Universidade Federal de Juiz de Fora
Programa de Pós-Graduação em
Engenharia Elétrica

David de Melo Souza

ESTIMAÇÃO DE DENSIDADES MULTIVARIADAS PARA A FILTRAGEM DE EVENTOS BASEADA EM UM DETECTOR DE ALTAS ENERGIAS COM FINA SEGMENTAÇÃO

Dissertação de Mestrado

Juiz de Fora
2015

David de Melo Souza

ESTIMAÇÃO DE DENSIDADES MULTIVARIADAS PARA A FILTRAGEM DE EVENTOS BASEADA EM UM DETECTOR DE ALTAS ENERGIAS COM FINA SEGMENTAÇÃO

Dissertação apresentada ao Programa de Pós–Graduação em Engenharia Elétrica, área de concentração: Sistemas Eletrônicos, da Faculdade de Engenharia da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Mestre.

Orientadores: Prof. Rafael Antunes Nóbrega, D.Sc.
Prof. José Manoel Seixas, D.Sc.

Juiz de Fora
2015

David de Melo Souza

ESTIMAÇÃO DE DENSIDADES MULTIVARIADAS PARA A FILTRAGEM DE EVENTOS BASEADA EM UM DETECTOR DE ALTAS ENERGIAS COM FINA SEGMENTAÇÃO

Dissertação apresentada ao Programa de Pós–Graduação em Engenharia Elétrica, área de concentração: Sistemas Eletrônicos, da Faculdade de Engenharia da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Mestre.

Aprovada em 10 de Setembro de 2015.

BANCA EXAMINADORA:

Prof. Rafael Antunes Nóbrega, D.Sc.
Universidade Federal de Juiz de Fora, UFJF

Prof. José Manoel Seixas, D.Sc.
Universidade Federal do Rio de Janeiro, UFRJ

Prof. Augusto Santiago Cerqueira, D.Sc.
Universidade Federal de Juiz de Fora, UFJF

Prof. Ernesto Kemp, D.Sc.
Universidade Estadual de Campinas, UNICAMP

Aos meus pais, minha esposa, minha irmã, aos meus familiares e amigos.

AGRADECIMENTOS

Primeiro, gostaria de agradecer a Deus, pelas permissões, possibilidades e determinações.

Aos meus pais, Adélia e Carlos, pelo amor, pela referência segura e dedicação genuína. Vocês fazem de cada detalhe existencial, uma lição de caráter e essência de vida.

À minha esposa, Laila, pelo amor, paciência e incentivo altruísta, metaforizado pela tristeza da ausência no olhar. Sem você os desafios seriam muito maiores e as conquistas não teriam o mesmo valor.

À minha irmã Priscila e sobrinhos, pela cumplicidade e amor. Impossível descrever como nos impacta o sorriso de uma criança.

Ao meu orientador, Rafael, por me ensinar diariamente os melhores caminhos pra se fazer ciência. Sem você a realização deste trabalho não seria possível.

Ao meu coorientador, Seixas, pela atenção e ajuda, ponderando entre o trabalho acadêmico e a ótica da colaboração.

Ao meu amigo, Igor. Por ser engrenagem fundamental deste trabalho, comemorar cada curva ROC e esbravejar a falta de estatística.

Aos colegas de trabalho, Denis e Werner, pelas ajudas pontuais e essenciais. Vocês me pouparam muito tempo de estudo.

Aos companheiros do LAPTEL e amigos de Laboratório Tony, Kátia e Tiago. Pelas conversas, ajudas e bom humor. É sempre bom ter alguém para compartilhar alegrias e desespero. Melhor ainda, para dividir a loucura.

Finalmente, agradeço à CAPES(Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), à Universidade Federal de Juiz de Fora e à Faculdade de Engenharia por todo o suporte e pelas ferramentas necessárias ao desenvolvimento deste trabalho.

RESUMO

Nas últimas décadas, a sinergia entre engenharia e física, nas áreas de aplicação da física moderna, tem sido crescente. Para o programa de física do ATLAS, no CERN, por exemplo, a identificação de elétrons é de fundamental importância, sendo uma demanda responsável por diversos estudos em engenharia. Esse trabalho se desenvolve nesse viés, tendo como base a técnica de verossimilhança utilizada pela colaboração ATLAS na identificação *offline* de elétrons relevantes, considerados sinal, em meio a diversas partículas, consideradas ruído de fundo. Atualmente, a verossimilhança tem sido aplicada pela colaboração de forma simplificada, supondo independência entre as variáveis discriminantes fornecidas pelo detector ATLAS. Essa consideração, possibilita que a formulação matemática da probabilidade conjunta seja feita pela utilização do produtório das densidades marginais das variáveis discriminantes. Entretanto, a simplificação promove um erro na reconstrução da probabilidade conjunta, visto que, algumas variáveis discriminantes possuem um certo grau de dependência entre si. Esse cenário, nos abre a possibilidade de melhora do método, a partir de técnicas capazes de mitigar a dependência entre tais varáveis. A principal contribuição desse trabalho se dá na implementação de um algoritmo baseado na técnica não-paramétrica para estimação de densidade multivariada conhecida como MKDE (do inglês, *Multivariate Kernel Density Estimation*), com o objetivo de minimizar o erro de estimação da probabilidade conjunta, que ocorre devido à consideração de independência acima citada. Dentro da realidade comparativa deste trabalho, foi possível observar a melhora na estimação da probabilidade conjunta via MKDE e a propagação desta melhora na identificação de elétrons.

Palavras-chave: Estimação de Densidade de Probabilidade, Verossimilhança, Identificação de Elétrons, Detector ATLAS, KDE Multivariado.

ABSTRACT

The electron identification is of fundamental importance to the ATLAS physics program, at CERN. This Master's Thesis planned to study and to reproduce one of the main offline algorithms, based on nonparametric maximum likelihood estimation, applied to identify electron/positron particles using the ATLAS Detector to then propose additional processing techniques that could improve its performance. The ATLAS Collaboration simplifies the likelihood method by considering independence between the discriminant variables. This approach opens possibilities for improving the method by means of applying techniques capable of mitigating the variables dependence. Our main contribution lies in the implementation of an algorithm based on Multivariate Kernel Density Estimation (MKDE). This algorithm should be able to decrease the error caused by the variables dependence, as mentioned above, improving the ATLAS electron identification performance. The impact of this new proposal was also compared to the most used algorithms developed by ATLAS group, known as Egamma and Likelihood.

Keywords: Electron identification, Likelihood, Multivariate KDE.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ALICE *A Large Ion Collider Experiment*

ATLAS *A Toroidal LHC Apparatus*

CERN *Conseil Européen pour la Recherche Nucléaire*

CMS *Compact Muon Solenoid*

EDF *Empirical Distribution Function*

EF Filtro de Eventos, (do inglês, *Event Filter*)

EM Calorímetro Eletromagnético, (do inglês, *Eletromagnetic Calorimeter*)

GSF *Gaussian Sum Filter*

HAD Calorímetro Hadrônico, (do inglês, *Hadronic Calorimeter*)

HLT Filtragem de Alto nível, (do inglês, *High Level Trigger*)

IBL *Insertable B-Layer*

ID Detector Interno, (do inglês, *Inner Detector*)

KDE Estimação de Densidade de Núcleo, (do inglês, *Kernel Density Estimation)*

MKDE Estimação de Densidade de Núcleo Multivariada, (do inglês, *Multivariate Kernel Density Estimation)*

L1 *Level 1*

L2 *Level 2*

LHC *Large Hadron Collider*

LH *Likelihood*

LHCb *Large Hadron Collider beauty*

LINAC 2 Acelerador Linear, (do inglês, *Linear accelerator 2*)

MC Monte Carlo

PDF Função Densidade de Probabilidade(do inglês, *Probability density function*)

PMT Tubo Fotomultiplicador, (do inglês, *Photomultiplier tubes*)

PS *Proton Synchrotron*

PSB *Proton Synchrotron Booster*

RMS Raiz do Valor Quadrático Médio, (do inglês, *Root Mean Square*)

SCT *SemiConductor Tracker*

SPD *Silicon Pixel Detector*

SPS *Super Proton Synchrotron*

TP *Tag and Probe*

TRT *Transition Radiation Tracker*

MISE Erro médio quadrático integrado, (do inglês, *Mean Integrated Squared Error*)

SUMÁRIO

1 INTRODUÇÃO

A teoria da estimação pode ser encontrada no cerne de vários problemas ligados à Engenharia Elétrica, mais especificamente, em sistemas de processamento de sinais projetados para extrair informação de alguma natureza. Esses sistemas incluem, radar, sonar, fala, imagem, biomedicina, comunicação, controle, entre outros. A maioria das aplicações é projetada para se estimar parâmetros desconhecidos a partir de uma coleção de observações contaminadas por ruído, tendo em vista uma hipótese ou modelo inicial. Quando o modelo dos dados é desconhecido, deve-se partir para o campo da estimação não-paramétrica, que considera métodos com uma grande generalidade de aplicação, uma vez que suas hipóteses subjacentes são pouco restritivas, permitindo-nos lidar com um número maior de situações. Dentro desse assunto, os dois problemas mais abordados na literatura são o de estimação de funções de regressão e de estimação de densidades; esta última, sendo o principal assunto dessa dissertação.

Estimar a densidade de probabilidade de uma ou mais variáveis aleatórias promove a possibilidade de fazer previsões ou inferências probabilísticas de determinado evento associado a essa ou essas variáveis, sendo possível, de acordo com a(s) característica(s) de cada evento, traduzida por cada variável em estudo, prever a qual grupo (em nosso caso, sinal e ruído de fundo) pertence. Essa busca, pela melhor representação probabilística de um grupo, serve como base para uma posterior diferenciação entre indivíduos (ou eventos) de diferentes grupos, utilizadas em algoritmos de seleção de eventos.

Em física de altas energias, o desempenho dos algoritmos de seleção de eventos é fundamental, devido a crescente quantidade de ruído de fundo em relação aos sinais de interesse. Podemos citar, como exemplo, uma partícula em estudo no CERN, mais divulgada pela mídia: o bóson de Higgs. Estima-se que apenas três Higgs sejam produzidos a cada 10^{10} colisões, sendo um de seus modos de decaimento quatro elétrons. Se a eficiência do algoritmo de identificação de elétrons for baixa, boa parte desses eventos raros serão perdidos. Por outro lado, se a rejeição de ruído de fundo for baixa, um grande número de Higgs será reconstruído erroneamente.

Geralmente, esses algoritmos, se baseiam em múltiplas variáveis, formadas a partir dos sinais gerados na interação entre detectores e partículas. Entretanto, o comportamento dessas variáveis não pode ser modelado facilmente através de distribuições de probabilidade conhecidas. Nesse contexto, estimadores não-paramétricos podem ser aplicados através do processamento conjunto das informações disponíveis, objetivando uma percepção mais profunda do problema, propondo um caminho a ser explorado na busca pela otimização das técnicas utilizadas em identificação de partículas.

1.1 MOTIVAÇÃO

A colaboração ATLAS publicou (COLLABORATION et al., 2013), em 2013, resultados iniciais aplicando o método da verossimilhança na identificação de elétrons, utilizando dados produzidos com uma energia de 8 TeV. Entretanto, constatou-se a possibilidade de melhoria no método, visto que, embora fosse conhecida a dependência entre algumas variáveis, sua formulação considerou uma simplificação, ao assumir independência entre tais variáveis discriminantes fornecidas pelo detector ATLAS. Portanto, surgem questões relativas ao nível de degradação da performance do algoritmo de identificação de elétrons, devido a essa presunção de independência. Paralelamente ao estudo do método, existe uma perspectiva de aprendizado das técnicas envolvidas em estimação não-paramétrica de densidades, inserida no contexto de reconhecimento de padrões em física de partículas.

1.2 O QUE FOI FEITO

Em primeira instância, foi feito um estudo do artigo da colaboração, citado acima, referente à verossimilhança aplicada na identificação de elétrons. Teorias utilizadas e termos específicos foram o foco desse estudo, mapeando possíveis soluções e problemas, inerentes ao estudo.

Posteriormente, foram construídos os algoritmos necessários para aplicação da verossimilhança, obedecendo as características citadas no artigo, com o intuito de reproduzir, tanto quanto possível, o método descrito. Sequencialmente, foi possível comparar a perfomance do algoritmo construído nessa dissertação com o artigo base, servindo como realimentação para ajustes e aprofundamento no método.

Sendo percebida a coerência entre a eficiência de identificação de elétrons e rejeição de ruído de fundo, o trabalho encerrou a etapa de reprodução do método de verossimi-

lhança utilizado pela colaboração, e iniciou a busca por técnicas capazes de melhorar sua performance.

Como técnica escolhida, e principal contribuição dessa dissertação, tem-se a estimação não-paramétrica multivariada, sendo objeto de estudo teórico mais aprofundado. Posteriormente, essa técnica foi implementada e associada a um novo algoritmo de verossimilhança, que agora, busca melhorar a estimação de densidade de probabilidade de variáveis dependentes.

Todas as análises foram feitas utilizando dados simulados e dados reais, levando a necessidade da implementação de algoritmos capazes de contemplar essas duas realidades. Além de todo estudo associado, os algoritmos implementados nessa dissertação foram:

- *Kernel* univariado
  - *l*argura de banda fixa
  - *l*argura de banda variável
- *Kernel* multivariado
  - *l*argura de banda fixa
  - *l*argura de banda variável
- Informação Mútua
- *Tag and Probe*
- Verossimilhança
  - *naive*: utilizando apenas densidades de probabilidade univariadas
  - M_{KDE}: utilizando densidades de probabilidade multivariadas

Através desses algoritmos foram feitas análises com diferentes densidades de probabilidade, com a construção de diferentes configurações de verossimilhanças; resultados em diferentes regiões em η, referente a posição do detector e E_t, referente a energia; avaliação dos algoritmos em diferentes valores de empilhamento N_{vtx}; comparação com os algoritmos da colaboração; comparação entre a verossimilhança *naive*, desconsiderando a dependência, e o método de verossimilhança M_{KDE} considerando a dependência.

1.3 ESTRUTURA DA DISSERTAÇÃO

O Capítulo 2, será dedicado a ambientação, apresentando alguns conceitos básicos da Física de Altas Energias e uma introdução sobre o *Conseil Européen pour la Recherche Nucléaire* (CERN). Será apresentado o ambiente em que a dissertação foi desenvolvida, através do detalhamento do Detector *A Toroidal LHC Apparatus* (ATLAS).

O Capítulo 3 expõe à formulação do problema de identificação de sinais. Será feita uma revisão sobre as variáveis empregadas no experimento. Além disso, serão descritos os métodos atuais implementados no ATLAS para a identificação de elétrons *offline*.

No Capítulo 4 será mostrada uma revisão bibliográfica das teorias matemáticas utilizados nessa dissertação.

No Capítulo 5 teremos a explicação do funcionamento dos algoritmos desenvolvidos, e sua utilização na identificação de elétrons.

O Capítulo 6 apresenta os resultados de identificação de elétrons pela técnica proposta e comparação com a técnica atual. Serão utilizados dados simulados e dados reais adquiridos durante operação nominal do LHC em 2012.

Por fim, o Capítulo 7 apresentará as principais conclusões do trabalho e algumas propostas de desdobramentos futuros.

2 O QUE É? MODELO PADRÃO, CERN, LHC E ATLAS

No ambiente de física de partículas, uma parte dos leitores encontra dificuldade na contextualização do problema, visto que, ferramentas, nomenclaturas, tecnologias, implementações e métodos, exigem uma especificidade teórica que, atualmente, não é comum na formação em Engenharia.

Esta seção será dedicada a apresentar uma visão geral do estudo, mostrando a relação entre alguns acrônimos, importantes para o entendimento do plano de fundo dessa dissertação. Portanto, como relacionar, modelo padrão, CERN , *Large Hadron Collider* (LHC) e ATLAS?

O modelo padrão é o que melhor descreve as partículas elementares que formam nosso universo e as forças que governam essas partículas. Algumas partículas, desse modelo, são estáveis e compõem a matéria conhecida, entretanto, a maioria dura frações de segundo antes de decaírem em partículas mais estáveis. A fim de observarmos estas partículas, instáveis, é necessário recriar um ambiente propício ao seu "surgimento". Com este propósito, o CERN utiliza um acelerador de partículas, chamado LHC, capaz de recriar tal ambiente, ao fazer partículas serem aceleradas em direções opostas e colidirem em altas energias.

Após a colisão, faz-se necessário detectores capazes de interpretar as informações do ocorrido. Nessa dissertação utilizaremos informações do detector ATLAS, usadas por grupos de performance da colaboração, vinculados à identificação de elétrons pela técnica de verossimilhança.

2.1 MODELO PADRÃO

O modelo padrão é um conjunto de teorias que descreve os mecanismos de interação regidos por três das quatro forças conhecidas, bem como a estrutura fundamental da matéria. As forças abrangidas pelo modelo padrão são: eletromagnética, fraca e forte, sendo que, em altas energias, a força eletromagnética e a força fraca são descritas como

uma única força, eletrofraca. O modelo padrão não considera a força gravitacional, e conforme suas teorias, existem duas partículas fundamentais: Os férmions e os bósons (PERKINS, 2000).

Os férmions são partículas que constituem a matéria e são subdivididas em léptons e quarks. Os léptons são: elétron, múon, tau, seus neutrinos e suas antipartículas. Os quarks são: up, down, charme, strange, top e bottom e suas antipartículas.

As interações entre partículas são mediadas através de trocas de partículas transportadoras de forças, chamadas bósons, que são: glúon (força forte), fóton (força eletromagnética), bósons W e Z (força fraca) , bóson de Higgs (responsável pela existência de massa inercial).

A Figura 1 resume algumas informações das partículas do Modelo Padrão.

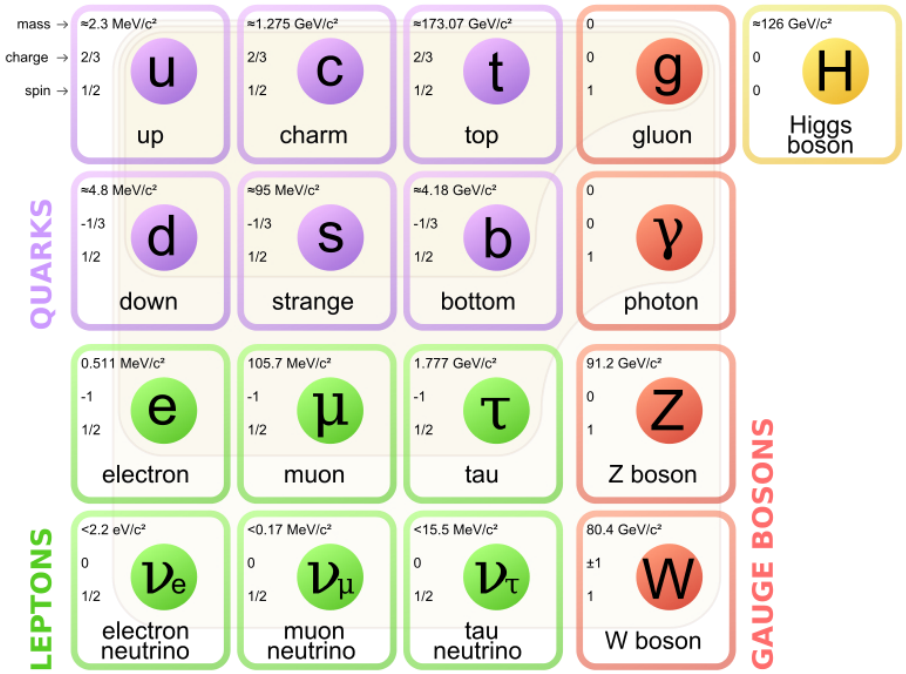

Figura 1: Modelo Padrão.

2.2 CERN

O CERN (do francês, *Conseil Européen pour la Recherche Nucléaire*) (CERN, 2015a) é o maior laboratório de física de partículas do mundo e conta com a colaboração de diversos países. Fundado em 1954, situa-se na fronteira franco-suiça, perto de Genebra. Neste laboratório, engenheiros e físicos trabalham em conjunto, a fim de estudar a estrutura fundamental da matéria, utilizando os maiores e mais complexos instrumentos científicos do mundo.

Os estudos desenvolvidos neste laboratório são bastante amplos. Alguns tópicos estudados são: O Modelo Padrão, Matéria Escura, Dimensões Extras, Grávitons, Super Simetria, Bóson Z, Bóson W e Bóson de Higgs. As inovações tecnológicas adquiridas com esses estudos promovem a inserção da ciência em aplicações mais próximas da sociedade, como: Tratamento de câncer e diversos estudos na Medicina, computação, criptografia, transferência de tecnologia para a Indústria, dentre outros (CERN, 2015b).

A necessidade experimental dos estudos levou a construção do acelerador de partículas LHC. Neste acelerador, feixes de prótons são acelerados em direções opostas até atingirem altas energias, antes de colidirem uns com os outros. No exato local onde ocorrem as colisões entre os feixes, faz-se necessária a utilização de detectores, capazes de identificar os subprodutos dessas colisões. O LHC conta com quatro detectores principais: *A Large Ion Collider Experiment* (ALICE), ATLAS, *Compact Muon Solenoid* (CMS), *Large Hadron Collider beauty* (LHCb). Neste estudo, daremos maior ênfase ao ATLAS, visto que, essa dissertação foi baseada em dados gerados por esse detector.

2.3 LHC

O LHC (CERN, 2015c) é composto por um túnel subterrâneo (Figura 2) circular com aproximadamente 27 km de extensão, que se encontra a 100 metros de profundidade, no qual, prótons são conduzidos e acelerados por campos eletromagnéticos. A fonte de prótons é um simples cilindro de gás hidrogênio. Um campo eletromagnético é utilizado para retirar seus elétrons, restando prótons. Esses prótons são inseridos no acelerador em pacotes, formando feixes de prótons, que são conduzidos ao primeiro acelerador da cadeia, o Acelerador Linear, (do inglês, *Linear accelerator 2*) (LINAC 2), que acelera os feixes até a energia de 50 MeV. Esse feixe é injetado no *Proton Synchrotron Booster* (PSB) que impulsiona os prótons até 1,4 GeV. Em seguida o feixe segue para o *Proton Synchrotron* (PS) que o leva até uma energia de 25 GeV. Então, o feixe é enviado para o *Super Proton Synchrotron* (SPS), onde é novamente acelerado, atingindo 450 GeV. Finalmente, os feixes de prótons são transferidos para os dois tubos do LHC e viajam em sentidos opostos até atingirem a energia de 4 TeV, por feixe, correspondendo a energia de colisão de 8 TeV.

Os feixes, que viajam em sentidos opostos, trafegam no acelerador em tubos separados, onde existem pontos de interseção, sem campos magnéticos, em que esses feixes percorrem uma trajetória retilínea, entrando em rota de colisão. Esses pontos de

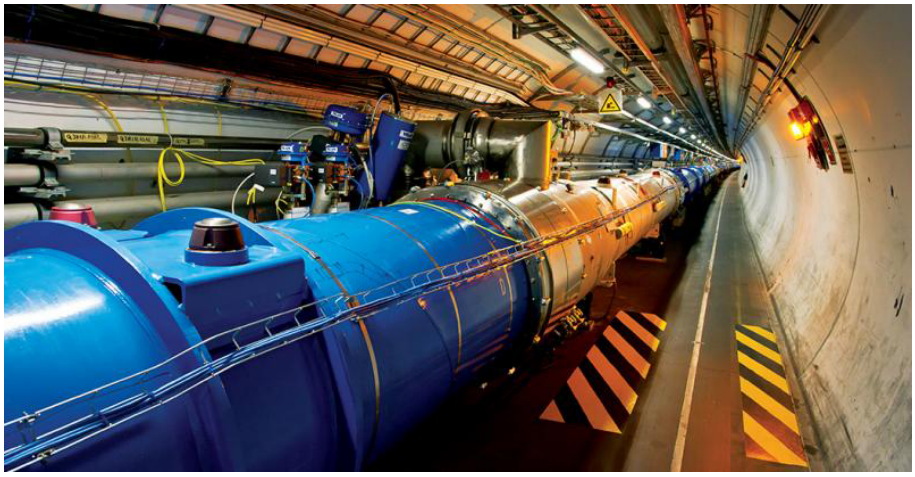

Figura 2: O acelerador de partículas, LHC. Extraído de (CERN, 2015c).

colisão ocorrem no interior dos detectores, em um ambiente a vácuo, onde o número de colisões, entre os prótons, depende da quantidade de partículas em cada pacote, da seção transversal dos prótons e do desenho do acelerador. A combinação dessas características, dependentes do projeto do acelerador, podem ser combinadas em um parâmetro chamado luminosidade.

A luminosidade, geralmente, é expressa em $cm^{-2}s^{-1}$, e tem como definição o número de partículas por unidade de área, por unidade de tempo, vezes a opacidade do alvo. Em um acelerador circular, a luminosidade é definida como:

$$L = fn\frac{N_1 N_2}{A} \tag{2.1}$$

onde f é a frequência de colisões, n é o número de pacotes de partículas que compõe o feixe, N_i é o número de prótons em cada pacote e A a seção transversal do feixe.

Em experimentos de física de partículas, atingir elevados níveis de luminosidade tem tanta importância quanto atingir elevados níveis de energia. A energia está relacionada a probabilidade de gerar distintos eventos raros de interesse, enquanto a luminosidade se relaciona a probabilidade de observação desses eventos.

A luminosidade nominal do LHC para o ATLAS e o CMS é de $10^{34}cm^{-2}s^{-1}$, para feixes de 2808 pacotes, com $1.1 * 10^{11}$ prótons cada. Nos outros dois detectores teremos uma luminosidade menor, com $10^{32}cm^{-2}s^{-1}$ no LHCb e $10^{30}cm^{-2}s^{-1}$ no ALICE (EVANS; BRYANT, 2008). A Figura 3 mostra a localização dos detectores (ALICE, ATLAS, CMS e LHCb) no LHC.

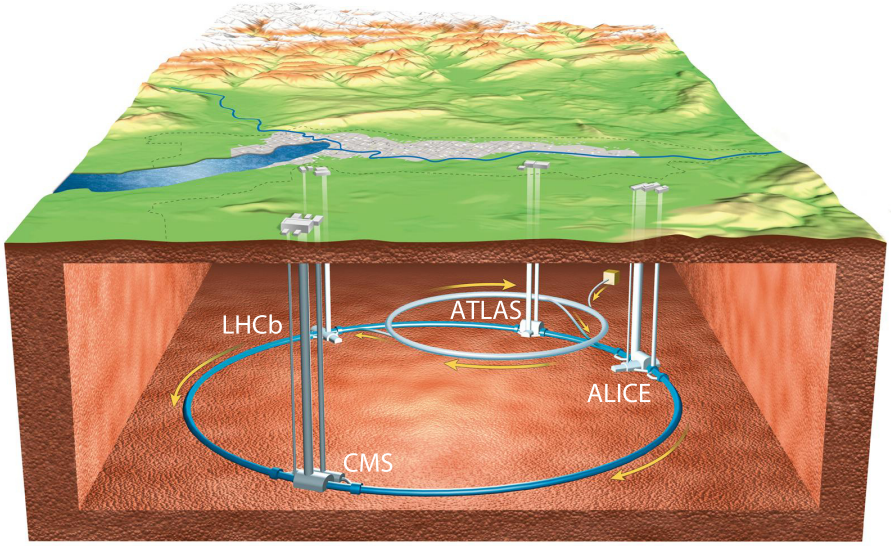

Figura 3: Uma visão geral do LHC. Extraído de CERN.

Ao passar por dois anos de modernização e manutenção, o LHC reiniciou suas atividades, em 2015, com uma energia de colisão de 13 TeV, apesar de ter sido projetado para um máximo de 14 TeV. Essa decisão ocorreu devido às exigências de campo magnético necessárias para 13 TeV demandarem um período de tempo menor do que para 14 TeV, sendo a melhor alternativa para chegarem rapidamente a novos resultados, a uma energia nunca antes alcançada.

2.4 O DETECTOR ATLAS

O detector ATLAS (AAD et al., 2008) é um dispositivo de formato cilíndrico com 44 metros de comprimento e 25 metros de altura, como mostrado na Figura 4. O ATLAS é um detector de uso geral. Portanto, precisa ser capaz de identificar diversos tipos de processos físicos de interesse, a partir dos subprodutos das colisões próton-próton. Sua posição em relação ao LHC pode ser vista na Figura 5.

Para a reconstrução e identificação das energias depositadas pelas partículas resultantes das colisões e suas respectivas trajetórias, o ATLAS é dividido em subdetectores independentes (com características diversas), de modo que, o conjunto das informações fornecidas por esses subdetectores nos permite conhecer o perfil de cada partícula. Como mostra a Figura 4, na parte mais interna, encontramos o Detector Interno, representado por três subdetectores: *Pixel detector*, *Semicondutor tracker* e *Transition radiation tracker*. Ao seu redor, o calorímetro eletromagnético (ou *LAr electromagnetic calorimeters*). Logo após, o calorímetro hadrônico (*Tile calorimeters*) seguido dos

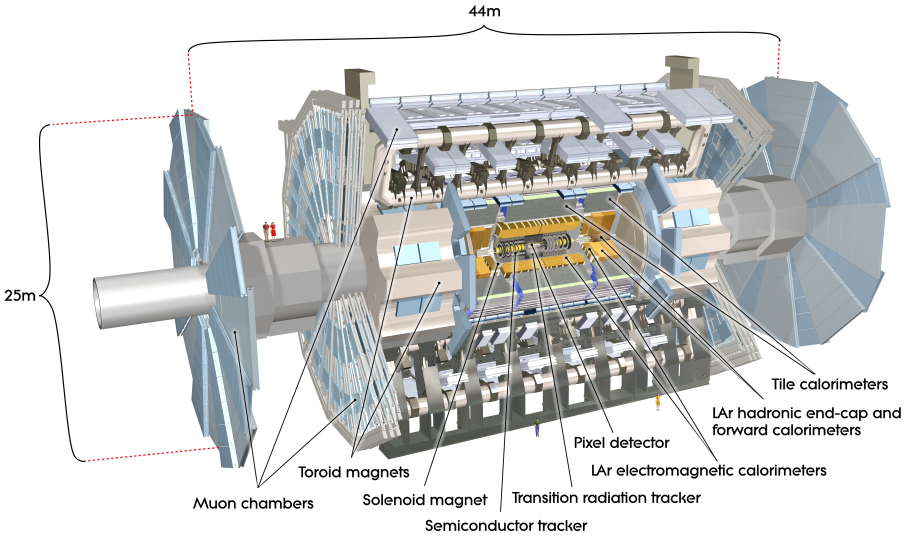

Figura 4: Detector ATLAS. Extraído de (www.atlas.ch).

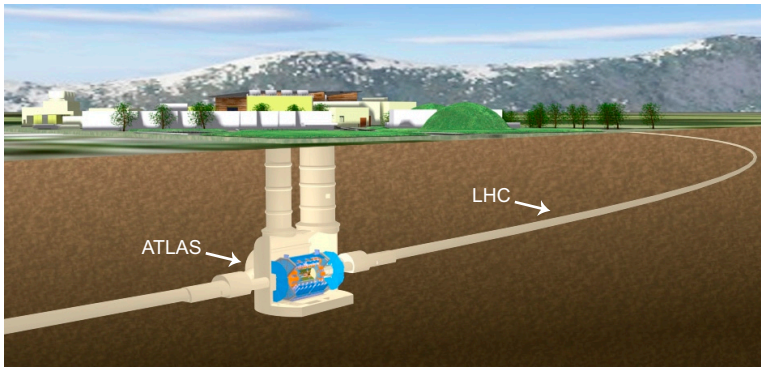

Figura 5: Vista subterrânea do detector ATLAS . Extraído de (www.atlas.ch).

toróides. Por fim, na parte mais externa, as câmaras de múons.

Como visto, o ATLAS tem formato cilíndrico, e para a identificação da posição das partículas no detector utiliza-se o sistema de coordenadas (AAD et al., 2008), apresentado na Figura 6. Geralmente, alguns estudos dos grupos de performance do ATLAS são encontrados em função de η (ou, pseudo-rapidez) e ϕ.

2.4.1 DETECTOR INTERNO

O Detector Interno, (do inglês, *Inner Detector*) (ID) (COLLABORATION; RYAN et al.,) é constituído por três tipos de detectores, como mostra a Figura 7:

- Detector de Pixels (*Silicon Pixel Detector* (SPD));
- Detector de Traços baseado em semicondutores (*SemiConductor Tracker* (SCT));

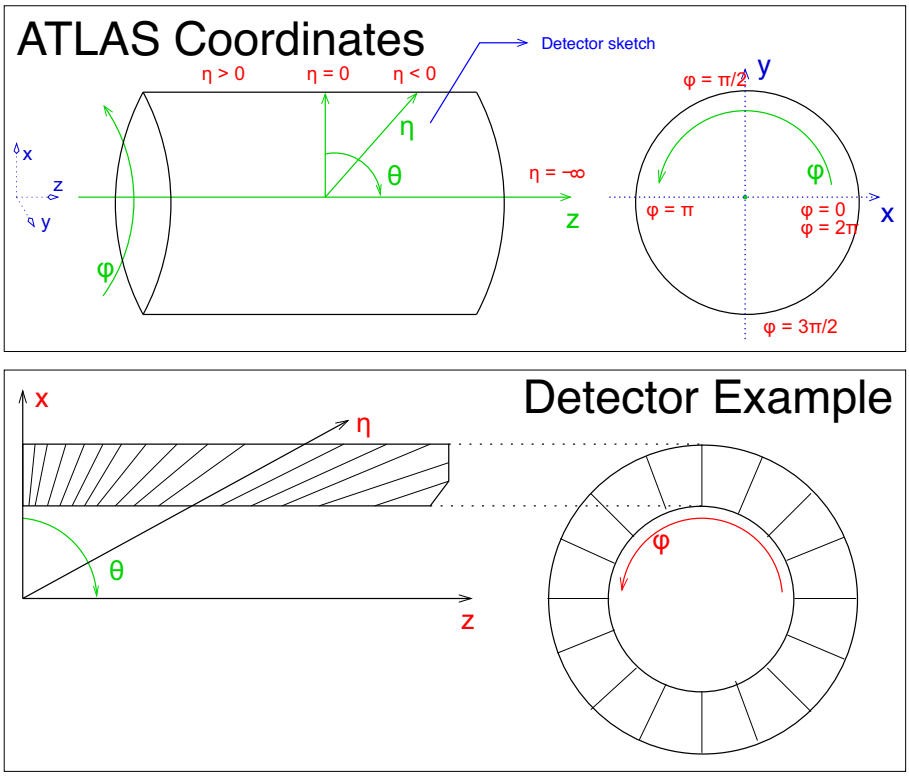

Figura 6: Sistema de coordenadas do detector ATLAS. Extraído de (ANJOS, 2006).

- Detector de Radiação de Transição (*Transition Radiation Tracker* (TRT)).

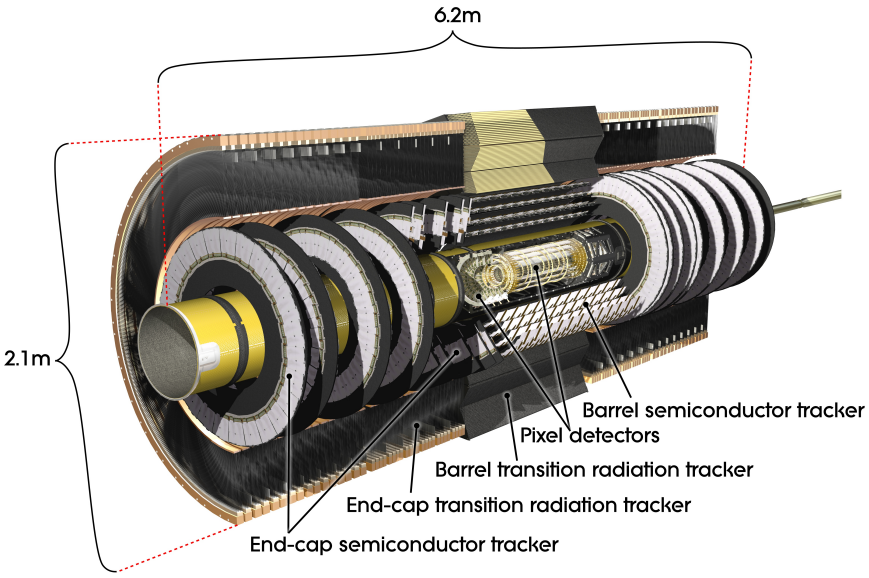

Figura 7: Detector Interno. Extraído de (cds.cern.ch).

Esses detectores permitem a medição da trajetória de partículas carregadas e estão contidos em um solenoide central, que fornece um campo nominal de 2T.

O detector de Pixels, que é feito de silício, contribui, principalmente, para a medição

precisa dos vértices. O SCT mede com precisão o momento das partículas. O TRT facilita o reconhecimento de padrões, auxiliando na identificação de elétrons.

Na Tabela 1, temos uma visão geral sobre a posição, cobertura em η, número de camadas, *hits* e resolução de todos os detectores do ID.

Tabela 1: Parâmetros principais do ID. Extraído de (PEETERS, 2003)

System	Position	η-coverage	Layers	Hits	Resolution (μm)
Pixel	B-layer	$\pm$ 2.5	1	1	Rϕ = 12, z = 66
	barrel layers	$\pm$ 1.7	2	2	Rϕ = 12, z = 66
	end-cap discs	1.7 - 2.5	3	3	Rϕ = 12, R = 77
SCT	barrel layers	$\pm$ 1.4	4	4	Rϕ = 23, z = 580
	end-cap discs	1.4 - 2.5	9	4	Rϕ = 20-26, R = 580
TRT	barrel straws (axial)	$\pm$ 0.7	73	36	Rϕ = 170
	end-cap straws (radial)	0.7 - 2.5	224	36	Rϕ = 170

A Figura 8 mostra um corte transversal do ID.

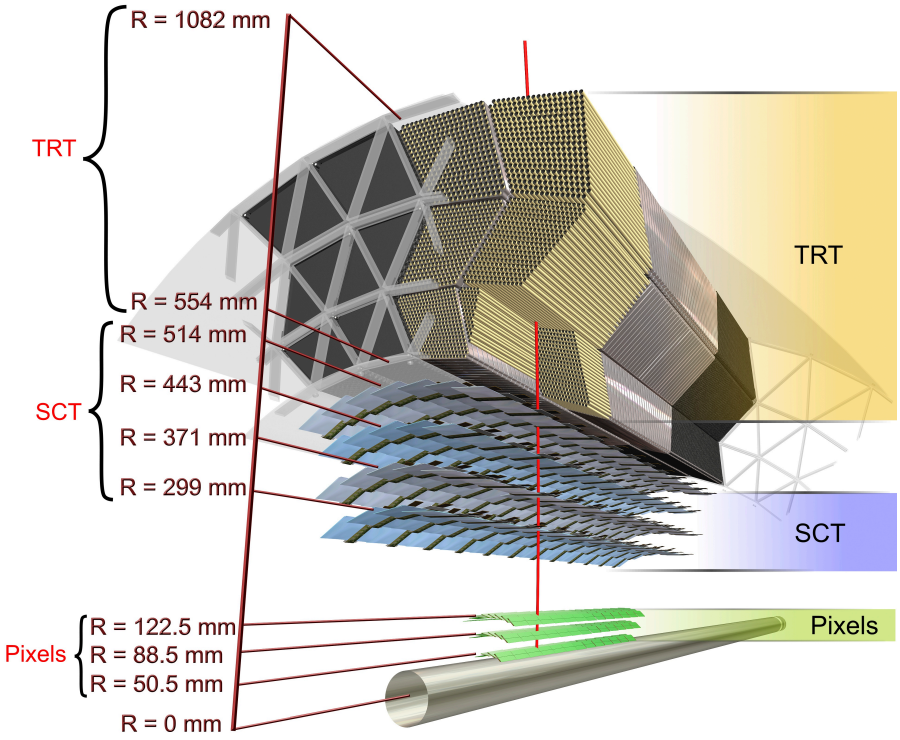

Figura 8: Detector Interno - corte transversal. Extraído de (cds.cern.ch).

2.4.1.1 DETECTOR DE PIXELS

O detector de pixels (AAD et al., 2008) tem um comprimento de, aproximadamente, 1,3 metros. Conforme é mostrado na Figura 8, esse detector possui três camadas no barril, sendo que uma delas está em volta do tubo de feixe, com raio de 50,5 mm. As outras duas possuem raio de 88,5 mm e 122,5 mm, respectivamente. Em cada extremidade do barril possui três discos, mostrados na Figura 9. Esta disposição proporciona três detecções na trajetória da partícula em $|\eta| < 2{,}5$. Como o ID fica próximo ao tubo de feixe, é necessário que seja bastante resistente à radiação.

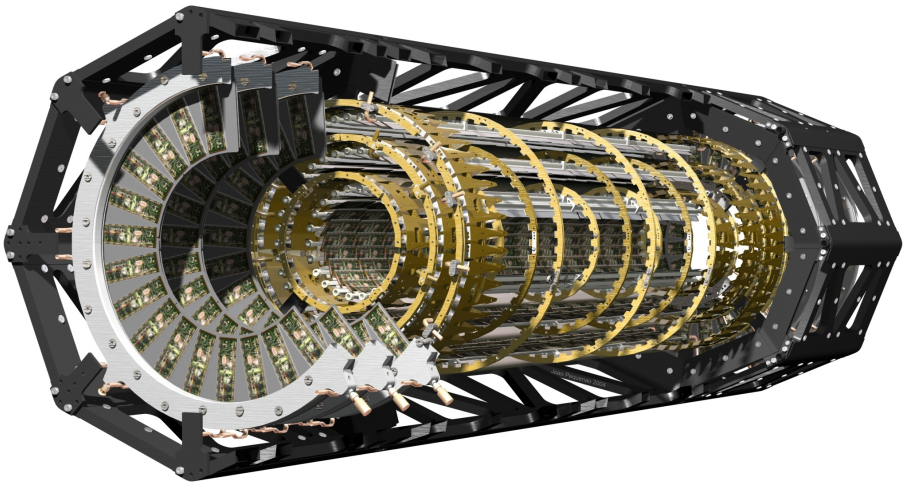

Figura 9: Detector de Pixels do ID. Extraído de (cds.cern.ch).

Esse detector possui, aproximadamente, 1700 módulos idênticos, composto de sensores e chips de leitura, conexo a 80 milhões de pixels. Essa fina granularidade possibilita grande acuidade na identificação do início da trajetória de partículas, dando capacidade ao ID de encontrar partículas de vida curta.

Em 2014, foi instalada uma camada adicional no interior da primeira camada do detector de pixels, denominada *Insertable B-Layer* (IBL). Essa camada contribuirá no rastreamento da trajetória e na posição do vértice das partículas, apesar dos efeitos da radiação, vida útil do hardware e luminosidade nesta região.

2.4.1.2 DETECTOR DE TRAÇOS BASEADO EM SEMICONDUTORES

O SCT (COLLABORATION, 2014) possui 5,6 m de comprimento e 1 m de diâmetro. É composto por 4.088 módulos de detectores de silício e 6 milhões de canais individuais de leitura dispostos em barril (Figura 10) e tampa.

O barril possui quatro camadas e cada tampa possui nove discos. Desta forma, o SCT fornece quatro pontos de precisão por trajetória na região do barril, sendo projetado para fornecer oito medições de precisão por trajetória. As medidas do SCT contribuem na medição do momento, parâmetro de impacto e vértice da partícula.

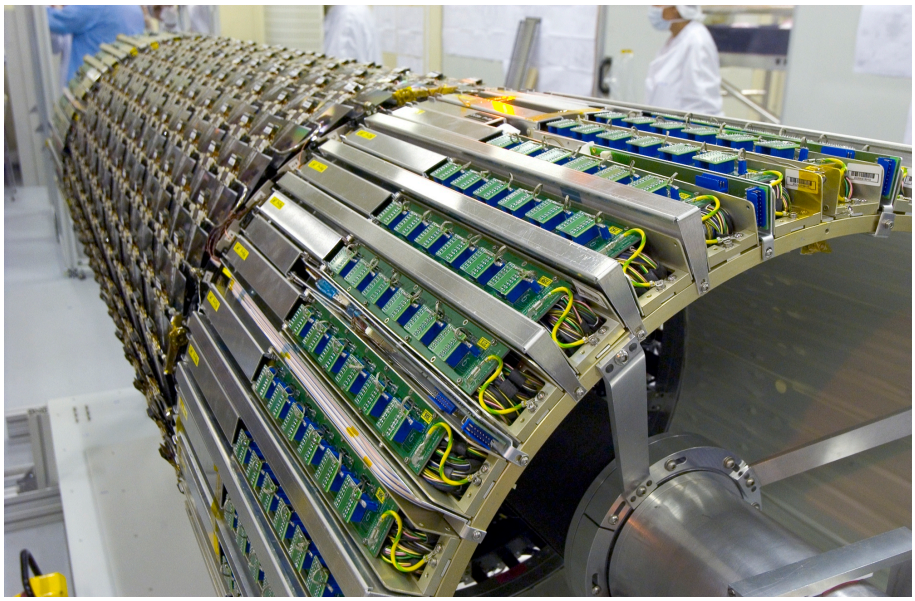

Figura 10: Foto do barril do SCT. Extraído de (cds.cern.ch)

2.4.1.3 DETECTOR DE RADIAÇÃO DE TRANSIÇÃO

O TRT (ABAT et al., 2008) é o componente mais externo do ID e também é dividido em barril (Figura 11) e tampa. Possui 6,8 m de comprimento e 2,2 m de diâmetro. Este detector baseia-se na utilização de detectores de traços em microtubos (do inglês Straw Tube Tracker). O barril e a tampa contêm 50 mil e 320 mil microtubos, respectivamente, que podem operar em altas taxas, devido ao seu pequeno diâmetro e isolamento do fio condutor por um gás individual.

No TRT, o aumento da capacidade de identificação de elétrons ocorre devido ao Gás Xenônio, utilizado para detectar radiações de transição de fótons.

2.4.2 CALORIMETRIA

Conceitualmente, em física de partículas, calorímetros são blocos de matéria com espessura suficiente para absorver completamente a energia de uma partícula, onde parte dessa energia é convertida em calor, daí o termo calorimetria. Em aceleradores modernos, calorímetros formam o coração e a alma do experimento (WIGMANS, 2000).

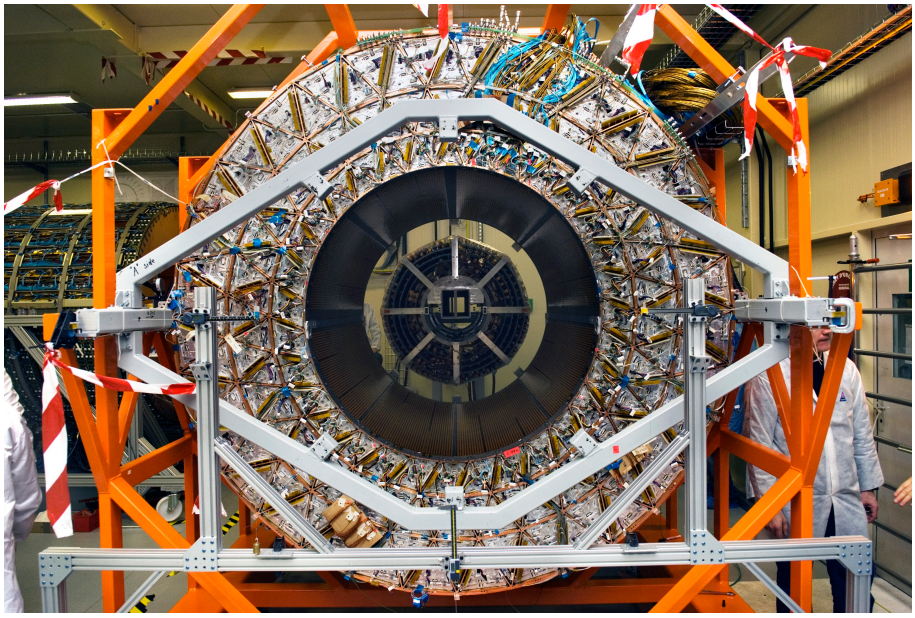

Figura 11: Foto do barril do TRT. Extraído de (cds.cern.ch)

Esses instrumentos geram informações, obtidas a partir de sinais elétricos, importantes na identificação de partículas.

Ao atravessar a matéria, durante o processo, uma partícula incidente interage e perde parte de sua energia. Esse processo de interação depende da energia e da natureza de tal partícula. Essas interações são um processo em cascata (ou chuveiro) em que, um número muito grande de partículas secundárias é produzido ao longo do calorímetro. Uma fração da energia, da partícula incidente, é depositada na forma de luz de cintilação ou Čerenkov, cuja intensidade é proporcional à energia incidente, produzindo um sinal elétrico detectável.

As características gerais dos calorímetros são:

- sensibilidade tanto para partículas neutras como para carregadas;
- a resposta é diferente para elétrons, múons e hádrons, o que permite sua utilização para identificação de partículas;
- o tempo de resposta é rápido, o que os torna adequados para seleção *online* de eventos em regimes de altas taxas;
- a segmentação permite medir a posição e o ângulo de incidência da partícula.

2.4.2.1 CHUVEIROS

A interação das partículas com o calorímetro é um processo em cascata. Existem dois tipos de cascatas (ou chuveiros) (WIGMANS, 2000): os iniciados por elétrons e fótons, chamados de chuveiros eletromagnéticos; e os iniciados por hádrons, chamados chuveiros hadrônicos. Cada um deles possui características peculiares, que determinam o *design* dos calorímetros.

CHUVEIROS ELETROMAGNÉTICOS E HADRÔNICOS Elétrons e fótons com alta energia, ao passar por um material absorvedor, dão início a um chuveiro eletromagnético, como ilustrado na Figura 12. Partículas carregadas podem sofrer diversos tipos de interações, criando fótons, que se convertem em pares elétron-pósitron. Havendo energia suficiente, existe a probabilidade desse processo produzir dois novos fótons, que produzirão outros pares elétron-pósitron, multiplicando o número de partículas. Porém, com o desenvolvimento do chuveiro, a média da energia do chuveiro de partículas diminui, até que, em algum ponto essa multiplicação se encerra.

Os chuveiros hadrônicos, como mostra a Figura 13), são governados pelo comprimento de interação nuclear e são em geral, muito maiores que os chuveiros eletromagnéticos. Por essa razão, os calorímetros hadrônicos são muito maiores que os eletromagnéticos, e não são apenas mais extensos, mas também, mais largos.

Enquanto no chuveiro eletromagnético o desenvolvimento lateral é ditado pelo espalhamento Coulombiano múltiplo, nos chuveiros hadrônicos o desenvolvimento lateral é causado pela grande transferência de momento típica de interações nucleares. Enquanto o chuveiro eletromagnético é composto por elétrons e pósitrons produzidos por dissociação de fótons, e por fótons originados de *Bremsstrahlung*, os chuveiros hadrônicos são compostos, basicamente, por píons.

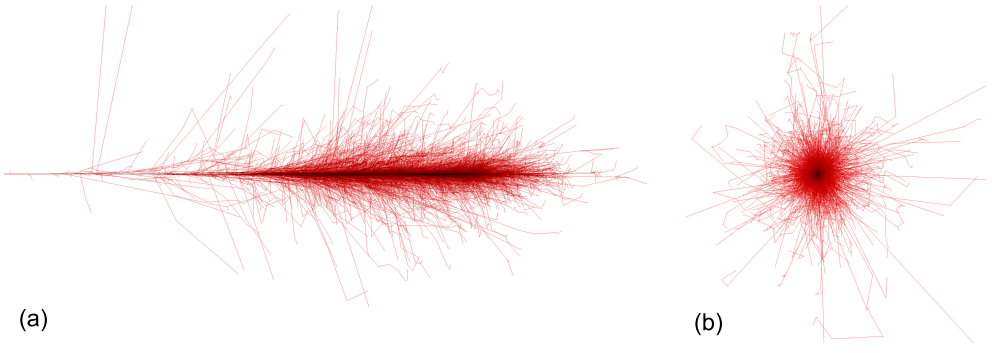

Figura 12: Simulação computacional utilizando algoritmo Corsika do Chuveiro Eletromagnético (100GeV), (a) vista lateral e (b) vista frontal.

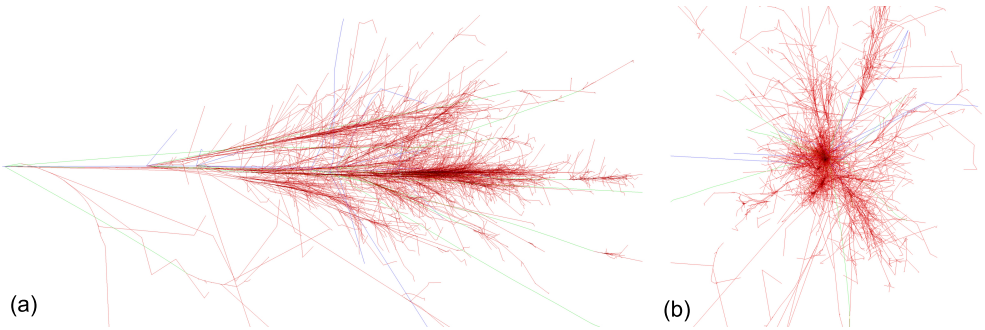

Figura 13: Simulação computacional utilizando algoritmo Corsika do Chuveiro Hadrônico (100GeV), (a) vista lateral e (b) vista frontal.

2.4.2.2 CALORÍMETRO ELETROMAGNÉTICO

O Calorímetro Eletromagnético, (do inglês, *Eletromagnetic Calorimeter*) (EM) (CALORIMETER et al., 2008) é a parte mais interna do sistema de calorimetria do ATLAS e cobre a região de $|\eta| < 3,2$. Ele é dividido em três camadas, com segmentações distintas (veja a Figura 14). A primeira camada possui a segmentação mais fina, permitindo a localização precisa da partícula; a segunda camada é a mais profunda e a terceira é a menos segmentada, com o intuito de absorver toda energia de elétrons e fótons incidentes.

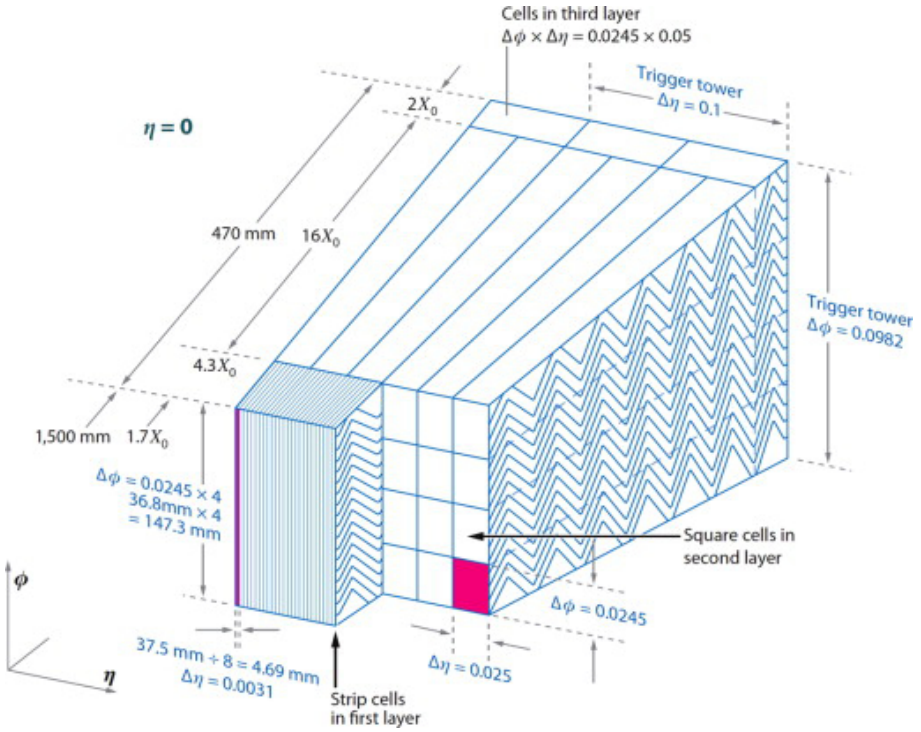

Figura 14: Segmentação do Calorímetro Eletromagnético. Extraído de (FRANCAVILLA; COLLABORATION et al., 2012).

O calorímetro EM também possui um pré-irradiador (do inglês, *pre-sampler*), que

funciona como um calorímetro muito fino, posicionado na parte mais interna do calorímetro e que tem, como função, recuperar a informação perdida no material morto da seção EM (ou seja, fios, encapamentos, etc).

O calorímetro Eletromagnético é baseado em absorvedores de chumbo e utiliza Argônio líquido como material ativo. As placas de chumbo estão imersas em um tanque de argônio líquido, sujeitas a um forte campo elétrico, e são cobertas por finos eletrodos de cobre.

Quando o chuveiro eletromagnético chega ao Argônio, elétrons são arrancados dos átomos de Argônio. Esses elétrons livres, sujeitos a um forte campo elétrico, migram rapidamente para o lado positivo do campo, fazendo com que os íons migrem para o lado negativo. Esse processo gera uma corrente elétrica detectável em um circuito externo conectado ao calorímetro.

2.4.2.3 CALORÍMETRO HADRÔNICO

O barril e tampas (Figura 15) do Calorímetro Hadrônico, (do inglês, *Hadronic Calorimeter*) (HAD) (AAD et al., 2010) são divididos em três camadas de diferentes segmentações (Figura 16).

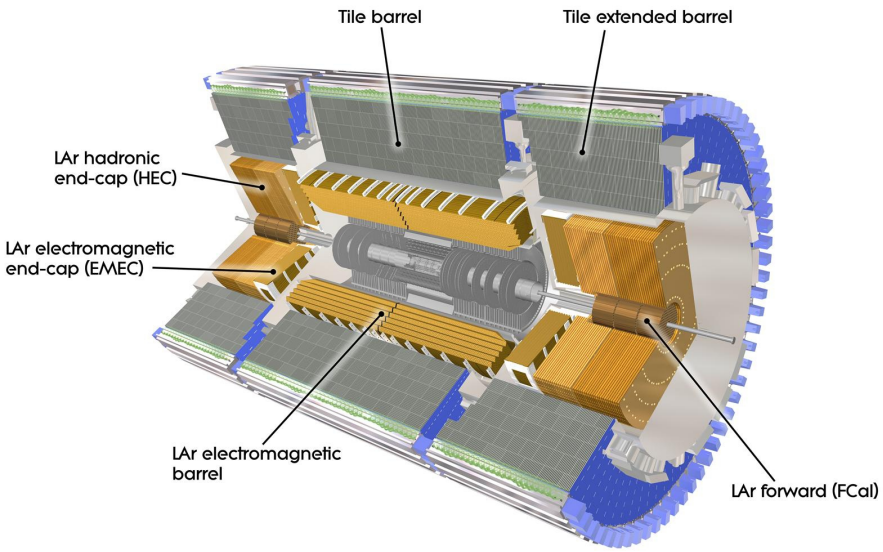

Figura 15: Modelo computacional do HAD e do EM. Extraído de (cds.cern.ch)

Este calorímetro é composto por módulos que possuem, em sua construção, placas de cintiladores alternadas com placas de aço. Os cintiladores, ao serem "excitados"

por partículas do chuveiro hadrônico produzem luz, que é conduzida por fibras óticas não cintilantes até cada Tubo Fotomultiplicador, (do inglês, *Photomultiplier tubes*) (PMT). Os Tubos Fotomultiplicadores, por um efeito em cascata, multiplicam os elétrons arrancados de seus dinodos, gerando então um sinal elétrico detectável.

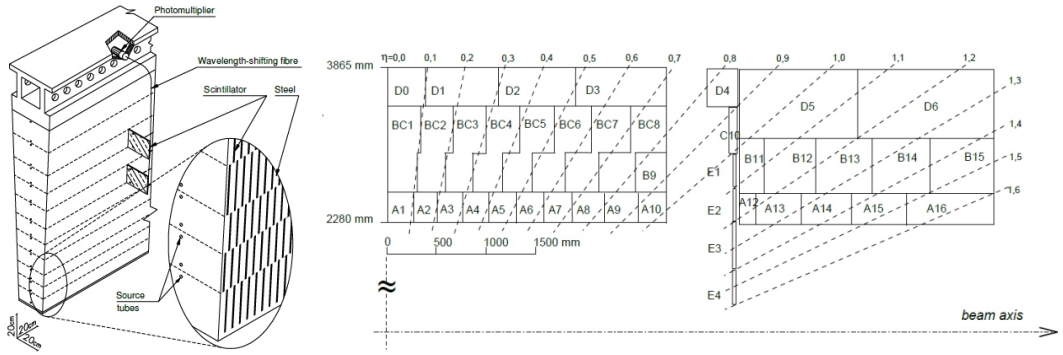

Figura 16: Segmentação do Calorímetro Hadrônico. Extraído de (cds.cern.ch)

2.4.3 CÂMARA DE MÚONS

A câmara de Múons é constituída por milhares de sensores para partículas carregadas, colocados em um campo magnético produzido por grandes bobinas toroidais supercondutoras. Os sensores são semelhantes aos descritos no TRT do ID, mas com os diâmetros dos tubos maiores.

Múons são partículas como os elétrons, mas, 200 vezes mais pesadas. Múons são as únicas partículas detectáveis que podem atravessar todos os absorvedores dos calorímetros. O espectrômetro de múons (COLLABORATION et al., 2010) envolve os calorímetros e mede trajetos dos múons para determinar os seus momentos com alta precisão, como visto na Figura 17 na parte externa do detector.

2.4.4 PERFIL DOS EVENTOS NO ATLAS

Após conhecermos os principais detectores do ATLAS é possível entender o perfil de alguns eventos neste detector. Partículas carregadas, como múons, prótons e elétrons, deixam sinal no detector de traço. O fóton e o elétron são completamente absorvidos no calorímetro eletromagnético. O próton deixa sinal no calorímetro eletromagnético e no hadrônico. O nêutron, deixa sinal apenas no calorímetro hadrônico e não deflete no campo eletromagnético no detector. Os múons deixam sinal em todo detector. Estas são características distintas que possibilitam perceber diferentes assinaturas para cada partícula. Essas assinaturas são mostradas na Figura 18.

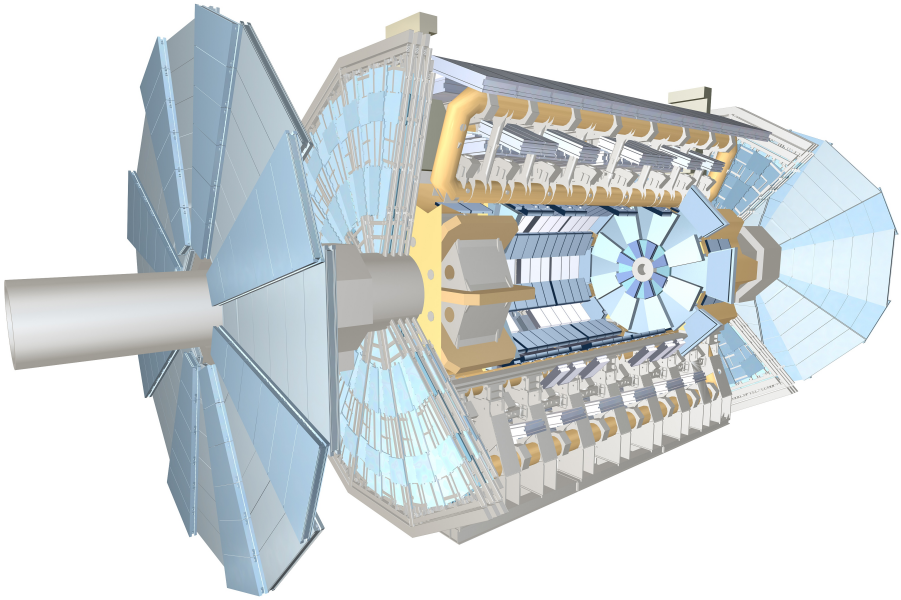

Figura 17: Câmara de Múons do detector ATLAS. Extraído de (cds.cern.ch).

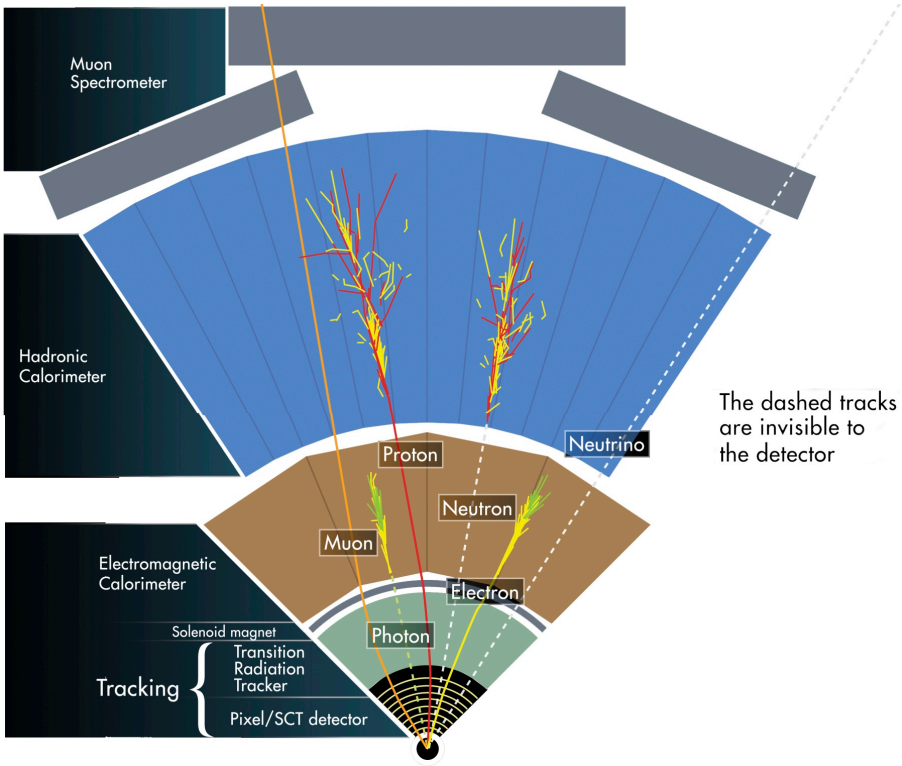

Figura 18: Assinatura das partículas no detector ATLAS. Extraído de (cds.cern.ch).

2.4.5 SISTEMA DE FILTRAGEM DO ATLAS

Como o detector ATLAS gera informações na ordem de 60 Tbytes por segundo, sendo grande parte dos eventos descartáveis (WATTS, 2003), é essencial que se tenha um

sistema de filtragem (*trigger*) capaz de selecionar "durante" (*online*) a colisão os eventos relevantes, e um sistema para análise permanente (*offline*), que através de algoritmos mais complexos, pode identificar, de uma forma mais criteriosa, os processos físicos de interesse.

2.4.5.1 *FILTRAGEM* ONLINE

O Sistema de Filtragem *Online* é baseado em níveis hierárquicos em cascata, onde o evento rejeitado em cada nível anterior não é avaliado pelo nível posterior. É interessante notar que, como um nível anterior está exposto a uma taxa muito maior de eventos do que um nível posterior, sua complexidade de análise aumenta a cada nível "vencido" pelo evento.

O sistema *online* pode ser dividido em *Level 1* (L1) (ACHENBACH et al.,) e Filtragem de Alto nível, (do inglês, *High Level Trigger*) (HLT) (TORRES, 2010), onde a Filtragem de Alto Nível é subdividida em *Level 2* (L2) e Filtro de Eventos, (do inglês, *Event Filter*) (EF). A Figura 19 mostra o diagrama de blocos do Sistema de Filtragem *Online* do ATLAS:

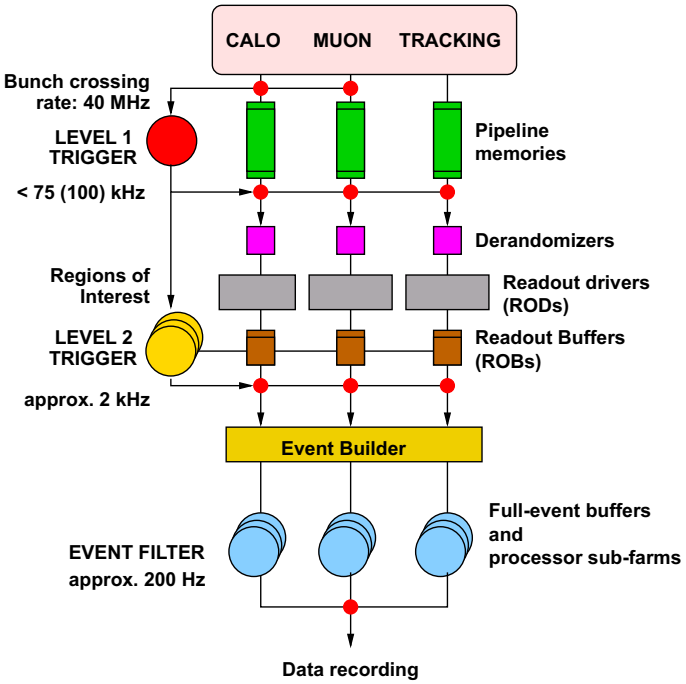

Figura 19: Fluxograma do sistema de Trigger Online do ATLAS. Extraído de (ANJOS, 2006).

O L1 é responsável pela seleção inicial de eventos e utiliza informação dos calorímetros e detectores de múons. Nessa etapa, a decisão precisa ser bem rápida. Portanto, a

quantidade de informação a ser processada precisa ser bastante reduzida, somando-se células dos detectores para reduzir a granularidade utilizada. Esse nível de filtragem (L1) é feito em *hardware*, devido ao tempo de latência ser muito curto ($\sim 2\mu$s) e a taxa de eventos ser bastante elevada. Nesse nível, apenas são descartados os eventos com características bem distintas dos canais de interesse.

O HLT é formado pelo L2 e EF, e tem um tempo conjunto de latência ($\sim$40ms), sendo implementado em *Software*. Utilizando uma granularidade mais fina, o L2 seleciona os eventos que serão avaliados pelo EF e por fim, teremos uma taxa de saída do sistema de filtragem *online* de, aproximadamente, 200Hz. Os eventos selecionados serão armazenados em mídias permanentes para uma análise futura.

Essa é uma visão geral da filtragem *online* do ATLAS, na Seção 3.1 veremos como acontece a filtragem especificamente para o elétron.

2.4.5.2 FILTRAGEM OFFLINE

Como explicado anteriormente, para um sistema *online* os eventos que foram rejeitados não são utilizados. Por isso, na filtragem *online* a eficiência deve ser elevada, considerando-se que não é desejado perder nenhum evento físico relevante para análise posterior. Como consequência, ao final do processo *online* de filtragem, uma considerável quantidade de ruído de fundo ainda é encontrado nos canais de interesse, sendo esses dados, armazenados, responsáveis por suprir os diferentes estudos vigentes no ATLAS.

Com o intuito de viabilizar esses estudos, um sistema de filtragem *offline* é empregado. Neste sistema, como o tempo de latência não é um fator determinante, algoritmos bastante complexos podem ser empregados. Nesta etapa, pode-se equilibrar melhor a eficiência e rejeição de ruido de fundo, visando uma melhor reconstrução do perfil dos eventos.

3 IDENTIFICAÇÃO DE ELÉTRONS

A identificação de elétrons é de fundamental importância para o programa de física do ATLAS. Léptons são as principais assinaturas dos processos eletrofracos (ALISON, 2014), sendo utilizados em uma vasta gama de análises físicas, compreendendo desde as medições de precisão do modelo padrão até a busca de uma nova física, além do modelo padrão.

Muitos aspectos da concepção global do detector ATLAS foram impulsionados pela exigência de que os elétrons fossem bem reconstruídos e identificados de forma eficiente.

3.1 RECONSTRUÇÃO DE ELÉTRONS

A Reconstrução dos elétrons no ATLAS segue os seguintes passos(AAD et al., 2012a):

1. Encontrar um conjunto de células (do inglês *cluster seed*) no calorímetro com energia acima de 2,5 GeV através de um algoritmo de janela móvel;

    - *Cluster seed* de tamanho 3x5 é procurado na segunda camada do calorímetro eletromagnético (granularidade 0,025 x 0,025 em η x ϕ). Na Figura 14 é possível perceber a segmentação na segunda camada do EM, contextualizando o Cluster seed.

2. Combinar o *cluster seed* com o traço do ID;

    - Para ser um candidato a elétron, é necessária a existência de pelo menos um traço dentro de $\Delta\eta < 0{,}05$ e $\Delta\phi < 0{,}1(0{,}05)$ do *cluster seed*. A variação em ϕ é maior devido as perdas *Bremsstrahlung* no ID. Começando nos dados obtidos em 2012, um algoritmo chamado *Gaussian Sum Filter* (GSF) (COLLABORATION et al., 2012) foi utilizado para melhorar a estimativa dos parâmetros de traço quando a radiação de *Bremsstrahlung* ocorre. O aumento da eficiência na reconstrução dos dados de 2012 em relação aos anteriores é devido ao GSF, e pode ser visto na Figura 20.

3. Reconstruir o *cluster* com tamanho otimizado;

   - $\Delta\eta$ x $\Delta\phi$ = 3x7 (5x5) barril (tampa).

4. Computar a energia medida;

   - O total de energia é determinado pela adição de quatro componentes distintos: a energia medida no *cluster*, a energia estimada do que foi perdido antes do *cluster*, a energia estimada para a perda na lateral do *cluster*, a energia estimada para a perda longitudinal atrás do *cluster*. Esses componentes são parametrizados em função das energias medidas em diferentes camadas longitudinais do calorímetro eletromagnético. Essas parametrizações são determinadas a partir de dados simulados de Monte Carlo e corrigidas nos dados reais, baseados em elétrons de decaimentos $Z \rightarrow$ ee (AAD et al., 2012a).

Os candidatos a elétrons, como mostrado na Figura 20, que chegam neste estágio são chamados de "*reconstructed electrons* " ou "*container electrons*", e a eficiência de reconstrução, para os elétrons que passam pela requisitos de *cluster* e de traço é alta.

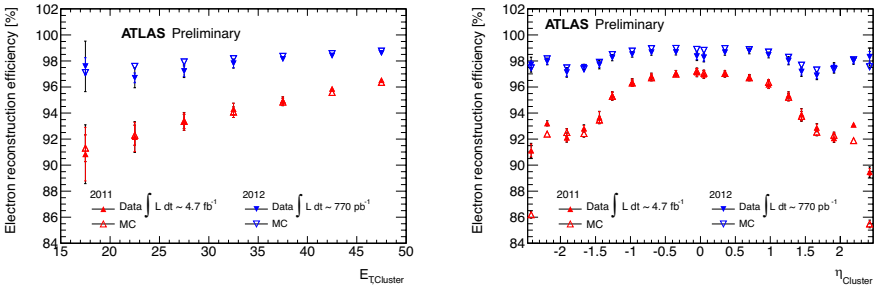

Figura 20: Eventos de elétrons reconstruídos a partir de candidatos a decaimento de W. Extraído de (ALISON, 2014).

Embora a eficiência na reconstrução dos elétrons seja grande, sua pureza é baixa. Elétrons reconstruídos sofrem com o elevado número de ruído de fundo, causado, principalmente, por três fontes: hádrons identificados erroneamente, conversões de fótons e decaimentos semi-leptônicos (*heavy-flavor*).

Nos casos de conversões de fótons e decaimentos semi-leptônicos, um elétron real está presente no estado final. Esses elétrons são considerados ruídos de fundo, pois eles não são produzidos isoladamente. Na sequência, ambos os hádrons identificados erroneamente como elétrons e os elétrons de fontes que não são de uma partícula de interesse serão considerados ruídos de fundo. Elétrons produzidos por decaimento de bósons como W ou Z, serão referidos como "reais", "verdadeiros", "sinal" ou *prompt electrons*.

3.1.1 FILTRAGEM DE ELÉTRONS

Uma das motivações para a utilização dos elétrons em aplicações de estudos físicos é o seu característico sinal de *trigger*. No L1, eventos são selecionados como possíveis elétrons quando sua energia excede um certo limiar, esse limiar varia em função de η. Devido a alta luminosidade instantânea, um veto hadrônico é aplicado em muitos *triggers* do L1, que requer que a energia no calorímetro hadrônico seja menor do que um determinado limiar. Cada *trigger* do L1 define uma região de interesse para a reconstrução de elétrons no HLT. Rápidos algoritmos dedicados de reconstrução do calorímetro são executados no L1.

Os algoritmos do L2 são similares aos utilizados na filtragem *offline*; um limiar mais refinado de energia pode ser definido, e variáveis discriminantes são utilizadas para reduzir a taxa de *trigger* do L2 a um nível aceitável. O EF usa algoritmos *offline* de reconstrução e identificação, com ligeiras diferenças nas configurações, para aplicar uma seleção final no *trigger* dos elétrons.

Existem dois tipos básicos de *trigger*: primário e de suporte. O *trigger* primário é utilizado principalmente para coletar eventos de sinal em análises usando elétrons. Eles aplicam critérios rígidos de identificação (como veremos na Seção 3.2) de partículas para reduzir a taxa de dados a um nível aceitável. *Triggers* primários são essencialmente utilizados em análises físicas que têm um elétron no estado final. Uma fração significativa da largura de banda de *trigger* do ATLAS é reservado para este tipo de *trigger*.

Outra classe crucial de *trigger* é a de suporte. Seu objetivo é coletar amostras de elétrons não polarizadas. No *trigger* primário, os elétrons passam por vários critérios de identificação, já no de suporte o critério é basicamente E_t, sem nenhum critério de identificação; e são referidos como *"et-cuts" triggers*. Eles são utilizados na construção da Função Densidade de Probabilidade(do inglês, *Probability density function*) (PDF) do ruído de fundo, para otimizar a identificação dos elétrons e outros estudos.

Existem temas que não foram abordados nesse capítulo, como a reconstrução de elétrons *forward* e elétrons com E_t abaixo de 5 GeV. Todas as análises futuras faremos cortes em $E_t > 5$ GeV e utilizaremos uma variável, do conjunto de dados, para eliminar elétrons *foward* (el_autor = 1 | 3). Esses cortes foram baseados no artigo (COLLABORATION et al., 2013) da colaboração.

3.2 VARIÁVEIS DISCRIMINANTES PARA IDENTIFICAÇÃO DE ELÉTRONS

Para a utilização de elétrons em análises *offline* alguns critérios adicionais de identificação são aplicados, com o intuito de aumentar a pureza dos elétrons reconstruídos. Esses critérios fornecem uma identificação de elétrons mais eficiente, com grande rejeição de ruído de fundo, e são fornecidas através de variáveis discriminantes que utilizam informações do ID e dos calorímetros (COLLABORATION et al., 2011).

3.2.1 VARIÁVEIS DE CALORIMETRIA

As distribuições das variáveis discriminantes, provenientes da calorimetria, podem ser observadas na Figura 21. Essas variáveis, geralmente, exploram a fina segmentação lateral e longitudinal dos calorímetros do ATLAS. Na figura percebemos as distribuições para: Elétrons verdadeiros, chamados "*Isolated Electrons*"; hádrons; conversões, chamados "*Background Electrons*" e decaimentos semi-leptônicos, chamados de "*Non-Isolated Electrons*" .

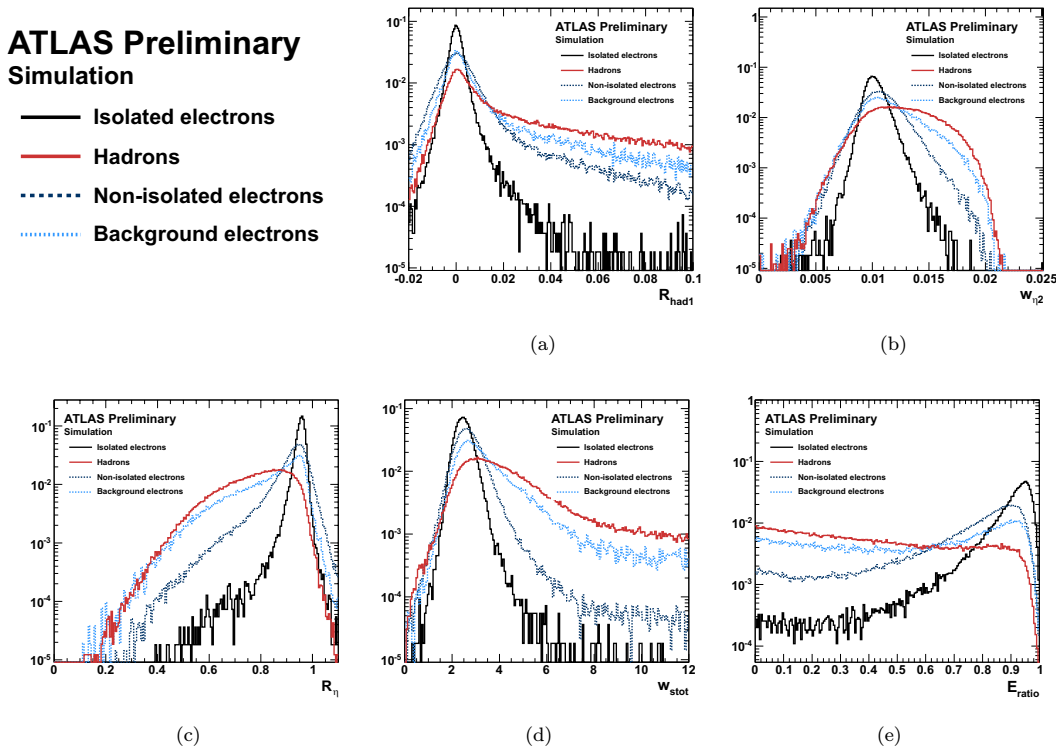

Figura 21: Variáveis de identificação de elétrons no calorímetro, "formato do chuveiro", apresentados separadamente para sinal e os vários tipos de ruído de fundo. As variáveis apresentadas são: (a) vazamento hadrônico R_{had}, (b) $W_{\eta 2}$, (c) R_{η}, (d) $W_{s,tot}$ e (e) E_{ratio}. Extraído de (ALISON, 2014).

A Figura 21.a mostra a variável de vazamento hadrônico, *Rhad1*. Essa variável é definida como a razão entre a primeira camada do calorímetro hadrônico, atrás do *cluster* de elétrons, sobre a energia do *cluster* de elétrons. Elétrons reais depositam a maior parte de sua energia no EM antes de atingir o HAD, apresentando pequenos valores de *Rhad1*. Grandes vazamentos hadrônicos indicam atividade hadrônica associada ao *cluster* de elétrons. Na região de $0{,}8 < |\eta| < 1{,}37$ o espalhamento hadrônico é calculado com todas as camadas do HAD, já em outras regiões de η a primeira camada é suficiente.

A Figura 21.b mostra a variável $\mathrm{W}_{\eta 2}$, que é a medida da largura do chuveiro em η ponderada pela Raiz do Valor Quadrático Médio, (do inglês, *Root Mean Square*) (RMS) da distribuição em η na segunda camada do EM. É definida como:

$$W_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i}\right)^2} \tag{3.1}$$

Onde $\mathrm{E}_i\eta_i$ é a energia(η) da i-ésima célula, em uma janela 3x5 da segunda camada do EM, centrada no elétron reconstruído. Essa variável exige largura estreita de chuveiro em η, suprimindo ruído de fundo de jatos e conversões de fótons, que tendem a ter chuveiros maiores do que elétrons verdadeiros.

Outra medida de largura do chuveiro está em R_η, mostrado na Figura 21.c. R_η é definida como a razão entre energia de uma janela 3x7 sobre uma janela 7x7, na segunda camada do EM. Ruídos de fundo tendem a ter uma maior fração de energia fora do núcleo 3x7, resultando em baixos valores de R_η, que é uma das mais poderosas variáveis para separação de ruído de fundo.

A largura do chuveiro na primeira camada do calorímetro é mostrada na Figura 21.d. Essa variável é chamada de $\mathrm{W}_{s,tot}$. Sua definição é:

$$W_{s,tot} = \sqrt{\frac{\sum_i E_i \left(i - i^{\max}\right)}{\sum_i E_i}} \tag{3.2}$$

Onde E_i é a energia na primeira *strip*, i é o índice das *strips*, e $\mathrm{i}^{\max}$ é a *strip* de maior energia. A soma das *strips* é executada ao longo de uma janela de 0,0625 x 0,2 centrada sobre o elétron, isso corresponde a 20 x 2 tiras em η x ϕ. A largura das *strips* é maior para ruído de fundo do que para o sinal.

A variável E_{ratio}, mostrada na Figura 21.e, também é utilizada para diminuir o ruído

Tabela 2: A variação do tamanho das *strips* em função de η. Extraído de (ALISON, 2014)

$\lvert\eta\rvert$-value	Detector Change
0,6	Change in depth of the 1^{st} sampling
0,8	Change in absorber thickness (1,53 mm to 1,13 mm)
1,37	Beginning of Barrel-end-cap transition
1,52	End of Barrel-end-cap transition
1,81	Strips width changes from $\frac{0,025}{8}$ units in $\lvert\eta\rvert$ to $\frac{0,025}{6}$
2,01	Strips width changes from $\frac{0,025}{6}$ units in $\lvert\eta\rvert$ to $\frac{0,025}{4}$
2,37	Strips width changes from $\frac{0,025}{4}$ units in $\lvert\eta\rvert$ to $0,025$
2,47	Strips width changes from $0,025$ units in $\lvert\eta\rvert$ to $0,1$

de fundo. É definida utilizando as células correspondentes as duas maiores energias nas *strips* do EM. A diferença entre a primeira e a segunda maior energia é comparada com sua soma:

$$E_{ratio} = \frac{E^1_{\max} - E^2_{\max}}{E^1_{\max} + E^2_{\max}} \tag{3.3}$$

Ruídos de fundo tendem a ter múltiplas incidências de partículas associadas ao *cluster*. Esses ruídos de fundo terão menores valores de E_{ratio} do que os elétrons verdadeiros, que tendem a deixar energia em um menor número de células.

A fração da energia da terceira camada do EM, chamada f_3, é outra variável adicional do calorímetro, semelhante ao R_{had}, a fração de energia na terceira camada do EM tende a ser menor para elétrons do que pra ruído de fundo, que penetram mais profundamente no calorímetro.

As variáveis discriminantes do ATLAS estão em função de η e E_t (Energia transversa) dos elétrons reconstruídos. A dependência de η é impulsionada pela mudança da geometria dos calorímetros. Por exemplo, na região de transição entre o barril e a tampa, $1,37 < \lvert\eta\rvert < 1,52$, muitas dessas variáveis perdem o seu poder de discriminação devido a perda de resolução. A maioria das análises excluem essa região por causa do *crack*.

Os tamanhos físicos das *strips* também variam com η, levando a uma forte dependência de η. A variação do tamanho das *strips* é dado pela Tabela 2.

A dependência de E_t, para elétrons reais, é devido ao tamanho dos chuveiros. Com o aumento de E_t, a largura do chuveiro tende a diminuir, o ruído de fundo tende a ter uma menor dependência de E_t. O resultado disto é que, a separação das formas do chuveiro com o aumento de E_t melhoram.

3.2.2 VARIÁVEIS DE TRAÇO

O Detector Interno também nos fornece variáveis discriminantes utilizadas na identificação dos elétrons. As variáveis do ID estão em sistemas distintos e são complementares as do calorímetro, como mostrado na Figura 22.

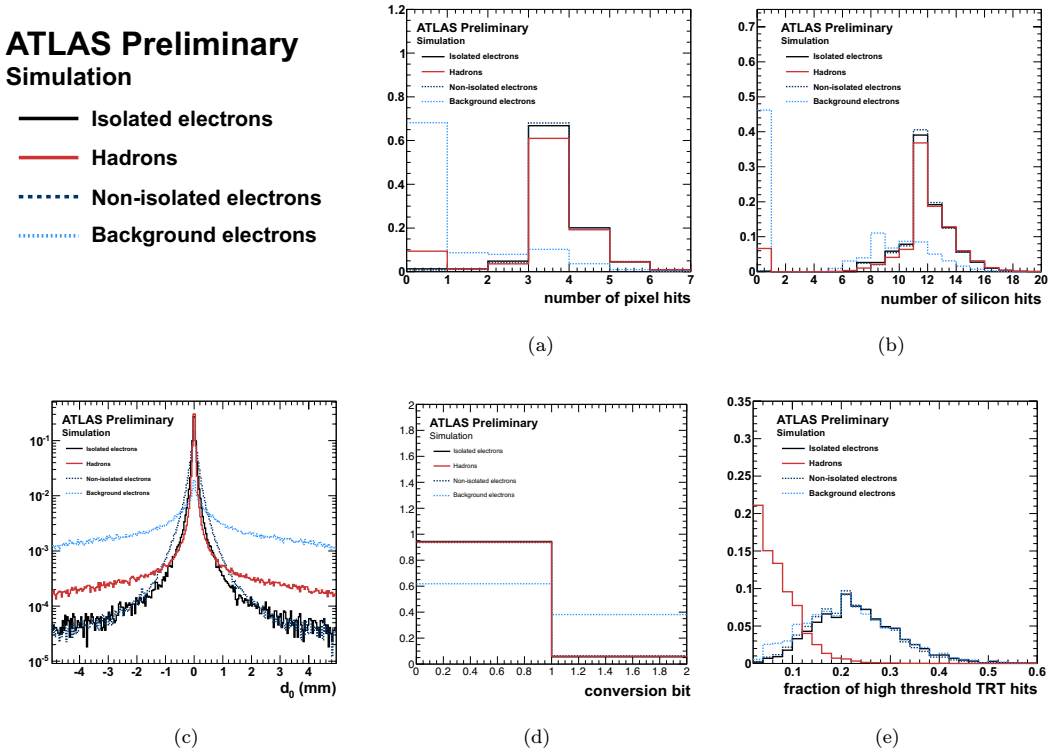

Figura 22: Variáveis de identificação elétron no ID, agrupados em sinal e vários tipos de ruído de fundo. As variáveis apresentadas são: (a) número de *hits* no detector Pixel, (b) número combinado de *hits* do detector de Pixels e SCT, (c) parâmetro de impacto transverso D_0, (d) *flag* de conversão, ou "bit conversão", e (e) fração *hits* de alto *threshold* no TRT. Extraído de (ALISON, 2014).

A Figura 22.a e Figura 22.b mostram o número de detecções ou *hits*, nos detectores de Pixels e SCT associados ao traço do elétron. Ao exigirmos um número mínimo de *hits* no detector de Pixels e SCT, para satisfazer os requisitos de "qualidade do traço", o ruído de fundo pode ser reduzido com pouca perda da eficiência do sinal. As camadas de detectores que fótons atravessam antes de serem convertidos não têm traços associados, resultando em um menor número de *hits* no detector de Pixels e SCT, do que elétrons verdadeiros.

Outra variável importante do Detector Interno é o número de *hits* na primeira camada do detector de Pixels ou *B-layer*. Os requisitos da camada *B-layer* são bastante efetivos na redução de ruído de fundo, pois é sensível a todas as conversões que ocorrem depois da primeira camada do detector de Pixels.

Na Figura 22.c temos a distribuição do parâmetro de impacto transverso, d_0. O parâmetro de impacto mede a distância mais próxima do traço do elétron até o vértice primário. Ele proporciona separação contra as conversões, que podem ter traços deslocados significativamente dos pontos de interação. O d_o é maior em decaimentos semileptônicos (*heavy-flavor*) devido ao longo tempo de vida do b-quark.

O bit de conversão é apresentado na Figura 22.d. Ele é definido se o traço do elétron corresponde a um vértice de conversão. Apesar de reduzir o número de elétrons reconstruídos de conversões, esta variável não é tão eficiente para elétrons verdadeiros.

A Figura 22.e mostra a fração de detecções que passaram no limiar do detector TRT. Isso indica a presença de radiação de transição de fótons. A existência de fótons no TRT fornece rejeição contra hádrons, mas não de conversões e decaimentos-leptônicos, que também possuem elétrons no seu estado final. Essa variável é uma das mais poderosas contra ruído de fundo de hádrons, e tem como característica ser descorrelacionada das variáveis discriminantes dos calorímetros.

Em geral, as variáveis de traço são independentes de η e E_t, com exceção do TRT, que é dependente de η devido a mudança do material utilizado no barril e tampas. Além disso, as variáveis do ID são pouco afetadas pelo *pileup*, ou empilhamento de partículas, devido a rápida leitura dos sistemas de identificação e a fina granularidade do ID.

3.2.3 VARIÁVEIS DE COMBINAÇÃO TRAÇO-CALORIMETRIA

A combinação das variáveis do ID e dos calorímetros nos fornecem outras variáveis discriminantes, que serão mostradas na Figura 23.

A Figura 23.a mostra a diferença entre o η do traço e do *cluster*. A comparação é feita extrapolando o traço até o calorímetro, e esta distribuição tem menor variância para os elétrons reais. As partículas adicionais produzidas pelos hádrons e conversões do ruído de fundo polarizam a posição do *cluster* em relação ao traço correspondente. A exigência de valores pequenos de $\Delta\eta$ reduz o ruído de fundo.

Uma variável semelhante em ϕ é mostrada na Figura 23.b, porém essa correspondência em ϕ é menos poderosa devido aos fótons da radiação de *Bremsstrahlung* causarem uma diferença entre a posição do traço e o *cluster* em ϕ.

Na Figura 23.c temos a variável E/P que é a razão entre a energia medida no calorímetro e o momento determinado no ID. Hádrons não irão depositar toda sua

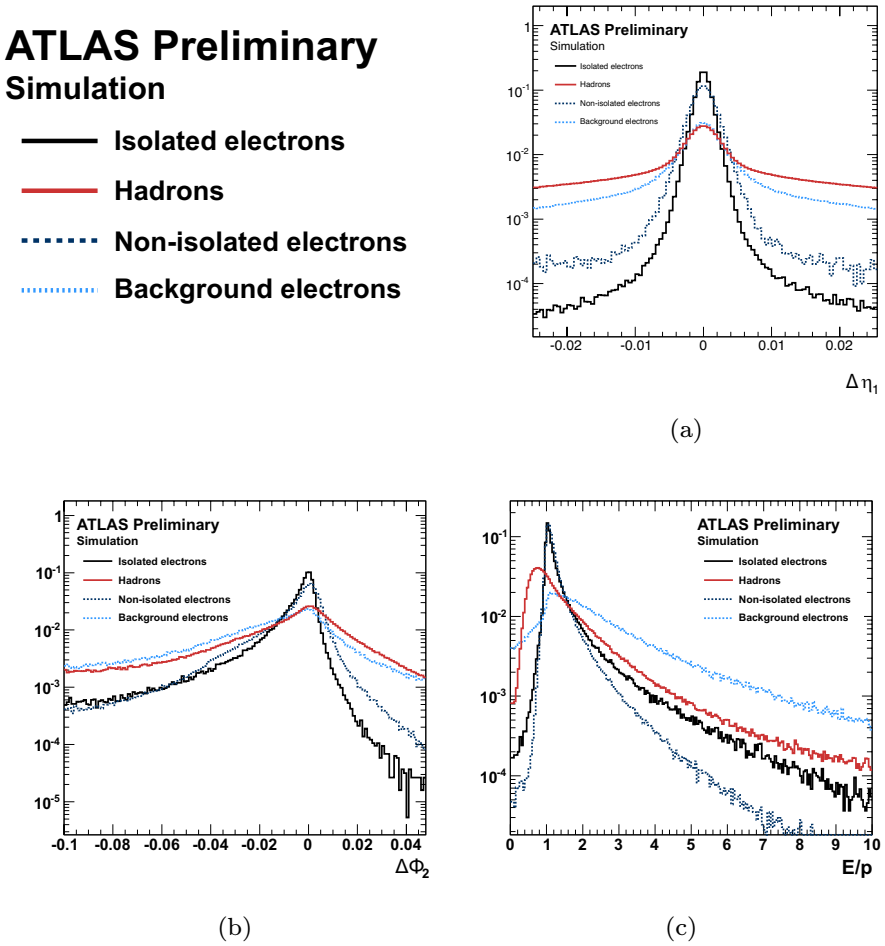

Figura 23: Variáveis combinadas de traço-calorimetria, mostrando a separação de vários tipos de ruído de fundo. As variáveis mostradas são: (a) diferença entre o traço e o *cluster* de energia em η, (b) diferença entre o traço e o *cluster* de energia em ϕ, e (c) razão da energia medida no calorimetro com o momento medido no traço. Extraído de (ALISON, 2014).

energia no EM, uma fração significativa será depositada no HAD. Portanto, a energia do cluster EM não irá refletir a energia total da partícula incidente. Conversões tendem a ter altos valores de E/P.

3.2.4 VARIÁVEIS DE ISOLAMENTO

A última classe de variáveis utilizada para discriminar sinal e ruído de fundo é a de isolamento, conhecidas como: *Etcone* e *Ptcone*. O isolamento mede a quantidade de energia depositada perto dos elétrons reconstruídos. Elétrons do ruído de fundo são produzidos juntamente com outras partículas, levando a valores de isolamento maiores que dos elétrons isolados. O isolamento é calculado pela soma de energia em um cone centrado em um candidato a elétron.

3.3 ALGORITMOS OFFLINE DE IDENTIFICAÇÃO DE ELÉTRONS

Essa seção descreverá os algoritmos utilizados pela colaboração ATLAS na identificação de elétrons. Primeiramente, será abordado o algoritmo padrão chamado e/γ, suas características e atuais configurações. Em seguida, será mostrado o método da verossimilhança, que tem como proposta atingir melhores resultados do que o algoritmo padrão.

3.3.1 ATLAS E/γ

A fim de padronizar a seleção de elétrons usados no *trigger* e em análises físicas, o grupo de performance ATLAS e/γ tem desenvolvido critérios de identificação para selecionar elétrons. Essa seleção de elétrons é um simples corte baseado (*cut-based*) nas variáveis descritas na seção 3.2, sendo referida como "menu isEM" ou "isEM" (AAD et al., 2012b).

A utilização de um mesmo algoritmo na seleção de elétrons, por diversos grupos de pesquisa do ATLAS, tem como vantagem a padronização de um *software*, melhorando a confiabilidade das análises. Este *software* deve ser capaz de aplicar diferentes critérios de identificação, capaz de contemplar as diversas necessidades de estudos, permitindo que as análises possam ser compartilhadas. A eficiência da seleção de elétrons é utilizada em diversas dessas análises físicas. Com isso, a seleção de elétrons isEM tem uma importância central no grupo de performance do ATLAS.

Para suportar uma diversidade de estudos físicos, três pontos de operação têm sido desenvolvidos, como mostra a Tabela 3. Os pontos de operação são:

- *Loose*: com a pior rejeição de ruído de fundo e maior probabilidade de detecção de sinal.
- *Tight*: tem a melhor rejeição de ruído de fundo e menor probabilidade de detecção de sinal.
- *Medium*: tem um nível de rejeição de ruído melhor que o *Loose* e uma maior probabilidade de detecção de sinal do que o *Tight*. O critério isEM *medium* é utilizado na reconstrução de elétrons no HLT, com alguns ajustes.

O isEM foi desenvolvido utilizando dados de Monte Carlo (MC) para as distribuições de sinal e ruído de fundo, antes da aquisição de dados reais. Os valores dos cortes

Tabela 3: Sumário das variáveis usadas nos critérios *Loose*, *Medium* e *Tight* do isEM. Extraído de (ALISON, 2014)

Loose
Middle-layer shower shapes: R_η, w_2
Hadronic leakage: R_{had1} (R_{had} for $0.8 < \|\eta\| < 1.37$)
Medium
Pass Loose selection
Strip-layer shower shapes: $w_{s,tot}$, E_{ratio}
Track quality
$\|\Delta\eta\| < 0.01$
$\|d0\| < 5$ mm
Tight
Pass Medium selection
$\|\Delta\eta\| < 0.005$
$\|d0\| < 1$ mm
Track matching: $\|\Delta\phi\|$ and E/P
High TRT HT fraction
NBL ≥ 1
Pass conversion bit

utilizados no menu foram otimizados para essa realidade, segmentando as PDF's em *bins* de η e E_t, como na Tabela 2.

Com a aquisição dos primeiros dados reais, em 2009 e 2010, ficou evidente que muitas PDF's de MC foram mal estimadas, principalmente devido ao alargamento dos chuveiros dos dados reais em relação ao MC. Isso implica em uma perda de eficiência do menu isEM para os dados reais, visto que, os cortes ficam deslocados nesses dados. Esses problemas foram contornados relaxando e alterando os cortes baseando-se, agora, nos dados reais.

Com a luminosidade instantânea alcançada de $10^{33} cm^{-2} s^{-1}$, a taxa de rejeição de ruído de fundo do reotimizado *medium* isEM não foi suficiente para fornecer taxas sustentáveis para o *trigger online* de elétrons, para a largura de banda disponível. Então, foi necessário desenvolver alternativas para aumentar a rejeição do ruído de maneira mais eficiente. A necessidade de utilizar somente as variáveis do *medium* serviu para romper com a filosofia tradicional do menu isEM. Ao invés de utilizar um conjunto de variáveis em cada ponto de operação, foi decidido utilizar todas as variáveis em todos os critérios e alterar somente os valores de cortes, como na Tabela 4. Em 2011 o menu isEM ganhou uma versão mais atualizada, referida como isEM++ e seus pontos de operação foram chamados: *Loose++*, *Medium++* e *Tight++*.

Tabela 4: Sumário das variáveis usadas nos critérios *Loose++*, *Medium++* e *Tight++* do isEM++. Extraído de (ALISON, 2014)

Loose++
Shower shapes: R_{η}, R_{had1} (R_{had}, w_2, E_{ratio}, $w_{s,tot}$
Track quality
$\|\Delta\eta\| < 0.015$

Medium++
Shower shapes: Same variables as Loose++, but at tighter values
Track quality
$\|\Delta\eta\| < 0.005$
$N_{BL} \geq 1$ for $\eta < 2.01$
$N_{Pix} > 1$ for $\eta > 2.01$
Loose TRT HT fraction cuts
$\|d0\| < 5$ mm

Tight++
Shower shapes: Same variables as Medium++, but at tighter values
Track quality
$\|\Delta\eta\| < 0.005$
$N_{BL} \geq 1$ for all η
$N_{Pix} > 1$ for $\eta > 2.01$
Tighter TRT HT fraction cuts
$\|d0\| < 1$ mm
E/P requirement
$\|\Delta\phi\|$ requirement
Conversion bit

A identificação de elétrons baseada em cortes nas variáveis discriminantes atingiu um limite de performance com o isEM++, qualquer melhora na rejeição de ruído de fundo viria em detrimento da detecção de sinal, segundo (ALISON, 2014).

3.3.2 VEROSSIMILHANÇA

A análise multivariada tem sido amplamente utilizada em análises físicas na separação de sinal e ruído de fundo. Em contraste com o corte do e/γ, essa abordagem permite uma avaliação simultânea das variáveis discriminantes antes da tomada de decisão (COLLABORATION et al., 2013). A técnica de verossimilhança, no contexto de classificação, utiliza uma PDF conjunta de sinal e outra de ruído de fundo, das variáveis discriminantes, para atribuir valores probabilísticos a um determinado evento. As respectivas probabilidades são combinadas e utilizadas para definir a qual grupo determinado evento pertence. As PDFs utilizadas na verossimilhança da colaboração ATLAS foram estimadas utilizando um método não-paramétrico, chamado Estimação de Densidade de Núcleo, (do inglês, *Kernel Density Estimation)* (KDE), a partir de uma ferramenta chamada TMVA *Tool*. Detalhes sobre esse método serão mostrados na seção 4.2.

A verossimilhança representa as PDFs conjuntas, utilizadas em inferências estatísticas do problema em análise. A equação geral para verossimilhança é:

$$L_s\left(\theta\right) = P_s(x|\theta) \qquad e \qquad L_b\left(\theta\right) = P_b(x|\theta) \tag{3.4}$$

Ou seja, a verossimilhança (L_s ou L_b) é denotada pela probabilidade conjunta que um evento x possui, sendo que essa função (P_s ou P_b) densidade de probabilidade (univariada ou multivariada) foi parametrizada por θ. Quando essa função é conhecida na literatura, e seus parâmetros adequados ao problema, é possível encontrar os valores de probabilidade do evento x, sendo esse raciocínio aplicado a variáveis aleatórias dependentes e independentes.

Se as variáveis aleatórias do problema forem independentes é possível fazer uma simplificação na formulação da verossimilhança multivariada, ou seja, podemos utilizar a multiplicação da probabilidade de cada dimensão do evento x para encontrar a probabilidade conjunta. Por exemplo, cada evento possui probabilidades associadas a cada variável do problema, e se possuirmos n variáveis teremos n dimensões em nossa análise, com n valores de probabilidade multiplicados associados à PDF de sinal e n va-

lores multiplicados associados à PDF de ruído de fundo. Ao assumirmos independência entre as variáveis, como a colaboração, teremos a possibilidade de fazer uma simplificação na formulação da verossimilhança multivariada para sinal (equação 3.5) e para ruído de fundo (equação 3.6), sendo posteriormente combinadas em um discriminante (equação 3.7), dadas por:

$$L_s\left(x\right) = \prod_{i=1}^{n} P_{s,i}(x_i) \tag{3.5}$$

$$L_b\left(x\right) = \prod_{i=1}^{n} P_{b,i}(x_i) \tag{3.6}$$

$$dL = \frac{Ls}{Ls + Lb} \tag{3.7}$$

Onde $P_{s,i}(x_i)$ e $P_{b,i}(x_i)$ são as probabilidades associadas a cada variável n da análise do evento x, L_s e L_b são os valores da multiplicação das probabilidades univariadas e dL o discriminante.

A verossimilhança para classificação de eventos, da colaboração ATLAS, consiste em três passos:

1. Escolher as variáveis a serem utilizadas nos cálculos das PDFs;
2. Selecionar variáveis discriminantes adicionais a serem aplicadas antes do dL da verossimilhança. (esses são cortes adicionais, separados do discriminante, referentes a qualidade do traço da partícula, se o evento falha nesses cortes ele é considerado ruído de fundo);
3. Escolher um valor limiar para o discriminante dL.

A eficiência da verossimilhança é resultado da combinação da probabilidade de detecção de sinal do discriminante dL, após a aplicação dos cortes adicionais.

A impossibilidade de classificar eventos de interesse nas caudas das PDFs está relacionada a característica do algoritmo padrão e/γ fazer cortes rígidos nas variáveis discriminantes. O classificador baseado na verossimilhança contorna esse problema, levando até o fim do processo a possibilidade de classificar qualquer evento, possibilitando uma melhora na probabilidade de detecção de sinal e rejeição do ruído de fundo.

A formulação da verossimilhança, vista anteriormente, é aplicada idealmente quando

as variáveis são independentes, portanto variáveis como $\Delta\phi_2$ e E/P não são utilizadas pela colaboração ATLAS, por serem fortemente dependentes a outras variáveis. Essa dependência degradaria a performance da técnica de classificação baseada na verossimilhança, pois ocasiona um erro na estimação da probabilidade conjunta das PDFs de sinal e ruído de fundo. Esse tema será abordado na seção 4.2.5.

3.3.2.1 VEROSSIMILHANÇA PARA ELÉTRONS

A verossimilhança oferece cinco diferentes pontos de operação: *Very Tight, Tight, Medium, Loose, Very Loose*. Com diferentes níveis de rejeição de ruído e probabilidade de detecção de sinal.

As variáveis utilizadas pela verossimilhança serão mostradas na Tabela 5. Note que, algumas variáveis são utilizadas somente pelo e/γ enquanto outras, como ΔP/P e $\Delta\phi_{res}$, são utilizadas pela verossimilhança. Diferente do e/γ, devido ao seu método construtivo, a verossimilhança faz uso de variáveis de isolamento fortemente sobrepostas como, Rϕ e f_1.

Existem diferenças conhecidas no formato do chuveiro eletromagnético entre os dados reais e de MC, afetando variáveis como Rη, Wη_2. Além disso, a radiação de transição no TRT é subestimada no MC. Portanto, ao fazer uso das informações das PDFs a verossimilhança torna-se sensível a sua má estimação, justificando o uso de dados reais, posteriormente, na construção das PDFs de sinal e ruído de fundo, na classificação de dados reais.

Tabela 5: Definição das variáveis discriminantes do elétron, que foram usadas no *cut-based* e *likelihood* em 2012. Extraído de (COLLABORATION et al., 2013)

Type	Description	Name	Cuts	LH
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ and $\|\eta\| > 1.37$)	R_{Had1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| > 0.8$ and $\|\eta\| < 1.37$)	R_{Had}	✓	✓
Third layer of EM calorimeter	Ratio of the energy in the third layer to the total energy	f_3	✓	✓
Middle layer of EM calorimeter	Ratio of the energy in 3x7 cells over the energy in 7x7 cells centered at the electron cluster position	R_η	✓	✓
	Ratio of the energy in 3x3 cells over the energy in 3x7 cells centered at the electron cluster position	R_ϕ		✓
	Lateral width of the shower	$W_{\eta 2}$	✓	✓
Strip layer of EM calorimeter	Total shower width	W_{stot}	✓	
	Ratio of the energy difference between the largest and second largest energy deposits in the cluster over the sum of these energies	E_{ratio}	✓	✓
	Ratio of the energy in the strip layer to the total energy	f_1		✓
Track quality	Number of hits in the pixel detector	nPixHits	✓	✓
	Number of total hits in the pixel and SCT detectors	nSiHits	✓	✓
	Transverse impact parameter	d_0	✓	✓
	Significance of transverse impact parameter	σ_{d_0}	✓	✓
Track–cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$	✓	✓
	$\Delta\phi$ between the cluster position in the middle layer and the extrapolated track	$\Delta\phi_2$	✓	
	Ratio of the cluster energy to the track momentum	E/p	✓	
TRT	Total number of hits in the TRT	nTRTHits		
	Ratio of the number of high-threshold hits to the total number of hits in the TRT	F_{HT}	✓	✓
Conversions	Number of hits in the Blayer	nBlayerHits	✓	✓
Bremsstrahlung (GSF output)	Momentum lost by the track between the perigee and the last measurement point divided by original momentum	$\Delta p/p$	(Multilepton)	✓
	Same as $\Delta\phi_2$, but the track momentum is rescaled to the cluster energy before extrapolating to the middle layer	$\Delta\phi_{Res}$	(Multilepton)	✓

Na tabela 6 percebemos que existem diferenças entre a utilização das variáveis discriminantes (*likelihood variables*) e os cortes adicionais (*add cuts*) entre os cinco pontos de operação do menu da verossimilhança. Por exemplo, no menu *Loose* as variáveis d_0 e σ_{d_0} não são utilizadas. Nos cortes adicionais, no menu *Very Tight* utiliza-se a variável isConv, para suprimir elétrons de conversões de fótons.

Tabela 6: Variáveis usadas na construção da *likelihood* para diferentes pontos de operação. Extraído de (COLLABORATION et al., 2013)

Menu	(VERY TIGHT), TIGHT	MEDIUM	LOOSE, (VERY LOOSE)
Likelihood Variables	R_{Had}	R_{Had}	R_{Had}
	R_{η}	R_{η}	R_{η}
	F_{HT}	F_{HT}	F_{HT}
	$\Delta\eta_1$	$\Delta\eta_1$	$\Delta\eta_1$
	$W_{\eta 2}$	$W_{\eta 2}$	$W_{\eta 2}$
	f_1	f_1	f_1
	f_3	f_3	f_3
	E_{ratio}	E_{ratio}	E_{ratio}
	R_{ϕ}	R_{ϕ}	R_{ϕ}
	$\Delta p/p$	$\Delta p/p$	$\Delta p/p$
	$\Delta\phi_{Res}$	$\Delta\phi_{Res}$	$\Delta\phi_{Res}$
	d_0	d_0	
	σ_{d_0}	σ_{d_0}	
Additional Cuts	nSiHits ≥ 7	nSiHits ≥ 7	nSiHits ≥ 7
	nPixHits ≥ 2	nPixHits ≥ 2	nPixHits ≥ 2 (≥ 1)
	Blayer	Blayer	Blayer (no Blayer)
	!(isConv)		
Compare to (macro)	isTightPlusPlus	MediumPlusPlus	isLoosePlusPlus Multilepton

Nos próximos capítulos, sempre que referirmos ao ponto de operação *tight*, *medium* e *loose* da verossimilhança, é importante entender que os menus foram construídos com as especificações da tabela 6. Os pontos de operação *tight*, *medium* e *loose* do e/γ são referentes aos menus da tabela 4.

4 REVISÃO BIBLIOGRÁFICA

Esta seção tem como objetivo descrever as teorias matemáticas utilizadas como base na construção dos algoritmos dessa dissertação, responsáveis pela identificação de elétrons isolados. Primeiro, será visto a teoria da informação mútua, responsável por identificar o nível de dependência entre as variáveis discriminantes. Depois, será mostrado o método não-paramétrico utilizado para estimar as densidades de probabilidade das variáveis discriminantes, esse método é conhecido como KDE. Após o embasamento teórico será possível compreender a utilização da dependência estatística entre as variáveis discriminantes como parâmetro de entrada para uma estimação de densidade multivariada.

4.1 INFORMAÇÃO MÚTUA

Informação mútua é a medida da dependência estatística entre duas variáveis aleatórias. A escolha do método deve-se a característica estatística das variáveis descritas na Seção 3.2. Como foi visto, estas variáveis não são Gaussianas, portanto, existem dependências de ordens superiores que não podem ser percebidas utilizando a análise de correlação.

4.1.1 ENTROPIA

Entropia é a medida da incerteza de uma variável aleatória. Dado uma variável aleatória discreta X, com M valores distintos e estatisticamente independentes, e se quando o valor de ordem j for transmitido a informação transportada for $I_j = -\log_b P_j$, nossa entropia associada aos M valores da variável aleatória X será a média ponderada das auto-informações de cada valor assumido por X.

Definição: A entropia $H(X)$ de uma variável aleatória discreta é definida como: (COVER; THOMAS, 2012)

$$H(X) = \sum_{j=1}^{M} P_j I_j = -\sum_{j=1}^{M} P_j \log_b P_j \tag{4.1}$$

Onde P_j é a probabilidade do valor j da variável aleatória.

Propriedade 1 : $H(X) \geq 0$

Prova: $0 \leq P_j \leq 1$ implica em $-\log_b P_j \geq 0$

Propriedade 2 : $H_b(X) = \log_b a H_a(X)$

Prova: $\log_b p = \log_b a \log_a p$

A segunda propriedade da entropia demonstra que é possível a mudança de base do logaritmo da definição, $b = 2$ (bit), $b = e$ (nat) e $b = 10$ (hartley).

4.1.2 ENTROPIA CONDICIONAL

Na maioria dos problemas, existem diversas variáveis aleatórias e em muitas vezes essas variáveis possuem informação comum entre si, portanto, é importante definir o conceito de entropia condicional.

Admitindo a existência de duas variáveis aleatórias X e Y, com M e N possibilidades de valores, respectivamente, então $P(x_i, y_j)$ é a probabilidade conjunta de ocorrência e $P(y_j|x_i)$ é a probabilidade condicional de y_j ocorrer dado que x_i ocorreu.

A entropia condicional de Y dado a ocorrência de x_i é definida como:

$$H(Y|x_i) = \sum_{j=1}^{N} P(y_j|x_i) \log_b \frac{1}{P(y_j|x_i)} = -\sum_{j=1}^{N} P(y_j|x_i) \log_b P(y_j|x_i) \tag{4.2}$$

A entropia condicional de Y dado X é definida como a média ponderada de $H(Y|x_i)$ para todos os valores de x_i:

$$H(Y|X) = \sum_{i=1}^{M} P(x_i) H(Y|x_i) = -\sum_{i=1}^{M} \sum_{j=1}^{N} P(x_i) P(y_j|x_i) \log_b P(y_j|x_i) \tag{4.3}$$

A entropia conjunta é definida como:

$$H(X,Y) = \sum_{i=1}^{M}\sum_{j=1}^{N} P(x_i, y_j) \log_b \frac{1}{P(x_i, y_j)} \tag{4.4}$$

A diferença entre $H(X) - H(X|Y)$ é a informação mútua média entre as variáveis X e Y , e representa a incerteza de uma variável em relação a outra, ou seja, é a redução da incerteza sobre X ao conhecer Y.

Definição: A informação mútua entre as variáveis X e Y é:

$$\begin{aligned} &I(X;Y) = H(X) - H(X|Y) \\ &= \sum_{i=1}^{M} P(x_i) \log_b \tfrac{1}{P(x_i)} - \sum_{i=1}^{M}\sum_{j=1}^{N} P(y_j)\, P(x_i|y_j) \log_b \tfrac{1}{P(x_i|y_j)} \end{aligned} \tag{4.5}$$

Mas, $P(x_i) = \sum_{j=1}^{N} P(x_i, y_j)$ e $P(x_i, y_j) = P(x_i)P(y_j|x_i)$(regra de Bayes), logo:

$$\begin{aligned} &I(X;Y) = \sum_{i=1}^{M}\sum_{j=1}^{N} P(x_i, y_j) \left[\log_b \tfrac{1}{P(x_i)} + \log_b \tfrac{P(x_i,y_j)}{P(y_j)}\right] \\ &= \sum_{i=1}^{M}\sum_{j=1}^{N} P(x_i, y_j) \log_b \tfrac{P(x_i,y_j)}{P(x_i)P(y_j)} = \sum_{i=1}^{M}\sum_{j=1}^{N} P(x_i, y_j)\, I(x_i, y_j) \end{aligned} \tag{4.6}$$

Onde, $I(x_i, y_j)$ é a informação mútua, sendo definida como:

$$I(x_i, y_j) = \log_b \frac{P(x_i, y_j)}{P(x_i)\, P(y_j)} = \log_b \frac{P(y_j|x_i)}{P(y_j)} = \log_b \frac{P(x_i|y_j)}{P(x_i)} \tag{4.7}$$

As relações entre entropia e informação mútua média podem ser visualizadas no diagrama de Venn (Figura 24) abaixo:

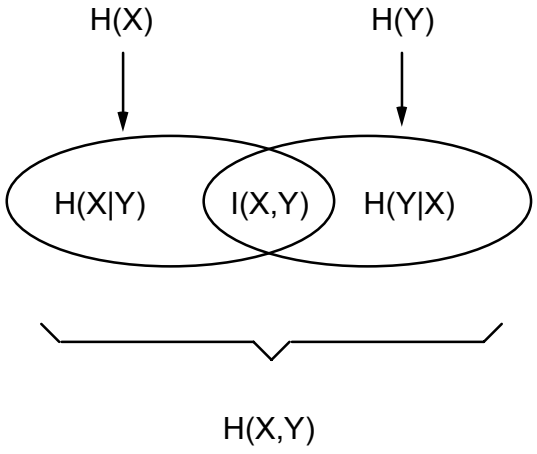

Figura 24: Diagrama de Venn: relação entre entropias condicionais, conjunta e informação mútua média.

Do diagrama de Venn é possível extrair as propriedades:

- $H\left(X,Y\right)=H\left(X\right)+H\left(Y|X\right)=H\left(Y\right)+H(X|Y)$;
- $H\left(X,Y\right)\leq H\left(X\right)+H(Y)$;
- $H\left(X|Y\right)\leq H(X)$;
- $H\left(Y|X\right)\leq H(Y)$;
- $I\left(X;Y\right)=H\left(X\right)-H\left(X|Y\right)=H\left(Y\right)-H(Y|X)$;
- $I\left(X;Y\right)=H\left(X\right)+H\left(Y\right)-H\left(X,Y\right)$.

Seja XY um conjunto discreto de pares (x_i,y_j) . A informação mútua média entre eles satisfaz $I\left(X;Y\right)\geq 0$, onde a igualdade só é satisfeita se e somente se X e Y forem estatisticamente independentes.

Demonstração de $I\left(X;Y\right)=0$:

$$I\left(X;Y\right)=\sum_{i=1}^{M}\sum_{j=1}^{N}P\left(x_i,y_j\right)I(x_i,y_j) \tag{4.8}$$

Onde,

$$I\left(x_i,y_j\right)=\log_b\frac{P\left(x_i,y_j\right)}{P\left(x_i\right)P\left(y_j\right)}=\log_b\frac{P\left(y_j|x_i\right)}{P\left(y_j\right)}=\log_b\frac{P\left(x_i|y_j\right)}{P\left(x_i\right)} \tag{4.9}$$

Então,

$$I\left(X;Y\right)=\sum_{i=1}^{M}\sum_{j=1}^{N}P\left(x_i,y_j\right)\log_b\frac{P\left(x_i|y_j\right)}{P\left(x_i\right)}=\sum_{i=1}^{M}\sum_{j=1}^{N}P\left(x_i,y_j\right)\log_b\frac{P\left(y_j|x_i\right)}{P\left(y_j\right)} \tag{4.10}$$

Como,

$$\sum_{i=1}^{M}\sum_{j=1}^{N}P\left(x_i,y_j\right)=1 \tag{4.11}$$

Para, $I\left(X;Y\right)=0\rightarrow P\left(x\right)=P\left(x|y\right)$ e $P\left(y\right)=P\left(y|x\right)$, ocorrendo somente quando x e y são estatisticamente independentes.

4.2 KERNEL DENSITY ESTIMATION

O KDE (TURLACH et al., 1993) é um método não-paramétrico utilizado para estimar funções densidade de probabilidade de modelos desconhecidas.

4.2.1 ESTIMADOR DISCRETO

Se uma variável aleatória X possui uma distribuição contínua $F(x)$ e densidade $f(x) = \frac{dF(x)}{dx}$, o objetivo é estimar $f(x)$ de uma amostra randômica $X_1, \ldots, X_n$ (HANSEN, 2009).

$F(x)$ é uma função de distribuição que pode ser estimada pela *Empirical Distribution Function* (EDF) $\hat{F}(x) = n^{-1} \sum_{k=1}^{n} 1(X_k \leq x)$. Como $\hat{f}(x) = \frac{d\hat{F}(x)}{dx}$, esse é um estimador de massa de probabilidade, portanto, não é utilizado para estimar uma função densidade de probabilidade.

Ao invés disso, a derivada discreta é considerada, para algum pequeno valor de $h > 0$, como:

$$\hat{f}(x_i) = \frac{\hat{F}(x_i + h) - \hat{F}(x_i - h)}{2h} \tag{4.12}$$

Que pode ser reescrita como:

$$\begin{aligned} &\tfrac{1}{2nh} \sum_{k=1}^{N} 1(x_i + h < X_k \leq x_i + h) = \tfrac{1}{2nh} \sum_{k=1}^{N} 1\left(\left|\tfrac{(x_i - X_k)}{h}\right| \leq 1\right) \\ &= \tfrac{1}{nh} \sum_{k=1}^{N} K\left(\tfrac{(x_i - X_k)}{h}\right) \end{aligned} \tag{4.13}$$

onde,

$$K(u) = \left\{ \begin{array}{l} \frac{1}{2}, |u| \leq 1 \\ 0, |u| > 1 \end{array} \right\} \tag{4.14}$$

é uma função uniforme em [-1,1].

O estimador $\hat{f}(x_i)$ conta a porcentagem de observações que estão próximas ao ponto x_i. O parâmetro h é conhecido como largura de banda.

$\hat{f}(x_i)$ é chamado de estimador *Kernel*. Sua forma geral é dada por:

$$\hat{f}(x_i) = \frac{1}{nh} \sum_{k=1}^{N} K\left(\frac{(x_i - X_k)}{h}\right) \tag{4.15}$$

onde *K(.)* é a função *Kernel*.

4.2.2 FUNÇÃO KERNEL

Uma função *kernel* é qualquer função que satisfaça $\int_{-\infty}^{+\infty} K(u)\, du = 1$. Para a função *kernel* ser uma função densidade de probabilidade ela precisa ser não negativa, ou seja, $K(u) \geq 0$, para todo u. Uma função *kernel* simétrica satisfaz: $K(u) = K(-u)$, para todo u.

Os momentos do *kernel* são: $k_j(k) = \int_{-\infty}^{+\infty} u^j K(u)\, du$. A ordem de *K(.)* é definida como a ordem do primeiro momento não zero, sendo representada por v. A função *kernel* terá partes negativa e não será função de probabilidade se for de alta ordem, ou seja, quando $v > 2$. Alguns exemplos de funções *kernel* de segunda ordem:

- Triangular: $K(u) = (1 - |u|)I(|u| \leq 1)$
- Epanechnikov: $K(u) = \frac{3}{4}(1 - u)I(|u| \leq 1)$
- Quartic (Biweight): $K(u) = \frac{15}{16}(1 - u)I(|u| \leq 1)$
- Triweight: $K(u) = \frac{35}{32}(1 - u)I(|u| \leq 1)$
- Gaussiana: $K(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}u\right)}$

A forma geral para o KDE é:

$$\hat{f}_h(x_i) = \frac{1}{n} \sum_{k=1}^{n} K_h(x_i - X_k) \tag{4.16}$$

onde, n é o número de pontos a estimar em torno de x_i, $K_h(.)$ é a função *kernel* com alguns termos implícitos.

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right) \tag{4.17}$$

onde, K é a função *kernel* vista anteriomente, u é uma variável com valor $x_i - X_k$ e h é a largura de banda.

4.2.3 LARGURA DE BANDA

A largura de banda h controla a suavidade da estimativa de probabilidade e sua escolha é um problema crucial. A largura de banda pode ser de dois tipos, fixa ou variável e existem diversas teorias que abordam o tema de otimização desse parâmetro.

Largura de Banda Fixa Como o próprio nome denota, somente uma banda será escolhida, fixa, para toda a estimação da função de probabilidade. Visualmente é possível perceber o impacto da variação de h, como ilustrado na Figura 25.

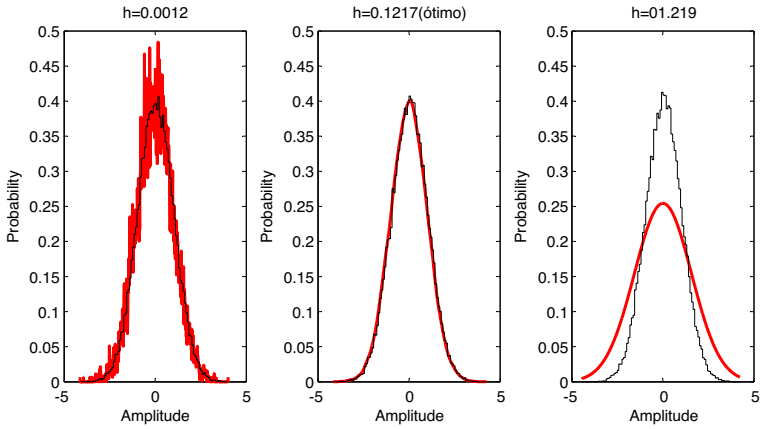

Figura 25: Variação de h de banda fixa na estimação da PDF.

Largura de Banda Variável Ao invés de utilizar somente um h, alguns autores têm considerado a possibilidade de se utilizar uma largura de banda $h(x_i)$ para cada ponto de x_i em que desejamos estimar a probabilidade $f_{h_i}(x_i)$. Esse estimador é conhecido como *balloon estimator* e tem a forma:

$$\hat{f}_{hi}(x_i) = \frac{1}{nh(x_i)} \sum_{k=1}^{n} K\left(\frac{(x_i - X_k)}{h(x_i)}\right) \tag{4.18}$$

Onde $h(x_i)$ é uma largura de banda que varia de acordo com o ponto x_i que desejamos estimar.

Outro método de banda variável é obtido diferentemente da variação da banda em cada ponto que desejamos estimar. Essa variação é feita a cada valor (ou evento) da variável aleatória. Esse estimador é conhecido como *sample-point estimator* e foi introduzido por Breimam, Meisel e Prucel (BREIMAN; MEISEL; PURCELL, 1977), e tem a forma:

$$\hat{f}_{hi}(x_i) = \frac{1}{nh(X_k)} \sum_{k=1}^{n} K\left(\frac{(x_i - X_k)}{h(X_k)}\right) \tag{4.19}$$

Onde $h(X_k)$ é uma largura de banda que varia de acordo com cada valor X_k da variável aleatória X.

Nessa dissertação utilizaremos esses dois tipos de estimadores de banda variável, porém existem variações das formulações, que podem ser encontradas em (HALL, 1992), (TARON; PARAGIOS; JOLLY, 2005), (WU; CHEN; CHEN, 2007) e (GINÉ; SANG, 2010).

4.2.3.1 MÉTODOS DE ESTIMAÇÃO DA LARGURA DE BANDA DO KERNEL

Existem lacunas importantes para a implementação deste método, visto que, não foi escolhido os melhores parâmetros para construção do KDE. O único parâmetro mencionado, passível de otimização, foi a largura de banda $h(.)$. (ABRAMSON, 1982) sugere:

$$h(x_i) = \frac{h}{\sqrt{f_p(x_i)}} \tag{4.20}$$

Onde, h é uma largura de banda fixa e $f_p(x_i)$ é a probabilidade de x_i na PDF. Porém, na tentativa de solucionar a escolha do parâmetro de banda variável, surgem dois problemas. Não conhecemos o h fixo otimizado e nem $f_p(x_i)$.

A natureza da estimação da largura de banda é não-supervisionada, visto que, não conhecemos a função geradora dos dados. Portanto, não seria possível, a princípio, minimizar uma função custo para obter um parâmetro ótimo (BARBOSA, 2013).

Apesar das restrições, (SILVERMAN, 1986) propôs um método de estimação da largura de banda, baseado na minimização do Erro médio quadrático integrado, (do inglês, *Mean Integrated Squared Error*) (MISE), para encontrar o melhor h que leva a melhor estimação possível, para uma função geradora desconhecida f . O MISE é dado pela equação abaixo:

$$\begin{aligned} &\text{MISE}(f_h(x)) = \int E\Big\{\hat{f}_h(x) - f(x)\Big\}^2 dx \\ &= \int E\Big\{\hat{f}_h(x) - f(x)\Big\}^2 dx + \int \text{var}\hat{f}_h(x)\, dx \end{aligned} \tag{4.21}$$

O MISE é formulado como a soma do viés integrado e da variância integrada.

Silverman propôs fazer o viés como $\frac{1}{2}h^2 f''(x)k_2$ e a variância como $\frac{1}{nh}\int K(t)^2$, podendo reescrever a equação 4.21 como:

$$\text{MISE}(\hat{f}_h(x)) = \frac{1}{4}h^4 k_2^2 \int f''(x)dx + \frac{1}{nh}\int K(t)^2 dt \tag{4.22}$$

Onde, k_2 é uma constante do segundo termo da expansão da Série de Taylor, $f''(x)$ é a derivada segunda da função geradora, n é o tamanho da amostra de dados e K é a função *kernel* usada.

Os cálculos anteriores dependem de uma função geradora desconhecida, e o tipo de *kernel* utilizado. Silverman assume que os dados foram gerados por uma distribuição normal e o *Kernel* é Gaussiano. Portando, o valor de h que minimiza o MISE é:

$$h = 1,06\sigma n^{-1/5} \tag{4.23}$$

Neste trabalho, foi escolhido como largura de banda fixa a proposta de (WAND; JONES, 1995), que é uma generalização de h para todas as dimensões.

$$h_j = \left(\frac{4}{d+2}\right)^{\frac{1}{(d+4)}} n^{\frac{-1}{(d+4)}}\sigma_j \tag{4.24}$$

Onde, d representa o número de dimensões do problema, j é o subíndice da respectiva dimensão, n é o número de eventos e σ_j é o desvio padrão dos eventos da dimensão j. Note que para *d=1* a fórmula coincide com o método de Silverman. Percebe-se que apesar de todas as formulações serem feitas para o caso unidimensional, o conceito de KDE Multidimensional começa a ser introduzido, e sua generalização será feita posteriormente.

A equação 4.20 mostra que além do parâmetro h fixo é preciso estimar a função de probabilidade $f_p(x_i)$. Uma escolha é feita baseada no algoritmo proposto por (SHIMAZAKI; SHINOMOTO, 2007) que estima a binagem ótima de um histograma. Então, esse histograma será normalizado e utilizado como uma estimativa, mesmo que flutuante devido ao erro, de $f_p(x_i)$.

Depois da escolha de h fixo e $f_p(x_i)$ é possível estimar um $h(x_i)$. A consideração inicial por uma função geradora normal e um *Kernel* Gaussiano na otimização do parâmetro h fixo não fornece uma boa estimativa para o KDE, devido a variedade nas

formas das PDFs. A teoria provê outro parâmetro para tornar o KDE mais resiliente, a solução foi encontrada em (COMANICIU; RAMESH; MEER, 2001) que insere um novo parâmetro λ, chamado de constante de proporcionalidade, sendo incorporado a formulação da banda variável da seguinte forma:

$$h\left(x_i\right) = h\left[\frac{\lambda}{f_p(x_i)}\right]^{\frac{1}{2}} \tag{4.25}$$

Segundo (COMANICIU; RAMESH; MEER, 2001), o parâmetro λ é dado por:

$$\lambda = e^{n^{-1}\sum_{i=1}^{n}\log(f_p(x_i))} \tag{4.26}$$

4.2.4 KDE COMO A SOMA DE VÁRIAS PROBABILIDADES

Até o momento, foi considerado a probabilidade do ponto x_i olhando a influência dos pontos ao seu redor, dados por X_k.

Supondo uma situação ideal, onde exista um total de pontos x infinitos dentro da faixa de X onde a PDF é estimada. Note que a área sob a curva de densidade estimada será sempre igual a um, independente do número de amostras m.

$$\begin{aligned}
&\int_{-\infty}^{+\infty} f_h\left(X\right) dX \\
&= \tfrac{1}{m}\sum_{i=1}^{m}\int_{-\infty}^{+\infty}\tfrac{1}{h}K\left(\tfrac{(x_i-X)}{h}\right) dX \\
&= \tfrac{1}{mh}\sum_{i=1}^{m} h\int_{-\infty}^{+\infty} K\left(u\right) du = \tfrac{1}{m}\sum_{i=1}^{m} 1 = 1
\end{aligned} \tag{4.27}$$

4.2.5 KDE MULTIVARIADO

Na maioria das vezes, as análises possuem mais do que uma variável aleatória, sendo necessário calcular a probabilidade conjunta entre as variáveis. Baseado na teoria anterior é possivel generalizar a formulação do KDE:

$$f_{h_1,h_2,h_n}\left(x_{1,2,n}\right) = \frac{1}{n}\sum_{k=1}^{N}\frac{1}{h_1}\frac{1}{h_2}\frac{1}{h_n}K\left(\frac{(x_1 - X_{k_1})}{h_1}\right)K\left(\frac{(x_2 - X_{k_2})}{h_2}\right)K\left(\frac{(x_n - X_{k_n})}{h_n}\right) \tag{4.28}$$

O KDE multivariado funciona como um acumulador, onde cada probabilidade no

espaço n dimensional é ponderada pela quantidade de eventos no entorno do ponto em que se deseja estimar, e suavizada pela função *kernel*, de acordo com sua respectiva largura de banda.

É possível reescrever as fórmulas anteriores matricialmente, tendo um raciocínio análogo. Ao escrever a largura de banda como uma matriz diagonal geramos uma função *kernel* gaussiana multidimensional sem correlação, como:

$$H = diag(h_1, h_2, n) \tag{4.29}$$

Sendo $x = x_{1,2,n}$ uma variável aleatória multidimensional, o KDE Multivariado é generalizado, matricialmente, por:

$$f_H(x) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\det(H)}K\left\{H^{-1}(x - X_k)\right\} = \frac{1}{n}\sum_{k=1}^{n}K_H(x - X_k) \tag{4.30}$$

com,

$$K_H(u) = \frac{1}{\det(H)}K\left(H^{-1}u\right) \qquad , u = (x - X_k) \tag{4.31}$$

4.2.6 "A MALDIÇÃO DA DIMENSIONALIDADE"

Embora, a princípio, problemas com múltiplas dimensões num espaço de amostra não sejam essencialmente diferentes de problemas unidimensionais, há uma diferença prática muito importante. Este problema é conhecido como a "maldição da dimensionalidade" (NARSKY; PORTER, 2013). Além da dificuldade na visualização, quando as dimensões aumentam, todo o volume de uma região delimitada aumenta para as fronteiras exponencialmente. Como ilustrado na Figura 26, começando pela reta da esquerda unidimensional, com uma região delimitada no centro com tamanho de 1/2 unidades. Quando uma dimensão é aumentada, a área do quadrado central representa 1/4 da área total. Ao aumentar mais uma dimensão, o volume do cubo central representa 1/8 do volume total do cubo. As frações sofrem um decréscimo de 2^{-n} com o aumento de cada dimensão. Como consequência os dados parecem cada vez mais esparsos, aumentando o erro de estimação do *kernel* multidimensional, devido a falta de estatística.

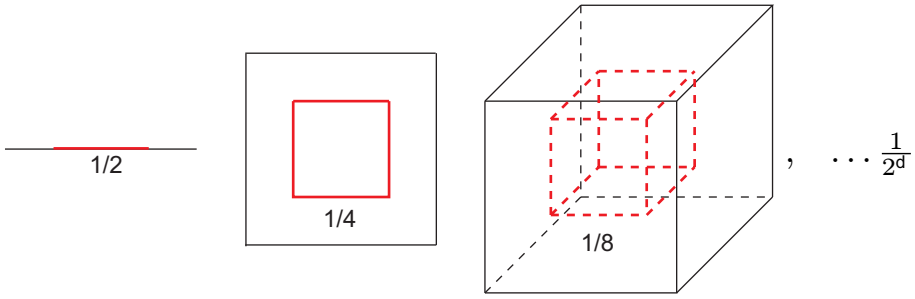

Figura 26: Demonstração gráfica da ”Maldição da dimensionalidade”.

A proposta de inserção da dependência estatística no algoritmo de verossimilhança, através de PDFs multivariadas, nos leva a pressupor uma evidente melhora no método da colaboração ATLAS, que considera todas as PDFs independentes, introduzindo um erro em sua estimação. Porém, de acordo com a “maldição da dimensionalidade”, com o aumento dimensional, é necessário mais estatística para descrever as PDFs. Portanto, sendo um método não-paramétrico, o KDE depende da estatística para uma boa estimação de densidade, evidenciando duas “forças” opostas na análise, dependência e estatística, responsáveis pela diferença entre a performance da nova proposta de verossimilhança e a proposta da colaboração.

5 DESENVOLVIMENTO

Este capítulo concatena teorias e técnicas matemáticas, vistas anteriormente, em algoritmos capazes de identificar elétrons isolados, ou seja, sistematiza as etapas necessárias para obter os resultados encontrados nessa dissertação. Para isto, como ponto de partida, é necessário conhecer os conjuntos de dados, sugeridos pelo artigo da colaboração (COLLABORATION et al., 2013), utilizados no desenvolvimento dos algoritmos de classificação baseados em verossimilhança. Os algoritmos serão detalhados em diagramas de blocos, facilitando a compreensão de cada etapa.

5.1 CONJUNTO DE DADOS

Neste trabalho usaremos dois tipos de banco de dados: Simulação MC e Dados Reais. Esses bancos de dados são referentes ao ano de 2012 e foram produzidos com uma energia de 8 TeV.

5.1.1 DADOS DE SIMULAÇÃO

É comum, durante o desenvolvimento de algoritmos de classificação de eventos, a utilização dos dados de simulação. Os algoritmos geradores de eventos de Monte Carlo descrevem, com a melhor reprodutibilidade possível, as características experimentais dos processos físicos de interesse (SJÖSTRAND, 1991).

As principais vantagens da utilização do MC são:

- Dar uma estimativa de qual tipo de evento espera-se encontrar, e em qual taxa;
- Ajudar no planejamento dos detectores, de forma que o desempenho desses seja otimizado, delimitando as restrições para o cenário físico de interesse;
- Ser utilizado como ferramenta para elaboração de estratégias das análises que devem ser utilizadas em dados reais, melhorando a relação sinal-ruído;

- Estimar as correções de aceitação do detector, que devem ser aplicadas nos dados reais, a fim de extrair um sinal físico mais próximo da realidade;
- Ser uma estrutura conveniente de interpretação para o significado de fenômenos, em termos de uma teoria mais fundamental (normalmente o Modelo Padrão).

Os dados MC utilizados estão listados abaixo, e seu perfil pode ser visto na Figura 27.

- Dados de sinal (Pacote Zee MC):

  mc12_8TeV.147806.PowhegPythia8_AU2CT10_Zee.merge.NTUP_PHOTON.e1169_s1469_s1470_r3542_r3549_p1344/

- Dados de ruído de fundo (Pacote JF17 MC):

  mc12_8TeV.129160.Pythia8_AU2CTEQ6L1_perf_JF17.merge.NTUP_EGAMMA.e1130_s1468_s1470_r3542_r3549_p1032/

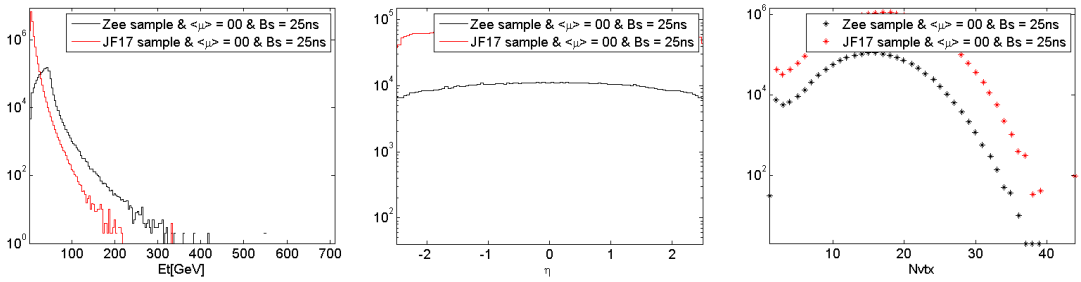

Figura 27: Perfil dos eventos de dados MC: Esquerda, Distribuição por E_t, por η (Centro) e por N_{vtx} (Direita).

5.1.2 DADOS REAIS

A etapa subsequente ao desenvolvimento dos algoritmos é a aplicação em dados reais. Esse tipo de análise difere da anterior logo na etapa inicial, pois nela não sabemos, à priori, qual é nosso conjunto de sinal e de ruído.

Nos dados MC, sabemos exatamente qual a identidade das partículas (através da variável el_truth_type) e de qual decaimento ela provém (através da variável el_truth_mothertype). Com isso, é fácil separarmos um conjunto de elétrons, que decaiu de uma partícula de interesse, do ruído de fundo, constituído de elétrons de conversões de fótons, de decaimentos semi-leptônicos e hádrons, como será abordado na Seção 3.

Na análise com dados reais a separação entre sinal e ruído de fundo é avaliada, tipicamente, pela técnica *Tag and Probe* (TP), que será explicado na seção seguinte.

O conjunto de dados reais utilizado está listado abaixo e o perfil dos dados é mostrado na Figura 28

data12_8TeV.00216399.physics_Egamma.merge.NTUP_PHOTON.r5203_p1644 _p1364

data12_8TeV.00216416.physics_Egamma.merge.NTUP_PHOTON.r5203_p1644 _p1364

data12_8TeV.00216432.physics_Egamma.merge.NTUP_PHOTON.r5203_p1644 _p1364

data12_8TeV.00214680.physics_Egamma.merge.NTUP_PHOTON.f489_m1261 _p1344_p1345

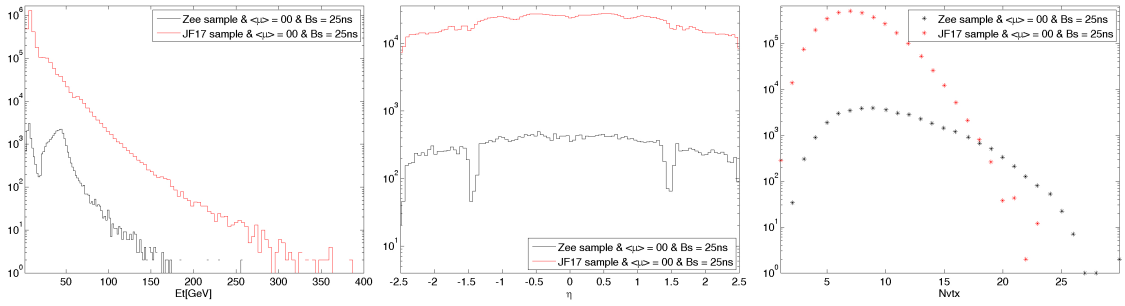

Figura 28: Perfil dos eventos de dados reais: Esquerda, Distribuição por E_t, por η (Centro) e por N_{vtx} (Direita).

5.1.2.1 TAG AND PROBE

Este método é utilizado para a identificação de elétrons que decaíram de uma partícula de interesse, em nosso caso, do bóson Z, Figura 29.

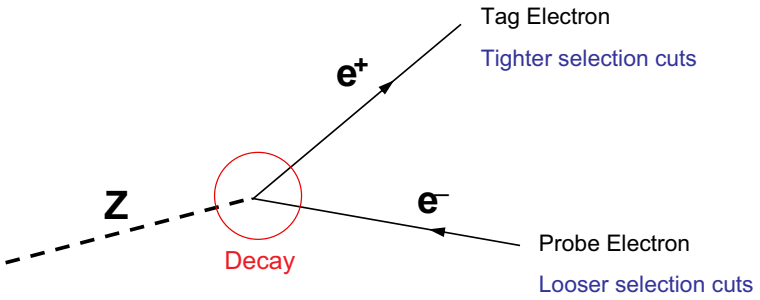

Figura 29: Ilustração do decaimento de Z.

Sabe-se que uma partícula reconstruída é identificada através da medida de sua massa invariante, pela combinação de partículas identificadas em seu estado final (SUNDARESAN, 2001).

O *Tag and Probe* (AAD et al., 2012a) utiliza esse artifício para reforçar a identificação de elétrons considerados como eventos de sinal. Para isto, o conjunto de dados é submetido a alguns cortes, restando eventos que serão divididos em dois conjuntos: *tag* e *probe*.

Para um evento ser considerado *tag*, é necessário ser aprovado pelos seguintes critérios:

- Tipo de partícula: Author 1 ou 3; (usado para excluir elétrons *forward*)
- Qualidade de Traço: número mínimo de hits no detector de *Pixels* e SCT;
- Regiões do Detector: Todo detector excluindo o *crack*;
- Aprovado pelo critério do isEM++: Tight++;
- Energia transversa: E_t >20 GeV;
- Aprovação pelo canal de trigger: EF_e24vh_loose1;
- No mínimo um vértice primário com três traços associados a ele.

Para o evento ser identificado como *probe*, é necessário ser aprovado pelos seguintes critérios:

- Tipo de partícula: Author 1 ou 3; (usado para excluir elétrons *forward*)
- Qualidade de Traço: número mínimo de hits no detector de *Pixels* e SCT;
- Regiões do Detector: Todo detector excluindo o *crack*;
- Energia transversa: $E_t cone20 < 6$ GeV;
- Aprovação pelo canal de trigger: EF_e24vh_loose1;
- No mínimo um vértice primário com três traços associados a ele.

* O significado dessas variáveis pode ser encontrado no site ntuple D3PD variables (COLLABORATION-NTUPLE,).

Agora, é possível combinar um evento do conjunto *tag* com o respectivo *probe*, ambos provenientes da mesma colisão, e calcular a massa invariante (IM):

$$IM = \sqrt{\left(\sum E^2 - \sum p^2\right)} \tag{5.1}$$

Sendo, E a energia e p o momento.

Como os momentos são dados nas coordenadas x, y e z o cálculo é feito da seguinte forma:

$$m_e = 511 * 10^{-3} \tag{5.2}$$

$$E_{tag} = \sqrt{m_e + px_{tag} + py_{tag} + pz_{tag}} \tag{5.3}$$

$$E_{probe} = \sqrt{m_e + px_{probe} + py_{probe} + pz_{probe}} \tag{5.4}$$

$$E_{sum} = E_{tag} + E_{probe} \tag{5.5}$$

$$IM = \sqrt{E_{sum} - [(px_{tag} + px_{probe})^2 + (py_{tag} + py_{probe})^2 + (pz_{tag} + pz_{probe})^2} \tag{5.6}$$

Onde, m_e é a massa do elétron em MeV, E_{tag} é a energia do *tag*, E_{probe} é a energia do *probe* e E_{sum} é a soma das energias do *tag* e *probe*. E *px*, *py* e *pz* os momentos em x, y e z de cada partícula.

A partir desse cálculo, um corte de 80 a 100 GeV é feito na massa invariante, restando apenas alguns pares dentro desta faixa. Destes pares, serão selecionados todos os eventos do conjunto *probe* como sendo o nosso sinal, devido ao conjunto ser menos polarizado em relação ao *tag*. Os eventos restantes serão considerados ruído de fundo.

5.2 ALGORITMOS

Os principais algoritmos da dissertação foram implementados em blocos, proporcionando maior controle na saída de cada etapa.

Podemos separar as implementações em duas análises principais: Análise Univariada e Análise Multivariada.

Na Análise Univariada, o modelo adotado foi o da verossimilhança do ATLAS (COLLABORATION et al., 2013), desconsiderando a dependência entre as variáveis. Para uma melhor compreensão, os processos foram organizados em um diagrama de blocos (Figura 30).

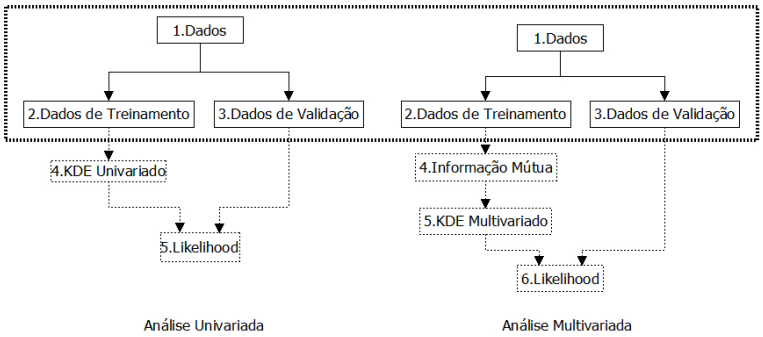

Figura 30: Diagrama de blocos das Análises Univariada e Multivariada.

Na análise Multivariada, o problema da dependência é colocado em foco. Portanto, existe a necessidade da inclusão do algoritmo de Informação Mútua para quantizar esse nível de dependência.

Os blocos inerentes ao entendimento das análises serão explicados nas seções posteriores.

5.2.1 ALGORITMO DE SELEÇÃO DE DADOS

O Algoritmo de Seleção de dados é formado pelos três blocos iniciais, *1.Dados*, *2.Dados de Treinamento* e *3.Dados de Validação*. Esses três blocos desempenham funções simples e essenciais para a validação da análise. Nestes blocos, garantimos que os eventos que treinam nosso algoritmo são diferentes dos eventos validados por ele.

Os três blocos iniciais são exatamente iguais nas análises, univariada e multivariada (como mostra a região destacada no diagrama da Figura 30). Portanto, os algoritmos de seleção de dados, serão tratados de forma conjunta entre as análises.

O primeiro algoritmo da cadeia é denotado pelo bloco *1.Dados*. Para entendê-lo melhor a Figura 31 evidencia suas principais funções.

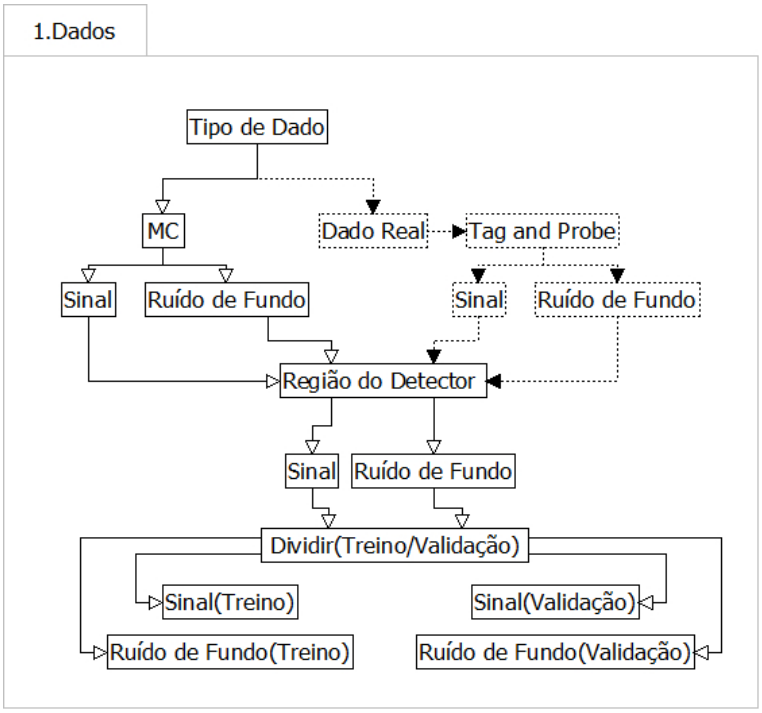

Figura 31: Diagrama do bloco *1.Dados*.

Como visto anteriormente, existem dois tipos de conjunto de dados: dados reais e dados de simulação MC.

O primeiro passo do algoritmo é definir qual tipo de conjunto de dados analisar. Caso seja os dados de simulação MC, segue a cadeia representada pelas setas contínuas.

Como esses dados são simulados, sabemos exatamente quais eventos são os sinais verdadeiros, decaídos de uma partícula de interesse, e quais são ruído de fundo. Portanto, torna-se simples separar os conjuntos de sinal e ruído de fundo (verificando as variáveis el_truth_type e el_truth_mothertype).

Logo após, é preciso decidir qual região do detector analisar, ou seja, qual região em η e E_t o algoritmo deve trabalhar, essa região foi definida anteriormente como *bin* e podem ser vistas na tabela 7.

Após a escolha da região do detector, sinal e ruído de fundo são limitados dentro da faixa desejada em η e E_t, com isso é possível dividir o conjunto total daquele *bin* em conjuntos de Treino e Validação, para sinal e ruído de fundo.

Nessa fase o algoritmo garante que os eventos para treino serão diferentes dos

eventos validados.

Porém, caso a escolha seja trabalhar com dados reais, ao invés seguir as setas contínuas, é preciso continuar a sequência mostrada pelas setas pontilhas, no diagrama da Figura 31, sendo necessário utilizar o algoritmo de *Tag and Probe* na separação dos conjuntos de sinal e ruído de fundo. O algoritmo *Tag and Probe* é mostrado na Figura 32:

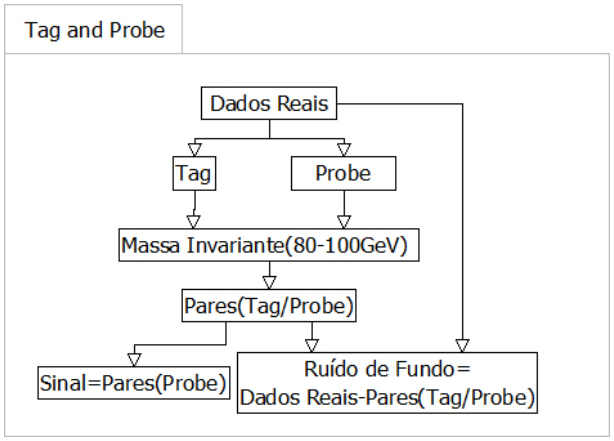

Figura 32: Diagrama do bloco *Tag and Probe*.

Dentro de um conjunto de eventos de dados reais é preciso selecionar dois conjuntos: *Tag* e *Probe*. As características específicas de cada conjunto foram vistas na Seção 5.1.2.1.

Ao selecionar os conjuntos é preciso combinar seus eventos na fórmula da massa invariante, neste caso, a proposta é encontrar a massa invariante do bóson Z, que está em torno de 90 GeV, selecionando apenas os pares entre $80 - 100$ GeV.

Com o conjunto de pares, são selecionados os eventos *Probe* para ser sinal (por ser um conjunto menos polarizado, devido aos critérios de seleção) e para ser o ruído de fundo são utilizados todos os eventos, exceto os pares *Tag* e *Probe*.

Após a separação de sinal e ruído de fundo o algoritmo retorna para etapa região do detector, descrita anteriormente.

O bloco *2.Dados de Treinamento* recebe os conjuntos de treino do bloco 1 e faz uma correção (ou adaptação) nos dados, para serem enviados ao algoritmo de KDE. O diagrama do bloco 2 é mostrado na Figura 33.

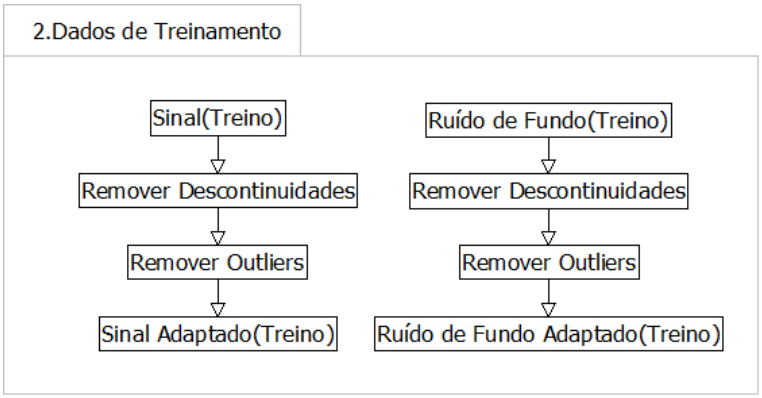

Figura 33: Diagrama do bloco *2.Dados de Treinamento*.

O sinal e ruído de fundo de treino são tratados separadamente, pois seus *outliers* não ocorrem nas mesmas regiões. Portanto, como primeira adaptação, todas as descontinuidades são removidas, sendo possível visualizar melhor as distribuições, por último os *outliers* são removidos das distribuições. Esses ajustes são feitos como indicado em (COLLABORATION et al., 2013). Neste ponto, existe um conjunto de treino pronto para ser transformado em PDF, pelo algoritmo KDE.

O bloco *3.Dados de Validação* é mostrado na Figura 34.

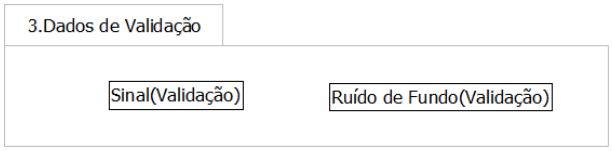

Figura 34: Diagrama do bloco *3.Dados de Validação*.

Este bloco foi separado em sinal e ruído de fundo, de validação, para facilitar a implementação da curva ROC, e será utilizado na verificação e comparação do funcionamento entre o algoritmo proposto nessa dissertação e o algoritmo proposto pela colaboração ATLAS.

5.2.2 ALGORITMO DA ANÁLISE UNIVARIADA

Foram definidos os conjuntos de treino adaptado e validação, que serão utilizados como parâmetros de entrada da nossa análise univariada.

Essa análise é definida como univariada devido a construção de seu KDE, ou seja, é construída uma PDF para cada dimensão ou variável discriminante, onde sua verossimilhança é dada pela multiplicação dessas PDFs unidimensionais. Nessa etapa a

verossimilhança é construída como a proposta da colaboração, através da simplificação vista na Seção 3.3.2.

Através do conjunto de treino adaptado, mostrado anteriormente, as entradas do bloco *4.Kernel N Dimensional* são obtidas. Esse algoritmo foi construído e generalizado pensando nas análises univariada e multivariada. Por exemplo, ao definir a variável $D = 1$ o algoritmo utilizará apenas PDFs unidimensionais na construção da PDF conjunta, o diagrama da Figura 35 descreve o processo.

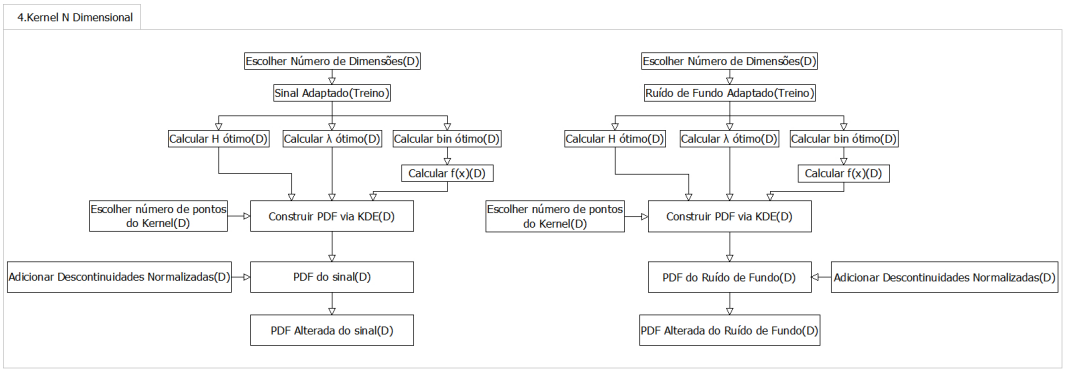

Figura 35: Diagrama do bloco *4.Kernel N Dimensional*.

O diagrama representa a construção da PDF para cada variável, ou seja, se o problema tiver 13 variáveis o algoritmo se repetirá 13 vezes. A etapa de construção das PDFs através do KDE é dividida em sinal adaptado na esquerda, e ruído de fundo adaptado na direita, onde cada variável é utilizada na construção de sua respectiva PDF, pra ruído de fundo e sinal.

Sabendo com qual conjunto trabalhar, formulações matemáticas como as descritas na Seção 4.2.3.1 utilizarão a estatística de cada variável para calcular seus parâmetros, responsáveis por ajustar a largura de banda h para cada variável aleatória. Através das teorias, é possível obter o h variável e o λ do nosso algoritmo, restando utilizar o histograma, com binagem "ótima", para calcular o valor de $f(x_i)$ de cada ponto que desejamos estimar. Definido o número de pontos da PDF que desejamos estimar, e os parâmetros calculados anteriormente, podemos construir a PDF através do KDE.

Após a estimação da PDF é preciso adicionar as descontinuidades removidas no bloco 2, que geralmente são picos acentuados em valores como 0, 1 e -9999. Como o KDE utiliza a característica estatística das variáveis aleatórias, esses picos distorcem os valores de média e desvio padrão, adicionando um erro no cálculo dos parâmetros iniciais do KDE. Por fim, as descontinuidades são adicionadas através da inserção desses picos acentuados normalizados pela área da PDF em foco. Muitas dessas des-

continuidades são geradas por questões numéricas (como razões ou falta de resolução) e precisam ser consideradas nas PDFs que serão utilizadas na reconstrução da probabilidade conjunta. As PDFs alteradas, de cada variável, serão utilizadas no algoritmo da verossimilhança, sendo representado pelo bloco 5 na Figura 36.

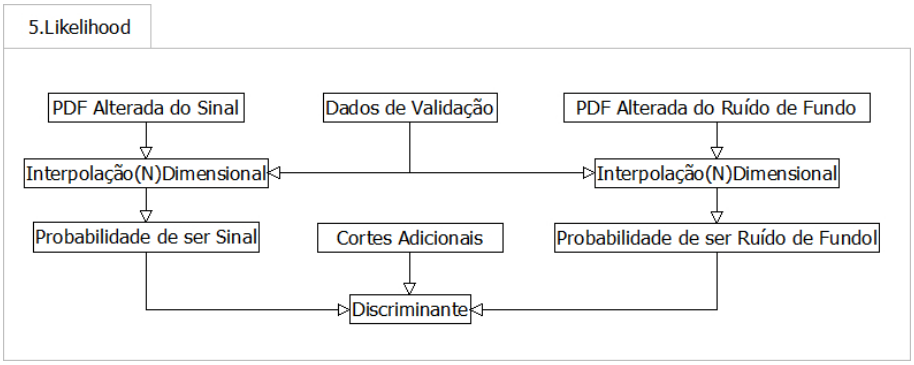

Figura 36: Diagrama do algoritmo da *Likelihood*.

O bloco 5 representa a parte final da nossa análise, onde existe a definição de qual grupo o evento pertence: sinal ou ruído de fundo.

Essa etapa têm como parâmetros de entrada as PDFs alteradas de sinal e ruído de fundo, as variáveis de cortes adicionais e os dados de validação. Cada evento dos dados de validação precisa quantificar a sua probabilidade de ser sinal, fazendo uma inferência estatística na PDF de sinal, através de uma interpolação linear. O mesmo evento precisa conhecer sua probabilidade de ser ruído de fundo, através do mesmo processo, na PDF de ruído de fundo. Note que, para cada evento existe um valor de sinal e ruído de fundo, para cada variável aleatória do problema.

Na análise de cada evento, cada valor de probabilidade de sinal encontrado em cada variável aleatória será multiplicado para formar a verossimilhança de sinal (L_s) , composta por todas as variáveis aleatórias do respectivo menu da tabela 6, a mesma metodologia acontece com a verossimilhança de ruído de fundo (L_b). As verossimilhanças L_s e L_b serão combinadas em um discriminante, mostrado na equação 3.7. O discriminante dL retornará um valor de 0 a 1 que classificará, de acordo com um limiar, a qual grupo o evento pertence, se sinal ou ruído de fundo.

Os cortes adicionais são definidos antes do discriminante dL, se o evento não for aprovado por eles, será considerado ruído de fundo automaticamente.

5.2.3 ALGORITMO DA ANÁLISE MULTIVARIADA

A análise multivariada foi proposta como alternativa ao problema de dependência entre as variáveis discriminantes. Cientes do problema, o grupo de performance responsável pela verossimilhança do ATLAS não utiliza algumas variáveis, como visto em (COLLABORATION et al., 2013).

Como descrito na seção 4.1, a informação mútua se torna uma ferramenta adequada a quantificação dessa dependência. O Algoritmo de informação mútua é uma das diferenças entre as análises uni e multivariada, e sua utilização será entendida com o auxílio da Figura 37, que representa o bloco 4.

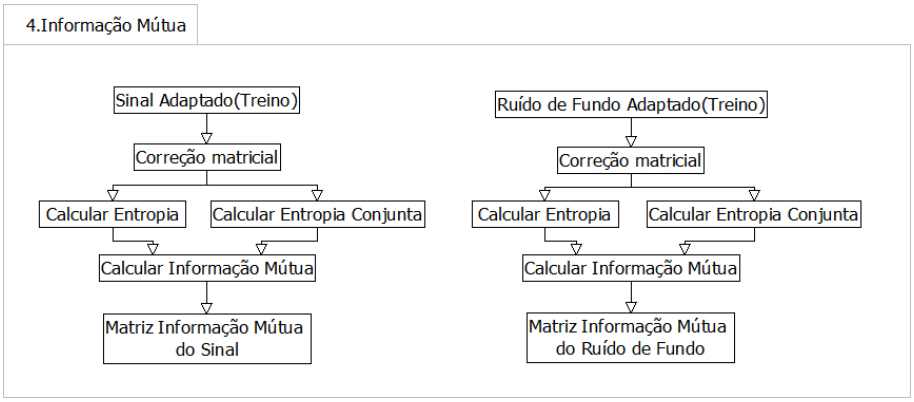

Figura 37: Diagrama do algoritmo de Informação Mútua.

Na seqüência do algoritmo de seleção de dados, o sinal adaptado e o ruído de fundo adaptado são utilizados como parâmetro de entrada do bloco *4.Informação Mútua*, a utilização desses conjuntos ocorre devido as descontinuidades, mencionadas anteriormente, diminuírem a resolução do algoritmo de informação mútua.

O conjunto (sinal e ruído de fundo) adaptado, de treino, é composto por 13 variáveis discriminantes com n eventos. Cada variável, após o algoritmo de remoção de *outliers* e descontinuidades, utilizado individualmente por variável, fica com um número diferente de eventos. Esse fato introduz um problema na análise, visto que, a obtenção da probabilidade conjunta das 13 variáveis depende da existência de valores em todas as dimensões durante a multiplicação, ou seja, é necessário que o evento tenha projeção em todas as dimensões da verossimilhança.

Como o conjunto adaptado de treino requer, para cada evento, um valor em cada variável discriminante, existe a necessidade de escolher somente os eventos que possuem projeções nas 13 variáveis, formando uma matriz. Essa matriz com 13-variáveis x n-eventos, é utilizada para calcular a entropia de cada uma das variáveis e a entropia

conjunta de 2 em 2 variáveis. Com esses valores, a informação mútua é calculada, como explica a Seção 4.1, sendo possível preencher uma matriz de informação mútua 13x13.

A matriz de informação mútua mostra o nível de dependência entre as variáveis aleatórias, sendo possível, teoricamente, classificar quais das variáveis propagam um maior erro de estimação, devido a simplificação do método de verossimilhança, durante a reconstrução da probabilidade conjunta.

A Figura 38 demonstra o método utilizado na escolha dos pares de variáveis com maior informação mútua dentre todas as possibilidades. O par ou pares escolhidos serão adicionados a nova proposta de verossimilhança, tendo suas respectivas variáveis constituintes, unidimensionais, removidas do cálculo da verossimilhança.

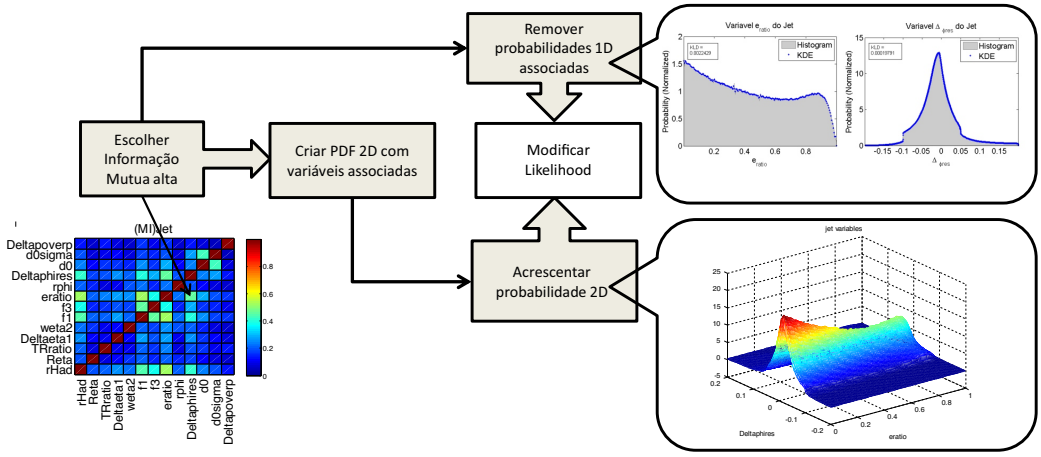

Figura 38: Fluxograma do uso da Informação mútua na escolha de variáveis dependentes para uso do MKDE.

O ideal seria utilizar todas as 13 variáveis na construção de uma PDF conjunta com 13 dimensões. Porém, é sabido que os dados se tornam cada vez mais esparsos com o aumento dimensional, criando o seguinte problema: Embora seja possível uma melhor estimação da PDF conjunta de variáveis dependentes via MKDE, ao invés da simplificação de verossimilhança, o aumento dimensional requer um aumento estatístico suficiente na estimação da PDF n dimensional, para a mesma qualidade de estimação na PDF com *n-1* dimensões.

Com a matriz de informação mútua, é possível definir quais PDFs tem maior dependência e criar pares, considerando a matriz de informação mútua de sinal e ruído de fundo. Sendo esse, um parâmetro fornecido ao bloco 5, na construção do vetor de

dimensões.

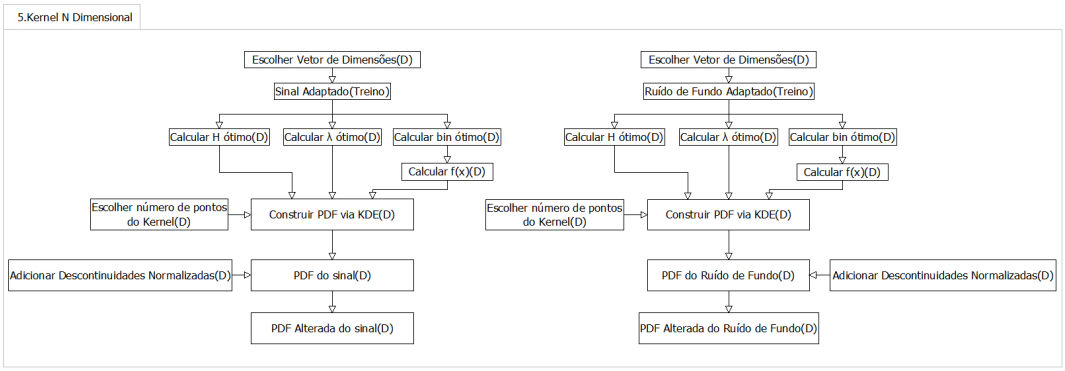

Figura 39: Diagrama do Kernel N-dimensional.

O vetor de dimensões define quais pares de variáveis devemos utilizar na construção de PDFs bivariadas, e quais PDFs devem continuar univariadas. Esse artifício possibilita uma diminuição do erro de estimação na verossimilhança, causado pelas variáveis com maior dependência. Definido o vetor de dimensões, o algoritmo acessa o conjunto adaptado de treino e seleciona qual foi o arranjo definido para a verossimilhança, ou seja, quais pares de variáveis discriminantes serão parâmetro de entrada para o MKDE construir PDFs bivariadas, e quais variáveis continuarão unidimensionais nesse arranjo. O mesmo processo é descrito na seção univariada, com alguns ajustes nos parâmetros, para generalizá-los a uma nova realidade n-dimensional, como explica a Seção 4.2.3.1.

Com a verossimilhança utilizando PDFs bivariadas e univariadas, o algoritmo segue para o bloco *6.Likelihood*, Figura 40 :

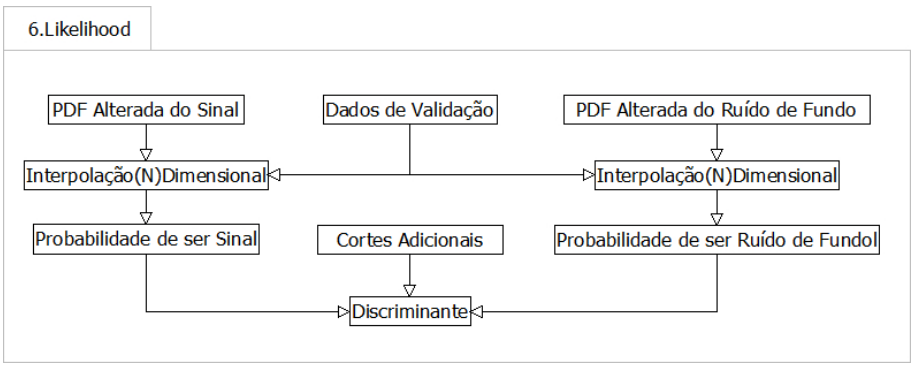

Figura 40: Diagrama do Kernel N-dimensional.

Como diferença entre a análise anterior, univariada, existe a interpolação multidimensional. Eventos que obtinham suas probabilidades através de PDFs univariadas utilizam, agora, projeções em duas dimensões, ou variáveis, para encontrar um único valor de probabilidade, que será adicionado na verossimilhança análogamente aos valores das PDFs univariadas das outras variáveis. Nesse novo contexto, a verossimilhança,

para cada evento, será criada pela multiplicação de valores de probabilidade oriundos de PDFs unidimensionais e bidimensionais, respeitando o pressuposto que, os pares de variáveis, selecionados pelo algoritmo de informação mútua, constituem apenas um valor de probabilidade (bivariada), portanto, seus valores de probabilidade univariados não podem ser utilizados, na construção da verossimilhança.

Essa lógica é respeitada para verossimilhança de sinal e ruído de fundo, que são combinadas em um discriminate dL, que varia de 0 a 1, respeitando um certo limiar de decisão, que indica a qual grupo o evento pertence, sinal ou ruído de fundo.

Os cortes adicionais são aplicados antes do discriminante dL, os valores reprovados por esses cortes são considerados ruído de fundo.

6 RESULTADOS

Nesse capítulo serão apresentados os resultados obtidos utilizando Dados de Simulação e Dados Reais, para as análises univariada e multivariada. Na análise univariada, o algoritmo e\γ será utilizado como referência na comparação com o método de verossimilhança univariada, que será mostrada nos gráficos como *Likelihood* (LH). Na análise Multivariada, a verossimilhança univariada será comparada com a verossimilhança multivariada.

Para comparação dos métodos foram utilizadas algumas medidas de performance bastante comuns na literatura: ROC (METZ, 1978) e o critério SP, utilizado na discriminação binária, onde seu resultado denota o equilíbrio entre a detecção de sinal e ruído de fundo. Sua fórmula é dada por:

$$SP = 100\% \sqrt{\sqrt{DET_{sinal} DET_{ruido}} \left(\frac{DET_{sinal} + DET_{ruido}}{2} \right)} \tag{6.1}$$

Onde, DET_{sinal} é a probabilidade de detecção de sinal e DET_{ruido} é a probabilidade de detecção de ruído.

6.1 SIMULAÇÃO MONTE CARLO

As características principais dos dados simulados, do conjunto de sinal e ruído de fundo, serão mostrados na Figura 41. Estes conjuntos estão em pacotes separados, Zee (sinal) e JF17 (ruído de fundo), portanto é possível aumentar a quantidade de elétrons isolados em relação ao ruído de fundo, apenas adquirindo mais pacotes de Zee do que JF17, uma vez que, elétrons isolados são estatisticamente mais raros do que ruído de fundo. Esse artifício nos permite obter estatística suficiente do conjunto de sinal necessária para segmentar a análise em η e Et, evidenciando a dependência dos métodos e variáveis discriminantes em diferentes energias e regiões do detector.

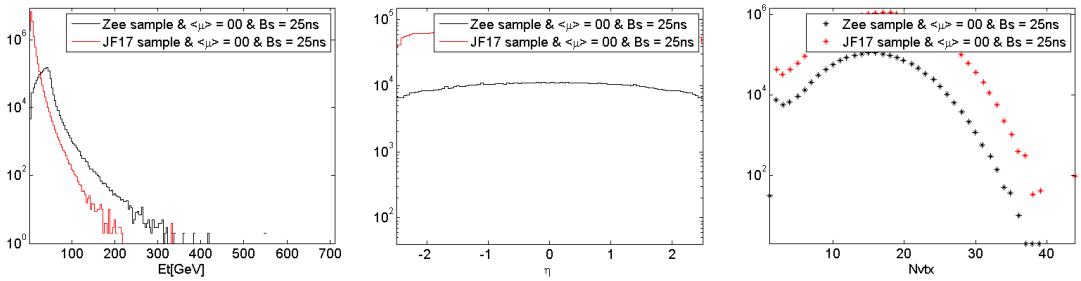

Figura 41: Perfil dos eventos de dados MC. (Esquerda) Gráfico de eventos por E_t, (Centro) Gráfico de eventos por η e (Direita) Gráfico de eventos por NVTX.

6.1.1 ANÁLISE UNIVARIADA

A análise univariada, com dados MC foi dividida em 12 regiões, como mostra a Tabela 7. Essas regiões foram decididas levando em consideração os padrões utilizados pelo e\γ nas regiões de η (sem as segmentações em 0.6, 1.81 e 2.37) e as regiões de E_t foram escolhidas de acordo com a estatística disponível.

Tabela 7: Tabela de divisão de regiões em η e E_t.

Regiões de Estudo				Número de Eventos	
n	η		E_t (GeV)	Sinal	Ruído de Fundo
1			5 - 20	98.974	2.222.932
2	0	0.8	20 - 30	142.168	120.237
3			> 30	489.638	45.339
4			5 - 20	66.264	1.600.333
5	0.8	1.37	20 - 30	98.404	83.452
6			> 30	329.811	31.008
7			5 - 20	61.774	1.444.673
8	1.52	2.01	20 - 30	99.586	81.168
9			> 30	325.878	28.938
10			5 - 20	32.599	845.501
11	2.01	2.47	20 - 30	55.539	51.550
12			> 30	204.215	18.689
Total				2.004.850	6.573.820

Nesta etapa será feita uma comparação entre o e\γ e a verossimilhança univariada. A eficiência e falso alarme de cada ponto de operação (LoosePP, MediumPP e TightPP)

do e\γ podem ser obtidos à partir do pacote de dados disponibilizado pela Colaboração ATLAS.

6.1.2 ESTIMAÇÃO DE PDFS UNIVARIADAS

A estimação das PDFs univariadas foram feitas pelo algoritmo KDE desenvolvido no âmbito dessa dissertação. A otimização dos parâmetros do KDE é um desafio, visto que a consideração inicial (função geradora normal) para o cálculo do comprimento de banda h utilizado pelo MISE, não é a realidade do problema.

No intuito de otimizar a estimação das PDFs, efetuou-se uma análise onde ao invés de utilizar as 13 variáveis para construção da probabilidade conjunta, retirou-se uma variável por vez, construindo assim 13 probabilidades conjuntas contendo cada uma 12 variáveis em sua estrutura e em seguida mensurou-se a performance de cada uma pela sua respectiva ROC, deu-se o nome a esse método de 'Análise $n-1$'.

A ideia básica dessa técnica consiste em retirar uma variável do discriminante e observar se sua saída degrada ou melhora o resultado. No caso de degradar o resultado, pode-se concluir que essa variável aumenta o poder de discriminação do algoritmo, não sendo necessária sua otimização. Já quando a retirada da mesma ocasiona melhora na ROC, entende-se que é necessário a busca por uma melhor estimação de sua PDF, feito a partir do ajuste fino do parâmetro λ.

Com isso, consegue-se melhorar a resposta do algoritmo, fazendo com que a melhor discriminação fosse efetuada com o uso das 13 variáveis. A Figura 42 mostra a Análise $n-1$ para a região 1, após a otimização do λ. Embora este método tenha contribuído para melhorar o algoritmo de identificação de elétrons em todas as regiões, algumas dessas não puderam ser completamente otimizadas, como mostra a Figura 43, onde pode-se ver que a ROC, quando retirada a variável $W_{\eta 2}$, apresenta uma melhora em relação ao uso das 13 variáveis. Este fato indica que além do parâmetro citado, outros aspectos podem degradar o resultado do algoritmo, como dependência entre as variáveis ou mesmo estatística insuficiente pra descrever de forma fiel a realidade do problema.

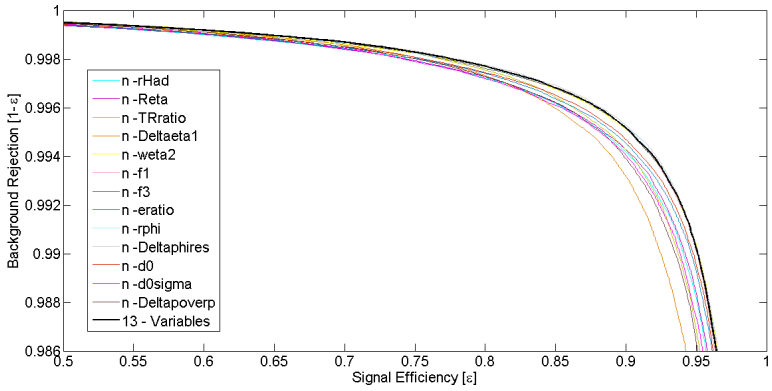

Figura 42: Gráfico da curva ROC removendo 1 variável por vez (mostrada na legenda) e construindo a LH utilizando as 12 variáveis que restaram. Para a região 1.

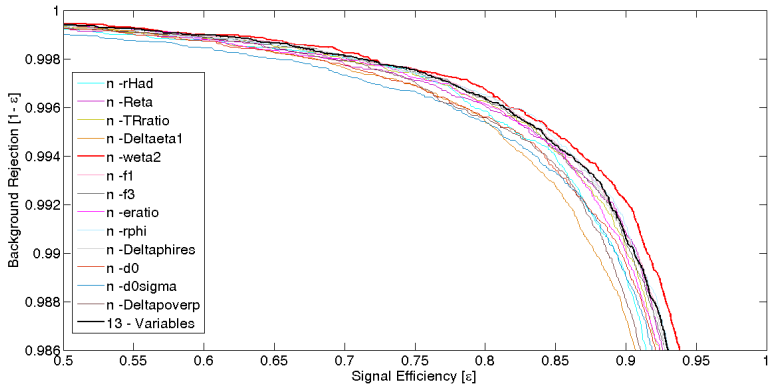

Figura 43: Gráfico da curva ROC removendo 1 variável por vez (mostrada na legenda) e construindo a LH utilizando as 12 variáveis que restaram. Para a região 8.

As Figuras 44 até 57 mostram as PDFs estimadas pelo algoritmo de KDE na região 1, para o conjunto de sinal (denominado *electron*, por ser formado de elétrons isolados) e o conjunto de ruído de fundo (denominado *Jet*, pela predominância de jatos), utilizando os parâmetros λ ajustados pela análise anterior.

A escolha da melhor binagem para visualização do histograma foi feita utilizando a minimização do χ^2, que foi calculado entre a PDF estimada pelo KDE e o histograma, com sua binagem variando de 2 a 2500 *bins*.

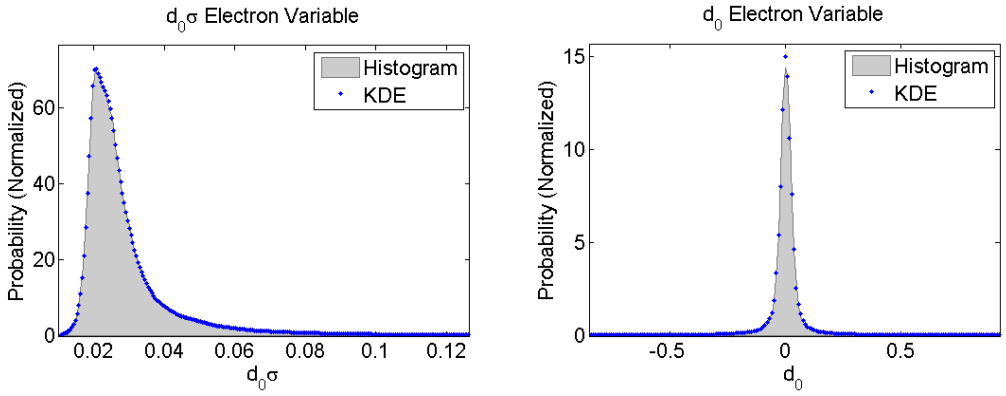

Figura 44: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $d_{0\sigma}$ e (Direita) Variável d_0.

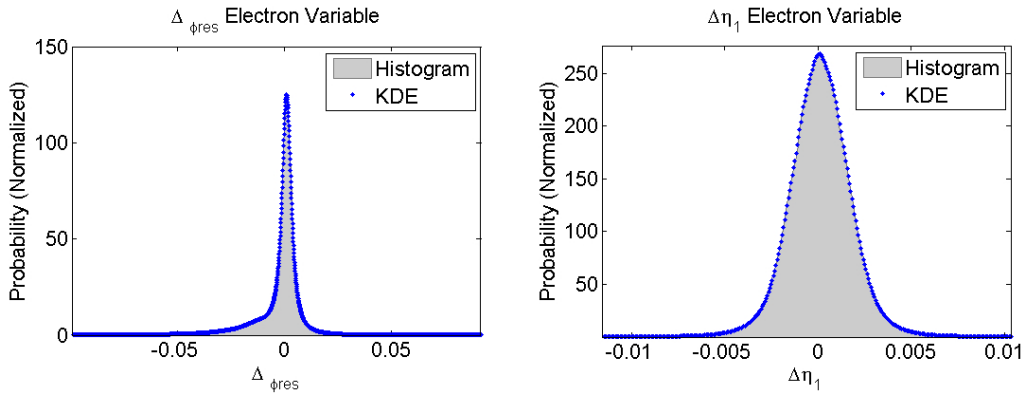

Figura 45: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $\Delta_{\phi res}$ e (Direita) Variável $\Delta_{\eta 1}$.

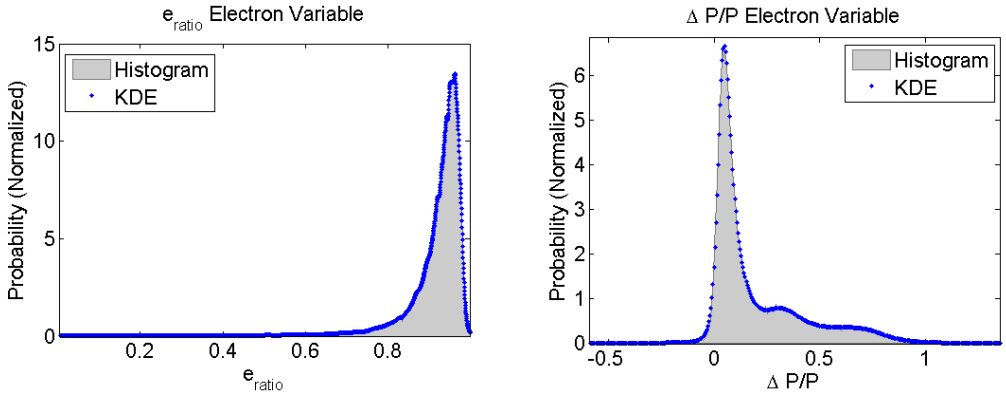

Figura 46: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável E_{ratio} e (Direita) Variável $\Delta P/P$.

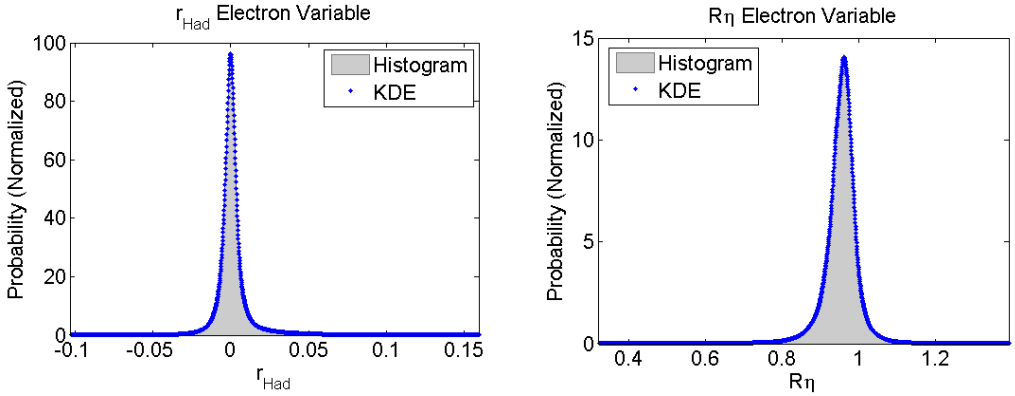

Figura 47: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável R_{Had} e (Direita) Variável r_{η}.

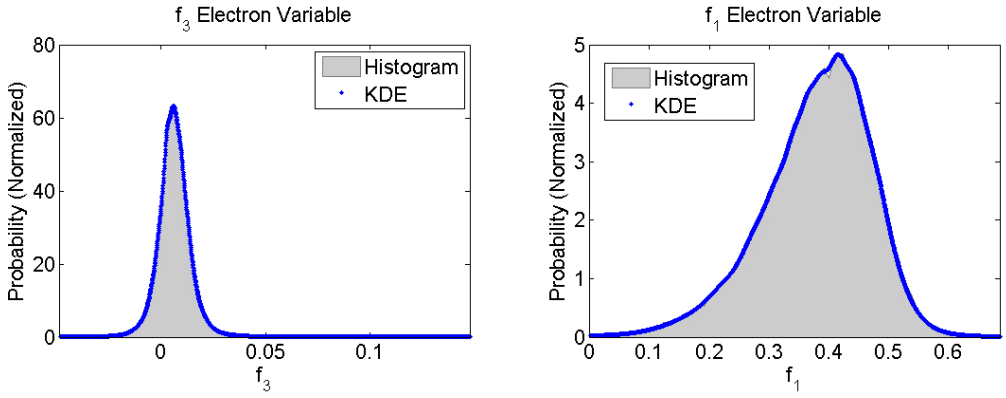

Figura 48: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável f_3 e (Direita) Variável f_1.

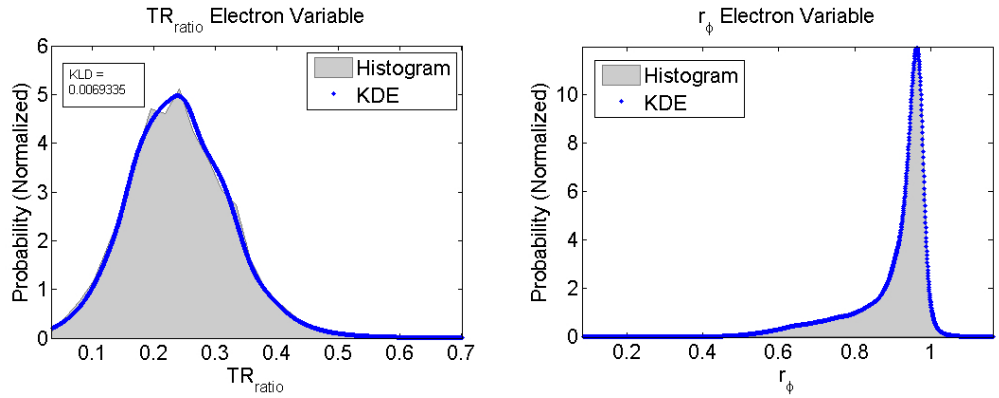

Figura 49: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável TR_{ratio} e (Direita) Variável r_{ϕ}.

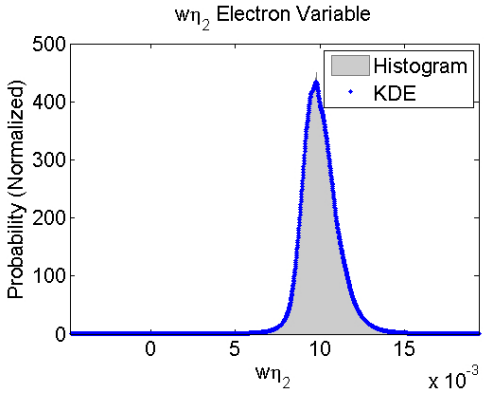

Figura 50: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). Variável $W_{\eta 2}$.

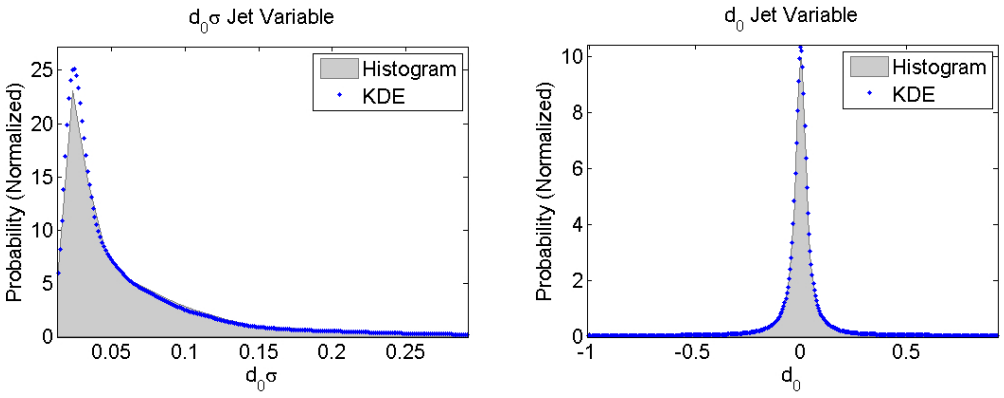

Figura 51: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $d_{0\sigma}$ e (Direita) Variável d_0.

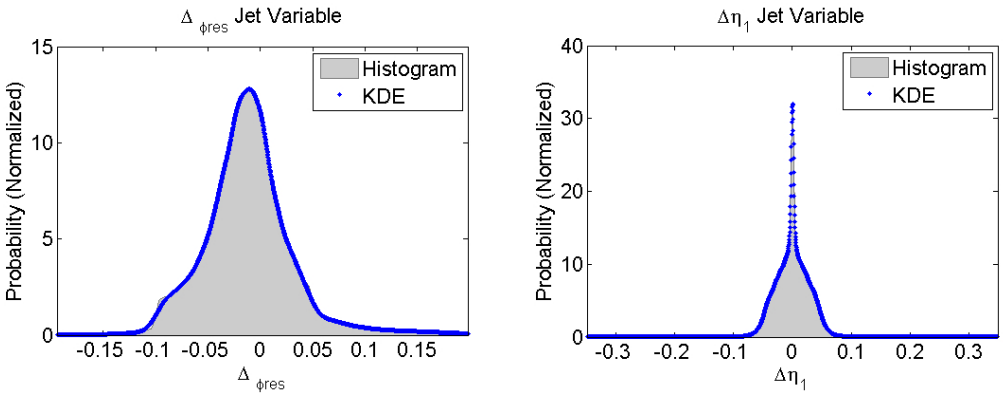

Figura 52: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $\Delta_{\phi res}$ e (Direita) Variável $\Delta_{\eta 1}$.

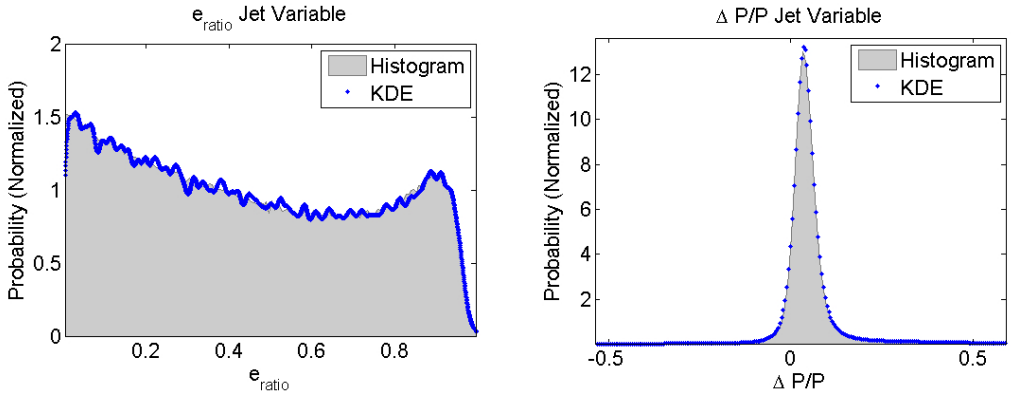

Figura 53: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável E_{ratio} e (Direita) Variável $\Delta P/P$.

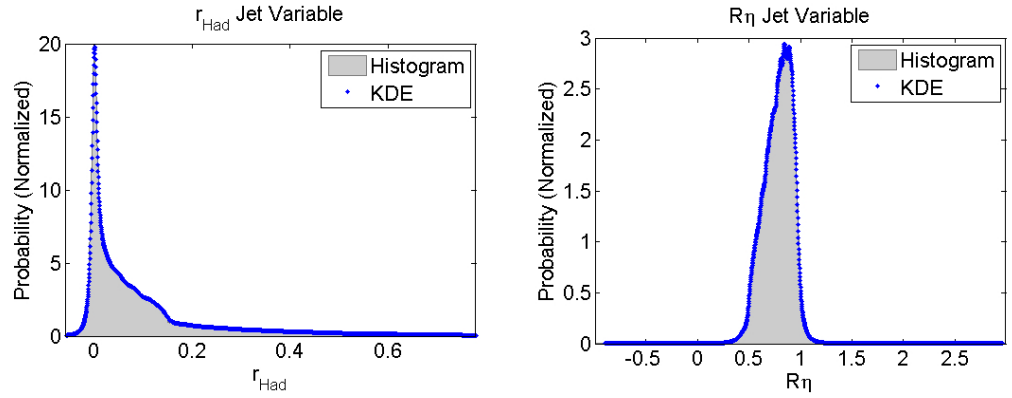

Figura 54: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável R_{Had} e (Direita) Variável r_η.

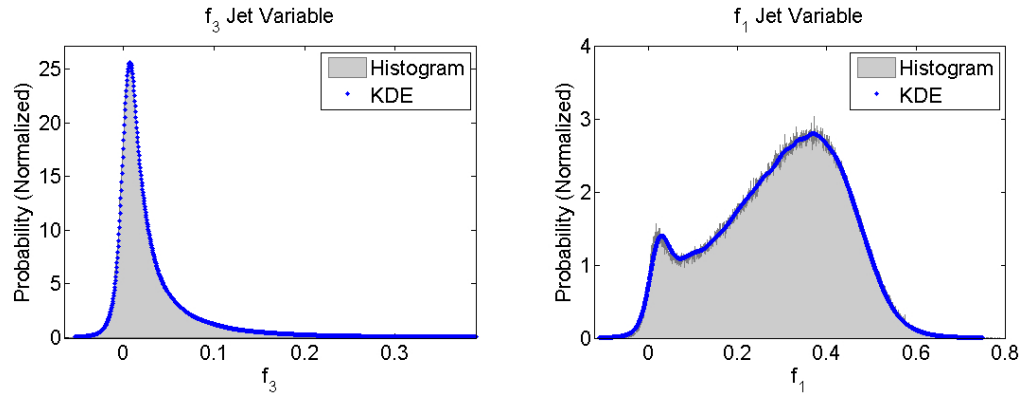

Figura 55: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável f_3 e (Direita) Variável f_1.

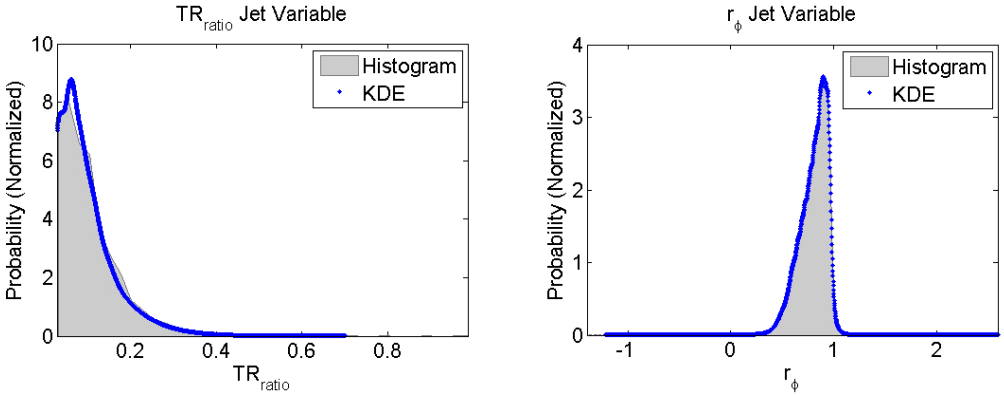

Figura 56: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável TR_{ratio} e (Direita) Variável r_{ϕ}.

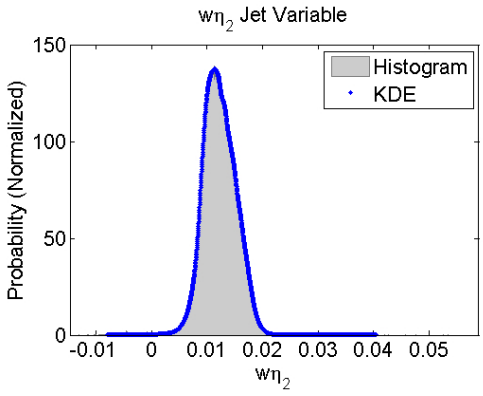

Figura 57: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). Variável $W_{\eta 2}$.

Utilizando as PDFs vistas anteriomente, é possível encontrar a probabilidade de cada evento ser sinal e ruído de fundo em cada uma das variáveis. Esses valores serão combinados num discriminante, como descrito na Seção 3.3.2, com isso, poderemos gerar as curvas ROC da análise.

6.1.3 RESULTADOS DA ANÁLISE UNIVARIADA COM DADOS MC

Como a verossimilhança é comparada aos pontos de operação do e\$\gamma$ utilizando diferentes menus, como mostrado na Seção 3.3.2, faz-se necessário a construção de um menu com 13 variáveis (utilizado para comparação com os pontos de operação *Tight* e *Medium* do e\$\gamma$) e um menu com 11 variáveis (utilizado para comparação com o ponto de operação *Loose* do e\$\gamma$). Na Figura 58 são mostradas as curvas ROC geradas por cada menu, na região 1.

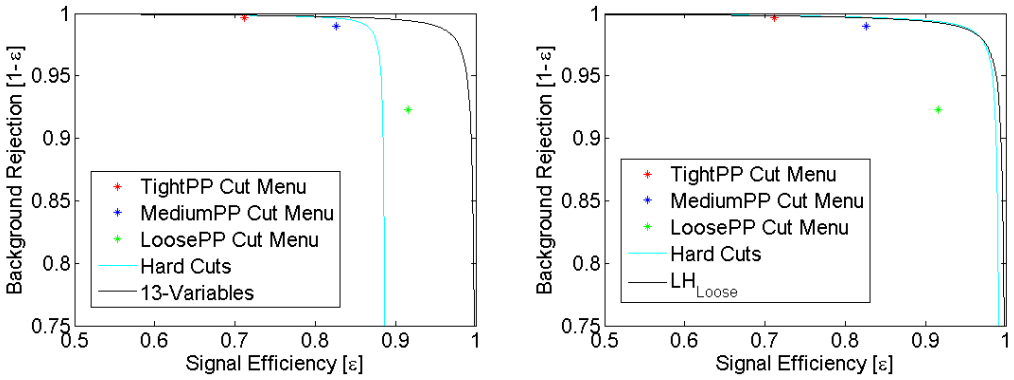

Figura 58: Curva ROC. (Esquerda) Menu utilizado para comparação da verossimilhança com os pontos de operação *Tight* e *Medium* do e\γ e (Direita) Menu utilizado para comparação da verossimilhança com o ponto de operação *Loose* do e\γ.

O impacto dos *hard cuts* nas curvas ROC são mostrados na Figura 58. Os *hard cuts* utilizam algumas variáveis de traço para aumentar a rejeição de ruído de fundo, como explicado na Seção 3.2.2. Entretanto, este aumento não ocorre em todas as regiões, como mostrado na Figura 59 à esquerda, onde pode-se observar que a ROC da verossimilhança apresenta um resultado melhor que a ROC da verossimilhança com a adição dos *hard cuts*. Na mesma figura à direita percebemos que os *hard cuts* melhoram a rejeição de ruído de fundo da verossimilhança em relação ao *Tight PP* do e\γ, porém, perde em eficiência e rejeição de ruído de fundo em relação ao *medium PP*. Pelos resultados obtidos, infere-se que nos casos onde se têm mais estatística e consequentemente PDFs melhores representadas, a adição dos *hard cuts* pode ser desconsiderada.

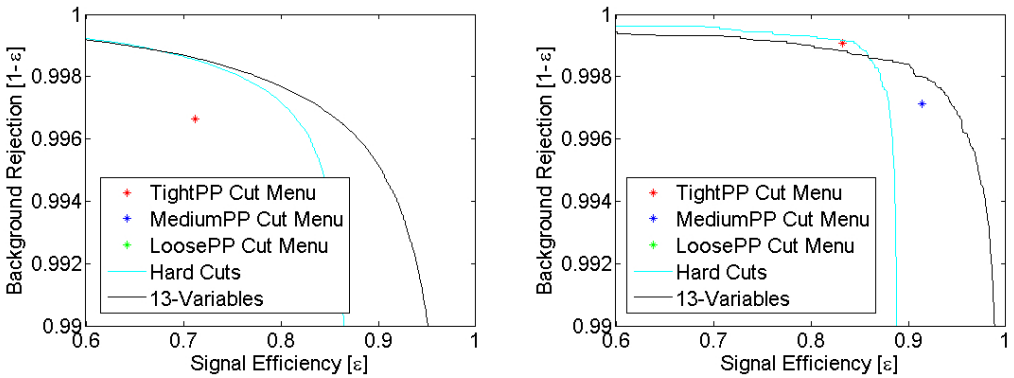

Figura 59: Zoom na Curva ROC. (Direita) Região 1 e (Esquerda) Região 2.

A Tabela 8 mostra uma comparação entre os pontos de operação da verossimilhança e o e\γ fixando a Eficiência de Sinal, bem como o Índice SP dos mesmos, em

$0 \leq |\eta| < 0.8$ na energia de 5 GeV$<E_t<$20 GeV. Pode-se ver que usando método da verossimilhança consegue-se uma Eficiência de Sinal melhor que do ponto de operação *Loose* mantendo uma Rejeição de Ruído de Fundo semelhante ao ponto de operação *Tight*.

Tabela 8: Eficiência de Sinal e Rejeição de Ruído de Fundo para a verossimilhança e o e\γ, para a Região 1 - $0 \leq |\eta| < 0.8$ e $5 \leq E_t < 20GeV$, fixando a Eficiência de Sinal.

Menu	Região 1 - $0 \leq \|\eta\| < 0.8$ e $5 \leq E_t < 20GeV$						
	Índice SP (%)	Eficiência de Sinal (%)			Rejeição de Ruído(%)		
TightPP Cuts	84,82	71,17	+0,40	-0,40	99,66	+0,01	-0,01
Tight LH	84,91	71,17	+0,40	-0,40	99,86	+0,01	-0,01
MediumPP Cuts	90,62	82,64	+0,33	-0,34	98,97	+0,02	-0,02
Medium LH	90,98	82,64	+0,33	-0,34	99,73	+0,01	-0,01
LoosePP Cuts	91,94	91,62	+0,24	-0,25	92,26	+0,05	-0,05
Loose LH	95,48	91,62	+0,24	-0,25	99,42	+0,01	-0,01

A Tabela 9 apresenta o resultado para mesma região da Tabela 8, porém ao invés de se fixar a eficiência de sinal, fixa-se a Rejeição de Ruído de Fundo. Nessa comparação a diferença entre os dois métodos fica mais evidente, confirmando que a verossimilhança demonstra melhores resultados.

Tabela 9: Eficiência de Sinal e Rejeição de Ruído de Fundo para a verossimilhança e o e\γ, para a Região 1 - $0 \leq |\eta| < 0.8$ e $5 \leq E_t < 20GeV$, fixando a Rejeição de Ruído de Fundo.

Menu	Região 1 - $0 \leq \|\eta\| < 0.8$ e $5 \leq E_t < 20GeV$						
	Índice SP (%)	Eficiência de Sinal (%)			Rejeição de Ruído(%)		
TightPP Cuts	84,82	71,17	+0,40	-0,40	99,66	+0,01	-0,01
Tight LH	92,57	85,74	+0,31	-0,31	99,66	+0,01	-0,01
MediumPP Cuts	90,62	82,64	+0,33	-0,34	98,97	+0,02	-0,02
Medium LH	97,10	95,25	+0,19	-0,19	98,97	+0,02	-0,02
LoosePP Cuts	91,94	91,62	+0,24	-0,25	92,26	+0,05	-0,05
Loose LH	95,74	99,29	+0,07	-0,08	92,26	+0,05	-0,05

Os resultados de Eficiência de Sinal e Rejeição de Ruído de Fundo, da verossimilhança e do e\γ, para os 3 pontos de operação (*LoosePP*, *MediumPP* e *TightPP*) das 12 regiões propostas, podem ser encontrados no Apêndice A.

Nas Figuras 60 a 62 temos uma análise de Eficiência *vs* η (à esquerda) e Falso Alarme vs η (à direita), para os pontos de operação *LoosePP*, *MediumPP* e *TightPP*,

respectivamente . No gráfico da esquerda foi fixado o Falso Alarme médio e no gráfico da direita a Eficiência média, dos pontos de operação do e\γ de cada região de η vistas na Tabela 7.

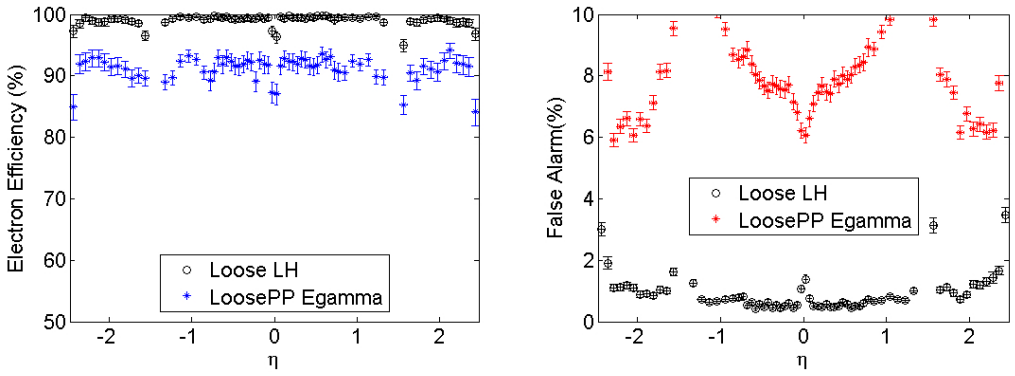

Figura 60: Gráfico de η no ponto de operação *Loose*, comparando LH e e\γ, com 5 GeV$<E_t<$20 GeV.

Como mostrado na Figura 60 à esquerda, o algoritmo de verossimilhança apresenta um ganho de eficiência de aproximadamente 10%, quando comparado com o *Loose PP*, fixados no mesmo falso alarme. Já quando a eficiência é fixada, temos uma diminuição do falso alarme de aproximadamente 6%, como mostrado à direita.

Nas Figuras 61 e 62, percebe-se à esquerda, que obtivemos um ganho de 12% em relação ao *Medium PP* e 10% em relação ao *Tight PP* em eficiência, nessas mesmas figuras, à direta, é mostrado a diminuição de 1% no *Medium PP* e 0,3% no *Tight PP* em falso alarme. Os gráficos para as outras regiões de E_t, estão disponíveis no Apêndice B.1.

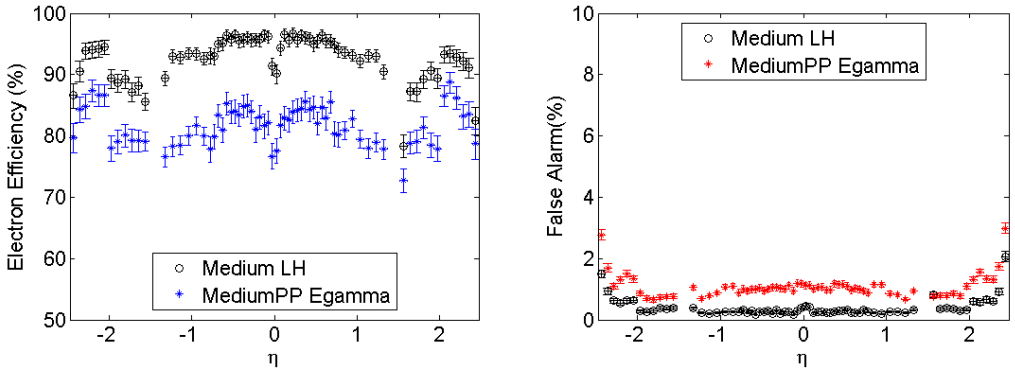

Figura 61: Gráfico de η no ponto de operação *Medium*, comparando LH e e\γ, com 5 GeV$<E_t<$20 GeV.

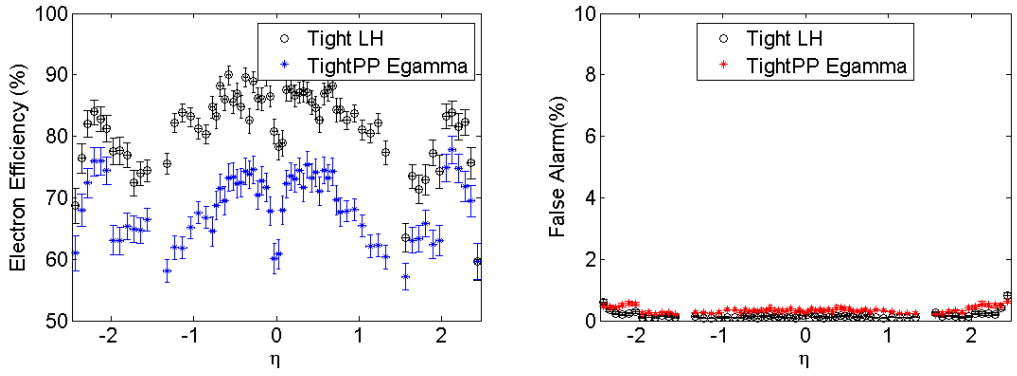

Figura 62: Gráfico de η no ponto de operação *Tight*, comparando LH e e\γ, com 5 GeV$<E_t<$20 GeV.

Nas Figuras 63 a 65, será mostrado o comportamento do algoritmo em 3 regiões de E_t definidas na Tabela 7, com $0 \leq |\eta| < 0.8$, considerando os pontos de operação *LoosePP*, *MediumPP* e *TightPP* do e\γ, respectivamente. O resultado se mantém parecido para as outras regiões de η, como pode ser visto nos gráficos disponíveis no Apêndice B.2.

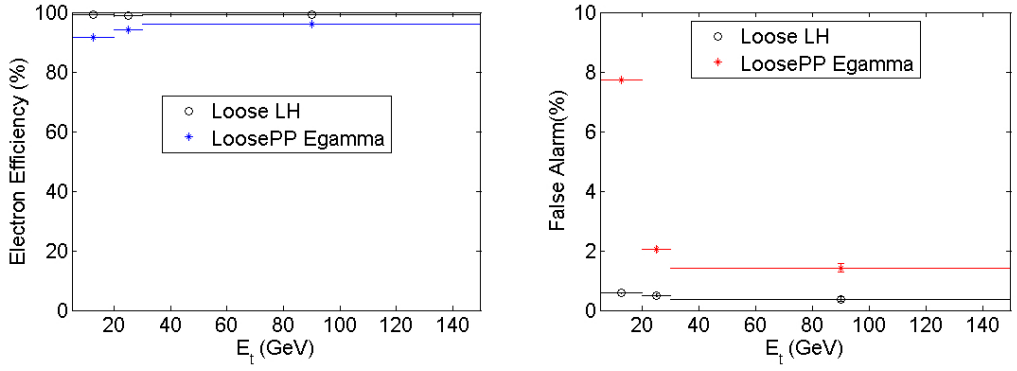

Figura 63: Gráfico de E_t no ponto de operação *Loose*, comparando LH e e\γ, com $0>|\eta|>0.8$.

Como visto na Figura 63, o método da verossimilhança mostra um aumento de eficiência de $\approx 6\%$, em relação ao *Loose PP*, com o mesmo falso alarme. Quando a eficiência é fixada, há uma redução do falso alarme de $\approx 2\%$, em média. Na Figura 64 observa-se um ganho de 4% na eficiência e diminuição de 0.15% no falso alarme em comparação com o *Medium PP*. Já em comparação ao *Tight PP*, as barras de erro mostradas na Figura 65 não permitem qualquer afirmação.

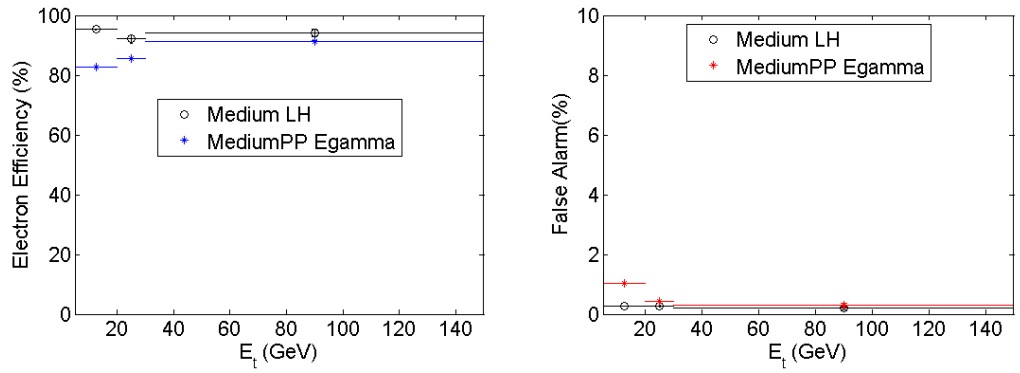

Figura 64: Gráfico de E_t no ponto de operação *Medium*, comparando LH e e\γ, com $0>|\eta|>0.8$.

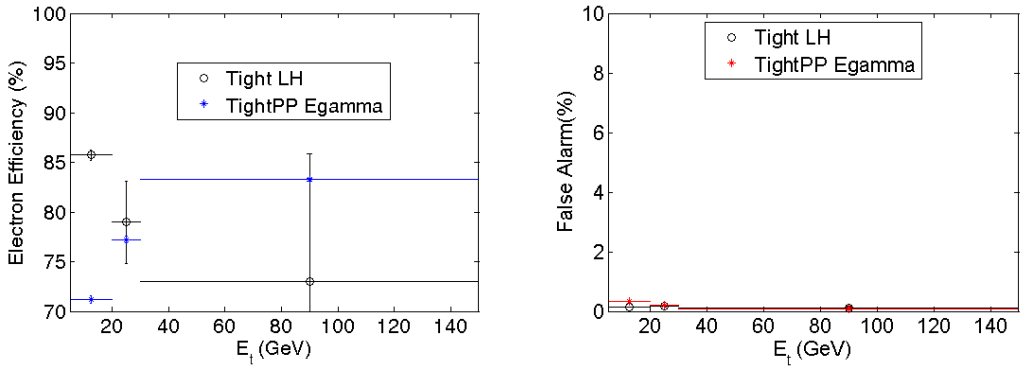

Figura 65: Gráfico de E_t no ponto de operação *Tight*, comparando LH e e\γ, com $0>|\eta|>0.8$.

Nas Figuras 66 a 68 é mostrado a Eficiência de sinal e Rejeição de ruído de fundo *vs* número de vértices primário, para a Região 1, definida na Tabela 7, para os pontos de operação *LoosePP*, *MediumPP* e *TightPP* do e\γ, respectivamente. Observa-se que existe uma dependência dos dois algoritmos em relação ao NVTX. Com o aumento do NVTX a Eficiência de Sinal diminuiu e o Falso Alarme aumenta. O resultado se mantém parecido para as outras regiões, como pode ser visto nos gráficos apresentados no Apêndice B.3.

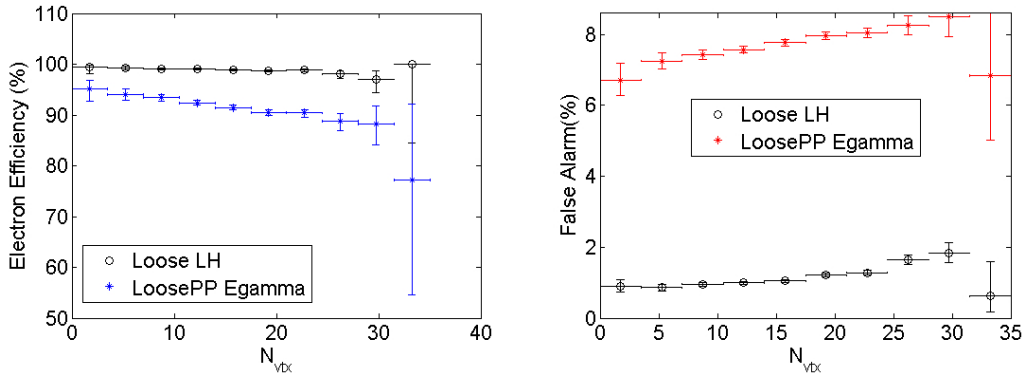

Figura 66: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 1.

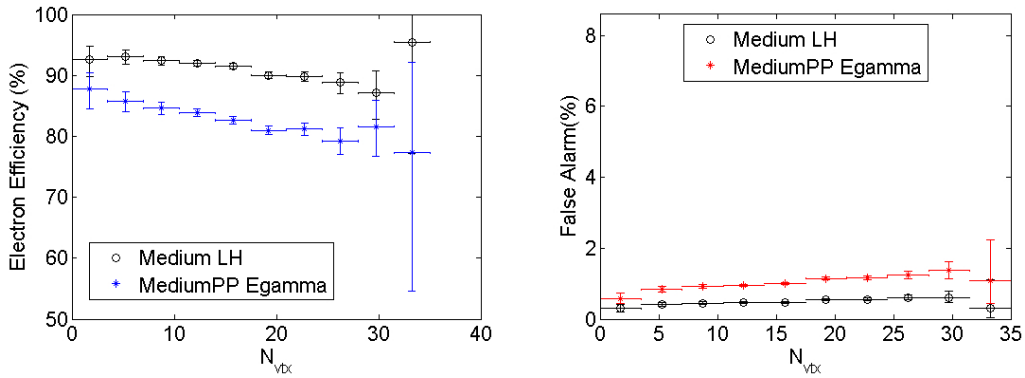

Figura 67: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 1.

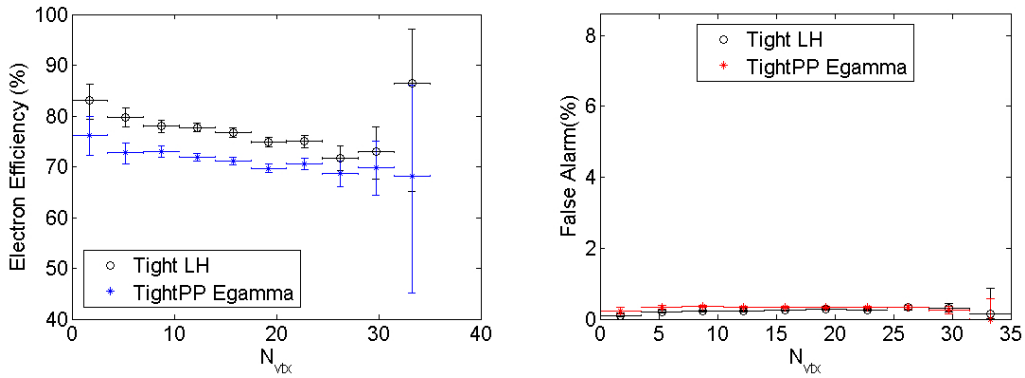

Figura 68: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 1.

6.1.4 ANÁLISE MULTIVARIADA

A análise multivariada com dados MC foi dividida em 12 regiões, mostradas na Tabela 7. A estratégia adotada nessa dissertação, para a análise multivariada, foi verificar a viabilidade da inclusão de PDFs bidimensionais no modelo de verossimilhança. Evitando assim, um possível impacto negativo na estimação das PDFs pelo MKDE, causado pela "maldição da dimensionalidade", como visto na Seção 4.2.6.

Nesta análise foram utilizadas somente as 13 variáveis discriminantes, mostradas na Tabela 6, e os *Hard Cuts* não foram considerados.

6.1.5 ESTIMAÇÃO DE PDFS MULTIVARIADAS

O ponto de partida para a estimação das PDFs multivariadas é a medida de Informação Mútua, cujo algoritmo foi desenvolvido para essa análise. Foi decidido fazer separadamente as matrizes de informação mútua dos conjuntos de sinal e ruído de fundo, visto que a PDF conjunta utilizada pelo algoritmo de verossimilhança é construída separadamente. A representação gráfica destas matrizes é mostrada na Figura 69.

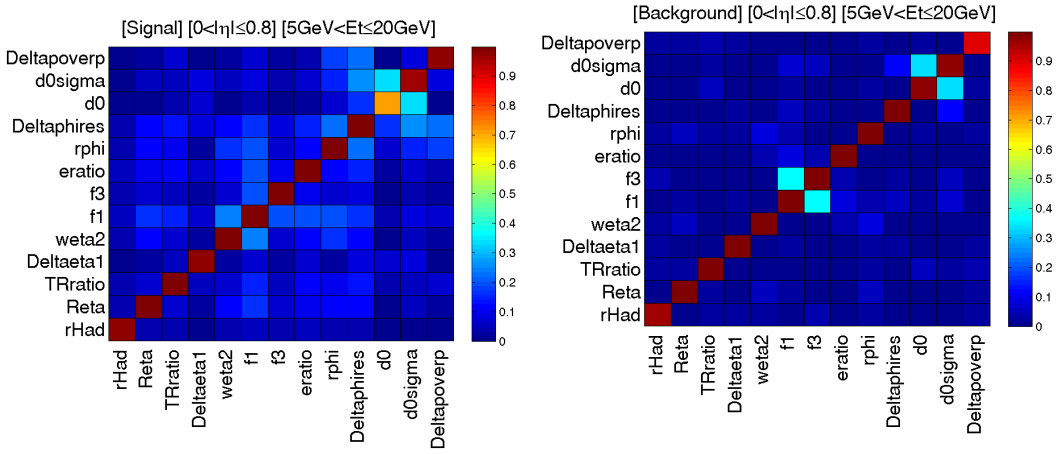

Figura 69: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 1.

Após analisar as Informações Mútuas do sinal e do ruído de fundo, percebeu-se que as variáveis d_0 e $d_0\sigma$, e as variáveis f_1 e f_3 possuem maior dependência, na região 1. Com isso, foi decidido estimar duas PDFs bidimensionais para as variáveis citadas acima. As Figuras 70 e 71 mostram as PDFs bidimensionais construídas pelo algoritmo MKDE desenvolvido nessa dissertação. A otimização do comprimento de banda h

N-Dimensional foi implementada de acordo com a Seção 4.2.3.1, o parâmetro λ foi considerado = 1.

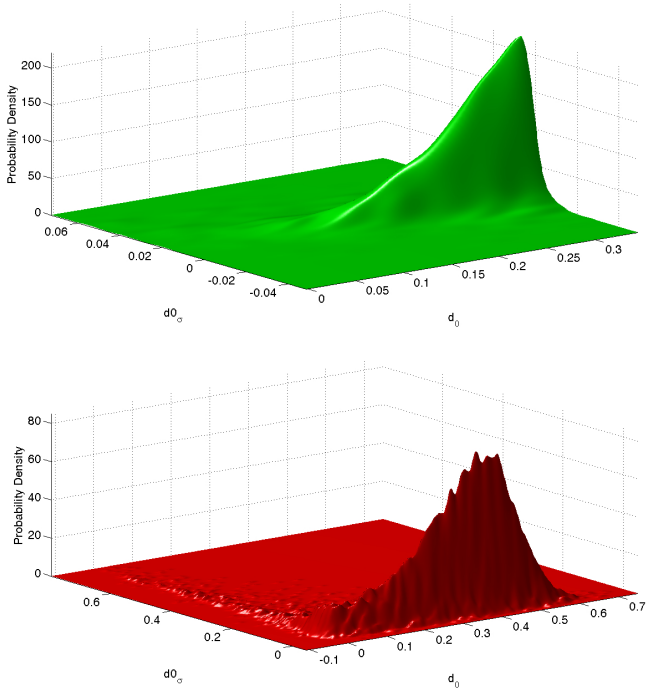

Figura 70: PDF conjunta de d_0 e $d_0\sigma$ do sinal (acima) e do ruído de fundo (abaixo).

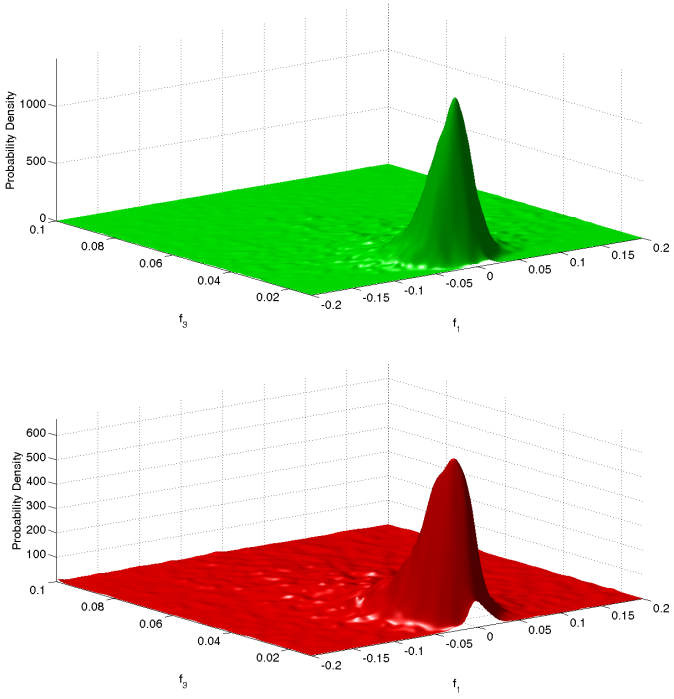

Figura 71: PDF conjunta de f_1 e f_3 do sinal (acima) e do ruído de fundo (abaixo).

6.1.6 RESULTADOS DA ANÁLISE MULTIVARIADA COM DADOS MC

Como foi percebido nos resultados da análise univariada, a verossimilhança possui um desempenho melhor do que o e\γ, por isso, iremos comparar apenas os dois métodos de verossimilhança, univariado e multivariado.

Durante o desenvolvimento deste algoritmo de MKDE analisamos sua performance com o conjunto utilizado para seu treinamento, esse conjunto foi denominado "Conjunto de Desenvolvimento", e posteriormente utilizamos o conjunto de validação, para, de fato, perceber o comportamento num contexto aplicável.

Na Figura 72 são mostradas as ROCs geradas pelos algoritmos unidimensional e multidimensional, do conjunto de desenvolvimento e validação, para a região 1, os gráficos de ROC das outras regiões se encontram no Apêndice B. O significado dos ítens na legenda são: LH(Univarida), representa a LH construída como proposto em (COLLABORATION et al., 2013); LH(Usando 1 PDF bidimensional), significa remover as PDFs unidimensionais das variáveis d_0 e $d_0\sigma$ e adicionar 1 PDF bidimensional gerada por elas; LH(Usando 2 PDFs bidimensionais), significa adicionar mais uma PDF bidimensional gerada por f_1 e f_3, juntamente com a gerada por d_0 e $d_0\sigma$.

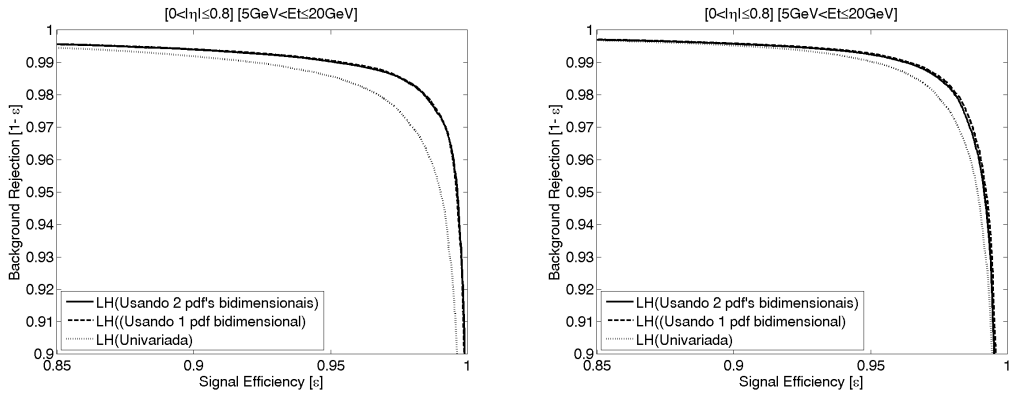

Figura 72: Gráfico comparando as ROCs da verossimilhança: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDFs Bidimensionais, para Região 1. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

As Tabelas 10 e 11 representam os resultados do SP, dos conjuntos de desenvolvimento e validação, respectivamente, gerados pelas ROCs das 12 regiões indicadas pela Tabela 7.

Tabela 10: Comparação do Índice SP das Likelihoods Univariada e Multivariada, para as 12 Regiões, utilizando os dados de desenvolvimento.

	Univariada	Multivariada
Regiões	Índice SP (%)	Índice SP (%)
1	97,51 ±0,05	98,21 ±0,04
2	97,96 ±0,07	98,50 ±0,07
3	98,78 ±0,08	99,09 ±0,07
4	97,05 ±0,06	97,90 ±0,06
5	97,67 ±0,09	98,26 ±0,09
6	98,24 ±0,12	98,75 ±0,11
7	96,45 ±0,09	97,62 ±0,07
8	96,58 ±0,13	97,68 ±0,10
9	97,44 ±0,17	98,27 ±0,14
10	96,65 ±0,25	98,01 ±0,18
11	96,33 ±0,41	97,58 ±0,35
12	97,42 ±0,53	98,26 ±0,46

Tabela 11: Comparação do Índice SP das Likelihoods Univariada e Multivariada, para as 12 Regiões, utilizando os dados de validação.

	Univariada	Multivariada
Regiões	Índice SP (%)	Índice SP (%)
1	97,70 ±0,05	98,06 ±0,04
2	98,51 ±0,04	98,82 ±0,04
3	98,96 ±0,05	99,22 ±0,04
4	97,22 ±0,06	97,70 ±0,06
5	98,04 ±0,06	98,60 ±0,05
6	98,69 ±0,06	99,08 ±0,05
7	96,18 ±0,09	97,56 ±0,07
8	96,78 ±0,09	98,10 ±0,07
9	97,94 ±0,09	98,75 ±0,07
10	95,70 ±0,11	97,68 ±0,08
11	96,82 ±0,10	98,10 ±0,07
12	97,80 ±0,10	98,57 ±0,08

Como visto nas Tabelas 10 e 11 a inclusão das PDFs bidimendionais, construídas pelo MKDE, na verossimilhança, aumentam o desempenho do índice SP em relação a verossimilhança desenvolvida apenas com PDFs unidimensionais. Esse resultado pode ser percebido em todas as 12 Regiões, sendo que para a análise com conjunto de validação, vista na Tabela 11, a maior melhora se dá na região 10, com aproximadamente 2% de ganho no índice SP, já a região 3 apresenta o menor ganho, aproximadamente 0.26%. Baseado nesses resultados, conclui-se que a inclusão de PDFs bidimensionais, construídas pelo MKDE, no cálculo da probabilidade conjunta, reduz o impacto negativo da consideração de independência entre as variáveis discriminantes utilizadas no método da verossimilhança para identificação de elétrons isolados, no contexto descrito nesta dissertação.

6.2 DADOS REAIS

Os dados reais utilizados neste estudo são mostrados na Figura 73. Como é possível notar, a estatística disponível para elétrons isolados, proporcionalmente, é bem menor nos dados reais do que nos dados MC. Com isso, o problema estatístico se torna mais acentuado, devido aos eventos de interesse serem uma pequena minoria imersa em uma grande quantidade de ruído de fundo. Desta forma, essa análise foi dividida em apenas 1 região, para todo η e E_t disponíveis.

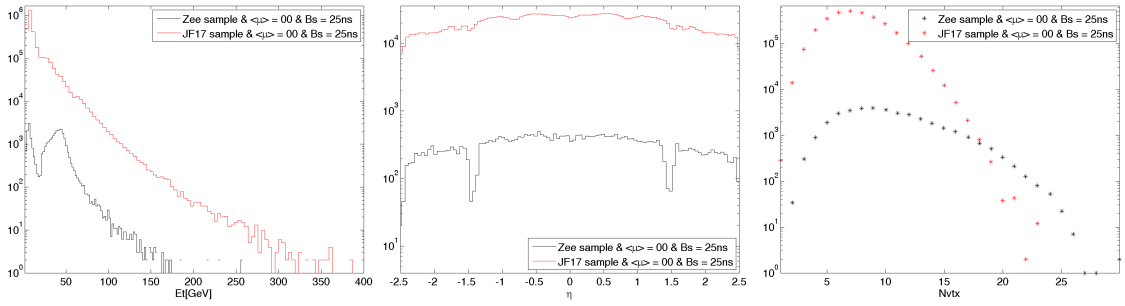

Figura 73: Perfil dos eventos de dados reais. (Esquerda) Gráfico de eventos por E_t, (Centro) Gráfico de eventos por η e (Direita) Gráfico de eventos por NVTX.

6.2.1 ESTIMAÇÃO DE PDFS UNIVARIADAS

Para estimar as PDFs das variáveis dos dados reais foi preciso recorrer ao algoritmo de *Tag and Probe* para separar os elétrons isolados do ruído de fundo. As Figuras 74 até 80 mostram a comparação das PDFs dos dados reais com o MC, dos elétrons isolados, obtidos pelo algoritmo de *TaP* desenvolvido para essa dissertação. As PDFs

não são exatamente iguais devido aos chuveiros não serem bem estimados nos dados MC simulados.

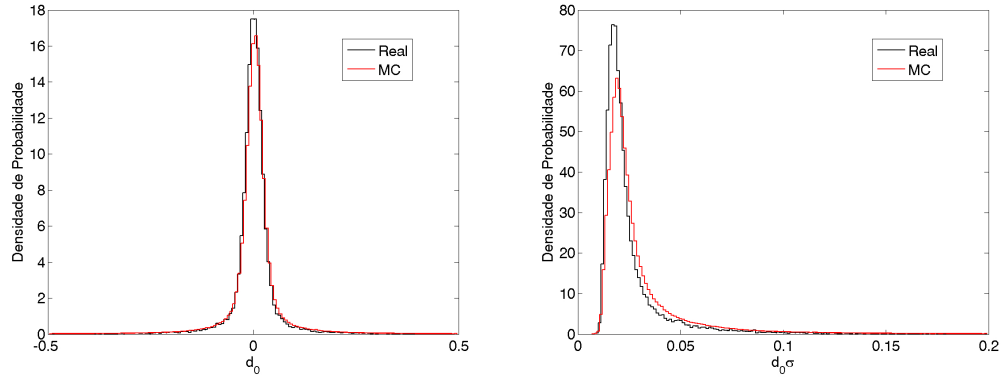

Figura 74: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável d_0 e (Direita) Variável $d_{0\sigma}$.

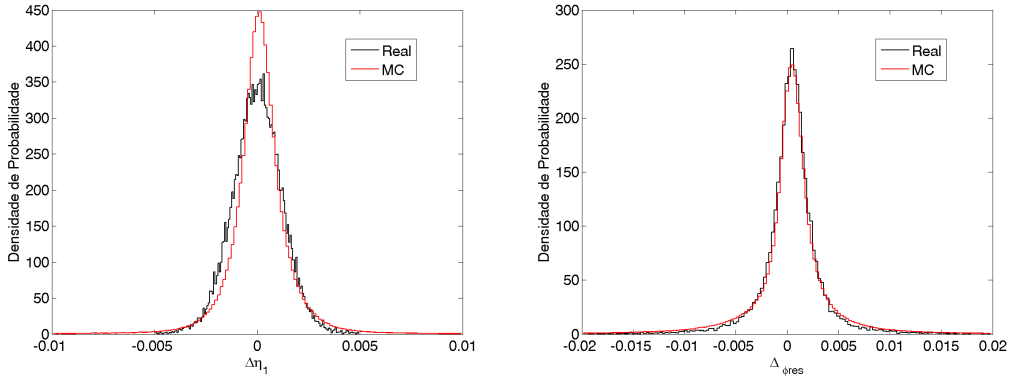

Figura 75: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável $\Delta_{\eta 1}$ e (Direita) Variável $\Delta_{\phi res}$.

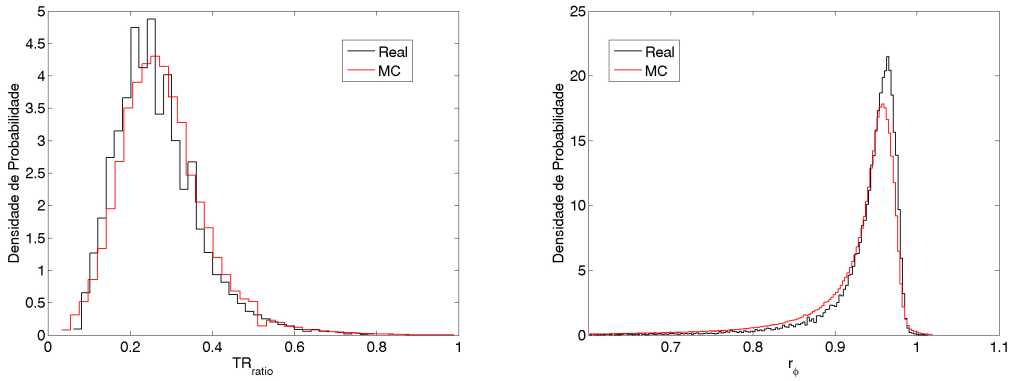

Figura 79: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável TR_{ratio} e (Direita) Variável r_ϕ.

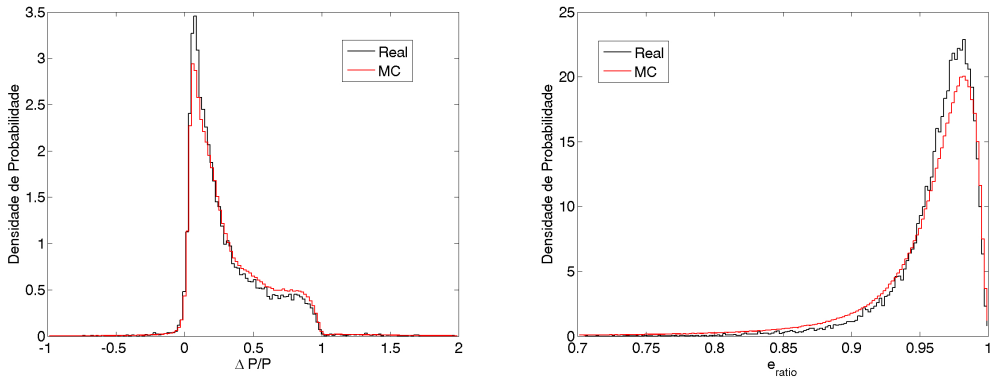

Figura 76: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável $\Delta P/P$ e (Direita) Variável E_{ratio}.

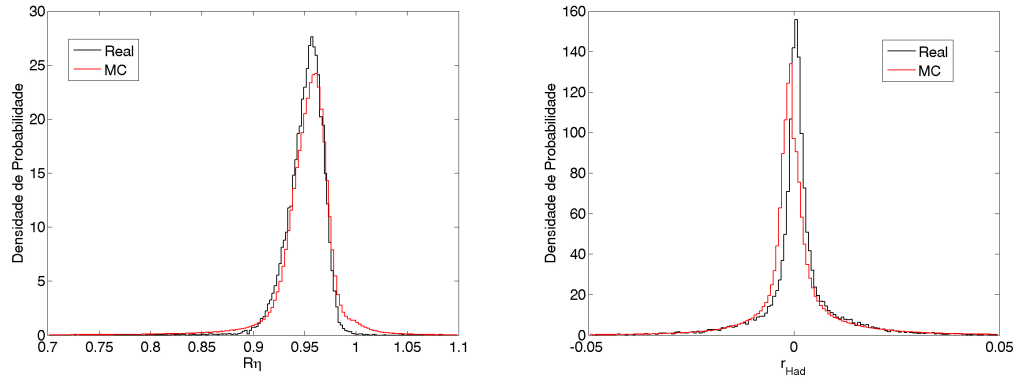

Figura 77: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável R_{Had} e (Direita) Variável r_η.

6.2.2 RESULTADOS DA ANÁLISE UNIVARIADA E MULTIVARIADA COM DADOS REAIS

As ROCs geradas pelos algoritmos univariado e multivariado, aplicados aos dados reais, serão mostradas na Figura 81, cabe ressaltar que as *Likelihoods* dessa comparação foram feitas utilizando o *Tight* Menu, descrito na Tabela 6.

Analogamente à análise multidimensional feita com dados MC, decidiu-se utilizar as mesmas variáveis (d_0, $d_{0\sigma}$, f_1 e f_3) para gerar as PDFs bidimensionais nos dados reais.

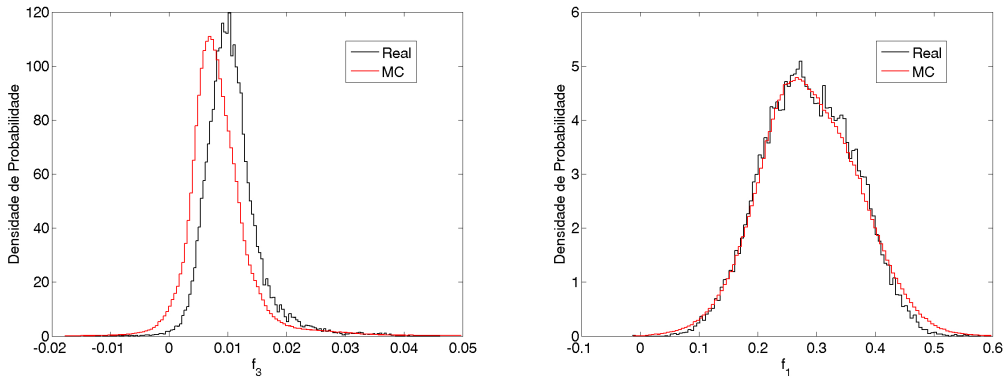

Figura 78: Comparação entre as PDFs de Dados Reais e MC. (Esquerda) Variável f_1 e (Direita) Variável f_3.

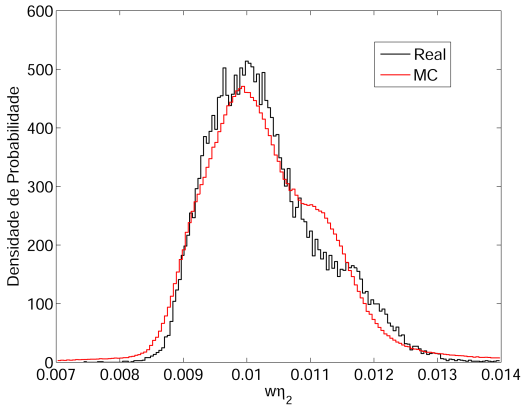

Figura 80: Comparação entre as PDFs de Dados Reais e MC. Variável $W_{\eta 2}$.

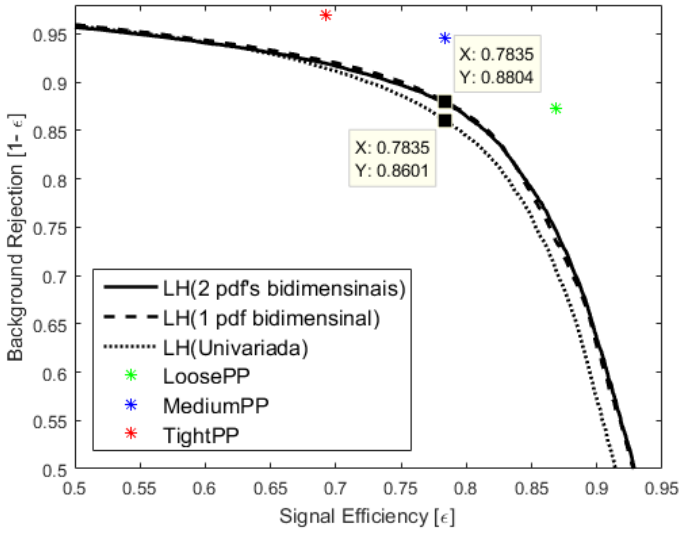

Figura 81: Comparação entre as ROCs da análise univariada e multivariada dos Dados Reais.

Observa-se na Figura 81, que o método multivariado apresenta melhores resultados em relação ao univariado, entretanto o algoritmo e\$\gamma$ tem a vantagem de ter sido construído com maior segmentação em η e E_t do que a verossimilhança, mostrando, como esperado, uma eficiência maior neste caso. Essa melhora é evidenciada na tabela 12, que mostra a relação entre a eficiência dos métodos univariado, multivariado, com o falso alarme do ponto de operação do e\$\gamma$ *tightPP*, para dados reais.

Tabela 12: Eficiência de Sinal para todos os métodos estudados, fixando a Eficiência do ponto de operação *MediumPP*, para dados reais.

	Eficiência de Sinal (%)			Rejeição de Ruído(%)		
Likelihood Univariada	78,35	+0,81	-0,83	86,01	+0,02	-0,02
Likelihood Multivariada	78,35	+0,81	-0,83	88,04	+0,02	-0,02
Tight PP	78,35	+0,81	-0,83	94,58	+0,02	-0,02

Como mostrado na Tabela 12 a proposta Multivariada apresenta um ganho de eficiência de sinal próximo de 4% em relação ao *Tight PP*, para a mesma rejeição de ruído de fundo. Embora não tivéssemos estatística suficiente para realizar a análise dividida em regiões de η e E_t, esses resultados apresentam fortes indícios de que o método multivariado proposto é uma alternativa real para diminuir o efeito de se utilizar a presunção de independência entre as variáveis discriminantes vistas.

7 CONCLUSÕES

Esta dissertação apresentou o estudo e implementação de um algoritmo de estimação de densidade não-paramétrico baseado no MKDE, aplicado à seleção de elétrons para o Experimento ATLAS, usando a técnica de *Likelihood*. Os resultados mostraram que a versão univariada desta técnica apresenta desempenho superior ao método utilizado pela Colaboração conhecido como e\γ, quando segmentados com a mesma resolução de η e E_t, e que a versão multivariada pode melhorar ainda mais o seu desempenho; mostrando que o método proposto pode ser de fato uma alternativa para o aprimoramento do sistema de seleção de eventos do ATLAS. A *Likelihood* multivariada foi implementada a partir de uma proposta, aqui desenvolvida, de se utilizar PDFs bidimensionais selecionadas a partir da medida de Informação Mútua, no intuito de minimizar o impacto negativo na estimação das PDFs multidimensionais provocadas por questões estatísticas que aparecem com o aumento dimensional das mesmas, e de reduzir ao mesmo tempo o efeito da dependência entre as variáveis na reconstrução das PDFs conjuntas, de sinal e ruído, em relação a *Likelihood* univariada.

O maior esforço deste trabalho se deu na implementação dos algoritmos de estimação de densidade univariada e multivariada, na sua aplicação em seleção de eventos pelo método da *Likelihood*, e na análise de desempenho a partir dos dados simulados e reais do Experimento ATLAS. Alguns procedimentos de otimização foram propostos e aplicados na construção das PDFs das variáveis discriminantes fornecidas pelo detector, servindo como alicerce no ganho de performance do método proposto. Além disso, vários algoritmos adicionais foram desenvolvidos, como Informação Mútua, *Tag and Probe* e *Likelihood*.

A partir dos estudos realizados nessa dissertação, foi possível ampliar o conhecimento do tema abordado e identificar algumas dificuldades relacionadas a sua realização experimental, possibilitando vislumbrar alguns possíveis caminhos de melhora na análise e implementação dos métodos expostos aqui.

7.1 TRABALHOS FUTUROS

O estudo teórico e a implementação dos métodos evidenciou possibilidades de melhoras, nas análises uni e multivariada, em diversas diretrizes. Dentre elas se destacam:

- Segmentar a análise da *Likelihood* em η e E_t como proposto pela colaboração;
- Aumentar a estatística para construção do *Kernel* em todas as regiões de η e E_t;
- Estudar estratégias para aumentar a estatística de elétrons isolados com dados reais;
- Incluir medidas de performance mais detalhadas no método do *Tag and Probe*;
- Estudar o impacto da interpolação N-Dimensional na obtenção das probabilidades nas PDF's verificando o número de pontos a serem estimados pelo KDE N-dimensional;
- Estudo de uma otimização robusta do KDE N-Dimensional para todas as variáveis;
- Avaliar os erros associados à medida de Informação Mútua;
- Comparação entre a Likelihood Univariada proposta nesta dissertação e a construída pela colaboração, tendo em vista que os resultados apresentados foram gerados através das PFD's construídas pelo algoritmo KDE N-dimensional, exposto nesse trabalho, e a colaboração utiliza o *TMVA Tool*.
- Aplicar os algoritmos mostrados em dados com *pileup* (ou empilhamento) maiores do que os apresentados neste trabalho.

REFERÊNCIAS

AAD, G. et al. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, IOP Publishing, v. 3, n. 08, p. S08003, 2008.

AAD, G. et al. Readiness of the atlas tile calorimeter for lhc collisions. *The European Physical Journal C*, Springer, v. 70, n. 4, p. 1193–1236, 2010.

AAD, G. et al. Electron performance measurements with the atlas detector using the 2010 lhc proton-proton collision data. *The European Physical Journal C*, Springer, v. 72, n. 3, p. 1–46, 2012.

AAD, G. et al. Performance of the atlas trigger system in 2010. *The European Physical Journal C*, Springer, v. 72, n. 1, p. 1–61, 2012.

AAD, G. et al. Atlas pixel detector electronics and sensors. *Journal of Instrumentation*, IOP Publishing, v. 3, n. 07, p. P07007, 2008.

ABAT, E. et al. The atlas transition radiation tracker (trt) proportional drift tube: design and performance. *Journal of Instrumentation*, IOP Publishing, v. 3, n. 02, p. P02013, 2008.

ABRAMSON, I. S. On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*, JSTOR, p. 1217–1223, 1982.

ACHENBACH, R. et al. *The ATLAS level-1 calorimeter trigger, 2008*. [S.l.]: JINST.

ALISON, J. *The road to discovery: Detector alignment, electron identification, particle misidentification, ww physics, and the discovery of the Higgs Boson*. [S.l.]: Springer, 2014.

ANJOS, A. dos. *Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos*. Tese (Doutorado) — Tese (Doutorado), COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

BARBOSA, M. F. *Estudos em Estimação de Densidade por Kernel: Métodos de Seleção de Características e Estimação do*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE MINAS GERAIS, 2013.

BREIMAN, L.; MEISEL, W.; PURCELL, E. Variable kernel estimates of multivariate densities. *Technometrics*, Taylor & Francis Group, v. 19, n. 2, p. 135–144, 1977.

CALORIMETER, A. E. L. A. E. et al. Construction, assembly and tests of the atlas electromagnetic end-cap calorimeters. *Journal of Instrumentation*, IOP Publishing, v. 3, n. 06, p. P06002, 2008.

CERN. *About CERN*. 2015. Disponível em: <http://home.web.cern.ch/about>.

CERN. *Knowledge Transfer*. 2015. Disponível em: $<http : //knowledgetransfer.web.cern.ch/>$.

CERN. *The Large Hadron Collider*. 2015. Disponível em: <http://home.web.cern.ch/topics/large-hadron-collider>.

COLLABORATION, A. Operation and performance of the atlas semiconductor tracker. 2014.

COLLABORATION, A. et al. Commissioning of the atlas muon spectrometer with cosmic rays. *arXiv preprint arXiv:1006.4384*, 2010.

COLLABORATION, A. et al. Expected electron performance in the atlas experiment. *ATLAS note: ATL-PHYS-PUB-2011-006*, 2011.

COLLABORATION, A. et al. Improved electron reconstruction in atlas using the gaussian sum filter-based model for bremsstrahlung. In: ATLAS-CONF-2012-047. [S.l.], 2012.

COLLABORATION, A. et al. *Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data*. [S.l.], 2013.

COLLABORATION, A.; RYAN, P. et al. The atlas inner detector commissioning and calibration, accepted by. *Eur. Phys. J.*

COLLABORATION-NTUPLE, A. *Ntuple analysis framework*. Disponível em: $<http : //atlas - saclay.in2p3.fr/doc_ntuple.pdf>$.

COMANICIU, D.; RAMESH, V.; MEER, P. The variable bandwidth mean shift and data-driven scale selection. In: IEEE. *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.], 2001. v. 1, p. 438–445.

COVER, T. M.; THOMAS, J. A. *Elements of information theory*. [S.l.]: John Wiley & Sons, 2012.

EVANS, L.; BRYANT, P. Lhc machine. *Journal of Instrumentation*, IOP Publishing, v. 3, n. 08, p. S08001, 2008.

FRANCAVILLA, P.; COLLABORATION, A. et al. The atlas tile hadronic calorimeter performance at the lhc. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2012. v. 404, n. 1, p. 012007.

GINÉ, E.; SANG, H. Uniform asymptotics for kernel density estimators with variable bandwidths. *Journal of Nonparametric Statistics*, Taylor & Francis, v. 22, n. 6, p. 773–795, 2010.

HALL, P. On global properties of variable bandwidth density estimators. *The Annals of Statistics*, JSTOR, p. 762–778, 1992.

HANSEN, B. E. Lecture notes on nonparametrics. *Lecture notes*, 2009.

METZ, C. E. Basic principles of roc analysis. In: ELSEVIER. *Seminars in nuclear medicine*. [S.l.], 1978. v. 8, n. 4, p. 283–298.

NARSKY, I.; PORTER, F. C. *Statistical analysis techniques in particle physics: Fits, density estimation and supervised learning*. [S.l.]: John Wiley & Sons, 2013.

PEETERS, S. J. M. *The ATLAS semiconductor tracker endcap*. [S.l.: s.n.], 2003.

PERKINS, D. H. *Introduction to high energy physics*. [S.l.]: Cambridge University Press, 2000.

SHIMAZAKI, H.; SHINOMOTO, S. A method for selecting the bin size of a time histogram. *Neural computation*, MIT Press, v. 19, n. 6, p. 1503–1527, 2007.

SILVERMAN, B. W. *Density estimation for statistics and data analysis*. [S.l.]: CRC press, 1986.

SJÖSTRAND, T. *Monte carlo event generation for lhc*. [S.l.], 1991.

SUNDARESAN, M. K. *Handbook of particle physics*. [S.l.]: CRC Press, 2001.

TARON, M.; PARAGIOS, N.; JOLLY, M.-P. Modelling shapes with uncertainties: Higher order polynomials, variable bandwidth kernels and non parametric density estimation. In: IEEE. *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. [S.l.], 2005. v. 2, p. 1659–1666.

TORRES, R. Sistema online de filtragem em um ambiente com alta taxa de eventos e fina granularidade. *Rio de Janeiro, UFRJ/COPPE*, 2010.

TURLACH, B. A. et al. *Bandwidth selection in kernel density estimation: A review*. [S.l.]: Université catholique de Louvain, 1993.

WAND, M.; JONES, M. *Kernel Smoothing, Vol. 60 of Monographs on statistics and applied probability*. [S.l.]: Chapman and Hall, London, 1995.

WATTS, G. Review of triggering. In: *Nuclear Science Symposium Conference Record, 2003 IEEE*. [S.l.: s.n.], 2003. v. 1, p. 282–287 Vol.1. ISSN 1082-3654.

WIGMANS, R. *Calorimetry: Energy measurement in particle physics*. [S.l.]: Oxford University Press, 2000.

WU, T.-J.; CHEN, C.-F.; CHEN, H.-Y. A variable bandwidth selector in multivariate kernel density estimation. *Statistics & probability letters*, Elsevier, v. 77, n. 4, p. 462–467, 2007.

APÊNDICE A – TABELAS

Os resultados apresentados a seguir são baseados na análise univariada. As tabelas tem como objetivo mostrar a Eficiência de Sinal e Rejeição de Ruído de Fundo, da *Likelihood* e do e\γ, nos pontos de operação do *Cut-Based* para as 12 Regiões da Tabela 7.

Tabela 13: Rejeição de Ruído de Fundo para o e\γ, fixando a Eficiência de Sinal do ponto de operação *Loose*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	92,26	+0,05	-0,05	97,951	+0,11	-0,10	98,59	+0,15	-0,14
$0.8 \leq \|\eta\| < 1.37$	89,36	+0,06	-0,06	96,873	+0,16	-0,15	97,99	+0,21	-0,20
$1.52 \leq \|\eta\| < 2.01$	92,23	+0,07	-0,07	97,331	+0,16	-0,15	97,81	+0,25	-0,23
$2.01 \leq \|\eta\| < 2.47$	92,46	+0,07	-0,07	95,007	+0,22	-0,22	96,28	+0,34	-0,32

Tabela 14: Rejeição de Ruído de Fundo para a *Likelihood*, fixando a Eficiência de Sinal do ponto de operação *Loose*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	99,42	+0,01	-0,01	99,50	+0,05	-0,06	99,62	+0,07	-0,08
$0.8 \leq \|\eta\| < 1.37$	99,23	+0,02	-0,02	99,27	+0,07	-0,08	99,48	+0,10	-0,12
$1.52 \leq \|\eta\| < 2.01$	98,83	+0,03	-0,03	98,84	+0,10	-0,11	98,94	+0,16	-0,18
$2.01 \leq \|\eta\| < 2.47$	97,67	+0,04	-0,04	98,25	+0,13	-0,14	98,64	+0,19	-0,21

Tabela 15: Rejeição de Ruído de Fundo para o e\γ, fixando a Eficiência de Sinal do ponto de operação *Medium*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	98,97	+0,02	-0,02	99,58	+0,05	-0,05	99,71	+0,07	-0,06
$0.8 \leq \|\eta\| < 1.37$	99,07	+0,02	-0,02	99,57	+0,06	-0,06	99,65	+0,10	-0,08
$1.52 \leq \|\eta\| < 2.01$	99,21	+0,02	-0,02	99,55	+0,07	-0,06	99,67	+0,11	-0,09
$2.01 \leq \|\eta\| < 2.47$	98,36	+0,04	-0,03	98,85	+0,11	-0,10	99,16	+0,17	-0,15

Tabela 16: Rejeição de Ruído de Fundo para a *Likelihood*, fixando a Eficiência de Sinal do ponto de operação *Medium*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	99,73	+0,01	-0,01	99,73	+0,04	-0,04	99,80	+0,05	-0,06
$0.8 \leq \|\eta\| < 1.37$	99,75	+0,01	-0,01	99,71	+0,04	-0,05	99,76	+0,07	-0,08
$1.52 \leq \|\eta\| < 2.01$	99,59	+0,02	-0,02	99,55	+0,06	-0,07	99,66	+0,09	-0,11
$2.01 \leq \|\eta\| < 2.47$	98,78	+0,03	-0,03	99,15	+0,09	-0,10	99,22	+0,14	-0,17

Tabela 17: Rejeição de Ruído de Fundo para o e\γ, fixando a Eficiência de Sinal do ponto de operação *Tight*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	99,66	+0,01	-0,01	99,81	+0,04	-0,03	99,91	+0,04	-0,03
$0.8 \leq \|\eta\| < 1.37$	99,72	+0,01	-0,01	99,79	+0,04	-0,04	99,89	+0,06	-0,04
$1.52 \leq \|\eta\| < 2.01$	99,74	+0,01	-0,01	99,82	+0,05	-0,04	99,91	+0,06	-0,04
$2.01 \leq \|\eta\| < 2.47$	99,50	+0,02	-0,02	99,74	+0,06	-0,05	99,80	+0,09	-0,07

Tabela 18: Rejeição de Ruído de Fundo para a *Likelihood*, fixando a Eficiência de Sinal do ponto de operação *Tight*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	99,86	+0,01	-0,01	99,82	+0,03	-0,03	99,88	+0,04	-0,05
$0.8 \leq \|\eta\| < 1.37$	99,90	+0,01	-0,01	99,83	+0,03	-0,04	99,84	+0,05	-0,07
$1.52 \leq \|\eta\| < 2.01$	99,82	+0,01	-0,01	99,82	+0,04	-0,05	99,89	+0,05	-0,07
$2.01 \leq \|\eta\| < 2.47$	99,55	+0,02	-0,02	99,64	+0,06	-0,07	99,74	+0,08	-0,10

Tabela 19: Eficiência de Sinal para o e\γ fixando a Rejeição de Ruído de Fundo do ponto de operação *Loose*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	91,62	+0,24	-0,25	94,07	+0,17	-0,18	95,96	+0,08	-0,08
$0.8 \leq \|\eta\| < 1.37$	91,27	+0,30	-0,31	94,28	+0,20	-0,21	96,44	+0,09	-0,09
$1.52 \leq \|\eta\| < 2.01$	90,06	+0,38	-0,39	91,66	+0,28	-0,29	95,46	+0,12	-0,12
$2.01 \leq \|\eta\| < 2.47$	91,34	+0,43	-0,44	93,45	+0,29	-0,30	95,13	+0,13	-0,13

Tabela 20: Eficiência de Sinal para a *Likelihood* fixando a Rejeição de Ruído de Fundo do ponto de operação *Loose*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	99,29	+0,07	-0,08	98,92	+0,07	-0,08	99,22	+0,03	-0,04
$0.8 \leq \|\eta\| < 1.37$	99,34	+0,08	-0,09	98,71	+0,10	-0,10	99,10	+0,04	-0,05
$1.52 \leq \|\eta\| < 2.01$	98,31	+0,16	-0,17	96,16	+0,19	-0,20	97,89	+0,08	-0,08
$2.01 \leq \|\eta\| < 2.47$	97,69	+0,22	-0,25	98,27	+0,15	-0,16	98,81	+0,07	-0,07

Tabela 21: Eficiência de Sinal para o e\γ fixando a Rejeição de Ruído de Fundo do ponto de operação *Medium*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	82,64	+0,33	-0,34	85,55	+0,26	-0,26	91,39	+0,11	-0,11
$0.8 \leq \|\eta\| < 1.37$	79,47	+0,43	-0,44	84,17	+0,32	-0,33	91,00	+0,14	-0,14
$1.52 \leq \|\eta\| < 2.01$	78,57	+0,53	-0,54	82,52	+0,39	-0,39	89,15	+0,17	-0,18
$2.01 \leq \|\eta\| < 2.47$	84,93	+0,55	-0,56	86,82	+0,40	-0,40	90,62	+0,18	-0,18

Tabela 22: Eficiência de Sinal para a *Likelihood* fixando a Rejeição de Ruído de Fundo do ponto de operação *Medium*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	95,25	+0,19	-0,19	92,10	+0,20	-0,20	94,18	+0,09	-0,09
$0.8 \leq \|\eta\| < 1.37$	92,55	+0,28	-0,29	88,86	+0,28	-0,28	94,00	+0,11	-0,12
$1.52 \leq \|\eta\| < 2.01$	86,48	+0,44	-0,45	82,68	+0,39	-0,39	87,69	+0,18	-0,19
$2.01 \leq \|\eta\| < 2.47$	88,11	+0,49	-0,51	89,78	+0,35	-0,36	91,41	+0,17	-0,17

Tabela 23: Eficiência de Sinal para o e\γ fixando a Rejeição de Ruído de Fundo do ponto de operação *Tight*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	71,17	+0,40	-0,40	77,24	+0,31	-0,31	83,29	+0,15	-0,15
$0.8 \leq \|\eta\| < 1.37$	64,07	+0,52	-0,52	72,46	+0,39	-0,40	80,16	+0,19	-0,19
$1.52 \leq \|\eta\| < 2.01$	63,57	+0,62	-0,63	68,56	+0,48	-0,48	76,87	+0,24	-0,24
$2.01 \leq \|\eta\| < 2.47$	71,81	+0,69	-0,70	75,16	+0,51	-0,51	78,94	+0,25	-0,25

Tabela 24: Eficiência de Sinal para a *Likelihood* fixando a Rejeição de Ruído de Fundo do ponto de operação *Tight*.

	$5 \leq E_t < 20GeV$			$20 \leq E_t < 30GeV$			$E_t > 30GeV$		
$0 \leq \|\eta\| < 0.8$	85,74	+0,31	-0,31	78,96	+0,30	-0,30	73,01	+0,18	-0,18
$0.8 \leq \|\eta\| < 1.37$	81,12	+0,42	-0,43	77,00	+0,37	-0,37	75,24	+0,21	-0,21
$1.52 \leq \|\eta\| < 2.01$	72,47	+0,58	-0,58	69,60	+0,47	-0,48	70,01	+0,26	-0,26
$2.01 \leq \|\eta\| < 2.47$	73,19	+0,68	-0,69	69,37	+0,54	-0,55	72,21	+0,28	-0,28

APÊNDICE B – FIGURAS

B.1 GRÁFICOS DE η

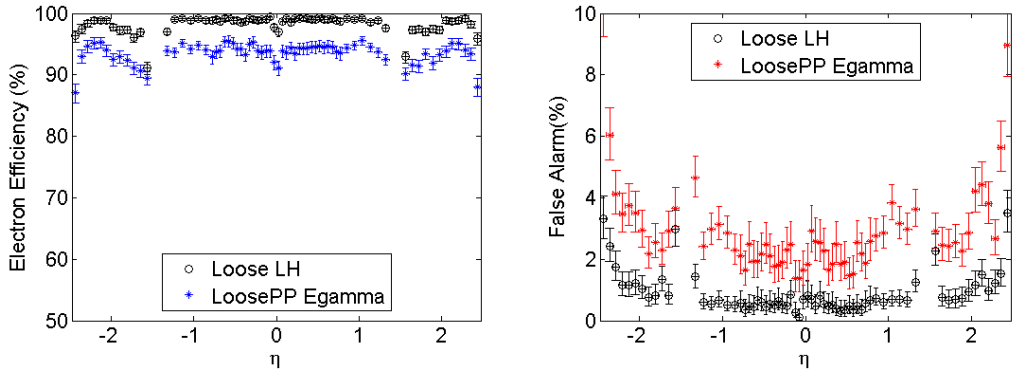

Figura 82: Gráfico de η no ponto de operação *Loose*, comparando LH e e\γ, com 20GeV$<E_t<$30GeV.

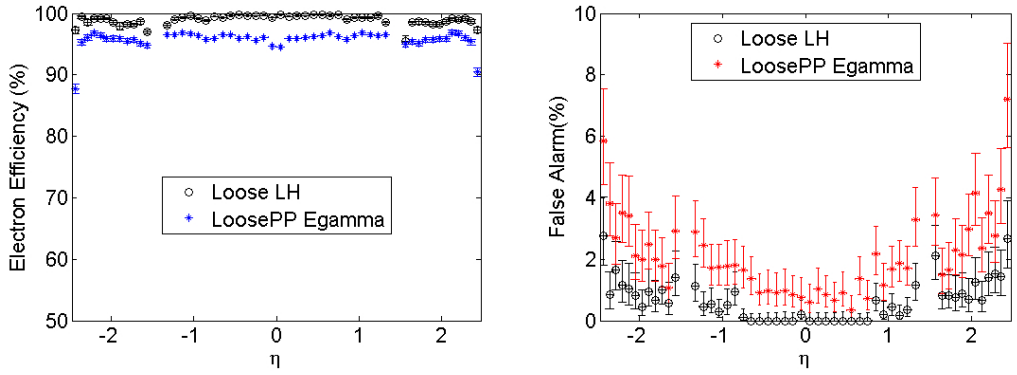

Figura 83: Gráfico de η no ponto de operação *Loose*, comparando LH e e\γ, com $E_t>$30GeV.

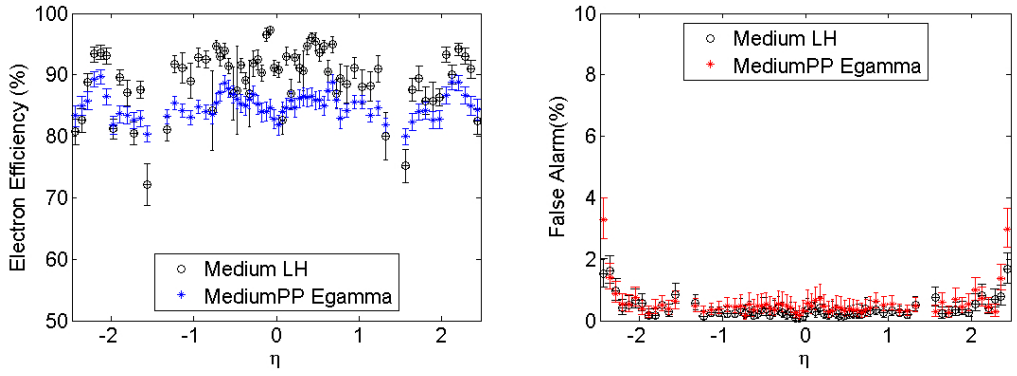

Figura 84: Gráfico de η no ponto de operação *Medium*, comparando LH e e\γ, com 20GeV$<E_t<$30GeV.

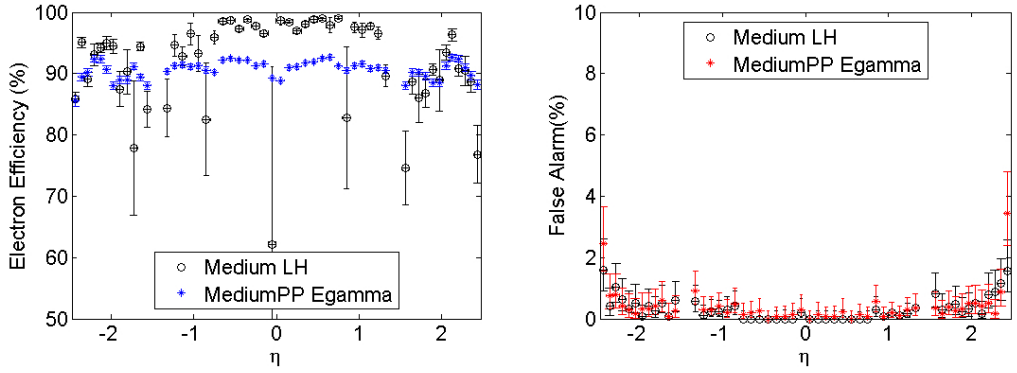

Figura 85: Gráfico de η no ponto de operação *Medium*, comparando LH e e\γ, com $E_t>$30GeV.

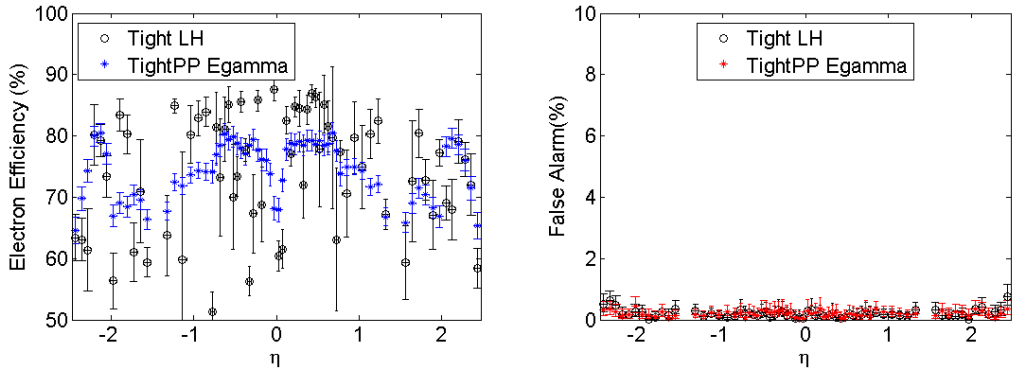

Figura 86: Gráfico de η no ponto de operação *Tight*, comparando LH e e\γ, com 20GeV$<E_t<$30GeV.

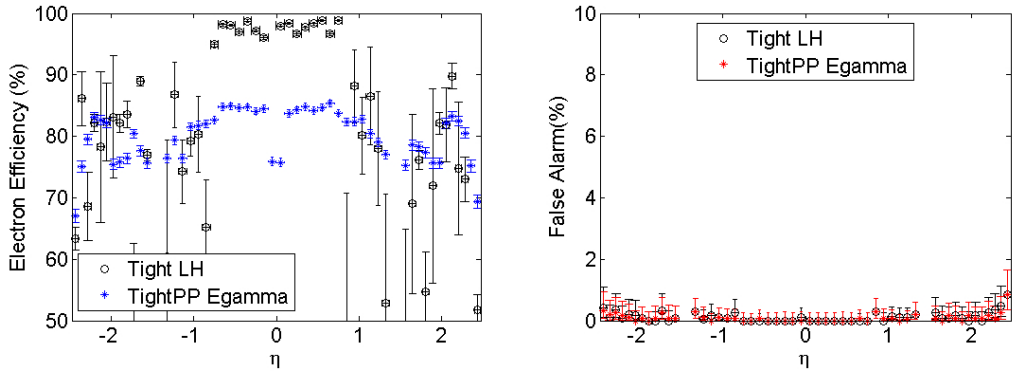

Figura 87: Gráfico de η no ponto de operação *Tight*, comparando LH e e\γ, com E_t>30GeV.

B.2 GRÁFICOS DE E_T

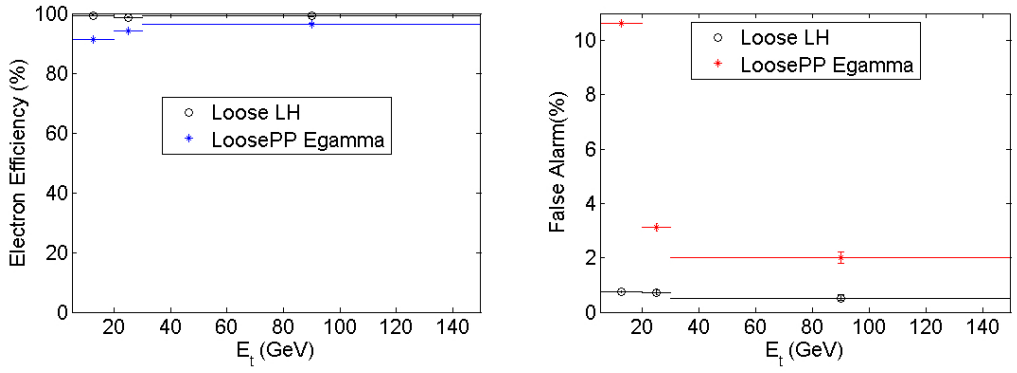

Figura 88: Gráfico de E_t no ponto de operação *Loose*, comparando LH e e\γ, com 0.8>$|\eta|$>1.37.

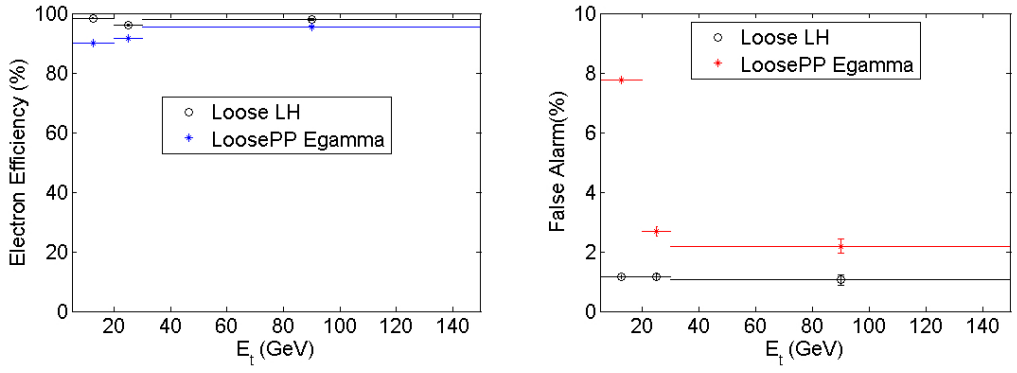

Figura 89: Gráfico de E_t no ponto de operação *Loose*, comparando LH e e\γ, com 1.52>$|\eta|$>2.01.

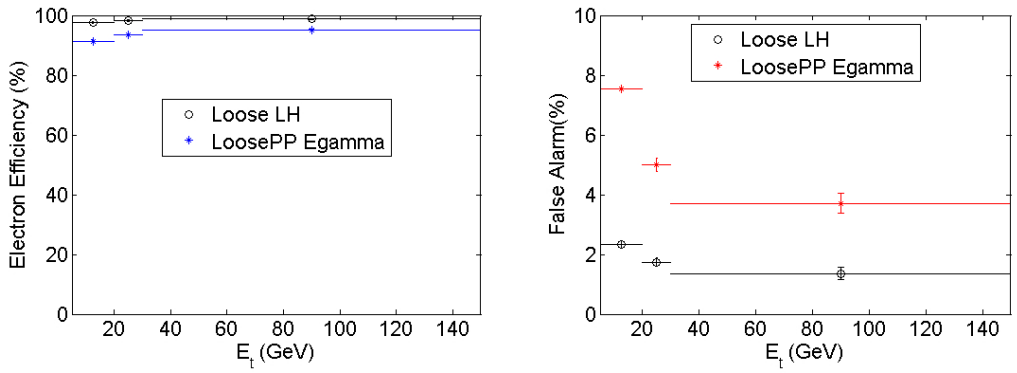

Figura 90: Gráfico de E_t no ponto de operação *Loose*, comparando LH e e\γ, com 2.01>$|\eta|$>2.47.

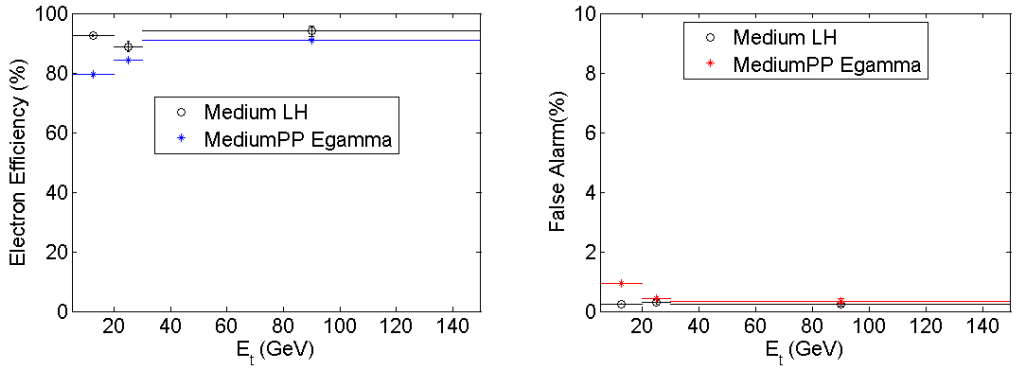

Figura 91: Gráfico de E_t no ponto de operação *Medium*, comparando LH e e\γ, com 0.8>$|\eta|$>1.37.

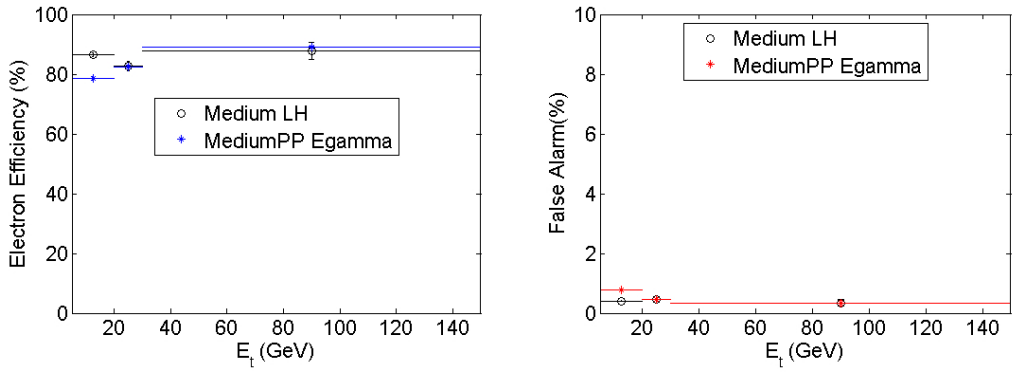

Figura 92: Gráfico de E_t no ponto de operação *Medium*, comparando LH e e\γ, com 1.52>$|\eta|$>2.01.

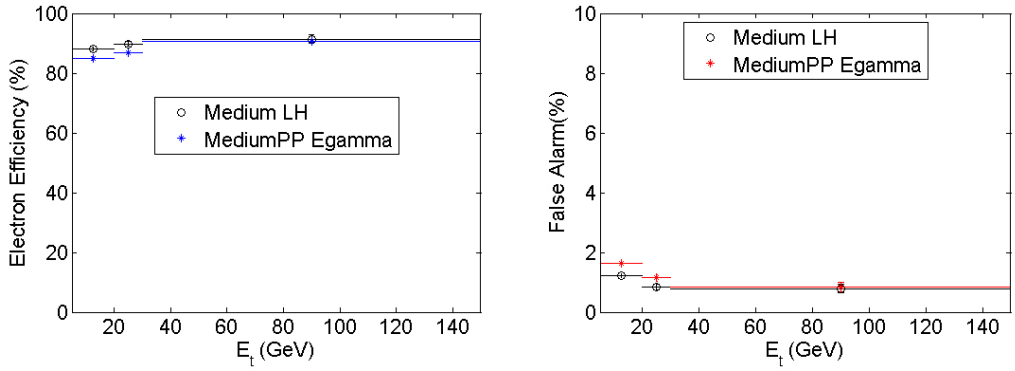

Figura 93: Gráfico de E_t no ponto de operação *Medium*, comparando LH e e\γ, com 2.01>$|\eta|$>2.47.

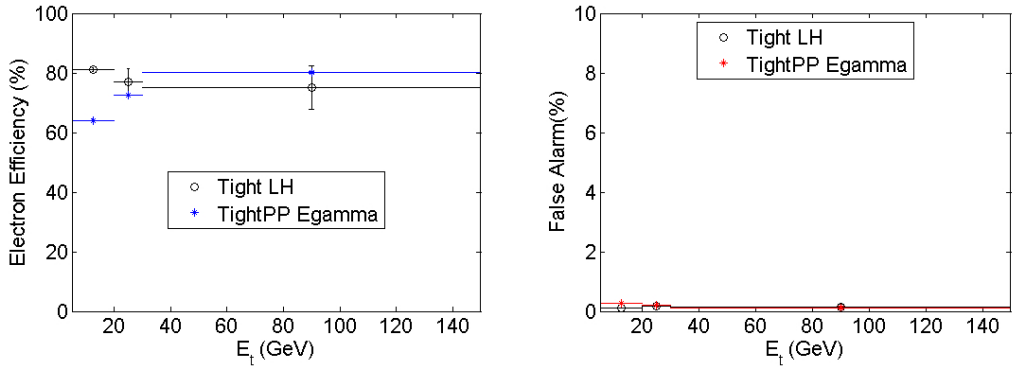

Figura 94: Gráfico de E_t no ponto de operação *Tight*, comparando LH e e\γ, com 0.8>$|\eta|$>1.37.

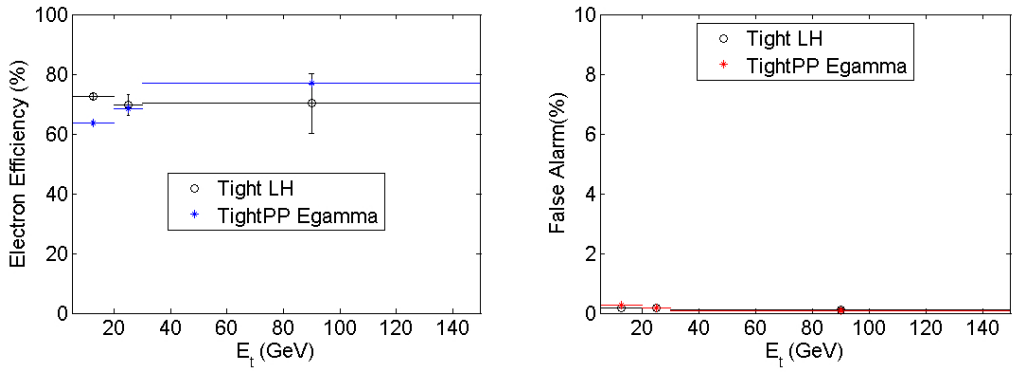

Figura 95: Gráfico de E_t no ponto de operação *Tight*, comparando LH e e\γ, com 1.52>$|\eta|$>2.01.

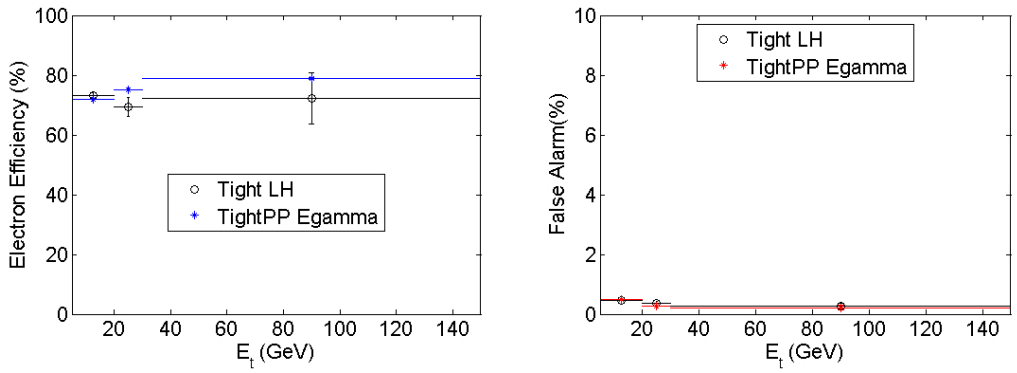

Figura 96: Gráfico de E_t no ponto de operação *Tight*, comparando LH e e\γ, com 2.01>$|\eta|$>2.47.

B.3 GRÁFICOS DE NVTX

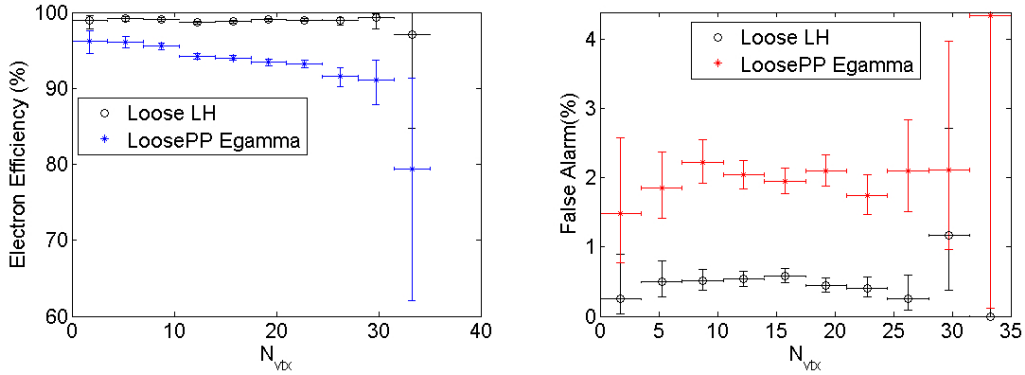

Figura 97: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 2.

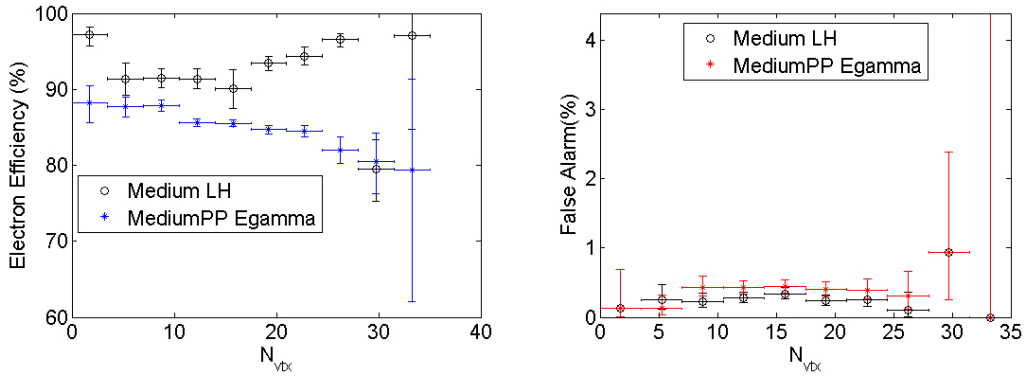

Figura 98: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 2.

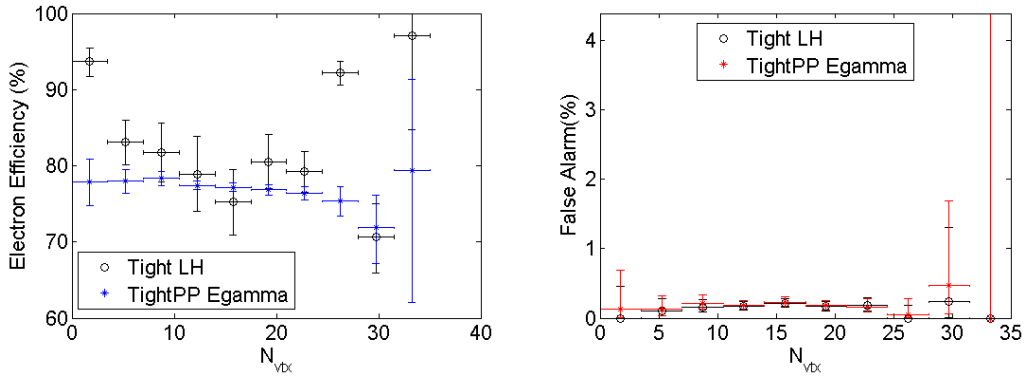

Figura 99: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 2.

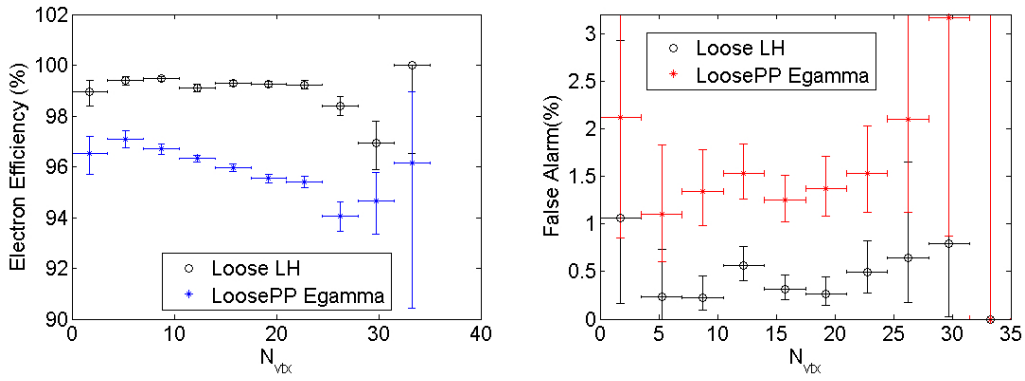

Figura 100: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 3.

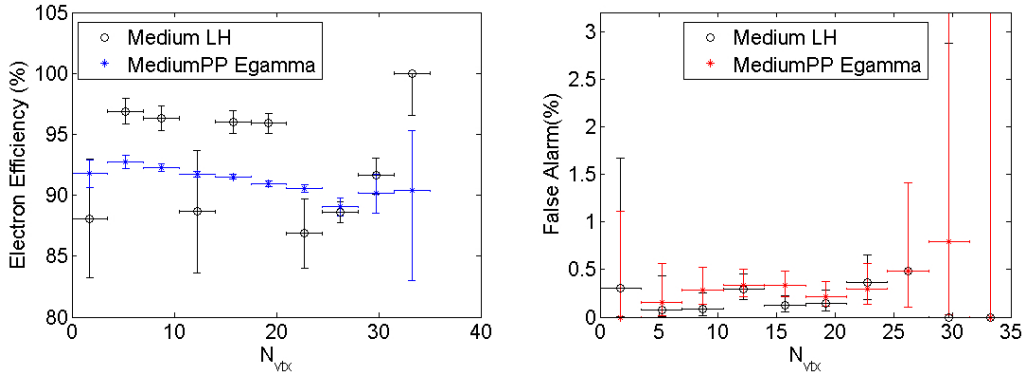

Figura 101: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 3.

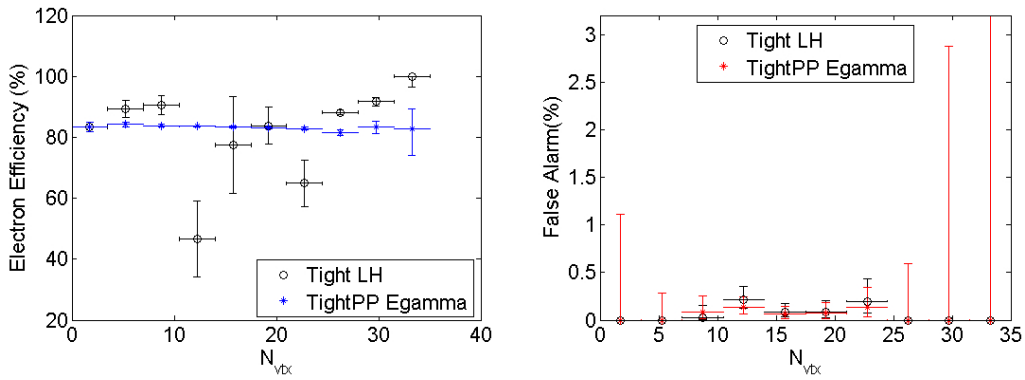

Figura 102: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 3.

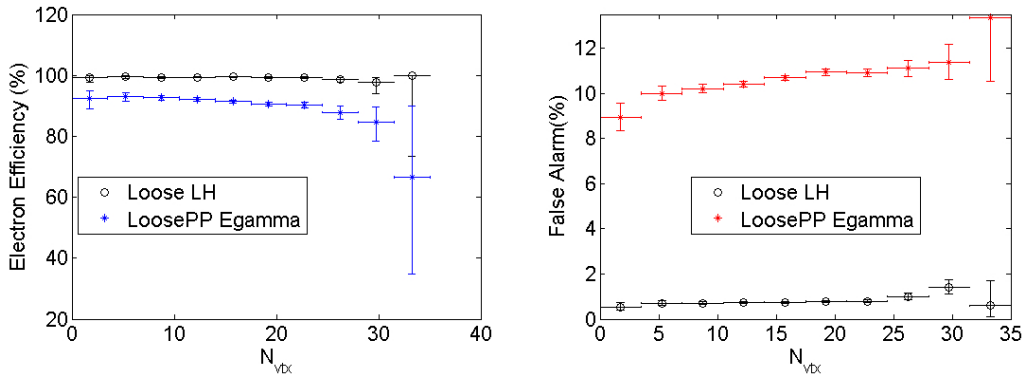

Figura 103: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 4.

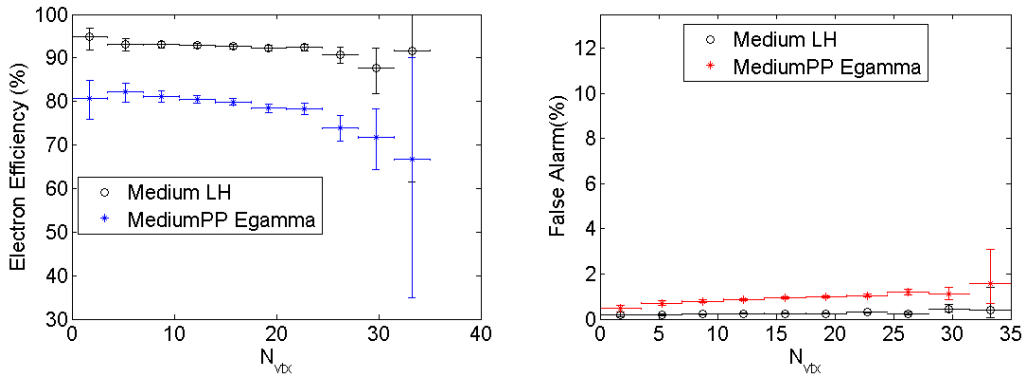

Figura 104: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 4.

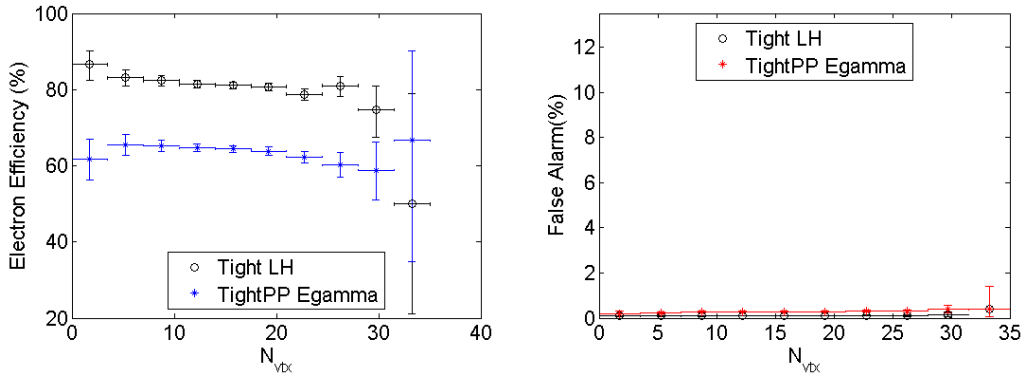

Figura 105: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 4.

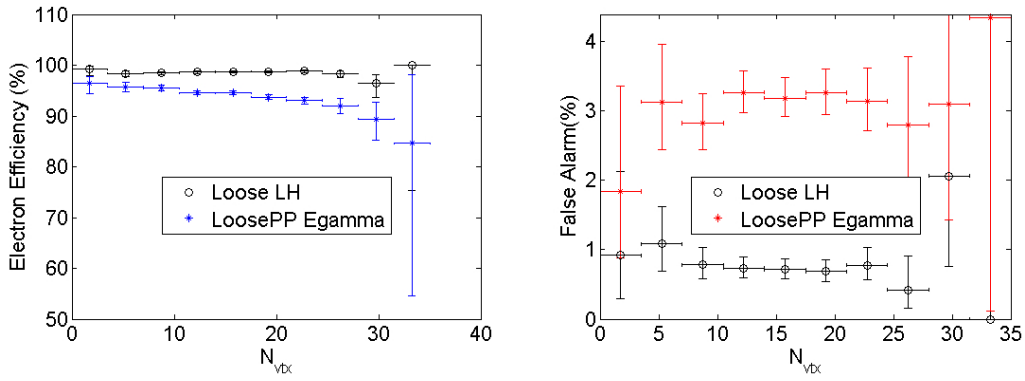

Figura 106: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 5.

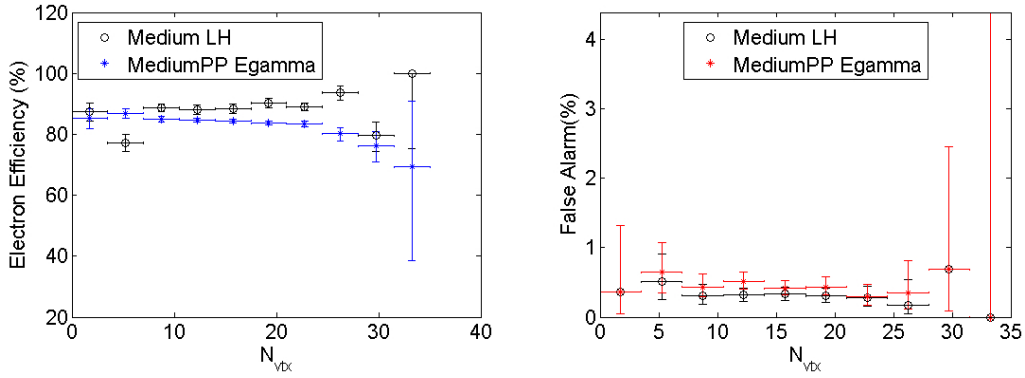

Figura 107: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 5.

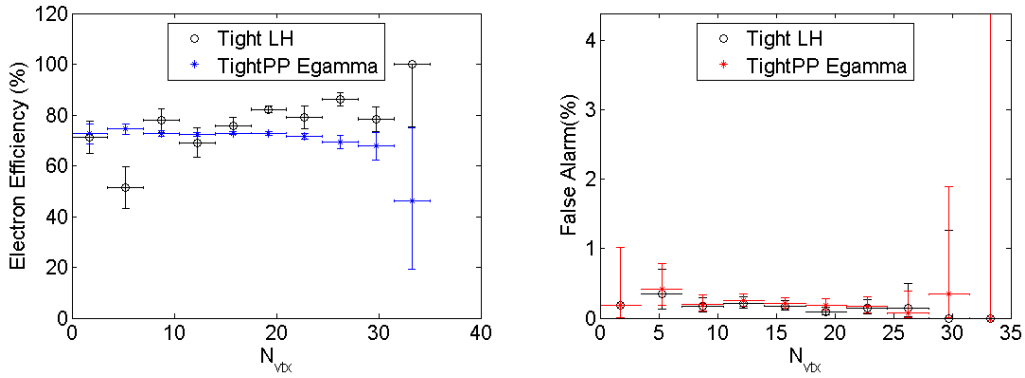

Figura 108: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 5.

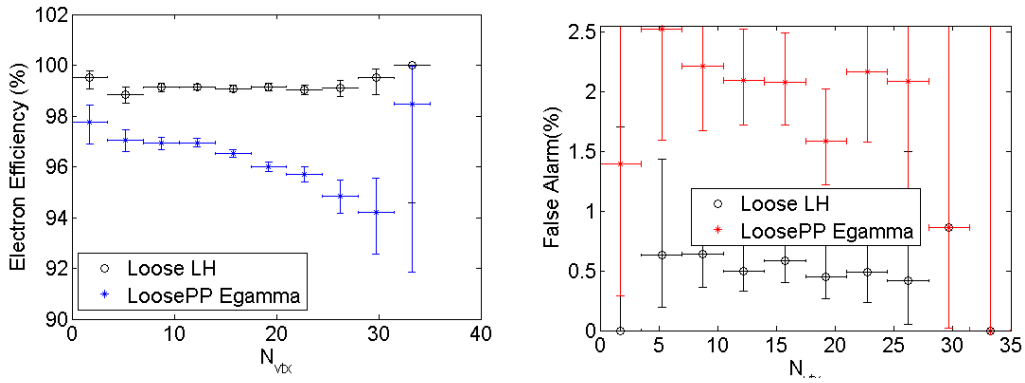

Figura 109: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 6.

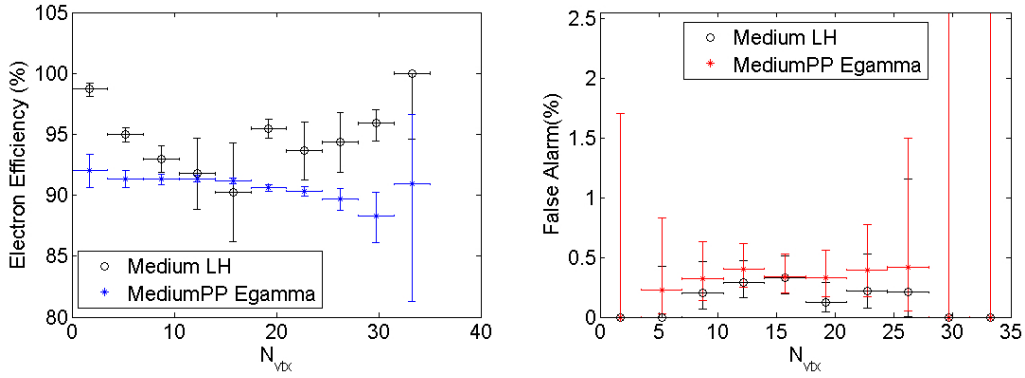

Figura 110: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 6.

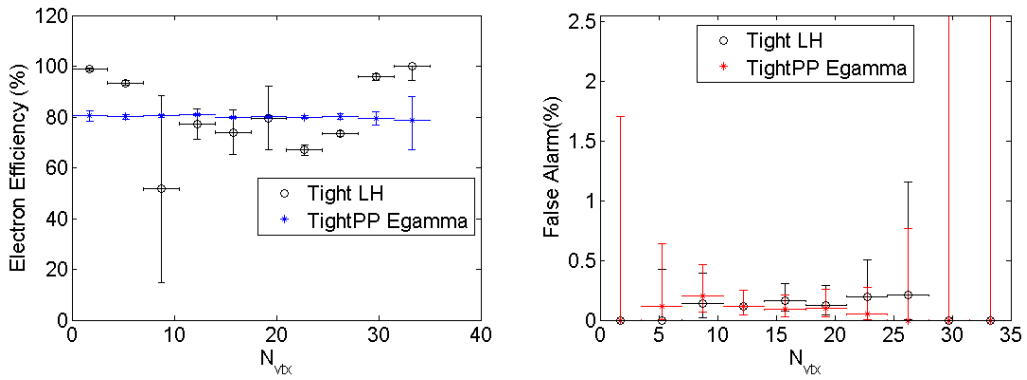

Figura 111: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 6.

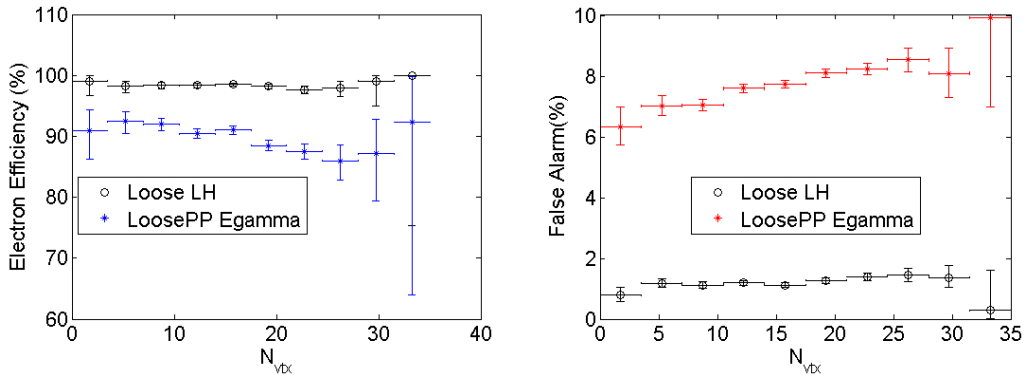

Figura 112: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 7.

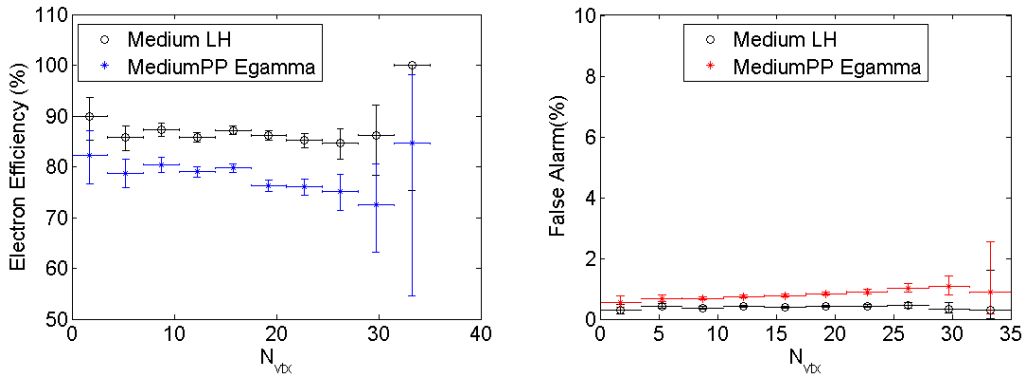

Figura 113: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 7.

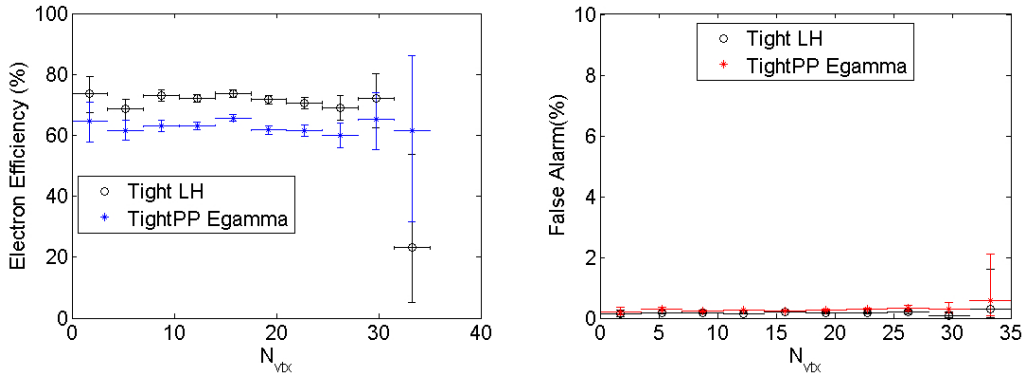

Figura 114: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 7.

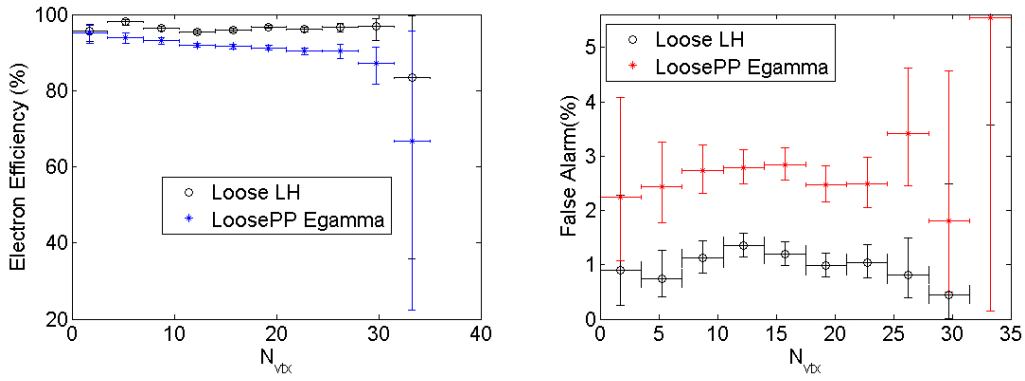

Figura 115: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 8.

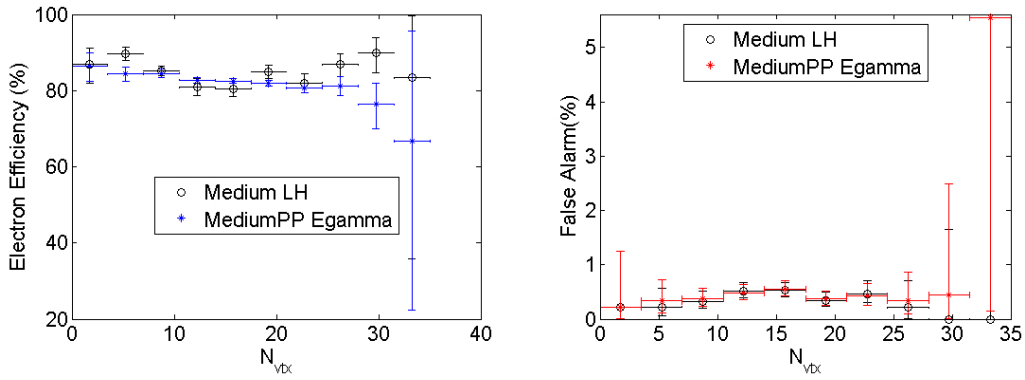

Figura 116: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 8.

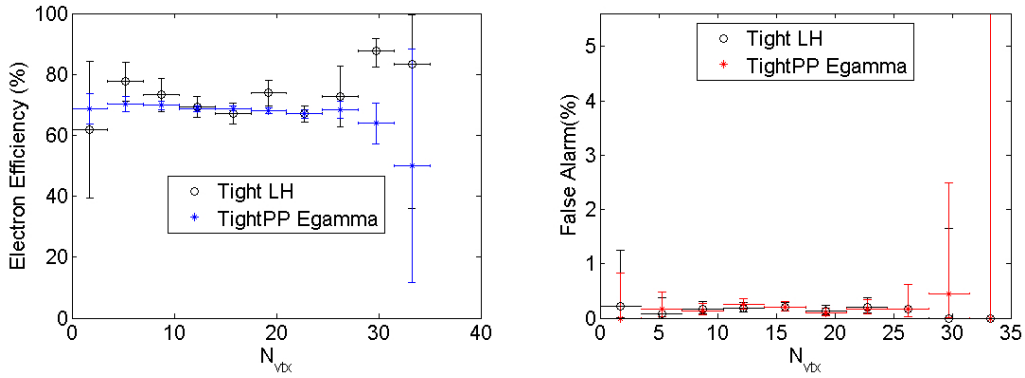

Figura 117: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 8.

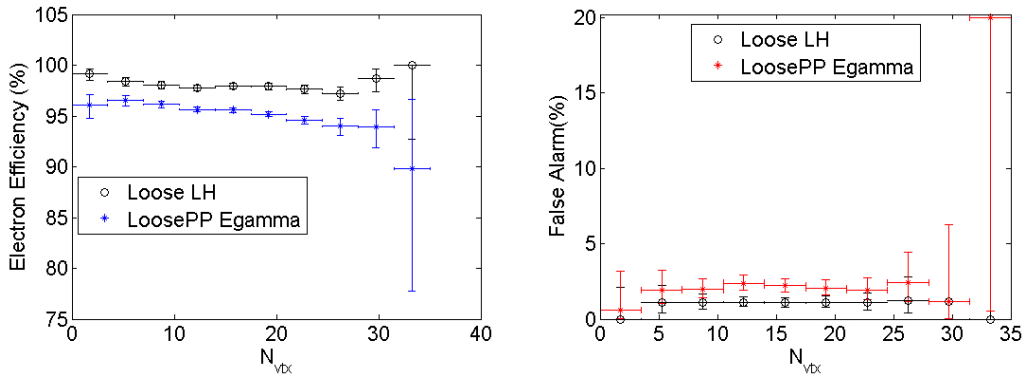

Figura 118: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 9.

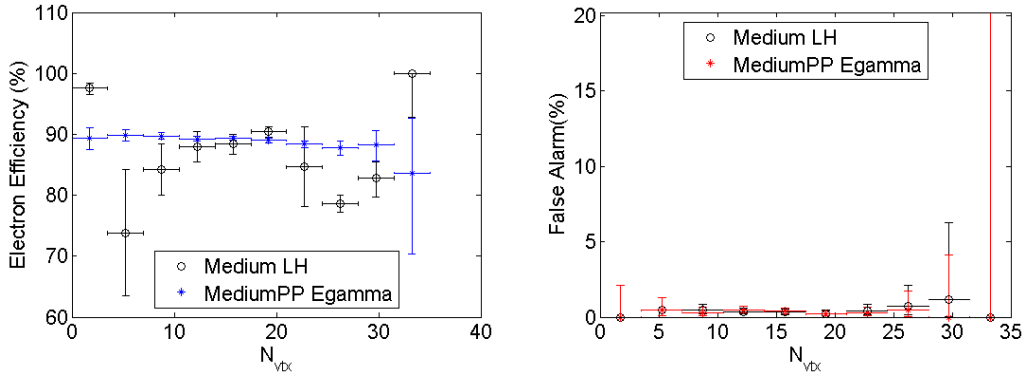

Figura 119: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 9.

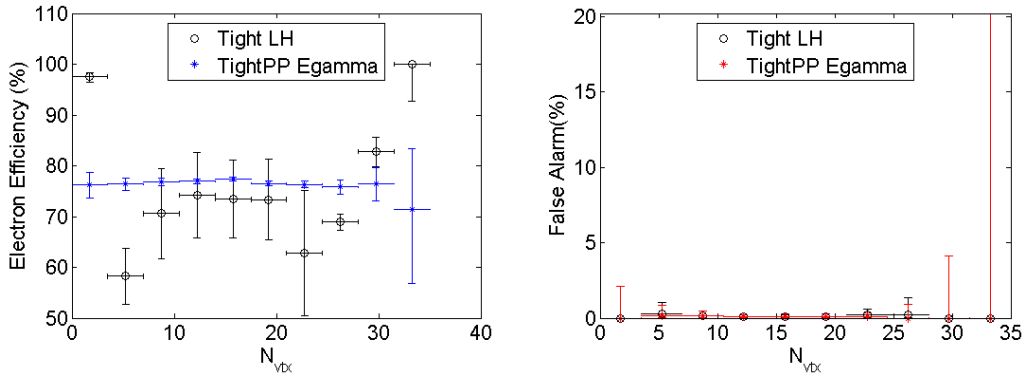

Figura 120: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 9.

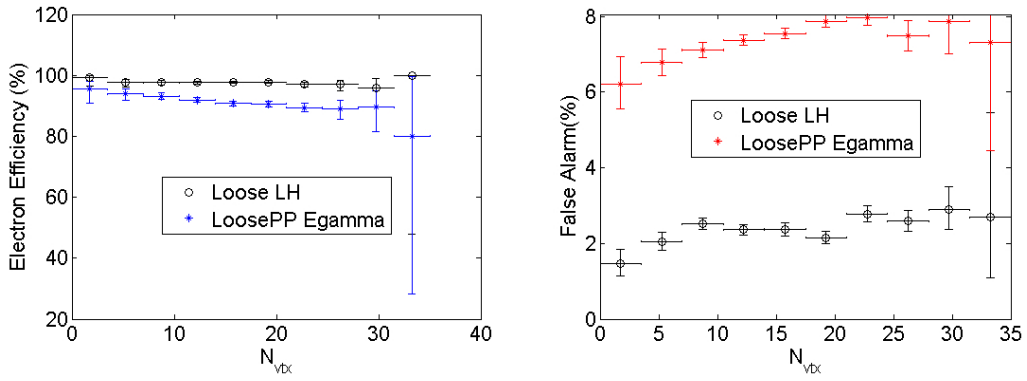

Figura 121: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 10.

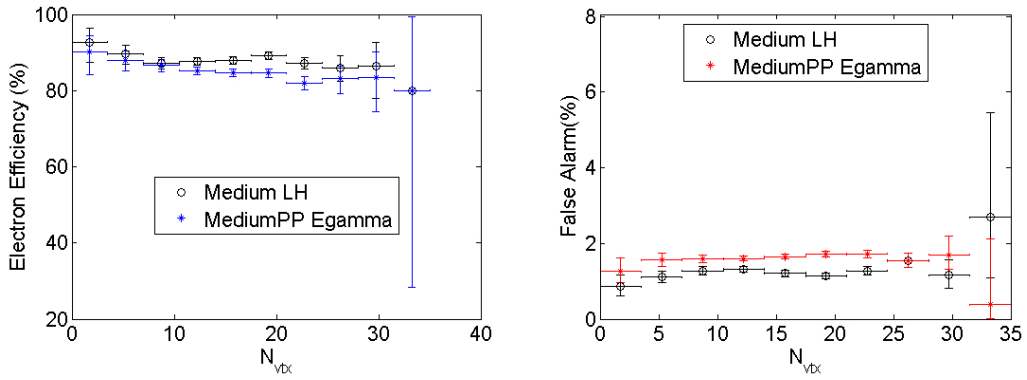

Figura 122: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 10.

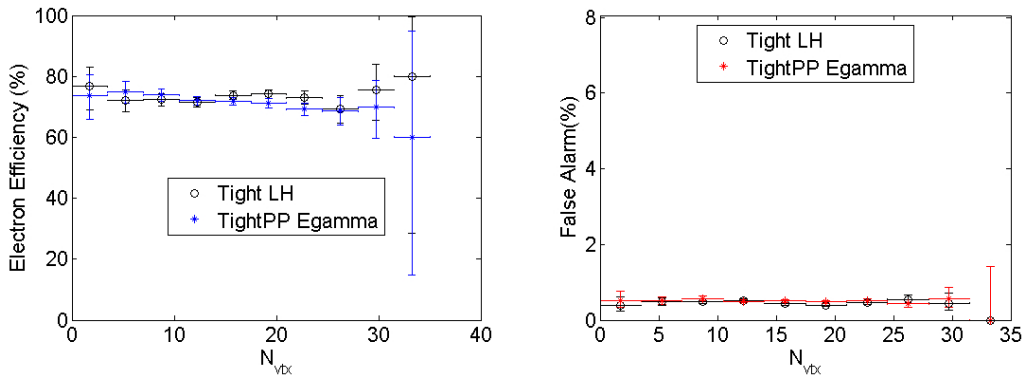

Figura 123: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 10.

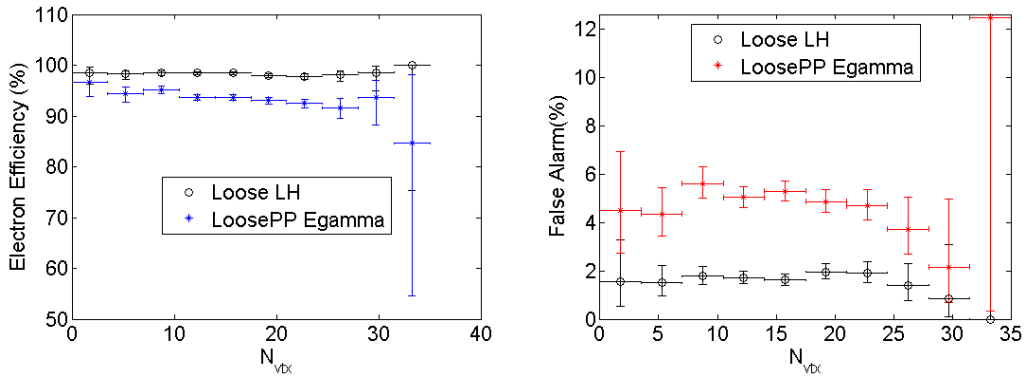

Figura 124: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 11.

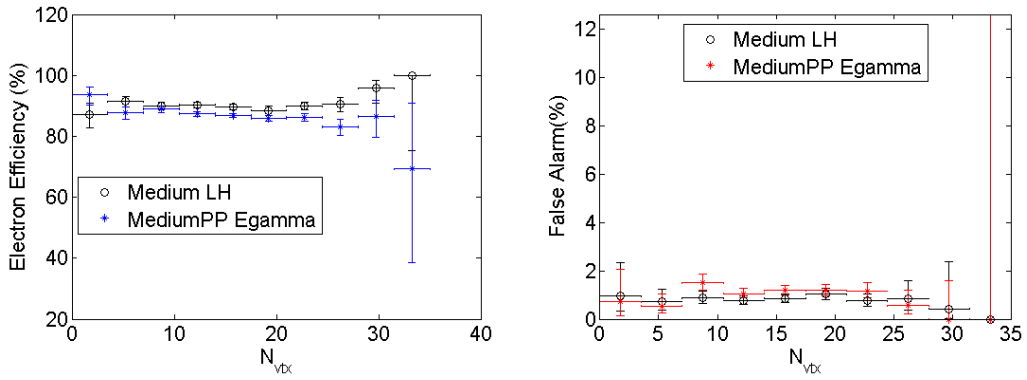

Figura 125: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 11.

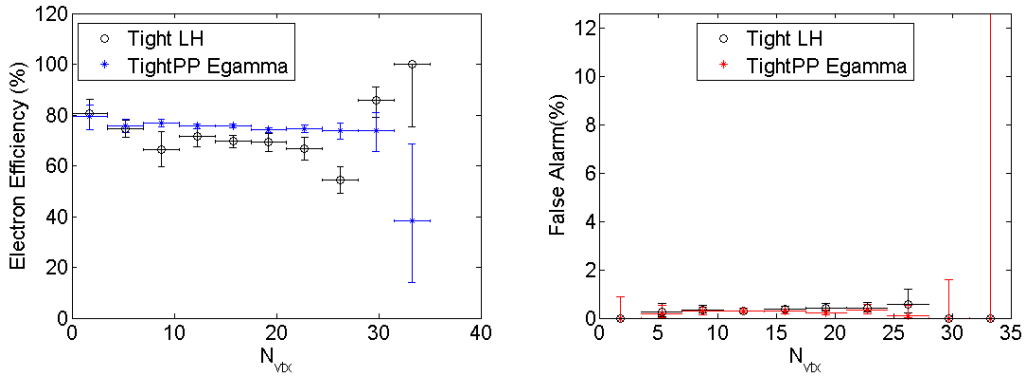

Figura 126: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 11.

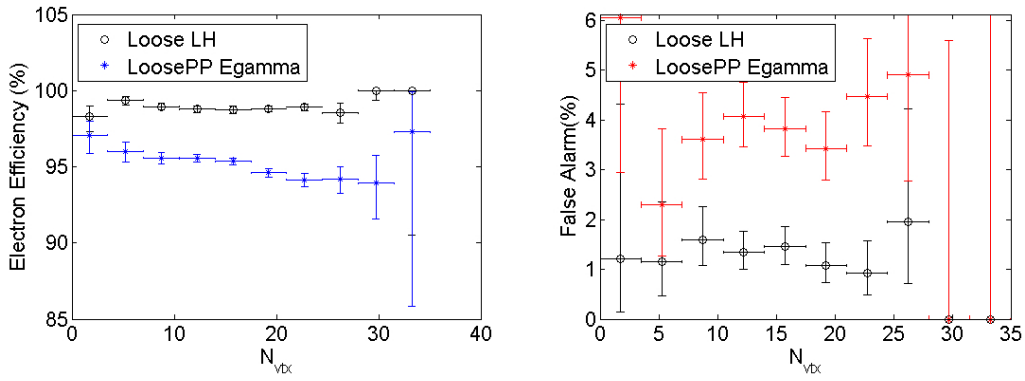

Figura 127: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Loose*, comparando LH e e\γ, para a Região 12.

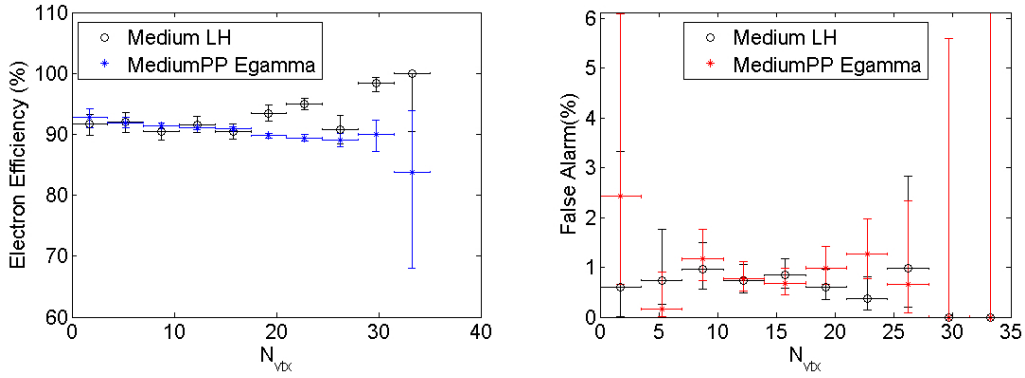

Figura 128: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Medium*, comparando LH e e\γ, para a Região 12.

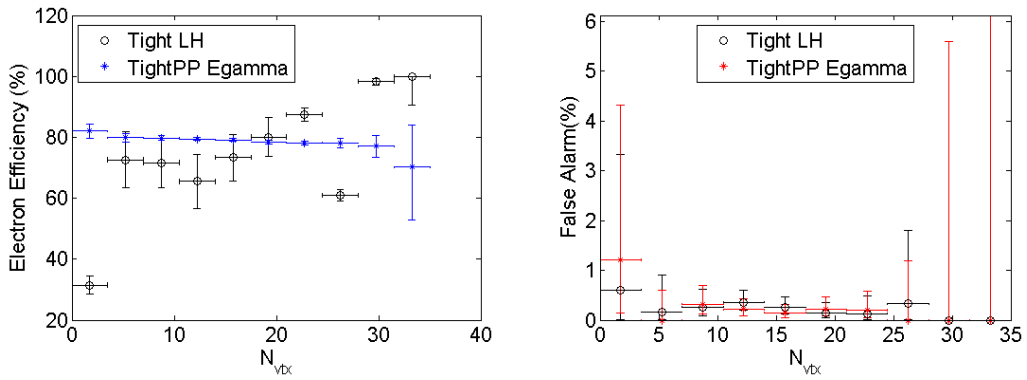

Figura 129: Gráfico de Eficiência de Sinal\Falso Alarme por NVTX, no ponto de operação *Tight*, comparando LH e e\γ, para a Região 12.

B.4 GRÁFICOS DE INFORMAÇÃO MÚTUA

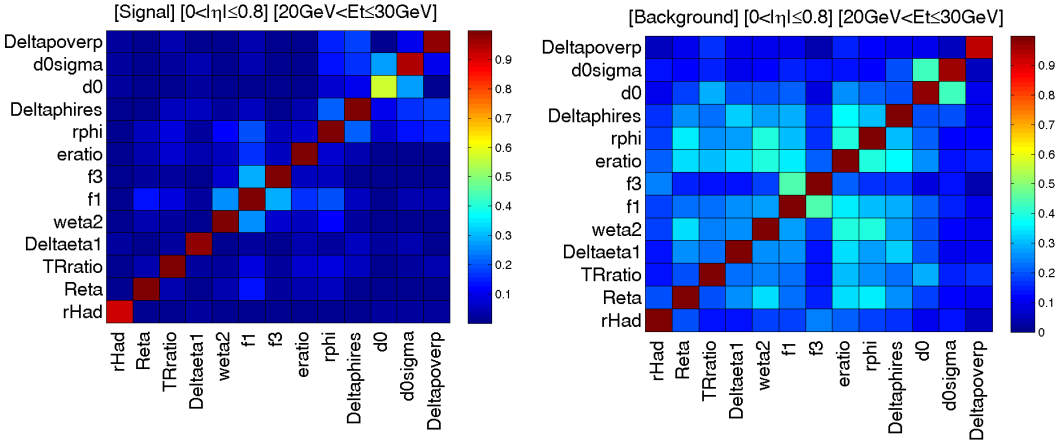

Figura 130: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 2.

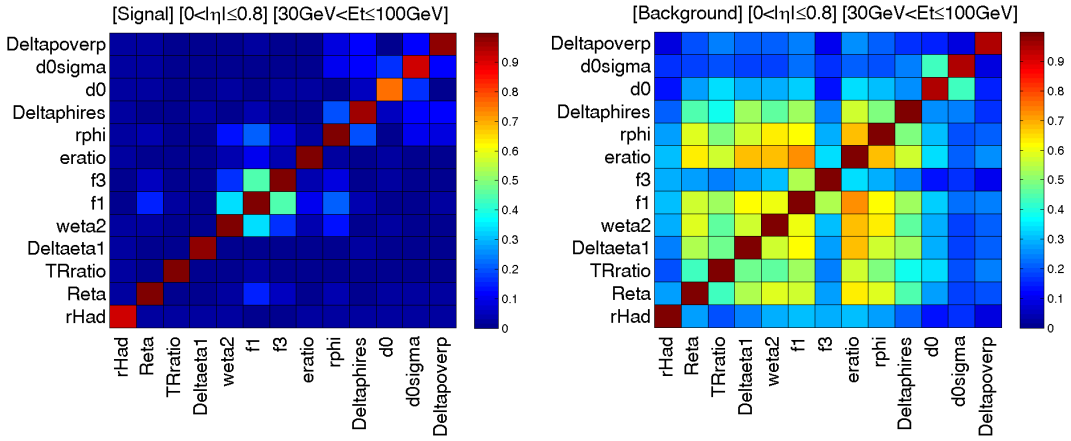

Figura 131: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 3.

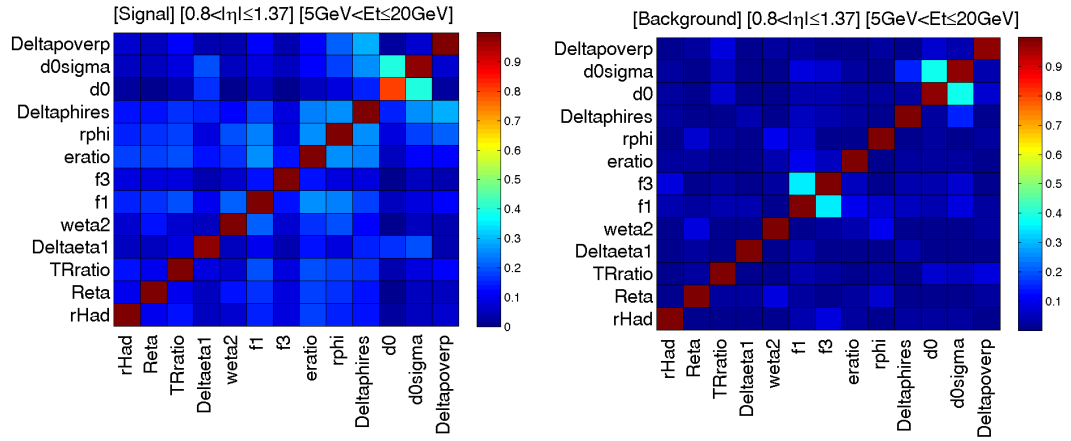

Figura 132: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 4.

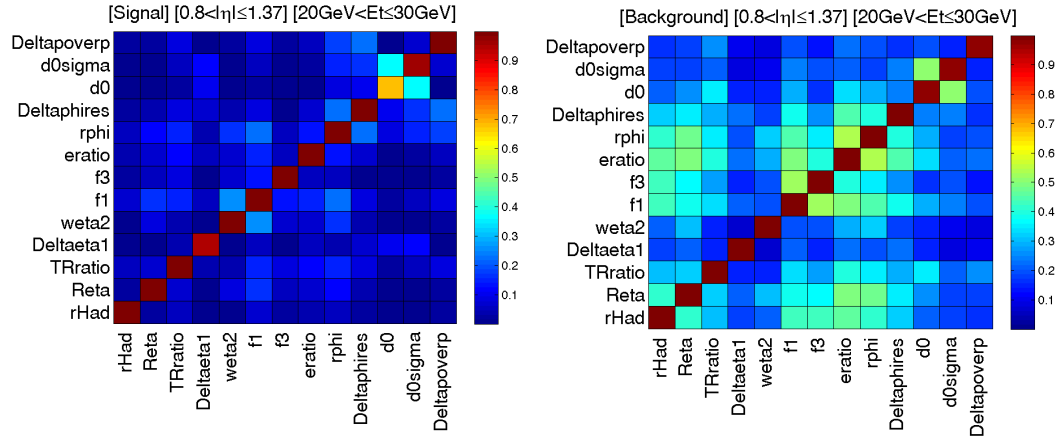

Figura 133: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 5.

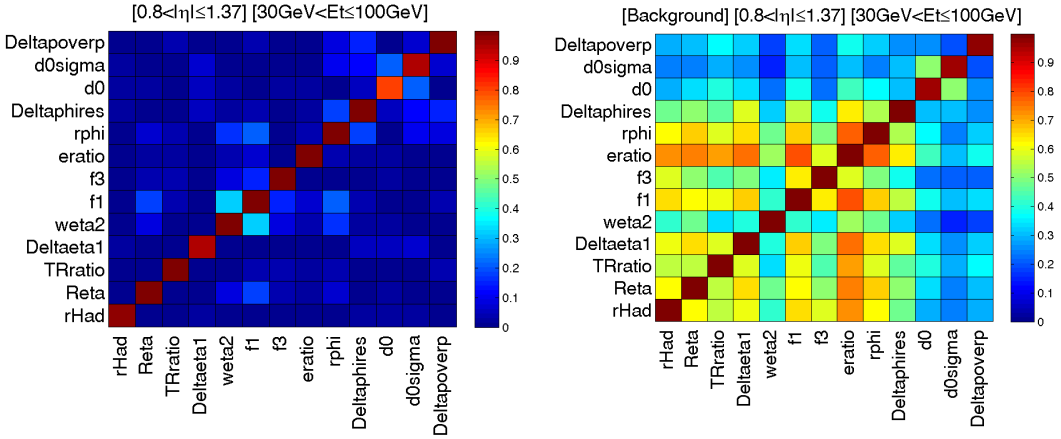

Figura 134: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 6.

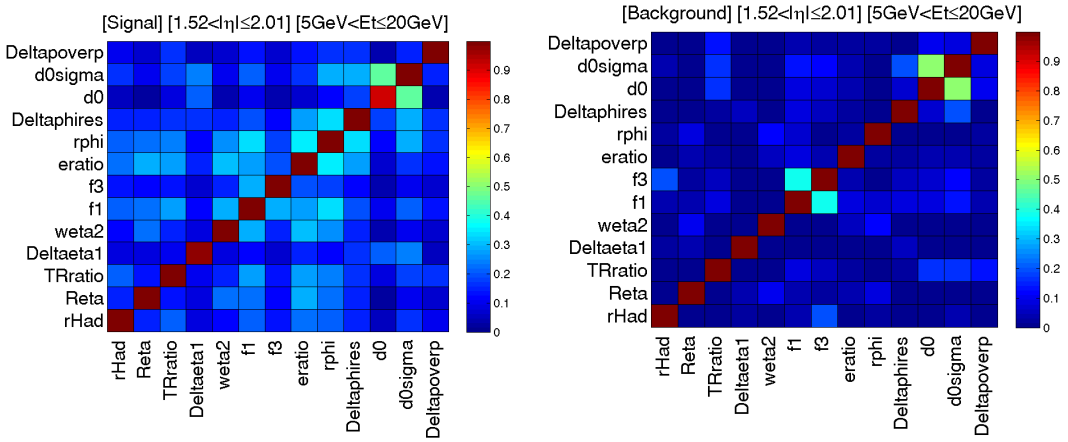

Figura 135: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 7.

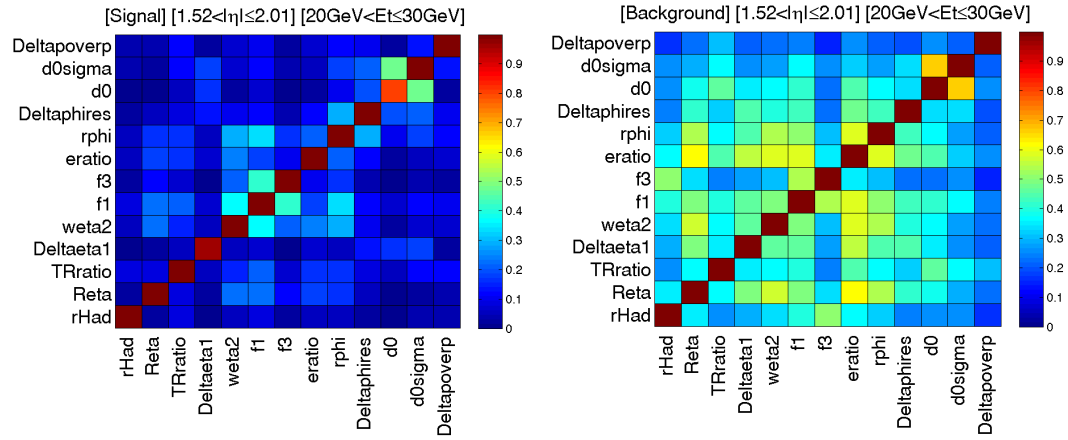

Figura 136: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 8.

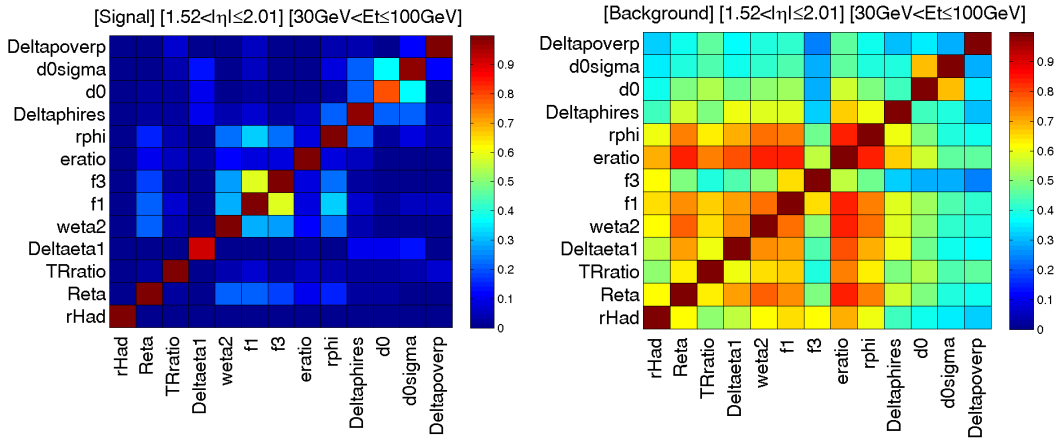

Figura 137: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 9.

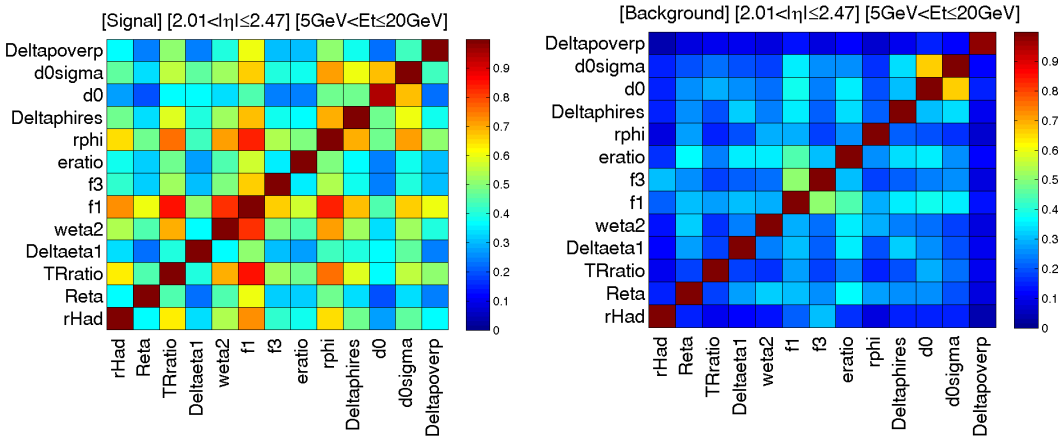

Figura 138: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 10.

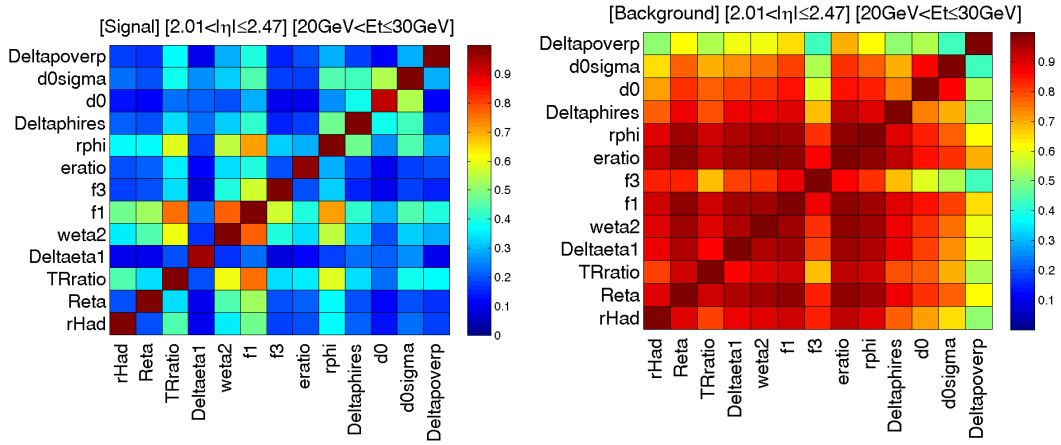

Figura 139: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 11.

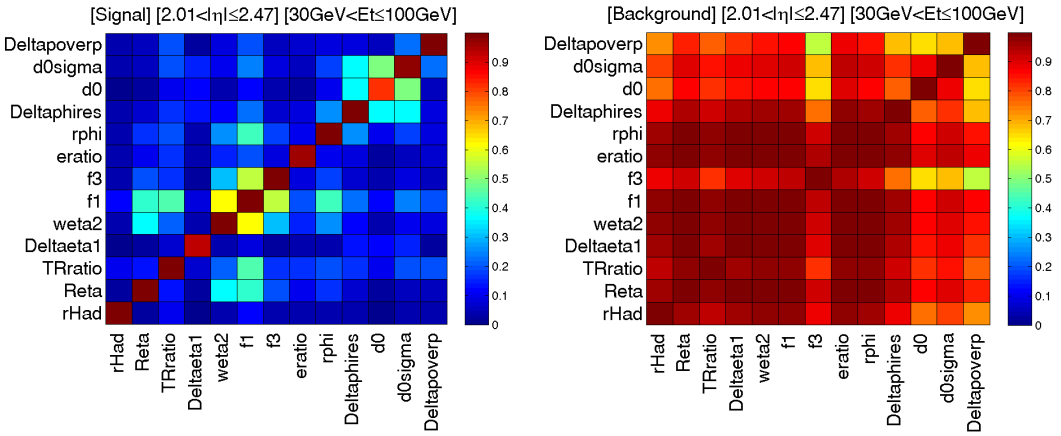

Figura 140: Gráfico de Informação Mútua do Sinal (Esquerda) e Ruído de Fundo (Direita), para a Região 12.

B.5 GRÁFICOS DE ROC DA ANÁLISE MULTIVARIADA

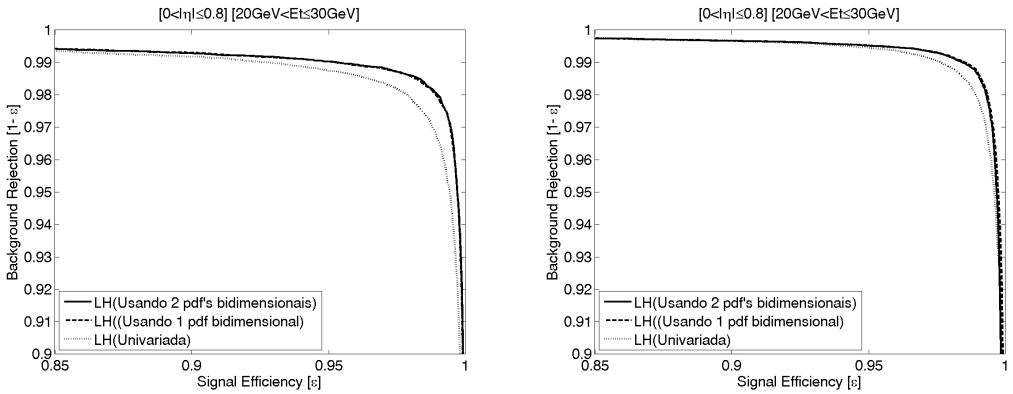

Figura 141: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 2. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

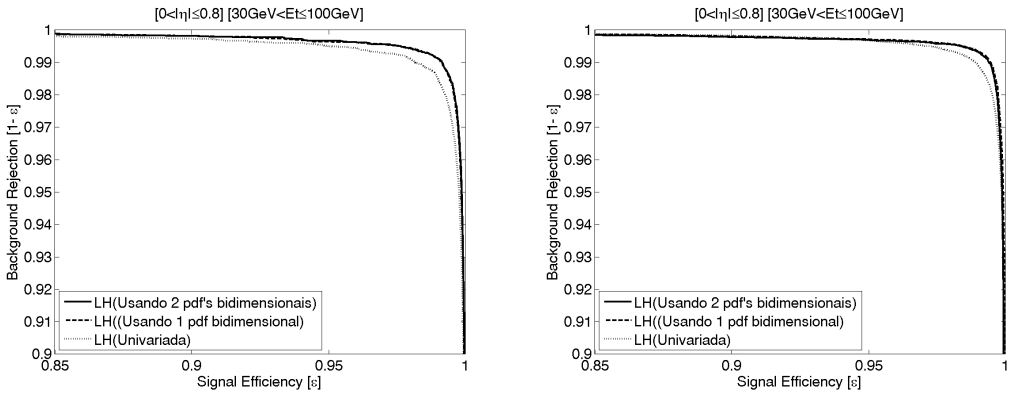

Figura 142: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 3. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

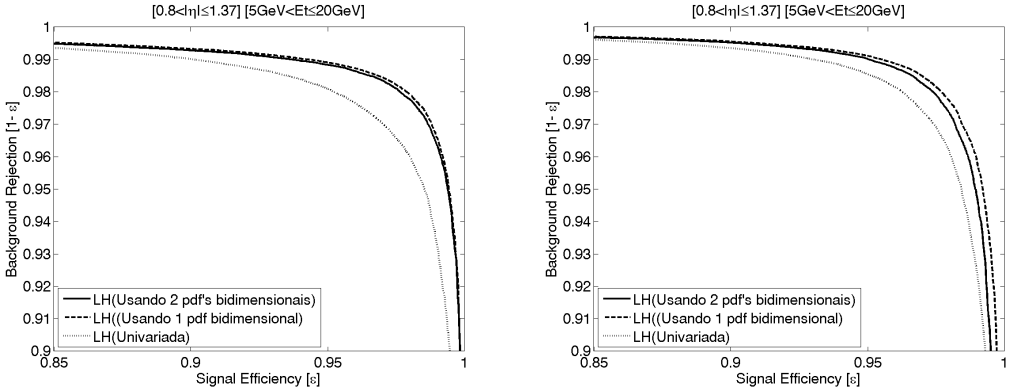

Figura 143: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 4. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

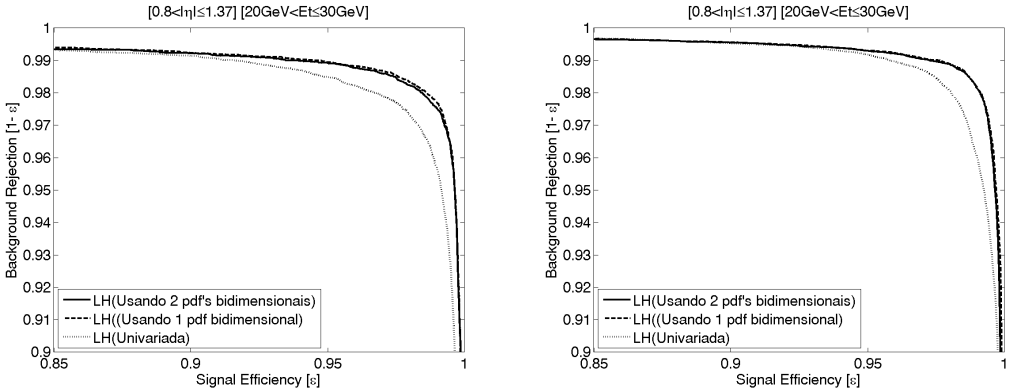

Figura 144: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 5. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

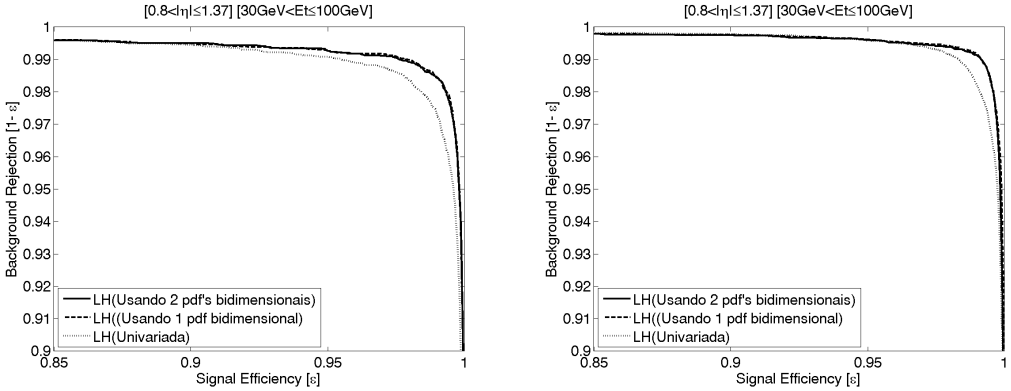

Figura 145: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 6. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

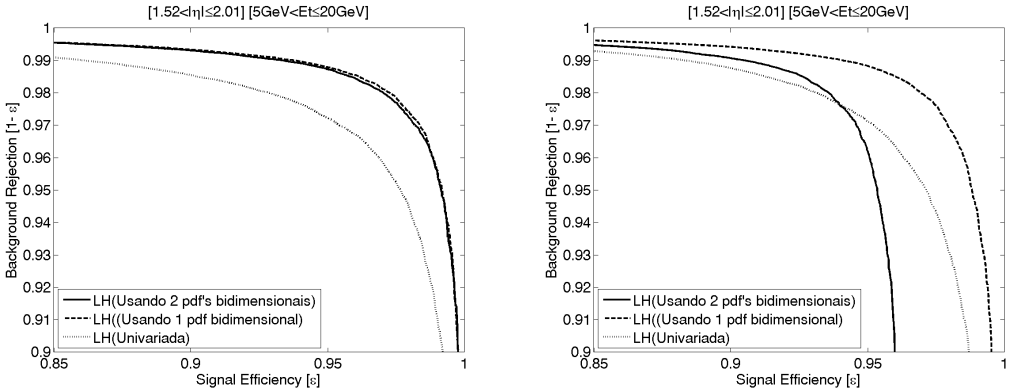

Figura 146: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 7. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

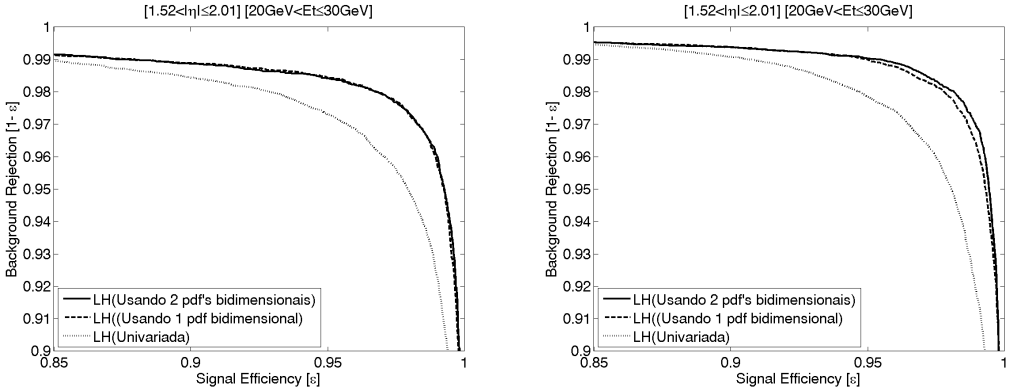

Figura 147: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 8. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

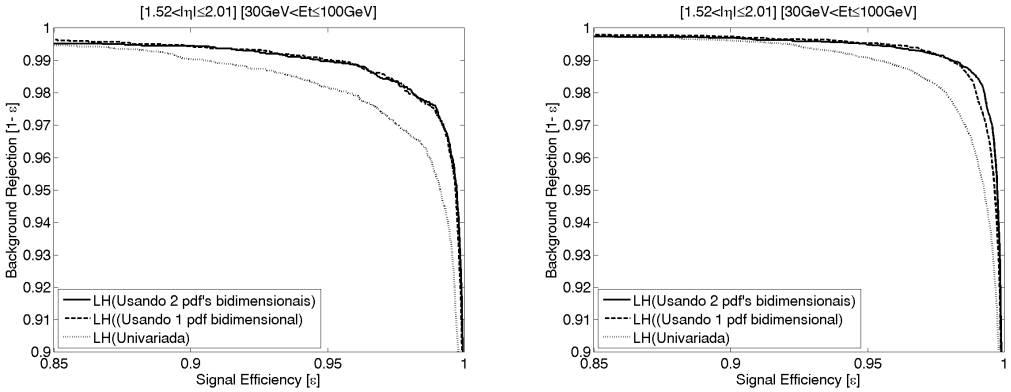

Figura 148: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 9. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

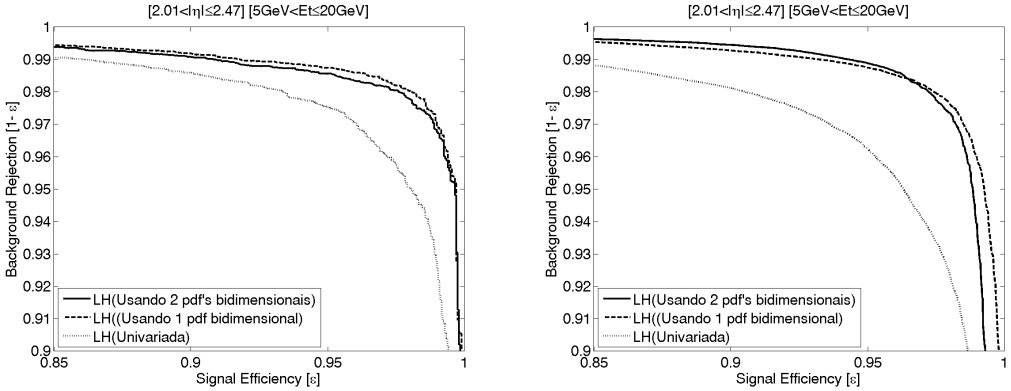

Figura 149: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 10. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

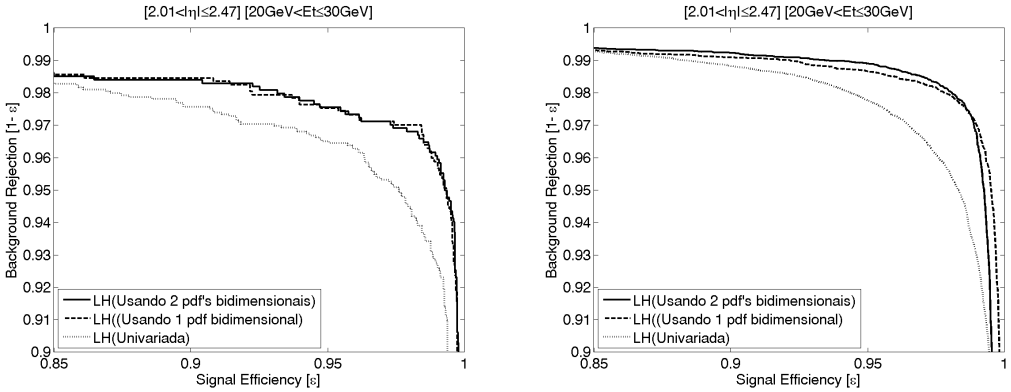

Figura 150: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 11. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.

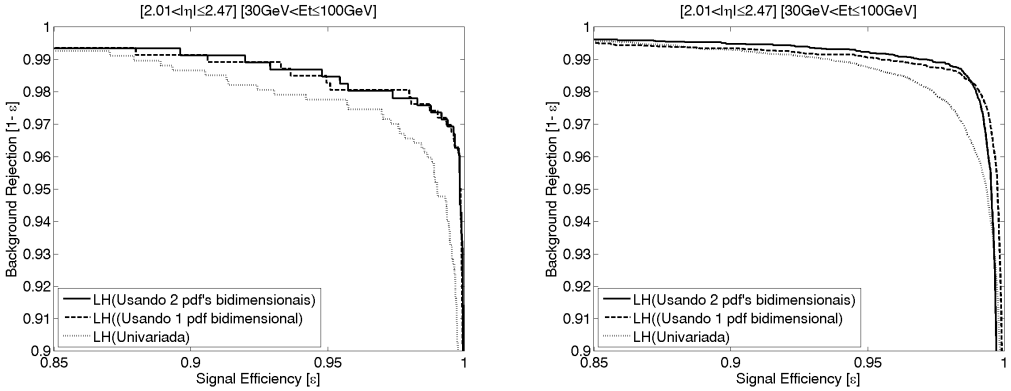

Figura 151: Gráfico comparando as ROC's da *Likelihood*: univariada, utilizando 1 PDF Bidimensional e utilizando 2 PDF's Bidimensionais, para Região 12. (Esquerda) Conjunto de Desenvolvimento e (Direita) Conjunto de Validação.