

Calibração de Energia para a Seleção Online de Elétrons com Alta Taxa e Utilizando um Conjunto de Árvores de Decisão com Reforço por Gradiente e Informação Especialista de Calorimetria

Arthur S. O. Alves¹, Paulo R. A. da Silva¹, Eduardo F. Simas Filho¹, Paulo C. M. A. Farias¹, Juan L. Marin², José M. Seixas³, Edmar E. P. Souza⁴ e Bertrand Laforge⁴

¹Laboratório de Sistemas Digitais - PPGEE / Politécnica - Universidade Federal da Bahia

²Coordenação de Eletrônica - Instituto Federal da Bahia - Campus Vitória da Conquista

³Lab. de Processamento de Sinais - PEE-Coppe / Poli, Universidade Federal do Rio de Janeiro

⁴LPNHE - Sorbonne Université, Paris, França

{arthursoa, araujop, eduardo.simas, paulo.farias}@ufba.br
juan.marin@ifba.edu.br, seixas@lps.ufrj.br
{edmar.egidio, bertrand.laforge}@cern.ch

Resumo—Em física experimental de altas energias, é preciso lidar com um grande volume de informações, sendo grande parte delas proveniente do ruído de fundo que dificulta a caracterização dos fenômenos de interesse particular de um dado experimento. Deste modo, é necessário um complexo processo de seleção online de eventos (trigger). No ATLAS, maior experimento do LHC (Large Hadron Collider), o sistema de trigger opera em duas etapas de seleção sequenciais, denominadas primeiro e alto nível. No caso de elétrons, importantes como mensageiros da nova física que se deseja observar, o sistema de trigger se apoia fortemente no sistema de calorimetria, que mede a energia da partícula incidente. Neste trabalho, é proposto um método de calibração de energia baseado em um conjunto de árvores de decisão com reforço por gradiente (*Gradient Boosted Decision Trees Ensemble* - GBDTE) para melhorar a acuidade da estimativa da energia na etapa rápida do trigger de alto nível do experimento ATLAS. Com esse método proposto, é possível reduzir os requisitos computacionais e aumentar a eficiência na seleção de partículas eletromagnéticas, como elétrons.

Index Terms—Experimento ATLAS, calibração de energia, *gradient boosting*, *decision trees*, *filtragem online*.

I. INTRODUÇÃO

Em experimentos de física de partículas, aceleradores desempenham um papel crucial ao possibilitar colisões entre feixes de partículas, permitindo a detecção e medição de propriedades fundamentais da matéria [1], [2]. Durante as colisões, uma quantidade elevada da informação produzida já é conhecida, e é tratada como ruído de fundo, em função do tipo de partícula ou interação que se deseja investigar. Os sistemas de filtragem *online* são projetados para preservar a informação dos eventos raros e de interesse e rejeitar o ruído de fundo, ainda que em algumas situações as assinaturas das diferentes classes sejam semelhantes. A fim de lidar com essa dificuldade, técnicas de aprendizado de máquina têm sido empregadas em várias aplicações relacionadas à detecção de partículas [3], [4], reconstrução de eventos [5] e

filtragem *online* [6]. Essas técnicas permitem que os sistemas de classificação identifiquem padrões e características relevantes nos sinais coletados, melhorando assim a capacidade de separar os eventos de interesse do ruído de fundo.

O Grande Colisionador de Hádrons (*Large Hadron Collider* - LHC) é um dos experimentos em atividade atualmente na área da física de partículas. Localizado na fronteira franco-suíça, perto de Genebra, o LHC acelera feixes de prótons em sentidos opostos ao longo de um túnel circular com cerca de 27 km de circunferência, a uma profundidade média de 100 m [7]. No LHC, colisões entre os feixes de prótons ocorrem em cada um dos quatro pontos de cruzamento de feixes, a uma taxa de 40 MHz, o que corresponde a uma colisão a cada 25 ns [8].

O detector ATLAS, posicionado em um dos pontos de colisão do LHC, é composto por vários sub-detectores e gera cerca de 1,3 MB de informação a cada colisão [9]. Considerando a taxa de colisão nominal do LHC, é produzido no ATLAS aproximadamente 52 TB/s, sendo necessário um sistema de seleção em online, conhecido como *trigger*, para escolher os eventos relevantes que são armazenados para análises posteriores [10].

Dentre os seus sub-detectores, os calorímetros (medidores de energia altamente segmentados) são muito importantes para o sistema de *trigger online*, pois possuem uma resposta rápida e apresentam características discriminantes nos perfis de deposição de energia para a detecção dos eventos de interesse [11]. O calorímetro do ATLAS é dividido em calorímetro eletromagnético e calorímetro hadrônico, conforme ilustrado na Figura 1. O calorímetro eletromagnético é composto por uma estrutura sequencial de argônio líquido (LAr - *Liquid Argon*) e placas de chumbo com formato de acordeão, enquanto o calorímetro hadrônico é constituído por telhas (*tiles*) de material plástico cintilante.

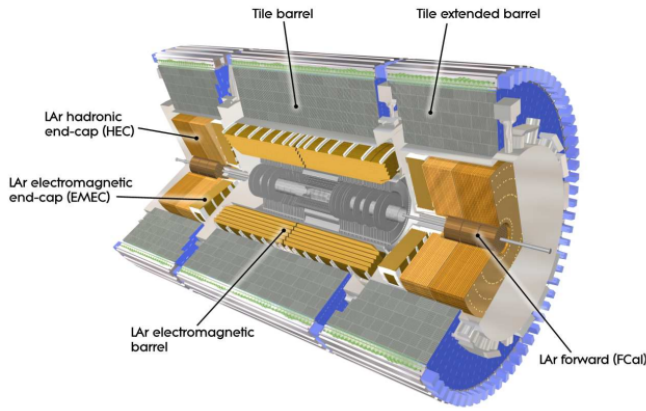


Figura 1: Diagrama ilustrativo do sistema de calorímetros do detector ATLAS [9].

O sistema de seleção *online* de eventos do ATLAS é composto por: primeiro nível (L1), implementado em *hardware* especializado, e alto nível (HLT - *High-Level Trigger*), desenvolvido em *software* e executado em um sistema de computação distribuída. O HLT consiste em uma etapa rápida seguida por uma etapa precisa. Como o sistema de *trigger* depende principalmente das informações dos calorímetros, ele é suscetível a erros de medição de energia. Tais erros podem ocorrer durante o processamento *online*, devido à limitação da resolução do detector. As etapas de calibração, que melhoram a acuidade das medições nos calorímetros, estão atualmente disponíveis apenas no processamento *offline* e na etapa precisa do HLT [12], [13].

Neste trabalho, é proposto um sistema de calibração de energia para operação na etapa rápida do HLT, específico para elétrons. A detecção de elétrons é de extrema importância para o experimento ATLAS, pois a partir dos elétrons pode-se inferir as partículas raras que originaram os decaimentos do Modelo Padrão [2]. Foi utilizado um regressor baseado em um conjunto de árvores de decisão com reforço por gradiente (*gradient boosted decision trees ensemble* - GBDTE), treinado com dados simulados. O conjunto é determinado pela segmentação de energia transversa (E_t) e pseudo-velocidade (η) do calorímetro. Essa abordagem teve seus resultados iniciais apresentados em [14]. Na nova proposta, os anéis concêntricos formados a partir das células do calorímetro foram incluídos como entrada para o GBDTE. Os anéis fornecem informações compactadas e altamente discriminantes para a classificação. Essa abordagem em anéis é utilizada também no NeuralRinger, que é algoritmo para classificação de elétrons operando atualmente na etapa rápida do sistema de *trigger* do ATLAS [15].

Com a implementação do sistema proposto, foi possível melhorar a acuidade na estimação de energia, contribuindo assim para aumentar a eficiência do sistema de detecção *online*. Essa nova abordagem de calibração para elétrons pode ajudar a reduzir a quantidade de eventos não relevantes processados nas cadeias de *trigger* no ATLAS.

II. SELEÇÃO *online* (*trigger*) DE ELÉTRONS

A Fig. 2 ilustra a composição do sistema de *trigger* do ATLAS, que consiste em dois níveis: Nível 1 (L1 - *Level 1*) e HLT (*High-Level Trigger*). O HLT, por sua vez, é dividido em duas etapas: etapa rápida (*Fast*) e etapa precisa (*Precision*).

O Nível 1 opera em *hardware* dedicado para lidar com a alta taxa de informação gerada pelo detector e fornecer uma resposta de aceitação ou rejeição do evento em até $2,5\mu s$. O HLT utiliza um conjunto de computadores para processar em paralelo os eventos aprovados no Nível 1. A etapa rápida, que é a primeira parte do HLT, utiliza apenas as informações da região de interesse (RoI) indicada pelo Nível 1. Os eventos aprovados nessa etapa são processados na etapa precisa, que utiliza todas as informações do detector para tomar a decisão final de aceitação ou rejeição do evento. Durante todo o processo de *trigger*, as informações são armazenadas em um *pipeline* de memórias voláteis, e os eventos aceitos são gravados em mídia permanente sequencialmente.

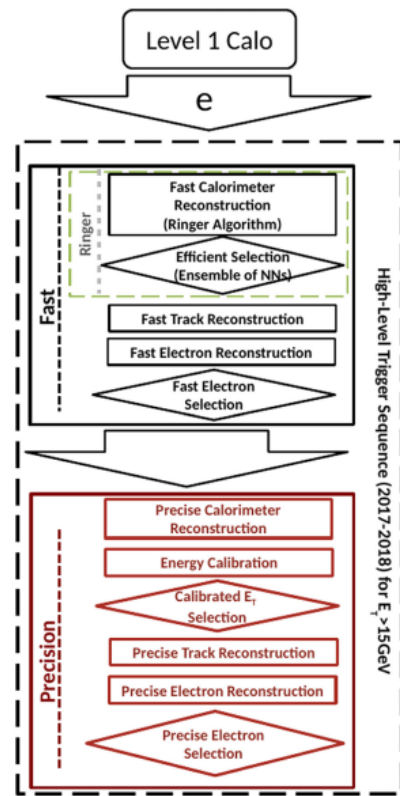


Figura 2: Diagrama do sistema de seleção *online* (*trigger*) de elétrons do detector ATLAS, baseado em [16].

Desde 2017, o algoritmo NeuralRinger [16] [15] funciona no começo da Etapa Rápida do processo de identificação de elétrons [16]. Atualmente, um conjunto de redes neurais

opera para diferentes faixas de energia e posição η ¹ [6]. O NeuralRinger executa dois procedimentos sequenciais: montagem de anéis e classificação do evento. A Fig. 3 descreve a primeira etapa desse processo, os anéis concêntricos são montados centrados na célula mais energética (em vermelho) em cada camada do calorímetro eletromagnético (EM1-EM3) e hadrônico (H1-H3), além do pré-amostrador (PS). Feito isso, as células pertencentes a cada anel são somadas, de tal forma a gerar um vetor que carrega características da propagação da energia em todas as camadas do calorímetro, bem como o espalhamento lateral dessa energia em uma única camada. A quantidade de anéis montados em cada camada do sistema de calorimetria é mostrada na Tabela I.

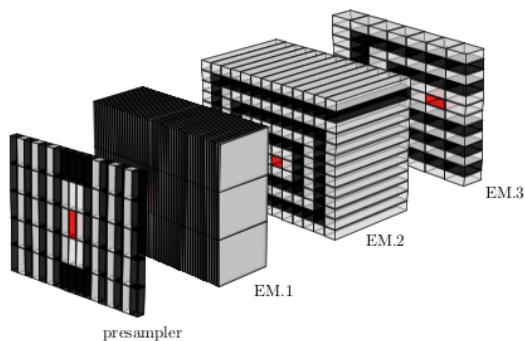


Figura 3: Exemplo da montagem dos anéis para as camadas do calorímetro eletromagnético, pode-se perceber a diferença de granularidade entre as diferentes camadas.

Tabela I: Anéis montados por camada do sistema de calorimetria do experimento ATLAS

Camadas	PS	EM1	EM2	EM3	H1	H2	H3
Anéis	8	64	8	8	4	4	4

O valor de energia de cada um dos 100 anéis gerados é utilizado como entrada para um *ensemble* de redes neurais artificiais [17] (treinadas para atuar em distintas faixas de energia e região do calorímetro). O modelo *ensemble* é responsável por discriminar as assinaturas candidatas a elétron ou jatos hadrônicos (ruído de fundo). O NeuralRinger, explora o formato lateral e longitudinal do chuveiro de partículas, utilizando os anéis para descrever o perfil de deposição de energia no sistema de calorimetria [18]. Uma versão do NeuralRinger utilizando 50 anéis começou a ser utilizada em 2023.

O HLT no ATLAS opera em paralelo a partir de um menu que tem como uma das principais variáveis a energia estimada

¹O ATLAS usa um sistema de coordenadas baseado na mão direita [9], com sua origem no ponto de interação (PI) no centro do detector e o eixo z ao longo do feixe de prótons. O eixo x é direcionado a partir do PI até o centro do anel do LHC e o eixo y é perpendicular a este plano. As coordenadas cilíndricas (r , ϕ) são usadas no plano transversal, com ϕ sendo o ângulo azimutal em torno do eixo z. A pseudo rapidez é definida em termos do ângulo polar θ como $\eta \equiv -\ln \left[\tan\left(\frac{\theta}{2}\right) \right]$.

da partícula medida no calorímetro. Ou seja, os eventos são separados a partir da energia estimada, havendo por exemplo cadeias de *trigger* E15, E26 e E60, nas quais idealmente devem ser selecionados eventos com energia estimada maior que 15, 26 e 60 GeV,² respectivamente. Deste modo, observa-se que a correta estimativa de energia nas etapas de seleção *online* é importante para otimizar as diferentes cadeias de processamento.

III. CALIBRAÇÃO DE ENERGIA USANDO GBDTE

Procedimentos de calibração executados nas etapas do HLT e também na análise *offline* dos dados são essenciais no processo de caracterização das assinaturas energéticas de interesse. A informação medida no calorímetro precisa ser convertida, por meio do fator de calibração, em um valor que seja aproximadamente igual à quantidade de energia perdida pela partícula ao interagir com o material sensor [19].

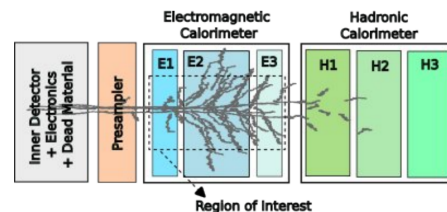


Figura 4: Ilustração das principais causas dos erros na estimativa da energia total da partícula ao interagir com o calorímetro do ATLAS: perda de energia antes (*upstream*) de interagir com o detector; vazamento lateral e longitudinal além da região de interesse.

A energia da partícula eletromagnética é estimada como sendo o somatório das informações das células das camadas EM1, EM2 e EM3 que ficam contidas em uma região de interesse (RoI) ao redor da célula de maior energia (célula mais quente), conforme ilustrado na Figura 4. Portanto, pode ocorrer perda de informação lateral (além da área da RoI), longitudinal (para as camadas hadrônicas) e pela interação da partícula com material do detector antes de chegar ao calorímetro (*upstream*).

De maneira simplificada, a energia estimada é obtida pela equação [14]:

$$E = \alpha_{ADC \rightarrow MeV} \times \alpha_{BDT} \times \alpha_{MC \rightarrow Data} \times \sum_{j=1}^N a_j (s_j - p) \quad (1)$$

onde:

- $\alpha_{ADC \rightarrow MeV}$ é um fator que incorpora as conversões de nível de conversão AD para corrente elétrica e depois de corrente elétrica para MeV. Este fator é determinado a partir de procedimentos experimentais de calibração [19];
- α_{BDT} é o fator de correção fornecido pela GBDTE proposta neste trabalho;

²Elétron-volt (eV) é a quantidade de energia cinética ganha por um único elétron quando acelerado por uma diferença de potencial elétrico de um Volt.

- $\alpha_{MC \rightarrow Data}$ é um fator que ajusta os dados simulados (Monte Carlo - MC) para dados experimentais (*Data*), este fator é necessário pois a BDT é desenvolvida para dados simulados;
- a_j e N são os pesos e a ordem do filtro ótimo utilizado na estimação da energia, a partir dos pulsos produzidos pelo calorímetro [19];
- s_j são as amostras do pulso produzido pelo calorímetro;
- p é o pedestal (linha de base estimada para os pulsos produzidos pelos sensores do calorímetro).

O fator de calibração a ser estimado neste trabalho (α_{BDT}) é calculado utilizando:

$$\alpha_{BDT} = \frac{Et_{\text{truth}}}{Et_{\text{Fast}}} \quad (2)$$

sendo Et_{truth} e Et_{Fast} os valores de energia transversa verdadeira (por Monte Carlo) e estimada na etapa rápida, respectivamente. A partir de agora, para simplificar a notação, α_{BDT} será referido apenas como α .

A Figura 5 ilustra a proposta de inserção de um sistema de calibração para a etapa rápida do HLT. Aqui, a BDT proposta procura trazer para a etapa rápida uma melhoria na estimação de energia similar a da etapa Precisa, atuando localmente na etapa rápida utilizando a informação de calibração online original do calorímetro. Com esse novo bloco de processamento será possível corrigir a estimação de energia na entrada do HLT, aumentando a probabilidade de uma operação otimizada da etapa rápida e, conseqüentemente, uma redução na demanda computacional do HLT.

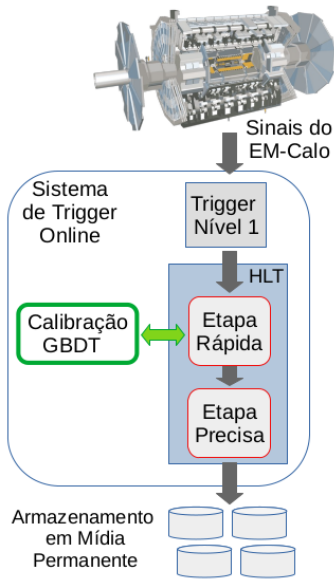


Figura 5: Diagrama indicando o bloco de calibração proposto, mostrando sua integração com o HLT do ATLAS.

Aqui serão descritas as etapas adotadas para estimação de α , na calibração da etapa rápida do HLT para detecção de elétrons utilizando as informações dos anéis (*Rings*). Conforme indicado na Figura 6, a etapa (i) consiste no treinamento

da GBDTE utilizando como entradas as informações do agrupamento (*cluster*) de células e também os anéis. A etapa (ii) realiza o ajuste entre as distribuições de simulação e de dados experimentais para corrigir possíveis discrepâncias advindas de imperfeições na simulação.

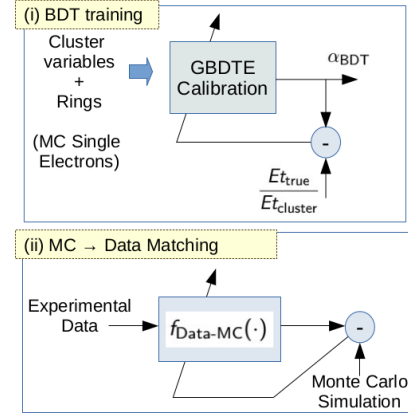


Figura 6: Diagrama das etapas de desenvolvimento do sistema de calibração proposto.

A Figura 7 ilustra a operação do sistema de calibração, incluindo as etapas de casamento de dados simulados com experimentais e o ajuste com a GBDTE.

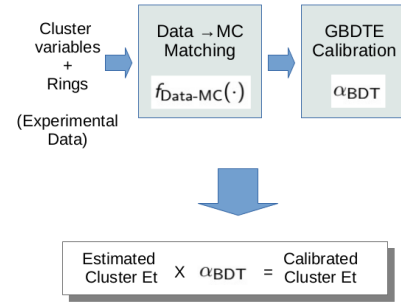


Figura 7: Diagrama das etapas de operação do sistema de calibração proposto.

O perfil de deposição de energia do calorímetro do ATLAS muda em função de E_t , sendo que, em baixas energias é usualmente mais difícil a identificação de elétrons. Também há uma variação em relação à elevação em relação ao feixe do LHC medida a partir da pseudo-rapidez η . Ou seja, $|\eta| = 0$ indica uma posição a 90° em relação ao feixe e $|\eta| \rightarrow \infty$ indica uma região adjacente ao feixe, ou seja fazendo um ângulo com o feixe que tende a 0° . Essa variação se dá devido às diferentes granularidades dos sensores com a variação de $|\eta|$.

Considerando tais características, todo o procedimento de calibração precisa ser projetado para diferentes intervalos de $|\eta|$ e E_t . A partir de estudos prévios [20], a colaboração ATLAS decidiu adotar a segmentação à seguir:

- $|\eta|$ - [0, 0.6, 0.8, 1.2, 1.37, 1.52, 1.8, 2.0, 2.2, 2.5].

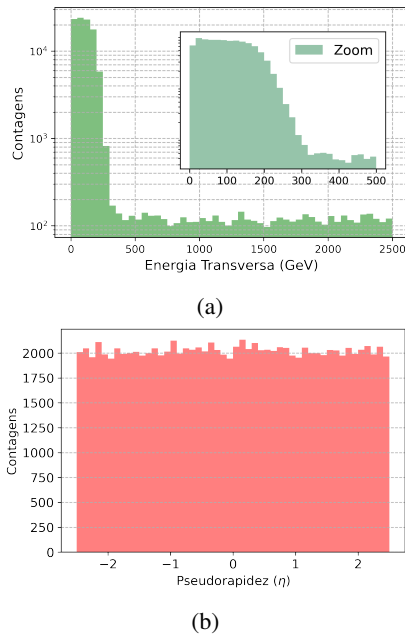


Figura 8: Histogramas característicos em (a) energia transversa e (b) pseudorapidez para os eventos utilizados.

- E_t - $[0, 5, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 900, 2500]$ GeV.

Para a implementação da GBTDE, foi utilizado um conjunto de dados de simulações de Monte Carlo de colisões que produzem elétrons isolados. Foram utilizados aproximadamente 500.000 eventos com distribuição de energia entre 0 e 2500 GeV. Conforme mostrado na Figura 8, os eventos estão mais concentrados em $E_t < 250 \text{ GeV}$ e são distribuídos de modo aproximadamente uniforme em $|\eta| < 2, 5$.

Assim, os fatores α definidos na Eq. 2 (e que representam a relação entre a energia verdadeira e a estimada) precisam ser determinados para cada combinação $|\eta| \times E_t$, o que produzirá $9 \times 13 = 117$ subconjuntos de treinamento distintos. Essa segmentação do α para faixas da pseudorapidez, $|\eta|$, e energia transversal, E_t , podem ser vistas nas figuras 9a e 9b, respectivamente para os dados produzidos por simulações Monte Carlo de colisões no detectores.

É importante notar na Fig. 9 que o desvio do alpha desejado, quando se considera a calibração online original das células do calorímetro. Além disso, tanto o nível de dispersão, como a diferença da mediana em relação a 1 (valor esperado ideal para α) mudam consideravelmente para os subconjuntos analisados. Quando maior essa dispersão, maior o nível de incerteza na estimativa da energia, em comparação com o valor verdadeiro. Observa-se que a qualidade na estimativa é particularmente deteriorada em baixas energias e na região do detector entre $1,37 < |\eta| < 1,8$.

Neste trabalho, utilizou-se um conjunto de árvores de decisão com reforço por gradiente (*Gradient Boosted Decision Trees Ensemble* - GBDTE) para realização da calibração objetivando a redução da dispersão no parâmetro α .

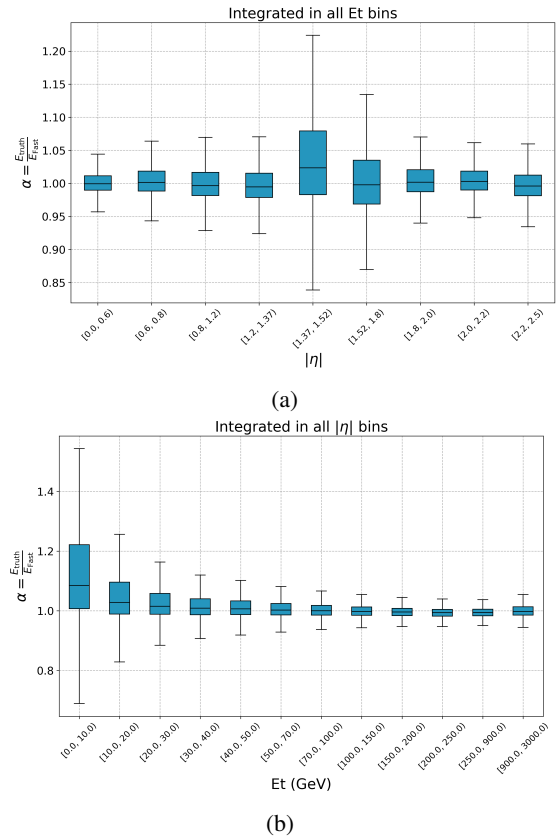


Figura 9: (a) Dispersão do α para a segmentação em $|\eta|$. (b) Dispersão do α para a segmentação em E_t . Dados obtidos a partir de simulações Monte Carlo.

Em geral, as árvores de decisão são utilizadas de forma recorrente em problemas de classificação e regressão [21]. Notadamente, a técnica do reforço por gradiente (*gradient boosting*) [22] possibilita incorporar diferentes funções de perda (*loss functions*), sendo aplicada em problemas de regressão e de classificação multi-classe. O princípio fundamental do *boosting* é a premissa de que modelos simples podem ser combinados e ajustados para ter seu desempenho melhorado (*boosted*).

Para implementação da GBDTE, foi utilizado o *LightGBM framework* [23]. Vários trabalhos anteriores da colaboração ATLAS já utilizam esse *framework* para calibração de energia na reconstrução *offline* [14], [20] bem como na etapa precisa do *High-Level-Trigger*.

A calibração GBDTE, na etapa rápida, utiliza como entradas algumas variáveis que descrevem o formato do perfil de deposição nos calorímetros (chuveiro de partículas). A Tabela II apresenta as variáveis de entrada e saída utilizadas no ATLAS. Ao todo 6 atributos são calculados a partir do perfil de deposição de energia nas diferentes camadas do calorímetro e são utilizados como entradas para as árvores de decisão.

Neste trabalho, propõe-se investigar a utilização da informação de calorimetria formatada em anéis concêntricos como entradas para a calibração GBDTE. Desta forma, os

resultados obtidos serão apresentados, em geral, de forma comparativa entre as duas abordagens: anéis (*rings*) e variáveis do chuveiro (*shower variables*).

Os eventos disponíveis foram divididas em conjuntos de treino e teste com respectivamente 70% e 30% do total de exemplos, para cada segmento $|\eta| \times E_t$ de calibração.

Tabela II: Atributos de entrada (*shower variables*) utilizados para o treinamento da GBDTE da colaboração ATLAS.

Entradas	
$E_{tcluster}$:	energia medida do cluster
$\eta_{cluster}$:	pseudo-velocidade do centro do cluster
$\phi_{cluster}$:	ângulo do centro do cluster
$E1_{raw}/E2_{raw}$:	razão entre as energias da EM1 e EM2
$E_{raw} = E1_{raw} + E2_{raw} + E3_{raw}$:	somatório de energia
$E0_{raw}/E1_{raw}$:	razão entre as energias da PS e EM1
E_{Tile1}/E_{raw} :	razão entre as energias da HAD1 e EM
Saída Alvo	
$\alpha_{BDT} = \frac{E_{true}}{E_{raw}}$:	fator de correção

IV. RESULTADOS

Neste trabalho, para escolha dos hiperparâmetros da GBDTE foram realizados exaustivos testes empíricos com diferentes configurações. Os melhores resultados foram obtidos para o seguinte conjunto de hiperparâmetros:

- Função de perda L1;
- 60 folhas em cada árvore;
- 2000 rodadas de *boosting*;
- Taxa de aprendizagem 0,05;
- Razão de bagging igual a 0,8;
- Parada antecipada de 5 rodadas sem mudança no conjunto de validação;

Uma comparação dos resultados obtidos para as duas configurações de calibração podem ser vistas por meio dos *boxplots* das Figuras 10 e 11. Nesse caso, as estimativas do α nos 2 cenários de calibração e no cenário sem calibração (*no BDT*) são comparadas. Pode-se observar na Figura 10 que a calibração utilizando tanto a informação do chuveiro quanto dos anéis reduz a dispersão da estimativa para praticamente toda a faixa de $|\eta|$ para o cenário simulado. Da mesma forma, a Figura 11 mostra que o GBDTE também reduz a dispersão quando avaliado nas faixas de energia transversa E_t , para qualquer uma das duas configurações de calibração e dentro do cenário simulado nesse experimento. Importante notar também, que além da redução da dispersão, o valor médio de α tende a se aproximar de 1 após a calibração, representando uma maior concordância com o valor verdadeiro.

De forma complementar à análise com *boxplots*, pode-se observar na Figura 12, para todas as faixas de $|\eta|$ a distância interquartílica ($IQR = Q3-Q1$) para as duas configurações de calibração comparadas. Ainda da Figura 12, parte inferior, pode-se verificar os valores de IQR normalizados (dividindo-se pela referência sem calibração). Nesse ponto, é importante salientar que até $|\eta| < 2,0$ os valores de IQR utilizando o BDT foram menores que os da configuração *no BDT*, indicando uma redução da dispersão na estimativa do α para essas faixas de $|\eta|$.

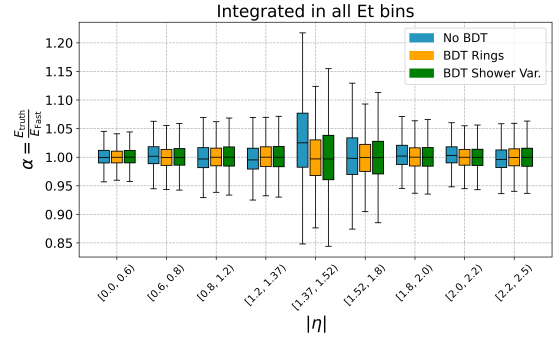


Figura 10: Comparação entre dispersões por faixa de $|\eta|$ com e sem calibração usando o BDT.

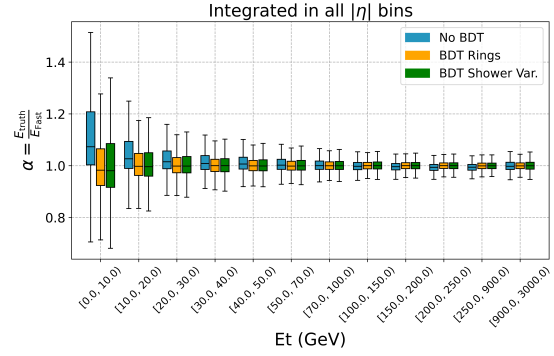


Figura 11: Comparação entre dispersões por faixa de E_t com e sem calibração usando o BDT.

Similarmente, a Figura 13 representa as distâncias interquartílicas ($IQR = Q3-Q1$) para as diferentes faixas da energia transversa E_t . Neste caso nota-se também uma redução da IQR praticamente em toda a faixa de energia. A maior redução da dispersão ocorre principalmente até 200 GeV, o que é destacado no gráfico normalizado.

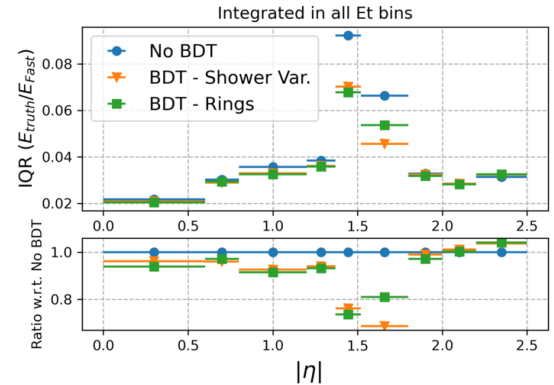


Figura 12: Distância interquartílica para a segmentação em $|\eta|$.

Analisando em mais detalhes os resultados obtidos para a faixa de energia entre 20 e 30 GeV, pode-se observar na Figura 14 que o histograma da energia antes da calibração usando GBDTE (Figura 14a) tem um considerável desca-

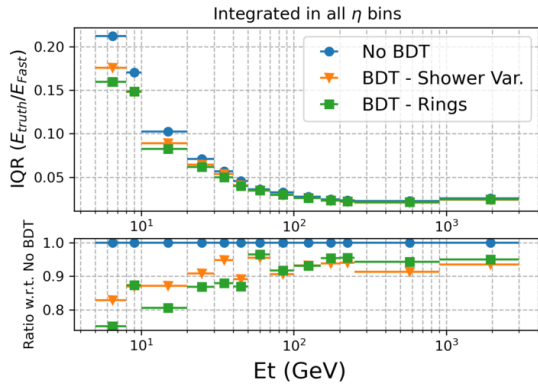
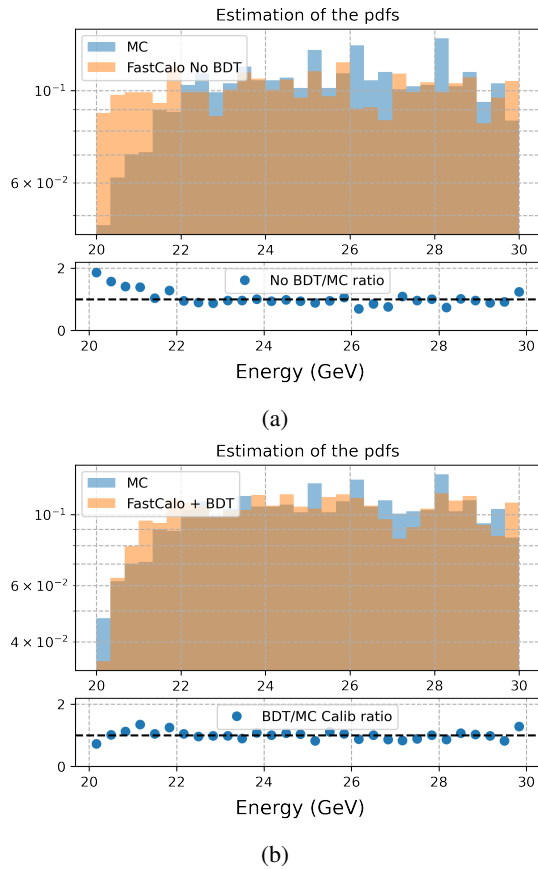

 Figura 13: Distância interquartil para a segmentação em E_t .


Figura 14: Comparação entre os histogramas da energia verdadeira versus a energia estimada na etapa rápida (a) antes e (b) depois da calibração com GBDTE.

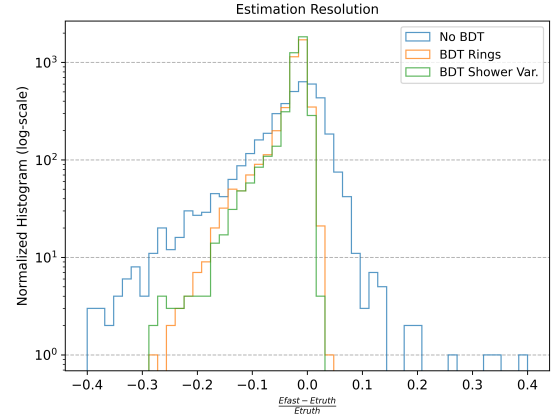
samento em relação ao histograma dos valores verdadeiros (MC). Esse descasamento é reduzido significativamente após a calibração, como pode-se observar na Figura 14b. Comportamento semelhante é observado para as demais faixas de energia.

Uma outra forma de observar o efeito da calibração é

calculando-se o erro relativo de estimação:

$$Er = \frac{E_{Fast} - E_{Truth}}{E_{Truth}}. \quad (3)$$

Pode-se observar na Figura 15 que a utilização da calibração em ambos os casos, com informações do chuva (*Shower Variables*) ou anéis (*Rings*) aumenta a quantidade de eventos com erro próximo a zero e diminui a dispersão de Er . Observa-se ainda que o uso da informações com anéis produz um pequeno ganho nos dois aspectos analisados.


 Figura 15: Histograma de resolução da estimação para E_t entre 20 GeV e 30 GeV.

Como a calibração está sendo projetada para operar no sistema de *trigger* do ATLAS, é importante observar seu efeito para a seleção de eventos. Pode-se observar nas curvas ROC da Figura 16, que a calibração utilizando anéis produz um aumento na probabilidade de detecção para um dado valor de taxa de aceitação de falso-alarme.

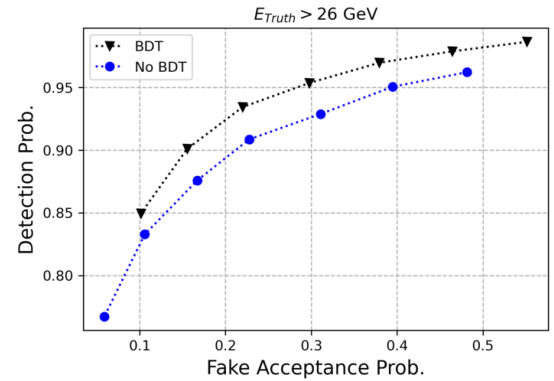


Figura 16: Curva ROC mostrando a probabilidade de detecção de elétrons com e sem o BDT, resultados para a calibração utilizando anéis.

Para ativação das cadeias de *trigger*, em muitos casos utiliza-se um patamar de energia mínima. Existem, por exemplo, as cadeias E15, E26 e E60, que idealmente deveriam processar candidatos a elétrons com energias maiores que 15, 26 e 60 GeV, respectivamente.

Para acomodar os erros de estimação de energia que existem atualmente na etapa rápida, o patamar de seleção das cadeias (λ_{Fast}) é definido 3 GeV abaixo do valor ideal para o menu (λ_{Menu}), fazendo:

$$\lambda_{\text{Fast}} = \lambda_{\text{Menu}} - 3 \quad (4)$$

Após a calibração proposta neste trabalho, foi possível mudar o patamar de corte da etapa rápida de *trigger* para:

$$\lambda_{\text{Fast}} = \lambda_{\text{Menu}} - 1,5 \quad (5)$$

de modo a manter a mesma probabilidade de detecção original, reduzindo consideravelmente a aceitação de elétrons fora da faixa desejada. A Tabela III apresenta os valores de elétrons fora da faixa desejada para diferentes cadeias de *trigger*. Pode-se observar que é possível obter reduções da ordem de 12%, 17% e 11% para as cadeias E15, E26 e E60, respectivamente.

Tabela III: Taxas (em %) de aceitação de elétrons com energia verdadeira menor que o corte do menu antes e depois da calibração, para três diferentes cadeias de *trigger* do ATLAS.

	No BDT	BDT - Rings
E15	40,0	27,8
E26	56,3	39,0
E60	32,0	21,0

V. CONCLUSÕES

Um novo método para calibração das medições de energia de elétrons na etapa rápida de *trigger* do detector ATLAS foi apresentado nesse trabalho. Um conjunto de árvores de decisão com reforço por gradiente foi gerado para produzir um fator de calibração α para valores estimados de energia transversas. Com o sistema proposto, foi possível reduzir a dispersão na estimativa para as diversas configurações geradas na avaliação dos segmentos da pseudorapidez, $|\eta|$ e energia transversa E_t . Também foi possível melhorar o desempenho da seleção das cadeias de *trigger*, reduzindo a aceitação de eventos com energia verdadeira menor que o patamar de seleção. Em trabalhos futuros pretende-se utilizar dados reais do experimento Atlas para calibração e classificação de energia na cadeia de elétrons utilizando a GBDTE apresentada neste trabalho.

AGRADECIMENTOS

Os autores agradecem à FAPESB, FAPERJ, CNPq e RE-NAFAE pelo apoio financeiro, e à colaboração ATLAS pelas contribuições ao trabalho. O presente trabalho foi apoiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- [1] H. M. Gray, "Future colliders for the high-energy frontier," *Reviews in Physics*, vol. 6, p. 22, 2021.
- [2] D. J. Griffiths, *Introduction to elementary particles; 2nd rev. version*, ser. Physics textbook. New York, NY: Wiley, 2008.
- [3] A. P. Morais, A. Onofre, F. F. Freitas, J. Gonçalves, R. Pasechnik, and R. Santos, "Deep Learning Searches for Vector-Like Leptons at the LHC and Electron/Muon Colliders," *Eur. Phys. J. C*, vol. 83, no. 3, p. 232, 2023, 26 pages, 11 figures, 10 tables, Published version. [Online]. Available: <https://cds.cern.ch/record/2853391>
- [4] M. Paganini and ATLAS Collaboration, "Machine learning algorithms for b-jet tagging at the ATLAS experiment," *Journal of Physics: Conference Series*, vol. 1085, p. 042031, sep 2018. [Online]. Available: <https://doi.org/10.1088/1742-6596/1085/4/042031>
- [5] D. J. V. J. e. a. Pata, J., "Mlpf: efficient machine-learned particle-flow reconstruction using graph neural networks," *Eur. Phys. J. C*, vol. 81, no. 381, pp. P03 017–P03 017, may 2021.
- [6] W. Spolidoro Freund, "An Ensemble of Neural Networks for Online Electron Filtering at the ATLAS Experiment," *J. Phys. Conf. Ser.*, vol. 1525, no. 1, p. 012076, 2020.
- [7] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 001–S08 001, aug 2008.
- [8] O. Brüning, H. Burkhardt, and S. Myers, "The Large Hadron Collider," *Progress in Particle and Nuclear Physics*, vol. 67, no. 3, pp. 705–734, 2012.
- [9] A. Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 003–S08 003, aug 2008. [Online]. Available: <https://doi.org/10.1088/1748-0221/3/08/s08003>
- [10] G. E. Orellana and A. Collaboration, "Projected ATLAS Electron and Photon Trigger Performance in Run 3," CERN, Geneva, Tech. Rep., 2021. [Online]. Available: <https://cds.cern.ch/record/2730417>
- [11] R. Wigmans, *Calorimetry: Energy Measurement in Particle Physics*, ser. International series of monographs on physics. Clarendon Press, 2017.
- [12] A. Collaboration, "Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data," *Journal of Instrumentation*, vol. 14, no. 03, pp. P03 017–P03 017, mar 2019.
- [13] A. Collaboration, "Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data," *Journal of Instrumentation*, vol. 14, no. 12, pp. P12 006–P12 006, dec 2019.
- [14] E. F. Simas Filho, P. C. M. . A. Farias, E. E. P. Souza, J. L. Marin, W. S. Freund, J. M. de Seixas, J. ao F. P. da Fonseca, and B. Laforge, "Uma proposta para calibração do sistema online de seleção de eventos no detector ATLAS utilizando gradient boosted decision trees," in *Anais do 15 Congresso Brasileiro de Inteligência Computacional*. Joinville, SC: SBIC, 2021, pp. 1–8.
- [15] J. Seixas, L. Caloba, M. Souza, A. Braga, and A. Rodrigues, "Neural second-level trigger system based on calorimetry," *Computer Physics Communications*, vol. 95, no. 2, pp. 143–157, 1996.
- [16] G. Aad, Abbott *et al.*, "Performance of electron and photon triggers in ATLAS during lhc run2," *Eur. Phys. J. C*, vol. 80, no. 47, pp. P2–P41, sep 2019.
- [17] S. S. Haykin, *Neural networks and learning machines*, 3rd ed. Upper Saddle River, NJ: Pearson Education, 2009.
- [18] T. Ciodaro, D. Deva, J. de Seixas, and D. Damazio, "Online Particle Detection by Neural Networks Based on Topologic Calorimetry Information," *14th International Workshop On Advanced Computing And Analysis Techniques In Physics Research, Uxbridge, West London, UK, 5 - 9 Sep 2011*, Sep 2011. [Online]. Available: <http://cds.cern.ch/record/1379508>
- [19] M. Aaboud *et al.*, "Electron and photon energy calibration with the ATLAS detector using lhc run 1 data," *The European Physical Journal C*, vol. 74, no. 14, p. 3071, 2014.
- [20] A. Collaboration, "Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data," *Journal of Instrumentation*, vol. 14, no. 03, pp. P03 017–P03 017, mar 2019. [Online]. Available: <https://doi.org/10.1088/1748-0221/14/03/p03017>
- [21] R. C. Barros, A. C. de Carvalho, and A. A. Freitas, *Automatic Design of Decision-Tree Induction Algorithms*. New York, NY: Springer, 2015.
- [22] R. F. J. H. Hastie, T.; Tibshirani, *Boosting and Additive Trees*. New York, NY: Springer, 2009.
- [23] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017.