

Uma Proposta para Calibração do Sistema Online de Seleção de Eventos no Detector ATLAS utilizando *Gradient Boosted Decision Trees*

Eduardo F. Simas Filho¹, Paulo C. M. A. Farias¹, Juan L. Marin¹, Edmar E. P. Souza¹, Werner S. Freund², José M. Seixas², João V. F. Pinto² e Bertrand Laforge³

¹*Laboratório de Sistemas Digitais - PPGEE - Universidade Federal da Bahia*

²*Lab. de Processamento de Sinais - PEE/COPPE/Poli, Universidade Federal do Rio de Janeiro*

³*LPNHE - Sorbonne Université, Paris, França*
{eduardo.simas, paulo.farias}@ufba.br

Resumo—Experimentos de física de partículas lidam com um enorme volume de informação e uma complexa cadeia sequencial de processamento para seleção *online* (*trigger*) de eventos. No detector ATLAS do LHC (*Large Hadron Collider*), o sistema de *trigger* é responsável por escolher os eventos que serão gravados em mídia permanente para futura análise e opera de modo sequencial em dois níveis de seleção. O valor estimado da energia depositada pelas partículas no detector é um importante parâmetro para o processo de seleção *online*. Neste trabalho é proposto um método de calibração baseado em *gradient boosted decision trees ensemble* (GBDTE) para melhorar a qualidade da energia estimada na segunda etapa de *trigger* do detector ATLAS. Com o método proposto é possível, ao mesmo tempo, reduzir requisitos computacionais e aumentar a eficiência de seleção de partículas eletromagnéticas (elétrons e fótons).

Index Terms—Física de partículas, calibração, *gradient boosting*, *decision trees*.

I. INTRODUÇÃO

Na física de partículas experimental, aceleradores são utilizados para produzir colisões entre feixes de partículas e, assim, permitir a detecção e medição das características fundamentais da matéria [1], [2]. Durante as colisões, grande parte da informação produzida já é conhecida (ruído de fundo), exigindo sistemas de filtragem *online* com capacidade de preservar a informação de interesse (geralmente eventos raros) e rejeitar o ruído de fundo. Em muitos casos, assinaturas de ruído de fundo apresentam perfil de deposição de energia semelhante ao evento físico de interesse, dificultando o desempenho de classificação. Neste contexto, técnicas de aprendizado de máquina têm sido empregadas em aplicações envolvendo sistemas de classificação para detecção de partículas [3] [4], reconstrução de eventos [5] e filtragem *online* [6].

Entre os experimentos em atividade atualmente, o Grande Colisionador de Hádrons (*Large Hadron Collider*) [7] acelera feixes de prótons em sentidos opostos no seu caminho circular de aproximadamente 27 km que fica a uma profundidade média de 100 m, na fronteira franco-suíça perto de Genebra. O LHC realiza colisões a cada 25 ns (a uma taxa de 40 MHz) em cada um de seus quatro pontos de cruzamento de feixe [8].

O detector ATLAS [9] está posicionado em um dos pontos de colisão do LHC, sendo formado de diversos sub-detectores (detector de traços, calorímetros e detector de múons), a cada colisão gera aproximadamente 1,3 MB de informação. Considerando a taxa de colisão nominal do LHC, há uma produção de aproximadamente 52 TB/s, deste modo é necessário um sistema *online* de seleção (*trigger*) dos eventos de interesse que são armazenados em mídia permanente [10].

Os calorímetros (medidores de energia altamente segmentados) são muito importantes para o sistema de (*trigger*) *online* no ATLAS, pois tem resposta bastante rápida e os perfis de deposição de energia apresentam características discriminantes para detecção dos eventos de interesse [11]. Conforme ilustrado na Fig 1, o calorímetro do ATLAS é dividido no calorímetro eletromagnético e calorímetro hadrônico. O primeiro é composto por uma estrutura sequencial de argônio líquido (LAR - *Liquid Argon*) e placas de chumbo com formato de acordeon, e o segundo por telhas (*tiles*) de material plástico cintilante. Os calorímetros foram projetados para medir a energia das partículas de modo destrutivo (por absorção total), sendo assim elas vão perdendo energia à medida que interagem com o material detector e gerando uma cascata de novas partículas menos energéticas. Apenas partículas como múons ou neutrinos conseguem se propagar além dos calorímetros, pois interagem muito pouco com a matéria.

O sistema de seleção *online* de eventos do ATLAS é dividido em dois níveis, o primeiro nível (L1), que opera em *hardware* dedicado e o alto nível (HLT - *High-Level Trigger*), projetado em *software*, que é executado em um sistema de computação distribuída. O HLT é dividido em uma etapa rápida, seguida de uma etapa precisa. Como depende essencialmente da informação dos calorímetros, o sistema de *trigger* é sensível a erros de medição de energia. Esses erros podem ocorrer na etapa de processamento *online*, pois nem toda a resolução do detector está disponível. Atualmente etapas de calibração, que aumentam consideravelmente a resolução da medição dos calorímetros, estão disponíveis apenas no processamento *offline* e na etapa precisa do HLT [12] [13].

Neste trabalho é apresentada uma proposta para um sistema

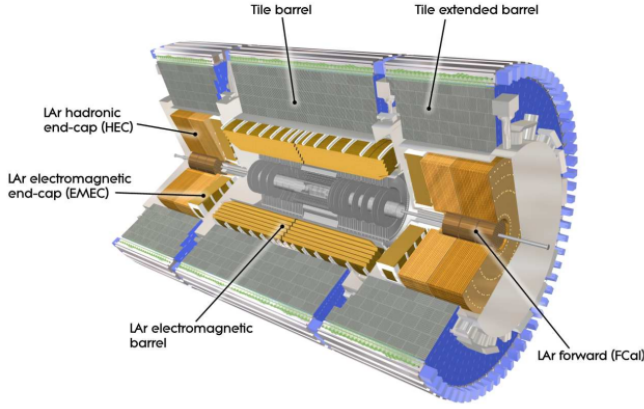


Figura 1: Diagrama ilustrativo do sistema de calorímetros do detector ATLAS com seus diferentes segmentos: eletromagnético de argônio líquido - LAr (barril e tampa, ou *end-cap*) e hadrônico de telhas cintilantes - Tile (barril, barril estendido e tampa), baseado em [9].

de calibração do calorímetro eletromagnético do ATLAS para operar na etapa rápida do HLT para elétrons. Para isso foi utilizado um regressor baseado em *gradient boosted decision trees ensemble* (GBDTE) treinado com dados simulados, essa é a mesma topologia empregada atualmente nas demais etapas de análise do ATLAS. Para a nova proposta deste trabalho foram incluídas como informações de entrada para a GBDTE os anéis concêntricos formados a partir das células do calorímetro. A informação em anéis é altamente discriminante para classificação, produz significativa compactação e é utilizada para o *trigger* de elétrons na etapa rápida do ATLAS.

Com o sistema proposto foi possível melhorar a precisão na estimação de energia e , portanto, contribuir para aumentar a eficiência do sistema de detecção *online*. Adicionalmente, são apresentados resultados preliminares sobre a investigação da assimetria do chuveiro de partículas a partir de estruturas geométricas inspiradas nos anéis concêntricos.

II. SELEÇÃO ONLINE (TRIGGER) DE ELÉTRONS

Conforme ilustrado na Fig. 2, o sistema de *trigger* do ATLAS é composto de dois níveis: Nível 1 (L1 - *Level 1*) e HLT (*High-Level Trigger*). O último, por sua vez é dividido em etapa rápida (*Fast*) e etapa precisa (*Precision*).

O nível 1 opera em *hardware* dedicado para conseguir lidar com toda a taxa de informação produzida no detector e fornecer uma resposta de aceitação ou rejeição do evento em até $2,5\mu s$. O HLT utiliza um cluster de computadores para realizar o processamento paralelo dos eventos aprovados no L1. A Etapa Rápida (primeira parte do HLT) opera utilizando apenas a informação da região de interesse (*Region of Interest* - RoI) indicada pelo L1. Os eventos aprovados nesta etapa são processados na Etapa Precisa, que utiliza agora toda a informação do detector para produzir a decisão final de aceitação ou rejeição do evento. Durante todo o processo de *trigger* as informações ficam armazenadas em um *pipeline*

de memórias voláteis e em sequencia os eventos aceitos são gravados em mídia persistente.

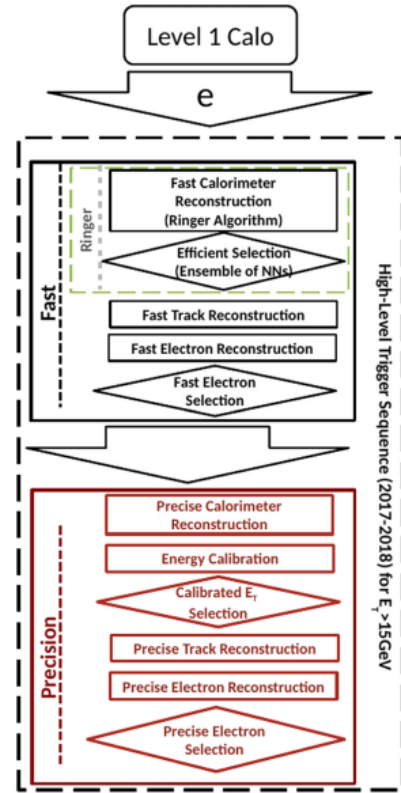


Figura 2: Diagrama do sistema de seleção *online* (*trigger*) de elétrons do detector ATLAS, baseado em [14].

O NeuralRinger [14] [15] opera no começo da Etapa Rápida do processo de identificação de elétrons desde 2017 [14]. 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 ilustra a primeira etapa do algoritmo, onde anéis concêntricos são construídos centrados na célula mais energética (em vermelho) em cada camada do calorímetro eletromagnético (EM1-EM3) e hadrônico (H1-H3), além do pré-amostrador (PS). Em seguida, as células pertencentes a cada anel são somadas, gerando 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.

A energia resultante em cada um dos 100 anéis gerados é utilizada como entrada para um *ensemble* de redes neurais

¹O ATLAS usa um sistema de coordenadas baseado na mão direita, 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 η é definida em termos do ângulo polar θ como $\eta \equiv -\ln [\tan(\theta/2)]$.

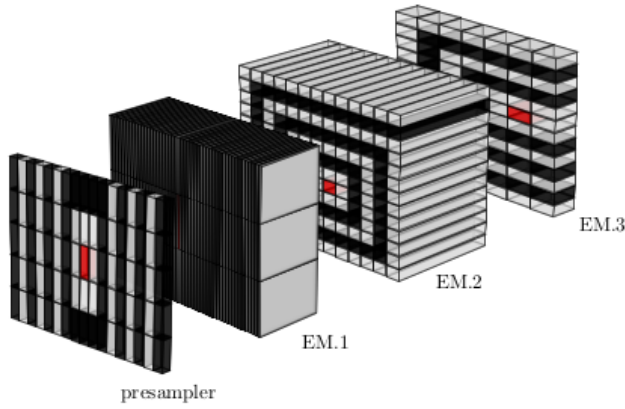


Figura 3: Exemplo da montagem dos anéis para as camadas do calorímetro eletromagnético, pode-se perceber a diferença de granularidade entre as diferentes camadas.

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

artificiais [16] (treinados para atuar em distintas faixas de energia e região do calorímetro). O *ensemble* é responsável por efetuar a discriminação entre 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 [17].

III. CALIBRAÇÃO DE ENERGIA USANDO GBDTE

Para ser útil no processo de caracterização das assinaturas de interesse, a informação medida no calorímetro precisa ser convertida em um valor que seja aproximadamente igual à quantidade de energia perdida pela partícula ao interagir com o material sensor. Para isso, existem procedimentos de calibração executados nas etapas do HLT e também na análise *offline* dos dados [18].

Conforme ilustrado na Figura 4, 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). Assim, 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*).

Neste contexto, o procedimento de calibração tenta corrigir os fatores de erro de medição. A Figura 5 ilustra a inserção da calibração proposta para a etapa rápida do HLT. Com esse novo bloco de processamento será possível corrigir a estimativa de energia do agrupamento na entrada do HLT, garantindo uma operação otimizada da etapa rápida e, con-

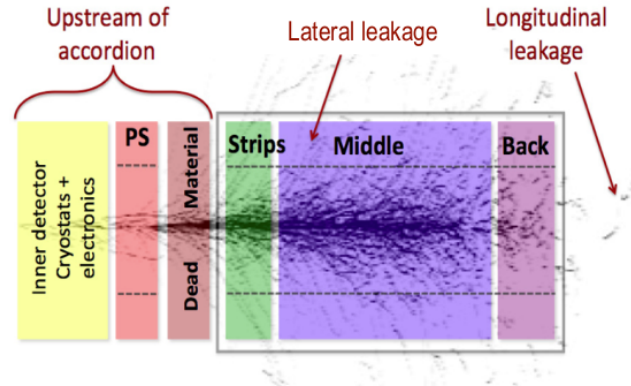


Figura 4: Ilustração das principais causas dos erros na estimativa da energia total da partícula ao interagir com o calorímetro do ATLAS: perda de energia antes (*upstream*) de interagir com o detector; vazamento lateral e longitudinal além da região de interesse, baseada em [19].

sequentemente, uma redução na demanda computacional do HLT.

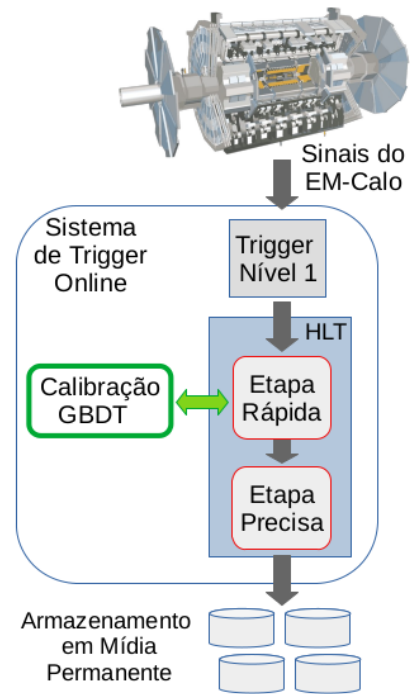


Figura 5: Diagrama indicando o bloco de calibração proposto, mostrando sua integração com o HLT do ATLAS.

De modo simplificado, pode-se considerar que a energia estimada é obtida a partir de:

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

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 [18];
- α_{BDT} é o fator de correção fornecido pela GBDTE proposta neste trabalho;
- $\alpha_{MC \rightarrow Data}$ 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 casado utilizado na detecção dos pulsos medidos no calorímetro [18];
- s_j são as amostras medidas no calorímetro;
- p é o pedestal (nível médio esperado nos sensores do calorímetro).

Neste trabalho serão descritos os procedimentos adotados para estimação de α_{BDT} e $f_{MC \rightarrow Data}(\cdot)$, na calibração da etapa rápida do HLT para detecção de elétrons utilizando as informações dos anéis, disponíveis na etapa rápida do HLT. Conforme indicado na Figura 6, a etapa (i) consiste no treinamento da GBDTE utilizando como entradas as informações do agrupamento de células e também os anéis. A etapa (ii) realiza o ajuste entre as distribuições de simulação e de dados experimentais. Essa etapa visa corrigir possíveis discrepâncias advindas de imperfeições na simulação.

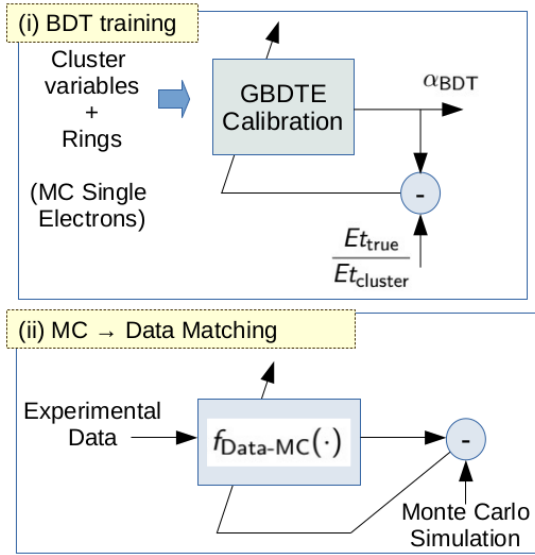


Figura 6: Diagrama das etapas de desenvolvimento do sistema de calibração proposto.

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.

A resposta do calorímetro do ATLAS tem características diferentes em função da energia transversa da partícula E_t , pois o perfil de deposição muda em função de E_t , sendo que, em baixas energias é usualmente mais difícil a identificação de elétrons. Também há um comportamento variável em termos da granularidade dos elementos sensores considerando

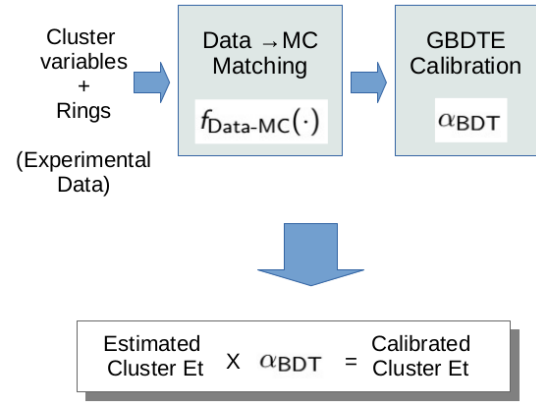


Figura 7: Diagrama das etapas de operação do sistema de calibração proposto.

a elevação em relação ao feixe do LHC, medida a partir da pseudo-rapidez η ($|\eta| = 0$ 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 $\rightarrow 0^\circ$). Em geral, pode-se afirmar que a granularidade dos sensores diminui com o aumento de $|\eta|$.

Neste sentido, todo o procedimento de calibração precisa ser projetado para diferentes intervalos de $|\eta|$ e E_t . A partir de estudos prévios [13] 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].
- E_t - [0, 5, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 900, 3000] GeV.

Assim, os fatores α definidos na Eq. 1 precisam ser determinados para cada par de combinação $|\eta| \times E_t$, o que produzirá um conjunto de $9 \times 13 = 117$ fatores distintos.

As árvores de decisão são muito utilizadas em problemas de classificação e regressão [20]. A técnica do reforço por gradiente (*gradient boosting*) [21] possibilita incorporar diferentes funções de perda (*loss functions*) sendo aplicação 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*).

Neste trabalho, foi utilizado um conjunto de árvores de decisão com reforço por gradiente (*Gradient Boosted Decision Trees Ensemble - GBDTE*). Para treinamento foi utilizado o LightGBM framework [22]. Essa escolha foi guiada por trabalhos anteriores da colaboração ATLAS, que utilizam GBDTE para calibração de energia na reconstrução *offline* [13], [18] e também na etapa precisa do HLT.

Nestes casos, a calibração GBDTE utiliza como entradas algumas variáveis descritoras do formato do perfil de deposição nos calorímetros. Na Tabela II são apresentadas as variáveis de entrada e saída, ao todo, 6 atributos calculados a partir do perfil de deposição de energia nas diferentes camadas do calorímetro são utilizados como entradas para as árvores de decisão.

Tabela II: Atributos de entrada utilizados para o treinamento da GBDTE padrão da colaboração ATLAS.

Entradas
$\eta_{cluster}$: posição 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
$\Delta\eta$ and $\Delta\phi$: assimetria em relação à célula mais energética
Saída Alvo
$\alpha_{BDT} = \frac{E_{true}}{E_{raw}}$: fator de correção

Para o ajuste dos dados simulados aos experimentais, o fator de conversão $\alpha_{MC \rightarrow Data}$ precisa ser estimado. Neste trabalho, foi proposta a realização de uma transformação linear $f_{MC \rightarrow Data}$ aplicada às amostras de simulação R_{MC} , de modo que:

$$\hat{R}_{MC} = aR_{MC} + b \quad (2)$$

com o objetivo de reduzir a discrepância entre as distribuições de probabilidade (*probability distribution function* - pdf) dos dados de simulação transformados \hat{R}_{MC} e as pdf experimentais

Assim, $f_{MC \rightarrow Data}$ foi projetada para tornar aproximadamente iguais o valores médio e o desvio padrão de \hat{R}_{MC} e R_{Data} .

Nesta transformação:

$$a = \sqrt{\frac{\sigma_{Data}^2}{\sigma_{MC}^2}} \quad (3)$$

$$b = E\{R_{Data}\} - aE\{R_{MC}\} \quad (4)$$

onde $E\{\cdot\}$ é o operador do valor esperado (média).

O efeito da transformação foi avaliado a partir do cálculo da distância de Wasserstein [23] entre as pdf dos anéis experimentais e simulados (originais e depois da transformação linear).

A GBDTE recebe como entrada os anéis concêntricos de deposição de energia, que descrevem a distribuição energética longitudinal e transversal. No entanto, estudos preliminares apontaram a existência de assimetrias na deposição de energia em relação aos eixos de referência espacial η e ϕ do calorímetro. O impacto no processo de calibração do uso de novas topologias que consigam descrever essas assimetrias está sendo estudado. A comparação entre os anéis concêntricos (anéis padrão) e as novas topologias está em estudo para avaliar a diferença em desempenho na calibração.

Duas topologias propostas são *Quarter Rings* (QR) e *Super Strips* (SS). No caso dos *Quarter Rings*, é feita uma divisão axial da estrutura dos anéis concêntricos, onde cada anel é dividido em quatro partes (excetuando-se o primeiro). Essa topologia aumenta a granularidade do mapeamento, permitindo capturar informações de assimetrias dos chuveiros. A Figura 8 traz uma representação simplificada do QR, mostrando 1 anel padrão, que é dividido nos 4 quadrantes arbitrariamente denominados $\eta_+\phi_+$, $\eta_+\phi_-$, $\eta_-\phi_-$ e $\eta_-\phi_+$.

A Figura 9 mostra a representação esquemática de uma *Super Strip*. Essa topologia agrupa células ao longo de um dos eixos de orientação (η ou ϕ).

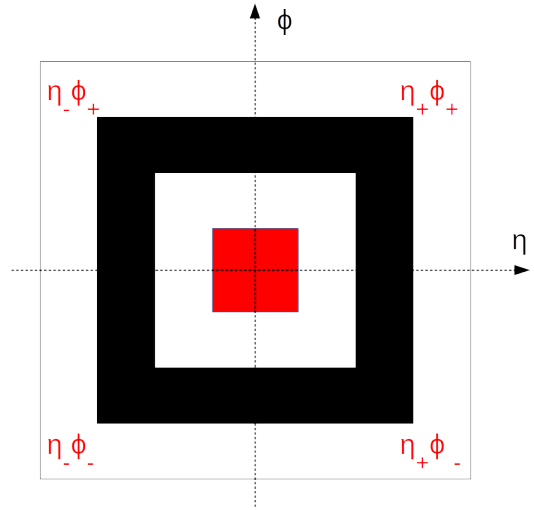


Figura 8: Estrutura do *Quarter Ring*.

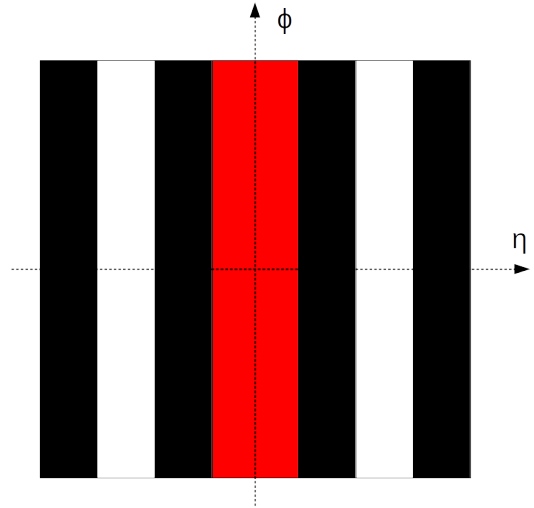


Figura 9: *Super Strip* orientada na direção ϕ .

IV. RESULTADOS

Para esse trabalho, no desenvolvimento da GBDTE foram utilizados dados simulados de elétrons isolados, com distribuição de energia aproximadamente uniforme entre 20 e 200 GeV. O conjunto de dados é composto de aproximadamente 100.000 assinaturas, divididas nos conjuntos de treino (70%), teste e validação (30%). Nas simulações de Monte Carlo são consideradas as características físicas de todo o detector ATLAS.

Na obtenção do mapeamento linear ($f_{MC \rightarrow Data}$) foram empregados eventos simulados e experimentais do decaimento do bóson Z em um elétron e um pósitron ($Z \rightarrow e^-e^+$). Ao todo, para esse estudo, foram utilizados 370.000 agrupamentos eletromagnéticos simulados e 420.000 experimentais.

Na faixa de energia para a qual os eventos de elétrons isolados estavam disponíveis, foi realizada a segmentação em $|\eta| \times E_t$ definida para calibração no ATLAS (indicada na

Sec. III) e realizado o treinamento da GBDTE. Os hiperparâmetros para treinamento foram obtidos após teste exaustivo de diferentes configurações. Deste modo, foram utilizadas 60 folhas em cada árvore, 2000 rodadas de *boosting*, taxa de aprendizagem 0,05 e razão de *bagging* igual a 0,8.

Os resultados combinados para todos os eventos de elétrons isolados utilizados estão indicados na Figura 10, em termos da estimação de energia dos eventos de elétrons isolados. Foram utilizadas diferentes configurações de entradas para as GBDT, conforme descrito a seguir:

- *Replicate (no rings)*: treinamento realizado de acordo com a configuração padrão da colaboração ATLAS (sem os anéis);
- *withRings (p1)*: todos os anéis em uma janela de $0, 1 \times 0, 1$ em $\eta \times \phi$. Os anéis são normalizados dividindo-os pela soma de todos os 100 anéis;
- *withRings (p1, c1N)*: igual ao modo *p1*, mas normalizado pela energia de todo o agrupamento eletromagnético;
- *withRings (p1, c1N, 8etacells)*: é utilizada uma janela reduzida de 0.05 em η .
- *withRings (p05)*: a região utilizada é reduzida a uma janela de $0, 05 \times 0, 05$ em $\eta \times \phi$.

Conforme pode-se observar na Figura 10 a distribuição da razão entre a energia estimada e a energia verdadeira ($E_{Reconstructed}/E_{MC,truth}$) é mais concentrada em torno de 1 para as estratégias de calibração que utilizam os anéis como entradas, o melhor caso foi obtido para o modo *withRings (p1, c1N)*. O pico da distribuição para essa configuração é aproximadamente 7% mais alto do que para a estratégia de calibração padrão (*no rings*)

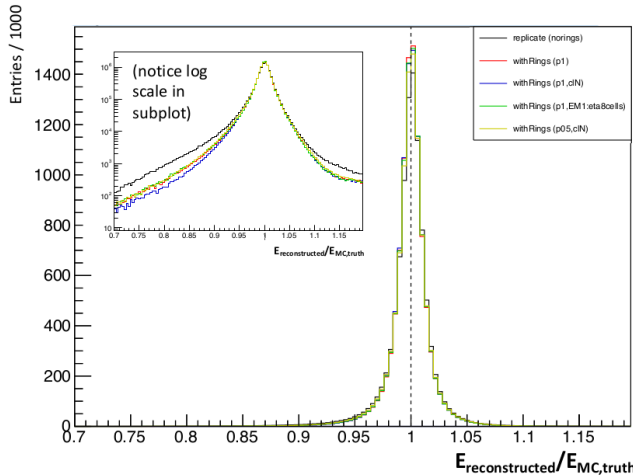


Figura 10: Razão entre a energia estimada e a energia verdadeira para diferentes configurações de calibração.

Outra forma de avaliar os resultados de calibração é comparar espalhamento da razão $E_{Reconstructed}/E_{MC,truth}$, pois quanto menos espalhada, está mais próximo do valor verdadeiro está o valor estimado. Neste trabalho será utilizada a distância inter-quartilica (DIQ), que indica a região em torno

do valor médio onde está concentrada 50% da área da pdf. Nas Figuras 11 e 12 pode-se observar o comportamento da DIQ normalizada pela pdf da calibração padrão (sem anéis) para diferentes intervalos de E_t e η , respectivamente.

Considerando especificamente o comportamento em E_t , observa-se na Figura 11 que pode-se reduzir o espalhamento em até 15% para energia acima de 40 GeV. Para energias mais baixas, configurações com janelas menores podem apresentar melhor resultado.

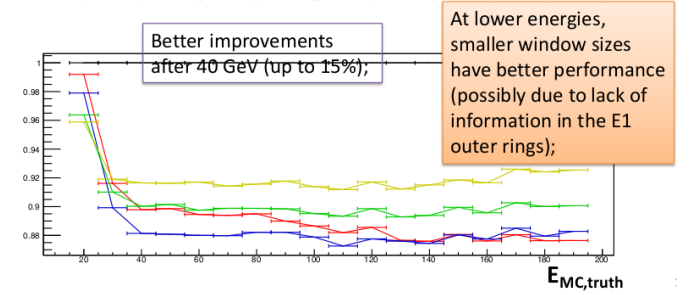


Figura 11: Distância inter-quartilica normalizada pela calibração padrão do ATLAS para diferentes intervalos de energia transversa (E_t).

No caso da variação da pseudorapidez (η), pode-se observar na Figura 12 que a calibração utilizando anéis pode apresentar resultados bastante interessantes para $|\eta| > 0, 8$, ou seja, se afastando da região central do detector (que fica próxima de $|\eta| = 0$).

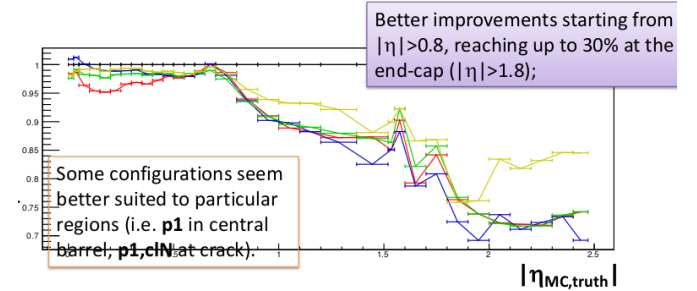


Figura 12: Distância inter-quartilica normalizada pela calibração padrão do ATLAS para diferentes intervalos de pseudorapidez (η).

Em relação à transformação que faz com que a média e o desvio padrão dos anéis simulados sejam aproximadamente iguais aos experimentais, ela também foi calculada para cada segmentação em $|\eta| \times E_t$ utilizando os dados (experimentais e simulados) de decaimento $Z \rightarrow e^-e^+$. As transformações foram calculadas para cada um dos 100 anéis e para avaliar o desempenho obtido, foi estimada a distância de Wasserstein (WD) entre as distribuições dos anéis experimentais e simulados (originais e após a transformação). Conforme pode-se observar na III, a distância entre as distribuições após a transformação (WD Transf) é reduzida consideravelmente. Na tabela são mostrados os 8 primeiros anéis, que estão na camada

pre-sampler (PS). Resultados semelhantes foram obtidos para os demais anéis.

Tabela III: Resumo dos resultados do mapeamento linear para ajuste dos dados simulados aos dados experimentais para os anéis da camada PS. Pode-se observar os valores médios (\bar{R}) e desvio padrão (σ) para distribuições de simulação (MC) e dados experimentais (Data), além das distâncias de Wasserstein entre as pdfs antes e depois da transformação.

Ring - PS	1	2	3	4	5	6	7	8
\bar{R}_{Data} (MeV)	844	328	74	33	22	17	16	10
\bar{R}_{MC} (MeV)	861	291	51	20	19	10	9	7
σ_{Data} (MeV)	501	430	339	267	252	250	235	212
σ_{MC} (MeV)	494	422	351	281	282	273	260	234
WD	21.0	40.8	25.9	18.0	23.8	18.9	20.9	18.1
WD Transf	8.7	14.9	12.3	5.8	2.0	2.8	2.8	2.0

Na Figura 13 pode-se observar o resultado da transformação linear para o anel 5 da camada PS. Na coluna da esquerda, parte superior, são mostradas as distribuições de simulação (MC) e experimental (Data) e na parte inferior a divisão entre as distribuições. Na coluna da direita são mostrados gráficos equivalentes após a transformação. Pode-se ver que após a transformação as duas distribuições estão mais sobrepostas, implicando na maior concentração da razão em torno de 1 e em um valor reduzido da WD.

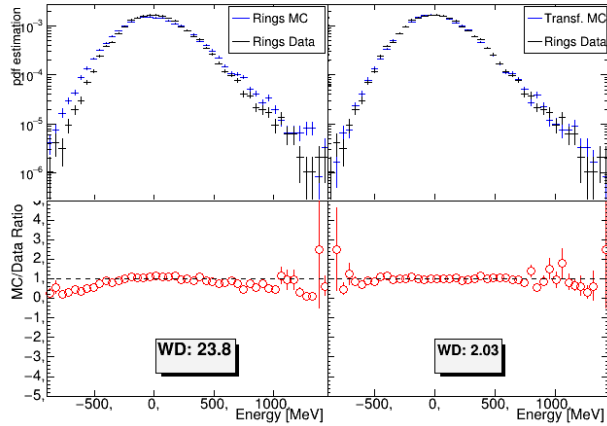


Figura 13: Avaliação da discrepância entre os dados simulados e experimentais para o anel 5 da camada PS, antes (esquerda) e depois (direita) da transformação linear. No topo são mostradas as pdf e na parte inferior a razão entre as pdf de simulação e experimental.

Como resultado preliminar da investigação da assimetria usando a topologia de *Super Strips*, foi realizada uma análise de acordo com o diagrama da Fig. 14, onde foi calculada a diferença de energia entre faixas equidistantes em relação à célula central. As distribuições de diferenças da Fig. 15 apontam para a existência de assimetrias na distribuição energética lateral. O gráfico da Fig. 15 mostra o resultado para a camada EM2 do calorímetro, na região do barril ($0 < |\eta| < 0,8$). Pode-se observar, por exemplo que para o SuperStrip Index = 1

a distribuição não é simétrica apresentando, um deslocamento para valores negativos perto de zero

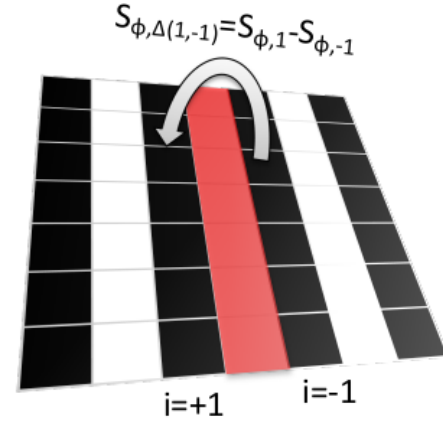


Figura 14: Análise de assimetria no fluxo de energia usando *Super Strips*.

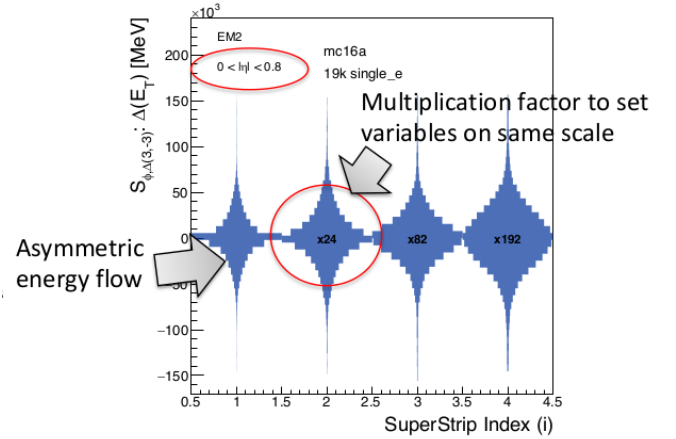


Figura 15: Distribuição energética com a *Super Strip*.

V. CONCLUSÕES

Neste trabalho foi apresentada uma nova metodologia para calibração das medições de energia de elétrons na etapa rápida de *trigger* do detector ATLAS. Um conjunto de árvores de decisão foi gerado para produzir um ajuste nos valores estimados, a partir do treinamento com dados simulados. Uma etapa adicional que proporciona um maior casamento entre os dados simulados e os experimentais também foi desenvolvida. Com o sistema proposto, foi possível melhorar a estimação em até 15% para alguns intervalos de energia e 30% para certos intervalos de pseudorapidez. A transformação linear proposta para reduzir a discrepância entre as pdf dos dados simulados e dos experimentais mostrou bons resultados para a redução da distância entre elas. Em próximos trabalhos pretende-se avaliar a utilização das informações assimétricas para o processo de calibração.

AGRADECIMENTOS

Os autores agradecem à FAPESB, CAPES, CNPQ e RE-NAFAE pelo apoio financeiro, e à colaboração ATLAS pelas contribuições técnicas ao trabalho. O presente trabalho foi realizado com apoio da 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] L.-G. Gagnon, “Searching for supersymmetry using deep learning with the ATLAS detector,” Ph.D. dissertation, Montreal U., 2020.
- [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] —, “Projected atlas electron and photon trigger performance in run 3,” *Proceedings of Science*, vol. ICHEP2020, p. 757, 2021.
- [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 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.
- [13] —, “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>
- [14] 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.
- [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] S. S. Haykin, *Neural networks and learning machines*, 3rd ed. Upper Saddle River, NJ: Pearson Education, 2009.
- [17] 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>
- [18] 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.
- [19] G. Pásztor *et al.*, “The upgrade of the atlas electron and photon triggers towards lhc run 2 and their performance,” *DPF 2015 Meeting of the American Physical Society Division of Particles and Fields*, Aug, Ann Arbor, Michigan 2015.
- [20] R. C. Barros, A. C. de Carvalho, and A. A. Freitas, *Automatic Design of Decision-Tree Induction Algorithms*. New York, NY: Springer, 2015.
- [21] R. F. J. H. Hastie, T.; Tibshirani, *Boosting and Additive Trees*. New York, NY: Springer, 2009.
- [22] LightGBM, “Lightgbm’s documentation webpage,” <https://lightgbm.readthedocs.io/en/latest/>, junho 2021.
- [23] P. Clement and W. Desch, “An elementary proof of the triangle inequality for the wasserstein metric,” *Proceedings of The American Mathematical Society - PROC AMER MATH SOC*, vol. 136, pp. 333–340, 01 2008.