Contents

Chapter 1

Search for neutral MSSM Higgs Bosons in the $A/h/H \to \tau^+\tau^- \to e\mu + 4\nu$ decays

Under the light of the recent discovery of a Higgs boson with mass of 125 GeV [], remains an open question wheter this new particle constitute all the pieces of the Higgs sector or wheter it is only one of several bosons predicted in some theories that go beyond the SM. The most recent measurements [] of its properties shows it to be, within experimental uncertainties, perfectly compatible with the SM Higgs boson, however such a new particle can be accomodated within several beyond the standard model (BSM) theories, this is particularly true for Super Symmetry.

There are two approach to explore the Higgs sector: one can study the cupling of the Higgs boson with vector bosons and fermions, those measure in fact are sensitive to new physics and can determine if this particle is fully responsible for the generation of all the SM particles masses. Another approach is to directly search for for additional Higgses in a well defined model, which is the approach followed in this thesis where new neutral bosons are sought within the MSSM (see chapter ??).

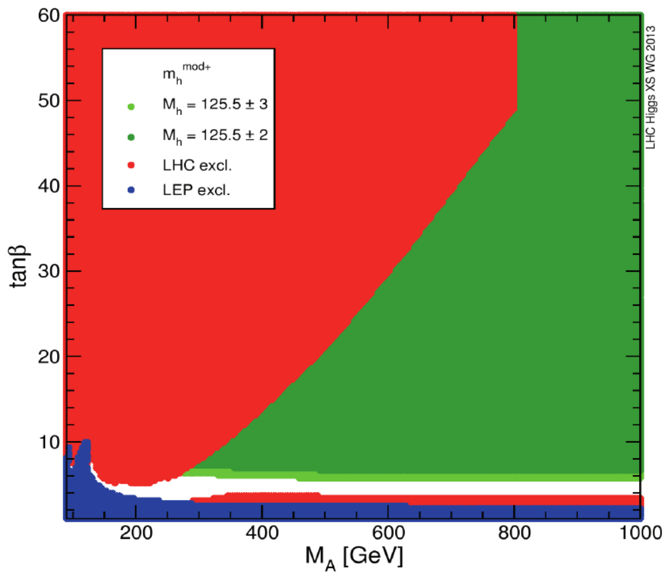


Figure 1.1: Excluded and allowed regions of the $m_A - \tan\beta$ parameter space for the m_h^{mod+} benchmark scenario. Excluded region are determined based on direct Higgs boson searches at LEP (blue) and LHC (red). The two green bands corresponds to the parameter regions which are compatible with the assumption that the lightest Higgs boson, h, has a mass respectively of $M_h = 125.5 \pm 2$ or ± 3 GeV. For more detail see [34].

1.1 Introduction

In the Minimal Supersymmetric extension of the Standard Model (MSSM) [7, 8] the Higgs sector is composed of two Higgs doublets of opposite hyper-charge, resulting in five observable Higgs bosons. Two of these Higgs bosons are neutral and CP-even (h,H), one is neutral and CP-odd (A) and two are charged ($H^{\pm}$). At tree level their properties such as masses, widths and branching ratios can be predicted in terms of only two parameters, often chosen to be the mass of the CP-odd Higgs boson m_A, and the ratio of the vacuum expectation values of the two Higgs doublets $\tan\beta$ (for more details see chapter **??**).

The MSSM predicts the existence of a Higgs boson with properties that resemble those of a SM Higgs boson in large regions of its parameter space. It is technically impossible, for an experimental search, to explore the full parameter space of the model, several benchmark scenarios are then introduced by fixing the parameters to values typical for most interesting physics scenarios. With the recent Higgs boson discovery, bechmark scenarios of the MSSM have been updated to accomodate for new experimental constraints (see chapter **??**). As an example, Figure 1.1 shows the currently excluded and allowed regions of the MSSM parameter space for the m_h^{mod+} updated benchmark scenario. In this scenario a supersymmetric SM-like Higgs boson is assumed, large region of the $m_A - \tan\beta$ parameter space is compatible with this assumption and remains still unexplored.

This chapter presents the search for the neutral MSSM Higgs bosons decaying into pairs of tau leptons in the fully leptonic final state. The search is based on 20.3 fb^{-1} of 8 TeV data recorded by the ATLAS experiment during 2012 at the Large Hadron Collider (LHC) [1]. Higgs boson candidated events are selected based on the topological properties of the Higgs boson production and decay. The two dominant Higgs boson production modes are via gluon-gluon fusion and in association with b-quarks, the search take advantage of that being performed separately for events without or with b-tagged jets in the final state, respectively. Two of the dominat background contributions, $Z \to \tau\tau$ and multi-jet process, are predicted with the corresponding signal-depleted control data samples. The final statistical interpretation of the data is based on the comparison of the observed invariant mass distributions, exclusion limits are set by means of a binned likelihood ratio test statistic and interpreted either in the MSSM m_h^{max} scenario and in a model independed.

In the MSSM for large region of parameter space one found that one of the CP-even neutral Higgses has properties that resemble the one of the SM Higgs, this is usually the case for the lightest Higgs, h, the other two, H and A, tend to be degenerate in mass and decouple from gauge bosons. An interesting fenomenological consequence is that the coupling of the latter two Higgses with down (up) type fermions are enhanced (suppressed) by $\tan\beta$, meaning that for large $\tan\beta$ bottom-quark and τ lepton will play a more important role than in the SM case either for production and decay.

The production of the neutral CP-even MSSM Higgs bosons at hadron colliders proceeds via the same processes as for the SM Higgs production. However, the pseudoscalar A instead cannot be produced in association with gauge bosons or in vector boson fusion (VBF) at tree-level, as this coupling is forbidden due to CP-invariance. At the LHC one of the most relevant production mechanisms for the MSSM Higgs bosons is gluon-gluon fusion, $gg \to A/H/h$. In addition, the production in association with b-quarks becomes important for large value of $\tan\beta$. Those are the two production mechanism that are considered in this analysis, Figure 1.2 shows the Feynman-diagram for those processes, while Figure 1.3 shows the production cross section of the neutral MSSM Higgses via these two processes. The search is divided in two category which are optimized for the two different production mode considered, in the gluon-fusion category is requred a b-jet veto (for definition of b-tagging algorithm see chapter **??**), in fact no b-jet in the final state are present for this production mode. In contrast a a b-jet tag is required for b-associated production, this category is expected to be very sensitive to $\tan\beta$. The two category are ortogonal and present different backgrounds contributions, which can be optimized separately.

The decays of the neutral MSSM Higgs bosons (in the assumption that all supersymmetric particle are heavy enough) are the same as for the SM one with the already cited exception of A. Figure 1.3 shows the decay branching fractions for H and A as a function of the mass, the decay into tau pair is the most important after $b\bar{b}$ and the one used in this analysis. The decay channel in $b\bar{b}$ is in fact very challenging due to the huge background from QCD multi-jet. In this thesis only cases in which the taus decay one in $e + 2\nu$ and the other in $\mu + 2\nu$ are considered, This final state corresponds to a total $\tau^+\tau^-$ branching ratio of approximately 6%.

The signal topology described in the previous section is common to many other processes, unfortunately, those have higher cross section than the sought signal and a set of additional selections is needed to enhance the sensitivity of the search. The most important backgrounds to this search are the production of $Z \to \tau\tau$ + jets, the top quark ($t\bar{t}$ and single top production is intended), diboson production (like WW or ZZ events) and Drell-Yan process or events with non-prompt leptons coming solely from hadron decay (in short QCD multi-jet). Vector bosons production like $W \to \ell\nu$ or $Z \to \ell\ell$ + jets (with ℓ here meaning either e or μ) are also considered, however those processes have a limited impact.

In section 1.1 an introduction to experimental searches and to the strategy of this particular analysis is given, in section 1.2 the background model estimation is described, while in section 1.3 methods to evaluate systematics uncertaities are discussed, finally in section 1.4 the result of the sarch are presented.

1.1.1 Simulated Event Samples

Signal production via the gluon fusion process, $gg \to A/H/h$, was simulated with POWHEG [30] and the associated $b\bar{b}A/H/h$ production with SHERPA [31]. The pseudoscalar Higgs boson samples were generated in the mass range from 90 GeV to 300 GeV and at $\tan\beta = 20$, the same kinematics are assumed for $A/h/H$ Higgs bosons decay products and at other $\tan\beta$ values, appropriate reweighting is applied according to the different cross-sections. The $m_h^{\max}$ MSSM benchmark scenario [35] is assumed.

The production of W and Z/γ^* bosons in association with jets was simulated with the ALPGEN [23] generator. The $t\bar{t}$ process was generated using the POWHEG generator. The single-top (s-channel, Wt) processes were generated using MC@NLO [25], while single-top (t-channel) processes were generated with AcerMC [26]. The production of diboson (WW, WZ, ZZ) were generated with HERWIG [27]. For all ALPGEN and MC@NLO samples described above, the parton shower and hadronisation were simulated with HERWIG and the activity of the underlying event with JIMMY [28]. Different parton density functions (PDFs) sets are used depending on the generator - CTEQ6L1 [32] is used by ALPGEN and AcerMC while CT10 [33] is used by SHERPA, POWHEG and MC@NLO.

TAUOLA [37] and PHOTOS [38] are used to model the tau lepton decay and additional photon radiation from charged leptons in the leading-log approximation, respectively, except for SHERPA samples.

All MC event samples were passed through the full simulation of the ATLAS detector using GEANT4 [39, 40] The effects of the simultaneous recording of several events from the same or neighbouring bunch crossings (pile-up) are considered in the simulation.

The cross-sections of the MC event samples used in this note are summarised in Table 1.1. The W/Z+jets and $b\bar{b}A/H/h \to \tau\tau$ cross sections are calculated to NNLO. Those for $t\bar{t}$ comes from direct cross section measurement []. The single top and diboson cross sections are calculated at NLO for single top and dibosons. Finally, the direct $gg \to A/H/h \to \tau\tau$ signal cross sections are calculated at NNLO and NLO for the top loop and the bottom loop and top/bottom loops interference,

Process	Cross-section (pb) [$\times$ BR]
$W \to \ell$+jets ($\ell = e, \mu, \tau$)	12.22×10^3
$Z/\gamma^* \to \ell\ell$+jets ($m_{\ell\ell} > 60$ GeV)	1.15×10^3
$Z/\gamma^* \to \ell\ell$+jets ($10 < m_{\ell\ell} < 60$ GeV)	4.35×10^3
$t\bar{t}$	137.3
Single top t-, s- and Wt-channels	28.4, 1.8, 22.4
Diboson WW, WZ and ZZ	20.6, 6.8, 1.55
Signal ($m_A = 150$ GeV, $\tan\beta = 20$, m_h^{max} scenario)	
$gg \to A\times\mathrm{BR}(A \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$	$16.8 \times 0.118 \times 0.062$
$gg \to H\times\mathrm{BR}(H \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$ ($m_H = 151$ GeV)	$18.4 \times 0.119 \times 0.062$
$gg \to h\times\mathrm{BR}(h \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$ ($m_h = 129$ GeV)	$13.7 \times 0.110 \times 0.062$
$b\bar{b}A\times\mathrm{BR}(A \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$	$39.4 \times 0.118 \times 0.062$
$b\bar{b}H\times\mathrm{BR}(H \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$ ($m_H = 151$ GeV)	$35.7 \times 0.119 \times 0.062$
$b\bar{b}h\times\mathrm{BR}(h \to \tau\tau)\times\mathrm{BR}(\tau\tau \to e\mu + 4\nu)$ ($m_h = 129$ GeV)	$4.71 \times 0.110 \times 0.062$

Table 1.1: The cross sections (multiplied by the relevant branching ratios (BR)) used in this note. Signal cross sections are shown for $m_A = 150$ GeV and $\tan\beta = 20$

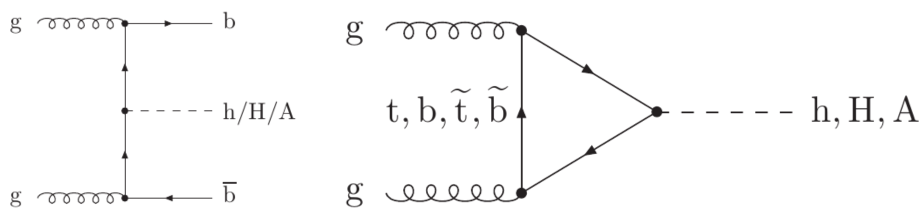


Figure 1.2: Feynman diagram for b-associated production and gluon-gluon fusion for MSSM neutral Higgs.

respectively.

1.1.2 Event Selections and Categorizzation

The signal events considered in this search are characterised by the presence of one electron, one muon and four neutrinos, the latter are associated to missing transverse energy of the event. Furthermore, additional b-tagged jets may be present if the the Higgs boson is produced in association with b-quark. According to the signal events caractheristics, each event either data and MC should satisfy the following selection criteria, these selections are shared by both analysis category and therefore referred in the following as "common selections":

(i) Trigger requiring the presence of an electron with $P_T > 24$ GeV, or alternativaley, an electron with $P_T > 12$ GeV togheter with a muon with $P_T > 8$ GeV. *Note to Sandra: should I mention that the trigger is a selection during data taking? Or is the trigger is supposed common knowledge?*

(ii) Exactly one reconstructed electron and one muon of opposite charge should be present in the event. The muon is required to have $P_T > 10$ GeV, while the electron should have $P_T > 15$ or 25 GeV depending on the trigger that

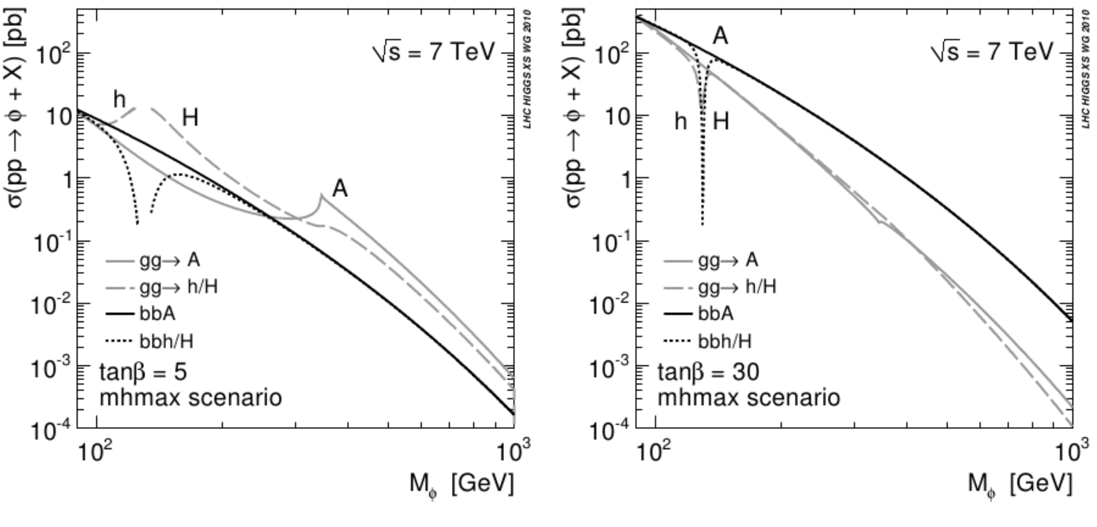


Figure 1.3: Production cross section for the $h/H/A$ MSSM neutral Higgs bosons via b-associated production and gluon-gluon fusion production mode. The calculation are for the m_h^{max} scenario and for $\tan\beta = 5$ (left) and $\tan\beta = 30$ (right).

selected the event. For definition of reconstructed electron and muon object see chapter **??**.

(iii) The two leptons should be isolated, meaning that in a cone around the lepton there should be little energy deposit (should not be sorraunded by other particle, common of non-promt leptons coming from jets). For more detail about isolation properties see section 1.2.

(iv) The events is rejected if at least one jet from hadronic τ decay is found with $P_T > 15$ GeV.

(v) The invariant mass of the sum of the electron and muon 4-vectors should be greather than 30 GeV.

This set of selection all togheter are referred in the following as *preselection.* More detail on preselections are reported in table 1.2, for details on object reconstruction and quality requirements see chapter **??**. The two analysis category, *b-tag* and *b-veto*, are defined adding on top of the preselections the request of "exactly one b-tagged jet" or "no b-tagged jet" in the event respectively, to be *taggable* a jet should have $P_T > 20$ GeV and $|\eta| < 2.5$.

Event Category

The final state of Higgs decaying into tau pair coincide with the one from $Z \to \tau\tau$ process, this is then an irreducible background. Exploiting the different kinematics of the Higgs decay with respect to other backgrounds it possible to disentangle between the two. In the Higgs decaying into $\tau^+\tau^- \to e\ \mu + 4\nu$ the taus are highly boosted and this feature is transferred to the final state leptons, their kinematics

Channel	Selection
Preselection	Trigger
	At least one reconstructed vertex
	Event cleaning
	Tau Veto
	Exactly one tight isolated electron with $P_T > 15$ or 25 GeV (trigger dependent)
	Exactly one Combined isolated muon with $P_T > 10$ GeV
	Opposite charge between the leptons
b-Tag	Exactly one b-tagged taggable jet
	$\Delta\phi(e-\mu) > 2$
	$\sum \cos\Delta\phi > -0.2$
	$\sum H_T < 100$ GeV
	$\sum L_T + E_T^{miss} < 100$ GeV
	Good MMC solution
b-Veto	Exactly zero b-tagged taggable jets
	$\Delta\phi(e-\mu) > 1.6$
	$\sum \cos\Delta\phi > -0.4$
	Good MMC solution

Table 1.2: Summary of the preselection and the full selections used for the b-tag and b-veto channels.

then result to be significantly different with respect to process like diboson or $t\bar{t}$. A first difference is that e and μ from the Higgs decay will be more likely "back-to-back", as it is shown in Figure 1.4(a) where the angle between the leptons in the transverse plane $\Delta\phi = |\phi_e - \phi_\mu|$ is reported. Furthermore the neutrinos will be more likely collinear with the charged leptons: this feature can be matematically seen as the sum of scalar product between missing energy and the leptons four-vectors in the transverse plane, if the vectors are normalised to unit versors then what remains is a relation only between angles:

$$\hat{E}_T^{miss} \cdot (\hat{P}_T^{\mu} + \hat{P}_T^{e}) = cos(\Delta\phi_{E_T,\mu}) + cos(\Delta\phi_{E_T,e}) = \sum_{\ell} cos(\Delta\phi_{E_T,\ell})$$

collinearity implies this sum to be equal to zero as it is shown in Figure 1.4(b). These two feature can be used to distinguish between mu-e coming from decay from highly boosted object and the one coming from W decays in top or in dibosons backgrounds which will have a more spread distribution. In b-veto category these two variables are sufficient to suppress contribution from dibosons, no other selection is applied in this category because it has been shown to not bring significant improvement.

In the b-tag category the situation is different, the request of b-jet enhance backgrounds with high jet activity as top production, given the relatively low jet activity of Higgs events (also in the case of b-associated production) it is possible to separate them from top production which instead is very likely to have two or more highly enegetic jets in the event. Little jet activity is achieved by requesting requesting the sum of the jets P_T in the event to be small, this variable is called H_T and is shown in Figure 1.4(c). Another feature that distinguish top pair production from Higgs is the much higher invariant mass of the former final state, in the transverse plane all the leptons will tend to have a higher momentum, the sum of lepton P_T and E_T^{miss} is then used as a discriminating variable. Figure 1.4(d) shows the distribution of this last analysis variable.

The above described variables defines the signal region in the b-tag and b-veto category, in table 1.2 a summary of the preselection and all the selection variable used with their optimized cut values is reported. Figure 1.5 shows the final state invariant mass distribution (here the MMC_{mass} discriminating variable is used see section 1.1.3) as a function of the selection stage, while in tables 1.4-1.3 the number of events that survises at each cut stage for different background is reported.

1.1.3 Mass Reconstruction with MMC Technique

Reconstructing the invariant mass of a di-τ resonace is a challenging task due to the presence of neutrinos from the τ lepton decay. In case the τ leptons decays leptonically a total of four neutrinos are involved in the final state, a pair of neutrinos for each τ lepton decay, the system presents then eight uknowns, which are the components of the two four-vectors related to each of the two pair of neutrinos. There are four kinematic constraint to the system which are summarized by following equation:

$$\begin{aligned} &\vec{E}_T^{miss} = \vec{P}_T^{mis_1} + \vec{P}_T^{mis_2} \\ &M_{\tau_i}^2 = m_{mis_i}^2 + m_{vis_i}^2 + 2\mathbf{P}_{vis_i} \cdot \mathbf{P}_{mis_i} \end{aligned} \tag{1.1}$$

where the index i runs over the two τ lepton of the event and assumes the value of 1 or 2, $\mathbf{P}_{mis}$ and $\mathbf{P}_{vis}$ represent respectively the four momentum of the pair of neutrinos and the four momentim of the charged lepton from the τ lepton decay, m_{miss}^2, therefore there are still four degrees of freedom in the system. Several approximation are possible to further constraint the momentum carried by neutrinos, for example assuming them collinear to the charged leptons from tau decay, however those approximation suffers of limitations.

In this analysis, the so-called "Missing Mass Calculator" (MMC) algorithm is used to calculate the most likely di-tau system invariant mass from its decay products and the E_T^{miss}, the implementation of this method is based on [56]. The concept of the MMC is to reconstruct the momentum of each neutrino pair by using all known kinematic constraints and performing a "scan" over the yet undetermined variables, the four degrees of freedom are fixed assigning values to four independent variables, which are chosen to be: $m_{mis_i}^2$ and $cos\theta_i^*$, the latter defined as the angle between the charged lepton from the τ lepton decay and the boost direction of the τ lepton. For a given event topology and given values of those variable, the system of equation 1.1 is solved and is possible to calculate the di-tau invariant mass, however, the solution are not all equally likely, using Pythia simulation supplemented with TAUOLA package, probability obtained from the matrix element of a tau lepton decay

The missing energy plays an important role in the mass reconstrution with this method, the mass resolution is connected with the missin energy resolution, the scan over the parameter space is then achieved in two steps: first a scan over six parameter is done (the one mentioned above and $\vec{E}_T^{miss}$), the scan over $E_{T_x}^{miss}$ and $E_{T_y}^{miss}$ is done chosing values within one sigma from the central missing energy value,

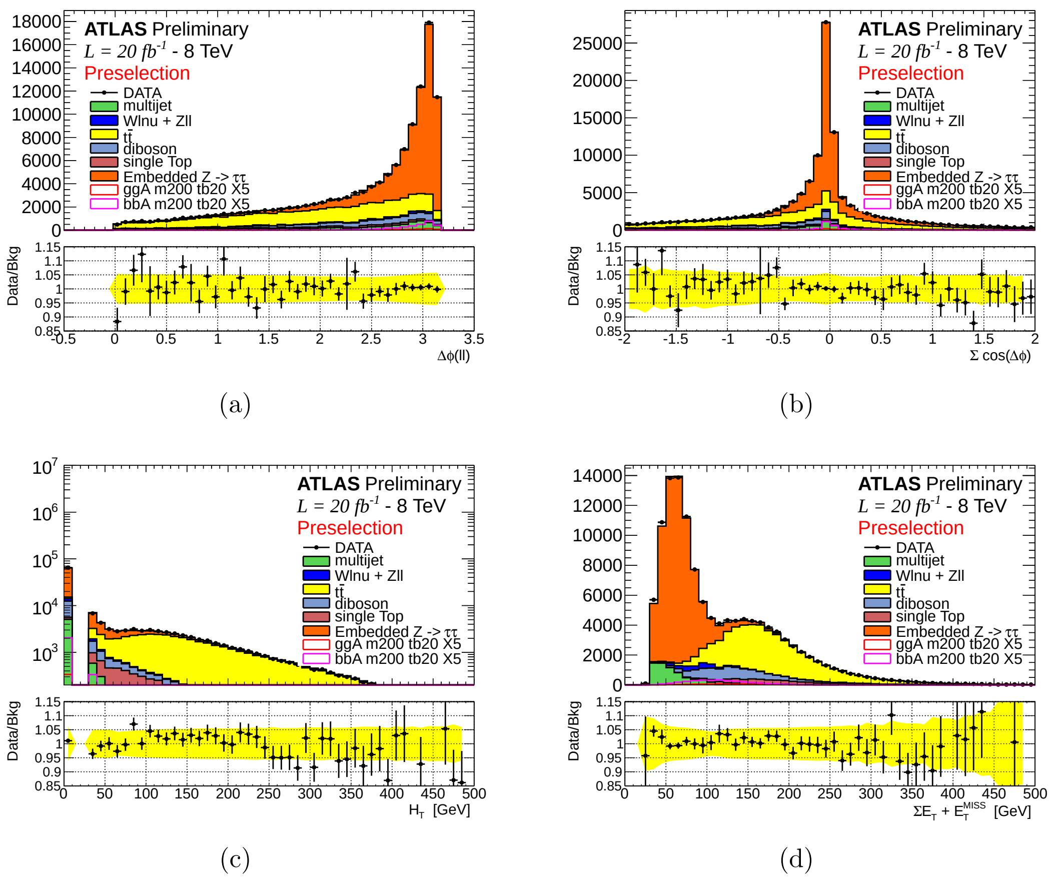


Figure 1.4: Distribution of analysis variable after preselection.

according to its uncertainty, the scan is made over the points which are physical solution for the system of equation 1.1, each point is then weighed by the probability obtained from the matrix element of a tau lepton decay, the weighted mean of $\vec{E}_T^{miss}$ is assumed as a better estimate and used in a four dimensional scan.

The performance of the missing transverse energy (Emiss) reconstruction has a large impact on the performance of the MMC. As the resolution

The MMC scans variables in the respective rest frame of the leptons. By scanning $cos\theta_i^*$, defined as the angle between the charged lepton from the lepton decay and the boost direction of the lepton, and the mass of the di-neutrino system, the system of equations can be solved. Each scan point is then weighted by the probability obtained from the matrix element of a tau decay decay at tree level. The estimator for the final discriminant, the mass of the di-tau system mMMC, is then defined as the maximum of the histogram that has been filled with the weighted scan points.

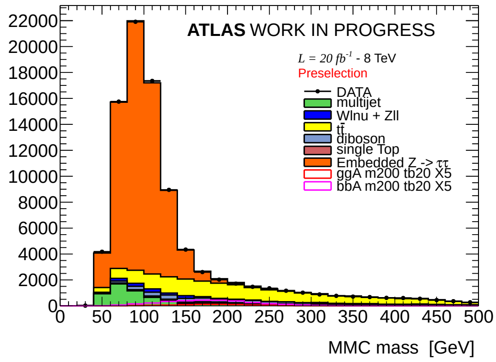
information from the well known tau decay to constraint the system: in the above mentioned four dimensional parameter space, in which lies the solution of the di-tau system, not all the points are equally likely for a given event topology, using Pythia simulation supplemented with TAUOLA package, a likelihood based on $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ distribution between the missing energy and each of the two "visible" leptons is build, minimising the likelihood for a given event topology it is possible to determine the most likely point in the parameter space and solve the system. Effect of resolution are taken into account by producing a mass distribution for each of the scanned points in the parameter space. As a result this method give a more precise measurement of the di-tau system invariant mass and a considerable improvement in resolution. The invariant mass distribution calculated with the MMC technique is referred in the following as MMC_{mass} and is used as discriminating variable in the limits setting.

1.2 Background Modeling and Validation

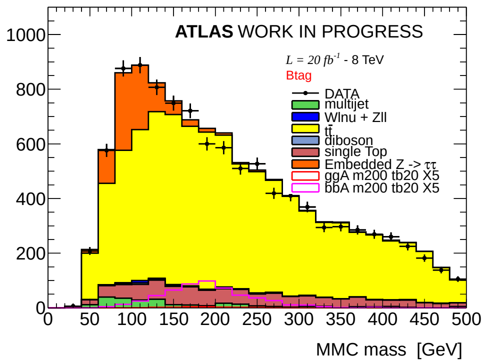
This section describes the strategies for background modeling and validation. Monte Carlo (MC) simulation is extesively used for model either background and signal, in section 1.1.1 a brief description of the simulated sample used is given.

Monte Carlo simulations of any process are usually prone to systematics uncertainties due to non-perfect descriptions of pileup effects, underlying event and detector performance, therefore, data-driven background esimation method are employed for the estimate of $Z \to \tau\tau$ and QCD multijet backgrounds, described respectively in section 1.2.3 and 1.2.2. Other background processes, such as $t\bar{t}$, single top, dibosons, $Z \to ll$ + jets (where $l = e, \mu$) and W + jets, are estimated using MC predictions. Given the particular importance of $t\bar{t}$ a dedicated study to validate this background has been made and described in section 1.2.1.

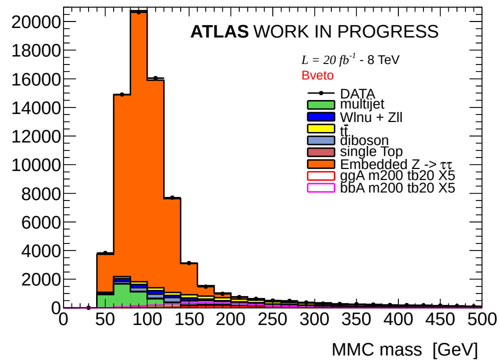
In this section we make use of analysis tools, quality requirements and object definition (like electrons, jet and muon) described in chapter **??**. Furthermore a set corrections is applied to simulated events to take into the non perfect description of detector performance and response, full detail on those corrections is reported



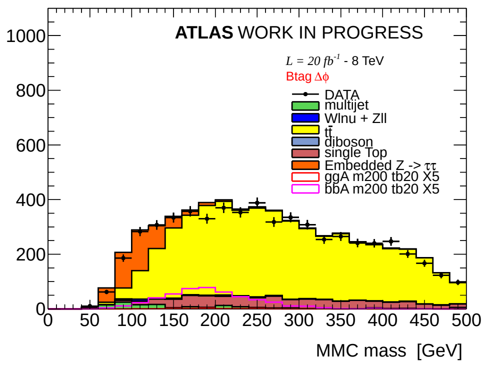
(a)



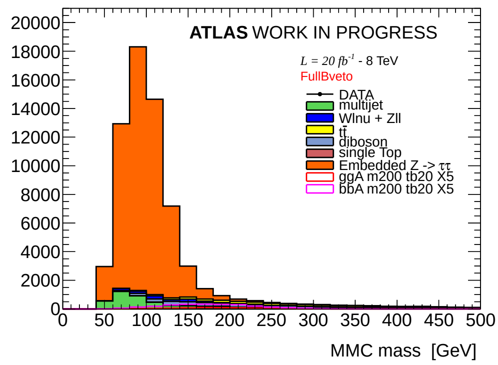
(b)



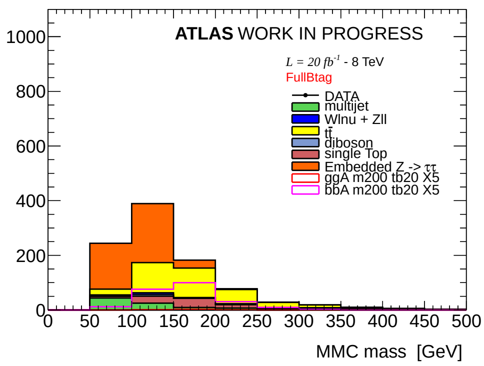
(c)



(d)



(e)



(f)

Figure 1.5: Distribution of the MMC_{mass} for different cuts stage, see text. Left column corresponds to b-tag category, right column to b-veto.

	Preselection	n(b-jet)=1	$\Delta\phi(e-\mu) > 2$	$\sum \cos\Delta\phi > -0.2$	$\sum L_T + E_T^{miss} < 100$ GeV	$\sum H_T < 100$ GeV	mmc
Data	125886	23352	-	-	-	-	-
Multijet	6700 ± 500	330 ± 40	208 ± 27	135 ± 22	114 ± 17	100 ± 15	100 ± 15
$Z \to \ell\ell$	570 ± 50	5.2 ± 1.8	2.3 ± 1.1	2.3 ± 1.1	1.7 ± 1.0	0.9 ± 0.8	0.9 ± 0.8
$W \to \ell\nu$	1630 ± 150	20 ± 6	15 ± 6	13 ± 6	10 ± 6	10 ± 6	10 ± 6
Diboson	9340 ± 50	99 ± 5	63 ± 4	36.4 ± 3.0	14.8 ± 1.8	13.3 ± 1.8	13.1 ± 1.8
$t\bar{t}$	40630 ± 110	19810 ± 70	9680 ± 50	6450 ± 50	808 ± 15	350 ± 10	330 ± 10
Single Top	4450 ± 40	2456 ± 33	1223 ± 23	784 ± 18	122 ± 7	99 ± 7	90 ± 6
$Z \to \tau\tau$	61500 ± 70	952 ± 9	625 ± 7	540 ± 7	482 ± 6	421 ± 6	418 ± 6
Signal			-	-	-	-	-

Table 1.3: Number of data and background events in the b-tag channel.

	Preselection	n(b-jet)=0	$\Delta\phi(e-\mu) > 1.6$	$\sum \cos\Delta\phi > -0.4$	mmc
Data	125886	89155	-	-	-
Multijet	6693 ± 456	6357 ± 461	5322 ± 438	4137 ± 339	3934 ± 335
$Z \to \ell\ell$	569 ± 48	564 ± 48	516 ± 47	434 ± 44	432 ± 44
$W \to \ell\nu$	1625 ± 155	1604 ± 155	1145 ± 125	714 ± 101	656 ± 100
Diboson	9338 ± 48	9235 ± 48	7358 ± 43	4002 ± 31	2925 ± 27
$t\bar{t}$	40632 ± 106	7707 ± 46	5044 ± 37	3416 ± 31	2159 ± 24
Single Top	4449 ± 44	1664 ± 27	1124 ± 22	682 ± 18	435 ± 14
$Z \to \tau\tau$	61503 ± 68	60440 ± 67	58078 ± 65	55303 ± 64	54683 ± 63
Signal			-	-	-

Table 1.4: Number of data and background events in the b-veto channel.

in appendix **??**. Systematic uncertainties on the background model predictions are detailed in Section 1.3.

A good agreement between data and background model is found after preselections, this is supported by figure 1.6 which shows few kinematic variables and figure 1.4 which shows analysis selection variables after preselection.

1.2.1 Top Quark Pair Production Validation

The background from top quark pair production is estimated using a sample of events from the POWHEG-PYTHIA MC generator. Since this is one of the major backgrounds for this analysis a careful validation is needed, for this purpose a $t\bar{t}$ rich control region is defined using events passing the preselection with the additional requirement of two b-tagged jets. Figures 1.7 and 1.8 show a set of kinematic and analysis selection variables in this CR, for both data and the MC prediction, good agreement between data and the background model is found: an overall data to background ratio of $0.998 \pm 0.011(\text{stat.}) \pm 0.110(\text{sys.})$ is observed. The total systematic uncertainty on the ratio is dominated by the uncertainty on the b-tagging efficiency.

1.2.2 Multi-jet Background

The QCD multi-jet background represents an important background, especially in the b-veto category, due to its high cross-section and the relatively low cut on lepton P_T used in this analysis. This background is evaluated by a data-driven technique, the so-called ABCD method. The ABCD method consists of splitting the data sample in four regions: the signal region (SR) and three control regions (CR), where the control regions are mutually orthogonal and designed to be enriched in multi-jets events. The four regions are defined by using the charge correlation between the leptons and isolation selections. With isolation is intended the sum of the energy deposit in a cone of fixed size around the lepton, this variable can be defined using calorimetric energy deposition or track momentum measurement done by the inner detector. To obtain regions rich in multi-jet background, the selections on both the calorimetric and tracking isolation are inverted with respect to the nominal ones defining anti-isolated leptons, is then possible to define four

Region	Lepton Charge	Lepton Isolation
A (signal region)	OS	isolated
B	SS	isolated
C	OS	anti-isolated
D	SS	anti-isolated

Table 1.5: QCD background estimation control regions, defined by having leptons with opposite signs (OS) or same signs (SS) and by having the leptons either isolated or anti-isolated.

regions: opposite sign (OS) or same sign (SS) with respectively isolated or anti-isolated leptons. Historically the letters A-D are assigned to this regions for a quicker reference as defined in Table 1.5.

An assumption of the ABCD method is that multi-jet backgrounds populate the OS and SS events independently of lepton isolation criteria, or in other words that the ratio of OS/SS events is uncorrelated with the lepton isolation selections. In this case, the number of QCD events in the signal region A can be estimated from the yield of multijet events in the control regions B, C and D, using the equation

$$N_A = N_B \times \frac{N_C}{N_D} = N_B \times R_{QCD} \tag{1.2}$$

To obtain the multijet yields in the data CRs, the contamination from electroweak (W+jets, Z+jets and dibosons) and top processes ($t\bar{t}$ and single top production) are subtracted in each control region using the MC prediction for their event yield. Tables 1.6 and 1.7 show the event yield for each CR throughout the full cut-flows along with the predictions of non-QCD multi-jets events which are subtracted. Signal contamination has been checked in all the three control regions for different mass points. For the range of m_A and tanβ considered in this analysis, the highest signal contamination is seen in region B for the mass point $m_A = 300$ GeV and tan$\beta = 50$, where a contamination of 0.2% is observed[1].

Shapes of kinematic distributions for QCD events are taken from region B, this region is expected to have similar kinematic property to the SR, however, suffers of either lower statistics and higher contamination with respect to region C or D. This choice is made to avoid a shape bias due to isolation requirements at trigger level (only the single-electron trigger ask for isolation), figure 1.9 shows the comparison between the electron P_T distributions in region B and D, in the latter high P_T electrons are suppressed, they do not pass trigger selections. Eventually the trigger isolation requirement could bias also the ratio OS/SS, this possibility has been checked carefully in a dedicated study and reported in Appendix **??**: to a good approximation, such trigger effects cancel out in the ratio OS/SS and no additional systematic is needed.

To test the ABCD method predictions an additional control region has been defined with the following selections:

[1] This value is mainly due to b-associated production and, as it scales with the cross section, for tan$\beta = 20$ would be an order of magnitude smaller.

- $E_T^{miss} < 20$ GeV
- $H_T < 70$ GeV and $\sum L_T + E_T^{miss} < 50$ GeV
- $0 < MMC_{mass} < 80$ GeV

This control region is designed to enhance multi-jet background with respect to $Z \to \tau\tau$ keeping the final state kinematics as similar as possible to the SR. Figure 1.10 shows the MMC_{mass} distribution for this CR with and without b-tagging requirements, agreement between data and the background model is found within statistical and detector related systematics uncertainty.

Systematic uncertainties are assigned on the scaling factor R_{QCD} and on the shape of the discriminating variable MMC_{mass} to take into account any correlation between isolation and charge of the leptons, details on the systematic uncertainty evaluation are addressed in Section 1.3.

1.2.3 $Z \to \tau\tau$ + Jets Background: Embedding Technique

The background from $Z \to \tau\tau$ decays is the major background to this analysis, a good uderstanding of it is then crucial. Unfortunately, for a light Higgs boson, it is impossible to completely separate $Z \to \tau\tau$ decays from the signal and a signal free data control region cannot be defined. However, thanks to the small Higgs coupling to muons, $Z \to \mu\mu$ decays provide a good starting point to model $Z \to \tau\tau$ events in a data-driven way. An hybrid Data-MC sample, known as "Embedding" is used to model the $Z \to \tau\tau$ background: $Z \to \mu\mu$ candidates are selected in data, then, the two muons from the Z decay are substituted with the decay products from simulated taus, this means that the energy deposit and tracks in a cone around the muon are subtracted and substituted with the one from τ decay, those taus have the same kinematics as the original muons. Further details on this technique may be found in [41, 42].

Trigger is not simulated in the embedding samples, the event yield is normalised to ALPGEN $Z \to \tau\tau$ at preselection stage. Furthermore a set of corrections, as described in [43], are applied to unfold from the original $Z \to \mu\mu$ trigger and reconstruction efficiency, then trigger and reconstruction efficiency for a $e - \mu$ final state are emulated by means of event weight.

The Embedding technique has been validated in several studies, detailed in [41, 43], which show a good description of data and $Z \to \tau\tau$ MC by Embedding. In the context of this analysis, figures 1.11 and 1.12 show comparisons of various kinematic variables between data, embedding and ALPGEN $Z \to \tau\tau$ events at preselection. No significant deviation is seen between the MMC_{mass} distribution of the embedding and ALPGEN samples, however other relevant variables for this analysis, such as the E_T^{miss} and the number of b-jets, are slightly better described by embedding.

The Embedding sample is based on selecting $Z \to \mu\mu$ candidates in data, the selections assure a rather pure $Z \to \mu\mu$ sample, however further selections used in this analysis, for example the b-tagging requirements, could enhance the contamination fraction from other processes. Dedicated studies have been made to estimate the $t\bar{t}$

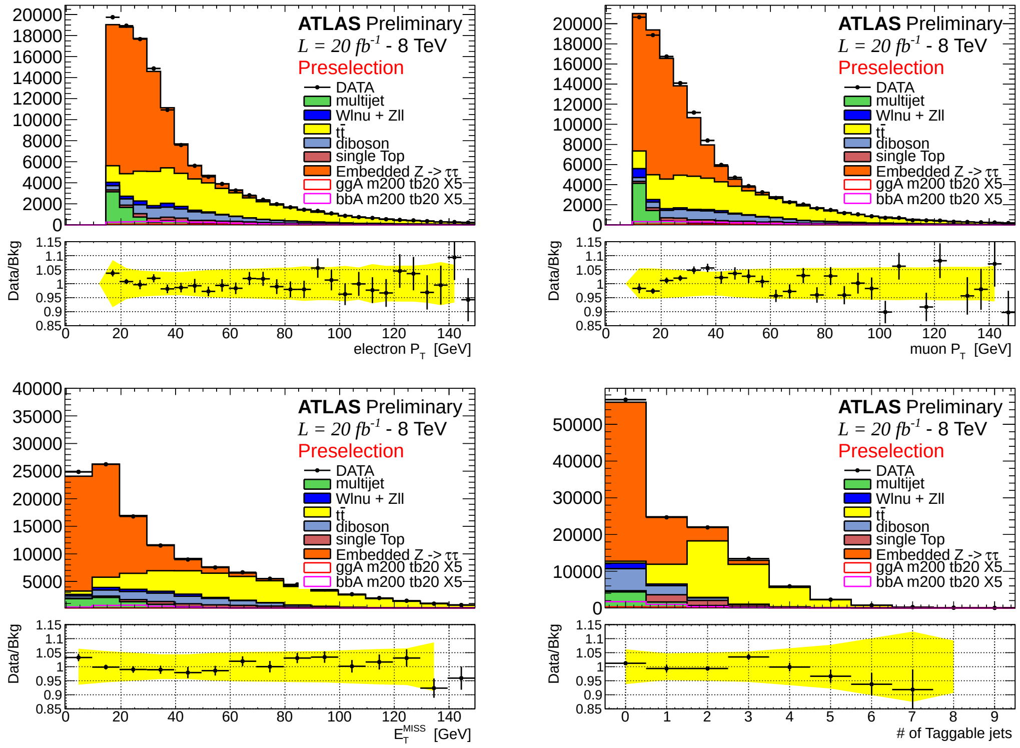
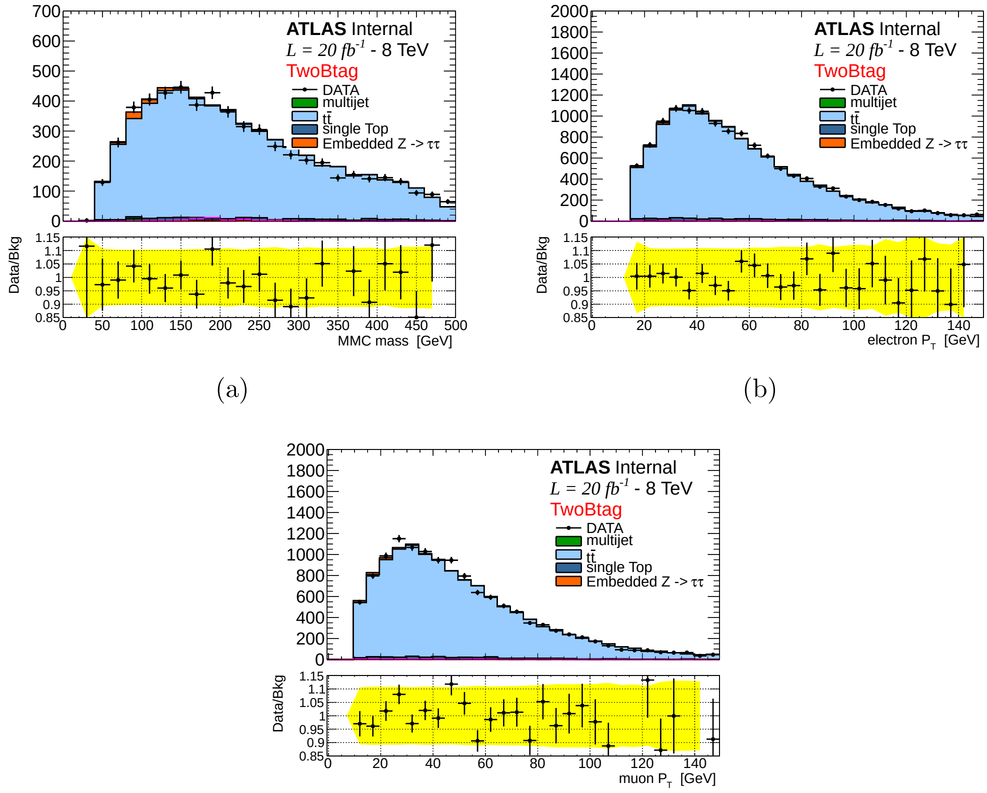


Figure 1.6: Distribution of some kinematic variables after preselection.

Selection		B	C	D	R_{QCD}
Preselection	Data	6189	604628	312901	1.929 ± 0.004
	non-QCD	2510 ± 180	1090 ± 30	730 ± 35	
B-tag	Data	419	44619	27257	1.64 ± 0.01
	non-QCD	215 ± 10	310 ± 12	277 ± 13	
$\Delta\phi(e-\mu)$	Data	230	38810	23316	1.67 ± 0.01
	non-QCD	104 ± 6	200 ± 10	175 ± 7	
$\sum\cos\Delta\phi$	Data	149	31379	18779	1.67 ± 0.02
	non-QCD	67 ± 5	127 ± 8	114 ± 6	
$\sum H_T$	Data	83	27781	15626	1.78 ± 0.02
	non-QCD	23 ± 4	25 ± 3	22 ± 3	
$\sum L_T + E_T^{miss}$	Data	71	27735	15590	1.78 ± 0.02
	non-QCD	10 ± 3	22 ± 3	18 ± 2	
$MMC_{mass} > 0.$	Data	70	27634	15522	1.78 ± 0.02
	non-QCD	9 ± 3	20 ± 3	17 ± 2	

Table 1.6: QCD background estimation as a function of the analysis selections for the b-tagged category. The yields for the different control regions, as well as the scaling factor R_{QCD}, are reported. The error on the R_{QCD} is statistical only.

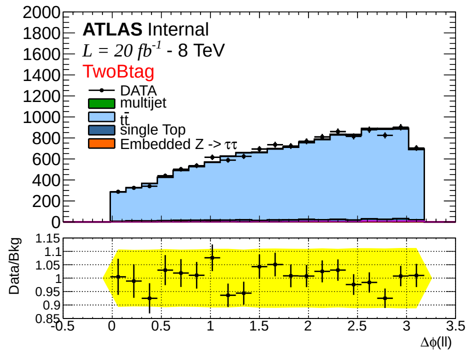


(a) (b) (c)

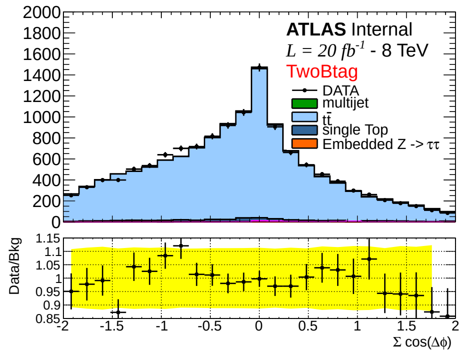
Figure 1.7: Distributions of a) the MMC mass, b) the transverse momentum of the electron $P_T(e)$ and c) the transverse momentum of the muon $P_T(\mu)$, for both data and MC in the $t\bar{t}$ control region. The uncertainties on the points for the ratio plot show the statistical uncertainty on the data to background ratio, whereas the yellow band show the total systematic uncertainty on this ratio.

Selection		B	C	D	R_{QCD}
Preselection	Data	6189	604628	312901	1.929 $\pm$ 0.004
	non-QCD	2510 $\pm$ 180	1090 $\pm$ 30	730 $\pm$ 35	
B-veto	Data	5673	558217	284847	1.960 $\pm$ 0.004
	non-QCD	2220 $\pm$ 180	710 $\pm$ 30	415 $\pm$ 30	
$\Delta\phi(e-\mu)i$	Data	4610	532583	271404	1.962 $\pm$ 0.005
	non-QCD	1700 $\pm$170	580 $\pm$ 30	345 $\pm$ 30	
$\sum\cos\Delta\phi$	Data	3417	486747	247712	1.965 $\pm$ 0.005
	non-QCD	1120 $\pm$ 100	370 $\pm$ 20	230 $\pm$ 20	
$MMC_{mass} > 0.$	Data	3177	479967	244276	1.965 $\pm$ 0.005
	non-QCD	1000 $\pm$ 100	300 $\pm$ 17	190 $\pm$ 20	

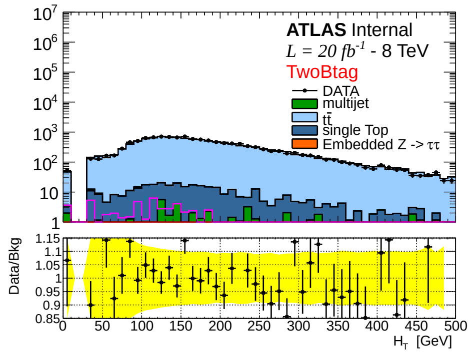
Table 1.7: QCD background estimation as a function of the analysis selections for b-veto category. The yields for the different control regions, as well as the scaling factor R_{QCD}, are reported. The error on the R_{QCD} is statistical only.



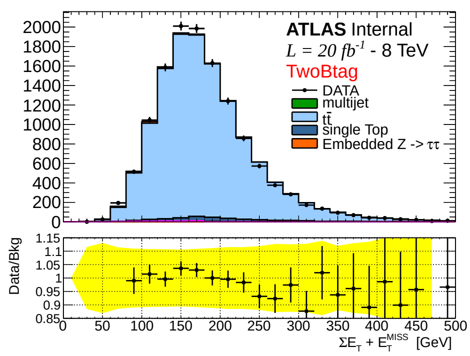
(a)



(b)



(c)



(d)

Figure 1.8: Distributions of a) $\Delta\phi(e-\mu)$, b) $\sum\cos\Delta\phi$, c) $\sum L_T + E_T^{miss}$ and d) H_T, for both data and MC in the $t\bar{t}$ control region. The uncertainty on the points for the ratio plot show the statistical uncertainty on the data to background ratio, whereas the yellow band show the total systematic uncertainty on this ratio.

and QCD multi-jet contamination in the embedding sample. The $t\bar{t}$ contamination is estimated by evaluating the embedding yield in a two b-tag control region (as described in Section 1.2.1), these events are assumed to be solely from $t\bar{t}$ and their yield in the signal region is extrapolated using MC simulation. Table 1.8 shows a summary for the top contamination in embedding The multi-jet contamination can be estimated starting from the embedding yield in (ABCD) region C, assuming all events in this CR as QCD multi-jet events, the contamination in the SR can be estimated by means of the ABCD method (see Section1.2.2). The R_{QCD} factor, in this case, is evaluated using a $\mu-\mu$ final state with same kinematic selections as for embedding $Z\rightarrow\mu\mu$ candidate. Table 1.9 shows the estimated contamination of QCD multi-jet in embedding. We consider contamination effects negligible.

1.3 Systematic Uncertainties

This section describes the range of systematic uncertainties that are relevant for this analysis. To account for differences in the detector responses between simulation

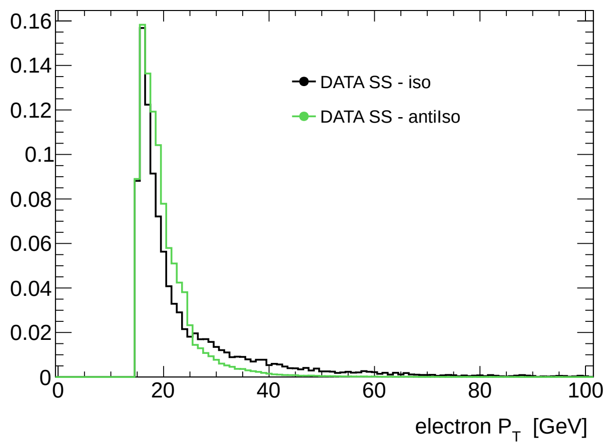
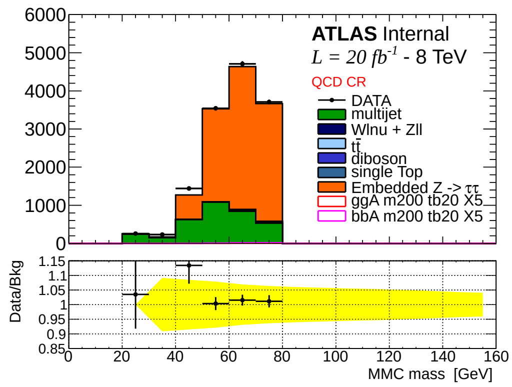
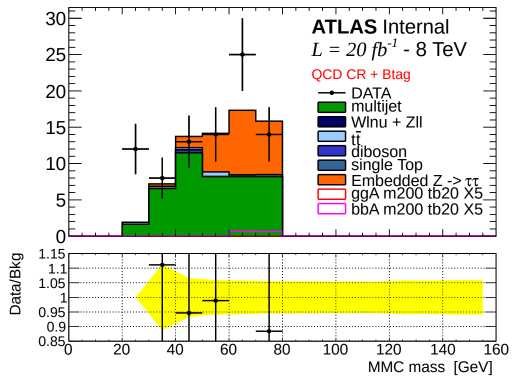


Figure 1.9: Comparison of the electron P_T distribution in region B and region D, showing the bias due to the trigger. The histograms are normalised to the same area.



(a)



(b)

Figure 1.10: MMC_{mass} distribution for QCD cross check regions defined in section 1.2.2 (a) and for the same CR when in addition one b-tagged jet is required (b).

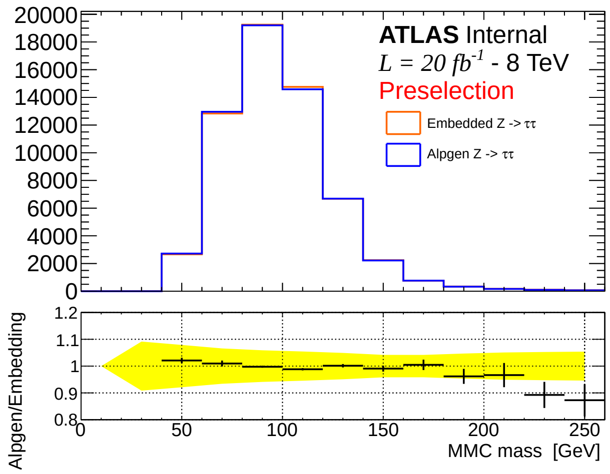


Figure 1.11: Comparison between the embedded $Z \to \tau\tau$ and ALPGEN for MMC_{mass} distributions.

	Embedding yield in CR	Transfer factor	Estimated events in SR	Contamination
b-tag	84 ± 9	$(2.6 \pm 0.1) \times 10^{-2}$	2.2 ± 0.2	0.5 %
b-veto	84 ± 9	$(1.74 \pm 0.02) \times 10^{-1}$	15 ± 2	0.03 %

Table 1.8: Evaluating embedding $t\bar{t}$ contamination using a two b-tag CR. The transfer factor is the multiplicative factor that allows to estimate events in SR from the CR.

	Embedding yield in CR	Transfer factor	Estimated events in SR	Contamination
B-tag	12 ± 3	$(7 \pm 1) \times 10^{-3}$	$(8.4 \pm 0.3) \times 10^{-2}$	0.03 %
B-veto	390 ± 20	$(2.5 \pm 0.1) \times 10^{-2}$	10.0 ± 0.5	0.02 %

Table 1.9: Evaluating embedding contamination due to QCD multi-jet using ABCD method, the CR here is with OS anti-isolated events (region C). The transfer factor is the multiplicative factor that allows to estimate events in SR from the CR, in this case is N_B/N_D and is evaluated using mu-mu final state with the same kinematic selection used in the definition of the embedding sample.

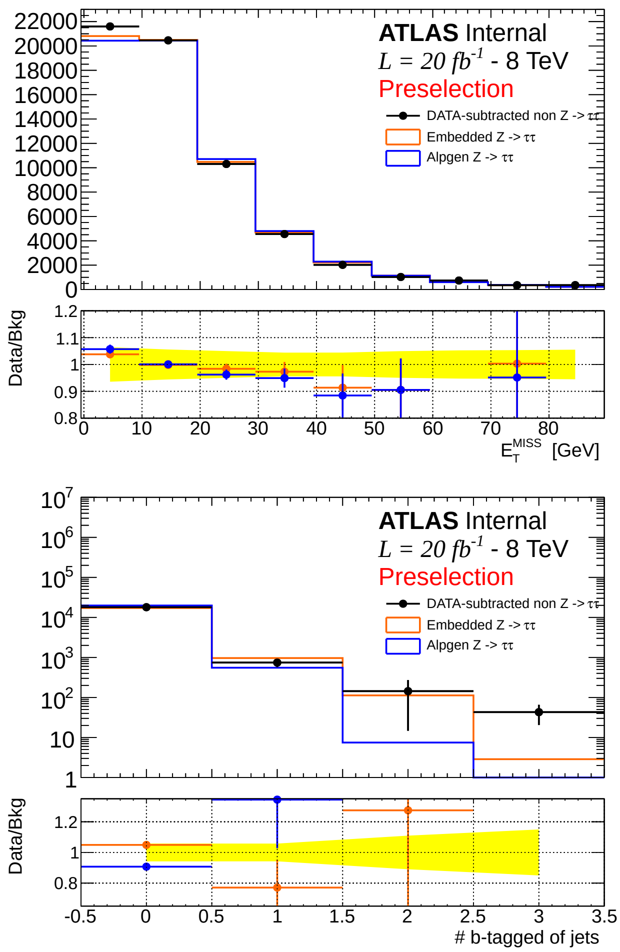


Figure 1.12: Comparison between embedded $Z \to \tau\tau$ and ALPGEN for E_T^{miss} and the number of b-tagged jets distributions. Data are superimposed, with the contribution of non-$Z \to \tau\tau$ are subtracted.

and data a set of corrections are applied either at object reconstruction level and at event level, the uncertainties on such corrections are considered as detector-related systematic uncertainties and are detailed in section 1.3.1. For samples which rely on MC simulation, theory-related systematics, which include uncertainties on the cross-section and uncertainties on the acceptance of analysis selections, are described in section 1.3.2. Further systematic uncertainties related to data-driven methods for backgrounds estimation are described in section 1.3.3 and 1.3.4.

Each single systematic can contribute separately to the uncertainty on the final event yield and on the shape of the MMC_{mass} distribution which is used as discriminating variable in limit derivation. Shape systematics are documented in appendix **??**, they are found to be negligible for all the samples except Embedding, for which significant deviation are found only in the b-veto category. Systematic uncertainties that do not effect the mass shape distribution and have an impact on the event yield of less than 0.5% (per sample) are neglected in the final limit calculations.

1.3.1 Detector-related Systematics Uncertainties

Here systematic uncertainty related to object reconstruction and event corrections are addressed, those corrections are based on the measure of some relevant parameter, each of those parameters correspond to a "nuissance paremeter" in our probability model as described in Section **??**. Each parameter is variated independently (one sigma up or down) according with its uncertainty and the impact on the analysis yield for each sample is evaluated. In the following, detector related uncertainty are described with some more details, table 1.10 and 1.11 briefly summarize the impact on the samples yield for the most significant systematic uncertainty considered.

Luminosity The integrated luminosity of the 8 TeV data recorded at ATLAS during 2012 is measured to be 20.3 fb^{-1} [68], its uncertainty is 2.8%.

Pileup Simulated events are re-weighted to reproduce the average interactions per bunch crossing, $< \mu >$, seen in data. Those event weights has an uncertainty wich is propagated to each simulated sample.

Trigger Efficiency is corrected in simulation to match (as a mean value) the one in data, those correction weights are evaluated as a function of P_T and η of the leptons and have assciated uncertainties. Systematic uncertainties on both the single electron and electron-muon trigger efficiency are considered independently, those uncertainty range aproximately 1-2%.

In the embedding sample, the trigger is emulated by applying weights to the event topology in order to recover the right trigger efficiency, those weights are related to the one just described above and have similar uncertainty. Trigger efficiency uncertainty for Embedding are considered uncorrelated with the one of other samples.

Electrons Two types of uncertainty on reconstructed electron objects are considered: the first are related to electron identification and reconstruction efficiencies ("Electron ID"), the second type are related to electron energy scale and resolution corrections. The energy scale uncertainties are split into a set of six different nuisance parameters, however, only few of them give a non negligible contribution. Two of them are found to effect the shape of the MMC_{mass} distribution and are considered independently, those are the uncertainty that arise from the $Z \to ee$ momentum measurement ("Electron Zee") and the one related to low momentum electrons ("Electron LOWPT"). All the other uncertainties related to energy scale and resolution are summed in quadrature ("Electron E").

Muons The uncertainty on muon identification efficiency depends on the charge and momentum of the muon. Typically these uncertainties are of the order of a fraction of percent, and are referred as "Muon ID". The uncertainties on the muon energy scale and resolution are considerd independently for the inner detector and muon spectrometer measurements, then are added in quadrature to eastimate the final effect ("Muon E").

Taus Hadronic tau object are only used in the analysis as a veto. Uncertainties on both tau energy scale and identification efficiency have been investigated and are found to be negligible for this analysis.

Jets The systematic uncertainties on the Jet Energy Scale (JES) are split up into multiple sets of nuisance parameters, which are related to different effects and components, for example the sensitivity to pileup or to the flavour composition of the jet. The overall uncertainty on the JES ranges between 3% and 7%, depending on the P_T and η of the jet. To give an idea of the effect that these uncertainty have on the analisis yield their sum in quadrature is reported in table 1.10 and 1.11 as "JES", however this is just a simplification for illustration purposes and in the limits extraction those uncertainties are considered uncorrelated. Systematic uncertainty due to jet resolution ("Jet Resolution") are obtained by smearing the jet energy according to its uncertainty.

b-Tagging is described in chapter **??**. Corrections are applied to simulation to match b-tagging efficiency in data, uncertainties on the knowledge of the b-tagging efficiencies for the 70% working point of the MV1 b-tagger are considered. The effect of those uncertainties is evaluated independently in the cases of b-quark, c-quark and light or gluon initiated jets and referred respectively to as "B Eff", "C Eff" and "L Eff". The tagging and mistagging efficiency uncertainties are considered to be totally anti-correlated.

Missing Transverse Energy The effect of the energy scale uncertainties for all the physics objects is propagated to the E_T^{miss} calculation. In addition uncertainty on the energy scale and resolution due to the remaining calorimeter energy deposit,

b-tag category uncertainties (%)					
Source	Signal bbH	Signal ggH	$Z \to \tau\tau$	Top	Other
Electron ID	2.3	2.6	2.8	1.8	2.0
Electron E	0.7	1.2	0.5	0.5	0.9
Electron LOWPT	0.4	0.0	0.4	0.1	0.4
Electron Zee	0.3	0.6	0.4	0.6	0.5
Muon ID	0.3	0.3	0.3	0.3	0.3
Muon E	0.5	0.8	0.1	0.1	0.2
Trigger Single Ele.	0.7	0.5	0.5	0.8	0.8
Trigger Dilepton	1.0	1.2	1.4	0.6	0.6
Embedding MFS	-	-	0.0	-	-
Embedding Iso.	-	-	1.3	-	-
JES	2.7	7.3	-	10.0	7.0
JER	1.4	6.3	-	2.9	3.0
B Eff	10.2	3.1	-	2.6	5.0
C Eff	0.2	4.3	-	0.0	1.2
L Eff	0.4	8.0	-	0.1	1.2
Pileup	0.4	0.7	0.4	0.4	0.9
MET	0.7	0.5	0.2	1.0	1.2
Luminosity	2.8	2.8	2.8	2.8	2.8

Table 1.10: Summary of the effect of the experimental systematic uncertainties on the yields of the different samples used in the b-tag channel. Here "Other" refers to the sum of all the remaining samples: $W \to \ell\nu$, diboson, $Z \to \ell\ell$ and single top. The signal samples listed here are b-associated production and gluon fusion with $m_A = 120$ GeV and $\tan\beta = 20$.

the so called "soft-terms", are considered. All the uncertainty on E_T^{miss} are independently propagated through the analysis and are added in quadrature, this final term is referred as "MET" uncertainty.

1.3.2 Theoretical Uncertainties

Uncertainties on the cross-sections that have been used to normalise simulation samples to data are reported in Table 1.13. These uncertainties include contributions due to parton distribution functions (PDFs), the choice of the value of strong coupling constant, and the renormalisation and factorisation scales. Furthermore the uncertainties on signal cross-section depends on $\tan\beta$, the Higgs boson type $(A/h/H)$ and mass.

The effect of systematic uncertainties due to various MC tuning parameters, underlying event and lepton kinematic description is considered. Since the effect on the invariant mass distribution of the di-tau system from these systematic uncertainties is negligible (as an example see Figure 1.13), only the variation in acceptance is considered as systematic uncertainty. The acceptance uncertainties for the ALPGEN Z MC, used for the normalisation of the embedded sample, are estimated at lepton preselection to be 4% [65]. Since additional selections are applied

b-veto category uncertainties (%)					
Source	Signal bbH	Signal ggH	$Z \to \tau\tau$	Top	Other
Electron ID	2.4	2.3	2.9 (**s**)	1.4	1.6
Electron E.	0.4	0.5	0.4	0.5	0.9
Electron LOWPT	0.3	0.5	0.4 (**s**)	0.0	1.2
Electron Zee	0.4	0.4	0.4 (**s**)	0.1	0.3
Muon ID	0.3	0.3	0.3	0.3	0.3
Muon E.	0.1	0.1	0.1	0.5	0.5
Trigger Single Ele.	0.6	0.6	0.5	0.9	0.9
Trigger Dilep.	1.0	1.0	1.3	0.2	0.3
Embedding MFS	-	-	0.1 (**s**)	-	-
Embedding Iso.	-	-	0.0 (**s**)	-	-
JES	0.6	0.7	-	1.0	1.2
JER	0.5	0.3	-	0.6	0.3
B Eff	1.8	0.0	-	12.0	0.8
C Eff	0.0	0.1	-	0.1	0.0
L Eff	0.0	0.1	-	0.2	0.1
Pileup	0.5	0.8	0.4	0.3	0.3
MET	0.2	0.8	0.1	0.2	0.5
Luminosity	2.8	2.8	2.8	2.8	2.8

Table 1.11: Summary of the effect of the experimental systematic uncertainties on the yields of the different samples used in the b-veto channel. Here ”Other” refers to the sum of all the remaining samples: $W \to \ell\nu$, diboson, $Z \to \ell\ell$ and single top. The signal samples listed here are b-associated production and gluon fusion with $m_A = 120$ GeV and $\tan\beta = 20$. Shape uncertainty are noted with the symbol (**s**).

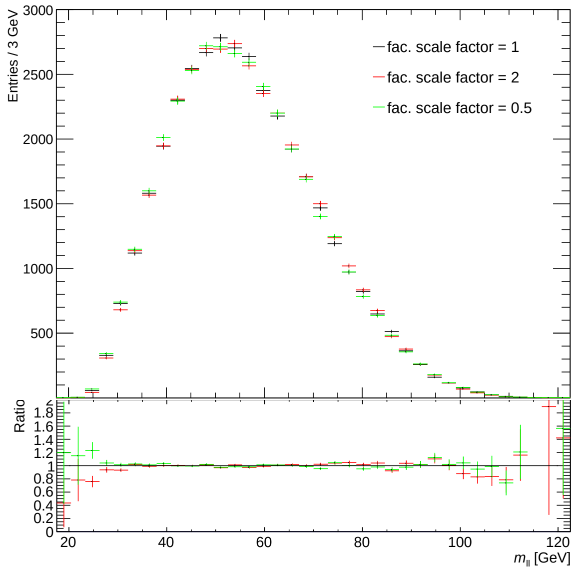


Figure 1.13: Comparison of the visible mass of tau decay products after factorisation scale variation for the b-veto category on a gluon fusion signal sample.

directly to the embedded sample, no further acceptance uncertainties is considered. Acceptance systematics on $t\bar{t}$ simulated events are estimated to be of 2%. The acceptance uncertainties on diboson and single top production are assumed to be 2%. Uncertainties on signal acceptance have been estimated by producing samples with varied MC generator parameters and evaluating, at truth-level, the effect of analysis selections on leptons, taus and jets. This truth-level study is implemented within the Rivet framework [70], where additionally b-tagging is performed by identifying b-quarks and applying a weighting according to the estimated ATLAS b-tagging efficiencies [66]. The variation of the acceptance with respect to the nominal MC tune has been considered as a source of systematic uncertainty. For signal a total acceptance uncertainty varies from 4% to 30% depending on m_A, production process and on the analysis category.

1.3.3 $Z \to \tau\tau$ Embedding Systematics

An important element of the embedding method is the subtraction of the calorimeter cells associated with the muons in the original $Z \to \mu\mu$ event and their substitution with those from the simulated tau decays. To make a conservative estimate of the systematic uncertainty on this procedure, the energy of the subtracted cells is scaled up or down by 30%. The analysis is repeated with those modified samples and the relative uncertainty is referred as "EMB_MFS", this uncertainty affects mainly the shape of the MMC_{mass} distribution, shown in figure 1.14.

In the selection of the $Z \to \mu\mu$ sample only a loose requirement on muon track isolation is required. A different selection on the muon isolation may effect the selected sample by modifying the topology of the event, changing the non-$Z \to \mu\mu$ contamination or the activity in the calorimeter. To estimate the importance of

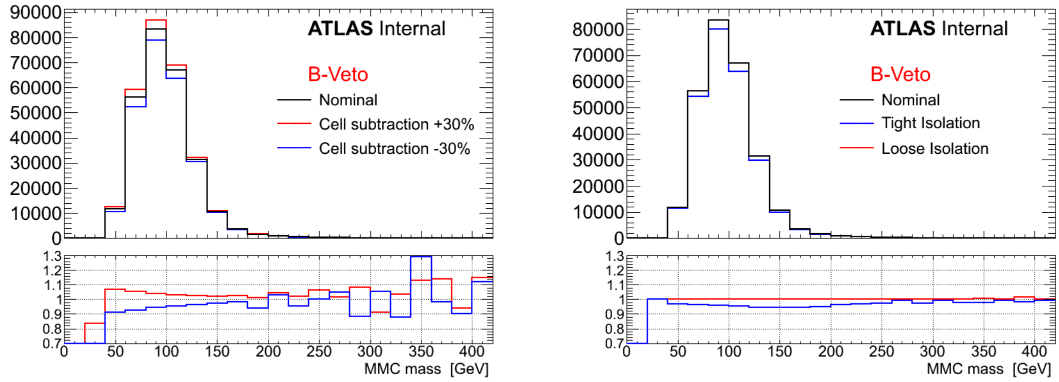


Figure 1.14: Impact of EMB_MFS (left) and EMB_ISO (right) systematic uncertainties MMC_{mass}distribution for Embedding sample. Only the b-veto category report significant deviations.

these effects in our embedding sample, the isolation selection on the muons in the original $Z \to \mu\mu$ events is tightened, a looser selection would have limited impact because of isolation requirements at trigger level. The resulting uncertainty, referred to as "EMB_ISO", affects both the yield and the MMC_{mass} shape of the embedding samples, as shown in figure 1.14.

Finally, because the normalisation of the embedding sample is determined by the use of the ALPGEN sample, the relative cross section and luminosity uncertainties are assigned. In addition all the detector-related systematic uncertainties relevant to the decay products of the simulated tau decay are propagated to the embedding sample.

1.3.4 QCD Multi-Jet Systematics

In this analysis the QCD multi-jet background is estimated via the ABCD method, as described in Section 1.2.2. This technique relies strongly on the assumption that the lepton isolation variables are independent from the charge correlation between the two leptons. Systematic uncertainties are assigned to take into account deviations from this assumption. First the correlation between R_{QCD} and the lepton isolation selections is considered, then the result is compared with an auxiliary method.

Figure 1.15 shows the R_{QCD} factor, the ratio between the QCD yields in region C and D, as a function of the lepton isolation selections (red points). As described previously, the expectation from non-QCD backgrounds is subtracted from the data in regions C and D. To estimate the uncertainty on the value of R_{QCD} an additional transfer factor is defined as follows: $R^{iso}_{QCD} = \hat{A}/\hat{B}$, where $\hat{A}$ and $\hat{B}$ are semi-isolated OS and SS regions defined with the lepton isolation larger than the standard requirement, but less than a sliding cut. Once more, the non-QCD contributions are subtracted from the data yields. The regions $\hat{A}$ and $\hat{B}$ are chosen to be semi-isolated due to the high contamination of non-QCD background and possible signal in region A and B. Figure 1.15 shows R^{iso}_{QCD} as a function of the lepton isolation

Selection	R_{QCD}	R^{AB}_{QCD}	R^{iso}_{QCD}
Preselection	1.929 ± 0.004	2.12 ± 0.17	2.22 ± 0.16
B-veto	1.965 ± 0.005	2.10 ± 0.16	2.22 ± 0.16
B-tag	1.78 ± 0.02	1.9 ± 0.9	2.0 ± 0.8

Table 1.12: Comparison between R_{QCD}, R^{AB}_{QCD} and R^{iso}_{QCD} for early stage in the cutflow, only b-tag and b-veto requirement are applied after preselections. Reported is statistical uncertainty only.

selections (black points). The difference between R_{QCD} and R^{iso}_{QCD} in the vicinity of the standard cut value is then assigned as a systematic uncertainty on R_{QCD}. Using the point where the cuts on the lepton isolation are twice their standard values, a systematic uncertainty of 15% is found. The plot in Figure 1.15 is made at preselection level, similar plots using the full selection for the two categories are in Appendix **??**.

An additional method, used as a crosscheck, considers calculating R_{QCD} as the ratio between the estimated QCD contributions in region A and B. Here the non-QCD contributions are once more subtracted from data. However the large contribution of this non-QCD background, along with lack of statistics and possible signal contamination, lead to this method being only used as a cross check. Table 1.12 shows a comparison between R_{QCD} and R^{AB}_{QCD} for the two categories at the preselection stage of the cutflow, where signal contamination is negligible. Agreement is seen between R_{QCD} values in the two regions, within statistical uncertainties.

The difference in MMC_{mass} shape observed between the OS and SS anti-isolated regions (C and D) is shown in Figure 1.16. This effect is within the uncertainty on R_{QCD} of the ABCD method, hence no correction factor is applied to the mass shape. We assume, however, that there could be the same shape difference in the isolated regions, a shape uncertainty is then assigned to region B to take into account this deviation. Further shape uncertainties due to non-QCD background subtraction are found to be negligible. The uncertainty due to the use of an isolation requirement at trigger level is discussed in Appendix **??** and is found to be negligible.

1.4 Results

1.4.1 LHC Procedure For Limits Setting

A detailed description of the LHC procedure for Higgs search can be found in [**?**, **?**], in the following a brief summary is given. Statistical tests are used for quantify an observation or to set an exclusion limit, in search for new phenomena, hypotesis testing is performed by means of two hypotesis: the *background only* H_0 and and the *signal+background* H_1. As its has allready been outlined in section **??**, any statistical test is based on probability distribution, once a probability density function (p.d.f.) is defined, one can calculate its value for a given set of data obtaining what is called a "likelihood". Taking the marked Poisson p.d.f. in equation (**??**)

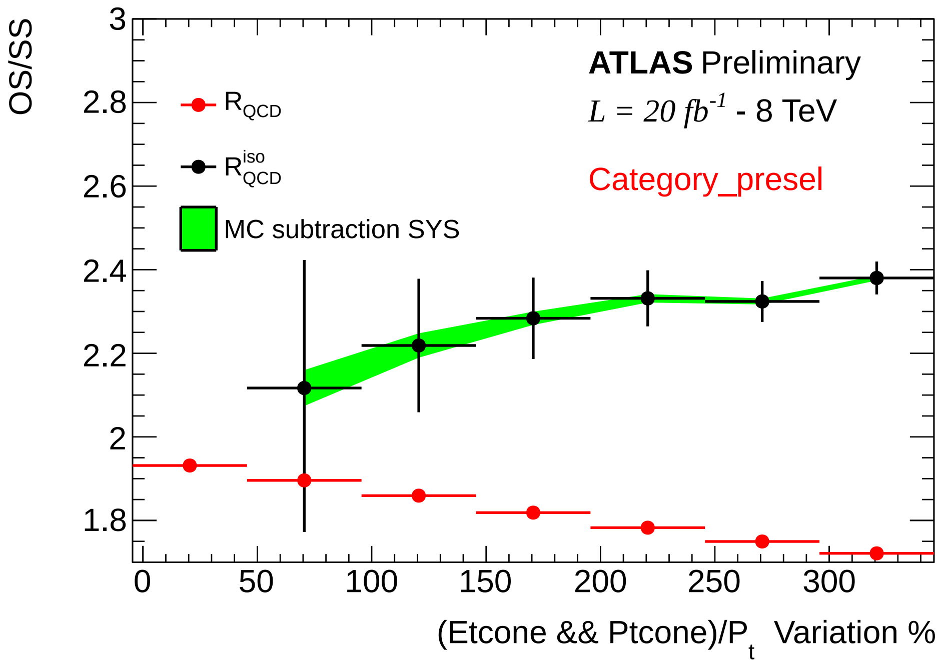


Figure 1.15: OS/SS ratio as a function of lepton isolation variable selections. The selections are varied as a percentage relative to the standard lepton isolation cut values (0 in the plot). The red points show the anti-isolated scale factor R_{QCD}, i.e. the ratio between regions C and D. The black points show the isolated scale factor, which is defined as the ratio between region $\hat{A}$ and $\hat{B}$, where the leptons have isolation values larger than the nominal value but smaller than the sliding cut on X axis.

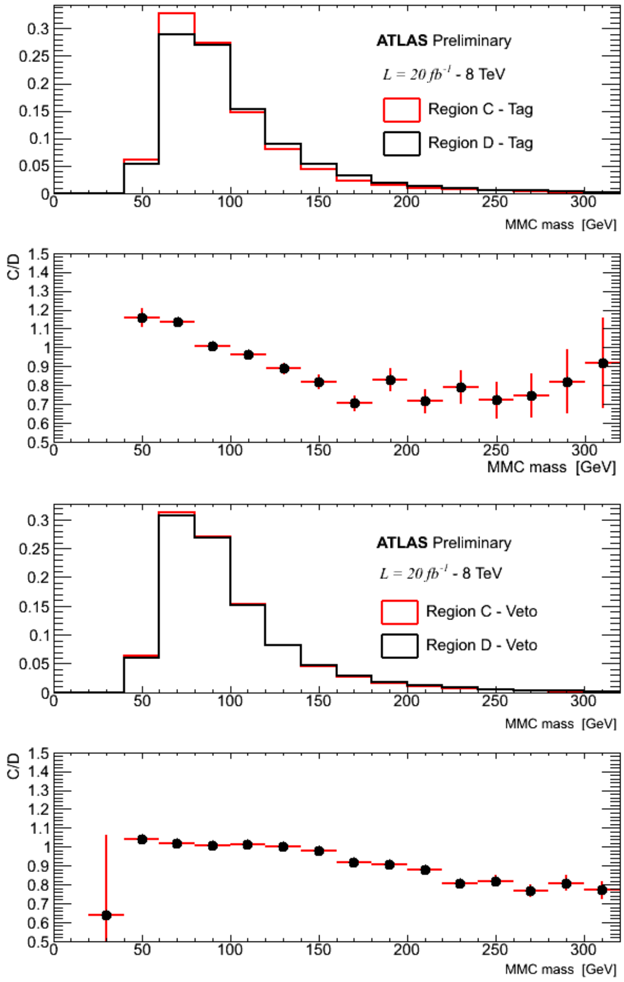


Figure 1.16: Shape differences for the b-tag and b-veto categories between the ABCD regions C and D.

Generator	Process	Uncertainty
ALPGEN	$Z \to \tau\tau/ee/\mu\mu$	$\pm 5\%$
POWHEG	$t\bar{t}$	$\pm 5.5\%$
ALPGEN	$W \to \tau\nu/e\nu/\mu\nu$	$\pm 5\%$
AcerMC	single top	$\pm 13\%$
HERWIG	dibosons	$\pm 6\%$
SHERPA	$bbA/h/H$ ($m_A \geq 120$ GeV)	$-(< 20)\%$, $+(< 9)$ %
SHERPA	$bbA/h/H$ ($m_A = 110$ GeV)	$-(< 25)\%$, $+(< 9)$ %
SHERPA	$bbA/h/H$ ($m_A = 100$ GeV)	$-(< 28)\%$, $+(< 9)$ %
SHERPA	$bbA/h/H$ ($m_A = 90$ GeV)	$-(< 30)\%$, $+(< 9)$ %
POWHEG	$ggA/h/H$ ($m_A \leq 300$ GeV)	$< 15\%$

Table 1.13: Cross-section uncertainties for background and signal samples.The reported signal samples are all for $\tan\beta = 20$.

one obtain the following likelihood function:

$$\mathcal{L}(\text{data}|\mu, \boldsymbol{\theta}) = \text{Poisson}(\text{data}|\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) \cdot f(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) \tag{1.3}$$

this now describes how likely are the data under a certain hypotesis and it is only function of the parameter μ and of the nuissance parameter $\boldsymbol{\theta}$. If the hypotesis under test is unlikely to happen with the given dataset the value of $\mathcal{L}$ is decreasing, one can define which is the best value of a parameter that describes the data via maximising the likelihood, obtaining a so called maximum likelihood estimator. The Poisson distribution in equation (1.3) stands for a product of Poisson probabilities to observe events in the bin i of an histogram:

$$\prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \, e^{-\mu s_i - b_i}$$

while the $f(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$ is the p.d.f. for a given set of nuissance parameter $\boldsymbol{\theta}$ with their best estimate $\hat{\boldsymbol{\theta}}$.

To compiute the compatibility of the data with the H_0 and H_1 hypotesis and then exclusion limits, one needs to define a test statistic. The test statistic, which has already been mentioned in section **??**, is a function of the data which returns a real value. One can in principle use any test statistic, however, given the size of the test (probability to reject the null hypotesis when is true) one would like to have a test statistic which has the highest power $1 - \beta$ possible (probability to reject the null hypotesys when it is false). Figure 1.17 shows an example of the distribution of an hipotetical test statistic for two hypotesis. It has been shown by Neuman-P [] that in case of simple hyptesis (probability model without any parameter), then the test statistic with the highest power is the ratio of the likelihood calculated with the two hypotesis. The standard procedure at the LHC is to use the following test statistic [] based on the likelihood ratio:

$$\tilde{q_\mu} \;=\; -2\ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad \text{with the constraint} \quad 0 \leq \hat{\mu} \leq \mu$$

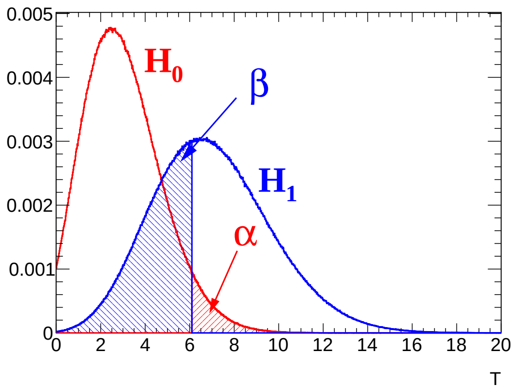


Figure 1.17: Example of a test statistic which in this case is just the total number of events of a counting experiment. Under the hypotesis H_0 are expected four events, while under the H_1 seven events are expected.

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the maximum likelihood estimators for μ and $\boldsymbol{\theta}$ given the data, whereas $\hat{\boldsymbol{\theta}}_\mu$ is the maximul likelihood estimator of $\boldsymbol{\theta}$ given the data but considering a signal streght of value μ, $\tilde{q}_\mu$ is increasing with increasing disagreement between data and the μ hypotesis under test. The procedure for limits setting follows five steps:

1. The signal hypotesis with signal streght μ is assumed, under this assumpion a set of *pseudo-data* is generated for different values of μ.

2. $\tilde{q}_\mu$ is calculated for each of the *pseudo-dataset* and each signal hypotesis generating the expected probability density function for $\tilde{q}_\mu$ given μ, $f(\tilde{q}_\mu \mid \mu, \hat{\boldsymbol{\theta}}_\mu, H_1)$.

3. One does the same thing for the null hypotesis, generate pseudo-data with the distribution of background only and obtain the $f(\tilde{q}_\mu \mid \mu = 0, \hat{\boldsymbol{\theta}}_0, H_0)$.

4. Once the p.d.f. for the signal and signal + background hypotesis is obtained, one can define for a given dataset (that can be this time real data or again pseudodata) two p-values for any given value of μ, which are the probability to obtain data less compatible with the hypotesis in consideration:

$$p_{s+b} = P(\tilde{q}_\mu > \tilde{q}_\mu{}^{observed} \mid H_1)$$

$$p_b = P(\tilde{q}_\mu > \tilde{q}_\mu{}^{observed} \mid H_0)$$

   The ratio of this two probability is what is called the $CL_s = p_{s+b}/p_b$ [].

5. If for a given μ is obtained $CL_s \leq \alpha$ one states that the signal hypotesis (with that μ) is excluded with (1 - α) CL_s confidence level. To get the 95% confidence level upper limit on μ, denoted as μ^{95} one adjust μ until $CL_s = 0.05$.

This is a quite complicated prescription, however its interpretation is not so different from the usual Neyman Costruction [] of confidence intervals: for each μ is possible to define $\tilde{q_\mu}^{95}$ for which the probability P($\tilde{q_\mu} \geq \tilde{q_\mu}^{95} \mid \mu,\ H_1) = 5\%$, this means that if H_1 is true one expects $\tilde{q_\mu} \geq \tilde{q_\mu}^{95}$ in 5% of the cases. With this definition μ^{95} would be the value of μ that for the observed data gives $\tilde{q_\mu} = \tilde{q_\mu}^{95}$, or in other words a p-value of 5%. By costruction, rejecting $\mu > \mu^{95}$ the hypotesis H_1 will be rejected, when is true, at most 5% of the time, given the fact that $\tilde{q_\mu}$ is increasing with increasing discrepancy of the hypotesis with data. The difference with the CL_s prescription is that there the ratio of p-values is used to define μ^{95}: it has been shown that this choice protect the upper limit from down fluctuation of the data, giving a conservative estimate in any case.

The expected median exclusion upper-limit and its error are evaluated by generating a large sample of *bacground only* pseudo-data and calculating CL_s and μ^{95} for each of them, from the distribution of μ^{95} one can get the mean excluded value and its error.

The actual implementation in the limit framework of the ABCD method follows that suggested in [63]. Here three free parameters are fitted: number of multi-jet events in region B, N_B^{QCD}, factor that extrapolates from SS region to OS regions, R_{QCD}, and the factor that extrapolates from isolated to anti-isolated regions R_{BD}. Neglecting signal contributions, the following equations can be written for the event yield of the B,C and D control regions:

$$N_B = N_B^{BKG} + N_B^{QCD}$$

$$N_C = N_C^{BKG} + N_B^{QCD} \times R_{QCD} \times R_{BD}$$

$$N_D = N_D^{BKG} + N_B^{QCD} \times R_{BD}$$

where N^{BKG} represent the prediction of non-QCD background in the relative regions. The estimate of multi-jet event yield in SR will be then $N_B^{QCD} \times R_{QCD}$. This method is particularly powerful because in the best fit of R_{QCD} the statistical and systematics uncertainty for non-QCD backgrounds and data will be considered.

1.4.2 Exclusion Limits

The procedure described in section 1.4.1 is the one used for the SM Higgs, for the MSSM further complication arises: one has to consider in the signal model three Higges, in a particular scenario the masses and cross section are defined for a given point in the $\tan\beta - m_A$ plane, so the procedure described previously has to be repeated for each point in that plane. For the m_h^{max} scenario exclusion limits are derived by calculating 95% CLs limits on the cross section of $bb/gg \rightarrow A/H/h \rightarrow \tau_{lep}\tau_{lep}$ for 15 tanβ values (between[2] tanβ = 5 and tanβ = 60), a point in the $\tan\beta - m_A$ plane is excluded if $\mu^{95} \leq 1$ for that point, a linear interpolation is used to determine the $\tan\beta$ excluded for a given m_A. The procedure is followed for a set of different CP-odd Higgs masses m_A: 90, 100, 110, 120, 125, 130, 140, 150, 170, 200, 250 and 300 GeV. The event yield has been compared between data and background expectation in bins of the MMC_{mass} distribution. The bin sizes were chosen such that there are enough events left for the asymptotic approximation [71]

[2]The set of tanβ values used is 5, 8, 10, 13, 16, 20, 23, 26, 30, 35, 40, 45, 50, 55, 60

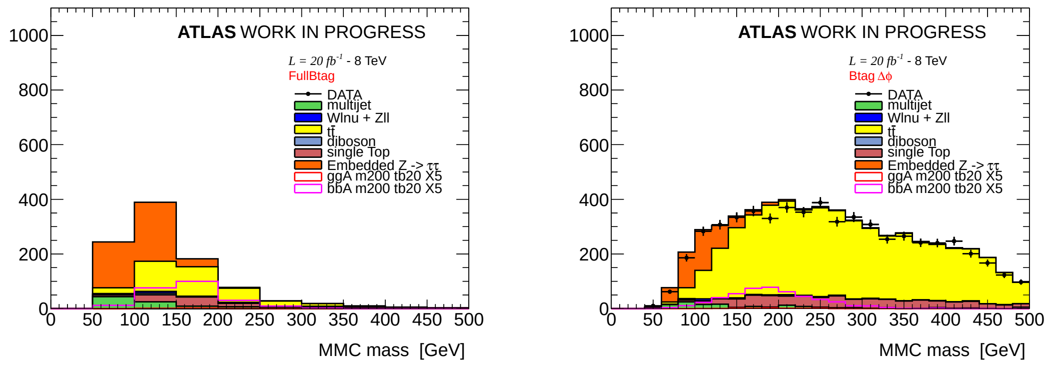
Sample	b-tag category			b-veto category		
	N(event)	Stat.	Syst.	N(event)	Stat.	Syst.
$Z \to \tau\tau$	418	$\pm$ 6		54680	$\pm$ 60	
$t\bar{t}$	330	$\pm$ 10		2228	$\pm$ 25	
Multijet	100	$\pm$ 15		3940	$\pm$ 330	
$W \to \ell\nu$	10	$\pm$ 6		650	$\pm$ 100	
Diboson	13.1	$\pm$ 1.8		2921	$\pm$ 27	
Single Top	90	$\pm$ 6		443	$\pm$ 15	
$Z \to \ell\ell$	0.9	$\pm$ 0.8		430	$\pm$ 40	
Total	962	$\pm$ 16		65290	$\pm$ 180	
Signal						
Data	-	-	-	-	-	-

Table 1.14: Comparison between yield in data and the one expected from our background model, b-tag and b-veto category are reported separately.

to hold. Table 1.14 compares yields between data and background model for the two categories at the final stage of the cut flow. Additionally, figure 1.18 shows the MMC_{mass} distributions for the full b-tag and b-veto categories.

The resulting exclusion limit on the MSSM parameter space (m_A vs $\tan\beta$ plane) are interpreted within the m_h^{max} benchmark scenario [35] and shown in Figure 1.19. The expected and observed 95% confidence-level limits are shown as solid and dashed black lines, the green and yellow bands correspond to the 1σ and 2σ error bands. The analysis is sensitive to MSSM Higgs production of $\tan\beta \geq 13$ for the range $90 < m_A < 200$ GeV. The observed limit is presently unknown.

The outcome of the search is also interpreted in the generic case of a scalar boson produced in the $pp \to gg \to \phi$ or $pp \to bb\phi$ mode and decaying to a di-tau pair. These limits are shown in Figure 1.20 for the b-associated and the gluon-gluon fusion production mechanisms separately. All signal systematic uncertainties are implemented in the likelihood for this limit derivation, more information about the limits and their validation can be found in Appendix **??**.



(a) (b)

Figure 1.18: Distributions of the MMC_{mass} mass for (a) the full b-tag category selection and (b) the full b-veto selection.

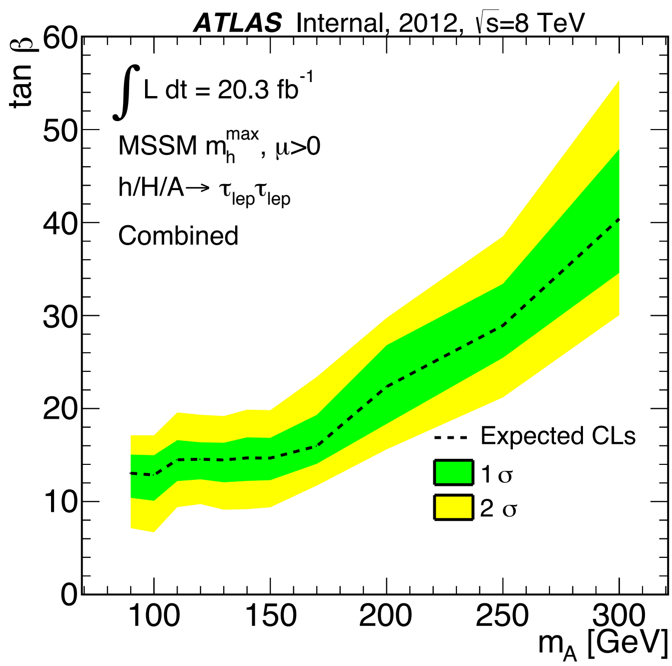


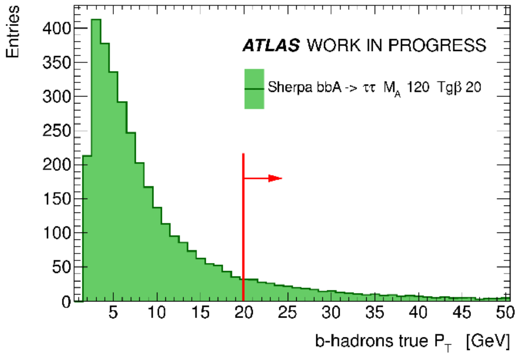
Figure 1.19: Expected exclusion limits for MSSM Higgs boson production in the MSSM m_A vs $\tan\beta$ parameter space. Combination between b-tag and b-veto category.

Figure 1.20: Limits on the production of a scalar particle decaying to a di-tau pair and produced in association with b quarks (left) or via gluon-gluon fusion (right). Still not produced...

Chapter 2

Prospects for Neutral MSSM Higgs Search Improvement

The neutral MSSM Higgs boson search, described in the previous chapter, suffers strongly of poor b-tagging performance due to the particular phase space required, this bound the potential of this search, improving b-tagging would result in a major improvement of the search sensitivity. This chapter investigates an alternative to the commonly used calorimeter jets in ATLAS, which is trackjets b-tagging. The prospects for successfully use trackjets b-tagging in the future neutral MSSM Higgs boson search are reported, b-tagging on trackjets was never attempted before. In section 2.1 an introduction to the b-tagging challenges of the analysis and to trackjets is given. Section 2.2 presents trackjets performance on b-tagging in comparison with calorimeter jets, preliminary results on the impact of trackjets to the analysis are also described here. Finally, in section 2.3 an evaluation of trackjets systematic uncertainties is presented.



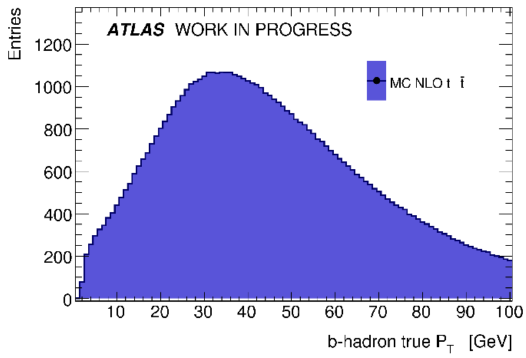


Figure 2.1: Comparison of simulated b-hadron distribution for signal b-associated production events (left) and $t\bar{t}$ events (right). The red line in the figure shows the acceptance region due to calibrated jet P_T requirements.

2.1 Introduction to Trackjets

The neutral MSSM Higgs search, as described in chapter 1, splits the dataset in two category by means of the presence or the absence of a b-tagged jet, the b-tagged category is optimized for the b-associated production mechanism, in which the Higgs is produced in association with two b-jets. Figure 2.1 shows a comparison between the P_T spectrum of simulated b-hadron in $bb/A/h/H$ production and $t\bar{t}$ events, the signal prefers b-hadron with relatively low transverse momentum, which is actually the major challenge for the b-tag category. Due to the high amount of pileup and ambient energy density in the events, calorimeter jets are not calibrated below 20 GeV in P_T (see chapter **??**), systematic uncertainties and performance are also not evaluated below this threshold, this means that, currently, the low transverse momentum phase space is not accessible to canonical calorimeter jets (*calojets* in the following). Calojets are then inconvenient for $bb/A/h/H$ production and one of the major reason for sensitivity lost in the b-tag category. Another challenge to this search is the drop in b-tagging performance at low transverse momentum, the MV1 tagger (see chapter **??**) efficiency , in fact, decreases rapidly with jet P_T, reaching a minimum of 50% at 20 GeV [66, 67] (using the tagging point with 70% efficiency).

A solution to access jets with low transverse momentum in to use *trackjets* instead of calojets. Trackjets are anti-kt jets (see chapter**??**) reconstructed by clustering inner detector tracks, for them it is possible to take advantage of the tracks longitudinal (z) impact parameter information and build trackjets in three dimensions $\eta - \phi - z$. Trackjets will then contains only tracks originating from the same interaction point (reconstructed vertex), this feature make them very robust with respect to pileup. B-tagging has never been tested before on trackjets, in section 2.2 the first study of b-tagging over trackjets performances is reported.

Trackjets are builded in the ATLAS reconstruction software by the *TrackZTool*, this runs the anti-kt clustering algorithm on a subset of tracks which can be defined by the user. For the purposes of this thesis trackjets are reconstructed out of tracks that passes the following quality selection criteria:

Process	MC Generator	Purpose
Minimum bias	Pythia	Systematics study
$b\bar{b}$	Alpgen	Performance for low P_T b-tagging
$Z \to \tau\tau$	Pythia	Impact on the MSSM Higgs search
$t\bar{t}$	MC@NLO	Impact on the MSSM Higgs search
MSSM $bb/A/h/H$	Sherpa	Impact on the MSSM Higgs search

Table 2.1: Monte Carlo simulation sample produced for the studies reported in this chapter.

- $|z_{track} - z_{PV}| < 2$ mm, The track should be associated to the primary vertex (PV).
- $|z_{PV} * sin(\theta)| < 1.5$ mm, which is a measure of how much the track is pointing to the PV in the plane that contain the beam axis.
- $d_{PV} < 1.5$ mm, where d_{PV} is the distance of minimum approach of the track to the primary vertex in the plane orthogonal to the beam axis.
- At least one pixel hit and at least 6 SCT hits (including SCT holes).
- At least one b-layer hit if expected (i.e. the module passed by the track was active).
- $|\eta| < 2.5$
- $P_T > 300$ MeV
- To build a trackjets is necessary to cluster at least two tracks
- A trackjet is produced and stored if the sum of its tracks has $P_T > 2$ GeV.

it has been shown that those selections, together with a maximum cone size for clustering of $\Delta R = 0.6$, are the best compromise between quality requirements, aimed to control fake tracks, and b-hadron reconstruction efficiency. Several MC simulation samples has been produced with the purpose of studying trackjets performance, trackjets were reconstructed and b-tagged using an ad-hoc implementation of the TrackZTool within the ATLAS software framework, table 2.1 reports a summary of the produced samples along with their usage in this thesis.

2.2 Trackjet Performance

Many analysis could profit from an enhanced b-jet reconstruction efficiency at low P_T, the studies presented in this section are aimed to compare performance of common b-tagging algorithm and b-jet reconstruction efficiency between calojets and trackjets, these studies are specially focused on low transverse momentum.

Despite trackjets are more robust with respect to pileup, which makes them appealing, they can only reconstruct the charged part of the jet, the neutral part is

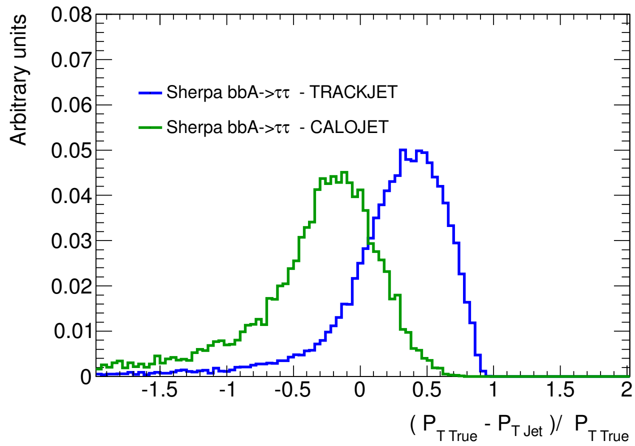


Figure 2.2: Residuals comparison of trackjet and calojet P_T with respect simulated jet P_T.

lost. According to isospin invariance the expected charged fraction in a jet is roughly 2/3 of the total, the trackjet momentum will be then shifted and its direction will have a larger uncertainty. Figure 2.2 shows a comparison of trackjet and calojet transverse momentum residuals with respect to *truthjet* P_T (reconstructed jets from truth particle), here truthjet are matched with jets within a ΔR cone of 0.4 (jet splitting effect are resolved by matching with the nearest jet). The trackjets energy shift may be critical for b-tagging algorithm since some of them strongly rely on the measurement of jet axis and jet P_T.

To compare performance of trackjet and calojet an anti-kt cone size of $\Delta R = 0.4$ is chosen, if a reconstructed jets lies within $\Delta R < 0.3$ from a simulated b-hadron in the event, this jet is said to *match* with a b-hadron. *Reconstruction efficiency* is then defined as the ratio between the number of matched b-hadron and the total number of b-hadron within inner detector acceptance. Figure 2.3 compare b-hadron reconstruction efficiency between calojet and trackjets, the latter shows a higher reconstruction efficiency for low transverse momentum due to their robustness to pileup.

2.2.1 B-tagging on Trackjets

Performance of b-tagging algorithms are usually described by means of tagging efficiency and rejection power. The *tagging efficiency* is the fraction of matched jets which passes a determined selection on a tagging algorithm, i.e. which are *tagged*. The *rejection* is the inverse of the misidentifying rate, i.e. the inverse of the fraction of the jets which are not matched with a b-hadron or c-hadron, but are tagged. Fixing the selection value for a given tagging algorithm will fix a point in the efficiency-rejection plane, this is a convenient way to compare performance of b-tagging algorithms and is shown in figure 2.4 for trackjets. Figure 2.5 instead shows the rejection as a function of trackjets P_T for the tagging point which gives 50% tagging efficiency. Mistagging rate is rapidly increasing for low transverse momentum trackjets, revealing the necessity of a dedicated tagging algorithm for

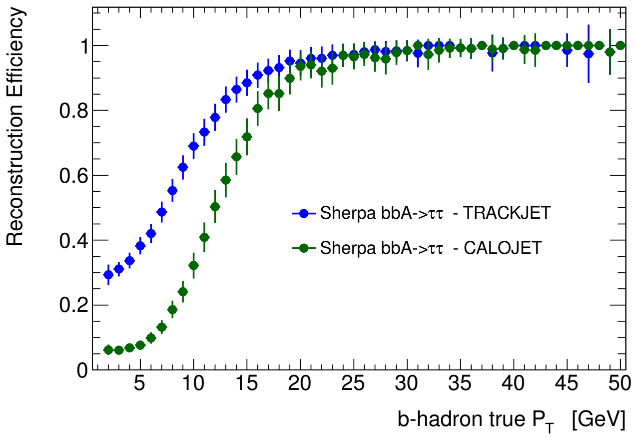


Figure 2.3: Comparison of b-hadron reconstruction efficiency for trackjet and calojet as a function of the simulated b-hadron P_T. Note that calojet and trackjet have a requirement at reconstruction level to be respectively with $P_T > 7$ and 2 GeV, a fair comparison in this plot is only possible above 10 GeV in P_T.

low P_T jets.

The previously introduced rejection and tagging efficiency do not allow a fair comparison between trackjets and calojets, the latter, in fact, can be reconstructed also in case no tracks are associated with them, in this case any tagging algorithm would likely fail altering the rejection distribution. It is convenient to use instead the following quantities: *effective rejection*, which is the inverse of the number of mistagged jets per event, and the b-hadron reconstruction efficiency, which is defined above. Figure 2.6 shows a comparison between calojets and trackjets for the two variables just defined, for a given b-hadron reconstruction efficiency trackjets can achieve higher rejection, which is quite promising. For a fair comparison with calojets, trackjets in figure 2.6 are selected in the transverse momentum range between 4 and 33 GeV, while calojets between 8 and 50 GeV, this corresponds to the same range: figure 2.2 in fact, is only valid for low P_T jets and the fraction of momentum lost approaches 1/3 for high P_T trackjets. In conclusion, thanks to the higher b-hadron reconstruction efficiency, trackjets are more suitable than calojets for low transverse momentum b-tagging.

2.2.2 Impact of Trackjet on the Analysis

The impact of trackjets on the neutral MSSM Higgs search is tested in a preliminary study and reported in what follows. Preselections[1], as defined in in section 1.1.2, are applied to MC samples of signal and backgrounds with the following exceptions on the definition of taggable jets:

- Calorimeter taggable jets should have $|\eta| < 2.5$ and $20 < P_T < 50$ GeV.
- Track taggable jets should have $|\eta| < 2.5$ and $5 < P_T < 33$ GeV.

[1]This study has not been updated with the newest version of the object reconstruction selections and corrections, a difference of the order of 10% is expected with respect the numbers in table 1.3.

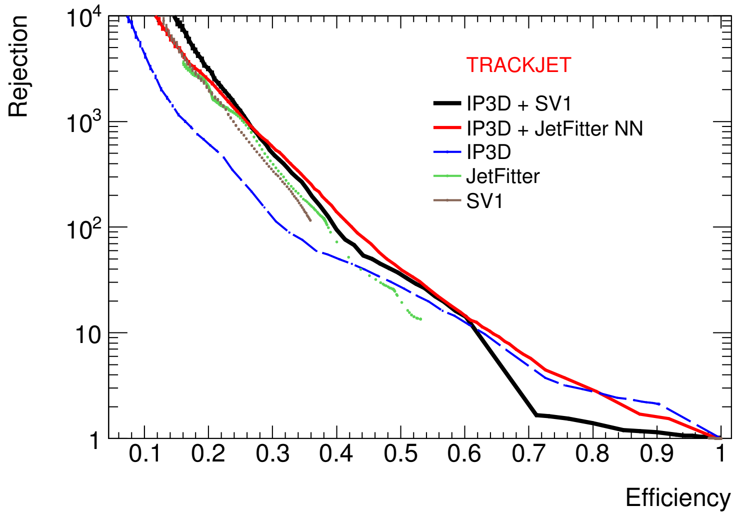


Figure 2.4: Rejection as a function of the tagging efficiency for different ATLAS tagging algorithm tested on trackjets.

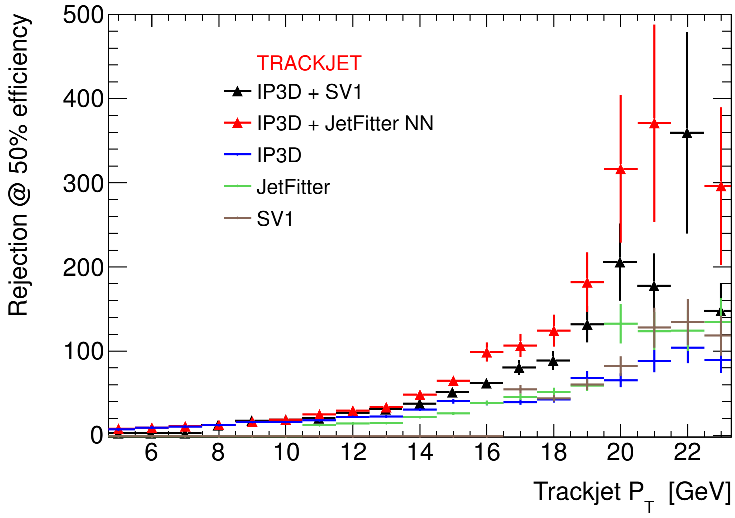


Figure 2.5: Rejection as a function of the transverse momentum of the trackjet for the tagging point which gives 50% tagging efficiency for that P_T value. Different ATLAS tagging algorithm are reported.

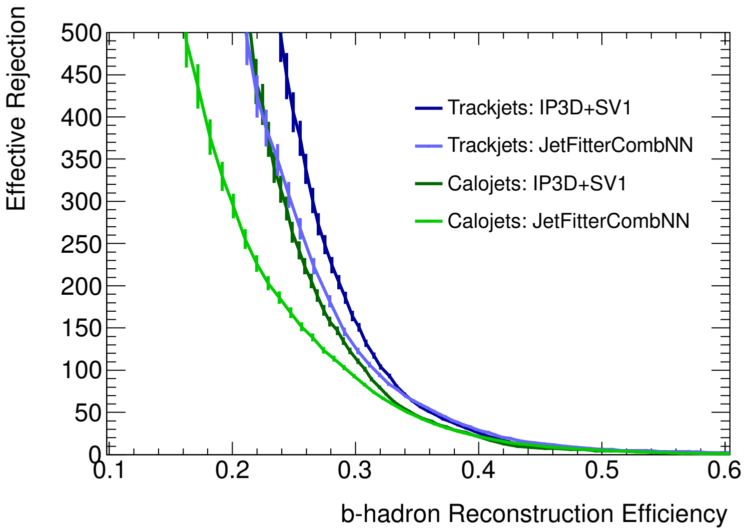


Figure 2.6: Effective rejection as a function of b-hadron reconstruction efficiency, trackjet and calo jets are compared for two different ATLAS tagging algorithms. Trackjets are selected in the transverse momentum range between 4 and 33 GeV, while calojets between 8 and 50 GeV.

- The tagging algorithm used is "IP3D+SV1" at its 70% tagging efficiency point.

the event yields for $bbA/h/H$ production, $Z \rightarrow \tau\tau$ and $t\bar{t}$ (the two most important backgrounds of the b-tag category) are reported in table 2.2 normalized to an integrated luminosity of 1 fb^{-1}. Along with the preselection yields other interesting selections are reported, those are X for the characterization of jet and b-tagging impact on the analysis, calojets and trackjets yields are compared for each of them. As expected, after requiring exactly one b-tagged jet, trackjets presents higher efficiency on signal and higher rejection of top background, which is the most important background for the b-tag category. However, lower transverse momentum requirements on trackjets implies higher tagging fake rates, which is seen as an increase of $Z \rightarrow \tau\tau$ background, this may be also a serious issue for QCD multi-jet background, even tough this is a minor background in b-tag category.

Concluding, the use of trackjets in the b-tag category is very promising and can bring up to twice better sensitivity[2], however, to exploit the full power of this technique a dedicated b-tagging calibration on trackjets is needed, study on algorithm improvements for low P_T b-tagging are also desirable, furthermore, systematics uncertainty on trackjets need to be evaluated. A preliminary study, addressing one of the most important systematics uncertainty for trackjets, is reported in section 2.3.

[2]Note that this estimate is done according to $s/\sqrt{b}$ ratio, considering a counting experiment without systematic uncertainties and only two backgrounds, it represent then the upper limit to the gain in sensitivity with the current b-tagging performance.

Selection	Signal $bbA/H/h$		$Z \to \tau\tau$		$t\bar{t}$	
Preselection	127.2 ± 2.2		3017 ± 8		2066 ± 5	
	Calojet	Trackjet	Calojet	Trackjet	Calojet	Trackjet
At least one taggable jet	47.3 ± 0.8	106.9 ± 1.8	1146 ± 3	2513 ± 7	1804 ± 4	2014 ± 5
Exactly one jet matched b-hadron	18.4 ± 0.3	46.7 ± 0.8	4.5 ± 0.3	18.2 ± 0.5	1054 ± 3	959.1 ± 2.3
Exactly one tagged jet	10.2 ± 0.1	21.0 ± 0.6	37.3 ± 0.5	107 ± 1	777 ± 4	630 ± 4

Table 2.2: Impact of trackjets on the analysis, the event yield is compared between trackjets and calojets. For signal b-associated production is simulated for $\tan\beta = 20$. The yields are normalized to an integrated luminosity of 1 fb^{-1}, all the selections are meant after preselection.

2.2.3 A Novel Technique for low-P_T b-Tagging

this small paragraph will be added if I manage to access trackjets data on MDTRaid16, I just need to reproduce one plot which is missing.

2.3 Systematic Uncertainties on Trackjets

2.3.1 Introduction to Trackjet Systematics

There are several sources of systematic uncertainties on trackjets that may contribute to physics observables mismodeling, those effects are briefly summarized in what follows, the focus is on energy scale and reconstruction efficiency systematic uncertainties.

Uncertainty can arise from MC generator details, like the particular choice of PDF and fragmentation functions, or details of the parton shower and underlying event, challenging to simulate for low transverse momentum object. Those uncertainty can be evaluated by means of a dedicated MC Rivet [70] analysis, they will be dependent on the specific use of trackjets and need to be evaluated case by case.

Energy scale and resolution for single tracks is found to be very well modeled by simulation for tracks above 500 MeV [74], thus, uncertainty on the energy scale and resolution that arise form mismodeling of the pattern recognition algorithm are considered to be negligible.

In dense track environment different tracks may share same hits and this can generate degradation of resolution, fake tracks, loss of track efficiency. Mismodeling of tracks shared hits may affect in general trackjet energy scale, resolution and reconstruction efficiency. This kind of effects has been checked in [76], where calojet energy scale uncertainty are measured using tracks, it has been shown that effects due to tracks hit merging are negligible for jets with $P_T < 300$ GeV.

Mismodeling of the inner detector material budged leads to track reconstruction efficiency mismodeling, which strongly affects trackjets. A methodology to estimate energy scale and reconstruction efficiency uncertainty on trackjets, due to material budget mismodeling, is presented for the first time in section 2.3.2.

2.3.2 Trackjets Uncertainty from Material Budget

An obvious, but rather inconvenient way, to estimate uncertainty due to inner detector (ID) material budget mismodeling, is to produce the relevant MC samples of a given analysis modifying the ID material budget in them. It can be shown that the primary effect of material budget mismodeling influences mainly track reconstruction efficiency (see section 2.3.3), an alternative approach would be then to modify the track efficiency in a given sample according to its uncertainty [75, 77] and build trackjets out of the new collection of tracks. A tool has been made which randomly removes tracks according to reconstruction efficiency uncertainty, trackjets which are build out of this subset of tracks are called in the following *INEF-trackjets*.

A minimum bias MC simulation sample is reproduced containing standard trackjets and INEF-trackjets. A set of "isolated" trackjet with cone size $\Delta R = 0.4$ are selected, isolated means that no other trackjet should be reconstructed within a distance of $\Delta R = 1$. INEF-trackjets are then matched with the original trackjet via cone matching in an event by event basis, the matching fails if no INEF-trackjet is found within $\Delta R = 0.8$ from the original one. Result on the deterioration of the trackjets efficiency and of the energy scale are presented respectively in figure 2.7 and 2.8, these results are based on the current knowledge of inner detector material budged [75]. For low transverse momentum trackjets, uncertainty on the material budget translates into an energy scale shift of 2-4% and in a reduction of the mean number of tracks. This method can only simulate excess of material (reduced track efficiency) but not a lack of material (increased track efficiency), however, for the latter case a symmetric effect is expected.

2.3.3 Track Subtraction Method Validation

The method described in section 2.3.2 depends strongly on the assumption that hadronic secondary interaction, within the inner detector, leads manly to lost of tracks and only in a marginal way to a decrease of tracks quality, a consequence is that material budget mismodeling influences mainly track reconstruction efficiency. In this section, effect of material budget uncertainty on tracks resolution and fake rate are evaluated, this is achieved by means of a simulated sample of minimum bias events, where extra material is added to the ID increasing uniformly of 10% the interaction length.

The requirement on tracks are the ones defined in section 2.1, furthermore a

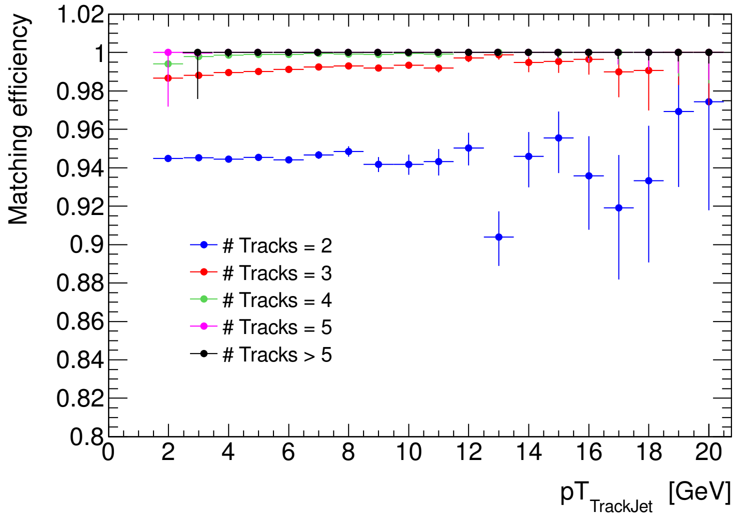


Figure 2.7: INEF-Trackjets are matched with standard trackjets, here is reported the matching efficiency as a function of P_T and number of track of standard trackjet.

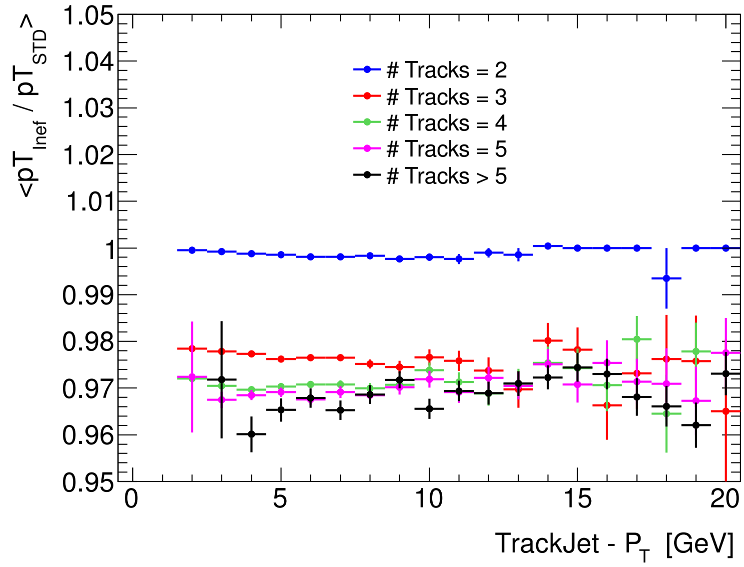


Figure 2.8: INEF-Trackjets are matched with standard trackjets, here is reported the effect on the energy scale as a function of P_T and of the number of tracks of the standard trackjet.

track should be matched within $\Delta R < 0.1$ with a stable[3] simulated particle which should be responsible (alone) of at least 80% of the track hits, tracks that do not fulfil these requirements are called fakes. Fake tracks are tracks that come from a random combination of hits generated from different particles. The track fake rate, shown in figure 2.10, is about 1-3%, the extra material sample has a total increase of the track fake rate of permille. Resolution as shown in figure 2.9 is about 1% for large range of tracks P_T, the total increase of resolution in the extra material sample is also of the order of permille. The deterioration of the tracks resolution and fake rate due to extra material is then negligible compared to the one of track reconstruction efficiency, which undergo to a total decrease in the extra material sample of 1-2%, decrease in efficiency has serious impact on trackjet energy scale. Figure 2.11 shows the ratio of the track reconstruction efficiency of primary particle between the standard and extra material sample.

Results from INEF-trackjets (builded in a standard sample) are also directly compared with trackjets from extra material sample, the comparison is done by means of trackjet-to-truthjet matching (see section 2.2 for truthjets matching) and is reported in figure 2.12 and 2.13 for reconstruction efficiency and energy scale respectively. INEF-trackjets shows to reproduce correctly the effect of extra material either on reconstruction efficiency and energy scale, giving, in most of the cases, a conservative estimate.

[3]Here is intended a Generator stable and interacting particle, which means a charged particle with decay length greater than 1m, also stable particle from secondary interactions are considered.

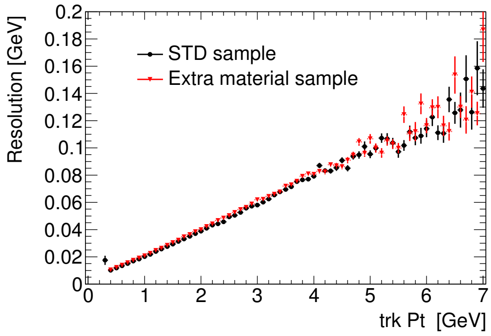


Figure 2.9: Track resolution with respect to matched truth particle as a function of truth particle P_T, for standard Pythia minimum bias sample and 10% inner detector extra material sample.

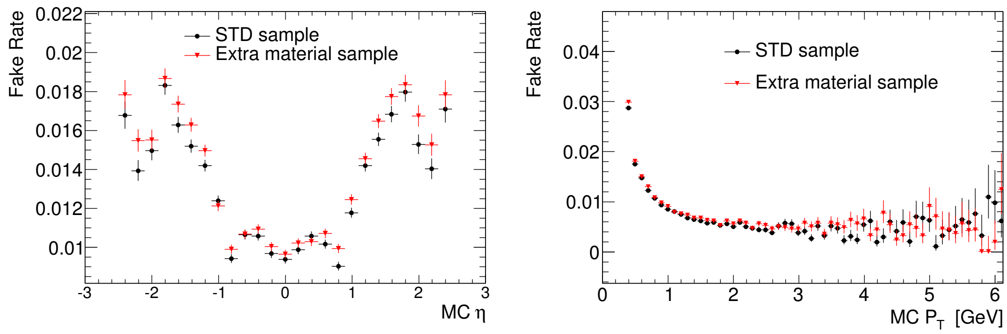


Figure 2.10: Track fake rate resolution as a function of track η (left) and track P_T (right), for standard Pythia minimum bias sample and 10% inner detector extra material sample.

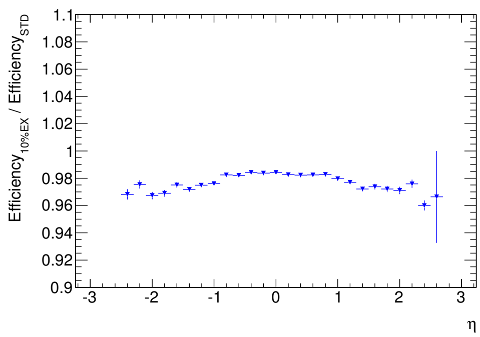


Figure 2.11: Track efficiency with respect to primary truth particle as a function of truth particle η, reported is the ratio between standard Pythia minimum bias sample and 10% inner detector extra material sample.

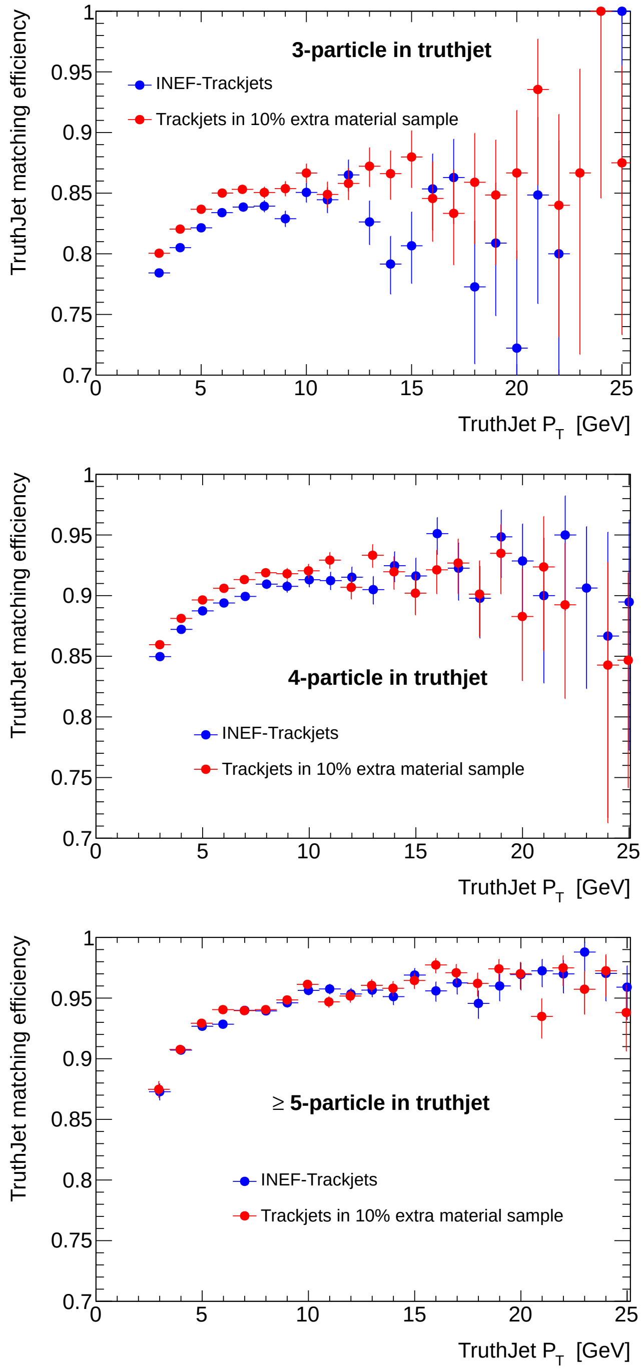


Figure 2.12: Jet reconstruction efficiency with respect to truthjet for INEF-trackjets and trackjets in a 10% extra material sample, in case of 3,4 and ≥ 5 truth-particle. INEF-trackjets always reproduce correctly the inefficiency or give a conservative estimate.

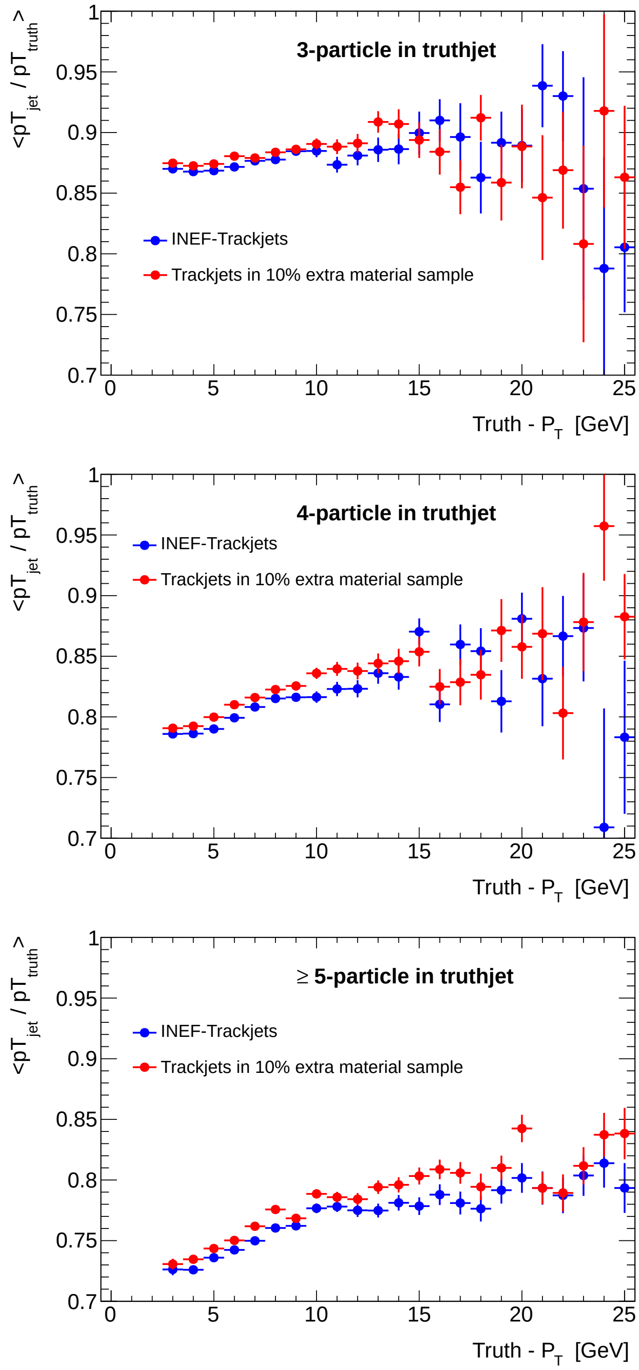


Figure 2.13: Fraction jet transverse momentum with respect to truthjet for INEF-trackjets and trackjets in a 10% extra material sample, in case of 3,4 and ≥ 5 truth-particle. INEF-trackjets always reproduce correctly the inefficiency or give a conservative estimate.

Bibliography

[1] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[2] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[3] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132.

[4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[5] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[6] G. S. Guralnik, C.R. Hagen and T. W. B. Kibble Phys.Rev.Lett. **13** (1964) 585.

[7] N. P. Nilles, *Supersymmetry, supergravity and particle physics*, Phys. Rep. **110** (1984) 1.

[8] H. E. Haber and G. L. Kane, *The search for supersymmetry: Probing physics beyond the standard model*, Phys. Rep. **117** (1985) 75.

[9] ALEPH, DELPHI, L3 and OPAL Collaboration, *Search for neutral MSSM Higgs bosons at LEP*, Eur. Phys. J. **C47** (2006) 547.

[10] *Combined CDF and D0 upper limits on MSSM Higgs boson production in tau-tau final states with up to 2.2 fb^{-1} of data*, `arXiv:1003.3363 [hep-ex]`.

[11] CDF Collaboration, T. Aaltonen et al. Phys. Rev. Lett. **103** (2009) 201801.

[12] D0 Collaboration, V. Abazov et al. Phys. Rev. Lett. **101** (2008) 071804.

[13] TNPWG (Tevatron New Physics Higgs Working Group), CDF and D0 Collaborations, *Search for Neutral Higgs Bosons in Events with Multiple Bottom Quarks at the Tevatron*, `arXiv:1207.2757 [hep-ex]`.

[14] CDF Collaboration, T. Aaltonen et al., *Search for Higgs Bosons Produced in Association with b-quarks*, Phys.Rev. **D85** (2012) 032005, `arXiv:1106.4782 [hep-ex]`.

[15] D0 Collaboration, V.M. Abazov et al., *Search for neutral Higgs bosons in the multi-b-jet topology in 5.2fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys.Lett. **B698** (2011) 97–104, `arXiv:1011.1931 [hep-ex]`.

[16] The CMS Collaboration, S. Chatrchyan et al., `arXiv:1104.1619 [hep-ex] [hep-ex]`.

[17] The ATLAS Collaboration , *Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, `arXiv:1211.6956 [hep-ex]`.

[18] T. A. Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1–29.

[19] T. C. Collatoration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30–61.

[20] S. Heinemeyer, O. Stål and G. Weiglein, *Interpreting the LHC Higgs search results in the MSSM*, Phys.Lett. **B710** (2012) 201–206, `arXiv:1112.3026 [hep-ph]`.

[21] A. Arbey, M. Battaglia, A. Djouadi and F. Mahmoudi, *The Higgs sector of the phenomenological MSSM in the light of the Higgs boson discovery*, JHEP **1209** (2012) 107, `arXiv:1207.1348 [hep-ph]`.

[22] The ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[23] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001.

[24] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. **C53** (2008) 473, `arXiv:0706.2569`.

[25] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, `hep-ph/0204244`.

[26] B. P. Kersevan and E. Richter-Was, *The Monte Carlo Event Generator AcerMC 2.0 with Interfaces to PYTHIA 6.2 and HERWIG 6.5*, `arXiv:0405247v1 [hep-ph]`.

[27] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010.

[28] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton Interactions in Photoproduction at HERA*, Z. Phys. **C72** (1996) 637.

[29] T. Binoth, M. Ciccolini, N. Kauer, and M. Kramer, *Gluon-induced W-boson pair production at the LHC*, JHEP **12** (2006) 046.

[30] A. S. et al., *Higgs boson production in gluon fusion*, JHEP **02** (2009) 029.

[31] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007.

[32] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung, "New generation of parton distributions with uncertainties from global QCD analysis," JHEP **0207** (2002) 012 [hep-ph/0201195].

[33] H. -L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C. -P. Yuan, "New parton distributions for collider physics," Phys. Rev. D **82** (2010) 074024 [arXiv:1007.2241 [hep-ph]].

[34] S. Heinemeyer *et al.* [LHC Higgs Cross Section Working Group Collaboration], "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," arXiv:1307.1347 [hep-ph].

[35] M. Carena, S. Heinemeyer, C. E. M. Wagner, and G. Weiglein, *Suggestions for benchmark scenarios for MSSM Higgs boson searches at hadron colliders*, Eur. Phys. J. **C26** (2003) 601–607, `hep-ph/0202167`.

[36] The ATLAS Collaboration, *ATLAS Monte Carlo Tunes for MC09*, ATL-PHYS-PUB-2010-002.

[37] S. Jadach, J. H. Kuhn and Z. Was, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1990) 275.

[38] E. Barberio, B. V. Eijk and Z. Was, *Photos - a universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66** (1991) 115.

[39] The GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 - a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[40] The ATLAS Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*, ATLAS-SOFT-2010-01-004, submitted to Eur. Phys. J. C., `arXiv:1005.4568`.

[41] The ATLAS Collaboration, *Estimation of $Z \to \tau\tau$ Background in VBF $H \to \tau\tau$ Searches from $Z \to \mu\mu$ Data using an Embedding Technique*, ATL-PHYS-INT-2009-109.

[42] The ATLAS Collaboration, *Search for the Standard Model Higgs boson in the $H \to \tau\tau$ decay mode with 4.7 $fb1$ of ATLAS detector*, Tech. Rep. ATLAS-CONF-2012-014, CERN, Geneva, Mar, 2012.

[43] The ATLAS Collaboration, *Search for the Standard Model Higgs boson $H \to \tau\tau$ decays with the ATLAS detector*, ATL-COM-PHYS-2013-722.

[44] T. S. et al., Z *physics at LEP 1*, CERN 89-08 **3** (1989) 143.

[45] The ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, CERN-OPEN-2008-020, `arXiv:0901.0512`.

[46] The ATLAS Collaboration, *ATLAS Muon Momentum Resolution in the First Pass Reconstruction of the 2010 p-p Collision Data at $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2011-046.

[47] The ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC p-p collision data recorded with the ATLAS detector*, ATLAS-CONF-2011-063.

[48] The ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*, ATLAS-PUB-2011-006.

[49] ATLAS egamma WG, *Electron efficiency measurements*, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EfficiencyMeasurements.

[50] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063.

[51] W. Lampl et al., *Calorimeter Clustering Algorithms : Description and Performance*, ATL-LARG-PUB-2008-002.

[52] T. Barillari et al., *Local Hadron Calibration*, ATL-LARG-PUB-2009-001.

[53] The ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV in ATLAS 2010 data*, ATLAS-CONF-2011-032.

[54] TheATLAS Collaboration, *Performance of the Reconstruction and Identification of Hadronic tau Decays in ATLAS with 2011 Data*, ATLAS-CONF-2012-142.

[55] The ATLAS Collaboration, *Reconstruction and Calibration of Missing Transverse Energy and Performance in Z and W events in ATLAS Proton-Proton Collisions at $\sqrt{s}$=7 TeV*, ATLAS-CONF-2011-080.

[56] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, `arXiv:1012.4686 [hep-ex]`. * Temporary entry *.

[57] ATLAS Jet/EtMiss Combined Performance Group, *Jet Energy Resolution Provider*, https://twiki.cern.ch/twiki/bin/view/Main/JetEnergyResolutionProvider.

[58] The ATLAS Collaboration, *Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV*, ATLAS-CONF-2010-038.

[59] T. A. Collaboration, *Search for neutral MSSM Higgs bosons decaying to $\tau\tau$ pairs in proton-proton collisions at with the ATLAS detector*, Physics Letters B **705** (2011) no. 3, 174 – 192.

[60] The ATLAS Collaboration, *Data-driven estimation of the background to charged Higgs boson searches using hadronically-decaying tau final states in ATLAS,* ATLAS-CONF-2011-051.

[61] The ATLAS Collaboration, *Measurement of the $Z \to \tau\tau$ cross section with the ATLAS detector*, Phys. Rev. D **84** (2011) 112006.

[62] T. A. Collaboration, *Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP , `arXiv:1211.6956`.

[63] Atlas statistics forum, *ABCD method in searches*, link

[64] The ATLAS Collaboration, *Search for Neutral MSSM Higgs Bosons H to $\tau\tau$ to $l\tau_h$ with the ATLAS Detector in 7 TeV Collisions*, ATL-COM-PHYS-2012-094.

[65] The ATLAS Collaboration, *Search for neutral Higgs Bosons in the decay mode $H \to \tau\tau \to ll{+}4\nu$ in proton proton collision at $\sqrt{7}$ TeV with the ATLAS Detector*, ATL-COM-PHYS-2011-758.

[66] The ATLAS Collaboration, *Measuring the b-tag efficiency in a ttbar sample with 4.7 fb^{-1} of data from the ATLAS detector* ATLAS-CONF-2012-097.

[67] The ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment* ATLAS-CONF-2014-004.

[68] The ATLAS Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011*, ATLAS-CONF-2011-116.

[69] T. Sjostrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026.

[70] A. B. et al., *Rivet user manual*, `arXiv:1003.0694 [hep-ph]`.

[71] E. G. G. Cowan, K. Cranmer and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, `arXiv:1007.1727 [hep-ex]`.

[72] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka (Eds.), et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, `arXiv:1101.0593 [hep-ph]`.

[73] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN-2012-002 (CERN, Geneva, 2012) , `arXiv:1201.3084 [hep-ph]`.

[74] ATLAS collaboration *Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at* $\sqrt{s} = 7$ *TeV*, ATLAS-CONF-2010-072

[75] The ATLAS Collaboration, *A measurement of the material in the ATLAS inner detector using secondary hadronic interactions*, arXiv:1110.6191, JINST 7 (2012) P01013

[76] The ATLAS Collaboration, *Validation of the ATLAS jet energy scale uncertainties using tracks in proton-proton collision* $\sqrt{s} = 7$ *TeV*, ATLAS-CONF-2011-067

[77] The ATLAS Collaboration, *Track Reconstruction Efficiency in* $\sqrt{s} = 7$ *TeV Data for Tracks with* $P_T > 100$ *MeV* , ATL-PHYS-INT-2010-112

[78] D. de Florian, G. Ferrera, M. Grazzini and D. Tommasini, *Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC*, JHEP **1111** (2011) , `arXiv:1109.2109 [hep-ph]`.

[79] Statistical twiki, NuissanceCheck. https://twiki.cern.ch/twiki/bin/view/AtlasProtected/Nuisan