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

Igor Abritta Costa

Otimização dos Algoritmos Univariados e Bivariados aplicados à identificação de elétrons no experimento ATLAS

Dissertação de Mestrado

Juiz de Fora
2016

Igor Abritta Costa

Otimização dos Algoritmos Univariados e Bivariados aplicados à identificação de elétrons no experimento ATLAS

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
2016

Igor Abritta Costa

Otimização dos Algoritmos Univariados e Bivariados aplicados à identificação de elétrons no experimento ATLAS

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 22 de Fevereiro de 2016.

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. Luciano Manhães de Andrade Filho, D.Sc.
Universidade Federal de Juiz de Fora, UFJF

Prof. Erica Ribeiro Polycarpo Macedo, D.Sc.
Universidade Federal do Rio de Janeiro, UFRJ

Aos meus pais, meus irmãos, aos meus familiares, aos meus amigos.

AGRADECIMENTOS

Agradeço a Deus, pelo dom da vida e por sempre iluminar meu caminho, possibilitando assim que eu chegasse aqui hoje.

Aos meus pais, Marcelino e Isabel, por serem sempre meu porto-seguro e por estarem sempre ao meu lado, me guiando no caminho certo. Pelo respeito, afeto e amor incondicional dedicados a mim. Ao meu pai, por ter me ensinado umas das lições mais importantes da minha vida, a ser humilde sempre e respeitar toda e qualquer pessoa. A minha mãe, por ter me ensinado a não desistir por mais difícil que seja o caminho, por ser meu espelho e, principalmente, por me amar em todos os momentos da minha vida. Meu muito Obrigado. Só tenho a dizer "Muito obrigado" e "Amo vocês".

Aos meus tios, Garonce, Maria Inês e Cláudio, e aos meus padrinhos, Júlio e Naílce pelo apoio e conselhos, não somente a mim, mas a toda minha família durante esta jornada. Muito obrigado, continuo contando com vocês.

Ao meu irmão José, por conseguir ser pai, irmão mais velho e amigo, nos momentos que eu mais preciso, por acreditar e confiar em mim, por ser meu exemplo e por me fazer sentir orgulho de responder "Sim" quando me fazem a pergunta "Você é irmão do José Abritta ?". Conta comigo sempre, irmão.

Ao meu irmão Lucas, por ter ajudado na formação dos meus princípios e por me mostrar que dedicação e esforço andam lado a lado com o sucesso. Obrigado e conta comigo sempre.

Aos meus tios e primos, que mesmo apesar da distância estão sempre presentes e fazendo a diferença na minha vida.

Aos meus nobres professores, que com profissionalismo, dedicação e empenho contribuíram para minha formação profissional e pessoal. Em especial ao meu orientador, Rafael, que foi imprescindível neste projeto. Meus sinceros agradecimentos.

Ao meu amigo e companheiro de colaboração David. Por cada *insight* e conselhos fundamentais para o progresso deste trabalho.

Aos companheiros do LAPTEL e amigos de Laboratório Tiago e Kátia. Pelas conversas, ajudas e bom humor. É sempre bom ter alguém para compartilhar alegrias

e desespero.

Finalmente, agradecemos 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.

Não vemos as coisas como elas são, mas como nós somos.

Anaïs Nin

RESUMO

A identificação de elétrons é de fundamental importância para os grupos de física do experimento ATLAS, devido à presença destes no processo final de decaimento de partículas de interesse. Nesse ambiente de física de partículas, a probabilidade de ocorrência de elétrons relevantes aos estudos propostos são baixíssimas em relação às partículas que formam o ruído de fundo, exigindo dos grupos de performance do ATLAS algoritmos com índices de eficiência de detecção dos sinais de interesse e rejeição de ruído de fundo cada vez melhores. Nessa dissertação, os métodos aplicados na identificação de elétrons no experimento ATLAS serão revisados e possíveis otimizações serão avaliadas a partir dos dados produzidos pelo ATLAS. Concentrado no contexto *offline*, o trabalho reproduz o método baseado em verossimilhança e propõe uma melhoria com o uso da técnica multivariada conhecida como MKDE (do inglês, *Multivariate Kernel Density Estimation*), capaz de mitigar o erro inserido na consideração de dependência entre as variáveis discriminantes inserida pelo método de *Likelihood* atualmente em uso pelo ATLAS. Inicialmente, este trabalho se propõe a implementar o método de verossimilhança em uso, que se baseia em densidade univariadas usadas na reconstrução da densidade conjunta das variáveis discriminantes, e a estudar o impacto de possíveis parâmetros relacionados à implementação do algoritmo de estimação de densidades univariadas. Este método será então comparado com o método padrão do ATLAS conhecido como e/γ. Em uma segunda etapa, a implementação do MKDE é inserida através de uma comparação direta com o método univariado.

Palavras-chave: Identificação de Elétrons, Máxima Verossimilhança, KDE Multivariado.

ABSTRACT

The electron identification is of fundamental importance for the ATLAS physics groups due to the presence of these in the final process of interest particles decay. In particle physics environment, the occurrence probability of relevant electrons to the proposed studies are very low compared to particles considered background, requiring ATLAS performance groups algorithms with identification efficiency index and background rejection each time better. In this dissertation, the methods applied in the electron identification in ATLAS experiment will be reviewed and possible optimizations will be evaluated from the data produced by the ATLAS experiment. Concentrated in the offline context, the work reproduces the method based on Likelihood and proposes an improvement with the use of multivariate technique known as MKDE (Multivariate Kernel Density Estimation), capable of mitigate the error inserted in consideration of dependence between discriminating variables entered by the method Likelihood currently in use by ATLAS. Initially, this work proposes to implement the method Likelihood in use, which is based on univariate density used in the reconstruction of the joint density of the discriminant variables, and to study the possible impact of parameters related to the implementation of univariate densities estimation algorithm. This method is then compared with the ATLAS standard method known as e/γ. In a second step the implementation of the MKDE is inserted through a direct comparison to the univariate method.

Keywords: Electron identification, Likelihood, Multivariate KDE.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ALICE *A Large Ion Collider Experiment*

AMISE *Asymptotic Mean Integrated Squared Error*

AMSE *Asymptotic Mean Squared Error*

ATLAS *A Toroidal LHC Apparatus*

AUC Área Sob a Curva, (do inglês, *Area Under the ROC Curve*)

BCV *Biased Cross-Validation*

CC *Center-Center*

CERN Centro Europeu de Pesquisa Nuclear, (do francês, *Conseil Européen pour la Recherche Nucléaire*)

CMS *Compact Muon Solenoid*

CT *Center-Tail*

EF Filtro de Eventos

EM Calorímetro Eletromagnético

HAD Calorímetro Hadrônico

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

IAE *Integrated Absolute Error*

ISE *Integrated Squared Error*

ID Detector Interno

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

L1 *Level 1*

L2 *Level 2*

LH *Likelihood*

LHC *Large Hadron Collider*

LHCb *Large Hadron Collider beauty*

LINAC 2 Acelerador Linear

LSCV *Least-Square Cross-Validation*

MISE *Mean Integrated Squared Error*

MKDE *Multivariate Kernel Density Estimation)*

MSE *Mean Squared Error*

PDF Função de Densidade de Probabilidade

PS *Proton Synchrotron*

PSB *Proton Synchrotron Booster*

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

RoI Regiões de Interesse, (do inglês,*Region of Interest*)

SCT *SemiConductor Tracker*

SPD *Silicon Pixel Detector*

SPS *Super Proton Synchrotron*

TC *Tail-Center*

TFM Tubos Fotomultiplicadores

TMVA *Toolkit for Multivariate Data Analysis*

TRT *Transition Radiation Tracker*

TT *Tail-Tail*

SUMÁRIO

1 INTRODUÇÃO

Em ambientes industriais ou em laboratórios de pesquisa, a análise multivariada tem se mostrado uma ferramenta poderosa na resolução de vários problemas ligados à estimação de densidades e seleção de eventos; assuntos estes de grande afinidade com a área de engenharia elétrica. Todavia, problemas desse tipo ocorrem em várias áreas do conhecimento e a busca por uma solução multidisciplinar se torna de grande interesse.

Nas últimas décadas, vários experimentos geradores de enorme quantidade de dados foram iniciados, fazendo com que a importância de uma modelagem estocástica por funções densidade de probabilidade, utilizando-se de métodos não paramétricos, crescesse consideravelmente. Os experimentos ligados ao LHC (do inglês Large Hadron Collider) representam alguns deles. Em geral, a física experimental de altas energias é um dos ramos da ciência que mais exige de sistemas de processamento, tendo em vista que os eventos de interesse são muito raros e contaminados com alto nível de ruído de fundo, devendo o sistema ser capaz de eliminar a maior quantidade possível dos eventos que formam o ruído de fundo sem descartar os raros eventos de interesse. Posteriormente, faz-se necessário que estes eventos de interesse sejam identificados e selecionados com um mínimo de contaminação possível. Isso faz com que o desempenho dos algoritmos aplicados para este fim seja essencial, e uma busca contínua pela otimização dos mesmos faz-se indispensável.

Aqui, mais uma vez, o trabalho conjunto entre físicos e engenheiros pode ser realizado em prol da ciência, para que os experimentos consigam tirar o máximo de seus resultados, melhorando a sua relação custo-benefício, devido ao seu tamanho, tecnologia aplicada e mão de obra envolvida, onerosos para os países participantes, e ao mesmo tempo, extremamente relevantes para a evolução da ciência fundamental e aplicada. Uma das implicações deste processo ocorre na formação de profissionais dentro das diversas áreas de conhecimento, uma vez que tais experimentos necessitam resolver problemas complexos que, muitas vezes, exigem a aplicação de conhecimento avançado, promovendo, naturalmente, uma interação entre estudantes, professores/pesquisadores

e profissionais qualificados de diversos países.

Em geral, para problemas cujas variáveis podem ser modeladas, a estimação das mesmas se torna paramétrica. Na impossibilidade ou ineficácia do emprego de modelos paramétricos, técnicas não paramétricas devem ser utilizadas. A aplicação de métodos não paramétricos se espalhou consideravelmente nos últimos anos devido às ferramentas recentemente desenvolvidas para análise estatística. Essas ferramentas oferecem alternativas aos tradicionais métodos paramétricos na exploração de enormes quantidades de dados, sem que seja necessário pressupor qualquer distribuição específica. Estimação de densidade não paramétrica é um dos tópicos nesta linha que se tornaram objetos relevantes de pesquisa. Em particular, pesquisas teórica e aplicada relacionadas a temas como regressão, discriminação e reconhecimento de padrões, por exemplo, foram bastante influenciadas pelos recentes desenvolvimentos em estimação de densidade não paramétrica. Histrogramas, estimadores por núcleo e series ortogonais, por exemplo, ainda são bastante populares. Para casos multivariados, normalmente generalizações a partir de métodos univariados podem ser aplicados; porém, o custo computacional pode se tornar grande, comprometendo assim sua aplicabilidade. Neste contexto, um dos caminhos possíveis é a busca por uma combinação criteriosa entre processamento estatístico baseado em múltiplas variáveis e técnicas de estimação de densidade não paramétricas, de modo que a solução resultante seja, ao mesmo tempo, factível e robusta.

1.1 MOTIVAÇÃO

Nos anos de 2013 (COLLABORATION et al., 2013) e 2014 (COLLABORATION et al., 2014), a colaboração ATLAS publicou análises e resultados da técnica de verossimilhança aplicada à identificação de elétrons. Contudo, o algoritmo proposto utiliza uma simplificação da formulação matemática desse método, assumindo independência entre as variáveis, mesmo sabendo que algumas das variáveis em uso são dependentes. Com efeito, estudos anteriores demonstraram que existe degradação do desempenho de classificação de eventos e apontam possíveis caminhos para a otimização desta metodologia, sendo um deles a inclusão de densidades multivariadas na aplicação do método de *Likelihood*.

1.2 O QUE FOI FEITO

Esta dissertação se concentra na identificação de elétrons, no detector ATLAS, analisando os subprodutos das colisões geradas no LHC. Mais especificamente, no problema da dependência entre as variáveis utilizadas pelo algoritmos de identificação por verossimilhança. Foi desenvolvido o método univariado semelhante ao utilizado pelo grupo ATLAS, descrito em (COLLABORATION et al., 2013), e, além de avaliar parâmetros que podem levar à otimização deste método, um algoritmo de estimação de densidades bivariadas foi desenvolvido e aplicado aos dados de simulação do experimento. O desempenho dos algoritmos propostos foram avaliados em detalhe e comparados com os métodos em uso pela Colaboração ATLAS.

1.3 ESTRUTURA DA DISSERTAÇÃO

Este documento está organizado da seguinte maneira: o Capítulo 2 apresenta uma breve introdução da física de altas energias, e descreve o LHC e o experimento ATLAS. O Capítulo 3 apresenta os fundamentos do problema de identificação de sinais apresentando as variáveis empregadas no experimento e os algoritmos de identificação de elétrons *offline* atualmente implementados no ATLAS. O Capítulo 4 apresenta os fundamentos teóricos dos conceitos utilizados nessa dissertação. O Capítulo 5 detalha o funcionamento do algoritmo de identificação de elétrons proposto, suas melhorias e análises. O Capítulo 6 traz os resultados do método proposto e comparações desta técnica com o algoritmo de identificação de elétrons do Experimento ATLAS. Por fim, as principais conclusões são feitas no Capítulo 7, além de mostrar algumas propostas futuras.

2 FÍSICA DE ALTAS ENERGIAS E O CERN

Desde os anos setenta, os físicos de partículas têm descrito a estrutura fundamental da matéria usando uma elegante série de equações denominado Modelo Padrão. O modelo tenta descrever tudo que pode ser observado no universo, a partir de alguns blocos básicos chamados partículas fundamentais (CERN, 2015c).

Portanto, para a comprovação experimental dessa teoria, equipamentos que são capazes de colidir partículas em altas energias foram construídos, recriando um ambiente onde é possível observar, mais profundamente, as partículas fundamentais e seus processos de interação. Desde 1911, quando o primeiro acelerador de partículas foi criado pelo físico britânico Ernest Rutherford (ARAUJO, 2015), aceleradores vem sendo construídos com energias de colisão cada vez maiores. Em 2008, o Centro Europeu de Pesquisa Nuclear, (do francês, *Conseil Européen pour la Recherche Nucléaire*) (CERN) inaugurou o maior acelerador de partículas do mundo, o *Large Hadron Collider* (LHC), projetado para colidir prótons a uma energia de centro de massa de 14 TeV.

Esta seção é dedicada a ambientar e dar uma visão geral sobre qual é o objeto de estudo e onde este estudo é realizado, ou seja, será feita uma breve explicação sobre o Modelo Padrão, o CERN, o experimento LHC e o detector ATLAS.

2.1 MODELO PADRÃO

O Modelo Padrão das partículas elementares, não é exatamente um modelo, mas sim uma teoria muito bem fundamentada, considerada a melhor teoria sobre a natureza da matéria por muitos físicos (MOREIRA, 2009).

De acordo com o Modelo Padrão, todas as partículas podem ser classificadas em Bósons e Férmions, sendo que os primeiros não obedecem o Princípio de Exclusão de Pauli, que é um princípio da mecânica quântica, que afirma que dois férmions idênticos não podem ocupar o mesmo estado quântico simultaneamente. Este modelo também descreve os mecanismos de interação regidos pelas forças: eletromagnética, fraca e forte;

a única força não abrangida por esta teoria é a força gravitacional. (PERKINS, 2000)

As partículas constituintes da matéria podem ser divididas e nomeadas como segue:

- Férmions: partículas que constituem a matéria e são subdivididas em léptons e quarks.
- Léptons: elétron, múon, tau e seus neutrinos e suas antipartículas.
- Quarks são: up, down, charm, strange, top e bottom e suas antipartículas.

As partículas transportadoras de força que mediam as interações entre partículas e são: glúon(força forte), fóton(força eletromagnética), bósons W e Z(força fraca) e o bóson de Higgs(responsável pela existência de massa inercial).

Um resumo das informações das partículas do Modelo Padrão é apresentado na Figura 1.

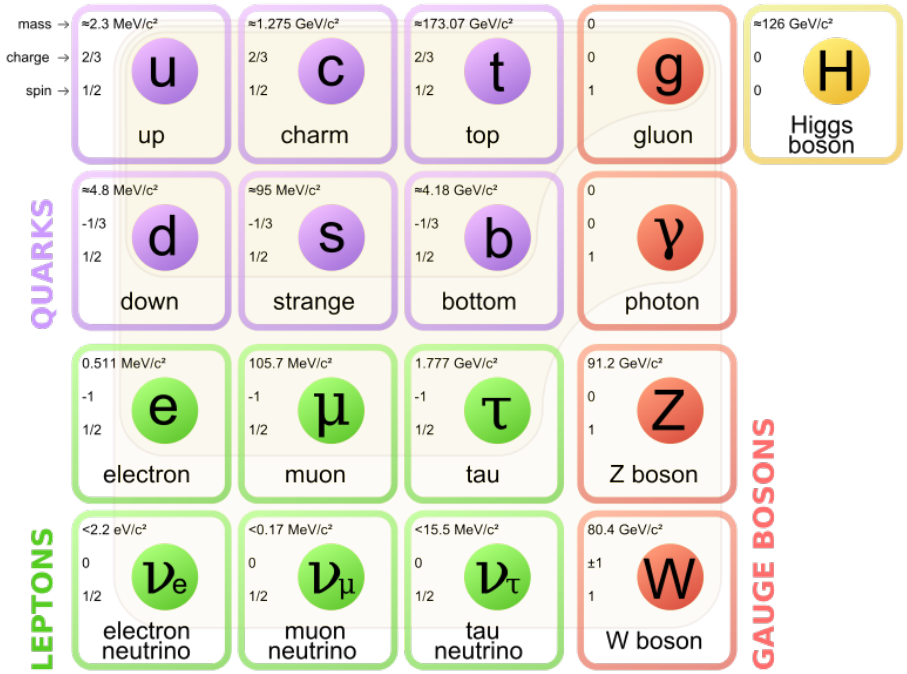

Figura 1: Modelo Padrão. Extraído de (RONQUI, 2015).

2.2 CERN

No CERN, físicos e engenheiros trabalham em conjunto com o objetivo de investigar a estrutura fundamental do universo. Fundada em 1954, o laboratório CERN foi construído na fronteira franco-suíça, em Genebra. Ele foi um dos primeiros empreendimentos conjuntos da Europa e tem agora 21 Estados membros (CERN, 2015a).

O principal foco desta organização é a física de partículas que abrange estudos como: composição, raios cósmicos, matéria escura, dimensões extras, grávitons, minúsculos buracos negros, íons pesados, plasma quark-glúon, entre outros (CERN, 2015c).

A necessidade de comprovar as teorias e estudar as partículas de maneira mais profunda tornou a construção do acelerador de partículas LHC imprescindível. Nele, feixes de prótons são acelerados em direções opostas até atingirem altas energias e colidirem uns com os outros. A Seção 2.3 abordará melhor a estrutura deste aparato.

Com o intuito de 'ler' e armazenar as informações geradas nas colisões dentro do acelerador, faz-se necessária a utilização de detectores, no local exato das colisões entre os feixes. Atualmente. o LHC conta com alguns detectores como: *A Large Ion Collider Experiment* (ALICE), *A Toroidal LHC Apparatus* (ATLAS), *Compact Muon Solenoid* (CMS), *Large Hadron Collider beauty* (LHCb). A Figura 2 mostra a localização dos detectores (ALICE, ATLAS, CMS e LHCb) no LHC.

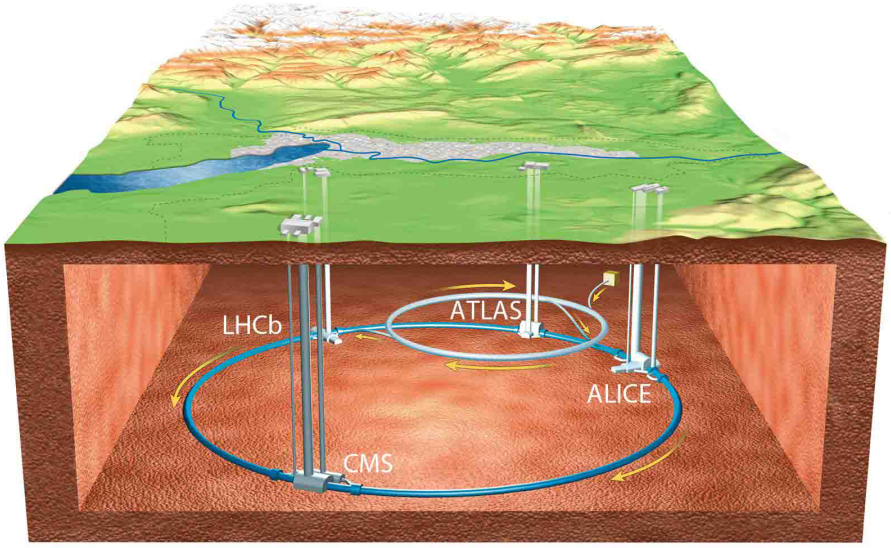

Figura 2: Uma visão geral do experimento LHC. Extraído de CERN (CERN, 2015a).

Na Seção 2.4, o detector de ATLAS será melhor detalhado, uma vez que os dados utilizados nessa dissertação foram gerados por esse detector.

2.3 LHC

O LHC„ Figura 3, tem cerca de 27 km de circunferência. Ele acelera prótons ou íons que viajam em direções opostas, e são colocados para colidir (LEFEVRE, 2009).

Um acelerador só pode acelerar certos tipos de partícula: em primeiro lugar, esses

elementos precisam ter carga, uma vez que os feixes são manipulados por dispositivos eletromagnéticos que só podem influenciar as partículas carregadas; e, em segundo lugar, exceto em casos especiais, estas não podem decair. Isso limita o número de partículas que podem ser acelerados para elétrons, pósitrons, prótons e íons. É necessário acrescentar que em um acelerador circular, como o LHC, partículas pesadas, como prótons, têm uma perda de energia, através de radiação síncrotron, muito menor que partículas leves, como elétrons. Portanto, para obter colisões com energias muito elevadas, o LHC faz uso de prótons.

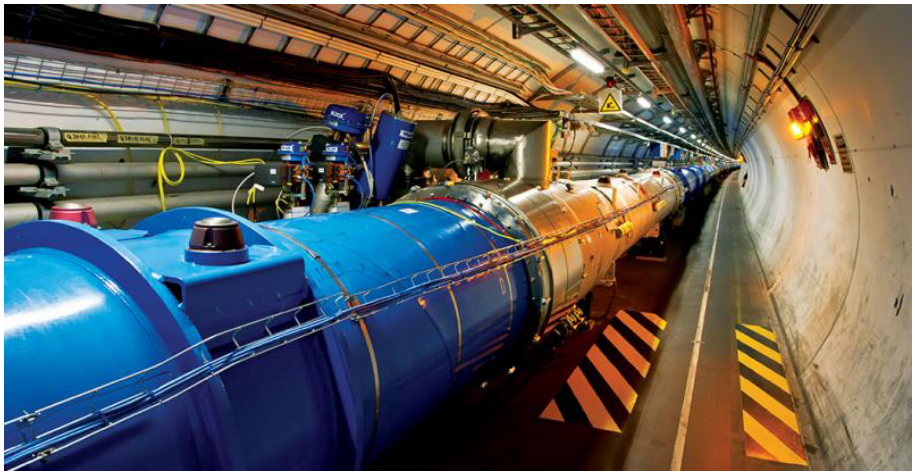

Figura 3: O LHC é o maior e mais poderoso acelerador de partículas do mundo. Extraído de (CERN, 2015b).

O complexo de aceleradores do CERN é uma sucessão de mecanismos que aceleram partículas a energias cada vez maiores. Cada um desses instrumentos aumenta a energia de um feixe de partículas, antes de injetar o feixe no LHC, propriamente dito.

Com a ajuda de um campo eletromagnético, os prótons dos átomos de hidrogênio são separados dos elétrons. Estes prótons, primeiramente, são acelerados a energia de 50 MeV pelo Acelerador Linear (LINAC 2). Esse feixe é então injetado no *Proton Synchrotron Booster* (PSB), que o leva a energia de 1,4 GeV, seguido pelo *Proton Synchrotron* (PS), que o impulsiona a 25 GeV. Esses prótons são enviados para o *Super Proton Synchrotron* (SPS), onde eles são acelerados para 450 GeV. No LHC, o último elemento nesta cadeia, feixes de partículas são aceleradas até a energia recorde de 7 TeV por feixe, nominal de operação, com que colidem (CERN, 2016a).

No ano de 2015, depois de passar por manutenção e modificações, o LHC retomou as atividades preparado para atingir energia de colisão no centro de massa de 13 TeV, aproximadamente o dobro da energia que vinha trabalhando. Este *upgrade* tem o intuito de alcançar novos resultados e descobertas, uma vez que funcionará a uma

energia nunca antes alcançada.

2.4 ATLAS

Com 46 metros de comprimento, 25 metros de altura e 25 metros de largura, e 7 mil toneladas, o detector ATLAS, mostrado na Figura 4, é o maior detector de partículas já construído. Ele se situa em uma caverna a 100 metros de profundidade perto do prédio principal do CERN, como pode ser visto na Figura 5, próximo da cidade de Meyrin, na Suíça (CERN, 2016b).

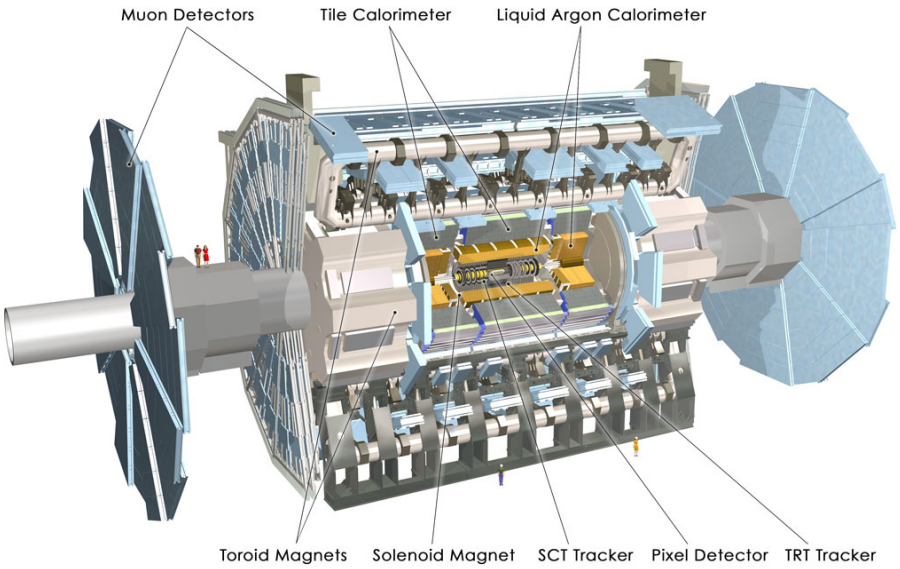

Figura 4: Modelo computacional do Detector ATLAS. Extraído de (www.atlas.ch).

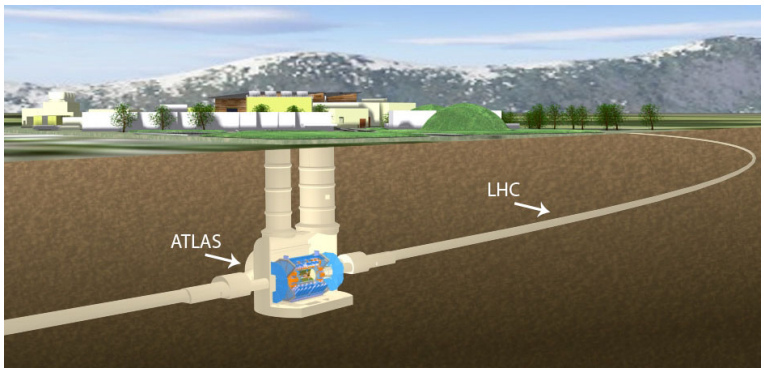

Figura 5: Vista subterrânea do detector ATLAS . Extraído de (www.atlas.ch).

No centro do detector ATLAS, feixes de partículas do LHC colidem gerando produtos da colisão sob a forma de novas partículas, que se espalham em todas as direções. Como este é um detector de uso geral, precisa ser capaz de identificar os mais diversos

tipos de partículas. O detector contém seis subsistemas de detecção diferentes dispostos em camadas ao redor do ponto de colisão, no intuito de gravar as trajetórias, momentos e energias das partículas, permitindo, assim, que sua identificação individual seja possível.

Esta seção tem o intuito de prover uma visão geral deste detector, suas característica básicas, detalhes internos e funcionamento, visando um melhor entendimento do ambiente onde está inserida essa dissertação.

2.4.1 SISTEMAS DE COORDENADAS

O detector ATLAS tem formato cilíndrico e, para a identificação da posição das partículas no detector, utiliza-se um sistema de coordenadas pre-estabelecido. Este sistema define o ponto de interação nominal como a origem do sistema de coordenadas; o eixo z é definido pela direção do feixe e o plano xy é transversal à direção do feixe. Alternativamente, usando coordenadas cilíndricas, o ângulo ϕ é medido, como de costume, em torno do eixo do feixe, e o ângulo polar θ é o ângulo a partir do eixo do feixe (AAD et al., 2008). Por fim, a *pseudorapidez* $\eta = -\ln\tan\left(\frac{\theta}{2}\right)$, sendo η e ϕ as principais coordenadas do ATLAS, como mostrado na Figura 6.

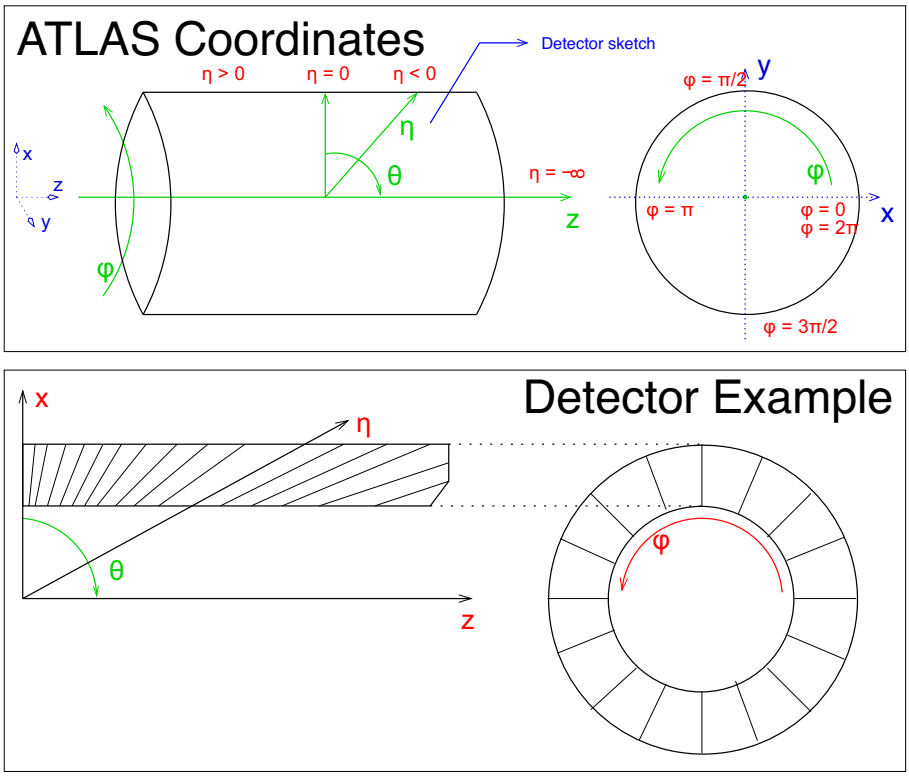

Figura 6: Sistema de coordenadas do Detector ATLAS. Extraído de (ANJOS, 2006).

2.4.2 PERFIL DOS EVENTOS DO ATLAS

O perfil dos eventos de interesse do ATLAS e suas particularidades ditaram as características de construção e modulação do detector, uma vez que a identificação dessas partículas é feita pelas características de sua assinatura, que é uma marca particular, deixada no aparato. Cada componente deste equipamento foi especificado para detectar um conjunto de propriedades das partículas (LIPPMANN, 2012).

Com o conhecimento sobre as peculiaridades de cada parte do detector, podemos entender como é o perfil de alguns desses eventos de interesse, mostrados na Figura 7 e traduzidos abaixo:

- Detector de traço: múons, prótons, elétrons, píons e káons;
- Calorímetro eletromagnético: múons, elétrons, fótons, prótons, píons e káons;
- Calorímetro hadrônico: múons, prótons, nêutrons, píons e káons.
- Câmara de múons: Múons;

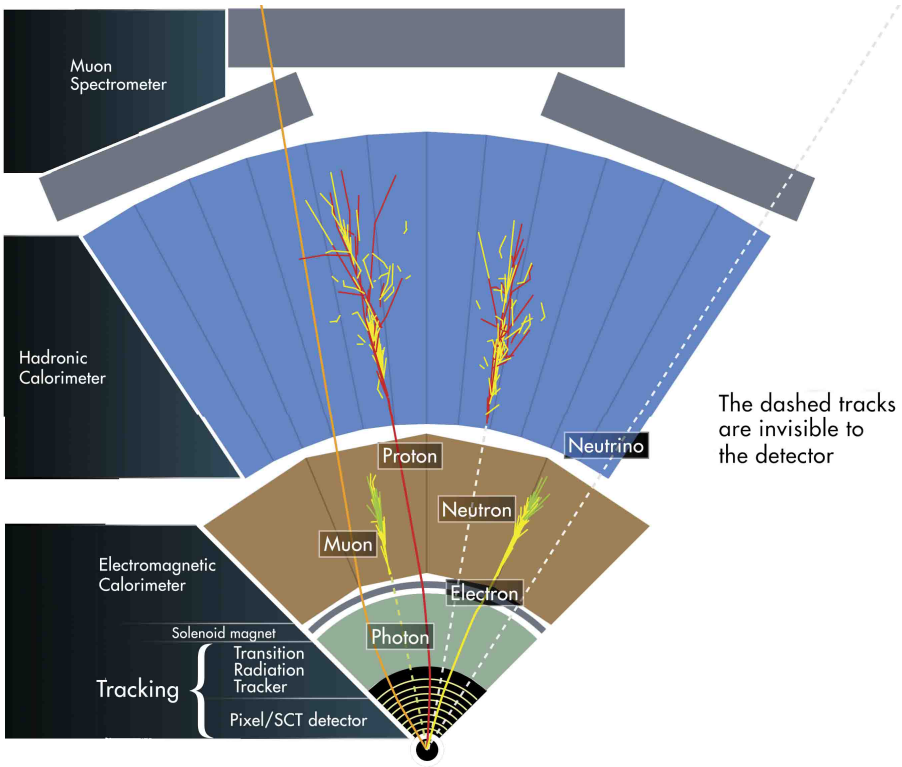

Figura 7: Modelo computacional da assinatura das partículas no detector ATLAS. Extraído de (cds.cern.ch).

2.4.3 DETECTOR INTERNO

O Detector Interno (ID) do ATLAS é composto de três partes: uma seção cilíndrica, chamada de Barril(*Barrel*), que cobre a região central ($|\eta| \leq 1$), e duas regiões em forma de disco, chamadas de tampas(*Endcaps*), abrangendo as regiões ($1 \leq |\eta| < 2.5$).

Os subprodutos das colisões primeiramente cruzam o tubo de feixe, em seguida, as três camadas do Detector de Pixels (*Silicon Pixel Detector* (SPD)), quatro camadas do Detector de Traços baseado em semicondutores (*SemiConductor Tracker* (SCT) e 36 tubos do Detector de Radiação de Transição (*Transition Radiation Tracker* (TRT) (BARBERIS, 2000).

- Detector de Pixels (SPD): Esse detector fornece medidas em duas dimensões com alta precisão perto do ponto de interação, que são especialmente importantes para a caracterização de partículas de decaimentos semi-leptônicos;
- Detector de Traços baseado em semicondutores (SCT): Faz 4 pares de medidas por traço e, combinado com o SPD, provê medidas de momento, parâmetro de impacto e posição de vértice;
- Detector de Radiação de Transição (TRT): Esse detector contribui de maneira significativa para a medida precisa do momento de todos os traços, bem como, proporciona uma capacidade inerente de identificação de elétron. (BENEKOS et al., 2003).

A Figura 8 mostra a disposição dos detectores e a Figura 9 mostra um corte transversal do ID.

2.4.4 CALORÍMETROS

O calorímetro é um dispositivo que absorve toda a energia cinética de uma partícula, que ao colidir com seu material inicia um chuveiro de partículas, cuja interação fornece, ao fim da cadeia, um sinal eletrônico proporcional ao valor da energia depositada (DAS; FERBEL; GLASHAUSSER, 1994).

Na interação com o calorímetro cria-se um processo em cascata, onde partículas secundárias são produzidas ao longo do detector. Uma fração dessa energia é entregue na forma de luz de cintilação que produz um sinal detectável.

As características básicas dos calorímetros são:

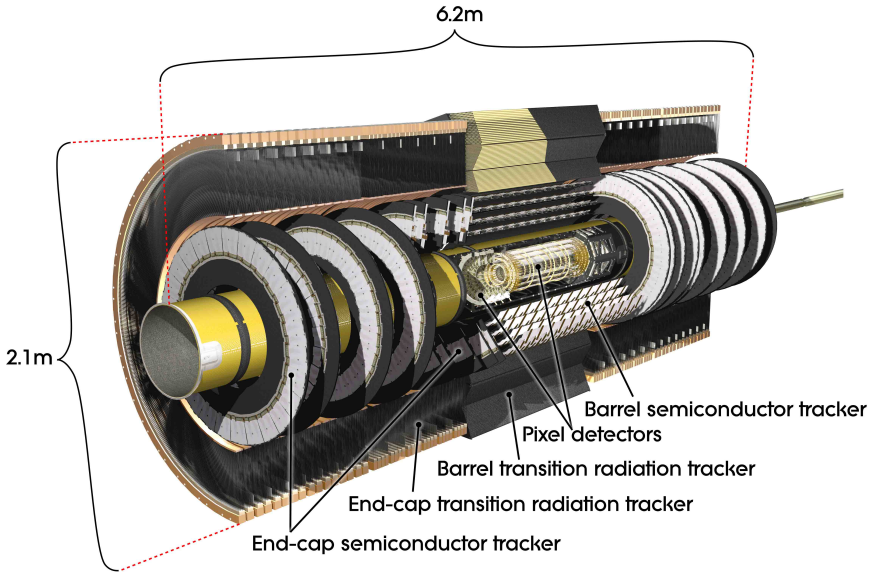

Figura 8: Modelo computacional do ID. Extraído de (cds.cern.ch).

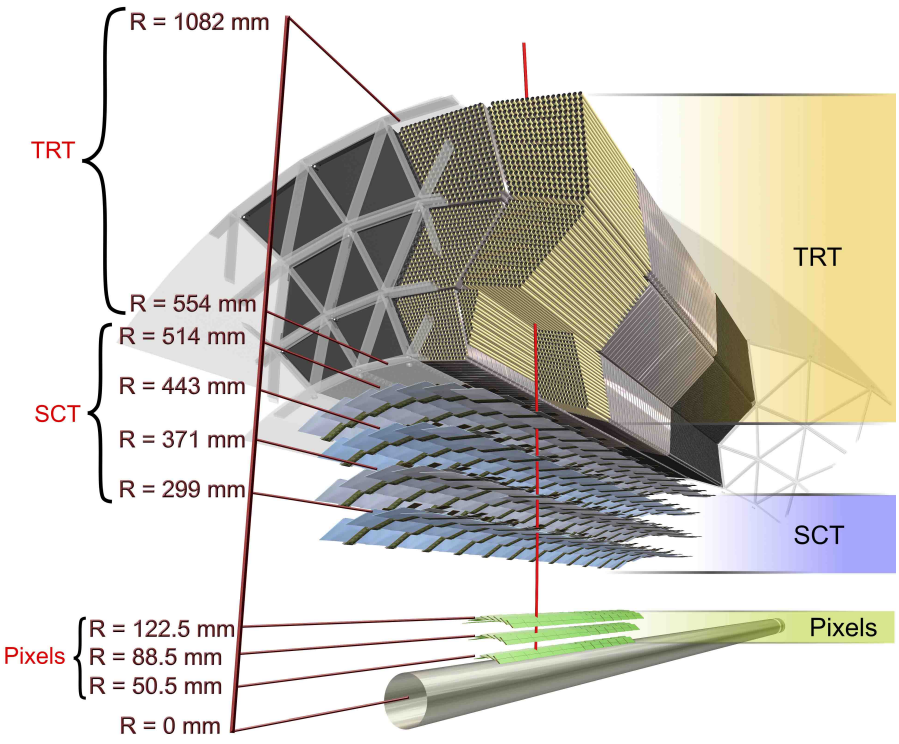

Figura 9: Modelo computacional do ID - corte transversal. Extraído de (cds.cern.ch).

- Calorímetros podem ser sensíveis tanto a partículas neutras quanto a carregadas;
- Pode ser utilizado para identificação de partículas, uma vez que há diferenças na forma de deposição de energia para elétrons, múons e hádrons, por exemplo;
- Permite tanto medida da energia quanto de trajetória das partículas, devido à

sua segmentação;

- Tempo de resposta rápido (menor que 50 ns), adequando-se a um ambiente com alta taxa de eventos (PERALVA, 2012).

2.4.4.1 CHUVEIROS ELETROMAGNÉTICOS E HADRÔNICOS

Em física de altas energias podemos destacar dois tipos de chuveiros (ou cascatas) (GRUPEN; SHWARTZ, 2008):

- Chuveiros eletromagnéticos (Figura 10): são iniciados por elétrons ou fótons com alta energia ao passarem pelo calorímetro. Essas partículas carregadas sofrem interações criando fótons que, por sua vez, se convertem em pares elétron-pósitron. Essa cascata aumenta até a energia dos elétrons ser menor que uma energia crítica;
- Chuveiros hadrônicos (Figura 11): decorrem do comprimento de interação nuclear, e geralmente são muito maiores que os chuveiros eletromagnéticos. Seu desenvolvimento lateral é causado pela grande transferência de momento típica de interações nucleares; e são basicamente compostos por píons.

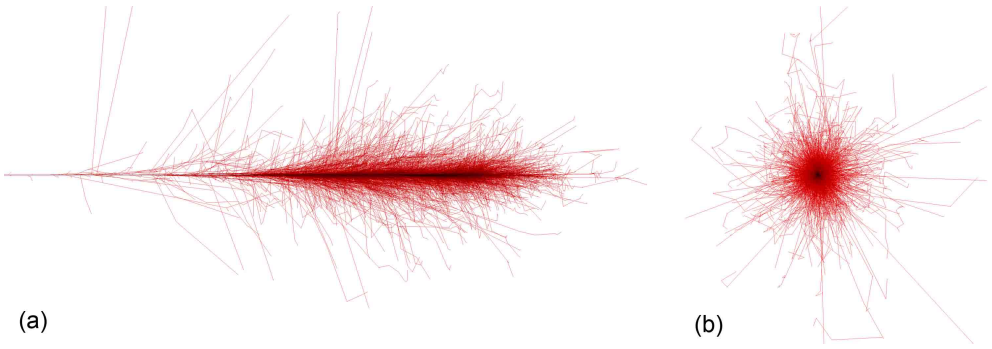

Figura 10: Simulação computacional utilizando algoritmo Corsika do Chuveiro Eletromagnético (100GeV), (a) vista lateral e (b) vista frontal.

2.4.4.2 CALORÍMETRO ELETROMAGNÉTICO

O Calorímetro Eletromagnético (EM) (CALORIMETER et al., 2008) é composto de absorvedores de chumbo e eletrodos intercalados em forma de acordeão, sendo utilizado Argônio líquido como material ativo. Esse aparato compõe a parte mais interna no sistema de calorimetria do ATLAS.

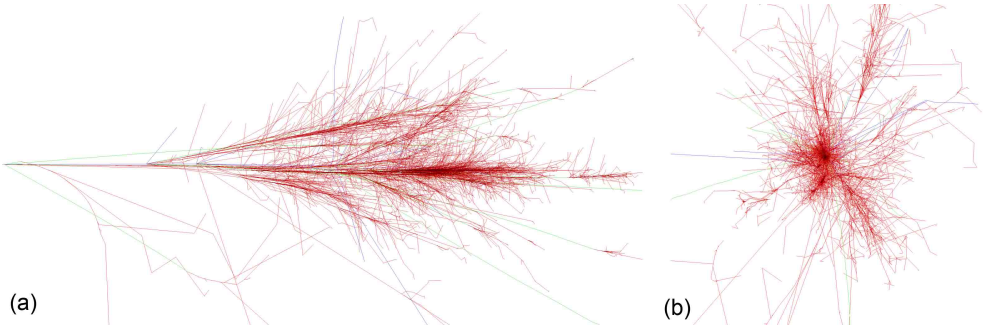

Figura 11: Simulação computacional utilizando algoritmo Corsika do Chuveiro Hadrônico (100GeV), (a) vista lateral e (b) vista frontal.

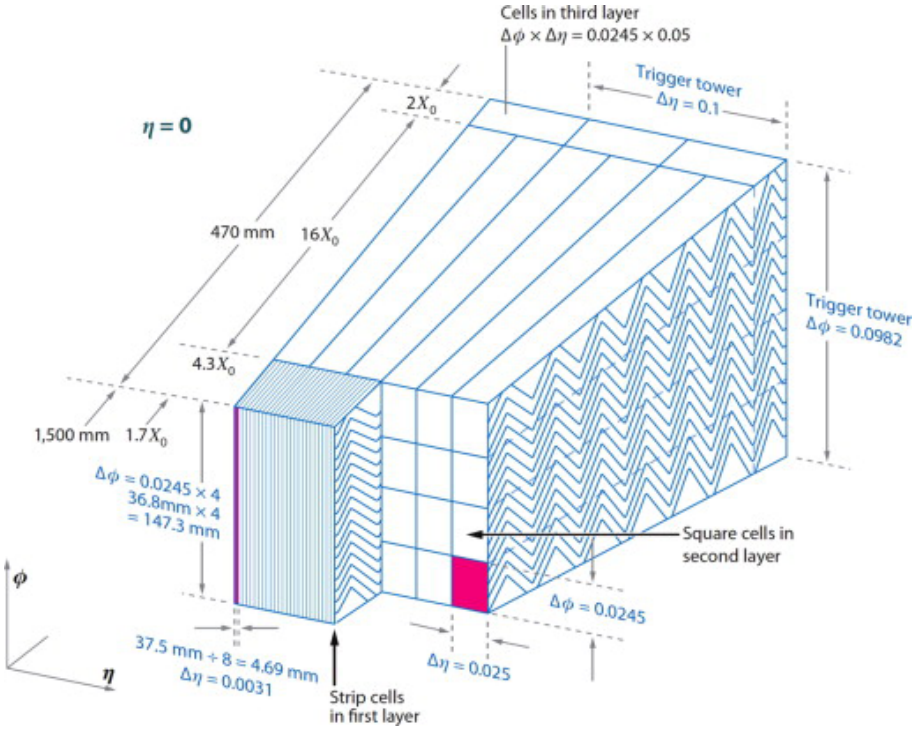

Figura 12: Modelo computacional do Calorímetro Eletromagnético. Extraído de (FRANCAVILLA; COLLABORATION et al., 2012).

A construção desse calorímetro foi dividida em três camadas, sendo a primeira mais segmentada, no intuito de efetuar uma localização precisa da partícula; a segunda é mais profunda e menos segmentada que a primeira; e, por fim, a terceira camada é a menos segmentada e tem a função de absorver completamente a energia da partícula incidente. Essas divisões e segmentações podem ser vistas na Figura 12.

Como existem perdas de informação devido ao 'material morto' (fios, encapamentos, etc) o calorímetro EM possui um pré-irradiador, que atua na recuperação dessas informações.

2.4.4.3 CALORÍMETRO HADRÔNICO

O Calorímetro Hadrônico (HAD) do detector ATLAS, chamado de Calorímetro de Telhas (do inglês *Tile Calorimeter*, ou *TileCal*), utiliza placas cintiladoras, em formato de telha, como material ativo e, como material absorvedor, faz o uso de placas de aço com baixo carbono. Esse equipamento é subdivido em 3 partes, como pode ser visto na Figura 13: o barril *(Tile Barrel)*, situado no região central e dois barris estendidos *(Tile Extended Barrel)*, um em cada lateral do barril. O Tilecal também possui 3 camadas, cada uma segmentada de uma forma diferente (AAD et al., 2010).

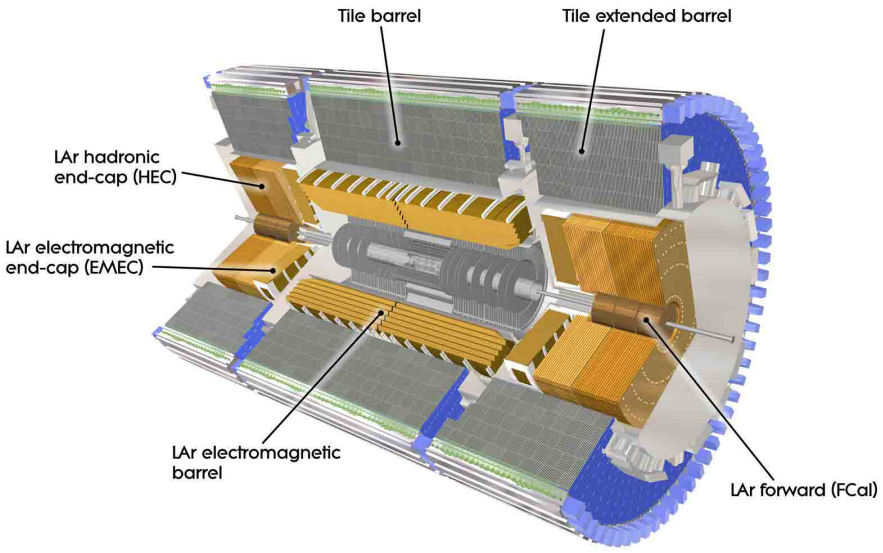

Figura 13: Modelo computacional do HAD e do EM. Extraído de (cds.cern.ch)

Quando partículas 'excitam' as telhas, este material cintilador produz luz, que é transmitida por fibras ópticas até Tubos Fotomultiplicadores (TFM). Este instrumento converte a luz em sinal elétrico, que é lido pela eletrônica do TileCal.

2.4.5 O DETECTOR DE MÚONS

A camada mais externa do detector ATLAS é a câmara de múons. Idealmente, essas partículas são as únicas, detectáveis, capazes de atravessar os calorímetros. O espectrômetro de múons (COLLABORATION et al., 2010) circunda o calorímetro e mede as trajetórias dessas partículas, sendo, assim, capaz de medir o seu momento junto com o ID. Esses traços sempre são normais a componente principal do campo eletromagnético, o que torna o resolução do momento transverso rudemente independente de η.

O sistema de detecção de múons é constituído por milhares de sensores de partículas carregadas, colocados em um campo magnético, produzido por grandes bobinas toroidais supercondutoras. Na Figura 14, é mostrado o modelo computacional do Detector de Múons.

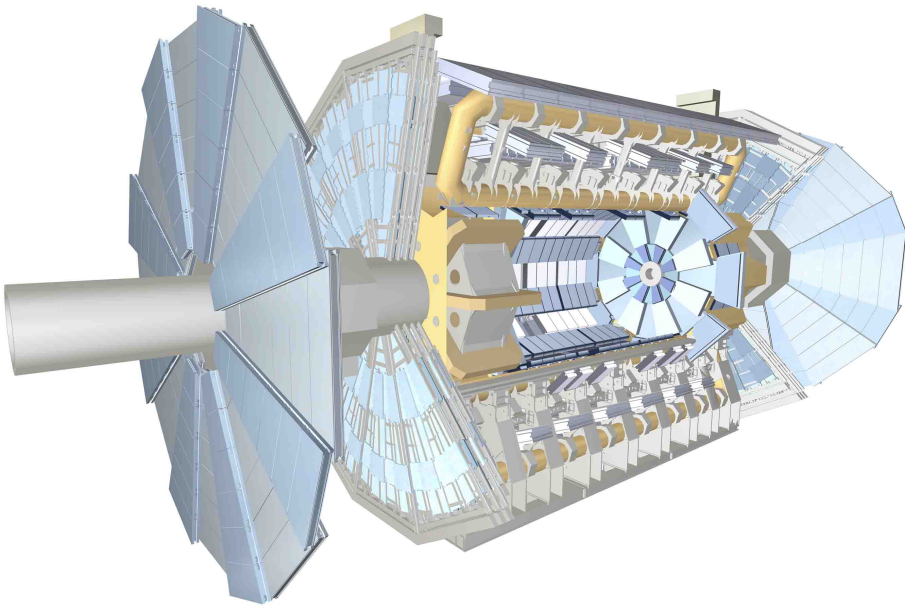

Figura 14: Modelo computacional da Câmara de Múons do detector ATLAS. Extraído de (cds.cern.ch).

2.4.6 SISTEMA DE FILTRAGEM DO ATLAS

No intuito de alcançar novas descobertas, utilizando-se de eventos raros, o experimento efetua colisões com uma taxa muito alta de operação, gerando um conjunto muito grande de informações, onde grande parte desses eventos pode ser descartada, a fim de evitar o armazenamento de dados não relevantes ou mesmo já bem explorados (WATTS, 2003). Nesse contexto, faz-se necessário um Sistema de Filtragem *Online* que seja capaz de separar os eventos considerados importantes e armazená-los para um análise posterior mediada por algoritmos mais complexos e criteriosos.

Como a Figura 15 apresenta, o sistema de filtragem *online* foi desenvolvido em 3 níveis consecutivos que, juntos, reduzem a taxa de eventos de 40 MHz para 200 Hz (ELSING; SCHÖRNER-SADENIUS, 2003). Esses níveis são o *Level 1* (L1), *Level 2* (L2) e Filtro de Eventos (EF), respectivamente.

- O primeiro nível de *trigger* é feito em hardware, uma vez que precisa trabalhar com uma latência de $\sim 2\mu$s. Esse sistema recebe os sinais dos calorímetros e

da câmara de múons, separando os conjuntos de eventos que ficaram dentro do limiar de corte estabelecido, chamados de Regiões de Interesse, (do inglês,*Region of Interest*) (RoI), reduzindo a taxa de eventos para $75kHz$ (GABALDON, 2012);

- O segundo nível de filtragem tem sua implementação baseada em softwares operando em uma rede de computadores e sua principal característica é observar as RoI pré-definidas pelo L1. Este tem a capacidade de reduzir para $\sim$2 kHz a taxa de eventos;

- Já o último nível de filtragem do ATLAS reduz essa taxa para 200 Hz, trabalhando com uma granularidade maior, re-selecionando as informações transmitidas pelo L2.

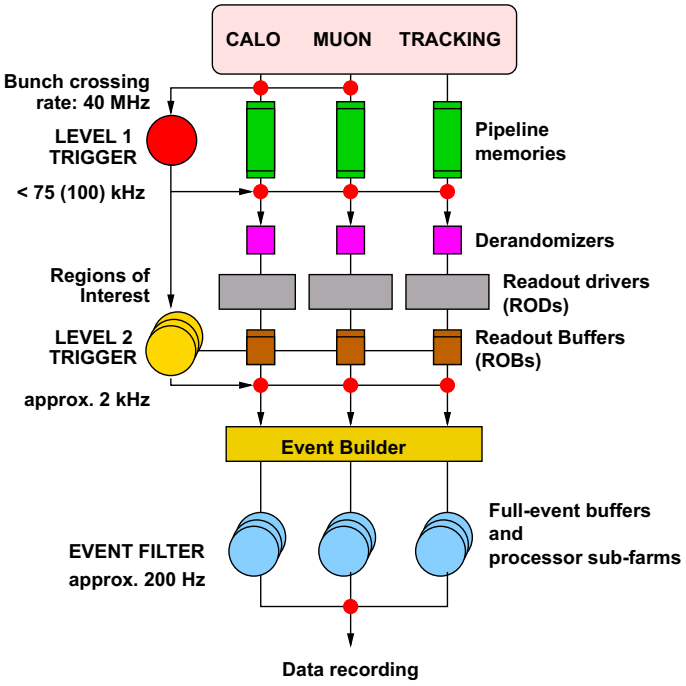

Figura 15: Fluxograma do sistema de Trigger Online do ATLAS. Extraído de (ANJOS, 2006).

Como foi dito na Seção 2.4, o ATLAS é um detector de uso geral; portanto, os dados armazenados são utilizados em diferentes estudos, que são realizados por filtragem *offline*. Uma vez que este sistema não tem como fator determinante o tempo de processamento, algoritmos mais bem elaborados e específicos para cada tipo de estudo podem ser empregados, possibilitando, assim, uma identificação mais robusta das partículas. O ambiente *offline* é também propício para implementação e testes de novos algoritmos que podem, eventualmente, ser aplicados futuramente no sistema de filtragem *online* do ATLAS.

3 IDENTIFICAÇÃO DE ELÉTRONS

Um bom desempenho na reconstrução e identificação de elétrons é um ingrediente fundamental para o sucesso do programa científico do experimento ATLAS, uma vez que as principais assinaturas dos processos eletrofracos são os léptons (ALISON, 2014) e são utilizados para inúmeras análises, como, por exemplo, as medidas de precisão do modelo padrão, a descoberta do bóson de *Higgs* e a busca por uma nova física além do Modelo Padrão (AAD et al., 2014).

Os elétrons isolados produzidos em muitos do processos físicos de interesse estão sujeitos a uma grande quantidade de ruídos de fundo provenientes de:

- Hádrons identificados equivocadamente;
- Fótons convertidos;
- Electrons não-isolados originados de decaimentos *heavy-flavour*.

Por essa razão, é de primordial importância alcançar uma eficiente identificação de elétrons, sobre todo o detector e, ao mesmo tempo, manter uma grande rejeição de ruído de fundo.

3.1 RECONSTRUÇÃO DE ELÉTRONS

O algoritmo que efetua a reconstrução dos elétrons da região central ($|\eta| < 2,5$) do detector ATLAS identifica as energias depositadas no calorímetro EM e as associa aos traços do ID (AAD et al., 2014), seguindo os três passos abaixo descritos:

1. Reconstrução do *Cluster*: o conjunto de células (do inglês *cluster seed*) do EM provém das energias depositadas que contêm um total de energia transversa superior a 2,5 GeV, através de um algoritmo de janela móvel, com janela de tamanho 3x5 em unidades de 0,025 x 0,025 em η x ϕ;

2. Combinar traço com conjunto de células: um traço e uma célula podem ser ditos combinados se a distância entre o ponto de impacto do traço e o baricentro da célula for $\Delta\eta < 0{,}05$. E o tamanho de $\Delta\phi$ necessariamente precisa estar dentro de uma janela de 0.1;

3. Candidato a elétron reconstruído: depois de combinar traço-célula, o tamanho do *cluster* é otimizado para $\Delta\eta$ x $\Delta\phi$ = 3x7 (5x5) barril (tampa). O total da energia do candidato a elétron reconstruído é determinado pela soma de 4 fatores (ABAT et al., 2008):

    - A energia estimada depositada na parte frontal do calorímetro EM;
    - A energia depositada no *cluster*;
    - A energia depositada fora do *cluster* (também chamadas de perdas laterais);
    - A energia depositada atrás do *cluster* (perdas longitudinais).

A eficiência da reconstrução para os elétrons que passam pelo procedimento acima descrito é alta. Nesse estágio, da-se o nome de "*reconstructed electrons* "aos candidatos que foram aprovados nos requisitos de *cluster* e traço.

3.1.1 TRIGGER *DE ELÉTRONS*

O sistema de *Trigger* do ATLAS, como já mencionado na Seção 2.4.6, é constituído de três níveis, sendo que o L2 e EF juntos compõem a Filtragem de Alto nível, (do inglês, *High Level Trigger*) (HLT). No primeiro nível, selecionam-se somente os elétrons que ultrapassem um limiar de energia e, devido à dependência em η, esse limiar sofre variações (ALISON, 2014).

O HLT utiliza as RoI preestabelecidas pelo L1; entretanto, um limiar mais refinado pode ser aplicado, bem como o uso de variáveis discriminantes, que serão apresentadas na Seção 3.2.

Os pontos de operação do *trigger* são definidos em três categorias:

- *Trigger* Primário: critérios rígidos são aplicados para coletar eventos de sinal em análise usando elétrons;
- *Trigger* de Suporte: coletam amostras de elétrons não polarizados, utilizando basicamente E_t como critério;

- *Trigger* para Monitoramento e Calibração: utilizados para coletar dados no intuito de garantir o correto funcionamento do *trigger* e do detector (AAD et al., 2012).

3.2 VARIÁVEIS DISCRIMINANTES PARA IDENTIFICAÇÃO DE ELÉTRONS

Tanto nas análises *online* quanto *offline*, critérios adicionais são aplicados no intuito de garantir uma melhor pureza dos elétrons reconstruídos. Estes critérios são informações retiradas dos calorímetros e do detector interno (COLLABORATION et al., 2011) a partir de um conjunto de variáveis discriminantes, que podem ser divididas em:

- Variáveis de Calorimetria;
- Variáveis de Traço;
- Variáveis de Traço-Calorimetria;
- Variáveis de Isolamento.

3.2.1 VARIÁVEIS DE CALORIMETRIA

As variáveis de calorimetria utilizam a fina segmentação lateral e longitudinal dos calorímetros do detector ATLAS:

- Variável de vazamento hadrônico, R_{had_1}:

  Definida como a razão entre as energias transversas da primeira camada do calorímetro hadrônico e do *cluster*. Elétrons reais depositam mais energia no EM do que no HAD, apresentando assim valores pequenos de R_{had_1};

- Variável de largura em η na segunda camada, $W_{\eta 2}$:

  É a medida da largura do chuveiro em η ponderada pelo Raiz do Valor Médio Quadrático, (do inglês, *Root Mean Square*) (RMS) da distribuição em η na segunda camada do EM. Esta variável contribui para suprimir ruído de fundo de jatos e conversões de fótons, que tendem a ter chuveiros maiores do que elétrons verdadeiros;

- Variável de largura do chuveiro, R_η:

É definida como a razão entre energia de uma janela 3x7 sobre uma janela 7x7, na segunda camada de amostragem. Ruídos de fundo tendem a ter uma maior fração de energia fora do núcleo 3x7, resultando em baixos valores de R_η.

- Variável de largura do chuveiro, R_ϕ:

  Semelhante à variável R_η; entretanto, definida como a razão entre a energia em uma janela 3x3 sobre uma janela 3x7.

- Variável de largura do chuveiro na primeira camada do calorímetro, w_{stot}:

  Mede a largura do chuveiro, que ajuda na identificação de elétrons porque apresenta maiores valores para ruído de fundo.

- Variável de razão de energia, E_{ratio}.

  Também é utilizada para diminuir o ruído de fundo. É definida utilizando as células correspondentes às duas maiores energia nas camadas. Ruídos de fundo tendem a ter múltiplas incidências de partículas associadas, apresentando assim valores de E_{ratio} menores do que partículas de sinal.

- Fração de energia da terceira camada do EM, f_3:

  Essa variável tende a ser menor para elétrons do que para ruído de fundo, uma vez que os elétrons não penetram tão profundamente no calorímetro.

- Fração de energia nas camadas do EM, f_1:

  Essa variável é definida como a razão de energia depositada nas camadas sobre a energia total do EM.

A Figura 16 mostra algumas das distribuições das variáveis acima apresentadas, bem como os vários tipos de ruídos de fundo.

Essas variáveis são dependentes de η e E_t. Em η devida à geometria dos calorímetros, que apresentam alguns pontos com menor resolução, como, por exemplo, a região, definida como *crack*, que se encontra entre o barril e a tampa, $1{,}37 < |\eta| < 1{,}52$, e que muitas vezes, é excluída de análises por conta da baixa resolução. Por outro lado, o poder de discriminação dessas variáveis melhora com o aumento de E_t, uma vez que a largura do chuveiro tende a diminuir e o ruído de fundo tende a ter uma menor dependência de E_t.

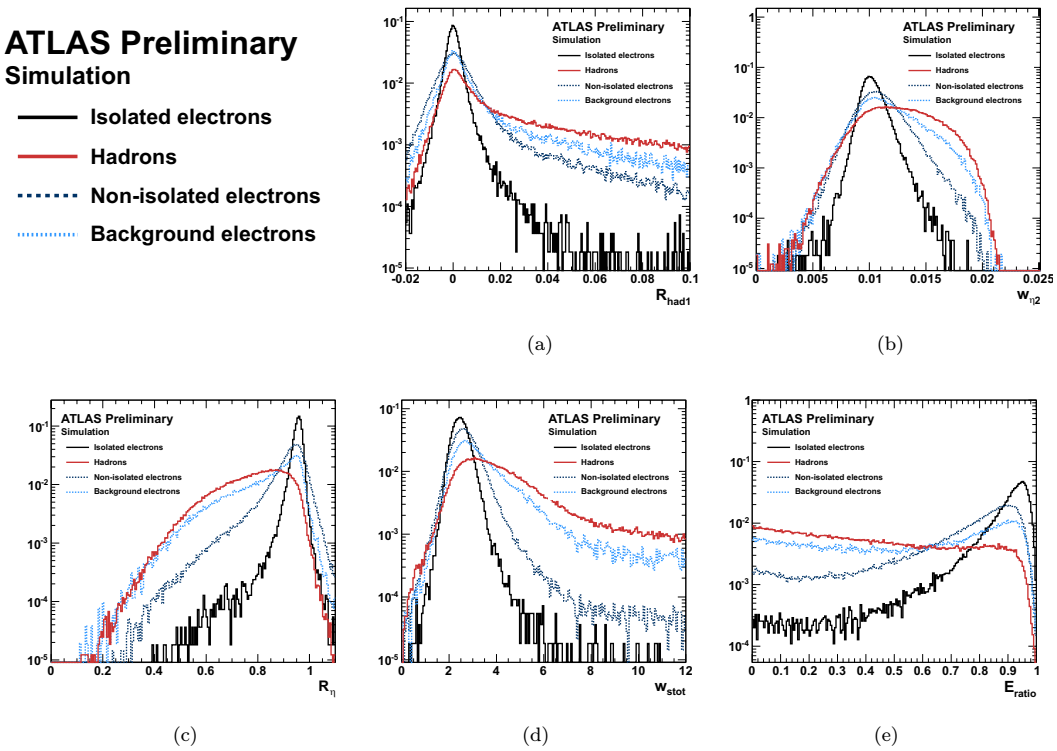

Figura 16: Variáveis de identificação de elétrons no calorímetro, formato do chuveiro, apresentados separadamente para sinal e os vários tipos de ruídos de fundo. As variáveis apresentadas são: (a) vazamento hadrônico R_{had}, (b) de largura em eta no segundo W_2 amostragem, (c) R_{η}, (d) largura em η nas $w_{s,tot}$, pequeno, e (e) E_{ratio}. Extraído de (ALISON, 2014).

3.2.2 VARIÁVEIS DE TRAÇO

As variáveis de traço são provenientes do ID e podem ser utilizadas de forma a complementar as do calorímetro.

- Número de *hits* no detector Pixel (nPixHits) e Número combinado de *hits* do Pixel e detectores de SCT (nSixHits):

  As camadas de detectores que são atravessadas por fótons antes de serem convertidos não têm traços associados a eles. Isso resulta em um menor número de *hits* no detector de Pixels e SCT, do que os elétrons verdadeiros.

- Número de *hits* na primeira camada do Detector de Pixels ou *B-layer*:

  Essa variável apresenta uma sensibilidade a todas as conversões que ocorrem depois da primeira camada do Detector de Pixels, sendo bastante efetiva na redução de ruído de fundo.

- Parâmetro de impacto transverso, D_0:

Mede a distância mais próxima do traço do elétron até o vértice primário e possibilita a separação de conversões, dado que estes podem ter traços deslocados significativamente dos pontos de interação.

- Significância do parâmetro de impacto transverso, σ_{d_0}:

  Mede a relevância da distância mais próxima do traço do elétron até o vértice primário.

- *Flag* de conversão, ou "bit conversão":

  É definido se o traço do elétron corresponde a um vértice de conversão. Reduz o número de elétrons reconstruídos de conversões, entretanto não é tão eficiente para elétrons verdadeiros.

- Fração de *hits* de alto *threshold* no TRT:

  Essa variável é uma das mais poderosas contra ruído de fundo provenientes de hádrons, em razão de mostrar a fração das detecção que passaram o limiar do detector TRT, indicando a presença de radiação de transição de fótons.

Na Figura 17 são apresentadas algumas das variáveis de traço. Diferente das variáveis de calorimetria, as variáveis de traço são independentes de η e E_t, com exceção do TRT, e são pouco dependentes do *pileup*.

3.2.3 VARIÁVEIS DE COMBINAÇÃO TRAÇO-CALORIMETRIA

Ao combinar as informações de traço e calorimetria, tem-se variáveis adicionais para discriminação de ruído de fundo. Essas são apresentadas na Figura 18.

- Variável de diferença entre o traço e o *cluster* de energia em η, $\Delta\eta_1$:

  A comparação é feita extrapolando o traço até o calorímetro EM e esta distribuição é mais reduzida para os elétrons reais, portanto, a exigência de valores pequenos de $\Delta\eta$ reduz o ruído de fundo.

- Variável de diferença entre o traço e o *cluster* de energia em ϕ, $\Delta\phi_2$:

  Semelhante a variável descrita anterior, entretanto, menos discriminante devido aos fótons da radiação de Bremsstrahlung causarem uma diferença entre a posição do traço e o cluster em ϕ.

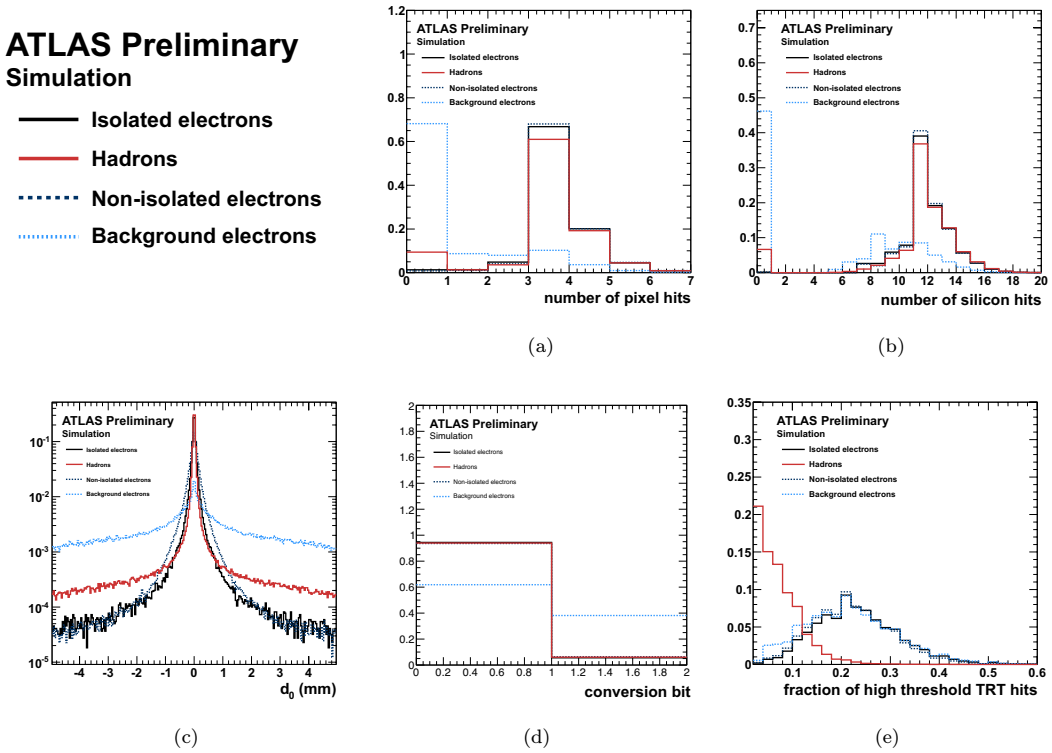

Figura 17: Variáveis de identificação elétron no ID, agrupados em sinal e vários tipos de ruídos de fundo. As variáveis apresentadas são: (a) número de *hits* no detector Pixel, (b) número combinado de *hits* do Pixel e detectores de SCT, (c) parâmetro de impacto transverso D_0, (d) *flag* de conversão, ou "bit conversão", e (e) fração *hits* de alto *threshold* no TRT. Extraído de (ALISON, 2014).

- Variável de diferença entre o traço e o *cluster* de energia em ϕ, reescalada, $\Delta\phi_{res}$:

  É a variável $\Delta\phi_2$, porém com o momento de traço reescalado para a energia do *cluster* depois da extrapolação para a camada central.

- Variável de razão entre a energia medida no calorímetro pelo momento determinado no ID, E/P:

  Como os hádrons não depositarão toda sua energia no EM, uma fração será depositada no HAD. A exigência de que E/P seja consistente com a expectativa de um elétron real pode suprimir tanto hádrons e quanto conversões.

3.2.4 VARIÁVEIS DE ISOLAMENTO

Por último, as variáveis de isolamento também são utilizadas para discriminar sinal e ruído de fundo, são elas:

- Et_{cone};

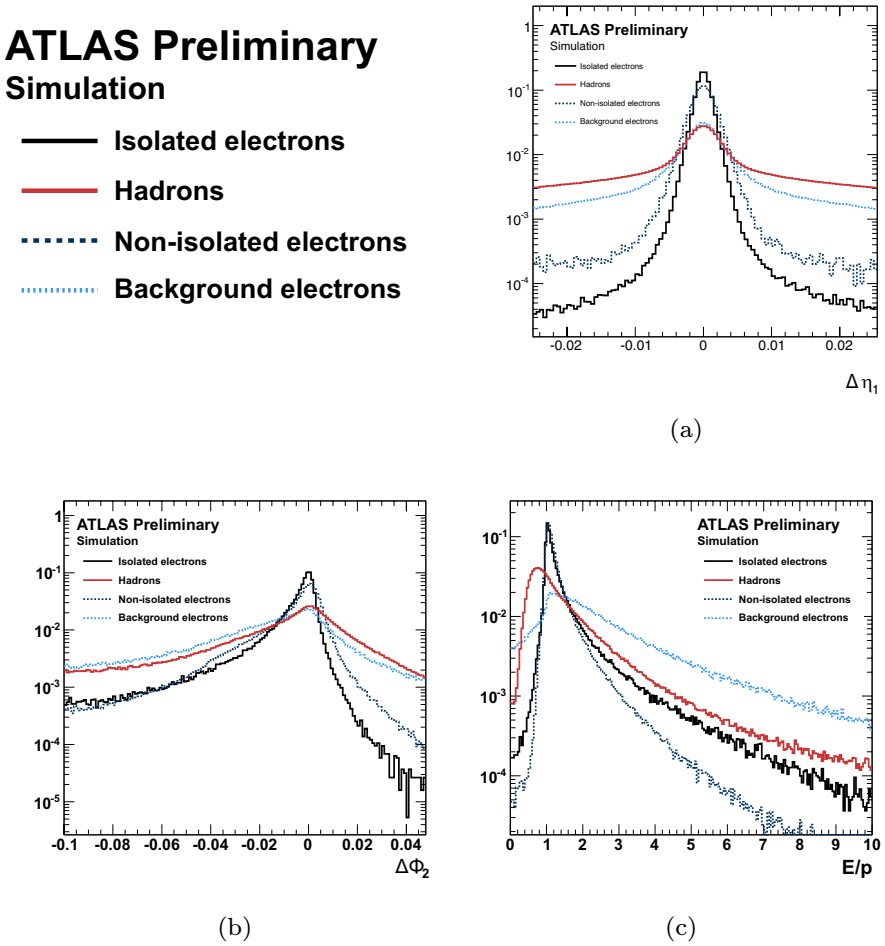

Figura 18: Variáveis combinadas de traço-calorimetria, mostrando a separação de vários tipos de background. As variáveis mostradas são: (a) diferença entre o traço e o *cluster* de energia em η, (b) diferença entre o traço e o *cluster* de energia em ϕ, e (c) razão da energia medida no calorimetro com o momento medido no traço. Extraído de (ALISON, 2014).

- Pt_{cone}.

O isolamento é medido pela quantidade de energia próxima do elétron reconstruído, uma vez que elétrons de ruído de fundo são produzidos juntamente com outras partículas, o que os leva a ter maiores valores nessas variáveis. Dessa forma, essas variáveis conseguem ajudar na identificação de sinal e ruído de fundo

3.3 ALGORITMOS OFFLINE DE REFERÊNCIA PARA A IDENTIFICAÇÃO DE ELÉTRONS

A colaboração ATLAS utiliza alguns algoritmos *offline* para identificação de elétrons. Nessa seção, serão apresentados dois algoritmos, o e/γ, que é o padrão da

colaboração e o algoritmo baseado em Verossimilhança, (do inglês, *Likelihood*), que é a metodologia utilizada nessa dissertação.

3.3.1 ATLAS E/γ

O algoritmo e/γ de seleção e identificação de elétrons, tem como princípio o corte baseado nas variáveis discriminantes. A utilização deste método como padrão traz a vantagem de compartilhamento e cruzamento de análises entre as diversas linhas de pesquisa no ATLAS (AAD et al., 2012).

Como o intuito desta ferramenta é de ser compatível com o maior número possível de pesquisas físicas, três pontos de operação são disponibilizados (ALISON, 2014):

Loose Nível de detecção de sinal elevado; entretanto, com a pior rejeição de ruído de fundo entre os três;

Tight Melhor rejeição de ruído de fundo, por conseguinte, menor nível de eficiência entre os três.

Medium Apresenta o ponto de equilíbrio entre os dois primeiros, com nível de rejeição de ruído melhor que o *Loose* e eficiência melhor do que o *Tight*.

Esses pontos de operação são configurados de forma que o *Loose* seja um subconjunto de *Medium* que é um subconjunto do *Tight*, como mostra a Tabela 1; entretanto, com valores de cortes um pouco diferentes.

O perfil dos pontos de operação do e/γ, em 2011, ganhou uma versão mais atualizada, e seus pontos de operação são chamados: *Loose++*, *Medium++* e *Tight++*.

3.3.2 VEROSSIMILHANÇA

Dentre as técnicas multivariadas existentes, a verossimilhança apresenta a vantagem de uma construção simples, no caso de independência entre as variáveis.

O método de verossimilhança faz uso de Função de Densidade de Probabilidade (PDF) das variáveis discriminantes de sinais e de ruído de fundo para encontrar a probabilidade total. No algoritmo de verossimilhança do ATLAS, as PDF foram feitas por uma ferramenta da colaboração chamada Toolkit for Multivariate Data Analysis *(TMVA) adaptative KDE*, que utiliza um método não-paramétrico chamado Estimação

Tabela 1: 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

de Densidade de Núcleo, (do inglês, *Kernel Density Estimation)* (KDE) (THERHAAG; TEAM, 2012).

3.3.2.1 O MÉTODO DA VEROSSIMILHANÇA

O método de verossimilhança é uma função dos parâmetros de um modelo estatístico que permite inferir sobre o seu valor a partir de um conjunto de observações. No caso de identificação de elétrons do ATLAS, esse conjunto de observações são as variáveis discriminantes. Primeiramente, são construídas as PDF a partir dos dados de sinal e ruído de fundo. Com a consideração de independência entre as variáveis, as probabilidades conjuntas de sinal e ruído de fundo podem ser calculadas através de uma simplificação do método de verossimilhança, como mostram as equações 3.1 e 3.2 (COLLABORATION et al., 2014).

$$L_s\left(x\right) = \prod_{a=1}^{m} P_{s,a}(x_a) \tag{3.1}$$

$$L_b\left(x\right) = \prod_{a=1}^{m} P_{b,a}(x_a) \tag{3.2}$$

onde $P_{s,a}(x_a)$ e $P_{b,a}(x_a)$ são as probabilidades associadas a cada uma das m variáveis (x_i) do evento analisado. L_s e L_b são os valores da multiplicação da verossimilhança para sinal e ruído de fundo.

Com as duas probabilidades conjuntas calculadas, faz-se o discriminante, utilizando a equação 3.3, onde dL é o discriminante.

$$dL = \frac{Ls}{Ls + Lb} \tag{3.3}$$

Construir o menu para a verossimilhança consiste em: escolher as variáveis, selecionar cortes adicionais e definir o valor de limiar do discriminante (COLLABORATION et al., 2013), sendo que a eficiência da verossimilhança será o resultado da eficiência do discriminante e os cortes adicionais combinados.

Uma das principais diferenças entre o algoritmo e/γ e a verossimilhança está nos eventos de cauda da PDF, uma vez que o primeiro efetua cortes rígidos, impossibilitando assim a classificação destes (COLLABORATION et al., 2013).

3.3.2.2 VEROSSIMILHANÇA PARA ELÉTRONS NO ATLAS

O método de verossimilhança para identificação de elétrons apresenta cinco pontos de operação: *Very Tight, Tight, Medium, Loose, Very Loose*, cada um com diferentes níveis de rejeição de ruído e eficiência de sinal, e estes pontos diferem entre si pelas variáveis e limiar dos cortes adicionais, como mostra a Tabela 2.

Tabela 2: Variáveis usadas na construção da verossimilhança para diferentes pontos de operação. Extraído de (COLLABORATION et al., 2013).

Menu	VERY TIGHT, TIGHT	MEDIUM	LOOSE, (VERY LOOSE)
Variáveis da Verossimilhança	R_{Had} R_{η} F_{HT} $\Delta\eta_1$ $W_{\eta 2}$ f_1 f_3 E_{ratio} R_{ϕ} $\Delta p/p$ $\Delta\phi_{Res}$ d_0 σ_{d_0}	R_{Had} R_{η} F_{HT} $\Delta\eta_1$ $W_{\eta 2}$ f_1 f_3 E_{ratio} R_{ϕ} $\Delta p/p$ $\Delta\phi_{Res}$ d_0 σ_{d_0}	R_{Had} R_{η} F_{HT} $\Delta\eta_1$ $W_{\eta 2}$ f_1 f_3 E_{ratio} R_{ϕ} $\Delta p/p$ $\Delta\phi_{Res}$
Cortes Adicionais	nSiHits ≥ 7 nPixHits ≥ 2 Blayer !(isConv)	nSiHits ≥ 7 nPixHits ≥ 2 Blayer	nSiHits ≥ 7 nPixHits ≥ 2 (≥ 1) Blayer (no Blayer)
Comparar com	isTightPlusPlus	MediumPlusPlus	isLoosePlusPlus Multilepton

4 ESTIMAÇÃO DE DENSIDADE POR NÚCLEO

O KDE é uma técnica não-paramétrica para estimação de densidades, onde cada observação é ponderada pela distância em relação a um valor central, o núcleo. Após a estimação da densidade de um conjunto de dados, tem-se conhecimento sobre a probabilidade de ocorrência de cada evento, o que possibilita análises sobre essas observações e, por exemplo, a aplicação deste método em problemas de classificação (LEDL, 2002).

Nas próximas seções, os métodos univariado e multivariado serão descritos, bem como os critérios de otimização dos mesmos.

4.1 ESTIMAÇÃO DE DENSIDADE UNIVARIADA

4.1.1 HISTOGRAMA PARA KDE

Uma das maneiras mais simples de abordar esse problema, muito utilizada na literatura, é a estimação por histograma, mostrada na equação 4.1:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(x_i \in B_j)\, I(x \in B_j) \tag{4.1}$$

Onde, n é o número de eventos, i e j são os subíndices de eventos, I é a função indicadora e,

$$\begin{aligned} &B_j = [x_o + (j-1)\,h, x_o + jh) \\ &j \in \mathrm{Z} \end{aligned} \tag{4.2}$$

Esse método lida com os parâmetros de origem (x_o) e tamanho do *bin* (h), sendo que ambos, se escolhidos de maneira errada, podem resultar em uma má estimação. No intuito de diminuir esse problema, é possível efetuar a estimação pela média de sub-histogramas, tornando o método independente da origem x_o (SEATHER, 1992), como mostra a equação 4.3.

$$\hat{f}_h(x) = \frac{1}{M}\sum_{l=0}^{M-1}\frac{1}{nh}\sum_{i=1}^{n}\sum_{j} I\left(x_i \in B_j\right) I\left(x \in B_j\right) \tag{4.3}$$

Onde,

$$\begin{aligned} &B_{j,l} = \left[\left(j-1+\tfrac{1}{M}\right)h, \left(j+\tfrac{1}{M}\right)h\right) \\ &l \in \{0,1,...,M-1\} \end{aligned} \tag{4.4}$$

Essa configuração produz uma estimação mais suave que o histograma, mas é somente uma maneira melhor de utilizar esse conceito básico, podendo não resultar em uma boa representação da densidade. Entretanto, o problema de otimização do parâmetro h continua e sua otimização foi considerada em vários trabalhos, como por exemplo: (CHEN; MEER, 2002), (ABRAMSON, 1982), (COMANICIU; RAMESH; MEER, 2001), (SILVERMAN, 1986) e (JONES, 1991).

4.1.2 MODELO DO ESTIMADOR DE DENSIDADE POR NÚCLEO

O KDE, de forma direta, aplica uma função de *Kernel* em cada evento, levando em conta a contagem de eventos em sua vizinhança. O estimador é dado pela equação 4.5 e foi Rosenblatt em 1956 o primeiro autor que considerou esse modelo.

$$\hat{f}(x_i) = \frac{1}{nh}\sum_{k=1}^{N} K\left(\frac{(x_i - X_k)}{h}\right) \tag{4.5}$$

onde *K(u)* é a função *Kernel* e o parâmetro h é conhecido como largura de banda.

Para garantir que esse modelo retorne uma função de densidade, o *Kernel* deve satisfazer $\int_{-\infty}^{+\infty} K(u)\,du = 1$ e $K(u) \geq 0 \quad \forall \quad u \in \Re$ (Funções Kernels de ordem maiores que 2 não satisfazem a última propriedade).

Algumas outras propriedades da função Kernel são:

- $K(u) = K(-u)$;
- $K(u)$ tem seu máximo quando $u = 0$;
- Os momentos do *kernel* são: $k_j(k) = \int_{-\infty}^{+\infty} u^j K(u)\,du$;
- A ordem do *kernel* v é definida como a ordem do primeiro momento não nulo.

Alguns exemplos de funções *kernel* de segunda ordem são mostradas abaixo, sendo que a função Gaussiana foi a escolhida para esse trabalho:

- 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: $K(u) = \frac{1}{\sqrt{2\pi}}e^{\left(-\frac{1}{2}u\right)}$

Em ambas as abordagens, histogramação e KDE, fica clara a dependência da estimação em relação ao parâmetro h de largura de banda. Nas próximas seções, serão discutidos os critérios para otimização da estimação de densidade e, consequentemente, da escolha ótima do parâmetro h.

4.1.3 CRITÉRIO DE OTIMIZAÇÃO

No que tange a otimização de estimadores de densidade, não existe uma maneira geral de otimização; toda otimização é feita olhando de maneira particular para um certo critério (LEDL, 2002). Essa seção tem o intuito de mostrar alguns métodos de otimização de estimadores e suas propriedades.

4.1.3.1 CRITÉRIO BASEADO NA DISTÂNCIA L_1

Um ideia bastante comum para medir a diferença entre duas funções quaisquer f e g é utilizando o método L_p (LEDL, 2002), definido como:

$$L_p = \left(\int |f - g|^p \right)^{1/p} \tag{4.6}$$

Onde p é o parâmetro a ser escolhido. No caso mais simples, $p = 1$, a equação 4.6 se torna a equação 4.7.

$$L_1 = \int |f - g| \tag{4.7}$$

A equação 4.7 é chamado de *Integrated Absolute Error* (IAE) e deve ser minimizada, no intuito de garantir uma boa estimação da densidade.

4.1.3.2 CRITÉRIO BASEADO NA DISTÂNCIA L_2

Quando $p = 2$ é escolhido na equação 4.6, temos:

$$ISE\left(\hat{f}_h\right) = \int \left[\hat{f}_h\left(x\right) - f\left(x\right)\right]^2 dx \tag{4.8}$$

A equação 4.8 é então chamada de *Integrated Squared Error* (ISE), sendo assim, para o cálculo do *Mean Integrated Squared Error* (MISE), utiliza-se o valor esperado de $f\left(x\right)$, equação 4.9. Em (JONES, 1991), o autor faz uma comparação do uso desses dois métodos e conclui que encontrar o valor otimizado da largura de banda utilizando o MISE apresenta melhores resultados que por ISE.

$$MISE\left(\hat{f}_h\right) = \int E\left\{\hat{f}_h\left(x\right) - f\left(x\right)\right\}^2 dx \tag{4.9}$$

Do mesmo modo que o *Mean Squared Error* (MSE), que é a medida do erro médio de um certo ponto x, a expressão do MISE pode ser decomposta em termos de *bias* e variância. A partir dessa decomposição, existem na literatura muitos estudos sobre a minimização dos critério MISE, alguns deles serão abordados na Seção 4.1.4.

4.1.3.3 CRITÉRIO BASEADO NA DISTÂNCIA L_∞

De acordo com a equação 4.6, qualquer valor de p pode ser escolhido. Entretanto, para escolhas de p maiores que dois, não existem propriedades muito úteis, no sentido de otimização de estimadores. Sendo assim, para $p > 2$, o único caso que pode ser interessante é quando $p \to \infty$, que é igual a minimizar o erro máximo absoluto entre duas funções, f e g por exemplo (LEDL, 2002).

Todas as medidas descritas são estritamente critérios matemáticos, que fazem medidas, de modos diferentes, de similaridade entre duas funções. Entretanto, essas medidas podem não levar ao tipo de otimização desejada para estimadores de densidade. Como exemplo, o critério L_∞, que ignora as caudas da distribuição, sendo que essas tornam-se mais importantes à medida em que se aumenta o número de dimensões.

4.1.4 CÁLCULO DOS CRITÉRIOS DE ERRO

4.1.4.1 MISE E AMISE

Como introduzido na Seção 4.1.3.2, o MSE pode ser dividido em dois termos, como mostrado na equação 4.10:

$$\begin{aligned} MSE\left(\hat{f}_h\right) &= E\Big\{\hat{f}_h(x) - f(x)\Big\}^2 \\ &= Bias\Big(\hat{f}_h(x)\Big)^2 + \operatorname{var}(f(x)) \end{aligned} \tag{4.10}$$

ESTIMAÇÃO DO *BIAS* Sabendo que o valor esperado da transformação por *kernel* pode ser escrita como:

$$E\left\{\hat{f}(x)\right\} = \int \frac{1}{h} K\left(\frac{(z-x)}{h}\right) f(z)\, dz \tag{4.11}$$

usando uma mudança de variáveis, $u = (z-x)/h$, temos:

$$E\left\{\hat{f}(x)\right\} = \int K(u) f(x+hu) du \tag{4.12}$$

A equação 4.12 mostra que o valor esperado é a média de $f(z)$ localmente sobre x. Essa integral não pode ser resolvida analiticamente, então utiliza-se uma aproximação, via expansão de Taylor, no termo $f(x+hu)$, que é válida quando $h \to 0$. Para uma função *kernel* de segunda ordem, a expansão toma a forma da equação 4.13:

$$\mathrm{f(x+hu)} = \mathrm{f}(x) + f'(x)\, hu + \frac{1}{2} f''(x)\, h^2 u^2 + o(h^2) \tag{4.13}$$

Integrando termo por termo da equação 4.13 e sabendo que $\int_{-\infty}^{+\infty} K(u)\, du = 1$, temos:

$$\int K(u) f(x+hu) du = \mathrm{f}(x) + \frac{1}{2} f''(x)\, h^2 \mu_2(K) + o(h^2) \tag{4.14}$$

onde, $\mu_2(K) = \int z^2 K(z)\, dz$ representa a variância do *kernel* de segunda ordem.

Isso significa que, para *kernels* de segunda ordem, podemos calcular o *Bias* como sendo (HANSEN, 2009):

$$\begin{aligned} Bias\left(\hat{f}\left(x\right)\right) &= E\hat{f}\left(x\right) - f\left(x\right) \\ &= \tfrac{1}{2}f''\left(x\right)h^2\mu_2(K) + o(h^2) \end{aligned} \tag{4.15}$$

ESTIMAÇÃO DA VARIÂNCIA Desde que o estimador kernel seja linear, podemos calcular a variância como mostrado na equação 4.16:

$$\begin{aligned} \text{var}\left(\hat{f}\left(x\right)\right) &= \tfrac{1}{nh^2}\text{var}\left(K\left(\tfrac{x_i - X}{h}\right)\right) \\ &= \tfrac{1}{nh^2}E\left\{K\left(\tfrac{x_i - X}{h}\right)^2\right\} - \tfrac{1}{n}\left(\tfrac{1}{h}E\left\{K\left(\tfrac{x_i - X}{h}\right)\right\}\right)^2 \end{aligned} \tag{4.16}$$

Utilizando a seguinte aproximação, proveniente da análise de *bias*, $\frac{1}{h}E\left\{K\left(\frac{(x_i - X)}{h}\right)\right\} = f\left(x\right) + o\left(1\right)$, temos que o segundo termo é igual a $o\left(\frac{1}{nh}\right)$. Para o primeiro termo da equação 4.16, faz-se mudança de variável, análoga aquela feita no cálculo do *bias*, e expansão de Taylor de primeira ordem, assim sendo, temos:

$$\begin{aligned} \tfrac{1}{h}E\left\{K\left(\tfrac{(x_i - X)}{h}\right)\right\}^2 &= \int K(u)^2\left(f\left(x\right) + o\left(h\right)\right) \\ &= f\left(x\right)R\left(K\right) + o\left(h\right) \end{aligned} \tag{4.17}$$

onde $R\left(K\right) = \int K(u)^2 du$ é a rugosidade de uma função (SCOTT, 2015).

Com isso, podemos calcular a variância como sendo:

$$\text{var}\left(\hat{f}\left(x\right)\right) = \frac{f\left(x\right)R\left(K\right)}{nh} + o\left(\frac{1}{n}\right) \tag{4.18}$$

Uma vez definidos os termos de *bias* e variância, podemos reescrever a equação 4.10, substituindo nela as Equações 4.15 e 4.18. Assim temos:

$$\begin{aligned} MSE\left(\hat{f}_h\right) &= E\left\{\hat{f}_h\left(x\right) - f\left(x\right)\right\}^2 \\ &= Bias\left(\hat{f}_h\left(x\right)\right)^2 + \text{var}\left(\hat{f}\left(x\right)\right) \\ &\simeq \left(\tfrac{1}{2}f''\left(x\right)h^2\mu_2(K)\right)^2 + \tfrac{f(x)R(K)}{nh} + o\left(\tfrac{1}{nh} + h^4\right) \end{aligned} \tag{4.19}$$

Na equação 4.19, o último termo é devido ao erro de truncamento da expansão de Taylor. Como sugerido em (HANSEN, 2009), faz-se uma aproximação assintótica no MSE, desconsiderando esse termo, que recebe o nome de *Asymptotic Mean Squared Error* (AMSE) . Quando se trata de distribuições e não de pontos, como visto na Seção 4.1.3.2, a medida de erro é chamada de MISE, e sua aproximação assintótica *Asymptotic Mean Integrated Squared Error* (AMISE), dada pela equação 4.20.

$$\begin{aligned} AMISE &= \int AMSE\left(\hat{f}\left(x\right)\right)dx \\ &= \tfrac{1}{4}h^4\mu_2(K)^2R\left(f''\right) + \tfrac{1}{nh}R\left(K\right) \end{aligned} \tag{4.20}$$

Temos então que o primeiro termo (*bias*) aumenta proporcionalmente ao parâmetro h, enquanto que o segundo termo (variância) diminui de maneira inversamente proporcional ao aumento de h. Com isso, vemos claramente um conflito entre reduzir a variância e o *bias* de forma simultânea, visto que a escolha de um h pequeno para garantir um menor bias ocasiona uma variância grande (LEDL, 2002).

Na equação 4.20, fazendo a primeira derivada igual a zero, pode-se calcular a largura de banda ótima pela equação 4.21.

$$h_{AMISE} = \left[\frac{R\left(K\right)}{\mu_2(K)^2R\left(f''\right)n}\right]^{1/5} \tag{4.21}$$

Como a equação 4.21 mostra, o problema de calcular a largura de banda ótima via AMISE, com o conhecimento somente da distribuição, é circular; portanto, para escolher esse parâmetro de forma automática, pode-se assumir uma distribuição normal com variância σ^2 (SILVERMAN, 1986). Com isso, o parâmetro de largura de banda fixa h que minimiza o AMISE é:

$$h = 1,06\sigma n^{-1/5} \tag{4.22}$$

4.1.4.2 LARGURA DE BANDA VARIÁVEL

A largura de banda h, como foi abordado, é um parâmetro crucial, e pode ser utilizada de duas formas: fixa (Seção 4.1.4.1) ou variável. Esta segunda abordagem possibilita a utilização de uma largura de banda $h(x_i)$ para cada ponto de x_i em que desejamos estimar a probabilidade $f_{h_i}\left(x_i\right)$. Esse estimador é conhecido como *balloon estimator* e tem a forma:

$$\hat{f}_{hi}\left(x_i\right) = \frac{1}{nh(x_i)}\sum_{k=1}^{n} K\left(\frac{(x_i - X_k)}{h(x_i)}\right) \tag{4.23}$$

Para calcular o parâmetro de largura de banda variável $h(x_i)$, utiliza-se a equação 4.24, conforme sugerido por (ABRAMSON, 1982).

$$h\left(x_i\right) = \frac{h}{\sqrt{f_p(x_i)}} \tag{4.24}$$

onde h é uma largura de banda fixa e $f_p(x_i)$ é a probabilidade de x_i na PDF.

Como o h fixo otimizado foi definido na Seção 4.1.4.1, necessita-se agora de otimizar a escolha do parâmetro $f_p\left(x_i\right)$. Nesse trabalho, fez-se um estudo sobre métodos de validação cruzada como: *Least-Square Cross-Validation* (LSCV) e *Biased Cross-Validation* (BCV), que utilizam os critérios de ISE e MISE, respectivamente. Entretanto, a escolha foi feita baseada no algoritmo proposto por (SHIMAZAKI; SHINOMOTO, 2007) de estimação da binagem ótima para um histograma, esse algoritmo aborda o problema fazendo uso de uma função custo que minimiza o critério MISE. Por sua vez, essa escolha não mostrou um bom resultado devido às variedades nas formas das distribuições analisadas nessa dissertação. Para contornar este problema, foi necessário inserir outro parâmetro, no intuito de tornar o KDE mais robusto. O novo parâmetro λ, chamado de constante de proporcionalidade foi proposto por (COMANICIU; RAMESH; MEER, 2001) e é incorporado à equação da banda variável da seguinte forma:

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

sendo λ dado por:

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

4.2 ESTIMAÇÃO DE DENSIDADE MULTIVARIADA

O conceito de KDE univariado precisa ser ampliado quando precisa-se investigar a dependência entre as variáveis do problema, ou quando trata-se de um problema de classificação. Tendo como base a teoria univariada, uma generalização do KDE é mostrada na equação 4.27:

$$\begin{aligned} &f_{h_1,h_2,\ldots,h_v}\left(x_{1,2,\ldots,v}\right) = \\ &= \tfrac{1}{n}\sum_{k=1}^{N}\tfrac{1}{h_1}\tfrac{1}{h_2}\ldots\tfrac{1}{h_v}K\left(\tfrac{(x_1-X_{k_1})}{h_1}\right)K\left(\tfrac{(x_2-X_{k_2})}{h_2}\right)\ldots K\left(\tfrac{(x_v-X_{k_v})}{h_v}\right) \end{aligned} \tag{4.27}$$

O raciocínio é análogo, e, em nosso algoritmo, utilizamos a forma generalizada matricial. Primeiro modelamos a largura de banda como:

$$\mathrm{H} = \mathrm{diag}(\mathrm{h}_1, \mathrm{h}_2, ..., \mathrm{h}_v) \tag{4.28}$$

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

$$h_j = \left(\frac{4}{d+2}\right)^{\frac{1}{(d+4)}} n^{\frac{-1}{(d+4)}} \sigma_j \tag{4.29}$$

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 equação 4.29 coincide com o método proposto por Silverman, equação 4.22.

Sendo $x = x_1, x_2, ..., x_v$ temos a fórmula para o KDE Multivariado, que será chamado, nessa dissertação, de *Multivariate Kernel Density Estimation)* (MKDE):

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

com,

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

4.2.1 'A MALDIÇÃO DA DIMENSIONALIDADE'

Como um problema adicional, existe uma dificuldade, conhecida como a "maldição da dimensionalidade"(NARSKY; PORTER, 2013), que ocorre quando se faz estimativa de densidade de maiores ordens. Silverman (SILVERMAN, 1986) aborda esse tema descrevendo a importância dos eventos de cauda em altas dimensões. Em uma PDF univariada, bem estimada, aproximadamente 1% dos eventos caem nas regiões de cauda. Já para PDF's de 10 dimensões, estima-se que mais da metade dos eventos caiam em regiões de baixa densidade e ignorá-los pode afetar consideravelmente os resultados (LEDL, 2002). A grosso modo, isso acontece porque, em mais dimensões, existe mais espaço para observações e essa fato torna a estimação não-paramétrica da densidade mais complicada que o caso univariado. A Figura 19 sugere como os eventos ficam cada vez mais esparsos com o aumento das dimensões e indica quantos eventos a mais seria necessário para preencher proporcionalmente os respectivos espaços amostrais.

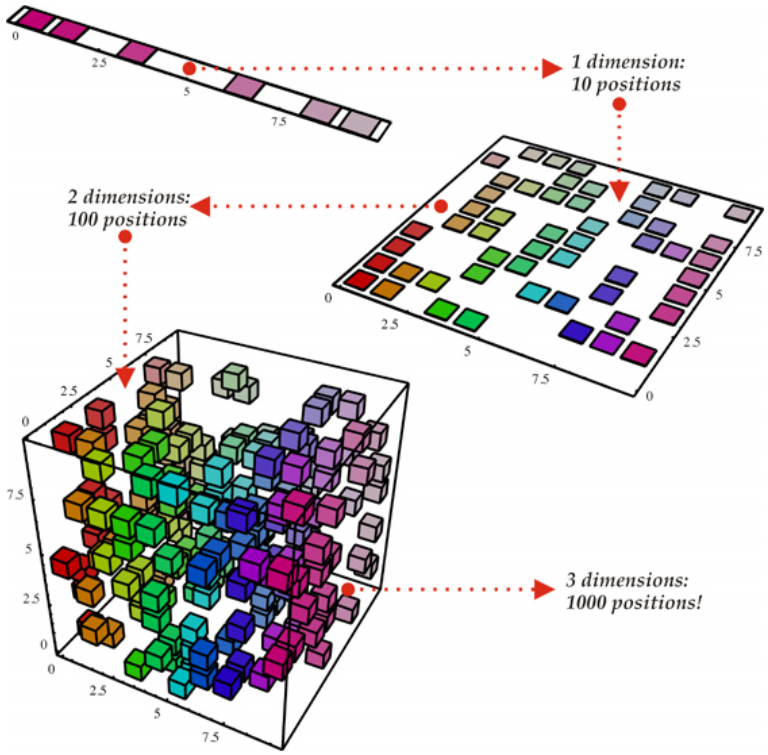

Figura 19: Demonstração gráfica da "Maldição da dimensionalidade", retirado de (BENGIO, 2016)

5 IMPLEMENTAÇÃO DO KDE

Neste capítulo, serão descritos os detalhes de construção do algoritmo de identificação de elétrons desenvolvido no âmbito dessa dissertação. A estimação da densidade conjunta das variáveis discriminantes é central para o método de Verossimilhança e, devido a isto, este trabalho se concentrou neste tópico, a fim de indicar possíveis caminhos na definição de uma metodologia a ser adotada na implementação do KDE. Para tal, dados de Monte Carlo gerados pela Colaboração ATLAS, conforme listado abaixo, foram utilizados. Suas características estão resumidas na Figura 20.

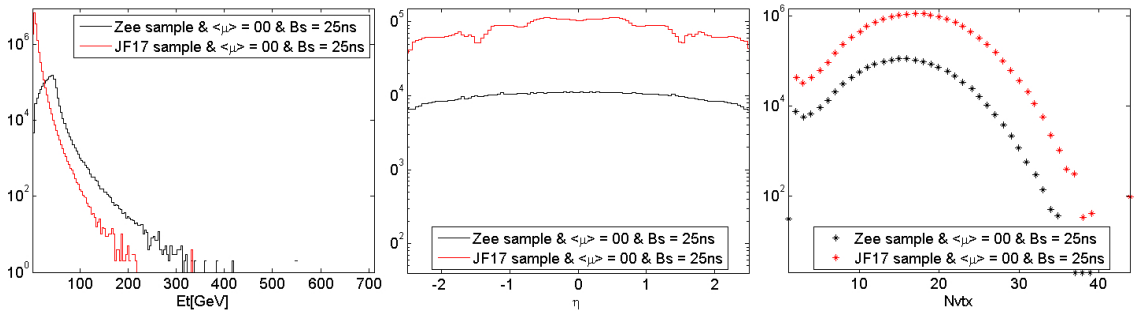

Figura 20: Perfil dos eventos gerados por MC. Gráfico de eventos por E_t (Esquerda), Gráfico de eventos por η (Centro) e Gráfico de eventos por NVTX (Direita).

Esses dados foram separados em dois conjuntos: desenvolvimento e validação (50% para cada conjunto). O primeiro utilizado para a estimação da densidade conjunta e o segundo para estudo e análise de desempenho do algoritmo proposto. Os dados de desenvolvimento foram reorganizados e pré-processados para que pudessem ser usados na estimação da densidade. A reorganização dos dados simplesmente cria subconjuntos de sinal e ruído, os separando por 4 regiões de η e 3 de E_t. Em princípio, quanto maior o número de divisões em η e E_t, melhor representado estará o detector, uma vez que esses parâmetros influenciam nos perfis de energia deixados pelas partículas no mesmo. No entanto, as regiões foram escolhidas considerando-se, principalmente, a estatística e as regiões de interesse da Colaboração ATLAS. Cabe ressaltar que a região de E_t de

$5 - 20GeV$ é bastante interessante devido ao alto nível de ruído de fundo e à medida que se analisa uma faixa de energia mais alta, esse ruído diminui. A Tabela 3 mostra as regiões selecionadas e o número de eventos em cada uma delas.

Tabela 3: Tabela de divisão de regiões em η e E_t.

Regiões de Estudo				Número de Eventos	
n	$\|\eta\|$		E_t (GeV)	Sinal	Ruído de Fundo
1	0	0,8	5 - 20	98.974	2.222.932
2			20 - 30	142.168	120.237
3			> 30	489.638	45.339
4	0,8	1,37	5 - 20	66.264	1.600.333
5			20 - 30	98.404	83.452
6			> 30	329.811	31.008
7	1,52	2,01	5 - 20	61.774	1.444.673
8			20 - 30	99.586	81.168
9			> 30	325.878	28.938
10	2,01	2,47	5 - 20	32.599	845.501
11			20 - 30	55.539	51.550
12			> 30	204.215	18.689
Total				2.004.850	6.573.820

O pré-processamento dos dados é utilizado para a remoção de descontinuidades presentes nos histogramas das variáveis discriminantes, efeito este que ocorre, principalmente, por indeterminações matemáticas no processo de cálculo das probabilidades, como por exemplo, divisão por zero. Uma das variáveis que sofre com este efeito é a r_{Had}, como pode ser visto na Figura 21.

Neste trabalho, foram seguidos dois caminhos para a implementação do método de Verossimilhança aplicado à seleção de elétrons. O primeiro considerou apenas densidades de probabilidade unidimensionais, em seguida densidades bidimensionais foram consideradas. Embora o primeiro caso seja, supostamente, uma reprodução do método aplicado atualmente na Colaboração ATLAS, todo o algoritmo foi desenvolvido de modo customizado, inclusive aplicando ferramentas diferentes daquelas utilizadas pela Colaboração ATLAS e buscando otimizar alguns parâmetros de implementação. Portanto, espera-se que os resultados finais sejam diferentes daqueles obtidos com o algoritmo da Colaboração ATLAS e, desta maneira, esperamos trazer algumas contribuições para o método atualmente empregado. Já a proposta de aplicação de densidades

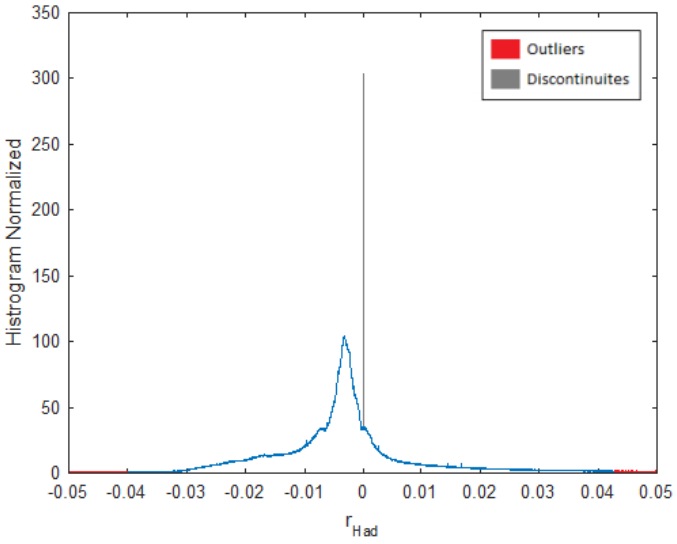

Figura 21: Gráfico da distribuição da variável r_{Had}, exemplificando os Valores Extremos (*Outliers*) e Descontinuidades (*Discontinuites*).

bidimensionais, foi feita com o objetivo de estudar o impacto das mesmas na mitigação de eventuais erros no cálculo de probabilidade, que podem ocorrer devido à reconstrução da densidade conjunta a partir de suas densidades unidimensionais, considerando que, possivelmente, exista dependência entre algumas das variáveis discriminantes.

Em particular, a estimação das densidades utilizando-se o método de Kernel é aplicada na parte central das distribuições e as regiões de cauda são desconsideradas nesta primeira etapa. Como cada variável apresenta uma densidade para sinal e outra para ruído de fundo, sendo que a combinação dessas probabilidades gera o discriminante, pode-se dividir os eventos de uma mesma variável em três regiões, como mostra a Figura 22. Embora esta figura ilustre o caso unidimensional, esta classificação também é válida para densidades bivariadas, seguindo a mesma lógica. As três regiões destacadas na Figura 22, são descritas abaixo:

***Center-Center* (CC)** Conjunto de eventos que pertencem a região central das densidades do sinal e do ruído de fundo;

***Center-Tail* (CT)** Conjunto dos eventos que caem na região central da PDF do sinal e na região de cauda da PDF do ruído de fundo, ou vice-versa (*Tail-Center* (TC)), sendo que os dois casos serão tratados com a nomenclatura CT nesse texto;

***Tail-Tail* (TT)** Conjunto dos eventos que caem na região das caudas das PDF's tanto de sinal quanto de ruído de fundo;

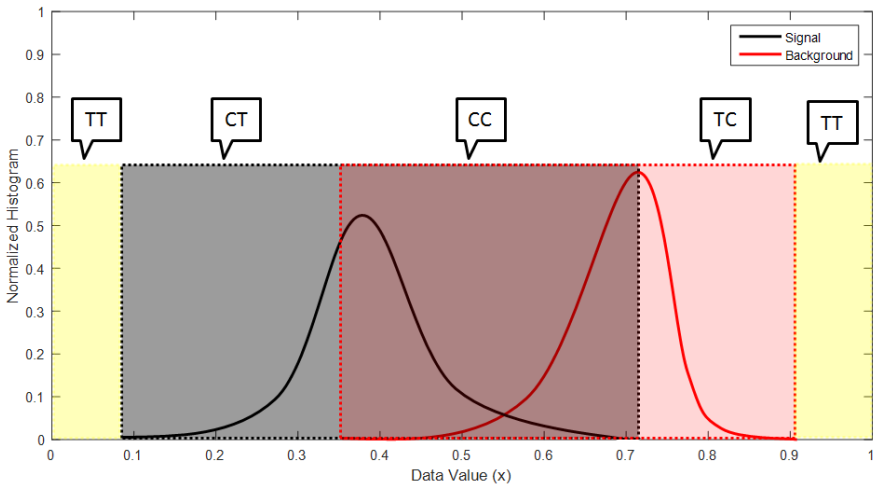

Figura 22: Exemplificação da sub-divisão dos eventos por regiões.

Como mencionado anteriormente, a estimação de densidade usando o método de kernel ocorre somente na região central, enquanto que, para resolver os eventos de cauda, foi proposto um algoritmo de extrapolação, baseado em um ajuste exponencial, a partir de valores estimados na região central localizados em sua periferia. O impacto desse algoritmo foi então avaliado frente a outras duas possibilidades: (1) a não utilização dos valores de cauda, denominado nesse trabalho de *Bypass*, e (2) a utilização dos valores mínimos obtidos a partir da região central das densidades estimadas pelo método de Kernel, chamado nesse trabalho de método do vizinho mais próximo. O primeiro item leva a desconsiderar uma dada variável na reconstrução das probabilidades dos eventos que caem na região da cauda, o que significa que a decisão de seleção será formada pelas demais variáveis discriminantes. O segundo item, de acordo com (COLLABORATION et al., 2013), representa o critério aplicado atualmente pelo Colaboração ATLAS.

No decorrer desse capítulo, mostraremos a implementação do kernel univariado na Seção 5.1 e do bivariado na Seção 5.2. Nos dois casos, as regiões de centro (CC) e de cauda (CT e TT) foram avaliados separadamente, uma vez que seus valores são estimados por dois processos diversos.

5.1 KERNEL UNIDIMENSIONAL

5.1.1 TRATAMENTO DOS EVENTOS CENTRAIS

A implementação do algoritmo que trata dos eventos centrais se baseia nas formulações matemáticas descritas na Seção 4.1. Nessa abordagem, é considerado um KDE

com banda variável otimizado através do parâmetro λ, de acordo com a otimização apresentada na Seção 4.1.4.2, restando utilizar o histograma, com a binagem escolhida baseada no algoritmo proposto por (SHIMAZAKI; SHINOMOTO, 2007), para calcular o valor de $f(x)$ de cada evento, como mostrado na equação 4.24.

Computacionalmente, no processo de estimação das densidades, deve-se escolher o número de pontos a ser usado para a sua representação discreta. Adicionando esse número aos parâmetros mencionados no parágrafo anterior, as PDF podem ser construídas através do método de KDE. O algoritmo implementado foi programado para gerar 2000 pontos para cada variável, de maneira a garantir uma boa representação de suas respectivas distribuições. Os valores das variáveis gerados pelos eventos de colisão são, então, interpolados linearmente a partir dos pontos da PDF discreta.

As Figuras 23 até 29 mostram as PDF estimadas pelo algoritmo de KDE na região 1, para o conjunto de sinal (denominado *electron*, por ser formado por elétrons isolados) e as Figuras 30 até 36 mostram o conjunto de ruído de fundo (denominado *Jet*, pela predominância de jatos). A escolha da melhor binagem para visualização do histograma foi feita utilizando a medida do χ^2.

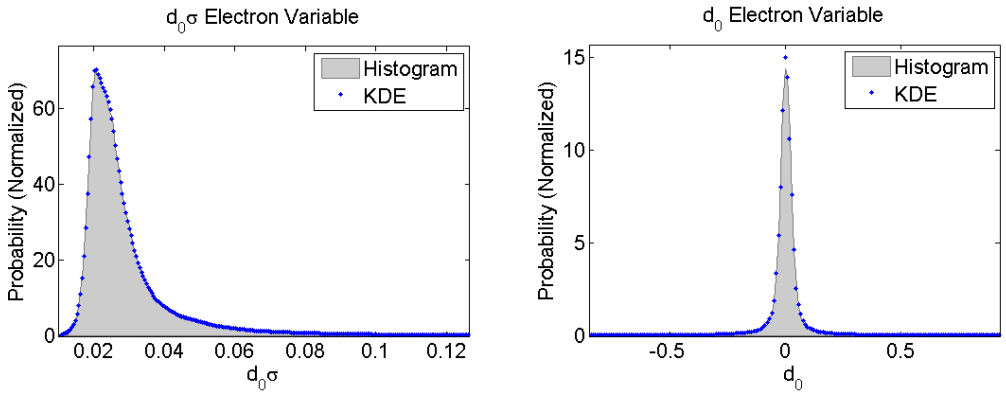

Figura 23: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $d_{0\sigma}$ e (Direita) Variável d_0.

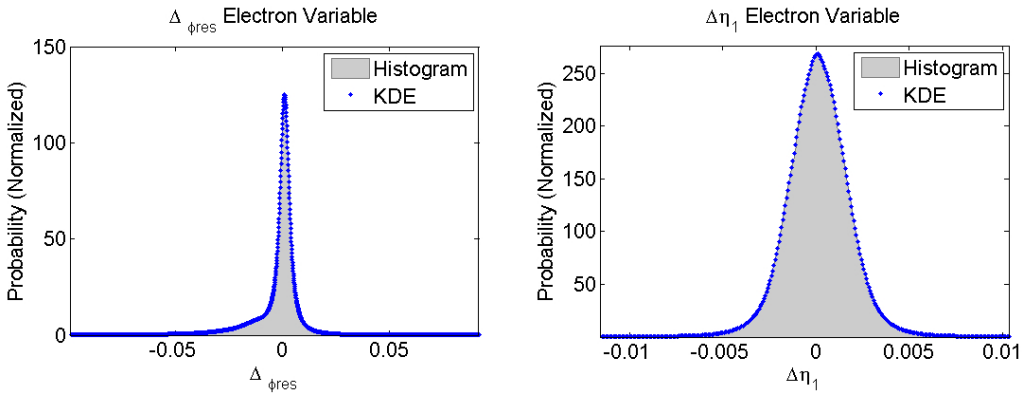

Figura 24: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $\Delta_{\phi res}$ e (Direita) Variável $\Delta_{\eta 1}$.

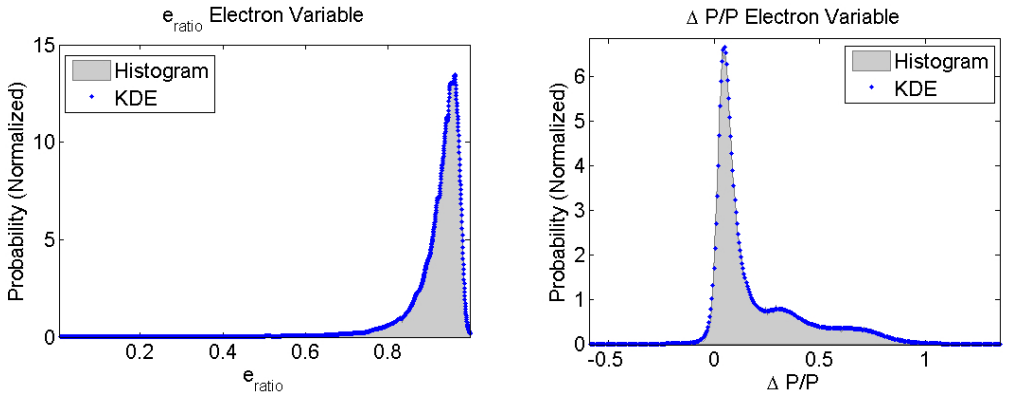

Figura 25: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável E_{ratio} e (Direita) Variável $\Delta P/P$.

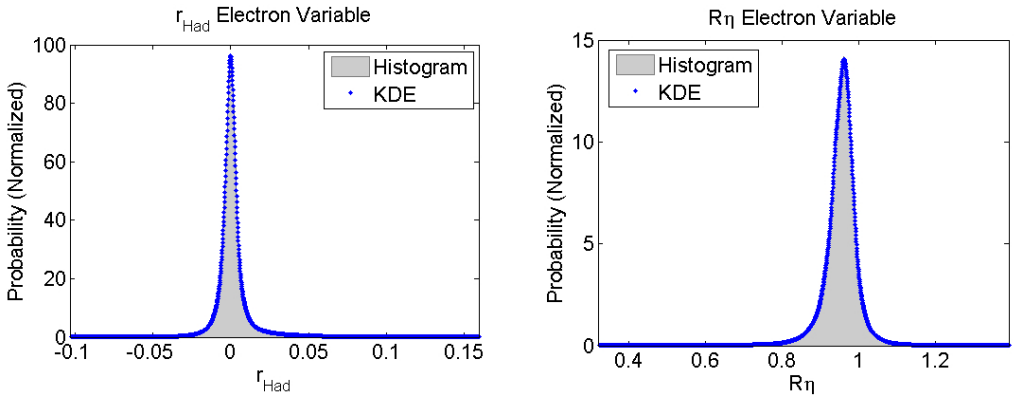

Figura 26: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável R_{Had} e (Direita) Variável r_{η}.

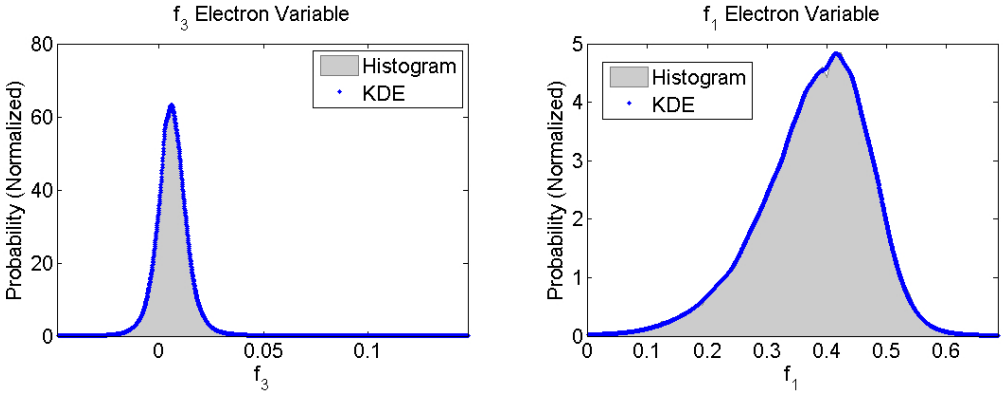

Figura 27: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável f_3 e (Direita) Variável f_1.

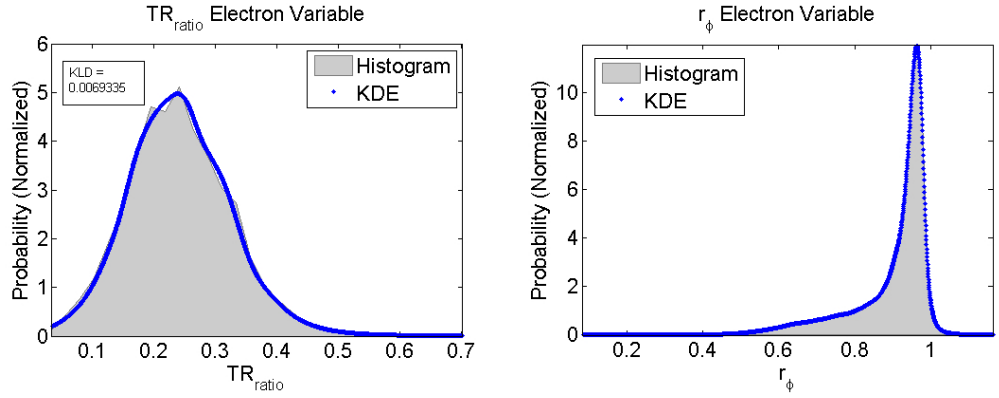

Figura 28: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável TR_{ratio} e (Direita) Variável r_ϕ.

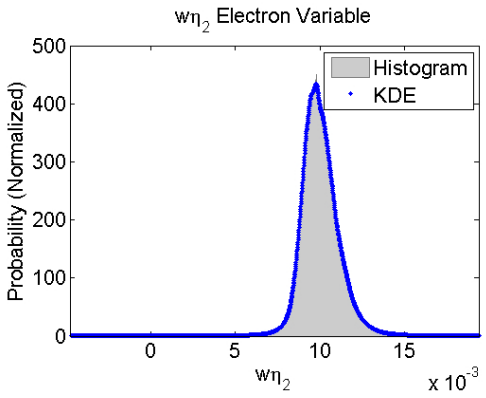

Figura 29: Gráfico da distribuição da variável de elétron e a sua PDF estimada pelo KDE (Dados de MC). Variável $W_{\eta 2}$.

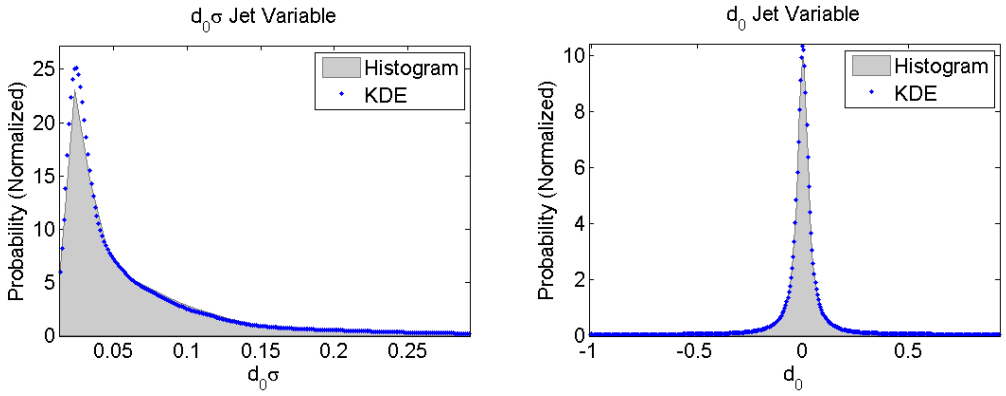

Figura 30: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $d_{0\sigma}$ e (Direita) Variável d_0.

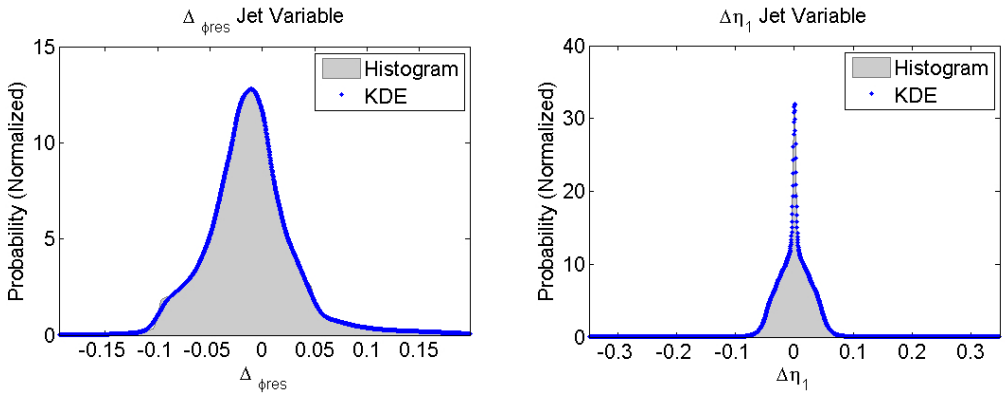

Figura 31: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável $\Delta_{\phi res}$ e (Direita) Variável $\Delta_{\eta 1}$.

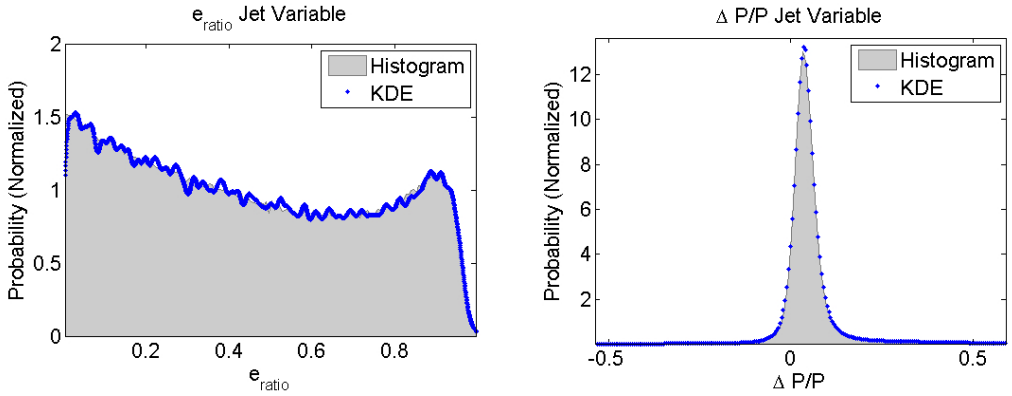

Figura 32: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável E_{ratio} e (Direita) Variável $\Delta P/P$.

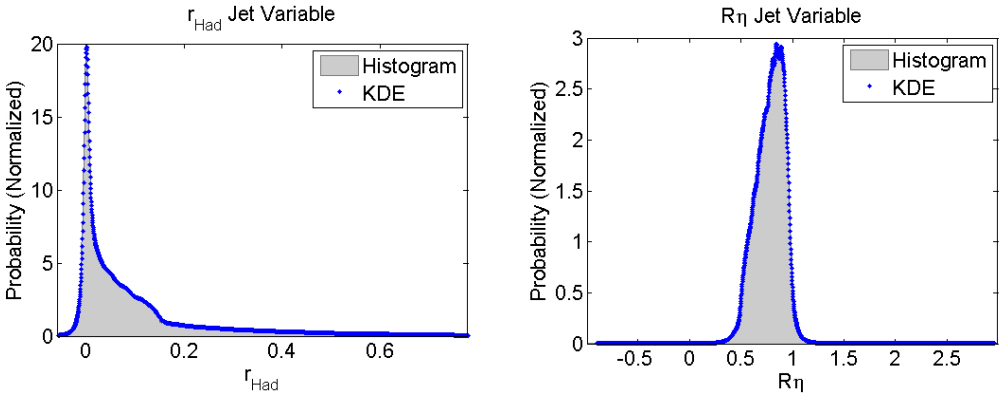

Figura 33: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável R_{Had} e (Direita) Variável r_η.

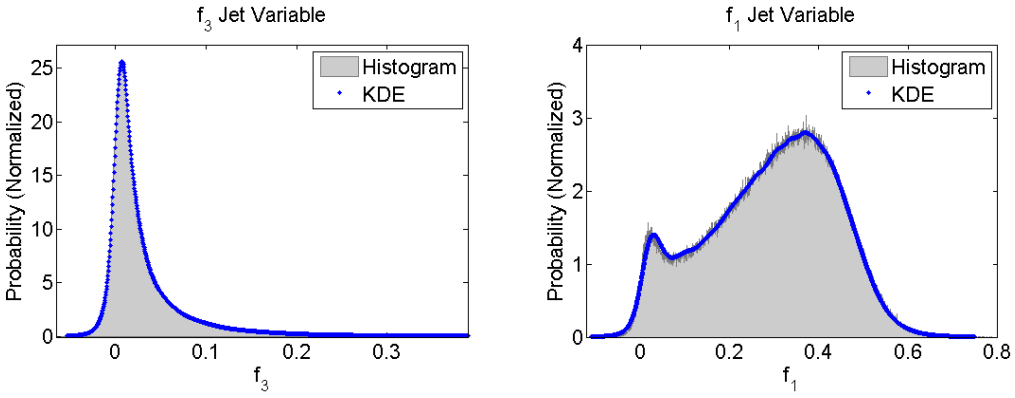

Figura 34: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável f_3 e (Direita) Variável f_1.

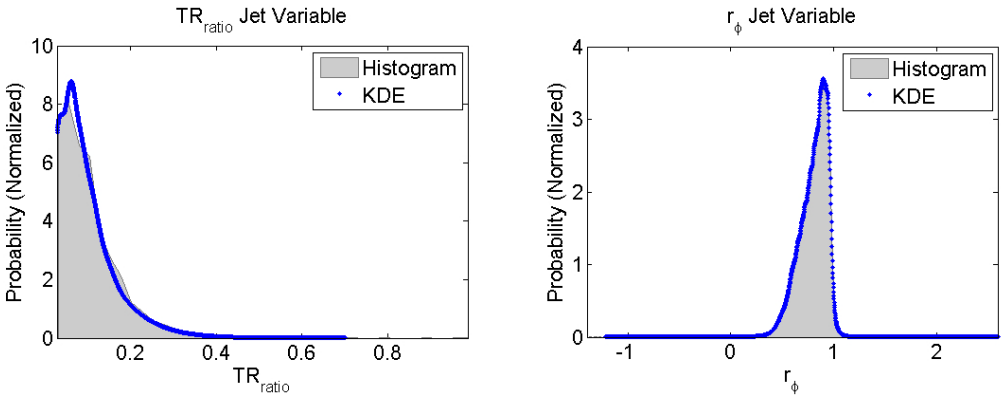

Figura 35: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). (Esquerda) Variável TR_{ratio} e (Direita) Variável r_ϕ.

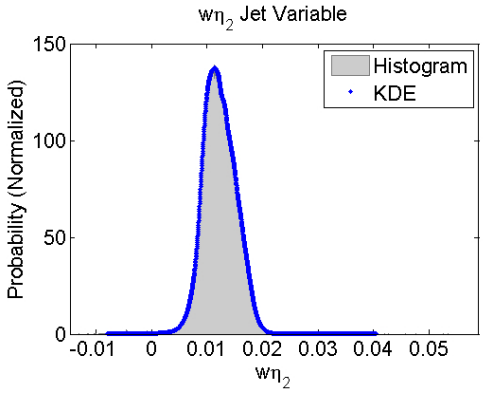

Figura 36: Gráfico da distribuição da variável de jato e a sua PDF estimada pelo KDE (Dados de MC). Variável $W_{\eta 2}$.

5.1.2 TRATAMENTO DOS EVENTOS DE CAUDA

A extrapolação para o cálculo das probabilidade dos eventos faz-se necessária quando esses ocorrem na região de cauda da PDF. Os dois casos contendo valores de cauda, CT e TT, foram avaliados separadamente.

O cálculo do discriminante dos eventos de CT é feito com a combinação das probabilidades geradas pelo interpolador linear (para a região central) e extrapolação (para a região de cauda). O desempenho da extrapolação exponencial foi comparada com o método do vizinho mais próximo, que assume que a probabilidade de um evento de cauda é igual a probabilidade do evento de centro mais próximo. A Figura 37 mostra quatro casos representativos. Apenas dados CT foram utilizados na construção dessas ROC individuais (considerando uma única variável discriminante por vez). Nota-se que a aplicação do ajuste exponencial pode melhorar consideravelmente o poder de discriminação das variáveis em uso e, portanto, decidiu-se pela aplicação da mesma para os eventos que caem na região CT.

Para o caso TT, duas abordagens foram consideradas: extrapolação exponencial e o método chamado de *Bypass*. Como explicado anteriormente, pelo método *Bypass*, as variáveis com eventos na região TT são ignoradas, fazendo com que o cálculo da probabilidade final seja obtido a partir das outras variáveis (cujos eventos caem nas regiões CT e CC). Portanto, o desempenho de ambas deve ser comparado utilizando-se todos os eventos, a partir de uma análise com as 13 variáveis discriminantes. Os gráficos da Figura 38 mostram que o método *Bypass* apresenta melhor desempenho, produzindo melhorias significativas no poder de discriminação do algoritmo na maioria das regiões. Sendo assim, a método de *Bypass* foi escolhido como solução padrão para

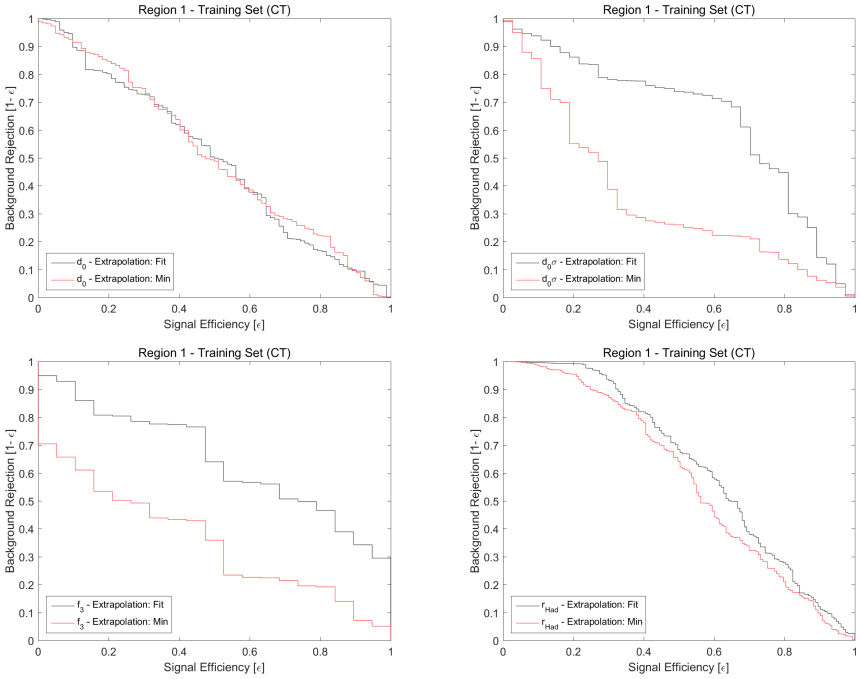

Figura 37: Gráfico comparando a extrapolação feita pelo método exponencial (*Fit*) e pelo método de vizinho mais próximo (*Min*) na Região 1, a partir das ROC individuais das seguintes variáveis: d_0, σ_{d_0}, f_3 e r_{Had}.

os eventos TT.

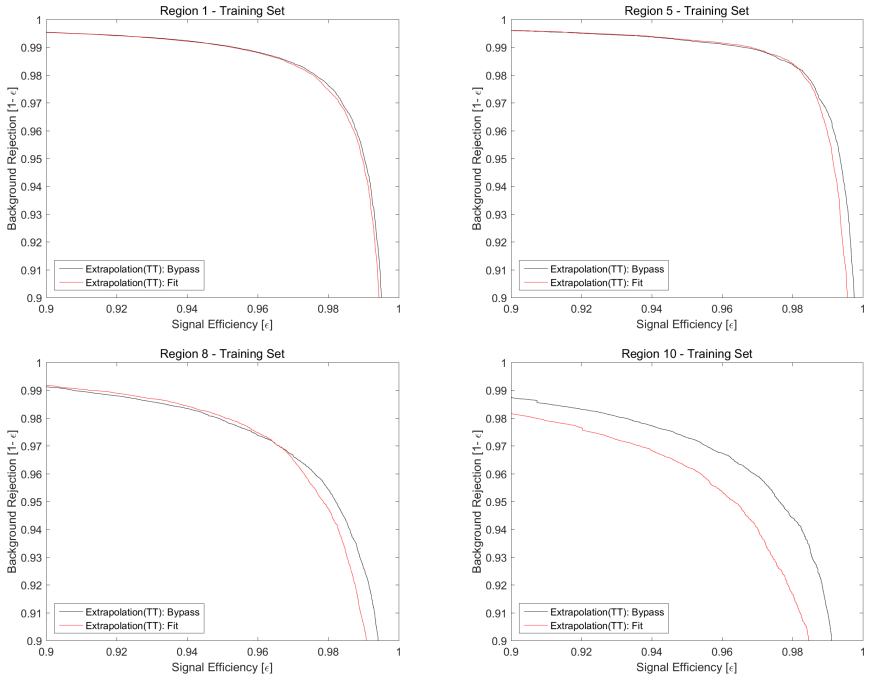

Figura 38: Gráfico comparando as ROC's de todos os eventos com a extrapolação feita pelo método exponencial (*Fit*) e pelo método de *Bypass* aplicados somente nos eventos de TT. (Superior Esquerda) ROC da Região 1; (Superior Direita) ROC da Região 5; (Inferior Esquerda) ROC da Região 8 e (Inferior Direita) ROC da Região 10.

5.1.3 PRINCIPAIS DIFERENÇAS EM RELAÇÃO AO MÉTODO UTILIZADO PELA COLABORAÇÃO ATLAS

Como visto nas seções precedentes, a busca por uma otimização do algoritmo baseado no método de Verossimilhança a partir da estimação de densidades univariadas desenvolvida neste trabalho levaram a uma implementação ligeiramente diversa daquela em uso atualmente pela Colaboração. A Tabela 4 resume as principais diferenças encontradas.

Tabela 4: Tabela de diferenças entre a LH da colaboração (COLLABORATION et al., 2013) e o algoritmo implementado nessa dissertação.

	LH Colaboração Atlas	LH desse trabalho
Método	Verossimilhança	Verossimilhança
Variáveis	13	13
Cortes Adicionais	Todos	Todos
Interpolação	*	Linear
Extrapolação	Vizinho mais próximo	Exponencial
PDF	KDE	MKDE
Regiões em η e E_t	9x6	4x3
Pontos na PDF	520**	2000

*A nota (COLLABORATION et al., 2013) não deixa clara a forma de interpolação utilizada. **Para a variável TRT, utilizaram-se 62 pontos.

5.2 KERNEL BIDIMENSIONAL

Nessa seção, serão apresentados os principais itens que conduziram ao algoritmo de classificação baseado em densidades bidimensionais proposto nesse trabalho. De forma análoga à análise unidimensional, avaliaremos separadamente o desempenho do algoritmo para os eventos centrais daqueles de cauda, de acordo com as Seções 5.2.1 e 5.2.2, respectivamente.

5.2.1 TRATAMENTO DOS EVENTOS CENTRAIS

Os eventos centrais da análise bidimensional são tratados da mesma forma que os eventos centrais na análise unidimensional; entretanto, as PDF bidimensionais são estimadas pelo algoritmo de estimação de densidades multivariada, que é feito utili-

zando a teoria descrita na Seção 4.2, de acordo com *MKDE* de banda fixa baseado na equação 4.30.

Aqui, seguindo a mesma lógica do caso unidimensional, as densidades foram discretizadas a partir de uma matriz 2000x2000 e as probabilidades dos eventos de colisão são obtidos a partir de uma interpolação linear aplicado a partir dos valores discretos da PDF.

As Figuras 39 até 42 mostram a saída do algoritmo de MKDE para alguns pares de variáveis da região 1, e as Figuras 43 até 46 os pares para região 2, trazendo, ao lado das mesmas, para fins de comparação, as densidades estimadas a partir do método univariado, utilizando-se de suas marginais (*Marginal*) e do histograma da distribuição.

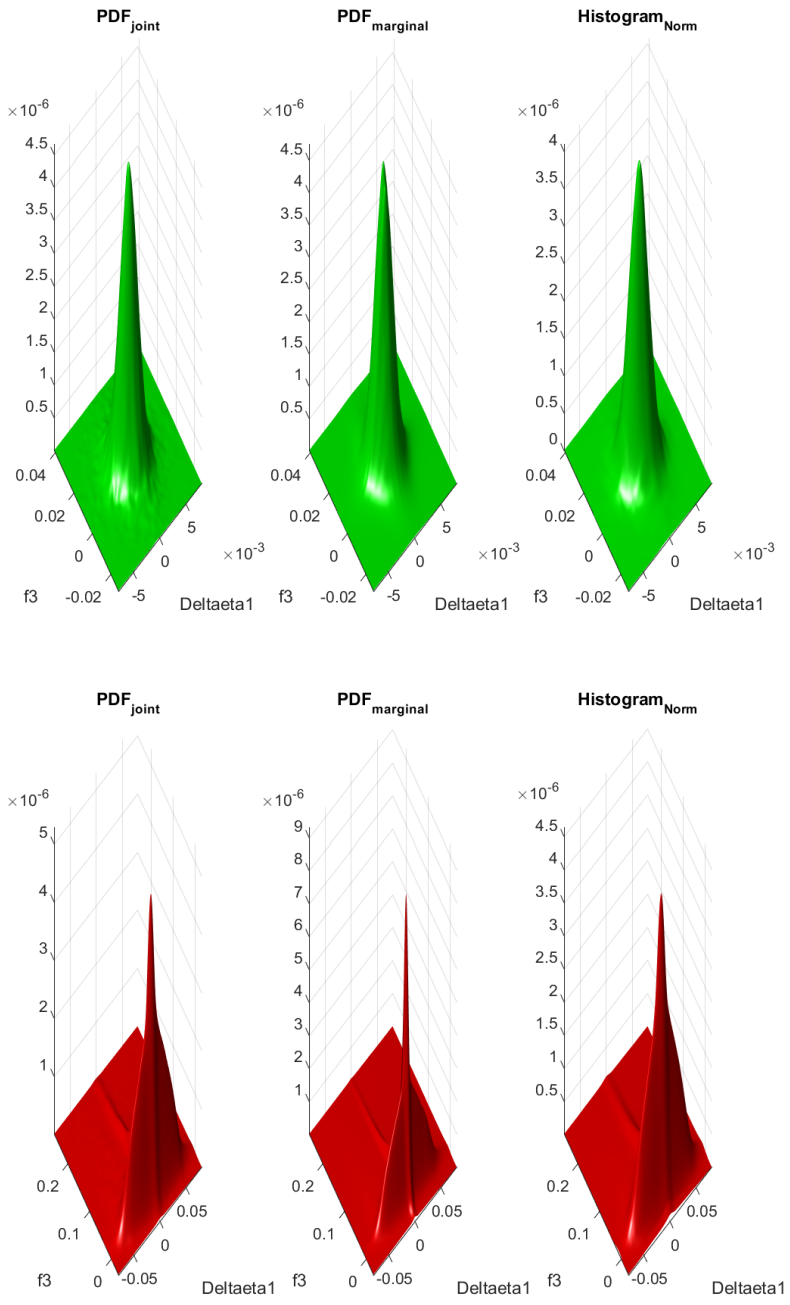

Figura 39: Gráfico que ilustra, para o par de variáveis $\Delta_{\eta 1}$ e f_3, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 1, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

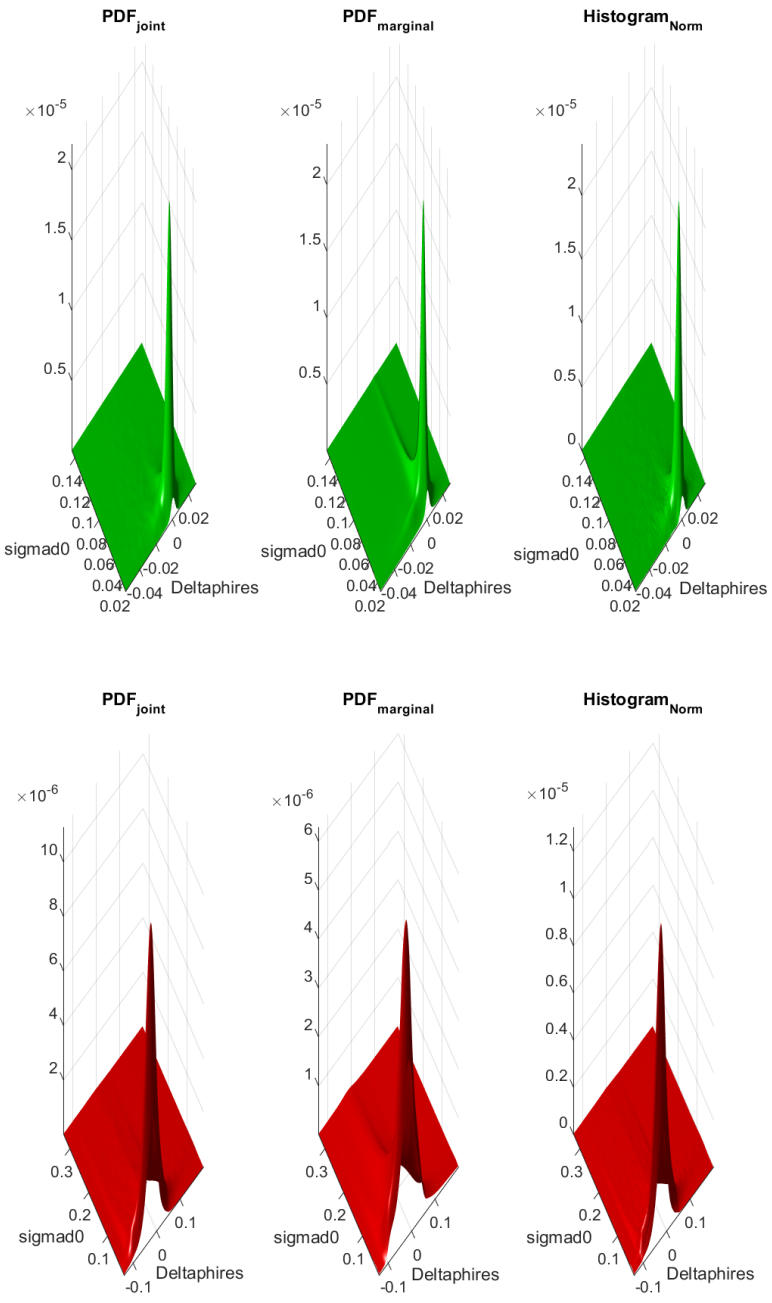

Figura 40: Gráfico que ilustra, para o par de variáveis $\Delta_{\phi res}$ e σ_{d_0}, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 1, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

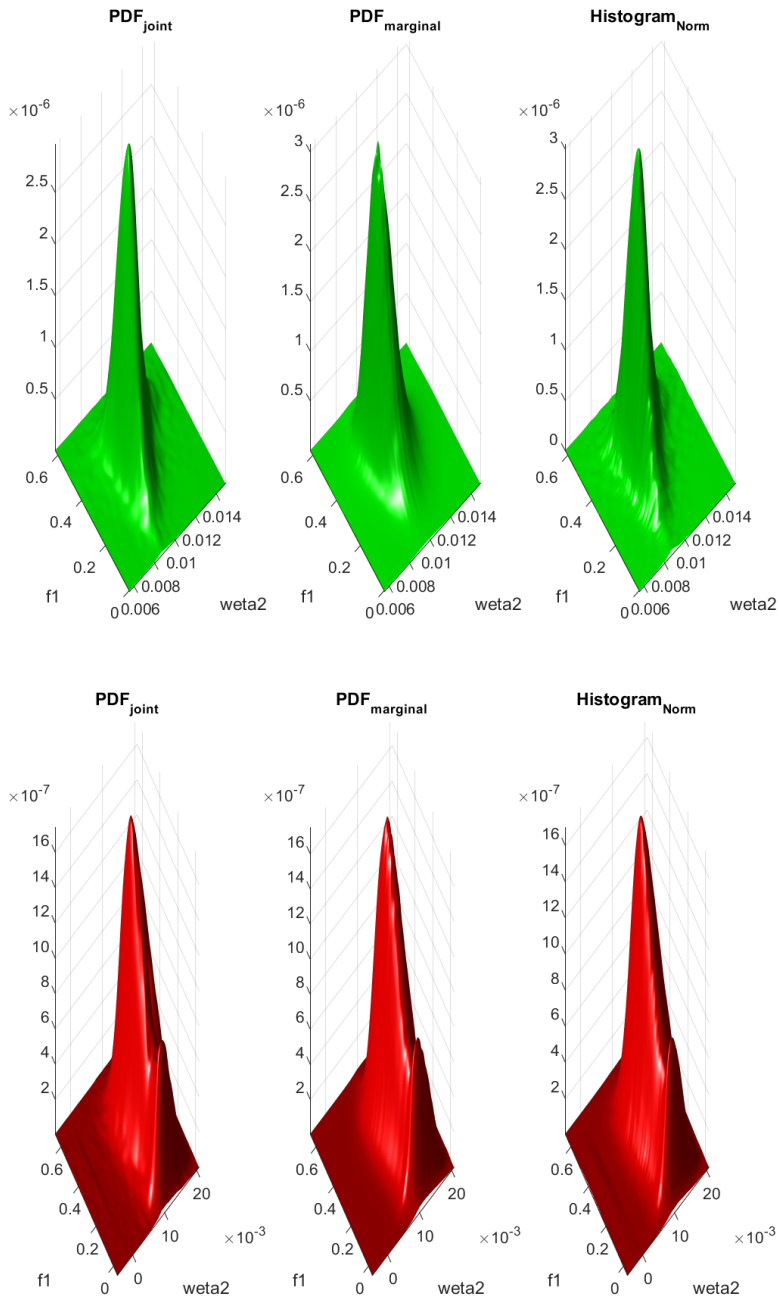

Figura 41: Gráfico que ilustra, para o par de variáveis $w_{\eta 2}$ e f_1, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 1, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

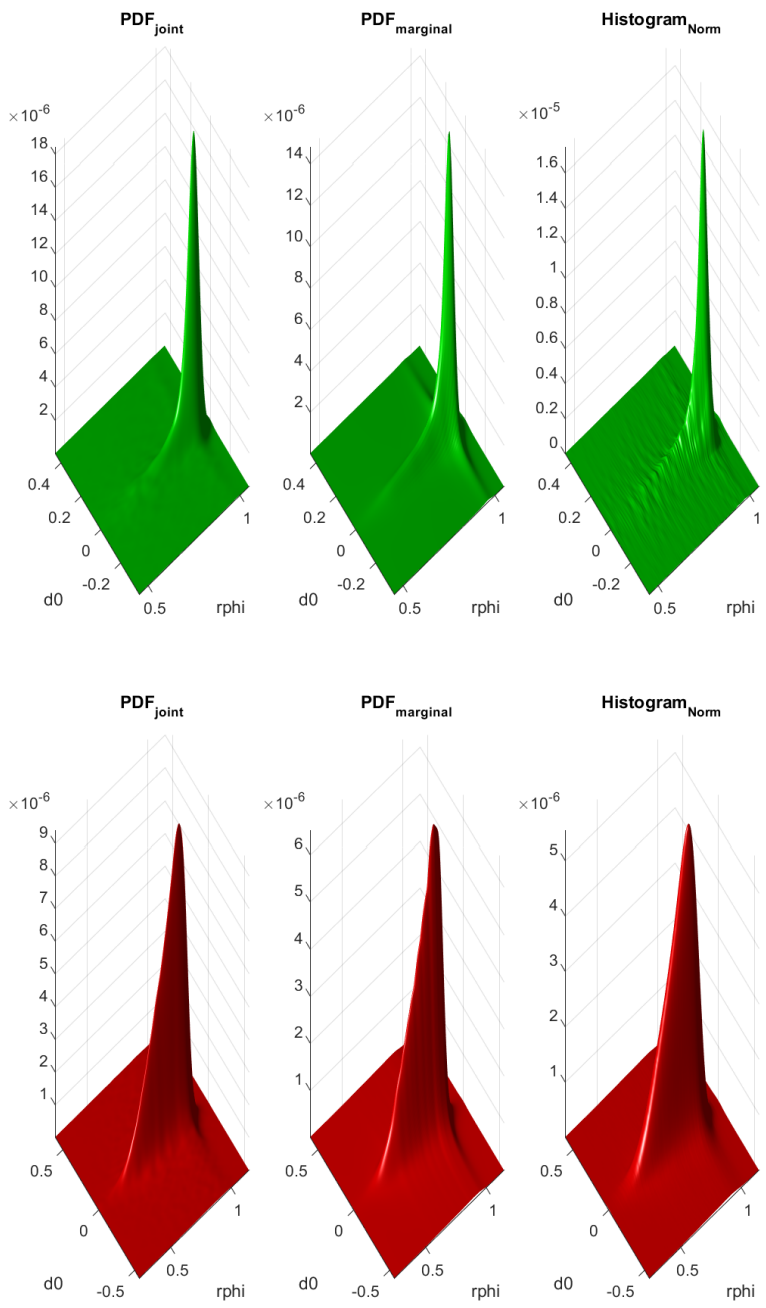

Figura 42: Gráfico que ilustra, para o par de variáveis r_ϕ e d_0, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 1, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

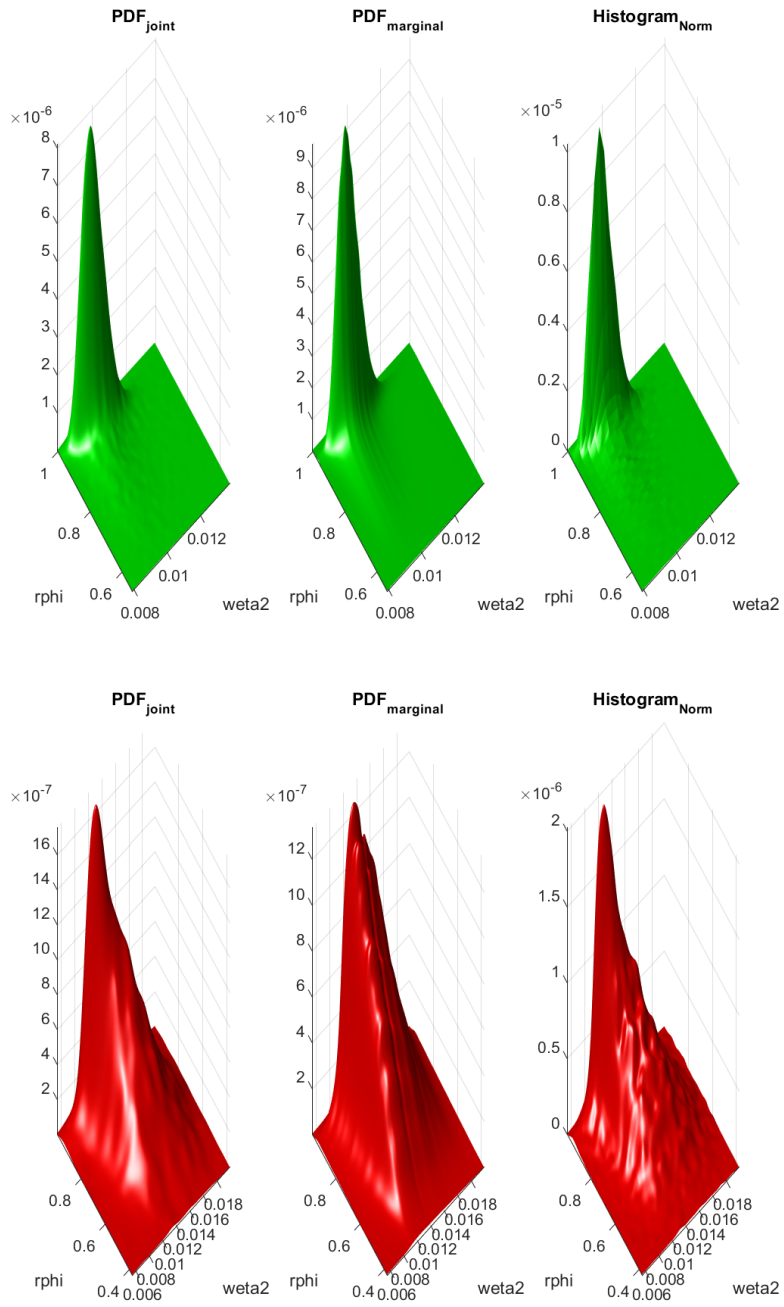

Figura 43: Gráfico que ilustra, para o par de variáveis r_ϕ e w_{η_2}, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 2, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

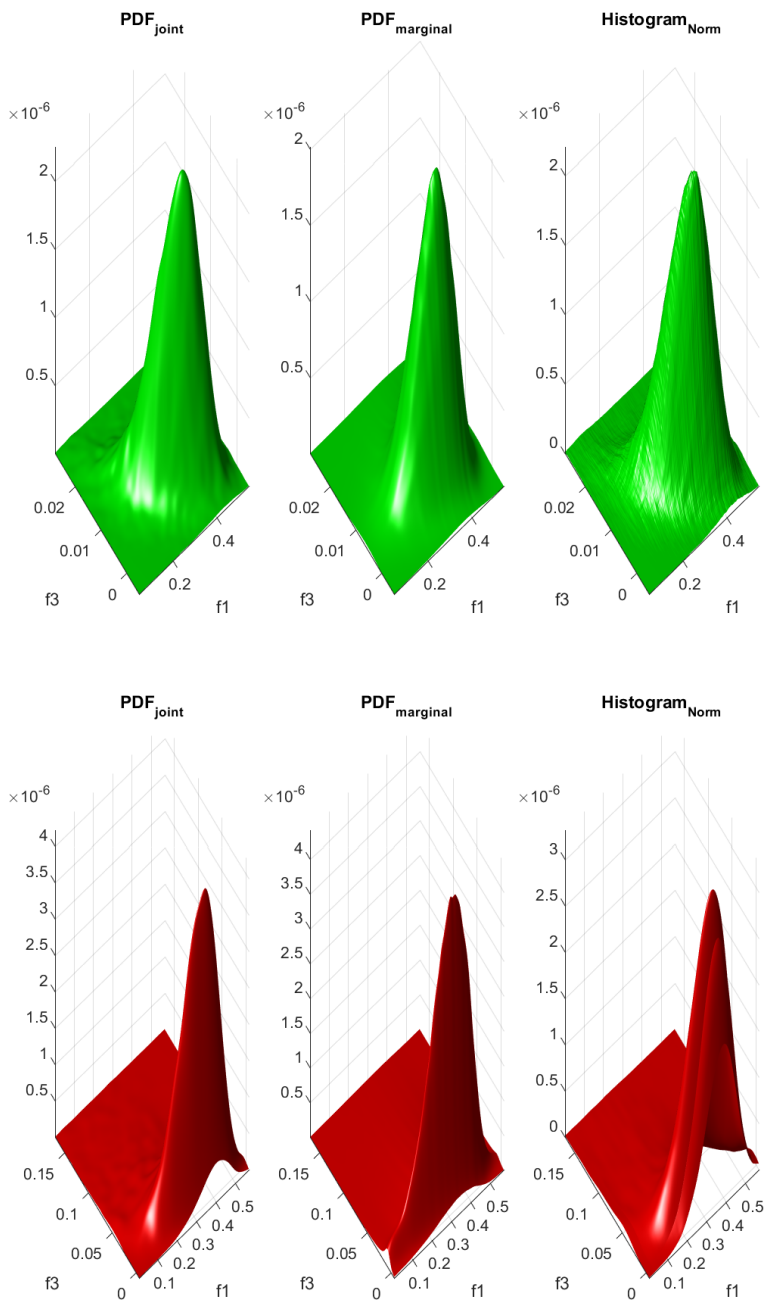

Figura 44: Gráfico que ilustra, para o par de variáveis f_1 e f_3, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 2, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

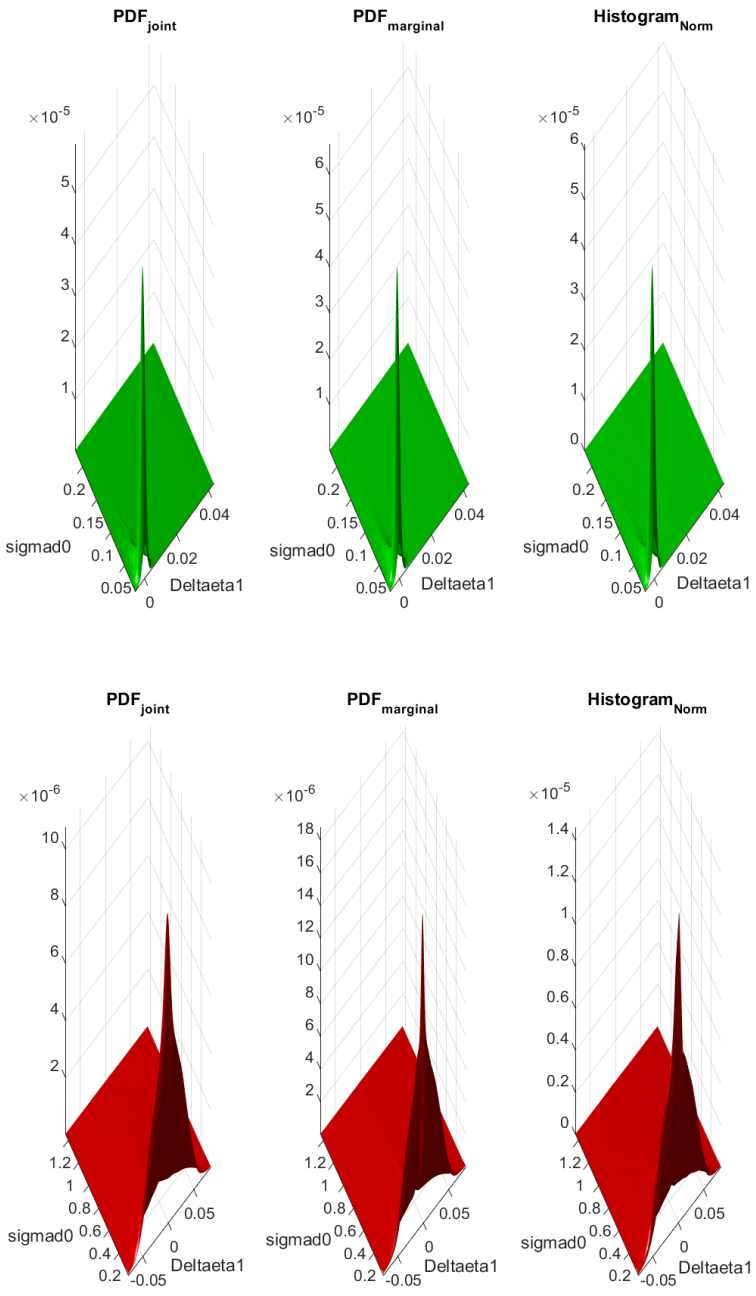

Figura 45: Gráfico que ilustra, para o par de variáveis $\Delta_{\eta 1}$ e σ_{d_0}, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 2, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

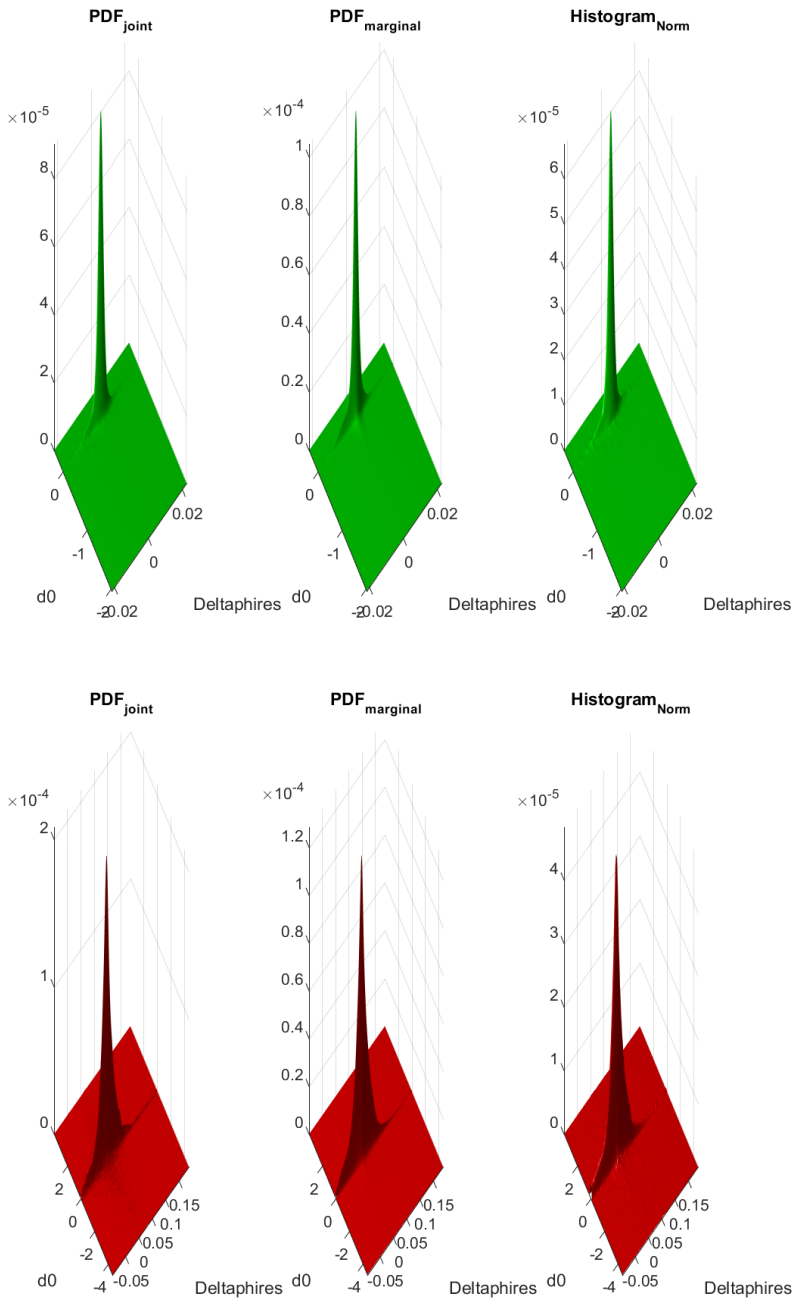

Figura 46: Gráfico que ilustra, para o par de variáveis $\Delta_{\phi res}$ e d_0, a diferença entre a estimação de densidade conjunta (*joint*) e marginal (*marginal*), com o histograma da distribuição, para a Região 2, utilizando somente os eventos de CC. (Superior) Sinal; (Inferior) Ruído de Fundo

.

5.2.1.1 ESCOLHA DAS VARIÁVEIS PARA DENSIDADE BIVARIADA

A escolha dos pares de variáveis discriminantes que serão utilizados para estimação de densidade bivariada é de suma importância para esse trabalho. Essa abordagem multivariada tem como intuito reduzir o impacto da aproximação de independência efetuada utilizando a simplificação da fórmula de verossimilhança.

Uma metodologia de análise proposta nessa dissertação é observar a diferença entre a Área Sob a Curva, (do inglês, *Area Under the ROC Curve*) (AUC) das análises conjunta (multivariada) e marginal (univariada) para todos os pares de variáveis e então escolher as que apresentarem os maiores índices de melhoria, no intuito de otimizar a performance do algoritmo de classificação de eventos multivariado. A Figura 47 apresenta a diferença de AUC das ROCs obtidas a partir dos algoritmos de estimação unidimensional e bidimensional aplicados a cada um dos pares possíveis entre as 13 variáveis discriminantes. Observa-se que algumas combinações de pares são possíveis. Os pares de variáveis escolhidos para a região 1 foram $\Delta_{\eta 1}$ e f_3; $\Delta_{\phi res}$ e σ_{d_0}; $w_{\eta 2}$ e f_1; e r_ϕ e d_0.

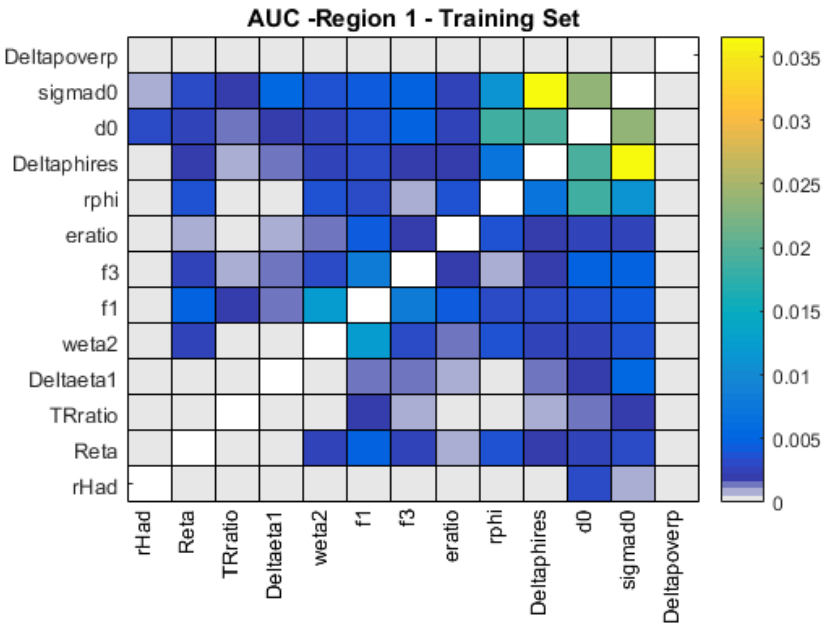

Figura 47: Gráfico de AUC para os pares de variáveis, considerando a Região 1, utilizando somente os eventos de CC.

A Figura 48 apresenta os gráficos das ROC's dos pares de variáveis escolhidos e a comparação com suas marginais. Pode-se observar que existe uma melhora, em alguns casos, considerável, que surge ao se utilizar a abordagem bivariada.

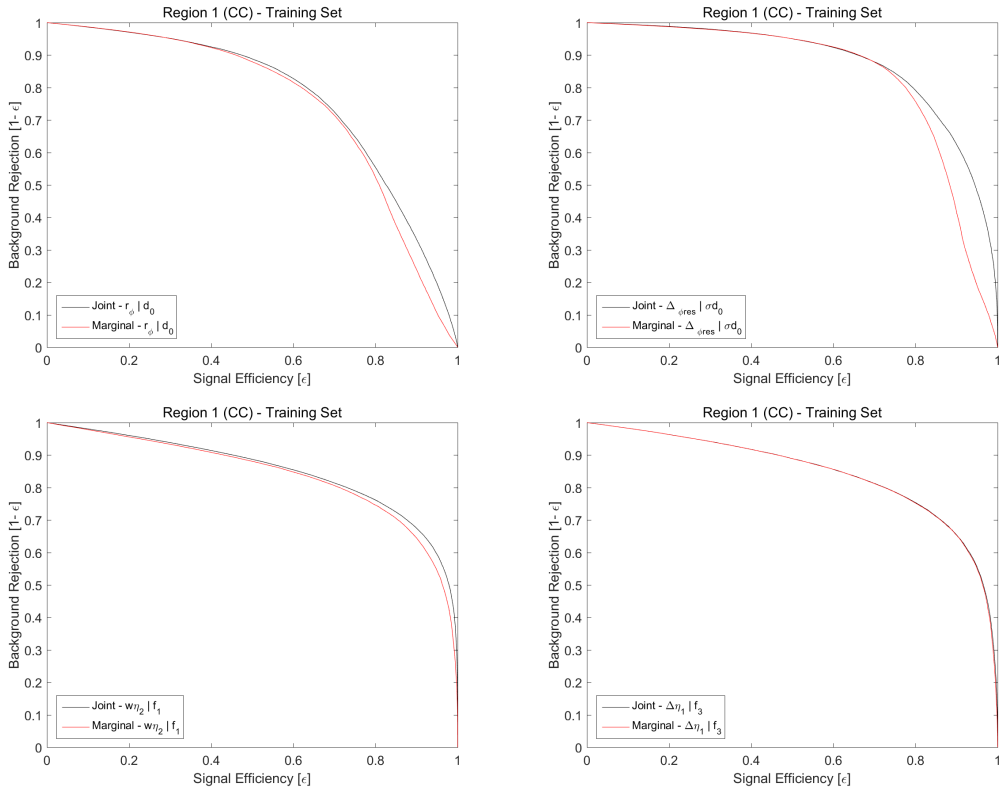

Figura 48: Gráfico comparando as ROC's as análises conjunta e marginal, para os eventos de CC da Região 1. (Superior Esquerda) Par de variáveis - r_ϕ e d_0; (Superior Direita) Par de variáveis - $\Delta_{\phi res}$ e σ_{d_0}; (Inferior Esquerda) Par de variáveis - $w_{\eta 2}$ e f_1 e (Inferior Direita) Par de variáveis - $\Delta_{\eta 1}$ e f_3.

A Figura 50 mostra os pares de variáveis escolhidos para a região 2 de acordo com a Figura 49. Note que as combinação de variáveis pode variar de acordo com a região do detector; para a região 2 os seguintes pares foram selecionados: w_{η_2} e r_ϕ; $\Delta_{\eta 1}$ e σ_{d_0}; $\Delta_{\phi res}$ e d_0; e f_1 e f_3.

5.2.2 TRATAMENTO DOS EVENTOS DE CAUDA

No caso bidimensional, os eventos de cauda são calculados utilizando-se uma extrapolação aplicada a uma projeção da PDF estimada passando pelo seu ponto de maior valor. Como ilustrado pela Figura 51. Os valores periféricos desta projeção são então usados em um ajuste exponencial utilizado para conferir um valor de probabilidade ao evento em questão. Entretanto, a realização dessa operação evento-a-evento se mostrou inviável computacionalmente, uma vez que uma grande quantidade dos eventos caem na região de cauda. No caso dos dados utilizados nessa dissertação, seriam aproximadamente 10^5 eventos.

Portanto, foi necessário a utilizar uma repartição por setores, onde cada setor teria

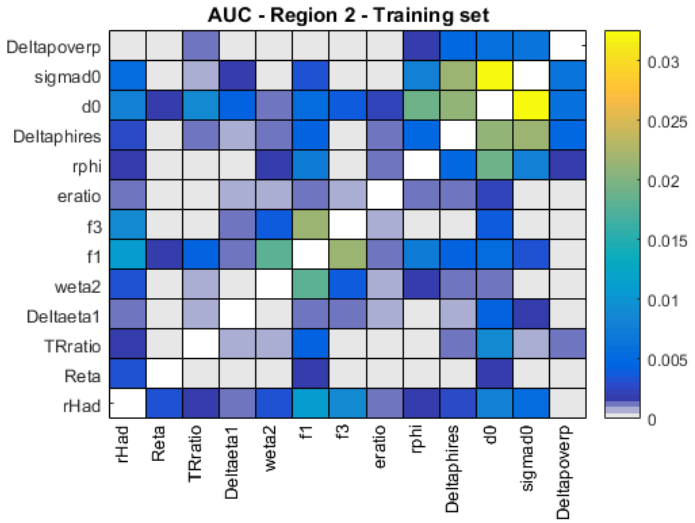

Figura 49: Gráfico de AUC com a combinação dos pares de variáveis, para a Região 2, utilizando somente os eventos de CC.

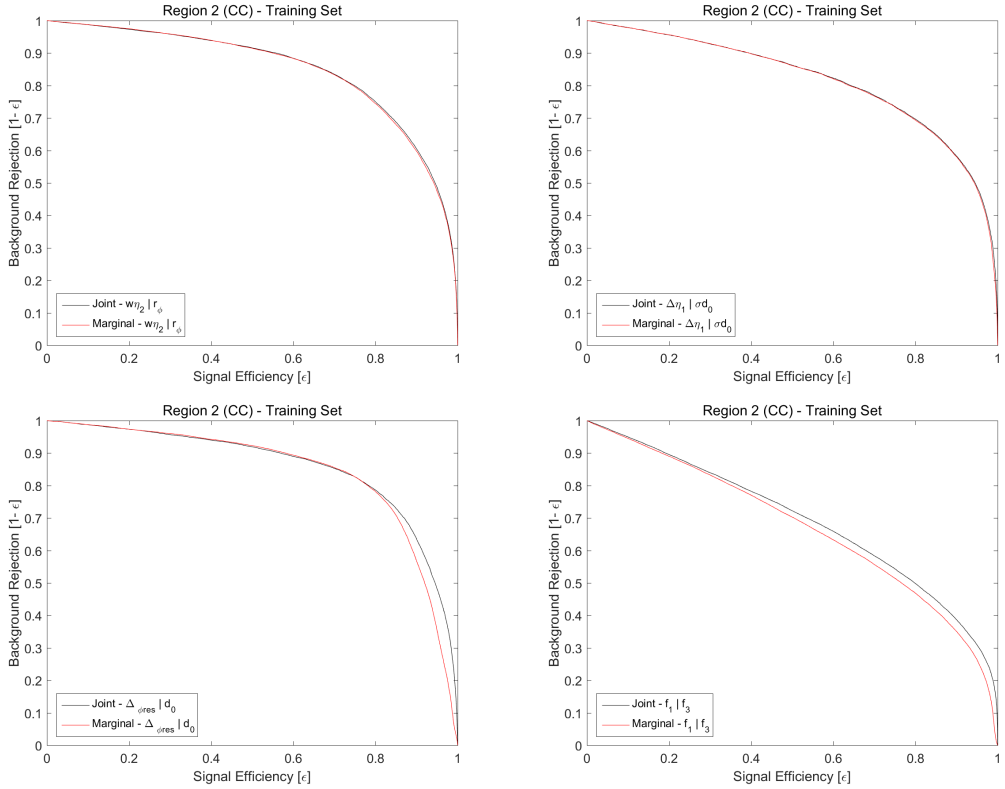

Figura 50: Gráfico comparando as ROC's as análises conjunta e marginal, para os eventos de CC da Região 2. (Superior Esquerda) Par de variáveis - w_{η_2} e r_ϕ; (Superior Direita) Par de variáveis - $\Delta_{\eta 1}$ e σ_{d_0}; (Inferior Esquerda) Par de variáveis - $\Delta_{\phi res}$ e d_0 e (Inferior Direita) Par de variáveis - f_1 e f_3.

uma única projeção a ser usada para todos os eventos que a ela pertencem. Isto foi feito dividindo os 360º de uma densidade bivariada em 2^n setores. A Figura 52 mostra

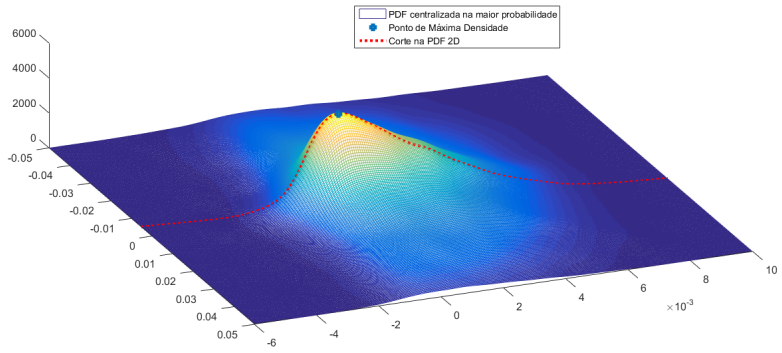

Figura 51: Gráfico exemplificando a extrapolação 2D feita pelo método exponencial (*Fit*).

um exemplo para n=4. Os setores estão indicados por diferentes cores, cada um com uma projeção própria representada pelas linhas vermelhas.

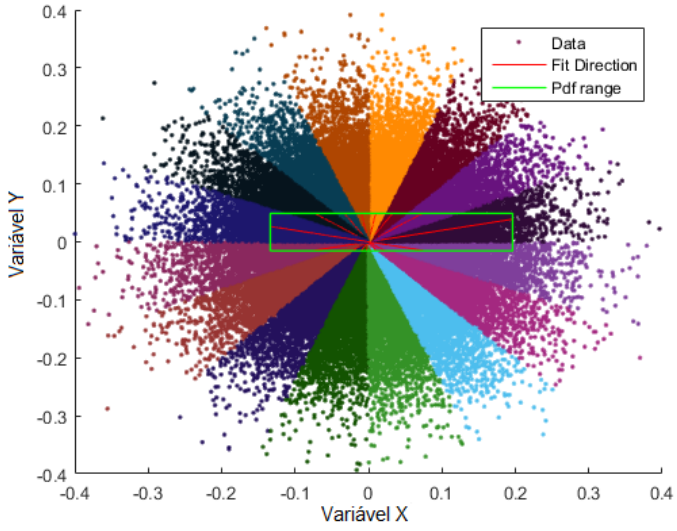

Figura 52: Gráfico exemplificando a setorização da extrapolação 2D feita pelo método exponencial (*Fit*).

Logo, a implementação da extrapolação via ajuste exponencial, no caso bivariado, depende de dois parâmetros que podem ser escolhidos com o objetivo de garantir uma boa performance do algoritmo de classificação de elétrons: o número de setores a ser criado e o maior valor de probabilidade da densidade a ser considerado no processo de ajuste, o qual todos os pontos da densidade discretizada abaixo deste valor serão usados para o processo de ajuste exponencial. Para encontrar os seus valores ótimos, a AUC das ROC criadas a partir das densidades bivariadas utilizando somente os eventos de CT foi analisada. Para tal, os valores dos dois parâmetros citados acima foram variados

para duas combinações de PDF, f_1 - f_3 e d_0 - σ_{d_0}, como mostra a Figura 53.

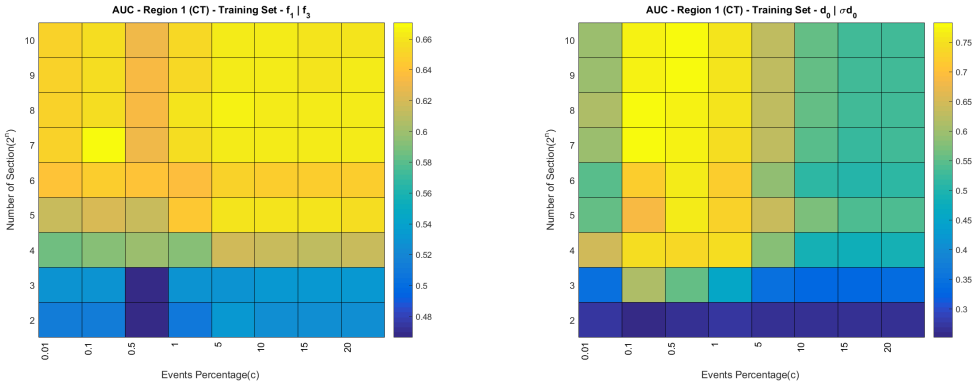

Figura 53: Gráfico de AUC de acordo com a variação dos parâmetros n e c. (Esquerda) AUC da análise multivariada de f_1 e f_3 da Região 1 e (Direita) AUC da análise multivariada de d_0 e σ_{d_0} da Região 1.

Como esperado, em ambas as combinações, os melhores resultados ocorrem quando maximizamos o parâmetro n, ou seja, aumentamos a quantidade de setores. Portanto, com os resultados demonstrados na Figura 53, foi decidido utilizar $n = 9$ por ter apresentado uma boa performance e um custo computacional aceitável. Já o parâmetro c, que é a porcentagem de eventos que o algoritmo de extrapolação deve observar para fazer o ajuste da exponencial, apresentou variações significativas para as duas densidades bidimensionais aqui consideradas. Sendo assim, optou-se por utilizar um c encontrado de forma empírica para cada conjunto de variáveis escolhidas para a abordagem multivariada.

Depois de escolhidos os parâmetros do ajuste exponencial, de forma análoga à otimização univariada, foi feita uma comparação entre esse método, com os parâmetros otimizados, e a abordagem de extrapolação feita pelo método do vizinho mais próximo. A Figura 54 mostra a comparação desses dois métodos para a Região 1, utilizando somente eventos de CT, para as variáveis combinadas f_1 e f_3 e d_0 e σ_{d_0}. Observa-se uma diferença de performance significativa entre os dois métodos. A melhora de performance no uso do método exponencial para a análise multivariada aparenta ser maior ainda que a melhora vista na análise univariada. Com isso, é feita a escolha de utilizar essa abordagem para o caso multivariado.

Nesse capítulo foram apresentados os principais testes de definições relacionados ao algoritmo de classificação de elétrons via método de verossimilhança com abordagens univariada e multivariada desenvolvidos nessa dissertação, bem como as otimizações realizadas nele no intuito de se obter o melhor desempenho na classificação dos eventos. No Capítulo 6 serão apresentados os resultados obtidos com a aplicação desses métodos

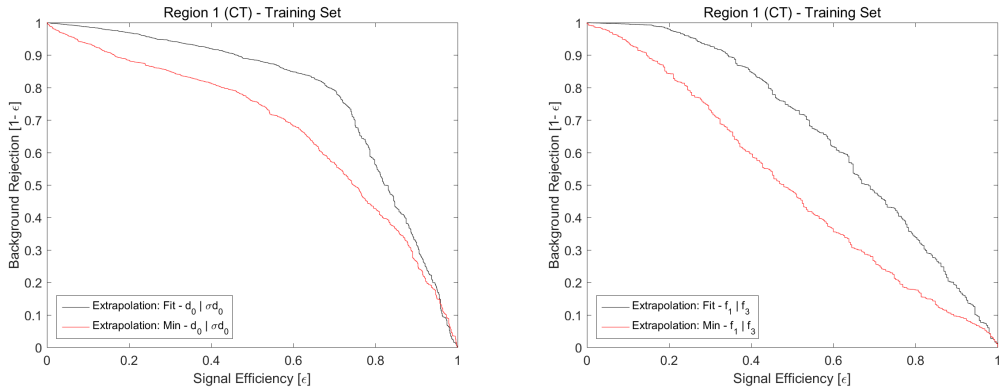

Figura 54: Gráfico comparando as ROC's da estimação de densidade de um par de variáveis, utilizando a extrapolação feita pelo método exponencial (*Fit*) e pelo método de vizinho mais próximo (*Min*) na Região 1, utilizando somente os eventos de CT. (Esquerda) ROC das variáveis f_1 e f_3 de forma conjunta; (Direita) ROC das variáveis d_0 e σ_{d_0} de forma conjunta.

nos dados de validação e as comparações entre o métodos baseado na estimação de densidades univariado e e/γ; e os métodos de estimação de densidades univariado e bivariado.

6 RESULTADOS

Nesse capítulo, serão apresentados os resultados obtidos utilizando o conjunto de validação dos dados de Simulação de Monte Carlo, para as análises univariada e multivariada. Na Seção 6.1, o algoritmo baseado nas densidades univariadas será avaliado e comparado com os pontos de operação do e\γ. Na Seção 6.2, a verossimilhança que utiliza de densidades bivariadas será avaliada e comparada com a verossimilhança univariada, ambas utilizando as otimizações propostas no Capítulo 5.

6.1 ANÁLISE UNIVARIADA

Nessa seção, serão apresentados os resultados do algoritmo de identificação univariada de elétrons pelo método de verossimilhança desenvolvido nessa dissertação e a comparação com o algoritmo de identificação do experimento ATLAS, e\γ. Como citado, na Seção 3.2, as variáveis utilizadas apresentam dependência em η e E_t, e ainda podem apresentar dependência por empilhamento (*Pile-up*), sendo representada, nessa dissertação, pela sigla N_{vtx} (do inglês, *Number of Reconstructed Primary Vertices in the Event*); portanto, serão apresentados resultados de desempenho em função de η, E_t e N_{vtx}, no intuito de comparar a robustez de cada um sobre esses aspectos.

Como forma de comparação, os resultados obtidos com a *Likelihood* (LH) serão apresentados junto aos três pontos de operação do e\$\gamma$ (*Tight*, *Medium* e *Loose*), sempre que possível. Na Figura 55, são mostradas as curvas ROC da LH e a mesma utilizando cortes adicionais, que fazem uso das variáveis de traço para aumentar a rejeição de ruído de fundo, como explicado na Seção 3.2.2, bem como os três pontos de operação do e\$\gamma$, para as Regiões 1 e 2 da Tabela 3. Na Figura 56 são apresentados os resultados para as Regiões 11 e 12.

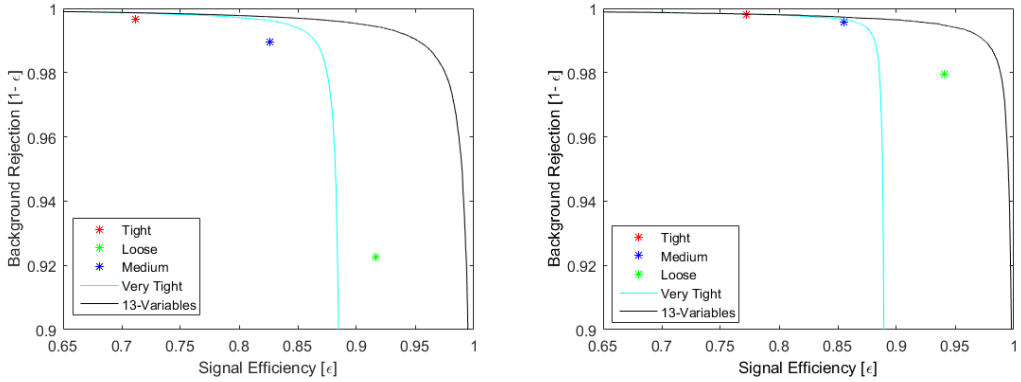

Figura 55: Curva ROC, comparação da *Likelihood* com os pontos de operação *Tight*, *Medium* e *Loose* do e\γ. (Esquerda) Região 1; (Direita) Região 2.

Como mostrado nas Figuras 55 e 56, o aumento da rejeição de ruído de fundo com uso dos cortes adicionais não ocorre em todas as regiões, como é o caso das regiões 1 e 2, mostradas na Figura 55. Pelos resultados obtidos, infere-se que nos casos onde se têm mais estatística e consequentemente PDF melhores representadas, o adição dos *hard cuts* pode ser facultativa, mas em casos onde não é possível ter estatística suficiente, esses cortes podem aumentar de maneira significativa a rejeição de ruído de fundo.

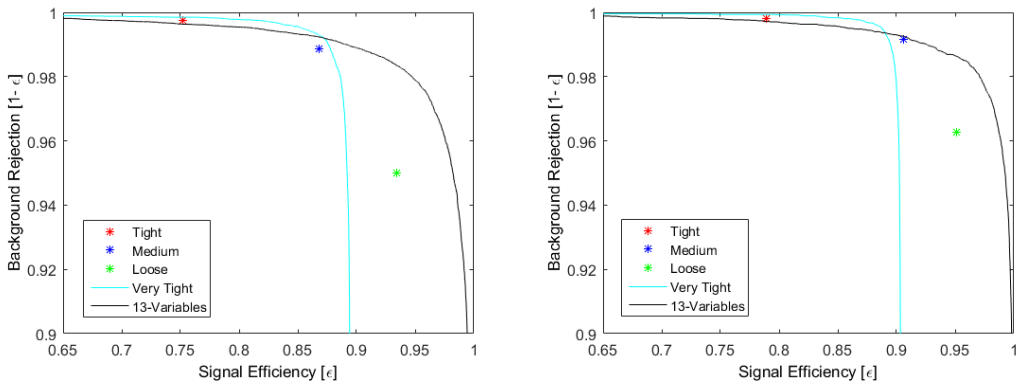

Figura 56: Curva ROC, comparação da *Likelihood* com os pontos de operação *Tight*, *Medium* e *Loose* do e\γ. (Esquerda) Região 11; (Direita) Região 12.

Como forma de comparar os resultados obtidos pelo algoritmo desenvolvido nessa dissertação com o e\γ, fez-se um ajuste dos pontos de operação. Utilizando os dados de treinamento, o algoritmo de Verossimilhança foi ajustado para ter a mesma eficiência de sinal que os pontos de operação *Tight*, *Medium* e *Loose* do e\γ, sendo que esses pontos foram denominados de *Very TightLH*, *MediumLH* e *LooseLH*, respectivamente.

As Figuras 57, 58 e 59 mostram a eficiência e falso alarme em função η, E_t e N_{vtx},

respectivamente, para a região de $0 \leq |\eta| < 2.47$ e $20 \leq E_t < 50GeV$. Sendo que na Figura 57, o gráfico da esquerda mostra as eficiências de sinal por η para os 3 pontos do e\γ e os 3 pontos da LH, pode-se observar que o ajuste realizado é coerente e que os recíprocos pontos de operação estão aproximadamente com a mesma eficiência de sinal. E no gráfico da direita, que mostra o falso alarme em função de η, podemos ver que o método de Verossimilhança, para a mesma eficiência de sinal, alcança melhores valores de falso alarme. Sendo que o maior ganho se encontra no ponto de operação *LooseLH*.

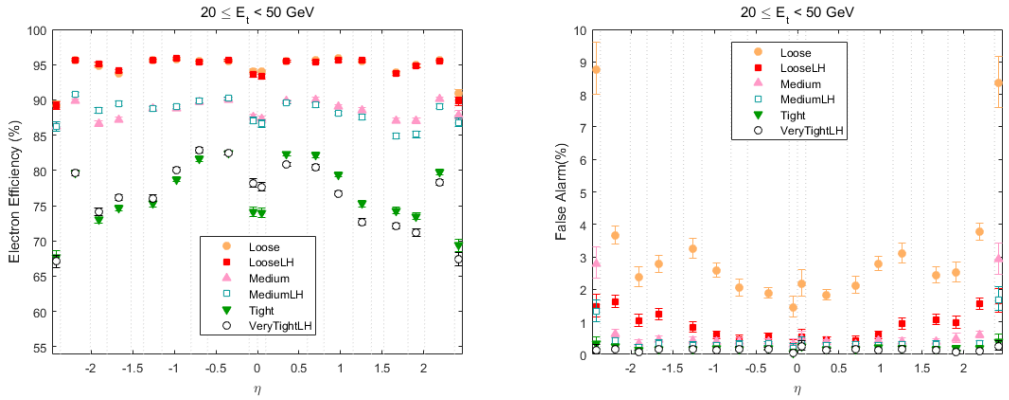

Figura 57: Gráfico de η, comparando a LH e o e\γ. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

Assim como mostrado para η, a Figura 58 apresenta a eficiência de sinal em função de E_t, no gráfico da esquerda podemos perceber o ajuste feito para os pontos de *Likelihood* e e\γ através da quase igualdade dos valores de eficiência de sinal, e na direita o ganho na medida de falso alarme do método descrito nessa dissertação.

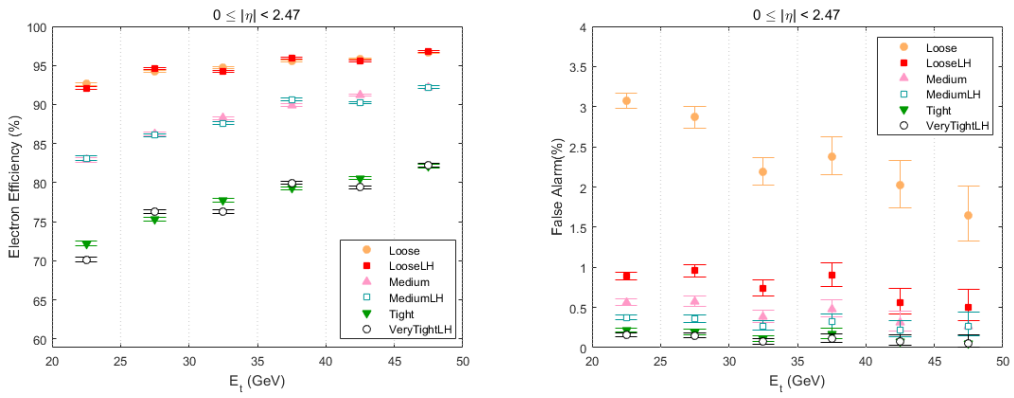

Figura 58: Gráfico de E_t, comparando a LH e o e\γ. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

A Figura 59 mostra a dependência de empilhamento dos dados utilizados, e confirma que o ajuste de eficiência dos pontos de operação está correto, bem como a melhora no falso alarme quando utilizada a LH.

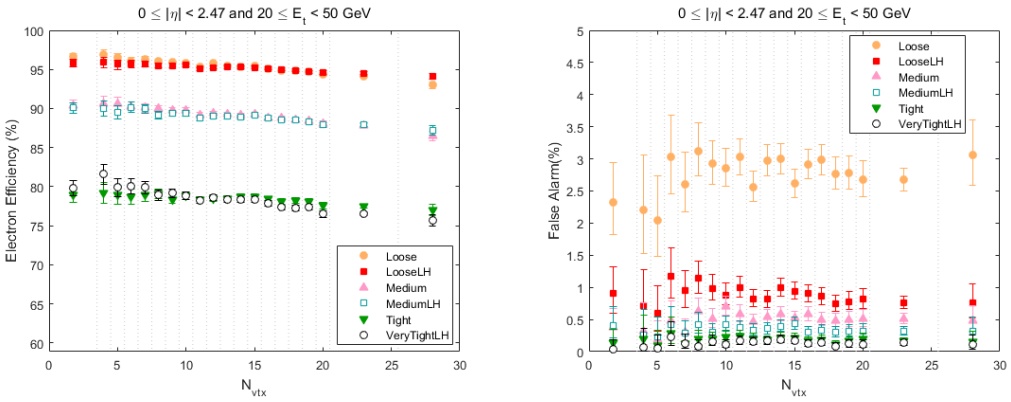

Figura 59: Gráfico de N_{vtx}, comparando a LH e o e\gamma. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

A Tabela 5 mostra uma comparação entre os pontos de operação da *Likelihood* e o e\gamma fixando a eficiência de sinal. Pode-se ver que usando método da *Likelihood*, os três pontos de operação conseguem um falso alarme melhor mantendo uma eficiência de sinal semelhante aos pontos de operação e\gamma.

Tabela 5: Eficiência de Sinal e Rejeição de Ruído de Fundo para a *Likelihood* e o e\gamma, para $0 \leq |\eta| < 2.47$ e $20 \leq E_t < 50GeV$, fixando a Eficiência de Sinal.

$0 \leq \|\eta\| < 2.47$ e $20 \leq E_t < 50GeV$						
Menu	Eficiência de Sinal			Falso Alarme		
Loose Cuts	95,10	+0,05	-0,05	2,81	+0,06	-0,06
Medium Cuts	88,87	+0,07	-0,07	0,53	+0,03	-0,03
Tight Cuts	78,29	+0,10	-0,10	0,19	+0,02	-0,02
Loose LH	95,05	+0,05	-0,05	0,87	+0,04	-0,04
Medium LH	88,70	+0,07	-0,07	0,35	+0,02	-0,02
VeryTight LH	77,85	+0,10	-0,10	0,14	+0,02	-0,01

Cabe ressaltar que comparando os pontos de operação *Loose* e *LooseLH*, temos que o falso alarme do último é aproximadamente 3 vezes menor que o primeiro. Entre os pontos *Medium* e *MediumLH*, a melhora é um pouco menor, aproximadamente 1,5 vezes. Já a melhora entre os pontos *Tight* e *VeryTightLH*, é de aproximadamente 1,3 vezes.

As figuras acima mostraram o desempenho do algoritmo para uma região específica

de energia. No intuito de apresentar resultados mais gerais, as Figuras 60, 61 e 62 mostram a eficiência e falso alarme em função de η, E_t e N_{vtx}, respectivamente, para a região de $0 \leq |\eta| < 2.47$ e $5 \leq E_t < 100GeV$.

A Figura 60 mostra que existe um pequeno aumento do falso alarme em relação a Figura 57 tanto para os pontos do e\γ quanto para os pontos da LH. Isso se deve ao fato de que, em energias mais baixas, a classificação de elétrons é mais difícil. Entretanto, podemos reparar que o aumento de falso alarme é menor para os pontos de LH, o que indica uma robustez maior desse método em relação ao algoritmo baseado em cortes.

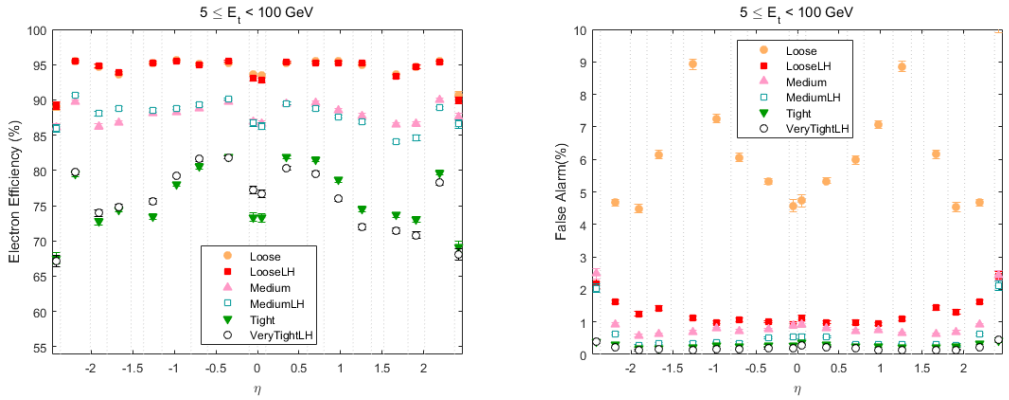

Figura 60: Gráfico de η, comparando a LH e o e\γ. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

Outro ponto a ser comentado é a queda de eficiência do algoritmo para eventos com energia menor que $20GeV$, como mostrado na Figura 61. No entanto, a melhora de falso alarme na utilização da LH continua notória.

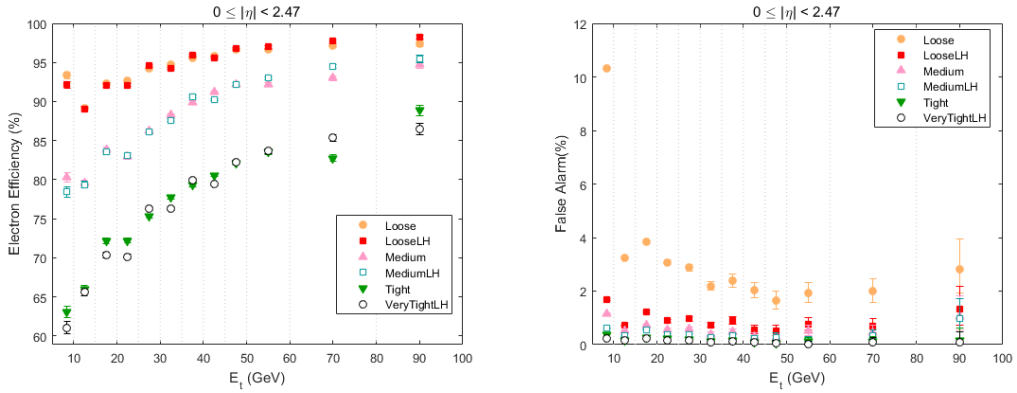

Figura 61: Gráfico de E_t, comparando a LH e o e\γ. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

A Figura 62, que utiliza mais estatística do que a Figura 59, por abranger uma

faixa maior de energia, deixa claro a dependência de eficiência de sinal e falso alarme ao empilhamento. Nota-se uma tendência de queda de eficiência e um leve aumento de falso alarme, à medida que o N_{vtx} aumenta.

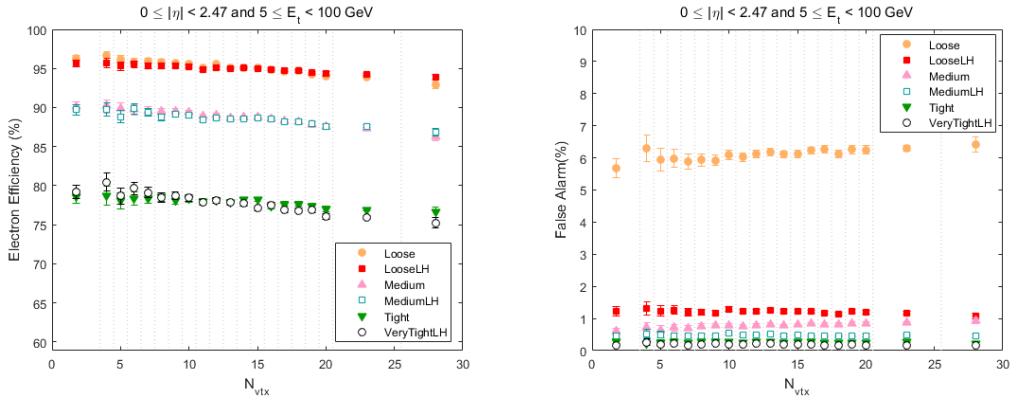

Figura 62: Gráfico de N_{vtx}, comparando a LH e o e\γ. (Esquerda) Eficiência de sinal; (Direita) Falso Alarme.

A Tabela 6 mostra a comparação entre os pontos de operação da *Likelihood* e o e\γ fixando a Eficiência de Sinal. Como se trata de um resultado analisando toda a região de η e uma grande faixa de E_t, podemos ver que a eficiência média dos métodos é menor em relação a quando se faz esta medida para uma área específica de energia onde é mais fácil a separação de elétrons do ruído de fundo.

Tabela 6: Eficiência de Sinal e Rejeição de Ruído de Fundo para a *Likelihood* e o e\γ, para $0 \leq |\eta| < 2.47$ e $5 \leq E_t < 100GeV$, fixando a Eficiência de Sinal.

$0 \leq \|\eta\| < 2.47$ e $5 \leq E_t < 100GeV$						
Menu	Eficiência de Sinal			Falso Alarme		
Loose Cuts	94,82	+0,05	-0,05	6,16	+0,03	-0,03
Medium Cuts	88,43	+0,07	-0,07	0,81	+0,01	-0,01
Tight Cuts	77,72	+0,08	-0,08	0,27	+0,01	-0,01
Loose LH	94,81	+0,05	-0,05	1,20	+0,01	-0,01
Medium LH	88,35	+0,07	-0,07	0,48	+0,01	-0,01
VeryTight LH	77,29	+0,09	-0,09	0,19	+0,01	-0,01

Podemos analisar que, para essa região, o fator de melhora de falso alarme dos pontos de operação *LooseLH*, *MediumLH* e *VeryTightLH* em comparação aos pontos *Loose*, *Medium* e *Tight*, é de aproximadamente 5, 1,7 e 1,4, respectivamente.

Com esses resultados, podemos concluir que o algoritmo de classificação de elétrons pelo método de verossimilhança implementado nesse trabalho apresenta grandes

melhorias em relação ao algoritmo e\γ utilizado pelo experimento ATLAS.

6.2 ANÁLISE MULTIVARIADA

Nessa seção, serão apresentados os resultados da implementação do algoritmo multivariado de identificação de elétrons pelo método de verossimilhança e a comparação com o algoritmo univariado, ambos desenvolvidos nessa dissertação.

Como mostrado na Seção 6.1, o método baseado em estimação de densidades univariadas apresentou desempenho superior ao algoritmo e\γ; portanto, nesta seção, será usado apenas o algoritmo de verossimilhança univariada para a comparação com o método baseado em estimação de densidades multivariadas.

A Figura 63 mostra a comparação, via curva ROC, entre a utilização da análise Univariada e Bivariada para os dados de validação da região 1. Observando esses resultados, temos uma indicativa que o método de análise multivariada pode apresentar o ganho de performance esperado, uma vez que este método diminui o impacto dos eventuais erros causados pela consideração de independência das variáveis discriminantes. Os resultados mostram um desempenho superior da Likelihood Conjunta (*Joint*), na maioria dos casos, quando utilizado somente o par escolhido em questão.

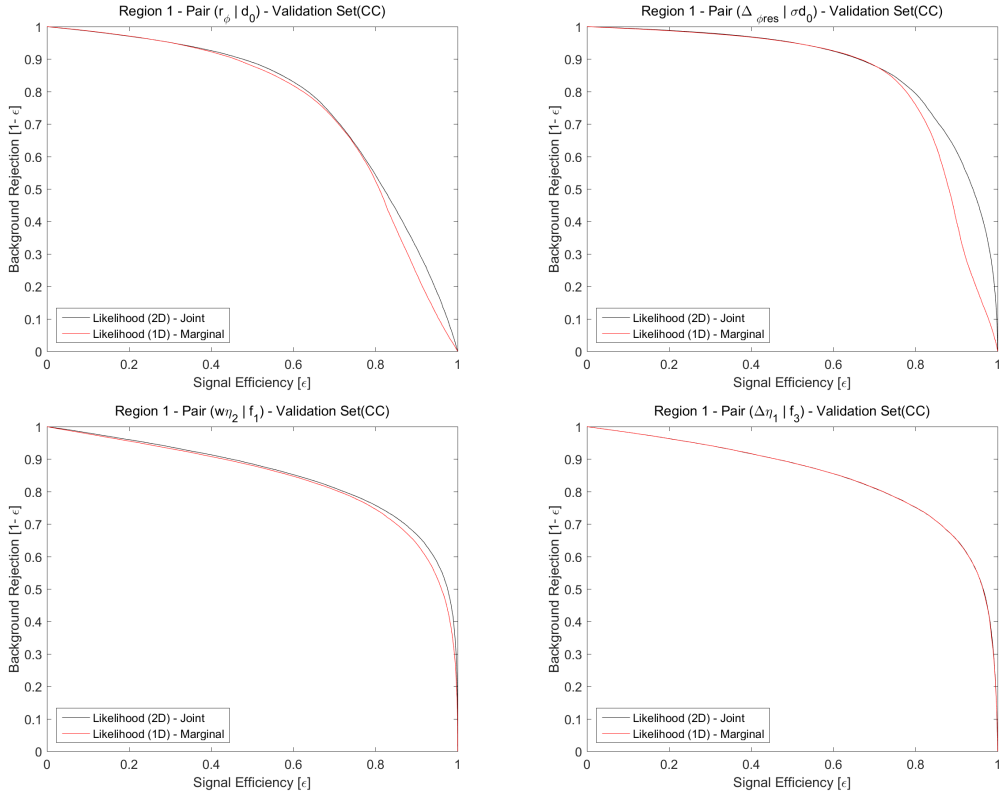

Figura 63: Curva ROC da Região 1 comparando a análise Univariada com a Bivariada para os eventos de CC.(Superior Esquerda) Par de variáveis - r_ϕ e d_0; (Superior Direita) Par de variáveis - $\Delta_{\phi res}$ e σ_{d_0}; (Inferior Esquerda) Par de variáveis - $w_{\eta 2}$ e f_1 e (Inferior Direita) Par de variáveis - $\Delta_{\eta 1}$ e f_3.

Este mesmo teste foi feito para a região 2, e são mostrados na Figura 64. Como pode-se observar, um resultado similar é obtido.

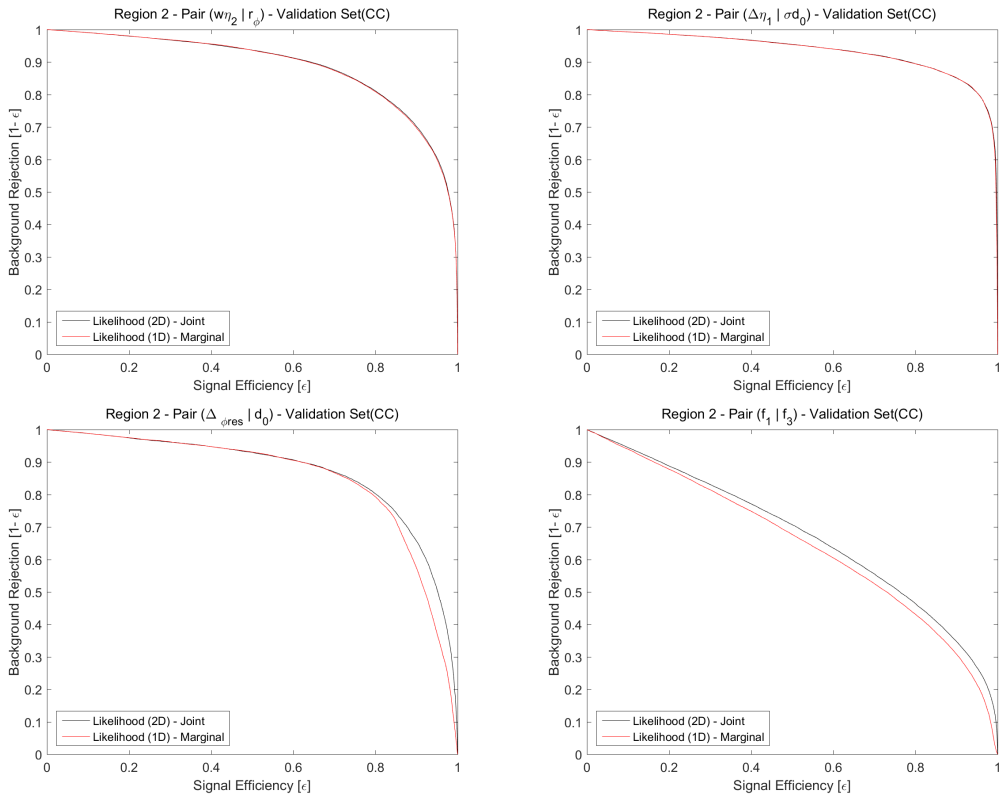

Figura 64: Curva ROC da Região 2 comparando a análise Univariada com a Bivariada para os eventos de CC. (Superior Esquerda) Par de variáveis - w_{η_2} e r_ϕ; (Superior Direita) Par de variáveis - $\Delta_{\eta 1}$ e σ_{d_0}; (Inferior Esquerda) Par de variáveis - $\Delta_{\phi res}$ e d_0 e (Inferior Direita) Par de variáveis - f_1 e f_3.

Estes resultados nos indicam que o uso de PDF bidimensionais é um possível caminho para o aumento do desempenho do algoritmo de identificação de elétrons. Sabendo do problema da 'Maldição da dimensionalidade', descrito na Seção 4.2.1, a metodologia proposta foi de fazer PDF bidimensionais somente para os casos que apresentem uma melhor significativa em relação as PDF univariadas.

A Figura 65 mostra a comparação entre a Verossimilhança Univariada e a Verossimilhança Bivariada utilizando as 4 PDF bidimensionais escolhidas, para os eventos de CC das regiões 1 e 2, gráficos da esquerda e da direita, respectivamente.

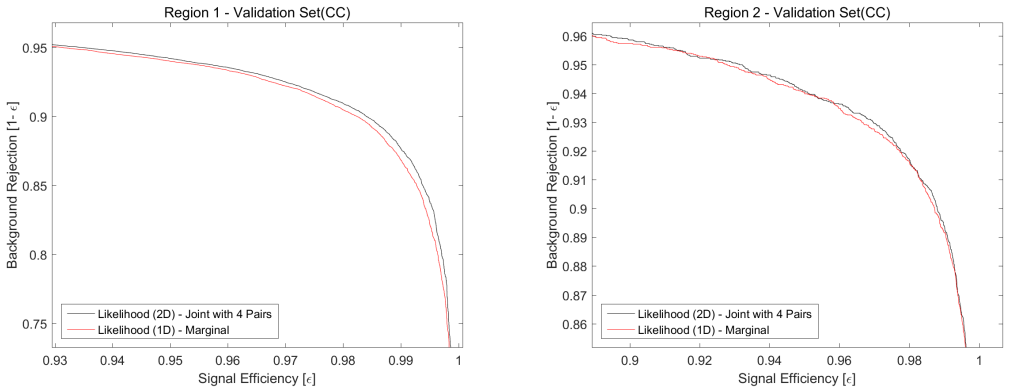

Figura 65: Curva ROC comparando a análise Univariada com a Bivariada, utilizando 4 PDFs Conjuntas, para os eventos de CC. (Esquerda) Região 1 e (Direita) Região 2.

Observa-se um ganho de performance do método bivariado na região 1, que encontra-se em energias abaixo de $20GeV$ onde a eficiência dos algoritmos são inferiores em comparação as outras regiões. Para a região 2, podemos assumir que houve uma melhora sutil, mas real. A região CC é onde se encontra a maior parte dos eventos e estes resultados representam o ganho real do uso das PDFs Bidimensionais pois como explicado na Seção 5.2.1, é nessa região que são utilizadas, de fato, as estimações de densidade bidimensionais feitas pelo algoritmo de KDE.

Depois de apresentado esses resultados utilizando os eventos de CC, a Figura 66 mostra os resultados para todos os eventos, onde entram em ação os algoritmos de tratamento de cauda unidimensionais e bidimensionais descritos nas Seções 5.1.2 e 5.2.2, respectivamente, retirando somente os casos de descontinuidades.

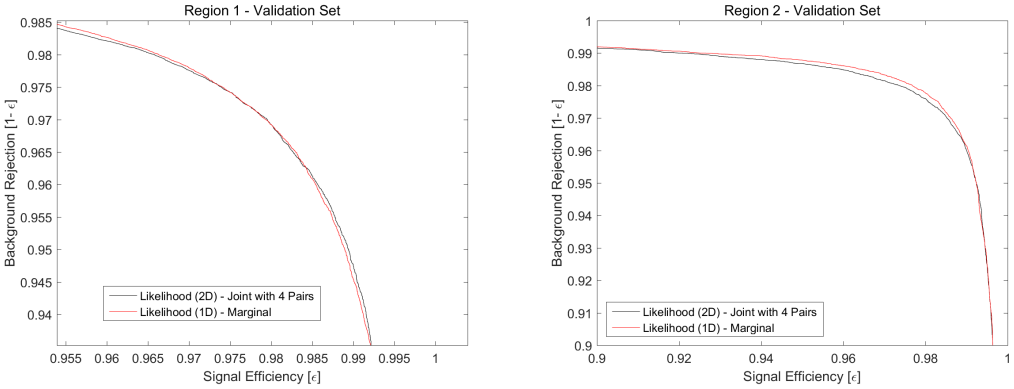

Figura 66: Curva ROC comparando a análise Univariada com a Bivariada, utilizando 4 PDFs Conjuntas, para todos os eventos exceto descontinuidades. (Esquerda) Região 1 e (Direita) Região 2.

Como podemos observar, para a região 1 existe um pequeno ganho na eficiência de sinal e uma pequena perda na reijeição de ruído de fundo com o uso da análise multivariada em comparação com a análise univariada. Entretanto, para a região 2, existe um pequena perda de performance de classificação.

Para completar as análises, a adição dos *Hard Cuts* é feita tanto para *Likelihood* Univariada quanto para a *Bivariada*, com o intuito de utilizar as variáveis de traço para aumentar a rejeição de ruído de fundo, como explicado na Seção 3.2.2.

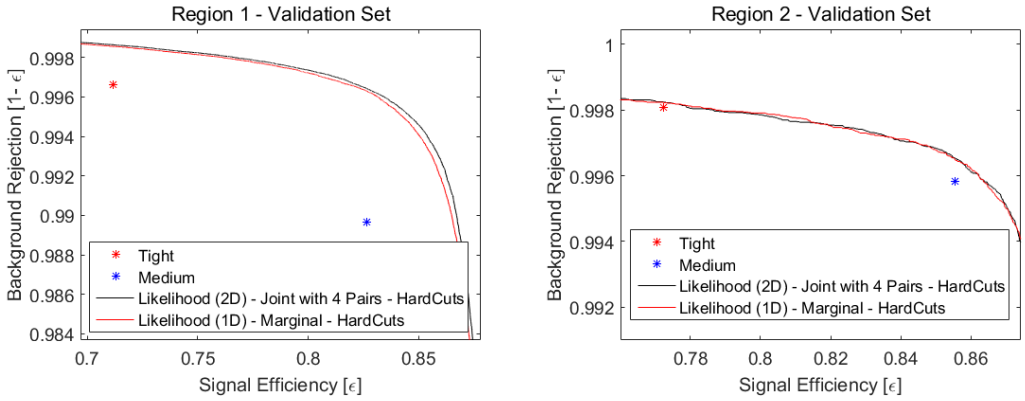

Figura 67: Curva ROC comparando a análise Univariada com a Bivariada, utilizando 4 PDFs Conjuntas, para todos os eventos, inclusive as descontinuidades. (Esquerda) Região 1 e (Direita) Região 2.

A Figura 67 apresenta as *Likelihood* 1D e 2D, com a adição dos cortes adicionais, e os pontos de operação *Tight* e *Medium* do e\γ. Pode-se perceber que existe um ganho de performance na medida de rejeição de ruído de fundo quando utilizamos a *Likelihood 2D*, quando comparada a *Likelihood* 1D.

7 CONCLUSÕES

Esta dissertação apresentou os estudos, implementação e análise relacionados a 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*. A identificação de elétrons é de fundamental importância para o programa de física do ATLAS, e tem impacto direto nos resultados das pesquisas desenvolvidas nesse ambiente. Portanto, tendo em vista a contínua evolução do experimento, o processo de otimização das técnicas de classificação de eventos é crucial.

Praticamente desde o início do Experimento ATLAS a classificação de elétrons é feita por um algoritmo baseado em cortes, conhecido como e\$\gamma$. Essa técnica impõe algumas restrições uma vez que o uso de cortes rígidos dificulta a classificação dos eventos de cauda. Para contornar este problema, um grupo da Colaboração ATLAS propôs a identificação de eventos via método de *Likelihood*, considerando independência entre as variáveis discriminantes. Esta proposta tem apresentado resultados superiores ao e\$\gamma$. Porém a consideração de independência pode causar eventuais erros na reconstrução das densidades conjuntas sendo este um dos pontos principais de investigação desta dissertação.

Esta dissertação mostrou que o algoritmo de classificação de elétrons univariado via método de *Likelihood* apresenta desempenho superior ao método utilizado pela Colaboração; e que o mesmo pode ser melhorado através do estudo dos eventos centrais e de cauda de forma separada, aplicando tratamento especial para cada caso; e que a versão multivariada tem o potencial de melhorar ainda mais o seu desempenho, mostrando que o método proposto pode ser de fato uma ferramenta útil na busca por um aprimoramento do sistema de seleção de eventos do ATLAS.

A *Likelihood* multivariada foi implementada a partir de uma proposta aqui desenvolvida de se utilizar os pares de PDF's bidimensionais que apresentam melhoras significativas de performance quando avaliada a diferença de AUC das ROC entre os métodos uni e multivariado.

Os maiores esforços deste trabalho foram concentrados no estudo aprofundado das teorias de estimação de densidade não-paramétricas, na busca de uma otimização de seus parâmetros, na implementação dos algoritmos de estimação de densidade uni e multivariada, na sua aplicação na seleção de eventos pelo método da *Likelihood*, e na análise de desempenho a partir dos dados simulados do Experimento ATLAS.

O principal ganho dos estudos realizados nessa dissertação foi a ampliação do conhecimento na área de processamento de sinais aplicado à física de altas energias. As análises, aqui apresentadas, possibilitaram reconhecer as dificuldades encontradas quando saímos do campo de análises teóricas e passamos para a aplicação experimental das técnicas estudadas. E, além disso, a identificação de caminhos que podem gerar melhorias dos métodos propostos neste trabalho.

7.1 PRÓXIMOS PASSOS

Ao final desse estudo, fica claro que existem possibilidades de melhorias nos métodos apresentados, tanto na análise univariada, quanto na multivariada, sendo que alguns desses pontos estão listados abaixo:

- Estudo de uma otimização robusta do KDE N-Dimensional com a aplicação de uma versão de banda variável;
- Utilização da mesma segmentação em η e E_t utilizada pelo ATLAS;
- Estudar possíveis otimizações na técnica de extrapolação bidimensional;
- Análise mais profunda do tratamento dos eventos de cauda;
- Aplicar os algoritmos mostrados em dados com pileup (ou empilhamento) maiores do que os apresentados neste trabalho.
- Aplicar essas técnicas estudadas nos dados reais do Experimento ATLAS.

REFERÊNCIAS

AAD, G. et al. Electron reconstruction and identification efficiency measurements with the atlas detector using the 2011 lhc proton–proton collision data. *The European Physical Journal C*, Springer, v. 74, n. 7, p. 1–38, 2014.

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

ABAT, E. et al. Expected performance of the atlas experiment-detector, trigger and physics. *arXiv preprint arXiv:0901.0512*, 2008.

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

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.

ARAUJO, T. *O que são os aceleradores de partículas?* 2015. Accessed: 2015-12-22.

BARBERIS, D. Atlas inner detector developments. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Elsevier, v. 446, n. 1, p. 331–337, 2000.

BENEKOS, N. C. et al. *ATLAS inner detector performance*. [S.l.], 2003.

BENGIO, Y. *Research*. 2016. Accessed: 2016-01-26.

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

CERN. *Physics*. 2015. Accessed: 2015-12-21.

CERN. *The accelerator complex*. 2016. Disponível em: <http://home.cern/about/accelerators>.

CERN. *ATLAS*. 2016. Disponível em: <http://home.cern/about/experiments/atlas>.

CHEN, H.; MEER, P. Robust computer vision through kernel density estimation. In: *Computer Vision—ECCV 2002*. [S.l.]: Springer, 2002. p. 236–250.

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

COLLABORATION, A. et al. Electron efficiency measurements with the atlas detector using the 2012 lhc proton-proton collision data. In: ATLAS-CONF-2014-032. [S.l.], 2014.

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.

DAS, A.; FERBEL, T.; GLASHAUSSER, C. *Introduction to nuclear and particle physics*. [S.l.]: Wiley, 1994.

ELSING, M.; SCHÖRNER-SADENIUS, T. Configuration of the atlas trigger system. *arXiv preprint physics/0306046*, 2003.

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.

GABALDON, C. Performance of the atlas trigger system. *Journal of Instrumentation*, IOP Publishing, v. 7, n. 01, p. C01092, 2012.

GRUPEN, C.; SHWARTZ, B. *Particle detectors*. [S.l.]: Cambridge university press, 2008.

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

JONES, M. The roles of ise and mise in density estimation. *Statistics & Probability Letters*, Elsevier, v. 12, n. 1, p. 51–56, 1991.

LEDL, T. *Kernel Density Estimation: Theory, Aspects of Dimension and Application in Discriminant Analysis*. 2002.

LEFEVRE, C. *LHC: the guide (English version)*. [S.l.], 2009.

LIPPMANN, C. Particle identification. *Nucl. Instrum. Meth.*, A666, p. 148–172, 2012.

MOREIRA, M. A. O modelo padrao da fisica de partıculas. *Revista brasileira de ensino de fisica*, SciELO Brasil, v. 31, n. 1, p. 1306, 2009.

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

PERALVA, B. S.-M. *Detecçao de Sinais e Estimaçao de Energia para Calorimetria de Altas Energias*. Tese (Doutorado) — Universidade Federal de Juiz de Fora, 2012.

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

RONQUI, C. M. *Modelo Padrão*. 2015. Accessed: 2015-12-22.

SCOTT, D. W. *Multivariate density estimation: theory, practice, and visualization*. [S.l.]: John Wiley & Sons, 2015.

SEATHER, S. The performance of six popular bandwidth selection methods on some real data sets. *COMPUTATIONAL STATISTICS QUARTERLY*, v. 7, p. 225–225, 1992.

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.

THERHAAG, J.; TEAM, T. C. D. Tmva-toolkit for multivariate data analysis. In: AIP PUBLISHING. *INTERNATIONAL CONFERENCE OF COMPUTATIONAL METHODS IN SCIENCES AND ENGINEERING 2009:(ICCMSE 2009)*. [S.l.], 2012. v. 1504, n. 1, p. 1013–1016.

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.