Truth-level based estimation of the sensitivity to pMSSM models in events with one hard lepton

Truth-level basierte Abschätzung der Sensitivität auf pMSSM Modelle in Ereignissen mit einem harten Lepton

Bacheloararbeit an der Fakultät für Physik
der
Ludwig-Maximilians-Universität München

vorgelegt von

Kilian Adriano Lieret

geboren in Augsburg

Genf, den 10. Juli 2015

Gutachter: Prof. Dr. Dorothee Schaile
Tag der mündlichen Prüfung: 15. September 2015

Abstract

Based on the search for supersymmetry in final states containing one isolated lepton, jets and missing transverse momentum with the proton-proton collision data recorded with the ATLAS detector at a center-of-mass energy $\sqrt{s} = 8$ TeV in 2012, this thesis presents an estimation of the sensitivity to phenomenological MSSM models using the signal shape of truth-level signal samples. The obtained sensitivity estimates are compared to the sensitivity as calculated with MC samples on which a full detector simulation and reconstruction was performed. The agreement is found to be generally low. Several sources of error are ruled out, showing the necessity of a more detailed study of the underlying truth- and reco-level signal models.

Contents

Chapter 1

Introduction

Most of our current knowledge about elementary particle physics is encapsulated in the so called *Standard Model* (SM), a theory that describes all known particles and their interactions with the exception of gravity. The Standard Model has been probed in numerous experiments and is generally considered a very successful theory.[1] However, besides not being a "theory for everything" (as it does not incorporate gravitation), there are a series of shortcomings and unexplained phenomena, which point to physics beyond the standard model. Most notably, astronomical studies concluded that the visible matter can only account for about 5% of the total energy content of the universe.[2] To address this problem, the concept of *dark matter* and *dark energy* is introduced.

An important class of theories that present possible candidates for dark matter are models featuring *supersymmetry* (SUSY), a new symmetry that introduces a new boson for each known fermion and vice-versa.

Searches for supersymmetry are currently conducted at the *Large Hadron Collider* (LHC) at the *European Organization for Nuclear Research* (CERN) in Geneva, Switzerland. Using the data of run 1[1], no significant evidence for supersymmetry could be found.[4] Instead, certain SUSY models and parameters were excluded. The searches will be continued in run 2 (started April 2015), now with a unprecedented energy of 13 TeV (6.5 TeV per beam).[5]

This work focuses on the study of the *phenomenological minimal standard model* (pMSSM, phenomenological MSSM), a supersymmetric extension of the standard model with a comparably small number of 19[6] free parameters[2]. Out of the different analysis strategies that probe for pMSSM, this work is based on the *hard one lepton analysis*, which is limited to collision events with only one hard (i.e. high transverse momentum) electron.

There are two general ways to perform this analysis, either based on *reco-samples* (reco-level analysis) or on *truth-samples* (truth-level analysis). The usage of the former allows for a involved but thorough analysis. As the dimensionality of the considered parameter space is high, computation time quickly becomes one of the limiting resources of the analysis. This motivates the usage of the faster truth-level analysis. In previous hard one lepton studies studies this strategy showed however only low sensitivity to pMSSM models.

In this thesis a truth-level analysis for the hard one lepton case is presented that does (unlike previous studies) also take the signal shape of the signal estimations into consideration.

[1] i.e. the timeframe from the March 2010 (first activation) to February 2013 (shut down for further upgrades)[3]

[2] that is, parameters that are not predicted by the theory itself, but have to be found experimentally

Overview

Chapter 2 presents a quick outline of the Standard Model and its flaws which lead to the concept of supersymmetry. After a short description of the LHC and the ATLAS detector in chapter 3, some general concepts of data analysis in high energy physics are introduced (chapter 4), before focusing more specifically on the hard one lepton case (chapter 5) and turning to the proposed truth level analysis in chapter 6. Finally, a short outlook is given in chapter 7.

Chapter 2

Supersymmetry and the Standard Model

2.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) describes all particles that were observed so far (listed in Fig. 2.1), as well as the strong, weak and electromagnetic interactions between them. Elementary particles can be classified with respect to their spin S as *Fermions* (half integer spin) or *Bosons* (integer spin).

- All fermions of the Standard Model have spin $1/2$. They are subdivided into two types, *Leptons* and *Quarks*, each of which come in three *generations*. Each generation consists of
  - one *charged lepton* of charge -1 (e, μ, τ),
  - one neutral *neutrino* (ν_e, ν_μ, ν_τ),
  - one quark of charge $+2/3$ (u, c, t), sometimes referred to as *up-type quark*, and
  - one quark of charge $-1/3$ (d, s, b), sometimes referred to as *bottom-type quark*.
- Interactions of fermions are described via the exchange of Bosons (therefore also called "force carriers"), some of which can also interact with themselves (see Fig. 2.2).
  - The *gluon* g corresponds to the strong force, which is the binding force of *Hardrons*, composite particles made of quarks (e.g. protons and neutrons). It is massless, but has a very limited effective range due to self interactions.
  - The *photon* γ conveys electromagnetic interactions and couples to all charged particles. It is massless and has infinite range.
  - The weak force is mediated by the *weak bosons*, the neutral Z boson and the charged $W^\pm$ bosons. Because of their mass of 80.4 GeV/c^2 resp. 91.2 GeV/c^2, the mediated interactions have very short range.
  - The *Higgs boson* is a particle corresponding to the *Higgs field* which gives masses to the weak bosons.

2.2 Shortcomings of the SM, motivation for SUSY

Several shortcomings of the SM can be fixed by introducing supersymmetriy:

- *Dark matter*: The 2015 data of the Planck satellite indicates that the universe consists of 69% dark energy, 26% dark matter, and 5% ordinary matter.[2] Supersymmetric extensions

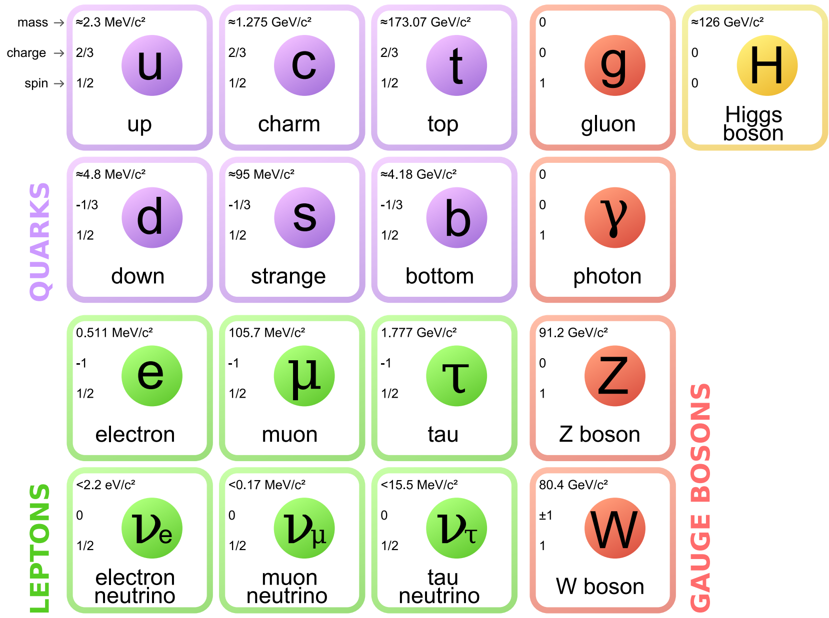


Figure 2.1: The particles of the Standard Model. Figure from [7], using data from [8] and [9] and slightly edited by the author.

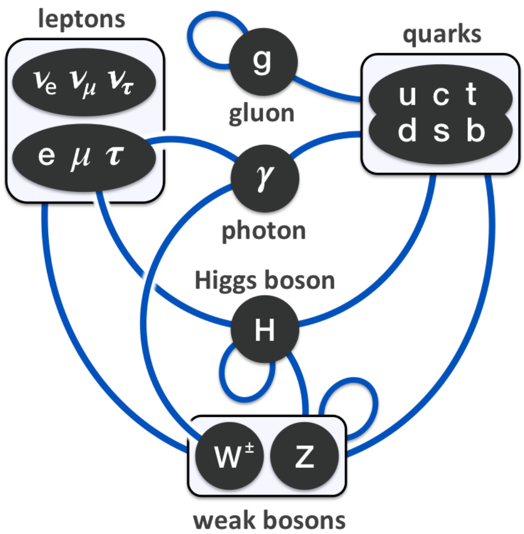


Figure 2.2: Interactions in the Standard Model. Blue lines connect particles to bosons with which they couple; loops represent bosonic self-interaction. Figure from [10].

of the Standard Model imply new particles and feature some promising candidates for dark matter.

- *Unification of gauge couplings*:[1] The *Grand Unification Hypothesis* proposes that all known interactions (electromagnetic, weak and strong) are in fact aspects of one single unique interaction. According to the hypothesis, the gauge symmetry increases with the energy until the strengths of all known interactions become equal at an energy scale Λ_{GUT}. Extrapolation of experimental data and theoretical expectations for the Standard Model shows however that such a unification of coupling constants is not possible with the Standard Model (left plot of Fig. 2.3). In supersymmetric extensions of the Standard Model the unification is possible (right plot of Fig. 2.3).

- *Hierarchy problem*:[2] The squared Higgs mass parameter m_H^2 receives quantum corrections from couplings to virtual particles. Figure 2.4 shows Feynman diagrams of one-loop

[1] based on [12]
[2] based on [1]

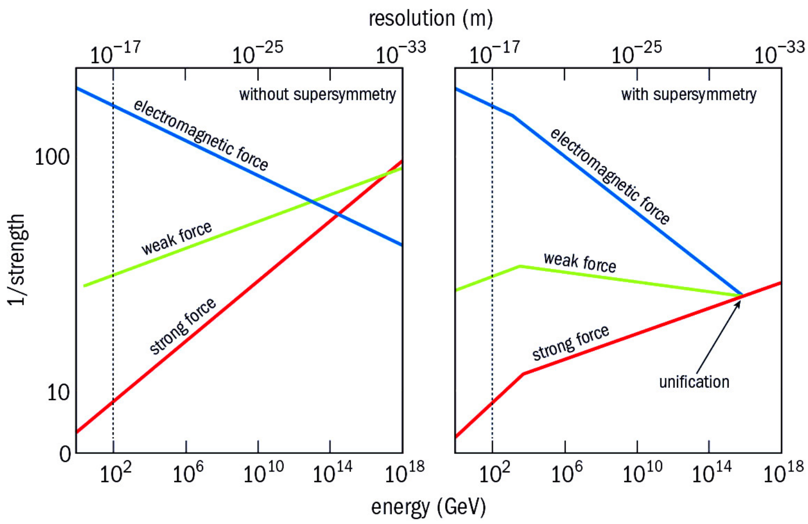
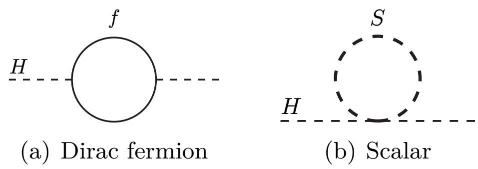


Figure 2.3: Unification of coupling constants. In the standard model, the graph of 1/strength vs the logarithmic energy of the three forces is given by straight lines . The slope of the lines decreases with ascending initial value, yet the the three lines have no common intersection point. In the right plot the influence of supersymmetric particles (with masses at the energy scale denoted by the dashed line) causes the three lines to bend downwards, ensuring the common intersection point. Figure from [11] (coloring modified by the author).

Figure 2.4: One-loop quantum corrections to m_H^2. Figure from [1].



quantum corrections due to coupling fermions and scalars, corresponding to coupling terms $-\lambda_f H\bar{f}f$ resp. $-\lambda_S|H|^2|S|^2$ in the Lagrangian. The corresponding first order corrections are

$$\underset{\text{(Ferm.)}}{\Delta}m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\mathrm{UV}}^2 + \cdots \quad \text{resp.} \quad \underset{\text{(Scalar)}}{\Delta}m_H^2 = +\frac{\lambda_S}{16\pi^2}\Lambda_{\mathrm{UV}}^2 + \cdots, \tag{2.1}$$

where Λ_{UV} is the cutoff value of an ultraviolet momentum cutoff, an add-hock approach to regulate otherwise diverging integrals which is taken to correspond to new physics at the energy scale Λ_{UV}. This interpretation implies that Λ_{UV} must be rather large – else the phenomenons causing the cutoff should have already been observed at the currently accessible energies. However, taking such a Λ_{UV} (say at the Planck scale), the fermionic corrections are of order 30 larger than the value of m_H^2 itself, which makes the small mass of m_H^2 feel “unnatural”.
This issue is called the *Hierarchy problem* and directly motivates supersymmetry as a connection of fermions and scalars: Noting that the corrections of fermions are negative and those of scalars positive, the introduction of new fermions for bosons and new bosons for fermions allows to balance out the sum and yields more “natural” corrections.

2.3 Supersymmetry

The Poincaré group is the group of spacetime isometries of the Minkowsky space $\mathbb{R}^{1,3}$. It describes the symmetries of special relativity, containing translations, rotations and Lorentz-Boosts. The symmetry group of SUSY theories extends the Poincaré group with an operator Q that maps fermions to bosons and vice-versa:

$$Q\,|\text{fermion}\rangle = |\text{boson}'\rangle \qquad \text{and} \qquad Q\,|\text{boson}\rangle = |\text{fermion}'\rangle\,. \tag{2.2}$$

Apart from spin, particles and their supersymmetric partners share all other properties. If SUSY were an unbroken symmetry, this would also apply to the mass of the particles and would have made SUSY particles accessible to previous experiments. As this is not the case, SUSY needs to be a broken symmetry.

2.3.1 Nomenclature

- $|\text{fermion}\rangle$ resp. $|\text{boson}\rangle$ of equation (2.2) are called the corresponding *superpartner* (spartner) of $|\text{boson}'\rangle$ resp. $|\text{fermion}'\rangle$. The term *SUSY particle* (sparticle) is used as a collective term for all superpartners of the SM particles.
- The spartners of the fermions are denoted with a preceding "s": *slepton* (*selectron*, *smu*, *tau*) and *squark* (*sup*, *sdown*, ..., *stop*, *sbottom*). The spartners of the bosons are denoted with the suffix "ino": *gluino*, *wino*, *higgsino* etc.
- In formulas, the spartner of a particle is denoted by a tilde "~", e.g. $\tilde{\ell}$ (slepton), $\tilde{e}$ (selectron), $\tilde{q}$ (squark), $\tilde{B}$ (bino) etc.

2.3.2 The Minimal Supersymmetric Standard Model

The *Minimal Supersymmetric Standard Model* (MSSM) is the supersymmetric extension of the SM with the smallest possible number of sparticles.

2.3.2.1 The sparticles of the MSSM

In the MSSM each of the SM particles has a superpartner with which it shares all quantum numbers with the exception of the spin. The superpartners of the spin-1/2 fermions are spin-0 scalar particles and the superpartners of the spin-1 gauge bosons are spin-1/2 fermions. In the MSSM five Higgs-Bosons exist instead of one: h^0, H^0, A^0 and $H^\pm$.

Note that not all of these spartners are mass eigenstates. Instead, the mixtures of the spartner of the photon (photino), the spartner of the B-boson[3] (Bino) and the spartners of the neutral Higgs-bosons (Higgsinos) form the four *neutralinos* $\tilde{\chi}_1$, $\tilde{\chi}_2$, $\tilde{\chi}_3$ and $\tilde{\chi}_4$ (ordered from light to heavy). Similarly, the spartner of the W-bosons (Wino) and the charged Higgsinos form four *charginos* $\chi_1^\pm$ and $\chi_2^\pm$.

2.3.2.2 R-Parity

General concepts of supersymmetry allow for processes violating the lepton number L and the barion number B.[4] This allows for squark mediated decay modes of the proton such as $p^+ \longrightarrow e^+\pi^0$ which would result in extremely short lifetimes of the proton. To forbid such processes, a new multiplicatively conserved quantum number, called R-parity, is introduced:

$$R = (-1)^{3B+L+2S} = (-1)^{3(B-L)+2S} = \begin{cases} -1 & \text{for sparticles} \\ 1 & \text{for SM particles.} \end{cases} \tag{2.3}$$

Note that this implies that sparticles are always produced in even numbers (usually two) and that a sparticle always decays into an odd number of sparticles!

[3] a mixture of W and Z bosons
[4] section section based on [1]

Variable	Description	#param
$\tan\beta$	ratio of the vacuum expectation values of the two Higgs doublets	1
M_A	pseudoscalar Higgs boson mass	1
μ	Higgs-higgsino mass parameter	1
M_1, M_2, M_3	bino (M_1), wino (M_2) and gluino (M_3) mass parameter	3
$m_{\tilde{q}}$, $m_{\tilde{u}_R}$, $m_{\tilde{d}_R}$	1$^{\text{st}}$ and 2$^{\text{nd}}$ generation squark masses	3
$m_{\tilde{l}}$, $m_{\tilde{e}_R}$	1$^{\text{st}}$ and 2$^{\text{nd}}$ generation slepton masses	2
$m_{\tilde{Q}}$, $m_{\tilde{t}_R}, m_{\tilde{b}_R}$	3$^{\text{rd}}$ generation squark masses	3
$m_{\tilde{L}}$, $m_{\tilde{\tau}_R}$	3$^{\text{rd}}$ generation slepton masses	2
A_t, A_b, A_τ	3$^{\text{rd}}$ generation trilinear couplings	3
		$\sum$ #param = 19

Table 2.1: The remaining 19 parameters of the pMSSM as specified in [6].

2.3.2.3 The Lightest Supersymmetric Particle

As a consequence, the lightest supersymmetric particle (LSP) is stable and a good candidate for dark matter.[5] Every other sparticle must have an odd number of LSPs (usually one) at the end of its decay chain. Neutral, colorless LSPs only interact weakly with SM particles and are therefore a plausible dark matter candidate. Sneutrino LSPs have been excluded by a combination of searches at the LEP and cosmological studies, leaving only the lightest neutralino $\tilde{\chi}_0$ and (for supergravity theories) the gravitino as possible LSPs.[13] This study focuses on models with $\tilde{\chi}_0$ as LSP.

2.3.3 The Phenomenological Minimal Supersymmetric Standard Model

Though introducing the minimal number of sparticles, the MSSM still has 105 free parameters, making phenomenological analyses almost impossible.[6] Therefore, several phenomenologically motivated constraints are imposed on the parameters, reducing their number to only 19:[14][6]

- There is no additional source of CP-violation.
- There are no flavor changing neutral currents. Only minimal flavor violation occurs at the electroweak scale; flavor physics is controlled by the CKM matrix.
- 1$^{\text{st}}$ and 2$^{\text{nd}}$ generation universality: The sfermion masses of the first and second generation are degenerate (i.e. equal).
- The Yukawa couplings and the A-terms for the first and second generation are neglible.

The remaining parameters are shown in table 2.1. The analysis described in this thesis only considers models of a certain subset of the 19-dimensional parameter space; this is further described in section 5.5.

[5] paragraph based on [1].

2.3.4 Sparticle decays

This section lists some possibilities for different sparticles to decay. A more complete account can be found in [1] (also the source of this section). Where applicable, the decay modes are roughly ordered by decreasing decay probability. Depending on the mass parameters of the considered pMSSM model, some of the listen decay channels are kinematically forbidden.[6] As index of a chargino resp. neutralino, the indizes i and j can take the values $1,\ 2$ resp. $1,\ 2,\ 3,\ 4$.

- Squarks: $\tilde{q} \longrightarrow q\tilde{g},\ q\tilde{\chi}_i^{\pm},\ q\tilde{\chi}_i^0$
- Gluinos: $\tilde{g} \longrightarrow q\tilde{q},\ qq\tilde{\chi}_i^0,\ qq'\tilde{\chi}_i^{\pm}$
- Charged Sleptons: $\tilde{\ell} \longrightarrow l\tilde{\chi}_i^{\pm},\ \nu\tilde{\chi}_i^{\pm}$
- Sneutrinos: $\tilde{\nu} \longrightarrow \nu\tilde{\chi}_i^{\pm},\ l\tilde{\chi}_j^{\pm}$
- Neutralinos: $\tilde{\chi}_i^0 \longrightarrow Z\tilde{\chi}_j^0,\ W\tilde{\chi}_j^{\pm},\ h^0\tilde{\chi}_j^0,\ \ell\tilde{\ell},\ \nu\tilde{\nu}$
- Chargino: $\tilde{\chi}_i^{\pm} \longrightarrow W\tilde{\chi}_j^0,\ Z\tilde{\chi}_1^{\pm},\ h^0\tilde{\chi}_1^{\pm},\ \ell\tilde{\nu},\ \nu\tilde{\ell}$

[6] e.g. for the squark decay: $\tilde{q} \longrightarrow q\tilde{g}$ is only possible if $\tilde{q}$ is heavier than $q\tilde{g}$. If this is the case, then then this process will dominate, otherwise $\tilde{q} \longrightarrow q\tilde{\chi}_i^{\pm}$ and $\tilde{q} \longrightarrow q\tilde{\chi}_i^0$ become the important decay channels.

Chapter 3

The ATLAS experiment at the LHC

The LHC (Large Hadron Collider), located near Geneva (Switzerland) is currently the most powerful particle accelerator in the world.[15] Two particle beams circle in a ring with a circumference of 26.7 kilometers and are brought to collision at four interaction points.[16] One of the big experiments of the LHC is ATLAS (A Toroidal LHC Apparatus), which recorded the data for the analysis described in this work.

3.1 The Large Hadron Collider

The LHC can operate in three different modes, either colliding two lead ion beams, one lead ion beam with one proton beam or two proton beams. For the latter, the machine is designed for a center-of-mass (CM) energy of 14 TeV at a luminosity of 10^{34} cm$^{-2}s^{-1}$.[16] Before its shutdown in 2012, the LHC reached a CM energy of only 8 TeV[17] but almost reached its design lumonisty with a peak luminosity of $7.7 \cdot 10^{33}$ cm$^{-2}s^{-1}$[18]. In LHC run 2, the CM energy is increased to 13 TeV[1], and on the 3$^{\mathrm{rd}}$ of June 2015 stable beams were realized for the first time[19].

Seven particle detectors are currently in use at the LHC. The two biggest[20] experiments CMS (Compact Muon Solenoid) and ATLAS employ general-purpose detectors, whereas the remaining five experiments use detectors tailored to the study of specific phenomena:[20]

- ALICE (A Large Ion Collider Experiment) focuses on the study of quark-gluon plasma (a state of matter present shortly after the big bang)[21]
- LHCb (Large Hadron Collider beauty) studies the properties of bottom quarks to probe for differences between matter and antimatter[22]
- TOTEM (Total Elastic and diffractive cross section Measurement) is devoted to precision measurements of proton properties[23]
- LHCf (Large Hadron Collider forward) detects neutral particles which are produced in proton-proton collisions and leave the ATLAS interaction point at extremely low angles[24]
- MOEDAL (Monopole and Exotics Detector at the LHC) searches for magnetic monopoles[25]

The tunnel of the LHC consists of eight straight sections which alternate with eight arcs.[2] The straight sections which hold the interaction points are about 500m long and can host experiments,

[1] to make the superconducting magnets of LHC fit for 14 TeV beams, they need to be "trained", a procedure that was skipped in order to start the operation of LHC earlier. However, this limited the maximum energy to 13 TeV, still slightly below the design energy.[17]

[2] paragraph based on [16].

Name	Year	Energy[a] [GeV]	Length [m]
Linac	1979	0.05	30
PSB	1972	1.4	157
PS	1959	26.0	628
SPS	1976	450.0	6,911
LHC	2008	7,000.0	26,657

[a] Maximal single beam energy (design parameter)

Table 3.1: Some facts about the different accelerators that make up the injection chain of LHC[28]

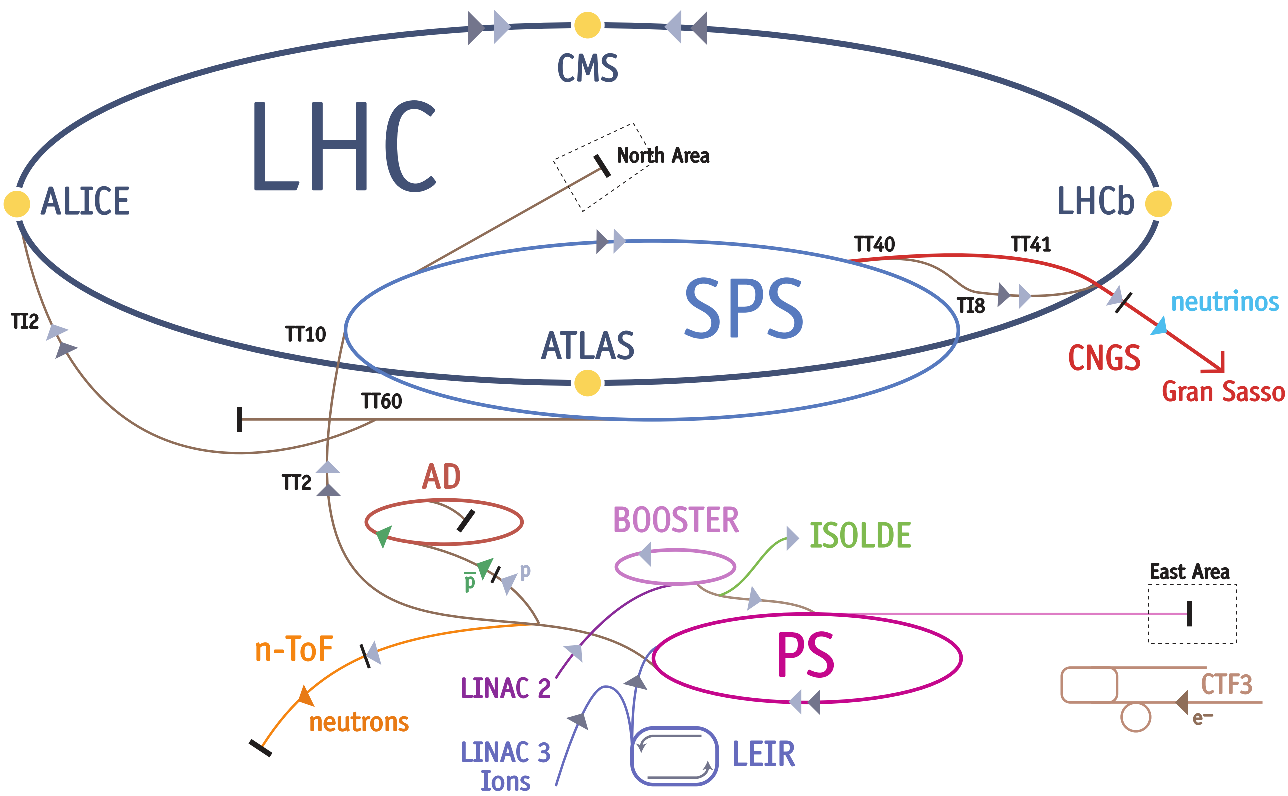


Figure 3.1: An overview of the injection chain and the different experiments at the LHC[26]

beam dumps, injection systems and other utility systems. The experiments use the interaction points 1 (ATLAS), 5 (CMS), 2 (ALICE) and 8 (LHCb).

The LHC features a complex injection chain.[3] Protons are isolated via the ionization of hydrogen and are accelerated to an energy of 50 MeV by the linear accelerator *Linac2*. They are then injected into the *Proton Synchroton Booster* (PS Booster, Booster), which consists of four independent superimposed synchroton rings and raises the energy to 4 GeV.[27] The *Proton Synchroton* (PS) and the *Super Proton Synchroton* (SPS) further accelerate the protons which are then finally injected into the LHC. Once the LHC is filled (which takes about 4 minutes and 20 seconds), the final acceleration to the design energy of 7 TeV begins, taking another 20 minutes.

[3] Paragraph based on [26]

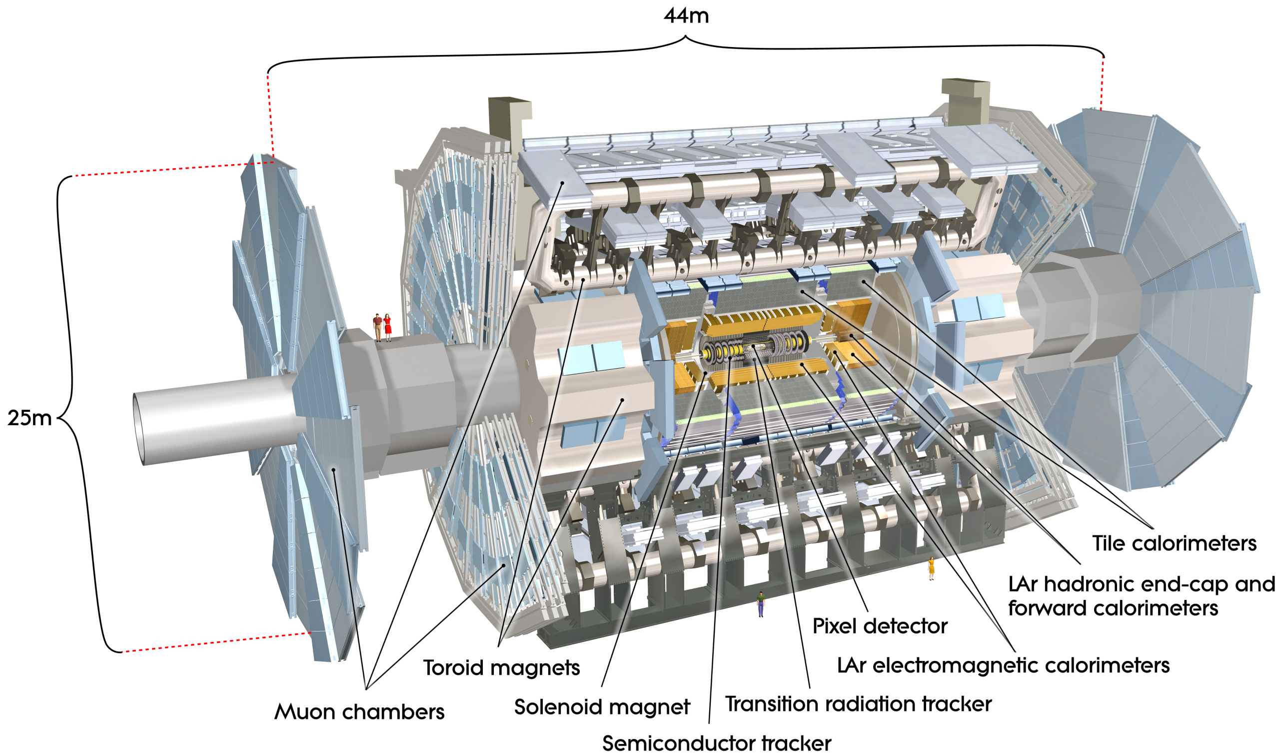


Figure 3.2: Cut-away view of the ATLAS detector[29].

3.2 The ATLAS detector

ATLAS (see Fig. 3.2) is one of the seven particle detectors currently in use at the LHC. Weighting 7,000 tons, it has a length of 45 meters with a height of 25 meters.[30] ATLAS was constructed as a general-purpose detector and – together with CMS – discovered a boson with a mass of $m_H = 125.09 \pm 0.21\ (\text{stat.}) \pm 0.11\ (\text{syst.})$ GeV, which was identified as a Higgs particle.[31] Its other purposes include the search for extra dimensions, dark matter and in particular SUSY particles.[32] With similar applications but independent design, CMS provides a way to cross-confirm the results of ATLAS.[20]

3.2.1 Coordinate system

The following right handed coordinate system is used[33]: The origin is defined to be the interaction point of both beams, the *z-axis* points in beam direction, the *x-axis* points to the center of the LHC ring and the *y-axis* points to the surface.

Looking in the direction of the beam, the *polar angle* θ of a point is set to be the angle between the beam axis and the connecting half-line from origin to the point (i.e. it is measured *from* the beam axis) and the *azimuthal angle* ϕ is the angle between the y-axis and the connecting half-line from origin to the point (i.e. it is measured *around* the beam axis).

The angle of a particle towards the beam axis θ is often described in terms of the spatial coordinate η (*pseudorapidity*):

$$\eta := -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \qquad \Longleftrightarrow \qquad \theta = 2\arctan\left(e^{-\eta}\right). \tag{3.1}$$

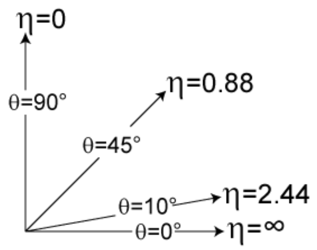


Figure 3.3: For a particle aproaching the beam axis, the pseudorapidity η diverges [34].

The closer a particle gets to the beam axis, the higher the pseudorapidity, approaching infinity for a polar angle θ that tends to zero (cf. Figure 3.3).

As a measure for the distance of two points in the pseudorapidity-azimuthal angle space, the quantity ΔR is used:

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

3.2.2 Detector Layout

This section is based on [33] and [35].

The ATLAS detector shows both forward-backward symmetry (with respect to the interaction point) and an eight-fold azimuthal symmetry (around the beam-line). The subdetectors either take the form of concentric barrels or of disks perpendicular to and centered at the beam line.

- The *inner detector* (Fig. 3.4) is closest to the interaction point. It consists of semiconductor pixel detectors and silicon strip detectors for precision tracking (allowing for a precise momentum and vertex measurement) as well as straw-tube tracking detectors for the measurement of transition radiation. As impulse and charge measurements rely on measuring the curvature of particle tracks in an external magnetic field, a strong magnetic field of 2 T is required inside of the inner detector.

- The *magnetic field* of the ATLAS detector is generated by one solenoid (surrounding the inner detector), one barrel toroid (visible as dominant silver "tubes") and two end-cap toroids (Fig. 3.5).

- The next layer after the inner detector is dedicated to *calorimetry* (energy measurement; Fig. 3.6). Two different types of calorimeters are employed: granularity liquid-argon (LAr) calorimeters and scintillator-tile calorimeters. Both of them are sampling calorimeters[4] and consist of one barrel and two end-caps. The scintillator-tile calorimeter is used as hadronic calorimeter, whereas the LAr calorimeter is primarily used as electromagnetic calorimeter (though its endcaps help out with high η hadronic measurements).

- The outermost detectors are the *muon chambers*. Three of them are assembled to concentric barell-like shapes and six of them (three on each sides) take the form of disks with increasingly large size (called small, big and outer/large wheel). Together, they act as high precision tracking chambers and, together with the strong and far-reaching toroid magnet system, allow for muon momentum measurements.

[4] i.e. consist of two different and alternating materials: one *passive*, absorption heavy material that creates showers and one *active* material that measures the energy deployed in itself[36]

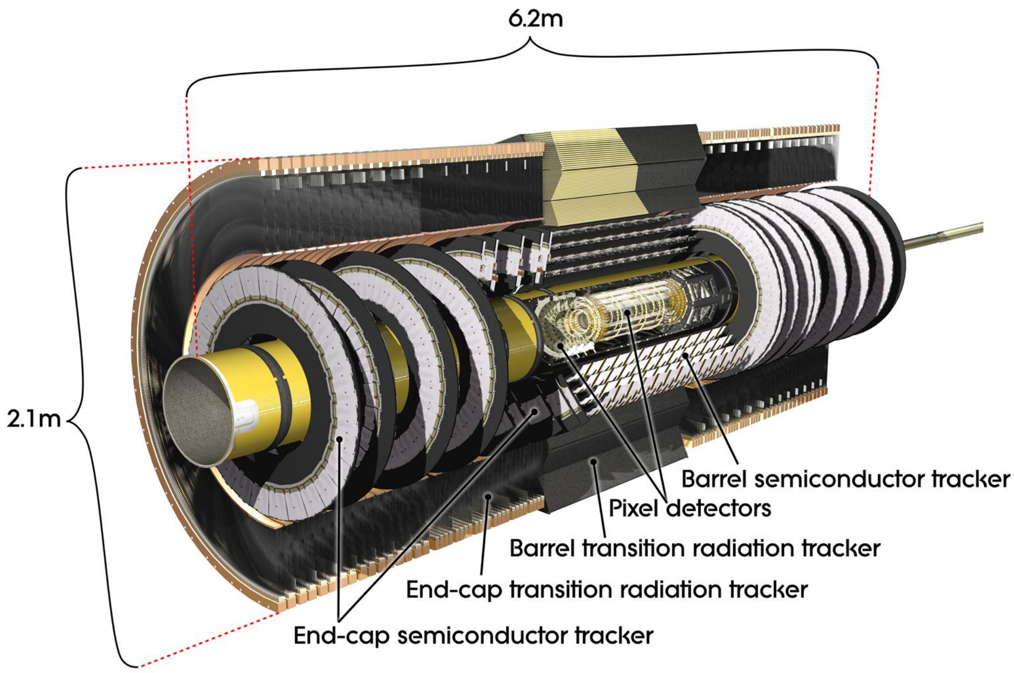
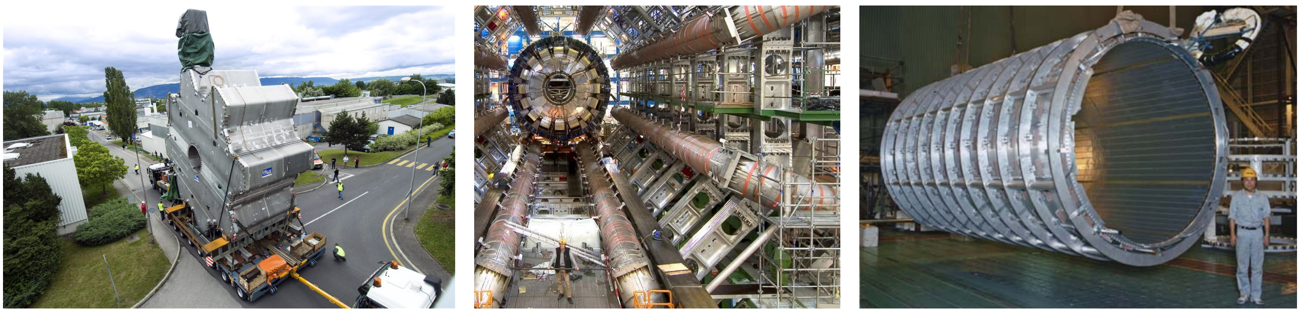


Figure 3.4: The inner detector of the ATLAS experiment [37]



(a) End-cap toroid [38] (b) Barrel toroid [39] (c) Central solenoid [40]

Figure 3.5: The three types of magnets used in the ATLAS experiment

3.3 Data acquistion and processing

Note that the informations of this section are based on the setup of 2012 of and disregards newer developments.

3.3.1 The trigger system

At the design luminosity of 10^{34} cm^{-2} s^{-1}, the bunch crossing rate is 40 MHz with about 25 interactions per bunch crossing.[42] The total amount of 140 million channels of the ATLAS detector produces a data output of order ~ 1 PB/s.[43] However, due to the limited availability of data storage and processing power, only an event rate of 400 Hz could be handled in 2012.[35]

The required rate reduction is achieved by letting *triggers* choose which events are to be recorded and which are to be discarded, filtering out events that are unlikely to be interesting for physics analyses. A random selection of events would make it impossible to study events with production rates smaller then 10^{-5} Hz.[43] Instead, the multi level trigger system sketched in Fig. 3.7 is used.

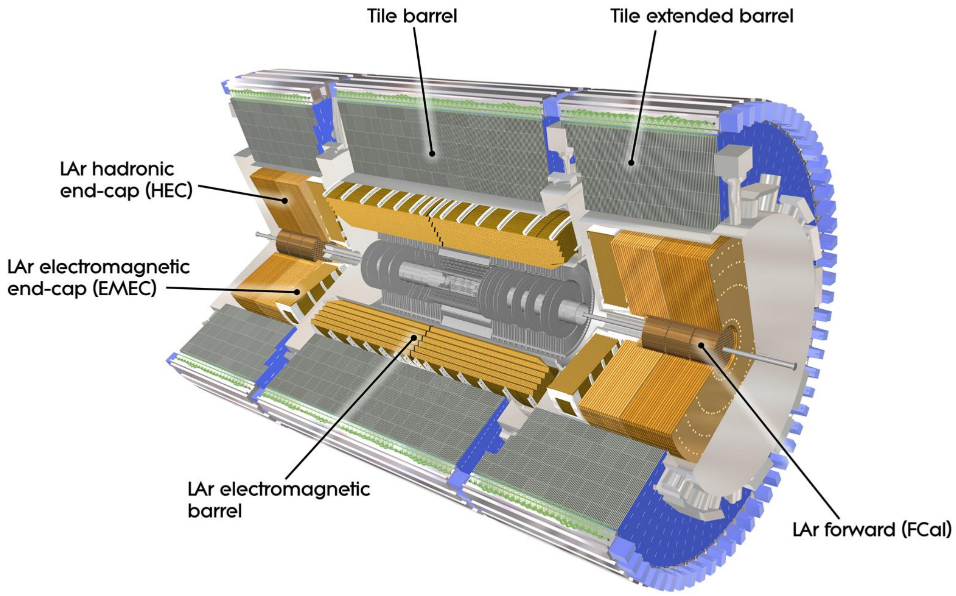


Figure 3.6: Different calorimeters of the ATLAS detector[41]. Note that the colors match those of Fig. 3.2. The inner detector is the gray cylinder (with transverse disks) in the center of the model.

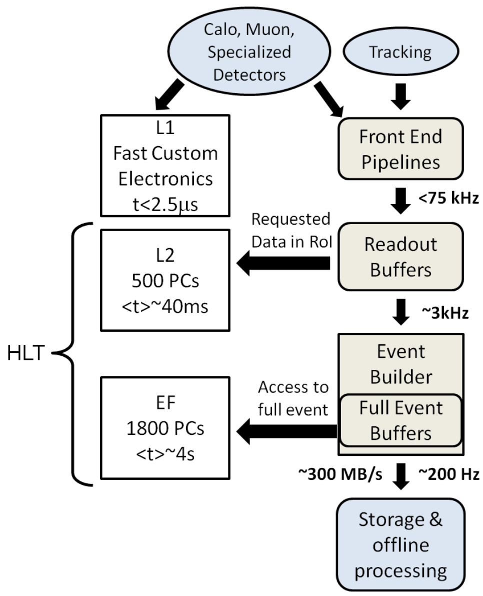


Figure 3.7: The ATLAS trigger system (note that some of the numbers of the figure are out-of-date). Events are filtered through 3 consecutive levels: Level 1 is implemented on specialised electronics to allow for a latency of less than 2.5 μs and reduces the event rate to $<$ 75 kHz; Level 2 and the Event Fiter (EF), together called *high level triggers* (HLT), are handled by a processor farm and reduce the event rate to 3.5 kHz resp. 400 Hz, at which point the events are stored for later ("offline") processing. Figure from [42], updated numbers from [35].

3.3.2 The Grid

The *Worldwide LHC Computing Grid* (WLCG) is the world's largest computing grid and a collaboration of over 170 computer centers accross 41 countries that are joined into a single LHC computing service.[5] The WLCG consists of four levels, called *Tier 0* to *Tier 3*, with Tier 0 being the CERN Data Centre which provides about 20% of WLCG's total computing resources. Tier 1, made up of 13 computer centers, handles a share of the storing and reprocessing tasks, whereas Tier 2, consisting of about 155 sites, performs specific analysis tasks. Finally, the Grid can be accessed by local computing resources (Tier 3).

[5] paragraph based on [44] and [45]

Chapter 4

Data analysis

This chapter introduces general methods and basic terms of data analysis in high energy physics, whereas chapter 5 will look more specifically into the hard one lepton analysis. This chapter is mainly based on [35] and [46], as well as [47] and [48] (describing the different fits in the context of the *HistFitter* software framework[1]).

4.1 Preparing the data for the analysis

The general workflow to compare experimental results with theoretical predictions is sketched in Fig. 4.1. There are two different strategies, *Reco-Level Analysis* and *Truth-Level Analysis*, with the former being more thorough but also more involved.

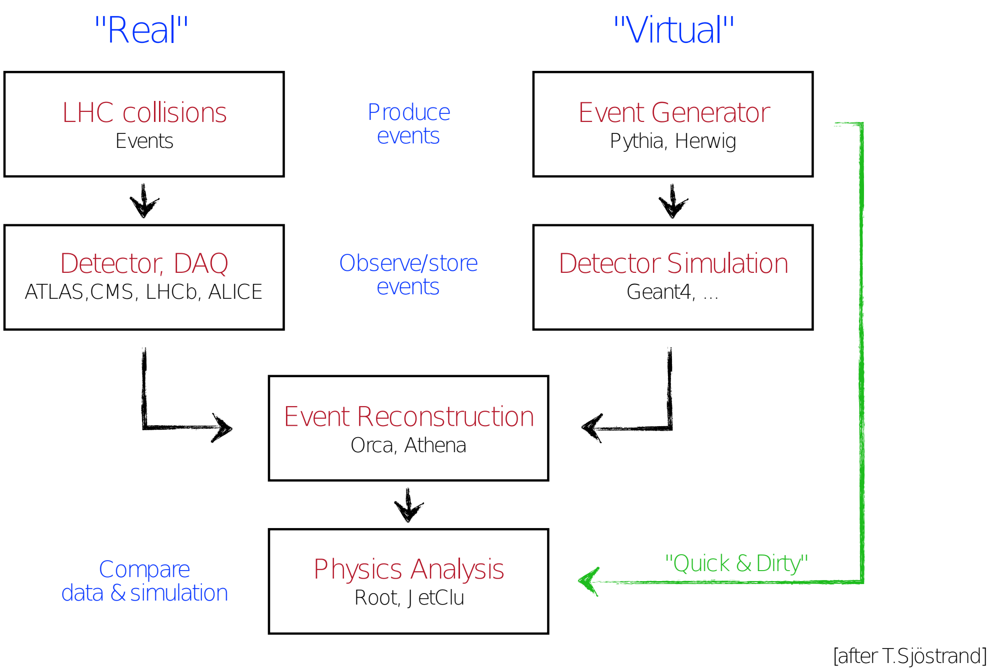


Figure 4.1: General Workflow of Data analysis in high energy physics. On the left side, the processing of real (i.e. measured) data is sketched, the right side summarizes the handling of simulated data. The green arrow represents the "shortcut" taken by the truth-level analysis. Figure taken from [50].

4.1.1 Reco-Level Analysis

Given a high energy physics process whose existence is in question, one starts by simulating its primary vertices. This is done using Monte-Carlo generators (MC/Event generators) that produce simulated events. Then, using another MC generator, dependent "soft physics" processes are simulated, such as initial and final state radiation, showering etc. Since all properties of the event (such as the precise type, energy and momentum of all involved particles) are known, those samples are also called *truth samples*.

The next step is the *detector simulation* that simulates the behavior and the output signals of the different detector components. Now the simulation data is of the same type as the data actually measured at the detector, enabling us to treat both samples equally by feeding them into the *detector/event reconstruction*, which tries to reconstruct the events that gave rise to the measured (resp. simulated) detector data. The reconstructed events of simulation and reality are then compared in the actual physics analysis.

For the truth sample the process of detector simulation followed by detector reconstruction might seem long-winded at first, given that the detector reconstruction tries to revert the detector simulation. Note however that this allows for a much better comparison between data and MC simulation, since all the effects of mismeasurements, experimental uncertainties etc. are applied equally to data and simulation.

4.1.2 Truth-Level Analysis

Detector simulation and reconstruction are not only very complicated but also require a high amount of computation time.[2] As we want to test as many models and parameter sets as possible, computation time quickly becomes one of the limiting resources of the analysis and motivates searches for ways to skip both steps, i.e. to compare the reconstructed detector data directly to the truth samples (green arrow in Fig. 4.1). This is called a *truth-level analysis*.

4.2 Comparing data and simulation

Regardless of whether a truth-level analysis or a reco-level analysis is used – in the end data and simulation have to be compared in a way that results in the confirmation or exclusion of the model that the simulations were based upon.

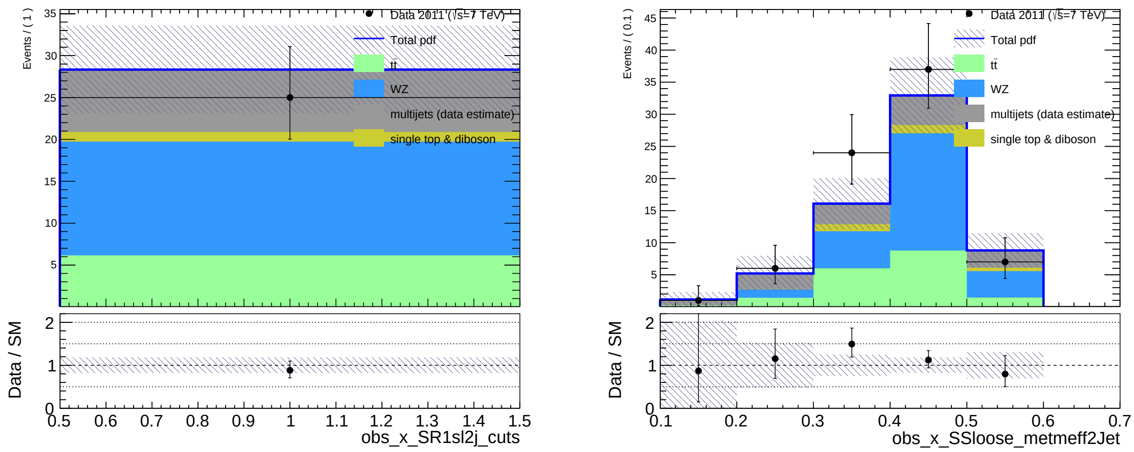
4.2.1 Some Common Notions

It is sensible to introduce some general used terms first:

- *Cuts* are restrictions on the considered events and might for example impose a lower limit on the transversal jet impulse or require a certain number of hard (i.e. high-energy) electrons in the final state.
- When trying to study certain processes/events (or when trying to find a new particle), the main problem is to distinguish those processes from similar looking "uninteresting" ones. The latter are called *background* processes, the former *signal* processes. E.g. for a typical SUSY analysis the background consists of the SM events, whereas the signal events are a subset of events with SUSY particles.

[1] Homepage: [49]

[2] For a typical detector at the LHC, performing the detector simulation of a single hadronic double top quark event with Geant4 can take up to 20 minutes of CPU time (which is a lot, given that some analyses need millions to billions of events).[51]



(a) Cut-and-Count Analysis (b) Shape fit Analysis

Figure 4.2: Comparison of a Cut-and-Count Analysis (left) and a sape fit Analysis (right). The blue line (marked "Total pdf") with diagonally striped error band is the result of the simulation, whereas the black dots represents the data with corresponding error bars. The quotient of Data and simulation is presented as black dots in the bottom plot (marked "Data/SM"). The error bars on the data points correspond to data errors, whereas the diagonally striped error bands are the uncertainties of the simulation. As one can see, data and simulation are compatible in both analyses (though the error bars are certainly used). Plots produced as part of the HistFitter tutorial [48].

- Different sets of cuts are said to define certain *regions* (a term that becomes more natural when picturing regions in phase-space as illustrated in figure 4.3), which are generally classified as signal- control- and validation regions (see section 4.2.3). *Signal regions* are the regions that will be probed for signal processes.

- A *signal model* is a model that quantitatively predicts an excess over background in a signal region. In this thesis the signal models are pMSSM models with fixed parameters.

4.2.2 Cut-And-Count vs Shape Fit

The simplest way to perform an analysis is a "Cut-and-count" study (single bin analysis, Fig. 4.2a), which simply compares the number of events in the signal region between data and simulation. If the measured number of events in the signal region far exceeds the expected number of background events, then this is a hint at new physics. If in contrast the number of simulated events for a certain SUSY model far exceeds the number of measured events, then this specific model can be excluded.

In a "shape fit", the signal region is subdivided into disjoint subregions. In most cases the subdivision is done by "binning" in one variable, i.e. the signal region is split upwith respect to different intervals in this variable – see Fig. 4.2(b) for an exemplary plot. Though oftentimes neither obvious from the plots nor stated explicitly, the last bin normally contains the overflow, i.e. all of the events in the signal region with values larger than the next to last bin.[3] The subregions are then used simultaneously, making the analysis more sensitive to signals with a shape distinct from the shape of the background.[35]

[3] for signal regions that are bounded with respect to the binning variable there might of course not be any overflow.

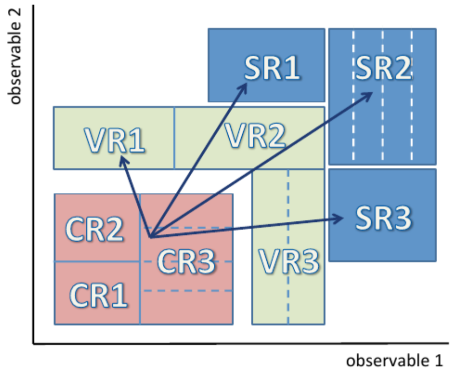


Figure 4.3: Sketch of multiple signal regions (blue), control regions (red) and validation regions (green) in a two dimensional projection of phase-space, spanned by two observables. All regions can contain multiple bins, as indicated by the dashed lines (horizontal resp. vertical lines correspond to binning in observable 2 resp. 1). Arrows symbolise the extrapolation of the background to validation and signal regions. Figure from [52]. See Fig. 5.3 for the actual regions used in the hard one lepton analysis.

Fit	Typical Result	Typical Goal	Used samples	Fit regions
Bkg Only	Bkg	Bkg. Estimation	bkg	CR(s)
Model-dep. Sig.	CL_s, CL_{s+b}	Model-dep. Limit	bkg, sig	CR(s), SR(s)
Model-indep. Sig.	Upper limit μ_{sig}	Model-indep. Limit	bkg, dummy sig	CR(s), SR

Table 4.1: Overview of the different fit configurations (extended version of [47]).

4.2.3 Controlling the background – control, signal and validation regions

While it is possible to simulate most SM background processes, the generated MC samples does not always allow for an accurate estimation of the background. Thus, many SUSY analyses in collider physics use a semi-data-driven strategy.[35]

To do so, it is common to define three different types of regions[47], sketched in Fig. 4.3:

- *Control regions* (CR): Regions that contain one type of background events of high purity with as few signal contamination as possible. Those regions are used to control the background estimation in the signal regions: First, MC simulated event yields are scaled to the actual event yield in the control regions.[4] Using the obtained correction factors on the MC simulated event yields in the signal region, the background yield in the signal regions can be estimated.

- *Signal regions* (SR): As mentioned above, those regions are chosen to contain as much signal processes as possible with only little background.

- *Validation regions* (VR) are used to validate the extrapolation of the background estimation from the control regions to the signal regions. Typically the validation regions are placed "in between" control and signal regions; close enough to the signal regions to provide a check of the estimation, while still only showing small signal contamination.

Simultaneous fits of several regions necessarily require the regions to be disjoint/statistically independent.

4.3 Fitting the data

There are three common fit strategies[47], summarized in Table 4.1:

[4] in a model dependent fit (see section 4.3) signal contamination is taken into account as well

- A *Background Only Fit* yields estimates of the number of background events in the control, validation and signal regions. Only background regions are used (and those are assumed to be completely free of signal contamination) and no signal model is taken as input. The Background Only Fit is commonly used to calculate the background estimates in the signal and validation regions, but it also allows for an easy comparison of the data and independently produced signal predictions.

- The *Model-dependent Signal Fit* ("Exclusion Fit") is used to run signal hypothesis tests and to set exclusion limits or signal upper limits on *specific* signal models. Hypothesis tests will be discussed in a bit more detail in section 4.4. In contrast to the other two fits, the model-dependent fit takes signal contamination in the control regions into account.

- Even without a specific signal model, upper limits for the signal strength of new physics processes can be set, using the *Model-independent Signal Fit* ("Discovery Fit"). The fit strategy itself is similar to a model-dependent fit running on a single bin signal region and using a freely scalable "dummy signal" instead of a specific signal model.

4.4 Hypothesis Tests

When probing for a signal, there are two contrary hypotheses:[5]

1. The event distribution in the signal region consists only of background events
2. The event distribution in the signal region consists of background *and* signal

As a measure for the likeliness of a hypothesis, the p-value is introduced: Under the assumption of a hypothesis, the p-value of an observation is the probability of obtaining the same observation or an observation that is even more extreme (in the sense that it is even more unlikely under the assumption of the the hypothesis) in repeated identical experiments.

Applied to the two assumptions above, the following two p-values are defined:

1. p_b: The probability of having a higher event yield in the signal region than currently observed, even though no signal is present in the signal region
2. p_{s+b}: The probability of having a lower event yield in the signal region than currently observed, even though both signal and background events are present in the signal region

If p_{s+b} is small (less than 5%), then this is generally taken as a hint to exclude the corresponding hypothesis. However this approach is biased if the signal expectation is very low in comparison to the background.[6] To protect against such situations, it is LHC convention to use the Variable CL_s instead:

$$\mathrm{CL}_s = \frac{p_{s+b}}{1 - p_b}. \tag{4.1}$$

A signal model with a CL_s value less than 5% is considered to be excluded. For a simple cut and count analysis, it is easy to calculate p_b, p_{s+b} and CL_s by evaluating integrals over Poisson distributions (see [46] and [35]). At the LHC, more complicated test statistics are being used.

[5] section based on [46] and [35]

[6] consider e.g. a signal model with a vanishing event yield in the signal regions (i.e. a model to which the study clearly is not sensitive). If the number of observed events in the signal regions is much smaller than the background expectation (i.e. if the background shows a downward fluctuation), then the CL_s-value will tend to zero (even though the signal model cannot be included).

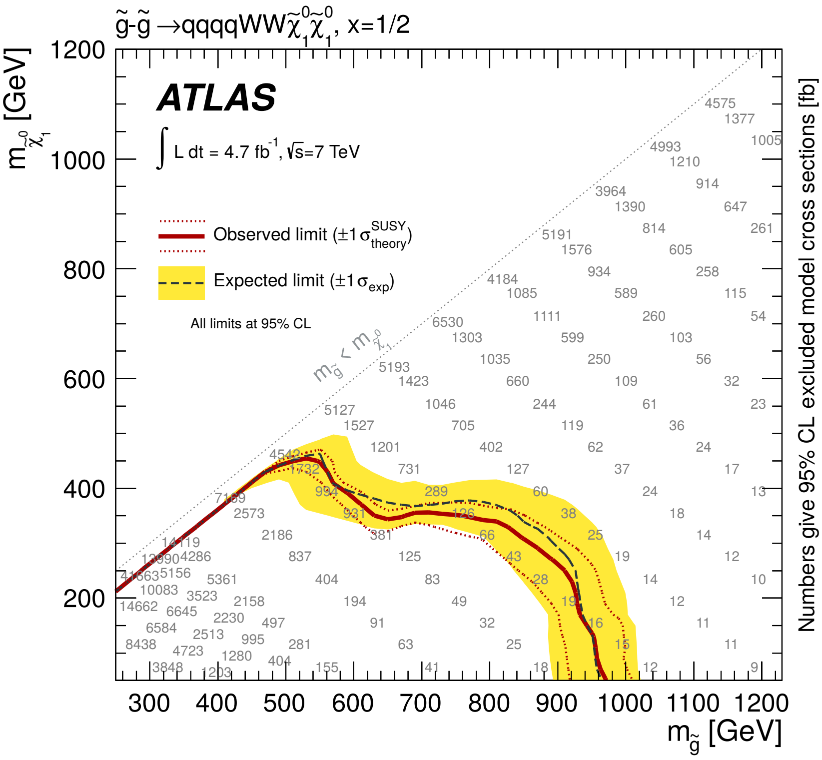


Figure 4.4: Example of an exclusion plot. The strong red (dashed blue) line indicates the observed (expected) limit, including all uncertainties with the exception of the theoretical uncertainties. All models below the observed limit are excluded.
Both exclusion limits are given at a 95% confidence level. The red dotted line (yellow error band) represent the $\pm 1\sigma$ variations of the observed (expected) limit. The gray numbers give the excluded cross section in fb of the probed models, i.e. a smaller number implies a more stringent limit.
The shown parameter space belongs to a simplified model with gluino pair production with the decay channel $\tilde{g} \to q\bar{q}'\chi_1^{\pm} \to q\bar{q}'W^{\pm}\chi_1^0$; the analysis was based on single isolated leptons. The considered decay topology implies $m_{\tilde{g}} < m_{\tilde{\chi}_1^0}$ which is marked with the gray dotted line, excluding the "upper left part" of the parameter space. Plot taken from [53].

4.5 Exclusion Contour Plots

Since most new physics scenarios have free parameters, an analysis will typically need to run signal model hypothesis tests for a whole grid of models and then extrapolate between the grid points to exclude a whole subregion of parameter space. Such analyses can be summarized with a plot like Fig. 4.4.

Chapter 5

The hard one lepton analysis

This thesis is based on a study that searched for supersymmetry in final states with at least one isolated hard electron or muon, jets and large missing transverse momentum with the ATLAS detector (hereafter referred to as *hard one lepton analysis*). The data corresponds to proton-proton collisions at a CM energy of $\sqrt{s} = 8$ TeV and an integrated luminosity of 20 fb^{-1} at the LHC in 2012.

Out of the different sub-analyses, this thesis focuses on the study of pMSSM models in the *hard one lepton* analysis, which requires exactly one hard ($p_\mathrm{T} > 25$ GeV) electron or muon (hereafter shortly referred to as lepton) in the final state.

5.1 Object selection

To cope with some peculiarities of the detector, counteract mismeasurements, exclude effects of pile-ups, cosmic rays, detector noise and other non-collision sources, a number of selection criteria, calibration and correction factors are applied to the objects in the analysis. Not all of them are mentioned in the following list. [54] offers a more complete description and is the source of this section.

5.1.1 Object Preselection

- It is required that the *primary vertex* is consistent with the beam-spot envelope. If more than one vertex fulfills that criterion, the one with the largest summed $|p_\mathrm{T}|$ is chosen.
- *Preselected Jets* must have $|p_\mathrm{T}| > 20$ GeV and $|\eta| < 2.5$.
- *Preselected muons* are required to fulfill $|p_\mathrm{T}| > 10$ GeV, $|\eta| < 2.4$ and $\Delta R(\mu, \mathrm{jet}) > 0.4$. Here and in the rest of this list, *jet* refers to the closest preselected jet.
- *Preselected electrons* have several requirements:
  - $|p_\mathrm{T}| > 10$ GeV and $|\eta| < 2.47$,
  - Due to ambiguity of jets and electrons with small $\Delta R(e, \mathrm{jet})$, the following procedure is applied:

Distance	Electron	Jet
$\Delta R(e, \mathrm{jet}) < 0.2$	kept	discarded
$0.2 < \Delta R(e, \mathrm{jet}) < 0.4$	discarded	kept
$0.4 < \Delta R(e, \mathrm{jet})$	kept	kept

- If there are two preselected electrons with an angular separation $\Delta R(e,e) < 0.05$, the electron with the lower p_{T} is discarded. Furthermore $\Delta R(e,\mu) > 0.01$ for selected electrons and muons is required.
- Finally, electrons with $1.37 < |\eta| < 1.52$ (the so called *crack region*, the transition region between the barrel and the end-cap electromagnetic calorimeters) are discarded.

5.1.2 Signal object selection

For an object to be considered a *signal object* (i.e. to be allowed to constitute a signal event), tighter criteria have to be fulfilled (from [54]):

- For electrons to be *Signal Electrons*, it is required that
  - the p_{T} of the electron is at least 25 GeV,
  - the sum of the $|p_{\mathrm{T}}|$ of tracks within a cone of size $\Delta R = 0.2$ around the electron (but excluding the electron itself), amounts to less than 10% of the electron $|p_{\mathrm{T}}|$,
  - $|z_0| \leq 2$ mm, with z_0 the longitudinal impact parameter with respect to the primary vertex,
  - $|d_0 \leq 1 \text{ mm}|$, with d_0 the distance in the transverse plane of the closest approach of the electron to the primary vertex.
- For preselected muons to be *Signal Muons*, it is required that
  - the p_{T} of the muon is at least 25 GeV,
  - the scalar sum of the p_{T} of tracks within a cone of size $\Delta R = 0.2$ is smaller than 1.8 GeV.
- For preselected jets to be *Signal Jets*, it is required that
  - Every signal jet has to fulfill $p_{\mathrm{T}} > 25$ GeV.
  - To remove jets from pile-ups, the *Jet-Vertex-Fraction cut* (JVF cut) is applied to all non-b-tagged jets with $|\eta| > 2.4$ and $p_{\mathrm{T}} \leq 50$ GeV, demanding that at least 25% of the sum of the $|p_{\mathrm{T}}|$ of all tracks associated with the jet comes from tracks associated with the primary vertex in the event.

5.1.3 Event selection

Events that pass the trigger system are still rejected, if they

- do not have a primary vertex with at least five associated tracks with $p_{\mathrm{T}} > 400$ MeV,
- contain preselected jets that fail to fulfill the quality requirements used to remove detector noise and non-collision backgrounds,
- contain a preselected muon with $|z_0| > 1.0$ mm and $|d_0| > 0.2$ mm (used to suppress cosmic-ray muons).

5.2 Variables

The hard one lepton analysis uses control- validation- and signal regions as described in chapter 4.2.3. The following list introduces the kinematic variables that are used to define the corresponding cuts and selection criteria:

- *Jet Multiplicity* N_{jet}: The number of all signal jets in an event.
- *Missing transverse energy* E_T^{miss} (also denoted $\not{E}_T$): Not all particles can be detected by ATLAS. In particular, neither neutrinos nor the LSP are recorded/reconstructed because they do not interact with the detector. This leads to "missing" Energy/Impulse (see Fig. 5.1).
However, using that – due to momentum conservation – the sum of the transverse[1] momenta of all particles must be zero (as none of the protons brought to collision can have non-vanishing transverse momentum), the sum of the transverse momenta of all non-detected particles, $\vec{p}_T^{\text{miss}}$ can be calculated:

$$\vec{p}_{\mathrm{T}}^{\text{miss}} = -\sum_{\substack{\text{reconstructed}\\ \text{particles}}} \vec{p}_{\mathrm{T}}. \tag{5.1}$$

The missing transverse energy $E_{\mathrm{T}}^{\text{miss}}$ is defined as the absolute value of $\vec{p}_{\mathrm{T}}^{\text{miss}}$:

$$E_{\mathrm{T}}^{\text{miss}} = |\vec{p}_{\mathrm{T}}^{\text{miss}}| = \Big| -\sum_{\substack{\text{reconstructed}\\ \text{particles}}} \vec{p}_{\mathrm{T}} \Big|. \tag{5.2}$$

- *Transverse mass* m_T:[2] This variable was used extensively in studies of the W Boson Mass via the decay channel $W \longrightarrow e\nu$ (e.g. [56]). Since neither the W Boson nor the electron can be detected, the mass of the W Boson cannot be calculated directly. Instead the variable $m_e xtT$ is used:

$$m_{\mathrm{T}}^2 = \left(E_{\mathrm{T}}^e + E_{\mathrm{T}}^{\text{miss}}\right)^2 - \left(\vec{p}_{\mathrm{T}}^{\,e} + \vec{p}_{\mathrm{T}}^{\text{miss}}\right)^2, \tag{5.3}$$

where the transverse energy E_{T}^e is defined as $\left(E_{\mathrm{T}}^e\right)^2 = \left(m^e\right)^2 + \left(\vec{p}_{\mathrm{T}}^{\,e}\right)^2$.[3] Squaring out equation (5.3) yields:

$$m_{\mathrm{T}}^2 = \left(m^\ell\right)^2 + \left(m^{\text{miss}}\right)^2 + 2\left(E_{\mathrm{T}}^e E_{\mathrm{T}}^{\text{miss}} - \vec{p}_{\mathrm{T}}^{\,e} \cdot \vec{p}_{\mathrm{T}}^{\text{miss}}\right). \tag{5.5}$$

Assuming that the rest masses of the electron and the neutrino are almost zero, that the neutrino is responsible for the whole missing transverse energy and using that without mass $E_{\mathrm{T}}^{e/\text{miss}} = |\vec{p}_{\mathrm{T}}^{\,e/\text{miss}}|$, we get:

$$m_{\mathrm{T}} = \sqrt{2E_{\mathrm{T}}^{\text{miss}}|\vec{p}_{\mathrm{T}}^{\,e}|\left(1 - \cos\phi\right)}, \tag{5.6}$$

[1] Using the coordinate system of section 3.2.1, the suffix transverse means that the corresponding quantity is projected into the x-y-plane.
[2] derivation and definition of m_{T} similar to [55], equation (5.6) can also be found in [54] and [35]
[3] Note that, if the measured quantities really do come from a $W \longrightarrow e\nu$ process, the following inequality holds:

$$m_{\mathrm{T}}^2 = \left(E_{\mathrm{T}}^e + E_{\mathrm{T}}^{\text{miss}}\right)^2 - \left(\vec{p}_{\mathrm{T}}^{\,e} + \vec{p}_{\mathrm{T}}^{\text{miss}}\right)^2 = E_Z^2 - \left(\vec{p}_{\mathrm{T}}^{\,W}\right)^2 \geq E_W^2 - \left(\vec{p}^{\,Z}\right)^2 = m_W, \tag{5.4}$$

with m_W the invariant mass of the W Boson. This means that, if all E_T^{miss} comes from $W \longrightarrow e\nu$ processes, the m_{T} will show a clear starting point at m_W. This fact is used to suppress W+jets background (see section 5.3).

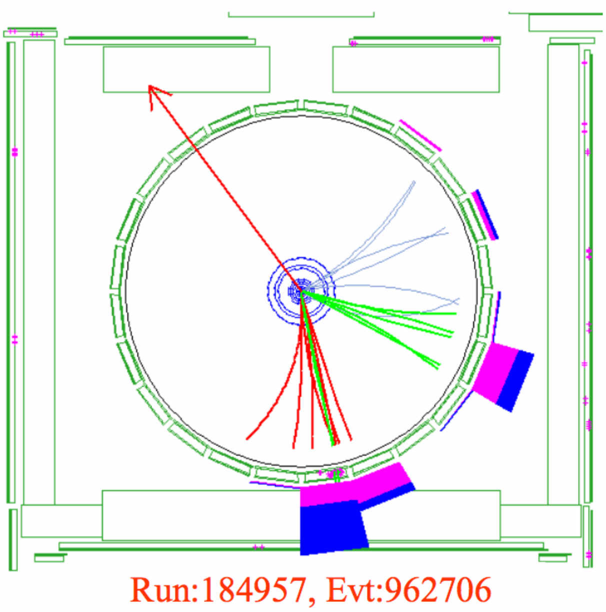


Figure 5.1: Missing transverse impulse at a proton-antiproton collision event, taken from [57]. The slightly bent lines represent reconstructed particles. The arrow points in the direction of the missing transverse impulse.

with ϕ being the angle in the transverse plane between the electron and the missing transverse momentum, $\phi = \phi(e) - \phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$. m_extT is also analogously used in the one hard lepton analysis (with an arbitrary lepton instead of the electron). The characteristic distribution of the m_{T} stemming from W decays and the fact that SUSY processes generally show high m_{T} can be used to surpress W-bosonic background[35][54].

- *Inclusive effective mass m_{eff}^{incl}*: The inclusive effective mass is the sum of the missing transverse energy, the transverse impulse of the lepton and the sum of the transverse impulses of the jets:[54]

$$m_{\mathrm{eff}}^{\mathrm{incl}} = E_{\mathrm{T}}^{\mathrm{miss}} + p_{\mathrm{T}}^{\ell} + \sum_{j=1}^{N_{\mathrm{jet}}} p_{\mathrm{T},\,j}. \tag{5.7}$$

- *Truncated effective mass m_{eff}^{excl}*: Calculated identically to the *inclusive* effective mass (5.7) but only considers the n jets with highest p_{T}[54]:

$$m_{\mathrm{eff}}^{\mathrm{excl}} = E_{\mathrm{T}}^{\mathrm{miss}} + p_{\mathrm{T}}^{\ell} + \sum_{j=1}^{3} p_{\mathrm{T},\,j}, \tag{5.8}$$

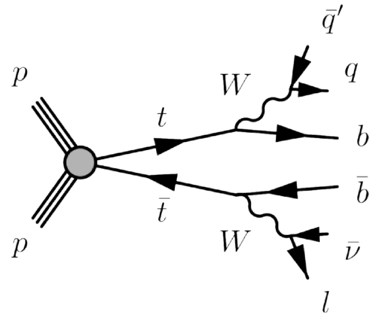
under the common assumption that jets are numbered with decreasing p_{T}, i.e. $p_{\mathrm{T},\,1} \geq p_{\mathrm{T},\,2} \geq \cdots$. In the hard one lepton analysis, we set $n = 3$. This variable is however not directly used in the cuts in favor of the ratio $E_{\mathrm{T}}^{\mathrm{miss}}/m_{\mathrm{eff}}^{\mathrm{ecl}}$.

5.3 Dominant Backgrounds

The tree most important backgrounds of the hard one lepton analysis are $t\bar{t}$ decays, W+jets processes and QCD processes:

- $t\bar{t}$ decays are responsible for the largest background in the signal regions of the hard one lepton analysis. An example is given in figure 5.2: both t quarks decay into W bosons and b- quarks (which lead to b-jets). With one W boson decaying in a lepton and the corresponding neutrino and the other one decaying into two quarks, at least four jets are found, one lepton and E_T^{miss} (caused by the neutrino).

Figure 5.2: A $t\bar{t}$ decay with one lepton, at least four jet and missing transverse energy, therefore contributing to the SM background in the signal regions of the hard one lepton analysis. Figure from [35].



Variable	3-jet (`SR3J`)	5-jet (`SR5J`)	6-jet (`SR6J`)
N_ℓ	1 e/μ	1 e/μ	1 e/μ
$p_{\mathrm{T}}^{\text{first } \ell}$	> 25	> 25	> 25
N_{jet}	≥ 3	≥ 5	≥ 6
$p_{\mathrm{T}}^{\text{jet } 5/6}$	$p_{\mathrm{T}}^{\text{jet } 5} < 40$ GeV	$p_{\mathrm{T}}^{\text{jet } 6} < 40$ GeV	—
$p_{\mathrm{T}}^{\text{jet } i}$	$> 80, 80, 30$	$> 80, 50, 40, 40, 40$	$> 80, 50, 40, 40, 40, 40$
$E_{\mathrm{T}}^{\text{miss}}$	> 300	> 300	> 250
m_{T}	> 150	> 150	> 150
$E_{\mathrm{T}}^{\text{miss}}/m_{\text{eff}}^{\text{ecl}}$	> 0.3	—	—
$m_{\text{eff}}^{\text{inc}}$	> 800	> 800	> 600

Table 5.1: Cuts that define the (exclusion) signal regions SR3J, SR5J and SR6J of the 3, 5 and 6 jet region of the hard one lepton analysis for a shape fit (for a single bin analysis, slightly different criteria are chosen, cf. [54]). The regions are Illustrated in Fig. 5.3. Note that all three signal regions are disjoint due to the required upper bounds for $p_{\mathrm{T}}^{\text{jet } 5/6}$ in SR3J and SR5J.
Units (if any) are GeV. The variables $p_{\mathrm{T}}^{\text{jet } i}$ refers to the p_{T} of the i^{th} jet (order of descending p_{T} assumed). All variables in this table are calculated with *signal* objects. Table similar to [54] (note that in this reference, the requirement $N_\ell = 1$ is implemented via $p_{\mathrm{T}}^{\text{second } \ell} < 10$ GeV).

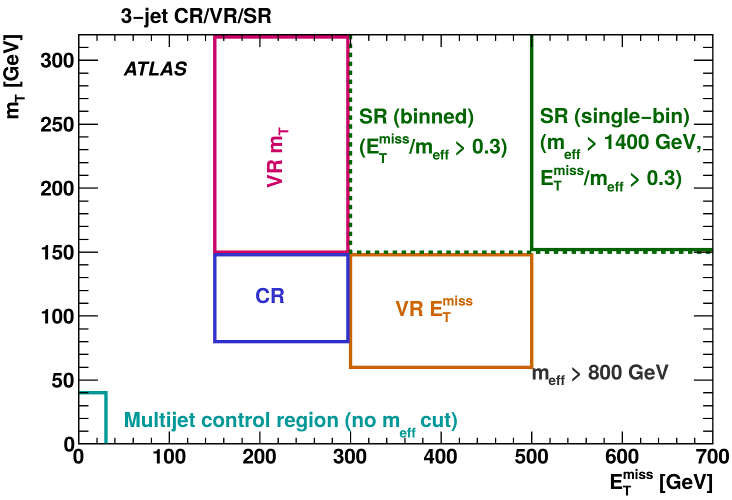
- The production of a W-boson together with jets is responsible for the W+jets background: If the W-boson decays into a charged lepton and a neutrino, events with jets, missing transverse energy and one isolated hard lepton can occur, having a similar signature than the signal.

- If a lepton is created in the decay of a heavy quark (especially a bottom quark) or if a jet is misidentified as a lepton, then QCD processes can contribute to the background as well.

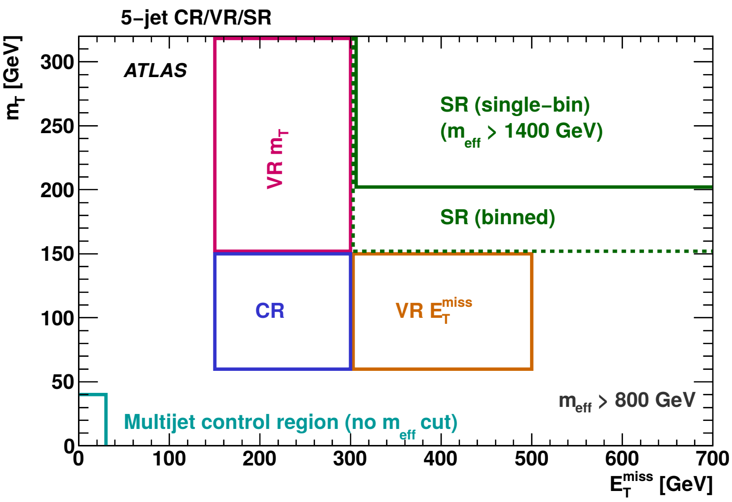
The three stated backgrounds are estimated in a (semi-)data-driven way via the use of control regions. Smaller background contributions are estimated entirely based on MC simulations.

5.4 Region definitions

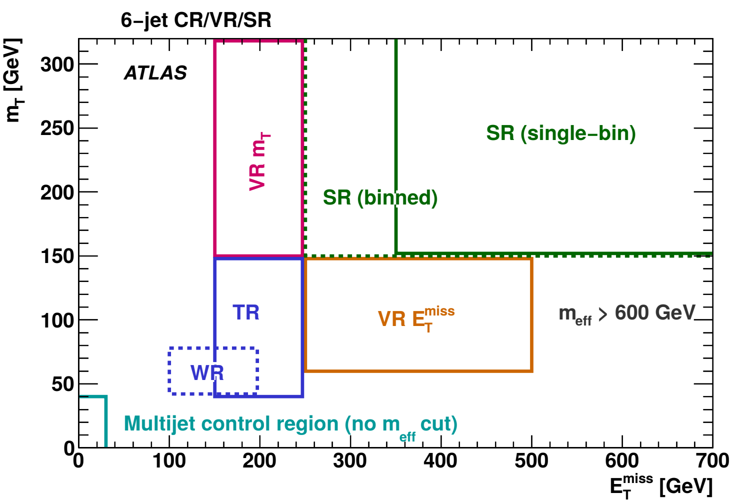
As described in section 4.2.3, control, validation and signal regions are defined. Different regions are used for 3-jet, 5-jet and 6-jet events. All regions are illustrated in Fig. 5.3. The detailed cuts for the signal region are shown in table 5.1, the binning of the signal regions in table 5.2.



(a) 3 jet regions



(b) 5 jet regions



(c) 6 jet regions

Figure 5.3: Illustration of the region definition of the hard one lepton analysis. Figure taken from [54]. For each of the signal regions `SR` one single-bin (e.g. for model-independent fits) and one binned region (e.g. for exclusion fits) is defined. The analysis in this thesis only employs the latter. `WR` are the control regions for the W+jets background, `TR` those of the $t\bar{t}$ background.

	3-jet (`SR3J`), 5-jet (`SR5J`)	6-jet (`SR6J`)
Binned variable	$m_{\mathrm{eff}}^{\mathrm{inc}}$	$E_{\mathrm{T}}^{\mathrm{miss}}$
Bin 1	[800, 1000]	[250, 350]
Bin 2	[1000, 1200]	[350, 450]
Bin 3	[1200, 1400]	[450, ∞)
Bin 4	[1400, ∞)	—

Table 5.2: The binning of the the binned hard one lepton signal regions.

5.5 Considered pMSSM models

The pMSSM models considered in the hard one lepton analysis belong to a subset of the general pMSSM parameter space on which several constraints have been applied.[4] Only models for which sensitivity can be expected and which have not already been excluded by other experiments are covered. This places upper and lower bounds on all of the parameters, but the parameter space is still to large to be covered using a grid. Instead, a set of set of models with random parameters is considered. The models are split into two categories depending on whether the LSP is wino-like (i.e. the W-Boson share in the LSP is more than 50%) or bino-like (i.e. the B-Boson share in the LSP is more than 50%).

[4] section based on [58].

Chapter 6

Truth-level studies with the hard one lepton analysis

In this chapter a truth-level shape fit analysis for the hard one lepton case is presented. The obtained truth-level CL_s values are compared to the corresponding reco-level CL_s values. The agreement is found to be very low. Several sources of error are ruled out, showing the necessity of a more detailed analysis of the underlying signal histograms.

6.1 Previous studies

Previous pMSSM studies tried to estimate the sensitivity to pMSSM models on a truth-level basis by considering the variable

$$r^{\mathrm{SR}} = \frac{N^{\mathrm{SR}}_{\text{expected, truth}}}{N^{\mathrm{SR}}_{\mathrm{UL}}}, \tag{6.1}$$

where $N^{\mathrm{SR}}_{\text{expected, truth}}$ is the truth-level expectation of the signal yield in a single-bin signal region and $N^{\mathrm{SR}}_{\mathrm{UL}}$ the corresponding model-independent upper limit that was set using reco samples.

Based on the r^{SR} value, the models were subdivided into three different categories as shown in table 6.1.

Category	Requirement	Interpretation
1	$r^{\mathrm{SR}} \leq r_1^{\mathrm{SR}}$	Model likely can not be excluded
2	$r_1^{\mathrm{SR}} < r^{\mathrm{SR}} \leq r_2^{\mathrm{SR}}$	Model can possibly be excluded
3	$r_2^{\mathrm{SR}} < r^{\mathrm{SR}}$	Model can certainly be excluded

Table 6.1: Model categories. Table taken from [59].

Here, r_1^{SR} and r_2^{SR} are constants that are estimated by comparing the r^{SR} values to the CL_s values based on reco samples. An exemplaric plot is shown in Fig. 6.1. It is clearly visible that the correlation between r^{SR} and the reco-level CL_s value is very low. For the hard one lepton analysis, the values $r_1^{\mathrm{SR}} = 1$ and $r_2^{\mathrm{SR}} = 5$ are obtained.

The amount of models that can be excluded by this method is very low for the one lepton case: 92% of the considered models belong to category 1 and less than 3% belong to category 3.[58] Most other analyses are far more sensitive, e.g. for the 2-6 jet zero lepton analysis, more than 45% of the models belong to cateogry 3.[58] This is surprising, since the reco-level one hard lepton analysis normally compares quite well to other analyses.

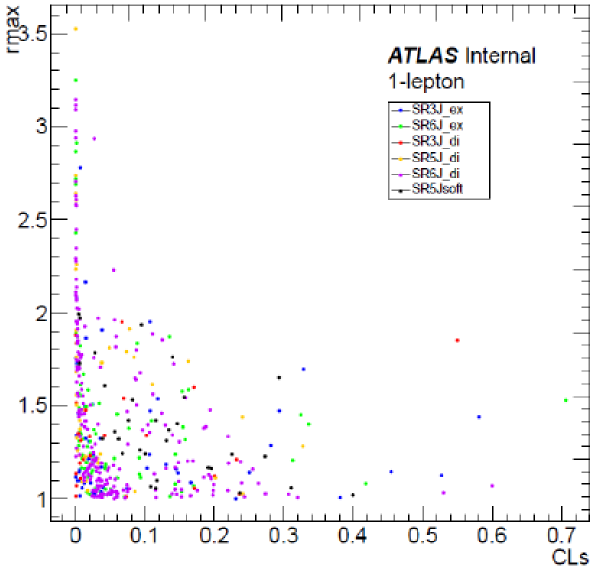


Figure 6.1: A comparison plot of the r^{SR} values to the CL_s values from reco samples. Plot from [59].

6.2 Motivation for a shape fit

By using the above described r^{SR} values, which are based on single bin signal regions, the potentially distinct shapes of signal and background are neglected. Using this information might result in an increased sensitivity of the study. In this thesis, an analysis with largely identical fit strategies for the reco and truth samples is performed. In particular a shape fit is performed on the truth sample.

There are several steps and conditions for this to work:

- The total signal event yields of the truth samples should match those of the reco samples. If necessary, this needs to be corrected by the use of appropriate scaling factors (discussed in section 6.4.1)
- The shapes of the signal distributions of reco truth and reco samples should match (studied in section 6.4.5)
- The uncertainties on the signal histograms should match; for the truth samples estimations for some uncertainties are necessary (see section 6.4.3)

Figure 6.2 shows overlayed truth-level and reco-level signal histograms for an exemplary model. The distributions seem to be in a generally good agreement, which was taken as a promising sign to pursue the idea of a truth shape fit. Note however that only two cuts were applied to the events, so that the differences – even though seemingly small in the presented plots – might still be substantial in the signal regions (which are subject to more restrictions).

6.3 Setup of the truth-level shape fit

The comparison routines and fits described in section 4.2 to 4.5 are implemented in `HistFitter`, a software framework that has become the standard statistical tool for the SUSY analyses performed by the ATLAS collaboration.[47] `HistFitter` acts as an interface that wraps around

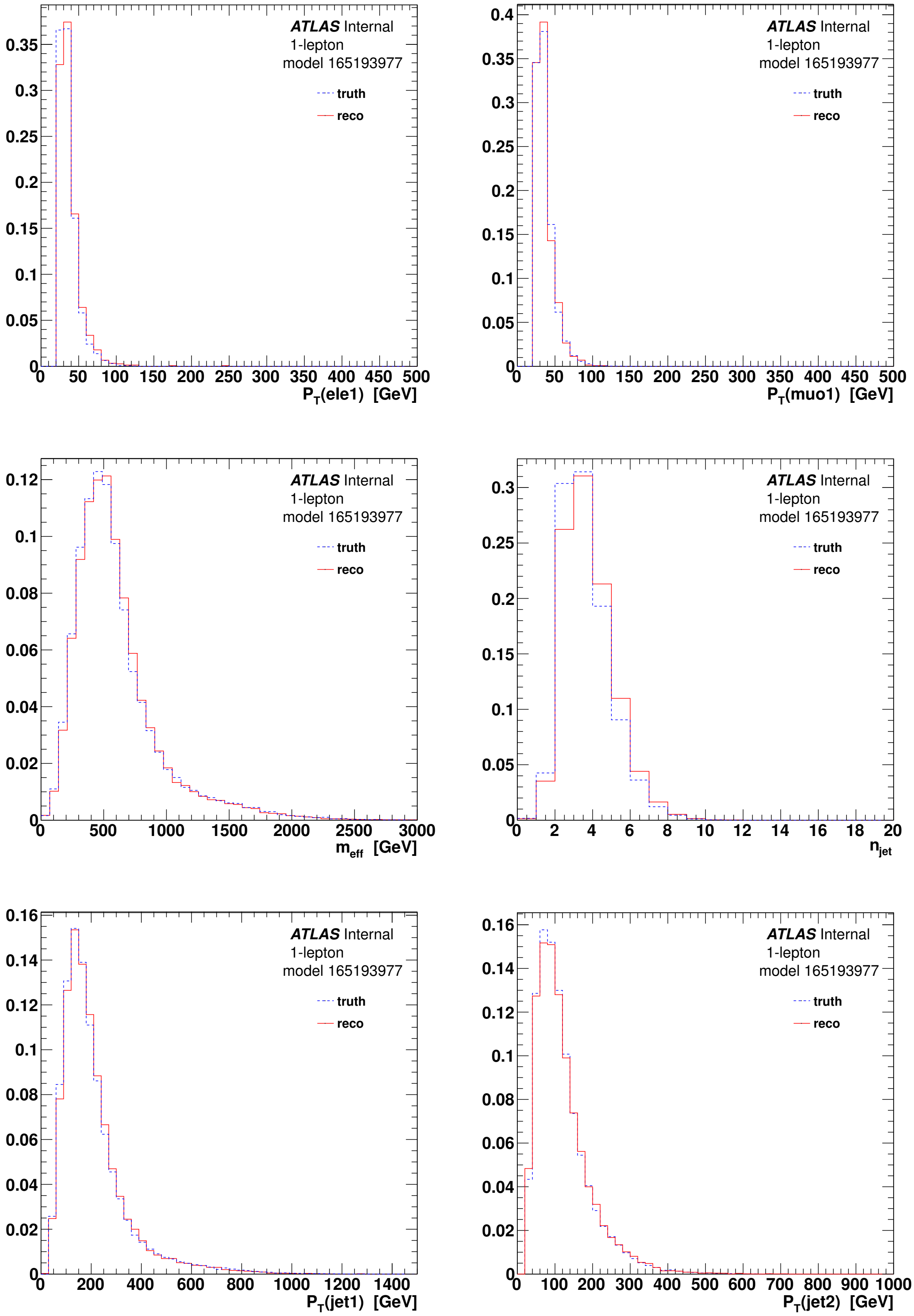


Figure 6.2: Overlayed truth and reco signal yield expectations, binned in various variables. The cuts $p_{\mathrm{T}}^{1^{\mathrm{st}}\,\ell} > 25$ GeV and $p_{\mathrm{T}}^{2^{\mathrm{nd}}\,\ell} < 10$ GeV were applied to the samples. Plot from [58].

underlying calls to `ROOT`[1] and its libraries `RooFit` and `RooStats`. It is governed by a single configuration file, written in python.

The configuration file for the reco-level hard one lepton analyses consists of about 1500 lines of code. It requires specialized `ROOT` files and loads numerous systematic uncertainties. To bypass its complexity, a trick was employed, modifying the configuration file in a way that independently generated signal histograms could easily be incorporated. For this to work, the source code of the `HistFitter` framework had to be adapted.[2]

6.4 Studies using a Truth Cutflow

With the workflow described above, the required truth histograms need to be produced independently. To that end, a script (*truth cutflow*) was adapted to mirror the cuts performed by `HistFitter`, extract the necessary histograms from the corresponding `ROOT` trees and to bring them into the exact form required by `HistFitter` (e.g. identical name, binning and use of overflow bins).

Nine pMSSM models that fulfilled all requirements to be used for both reco-level and truth-level analysis were at hand and were used to perform initial tests.

6.4.1 Scaling factors

As the detector efficiency is generally lower than one, scaling factors < 1 need to be applied to the truth samples. The signal event yields in the signal regions of reco and truth samples were first compared using the script `YieldsTable.py` that comes with the `HistFitter` bundle. While the results of some models (e.g. `pMSSM_Wino_3368912`) are acceptable, the agreement is generally low. However, due to the small number of considered models, no general conclusion could be drawn.

Figure 6.3 shows the event yield in the signal regions of reco and truth samples for a higher number of pMSSM models. A linear connection between both values is observable. Fitting of the corresponding lines returns the slopes 0.71 (`SR3J`), 0.78 (`SR5J`) and 0.69 (`SR6J`) (as calculated by Brian Petersen). These *scaling factors* were implemented in the truth cutflow.

6.4.2 B-Tagging

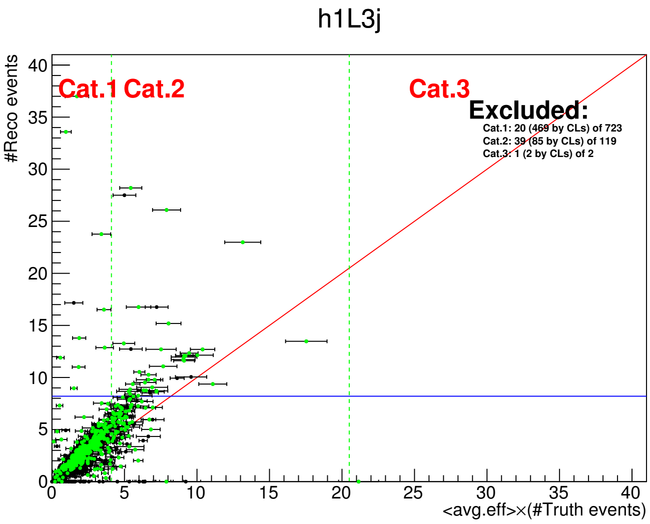
The previous configuration of the truth cutflow did not contain cuts based on quark flavors so that the the b-tagging[3] (which is necessary for the definition of the control regions) had to be implemented. To validate the implementation, the ratio of the events that passed all control region cuts to the events that passed the same conditions without a requirement for the quark flavor was compared between truth and reco samples. To that end, a script in `pyROOT`[4] was written to loop over the models and extract the control region event yields with and without b-tagging from the corresponding `ROOT` trees.

[1] a `C++` written library developed by CERN that specializes in efficient handling and analyzing of large data amounts[60]

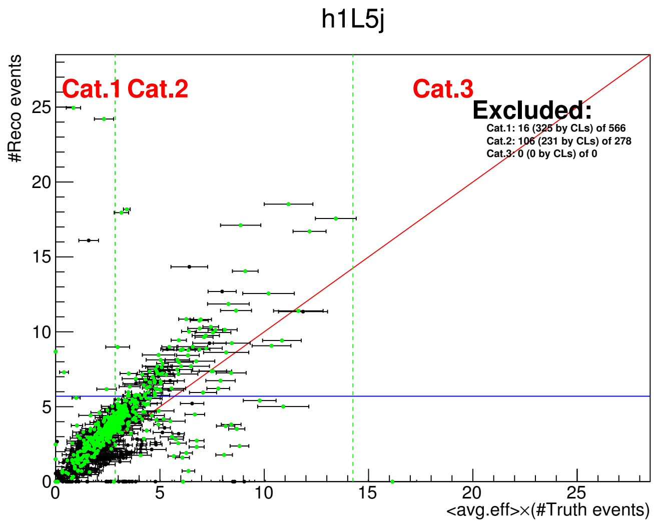
[2] the following "hack" was employed: To speed up the computation process, `HistFitter` is set up to search a *cache file* for relevant histograms before trying to recreate them from `ROOT` trees. Thus, replacing the reco-level signal histograms of the cache file with truth-level signal histograms makes `HistFitter` run on truth samples with but few necessary changes in the configuration file. The current release of `HistFitter` only allows for one arbitrary cache file. To allow for more flexibility, this functionality was extended by the possibility to attach *multiple* cache files, which required some alterations of the `HistFitter` source.

[3] b-tagging is the identification of jets that originated from bottom quarks

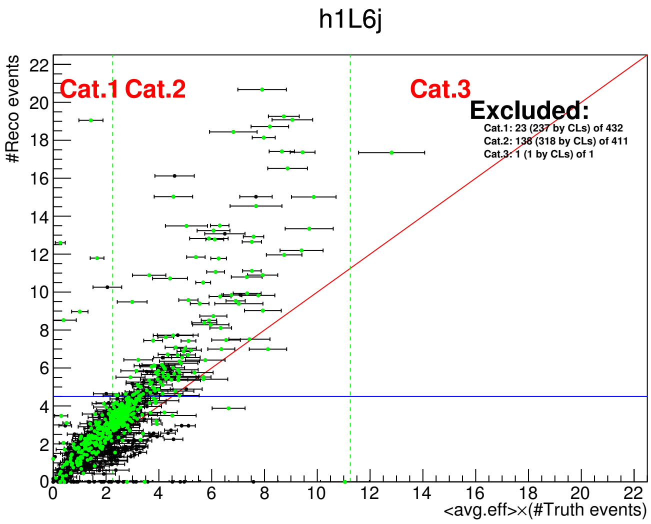
[4] a python wrapper for `ROOT`



(a) 3-jet



(b) 5-jet



(c) 6-jet

Figure 6.3: Event yields in the signal regions for truth and reco samples. A scaling factor (`<avg.eff>`) of 0.5 has already been applied to the truth samples. Since the majority of models lies above the red line with slope 1, the ideal scaling factors are a higher. Plots by Brian Petersen [61].

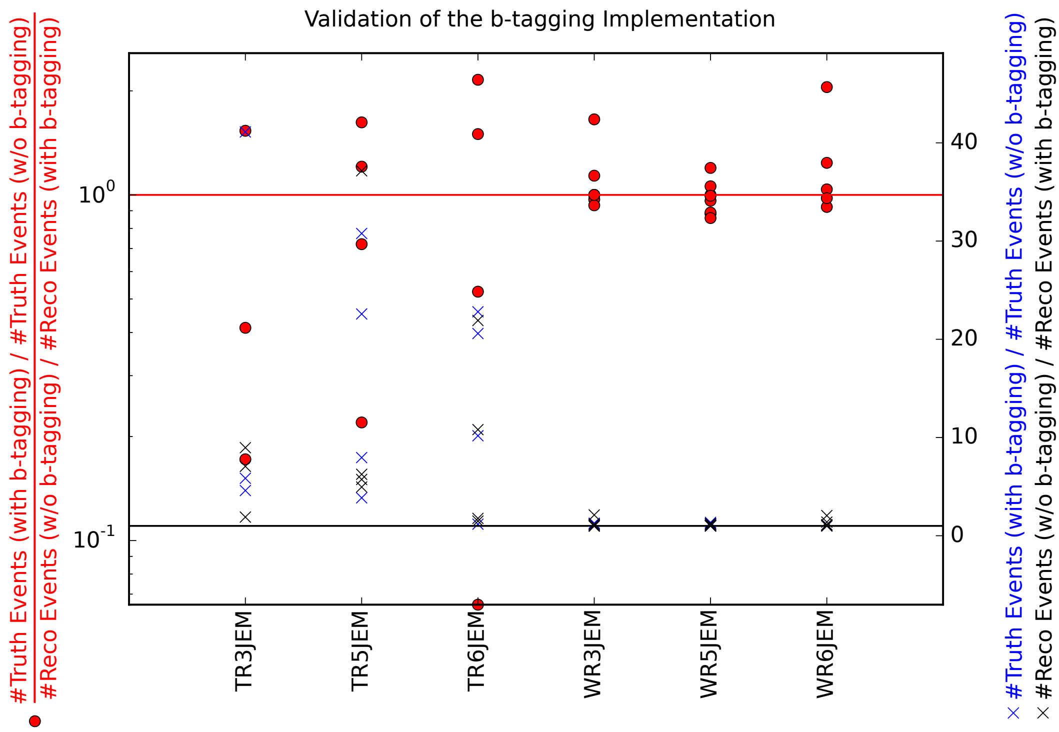


Figure 6.4: The ratio of the events that passed all control region cuts to the events that passed the same conditions without a requirement for the quark flavors is represented by black crosses (reco-level) resp. blue crosses (truth-level), using the right scale. The ratio of both is drawn as red dots using the left (logarithmic) scale. The red resp. black horizontal line is at 1 (seen from the left resp. right scale).

As shown in figure 6.4, the respective ratios only agree well in regions with few b-quarks. However, due to low statistics (both as far as the number of considered models is concerned and with respect to very low signal levels), many of the outliers may be the results of statistical fluctuations.

6.4.3 Systematic Uncertainties

Using the script `SysTable.py` that comes with the `HistFitter` bundle, the uncertainty of the background estimate in the different signal regions was broken down into the dominant systematic uncertainties. The result for one of the reco samples is shown in table 6.2. For the truth sample (table 6.3), the detector-related uncertainties cannot be calculated, as the truth-level analysis bypasses the detector simulation and reconstruction. To make up for this, an additional user defined fixed systematic uncertainty is added. This also has the reason that the truth analysis is motivated by the search for practical and light-weighed analysis strategies that do not require the time consuming loading of numerous systematic uncertainties.

Based on the results of table 6.2, the new uncertainty is set to a rough estimate of 30% for all signal regions. In section 6.5.2, the effects of different uncertainties on the corresponding truth-level CL_s values are discussed.

Table 6.2: Breakdown of the dominant systematic uncertainties on background estimates in the signal regions for the reco sample pMSSM_Wino_8892390 (before any fit was performed). The uncertainty of the cross section was set to zero. Note that the individual uncertainties can be correlated, and do not necessarily add up quadratically to the total background uncertainty. The percentages show the size of the uncertainty relative to the total expected background. Uncertainities that are nonzero after rounding are highlighted in blue.

Uncertainty of channel	SR3J[5]	SR5J[6]	SR6J[7]
Total background expectation	6.81	3.76	3.30
Total statistical ($\sqrt{N_{\mathrm{exp}}}$)	±2.61	±1.94	±1.82
Total background systematic	±0.79 [11.56%]	±1.59 [42.21%]	±0.79 [23.81%]
alpha_JER	±0.62 [9.1%]	±0.66 [17.5%]	±0.24 [7.3%]
alpha_h1L_JVF	±0.28 [4.1%]	±0.28 [7.4%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_1	±0.19 [2.8%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_pileup	±0.17 [2.5%]	±0.09 [2.3%]	±0.24 [7.2%]
alpha_JES_FlavorCompUncert	±0.17 [2.5%]	±0.47 [12.5%]	±0.27 [8.2%]
alpha_JES_EffectiveNP2	±0.17 [2.5%]	±0.47 [12.5%]	±0.26 [8.0%]
alpha_JES_PileupRhoTopology	±0.14 [2.0%]	±0.13 [3.5%]	±0.00 [0.00%]
alpha_JES_EffectiveNP1	±0.09 [1.4%]	±0.54 [14.3%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_2	±0.07 [0.98%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_3	±0.06 [0.85%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_FlavorResponseUncert	±0.04 [0.54%]	±0.53 [14.1%]	±0.26 [8.0%]
alpha_JES_BJes	±0.04 [0.54%]	±0.75 [19.8%]	±0.26 [8.0%]
alpha_JES_EtaIntercalibrationModelling	±0.00 [0.07%]	±0.54 [14.3%]	±0.26 [7.9%]
mu_SIG	±0.00 [0.01%]	±0.00 [0.01%]	±0.00 [0.01%]
alpha_h1L_SCALEST	±0.00 [0.00%]	±0.27 [7.3%]	±0.26 [8.0%]
alpha_h1L_ktfacZ	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_SingleTopTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_W_5J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_qfacW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_2	±0.00 [0.00%]	±0.00 [0.00%]	±0.26 [7.9%]
alpha_h1L_ktfacW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_BT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_0	±0.00 [0.00%]	±0.11 [2.9%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_1	±0.00 [0.00%]	±0.09 [2.3%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_2	±0.00 [0.00%]	±0.08 [2.1%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_3	±0.00 [0.00%]	±0.15 [4.0%]	±0.00 [0.00%]
alpha_ttbarVTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_6J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]

[5] h1L_SR3JEM_meffInc30_JVF25pt50_pMSSM_Wino_889239000
[6] h1L_SR5JEM_meffInc30_JVF25pt50_pMSSM_Wino_8892390
[7] h1L_SR6JEM_met_pMSSM_Wino_8892390

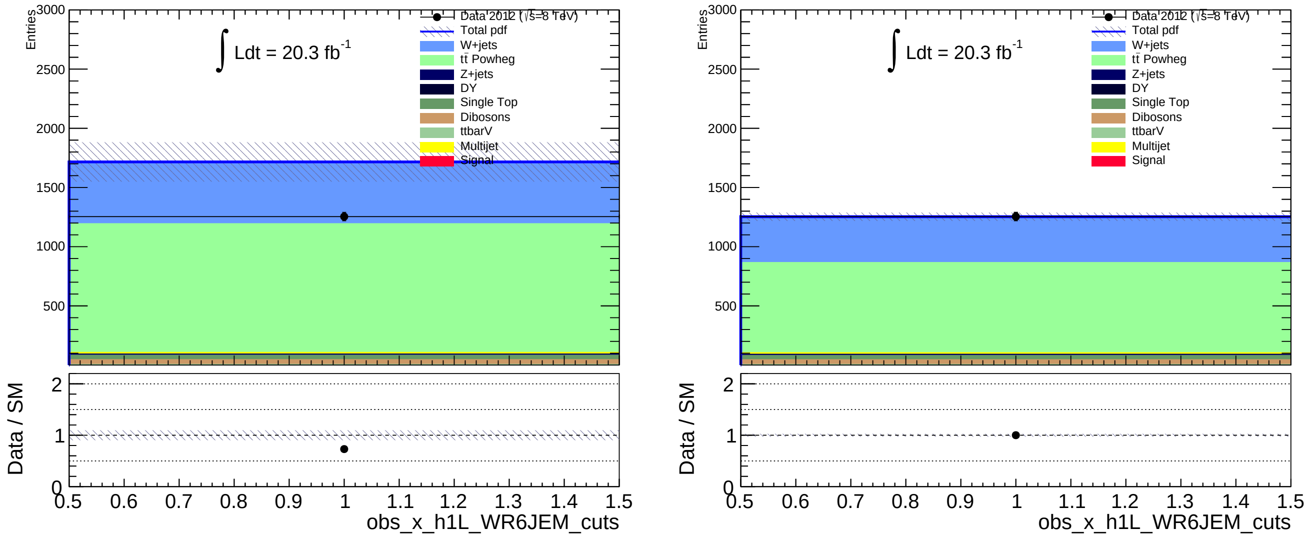
h1L_mu_W_6J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_TR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.06 [1.8%]
alpha_h1L_SingleTopWtXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_WTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ZTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarWWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_qfacZ	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_WTheoNpart	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_topTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopTXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarZWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_TR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopSXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_RESOST	±0.00 [0.00%]	±0.01 [0.31%]	±0.00 [0.00%]
alpha_h1L_DBWZXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_5J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_MisT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_DBWWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_3J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_CT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_dbTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_TR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_AlpgenJimmyTTbar	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ZTheoNpart	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_W_3J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarVTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_PowhegJimmyTTbar	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_topTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.11 [3.4%]
alpha_h1L_SingleTopZXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_dbTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]

Table 6.3: Breakdown of the dominant systematic uncertainties on background estimates in the signal regions for the truth sample `pMSSM_Wino_8892390` (before any fit was performed). Note that the individual uncertainties can be correlated, and do not necessarily add up quadratically to the total background uncertainty. The percentages show the size of the uncertainty relative to the total expected background. Uncertainities that are nonzero after rounding are highlighted in blue.
Due to the nature of the truth analysis, all of the detector-related uncertainties are zero; only statistical errors (all of them having the prefix `gamma_stat`) and the uncertainties of the fits of the signal strength parameters (all of them having the prefix `mu`) remain. To make up for this, an additional user defined fixed systematic uncertainty is added: `alpha_myOverallSys`.

Uncertainty of channel	SR3J[8]	SR5J[9]	SR6J[10]
Total background expectation	7.66	3.22	2.50
Total statistical ($\sqrt{N_{\text{exp}}}$)	±2.77	±1.79	±1.58
Total background systematic	±2.33 [30.42%]	±1.01 [31.26%]	±0.78 [31.23%]
alpha_myOverallSys	±2.30 [30.0%]	±0.96 [30.0%]	±0.75 [30.0%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_0	±0.31 [4.1%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_1	±0.19 [2.5%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_3	±0.10 [1.2%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR3JEM_meffInc30_JVF25pt50_bin_2	±0.07 [0.94%]	±0.00 [0.00%]	±0.00 [0.00%]
mu_SIG	±0.00 [0.01%]	±0.00 [0.01%]	±0.00 [0.01%]
alpha_h1L_ktfacZ	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_PileupRhoTopology	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_pileup	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_EtaIntercalibrationModelling	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_SingleTopTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_W_5J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_JVF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_qfacW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_2	±0.00 [0.00%]	±0.00 [0.00%]	±0.05 [2.2%]
alpha_h1L_ktfacW	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_FlavorCompUncert	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_BT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_0	±0.00 [0.00%]	±0.12 [3.8%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_1	±0.00 [0.00%]	±0.22 [6.9%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_2	±0.00 [0.00%]	±0.09 [2.9%]	±0.00 [0.00%]
gamma_stat_h1L_SR5JEM_meffInc30_JVF25pt50_bin_3	±0.00 [0.00%]	±0.08 [2.5%]	±0.00 [0.00%]
alpha_ttbarVTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_6J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_W_6J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]

[8] h1L_SR3JEM_meffInc30_JVF25pt50_pMSSM_Wino_889239000
[9] h1L_SR5JEM_meffInc30_JVF25pt50_pMSSM_Wino_8892390
[10] h1L_SR6JEM_met_pMSSM_Wino_8892390

alpha_QCDNorm_h1L_TR3JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_WR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_1	±0.00 [0.00%]	±0.00 [0.00%]	±0.15 [5.9%]
alpha_h1L_SingleTopWtXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_WTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_FlavorResponseUncert	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ZTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarWWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_qfacZ	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_WTheoNpart	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_topTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopTXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarZWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_BJes	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_TR5JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopSXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_RESOST	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_DBWZXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_5J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_MisT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SCALEST	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_DBWWXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_Top_3J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_SingleTopTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_CT	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_dbTheo	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_TR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_EffectiveNP2	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JES_EffectiveNP1	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_AlpgenJimmyTTbar	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_ZTheoNpart	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
h1L_mu_W_3J	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_ttbarVTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_PowhegJimmyTTbar	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_topTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_QCDNorm_h1L_SR6JEM	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
gamma_stat_h1L_SR6JEM_met_bin_0	±0.00 [0.00%]	±0.00 [0.00%]	±0.15 [6.0%]
alpha_h1L_SingleTopZXsec	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_h1L_dbTheoPDF	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]
alpha_JER	±0.00 [0.00%]	±0.00 [0.00%]	±0.00 [0.00%]



(a) WR6J before fit

(b) WR6J after fit

Figure 6.5: Exclusion fit of the W+jets background in the corresponding 6-jet control region `WR6J` for the truth-level signal model `pMSSM_Wino_8892390`. Plots produced with `HistFitter`.

6.4.4 Fit results

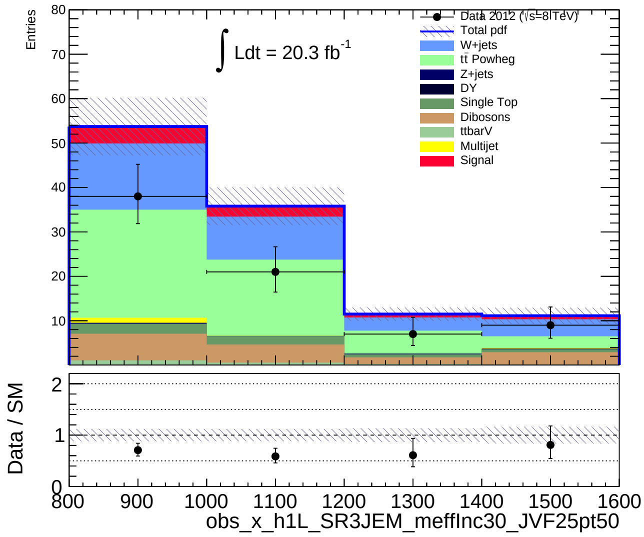
Fig. 6.5 and 6.6 present event yields in different regions before and after the exclusion fit. Fig. 6.5 illustrates how the W+jets background in the 6-jet signal region `SR6J` is estimated based on the event yield in the conrol region `WR6J`. The left plot shows the simulated event yield together with the measured event yield in `WR6J` before any fit. The fit then adapts the background estimation to exactly match the measured event (right plot). The background estimation is transferred to the signal region `SR6J` which can be seen in a decreased background after the fit, also eliminating the SUSY signal (Fig. 6.6e and 6.6f). The before-after plots of the 3-jet and 5-jet signal region are shown in Fig. 6.6a to 6.6d.

6.4.5 Comparison of signal shapes between truth and reco samples

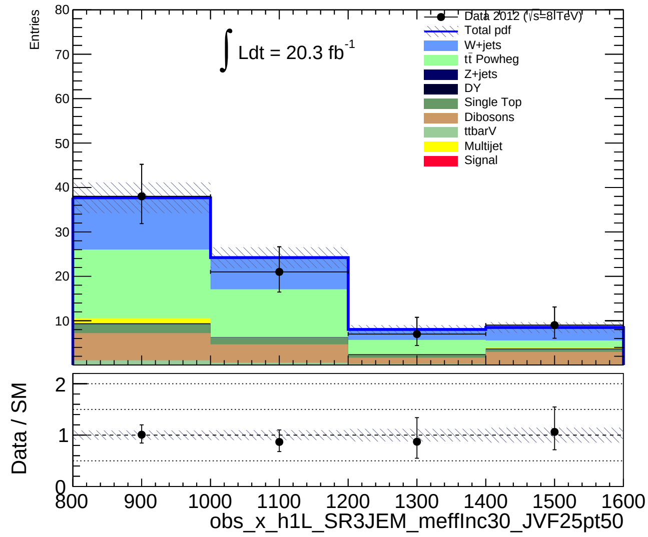
In order for the reco- and truth-level analysis to consistently show a similar sensitivity, the signal expectations in the signal regions need to match. To test this for the 9 signal models at hand, a `pyROOT` script was written to extract the relevant histograms out of the `HistFitter` fit results output.

An example for a model that shows good comparability between reco and truth samples is given in Fig. 6.7. In the 3-jet and 5-jet signal region, the scaling could be improved, but the truth- and reco-level signal shapes show a good agreement. The signal shapes in the 6-jet signal region show less agreement.

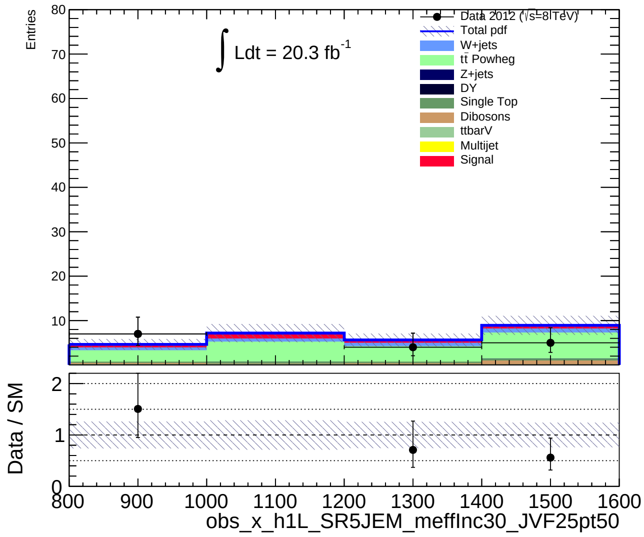
Fig 6.8 shows a model with non-agreeing signal shapes for truth- and reco-level samples (though the scaling roughly agrees). An even more extreme example is given in Fig. 6.9, where for each region either the truth- or reco-level event yield is vanishing. The plots for the remaining 6 signal models can be found in appendix 8.



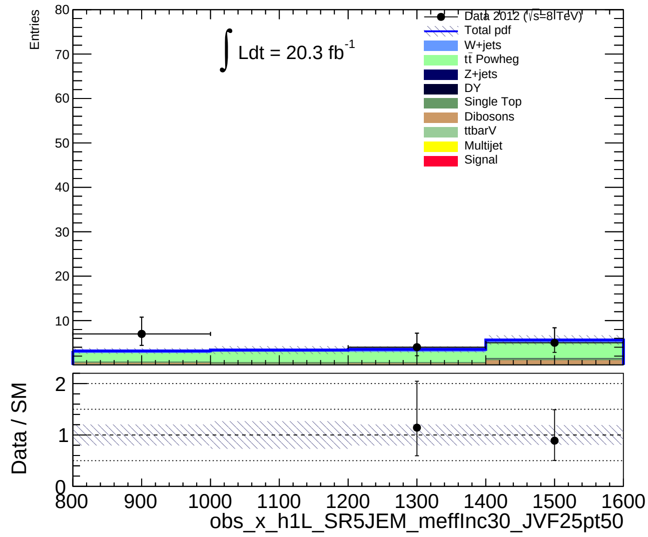
(a) SR3J before fit



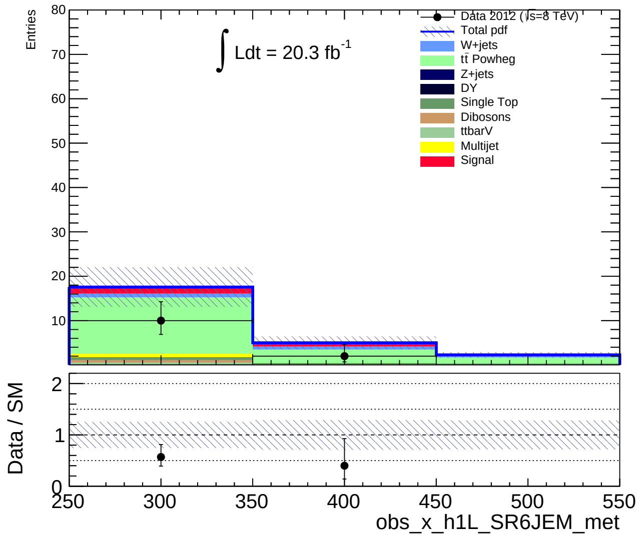
(b) SR3J after fit



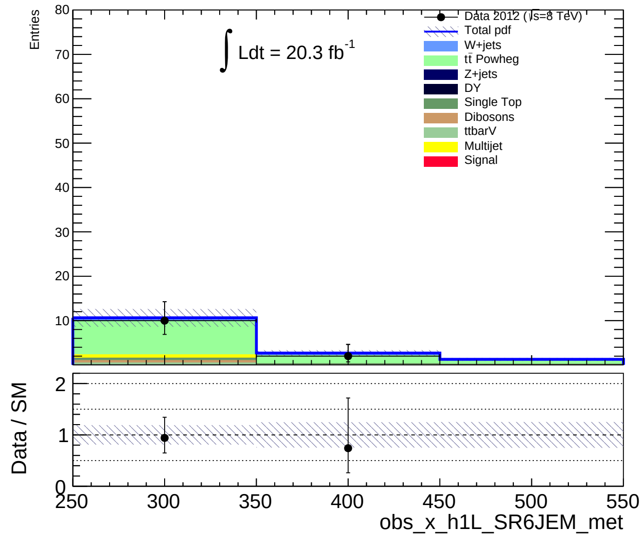
(c) SR5J before fit



(d) SR5J after fit

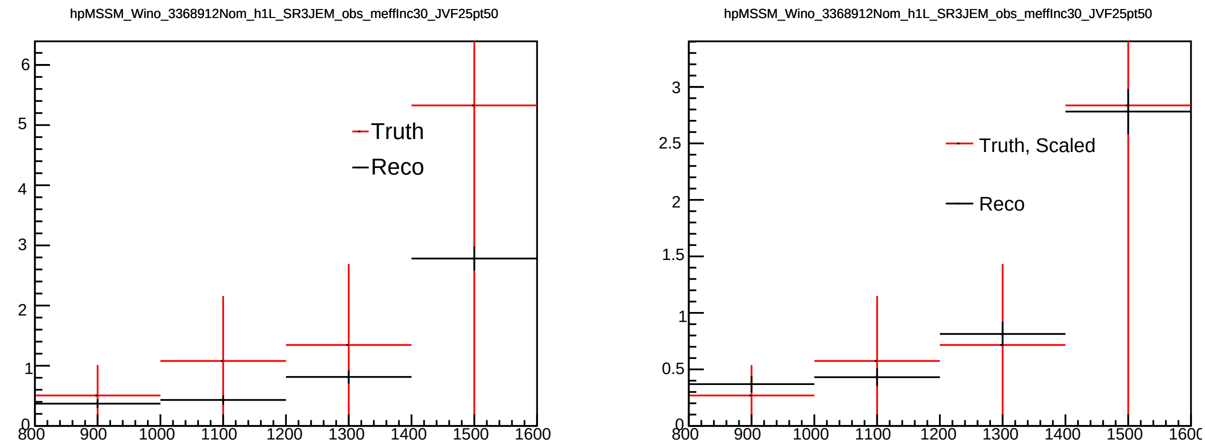


(e) SR6J before fit

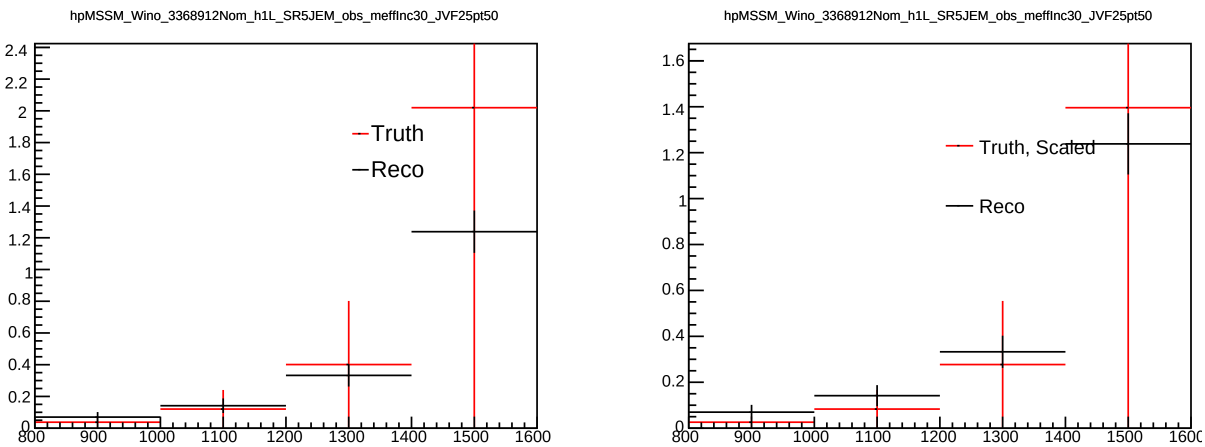


(f) SR6J after fit

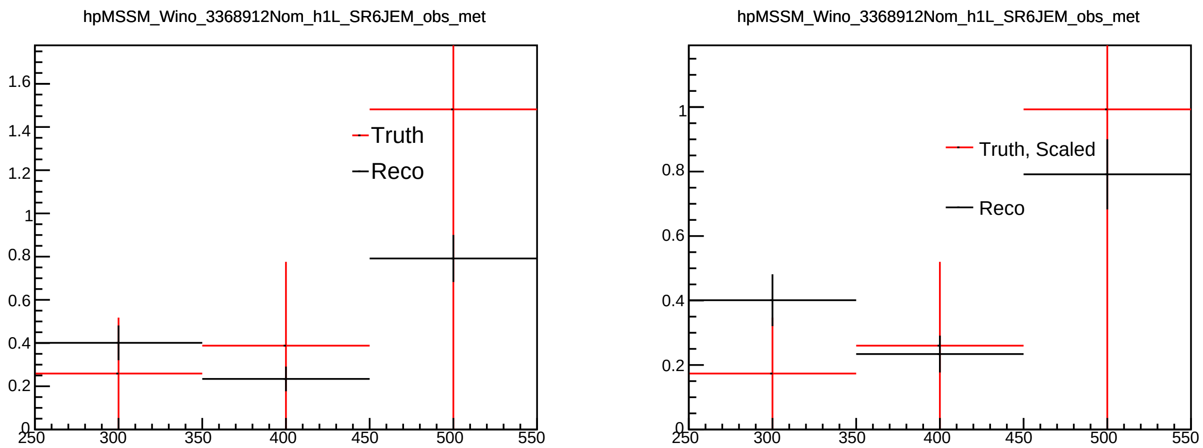
Figure 6.6: Exclusion fit of the 6-jet signal region `SR6J` for the truth-level signal sample `pMSSM_Wino_8892390`. Plots produced with `HistFitter`.



(a) SR3J

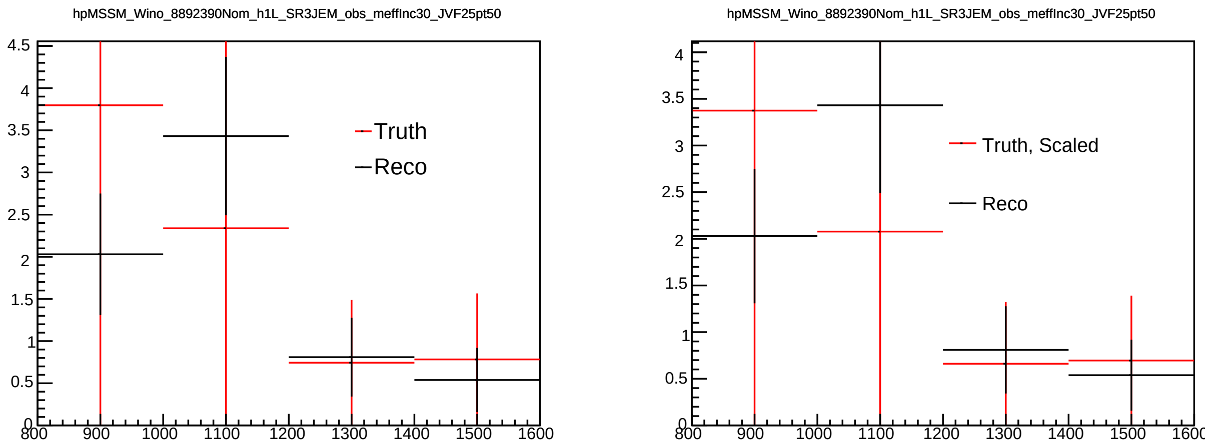


(b) SR5J

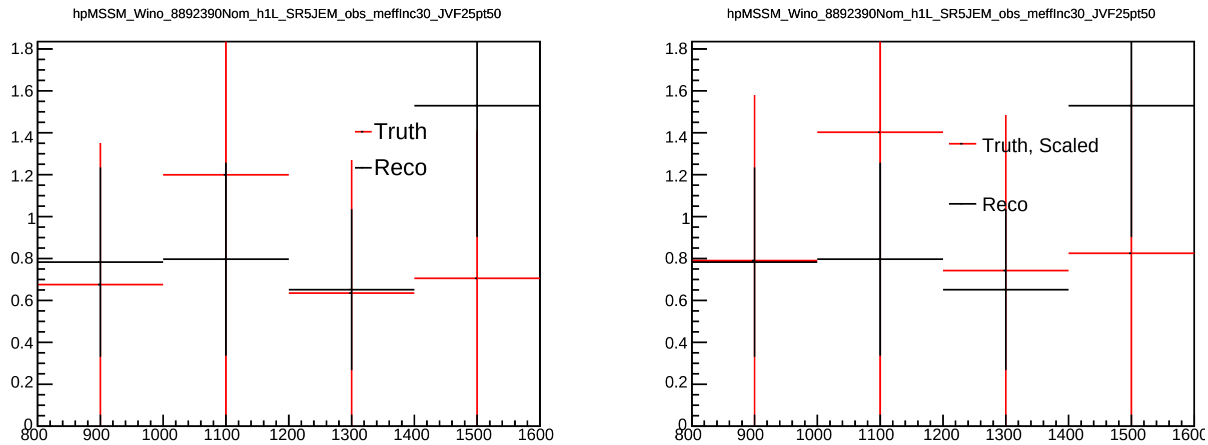


(c) SR6J

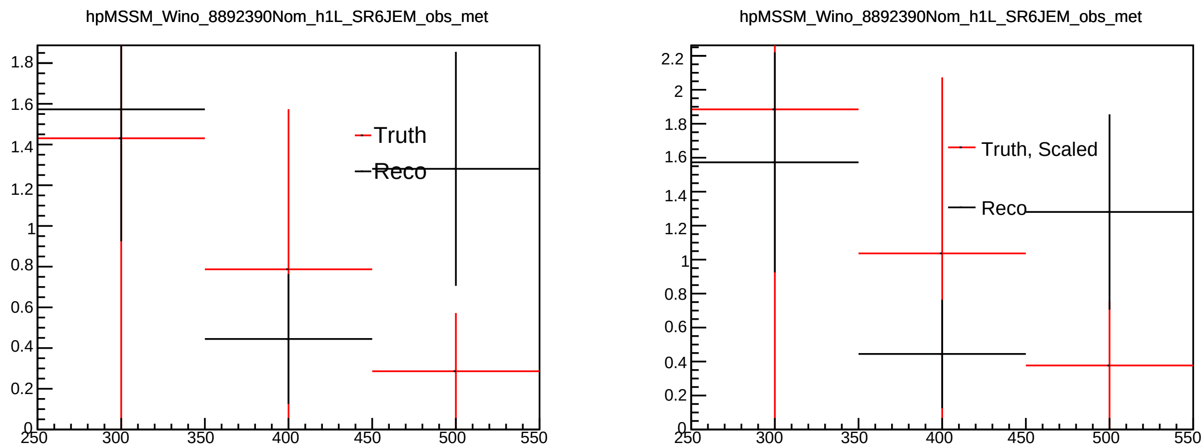
Figure 6.7: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_3368912`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J

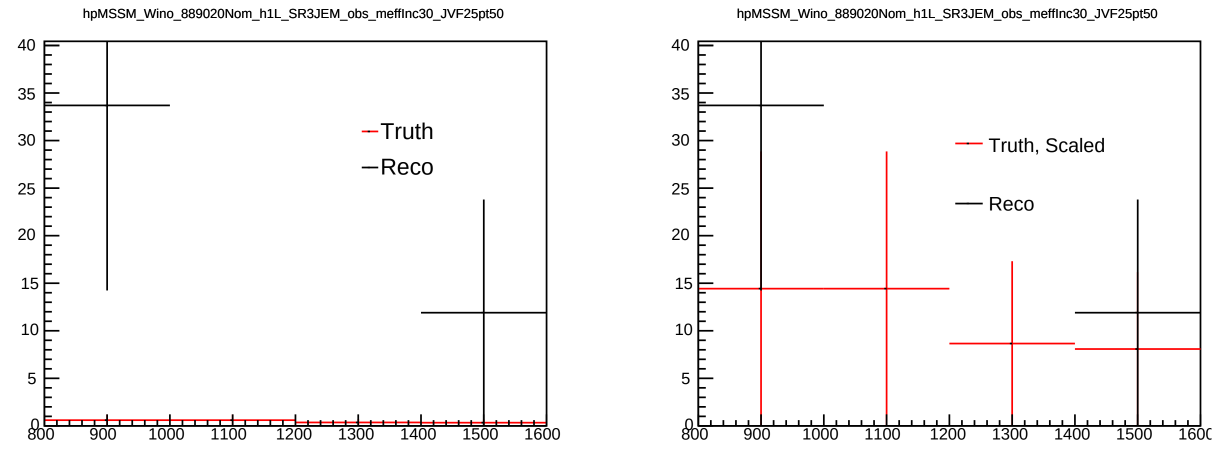


(b) SR5J

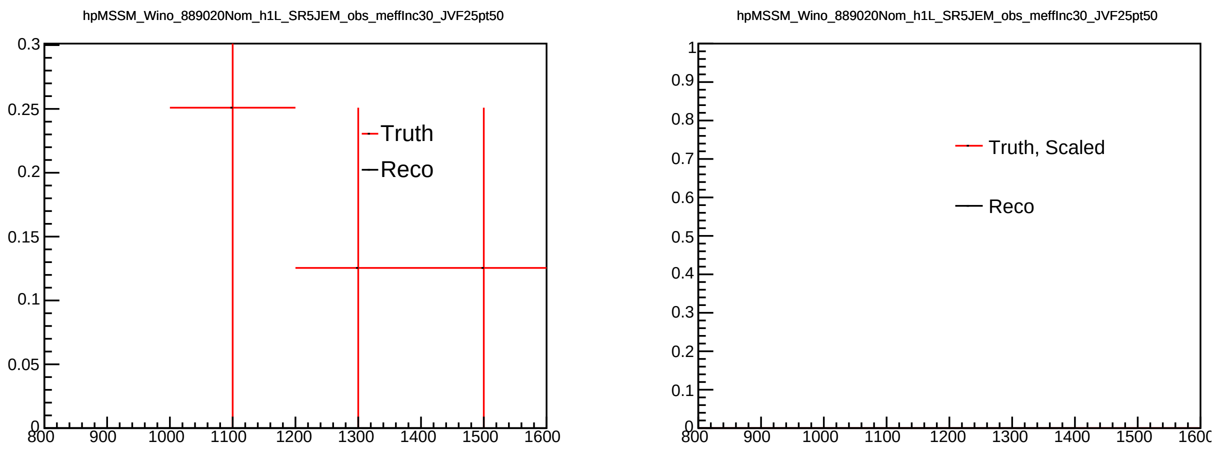


(c) SR6J

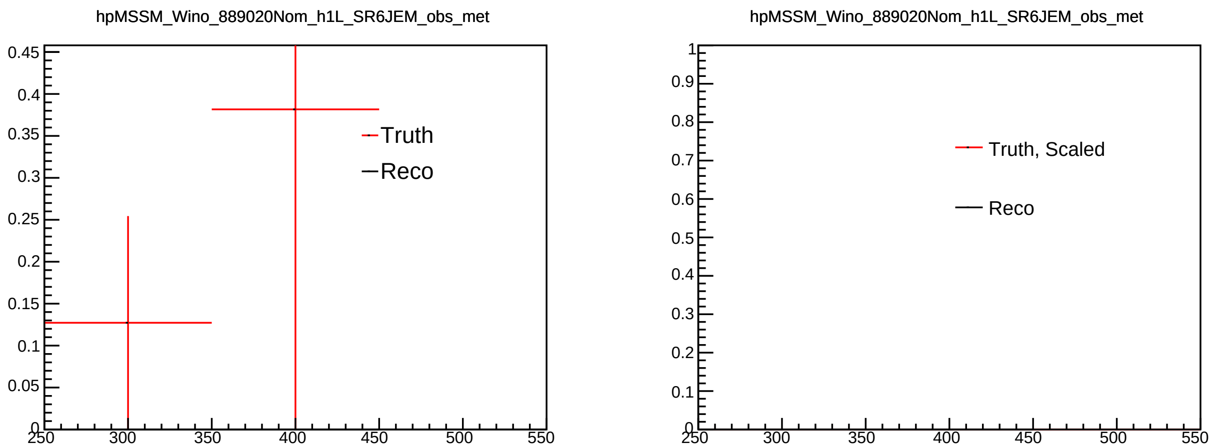
Figure 6.8: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_8892390`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J



(b) SR5J



(c) SR6J

Figure 6.9: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_889020`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.

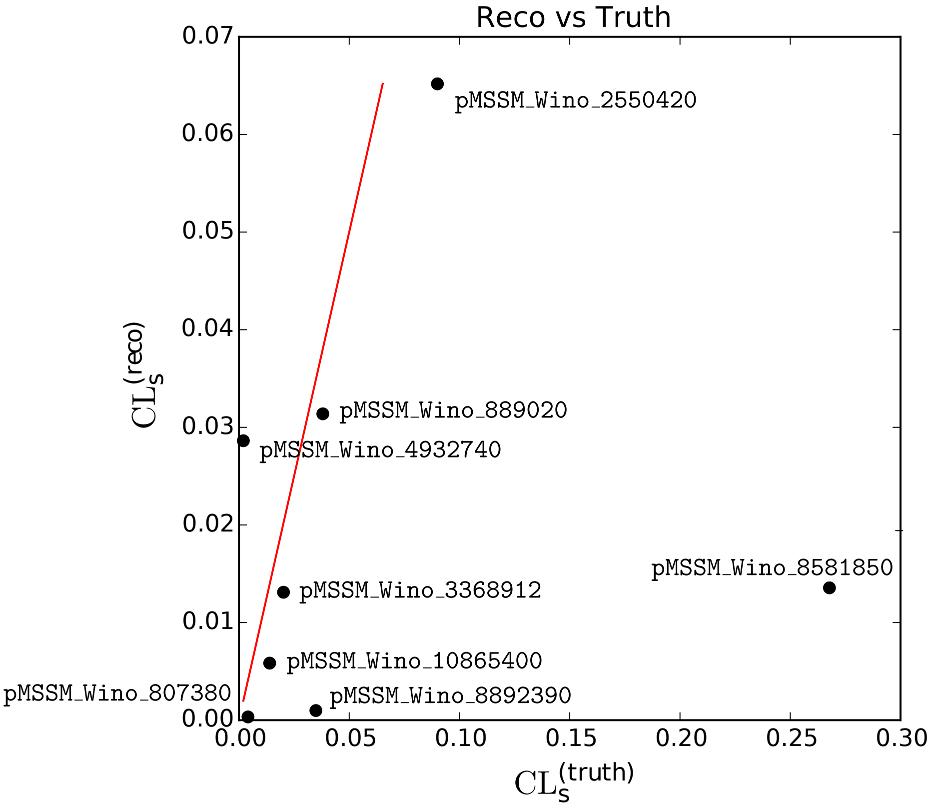


Figure 6.10: CL_s values of reco and truth samples of the 8 signal points. The red line has slope 1.

Model	$\mathrm{CL}_s^{(\mathrm{reco})}$	$\mathrm{CL}_s^{(\mathrm{truth})}$	$\mathrm{CL}_s^{(\mathrm{reco})}/\mathrm{CL}_s^{(\mathrm{truth})}$
pMSSM_Wino_10865400	0.00584614	0.0139343	0.420
pMSSM_Wino_13370909	nan	0.012229	nan
pMSSM_Wino_2550420	0.0651818	0.0899978	0.724
pMSSM_Wino_3368912	0.0131067	0.0202039	0.649
pMSSM_Wino_4932740	0.0286261	0.00199311	14.363
pMSSM_Wino_807380	0.00033287	0.00410059	0.081
pMSSM_Wino_8581850	0.0135577	0.267772	0.052
pMSSM_Wino_889020	0.0313822	0.0380319	0.8251
pMSSM_Wino_8892390	0.000985143	0.0348824	0.028

Table 6.4: Truth- and reco-level CL_s values for the 9 considered models

6.4.6 Hypothesis tests

Hypothesis tests for the reco and truth samples of the 9 signal samples were run; the results are presented in table 6.4. The hypothesis test failed for the reco samples of model `pMSSM_Wino_13370909`; the remaining 8 comparison points are shown in Fig. 6.10. The CL_s values for 5 of the 8 models match quite well, while 3 values are outliers.

Note that high exclusion limits (i.e. small CL_s values) are obtained for high signal expectations and signal shapes that are distinct from the background. Furthermore, the hypothesis test uses the data of all three signal regions simultaneously. This means that, while similar truth- and reco-level signal expectations imply similar reco- and truth-level CL_s values, the reverse statement does not hold. This can be illustrated with the three models `pMSSM_Wino_3368912`, `pMSSM_Wino_8892390` and `pMSSM_Wino_889020`:

- Model `pMSSM_Wino_3368912` showed excellently agreeing signal histograms (Fig. 6.7) and thus it is not surprising that similar reco/truth CL_s values are obtained. By comparing the signal expectations with the background expectations shown in Fig. 6.6, it can already been seen that the model is likely to be excluded: the shape of the signal histograms is completely opposed to that of the background estimation and the signal yield is not negligible. The CL_s values of 0.013 (reco) resp. 0.020 (truth) confirm those considerations, excluding the model.

- Model `pMSSM_Wino_8892390` has barely matching truth/reco signal histograms (Fig. 6.8) and, as one might expect, the reco/truth CL_s values disagree. However, the CL_s values of model `pMSSM_Wino_889020` (Fig. 6.9) are in good agreement, even though the truth/reco signals do not match at all: the reco-level signal expectation is high in the 3-jet region but the truth-level signal expectation vanishes, whereas the reco-level signal is vanishing in the 5- and 6-jet regions. Thus, the sensitivity of the reco-level analysis stems only from the 3-jet region, while the sensitivity of the truth-level analysis stems only from he 5- and 6-jet region and both CL_s values happen to be similar.

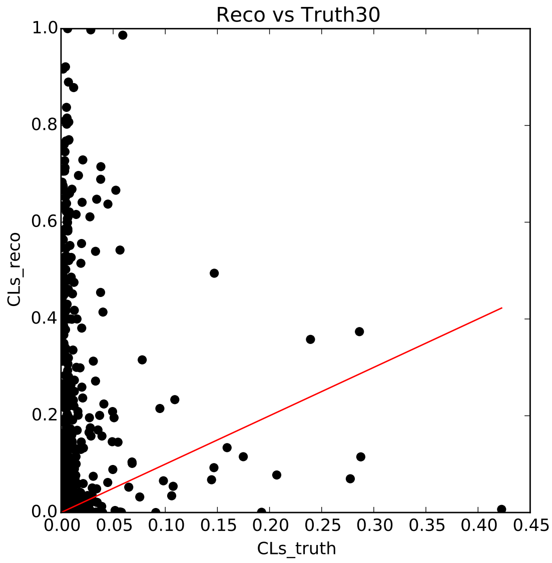
6.5 Studies using Summary NTuples

As the 9 signal models obviously do not allow for a statistically significant analysis, the number of models needed to be increased. To speed up the analysis, summary `NTuples`[11] and reduced summary `NTuples` were used, with the former containing the truth-level event yield of various signal models in the signal regions and the latter the CL_s values of corresponding reco-level one lepton analyses. Unfortunately, the CL_s values were not listed separately for the hard and soft one lepton analysis, but only the minimum of the two analyses was available. This means that the reco-level CL_s values that are used from this point on are lower than the reco-level hard one lepton CL_s value (this is further discussed in section 6.5.4).

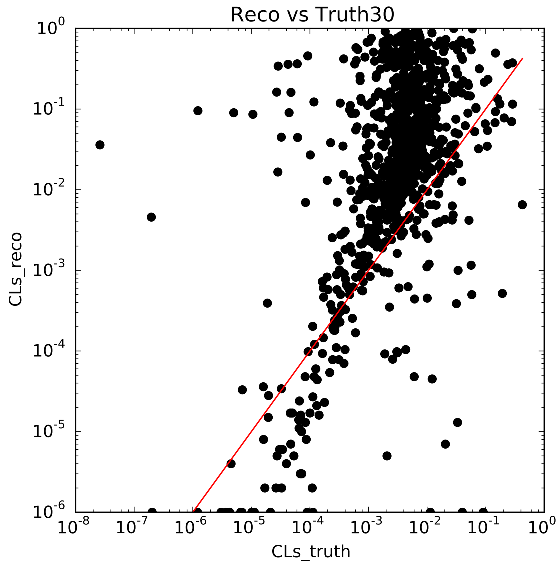
6.5.1 Comparing Reco and Truth CL_s values

The reco-level hard one lepton analyses had only been performed on a small subset of the total amount of models in the `NTuples`, leaving 1043 models to be considered.

[11] `NTuples` are `ROOT` trees that only contain float variables[62]. In this context *Summary* ***NTuples*** refer to `NTuples` that summarize the analyses performed for a certain set of models.



(a) simple scales



(b) doubly logarithmic scales

Figure 6.11: CL_s values of reco and truth samples for the 1043 signal points (of which the hypothesis test failed/was aborted for 9 points). The red line has slope 1.

To extract the necessary data from the `NTuples`, calculate truth-level CL_s values and to test, compare and plot the results, more than 20 bash, python and and `ROOT` scripts were written with a total of about 2000 lines of code.[12]

The initial result is shown in Fig. 6.11(a). If the CL_s values of reco and truth-level analysis would match, the data points should form a straight line with slope 1 (indicated by the red line). This is clearly not the case: the CL_s value of the truth samples is far below the CL_s values of the reco samples. There are almost no models with a reco CL_s value higher then the truth CL_s value. As recognizable in the zoomed out plot 6.12(b) as well as in a double logarithmic plot 6.11(b), the data points do however very roughly form a line but with a slope fare larger than 1.

6.5.2 The influence of systematic uncertainties

Since the uncertainties of the truth samples were only estimated to be 30% based on the consideration of only few models, these might not be representative for the whole data set. Raising the uncertainties of the truth samples results in raising CL_s values and could thus improve the agreement of truth/reco CL_s values.

Additional simulations with 60% and 90% uncertainties were conducted. The effect on the distribution of the truth-levels CL_s values is shown in Fig. 6.14c and 6.14d, with Fig. 6.14b

[12] Some technical details: To extract the variables of interest of the `NTuples`, two `ROOT` macros were written, which were repeatedly called for the different models by a python script. To calculate the truth-level CL_s values, the same `HistFitter` configuration file as in section 6.4 was used (with slight extensions to account for the increased number of input files).
Due to the high computation time, a combination of python and bash scripts was written to employ the batch system of the local computer network, providing around 70 cores to be used by the analysis. To save disk space, only the logfiles of the `HistFitter` runs were kept. A combination of bash and python scripts was written to extract the CL_s values and join them in text files together with the reco-level CL_s values from the reduced summary `NTuples`. Finally, several python scripts were written for the graphical representation of the data points.

as comparison. As can be seen in figure 6.12 a change in the systematic uncertainties can not resolve the issue.

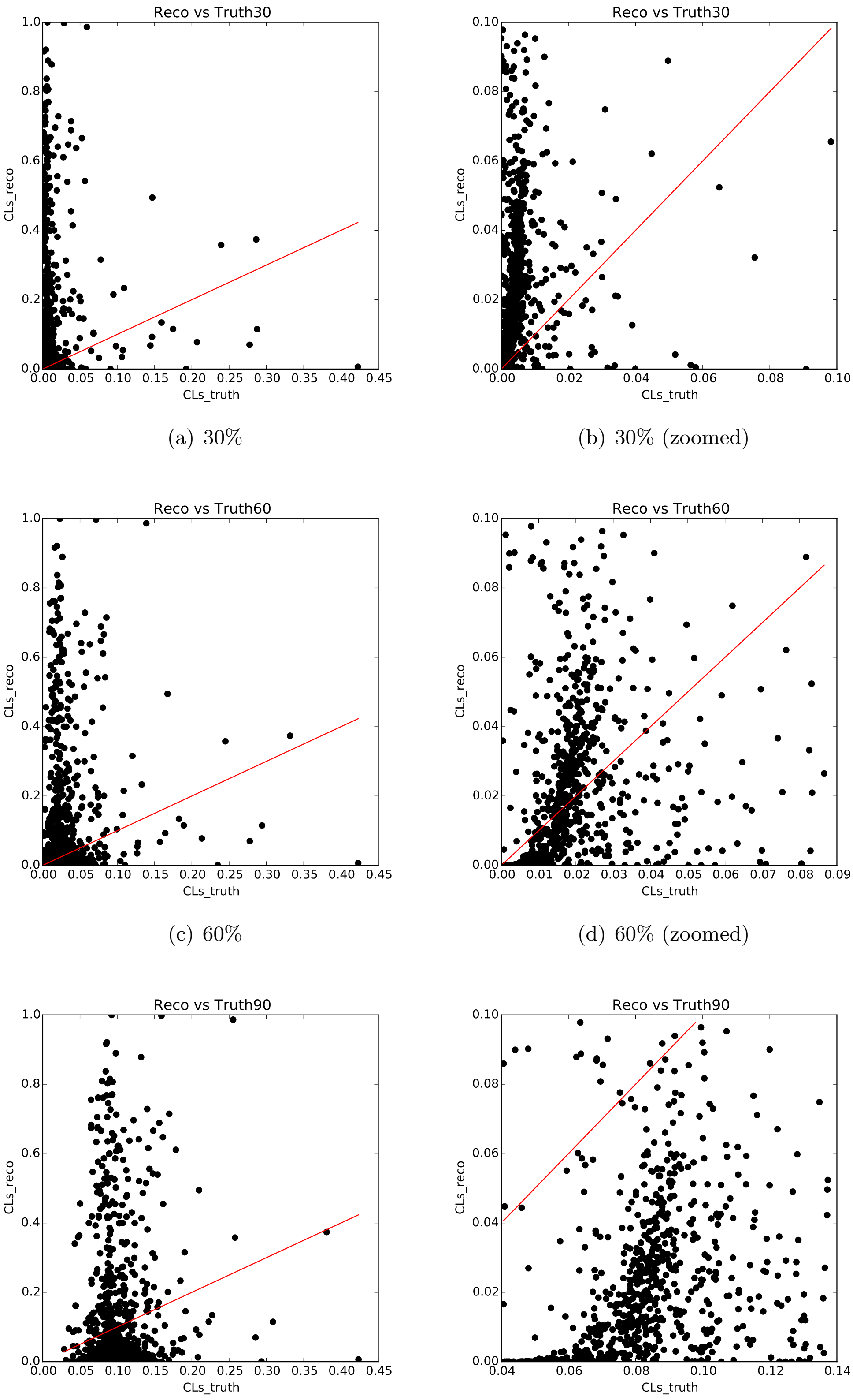
6.5.3 The influence of scaling factors

Applying additional scaling factors < 1 to the truth-level sample results in raising CL_s values. The calculation of the scale factors was based on the full set of models and is therefore more accurate than the estimation of the systematic uncertainties, but the sensitivity of the analysis to the scaling factors in interesting nonetheless. The outcome is shown in figures 6.14(e) and 6.14(f). The effect is similar to the the effect of changed systematics and does not improve the matching of the reco and truth CL_s values much.

6.5.4 Other sources of error

As already mentioned, the reco-level CL_s values used in the previous plots are the minimum of the CL_s values obtained in (reco-level) hard *and* soft one lepton studies but are compared to the truth-level CL_s values of only the hard one lepton analysis. However, this results in reco-level CL_s values that are *smaller* than they should be, so this systematic error (which was assumed to be small anyway) can not amend the mismatch.

Thus, it seems most likely that the general agreement of truth- and reco-level signal histograms is too low to allow for consistent truth-level based sensitivity estimations.

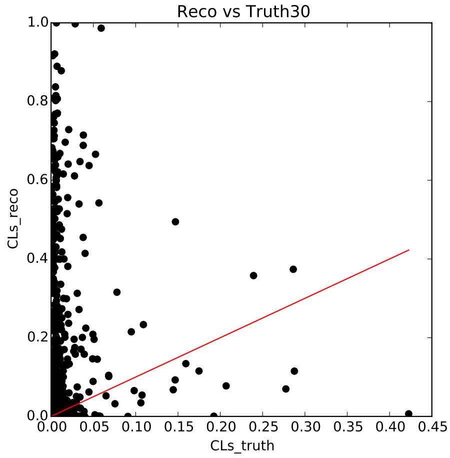


(a) 30% (b) 30% (zoomed)

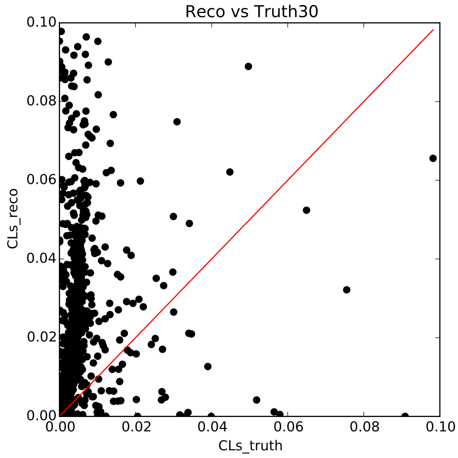
(c) 60% (d) 60% (zoomed)

(e) 90% (f) 90% (zoomed)

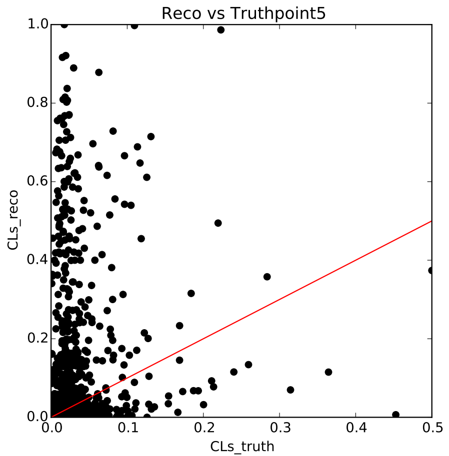
Figure 6.12: CL_s values of Reco and Truth with varying systematic uncertainties. The red line has slope 1. The computation time of the data for this figure amounts to a total of about 150 hours. The calculation of the truth CL_s value of a specific value was aborted if the computation time exceeded 5 minutes (average is about 2.5 minutes). This resulted in aborted calculations for less than 1% of the models.



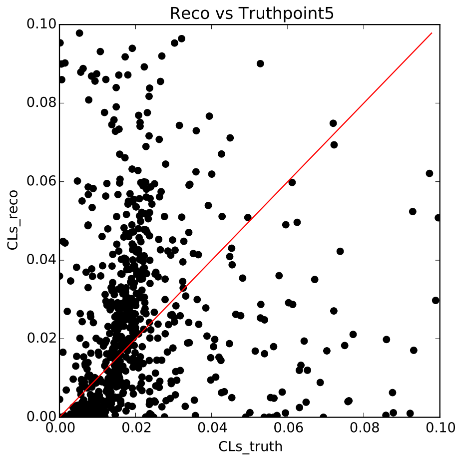
(a) Standard scaling factors



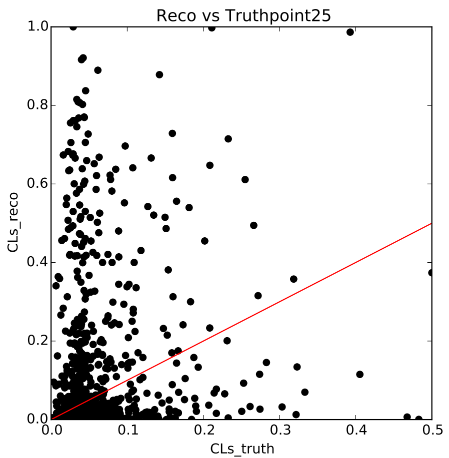
(b) Standard scaling factors (zoomed)



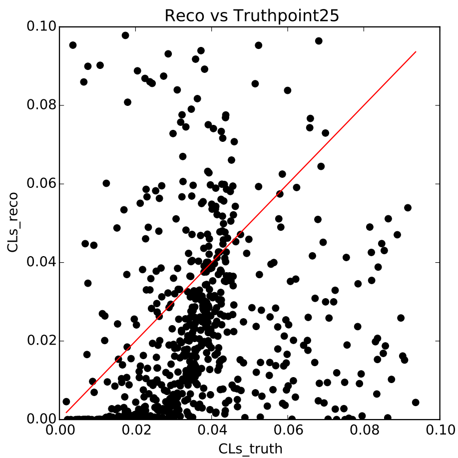
(c) 0.5×Standard scaling factors



(d) 0.5×Standard scaling factors (zoomed)

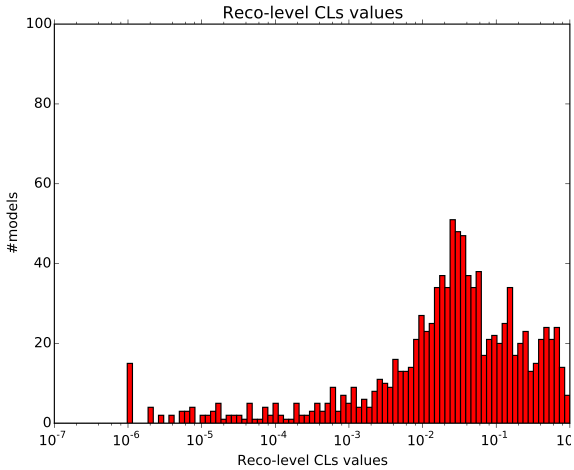


(e) 0.25×Standard scaling factors

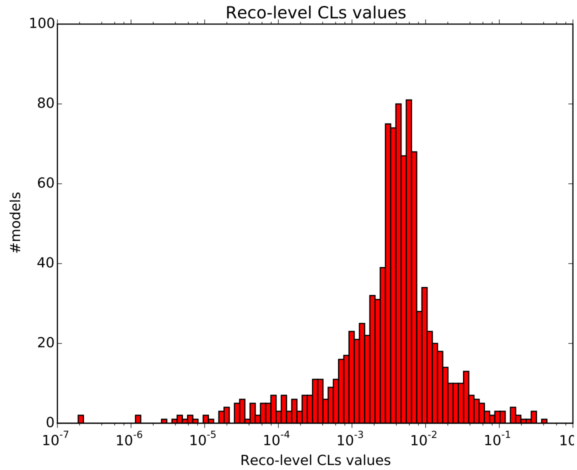


(f) 0.25×Standard scaling factors (zoomed)

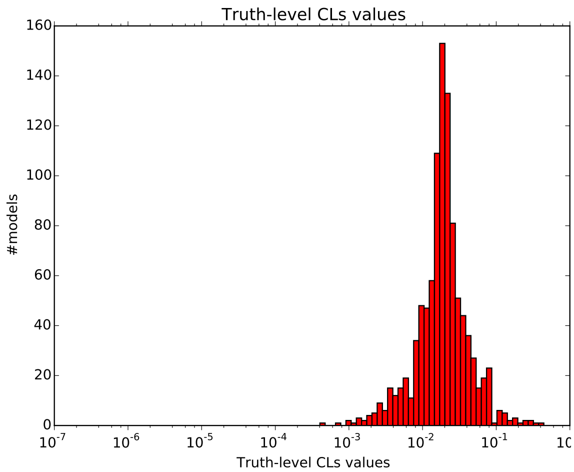
Figure 6.13: The influence of scaling factors on the reco and truth-level CL_s values (with 30% uncertainty on the truth samples). The red line has slope 1. The computation time for the data in this figure amounts to a total of about 150 hours. The calculation of the truth CL_s value of a specific value was aborted if the computation time exceeded 5 minutes (average is about 2.5 minutes). For the 0.25 scaling factors this resulted in aborted calculations for 6.6% of the models, in all other cases the number was below 1%.



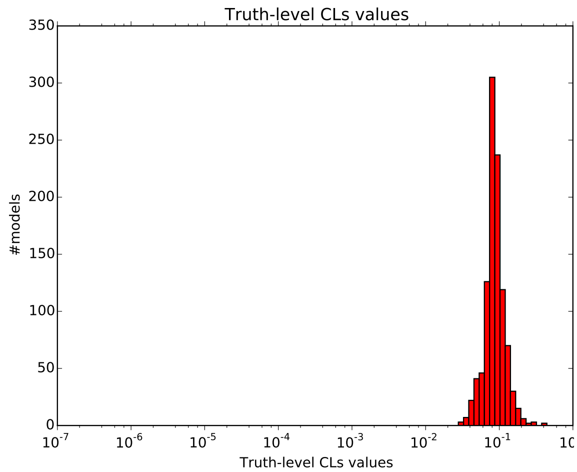
(a) Reco



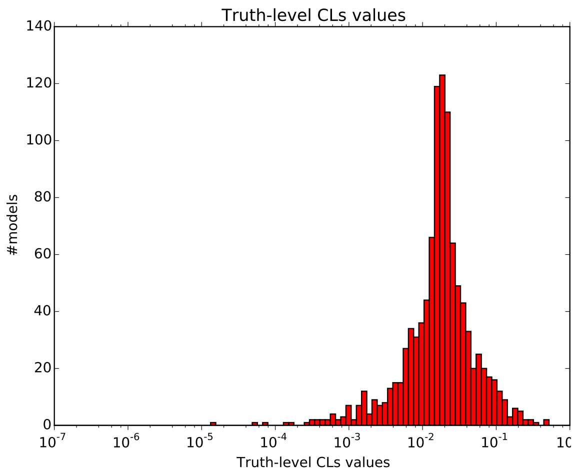
(b) Truth (30% systematics, standard scaling)



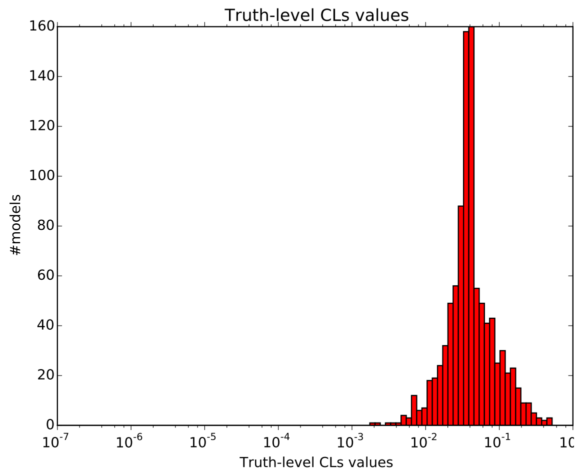
(c) Truth (60% systematics)



(d) Truth (90% systematics)



(e) Truth (0.5×standard scaling)



(f) Truth (0.25×standard scaling)

Figure 6.14: Histograms of truth- and reco-level CL_s values as obtained by “projection” of the plots 6.12 and 6.13 to the x-Axis (truth) resp. y-Axis (reco). The same notes regarding aborted calculations apply.

Chapter 7

Summary and Conclusion

An estimation of the sensitivity to phenomenological MSSM models using the signal shape of truth-level signal samples was presented. An initial set of 9 pMSSM models was used for initial tests. Systematic uncertainties and scaling factors were estimated. A comparison between the truth- and reco-level signal histograms showed varying levels of agreement, ranging from almost perfect to essentially non-existent. The corresponding truth- and reco-level CL_s values were calculated and compared, showing mixed results.

The study was continued with a set of 1043 pMSSM models. The agreement of truth- and reco-level CL_s values was found to be generally very low, with the majority of the truth level CL_s values being far smaller than the corresponding reco-level CL_s values. Increased systematic uncertainties on the truth samples were tested, resulting in increased truth-level CL_s values but this did not resolve the mismatch. Similarly, additional scaling factors < 1 were discussed, also yielding higher truth-level CL_s values but not resulting in an improved agreement of both analysis strategies.

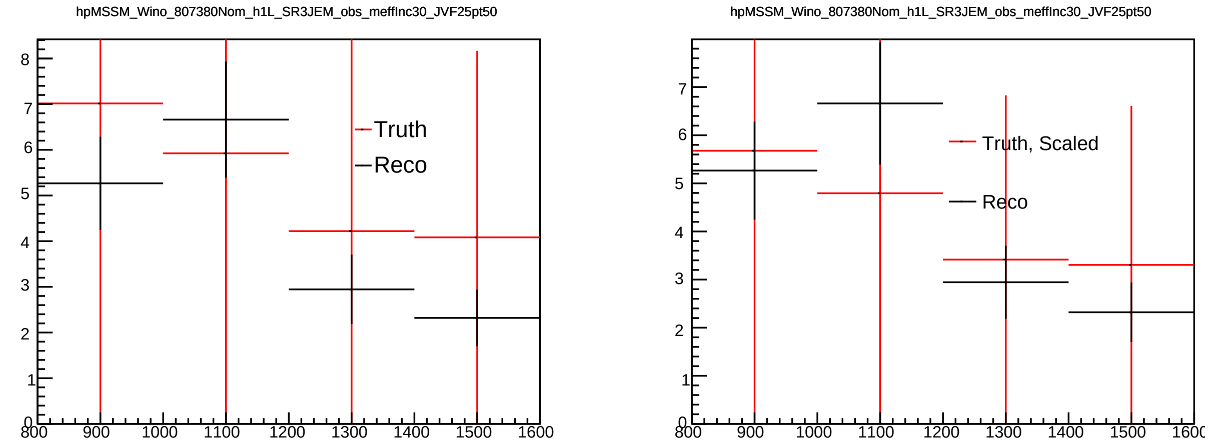
As the tests on the 9 pMSSM models showed agreeing truth/reco CL_s values for the model with nearly matching truth- and reco-level signal histograms, it is assumed that the fraction of models with equally matching truth/reco-level signal histograms is very low. However a more thorough inspection of the corresponding histograms will be needed to provide more quantitative results. By studying subsets of models that fulfill certain requirements on the similarity of reco/truth signal shapes or overall signal yields, quantitative criteria that lead to similar reco/truth CL_s values could be extracted.

Most importantly however, the reason for the truth/reco mismatch needs to be found: Since plots like 6.2 (with but few cuts on the considered region) show good truth/reco agreement, the effects of the additional cuts that lead to the definition of the signal regions should to studied. Subsequently, it can be tried to improve the agreement with the employment of additional scaling or smearing factors.

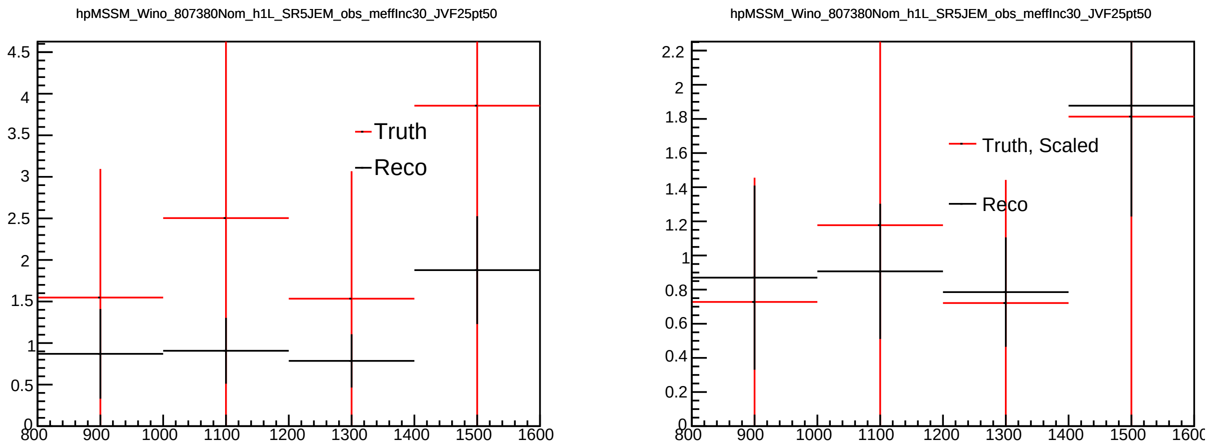
Chapter 8

Comparison of signal shapes between truth and reco samples

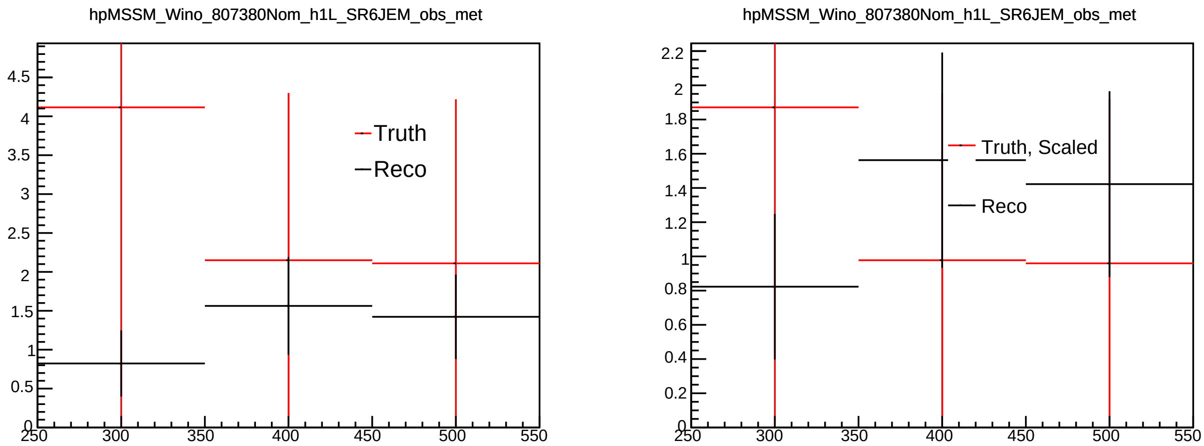
This chapter shows the remaining plots that compare the signal event yields between truth and reco samples. Note that no point/error bar is drawn for empty bins. In the right handed plots, the truth sample was scaled to have the same integral as the reco sample. For empty reco samples, the right handed plots thus remain empty.



(a) SR3J

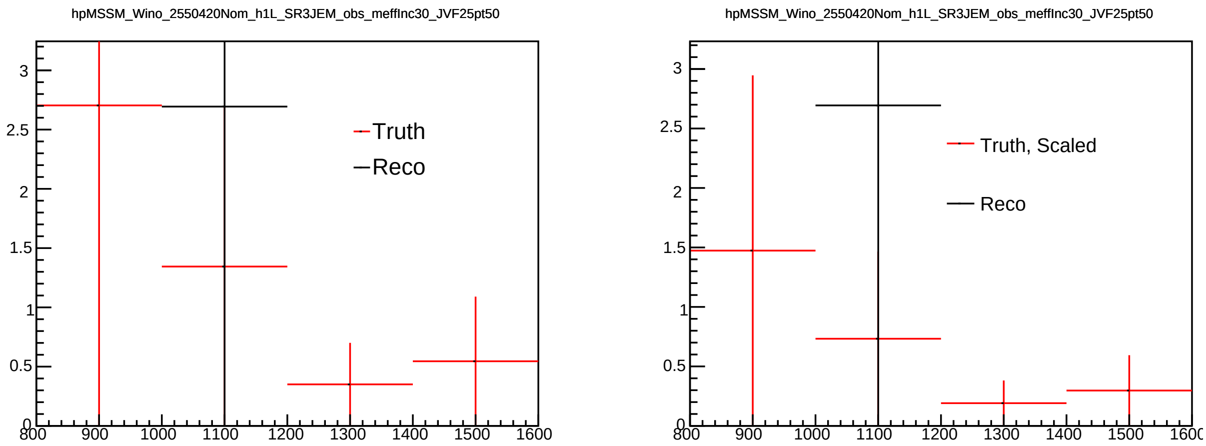


(b) SR5J

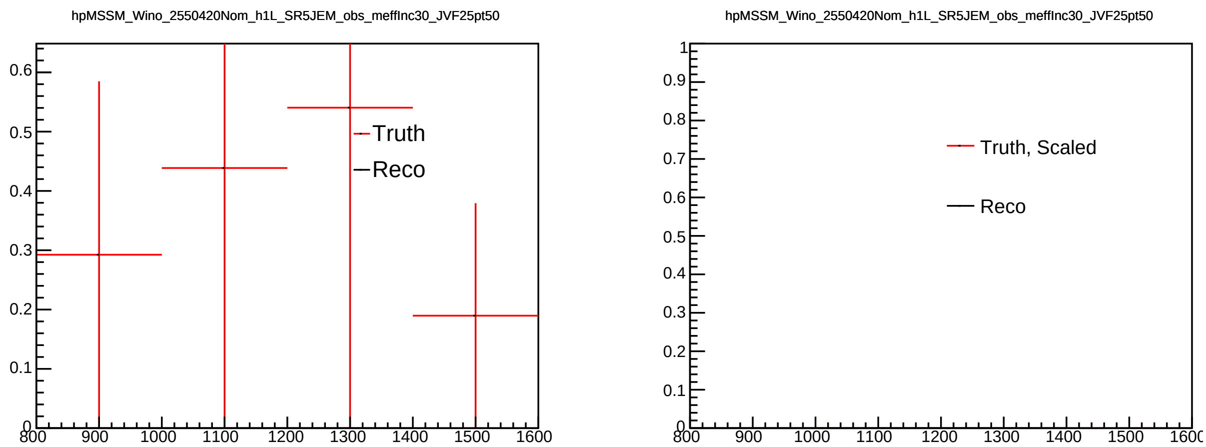


(c) SR6J

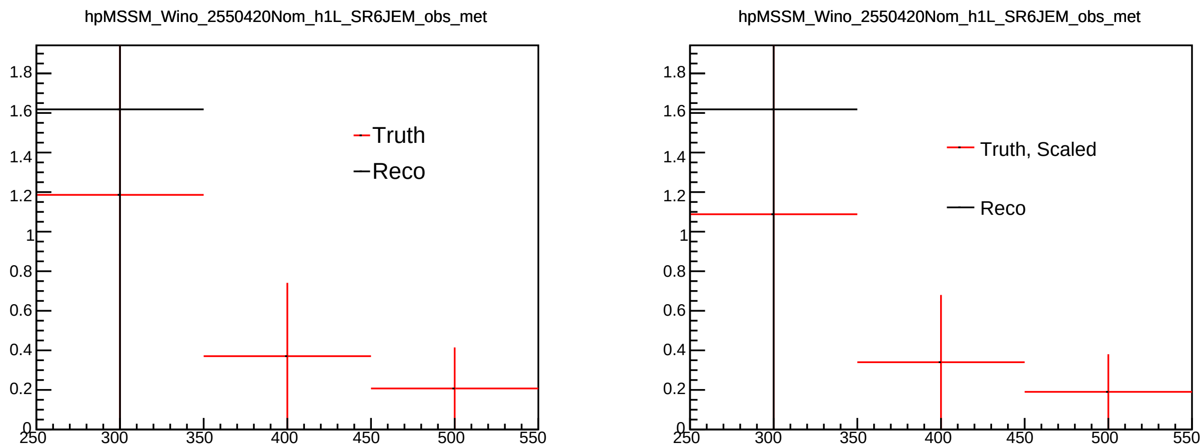
Figure 8.1: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_807380`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J

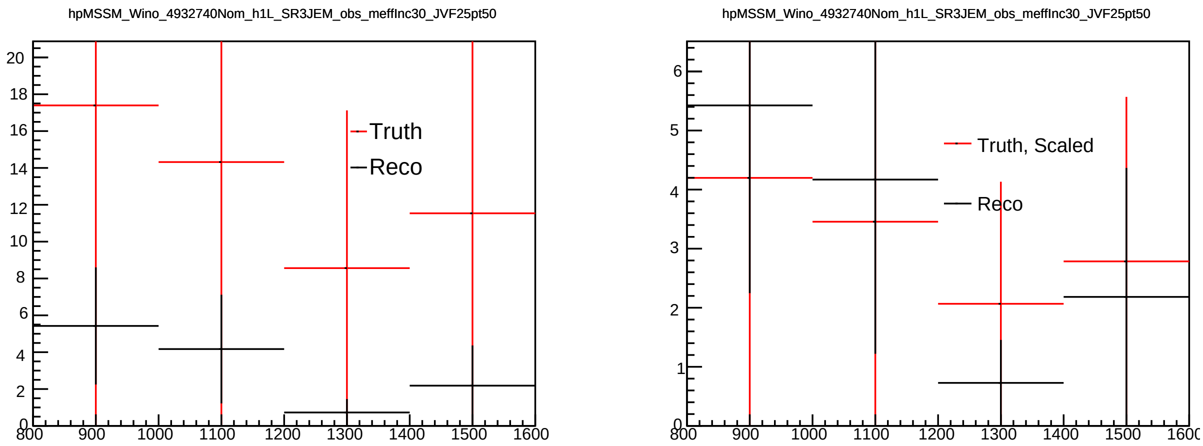


(b) SR5J

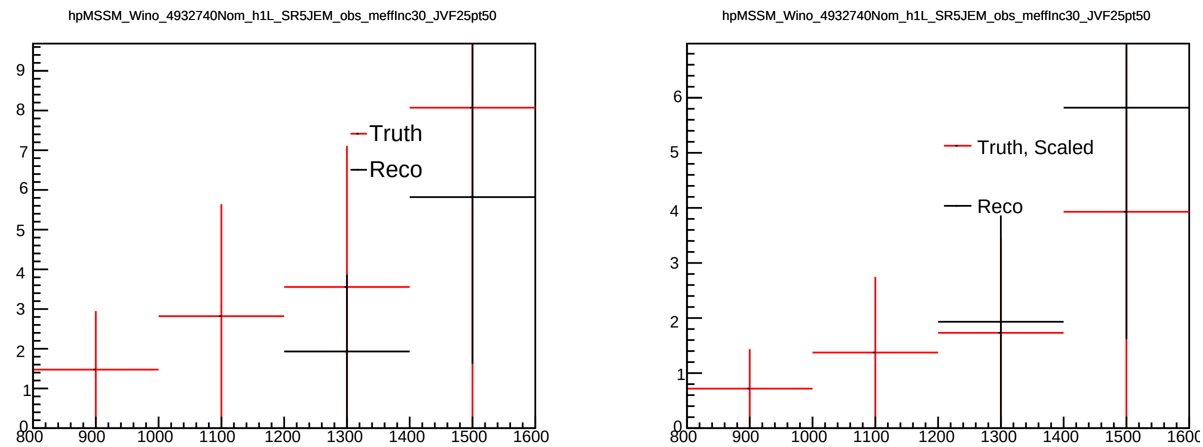


(c) SR6J

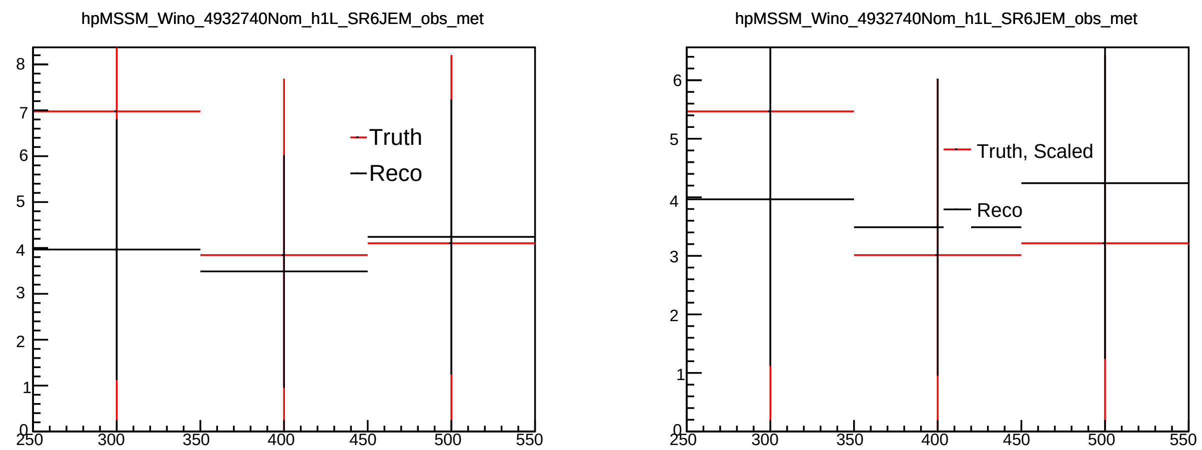
Figure 8.2: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_2550420`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J

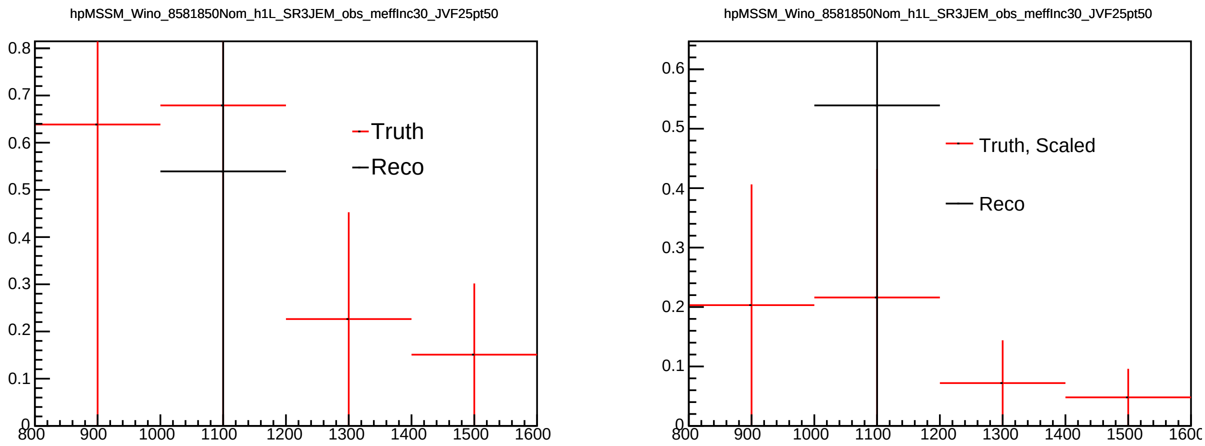


(b) SR5J

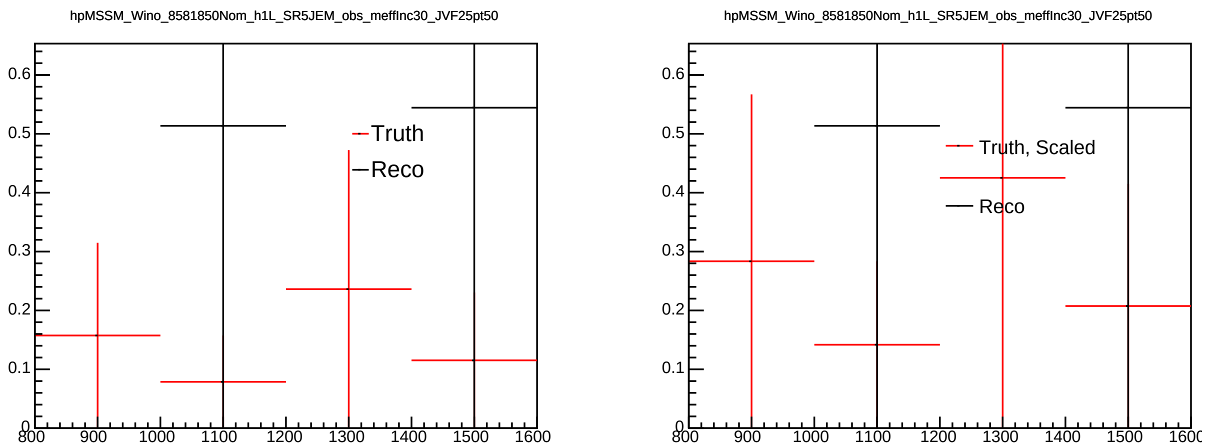


(c) SR6J

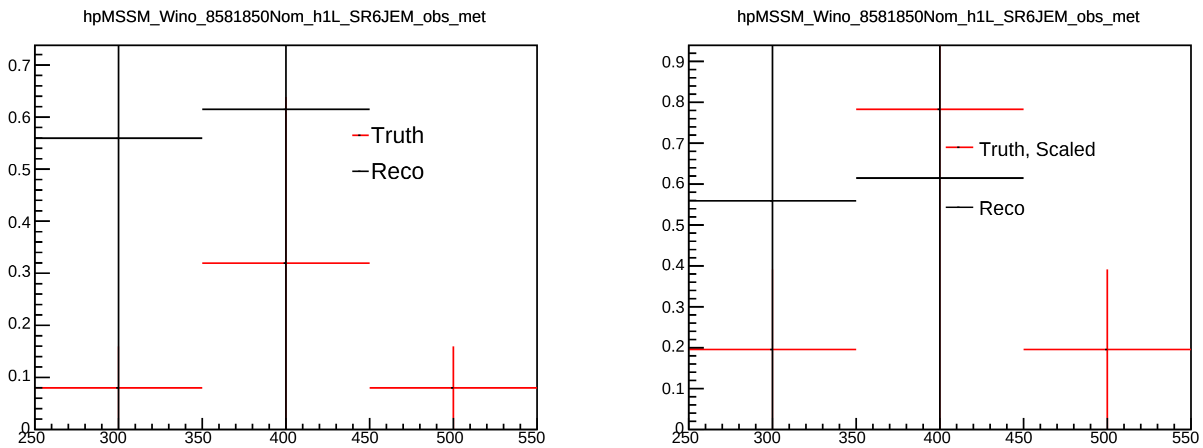
Figure 8.3: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_4932740`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J

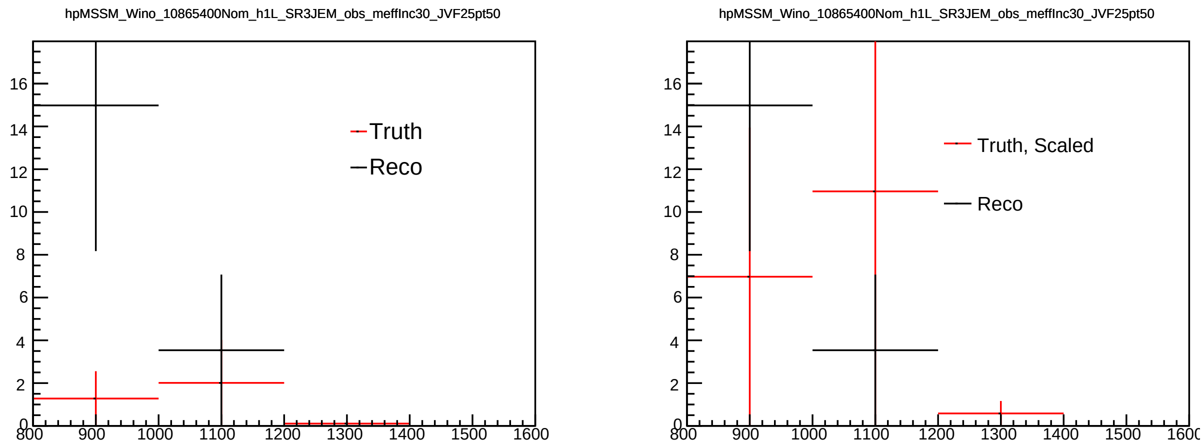


(b) SR5J

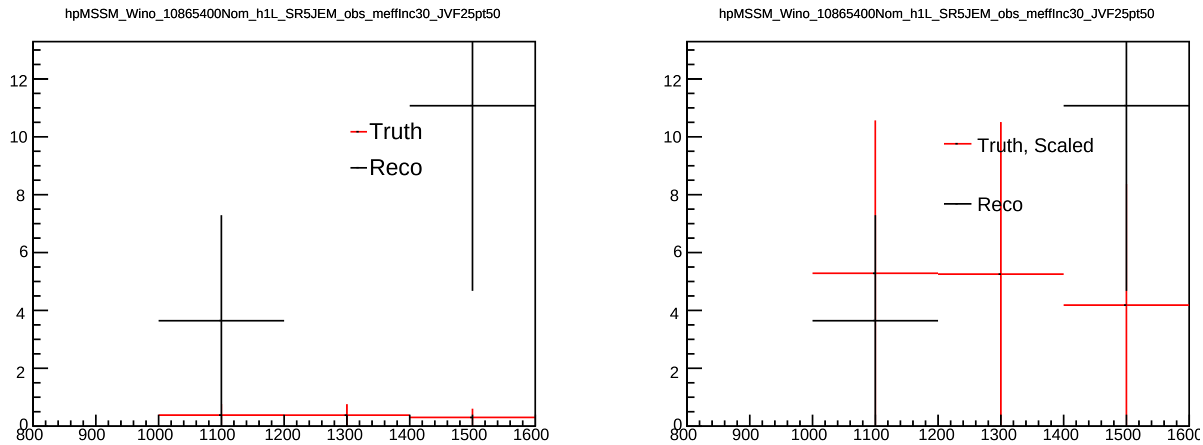


(c) SR6J

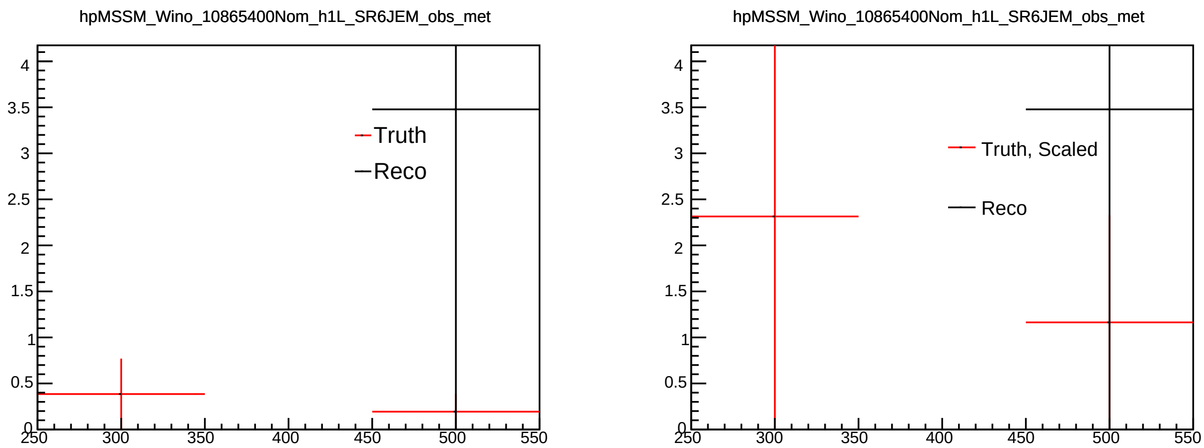
Figure 8.4: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_8581850`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J

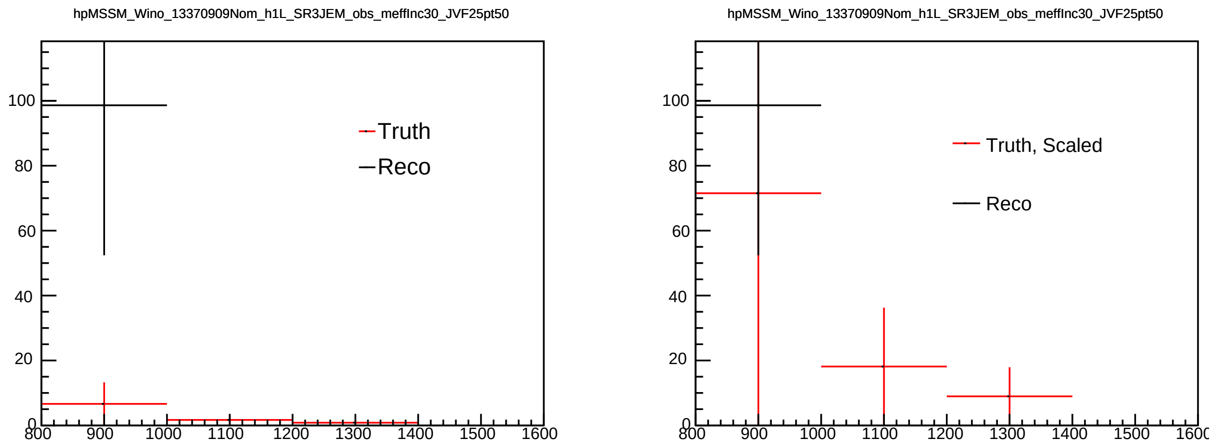


(b) SR5J

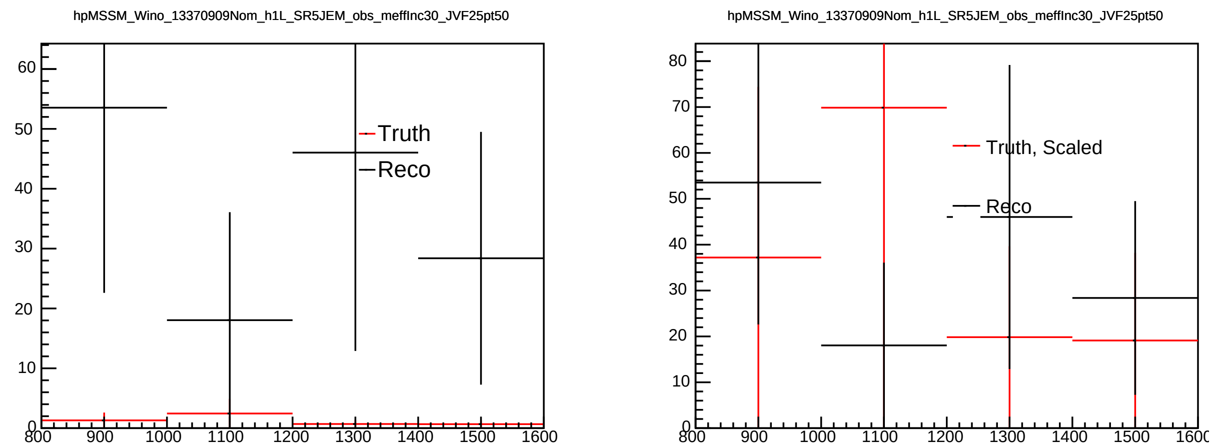


(c) SR6J

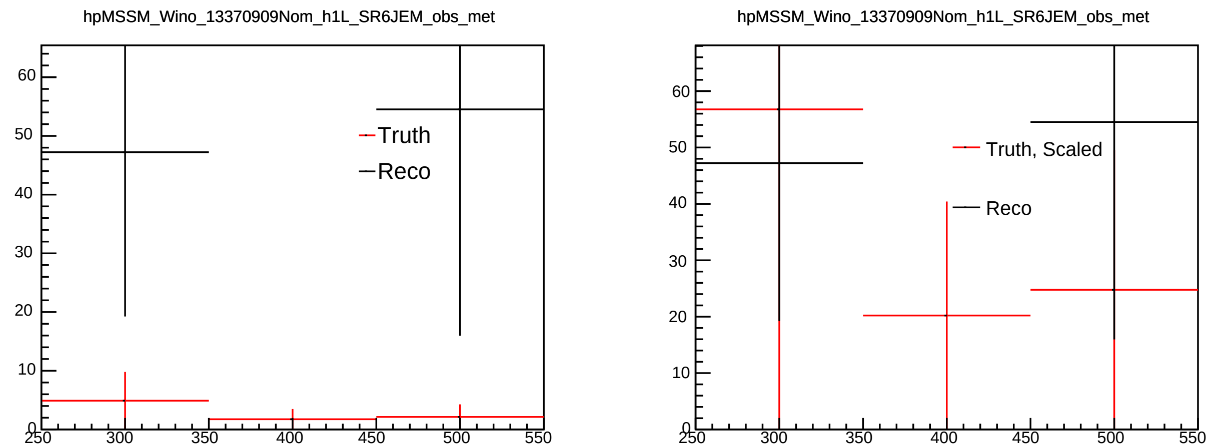
Figure 8.5: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_10865400`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.



(a) SR3J



(b) SR5J



(c) SR6J

Figure 8.6: Comparison of signal event yields between truth and reco samples for the signal model `pMSSM_Wino_13370909`. On the right hand side, the signal is normalized in a way that the truth and reco signal have the same integral.

List of Figures

List of Tables

Bibliography

[1] S. P. Martin. *A Supersymmetry Primer*. Sept. 2011. arXiv: hep-ph/9709356 [hep-ph].

[2] S. Perrenod. *Planck 2015. Constraints on Dark Energy and Inflation*. Apr. 2015. URL: http://darkmatterdarkenergy.com/2015/04/06/planck-2015-constraints-on-dark-energy-and-inflation/.

[3] M. Lamont. "The LHC's first long run". In: *CERN Courier* (Aug. 2013). URL: http://cerncourier.com/cws/article/cern/54381.

[4] K. A. Ulmer. "Future Sensitivity Studies for Supersymmetry Searches at CMS at 14 TeV". In: (2013). arXiv: 1310.0781 [hep-ex].

[5] CERN. *LHC Season 2. A stronger machine*. Mar. 2015. URL: http://press.web.cern.ch/sites/press.web.cern.ch/files/file/backgrounder/2015/03/a_stronger_machine-en.pdf.

[6] A. Djouadi et al. *The Minimal Supersymmetric Standard Model: Group Summary Report*. Report of the MSSM working group for the Workshop "GDR-Supersymmetrie", France. Jan. 1999. arXiv: hep-ph/9901246 [hep-ph].

[7] Wikimedia Commons. *Standard Model of Elementary Particles.svg*. Oct. 2014. URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.

[8] J. Beringer et al. *2013 partial update for the 2014 edition*. 2013. URL: http://pdg.lbl.gov/2013/tables/rpp2013-sum-quarks.pdf.

[9] J. Beringer et al. "Review of Particle Physics". In: *Phys. Rev. D* 86 (1 July 2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.

[10] Wikimedia Commons. *Elementary particle interactions in the Standard Model*. May 2014. URL: https://en.wikipedia.org/wiki/File:Elementary_particle_interactions_in_the_Standard_Model.png.

[11] S. J. Gates, Jr. *Sticking with SUSY*. 2014. URL: http://live.iop-pp01.agh.sleek.net/2014/09/25/sticking-with-susy/.

[12] A. Gladyshev and D. Kazakov. *IS (Low Energy) SUSY STILL ALIVE?* Dec. 2012. arXiv: hep-ph/1212.2548 [hep-ph].

[13] J. R. Ellis. "Beyond the standard model for hill walkers". In: (1998). arXiv: hep-ph/9812235 [hep-ph].

[14] M. Cahill-Rowley, J. Hewett, A. Ismail, and T. Rizzo. *pMSSM Studies at the 7, 8 and 14 TeV LHC*. Sept. 2013. arXiv: 1307.8444 [hep-ph].

[15] CERN. *The Large Hadron Collider*. 2015. URL: http://home.web.cern.ch/topics/large-hadron-collider.

[16] L. Evans and P. Bryant. "LHC Machine". In: *Journal of Instrumentation* (Aug. 2008). URL: http://iopscience.iop.org/1748-0221/3/08/S08001.

[17] CERN. *Restarting the LHC: Why 13 TeV?* June 2014. URL: http://home.web.cern.ch/about/engineering/restarting-lhc-why-13-tev.

[18] M. Lamont. *LHC Operation: Past and Future (3/3)*. Academic Training Lecture Regular Programme. Sept. 2013. URL: https://indico.cern.ch/event/266732/material/slides/1.pdf.

[19] CERN. *LHC experiments are back in business at a new record energy*. June 2015. URL: http://press.web.cern.ch/press-releases/2015/06/lhc-experiments-are-back-business-new-record-energy.

[20] CERN. *Experiments. Range of experiments at CERN investigate physics from cosmic rays to supersymmetry*. 2015. URL: http://home.web.cern.ch/about/experiments.

[21] CERN. *ALICE. ALICE detects quark-gluon plasma, a state of matter thought to have formed just after the big bang*. 2015. URL: http://home.web.cern.ch/about/experiments/alice.

[22] CERN. *LHCb. The LHCb experiment will shed light on why we live in a universe that appears to be composed almost entirely of matter, but no antimatter*. 2015. URL: http://home.web.cern.ch/about/experiments/lhcb.

[23] CERN. *TOTEM. The "Total, elastic and diffractive cross-section measurement" experiment studies particles thrust forward by collisions in the LHC*. 2015. URL: http://home.web.cern.ch/about/experiments/totem.

[24] N. U. Solar-Terrestrial Environment Laboratory. *LHC forward experiment*. 2015. URL: http://www.stelab.nagoya-u.ac.jp/LHCf/LHCf/index.html.

[25] CERN. *The MOEDAL experiment is looking for a hypothetical particle with magnetic charge: the magnetic monopole*. 2015. URL: http://home.web.cern.ch/about/experiments/moedal.

[26] CERN. *LHC – the guide*. 2009. URL: https://cdsweb.cern.ch/record/1165534/files/CERN-Brochure-2009-003-Eng.pdf.

[27] CERN. *The Proton Synchrotron Booster*. Accessed 20 June 2015. URL: http://home.web.cern.ch/about/accelerators/proton-synchrotron-booster.

[28] S. Redaelli. *The operation of the LHC accelerator complex (1/2)*. LPCC student lectures. Accessed 19 June 2015. Apr. 2010. URL: http://indico.cern.ch/event/87719/material/slides/2.pdf.

[29] ATLAS Collaboration. 2015. URL: http://www.atlas.ch/photos/atlas_photos/selected-photos/full-detector/ATLAS_Silver_White_MK.jpg.

[30] ATLAS. *What is ATLAS*. 2015. URL: http://atlas.ch/what_is_atlas.html.

[31] ATLAS, CMS Collaborations. *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*. Mar. 2015. arXiv: 1503.07589 [hep-ex].

[32] CERN. *ATLAS. From a cavern 100 metres below a small Swiss village, the 7000-tonne ATLAS detector is probing for fundamental particles*. 2015. URL: http://home.web.cern.ch/about/experiments/atlas.

[33] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[34] Wikimedia Commons. URL: https://upload.wikimedia.org/wikipedia/commons/3/30/Pseudorapidity2.png.

[35] J. Lorenz. "Search for strongly interacting supersymmetric particles decaying to final states with an isolated lepton with the ATLAS detector at the LHC". PhD thesis. Ludwig-Maximilians-Universität München, Jan. 2014. URL: http://www-static.etp.physik.uni-muenchen.de/dokumente/thesis/phd_jlorenz.pdf.

[36] R. K. Bock. *Calorimeter.* Accessed 23 Jun 2015. Apr. 1998. URL: http://rd11.web.cern.ch/RD11/rkb/PH14pp/node19.html.

[37] ATLAS. URL: http://www.atlas.ch/photos/atlas_photos/selected-photos/inner-detector/combined/0803014_01-A4-at-144-dpi.jpg.

[38] URL: http://photo.sf.co.ua/g/7/11.jpg.

[39] URL: https://makanaka.files.wordpress.com/2010/11/rg-cern-0511013_01.jpg.

[40] ATLAS. URL: http://atlas-cs.web.cern.ch/atlas-cs/images/atlas-cs.jpg.

[41] ATLAS. URL: http://www.atlas.ch/photos/atlas_photos/selected-photos/calorimeters/combined-barrel/0803015_01-A4-at-144-dpi.jpg.

[42] ATLAS Collaboration. *Performance of the ATLAS Trigger System in 2010.* Dec. 2011. arXiv: 1110.1530 [hep-ex].

[43] M. zur Nedden. *The ATLAS Trigger System.* Oct. 2011. URL: https://www.physik.hu-berlin.de/de/gk1504/events/blockcourse_10_11/23_2011-10_zurneddenmartin.pdf.

[44] CERN. *The Worldwide LHC Computing Grid.* Accessed 23 Jun 2015. URL: http://home.web.cern.ch/about/computing/worldwide-lhc-computing-grid.

[45] CERN. *The Grid: A system of tiers.* Accessed 23 Jun 2015. URL: http://home.web.cern.ch/about/computing/grid-system-tiers.

[46] M. Baak, G.-J. Besjes, J. Lorenz, and S. Pataraia. *Introduction to statistics / statistical tools & the HistFitter framework.* Apr. 2015. URL: http://www-static.etp.physik.uni-muenchen.de/kurs/comp15/docs/HistFitter.pdf.

[47] M. Baak, G. Besjes, D. Cote, A. Koutsman, J. Lorenz, and D. Short. *HistFitter software framework for statistical data analysis.* Oct. 2014. arXiv: 1410.1280 [hep-ex].

[48] J. Lorenz. *HistFitter Tutorial.* 2015. URL: https://twiki.cern.ch/twiki/bin/view/Main/HistFitterTutorialOutsideAtlas.

[49] The HistFitter Group. *HistFitter. A software framework for statistical data analysis.* URL: http://histfitter.web.cern.ch/histfitter/.

[50] H.-C. Schultz-Coulon. "Monte Carlo Generators. Simulating high energy pp-collisions ..." Nov. 2010. URL: http://www.kip.uni-heidelberg.de/~coulon/Lectures/LHC/Free_PDFs/Lecture6.pdf.

[51] V. Erkcanözcan. *Monte Carlo Simulation Techniques.* ISTAPP 2011. Feb. 2011. URL: http://indico.cern.ch/event/115514/contribution/41/material/slides/0.pdf.

[52] G. Besjes. *HistFitter: a flexible framework for statistical data analysis.* 21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015), Okinawa (Japan). Apr. 2015. URL: http://histfitter.web.cern.ch/histfitter/Files/HistFitter_CHEP2015_150316.pdf.

[53] ATLAS Collaboration. *Further search for supersymmetry at $\sqrt{s} = 7$ TeV in final states with jets, missing transverse momentum and isolated leptons with the ATLAS detector.* Aug. 2012. arXiv: 1208.4688 [hep-ex].

[54] ATLAS Collaboration. *Search for squarks and gluinos in events with isolated leptons, jets and missing transverse momentum at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*. May 2015. arXiv: 1501.03555 [hep-ex].

[55] C. Amsler et al. “Review of Particle Physics: Kinematics”. In: *Physics Letters B667* (June 2008). URL: http://pdg.lbl.gov/2008/reviews/kinemarpp.pdf.

[56] B. Abbott, M. Abolins, B.S. Acharya, et al. *A Measurement of the W Boson Mass*. Dec. 1997. arXiv: hep-ex/9712029 [hep-ex].

[57] CDF group. Accessed 22 Jun 2015. 2013. URL: http://www.phys.sinica.edu.tw/~smwang/cdf/physics/sm_higgs/pic/zhvvbb_candidate_event_display.gif.

[58] ATLAS Collaboration. *Interpretation of ATLAS results in phenomenological MSSM*. Draft version 0.85, ATL-COM-PHYS-2014-952. July 2015.

[59] V. Tudorache. *One Lepton Analysis in pMSSM*. Mar. 2015.

[60] CERN. *What is ROOT?* URL: https://root.cern.ch/drupal/content/about.

[61] B. Petersen. Accessed Jul 08 2015. URL: https://cernbox.cern.ch/public.php?service=files&t=41952654dbfe7e12991ac47722f32966.

[62] CERN. *class TNtuple*. URL: https://root.cern.ch/root/html/TNtuple.html.