UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Search for the Higgs boson at ATLAS/LHC, in associated production with a Z boson

Doutoramento em Física

Mário José da Cunha Sargedas de Sousa

Tese orientada por:
Doutor Agostinho Gomes
Doutora Patricia Conde Muíño

Documento especialmente elaborado para a obtenção do grau de doutor

2017

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Search for the Higgs boson at ATLAS/LHC, in associated production with a Z boson

Doutoramento em Física

Mário José da Cunha Sargedas de Sousa

Tese orientada por:
Doutor Agostinho Gomes
Doutora Patricia Conde Muíño

Júri:
Presidente:
  Doutor António Joaquim Rosa Amorim Barbosa
Vogais
  Doutor Paolo Francavilla
  Doutor Nuno Filipe da Silva Fernandes de Castro
  Doutor José Carvalho Maneira
  Doutora Patricia Conde Muíño

Documento especialmente elaborado para a obtenção do grau de doutor

IDPASC / FCT PhD grant SFRH/BD/51372/2011

2017

Abstract

A mechanism of spontaneous symmetry breaking was used to explain the mass of elementary particles and predicted the existence of the Higgs boson. The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments at the LHC with a mass of about 125 GeV. It now becomes necessary to study this new boson in order to validate the Standard Model of elementary particles.

The Standard Model Higgs boson with a mass of 125 GeV decays most of the times to a pair of b-quarks. However, this decay is very difficult to study in a proton-proton collider like the LHC, due to the production of a huge background of b-jets (and also non-b-jets). In the LHC, the only production process with some chance to be used in this study is the associated production with a vector boson, which can decay leptonically allowing the identification of the event. One can use three possibilities: a Z boson decaying to neutrinos (0-lepton channel), a W boson decaying to an electron or muon and a neutrino (1-lepton channel) or a Z boson decaying to a pair of electrons or muons (2-lepton channel). The latter channel is the subject of this thesis. It is an interesting channel because the two charged leptons reduce the multi-jet background more than the other channels and because this channel reconstructs all the particles in the event.

The analysis of the 2-lepton channel, known as ZH analysis, considers only events with high quality reconstructed particles. The two charged leptons are required to reconstruct the Z boson and the b-quarks which are reconstructed as jets, are required to be identified as b-jets (b-tagging) for the reconstruction of the Higgs boson. A statistical procedure uses the events selected as input to a likelihood fit to obtain normalisations of the simulated backgrounds from data and simulation comparisons and to extract the signal strength parameter, μ, that multiplies the cross section times the branching ratio from the prediction for the Higgs boson from the Standard Model to adjust the ZH signal simulation process to the data.

Both muons and electrons, used in the ZH analysis, are much better reconstructed than jets. So, most of the effort in this thesis was dedicated to the improvement of the jet reconstruction, in particular to the energy scale and uncertainty of the jets. I studied two methods to help reduce the systematic uncertainty of the jet energy scale, the leading experimental uncertainty of several physics analysis, with a important contribution to the ZH search, since the Higgs decays to b-quarks that hadronize into jets.

First, I created a new trigger with the objective of selecting high-momenta isolated charged hadrons to be used in the determination of the calorimeter response to single hadrons, known as the single hadron response. In addition to extending the momentum range from 10 to 30 GeV, it did so in a high-pileup environment (pileup - number of interactions per proton bunch crossing). This contributed to the validation of the test beam determination of the jet energy scale uncertainty. For jets with p_T above 1.8 TeV, for which the current *in situ* techniques are not able to set the jet energy scale uncertainty, the single hadron response is used to obtain this uncertainty.

I have tested a new jet calibration in the ZH analysis: the global sequential calibration (GSC). This method was first developed to reduce the jet energy response dependence on the jet flavour, for gluon and light quark jets. My contribution was the validation and detailed performance studies for b-jets in general, and in particular for b-jets coming from the Higgs boson decay. This new calibration improved the jet energy response by about $2-3\,\%$ and the jet energy resolution by about $20\,\%$. Systematic uncertainties were evaluated for this calibration. I derived specific b-jet systematic uncertainties for this calibration as well as the uncertainties associated to differences between the ATLAS fast and full simulations.

The Standard Model Higgs boson decaying in $b\bar{b}$ is expected to produce a very small excess of events in the invariant mass distribution of the two b-quarks. We need to have the best possible resolution on this variable and since GSC was found to reduce the b-jet jet energy response resolution, it was applied to the $m_{b\bar{b}}$ distribution. GSC improves the invariant mass resolution up to $18\,\%$.

I performed the ZH analysis using a cut-based method and a statistical procedure to obtain the value of the signal strength parameter. A value of $\mu = -0.69 \pm 1.29$ was obtained. This value is negative, but there is an enormous uncertainty, dominated by the statistical uncertainty, and the result is still compatible with the Standard Model prediction for the Higgs boson.

Key words: ATLAS, Higgs decays to b-quarks, ZH associated production, b-jet calibration, Global Sequential Calibration.

Resumo

O mecanismo de quebra espontânea de simetria, que foi utilizado para explicar a massa das partículas elementares do Modelo Padrão, prevê a existência do bosão de Higgs. O bosão de Higgs foi apenas descoberto em 2012 pelas experiências ATLAS e CMS no LHC

com uma massa de cerca de 125 GeV. Torna-se agora necessário estudar este novo bosão de modo a validar o Modelo Padrão das partículas elementares.

O bosão de Higgs do Modelo Padrão, com uma massa de 125 GeV, vai decair para um par de quarks b na maioria das vezes. No entanto, este decaimento é bastante difícil de estudar num colisionador de protões como o LHC, devido ao enorme fundo de jatos b (e jatos em geral) que são produzidos. No LHC, o único processo de produção com alguma viabilidade para ser usado para estudar este decaimento é a produção associada a um bosão vetorial, que pode decair leptonicamente, proporcionando uma assinatura limpa para identificar o evento. Aqui existem três possibilidades: um bosão Z que decai para neutrinos (canal de 0 leptões), um bosão que W decai para um eletrão ou um muão carregado e um neutrino (canal de 1 leptão), ou um bosão Z que decai para dois eletrões ou dois muões carregados (canal de 2 leptões). Este último canal foi o escolhido para esta tese. Neste canal a presença de dois leptões carregados reduz o fundo de multi-jatos mais de que nos outros canais e todas as partículas do acontecimento são reconstruidas.

A análise do canal de dois leptões, referida como análise ZH, considera apenas acontecimentos com partículas reconstruidas de elevada qualidade. Aos dois leptões é exigido que reconstruam o bosão Z, e os quarks b, que são reconstruidos como jatos, têm de ser identificados como jatos b (b-tagging). Um processo estatístico utiliza os eventos candidatos num ajuste de verosimilhança (likelihood) para obter as normalizações dos fundos simulados e extrair o parâmetro de força do sinal, μ, definido como o quociente entre o produto da seção eficaz e a razão de decaimento observados com respeito ao esperado no Modelo Padrão.

Tanto os muões como os eletrões, utilizados na análise ZH, são muito melhor reconstruidos do que os jatos. Assim, o maior esforço desta tese concentrou-se para melhorar a reconstrução de jatos, em particular a escala de energia dos jatos e a sua incerteza. Estudei dois métodos que ajudaram a reduzir as incertezas sistemáticas da escala de incerteza dos jatos, que é a maior incerteza sistemática experimental de várias análises de pesquisa, com uma boa contribuição para análise ZH, uma vez que o bosão de Higgs decai para quarks b que hadronizam para jatos.

Comecei por criar um novo mecanismo de seleção de acontecimentos (trigger) com o objetivo de escolher hadrões carregados isolados de elevado momento, que seriam utilizados na determinação da resposta em energia do calorímetro a hadrões isolados. Graças a este trigger, passou a ser possível utilizar hadrões isolados produzidos em colisões até um

momento de 30 GeV/c, onde este limite era anteriormente de 10 GeV. Com esta amostra de hadrões isolados, pode calibrar-se a resposta do calorímetro em energia, obtendo-se simultaneamente um zona de validação da incerteza determinada para a escala de energia dos jatos hadrónicos a partir de um teste de feixe de partículas com energias bem definidas a partir de 20 GeV. O Trigger mostrou ainda a estabilidade da resposta num ambiente de elevado empilhamento de sinais (pileup - número de interações por colisão de pacotes de protões).

Testei uma nova calibração da energia de jatos na análise ZH: a calibração sequencial global (GSC). Este método, foi em primeiro lugar, desenvolvido para reduzir a dependência da resposta em energia dos jatos no sabor dos mesmos (jatos de gluões ou de quarks leves). A minha contribuição foi a validação e o estudo detalhado do desempenho desta calibração para jatos b em geral, e em particular para jatos b provenientes do decaimento do bosão de Higgs. Esta nova calibração melhorou a resposta da energia dos jatos em $2-3\,\%$ e a resolução em energia destes em $20\,\%$. Avaliei os erros sistemáticos, derivei incertezas sistemáticas específicas para jatos b para esta calibração e ainda uma incerteza associada a diferenças entre simulação rápida e completa do detetor ATLAS.

Espera-se que o bosão de Higgs do Modelo Padrão a decair em $b\bar{b}$ produza um pequeno excesso de eventos na distribuição de massa invariante do par de jatos b. Necessitamos obter a melhor resolução possível nesta variável e, como a calibração GSC mostrou reduzir a resolução em energia dos jatos b, a massa invariante foi também testada com GSC. A utilização desta nova calibração melhorou a resolução em massa até $18\,\%$. Os estudos de validação e desempenho do GSC realizados permitiram a utilização desta calibração na análise do ZH.

Desenvolvi uma análise ZH independente utilizando um método de cortes e procedimento estatístico. Obtive para o parâmetro força de sinal o valor $\mu = -0.69 \pm 1.29$. Embora negativo, este valor possui uma enorme incerteza, dominada pela incerteza estatística, sendo o resultado ainda compatível com a previsão para o bosão de Higgs feita pelo Modelo Padrão.

Palavra-chave: ATLAS, decaimento do Higgs para quarks-b, produção associada ZH, calibração de jatos-b, Calibração Sequencial Global.

Sumário

O Modelo Padrão das partículas elementares é uma teoria quântica de campo que explica três das interações fundamentais da matéria: nuclear fraca, nuclear forte e eletromagnética. A interação nuclear forte é descrita pela simetria de gauge $SU(3)$. As interações eletromagnética e fraca foram unificadas através da simetria de gauge $SU(2) \times U(1)$, pelo trabalho de Sheldon Glashow, Abdus Salam, e Steven Weinberg que lhes valeu o prémio Nobel em 1979. O mecanismo de Englert-Brout-Higgs-Guralnik-Hagen-Kibble, proposto separadamente por três grupos distintos na década de 60 do século XX, explicava a massa das partículas elementares através da quebra espontânea da simetria $SU(2) \times U(1)$. Este modelo prevê a existência de uma nova partícula escalar, o bosão de Higgs. No entanto, o modelo não conseguiu prever o valor da massa desta nova partícula, e apenas existiam alguns constrangimentos teóricos e experimentais. Os aceleradores LEP e Tevatrão apenas conseguiram fixar limites de exclusão com uma um nível de confiança de 95%, para massas menores que 115 GeV e entre 158 e 175 GeV. O acelerador LHC colidiu protões com uma energia de centro de massa de $\sqrt{s} = 7$ TeV e 8 TeV em 2011 e 2012, respetivamente, conseguindo excluir praticamente toda a região de massa e, em 4 de Julho de 2012, as experiências ATLAS e CMS anunciaram finalmente a descoberta do bosão de Higgs com uma massa de cerca de 125 GeV. No ano seguinte Peter Higgs e François Englert partilharam o prémio Nobel da Física.

Após a descoberta do bosão de Higgs e conhecida a sua massa, o Modelo Padrão prevê com grande precisão as suas propriedades. Para esta massa de 125 GeV, o bosão de Higgs possui um vasto leque de possibilidade de decaimento, sendo que o que tem maior probabilidade é o decaimento para um par de quarks-b com 56.7%. No entanto, este decaimento não foi ainda observado devido ao elevado fundo de pares de quarks-b que existe no LHC. Por cada bosão de Higgs produzido no LHC, este produz cerca de 10^7 pares destes quarks, indistinguíveis do decaimento $H \to b\bar{b}$. Apenas utilizando a produção associada do bosão de Higgs com um bosão vetorial ou um quark top, será possível estudar

este decaimento no LHC.

A produção do bosão de Higgs associada à produção simultânea de um bosão Z foi utilizada para pesquisar o decaimento $H \to b\bar{b}$, quando o bosão Z decai para dois leptões carregados. Isto permite escolher assinaturas limpas destes acontecimentos utilizando o sistema de disparo (Trigger) do detetor ATLAS para selecionar eletrões e muões isolados de alto momento transverso. Seleções adicionais na topologia dos acontecimentos permitem ainda aumentar a fração destes provenientes do processo $ZH \to \ell\ell b\bar{b}$. A análise destes acontecimentos, denominada análise ZH, constituiu o objetivo principal desta tese, onde desenvolvi software para esta análise que serviu como uma validação independente das outras análises que existem na colaboração ATLAS para o mesmo canal, com diferenças inferiores a um por mil.

O canal de pesquisa $ZH \to \ell\ell b\bar{b}$ inclui ainda a seleção de dois leptões carregados, eletrões ou muões, com momento transverso superior a 25 GeV e um isolamento ao nível dos traços de 10%. A massa invariante reconstruida com os dois leptões deverá estar entre 83 e 99 GeV, compatível com a massa do bosão Z, e a energia transversa em falta no acontecimento deverá ser inferior a 60 GeV. Deverão existir pelo menos dois jatos no acontecimento, em que exatamente dois deles serão identificados pelo algoritmo de identificação de jatos-b (b-tagging) com uma eficiência de 80%. Após esta seleção, a análise é dividida em 30 regiões ortogonais consoante a multiplicidade de jatos no acontecimento, a eficiência com que estes são identificados e o momento transverso do bosão Z. Existem ainda outras regiões onde se prevê que os fundos sejam dominantes. Em particular, consideram-se as regiões em que apenas um jato é identificado como proveniente de um quark-b ou em que se requer que os dois leptões tenham sabor diferente, *i.e.*, um eletrão e um muão.

As distribuições da massa invariante reconstruida pelos dois jatos-b são fornecidas a um ajuste global (fit), juntamente com as incertezas sistemáticas teóricas e experimentais para obter as normalizações das amostras dos fundos simulados e encontrar o parâmetro de “força de sinal” (μ), que é um factor multiplicativo da normalização da secção eficaz do processo $ZH \to \ell\ell b\bar{b}$, de modo a que este parâmetro será 1 se o bosão de Higgs se comportar como o previsto pelo Modelo Padrão. O ajuste dos dados reais de 2012 tiveram uma flutuação estatistica negativa com respeito aos fundos e obtive o valor $\mu = -0.69 \pm 1.29$ para o valor do parâmetro da força de sinal. Embora negativo, este valor encontra-se ainda compatível com a previsão do Modelo Padrão dado que existe uma grande incerteza, dominada pela incerteza estatistica.

Das incertezas experimentais que afectam a análise ZH, a incerteza na escala de energia dos jatos é das mais relevantes. O mesmo se verifica para todas as pesquisas e medições em ATLAS que utilizam jatos para aceder à informação dos quarks e gluões existentes no estado final. Dediquei por isso parte do trabalho desta tese ao estudo destas incertezas. Em particular, contribuí para a validação da resposta em energia do calorímetro de ATLAS a hadrões isolados, a principal responsável pela incerteza na escala de energia dos jatos para momento transverso superior a 1.8 TeV. Fiz ainda a validação da utilização de uma nova calibração de jatos com o propósito de melhorar a sua resolução em energia quando aplicada a jatos-b.

Antes do LHC entrar em funcionamento, uma secção do calorímetro foi usada em testes com um feixe de piões com várias energias entre 20 GeV e 350 GeV, estabelecendo a resposta em energia do calorímetro a hadrões isolados. Ao começarem as colisões no LHC, foi possível obter também esta resposta com colisões protão-protão, para hadrões com energia no intervalo $0.4 < p < 10$ GeV, mas não foi possível estender esta região à zona entre 10 e 20 GeV, devido à falta de estatística acima de 10 GeV. Não existia nenhum trigger capaz de obter os traços alto momento isolados no calorímetro. Criei um novo trigger que reconstruia todos os traços com momento superior a 500 MeV e aplicava as mesmas condições de isolamento que a análise da resposta a hadrões isolados. Este novo trigger permitiu a validação da incerteza na escala de energia dos jatos, utilizando a resposta do calorímetro a hadrões isolados.

A análise ZH utilizou uma nova calibração de jatos, a calibração sequencial global (GSC), desenvolvida para reduzir a dependência na resposta em energia de jatos iniciados por quarks e por gluões. Complementarmente ao trabalho já referido, testei a aplicação desta calibração a jatos-b, quando resultantes de uma ressonância sem cor, como é o caso do decaimento do bosão de Higgs para $b\bar{b}$. Ficou demonstrado que esta calibração melhorou a resposta em energia destes jatos em $2-3\%$ e que a sua resolução em energia obteve melhorias de 20%, em concordância com o esperado para uma amostra de jatos inclusiva em sabor.

Foram avaliadas duas incertezas sistemáticas relacionadas com a utilização da calibração GSC em amostras simuladas específicas. Primeiro, a incerteza medida utilizando a calibração anterior foi comparada com as diferenças entre as respostas em energia dos jatos para vários processos de produção de jatos-b, sensível a alterações de modelos de chuveiro partónico e de hadronização, confirmando que a incerteza calculada anteriormente se mantinha quando GSC era aplicada. Adicionalmente, avaliei a utilização desta nova calibração na simulação rápida do detetor, que incluiu uma parametrização da resposta em energia do calorímetro. Esta simulação foi utilizada de forma geral nas amostras dos diversos processos de sinal e fundo da análise ZH. Descobri que no caso de se utilizar a simulação rápida do detetor, a incerteza estava subestimada por 0.5% e portanto tinha de ser corrigida. No entanto, a melhoria obtida na utilização de GSC foi suficiente para que esta incerteza extra não penalizasse a análise ZH.

Foi estudado ainda o impacto direto da nova calibração na resolução da distribuição da massa invariante reconstruida pelos dois jatos-b. A amostra de sinal que simulou o processo $ZH \to \ell\ell b\bar{b}$ foi testada com e sem a calibração GSC, tendo-se obtido melhorias na resolução da massa $m_{b\bar{b}}$ até 18% após a nova calibração. Os estudos detalhados do desempenho da calibração GSC e das incertezas associadas permitiu a sua utilização não só na análise do ZH como também nos outros canais de pesquisa do $H \to b\bar{b}$ WH (1 leptão) e o ZH (0-leptões), contribuindo para melhorar os resultados existentes.

Contents

Agradecimentos / Acknowledges

Não posso deixar de agradecer a todos os que me ajudaram no trabalho que resultou nesta tese.

Agradeço em primeiro lugar aos meus orientadores, Doutor Agostinho Gomes e Doutora Patricia Conde-Muiño pelo apoio constante durante o desenvolvimento do meu trabalho. Agradeço a paciência que me mostraram em particular nas inúmeras revisões do meu trabalho nesta fase final da escrita.

Agradeço à Fundação para a Ciência e Tecnologia pela bolsa de doutoramento através do programa IDPASC que me permitiu desenvolver este trabalho e ao LIP que me forneceu também bolsa de doutoramento nestes últimos tempos.

Um grande obrigado à Professora Amélia Maio, como responsável do grupo de ATLAS português que me abriu as portas à Física de Partículas, que embora reformada continua a exercer uma presença regular nos trabalhos de todo o grupo. Agradeço também a oportunidade concedida para lecionar na Faculdade de Ciências as suas aulas laboratoriais da disciplina Física para Biólogos.

Ao Doutor José Maneira e ao Doutor Ricardo Gonçalo pela também sua disponibilidade na revisão de algumas partes desta tese.

Agradeço também a todos os colegas que também participaram nas análises em que estive envolvido: o Alberto e a Rute, na construção conjunta do software de análise dos canais do ZH e o WH; e ao Emiliano e ao Doutor João Saraiva, no estudo da nova calibração de jatos, GSC.

Agradeço a todos os membros do grupo português de ATLAS pelo companheirismo sentido e pelas suas críticas construtivas que me proporcionaram ao meu trabalho.

I would like also to thank to all those people that have help me in some manner during the time I spent at CERN. It would be impossible to enumerate here all of them, since we are a collaboration of several thousand and only those with direct help will be referred.

For the time during the development of the HadCalib triggers and the E/p studies, I

would like to thank Doctor Michael Begel, Doctor Mario Campanelli, Doctor Tancredi Carli, Doctor Iacopo Vivarelli and Doctor Garabed Halladjian.

I would like to thank Doctor Andrew Mehta, Doctor Paolo Francavilla, Doctor Giacinto Piacquadio, Doctor Heather Gray, and again Doctor Ricardo Gonçalo that were the conveners of the ATLAS $H \to b\bar{b}$ sub-group, during the time of the Run-1 VH analysis. Additionally, I thank those more evolved with my work developed for GSC: Doctor David Mateos, Doctor Reina Camacho, Doctor Adrian Buzatu and Manuel Proissl.

I also would like to thank Doctor Martin Zur Nedden, Doctor Catrin Bernius and Doctor Mark Stockton, conveners these last few years of the Trigger Operations group, for the help provided during the remote shifts taken for LHC Run-2.

Quero agradecer também ao pessoal da secretaria que me ajudou com a parte burocrática e o pessoal da computação pela resolução eficiente dos problemas que lhes fui colocando ao longo dos anos.

Um agradecimento especial a todos os que durante estes anos partilharam ou ainda continuam na condição de estudantes no LIP pela vossa camaradagem. Espero que brevemente consigam completar os vossos projetos.

Quero agradecer a todos os meus amigos o apoio demonstrado ao longo dos anos, quer na comunidade de Taizé, quer na minha paróquia. Em particular os grupos de jovens e da catequese onde estive inserido e o grupo dos acólitos que liderei desde o liceu. Agradeço àquelas três pessoas que me acompanharam sempre durante o meu doutoramento e mesmo antes deste.

Agradeço à minha família todo o suporte que me deu durante toda a minha vida académica e, acima de tudo, durante a minha vida. A esta dedico todo o trabalho documentado presente na tese e também o que não foi cá incluído.

1. The Higgs boson in the Standard Model

> "For He gave me sound knowledge of what exists, that I might know the structure of the universe and the force of its elements."
>
> *(Book of Wisdom,7,17 (NABRE))*

1.1. Introduction

Far from giving a detailed description of the Standard Model (SM), this chapter will give but a small set of considerations needed to understand the main topic of this dissertation. A detailed description of electroweak theory can be in the historical references [1, 2, 3] and particle physics reviews [4, 5, 6]. Those provided in the bibliography are but a few of the available references.

1.2. Particles in the Standard Model

Figure 1.1 summarises the list of elementary particles in the Standard Model. The fermions are represented by a square box, where $\{d,u,s,c,b,t\}$ are the quarks *down*, *up*, *strange*, *charm*, *bottom* and *top*, respectively. The quarks and the gluons are the only particles that interact with the strong force, whose carriers are the eight gluons, g, that form the base of the quantum chromodynamics. In the last two columns, the leptons are presented. The charged leptons are the electron, the muon and the tau, represented by $\{e,\mu,\tau\}$, respectively. Together with the quarks, they interact through the electromagnetic force, whose carrier is the photon, γ. The neutral leptons are the neutrinos and are represented by the Greek letter ν with an index indicating the flavour of the neutrino. The neutrinos can only interact through

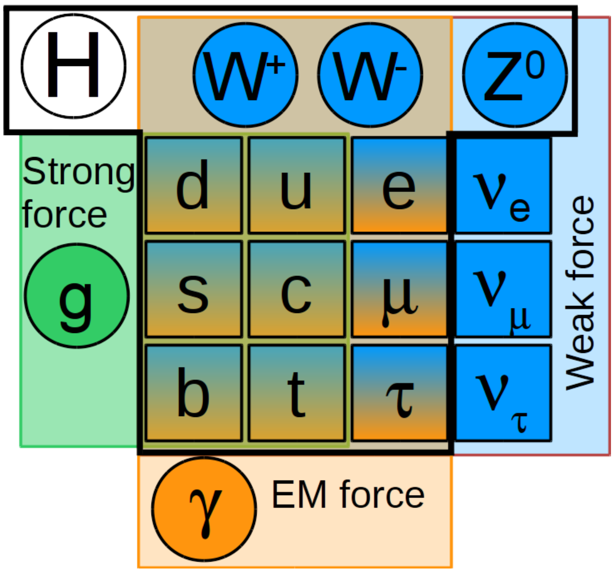


Figure 1.1.: The Standard Model elementary particles. Squares represent fermions and circles represent bosons. The rectangles enclose the particles subjected to the three forces contemplated in the Standard Model: the strong force in green, the weak force in blue and the electromagnetic force in yellow. The direct interaction of the Higgs boson with the particles depends on the mass of the particles. The neutrinos are considered massless in the Standard Model.

the weak force, whose carriers are the W and the Z bosons. Furthermore, all other particles interact with the weak bosons except the gluon. All elementary particles interact with the Higgs field, proportionally to their masses. Since the gluon and the photon are massless, they do not interact directly with the Higgs boson at *leading-order*, although they can interact at higher order. The neutrinos are also considered to be massless in the Standard Model and therefore do not interact with the Higgs field. [1]

Table 1.1 summarises the charge, mass and spin for each elementary particles. The particles of each type are presented in order of increasing mass. The uncertainties on the electron and muon masses are in the 8th and 9th significant figure, respectively, and for this reason, are not presented. Reference [4] also provides mass limits to neutrinos, photons and gluons but, in this thesis, their masses are considered to be zero for all practical matters. Relative to the electron charge, all quarks have a fractional charge. For each of the fermions, there is also a corresponding anti-fermion, with the same properties but opposite charge. In

---

[1] Neutrinos were found to have very small masses, resulting in a negligible interaction with the Higgs field.

	Name	Symbol	Type	Spin	Mass	Charge
Fermions	*up*	u	Quark	$1/2$	$2.2^{+0.6}_{-0.4}$ MeV	$+2/3$
	down	d	Quark	$1/2$	$4.7^{+0.5}_{-0.4}$ MeV	$-1/3$
	strange	s	Quark	$1/2$	96^{+8}_{-4} MeV	$-1/3$
	charm	c	Quark	$1/2$	1.27 ± 0.03 GeV	$+2/3$
	bottom	b	Quark	$1/2$	$4.18^{+0.04}_{-0.03}$ GeV	$-1/3$
	top	t	Quark	$1/2$	173.21 ± 0.87 GeV	$+2/3$
	e-neutrino	ν_e	Lepton	$1/2$	0	0
	μ-neutrino	ν_μ	Lepton	$1/2$	0	0
	τ-neutrino	ν_τ	Lepton	$1/2$	0	0
	electron	e	Lepton	$1/2$	0.511 MeV	-1
	muon	μ	Lepton	$1/2$	105.7 MeV	-1
	tau	τ	Lepton	$1/2$	1776.86 ± 0.12 MeV	-1
Bosons	*photon*	γ	Boson	1	0	0
	gluon	g	Boson	1	0	0
	W boson	W^-	Boson	1	80.385 ± 0.015 GeV	-1
		W^+	Boson	1	80.385 ± 0.015 GeV	$+1$
	Z boson	Z^0	Boson	1	91.1876 ± 0.0021 GeV	0
	Higgs boson	H	Boson	0	125.09 ± 0.24 GeV	0

Table 1.1.: List of properties of the particles in the Standard Model. All non-null mass values were taken from reference [4].

the charged leptons, the particles have a negative charge and are thus identified by the sign they have. The remaining anti-particles are identified with a bar on top of the symbol (*i.e.* $\bar{b}$ or $\bar{\nu}_e$). All fermions have semi-integer spin, $1/2$, and all bosons have an integer spin: 0 for the Higgs boson and 1 for the remaining.

Only quarks and gluons interact through the strong force as they are the only particles that have a colour quantum number different from zero. In nature, no coloured particle exists freely. If a collision creates either quarks or gluons, they will have to recombine so that any final particles will be colourless. This recombination process is referred as hadronization and can create mesons, formed by one quark and one anti-quark, or baryons, formed by three quarks or three anti-quarks. The only exception is the top quark, with decay width of 1.41 GeV, that decays to a W boson and a b-quark before hadronizing. Other top decays are suppressed but allowed. The name hadron is given to any particle composed of quarks. All hadrons are unstable and decay into other particles except the proton. The neutron is also considered to be stable since its decay to the proton, $n \rightarrow pe^{-}\bar{\nu}_e$, has a half-live of 880.2 ± 1.0 s [4]. Any neutron produced in the proton-proton collision, that does not interact with the detector, will decay outside of it. Very important for the subject of this thesis are the B-hadrons, which are hadrons that contain at least one *bottom* quark.

1.3. Description of electroweak theory

There are four fundamental interactions in nature: electromagnetic, weak, strong and gravitation. Except for the latter, they are mathematically well described using quantum field theories, which combine both quantum mechanics and special relativity. Glashow, Salam and Weinberg introduced the unification of the electromagnetic and weak interaction, into the electroweak theory, which could be described by the $SU(2)_L \times U(1)_Y$ symmetry group [7, 8]. The electroweak theory established the foundation for the strong quantum field theory using the $SU(3)_c$ symmetry group, which is known separately as quantum chromodynamics (QCD). The Standard Model of Elementary Particles then combines the three symmetry groups and the Lagrangian of equation 1.1 that is used to describe the interactions between all elementary particles.

In this section, the indices μ and ν vary from 0 to 3, indicating the time-like and the three space-like coordinates of the Lorentz quadri-vectors. The combination of two indices result in a Lorentz tensor that corresponds to a 4×4 matrix.

$$\begin{aligned}\mathscr{L} = &-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\vec{W}_{\mu\nu}\vec{W}^{\mu\nu} - \frac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a} \\ &+ \sum_{\alpha=1}^{3}\{\overline{Q_\alpha}\not{D}Q_\alpha + \overline{u_{\alpha,R}}\not{D}u_{\alpha,R} + \overline{d_{\alpha,R}}\not{D}d_{\alpha,R} + \overline{L_\alpha}\not{D}L_\alpha + \overline{e_{\alpha,R}}\not{D}e_{\alpha,R} + h.c.\} \qquad (1.1)\end{aligned}$$

In this Lagrangian, the triplet $\vec{W}_{\mu\nu}$ and the singlet $B_{\mu\nu}$ are the electroweak field strength tensors obtained from the massless gauge fields of the $SU(2)_L$ and $U(1)_Y$ groups, $\vec{W}_\mu$ and B_μ

respectively, defined as:

$$\begin{aligned}\vec{W}_{\mu\nu} =& \partial_\mu \vec{W}_\nu - \partial_\nu \vec{W}_\mu + g_W \vec{W}_\mu \times \vec{W}_\nu \\ B_{\mu\nu} =& \partial_\mu B_\nu - \partial_\nu B_\mu\end{aligned} \tag{1.2}$$

where g_W is the electroweak coupling. The third term in equation 1.1 represents the strong field tensors and describes the interactions between the eight gauge bosons of the $SU(3)$ group, the gluons, represented by G^a in equation 1.3.

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \quad \text{with } a = 1, \ldots, 8 \tag{1.3}$$

where g_S is the strong interaction coupling and the f^{abc} are the 3×3 matrices equivalent to the Pauli matrices, known as the Gell-Mann "λ-matrices".

The second line of the Lagrangian describes the interactions between the gauge fields and the three fermions families ($\alpha = 1, 2, 3$), where the covariant derivative to obtain an invariant Lagrangian is given by:

$$\not{D}\Psi = \gamma^\mu D_\mu \Psi = \gamma^\mu \left(\partial_\mu - ig\vec{T} \cdot \vec{W}_\mu - \frac{1}{2} ig' Y B_\mu - \frac{i}{2} g_s \lambda_a G^a \right) \Psi \tag{1.4}$$

The three components of weak isospin, $\vec{T}$, is given by the Pauli matrices as $\vec{T} = \frac{\vec{\sigma}}{2}$ and Y is the hypercharge obtained from a linear combination of the particle charge and the third component of the weak isospin. Y and $\vec{T}$ are the generators of the electroweak field. The λ_a represent the eight generators of the QCD field, associated to the eight gluons.

In equation 1.1, the u, d and e symbols represent the right-handed weak isosinglets *up* quark, *down* quark and electron, respectively and the Q and L symbols represent the left-handed weak isodoublets as:

$$L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, Q = \begin{pmatrix} u \\ d \end{pmatrix}_L, \tag{1.5}$$

Finally, the Standard Model does not predict right handed terms to neutrinos, and these are considered to be massless in this thesis.

1.4. The Higgs mechanism

The Standard Model Lagrangian in equation 1.1 predicts a zero mass for all elementary particles, ensuring an invariant Lagrangian under local gauge transformations. Including explicit mass terms for these particles breaks the symmetry and the Lagrangian is no longer invariant. However, experimentally it is known that many elementary particles have mass and, in particular, the W and Z bosons and the top quark are extremely massive.

The Higgs mechanism [2] [9, 10, 11] uses spontaneous breaking of the $SU(2)_L \times U(1)_Y$ symmetry to provide mass to the W and Z bosons, maintaining the photon massless. It predicts the existence of a new scalar field, ϕ. This scalar field has a potential form as shown in equation 1.6

$$V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2 \tag{1.6}$$

with a negative value for the μ^2 parameter.

The Lagrangian for the interaction between the gauge fields and this new field is shown in equation 1.7:

$$\mathscr{L} = \left(D_\mu\phi\right)^\dagger (D^\mu\phi) - V(\phi) \tag{1.7}$$

where the covariant derivative, from equation 1.4, does not include the term related to the quantum chromodynamics, since the Higgs boson does not have strong interaction.

If there would be only one massive gauge boson, the simplest scalar field that can be considered is a complex scalar, $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$. It is possible to represent the potential from equation 1.6 in a two-dimensional graphic as the one in figure 1.2. The ball placed at the origin of the referential represents the universe where the interactions obtained by the Lagrangian preserve the symmetries and all particles are massless. As the system cools down, the universe will choose a ground state from the circle of minima that satisfies the equation $\phi_1^2 + \phi_2^2 = \mu^2/\lambda^2$. By changing the coordinate system to be centred in the considered vacuum minimum, the system breaks the symmetry spontaneously. From the interaction of the scalar field with the single massive gauge boson in this hypothetical universe, the gauge boson gains mass, and an additional degree of freedom appears which is associated with the mass term of the scalar field.

Since there are three massive gauge bosons, the minimum degrees of freedom required

---

[2] The Higgs mechanism should, in fact, be denominated as the Englert-Brout-Higgs-Guralnik-Hagen-Kibble mechanism to account all the developers that, working in three independent groups, arrived to the same conclusion.

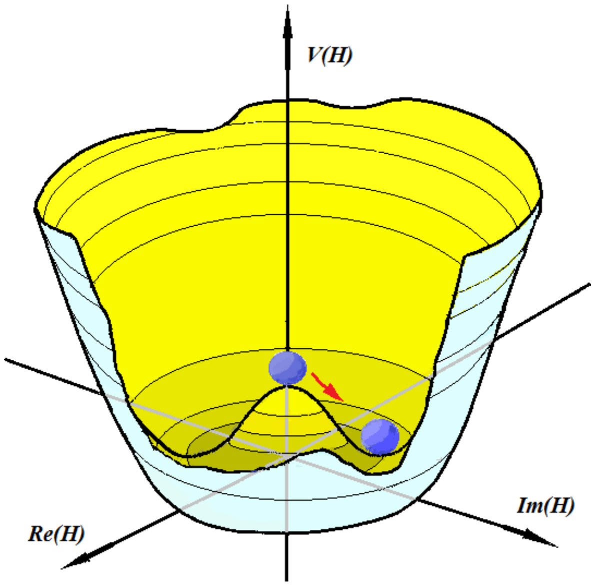


Figure 1.2.: Illustration of the spontaneous symmetry breaking in two dimensions. The particle at the local maximum of energy has symmetry and no mass. A perturbation on the system sends the particle to a state of minimum energy, and it acquires mass from the symmetry breaking [12].

for the scalar field is four, and it is usually represented by a complex doublet, shown in equation 1.8.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{1.8}$$

The right term in this equation, shows already the coordinate change to a specific (simplest) vacuum, where v is the vacuum expected value and h is the real scalar field. With the complex doublet, one obtains the mass terms for the three massive gauge bosons and again a single mass term for the scalar field associated with a new particle, the Higgs boson. The mass of this new boson is given by:

$$m_H = \sqrt{-2\mu^2} \tag{1.9}$$

The interaction term between the gauge bosons and this scalar field appears in equation 1.7 in the covariant derivative. Applying the spontaneous symmetry breaking to the Higgs Lagrangian part in equation 1.7, the mass terms for the $W^{\pm}$ boson result from the linear combination of the gauge fields W^1_μ and W^2_μ and the mass terms for the Z boson and the photon from the linear combination between the W^3_μ and B_μ.

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp i\,W_\mu^2\right) \tag{1.10}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \tag{1.11}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W \tag{1.12}$$

Equations 1.10 to 1.12 summarise the relations between gauge bosons and physical particles, with the weak mixing angle, θ_W, relating the electroweak couplings in equation 1.13.

$$g_e = g_W \sin\theta_W = g_Z \sin\theta_W \cos\theta_W \tag{1.13}$$

The Lagrangian in equation 1.7 predicts also triple and quartic couplings between the weak bosons and also triple and quartic self-interacting couplings of the Higgs boson.

A final interaction is introduced between the new scalar field and the fermions, which is known as the Yukawa interaction and it is represented by the Lagrangian in equation 1.14.

$$\mathscr{L} = \sum_{\alpha=1}^{3}\{-\lambda_{d,\alpha}\bar{Q}\phi d_\alpha - \lambda_{u,\alpha}\bar{Q}\widetilde{\phi} u_\alpha - \lambda_{e,\alpha}\bar{L}\phi e_\alpha + h.c.\} \tag{1.14}$$

The scalar field links the left-hand-side with the right-hand-side fermions providing the mass term and the interactions between each fermion and the Higgs boson, which obeys to equation 1.15.

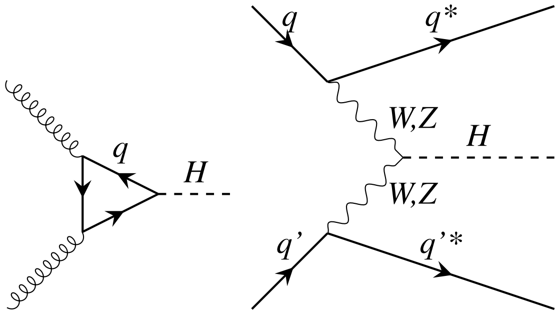
$$\lambda_f = \sqrt{2}\frac{m_f}{v} \tag{1.15}$$

Therefore, the Higgs boson probability of interaction with the fermion is proportional to the mass of the fermion (m_f), with the vacuum expectation value: $v = 246$ GeV. This Lagrangian is again considered for the three families of fermions: $\alpha = 1,2,3$.

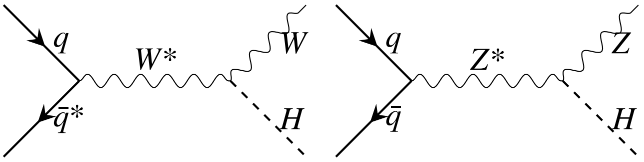
1.5. Higgs boson phenomenology

1.5.1. Higgs boson production at LHC

The Large Hadron Collider (LHC) at CERN collided protons with a centre of mass energy $\sqrt{s} = 8$ TeV during 2012 data taking. Protons have internal structure, as they are composed of three valence quarks, *up-up-down*, and a sea of gluons, that mediate their interactions. Additionally, a gluon can give origin to two or three other gluons and can also split into a quark-anti-quark pair with the same flavour. Parton distribution functions, $\mathrm{PDF(x,Q^2)}$, represent the probability of finding a certain quark or gluon (parton) inside the proton with a specific momentum fraction of the proton, x, when probed with an energy squared of Q^2. Cross section is the probability that two particles will collide and react in a certain way, so in a proton-proton collision it is the probability of interaction between a parton from each proton in a process to produce some possible final state. So, the cross section is in turn proportional to the probability that the process between two partons occurs, which can be calculated by perturbation theory, and to the PDF of each parton. When two protons collide, there can be one or more interactions between the two gluons, between a gluon and a quark or between two quarks. These interactions can produce any number of particles of any type.



(a) gluon fusion (b) vector boson fusion



(c) W boson association (d) Z boson association

Figure 1.3.: Feynman diagrams for the main Higgs boson production processes at the LHC.

Figure 1.3 shows the leading Feynman diagrams of the four Higgs boson production processes with the largest cross sections for collisions at a centre of mass energy of $\sqrt{s} = 8$ TeV and, in Figure 1.4, the production cross section corresponding to these processes is shown as a function of the Higgs boson mass. The most important production mode is

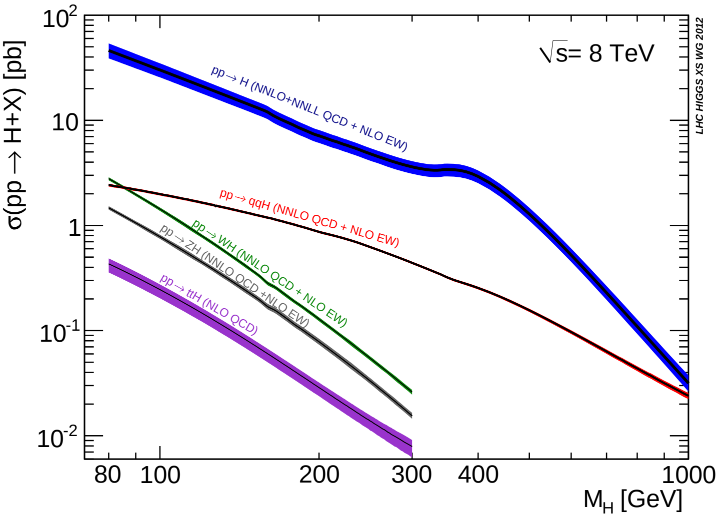
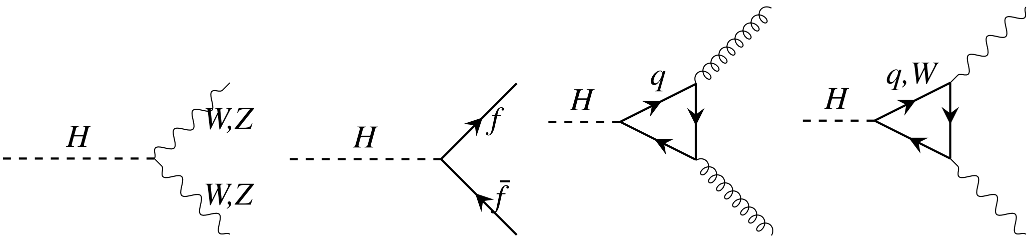


Figure 1.4.: Higgs production cross section and respective uncertainties as a function of the Higgs boson mass [13].

the gluon fusion process shown by the blue curve. The Higgs can only interact indirectly with gluons, through a quark loop, as shown in Figure 1.3 (a). Although the loop reduces significantly the process $gg \to H$, the gluons are the most abundant particles in the proton at high energies. Since the probability of the Higgs boson interaction with the quark in the loop scales with the square of the mass of that quark, other quark contributions than the top quark are negligible ($< 0.1\,\%$). The graph identified as $pp \to H$ in figure 1.4 also shows an increase in cross section if the Higgs boson had twice the top quark mass, in which the top quarks in the loop would become real. The second leading production process, corresponding in first order to the Feynman diagram shown in Figure 1.3 (b), is vector boson fusion. Two quarks radiate vector bosons W or Z, which interact, giving origin to a Higgs boson. This production mechanism, identified as $pp \to qqH$ in figure 1.4 has about one order of magnitude lower cross section, except at very large Higgs boson mass. The next leading process is the associated production of the Higgs boson with a W or a Z boson. Their dominant Feynman diagrams are shown in figures 1.3 (c) and 1.3 (d), respectively. Figure 1.4 shows a faster decrease in cross section for these two production processes as the mass of the Higgs boson increases than for vector boson fusion. The cross section production for the WH process is roughly twice that for ZH. These two production processes are commonly referred to as VH production. Figure 1.4 also shows the Higgs boson cross section production associated with a pair of top quarks, with about a factor of five below the value of the cross section production of ZH.

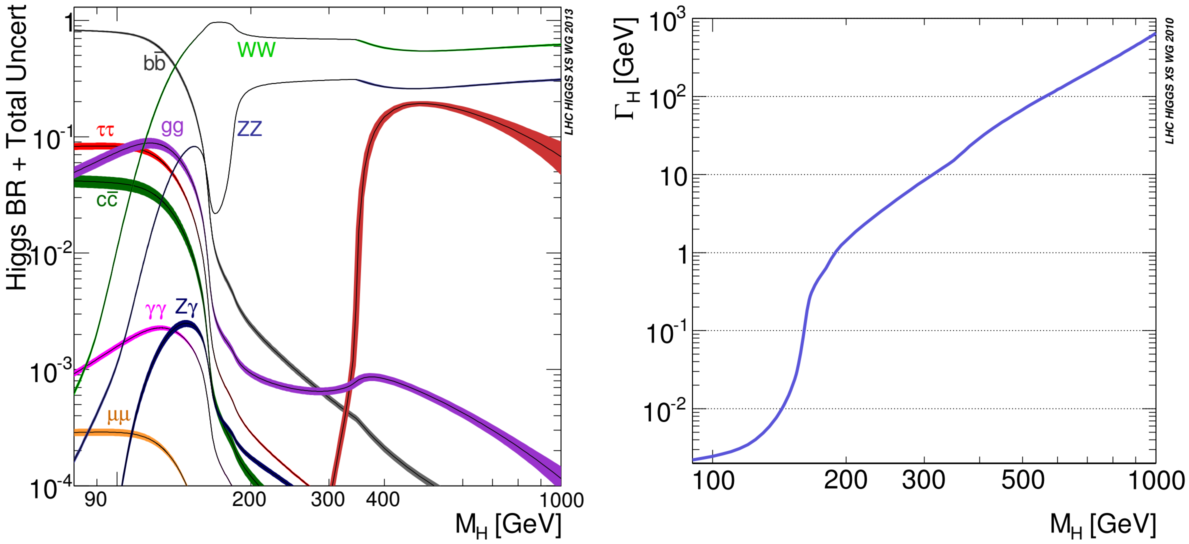
1.5.2. Higgs boson decay modes



(a) $H \to VV$ (b) $H \to f\bar{f}$ (c) $H \to gg$ (d) $H \to \gamma\gamma$

Figure 1.5.: Feynman diagrams for the Higgs boson decays.

Figure 1.5 shows the four possible Higgs boson decays to vector bosons or fermions. The decay to gluons mirrors the gluon fusion production and, in the decay to photons, the loop may include W bosons or quarks.



(a) (b)

Figure 1.6.: Branching ratios with total uncertainty (a) and total decay width, Γ, of the Higgs boson as a function of its mass (b) [13].

Figure 1.6 (a) shows the branching ratios, as a function of the Higgs boson mass. The branching ratio is the ratio between the individual decay width, Γ_X, and the total decay width of the Higgs boson. The total uncertainty is also shown by the coloured bands for each decay. Figure 1.6 (b) shows the total decay width of the Higgs boson as a function of the Higgs mass. As the mass of the Higgs boson increases, decays to particles with larger masses become possible. At very low masses, below 100 GeV, the Higgs decays in the majority of the cases to a b-quark pair, with contributions from $\tau\tau$, gg and $c\bar{c}$ in the 5-10 % range. Remaining contributions account for less than 1 %. In the mass range from 100 to 180 GeV, there are several alternatives, as the Higgs boson mass starts to allow the production of W and Z boson pairs. In the high-mass region, between 180 and 350 GeV, the decay to WW and ZZ dominates the possibilities, with other decay alternatives having a branching ratio below 1 %. At 350 GeV, the Higgs boson can also decay to $t\bar{t}$ and the decay to gluon-gluon also rises. The branching ratio for the $t\bar{t}$ decay decreases afterwards because the partial width from the vector boson decays increase faster for higher masses.

Since partial decay widths of the vector bosons rise faster than those for fermions as the Higgs boson mass increases, the branching ratio to the latter decreases again.

1.6. Higgs discovery and measurements

Although the Higgs boson mass was not known a priori in the Standard Model, there were some theoretical and experimental constraints on the Higgs boson mass before the LHC. From the experimental constraints such as vector bosons and top quark masses, and using the GFITTER tool, a Higgs mass of $m_H = (93^{+25}_{-21})$ GeV [14] was obtained, as shown in figure 1.7. The exclusion limit from searches for the Higgs boson at LEP was 114.4 GeV [15]. Additionally, Tevatron excluded the Higgs boson mass region between 158 and 175 GeV with a confidence level of 95 % [16]. The theoretical constraints include the requirement that the $VV \rightarrow VV$ scattering amplitude should conserve unitary, the quartic coupling of the Higgs boson and the stability of the vacuum [5].

The LHC expected to either find the Higgs boson or to set exclusion limits on its mass up to 1 TeV, using different channels. A channel is the combination of one production process and one specific decay mode characterized by certain experimental signatures. The most relevant search channels combined the gluon-fusion production with the Higgs decays WW, ZZ, with leptonic decays of the vector bosons, which offer clear signatures. Also with

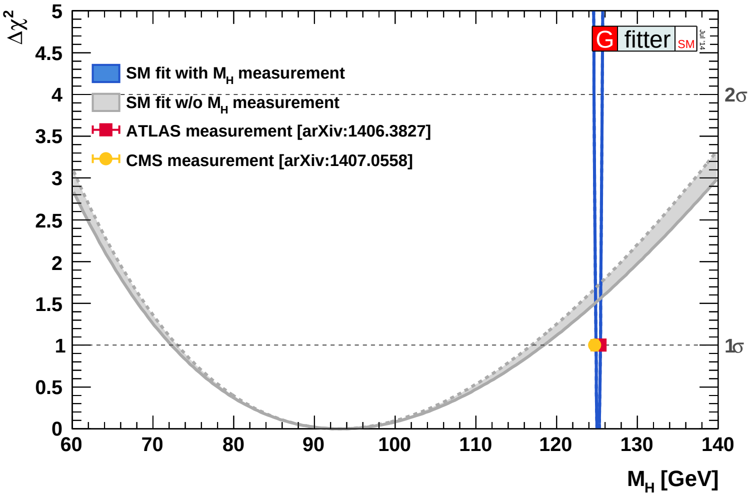


Figure 1.7.: Standard Model best fit for the Higgs boson mass with and without considering the measurements of ATLAS (pink marker) and CMS (yellow marker) experiments [14].

clear signature is the $H \rightarrow \gamma\gamma$ decay, important for low Higgs boson masses. The LHC was able to set exclusion limits to nearly all the Higgs mass range and, in 2012, the ATLAS and CMS [17, 18] experiments finally announced the discovery of the Higgs boson with a mass near 125 GeV. Figure 1.7 shows also the two mass values obtained for these two experiments and the best fit including the most recent measurements. The final combination for the Higgs mass from the two experiments, at the end of LHC Run-1 data taking, was $125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst})$ GeV [19].

The observation of Higgs couplings to fermions, in the $H \rightarrow \tau\bar{\tau}$ channel, was recently done combining the results from ATLAS and CMS experiments. However, there is no direct observation of the coupling of the Higgs boson to quarks, although the $H \rightarrow b\bar{b}$ decay has the leading branching ratio for a 125 GeV Higgs boson, $(57.5 \pm 1.9)\,\%$. This is due to the very large background of b-quark pairs produced in the proton-proton collision at the LHC. Other missing measurements are the Higgs coupling to the first and second family of fermions and the Higgs self-coupling. The Higgs decay to muons and its self-coupling are some of the objectives of the High Luminosity LHC (HL-LHC) [20]. Since the first family of fermions have very small masses, their interaction with the Higgs boson should be very faint, and such measurements are impossible in a hadron collider, due to the significant background.

The compatibility of the experimental results with the theoretical results is measured by the signal strength parameter, which is defined by the ratio between what is observed and what is predicted by the Standard Model, combining the cross section production and the branching ratios of the Higgs boson of all possibilities as described by equation 1.16.

$$\mu = \frac{\sigma_{exp} \times BR_{exp}}{\sigma_{theo} \times BR_{theo}} \tag{1.16}$$

The product of the cross section times the branching ratio of a certain search channel, with a production i and decay f of the Higgs boson ($i \rightarrow$H$\rightarrow f$), can also be parametrized by a set of coupling modifiers, the $\vec{\mathrm{k}}$-factors, that measure possible deviations on the fermionic and bosonic couplings with the Higgs boson expected by the Standard Model [21, 22], as shown in equation 1.17:

$$\sigma_i \times \mathrm{BR}^f = \frac{\sigma_i(\vec{\mathrm{k}}) \times \Gamma^f(\vec{\mathrm{k}})}{\Gamma_\mathrm{H}} \tag{1.17}$$

where Γ_H is the total width of the Higgs boson and Γ^f is the partial width of each decay channel. The coupling modifiers are obtained from equation 1.18 for a given production process i or decay mode f.

$$\mathrm{k}_i^2 = \frac{\sigma_i}{\sigma_i^{SM}} \qquad \text{or} \qquad \mathrm{k}_f^2 = \frac{\Gamma_f}{\Gamma_f^{SM}} \tag{1.18}$$

The best available higher-order QCD and EW corrections are used to take into account possible correlations between different production (decay) mechanism that contribute to the cross-section (partial decay width). For example, in the decay of the Higgs boson to a pair of photons, one needs to consider in interference between the top quark and the W boson in the loop [22]. The data collected by the LHC is then fitted to the Standard Model prediction and the coupling modifiers are extracted. This is often simplified by considering all fermionic ($\mathrm{k_F}$) from bosonic ($\mathrm{k_V}$) coupling modifiers to be the same:

$$\mathrm{k_V} = \mathrm{k}_Z = \mathrm{k}_W = \mathrm{k}_\gamma = \mathrm{k}_g \qquad \text{and} \qquad \mathrm{k_F} = \mathrm{k}_t = \mathrm{k}_b = \ldots = \mathrm{k}_\tau = \mathrm{k}_\mu = \mathrm{k}_e \tag{1.19}$$

Figure 1.8 shows the $\mathrm{k_F}$-$\mathrm{k_V}$ fit, combining ATLAS and CMS data, obtained for the individual Higgs boson decay modes (coloured crosses) and the best combined fit (gray cross). The 68 % confidence level is shown for each marker by the filled lines and the 95 % confidence level is shown for the combined fit. The Standard Model, shown by the star marker, is inside

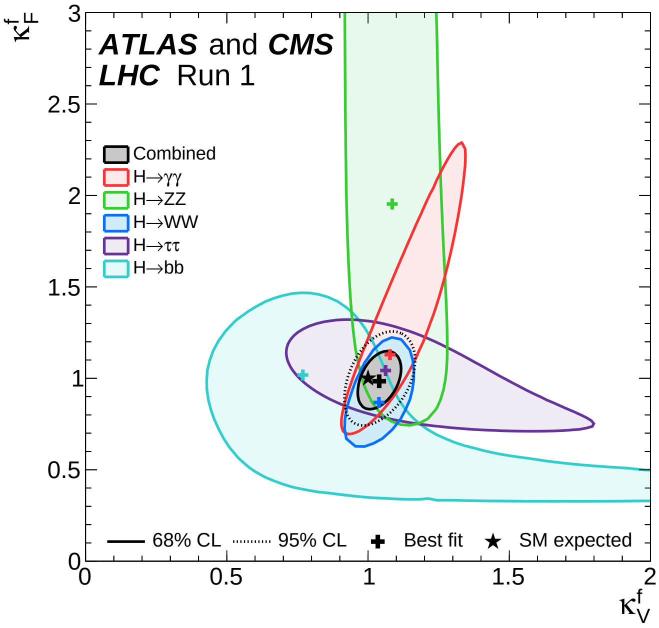


Figure 1.8.: Measured k-factors relation between fermions (k_F^f) and bosons (k_V^f), for the different decay modes of the Higgs boson, shown by different colours. The region delimited by the black line indicates the combination of all channels. The prediction for the Standard Model Higgs boson, shown by the *star* marker is inside the 68 % confidence level from the best fit of the ATLAS and CMS combined measurement, shown by the black *cross* marker [22].

the 1 standard deviation region from all fits.

According to the Standard Model prediction, the Higgs boson should have a spin-parity $J^P = 0^+$. Both ATLAS and CMS [23, 24] have tested other possible spin-parity combinations against this prediction and have rejected all the considered alternatives with a confidence level larger than 95 %, using the Higgs bosonic decays and LHC collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. Most of these measurements are based on angular correlations of the decay products that are then combined in multivariate analysis.

1.7. The $H \to b\bar{b}$ decay search

The largest expected branching ratio of the Higgs boson corresponds to the decay to a b-quark pair, with 57.5 %. The $H \to b\bar{b}$ channel is the best one to measure the Higgs boson

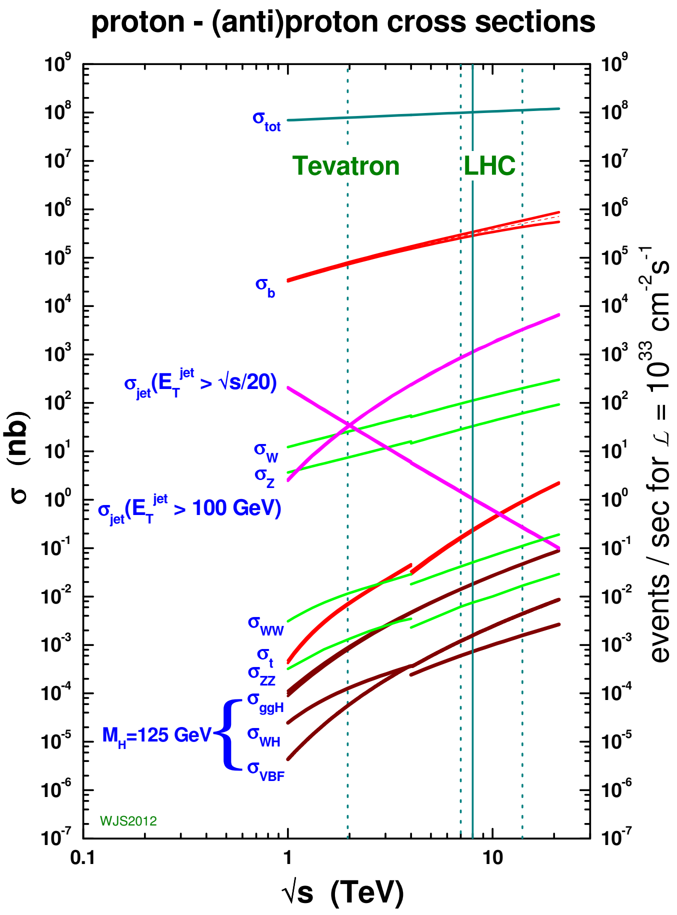


Figure 1.9.: Production cross section for a given channel and correspondent event rate estimate production for the nominal LHC Luminosity [25].

direct coupling to quarks and it also constrains experimentally the decay width of the Higgs boson that can be used to infer if this is indeed the Higgs boson predicted by the Standard Model.

There is, however, a significant caveat in the $H \to b\bar{b}$ searches: the high background of this decay channel. For every $H \to b\bar{b}$ decay, LHC produces additionally about 10^7 $b\bar{b}$ pairs, as shown in figure 1.9 and also an order of magnitude more of lighter quark pairs. It is, therefore, impossible to detect the Higgs boson in the $b\bar{b}$ decay channel in the gluon-fusion dominant process. The Higgs boson production process with the second largest cross section, vector boson fusion, only has quarks in the final state with no clear signature. In the associated production with a vector boson, leptons from the W/Z decays can provide a clean trigger signature. The associated production with a b-quark pair has the same problem again as the gluon-fusion, although with a smaller background. The associated production with a top pair can use the top decay $t \to Wb$ with large branching ratio and choose events in which one or both W bosons decays to a charged lepton and a neutrino to trigger the event. This production process, however, has a cross section of about an order of magnitude lower than the associated production with a vector boson at the centre of mass energy $\sqrt{s} = 8$ TeV. Therefore, the most promising channel to study this decay is the Higgs boson associated production with a W or a Z boson. This thesis focus the latter, in which the Z boson decays to an electron or a muon pair.

2. The ATLAS Experiment

> "(A light-year) it's something they invented, since it would take all the paper in the world to write such large distances between stars."
>
> *(Mário Sargedas, grandfather)*

2.1. CERN and the LHC

Founded in 1954, CERN has been since then responsible for a remarkable progress in physics and technology [26]. CERN's first accelerator, the Synchro-Cyclotron (SC), reached an energy of 600 MeV and the first experiments in nuclear and particle physics at CERN used its beam.

In 1964, the SC accelerator was assigned to the nuclear physics area and since 1967 up to its end, supplied beam to the ISOLDE facility for the production of short-lived isotopes.

Particle physics abandoned the SC as a new generation of accelerators was born. The Proton Synchrotron (PS) accelerates protons since 1959, reaching a beam energy of 28 GeV. After the seventies, CERN started to construct larger accelerators, so the PS was no longer the primary place to do experiments. The PS became the supplier of particles for the new machines, something that it still does in our days.

Scientists realised that the available collision energy could almost be doubled if, instead of colliding a beam into a target, two beams were made to collide with each other. Thus, the Intersecting Storage Rings (ISR) collider began its operation in 1971 with the objective of using two beams from PS, achieving the first collisions between two proton beams. The ISR had a diameter of 300 meters. From then on, new collidors with increasing diameter rings were built, increasing the collision energy.

In 1976, the Super Proton Synchrotron (SPS) was finished, with a perimeter of 7 kilometres and with two main experimental sites: Meyrin and Prevessin. Using the protons provided by the PS, this accelerator increased the beam energy up to 300 GeV (working nowadays with energies of 450 GeV). With the SPS, it was possible to measure direct CP violation in the kaon-sector, to probe for the first time the internal structure of the proton and to look to conditions similar to the seconds following the Big-Bang. In 1979, SPS was converted into a proton-anti-proton collider ($SP\bar{P}S$), to enable the same electric field to accelerate both beams. The primary goal was to find the Z and W bosons responsible for the weak interaction. Their masses were predicted, and collisions started. Finally, they were discovered in 1983, and for this accomplishment, the Nobel Prize in Physics was awarded the following year to Carlo Rubbia and Simon van der Meer.

A still more daring project was planned: to build an electron-positron collider with a perimeter of 27 kilometres. The construction started in 1985 and took four years to complete. The Large Electron-Positron collider (LEP) was in operation during seven years at 100 GeV, having produced about 17 million Z bosons. To bend the beam trajectories, LEP had 5176 magnets and 218 accelerating cavities. Increasing the number of cavities to 288 and using superconductor technology it doubled the energy, allowing the production of W boson pairs. This accelerator was closed at the end of 2000 so that the LHC installation could start.

Meanwhile, the Tevatron collider at FermiLab was the direct competition of CERN, and it produced and detected the top quark, in 1995, completing the third generation of quarks [27, 28].

Finally, the current accelerator in use is the Large Hadron Collider (LHC). Reusing the tunnel located 100 meters deep constructed for LEP, the LHC accelerates protons very close to the speed of light, producing collisions with a design centre of mass energy of $\sqrt{s} = 14$ TeV.

The LHC operation started in September 2008, only to stop one week later due to a liquid helium leak which caused one of the magnets to increase its temperature over 100 °C. The experiments used the following year downtime to commission the detectors using data from cosmic rays. After this stop, in November 24^{th} 2009, the LHC operation was finally restarted, having registered the first collisions at $\sqrt{s} = 2.36$ TeV. A milestone was achieved in March 19^{th} 2010 when the beam reached a record energy of 3.5 TeV. At the end of March, the two beams were colliding with a centre of mass energy of $\sqrt{s} = 7$ TeV which was used until the end of 2011 to search for new physics and to perform many measurements. A total of $5.46 - 6.10$ fb^{-1} of proton-proton collisions were delivered to the experiments during this period. In 2012, the accelerator increased the centre of mass energy to $\sqrt{s} = 8$ TeV, which led to the discovery of the Higgs boson. During this year, LHC delivered to ATLAS proton-proton collisions with an integrated luminosity of 22.8 fb^{-1}. In 2015, the LHC increased its centre of mass energy to $\sqrt{s} = 13$ TeV, after about a 2-year shut-down period. The total integrated luminosity delivered to the experiments in that year was about 4 fb^{-1}.

2.1.1. Main experiments

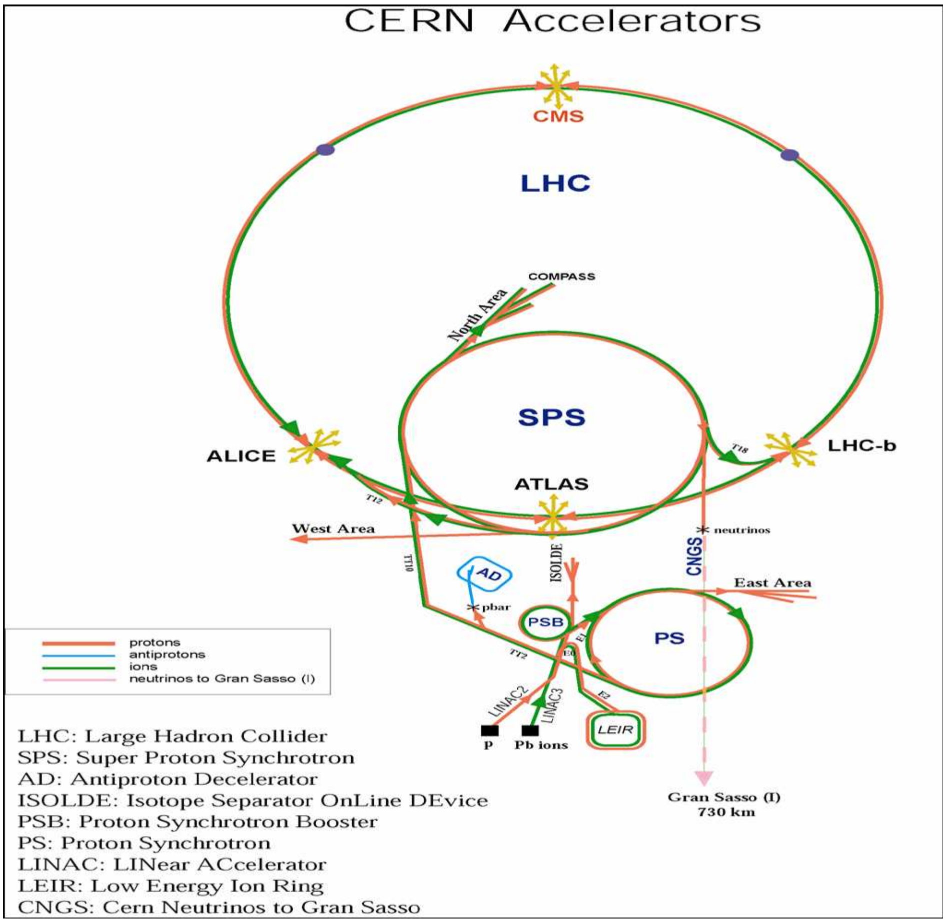


Figure 2.1.: The CERN accelerator complex and locations of the main experiments [29].

The previous CERN accelerators are used now as pre-accelerators as shown in figure 2.1. There are several experiments installed at the points where the LHC beams intersect, and the collision symbols represent their positions. The most important are ATLAS, CMS, ALICE, and LHCb. ATLAS and CMS are both general purpose experiments, and ALICE and LHCb are dedicated ones. The purpose of ALICE is to study the collisions between heavy ions, with the primary objective of studying the quark-gluon plasma. ALICE, ATLAS and CMS try to have a full coverage around the interaction point. LHCb has only forward coverage, as it studies CP violation in the B-meson sector and the physics of the bottom quark, which events are produced close to the beam direction.

2.2. The ATLAS detector

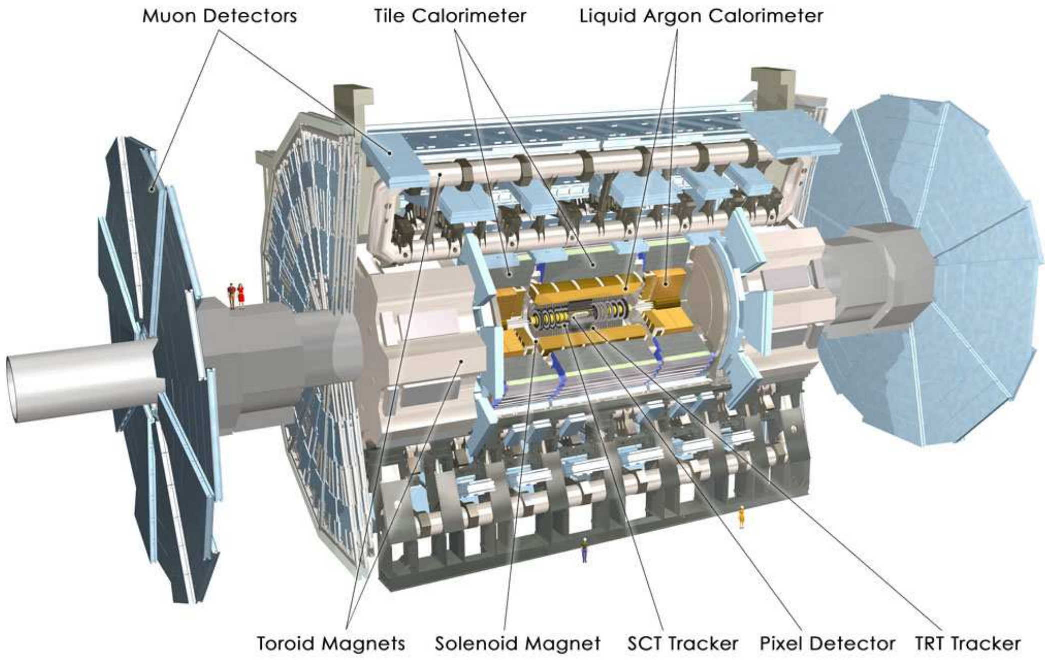


Figure 2.2.: The ATLAS detector [29].

The ATLAS (A Toroidal LHC ApparatuS) detector, shown in figure 2.2, is a general purpose detector that studies the fundamental interactions of the elementary particles. An extensive description of the ATLAS detector and its performance can be found in references [30, 31].

ATLAS is one of the experiments that have discovered the Higgs boson, validating the mechanism responsible for the origin of the mass of the elementary particles according to the Standard Model, discussed in chapter 1. ATLAS has an extensive physics program. In addition to revalidating the measurements of the Standard Model and continuing the study of this new particle, it also tries to find evidence of the dark matter, of supersymmetric particles, and other physics beyond the Standard Model.

ATLAS is 44 metres long and 25 m in diameter, weighing about $7,000$ tonnes. Figure 2.2 also displays the ATLAS sub-detectors, placed in several layers around the interaction point. It is composed of an inner detector tracking system, electromagnetic and hadronic calorimeters, a muon spectrometer, and a solenoidal and toroidal magnetic systems.

ATLAS uses the following system of coordinates: the origin is at the centre of the detector, the Y axis is orientated towards the surface, X towards the centre of the LHC accelerator and Z along the beam direction keeping the right-hand rule to complete the axis system. The spherical coordinate system (r, θ, ϕ) is frequently used, where ϕ is the azimuthal angle with the X-axis and θ is the polar angle, with the Z-axis. In experimental particle physics, the rapidity (y) is used to describe the behaviour of particles that come from collisions. It is defined by:

$$y = -\frac{1}{2}\log\left(\frac{E+p_L}{E-p_L}\right) \tag{2.1}$$

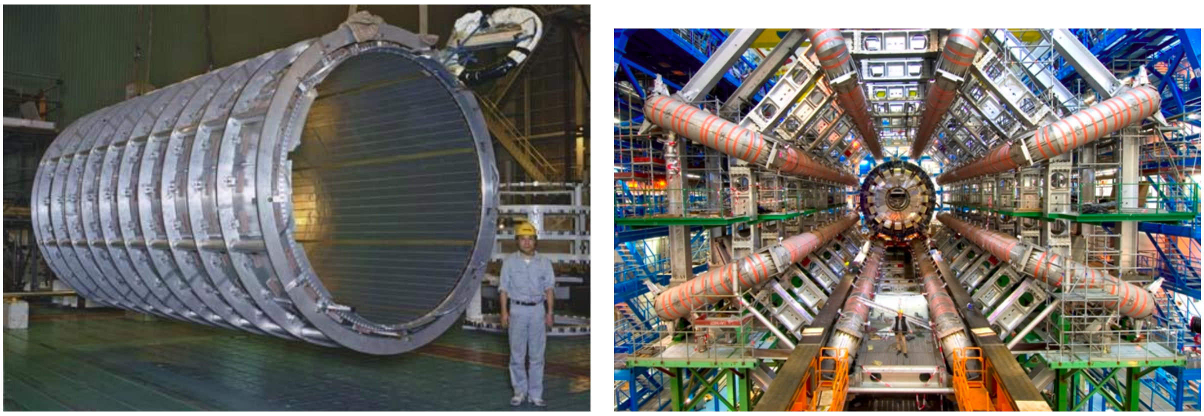
where E is the particle energy and p_L is the particle momentum in the direction of the beam. For relativistic particles, as those produced in the LHC, the mass is negligible relatively to the energy and the rapidity is approximated by the pseudorapidity, η, which depends only on the polar angle θ as shown in equation 2.2.

$$\eta = -\log\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.2}$$

2.2.1. Magnetic system

The ATLAS magnetic system is needed to bend the trajectory of charged particles, allowing the trackers to measure their momenta. A 5.8-metre long superconducting solenoid magnet (see figure 2.3 (a)) is located between the inner detector and the barrel electromagnetic calorimeter. It was designed to apply a strong magnetic field of 2 T near the interaction point, decreasing to a value of 0.5 T at the inner detector edge.

The magnetic field system of ATLAS also includes the superconducting air-core toroid magnet system, with a barrel toroid, presented in figure 2.3 (b), and two endcap toroids at each side of the detector. This system provides a magnetic field coverage of $|\eta| < 2.7$. Each of the eight coils of the barrel extends radially from 4.7 to 10.05 m and has an axial length of 25.3 m. Each endcap coil has an axial length of 5 m and extends radially from 82.5 cm to 5.35 m. The main objective of this magnetic system is to improve the resolution of the muon momentum measurement, particularly for high-p_T muons. The magnetic field increases its



(a) (b)

Figure 2.3.: Solenoid (a) and toroid barrel (b) of the ATLAS magnetic system [32].

power with the pseudorapidity, up to 4.1 T. The main advantage of this type of magnetic system is that, since it has an air core, the muon momentum measurement is less affected by multiple scattering, when compared with a similar system based on iron core like the CMS one.

The two magnetic field systems bend the trajectory of particles in different directions: the solenoid (toroid) system bends them in a transversal (longitudinal) plane.

2.2.2. ATLAS Trackers

ATLAS has two tracking detectors: the inner detector and the muon spectrometer. The inner detector is the detector closest to the beam, to track charged particles and measure their momenta with high precision. The muon spectrometer is positioned in the region farthest from the beam. The particles that reach the muon spectrometer are mostly muons and neutrinos, but neutrinos have no charge and interact very weakly, so they are not detected.

The inner detector [33] reconstructs tracks from particles and by doing so, allows for the identification of decay vertices with high efficiency. Its outer radius is 115 cm. The inner detector, shown in figure 2.4, uses three technologies: silicon pixels, semiconductor strip tracker, and transition radiation tracker. The inner detector provides a combined coverage of $|\eta| < 2.5$ and has a resolution of the momentum measurement of:

$$\frac{\sigma(p_\mathrm{T})}{p_\mathrm{T}} = 0.04\,\% \times p_\mathrm{T}[GeV] \oplus 1.5\,\% \text{ [33]} \tag{2.3}$$

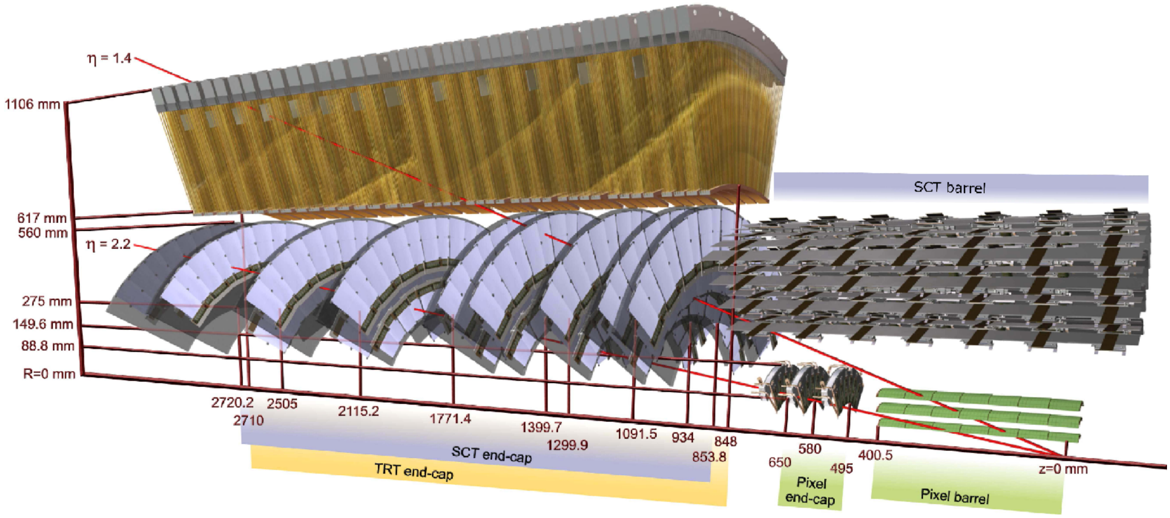
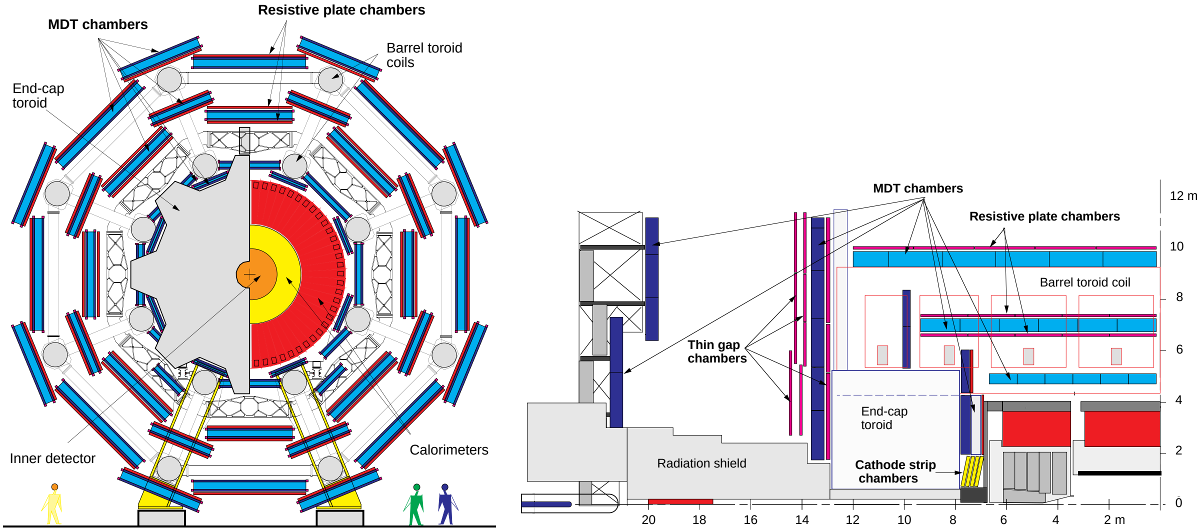


Figure 2.4.: Inner detector and its sub-detectors [30].

The two inner layers of the inner detector, shown in figure 2.4, use silicon technology. The innermost layer is a pixel detector with 250 μm thickness sensors with an $R-\phi$ inherent accuracy of 10 μm and 115 μm in the beam direction. The leading-edge technology was used, despite its cost, to withstand the radiation from the collisions and ensuring an excellent tracking and vertexing resolution despite the heavy detector occupancy. The second sub-detector, the semiconductor tracker (SCT), uses a standard silicon wafer with microstrips with almost twice the thickness of the pixels sensors, but less expensive and less complex. The intrinsic angular accuracy is 17 μm in the longitudinal (barrel) and radial (endcap) directions.

The technology used in the outer part of the inner detector is a straw detector filled with a gaseous mixture of xenon, carbon dioxide, and oxygen. The straws are the cathode of this detector and a tungsten wire in the centre of the straw is the anode. Charged particles ionise the gas mixture, producing electrons that are collected. Between each straw there is another material and transition radiation is produced proportionally to the energy-mass ratio (E/m) of the charged particle, which ionise further the gas. Electrons produce more radiation, so their passage generates high-threshold hits, which can then be used for their identification. Only a section of the endcap of the transition radiation tracker is shown in figure 2.4. The barrel extends until $|z| = 712$ mm, with similar radius. In both barrel and endcap, the straws' diameter is about 4 mm. Therefore, the transition radiation tracker (TRT) helps the pattern recognition of tracks thanks to many close-by hits.

The muon spectrometer [34] outer diameter of the barrel is about 22 m, reaching 25 m in the endcap wheels. Since all other charged particles are expected to be stopped by the calorimeters, the objective of the muon spectrometer is to reconstruct the muon tracks and improve their measurement in combination with the inner detector. The combined nominal momentum resolution is about $2-4\,\%$ for muons with $p_T < 300$ GeV rising to $10\,\%$ for $p_T > 1$ TeV [34].



(a) (b)

Figure 2.5.: Transverse (a) and one quadrant side (b) views of the muon spectrometer [32].

The cathode strip chambers are multi-wire proportional chambers for precision measurements of the muon tracks which are placed in the innermost ring of the muon spectrometer endcap toroid. They are specially designed to support the high particle fluxes of that region. Moreover, for precision measurements, the monitored drift tubes (MDT) chambers are placed in all the regions where the flux of particles is not so intense (barrel and outer endcap). In addition to these chambers, two types of precision muon trigger chambers were installed near the monitored drift tubes chambers providing a fast trigger signal. They are the resistive plate chambers (RPC) in the barrel and the thin gap chambers (TGC) in the endcap. Figure 2.5 presents the location of the each muon spectrometer system in ATLAS.

2.2.3. Calorimeters

The ATLAS calorimeter system [35] comprises an electromagnetic calorimeter to measure the energy and direction of electrons, positrons and photons, and a hadronic calorimeter to identify and measure the energy and direction of jets and hadronically interacting particles. Figure 2.6 shows the location of the electromagnetic and hadronic calorimeters in the ATLAS detector. Given the excellent coverage of the calorimeter system (up to $|\eta| = 4.9$), the calorimeters provide a measurement of the total missing transverse energy, necessary for any process involving neutrinos or other non-interacting particles predicted by new physics models such as supersymmetry.

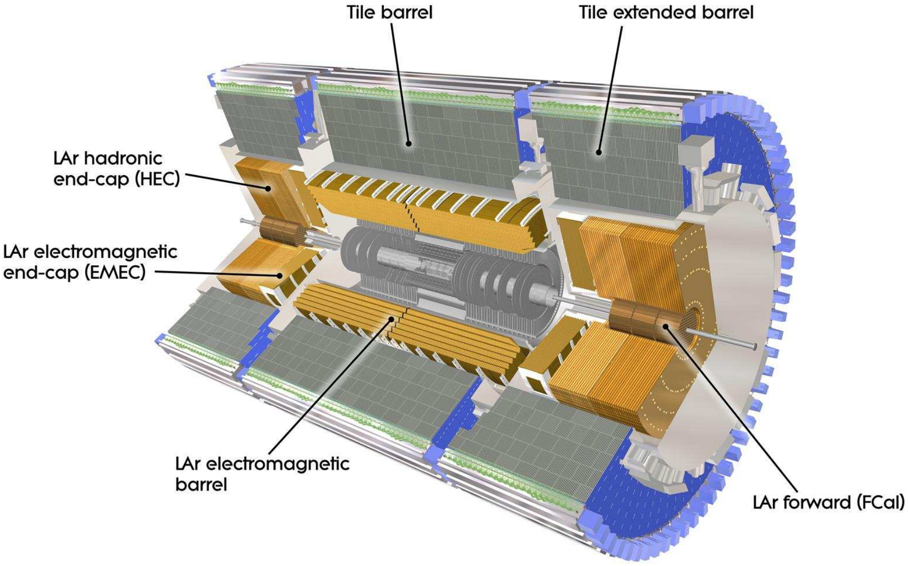
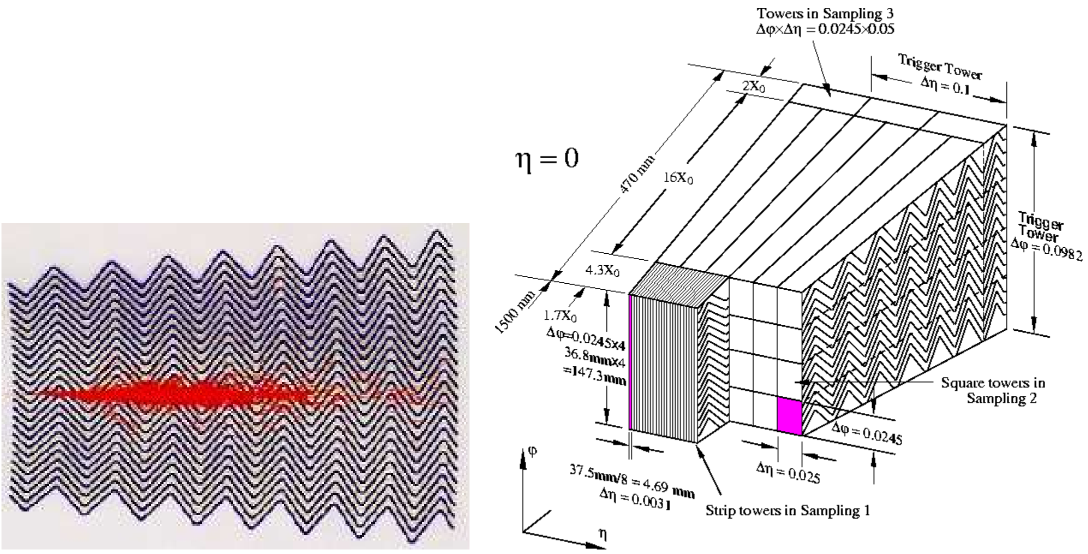


Figure 2.6.: ATLAS calorimeter system [32].

The electromagnetic barrel and endcap calorimeters [36] are sampling liquid argon (LAr) calorimeters with lead absorbers implemented in an accordion geometry. Figure 2.7 (a) presents an image of an electromagnetic shower traversing the electromagnetic calorimeter. An energetic electron entering on the electromagnetic calorimeter produces photons by bremsstrahlung radiation. An energetic photon can convert into an electron-positron pair. In both these processes, the energy of the first particle is split in two in average. Charged particles in the LAr calorimeter will ionise the liquid argon. An electric field then amplifies the shower from this ionisation that is collected in the electrodes. The liquid argon, which is intrinsically radiation hard, is cooled down to a temperature of 80 K using liquid nitrogen.

The coverage in η is up to 1.475 in the barrel and from 1.375 to 3.2 in the endcaps. The electromagnetic calorimeter is segmented longitudinally in 4 layers, with different granularities.

The third layer, sampling 2, of the electromagnetic calorimeter, collects most of the energy of electrons and photons. The $\Delta\eta \times \Delta\phi$ segmentation in this layer is 0.025×0.0245. Sampling 1 has a finer segmentation 0.1×0.0031 to allow the separation of the signals of photons from $\pi^0 \rightarrow \gamma\gamma$ decays. The fourth layer, sampling 3, has a segmentation of 0.0245×0.05. Figure 2.7 (b), presents a segment of the barrel electromagnetic calorimeter. The first layer, sampling 0 or presampler, absent from the figure, has just active material and improves the overall resolution of the electromagnetic calorimeter by correcting energy losses in the material in front of it [37]. Also, the presampler may help to distinguish electrons from unconverted photons, since the latter are insensitive to the inner detector and electrons can start the shower development already in the inner detector. The presampler granularity is 0.025×0.1 and covers the pseudorapidity of $|\eta| < 1.82$ [38].



(a) (b)

Figure 2.7.: Electromagnetic shower in the accordion shaped electromagnetic calorimeter (a). Segmentation of the LAr electromagnetic barrel calorimeter [32] (b).

The energy resolution for electrons in the barrel electromagnetic calorimeter is:

$$\frac{\sigma_E}{E} = \frac{10.1\,\%}{\sqrt{E(MeV)}} \oplus 0.17\,\%\ [30] \tag{2.4}$$

The central hadronic Tile calorimeter is a large sampling calorimeter that uses steel as absorber and scintillators made of polystyrene doped with PTP and POPOP [1] as active material [39]. It is composed of a long barrel and two extended barrel cylinders, covering a pseudorapidity of $|\eta| < 1.7$. Figure 2.8 shows the scheme of one module for half the long barrel and one extended barrel. The gap between the barrel and the extended barrel allows the passage of cabling and services for the inner detector and electromagnetic calorimeter. Radially, each module is further segmented into three layers, referred as A, BC, and D, or 0, 1 and 2, with increasing radial distance. Their thickness is respectively, 1.5, 4.1 and 1.8 λ_I (nuclear interaction length) in the long barrel and 1.5, 2.6 and 3.3 λ_I in the two extended barrels. It is composed of 5,184 cells, each one readout by two photomultipliers. The Tile calorimeter is also divided into four electronic partitions, each composed of 64 modules. The segmentation in ϕ, η and radius defines three-dimensional Tile calorimeter cells, with a segmentation in $\Delta\eta \times \Delta\phi$ of 0.1×0.1 for the layers A and BC and of 0.2×0.1 for layer D in the long barrel. In the two extended barrels, the segmentation is slightly different.

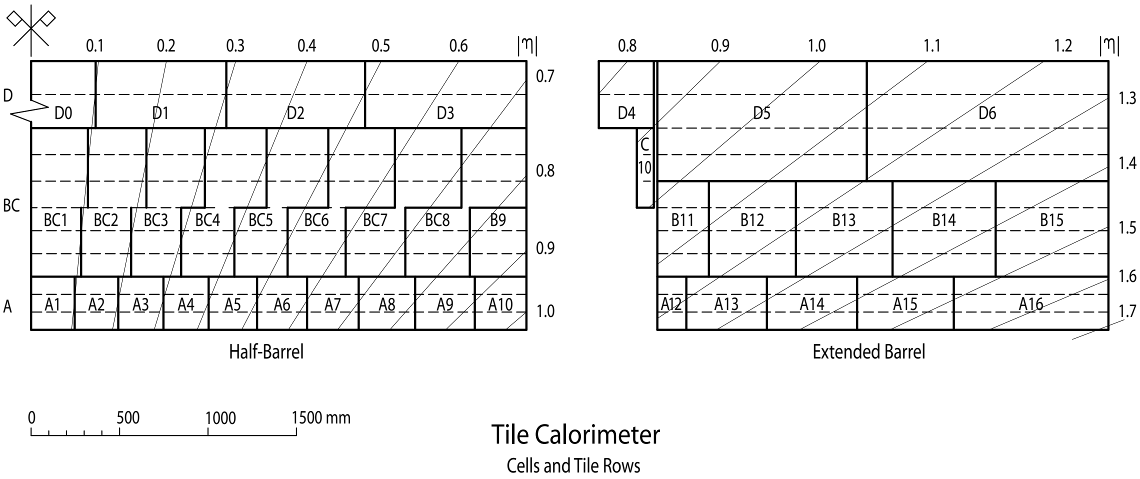


Figure 2.8.: Representation of Tile calorimeter cells in one half barrel and one extended barrel [32].

---

[1] PTP and POPOP are the para-terphenyl-benzene and the 1,4-bis-(2-(5-phenyloxazol-2-yl))-benzene molecules, respectively.

The nominal energy resolution obtained for the Tile calorimeter is:

$$\frac{\sigma_E}{E} = \frac{56.4\,\%}{\sqrt{E(GeV)}} \oplus 5.5\,\% \quad \text{for pions [30].} \tag{2.5}$$

The hadronic endcap calorimeter has a pseudorapidity coverage of $1.5 < |\eta| < 3.2$ and is another sampling calorimeter with a copper absorber and liquid argon as active medium. It is composed of two wheels on each side of the ATLAS detector with a ϕ segmentation of 0.2. The hadronic endcap calorimeter has an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{21.4\,\%}{\sqrt{E(GeV)}} \qquad \text{for electrons and} \tag{2.6}$$

$$\frac{\sigma_E}{E} = \frac{70.6\,\%}{\sqrt{E(GeV)}} \oplus 5.8\,\% \qquad \text{for jets [30].} \tag{2.7}$$

To have the best possible hermetic detector, a forward calorimeter covers the pseudorapidity region of $3.1 < |\eta| < 4.9$. Again, due to the high energy and particle flux, the active medium is liquid argon. The forward calorimeter has three layers. The passive material for the innermost layer is copper and for the remaining layers is tungsten. This calorimeter has an energy resolution of:

$$\frac{\sigma_E}{E} = \frac{28.5\,\%}{\sqrt{E(GeV)}} \oplus 3.5\,\% \qquad \text{for electrons.} \tag{2.8}$$

$$\frac{\sigma_E}{E} = \frac{94.2\,\%}{\sqrt{E(GeV)}} \oplus 7.5\,\% \qquad \text{for pions [30].} \tag{2.9}$$

2.3. Trigger and data acquisition

2.3.1. Introduction

The trigger and data acquisition system of ATLAS [40] plays a crucial role in the physics program. The proton bunches from LHC collide head-on every 50 ns in 2012, which represents a rate of 20 million collisions per second (half the LHC designed rate) while it is possible to save only around 600 events per second. The trigger system is responsible for selecting in real time events interesting for the various physics searches and measurements, while rejecting the non-interesting ones. Figure 1.9 shows that this kind of attractive events do not often occur. The total cross section is around 10^8 nb at the centre of mass energy used for this thesis studies, $\sqrt{s} = 8$ TeV, and is dominated by soft QCD interactions. The cross section production of b-jets is less than 1 % of the total cross section. It is necessary to go down an additional three orders of magnitude in cross section to start seeing rarer physics processes like jet production with transverse energy above 100 GeV, which have a rate around 100 Hz. The production of vector bosons, top, Higgs, are decreasingly frequent, and the trigger must be tuned to recognise their specific signatures. For example, in the decay $Z \rightarrow \mu\mu$, the trigger should be able to identify at least one isolated high-p_T muon. A high-p_T isolated lepton is an indication of a heavy-particle decay.

In addition to the already known Standard Model physics presented in figure 1.9, there can be a large variety of expected or unexpected new physics processes. For this reason, the trigger also needs to record events where unknown physics may take place, by selecting generic high-p_T signatures.

Another challenge imposed by the small bunch spacing is the size of the detector. A muon travelling close to the speed of light will exit the detector after two to three bunch-crossings have taken place.

In Run-1, three levels composed the trigger system: Level 1, Level 2 and Event Filter. An event is rejected if it does not pass one of the trigger levels. If the event passes the criteria of a trigger level, it seeds the following level or is saved on tape if it passes the last level. Also, the trigger system makes use of early rejection, *i.e.* as soon as the event does not satisfy any of the selection signatures, it is rejected, without further processing.

Figure 2.9 shows the schematics of the trigger and data acquisition (TDAQ) system of ATLAS in the year 2012, with the trigger scheme on the left side and the data acquisition (DAQ) scheme on the right. The design values for latency, rate and data volume are also shown in this figure. The grey boxes represent the actual values from 2012, which will be summarised next.

With a maximum latency of 2.5 μs, the Level 1 (L1) trigger uses only the muon detectors and the calorimeter to identify electrons/photons, jets, $\not{E}_T$, hadronic tau decays and muons, using simplified algorithms running on dedicated hardware. If the event passes the first level, the position (η, ϕ) of these objects is sent to the next level. The output rate was 70 kHz in 2012.

In Level 2 (L2), software based algorithms are processed by a farm of commodity computers in about 75 ms. It reduces the rate by one order of magnitude. L2 starts by

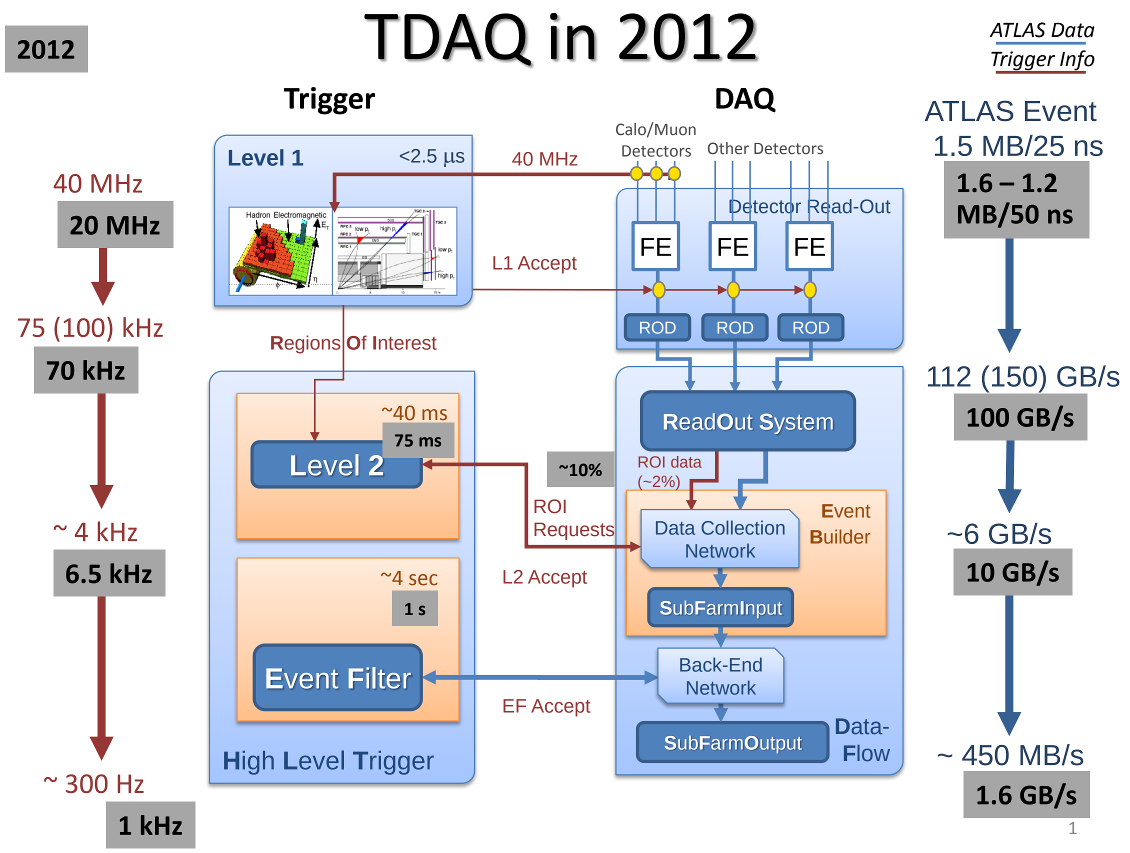


Figure 2.9.: Overview of the trigger and data acquisition system [41]. The design values for latency, rate and data volume are shown in red for the trigger parameters and in blue for the DAQ parameters. The grey boxes show the average values of these parameters in 2012.

using the positions of the high p_T signatures identified by L1 and requesting to the DAQ system the detector information, including tracking, in a Region of Interest (ROI) around those seed positions, corresponding to around 10 % of the full detector data. It then executes a sequence of algorithms designed to verify or reject the signal identified by the L1. It uses the full detector granularity and improved calibrations.

If L2 accepts the event, the Event Builder (EB) is executed, collecting all the information fragments from the sub-detectors. The Event Filter (EF) is the last step of the trigger, which runs offline-like algorithms within around one second to make the final decision to discard or to save the event on tape. The final output rate was 1 kHz in 2012, which represents about 1.6 GB of information saved per second.

2.3.2. Level 1

The first level of the trigger is fully implemented in the front-end electronics of the detector read-out system, which executes simple algorithms. It uses information from the muon detector and the calorimeter, each with different algorithms.

Muons with energy above a few GeV can reach the muon chambers. The key sub-detectors for triggering are the RPC, in the barrel, and the TGC, in the endcap. The trigger algorithms use the coincidence hits in two layers to reconstruct the trajectory of the muon, coming from the origin of the detector. The curvature of the trajectory of the muon determines which trigger threshold is fired.

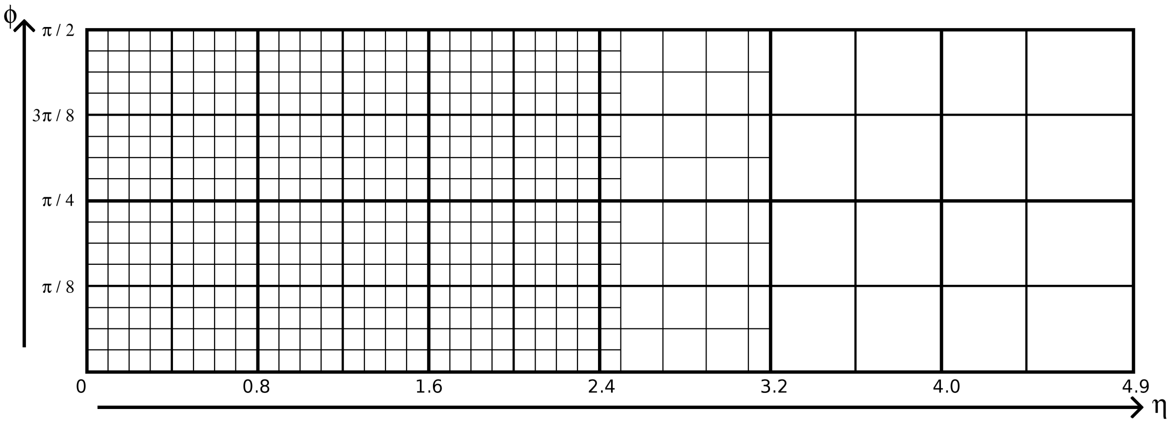


Figure 2.10.: Trigger towers distribution for one octant of the calorimeter in the $\eta - \phi$ plane.

The calorimeter finds the rest of the particle signatures. The granularity used for the calorimeter in the L1 is very coarse and is sketched in figure 2.10. The squares in $\eta - \phi$ space shown in this figure are referred to as trigger towers (TT) and have sizes of $\Delta\eta \times \Delta\phi$: 0.1×0.1 in the $|\eta| < 2.5$ region, 0.2×0.2 in $2.5 < |\eta| < 3.2$ and 0.4×0.4 in $3.2 < |\eta| < 4.9$. Furthermore, these TT are evaluated for the electromagnetic and hadronic calorimeters separately.

A sliding-window algorithm [42] is then run to find local maxima of transverse energy, with a window of size $N_\phi \times N_\eta$, where N_ϕ and N_η are the numbers of towers in the window in ϕ and η, respectively.

For electrons, photons, taus and hadrons, the window has size 4×4, shown in figure 2.11. Horizontal or vertical sums of two TT are performed with size 1×2 or 2×1, respectively, in the centre of the electromagnetic towers forming four electromagnetic

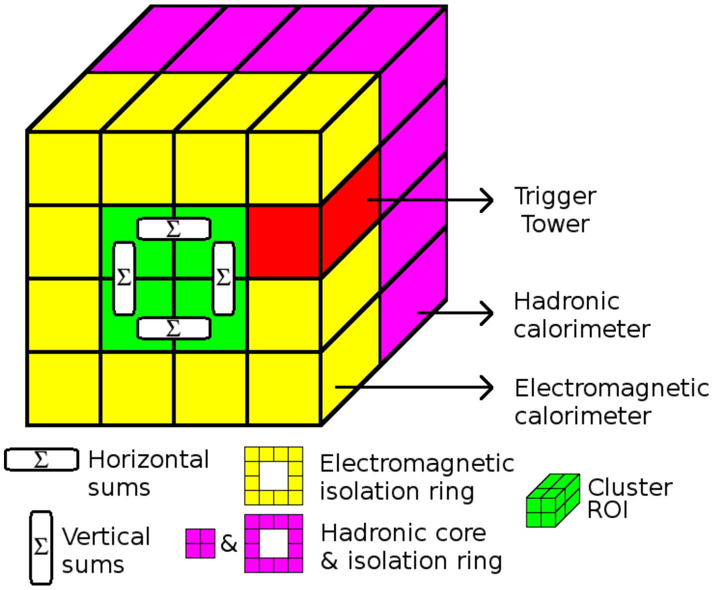


Figure 2.11.: Sliding window representation with 4×4 trigger towers. Adapted from reference [43].

clusters. The twelve TT, around that centre, form the isolation ring. The hadronic core and isolation ring are formed from the four hadronic central TT and the twelve hadronic TT around them, respectively. Finally, the centre cluster ROI is composed of four hadronic clusters, each composed by one electromagnetic TT and one hadronic TT adjoined. Level 1 algorithms use these calorimeter cell sums to identify different particles. Some of these particles are now briefly addressed.

Since no tracking is available in L1, electrons and photons are identified in the same way. Their signatures require one electromagnetic cluster with transverse energy higher than a given threshold, E_{T}^{EM}, the sum $\mathrm{E_T}$ in the electromagnetic isolation ring to be smaller than a given threshold, E_{T}^{EMiso}, the same for the hadronic core, $E_{\mathrm{T}}^{HADcore}$, and for the hadronic isolation ring, E_{T}^{HADiso}. The taus and hadrons are also considered together and differ from the electron/photon by replacing the threshold in the electromagnetic cluster and the hadronic core with the threshold in the hadronic cluster, E_{T}^{HAD}.

Regarding the jet and neutrino signatures, the algorithms use jet elements, which have a size of 0.2×0.2 in $\Delta\phi \times \Delta\eta$, summing over hadronic and electromagnetic TT. Three windows sizes are available: 2×2, 3×3 and 4×4, given in terms of jet elements. The jet algorithm runs in the $|\eta| < 3.2$ region, whereas the algorithm that sums the energy in the calorimeters, and evaluates the value of $\not{E}_{\mathrm{T}}$, uses the whole calorimeter region: $|\eta| < 4.9$.

The position of these objects is then saved and passed to Level 2 for further processing.

2.3.3. High level trigger

During Run-1, the high-level trigger (HLT) was subdivided in L2 and EF. The HLT is software based with algorithms that can be either "feature extraction" or "hypothesis". The former is used to reconstruct detector information, and the latter will test if a specific criterion is passed.

The L2 has access to the full granularity of the calorimeter for better determination of the positions and energies. The inner detector information is now also available to the trigger system. The L2 reconstructs ROIs of variable size $\Delta\phi \times \Delta\eta$ around the positions received by L1.

A quick overview of the trigger chains used for the ZH analysis, presented in chapter 4, is now discussed.

Electron and photon signatures [44] at the L2 use a window with size $\Delta\phi \times \Delta\eta = 0.4 \times 0.4$ and the reconstruction algorithm uses the knowledge that they deposit most of their energy in the second layer of the electromagnetic calorimeter. If there is a track associated with the electromagnetic cluster, it is considered an electron. Otherwise, it is a photon. The EF reconstructs the electron with offline-like tools.

For muon trigger signatures [45] at L2, a region of interest is formed around the L1 seed. An algorithm then chooses the closest inner detector track to match the segment found in the muon spectrometer. A weighted average between the information of the two detectors recalculates the p_T of the L2 combined muon. EF muons are also obtained using the ROI based method, but now implemented with the offline-like reconstruction algorithms.

Electron (muon) triggers are isolated if the scalar sum of the p_T of the tracks within a cone of size $\Delta R = 0.2$ divided by the p_T of the electron (muon) is smaller than a certain value, typically 0.1.

2.3.4. Further details

A trigger chain is a sequence of algorithms, feature extraction and hypothesis algorithms, starting from L1 until the EF to identify physics objects candidates which can be combined to form physics signatures.

A trigger menu is a collection of configurable triggers containing information on:

- which sequences and chains of algorithms to run;
- the selection condition and threshold;
- isolation;
- prescale (PS);
- pass-through (PT).

To reduce the rate, a prescale might be used so that only one in a certain number of events (n_{PS}) passing the trigger is saved to tape. It is also frequent to use triggers in pass-through, for determination of trigger efficiencies for example, meaning that one in a certain number of events, n_{PT}, might be saved even if it does not pass the selection conditions.

At the output of the EF, the events are classified into streams according to which trigger type selected the event. The main streams are:

- Muon stream that collects events passing the muon triggers.
- Egamma stream that collects events passing the electron or the gamma triggers.
- JetTauEtMiss stream that collects events passing the jet, the tau or missing transverse energy triggers.
- MinBias stream that collects events passing the minimum bias triggers, composed of just random events in 2012.

It happens with some frequency that an event survives different trigger types, *e.g.* passes a muon and an electron trigger. In this case, in 2012, the event was saved both in the Muon and Egamma streams.

2.4. The pileup challenge

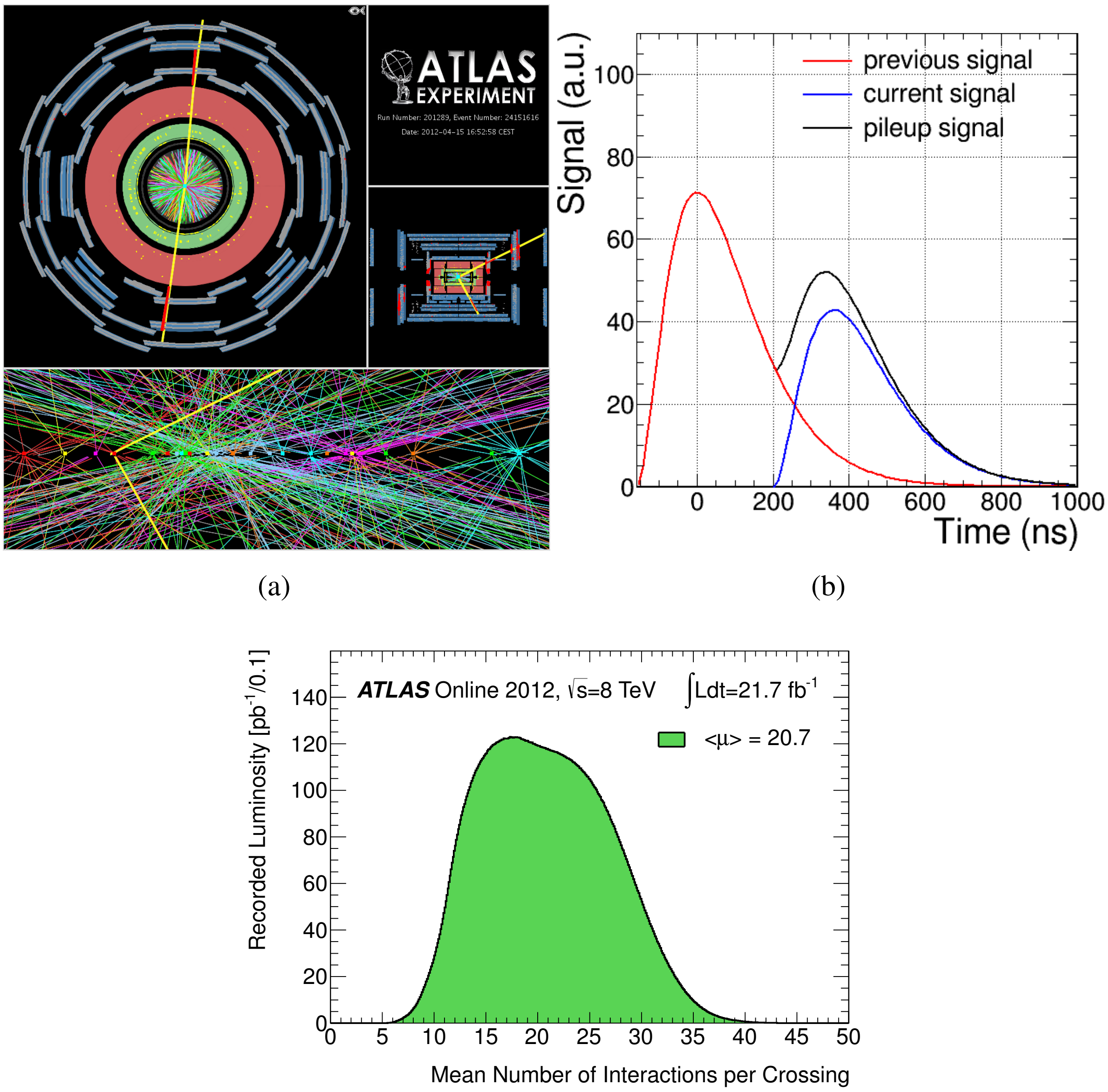
Each time that two bunches of protons intersect, there can be more than one collision. These additional collisions in the event are known as in-time pileup. Figure 2.12 (a) displays one event with three images. The top left image shows the projection of the ATLAS detector in the transverse plane; the centre right image shows the detector projected in the longitudinal plane and the bottom image shows a zoom of the longitudinal plane to see the detail of the inner detector track and vertex reconstruction. The electromagnetic calorimeter appears in green, the hadronic calorimeter in red and the muon spectrometer in blue. This figure shows an event with 25 primary reconstructed vertices, each representing one proton-proton collision. In one of them, two muons were reconstructed and identified as originating from a

Z boson decay. The large number of particle tracks shown, illustrates the complexity of the event reconstruction.

In addition to the in-time pileup, characterized by the number of primary vertices, another challenge exists in ATLAS. The detector time response is longer than the inter-bunch separation, so the signal from a previous collision event can overlap the signal of the current event. Figure 2.12 (b) shows an example, where the calorimeter signal pulse of the current event, in blue, overlaps with the previous signal pulse. The sum of the two distributions appears in black, which will be the pulse with pileup that the calorimeter cell will record. This overlap is known as out-of-time pileup. A parameter that is often used to study the out-of-time pileup is the mean number of interactions per bunch crossings, $\langle\mu\rangle$, is defined by equation 2.10:

$$\langle\mu\rangle = \frac{L_{bunch} \times \sigma_{inel}}{f_{LHC}} \tag{2.10}$$

where L_{bunch} is the instantaneous luminosity per bunch crossing, σ_{inel} is the cross section for inelastic processes (73 mb) and the f_{LHC} is the LHC revolution frequency. The $\langle\mu\rangle$ distribution for 2012 is shown in figure 2.12 (c). It reached values up to 40 in 2012, with an average of 20.7.



(a) (b)

(c)

Figure 2.12.: ATLAS event display of an event with 25 reconstructed vertices (a) [46]. An example of the overlap between two consecutive pulses in a calorimeter cell (b). The pileup pulse is the sum of the two individual pulses. Mean number of interactions per bunch crossings during 2012 data taking period (c) [47].

2.5. Data reconstruction for real events

Figure 2.13 shows the data reconstruction scheme for real collision events, as it was in the 2012 data taking. When the trigger saved the events, the detector information in byte-stream format was stored in files for offline processing. The offline reconstruction chain started with the byte-stream conversion step, which transformed the byte-stream into C++ objects such as cells or hits that would be used by the next algorithms in the reconstruction chain. These objects, called raw data objects (RDO), were stored in files. Next, detector objects like clusters, tracks and vertices were reconstructed and stored in the event summary data (ESD) files. Then, these detector objects were combined to obtain the physics objects, leptons, jets and $\not{E}_{\mathrm{T}}$, and stored in the analysis object data (AOD) format. These data formats could only be accessed within the Athena framework [48]. In the last step, the event summary information was stored in ROOT [49] trees for analysis. ROOT is a lighter analysis framework.

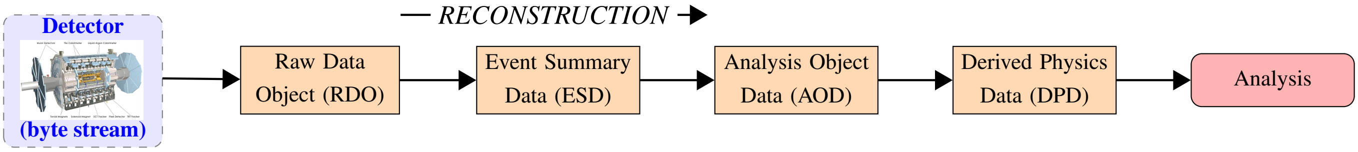


Figure 2.13.: Data flow from real events from the detector until the analysis.

2.6. Detector object reconstruction

Physics analysis objects, like electrons, photons, muons, taus or jets are reconstructed starting from calorimeter clusters and tracks.

2.6.1. Sliding window and topological clusters

Clusters are groups of calorimeter cells that represent energy depositions of single particles. Analyses in ATLAS can reconstruct them by either a sliding-window or a topological clustering algorithm.

The sliding-window algorithm [42] begins by dividing the calorimeter towers of a certain size in a rectangular region of the $\eta - \phi$ plane. The towers are either electromagnetic with a 0.025×0.025 granularity of the electromagnetic calorimeter, with a pseudorapidity limit of 2.5, or combined with the hadronic calorimeter with the $|\eta| < 5.0$ limit but with 0.1×0.1 granularity. In the second step, this algorithm will "slide" through all the towers with a particular window of programmable size $N_\eta \times N_\phi$. If the sum of the tower energy in the window is above a determined threshold, a pre-cluster is formed, and its position is calculated using the centre position of each tower in a smaller window, weighted by their energy. This smaller window has the same centre as the initial window. If the distance between two pre-clusters is smaller than 2×2 towers, the one with smaller energy is discarded. This algorithm is faster than the topological clustering algorithm and is used for electron/photon reconstruction offline and at the trigger level. Particularly, EF trigger definitions of electrons/photons use the window $N_\eta \times N_\phi = 3 \times 5$ with the electromagnetic calorimeter granularity, and of taus definitions use the $N_\eta \times N_\phi = 8 \times 8$ window with the combined granularity [50].

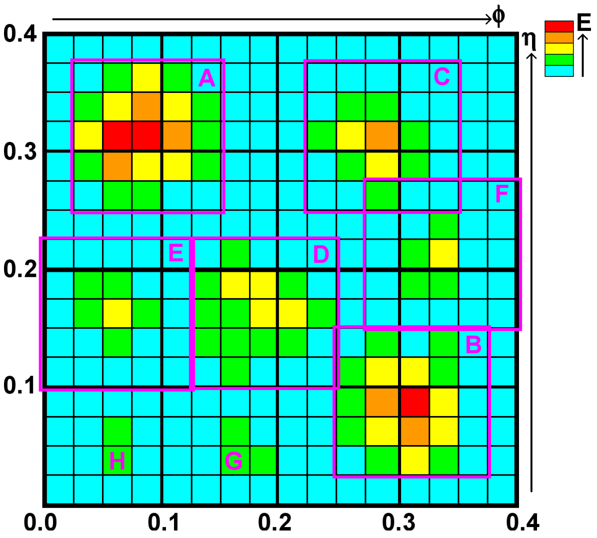
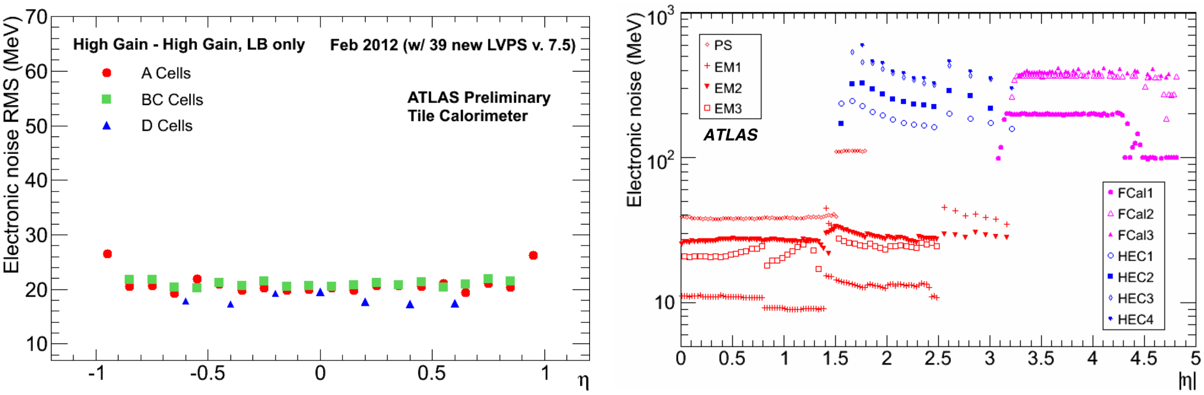


Figure 2.14.: Example of application of the sliding-window algorithm to find clusters.

Figure 2.14 shows an example of the application of the sliding-window clustering algorithm, which uses the electromagnetic calorimeter granularity. In this example, six clusters were found, labelled from A to F in decreasing energy order. The G and H tower groups did not have enough energy to pass the threshold requirement. In the case of the A and B tower groups, several windows with distances smaller than 2×2 could be formed but the one selected ought to have the highest energy. In the case of the C and F towers groups, they make two separate windows since the distance is 4×2, although they have three overlapping towers.

The other algorithm is topological clustering [42]. The objective is to reconstruct 3-dimensional clusters representing single particle energy depositions. First, the algorithm classifies the cells according to their signal to noise ratio, where the noise value takes into account the electronic and pileup noise. The former is given by the root mean squared (RMS) of the cell energy in calibration runs (no collisions) and evaluated for each cell. The electronic noise for the three layers of the Tile calorimeter is presented in figure 2.15 (a), with an average value around 21 MeV. In figure 2.15 (b), the electronic noise is shown for the LAr calorimeters, where the hadronic endcap and the forward calorimeter have larger noise due mainly to the size of the cells in this region.



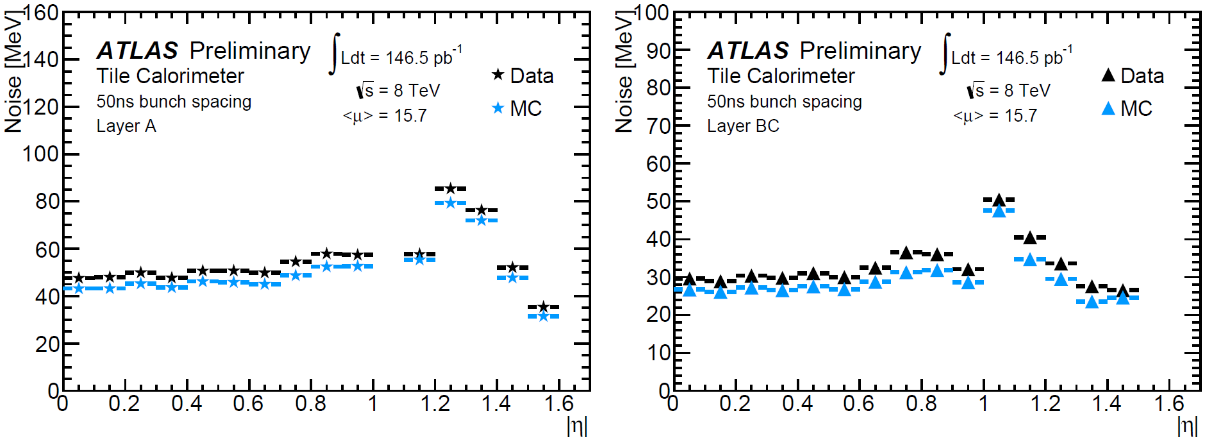
(a) (b)

Figure 2.15.: Electronic noise for the Tile calorimeter for each of the cells from the long barrel (a) [51] and for the LAr calorimeters (b) [52].

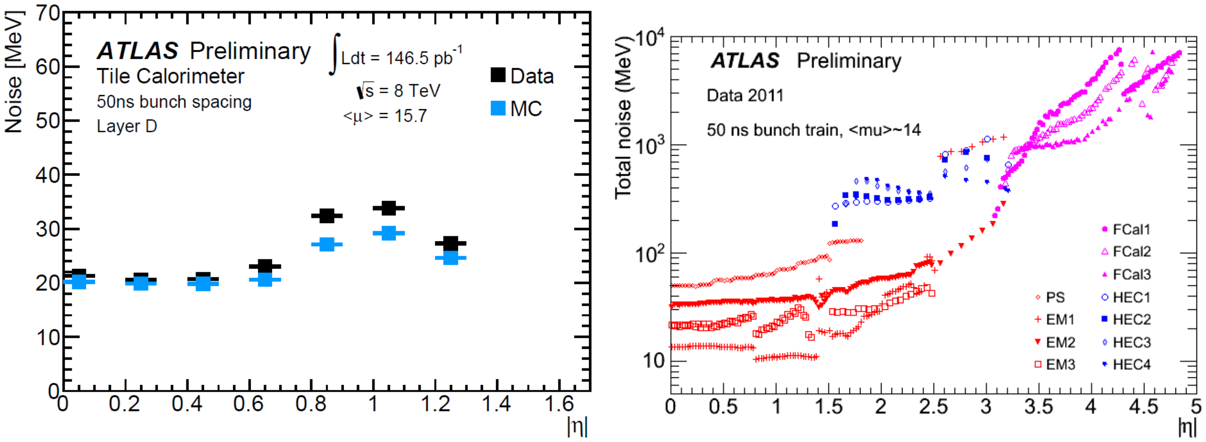
The pileup noise results from the effect of the out-of-time pileup in the calorimeter. A comparison between real and simulated pileup noise is shown in figures 2.16 (a), 2.16 (b) and 2.16 (c) for the Tile calorimeter layers A, BC and D, respectively. The pileup noise is always higher than the electronic noise. The total noise, evaluated for each cell, is given by the quadratic sum of the electronic and the pileup noise. Figure 2.16 (d) shows this quantity for the LAr-based calorimeters.

The ratio between the cell measured energy and its total noise energy values is referred as cell energy significance, σ_{cell}.

The topological cluster (topocluster) algorithm starts by finding *seed* cells with an energy significance above four: $\sigma_{cell} > 4$. Neighbouring *seed* cells are added to the same



(a) TileCal Layer A (b) TileCal Layer BC



(c) TileCal Layer D (d) LAr calorimeters

Figure 2.16.: Pileup noise for the three Tile calorimeter (TileCal) layers for $\langle\mu\rangle = 15.7$ (a-c) averaged over the 64 modules [53]. Total noise for the LAr calorimeters for $\langle\mu\rangle = 14$ (d) [52].

cluster. In a second step, the algorithm looks for neighbouring cells that have an energy significance above two: $\sigma_{cell} > 2$. These are referred as *neighbour* cells and included in the cluster. This step will iterate until there are no more *neighbours* to be added to the cluster. In the final step, all cells touching the cluster, independently of their energy significance, are added to the cluster.

The example used previously in figure 2.14 is shown for the topocluster algorithm in figure 2.17. The colour scheme selected is:

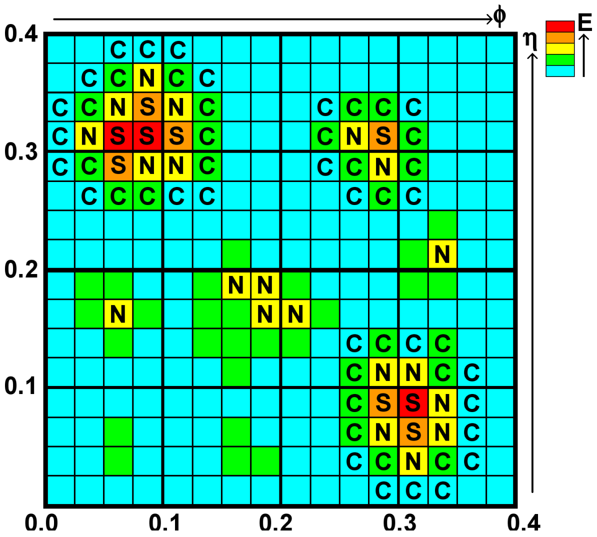


Figure 2.17.: Example of application of the topological algorithm to some clusters.

🟥 $\sigma_{cell} > 5$ 🟧 $\sigma_{cell} > 4$ 🟨 $\sigma_{cell} > 2$ 🟩 $\sigma_{cell} > 1$ 🟦 $\sigma_{cell} > 0$

The *seed*, *neighbour* and remaining cells are marked with “S”, “N” and “C” for visual clarity, forming three clusters. The unused yellow cells were also labelled with “N”, even though they are not part of any cluster since they have a cell energy significance larger or equal to two.

The cluster centre is calculated as the average position of the cells, weighted by the respective cell energy. The energy of the cluster is the sum of the individual cells that belong to the cluster.

Events with separate clusters become rarer with the increasing number of interactions per bunch crossing, in particular in the forward detector region, and the cluster might reach a vast size. When two clusters share the same cell, its energy is divided by them depending on their energies and on the distance from the cell to each centre.

2.6.2. Track and vertex finding

Tracks are obtained by grouping the *hits* deposited by charged particles in the inner detector and the muon spectrometer. There are two sensors in each silicon module. A module *hit* occurs when both sensors have recorded the hit. The name sensor hit or *hole* is given when only one of the sensors records the passage of the particle. In ATLAS, there are two approaches for track reconstruction [54]. The inside-out approach is the main track reconstruction and searches for three hits in the silicon detectors (pixel and SCT) to seed the reconstruction. This seed is then propagated to the remaining layers of the silicon trackers. Other hits are assigned to the track or rejected as part of the track, using a Kalman filter [55]. The ambiguity of having a hit belonging to more than one track is solved through a score attributed to each track. A reward (penalty) is given to the score of the track for each module hit (hole) that is a part of the track. An exception to the penalty occurs when the sensor is not available. Such hits are referred as *dead sensors*. In the end, the ambiguous hit will belong to the track with the highest score. Finally, tracks that have at least three silicon hits are then extended to the TRT. This approach has the drawback that it only uses tracks that survive the scoring from the silicon detectors. Particle decays may produce secondary particle tracks with a few hits in the silicon detectors. Also, a photon might convert into an electron-positron pair even after traversing the silicon-based tracker. Such tracks would be lost since there was no silicon seed. The track reconstruction may also use an outside-in approach, based on the Hough transform mechanism [56, 57], using only those hits that were not considered in the first approach.

Each track has information on the number of module hits, usually referred to as silicon hits, the number of holes and the number of dead sensors in the trajectory of the track. Together with the number of hits from the TRT, the number of high-threshold hits is available and used in the identification of electrons.

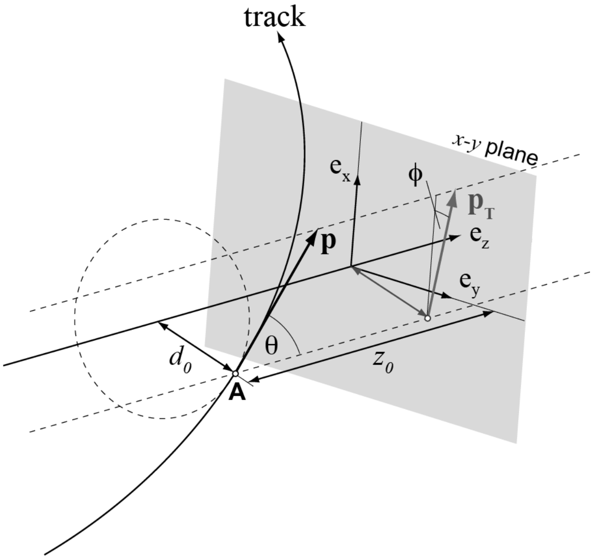


Figure 2.18.: Sketch of a track trajectory with its main information. Adapted from reference [58].

Due to the magnetic field provided by the solenoid, a charged particle bends, forming a circle in the x-y plane. Since particles have also momentum along the Z axis, the final track has a helix shape. Figure 2.18 shows a sketch of a track decomposed in its parameters, which are shortly described below:

A	The closest point of the track's propagation to the Z axis, which is called perigee;
d_0	Distance between the track at the perigee to the Z axis, referred as the transverse impact parameter;
z_0	Z coordinate of the track at the perigee, referred as the longitudinal impact parameter;
ϕ	Azimuth angle between the track direction at the perigee (tangent) with the X axis;
θ	Polar angle of the track direction measured with the Z axis;
η	The pseudorapidity of the track is defined from equation 2.2 using the θ angle.

The track fitting algorithm does the calculation of these track parameters.

Finally, the tracks are combined to form vertices. Most of them will be along the beam direction where the $p-p$ interactions took place and are referred as primary vertices. The event can also have secondary vertices, resulting from decays of particles or pair production as mentioned before. The primary vertex with the largest scalar sum of the transverse momentum of its associated tracks is designated as the hard scattering process vertex, while the remaining are considered vertices from in-time pileup.

2.7. Reconstruction and performance

The identification of particles is essential for any analysis. This section describes in detail the reconstruction of each particle used for the analysis presented in the thesis.

2.7.1. Muons

Muons are very penetrating charged particles, the only ones that can cross the calorimeters, depositing a small fraction of the energy there, and crossing the muon chambers before leaving the detector. They are, therefore, identified due to the presence of a track in the muon spectrometer.

Three reconstruction algorithms, called chains, were used in Run-1 [31, 59]. *Chain 1* used a statistical recombination between the tracks reconstructed in the muon spectrometer and in the inner detector. *Chain 2* used a refit of the track using the hits in both muon spectrometer and inner detector. *Chain 3* tried to combine both approaches. The ZH analysis used only muon reconstructed with the *Chain 2*.

Muons reconstructed by both the inner detector and the muon spectrometer have the designation of *combined* muons and typically have the best energy measurement. Given the acceptance of the inner detector, *combined* muons are only reconstructed up to $|\eta| < 2.5$. When muons have a pseudorapidity near 2.5, they only leave a track segment in the inner detector and are referred as *silicon associated forward* muons.

If a muon is identified and reconstructed only by the muon spectrometer, it has the designation of *standalone* muon. These muons extend the acceptance of muons to 2.7 in $|\eta|$, where only the muon spectrometer is available.

Muons might leave only one track segment in the muon spectrometer if they are low-p_T muons and lose most of their energy in the calorimeter or if the muon is too close to the edge of the muon spectrometer acceptance. Such muons are reconstructed if the inner detector track extrapolated to the muon spectrometer is associated with a track segment in this detector and is labelled as *segment-tagged* muon.

Finally, the *calorimeter-tagged* muon is identified and reconstructed when a track in the inner detector is extrapolated to the calorimeter and associated with an energy deposit compatible with a minimum ionising particle. This reconstruction allows the identification of muons in the central region ($|\eta| < 0.1$) that is not covered by the muon spectrometer, due to cabling access for the calorimeter readout.

2.7.1.1. Muon quality conditions

The combined and calorimeter-tagged muons are required to fulfill some inner detector quality requirements based on the number of hits and maximum number of holes in the different tracking sub-detectors [59]:

- At least 1 pixel hit;
- At least 5 SCT hits;
- Maximum of 2 combined pixel and SCT holes;
- At least 9 TRT hits if the muon track is in TRT acceptance;
- Maximum of 1 dead sensor in both pixel and SCT.

2.7.1.2. Muon performance

Reconstruction efficiency The methodology to measure the muon reconstruction efficiency differs between the central region, $|\eta| < 2.5$, where the information from the inner detector and the muon spectrometer is available, and the forward region, $2.5 < |\eta| < 2.7$, with only the muon spectrometer [59]. Both regions use the tag-and-probe method [60, 61] that consists on using one better-measured object called "tag" to correct the other object called "probe".

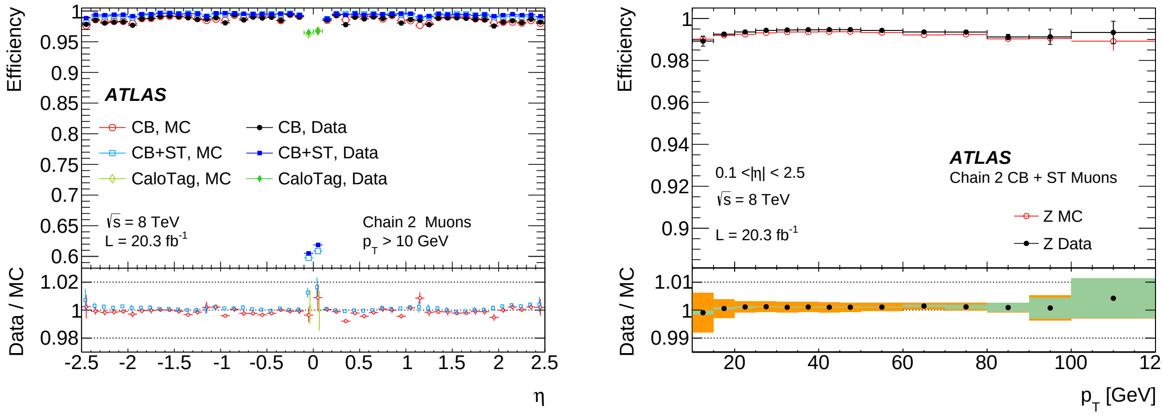
In the central region, the efficiency is calculated by the fraction between the probe muons reconstructed as one type of muons (combined, segmented-tagged or calorimeter-tagged) from all muons reconstructed in the inner detector, using $Z \to \mu\mu$ events.

The tag muon is a combined muon with $\mathrm{p_T} > 25$ GeV, passing the good quality criteria. This muon is also required to be isolated: an $R = 0.4$ cone-region is open around the track and the $\mathrm{p_T}$ scalar sum of tracks with $\mathrm{p_T} > 1$ GeV within the region must be smaller than 15 % of the reconstructed muon $\mathrm{p_T}$. The probe muon must have an opposite charge, an azimuth distance to the tag muon $\Delta\phi > 2.0$, and $\mathrm{p_T} > 10$ GeV. The invariant mass of the two muons, defined as $m = \sqrt{(E_{\mu 1} + E_{\mu 2})^2 - (\vec{p}_{\mu 1} + \vec{p}_{\mu 2})^2}$, must be compatible with the Z boson mass within 10 GeV.

Figure 2.19 (a) shows the efficiencies obtained for data and simulation, for different muon selections: for combined muons (CB), for the combination between combined muons and segmented-tag muons (CB+ST), and for calorimeter-tagged. The information from calorimeter-tagged muons is only shown at $|\eta| < 0.1$, the relevant region for physics analysis such as the ZH search discussed in chapter 4. Including the calorimeter-tag muons, the reconstruction efficiency is larger than 95 % in the range $|\eta| < 2.5$. The bottom panel of figure 2.19 (a) shows the ratio between data and MC for the three muon choices. For the combination of combined and segmented-tag muons, data and MC agree within 1 %.

Besides the Z$\to \mu\mu$ channel, the efficiency is also determined with $J/\psi \to \mu\mu$ decay to explore muons with transverse momentum below 20 GeV [59]. The efficiency as a function of the muon $\mathrm{p_T}$ is shown for the two channels in figure 2.19 (b), excluding the $|\eta| < 0.1$ region. A muon reconstruction efficiency larger than 99 % is observed. The bottom panel shows again the data/MC ratio. All the differences between data and simulation are below 1 %.

The difference between the measured efficiency in data and MC is used as an efficiency



(a)

(b)

Figure 2.19.: Muon reconstruction efficiencies for different muon types as a function of the pseudorapidity (a) and transverse momentum (b) for data and simulation [59].

scale factor or *SF* to correct the simulation:

$$SF = \frac{\varepsilon^{Data}}{\varepsilon^{MC}} \tag{2.11}$$

This scale factor is obtained (and applied) as a function of the direction of the muons (η, ϕ) and their transverse momentum. The efficiency in the forward region, $2.5 < |\eta| < 2.7$, can not use combined muons since the inner detector does not cover this region. An alternative method for the efficiency *SF* calculation is applied, that uses the information from the last $|\eta|$-bin with inner detector coverage [59].

The *SF* is shown in figure 2.20 in the form of a $\eta - \phi$ map for muon *Chain 1*. Analyses then implement these maps as a weight applied to the muon. Scale factors for the muons reconstructed with *Chain 2*, used in the ZH analysis, have similar result [59].

The dominant systematic uncertainties for the reconstruction efficiencies depend on the transverse momentum of the muon. For muon p_T up to 30 GeV, the dominant systematic source with about 0.6 % is the maximum matching distance between reconstructed and the probe muon. In the intermediate p_T region, between 30 and 60 GeV, it results from biases in the tag-and-probe method with an uncertainty of 0.1 %. Above 60 GeV, the uncertainty rises to 0.3 % and are mainly due to the $Z \to \mu\mu$ background. The uncertainty variation of the efficiency with pseudorapidity varies from 0.1 % to 0.5 % except for the calorimeter-tagged

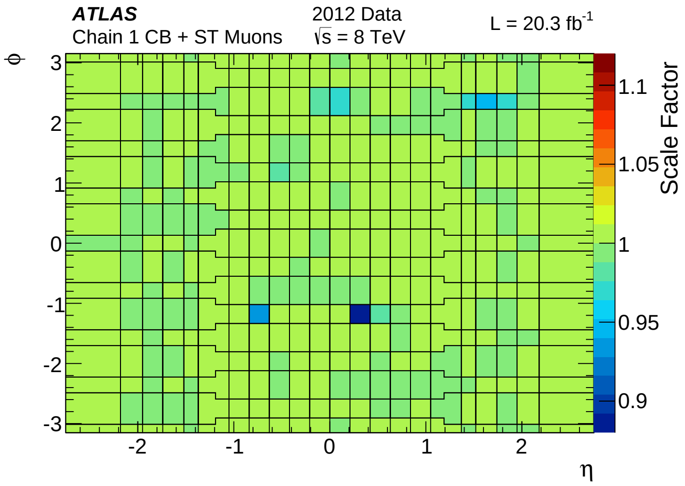


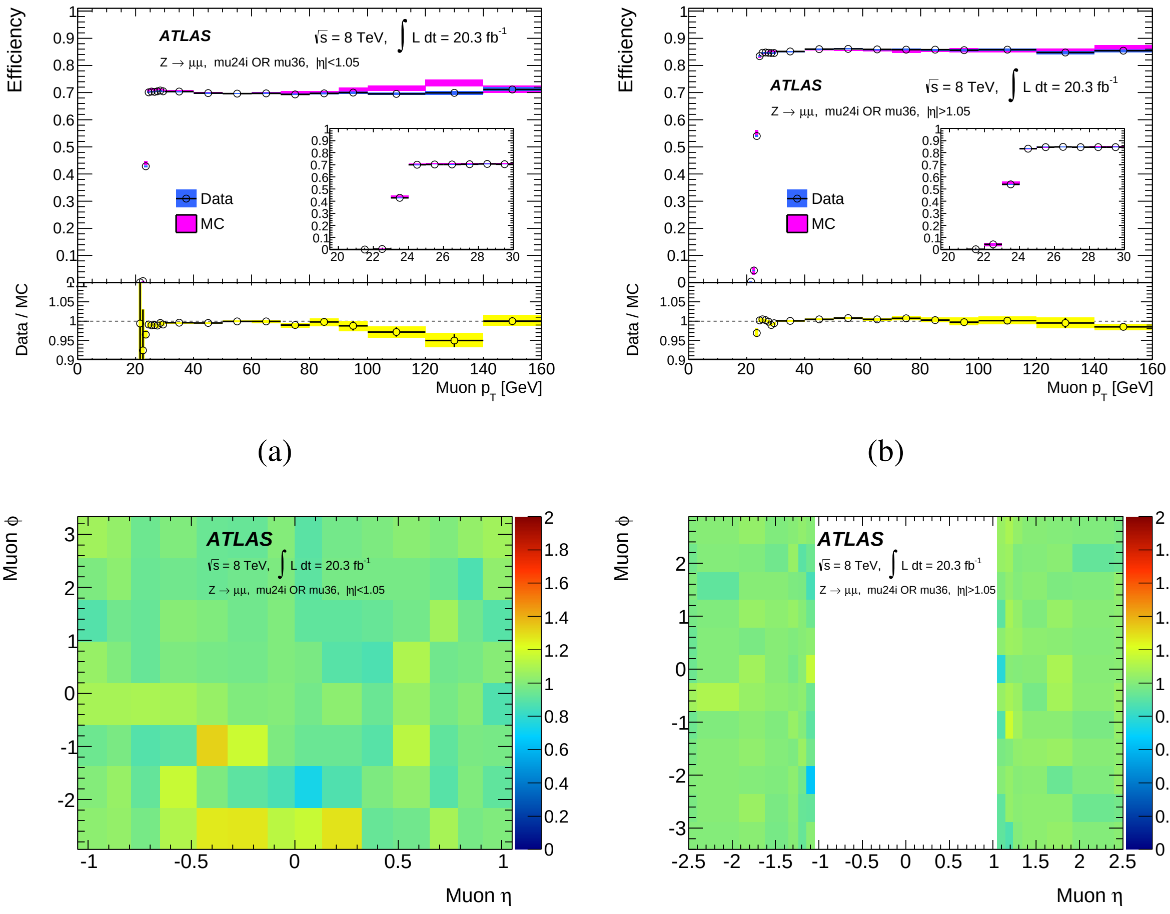
Figure 2.20.: Reconstruction efficiency scale factor obtained as a function of η and ϕ for combined muons with $p_T > 10$ GeV [59].

muons ($|\eta| < 0.1$), where it reaches 0.8 % [59].

Trigger efficiency For the trigger efficiency determination [45], the tag-and-probe method is used. The tag muon is an offline reconstructed muon and the probe muon is the trigger reconstructed muon. The tag muons are required to have tracking isolation (see section 2.3.3) and $p_T > 25$ GeV. Figure 2.21 shows, for muons with pseudorapidity below (left) and above (right) 1.05, the trigger efficiency (top) as a function of the probe muon's p_T, calculated with respect to offline selected muons of a combination of two muon triggers: either an isolated muon with $p_T > 24$ GeV or a muon with $p_T > 36$ GeV. The bottom panels on both figures show the ratio data/MC, which never deviates more than 5 % relative to the unity. From the resulting ratio between data and MC, an $\eta - \phi$ scale factor map is obtained, which is calculated as a function of the detector position and shown in the bottom figures.

The systematic uncertainty sources that affect the trigger efficiency calculation are the same as for the reconstruction efficiency and result in 0.6 % for the region $25 < p_T < 100$ GeV.

Momentum scale and resolution There is limitation to the precision on the description of the solenoid and toroid magnetic field, the amount of material traversed by muons and the interaction that they have with those materials, namely the energy loss. Figure 2.22 shows the invariant mass distribution of $Z \to \mu\mu$ candidate events in data and simulation. There is an apparent discrepancy between data and the uncorrected simulation, which is more visible



(a)

(b)

(c)

(d)

Figure 2.21.: Trigger efficiency as a function of transverse momentum of the probe muon data and simulation for events with either *mu24i* or *mu36* muon triggers for offline reconstruction muons with $p_T > 25$ GeV (top figures) in the $|\eta| < 1.05$ (left) and $|\eta| > 1.05$ (right) regions. The ratio between data and MC is shown on the bottom panels. Trigger efficiency scale factor as a function of the muon $\eta - \phi$ position (bottom figures) for the same muon triggers and pseudorapidity regions [45].

in the ratio data/MC shown in the bottom panel. The blue curve has a wavy structure with data-MC differences up to 7 %.

The invariant mass distributions of the muon pairs in $Z \rightarrow \mu\mu$ and events are obtained as a function of the p_T of the leading muon. These distributions are then used as input for a binned likelihood fit to extract a parametrization of the scale and resolution correction for

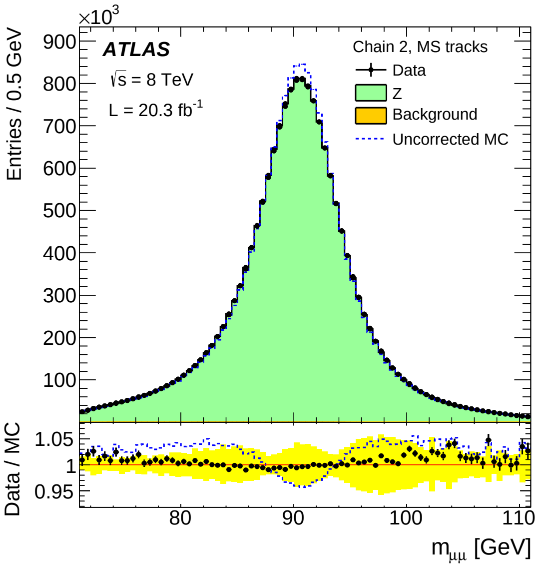


Figure 2.22.: Di-muon invariant mass distribution of $Z \rightarrow \mu\mu$ candidate events reconstructed in data (black points) and MC simulation with (filled green histogram) and without (blue dash line) the scale and resolution correction. The data/MC comparison is shown below. The yellow band represents the systematic uncertainty for the scale and resolution [59].

the simulated muons p_T, for each of the detectors [59].

The green region in figure 2.22 shows the effect of correcting the momentum and resolution of the muon. In the data/MC ratio plot shown in the bottom panel, it is possible to see that the discrepancy falls to less than 1 % with the correction.

The systematic uncertainties considered on the muon scale and resolution are shown with the yellow band in the bottom panel of figure 2.22 with a typical value of 1 % and it never larger than 5 %.

2.7.2. Electrons and photons

The electromagnetic calorimeter measures the energies of electrons and photons. The main difference between them is the presence of a track in the inner detector in the case of the electrons. Although the ZH analysis did not use photons directly, their reconstruction is presented here together with that of electrons due to their similarity.

The reconstruction [38, 62] of both electrons and photons makes use of the high granularity of the electromagnetic calorimeter in the region $|\eta| < 2.47$. Calorimeter towers with a transverse energy, summed over the layers, of 2.5 GeV and size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ are the seeds to make clusters. A sliding-window of size 3×5 calorimeter towers in (η, ϕ) phase space searches for the seeds.

For the tracking, all the tracks with $\mathrm{p_T} > 500$ MeV are extrapolated to the second layer of the calorimeter. If the impact point of the extrapolated track is within a distance of 0.05 (0.1) in $\Delta\eta$ ($\Delta\phi$) from the cluster barycentre, that track is associated with the cluster. This looser criterium in $\Delta\phi$ minimises the effect of bremsstrahlung losses. If there is no association between the cluster and any track, that cluster is labelled as an unconverted photon. Both converted photons and electrons have the cluster associated with a track. These are separated by identifying a close-by track that might indicate the production of an electron-positron pair.

Photons are identified based on their properties using a selection of variables that describe the longitudinal and lateral electromagnetic shower development [63]. Two categories, *loose* and *tight*, are defined in the identification of photons. The core difference is that the *tight* category has a selection on the layer 1 of the electromagnetic calorimeter to distinguish photons from the π^0 decay. This decay is identified by two local maxima in this layer.

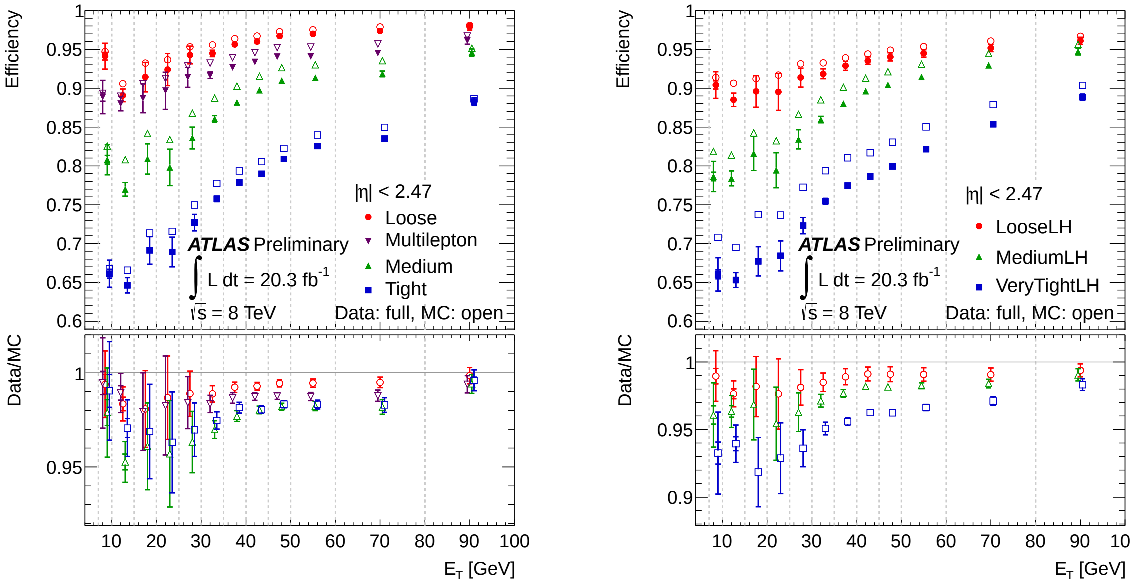
Electrons can be identified in a similar way using a cut-based scheme but applying also conditions on the track quality and the matching between the track and the cluster. For cut-based identified electrons, three categories, *loose*, *medium* and *tight* are defined with increasingly tight selection criteria. A fourth category, *multilepton*, is considered with similar identification efficiency as the *loose* but better background rejection. An alternative method for the electron identification uses a likelihood multivariate analysis. The same discriminating variables, used in the cut-based analysis, are chosen. There are also three likelihood categories defined: *looseLH*, *mediumLH* and *verytightLH*. These likelihood categories have better light-jet and photon conversion rejections.

Electrons and photons are indistinguishable outside the inner detector coverage and are not used in the ZH analysis.

2.7.2.1. Electron performance

Reconstruction and identification efficiencies Analogously to the muon case, the reconstruction and the identification efficiencies are estimated from data using the tag-and-probe method. The electrons that contribute to the ZH analysis have transverse energy higher than 25 GeV and their efficiencies use only the Z→ *ee* decay [64].

The *tag* electron is identified as a *tight* electron with $E_T > 25$ GeV. The *tag* electron is required to match an electron firing a trigger within a distance of 0.15 and to be outside the barrel-endcap transition region of the electromagnetic calorimeter ($1.37 < |\eta| < 1.52$). The *probe* electron requirements are to have a charge opposite to the *tag* electron, $E_T > 15$ GeV and that the invariant mass of the two electrons to be within 15 GeV of the Z boson mass.

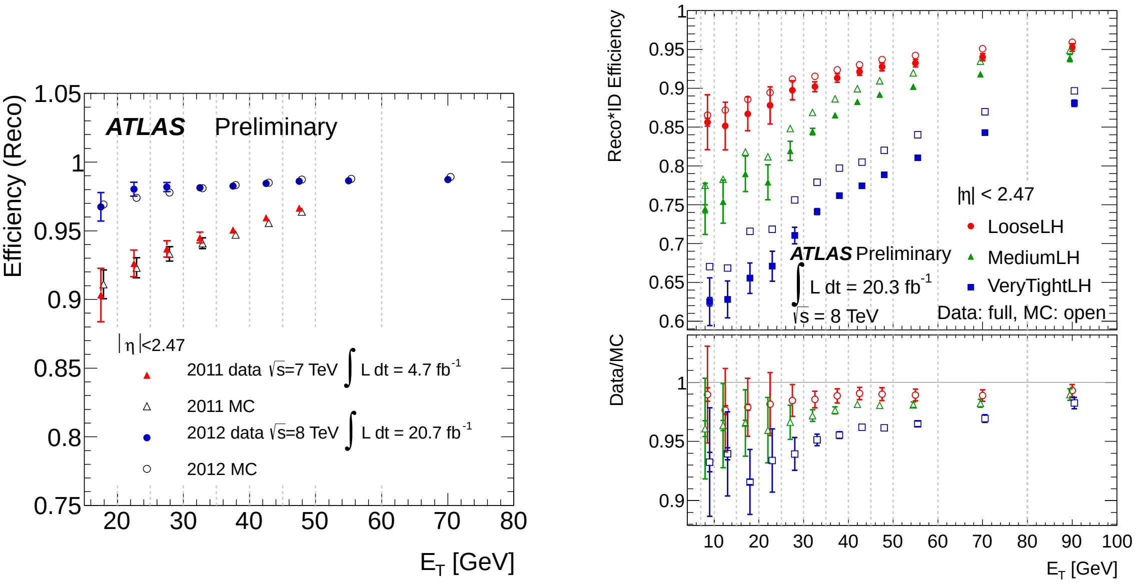


(a) (b)

Figure 2.23.: Electron identification efficiency in 2012 Z→ *ee* events as a function of E_T with $|\eta| < 2.47$ for data (full markers) and MC (open markers), using the cut-based (a) and the likelihood (b) identification methods [65].

The identification efficiency is calculated as the fraction of the *probe* electrons that pass each category (cut-based or likelihood) from all the *probe* electrons that have at least one pixel hit and at least seven silicon detector hits. Figure 2.23 shows the electron identification efficiencies obtained as a function of the electron E_T, for the cut-based (a) and the likelihood (b) alternatives. The three likelihood identification possibilities, *looseLH*, *mediumLH* and *verytightLH*, have roughly the same identification efficiencies as *multilepton*, *medium* and *tight* cut-based options. There was also a fourth likelihood option to match the *loose* cut-based efficiency which was called *verylooseLH* and was used in the ZH analysis. There is a clear dependence on E_T for data and MC, in particular for tighter categories. The data/MC ratio in the bottom panel shows differences to unity below 5 % within the statistical uncertainty for all categories. A scale factor is extracted from these ratios to correct the identification efficiency in simulation and is applied to each electron in the simulation.

The reconstruction efficiency for electrons use the same *tag* conditions as for the identification efficiency. However, the *probe* objects are now all electromagnetic clusters with $\mathrm{E_T} > 15$ GeV. The same invariant mass conditions are required, but no charge requirement is applied. The electron reconstruction efficiency is obtained from the fraction of the probe objects that match a track within a distance in η and ϕ smaller than 0.05 and 0.2, respectively.



(a)

(b)

Figure 2.24.: Data and MC comparisons of the electron reconstruction efficiency from 2011 and 2012 (a) and of the reconstruction-identification efficiency combination as a function of $\mathrm{E_T}$ (b) [65].

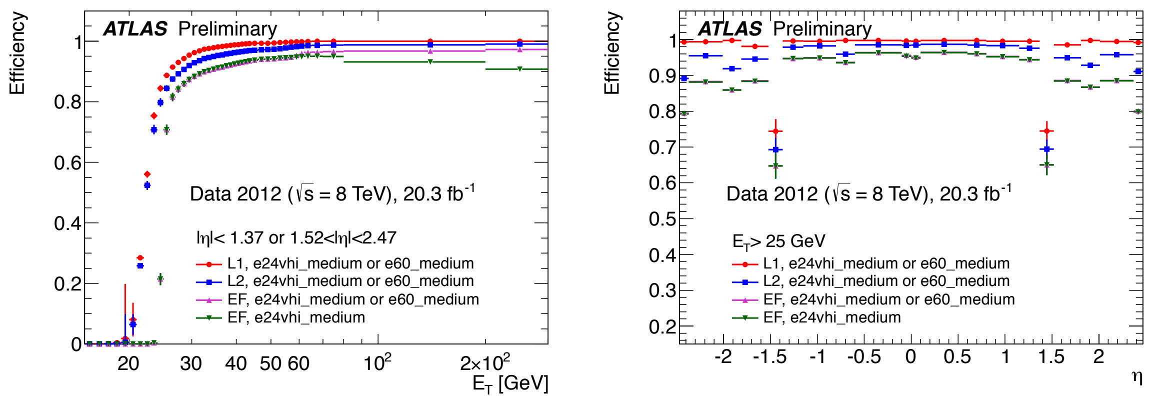
Figure 2.24 (a) shows the reconstruction efficiency as a function of $\mathrm{E_T}$, comparing data and MC for 2011 and 2012. The improvement in the efficiency between 2011 and 2012 is evident as well as the decrease of the dependence on $\mathrm{E_T}$ as the understanding of the detector improved. An excellent agreement between data and simulation is also observed. Figure 2.24 (b) shows the combination between the reconstruction efficiency and the identification efficiency with the likelihood method. The data-MC agreement, shown in the bottom panel, is better than 5 %.

The systematic uncertainties considered for the electron identification and reconstruction efficiencies are estimated as the RMS of the data/MC measurement [66]. Additional contributions account for the variation of the size of the mass window around the Z boson mass peak in Z$\rightarrow$ *ee* events, the requirement of a calorimeter isolation in a cone with $\Delta R = 0.4$ to have an energy deposition of less than 5 GeV, and the loosening of the electron identification criteria. If the RMS of the data/MC does not cover the differences observed, an additional systematic uncertainty is added to enclose them. The uncertainty on the reconstruction (identification) efficiencies are $0.5\,\%$ $(1-2\,\%)$ for electrons with $\mathrm{E_T} > 25$ GeV.

Trigger efficiency The trigger efficiency for electrons has been calculated, using the tag-and-probe method, for the logical OR between two triggers: e24vhi_medium1 and e60_medium1. Both triggers use the *medium1* identification properties described in reference [44]. The former trigger is the unprescale trigger with the lowest transverse momentum threshold, 24 GeV, and requires tracking isolation (see section 2.3.3) and no more than 1 GeV in the hadronic core region (see figure 2.11). The latter trigger requires an electron with $\mathrm{E_T} > 60$ GeV to increase the efficiency about this $\mathrm{E_T}$, as shown in figure 2.25 (a), by comparing the magenta up triangle to the green down triangles. Additionally, this figure shows the efficiencies for trigger L1 (blue squares) and L2 (red circles) as a function of the transverse momentum (a) and the pseudorapidity (b) of the electrons. The *tag* electron is the reconstructed one and the trigger electron used as *probe* has the additional requirement of having the opposite charge and the invariant mass of the two electrons in a window with size 30 GeV around the Z boson mass. The efficiency is obtained from the fraction of *probe* electrons that have a distance in the (η,ϕ) phase space to a trigger electron inferior to 0.15.

The systematic uncertainty sources considered in the trigger efficiency determination are similar to the ones used for the offline reconstruction and identification: the mass window of the Z boson and identification tag, both with tighter criteria [44].

Momentum scale and resolution The electron response scale and resolution are obtained from the ratio between reconstructed and truth energy in simulation using the processes $J/\psi\rightarrow ee$, Z$\rightarrow$ *ee* and $W\rightarrow e\nu$. The scale and resolution are obtained in the $1 < \mathrm{E_T} < 3000$ GeV range using multivariate analysis (MVA) [38]. The inputs for the MVA



(a) (b)

Figure 2.25.: Trigger efficiency obtained at the three levels in the early data for electrons as a function of $\mathrm{E_T}$ with $|\eta| < 2.47$ (excluding crack region) (a) and η position with $\mathrm{E_T} > 25$ GeV (b) [67].

are the energy measured in the calorimeter, the fraction between the presampler layer of the electromagnetic calorimeter (f_{em0}), the cluster barycentre position (η, ϕ) and the shower depth defined by equation 2.12.

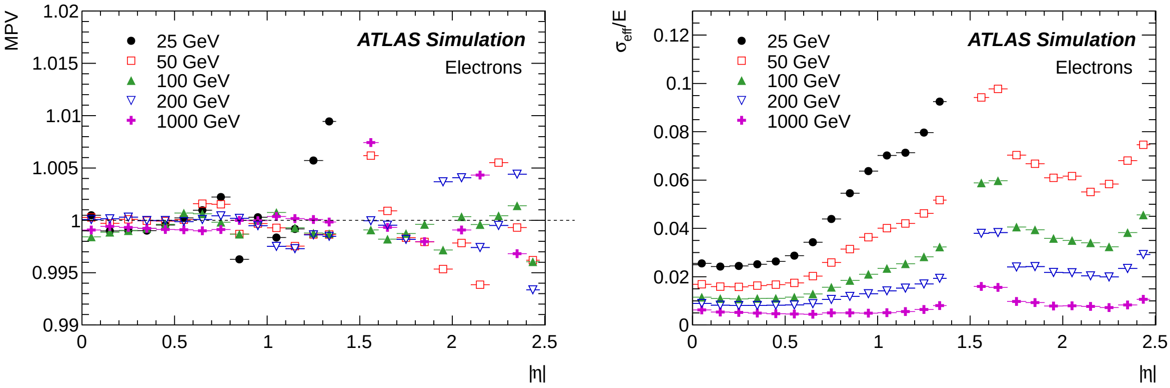
$$X = \frac{\sum\limits_i X_i\, E_i}{\sum\limits_i E_i} \tag{2.12}$$

where X_i is the calorimeter thickness in radiation lengths for the i^{th} calorimeter layer.

The electron energy response distribution is obtained for several bins of η and $\mathrm{E_T}$ and a Gaussian function is fitted to these distributions. The most probable value (MPV), given by the centre of the fit, is shown in figure 2.26 (a) as a function of the pseudorapidity. The electron energy resolution is calculated from the interquartile of the distribution and is shown in figure 2.26 (b).

The electron energy scale and resolution differ from data mainly due to some mismodelling on the material upstream before the calorimeter and on the calorimeter cell energy measurements. Corrections are derived from data-MC comparisons to account for this imperfect knowledge of the detector [38]. The energy scale comparison between data and MC is described by:

$$E^{\mathrm{data}} = E^{\mathrm{MC}}(1 + \alpha_i) \tag{2.13}$$



(a) (b)

Figure 2.26.: Electron energy scale most probable value (MPV) and energy resolution as a function of the pseudorapidity [38].

where i is a given bin of pseudorapidity, E^{data} is the energy measured in data, E^{MC} is the calibrated energy of the electron and the α constant is the *in situ* calibration factor. Z→ *ee* events are used to determine α, where the two electrons from the decay have η-bins i and j. The invariant mass distribution is obtained for every η_i, η_j combination in data and MC. The invariant mass data-MC comparison is given by:

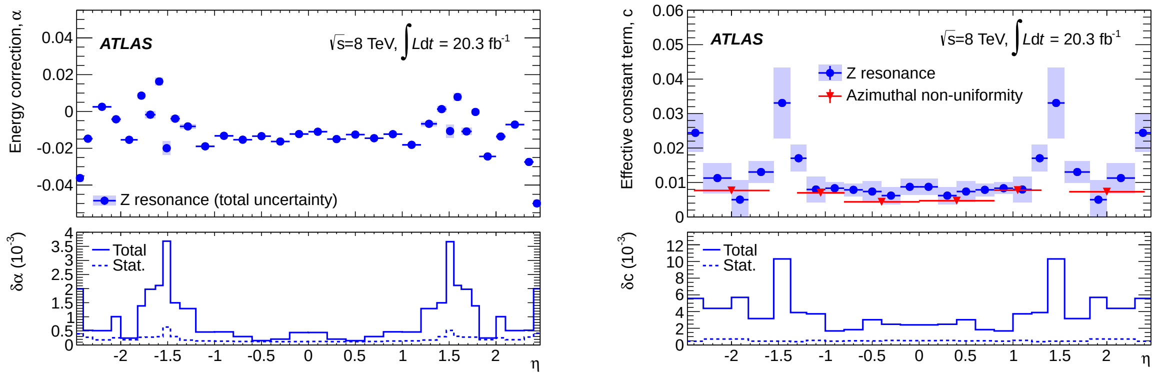
$$m_{ij}^{\text{data}} = m_{ij}^{\text{MC}} \left(1 + \frac{\alpha_i + \alpha_j}{2}\right) \tag{2.14}$$

The combination of the two α is obtained from a χ^2 minimization between the invariant mass distributions and the individual constants follow directly from solving the equation system. Figure 2.27(a) shows the value obtained for the energy correction, α, as a function of η. The bottom panels show the statistical and the total uncertainty, which is lower than 0.3×10^{-3} for $|\eta| < 1.47$, 2×10^{-3} for $1.47 < |\eta| < 1.82$ region (that includes the barrel-endcap discontinuity), and 5×10^{-3} for $|\eta| > 1.82$.

For the energy resolution of electrons, a similar procedure is applied and a "C_{ij}" term is added in quadrature to the resolution of the simulation:

$$\left(\frac{\sigma_m}{m}\right)_{ij}^{\text{data}} = \left(\frac{\sigma_m}{m}\right)_{ij}^{\text{MC}} \oplus \frac{c_i \oplus c_j}{2} \tag{2.15}$$

The electron energy resolution correction factor, c is shown in figure 2.27 (b) as a function of η. In average, the correction was 0.8 % for $|\eta| < 1.2$ with an uncertainty of 0.3 % and was 1 % in the endcap with an uncertainty of 0.5 %.



(a) (b)

Figure 2.27.: Correction parameters of the electron momenta for scale α (a) and resolution c (b). The top plots show the values of the parameters and the bottom panels give the total and the statistical uncertainty [38].

Figure 2.28 shows the effect of the scale and resolution corrections in the di-electron invariant mass distribution for Z→ ee events. The uncorrected MC, plotted with a dashed line, has differences of 5 % in the Z mass bin (90 − 92 GeV) and up to 8 % in the tails. After the scale and resolution correction, the data/MC agreement is within 1 % except in the low-mass tail (80 − 85 GeV) where the difference is 3 % pointing to some mismodelling in the MC.

The systematic uncertainties on the scale and resolution result mostly from the knowledge of the dead material that the electron needs to traverse, for low transverse energy electrons. For high E_T, the leading contributions are the layer calibration and the pedestal noise estimated from pileup events. For intermediate E_T values, these contributions are negligible, and the Z→ ee calibration dominates the uncertainty. For the region $|\eta| < 0.6$, electrons with transverse energies of 40 GeV and 200 GeV have energy uncertainties of 0.03 % and 0.27 %, respectively [38]. The green band in figure 2.28 shows the systematic uncertainty applied to the invariant mass of the two electrons.

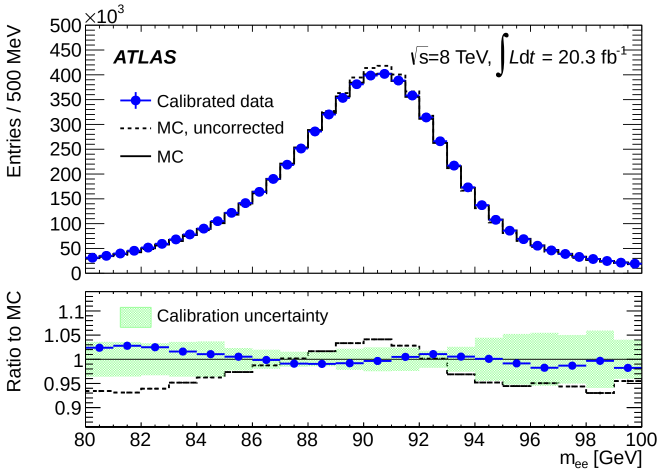


Figure 2.28.: Di-electron invariant mass distribution of Z→ ee candidate events reconstructed for data (blue points) and MC simulation with (black line) and without (black dash line) the scale and resolution correction applied. The data/MC comparison is shown below with the respective calibration uncertainty (green band) for corrected MC [38].

2.7.3. Taus

Taus were not used in the analysis presented in this thesis. Therefore, they are only briefly described here. A tau has a decay time of 290.3 fs ($c\tau \approx 87.11\ \mu$m) [4], which means it can travel a couple of millimetres in the detector. Either it decays into a lepton (electron or muon) and neutrinos, with a 35.24 % fraction, or it decays hadronically. In the first case, the lepton is reconstructed from the electron or muon tracks, while in the latter is reconstructed as a jet [2]. The reconstruction and calibration of taus are described in reference [68].

2.7.4. Neutrinos and missing transverse energy

Neutrinos are the only particles that leave no information in any of the sub-detectors of ATLAS. The only information, which can be extracted from the neutrino, comes from the conservation of the momentum/energy before and after the collision. Since the beams collide head-on, there is no transverse momentum from the beam before the collision. In the collision that creates a neutrino, it leaves the detector without interacting, and there will be a momentum imbalance in the transverse plane.

---

[2] Jets result from the reconstruction of the hadronization of quarks and gluons and is explained in the section 2.7.5.

The energy-momentum Lorentz vector is obtained for the identified particles that leave energy in the calorimeter based on the calorimeter energy, direction (position on the detector) and mass. These particles are electrons, photons, hadronic taus and jets. Additionally, there are energy depositions in the calorimeter that do not belong to any reconstructed particle. These *soft* energy deposits come from low energy particles that do not pass the reconstruction threshold. Finally, muons deposit a small amount of their energy in the calorimeter and only very low energy muons are stopped in the calorimeter volume. Their contribution is obtained directly from the momentum measured in the inner detector for isolated muons, discarding the energy deposited in the calorimeter. Non-isolated muons use only the momentum measurement from the muon spectrometer and the calorimeter term of the muon is included in the physics object reconstructed (usually jets). The information of the neutrino is referred to as missing transverse energy [69, 70] and is obtained from the sum of the energy-momentum of the contributions just referred, taking its projection onto the transverse plane and inverting the direction of the vector, as expressed in equation 2.16.

$$\vec{\not{E}}_{\mathrm{T}} = -\sum_{i}^{N_{obj}} \vec{E}_{\mathrm{T},i} = -\sum \vec{E}_{\mathrm{T}}^{e} - \sum \vec{E}_{\mathrm{T}}^{\gamma} - \sum \vec{E}_{\mathrm{T}}^{\tau} - \sum \vec{E}_{\mathrm{T}}^{\mathrm{jets}} - \sum \vec{E}_{\mathrm{T}}^{\mathrm{soft}} - \sum \vec{p}_{\mathrm{T}}^{\mu} \tag{2.16}$$

Figure 2.29 shows the missing transverse energy for events with two muons, with reconstructed invariant mass between 66 and 116 GeV for data and simulation. The bottom panel shows the data/MC ratio. Differences from the unity of about 5 % are found for missing transverse energy up to 60 GeV and up to 20 % after. Missing transverse energy is used in the ZH analysis, selecting events with small $\not{E}_{\mathrm{T}}$ as it is expected in the $ZH \to \ell\ell b\bar{b}$ process (see section 4.5.3).

2.7.5. Jets

Quarks and gluons produced in the hard scattering process can not exist freely in nature because they have a non-zero colour quantum number. The hadronization is a model that physicists use to explain and to mimic how the quarks and gluons transform into hadrons and will be discussed further in section 2.8. After the hadronization, there will be a set of particles emitted in the same direction, to which the denomination "jet" is given. Jets are the best tool to probe the quark or gluon resulting from the hard scattering, assuming that the

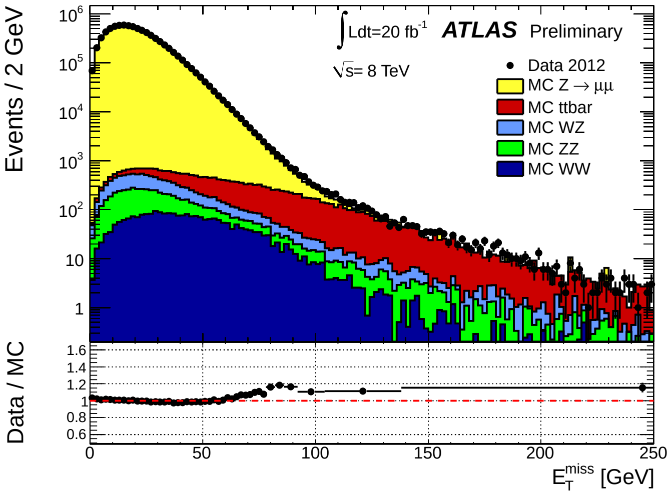


Figure 2.29.: Missing transverse energy distribution for $Z \rightarrow \mu\mu$ event [70].

energy and direction of the jet are related to the energy and direction of the initial particle.

Jets [71] can be defined at different levels: at parton level, at particle level and calorimeter level. A sketch of a proton-proton collision is depicted in figure 2.30. Parton level jets are composed of the quarks or gluons resulting from the hard scattering process hadronization. Jets at particle level consist of stable particles (half-life larger than 33.4 ps) after the hadronization process takes place. The information from these two levels can only be accessed at truth level, *i.e.* in simulated samples. Calorimeter level jets are composed of the energy depositions in the calorimeter. Since non-simulated events have only calorimeter level jets, it is necessary to compare these with those at truth level to probe the initial quarks or gluons. Jets can also be obtained from the combination of tracks but since there is no track from neutral particles in this way, the energy carried by neutral particles will not be measured correctly.

2.7.5.1. Jet finding algorithms

Jets are reconstructed using jet algorithms which combine either the quarks and gluons, or the stable particles at the generator level, or the calorimeter energy clusters. A good jet algorithm should provide almost identical results at each of the three levels and be relatively fast. The former requirement is assessed by evaluating if the algorithm is collinear and infrared safe. The latter was optimised since the FastJet program [73] was introduced to speed up jet reconstruction. There are two types of jet algorithms: cone and sequential

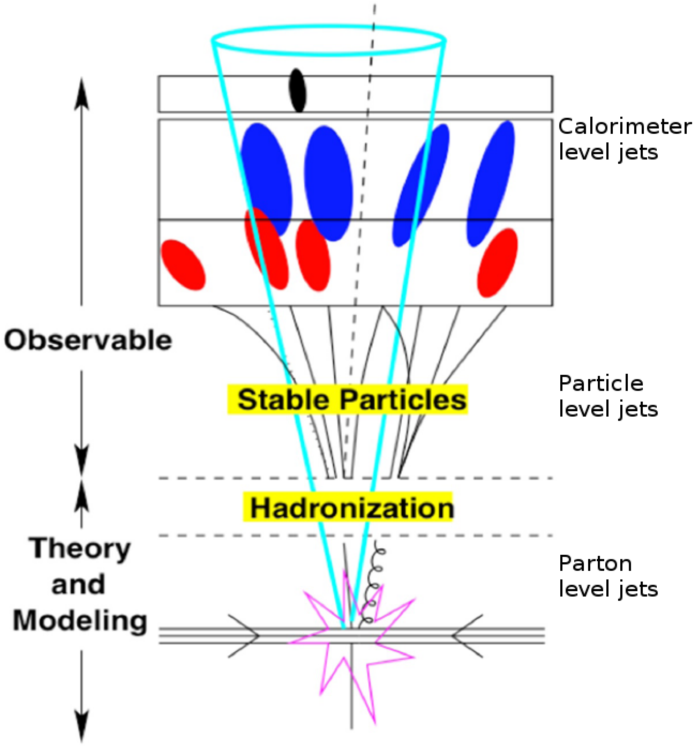


Figure 2.30.: Sketch of a collision showing the different types of jets. Adapted from reference [72].

clustering. Most cone algorithms require a seed, *i.e.* an energy deposit above a certain threshold, to start the iteration. These seeded algorithms reconstruct different jets if a particle deposited its energy in a single (with seed) or two collinear (no seed) calorimeter cells. Therefore, the algorithm is said to be collinear unsafe. Additionally, jet algorithms should be infrared safe, which is tested by adding additional soft radiation to the event and checking if there is impact on the algorithm performance. In algorithms which are not infrared safe, such as the seeded cone algorithms, it becomes impossible to make perturbative QCD calculations, and jets become useless because the three levels produce different results. From the cone algorithms, only the seedless SISCone algorithm [74] is both infrared and collinear safe. All sequential clustering algorithms are infrared and collinear safe and, for this reason, they are the most used.

The sequential clustering algorithms try to recombine two jet constituents (calorimeter clusters or particles) into one until reaching the final jets that correspond to the initial quark or gluon. Figure 2.31 shows a fluxogram of the sequential clustering method to find jets. Starting with a list of N jet constituents, the algorithm looks first if the list is currently empty.

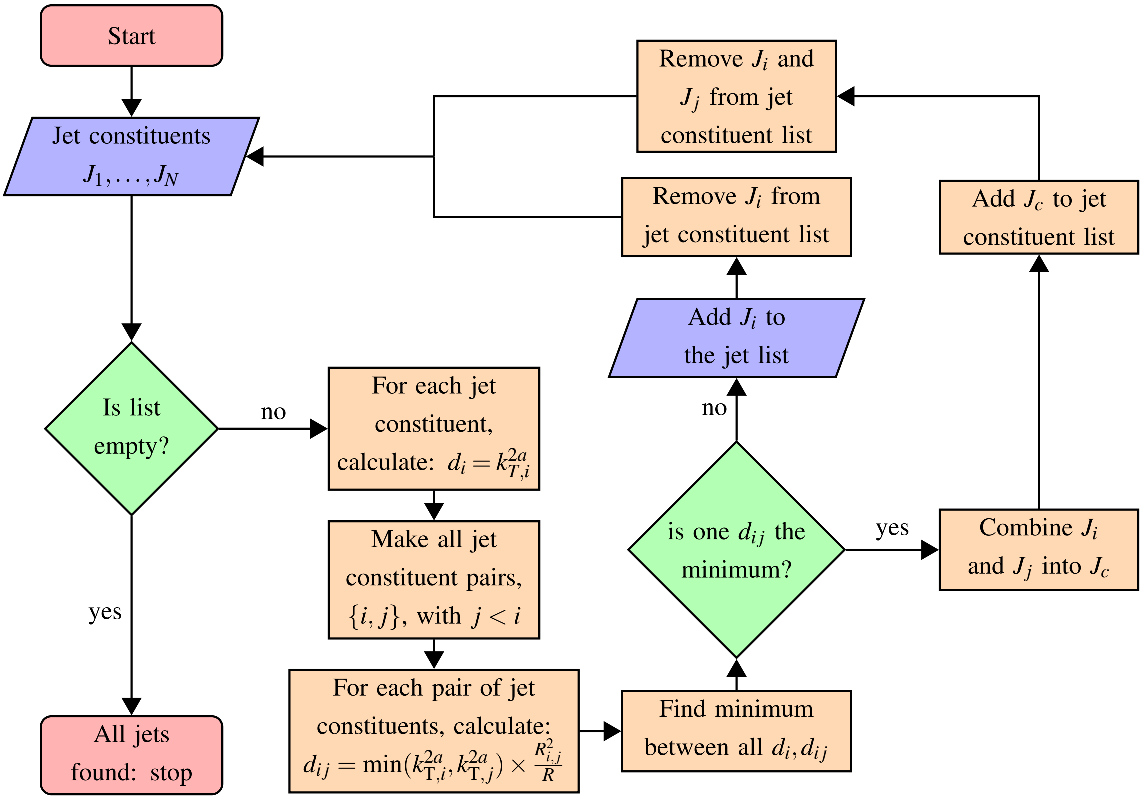


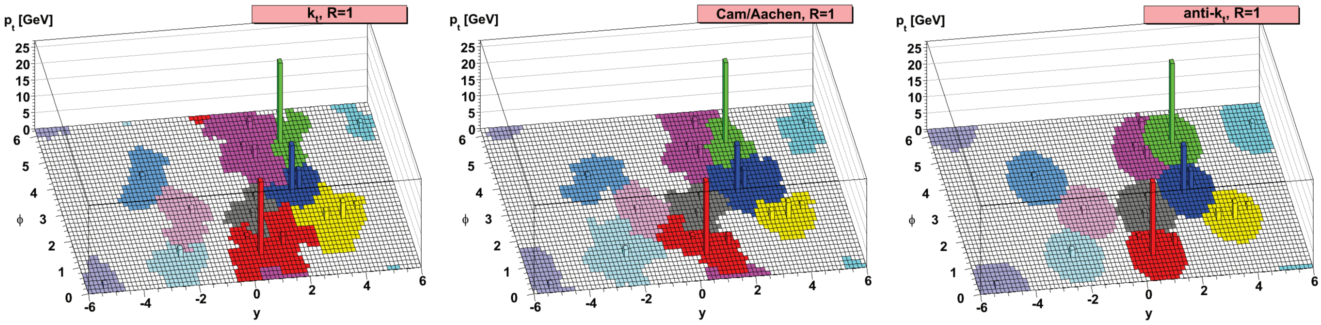
Figure 2.31.: Fluxogram of the sequential clustering method to reconstruct jets.

If it is not, the “distances” $d_{i,j}$ and d_i are evaluated from equation 2.17.

$$d_{ij} = \min(k_{\mathrm{T},i}^{2a}, k_{\mathrm{T},j}^{2a}) \times \frac{R_{i,j}^2}{R} \qquad\qquad d_i = k_{\mathrm{T},i}^{2a} \tag{2.17}$$

The a parameter is algorithm-specific and $R_{i,j}$ is the distance in the (η, ϕ) phase space between the two jet constituents. The R parameter is what defines the size of the jet. A large R value ensures that the jet will contain all the particles which were created by the initial parton but may also contain contributions from pileup and the underlying event. A small value of R is less sensitive to these effects but might not contain all the jet energy. The parameter k_T can be either $\mathrm{E_T}$ or $\mathrm{p_T}$ of the jet constituent. The distances $d_{i,j}$ and d_i are evaluated for every pair of jet constituents $\{i, j\}$ and for every jet constituent i, respectively. If the distance between two jet constituents $d_{i,j}$ is the smallest distance, those two constituents, i and j, are combined. This combination consists of determining the centre of the new jet constituent, as the weighted sum of the two primary jets constituents. In this case, the new combined jet constituent is added to the list, and the two input jet constituents are removed from the list. If the minimum is one d_i instead, this jet constituent is considered to be one jet and is removed from the list. The process will continue until the list of jet constituents is empty, and the clustering algorithm finishes with all jets found.

The three main sequential clustering jet algorithms are k_T $(a=1)$, Cambridge/Aachen $(a=0)$ and anti-k_T $(a=-1)$ [75]. The k_T algorithm combines first the softer elements. Figure 2.32 (a) shows one simulated event with this algorithm indicating that the found jets fluctuate considerably in size and shape. The Cambridge/Aachen algorithm has no dependence on the energy of the elements, and simply merges close-by constituents until no distance is larger than R. The same event is shown in figure 2.32 (b) with this algorithm. It is favourable when probing the sub-structure of a jet as it only uses the distance in the (η, ϕ) phase space. Finally, the anti-k_T algorithm will combine first the harder particles. It will result in cone-shaped jets, circles in the (η, ϕ) phase space as observed in figure 2.32 (c). Although this combination scheme is not adequate to study the substructure of jets, the distinct areas are the best for calibration purposes.



(a) k_T algorithm (b) Cambridge/Aachen algorithm (c) anti-k_T algorithm

Figure 2.32.: Reconstruction of the jets of an event using three sequential clustering algorithms [75].

2.7.5.2. Jet reconstruction and calibration in ATLAS

The ATLAS experiment uses the anti-k_T algorithm with an R parameter of either 0.4 or 0.6, depending on the physics searches involved. Also used by the experiment are large-R jets ($R = 1.0$) when studying boosted regime. The ZH analysis uses only $R = 0.4$ and so does the work developed for this thesis.

The ATLAS calorimeter is not compensated, which means that the energy response to hadrons is not the same as to electrons. Therefore, it is required to calibrate the energy of the calorimeter jets. There are three possible calibration levels of the calorimeter. The baseline calibration uses events from $Z \rightarrow ee$ decay to obtain the electromagnetic energy scale. These calorimeter jets are then calibrated at the hadronic scale, known as EM+JES, to correct for the calorimeter non-compensation. The derivation of the jet energy scale will be extensively covered in chapter 3. The third calibration level is applied to the topoclusters before running the jet finding algorithm. The topoclusters are divided into electromagnetic or hadronic clusters, depending on some shower shape variables and the energy density in the two calorimeters. This calibration also tries to correct for the lost energy around the cluster due to the noise thresholds and in the non-sensitive regions of the detector. It is known as the local hadronic cluster weighting calibration (LCW). Similarly, this calibration is improved further by applying a specific jet energy calibration, obtaining LCW+JES. ATLAS also uses the LCW based calibration, but since neither the ZH analysis nor the calibration studies documented in this thesis used it, it is not discussed further.

Most analysis are not interested in jets other than those resulting from the leading interaction. The ZH analysis used a pileup suppression method known as jet vertex fraction (JVF). This method associates tracks to jets and calculates the fraction of the scalar sum of transverse momentum of these tracks inside the jet that belong to the leading primary vertex [76]. If this fraction is below 0.5, the jet is considered as a pileup jet and will not be included in the analysis. The systematic uncertainty associated to this method is evaluated by changing the threshold by about 5 % and the effect will be studied in the analysis.

2.7.5.3. Jet categorisation

Jets in simulation can be divide into four categories:

b-jet If any B-hadron is found within a distance of 0.4 from a jet axis, that jet is considered to be a b-jet.

c-jet If the jet is not a b-jet and there is at least one C-hadron within a distance of 0.4 from the jet axis, that jet is considered to be a c-jet.

τ jet If the jet is neither a b-jet nor a c-jet and a τ lepton is found within a distance of 0.4 from the jet axis, the jet is considered to be a hadronic τ jet. These are not considered in the ZH analysis.

light jet If none of the previous conditions is verified, the jet is considered to be a light jet.

The latter category is further divided into light-quark or gluon initiated jet by checking what is the particle (quark or gluon) that has the largest transverse momentum in the jet.

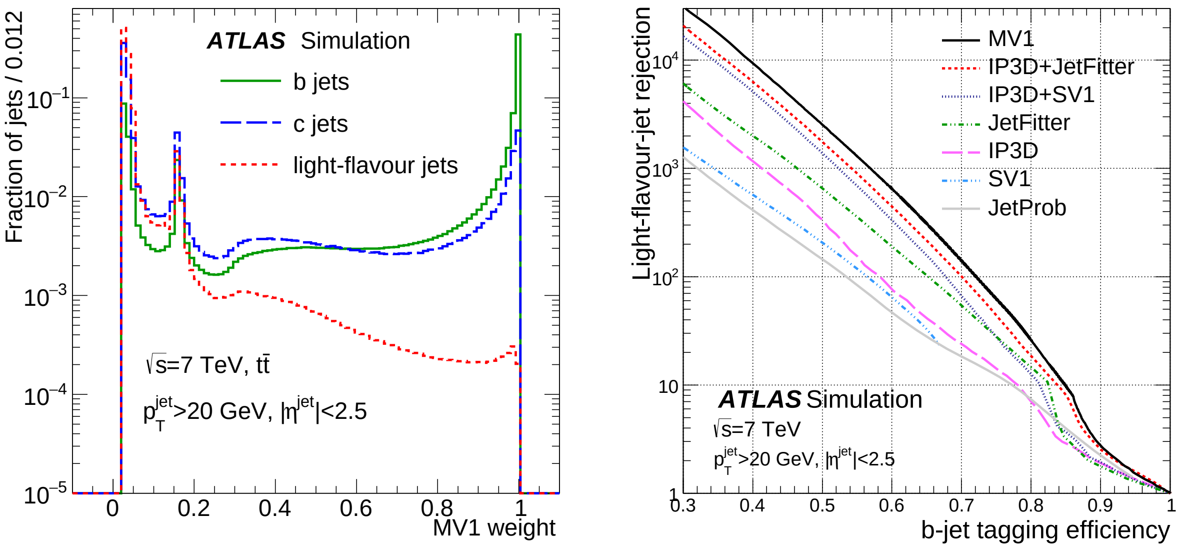
2.7.6. Identification of b-jets

The identification of b-jets in ATLAS is based on the large lifetime of B-hadrons, of the order of 1.5 ps ($c\tau \approx 450\mu$m). A B-hadron with $p_T = 45$ GeV may travel about 4 mm in the transverse plane before decaying. There are two classes of b-tagging algorithms that explore this characteristic of b-jets: impact parameter based, or secondary vertex based. A better description of each can be found in references [77, 78]. The combination of both classes of algorithms results on a better discriminator to identify b-jets.

Impact parameter based algorithms use the significance of the track impact parameters, defined as the ratio between the impact parameter value and its reconstruction uncertainty, as inputs for a log-likelihood ratio formalism to define the probability that the track originates from the primary vertex. A sign is given to the impact parameter depending on whether it is located along the jet direction or in the opposite direction. Since the B-hadron that produced the b-jet travels before decaying, the average sign of the impact parameter will be positive for tracks in the b-jet, whereas light flavour jets have a more evenly distribution.

Vertex based algorithms in ATLAS use either a log-likelihood ratio as previously or a neural network. The variables used for these algorithms are:

- the invariant mass calculated from the charged particles, represented by the reconstructed tracks, in the secondary vertex;
- the ratio between the momentum of the charged particles in the secondary vertex and the total momentum of those that are matched with the jet;
- the number of vertices that have exactly two charged particles associated;
- the number of vertices that have two or more charged particles associated;
- the total number of charged particles in the vertices referred in the previous bullet;
- the number of vertices, with one reconstructed track, in the flight direction of the B-hadron;
- the flight-length significance, defined as the distance between the primary and the secondary vertices divided by its uncertainty;
- Angular distance, ΔR, between the flight direction and the jet axis.

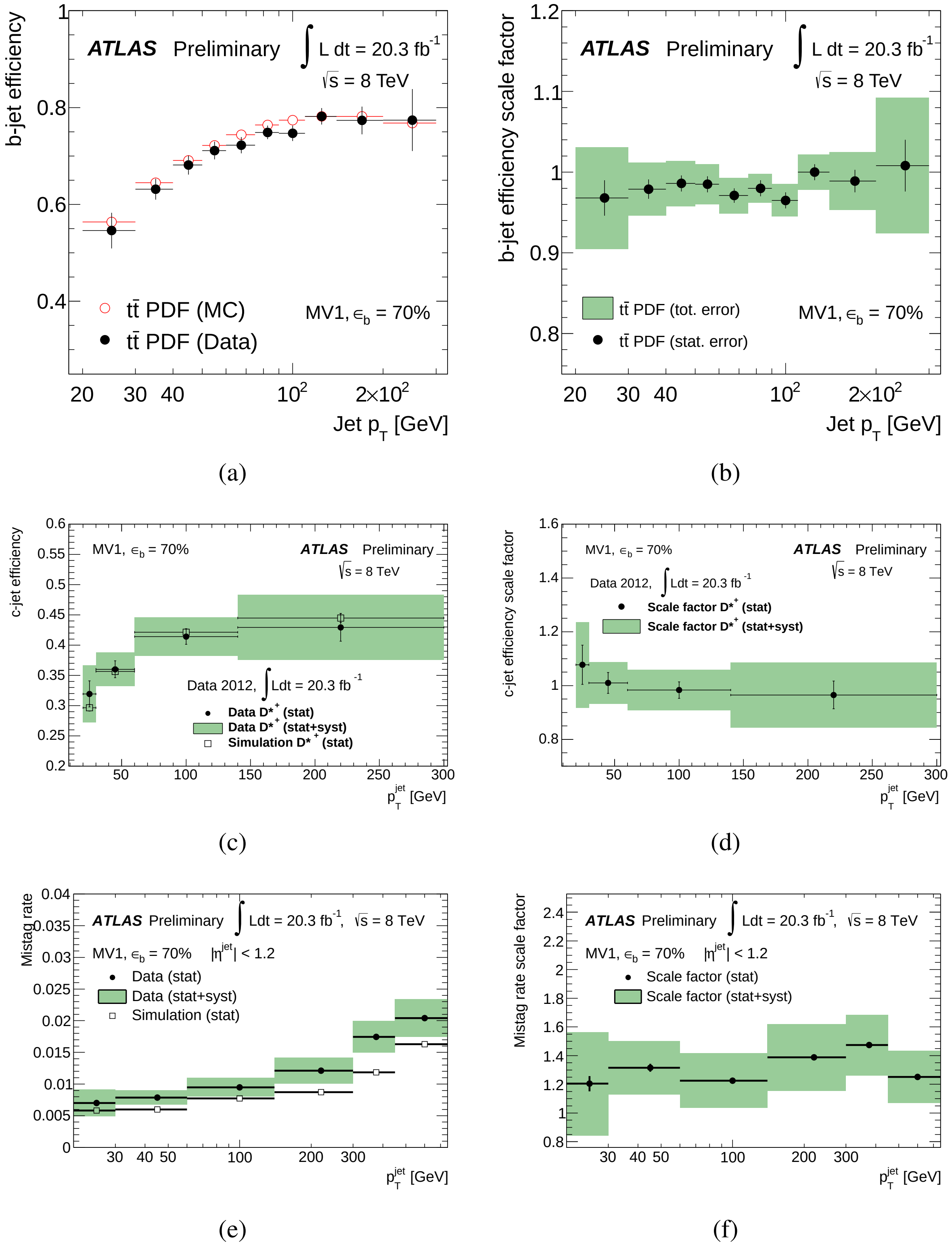


(a) (b)

Figure 2.33.: Performance of the MV1 b-tagging algorithm [78]. Distribution of the fraction of b-jets and c-jets and light flavour jets as a function of the MV1 weight (a). Rejection of light flavour jet as a function of the b-jet efficiency for some b-tagging algorithms (b).

The MV1 tagger algorithm [80] uses a neural network with inputs from both classes of algorithms and gives a final result, referred as the MV1 weight. Figure 2.33 (a) shows the fractions of b-jets, c-jets and light flavour jets as a function of the MV1 weight for a simulated $t\bar{t}$ sample. A jet is b-tagged if it has the MV1 weight above a particular value. The threshold is selected to accept a fraction of b-tagged jets, referred to as the operating point, and rejecting a corresponding amount of c-jets and light flavour jets. Figure 2.33 (b) shows the light-flavour jet rejection as a function of the b-tagging efficiency for impact parameter based algorithms (JetProb and IP3D), vertex based algorithms (SV1 and JetFitter), some impact parameter-vertex algorithm combinations, and finally for the MV1 tagger. The operating point of the MV1 algorithm must be chosen as a compromise between the b-tag efficiency and the light flavour jet rejection.

It is necessary to understand how adequately the simulated efficiencies correspond to reality, through comparisons between simulation and data. The methods for obtaining the efficiencies are detailed in references [78, 79, 80]. The efficiencies obtained for data and simulation as a function of the jet p_T are shown in figure 2.34 for b-jets (a), c-jets (c) and



(a) (b)

(c) (d)

(e) (f)

Figure 2.34.: Measured b-tagging efficiencies (left) and scale factors (right) obtained for b-jets (top), c-jets (middle) and light flavoured jets (bottom) for data and MC [79, 80].

light flavoured jets (e), for the 70 % operating point. The light flavoured jet efficiency is referred to as the mistag rate. The efficiencies were obtained in simulation: for b-jets using a $t\bar{t}$ simulated sample with a leptonic filter, for c-jets using a di-jet sample, filtering events with the decay of the D^{*+} meson, and for light-flavoured jets using also the same di-jet sample but without this filter. The efficiencies are then validated comparing the efficiency measured in data with the one for MC. Data events are selected by a leptonic trigger for b-jets and a jet trigger for the remaining. Scale factors are derived from data/MC ratios and are applied at analysis level to correct the simulations, which are shown on the right side of the figure 2.34 for the corresponding jets.

The systematic uncertainties shown by green bands around the calculated scale factors (black markers) include contributions from the systematic uncertainties from the detector resolution, from the methods adopted, and from the statistical uncertainties in each bin. The systematic uncertainty in the simulation contains the modelling and normalisation of signal and background samples when assessing the b-jet efficiency calculation and other contributions. For the detector uncertainties, the most notable are the resolution of the impact parameters required to obtain their significances. The uncertainty on the method is most important in the c-jet case since it has a very complicated method for extrapolating the efficiency for c-jets with the $D^{*+} \to D^0\pi^+$ decay to an inclusive set of c-jets.

Name	Operating point	MV1c weight	Efficiency (%)	Rejection efficiency (%)	
			b-jet	c-jet	light-jet
Loose	80	0.4050	79.85	67.11	96.57
Medium	70	0.7028	70.00	81.27	99.26
Tight	50	0.9237	49.99	96.19	99.93

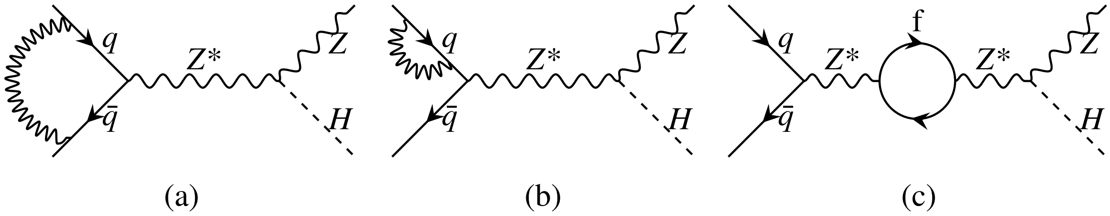
Table 2.1.: Summary of the MV1c properties for the three operating points. The rejection efficiencies give the fraction, in percentage, of the c-jets and light flavour jets not accepted as a b-jet [81].

The MV1 b-tagging algorithm was optimised for a better rejection of c-jets (MV1c), with a smaller rejection of light flavour jets. The MV1c algorithm was implemented in the ZH analysis [82], using three operating points, loose, medium and tight, with b-jet efficiencies near 80 %, 70 % and 50 %, respectively. Table 2.1 summarises the efficiencies for each operating point, with the respective light and c-quark rejections and the weight value.

2.8. Monte Carlo simulation

Deep inelastic scattering experiments discovered that the proton is not an elementary particle and has an internal structure with three valence quarks [83]. The strong interaction of quarks inside the proton is mediated by gluons. The momentum distributions of quarks and gluons inside the proton, referred to as partons, are described by various published parton distribution functions (PDFs). The main PDF sets used for this thesis were CT10 [84] and CTEQ6L1 [85].

A Monte Carlo (MC) event generator [86, 87] uses the PDF sets to select the momentum of the particles participating in the hard scattering process and to generate the events based on some properties distributions. Additionally, it can use the quantum mechanics matrix element, $\mathscr{M}^2$, integrated in the process phase space to determine the cross section, σ, for the simulated process at leading order (LO). The three most complete event generators are PYTHIA [88, 89], HERWIG [90, 91] and SHERPA [92]. In this thesis, the main parton shower event generator versions used for the ZH analysis were: PYTHIA-6, PYTHIA-8, SHERPA-1.4.1, HERWIG++.



(a) (b) (c)

Figure 2.35.: Next-to-leading order in electroweak Feynman diagrams of the ZH production mechanism.

Higher order calculation results from adding particle loops to the tree-level Feynman diagrams. Next-to-leading-order (NLO) electroweak (EW) corrections results from adding a virtual photon interaction between two electric charged particles or in the same particle. Two possible NLO EW corrections for the ZH production are shown in the Feynman diagrams on figure 2.35. Figure 2.35 (c) shows additionally a fermionic loop in the offshell Z boson. NLO QCD corrections results from adding to the Feynman diagrams a gluon interaction, instead of the photon, to colour charged particles and adding a quark or a gluon loop to a gluon. Because of the gluon self quartic coupling, the gluon loop can also begin and end in the same vertex. Analogous considerations are made for next-to-next-to-leading-order (NNLO), and so on. The higher order generator referred in this thesis is POWHEG [93]. Also used in the thesis are the ALPGEN [94] and ACERMC [95] generators, obtained at leading-order but with multileg (more parton shower). There are also specific programs that provide higher order calculations for particular processes like FEWZ [96] used for the calculation of the cross section production of Z and W bosons, TOP++ [97] for the $t\bar{t}$ process, or HAWK [98] for the $ZH \rightarrow \ell\ell b\bar{b}$ process. Alternatively, higher order corrections can also be considered to leading order generators by multiplying their cross section by some tabulated correction factor. Such is the case of the SHERPA simulations used in this thesis with a correction multiplicative factor of 1.12 (1.10) for Z+jet (W+jet) process.

Event generators simulate additional parton shower in the event by splitting out-going quarks or gluons. A gluon can split into two quarks, referred as gluon-splitting. All electric charged particles can radiate photons and all coloured charged particles, quarks and gluons, can radiate gluons. The term initial (final) state radiation is usually called to additional photons or gluons that are produced in the initial (final) state of the hard scattering process. The PHOTOS [99] program is used to address the final state radiation from QED. No quark or gluon can exist freely in nature because they have colour charge quantum number different from zero and have to be recombined. This recombination is referred to as hadronization. Event generators use two alternative hadronization models based on some properties of QCD. The Lund string model [100], used by the PYTHIA generator, is based on the direct relation between the potential energy of quarks or gluons and the distance between them. The cluster string model, used by the HERWIG and SHERPA generators, is based on partons in a shower to be clustered into colour singlet groups. In this process, called preconfinement [101], gluons are split into quarks before clusters are formed.

The so-called hard scattering interaction involves only two of the partons in the proton. Other partons also participate in the collision and the interaction of these proton remnants needs to be simulated as well. For that, several underlying event (UE) are used by the event generator and tuned to collision data. The underlying event is simulated before the hadronization since some times it is necessary to obtain an overall neutral colour quantum number from outgoing particles. Two main independent UE tunes are used for the processes addressed in this thesis: the ATLAS underlying event tune 2 (AUET2) [102, 103] and the PERUGIA tune P2011C [104], which was adapted to the PYTHIA6 generator and the CTEQ6L1 PDF set. SHERPA handles the underlying event internally.

The final event generation step takes care of all unstable particles before entering the detector simulation. All hadrons with a half life smaller than 33.4 ps are forced to decay using the default generators. Other programs can be used for handling the decay of particular particles. Examples of this used in this thesis are the TAUOLA [105] that handles the decay of τ-leptons and HDECAY [106] that handles the decay of the Higgs boson. The latter includes the Standard Model decays to fermions, gluons, $\gamma\gamma$, $Z\gamma$, WW, ZZ and also some beyond the Standard Model decay possibilities like supersymmetric particles.

On top of the leading proton-proton collision, other softer collisions occur, as discussed in section 2.4. The pileup is simulated through the superposition of minimum-bias interactions simulated with PYTHIA8 and using the MSTW2008LO PDF set [107] and the AUET2 tune.

After passing this last stage, particles are called truth particles in this thesis. Despite this, they retain the information of which non-stable particle they came from. This allows to apply filters to select if specific particles were created during the hadronization process. This filtering can happen during event generation or be used at analysis level. This thesis uses particular filtering for B-hadrons and C-hadrons. The particles considered for each filter are presented in table 2.2 with the MC particle identification, the symbol and the leading quark composition. This filtering helps to increase the statistics for rarer events. Another filter used was the production of separated samples for different transverse momentum windows of vector boson in a V+jets sample and of leading jet in di-jet sample. Another filter considered selects leptonic decays from W or Z bosons.

Bottom hadrons						Charmed hadrons					
Mesons			Baryons			Mesons			Baryons		
MCID	Symbol	qq	MCID	Symbol	qqq	MCID	Symbol	qq	MCID	Symbol	qqq
511	B^+	$u\bar{b}$	5122	Λ_B^0	udb	411	D^+	$c\bar{d}$	4122	Λ_C^+	udc
521	B^0	$d\bar{b}$	5132	Ξ_B^-	dsb	421	D^0	$c\bar{u}$	4132	Ξ_C^0	dsc
531	B_s^0	$s\bar{b}$	5232	Ξ_B^0	usb	431	D_s^+	$c\bar{s}$	4232	Ξ_C^+	usc
541	B_c^+	$c\bar{b}$	5112	Σ_b^-	ddb				4212	$\Xi_C^{\prime 0}$	dsc
			5212	Σ_b^0	udb				4322	$\Xi_C^{\prime +}$	usc
			5222	Σ_b^+	ddb				4332	Ω_C^0	ssc
			5332	Ω_B^-	ssb						

Table 2.2.: Bottom and charmed hadrons selected by the filter present in the SHERPA V+jets backgrounds.

Once particles leave the event generator, it is necessary to simulate how they interact with the detector. The ATLAS sub-detectors geometry are simulated with the GEANT4 tool-kit [108, 109]. Additionally, GEANT4 is also responsible for the interactions between particles and the detector. Particularly, for the calorimeter, there are several tools to simulate this interaction called *physics lists* [109, 110]. This thesis considers two physics lists: QGSP_BERT and FTFP_BERT. In the QGSP_BERT physics list, the final state of pions, kaons, protons and neutrons is modelled with the quark-gluon string model, followed by precompound and evaporation models for energies above 12 GeV, with the low energy parametrized model (LEP) for energies between 9.5 and 25 GeV and the Bertini intra-nuclear model below 9.9 GeV. Smooth transitions are applied between models in the overlap energy regions. The FTFP_BERT physics list uses the FRITIOF model with precompound/evaporation models for energies above 4 GeV and the Bertini model below 5 GeV, for the final state of the same particles.

Simulations where GEANT4 is used to describe the whole detector are referred as “full simulations” and contain the best description of the ATLAS detector. The simulation time for the propagation of particles in the calorimeter with GEANT4 accounts for almost 80 % of the total time of the detector simulation time and is dominated by the simulation of electromagnetic particles [111]. For this reason, the much faster FASTCALOSIM was used to parametrize the longitudinal and lateral energy profile in (η, ϕ) bins corresponding to the granularity of the calorimeter cells. Several simulation samples, considered in this thesis, used this “fast simulation”, called ATLFAST-II. This detector simulation reduced the simulation time by a factor of $20-40$ [111] and allowed for simulated process samples with larger statistics.

The event, after passing the detector simulation, contains the cell energy depositions and the tracking hits. The L1 trigger system, discussed on section 2.3, is also simulated. After this step, the event is again in the RDO format, and the reconstruction will take place in the same way as for data (see section 2.5).

3. Jet calibration and performance

> “. . . a single drop in a limitless ocean. | What is an ocean but a multitude of drops?”
>
> *(David Mitchell, “Cloud ATLAS”)*

3.1. Introduction

The properties of quarks and gluons are very complicated to measure because of the hadronization process, discussed in section 2.8. LHC is a hadronic collider and the outcome of nearly every collision will have these elementary particles which hadronize into a set of stable hadrons. The latter interact in the detector and it is their energy and properties that are measured. The final state particles have to be combined to measure the properties of the initial quark or gluon. The jet algorithm anti-k_T, described in section 2.7.5.1, combines these particles to access the information of the initial quarks and gluons.

3.1.1. Motivation

There are many different effects that affect the measurement of the jet energy, that will require dedicated calibrations to correct for them. The ATLAS calorimeter is a non-compensating calorimeter, meaning that, in a hadronic shower, the energy response to the electromagnetic component of the shower is different from the hadronic component. Despite the fact that the majority of particles in the shower are hadrons, the hadronic cascade has an electromagnetic component mainly from neutral pion decays, $\pi^0 \to \gamma\gamma$. However, some of the initial hadron energy is lost in nuclear binding energy. There are also energy losses due to spallation of slow nucleons and production of muons and neutrinos that escape the calorimeter. It may also happen that very energetic jets lose part of their energy because the calorimeter volume cannot contain the jet, which is usually referred as leakage of the calorimeter. Other energy losses result from the absorption in non-instrumented regions of the detector such as the cabling system or the cryostat between the two calorimeters. Additionally, since the anti-k_T algorithm reconstructs jets with a finite size, R, some of the energy of the original quark or gluon might be deposited outside of the calorimeter reconstructed jet, which is referred as out of cone energy. On the contrary, the energy of a reconstructed jet might be increased with the energy deposition from other proton-proton collisions, referred as pileup. Finally, besides the hard scattering process between the two partons, the protons remnants will also have softer interactions, which is known as the underlying event. Both the underlying event and pileup produce additional particles and if their direction is the same as the jet, the jet will include their energy in addition to the energy from the quark and gluons from the hard scattering parton-parton interaction.

Only a good accuracy in the energy measurement of the calorimeter jets will allow to correct jets to particle level and to probe final state quarks and gluons.

3.1.2. Jet calibration scheme

In 2012, the ATLAS jet calibration was performed in several stages. An overview of this procedure is shown in figure 3.1. The work described in this thesis uses the calorimeter clusters calibrated at the electromagnetic scale as the jet elements. The calorimeter jets are obtained with the anti-k_T algorithm with an R-parameter of 0.4. In the first stage, calorimeter jets are corrected for the energy offset caused by pileup. This correction is derived from MC and corrects the dependence of the jet energy on the average number of interactions per bunch crossing, μ, and the number of primary vertices (N_{PV}). The jet is by default reconstructed as coming from the origin of the ATLAS detector, and the origin correction takes instead the primary vertex position to define the direction of the jet. This correction does not affect the energy of the jet but only its direction. Next, an energy and η based calibration derived from MC is applied for each jet to correct for the non-compensation of the calorimeter referred in the previous section. *In situ* comparisons from data and MC are used to obtain a residual calibration which is applied only to data. Following these steps results in calorimeter jets calibrated at the jet energy scale: EM+JES.

An overview of the ATLAS jet calibration procedure is shown in figure 3.1. The jet calibration is performed in several phases. The starting point is the calorimeter jets obtained with the anti-k_T algorithm run over the calorimeter clusters calibrated at the electromagnetic scale. Alternatively, the clusters can also be calibrated at the LCW scale, defined in section 2.7.5.2.

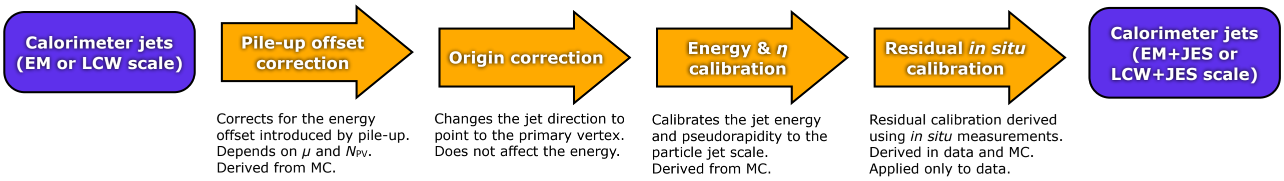


Figure 3.1.: Overview of the ATLAS jet calibration in 2011 [71].

A second pileup correction was added in 2012 to obtain event-by-event corrections based on the average energy density in the event to improve the pileup calibration procedure. Also, an additional calibration, the global sequential calibration (GSC), was designed to reduce the quark-gluon response difference. The study of the GSC was one of the objectives of this thesis.

3.2. Pileup offset correction

The pileup, discussed in section 2.4, imposed substantial challenges in the calibration of jets. Pileup increased the jet energy due to contributions from further proton-proton interactions. The transverse momentum of the jet was corrected by removing the energy contribution from the pileup:

$$p_{\mathrm{T}}^{Corr} = p_{\mathrm{T}}^{jet} - \mathscr{O} \tag{3.1}$$

During 2012, the technique known as the jet area correction [76] estimated this offset energy $\mathscr{O}$ from the area of the jet, A_{jet}, and the event-by-event pileup energy density, ρ, as shown in equation 3.2.

$$\mathscr{O} = \rho \times A_{jet} \tag{3.2}$$

The FASTJET software package [73], used for the jet finding algorithms was used for the area calculation. It applied a very dense and uniform net of particles with density, ν_g, on top of the event. These particles were ultra-soft with $\mathrm{p_T} \sim 10^{-100}$ GeV and referred to as ghost particles. After the jet finding algorithm was run, the ghost particles would be a part of the jet with a negligible contribution to its energy. The quadri-momenta of the ghost particles, g_i^{μ}, belonging to a jet are summed. Then, a four-momentum area of the jet, A_{jet}^{μ}, is defined by dividing this sum by the density of the net, ν_g and by the average $\mathrm{p_T}$ of the ghost particles:

$$A^{\mu}_{jet} = \frac{1}{\nu_g \langle p_{\mathrm{T}}^{g} \rangle} \sum_{g_i^{\mu} \in \mathrm{jet}} g_i^{\mu} \tag{3.3}$$

Finally, the scalar area of the jet is the transverse component of A^{μ}_{jet}.

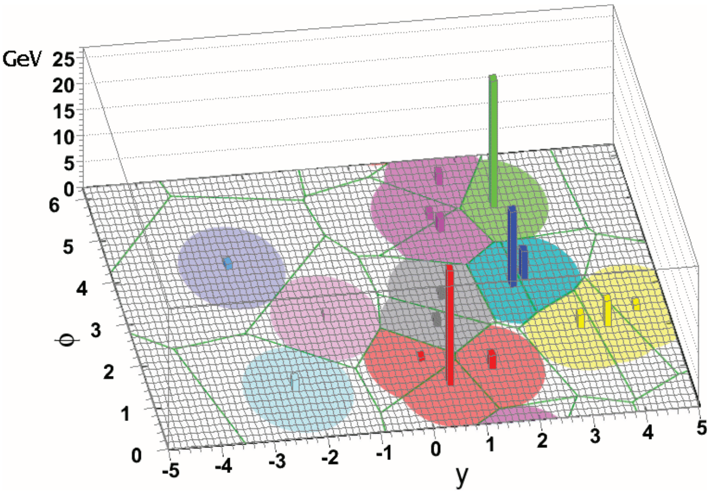
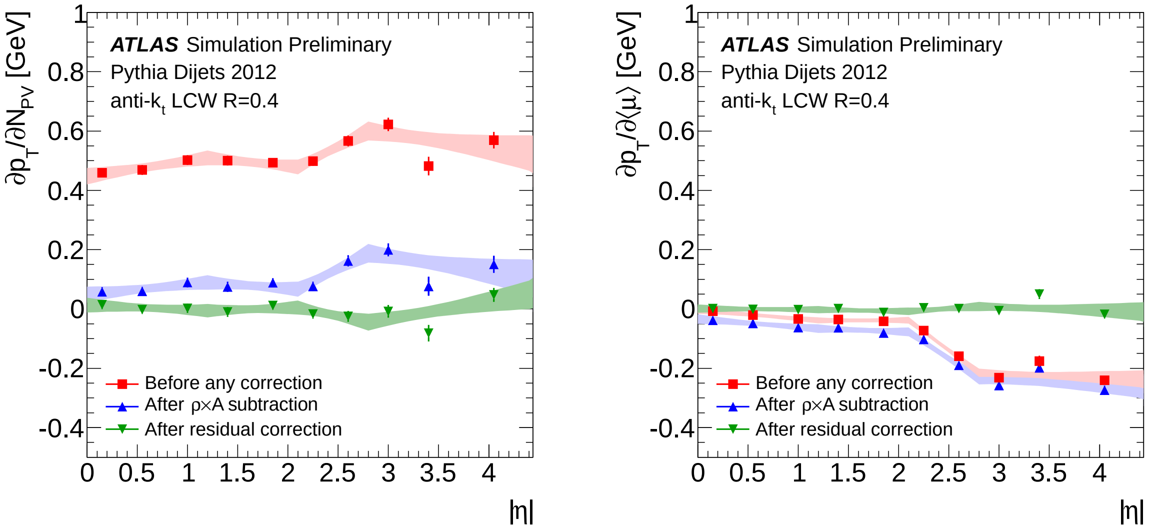


Figure 3.2.: Voronoi regions for one event seeded by the clusters in the calorimeter. The edges of these regions are shown with the green line. The jets are shown by the different coloured regions [112].

In the pileup energy density calculation for the event, ρ, jets are reconstructed with the k_{T} algorithm with the parameter $R = 0.4$. The choice of this algorithm instead of the anti-k_{T} is due to its sensitivity to pileup, as it joins first the soft particles with the closest hard particle. These hard particles, defined by the topological clusters, are used to seed Voronoi regions, that are mathematically defined by the cells which are closest to each seed in the $\eta - \phi$ phase space. Figure 3.2 shows an example of these regions, limited by the green lines, which represent the line segment bisector of two neighbour seeds. The density of each jet was obtained from the ratio between the jet p_{T} and the Voronoi area of the jet. The median of the density of these jets defined the density of the event.

The average jet p_{T} depends linearly on the number of primary vertices, N_{PV}, and the average number of interaction per bunch crossing, $\langle \mu \rangle$, and can be fitted to a straight line with a slope, $\frac{\partial p_{\mathrm{T}}}{\partial N_{PV}}$ and $\frac{\partial p_{\mathrm{T}}}{\partial \langle \mu \rangle}$, respectively [71]. Figure 3.3 shows the fitted slopes as a function of the pseudorapidity of the jet. The red squares refer to the dependence before applying any correction, and the blue up triangles after applying the $\rho \cdot A$ subtraction. The dependence on N_{PV}, sensitive to the in-time pileup, was drastically reduced to a residual dependence smaller



(a)

(b)

Figure 3.3.: Dependence on the N_{PV} (a) and on the $\langle\mu\rangle$ (b) of the reconstructed jet p_T as a function of the pseudorapidity of the jet [76]. The error bands represent the uncertainty of the linear fits. The dependence is obtained for before any correction, in red, after the $\rho \cdot A$ correction, in blue, and after both corrections, in green.

than 0.1 for $|\eta| < 2.5$ and smaller than 0.2 in the forward region. Since the $\rho \cdot A$ correction was designed to correct for the in-time pileup, the effect on the out-of-time pileup estimation variable, $\langle\mu\rangle$, is negligible, as observed in figure 3.3 (b).

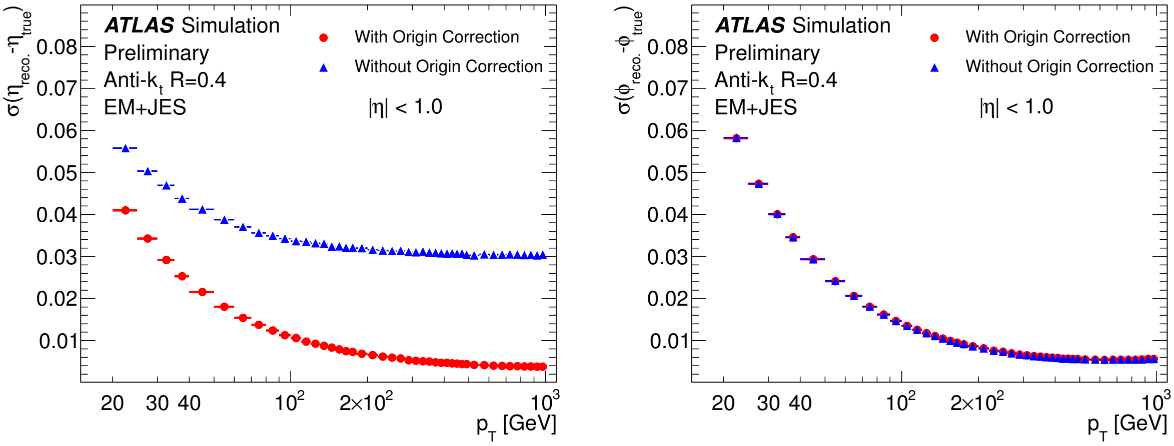
A second correction is implemented to correct for the out-of-time pileup similar to the one used in 2011, which used a parametrization of the dependences of the slopes referred above with N_{PV}, $\langle\mu\rangle$ and η, after the $\rho \cdot A$ correction [71] to further correct the jet transverse momentum. The green down triangles in figure 3.3 show the combination of these two methods, which reduce the dependence of the reconstructed jet p_T on both pileup variables to negligible values. The $\rho \cdot A$ correction was only obtained for the central pseudorapidity region. However, it was also able to correct most of the pileup dependence of the jet p_T on N_{PV}. For the dependence with $\langle\mu\rangle$, there was almost no change and this dependence was only corrected by the residual correction, as shown in the figure.

The systematic uncertainty on the pileup correction from jets comes mainly from the residual dependence of the pileup correction with the energy density, the number of primary vertices, the mean number of collisions per bunch crossing, and extrapolation to other bins of the corrections from the p_T region where it was calculated. The total systematic uncertainty is about 2 % in the $|\eta| < 2.1$ region for jets with $p_T > 40$ GeV, reaching up to 6 % in the forward region for jets with lower p_T [113].

3.3. Origin correction

By default, the (η, ϕ) coordinates of jets are calculated assuming that the collision point is at the origin of the coordinates' system (0,0,0). Although this happens in average, the nature of the train of bunches would produce collisions spread over the z-axis in a region of about 5 cm in 2012. To correct this effect, the calorimeter jet constituents, the clusters, are first corrected to point to the primary vertex and the jet η position is recalculated. The total jet energy remains unchanged, and only the transverse energy is recalculated as it depends on the pseudorapidity.

The effect of the origin correction is shown in figure 3.4, where the resolution of the pseudorapidity (a) and the azimuth angle (b) difference to the truth value are shown as a function of the jet p_T, with (red triangles) and without (blue circles) the origin correction. Almost no effect is observed in the azimuth resolution, which shows in both cases differences of up to 6 % for low-p_T jets, due to the magnetic field which has the direction of the beam.



(a) (b)

Figure 3.4.: Resolution on the η (a) and ϕ (b) position of the jet, as a function of the jet p_T, before and after the origin correction [114].

Before the origin correction, the resolution on the difference between the reconstructed and truth pseudorapidity of the jet, as a function of p_T, is always greater than 3 %, up to the TeV scale. With the origin correction, this resolution improves significantly, becoming smaller than 0.01 for $p_T > 100$ GeV.

3.4. Jet energy scale calibration

The jet energy scale calibration is the most important step of the jet energy calibration and corrects the measured jet energy to account for a series of effects, such as the non-compensation of the ATLAS calorimeters, the energy lost in the cryostat and material upstream of the calorimeter and the energy lost outside the jet cone.

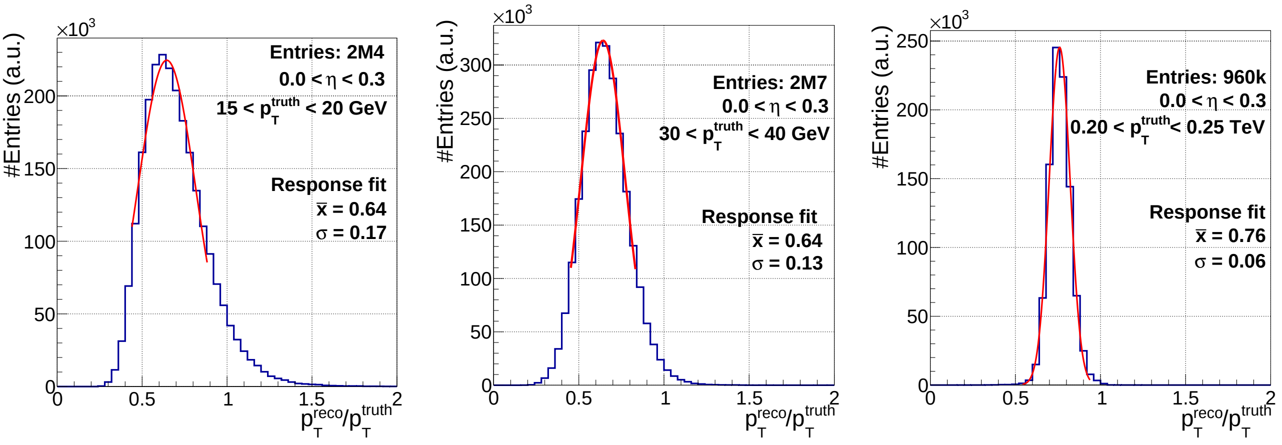
3.4.1. Calibration procedure

The correction is applied as an (η, p_{T}) dependent weight using numerical inversion. This calibration scheme was the default calibration applied by ATLAS analyses in Run-1 [71]. The weights are derived comparing the reconstructed jet E_T with the truth jet E_T in the ratio $E_{\mathrm{T}}^{\mathrm{reco}}/E_{\mathrm{T}}^{\mathrm{truth}}$. Equation 3.4 gives the relationship between E_T and p_T of the jet. For negligible jet masses with respect to the transverse energy, the ratio between transverse energies is the same as the ratio between transverse momentum $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$.

$$E_{\mathrm{T}}^2 = p_{\mathrm{T}}^2 + m^2 \underset{m \ll 1}{\Longrightarrow} E_{\mathrm{T}} = p_{\mathrm{T}} \tag{3.4}$$

Figure 3.5 shows the distributions of the jet energy response for the pseudorapidity region $0.0 < \eta < 0.3$, for three p_T bins. The $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distribution is narrower for the high-p_T jets because they are more collimated than low-p_T jets. The fraction of the energy lost outside the jet cone and in the non-sensitive materials is also smaller for high-p_T jets. For each p_T and η regions, the jet energy response is estimated as the average of the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distributions. To better account for non-Gaussian shaped distributions, particularly in the low-p_T range, a Gaussian function is adjusted to the distributions in the window centred in the mean value with a range extending up and down by 2.5 times the RMS value. If the average, $\bar{x}'$, and standard deviation, σ', are very different from the mean and RMS obtained before, a second fit is performed in the window defined by $[\bar{x}' \pm 2.5\sigma']$.

The mean parameter of the fit as defined above is referred as the jet energy response,

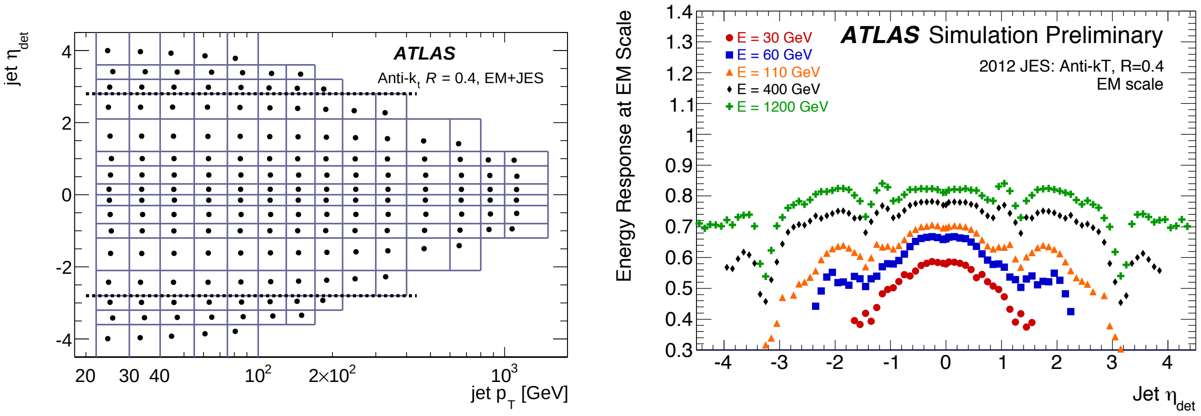


(a) $15 < p_T^{truth} < 20$ GeV (b) $30 < p_T^{truth} < 40$ GeV (c) $0.2 < p_T^{truth} < 0.25$ TeV

Figure 3.5.: p_T^{reco}/p_T^{truth} distributions obtained in the $0.0 < \eta < 0.3$ region for three p_T regions.

$\bar{x}$, and the ratio between the standard deviation of the fit, σ, and the jet energy response is referred as the jet energy resolution.

The jet energy response is obtained in bins of η and p_T of the jet, shown in figure 3.6 (a), which are delimited by the solid lines. The bullets mark the average (η, p_T) of each bin and the dashed line marks the coverage of the inner detector.



(a) (b)

Figure 3.6.: Definition of the η-p_T bi-dimensional bins used to evaluate the jet energy scale corrections (a) [71]. Response at the EM scale as a function of the detector jet η for five different truth jet energy values (b) [114].

Figure 3.6 (b) shows the jet energy response at the electromagnetic scale as a function of the pseudorapidity of the jet for five values of the energy of the truth jet. The sudden changes in the jet energy response along the pseudorapidity in the $0.8 < |\eta| < 1.4$ and $2.8 < |\eta| < 3.4$ regions are due to the transitions between the barrel and endcap calorimeters and between the hadronic endcap and the forward calorimeters, respectively. As observed the energy response of jets at the EM scale is smaller than 1, *i.e.* the reconstructed jet $\mathrm{E_T}$ is, in average, always lower than the truth jet $\mathrm{E_T}$.

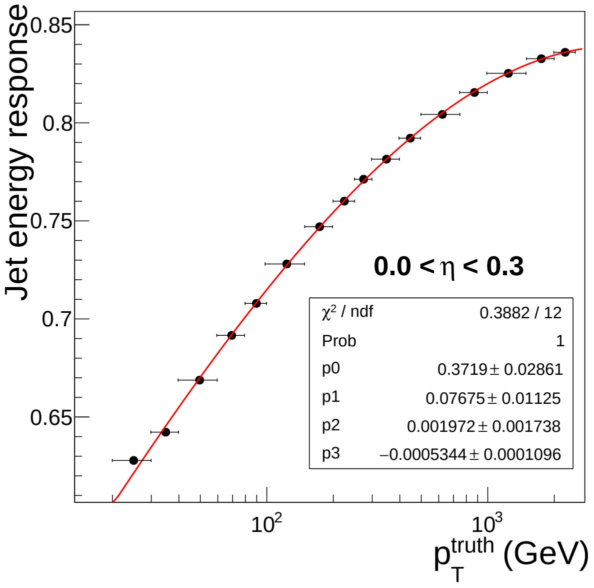


Figure 3.7.: Jet energy response as a function of $\mathrm{p_T}^{truth}$, with the polylogarithmic function fitted with $N_{MAX} = 3$ as defined in equation 3.5. The p_0 to p_3 parameters shown correspond to the $a_{n,0}$ to $a_{n,3}$ parameters in the definition, respectively.

Figure 3.7 shows the jet energy response as a function of the truth jet $\mathrm{p_T}$, for a central η-bin. To obtain the jet energy corrections, the jet energy response is parametrised using six polylogarithmic functions, $\mathscr{F}$, given by equation 3.5.

$$\mathscr{F}_i(E_i^{EM}) = \sum_{n=0}^{N_{\max}} a_{n,i} \left(\ln E_i^{EM}\right)^n \tag{3.5}$$

a_n are the parameters of the function, which will be used to fit the variation of the jet energy response with $\mathrm{p_T}$. For every η bin i, six fits are made, where the number of parameters is varied by ranging the upper bound of the summation, $N_{\max}$, from 1 to 6. The function which has the best goodness of fit, χ^2/ndf, is chosen to represent the jet response at the electromagnetic scale [115]. For the η region shown in figure 3.7, the minimum is obtained for $N_{\max} = 3$ for this η region and the corresponding fit function is shown in red. The parameters of the fit are also presented in the figure.

The correction is then obtained by inverting the value of the function $\mathscr{F}$ for a pseudorapidity bin:

$$E_i^{EM+JES} = \frac{E_i^{EM}}{\mathscr{F}_i(E_i^{EM})} \tag{3.6}$$

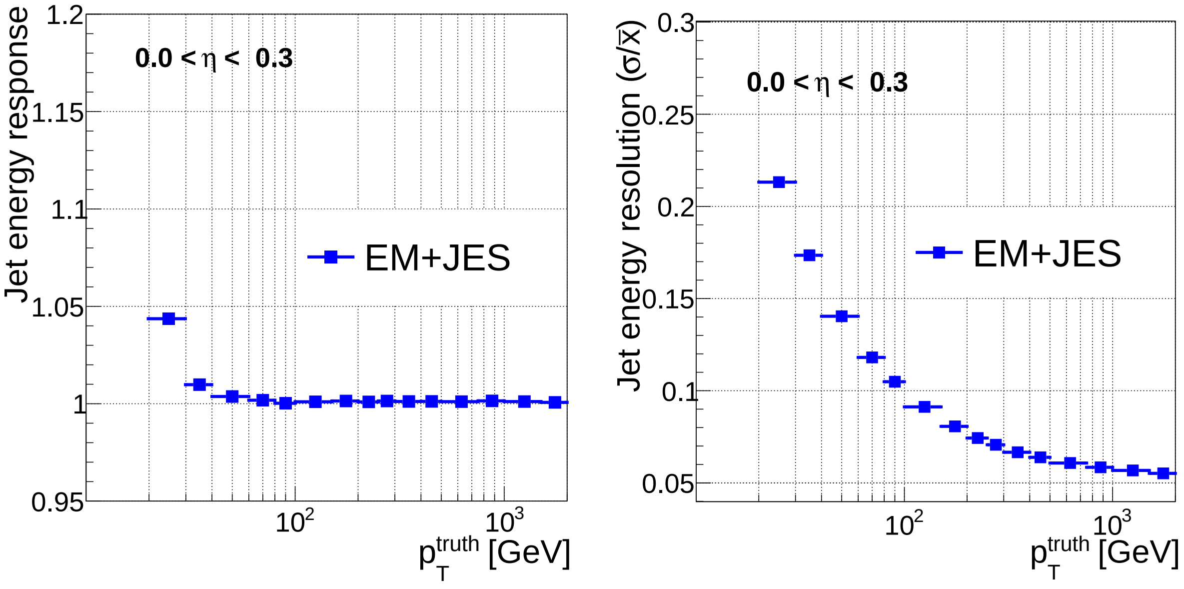
The jet became corrected for the hadronic scale after this procedure.

The jet energy scale (JES) calibration constants are derived using non-diffractive proton-proton collisions in a di-jet sample generated at leading order with PYTHIA6, with an MRST LO** [116] parton distribution function set and AUET2 for the simulation of the underlying event.

Events selected for the JES evaluation were considered if the average transverse momentum between the two leading reconstructed jets did not differ from the leading truth jet by more than 40 % to remove events where one of the jets originated from a pileup collision. Furthermore, a strict isolation is required between all jets that enter on the derivation to be able to make direct “reconstructed jet” to “truth jet” comparisons suppressing energy deposits from other jets. This isolation is achieved by requiring no other jet to be found within a distance of 1.0 in the $\eta - \phi$ phase space. For each reconstructed jet, the distance in the η, ϕ plane to all the isolated truth jets is calculated. If the distance to the closest truth jet is smaller than 0.3, the reconstructed jet is referred as matched to the truth jet.

3.4.2. Jet energy scale calibration performance

After performing the method just described, the jet energy response becomes closer to unity, as shown in figure 3.8 (a) for the pseudorapidity region $0.0 < \eta < 0.3$. Only in the first $\mathrm{p_T}$ bin the jet energy response differs from the unity by more than 1 %. Figure 3.8 (b) shows the jet energy resolution obtained from the ratio between the standard deviation and the mean from the Gaussian fit to the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ ratio distribution, in the same η region. The jet energy resolution is found to improve with the transverse momentum from about 21 % for $\mathrm{p_T}{\sim}25$ GeV to less than 10 % after 100 GeV.



(a) (b)

Figure 3.8.: Jet energy response (a) and resolution (b) as a function of p_T^{truth} for the $0.0 < \eta < 0.3$ region.

3.5. Residual *in situ* calibration and systematic uncertainties

At this stage, both jets reconstructed in data and simulation are calibrated at EM+JES scale. However, there is one additional calibration to correct for differences between data and simulations. This residual *in situ* calibration derivation results from comparing the reconstructed jet p_T against reference objects in data and simulation. Since the protons collide almost head-on, the transverse momentum before the collision should be zero and the same should hold after the collision due to momentum conservation. This property is used by three different techniques to determine the *in situ* residual corrections.

First, a jet can be balanced against a Z boson or a photon. This technique allows the study of the jet at parton level because the Z boson or the photon are produced in direct balance (DB) with the quark or gluon that have initiated the jet. Figure 3.9 shows the Feynman diagrams for the production of these processes. Since the Z boson or the photon are balanced against the quark or gluon and not the jet, a correction for out of cone radiation is applied to the jet in these cases [71, 117]. These two methods are used to correct jets with

p_T up to 800 GeV.

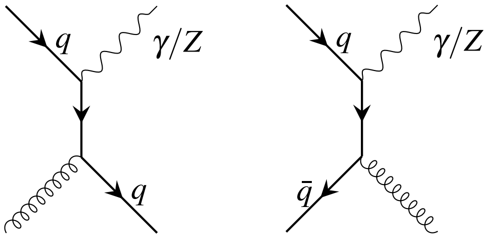


Figure 3.9.: Feynman diagrams for the leading order production of single jets associated to a photon or a Z boson.

In the *in situ* corrections for higher transverse momentum, jets are balanced against a multi-jet recoil system. This is an iterative process that first corrects the p_T of the jets in the recoil system with the direct balance techniques. Next, the method uses the balance between the vectorial sum of the jet's p_T in the recoil system, ${p_T}^{\mathrm{recoil}}$, and the leading jet p_T to correct the latter. Finally, these new corrections are used to correct jets with even higher p_T, up to 1.7 TeV [118].

The average of the distribution of the ratio between the jet p_T and reference object p_T was evaluated for data and simulation, obtaining the double ratio defined in equation 3.7.

$$\frac{R_{Data}}{R_{MC}} = \frac{\langle p_{\mathrm{T}}^{jet}/p_{\mathrm{T}}^{ref}\rangle_{Data}}{\langle p_{\mathrm{T}}^{jet}/p_{\mathrm{T}}^{ref}\rangle_{MC}} \tag{3.7}$$

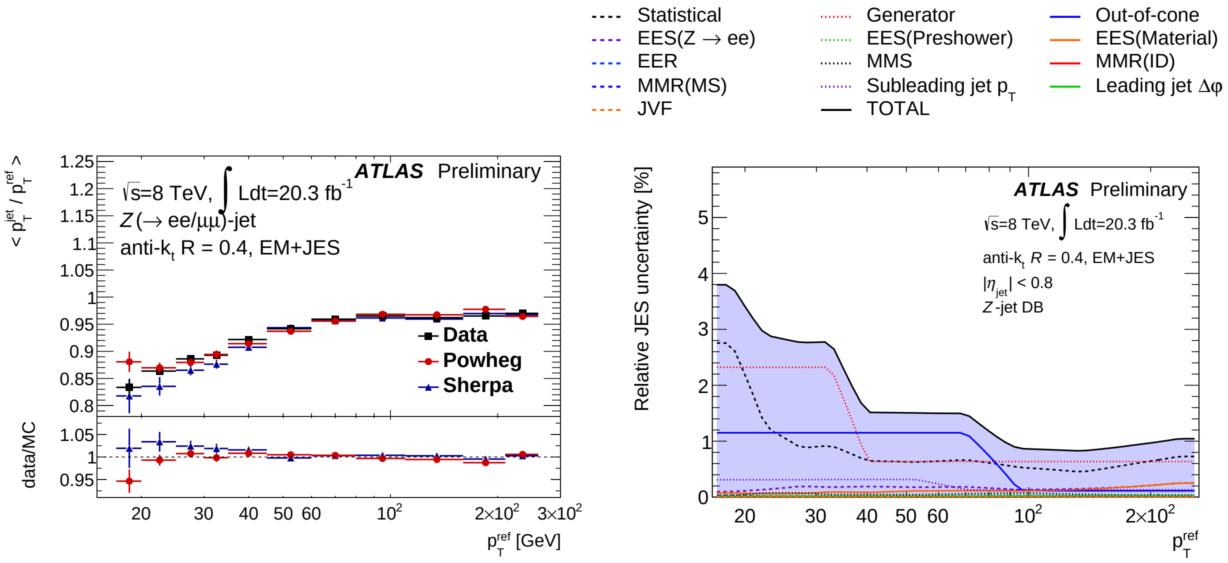
The reference object can be the Z boson, the photon or the multi-jet recoil system. Additionally, the p_T of the reference object is projected to the direction of the jet to improve the p_T balance.

3.5.1. Jet balanced against a Z boson

In the *in situ* calibration with the production of a single jet associated with a Z boson, the Z boson is chosen from its leptonic decay to an electron-positron or a muon pair. Both the energy scale of electrons [38] and the momentum scale of muons [59] are precisely known.

The average of the $p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{Z}}$ distribution is obtained for data and MC [117]. There are two generators used to describe the Z+jet process. The baseline generator is POWHEG interfaced with PYTHIA8 for hadronization process, using the CT10 PDF set and the AUET2 tune. This generator is used to obtain the *in situ* corrections for a jet balanced with a Z boson.

However, there are other generators which are used to simulate the multitude of processes in ATLAS. The distribution of $\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{Z}}\rangle$ is shown in figure 3.10 (a) as a function of $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ for data (black squares) and for the Z+jet samples simulated with Powheg+Pythia8 (red circle) and with Sherpa (blue triangles). Sherpa used also the CT10 PDF set but had different internal implementations of each event generator step and therefore was suitable to test the theoretical uncertainties. The Powheg generator had compatible $\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{Z}}\rangle$ with data expect for the first bin, whereas Sherpa discrepancies occur below 35 GeV.



(a) (b)

Figure 3.10.: $\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{Z}}\rangle$ as a function of the $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ (a). The *in situ* correction is obtained for jets balanced against a Z boson from the ratio of the Powheg with respect to data. Only statistical uncertainties are presented. Systematic uncertainty on the evaluation of the *in situ* correction on jets balanced against a Z boson (b). The individual uncertainties are summed in quadrature to obtain the total uncertainty [117].

There are sources of systematic uncertainties due to limitations on the methods which produce an uncertainty on the measurement of the jet energy scale. These systematic uncertainties are then propagated to all analysis in ATLAS. The systematic uncertainties are obtained from uncertainties on the detector, on the physics modelling or on the methods used to evaluate the jet energy scale.

The largest contribution to the systematic uncertainty of the direct balance with a Z boson came from the MC generator modelling uncertainty, obtained from the POWHEG+PYTHIA8 response difference with SHERPA. The difference between the two generators is about 4 % for jets with $p_T < 20$ GeV and falls below 1 % after 40 GeV. The second leading contribution to the systematic uncertainty is due to the out-of-cone correction applied in the Z+jet balance. It was evaluated comparing the $\langle p_T^{\text{jet}}/p_T^Z \rangle$ obtained with the baseline simulation against data. The estimative of the data-MC difference is about 1.2 % for $p_T^Z < 70$ GeV and becomes about 0.1 % above 100 GeV. The remaining systematic uncertainties include variations on additional radiation, the energy scale and resolution of the decays of the Z boson (electrons and muons), and the quality selection conditions of jets. These uncertainty contributions are smaller than 0.3 % [71, 117]. The statistical uncertainties of the baseline generator of Z+jets are only a significant systematic uncertainty source for $p_T^Z < 30$ GeV, becoming the dominant contribution below 20 GeV with up to 2.8 % uncertainty. For $p_T^Z > 40$ GeV, this contribution stabilises at about 0.6 %.

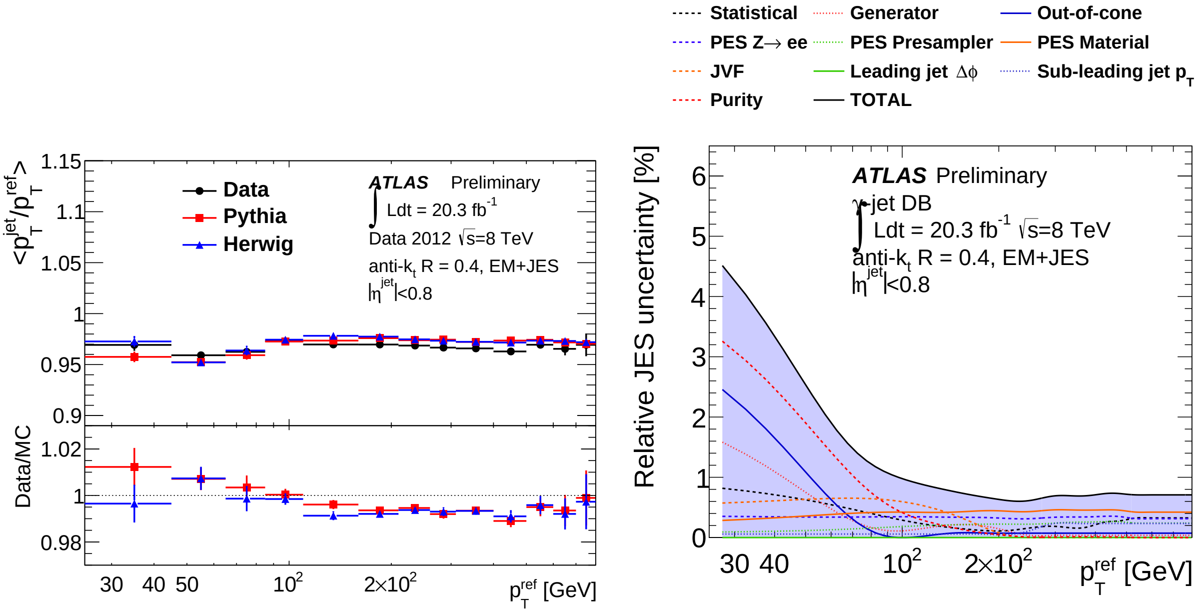
Figure 3.10 (b) shows each of the components tested for the jet energy scale uncertainty as a function of the transverse momentum of the Z boson. The total uncertainty, represented by the black line, is calculated from the quadratic sum of each independent component and ranges from 3.8 % for low p_T^Z to 0.9 % at about 100 – 150 GeV.

3.5.2. Jet balanced against a photon

For the γ+jet process, the baseline generator chosen is PYTHIA8 with the AUET2 tune. The resulting average of the $p_T^{\text{jet}}/p_T^{\gamma}$ distribution as a function of p_T^{γ} is shown in figure 3.11 (a) for data (black circles) and for the γ+jet process simulated with PYTHIA8 (red squares) and with HERWIG (blue triangles) C++ version. HERWIG is a different event generator and uses also different implementation of the underlying event tune, which is UE-EE-3 [119], and of the hadronization process, making it a suitable generator to test theoretical uncertainties. Both event generators used the CTEQ6L1 PDF set.

There are differences between data and the simulation with PYTHIA8 of the order of 0.5 % for $p_T^{\gamma} > 100$ GeV and slightly larger for the low-p_T^{γ} regime. Differences ranging from negligible to 0.5 % are found when comparing the two generators except for the first p_T^{γ} interval, $25 < p_T^{\gamma} < 45$ GeV.

The uncertainties obtained for this *in situ* calibration, include the same effects of out-of-cone radiation, underlying event, additional parton radiation and JVF that were discussed for the Z+jets case, in section 3.5.1. The theoretical uncertainties are obtained from comparisons



(a) (b)

Figure 3.11.: $\langle p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\gamma}\rangle$ as a function of the $\mathrm{p}_{\mathrm{T}}^{\gamma}$ (a). The *in situ* correction is obtained for jets balanced against a photon from the ratio of the PYTHIA8 simulation with respect to data. Only statistical uncertainties are presented. Systematic uncertainty on the evaluation of the *in situ* correction on jets balanced against a photon (b). The individual uncertainties are summed in quadrature to obtain the total uncertainty [117].

between the baseline generator, PYTHIA8, and the alternative generator, HERWIG++.

In addition to the systematic uncertainty sources referred above, there is another uncertainty associated with the contamination of di-jet events, where one of the jets is misidentified as a photon [120]. This effect, labelled as *purity* in figure 3.11 (b), dominates the total uncertainty on the γ+jet *in situ* technique up to photons with $\mathrm{p}_{\mathrm{T}}^{\gamma} = 80$ GeV, as the di-jet background is relevant particularly at low $\mathrm{p}_{\mathrm{T}}^{\gamma}$. The *purity* relative uncertainty is as high as 3.2 % in the first $\mathrm{p}_{\mathrm{T}}^{\gamma}$-bin, becomes smaller than 1 % at 70 GeV and negligible above 200 GeV. The systematic uncertainty due to the out-of-cone radiation is smaller than 0.1 % for $\mathrm{p}_{\mathrm{T}}^{\gamma} > 80$ GeV, and it is only relevant at low-$\mathrm{p}_{\mathrm{T}}^{\gamma}$, reaching 2.4 % in the first $\mathrm{p}_{\mathrm{T}}^{\gamma}$-bin. The average of the $p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\gamma}$ distribution as a function of $\mathrm{p}_{\mathrm{T}}^{\gamma}$ for HERWIG, shown by the blue triangles in figure 3.11 (a), have a relative difference to PYTHIA8 of about 1.6 % at low-$\mathrm{p}_{\mathrm{T}}^{\gamma}$ and 0.2 % for $\mathrm{p}_{\mathrm{T}}^{\gamma} > 80$ GeV. This difference is taken as the uncertainty due to generator.

The remaining systematic uncertainty components are no larger than 0.5 % [71, 117]. The statistical uncertainty of the baseline generator of γ+jets is less than 0.3 % for $\mathrm{p}_{\mathrm{T}}^{\gamma} > 100$ GeV, reaching 0.8 % in the lowest $\mathrm{p}_{\mathrm{T}}^{\gamma}$-bin.

Figure 3.11 (b) shows each of the components tested for the jet energy scale uncertainty as a function of the transverse momentum of the photon. The total uncertainty, represented by the black line, is calculated from the quadratic sum of each independent component. The total uncertainty is below 1 % for $\mathrm{p}_{\mathrm{T}}^{\gamma} > 100$ GeV and can reach up to 4.5 % for the lowest $\mathrm{p}_{\mathrm{T}}^{\gamma}$-bin.

3.5.3. Jet balanced against multi-jet recoil system

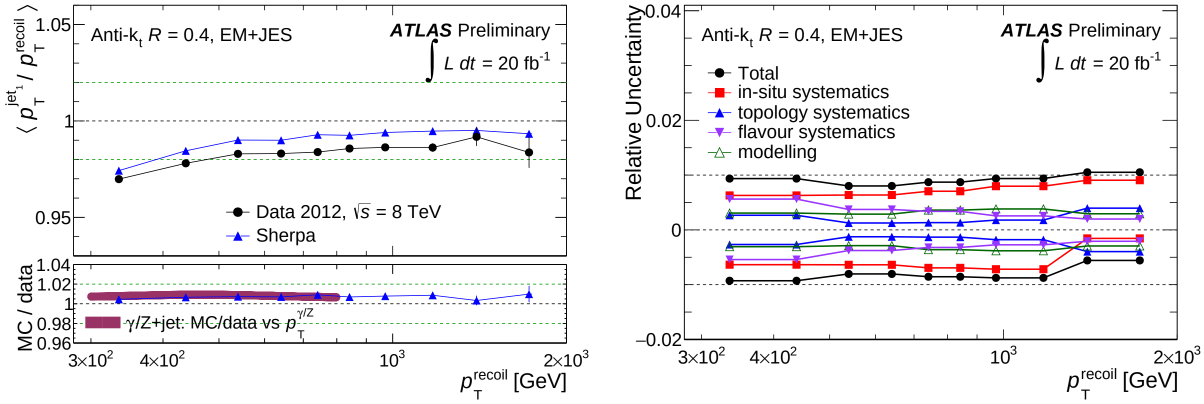
The multi-jet balance, MJB, is defined as the ratio in equation 3.8:

$$\mathrm{MJB} = \frac{p_{\mathrm{T}}^{\mathrm{leading}}}{|\vec{p}_{\mathrm{T}}{}^{\mathrm{recoil}}|} \tag{3.8}$$

where $p_{\mathrm{T}}^{\mathrm{leading}}$ is the transverse momentum of the jet to be calibrated and $\vec{p}_{\mathrm{T}}{}^{\mathrm{recoil}}$ is the vectorial sum of all other sub-leading jets. The multi-jet balance is obtained for data and simulation and the ratio between them defines the correction applied to data [118].

Figure 3.12 (a) shows the average of the multi-jet balance distribution as a function of the transverse momentum of the recoil system for data (black circles) and simulation with the SHERPA generator (blue triangles). The ratio between these distributions is shown in the bottom panel and provides the correction required to calibrate the leading jet. The average contribution from the two remaining *in situ* corrections is superimposed with the purple line. The discrepancies between MC and data are always smaller than 1 % and, in the region $300 < p_{\mathrm{T}}^{\mathrm{recoil}} < 800$ GeV, the MC/data ratio are compatible to what was obtained for the other two *in situ* calibrations.

The most important systematic uncertainties for the multi-jet method come from the uncertainties of the other *in situ* methods due to the jets present in the recoil system. Their impact on the multi-jet balance correction increase with $p_{\mathrm{T}}^{\mathrm{recoil}}$ from 0.6 % to 0.9 %, as shown in figure 3.12 (b) with red squares. The effect of the event selection conditions used in this method, called *topology systematic* uncertainties, is below 0.4 % in all $p_{\mathrm{T}}^{\mathrm{recoil}}$ range, as shown with up filled blue triangles. To evaluate the impact on the generator selection, the SHERPA default generator for the multi-jet balance is compared against POWHEG+PYTHIA8,



(a) (b)

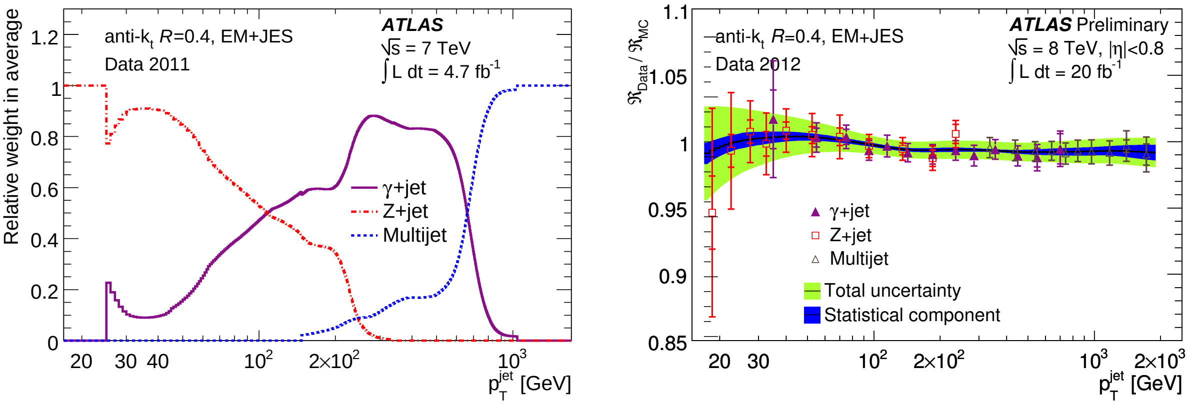
Figure 3.12.: $\langle p_T^{\mathrm{jet}} / p_T^{\mathrm{recoil}} \rangle$ as a function of the p_T^{recoil} (a). The *in situ* correction is obtained for jets balanced against a multi-jet system from the ratio of the SHERPA with respect to data. Only statistical uncertainties are presented. Systematic uncertainty on the evaluation of the *in situ* correction on jets balanced against a multi-jet system (b). The individual uncertainties are summed in quadrature to obtain the total uncertainty [118].

PYTHIA8 and HERWIG. This contribution is referred as *modelling* (green open triangles) and the contribution is 0.3 % to 0.4 % depending on the p_T^{recoil} region. Finally, there is an uncertainty associated with the flavour composition of the jets in the recoil system, since jets initiated by gluons have a calorimeter energy response different from jets initiated by quarks [115]. The effect of this *flavour* uncertainty decreases with p_T^{recoil} from 0.6 % to 0.2 %, as shown in purple down triangles in the same figure. The total uncertainty (black circles) is calculated from the quadratic sum of these components and values about 1 %, with no significant variation with p_T^{recoil}.

3.5.4. Combination of the *in situ* measurements

Each of the *in situ* techniques provides data-MC comparisons just for one region of p_T, where statistics are large enough. The Z-jet method evaluated the correction for jets with p_T between 17 and 250 GeV, while the γ+jet method covered the intermediate region, $25 < p_T < 800$ GeV and the jet balance with the recoil multi-jet system was used for p_T between 300 GeV and 1.7 TeV.

For the application of the *in situ* correction to data events, the information from these three techniques has to be combined. This combination takes into account also the systematic uncertainties of each method [71]. First, for each of the methods, the $\langle p_T^{jet}/p_T^{ref}\rangle$ is obtained in p_T^{ref} bins of 1 GeV, interpolating with second order polynomial splines. Since the jet energy response is close to unity after the $\eta - p_T$ calibration (see section 3.4), it is assumed that p_T^{ref} corresponds, in average, to the p_T of the jet. The residual correction for the *in situ* combination is then obtained for each bin through a χ^2 minimization of the jet energy response ratios. A weight is evaluated for each of the methods based on the distance to the combination value. On figure 3.13 (a), these weights are shown for the 2011 data taking for the three *in situ* techniques. Z+jet is the most important method up to about 100 GeV, in which γ+jet starts to dominate until 600 GeV. The multi-jet method takes over at this p_T value [71].

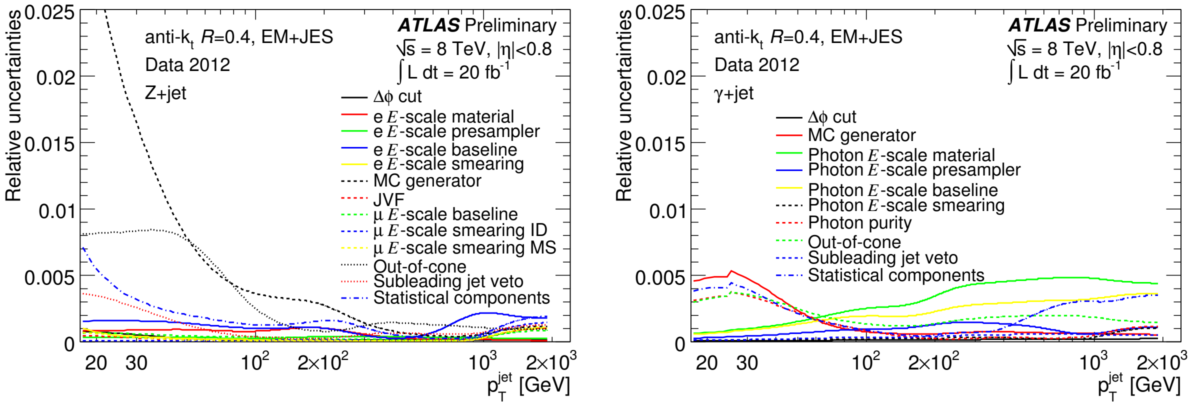


(a) (b)

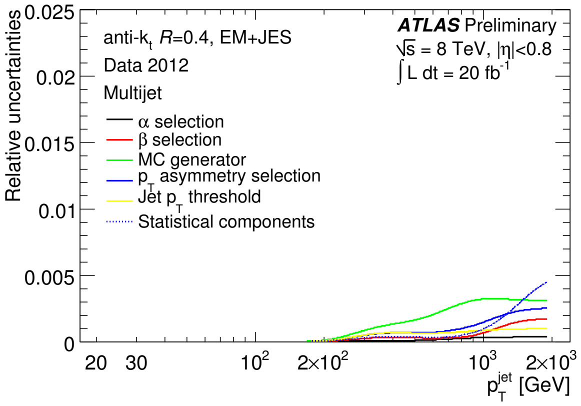
Figure 3.13.: Average relative weight of each *in situ* method for the combined residual correction as a function of the p_T of the jet (a) [71]. Inverse of the residual correction obtained for the three method (markers) and for the combination of them (black line) (b) [114].

The measurements of the ratios between the jet energy response in data and the default MC are summarised in figure 3.13 (b) for the three *in situ* methods. The smaller and the larger error bars for each of the measurements represent the statistical uncertainty and the total uncertainty, respectively. The latter ranges from 3 % for jets with p_T of 20 GeV to 1 % after about 100 GeV. The solid black line is the combination of the three methods. The blue and green bands are the statistical and total uncertainties of the combination, respectively.

The weights presented in figure 3.13 (a) are also applied to the uncertainties obtained in sections 3.5.1, 3.5.2 and 3.5.3 to obtain the systematic uncertainties for the combination of the three *in situ* methods. The resulting systematic uncertainties are those shown in figure 3.14. With the weights applied, the uncertainty contributions from the Z+jet method decrease rapidly for $p_T^{jet} > 200$ GeV. The same is also observed for the multi-jet method, with no contribution for p_T^{jet} smaller than 200 GeV. The flavour uncertainty component is absent because it will be handled as a separate systematic uncertainty for jets in section 3.7.1. The modelling of the MC generator becomes the leading contribution to the uncertainty, apart from the highest p_T bin, where the statistical component of the simulation becomes the largest effect for this method.



(a) Z+jet (b) γ+jet



(c) Multi-jet

Figure 3.14.: Relative systematic uncertainties for the three methods for the residual *in situ* correction weighted by their contribution to the combined *in situ* correction [114].

A total of sixty-five uncertainty components arise from the *in situ* residual calibration, but there are some correlations between them. To reduce the sources of uncertainty to the relevant ones, the covariance matrix of the correction factors is diagonalized, obtaining the non-correlated uncertainties from the eigenvectors of the matrix. The top five uncertainties raised from this procedure are taken as systematic uncertainty sources. All the remaining eigenvectors are combined into a single residual uncertainty source. Figure 3.15 shows the result of this procedure. The five leading uncertainties corresponding to the leading eigenvectors are shown with solid lines, and the dashed line indicates the combination of the remaining.

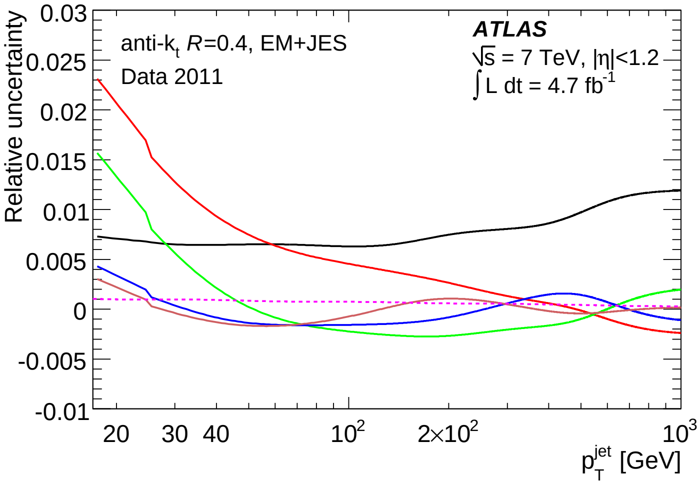


Figure 3.15.: Relative uncertainty as a function of the jet transverse momentum for the reduced set of uncertainty components. The top five are shown with filled lines and the combined residual uncertainty is shown with the dashed line [71].

3.6. Calibration of jets in the forward region

The *in situ* residual corrections presented in the section 3.5 with the Z+jet, γ+jet and multi-jet methods are applied only to jets in the central detector region. The inner detector is necessary to evaluate the p_T of the muons and electrons originating from the Z boson decay. Beyond the inner detector coverage, there is no difference in the reconstruction of electrons and photons. In the multi-jet method, the jets in the recoil system are also calibrated using the first two methods, valid only at central pseudorapidity regions.

However, the balance in the transverse plane is still valid if one of the objects is forward and the other is central, which is the base for the di-jet balance for the forward *in situ* correction of jets [118]. Here, the energy response of a jet in the forward region is compared with the energy response of the jet in the central region, and the distribution of the ratio between their transverse momenta is obtained:

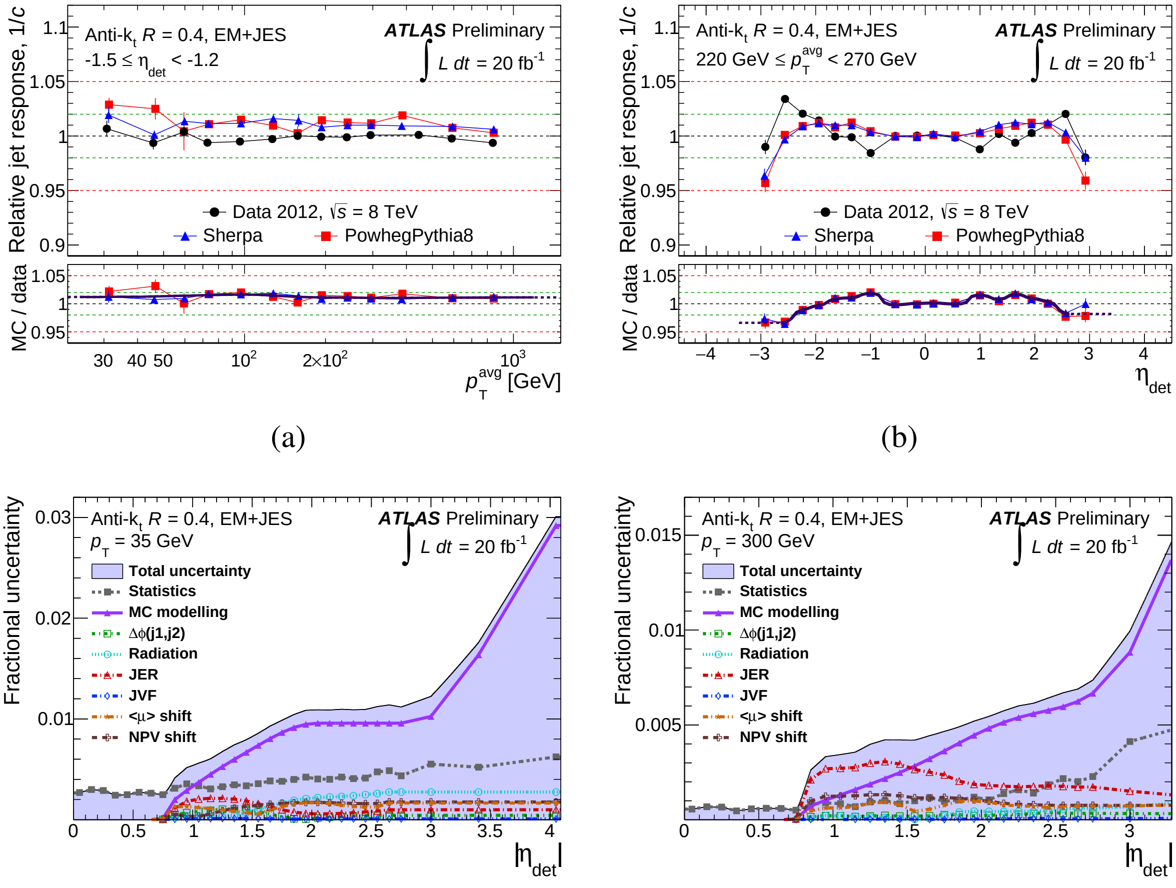
$$\mathscr{R} = \frac{p_{\mathrm{T}}^{\mathrm{forw}}}{p_{\mathrm{T}}^{\mathrm{ref}}} \tag{3.9}$$

The correction, c, to be applied to the forward jet is the inverse of the average of the $\mathscr{R}$ distribution.

The statistics that results from the usage of the events with central-forward jet pairs are low, so the method uses any reference jet to correct another jet with larger η.

Figure 3.16 shows the relative jet energy response as a function of the average $\mathrm{p_T}$ of the two jets for the $-1.5 < \eta < -1.2$ range (c) and of the jet pseudorapidity when the two jets have their average transverse momentum in the $220 < p_{\mathrm{T}}^{\mathrm{avg}} < 270$ GeV range (d). The data results are shown with black circles and compared against the simulations obtained with SHERPA (blue triangles) and with POWHEG+PYTHIA8 (red squares). The bottom panel shows the MC/data ratio for each generator. The black line in the bottom panel shows the residual correction.

Limitations on the di-jet method lead to systematic uncertainties due to the theoretical modelling of the predictions, the topology selection for the event, to the pileup and the jet energy resolution. The physics modelling uncertainty, shown by the purple continuous line in figure 3.16 (bottom), is the dominant uncertainty always for large pseudorapidity, reaching almost 3 % for jets with $\mathrm{p_T} = 35$ GeV (a). For jets with $\mathrm{p_T} = 300$ GeV (b), this uncertainty is reduced by half but still dominates for $|\eta| > 1.5$. The remaining uncertainties account for variations on the event selection: the azimuth distance between the two jets, additional radiation in the event and JVF. Also included are some pileup effects and the jet energy resolution. Most of these uncertainties are negligible and always smaller than 0.3 %. The statistical uncertainty is always smaller than 0.5 %, reaching this value at $|\eta| = 3.5$ for jets with $\mathrm{p_T} = 300$ GeV. The black line that delimits the blue area is the total uncertainty obtained from the quadratic sum of the individual systematic uncertainty components, that ranges between 0.2 % and 3 % for central and forward low-$\mathrm{p_T}$ jets, respectively, and about half in the high-$\mathrm{p_T}$ region.



(a) (b)

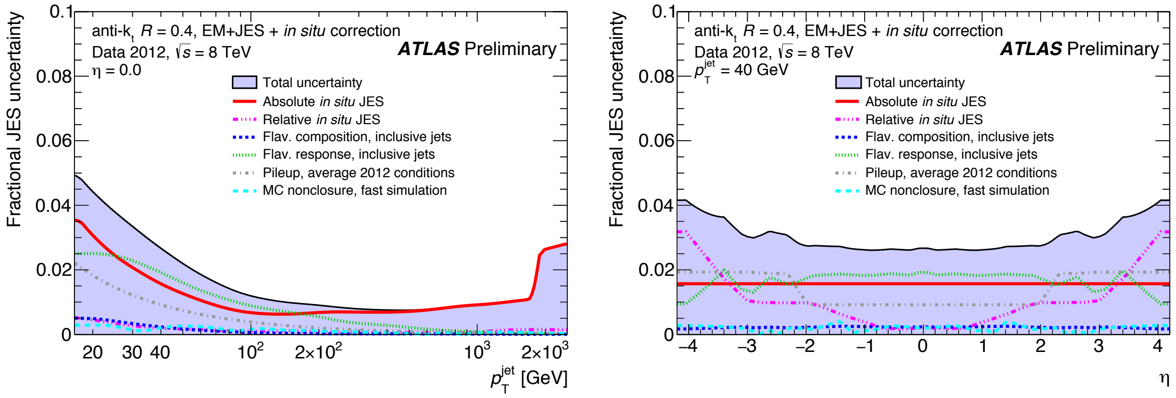
(c) (d)

Figure 3.16.: Relative jet energy response as a function of p_T^{avg} (a) and η (b) for data and two generators. Relative systematic uncertainties from each of the sources for $p_T = 35$ GeV (c) and $p_T = 300$ GeV (d) [118].

3.7. Systematic uncertainties on the jet energy scale

Figure 3.17 shows the total systematic uncertainty on the jet energy response as a function of the jet transverse momentum (a) and the jet pseudorapidity (b), grouping the several systematic components. The uncertainty from the three *in situ* calibrations discussed in section 3.5 is referred as the “Absolute *in situ* JES”. This uncertainty only depends on the jet p_T and is the leading uncertainty for jets with $p_T < 25$ GeV and with $p_T > 200$ GeV. In the most central η-bin, this uncertainty goes as high as 3.5 % and 3 % for the lowest and highest p_T regions, respectively.

The uncertainty from di-jet method of the forward pseudorapidity region is referred as



(a) (b)

Figure 3.17.: Relative jet energy response systematic uncertainty as a function of transverse momentum (a) and pseudorapidity (b) of the jet for the systematic terms [114]. The area delimited by the black line corresponds to the quadratic sum of the different terms and represents the total systematic uncertainty on the jet energy scale

the “Relative *in situ* JES” discussed in the previous section. This uncertainty rivalizes with the “Absolute *in situ* JES” at $|\eta| = 3.1$, for a $\mathrm{p_T} = 40$ GeV jet and becomes dominant for jets with $|\eta| > 3.3$ with a contribution up to 3.2 %.

The pileup contribution, discussed in sections 3.2 and 3.6 for the central and forward regions, respectively, contributes with an uncertainty of up to 2 % in both the low-$\mathrm{p_T}$ and forward regimes, becoming smaller in the remaining $\eta -$ $\mathrm{p_T}$ regions.

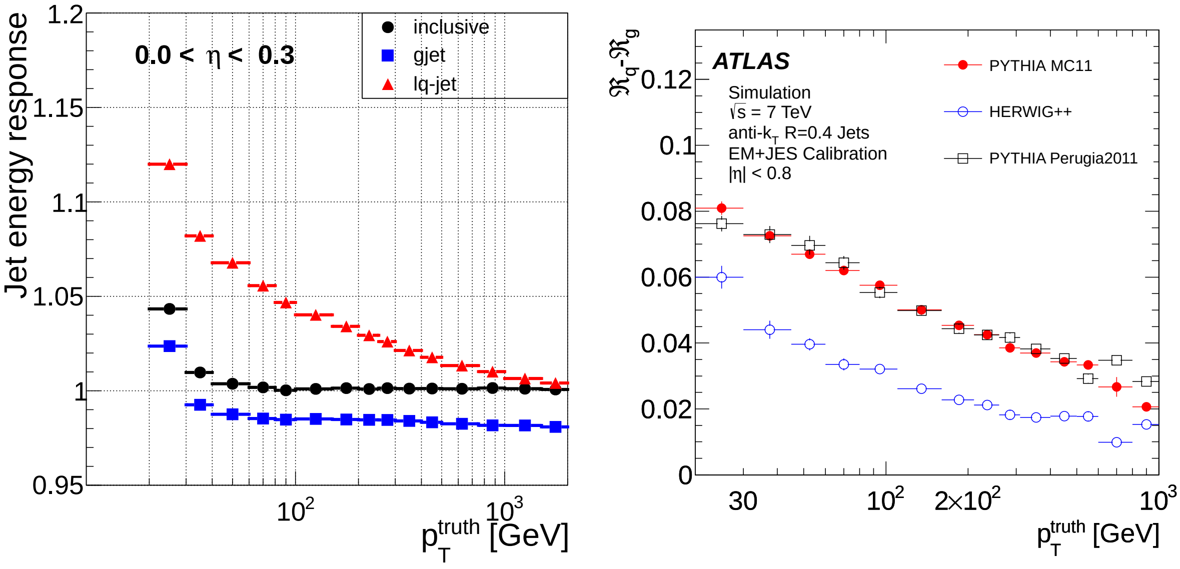
There are two flavour uncertainties, related to the jet flavour composition and response, that is discussed in section 3.7.1 and their uncertainties are represented in the figure with the dark blue and green lines. The flavour composition was found to have negligible contribution in most $\mathrm{p_T}$ and η regions. On the contrary, the flavour response is the dominant contribution in the intermediate $\mathrm{p_T}$ regime, with a contribution which decreases with $\mathrm{p_T}$ from 2.25 % to 1.5 %.

The last systematic uncertainty in the figure, represented by the light-blue line is related to the non-closure observed in the jet energy response for simulated events when the fast simulation, referred in section 2.8, is used with the parametrization of the calorimeter response and is discussed in section 3.7.3. This uncertainty has a negligible contribution regardless the $\eta -$ $\mathrm{p_T}$ region.

There is an additional uncertainty, which is obtained for the b-jet energy response and is presented in section 3.7.2. This uncertainty is not included in the figure because it is applied only to b-jets and replaces the two flavour uncertainties.

3.7.1. Light flavour systematic uncertainties

The jet energy calibration constants discussed in section 3.4 are obtained using the PYTHIA simulated di-jet sample, known as the inclusive sample, in which there is no separation based on the flavour of the jets. However, if the study is performed separately for jets initiated by quarks and by gluons, the jet energy response is known to be very different [71, 121]. These two types of jets are distinguished in simulation only, through the evaluation of which particle, quark or gluon, has the largest transverse momentum within the jet, which will tag the jet as being a quark or a gluon initiated jet, respectively.



(a) (b)

Figure 3.18.: Jet energy response as a function of p_T^{truth} for light quark and gluon initiated jets (a). Difference between the responses for light quark and gluon initiated jets for three simulations (b) [71].

Figure 3.18 (a) shows the jet energy response, defined in section 3.4.1, for the two types of jets: light-quark initiated jets (red triangles) and gluon initiated jets (blue squares). The jet energy response for the inclusive jet case, which combines the two flavours, is also shown with black circles for comparison. The jet energy response for gluon initiated jets is nearly always below one: the reconstructed transverse momentum of the jet in the calorimeter is in average 1.5 % smaller than for the jet obtained from the stable particles at the generator level. The discrepancies between the energy of the reconstructed and the truth jets for light-quark initiated jets are even larger since the jet energy response exceeds 1 by 12 % in the $\mathrm{p}_{\mathrm{T}}^{truth}$ range between 20 and 30 GeV, decreasing slowly to 1 % for jets with $\mathrm{p}_{\mathrm{T}}^{truth} = 1$ TeV.

Assuming that the jet energy scale for jets initiated by quarks is independent of the quark type, the systematic uncertainty for the gluon-quark dependence, $\Delta\mathscr{R}$, is given by equation 3.10.

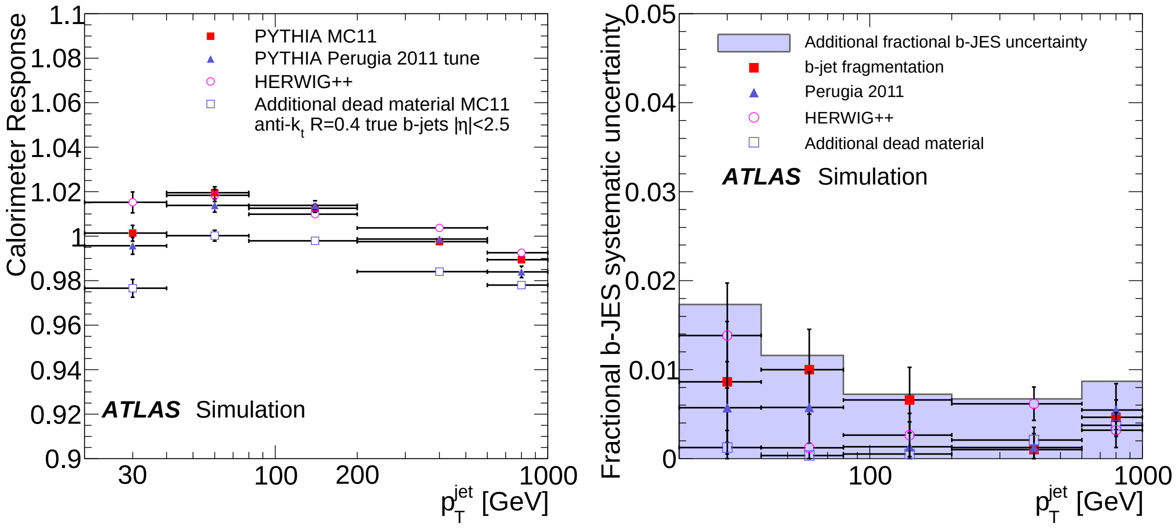
$$\Delta\mathscr{R} = \Delta f_{\mathrm{q}} \left(\mathscr{R}_{\mathrm{q}} - \mathscr{R}_{\mathrm{g}}\right) + \Delta\mathscr{R}_{\mathrm{q}} + f_{\mathrm{g}} \Delta\mathscr{R}_{\mathrm{g}}^{\mathrm{ex}} \tag{3.10}$$

where Δ represents the uncertainty on a given measurement and the fraction of jets initiated by quarks and gluons (f_{q} and f_{g}, respectively) is complementary: $f_{\mathrm{q}} + f_{\mathrm{g}} = 1$. $\mathscr{R}_{\mathrm{q}}$ and $\mathscr{R}_{\mathrm{g}}$ are the jet energy response for quarks and gluons, respectively. Figure 3.18 (b) shows the difference in the jet energy response between jets initiated by quarks and by gluons for three different simulations with the objective of testing the different models for the hadronization, which are observed to cause different jet energy response for quarks and gluons. The default PYTHIA simulation (red closed circles) is used for the evaluation of the difference in equation 3.10. The same PYTHIA generator is shown but using the PERUGIA tune instead (black squares). The HERWIG++ generator (blue open circles) that has an independent method for the hadronization process is also shown. The difference between the jet energy response for quark and gluons initiated jets obtained with two PYTHIA simulations is not very significant but the difference between PYTHIA and HERWIG++ can reach up to 3 % in some $\mathrm{p}_{\mathrm{T}}^{truth}$-bins. This difference represents the uncertainty on the jet energy response of gluons, $\Delta\mathscr{R}_{\mathrm{g}}^{\mathrm{ex}}$.

3.7.2. Heavy flavour systematic uncertainty

An additional jet energy scale systematic uncertainty, accounting for the uncertainty on the simulation model for the production and hadronization of b-quarks has to be considered. Jets are labelled as truth b-jets if there is a truth b-quark, within a $\Delta R = 0.3$ distance of the jet axis. c-jets are treated as light-quark jets.

The main physics process to evaluate the jet energy response to b-jets is di-jet events, where b-jets are produced from gluon-splitting, $g \to b\bar{b}$. In 2011, the generators used



(a) (b)

Figure 3.19.: Fractional b-jet systematic uncertainty as a function of the transverse momenta of the jet [71].

for di-jet events [71] were PYTHIA6 (red filled squares) and HERWIG++ (magenta open circles) and figure 3.19 (a) shows their b-jet energy response as a function of the p_T of the jet. Additionally, there is the b-jet energy response difference from replacing the default underlying event tune of PYTHIA6, AUET2B, with PERUGIA 2011 to test the effect of different underlying event tunes, which is shown with the purple triangles. Finally, the b-jet energy response is also obtained with some additional dead material (purple open squares) simulated in front of the calorimeter to evaluate the uncertainty of the detector knowledge. Figure 3.19 (b) shows the contribution of each systematic term to the total b-jet energy scale systematic uncertainty as a function of the jet transverse momenta. This evaluation is taken from the relative difference in the b-jet response between the default simulation and the alternatives. The impact of b-jet fragmentation on the systematic uncertainty is obtained by replacing nominal PYTHIA6 parameters with the BOWLER-LUND and PROFESSOR tunes [115] and is shown with the red filled squares. The total b-jet uncertainty, obtained from the quadratic sum of the individual components is about $1.5-2$ % for p_T below 80 GeV and less than 1 % above that.

In addition to the di-jet process, $t\bar{t}$ events were also analysed, which were simulated with POWHEG+HERWIG++ and MC@NLO+PYTHIA6. The PYTHIA6 to HERWIG++ comparison tests the hadronization modelling of the b-quarks from the top decays. The POWHEG to MC@NLO comparison evaluates the systematic uncertainty with different NLO generators.

The b-jet energy response and systematic uncertainty are also validated in data using tracks to calculate the fraction of the transverse momentum of the jet carried by charged particles as presented in equation 3.11:

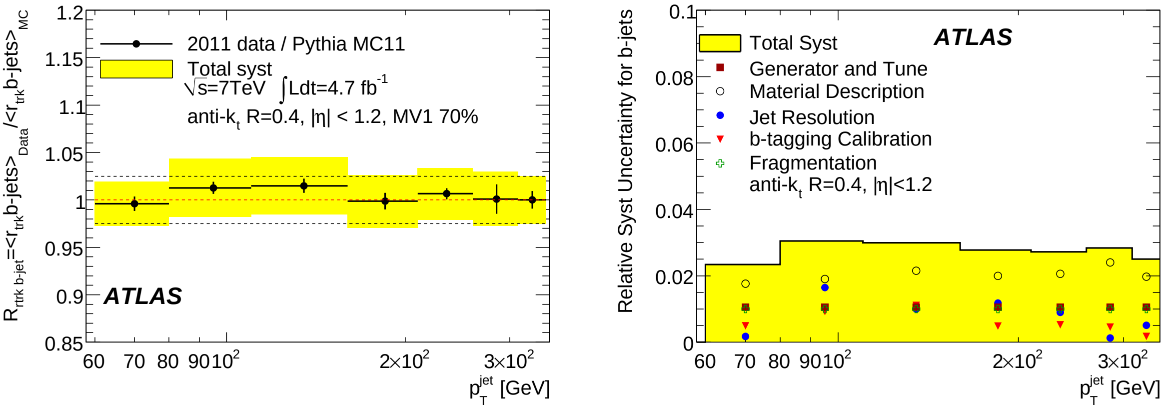
$$r = \frac{|\sum \vec{p}_{\mathrm{T}}^{\,\mathrm{track}}|}{p_{\mathrm{T}}^{\mathrm{jet}}} \tag{3.11}$$

This is obtained for jets that have been b-tagged with the MV1 algorithm with a weight $w_{\mathrm{MV1}} > 0.6$, corresponding to the 70 % b-tagging efficiency discussed in section 2.7.6. The r parameter is evaluated for data and MC and the ratio between both can probe the calorimeter energy scale to b-jets. Figure 3.20 shows the double ratio $r_{\mathrm{data}}/r_{\mathrm{MC}}$ for di-jet events (a) and $t\bar{t}$ events (c), obtained for PYTHIA6 simulated samples with data from 2011 with $\sqrt{s} = 7$ TeV. The yellow band represents the total uncertainty on the r parameters, which include uncertainty terms from b-tagging and jet energy resolution, in addition to those referred above. On the right side of figure 3.20, the b-jet energy scale systematic uncertainty is decomposed and shown separately with different markers for the di-jet sample (b) and for $t\bar{t}$ events (d). The total systematic uncertainty, obtained from the quadratic sum of the individual components, is about 2 % for di-jets and 3 % for $t\bar{t}$ events. The dependence on the jet $\mathrm{p_T}$ is small.

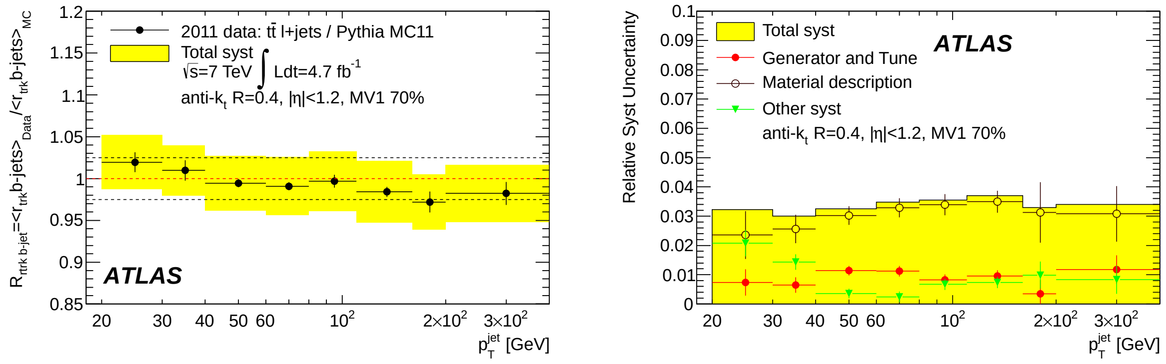
3.7.3. Uncertainty on the parametrized simulation of the calorimeter

As discussed in section 2.8, full simulation, in which GEANT4 is used to describe the full detector, takes too long and the ATLASFASTII, or simply AF, with a parametrization of the calorimeter response was used to reduce the processing time per event and increase the available statistics.

Since the calibration factors of the jet energy scale and their uncertainties were evaluated from full detector simulated samples, an additional uncertainty is taken into account for the difference in the jet energy response when the fast parametrised simulation is used. This uncertainty, evaluated with a di-jet sample, is required for analyses that use the fast simulation. It has a negligible contribution, except for the transition region between the endcap and the forward calorimeters ($|\eta| \approx 3.2$) where its value is 3.5 % for 2011 [122]. For the final 2012 calibration, figure 3.17 shows that this systematic uncertainty is smaller than



(a) di-jets results (b) di-jets systematics



(c) $t\bar{t}$ results (d) $t\bar{t}$ systematics

Figure 3.20.: Calorimeter jet energy measurement (left) and systematic uncertainties using tracks (right) in b-tagged jet event [71] for di-jet (a-b) and $t\bar{t}$ (c-d) events.

0.3 %, independently from the transverse momentum and pseudorapidity of the jet.

The difference in the jet energy response between the full detector simulation, FS, and that obtained with the parametrization of the calorimeter response, AF, observed for EM+JES is about $1-2$ % for low-p_T and smaller than 0.2 % for $p_T > 80$ GeV, as shown in figure 3.21. The jet energy response with the AF simulation is always larger than the response with FS simulation and therefore the difference is negative.

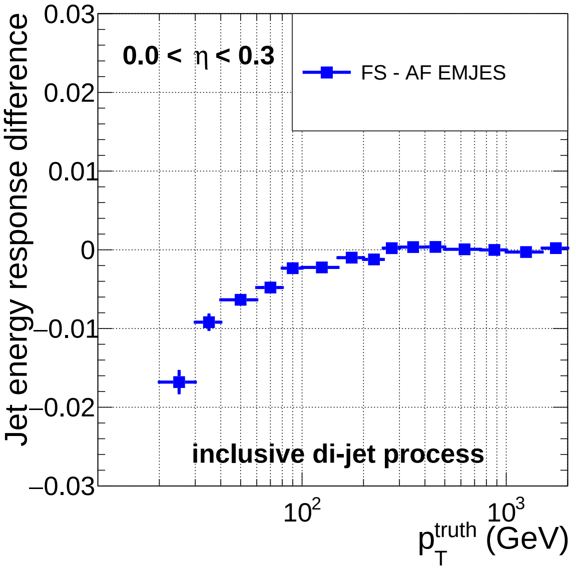


Figure 3.21.: Jet energy response difference between the full simulation (FS) of the detector and the fast simulation (AF) with the parametrization of the calorimeter as a function of p_T^{truth} for jets calibrated at the EM+JES scale.

3.8. Single hadron response

The calorimeter response to hadrons was tested for the first time in test beam from 2000 to 2003, in which the ATLAS sub-detectors were tested individually with several beams [123, 124]. Additionally, a full slice of the ATLAS detector was also exposed to high energy particles in a combined detector test beam (CTB) in 2004 [125, 126, 127]. The calorimeter energy response was measured with electrons and charged pions, setting the base for the electromagnetic and hadronic energy scale. The calorimeter energy response to isolated hadrons is also called the single hadron response and was measured with charged pions with a pseudorapidity in the 0.2 to 0.65 window and a momentum from 20 GeV to 350 GeV, using CTB.

The uncertainty on the jet energy scale discussed in the previous sections can be used for jets with p_T up to 1.7 TeV. However, at the beginning of data taking in 2009 and 2010, there was not enough statistics of Z+jet, γ+jet, di-jet or multi-jet, to evaluate this uncertainty. At the time, the uncertainty was instead measured using data-MC comparisons from the calorimeter response to hadrons. Jets are composed of hadrons and it is possible to extrapolate the uncertainty of the charged hadrons to the jet uncertainty. Proton-proton collisions were used to obtain the uncertainty for charged hadrons for the low-momentum range: up to 10 GeV. When combined with the results of CTB, it was possible to obtain the jet energy scale uncertainty for jets with p_T up to the TeV scale, early in the data taking. To date, the methods to determine the jet scale uncertainty have limited statistics in the jet very high p_T range (see figure 3.14 (c)) and the uncertainty is still obtained using the results from single hadron response. In figure 3.17 (a), there is a sudden change in the derivative of the function describing the "absolute *in situ* JES" (red line) located at 1.7 TeV, in which the single hadron response uncertainty takes over.

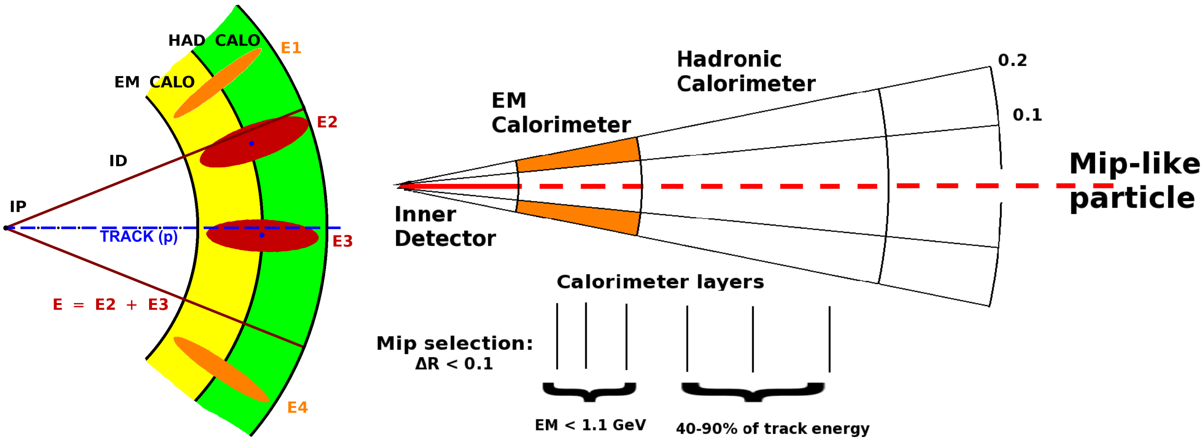
Section 3.8.1 describes the method used for obtaining the single hadron response in collisions. This measurement had low statistics of isolated tracks in the momentum range between 10 GeV and 20 GeV. Part of the work done for this thesis was the development of a trigger that would select such tracks and would be able to extend the method beyond this momentum range. That work is discussed in section 3.8.2 and allowed a direct comparison of the response between CTB and that obtained with collisions, validating the full p_T range of the jet energy scale uncertainty with the single hadron response. Additionally, this trigger was able to collect enough data to validate the method also in a high-pileup regime. Section 3.8.4 discusses this validation together with the results of the single hadron response method.

3.8.1. Method description

The single hadron response is evaluated through the comparison between the energy deposited in the calorimeter (E) by isolated hadrons and the precise measurement of its momentum (p) in the inner detector [128]. The distribution of the energy to momentum ratio is then assessed for different momentum and pseudorapidity bins. The single hadron response is also referred to as $\mathrm{E/p}$. Data and Monte Carlo are compared to understand the accuracy of the detector simulation.

The most interesting events to study the $\mathrm{E/p}$ observable are those in which a collision occurs. During 2009 and 2010 data taking, most of the LHC bunches were empty. Events were triggered if there was a simultaneous hit on both Minimum Bias Trigger Scintillators (MBTS), located just before the endcap of the electromagnetic calorimeter. During 2011 and 2012, these Scintillators saturated, as more and more collisions took place. Almost every LHC bunch was filled, and a random trigger was sufficient to select events with filled bunches. This trigger chain started at L1 with a random trigger called RD0_filled. Both L2 and EF had no algorithm run. A large prescale was set at L1 to limit the rate. This trigger chain was denominated rd0_filled_NoAlg and called *random* trigger in the remaining of the thesis. Events are further selected for analysis if there is at least one vertex, with more than three reconstructed tracks with $p_T > 500$ MeV.

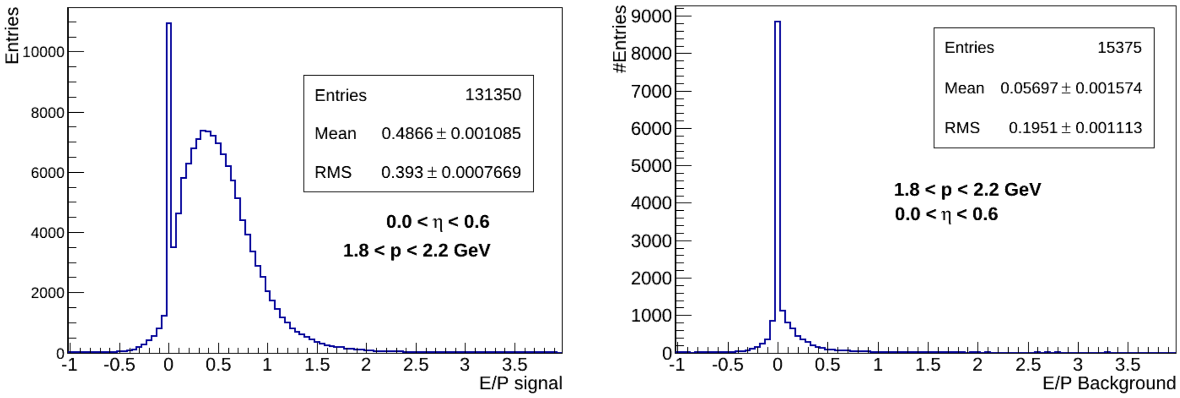
The calorimeter response to single hadrons was studied only for calorimeter isolated charged particles. This isolation is tested by propagating all the tracks reconstructed in the inner detector to the second layer of the electromagnetic calorimeter. A track is accepted for the E/p estimation if the impact point of the considered track is further than $\Delta R = 0.4$ to the closest track extrapolation. Once the isolated tracks are found, some additional quality criteria are required. The track should have at least one hit in the pixel detector, at least six hits in the semiconductor tracker, a transverse momentum above 500 MeV and small impact parameter with respect to the primary vertex: $|d0| < 1.5$ mm and $|z0\, sin\theta| < 1.5$ mm.



(a) (b)

Figure 3.22.: Sketches of the E/p observable definition (a) adapted from reference [129]. Estimation of the neutral background from the E/p observable (b).

After a track is accepted, the energy of all the topological clusters is summed if their centres are inside a cone of $\Delta R = 0.2$ around the extrapolation position of the track in the calorimeter. A sketch of this sum is shown in figure 3.22 (a), where clusters 2 and 3 are inside the cone and their energy is summed: $E = E_2 + E_3$. The E/p observable is obtained from the ratio between this energy and the momentum of the track (p). This quantity is calculated in bins of η and momentum of the track. Figure 3.23 (a) shows the distribution of the E/p observable for tracks with momentum between 1.8 and 2.2 GeV. The spike in the distribution at zero corresponds to tracks that have no associated topocluster in the $\Delta R = 0.2$ cone. This effect is only non-negligible for low momentum tracks.



(a) (b)

Figure 3.23.: Distribution of the raw E/p observable (a) and the background E/p estimation (b) for the tracks with momentum between 1.8 and 2.2 GeV for the pseudorapidity region $0.0 < \eta < 0.6$.

The main background for the E/p measurement is due to neutral particles that leave no track in the inner detector, but their energies are collected in the calorimeter. It is composed mainly, by π^0 mesons. The method developed to subtract this background is based on the identification of charged hadrons that induce late-showering in the calorimeter. Charged hadrons will have low energy deposits in the EM calorimeter and will deposit most of their energy in the hadronic calorimeter. Around them, an annulus is defined: $0.1 < \Delta R < 0.2$. Events considered for the background estimation are required to have less than 1.1 GeV in the electromagnetic calorimeter inside $\Delta R = 0.1$ around the track (*i.e.* centre circle inside the annulus) to resemble the minimum ionising like particle (MIP). The effect of the late-showering is increased by requiring that the hadronic E/p, *i.e.* the ratio between the energy deposited in the hadronic calorimeter and the track's momentum, lies in the window between 0.4 and 0.9. The energy of the clusters from the electromagnetic calorimeter inside the defined annulus is summed and extrapolated to the core cone, $\Delta R < 0.1$, assuming a flat distribution for the energy density of neutral background. The sketch in figure 3.22 (b) shows in orange, the area of the electromagnetic calorimeter region used for the background estimation that is estimated to contain 75 % of the background contribution. The same η and p-bins as before were used to evaluate the background. Figure 3.23 (b) shows the background

E/p estimation distributions for the same pseudorapidity and momentum region as before. Only about one tenth of the total tracks can be used to evaluate the background E/p. The corrected value of the E/p in each $\eta - p$ bin is calculated by the difference between the averages of the two distributions:

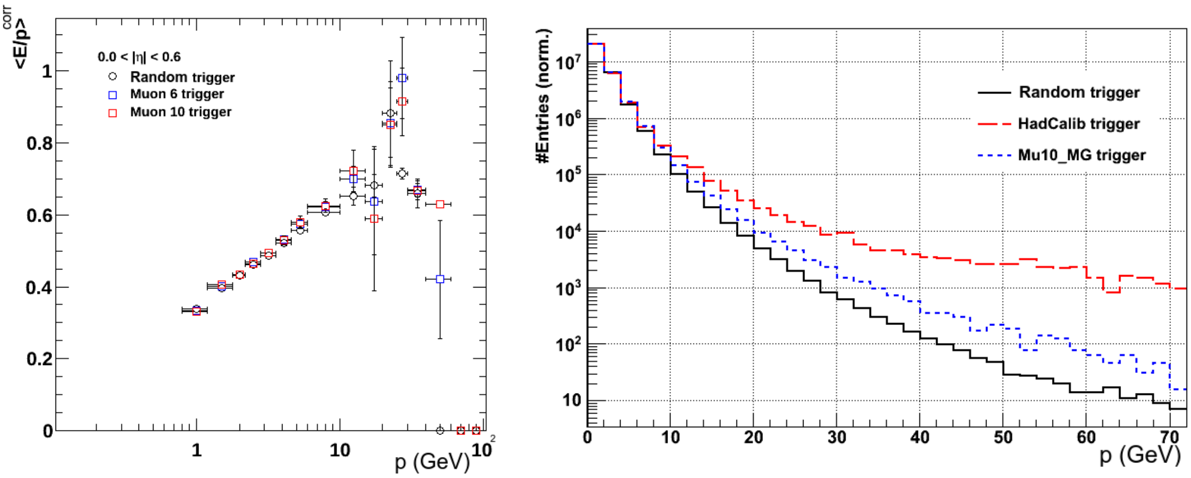
$$\langle E/p \rangle^{corr} = \langle E/p \rangle^{raw} - \langle E/p \rangle^{BG} \tag{3.12}$$

3.8.2. Trigger studies for E/p

The E/p study suffers from low statistics for tracks with momentum above 10 GeV when using the random trigger, especially for the background measurement. One objective of this thesis work was to search for different triggers to enhance the content of high-p hadrons. Several already existent triggers were evaluated to try to provide unbiased results for E/p, increasing the statistics.

The first possibility considered was muon triggered events using the existent thresholds of 6 and 10 GeV. For this study, the isolated tracks were selected if they were separated by more than $\Delta R = 0.4$ from the muon that triggered the event. Figure 3.24 (a) shows the E/p after background correction for the minimum bias trigger and the muon triggers with the two referred thresholds. There is a systematic bias of the corrected E/p measurement for tracks with $p > 5$ GeV. Figure 3.24 (b) shows the distribution of the momentum of the tracks used for E/p for the minimum bias trigger (black) and the muon trigger (blue). Since there was a difference between the E/p corrected values for the default and the muon triggers and the gain in high momentum tracks was not significant, the muon triggers were abandoned.

Another possibility was the usage of a tau-based trigger. The L1 required a hadronic tau with an energy threshold of 5 GeV. At L2, a feature extraction algorithm, called reverse ROI, retrieved the detector information from a square window of 0.6×0.6 in the $\eta - \phi$ phase space, opened in the opposite direction of the L1 tau and the trigger hypothesis algorithm searched for a track with a transverse momentum larger than 9 GeV. The top sketch of figure 3.25 (a) shows this L2 requirement. The events in which this high-p_T track was found would be sent to EF, where two isolation regions (bottom sketch) were defined. No other track with p_T larger than 1 GeV could be found within a distance of $\Delta R = 0.1$ from the track found at L2. The scalar sum of the p_T of the tracks in the region $0.1 < \Delta R < 0.3$ could not be larger than 10 % of the p_T of the selected track. The reverse ROI algorithm could be applied with the opposite η and ϕ direction or just ϕ direction, and it formed the

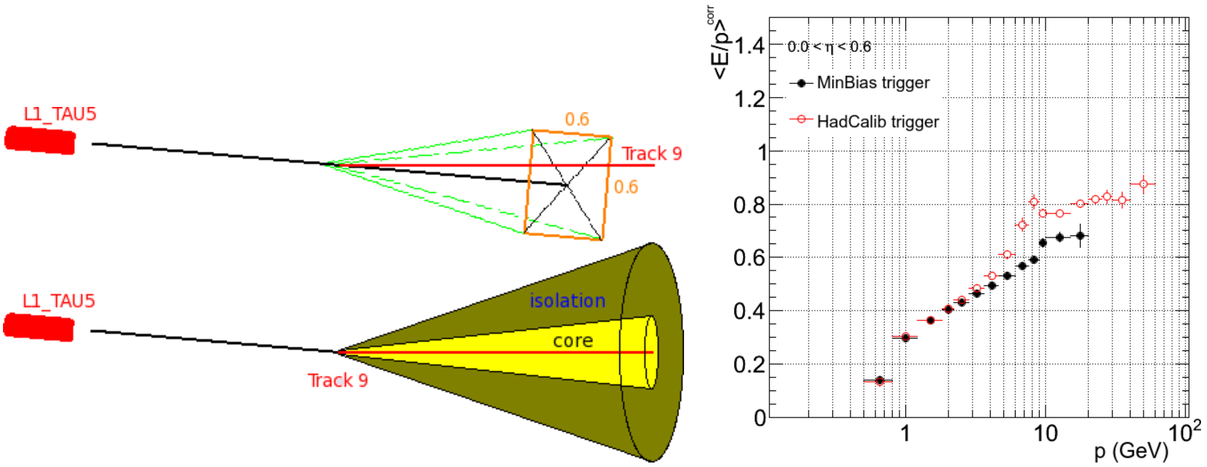


(a) (b)

Figure 3.24.: Corrected E/p as a function of the track momentum in the most central pseudorapidity region for the minimum bias trigger in black and the muon trigger with 6 GeV (10 GeV) threshold in blue (red) (a). Momentum distribution of the tracks chosen for E/p studies using the minimum bias, the muon and the hadCalib triggers shown by the black, red and blue lines, respectively, with the number of entries normalised (b).

hadCalib_trk9 and hadCalib_trk9Phi triggers, respectively. The red line in figure 3.24 (b) shows the momentum distribution of the tracks selected by these triggers, with a much larger increase of high-p tracks. However, the hadCalib trigger introduced a bias once more, as shown in figure 3.25 (b). It shows different E/p corrected values for minimum bias (black) and hadCalib (red) triggers, with the most significant observed for tracks with about 9 GeV, where the trigger should start improving the results.

Although the hadCalib trigger inserted a bias in the E/p analysis, it increased the statistics as intended and an extra effort was made to understand the origin of the bias. The possible sources of the bias were the tracking and isolation in the EF, the tracking and isolation in L2 and the L1 tau condition. Three new triggers were developed to test the origin of the bias, using a new hypothesis algorithm at EF. This new algorithm extrapolated the tracks inside the region of interest to the second layer of the calorimeter for the isolation to be evaluated there, replacing the track isolation from the initial hadCalib triggers. The objective was to have the trigger with the same isolation of the E/p analysis, discussed in section 3.8.1.

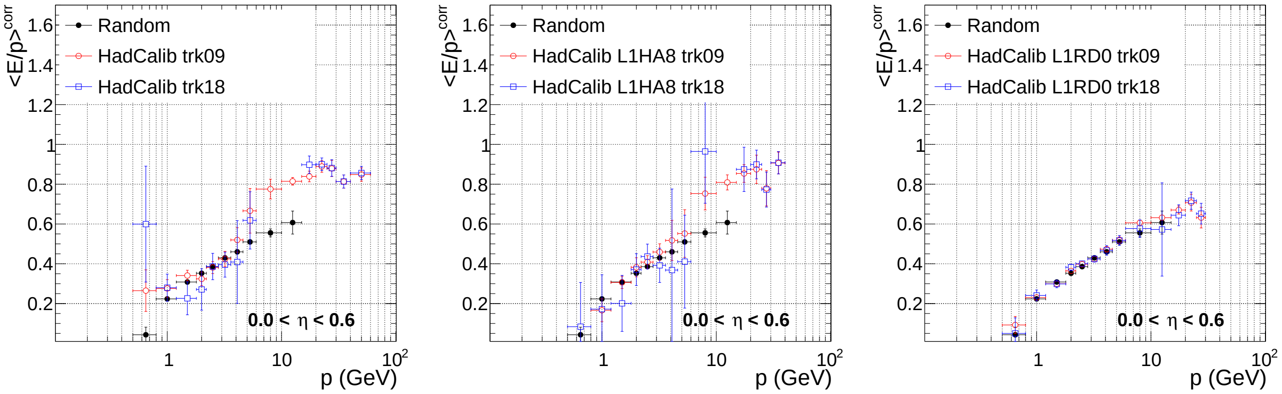


(a) (b)

Figure 3.25.: Sketches of the steps of the hadCalib trigger (a). Corrected E/p as a function of tracks momenta for the minimum bias trigger (black closed circles) and the hadCalib trigger (red open circles) (b).

The first trigger only had the new EF hypothesis algorithm applied, and the tau trigger threshold increased to 8 GeV, using the same name: hadCalib_trk9. A second track threshold of 18 GeV was used to increase further the high-p statistics: hadCalib_trk18. Figure 3.26 (a) shows the corrected E/p as a function of the momentum of tracks for the beginning of data taking in 2012. Although with reduced statistics, this figure shows clearly that the bias is still present with this trigger. The second trigger chain used the same criterium at L1, removed the L2 condition and applied the new hypothesis algorithm at EF. The trigger chains were denominated hadCalib_L1HA8_trk9 and hadCalib_L1HA8_trk18, and their results are shown in figure 3.26 (b). The bias still appeared with this trigger.

The last created trigger started at L1 trigger by selecting filled bunch crossings randomly, RD0_Filled, the same seed as the random trigger used for the default E/p analysis. At L2, the trigger uses no algorithm, since there is no object to select a region of interest. At EF, when the event builder is run, a feature extraction algorithm reconstructs all tracks with p_T larger than 500 MeV, referred full-scan tracking, and applies the new EF hypothesis algorithm to them. The full-scan tracking algorithm with such a low-p_T threshold was implemented for the first time for this trigger. Finally, the trigger accepts the event if there is any calorimeter isolated track in the event with $p_T > 9$ GeV. Like for the seeded-based triggers, another threshold of 18 GeV is chosen to increase high-p statistics. These



(a) hadCalib (b) hadCalib_L1HA8 (c) hadCalib_RD0

Figure 3.26.: Corrected E/p as a function of the tracks momenta for the random trigger (black closed circles) and the hadCalib triggers with a track threshold of 9 GeV (red open circles) and 18 GeV (blue squares) for the three improved hadCalib triggers.

triggers were named hadCalib_L1RD0_trk9 and hadCalib_L1RD0_trk18. Figure 3.26 (c) show the preliminary studies from the beginning of 2012 data taking for the corrected E/p measurement as a function of the momentum of the tracks for the random trigger and the two L1RD0-based hadCalib triggers. The differences between random and these new triggers were negligible, and therefore it was concluded that the bias originated from the L1 condition. Looking only at this preliminary data, one sees that the hadCalib triggers already introduce two new p-bins compared to the random triggers. Figure 3.27 shows a diagram for the L1RD0-based hadCalib trigger.

For the rest of the year, the two seeded triggers were discarded and the L1RD0-based hadCalib triggers were kept. In the remaining of this section, the term “hadCalib” trigger refers to the L1RD0-based triggers.

3.8.3. Performance of hadCalib in E/p measurement

A total luminosity of 911 nb^{-1} and 1037 nb^{-1} was used to study the trigger performance for the 9 GeV and 18 GeV trigger thresholds, respectively. Additionally, 38.71 nb^{-1} of low-pileup proton-proton collisions selected with the random trigger were also used for comparisons. Figure 3.28 shows the prescales applied to the triggers used in the E/p analysis during this period as a function of the runs considered. L1 random trigger, shown in green,

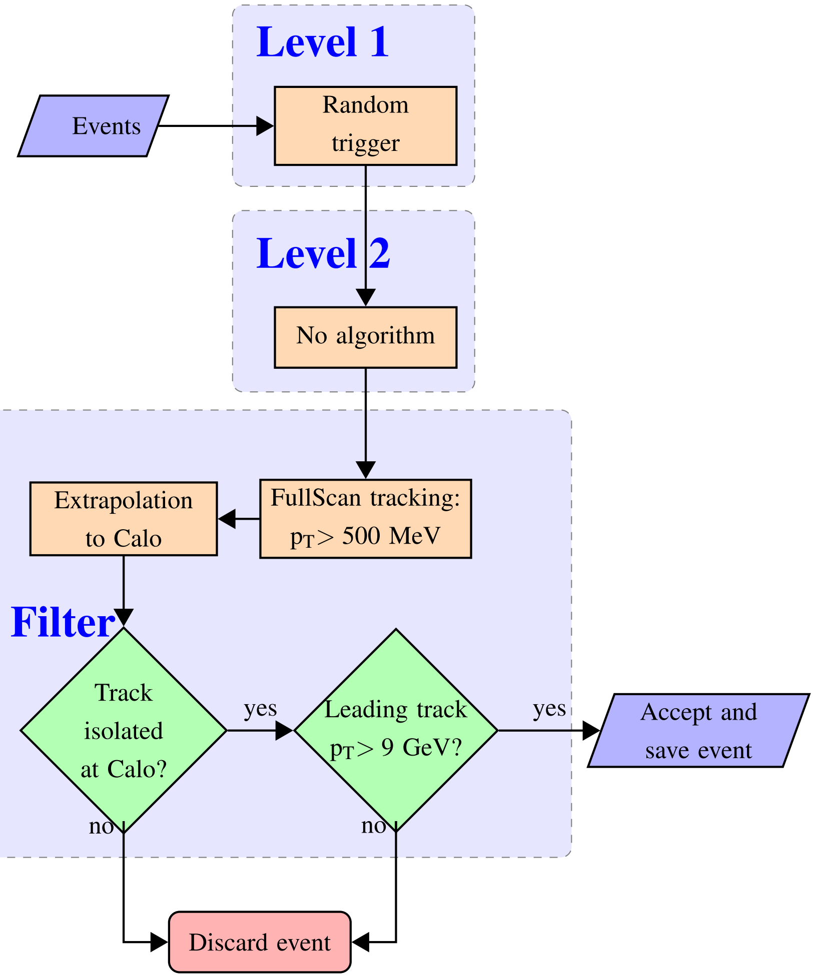


Figure 3.27.: Diagram of the L1RD0-based hadCalib trigger algorithm with a 9 GeV threshold. The other trigger chain is obtained just by changing the track $\mathrm{p_T}$ threshold.

had very large prescales up to the several tens of thousand, as referred in section 3.8.1, to reduce the rate of these triggers. The two hadCalib triggers used a prescale of 1 up to run 95 and then they started to use a prescale of 2 in the beginning of each run, in which the pileup was larger, and then going back to 1. A similar approach was used on the prescale for the L1 random and the EF random.

Figure 3.29 (a) shows the corrected average $\langle E/p\rangle^{corr}$ as a function of the single hadron's momentum for the random trigger, for the hadCalib trigger with threshold 9 GeV and for the hadCalib trigger with threshold 18 GeV. There is a good agreement between the $\langle E/p\rangle^{corr}$ measured with the three different triggers, demonstrating that the two hadCalib

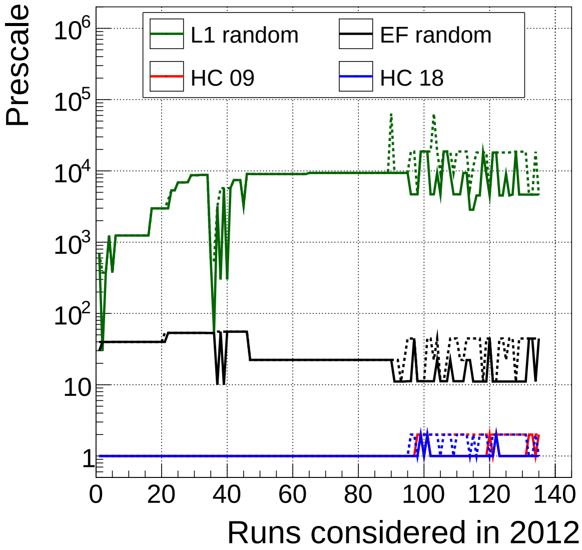
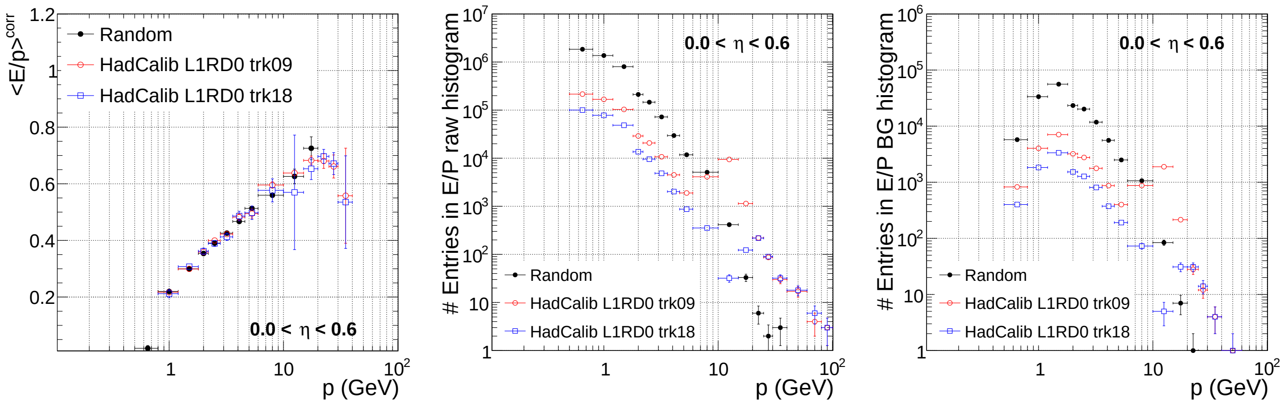


Figure 3.28.: Prescales applied during the runs considered for the E/p analysis in this thesis for L1 random (green), EF random (black), and hadCalib with tracking thresholds of 9 (red) and 18 GeV (blue). The solid (dash) line represent the minimum (maximum) prescale applied for each run.

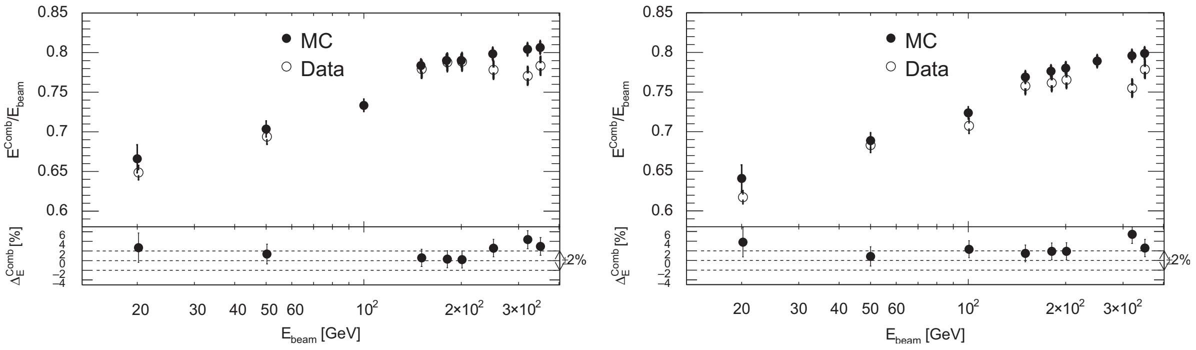
triggers do not introduce any bias in the E/p measurement. Figures 3.29 (b) and 3.29 (c) show the distribution of the number of entries in raw and background histograms, respectively, for the most central pseudorapidity region $|\eta| < 0.6$. The results for other detector regions are similar and are presented in the appendix A. These triggers increase the number of events accepted for the single hadron response evaluation with momentum above 10 GeV by a factor of about 25, achieving their objective.

The value obtained for the E/p measurement with CTB was between 0.63 and 0.67, depending on the pseudorapidity, for a pion beam with $E = 20$ GeV [127] and is compatible with the values obtained in proton-proton collisions. Figure 3.30 shows the ratio between the reconstructed energy at the calorimeter and the beam energy from CTB as a function of the beam energy, for the two extreme pseudorapidity tested: $\eta = 0.20$ (a) $\eta = 0.65$ (b). Additionally, although the second energy beam from CTB (50 GeV) is not reached in collisions, the new hadCalib triggers can already probe the 20-30 GeV region with a $\langle E/p \rangle^{corr} = 0.70$, which was the results obtained for the second energy beam from CTB [127]. The next E/p measurement for the 30-40 GeV momentum range starts to have a very large statistical uncertainty, as the number of accepted tracks for the background estimation fall below ten.



(a) (b) (c)

Figure 3.29.: Corrected average of the E/p distribution for the random (black closed circles), hadCalib_tr9_L1RD0 (red open circles) and hadCalib_tr18_L1RD0 (blue squares) triggers as a function of the single hadron momentum (a). Statistics available in each bin of the $\langle E/p^{raw}\rangle$ (b) and $\langle E/p^{BG}\rangle$ (c) histograms.



(a) $\eta = 0.20$ (b) $\eta = 0.65$

Figure 3.30.: E/p measurement in CTB as a function of the beam energy for two pseudorapidities [127]

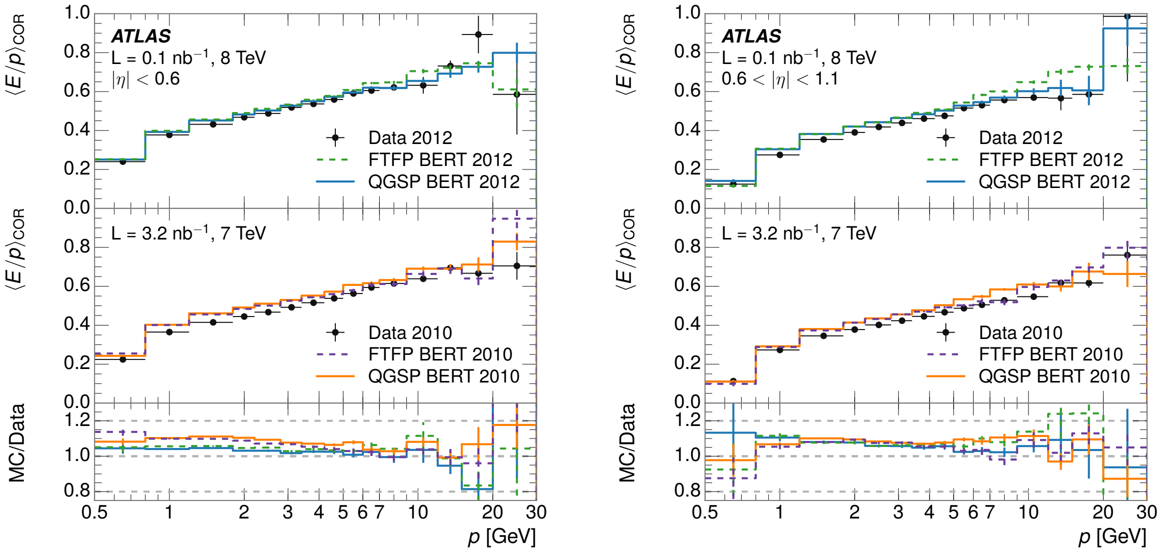
.

Thanks to the development of these triggers, the single hadron response is now validated in the momentum range near the first beam intensity from CTB for proton-proton collisions, and the jet systematic uncertainty was obtained for the full p_T-range.

The hadCalib triggers were also used to obtain the single hadron response in events with a non-negligible pileup environment, which is discussed in section 3.8.4.

3.8.4. Single hadron results

The single hadron response was studied for pileup-free environments using the minimum bias triggers (MBTS and random), which occurred in 2010 and in one run in the beginning of 2012 [130] with an integrated luminosity of 3.2 nb^{-1} at $\sqrt{s} = 7$ TeV and 0.1 nb^{-1} at $\sqrt{s} = 8$ TeV, respectively. The $\langle E/p \rangle^{corr}$ for the two most central $|\eta|$ regions is shown in figure 3.31. The results obtained for real data are compared to the two physics lists introduced in section 2.8: QGSP+BERT and FTFP+BERT [130]. The bottom panel in the figure shows the ratio between $\langle E/p \rangle^{corr}$ in data and MC, for the two physics lists considered. There was a large improvement on the physics lists models from 2010 to 2012, in particular in the most central $|\eta|$ region, resulting in an improvement of the data-MC disagreement from 10% in 2010 to 5% in 2012. In the next $|\eta|$ region ($0.6 < |\eta| < 1.1$), the physics lists models for the two years present the same discrepancy to data.



(a) (b)

Figure 3.31.: $\langle E/p \rangle^{corr}$ as a function of the single hadron momentum for tracks with $|\eta| < 0.6$ (a) and $0.6 < |\eta| < 1.1$ (b) [130]. These results are shown for 2012 in the top panel and 2010 in the middle panel, for data (black markers), and two MC simulation models: FTFP+BERT (dashed line) and QGSP+BERT (filled line). The bottom panel shows for both cases the ratio between each simulation and the corresponding data.

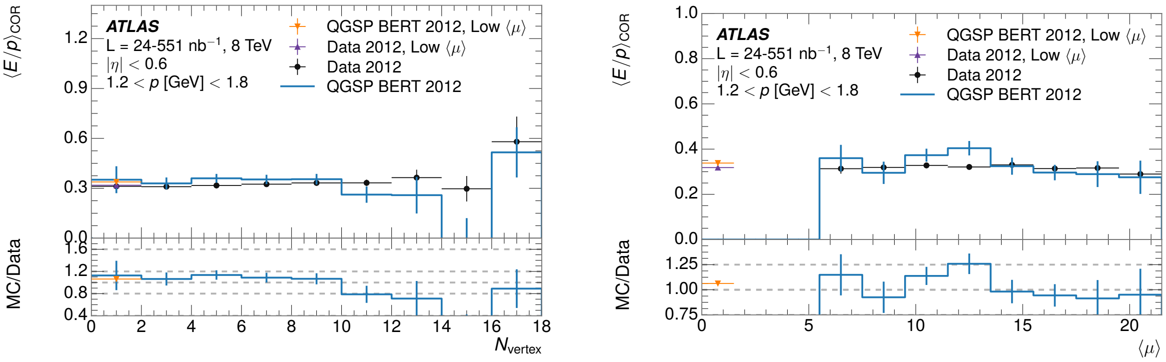
The extrapolation of the energy response uncertainty from the single hadrons to a jet environment requires the understanding of some additional calorimeter energy responses systematic sources, such as the difference in the response between different hadrons and between isolated and non-isolated hadrons. These were studied using charged pions and (anti-)protons produced in K_S and (anti-)Λ decays without isolation requirements first and then requiring isolation for the K_S and (anti-)Λ particles but not for their decay products [131]. Further knowledge is needed for calorimeter response difference between:

- positive and negative charged hadrons;
- using or not events in which there was no cluster associated with the track (the $\mathrm{E/p} = 0$, shown in figure 3.23);
- including or not the estimate of the pileup noise when building up the clusters;
- events with and without pileup.

For the last comparison, the hadCalib trigger played also an important role, as it run also in events with a non-negligible pileup environment. This study used an integrated luminosity of 499 nb^{-1} and 551 nb^{-1} for the 9 GeV and 18 GeV threshold, respectively, for $p-p$ collisions at $\sqrt{s} = 8$ TeV [130].

The resulted $\langle E/p\rangle^{corr}$ is shown in figure 3.32 (a) and 3.32 (b) as a function of the number of primary vertices, N_{PV}, and the average number of interactions per bunch crossing, $\langle\mu\rangle$, respectively, for the most central pseudorapidity bin and the tracks' momentum ranging from 1.2 to 1.8 GeV. The low-$\langle\mu\rangle$ data, taken with the random trigger, is shown in the first bin. There was no dependence of the $\langle E/p\rangle^{corr}$ with either of the pileup variables. The bottom panel shows the ratio between the data recorded and the QGSP+BERT physics list, with differences up to 25 % resulting mainly from fluctuations in the simulation model: the data collected has a flat response with both pileup variables.

Figure 3.33 shows the average jet energy response (black line) and the systematic uncertainties obtained with the single hadron response method, decomposed in the individual components, as a function of the transverse momentum of the jet. The total uncertainty, shown with the darkest blue shade, results from the quadratic sum of the components. Comparing this result with the combination of the *in situ* techniques found in figure 3.13 (b), the two *in situ* measurements provide compatible uncertainties. The uncertainty with the E/p method was an important validation of the systematic uncertainty derived with the other *in situ* techniques.



(a)

(b)

Figure 3.32.: $\langle E/p \rangle^{corr}$ as a function of the pileup estimators N_{PV} (a) and $\langle \mu \rangle$ (b). The low-$\langle \mu \rangle$ value of E/p, obtained with the random trigger is shown in the first bin [130].

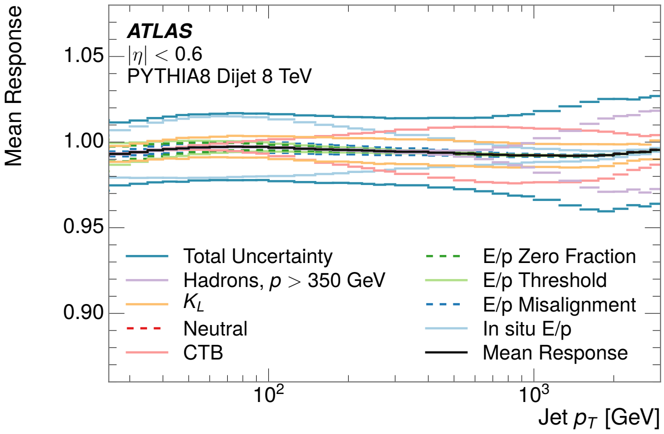


Figure 3.33.: Mean jet response as a function of the jet transverse momentum for the $|\eta| < 0.6$ region in black. The jet energy scale uncertainty is shown for the different sources of uncertainty and for their combination [130].

In the low-$\mathrm{p_T}$ regime, $\mathrm{p_T} < 250$ GeV, the E/p technique in collisions dominates the systematic uncertainty. In the intermediate region the uncertainty obtained with CTB is the most important, up to 1.5 TeV, where the extrapolation from CTB to charged hadrons with larger momentum dominates. As discussed already, the default *in situ* methods only allowed the uncertainty to be obtained up to 1.7 TeV. Above this limit, the single hadron response sets the jet energy scale uncertainty at about 3.5 %, for jets with larger $\mathrm{p_T}$. Figure 3.33 shows the uncertainty for jets with p_T up to 3 TeV, almost 40% of the maximum available energy from the proton-proton collision: $\sqrt{s} = 8$ TeV.

3.9. The global sequential calibration

The jet default calibration, EM+JES, was discussed in section 3.1.2. The global sequential calibration (GSC) was introduced to reduce the discrepancy in the jet energy response between jets initiated by quarks and jets initiated by gluons [115, 132], shown in figure 3.18 (a). The objective of this calibration was to remove the dependence of the jet energy response on some jet properties.

In this section, the GSC calibration method and its performance for different jets are presented (sections 3.9.1 to 3.9.3) followed by a discussion on the systematic uncertainties. The section ends with a detailed evaluation of the improvements on the invariant mass of b-jet pairs using GSC.

3.9.1. Introduction

Even after the calibration of jets with EM+JES, as described in previous sections, the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distribution is still dependent on different jet properties. The global sequential calibration (GSC) was designed to remove the residual dependence of the jet energy scale on certain properties of the jets, applying sequential corrections as a function of these jet properties. The numerical inversion technique, used for the evaluation of the EM+JES correction factors, is also used for the derivation of the GSC calibration constants.

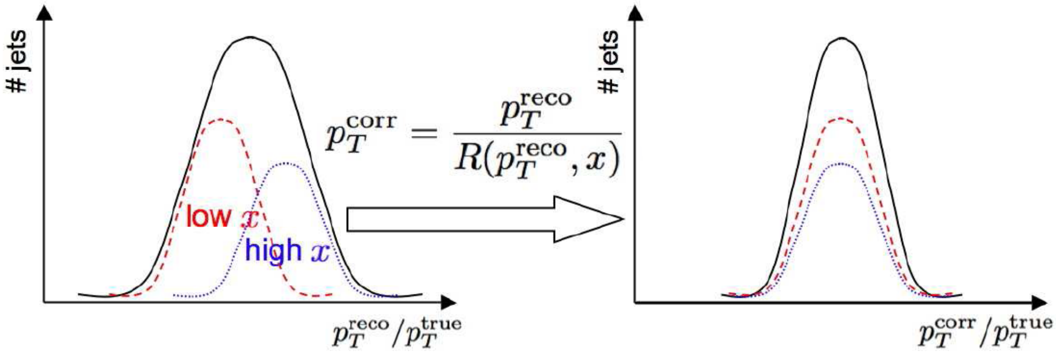


Figure 3.34.: Illustration of the GSC method.

The GSC calibration principle is illustrated schematically in figure 3.34. Suppose that the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distribution depends on a certain property “x” and it is obtained independently for two subsets of jets: one with low values of “x”, in red, and the other one with high values of “x”, in blue. Suppose also that the “low x” distribution happens to have an average value lower than the one marked “high x”. These two distributions sum up to the distribution with a large width, shown in black in figure 3.34. The GSC method intends to correct the two distributions in such a way that they will be aligned to the initial average jet energy response, by applying a multiplicative factor in each bin of “x”. After GSC, the sum of the distributions becomes narrower and therefore the resolution as a function of the jet $\mathrm{p_T}$ is improved.

The variables used for the GSC calibration are now described in the same order as they are derived and applied. The first two properties are the energy fraction that the jet deposits in the first layer of the hadronic barrel calorimeter, f_{Tile0}, and in the last layer of the electromagnetic calorimeter, f_{em3}. The former is only used within the pseudorapidity coverage of TileCal, $|\eta| < 1.7$, whereas the latter within the coverage of the electromagnetic calorimeter up to $|\eta| = 3.5$. These two properties of the jet measured in the calorimeter are sensitive to differences in the electromagnetic and hadronic components of the jet and to energy losses in the cryostat of the Liquid Argon electromagnetic calorimeter. Both fractions are obtained at the electromagnetic energy scale, *i.e.* before the jet energy scale calibration. The other two properties of the jet for which the dependencies are removed, use the information of the inner detector to get track-based jet measurements. The number of tracks associated with the jet, *nTrk*, is sensitive to the number of charged particles in the jet. The other track-based property is referred as the track width, $TrkWidth$, and reflects the transversal distribution of charged particles in the jet. $TrkWidth$ is calculated from the sum of the distances between charged particles (tracks), *i*, and the centre of the jet, *j*, weighted by the fraction of the charged particle $\mathrm{p_T}$ with respect to the jet $\mathrm{p_T}$, as shown in equation 3.13:

$$TrkWidth(j) = \frac{\sum_i p_{\mathrm{T}}^i \, \Delta R(i,j)}{\sum_i p_{\mathrm{T}}^i}, \tag{3.13}$$

The association of tracks to jets uses again the ghost association that was described in section 3.2. These two track based properties are derived for jets with $|\eta| < 2.5$ and are also sensitive to the difference in the jet energy response between quark and gluon initiated jets. For the same energy, light-quark initiated jets are more likely to have a lower number of charged particles associated and their spread within the jet is also expected to be smaller than gluon initiated jets [121]. Jets with a smaller number of charged particles should have a larger contribution from neutral particles. Most of these particles are neutral pions that decay into photon pairs. Since the electromagnetic energy scale is better known and the JES corrections always increase the jet energy response, the light-quark initiated jets will have a larger jet energy response than gluon initiated jets.

3.9.2. GSC derivation and performance for di-jet inclusive MC samples

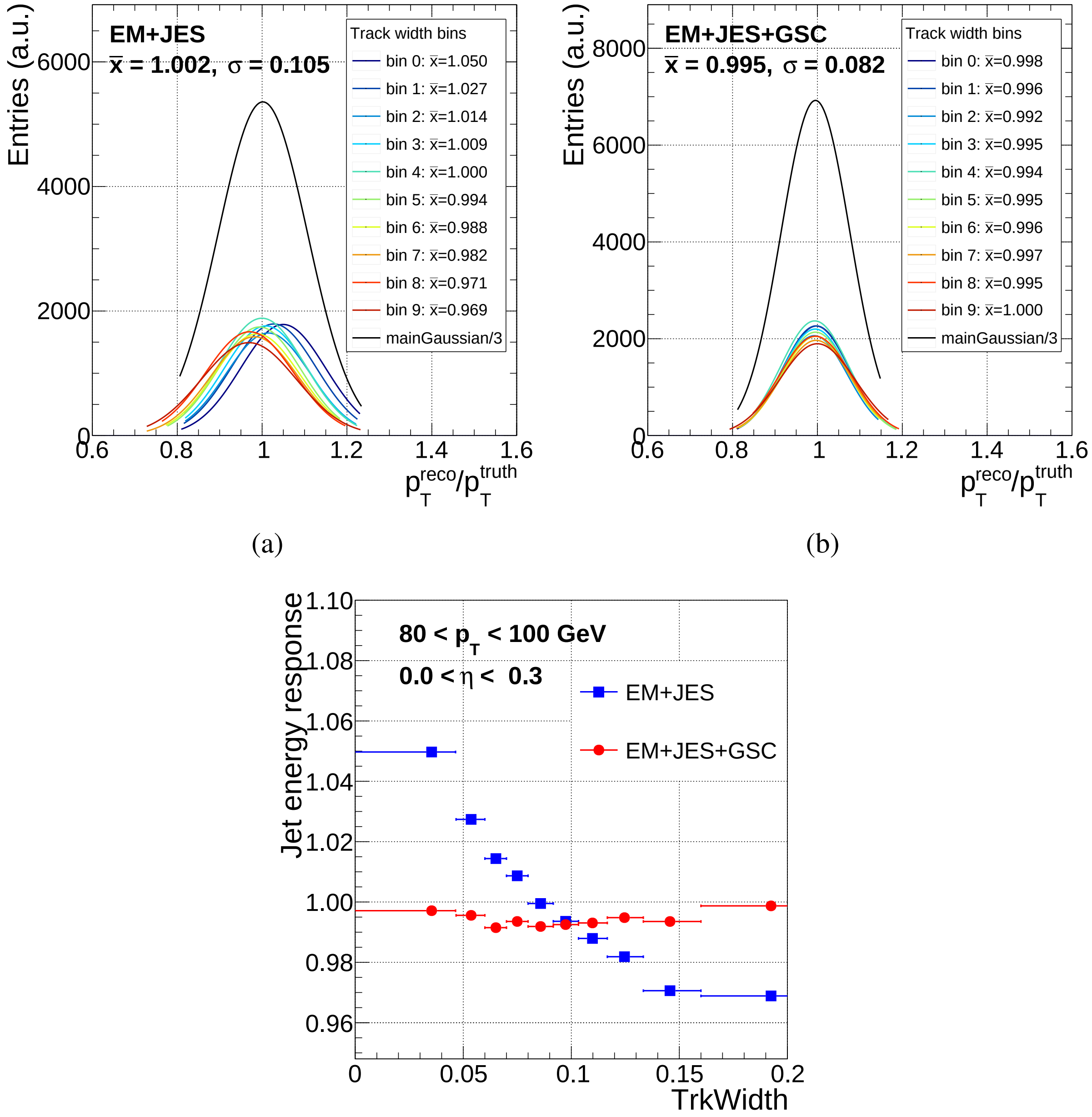
The derivation of the calibration factors for GSC is obtained sequentially using the same technique described in section 3.4.1 but now, instead of obtaining the jet energy response as a function of the transverse momentum of the jet, the jet energy response is obtained as a function of the jet properties described in section 3.9.1.

For the derivation of the GSC calibration factors, η is binned in steps of 0.1 and the bins of $\mathrm{p_T}$ are adjusted to 30 equal logarithmic bins between 20 GeV and 1.2 TeV. A low and a high threshold for each property are selected to have most of the statistics in all $\mathrm{p_T}$ and η bins. Between the two thresholds of each property, 30 bins of the same size are selected. The calibration factor for GSC is obtained from the ratio between the jet energy response combined in the 30 bins of one given property and the jet energy response for the particular property bin for which the calibration factor is being evaluated. This ensures that the overall jet energy response does not change with the GSC calibration. The calibration factors are obtained first for the f_{Tile0} property, using the jet energy response at the EM+JES scale. After this first correction is applied, the jets are calibrated with EM+JES+f_{Tile0}. This is then used as input for the evaluation of the correction factors for the f_{em3} property, obtaining the jets calibrated with EM+JES+f_{Tile0}+f_{em3}. The same procedure is applied sequentially to *nTrk* and to *TrkWidth*. After the four corrections, the jets are calibrated to the EM+JES+GSC scale or just GSC for simplicity.

To study the GSC calibration performance, the same bins of η and $\mathrm{p_T}$ that were shown for EM+JES in figure 3.6 (a), are used. For the closure tests, the jets are divided into 10 bins of each of the properties in order to have the same amount of statistics in each bin. For each of the sub-regions and for each property, the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distribution is obtained and the mean and the width of the distribution is calculated using the same Gaussian fits, as described in section 3.4.1. Then, the jet energy response is obtained as a function of that property.

An example is given for the di-jet inclusive sample in figure 3.35 (a), showing the Gaussian fits of the $p_{\mathrm{T}}^{reco}/p_{\mathrm{T}}^{truth}$ distributions for the ten selected $TrkWidth$ bins for $0.0 < \eta < 0.3$ for $\mathrm{p}_{\mathrm{T}}^{truth}$ in the range between 80 and 100 GeV. This figure uses a rainbow palette for the Gaussian distribution fits using violet for the smallest $TrkWidth$ bin up to the red for the highest bin of $TrkWidth$. The Gaussian fit of the combined distribution of this particular $\eta-\mathrm{p}_{\mathrm{T}}$-bin is shown in black also in the same plot, scaled by a factor of 1/3, for comparison. There is a clear misalignment of the Gaussian fits in each $TrkWidth$ bin with respect to the main Gaussian. Figure 3.35 (b) shows the same Gaussian distributions of the 10 bins of $TrkWidth$ after the application of GSC. The Gaussian functions now became aligned and the sum of all bins is a narrower Gaussian distribution. The same is shown in figure 3.35 (c) with the jet energy response plotted as a function of $TrkWidth$ for EM+JES in blue and for GSC in red. After applying the GSC correction factors, the jet energy response dependence on $TrkWidth$ nearly disappears. The difference observed between the average of the main Gaussian before and after GSC is applied is below 1 %. The improvement observed in the relative resolution is about 20 %.

The jet energy response after each of the sequential calibration steps is plotted for two representative jet transverse momentum ranges, [40,60] GeV and [250,300] GeV in figures 3.36 and 3.37 respectively, as a function of each calibration variables: f_{Tile0} (a), f_{em3} (b), $nTrk$ (c) and $TrkWidth$ (d). Before any GSC correction (blue squares), there is a visible jet energy response dependence on each of the variables. After the first calibration step, shown with the black up triangles, the dependence on f_{Tile0} is removed as shown in figures 3.36 (a) and 3.37 (a). None of the subsequent calibration steps modifies the jet energy response as a function of f_{Tile0}. Variations on the jet energy response when applying the subsequent calibration steps might have been visible if there were correlations between the jet energy response as a function of f_{Tile0} or any of other three variables. These figures show the application of GSC and therefore are considered to be a closure test of the procedure. In figures 3.36 (b) and 3.37 (b), the jet energy response is plotted as a function of the variable f_{em3}. After the first GSC correction factor is applied, there was some reduction in the dependence of the jet energy response with f_{em3}, but only the second correction, shown by the green-down triangles, removes the dependence on this jet property. This might suggest some small correlations between the f_{Tile0} and f_{em3} properties. In fact, they were chosen because they are sensitive to the energy deposited in the cryostat and to the development of the jet shower, so this is expected. The sub-figures 3.36 (c) and 3.37 (c), show the jet energy

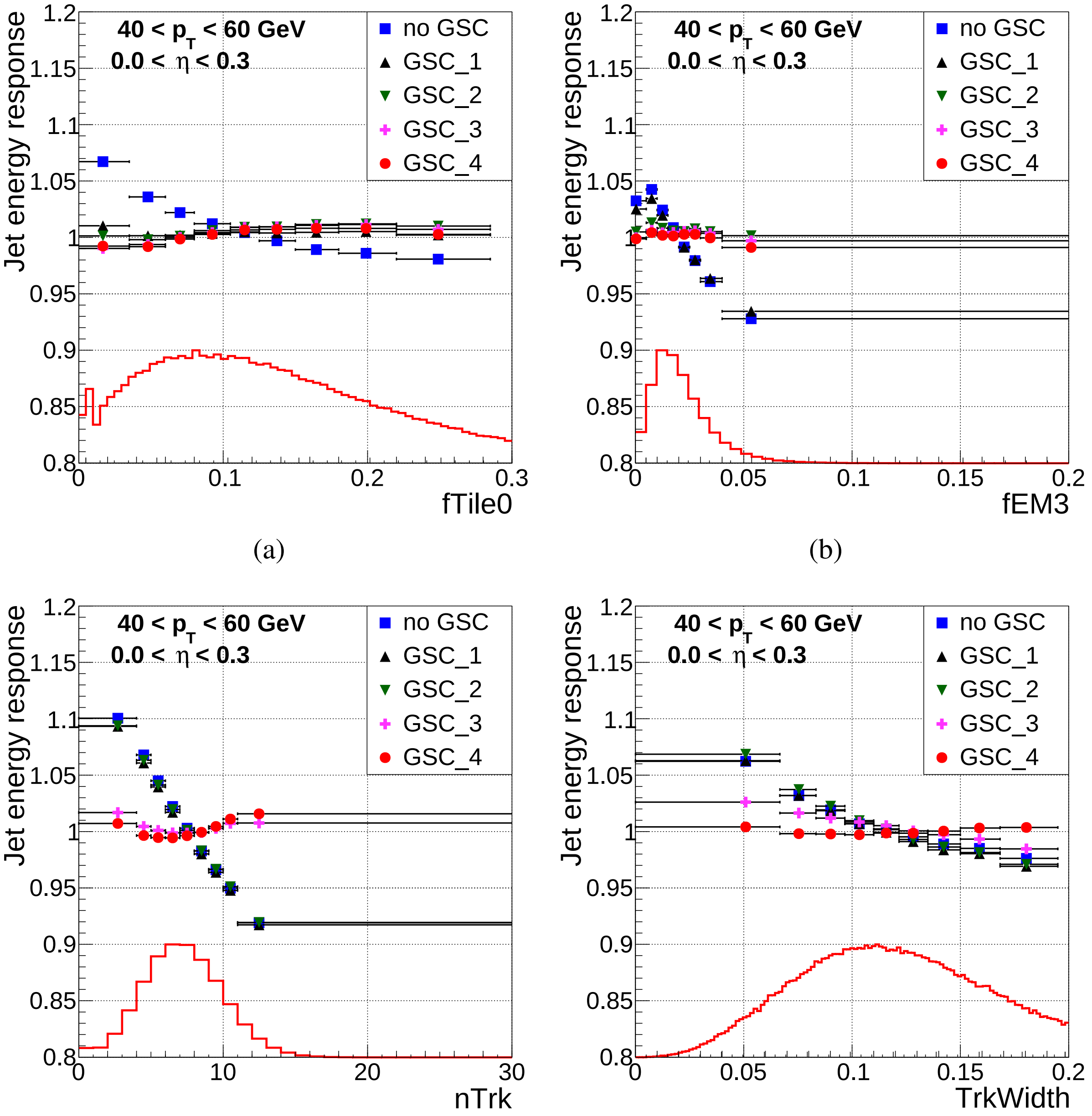


(a)

(b)

(c)

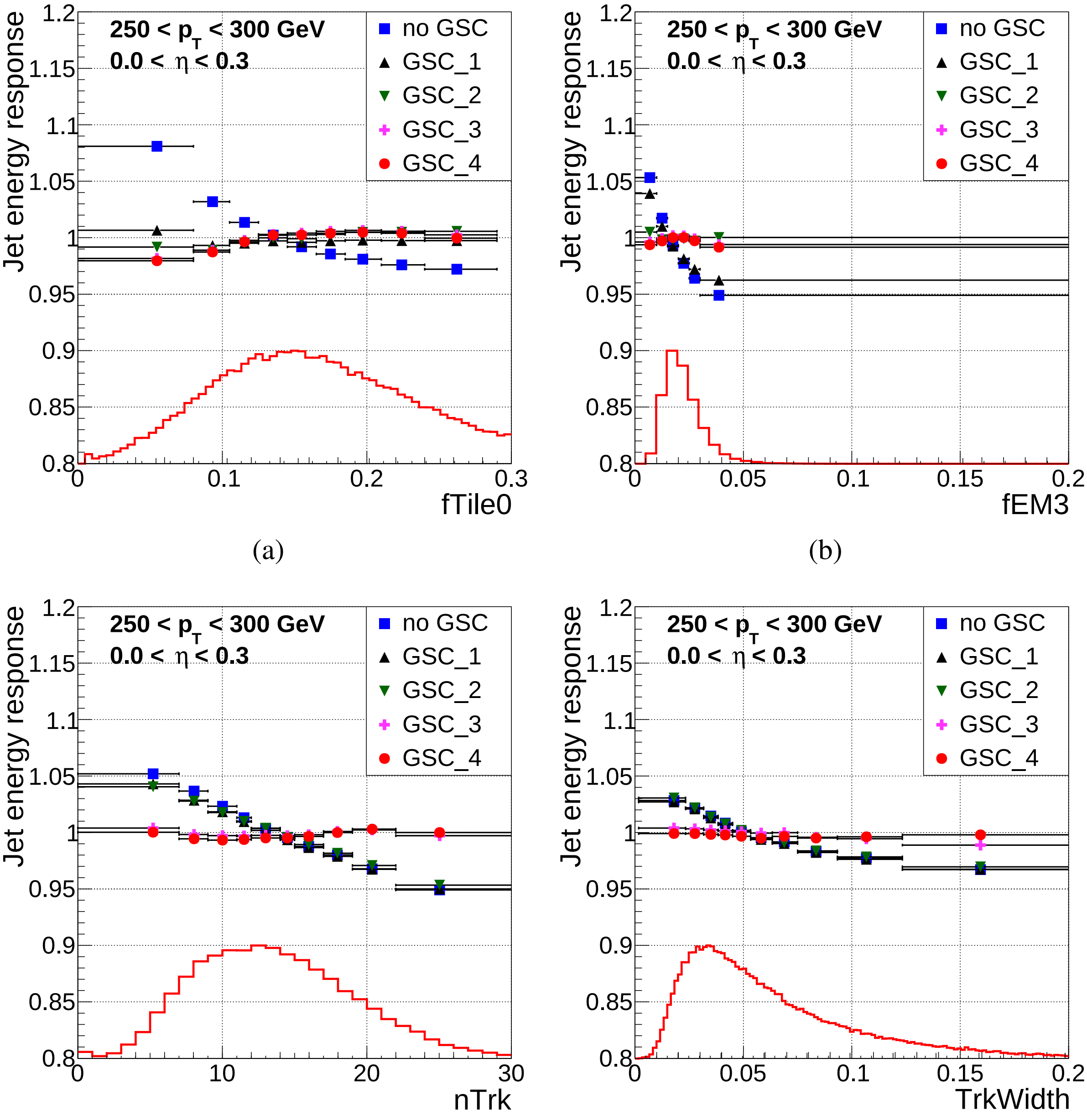
Figure 3.35.: Example of the application of the GSC calibration to correct the jet energy response with respect to *TrkWidth*, using inclusive di-jet events simulated with PYTHIA8. The p_T^{reco}/p_T^{truth} distributions for EM+JES (a) and GSC (b) use a rainbow colour palette for the different bins of *TrkWidth*. Jet energy response as a function of the *TrkWidth* (c), before (blue squares) and after (red circles) applying GSC.



(a) (b)

(c) (d)

Figure 3.36.: Jet energy response as a function of fTile0 (a), fEM3 (b), nTrk (c) and *TrkWidth* (d) for jets with $0.0 < \eta < 0.3$ and $40 < p_{\mathrm{T}}^{truth} < 60$ GeV, using inclusive di-jet events simulated with PYTHIA8. The histogram in red represents the distribution of each of these variables. The blue squares show the jet energy response without applying GSC and the black up triangles, green down triangles, pink crosses and red circles show the jet energy response after applying the first (fTile0), second (fEM3), third (nTrk) and fourth (*TrkWidth*) GSC corrections, respectively.



(a) (b)

(c) (d)

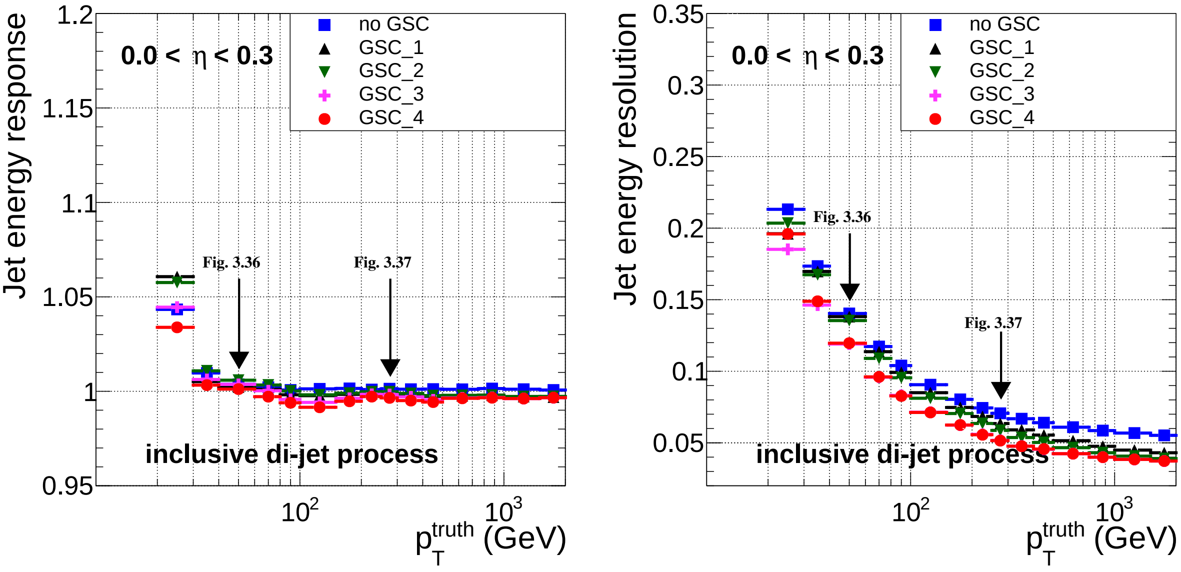
Figure 3.37.: Jet energy response as a function of fTile0 (a), fEM3 (b), nTrk (c) and $TrkWidth$ (d) for jets with $0.0 < \eta < 0.3$ and $250 < p_{\mathrm{T}}^{truth} < 300$ GeV, using inclusive di-jet events simulated with PYTHIA8. The histogram in red represents the distribution of each of these variables. The blue squares show the jet energy response without applying GSC and the black up triangles, green down triangles, pink crosses and red circles show the jet energy response after applying the first (fTile0), second (fEM3), third (nTrk) and fourth ($TrkWidth$) GSC corrections, respectively.

response as a function of the third GSC correction variable, *nTrk*. The effect of the previous corrections adjusts the jet energy response by less than 1 %. The application of the *nTrk* correction, shown with the pink crosses, improves the spread of the jet energy response from almost ±10 % in the $40 < p_T < 60$ GeV bin to less than 2 % and in the $250 < p_T < 300$ GeV from ±5 % to less than 1 %. Finally, in sub-figures (d), the jet energy response is shown as a function of the property *TrkWidth*. Although, the first and second variables contribute almost nothing to the reduction of the dependence with *TrkWidth*, the correction of the third property, *nTrk*, already reduces significantly the dependence on the *TrkWidth* variable, due to correlations between these two variables. The spread of *TrkWidth* with GSC applied with the *nTrk* property is reduced from about 10 % to 5 %. With the application of the last GSC correction, *TrkWidth*, shown by the red circles, the discrepancy in the jet energy response is reduced to less than 1 %. Looking back to the dependences of the jet energy scale with all the GSC corrections, the spread of the red circles is smaller than 1 % in most of the cases and it is always smaller than the initial spread of the blue squares which represent the jet energy response from jets calibrated at the EM+JES scale. In the appendix in figure B.1, two more p_T^{truth} bins and other pseudorapidity region are presented, with similar conclusions.

The overall jet energy response as a function of the truth jet transverse momentum before and after the GSC is shown in figure 3.38 (a) and the jet energy resolution evaluated as described in section 3.4.1 is shown in figure 3.38 (b) for jets in the central pseudorapidity range: $0.0 < \eta < 0.3$. On these plots, the two p_T^{truth} bins discussed previously are indicated with the arrows and the jet energy response and resolution are shown before and after each of the GSC calibration steps. The jet energy response does not change by more than 1 % with GSC, as intended. The jet energy resolution improves about 20 % throughout the p_T^{truth} range with GSC. Similar results are obtained for the other η ranges, which can be seen in the appendix B, figure B.2.

3.9.3. Flavour performance

The detector energy response for jets originating from different flavoured quarks or from gluons can be different, due to differences in the hadronization process that will translate into differences in the jet properties, as for instance the number of charged versus neutral particles produced. This was already observed when the light flavour systematic uncertainty was discussed in section 3.7.1 and the jet energy response was shown as a function of the truth jet p_T in figure 3.18 (a) for light-quark and gluon initiated jets.



(a) (b)

Figure 3.38.: Jet energy response (a) and resolution (b) as a function of the truth jet transverse momentum for inclusive di-jet events simulated with PYTHIA8. The blue squares show the jet energy response without applying GSC and the black up triangles, green down triangles, pink crosses and red circles show the jet energy response after applying the first (fTile0), second (fEM3), third (nTrk) and fourth (*TrkWidth*) GSC corrections, respectively. Each figure shows the corresponding p_T^{truth}-bin from figures 3.36 and 3.37.

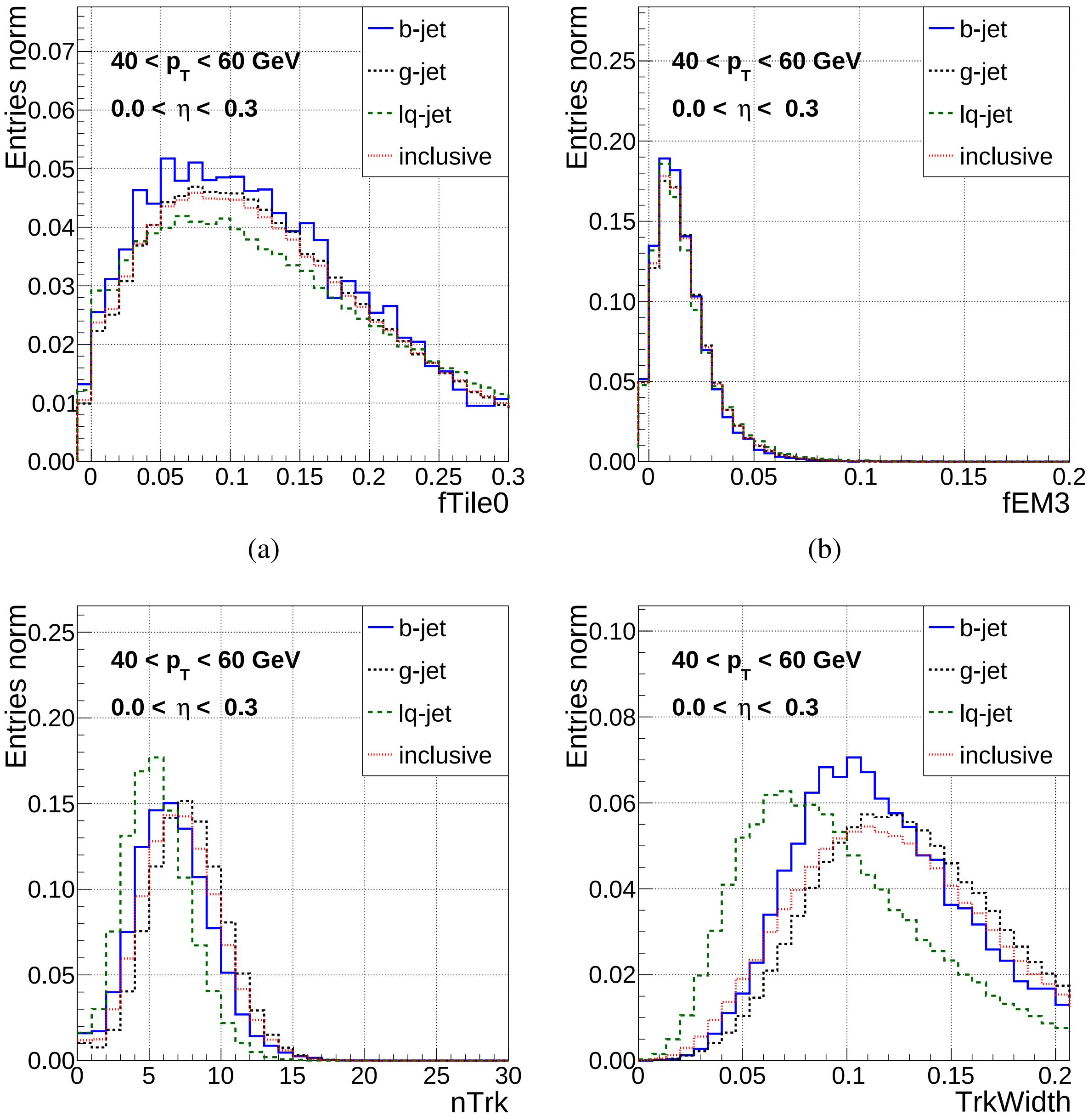
The identification of the jet flavour can only be done at truth level, using the generator information. The classification of jets at truth level uses the same criteria discussed in section 3.7.1. For the purpose of this thesis, c-jets are discarded and not used for further studies.

The distributions of f_{Tile0}, f_{em3}, $nTrk$ and $TrkWidth$ are shown in figures 3.39 and 3.40 for three types of jets (light-quark, gluon and b-jets) and for inclusive jet flavour, for the same η and p_T-bins as before. In order to make shape comparisons, all the distributions are normalised to unity. In the LHC, the dominant production process of di-jets is $gg \to gg$ and for this reason, the inclusive di-jet sample simulated with PYTHIA8 has a large content of gluon initiated jets, about 75 % [71]. Therefore, the black curve, representing the gluon initiated jet distributions is always closest to the red curve (inclusive distribution). The light-quark initiated jets distribution is shown by the green curves. These jets have, in particular, a smaller number of charged particles associated with the jet and the weighted p_T average of the position of the charged particles is found to be closer to the jet centre with respect to gluon initiated jets, as observed in the $nTrk$ and $TrkWidth$ distributions, respectively. In the distributions of f_{Tile0} and f_{em3}, the light-quark initiated jets distributions are more spread over than for gluon initiated jets but the average value is consistent with both types and properties. Additional p_T^{truth} and η-bins are shown in appendix B in figure B.3, and the shapes are similar.

The jet energy response is shown in figure 3.41 (top) as a function of p_T^{truth} for $0.0 < \eta < 0.3$ before (a) and after (b) applying GSC, separately for b-jets (black circles), for jets initiated by gluons (blue squares), and for jets initiated by light-quarks (red up triangles). The inclusive response is also shown as in figure 3.38, for comparison with the green down triangles. The default EM+JES jet energy response for quark initiated jets is always above the jet energy response for gluon initiated jets, with the light-quark initiated jet energy response being slightly higher than the response for b-jets. The jet energy response variation for the different jet types can be as large as 10 % for low-p_T jets, decreasing to about 4 % for jets with $p_T > 500$ GeV. There are also differences in the jet energy response after GSC for the different jet types but considerably smaller, within 2 % for $p_T > 80$ GeV and up to 5 % for very low p_T jets, close to the reconstruction threshold. GSC is, therefore, able to reduce the flavour dependence in the di-jet sample. The jet energy response for b-jets after GSC is applied is also reduced to values closer to unity.

B hadrons have about 20 % probability of decaying leptonically for a muon or an electron and the corresponding neutrino [4]. However, neither muons nor neutrinos are included in the reconstruction of the truth jet when the calorimeter energy response to jets is evaluated. Therefore, the energy of b-jets calibrated with GSC is always inferior to the energy of the original b-quark, and therefore further jet energy calibration steps will be required. One of these steps is implemented in the ZH analysis to correct for the energy of the muon in the jet (see section 4.8.2). This correction is not applied in the re-evaluation of GSC-specific systematic uncertainties for b-jets, which is discussed in the section 3.9.4.1.

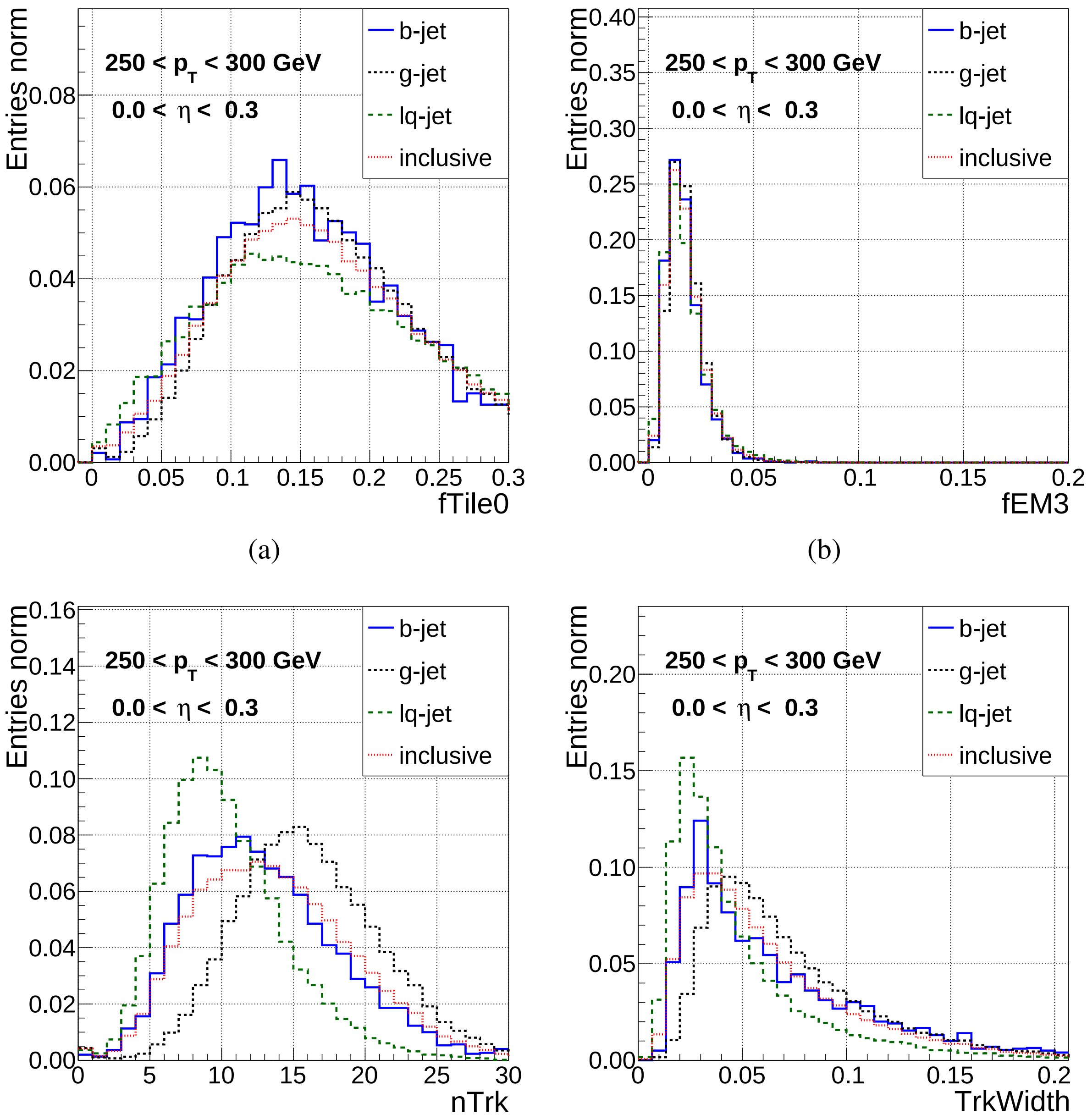
The bottom plots of figure 3.41 show the jet energy resolution as a function of p_T^{truth}, for each jet flavour before (c) and after (d) applying GSC. The jet energy resolution does not have a large dependence on the jet flavour. Therefore, the improvement of 20%, discussed in section 3.9.2, is similar in all jet flavours. Additional p_T^{truth}-bins are shown in appendix B in figure B.4, with similar behaviour.



(a) (b)

(c) (d)

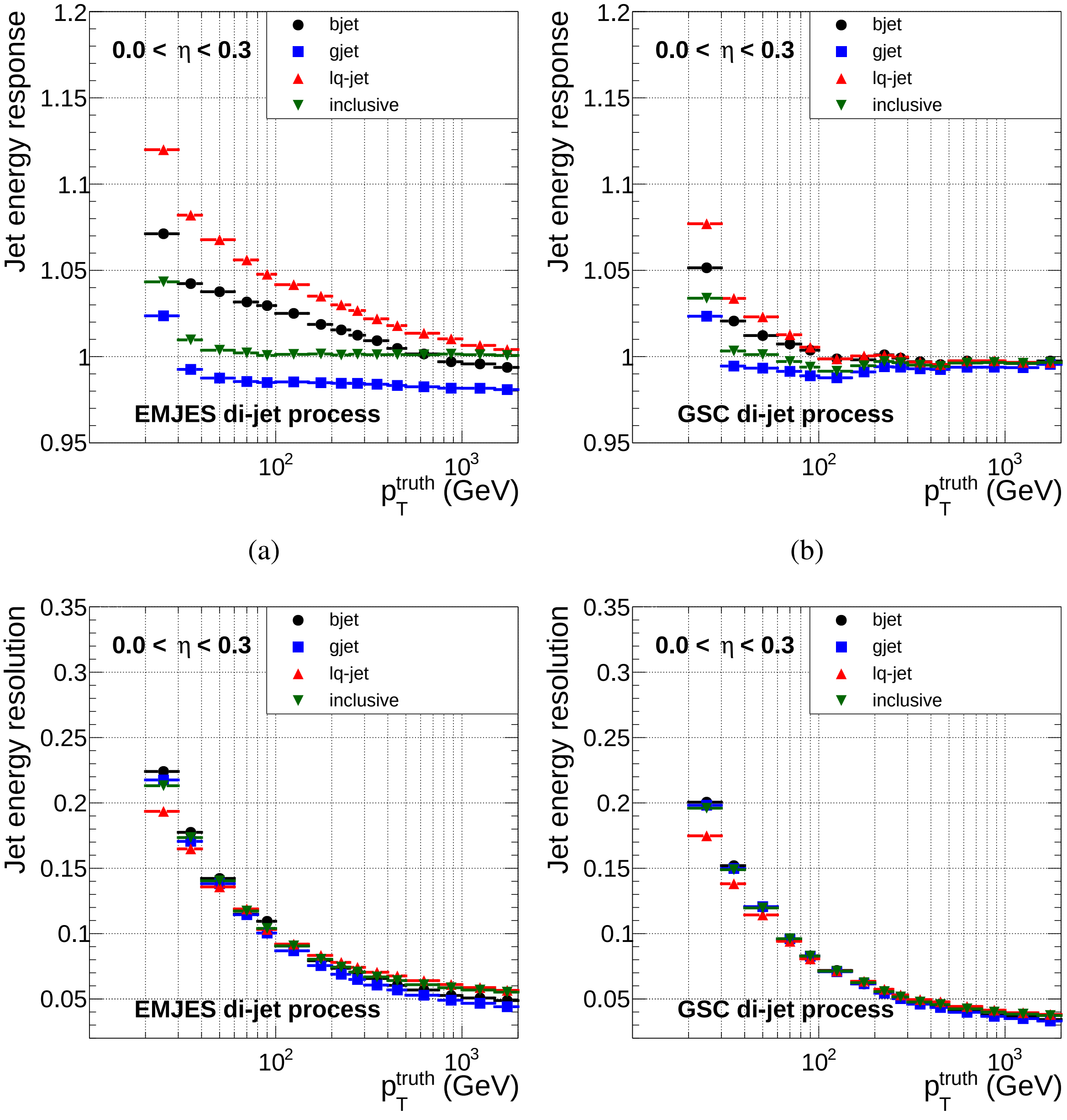
Figure 3.39.: Distributions of f_{Tile0} (a), f_{em3} (b), $nTrk$ (c) and $TrkWidth$ (d) for jets with $0.0 < \eta < 0.3$ and $40 < \mathrm{p}_{\mathrm{T}}^{truth} < 60$ GeV, normalised to unity, for different jet flavour compositions of di-jet events simulated with PYTHIA8. The b-jet flavour is shown in blue, gluon initiated jets in black, light-quark initiated jets in green and the inclusive jet flavour in red, with different line styles.



(a) (b)

(c) (d)

Figure 3.40.: Distributions of f_{Tile0} (a), f_{em3} (b), $nTrk$ (c) and $TrkWidth$ (d) for jets with $0.0 < \eta < 0.3$ and $250 < \mathrm{p}_{\mathrm{T}}^{truth} < 300$ GeV, normalised to unity, for different jet flavour compositions of di-jet events simulated with PYTHIA8. The b-jet flavour is shown in blue, gluon initiated jets in black, light-quark initiated jets in green and the inclusive jet flavour in red, with different line styles.



(a) (b)

(c) (d)

Figure 3.41.: Jet energy response (top) and resolution (bottom) as a function of the truth jet transverse momenta, obtained with inclusive di-jet events simulated with PYTHIA8, for the different jet flavour compositions: b-jets (black circles), gluon initiated jets (blue squares), light-quark initiated jets (red up triangles), and inclusive jets (green down triangles). The default calibration, EM+JES, is shown on the left and the improved calibration, GSC, is shown on the right.

3.9.4. Evaluation of GSC specific systematic uncertainties

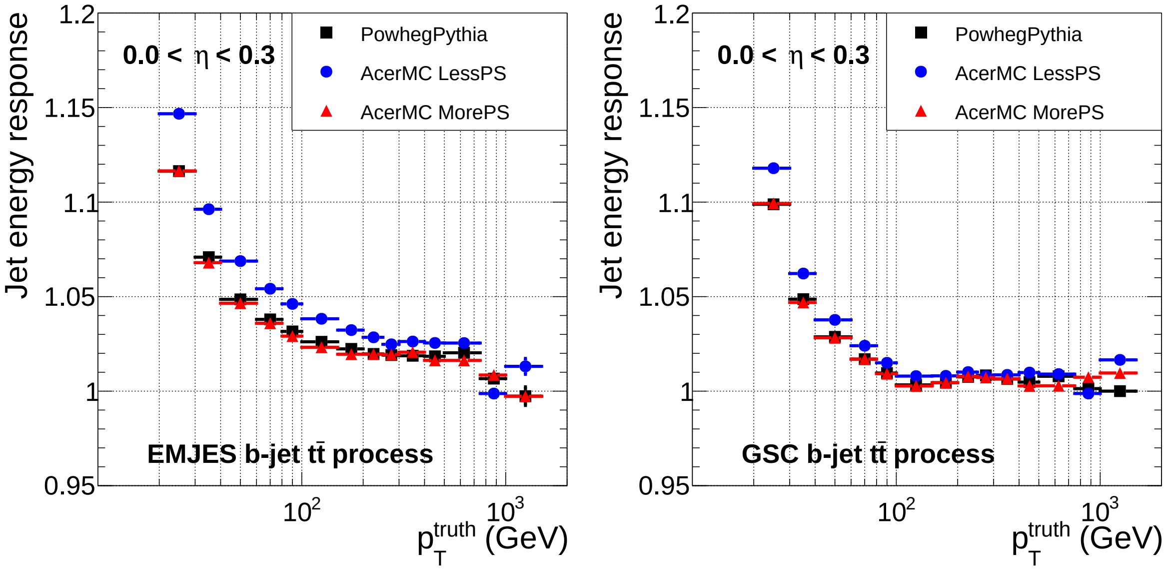
The jet systematic uncertainties were measured by comparing the jet response between data and simulation using *in situ* techniques [71], discussed in section 3.7. In addition to the EM+JES baseline uncertainty, other uncertainties have to be considered. They are estimated with simulation and validated with data. The uncertainties obtained for EM+JES were validated for GSC again by data and MC comparisons except for the MC specific systematic uncertainties: b-jet response and jet energy response differences of the fast versus full simulations. In the following sections, the evaluation of these uncertainties is presented.

Another MC based systematic uncertainty was discussed in section 3.7.1 and is related to the difference in the jet energy response for jets initiated by light-quarks and jets initiated by gluons. Section 3.9.3 shows that this difference is greatly reduced after GSC.

3.9.4.1. Uncertainty on the response to b-jets

The jet energy response systematic uncertainty, due to differences in the fragmentation model of the b-quarks, was determined by comparing, for different simulation models, the jet energy response for b-jets from the di-jet and $t\bar{t}$ processes. The b-quarks are hadronized with different simulations and their jet energy response is validated using jets which have been b-tagged and comparing data to simulations [133]. This systematic evaluation used truth b-jets, as discussed in section 3.7.2. However, in section 3.9.3, there were some differences observed between the GSC property distributions for inclusive jets and those from b-jets, specially for low-p_T^{truth} jets. For this reason, it becomes necessary to evaluate if the estimation of the b-jet flavour uncertainty is still valid for GSC.

In the $t\bar{t}$ process, two b-quarks are created in each event, which results in a sample enriched on b-jets. The chosen simulations for this process used the POWHEG generator interfaced with PYTHIA6 to simulate the hadronization. This is the same sample used for the simulation of the $t\bar{t}$ process in the ZH analysis and more details will be given in section 4.4. This simulation is compared to other $t\bar{t}$ samples, generated with ACERMC interfaced with PYTHIA6 for the hadronization but tuned with “more” and “less” parton shower by changing the parameters of the PYTHIA6 settings for initial and final state radiation [134]. The energy response for b-jets is evaluated for each sample as a function of p_T^{truth} in figure 3.42 for the most central η-bin. The $t\bar{t}$ process simulated with POWHEG+PYTHIA6 is shown with the black square, and the sample simulated with the ACERMC+PYTHIA6 with more (less) parton shower is shown with the blue circles (red triangles). The b-jet energy response,

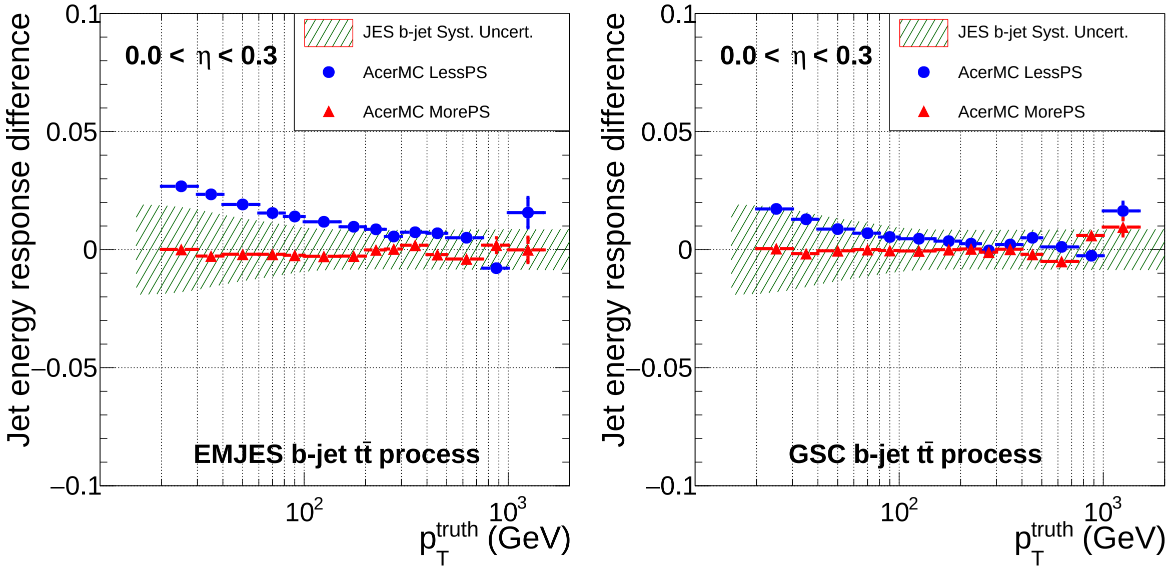


(a) (b)

Figure 3.42.: b-jet energy response as a function of $\mathrm{p}_{\mathrm{T}}^{truth}$, for three $t\bar{t}$ samples simulated with POWHEG+PYTHIA6 (black squares), ACERMC+PYTHIA6 with less (blue circles) and more (red triangles) parton shower, for EM+JES (a) and GSC (b) for jets with $0.0 < \eta < 0.3$.

obtained with jets calibrated before (after) the application of the GSC calibration, is shown on the left (right). Some additional η-bins are shown in appendix B in figure B.5. After GSC is applied, the jet energy response for b-jets decreases, similarly to what happened already in the di-jet sample. To make the evaluation easier, the difference between the energy responses for b-jets for the two ACERMC samples with respect to the one obtained with the POWHEG sample are shown in figure 3.43, for EM+JES (a) and GSC (b). The green band represents the b-jet systematic uncertainty evaluated for EM+JES [135]. After applying the GSC calibration, the difference in the energy response for b-jets relative to POWHEG+PYTHIA6, produced with different MC generators and parton shower tunes, becomes smaller than the uncertainty band. This means that the uncertainty derived for EM+JES can also be used for GSC.

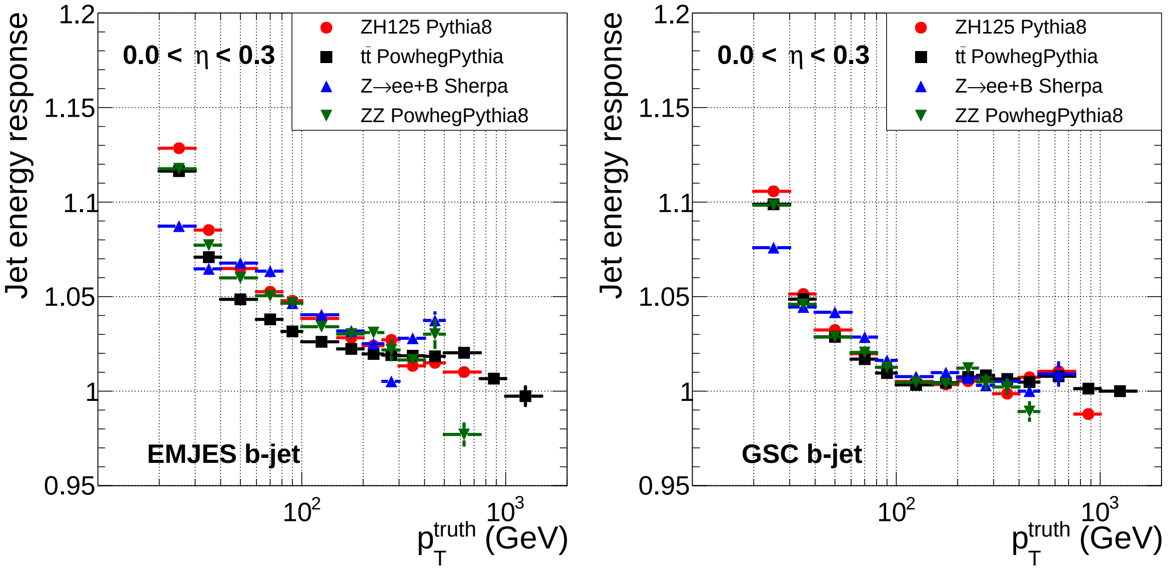
It is also necessary to evaluate possible b-jet energy scale uncertainties when considering b-quarks originating in the decay of a colour singlet resonance. The energy response for b-jets is compared using four different processes: ZH, Z+jets, $t\bar{t}$ and ZZ. These four processes are the most relevant processes in the ZH analysis: the signal and the three



(a) (b)

Figure 3.43.: b-jet energy response difference between the two $t\bar{t}$ samples simulated with ACERMC+PYTHIA6 with less (blue circles) and more (red triangles) parton shower and the POWHEG+PYTHIA6 sample, for EM+JES (a) and GSC (b) for jets with $0.0 < \eta < 0.3$. The green band represents the b-jet systematic uncertainty obtained for the EM+JES calibration.

leading backgrounds, respectively, as it will be discussed in chapter 4. The objective of this evaluation is to test different b-jet production mechanisms. In the ZH and ZZ processes, a b-jet pair results from the decay of a colourless resonance, the Higgs and the Z boson, respectively. In the $t\bar{t}$ process, a b-quark is created from the decay of each of the top quarks. Finally, in the associated production of b-jets with a Z boson, the b-quarks result mainly from gluon-splitting, similar to the di-jet production used in the GSC derivation. For this evaluation, the energy response for b-jets using the ZH signal and three background processes is shown in figure 3.44 for the central pseudorapidity bin, before (a) and after (b) applying the GSC calibration. The ZH signal process is simulated with PYTHIA8 and shown with the red circle, the $t\bar{t}$ process is shown again with the black squares, the Z+b-jets simulated with the SHERPA generator is shown with the blue up triangles, and the diboson ZZ production is simulated with POWHEG+PYTHIA8 and shown by the green down triangles. More details on these samples are given in section 4.4. The differences between the jet energy response

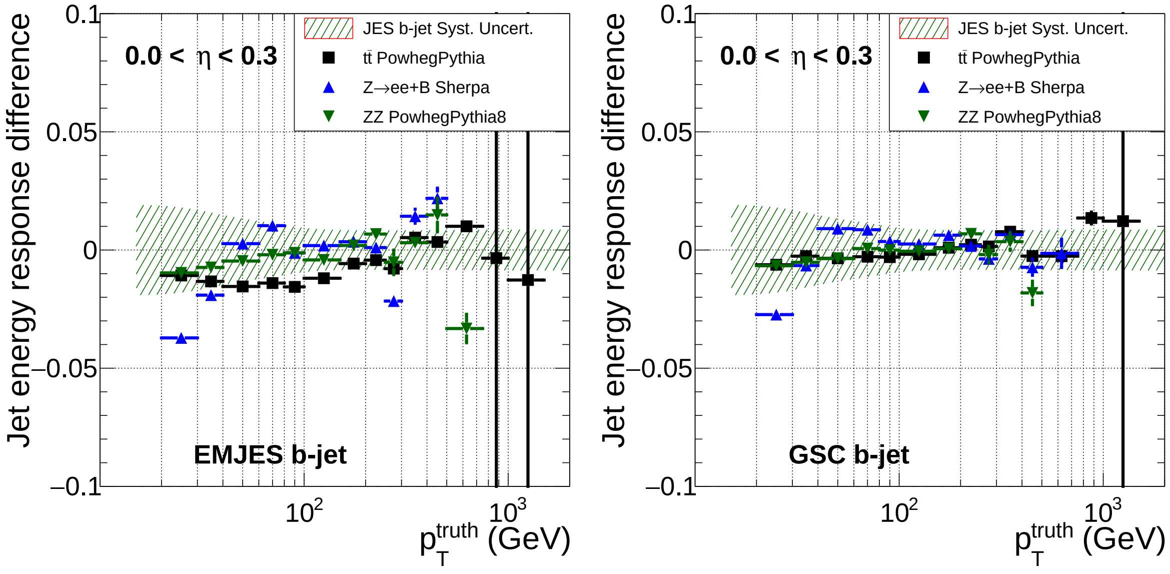


(a) (b)

Figure 3.44.: b-jet energy response with EM+JES (a) and GSC (b), for jets with $0.0 < \eta < 0.3$, for the signal process for the ZH analysis (red circles) and for the three leading backgrounds: $t\bar{t}$ (black squares), Z+b-jets (blue up triangles), and ZZ (green down triangles).

obtained for different samples in each p_T^{truth} bin is smaller after applying GSC. The b-jet energy response difference of each of the background processes is calculated relative to the signal process, ZH, and compared again with the b-jet uncertainty obtained for the EM+JES calibration. The relative difference with ZH is shown in figure 3.45 before (a) and after (b) applying GSC. Before applying GSC, the only b-jet energy response within the b-jet uncertainty band for the ZH process is the diboson process ZZ as in both the b-quarks result from a colourless resonance decay. After the GSC calibration is applied, the discrepancies in the energy responses for b-jets, originating from different processes are much smaller and nearly all the differences are located inside the b-jet systematic uncertainty band obtained for EM+JES. Figure B.6 in the appendix B shows other η-bins with similar conclusions.

Based on these two evaluations, no additional systematic uncertainty was added.



(a) (b)

Figure 3.45.: b-jet energy response difference between the signal process for the ZH analysis and each of the three leading backgrounds: $t\bar{t}$ (black squares), Z+b-jets (blue up triangles), and ZZ (green down triangles), with EM+JES (a) and GSC (b), for jets with $0.0 < \eta < 0.3$. The green band represents the b-jet systematic uncertainty obtained for the EM+JES calibration.

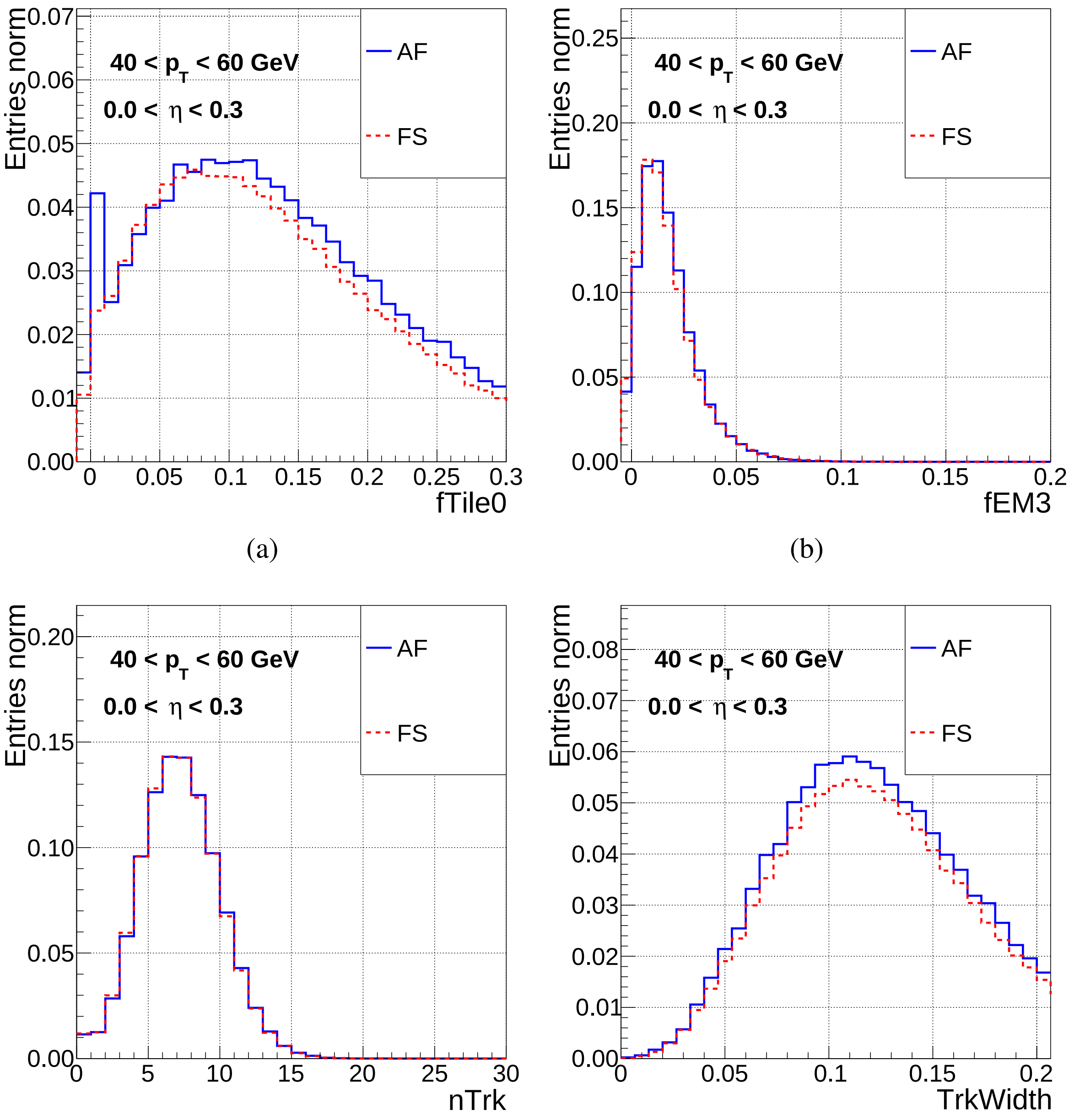
3.9.4.2. Difference in response between full and fast simulation

Section 3.7.3 referred already the importance of the parametrised version of the calorimeter energy response in the production of nearly all the samples used for the ZH search. A systematic uncertainty accounting for the difference in the response between full simulation (FS) of the detector with GEANT4 and the parametrised version ATLASFASTII (AF) has to be evaluated and considered in any analysis using jets. This uncertainty was evaluated for EM+JES with the inclusive di-jet sample. It is, therefore, necessary to understand if this systematic uncertainty is also enough to cover the differences between the full and the fast simulation, when using GSC.

Figures 3.46 and 3.47 show the distributions of the four jet properties used in the GSC calibration, for jets with $0.0 < \eta < 0.3$, using the same $\mathrm{p}_{\mathrm{T}}^{truth}$-bins as before. The lines represent the inclusive distributions for the di-jet sample obtained with the full detector simulation (blue full line) and for the ATLASFASTII parametrised simulation of the calorimeter (red dashed line). Small discrepancies are found in every variable distribution and in every $\mathrm{p}_{\mathrm{T}}^{truth}$-bins. Additional η and $\mathrm{p}_{\mathrm{T}}^{truth}$-bins are found in the appendix B in figure B.7.

The jet energy response as a function of the $\mathrm{p}_{\mathrm{T}}^{truth}$ is shown in figure 3.48 (a) for jets with $0.0 < \eta < 0.3$, for the di-jet inclusive flavour samples before (after) applying GSC shown by the blue squares (red circles), obtained with the full simulation (open markers) and fast simulation (closed markers). The jet energy response with the ATLASFASTII is observed to be systematically higher when compared with the full simulation of the detector for both EM+JES and GSC. Figure 3.48 (b) shows the relative difference in the jet energy response between the full simulation (FS) and ATLASFASTII (AF), FS-AF. The AtlasFastII non-closure systematic uncertainty obtained for EM+JES is shown by the blue shaded band. The difference for EM+JES between the jet energy response with FS and with AF lies inside the expected systematic uncertainty, except for a couple of bins at low $\mathrm{p}_{\mathrm{T}}^{truth}$. Conversely, after GSC is applied, the jet energy response difference is around 0.5 % larger than the EM+JES based systematic uncertainty at high-$\mathrm{p}_{\mathrm{T}}^{truth}$, starting around 200 GeV. For this reason, an additional 0.5 % term was added to this systematic uncertainty.

This systematic uncertainty was then validated using the POWHEG+PYTHIA6 generated $t\bar{t}$ sample. Despite having a jet flavour composition completely different from the di-jet sample, since the GSC calibration objective is the removal of flavour dependence, the $t\bar{t}$ sample is a good candidate to test this new systematic uncertainty. Additionally, since each



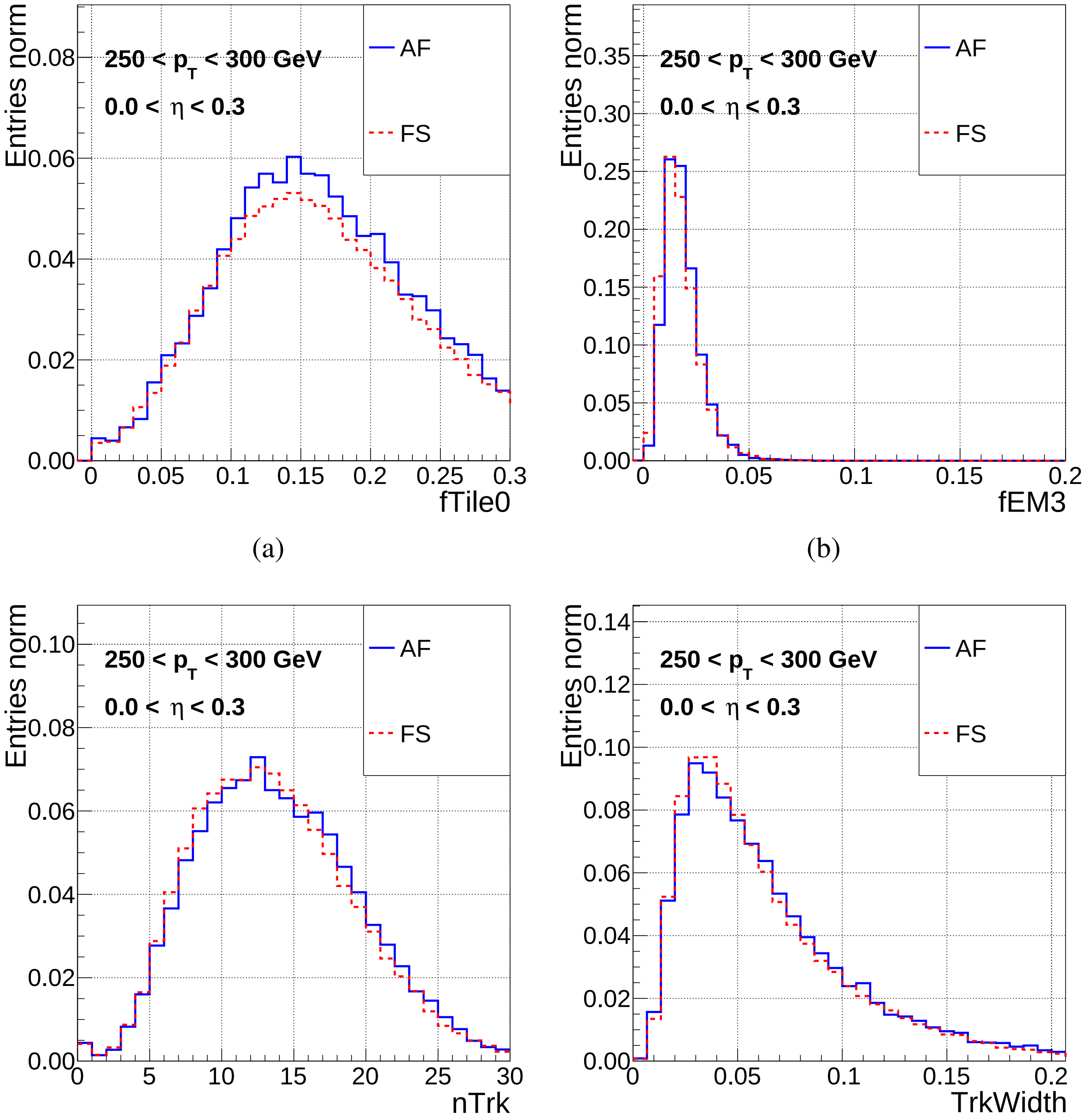
(a)

(b)

(c)

(d)

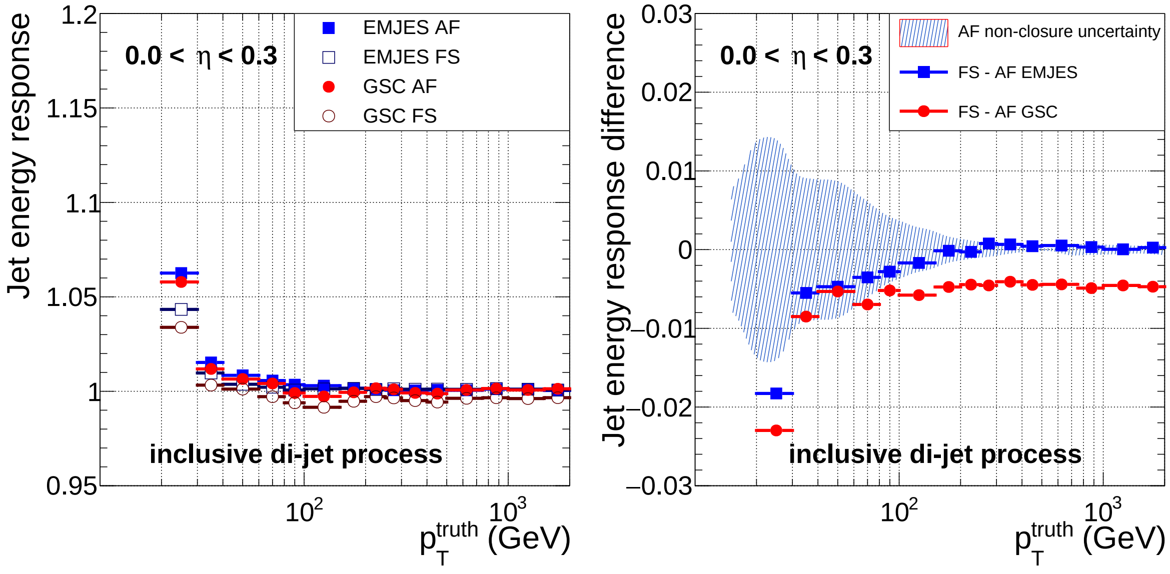
Figure 3.46.: Distributions of the jet properties, fTile0 (a), fEM3 (b), nTrk (c) and *TrkWidth* (d), for inclusive di-jet events simulated with PYTHIA8, for jets with $0.0 < \eta < 0.3$ and $40 < p_T < 60$ GeV, normalised to unity and obtained for different detector simulations: full simulation (FS) and ATLASFASTII (AF) simulation shown by the blue filled line and red dash line, respectively.



(a) (b)

(c) (d)

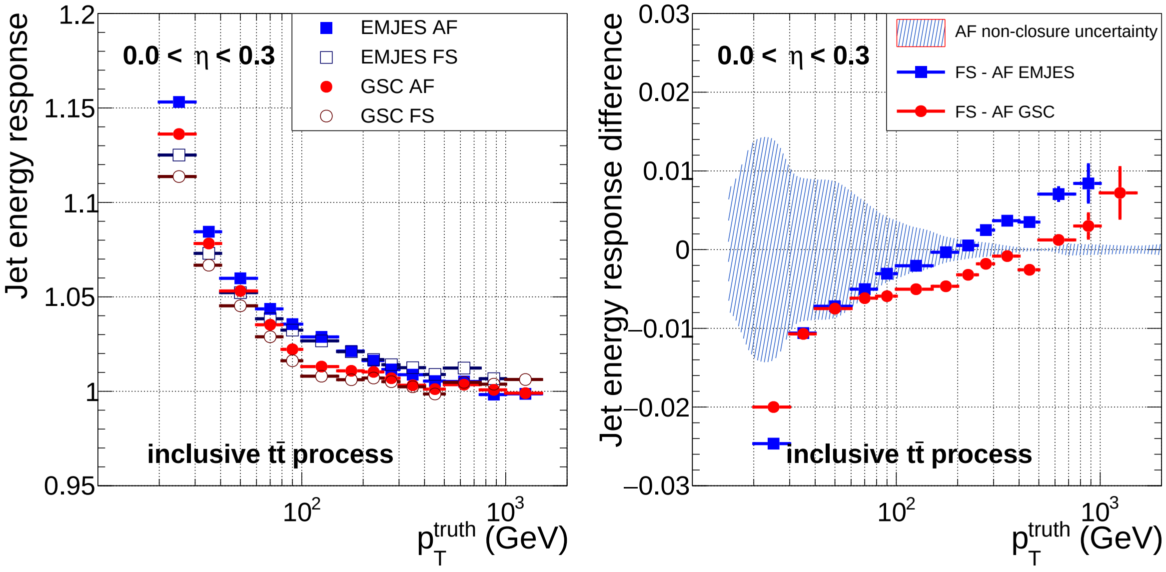
Figure 3.47.: Distributions of the jet properties, fTile0 (a), fEM3 (b), nTrk (c) and *TrkWidth* (d), for inclusive di-jet events simulated with PYTHIA8, for jets with $0.0 < \eta < 0.3$ and $250 < p_{\mathrm{T}} < 300$ GeV, normalised to unity and obtained for different detector simulations: full simulation (FS) and ATLASFASTII (AF) simulation shown by the blue filled line and red dash line, respectively.



(a) (b)

Figure 3.48.: Jet energy response (a) as a function of the p_T^{truth}, for the parametrised and the full simulation for jets with $0.0 < \eta < 0.3$, using inclusive di-jet events simulated with PYTHIA8. Comparison between the ATLASFASTII (closed markers) and the full simulation (open markers) of the detector. Jet energy response difference (b) between the full and fast simulation, FS-AF. The jet energy response for the EM+JES (GSC) calibration is shown by blue squares (red circles). The blue band in (b) is the measured systematic uncertainty obtained for the ATLASFASTII non-closure with the EM+JES calibration.

event in the $t\bar{t}$ samples should contain at least two b-quarks from the top quark decay, it can validate the additional fast simulation systematic uncertainty for a sample enriched with b-jets. Figure 3.49 (a) shows the jet energy response as a function of p_T^{truth} in the central η-bin. The difference in the jet energy response between ATLASFASTII and full simulations is shown in figure 3.49 (b). It shows that the difference is strongly dependent on p_T^{truth} when using EM+JES. The same systematic uncertainty band for EM+JES is shown for comparison. After GSC is applied, the difference falls outside this uncertainty band but the extra 0.5 % determined previously for the di-jet sample still accounts for the observed discrepancies, validating this new uncertainty. Appendix B shows other η-bins in figure B.8.



(a) (b)

Figure 3.49.: Jet energy response (a) and response difference between the parametrised and the full simulation (b) as a function of the p_T^{truth} for jets with $0.0 < \eta < 0.3$, using the flavour inclusive $t\bar{t}$ sample. The figures show the comparison between the ATLASFASTII (closed markers) and the full simulation (open markers) of the detector. The jet energy response for the EM+JES (GSC) calibration is shown by blue squares (red circles). The blue band is the measured systematic uncertainty obtained for the ATLASFASTII non-closure with the EM+JES calibration.

3.9.5. Invariant mass studies

It has already been shown that GSC improves the jet energy resolution (section 3.9.2). In this section, the effect of GSC on the resolution of the invariant mass of a b-jet pair (and therefore on the $H \to b\bar{b}$ searches) will be evaluated. The estimation of the improvement in the invariant mass resolution is studied for a narrow resonance, decaying to b-quarks, when the jets are reconstructed with the anti-k_T algorithm and calibrated with EM+JES and with the EM+JES+GSC. Since this work has a special focus on the Higgs boson search, the same jet selection used in the ZH analysis is applied in this study.

Sections 4.5 will discuss the ZH selection in detail, whereas here only the selection of the reconstructed jet and the event are referred. This work uses only the ZH process simulated with the PYTHIA8 generator.

Jets are required to be in the calorimeter acceptance region, $|\eta| < 4.5$, and to have a transverse momentum above 25 GeV, when they are within the inner detector acceptance ($|\eta| < 2.5$) or above 30 GeV for $|\eta| > 2.5$. The JVF criterium, discussed in section 2.7.5, is applied to reject pileup jets. The event is required to have exactly two reconstructed b-tagged jets with MV1 value corresponding to the 80 % operating point of b-tagging efficiency. Events with more than two jets are vetoed. Both reconstructed and truth jets are required to be isolated, as described in section 3.4.1 and the reconstructed jet is required to be matched to the closest truth b-jet within the distance $\Delta R < 0.3$. These two selections ensure that the two reconstructed jets which were b-tagged are the same produced in the hadronization of the two b-quarks from the Higgs boson decay.

3.9.5.1. Invariant mass with reconstructed jets

The distributions of the invariant mass of the di-b-jet system are not symmetrical due to initial state radiation (ISR) and final state radiation (FSR) effects discussed in section 2.8. In ISR, a gluon is emitted in the same direction as one of the b-jets, increases the energy of that b-jet and consequently the invariant mass of the b-jet system. Such events will populate the high invariant mass region. In FSR, one of the b-quarks radiates one gluon which is reconstructed as a different jet and the resulting invariant mass of the b-jets will have a smaller value, populating the low invariant mass region. For this reason, a Bukin function [136] is used to fit the invariant mass distributions. This function is composed of a Gaussian core with two asymmetric exponential shaped tails. The fit is performed centred on the maximum of the distribution with a window of four times the RMS of the $m_{b\bar{b}}$ distribution. The width of the Gaussian core of the Bukin fit is used to evaluate the resolution in the invariant mass distribution before and after applying GSC.

The invariant mass distributions are obtained for five bins of the Higgs boson transverse momentum, using generator level information, with boundaries of 90 GeV, 120 GeV, 160 GeV and 200 GeV (see figure 3.50). These will be the same boundaries of the regions defined in section 4.5 for the transverse momentum of the Z boson. Figure 3.50 shows the distributions of the invariant mass of the two reconstructed b-jets, with the b-jets calibrated at EM+JES (GSC) scale with the blue (red) markers. The lines represent the Bukin fit applied to the distribution with the respective colour. The figures present also the parameter retrieved from the Bukin fit for the average and for the width of the fitted distribution.

The ratio between the widths, σ, from the distributions for EM+JES and GSC evaluates the improvement in the $m_{b\bar{b}}$ resolution due to the GSC calibration from equation 3.14.

$$\text{Improvement} = \left(\frac{\sigma^{\text{EM+JES}}}{\sigma^{\text{GSC}}} - 1 \right) \times 100\% \tag{3.14}$$

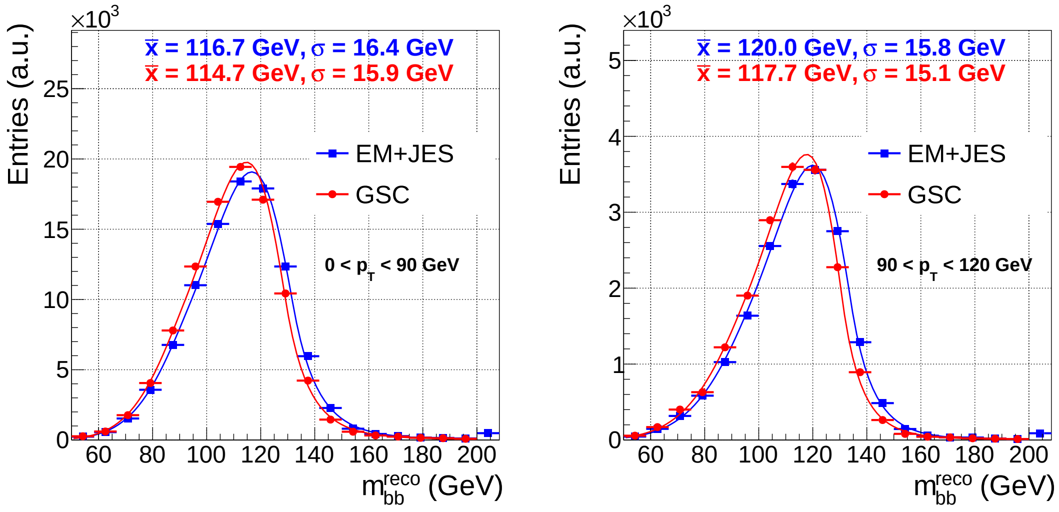
The mean value of the Gaussian core of the Bukin fit that represents the peak central position and the width of the fit shown are summarised in table 3.1. These central position values provide an estimate of the measured Higgs boson mass.

As shown in section 3.9.3, the b-jet response with GSC becomes closer to unity. However, the truth jets, used in the evaluation of the GSC calibration, do not include muons or neutrinos, since only the particles that deposit their energy in the calorimeter are included. After applying the GSC calibration, the mean value of the $m_{b\bar{b}}$ distribution decreases by around 2.3 GeV, moving away from the truth Higgs boson simulated mass, 125 GeV. This effect is due to the energy of muons and neutrinos, which were not included in the jets and no correction was yet applied for that.

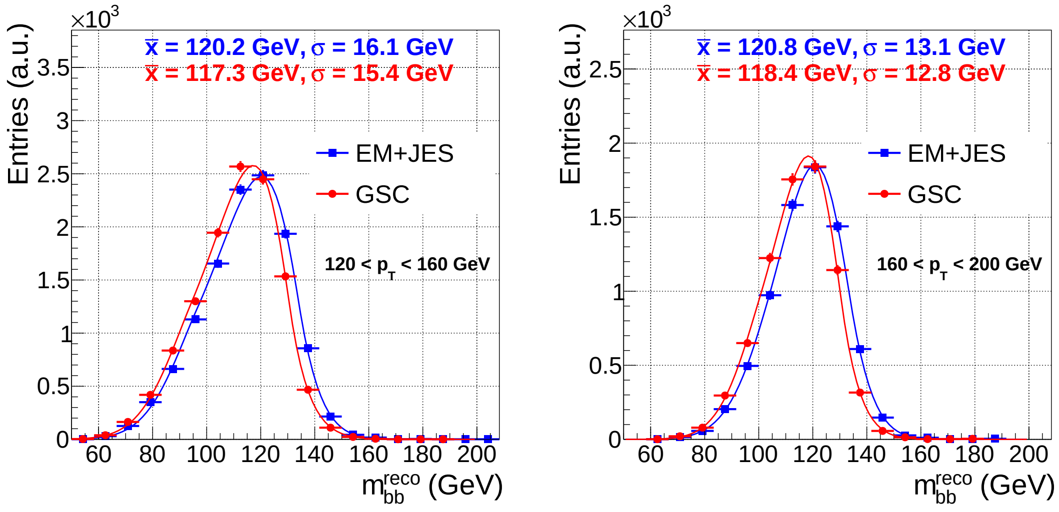
Nevertheless, both ISR and FSR affect the calculation of the di-b-jet invariant mass resolution so, in the next section, another method is used to separate these effects from the improvement.

$\mathrm{p_T}$-bin	EM+JES		GSC		Improvement
	mean	width	mean	width	(%)
$p_T^H < 90$ GeV	116.7	16.4	114.7	15.9	3.2 ± 0.0
$90 < p_T^H < 120$ GeV	120.0	15.8	117.7	15.1	4.3 ± 0.1
$120 < p_T^H < 160$ GeV	120.2	16.1	117.3	15.4	4.7 ± 0.1
$160 < p_T^H < 200$ GeV	120.8	13.1	118.4	12.8	2.7 ± 0.1
$p_T^H > 200$ GeV	125.4	10.8	121.5	9.9	9.4 ± 0.3
Combined	118.6	16.0	116.3	15.4	3.9 ± 0.0

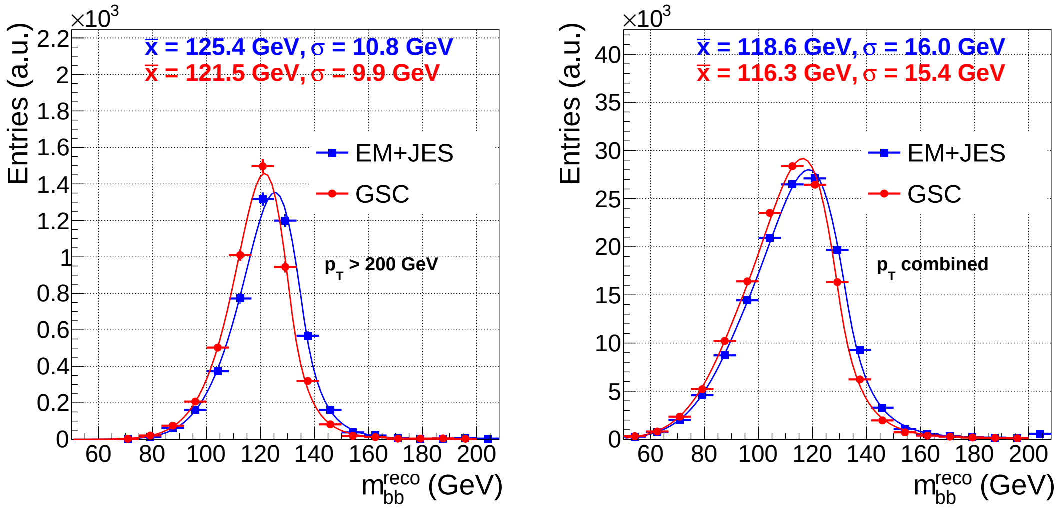
Table 3.1.: Mean and width of a fit to a Bukin function of the $m_{b\bar{b}}$ distributions when the b-jets are calibrated with EM+JES and GSC. The last column shows the improvement observed in the invariant mass resolution due to the GSC calibration.



(a) $p_{\mathrm{T}}^{\mathrm{H}} < 90$ GeV (b) $90 < p_{\mathrm{T}}^{\mathrm{H}} < 120$ GeV



(c) $120 < p_{\mathrm{T}}^{\mathrm{H}} < 160$ GeV (d) $160 < p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV



(e) $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV (f) Combined

Figure 3.50.: Invariant mass of the two reconstructed b-tagged jets in bins of the truth Higgs boson transverse momentum for jets calibrated with EM+JES (blue squares) and GSC (red circles). The distributions are fitted with a Bukin function. The mean and width obtained from the fit are shown.

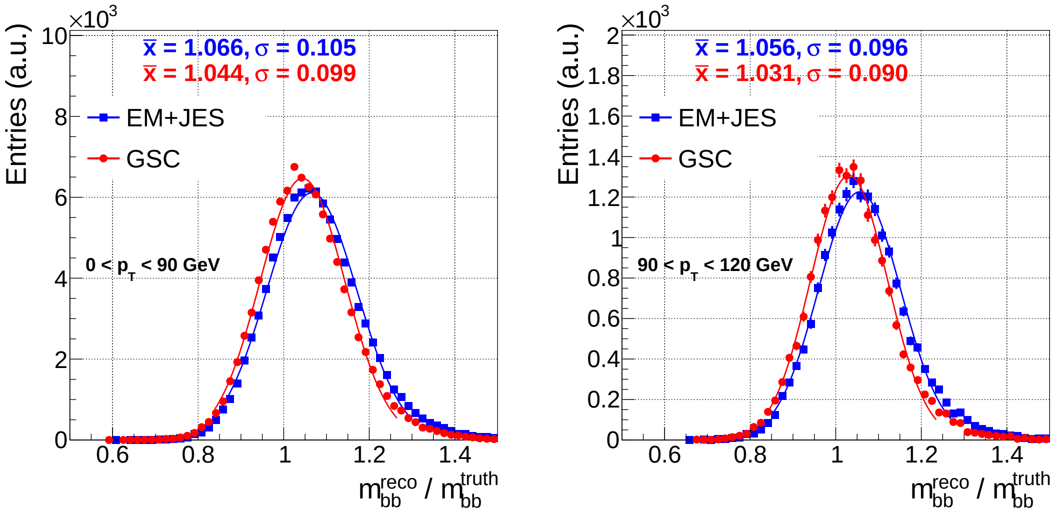
3.9.5.2. Invariant mass ratio

To disentangle the improvement of the invariant mass resolution from the initial and final state radiation effects, an invariant mass ratio between the reconstructed and truth $b\bar{b}$-invariant masses is defined: $\mathrm{m}_{b\bar{b}}^{\mathrm{reco}}$ / $\mathrm{m}_{b\bar{b}}^{\mathrm{truth}}$. The truth mass of the Higgs boson is calculated using the truth b-jets, *i.e.* jets that include a b-quark and are reconstructed at particle level using the anti-k_{T} algorithm with $R = 0.4$ including all particles except muons and neutrinos. Figure 3.51 shows the distributions of the $\mathrm{m}_{b\bar{b}}^{\mathrm{reco}}$ / $\mathrm{m}_{b\bar{b}}^{\mathrm{truth}}$ ratio with the same colours used before. Now a Gaussian fit is adopted instead of the Bukin function since one expects symmetric tails. The fit is done in an interval obtained of $\pm 2\times$ RMS from the centre of the distribution. Table 3.2 summarises the parameters, mean and width obtained from the Gaussian fit.

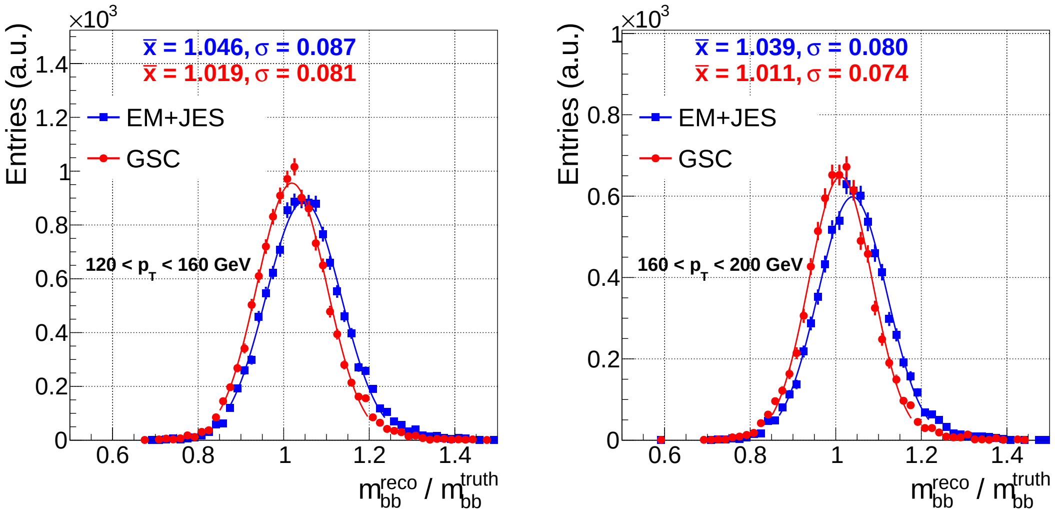
The mean value of the Gaussian distribution becomes closer to 1 after the GSC calibration, which is a consequence of the better jet energy scale obtained with GSC, as shown in section 3.9.3. The width of the $\mathrm{m}_{b\bar{b}}^{\mathrm{reco}}$ / $\mathrm{m}_{b\bar{b}}^{\mathrm{truth}}$ improves by around 6.2 % for the lowest $\mathrm{p_T}$-bin of the Higgs boson. The improvement increases with the Higgs boson $\mathrm{p_T}$ up to 11.5 % in the most sensitive range $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV, with an average improvement of 6.1 % when all the Higgs boson $\mathrm{p_T}$-bins are considered together.

$\mathrm{p_T}$-bin	EM+JES		GSC		Improvement
	mean	width	mean	width	(%)
$p_{\mathrm{T}}^{\mathrm{H}} < 90$ GeV	1.066	0.105	1.044	0.099	6.2 ± 0.0
$90 < p_{\mathrm{T}}^{\mathrm{H}} < 120$ GeV	1.056	0.096	1.031	0.090	6.9 ± 0.1
$120 < p_{\mathrm{T}}^{\mathrm{H}} < 160$ GeV	1.046	0.087	1.019	0.081	7.8 ± 0.1
$160 < p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV	1.039	0.080	1.011	0.074	8.4 ± 0.1
$p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV	1.035	0.083	1.008	0.074	11.5 ± 0.3
Combined	1.060	0.100	1.036	0.095	6.1 ± 0.0

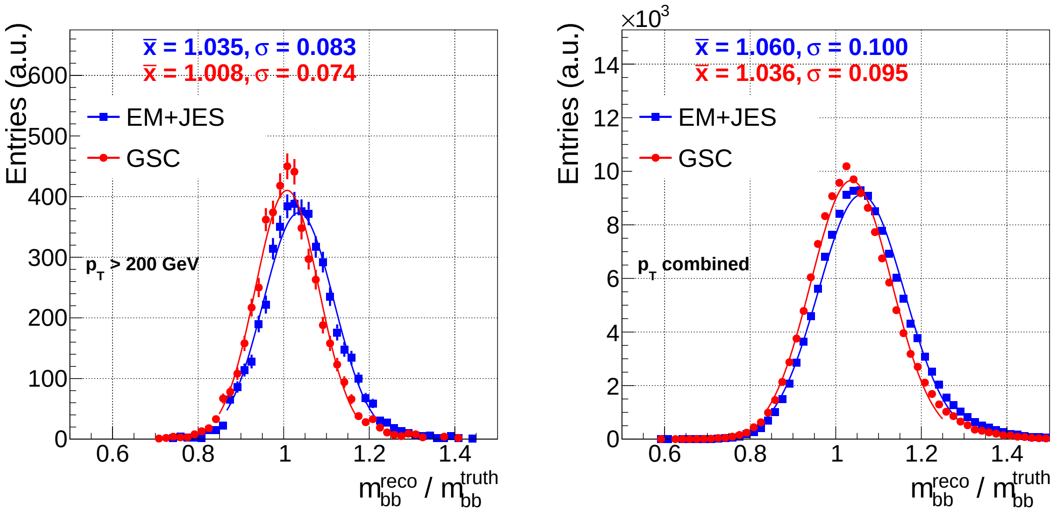
Table 3.2.: Mean and width of a Gaussian fit to the $\mathrm{m}_{b\bar{b}}^{\mathrm{reco}}$ / $\mathrm{m}_{b\bar{b}}^{\mathrm{truth}}$ ratio distribution obtained when jets are calibrated with EM+JES and GSC. The improvement in the resolution (width) with GSC with respect to EM+JES is shown in the last column.



(a) $p_{\mathrm{T}}^{\mathrm{H}} < 90$ GeV (b) $90 < p_{\mathrm{T}}^{\mathrm{H}} < 120$ GeV



(c) $120 < p_{\mathrm{T}}^{\mathrm{H}} < 160$ GeV (d) $160 < p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV



(e) $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV (f) Combined

Figure 3.51.: Invariant mass ratio between the reconstructed and the Higgs boson invariant mass calculated using particle level truth b-jets for jets calibrated with EM+JES (blue squares) and GSC (red circles). The mean and width of the Gaussian fit is also shown.

3.9.6. Summary of the GSC conclusions

The global sequential calibration (GSC) was developed to remove the dependence of the jet energy response on some jet properties. The energy fraction deposited by the jet in the first layer of the Tile calorimeter and in the last layer of the electromagnetic calorimeter, are sensitive to the energy losses in the cryostat between the calorimeters and to differences in the jet energy response between the electromagnetic and the hadronic cascade developments in the hadronic calorimeter. Two other jet properties use the number of charged particles and their transverse distribution in the jet to distinguish between light-quark and gluon initiated jets and to reduce the jet energy response dependence on the jet flavour.

GSC was derived as a sequence of corrections for each of the properties. Using the same simulation as for the jet energy scale derivation, the flavour inclusive di-jet sample, the jet energy response as a function of each jet property was studied, before and after each of the corrections and was found to change only when the correction was applied. This study concluded that the jet energy response in each η-p_T interval did not differ more than 1 % with respect to the values obtained with EM+JES, as intended. GSC improved the jet energy resolution by 20 % relatively to the result obtained with the EM+JES default jet calibration.

Next, the performance of GSC was studied for different jet flavours. The jet energy response had a strong dependence on the jet flavour after the EM+JES calibration for the di-jet sample used to derive it. This sample is dominated by gluon initiated jets, which were found to have a jet energy response below 1 by about 2 %. About 25 % of the jets in the sample were initiated by light quarks, which had a jet energy response larger than 1 by more than 8 % at low-p_T decreasing to a 1-2 % for $p_T > 400$ GeV. The b-jets that exist in this sample result from gluon-splitting and their jet energy response was above 1 by up to 4 %. In all cases, the jet energy response with EM+JES had larger absolute differences to 1 than GSC. This study concluded that, after GSC was applied, the jet energy response becomes closer to 1 for all jet flavours. The maximum difference observed from 1 after GSC was 4 % for light-quark initiated jets for jets with $p_T < 30$ GeV.

After this, it was necessary to evaluate the systematic uncertainties. These were obtained using the same *in situ* techniques to derive the uncertainties at the EM+JES scale but applied to GSC. The EM+JES uncertainties were found valid with GSC. The only systematic uncertainties that required further corrections were simulation related, which accounted the differences in the jet energy response for heavy flavour jets and different detector simulation conditions.

The systematic uncertainty on the jet energy response for heavy flavoured jets results from differences in the hadronization models of various simulations. Two evaluations were carried out to assess this systematic uncertainty. The hadronization model was tested using different simulations of the $t\bar{t}$ process, which feature three different hadronization parametrisations of PYTHIA6. The jet energy response from the default parametrisation was compared with hadronization parametrisations simulating less and more parton shower in the final state. The differences in the jet energy response with respect to the default parametrisation were compared against the b-jet flavour systematic uncertainty derived for EM+JES. This evaluation showed that, after GSC is applied, the heavy flavour uncertainty from EM+JES was enough to cover the differences in the jet energy response from the hadronization differences. The second evaluation had the objective of studying the energy response for b-jets for the most relevant processes of the ZH analysis: the ZH process and the three most important background processes, Z+jets, ZZ and $t\bar{t}$. This evaluation allowed the comparison of different processes that produce b-jets: colourless resonance decay in the ZH and ZZ processes, the decay of top quark in the $t\bar{t}$ process and gluon-splitting (like in the di-jet main process) in the Z+jets process. In this case, the difference in the jet energy response between each of the background processes relative to the ZH sample was compared to the same b-jet flavour systematic uncertainty obtained for EM+JES and validated with GSC in the previous evaluation. After GSC was applied, the discrepancy on the jet energy response between the four processes was reduced, such that the differences in the jet energy response were contained in the systematic uncertainty band obtained for EM+JES. These two evaluations concluded that the EM+JES heavy flavour systematic uncertainty can be used for GSC. Moreover, this study concluded that GSC was able to reduce the differences in the jet energy response for b-jets produced from different processes.

Concerning the different detector simulations, the di-jet sample is obtained using a full simulation of the detector and a fast simulation, which uses a parametrisation of the energy response of the calorimeter. Many analyses use the fast simulation that increases the statistics. One of them is the ZH analysis, which simulated almost every process with this fast simulation version of the detector. The difference in the jet energy response was evaluated by comparing the full simulation (FS) and the ATLASFASTII (AF) fast simulation of the detector for the EM+JES and GSC calibrations using the di-jet sample. With the default calibration, the difference in the jet energy response followed closely the uncertainty band obtained by the ATLAS jet performance group for EM+JES, as expected. However, after applying GSC, an additional contribution to this uncertainty of 0.5 % was necessary. As GSC reduces the flavour dependence, a $t\bar{t}$ sample was also used to validate this effect. The 0.5 % additional uncertainty was also enough to accommodate the FS-AF response differences in the $t\bar{t}$ sample after GSC is applied.

Finally, the ZH analysis is particularly interested in the GSC calibration to improve the invariant mass $m_{b\bar{b}}$ resolution and hence enhance the sensitivity of the ZH signal. Two tests were performed to assess the improvements in the invariant mass resolution. In the first one, a Bukin function was used to fit the asymmetric $m_{b\bar{b}}$ distributions of the Higgs boson. The width of this distribution improved by up to 9.4 % for the most sensitive $\mathrm{p_T}$-bin of the Higgs boson, $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV. Additionally, to suppress the effects of FSR and ISR, the ratio between the reconstructed Higgs boson mass and truth Higgs boson mass, defined as the invariant mass reconstructed with truth b-jets from the Higgs boson decay (excluding muons and neutrinos), was used. The improvements were up to 11.5 % for $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV, in this case.

4. ZH analysis

> "Ultimately, it's up to the fish..."
>
> *(Lee David Zlotoff, MacGyver*
> *Season 1, Episode 11,*
> *"Nightmares")*

4.1. Introduction

The $H \to b\bar{b}$ decay has not been observed yet, and its observation is the only way to probe the coupling of the Higgs boson to the b-quark [82]. $H \to b\bar{b}$ is the most probable Higgs boson decay, with a branching ratio of 57.5 % at the 125.09 GeV [19] value of the Higgs boson mass. Therefore, it is essential to measure this decay to constrain its total decay width, $\Gamma(H)$, and to understand if the recently found Higgs boson is the Standard Model Higgs boson.

In hadron colliders such as LHC, it is impossible to distinguish the $H \to b\bar{b}$ signal from the direct $b\bar{b}$ production, with several orders of magnitude larger cross section (see figure 1.9). For this reason, this search can only be performed using channels which provide additional clean signatures. In the VH associated production of the Higgs boson with a vector boson, W or Z, the leptonic decay of the vector boson supplies the signatures that allow distinguishing the Higgs boson events from the huge backgrounds.

The VH association production channel is divided into three channels: in the 0-lepton channel, a Z boson decays into a pair of neutrinos and the event is selected by a large missing transverse energy; the 1-lepton channel corresponding mainly to WH associated production in which the W boson decays to a charged lepton and a neutrino; while the 2-lepton channel corresponds mainly to the ZH associated production, when the Z boson decays into a pair of charged leptons. In all these cases, there are in addition two b-jets from the Higgs boson decay. In this thesis, only muons and electrons are considered as charged leptons. The combined study of the three channels is referred as VH analysis in this thesis, while the

2-lepton channel study, which is the object of study in the thesis, is referred as ZH analysis.

The main disadvantage of the Higgs associated production with a vector boson compared to the other production mechanisms is the lower cross section. For the WH associated production, the cross section is 0.703±0.018 pb, whereas for ZH it is nearly a factor of two smaller: 0.414±0.016 pb. These values are calculated at *next-to-next-leading-order* in QCD and *next-leading-order* in electroweak corrections at the centre of mass energy of $\sqrt{s} = 8$ TeV [22]. Also, the W and Z bosons decay further, so the product between the cross section, σ, and the branching ratio, BR, describes better the difference between channels. The BR of the $W \to \ell\nu$ decay is around 10 %, while for the Z boson decaying to neutrinos, the BR is about 20 % and for the Z boson decay to charged leptons the BR is around 3 %. Taking this into account, the $\sigma\times$ BR product for the $WH \to \ell\nu b\bar{b}$ is around four times larger than for the $ZH \to \ell\ell b\bar{b}$ channel, as shown in table 4.1.

Channel	σ (pb)	BR (V→ leptons)	$\sigma \times BR$ (fb)
$ZH \to \nu\nu b\bar{b}$	0.414±0.016	20.00±0.06 %	47± 3
$WH \to e\nu b\bar{b}$	0.703±0.018	10.75±0.13 %	42± 2
$WH \to \mu\nu b\bar{b}$	0.703±0.018	10.57±0.15 %	42± 2
$ZH \to eeb\bar{b}$	0.414±0.016	3.363±0.004 %	8.6±0.5
$ZH \to \mu\mu b\bar{b}$	0.414±0.016	3.366±0.007 %	8.7±0.5

Table 4.1.: Calculation of the product $\sigma \times$ BR. Branching ratio includes also the leptonic decays of taus [4].

Notwithstanding, the presence of two high-p_T isolated leptons reduces considerably the multi-jet background so in the 2-lepton channel this background will be very small. Also, the ZH event is entirely reconstructed (two charged leptons and two b-jets), allowing the improvement of the background rejection and the application of a global kinematic fit to improve the $b\bar{b}$ mass resolution.

4.2. Overview of the $ZH \to \ell\ell b\bar{b}$ analysis

Figure 4.1 shows the manipulation process of events in the ZH analysis. The first step is to select only events that have fired the leptonic triggers used in the analysis. Quality criteria are applied to select events in which the detector was working in good conditions, and well-reconstructed electrons, muons and jets, which are referred to as physics objects.

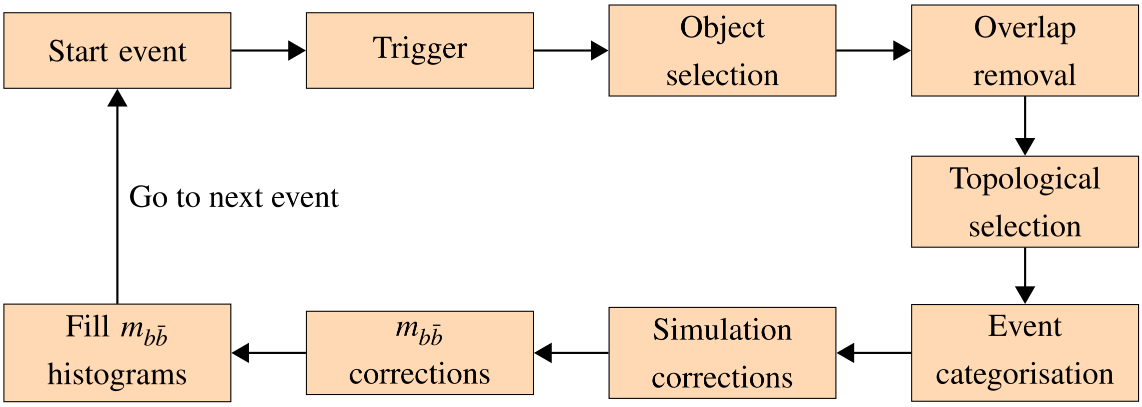


Figure 4.1.: Scheme of the treatment of events in the ZH analysis.

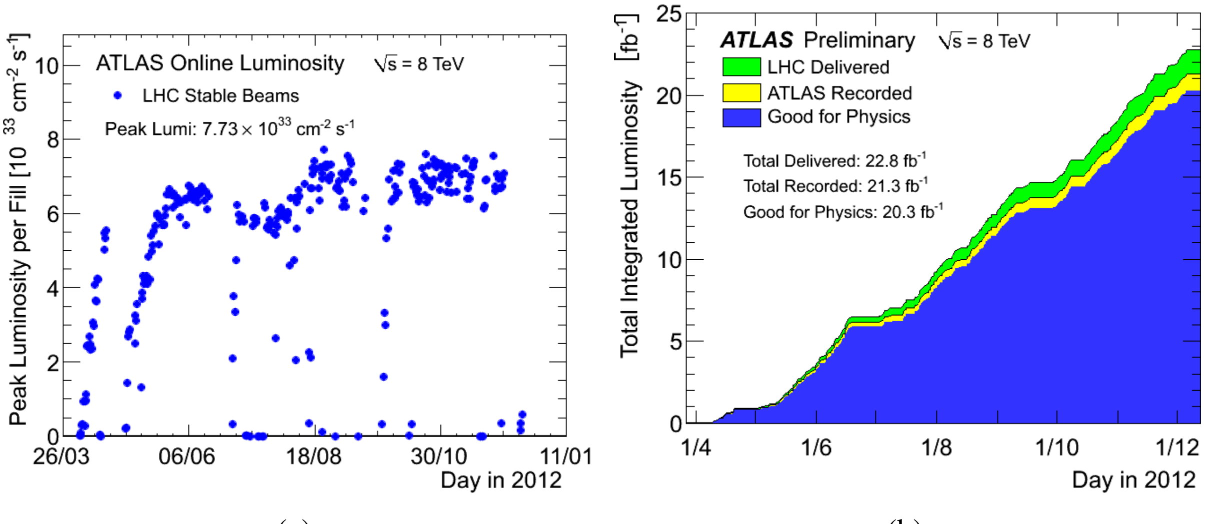
Since the reconstruction and identification algorithms do not apply any overlap removal by default, an electron might also be reconstructed as a jet, or if a muon radiates it might be reconstructed as an electron. Therefore, overlap removal criteria are used to reject repeated physics objects. The particular properties of the $ZH \to \ell\ell b\bar{b}$ events are used to distinguish between the ZH signal and background events and referred to as topology selection. Events are then categorised according to the multiplicity of jets, the transverse momentum of the reconstructed Z boson, and the MV1c weight of each jet used for the identification of b-jets. Some physics object properties are not modelled correctly in the simulations and corrections are derived from data-MC comparisons or obtained from higher order simulations. The analysis is further enhanced by improving the resolution of the reconstructed jets that result in the Higgs candidate. One of them, the global sequential calibration, is one of the objects of study of this thesis (see section 3.9). The invariant mass of the reconstructed Higgs boson is then registered in histograms for each of the categorisation regions for further analysis.

The effect of the systematic uncertainties is evaluated by running the full analysis but varying in the simulation each source of systematic uncertainty by one standard deviation. The resulting modified distributions of the Higgs boson invariant mass are also registered for further analysis. Other systematic uncertainties result from the modelling of signal and backgrounds, which include those obtained from the corrections to the simulation.

Finally, a statistical procedure uses all the invariant mass distributions from the nominal analysis and the systematic uncertainties to normalize simultaneously all the backgrounds to fit the data. At the same time, this global fit will extract the value of the signal strength parameter, μ, as discussed in section 1.6.

4.3. Data sample and integrated luminosity

The LHC had an outstanding performance during 2012: the instantaneous luminosity increased from around 0.5×10^{33} $\mathrm{cm^{-2}s^{-1}}$ [47] at the beginning of 2011 to almost 8×10^{33} $\mathrm{cm^{-2}s^{-1}}$ in 2012. In figure 4.2 (a), the evolution of the peak luminosity per LHC fill along 2012 is shown, reaching a maximum of 7.73×10^{33} $\mathrm{cm^{-2}s^{-1}}$. The total integrated luminosity as a function of the day in 2012 is shown in figure 4.2 (b). The evolution of the integrated luminosity delivered by LHC to ATLAS appears in green with a cumulative value of 22.8 $\mathrm{fb^{-1}}$, whereas the ATLAS trigger system recorded 21.3 $\mathrm{fb^{-1}}$ (in yellow) corresponding to a data taking efficiency of 93.4%. Table 4.2 gives the percentage of each sub-detector up-time and good quality data. The combined efficiency was 95.5%, meaning that 20.3 $\mathrm{fb^{-1}}$ of data has good quality to be used in physics analysis, as shown in blue in the figure 4.2 (b).



(a) (b)

Figure 4.2.: Peak luminosity per fill during 2012 data taking period (a) and total integrated luminosity delivered, recorded and good for analysis (b) [47].

4.4. ZH signal and background characterization

In the ZH analysis, the production of a Higgs boson with a Z boson is called the signal process. There are however other physics processes similar to the signal event, known

ATLAS $p-p$ run: April-December 2012										
Inner Tracker			*Calorimeters*		*Muon Spectrometer*				*Magnets*	
Pixel	SCT	TRT	LAr	Tile	MDT	RPC	CSC	TGC	Solenoid	Toroid
99.9	99.1	99.8	99.1	99.6	99.6	99.8	100.	99.6	99.8	99.5
All good for physics: 95.5%										

Table 4.2.: Percentage of the up-time with good quality data of each sub-detector obtained during 2012 data taking [137].

as backgrounds. The leading or irreducible backgrounds are those that present the same signature as the signal process and have usually a higher cross section production. Other backgrounds, known as reducible, result from a misreconstruction of any physics object, faking the same signature as the signal. Although the efficiency of the event selection of the analysis rejects most of such events, their cross section is even higher than the irreducible backgrounds and several events can be accepted.

Different Monte Carlo (MC) generators were used to simulate the hard scattering process for the signal and each of the backgrounds in the ZH analysis. Table 4.3 summarises for each sample, the corresponding generator, parton density function (PDF), underlying event (UE) tune and order of cross section computation and will be described in next sections.

Almost all samples use the ATLAS FAST II detector simulation to speed up simulations and increase the available statistics for each MC sample as discussed in section 2.8.

The simulations have to be normalised to the total integrated luminosity of the data, to allow for data-MC comparisons. The luminosity of an MC simulated sample, *MClumi*, is calculated as:

$$MClumi = \frac{\sum w^i}{\sigma \times \mathrm{BR} \times \varepsilon_F}, \tag{4.1}$$

where σ is the cross section of the given process, BR is the branching ratio of the produced particles decay, ε_F is the generator level filter efficiency of the process and $\sum w^i$ is the sum of the product of all MC weights. The MC weights include only the pileup, the vertex weight and the $\mathrm{p_T^Z}$ weight for the signal sample, that will be discussed in section 4.7.

Process	Generator	PDF	UE tune	Calculation order	Nevents	σ (fb)	$\sum w^i$
ZH signal with Higgs mass of 125 GeV							
qqZH ($\rightarrow \ell\ell b\bar{b}$)	Pythia8	CTEQ6L1	AUET2	NNLO QCD/NLO EW	2998998	22	2.92×10^{6}
ggZH ($\rightarrow \ell\ell b\bar{b}$)	Powheg+Pythia8	CT10		NLO QCD	300000	1.9	194.0
Top backgrounds							
$t\bar{t}(\rightarrow \ell\nu b\ell\nu\bar{b})$	Powheg+Pythia6	CTEQ6L1	P2011C	NNLO+NNLL	99930891	1.4×10^{5}	9.951×10^{7}
single-t (s channel)					5995993	1.8×10^{3}	5.971×10^{6}
single-t (Wt channel)					19937980	2.2×10^{4}	1.985×10^{8}
single-t (t channel)	AcerMC+Pythia6				8996990	2.8×10^{4}	7.672×10^{6}
Diboson backgrounds							
ZZ ($\rightarrow \ell\ell jj$)	Powheg+Pythia8	CT10	AUET2	NLO QCD	3999995	1.3×10^{3}	4.561×10^{6}
WZ ($\rightarrow jj\ell\ell$)					1500000	1.6×10^{3}	2.382×10^{6}
WZ ($\rightarrow \ell\nu jj$)					9999988	4.9×10^{3}	4.849×10^{7}
WW ($\rightarrow \ell\nu\ell\nu$)					9999994	5.2×10^{4}	5.081×10^{7}
V+jets backgrounds							
Z ($\rightarrow \ell\ell$) + b	Sherpa 1.4.1	CT10	Sherpa	NNLO QCD NLO EW	11994991	3.5×10^{4}	3.274×10^{6}
Z ($\rightarrow \ell\ell$) + c					8936988	3.5×10^{5}	4.272×10^{6}
Z ($\rightarrow \ell\ell$) + l					14962997	8.5×10^{5}	8.472×10^{6}
W ($\rightarrow \ell\nu$) + b					44913447	1.5×10^{5}	1.298×10^{7}
W ($\rightarrow \ell\nu$) + c					29985457	5.5×10^{5}	8.860×10^{6}
W ($\rightarrow \ell\nu$) + l					149653900	1.1×10^{7}	8.356×10^{7}

Table 4.3.: Information on the simulation samples for the signal and backgrounds. The generator's column include also the simulation used for the hadronization. Also present is the parton distribution functions (PDF), the underlying event (UE) tune, the number of generated events (Nevents), calculation order, cross-section (σ) and the number of generated events taking into account all coorections and weigths ($\sum w^i$) are presented. The cross-section include the filter efficiency, branching ratio and the k-factors discussed in section 2.8. Further details are given in section 4.4.

Table 4.3 also provides information on the absolute number of simulated events, Nevents. The cross-section, σ, of each process in this table is the product of the inclusive cross-section by the leptonic/hadronic filter efficiency and the higher-order k-factors discussed in section 2.8. Uncertainties on the cross-section of the signal and background samples are discussed in section 4.10. The MC weight sum, in the last column of the table is used for the calculation of the simulation luminosity for each sample using equation 4.1. Finally, ℓ refers to the three charged leptons and ν is the corresponding neutrino and jj corresponds to a hadronic decay of a vector boson, $W \to q'\bar{q}$ or $Z \to q\bar{q}$, where q' is a quark with a different flavour and absolute charge from q.

The distributions for each MC sample will then be scaled by:

$$\text{scale} = \frac{\int \mathscr{L}^{\text{data}}}{MClumi} = \frac{20.3}{MClumi(fb^{-1})} \tag{4.2}$$

4.4.1. Signal process

The primary production Feynman diagram of the Higgs in associated production with a Z boson is shown schematically in figure 4.3: a quark-anti-quark interaction produces a virtual Z boson, which decays into a real Z boson and a Higgs boson. Additionally, in the search presented in this thesis, the Z boson decays either to an electron-positron or a muon-anti-muon pair and the Higgs boson decays to a pair of b-quarks. The total cross section also receives a non-negligible contribution from two other Feynman diagrams, both initiated by gluons. In figure 4.4 (a), a triangle quark loop will produce the virtual Z boson. In figure 4.4 (b), the final Z and Higgs bosons are produced directly in the square loop. Together, the gluon-initiated ZH production that includes the Feynman diagrams shown in figure 4.4, contributes with 32.46 fb (8.5 % of the total cross section) [138].

The $ZH \to \ell\ell b\bar{b}$ process is referred as signal and the Feynman diagrams shown in figures 4.3 and 4.4 represent the tree-level production and decay of this process. Higher order diagrams are also accounted as the signal, although they were calculated only with *next-to-next-to-leading-order* (NNLO) and *next-to-leading-order* (NLO) for QCD and electroweak corrections, respectively.

Both signal samples use the ATLAS underlying event tune 2 (AUET2) [102, 103] to simulate the underlying event. The branching ratio and decay widths are calculated with HDECAY [106], described in section 2.8.

The Higgs associated production with a Z boson initiated by gluons was simulated with

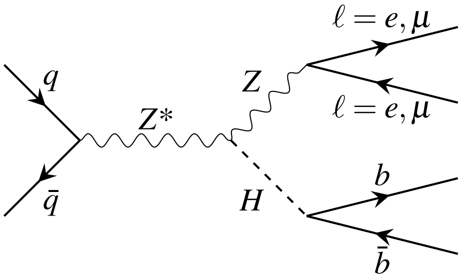
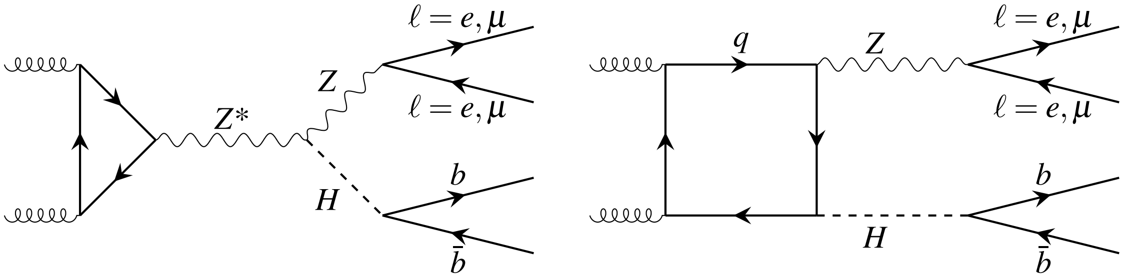


Figure 4.3.: Feynman diagram of the production of the Higgs boson associated to a Z boson, initiated by quarks.



(a) (b)

Figure 4.4.: Feynman diagrams of the production of the Higgs boson associated to a Z boson, initiated by gluons through a triangle (a) and a squared (b) loop.

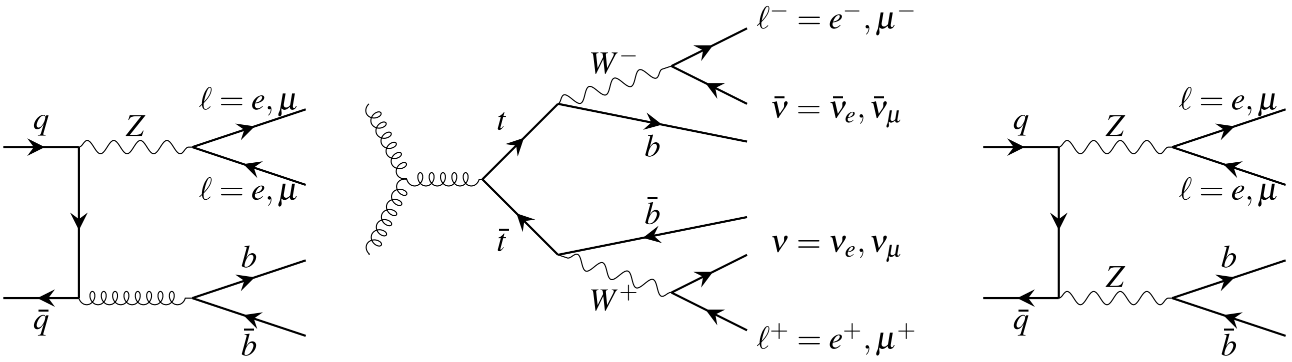
the POWHEG [93] event generator for the hard scattering process, interfaced with PYTHIA8 for the shower propagation and the cross-section calculation was obtained at *next-to-leading-order* in QCD. This sample uses the PDF set CT10 [84]

In the ZH sample, the Z boson is forced to decay into a pair of charged leptons (including tau leptons) and the Higgs boson to a b-quark pair.

4.4.2. Leading backgrounds

The leading backgrounds for the ZH analysis, by the order of importance, are the production of a bottom-anti-bottom quark pair together with a Z boson (Z+b-jets), the production of a top-anti-top pair of quarks ($t\bar{t}$) and the production of a pair of Z bosons (ZZ).

The production of Z+b-jets is the most important background. A quark-anti-quark pair



(a) Z+b-jets production (b) $t\bar{t}$ production (c) ZZ production

Figure 4.5.: Leading order Feynman diagrams of the three irreducible backgrounds of the ZH channel: Z+b-jets (a), $t\bar{t}$ (b) and ZZ (c).

interacts producing a Z boson which decays into an electron-positron or a muon pair. One of the quarks radiates a gluon which is split into two b-quarks. The Feynman leading order diagram for this process is shown in figure 4.5 (a). Also considered as part of this background is the decay from the Z boson to tau-leptons, which can also decay to electrons or muons. The simulation of the Z+b-jets sample is performed with SHERPA 1.4.1 [92]. The cross section is calculated at *next-to-next-to-leading-order* in QCD and uses the CT10 PDF set. SHERPA deals with the underlying event internally. The simulation of the Z+jets background is divided into three exclusive samples: events with at least one bottom hadron (Z+b), events with at least one charmed hadron but no bottom hadron (Z+c) and events with neither a bottom hadron nor a charmed hadron (Z+light). The bottom and charmed hadrons selected by the filter are presented in table 2.2. In addition, there are six exclusive slices depending on the p_T of the Z boson, 0 - 40 GeV, 40 - 70 GeV, 70 - 140 GeV, 140 - 280 GeV, 280 - 500 GeV and above 500 GeV to increase the number of events with high transverse momentum of the vector boson. Table 4.3 shows the total number of events for all p_T^Z slices and the full list will be shown in appendix C.1.

In the $t\bar{t}$ production, each of the top quarks decays to a b-quark and a W boson ($t \rightarrow Wb$) with a *BR* in the Standard Model close to 100 %. This process is a background in the ZH analysis when both W bosons decay to the same lepton flavour and a neutrino ($W \rightarrow \ell\nu$). One of the Feynman leading order diagram for this process is presented in figure 4.5 (b). For the simulation of the $t\bar{t}$ sample, the hard scattering process for this sample is processed with POWHEG, interfaced with PYTHIA6 [88] for the hadronization procedure. The $t\bar{t}$ cross section is obtained at *next-to-next-to-leading-order* in QCD including *next-to-next-to-leading-log* (NNLL) of soft gluon terms with TOP++. The uncertainties on the cross-section from PDF and α_S are calculated using the PDF4LHC prescription [139] using the MSTW2008 68% CL NNLO [140, 141], CT10 NNLO [84, 142] and NNPDF2.3 5f FFN [143] PDF sets, added in quadrature to the scale uncertainty. The PDF set chosen was CTEQ6L1 and the underlying event used was the Perugia2011C tune [102, 103]. A leptonic filter was applied in this sample: at least one of the W bosons produced in the top decay had to decay leptonically. The inclusive cross-section for the $t\bar{t}$ production in 8 TeV is $\sigma_{t\bar{t}} = 253^{+13}_{-15}$ pb and after applying the leptonic filter the value is about 140 pb.

In the diboson ZZ production, one of the Z bosons decays into a pair of charged leptons ($Z \rightarrow \ell\ell$) and the other decays to a pair of b-quarks. The Feynman leading order diagram for this process is presented in figure 4.5 (c). The simulation of the ZZ process is performed using the POWHEG generator, with cross sections calculated at *next-to-leading-order*, and the hadronization was dealt with PYTHIA8. It uses the CT10 PDF set and the AUET2 tune to simulate the underlying event. A filter was applied to this sample to ensure that one of the Z bosons has a leptonic decay and the other a hadronic one.

The multi-jet background can not be simulated in particular due to the difficult of the fake leptons modeling. This background is therefore estimated from data with the method discussed in section 4.4.4.

4.4.3. Remaining backgrounds

From the remaining backgrounds, the only one with a relevant contribution to the ZH analysis is the production of a Z boson with two jets not resulting from b-quarks but nonetheless misidentified as b-jets. The leading order Feynman diagrams for the production of non-b-jets in association with a Z boson are shown in figure 4.5 (a), if the b-quarks are replaced by a c-quark or by a light flavoured quark. As discussed in section 2.7.6, the b-tagging algorithm has an inefficiency of 9.57 % for c-jets and 0.74 % for light-jets at the intermediate operating point. However, the cross section to produce Z+non-b-jets events is about one order of magnitude higher that to produce Z+b-jets events. The SHERPA generator with the veto on B-hadrons, discussed in section 4.4.2, is used to simulate this background.

Other two backgrounds with still some relevance to the ZH analysis are the production of a top quark and a W boson (Wt) and the production of a Z and a W bosons (ZW). In the former, the top quark decays nearly 100 % of the times to Wb. The event contains two W bosons that may decay to a lepton of the same flavour. If there is another jet in the event, this might be (mis)identified as a b-jet and produce the same final state as the ZH process. In the ZW, the Z boson decays into a charged lepton pair and the W boson decays into two quarks. If the jets originated from these quarks are misidentified as b-jets, the event becomes a background to the ZH channel. The Wt process is simulated with the ACERMC [95] generator and the hadronization is handled by PYTHIA6. The PDF set and underlying tune are the same used for $t\bar{t}$ (CTEQ6L1 and P2011C, respectively) and the cross sections are calculated at *leading-order*. The simulation details of the WZ process are the same as for the ZZ diboson production and a filter is applied to select events where the W boson decays leptonically.

Other backgrounds have a negligible impact in the ZH analysis. However, since the final VH search includes also the 0-lepton and the 1-lepton channels, their backgrounds are shortly described.

The leading backgrounds of the 0-lepton channel is also the Z+jets production but with the Z boson decaying to neutrinos. A similar change occurs in the ZZ background. The same generator, PDF set, underlying event tune and hadronization as the corresponding simulation for the 2-lepton channel are used. Since two charged leptons will be required for the ZH analysis (see section 4.5) and these samples are created without them, they were ignored for the ZH analysis.

The leading backgrounds of the 1-lepton channel is the W+jets production. This process differs from the Z+jets process by only one less charged lepton. However, the cross section production of W+jets is about a factor of 10 higher than Z+jets. If there is an additional charged lepton in the event, this process can produce the same final state as the signal ZH. In particular, if the W boson is produced with a pair of b-jets and one of the b-jets decays leptonically, this new lepton might be reconstructed as isolated. If the b-jet decays hadronically instead, one of the charged hadrons might be reconstructed as a fake electron. The simulation of the W+jets background is the same as that for Z+jets.

Three more processes are considered as backgrounds for the WH channel: the production of a pair of W bosons (WW), and the production of a single top quark in the s- and t-channels. For the WW process, with a similar simulation as other diboson processes, the background to the ZH analysis, results from the combination of two effects, the (mis)identification of additional jets in the event as b-jets and the reconstruction of the two leptons from the W boson decays as originated from the Z boson decay. In the production of a single top in the t-channel, a gluon splits in two b-quarks and one of them interacts with another quark through a W boson producing a top quark. The top decays as before, resulting in an event with two b-jets and a lepton. If one of the b-quarks decays also leptonically, that lepton might be reconstructed as isolated. Finally, in the single top s-channel, a top quark is produced together with a b-quark. The top decays into Wb and the W may decay into a charged lepton and the corresponding neutrino. One of the B-hadron decay with electrons or muons (BR $\sim 21\,\%$ combined) provides the additional lepton of the same flavour if it is reconstructed as isolated. These backgrounds are expected to give a small to negligible contribution on the $ZH \to \ell\ell b\bar{b}$ analysis, being more important for the 1 lepton channel. These two single top production processes are simulated with the POWHEG generator and the same configuration as the Wt-channel process for the remaining simulation steps.

4.4.4. Multi-jet background estimation and uncertainty

The multi-jet background can not be simulated and is estimated by data-driven methods. Multi-jet background arise from two jets that are misidentified as leptons or from leptonic decays in the jets when the leptons are reconstructed as isolated.

A multi-jet dominated sample is obtained by requiring non-isolated leptons. This test uses all tracks inside a $\Delta R = 0.2$ cone, excluding the lepton track, around the lepton and requires that the scalar sum of their transverse momentum is larger than $10\,\%$ of the transverse momentum of the lepton:

$$\frac{p_{\mathrm{T}}^{\Delta R=0.2},\mathrm{tracks}}{p_{\mathrm{T}}^{\mathrm{lepton}}} > 0.1 \tag{4.3}$$

This region defines templates for the multi-jet background for different distributions, assuming that the shape of those distributions does not depend on the isolation cut.

To obtain the correct normalisation of this multi-jet template, a fit is performed using the distribution of the invariant mass of the two leptons. Events selected for this fit are required to contain exactly two leptons and at least two jets. The fit uses $m_{\ell\ell}$ distributions from three processes: the Z+jets background, the multi-jet template and all other backgrounds considered together. The normalisation of Z+jets and the multi-jet template are allowed to float to fit the number of background events to the data distribution, keeping all other backgrounds fixed. This fit is done independently in events where two, one or no jets are identified with the b-tagging algorithm MV1c, obtaining normalization factors of

2.22 ± 0.20, 1.89 ± 0.04 and 2.36 ± 0.02, respectively.

The multi-jet contribution is of the order of a few percent for electrons and negligible for muons.

4.5. Event selection

4.5.1. Triggers

Table 4.4 presents the muon and electron triggers used for the event selection. The ZH analysis uses the lowest unprescaled triggers threshold for the main triggers of the analysis: mu24i_tight for muons and e24vhi_medium1 for electrons. Both triggers use a track isolation requirement to reduce the sensibility to pileup and have a threshold of 24 GeV on the reconstructed transverse energy (momentum) of the electron (muon) in the HLT. The isolation condition is achieved if the ratio between the scalar sum of the transverse momentum of all charged particles (except the lepton) within a $\Delta R = 0.2$ cone from the lepton and the transverse momentum of the lepton is below the threshold present in the table. Furthermore, to recover the loss of efficiency at high-p_T due to the isolation criteria, the ZH analysis uses the mu36_tight muon and e60_medium1 electron triggers without isolation, with a threshold of 36 GeV and 60 GeV, respectively. Finally, the 2-lepton analysis also uses events selected by the di-lepton triggers 2mu13 and 2e12Tvh_loose1. These require at least two leptons with smaller thresholds: 13 GeV for muons and 12 GeV for electrons to improve the efficiency for low-p_T leptons, below 24 GeV. Finally, if one event is recorded in both *Muon* and *Egamma* streams (see section 2.3), it is removed from the latter.

4.5.2. Quality criteria

Data events are grouped into runs that lasted for several hours and are separated in luminosity blocks with a duration of about one minute each. Data quality cross-checks of all data taken are used to produce a good runs list (GRL) to ensure that only data with good performance of the full detector is used in the analysis.

Additionally, events might be flagged as problematic during the reconstruction due to detector malfunction and are rejected at analysis level:

- LAr error: noise burst in the liquid Argon electromagnetic and hadronic calorimeters.

Stream	Trigger name	Number	Threshold	Track isolation
Muon	mu24i_tight	$\geq$1	24 GeV	0.12
	mu36_tight	$\geq$1	36 GeV	none
	2mu13	$\geq$2	13 GeV	none
Egamma	e24vhi_medium1	$\geq$1	24 GeV	0.10
	e60_medium1	$\geq$1	60 GeV	none
	2e12Tvh_loose1	$\geq$2	12 GeV	none

Table 4.4.: List of triggers used in the ZH analysis in 2012. The track isolation is defined as the fraction between the sum of the tracks p_T in a $\Delta R = 0.2$ cone and the p_T of the particle that triggers the event.

- Tile error: noise burst in the Tile calorimeter.
- Tile trips: a low voltage power supply in a module was sporadically switched off [144].
- Incomplete events resulting from hardware problems.
- Bad jet events: reconstructed jets that do not correspond to real jet energy deposits in the calorimeter but to hardware problems, to the interaction of the beam with residual gas or the collimator or cosmic muons overlapping with a jet [115].

Additionally, the event is required to have at least one vertex with three or more tracks associated to ensure that at least one collision took place.

4.5.2.1. Muons selection

In section 2.7.1, five types of muons were described. For the muon selection, the *Combined* and *Segment Tagged* muons are treated in the same way, as single category: μ^{CB+ST}. The same happens for the *Standalone* and *Silicon Associated Forward* muons, combined in the category μ^{SA+FW}. The *Calorimeter-tagged* muons, μ^{Calo}, are treated as a single category.

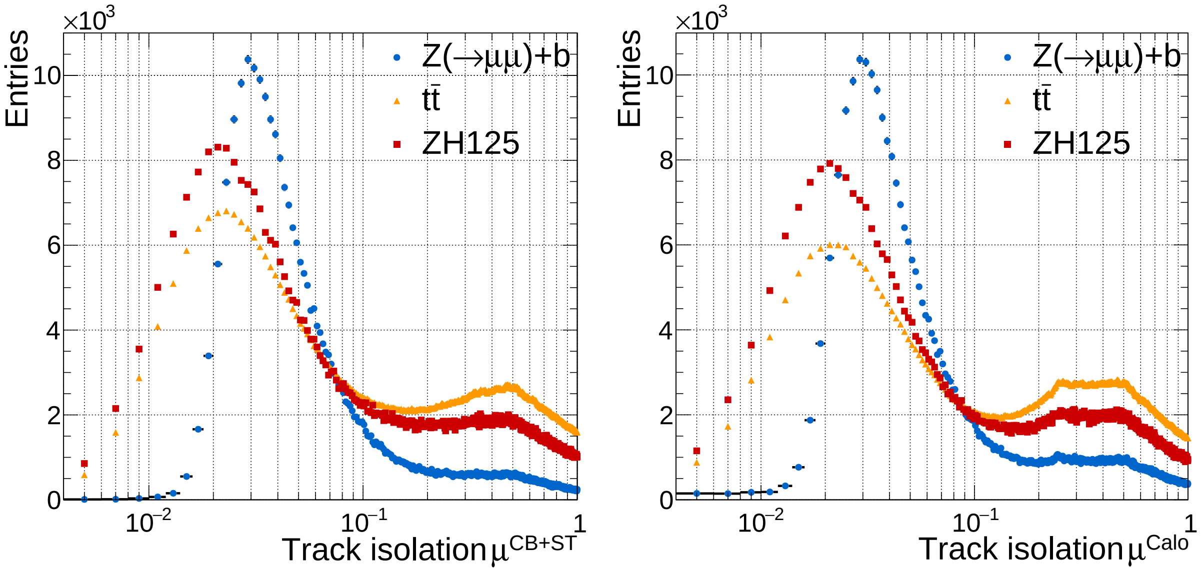
Table 4.5 indicates the selection conditions for the three categories of muons. Since the μ^{CB+ST} and the μ^{Calo} muon categories use the inner detector, the tracking quality requirements described in section 2.7.1.1 are applied to them: at least one pixel and five SCT hits, no more than two silicon holes and, for μ^{CB+ST} muons, at least five TRT hits [59]. The number of pixel hits includes the number of dead sensors. The same happens for the number

Selection	Muon category		
	μ^{CB+ST}	μ^{SA+FW}	μ^{Calo}
Inner Detector track quality criteria			
Number of pixel hits	$\geq$1	none	$\geq$1
Number of SCT hits	$\geq$5	none	$\geq$5
Number of Si holes	$\leq$2	none	$\leq$2
Number of TRT hits	$\geq$5	none	none
Impact parameter criteria			
$\lvert d_0 \rvert$	< 0.1 mm	none	< 0.1 mm
$\lvert z_0 \sin\theta \rvert$	< 10 mm	none	< 10 mm
Selection conditions			
Pseudorapidity	$\lvert\eta\rvert < 2.7$	$2.5 < \lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 0.1$
p_T	> 7 GeV	> 7 GeV	> 20 GeV
Track isolation	0.1	none	0.1

Table 4.5.: List of selection conditions for the muons.

of SCT hits. The TRT hits include the TRT outliers, which cannot be more than 90 % of the value. These criteria ensure the quality of the muon track reconstruction. The μ^{CB+ST} and μ^{Calo} muon types are also required to have a small impact parameter with respect to the leading primary vertex, $|d_0| < 0.1$ mm and $|z_0 \sin\theta| < 10$ mm to reduce the probability that the muon originates from a leptonic decay of a heavy hadron or from a pileup interaction. A μ^{Calo} muon is rejected if it distances less than $\Delta R = 0.1$ from a μ^{CB+ST} muon, as the two algorithms could reconstruct the same muon. The muon with higher quality, μ^{CB+ST}, is kept and the other one discarded. If a muon trigger selects the event, the muon(s) that pass the selection is (are) required to be the one(s) that fired the trigger, *i.e.*, if the distance between the EF muon and the reconstructed muon is smaller than 0.1, in the $\eta - \phi$ plane. μ^{CB+ST} and μ^{SA+FW} muons are required to have $p_T > 7$ GeV, to be reconstructed in the muon spectrometer. The identification of μ^{Calo} muons was optimised for $p_T > 20$ GeV [59], so a tighter p_T threshold of 20 GeV is used for those.

Finally, a muon which has a track in the inner detector (μ^{CB+ST} and μ^{Calo}), must have its track isolated from other tracks. The isolation parameter is defined as the ratio between the scalar sum of the p_T of tracks, excluding the muon, that fall in a cone of $\Delta R = 0.2$ around the muon and the p_T of the muon. The distribution of this parameter is shown in figure 4.6



(a) μ^{CB+ST} muons (b) μ^{Calo} muons

Figure 4.6.: Normalised distributions to the same number of entries of the track isolation for μ^{CB+ST} (a) and μ^{Calo} (b) muons. The signal ZH sample (red circles), $t\bar{t}$ (yellow triangles) and Z+b-jets (blue squares).

for the ZH, $t\bar{t}$ and $Z \to \mu\mu$ samples, normalised to the number of events of the latter sample. Muons accepted for the ZH analysis must have this ratio below 0.1 to reject multi-jet events [1]. This selection is also efficient to reject part of the $t\bar{t}$ background.

All the muons fulfilling the criteria referred are labelled *loose muons*. Furthermore, to select muons originating from the leptonic decay of a Z boson, the transverse momentum of at least one muon is required to be larger than 25 GeV. Muons that fulfil this last requirement and belong to the μ^{CB+ST} category are labelled *signal muons*.

4.5.2.2. Electron selection

Electrons are reconstructed with $E_T > 7$ GeV and $|\eta_{Cluster}| < 2.47$ (see section 2.7.2). Additionally, electrons in the ZH analysis must be reconstructed using the *VeryLooseLH* identification criteria [66].

If an electron trigger selects the event, the reconstructed electron(s) is (are) required to be the one(s) that fired the trigger. Like in the muon selection, this implies that it should be

---

[1] The complementary selection is used when choosing muons for multi-jet events as it was described in section 4.4.4.

closer than 0.1 in the η,ϕ plane to the closer EF reconstructed electron.

Electrons must be isolated from other charged particles, requiring the scalar sum of the p_T of the tracks in an $R = 0.2$ cone around the electron track to be inferior to the 10 % of the electron E_T. This requirement removes most of the multi-jet background, in which electrons originate from leptonic decays within jets and photon conversion processes. Figure 4.7 shows the track isolation variable distribution for the same three samples as for muons, Z+jets, $t\bar{t}$ and ZH, before the electron selection conditions. The distributions are normalised to the number of events of the Z+b-jets sample to compare their shapes. Besides removing the multi-jet events, this criterion is also very efficient to reject $t\bar{t}$ events. As for the muon case, inverting this selection is used to form the multi-jet template.

An electron passing these selection criteria is labelled *loose electron*. The electrons in this analysis result from the leptonic decay of a Z boson, so *loose electrons* with $E_T > 25$ GeV are also labelled *signal electrons*.

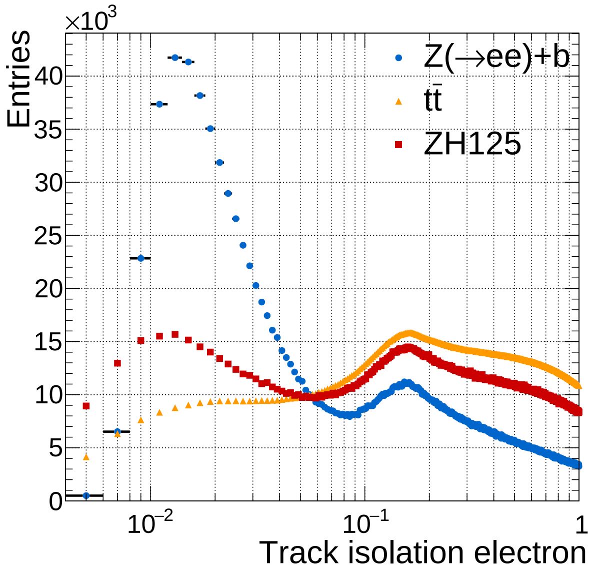


Figure 4.7.: Distributions for electrons of the track isolation obtained for the signal ZH sample in red, $t\bar{t}$ in dark yellow and Z+b-jets in blue, normalised to the number of entries from Z+b-jets.

4.5.2.3. Jets

Jet reconstruction and performance were discussed in some detail in chapter 3. Here, the particular selection for the ZH analysis is presented.

Jets are reconstructed with the anti-k_T algorithm [75] with parameter $R = 0.4$ and calibrated at the EM+JES scale. The global sequential calibration (GSC) is applied on top of the EM+JES to improve the jet energy resolution.

In the ZH analysis, jets are defined in the pseudorapidity range of the calorimeter, $|\eta| < 4.5$. The $\mathrm{p_T}$ of jets must be above 20 GeV in the central region ($|\eta| < 2.5$) and above 30 GeV in the forward region, in which the jet energy scale and uncertainty are well set (see chapter 3). For low-$\mathrm{p_T}$ jets, $\mathrm{p_T} < 50$ GeV, a JVF threshold of 0.5 was applied to reject jets from pileup. Jets that fulfil these criteria are labelled *loose jets*. If a loose jet is in the central region, it is also designated *signal jet*.

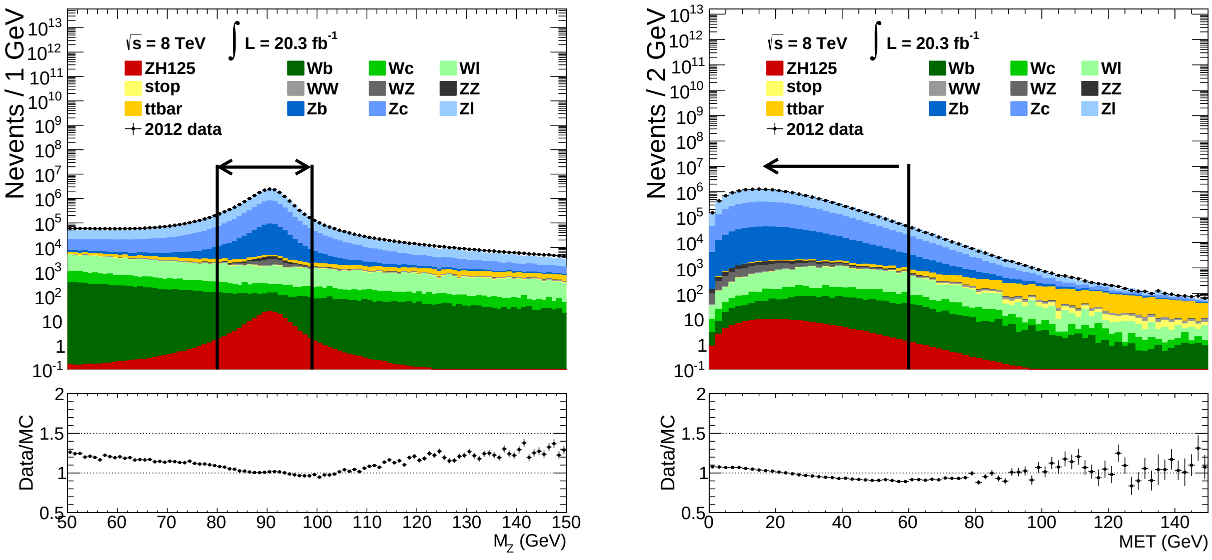
4.5.2.4. Overlap removal

When the detector reconstructs two different objects in the same position, an overlap removal procedure is used to decide which one to keep and which one to discard. Electrons leave an energy deposit in the EM calorimeter, and most of them are reconstructed and pass the selection criteria of jets. The same effect can happen with muons that radiate hard Bremsstrahlung. In some cases, there can be a semi-leptonic decay inside a jet and a muon emerging might be reconstructed. Finally, muon Bremsstrahlung radiation might be reconstructed as an electron. For the particle not to be counted twice, the ZH analysis performs the following overlap removal steps in sequence:

el-jet	All jets within a distance $\Delta R < 0.4$ from an electron are removed.
mu-jet	A jet with less than four associated tracks is removed if it is within a distance $\Delta R < 0.4$ from a muon.
jet-mu	If the jet has more than four associated tracks the muon is removed instead.
mu-el	The electron is removed if it is within a distance $\Delta R < 0.2$ from a μ^{CB+ST} or μ^{SA+FW} muon.
el-mu	If the electron overlaps with a μ^{Calo} muon, the muon is removed because it was not reconstructed in the muon spectrometer and so the electron is considered.

4.5.3. Event topology selection

Events with both types of leptons, electrons and muons, are discarded. There must be at least one signal lepton and one additional loose lepton with the same flavour, in the event. If there are any other additional loose leptons, the event is rejected. At this stage, the topology of the event has exactly two loose leptons, where at least one of them is a signal lepton.



(a) $m_{\ell\ell}$ (b) $\not{E}_{\rm T}$

Figure 4.8.: Distributions of the invariant mass of the lepton pair (a) and the missing transverse energy (b). The arrows indicate the accepted events for the nominated criteria. The histograms are cumulative for the signal and background and the black markers are data points.

The reconstructed invariant mass of the two leptons, $m_{\ell\ell}$, should be compatible with the mass of the Z boson. Only events falling inside the mass region $83 < m_{\ell\ell} < 99$ GeV are accepted. Figure 4.8 (a) presents the reconstructed lepton invariant mass in events passing the quality criteria, trigger and lepton selection. The coloured histograms show the signal and the background samples stacked one after the other, with the black bullets corresponding to the data points with its statistical uncertainty. Each simulation sample is normalised to the integrated luminosity of $\mathscr{L} = 20.3$ fb^{-1}, as discussed in section 4.4. The colour scheme applied to the simulation samples is uniform throughout this chapter and is detailed here: the ZH signal appears in red, top quark samples appear in yellow, Z+jets events in blue, W+jets in green and dibosons in grey. Different shades of green and blue may be used to separate light flavour from heavy flavour jets. Particularly, in figure 4.8 (a), the three shades for V+jets correspond, to the three different filters applied to these backgrounds, as it was explained in section 4.4, darker shade corresponding to a heavier quark in the filter. The same applies to the top (separation between single top and $t\bar{t}$) and the diboson (separation in ZZ, WZ and WW) backgrounds. The bottom panel shows the ratio between the data and the sum of all simulation samples for comparison. The discrepancy of $20 - 30\,\%$ in the data/MC ratio for the regions outside the Z mass window is due to the multi-jet background that results from the data driven method as discussed in section 4.4.4 and was not included. The mass window selects the region where the data/MC ratio is closer to unity, and the multi-jet background has a small contribution. In the $t\bar{t}$ background, where both tops decay leptonically, the two emerging leptons should have no correlation so their invariant mass is spread along the whole range of masses displayed, as it is seen in figure 4.8 (a). The other backgrounds suppressed by the mass window requirement are W+jets, WW and single top, that do not have a Z boson decaying to leptons and therefore the di-lepton invariant mass also spreads evenly across the spectrum.

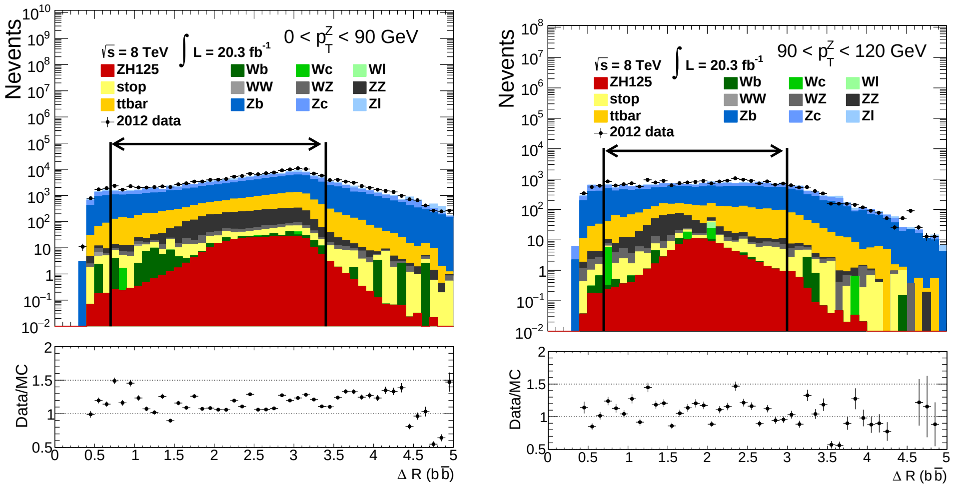
In the $ZH \to \ell\ell b\bar{b}$ analysis, all the final state particles in the event can be reconstructed. Therefore, the missing transverse energy, defined in section 2.7.4, is expected to be small. For that reason, the missing transverse energy is required to be below 60 GeV. This condition reduces further the background with neutrinos in the final state: $t\bar{t}$, single top, W+jets and dibosons WW and WZ. Figure 4.8 (b) shows the $\not{E}_\mathrm{T}$ distribution for events passing the $m_{\ell\ell}$ requirement. The vertical line indicates the $\not{E}_\mathrm{T}$ selection condition. The excess of events observed in data in comparison to MC at low $\not{E}_\mathrm{T}$ is attributed to the multi-jet background. Since neutrinos are not produced in multi-jet events, their missing transverse energy should also be small.

The event is required to have at least two signal jets, *i.e.* two jets reconstructed with the acceptance of the inner detector (so they can be b-tagged) and with $\mathrm{p_T} > 20$ GeV. Events with jets in the forward region ($|\eta| < 2.5$) are vetoed to reduce the $t\bar{t}$ background [2].

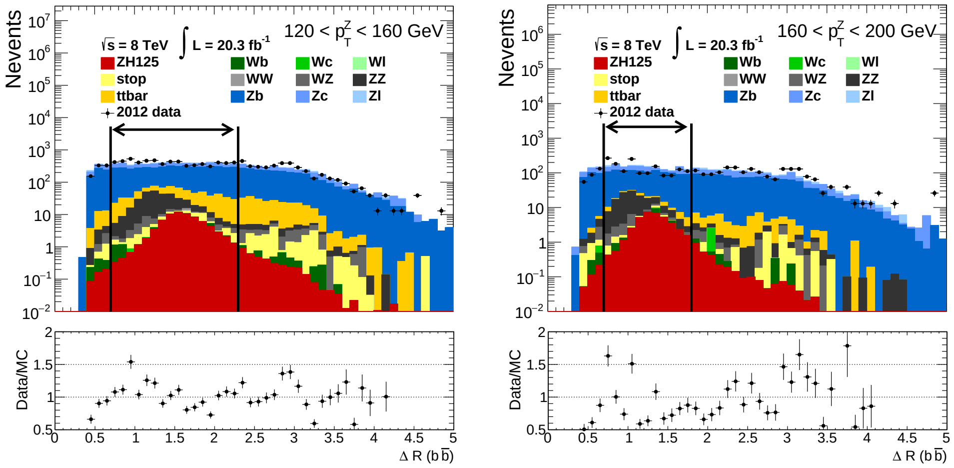
There must be exactly two jets identified as originating from b-quarks (b-tagged jets), and they are required to be the two jets with higher $\mathrm{p_T}$. They are obliged to have an MV1c value greater than 0.405, which corresponds to the 80 % operating point efficiency, as discussed in section 2.7.6. This criterium reduces the Z+jets background, in which either jet is not initiated by a b-quark, and the W+b-jet events, in which a B-hadron decays leptonically and produces an isolated lepton.

---

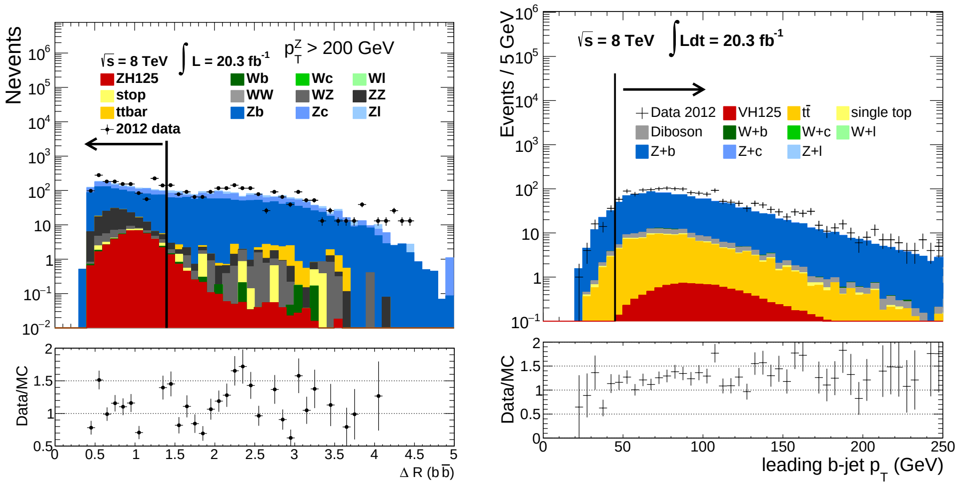
[2] This forward jet veto criteria is driven by the 0- and 1-lepton analysis and is not expected to improve the signal-to-background ratio in this case. In fact, table 4.10 will show a decrease in the $S/\sqrt{S+B}$ ratio for this selection. However, this cut is applied to the 2-lepton analysis to keep the same jet selection as in the other two analysis.



(a) $p_{\mathrm{T}}^{Z} < 90$ GeV (b) $90 < p_{\mathrm{T}}^{Z} < 120$ GeV



(c) $120 < p_{\mathrm{T}}^{Z} < 160$ GeV (d) $160 < p_{\mathrm{T}}^{Z} < 200$ GeV



(e) $p_{\mathrm{T}}^{Z} > 200$ GeV (f) leading b-jet p_{T}

Figure 4.9.: Distributions of the distances between the two b-jets for the five p_{T}^{Z} bins (a-e). Distribution of the transverse momentum of the leading b-jet (f). The arrows indicate the accepted events. The histograms are cumulative for the signal and background, and the black markers are data.

The distribution of the distance ΔR between the two b-jets, in the (η, ϕ)-plane, is shown in figure 4.9 at this stage. The ΔR separation between the two b-jets is required to be above 0.7 if the transverse momentum of the reconstructed Z boson, p_T^Z, is below 200 GeV as shown in the sub-figures 4.9 (a-d), signalised by the vertical line at 0.7. This criterion helps to reduce the Z+b-jets and multi-jet backgrounds, as b-jets from gluon splitting tend to be emitted in the same direction, which does not happen if the b-jet pair results from a colourless particle decay such as the Higgs or the Z boson. The distance between the two b-jets is also required to be below a given p_T^Z-dependent value as shown in table 4.6, which helps to reduce further the $t\bar{t}$ and Z+jet backgrounds. The motivation for this cut is the phase space of the Higgs boson mass which constrains the maximum aperture between the two b-jets. Figures 4.9 (a-e) show the ΔR distribution between the two b-jets and, the vertical line corresponds to the selections.

p_T^Z window (GeV)	< 90	$90 - 120$	$120 - 160$	$160 - 200$	> 200
ΔR upper selection	3.4	3.0	2.3	1.8	1.4

Table 4.6.: p_T^Z dependent selection on the distance ΔR between the two b-jets.

Given that the mass of the Higgs boson is 125 GeV, each of the b-jets is produced on average with an energy around 60 GeV, while the p_T spectrum of the Z+b-jets sample, the dominant background at this point, decreases more rapidly. Hence, the leading jet p_T is required to be larger than 45 GeV. This condition will also reduce the ZZ background by about 30 %. Figure 4.9 (f) shows the distribution of the leading b-jet p_T, with an arrow indicating the events that are accepted. This selection was historically placed between the two ΔR selections just discussed.

Table 4.7 indicates the expected number of events that survive each of the selection conditions, for the MC signal and the main backgrounds samples, Z+jets, ZZ and $t\bar{t}$, normalised to the integrated luminosity. The remaining backgrounds are summed together in the "Other BG" column. The topological selection referred in this section is applied sequentially. The reduction of the number of events for each sample is shown in the bottom row in percentage and calculated by the ratio between the survivals of the full topological selection and the initial number of events. The signal efficiency referred as acceptance, is about 8 %. The event reduction in the irreducible backgrounds samples varies from about 0.887 % in the dibosons ZZ sample to 0.181 % in the Z+b-jet sample. The reducible backgrounds, present in the "Other BG" column, are reduced by up to six orders of magnitude.

Selection	ZH	Z+b	$t\bar{t}$	ZZ	Others BG
Before selection	491 ± 3	2109228 ± 1094	2783458 ± 279	24456 ± 11	806674256 ± 77021
trigger	478 ± 3	2070528 ± 1084	2630458 ± 271	24040 ± 11	801418304 ± 76781
Lepton type	315 ± 2	1146348 ± 796	1316863 ± 192	14037 ± 9	271253412 ± 44040
1 Signal lepton	223 ± 2	821575 ± 675	123881 ± 59	9846 ± 7	22845114 ± 11083
2 Loose leptons	207 ± 2	765068 ± 652	111919 ± 56	9161 ± 7	21476969 ± 10747
Trig matching	206 ± 2	764637 ± 652	111483 ± 56	9154 ± 7	21474627 ± 10747
$83 < m_{\ell\ell} < 99$ *GeV*	171 ± 2	621686 ± 588	11340 ± 18	7221 ± 6	16945794 ± 9540
$\not{E}_{\mathrm{T}} < 60$ *GeV*	162 ± 2	609047 ± 583	5128 ± 12	7046 ± 6	16744920 ± 9488
≥ 2 *signal jets*	126 ± 2	107800 ± 211	4395 ± 11	4547 ± 5	1046976 ± 1860
forward jet veto	116 ± 1	96306 ± 200	3698 ± 10	4197 ± 5	965080 ± 1792
exactly 2 bjets	43 ± 1	7960 ± 56	1173 ± 6	308 ± 1	5239 ± 134
ΔR *low cut*	43 ± 1	7592 ± 55	1151 ± 6	307 ± 1	4962 ± 132
$p_T^{b1} > 45$ *GeV*	39 ± 1	5009 ± 42	1072 ± 5	224 ± 1	2776 ± 84
ΔR *high cut*	38 ± 1	3811 ± 37	955 ± 5	217 ± 1	2093 ± 74
Selected events fraction (%)	8±1	0.181±0.003	0.343±0.001	0.89±0.06	$(2.59\pm0.06)\times10^{-4}$

Table 4.7.: Expected number of surviving events after each selection condition for the signal and the main backgrounds. Numbers presented are normalised to the integrated luminosity of $\mathscr{L} = 20.3$ fb^{-1}. The last row shows the fraction of selected events in percentage.

Additionally, tables 4.8 and 4.9 presents individually the number of events that survive for each of the other backgrounds. 97 % of the events that survive all the selection come from the samples Z+c-jets and Z+light jets, with 69 % (1145 events) and 28 % (583 events), respectively. In the remaining, the dibosons WZ and the channels Wt and t from single top production, with about 1 % each. The survival rate of the event selection reaches 1×10^{-7} % in W+light jets. The event selection removes all generated events from the diboson WW sample and the s-channel from single top production.

Selection	Z+jets		W+jets		
	Z+c	Z+l	W+b	W+c	W+l
Before selection	21380960 ± 10069	51358227 ± 17245	9382847 ± 2576	33669708 ± 10342	688568458 ± 73620
trigger	21135540 ± 10012	50796649 ± 17152	9289793 ± 2563	33435185 ± 10306	684497920 ± 73403
lepton type	10146903 ± 6879	23218045 ± 11475	3242931 ± 1506	13761616 ± 6520	220191593 ± 41422
1 signal lepton	6864718 ± 5660	15346278 ± 9325	35043 ± 155	60394 ± 422	490795 ± 1910
2 loose leptons	6471105 ± 5498	14509361 ± 9074	27894 ± 138	48602 ± 377	376489 ± 1668
trig matching	6470352 ± 5498	14508033 ± 9074	27865 ± 138	48567 ± 377	376338 ± 1668
$83 < m_{\ell\ell} < 99$ *GeV*	5163988 ± 4914	11744574 ± 8168	1641 ± 33	2794 ± 87	20157 ± 381
$\not{E}_{\mathrm{T}} < 60$ *GeV*	5093886 ± 4882	11618702 ± 8127	1343 ± 30	2320 ± 80	17418 ± 357
≥ 2 *signal jets*	436405 ± 1197	602097 ± 1421	379 ± 15	539 ± 36	1259 ± 81
forward jet veto	397633 ± 1146	559913 ± 1375	334 ± 15	477 ± 34	1170 ± 79
exactly 2 b-jets	3346 ± 104	1794 ± 84	12 ± 3	3 ± 2	1 ± 2
ΔR *low cut*	3156 ± 102	1710 ± 83	12 ± 3	3 ± 2	1 ± 2
$p_T^{b1} > 45$ *GeV*	1915 ± 71	782 ± 46	9 ± 2	3 ± 2	1 ± 2
ΔR *high cut*	1445 ± 62	583 ± 41	7 ± 2	3 ± 2	1 ± 2
Selected events fraction ($\times 10^{-3}$ %)	6.8±0.2	1.14±0.05	0.07±0.03	0.009±0.005	$(1\pm1)\times10^{-4}$

Table 4.8.: Expected number of surviving events after each selection condition for the V+jet backgrounds. Numbers presented are normalised to the integrated luminosity of $\mathscr{L} = 20.3$ fb^{-1}. The last row shows the fraction of selected events in percentage.

Selection	Dibosons		single top		
	WZ	WW	s-channel	t-channel	Wt-channel
Before selection	187307 ± 32	1062473 ± 47	36922 ± 15	573568 ± 208	453785 ± 102
trigger	185597 ± 32	1035955 ± 47	36424 ± 15	565992 ± 206	439248 ± 100
lepton type	59017 ± 18	251549 ± 23	15266 ± 10	244812 ± 136	121678 ± 53
1 signal lepton	13197 ± 13	19559 ± 6	235 ± 1	3684 ± 17	11210 ± 16
2 loose leptons	12258 ± 13	17870 ± 6	194 ± 1	3041 ± 15	10155 ± 15
trig matching	12252 ± 13	17864 ± 6	193 ± 1	3032 ± 15	10132 ± 15
$83 < m_{\ell\ell} < 99$ *GeV*	9471 ± 11	1900 ± 2	9 ± 0	203 ± 4	1055 ± 5
$\not{E}_{\mathrm{T}} < 60$ *GeV*	9226 ± 11	1352 ± 2	7 ± 0	150 ± 3	515 ± 3
≥ 2 *signal jets*	5829 ± 9	128 ± 1	3 ± 0	75 ± 2	262 ± 2
forward jet veto	5143 ± 8	120 ± 1	3 ± 0	55 ± 2	231 ± 2
exactly 2 b-jets	38 ± 1	1 ± 0	0 ± 0	2 ± 0	41 ± 1
ΔR *low cut*	36 ± 1	1 ± 0	0 ± 0	2 ± 0	39 ± 1
$p_T^{b1} > 45$ *GeV*	29 ± 1	0 ± 0	0 ± 0	2 ± 0	34 ± 1
ΔR *high cut*	24 ± 1	0 ± 0	0 ± 0	2 ± 0	28 ± 1
Selected events fraction ($\times 10^{-3}$ %)	13±3	0	0	0.3±0.2	6±1

Table 4.9.: Expected number of surviving events after each selection condition for the single top and diboson WW and WZ backgrounds. Numbers presented are normalised to the integrated luminosity of $\mathscr{L} = 20.3$ fb^{-1}. The last row shows the fraction of selected events in percentage.

The topological event selection to the ZH analysis was designed to reduce the number of background events, while keeping the number of expected signal events as large as possible. The ratio in equation 4.4 is often used to characterise the signal-to-background relationship, where S (B) is the number of signal (all background) events.

$$\frac{S}{\sqrt{S+B}} \tag{4.4}$$

Table 4.10 shows the expected number of events surviving each event selection, for signal (ZH) and for all simulated events (S+B), which can be compared to the surviving events from data. The disagreement observed between the number of data and simulated events is attributed to the missing multi-jet events that are not simulated and possible mismodelling of the normalization of some simulated samples. The last column shows the ratio defined in equation 4.4 that increases from 0.044 after the trigger and lepton selection, to 0.449 after all the selection criteria.

Selection	ZH	S+B	data	$S/\sqrt{S+B}$
lepton and trigger	206 ± 2	22360107 ± 10767	24714836	0.044
$83 < m_{\ell\ell} < 99$ *GeV*	171 ± 2	17586212 ± 9558	17736618	0.041
$\not{E}_{\mathrm{T}} < 60$ *GeV*	162 ± 2	17366303 ± 9506	17532357	0.039
$\geq$ *2 Signal jets*	126 ± 2	1163843 ± 1872	1198934	0.117
forward jet veto	116 ± 1	1069396 ± 1803	1100965	0.112
exactly 2 b-jets	43 ± 1	14724 ± 145	15677	0.354
ΔR *low cut*	43 ± 1	14054 ± 143	15043	0.362
$p_T^{b1} > 45$ *GeV*	39 ± 1	9119 ± 94	10093	0.408
ΔR *high cut*	38 ± 1	7115 ± 83	7903	0.449

Table 4.10.: Number of surviving events for the signal, signal+total background and data. The last column shows the improvement on the signal-to-background ratio obtained from equation 4.4. These numbers are normalised to the integrated luminosity of $\mathscr{L} = 20.3$ fb^{-1} during 2012 data taking.

4.5.4. Event categorization

The signal region is composed of events that pass all the event selection, discussed in section 4.5.3. It is referred simply by *2tag* region. A jet is considered to be b-tagged if its MV1c weight is above 0.405, which corresponds to a b-tagging efficiency of 80 %. Two additional operating points are used with higher thresholds of MV1c of 0.7028 and 0.9237, which correspond to b-tagging efficiencies of 70 % and 50 %, respectively. With this division, the ZH analysis defines three signal regions based on the value of the MV1c weight of the two b-tag jets, as schematically shown in figure 4.10. If any of the b-jets have MV1c values corresponding to efficiencies between 80 % and 70 %, the event is considered to be in the loose-loose tag region or *LLtag*. If not and any of the b-tag jets have MV1c values corresponding to efficiencies between 70 % and 50 %, the event is considered to be in the medium-medium tag region or *MMtag*. Finally, if both b-tag jets have their MV1c values corresponding to efficiencies below 50 %, the event is considered to be in the tight-tight tag region or *TTtag*. The events with only one b-tagged jet form the *1tag* control region, while events with no b-tagged jet form the *0tag* control region.

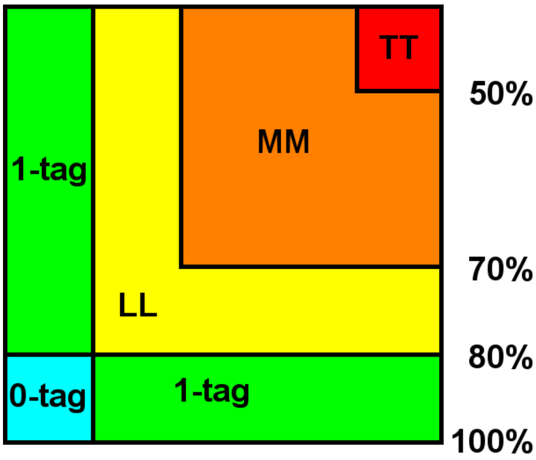


Figure 4.10.: Schematic diagram of the three b-tagging signal regions and the two control regions. The b-tagging efficiencies are shown on the right.

To reduce the large statistical uncertainty on the simulated *Z+c* and *Z+light* backgrounds after the b-tagging conditions, a parametrized b-tagging was applied to jets from events in the *0tag* region to simulate the distribution in the *2tag* region. Three possible values of the MV1c weight are considered to these jets: 0.5539, 0.81325 and 0.9618, which correspond to the midpoint between the MV1c weights attributed to the operating points referred in table 2.1. A randomly MV1c weight, chosen from the three values according to the efficiency distribution, is attributed to the *0tag* jets and an additional weight is added to the event to account for the efficiency of the parametrized b-tagging. This method allowed the reduction of the statistical uncertainty in the *2tag* region and to obtain smooth distributions of the backgrounds.

As one of the b-quarks can radiate a gluon, which can be reconstructed as a different jet, the ZH analysis also considered regions depending on the multiplicity of the jets in the event, *2jet* and *3jet*. The signal and control regions are then sub-divided into the multiplicity of jets and the number of b-tagged jets, according to the three b-tagging conditions: loose, medium and tight. These regions often appear combined. For example, the label *MMtag2jet* corresponds to the signal region with two jets, where both have their MV1c weight values between the 50 % and the 70 % b-tagging efficiencies.

The ZH analysis uses a particular control region to study the $t\bar{t}$ background. It is defined by requiring one electron and one muon in the event, with the remaining event selection conditions applied. It is called the *top* e-μ control region. $t\bar{t}$ events dominate such region, while the remaining backgrounds have a negligible contribution.

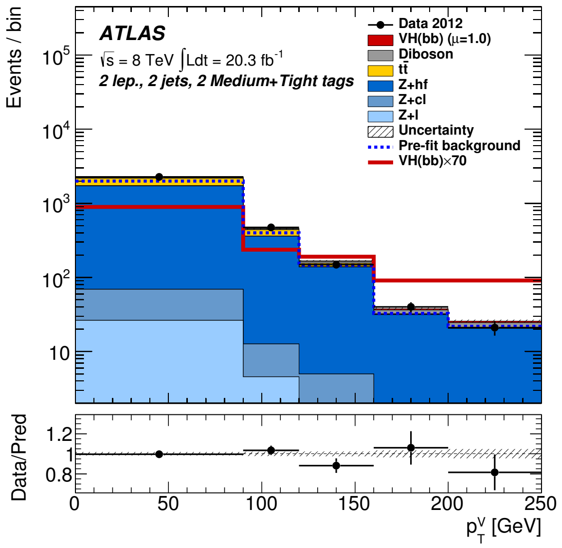


Figure 4.11.: Comparison of the distribution of the transverse momentum of the Z boson between the signal, ZH, and the main background, Z+jets, samples. The signal multiplied by 70 is also shown with the red line [82].

There are five p_T regions of the Z boson candidate because the p_T^Z spectrum of the ZH sample is very different from the dominant background (Z+jets), as observed in figure 4.11. Although the highest p_T^Z bin has the least statistics, it is where the ratio from equation 4.4 is expected to be higher. The p_T^Z bins are: $p_T^Z < 90$ GeV, $90 < p_T^Z < 120$ GeV, $120 < p_T^Z < 160$ GeV, $160 < p_T^Z < 200$ GeV and $p_T^Z > 200$ GeV.

Considering all these divisions, the VH analysis defines almost two hundred control and signal regions when considering the three channels.

4.6. ZH analysis software validation

Different ATLAS groups performed the VH analysis using different analysis software. Additionally, groups could choose either an *Athena*-based analysis and use the AOD object format or the *ROOT* framework, to analyse the D3PD format. Since it is essential that all the groups obtain the same result, the groups performed detailed comparisons of the number of events surviving each of the selection conditions on specific signal samples. This comparison was called “cut flow challenge” and was divided into three stages. First, 10 % of the events simulated of the Higgs signal sample from the 1-lepton analysis (WH) were used to validate the object classification, that is the same in all three channels. Then, 10 % of the qqZH signal sample was used to validate the event yields after each of the event topology selection of the 2-lepton analysis, discussed in the last section. Finally, one real data run was also used to validate the event yields after each condition of the ZH topology selection, corresponding to an integrated luminosity of about 5 pb^{-1}.

Although twelve groups participated in the VH analysis, only four, referred as *LIP* (my group), *A*, *B* and *C*, were involved in the 2-lepton analysis and are mentioned in this thesis. To facilitate the reading, tables for the object comparison between LIP and other groups were placed in appendix C.2. Very small differences are observed, always in the *permille* level. The second and third stages are particular for the ZH analysis. Table 4.11 is similar to that shown in section 4.5.3, with each selection applied sequentially. However, the values shown in this table are the absolute number of events generated for the signal ZH sample, *i.e.*, without the normalisation to the integrated luminosity. Columns *A* to *C* show the absolute difference with the LIP group in the number of events after each requirement. The same table shows the number of events for data, from the Muon and the Egamma streams. For these streams, group B have not filled out the numbers so that column does not appear in the table.

Selection	qqZH sample				Muon stream			Egamma stream		
	LIP	A	B	C	LIP	A	C	LIP	A	C
none	298773	-52	0	0	147283	-8	0	159596	-3	-1
Lepton type	292250	-47	0	0	147283	-8	0	159065	-4	0
1 Signal lepton	191297	-40	0	0	33894	-6	0	56634	-3	0
2 Loose leptons	133942	-18	0	0	3107	1	0	3518	0	0
Leptons veto	131290	-180	0	0	3104	-16	0	3486	0	0
Trigger	123137	-178	0	0	3047	-15	0	3343	0	0
Trig matching	122805	-142	0	0	3042	-15	0	3328	1	0
$83 < m_{\ell\ell} < 99$ *GeV*	101619	-25	0	0	2302	0	0	2087	1	0
$\not{E}_{\mathrm{T}} < 60$ *GeV*	95674	-81	-1	-1	2221	1	0	2024	1	0
$\geq$ *2 Signal jets*	72720	-43	-2	-2	143	0	0	127	0	0
Forward jet veto	67405	-38	-2	-2	136	0	0	113	0	0
Exactly 2 b-jets	25348	-17	-1	-1	0	0	0	1	0	0
ΔR *low cut*	25260	-17	-1	-1	0	0	0	1	0	0
$p_T^{b1} > 45$ *GeV*	23418	-15	-1	-1	0	0	0	1	0	0
ΔR *high cut*	22698	-16	-1	-1	0	0	0	0	0	0

Table 4.11.: Validation of the event selection for the analysis performed in this thesis (labelled LIP) comparing with three other groups that contributed for the 2-lepton analysis. The identification of the selection number was given in section 4.5. Only the LIP columns have the absolute number of events, while the remaining columns show the absolute difference to this value. Group B did not fill out the numbers for data validation and hence was removed from the table.

Group *A* shows the largest differences with respect to other groups. The difference between LIP and the B and C groups is, for the ZH sample, 4.4×10^{-3} % for the final selection (and less before it). No differences were found between this analysis and that performed by group C with the data sample.

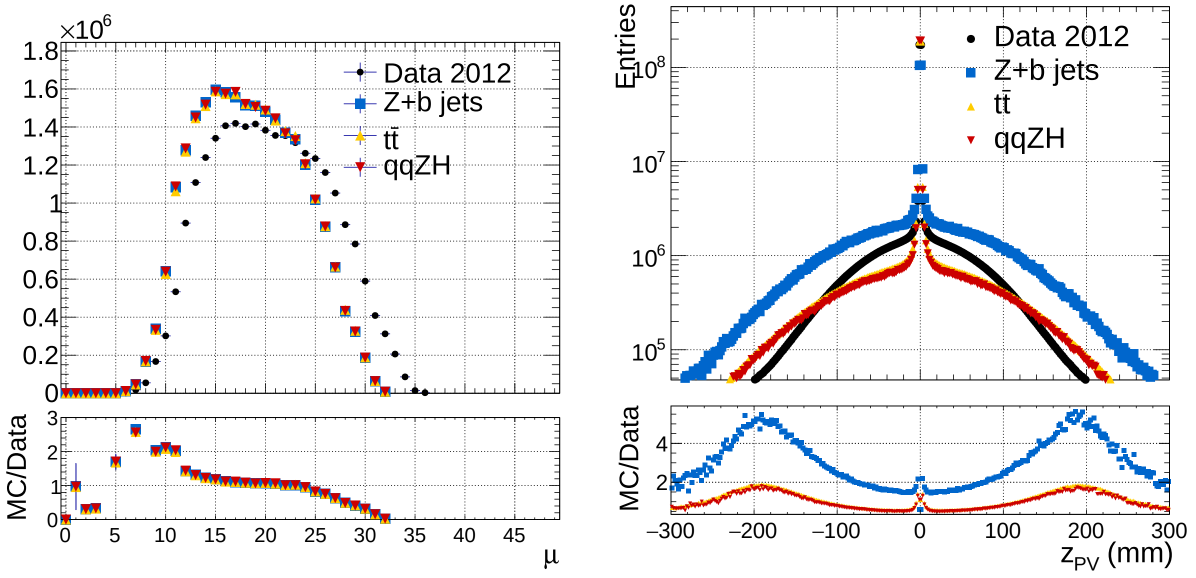
4.7. Corrections to the simulation

Simulation can only explain the reality of the detector and the different processes with a limited precision. There were corrections derived to improve the simulation, for the efficiencies (reconstruction, identification and trigger) and for the energy/momentum resolution of leptons used in the analysis (see section 2.7). These corrections are applied to the MC simulations as event weights. Additionally, there are also corrections related to the particle generation. They result in a generator related weight that is the product of various corrections, which can come either from *a priori* conditions imposed to the simulation or from mismodelling of any physical parameter. Table 4.12 presents the list of corrections applied, with a short description.

The production of the simulated samples happened before 2012 data taking started. Therefore, neither the pileup nor the z-position of the primary vertex distributions was available for the production. The best possible estimate was input to the simulations. The comparison of these two distributions between the data and the three most important simulation samples for the ZH analysis is shown in figure 4.12. The three MC distributions are normalised to the data. The ratio between the data and each of the MC distributions is shown in the lower panels in figure 4.12 (a) and 4.12 (b) for the average number of interactions per bunch crossing and for the z-position of the leading primary vertex. These

Corrections	Name	Short description
Leptons	Reconstruction Identification Trigger	Efficiency corrections for muons and electrons, described in section 2.7.1.2 and 2.7.2.1 respectively.
Global	Pileup	Corrects the pileup in MC to data
	Vertex z	Corrects the z coordinate of the primary vertex in MC to data.
Process dependent	$\mathrm{p_T^Z}$ in ZH	Corrects the distribution of the truth $\mathrm{p_T^Z}$ in the ZH sample.
	$\mathrm{p_T^Z}$ in Z+jets	Corrects the distribution of the truth $\mathrm{p_T^Z}$ in the Z+jets sample.
	$\Delta\phi$ in Z+jets	Corrects the $\Delta\phi$ distribution in the Z+jets sample.
	p_{T}^{top} in $t\bar{t}$	Corrects the distribution of the average truth $\mathrm{p_T}$ of the two top quarks in the $t\bar{t}$ sample.
b-tagging	Efficiency	Corrects for the efficiency of the b-tagging algorithm.
	MC-to-MC	MC corrections to account for differences in the b-tagging efficiency for simulations different than PYTHIA6

Table 4.12.: Summary table enumerating the corrections applied to the MC simulated samples in the ZH analysis.



(a) (b)

Figure 4.12.: Distribution of the average number of interactions per bunch crossings (a) and of the longitudinal position of the leading primary vertex (b) for signal (red down triangles), Z+b-jets (blue squares) and $t\bar{t}$ (yellow down triangles) MC samples compared to the data (black circles), evaluated before the pileup and z vertex corrections.

ratios are then used to derive the corrections, which are applied to each of the simulated data samples.

Since the main signal sample was generated at *leading-order* with PYTHIA8, corrections have to be applied to take into account higher order terms in the cross section calculation. The distribution of the $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ at truth level in the signal ZH sample is calculated by the HAWK Monte Carlo program [145], with *next-to-leading-order* electroweak (EW) corrections. Figure 4.13 shows the correction to the *leading-order*, as a function of the $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ for $ZH \rightarrow \nu\nu b\bar{b}$, in green, and for $ZH \rightarrow \ell\ell b\bar{b}$ (red and black). Two alternative methods to handle photons collinear with the charged leptons are shown for the ZH sample: either assuming perfect isolation (black, bare), which is applied only to the Z$\rightarrow \mu\mu$ case, or performing a recombination of the photon with the lepton, mimicking the approach of electrons in electromagnetic showers (red, rec) and is applied to the Z$\rightarrow ee$ case. The correction of the number of signal simulated events ranges between 5 and 15 % for Z boson with p_T^Z up to 300 GeV.

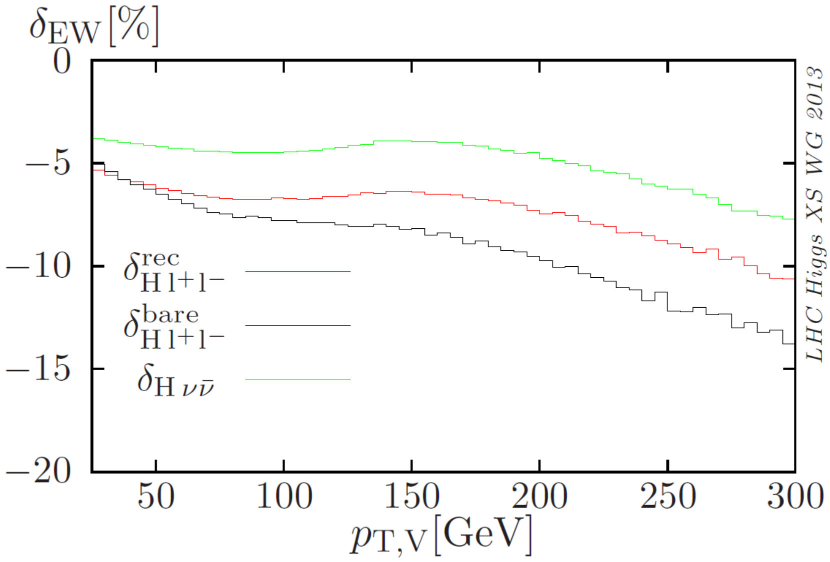
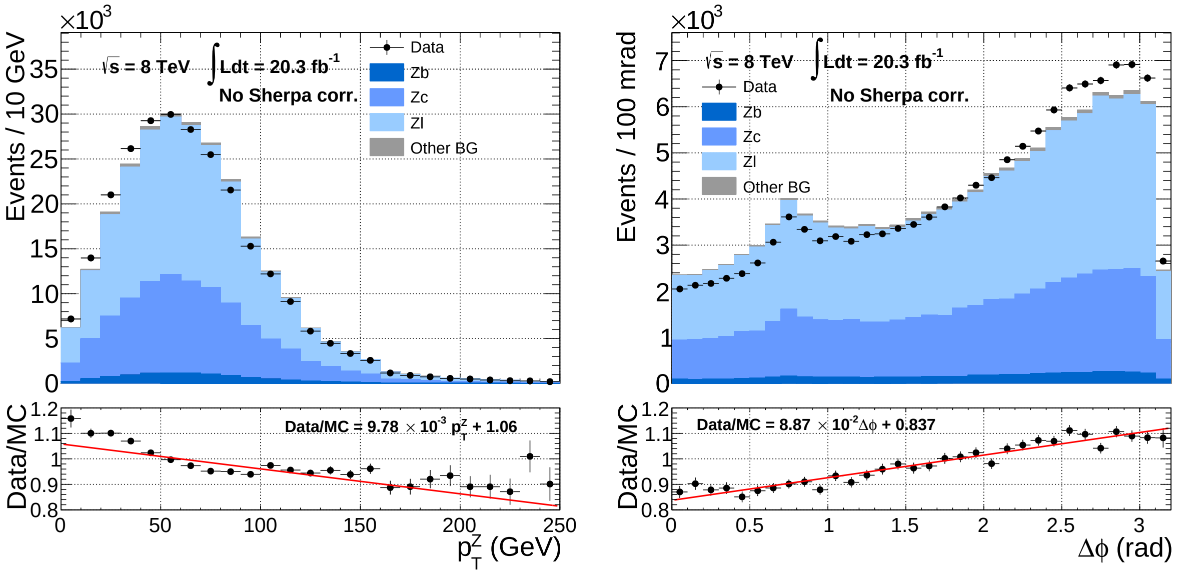


Figure 4.13.: Relative electroweak NLO corrections for the p_T^Z distribution in $ZH \to \nu\nu b\bar{b}$ (green) and $ZH \to \ell\ell b\bar{b}$ (red and black) [21]. The "bare" approach treats leptons as perfectly isolated and the "rec" includes a recombination between the lepton and the photon.

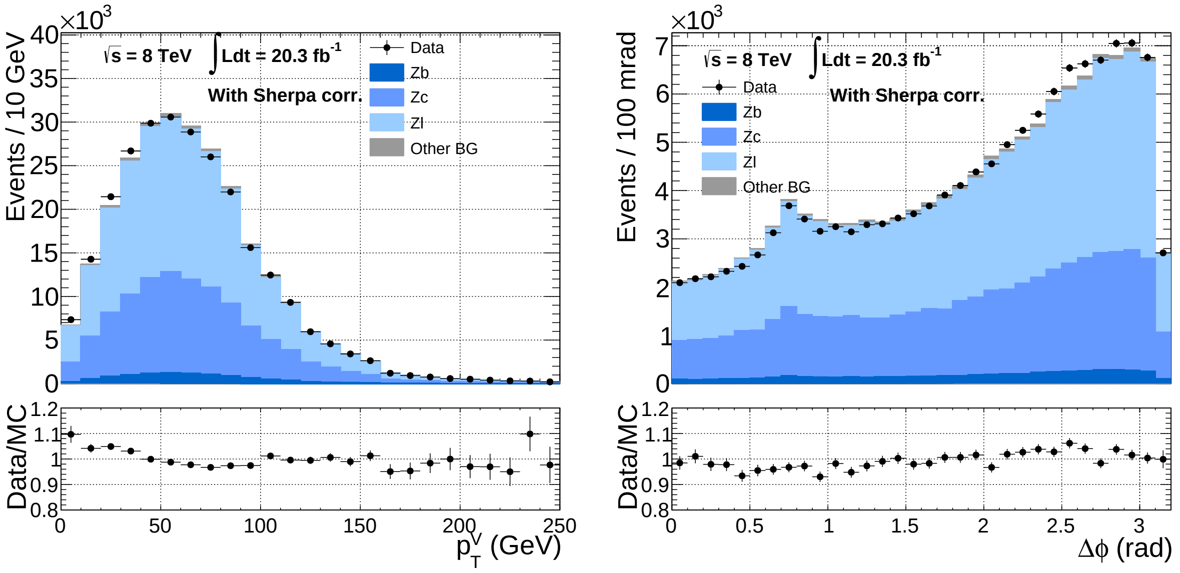
The SHERPA simulation of V+jets was found to have a mismodelling in the distributions of the transverse momenta of the vector boson and of the azimuth distance between the two jets [82], as observed in figures 4.14 (a) and 4.14 (b), respectively. These distributions use the 0-tag control region, where the Z+jet backgrounds dominate and the remaining backgrounds are negligible. Three different shades of blue represent the three different filters applied to the Z+jet samples, corresponding the darkest shade to the heaviest quark. The discrepancy between data and MC is visible. The correction, taken from a linear fit applied to the ratio between the data and simulation distributions, appears at the bottom. Applying the corresponding weight to the simulation distributions lead to a better data-MC agreement as it is observed in figure 4.14 (bottom).

The $t\bar{t}$ POWHEG+PYTHIA6 sample predicts a harder top-quark p_T spectrum than data [146]. The spectrum of the average p_T of the two top quarks is compared between data and MC, the later using the top quark truth p_T value. The data-MC ratio distributions provide a weight to be applied to this sample.

For each b-tagged jet in the event, a p_T and η dependent b-tagging efficiency correction is applied. The objective of this correction is to match the efficiencies in simulation to those



(a) p_T^Z before correction (b) $\Delta\phi$ before correction



(c) p_T^Z after correction (d) $\Delta\phi$ after correction

Figure 4.14.: Comparison between the distributions of p_T^Z (left) and $\Delta\phi$ (right) of data and the simulation before (top) and after (bottom) applying the SHERPA corrections. Data is shown with black markers and the three filters of the Z+jets background sample is shown in three shades of blue.

in data (see section 2.7.6). Additionally, since the b-tagging efficiencies were obtained with a POWHEG+PYTHIA6 sample of $t\bar{t}$ and such efficiencies are dependent on the parton shower and hadronization models, two other samples are tested, SHERPA and PYTHIA8, to evaluate sample dependent b-tagging efficiencies. A specific MC-to-MC additional correction is then applied to samples that model parton shower and hadronization different from PYTHIA6 as another weight which is applied to each of the jets, depending on the truth flavour of the jet, defined in section 2.7.5.3.

4.8. Improvements to $m_{b\bar{b}}$ resolution

Table 4.10 showed that the number of signal events is diminutive when compared to the total number of background events, even after applying all the topological selection. It is, therefore, crucial to obtain the best possible resolution in the invariant mass of the $b\bar{b}$ pair to improve the sensitivity to the Higgs boson signal.

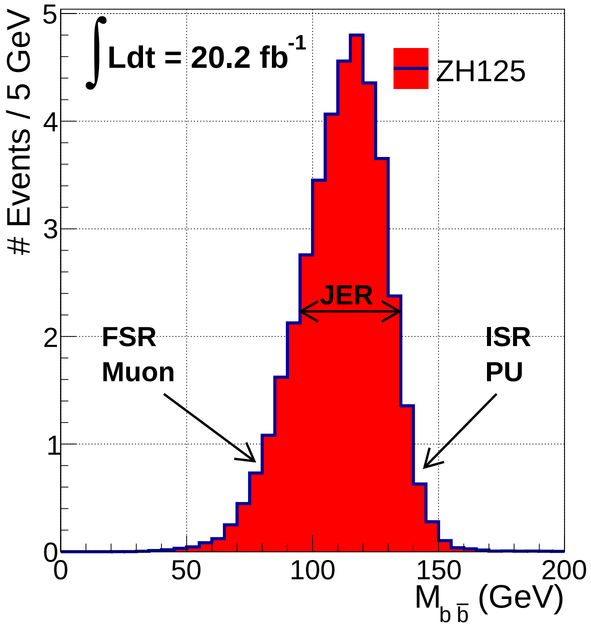


Figure 4.15.: Distribution of the invariant mass, $m_{b\bar{b}}$, for the qqZH signal sample. The arrows indicate effects that contribute to the limited resolution.

Figure 4.15 shows the invariant mass distribution for the b-jet pair, $m_{b\bar{b}}$, for the qqZH signal sample, after the full event selection is applied. Some effects contribute to the limited resolution. The high-mass tail is due mainly to the initial state radiation (ISR). Energy from pileup may also enter the jet radius and contribute to this tail. The final state radiation (FSR) is one of the effects responsible for the low-mass tail. Another effect results from the B-hadrons that decay leptonically to a muon and a neutrino that are not measured in the calorimeter. The jet energy resolution (JER) is the main responsible for the width of the distribution.

The following procedures were applied to improve the invariant mass $m_{b\bar{b}}$:

- JER is reduced with the global sequential calibration (GSC).
- A special correction is applied to correct for the muon in the case of leptonic B-meson decays reducing the effect observed in the low-mass tail.
- An event kinematic fit is used to improve even further the jet energy resolution, since all objects are well reconstructed in the event.

4.8.1. Global sequential calibration

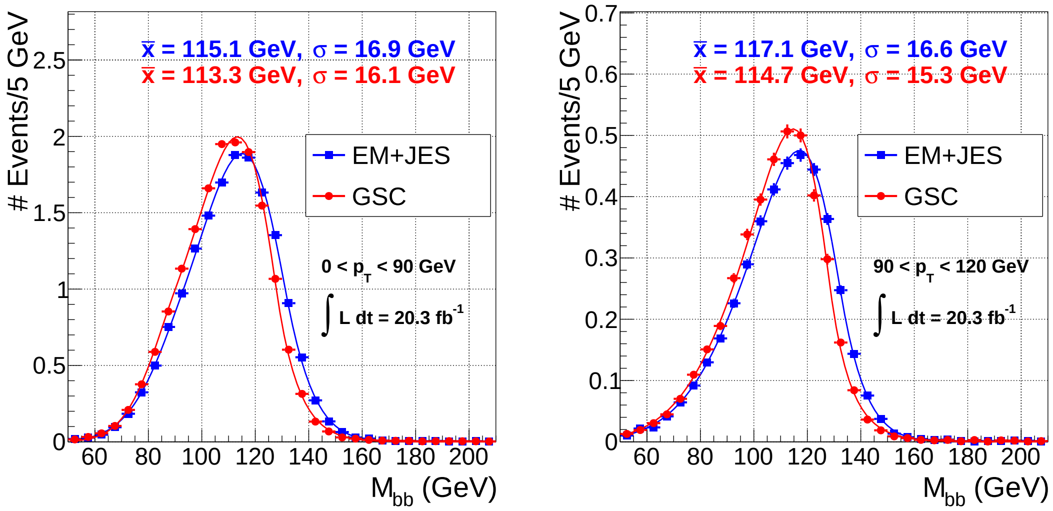
The sensitivity to the Higgs signal is directly related to how the invariant mass peak of the b-jet pair, $m_{b\bar{b}}$, is resolved. To improve the mass resolution, the global sequential calibration, GSC, studied in section 3.9, was applied in the ZH analysis.

Figure 4.16 shows the $m_{b\bar{b}}$ distribution obtained with the EM+JES calibration (blue squares) and with GSC (red circles) for each of the $\mathrm{p_T^Z}$ bins considered in the ZH analysis. This distribution is not symmetrical due to ISR and FSR effects so a Bukin function [136] was used to fit them. Applying GSC, the width of the Gaussian core of the Bukin function becomes narrower, as shown in table 4.13, that summarises the fit parameters obtained for the Bukin mean and width. The improvement is again obtained from equation 3.14. The last column of table 4.13 shows these improvements in percentage. The resolution enhancement is larger for high transverse momenta of the Z boson, up to 18.5 % in the most sensitive $\mathrm{p_T^Z}$ bin and about 1 %, for the five bins combined.

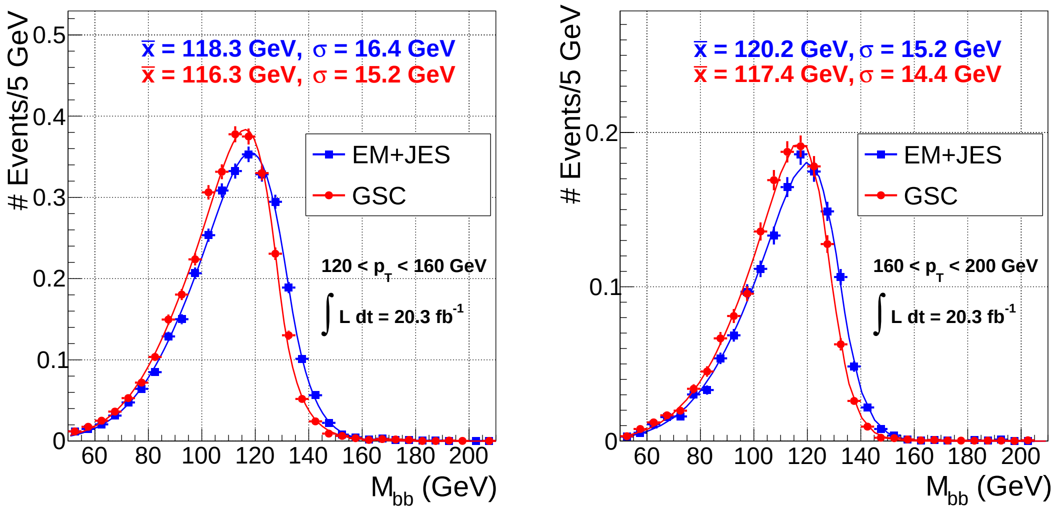
The mean value of the invariant mass distribution with GSC becomes smaller, since the jet energy scale was derived using truth jets, which contain neither muons nor neutrinos, as discussed in section 3.4.

4.8.2. Correction for muon inside the jet

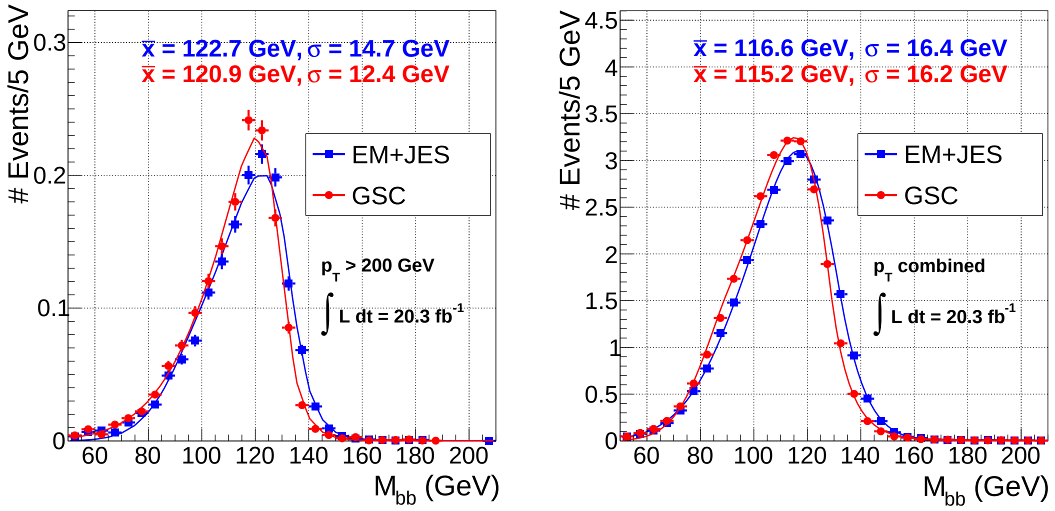
B hadrons can decay leptonically, emitting a muon and a neutrino which is not reconstructed in the calorimeter. Such jets are reconstructed with $\mathrm{p_T}$ lower than it should



(a) $p_{\mathrm{T}}^{Z} < 90$ GeV (b) $90 < p_{\mathrm{T}}^{Z} < 120$ GeV

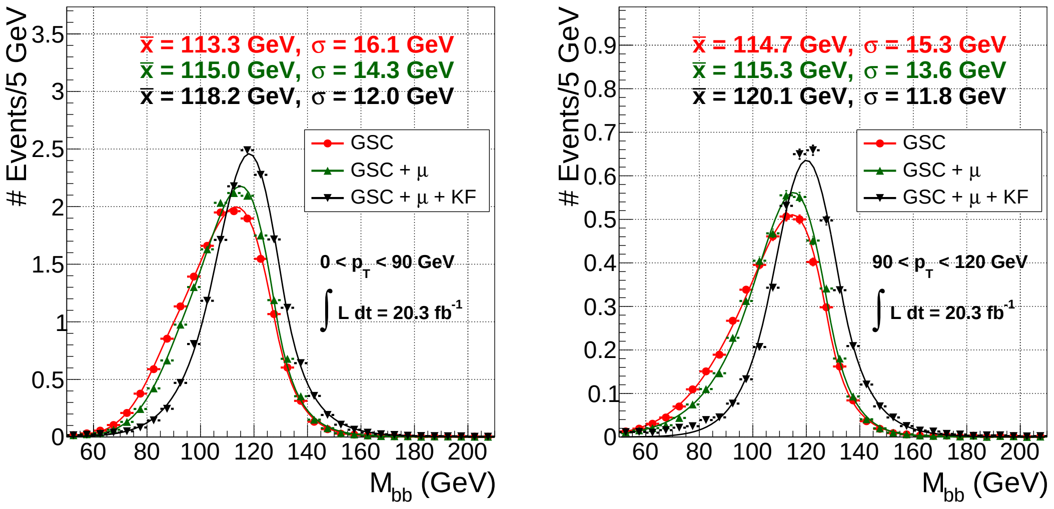


(c) $120 < p_{\mathrm{T}}^{Z} < 160$ GeV (d) $160 < p_{\mathrm{T}}^{Z} < 200$ GeV

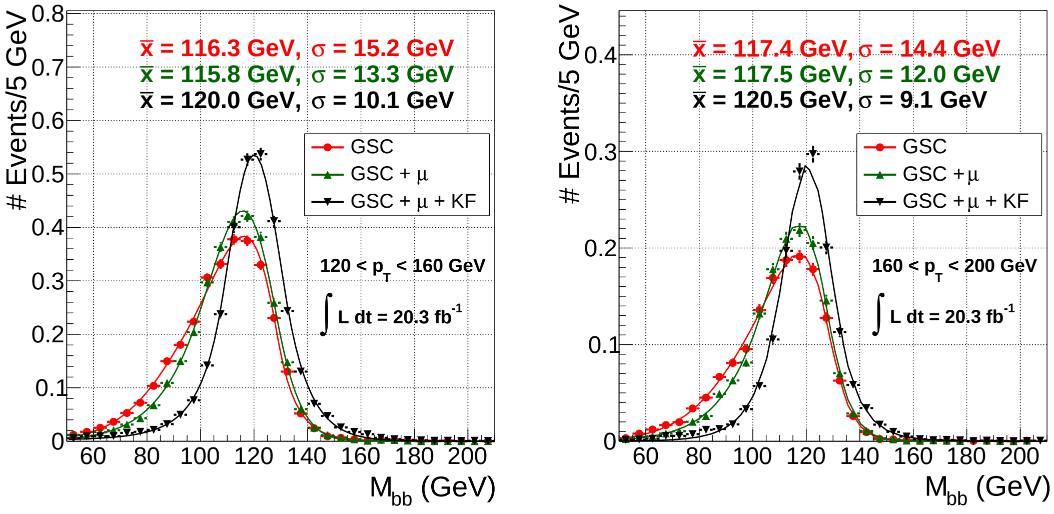


(e) $p_{\mathrm{T}}^{Z} > 200$ GeV (f) $\mathrm{p}_{\mathrm{T}}^{Z}$ combined

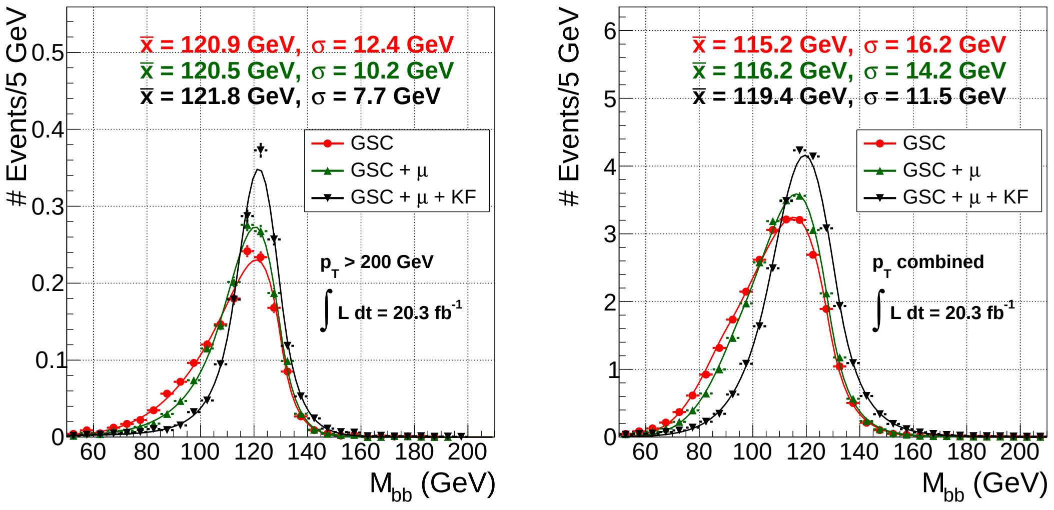
Figure 4.16.: Invariant mass of the b-jet pair in the signal sample qqZH after the ZH event selection, for five $\mathrm{p}_{\mathrm{T}}^{Z}$ regions (a-e) and combining all regions (f). The default jet calibration, EM+JES, is shown in blue circles and GSC in red squares.



(a) $p_{\mathrm{T}}^{Z} < 90$ GeV (b) $90 < p_{\mathrm{T}}^{Z} < 120$ GeV



(c) $120 < p_{\mathrm{T}}^{Z} < 160$ GeV (d) $160 < p_{\mathrm{T}}^{Z} < 200$ GeV



(e) $p_{\mathrm{T}}^{Z} > 200$ GeV (f) $\mathrm{p}_{\mathrm{T}}^{Z}$ combined

Figure 4.17.: Invariant mass of the b-jet pair, in the signal sample qqZH, after the ZH event selection for the five $\mathrm{p}_{\mathrm{T}}^{Z}$ regions (a-e) and combining all regions (f). GSC is shown with red circles, the result of applying the *μ-in-jet* correction is shown with green up triangles and the kinematic fit is shown with the black down triangles.

	EM+JES		GSC		Improvement in
p_T bin	mean	width	mean	width	the resolution (%)
$p_T^Z < 90$ GeV	115.1 ± 0.2	16.9 ± 0.1	113.3 ± 0.2	16.1 ± 0.1	4.97 ± 0.04
$90 < p_T^Z < 120$ GeV	117.1 ± 0.4	16.6 ± 0.2	114.7 ± 0.3	15.3 ± 0.2	8.5 ± 0.2
$120 < p_T^Z < 160$ GeV	118.3 ± 0.4	16.4 ± 0.2	116.3 ± 0.3	15.2 ± 0.2	7.9 ± 0.1
$160 < p_T^Z < 200$ GeV	120.2 ± 0.6	15.2 ± 0.3	117.4 ± 0.5	14.4 ± 0.3	5.6 ± 0.2
$p_T^Z > 200$ GeV	122.7 ± 0.4	14.7 ± 0.3	120.9 ± 0.3	12.4 ± 0.2	18.5 ± 0.5
Combined	116.6 ± 0.2	16.4 ± 0.1	115.2 ± 0.1	16.2 ± 0.1	1.23 ± 0.01

Table 4.13.: Result from applying a Bukin fit to the invariant mass distributions, before and after implementing the GSC calibration. The mean and width values correspond to Gaussian function for the core of the distribution and are presented in GeV. The improvement is calculated from equation 3.14.

be since part of their energy is taken away by the muon. A muon is searched for, matching the jet within a distance $\Delta R < 0.4$ and with $p_T > 4$ GeV. The energy of the muon is added to the jet energy. The equivalent energy deposit in the calorimeter of a minimum ionising particle is removed from the total energy, to avoid double counting. This method is known as *μ-in-jet* correction. Figure 4.17 shows the effect of this correction: the low-mass tail of the distribution is reduced. In each of the p_T^Z bins, the distribution obtained with just GSC is shown once more for comparison. A Bukin function is also adjusted to the $m_{b\bar{b}}$ distribution and table 4.14 shows the Bukin fit parameters, mean and width, after applying the *μ-in-jet* correction on the left, with the improvement exhibited with GSC (see table 4.13) shown in the penultimate column. The mean value of the Bukin fit is closer to the truth value of the Higgs boson mass and the resolution improvement with respect to that obtained with GSC is 14 % inclusive in p_T^Z and up to 21 %, if considering the $p_T^Z > 200$ GeV region.

4.8.3. Kinematic fit

One of the biggest advantages of the ZH analysis with respect to the 0-lepton and 1-lepton analysis is the possibility to reconstruct all the particles that form the ZH event: two muons or two electrons, and two b-jets. These four objects should be balanced in the transverse plane and the missing transverse energy measured should result mainly from miscalibrations of the b-jets. A kinematic likelihood fit (KF) uses the good resolution of

	w/ *μ-in-jet*		w/ kinematic fit		Improvement w.r.t. GSC (%)	
p_T bin	mean	width	mean	width		
$p_T^Z < 90$ GeV	115.0 ± 0.1	14.3 ± 0.1	118.2 ± 0.1	12.0 ± 0.1	12.6 ± 0.1	34.2 ± 0.4
$90 < p_T^Z < 120$ GeV	115.3 ± 0.3	13.6 ± 0.2	120.1 ± 0.2	11.8 ± 0.1	12.5 ± 0.2	29.7 ± 0.5
$120 < p_T^Z < 160$ GeV	115.8 ± 0.3	13.3 ± 0.1	120.0 ± 0.2	10.1 ± 0.1	14.3 ± 0.2	50.5 ± 0.8
$160 < p_T^Z < 200$ GeV	117.5 ± 0.5	12.0 ± 0.2	120.5 ± 0.3	9.1 ± 0.1	20.0 ± 0.5	58 ± 1
$p_T^Z > 200$ GeV	120.5 ± 0.3	10.2 ± 0.2	121.8 ± 0.2	7.7 ± 0.1	21.6 ± 0.5	61 ± 1
Combined	116.2 ± 0.1	14.2 ± 0.1	119.4 ± 0.1	11.5 ± 0.0	14.1 ± 0.1	40.9 ± 0.3

Table 4.14.: Improvement observed in the invariant mass resolution, resulting from applying the Bukin fit to the invariant mass distributions obtained with the *μ-in-jet* correction and the kinematic fit on top of GSC. The mean and width values are in GeV. The improvement is calculated from equation 3.14.

muons and electrons and the knowledge of the event p_T balance to apply a correction to the b-jets energy.

The inputs for the likelihood are the energy/momentum, pseudorapidity and azimuth angle of the four particles. The fit uses as constraints the invariant mass of the two leptons with a Breit-Wigner distribution, considering the mass and width of the Z boson, the transverse components of the momentum (p_x and p_y) of the $\ell\ell b\bar{b}$ system to be 0 ± 9 GeV. The energy/momentum of the leptons are constraint to with a Gaussian function with the uncertainty on the scale as the standard deviation.

Two constraints are applied to the energy of jets. Using the qqZH sample with the event selection up to the “Exactly 2 b-jets” requirement, reconstructed jets are compared to anti-k_T truth jets including muons and neutrinos. Truth and reconstructed jets are matched using the same criterium as described in section 3.4.1, without isolation. The constraint is obtained from transfer functions such as the one in figure 4.18. Additionally, the kinematic fit uses the p_T spectrum obtained for truth b-jets (excluding muons and neutrinos) from the qqZH sample after the full ZH event selection as another prior.

The result from applying both the *μ-in-jet* correction and the kinematic fit is shown in figure 4.17 with the black markers. Table 4.14 also shows the Bukin mean and width after applying the kinematic fit.

The improvement of applying both the *μ-in-jet* and the kinematic fit to the invariant mass $m_{b\bar{b}}$ distribution with respect to just applying GSC is calculated from equation 3.14 and

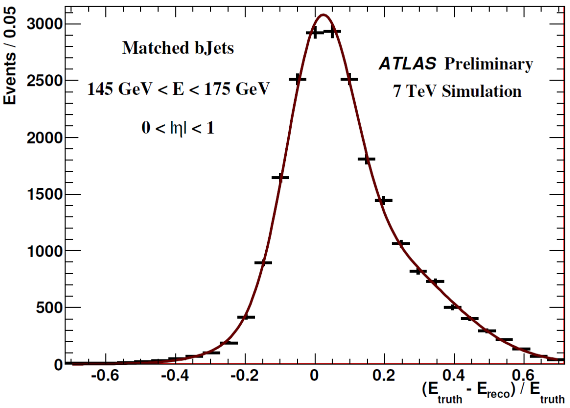


Figure 4.18.: Transfer function used as input for the kinematic fit [147].

shown in the last column. As the kinematic fit corrects also the neutrino from semi-leptonic decays, the mean value becomes closer to the simulated value for the Higgs boson mass (125 GeV). The improvement including the kinematic fit with respect to the GSC was about 41 % when combining all $\mathrm{p_T^Z}$ bins, and reaching 61 % when considering events in which the Z boson had a $\mathrm{p_T^Z}$ larger than 200 GeV.

Considering the total improvement resulting from the three methods with respect to the nominal jet calibration, EM+JES, it is 42.6 % for combining all $\mathrm{p_T^Z}$ regions and 90.9 % just for the $p_{\mathrm{T}}^{\mathrm{Z}} > 200$ GeV region.

4.9. Signal extraction and background normalization

The signal-strength parameter, μ, is the parameter of interest of the ZH analysis and compares the theoretical to the observed value of the product between the cross section production of the ZH process and the branching ratio of $H \to b\bar{b}$, as discussed in section 1.6. The cross section production of the ZH backgrounds are several orders of magnitude higher than the ZH signal, as it was shown in figure 1.9. Therefore, it is essential to have the best possible knowledge of these backgrounds.

A statistical fitting procedure, known as global fit, uses a binned profile likelihood function, which is constructed as the product of Poisson-probability terms [82]. The input of this method is the invariant mass distribution of the two b-jets in events that survive the topological selection and those selected by the *top* e-μ criteria, and the distribution of the

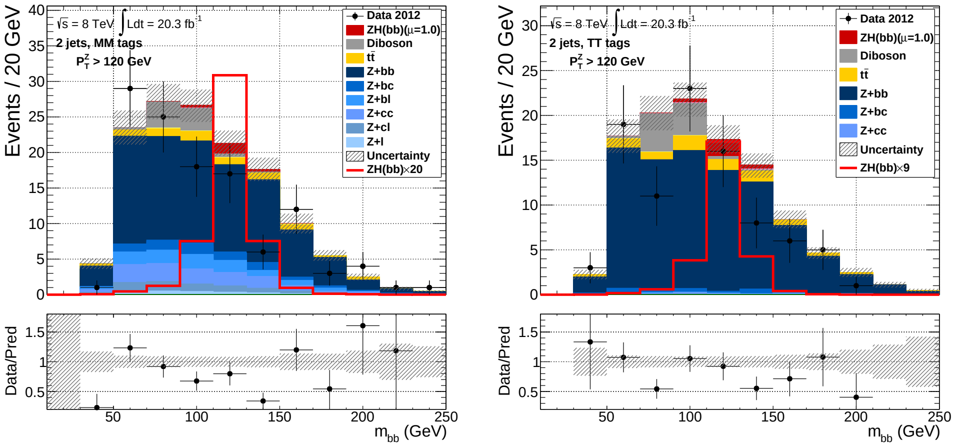
MV1c values of the leading jet in events from the *1tag* region.

The simulated signal and background distributions are stacked together and compared to the data. In each figure, the bottom panel shows the ratio between the number of events observed in data and the number of expected signal ($\mu = 1$) and background events summed together. Using the truth flavour, discussed in section 2.8, of the each jet that form the Higgs boson candidate, the Z+jets background is separated in six subsamples: Z+bb, Z+bc, Z+bl, Z+cc, Z+cl, and Z+l. The first four subsamples are sometimes combined into Z+hf (heavy flavour). The quadratic sum of the experimental systematic uncertainties, that will be discussed in section 4.10.1, is represented by the grey band. The uncertainty shown on the bullet points of the ratio corresponds to the statistical uncertainty of the data.

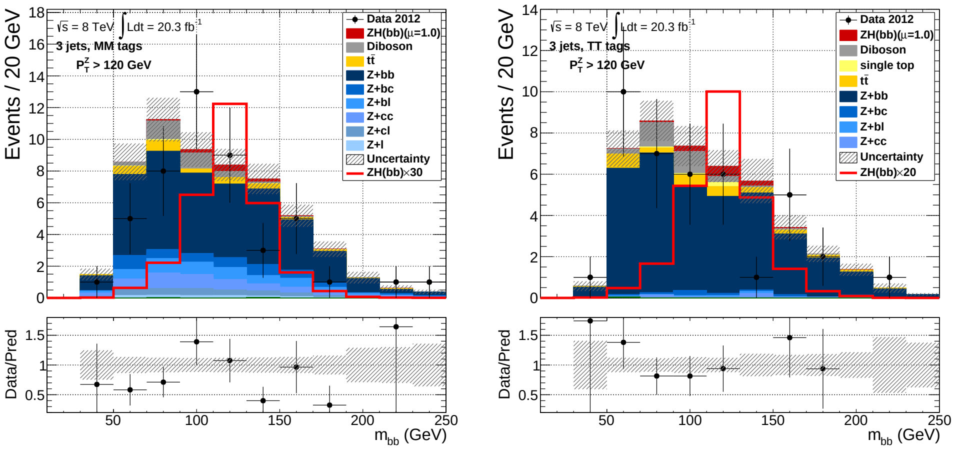
The distributions of the invariant mass $m_{b\bar{b}}$ were re-binned in intervals of 20 GeV to accommodate the expected maximum number of signal events in one bin: $110 < m_{b\bar{b}} < 130$ GeV. The shape of the signal distribution is also shown in these plots, with a red line with different normalisations for visibility purposes, to allow comparisons with the background. For events with three jets the $m_{b\bar{b}}$ distribution considers the invariant mass of the two b-jets.

Figure 4.19 shows the invariant mass $m_{b\bar{b}}$ distribution for four of the regions, where the signal-to-background ratio should be larger: the MMTag and TTtag and combining the three most significant intervals of $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$: $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}} > 120$ GeV. The contribution from the Z+jet background in the *TT* region, other than Z+bb is negligible for both two (a) and three (b) jet multiplicities. There are some differences between the distributions obtained with data and those from the simulation but there is still very large statistical uncertainties. Just these four histograms are expected, from the Standard Model ($\mu = 1$), to contain only seven events originated from the ZH process.

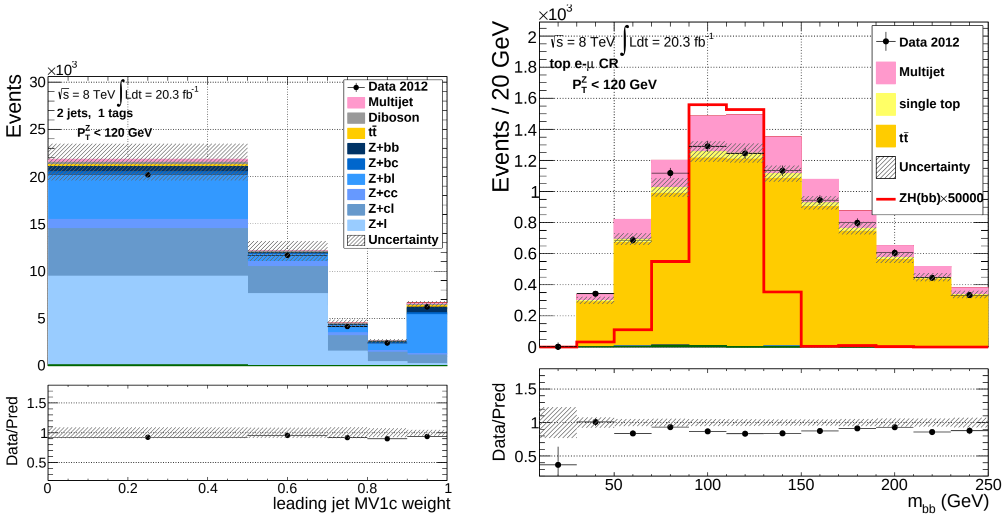
Another distribution that helps to set the normalisation of the Z+jets background is the MV1c weight of the b-jet in events from the *1-tag* control region. Figure 4.19 (e) shows this variable distributions for events with a reconstructed Z boson with $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}} < 120$ GeV. The *top* e-μ region is shown in figure 4.19 (f) for the same $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ region. The major contribution in this region is the $t\bar{t}$ background, as expected. There is also a large multi-jet background in this plot. However, the data seems to have the same number of entries as the $t\bar{t}$ sample, which might suggest some problem in the estimation of the multi-jet background in this region, which was estimated without correlation with other regions.



(a) MM tag 2 jet region (b) TT tag 2 jet region



(c) MM tag 3 jet region (d) TT tag 3 jet region



(e) MV1c in 1tag region (f) *top* e-μ region

Figure 4.19.: Invariant mass of the two b-tagged jets for four signal regions. The top (centre) histograms correspond to events with two (three) jets, combining those with a reconstructed Z boson with $p_T^Z > 120$ GeV for the Medium (left) and Tight (right) b-tag region. Same distribution in the *top* e-μ control region for $p_T^Z < 120$ GeV (bottom-right) and MV1c weight of the leading b-tagged jet in the *1tag* region (bottom-left). The multi-jet background is not normalized and only its shape should be considered.

4.9.1. The statistical model

The distributions of the di-b-jet invariant mass and the MV1c of the leading b-jet are fed to a complex binned profiled likelihood fit [82, 148]. Equation 4.5 shows the expression of the likelihood, which is composed of two terms.

$$L = \prod_r \prod_b \text{Pois}\left(n_{br} | \nu_{br}\right) \cdot \prod_p f_p(a_p | \alpha_p) \tag{4.5}$$

The first term is a product of Poisson functions for each bin of the signal and control regions, which were defined in section 4.5.4. It gives the probability of finding n_{br} events of data on a given bin b of a region r when a particular number is expected from simulation, ν_{br}. This per-bin prediction is given as:

$$\nu_{br} = \mu S_{br}(\alpha_p) + B_{br}(\alpha_p) \tag{4.6}$$

The signal-strength parameter (μ), introduced in section 1.6, multiplies the number of events predicted for the ZH signal for that bin and region, S_{br}, and B_{br} is the number of background events for the same bin and region. The number of expected signal and background events in each bin depends on some nuisance parameters p, associated to the systematic uncertainties that will be described in section 4.10.

The second term of the likelihood is used to constrain these nuisance parameters, with a penalty term that will decrease the likelihood if any of the parameters deviates from its nominal value. It uses a Gaussian centred at a_p, the average value with the respective uncertainty (σ_p), obtained by the performance groups in the ATLAS collaboration and from priors obtained in the modeling of the simulated signal and backgrounds, and α_p represents the shift from the nominal value: $Gaus(\alpha_p; a_p, \sigma_p)$. However, some nuisance parameters are normalisations of the diverse samples of the analysis so a negative α_p is not physical. For this reason, a log-normal probability distribution function is used instead, which corresponds to a change of variable:

$$\alpha_p \rightarrow ln\alpha_p \tag{4.7}$$

The normalization of the $t\bar{t}$, Z+bb and Z+cl processes in two jet events enter with Poisson terms in the likelihood and are allowed to float. When combined with the other lepton channels of the VH search, the normalisation W+bb and W+cl are also unconstrained, and the $t\bar{t}$ normalisation is obtained separately for the three lepton channels. In this case, the parameter of interest, μ, can be obtained for each channel separately, combining 0 and 2 leptons or combining the three channels.

The di-b-jet invariant mass distribution is re-binned for the fit, to reduce bin-by-bin statistical fluctuations and uncertainties associated to them. In addition to the improvement on the statistical fluctuation from simulation, this rebinning concentrates the bins where the signal-to-background ratio is larger in the middle of the histogram. The bins of the $m_{b\bar{b}}$ distributions are performed according to the Z-function defined in equation 4.8.

$$Z = z_S f_S + z_B f_B \tag{4.8}$$

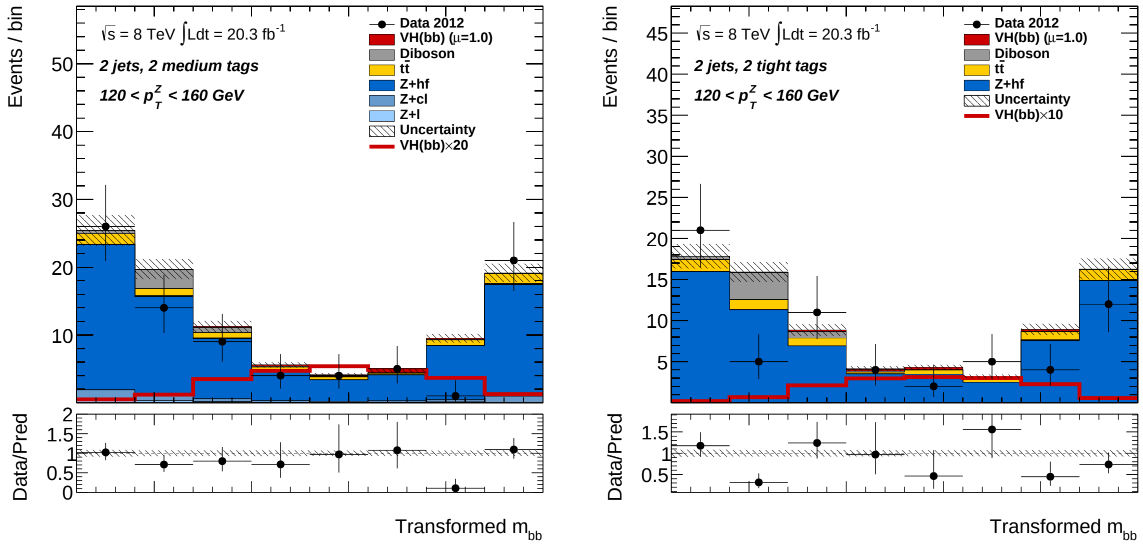
where f_S (f_B) is the fraction of the number signal (background) events in a certain interval and z_S and z_B are parameters of the rebinning. Considering $m_{b\bar{b}}$ intervals finishing on the last bin, the Z-function is calculated after adding each bin. When $Z > 1$, the interval is considered as one of the transformed $m_{b\bar{b}}$ bins. The following bins then finished on the bin before the beggining of the just selected interval until all bins are considered. The parameters were optimize for this analysis with values of $z_S = z_B = 4$ for 2 jets and $z_S = z_B = 2$ for 3 jets multiplicity events. The transformed $m_{b\bar{b}}$ distribution is shown in figure 4.20 for two regions. Due to technical requirements, histograms entering the fit are required to have the same bin size as is also shown in this figure and, for that reason, there are no scale in the x-axis.

4.10. Systematic uncertainties in the ZH analysis

The systematic uncertainty sources for the ZH analysis can be divided into experimental and theoretical. The former are the uncertainties associated with the detector and the reconstruction of the physics objects. The latter are related to the limited theoretical knowledge of the signal and background processes and their simulations. The following sub-sections present the list of systematic uncertainty sources considered in the ZH analysis. To evaluate the total systematic uncertainty, all the systematic uncertainties are summed quadratically.

4.10.1. Experimental systematic uncertainties

The experimental systematic uncertainties result from the uncertainty that exists on the energy scale and resolution of each physics objects used in the ZH analysis. To evaluate



(a) MM 2 jets (b) TT 2 jets

Figure 4.20.: Example of two $m_{b\bar{b}}$ distributions entering the global fit after passing the re-bin transformation for the MMtag2jet (a) and TTtag2jet (b) regions with p_T^Z between 120 and 160 GeV. The red line indicates the number of signal events multiplied by 20 in (a) and 10 in (b).

the effect on the analysis of each source of experimental systematic uncertainty, the affected quantity is shifted up or down according to the uncertainty parameter and the full analysis is redone. The variations in the $m_{b\bar{b}}$ distribution are registered and are used in the global fit, discussed in section 4.9.1, to evaluate the systematic uncertainties on the expected signal strength.

Figure 4.21 shows the relative experimental systematic uncertainties as a function of the invariant mass distribution of the two b-tagged jets. It considers events with 2 or 3 jets, in which the two leading jets are b-tagged and inclusive in the transverse momentum of the Z boson. The different lines represent the relative uncertainty estimated from the different physics objects: missing transverse energy (soft term) in blue, muons in red, electrons in green, jets in pink and the b-tagging in brown, using different dashed lines. The details of the uncertainty on each physics objects will be explained next. Additionally, the light green line shows the uncertainty in the integrated luminosity from the 2012-Run at a centre of mass of $\sqrt{s} = 8$ TeV, which had a value of 20.2 fb^{-1}, with an uncertainty of 2.8 % [82]. The black solid line represents the total experimental systematic uncertainty. It varies between 6.5 %

and 14.5 % for invariant masses $m_{b\bar{b}}$ in the range $10-250$ GeV.

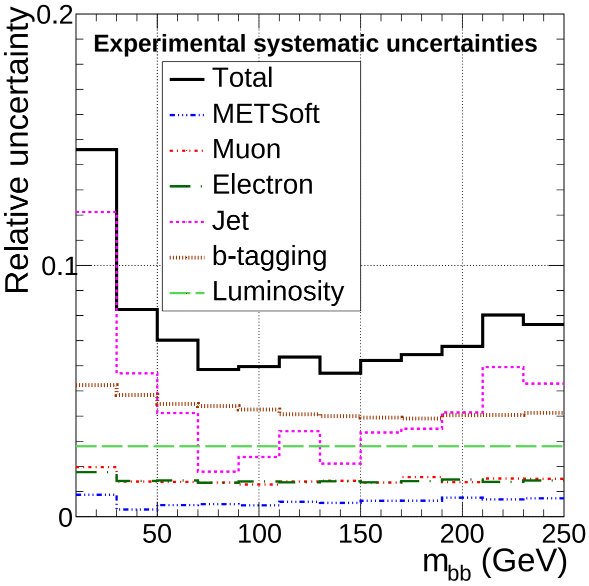


Figure 4.21.: Experimental systematic uncertainty on the invariant mass distribution obtained separately for different physics objects: $\not{E}_{\mathrm{T}}$ soft (blue), muon (red), electron (dark green), jet (pink), b-tagging (brown) and luminosity (light green). Histogram obtained from the quadratic sum of the deviations from each experimental uncertainty to the nominal sum of the signal and background simulated samples.

Uncertainties on the reconstruction, identification and trigger efficiencies, obtained from data-MC comparison, are taken into account for muons and electrons. Uncertainties on the momentum/energy scale of these leptons are also considered. These uncertainties have typically a small effect in the analysis, as shown in figure 4.21 (about 1.5 % for both electron and muon).

The propagation of the other systematic sources, such as energy calibration of the physics objects, to $\not{E}_{\mathrm{T}}$ is obtained by recalculating $\not{E}_{\mathrm{T}}$ for every systematic variation. There is also a systematic uncertainty on the energy scale and resolution of the soft terms of the missing transverse energy, defined in section 2.7.4. However, it has a negligible impact on this analysis, as shown in figure 4.21 with a variation on the nominal invariant mass distribution smaller than 1 %.

There are several sources of systematic uncertainties on the jet energy scale. Since each one changes the p_{T} of the jets, the number of jets that pass the 20 GeV transverse momentum requirement also changes. The ZH analysis considers the uncertainties associated to the *in situ* calibration (6 sources after the decorrelation described in section 3.5.4), the energy response difference between light and gluon jets (described in section 3.7.1), the η-intercalibration (described in section 3.6), the pileup (described in section 3.2) and the additional uncertainties derived for the global sequential calibration (see section 3.9). Each of the referred systematic uncertainties was propagated to the ZH analysis. The effect of all these systematic uncertainties in the invariant mass distribution ranges from 2 to 6 %, depending on the bin. In some category regions, this is the leading experimental systematic uncertainty.

The corrections on the b-tagging efficiency (see section 4.7) also have sources of systematic uncertainty that affect this analysis. These uncertainties are also obtained from data-MC comparisons, separately for b-jets, c-jets and light flavoured jets, in bins of MV1c operating point and jet transverse momentum. Light flavoured jets had also an η-dependent uncertainty. A method, similar as for jets, decorrelated the b-tagging uncertainties [78] into 35 components that are considered for the ZH analysis: 10 for b-jets, 15 for c-jets and 10 for light flavoured jets [82]. Another b-tagging uncertainty is considered for samples simulated with generators different than PYTHIA6 due to the MC-to-MC correction, discussed in section 4.7. The uncertainty considered was 50% of the correction. Figure 4.21 shows a relative uncertainty in this $m_{b\bar{b}}$ distribution up to 5 %. The b-tagging uncertainty is the dominant experimental systematic uncertainty in the ZH analysis.

4.10.2. Uncertainties in the modelling of the ZH signal process

Whereas the experimental systematic uncertainties relate on how well known are the reconstructed physics objects, the theoretical or modelling uncertainties, result from the knowledge of the simulated processes.

An uncertainty of 3.3 % is considered for the branching ratio of the Higgs decay to $b\bar{b}$, determined with HDECAY, accounting for N4LO QCD and NLO EW corrections [82, 138]. After the topological selection, the contributions from other decay modes besides to $b\bar{b}$ are less than 1 % [82].

The calculated cross section obtained at NNLO QCD has an uncertainty of 3.1 %, evaluated from varying the factorization and renormalization scales up and down by a factor of 2 for this calculation [138]. Two additional uncertainties to the cross section were evaluated from the PDF4LHC recommendations [139] following the NLO prescription. One of the uncertainties is on the α_S parameter at the Z mass scale. Comparing the cross sections obtained with three different PDF sets, another uncertainty results from the relative differences between them. These two uncertainties are combined in quadrature resulting in a value of 2.5 %. Table 4.15 shows the separation of the uncertainties on the cross section for the quark and gluon initiated process.

The selection, described in section 4.5.3, chose only a small fraction of events. A similar procedure to that just described for the cross section was used to derive uncertainties of the effect of scale/PDF variations on the fraction of signal accepted events (acceptance). These uncertainties were obtained separately for two and three jet multiplicities, and for the gluon and quark initiated processes, ranging from 1.5 % to 5 %. Table 4.15 shows again the uncertainties on acceptance for each category. Additionally, when the scale was varied, the p_T^Z distribution was also affected and a shaped systematic uncertainty was added to the acceptance, based on a linear envelope on the p_T^Z ratio distribution between the different scales and the nominal ($\mu_R/\mu_F = 1$). The uncertainty varied from 1 % (2 %) for $p_T^Z = 50$ GeV to 3 % (8 %) for $p_T^Z = 200$ GeV in the quark (gluon) initiated process. The same exercise was carried out for the PDF uncertainties, with no evidence of a p_T^Z dependence.

A systematic uncertainty is considered for changes in the parton shower and hadronization by comparing two simulations with the NLO POWHEG generator, one of them interfaced with PYTHIA8 and the other with HERWIG for propagation processes. Variations on the acceptance were found to be about 8 %, reaching 13 % for 3-jet and $p_T^Z > 120$ GeV.

Finally, a shape systematic uncertainty is considered for the electroweak *NLO* correction (discussed in section 4.7) of 2.5 % for $p_T^Z > 200$ GeV and about 2 % in other p_T^Z ranges [82].

Table 4.15 summarises the systematic uncertainty sources considered for the two signal samples and their associated uncertainties. The signal strength parameter itself is left to float freely in the global fit, since it is the objective of this analysis.

4.10.3. Uncertainties in the modelling of the Z+jets background

The Z+jet is the leading background of the ZH analysis and is expected to dominate also the systematic uncertainties.

A 50 % uncertainty was estimated for the shape correction on the p_T^Z distribution. The shape correction on the $\Delta\phi$ distribution for Z+jets vetoing C and B-hadrons was also evaluated with a 50 % uncertainty. After requiring the two b-tagged jets condition, the slope in the ratio between the $\Delta\phi$ distribution for data and MC vanished, which resulted

Uncertainty source	Value
Branching ratio	3.3 %
Cross section (scale)	1 % (qqZH) 50 % (ggZH)
Cross section (PDF)	2.4 % (qqZH) 17 % (ggZH)
Acceptance (scale)	1.5 – 3.4 %
Acceptance (3-jet)	3.3 – 3.6 %
Scale of p_T^Z	Shape
Acceptance (PDF)	2 – 5 %
Parton shower	8 – 13 %
NLO EW correction on p_T^Z	Shape

Table 4.15.: Summary of the uncertainty sources considered in the signal samples.

in considering an uncertainty of 100 % on this shape correction, for Z+*hf* events.

The jet composition of the Z+jets background was not accurate enough and the normalization of each sub-sample was obtained from data-MC comparisons in control regions, using the global fit (see section 4.9.1). The *0-tag* control region was highly enriched with events from the *Z+l* category, as it was seen in figure 4.14, and a systematic of 5 % on the normalization of this region was established from the data-MC comparison. For the remaining categories, the systematic uncertainties were not so trivial. A different simulation of the Z+jets process, generated with ALPGEN [94] and interfaced with HERWIG for parton shower and hadronization, was used to estimate the systematic uncertainties. The comparison between these two simulated samples provided uncertainties on the ratios between two components: 12 % on the normalization ratio between the *Z+bb* category and the remaining heavy flavour categories (*Z+bc*, *Z+bl* and *Z+cc*). An additional 20 % uncertainty was estimated for the ratio in the normalizations between 2 and 3 jet multiplicities for the *Z+hf* category, and of 26 % for the *Z+cl* category.

A comparison of the $m_{b\bar{b}}$ distributions in the 2-tag combined region between data and the default simulation (SHERPA), shows that another shape systematic uncertainty on this process has to be considered. The invariant mass window $100 < m_{b\bar{b}} < 150$ GeV is excluded from the comparison. An increase in the $m_{b\bar{b}}$ distribution at 50 GeV by 3%, would decrease the distribution by 5% at 200 GeV, with a linear dependence on $m_{b\bar{b}}$. This uncertainty covers the differences between the two simulations discussed, SHERPA and ALPGEN, and no other uncertainty was added.

Table 4.16 summarises the systematic uncertainty sources considered for Z+jets.

Uncertainty source	Value
$\Delta\phi\ (b_1, b_2)$	Shape
$\mathrm{p}_{\mathrm{T}}^{Z}$	Shape
Z+*l* normalisation	5 %
Z+*l* 3/2-jet ratio	5 %
Z+*cl* 3/2-jet ratio	26 %
Z+*hf* 3/2-jet ratio	20 %
Z+*bc* / *Z*+*bb* ratio	12 %
Z+*bl* / *Z*+*bb* ratio	12 %
Z+*cc* / *Z*+*bb* ratio	12 %
$m_{b\bar{b}}$	Shape

Table 4.16.: Summary of the uncertainty sources considered in the Z+jets samples.

4.10.4. Uncertainties in the modelling of the $t\bar{t}$ background

About 13 % of the events that survive the topological selection are expected to originate from the $t\bar{t}$ process. So any systematics resulting from modelling of this process are expected to have a non-negligible effect on the signal strength parameter.

A correction was applied to the average p_{T} of the two top quarks in the $t\bar{t}$ process, as discussed in section 4.7. The systematic uncertainty considered for this correction is 50 % of the correction.

The normalization of the $t\bar{t}$ background is allowed to float freely in the global fit and is obtained from the *top* e-μ control regions. Events in this region mostly lie in the 2-jet region, with a $\mathrm{p}_{\mathrm{T}}^{Z} < 120$ GeV. Extrapolation to the 3-jet and high-$\mathrm{p}_{\mathrm{T}}^{Z}$ regions result from additional uncertainties, obtained by comparing the nominal POWHEG+PYTHIA6 simulation that uses the CT10 PDF set (see section 4.4), with other simulated samples:

- the same generator, but using the HERAPDF [149] PDF set.
- the same generator, but using HERWIG for hadronization.
- the same hadronization scheme, but different tree-level generators (MC@NLO [150] and ACERMC [95]).
- Multi-leg LO generator (ALPGEN) with PYTHIA6 for hadronization.

- the effect of changing the amount of ISR and FSR by changing the default parameters in PYTHIA6, with the ACERMC generator.

The uncertainties estimated from these comparisons were 20 % for the $t\bar{t}$ normalisation of events with three jets and 7.5 % for the normalisation of events with $\mathrm{p}_\mathrm{T}^\mathrm{Z} > 120$ GeV.

A linear shape systematic uncertainty was evaluated on the invariant mass distribution from the comparison of the $t\bar{t}$ simulations referred. If there was an increase on the $m_{b\bar{b}}$ distribution of 3 % at $m_{b\bar{b}} = 50$ GeV, the distribution would decrease by 1 % at $m_{b\bar{b}} = 200$ GeV for events with two jets and $\mathrm{p}_\mathrm{T}^\mathrm{Z} > 120$ GeV. A similar behaviour was observed for events with three jets but with opposite sign. For the low-$\mathrm{p}_\mathrm{T}^\mathrm{Z}$ region, the shape uncertainty on the $m_{b\bar{b}}$ distribution is smaller.

Table 4.17 summarises the four sources of uncertainty found for the $t\bar{t}$ background.

Uncertainty source	Value
Top-quark p_T	Shape
3/2-jet ratio	20 %
High/low-$\mathrm{p}_\mathrm{T}^\mathrm{Z}$ ratio	7.5 %
$m_{b\bar{b}}$	Shape

Table 4.17.: Summary of the uncertainty sources considered in the $t\bar{t}$ process.

4.10.5. Uncertainties in the modelling of the ZZ background

The ZZ background is expected to be only a small fraction of the events that survive the topological selection. However, this background still have about six times more events than the signal ZH sample and there ought to be some overlap between the invariant mass distributions of this background and the signal, from the large widths observed in this distribution, in section 4.8. Therefore, the contribution to the systematic uncertainty on the signal strength parameter should also be non-negligible.

The scale and PDF uncertainties on the cross section and acceptance of the ZZ background were evaluated using the same procedure as for the ZH. The renormalization and factorization scales are also varied by a factor of two and $1/2$ and the differences on the cross section are obtained in the five $\mathrm{p}_\mathrm{T}^\mathrm{Z}$ bins and the two jet multiplicities (two and three jets) and considered as the systematic uncertainty. This uncertainty increases with $\mathrm{p}_\mathrm{T}^\mathrm{Z}$, reaching for $\mathrm{p}_\mathrm{T}^\mathrm{Z}$ larger than 200 GeV, about 12 % for 2 jet events and 17 % for three jet events. The uncertainty on the PDF, following the same PDF4LHC procedure defined for the signal process, is 3 % in both jet multiplicities, with no correlation found between the p_T^Z bins. Different hadronization models are evaluated also, replacing the default simulation, PYTHIA8, with HERWIG. A shape uncertainty is considered in the $m_{b\bar{b}}$ distribution in this case. Table 4.18 summarizes uncertainty sources evaluated for the ZZ background.

Uncertainty source	Value
Cross section and acceptance (scale)	$3-17$ %
Cross section and acceptance (PDF)	3 %
$m_{b\bar{b}}$	Shape

Table 4.18.: Summary of the uncertainty sources considered in the ZZ process.

4.10.6. Uncertainties in the modelling of the remaining backgrounds

The signal and the Z+jets, $t\bar{t}$ and ZZ processes are the only backgrounds with a significant contribution in the ZH analysis. Tables 4.8 showed that the contribution of the remaining backgrounds accounts for less than 1 % altogether, after applying the topological event selection. These backgrounds affect, however, the 1-lepton and the 0-lepton analysis and were evaluated in detail and described in reference [82].

For the estimated multi-jet background, that was estimated from data, as described in section 4.4.4, a systematic uncertainty of 100 % is assumed, uncorrelated between the normal analysis regions and the *top* e-μ control region.

4.11. ZH analysis results

The signal strenght parameter obtained with the analysis described in this thesis was $\mu = -0.69^{+1.33}_{-1.25}$. The negative value of the signal strenght parameter indicates a statistic flutuation of the data. The Standard Model prediction, $\mu = 1$, is still within two stardard deviations from this result. Table 4.19 shows the contributions from the statistical and some systematic uncertainties to the total uncertainty. The impact on the total error from the individual uncertainty components was estimated by fixing the nuisance parameter to nominal, and calculating μ. Then, the same nuisance parameter is shifted up and down by its uncertainty, and calculating μ again. The quadratic difference to the nominal calculated

μ is the impact of this nuisance parameter. The same procedure is followed for the other nuisance parameters. The second column of table 4.19, shows the impact on the total error thus evaluated, summing quadratically the uncertainties. The third column shows the fractional error impact from the quadratic difference from total uncertainty and each of the impacts. The uncertainty on the ZH analysis is dominated by the statistical uncertainty with a contribution of 63 %. The specific uncertainties evaluated for the new calibration for jets (see section 3.9.4), *Jet b-jets* and *Jet AFII*, are shown in the table and have small contributions to the uncertainty of μ. The other jet systematic uncertainties and the b-tagging uncertainties are the dominant sources of experimental uncertainties, contributing 1-2% to the total uncertainty of the result. The modelling of the simulations is dominated by the normalization uncertainty of the Z+jet background.

The expected signal significance obtained from the fit was 0.84. Since the signal strenght parameter was negative the observed significance is zero by default. In addition to the signal strenght parameter, the background normalizations and respective uncertainties are shown in Table 4.20.

Figure 4.22 shows the $m_{b\bar{b}}$ distributions for the two regions where the signal-to-background ratio is expected to be the largest: events in the boosted region, $p_{\mathrm{T}}^{\mathrm{Z}} > 200$ GeV, with the medium and tight b-tagging criteria, after the global fit. These regions are therefore those most relevant for the signal strenght parameter extraction. However, in these plots the number of observed events is much smaller that those expected with and even without the Higgs boson predicted by the stardard model. Particularly, the number of events expected with (without) the Standard Model Higgs boson, in the three bins centred at 125 GeV, was 10.6 (9.2) and the observed number of events was 8. This is the reason why the signal strenght parameter was negative.

Figure 4.23 shows the MV1c weight distribution of the b-jet in the *1-b-tagged* control region (a) and the $m_{b\bar{b}}$ distribution in the *top* e-μ control region (b). In the latter, the overflow is included in the last bin. After the global fit, simulation and data agree within uncertainties. Particularly, the large change in the *top* e-μ control region results from the multi-jet background becoming negligible after the fit.

Appendix C.3 shows all the distributions used in the fit: prefit and post-fit, before and after applying the $m_{b\bar{b}}$ tranformation discussed in section 4.9. The first two $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$ regions have enough statistics to make precise comparisons between the data and the predicted simulation distributions. The difference to the unity of the data/MC ratio is often smaller than 10 %

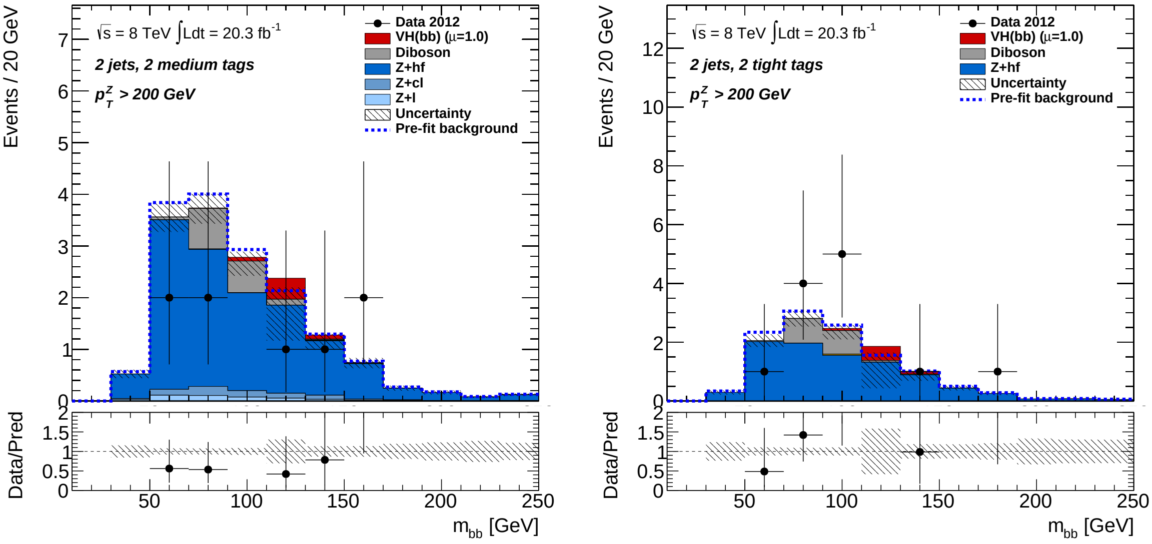
$\mu = -0.69^{+1.33}_{-1.25}$		
Uncertainty source	Impact on error	Fractional error impact
Statistical	$+1.05$ -0.99	$+63\%$ -63%
Systematics	$+0.812$ -0.761	$+37\%$ -37%
Jets	$+0.141$ -0.153	$+1.1\%$ -1.5%
Jet AFII	$+0.0114$ -0.00446	$+0.0073\%$ -0.0013%
Jet b-jets	$+0.0667$ -0.0544	$+0.25\%$ -0.19%
Jet Flavour comp	$+0.0623$ -0.0717	0.219% 0.330%
Jet Flavour resp	$+0.0517$ -0.0595	0.151% 0.227%
MET	$+0.0104$ -0.0051	$+0.0062\%$ -0.0017%
Muons	$+0.0142$ -0.0104	$+0.011\%$ -0.0069%
Electrons	$+0.0582$ -0.0469	$+0.19\%$ -0.14%
All b-tagging	$+0.19$ -0.136	$+2\%$ -1.2%
Luminosity	$+0.0283$ -0.0376	$+0.045\%$ -0.091%
VH	$+0.098$ -0.23	$+0.54\%$ -3.31%
Z+jets	$+0.428$ -0.441	$+10\%$ -12%
ttbar	$+0.0885$ -0.0619	$+0.44\%$ -0.25%
ZZ	$+0.0090$ -0.012	$+0.0046\%$ -0.091%
All normalizations	$+0.494$ -0.461	$+14\%$ -14%

Table 4.19.: Signal strenght parameter with the total uncertainty resulted from the global fit. The impact on some of the uncertainties sources and the fraction from the total uncertainty estimated from the global fit.

Parameter	Nominal value	Total uncertainty
Zbb normalisation	0.95	±0.05
Zcl normalisation	0.75	±0.11
$t\bar{t}$ normalisation	1.05	±0.06

Table 4.20.: Floating background normalisations and uncertainties.

and never larger than 20 %. The diboson ZZ contribution is always present, as it behaves with the same signature as the signal process. The $t\bar{t}$ background is relevant in the low-$\mathrm{p}_\mathrm{T}^\mathrm{Z}$ regime with a negligible contribution for $\mathrm{p}_\mathrm{T}^\mathrm{Z} > 160$ GeV. The multi-jet background, evaluated with the procedure in section 4.4.4, was very small and becomes negligible after the fit in



(a) MM 2 jets

(b) TT 2 jets

Figure 4.22.: Invariant mass of the two b-tagged jets after the fit for $p_T^Z > 200$ GeV for events with two jets for the b-tagging regions: MM (a) and TT (b). The dashed line corresponds to the background distribution before the fit.

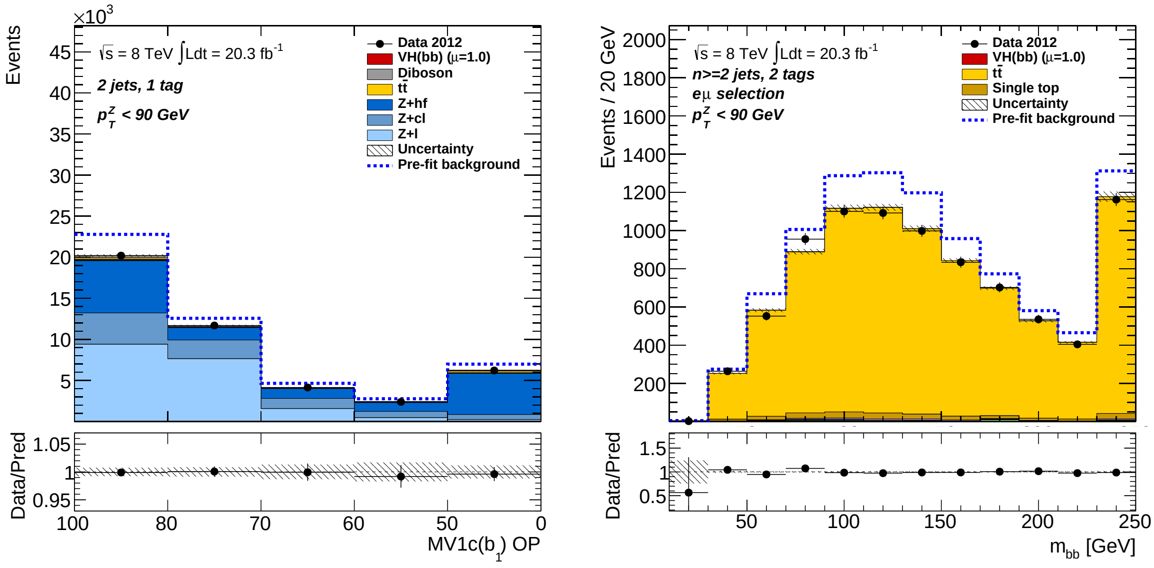
all regions. The lighter components of the Z+jet background, Z+cl and Z+l, have negligible contributions in the MM and TT b-tagging regions.

4.12. Combined VH analysis

Section 4.1 introduced the VH analysis as composed of three different channels: 0, 1 and 2-lepton analyses. The objective of this section is to refer the global results from the combined analysis, describing the analysis methods used. In addition to the ZH analysis, the LIP group also made contributions to the WH analysis (1-lepton channel) [151].

4.12.1. Multivariate analysis

The ZH analysis contemplated in this thesis uses a set of event selection criteria, described in section 4.5, to select a region, where the number of events from the signal divided by those from the backgrounds was as large as possible. This technique is called



(a) (b)

Figure 4.23.: MV1c weight distribution in the *1-tag* control region in events with two jets, after the fit (a). Invariant mass of the two b-tagged jets after the fit in the *top* e-μ control region (b). Both distributions refer to events with $p_{\mathrm{T}}^{Z} < 90$ GeV. The dashed line corresponds to the background distribution before the fit.

a cut-based analysis. There was another complementary analysis which used a multivariate procedure, the boosted decision trees (BDT) method [152, 153], present in the ROOT toolkit for multivariate analysis (TMVA) [154]. Softer selection criteria are applied to the event topology to prepare these trees. The physics objects selection and overlap removal are the same in both cut-based and multivariate analysis. The differences in the event selection are summarised on table 4.21. The di-lepton invariant mass is loosened to the 71-121 GeV window, and the selections of the missing transverse energy and the high cut of the angular distance, ΔR, between the two b-tagged jets are removed. This allows the BDT to better optimise the selection. Furthermore, the multivariate analysis considers only two $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$-regions, above and below 120 GeV, instead of the original five regions.

A set of 14 variables was used in the BDT:

- the transverse momentum of the Z boson, $\mathrm{p}_{\mathrm{T}}^{\mathrm{Z}}$;
- the invariant mass between the two leptons, $m_{\ell\ell}$;
- the missing transverse energy, $\not{E}_{\mathrm{T}}$;

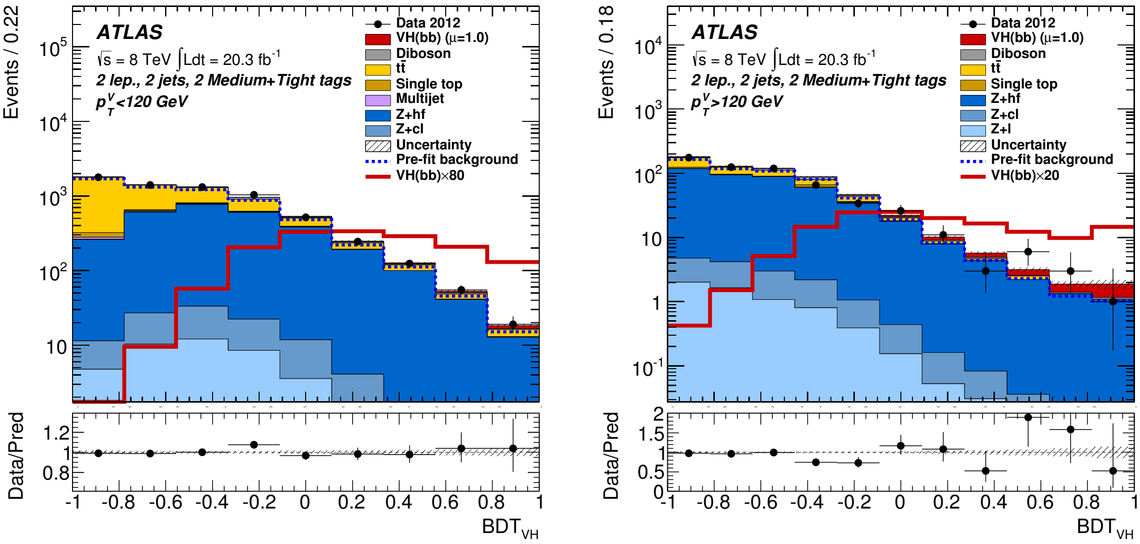
Selection	Cut-based analysis	Multivariate analysis
Di-lepton invariant mass	$83 < m_{\ell\ell} < 99$ GeV	$71 < m_{\ell\ell} < 121$ GeV
Missing transverse energy	$E\!\!\!/_{\mathrm{T}} < 60$ GeV	Removed
ΔR *high cut*	$\mathrm{p}_{\mathrm{T}}^{Z}$-dependent	Removed

Table 4.21.: Changes in the selection between the cut-based analysis and the multivariate analysis [82].

- the transverse momentum of the b-tagged jets, ${\mathrm{p}_{\mathrm{T}}}^{b1}$ and ${\mathrm{p}_{\mathrm{T}}}^{b2}$;
- the b-jets MV1c weight, w_{MV1c}^{b1} and w_{MV1c}^{b2};
- the invariant mass between the two b-jets, $m_{b\bar{b}}$;
- the angular distance of the two b-jets, $\Delta R(b1,b2)$;
- the difference of the pseudorapidity between the two b-jets, $\Delta\eta(b1,b2)$;
- the difference of the pseudorapidity between the Z boson and the Higgs candidate, $\Delta\eta(Z,H)$;
- the difference of the azimuth angle between the Z boson and the Higgs candidate, $\Delta\phi(Z,H)$;
- in three jet events, the transverse momentum of the third jet, ${\mathrm{p}_{\mathrm{T}}}^{j3}$;
- in three jet events, the invariant mass of the three jet system, $m_{b\bar{b}j}$;

The starting point of the BDT method is the training procedure, that uses a set of the signal and background events after the loose selection discussed above and normalised to the same value. A decision tree looks to one of the variables and evaluates the selection in iterative steps to choose the best threshold that separates the signal from the background events. TMVA does this for every variable in the list. The variable and cut that has the most discriminating power is chosen to split the tree into two *branches*. The same process is repeated for each of the branches until some specific conditions stop the growing of the tree and the final member of the tree is called *leaf*. The condition to stop the growth of the tree can be the events falling below a certain number, reaching the maximum number of selections (layers), or if the chosen variable-cut pair no longer has a significant separation power. Each of the leafs is considered to be either signal or background. Depending on the signal/background fraction in the leaf, events are given a weight +1 or -1, respectively. Some of the leafs will have events that were misclassified by the decision tree. The boosting technique gives a larger weight (a boost) to such events and executes the optimisation procedure again as described above. This process is repeated several (even hundreds) of times saving each decision tree. A final BDT output variable is constructed using the weighted average of the classification weights. The BDT output distribution, which varies from -1 to +1, is obtained for data and each of the backgrounds.

The BDT is trained independently for different signal regions, considering events with p_T^Z below/above 120 GeV, events with two/three jets and even/odd event number. Since simulation samples are not infinite, the training of odd and even event number does not produce the same BDT result. The BDT that is trained with events with the odd number is applied to the events with even number to estimate this bias. To minimise the statistical uncertainty, the opposite also happens: trained BDT with odd event number applied to even number events. Finally, the same BDTs are also applied to data events.



(a) $p_T^Z < 120$ GeV

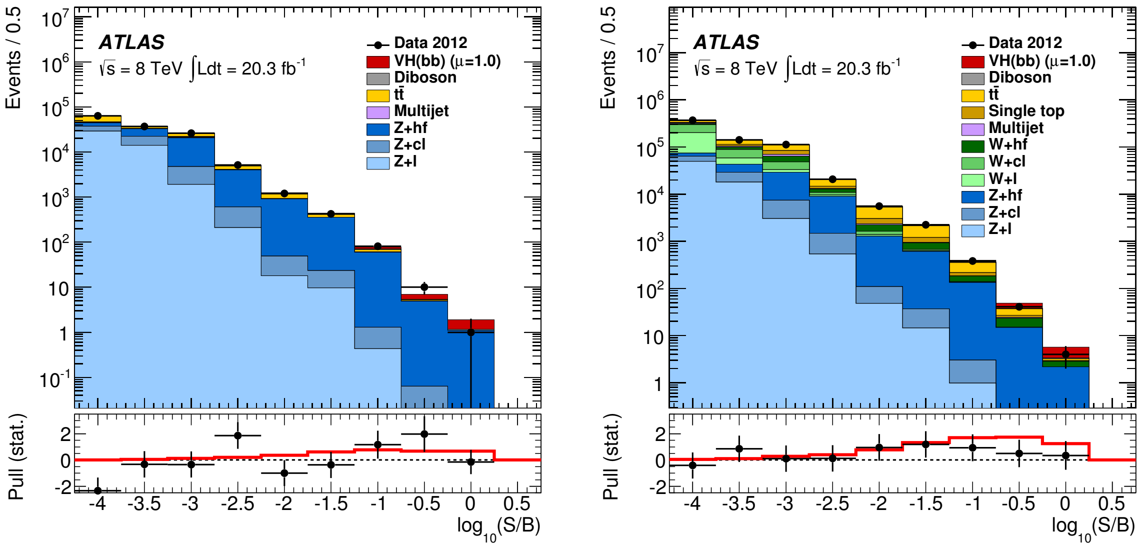
(b) $p_T^Z > 120$ GeV

Figure 4.24.: BDT output distribution in the 2-lepton channel for events with two jets and combining the MM and TT b-tagged regions, after applying the global fit [82].

In the multivariate analysis, the BDT output distributions replace the invariant mass distributions in the two b-tagged jet regions to be used in the global fit discussed in section 4.9.1. Additionally, to reduce the statistical uncertainty, the MM and TT b-tagged regions are combined into a single BDT output. Figure 4.24 shows the BDT distributions for this combined b-tagged region for low-(a) and high-p_T^Z (b) region, after the fit is performed. In the first bins, the $t\bar{t}$ background dominates, so its normalisation is constrained from this part of the BDT distribution and the *top* e-μ region is no longer needed. The signal events are expected to populate the region with a larger BDT output variable.

4.12.2. Run-1 VH results

The best fit results for the VH analysis were obtained with the multivariate analysis and using the three lepton channels for the 2012 data taking. For the 2011 data, the integrated luminosity was smaller leading to large statistical uncertainties, so only the cut-based analysis was used in the Run-1 combination [82]. The cut-based analysis from 2012 data taking was used as one of the validations of the MVA results.

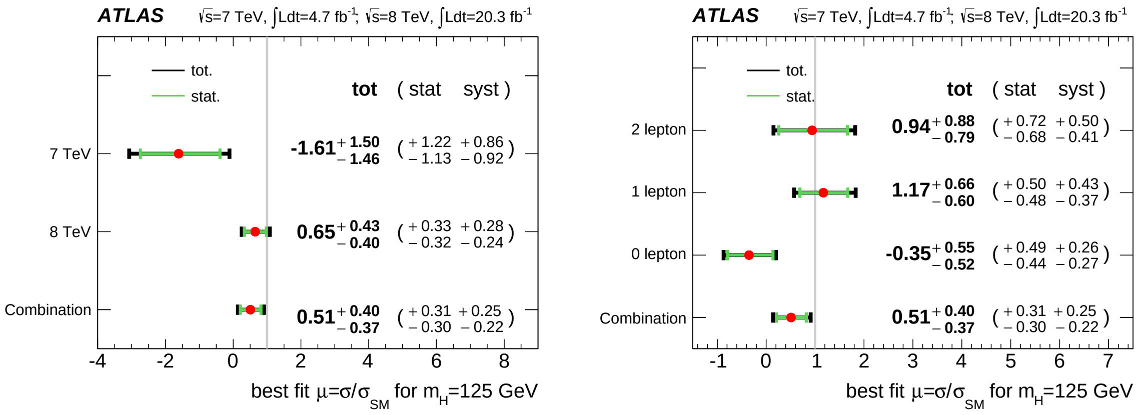


(a) 2-lepton

(b) 0+1+2 combined

Figure 4.25.: Distribution of the logarithm of the ratio between the number of events from signal to those from background for the 2-lepton channel (a) and the combination of the 0-, 1- and 2-lepton channels (b) [82].

Figure 4.24 showed the BDT output distribution for two regions of the 2-lepton analysis. To obtain a single histogram that summarises the BDT outputs information from all regions, the ratio between the number of signal and background events is evaluated in each of the BDT output bins. Figure 4.25 (a) shows the distribution of the logarithm of this ratio for the 2-lepton analysis. The bottom panel shows the relative difference to the background-only hypothesis of data (black markers) and with the hypothesis of the Standard Model Higgs boson at 125 GeV (red line). Figure 4.25 (b) shows the distribution of the logarithm of the ratio between the number of signal and the number of background events for the combination of the three lepton channels. This figure shows smaller fluctuations when combining the three channels. The histograms show that the data observed is compatible with both the null hypothesis (no Higgs boson) and the existance of the Standard Model Higgs boson, due to the large uncertainties observed.



(a) (b)

Figure 4.26.: Signal-strength parameter, μ, obtained from a global fit combining the 0-, 1- and 2-lepton channels and separating the $\sqrt{s} = 7$ TeV pp collisions (2011 data) from $\sqrt{s} = 8$ TeV (2012 data) (a). Signal-strength parameter, μ, obtained from a combined fit of the $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV pp collisions (full Run-1 dataset) for the three analysis channels (0-, 1- and 2-leptons) separately and combined (b) [82].

Figure 4.26 (a) shows the signal-strength obtained from the global fit combining the 0-, 1- and 2-lepton channels for cut-based analysis in 2011 and for the multivariate analysis in 2012. Figure 4.26 (b) shows alternatively the combination of the two data taking periods but separating the three analysis channels. The $\sqrt{s} = 8$ TeV analysis with increased statistics, a better understanding of the systematics and the usage of a multivariate analysis was able to have a much smaller uncertainty than $\sqrt{s} = 7$ TeV. The combined result of the signal-strength for Run-1 with the three lepton analysis was $0.52 \pm 0.32(\text{stat.}) \pm 0.24(\text{syst.})$, for a Higgs boson mass of 125.36 GeV [82].

The combined cut-based analysis of the three channels obtained a signal-strength parameter of $1.23 \pm 0.44(\text{stat.}) \pm 0.41(\text{syst.})$ and a observed (expected) significance of 2.2 σ

(1.9 σ). The multivariate analysis expected significance was of 2.5 σ. For this reason, the latter was chosen as the nominal analysis.

The results obtained in the ATLAS collaboration were also compatible with those from the CMS collaboration, that reported a signal strength parameter of 1.0 ± 0.5, with an expected and observed significances of 2.1 [155]. The analysis of Run-2 data is now underway and preliminary $\sqrt{s} = 13$ TeV results from 2016 summer conferences set the value in $\mu = 0.21^{+0.36}_{+0.35}(\text{stat.}) \pm 0.36(\text{syst.})$ [156].

5. Conclusions and outlook

The LHC started a golden age of experimental particle physics in the testing of the theoretical predictions of the Standard Model. The discovery of the Higgs boson in 2012, by the ATLAS and CMS experiments, was undoubtedly one of the greatest marks of particle physics. As a physicist from one of these experiments, I analysed LHC proton-proton collisions data to determine if this particle is compatible with the Standard Model Higgs boson or if it brings hints of any new physics.

A particular physics search is the measurement of the $H \to b\bar{b}$ decay, the major branching ratio for the 125 GeV Higgs boson. This decay is, however, one of the most challenging channels to be studied for in a proton-proton collider such as LHC. Nearly every inelastic collision produces quarks and gluons that hadronize and are only well reconstructed as jets. It is impractical for any detector to save all the events from the LHC collisions at a rate of 20 million per second. One Higgs boson is produced every $\sim$17 seconds (assuming an instantaneous luminosity of $\mathscr{L} = 10^{33}$ cm^{-2}s^{-1}) but this decay mode can only be studied with a trigger that provides a clear signature, such as the associated production. The objective of this thesis was the search for the $H \to b\bar{b}$ decay when the Higgs boson is produced in association with a Z boson. The leptonic decay of the Z boson to muons or electrons provides the required signature in an event where all the physics objects are fully reconstructed.

The analysis of the ZH search channel is, however, still dominated by background processes, where the largest contributions are the production of jets associated with a Z boson, the production of a pair of top quarks and the production of a pair of Z bosons. For this reason, a clear understanding of the detector is crucial. This thesis makes use of the leptonic decay of the Z boson to an electron pair or a muon pair. As the reconstruction of these pairs of particles is relatively straight forward, I dedicated most of my effort to the improvement of the knowledge of the reconstructed jets from the Higgs boson decay $H \to b\bar{b}$, enhancing two aspects which are discussed in the following.

The first one is the calorimeter response to single hadrons, discussed in section 3.8. It plays a major role in the validation of the jet energy scale systematic uncertainty and in the setting of the jet energy scale uncertainty for jets above 1.8 TeV. Before my work developed for this thesis, there was no information on the single hadron response in the energy range from 10 to 20 GeV, due to lack of statistics in collision events. The development of two triggers that unbiased the E/p measurement while selecting high momentum and isolated tracks was one of the objectives of my work. The triggers extended the momentum coverage from the E/p collision measurements up to 30 GeV, which provided an overlap with the method used before collisions, validating it. Additionally, the triggers developed were run in a large pileup environment and validated the uncertainty up to an average number of interactions per bunch crossing of $\langle\mu\rangle = 20$.

The other main contribution to the knowledge of the calibration of jets was the Global Sequential Calibration (GSC) that reduced the jet energy response dependence with the jet flavour. I tested and used it in the context of the ZH analysis, *i.e.* using it for b-jets. GSC removed the jet energy response dependence from some of the jet properties that are usually used to identify the flavour of the jets. The complete set of conclusions from this study was already discussed in section 3.9.6 and I just refer here the most relevant ones. The systematic uncertainty from the jet energy response difference between jets initiated by light quarks and jets initiated by gluons was reduced to negligible values, after using GSC as intended. The b-jet systematic uncertainty was verified to be still valid after the application of GSC, when comparing the jet energy responses among different generators, among the various hadronization models and among the three leading background processes of the ZH analysis (referred above) when comparing with the signal process. The GSC calibration required, however, an additional systematic uncertainty of 0.5 % for MC samples that use the fast simulation of the detector, where a parametrization of the calorimeter response is used instead of the full simulation of the detector with GEANT4.

The main result of the global sequential calibration was the improvement of the jet energy resolution by about 20 %, and I also studied the improvement obtained on the distributions of the invariant mass between the two reconstructed b-jets, when applying this calibration. This calibration was able to enhance the invariant mass resolution up to 18.5 % in the most sensitive region of the ZH analysis, where the reconstructed Z boson has a transverse momentum larger than 200 GeV.

An independent ZH analysis, discussed in chapter 4, was performed in close collaboration with colleagues from other institutes. A comparative analysis if all the results allowed the validation of the performance software tools for the physics objects and the validation of the independent software of the analysis, with differences evaluated at the *per-mille* level. From a statistical combination fit, the signal strength parameter was obtained with a value of $\mu = -0.69^{+1.33}_{-1.25}$. The total uncertainty obtained still shows compatibility with the Standard Model Higgs boson ($\mu = 1$) and with the strength parameter value achieved with the VH combination for $\sqrt{s} = 8$ TeV: $\mu = 0.65^{+0.43}_{-0.40}$. The work developed in this thesis resulted in a 0.85σ of expected signal significance. The final value using the three alternatives of VH was 2.5σ when considering a multivariate analysis.

The LHC increased in 2015 the centre of mass energy from $\sqrt{s} = 8$ TeV to $\sqrt{s} = 13$ TeV, which increases the cross section production of all processes, particularly the production of the $t\bar{t}$ process. Both the WH analysis and the 0-lepton analysis have an enormous background contribution from this process. However, in the ZH analysis, by applying a very tight criterium on the b-tagging and requiring that the Z boson is produced boosted (large-p_T), this background is reduced to negligible values. Since the cross section production of the ZH process is expected to increase by the same amount as the Z+jets background, for the same amount of collected data the ZH analysis should have an improvement on the expected significance.

Bibliography

[1] S. Weinberg. A model of leptons. *Phys Rev Lett*, 19(21):1264–1266, November 1967.

[2] S.L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22(4):579–588, February 1961. `https://doi.org/10.1016/0029-5582(61)90469-2`.

[3] A. Salam. Gauge unification of fundamental forces. *Rev. Mod. Phys.*, 52:525–538, 1980.

[4] C. Patrignani *et al.*. The Review of Particle Physics, 2016.

[5] A. Djouadi. The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model. *Phys.Rept.457:1-216,2008*, 2008. arXiv:hep-ph/0503172.

[6] R. Kleiss. Field Theory & the E-W Standard Model. `https://indico.cern.ch/event/287744/contributions/1641244/attachments/535746/738660/qft.pdf`, 2014. European school of high-energy physics.

[7] D. J. Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.

[8] M. Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.

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

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

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett. 13, 585*, 1964.

[12] N.A. Ky and N.T.H. Van. Was the Higgs boson discovered? *arXiv:1503.08630*, 2015.

[13] LHC Higgs Cross Section Working Group (2012-2013). `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections`, 2013.

[14] H. Flacher *et al.*. Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter. *Eur.Phys.J. C60 (2009) 543-583*, 2009. arXiv:0811.0009.

[15] G. Abbiendi *et al.*. Search for the Standard Model Higgs Boson at LEP. *Phys.Lett.B565:61-75,2003*, 2003. arXiv:hep-ex/0306033.

[16] J. Baglio *et al.*. The Tevatron Higgs exclusion limits and theoretical uncertainties: a critical appraisal. *Phys.Lett.B699:368-371,2011*, 2011. arXiv:1101.1832v3.

[17] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett. B716 (2012) 1-29*, 2012. arXiv:1207.7214 [hep-ex].

[18] The CMS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B 716 (2012) 30-61*, 2012. arXiv:1207.7214 [hep-ex].

[19] ATLAS and CMS Collaboration. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett. 114, 191803 (2015)*, 2015. arXiv:1503.07589.

[20] The ATLAS Collaboration. Projections for measurements of Higgs boson cross sections, branching ratios and coupling parameters with the ATLAS detector at a HL-LHC. *ATL-PHYS-PUB-2013-014*, 2013.

[21] LHC Higgs Cross Section Working Group, S. Heinemeier, C. Mariotti, G. Passarino, R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, 2013. arXiv:1307.1347.

[22] ATLAS and CMS Collaboration. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$= 7 and 8 TeV. *CERN-EP-2016-100*, 2016. arXiv:1606.02266 [hep-ex].

[23] The ATLAS collaboration. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J. C75 (2015) 476*, 2015. arXiv:1506.05669v2.

[24] The CMS Collaboration. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev. D 92, 012004 (2015)*, 2015. arXiv:1411.3441.

[25] W. J. Stirling. `http://www.hep.ph.ic.ac.uk/~wstirlin/plots/crosssections2012_v5.pdf`. private communication.

[26] Copyright ©2008-2016 CERN. About CERN webpage. `http://home.cern/about`.

[27] D0 Collaboration. Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV. *Phys. Rev. Lett. 74, 2422*, 1995.

[28] CDF Collaboration. Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett. 74, 2626*, 1995.

[29] Copyright ©2008-2016 CERN. `https://www.cern.ch`.

[30] G. Aad *et al.*. The ATLAS experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[31] ATLAS Collaboration. Expected performance of the ATLAS experiment: detector, trigger and physics. *CERN-OPEN-2008-020*, 2008. arxiv:0901.0512.

[32] ATLAS Experiment ©2014 CERN. `https://www.atlas.ch/photos/index.html`.

[33] ATLAS Inner Detector Community. Inner Detector Technical Design Report. *CERN/LHCC*, 97-16, 1997.

[34] ATLAS Muon Collaboration. ATLAS Muon Spectrometer Technical Design Report. *CERN/LHCC*, 97-22, 1997.

[35] ATLAS Collaboration. Calorimeter Performance Technical Design Report. *CERN/LHCC*, 96-40, 1997.

[36] ATLAS LARG Unit. Liquid Argon Calorimeter Technical Design Report. *CERN/LHCC*, 96-41, 1996.

[37] M. L. Andrieux *et al.*. Construction and test of the first two sectors of the ATLAS barrel liquid argon presampler. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 479(2-3):316–333, 2002.

[38] ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *The European Physical Journal C*, 74(10), 2014. arXiv:1407.5063v2 [hep-ex].

[39] ATLAS/Tile Calorimeter. Tile Calorimeter Technical Design Report. *CERN/LHCC*, 96-42, 1996.

[40] R. Bartoldus *et al.*. Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System. `https://cds.cern.ch/record/1602235`, Sep 2013. Final version presented to December 2013 LHCC.

[41] ATLAS Collaboration. TDAQ in 2012. `https://twiki.cern.ch/twiki/pub/AtlasPublic/ApprovedPlotsDAQ/12.pdf`.

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

[43] K. Benslama *et al.*. Tau trigger at the ATLAS experiment. In *Proceedings, 34th International Conference on High Energy Physics (ICHEP 2008)*, 2008. arXiv:0810.0465.

[44] ATLAS Collaboration. Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at $\sqrt{s} = 7$ TeV in 2011. *ATLAS-CONF-2012-048*, 2012.

[45] ATLAS Collaboration. Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV. *CERN-PH-EP-2014-154*, 2014. arXiv:1408.3179v1 [hep-ex].

[46] ATLAS Experiment ©2015 CERN. ATLAS Stand-Alone Event Displays. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone`.

[47] ATLAS Collaboration. Luminosity public results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[48] ATLAS Collaboration. ATLAS Computing Techincal Design Report. *CERN-LHCC-2005-022*, 2005.

[49] ROOT Framework. `https://root.cern.ch`.

[50] ATLAS Collaboration. Performance of the ATLAS Trigger System in 2010. *Eur.Phys.J.C 72 (2012) 1849*, 2012. arXiv:1110.1530v2 [hep-ex].

[51] O Lundberg. Eta dependence of electronic noise and noise in a single channel, with changed LVPS. *ATLAS-PLOT-TILECAL-2012-005*, Apr 2012.

[52] ATLAS Collaboration. Noise in Liquid Argon at the electron scale. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LArCaloPublicResultsDetStatus`.

[53] ATLAS Collaboration. Pile-up Noise Performance for 2012. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsTileNoise`.

[54] ATLAS Collaboration. Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. *ATLAS-CONF-2012-042*, 2012.

[55] R. Frühwirth *et al.*. Application of Kalman filtering to track and vertex finding. *Nucl. Inst. Meth.*, A(262), 1987.

[56] P. Hough. Method and Means for Recognizing Complex Patterns. *U.S. Patent No. 3069654*, 1962.

[57] R. Duda and P. Hart. Use of the Hough Transform to detect lines and curves in pictures. *Comm. ACM 15*, pages 11–15, 1972.

[58] T. G. Cornelissen *et al.*. Updates of the ATLAS Tracking Event Data Model (Release 13). *ATL-SOFT-PUB-2007-003*, 2007.

[59] ATLAS Collaboration. Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data. *Eur.Phys.J. C74 (2014) 3130*, 2014. arXiv:1407.3935v2 [hep-ex].

[60] D0 Collaboration. Measurement of the shape of the boson rapidity distribution for $p\bar{p}->Z/\gamma*\rightarrow e^+e^-+X$ events produced at $\sqrt{s}$ of 1.96 TeV. *Phys.Rev.D76:012003*, 2007.

[61] CDF Collaboration. First Measurements of Inclusive W and Z Cross Sections from Run II of the Tevatron Collider. *Phys.Rev.Lett.94:091803*, 2005.

[62] ATLAS Collaboration. Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data. *Eur. Phys. J. C (2014) 74:2941*, 2014. arXiv:1404.2240.

[63] ATLAS Collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data. *CERN-EP-2016-110*, 2016. arXiv:1606.01813.

[64] The ATLAS collaboration. Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data at $\sqrt{s} = 8$ TeV. *PERF-2016-01*, 2016. To be submitted to: EPJC.

[65] ATLAS Collaboration. Electron Efficiency Measurements for 2012 and 2011 Data. *ATL-COM-PHYS-2013-1287*, 2013.

[66] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data. *ATLAS-CONF-2014-032*, 2014.

[67] ATLAS Collaboration. Egamma Trigger Public Results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EgammaTriggerPublicResults`, 2015.

[68] ATLAS Collaboration. Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV. *CERN-PH-EP-2014-227*, 2014. arXiv:1412.7086.

[69] ATLAS Collaboration. Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with ATLAS. *Eur.Phys.J.C 72 (2012) 1844*, 2012.

[70] ATLAS Collaboration. Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at $\sqrt{s} = 8$ TeV. *ATLAS-CONF-2013-082*, 2013.

[71] ATLAS Collaboration. Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C (2015) 75:17*, 2015. arXiv:1406.0076 [hep-ex].

[72] R. Atkin. Review of jet reconstruction algorithms. In *High Energy Particle Physics Workshop (HEPPW2015)*, number 645 in Journal of Physics: Conference Series. HEPPW2015, IOP Publishing, 2015. 012008.

[73] G. Soyez M. Cacciari, G. P. Salam. FastJet user manual. *CERN-PH-TH/2011-297*, C(72):1896, 2012. arXiv:1111.6097.

[74] G. P. Salam and G. Soyez. A practical Seedless Infrared-Safe Cone jet algorithm. *JHEP 0705:086,2007*, 2007. arXiv:0704.0292.

[75] G. Soyez M. Cacciari, G. P. Salam. The anti-k_T jet clustering algorithm. *arXiv:0802.1189 [hep-ph]*, 2008.

[76] The ATLAS collaboration. Pile-up subtraction and suppression in ATLAS. *ATLAS-CONF-2013-083*, 2013.

[77] ATLAS Collaboration. Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. *ATLAS-CONF-2011-102*, 2011.

[78] ATLAS Collaboration. Performance of b-jet Identification in the ATLAS Experiment. *CERN-PH-EP-2015-216*, 2015. arXiv:1512.01094.

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

[80] ATLAS Collaboration. Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data. *ATLAS-CONF-2014-046*, 2014.

[81] ATLAS Collaboration. Expected performance of the ATLAS b-tagging algorithms in Run-2. *ATL-PHYS-PUB-2015-022*, 2015.

[82] ATLAS Collaboration. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector. *JHEP01(2015)069*, 2014. arXiv:1409.6212v1.

[83] "Illustrated Information". Nobelprize.org. The Nobel Prize in Physics 1990. `http://www.nobelprize.org/nobel_prizes/physics/laureates/1990/illpres/index.html`, 2014. Nobel Media AB.

[84] H.-L. Lai *et al.*. New parton distributions for collider physics. *Phys. Rev. D*, 82(074024), 2010. arXiv:1007.2241.

[85] J. Pumplin *et al.*. New Generation of Parton Distributions with Uncertainties from Global QCD Analysis. *JHEP 0207:012*, 2002. arXiv:hep-ph/0201195.

[86] M.H.Seymour and M. Marx. Monte Carlo Event Generators. *MCnet-13-05*, 2013. 1304.6677v1.

[87] B. Webber. Parton shower Monte Carlo event generators. *Scholarpedia*, 6(12):10662, 2011. `http://www.scholarpedia.org/article/Parton_shower_Monte_Carlo_event_generators`.

[88] T. Sjöstrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05(026), 2006. arXiv:hep-ph/0603175v2.

[89] T. Sjöstrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008. arXiv:0710.3820 [hep-ph].

[90] G. Corcella *et al.*. HERWIG 6.5 Release Note. *arXiv:hep-ph/0210213v2*, 2005.

[91] K. Arnold *et al.*. HERWIG++ 2.6 Release Note. *arXiv:1205.4902*, 2012.

[92] T. Gleisberg *et al.*. Event generation with SHERPA 1.1. *JHEP*, 02(007), 2009. arXiv:0811.4622.

[93] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11(070), 2007.

[94] M. Mangano *et al.*. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP 0307 (2003) 001*, 2003. arXiv:hep-ph/0206293.

[95] B. P. Kersevan and E. Richter-Was. The Monte Carlo Event Generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1. *TPJU-6/2004*, 2012. arXiv:hep-ph/0405247.

[96] R. Gavin *et al.*. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput.Phys.Commun.182:2388-2403,2011*, 2011. arXiv:1011.3540.

[97] M. Czakon and A. Mitov. Top++: a program for the calculation of the top-pair cross-section at hadron colliders. *Computer Physics Communications 185 (2014) 2930*, 2014. arXiv:1112.5675.

[98] A. Denner *et al.*. Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with HAWK. *JHEP 1203 (2012) 075*, 2012. arXiv:1112.5142.

[99] P. Golonka and Z. Was. PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays. *Eur. Phys. J. C*, 45:97–107, 2006. hep-ph/0506026.

[100] M. Bowler. e^+ e^- Production of Heavy Quarks in the String Model. *Z. Phys. C11 (1981) 169*, 1981.

[101] D. Amati and G. Veneziano. Preconfinement as a property of perturbative QCD. *Physics Letters B*, 83(1):87–92, 1979.

[102] ATLAS Collaboration. ATLAS tunes of PYTHIA6 and PYTHIA8 for MC11. *ATL-PHYS-PUB-2011-009*, 2011.

[103] ATLAS Collaboration. New ATLAS event generator tunes to 2010 data. *ATL-PHYS-PUB-2011-008*, 2011.

[104] P. Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys.Rev.D82:074018,2010*, 2010. arXiv:1005.3457.

[105] Z. Was. TAUOLA for simulation of tau decay and production: perspectives for precision low energy and LHC applications. *Nucl.Phys.Proc.Suppl.218:249-255,2011*, 2011. arXiv:1101.1652.

[106] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56–74, 1998. hep-ph/9704448.

[107] A. D. Martin *et al.*. Parton distributions for the LHC. *Eur. Phys. J.*, 63:189–285, 2009. arXiv:0901.0002.

[108] GEANT4 Collaboration. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A*, pages 250–303, 2003.

[109] A. Ribon *et al.*. Status of GEANT4 hadronic physics for the simulation of LHC experiments at the start of LHC physics program. *CERN-LCGAPP-2010-02*, 2010.

[110] D.H. Wright. Geant4 User's documentation: Physics Reference Manual. `http://geant4.web.cern.ch/geant4/G4UsersDocuments/UsersGuides/PhysicsReferenceManual/html/PhysicsReferenceManual.html`, 2005. 4.12.01.

[111] ATLAS Collaboration. Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles. *ATL-SOFT-PUB-2014-01*, 2014.

[112] G. P. Salam M. Cacciari. The Catchment Area of Jets. *arXiv:0802.1188 [hep-ph]*, 2008.

[113] ATLAS Collaboration. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *arXiv:1510.03823 [hep-ex]*, 2015.

[114] ATLAS Collaboration. Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS . *ATLAS-CONF-2015-037*, 2015.

[115] ATLAS Collaboration. Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J. C, 73 3 (2013) 2304*, 2013. arXiv:1112.6426 [hep-ex].

[116] A. Sherstnev and R.S. Thorne. Parton Distributions for LO Generators. *Eur.Phys.J.C55:553-575*, 2008. arXiv:0711.2473 [hep-ph].

[117] ATLAS Collaboration. Determination of the jet energy scale and resolution at ATLAS using Z/γ|jet events in data at $\sqrt{s} = 8$ TeV. *ATLAS-CONF-2015-057*, 2015.

[118] ATLAS Collaboration. Data driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at $\sqrt{s} = 8$ TeV. *ATLAS-CONF-2015-017*, 2015.

[119] S. Gieseke *et al.*. Colour reconnections in Herwig++. *Eur.Phys.J. C72 (2012) 2225*, 2012. arXiv:1206.0041.

[120] The ATLAS collaboration. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Phys.Rev.D83:052005,2011*, 2011. arXiv:1012.4389.

[121] ATLAS Collaboration. Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Eur. Phys. J. C (2014) 74: 3023*, 2014. arXiv:1405.6583 [hep-ex].

[122] ATLAS Collaboration. Jet energy scale and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS 2011 data. *ATLAS-CONF-2013-004*, 2013.

[123] P. Adragna *et al.*. Testbeam Studies of Production Modules of the ATLAS Tile Calorimeter. *NIM A 606 (2009) 362-394*, 2009. ATL-COM-TILECAL-2009-002.

[124] M. Aharrouche *et al.*. Energy linearity and resolution of the ATLAS electromagnetic barrel calorimeter in an electron test-beam. *NIM A 568 (2006) 601-623*, 2006. arXiv:physics/0608012 [physics.ins-det].

[125] B. Di Girolamo *et al.*. Beamline instrumentation in the 2004 combined ATLAS testbeam. *TECH-PUB-2005-001*, 2005.

[126] ATLAS Collaboration. Response and Shower Topology of 2 to 180 GeV Pions Measured with the ATLAS Barrel Calorimeter at the CERN Test-beam and Comparison to Monte Carlo Simulations. *ATL-CAL-PUB-2010-001*, 2010.

[127] E. Abat *et al.*. Study of energy response and resolution of the ATLAS barrel calorimeter to hadrons of energies from 20 to 350 GeV. *NIM A 611*, pages 134–150, 2010.

[128] ATLAS Collaboration. Single hadron response measurement and calorimeter jet energy scale uncertainty with the ATLAS detector at the LHC. *Eur. Phys. J. C, 73 3 (2013) 2305*, 2013. arXiv:1203.1302 [hep-ex].

[129] M. J. Sousa. Single hadron response measurements in ATLAS. Poster, 2012. 12th Pisa Meeting on Advanced Detectors.

[130] The ATLAS Collaboration. A measurement of the calorimeter response to single hadrons and determination of the jet energy scale uncertainty using LHC Run-1 *pp*-collision data with the ATLAS detector. *arXiv:1607.08842v1*, 2016.

[131] M. J. Sousa. Single hadron response measurements in ATLAS and evaluation of the uncertainty on the jet energy measurement. In *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 718, pages 112 – 114, 2013. Proceedings of the 12th Pisa Meeting on Advanced Detectors, La Biodola, Isola d'lba, Italy, May 20-26, 2012, ATL-PHYS-PROC-2014-077.

[132] ATLAS Collaboration. Jet global corrections with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *ATLAS-CONF-2015-002*, 2015.

[133] ATLAS Collaboration. Jet energy measurement and systematic uncertainties using tracks for jets and for b-quark jets produced in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector. *ATLAS-CONF-2013-002*, 2013.

[134] The ATLAS collaboration. A study of the sensitivity to the Pythia8 parton shower parameters of $t\bar{t}$ production measurements in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS experiment at the LHC. *ATL-PHYS-PUB-2015-007*, 2015. http://cds.cern.ch/record/2004362.

[135] private communication.

[136] A. D. Bukin. Fitting function for asymmetric peaks, 2007.

[137] ATLAS Collaboration. Data Quality Information. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010`.

[138] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, 2012. arXiv:1201.3084 [hep-ph].

[139] M. Botje *et al.*. The PDF4LHC Working Group Interim Recommendations. *arXiv:1101.0538 [hep-ph]*, 2011.

[140] A.D. Martin *et al.*. Parton distributions for the LHC. *Eur.Phys.J.C63:189-285,2009*, 2009. 0901.0002v3.

[141] A. D. Martin *et al.*. Uncertainties on α_s in global pdf analyses and implications for predicted hadronic cross sections. *Eur. Phys. J.C64 (2009) 653-680*, 2009. arXiv:0905.3531 [hep-ph].

[142] J. Gao *et al.*. The ct10 nnlo global analysis of qcd. *Phys. Rev. D 89, 033009 (2014)*, 2014. arXiv:1302.6246 [hep-ph].

[143] NNPDF Collaboration. Parton distributions with LHC data. *Nucl. Phys. B867, 244*, 2013. arXiv:1207.1303.

[144] M. Fiascaris, on behalf of the ATLAS Collaboration. Performance of the ATLAS Tile Calorimeter in pp collisions at the LHC. *Journal of Physics: Conference Series 587 (2015) 012010*, 2015.

[145] A. Denner, S. Dittmaier, and A. Mück. HAWK: A Monte Carlo generator for the production of Higgs bosons Attached to WeaK bosons at hadron colliders. `http://omnibus.uni-freiburg.de/~sd565/programs/hawk/hawk.html`, 2010.

[146] The ATLAS collaboration. Measurements of normalized differential cross-sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *CERN-PH-EP-2014-099*, 2014. arXiv:1407.0371v3.

[147] The ATLAS collaboration. Measurement of the Top-Quark Mass using the Template Method in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2011-033, CERN, 2011.

[148] K. Cranmer *et al.*. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. *CERN-OPEN-2012-016*, 2012.

[149] H1 and ZEUS collaborations. Combined measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA. *JHEP 1001:109,2010*, 2010. arXiv:0911.0884.

[150] S. Frixione and B. R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP 0206:029,2002*, 2002. arXiv:hep-ph/0204244.

[151] R. Pedro. *Search for the Higgs boson at ATLAS/LHC in WH associated production and decay to b-quark pairs*. PhD thesis, Faculdade de Ciências da Universidade de Lisboa, 2017.

[152] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[153] Y. Freund and R. Schapire. Experiments with a new boosting algorithm, in Machine Learning. *Proceedings of the Thirteenth International Conference (ICML)*, 1996. Morgan Kaufmann.

[154] A. Hoecker *et al.*. TMVA - Toolkit for Multivariate Data Analysis. *PoS ACAT:040,2007*, 2007. arXiv:physics/0703039v5.

[155] The CMS Collaboration. Search for the Standard Model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks. *Phys. Rev. D 89, 012003 (2014)*, 2014. arXiv:1310.3687.

[156] ATLAS Collaboration. Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector. *ATLAS-CONF-2016-091*, 2016.

Appendices

A. Appendix E/p

This appendix shows additional η-ranges for the neutral background corrected $\langle E/p \rangle$ for the random trigger and the two hadCalib triggers.

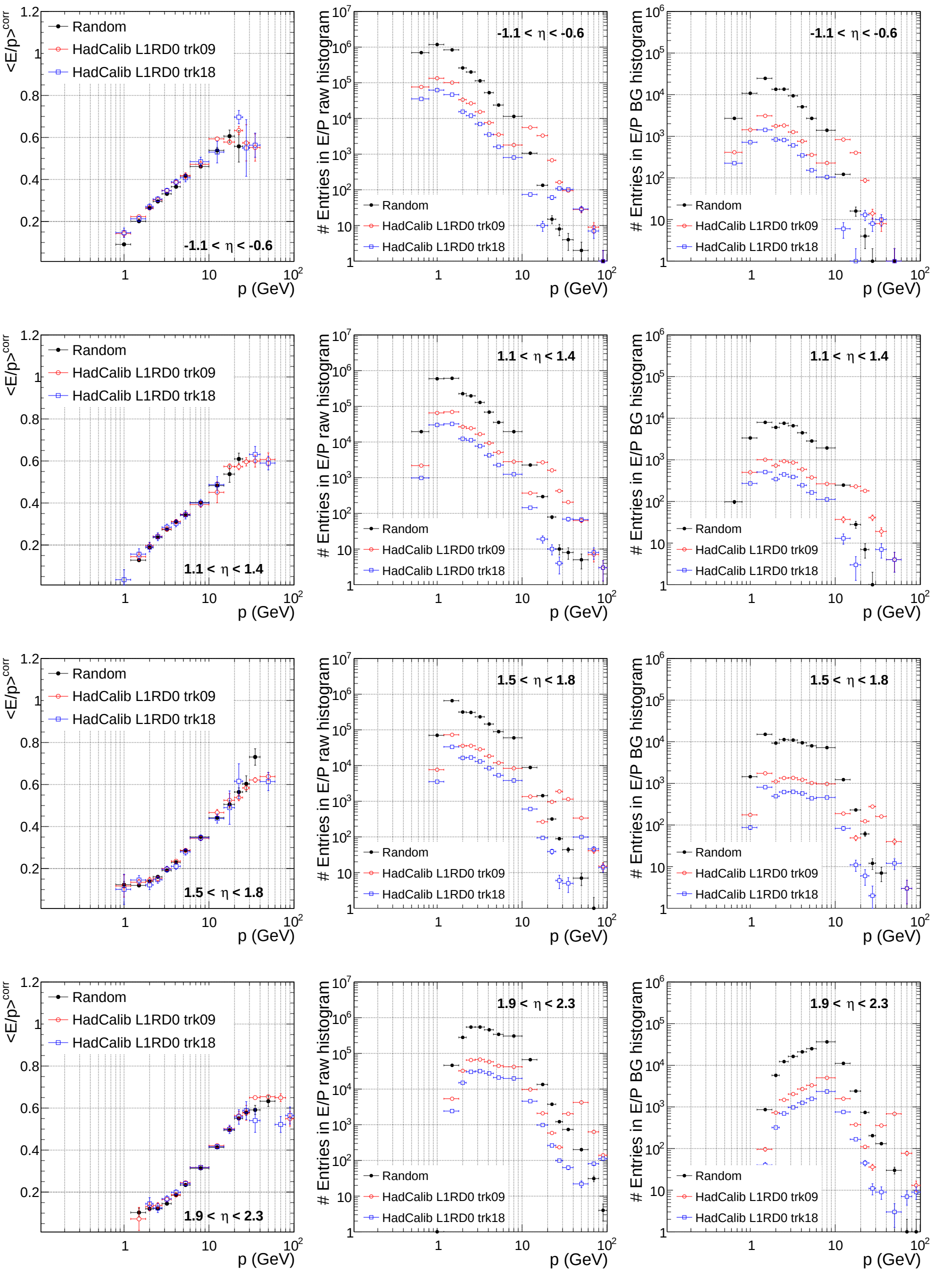


Figure A.1.: Corrected average of the E/p distribution for the random (in black), hadCalib_tr9_L1RD0 (in red) and hadCalib_tr18_L1RD0 (in blue) triggers as a function of the single hadron momentum (left). Statistics available for the $\langle E/p^{raw}\rangle$ (centre) and $\langle E/p^{BG}\rangle$ (right). See section 3.8.3 for further details.

B. Appendix GSC

This appendix shows additional transverse momentum and η-bins for the jet energy response and jet energy resolution, with respect to the figures discussed in section 3.9.

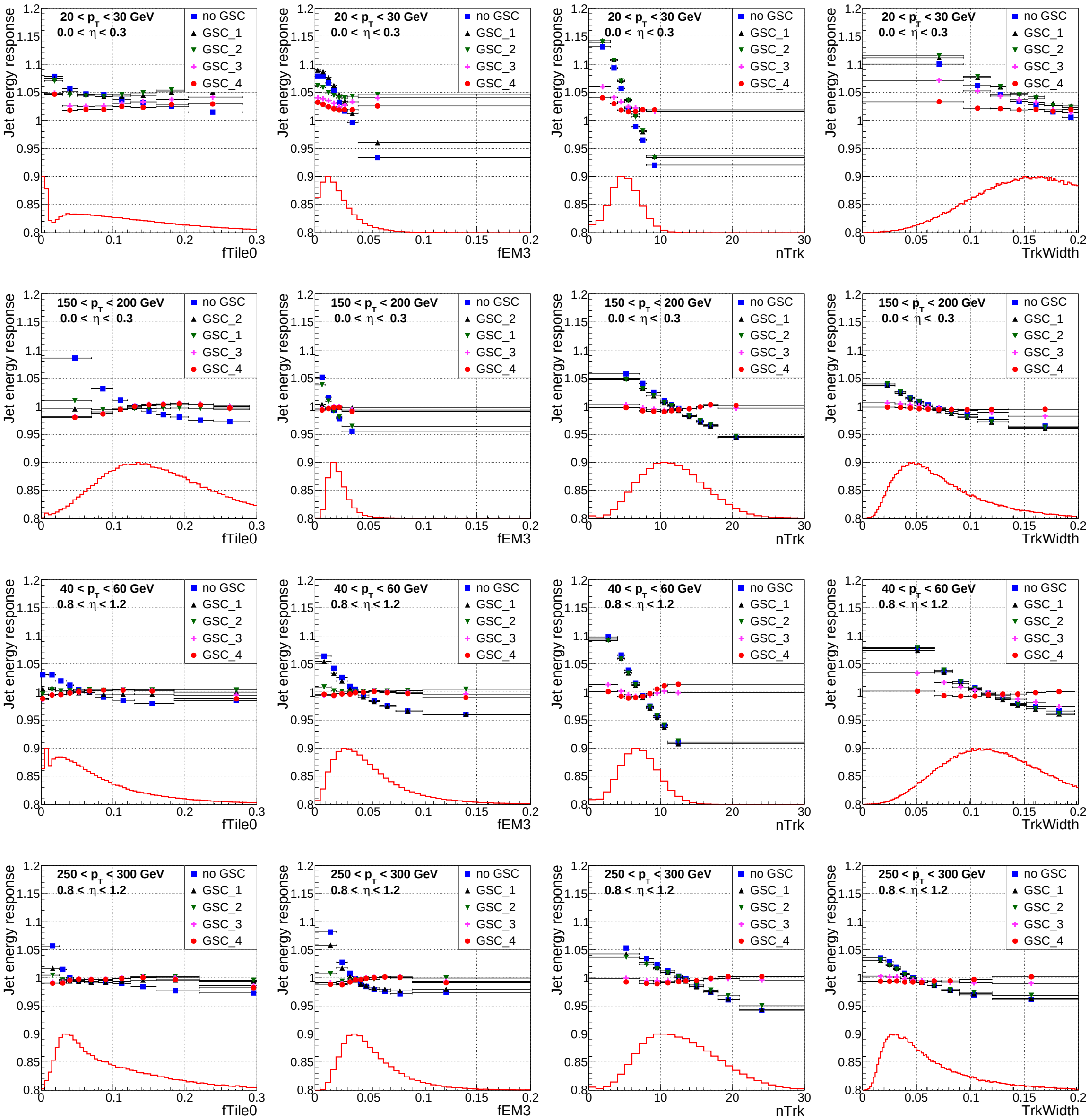


Figure B.1.: Jet energy response as a function of the GSC variables, f_{Tile0}, f_{em3}, $nTrk$ and $TrkWidth$, in the first, second, third and fourth columns, respectively. The blue squares show the jet energy response without applying GSC and the black up triangles, green down triangles, pink crosses and red circles show the jet energy response after applying the first (f_{Tile0}), second (f_{em3}), third ($nTrk$) and fourth ($TrkWidth$) GSC corrections. The histogram in red represents the distribution of each of these variables. Two additional p_T-bins for the $0.0 < \eta < 0.3$ (first two rows) and one additional η region: $0.8 < \eta < 1.2$ (last two rows). Further details are found in section 3.9.2.

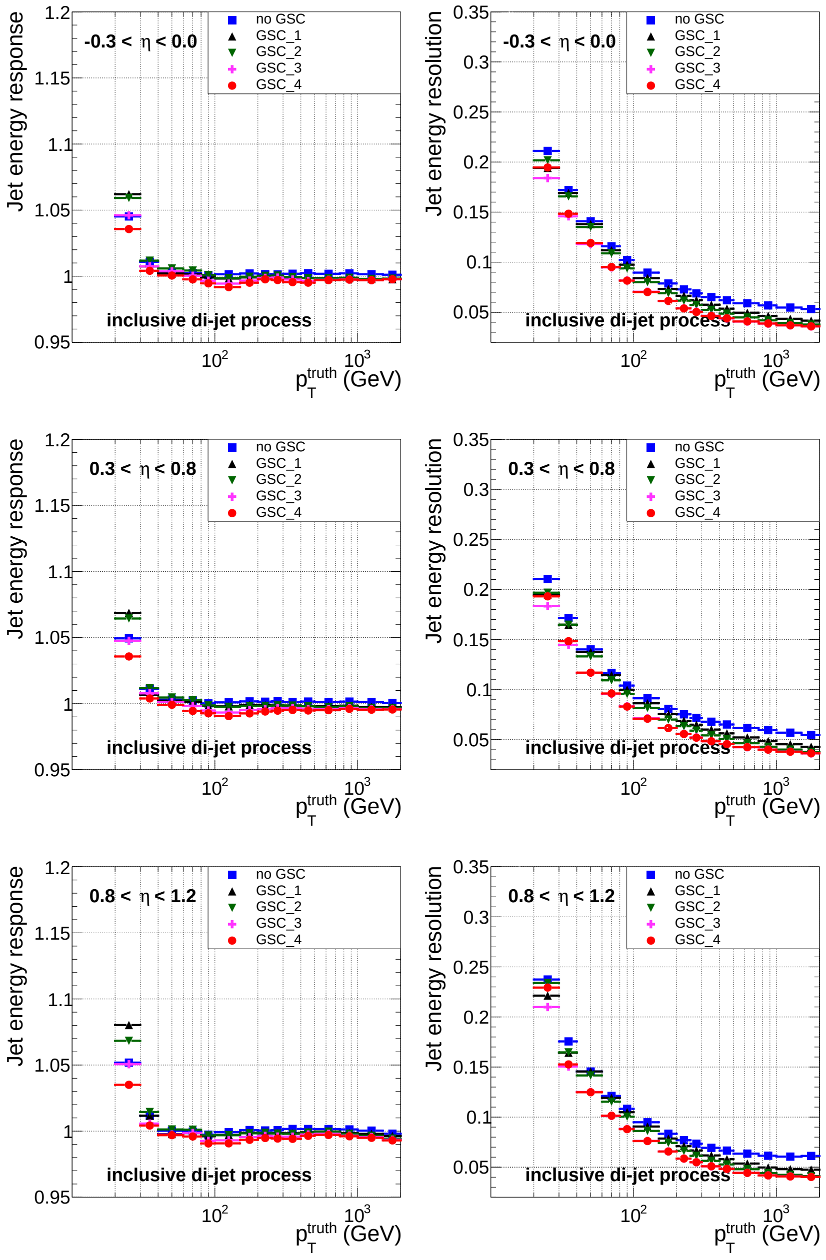


Figure B.2.: Jet energy response (left) and resolution (right) as a function of the truth jet transverse momentum for the inclusive di-jet events simulated with PYTHIA8 for $-0.3 < \eta < 0.0$ (top), $0.3 < \eta < 0.8$ (centre) and $0.8 < \eta < 1.2$ (bottom). The blue squares show the jet energy response and resolution without applying GSC and the black up triangles, green down triangles, pink crosses and red circles show the jet energy response after applying the first (fTile0), second (fEM3), third (nTrk) and fourth (*TrkWidth*) GSC corrections. Further details are found in section 3.9.2.

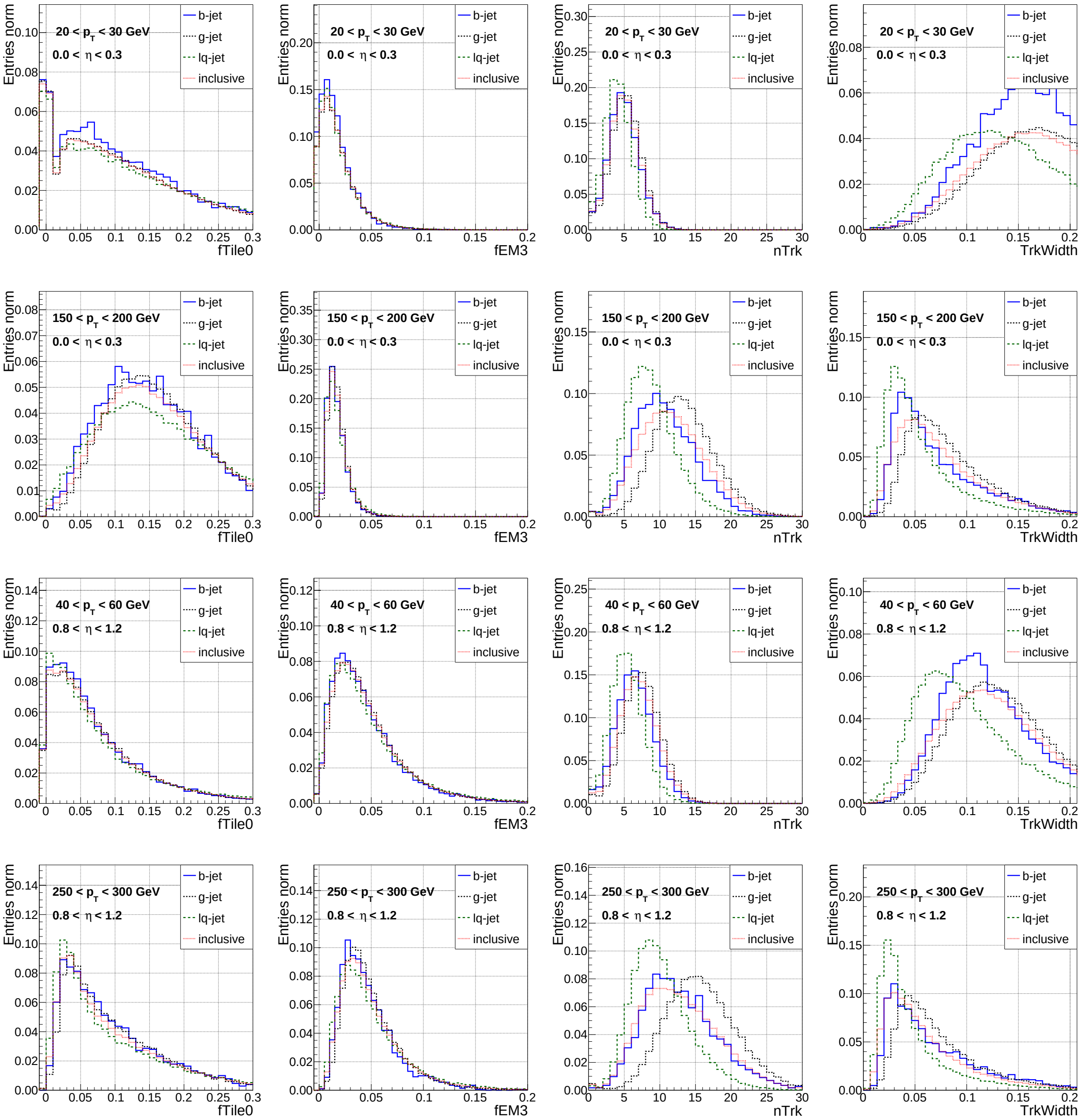


Figure B.3.: Distributions of f_{Tile0} (first column), f_{em3} (second column), $nTrk$ (third column) and $TrkWidth$ (fourth column), normalised to unity, for different jet flavour compositions of di-jet events simulated with PYTHIA8. The b-jet flavour is shown in blue, gluon initiated jets in black, light-quark initiated jets in green and the inclusive jet flavour in red, with different line styles. Two additional p_T^{truth}-bins for the $0.0 < \eta < 0.3$ (first two rows) and one additional η region: $0.8 < \eta < 1.2$ (last two rows). Further details are found in section 3.9.2.

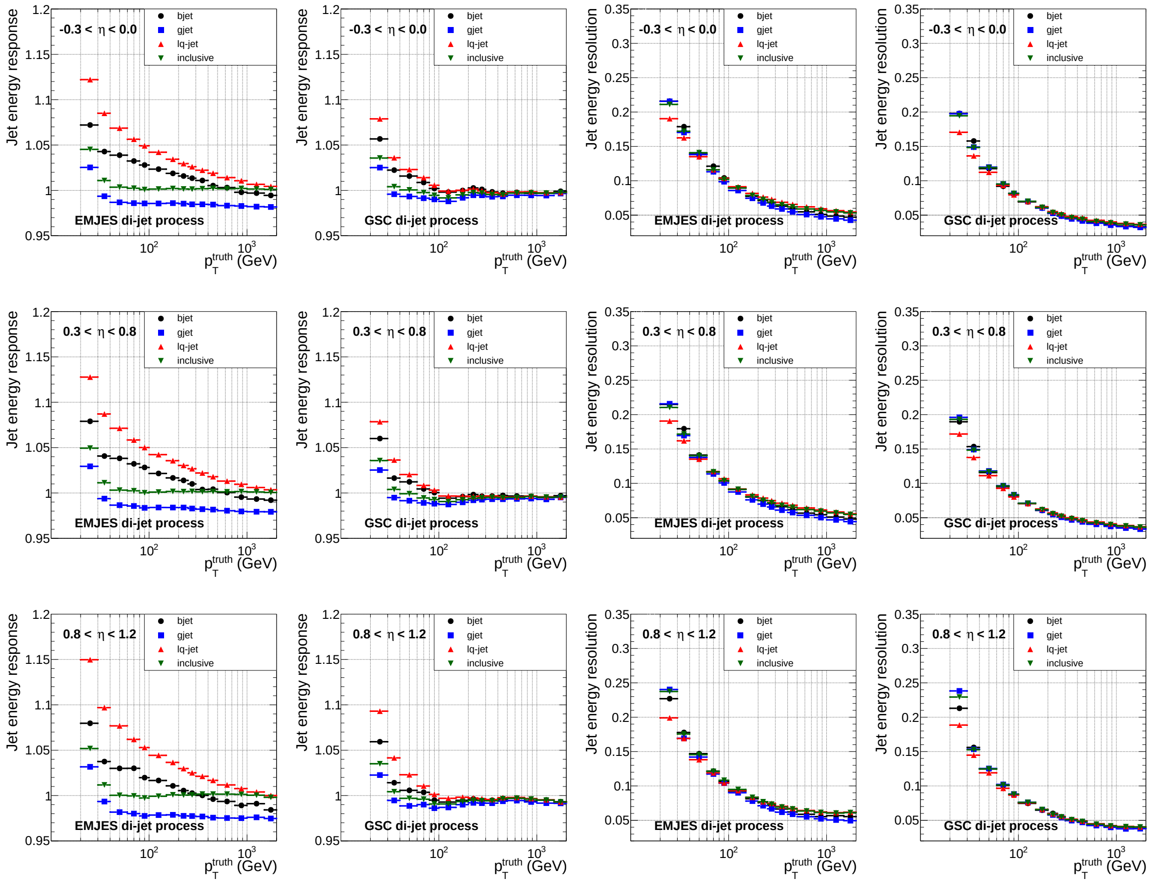


Figure B.4.: Jet energy response as a function of the jet p_T^{truth} for the different jet type samples. In the first (second) column the jets are calibrated with EM+JES (GSC). The third and forth columns are the resolutions as a function of the jet p_T^{truth} for jets calibrated at EM+JES and GSC, respectively. Jet energy response and resolution obtained for the di-jet simulated sample with PYTHIA8 separating light-quark initiated jets (red up triangles) gluon initiated jets (blue squares), b-jets (black circles) and inclusive jets (green down triangles). The plots are shown for $-0.3 < \eta < 0.0$ (top), $0.3 < \eta < 0.8$ (centre) and $0.8 < \eta < 1.2$ (bottom). Further details are found in section 3.9.3.

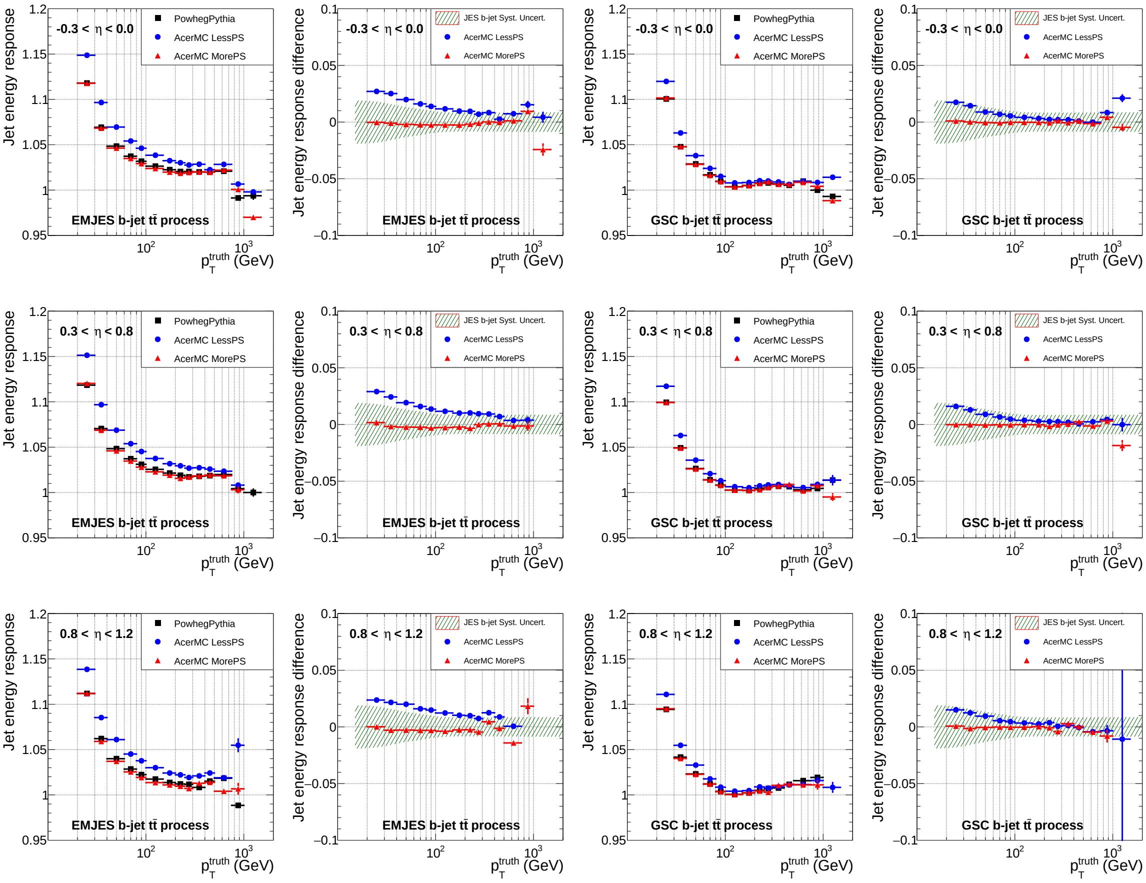


Figure B.5.: b-jet energy response and response difference with EM+JES (two left columns) and GSC (two right columns), for jets with $-0.3 < \eta < 0.0$ (top row), $0.3 < \eta < 0.8$ (middle row) and $0.8 < \eta < 1.2$ (bottom row), with the $t\bar{t}$ process simulated with POWHEG+PYTHIA6 (black squares) and with ACERMC+PYTHIA with more parton shower (blue circles) and less parton shower (red triangles). The jet energy response differences are done with respect to the POWHEG+PYTHIA6 sample. The green band represents the b-jet systematic uncertainty obtained for the EM+JES calibration. Further details are found in section 3.9.4

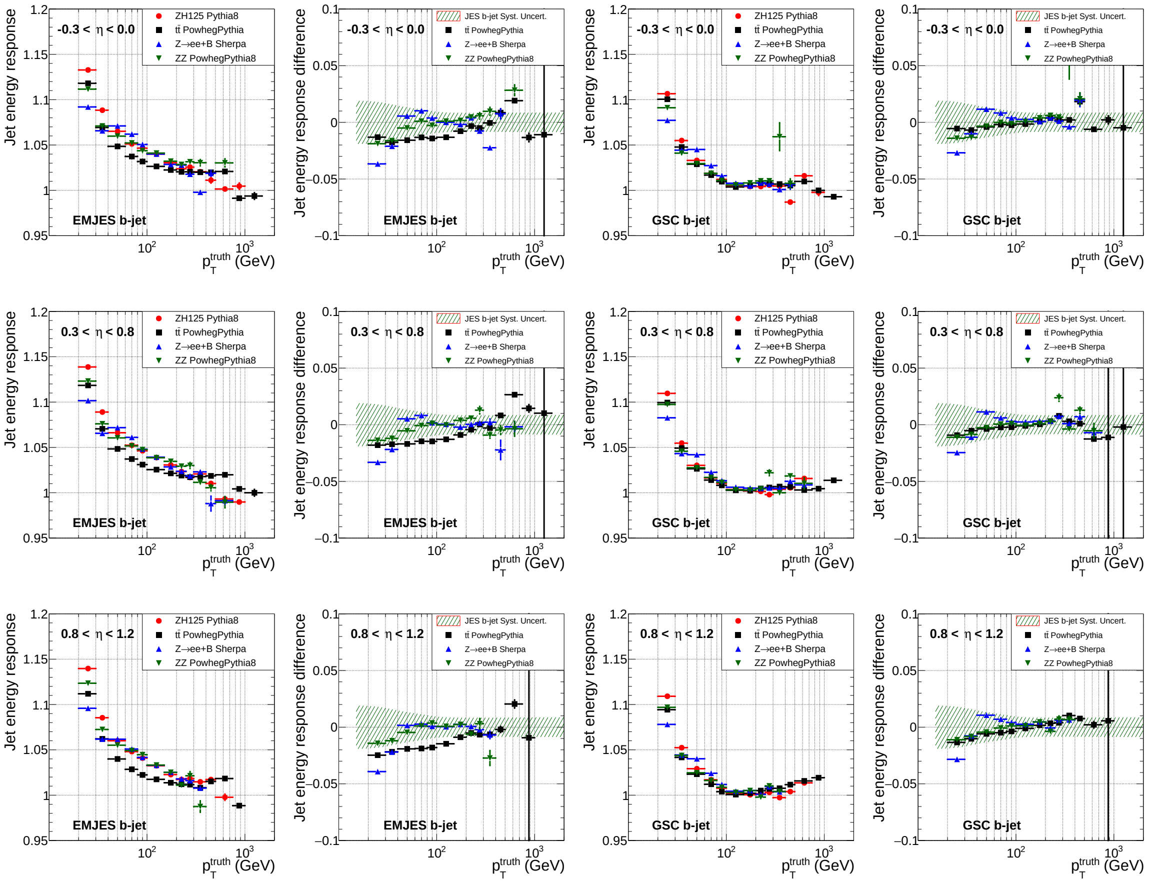


Figure B.6.: b-jet energy response and response difference with EM+JES (two left columns) and GSC (two right columns), for jets with $-0.3 < \eta < 0.0$ (top row), $0.3 < \eta < 0.8$ (middle row) and $0.8 < \eta < 1.2$ (bottom row), for the signal process for the ZH analysis (red circles) and for the three leading backgrounds: $t\bar{t}$ (black squares), Z+b-jets (blue up triangles), and ZZ (green down triangles). The jet energy response differences are done with respect to ZH process. Further details are found in section 3.9.4

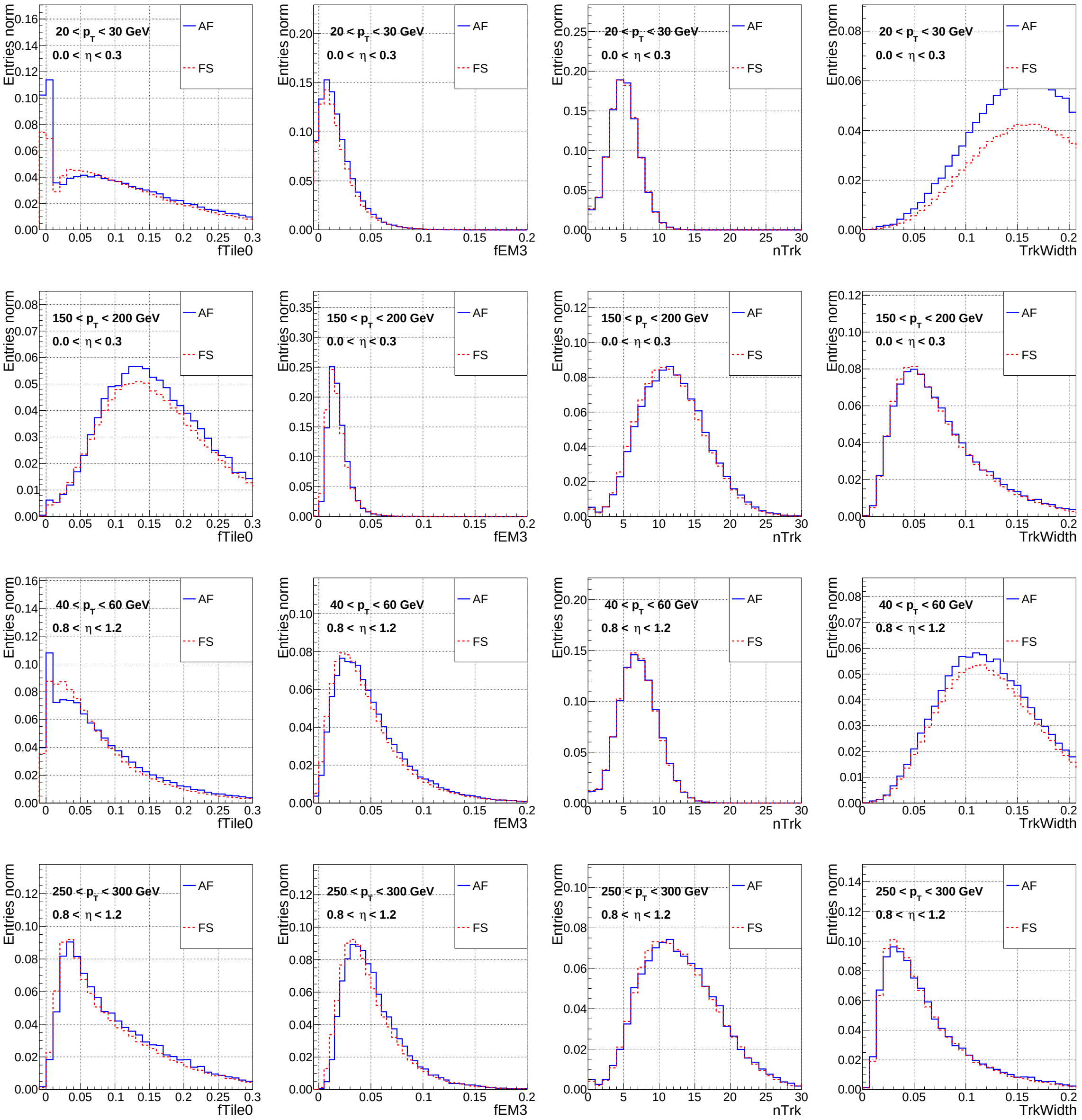


Figure B.7.: Distributions of f_{Tile0} (first column), f_{em3} (second column), $nTrk$ (third column) and $TrkWidth$ (fourth column), normalised to unity, for the inclusive di-jet events simulated with PYTHIA8 using the full simulation (filled blue line) and fast simulation (dashed red line) of the detector Two additional p_T-bins for the $0.0 < \eta < 0.3$ (first two rows) and one additional η region, $0.8 < \eta < 1.2$ (last two rows). Further details are found in section 3.9.4.2.

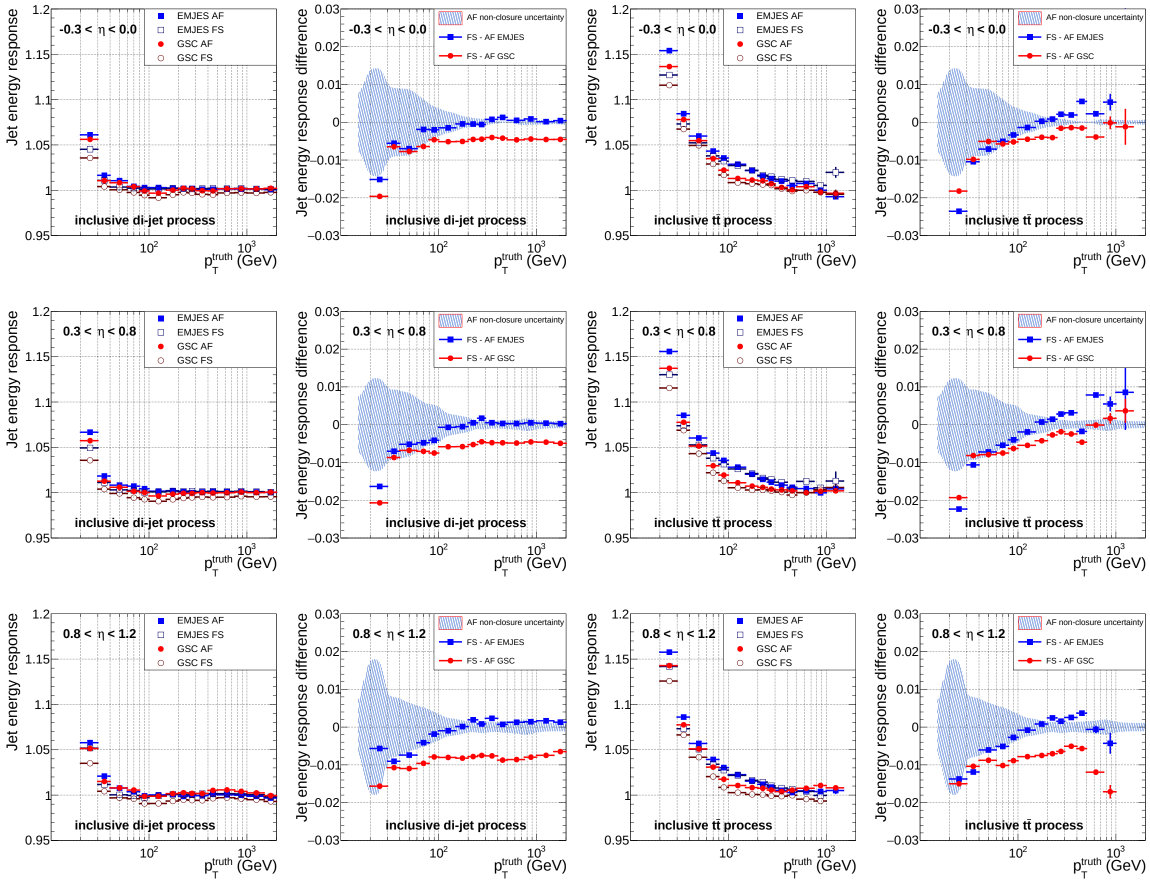


Figure B.8.: Jet energy response for the inclusive di-jet events simulated with Pythia8 (first column) and the inclusive $t\bar{t}$ sample simulated with Powheg+Pythia6 (third column). Jet energy response difference between the parametrised and the full simulation as a function of the p_T^{truth} for the di-jet sample (second column) and the $t\bar{t}$ sample (fourth column). Jets in the pseudorapidity region $-0.3 < \eta < 0.0$ (top row), $0.3 < \eta < 0.8$ (middle row) and $0.8 < \eta < 1.2$ (bottom row). Comparison between the AtlasFastII (closed markers) and the full simulation (open markers) of the detector. In all figure, the jet energy response for the EM+JES (GSC) calibration is shown by blue squares (red circles). The shaded blue band in the jet response difference plots corresponds to the measured systematic uncertainty obtained for the AtlasFastII non-closure with the EM+JES calibration. Further details are found in section 3.9.4.2.

C. Appendix ZH

C.1. Information on simulated samples

In this section, additional informations of the simulated samples used in the ZH analysis are given. Samples that have the * in the process refer to those for which there was no AFII fast simulation available and the full simulation of the detector, discussed in secion 2.8 was used instead.

ZH signal with Higgs mass of 125 GeV

MCid	Process	Nevents	$\sigma\times$BR	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
161827	qqZH($\to \ell\ell b\bar{b}$)	2998998	2.231×10^{-2}	1	2.920×10^{6}	1.309×10^{5}
189340	ggZH($\to eeb\bar{b}$)	100000	6.300×10^{-1}	1	64.72	102.7
189341	ggZH($\to \mu\mu b\bar{b}$)	100000	6.304×10^{-1}	1	64.46	102.5
189342	ggZH($\to \tau\tau b\bar{b}$)	100000	6.312×10^{-1}	1	64.83	102.7

Top backgrounds

MCid	Process	Nevents	$\sigma\times$BR	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
117050	$t\bar{t}$($\to \ell\nu b\ell\nu b$)	99930891	253.0	0.543	9.951×10^{7}	724.7
110101	single-t (t channel)	8996990	87.80	0.324	7.672×10^{6}	269.8
110119	single-t (s channel)	5995993	5.610	0.324	5.971×10^{6}	3284
110140	single-t (Wt channel)	19937980	22.40	1	1.985×10^{8}	887.4

Diboson backgrounds

MCid	Process	Nevents	$\sigma\times$BR	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
181966	ZZ($\to \ell\ell hh$)	3999995	1.207	1	4.561×10^{6}	3779
181968	WZ($\to hh\ell\ell$)	1500000	1.594	1	2.382×10^{6}	1494
181970	WZ($\to \ell\nu hh$)	9999988	4.870	1	4.849×10^{7}	9957
181971	WW	9999994	52.44	1	5.081×10^{7}	9690

Table C.1.: List of the samples for signal, and the top and diboson backgrounds with the values for MCid, number of generated events, $\sigma\times$ BR, filter efficiency, $\sum w^i$ and $MClumi$ calculation. Further details are given in section 4.4.

$p_T^Z < 40$ GeV: $\sigma\times$BR= 1242 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
167749	$\mathrm{Z}(\rightarrow ee)+b$	3999000	0.02797	1.092×10^6	31.43
167750	$\mathrm{Z}(\rightarrow ee)+c$	2999995	0.2832	1.433×10^6	4.068
167751	$\mathrm{Z}(\rightarrow ee)+l$	4978999	0.6886	2.820×10^6	3.302
167752	$\mathrm{Z}(\rightarrow \mu\mu)+b$	3997997	0.02796	1.091×10^6	31.40
167753	$\mathrm{Z}(\rightarrow \mu\mu)+c$	2937995	0.2835	1.405×10^6	3.988
167754	$\mathrm{Z}(\rightarrow \mu\mu)+l$	4993999	0.68872	2.827×10^6	3.311
167755	$\mathrm{Z}(\rightarrow \tau\tau)+b$	3997994	0.02794	1.091×10^6	31.40
167756	$\mathrm{Z}(\rightarrow \tau\tau)+c$	2998998	0.2833	1.434×10^6	4.070
167757	$\mathrm{Z}(\rightarrow \tau\tau)+l$	4989999	0.6889	2.825×10^6	3.309

$40 < p_T^Z < 70$ GeV: $\sigma\times$BR=78.95 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
180543	$\mathrm{Z}(\rightarrow ee)+b$	1199999	0.07067	2.636×10^5	47.24
180544	$\mathrm{Z}(\rightarrow ee)+c$	600000	0.3420	1.580×10^5	5.856
180545	$\mathrm{Z}(\rightarrow ee)+l$	1399998	0.5876	3.958×10^5	8.531
180546	$\mathrm{Z}(\rightarrow \mu\mu)+b$	1199000	0.07067	2.637×10^5	47.25
180547	$\mathrm{Z}(\rightarrow \mu\mu)+c$	599000	0.3414	1.576×10^5	5.839
180548	$\mathrm{Z}(\rightarrow \mu\mu)+l$	1398999	0.5877	3.955×10^5	8.525
180549	$\mathrm{Z}(\rightarrow \tau\tau)+b$	1198999	0.07078	2.637×10^5	47.24
180550	$\mathrm{Z}(\rightarrow \tau\tau)+c$	600000	0.3416	1.579×10^5	5.852
180551	$\mathrm{Z}(\rightarrow \tau\tau)+l$	1399996	0.5876	3.954×10^5	8.523

$70 < p_T^Z < 140$ GeV: $\sigma\times$BR=33.03 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
167797	$\mathrm{Z}(\rightarrow ee)+b$	1366999	0.08252	2.891×10^5	106.0
167798	$\mathrm{Z}(\rightarrow ee)+c$	999999	0.3550	2.389×10^5	20.38
167799	$\mathrm{Z}(\rightarrow ee)+l$	1999998	0.5626	5.079×10^5	27.34
167800	$\mathrm{Z}(\rightarrow \mu\mu)+b$	1394999	0.08259	2.952×10^5	108.3
167801	$\mathrm{Z}(\rightarrow \mu\mu)+c$	1000000	0.3549	2.391×10^5	20.40
167802	$\mathrm{Z}(\rightarrow \mu\mu)+l$	1996998	0.5620	5.071×10^5	27.29
167803	$\mathrm{Z}(\rightarrow \tau\tau)+b*$	1399396	0.08256	2.961×10^5	108.6
167804	$\mathrm{Z}(\rightarrow \tau\tau)+c*$	999998	0.3551	2.392×10^5	20.41
167805	$\mathrm{Z}(\rightarrow \tau\tau)+l*$	1969693	0.5625	4.999×10^5	26.90

Table C.2.: List of the samples for the three lower $\mathrm{p_T^Z}$ slices of the Z+jets background with the values for MCid, number of generated events, $\sigma\times$ BR, filter efficiency, $\sum w^i$ and *MClumi* calculation. Further details are given in section 4.4.

$140 < p_T^Z < 280$ GeV: $\sigma\times$BR=4.447 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\text{fb}^{-1})$
167809	$Z(\rightarrow ee)+b$	999999	0.09526	2.055×10^5	482.0
167810	$Z(\rightarrow ee)+c$	399999	0.3692	8.881×10^4	53.82
167811	$Z(\rightarrow ee)+l$	600000	0.5343	1.395×10^5	58.50
167812	$Z(\rightarrow \mu\mu)+b$	987999	0.09532	2.031×10^5	476.4
167813	$Z(\rightarrow \mu\mu)+c$	399000	0.3700	8.848×10^4	53.62
167814	$Z(\rightarrow \mu\mu)+l$	599500	0.5344	1.392×10^5	58.36
167815	$Z(\rightarrow \tau\tau)+b$	798998	0.09546	1.642×10^5	385.3
167816	$Z(\rightarrow \tau\tau)+c*$	399999	0.3695	8.865×10^4	53.72
167817	$Z(\rightarrow \tau\tau)+l*$	598897	0.5333	1.389×10^5	58.26

$280 < p_T^Z < 500$ GeV: $\sigma\times$BR=0.2707 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\text{fb}^{-1})$
167821	$Z(\rightarrow ee)+b$	180000	0.1071	3.639×10^4	1253
167822	$Z(\rightarrow ee)+c*$	49899	0.3874	1.059×10^4	101.4
167823	$Z(\rightarrow ee)+l*$	49999	0.5062	1.092×10^4	79.63
167824	$Z(\rightarrow \mu\mu)+b$	175000	0.1073	3.542×10^4	1220
167825	$Z(\rightarrow \mu\mu)+c*$	50000	0.3864	1.060×10^4	101.6
167826	$Z(\rightarrow \mu\mu)+l*$	50000	0.5055	1.095×10^4	79.79
167827	$Z(\rightarrow \tau\tau)+b$	180000	0.1072	3.645×10^4	1255
167828	$Z(\rightarrow \tau\tau)+c*$	50000	0.3848	1.057×10^4	101.2
167829	$Z(\rightarrow \tau\tau)+l*$	49899	0.5072	1.088×10^4	79.31

$p_T^Z > 500$ GeV: $\sigma\times$BR=0.01490 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\text{fb}^{-1})$
167833	$Z(\rightarrow ee)+b$	90000	0.1152	1.810×10^4	1.051×10^4
167834	$Z(\rightarrow ee)+c*$	10000	0.3985	2067	346.7
167835	$Z(\rightarrow ee)+l*$	50000	0.4848	1.062×10^4	1470
167836	$Z(\rightarrow \mu\mu)+b$	100000	0.1161	2.012×10^4	1.169×10^4
167837	$Z(\rightarrow \mu\mu)+c$	10000	0.3986	2073	347.8
167838	$Z(\rightarrow \mu\mu)+l$	10000	0.4869	2129	294.5
167839	$Z(\rightarrow \tau\tau)+b$	90000	0.1163	1.812×10^4	1.052×10^4
167840	$Z(\rightarrow \tau\tau)+c*$	10000	0.3932	2060	345.7
167841	$Z(\rightarrow \tau\tau)+l*$	149900	0.4856	3.174×10^4	4391

Table C.3.: List of the samples for the three higher $\mathrm{p_T^Z}$ slices of the Z+jets background with the values for MCid, number of generated events, $\sigma\times$ BR, filter efficiency, $\sum w^i$ and *MClumi* calculation. Further details are given in section 4.4.

$p_T^W < 40$ GeV: $\sigma\times$BR=1.207×10^5 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
167740	$\mathrm{W}(\rightarrow e\nu)+b$	14997980	0.01279	4.342×10^6	28.12
167741	$\mathrm{W}(\rightarrow e\nu)+c$	9998989	0.04900	3.036×10^6	5.481
167742	$\mathrm{W}(\rightarrow e\nu)+l$	49885967	0.9380	2.7867×10^7	2.452
167743	$\mathrm{W}(\rightarrow \mu\nu)+b$	14989485	0.01280	4.318×10^6	27.97
167744	$\mathrm{W}(\rightarrow \mu\nu)+c$	9992484	0.04251	2.864×10^6	5.582
167745	$\mathrm{W}(\rightarrow \mu\nu)+l$	49846965	0.9446	2.7820×10^7	2.448
167746	$\mathrm{W}(\rightarrow \tau\nu)+b$	14925982	0.01279	4.323×10^6	28.00
167747	$\mathrm{W}(\rightarrow \tau\nu)+c$	9993984	0.04615	2.960×10^6	5.344
167748	$\mathrm{W}(\rightarrow \tau\nu)+l$	49920968	0.9409	2.7875×10^7	2.453

$40 < p_T^W < 70$ GeV: $\sigma\times$BR=718.2 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
180534	$\mathrm{W}(\rightarrow e\nu)+b$	2999998	0.03455	6.740×10^5	27.17
180535	$\mathrm{W}(\rightarrow e\nu)+c$	4499994	0.1718	1.140×10^6	9.400
180536	$\mathrm{W}(\rightarrow e\nu)+l$	16997491	0.7934	4.774×10^6	8.344
180537	$\mathrm{W}(\rightarrow \mu\nu)+b$	2996996	0.03455	6.732×10^5	27.13
180538	$\mathrm{W}(\rightarrow \mu\nu)+c$	4498998	0.1658	1.137×10^6	9.379
180539	$\mathrm{W}(\rightarrow \mu\nu)+l$	16988984	0.7998	4.771×10^6	8.339
180540	$\mathrm{W}(\rightarrow \tau\nu)+b$	2998997	0.03455	6.738×10^5	27.16
180541	$\mathrm{W}(\rightarrow \tau\nu)+c$	4498999	0.1692	1.138×10^6	9.382
180542	$\mathrm{W}(\rightarrow \tau\nu)+l$	16996492	0.7962	4.776×10^6	8.347

$70 < p_T^W < 140$ GeV: $\sigma\times$BR=275.7 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi$(fb^{-1})
167761	$\mathrm{W}(\rightarrow e\nu)+b$	2000000	0.04593	4.260×10^5	33.65
167762	$\mathrm{W}(\rightarrow e\nu)+c$	2996497	0.2010	6.893×10^5	12.61
167763	$\mathrm{W}(\rightarrow e\nu)+l$	4998998	0.7527	1.265×10^6	6.074
167764	$\mathrm{W}(\rightarrow \mu\nu)+b$	1998999	0.04592	4.257×10^5	33.63
167765	$\mathrm{W}(\rightarrow \mu\nu)+c$	2995999	0.1951	6.877×10^5	12.58
167766	$\mathrm{W}(\rightarrow \mu\nu)+l$	4998992	0.7587	1.266×10^6	6.079
167767	$\mathrm{W}(\rightarrow \tau\nu)+b*$	1999893	0.04594	4.262×10^5	33.67
167768	$\mathrm{W}(\rightarrow \tau\nu)+c*$	2999890	0.1989	6.894×10^5	12.61
167769	$\mathrm{W}(\rightarrow \tau\nu)+l*$	4999786	0.7548	1.266×10^6	6.080

Table C.4.: List of the samples for the three lower $\mathrm{p_T^W}$ slices of the W+jets background with the values for MCid, number of generated events, $\sigma\times$ BR, filter efficiency, $\sum w^i$ and *MClumi* calculation. Further details are given in section 4.4.

$140 < p_T^W < 280$ GeV: $\sigma\times$BR=34.29 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\mathrm{fb}^{-1})$
167770	$\mathrm{W}(\to e\nu)+b$	4998995	0.06314	1.030×10^6	475.6
167771	$\mathrm{W}(\to e\nu)+c$	1999997	0.2220	4.290×10^5	56.99
167772	$\mathrm{W}(\to e\nu)+l$	2000000	0.7150	4.636×10^5	18.87
167773	$\mathrm{W}(\to \mu\nu)+b$	4983993	0.0632	1.026×10^6	473.9
167774	$\mathrm{W}(\to \mu\nu)+c$	1995998	0.2165	4.274×10^5	56.79
167775	$\mathrm{W}(\to \mu\nu)+l$	1993999	0.7203	4.622×10^5	18.81
167776	$\mathrm{W}(\to \tau\nu)+b$	3998996	0.06317	8.237×10^5	380.4
167777	$\mathrm{W}(\to \tau\nu)+c*$	1998688	0.2202	4.286×10^5	56.94
167778	$\mathrm{W}(\to \tau\nu)+l*$	1999994	0.7161	4.641×10^5	18.88

$280 < p_T^W < 500$ GeV: $\sigma\times$BR=2.023 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\mathrm{fb}^{-1})$
167779	$\mathrm{W}(\to e\nu)+b$	899999	0.08299	1.823×10^5	1084
167780	$\mathrm{W}(\to e\nu)+c*$	199898	0.2345	4.134×10^4	88.13
167781	$\mathrm{W}(\to e\nu)+l*$	499891	0.6820	1.093×10^5	78.85
167782	$\mathrm{W}(\to \mu\nu)+b$	898000	0.08312	1.820×10^5	1083
167783	$\mathrm{W}(\to \mu\nu)+c*$	199998	0.2285	4.127×10^4	87.99
167784	$\mathrm{W}(\to \mu\nu)+l*$	499698	0.6878	1.093×10^5	78.88
167785	$\mathrm{W}(\to \tau\nu)+b$	898999	0.08305	1.820×10^5	1083
167786	$\mathrm{W}(\to \tau\nu)+c*$	199998	0.2327	4.125×10^4	87.96
167787	$\mathrm{W}(\to \tau\nu)+l*$	499998	0.68400	1.092×10^5	78.77

$p_T^W > 500$ GeV: $\sigma\times$BR=0.1118 pb					
MCid	Process	Nevents	ε_F	$\sum w^i$	$MClumi(\mathrm{fb}^{-1})$
167788	$\mathrm{W}(\to e\nu)+b$	100000	0.09952	2.011×10^4	1804
167789	$\mathrm{W}(\to e\nu)+c$	10000	0.2444	2034	75.36
167790	$\mathrm{W}(\to e\nu)+l$	10000	0.6574	2097	28.48
167791	$\mathrm{W}(\to \mu\nu)+b$	90000	0.09946	1.810×10^4	1624
167792	$\mathrm{W}(\to \mu\nu)+c*$	10000	0.2385	2013	74.60
167793	$\mathrm{W}(\to \mu\nu)+l*$	49700	0.6584	1.050×10^4	142.6
167794	$\mathrm{W}(\to \tau\nu)+b$	90000	0.09973	1.802×10^4	1616
167795	$\mathrm{W}(\to \tau\nu)+c*$	10000	0.2422	2030	75.22
167796	$\mathrm{W}(\to \tau\nu)+l*$	49998	0.6600	1.055×10^4	143.4

Table C.5.: List of the samples for the three higher $\mathrm{p_T^W}$ slices of the W+jets background with the values for MCid, number of generated events, $\sigma\times$ BR, filter efficiency, $\sum w^i$ and *MClumi* calculation. Further details are given in section 4.4.

C.2. ZH software validation with objects and regions.

The number of selected objects (muons, electrons and jets) is evaluated independently for each of the selections applied. The rows in **bold** show the number of events passing the combination of all the cuts above. Tables C.6, C.7 and C.8 present the validation of muon, electron and jet selections, respectively. The overlap removal is also validated and the results are shown in table C.9. Very few small differences are observed. After all the selections are applied, these differences are always well below one *permille*.

μ^{CB+ST} muons					μ^{SA+FW} muons				
Selection	LIP	A	B	C	Selection	LIP	A	B	C
None	308776	0	0	0	None	23856	0	0	0
Number of pixel hits	283019	0	0	0	Pseudorapidity	7926	0	0	0
Number of SCT hits	285751	0	0	0	$\mathrm{p_T}$	4778	13	0	0
Number of Si holes	288450	0	0	0	**Loose**	**2704**	**0**	**0**	**0**
Number of TRT hits	284841	0	0	0	μ^{Calo}muons				
Track quality	277351	0	0	0	None	256436	0	0	0
$\lvert d_0 \rvert$	189286	0	0	0	Track quality	284140	0	0	0
$\lvert z_0 \sin\theta \rvert$	226882	0	0	0	$\mathrm{p_T}$	103821	0	0	0
Pseudorapidity	288450	0	0	0	$\lvert d_0 \rvert$	220896	0	0	0
$\mathrm{p_T}$	152297	89	0	0	$\lvert z_0 \sin\theta \rvert$	198821	0	0	0
Track isolation	183454	-15	0	0	Pseudorapidity	13264	0	0	0
Loose	**97162**	**5**	**0**	**0**	Track isolation	244383	0	0	0
$p_T > 25$ GeV	87240	36	0	0	OR μ^{CB+ST}	229096	-11	-2	-1
Signal	**73467**	**-1**	**0**	**0**	**Loose**	**1653**	**-2**	**0**	**0**
Loose muons total (μ^{CB+ST}+μ^{SA+FW}+μ^{Calo})						101519	3	0	0

Table C.6.: Validation of the muon selection for the analysis performed in this thesis (labelled LIP) comparing with three other groups that contributed to the ZH analysis. The selection column shows each of the cuts applied. More details in the muons selection were given in table 4.5 and in section 4.5.2.1. The OR μ^{CB+ST} relates to the overlap removal made between μ^{CB+ST} and μ^{Calo} muons. The LIP columns show the number of muons of each type that pass the corresponding selection alone for the LIP group. The remaining columns, labelled A to C refer to the absolute differences in the number of muons from three different groups relative to the LIP group, as explained in section 4.6.

Selection	LIP	A	B	C
None	3388073	0	0	0
Algorithm	1293957	0	0	0
VeryLooseLH	156859	0	0	0
Pseudorapidity	2293302	2	0	0
p_T	868224	0	0	0
E_T	1546277	0	0	0
Loose	**100885**	**0**	**0**	**0**
$E_T > 25$ GeV	365261	0	0	0
Signal	**72460**	**0**	**0**	**0**

Table C.7.: Validation of the electron selection for the analysis performed in this thesis (labelled LIP) comparing with three groups that contribute to the ZH analysis. The selection column shows each of the cuts applied. The LIP column shows the number of electrons that pass the corresponding selection alone for the LIP group. The remaining columns, labelled A to C refer to the absolute differences of the number of electrons from three different groups relative to the LIP group, as explained in section 4.6. More details on the electron selection are given in section 4.5.2.2.

Selection	LIP	A	B	C
None	1935448	0	0	0
Non-negative energy	1935448	0	0	0
Pseudorapidity	1908125	0	0	0
p_T	933006	-5	0	0
JVF	1351066	0	0	0
Loose	**829169**	**-5**	**0**	**0**
$\eta < 2.5$	1528019	-2	0	0
Signal	**766557**	**-5**	**0**	**0**
B-hadron	432065	-1	0	0
C-hadron	13849	-2	0	0
Light-hadron	273676	-2	0	0
Tau particle	46967	0	0	0

Table C.8.: Validation of the jet selection for the analysis performed in this thesis (labelled LIP) comparing with three other groups that contributed to the ZH analysis. More information on the selection of jets is given in section 4.5.2.3. The LIP column displays the number of jets of each type that pass the corresponding selection alone for the LIP group. The remaining columns, labelled A to C refer to the absolute differences in the number of jets from three different groups relative to the LIP group, as explained in section 4.6.

Selection	LIP	A	B	C
Loose jets after *(el-jet)*	735501	-5	0	0
Loose jets after *(mu-jet)*	730537	-2	0	0
Loose muons after *(jet-mu)*	98578	0	0	0
Loose electrons after *(mu-e)*	100572	0	0	0
μ^{Calo} muons after *(e-mu)*	1506	-2	0	0

Table C.9.: Validation of the overlap removal selection for the analysis performed in this thesis (labelled LIP) comparing with three other groups that contributed to the ZH analysis. The selection column represents the steps taken in the overlap removal and are described in section 4.5.2.4.

$\mathrm{p}_\mathrm{T}^\mathrm{Z}$-region:	$p_\mathrm{T}^\mathrm{Z} < 90$ GeV				$90 < p_\mathrm{T}^\mathrm{Z} < 120$ GeV				$120 < p_\mathrm{T}^\mathrm{Z} < 160$ GeV			
b-tag-region	LIP	A	B	C	LIP	A	B	C	LIP	A	B	C
lltag2jets	2335	-1	0	0	614	-1	0	0	443	0	0	0
mmtag2jets	3823	-1	0	0	955	-2	0	0	718	0	0	0
tttag2jets	4159	-3	0	0	1061	-2	0	0	800	-1	-1	-1
lltag3jets	589	-1	0	0	172	0	0	0	134	0	0	0
mmtag3jets	893	-1	0	0	283	0	0	0	222	0	0	0
tttag3jets	937	0	0	0	340	0	0	0	267	-1	0	0
≥ 4 jets	473		0	0	184		0	0	165		0	0

$\mathrm{p}_\mathrm{T}^\mathrm{Z}$-region:	$160 < p_\mathrm{T}^\mathrm{Z} < 200$ GeV				$p_\mathrm{T}^\mathrm{Z} > 200$ GeV			
b-tag-region	LIP	A	B	C	LIP	A	B	C
lltag2jets	214	0	0	0	174	0	0	0
mmtag2jets	332	0	0	0	405	0	0	0
tttag2jets	412	0	0	0	445	-1	0	0
lltag2jets	82	0	0	0	96	0	0	0
mmtag2jets	135	0	0	0	199	0	0	0
tttag2jets	175	0	0	0	183	-1	0	0
≥ 4 jets	107		0	0	172		0	0

Table C.10.: Validation of the b-tagging and $\mathrm{p}_\mathrm{T}^\mathrm{Z}$-regions for the analysis performed in this thesis (labelled LIP) comparing with three other groups that contributed to the ZH analysis. These regions are described in section 4.5.4. The LIP column displays the number of events of each region the LIP group. The remaining columns, labelled A to C refer to the absolute differences in the number of events from three different groups relative to the LIP group. Missing values correspond to values not reported by a group.

C.3. Distributions for the fit

This section shows the distributions that are used in the global fit discussed in section 4.9 and it includes those presented for the ZH results in section 4.11. The first (last) column of figures shows the invariant mass distributions for the two b-tagged jets before (after) the global fit. The figures in the centre are the transformed $m_{b\bar{b}}$ distributions, as discussed in section 4.9.1, that are used for the background normalization and for the signal strength parameter extraction. The five rows in each figure correspond to the 5 $\mathrm{p_T^Z}$-bins used in the fit (see section 4.5.4). Table C.11 makes the correspondence between the signal and control regions of the analysis and the respective distributions / figures. For the region of *1-tag*, the distributions that enter in the global fit are the MV1c weight of the jets with the largest transverse momentum. There is no transformation on this variable, so the figures only show the distributions before the fit on the left, and after it on the right. Since the main purpose of this distribution is the evaluation of the normalization of the Z+non-b-jet background, only the first $\mathrm{p_T^Z}$ bins are used. Each figure shows the distribution of $m_{b\bar{b}}$ or MV1c of the backgrounds stack together, with the expected number of signal events ($\mu = 1$) on top. Additionally, the prefit histograms (left) show with the red line the ZH distribution alone multiplied with a factor for visibility. The post fit distributions have also a dashed blue which represents the initial simulation distribution before the global fit.

Region	Figure
LL signal region with events with 2 jets	Figure C.1
LL signal region with events with 3 jets	Figure C.4
MM signal region with events with 2 jets	Figure C.2
MM signal region with events with 3 jets	Figure C.5
TT signal region with events with 2 jets	Figure C.3
TT signal region with events with 3 jets	Figure C.6
Top $e-\mu$ control region with events with 2 jets	Figure C.7
1-tag control region with events with 2 jets	Figure C.8
1-tag control region with events with 3 jets	Figure C.9

Table C.11.: Correspondence between signal and control regions and the respective figures.

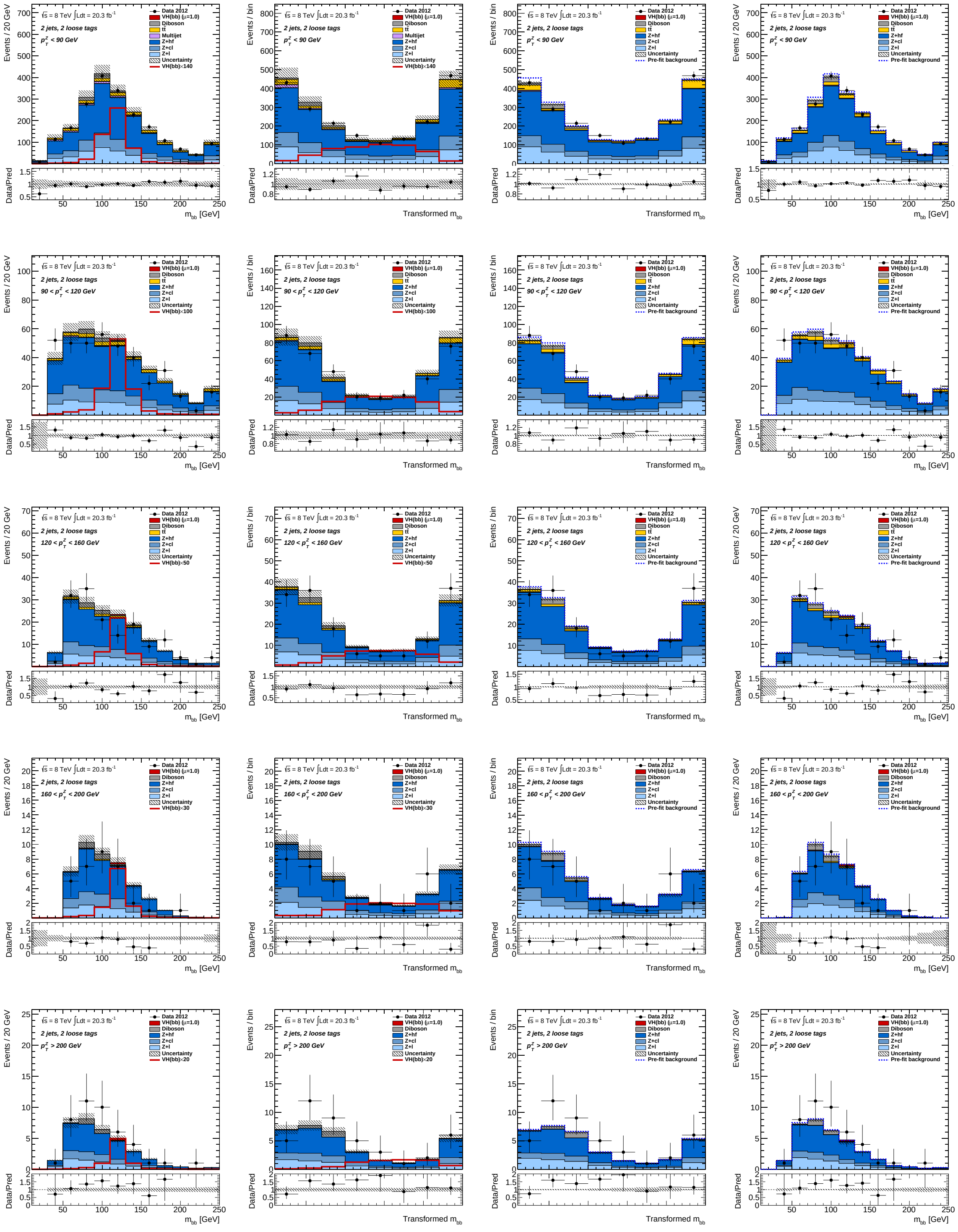


Figure C.1.: Invariant mass distributions for the two b-tagged jets. Further details are given in appendix C.3.

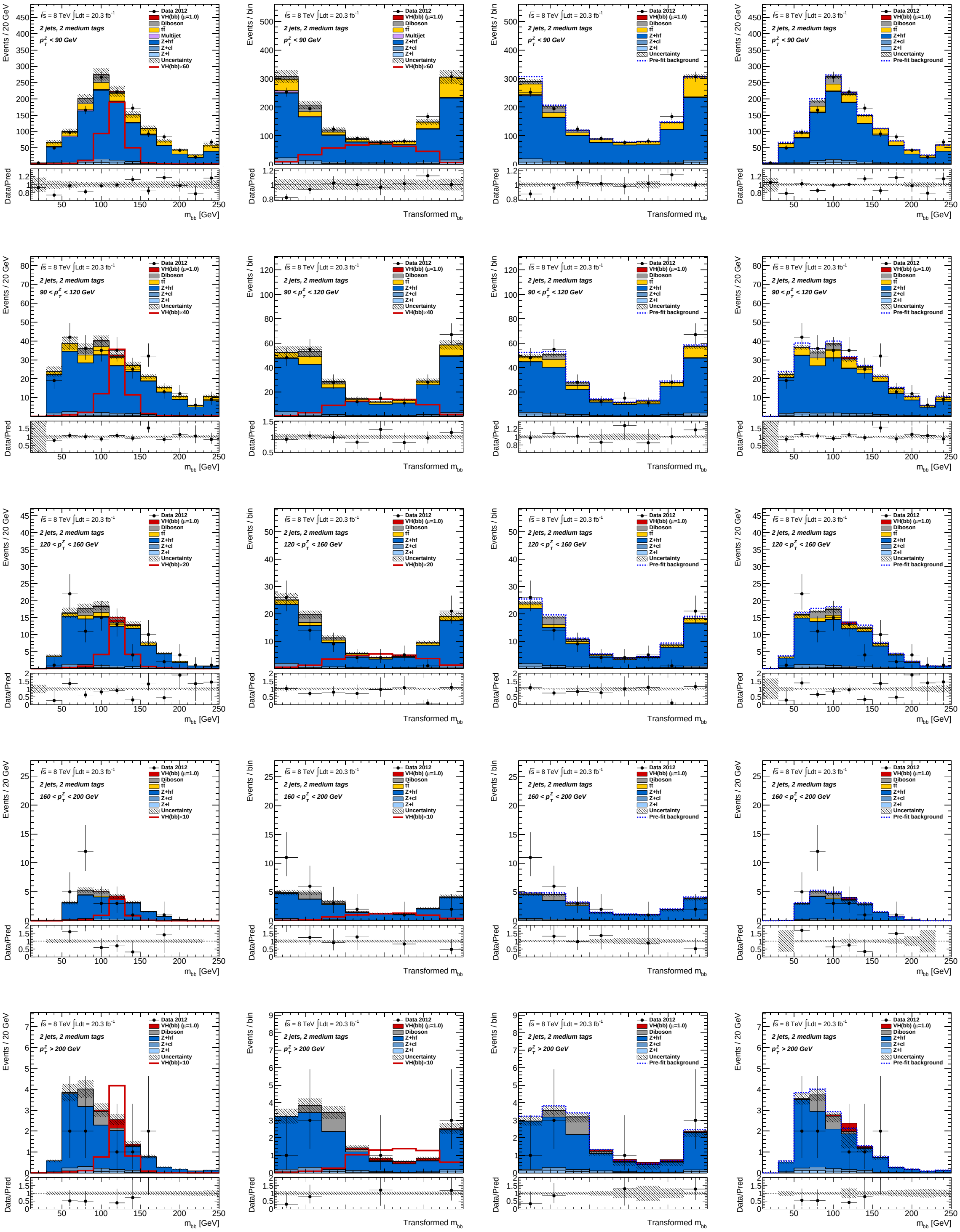


Figure C.2.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

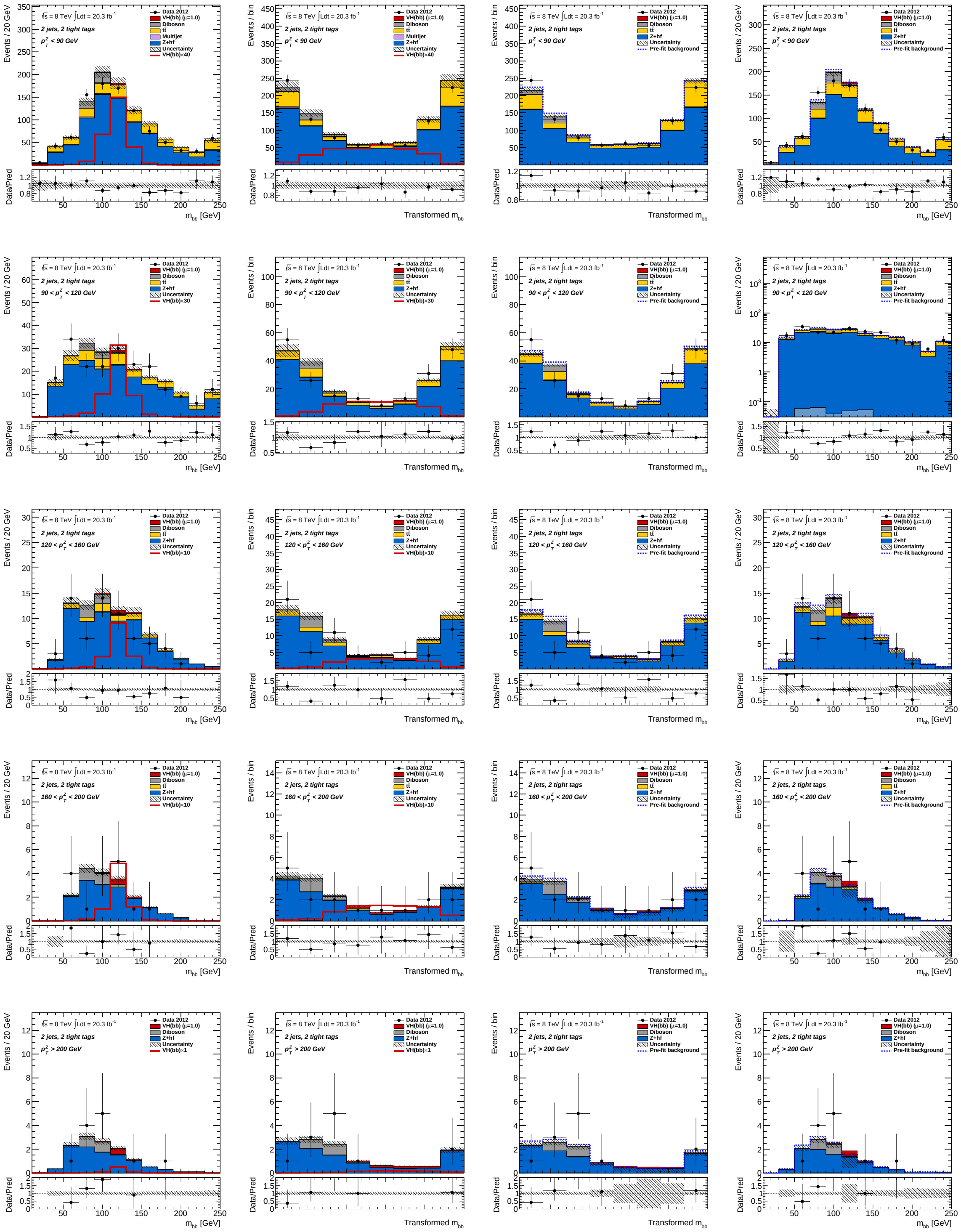


Figure C.3.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

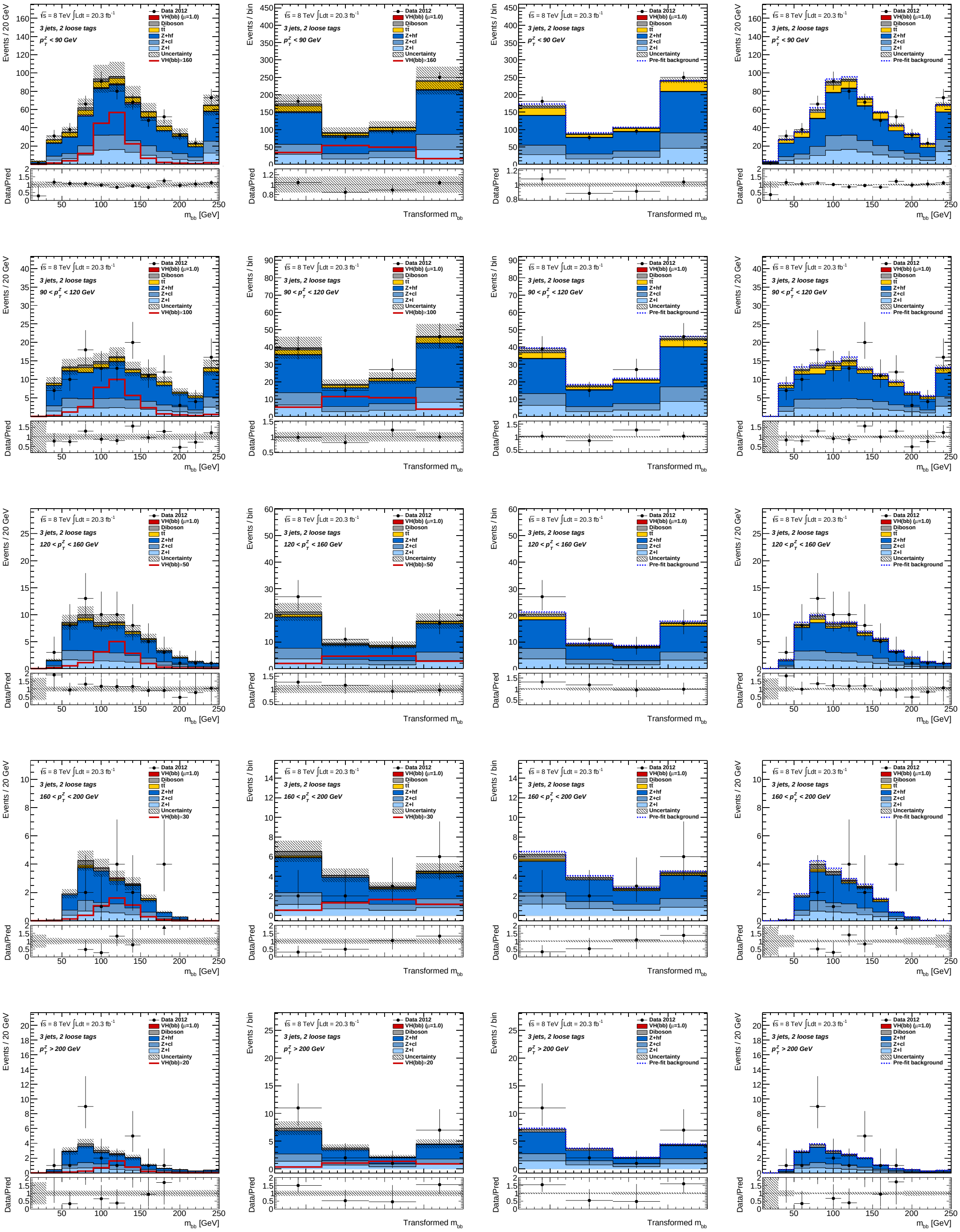


Figure C.4.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

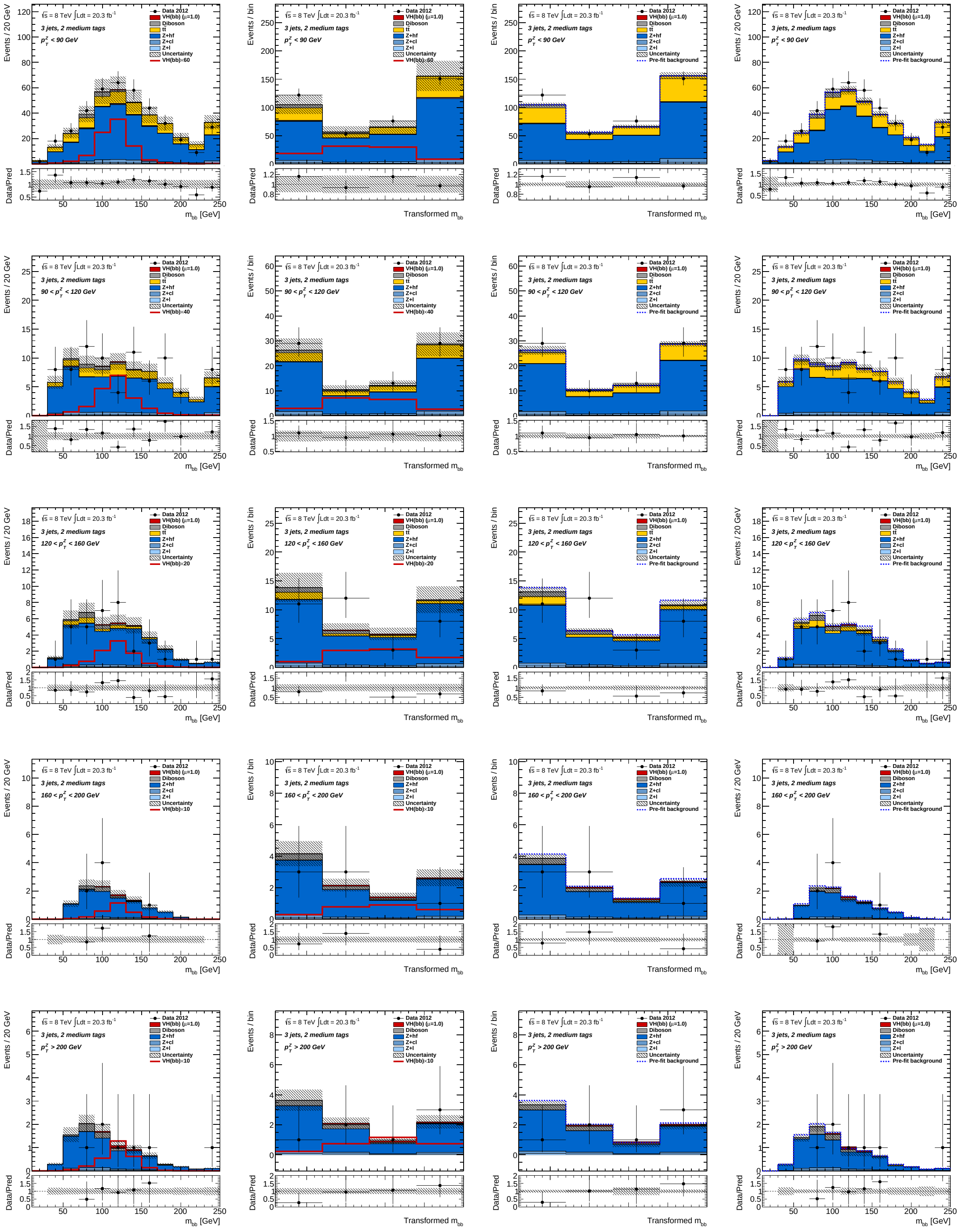


Figure C.5.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

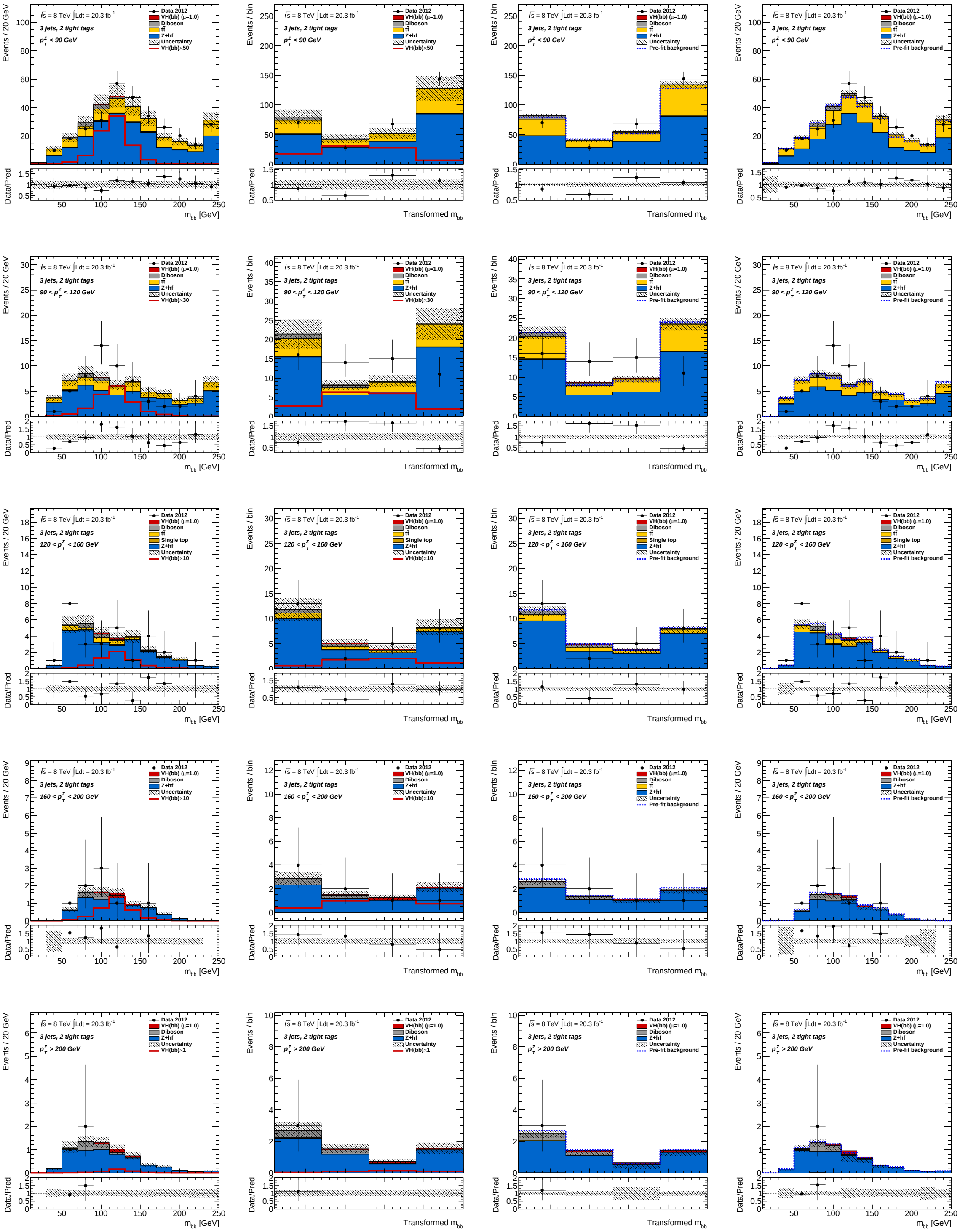


Figure C.6.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

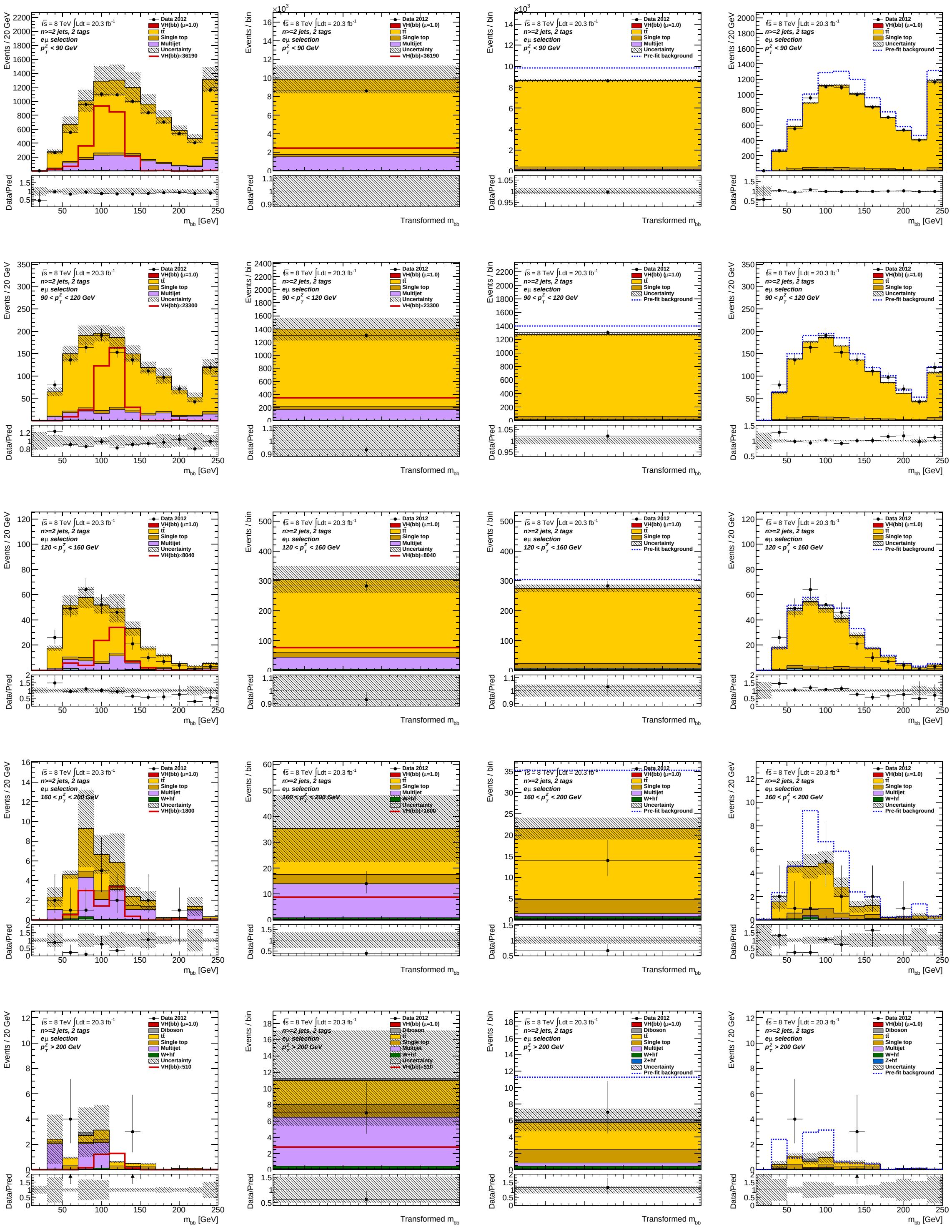


Figure C.7.: Invariant mass distributions for the two b-tagged jets (continued). Further details are given in appendix C.3.

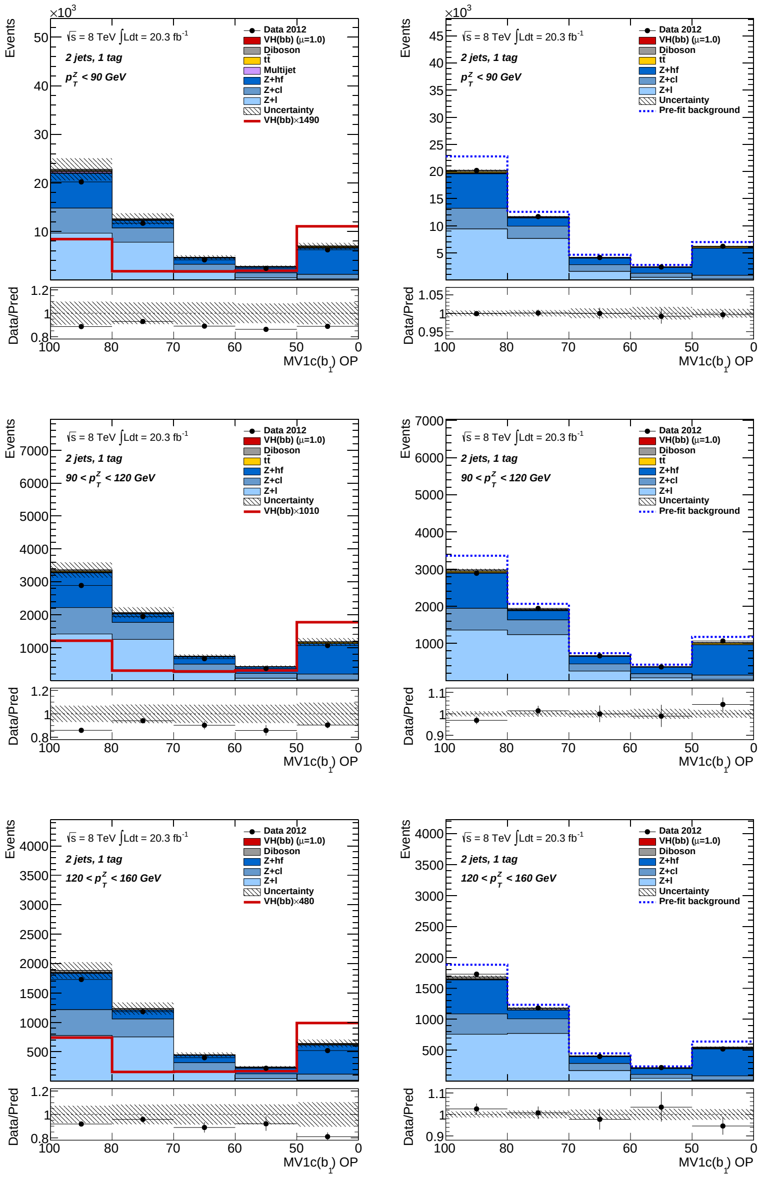


Figure C.8.: MV1c weight distributions of the b-jet in events in the *1tag* region with 2 jets. Further details are given in appendix C.3.

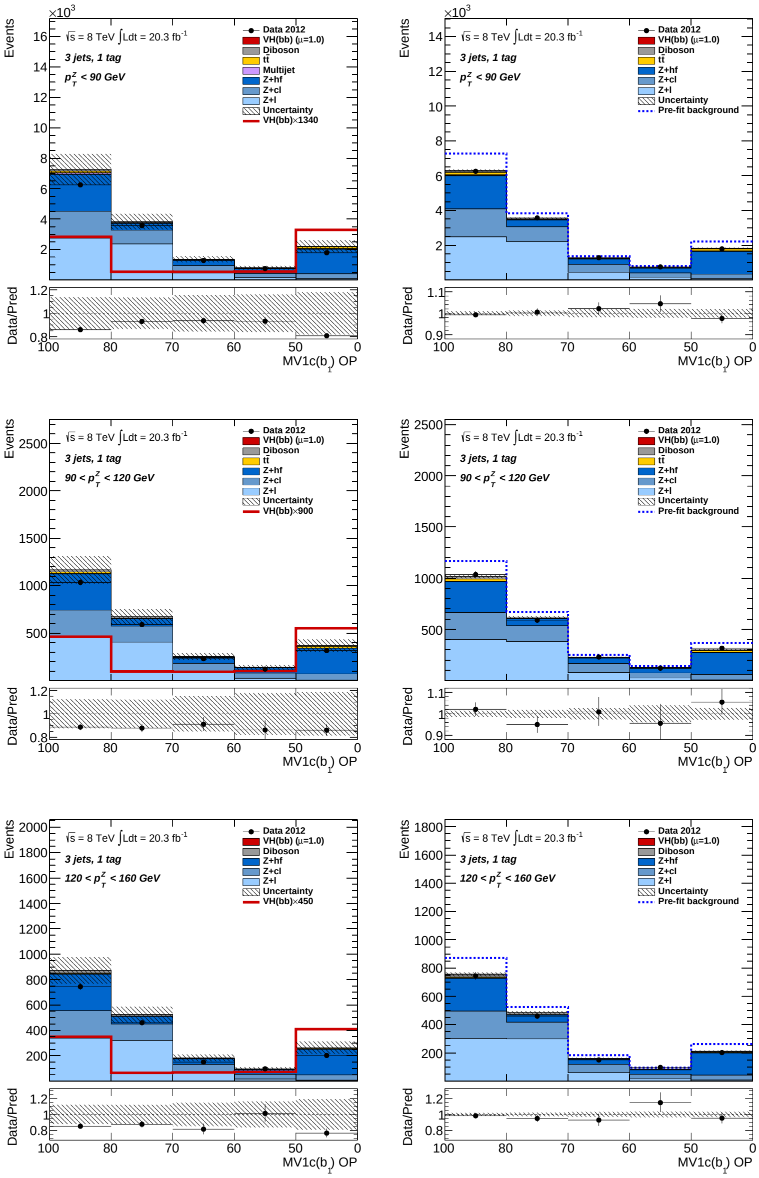


Figure C.9.: MV1c weight distributions of the b-jet in events in the *1tag* region with 3 jets. Further details are given in appendix C.3.

D. Acronyms

Γ	Width of a particle - inverse of decay rate
$TrkWidth$	Track width of the jet
$nTrk$	Number of tracks inside the jet
AF	ATLAS Fast Simulation with FastCaloSim
AFII	ATLAS Fast Simulation with FastCaloSim
ATLFAST-II	ATLAS Fast Simulation with FastCaloSim
AOD	Analysis Object Data
ATLAS	A Toroidal LHC ApparatuS. Refers to the detector, to the experiment or to the collaboration
AUET2	ATLAS underlying event tune 2
BDT	Boost Decision Tree
BERT	Bertini model for nucleon-nucleon classic scattering
BR	Branching ratio
C/A	Cambridge/Aachen jet cluster algorithm
CaloTag	Calorimeter-tagged muon
CB	Combined muon
CERN	Nuclear Research European Centre
CMS	Compact Muon Solenoid. Refers to the detector or to the collaboration. The direct competition of the ATLAS experiment
CTB	Combined Test Beam
DAQ	Data Acquisition
DB	Direct balance method
EB	Event Builder
EF	Event Filter
Egamma	(stream) events selected by electron and photon based triggers
EM	Electromagnetic
EMJES	Jet energy scale based on the electromagnetic calibration
EM+JES	Jet energy scale based on the electromagnetic calibration
EM+JES+GSC	GSC applied on top of EM+JES
EMEC	LAr electromagnetic end-cap calorimeter of ATLAS
EMiso	Electromagnetic isolation ring
ESD	Event Summary Data
EW	Electroweak
FastCaloSim	Fast Calorimeter Simulation
FCal	Forward calorimeter of ATLAS
f_{EM3}	Ratio between energy deposited in last layer of EM calorimeter and the total energy of the jet at the EM scale
FSR	Final state radiation
FTFP	FRITIOF model with Precompound
f_{Tile0}	Ratio between energy deposited in first layer of TileCal and the total energy of the jet at the EM scale
g-jet	gluon initiated jet
GRL	Good Runs List
GSC	Global Sequential Calibration

hadCalib	Hadronic calibration triggers
HADcore	Hadronic core region
HADiso	Hadronic isolation ring
HEC	Hadronic end-cap calorimeter of ATLAS
HL-LHC	High Luminosity LHC
HLT	High Level Trigger
ID	Inner detector - inner tracker
ISR	Intersecting Storage Rings
ISR	Initial state radiation
JER	Jet energy resolution
JetTauEtMiss	(stream) events selected by jet, tau and missing transverse energy based triggers
JVF	Jet vertex fraction
L1	Level 1 of trigger in ATLAS during Run-1
L2	Level 2 of trigger in ATLAS during Run-1
LAr	Liquid Argon calorimeters
LC+JES	Jet energy scale based on the LCW
LCW+JES	Jet energy scale based on the LCW
LCW	Local hadronic cluster weightning
LEP	Large Electron-Positron collider
LH	Likelihood
LHC	Large Hadron Collider
LIP	Laboratório de Instrumentação e Física Experimental de Partículas
LL	Light-light tagging
LLH	log-Likelihood
LO	Leading order
lq-jet	Light-quark initiated jet
MBTS	Minimum Bias Trigger Scintillators
MC	Monte Carlo simulation
MDT	Monitored Drift Tubes of ATLAS
MET	Missing transverse energy
MIP	Minimum ionising like particle
MM	Medium-medium tagging
MPV	Most probable value
MS	Muon Spectrometer of ATLAS
MV1	Default b-tagger based on multivariate analysis
MV1c	Same as MV1 but with better c-jet rejection - used in ZH analysis
MVA	Multivariate analysis
NLO	Next to leading order
NNLL	Next to next to leading log
NNLO	Next to next to leading order
N_{PV}	Number of primary vertices
OR	Overlap removal
PES	Photon energy scale

PDF	Parton distribution function
pdf	Probability distribution function
POPOP	1,4-bis-(2-(5-phenyloxazol-2-yl))-benzene
PS	Proton Synchrotron accelerator in CERN
PS	Prescale
PT	Pass-through
PTP	para-terphenyl-benzene
PU	Pile-up
QCD	Quantum chromodynamics
QED	Quantum electromagnetic dynamics
QGSP	Quark-Gluon String model with Precompound
RD0	Random triggers at L1
RDO	Raw Data Object
RMS	Root mean squared
ROI	Region of interest
ROS	Readout system
RPC	Resistive Plate Chambers of ATLAS
Run 1	First LHC data taking from 2010 to 2012
Run 2	Second LHC data taking from 2015 and still on-going
SA	Standalone muon
SC	Synchro-Cyclotron
SCT	Semiconductor tracker of ATLAS
SF	Scale factor
SM	Standard Model of particle physics
SP$\bar{P}$S	Super Proton-Anti-Proton Synchrotron
SPS	Super Proton Synchrotron
ST	Segmented-Tagged muon
TDAQ	Trigger and Data Aquisition of ATLAS
TGC	Thin Gap Chambers of ATLAS
TileCal	Tile hadronic barrel Calorimeter of ATLAS
TMVA	Toolkit for MVA
TRT	Transistion Radiation Tracker of ATLAS
TT	Trigger towers
TT	Tight-Tight tagging
UE	Underlying event
VH	channel or analysis of the search for the Higgs boson in association with a vector boson (W or Z boson), in which the Higgs decays into a b-quark pair and the vector boson decays leptonicaly
WH	channel or analysis of the search for the Higgs boson in association with a W boson, in which the Higgs decays into a b-quark pair and the W boson decays into a charge lepton (muon or electron) and a neutrino
ZH	channel or analysis of the search for the Higgs boson in association with a Z boson, in which the Higgs decays into a b-quark pair and the Z boson decays into two charge leptons (muons or electrons)

List of Figures

List of Tables