess is visible for the region $100 < \not{H}_T < 150\,GeV$. This can not be significantly improved by making use of the other two available QCD samples, where in addition the integrated number of events is underestimating the data (compare the numbers of the QCD-enriched selection with no $\Delta\phi_{min}$ cut in tab. 4.2).

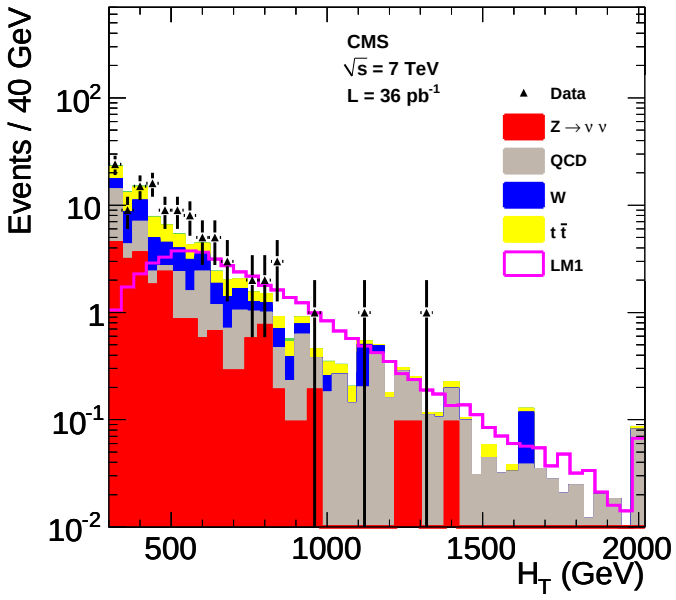
The H_T distribution after the baseline selection (fig. 4.3b) reveals a moderate data excess of ~ 20 % for the whole high-H_T region. This discrepancy can not be sufficiently explained with an inaccurate QCD simulation, since QCD

contributes only $\sim$ 25 % to the total background. A comparison of the numbers after removing the lepton veto (second column in tab. 4.2) indicates that also the $t\bar{t}$ and $W+$jets simulation slightly underestimates the data. Some deviations between data and simulation are expected here since the recommended Z2-tune for some of the $W+$jets was not available and the D6T-tune had to be used instead.

In summary, the size of the discrepancies are at an expected level and a further discussion would require a quantification of the uncertainties of the MC simulated samples. In the following, the background estimation will be exclusively based on data-driven methods, which will be introduced in the next two chapters.



(a)



(b)

Figure 4.3: $\not{H}_T$ and H_T distributions for background and signal with all other cuts from the baseline selection applied. The samples (and sum of SM MC) correspond to tab. 4.2. From the paper [1].

5 Estimation of Electroweak and Top Background from Data

In this chapter, the methods (first published in [45]) that are used to estimate the number of remaining background events from SM processes which produce "real" $\not{H}_T$ via neutrinos in the final state are reviewed. These neutrinos can either be produced together with a charged lepton (via a W boson) or pair-produced from the neutral electroweak process (via a Z boson).

For the first case semileptonic $t\bar{t}$ and $W+$jets events have to be taken into account. It has been shown in tab. 4.2 that the direct veto on electrons and muons can efficiently reduce the number of events with a single lepton and almost completely rejects the dileptonic background. The remaining background events have either an electron or muon that has not been identified for veto, or a hadronically decaying tau lepton. For both types one background estimation method each has been established. The "lost lepton method" (detailed in [46]) and the "hadronic tau method" (detailed in [47]) are summarized in sec. 5.1.

In the second case, an irreducible background arises from $Z \to \nu\bar{\nu} +$ jets events. In the course of the analysis multiple methods to estimate this background from data have been tested [48] [49]. These methods suffer from low statistics in the control regions. Therefore only the method which makes use of the similarity between Z boson and photons at high p_T is incorporated. Section 5.2 describes the $Z \to \nu\bar{\nu}$ prediction from γ+jets [50].

5.1 W and Top Quark Background

5.1.1 The Lost Lepton Background Estimation

This data-driven method developed in [46] is able to estimate the number of SM events with an electron or muon from a W boson which passes the explicit lepton veto requirement (sec. 4.3). These leptons are either not identified because they are out of the acceptance region of the detector or they do not fulfill the isolation conditions or they do not pass the ID requirements (quality cuts on μ and e).

The non-isolated (eq. 5.1) and non-identified (eq. 5.2) leptons are modeled using appropriately weighted data events from a control region containing $t\bar{t}$ and $W+$jets events. For this control region, the standard event selection is used except for the lepton veto which is replaced by requiring exactly one well identified and well isolated muon.

Due to lepton universality, the same isolated muon control sample can also be used for electrons (with a correction for the different efficiencies). For all further steps, the method is applied to electrons in an analogue way compared to the muons.

The control sample (CS) is weighted according to the lepton isolation efficiency in order to model the non-isolated (but identified) leptons (electron or muon separately) in the signal region (!ISO). For muons the calculation is:

$$!\text{ISO} = \text{CS} \cdot \frac{1-\epsilon_{\text{ISO}}}{\epsilon_{\text{ISO}}} \tag{5.1}$$

To model the sample containing not identified electron or muons in the signal region (!ID), the control-sample is weighted as follows:

$$!\text{ID} = \text{CS} \cdot \frac{1}{\epsilon_{\text{ISO}}} \cdot \frac{1-\epsilon_{\text{ID}}}{\epsilon_{\text{ID}}} \tag{5.2}$$

The lepton ID- and isolation-efficiencies must be sample independent, as their contribution is estimated on data Z-events using a tag&probe method, and is then applied to $t\bar{t}$ and $W+$jets events. Since the event topologies of these processes are different, the lepton isolation efficiencies are parametrized in transverse lepton momentum and in the angular distance ΔR between the lepton and the nearest jet. The lepton identification efficiency is parametrized in p_{T} and η. The remaining differences in the p_{T}- and the $|\eta|$-spectrum of signal- and control region have been studied and have been found to be smaller than 10%. This is included in the systematic uncertainty below.

The inefficiency due to events with leptons out of acceptance in p_{T} or η is calculated using Monte Carlo simulation. The not-accepted lepton background events are then derived with:

$$!\text{Acc} = \text{CS} \cdot \frac{1}{\epsilon_{\text{ISO}}} \cdot \frac{1}{\epsilon_{\text{ID}}} \cdot \frac{1-\epsilon_{\text{Acc}}}{\epsilon_{\text{Acc}}} \tag{5.3}$$

Systematic Uncertainties

The systematic uncertainties on the prediction are summarized in tab.6.4. The dominant uncertainties ($\sim$ 15 %) arise from the limited statistics of the muon

control sample and the Z-sample on which the lepton efficiencies have been determined.

Isolation & identification eff.	-13%	$+14\%$
Kinematic differences between W, $t\bar{t}$, Z samples	-10%	$+10\%$
SM background in μ control sample	-3%	$+0\%$
MC use for acceptance calculation	-5%	$+5\%$
Total systematic uncertainty	-17%	$+18\%$

Table 5.1: Systematic uncertainties for the prediction of the lost lepton background from the μ+jets control sample.

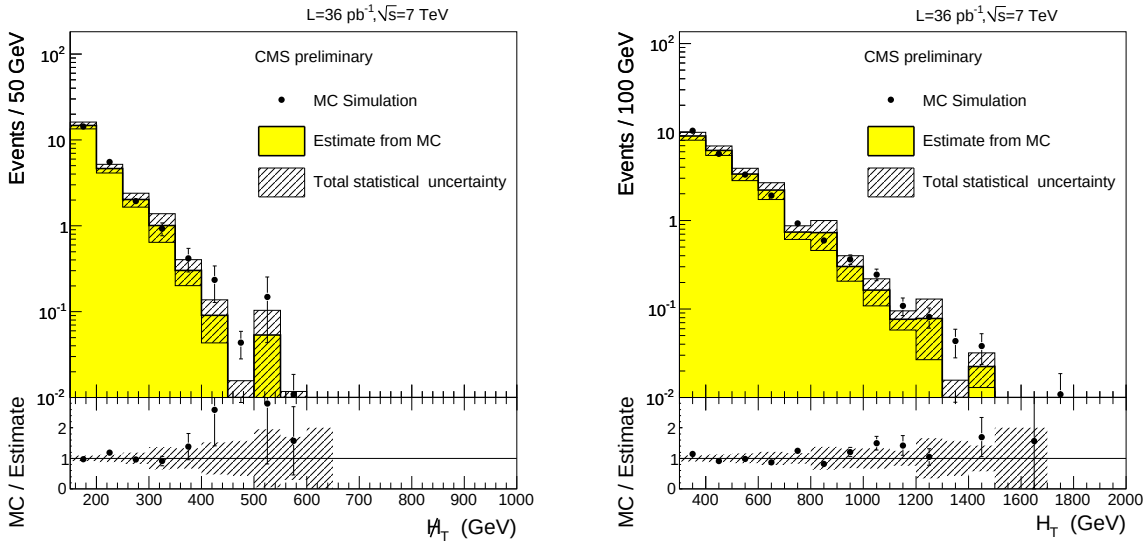
Closure Test and Resulting Prediction

A closure test is performed on Monte Carlo $t\bar{t}$ and $W+$jets simulation. The result of the comparison is shown in fig. 5.1. The estimate and the MC truth numbers agree within the expected uncertainties.

The method discussed above is applied on data corresponding to an integrated luminosity of $36\,\mathrm{pb}^{-1}$. The final prediction is shown in tab. 5.2 and compared to a prediction on MC events using the same data driven method, and to plain MC simulation.

	Baseline selection	High-$\not{H}_T$ selection	High-H_T selection
Estimate from data	$33.0 \pm 5.5\,^{+6.0}_{-5.7}$	$4.8 \pm 1.8\,^{+0.8}_{-0.6}$	$10.9 \pm 3.0\,^{+1.7}_{-1.7}$
Estimate from MC (PYTHIA)	$22.9 \pm 1.3\,^{+2.7}_{-2.6}$	$3.2 \pm 0.4\,^{+0.5}_{-0.5}$	$7.2 \pm 0.7\,^{+1.1}_{-1.1}$
MC expectation (PYTHIA)	23.6 ± 1.0	3.6 ± 0.3	7.8 ± 0.5
Estimate from MC (MADGRAPH)	$22.9 \pm 1.4\,^{+2.9}_{-2.8}$	$2.7 \pm 0.4\,^{+0.4}_{-0.4}$	$5.4 \pm 0.5\,^{+0.7}_{-0.6}$
MC expectation (MADGRAPH)	23.7 ± 0.8	3.4 ± 0.3	6.5 ± 0.5

Table 5.2: Estimates of the number of lost lepton background events from data and simulation for the baseline and search selections, with their statistical and systematic uncertainties.



(a) (b)

Figure 5.1: Closure test of the method prediction compared to Monte Carlo $t\bar{t}$ and $W+$jets simulation. The shown variables are: MHT (left), HT (right). All numbers are scaled to a luminosity of $100\,pb^{-1}$. From [45].

5.1.2 Hadronic τ Background Estimation

Electroweak tau lepton production with a hadronic decay ($W \to \tau_h\nu+$ jets, $t\bar{t} \to \tau_h\nu+$jets and $t\bar{t} \to \tau_h\nu + \tau_h\nu+$jets) constitutes an important background to the presented analysis. A method was developed [47] which is able to predict the hadronic tau background from a muon+jets control sample, mainly composed of $W \to \mu\nu+$ jets, $t\bar{t} \to \mu\nu+$jets and $t\bar{t} \to \mu\nu + \tau_h\nu+$jets processes.

The basic idea is to substitute the muon with a tau using a template which models the visible energy fraction of the tau jet. The muon p_T is smeared according to the template, which has been taken from the Monte Carlo simulation. Subsequently, the H_T, $\not{H}_T$ and other variables that use jets are recomputed for the event and the full event selection is applied.

The muon control sample is selected from data applying the following selection.

- The single muon triggers are required.
- At least 2 jets are required (jets as defined in sec. 4.3).

- Events are required to have exactly one isolated muon with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.1$.

- Events with an additional muon or electron are rejected

The multijet background and a possible contamination from physics beyond the SM in the muon control sample have been studied and found to be very small.

Tau jets are characterised by a low multiplicity of particles, typically a few pions and neutrinos. The $\not{H}_T$, H_T and jet composition of the muon and tau event types are similar except for the tau jet visible energy in the detector. To correct for this, the visible energy fraction template is applied to the measured muon p_T. In order to derive the template from Monte Carlo, reconstructed jets are matched in the $\eta - \phi$ plane ($\Delta R(\text{jet},\tau) < 0.1$) to generated tau leptons ($p_T >$20 and $\eta < 2.1$). For these matches, the fraction of visible energy (f_{VE}), defined as the ratio of the reconstructed tau jet energy and the simulated tau lepton p_T is computed. This energy is added along the direction of the muon to the measured energy depositions and the η and p_T dependent jet energy corrections (JES) are applied afterwards. A tau jet is accounted for if $f_{VE} \times \text{JES} \times p_T(\mu)$ is above the jet threshold of 30 GeV. H_T and $\not{H}_T$ are then computed starting from the resulting new jet collection.

For each event in the muon control sample the visible energy template is sampled 100 times to emulate tau+jets events. The statistical error associated to the prediction is studied with a set of 200 pseudo experiments and is of the order 20 % for the baseline selection and 30 % for the evolved search regions.

Systematic Uncertainties

All considered systematic uncertainties and their impact on the prediction using the 2010 data sample, corresponding to 36 pb^{-1}, are summarised in Tab. 5.3.

Resulting Prediction

In tab. 5.4 the number of predicted W/ $t\bar{t} \to \tau_{had}$ is shown for the different signal regions considered in this analysis.

	Baseline selection	High-$\not{H}_T$ selection	High-H_T selection
τ response distribution	2%	2%	2%
Acceptance	+6%/−5%	+6%/−5%	+6%/−5%
Muon efficiency in data	1%	1%	1%
SM backgr. subtraction	5%	5%	5%

Table 5.3: Systematic uncertainties for the hadronic-τ background prediction from the μ+jets control sample for the baseline and search selections.

	Baseline selection	High-$\not{H}_T$ selection	High-H_T selection
$W/t\bar{t} \to \tau_h$ estimate	$22.3 \pm 4.0 \pm 2.2$	$6.7 \pm 2.1 \pm 0.5$	$8.5 \pm 2.5 \pm 0.7$
$W/t\bar{t} \to \tau_h$ MC	19.9 ± 0.9	3.0 ± 0.4	5.5 ± 0.5

Table 5.4: Predicted number of hadronic-τ background events from data and simulation for the baseline and search selections, with their statistical and systematic uncertainties.

5.2 Invisible Z Background Estimation

5.2.1 Estimation of $Z \to \nu\bar{\nu}$ Background from γ + jets

This method estimates the background that arises from Z bosons decaying into two neutrinos which can not be measured with the detector (also referred to as invisible Z events). Only high-p_T Z+jets events produce enough $\not{H}_T$ to pass the event selection. These events have kinematical similarities to γ+jets events which can therefore be used as substitutes for the measurement.

The major steps of the presented method, the selection of the γ+jets control sample and the corrections that have to be applied are summarized in the following.

Selection of the Photon Control Sample

A highly pure γ+jets control sample is needed for the prediction of the $Z \to \nu\bar{\nu}$ background. To reach this goal, standard cleaning criteria and residual ECAL spike cleaning [51] are applied, and photon candidates are selected with a pre-selection cut of $E_T > 100\,\text{GeV}$. A veto on the presence of a pixel

seed removes photons which are part of an electron shower. Next, prompt photon candidates are selected by requiring tracker and calorimeter isolation requirements (defined in [51]), combined with a cut on the shower shape variable in the η coordinate as introduced in [52].

The data is selected using single-photon triggers, with transverse energy thresholds increasing during the run up to a maximum of 70 GeV. Well above this threshold the trigger has been measured to be quasi 100% efficient [53].

Corrections and Systematic Uncertainties

In tab. 5.5 the full list of corrections is summarized for the baseline and search selections, along with the corresponding systematic uncertainties.

		Baseline selection		High-$\not{H}_T$ selection		High-H_T selection	
Z/γ correction	$\pm$theory	0.41	$\pm$6 %	0.48	$\pm$6 %	0.44	$\pm$4 %
	$\pm$acceptance		$\pm$5 %		$\pm$5 %		$\pm$5 %
	$\pm$MC stat.		$\pm$7 %		$\pm$13 %		$\pm$13 %
Fragmentation		0.95	$\pm$1 %	0.95	$\pm$1 %	0.95	$\pm$1 %
Secondary photons		0.94	$\pm$9 %	0.97	$\pm$10 %	0.90	$\pm$9 %
Photon mistag		1.00	$\pm$1 %	1.00	$\pm$1 %	1.00	$\pm$1 %
Photon identification and isolation efficiency		1.01	$\pm$2 %	1.01	$\pm$2 %	1.01	$\pm$2 %
Total correction		0.37	$\pm$14 %	0.45	$\pm$18 %	0.38	$\pm$17 %

Table 5.5: Overview of all correction factors and corresponding systematic uncertainties for the prediction of the $Z \to \nu\bar{\nu}$+jets background from the γ+jets control sample for each of the selections.

Prediction of the $Z \to \nu\bar{\nu}$+jets background

The prediction for the $Z \to \nu\bar{\nu}$+jets from the γ+jets data control sample are summarized in tab. 5.6 and is found to be in agreement with the plain Monte Carlo simulation.

At this point, the electroweak and top background has been estimated by combing the above results of the three data-driven methods. The dominant source of uncertainty for all three methods and both evolved selections arises from statistics, which is promising for succeeding analyses with more data available.

	Baseline selection	High-$\not{H}_T$ selection	High-H_T selection
γ+jets data sample	72	16	22
$Z \to \nu\bar{\nu}$ estimate	$26.3 \pm 3.2 \pm 3.6$	$7.1 \pm 1.8 \pm 1.3$	$8.4 \pm 1.8 \pm 1.4$
$Z \to \nu\bar{\nu}$ MC	21.1 ± 1.4	6.3 ± 0.8	5.7 ± 0.7

Table 5.6: Number of γ+jets events in the data and the resulting estimate of the $Z \to \nu\bar{\nu}$+jets background, as well as the prediction from the MC simulation, for each of the selections, with their statistical and systematic uncertainties. The estimate from data is obtained by multiplying the number of events in the γ+jets sample with the total correction factor from Table 5.5.

6 Data-driven QCD Background Estimation

QCD multijet production is the most difficult background to model for new-physics searches in the all-hadronic channels. Current theoretical knowledge of the underlying "true" spectrum of particle jets has large uncertainties, especially at high $\not{H}_T$ and high jet multiplicity. Given the complexities of QCD multijet events and the importance of modeling this background well, two data-driven methods have been pursued to estimate the multijet contamination for this analysis.

In the presented thesis, the focus lies on the development of the factorization method (sec. 6.1) and its verification and application in the $36\,\mathrm{pb}^{-1}$ of 2010 data (sec. 6.2). A short overview of the rebalance and smear (R&S) method will be given and concluded with a comparison of the two methods (sec. 6.3).

6.1 The Factorization Method

Data-driven estimations of backgrounds are vital for searches which make use of variables that have large uncertainties in the Monte Carlo simulation. While the signal region for this search can be defined by the application of a few subsequent cuts on discriminating variables, control regions, which are ideally signal free, can be used to measure some of these variables in background events. The use of the factorization method makes it possible to trade the large uncertainty in the Monte Carlo simulation for the smaller uncertainty on the measurement of a ratio in control regions and its extrapolation to a signal region.

6.1.1 A simple Idea how to estimate QCD from Data

In order to reach a sufficient separation of expected signal events from QCD, special variables have been designed that efficiently reduce the huge number of QCD events during the selection (sec. 4.3). The high QCD cross section together with the good discrimination power, makes it possible to directly

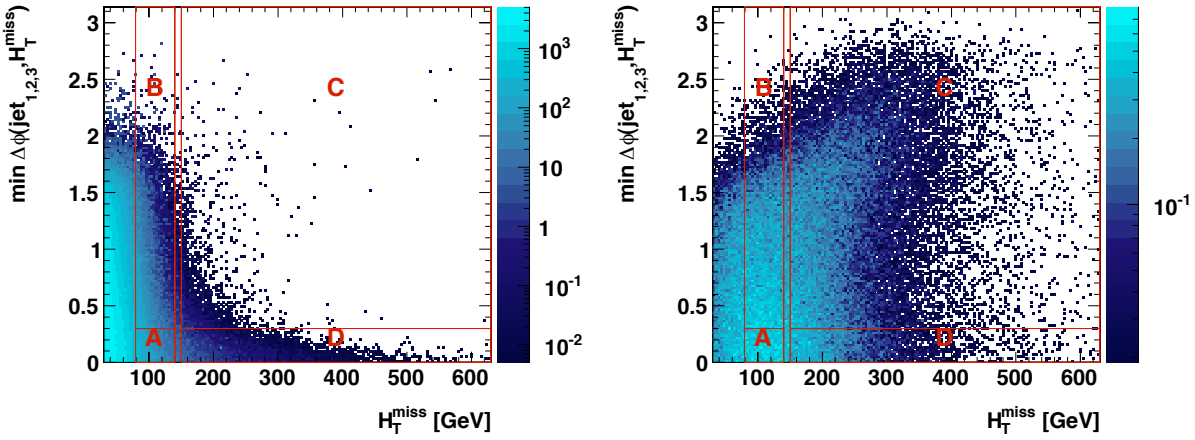


Figure 6.1: Distribution of $\Delta\phi_{\text{min}}$ vs. $\not{H}_{\text{T}}$ for QCD events (left) and a typical Susy sample (CMS benchmark point LMo - right). The different relevant regions in the $\not{H}_{\text{T}}$-$\Delta\phi_{\text{min}}$ plane are marked with capital letters: (A/B) fit region of $r(\not{H}_{\text{T}})$; (C) signal region; (D) control region, to which the extrapolated ratio is applied as weight.

measure QCD distributions with low signal contamination in data by simply inverting the cuts on these variables.

By using the two best discriminating variables, three QCD dominated control regions are obtained, which are illustrated in fig. 6.1 for the variables $\not{H}_{\text{T}}$ and $\Delta\phi_{\text{min}}$. These plots show the discriminative power of the two variables.

The basic idea of the factorization method is to make use of the ratio $r(\text{var}_1)$ of the two distributions that are measured by once applying the cut on var_2 and once inverting it. Whenever it is possible to predict the functional form of $r(\text{var}_1)$ and measure the parameters in the region (inversion of cut_1: $!\text{cut}_1$) with a sufficient precision, this function $r(\text{var}_1)$ can be extrapolated to the region (cut_1) and used to calculate the weights of events with cut_1 & $!\text{cut}_2$ to estimate the total number of events in the signal region $(\text{cut}_1$ & $\text{cut}_2)$:

$$N(\text{cut}_1 \;\&\; \text{cut}_2) = \sum_{N(\text{cut}_1 \;\&\; !\text{cut}_2)} r(\text{var}_1) \tag{6.1}$$

Equation 6.1 has evolved from the trivial case where the two variables (var_1

and var_2) are uncorrelated after all cuts. Then the number of events in the signal region is given by the well known formula of the "ABCD method":

$$N(\text{cut}_1 \,\&\, \text{cut}_2) = N(\text{cut}_1 \,\&\, !\text{cut}_2) \cdot \frac{N(!\text{cut}_1 \,\&\, \text{cut}_2)}{N(!\text{cut}_1 \,\&\, !\text{cut}_2)} \tag{6.2}$$

The goal of the here presented factorization method is to use this concept by establishing a functional form $r(\not{H}_T)$ for the ratio of events with low $\Delta\phi_{\text{min}}$ against large $\Delta\phi_{\text{min}}$. This can be achieved by using characteristics in the topology of QCD events that are required to have non-vanishing $\not{H}_T$.

6.1.2 Topology of QCD Events in hadronic SUSY Searches

The key signature of SUSY searches is the presence of missing transverse energy or momentum. On the other hand, QCD events have no intrinsic $\not{H}_T$.

The main source of multijet QCD events in the signal region of large $\not{H}_T$ and $\Delta\phi_{\text{min}}$ are non Gaussian fluctuations of the jet response. The origins of such fluctuations are

- Electroweak decays of heavy quarks: The ν and μ component of a jet deposits no or only a small amount of energy in the calorimeter. On average the jet energy corrections account for this, but single jets can be measured significantly too low.

- Punch through of very high energetic jets: In the barrel region of the detector the thickness of the hadronic calorimeter is about five interaction length λ. It is possible that the energy deposition of a jet is not completely contained in the hadronic calorimeter, but also in the coil, the outer hadronic calorimeter (HO), or the muon system. Such jets could be identified and maybe even corrected by using signals in the HO and/or the muon systems. However, in early data this effect is expected to be small compared to the effect by heavy flavor jets. With increasing statistics at very high energies this effect will become more important.

- Dead electromagnetic calorimeter cells: Although these cells are identified, it is important to study the influence on the jet response, since a rejection of all events with jets containing dead cells will result in a low efficiency. This effect should be covered on average by the η binning of the jet response correction.

- Cracks in the calorimeter: Similar to the previous effect this should be covered on average by the η binning of the jet response correction.

- Hot cells: This effect may vary on run-by-run basis.

The cut on $\not{H}_T$ positively selects those QCD events that have at least one severely mismeasured jet. The mismeasurements of jets is also responsible for the value of $\Delta\phi_{\text{min}}$. By construction $\not{H}_T$ and $\Delta\phi_{\text{min}}$ are correlated for QCD events since large $\not{H}_T$ values require heavily mismeasured jets and therefore relatively small values of $\Delta\phi_{\text{min}}$. The effects of the correlation are visualized in distributions of the two variables in slices of the other (fig. 6.2).

In the following, three different types of jet mismeasurement configurations in QCD events are considered, which cannot be separated in data in a straight forward way. The list below gives the definitions and cuts which are used to classify the types in simulated QCD events as shown in fig. 6.3:

type 1: The most mismeasured jet is reconstructed with too high energy. In the simulation, this is defined as events with at least one of the leading three jets with $p_T^{measured} - p_T^{true} > 50\,\text{GeV}$.

type 2: The most mismeasured jet is reconstructed with too low energy but still belongs to one of the leading three jets. The requirement is fulfilled if for at least one of the leading three jets $p_T^{measured} - p_T^{true} < -50\,\text{GeV}$ and the event is not of type 1.

type 3: The most mismeasured jet is not reconstructed as one of the leading three jets. The three leading jets are required to be within $|p_T^{measured} - p_T^{true}| < 50\,\text{GeV}$ which is complementary to the first two types and assumes that the $\not{H}_T$ originates from another jet mismeasurement.

In fig. 6.3 the ratio $r(\not{H}_T)$ of events with low $\Delta\phi_{\text{min}}$ against large $\Delta\phi_{\text{min}}$ for the three types is visualized by using generator jets and smearing them with Monte Carlo jet response histograms. The full smearing procedure is described in sec. 6.1.4 where it is used for the closure test of the method. Since we are comparing ratios in fig. 6.3 the relative sizes of the contributions are not directly visible (only the statistical error bars give a hint). The fraction of events of type 1 and 2 are of the same order while the contribution of type 3 events is about a few % which has only a small dependence on $\not{H}_T$. Nevertheless, these events with "lost leading jets" are clearly outliers in a coherent description of the correlation between $\Delta\phi_{\text{min}}$ and $\not{H}_T$. This effect will be discussed later on in this section. The first two types can be effectively approximated by a model which is now introduced as "Gaussian resolution model".

First, consider events which are perfectly measured except for one of the leading three jets where the measured p_T fluctuates to lower values. Here, the

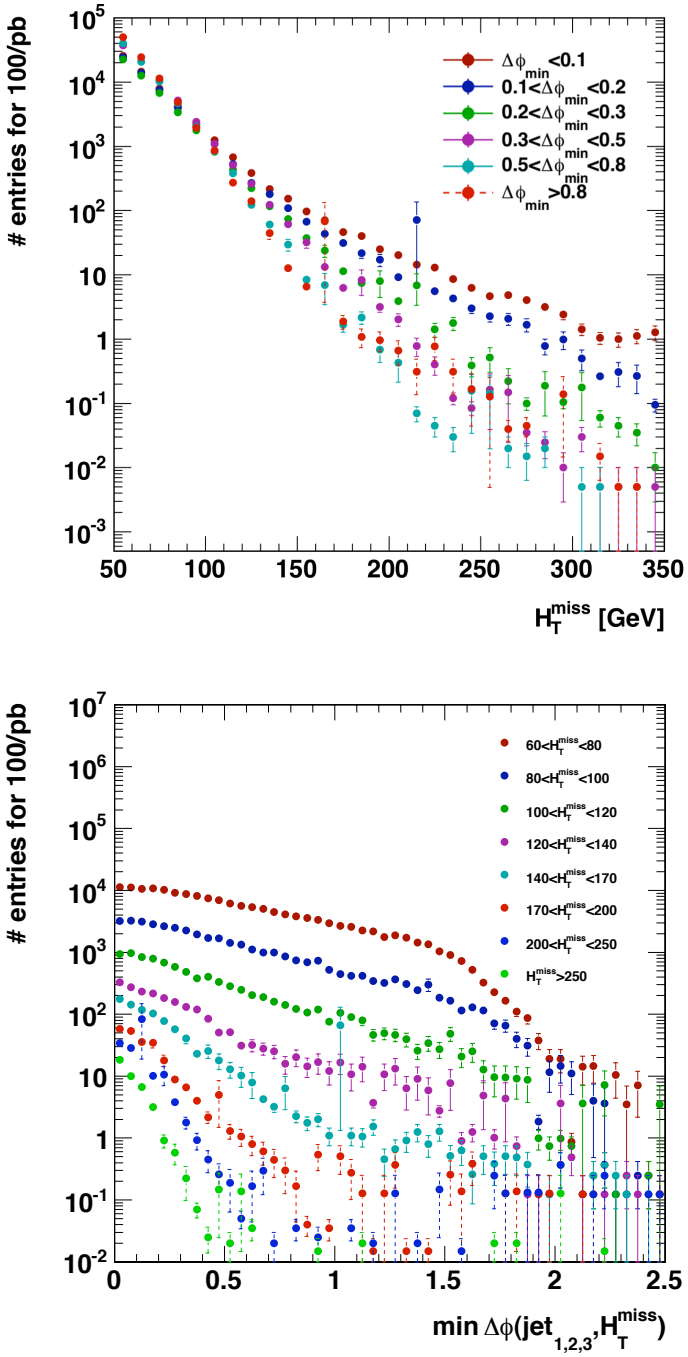


Figure 6.2: Distribution of $\not{H}_T$ in slices of $\Delta\phi_{\min}$ (left) and distribution of $\Delta\phi_{\min}$ in slices of $\not{H}_T$ for QCD events (right). All other cuts of the standard selection have been applied (see sec. 4.3).

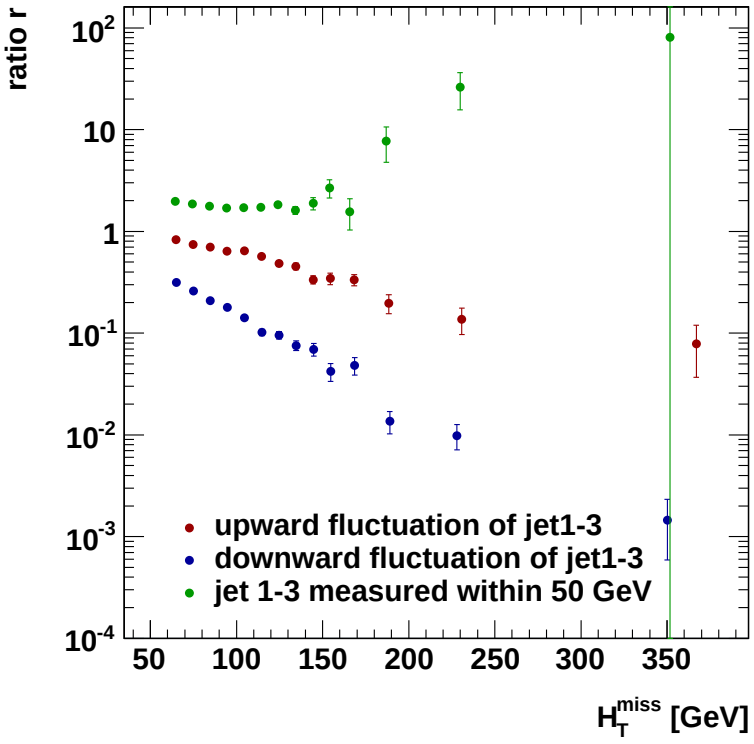


Figure 6.3: The ratio $r(\not{H}_T)$ for three different types of jet mismeasurement configurations as explained in the text. The events are categorized by taking jet $p_T^{measured}$ - jet p_T^{true} for the leading three jets. The jet and H_T cuts of the standard event selection have been used. Note the variable bin width and that the points are plotted at the center of mass in each bin.

direction of the mismeasured jet and $\not{H}_T$ would be identical and $\Delta\phi_{\min} = 0$ which means that the cut on $\Delta\phi_{\min}$ is 100% efficient in suppressing QCD.

Now, allowing all the other jets to have small fluctuations, that can be described by Gaussian resolutions, as a result also $\Delta\phi_{\min}$ is smeared in the approximation of small angles with a Gaussian distribution around zero (fig. 6.4, left). The σ of the Gaussian resolution is a function of $\not{H}_T$ since larger $\not{H}_T$ leads to less influence from the fluctuations of the other jets in the described model.

This relation is shown in the range $60 < \not{H}_T < 250\,\text{GeV}$where the model is believed to have dominant influence and it is found that it can be approximately described by a falling exponential function (fig. 6.4, right).

The Gaussian resolution model for $\Delta\phi_{\min}$ provides a functional form for r that only depends on σ_{Gauss} and cut_1:

$$\tilde{r}(\sigma_{Gauss}) = \frac{1}{erf(\frac{\text{cut}_1}{\sqrt{2}*\sigma_{Gauss}})} - 1 \tag{6.3}$$

where erf is the error function.

The idealized model breaks down in the region where $\not{H}_T$ is very low ($\lesssim 60\,\text{GeV}$) and the direction of the $\not{H}_T$ vector is influenced by many jets and generally not aligned with the direction of one of the leading three jets. Nevertheless, due to the construction of the variable as the minimum of the three $\Delta\phi$'s, smaller values are preferred and high values are very unlikely since the event topology of QCD forbids that all leading jets point in the same direction. This property of the $\Delta\phi_{\min}$ distribution can be seen in the sharp bend around $\pi/2$ in the first $\not{H}_T$ slice (60-80 GeV) of fig. 6.2. At higher $\not{H}_T$ values, this effect becomes negligible due to the smaller width of the $\Delta\phi_{\min}$ distribution.

The Gaussian resolution model can also be used for upward fluctuations in the p_T of the most mismeasured jet, since this causes $\not{H}_T$ in the opposite direction where in QCD very often one of the other two leading jets could be found in close vicinity. Compared to the downward fluctuation, this scenario produces a smeared out resolution of $\Delta\phi_{\min}$ which alters the slope in the model. Figure 6.3 shows that both types are monotone falling but not with the same slope. The unknown mixture of the two types of fluctuations introduces a large uncertainty. The sizes of the contributions are not the same since the jet response functions are not perfectly symmetric. This is caused by electroweak decays of heavy quarks and other effects discussed at the beginning of this section. In this method, this uncertainty will be controlled by using two types of functional forms (see sec. 6.1.3).

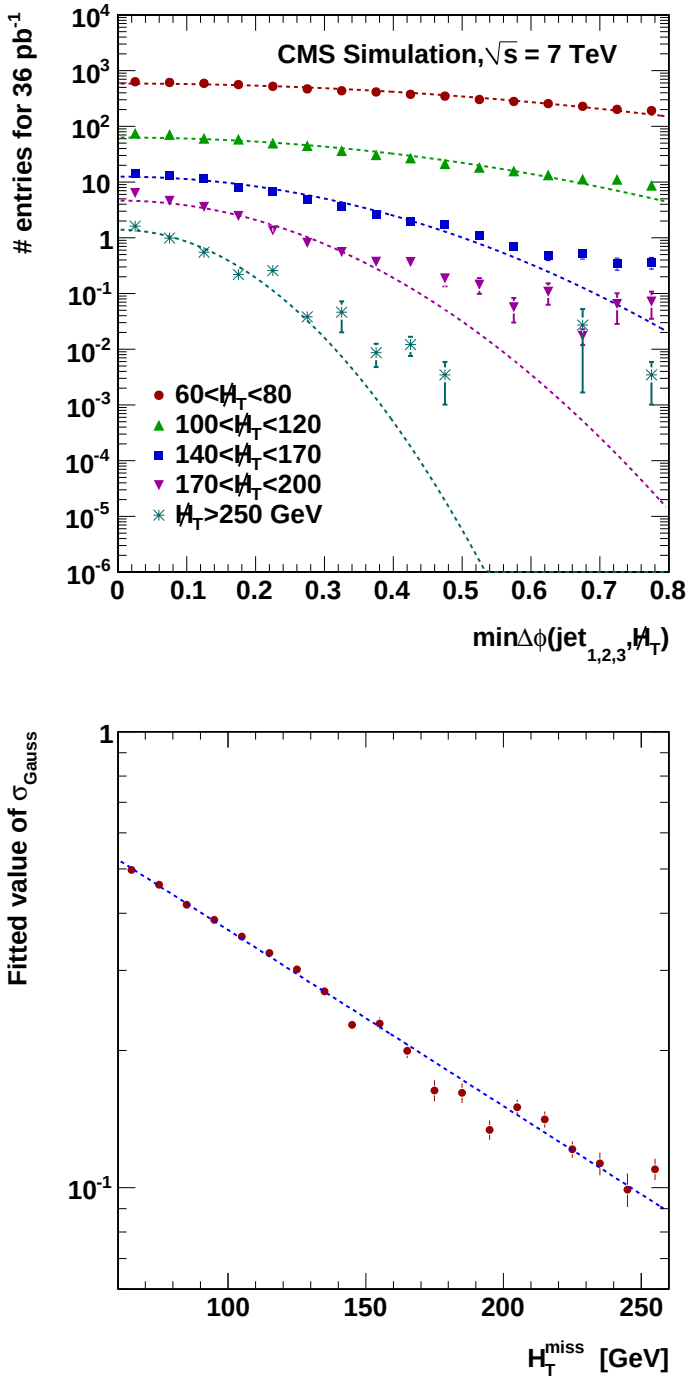


Figure 6.4: Distribution of $\Delta\phi_{\text{min}}$ in slices of $\not{H}_{\text{T}}$ for QCD events (MC simulation with MadGraph). Gaussian fits with a fixed mean $= 0$ are applied. Lower: σ width of the Gaussian fits of $\not{H}_{\text{T}}$ slices in steps of 10 GeV.

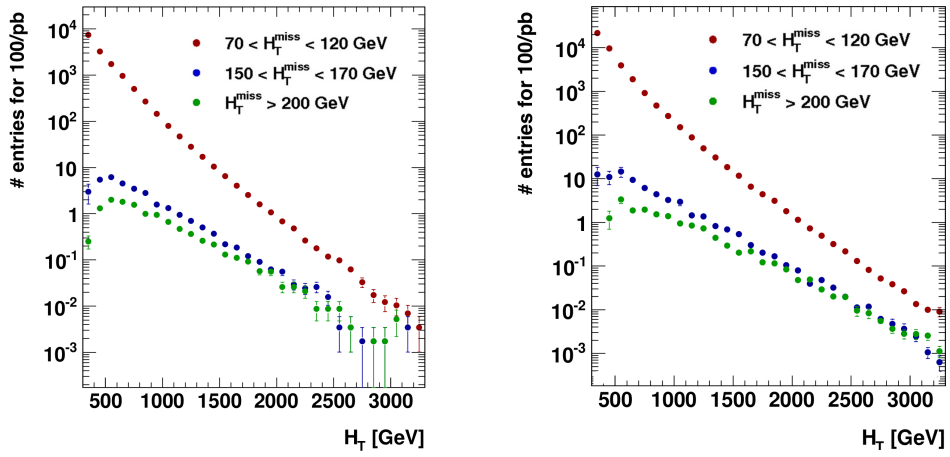


Figure 6.5: Distribution of H_T in three slices of $\not{H}_T$ for QCD generated with MadGraph (a) and pythia (b). The y-axis shows number of entries for $100\,\mathrm{pb}^{-1}$.

Events in the vicinity of the jet cuts and the H_T cut have different probabilities for the two types to pass the selection. This is because upward fluctuations can promote low-energy events to the control region by letting them pass the jet selection and the H_T cut, the downward fluctuations can not. As a result the contribution of type 1 events is more pronounced in low-energy events. Furthermore, there is a strong correlation of H_T and $\not{H}_T$, and low H_T values, which can be seen in fig. 6.5. The size of this effect can be reduced by introducing a higher cut on H_T.

Since the method depends on a monotone falling behavior of the ratio $r(\not{H}_T)$ the type of events where the most mismeasured jet is not reconstructed as one of the leading three jets (green points in fig. 6.3) is critical for its operation. A fraction of these events has the original jet ranking in p_T from the generator jets, but in the considered kinematic region of the method these events are greatly outnumbered by events with a fluctuation of one of the leading three jets. Important are only the extreme tails in the jet response causing the mismeasured jet not to be reconstructed as one of the leading three jets. Since the $\Delta\phi_{\mathrm{min}}$ is not sensitive to this kind of jet mismeasurement, these events will appear signal-like in the ratio $r(\not{H}_T)$. While it is clear that for a wide range

in $\not{H}_T$ this type is insignificant compared to the first two, there remains a big uncertainty for the high $\not{H}_T$ case ($\not{H}_T \gtrsim 250\,\text{GeV}$). Since at $\not{H}_T \gtrsim 250\,\text{GeV}$ the Gaussian part of the $\Delta\phi_{\text{min}}$ is completely outside the signal region (fig. 6.2), the efficiency of suppressing QCD flattens out.

If the calorimeter noise level can be kept under control, only high-H_T events can reach high $\not{H}_T$. As long as there are no large effects from punch trough visible, there is no apparent reason why QCD events of the third type should become more likely with increasing energy. Then, it is save to assume that the ratio $r(\not{H}_T)$ is reaching a constant value in the limit of very high $\not{H}_T$. Though the size of this effect with respect to the resulting event yield of this factorization method is only significant in a signal region with $\not{H}_T > 200\,\text{GeV}$ where QCD is only a minor background, an additional constant term in the ratio r($\not{H}_T$) determined using Monte Carlo simulation is considered (eq. 6.4).

$$r(\not{H}_T) = \tilde{r}(\not{H}_T) + c \tag{6.4}$$

6.1.3 The full Method and its technical Application

The above described Gaussian resolution model treats the decrease in the Gaussian width of $\Delta\phi_{\text{min}}$ with increasing $\not{H}_T$ in an idealized manner (exponential dependency as shown in fig. 6.4). The mixture of different types of jet mismeasurement configurations smears the resolution resulting in higher $\Delta\phi_{\text{min}}$ values. Therefore, the Gaussian resolution model (eq. 6.5) yields a safe lower boundary on the ratio, while, on the other hand, it would be very difficult to correct appropriately for these effects.

$$r(\not{H}_T) = \frac{1}{erf(\frac{1}{a\cdot\exp(-b\cdot\not{H}_T)})} - 1 + c \tag{6.5}$$

The description of the ratio is completed by a second functional form that is used as the upper boundary of the model. If we assume that the resolution of $\Delta\phi_{\text{min}}$ is not more than marginally improved in the region of interest in $\not{H}_T$, which means that the argument of the error function in eq. 6.5 becomes small, then this error function can be approximated by a linear function resulting in a plain exponential fit of the ratio r:

$$r'(\not{H}_T) = a' \cdot \exp(-b' \cdot \not{H}_T) + c' \,, \tag{6.6}$$

with three transformed parameters a', b' and c'. This functional form for the upper boundary takes into account that the effects on the resolution of $\Delta\phi_{\text{min}}$ described by the Gaussian resolution model might be superposed by

the effects of jet mismeasurement configurations that slow the improvement in the resolution down.

For both functional forms (eq. 6.5 and eq. 6.6) two free parameters (a, b) are used in the fit and a third (c) is fixed at the value of the ratio r at very high $\not{H}_T$ ($> 300\,\text{GeV}$) found in simulated QCD events. This value corresponds to the fraction of type 3 events discussed in sec. 6.1.2 and is between 1-3 %. The introduction of parameter c has a negligible impact on the fit results but corrects the ratio at large $\not{H}_T$.

In order to reach a closure for the factorization method all reasonable variations of QCD samples have to be investigated. This program together with a robustness check against all cut variations is accomplished in sec. 6.1.4. First, an overview of the basic steps in the application of the factorization method is given.

Extrapolation and Uncertainties

The events at large $\not{H}_T$ and small $\Delta\phi_{\text{min}}$ (region D as shown in fig. 6.1) are used in order to model the events in the signal region at large $\not{H}_T$ and large $\Delta\phi_{\text{min}}$ (region C) by weighting them according to the extrapolation function r. In addition to the uncertainties originating from the choice of the parameterization of the fit function there are purely statistical uncertainties on the background estimate from the extrapolation and the statistics in the control region. These are calculated from the variance of the fit function and the statistical uncertainty on the number of events in the control region D.

For the fit region an adapted selection compared to the event seclection in sec.4.3 had to be used:

- The lower boundary in $\not{H}_T$ of the fit region is $x_{min} = 70\,\text{GeV}$.
- The upper boundary in $\not{H}_T$ of the fit region is set to $x_{max} = 120\,\text{GeV}$ in order to avoid a significant number of other SM background and possibly signal events. The high QCD cross section at low $\not{H}_T$ values is a natural protection against such contaminations. The possible remaining contamination is considered as systematic uncertainty.
- The upper boundary in $\Delta\phi_{\text{min}}$ is set to $y_{max} = 0.2$, also in order to minimize the contamination from other SM background events in region D.
- An additional cut on $H_T > 600\,\text{GeV}$ is applied which is discussed subsequently.

- The fit region is divided into 10 bins (only 5 in data due to the lower statistics) in $\not{H}_T$ and the bin center is defined as the mean of entries.

All the definitions above are set such, that they are robust against reasonable variations (see sec. 6.1.4). Figure 6.6 shows that the two functional forms for the description of the ratio $r(\not{H}_T)$ bracket the different Monte Carlo QCD samples.

For the presented analysis, it has been decided on a relatively low H_T cut of 300 GeV for the baseline selection. As discussed in sec.6.1.2 this introduces a bias for the fit in the low $\not{H}_T$ region. The proposed solution is an additional cut on H_T that reduces this effect and gives both fit and signal region a similar composition of jet fluctuations.

Figure 6.7 shows the results of the factorization method with additional cuts on H_T. Instead of using a higher global cut on H_T, only the H_T cut for the fit region is increased in steps of 100 GeV while for the signal region the baseline selection is used. For high H_T the bias of the estimation is removed for both PYTHIA and MADGRAPH QCD. Due to the limited statistics in the ~ 36 pb^{-1} data sample the choice of the H_T cut for the fit region is restricted to ~ 600 GeV (details in sec. 6.2.1).

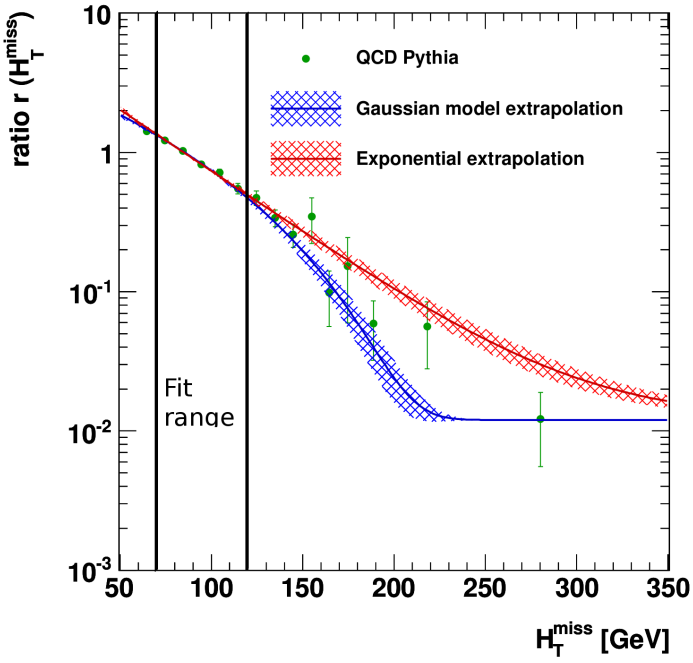
Later, for the application to data (sec. 6.2) three additional sources of systematic uncertainties will be considered for both chosen functional forms individually:

- The constant term in the functional form of the ratio $r(\not{H}_T)$
- The resolution effects from the additional cut which has to use measured H_T
- SM background and signal contamination

6.1.4 Closure and Robustness Check

A robustness check in order to verify that the chosen default fit scenario (sec. 6.1.3) produces stable results for the factorization method and a closure test of the method are presented, both using Monte Carlo simulation.

Since the factorization methods depends on some general assumptions on the jet energy fluctuations that influence the $\Delta\phi_{\text{min}}$ distributions (discussed in sec. 6.1.2), the method has to be tested with variations within assumed uncertainties of these fluctuations, especially considering the non Gaussian fluctuations of the jet response. A procedure of modeling these fluctuations within appropriate uncertainties by constructing pseudo simulated QCD samples is described here.



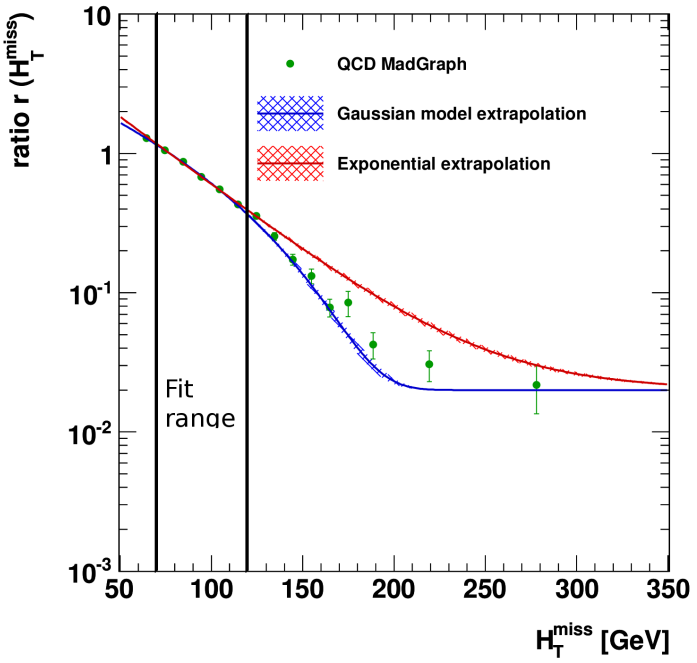


Figure 6.6: The extrapolation of the two chosen models to the signal region of ratio r for two QCD samples. The fit has been performed in the region ($70\,\text{GeV} < \not{H}_T < 120\,\text{GeV}$). The extrapolation of the fit function and error bands for the fit error propagation from the covariance matrix are shown.

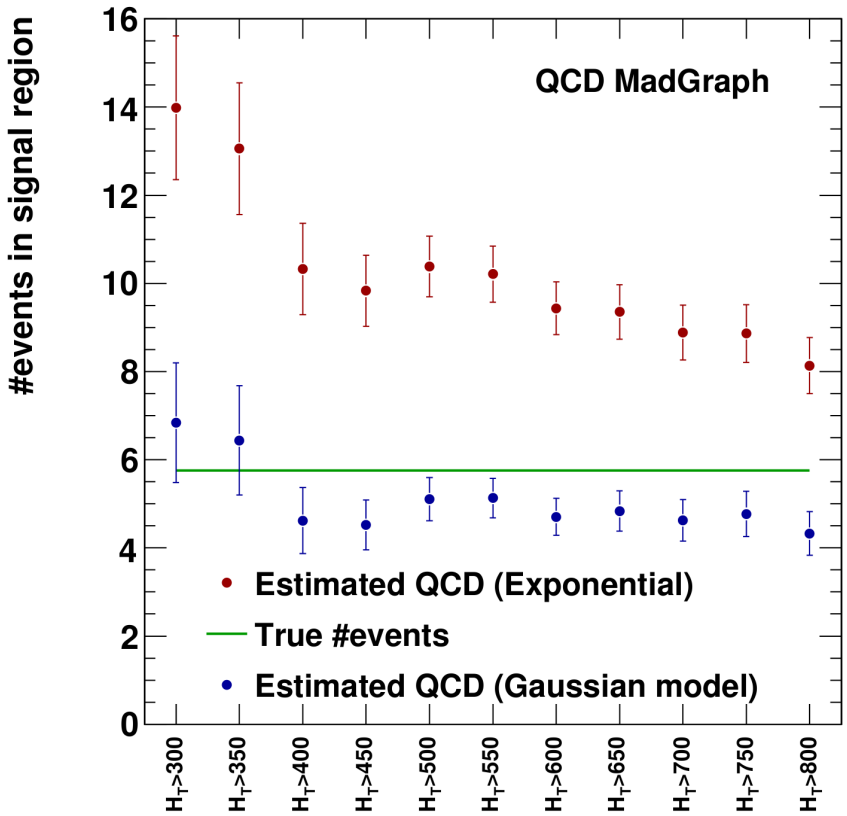


Figure 6.7: Number of estimated QCD events in the signal region with a MadGraph QCD simulation using different cuts on H_T in the fit region from 300 GeV - 800 GeV (shown as labels of the x-axis). The statistical errors arise from the error propagation from the fit covariance matrix and the limited statistics in control region D.

Robustness Check

The procedure of the factorization method is carried out multiple times for each of the two extrapolation models. Each time one parameter of the default scenario is varied. The results are summarized in tab. 6.1 and visualized in fig. 6.8. For QCD MadGraph the largest deviation from the default scenario is within 15% for both models while the statistical uncertainty is $\sim$ 10%. This demonstrates, that the extrapolations from both models are robust against reasonable changes of the fit scenario.

Creation of pseudo-simulated QCD Samples for a Closure of the Method

The sources of non-Gaussian fluctuations of the jet response have been discussed in sec. 6.1.2. Uncertainties in the flavor composition of the jets in multijet QCD events as well as punch through effects may be modeled by a variation of the non-Gaussian tails in the jet response. These tails can be estimated from Monte Carlo by subtraction of the Gaussian part which is determined by a fit around the mean of the distribution within a range of three RMS. The non-Gaussian tail is then added to the remaining core of the distribution with an appropriate chosen scaling factor (see fig. 6.9). Since this procedure would also scale statistical uncertainties in the core of the distribution, each bin of the tail is weighted with a pre-factor containing the Gaussian distribution

$$f_0 = 1 - e^{\frac{1}{2}\left(\frac{\text{Mean}-x}{\text{RMS}}\right)^2}$$

for response values x smaller than the mean and $f_0 = 0$ otherwise.

A variation of the non Gaussian part with the scaling factors $f = 5 \cdot f_0$ and $f = 0.2 \cdot f_0$ is performed. These scaling factors represent the maximal data to Monte Carlo simulation discrepancy that has been found in a measurement of the jet p_T response in QCD dijet events [54].

The resulting responses are used to smear generator jets to obtain a pseudo simulated QCD sample. For this purpose the QCD pythia sample is used. Statistical uncertainties are kept small by smearing the generator jets of each QCD event 5 times.

The resulting $\not{H}_T$ distributions scaled to 100 pb^{-1} are shown in fig. 6.10. The distribution obtained with the unmodified responses is in good agreement with the full detector simulation. By increasing the tails the $\not{H}_T$ distribution is shifted to higher values, and downscaling of the tails leads to lower $\not{H}_T$ values.

Since it has also been found that the jet resolution is generally worse in data compared to Monte Carlo simulation, a further pseudo simulated QCD

Variation	χ^2/ d.o.f.	# estimated QCD	estimate/truth
MC truth = 5.9 events			
Exponential extrapolation:			
$Default$	16/8	9.4 ± 0.6	1.6 ± 0.14
$x_{min} = 60\,\text{GeV}$	18/8	10 ± 0.6	1.8 ± 0.14
$x_{min} = 80\,\text{GeV}$	10/8	8.4 ± 0.6	1.5 ± 0.14
$x_{max} = 110\,\text{GeV}$	19/8	10 ± 0.7	1.8 ± 0.16
$x_{max} = 130\,\text{GeV}$	16/8	9.7 ± 0.6	1.7 ± 0.14
$N_{bins} = 5$	3/3	9.5 ± 0.6	1.6 ± 0.14
$N_{bins} = 20$	28/18	9.4 ± 0.6	1.6 ± 0.14
$x_{min}\&x_{max} - 10\%$	29/8	10 ± 0.6	1.7 ± 0.15
$x_{min}\&x_{max} + 10\%$	12/8	8.6 ± 0.6	1.5 ± 0.13
$y_{max} = 0.15$	19/8	9.6 ± 0.6	1.7 ± 0.15
$y_{max} = 0.25$	17/8	9.5 ± 0.6	1.6 ± 0.14
$H_T(Fit) - 10\%$	16/8	9.4 ± 0.6	1.6 ± 0.14
$H_T(Fit) + 10\%$	10/8	9.1 ± 0.6	1.6 ± 0.14
Gaussian model:			
$Default$	13/8	4.7 ± 0.4	0.82 ± 0.087
$x_{min} = 60\,\text{GeV}$	9/8	4.7 ± 0.4	0.82 ± 0.079
$x_{min} = 80\,\text{GeV}$	11/8	4.4 ± 0.5	0.77 ± 0.097
$x_{max} = 110\,\text{GeV}$	15/8	4.8 ± 0.5	0.84 ± 0.1
$x_{max} = 130\,\text{GeV}$	11/8	5.3 ± 0.4	0.92 ± 0.092
$N_{bins} = 5$	2/3	4.7 ± 0.4	0.82 ± 0.091
$N_{bins} = 20$	22/18	4.7 ± 0.4	0.81 ± 0.086
$x_{min}\&x_{max} - 10\%$	25/8	4.4 ± 0.4	0.76 ± 0.083
$x_{min}\&x_{max} + 10\%$	14/8	4.9 ± 0.4	0.85 ± 0.09
$y_{max} = 0.15$	13/8	4.5 ± 0.4	0.79 ± 0.087
$y_{max} = 0.25$	13/8	4.8 ± 0.4	0.83 ± 0.088
$H_T(Fit) - 10\%$	13/8	4.7 ± 0.4	0.82 ± 0.087
$H_T(Fit) + 10\%$	11/8	4.7 ± 0.4	0.81 ± 0.091

Table 6.1: Robustness check for QCD MadGraph. All boundaries of the control regions have been varied independently as well as simultaneously for the fit region in $\not{H}_T$. The default scenario is described in sec. 6.1.3. The χ^2/ d.o.f. values denote the quality of the fit. The number of estimated events corresponds to an integrated luminosity of $36\,\text{pb}^{-1}$.

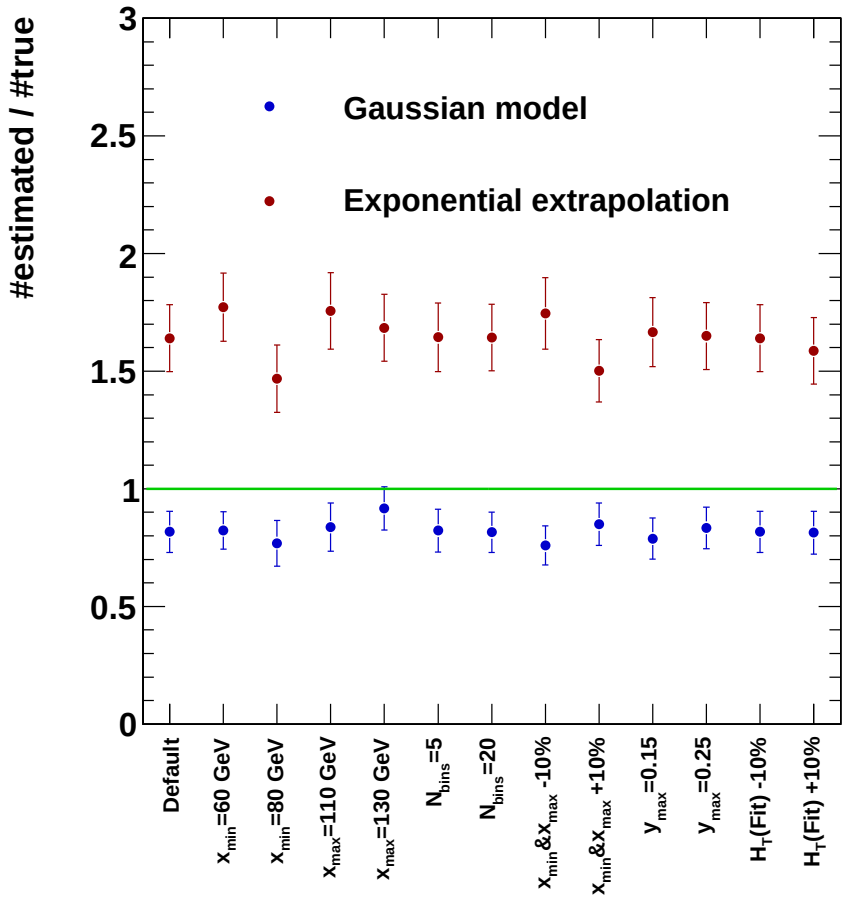


Figure 6.8: Number of estimated QCD events in the signal region for the robustness check with a MadGraph QCD simulation. All variations corresponding to the x-axis labels can be found in numbers in tab 6.1.

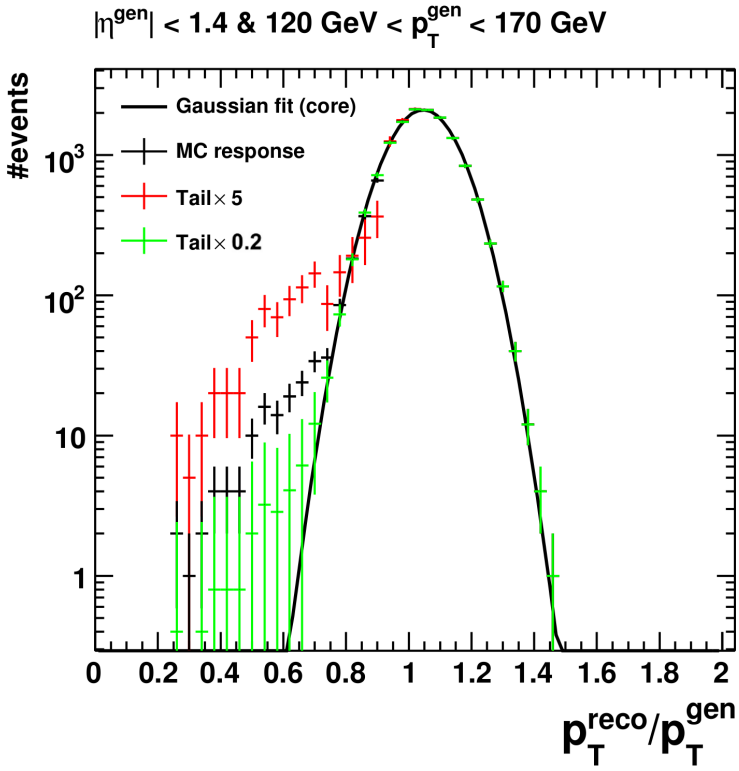


Figure 6.9: Monte Carlo response for jets in one p_T-η bin (QCD PYTHIA). The non-Gaussian tail is added to the core with different scaling factors $f = 5 \cdot f_0$ and $f = 0.2 \cdot f_0$ as described in the text.

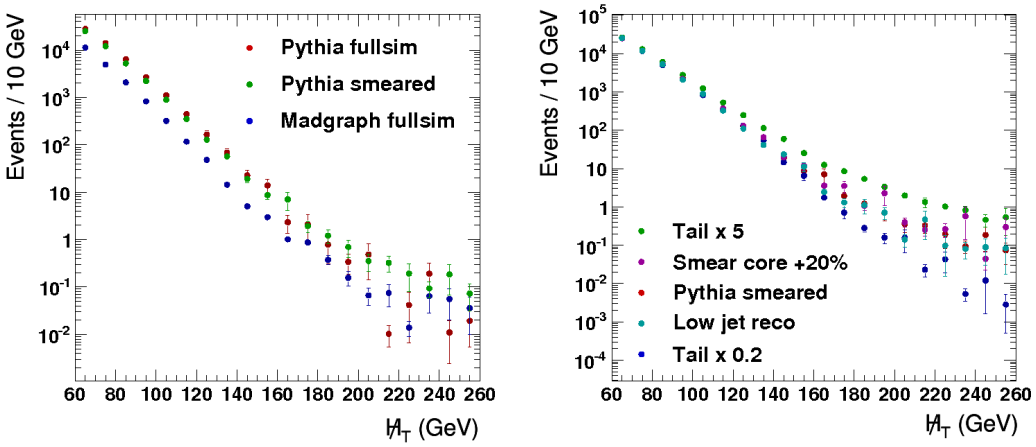


Figure 6.10: $\not{H}_T$ distribution after smearing of the generator jets. Left: Comparison of the smeared PYTHIA sample to the full simulation with PYTHIA and MADGRAPH. Right: Different smearing scenarios for the closure test as explained in the text.

sample with an additional smearing of 20% is used. The effect on the $\not{H}_T$ distribution (fig. 6.10) is small compared to the scaling of the tails.

One further scenario is added to account for possible not yet understood outliers in the jet resolution that produce jets that are reconstructed with a far too low jet energy and can not be described by the so far used scaling of the tails. Since in the tails of the data dijet asymmetry distribution no such extreme outliers have been found in $36\,\mathrm{pb}^{-1}$ data ([54]) an upper limit can be derived. This upper limit is then converted in a probability for jets to be reconstructed with extremely low energy ($P \approx 10^{-6}$) and applied during the smearing of generator jets for this scenario, which is also part of the closure test of the method.

In fig. 6.11 the influence from the scaling of the non-Gaussian tails and the additional smearing on the ratio $r(\not{H}_T)$ is visible. Both upscaling and downscaling results in a smaller ratio r which contradicts the naive expectation. But, since r is shown as a function of $\not{H}_T$ downscaling of the tails, thus reducing the average $\not{H}_T$ in an event, shifts the default to the left. For the upscaling the reverse effect is more than compensated by the large increase in the efficiency of the cut on $\Delta\phi_{\mathrm{min}}$.

The factorization method can be applied to the four pseudo simulated QCD

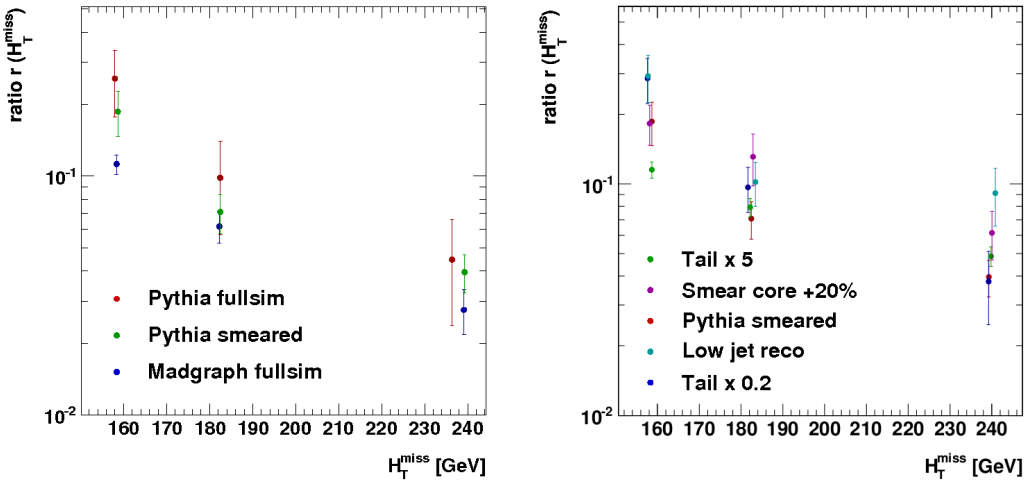


Figure 6.11: Ratio $r(\not{H}_T)$ for the fully simulated and different pseudo simulated QCD samples (by smearing generator jets). For details see fig. 6.10.

samples. By using the relative differences in the estimates from the default (no tail scaling) to the tail-scaled and additionally smeared samples the influence of these variations on the method can be evaluated. This procedure is used for the closure test.

The factorization method is applied to QCD MadGraph and Pythia and one default plus four extreme scenarios that have been described above. Figures 6.12 and 6.13 show the extrapolation for the two models in comparison to the true QCD ratio for all these MC samples. The QCD event yields for the baseline selection enter the last plot of fig. 6.13 which verifies that for all scenarios the two models bracket the QCD truth. This, together with the results from the two evolved selections is summarized in tab. 6.2.

The precision of these closure tests is limited to about 5-20 % for the baseline selection (20-50 % for the evolved $\not{H}_T$ selection) by the statistical errors of the simulated samples, especially the errors propagated from the fit covariance matrices. The largest deviation is found for the extreme low jet response scenario in both evolved selections (last row of tab. 6.2) and is still within two standard deviations.

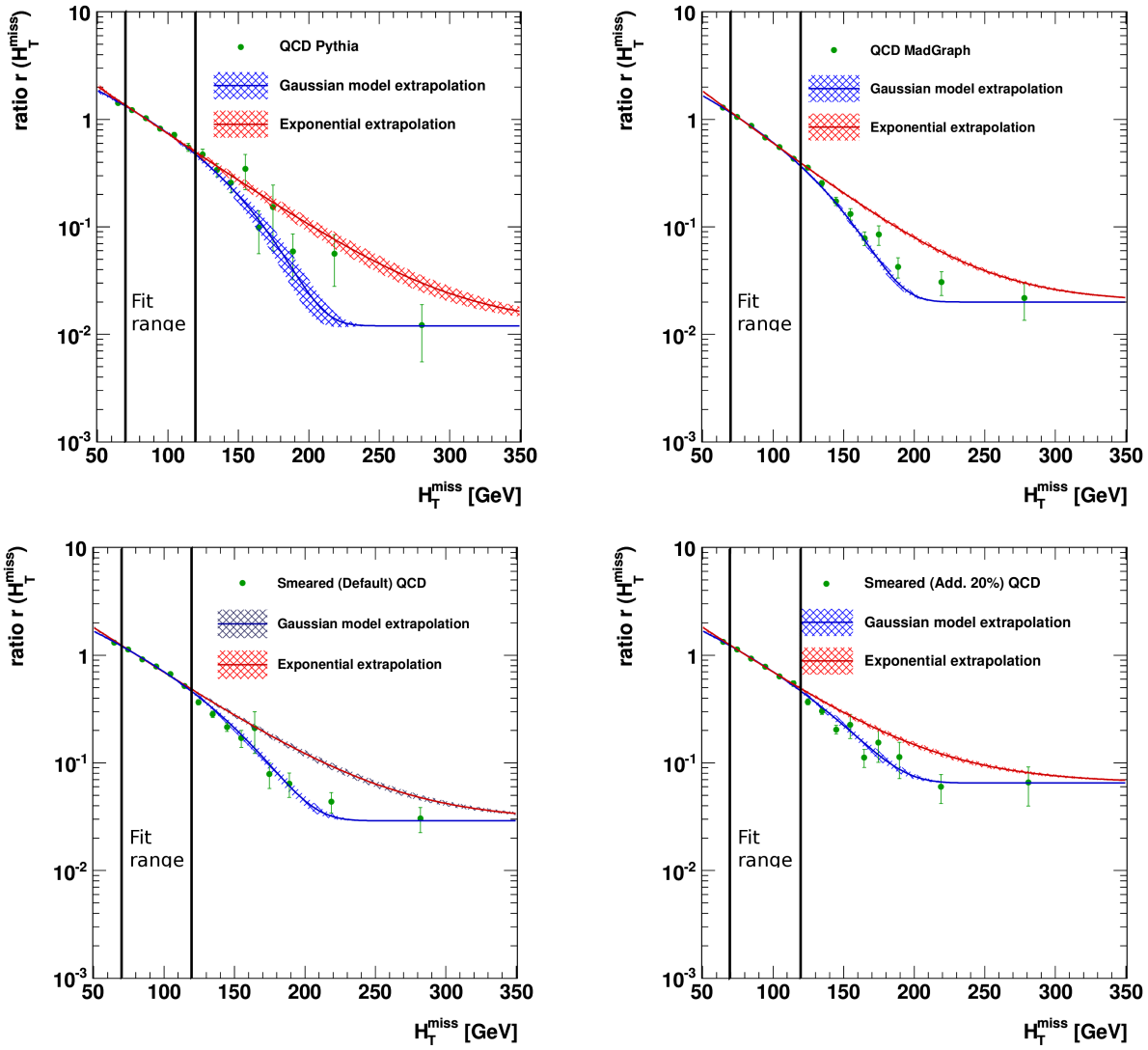


Figure 6.12: Ratio $r(\not{H}_T)$ for the two extrapolation methods for two QCD samples and for the first two of five different pseudo-simulated QCD variations described in the text (the others can be found in fig. 6.13). In each case, the functional form of the ratio $r(\not{H}_T)$ is fitted in the range $70\,\text{GeV} < \not{H}_T < 120\,\text{GeV}$ and then extrapolated to large $\not{H}_T$. The extrapolation of the fit function and error bands for the fit error propagation from the covariance matrix are shown.

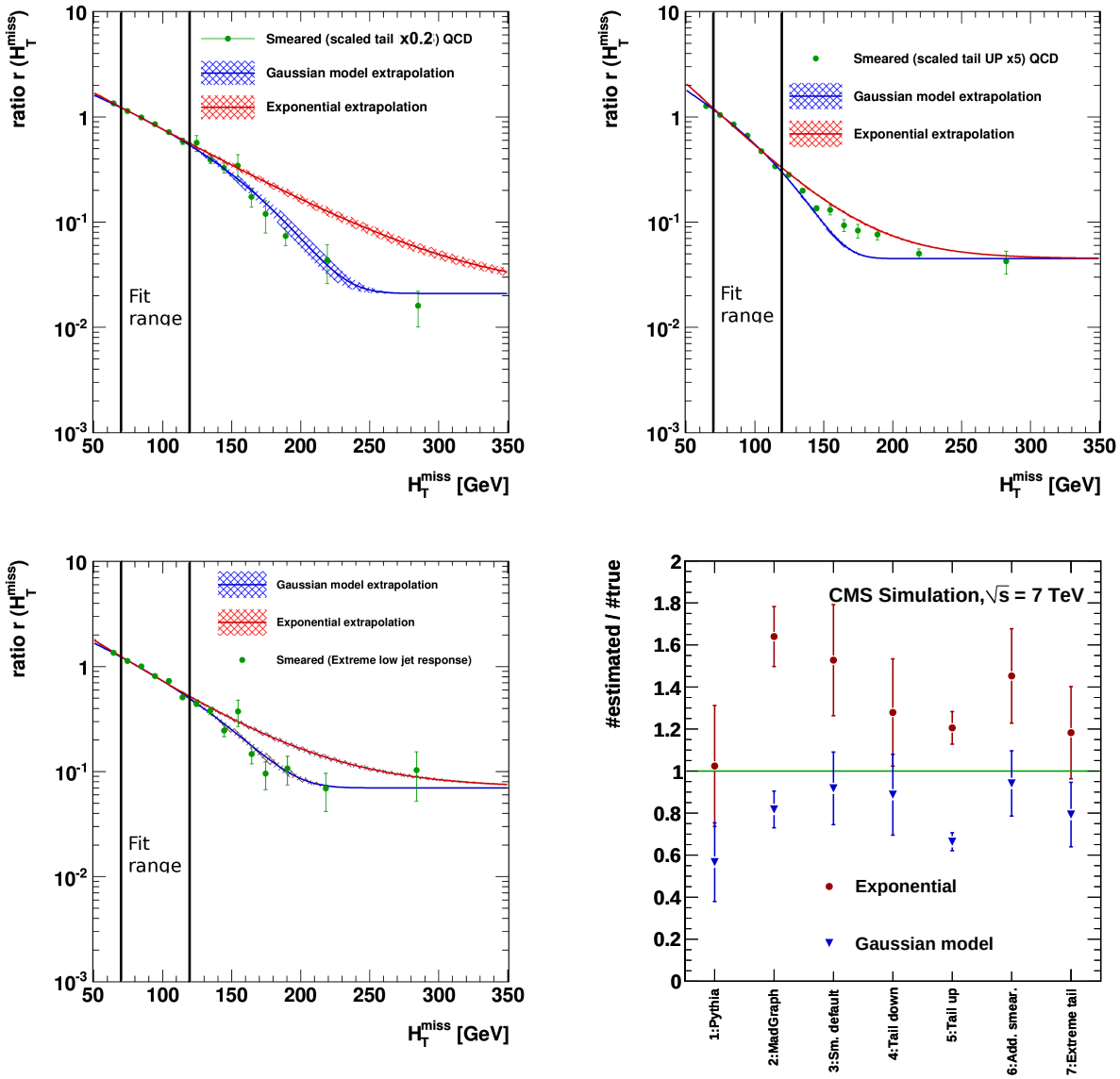


Figure 6.13: Ratio $r(\not{H}_T)$ for the two extrapolation methods for three different pseudo-simulated QCD variations described in the text. In each case, the functional form of the ratio $r(\not{H}_T)$ is fitted in the range $70\,\text{GeV} < \not{H}_T < 120\,\text{GeV}$ and then extrapolated to large $\not{H}_T$ (see also fig 6.12). The corresponding numbers can be found in tab. 6.2.

method	baseline			high-$\not{H}_T$			high-H_T		
		sys	stat		sys	stat		sys	stat
QCD PYTHIA									
exponential pred.	20.9	$^{+2.4}_{-2.4}$	$\pm$1.4	0.14	$^{+0.02}_{-0.02}$	$\pm$0.01	14.4	$^{+1.7}_{-1.7}$	$\pm$0.6
Gaussian model pred.	11.5	$^{+2.4}_{-2.4}$	$\pm$0.9	0.06	< 0.01	$\pm$0.01	7.8	$^{+1.6}_{-1.6}$	$\pm$0.4
plain MC simul.	20.4		$\pm$5.0	0.05		$\pm$0.03	12.6		$\pm$3.2
QCD MADGRAPH									
exponential pred.	9.4	$^{+0.4}_{-0.4}$	$\pm$0.4	0.11	< 0.01	$\pm$0.01	6.7	$^{+0.3}_{-0.3}$	$\pm$0.2
Gaussian model pred.	4.7	$^{+0.3}_{-0.3}$	$\pm$0.2	0.07	< 0.01	$\pm$0.01	3.3	$^{+0.2}_{-0.2}$	$\pm$0.1
plain MC simul.	5.8		$\pm$0.3	0.07		$\pm$0.03	4.8		$\pm$0.3
pseudo-sim QCD (Default)									
exponential pred.	32.1	$^{+1.9}_{-1.9}$	$\pm$0.7	0.48	$^{+0.03}_{-0.03}$	$\pm$0.02	21.2	$^{+1.2}_{-1.2}$	$\pm$0.4
Gaussian model pred.	19.2	$^{+1.8}_{-1.8}$	$\pm$0.4	0.29	< 0.01	$\pm$0.01	12.6	$^{+1.2}_{-1.2}$	$\pm$0.2
plain MC simul.	21.0		$\pm$3.4	0.29		$\pm$0.07	11.5		$\pm$1.4
pseudo-sim QCD (tail $\times$0.2)									
exponential pred.	12.3	$^{+0.8}_{-0.8}$	$\pm$0.6	0.11	$^{+0.01}_{-0.01}$	$\pm$0.01	8.9	$^{+0.6}_{-0.6}$	$\pm$0.3
Gaussian model pred.	8.6	$^{+0.9}_{-0.9}$	$\pm$0.4	0.04	< 0.01	$\pm$0.01	6.1	$^{+0.7}_{-0.7}$	$\pm$0.2
plain MC simul.	9.7		$\pm$1.8	0.04		$\pm$0.01	6.6		$\pm$1.2
pseudo-sim QCD (tail $\times$5)									
exponential pred.	75.1	$^{+2.4}_{-2.4}$	$\pm$0.8	2.04	$^{+0.02}_{-0.02}$	$\pm$0.03	47.1	$^{+1.4}_{-1.4}$	$\pm$0.4
Gaussian model pred.	41.3	$^{+1.3}_{-1.3}$	$\pm$0.4	1.91	< 0.01	$\pm$0.03	26.5	$^{+0.8}_{-0.8}$	$\pm$0.2
plain MC simul.	62.3		$\pm$3.4	1.91		$\pm$0.39	35.6		$\pm$1.6
pseudo-sim QCD (smear. +20%)									
exponential pred.	35.5	$^{+1.9}_{-1.9}$	$\pm$1.0	0.78	$^{+0.03}_{-0.03}$	$\pm$0.02	23.0	$^{+1.2}_{-1.2}$	$\pm$0.4
Gaussian model pred.	23.0	$^{+1.8}_{-1.8}$	$\pm$0.7	0.63	< 0.01	$\pm$0.02	14.7	$^{+1.1}_{-1.1}$	$\pm$0.3
plain MC simul.	24.4		$\pm$3.5	0.63		$\pm$0.22	15.8		$\pm$2.4
pseudo-sim QCD (low jet reco)									
exponential pred.	21.8	$^{+1.1}_{-1.1}$	$\pm$0.7	0.40	$^{+0.01}_{-0.01}$	$\pm$0.04	15.3	$^{+0.8}_{-0.8}$	$\pm$0.4
Gaussian model pred.	14.6	$^{+1.1}_{-1.1}$	$\pm$0.5	0.32	< 0.01	$\pm$0.03	10.2	$^{+0.8}_{-0.8}$	$\pm$0.2
plain MC simul.	18.4		$\pm$3.2	0.69		$\pm$0.26	8.4		$\pm$0.8

Table 6.2: Event yields for the different QCD samples for the baseline selection and for the evolved search selections. The column named systematics gives only the error propagation from the fit covariance matrix while the statistical uncertainty arises from control region D.

6.1.5 Contamination from SUSY and SM Processes

The method is safe against contamination of electroweak and top processes or SUSY signal in the fit region because of the very high QCD-multijet production cross section compared to other processes. Only if the upper border of the fit region is set to a too large $\not{H}_T$ value, the fitted ratio r will be different from the QCD-only case, because the non QCD events (SUSY signal, $W + jets$, $t\bar{t}$ or $Z \to \nu\bar{\nu}$) have "real" $\not{H}_T$ which is less correlated with $\Delta\phi_{min}$. The negative slope of the fitted function will be increased and consequently the extrapolation predicts too much background events in the signal region (see fig. 6.14).

While the effective over-prediction from the electroweak and top processes can be subtracted from the QCD estimation using Mont Carlo simulation for the relevant processes, the effect from SUSY events depends obviously on the kind of signal. However, the most optimistic signal in cMSSM models would give the largest effect, which turns out to be negligible for a fitting region with $x_{max} = 120\,\text{GeV}$. In addition, a variation of x_{max} in data would make large impacts on the fit visible, since an increase of x_{max} in the presence of signal would result in a higher QCD estimation compared to the case of decreasing x_{max}.

Furthermore, at large $\not{H}_T$ values signal and signal-like SM background events can populate the region at small $\Delta\phi_{min}$ (D) used for the application of the extrapolated ratio r which results in an overestimation of the background as described next.

Signal Contamination

The ratio r is fitted at low $\not{H}_T$ values (fit region) in order to be dominated by QCD events. Although $\Delta\phi_{min}$ is an important variable to discriminate between QCD and a supersymmetric signal, there are signal events expected to be in the control region at large $\not{H}_T$ and small $\Delta\phi_{min}$ and, depending on the cross section, signal contamination may lead to a significant overestimation of background events in the signal region. It is important to quantify this overestimation, in particular if a cross section measurement is performed. However, for the discovery of supersymmetry this is not a critical issue, since larger signal contamination is always accompanied by a much larger signal event yield in the signal region. In this context signal contamination can be called a "luxury problem".

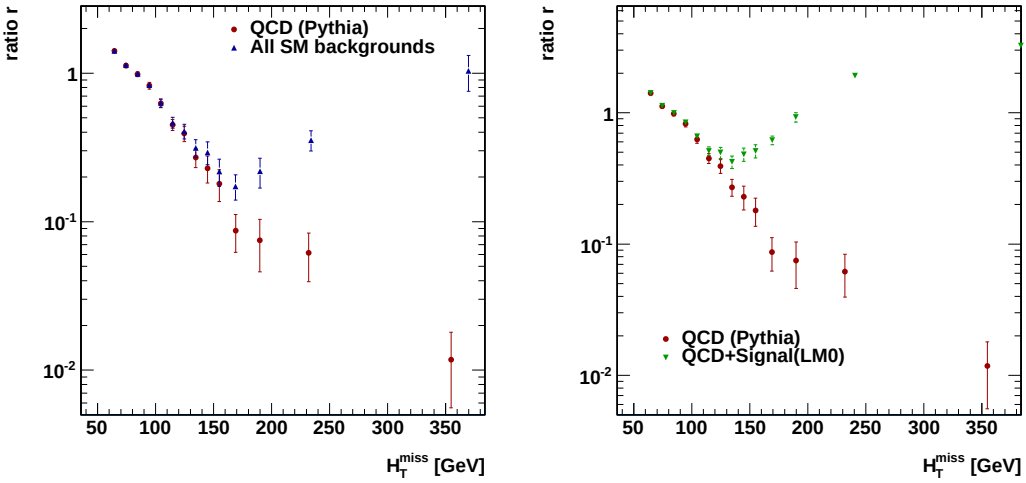


Figure 6.14: Left: The ratio r for QCD compared to all SM backgrounds(QCD, $W + \text{jets}$, $t\bar{t}$, $Z \to \nu\nu$). Right: Addition of an example signal (LMo) to the QCD. Additional to the standard selection a cut of $H_T > 700\,\text{GeV}$ is used.

SM Background

Similar to the case of contamination by signal events, standard model background will lead to an overestimation of the QCD background. The number of over-predicted events from $W + \text{jets}$ and $t\bar{t}$ processes depend on the definition and performance of the direct lepton veto. Fully hadronic decay modes of $W + \text{jets}$ or $t\bar{t}$ events have no intrinsic $\not{H}_T$ and are similar to QCD events. Therefore, their contribution will be automatically included by the presented method. However, the number of such events in the signal region is expected to be very small.

6.2 Application of the Factorization Method in Data

6.2.1 Control Regions

The control regions for the factorization method are defined by the baseline selection (sec. 4.3) and the scenario established in sec. 6.1.3. The fit region contains region A (events with $\Delta\phi_{\text{min}} < 0.2$) and B (events passing the $\Delta\phi_{\text{min}}$ criteria for the signal region). The additional cut $H_T > 600\,\text{GeV}$ is chosen such

that each bin in the $\not{H}_T$ distributions for the control regions (fig.6.15) has a minimum statistics of at least 25 events.

6.2.2 Verification of the Model in Data

The basic assumptions of the Gaussian resolution model discussed in sec. 6.1.2 can be directly verified in data (fig. 6.16). Here, in contrast to the final QCD event yields the electroweak and top background is not subtracted from the data, which makes the impact of such a contamination visible, and would imply the danger of fake signal contamination. The application to data demonstrates that the parameters of the Gaussian resolution model are very similar to the ones found in simulation. Especially in the QCD dominated region ($\not{H}_T < 150\,\text{GeV}$) the agreement between data and QCD simulation is good. In the high $\not{H}_T$ region, where electroweak and top processes lead to deviations, the exponential form of the distribution is still preserved.

The robustness checks previously carried out for Monte Carlo QCD samples (sec. 6.1.4) are also accomplished for data. Figure 6.17 shows the fit for the two chosen models with the default setting and also one variation is picked out of the various test. With the amount of data taken so far, no indication of a preferred model can be seen in the fit region.

Table 6.3 summarizes the results of the robustness checks in data. The event yields for all considered variations are also visualized in fig. 6.18. For both chosen models the deviation from the default scenario stay well within 20% while the statistical uncertainty is of the order of 20-40% for the different variations. A shift of the $\not{H}_T$ boundaries to higher values cause a slight rise in the QCD prediction that can be explained by contamination from other SM backgrounds. The size of the effect indicates that the correction and associated uncertainty (sec. 6.2.3) only have minor impact on the final result. Furthermore, no sign of a large signal contamination has been spotted, which should have resulted in a rise of the event yield from $y_{max} = 0.15$ to $y_{max} = 0.25$.

6.2.3 Systematic Uncertainties

Three sources of systematic uncertainties are considered for both functional forms of the factorization method independently. The total uncertainty is then taken from the lower edge of the uncertainty band of the lower boundary (Gaussian resolution model) and the upper edge of the uncertainty band of the upper boundary model (exponential fit) and used in the following analysis.

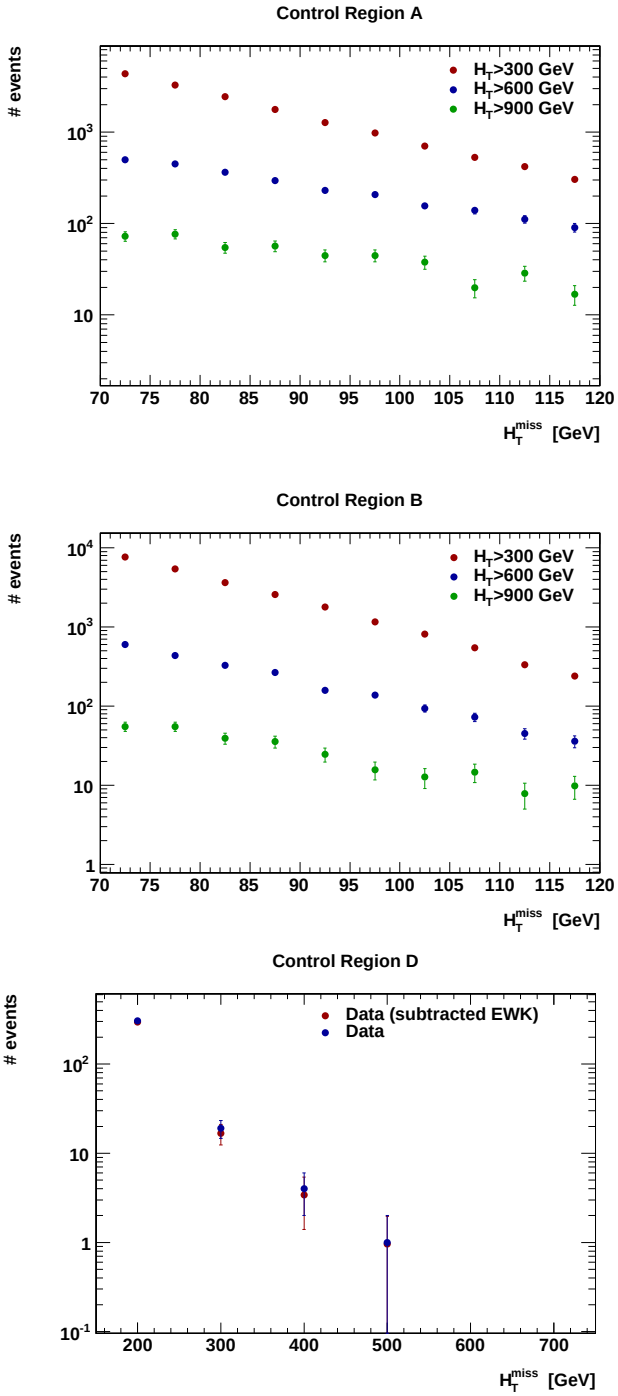


Figure 6.15: $\not{H}_T$ distributions of the control regions defined in sec. 6.1.3 for data. The subtraction of non-QCD backgrounds ($t\bar{t}$, W + jets $\rightarrow$ $l\nu$ + jets, and Z + jets $\rightarrow$ $\nu\nu$ + jets) using MC is only visualized for region D where it has the biggest impact.

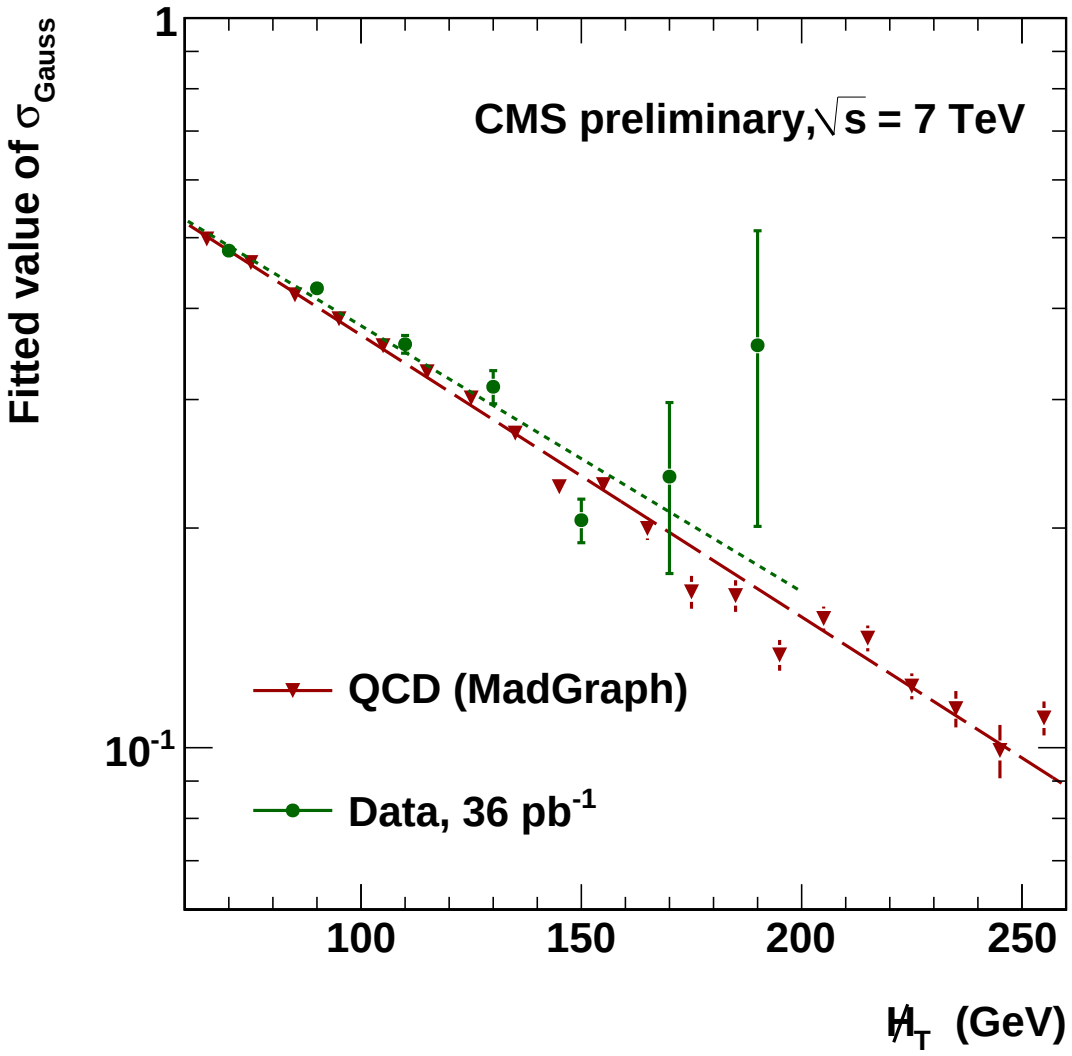


Figure 6.16: Data to simulation comparison of the σ of the Gaussian fits with a fixed mean=0 of $\Delta\phi_{\text{min}}$ slices distributions as described in sec. 6.1.2 and shown in fig. 6.4.

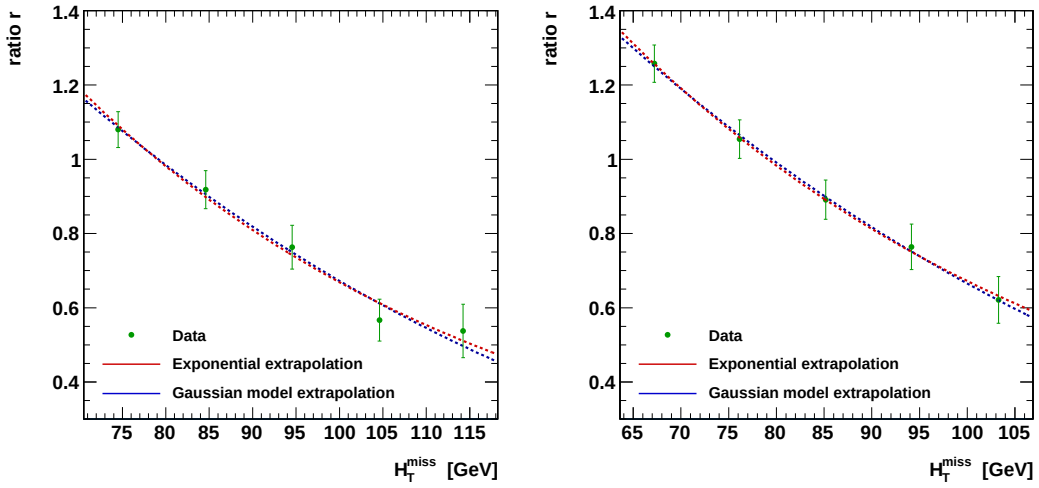


Figure 6.17: The two models fitted to data points in the default fit region (left) and (as example of the robustness check) with the boundaries in $\not{H}_T$ varied by -10% (right).

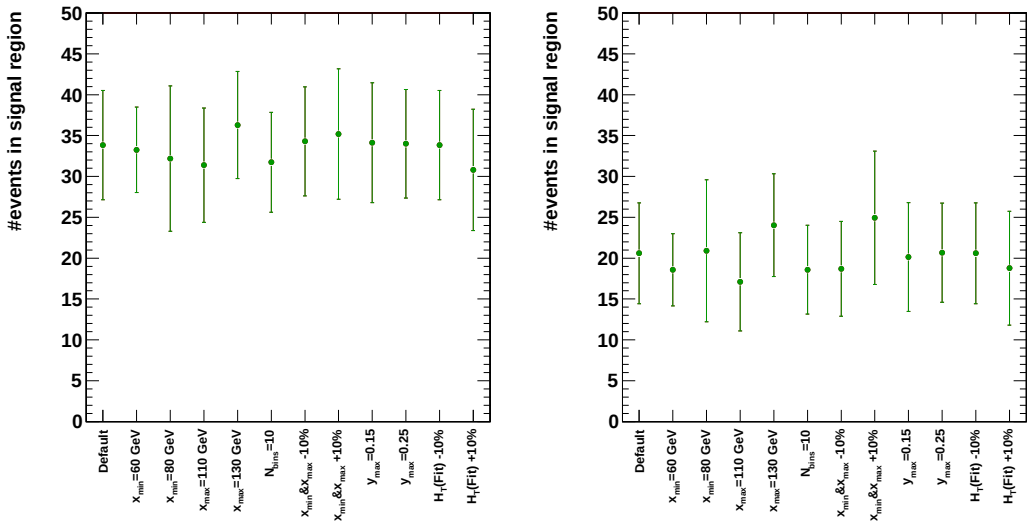


Figure 6.18: Results of the robustness check for data for the exponential extrapolation (left) and the Gaussian model (right). All variations corresponding to the x-axis labels can be found in numbers in tab. 6.3. The statistical errors arise from the error propagation from the fit covariance matrix and the statistics in control region D.

Variation	χ^2/ d.o.f.	# estimated QCD	estimated QCD fraction
Exponential extrapolation:			
$Default$	1.1/3	34 ± 7	0.3 ± 0.067
$x_{min} = 60\,\text{GeV}$	0.8/3	33 ± 5	0.3 ± 0.055
$x_{min} = 80\,\text{GeV}$	2.3/3	32 ± 9	0.29 ± 0.085
$x_{max} = 110\,\text{GeV}$	2.2/3	31 ± 7	0.28 ± 0.069
$x_{max} = 130\,\text{GeV}$	3.1/3	36 ± 7	0.33 ± 0.067
$N_{bins} = 10$	7.4/8	32 ± 6	0.29 ± 0.061
$x_{min}\&x_{max} - 10\%$	0.1/3	34 ± 7	0.31 ± 0.067
$x_{min}\&x_{max} + 10\%$	4.6/3	35 ± 8	0.32 ± 0.078
$y_{max} = 0.15$	0.2/3	34 ± 7	0.31 ± 0.072
$y_{max} = 0.25$	1.1/3	34 ± 7	0.31 ± 0.067
$H_T(Fit) - 10\%$	1.1/3	34 ± 7	0.3 ± 0.067
$H_T(Fit) + 10\%$	2.3/3	31 ± 7	0.28 ± 0.072
Gaussian model:			
$Default$	1.1/3	21 ± 6	0.19 ± 0.058
$x_{min} = 60\,\text{GeV}$	1.3/3	19 ± 4	0.17 ± 0.043
$x_{min} = 80\,\text{GeV}$	2.5/3	21 ± 9	0.19 ± 0.08
$x_{max} = 110\,\text{GeV}$	1.7/3	17 ± 6	0.15 ± 0.056
$x_{max} = 130\,\text{GeV}$	3.2/3	24 ± 6	0.22 ± 0.06
$N_{bins} = 10$	7.3/8	19 ± 5	0.17 ± 0.052
$x_{min}\&x_{max} - 10\%$	0.1/3	19 ± 6	0.17 ± 0.055
$x_{min}\&x_{max} + 10\%$	5.6/3	25 ± 8	0.23 ± 0.077
$y_{max} = 0.15$	0.1/3	20 ± 7	0.18 ± 0.062
$y_{max} = 0.25$	1.0/3	21 ± 6	0.19 ± 0.058
$H_T(Fit) - 10\%$	1.1/3	21 ± 6	0.19 ± 0.058
$H_T(Fit) + 10\%$	3.0/3	19 ± 7	0.17 ± 0.065

Table 6.3: Robustness check for data ($36\,\text{pb}^{-1}$) using the baseline selection. All boundaries of the control regions have been varied independently as well as simultaneously for the fit region in $\not{H}_\text{T}$. The default scenario corresponds to the default scenario established for the QCD samples in sec. 6.1.4. For these checks the subtraction of electroweak and top background from the data control regions has not been applied. The last column represent the ratio of estimated QCD events to all measured data events in the signal region.

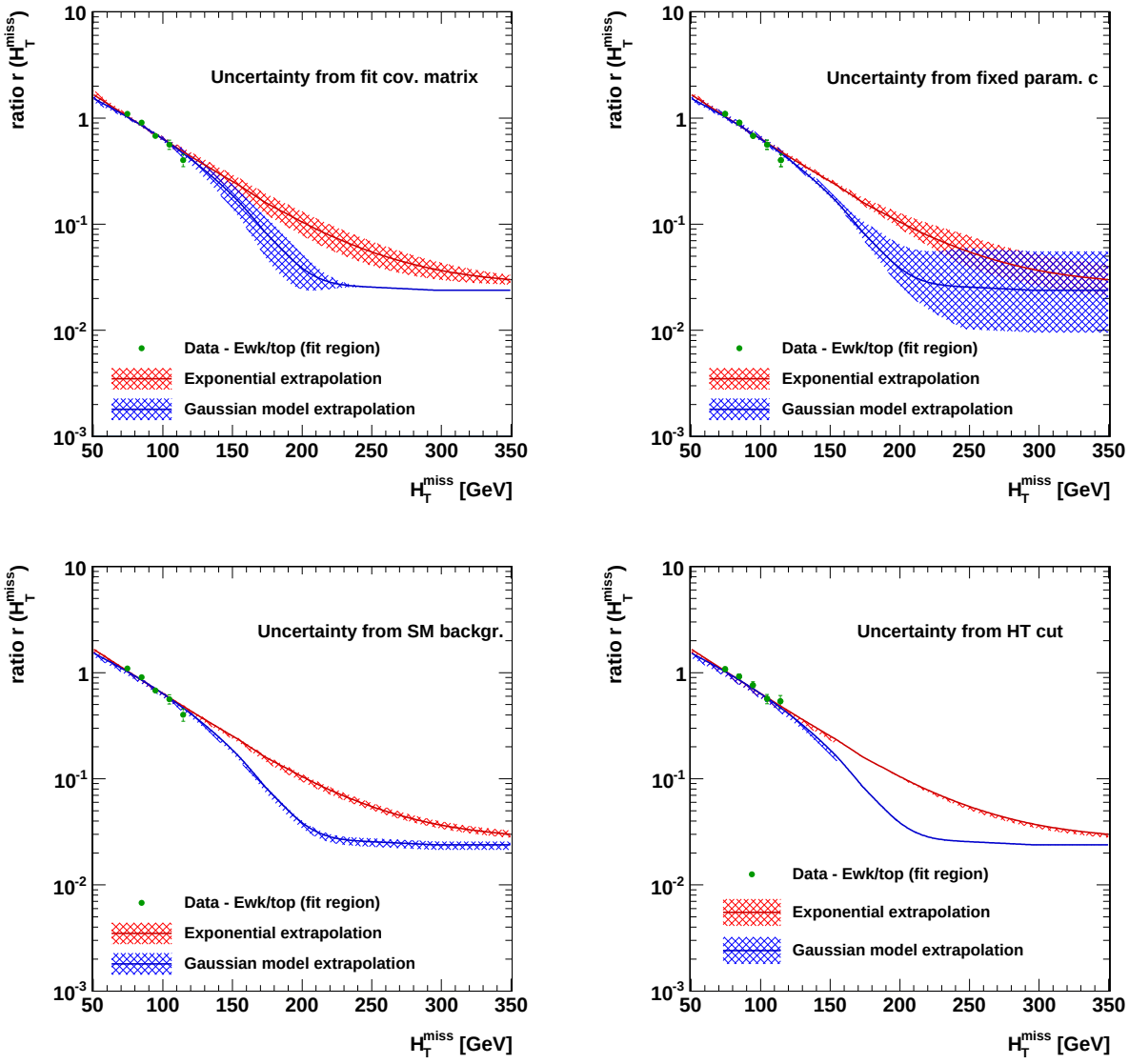


Figure 6.19: Uncertainty bands for the two chosen models from the different systematic effects. The upper and lower edges of the error bands in the signal region of $\not{H}_T > 150\,\text{GeV}$ are used to calculate the different uncertainties of the QCD estimation.

Uncertainty of the fixed Parameter

The parameter c in eq. 6.6 and eq. 6.5 has to be taken from Monte Carlo since it can not be derived from the control regions. From fig. 6.12 and fig. 6.13 we see that for the different QCD samples the values of c vary between $c = 0.012$ for PYTHIA and $c = 0.07$ for the last pseudo-simulated scenario with the extreme low jet response. These values are used for the lower and upper uncertainty while the estimate is done with $c = 0.03$ which corresponds to the default pseudo-simulated scenario.

The resulting uncertainty bands of both models for $r(\not{H}_T)$ are visualized in the upper right plot of fig. 6.19. The uncertainty becomes the dominant one for the exponential extrapolation at $\not{H}_T \gtrsim 250\,\text{GeV}$ and for the Gaussian model even at $\not{H}_T \gtrsim 200\,\text{GeV}$. However, for the baseline selection the uncertainty is not dominant since the region $150 < \not{H}_T < 200\,\text{GeV}$ gives by far the largest contribution to the QCD estimate.

Uncertainty of other SM Background Contamination

Contamination of the control regions with $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets} \to \nu\nu + \text{jets}$ events lead to an over-estimation of the factorization method (see sec. 6.1.5). The Monte Carlo expectations of these processes are subtracted from the data sample. The effect of this procedure can be seen for region D in fig.6.15 which has the biggest SM background contamination.

Scaling the SM Monte Carlo expectations by a factor 2, respectively 0.5 gives a worst case scenario of the uncertainty on the SM background contamination. The results for the two models can be seen in the uncertainty bands in the lower left plot of fig. 6.19.

Uncertainty of the H_T Correlation

In sec. 6.1.3 the conclusion has been drawn that a high H_T ($\gtrsim 600\,\text{GeV}$) cut in the fit region minimizes the influence due to the correlation of H_T and the ratio r. This can be confirmed by investigating the corresponding results in data (fig. 6.20) which are on the other hand less reliable due to the limited statistics. To account for a difference between Monte Carlo simulation and data, the H_T cut for the fit region is varied by $\pm 10\%$. The resulting discrepancies of the estimate in bins of $\not{H}_T$ are then used to evaluate this uncertainty (fig. 6.19, lower right).

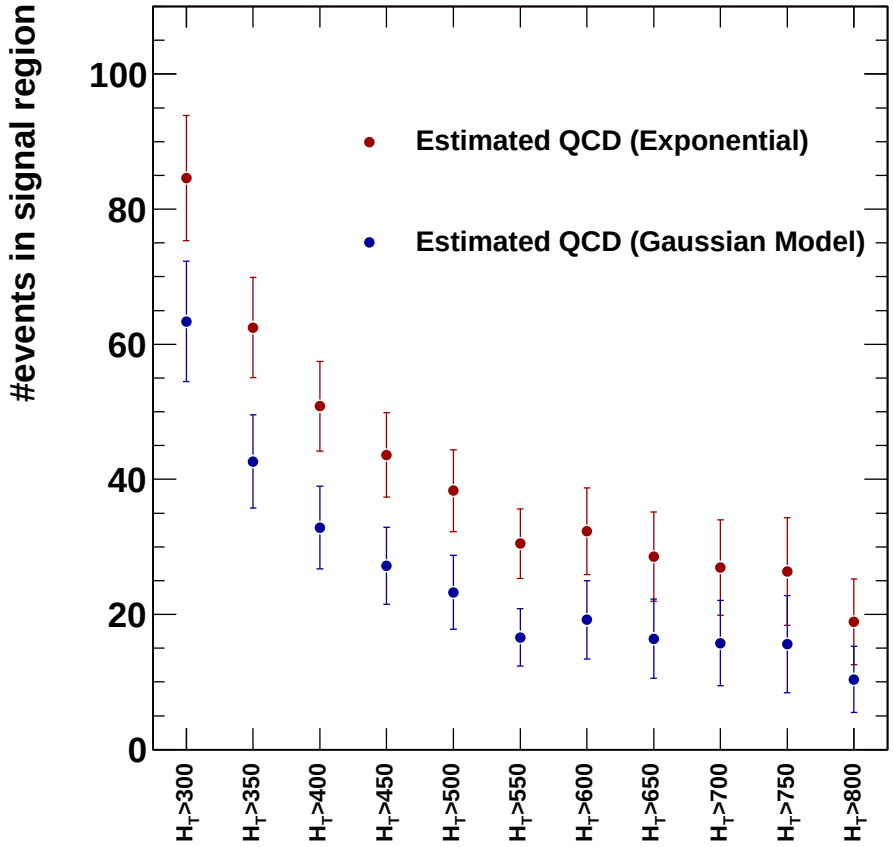


Figure 6.20: Results of the factorization method in data using different cuts on H_T in the fit region.

Uncertainty	baseline	high-$\not{H}_T$	high-H_T
Exponential extrapolation:			
Fit cov. matrix	$\pm 6.1[19\%]$	$\pm 0.09[18\%]$	$\pm 4.2[19\%]$
Fixed parameter c	$+3.3[11\%]$	$+0.20[40\%]$	$+2.2[10\%]$
	$-1.9[6\%]$	$-0.14[28\%]$	$-1.4[6\%]$
SM background cont.	$+1.1[4\%]$	$+0.04[8\%]$	$+0.5[2\%]$
	$-2.4[8\%]$	$-0.04[8\%]$	$-1.5[7\%]$
H_T cut	$+0$	$+0$	$+0$
	$-1.1[4\%]$	$-0.03[6\%]$	$-0.9[4\%]$
Gaussian model:			
Fit cov. matrix	$\pm 5.9[31\%]$	$\pm < 0.01$	$\pm 4.0[31\%]$
Fixed parameter c	$+4.0[21\%]$	$+0.39[130\%]$	$+2.8[22\%]$
	$-2.0[11\%]$	$-0.18[60\%]$	$-1.4[11\%]$
SM background cont.	$+0.6[3\%]$	$+0.02[7\%]$	$+0.4[3\%]$
	$-1.7[9\%]$	$-0.03[10\%]$	$-1.0[8\%]$
H_T cut	$+0$	$+0$	$+0$
	$-0.8[4\%]$	$- < 0.01$	$-0.5[4\%]$

Table 6.4: Systematic uncertainties (in number of events and relative to the estimated number) of the factorization method in data for the baseline selection and for the evolved search selections.

6.2.4 Summary of Results

The uncertainty bands of fig. 6.19 interpolate between four bins ($\not{H}_T = 150-160$, $160-180$, $180-200$, $200-2000$ GeV) in the signal region. A combination of the statistical and systematic uncertainties is shown in fig. 6.21. In the signal region ($\not{H}_T > 150\,\text{GeV}$) the influence from the electroweak and top backgrounds is clearly visible.

The upper and lower edges of these uncertainty bands are then used to weight the events in control region D according to the procedure in sec. 6.1.3. The results are summarized in tab. 6.4 for the baseline selection and for the evolved selections.

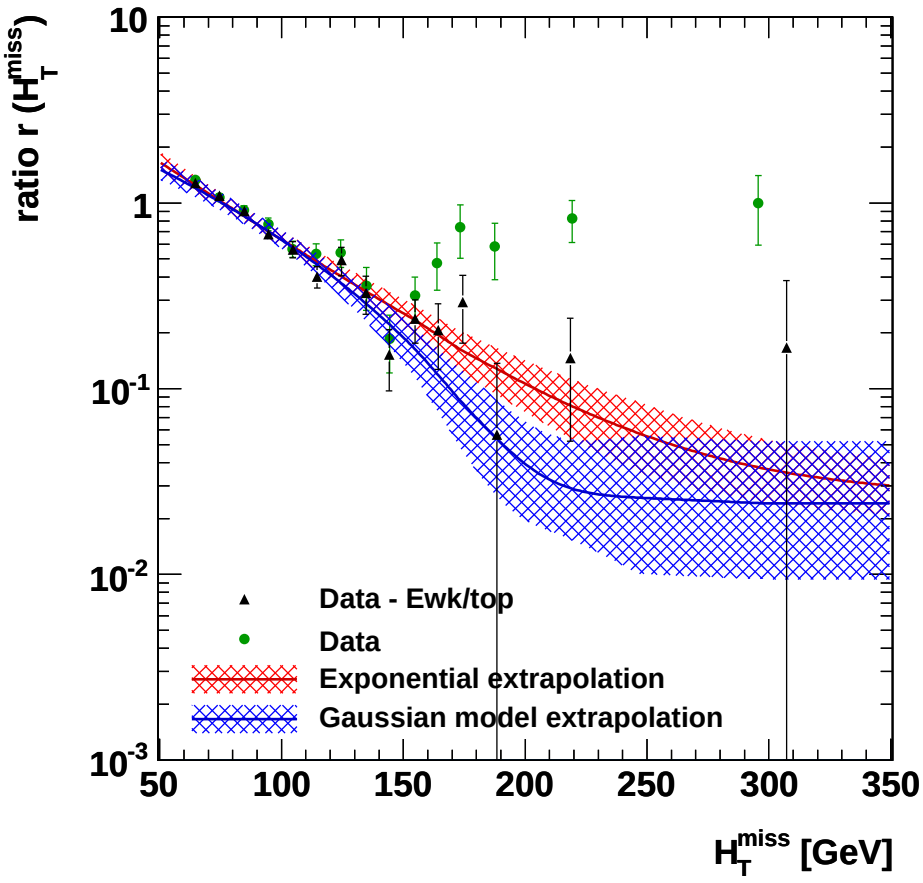


Figure 6.21: The fit and the extrapolation to the signal region of ratio r for the two chosen models using data with an integrated luminosity of 36 pb^{-1}. The fit is performed in the region ($70 < \not{H}_T < 120\,GeV$). The error bands of the extrapolations represent combined statistical and systematic uncertainties. Also shown is the data and the data with subtracted electroweak and top background (using MC).

The results of the factorization method for the three different selections are summarized in tab. 6.5 and visualized in fig. 6.22. The specific uncertainties are taken to be independent from one another, which means that the total systematic uncertainty, for each upper and lower uncertainty, is derived as the quadratic sum. For each selection the QCD estimate, the systematic uncertainty and the statistical uncertainty from the measurement in control region D is given. This can be compared with both PYTHIA and MADGRAPH Monte Carlo expectations and with the predictions from the method on Monte Carlo.

The best background estimate from this method is calculated from the average of both bracketing models. This is justified by the robustness tests in MC, where the true QCD event yield is distributed around the mean of both models. Half of the difference of the two models is assigned as additional systematic uncertainty which is combined linearly with the other uncertainties.

6.3 The Rebalancing and Smear Method

6.3.1 Basic Concept of the Method

The goal of the R&S method [55] is to construct a pseudo-simulated QCD sample from data seed events using parametrized jet resolution functions which can be measured in data. The final selections can then be applied to this sample in order to derive QCD background estimations.

A prerequisite of the method is a full measurement of jet response including the non Gaussian tails of the distributions. This can be achieved in different ways and is discussed later in this section. Once the jet resolution functions are available, the two main steps of the method are the construction of the seed events (rebalancing) and the application of the jet resolution functions to these seed events (smearing).

The rebalancing of events is done with an inclusive multijet data sample as input (seed sample). In each event, all measured n jet momenta are adjusted to bring the event into transverse momentum balance. For this a likelihood function of the true jet momenta $p_{\mathrm{T},i}^{\mathrm{true}}$ is constructed:

$$L = \prod_{i=1}^{n} r(p_{\mathrm{T},i}^{\mathrm{reco}} | p_{\mathrm{T},i}^{\mathrm{true}}), \tag{6.7}$$

where the jet resolution function r is taken to be a Gaussian distribution. The likelihood is maximized using the transverse momentum balance constraint

method	baseline			high-$\not{H}_T$			high-H_T		
		sys	stat		sys	stat		sys	stat
Data									
Exponential pred.	31.4	$^{+7.0}_{-6.9}$	± 2.4	0.5	$^{+0.2}_{-0.2}$	± 0.1	21.6	$^{+4.8}_{-4.8}$	± 2.0
Gaussian model	19.0	$^{+7.2}_{-6.5}$	± 1.6	0.3	$^{+0.4}_{-0.2}$	± 0.1	13.0	$^{+4.9}_{-4.4}$	± 1.3
Combined	25.2	$^{+14.0}_{-12.7}$	± 2.4	0.4	$^{+0.3}_{-0.3}$	± 0.1	17.3	$^{+9.4}_{-9.0}$	± 2.0
QCD Pythia									
Exponential pred.	20.9	$^{+2.4}_{-2.4}$	± 1.4	0.14	$^{+0.02}_{-0.02}$	± 0.01	14.4	$^{+1.7}_{-1.7}$	± 0.6
Gaussian model	11.5	$^{+2.4}_{-2.4}$	± 0.9	0.06	< 0.01	± 0.01	7.8	$^{+1.6}_{-1.6}$	± 0.4
Plain MC simul.	20.4		± 5.0	0.05		± 0.03	12.6		± 3.2
QCD MadGraph									
Exponential pred.	9.4	$^{+0.4}_{-0.4}$	± 0.4	0.11	< 0.01	± 0.01	6.7	$^{+0.3}_{-0.3}$	± 0.2
Gaussian model	4.7	$^{+0.3}_{-0.3}$	± 0.2	0.07	< 0.01	± 0.01	3.3	$^{+0.2}_{-0.2}$	± 0.1
Plain MC simul.	5.8		± 0.3	0.07		± 0.03	4.8		± 0.3

Table 6.5: Final event yield of the estimated QCD background for the baseline selection and for the evolved search selections high-$\not{H}_T$ and high-H_T, all for $36\,\text{pb}^{-1}$. Combined systematic uncertainties are shown for data while for the QCD MC samples only the error propagation from the fit covariance matrix is considered.

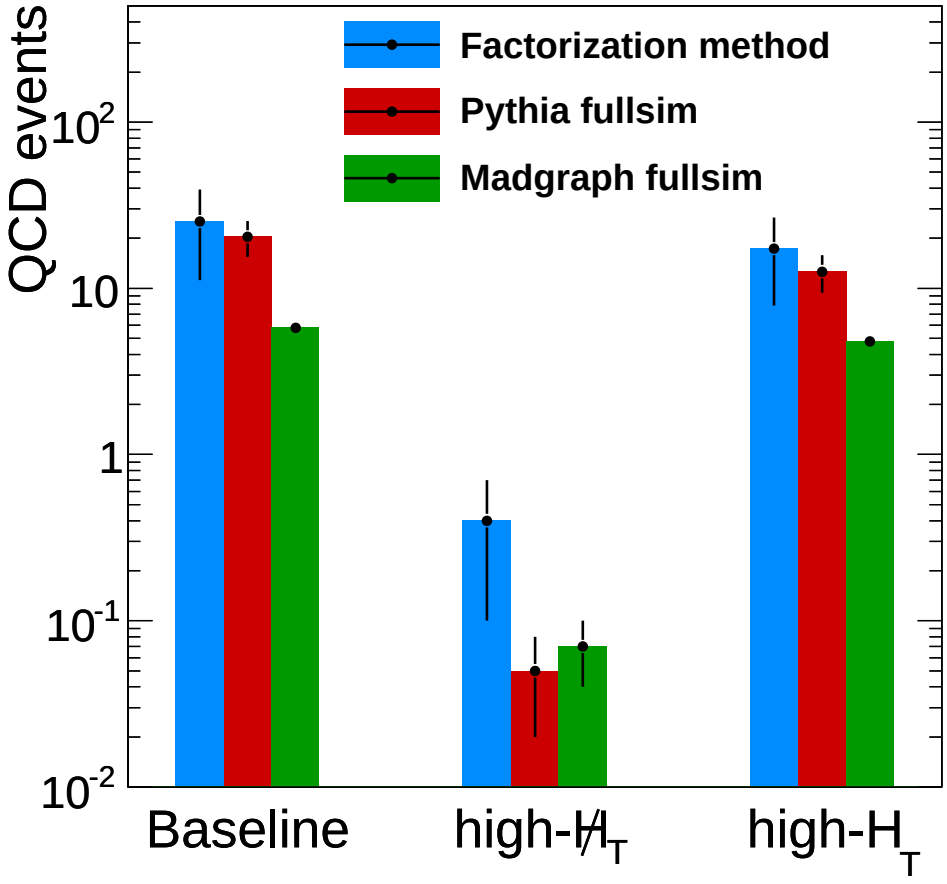


Figure 6.22: Final QCD background prediction for the baseline selection and for the evolved search selections high-$\not{H}_T$ and high-H_T for 36 pb^{-1}. The combined result of the factorization method is compared to PYTHIA and MADGRAPH full simulation. The numbers correspond to tab. 6.5.

$$\sum_{i=1}^{n} \vec{p}_{\mathrm{T},i}^{\,\mathrm{true}} + \vec{p}_{\mathrm{T,soft}}^{\,\mathrm{reco}} = 0, \tag{6.8}$$

where $\vec{p}_{\mathrm{T,soft}}^{\,\mathrm{reco}}$ comprises all particles not included in the jets. The assumption of a Gaussian resolution is justified by the vast majority of events that consist of jets with responses well within the core of the resolution distribution.

In QCD events, the four-momenta of rebalanced jets are good estimators of particle level jets. Events from SM processes that have true $\not{H}_{\mathrm{T}}$ or possible signal events are made QCD-like by the rebalancing procedure. A bias from these contributions can be safely neglected due to the huge QCD cross section that dominates the composition of the seed sample.

In the second step of the method, the momentum of each seed jet is smeared using the jet resolution distribution. After the event selection, the smeared sample can be used to predict all jet kinematic properties of QCD events in the search region.

6.3.2 Measuring the Jet Response

Two methods are used to measure from data a scaling factor for the Gaussian core of the jet momentum resolutions determined from simulation. At low p_{T}, γ+jet events are used [56] because the photons are reconstructed with excellent energy resolution and the p_{T} balance makes the photons good estimators of the true p_{T} scale of the event.

At larger p_{T}, dijet events are used [57] due to statistical reasons. An unbinned maximum likelihood fit is performed on the dijet asymmetry, $(p_{\mathrm{T}}^{\mathrm{jet1}} - p_{\mathrm{T}}^{\mathrm{jet2}})/(p_{\mathrm{T}}^{\mathrm{jet1}} + p_{\mathrm{T}}^{\mathrm{jet2}})$, with random ordering of the two highest-p_{T} jets.

For both measurements the presence of additional jets in the event destroys the momentum balance and an extrapolation to no-additional-jet activity is performed. These methods measure the core of the Gaussian resolution as a function of jet η to be $5-10\%$ larger in data compared to simulation, with systematic uncertainties of similar size as the deviation. No significant dependence on the p_{T} of the jet is observed.

No significant non-Gaussian tails are observed in γ+jet events. At higher p_{T}, the dijet asymmetry distributions show compatibility within uncertainties of the resolution tails in data and simulation. Using the ratio of these asymmetry distributions in data and simulation, correction factors to the jet resolution tails from simulation are derived. For the nominal resolution function, upper and lower tails are equally scaled. A systematic uncertainty band is taken from the envelope of varying the scaling from only low- to only high tail

scaling.

The resolution distributions are parametrized as function of p_T and η. Furthermore, an exceptionally low response arises at the specific $\eta - \phi$ locations where ECAL channels have been masked due to hardware problems. This effect is taken into account by parametrizing the jet response as a function of the fraction f_{ECAL}^{bad} of jet momentum lost in the masked area of the detector, computed using a template for the p_T-weighted distribution of particles as a function of the distance in η and ϕ to the jet axis. The dependence of the jet resolution on f_{ECAL}^{bad} is shown in fig. 6.23 (left). Note that $f_{ECAL}^{bad} < 0.1$ for 99 % of all events.

Finally, heavy-flavour b or c quarks and also gluons exhibit different jet resolution shapes than light jets, as shown in fig. 6.23 (right). For high jet p_T, decays of heavy-flavour hadrons into neutrinos become one of the dominant sources of significant jet energy loss. The jet resolution functions are determined for bottom, charm, gluon, and other light-flavour quarks separately. The flavour dependence is then accounted for by using these resolution functions in the smearing procedure according to the flavour fractions from simulation.

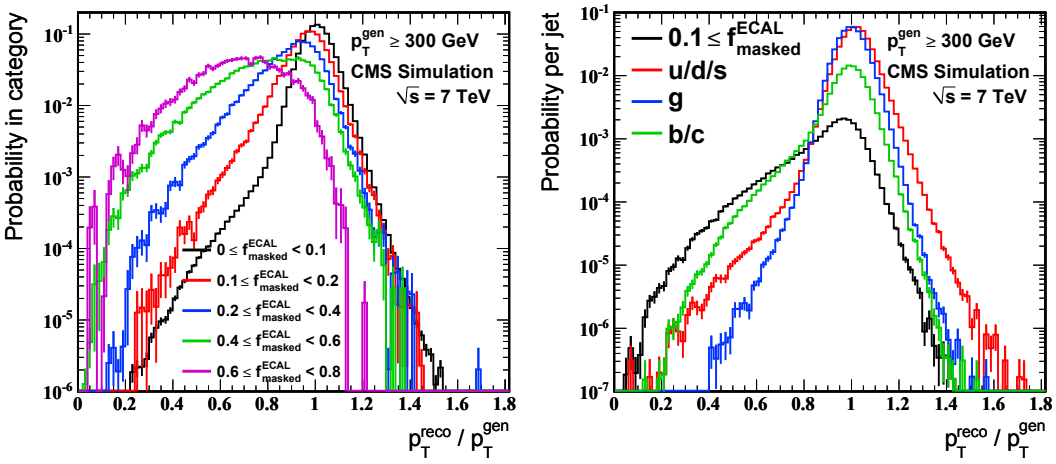


Figure 6.23: Ratio of the reconstructed jet transverse momentum and the generated transverse momentum for jets with $p_T^{gen} \geq 300\,\text{GeV}$. Distributions are shown for (left) different values of f_{ECAL}^{bad} and (right) gluons and different quark flavours. From the paper [1].

6.3.3 Results in Monte Carlo and Data

The distributions predicted by the R&S procedure are compared with those from MC simulation in fig. 6.24 and the corresponding numbers are given in tab. 6.6. The predicted $\not{H}_T$ and H_T distributions are within 40% of the plain MC distributions in the search regions.

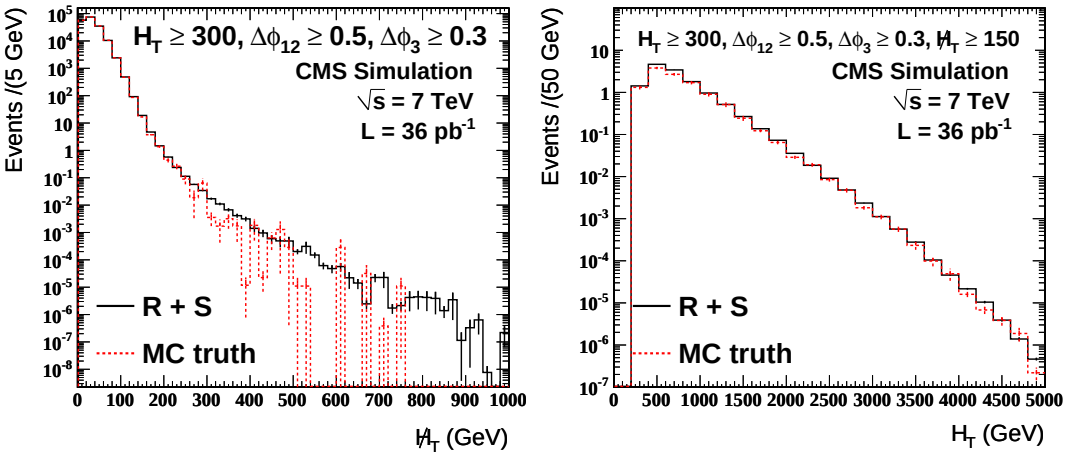


Figure 6.24: The (left) $\not{H}_T$ and (right) H_T distributions from the R&S method applied to simulation events, compared to MC distributions (MC truth), for events passing $\geq$ 3 jets, $H_T \geq 300\,\text{GeV}$, and $\Delta\phi(\not{H}_T, \text{jet 1-3})$ selections, and additionally $\not{H}_T > 150\,\text{GeV}$ for the right plot. From the paper [1].

For the QCD prediction in data the events selected by the H_T triggers described in sec. 4.2 are used. The R&S procedure applies jet energy resolution functions and the core and tail scale factors described above.

In tab. 6.7 the number of predicted events is listed for the search regions, along with the corrections of known biases of the method and the considered systematic uncertainties.

The largest correction pertains to the smearing step, and arises from ambiguities in how the jet resolution is defined and from limitations in the parametrization. It is obtained in simulation by comparing the prediction from smeared particle jets with the corresponding one from the detector simulation. The size of the difference is taken as both a bias correction and a systematic uncertainty.

A second bias is intrinsic to the rebalancing procedure, and is studied by iterating the R&S method. A first iteration $(\text{rebalance} + \text{smear})^{N1}$ of the

	Baseline selection No $\Delta\phi$ cuts	Baseline selection	high-$\not{H}_T$ selection	high-H_T selection
N(PYTHIA)	138.6 ± 1.3	11.4 ± 0.4	0.13 ± 0.04	8.46 ± 0.32
N(R&S)	160.2 ± 0.1	13.2 ± 0.1	0.177 ± 0.004	9.57 ± 0.04
Ratio	1.16 ± 0.01	1.15 ± 0.04	1.4 ± 0.4	1.13 ± 0.05

Table 6.6: Number of events passing the various event selections from the PYTHIA multijet sample, the R&S method applied to the same simulated sample, and their ratio. The uncertainties quoted are statistical only. From the paper [1].

method gives a sample of pure QCD multijet events with known true jet resolution, i.e., by construction the one used in the smearing step. Performing a second iteration $(\text{rebalance} + \text{smear})^{N2}$ of the method on this $(\text{rebalance} + \text{smear})^{N1}$ sample, using the same resolutions, provides a closure test of just the rebalancing part when compared to the input $(\text{rebalance} + \text{smear})^{N1}$ events. The degree of non-closure is measured to be 10%, which is also assigned as a systematic uncertainty.

The same $(\text{rebalance} + \text{smear})^{N2}/(\text{rebalance} + \text{smear})^{N1}$ procedure is employed to study the bias caused by using $\vec{p}^{\,\text{reco}}_{\text{T,soft}}$ as an estimator of $\vec{p}^{\,\text{true}}_{\text{T,soft}}$. The true value of $\vec{p}^{\,\text{true}}_{\text{T,soft}}$ in the second iteration is equal to the $\not{H}_T$ value calculated from the rebalanced jets in the first iteration. The difference between the $(\text{rebalance} + \text{smear})^{N2}$ predictions with $\vec{p}^{\,\text{reco}}_{\text{T,soft}}$ and $\vec{p}^{\,\text{true}}_{\text{T,soft}}$ as input is used as a third bias correction, with corresponding systematic uncertainty.

The largest systematic effect arises from uncertainties on the jet momentum resolution.

Further uncertainties which arise from the event selection and a contribution of pile-up events are found to be small.

The statistical uncertainty is associated with the size of the seed event sample. As prescribed by the bootstrap method [58], an ensemble of pseudo-datasets is selected randomly from the original seed sample, allowing repetition. The ensemble spread of predictions made from these pseudo-datasets is taken as the statistical uncertainty.

The uncertainties of the method are combined using the assumed shapes stated after the names in tab. 6.7. The mean and r.m.s. deviation of the resulting distributions are taken as the central values and uncertainties of the final R&S QCD prediction, which is stated in the last row of tab. 6.7.

	Baseline selection	high-$\not{H}_T$ selection	high-H_T selection
Nominal prediction (events)	39.4	0.18	19.0
Particle jet smearing closure (box)	+14%	+30%	+7%
Rebalancing bias (box)	+10%	+10%	+10%
Soft component estimator (box)	+3%	+19%	+4%
Resolution core (asymmetric)	+14% −25%	+0% −52%	+15% −21%
Resolution tail (asymmetric)	+43% −33%	+56% −78%	+48% −34%
Flavour trend (symmetric)	±1%	±12%	±0.3%
Pileup effects (box)	±2%	±10%	±2%
Control sample trigger (box)	−5%	−5%	−5%
Search trigger (symmetric)	±1%	±1%	0%
Lepton veto (box)	±5%	±0.05%	±0.2%
Seed sample statistics (symmetric)	±2.3%	±23%	±3.3%
Total uncertainty	51%	64%	49%
Bias-corrected prediction (events)	29.7 ± 15.2	0.16 ± 0.10	16.0 ± 7.9

Table 6.7: Number of QCD multijet events predicted with the R&S method, before and after bias corrections, along with all considered uncertainties and the type of uncertainty (uniform "box"-like, symmetric or asymmetric Gaussian distribution). Effects in italics are the biases corrected for, with the full size of the bias taken as the systematic uncertainty. From the paper [1].

6.3.4 Comparision of the two Methods

For the presented analysis two data-driven methods for the estimation of the important QCD background have been developed. The application to the $36\,\mathrm{pb}^{-1}$ data collected in 2010 results in very similar QCD predictions of both methods for all selections (see fig. 6.25). Comparing these predictions to tab. 4.2, both methods agree that the Monte Carlo simulation (PYTHIA and MADGRAPH) underestimates the number of QCD events.

In the end, also the sizes of the total uncertainties coincide. Both, the factorization method and R&S, assign about 50% total uncertainty to the evolved selection with increased H_T cut, where QCD has the largest relative contribution.

Since R&S can be seen as the more complete method which produces a pseudo-simulated QCD sample from data for further investigations and is safe against signal contamination, it has been set as the primary method for the limit calculation in this analysis (see ch. 7). The factorization method has served as a cross-check and was able to confirm the QCD estimations in all search regions.

It is known, that the QCD background is extremely difficult to model and also both data-driven methods are confronted with several biases which have complicated the procedures and increased the uncertainties. While the R&S has tried to disentangle different sources of biases and correct for them, the factorization method has found two enveloping functional forms for the prediction. Both procedures give rise to box-shaped systematic uncertainties which result in relatively large total uncertainties of the methods. While the statistical overlap of the control regions of the two methods have not been studied yet, it is reasonable to assume that the total systematic uncertainties are largely independent.

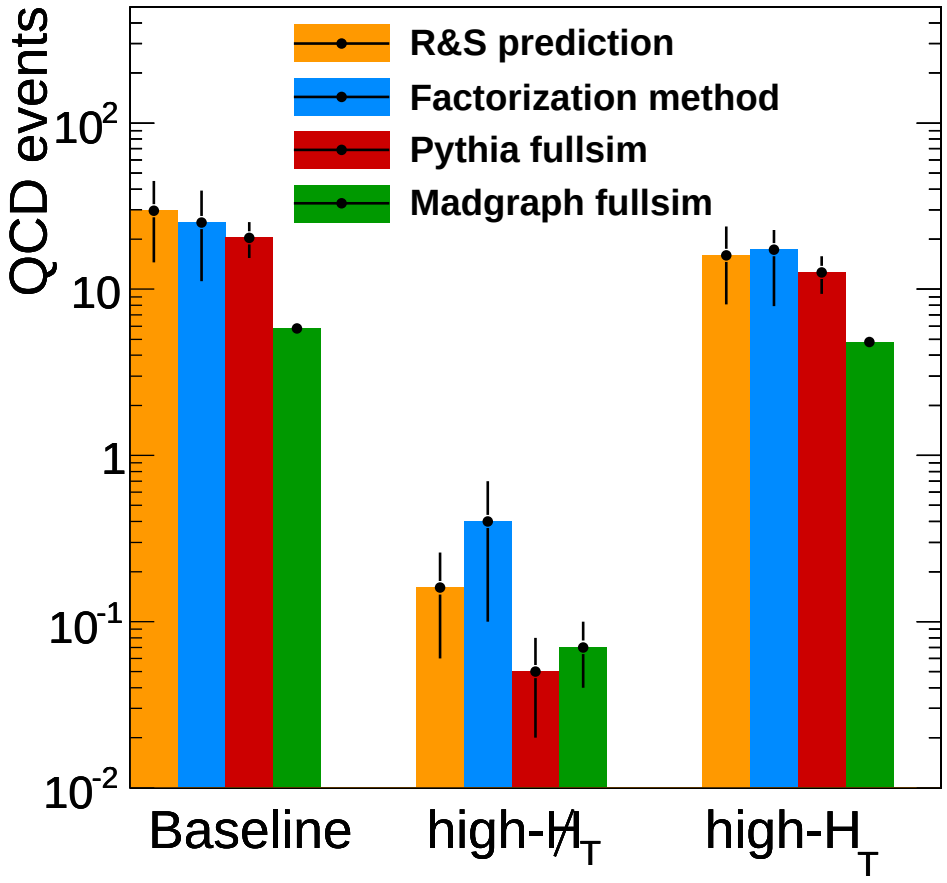


Figure 6.25: Final QCD background prediction for the baseline selection and for the evolved search selections high-$\not{H}_T$ and high-H_T for 36 pb^{-1}. The results of the R&S method are compared to the combined results of the factorization method and to the full simulation with PYTHIA and MADGRAPH. Numbers correspond to tab. 6.5 and tab. 6.7.

7 Search Results

In this chapter, the results of the search in the multijet and no lepton channel with CMS data taken in 2010 is presented (published in [1]). In previous chapters, the data event selection has been discussed (ch. 4) together with the cut-based reduction of the background. For each of the remaining SM backgrounds one or more data-driven methods have been used ($Z \to \nu\bar{\nu}$ and $W/t\bar{t}$ in ch. 5 and QCD in ch. 6) in order to obtain a reliable total background estimation which is given in sec. 7.1.

An extensive production of simulated signal samples is used to derive 95% C.L. exclusion limits for the important parameters of the cMSSM model. A short description of the hybrid CLs method applied for limit calculation as well as an interpretation of the results is presented in sec. 7.2.

While the search regions in this analysis are strictly limited by the relatively small amount of data available in 2010, subsequent analyses in the same channel can hugely benefit from search regions optimized for the best sensitivity. One possibility of finding and testing optimal cut scenarios is further investigated in sec. 7.3.

7.1 Combination of Background Estimations

The SM prediction for the number of events in the previously defined search regions is obtained as a combination of data-driven estimations of all contributing processes. The results are presented in tab. 7.1 together with the events observed in $36\,\mathrm{pb}^{-1}$ of data.

For a combination of the individual uncertainties correlations between the background estimations have to be taken into account. In case of the presented data-driven methods, possible overlaps between different control regions have been checked and found to be negligible.

Since some of the uncertainty sources can not be acceptably described by a Gaussian a convolution of different uncertainty shapes is used.

A detailed accounting of the possible correlations between the background estimations is essential for the total uncertainty combination. For this, all sources and corresponding uncertainties and the corresponding probability

distribution for each uncertainty were identified and combined using Monte Carlo integration.

Background	Baseline selection		High-$\not{H}_T$ selection		High-H_T selection	
$Z \to \nu\bar{\nu}$ (γ+jets method)	26.3	$\pm$4.8	7.1	$\pm$2.2	8.4	$\pm$2.3
$W/t\bar{t} \to e, \mu$+X	33.0	$\pm$8.1	4.8	$\pm$1.9	10.9	$\pm$3.4
$W/t\bar{t} \to \tau_{hadr}$+X	22.3	$\pm$4.6	6.7	$\pm$2.1	8.5	$\pm$2.5
QCD (R+S method - default)	29.7	$\pm$15.2	0.16	$\pm$0.10	16.0	$\pm$7.9
QCD (factorization method)	25.2	$\pm$13.4	0.4	$\pm$0.3	17.3	$\pm$9.4
Total background estimate	111.3	$\pm$18.5	18.8	$\pm$3.5	43.8	$\pm$9.2
Observed in 36 pb^{-1} of data	111		15		40	
95% C.L. limit on signal events	40.4		9.6		19.6	

Table 7.1: Predicted event yields from the different background estimation methods for the baseline selection and for the high-$\not{H}_T$ and high-H_T search selections. The total background estimate from data, used in the limit calculations, uses the R&S for QCD, the $Z \to \nu\bar{\nu}$ from photons and the $W/t\bar{t}$ lost-lepton and hadronic-tau estimates. The background combination is performed as explained in the text. The last line reports the 95% C.L. limit on the number of signal events given the predicted events of background and the observed events in data.

In this analysis, the number of observed events in the 2010 collision data set is in agreement with the SM prediction for both the high-$\not{H}_T$ and the high-H_T search region. Consequently, upper limits on model parameters can be set which will be discussed in the following sections.

7.2 Limits on SUSY Signals

Since no sign of a manifestation of a SUSY signal has been found with the analyzed data the search results are used to derive statistically significant constraints on the parameter space of SUSY models. It has been explained in sec. 2.2.3 that the presented search is expected to provide sensitivity to signatures from a wide range of different SUSY models. However, the full potential of the search can only be exploited with the higher luminosity data that will be recorded at the LHC in 2011 and the following years. An excellent starting point and reference for the early LHC data SUSY searches offers the

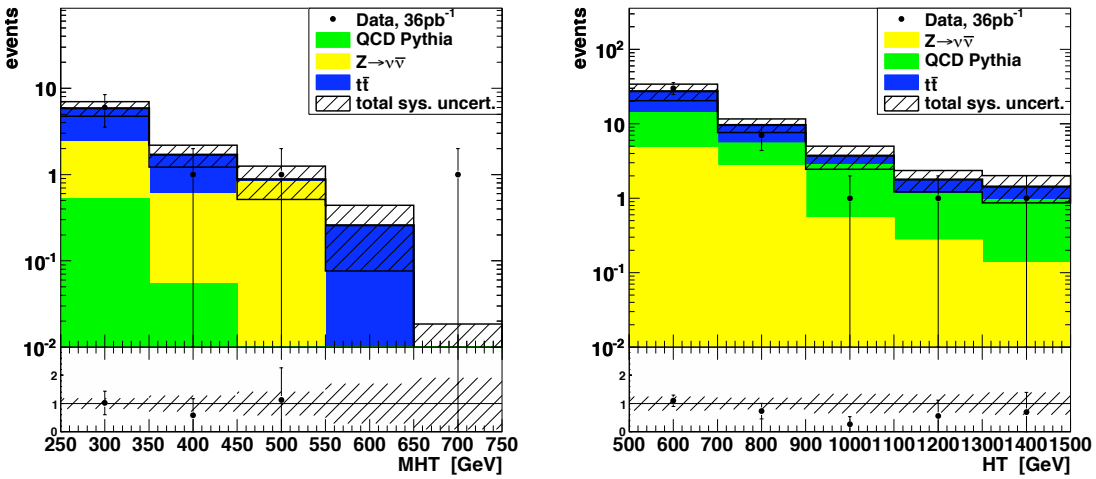


Figure 7.1: Stacked plot of the MC background distributions for $\not{H}_T$ and H_T for an integrated luminosity of 36 pb^{-1}.

cMSSM model which constraints the MSSM to only 4 parameters and a sign (discussed in sec. 2.2.2).

7.2.1 Signal Simulation and Uncertainty

For this, the cMSSM parameters m_0 and $m_{1/2}$ have been scanned in 10 GeV steps for three different values of $\tan\beta = 3$, 10, and 50. For each point 10 k signal events have been simulated and reconstructed with the CMS fast-simulation.

To be able to derive limits on new physics, the expected inefficiencies of the event selections need to be estimated on simulated signal samples, taking into account uncertainties corresponding to the selection and an overall luminosity uncertainty.

In tab. 7.2 all considered uncertainties are summarized for the LM1 signal benchmark point.

The largest contribution comes from the luminosity uncertainty. A maximal uncertainty of 1% was assigned to the trigger efficiency. What concerns jet energy scale and resolution uncertainties, the evaluation of the selection uncertainty brings in a dependence on p_T and η of the jets, and hence a model dependence. The evaluation of the uncertainties from the lepton veto is smaller than 2%. The inefficiency of the ECAL dead-cell filters was determined on the LM1 samples to be about 1.5% [44]. This full inefficiency is taken as uncertainty. For other event cleaning the systematic uncertainty is negligible, which is supported by very small inefficiencies observed in events passing the event selection before the $\not{H}_T$ requirement. These events can be seen as kinematically representative for potential signal, except for the lack of $\not{H}_T$.

7.2.2 The Hybrid CLs Method

The limit calculation in this analysis uses the modified frequentists procedure CLs together with a Bayesian-like integration of the systematic uncertainties and is therefore called a hybrid method. The CLs method has already been applied in several searches at the Tevatron and at LEP, especially in the Higgs searches, which is documented in [59]. The Cousins-Highland approach for incorporating systematic uncertainties into upper limits was suggested in [60].

Starting with a more general formulation, the presented search results are based on hypothesis testing. While the null hypothesis (background only) expresses the absence of a signal the alternate hypothesis (signal+background) claims that it exists. The confidence levels (C.L.) are then constructed to

Source	Uncertainty
Luminosity	11%
Trigger efficiency	2%
Jet energy scale & resolution	7%
Lepton veto	2%
ECAL dead-cell filters	1.5%
Statistical	2.5%

Table 7.2: Systematic effects on the signal efficiency and corresponding uncertainties, evaluated on the LM1 signal benchmark point. The statistical uncertainty corresponds to 10 k generated events, as the cMSSM scan points. Some uncertainties, like the jet energy scale, depend on the scan point.

quantify the level of agreement or exclusion of the hypotheses with the experimental observation.

The hypotheses are described by a function of the observables and model parameters (in our case systematic uncertainties for background and signal) which is called *test-statistic* Q. For counting experiments, like this analysis, the observables are simply number of events n_k in the investigated search channels.

A hypothesis test is called most powerful if it minimizes the probability of falsely excluding a true signal (type-II error) under a fixed false discovery rate (type-I error) which is known as the Neyman-Pearson criterion [61]. The optimal test-statistic Q is the ratio of the probability density functions (p.d.f.) for the signal+background hypothesis over the background-only hypothesis:

$$Q(n_k) = -2 \ln \frac{L_{S+B}(n_k)}{L_B(n_k)} \tag{7.1}$$

It is conventional to use the negative logarithmic likelihood ratio. In this notation the experimental results rank from most signal-like to most background-like (as at the x-axis of fig. 7.2).

For counting experiments, the observables are distributed according to Poisson statistics which already gives the full p.d.f. as long as the uncertainties on the signal and the background are neglected. Multiple exclusive channels, or independent bins of a histogram are represented as a product of the individual p.d.f.'s:

$$L_{S+B}(n) = \prod_{k}^{N_{ch}} Poiss(n_k|s_k + b_k) \tag{7.2}$$

Systematic uncertainties are incorporated into L_{S+B} and L_B with help of nuisance parameters δ_j that are distributed according to their own p.d.f.'s $P_j(\delta)$. The Hybrid method uses the distribution of the systematics $P(\delta)$ as prior in the integration of the p-values[1]:

$$L_{S+B}(n) = \int L(n|\mu)P(\delta)d\delta \ , \tag{7.3}$$

where for a precisely known absolute signal s and and only one source of uncertainty of the total background $\mu = s + b(1+\delta)$. In practice, the total background estimate b is the sum of different backgrounds which can have common and individual sources of uncertainty. Examples of common systematic uncertainties are instrumental effects like the luminosity uncertainty or the jet energy scale uncertainty. These sources of uncertainties are also shared between signal and background prediction in analyses that completely rely on Monte Carlo simulation. In this case, individual uncertainties of the different backgrounds and the signal would be caused by theory uncertainties on the cross-sections.

Technically, the common systematics can be described by one nuisance parameter each, introducing scaling factors f_i^0 for the individual backgrounds b_i:

$$f_i^0 = \frac{\Delta_i}{max(\Delta_i)} \ , \tag{7.4}$$

where $max(\Delta_i)$ is the maximum deviation due to the uncertainty j of all backgrounds and it is also used as σ_j. An explicit example of a single channel likelihood with one common Gaussian uncertainty and one individual Gaussian uncertainty for each background can then be written as:

$$L_{S+B}(n) = Poiss(n|s + \sum_{i}^{N_{bkgr}} b_i(1 + f_i^0\delta_0)(1+\delta_i))Gauss(\delta_0|0,\sigma_0) \prod_{j}^{N_{bkgr}} Gauss(\delta_j|0,\sigma_j) \tag{7.5}$$

The integral in eq. 7.3 has become multidimensional and is calculated nu-

[1]The p-value is defined as the probability of obtaining a test statistic which is at least as signal-like as the observed one under the assumption that the background-only hypothesis is true.

merically using Monte Carlo "toy" experiments. First, the nuisance parameters are simulated according to their distributions. Then, the expected average of the Poisson distribution is calculated using the nuisance parameters. For each 'toy' experiment the test-statistic Q (eq. 7.1) and a resulting p-value is calculated by this procedure. Half of the toy experiments take as input n_B which is distributed according to the background-only model and the other half takes n_{S+B} which represents signal+background input.

An advantage of the Hybrid method is that it can accurately describe not only Gaussian distributed systematics but also also others especially asymmetric ones. The precision, however, is determined by the number of toy experiments and large number of nuisance parameters can significantly increase the computing time.

Figure 7.2 shows two examples of the test statistics distributions that were obtained for the evolved $\not{H}_{\mathrm{T}}$ selection for two different signals in the m_0-$m_{1/2}$ parameter plane where the test-statistic Q is defined as in eq. 7.1. While in the first example case ($m_0 = 120\,\mathrm{GeV}$, $m_{1/2} = 240\,\mathrm{GeV}$, $\tan\beta = 10$) the two distributions are well separated, the second picture shows a case ($m_0 = 120\,\mathrm{GeV}$, $m_{1/2} = 340\,\mathrm{GeV}$, $\tan\beta = 10$) where the distinction is much less pronounced.

The confidence intervals of B and $S + B$ are defined as probability intervals

$$CL_x = P_x(Q \geq Q_{obs}) \,, \tag{7.6}$$

and CL_s is constructed as the ratio:

$$CL_s = \frac{CL_{S+B}}{CL_B} \,. \tag{7.7}$$

Then, the signal hypothesis is regarded as excluded at the confidence level (CL) if:

$$1 - CL_s \leq CL \,. \tag{7.8}$$

Strictly speaking, CLs itself is not a confidence level, which means that it has not the property of being flatly distributed in the limit of infinite number of experiments. Nevertheless, it can be interpreted as an approximation of the confidence in the signal hypothesis which is in practice not possible to obtain directly since one would have to exclude the background with certainty.

The use of CLs prevents the exclusion of signals to which the analysis in not sensitive to by normalizing the S+B confidence to the confidence in the background-only hypothesis. In this sense, the CLs method can be seen as more conservative, but it gives in fact a much closer answer to the question

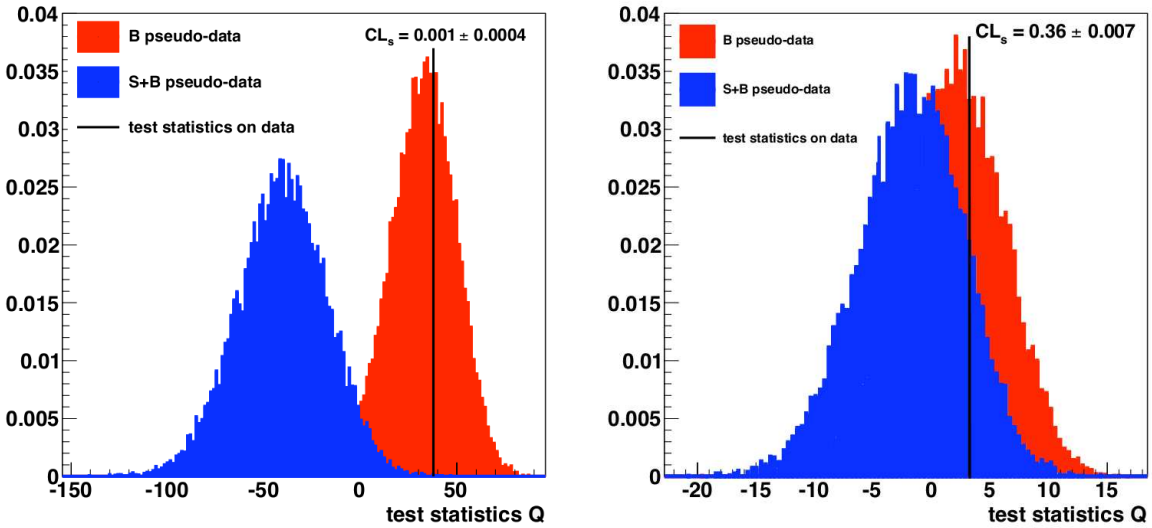


Figure 7.2: Distributions of the test-statistic Q for two different signals in the m_0-$m_{1/2}$ parameter plane. The precision of the stated CLs values is determined by the finite number of pseudo-experiments (here: 20k for each $S + B$ and B).

that an analyst expects: Can the signal be safely excluded (and not: Is the S+B model not in agreement with the data).

For the exclusion limits of the 2010 analysis (presented in the following section) the uncertainties of the individual backgrounds have been combined and symmetrized as it has been described in sec. 7.1. This is possible since the correlation between the systematic uncertainties of the data-driven background estimations is assumed to be negligible.

For the future, the analysis would gain in accuracy if all different sources of uncertainties with their partly asymmetric shapes would be directly incorporated in the calculation via nuisance parameters. Also possible correlations could be accounted for with the above described procedure. On the other hand, a complete integration of the full covariance matrix, though technically possible, would be a huge challenge both from the experimentalist aspect of determining its parameters and also in terms of computing resources needed to deal with the large number of nuisance parameters.

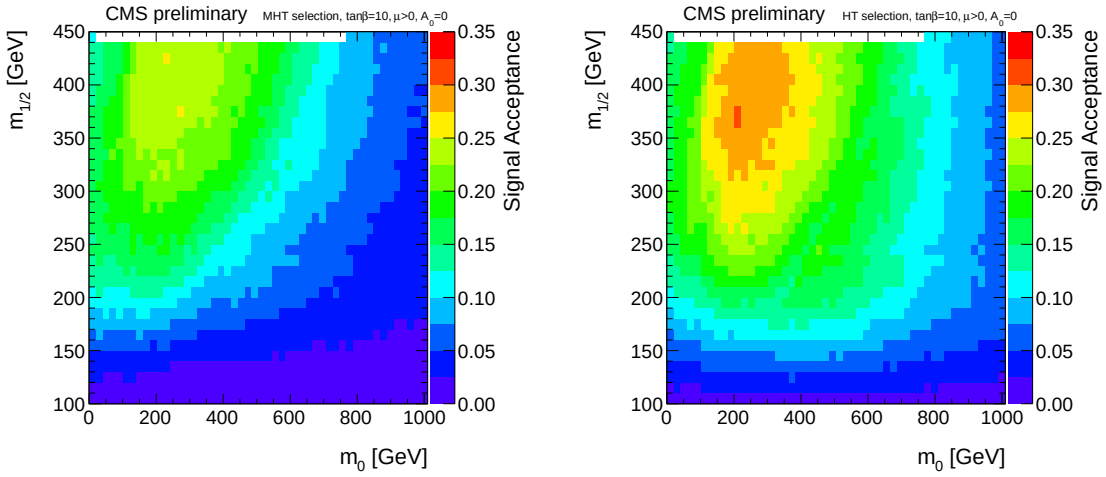
The concept of correlated uncertainties is again needed if the limit calculation is done for different kinematically separated sub-channels of the total search region as it is used in the follow-up analysis with 2011 data [14]. One part of the uncertainties is totally correlated between the different sub-channels (but not between the backgrounds) the other part is not since it comes from statistically independent control regions. For each of the backgrounds, one nuisance parameter is defined for the common systematics and for each sub-channel and each background a nuisance parameter describes the independent uncertainties. The common systematics use scaling factors for the sub-channels in the same way it has been introduced in eq. 7.4.

The implementation of the Hybrid CLs method in the RooStats package has been used.

7.2.3 Interpretation within cMSSM

The event yields used for the limit calculation are taken from tab. 7.1. The signal acceptance varies in the cMSSM phase-space as shown in fig. 7.3. In general, the signal acceptance is $20-30\%$ for the $H_{\mathrm{T}} > 500$ GeV selection and $10-20\%$ for the $\not{H}_{\mathrm{T}} > 250$ GeV selection.

The signal contamination in the isolated muon control-region of the lost-lepton method has been calculated, and removed from the background estimation for each parameter point. For both selections the background event estimate due to signal are $2-3$ events. The signal contamination in the γ-jet control region has been evaluated and was found to be smaller than 0.2 events for the scanned phase-space. Systematic uncertainties of the signal as

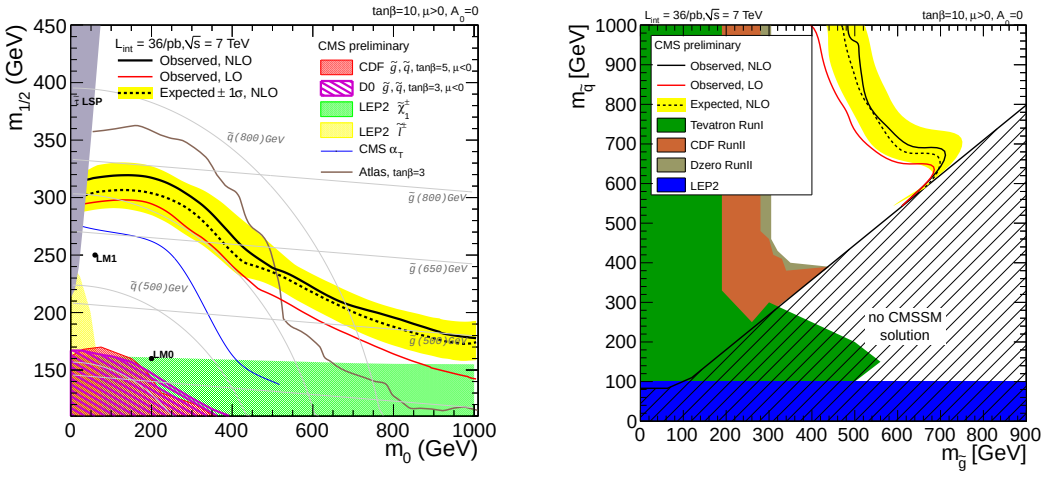


(a) (b)

Figure 7.3: The total signal efficiency (selection efficiency times acceptance) is shown in the cMSSM m_0-$m_{1/2}$ plane for the $\not{H}_T$ selection (a) and the H_T selection (b) . The other cMSSM parameters are $\tan\beta = 10$, $\mu > 0$, and $A_0 = 0$ for both figures. From the paper [1].

summarized in Tab. 7.2 have been considered.

In Figure 7.4 the observed and expected limits for both selections are shown in the cMSSM m_0-$m_{1/2}$ and the squark-gluino mass planes for $\tan\beta = 10$.

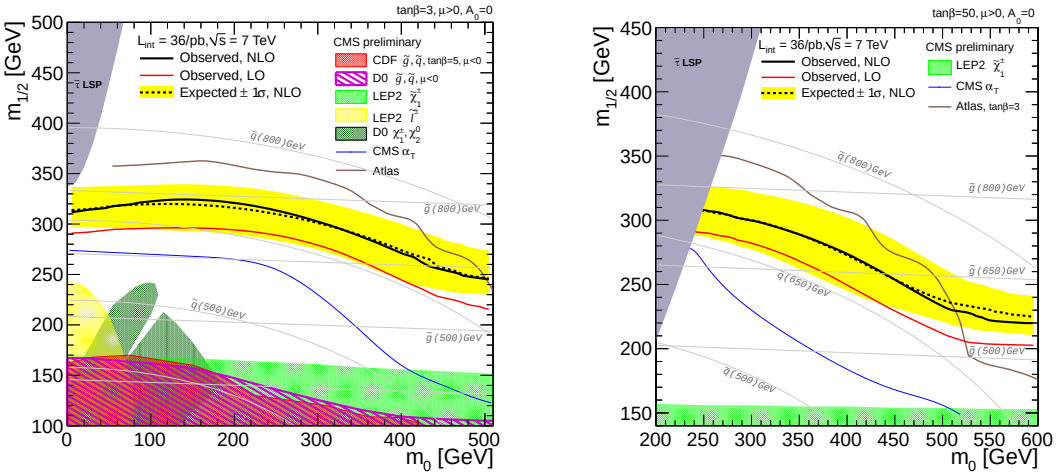


(a) (b)

Figure 7.4: The expected and observed 95% C.L. limits in the cMSSM m_0-$m_{1/2}$ parameter plane are shown in (a) and in the gluino mass–squark mass plane in (b). The yellow 1σ-uncertainty band corresponds to the expected limit. The shown contours are the combination of the H_T and the $\not{H}_T$ selection such that the contours are the envelope with respect to best sensitivity. The other cMSSM parameters are $\tan\beta = 10$, $\mu > 0$, and $A_0 = 0$ for both figures. From the paper [1].

The corresponding limits for $\tan\beta = 3$ and 50 can be found in fig. 7.5. The dependence of the parameter $\tan\beta$ on the exclusion limits is weak which is demonstrated in fig. 7.6 by overlaying the results with the $\not{H}_T > 250$ GeV selection for the three different choices of $\tan\beta$.

Figure 7.6 also shows that using Bayesian upper limit calculation results in very similar 95%CL exclusion contours compared with the Hybrid CLs technique (sec. 7.2.2) which serves as standard method for this analysis.



(a) $\tan\beta = 3$ (b) $\tan\beta = 50$

Figure 7.5: The expected and observed 95%CL limits in the cMSSM parameter plane are shown. The figure (a) corresponds to $\tan\beta = 3$ and (b) corresponds to $\tan\beta = 50$. The remaining cMSSM parameters are $\mu > 0$, and $A_0 = 0$ for both figures. From [45].

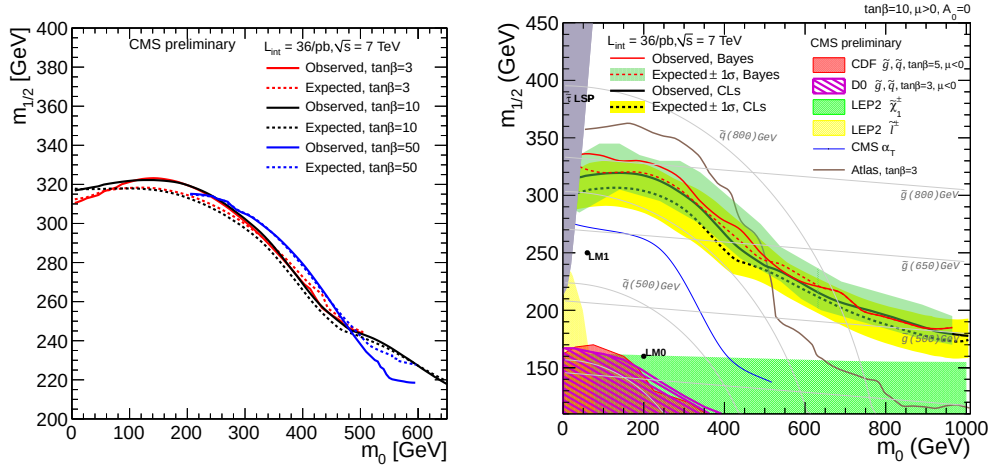


Figure 7.6: Exclusion regions expected (dashed) and observed (solid) at 95% C.L. in the cMSSM m_0-$m_{1/2}$ plane for three different values of $\tan\beta$ with the $\not{H}_T > 250$ GeV selection (a). And a comparison of the exclusion regions for the Hybrid CLs method (standard for this analysis) with Bayesian upper limits for $\tan\beta = 10$ (b). From [45].

7.3 Studying the Search Sensitivity

The two most sensitive variables of the analysis are H_T which corresponds to the total transverse energy in the event and missing transverse momentum $\not{H}_T$. Most of the SUSY signatures tend to have high values whereas the SM background either has less high energetic jets (e.g. $Z \to \nu\bar{\nu}$) or no intrinsic $\not{H}_T$ (like QCD).

An optimal suppression of the SM backgrounds leads to non-optimal signal efficiency in those regions of the m_0-$m_{1/2}$ parameter plane where the signal acceptance is too much reduced. Furthermore, large portions of the background uncertainty are due to the statistical uncertainty of control regions which become more important with higher background suppression.

For these reasons a systematic study of different inclusive search regions applied to the full m_0-$m_{1/2}$ parameter plane including the full systematic uncertainty on the background and the signal is presented in this section.

7.3.1 Variation of the inclusive Search Regions

In this part of the analysis, the 95% C.L. described above is used as a criterion to find the optimal search cuts for future analyses at different points of the m_0-$m_{1/2}$ parameter plane.

The intended variation of the search regions with respect to the two variables H_T and $\not{H}_T$ cannot be accomplished with the $36\,\text{pb}^{-1}$ of 2010 data. Nevertheless, an extrapolation of the data-driven background estimations from the 2010 evolved selections to stricter selections, that can be used with more data, is possible using the shapes of the Monte Carlo distributions of the individual backgrounds. For the new selections, the SM background distributions shown in fig. 7.1 are used which are made from Monte Carlo background samples that are scaled to data-driven background estimations from the 2010 analysis, based on numbers in tab. 7.3.

The new search regions are based on the evolved H_T selection which has been defined for the 2010 analysis. From this new baseline the $\not{H}_T$ (or H_T) cut is increased in steps of 50 GeV (respectively 200 GeV). As an example, the resulting total background expectation and uncertainties are shown for the case of $H_T = 500\,\text{GeV}$ in tab. 7.3.

For a full prediction of the background in $5\,\text{fb}^{-1}$ with the presented data-driven methods, some assumptions on the uncertainties are needed. Firstly, the relative uncertainties that do not depend on the size of control regions are taken to stay the same for future analyses. The statistically dependent uncertainties are each derived from the estimated size of the particular control

region. A detailed treatment of the extrapolation of the event yields and uncertainties for the individual backgrounds is listed below.

Final selection cuts	Total background estimation	Statistical unc.	Systematic unc.
$H_T > 500\,\text{GeV}$ $\not{H}_T > 150\,\text{GeV}$	40.3	9.6%	20.4%
$H_T > 500\,\text{GeV}$ $\not{H}_T > 200\,\text{GeV}$	14.8	18.6%	12.5%
$H_T > 500\,\text{GeV}$ $\not{H}_T > 250\,\text{GeV}$	7.53	27.0%	10.0%
$H_T > 500\,\text{GeV}$ $\not{H}_T > 300\,\text{GeV}$	4.31	35.5%	10.8%
$H_T > 500\,\text{GeV}$ $\not{H}_T > 350\,\text{GeV}$	2.47	46.1%	12.1%
$H_T > 500\,\text{GeV}$ $\not{H}_T > 400\,\text{GeV}$	1.57	56.3%	13.4%

Table 7.3: Predictions for the total background numbers for various new search regions and estimations of the uncertainties at $36\,\text{pb}^{-1}$. The results are extrapolated from the data-driven background estimations partly using the shape of MC simulations as explained in the text.

$Z \to \nu\bar{\nu}$: The events yields are obtained by applying the final search cuts to the Monte Carlo which is preselected for the evolved H_T selection and normalized to the data-driven background prediction with the γ+jets method. The systematic uncertainty is expected to be ~ 20 % and the relative size of the control sample is adjusted for the different evolved selections (high-H_T or high-$\not{H}_T$).

$W/t\bar{t}$: Here, the Monte Carlo is normalized to the sum of the data-driven estimates from the lost-lepton method ($W/t\bar{t} \to e, \mu$+X) and the hadronic τ ($W/t\bar{t} \to \tau_{\text{hadr}}$+X). The systematic uncertainties from the evolved H_T selection are added in quadrature and approximated to 20% for all selections. For the statistical uncertainty the simple assumption is used that the relative sizes of the control samples are roughly the same as for the evolved H_T selection.

QCD: Since the Monte Carlo shape for QCD is not as reliable as for the other backgrounds and on the other hand the control samples have much more statistics, the QCD estimates for all selections are directly derived from the 36 pb^{-1} data using the factorization method. Since the dominant uncertainties arise from the difference of the two enveloping functional forms (sec. 6.2.3) and the extrapolation of the fit uncertainty to the signal region, the two are added in quadrature and taken as total uncertainty for the QCD part.

For this study, the same simulated signal samples in the cMSSM plane with $\tan\beta = 10$ and the corresponding uncertainties are used as before (see sec. 7.2). As a simplification the total background is taken to follow a Gaussian distribution with the estimated event yield as the mean and the total uncertainty (adding in quadrature statistical and systematic) as the width.

Three characteristic points (defined in tab. 7.4) in the m_0-$m_{1/2}$ parameter plane are investigated to cover the regions of interest for setting 95% C.L. limit contours. Each of the three points is used to scan for an optimal $\not{H}_T$ cut. For this purpose, the $\not{H}_T$ is varied in steps of 50 GeV for three different H_T cuts. For each cut setting the expectation on the number of signal events is scaled in an iterative process to find the value of the scaling factor that corresponds to a 95% C.L. exclusion.

	m_0	$m_{1/2}$
Point A	200 GeV	430 GeV
Point B	800 GeV	340 GeV
Point C	1800 GeV	200 GeV

Table 7.4: Definition of the three selected points in the m_0-$m_{1/2}$ parameter plane with $\tan\beta = 10$ for the sensitivity study.

Point A represents the region with low m_0 and high $m_{1/2}$ where the $\not{H}_T$ cut is known to work best, due to a high production rate and high average momentum of invisible particles. Medium values of m_0 and $m_{1/2}$ characterize point B. The ratio of the limit cross section to the signal cross section for these two points are evaluated both for the 36 pb^{-1} and for a ten times higher integrated luminosity (see fig. 7.7).

For scenario A with 36 pb^{-1} the results support the choice of the evolved $\not{H}_T$ selection ($\not{H}_T > 250$ GeV) used in 2010 analysis. The luminosity extrapolation for follow-up analyses shows that the highest $\not{H}_T$ cut is favored and would

lead to an expected exclusion at 95% C.L. for point A. The scenario B exhibits the increasing impact of the H_T cut at higher values of m_0. The evolved $\not{H}_T$ selection of the 2010 analysis is still supported but a small gain in sensitivity for higher H_T cuts is clearly visible.

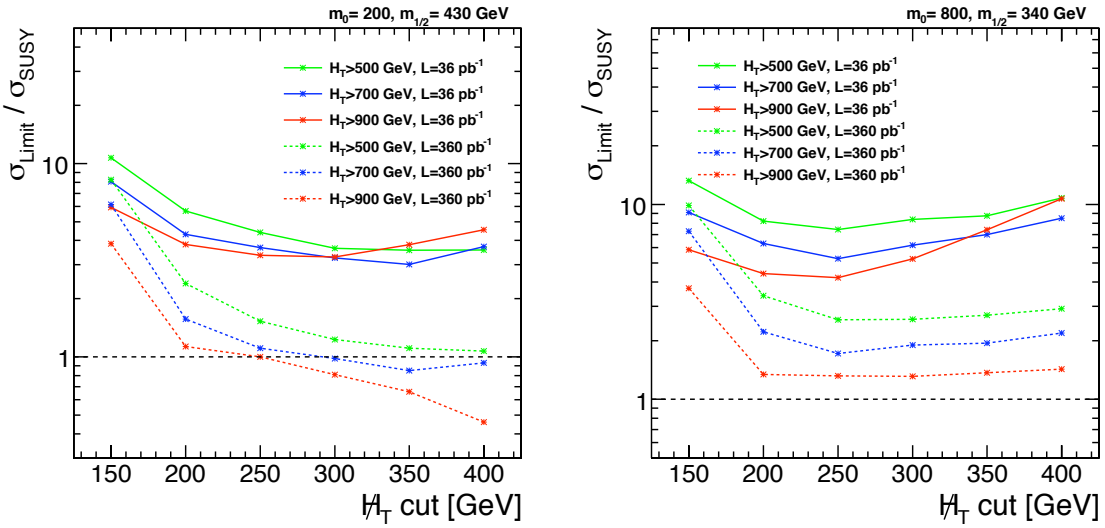


Figure 7.7: The ratio of the limit cross section and the signal cross section for two points of the m_0-$m_{1/2}$ parameter plane, testing different cuts of $\not{H}_T$ and H_T. The background and signal event yields and uncertainties are scaled as described in the text. The $36\,\text{pb}^{-1}$ corresponding to 2010 data is compared to a ten times higher luminosity.

The last scenario is C which represents a high m_0 and low $m_{1/2}$ region. In this region the relative QCD fraction of the total background is important since a high $\not{H}_T$ cut would dramatically reduce the signal efficiency resulting in a rise of the ratio of the limit cross section and signal cross section (see fig. 7.8). From this figure, it is evident that at least to different cut scenarios have to be used in order to get optimal limits in the whole m_0-$m_{1/2}$ parameter plane.

In the following scenario C is used to perform a full prediction of the sensitivity of different cut scenarios with increasing amount of data. Since, here, the $\not{H}_T$ cut does not strongly depend on the assumed integrated luminosity it is possible to fix the cut at $\not{H}_T = 200\,\text{GeV}$ and perform a finer and wider scan of the H_T cut (see right fig. of 7.8).

Three H_T cut scenarios are now tested for increasing integrated luminosity

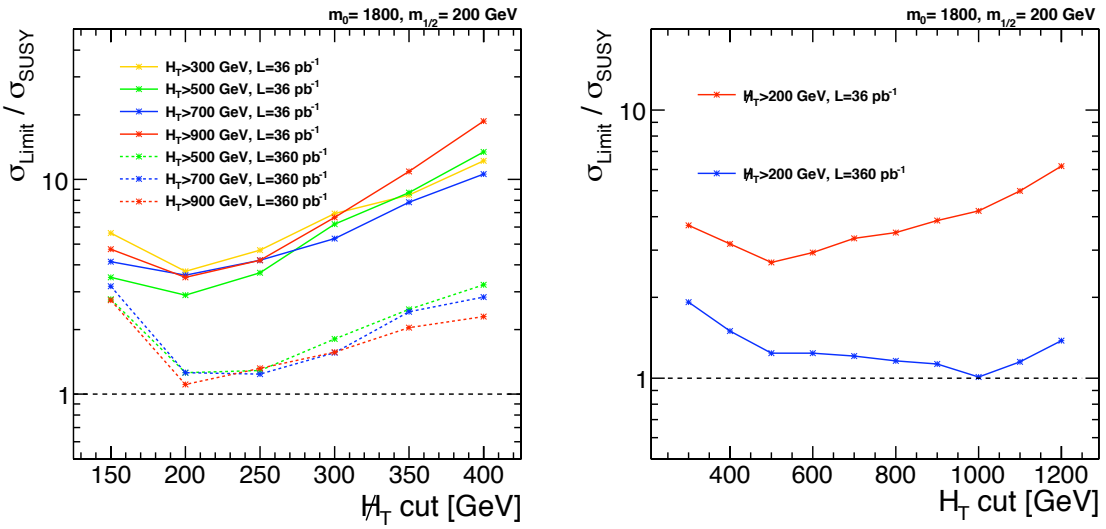


Figure 7.8: The ratio of the limit cross section and the signal cross for an example signal with high m_0 and low $m_{1/2}$. In the left figure the $\not{H}_T$ cut is varied for three H_T cuts and in the right the optimal cut ($\not{H}_T$ =200 GeV) is used for a finer scan of H_T. The 36 pb^{-1} corresponding to 2010 data is compared to a ten times higher luminosity.

(see fig. 7.9). In the presented double logarithmic scale the gain in sensitivity is almost linear before it saturates and the gain in sensitivity becomes smaller. The result that for a given cut scenario a given signal might not be possible to exclude even for an arbitrary amount of data is explained by the systematic uncertainties of the backgrounds which were taken to stay constant over time. While this assumption can turn out to be wrong, the presented method shows an opportunity to find and test cut scenarios for the momentarily available assumptions on the systematic uncertainties. The prediction of a value of the integrated luminosity where the used cut scenario is no longer optimal in terms of sensitivity could clearly help in future analyses.

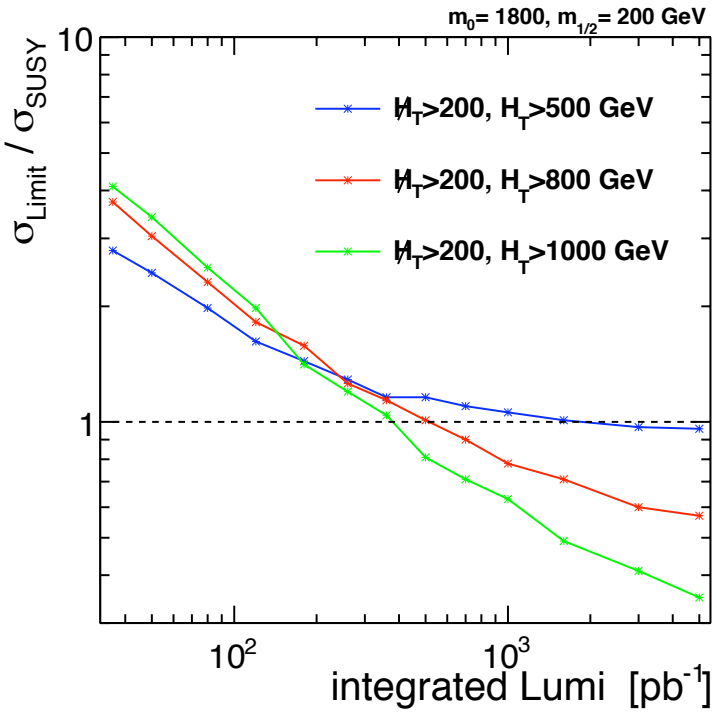


Figure 7.9: The predicted influence of the increasing amount of data on the sensitivity of three cut scenarios for an example signal with high m_0 and low $m_{1/2}$.

The performance of this sensitivity study can be tested by comparing the results of fig. 7.9 with the results from the 2011 analysis with 1.1 fb^{-1} of data shown in fig. 7.10. The figure shows a combination of different selections, but in the high-m_0 region the best sensitivity is provided by a high-H_T selection ($\not{H}_\text{T} > 200\,\text{GeV}$, $H_\text{T} > 800\,\text{GeV}$) which matches the central scenario in fig. 7.9.

The point C ($m_0 = 1800\,\text{GeV}$, $m_{1/2} = 200\,\text{GeV}$) is excluded but close to the 95% C.L. which is in agreement with the prediction. The prediction also suggests that a higher H_T would improve the sensitivity. This has not been pursued by the 2011 analysis but could be reasonable since the high-H_T selects a total of 70 events for $1.1\,\text{fb}^{-1}$.

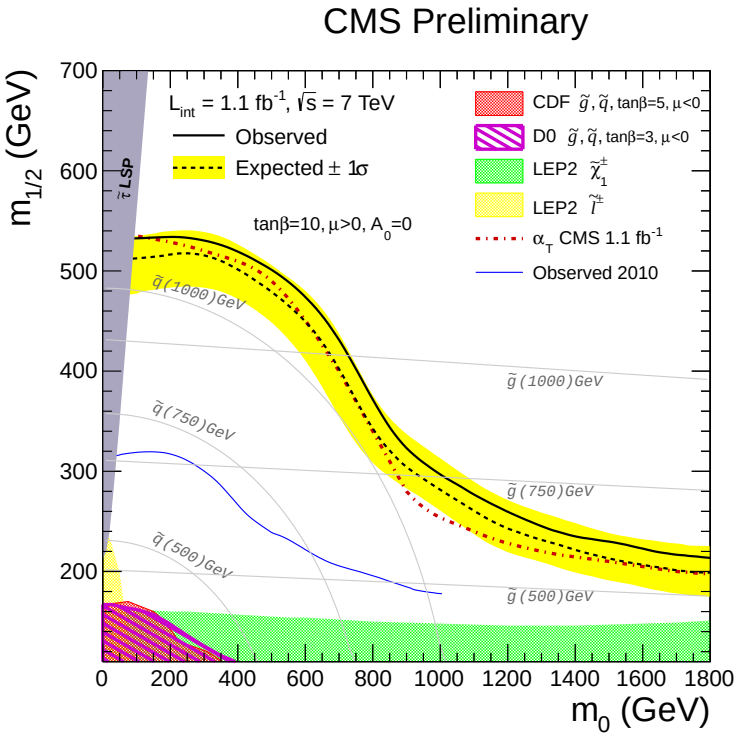


Figure 7.10: The observed and expected 95% C.L. exclusion contours in the cMSSM m_0-$m_{1/2}$ parameter plane obtained by the 2011 analysis [14]. The shown contours are the combination of the different selections, such that the shown contours are the envelope with respect to the best sensitivity.

A further interesting aspect for a signal search would be to predict the regions where a discovery would be possible. The mean of the signal expectations can be used to define a discovery contour in the m_0-$m_{1/2}$ parameter plane where one would expect to find the given signal with a chance of 50% (see fig. 7.11).

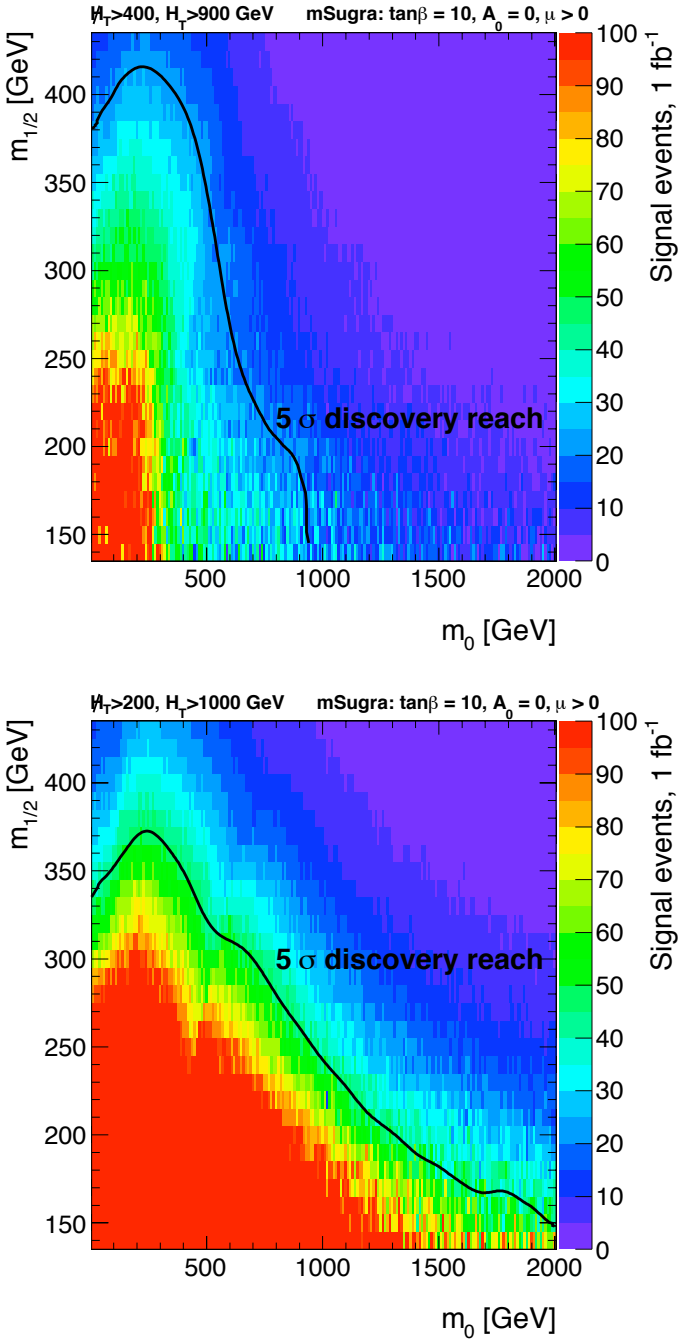


Figure 7.11: Expected discovery contour for two cut scenarios, high $\not{H}_T$ cut (left) and low $\not{H}_T$ cut (right) for the integrated luminosity of 1 fb^{-1}. The background and signal event yields and uncertainties are scaled as described in the text.

8 Summary

This thesis has presented a multijet search for supersymmetry with the first $\sqrt{s} = 7$ TeV pp collision data delivered by the LHC in 2010. The two key search variables have been the missing transverse momentum ($\not{H}_T$) and a measure for the total hadronic energy in the event (H_T). The results of the search have been interpreted as counting experiments in two search regions evolved from the baseline event selection, one with an increased cut on $\not{H}_T$ ($> 250\,\text{GeV}$) and the other with an increased cut on H_T ($> 500\,\text{GeV}$). In the absence of a signal, upper limits on the main parameters (m_0 and $m_{1/2}$) of the widely known cMSSM have been calculated. The results of the search have been published in 2011 [1].

The major challenge of the analysis presented here was the development and application of data-driven methods for the estimation of all contributing SM background processes, which were: $t\bar{t}$, $W + \text{jets}$, $Z \to \nu\bar{\nu} + \text{jets}$ and QCD events. These methods have been accomplished by different members of the CMS collaboration which have worked together on this analysis. This thesis has contributed with the investigation of QCD processes and the development of a data-driven method for estimating the number of remaining QCD events in the defined search regions.

It has been shown that the factorization method, which makes use of the correlation between the two variables $\not{H}_T$ and $\Delta\phi_{\text{min}}$, produces reliable QCD predictions for a number of variations of simulated QCD samples. As a central part of the concept, it is assumed that the cut efficiency of $\Delta\phi_{\text{min}}$ as a function of $\not{H}_T$ can ideally be described by a simple functional form for the ratio $r(\not{H}_T)$ ($\not{H}_T$ distribution of events which pass the cut on $\Delta\phi_{\text{min}}$ divided by the $\not{H}_T$ distribution of events with small $\Delta\phi_{\text{min}} < 0.2$). This is possible, since in QCD processes $\not{H}_T$ is the result of jet mismeasurements and the important sources of jet mismeasurement contribute to the ratio $r(\not{H}_T)$ in a similar way. The QCD prediction is done by measuring $r(\not{H}_T)$ at $70 < \not{H}_T < 120\,\text{GeV}$ and applying the extrapolated function to a QCD-dominated control region of high $\not{H}_T$ and small $\Delta\phi_{\text{min}}$.

The a priori unknown mixture of QCD events with different jet mismeasurement configurations, lead to an inherent model uncertainty which lower and upper limit can best be described by two bracketing models (referred

to as Gaussian resolution model and exponential model). The dominant uncertainties of the models themselves are statistical uncertainties and arise mainly from the error propagation of the fitted ratio $r(\not{H}_T)$ to high $\not{H}_T$ values.

The application to 36 pb^{-1} of data has been a successful test of the factorization method. The data can be described by the same functional form in the region $70 < \not{H}_T < 120\,GeV$ and the fit parameters have similar values compared to the simulation. It has also been found that the QCD prediction is robust against reasonable variations of the fit region and control region boundaries. A completely independent method, the R&S method, produced very similar numbers of estimated QCD events for all three final selections. The total uncertainties of the two data-driven QCD estimation methods are of the same level. For the baseline selection and for the high H_T selection the total uncertainty of the QCD prediction is ~ 50 %. Only for the high $\not{H}_T$ selection, the factorization method has a higher uncertainty of ~ 80 %, but the total QCD prediction is for both method below 1 event and the expected contribution of QCD to the total SM background is ~ 3 %.

For this analysis, it has been decided to employ the R&S as primary method for the final numbers used in the limit calculation and to cross-check the results with the factorization method. Also the succeeding analysis with $1.1\,fb^{-1}$ data [14] proceeded with the R&S method, while there is no principle objection against continuing with the factorization method.

With the rapidly increasing amount of data available for the analysis presented here, the cut on the two key search variables $\not{H}_T$ and H_T have to be increased in order to optimize the sensitivity for supersymmetry searches. It is vital to maintain at least two final selections, since a too high cut on $\not{H}_T$ would reduce the sensitivity in a large region of SUSY parameter space (i.e. the high m_0-region of the cMSSM). Using the results of data-driven SM background predictions and some assumptions on the development of the uncertainties of the methods, the search cuts on $\not{H}_T$ and H_T has been investigated in terms of optimal search sensitivity for future searches. The technique introduced in this thesis could help to define optimized search regions for an expected amount of data well in advance of the statistical interpretation of the final results.

The multi-jet searches for new physics at the LHC look forward to promising next years which could bring the discovery of supersymmetry. The analysis presented here, and to which this thesis could contribute, served as one of many steps that would be necessary on the way to achieve such a goal.

Danksagung

Ich möchte mich ganz herzlich bei Peter Schleper bedanken, der mich auf die Teilchenphysik neugierig gemacht hat, mir diese Doktorarbeit ermöglicht und am Ende auch sichergestellt hat, dass ich sie vollende. Für sein persönliches Engagement bezüglich des Letztgenannten und die stets freundschaftliche Zusammenarbeit und Betreuung, möchte ich mich ausdrücklich bei Christian Sander bedanken.

Ich habe die Zeit in unserer Arbeitsgruppe sehr genossen, was an den vielen netten Menschen gelegen hat mit denen ich zusammen arbeiten und Pause machen durfte.

Bei den weiteren Gutachtern Johannes Haller und Isabell Melzer-Pellmann möchte ich mich für ihre Mühe bedanken.

Mein letzter Dank geht an meine Familie ohne die natürlich gar nichts möglich gewesen wäre und denen ich hoffentlich noch viel von dem zurückgeben kann was sie für mich getan haben.

Bibliography

[1] CMS Collaboration. Search for New Physics with Jets and Missing Transverse Momentum in pp collisions at sqrt(s) = 7 TeV. *JHEP*, 08:155, 2011. doi: 10.1007/JHEP08(2011)155.

[2] W.-M. Yao et al. Review of Particle Physics. *Journal of Physics G*, 33:1+, 2006. URL http://pdg.lbl.gov.

[3] CMS Collaboration. Combined results of searches for the standard model higgs boson in pp collisions at. *Physics Letters B*, 710(1):26 – 48, 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.02.064. URL http://www.sciencedirect.com/science/article/pii/S0370269312002055.

[4] ATLAS Collaboration. Combined search for the standard model higgs boson using up to 4.9 fb^{-1} of pp collision data at with the atlas detector at the lhc. *Physics Letters B*, 710(1):49 – 66, 2012. ISSN 0370-2693. doi: 10.1016/j.physletb.2012.02.044. URL http://www.sciencedirect.com/science/article/pii/S0370269312001852.

[5] J. W. F. Valle. Neutrino physics overview. *J. Phys. Conf. Ser.*, 53:473–505, 2006. doi: 10.1088/1742-6596/53/1/031.

[6] W.N. Cottingham and D.A. Greenwood. An introduction to the standard model of particle physics. 2007.

[7] H. Nishino et al. Search for Proton Decay via $p \to e^+ \pi^0$ and $p \to \mu^+ \pi^0$ in a Large Water Cherenkov Detector. *Physical Review Letters, vol. 102, Issue 14, id. 141801*, 102(14):141801, April 2009. doi: 10.1103/PhysRevLett.102.141801.

[8] Stephen P. Martin. A Supersymmetry Primer. 1997.

[9] L. Pape and D. Treille. Supersymmetry facing experiment: Much ado (already) about nothing (yet). *Rept. Prog. Phys.*, 69:2843–3067, 2006. doi: 10.1088/0034-4885/69/11/R01.

[10] Oleg Brandt. Supersymmetry Searches at the LHC. 2008.

[11] Howard Baer, Vernon Barger, Andre Lessa, and Xerxes Tata. Supersymmetry discovery potential of the LHC at $\sqrt{s}$ = 10 TeV and 14 TeV without and with missing E_T. *JHEP*, 0909:063, 2009. doi: 10.1088/1126-6708/2009/09/063.

[12] S. Abdullin et al. Discovery potential for supersymmetry in CMS. *J.Phys.G*, G28:469, 2002. doi: 10.1088/0954-3899/28/3/401.

[13] Howard Baer, Vernon Barger, Andre Lessa, and Xerxes Tata. Capability of LHC to discover supersymmetry with $\sqrt{s} = 7$ TeV and 1 fb^{-1}. *JHEP*, 1006:102, 2010. doi: 10.1007/JHEP06(2010)102.

[14] CMS Collaboration. Search for supersymmetry in all-hadronic events with missing energy. *CMS-PAS-SUS-11-004*, 2011.

[15] The CMS experiment at the CERN LHC. *JINST*, 0803:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.

[16] (Ed.) Bruning, Oliver S. et al. LHC design report. Vol. I: The LHC main ring. CERN-2004-003-V-1.

[17] (Ed.) Buning, O. et al. LHC Design Report. 2. The LHC infrastructure and general services. CERN-2004-003-V-2.

[18] (Ed.) Benedikt, M., (Ed.) Collier, P., (Ed.) Mertens, V., (Ed.) Poole, J., and (Ed.) Schindl, K. LHC Design Report. 3. The LHC injector chain. CERN-2004-003-V-3.

[19] CMS Collaboration. CMS Luminosity Collision Data, 2010. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults2010.

[20] ATLAS: Detector and physics performance technical design report. Volume 1. . CERN-LHCC-99-14.

[21] ATLAS detector and physics performance. Technical design report. Vol. 2. . CERN-LHCC-99-15.

[22] CMS Collaboration. CMS detector performance and software, physics technical design report. *CERN/LHCC 2006-001 CMS TDR 8.1*, CERN 2006.

[23] CMS Collaboration. Detector Drawings, 2012. URL http://cms.cern.ch/iCMS/.

[24] CMS Collaboration. Tracking and primary vertex results in first 7 tev collisions. *CMS-PAS-TRK-10-005*, 2010.

[25] CMS Collaboration. Performance of muon identification in pp collisions at $\sqrt{s} = 7$ tev. *CMS-PAS-MUO-10-002*, 2010.

[26] CMS Collaboration. Particle flow reconstruction of jets, taus, and met. *CERN-CMS-NOTE-2009-039*, 2009.

[27] CMS Collaboration. Particle-flow commissioning with muons electrons from J/Psi and W events at 7 TeV. *CMS-PAS-PFT-10-003*, 2010.

[28] CMS Collaboration. Search for New Physics with a Mono-Jet and Missing Transverse Energy in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys.Rev.Lett.*, 107: 201804, 2011.

[29] CMS Collaboration. Search for new physics with same-sign isolated dilepton events with jets and missing transverse energy at the LHC. *JHEP*, 1106:077, 2011.

[30] CMS Collaboration. Search for new physics in events with opposite-sign dileptons and missing transverse energy with the CMS experiment. 2012.

[31] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual; v6.420, tune D6T. *JHEP*, 05:026, 2006.

[32] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.

[33] Johan Alwall et al. MadGraph/MadEvent v4: The New Web Generation. *JHEP*, 09:028, 2007.

[34] Geant4 - a simulation toolkit. *Nucl. Inst. Meth. A*, 506:250–303, 2003.

[35] Geant4 - developments and applications. *IEEE 53*, 1:270–278, 2006.

[36] W. Beenakker, R. Hopker, and M. Spira. PROSPINO: A program for the PROduction of Supersymmetric Particles In Next-to-leading Order QCD. 1996.

[37] CMS Collaboration. Inclusive search for new physics at CMS with the jets and missing momentum signature. *CMS Physics Analysis Note (AN-2010/417)*, 2011.

[38] M. Cacciari, G. P. Salam, and G. Soyez. The anti-kt jet clustering algorithm. *JHEP*, 0804:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[39] The CMS Collaboration. Jet energy corrections determination at 7 tev. *CMS-PAS-JME-10-010*, 2010.

[40] The CMS Collaboration. Electron reconstruction and identification at sqrt(s) = 7 tev. 2010.

[41] Riccardo Bellan, Mariarosaria D'Alfonso, Sue Ann Koay, Jia Fu Low, Steven Lowette, Roberto Rossin, Joseph Incandela, Colin Bernet, and Patrick Janot. Tail investigations for the ra2 analysis. 2010.

[42] The CMS Collaboration. Beam Halo Event Identification in CMS using CSCs, ECAL, and HCAL. *CMS-AN-10-111*, 2010.

[43] HCAL performance from first collisions data. *CMS Detector Performance Summary*, DPS-2010/025, 2010.

[44] Hongxuan Liu, Kenichi Hatakeyama, Ulla Gebbert, Konstantinos Theofilatos, and Will Flanagan. Studies on ecal dead and masked channel contributions to high met and mht. 2010.

[45] CMS Collaboration. Search for new physics at CMS with jets and missing momentum. *CMS-PAS-SUS-10-005*, 2011.

[46] Jan Thomsen, Jula Draeger, Christian Autermann, Christian Sander, and Peter Schleper. W and ttbar background estimation for all-hadronic susy searches. 2010.

[47] Riccardo Bellan, Mariarosaria D'Alfonso, Sue Ann Koay, Steven Lowette, Nick McColl, Roberto Rossin, and Joseph Incandela. -driven prediction of the hadronically decaying tau background for the ra2 inclusive hadronic susy search. 2010.

[48] Anwar Bhatti, Mariarosaria D'Alfonso, Kenichi Hatakeyama, Hongxuan Liu, Steven Lowette, Gheorghe Lungu, and Sarah Alam Malik. Estimation of the invisible z background to the susy jets plus missing momentum signature using w plus jet events. 2010.

[49] Anwar Bhatti, Kenichi Hatakeyama, Hongxuan Liu, Gheorghe Lungu, and Sarah Alam Malik. Estimation of the invisible z background to the susy jets plus missing momentum signature using z plus jet events. 2010.

[50] Riccardo Bellan, Mariarosaria D'Alfonso, Sue Ann Koay, Steven Lowette, Nick McColl, Roberto Rossin, and Joseph Incandela. Data-driven prediction of the invisible z background for the ra2 inclusive hadronic susy search. 2010.

[51] CMS Collaboration. Isolated photon reconstruction and identification at $\sqrt{s} = 7$ TeV. *CMS-PAS-EGM-10-006*, 2010.

[52] CMS Collaboration. Photon reconstruction and identification at $\sqrt{s} = 7$ TeV. *CMS-PAS-EGM-10-005*, 2010.

[53] CMS Collaboration. Electron reconstruction and identification at $\sqrt{s} = 7$ TeV. *CMS-PAS-EGM-10-004*, 2010.

[54] C. Autermann, C. Sander, P. Schleper, M. Schroeder, and H. Stadie. Measurement of the jet p_{T} response function in qcd dijet events using an unbinned maximum likelihood method. *CERN-CMS-AN-2010-341*.

[55] Riccardo Bellan, Mariarosaria D'Alfonso, Sue Ann Koay, Steven Lowette, Nick McColl, Roberto Rossin, and Joseph Incandela. Data-driven prediction with the rebalance+smear method of the qcd background for the ra2 inclusive hadronic susy search. 2010.

[56] E. Albayrak, A. Bhatti, D. Elvira, S. Sharma, and M. Zielinski. Measurement of full jet energy resolution using photon+jets events at $\sqrt{s} = 7$ tev. 2010.

[57] Christian Autermann, Christian Sander, Peter Schleper, Matthias Schroder, and Hartmut Stadie. Measurement of the jet p_t response function in qcd dijet events using an unbinned maximum likelihood method. 2010.

[58] B. Efron. *The Jackknife, The Bootstrap and Other Resampling Plans*, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1982.

[59] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys.*, G28:2693–2704, 2002. doi: 10.1088/0954-3899/28/10/313.

[60] Robert D. Cousins and Virgil L. Highland. Incorporating systematic uncertainties into an upper limit. *Nucl. Instrum. Meth.*, A320:331–335, 1992. doi: 10.1016/0168-9002(92)90794-5.

[61] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933. doi: 10.1098/rsta.1933.0009. URL http://rsta.royalsocietypublishing.org/content/231/694-706/289.short.