

MODELO NEURAL PARA CORREÇÃO DA ESTIMAÇÃO DE ENERGIA
PARA UM DETECTOR DE PARTÍCULAS DE PROPÓSITO GERAL

Natanael Nunes de Moura Junior

Dissertação de Mestrado apresentada ao
Programa de Pós-graduação em Engenharia
Elétrica, COPPE, da Universidade Federal
do Rio de Janeiro, como parte dos requisitos
necessários à obtenção do título de Mestre
em Engenharia Elétrica.

Orientadores: José Manoel de Seixas
Marcia Begalli

Rio de Janeiro
Março de 2014

MODELO NEURAL PARA CORREÇÃO DA ESTIMAÇÃO DE ENERGIA
PARA UM DETECTOR DE PARTÍCULAS DE PROPÓSITO GERAL

Natanael Nunes de Moura Junior

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA
ELÉTRICA.

Examinada por:

Prof. José Manoel de Seixas, D.Sc.

Prof. Marcia Begalli, Dr. rer. nat

Prof. Ignácio Alfonso de Bediaga e Hickman, Ph.D.

Prof. Marcello Luiz Rodrigues de Campos, Ph.D.

Prof. Ivan Nunes da Silva, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2014

Moura Junior, Natanael Nunes de

Modelo Neural para Correção da Estimação de energia
para um detector de partículas de propósito geral/Natanael
Nunes de Moura Junior. – Rio de Janeiro: UFRJ/COPPE,
2014.

XVIII, 99 p.: il.; 29, 7cm.

Orientadores: José Manoel de Seixas

Marcia Begalli

Dissertação (mestrado) – UFRJ/COPPE/Programa de
Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 93 – 99.

1. Inteligência Computacional. 2. Física de Altas Energias. 3. Jatos de Partículas. 4. Redes Neurais Artificiais. I. Seixas, José Manoel de *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Mas a manifestação do Espírito
é dada a cada um, para o que for
útil.*

*Porque a um pelo Espírito é
dada a palavra da sabedoria; e a
outro, pelo mesmo Espírito, a
palavra da ciência;
E a outro, pelo mesmo Espírito,
a fé; e a outro, pelo mesmo
Espírito, os dons de curar
Mas um só e o mesmo Espírito
opera todas estas coisas,
repartindo particularmente a
cada um como quer.*

1 Coríntios 12: 7-9, 11

Agradecimentos

Primeiramente a Deus, que mesmo com toda a minha infidelidade sempre esteve ao meu lado e caminhou comigo durante todo o tempo em que desenvolvi o meu trabalho. Deus sempre foi o meu refugio e minha fortaleza e não poderia deixar de agradecer ao Deus que tudo fez.

Aos meus pais, Natanael e Neidimar, pelos conselhos, pelas orientações e acima de tudo pela paciência e carinho comigo. Meu pai me ensinou o valor da vida e mesmo passando por muitos problemas de saúde sempre se manteve firme e forte em seus convicções e propósitos, nunca mudando os seus discursos. Minha mãe sempre flexível e amável me mostrou que é possível ter sucesso sem passar por cima de ninguém, vivendo uma vida integra e servindo a Deus.

Ao meu orientador, José Manoel de Seixas que foi um orientador dedicado, um mestre solícito e sempre um grande amigo. Eu gostaria de agradecer por sempre me fazer extrair o melhor do meu trabalho e sempre acreditar que mesmo com dificuldades podemos ser ótimos profissionais.

À minha co-orientadora, Marcia Begalli, com quem eu tive o imenso prazer de trabalhar, dividir problemas e experiências e com quem eu aprendi muito, não somente sobre física ou análise, mas de tudo um pouco.

Aos meus amigos de faculdade, que ouviram pacientemente as minhas queixas e minhas frustrações e que me deram, muitas vezes, idéias que salvaram este trabalho. Muito obrigado ao Daniel Dias, Gustavo Moraes, Luiz Eduardo Balabram Filho, Gustavo Campos, Rodrigo Perreira, Luiz Évora e outros bravos soldados do DEL.

Aos meus amigos do Laboratório de Processamento de Sinais, que sempre estiveram presentes nessa caminhada. Muito obrigado aos garotos João Victor Pinto, Felipe Gonzales, Julio de Castro, Werner Freund, Hellen Lima, Victor Costa, Diego Cruz, Luciano Vargas, Ana Fernanda Quaresma, Camila Gussen, Fabiano Castoldi, Felipe Fink Grael, Fernando Ferreira, Andressa Sivollela, Thiago Ciodaro, Laura Moraes, Luiz Frias, Victor Zatorre e tantos outros que passaram pelo LPS.

A todos os meus professores, que foram verdadeiros mestres, além de informações, me ensinaram sobre respeito e dignidade e são exemplos do que eu quero me tornar um dia.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MODELO NEURAL PARA CORREÇÃO DA ESTIMAÇÃO DE ENERGIA PARA UM DETECTOR DE PARTÍCULAS DE PROPÓSITO GERAL

Natanael Nunes de Moura Junior

Março/2014

Orientadores: José Manoel de Seixas

Marcia Begalli

Programa: Engenharia Elétrica

Experimentos com colisões de partículas visam aprofundar nossos conhecimentos sobre a matéria e identificar processos complexos previstos teoricamente. No CERN, o LHC começou a colidir feixes de prótons em 2011 e os colide a 14 TeV de energia nominal no centro de massa com colisões ocorrendo em um período de 25 ns. A fim de medir os subprodutos resultantes das colisões, detectores de partículas são acoplados em torno dos pontos de colisão, sendo o ATLAS, o maior é deles. Em julho de 2012, foi anunciada a observação de uma partícula que tem o comportamento similar ao bóson de Higgs. Para que a partícula anunciada seja, comprovadamente, o bóson de Higgs, a mesma deve ser observada em todos os seus canais de decaimento, com as mesmas características anunciadas. A leitura e identificação de jatos do tipo b (jatos cuja origem é um quark b) é de grande importância na física de altas energias, pois diversos processos físicos podem ser observados, a partir deles, dentre estes, um dos decaimentos teóricos do bóson de Higgs, o $H \rightarrow b\bar{b}$. Um modelo neural foi desenvolvido e utilizado com sucesso para a correção das não-linearidades intrínsecas à resposta do detector ATLAS para dois jatos do tipo b identificados como provenientes de um bóson de Higgs. Neste trabalho, além do processo de desenvolvimento do modelo neural, será apresentada a modelagem estatística feita para a distribuição de massa do bóson de Higgs, os índices de qualidade para essa modelagem e dentro desses índices, será introduzido o conceito de divergência estatística para física de altas energias.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

NEURAL MODEL FOR ENERGY ESTIMATION CORRECTION FOR
GENERAL PROPOSE PARTICLE DETECTOR

Natanael Nunes de Moura Junior

March/2014

Advisors: José Manoel de Seixas
Marcia Begalli

Department: Electrical Engineering

Collider experiments are probing the human knowledge about the inner structure of the matter, trying also to identify complex processes predicted by different theoretical models. At CERN, the LHC started to collide protons in 2011 reaching 14 TeV of center of mass nominal energy. The collisions occur in a period of 25 ns. To measure the products of those collisions, particle detectors were assembled around the collision points, being ATLAS the largest one in LHC. On July 2012, it was announced the observation of a Higgs-like particle. To prove the observed particle is, indeed the Higgs boson, it must be observed in all different decay channels with the same predicted features. In high energy physics it is very important to correctly tag the b-jets because many important physical processes in particle physics can be studied using them, including the Higgs boson decay channel $H \rightarrow b\bar{b}$. An Artificial Neural Networks were successfully used to compensate inherent non-linearities in ATLAS detector response. This Master Thesis will present the statistical modeling of Higgs boson mass distribution, quality indexes and will be introduced the concept of divergence for high energy physics.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
Lista de Símbolos	xv
Lista de Abreviaturas	xvii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivo	3
1.3 Contribuições do Trabalho	3
1.4 Organização da dissertação	3
2 Física de Altas Energias	5
2.1 Física Experimental de Altas Energias	5
2.2 CERN e o LHC	7
2.3 O Detector ATLAS	10
2.3.1 Sistema de Coordenadas do ATLAS	11
2.3.2 Detector Interno	12
2.3.3 Sistema de Calorimetria	13
2.3.4 Sistema de Múons	20
2.3.5 Sistema de Trigger	21
3 Técnicas para Correções de Energia	24
3.1 Correções de Energia	24
3.1.1 Resolução em Energia	24
3.2 Tevatron	26
3.2.1 Detector CDF	26
3.2.2 Detector $D\emptyset$	31

4 Metodologia de Análise	36
4.1 Modelo de Dados do detector ATLAS	36
4.1.1 Dados Utilizados	38
4.2 Redes Neurais Artificiais	40
4.2.1 Validação Cruzada	44
4.2.2 Aplicação do Modelo Neural	44
4.2.3 Treinamentos Realizados	44
4.3 Massa Invariante do Bóson de Higgs	45
4.3.1 Correção aplicada à Massa	46
4.3.2 Modelos de Ajuste	47
4.3.3 Determinação da Massa Invariante do Bóson de Higgs	51
4.4 Métodos de Análise Estatísticos	51
4.4.1 Estimação da função densidade de probabilidade	51
4.4.2 Divergência de Kullback-Leibler	52
4.4.3 Divergência de Jensen-Shannon	53
4.4.4 Teste de Kolmogorov-Smirnov	53
5 Resultados Obtidos	55
5.1 Massa reconstruída do bóson de Higgs	55
5.1.1 Análise de massa do bóson de Higgs para o detector ATLAS .	57
5.2 Primeiro Modelo Neural aplicado à Massa do Bóson de Higgs	58
5.3 Segundo Modelo Neural aplicado à Massa do Bóson de Higgs	60
5.4 Análise de Relevância do Modelo Neural	63
5.4.1 Análise de Relevância tendo como Referência as Saídas	64
5.4.2 Análise de Relevância tendo como Referência o Alvo	67
5.5 Terceiro Modelo Neural aplicado a Massa do Bóson de Higgs	69
5.6 Quarto Modelo Neural aplicado à Massa do Bóson de Higgs	71
5.7 Ajustes dos Modelos descritos	72
5.7.1 Resultados para Divergência de Kullback-Leibler	76
5.7.2 Resultados para Divergência de Jensen-Shannon	80
5.7.3 Resultados para teste de Kolmogorov-Smirnov	83
5.8 Discussão dos Resultados Obtidos	84
6 Conclusões e Trabalhos Futuros	90
6.1 Trabalhos Futuros	91
Referências Bibliográficas	93

Listas de Figuras

2.1	Tabela com os componentes do Modelo Padrão (Extraído de [1]).	6
2.2	Decaimentos do Bóson de Higgs e seus respectivos <i>branching ratios</i> . Extraído de [8]	7
2.3	Visão Aérea do CERN, com destaque para os dois detectores de propósito geral (Extraído de [1]).	9
2.4	O Detector ATLAS (Extraído de [1]).	10
2.5	Sistema de Coordenadas do ATLAS.	11
2.6	Visão Geral do Inner Detector (Extraído de [1]).	12
2.7	Visão do detector geral em um corte perpendicular a linha do feixe (Extraído de [1]).	13
2.8	Visão Geral do Sistema de Calorimetria do detector ATLAS (Extraído de [1]).	14
2.9	Visão Geral do Calorímetro Eletromagnético do ATLAS (Extraído de [1]).	15
2.10	Células de Leitura da parte central do Calorímetro LAr (Extraído de [1]).	16
2.11	Visão Geral do Calorímetro Hadronico do ATLAS (Extraído de [1]). .	16
2.12	Descrição de um Módulo de leitura do TileCal (Figuras extraídas de [1]).	18
2.13	Componentes do Calorímetro Hadronico do ATLAS (Figuras ex- traídas de [1]).	18
2.14	Visão Geral do Muon System (Extraído de [1]).	20
2.15	Distribuição das câmaras do Muon System (Extraído de [1]).	21
2.16	Representação gráfica do sistema de <i>trigger</i> do detector ATLAS (Ex- traído de [1]).	23
3.1	Esquema do detector CDF	27
3.2	Detector CDF: Visão interna do Central Outer Tracker	28
3.3	Esquema do detector $D\emptyset$	32
3.4	Detector $D\emptyset$: Correção do offset de energia com diversas luminosida- des e raio de reconstrução de cone de 0,5.	34

3.5	Detector $D\emptyset$: Correção do offset de energia com diversas luminosidades e raio de reconstrução de cone de 0,7	34
4.1	Metodologia de Análise utilizada neste trabalho	36
4.2	Modelo de uma rede neural artifical proposta por McCulloch e Pitts em 1943	41
4.3	Modelo de uma rede neural artifical com uma única camada escondida .	42
4.4	Modelo de uma rede neural artifical gerando duas correções (uma para cada jato)	45
4.5	Distribuições dos kernels <i>box</i> e de Epanechnikov	52
5.1	Distribuições para a Massa do bóson de Higgs	56
5.2	<i>Scatter plot</i> da massa simulada <i>versus</i> massa reconstruída	57
5.3	Distribuições de massa do bóson de Higgs com o ajuste gaussiano .	58
5.4	Resultados para o conjunto de variáveis sugeridos pela colaboração ATLAS, com o treinamento de primeira ordem	59
5.5	<i>Scatter plot</i> da massa simulada <i>versus</i> massa reconstruída com e sem a correção neural	61
5.6	Resultados para o conjunto de variáveis sugeridos pela colaboração ATLAS, com o treinamento de segunda ordem	62
5.7	<i>Scatter plot</i> da massa simulada <i>versus</i> massa reconstruída com e sem a correção neural (treinamento de segunda ordem)	63
5.8	Distribuições das Variáveis da tabela 5.1	65
5.9	Resultados para a Análise de Relevância com referência a Saída do Modelo Neural para ambos os jatos	66
5.10	Resultados para a Análise de Relevância tendo como referência o Alvo do Modelo Neural para ambos os jatos	68
5.11	Resultados para o conjunto de variáveis sugeridos pela análise de relevância com treinamento de primeira ordem	70
5.12	<i>Scatter plot</i> da massa simulada <i>versus</i> massa reconstruída com e sem a correção neural (treinamento de primeira ordem)	71
5.13	Resultados para o conjunto de variáveis sugeridos pela análise de relevância com treinamento de segunda ordem	73
5.14	<i>Scatter plot</i> da massa simulada <i>versus</i> massa reconstruída com e sem a correção neural (treinamento de segunda ordem)	74
5.15	Modelos ajustados as diferentes distribuições.	75
5.16	Modelos ajustados as diferentes distribuições.	77
5.17	Distribuições de Massas Simulada, Reconstruída, Reconstruída com o Corretor neural e o melhor ajuste de modelo para cada distribuição com cálculo da divergência KL	78

5.18 Divergência KL <i>bin a bin</i> entre as distribuições de massa simulada e massa reconstruída	79
5.19 Divergência KL <i>bin a bin</i> entre as distribuições de massa simulada e massa reconstruída com a correção neural	80
5.20 Divergência JS <i>bin a bin</i> entre as distribuições de massa simulada e massa reconstruída	82
5.21 Divergência JS <i>bin a bin</i> entre as distribuições de massa simulada e massa reconstruída com a correção neural	83
5.22 Resultados obtidos com o melhor ajuste dentre dos modelos propostos para o primeiro e segundo modelos neurais.	87
5.23 Resultados obtidos com o melhor ajuste dentre dos modelos propostos para o terceiro e o quarto modelos neurais.	88

Listas de Tabelas

2.1	Decaimentos do Bóson de Higgs e seus respectivos <i>branching ratios</i> . Valores extraídos de [8]	8
4.1	Parametros de treinamento de Redes Neurais	46
4.2	Eficiência das principais funções de Kernel, valores extraídos de [90]	52
5.1	Características das Variáveis provenientes do banco de dados disponibilizado	64
5.2	Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições	76
5.3	Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições	81
5.4	Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições	81
5.5	Divergência JS: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições	81
5.6	Divergência JS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições	82
5.7	Divergência JS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições	82
5.8	Teste KS: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições	83
5.9	Teste KS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições	84
5.10	Teste de KS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições	84
5.11	Comparação dos resultados obtidos com diferentes conjuntos de entradas para a modelagem gaussiana	85

5.12 Comparação e escolha do melhor modelo em diferentes análises estatísticas	86
5.13 Comparação dos resultados obtidos com diferentes conjuntos de entradas para a melhor modelagem dentro dos quatro modelos testados.	89

List a de Símbolos

E	Energia, p. 46
E_t	Energia Transversa do Jato, p. 64
E_{reco}	Energia Reconstruída, p. 47
E_{truth}	Energia Simulada, p. 47
F_{custo}	Função custo do treinamento de rede neural, p. 41
GeV	Giga eletron-volt - 10^9 eletron-volt, p. 6
H	Símbolo do bóson de Higgs, p. 44
MHz	Megahertz - 10^6 Hz, p. 22
M_t	Massa Transversa do Jato, p. 64
Pt	Momento Transversa do Jato, p. 64
TeV	Tera eletron-volt - 10^{12} eletron – volt, p. 7
V	Energia Potencial, p. 48
W	bóson W, p. 6
Z	bóson Z, p. 6
Γ	Inverso da meia-vida de uma partícula, p. 48
χ^2	Valor calculado para o teste de χ^2 , p. 47
η	Eta - 1 ^a coordenada do sistema ATLAS, p. 11
∞	Infinito, p. 48
μm	Micrômetro - 10^{-6} metro, p. 12
ϕ	Phi - 2 ^a coordenada do sistema ATLAS, p. 11

π	Píon, p. 5
ψ	Função de Onda, p. 48
τ	Meia-vida de uma partícula, p. 48
θ	Fator de Correção de Massa, p. 47
θ	O ângulo formado entre a direção do feixe e a trajetória da partícula, p. 11
ξ	Constante de Normalização, p. 48
b	Quark b , p. 7
c	Constante da velocidade da Luz - $3 \cdot 10^9$ m/s, p. 46
d	Quark d , p. 5
m	Massa Invariante, p. 46
m	Metros, p. 12
m_H	Massa do bóson de Higgs, p. 38
m_p	Massa da Partícula, p. 48
m_{reco}	Massa Reconstruída, p. 47
m_{truth}	Massa Simulada, p. 47
$n(t)$	Número de partículas em um determinado tempo t , p. 48
p	Momento, p. 46
s	Segundos, p. 10
u	Quark Up , p. 5
$w_{ij}(t)$	Valor do peso ij no tempo t , p. 41
Hz	Hertz - $1/s$, p. 22
Mb	Megabits - 10^6 bit, p. 37
kB	quilo Bytes - 10^3 Bytes, p. 37
mm	Milimetro, p. 12

Listas de Abreviaturas

D^1PD	Primary Derived Physics Dataset, p. 37
D^2PD	Secondary Derived Physics Dataset, p. 37
D^3PD	Tertiary Derived Physics Dataset, p. 38
ALICE	A Large Ion Collider Experiment, p. 10
AMISE	Asymptotic Mean Integrated Squared Error, p. 52
AOD	Analysis Object Data, p. 37
ATLAS	A Toroidal LHC ApparatuS, p. 1
CDF	Collider Detector at Fermilab, p. 24
CDF	Função de Distribuição Cumulativa, p. 53
CERN	Centro Europeu de Pesquisas Nucleares, p. 1
CFT	Central Fiber Tracker, p. 32
CMS	Compact Muon Solenoid, p. 10
COT	Central Outer Tracker, p. 27
EDM	Event Data Model, p. 36
EMEC	<i>ATLAS Electromagnetic End-cap</i> - End-cap Eletromagnético do detector ATLAS, p. 15
ESD	Event Summary Data, p. 37
FOOA	<i>first-order optimization algorithm</i> , p. 43
FermiLab	Fermi National Accelerator Laboratory, p. 24
HEC	<i>ATLAS Hadronic End-cap</i> - End-cap Hadrônico do detector ATLAS, p. 15

HOS	High Order Statistics, p. 3
JER	Jet Energy Resolution, p. 51
JS	Jensen-Shannon, p. 53
KL	Kullback-Leibler, p. 52
LAr	Liquid Argon Calorimeter - Calorímetro de Argônio Líquido, p. 11
LHC	Large Hadron Collider, p. 2
LHCb	Large Hadron Collider beauty, p. 10
LM	Algoritmo de Otimização de Levenberg-Marquardt, p. 44
MDC	Muon Drift Chambers, p. 29
MISE	Mean Integrated Squared Error, p. 52
MSE	Mean Squared Error, p. 43
PDF	Função densidade de probabilidade, p. 51
SLAC	Stanford Linear Accelerator Center, p. 49
SMT	Silicon microstrip Tracker, p. 32
SM	Standard Model, p. 38
SM	Standard Model ou Modelo Padrão, p. 5
SOOA	<i>second-order optimization algorithm</i> , p. 44
SPEAR	Stanford Positron Electron Asymmetric Rings, p. 49
SVX	Silicon VerteX Detector, p. 26
TileCal	Tile Calorimeter - Calorímetro de Telhas Cintilantes, p. 11

Capítulo 1

Introdução

Em um sistema de instrumentação eletrônica, a medição feita pelos sensores ou transdutores, a transmissão dos sinais obtidos e o armazenamento e/ou análise dos mesmos possuem papel fundamental. Nos dias atuais, existe uma grande necessidade de sistemas que forneçam medidas mais rápidas e com maior acurácia e precisão.

Atualmente, encontram-se sistemas de instrumentação que são compostos por diversos detectores e, muitas vezes, os detectores são compostos por diversos sensores, o que torna necessária a fusão de dados proveniente das leituras dos mesmos.

Em sistemas de instrumentação, deve-se observar atentamente a resolução da medida obtida que, quando comparada a um modelo matemático, pode fornecer o grau de confiabilidade do dado lido. Em um sistema de instrumentação real, pode-se utilizar diversos modelos, muitas vezes complexos, que descrevem o processo físico envolvido. Por isso, deve-se testar a aderência da medida ao modelo.

Outro fator que deve ser ponderado, no que concerne sistemas de instrumentação, é a linearidade associada à técnica de medida. As não-linearidades podem também ser influenciadas pela forma como o projeto foi implementado e diversos algoritmos buscam compensar seus efeitos.

1.1 Motivação

No contexto de física experimental de altas energias, sistemas de instrumentação são desenvolvidos para um experimento específico, o que faz com que laboratórios de física experimental se tornem, também, centros de desenvolvimento em sistemas de instrumentação.

Um exemplo de centro de desenvolvimento em sistemas de instrumentação é a colaboração ATLAS,¹ pertencente ao CERN. O detector ATLAS é o maior dos

¹Por vezes, a colaboração ATLAS pode ser chamada de experimento ATLAS ou mesmo de detector ATLAS, mesmo sendo composto por diversos sub-sistemas que são responsáveis por detectar partículas.

detectores colocados ao redor dos pontos de colisão do acelerador de partículas LHC, o grande colisor de hadróns que acelera feixes de prótons a 7 TeV, permitindo colisões próton-próton a uma energia de centro de massa de 14 TeV. Além de prótons, o LHC também acelera íons de chumbo a energia de 2,76 TeV por *nucleon*, perfazendo 575 TeV por íon.

No ATLAS, os subprodutos da colisão dos feixes de prótons são lidos por seus diversos sub-sistemas, cada um com um propósito diferente. Assim, o processo de reconstrução do evento se dá a partir das leituras registradas em cada um desses sub-sistemas. Cada sub-sistema do ATLAS é responsável pela leitura de uma característica proveniente dos sub-produtos da colisão.

O detector de traços registra o traço ou caminho percorrido por partículas carregadas que, quando sujeitas ao campo magnético do Solenóide, tem sua trajetória curvada, com raio de curvatura maior ou menor dependendo do momento da partícula. Já o sistema de calorimetria é o responsável pela leitura de energia das partículas incidentes, enquanto o Sistema de Múons deve detectar a presença de múons provenientes da colisão.

Para obtermos o valor de uma determinada quantidade, uma característica básica deve ser considerada: a resolução das medidas. No caso dos calorímetros, equipamentos desenvolvidos para medição de energia de partículas de elevada energia por meio de absorção total, o problema se torna ainda mais complexo, pois os mesmos atuam em uma ampla faixa de energia (de MeV a TeV), possuem não-linearidades inerentes a um projeto prático e precisam atingir uma resolução elevada, para que seja possível a medida de energia nos experimentos com colisão de partículas.

Quando ocorre uma colisão entre partículas sub-atômicas, como prótons ou elétrons a altas energias, podemos observar com frequência a formação de jatos de partículas. No caso da colisão entre dois prótons, como no ATLAS, sabemos que os quarks que formam os prótons irão interagir, teremos também a interação entre quarks e glúons, ou entre glúons. Nessas interações, outros quarks serão produzidos com alto momento e energia, possibilitando que os mesmos emitam glúons que, por sua vez irão decair em mésons e hadrons ao redor da direção de propagação de cada quark criado na interação próton-próton. Desta forma, cada quark primário dará origem a um jato de partículas.

Jatos também estão presentes no decaimento de partículas. Um bom exemplo é o bóson de Higgs, que entre seus canais de decaimento possui o $H \rightarrow b\bar{b}$, onde cada um dos quarks b irá gerar um jato. Reconstruindo a informação proveniente de cada um desses jatos, nos permitirá reconstruir a massa do bóson de Higgs.

A reconstrução da energia é feita utilizando técnicas de análise *offline* a partir das informações provenientes do sistema de calorimetria, portanto faz-se necessária uma análise das informações mais relevantes para a reconstrução *offline* de energia.

Partículas interagem com os calorímetros do sistema de calorimetria, criando uma cascata de interações secundárias, as partículas resultantes dessas interações são lidas pelo sistema de calorimetria e assim se pode reconstruir o processo de decaimento da partícula original.

1.2 Objetivo

Ao se realizar uma medida, o erro associado a ela pode ser decomposto em dois componentes: um componente associado à técnica de medida e outro componente ao erro estatístico. O segundo componente pode ser modelado com dependência do inverso da raiz quadrada do número de eventos medidos, o que faz com que à medida que o número de eventos aumenta, o erro estatístico sofra uma redução. Já o erro associado à técnica de medida não pode ser reduzido a não ser por um processo de correção externa.

As redes neurais artificiais são amplamente utilizadas no ambiente de física de altas energias, uma vez que são mapeadores universais. Sabe-se que a distribuição de massa do bóson de Higgs é não-gaussiana, ou seja, para que se consiga acessar todos os níveis de informações, deveremos nos utilizar de estatística de ordem superior (HOS).

Este trabalho tem como objetivo melhorar a resolução de energia para os jatos do quark b registrados pelo detector ATLAS. Para isso, serão utilizadas as informações reconstruídas a partir de simulações de decaimentos do bóson de Higgs e de simulações desses decaimentos propagados através do detector ATLAS. As informações reconstruídas propagadas através do detector alimentam uma rede neural que será responsável por gerar um fator de correção a partir de um mapeamento não-linear.

1.3 Contribuições do Trabalho

Este trabalho tem como contribuições a aplicação da técnica de redes neurais artificiais em um ambiente de análise de física de altas energias, bem como a inserção do conceito de divergência aplicado física de altas energias e, também, o aprofundamento da modelagem de massa do bóson de Higgs.

1.4 Organização da dissertação

No segundo capítulo serão apresentados o CERN, o detector ATLAS e seus subsistemas e suas características. No terceiro capítulo serão apresentados o Tevatron, os

detectores CDF e $D\emptyset$, suas características, bem como técnicas de correções para jatos do quark b aplicadas nestes detectores, que serviram de inspiração para a correção gerada nesta dissertação. A metodologia de análise será descrita no capítulo quatro, bem como os dados utilizados, os modelos neurais em que foram baseados os corretores e os modelos de ajuste e cálculo de massa na física de altas energias. No quinto capítulo serão expostos e analisados os resultados obtidos com os modelos neurais e por fim, as conclusões e trabalhos futuros se encontram no capítulo 6, que é seguido da bibliografia utilizada.

Capítulo 2

Física de Altas Energias

Neste capítulo, alguns conceitos de física de altas energias serão introduzidos, o CERN [2, 3], bem como o ATLAS, um dos experimentos do LHC e seus subsistemas, também serão detalhamente apresentados.

2.1 Física Experimental de Altas Energias

A física experimental de altas energias, ou simplesmente física de altas energias, é a área da física que destina-se ao estudo das propriedades fundamentais da matéria, bem como seus componentes e as interações realizadas pelos mesmos. Pode-se dizer que desde a Grécia Antiga, o ser humano tenta estabelecer uma teoria que explicasse a matéria a sua volta, como, por exemplo, a definição do átomo por Demócrito, que posteriormente seria contestada por pesquisadores como Dalton e Faraday [4].

Atualmente, a teoria mais aceita na comunidade científica é o Modelo Padrão, ou SM (*Standard Model*) [4]. Esse modelo descreve as interações entre as partículas mais fundamentais, a partir de duas classes básicas: os férmions e os bósons. Todas as partículas elementares, ou seja, as que originam outras, estão classificadas dentro desses dois grupos principais.

Os férmions são partículas que obedecem à estatística de Fermi-Dirac [4]. Este grupo é modelado como sendo composto por quarks e léptons. Os léptons são partículas de spin semi-inteiro supostamente pontuais, sem estrutura interna e detectáveis experimentalmente. Por exemplo, o elétron é classificado como lépton. Já os quarks não são detectáveis experimentalmente enquanto partículas isoladas.

Os quarks formam os hadrons, e estes sim, são detectáveis experimentalmente. Os hadrons se subdividem em bárions e mésons, sendo os bárions formados por três quarks e os mésons por um par quark-antiquark. Nos mésons, o antiquark não necessariamente é a anti-partícula do quark formador do par. Por exemplo, o π^+ é formado por um par $u\bar{d}$. Lembrando que o píon foi proposto por H. Yukawa [5] e

observado pela primeira vez por uma equipe da qual fazia parte o físico brasileiro César Lattes [6].

Os bósons são partículas de spin inteiro que obedecem a estatística de Bose-Einstein [4]. Os mediadores das interações fundamentais entre partículas são bósons. O mediador da força eletromagnética é o fóton, os mediadores da interação fraca são W^\pm e o Z^0 , os mediadores da interação forte são os glúons e o mediador da interação gravitacional é o gráviton, até hoje não detectado experimentalmente. Na figura 2.1 é mostrada uma tabela com os quarks, léptons e os bósons que formam o Modelo Padrão.

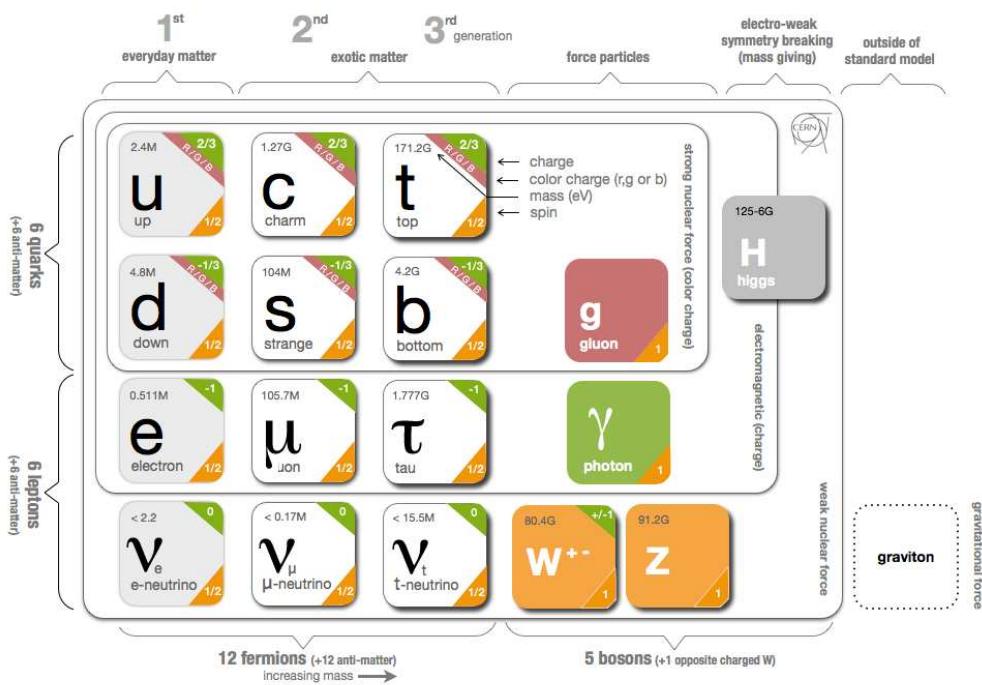


Figura 2.1: Tabela com os componentes do Modelo Padrão (Extraído de [1]).

Em julho de 2012, foi anunciada, por pesquisadores do CERN, a observação de uma partícula que pode ser o bóson de Higgs [7, 8]. Esta partícula foi chamada de “*higgs-like*”. Tal partícula foi observada a partir de seu decaimento em dois fótons e sua massa foi medida como sendo 125 GeV.

Quando uma partícula instável decai de diversas maneiras diferentes, cada uma dessas possibilidades de decaimento é chamada de canal de decaimento. O *branching ratio* (razão de ramificação) de um canal de decaimento é obtido a partir da quantidade de vezes que o referido canal de decaimento ocorreu dividido pela quantidade total de decaimentos. Ou seja, ele pode ser interpretado como sendo a probabilidade de uma partícula decair em um determinado canal de decaimento.

A teoria do bóson de Higgs foi elaborada para permitir a geração de massa

no Modelo Padrão. Segundo essa teoria, o bóson de Higgs possui diversos canais de decaimento, como mostrado na figura 2.2, onde podemos ver as probabilidades teóricas de ocorrência de cada decaimento em função da massa do bóson de Higgs. A massa de 125 GeV está assinalada para facilitar a visualização. Já na tabela 2.1 pode-se ver os valores teóricos de *branching ratios* para cada um dos decaimentos do bóson de Higgs [9]. Para que a partícula anunciada em julho do 2012 seja atestada como o bóson de Higgs, ela deve ser observada em todos esses canais de decaimento, com essas probabilidades.

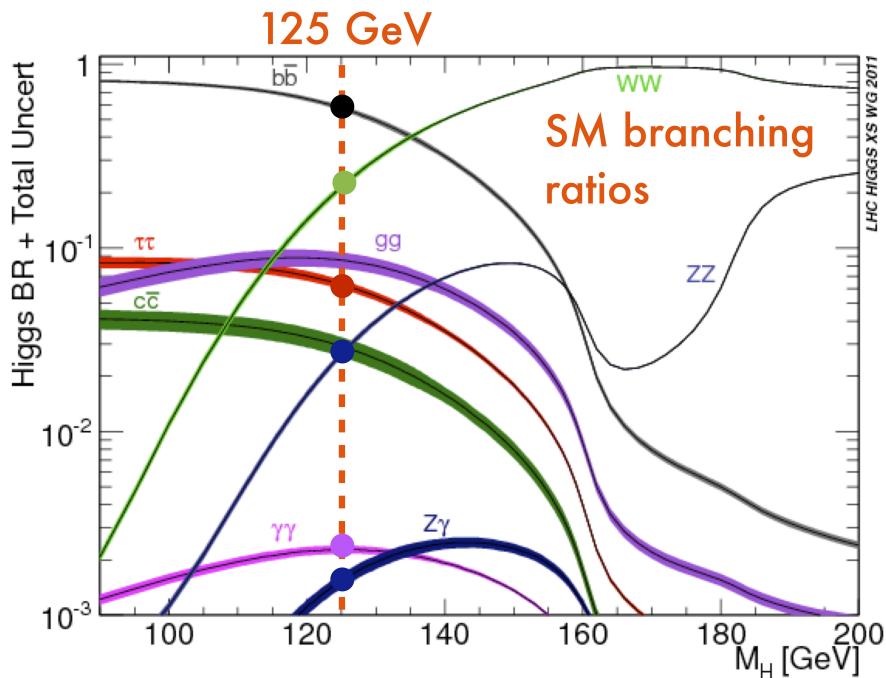


Figura 2.2: Decaimentos do Bóson de Higgs e seus respectivos *branching ratios*. Extraído de [8]

Como pode ser visto na figura 2.2 e na tabela 2.1, o decaimento mais provável para a massa de $125 \text{ GeV}/c^2$ é $b\bar{b}$ com o valor de aproximadamente 68,4%.

2.2 CERN e o LHC

O CERN [2], criado no ano de 1954, está localizado a noroeste de Genebra, na fronteira da França com a Suíça. É o maior laboratório de física de partículas do mundo. No CERN, encontram-se os detectores e aceleradores de partículas, utilizados para pesquisa de partículas fundamentais. Os aceleradores impulsionam feixes de partículas para altas energias, para que colidam umas com as outras ou contra alvos estacionários, enquanto detectores registram os resultados dessas colisões. Atualmente, o acelerador em operação é o LHC [3, 10] que tem 27 km de perímetro e permite que prótons possam colidir com até 14 TeV de energia no centro de massa.

Decaimento	Possível Massa (GeV)	<i>branching ratio</i>
$H \rightarrow b\bar{b}$	120	$6.48 \cdot 10^{-01}$
	150	$1.57 \cdot 10^{-01}$
	200	$2.40 \cdot 10^{-03}$
	500	$1.09 \cdot 10^{-04}$
$H \rightarrow \tau^+\tau^-$	120	$7.04 \cdot 10^{-02}$
	150	$1.79 \cdot 10^{-02}$
	200	$2.87 \cdot 10^{-04}$
	500	$1.53 \cdot 10^{-05}$
$H \rightarrow \mu^+\mu^-$	120	$2.44 \cdot 10^{-04}$
	150	$6.19 \cdot 10^{-05}$
	200	$9.96 \cdot 10^{-07}$
	500	$5.31 \cdot 10^{-08}$
$H \rightarrow c\bar{c}$	120	$3.27 \cdot 10^{-02}$
	150	$7.93 \cdot 10^{-03}$
	200	$1.21 \cdot 10^{-04}$
	500	$5.47 \cdot 10^{-06}$
$H \rightarrow t\bar{t}$	350	$1.56 \cdot 10^{-02}$
	360	$5.14 \cdot 10^{-02}$
	400	$1.48 \cdot 10^{-01}$
	500	$1.92 \cdot 10^{-01}$
$H \rightarrow gg$	120	$8.82 \cdot 10^{-02}$
	150	$3.46 \cdot 10^{-02}$
	200	$9.26 \cdot 10^{-04}$
	500	$6.04 \cdot 10^{-04}$
$H \rightarrow \gamma\gamma$	120	$2.23 \cdot 10^{-03}$
	150	$1.37 \cdot 10^{-03}$
	200	$5.51 \cdot 10^{-05}$
	500	$3.12 \cdot 10^{-07}$
$H \rightarrow Z\gamma$	120	$1.11 \cdot 10^{-03}$
	150	$2.31 \cdot 10^{-03}$
	200	$1.75 \cdot 10^{-04}$
	500	$7.58 \cdot 10^{-06}$
$H \rightarrow WW$	120	$1.41 \cdot 10^{-01}$
	150	$6.96 \cdot 10^{-01}$
	200	$7.41 \cdot 10^{-01}$
	500	$5.46 \cdot 10^{-01}$
$H \rightarrow ZZ$	120	$1.59 \cdot 10^{-02}$
	150	$8.25 \cdot 10^{-02}$
	200	$2.55 \cdot 10^{-01}$
	500	$2.61 \cdot 10^{-01}$

Tabela 2.1: Decaimentos do Bóson de Higgs e seus respectivos *branching ratios*. Valores extraídos de [8]

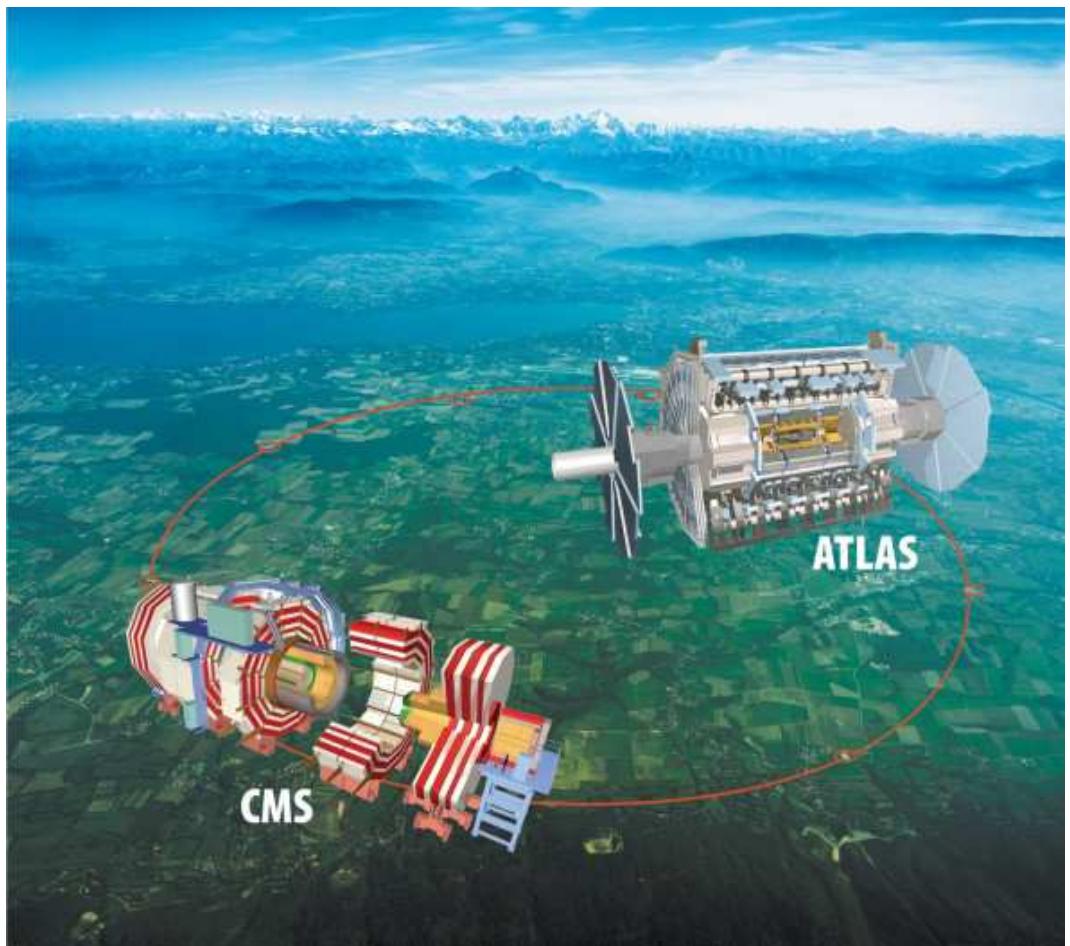


Figura 2.3: Visão Aérea do CERN, com destaque para os dois detectores de propósito geral (Extraído de [1]).

Íons de chumbo também são acelerados no LHC, com energia de 2,76 TeV por *nucleon* (575 TeV por íon). Na figura 2.3 pode ser vista a visão aérea do LHC com destaque para os dois detectores de propósito geral.

O LHC tem capacidade para acelerar partículas (prótons ou íons de chumbo) a até 99,999991% da velocidade da luz e, segundo os cálculos, seremos capazes de reconstruir um ambiente similar ao Universo apenas 10^{-12} s após o Big Bang, a suposta explosão que deu origem ao Universo [11]. Em cada um dos pontos de colisão do LHC, estão colocados detectores. Os principais são: o ATLAS [12–15], o CMS [16], o LHCb [17] e o ALICE [18].

O ATLAS e o CMS são detectores de propósitos gerais, ou seja, estudam toda a física envolvida nas colisões próton-próton a 14 GeV de energia de centro de massa. Entre os tópicos estudados, podemos citar a procura do bóson de Higgs, a detecção de uma partícula “higgs-like” em 2012, a procura de partículas previstas por modelos além do Modelo Padrão. O LHCb visa estudar minuciosamente todos os decaimentos e características do quark *b* e o ALICE, a física de íons pesados.

2.3 O Detector ATLAS

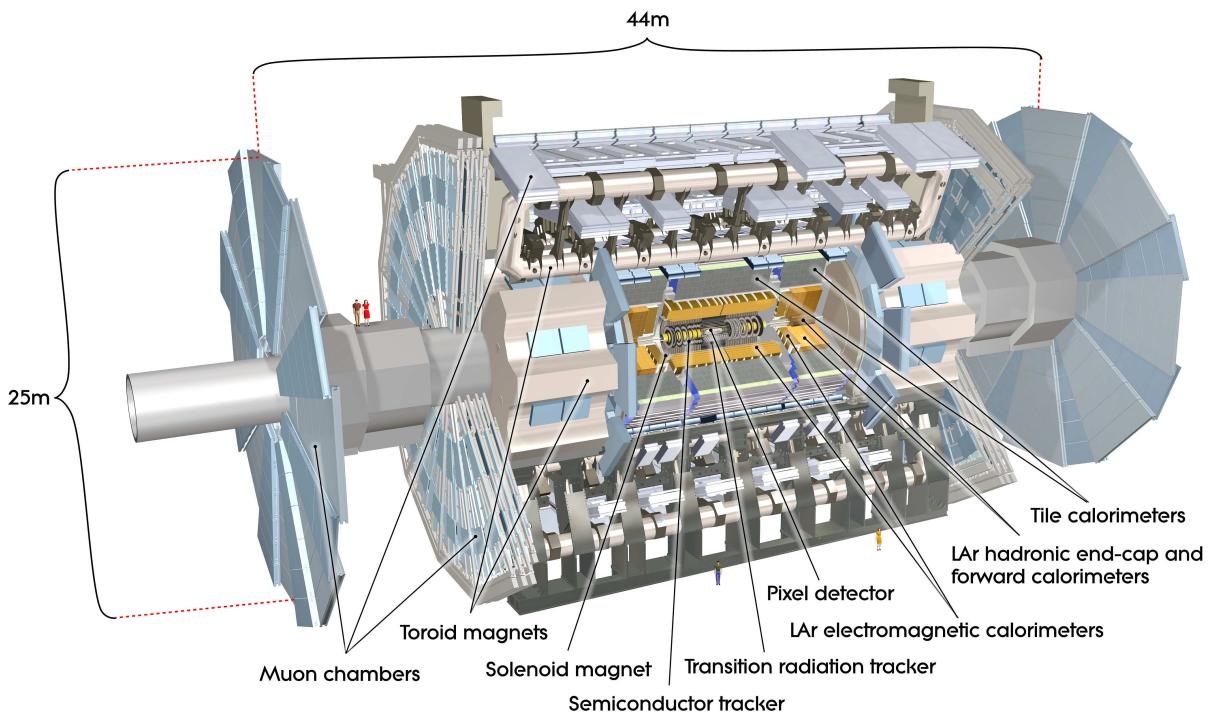


Figura 2.4: O Detector ATLAS (Extraído de [1]).

O ATLAS tem 44 metros de comprimento por 25 metros de altura e aproximadamente 7.000 toneladas distribuídas em seu formato cilíndrico, como pode ser visto na figura 2.4. O detector é formado por um conjunto de sub-sistemas, cada um com suas características próprias e uma infinidade de funcionalidades. Os principais são: o Inner Detector [19, 20], o calorímetro eletromagnético de Argônio Líquido (LAr) [21, 22], o calorímetro hadrônico de Telhas cintilantes (TileCal) [23–25] e o Sistema de Múons [26–28].

2.3.1 Sistema de Coordenadas do ATLAS

O detector ATLAS possui um sistema de coordenadas cilíndricas próprio [14], que visa facilitar os cálculos na área de física de altas energias. Suas coordenadas estão baseadas em duas variáveis: η e ϕ .

Define-se ϕ como sendo o ângulo formado pela projeção da trajetória da partícula no plano perpendicular ao eixo formado pelo feixe de partículas, no ponto de colisão, isto é, o ângulo radial, e θ é o ângulo formado entre a direção do feixe e a trajetória da partícula, como mostrado na figura 2.5.

Sendo assim, η é obtido por meio da equação 2.1. Como as duas coordenadas são baseadas em dois ângulos, podemos mapear qualquer ponto no espaço dentro do detector.

$$\eta = -\log [\tan(\theta/2)] \quad (2.1)$$

Por definição, o ATLAS foi dividido em duas partes para uma melhor manutenção e monitoramento. O lado A do detector representa o lado que possui η positivo, enquanto o lado C possui η negativo.

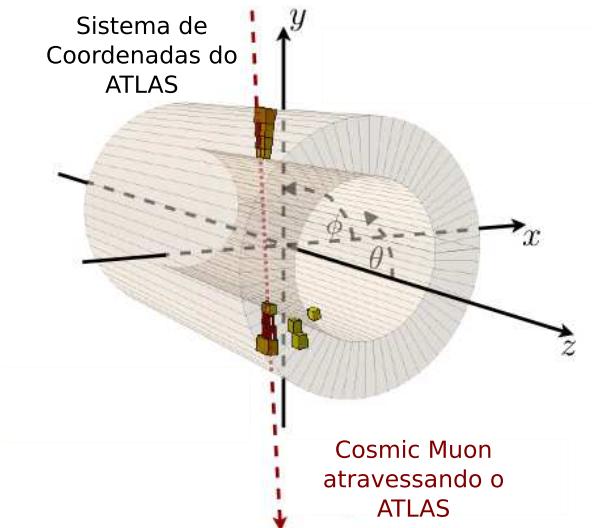


Figura 2.5: Sistema de Coordenadas do ATLAS.

2.3.2 Detector Interno

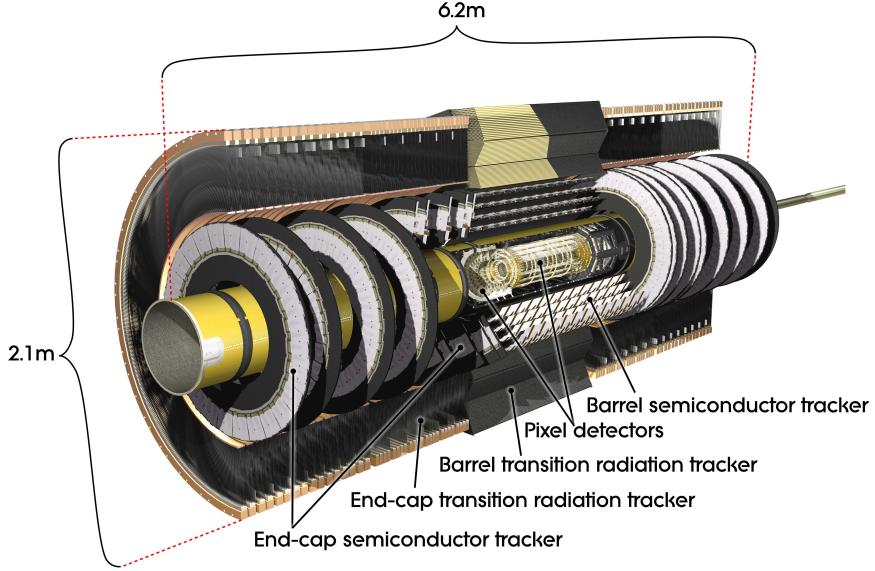


Figura 2.6: Visão Geral do Inner Detector (Extraído de [1]).

O Detector Interno ou *Inner Detector* [20] é a parte mais interior do detector e tem como principal objetivo a reconstrução da posição e do momento de partículas carregadas provenientes das colisões e, assim, permitir a reconstrução do traço, ou caminho percorrido pelas mesmas. Devido à sua alta resolução, que varia para cada um dos seus três sub-sistemas, o Inner Detector é capaz também de reconstruir os vértices, que por definição, são pontos onde ocorrem interações ou ínicio do decaimento de partículas. O Detector Interno possui 7 metros de comprimento, 115 centímetros de diâmetro e com sua configuração cilíndrica, cobre a região com $|\eta| < 2,5$.

O primeiro sub-sistema do detector de traços, como também é conhecido o Detector Interno, é o Pixel Detector [29]. Este sub-sistema é o que possui a mais fina resolução dentro do sistema, da ordem de 10^{-6} m. Utiliza uma estrutura de *pixels* de silício, com $25 \mu\text{m}$ de espessura como célula de leitura.

O segundo sub-sistema é o Semi Condutor Tracker (SCT) [30] que possui a mesma estrutura de leitura do Pixel Detector, com células de fibra de silício semelhante às do Pixel Detector. As células que compõem o SCT têm granularidade menos fina que as células que compõem o Pixel Detector, o que faz com que o SCT tenha uma resolução total menor. Somados, os dois primeiros sub-sistemas possuem 56 cm de diâmetro.

O último sub-sistema do Inner Detector é o Transition Radiation Tracker (TRT) [31], que é composto por 420.000 fibras de 4 mm de diâmetro cada. Estas fibras são

preenchidas por um material que emite radiação de transição quando uma partícula o transpassa. A reação gerada pela partícula carregada ao passar pela fibra preenchida permite a leitura dos dados do traço nessa região. Cada fibra possui $170\ \mu\text{m}$ de resolução.

O Inner Detector está localizado dentro do solenóide do ATLAS, como mostrado na figura 2.6, isto é, o sistema de detecção de traços está sujeito ao campo de 2 Teslas gerado pelo solenóide. Já na figura 2.7 vemos como cada uma das partes do Detector Interno está localizada em um corte perpendicular a linha do feixe.

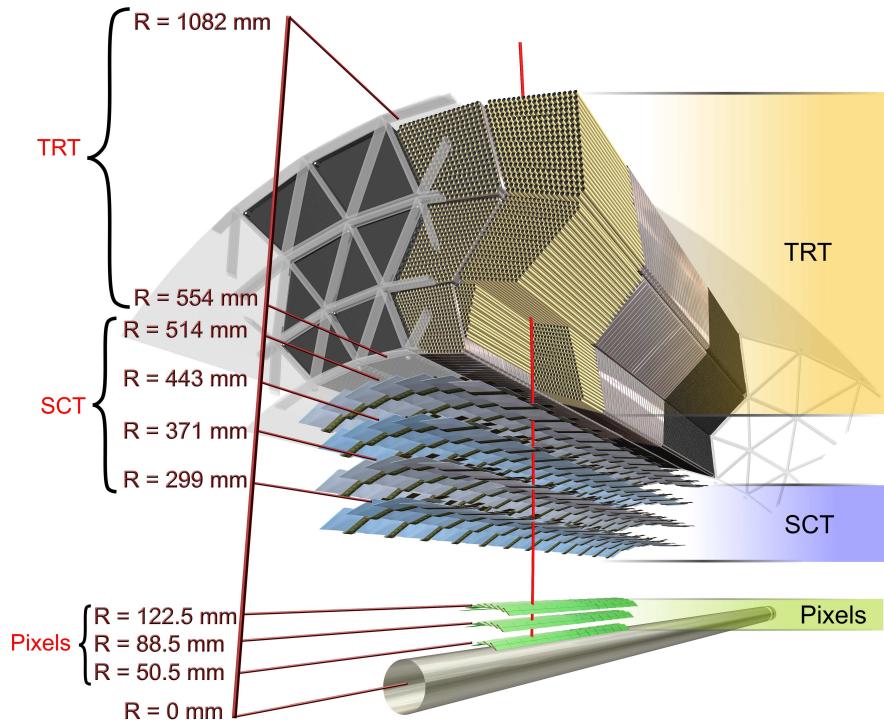


Figura 2.7: Visão do detector geral em um corte perpendicular a linha do feixe (Extraído de [1]).

2.3.3 Sistema de Calorimetria

O calorímetro, em física de altas energias, tem por finalidade a medição da energia de partículas provenientes de colisões por meio de sua absorção total por um bloco de matéria [32].

Existem, basicamente, dois tipos de calorímetros: os homogêneos e os calorímetros de amostragem. Nos calorímetros homogêneos, todo o material que constitui o calorímetro é sensível a interações com as partículas e, por isso, todo

o material irá contribuir para os sinais produzidos pelo calorímetro, ou seja, todo o material que compõe o calorímetro irá absorver e detectar a energia produzida pelos subprodutos da colisão. Para que este material desempenhe estas duas tarefas, o mesmo deve possuir uma alta densidade.

Nos calorímetros de amostragem, apenas uma parte do material constituinte interage com as partículas. Este material é chamado material passivo e possui uma alta densidade (por exemplo ferro, chumbo ou urânio), já a outra parte do material é responsável por amostrar a informação de energia das partículas incidentes, sendo este material conhecido como material ativo. Nos calorímetros de amostragem apenas uma pequena fração da energia é amostrada para gerar o sinal de saída. No sistema de calorimetria do ATLAS, os dois calorímetros são de amostragem, como pode ser visto na figura 2.8.

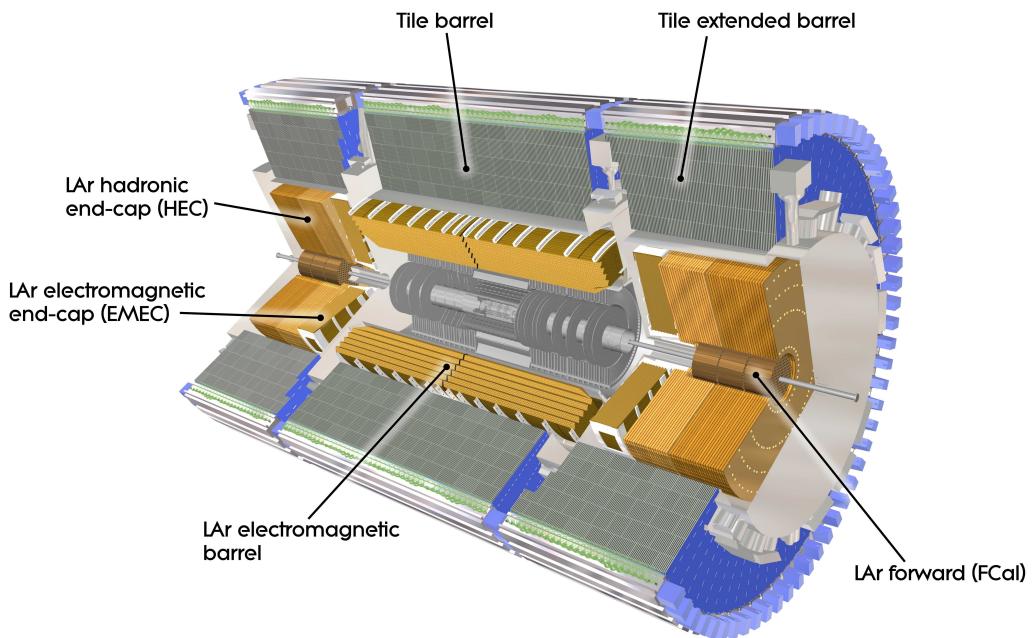


Figura 2.8: Visão Geral do Sistema de Calorimetria do detector ATLAS (Extraído de [1]).

Calorímetro Eletromagnético de Argônio Líquido

Como dito anteriormente, o sistema de calorimetria do ATLAS é dividido em duas partes. A parte eletromagnética é composta pelos calorímetros de Argônio Líquido, também chamados de LAr. Os calorímetros de Argônio Líquido são divididos em uma parte central ou barril, dois *end-caps* e dois calorímetros especiais, chamados de *Forward Calorimeters*.

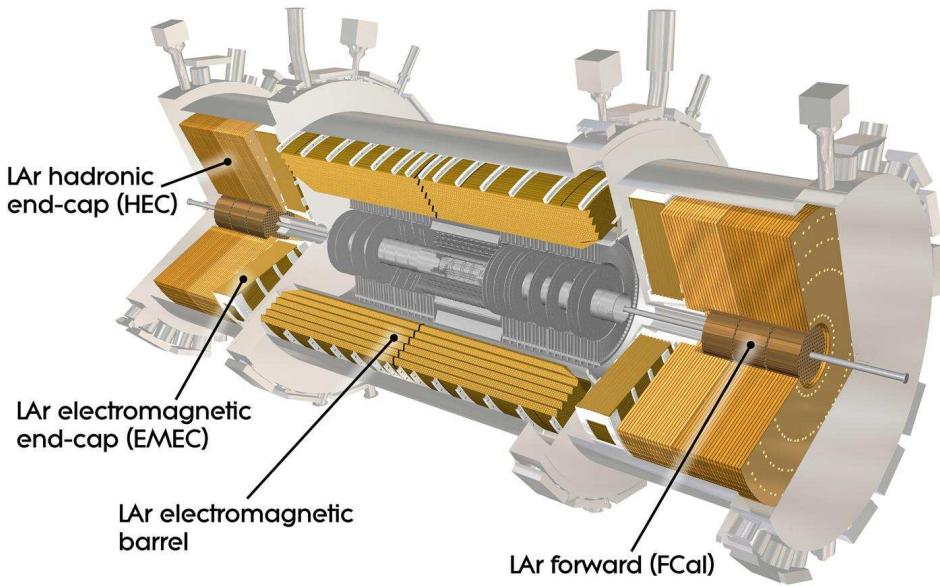


Figura 2.9: Visão Geral do Calorímetro Eletromagnético do ATLAS (Extraído de [1]).

O sistema de calorimetria do ATLAS é um sistema de amostragem, isto é, apenas uma parte da energia das partículas será lida pelos sensores.

Os calorímetros de Argônio líquido são divididos em várias partes: a parte central, também chamada de barril, que cobre $|\eta| < 1,475$; o *end-cap*, é dividido em: *end-cap* eletromagnético chamado de EMEC e em *end-cap* hadrônico chamado de HEC, que cobre $1,5 < |\eta| < 3,2$ e o *forward calorimeter*, também chamado de FCAL, que cobre $|\eta|$ entre 3,1 e 4,9. Em cada calorímetro temos diferentes materiais passivos bem como temos diferentes materiais ativos, que serão discriminados e explicados a seguir.

O barril do LAr, os *end-caps* e os *forward calorimeters* possuem como material ativo, o argônio líquido. Como material passivo temos, chumbo no LAr, cobre nos *end-caps* e tungstênio nos *forward calorimeters*. Como o material de leitura é o argônio líquido são necessárias câmaras criostáticas para manter o calorímetro na temperatura ideal de funcionamento. A escolha do argônio líquido está diretamente ligada a robustez, no que concerne a radioatividade, a técnica de leitura do calorímetro e a redução da presença de ruído térmico.

O barril central é composto por 2 barris idênticos no formato de acordeão, como mostrado na figura 2.9. Todo barril é feito em 1024 camadas de metal intercaladas com 1024 eletrôdos de leitura e toda essa estrutura está comprimida entre 2 camadas de aço inoxidável, como mostrado na figura 2.10. Além disso, o barril central tem

como característica uma boa precisão, isto é, tem a capacidade de determinar muito precisamente a quantidade de energia depositada e o ponto de sua deposição. A leitura dos dados provenientes do barril é feita em três camadas, que são chamadas *front*, *middle* e *back* (anterior, central e posterior).

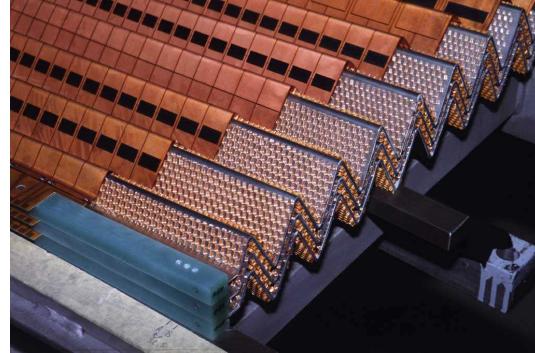


Figura 2.10: Células de Leitura da parte central do Calorímetro LAr (Extraído de [1]).

Calorímetro Hadrônico de Telhas Cintilhantes

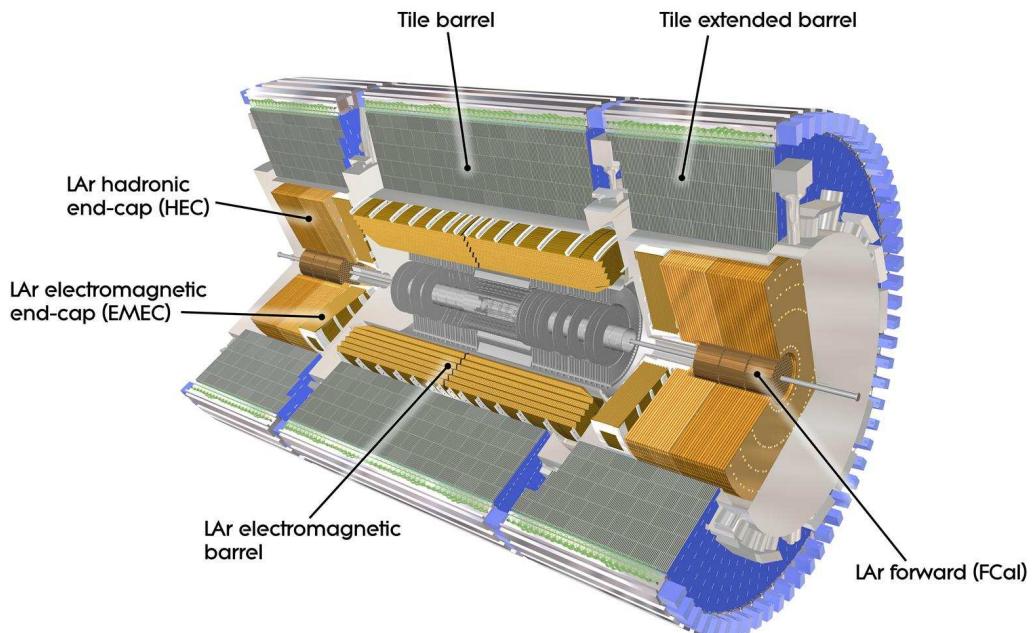


Figura 2.11: Visão Geral do Calorímetro Hadrônico do ATLAS (Extraído de [1]).

O TileCal ou *Tile Calorimeter* é um calorímetro hadrônico, composto 3 barris, um central, que tem aproximadamente 5,6 metros de extensão e outros dois externos com 2,9 metros, cada um, como pode ser visto na figura 2.11. Cada Barril Externo

(*Extended Barrel - EBA e EBC*) possui 64 *Super-Drawers* que suportam toda a eletrônica necessária para a aquisição dos dados, também chamada de eletrônica de *front-end*. Cada *Super-Drawer* é por sua vez dividida em 2 *Drawers* (Interna e Externa) com cerca de 1,5 metros de comprimento, facilitando assim seu manuseio e transporte. O Barril Central possui 128 *Super-Drawers* de modo a registrar os dados dos seus 64 módulos de 5,6 metros de comprimento. Os Módulos do TileCal são dispositivos de amostragem, compostos por chapas de ferro (parte passiva) que se intercalam com chapas do material cintilante (parte ativa) que amostram a componente hadrônica das partículas que interagem com o detector ATLAS.

Cadeia de Leitura do TileCal

As partículas (hádrons) que incidirem no TileCal excitarão o material cintilante que por sua vez emitirá luz [33]. A luz gerada por esse evento é diretamente proporcional à energia da partícula incidente. Esse sinal luminoso é captado por fibras ópticas, chamadas de *Wavelength Shifting Fibers* (WLS fibers). Essas fibras foram projetadas para mudar o comprimento de onda do sinal recebido, pois as telhas cintilantes do TileCal, quando excitadas, emitem luz na faixa do azul/violeta, embora os PMTs¹ tenham a resposta ótima na faixa do verde. As WLS fibers foram projetadas para mudar o comprimento de onde do sinal emitido pelas telhas afim de maximizar a resposta das PMTs. Estas fibras estão instaladas junto ao material cintilante, seguindo uma organização pré-definida.

Cada sinal luminoso amostrado pelas telhas cintilantes é transmitido por duas fibras ópticas independentes, criando-se assim uma redundância do sinal amostrado. Estas fibras são organizadas seguindo uma formatação em células, como esquematizado visto na figura 2.12(b), onde pode ser visto cada região dentro de um módulo. A estrutura de um módulo montado pode ser visto na figura 2.12(a). O agrupamento das fibras ópticas provenientes de uma célula é recebido como entrada por um único bloco fotomultiplicador [24], como mostrado na figura 2.13(a).

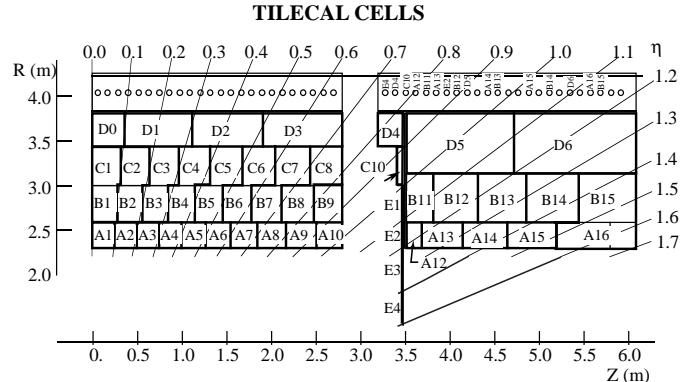
Uma vez que o sinal luminoso é recebido o bloco o transforma em um sinal elétrico, proporcional à intensidade luminosa recebida e que posteriormente será lido por um circuito eletrônico. Cada bloco fotomultiplicador é formado por um Light Mixer, uma PMT (modelo Hamamatsu R7877), um *HV divider* e uma placa 3 em 1 [34]. O bloco fotomultiplicador é o primeiro elemento da eletrônica de *front-end* e é ilustrado na figura 2.13(b).

Cada PMT tem uma tensão nominal de operação e é nessa tensão que a mesma opera com eficiência, isto é, atinge seu ganho nominal, que é da ordem de 10^6 . Como o modelo Hamamatsu possui tensões nominais variando entre 600 e 800 Volts, foi

¹PMT ou *Photo Multiplier Tube* são os elementos responsáveis pelo ganho luminoso na cadeia de leitura do TileCal.



(a) Módulo do Barril Extendido do TileCal



(b) Distribuição das células nos módulos do TileCal

Figura 2.12: Descrição de um Módulo de leitura do TileCal (Figuras extraídas de [1]).



(a) Fibras Ópticas do TileCal



(b) Bloco PMT

Figura 2.13: Componentes do Calorímetro Hadrônico do ATLAS (Figuras extraídas de [1]).

desenvolvido um sistema de alimentação que consiste, basicamente, de três placas: a *HVMicro*, a *HVOpto* e a *HVBus*. A *HVMicro* possui os valores nominais de voltagem de cada PMT em seu raio de ação e assim, quando recebe aproximadamente -830 Volts do sistema de alimentação central, os repassa para a *HVOpto* juntamente com o valor de tensão e esta faz com que a PMT seja alimentada a partir da *HVBus*. Cada *Drawer* possui uma *HVMicro* e duas *HVOpto* e duas *HVBus*.

O sinal da PMT é lido pela placa 3em1, que possui três funções como o nome diz: i) Condicionar o sinal para ser integrado. ii) Permitir a injeção de carga para calibração e teste dos sistemas eletrônicos do TileCal. iii) Integrar o impulso de corrente (saída da PMT) para a saída em tensão em uma janela de tempo pré-definida.

A placa 3 em 1 possui 2 saídas primárias: uma para baixo ganho e outra para alto ganho. A saída de baixo ganho é lida pelos somadores produzidos pela UFRJ e pelos digitalizadores, enquanto a saída de alto ganho é lida apenas pelos digitalizadores. Os somadores integram o primeiro nível de validação de eventos do detector, este é responsável por determinar a $40MHz$, se houve ou não algum evento relevante para análise.

O Digitalizador recebe o sinal de ganho alto e ganho baixo vindo de 6 placas 3em1 distintas, o que totaliza 60 bits de sinais, que são alinhados em uma pipeline aguardando a decisão do primeiro nível de validação. É no digitalizador onde é decidido se o sinal deve ser tratado como ganho baixo ou ganho alto. Se o evento for aceito, um número fixo de amostras será enviado através de um link de fibra óptica para os RODs (“*Read-Out Drivers*”) juntamente com o sinal de validação ou sinal TTC (“*Timing, Trigger and Control*”).

A eletrônica de *back-end* é composta por dois sistemas básicos: o ROD e a TTC. O sistema ROD é composto por 8 *Optical Receivers* (ORs), por 8 deserializadores HDMP, por 4 *Staging FPGA*, por 2 *Processing Units* (PU), 4 *Output Controllers*, uma VME FPGA e uma TTC FPGA. Já o sistema TTC é composto pela TBM, o LTP, a TTCvi, a TTCEx, o TTCPri e pela ROD Busy. Ambos os calorímetros do detector ATLAS utilizam a mesma estrutura de leitura de dados, isto é, o mesmo sistema ROD.

Os ORs recebem o sinal luminoso proveniente do último estágio da eletrônica de *front-end* e o transformam em sinal elétrico que será repassado para os deserializadores, que por sua vez, como o nome diz, irão deserializar o sinal por intermédio do chip HDMP-1024 com clock de 40 MHz. Uma vez que o sinal esteja deserializado, as *Staging FPGAs* criam uma rota direta para a entrada de dados provenientes de 4 ORs em uma única PU.

Cada ROD possui 4 PUs, pois foram inicialmente projetadas para atender as necessidades do calorímetro eletromagnético do ATLAS, mas seu projeto foi reutili-

zado para o calorímetro hadrônico. Sendo assim, para esta finalidade, são utilizadas apenas 2 PUs.

Cada PU é equipada com 2 DSPs (*Digital Signal Processor*) do modelo TMS320C6414 @720 MHz [35], 2 Input FPGAs, 2 FIFOs e uma *Output FPGA*. Cada DSP é responsável pelo processamento dos dados provenientes de 2 ORs, segundo um código previamente estabelecido, as Input FPGAs estão encarregadas de verificar se houve alguma corrupção dos dados recebidos e por um pré-processamento que preparará os dados para os DSPs.

A *Output FPGA* é responsável pelo armazenamento dos eventos na FIFO (*First In, First Out*), que suporta até 16 eventos. A FIFO só pode enviar seus eventos para o *Output Controller* quando receber o sinal de validação gerado pela TTC e enviado através do barramento VME para a *Output FPGA*.

O *Output Controller* (OC) é o responsável por exportar os dados do sistema ROD. Uma vez que o sistema receba o sinal de aceitação do evento, o OC gera uma estrutura de compactação de dados chamada ROD Fragment que contém as informações sobre o evento lido pelo TileCal. Após o sinal de aceitação do evento essa estrutura é exportada, através de um link VME, a taxa máxima de 75 kHz.

2.3.4 Sistema de Múons

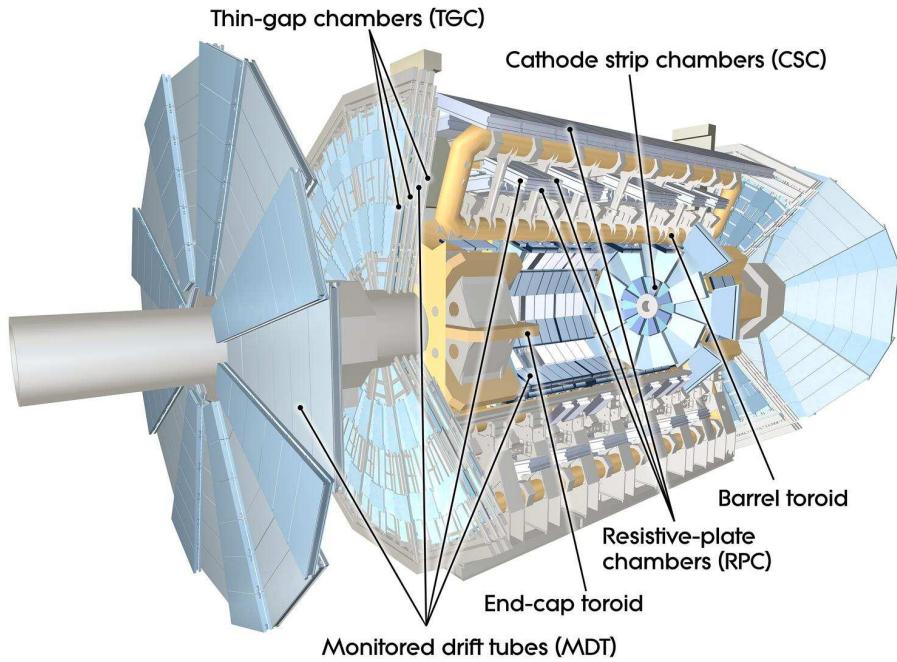


Figura 2.14: Visão Geral do Muon System (Extraído de [1]).

O sistema de múons ou *Muon System* [27] é o sub-sistema mais afastado do ponto de colisão, e tem por objetivos a reconstrução e a identificação de traços dos múons,

bem como, a medição de seus momentos. Como mísseis são partículas que depositam uma pequena quantidade de energia nos calorímetros [4], a presença de um sistema, que possa fazer uma leitura mais precisa do sinal emitido pela passagem de um mísseis, se faz necessária. Uma visão geral do sub-sistema é mostrada na figura 2.14.

O Muon System também é composto por 3 tipos de câmaras como podemos ver na figura 2.15. As câmaras Monitored Drift Tube (MDT) são utilizadas para a medição da trajetória dos mísseis. As Resistive Plate Chambers (RPC) são utilizadas, em conjunto com as informações provenientes de outros sub-detectores, para filtragem de primeiro nível, enquanto as Cathode Strip Chambers (CSC) contribuem para a taxa de amostragem na região de end-cap.

Como pode ser visto na figura 2.14, algumas partes do Sistema de Mísseis estão imersas no segundo sistema magnético do detector (o toróide) que possui campo nominal de 0,5 Teslas.

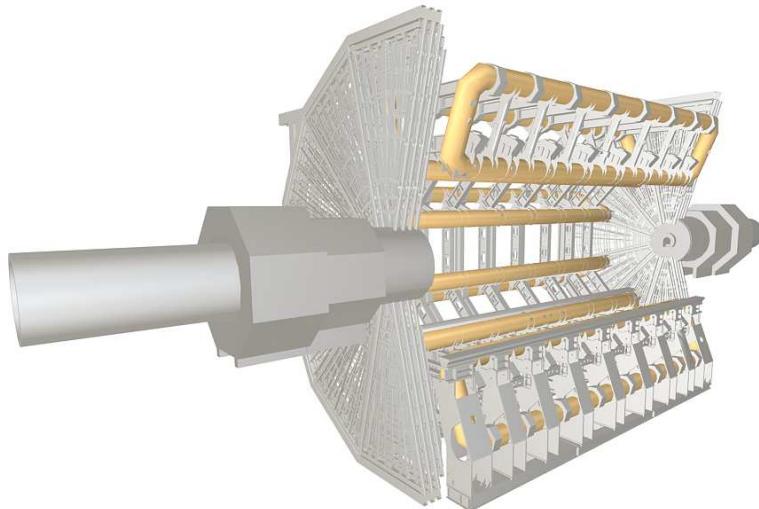


Figura 2.15: Distribuição das câmaras do Muon System (Extraído de [1]).

2.3.5 Sistema de Trigger

A maioria das reações físicas de interesse estudadas nos modernos colisionadores de partículas ocorrem com frequência bastante reduzida. Processos relevantes, como por exemplo $H \rightarrow \gamma\gamma$ (canal da observação da partícula “*higgs-like*”), ocorrem com uma frequência relativamente pequena, pois a sua probabilidade, como pode ser visto na figura 2.2 é bem pequena.

O problema da alta taxa de eventos, somado à raridade com que certos canais

físicos ocorrem é um problema compartilhado por todos os experimentos modernos envolvendo colisionadores de partículas. Consequentemente, os detetores desenvolvidos para estes colisionadores precisam ser dotados de um sofisticado sistema de filtragem [36], responsável por descartar eventos sem interesse para o experimento.

Este sistema de filtragem precisa ser cuidadosamente desenvolvido, de forma a atingir uma alta taxa de rejeição de ruído de fundo, ao mesmo tempo que possui uma alta eficiência em relação à seleção dos canais físicos importantes.

Sistema de *Trigger Online*

Em sistemas de filtragem online [37], normalmente, tem-se um sistema hierárquico de análise. Neste caso, os estágios iniciais são validados pelos estágios posteriores. A hierarquia de análise de um sistema de filtragem online é desenvolvida de forma que os níveis mais baixos façam a seleção com base em critérios mais simples, enquanto os níveis mais elevados possuem critérios mais refinados para a tomada de decisão, uma vez que estes níveis, possuem mais tempo para a decisão.

Como os estágios iniciais possuem menos tempo para a tomada de decisão, e estes sistemas de filtragem hierárquicos não podem desfazer a rejeição aplicada por um nível mais baixo, o projeto dos primeiros estágios se torna bastante complexo.

O sistema de filtragem (*trigger*) do ATLAS e o sistema de aquisição de dados são baseados em três estágios sequenciais de seleção de eventos online. Partindo de uma taxa de eventos de 40 MHz, a taxa de eventos armazenados precisa ser reduzida, ao final do processo de seleção, para aproximadamente 100 Hz, para uma posterior análise offline.

Observa-se na figura 2.16 o diagrama em blocos do sistema de trigger do ATLAS. Como se percebe, o mesmo está dividido em três níveis principais. A informação gerada pelos detetores é passada para o primeiro nível, que, com pouca informação, realiza uma análise inicial. Os eventos aprovados pelo primeiro nível são armazenados em buffers de leitura e escrita rápidas, para que possam ser acessados pelos níveis seguintes.

O segundo nível opera somente sobre as regiões de interesse marcadas pelo primeiro nível, validando sua decisão, desta vez utilizando toda a informação registrada pelo detector. Em seguida, os eventos aprovados no segundo nível são enviados ao filtro de eventos (terceiro nível), onde a informação completa de cada evento é utilizada. Por fim, os eventos que são aprovados pelo filtro de eventos são enviados para armazenamento em disco para posterior análise.

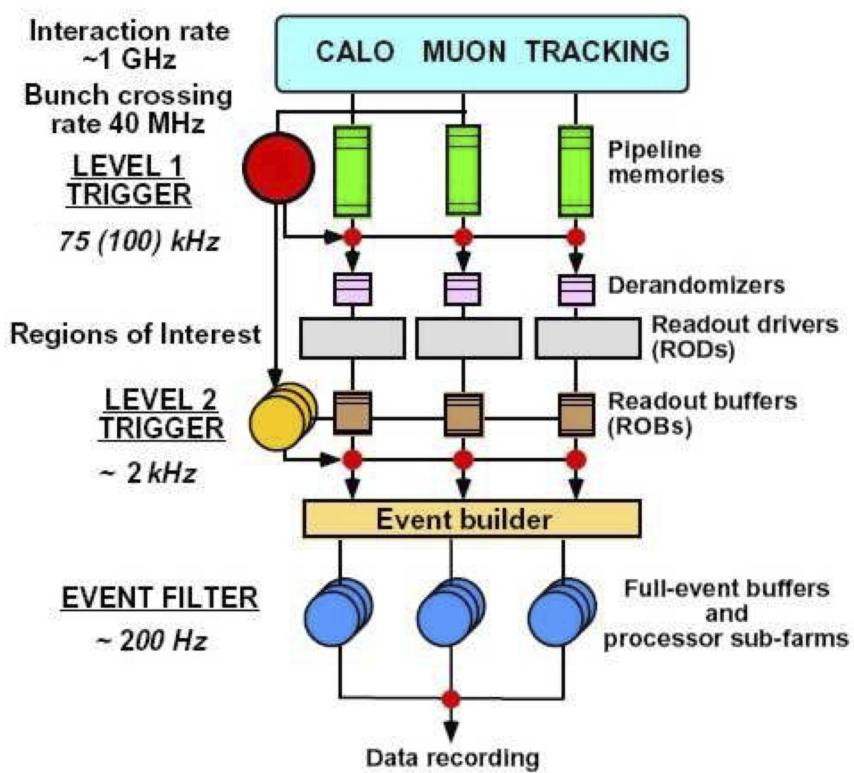


Figura 2.16: Representação gráfica do sistema de *trigger* do detector ATLAS (Extraído de [1]).

Capítulo 3

Técnicas para Correções de Energia

Neste capítulo serão apresentadas as técnicas de correção para a medida da energia dos jatos produzidos em colisões de altas energias. Primeiramente, vamos apresentar um exemplo de correção de jatos em diversos níveis que é feita no detector CDF (*Collider Detector at Fermilab*) do acelerador Tevatron, no Fermilab ou Fermi National Accelerator Laboratory, EUA. Apresentaremos também a correção do detector $D\emptyset$, também do Tevatron [38].

3.1 Correções de Energia

Como os detectores são sistemas de medições reais tem-se, naturalmente, regiões sem leitura, por onde passam os cabos elétricos, ou nas junções entre 2 módulos, bem como a presença de não-linearidades inerentes. Todos esses efeitos devem ser compensados e/ou corrigidos para que possamos ter a reconstrução da trajetória, do momento e da energia das partículas com acurácia e precisão necessárias a um experimento desse porte.

3.1.1 Resolução em Energia

Uma das principais características de uma partícula é sua energia. Por essa razão este parâmetro possui uma grande importância na reconstrução dos eventos, bem como em todas as análises dos mesmos. Ao longo dos anos, novas técnicas vêm sendo desenvolvidas e aperfeiçoadas a fim de melhorar a resolução na medida de energia. A construção dos calorímetros tem esse como um de seus pontos principais. Além disso, *softwares* específicos para corrigir e/ou compensar perdas nos dados registrados pelos calorímetros também são utilizados na reconstrução dos dados. Diversos fatores entram na composição da resolução em energia e afetam diretamente

a performance do detector.

A resolução está relacionada com as flutuações na resposta da medida da energia do calorímetro para partículas do mesmo tipo e mesma energia nominal. Em um calorímetro ideal, com comprimento infinito e homogêneo, a resolução de energia só é influenciada pela quantidade de partículas que incide no mesmo, como mostrado na equação 3.1, onde N é o número de partículas. Podemos definir também um fluxo de partículas, conhecido como ι na equação 3.2, onde N_{MAX} é uma constante que depende do número de partículas e não dos componentes do detector, assim sendo, com esta nova variável podemos definir a resolução de energia de um calorímetro ideal como na equação 3.3.

$$\frac{\sigma(E)}{E} = 1/\sqrt[3]{N} \quad (3.1)$$

$$\iota = E/N_{MAX} \quad (3.2)$$

$$\frac{\sigma(E)}{E} = \frac{1}{\iota} = 1/\sqrt[3]{E} \quad (3.3)$$

Calorímetros reais geralmente são de amostragem e assim sendo detectam apenas uma parte da energia, possuem comprimento finito e como as partículas devem depositar toda a sua energia dentro do calorímetro, tem-se outro tipo de flutuação, a flutuação de amostragem, que degrada a resolução de energia. Empiricamente, tem-se que a resolução de energia de um calorímetro real, de amostragem, com comprimento finito é dada pela equação 3.4.

$$\frac{\sigma(E)}{E} \approx \sqrt{\left(\frac{c_1}{\sqrt{E}}\right)^2 + \left(\frac{c_2}{E}\right)^2 + (c_3)^2} = \frac{c_1}{\sqrt{E}} \oplus \frac{c_2}{E} \oplus c_3 \quad (3.4)$$

Na equação 3.4 temos o termo c_1 , que representa a parte estocástica, que tenta modelar flutuações estatísticas como flutuações intrincicas ao chuveiro, contribuições na deposição de energia que não foram reconstruídas devido a materiais mortos (*dead materials*) e flutuações na amostragem do detector.

O termo c_2 é derivado do ruído eletrônico e de sua propagação na cadeia de aquisição de dados. A estimativa desse termo é feita uma vez, para cada análise específica, e seu valor é mantido constante a partir de então.

As principais contribuições para o termo c_3 são as não-homogeneidades do detector e as incertezas na calibração do mesmo. No caso de calorímetros hadrônicos, a não-compensação também contribui com o termo c_3 . Vale lembrar que a não-compensação indica a razão entre a resposta da componente eletromagnética e da componente hadrônica das partículas hadrônicas incidentes. O parâmetro e/h não

pode ser obtido diretamente das medições de energia feitas por um calorímetro, porém pode ser calculado a partir da razão e/π , que é a razão entre a energia medida de um elétron e a energia medida de um píon por um calorímetro hadrônico. Dizemos que um calorímetro é compensado quando $e/h = 1$, sendo assim a sua resposta é linear tanto para elétrons quanto para partículas hadrônicas. Quando temos diversos calorímetros em um mesmo sistema, é necessária a aplicação de uma calibração e inter-calibração dos calorímetros.

3.2 Tevatron

O Tevatron é um acelerador circular de partículas nos Estados Unidos. Está localizado no Fermilab, em Illinois, e sendo, atualmente, o segundo maior colisor de partículas do mundo, depois do LHC. O Tevatron é um sincrotron onde os prótons e antiprotons são acelerados em um anel de 6,86 km de comprimento. As colisões atingem energias de até 2 TeV de energia de centro de massa. Seu nome é derivado de 1 TeV ou um trilhão de elétrons-Volt¹.

3.2.1 Detector CDF

A figura 3.1 mostra um esquema do CDF [39, 40] no Tevatron. O detector é composto pelos subsistemas:

- Sistema de Traços;
- Sistema Magnético;
- Sistema de Calorimetria;
- Câmara de Múons.

Sistema de Traços

O sistema de tracking do CDF é dividido em dois subdetectores: o *Silicon Vertex Detector* [41] e o *Central Outer Tracker* [42].

Silicon Vertex Detector

O primeiro estágio a ser encontrado pelos subprodutos da colisão é o Silicon Vertex Detector ou SVX . O detector é formado por sete cilindros concêntricos de silício, dispostos como barris, que se encaixam ao redor do tubo de feixe. A camada

¹O elétron-Volt é uma unidade de medida de energia. Equivale a $1,602 \times 10^{-19}$ Joules.

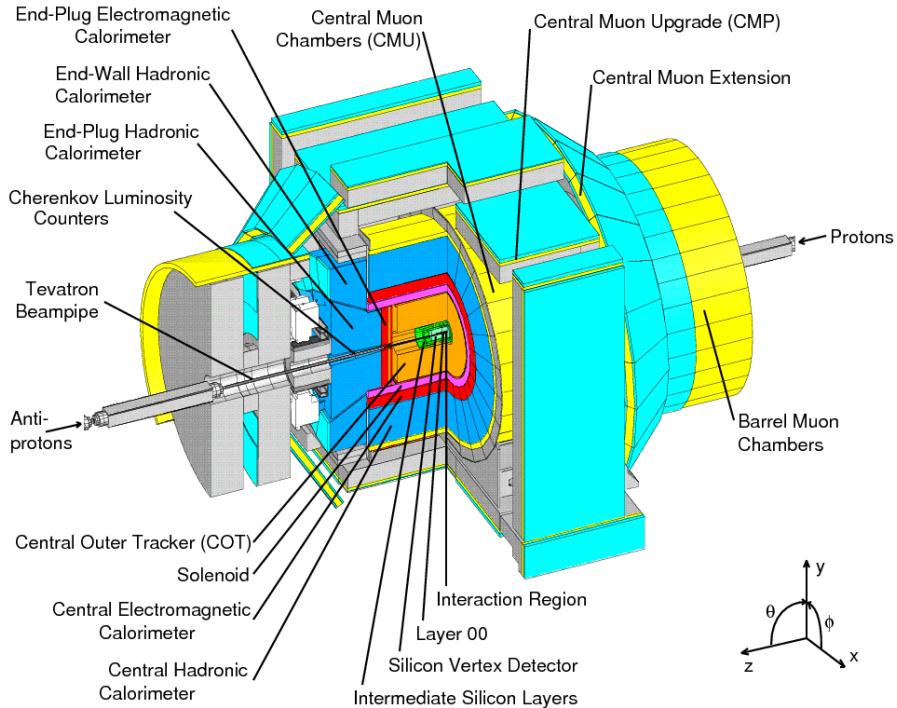


Figura 3.1: Esquema do detector CDF

mais interna toca o tubo de feixe, a camada mais externa começa a 28 centímetros (cerca de 11 polegadas) do ponto de colisão. As partículas que passam pela região do barril ionizam o material e deixam um traço ou rastro de elétrons, que é lido, camada a camada. Este traço, ou *track*, é registrado como um “*hit*” na camada de silício correspondente. Com a ligação dos pontos deixados pelos *hits*, pode-se determinar o traço da partícula.

Uma vez que o SVX está localizado dentro de um campo magnético, as partículas que possuem carga tem sua trajetória curvada. Com a informação sobre a concavidade e direção da curva, podemos determinar o momento dessas partículas.

Central Outer Tracker

O *Central Outer Tracker* ou COT possui 3 metros quadrados e está sujeito ao campo magnético produzido pelo solenóide. Tal qual um detector de SVX, o COT rastreia partículas carregadas que passam por ele, deixando um rastro de elétrons em sua trajetória, através de um processo de ionização ². O processo de aquisição de dados do COT é totalmente diferente do SVX. O COT utiliza uma câmara com Argônio e Etano, em forma gasosa, com dezenas de milhares de fios dispostos de

²Quando uma partícula carregada atravessa um meio, ela interage diretamente com os elétrons que estão na órbita dos átomos que compõem o meio, uma vez que estes elétrons adquirem uma quantidade de energia suficientemente grande para que possam escapar da órbita atômica. Assim, este átomo se torna ionizado e a este efeito chamamos de Ionização. Os elétrons liberados neste processo podem formar uma corrente elétrica de tal forma que esta pode ser lida através de um sistema próprio.



Figura 3.2: Detector CDF: Visão interna do Central Outer Tracker

maneira a capturar os elétrons e transportar a informação ao sistema de leitura para o processamento.

Na figura 3.2, vemos que existem dois tipos de fios. Os mais finos são chamados fios de sensoriamento (*sense wires*)[42]. O trabalho deles é capturar os elétrons que são liberados a partir dos átomos dos gases, quando ionizados por uma partícula. Os fios mais grossos, chamados de fios de campo (*field wires*), atraem íons positivos que são deixados quando os elétrons percorrem a extensão do subdetector. Ao se aplicar uma tensão mais negativa para os fios de campo do que para os fios de sensoriamento, cria-se um campo elétrico que faz com que elétrons movam-se numa direção (no sentido dos fios mais positivos) e os íons positivos se movam para outra (em direção ao potencial mais negativo). Um elétron, uma vez liberado, vai viajar para o fio com potencial mais positivo.

Sistema Magnético

Após o COT, tem-se o solenóide [43], que cria um forte campo magnético, da ordem de 1,4T, no volume ocupado pelos sistemas de COT e SVX. Este campo magnético curva as partículas carregadas que passam por ele, permitindo que o sentido e a magnitude da curva possam ser medidos e utilizados para a determinação do momento. Quanto maior for o momento da partícula, menor a curvatura.

Sistema de Calorimetria

O sistema de calorimetria do detector CDF é composto por dois calorímetros independentes. O primeiro é o calorímetro eletromagnético [44, 45], feito de camadas de cintilador plástico, que absorve a energia e emite luz, intercaladas entre camadas de 2 cm (aproximadamente 3/4 de polegada) de chumbo.

O segundo calorímetro é o calorímetro hadrônico[44]. Este subsistema do detector usa aço intercalado com o cintilador.

Câmara de Múons

Os detectores de múons consistem em duas partes [46]. As *Muon Drift Chambers* ou MDC [47], localizadas nas bordas exteriores do detector, funcionam de modo semelhante ao COT: cada câmara é constituída por um único fio suspenso através de um cilindro de alumínio cheio com gás. Diferentes voltagens são aplicadas ao fio e ao cilindro de alumínio, fazendo-os trabalhar da mesma maneira que os fios de campo e os fios de sensoriamento, visando criar uma diferença de tensão que atrai eletrons liberados, assim que um múon ultrapassa o subdetector. O múon entra na câmara e ioniza o gás, que libera eletrons. Os eletrons livres são atraídos para o fio, devido à diferença de potencial, e o fio transmite o sinal elétrico para o sistema de leitura.

As MDC permitem uma medição com bastante acurácia da posição do múon, mas a leitura do momento exato da ocorrência do múon também se faz necessária. Para obter uma medida precisa de tempo, foram dispostas camadas de cintiladores por trás das câmaras de múons.

A segunda parte é composta por Cintiladores. Os cintiladores [48] não dão uma medida precisa, no que tangue a posição da partícula incidente, mas trabalham muito mais rapidamente: a energia do múon é convertida quase que instantaneamente para luz e transmitida por fibras ópticas. Trabalhando em conjunto, as câmaras de múons e os cintiladores permitem a reconstrução do múon tanto no tempo quanto na quantidade de energia depositada pelo mesmo.

Correções Aplicadas a jatos b no CDF

Para que possamos analizar os sinais de jatos reconstruídos, se fazem necessárias algumas correções de erros instrumentais, isto é, erros devido às técnicas de leitura dos subprodutos gerados depois da colisão. Tais correções, no detector CDF, se aplicam em seis níveis [49], que serão apresentados a seguir.

Nível 1 - Correções na Escala Absoluta (*Absolute Scale Corrections*)

Primeiro, avaliam-se as principais incertezas sistemáticas sobre a escala absoluta de energia. Tais incertezas são obtidas através da propagação das incertezas sobre a resposta de uma única partícula. Outros fatores que introduzem menores incertezas são incluídos, a partir da comparação dos dados obtidos com as simulações de Monte Carlo da resposta do calorímetro.

Nível 2 - Correções na Escala Relativa (*Relative Scale Corrections*)

Como os calorímetros centrais são melhor calibrados e instrumentados, esta correção tenta igualar as escalas dos calorímetros da tampa (*forward calorimeters*) à escala do calorímetro central. Algumas propriedades físicas são utilizadas para corrigir a energia de jatos lidos fora da região $0,2 < |\eta| < 0,6$ para a mesma escala de energia dos jatos dentro dessa região. Esta região foi escolhida, porque está longe das regiões não instrumentadas (“*cracks*”). Após este nível de correção, a resposta do calorímetro é quase plana com relação a η .

Nível 3 - Correções de Interações Múltiplas (*Multiple interactions*)

A energia registrada nos calorímetros, proveniente de diferentes interações $p\bar{p}$ durante o mesmo *bunch cross* (janela de amostragem), pode estar dentro do mesmo cluster de reconstrução de jatos.

Um jato é um cone estreito de hadrons e outras partículas produzidas pela hadronização de um quark ou glúon gerados em uma colisão, por isso, uma das técnicas para a reconstrução dos eventos de um jato é a clusterização cônica. Para essa clusterização, faz-se um cone a partir do vértice da colisão, e classifica-se as partículas do interior desse cone como sendo pertencentes ao jato criado a partir do vértice.

Com isso, podemos ter, um aumento na leitura da energia do jato. Esta correção visa retirar, em média, esse aumento errôneo. O valor a ser subtraído é obtido a partir de dados de minimum bias e é parametrizado como uma função do número de vértices no evento, o que faz esta correção ser extremamente dependente da taxa de luminosidade.

O feixe de um anel de colisão é formado por *bunches* (pacotes) de partículas que são acelerados e colidem nos pontos onde estão instalados os experimentos. Quando ocorre a interação, temos um bunch de um dos feixes colidindo com outro bunch do outro feixe. Cada um desses *bunches* possui um determinado número de partículas, N_1 e N_2 . A Luminosidade é definida como

$$L = \frac{n \cdot N_1 \cdot N_2 \cdot f}{A}. \quad (3.5)$$

onde n é o número de *bunches* no feixe, f a frequência de colisão e A a área da seção transversal dos *bunches*.

Nível 4 - Correções em Eventos adjacentes (*Underlying events Corrections*)

Eventos adjacentes são definidos como eventos associados com o mesmo parton, embora sejam provenientes de outro subproduto da colisão.

Nível 5 - Correções de Eventos fora do Cone (*Out-of-Cone Corrections*)

Nesse nível de correção, a energia proveniente do vazamento para fora do cone de reconstrução de jatos é retirada, tentando-se fazer com que a energia do jato seja igual a energia do “parton pai”. Comparando os dados reais com as simulações de Monte Carlo, busca-se encontrar fatores corretores para diversos valores de η e ϕ .

Nível 6 - Correções para Incertezas sistemáticas totais (*Total Systematic Uncertainties Corrections*)

Para baixos valores de momento transverso , Pt , a principal contribuição é a incerteza fora do cone (correção de nível 5). Para altos Pt , a principal incerteza é proveniente da escala absoluta (correção de nível 1). O último nível busca reduzir as incertezas totais do sistema, sendo estas provenientes do sistema das simulações de Monte Carlo. Como a maioria das correções anteriores são dependentes de simulações, as incertezas totais podem ser reduzidas com a melhoria da simulação do CDF e maiores estatísticas.

Outros Tipos de Correção

Um outro tipo de correção que não pode ser classificada dentro de um nível de correção descrito anteriormente é aquela feita com redes neurais artificiais. Tal correção foi proposta em [50] e obteve resultados interessantes e promissores para o detector CDF.

3.2.2 Detector $D\emptyset$

O detector $D\emptyset$ [51] está localizado em um dos pontos de colisão de feixes acelerados pelo Tevatron e sua estrutura pode ser vista na figura 3.3. O $D\emptyset$ foi projetado para a observação de subprodutos criados pela colisão de prótons e anti-prótons, além do estudo de todos os processos físicos das interações próton-antiproton a 2 TeV de energia de centro de massa e da procura por partículas previstas por modelos ”além do Modelo Padrão“. Vale salientar que o quark *top* foi detectado pela primeira vez nos experimentos CDF e $D\emptyset$. No $D\emptyset$ foi descoberto o bárion Ξ_b , um bárion que contém os quarks *b* e *s*. Esse detector é constituído pelos seguintes sub-sistemas:

- Sistema de Traços;
- Sistema Magnético;
- Sistema de Calorimetria;
- Sistema de Múons.

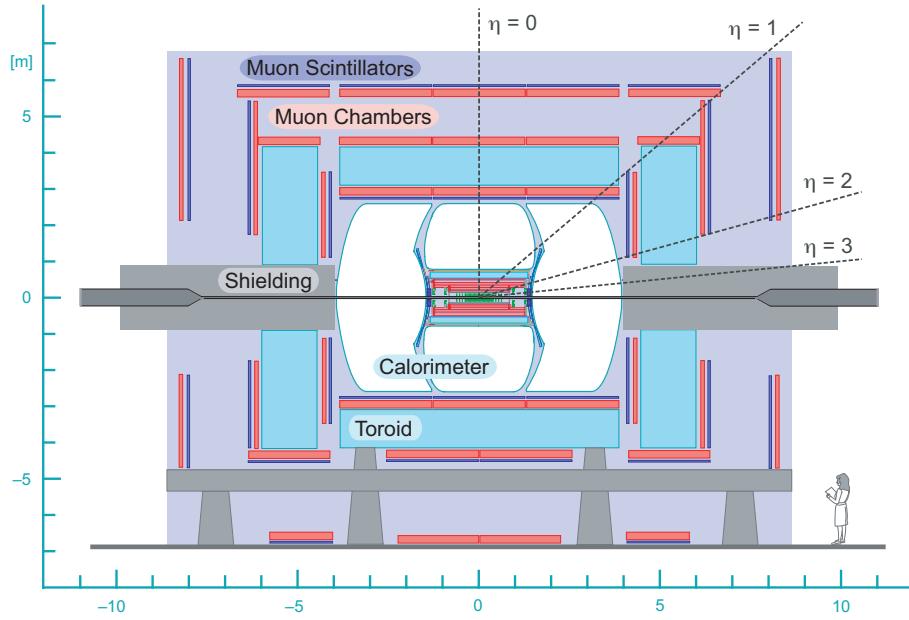


Figura 3.3: Esquema do detector $D\emptyset$

Sistema de Traços

Tal como no detector CDF, o ponto onde os feixes colidem está rodeado por um sistema de detectores de traços, que tem por objetivo ler o rastro deixado pelas partículas carregadas produzidas em cada evento. Para isso, como no detector CDF, as medições mais próximas da colisão são realizadas no SMT (Silicon Microstrip Tracker) [52].

Fora do detector de silício, o detector $D\emptyset$ possui um segundo nível de tracking formado por fibras cintilantes, que produzem fôtons quando uma partícula atravessa seu interior. Este nível de detecção é chamado de CFT (Central Fiber Tracker [53]). Todo sistema de traços é imerso num campo magnético para que partículas carregadas possam ser curvadas e, a partir das curvaturas, medirem-se os seus momentos.

Sistema Magnético

O detector $D\emptyset$ possui um solenóide que envolve tanto o Silicon Microstrip Tracker quanto o Central Fiber Tracker. Este solenóide produz um campo magnético de 2 T[54] que é responsável por curvar partículas carregadas em seu interior.

Sistema de Calorimetria

Depois do Sistema Magnético, na linha de propagação dos subprodutos da colisão, temos o sistema de calorimetria do $D\emptyset$, composto por dois calorímetros que possuem diferentes materiais ativos³ imersos em Argônio líquido. Na parte eletromagnética,

³O material ativo, em um calorímetro de amostragem, é o material (geralmente de alto número atômico) responsável pela amostragem (leitura) do sinal proveniente do chuveiro.

temos Urânio como material passivo, sendo que na parte hadrônica esse material é o aço inoxidável [55].

Sistema de Múons

A camada mais externa do detector [56] funciona de maneira similar ao sistema de múons do CDF.

Correções Aplicadas a jatos b no $D\emptyset$

O procedimento de correção da escala de energia de jatos no detector $D\emptyset$ [57] envolve uma série de subcorreções que são aplicadas de forma seqüencial, tal como no detector CDF. Estas correções são estimadas através de simulações do detector (Monte Carlo) [58].

Nível 1: Correção de Offset - *Offset Correction*

O “offset de energia”, neste caso, é definido como a energia depositada no interior do cone do jato reconstruído que não está relacionada com a interação primária formadora do jato, ou seja, a energia adicional que foi reconstruída como energia do jato, mas, na realidade não faz parte da energia de jato. Ruído eletrônico e o chamado ruído de Urânio⁴, ruído proveniente do sistema de leitura do calorímetro, e a energia de colisões anteriores (pile-up) contribuem para o offset de energia.

A correção é extraída de eventos simulados de “minimum-bias” (eventos definidos com exigência da ocorrência da *flag* de nenhum vértice, ou seja, nenhuma interação dentro do detector), assim, podemos avaliar a componente de ruído eletrônico e ruído de Urânio. Um fator dependente da luminosidade e do raio do cone de reconstrução é calculado através de simulação de eventos com diversas luminosidades, tentando descrever um cenário com múltiplas interações (MI), como pode ser visto na figura 3.4 com o raio de reconstrução de cone de 0,5 e na figura 3.5 com o raio de reconstrução de cone de 0,7, ambas as figuras mostram o offset de energia para diferentes valores de η . O raio de reconstrução é dado por $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

Nível 2: Correção de Resposta Absoluta - *Absolute Response Correction*

A correção da resposta absoluta é medida através da simulação de eventos de fótons. Assim, podemos propagar o sinal de uma única partícula e medir a diferença entre a resposta do detector e a resposta esperada (Monte Carlo) e, assim, obtermos a correção indicada para cada energia.

⁴O ruído de Urânio é o ruído proveniente do decaimento radiativo natural do Urânio usado como material passivo [59]

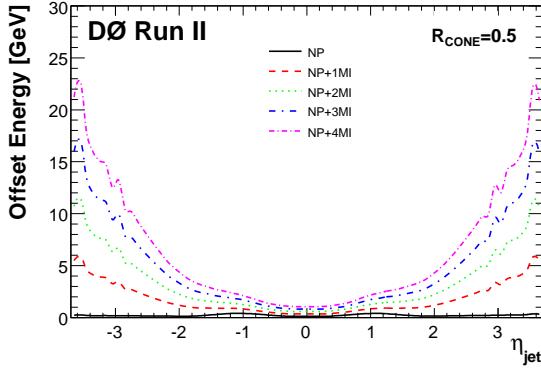


Figura 3.4: Detector $D\emptyset$: Correção do offset de energia com diversas luminosidades e raio de reconstrução de cone de 0,5.

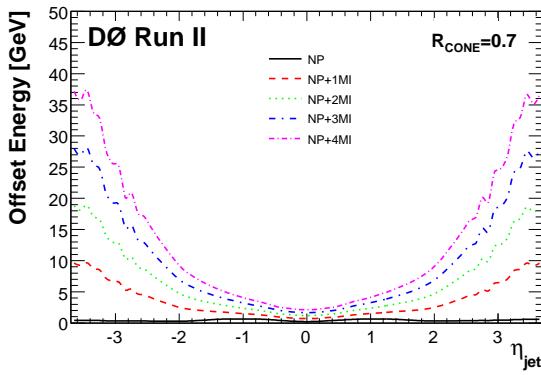


Figura 3.5: Detector $D\emptyset$: Correção do offset de energia com diversas luminosidades e raio de reconstrução de cone de 0,7.

Nível 3: Correção de Resposta Relativa - *Relative Response Correction*

Enquanto o calorímetro do detector $D\emptyset$ é relativamente uniforme em sua parte central (CC) e em sua parte final (CE), a lacuna entre as câmaras de resfriamento ($0.5 < |\eta| < 1.8$) não possui instrumentação, ou seja, gera uma não linearidade na resposta em função de η . Com a simulação de jatos com diferentes energias para diferentes valores de η , podemos alcançar a correção da resposta relativa (também chamada de “intercalibração em η ”), que visa calibrar a resposta para jatos, uniformizando a resposta do calorímetro para todos os valores de η .

Nível 4: Correção na Reconstrução do Chuveiro - *Showering Correction*

O principal objetivo da correção de reconstrução de chuveiro é corrigir a reconstrução errônea de energia do cone do jato. Partículas pertencentes ao cone podem vir a ser classificadas (clusterizadas) como partículas não-pertencentes ao jato, com isso, suas energias não serão adicionadas à reconstrução do jato. Este fenômeno é chamado de vazamento para fora do cone. Por outro lado, partículas que deve-

riam ser classificadas como não-pertencentes ao jato podem ser classificadas como partículas pertencentes a ele, e as energias de tais partículas serão reconstruídas como energia do jato. Para eliminar os efeitos desse erro de classificação e adição de partículas, foi criada a correção de reconstrução do chuveiro.

Esta correção é calculada através de simulação de Monte Carlo, utilizando eventos de fôtons e com um único vértice principal e diferentes valores de η , para reduzir o impacto de interações múltiplas. A correção é calculada para cada célula de acordo com o perfil levantado através das simulações.

Nível 5: Correção de Tendência Adicional - *Additional Biases Correction*

Esta correção visa aumentar a energia em células de leitura específicas, para que estas consigam superar o limiar de supressão por zero. Esta supressão visa zerar a leitura de células que não atinjam um limiar mínimo de energia. Esta correção se faz importante, pois, diferente de fôtons e elétrons, o chuveiro de um jato é muito complexo, de modo que, em alguns casos, informações de energia contidas em células de menor energia se fazem necessárias.

Nível 6: Correção de Total de Incertezas - *Total Uncertainties Correction*

Assim como no CDF, este nível de correção busca reduzir as incertezas totais do sistema, estas incertezas são provenientes diretamente das simulações de Monte Carlo, que é a principal referência para o cálculo das correções.

Capítulo 4

Metodologia de Análise

Neste capítulo a metodologia de análise será descrita bem como as técnicas utilizadas na dissertação. A metodologia de análise pode ser vista na figura 4.1. Os dados utilizados serão descritos na seção 4.1 enquanto a seção 4.2 descreve os modelos de corretores baseados em redes neurais artificiais. Os modelos de ajuste para a massa invariante, bem como o seu cálculo, serão descritos na seção 4.3 e a aplicação da correção será descrita na subseção 4.3.1. Na seção 4.4 serão mostradas as avaliações estatísticas de modelos ajustados.

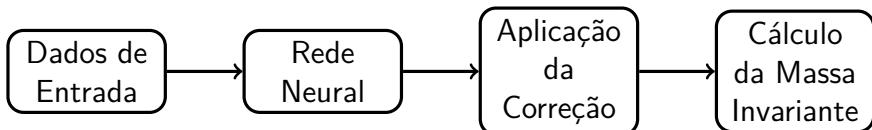


Figura 4.1: Metodologia de Análise utilizada neste trabalho

4.1 Modelo de Dados do detector ATLAS

O Modelo de Dados do detector ATLAS [60, 61], também chamado de **EDM**, de *Event Data Model*, provê, de forma simples e fácil, o compartilhamento de diferentes tipos de informação, por meio de todos os diferentes grupos pertencentes ao detector ATLAS¹. Além disso, o EDM permite o uso de ferramentas de software comuns entre o processamento de dados on-line e a reconstrução offline [61].

O detector ATLAS produz até um petabyte de dados por ano [62]. Este volume de dados não permite a ampla distribuição de dados brutos para colaboradores em todo o mundo. Para permitir o acesso a esses dados por todos os pesquisadores da colaboração situados em vários países, os dados são processados em três etapas:

¹O detector possui diversos grupos de trabalho com diversos objetivos, totalizando em torno de 4000 pessoas trabalhando em conjunto. Dentre esses grupos, podemos citar os grupos de análises físicas, de manutenção de sistemas e o de comissionamento (montagem do detector).

- Dados Brutos (*Raw Data*): Possui todas as informações brutas dos eventos reconstruídos pelo ATLAS. Nesta etapa, não há compactação ou perda de informação do detector. O tamanho final para um arquivo do tipo Raw Data é de 1,6 Mb por evento.
- Dados de Eventos Resumidos (**ESD** - *Event Summary Data*): contém a saída detalhada da reconstrução do detector e são produzidos a partir dos dados brutos. Irão conter informação suficiente para permitir a identificação de partículas, a calibração de jatos e outras informações que permitam o ajuste dos algoritmos de reconstrução e calibrações. O tamanho final para um ESD é de 500 kB por evento.
- Dados de objetos de análise (**AOD** - *Analysis Object Data*): um resumo do evento reconstruído, com informação suficiente para análises comuns. Um arquivo do tipo AOD pode ser produzido a partir de arquivos do tipo ESD, o que implica em benefícios de custo e de tempo significativos. O tamanho final para um AOD é de 100 kB por evento.

Em nenhuma das etapas descritas acima, os eventos e suas informações podem ser lidas pelo ROOT, o sistema de análises do CERN. Tem-se que converter os arquivos para um dos três formatos disponíveis no CERN, chamados de **DPD** (*Derived Physics Dataset*), para que possam ser analisados no ROOT. Os possíveis tipos são:

- *Primary Derived Physics Dataset - D¹PD*: basicamente tem o mesmo formato que as AOD, entretanto, são menores, contendo cerca de 10 kB por evento. Neste tipo de arquivo, são permitidas operações para a redução de tamanho que não são permitidas nos arquivos do tipo AOD, tais como:
 - Escolha de eventos relevantes²
 - Escolha de eventos relevantes para uma análise específica, diretamente dependente da análise de destino do arquivo gerado.
 - Seleção de propriedades de interesse de um objeto, por exemplo, um corte feito em função da posição de um traço ou chuveiro. Mais uma vez, aqui, temos uma forte dependência da análise de destino do arquivo gerado.
- *Secondary Derived Physics Dataset - D²PD*: possui o mesmo formato que os arquivos AODs, mas contendo informações mais específicas para análise, sendo geralmente produzidas a partir de arquivos da etapa anterior, os *D¹PD*.

²A escolha relevante é feita com base em eventos que foram identificados com uma confiabilidade acima de um limiar de escolha, tal limiar é avaliado em uma análise (*TAG*), e para essa análise, o tamanho médio de um evento é 1 kB.

- *Tertiary Derived Physics Dataset - D³PD*: é o tipo de arquivo mais utilizado para análise atualmente, facilitando a sua análise no ROOT. Para sua geração, o usuário especifica o nível de detalhes das informações e, assim, as variáveis são automaticamente configuradas e o arquivo é gerado.

4.1.1 Dados Utilizados

Durante décadas, o bóson de Higgs do Modelo Padrão (SM, de *Standard Model*) permaneceu sem observação experimental. Até que em julho de 2012, as colaborações ATLAS e CMS, anunciaram a observação de uma nova partícula com massa em torno de 125 GeV e com propriedades consistentes com aquelas esperadas para o SM Higgs - ou bóson de Higgs do Modelo Padrão. Desde então, medidas mais precisas estão sendo feitas para confirmar se, de fato, a partícula observada é o famoso bóson.

As medidas realizadas pelo detector ATLAS, para a observação da partícula, foram feitas em decaimentos bosônicos, isto é: $H \rightarrow ZZ$, $H \rightarrow \gamma\gamma$ e $H \rightarrow WW$. O objetivo é estudar todos os decaimentos do bóson de Higgs, inclusive os fermiônicos, para verificar todas as propriedades previstas pelo Modelo Padrão e para observar se há desvios que não foram previstos por este modelo.

Já o decaimento $H \rightarrow b\bar{b}$ é previsto no SM com o *branching ratio*³ de 68,4% para a $m_H = 125$ GeV [63], porém a análise inclusiva desse decaimento ($pp \rightarrow H^0 + X \rightarrow b\bar{b} + X$) é muito difícil devido à grande produção de eventos de QCD (Cromodinâmica Quantica) que produzem vários jatos, os chamados *multijet events*, responsáveis pelo alto *background* desse processo. O experimento ATLAS tem analisado eventos com produção associada ($W/Z + H$), que auxiliam, de maneira eficiente, a redução de *background*, sendo usados também no disparo do processo de aquisição, o *trigger*.

Os dados utilizados neste trabalho contêm eventos com zero, um ou dois léptons carregados (eletróns ou mûons). Os eventos reconstruídos são $Z \rightarrow \nu\nu$, $W \rightarrow \ell\nu$ e $Z \rightarrow \ell\ell$. Os eventos com zero lépton possuem uma reduzida, mas não insignificante, contribuição para o decaimento $W \rightarrow \ell\nu$, quando o lépton atravessa uma região não sensível do detector ou não é corretamente identificado.

Um algoritmo de identificação de b (*b-tagging*) é utilizado para selecionar cada um dos jatos produzidos pelo decaimento $H \rightarrow b\bar{b}$. Ao final do processo de reconstrução, a massa dos jatos identificados (*b-tagged*) é utilizada como variável de seleção de eventos e também como um qualificador dos eventos e seus respectivos ruídos de fundo (*background*).

³O *branching ratio* de um decaimento nada mais é que a probalidade daquele decaimento ocorrer para uma dada energia ou massa

As amostras de MC [58] para o sinal e para o ruído de fundo foram obtidas utilizando a simulação **ATLFAST-II** [64], que possui a simulação completa do detector ATLAS baseado no **GEANT4** [65], com exceção da resposta dos calorímetros, para a qual uma simulação parametrizada foi utilizada. A luminosidade nominal dos eventos foi de $4,7 \text{ fb}^{-1}$ simulada a $\sqrt{s} = 7 \text{ TeV}$ no centro de massa. O número total de eventos é de 5.524 e os dados foram reconstruídos durante o *Run 1*, que ocorreu em 2011 no CERN [66].

O gerador de MC utilizado para os processos de decaimento WH e ZH foi o **PYTHIA8** [67] com o **CTEQ6L1** [68] como gerador de função de distribuição de partons. O processo de equalização (*tune*) utilizado para o desenvolvimento do chuveiro do parton, para a hadronização⁴ [69] e para as múltiplas interações entre partons foi o **AU2** [70]. As probabilidades de cada decaimento ocorrer - *branching ratios* - foram calculadas utilizando o **HDECAY** [71].

As amostras de sinal utilizadas para a massa do bóson de Higgs foram simuladas de 100 a 150 GeV em passos (*steps*) de 5 GeV, para uso geral da colaboração, porém como já foi dito anteriormente, a massa anunciada do bóson de Higgs foi 125 GeV, por isso, utilizamos os eventos simulados com bósons de Higgs de 125 GeV de massa. Todos os tipos (sabores) de léptons carregados (*charged lepton flavours*) foram simulados nos decaimentos W e Z , bem como os decaimentos do lepton τ . Embora, para o bóson de Higgs, apenas o decaimento $b\bar{b}$ tenha sido gerado e registrado.

Para os processos que geraram o ruído de fundo (background), temos 2 principais responsáveis. O primeiro é o decaimento do W ou do Z com jatos de quarks c ou b massivos, simulados utilizando o gerador **SHERPA** [72] com o **CT10** [73] sendo responsável pelas funções de distribuição de partons. O segundo é a produção de $t\bar{t}$, que foi simulada com o gerador **POWHEG** [74] e, também, com o **CT10** gerando as funções de distribuição de partons, embora, neste caso, o **CT10**, tenha sido interfaceado pelo **PYTHIA6** [75] e tenha utilizado como *tunning* o **Perugia2011C** [76].

Outros ruídos de fundo podem ser provenientes da produção de um único top (*single-top*) ou da produção de um diboson (WW , WZ ou ZZ). Eventos de *mininum bias* foram simulados com o **PYTHIA8**, **CTEQ6L1** (responsável pelo cálculo das funções de distribuições de parton e com a equalização A2). Estes eventos são somados ao sinal e ao ruído de fundo durante a simulação dos mesmos. Os efeitos do *mininum bias* se propagam por toda a cadeia de aquisição e são reconstruídos pela mesma

⁴Em física de partículas, hadronização é o processo de formação de hadrons a partir de quarks e glúons. Após colisões de altas energias em aceleradores de partículas, quarks e glúons são criados. Devido ao confinamento, quarks e glúons não podem existir isoladamente, segundo Modelo Padrão os quarks e antiquarks criados emitem glúons, que decaem em quarks, produzindo assim vários quarks, que se combinam espontaneamente para formar hadrons. Dois programas exploram a parametrização atual da hadronização: **PYTHIA** e **HERWIG**

cadeia de reconstrução dos dados reais

Os hadrons que possuem $Pt > 5$ GeV são agrupados em jatos, que consistem de um cone de raio $\Delta R = 0,4^5$ ao redor da partícula de maior energia que compõe o jato. Se um hadrôn b é identificado, o jato é etiquetado como jato b (b jet). Atualmente, existem duas maneiras de se identificar um hadrôn b : identificando léptons de baixo Pt , elétrons ou mûons, provenientes do decaimento semi-leptônico do hadrôn b , ou pelo método do vértice, baseado na vida-média relativamente longa do hadrôn b , que mede o parâmetro de impacto dos traços do jato em relação ao vértice primário.

4.2 Redes Neurais Artificiais

Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de seres vivos que adquirem conhecimento por meio da experiência.

Uma rede neural artificial é composta por várias unidades de processamento, cujo funcionamento é bastante simples. Essas unidades, geralmente são conectadas por canais de comunicação que estão associados a determinado peso, que funciona de maneira análoga à intensidade de transmissão do sinal de um neurônio biológico para outro. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma Rede Neural Artificial vem das interações entre as unidades de processamento da rede.

A operação de uma unidade de processamento, proposta por McCulloch e Pitts em [77], pode ser considerada o primeiro modelo de rede neural que se tem notícia e pode ser visto na figura 4.2. Nesse modelo, vemos que o setor de entrada, que geralmente é o lugar onde os padrões são apresentados à rede, é representado em azul, enquanto o setor intermediário, que é o lugar onde é feita a maior parte do processamento, através das conexões ponderadas, é representada em vermelho e o setor de saída, que, geralmente, é o lugar que apresenta o resultado final do processamento, é representado em verde. Uma rede neural com esta estrutura, não consegue, a princípio, executar uma separação de classes que não sejam linearmente separáveis [78]. Um exemplo clássico é a função ou-exclusivo (XOR).

Essa restrição só foi superada em 1986 com a criação do algoritmo *backpropagation* [79], que conseguia ajustar pesos de uma rede com mais de uma camada retropropagando o erro da saída em função do peso em questão. Criando assim o modelo de Redes Neurais Artificiais mais utilizado atualmente, as redes Perceptron Multi-Camadas (MLP), que pode ser vista na figura 4.3, são treinadas eficientemente treinadas com o algoritmo backpropagation. Durante o treinamento com o

⁵O raio do cone de detecção é calculado segundo a relação $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$

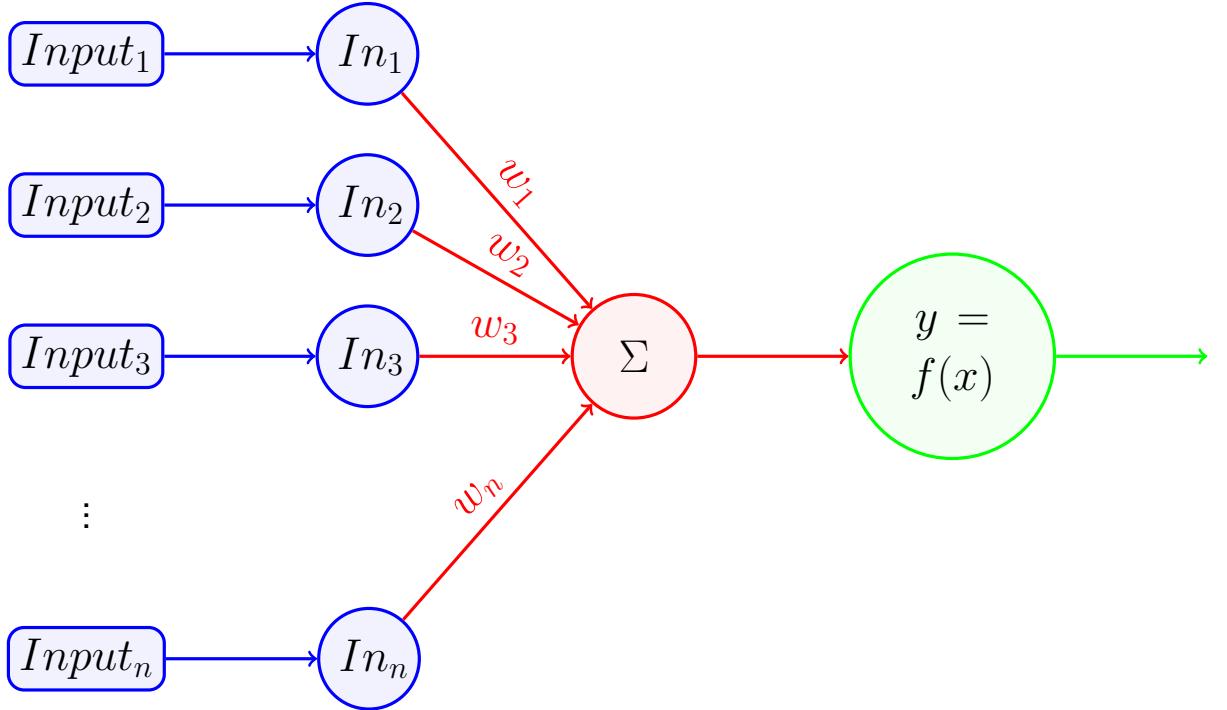


Figura 4.2: Modelo de uma rede neural artifical proposta por McCulloch e Pitts em 1943

algoritmo backpropagation, a rede opera em uma sequência de dois passos. Primeiro, um padrão é apresentado à camada de entrada da rede. Este padrão flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, o erro é calculado.

O erro é repropagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado, de acordo com a equação 4.1, onde F_{custo} é uma função do erro, δ é uma constante de proporcionalidade contida entre $[0,1]$, $w_{ij}(t)$ é o peso antes da atualização e $w_{ij}(t + 1)$ é o peso depois da atualização.

$$w_{ij}(t + 1) = w_{ij}(t) + \delta \left(-\frac{\partial F_{custo}}{\partial w_{ij}} \right) \quad (4.1)$$

Estes modelos, aliados ao teorema da aproximação universal (*Universal Approximation Theorem*) [80], que diz que qualquer função continua pode ser aproximada por uma rede neural com uma única camada escondida com um número finito de neurônios, faz com que as redes neurais sejam ferramentas poderosas no mapeamento de funções de separação e de funções de estimativa.

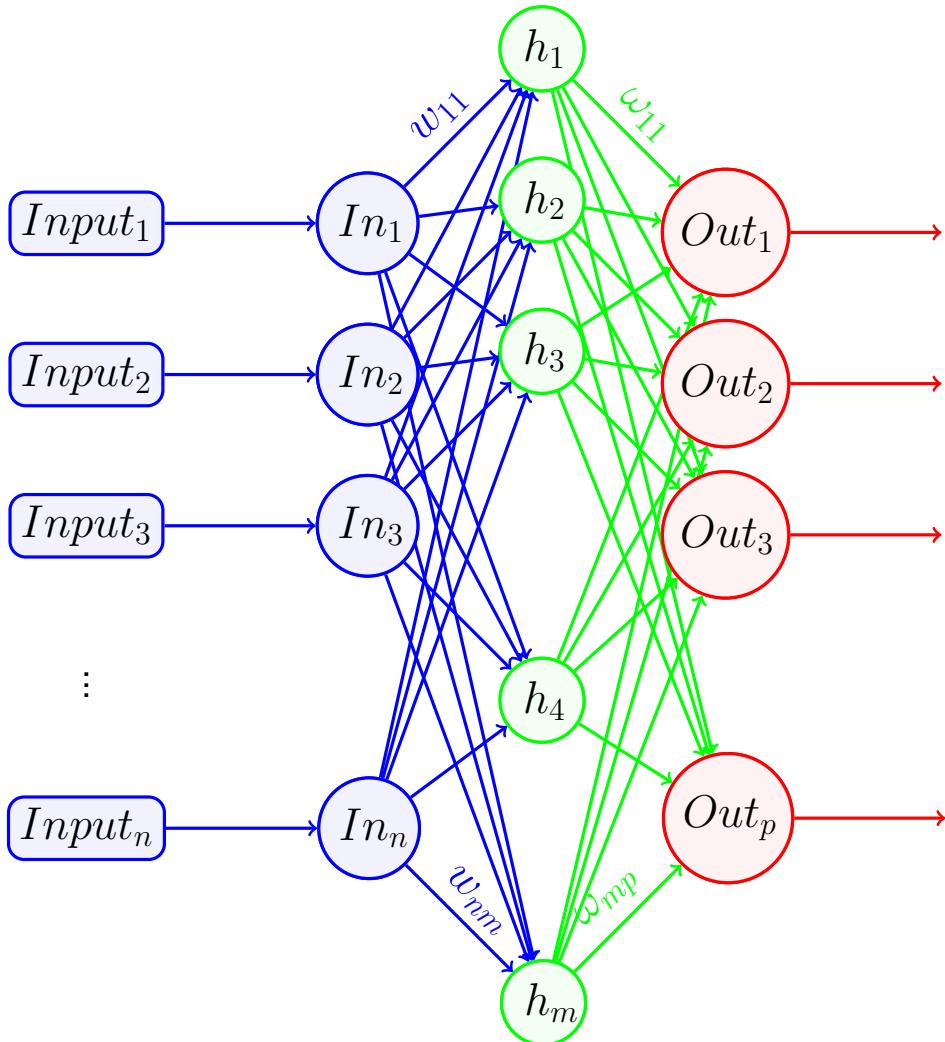


Figura 4.3: Modelo de uma rede neural artifical com uma única camada escondida

Algoritmos de Treinamento

A maioria dos modelos de redes neurais possui alguma regra de treinamento, onde os pesos de suas conexões são ajustados de acordo com os padrões apresentados. Em outras palavras, elas aprendem através de exemplos. Arquiteturas neurais são tipicamente organizadas em camadas, sendo que neste caso, os neurônios de uma dada camada tem suas saídas ligadas apenas a entradas de neurônios da próxima camada.

Uma das propriedades mais importantes de uma rede neural é a habilidade de aprender e se adaptar aos dados que lhe são apresentados e com isso melhorar seu desempenho. Isso é feito através de um processo iterativo de ajustes de seus pesos, o treinamento.

Denomina-se algoritmo de aprendizado, ou algoritmo de treinamento, um conjunto de regras para o ajuste de pesos visando a solução de um dado problema.

Existem muitos tipos de algoritmos de aprendizado específicos para determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são atualizados.

De modo geral, existem três classes de treinamento para uma rede neural artificial:

- o treinamento supervisionado - quando um tutor ou especialista é utilizado como um agente externo que indica à rede a resposta desejada para o padrão de entrada;
- o treinamento não-supervisionado (auto-organização) - neste caso, a rede neural busca dentro do conjunto de dados padrões similares e busca agrupá-los de acordo com uma métrica de similaridade.
- o treinamento semi-supervisionado (reforço de treinamento) - que nada mais que uma mistura dos dois anteriores, onde a rede neural aprende com as informações do tutor (treinamento supervisionado) e ao mesmo tempo busca o reconhecimento e separação de padrões de forma autônoma (treinamento não-supervisionado).

Neste trabalho, serão apresentados três algoritmos de treinamento supervisionado: o treinamento *Steepest Descent*, o treinamento BFGS e o treinamento de Levenberg-Marquardt.

O treinamento *Steepest Descent* é baseado puramente na regra delta. Sua função custo geralmente é o MSE (*Mean Squared Error*) que calcula a cada *epoch*⁶ de treinamento, um valor médio do gradiente e faz a atualização dos pesos para esse valor.

O treinamento *Steepest Descent* ou método de minimização *Steepest Descent* [81] é, como diz o nome, um método de minimização que utiliza um algoritmo de otimização de primeira ordem (*first-order optimization algorithm* - FOOA) para encontrar mínimos locais da função custo.

Para este fim, a função atualiza os pesos a partir de passos (*steps*) negativos do gradiente (ou de uma aproximação do gradiente) da função em determinado ponto. Como o vetor gradiente “aponta” para o máximo da função, com passos negativos na direção contrária do gradiente, tem-se um “caminho” em direção ao mínimo da função.

Esta técnica apresenta algumas limitações, como por exemplo: se a função custo (F_{custo}) possuir indeterminações, o processo de atualização pode se tornar muito instável. Uma outra limitação é que esse algoritmo tende a ser mais lento para a convergência.

⁶Uma *epoch* consiste da apresentação completa de todas as amostras do conjunto de treinamento a fim de computar o valor médio da função custo, visando-se a atualização dos pesos da rede.

O treinamento BFGS [82] é um método que utiliza um algoritmo de otimização de segunda ordem (*second-order optimization algorithm* - SOOA). Esse método aproxima a matriz de segundas derivadas, a matriz Hessiana⁷, usando atualizações que são obtidas de acordo com a evolução do gradiente [83].

O treinamento de Levenberg-Marquardt, ou LM [84] interpola entre o algoritmo de quasi-Newton e o algoritmo *Steepest Descent*. Por esse motivo, o algoritmo LM é mais robusto do que o BFGS, o que significa, em muitos casos, que mesmo que os valores iniciais sejam distantes da melhor solução final, esta pode ser encontrada com esse algoritmo.

4.2.1 Validação Cruzada

O processo de validação cruzada, utilizada nesse trabalho, consiste, basicamente, na divisão do conjunto de dados em vários subconjuntos de tamanho fixo. Os conjuntos de treinamento, teste e validação são criados a partir de vários sorteios aleatórios desses subconjuntos.

Redes Neurais Artificiais são altamente dependentes de suas entradas. Para que um treinamento seja comprovadamente eficaz, o mesmo deve obter resultados satisfatórios para diversas entradas diferentes. Para isso, neste trabalho, foram feitos 30 sorteios dentro dos subconjuntos separados a partir do conjunto de dados, o chamado método das K-partições.

Os pesos iniciais de uma rede neural, geralmente, são sorteados de forma aleatória, com isso, a cada inicialização, temos também um novo treinamento. Para que tenhamos uma análise independente dos pesos iniciais, temos que iniciar os pesos diversas vezes. Neste trabalho, o modelo neural, para cada sorteio das variáveis de entrada, foi iniciado com 100 valores iniciais de pesos diferentes.

4.2.2 Aplicação do Modelo Neural

O modelo neural a ser aplicado aos dados pode ser visto na figura 4.4. Nele, temos 2 saídas (uma para cada jato) que são as correções a serem aplicadas a cada jato separadamente, visando assim a melhora da reconstrução de energia do detector para o decaimento $H \rightarrow b\bar{b}$.

4.2.3 Treinamentos Realizados

Os treinamentos realizados neste trabalho utilizam a *toolbox* de redes neurais artificiais que integra o *software* MatLab, disponibilizado pelo Laboratório de Processa-

⁷A matriz Hessiana é uma matriz quadrada com todas as segundas derivadas parciais de uma função.

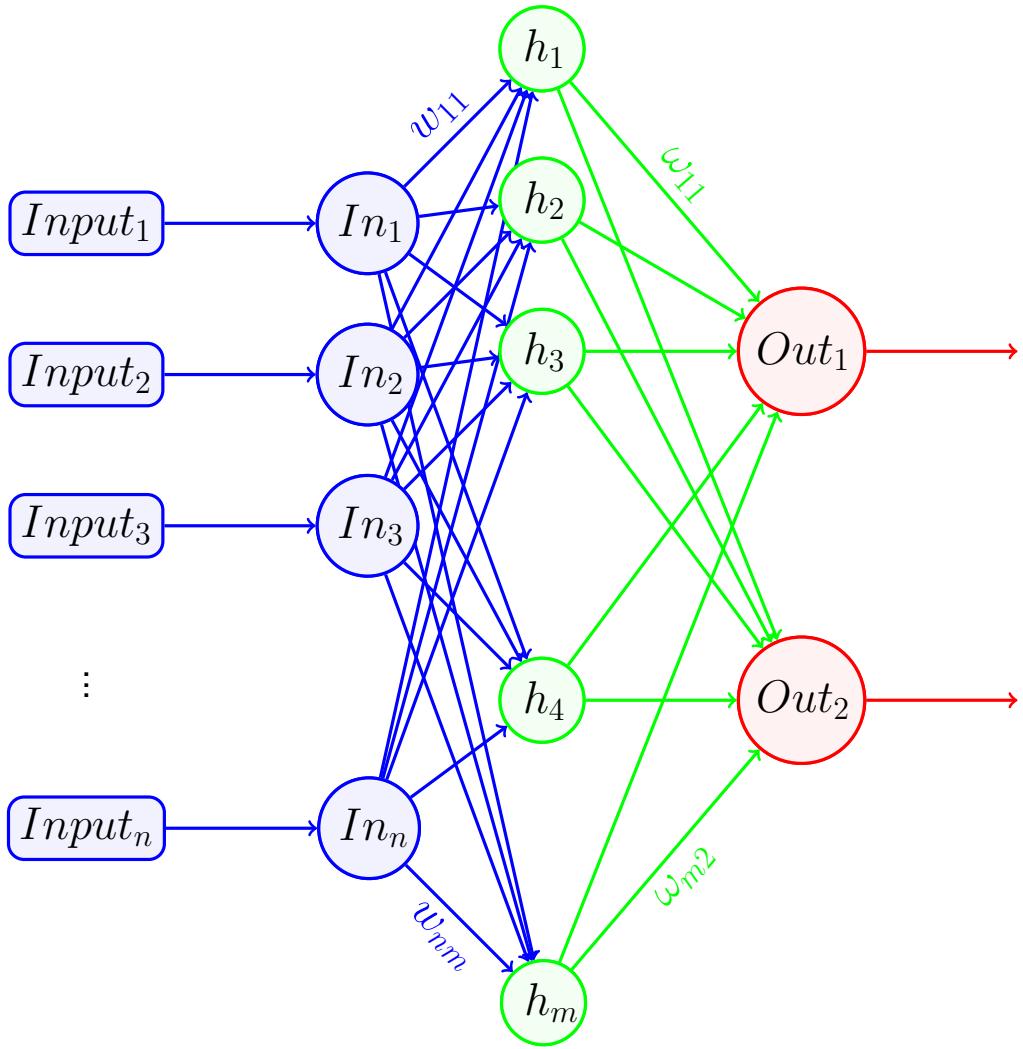


Figura 4.4: Modelo de uma rede neural artifical gerando duas correções (uma para cada jato)

mento de Sinais.

Para os treinamentos, uma porcentagem do conjunto amostral foi separada para o conjunto de treinamento (no caso 60%) e outra porcentagem para o conjunto de teste (no caso 40%). Como o conjunto de eventos utilizados foi simulado em grande quantidade, o conjunto de teste foi replicado para o conjunto de validação.

O conjunto de parametros utilizados para os treinamentos com os dois algoritmos pode ser visto na tabela 4.1.

4.3 Massa Invariante do Bóson de Higgs

A equivalência massa-energia surgiu da teoria da relatividade, proposta em 1905 por Albert Einstein, e pode ser descrita pela famosa equação 4.2, onde E é a energia

Parametro	Valor
<i>Learning Ratio</i>	0,000001
Número Máximo de Falhas na Validação	50
Valor Mínimo do gradiente	$1 \cdot 10^{-10}$
Número Máximo de Épocas	1000

Tabela 4.1: Parâmetros de treinamento de Redes Neurais

do corpo, m é a massa do corpo e c é a velocidade da luz no vácuo.

$$E = m \cdot c^2 \quad (4.2)$$

Para partículas em movimento, esta equação adquire mais um termo e se torna a equação 4.3, para manter a conservação de energia, o termo $p \cdot c$ foi adicionado sendo dependente da velocidade da partícula em questão. Portanto, temos dois componentes de energia, um estático - $m \cdot c^2$, e um componente dinâmico - $p \cdot c$, onde p é o vetor momento da partícula.

$$E^2 = (m \cdot c^2)^2 + (p \cdot c)^2 \quad (4.3)$$

Com isso, podemos calcular a massa partícula de acordo com a equação 4.4. Para facilitar a compreensão, a forma mais utilizada de obtenção da massa invariante é dada pela equação 4.5, onde $c = \hbar = 1$.

$$m = \frac{\sqrt[2]{E^2 - (p \cdot c)^2}}{c^2} \quad (4.4)$$

$$m = \sqrt[2]{E^2 - p^2} \quad (4.5)$$

4.3.1 Correção aplicada à Massa

Supondo que a massa m_{truth} seja a massa verdadeira depositada, dependente de um valor de energia E_{truth} e de momento p_{truth} e que m_{reco} seja a massa reconstruída pelo detector, dependente de um valor de energia E_{reco} e de momento p_{reco} e supondo ainda que o detector possua erros de medida, podemos gerar uma correção θ que

mapeia a massa reconstruída para a massa verdadeira.

$$\theta = \frac{m_{truth}}{m_{reco}} \quad (4.6)$$

A partir das equações 4.5 e 4.6 podemos verificar que uma simples multiplicação pode corrigir a massa m_{reco} para a massa m_{truth} , como pode ser visto na equação 4.7.

$$\theta \cdot m_{reco} = \theta \cdot \sqrt[2]{E_{reco}^2 - p^2} = \sqrt[2]{(\theta \cdot E_{reco})^2 - (\theta \cdot p)^2} \quad (4.7)$$

A seguir, serão apresentados diferentes modelos para a reconstrução de massa do bóson de higgs, e cada modelo será detalhado e aprofundado separadamente.

4.3.2 Modelos de Ajuste

Idealmente, a deposição de energia nos calorímetros não segue uma distribuição gaussiana. Apesar disso, a primeira abordagem da colaboração ATLAS foi o ajuste de um modelo gaussiano à distribuição de massa simulada e à distribuição de massa reconstruída pelo detector. Essa abordagem não foi aprovada no teste estatístico do χ^2 , que, neste caso, consiste em um teste aplicado a uma distribuição que foi modelada por um histograma. De posse do histograma, a valor de χ^2 é calculado seguindo a equação 4.8, onde O_i é o valor do modelo a ser ajustado no ponto i e o valor E_i é o valor estimado da distribuição pelo histograma no ponto i .

O número de graus de liberdade (*degree of freedom*) é igual a quantidade de células com as quais o histograma foi criado. Cada célula do histograma é chamada de *bin*. Depois do cálculo do valor de χ^2 , este é normalizado pelo número de graus de liberdade e assim é obtido o grau de confiabilidade do ajuste do modelo.

$$\chi^2 = \sum_{i=1}^{N_{eventos}} \frac{(O_i - E_i)^2}{E_i} \quad (4.8)$$

Breit-Wigner

Na mecânica clássica, a equação de movimento é dada pela segunda lei de Newton, ou $F = m \cdot a$. Esta equação e suas derivadas são usadas para resolver o movimento de um sistema mecânico e matematicamente prever o que o sistema fará a qualquer instante de tempo, independente das condições iniciais e da configuração do sistema.

Analogamente, a equação de Schrödinger (4.9, em sua forma completa), descreve matematicamente a evolução do estado quântico de um sistema ao longo do tempo. Na equação 4.9, temos que m_p é a massa da partícula, V é a energia potencial e ψ

é a função de onda.

$$i\hbar \frac{\partial}{\partial t} \psi(r, t) = \left[\frac{-\hbar^2}{2m_p} \nabla^2 + V(r, t) \right] \psi(r, t) \quad (4.9)$$

O quadrado da função de onda é a probabilidade da partícula se encontrar em um determinado estador de energia-momento (ou posição-tempo) tal como podemos ver na equação 4.10.

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 dx = \int_{-\infty}^{+\infty} \psi(x) \psi(x)^* dx = 1 \quad (4.10)$$

Quando um conjunto de partículas decai, podemos modelar esse decaimento como visto na equação 4.11, onde $n(t)$ é o número de partículas em um determinado instante de tempo, $n(0)$ é o número de partículas no instante inicial e Γ é uma constante, sendo a vida-média das partículas dada por $\tau = 1/\Gamma$

$$n(t) = n(0) \cdot e^{-\Gamma t} \quad (4.11)$$

A probabilidade de uma partícula decair no instante t é dada pela equação 4.12.

$$|\psi(t)|^2 = |\psi(0)|^2 e^{-\Gamma t} \quad (4.12)$$

Da equação 4.12 deriva-se a equação 4.13.

$$\psi(t) \sim \exp^{-iMt} \cdot e^{\frac{-\Gamma t}{2}} \quad (4.13)$$

Podemos escrever o estado de energia $\chi(E)$ como sendo a transformada de Fourier da função de onda $\psi(t)$, obtendo a equação 4.14.

$$\chi(E) = \int \psi(t) e^{iEt} dt \sim \frac{1}{E - M + i(\Gamma/2)} \quad (4.14)$$

Calculado-se a probabilidade $|\chi(E)|^2$, equação 4.15, chega-se à função de probabilidade de Breit-Wigner, comumente utilizada para descrever a distribuição de massa reconstruída a partir dos produtos do decaimento de partículas instáveis.

À função de probabilidade de Breit-Wigner, um fator ξ foi adicionado para ponderar toda a função preservando sua área igual a 1, independente da massa M , da energia E e do tempo de meia-vida $\tau = 1/\Gamma$

$$|\chi(E)|^2 = \frac{\xi}{(E - M^2)^2 + (\Gamma/2)^2} \quad (4.15)$$

A equação 4.15 apresenta a função de probabilidade de Breit-Wigner elevada ao quadrado. Na equação 4.16 temos a função de probabilidade de Breit-Wigner em

sua forma geral.

$$\chi(E) = \frac{\Gamma \cdot M}{(E - M) - i(\Gamma \cdot M)} \quad (4.16)$$

Crystall Ball

SPEAR (*Stanford Positron Electron Asymmetric Rings*) foi um colisor de partículas no SLAC (Stanford Linear Accelerator Center). Começou a operar em 1972, colidindo elétrons e pósitrons com energia igual a 3 GeV no centro de massa. Durante a década de 1970, o acelerador foi de grande importância para a física de altas energias, tendo participação fundamental na descoberta do méson J/ψ , que rendeu o prêmio nobel de física de 1976, e a descoberta do lépton τ , que foi premiado com o nobel de física em 1995.

Um dos experimentos do SPEAR foi o Crystal Ball [85, 86], um detector de partículas hermético projetado para detectar somente partículas neutras. O méson η_c foi detectado pela primeira vez nesse experimento. A sua parte central é composta por uma *spark chamber*, envolvida por uma esfera de completa de cristais cintilantes de NaI(Tl) cobrindo 98% do ângulo sólido. Seu nome deriva dessa esfera de cristais. Esse detector também foi usado em experimentos de física do quark b no DESY, Alemanha, e atualmente é usado pela colaboração A2, no Microtron de Mainz, Alemanha, em experimentos com fótons de *bremsstrahlung*.

A Função Crystal Ball [87] é uma função de densidade de probabilidade comumente utilizada para modelar processos com perdas em física de altas energias. Trata-se de um modelo realizado com duas funções convoluidas: uma gaussiana e uma função erro de Gauss, também chamada de *erf*, dada na equação 4.17. A função e sua primeira derivada são, ambas, contínuas.

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.17)$$

Esta função foi proposta pela colaboração Crystal Ball em [88] e é apresentada na equação 4.18.

$$f(x, \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, & \text{para } -\frac{(x-\bar{x})}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{para } -\frac{(x-\bar{x})}{\sigma} \leq -\alpha. \end{cases}$$

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot e^{\left(\frac{-|\alpha|^2}{2}\right)}$$

$$B = \frac{n}{|\alpha|} - |\alpha|$$

$$N = \frac{1}{\sigma(C+D)} \quad C = \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot e^{\left(\frac{-|\alpha|^2}{2}\right)}$$

$$D = \sqrt{\frac{\pi}{2}} \left(1 + erf \left(\frac{|\alpha|}{\sqrt{2}} \right) \right) \quad (4.18)$$

Função de Bukin

Durante o processamento de dados experimentais, por vezes nos deparamos com a tarefa de ajustar uma distribuição a um modelo com um conjunto finito de parâmetros ajustáveis. Em distribuições com picos, esta tarefa se torna um pouco mais complexa, pois a posição do pico deve ser ajustada de forma precisa, bem como a determinação da largura da distribuição.

Na prática, podemos utilizar a soma ponderada de várias gaussianas para ajustar a distribuição de dados observados. Nesse caso, o pico será dado pelo valor máximo do modelo ajustado e a largura pela largura de meia altura do modelo.

Uma outra técnica que é amplamente utilizada é a divisão da distribuição a ser ajustada em regiões de ajuste e, tentando-se ajustar cada região com um modelo diferente. No caso do decaimento do bóson de Higgs, o pico da distribuição segue uma Breit-Wigner, enquanto suas caudas tem comportamento gaussiano, seguindo o erro de medida do detector, que é modelado como gaussiano.

Proposta por A. Bukin em 2008, a função de Bukin [89] foi desenvolvida para modelar distribuições com pico, que possuam assimetrias e caldas com comportamento gaussiano. Para isso, a primeira abordagem do autor foi a tentativa de convoluir uma gaussiana com uma função exponencial (função erro de Gauss - erf), seguindo o mesmo princípio da Crystal Ball, como pode ser visto na equação 4.19.

$$F(x, x_g, \sigma_g, \lambda) = \frac{1}{2|\lambda|} \left[1 - erf \left(-\frac{(x - x_g)\lambda}{\sigma_g |\lambda| \sqrt{2}} + \frac{\sigma_g}{|\lambda| \sqrt{2}} \right) \right] \times e^{\left[-\frac{(x - x_g)}{\lambda} \right] + \frac{\sigma_g^2}{2\lambda^2}} \quad (4.19)$$

Na equação 4.19, a primeira abordagem de Bukin apresenta uma indeterminação do tipo $0 \cdot \infty$ quando $x \rightarrow \infty$. Para evitar essa indeterminação, foi proposta a aproximação dada na equação 4.20.

$$\left[1 - erf \left(-\frac{(x - x_g)\lambda}{\sigma_g |\lambda| \sqrt{2}} + \frac{\sigma_g}{|\lambda| \sqrt{2}} \right) \right] \approx \frac{e^{-z^2}}{z \sqrt{\pi}} \left[1 + \sum_{k=1}^{k_0} \frac{(-1)^k (2k-1)!}{2^k z^{2k}} \right], k_0 \leq z^2 \quad (4.20)$$

Utilizando a variável z , descrita na expressão 4.21, em conjunto com a apro-

ximação vista na equação 4.20, temos a definição da função Bukin na equação 4.22

$$z = -\frac{(x - x_g) \lambda}{\sigma_g |\lambda| \sqrt{2}} + \frac{\sigma_g}{|\lambda| \sqrt{2}} \quad (4.21)$$

$$F_{Bukin}(x, x_g, \sigma_g, \lambda) = \frac{1}{2|\lambda|z\sqrt{\pi}} e^{-\left(z - \frac{\sigma}{\sqrt{2}|\lambda|}\right)^2} \left[1 + \sum_{k=1}^{k_0} \frac{(-1)^k (2k-1)!}{2^k z^{2k}} \right] \quad (4.22)$$

4.3.3 Determinação da Massa Invariante do Bóson de Higgs

Uma vez que um dos modelos descritos acima tenha sido ajustado para a distribuição de massa do bóson de Higgs, a determinação do valor da massa se torna o pico da distribuição e seu erro de medida, a largura do modelo ajustado à distribuição.

Já a Resolução de Energia para Jatos (*Jet Energy Resolution* - JER) pode ser calculada como sendo a estimativa da variância do modelo ajustado dividido pela estimativa da média do modelo ajustado e seu erro pode ser obtido com a propagação do erro de cada uma das estimativas.

4.4 Métodos de Análise Estatísticos

Como mostrado anteriormente, temos 4 modelos que podem ser usados na determinação de massa do bóson de Higgs no detector ATLAS. Para medir a aderência de cada um dos modelos aos dados e às distribuições algumas análises estatísticas serão aplicadas.

4.4.1 Estimação da função densidade de probabilidade

Com os modelos ajustados, temos que comparar cada um deles utilizando métricas estatísticas para comprovar, de maneira eficaz, a sua aderência à distribuição e sua correta extração de características do modelo.

Para a avaliação de aderência, faz-se necessária a estimação da função densidade de probabilidade (PDF). Para isso, neste trabalho, a estimação foi feita utilizando uma função de *kernel*, que são funções não negativas e simétricas que podem ser ponderadas e somadas às outras para a estimação não-paramétrica de PDF. A função utilizada para a estimação foi a função de Epanichkov [90] e utilizados 50 núcleos de funções (quantidade de inicializações da função). Na figura 4.5, a comparação entre o kernel de Epanechnikov e o Kernel Box é apresentada

O Kernel Epanechnikov foi escolhido pois o mesmo minimiza o MISE (erro quadrático médio da área sob a função), que é uma métrica de extrema importância para se estimar a PDF. Uma outra característica que foi levada em conta, foi a sua eficiência⁸ que pode ser vista na tabela 4.2, com valores extraídos de [91].

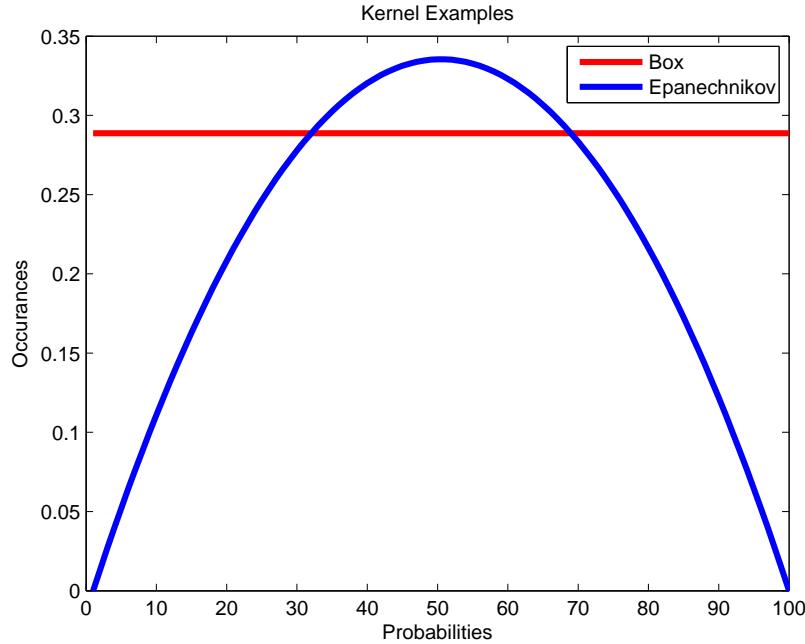


Figura 4.5: Distribuições dos kernels *box* e de Epanechnikov

Função Kernel	Eficiência
Epanechnikov	1.000
Biweight	0.994
Triangular	0.986
Gaussiano	0.951
Uniforme (Box)	0.930

Tabela 4.2: Eficiência das principais funções de Kernel, valores extraídos de [90]

4.4.2 Divergência de Kullback-Leibler

A divergência de Kullback-Leibler ou divergência KL [92] é uma divergência estatística que foi criada para medir a divergência no espaço de funções de densidade de probabilidade entre duas PDFs e/ou funções e, assim, pode ser utilizada como avaliação de quão próximo a distribuição da massa corrigida está da distribuição da massa simulada.

Um conceito importante para a divergência KL é a entropia contida nas variáveis aleatórias envolvidas. A entropia diferencial é uma identidade matemática que busca

⁸A eficiência para um Kernel é medida utilizando o AMISE (MISE assintótico)

medir a quantidade de aleatoriedade que uma distribuição possui e pode ser definida na equação 4.23, onde $H(X)$ é a entropia, X é a variável aleatória discreta e $P(x_i)$ é a função densidade de probabilidade no ponto x_i .

$$H(X) = - \sum_i P(X = x_i) \log P(X = x_i) \quad (4.23)$$

Um outro conceito utilizado para definir a divergência KL é a entropia cruzada (*cross entropy*), que pode vista na equação 4.24, onde P e Q são distribuições de variáveis aleatórias distintas.

$$H(P, Q) = - \sum_i P(X = x_i) \log Q(X = x_i) \quad (4.24)$$

De posse das definições de entropia e entropia cruzada, podemos definir a divergência KL como na equação 4.25.

$$D_{KL}(P\|Q) = H(P, Q) - H(P) = \sum_i P(X = x_i) \cdot \log \left(\frac{P(X = x_i)}{Q(X = x_i)} \right) \quad (4.25)$$

A divergência KL é não simétrica porque $D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$ e por isso não pode ser chamada de distância. Devido a isso, foi criada a divergência SKL (*Symmetrised Kullback-Leibler Divergence* ou Divergência KL simetrizada), dada na equação 4.26.

$$D_{SKL}(P\|Q) = \frac{D_{KL}(P\|Q) + D_{KL}(Q\|P)}{2} \quad (4.26)$$

4.4.3 Divergência de Jensen-Shannon

A divergência de Jensen-Shannon (JS) [93] é baseada na divergência KL mas é simétrica, não-negativa, possui cota superior em 1 e definida segundo a equação 4.27, onde $M = \frac{P(X)+Q(X)}{2}$, eliminando assim o problema da falta de limite superior das divergências KL e SKL.

$$D_{JS}(P\|Q) = \frac{D_{KL}(P\|M) + D_{KL}(Q\|M)}{2} \quad (4.27)$$

4.4.4 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov (T_{KS}) compara duas PDFs medindo a distância entre suas funções de distribuições cumulativas (CDFs). O teste pode ser definido

através da divergência de Kolmogorov-Smirnov (D_{KS}).

$$D_{KS}(P\|Q) = \max_x \left\| \int_{-\infty}^x (P(X=x) - Q(X=x)) dx \right\| \quad (4.28)$$

A probabilidade de duas PDFs serem similares, utilizando a divergência KS, pode ser obtida através da equação 4.29.

$$T_{KS}(D_{KS}) = 2 \cdot \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 D_{KS}(P\|Q)^2} \quad (4.29)$$

Capítulo 5

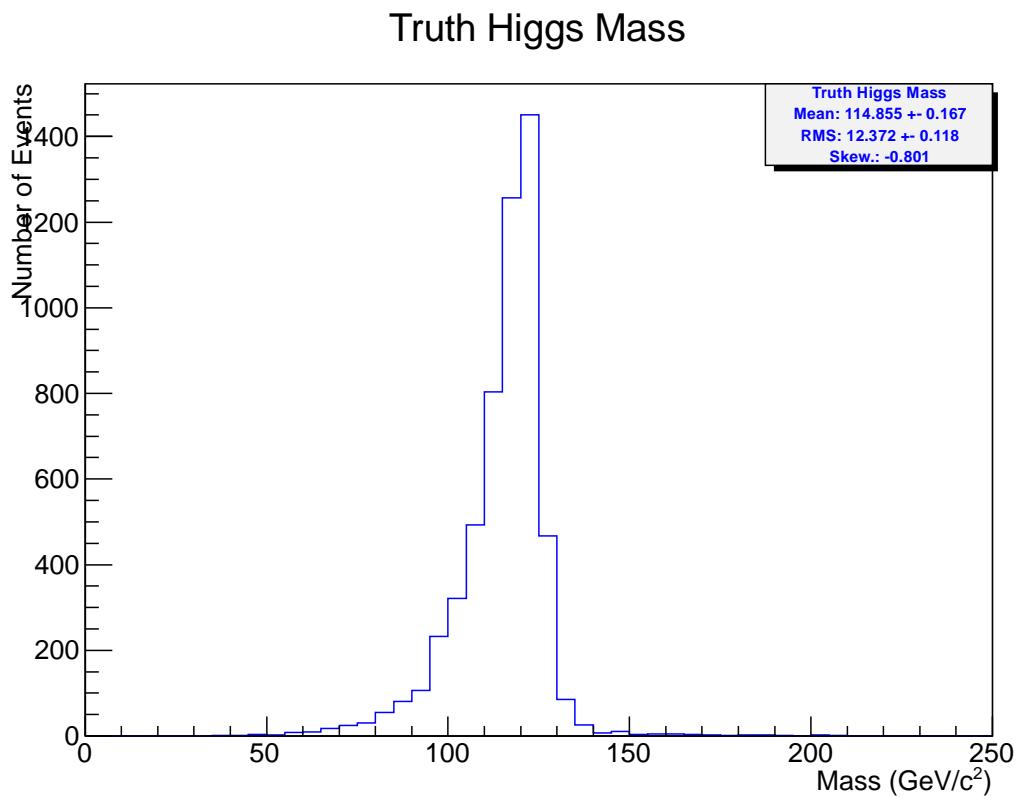
Resultados Obtidos

Neste capítulo, serão mostrados os resultados obtidos. Primeiramente serão mostradas a distribuição da massa para os bósons de Higgs simulados com 125 GeV de massa e a distribuição da massa reconstruída a partir de dados registrados pelo detector ATLAS, chamada apenas de massa reconstruída. Depois, a modelagem de *JER* será demonstrada, bem como dois modelos neurais utilizados como corretores de jatos. Em seguida, será desenvolvida a análise de relevância, bem como os corretores neurais que utilizaram as entradas da análise de relevância. A última parte do capítulo se destinará a apresentar os resultados das análises estatísticas.

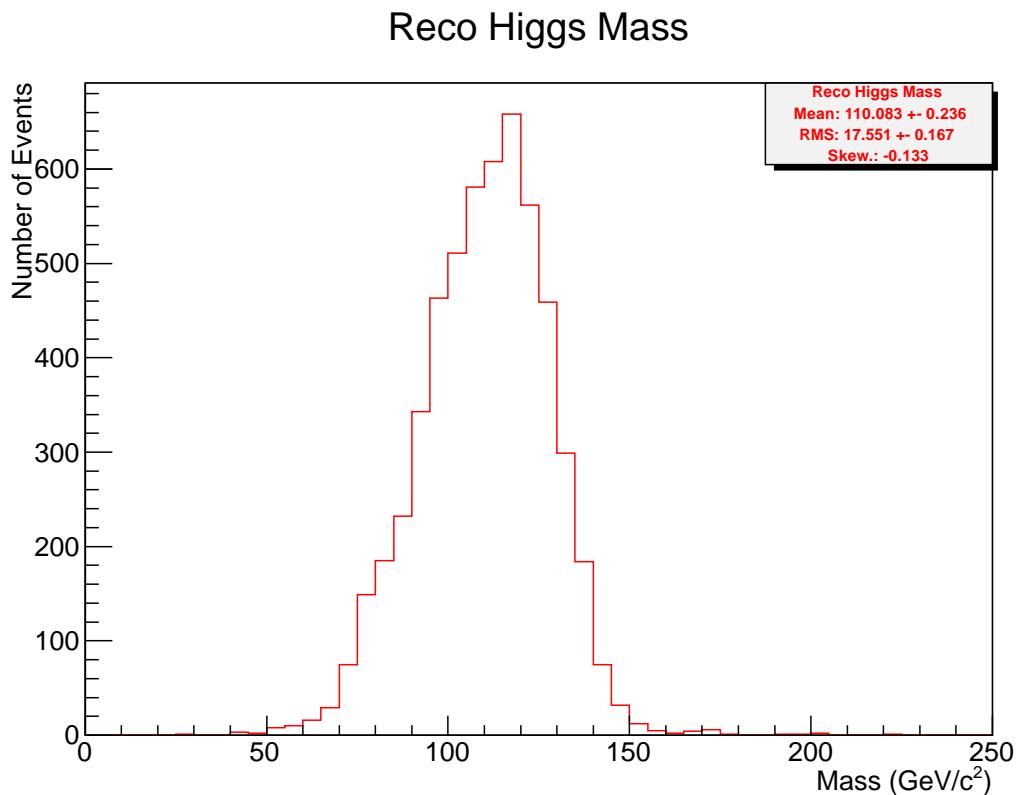
5.1 Massa reconstruída do bóson de Higgs

Como dito no capítulo anterior, a distribuição de massa do bóson de Higgs segue um modelo de Breit-Wigner que, devido a erros instrumentais do detector, deve ser convoluída com uma gaussiana. Na figura 5.1(a) podemos ver em azul a distribuição esperada para a massa do bóson de Higgs proveniente de dois jatos b .

A reconstrução feita a partir dos dados registrados pelo detector possui não-linearidades inerentes ao projeto. As não-linearidades podem ser modeladas como sendo provenientes, principalmente, de regiões sem instrumentação (os criostatos e cabeamentos que se encontram dentro do detector) e de dispersão do chuveiro nas células vizinhas a célula medida. Tal fato faz com que a resposta do detector se apresente, por vezes, muito diferente do que é esperado. Na figura 5.1(b) podemos ver a distribuição da massa reconstruída do bóson de Higgs e na figura 5.2 temos o *scatter plot* da massa simulada *versus* a massa reconstruída. Ainda nesta figura, vemos, através do ajuste de uma reta, que a reconstrução do ATLAS subestima a massa do bóson de Higgs (o coeficiente angular da reta é 0,402, ou seja, menor do que 1)



(a) Distribuição para massa simulada (massa verdadeira) do bóson de Higgs



(b) Distribuição da massa reconstruída do bóson de Higgs

Figura 5.1: Distribuições para a Massa do bóson de Higgs

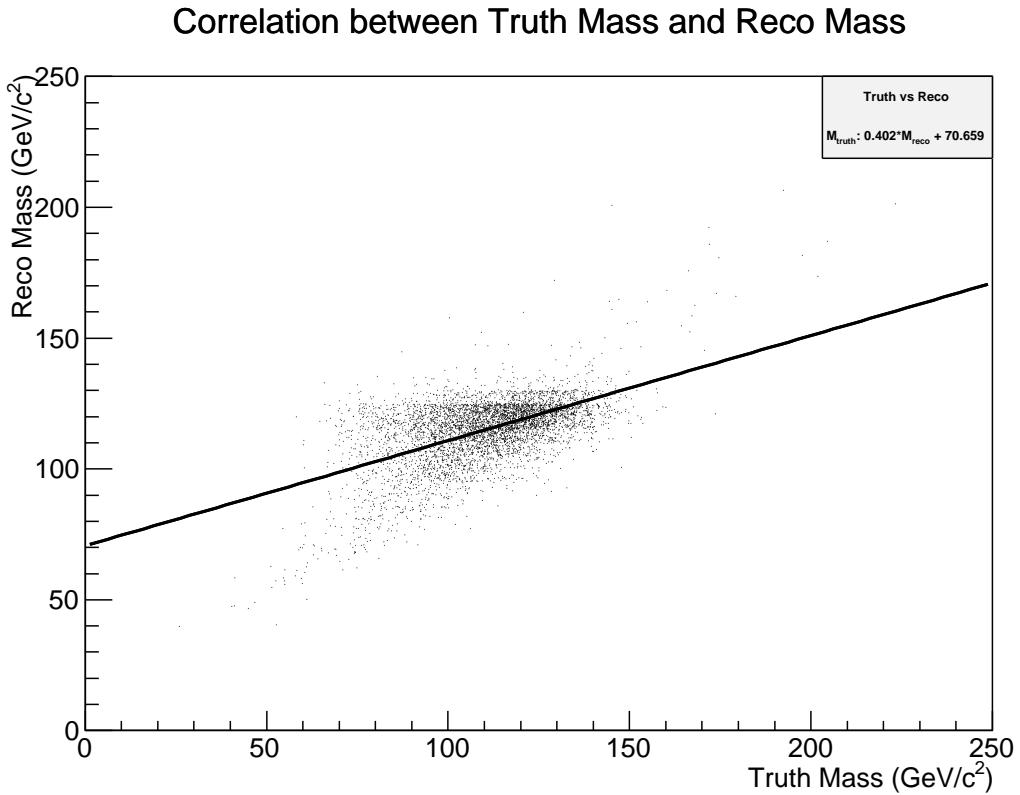


Figura 5.2: *Scatter plot* da massa simulada *versus* massa reconstruída

5.1.1 Análise de massa do bóson de Higgs para o detector ATLAS

No capítulo anterior, foi introduzido o conceito de *JER* (*Jet Energy Resolution*), que busca avaliar a resolução na medida dos jatos pelo detector e pode ser modelado como visto na equação 5.1.

$$\frac{\sigma_{Pt}}{Pt} = \sqrt{\left(\frac{a}{Pt}\right)^2 + \left(\frac{b}{\sqrt{Pt}}\right)^2 + (c)^2} = \frac{a}{Pt} \oplus \frac{b}{\sqrt{Pt}} \oplus c \quad (5.1)$$

A equação 5.1 é muito semelhante à equação de resolução de energia apresentada no capítulo 3. Nesta equação, o termo a agrupa as correções dependentes do tamanho das diferentes células que compõem o sistema de medida do momento. O termo b é dependente do sistema de amostragem e corrige, de forma sequencial, as não-linearidades ao longo da linha de propagação do jato. O termo c é dependente de erros sistemáticos do detector [11].

Para modelar os efeitos que alteram o valor do *JER*, a colaboração ATLAS faz o ajuste de um modelo gaussiano à massa simulada e à massa reconstruída. De posse do modelo, podemos calcular o valor do *JER*, como pode ser visto na figura 5.3, que no caso da massa simulada é de 0,108, para o histograma e de 0,066, para o

modelo gaussiano. Já para a massa reconstruída, o valor de *JER* é de 0,159, para o histograma e de 0,151, para o modelo gaussiano. Ou seja, da distribuição da massa simulada para a distribuição da massa reconstruída, o *JER* obteve uma variação de 32,07% (histograma) e 56,29% (modelo).

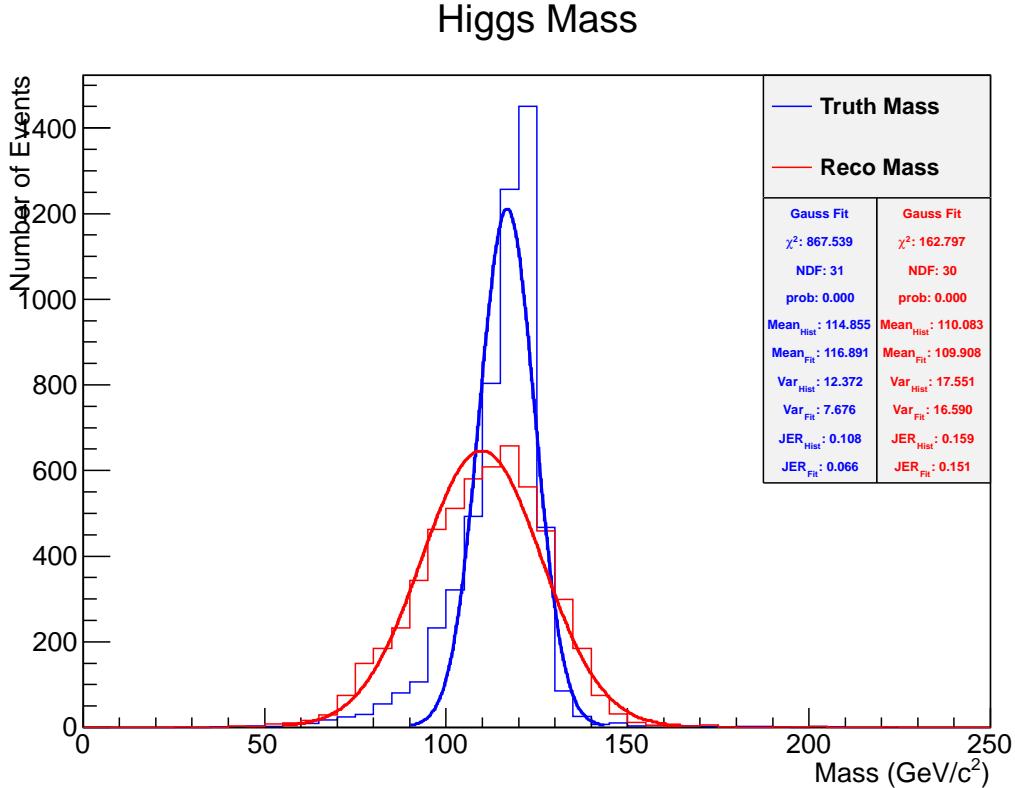
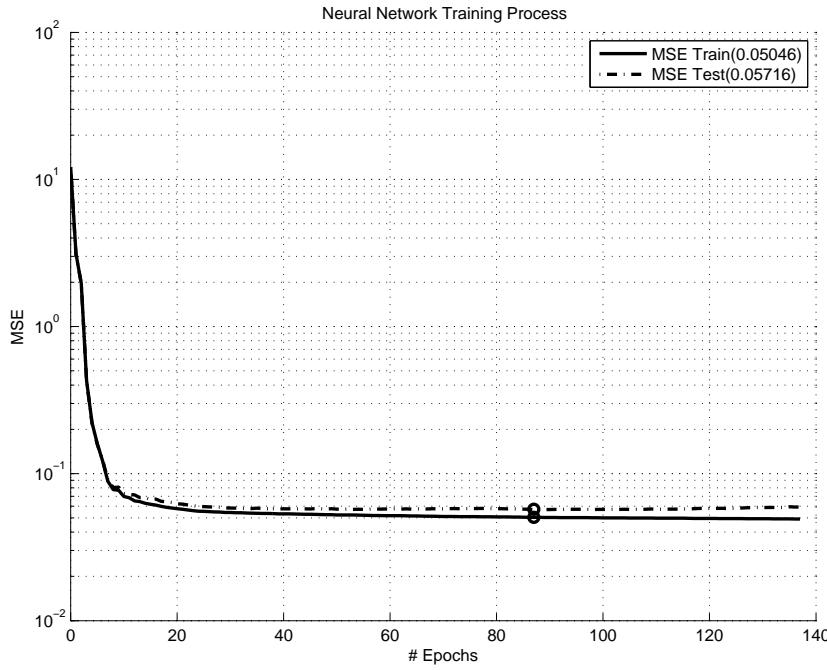


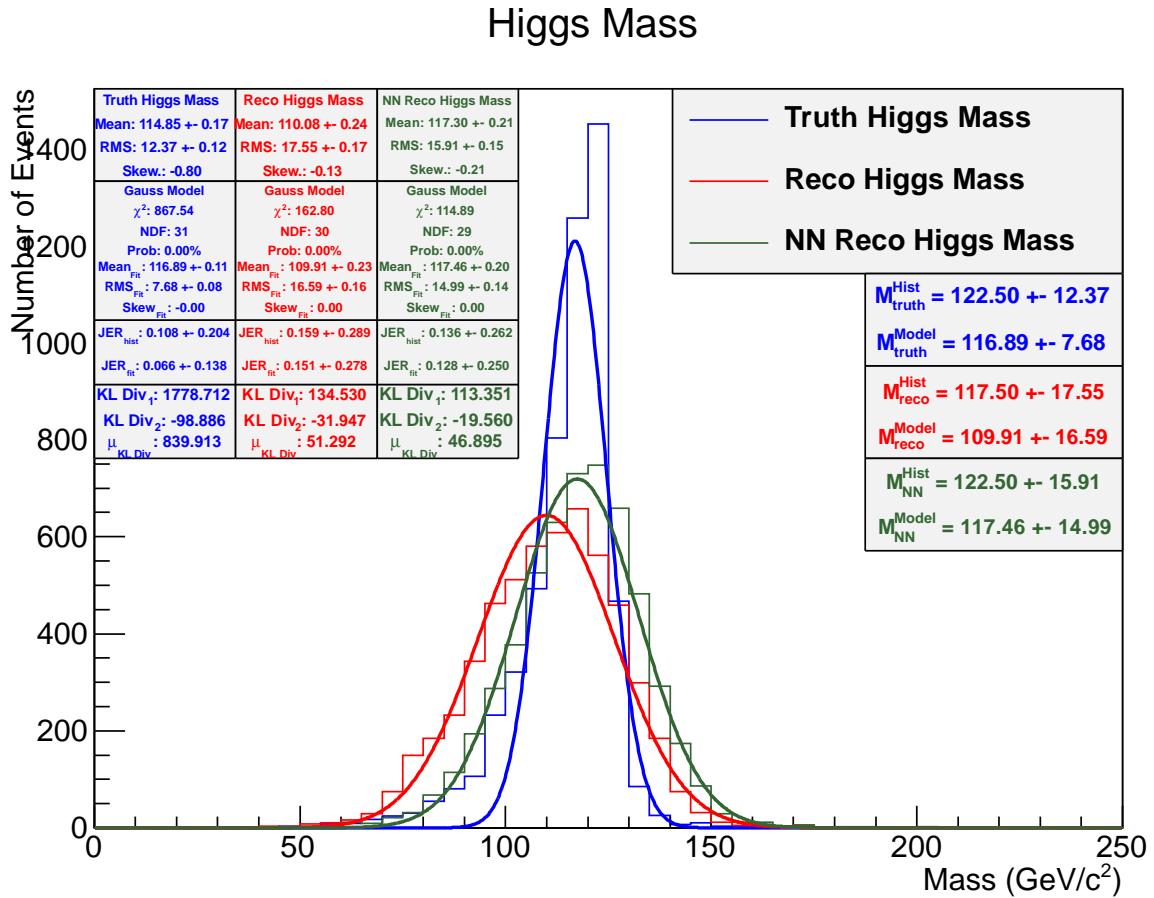
Figura 5.3: Distribuições de massa do bóson de Higgs com o ajuste gaussiano

5.2 Primeiro Modelo Neural aplicado à Massa do Bóson de Higgs

No capítulo anterior foi descrito como uma correção multiplicativa pode corrigir a massa do bóson de Higgs e também foi apresentado um modelo neural que poderia gerar essa correção multiplicativa. Com essas informações, a colaboração ATLAS propôs um modelo neural através do grupo de pesquisas da Universidade de Glasgow [94]. Este modelo utilizava como variáveis de entrada: o momento transverso reconstruído (*rPt*), a soma reconstruída dos momentos associados ao jato pelo *Inner Detector* (*rSumPtTrk*) e a largura do jato (*rWidth*), para ambos os jatos. A topologia utilizada foi de uma única camada escondida com 10 neurônios na camada, a função de treinamento foi baseada no Gradiente Descente e a função custo foi o erro médio quadrático (MSE).



(a) Erros médios quadráticos de treinamento e de teste no durante o processo de treinamento



(b) Distribuições de Massas Simulada, Reconstruída, Reconstruída com Rede Neural

Figura 5.4: Resultados para o conjunto de variáveis sugeridos pela colaboração ATLAS, com o treinamento de primeira ordem

Na figura 5.4(a) podemos observar que com o conjunto de variáveis, o treinamento não apresenta o processo de *overfitting*, pois o erro no conjunto de teste não aumentou ao longo do desenvolvimento do treinamento. Vemos também que os valores dos mínimos da função custo encontram-se na mesma ordem de grandeza, o que mostra que temos um treinamento consistente e generalizado (pelo menos para os conjuntos de treinamento e teste).

Na figura 5.4(b) temos o principal resultado da correção neural da massa do bóson de Higgs. Comparando os histogramas para a massa simulada (azul) e massa reconstruída (vermelho) podemos ver que as distribuições são bastante diferentes. No que tange o ajuste do modelo, o gaussiano foi rejeitado pelo teste do χ^2 (ajuste do modelo com 0,000% de confiabilidade) em todas as distribuições, com isso, os valores ajustados pelo modelo gaussiano não representam as distribuições em questão. Observando o resultado do *JER*, vemos que, tanto para os modelos quanto para as distribuições, os valores obtidos com a correção neural se aproximam dos valores obtidos a partir dos dados simulados, o que nos faz concluir que, pelo menos para esse parâmetro, a correção neural proposta pela colaboração ATLAS apresenta um resultado positivo.

Tendo como referência a massa estimada do bóson de Higgs podemos ver que a correção neural acerta o valor da massa simulada ($122,5 \text{ GeV}/c^2$), embora o espalhamento em torno desse valor seja levemente mais elevado, em torno de 22,25%, enquanto a massa reconstruída sem a correção neural, tem um valor cerca de 4,08% menor do que o valor simulado.

Comparando os *scatter plots* (figura 5.5) da reconstrução sem a correção neural com a reconstrução com a rede neural, vemos que a correção neural aproxima o coeficiente angular em 2,66% do valor ideal. Temos também uma redução no coeficiente linear de $70,65 \text{ GeV}/c^2$ para $66,42 \text{ GeV}/c^2$.

5.3 Segundo Modelo Neural aplicado à Massa do Bóson de Higgs

De posse dos resultados obtidos anteriormente, um treinamento de segunda ordem foi aplicado ao modelo neural para uma melhora nos resultados. O treinamento de segunda ordem tem se mostrado mais rápido e eficiente no alcance dos mínimos globais da função custo ao longo do treinamento. A função de treinamento aplicada foi a função de Levenberg-Marquardt e seu resultado pode ser visto na figuras 5.6(a) (treinamento) e 5.6(b) (distribuições).

Observamos na figura 5.6(a) que o processo de *overfitting* que não afeta o resultado obtido, uma vez que os pesos finais da rede neural foram salvos antes da

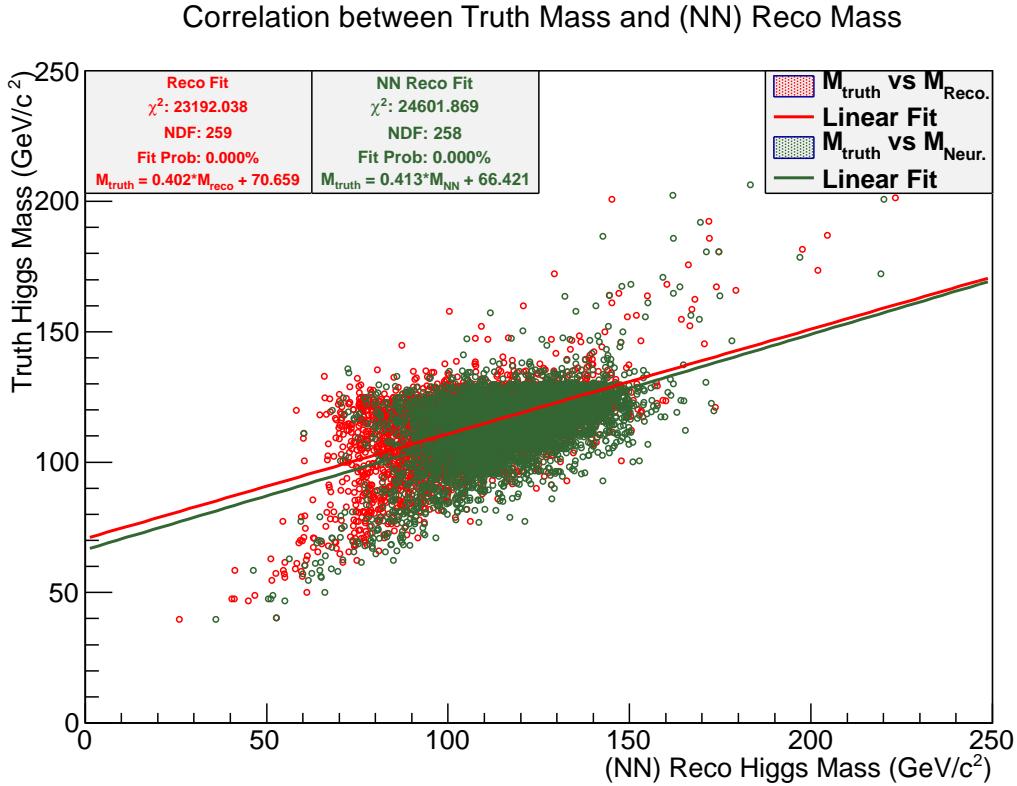


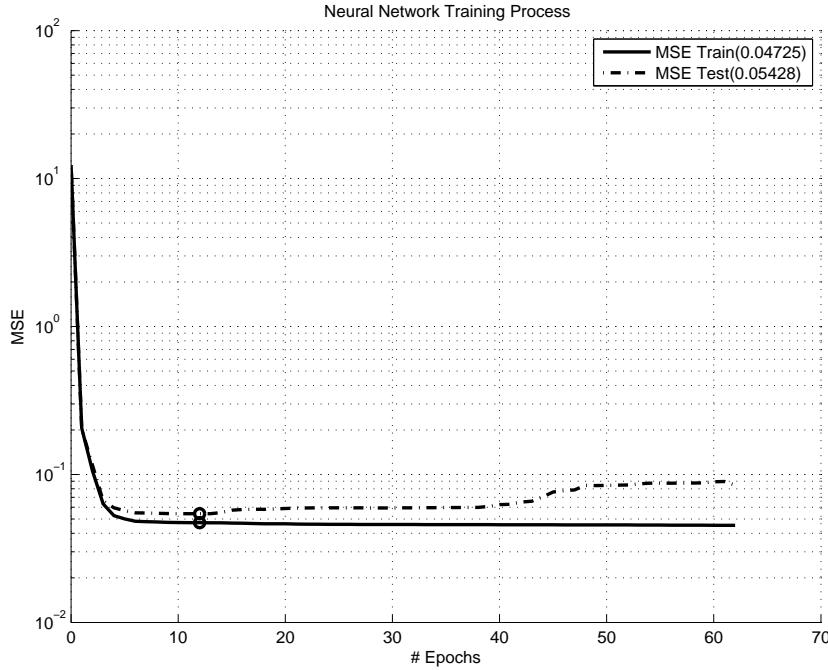
Figura 5.5: *Scatter plot* da massa simulada *versus* massa reconstruída com e sem a correção neural

ocorrência de *overfitting*. Vemos aqui, também, que o processo de treinamento converge mais rapidamente para o mínimo, tal como previsto teoricamente.

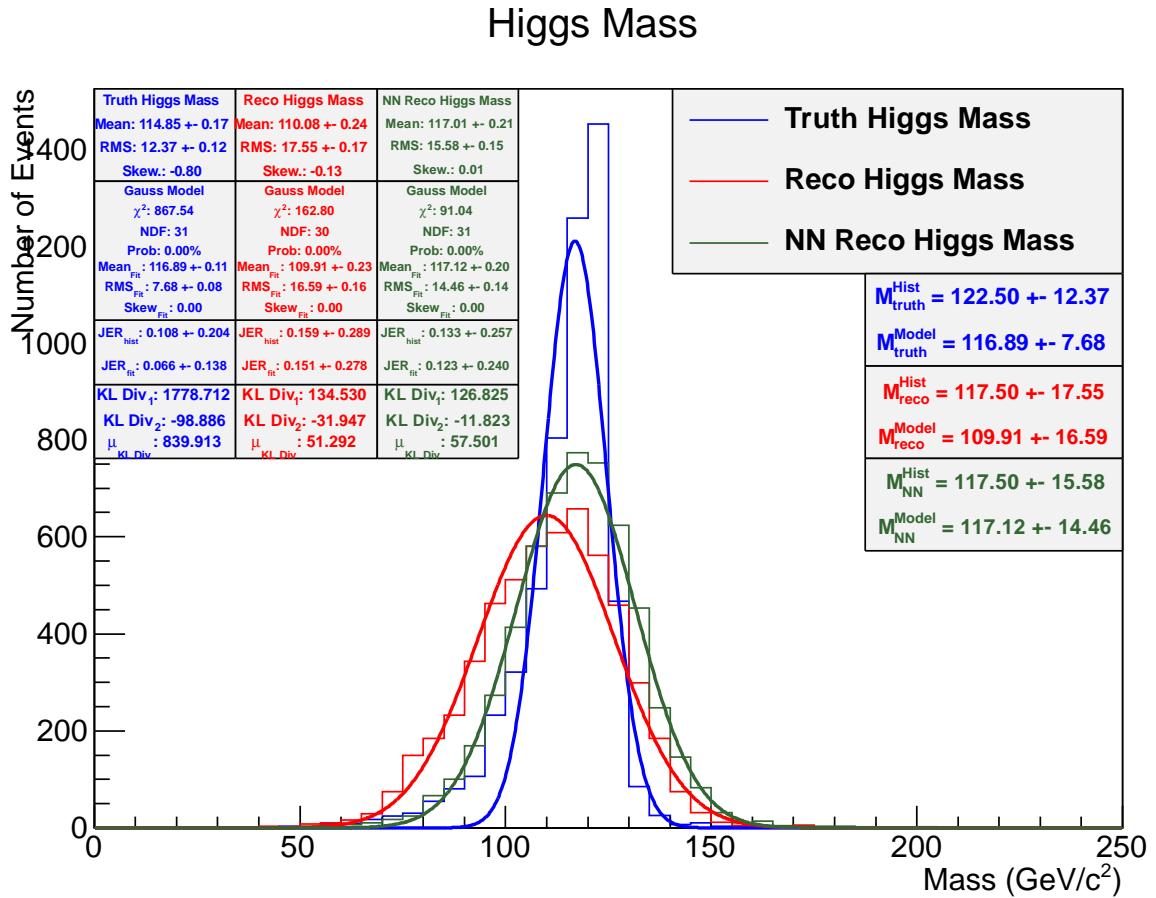
Com relação aos resultados expostos na figura 5.6(b), temos uma redução nos valores de *JER* obtidos com a correção neural, o que sugere que a reconstrução com a correção baseada no primeiro algoritmo de treinamento obteve um resultado melhor do que a reconstrução sem a correção neural. Isto pode ser comprovado com o valor da massa do bóson estimada, que se iguala, neste treinamento, ao valor de massa simulada. Embora, com a observação mais detalhada da figura 5.6(b) podemos ver que a distribuição de massa reconstruída com a correção neural (verde), bem como o seu modelo, obtiveram valores mais próximos da distribuição de massa simulada para o bóson de Higgs.

No *Scatter plot* apresentado na figura 5.7, vemos que o coeficiente angular aumentou 3.05% do anteriormente obtido, aproximando-se assim do valor ideal, o que pode ser considerado um pequeno ganho, enquanto o coeficiente linear obteve uma diminuição de, aproximadamente, $1 \text{ GeV}/c^2$.

Com os testes de diferentes funções de atualização de pesos, um novo conjunto de variáveis foi inserido no modelo neural para que, com isso, seja obtido um melhor resultado.



(a) Erros médios quadráticos de treinamento e de teste no durante o processo de treinamento



(b) Distribuições de Massas Simulada, Reconstruída, Reconstruída com a Rede Neural

Figura 5.6: Resultados para o conjunto de variáveis sugeridos pela colaboração ATLAS, com o treinamento de segunda ordem

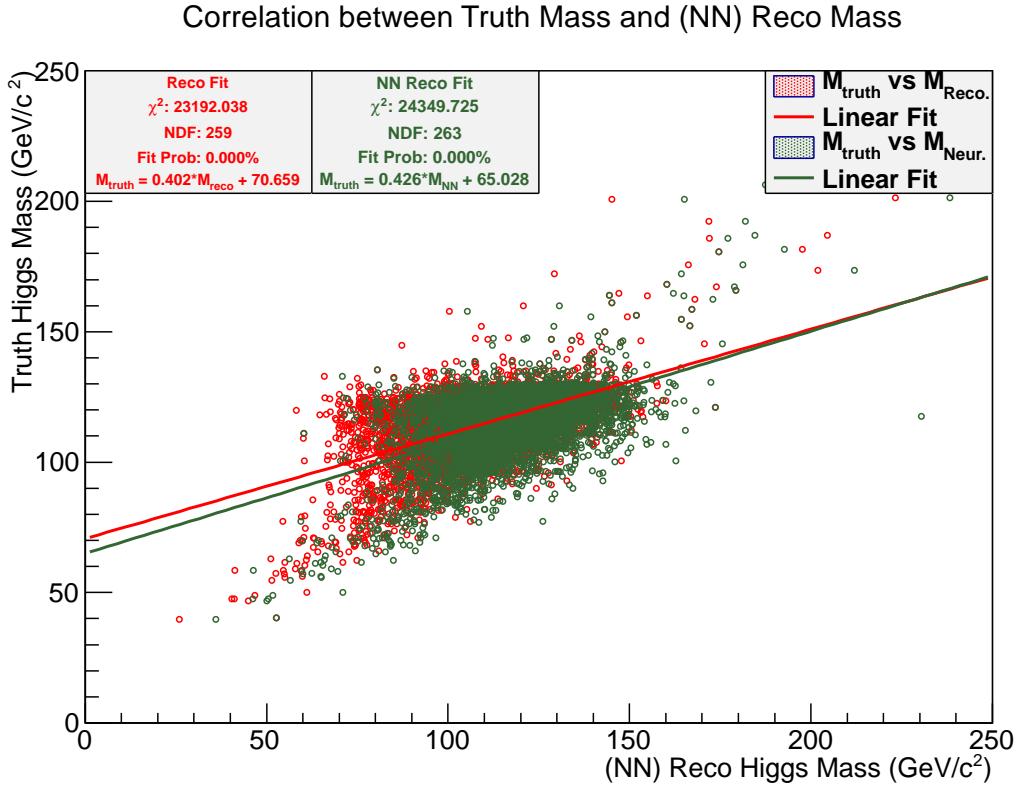


Figura 5.7: *Scatter plot* da massa simulada *versus* massa reconstruída com e sem a correção neural (treinamento de segunda ordem)

5.4 Análise de Relevância do Modelo Neural

O modelo neural tem uma grande dependência do conjunto de entrada. Para a criação de um modelo neural que gere uma correção para os dois jatos b , a colaboração ATLAS disponibilizou um conjunto de variáveis a serem estudadas. Na tabela 5.1 podemos ver esse conjunto detalhado em suas principais características, enquanto na figura 5.8 podemos ver as distribuições das variáveis tanto para o jato 1 (azul, linha sólida) quanto para o jato 2 (vermelho, linha pontilhada).

Antes de iniciarmos a análise do modelo neural, foi feita uma análise de topologia. Para isso, usamos o processo de validação cruzada (descrito no capítulo anterior) com 100 inicializações e 30 diferentes sorteios de eventos, variando a topologia de 5 a 20 neurônios. Com esse processo, obtivemos que a melhor rede neural treinada, isto é, a rede que apresentou o menor MSE médio¹, foi a rede neural de 10 neurônios na camada escondida.

De posse dessa rede, foi iniciado um processo de análise de relevância para a retirada de variáveis que não sejam relevantes ao modelo neural.

¹Média extraída nas 100 inicializações e 30 sorteios de eventos de treinamento.

Nome da Variável	Interpretação física
E	Energia total do Evento
E_t	Energia Transversa do Jato
Pt	Momento Transverso do Jato
η	Posição de Ocorrência do Jato (1 ^a coordenada)
ϕ	Posição de Ocorrência do Jato (2 ^a coordenada)
M	Massa associada ao Jato
M_t	Componente Transverso da Massa associada ao Jato
SumPtTrk	Soma dos momentos transversos associados ao Jato que foram detectados no Inner Detector
Width	Máximo valor dentro das distâncias dos traços associados ao Jato

Tabela 5.1: Características das Variáveis provenientes do banco de dados disponibilizado

5.4.1 Análise de Relevância tendo como Referência as Saídas

Para a análise da relevância de uma variável em um modelo neural, a primeira abordagem desenvolvida foi a Análise de Relevância tendo como Referência as Saídas.

Essa análise se baseia no fato que o modelo neural será sensível, isto é, apresentará uma maior variação em sua saída, quando variáveis relevantes sejam retiradas de suas entradas.

Para comprovar esse modelo, as redes treinadas com o processo de validação cruzada foram utilizadas. Assim, tomando-se uma rede neural por vez, as suas variáveis de entrada foram substituídas por suas médias e uma avaliação da saída foi obtida, tal como na equação 5.2

$$R_{output}(i) = \frac{\sum_{i=1}^{N_{events}} (Out_{x_i=x_i} - Out_{x_i=\bar{x}_i})^2}{N_{events}} \quad (5.2)$$

Depois da avaliação feita para todas as redes neurais de 10 neurônios na camada escondida, foi extraída a média e a variância de cada um dos valores de relevância e podemos ver nas figuras 5.9(a) e 5.9(b), os valores da análise de relevância para as saídas correspondentes às correções do jato 1 e do jato 2, respectivamente.

Como pode ser visto na figura 5.9(a), as entradas $j2rPhi$ (que é a posição reconstruída na coordenada ϕ em que o jato 2 foi detectado), $j2rPt$ (que é o valor do momento transverso reconstruído para o jato 2), $j1rEta$ (que é a posição reconstruída na coordenada η em que o jato 1 foi detectado), e $j2rEt$ (que é a energia transversa reconstruída para o jato 2) são as variáveis mais relevantes tendo como referência a correção para o jato 1. Vale ressaltar que as entradas $j2rPt$ e $j2rEt$ possuem uma barra de erro maior do que as das outras duas variáveis apontadas como relevantes,

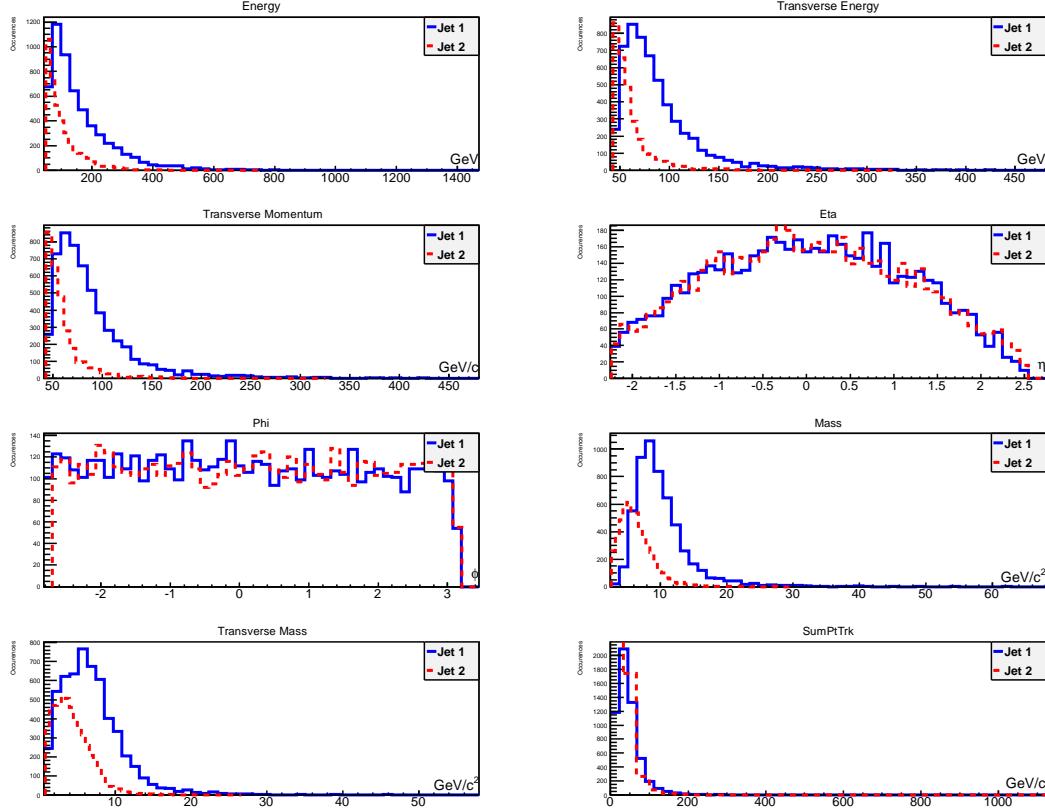
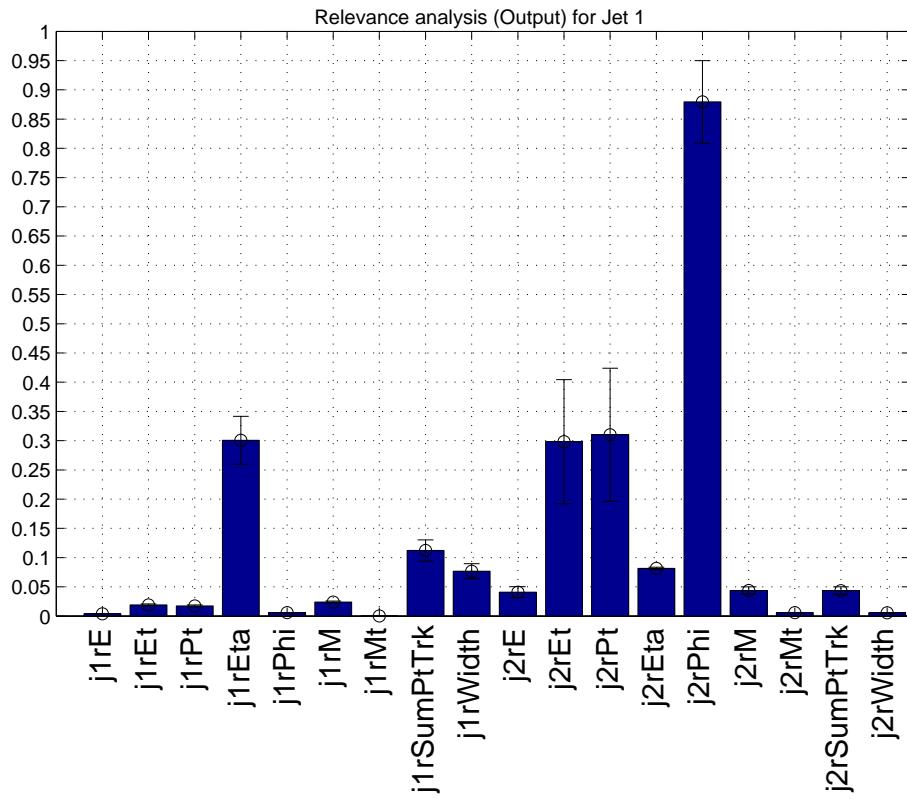


Figura 5.8: Distribuições das Variáveis da tabela 5.1

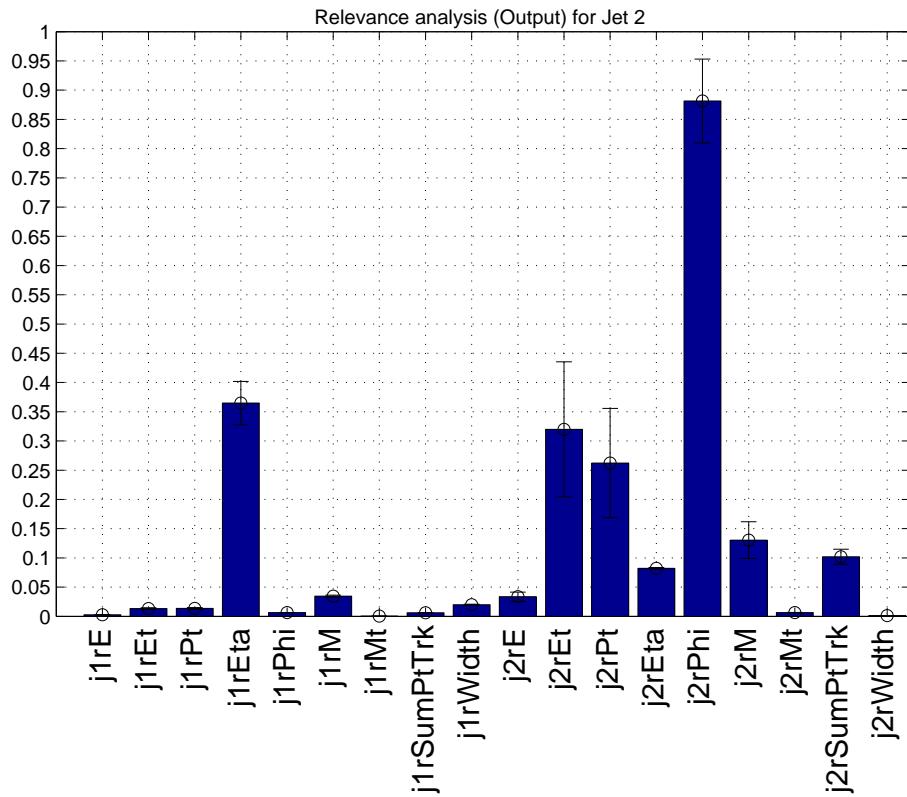
o que indica uma variância maior do valor de relevância calculado para cada um dos sorteios e conjuntos de entradas. Na figura 5.9(b), as mesmas entradas podem ser vistas como sendo relevantes.

Quando observamos a análise de relevância em conjunto com as distribuições apresentadas na figura 5.8, podemos notar que as variáveis com as distribuições mais próximas à uniforme, isto é $j2rPhi$ e $j1rEta$, foram mais relevantes para o modelo neural, pois, em uma distribuição uniforme (valores de *kurtosis*² menores que 1), a substituição da distribuição por seu valor médio reduziria consideravelmente a quantidade de informação introduzida por essa variável, uma vez que estas variáveis poderiam assumir qualquer valor (com probabilidades relativamente próximas) dentro do alcance (*range*) de suas distribuições, que são de $|\phi| < \pi$ e $|\eta| < 2,5$, respectivamente. Já as variáveis $j2rEt$ e $j2rPt$, na média, obtiveram valores de relevância maiores mesmo sendo variáveis com distribuições próximas a distribuições supergaussianas (distribuições com valores de *kurtosis* maiores que 1), o que indica que estas variáveis, mesmo com pequenas variações em função de suas

²A *kurtosis* é o quarto momento estatístico, como a média é o primeiro, a variância é o segundo e *Skewness* é o terceiro, e expressa o comportamento da distribuição ao redor da média, se a distribuição for uma gaussiana, a *kurtosis* assume o valor unitário, caso a distribuição seja próxima a uniforme, a *kurtosis* assume valores menores do que um e para supergaussianas temos valores de kurtosis maiores do que um.



(a) Análise de Relevância tendo como referência a Saída do Modelo Neural para o Jato 1



(b) Análise de Relevância tendo como referência a Saída do Modelo Neural para o Jato 2

Figura 5.9: Resultados para a Análise de Relevância com referência a Saída do Modelo Neural para ambos os jatos

médias, agregam informação ao modelo neural, tendo como referência a sua saída.

5.4.2 Análise de Relevância tendo como Referência o Alvo

Embora a primeira abordagem seja útil, esta tem como referência os valores de saída obtidos pela rede neural. Caso o mapeamento da rede neural contenha erros, esses erros podem ser propagados para a análise de relevância, comprometendo seus resultados.

Assim sendo, foi desenvolvida uma outra análise de relevância, que preserva o mesmo princípio da anterior, mas dessa vez, tendo como referência o alvo do treinamento, pois, a análise poderá observar qual variável faz com que o modelo se aproxime do alvo e consequentemente reduza o erro, como podemos ver na equação 5.3.

$$R_{target}(i) = \frac{\sum_{i=1}^{N_{events}} (Target - Out_{x_i=\bar{x}_i})^2}{N_{events}} \quad (5.3)$$

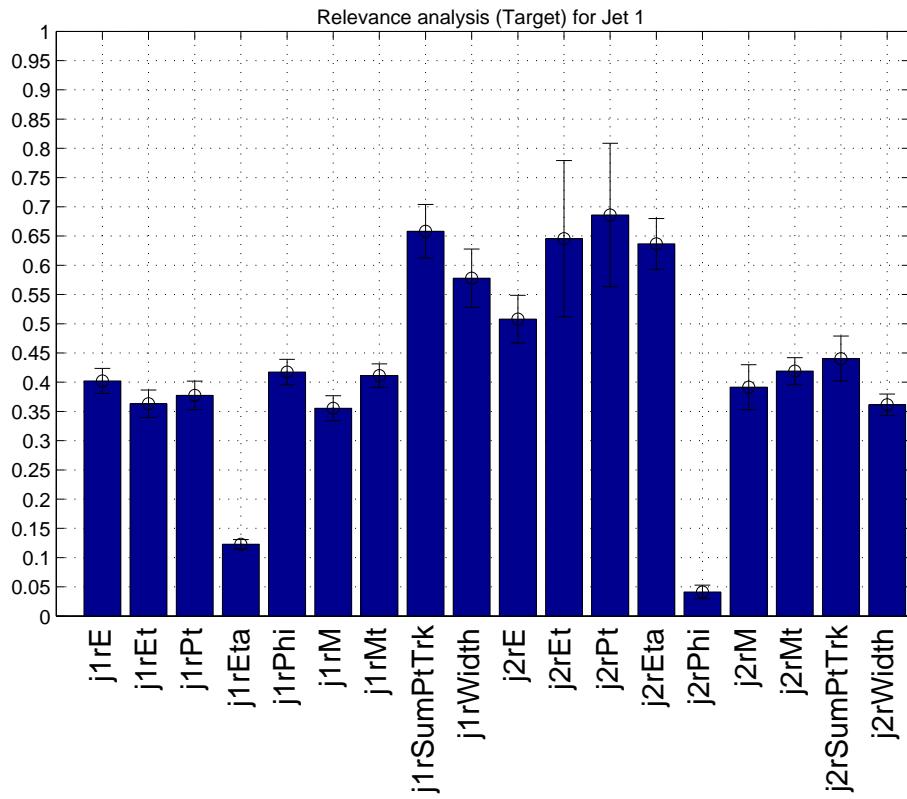
Nas figuras 5.10(a) e 5.10(b) podemos ver os resultados da análise de relevância tendo como Referência o Alvo do treinamento, para os jatos 1 e 2, respectivamente.

Na figura 5.10(a) vemos um comportamento bastante diferente do descrito nas figuras 5.9(a) e 5.9(b), e este pode ser descrito como um comportamento mais uniforme de cada uma das relevâncias apresentadas. Curiosamente, as variáveis *j1rEta* e *j2rPhi* são apontadas como as menos relevantes quanto tomamos a relevância tendo como referência o alvo. Outro ponto que pode ser ressaltado é que, com exceções *j2rEt* e *j2rPt*, as barras de erro são menores do que as apresentadas na análise tendo como referência as saídas.

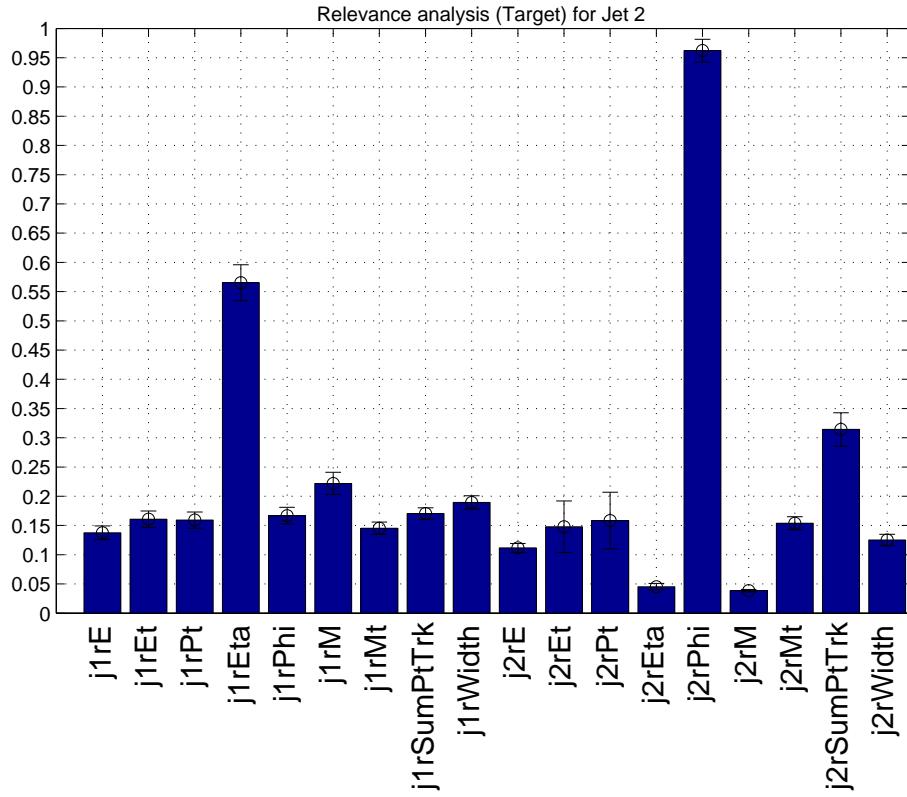
Observando as distribuições das variáveis (figura 5.8) podemos notar que mesmo variáveis que possuem comportamento supergaussiano (*kurtosis* maiores que 1), ou seja, tem sua distribuição bastante concentrada ao redor de sua média, obtiveram valores de relevância média alto. Este fato, nos permite concluir que estas variáveis são, de fato, relevantes para o modelo neural.

Já na figura 5.10(b), vemos que o comportamento apresentado na análise tendo como referência as saídas, volta a ser observado, isto é, mais uma vez as variáveis *j2rPhi* e *j1rEta* foram apontadas pela análise como sendo mais relevantes, embora, na figura 5.10(b) possa ser observado que as outras variáveis possuem relevância considerável, o que pode ser justificado, no caso de variáveis subgaussianas, pelo fato de termos pequenas variações em relação à sua média, ou seja, de maneira geral a sua relevância se mantém constante para quaisquer valores ao longo de sua distribuição.

Com esses resultados, vemos que as variáveis indicadas pela colaboração ATLAS são, de fato, variáveis relevantes para o modelo neural, já que nenhuma variável foi



(a) Análise de Relevância tendo como referência o Alvo do Modelo Neural para o Jato 1



(b) Análise de Relevância tendo como referência o Alvo do Modelo Neural para o Jato 2

Figura 5.10: Resultados para a Análise de Relevância tendo como referência o Alvo do Modelo Neural para ambos os jatos

excluída do conjunto de entradas pois nenhuma apresentou, em todas as análises, um nível de relevância que fosse suficientemente desprezível a ponto dessa variável ser excluída.

5.5 Terceiro Modelo Neural aplicado a Massa do Bóson de Higgs

De posse dos resultados da análise de relevância, um novo conjunto de variáveis de entrada foi aplicado ao treinamento neural de primeira ordem, que obteve melhores resultados de *JER* e de reconstrução de massa com as variáveis de entrada do modelo anterior.

Na figura 5.11(a) vemos que o treinamento de primeira ordem convergiu para um ponto de mínimo e no processo apresentou *overfitting*. Os erros médios quadráticos finais de treinamento e de teste se encontram na mesma ordem de grandeza, o que indica que a rede conseguiu agregar as características estatísticas do treinamento para os eventos de teste.

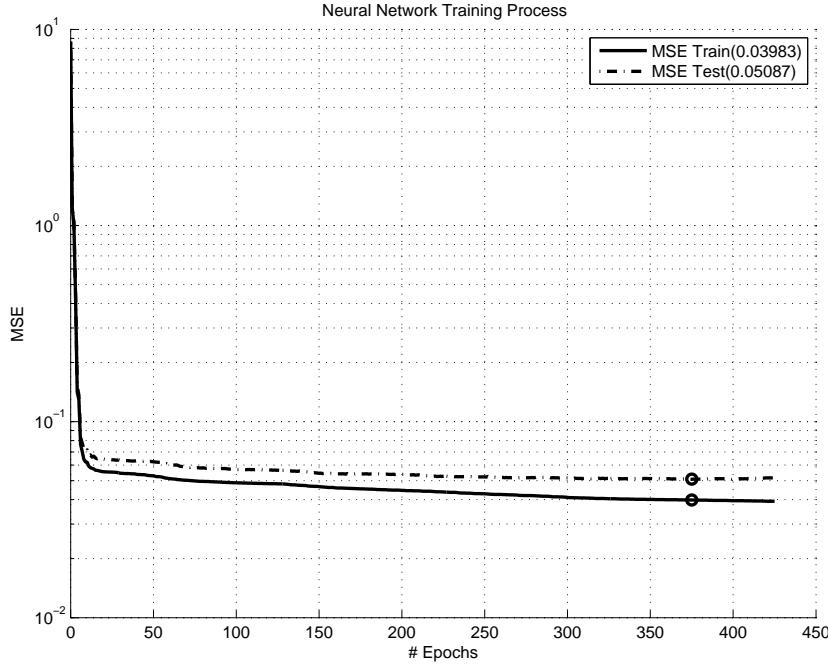
Os resultados para a reconstrução de massa do bóson de Higgs podem ser vistos na figura 5.11(b). Como podemos observar, a distribuição da massa reconstruída com a correção neural (distribuição em verde) consegue caracterizar o comportamento não gaussiano apresentado pela distribuição simulada (azul) em sua parte inferior (de 50 a 120 GeV), embora, não consiga obter uma boa caracterização na parte superior da distribuição (de 120 a 200 GeV).

Já no que tange a modelagem, vemos que o ajuste não foi aprovado no teste de χ^2 para nenhuma das distribuições. Embora o modelo da massa reconstruída com a correção neural se aproxime, visualmente, do modelo ajustado para a massa simulada.

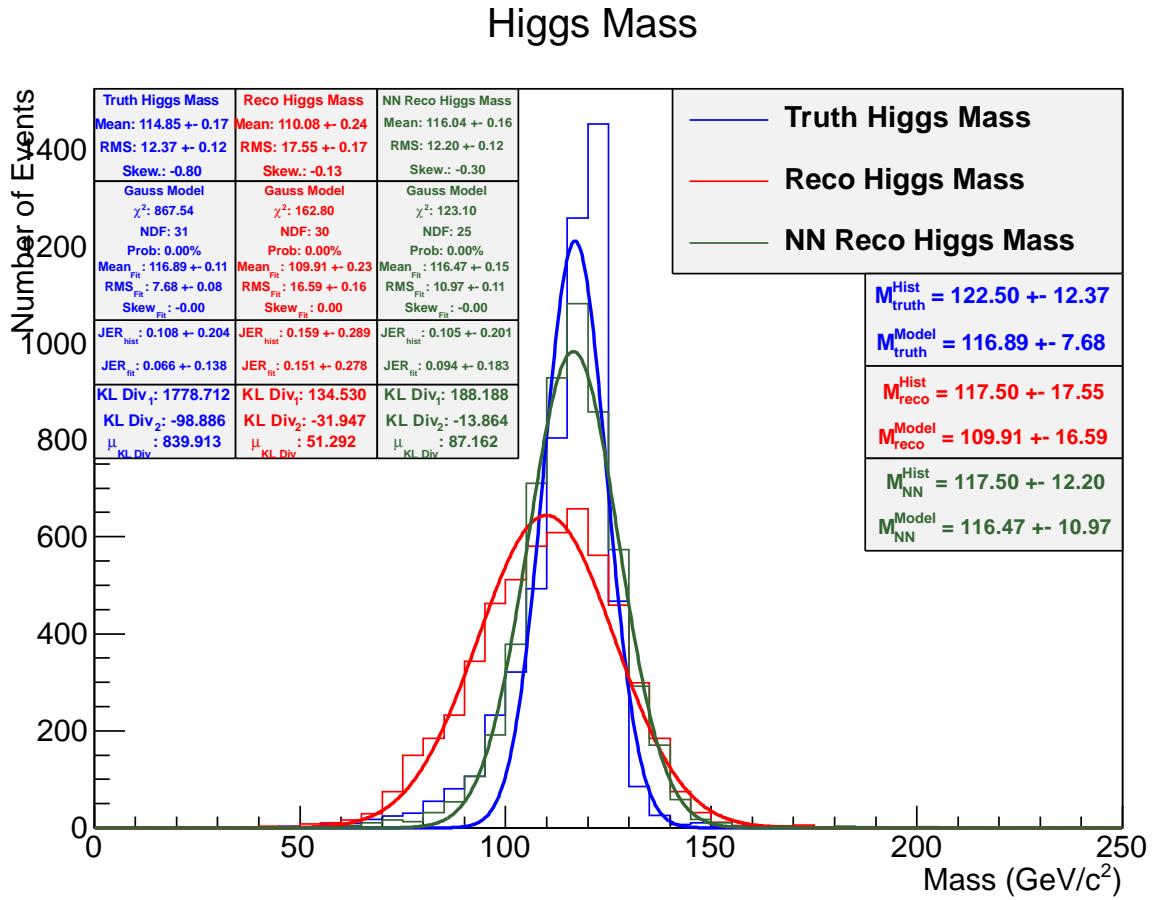
Para os valores obtidos de *JER* nas distribuições, podemos observar que o valor de *JER* obtido com a correção neural difere em 2,78% do valor de *JER* da massa simulada, com, aproximadamente, o mesmo erro de medida. Já o valor de *JER* da distribuição da massa reconstruída sem a correção neural (vermelho) difere em 32,07% em relação a distribuição da massa simulada.

Para os valores de *JER* obtidos tendo como referência os modelos ajustados, vemos que o valor de *JER* para a massa simulada é de 0,066, para a massa reconstruída sem a rede neural é de 0,151 e para a massa reconstruída com a correção neural é de 0,094. O que mostra que a correção neural apresentou um melhor resultado.

Com relação à reconstrução de massa, podemos ver que a massa estimada pelo modelo gaussiano se difere em 4,57% em relação ao valor de massa da distribuição simulada. O valor reconstruído de massa com correção neural difere em 5 GeV do



(a) Erros médios quadráticos de treinamento e de teste no durante o processo de treinamento



(b) Distribuições de Massas Simulada, Reconstruída, Reconstruída com a Rede Neural

Figura 5.11: Resultados para o conjunto de variáveis sugeridos pela análise de relevância com treinamento de primeira ordem

valor simulado, este valor é o valor de um *bin* neste histograma e também o valor de um *step* de energia na reconstrução dos dados. Os valores de massa reconstruídos sem a correção neural apresentam um maior espalhamento em relação à sua média, o que se reflete na maior incerteza da estimativa da massa.

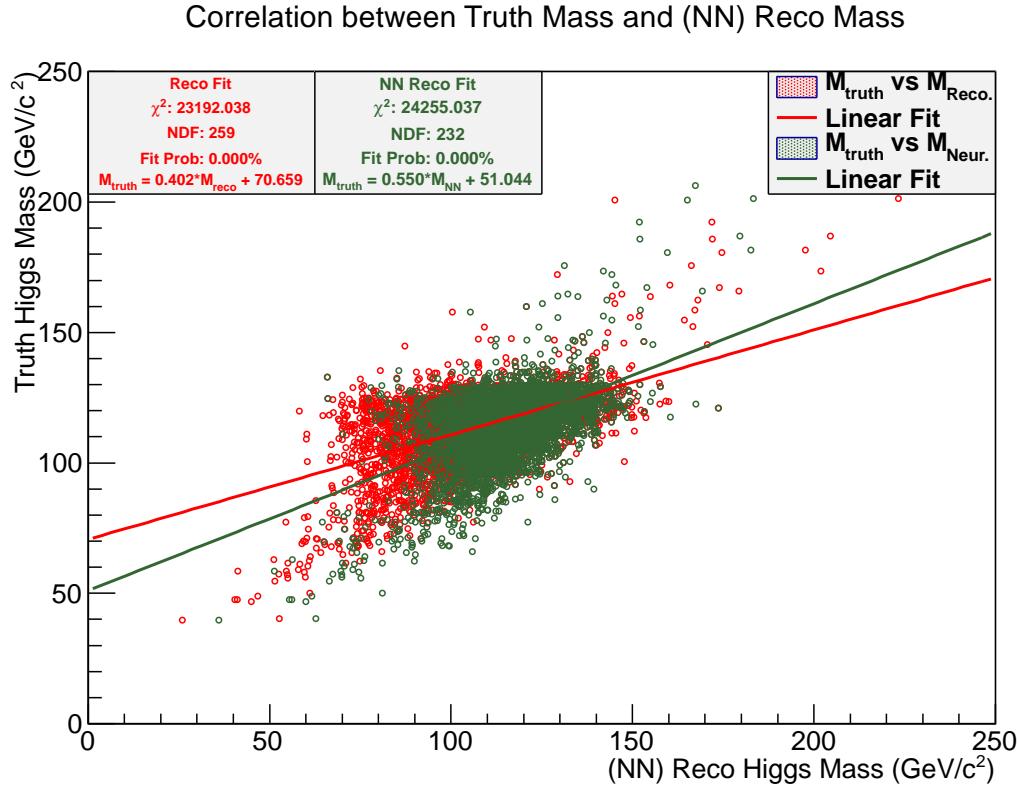


Figura 5.12: *Scatter plot* da massa simulada *versus* massa reconstruída com e sem a correção neural (treinamento de primeira ordem)

A correção neural de primeira ordem, aproximou do valor ideal tanto o coeficiente angular da reta, que passou de 0,402, na reconstrução sem correção neural, para 0,550 no caso da reconstrução com correção neural, com uma variação de 26,90%, quanto ao coeficiente linear, que passou de 70,659 GeV, no caso da reconstrução sem a correção neural, para 51,044 GeV, no caso da reconstrução com a correção neural, com uma variação de 27,76%. Esses valores podem ser observados na figura 5.12.

5.6 Quarto Modelo Neural aplicado à Massa do Bóson de Higgs

Com o mesmo conjunto de variáveis do terceiro modelo neural, mas com o treinamento de segunda ordem, um quarto modelo neural foi proposto para efeito de

comparação do efeito que um conjunto de variáveis que agrupa mais informação tem sobre um treinamento de segunda ordem.

Na figura 5.13(a) vemos que o processo de treinamento convergiu em torno de 75 *epochs*, consideravelmente mais rápido que as 375 *epochs* do treinamento anterior, sendo que os melhores pesos foram calculados na *epoch* de número 25.

No que diz respeito à reconstrução da massa, exposta na figura 5.13(b), mais uma vez o modelo gaussiano foi rejeitado pelo teste de χ^2 . Os modelos gaussianos para a massa simulada (linha em verde) e para a massa reconstruída com a correção neural praticamente coincidem, com pequenas variações.

Ainda nesta figura, podemos ver que a variação dos valores de *JER* obtidos para a distribuição de massa simulada e para a distribuição de massa reconstruída é de 18,51% e de 14,28% para os valores de *JER* obtidos com base nos modelos gaussianos.

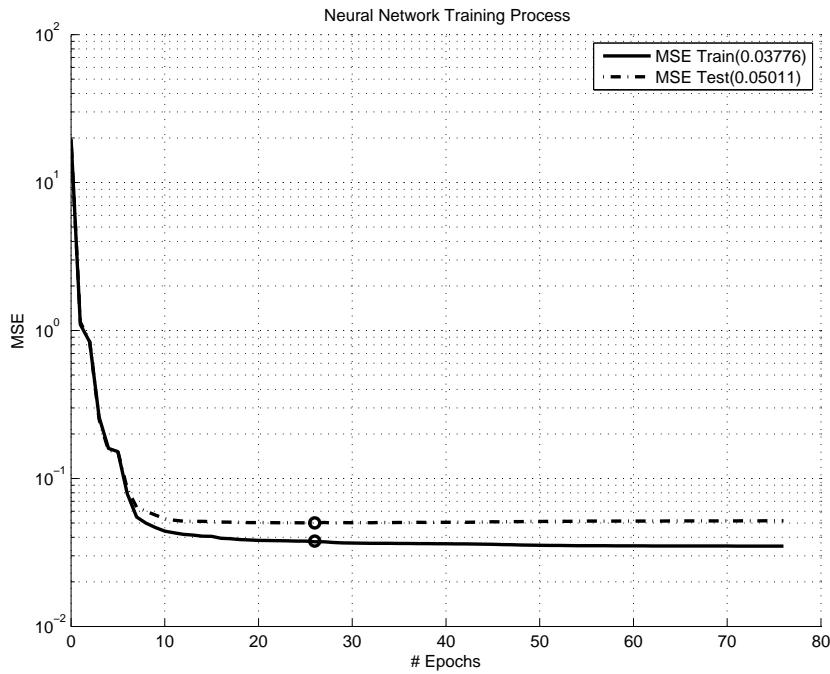
Os valores de massa reconstruídos para a massa com a correção neural e sem a correção neural coincidiram no valor de 117,50 GeV e este tem a variação de um *bin* (5 GeV), embora a incerteza associada a medida tenha diminuído de 17,55 (massa reconstruída sem a correção neural) para 10,18 (massa reconstruída com a correção neural).

Para os modelos gaussianos ajustados, os valores de massa foram 116,68, 109,91 e 116,04 GeV, para os modelos ajustados para a massa simulada, para a massa reconstruída sem o modelo neural, e para a massa reconstruída com o modelo neural. Sendo assim, temos 5,80% de variação do modelo de massa simulada para o modelo de massa reconstruída sem a correção neural. Já do modelo de massa simulada para o modelo de massa reconstruída com a correção neural, a variação é de 0,599%, o que mostra um melhor resultado da reconstrução com a correção neural com modelagem gaussiana.

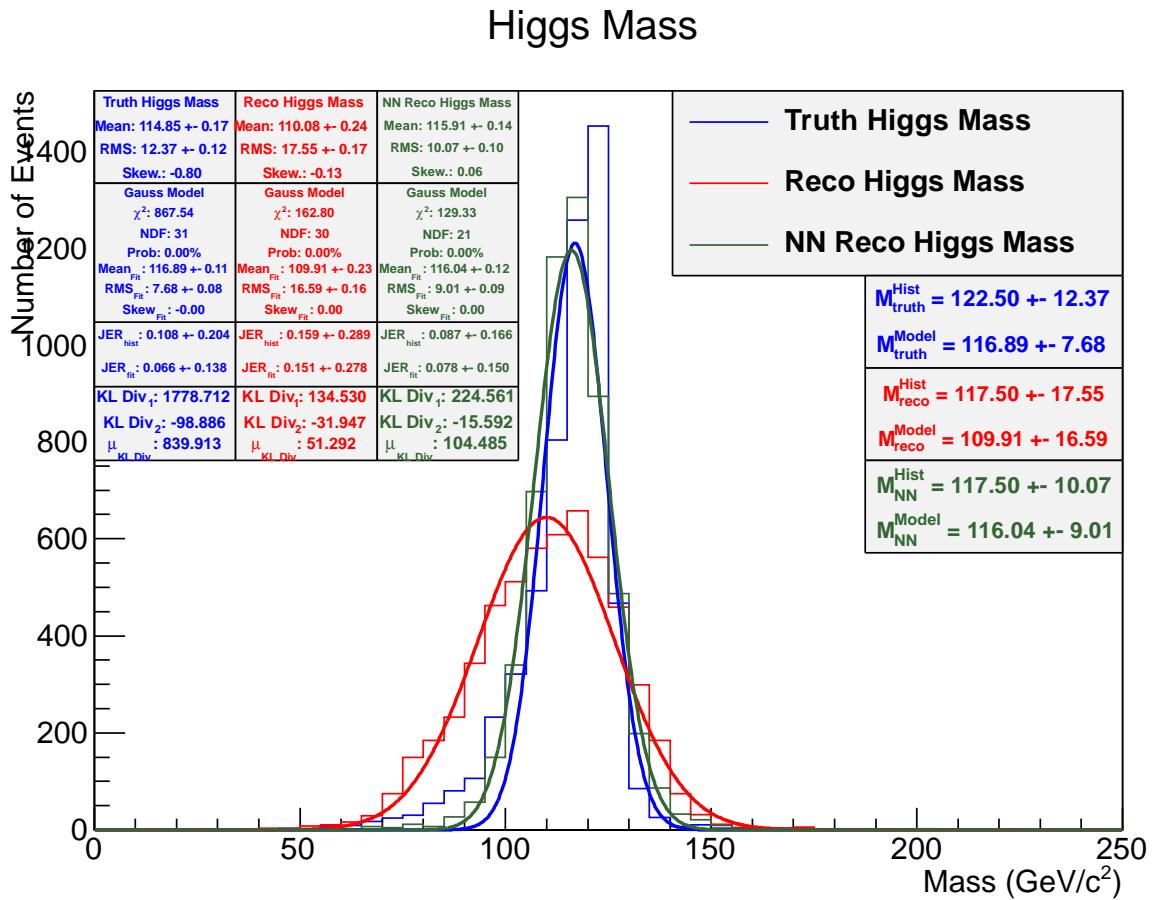
No que tangue a correlação entre a massa simulada e a massa reconstruída com a rede neural, temos que os valores do coeficiente angular se aproximou ainda mais do ideal, como pode ser visto na figura 5.14, obtendo um valor de 0,693. O coeficiente linear também se aproximou do valor ideal, que neste caso é 0, alcançando o valor de 34,57 GeV, contra os 70,659, atinguidos pela resconstrução sem correção neural.

5.7 Ajustes dos Modelos descritos

Os resultados apresentados anteriormente se baseam na modelagem gaussiana da distribuição de massa do bóson de Higgs. Como pode ser visto, esta modelagem não foi confirmada pelo teste do χ^2 , que rejeitou todos os modelos gaussianos propostos. Devido a este fato, foram encontrados na literatura, e apresentados no capítulo anterior, 3 novos modelos que podem ajustar a distribuição de massa do bóson de Higgs.



(a) Erros médios quadráticos de treinamento e de teste no durante o processo de treinamento



(b) Distribuições de Massas Simulada, reconstruída, reconstruída com a Rede Neural

Figura 5.13: Resultados para o conjunto de variáveis sugeridos pela análise de relevância com treinamento de segunda ordem

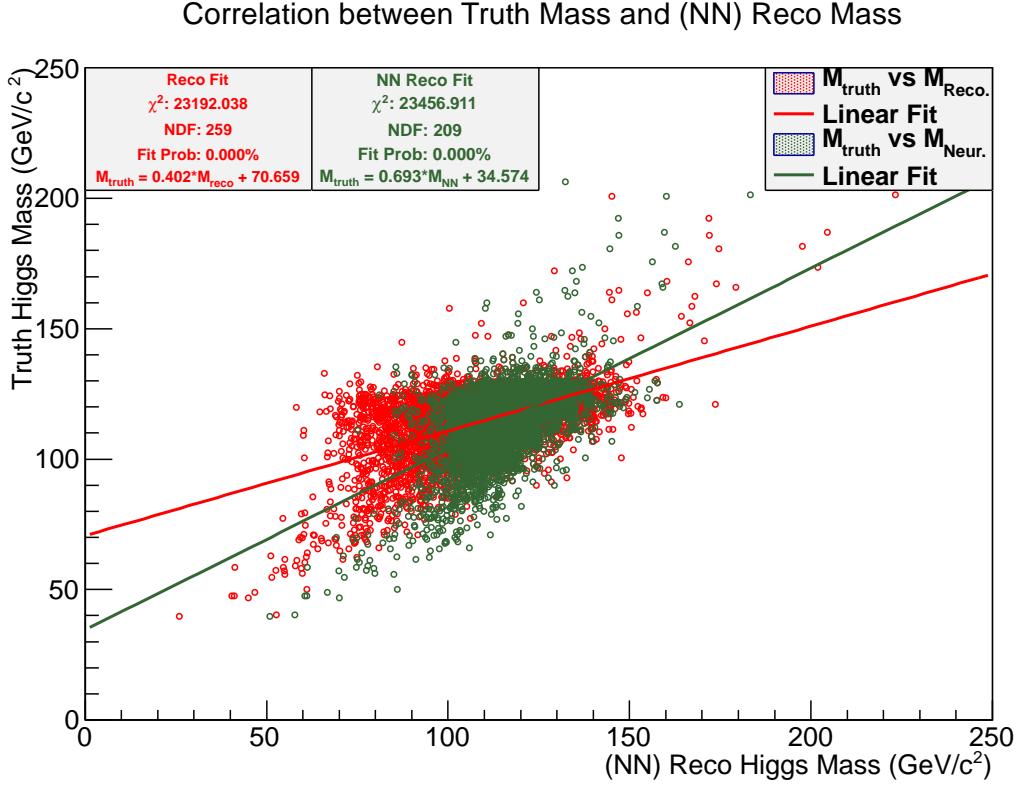


Figura 5.14: *Scatter plot* da massa simulada *versus* massa reconstruída com e sem a correção neural (treinamento de segunda ordem)

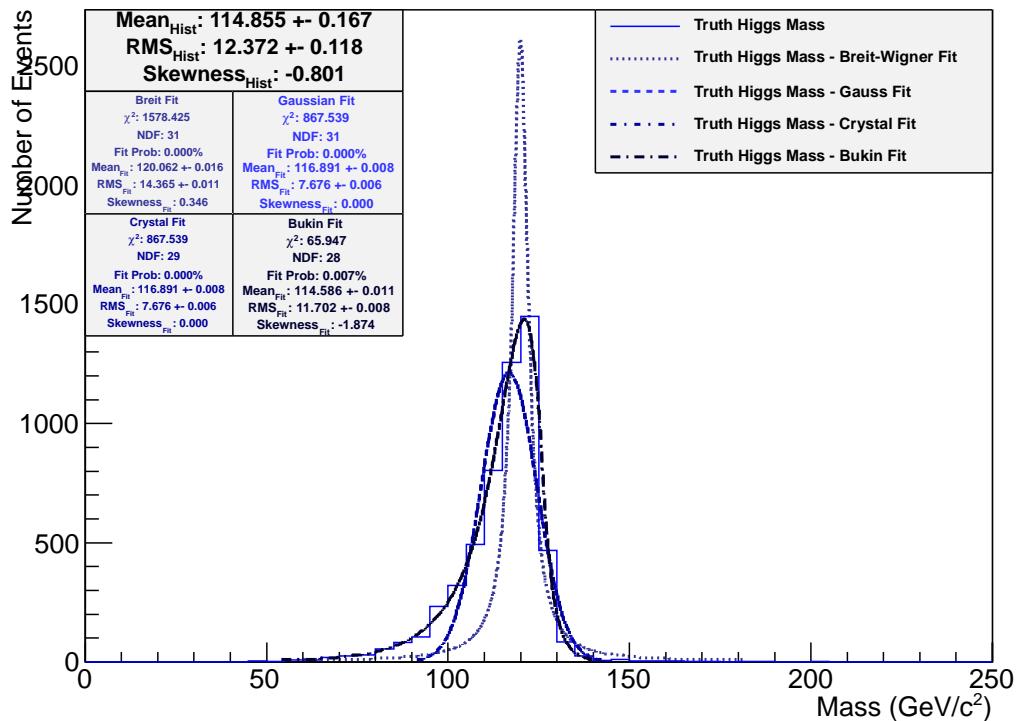
Esses modelos foram ajustados através do algoritmo MIGRAD [95] disponibilizado pelo ROOT. A modelagem foi fiel a cada um dos modelos apresentados resultando em 4 modelos ajustados: o gaussiano, o de Breit-Wigner, o da função Crystal Ball e o da função de Bukin. Esses ajustes se fazem necessários para o cálculo da massa do bóson de Higgs e também podem ser utilizados como uma validação da aderência da correção neural aos modelos físicos envolvidos no processo de decaimento.

Na figura 5.15(a) podemos ver os ajustes feitos na distribuição de massa do bóson de Higgs, a partir dos eventos simulados com massa de $125 \text{ GeV}/c^2$. O melhor ajuste (ajuste que apresentou a maior probabilidade de não-excludência do modelo) é a função de Bukin. Outra observação a ser feita é que o melhor resultado para a função Crystal Ball é quando a mesma assume valores para seus parâmetros que a tornam praticamente uma gaussiana.

Já na figura 5.15(b), temos os ajustes feitos na distribuição de massa para bósons de Higgs reconstruídos a partir dos dados registrados no detector. Mais uma vez, vemos que a função de Bukin obteve o melhor resultado para o teste do χ^2 e novamente, que o melhor resultado da função Crystal Ball é quando a mesma se torna uma gaussiana.

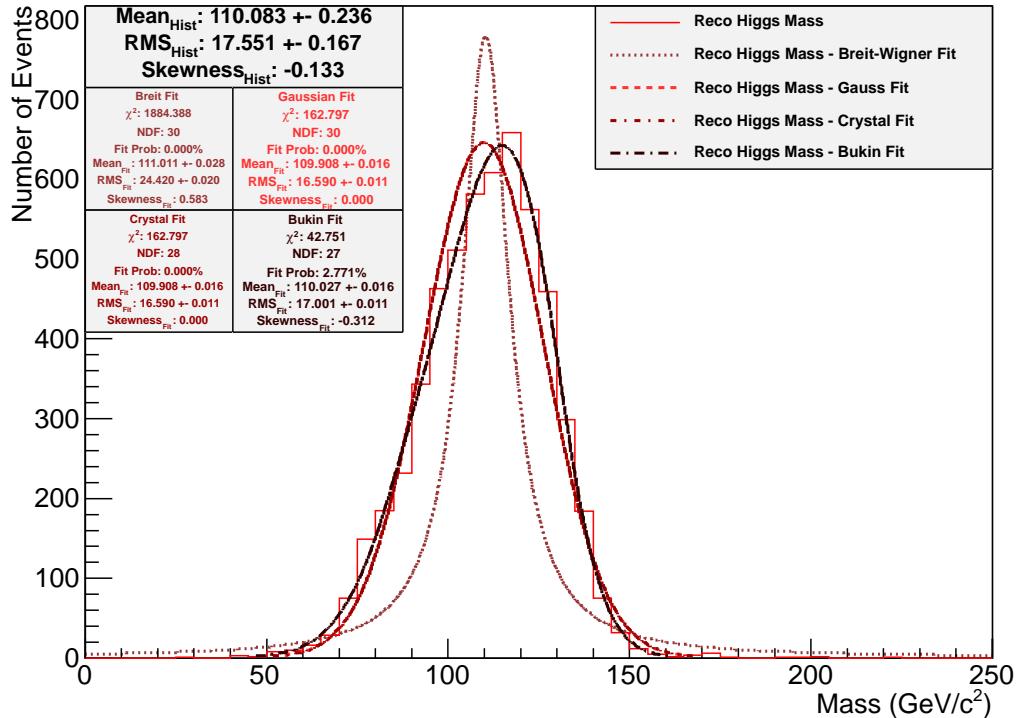
Na figura 5.16(a) apresentamos a comparação de todos os ajustes dos modelos

Truth Higgs Mass



(a) Ajuste de Modelos para Distribuição da Massa simulada do bóson de Higgs

Reco Higgs Mass



(b) Ajuste de Modelo para a Massa reconstruída sem a correção neural

Figura 5.15: Modelos ajustados as diferentes distribuições.

descritos no capítulo anterior para a distribuição da massa corrigida pelo quarto modelo neural descrito anteriormente e vemos mais uma vez que o melhor ajuste é o da função de Bukin. Novamente, observamos a similaridade da função Crystal Ball com uma gaussiana e vemos na figura 5.16(b) vemos as distribuições de massa com os melhores ajustes.

5.7.1 Resultados para Divergência de Kullback-Leibler

Na figura 5.17 podemos ver o resultado da divergência KL das distribuições para os seus respectivos ajustes para a função Bukin, função que apresentou a maior probabilidade de não-negação do modelo segundo o teste de χ^2 . Nesta figura, também vemos o resultado da divergência SKL (μKL).

Na figura 5.18 vemos o cálculo da divergência KL *bin a bin* da distribuição de massa simulada para distribuição de massa reconstruída e na figura 5.19 a divergência KL para a distribuição de massa simulada para distribuição de massa reconstruída com a correção neural baseada no quarto modelo neural.

Quando observamos a divergência KL mostrada na figura 5.18, vemos que o pico da divergência coincide com a maior diferença entre as distribuições e assume o valor de 600 *unidades de medida de D_{KL}* . Outro ponto que deve ser salientado é a inversão do sinal da divergência nas regiões de $50 < M < 110$ e $130 < M < 150$, onde a M_{reco} é maior que M_{truth} .

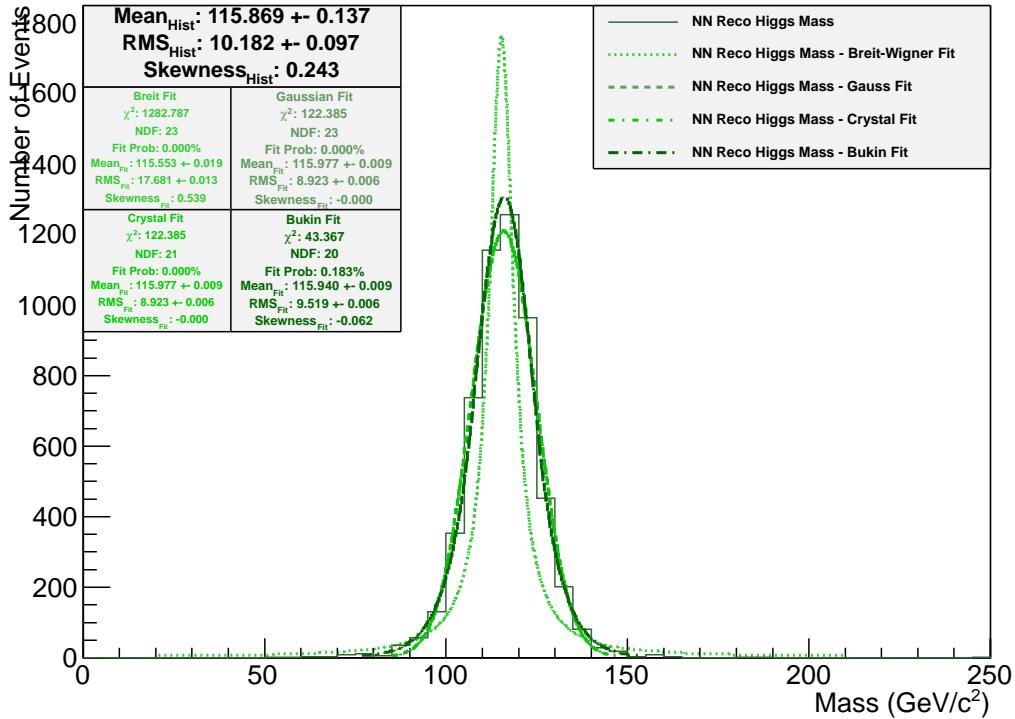
Observando a divergência KL mostrada na figura 5.19, podemos ver que o pico da divergência atinge um valor menor (no caso o valor de 250 *unidades de medida de D_{KL}*), embora, tenhamos inversões de sinais mais abruptas, pois a distribuição de massa reconstruída com a correção neural apresenta um pico deslocado do pico da distribuição de massa simulada em um *bin*.

-	M_{truth}	$M_{truth_{Gauss}}$	$M_{truth_{Breit}}$	$M_{truth_{Crystal}}$	$M_{truth_{Bukin}}$	M_{reco}	M_{neur}
M_{truth}	-	98.88	252.07	98.89	6.81	745.42	246.08
$M_{truth_{Gauss}}$	1778.71	-	2143.03	0.01	1518.76	4514.56	852.04
$M_{truth_{Breit}}$	1340.65	965.53	-	965.51	1348.20	2665.93	1526.07
$M_{truth_{Crystal}}$	1778.71	0.01	2143.02	-	1518.77	4514.56	852.02
$M_{truth_{Bukin}}$	131.79	111.97	193.36	111.98	-	980.80	349.69
M_{reco}	639.14	529.93	493.99	529.92	643.55	-	612.59
M_{neur}	211.15	243.35	166.93	243.36	183.44	945.60	-

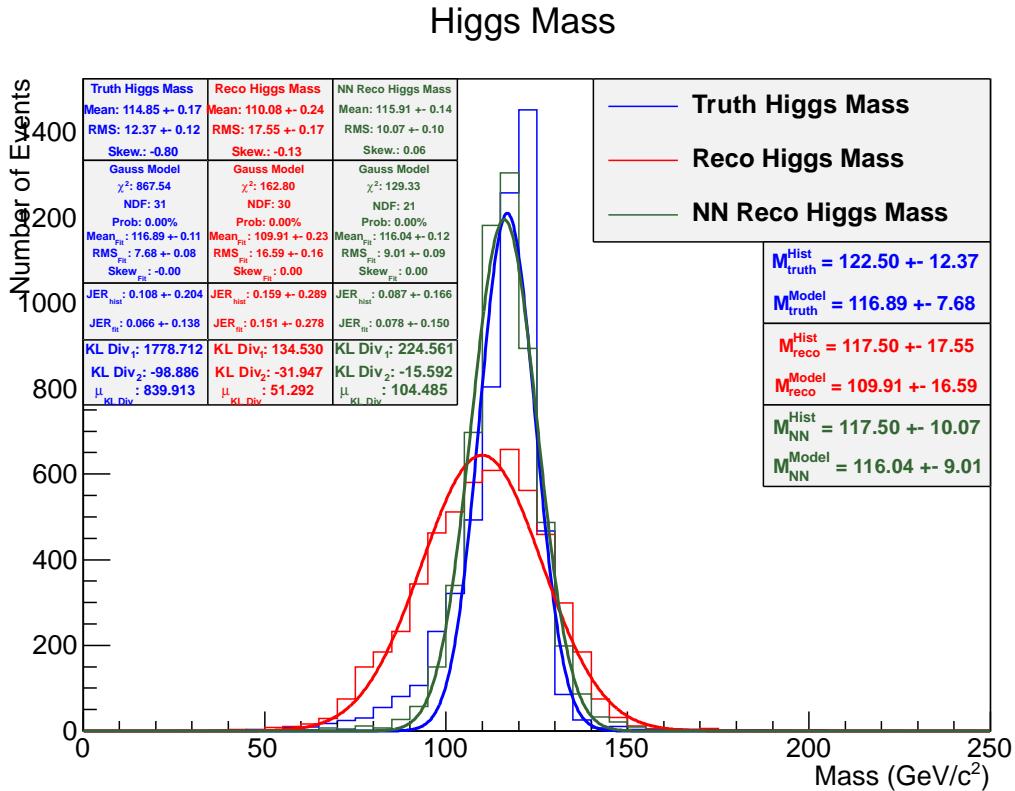
Tabela 5.2: Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições

Na tabela 5.2 podemos ver os valores da Divergência KL para a distribuição M_{truth} ou a massa simulada. O menor valor é $D_{KL}(M_{truth} \| M_{truth_{Bukin}}) = 6,81$ (Divergência KL de M_{truth} para $M_{truth_{Bukin}}$). No sentido contrário, ou seja,

NN Reco Higgs Mass



(a) Ajuste de Modelo para a Massa reconstruída com a correção neural



(b) Distribuições de Massas Simulada, Reconstruída, Reconstruída com a Rede Neural com o melhor ajuste de modelo para cada distribuição

Figura 5.16: Modelos ajustados as diferentes distribuições.

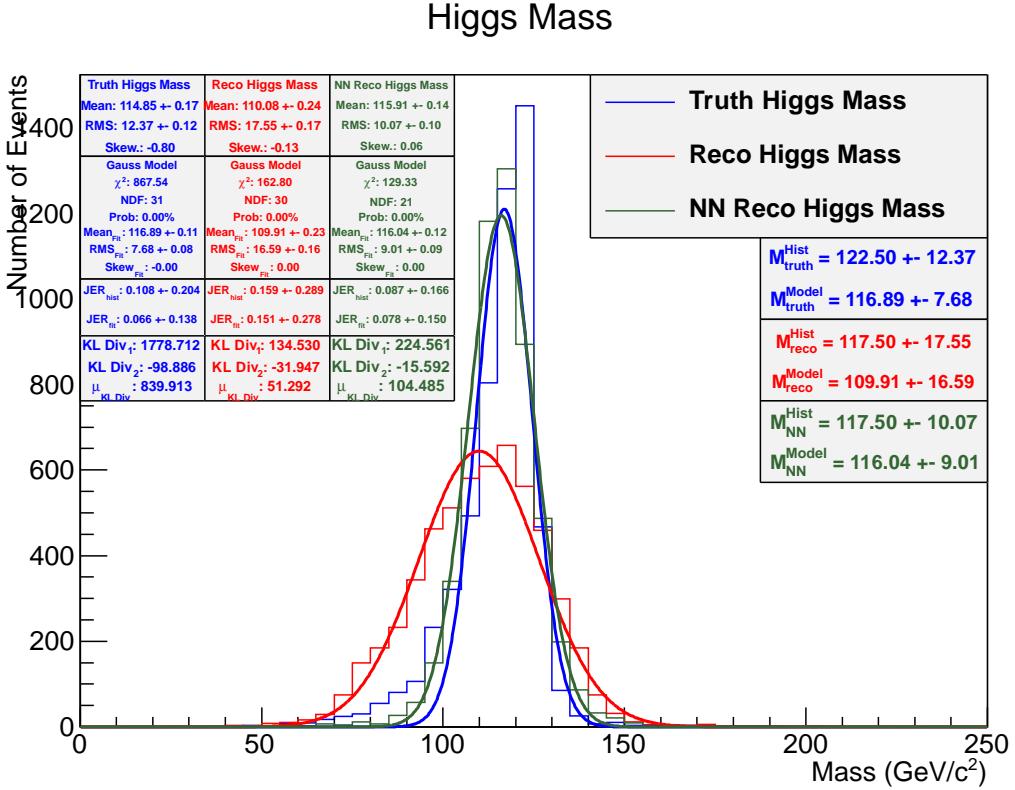


Figura 5.17: Distribuições de Massas Simulada, Reconstruída, Reconstruída com o Corretor neural e o melhor ajuste de modelo para cada distribuição com cálculo da divergência KL

$D_{KL}(M_{truth_{Bukin}} \| M_{truth})$ temos que a divergência possui valor de 131,8 e é o menor valor nessa coluna da tabela, ou seja, o menor valor no sentido contrário dessa divergência, e este resultado da divergência KL se mostra de acordo com o resultado do teste de χ^2 mostrado na sessão anterior.

O maior valor da coluna de divergências de M_{truth} é 1778,71 e é atingindo em dois modelos: no Crystal Ball e no Gaussiano. Ou seja, os modelos mais divergentes da distribuição de massa simulada são os modelos gaussiano e o modelo Crystall Ball.

Ainda na tabela 5.2, vemos que a distribuição de massa reconstruída com a correção neural (M_{neur}) é menos divergente que a massa reconstruída sem correção neural (M_{reco}), em ambos os sentidos, o que mostra que a correção neural, pelo menos na análise de D_{KL} , se mostra mais próxima da massa simulada.

Uma outra observação é que a divergência entre os modelos gaussiano e Crystal Ball é bastante pequena (0,01), o que mostra que o modelo Crystal Ball se aproxima muito do modelo gaussiano nesse ajuste.

Na tabela 5.3 podemos ver os valores de D_{KL} para M_{reco} comparada com outras funções e distribuições nos dois sentidos. Mais uma vez, à semelhança do que aconteceu com M_{truth} , o menor valor de divergência é $D_{KL}(M_{reco} \| M_{reco_{Bukin}})$, o que

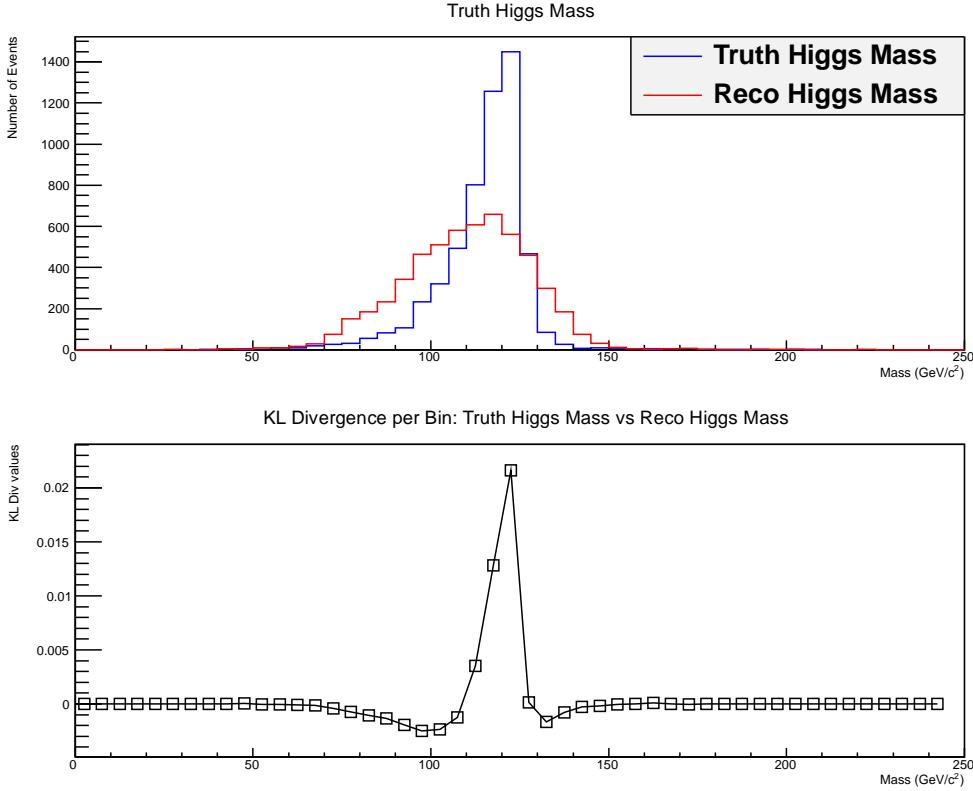


Figura 5.18: Divergência KL *bin a bin* entre as distribuições de massa simulada e massa reconstruída

comprova que o modelo de Bukin é o menos divergente.

Neste caso, os modelos Crystal Ball e Gaussiano são um pouco mais divergentes do que no caso de seus ajustes à distribuição de massa simulada. O comportamento da distribuição de massa reconstruída com a correção neural é bastante divergente do comportamento da distribuição da massa sem a correção.

Ainda na tabela 5.3 vemos que o modelo menos divergente neste caso é, mais uma vez, o modelo de Bukin, alcançando o valor de divergência de 5,69 no sentido da distribuição para o modelo e de 75,40 no sentido do modelo para distribuição. O modelo mais divergente da distribuição neste caso, foi o de Breit-Wigner, que atingiu o valor de 1380,65 no sentido da distribuição para o modelo e de 382,21 no sentido do modelo para a distribuição.

Outro ponto que deve ser ressaltado é que as $D_{KL}(M_{reco} \| M_{truth})$ e $D_{KL}(M_{reco} \| M_{neur})$ possuem a mesma ordem de grandeza (assumindo os valores 639,14 e 612,59, respectivamente). Sendo que a correspondência não se faz válida no sentido oposto da divergência, ou seja, para $D_{KL}(M_{truth} \| M_{reco})$ e $D_{KL}(M_{neur} \| M_{reco})$ não temos a mesma ordem de grandeza, pois os valores assumidos pelas divergências são 745,42 e 1083,03, respectivamente.

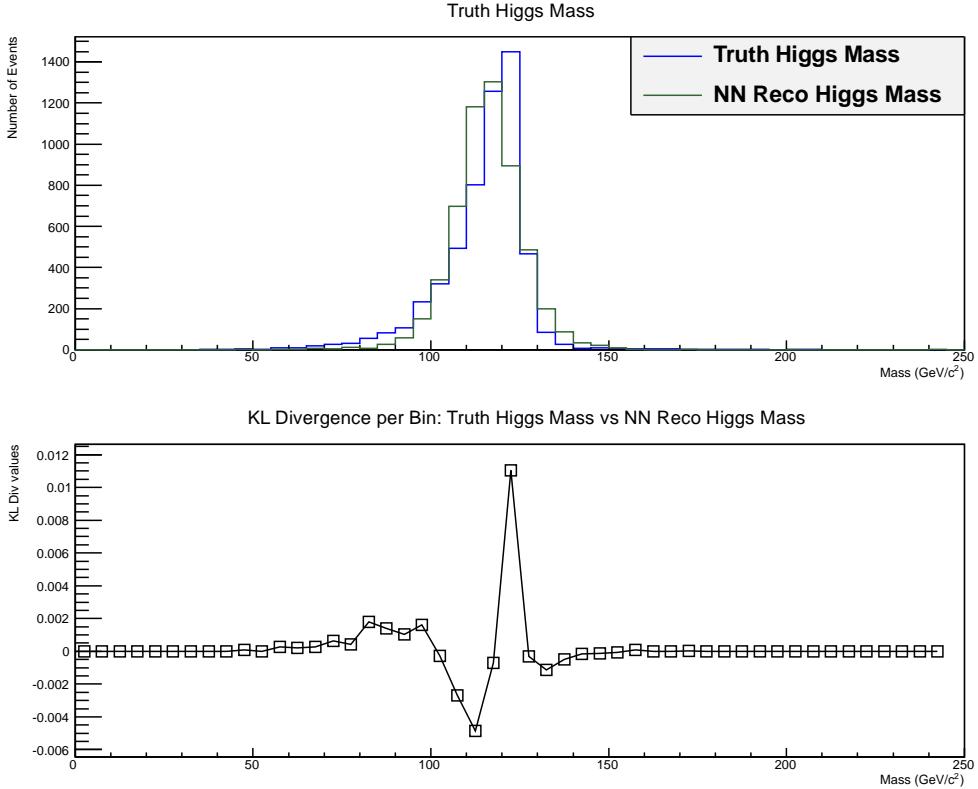


Figura 5.19: Divergência KL *bin a bin* entre as distribuições de massa simulada e massa reconstruída com a correção neural

5.7.2 Resultados para Divergência de Jensen-Shannon

Nas figuras 5.20 e 5.21 podemos ver os valores *bin a bin* da divergência JS da distribuição de massa simulada para massa reconstruída e massa simulada para massa reconstruída com a correção neural, respectivamente.

Como pode ser visto na figura 5.20,a divergência de JS, por ser normalizada, possui valores menores do que a divergência KL. Devido a este fato, o pico de ocorrência da divergência não se assentua tanto quanto o da divergência KL mostrado na figura 5.18, o que faz com que as variações ao redor desse pico seja mais acentuadas. Quando somamos todos os valores estimados na figura 5.20, obtemos o valor de divergência JS de 0,3678.

Já na figura 5.21 vemos que a divergência JS da distribuição de massa simulada para distribuição de massa reconstruída com a correção neural possui muitas oscilações de valores. Este fato pode ser explicado pois para o cálculo da divergência JS estimamos uma subdistribuição que é a média aritmética das duas distribuições a serem comparadas. Com isso, os valores dessa média ora são maiores do que os valores da distribuição, ora são menores, assim sendo, temos as oscilações apresentadas na figura 5.21. Com a soma de todos os valores de divergênciaJS apresentados na figura, temos um valor final de divergência JS de 0,1287.

-	M_{reco}	$M_{recoGauss}$	$M_{recoBreit}$	$M_{recoCrystal}$	$M_{recoBukin}$	M_{truth}	M_{neur}
M_{reco}	-	31.94	382.21	31.93	5.69	639.14	612.59
$M_{recoGauss}$	134.52	-	205.84	0.08	80.74	834.83	751.03
$M_{recoBreit}$	1380.65	1216.52	-	1216.53	1360.20	1847.02	1672.56
$M_{recoCrystal}$	134.52	0.08	205.83	-	80.73	834.83	751.03
$M_{recoBukin}$	75.40	23.25	418.55	23.24	-	693.43	631.17
M_{truth}	745.42	736.92	106.43	736.93	712.93	-	246.08
M_{neur}	1083.03	935.13	642.15	935.15	991.12	37.96	-

Tabela 5.3: Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições

-	M_{neur}	$M_{neurGauss}$	$M_{neurBreit}$	$M_{neurCrystal}$	$M_{neurBukin}$	M_{truth}	M_{reco}
M_{neur}	-	18.01	268.55	18.01	4.82	211.15	945.60
$M_{neurGauss}$	211.68	-	1945.74	0.04	96.73	808.99	2143.00
$M_{neurBreit}$	812.51	888.04	-	888.03	808.60	957.82	1519.23
$M_{neurCrystal}$	211.68	0.04	1945.73	-	96.73	808.99	2142.99
$M_{neurBukin}$	37.96	12.43	274.05	12.43	-	306.47	1083.03
M_{truth}	246.08	241.95	108.32	241.95	229.63	-	745.42
M_{reco}	612.59	604.03	248.86	604.03	616.07	639.14	-

Tabela 5.4: Divergência KL: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições

Como a divergência JS é uma divergência normalizada e os seus valores simetrizados, as tabelas 5.5, 5.6 e 5.7 se apresentam em um formato diferente das tabelas 5.2, 5.3 e 5.4, pois sua análise não tem necessidade de ser feita em um sentido e posteriormente no sentido reverso.

Na tabela 5.5 vemos que mais uma vez o modelo de Bukin apresentou o menor valor de divergência dentre os modelos ajustados, o que mostra que o modelo é o menos divergente dentro dos modelos ajustados. Outra observação que deve ser feita é que a distribuição de massa reconstruída com a correção neural é quase 3 vezes menos divergente da distribuição de massa simulada que a distribuição de massa reconstruída.

Na tabela 5.6 vemos que o modelo Bukin é o modelo menos divergente, atingindo o valor de 0,0088 na divergência JS, sendo que esse valor é aproximadamente 2,67 vezes menor do que o valor de divergência apresentado pelo segundo modelo

-	$M_{truthGauss}$	$M_{truthBreit}$	$M_{truthCrystal}$	$M_{truthBukin}$	M_{reco}	M_{neur}
M_{truth}	0.1921	0.2891	0.1921	0.0165	0.3678	0.1287

Tabela 5.5: Divergência JS: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições

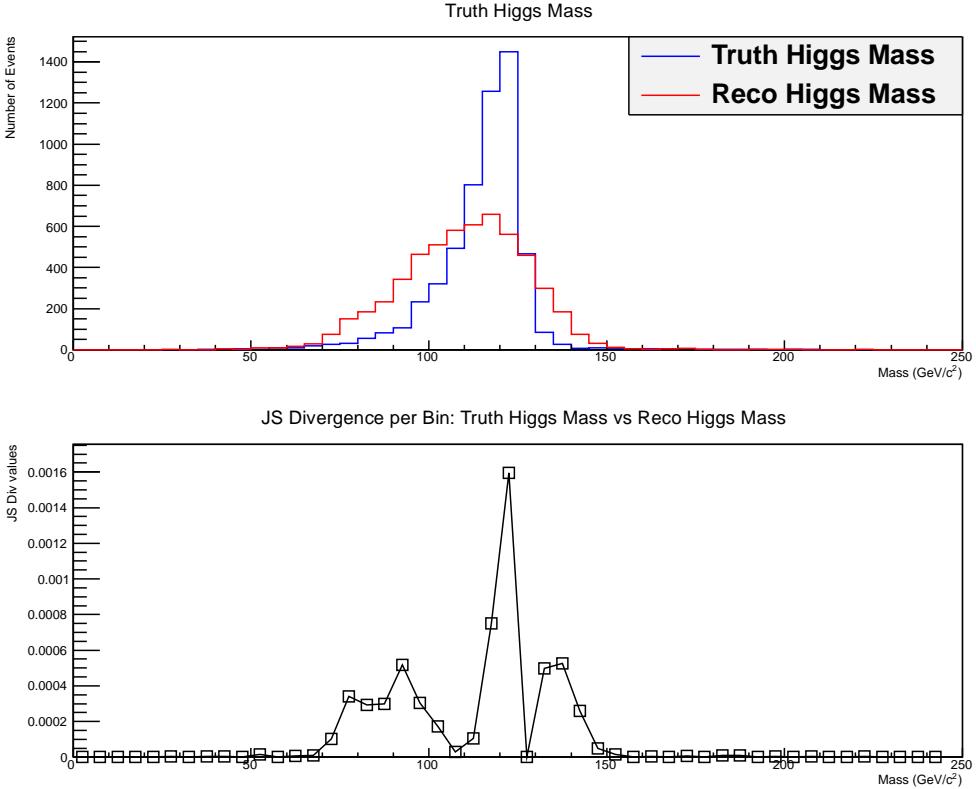


Figura 5.20: Divergência JS bin a bin entre as distribuições de massa simulada e massa reconstruída

-	$M_{recoGauss}$	$M_{recoBreit}$	$M_{recoCrystal}$	$M_{recoBukin}$	M_{truth}	M_{neur}
M_{reco}	0.0232	0.2953	0.0232	0.0088	0.3678	0.3847

Tabela 5.6: Divergência JS: Comparaçāo entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições

menos divergente, que neste caso são os modelos gaussianos e de Crystal Ball (que obtiveram os mesmos valores de divergência).

-	$M_{neurGauss}$	$M_{neurBreit}$	$M_{neurCrystal}$	$M_{neurBukin}$	M_{truth}	M_{reco}
M_{neur}	0.0271	0.1837	0.0271	0.0067	0.1287	0.3847

Tabela 5.7: Divergência JS: Comparaçāo entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições

O modelo de Bukin, mais uma vez, foi confirmado pela divergência JS como sendo o melhor modelo ajustado para a distribuição neural, uma vez que este modelo obteve valor de divergência JS três vezes menor do que o segundo modelo menos divergente, como pode ser visto na tabela 5.7. E mais uma vez, os modelos gaussiano e de Crystal Ball foram igualmente divergentes, mostrando que o ajuste do modelo de Crystal Ball é bastante próximo do modelo gaussiano (fato que também pode ser

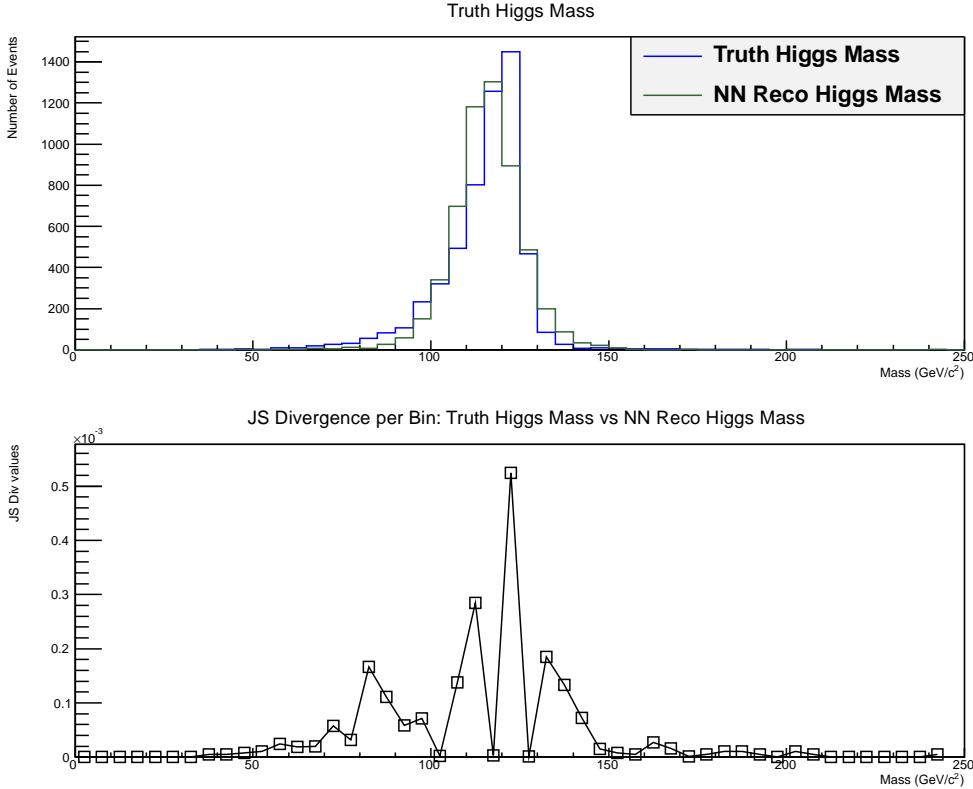


Figura 5.21: Divergência JS bin a bin entre as distribuições de massa simulada e massa reconstruída com a correção neural

visto nas tabelas 5.6 e 5.5).

5.7.3 Resultados para teste de Kolmogorov-Smirnov

Os resultados numéricos obtidos com os testes de Kolmogorov-Smirnov (KS) encontram-se nas tabelas 5.8, 5.9 e 5.10. Nestas tabelas, os valores são sempre expressos com 4 algarismos significativos, salvo o caso em que o valor do teste foi exatamente igual a zero. Para o T_{KS} , ou o valor do teste de Kolmogorov-Smirnov, quanto mais próximo do valor unitário, mais similares serão as distribuições.

-	$M_{truth_{Gauss}}$	$M_{truth_{Breit}}$	$M_{truth_{Crystal}}$	$M_{truth_{Bukin}}$	M_{reco}	M_{neur}
M_{truth}	0.0000	0	0.0000	0.8452	0	0.0000

Tabela 5.8: Teste KS: Comparação entre os ajustes de modelos para a distribuição de massa simulada e outras distribuições

Na tabela 5.8 encontram-se os valores dos testes de KS realizados tendo como referência a distribuição de massa simulada. É observado que o teste KS mostra uma proximidade do modelo de Bukin ajustado a esta distribuição (o resultado do teste foi de 0,8452) enquanto para os outros modelos os valores do teste KS se encontram

abaixo de 10^{-3} . Já para as distribuições M_{reco} e M_{neur} , o teste também apresentou valores menores que 10^{-3} .

-	$M_{recoGauss}$	$M_{recoBreit}$	$M_{recoCrystal}$	$M_{recoBukin}$	M_{truth}	M_{neur}
M_{reco}	0.0007	0.0000	0.0007	1.0000	0	0

Tabela 5.9: Teste KS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída e outras distribuições

Os resultados do teste KS tendo como referência a distribuição de massa reconstruída sem a correção neural podem ser vistos na tabela 5.9. Aqui, podemos ver que o teste KS mostra total compatibilidade do modelo Bukin ajustado a essa distribuição (o resultado do teste KS foi 1,000) e também vemos que esta distribuição é bastante diferente, no que tange o teste KS, das distribuições M_{truth} e M_{neur} , uma vez que o teste foi exatamente igual a zero.

-	$M_{neurGauss}$	$M_{neurBreit}$	$M_{neurCrystal}$	$M_{neurBukin}$	M_{truth}	M_{reco}
M_{neur}	0.6860	0.0000	0.6860	1.0000	0.0000	0

Tabela 5.10: Teste de KS: Comparação entre os ajustes de modelos para a distribuição de massa reconstruída com a correção neural e outras distribuições

Já na tabela 5.10 vemos os valores obtidos pelo teste KS a partir da distribuição M_{neur} . Podemos notar que os modelos gaussiano e Crystall Ball mais uma vez apresentaram valores similares, sendo desta vez, o valor de 0,6860. O modelo de Bukin foi o modelo com o melhor resultado para este teste, apresentando o valor de 1,0000, o que o torna o melhor modelo ajustado a esta distribuição.

5.8 Discussão dos Resultados Obtidos

Nesta sessão serão abordados, de maneira suscinta todos os resultados descritos neste capítulo.

Na tabel 5.11 temos os resultados dos quatro modelos neurais apresentados anteriormente. No que tange a reconstrução da massa do bóson de Higgs, o primeiro modelo (com as variáveis sugeridas pela colaboração ATLAS) obteve a melhor resposta, com o valor de pico da distribuição da massa reconstruída com a correção neural exatamente igual ao valor de pico da distribuição simulada, embora o erro associado à medida seja um pouco maior no caso da massa reconstruída.

Os outros modelos neurais obtiveram valores médios com uma diferença de 5 GeV, que é o valor de um *bin* no histograma que respresenta a distribuição de massa. Ainda no que concerne a reconstrução de massa, podemos ver que, geralmente, os treinamentos de segunda ordem obtiveram erros associados às medidas menores.

Com relação a modelagem, como dito anteriormente, nenhum dos modelos gaussianos que foram ajustados passou no teste do χ^2 . Mas podemos ver que o segundo modelo neural obteve um resultado mais próximo da média do modelo gaussiano ajustado aos dados simulados, mesmo que esse modelo tenha um espalhamento maior em relação a sua média.

Quando analizamos os valores do *Jet Energy Resolution*, vemos que o terceiro modelo neural obteve uma resposta aproximadamente igual ao *JER* da distribuição dos dados simulados, sendo que a variação do valor de *JER* da distribuição de massa simulada para o valor de *JER* da distribuição de massa reconstruída com o terceiro modelo neural é de 2,78%.

Na análise de valores de *JER* do modelo neural ajustado a cada uma das distribuições, vemos que os valores mais próximos aos dados simulados é o quarto modelo neural, que apresenta uma variação de 1,67% em relação ao modelo ajustado aos dados simulados.

-	M_{Higgs} Distribuição	M_{Higgs} Modelo (Gauss)	<i>JER</i> Distribuição	<i>JER</i> Modelo (Gauss)
Dados Simulados	122.50 ± 12.37	$116,89 \pm 7.68$	$0,108 \pm 0.024$	$0,066 \pm 0,013$
Dados Reconstruídos s/ mod. neural	$117,50 \pm 17,55$	$109,91 \pm 16,59$	$0,159 \pm 0,028$	$0,151 \pm 0.027$
Dados Reconstruídos c/ 1º mod.	$122.50 \pm 15,91$	$117,46 \pm 14,99$	$0,136 \pm 0,026$	$0,128 \pm 0,025$
Dados Reconstruídos c/ 2º mod.	$117,50 \pm 15,58$	$117,12 \pm 14,46$	$0,133 \pm 0,025$	$0,123 \pm 0,024$
Dados Reconstruídos c/ 3º mod.	$117,50 \pm 12.20$	$116,47 \pm 10,97$	$0,105 \pm 0,020$	$0,094 \pm 0,018$
Dados Reconstruídos c/ 4º mod.	$117,50 \pm 10,18$	$116,04 \pm 9,01$	$0,088 \pm 0.016$	$0.077 \pm 0,014$

Tabela 5.11: Comparaçāo dos resultados obtidos com diferentes conjuntos de entradas para a modelagem gaussiana

Como dito anteriormente, como nenhum dos modelos gaussianos foi aceito pelo teste do χ^2 , foram então avaliados outros testes estatísticos e outros modelos. Na tabela 5.12 temos os melhores resultados obtidos em cada um dos testes estatísticos de ajustes dos modelos descritos neste trabalho. Os testes foram realizados para o quarto modelo neural descrito.

	Melhor Ajuste			M_{Higgs}^{Modelo}	JER^{Modelo}
M_{truth}	χ^2	Bukin	$prob = 0,007\%$	$121,04 \pm 11,70$	$0,102 \pm 0,019$
	D_{KL}	Bukin	$D_{KL} = 6,81 \ 131,79$		
	D_{JS}	Bukin	$D_{JS} = 0,0165$		
	T_{KS}	Bukin	$T_{KS} = 0,8452$		
M_{reco}	χ^2	Bukin	$prob = 2,771\%$	$114,97 \pm 17,00$	$0,155 \pm 0,028$
	D_{KL}	Bukin	$D_{KL} = 5,69 \ 75,40$		
	D_{JS}	Bukin	$D_{JS} = 0,088$		
	T_{KS}	Bukin	$T_{KS} = 1,0000$		
M_{neur}	χ^2	Bukin	$prob = 0,183\%$	$116,12 \pm 9,52$	$0,082 \pm 0,0015$
	D_{KL}	Bukin	$D_{KL} = 4,82 \ 37,96$		
	D_{JS}	Bukin	$D_{JS} = 0,067$		
	T_{KS}	Bukin	$T_{KS} = 1,000$		

Tabela 5.12: Comparaçāo e escolha do melhor modelo em diferentes análises estatísticas

Em todos os testes estatísticos realizados em todas as distribuições, o modelo que foi aprovado foi o modelo de Bukin, que provou agregar corretamente os efeitos da convolução gaussiana (erros do detector) a Breit-Wigner (comportamento do decaimento do bóson de Higgs).

Também podemos ver nesta tabela que o valor da massa reconstruída com a correção neural e modelada por uma função Bukin se aproxima mais do valor de massa simulada e modela por outra função Bukin do que o valor de massa reconstruída sem a correção neural.

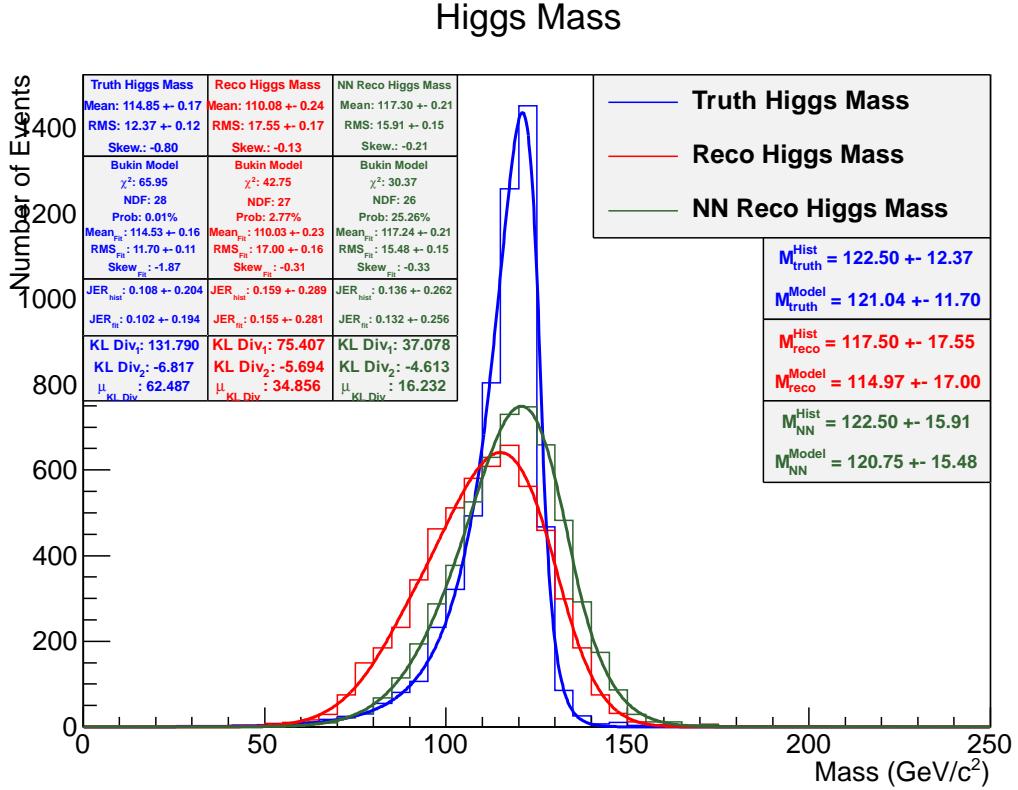
Ainda na tabela 5.12, com relação ao valor de JER modelado, vemos que o valor não só se aproximou do valor modelado para a simulação, mas também foi reduzido, o que mostra um baixo espalhamento da medida em torno da sua média.

Nas figuras 5.22(a), 5.22(b), 5.23(a) e 5.23(b) e na tabela 5.13 temos os resultados de todos os modelos neurais ajustados com o melhor modelo selecionado pelo teste do χ^2 .

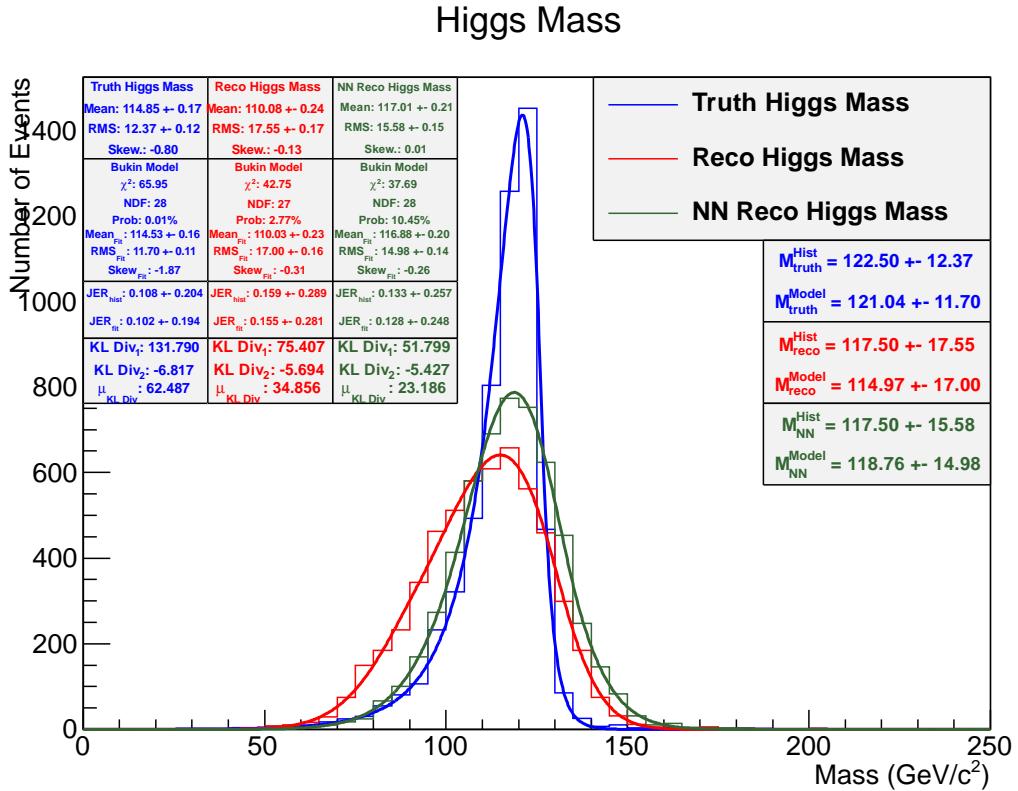
Como podemos observar mais uma vez, de maneira geral, a medida que adicionamos informações aos modelos neurais, o modelo de Bukin se agraga às distribuições de maneira mais eficiente, o que reduz o erro associado à medida de massa. Esse comportamento apresenta uma excessão: o terceiro modelo neural.

A modelagem de Bukin, apresentada na tabela 5.13, apresenta um desvio médio em relação à média (média dos valores de desvio em relação a média do modelo ajustado em relação à média da distribuição, que nesse caso é de 1,475%, e de 2,935%), mais uma vez mostrando que o modelo de Bukin ajusta, na média, as distribuições em questão com mais qualidade.

Ainda na tabela 5.13 vemos que o desvio máximo em relação ao valor de JER^{Modelo} em relação a JER é de 10,47% (calculado no terceiro modelo neural)

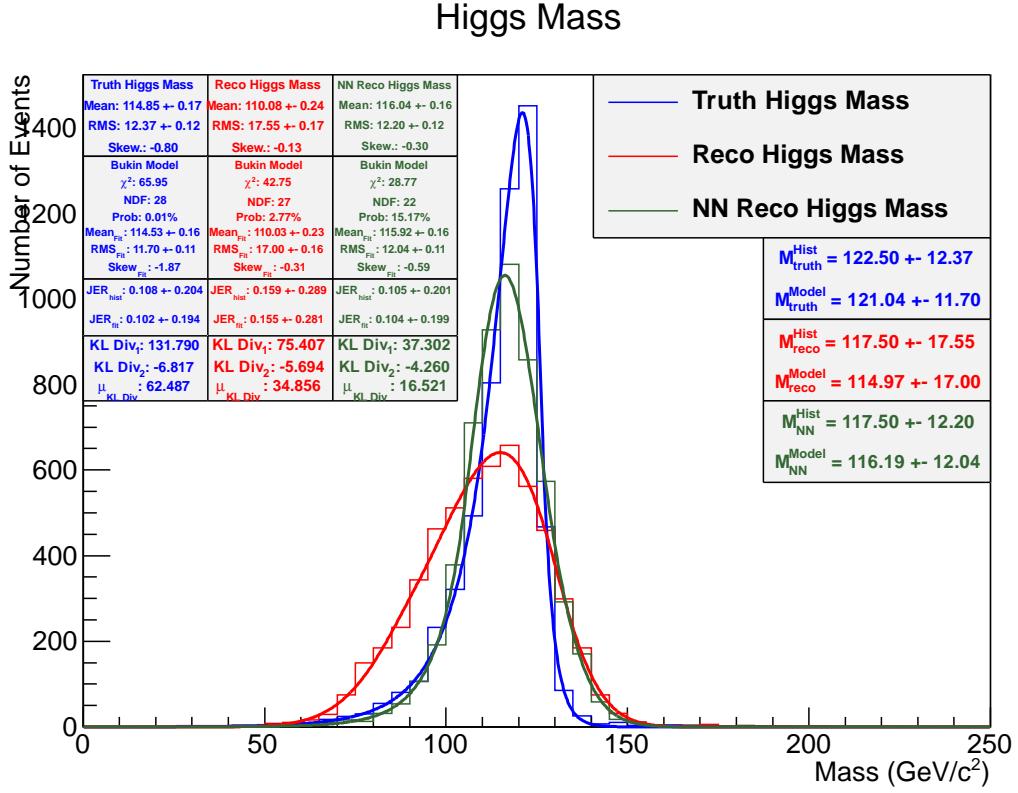


(a) Resultado obtido com o melhor ajuste dentre dos modelos propostos para o primeiro modelo neural.

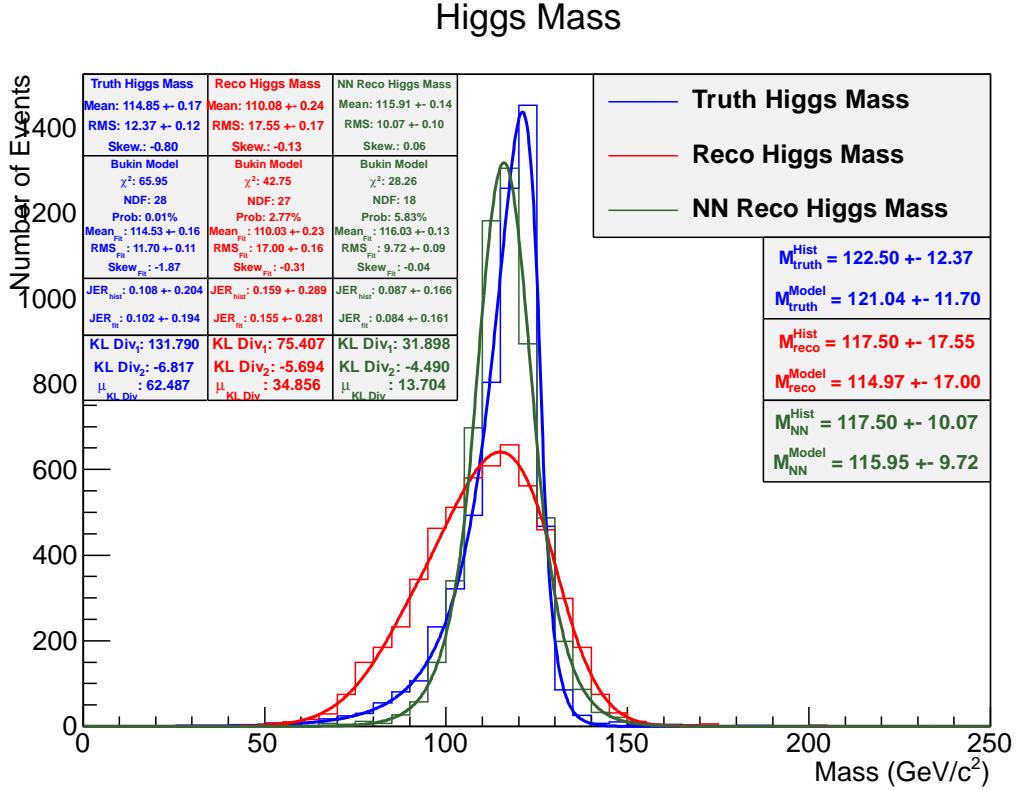


(b) Resultado obtido com o melhor ajuste dentre dos modelos propostos para o segundo modelo neural.

Figura 5.22: Resultados obtidos com o melhor ajuste dentre dos modelos propostos para o primeiro e segundo modelos neurais.



(a) Resultado obtido com o melhor ajuste dentre dos modelos propostos para o terceiro modelo neural.



(b) Resultado obtido com o melhor ajuste dentre dos modelos propostos para o quarto modelo neural.

Figura 5.23: Resultados obtidos com o melhor ajuste dentre dos modelos propostos para o terceiro e o quarto modelos neurais.

-	M_{Higgs} Distribuição	M_{Higgs} Melhor Modelo	JER Distribuição	JER Melhor Modelo
Dados Simulados	122.50 ± 12.37	$121,04 \pm 11,70$	$0,108 \pm 0.024$	$0,102 \pm 0,019$
Dados Reconstruídos s/ mod. neural	$117,50 \pm 17,55$	$114,97 \pm 17,00$	$0,159 \pm 0,028$	$0,155 \pm 0.028$
Dados Reconstruídos c/ 1º mod.	$122.50 \pm 15,91$	$120,75 \pm 15,48$	$0,136 \pm 0,026$	$0,132 \pm 0,025$
Dados Reconstruídos c/ 2º mod.	$117,50 \pm 15,58$	$118,76 \pm 14,98$	$0,133 \pm 0,025$	$0,128 \pm 0,024$
Dados Reconstruídos c/ 3º mod.	$117,50 \pm 12,20$	$114,97 \pm 17,00$	$0,105 \pm 0,020$	$0,094 \pm 0,018$
Dados Reconstruídos c/ 4º mod.	$117,50 \pm 10,07$	$115,95 \pm 9,72$	$0,087 \pm 0.016$	$0.084 \pm 0,016$

Tabela 5.13: Comparação dos resultados obtidos com diferentes conjuntos de entradas para a melhor modelagem dentro dos quatro modelos testados.

enquanto o maior desvio de JER^{Modelo} em relação ao valor de JER da distribuição (quando esta é modelada por uma gaussiana) é de 38,88% e é calculado entre distribuição e modelagem da massa simulada.

Capítulo 6

Conclusões e Trabalhos Futuros

A resposta do detector ATLAS pode ser melhorada com a utilização da técnica de redes neurais artificiais. Além do modelo mais simples já apresentar uma melhora considerável na reconstrução de massa do bóson de Higgs, temos que o erro associado à medida tende a diminuir quando inserimos mais variáveis.

Essa dissertação de mestrado mostrou o uso de diferentes técnicas para a avaliação estatística de um modelo ajustado. As divergências KL e JS foram aplicadas a diferentes distribuições tendo os seus resultados analisados e com total concordância com outros testes estatísticos consolidados na colaboração ATLAS.

Os treinamentos de primeira e de segunda ordem foram avaliados em diversos critérios, sendo que cada treinamento um obteve um melhor resultado para um determinado critério. Em todos os modelos apresentados, o jato 1 obteve bons valores de reconstrução, enquanto que o jato dois, obteve melhores valores para o primeiro e terceiro modelo. Este fato justifica os melhores valores de reconstrução de massa apresentados no primeiro e terceiro modelo, motivando assim a correção neural feita jato a jato.

O primeiro modelo neural (modelo com as variáveis de entrada sugeridas pela colaboração ATLAS, com treinamento de primeira ordem) obteve um bom resultado no que tange a reconstrução de massa da distribuição de massa do bóson de Higgs atingindo o valor de massa simulado, $125 \text{ GeV}/c^2$, embora o seu erro associado à medida tenha sido maior do que o valor obtido para a massa simulada. Quando analisamos o resultado da modelagem gaussiana, vemos que o modelo gaussiano ajustado à massa reconstruída com a correção neural tem o resultado melhor do que o modelo gaussiano ajustado a massa reconstruída sem a rede neural, tanto no valor da massa (117,46 GeV contra 109,91 GeV) como no valor do erro associado a medida (14,99 GeV contra 16,59 GeV). O mesmo pode ser observado nos valores de JER da distribuição e de JER do modelo gaussiano ajustado às respectivas massas.

O segundo modelo neural (modelo com as variáveis de entrada sugeridas pela colaboração ATLAS, com treinamento de segunda ordem) visava diminuir o erro

associado à medida, o que de fato ocorreu (o erro associado teve uma variação de 2,07%, com relação ao valor de erro do primeiro modelo) mas ao custo da inserção de um erro médio de 5 GeV ou um *bin* na distribuição. Mesmo com esse erro médio de 5 GeV, a modelagem gaussiana se mostra mais similar ao modelo gaussiano ajustado à massa simulada do que o modelo gaussiano ajustado à massa reconstruída com o primeiro modelo de correção.

De posse das demais entradas, uma análise de relevância foi feita. Um ponto a ser ressaltado foi a análise de relevância que trouxe como contribuição uma melhor caracterização da importância das variáveis de entrada de um corretor baseado em redes neurais. As duas análises de relevância usadas são mais comumente aplicadas a classificadores baseados em redes neurais, mas com os resultados demonstrados nessa dissertação, temos a motivação para a aplicação destas análises em estimadores e corretores.

O terceiro modelo neural (modelo com todas as variáveis utilizadas na análise de relevância, com treinamento de primeira ordem) apresentou, assim como o segundo modelo, o erro médio na distribuição de massa reconstruída de 5 GeV, embora, este modelo tenha apresentado a menor distância do modelo gaussiano para o modelo ajustado à massa simulada. E seu valor de *JER* da distribuição foi o mais próximo ao valor do *JER* da distribuição da massa simulada.

Por fim, o quarto modelo neural (modelo com todas as variáveis utilizadas na análise de relevância, com treinamento de segunda ordem) apresentou os melhores resultados dentre os outros modelos neurais no que concerne o valor de *JER* modelado por uma gaussiana.

Podemos com isso concluir que os modelos neurais, dentro de suas características, melhoraram a performance do detector no que tange a massa do bóson de Higgs decaindo em dois jatos *b*.

6.1 Trabalhos Futuros

Podemos apontar como trabalhos futuros a continuação dos testes com os diferentes treinamentos que não utilizem derivadas como funções de atualização, como por exemplo uma função de atualização de pesos por enxames.

Outro trabalho futuro é a utilização de um modelo neural que corrija cada um dos jatos separadamente. A correção dos dois jatos em um único corretor pode ser prejudicada pois as informações de um jato pode influenciar negativamente a correção do segundo, portanto, o processo de isolamento dos jatos pode trazer benefícios para os corretores.

Outro trabalho futuro é a criação de uma nova função de ajuste para as distribuições de massa do ATLAS. Entre os modelos apresentados no capítulo 4 e

testados no capítulo 5, temos o modelo de Bukin (modelo que obteve os melhores ajustes), que foi criado com uma aproximação da convolução de uma gaussiana com uma função exponencial (do tipo erf). Idealmente, o decaimento de partículas segue um comportamento Breit-Wigner, sendo que o detector tem resposta gaussiana, ou seja, o comportamento do decaimento de uma partícula em um detector tenderia a se comportar como uma Breit-Wigner convoluída com uma gaussiana.

Referências Bibliográficas

- [1] “CERN Document Server”, <https://cdsweb.cern.ch>, 2014, (acessado em Março de 2014).
- [2] “CERN - European Organization for Nuclear Research”, <http://www.cern.ch>, acessado em outubro de 2009.
- [3] HEUER, R.-D., “The future of the Large Hadron Collider and CERN”, *Phil. Trans. Roy. Soc. Lond.*, v. A370, pp. 986–994, 2012.
- [4] NE’EMAN, Y., KIRSH, Y., *Particle Hunters*. 2 ed. Cambridge University Press, 1996.
- [5] WEISE, W., “Yukawa’s Pion, Low-Energy QCD and Nuclear Chiral Dynamics”, *Progress of Theoretical Physics Supplement*, v. 170, pp. 161–184, 2007.
- [6] “Processes Involving Charged Mesons”, *Nature*, v. 159, pp. 694–697, 1947.
- [7] AAD, G., OTHERS, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys.Lett.*, v. B716, pp. 1–29, 2012.
- [8] CHATRCHYAN, S., OTHERS, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys.Lett.*, v. B716, pp. 30–61, 2012.
- [9] DENNER, A., HEINEMEYER, S., PULJAK, I., *et al.*, “Standard Model Higgs-Boson Branching Ratios with Uncertainties”, *Eur.Phys.J.*, v. C71, pp. 1753, 2011.
- [10] “LHC - The Large Hadron Collider”, <http://lhc.web.cern.ch>, acessado em agosto de 2013.
- [11] SILVA, P. V. M. D., *Processamento de Sinais para Calibração, Teste e Estimação de Energia com um Calorímetro Cintilante de Alta Velocidade*. Tese de doutorado, Instituto Alberto Luiz Coimbra, COPPE, Universidade Federal do Rio de Janeiro, julho 2007.

- [12] “ATLAS - A Toroidal Lhc AparatuS”, <http://atlas.web.cern.ch>, acessado em agosto de 2013.
- [13] AAD, G., OTHERS, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST*, v. 3, pp. S08003, 2008.
- [14] CERN, A., *ATLAS: Detector and physics performance technical design report. Volume 1*, Report CERN-LHCC-99-14, ATLAS-TDR-14, ATLAS Collaboration, 1999.
- [15] CERN, A., *ATLAS: Detector and physics performance technical design report. Volume 2*, Report CERN-LHCC-99-15, ATLAS-TDR-15, ATLAS Collaboration, 1999.
- [16] “CMS - Compact Muon Solenoid”, <http://cms.web.cern.ch>, acessado em agosto de 2013.
- [17] “LHCb - Large Hadron Collider beauty”, <http://lhcb.web.cern.ch/>, acessado em agosto de 2013.
- [18] “ALICE - A Large Ion Collider Experiment”, <http://aliceinfo.cern.ch>, acessado em agosto de 2013.
- [19] “ATLAS Inner Detector”, <https://twiki.cern.ch/twiki/bin/view/Atlas/InnerDetector>, acessado em agosto de 2013.
- [20] ALAM, M., OTHERS, *ATLAS pixel detector: Technical design report*, Report, ATLAS Collaboration, 1998.
- [21] “Calorímetro Eletromagnético de Argônio Líquido do detector ATLAS”, <http://atlas.web.cern.ch/Atlas/GROUPS/LIQARGON/Organization/index.html>, acessado em agosto de 2013.
- [22] CERN, A., *ATLAS liquid argon calorimeter: Technical design report*, Report CERN-LHCC-96-41, ATLAS Collaboration, 1996.
- [23] “Calorímetro Hadrônico de Telhas Cintilantes”, <https://twiki.cern.ch/twiki/bin/view/Atlas/TileCal>, acessado em agosto de 2013.
- [24] CERN, A., *ATLAS tile calorimeter: Technical design report*, Report CERN-LHCC-96-42, ATLAS Collaboration, 1996.
- [25] VARANDA, M., “The tile hadronic calorimeter for the ATLAS experiment”, *IEEE Trans.Nucl.Sci.*, v. 48, pp. 367–371, 2001.

- [26] “Câmara de Muons do detector ATLAS”, <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonSpectrometer>, acessado em agosto de 2013.
- [27] ATLAS Muon Collaboration, “ATLAS Muon Spectrometer Technical Design Report”, 1997, CERN/LHCC 97-22.
- [28] PALESTINI, S., “The muon spectrometer of the ATLAS experiment”, *Nucl.Phys.Proc.Suppl.*, v. 125, pp. 337–345, 2003.
- [29] ALAM, M., OTHERS, “ATLAS pixel detector: Technical design report”, , 1998.
- [30] AHMAD, A., ALBRECHTSKIRCHINGER, Z., ALLPORT, P., *et al.*, “The Silicon microstrip sensors of the ATLAS semiconductor tracker”, *Nucl.Instrum.Meth.*, v. A578, pp. 98–118, 2007.
- [31] ABAT, E., OTHERS, “The ATLAS TRT barrel detector”, *JINST*, v. 3, pp. P02014, 2008.
- [32] WIGMANS, R., *Calorimetry: Energy Measurement in Particle Physics*. Clarendon Press, 2003.
- [33] F. ARIZTIZABAL, M. BOSMAN, M. C.-S., “Construction and performance of an iron-scintillator hadron calorimeter with longitudinal tile configuration”. In: *Nuclear Instruments and Methods in Physics Research*, pp. 384 – 397, 1994.
- [34] K. Anderson, J. Pilcher, H. Sanders, F. Tang, “Front-end Electronics for the ATLAS Tile Calorimeter”. In: *4th Workshop on Electronics for the LHC Experiments*, 1998.
- [35] “TMS320C6414,Fixed-Point Digital Signal Processor, Texas Instruments”, <http://focus.ti.com/docs/prod/folders/print/tms320c6414.html>, acessado em agosto de 2013.
- [36] OLIVEIRA DAMAZIO, D., “ATLAS High Level Calorimeter Trigger Software Performance for Cosmic Ray Events”, , 2009.
- [37] PETERSEN, B. A., “The ATLAS Trigger Performance and Evolution”, *PoS*, v. ICHEP2012, pp. 493, 2013.
- [38] FERMILAB, “Fermilab History and Archives Project”.
- [39] DETECTOR, C., “The Collider Dector at FermiLab - Web Site”.

- [40] ABE, F., OTHERS, “The CDF Detector: An Overview”, *Nucl.Instrum.Meth.*, v. A271, pp. 387–403, 1988.
- [41] AMIDEI, D. E., OTHERS, “The Silicon Vertex Detector of the Collider Detector at Fermilab”, *Nucl.Instrum.Meth.*, v. A350, pp. 73–130, 1994.
- [42] AFFOLDER, A. A., OTHERS, “CDF central outer tracker”, *Nucl.Instrum.Meth.*, v. A526, pp. 249–299, 2004.
- [43] JENSEN, H. B., “The Collider Detector at Fermilab (CDF)”, *IEEE Trans.Nucl.Sci.*, v. 33, pp. 40–45, 1986.
- [44] JENSEN, H. B., “CDF Calorimetry”, *Nucl.Instrum.Meth.*, v. A265, pp. 186–199, 1988.
- [45] PAULETTA, G., “Operation and performance of the CDF calorimeters”, *J.Phys.Conf.Ser.*, v. 160, pp. 012007, 2009.
- [46] GINSBURG, C., “CDF Run 2 muon system”, *Eur.Phys.J.*, v. C33, pp. S1002–S1004, 2004.
- [47] BYRUM, K., OTHERS, “The CDF Forward Muon System”, *Nucl.Instrum.Meth.*, v. A268, pp. 46, 1988.
- [48] DORIGO, T., “The muon system upgrade for the CDF II experiment”, *Nucl.Instrum.Meth.*, v. A461, pp. 560–562, 2001.
- [49] Bhatti, A., Canelli, F., Heinemann, B., *et al.*, “Determination of the jet energy scale at the Collider Detector at Fermilab”, *Nuclear Instruments and Methods in Physics Research A*, v. 566, pp. 375–412, Oct. 2006.
- [50] AALTONEN, T., BUZATU, A., KILMINSTER, B., *et al.*, “Improved b -jet Energy Correction for $H \rightarrow b\bar{b}$ Searches at CDF”, *Internal Note*, , n. FERMILAB-TM-2513-PPD, 2011.
- [51] ABACHI, S., OTHERS, “The $D\emptyset$ Detector”, *Nucl.Instrum.Meth.*, v. A338, pp. 185–253, 1994.
- [52] AHMED, S., OTHERS, “The $D\emptyset$ Silicon Microstrip Tracker”, *Nucl.Instrum.Meth.*, v. A634, pp. 8–46, 2011.
- [53] CLARK, A., OTHERS, “The Central tracking detectors for $D\emptyset$ ”, *Nucl.Instrum.Meth.*, v. A279, pp. 243–248, 1989.
- [54] YAMADA, R., EARTLY, D. P., OSTIGUY, J., *et al.*, “Magnetic characteristics of the $D\emptyset$ detector”, *IEEE Trans.Magnetics*, v. 28, pp. 520–523, 1992.

- [55] FRANZINI, P., “Performance of the $D\emptyset$ uranium - liquid argon calorimeter modules. D0 Calorimeter Group”, *Nucl.Instrum.Meth.*, v. A289, pp. 438–445, 1990.
- [56] ABAZOV, V., ACHARYA, B. S., ALEXEEV, G., *et al.*, “The Muon system of the run II $D\emptyset$ detector”, *Nucl.Instrum.Meth.*, v. A552, pp. 372–398, 2005.
- [57] ABBOTT, B., OTHERS, “Determination of the absolute jet energy scale in the $D\emptyset$ calorimeters”, *Nucl.Instrum.Meth.*, v. A424, pp. 352–394, 1999.
- [58] METROPOLIS, N., “The beginning of the Monte Carlo method”. In: *Special Issue dedicated to Stanislaw Ulam*, 1987.
- [59] GUIDA, J. [STATE UNIV. OF NEW YORK, S. B. N. U. S., COLLABORATION, T. D., , Dec 1992.
- [60] BHIMJI, W., CRANSHAW, J., GEMMEREN, P. V., *et al.*, “The ATLAS ROOT-based data formats: recent improvements and performance measurements”, *Journal of Physics: Conference Series*, v. 396, n. 2, pp. 022006, 2012.
- [61] AKESSON, F., COSTA, M., DOBOS, D., *et al.*, “ATLAS Inner Detector Event Data Model”, , 2007.
- [62] ARMSTRONG, W., OTHERS, “ATLAS: Technical proposal for a general-purpose p p experiment at the Large Hadron Collider at CERN”, , 1994.
- [63] DJOUADI, A., KALINOWSKI, J., SPIRA, M., “HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension”, *Comput.Phys.Commun.*, v. 108, pp. 56–74, 1998.
- [64] LUKAS, W., “Fast Simulation for ATLAS: Atlfast-II and ISF”, , 2012.
- [65] AAD, G., OTHERS, “The ATLAS Simulation Infrastructure”, *Eur.Phys.J.*, v. C70, pp. 823–874, 2010.
- [66] ANTCHEV, G., ASPELL, P., ATANASSOV, I., *et al.*, “First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s} = 7$ TeV”, *Europhys.Lett.*, v. 96, pp. 21002, 2011.
- [67] SJOSTRAND, T., MRENNNA, S., SKANDS, P. Z., “A Brief Introduction to PYTHIA 8.1”, *Comput.Phys.Commun.*, v. 178, pp. 852–867, 2008.
- [68] “First tuning of HERWIG/JIMMY to ATLAS data”, , 2010.

- [69] KERSEVAN, B. P., RICHTER-WAS, E., “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”, *Comput.Phys.Commun.*, v. 184, pp. 919–985, 2013.
- [70] *Summary of ATLAS Pythia 8 tunes*, Report ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug 2012.
- [71] DJOUADI, A., KALINOWSKI, J., SPIRA, M., “HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension”, *Comput.Phys.Commun.*, v. 108, pp. 56–74, 1998.
- [72] GLEISBERG, T., HOECHE, S., KRAUSS, F., *et al.*, “Event generation with SHERPA 1.1”, *JHEP*, v. 0902, pp. 007, 2009.
- [73] GUZZI, M., NADOLSKY, P., BERGER, E., *et al.*, “CT10 parton distributions and other developments in the global QCD analysis”, , 2011.
- [74] FRIXIONE, S., NASON, P., OLEARI, C., “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP*, v. 0711, pp. 070, 2007.
- [75] SJOSTRAND, T., MRENNA, S., SKANDS, P. Z., “PYTHIA 6.4 Physics and Manual”, *JHEP*, v. 0605, pp. 026, 2006.
- [76] SKANDS, P. Z., “Tuning Monte Carlo Generators: The Perugia Tunes”, *Phys.Rev.*, v. D82, pp. 074018, 2010.
- [77] MCCULLOCH, W., PITTS, W., “A logical calculus of the ideas immanent in nervous activity”, *The bulletin of mathematical biophysics*, v. 5, n. 4, pp. 115–133, 1943.
- [78] MINSKY, M., PAPERT, S., *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA, MIT Press, 1969.
- [79] RUMELHART, D. E., HINTON, G. E., WILLIAMS, R. J., “Neurocomputing: Foundations of Research”. chapter Learning Representations by Back-propagating Errors, Cambridge, MA, USA, MIT Press, pp. 696–699, 1988.
- [80] CYBENKO, G., “Approximation by Superpositions of a Sigmoidal Function”, *Mathematics of Control, Signals, and Systems*, v. 2, pp. 303–314, 1989.
- [81] GOU-JEN WANG, J.-J. C., “Cascade steepest descent learning algorithm for multilayerfeedforward neural network”. In: *International Joint Conference on Neural Networks*, pp. 1889 – 1894 Vol.3, 1999.

- [82] V.S. ASIRVADAM, S.F. MCLOONE, G. I., “Memory efficient BFGS neural-network learning algorithms using MLP-network: a survey”. In: *Proceedings of the 2004 IEEE International Conference on Control Applications*, pp. 586 – 591 Vol.1, 2004.
- [83] GILL, P. E., MURRAY, W., WRIGHT, M. H., *Practical optimization*. London, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], 1981.
- [84] MARQUARDT, D., “An Algorithm for Least-Squares Estimation of Nonlinear Parameters”, *Journal of the Society for Industrial and Applied Mathematics*, v. 11, n. 2, pp. 431–441, 1963.
- [85] BLOOM, E. D., PECK, C., “Physics with the Crystal Ball Detector”, *Ann.Rev.Nucl.Part.Sci.*, v. 33, pp. 143–197, 1983.
- [86] BLOOM, E. D., PECK, C., “Physics with the Crystal Ball Detector”, *Ann.Rev.Nucl.Part.Sci.*, v. 33, pp. 143–197, 1983.
- [87] Martinez Santos, D., Dupertuis, F., “Mass distributions marginalized over per-event errors”, *ArXiv e-prints*, , Dec. 2013.
- [88] SANTOS, D. M., DUPERTUIS, F., “Mass distributions marginalized over per-event errors”, , 2013.
- [89] Bukin, A. D., “Fitting function for asymmetric peaks”, *ArXiv e-prints*, , Nov. 2007.
- [90] SAIN, S. R., “Adaptive Kernel Density Estimation”, 1994.
- [91] FALK, M., “Relative efficiency and deficiency of kernel type estimators of smooth distribution functions”, *Statistica Neerlandica*, v. 37, n. 2, pp. 73–83, 1983.
- [92] KULLBACK, S., LEIBLER, R. A., “On Information and Sufficiency”, *The Annals of Mathematical Statistics*, v. 22, n. 1, pp. 79–86, 03 1951.
- [93] LIN, J., “Divergence measures based on the Shannon entropy”, *Information Theory, IEEE Transactions on*, v. 37, n. 1, pp. 145–151, Jan 1991.
- [94] “University of Glasgow”, <http://www.gla.ac.uk/>, acessado em fevereiro de 2014.
- [95] Eadie, W. T., Drijard, D., James, F. E., *Statistical methods in experimental physics*, 1971.