

Combinando Filtro Casado e Topologia para Aprimoramento da Reconstrução de Múons no Calorímetro Hadrônico do ATLAS

D. O. Gonçalves, A. C. N. de Oliveira, J. M. de Seixas, A. S. Cerqueira e L. M. Andrade Filho

Resumo—Este trabalho apresenta o estudo de técnicas de discriminação de eventos de partículas do experimento ATLAS. O estimador de energia de máxima verossimilhança Gaussiano (MLEG) foi aplicado no contexto deste projeto e comparado com a técnica usada atualmente no calorímetro, o Filtro Ótimo (OF2). Diferentes topologias de fusão da informação de interesse também foram avaliadas. Os resultados adquiridos utilizando dados experimentais mostraram que o estimador MLEG, para o caso em que se considera somente a informação de maior valor estimado entre quatro informações avaliadas, apresenta maior eficiência na discriminação dos eventos de interesse, sendo o mais indicado para a aplicação.

Palavras-Chave—Estimação, Processamento Digital de Sinais, Fusão de Dados, Física de Altas Energias.

Abstract—This work presents the study of event discrimination techniques for particles in the ATLAS experiment. The Gaussian energy estimator for maximum likelihood (MLEG) was applied in this project's context and compared with the technique currently used in the calorimeter, the Optimum Filter (OF2). Different topologies for fusion of the information of interest were also evaluated. The results acquired using experimental data showed that the MLEG estimator, for the case in which it is considered only the highest value estimated among four selected information, presents the maximum discrimination efficiency of the events of interest, being therefore the method best suited for the task.

Keywords—Estimation, Digital Signal Processing, Data Fusion, High Energy Physics.

I. INTRODUÇÃO

A tarefa de discriminação de sinais é de suma importância para diversas áreas do conhecimento. Devido ao desenvolvimento tecnológico observado nas últimas décadas, a quantidade de informações a serem processadas vem se tornando cada vez maior e, consequentemente, o desenvolvimento de complexas técnicas de distinção da informação de interesse em um conjunto de dados se faz cada vez mais necessário.

No ambiente de física de partículas, tais técnicas são empregadas na identificação de fenômenos de interesse que se encontram imersos a um ruído de fundo. No LHC (*Large Hadron Collider*) [1], o maior colisor de partículas do mundo, esta tarefa se torna complexa devido a altas taxas de eventos,

A. C. N. de Oliveira, Escola Politécnica, Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. e-mail: amandaconliveira@poli.ufrj.br; D. O. Gonçalves, e-mail: dayane.goncalves@engenharia.ufjf.br, A. S. Cerqueira, e-mail: augusto.s.cerqueira@gmail.com, L. M. Andrade, e-mail: luciano.ma.filho@gmail.com, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Juiz de Fora, Juiz de Fora - MG, Brasil; J. M. de Seixas, Programa de Engenharia Elétrica, Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. e-mail: seixas@lps.ufrj.br

provenientes da colisão de feixes de protóns, a serem processados em um pequeno intervalo de tempo. O LHC é composto por quatro principais experimentos, responsáveis pela detecção das partículas incidentes no acelerador: ALICE [2], LHCb [3], CMS [4] e ATLAS [5].

O ATLAS é um experimento de propósito geral otimizado para detectar o maior número possível de eventos físicos que ocorrerão no LHC. Ele é composto por três sub-camadas: o detector de trajetória, os calorímetros eletromagnético e hadrônico e o espectrômetro de múons [5]. Os sub-detectores do ATLAS sofrem interações a uma taxa de 40MHz provenientes da colisão de dois feixes de prótons, resultando em uma enorme quantidade de dados gerados pelo detector. Este fato torna o processo de armazenamento e processamento de tais dados inviáveis e um sistema de filtragem na aquisição dos dados é necessário a fim de evitar o armazenamento de dados irrelevantes para os eventos físicos de interesse. O sistema de filtragem do ATLAS é composto por dois níveis, reduzindo de 40 MHz para 1 kHz a quantidade de eventos salvos na base de dados do experimento.

Inserido no primeiro nível de filtragem do ATLAS está o projeto TileMuon. Este sistema tem como objetivo melhorar a eficiência de discriminação de partículas de múons através da interseção das informações da energia depositada em dois sistemas do ATLAS, o Calorímetro Hadrônico de Telhas (TileCal) [6] e o Espectrômetro de Múons [7].

Este trabalho tem o intuito de avaliar técnicas de discriminação de eventos no TileCal dentro do contexto do projeto TileMuon. Uma técnica de estimação de energia foi proposta e sua eficiência na seleção dos eventos de interesse foi comparada com a técnica atualmente utilizada no TileCal. Diferentes topologias de fusão da informação de interesse também foram avaliadas e a melhor topologia foi escolhida considerando os resultados de probabilidade de detecção de sinais e rejeição de falsos alarmes.

II. O TILECAL E O PROJETO TILEMUON

Esta seção tem o intuito de apresentar o ambiente no qual este trabalho está inserido. Primeiramente, o TileCal é abordado, seguido por uma breve descrição do projeto TileMuon.

A. O TileCal

Os calorímetros são instrumentos responsáveis por medir a energia das partículas à medida que interagem com o material do dispositivo. O TileCal é um dos calorímetros que

compõem o sistema de calorimetria do ATLAS. Ele utiliza aço para absorver as partículas, e telhas cintilantes como material ativo para amostrar a energia depositada [6]. O sinal luminoso gerado pelas telhas cintilantes é convertido em sinal elétrico por tubos fotomultiplicadores (PMTs).

O TileCal é composto por três aparelhos cilíndricos em torno do feixe de colisões, o Barril Longo, separado em duas partições de leitura LBA e LBC, e dois Barris Estendidos EBA e EBC. Cada partição é dividida radialmente em 64 módulos, sendo os módulos do Barril Longo e do Barril Estendido divididos em 24 e 16 células, respectivamente. No total, o TileCal consiste em aproximadamente 5000 células que são lidas em 10000 canais. O calorímetro também é segmentado transversalmente em 3 camadas principais, A, BC e D, como pode ser observado na Figura 1.

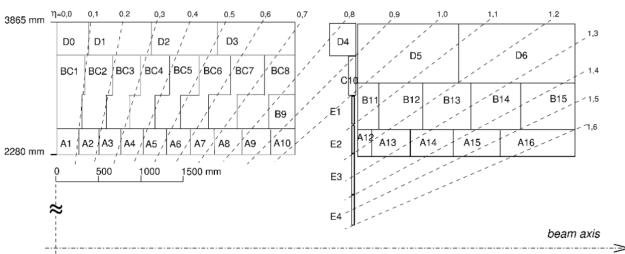


Fig. 1: Segmentação em células de um módulo LB e um EB do TileCal

O sinal elétrico fornecido pelas PMTs de cada leitura do TileCal é condicionado por uma eletrônica de aquisição, a qual fornece um pulso de forma fixa e amplitude proporcional à energia depositada na célula do calorímetro [6]. Este sinal é digitalizado à uma taxa de 40 MHz e uma janela de 7 amostras é adquirida. Posteriormente, durante a etapa do processamento digital, a amplitude e a fase do sinal recebido são estimadas.

Em ambientes como o TileCal, o sinal de interesse é, principalmente, corrompido por um ruído eletrônico. Como este ruído pode ser modelado por uma distribuição Gaussiana, as técnicas de estimativa de amplitude em calorímetros modernos são tipicamente baseadas em minimização da variância [8], [9].

B. O Projeto TileMuon

O projeto TileMuon foi desenvolvido com o principal objetivo de melhorar a eficiência da identificação de mísseis no primeiro nível de filtragem online do ATLAS [10]. Foi observado que partículas de baixa energia provenientes da eletrônica de blindagem do LHC estavam interagindo fortemente com o sistema dedicado a detectar a passagem de mísseis, o Espectrômetro de Mísseis, resultando em uma alta taxa de falso alarme de sinais de mísseis no sistema de validação em tempo real de eventos do ATLAS. Como o TileCal não está exposto às partículas geradas pela blindagem do LHC, surgiu a ideia de utilizar o sinal da camada mais externa do TileCal (células D, Figura 1) em coincidência com o sinal do Espectrômetro de Mísseis para melhorar a eficiência na identificação de mísseis. Desta forma, em meados de 2013, teve início o projeto TileMuon.

Como o TileCal não foi desenvolvido para a detecção de mísseis, um módulo eletrônico chamado TMDB (do inglês *TileMuon Digitizer Board*) foi elaborado para a aquisição, digitalização e processamento dos sinais da camada D do TileCal, além de prover comunicação com o Espectrômetro de Mísseis para que a coincidência seja realizada. Assim, a estimativa da energia depositada do sinal de mísseis, através da amplitude do pulso digitalizado, deve ser realizado pela TMDB. O algoritmo de estimativa de energia da TMDB é implementado em um FPGA, operando em tempo real. Dessa forma, a baixa complexidade computacional é um importante requisito para o algoritmo de estimativa da TMDB.

Os módulos eletrônicos foram instalados no primeiro nível de filtragem do ATLAS no final de 2015 e iniciou-se a fase de comissionamento do projeto. Embora as placas instaladas no Barril Longo já estejam operando desde de 2018, os benefícios de utilizar a informação das células do Barril Longo do TileCal em coincidência com o Espectrômetro de Mísseis ainda estão sendo avaliados. Dessa forma, as informações da energia depositada nas células do Barril Longo do TileCal são salvas e técnicas de estimativa estão sendo estudadas. Este trabalho se insere no contexto do projeto TileMuon no Barril Longo, através do estudo da eficiência de técnicas de discriminação de eventos de interesse imersos em um ruído de fundo.

III. DISCRIMINAÇÃO DE EVENTOS NA TMDB

Para discriminar os eventos de mísseis no TileCal, primeiramente é necessário estimar o valor da energia depositada nas células do módulo em que o mísseis ocorreu, bem como no módulo adjacente. Em seguida, as informações das células são fundidas com o objetivo de obter o valor da energia total depositada pelo mísseis no TileCal. Desta forma, duas técnicas de estimativa de energia são consideradas neste trabalho, o estimador de máxima verossimilhança Gaussiano (MLEG) e o Filtro Ótimo (OF2), uma técnica usada no TileCal desde o início da operação do LHC baseada na minimização da variância. Diferentes topologias de fusão de informação também são avaliadas.

A. Técnicas de estimativa da energia

O Filtro Ótimo é um método utilizado no ATLAS para reconstrução de energia. Esta técnica usa a função de autocorrelação do ruído para minimizar a variância do estimador. A versão utilizada neste trabalho, chamada Filtro Ótimo 2 (OF2) [11], está sendo atualmente utilizado no TileCal para estimar energia. O método é baseado em uma soma ponderada de amostras. A estimativa da amplitude pode ser calculada como:

$$\hat{A} = \sum_{i=1}^N s_i w_i \quad (1)$$

Os vetores s e w correspondem às amostras do sinal ADC recebidos e os pesos do filtro, respectivamente. O parâmetro N é o número de amostras.

A variância a ser minimizada pode ser calculada como:

$$\text{var}(\hat{A}) = \mathbf{w}^T \mathbf{C} \mathbf{w} \quad (2)$$

Onde \mathbf{C} é a matriz de covariância do ruído.

A minimização deve ser executada sob algumas restrições

$$\sum_{i=1}^N g_i w_i = 1 \quad (3)$$

$$\sum_{i=1}^N g'_i w_i = 0 \quad (4)$$

$$\sum_{i=1}^N w_i = 0 \quad (5)$$

Onde g_i é o pulso de referência do TileCal normalizado e g'_i é a sua derivada. A primeira restrição é para que o estimador seja não-tendencioso, as outras duas restrições tornam o estimador robusto a flutuações da linha de base e desvios de fase, respectivamente.

O conjunto ótimo de coeficientes do estimação é calculado minimizando a Equação (2) sujeita às restrições impostas pelas Equações (3), (4) e (5) utilizando os multiplicadores de Lagrange [12].

B. Estimador de Máxima Verossimilhança Gaussiano

O Estimador de Máxima Verossimilhança Gaussiano utiliza as funções densidade de probabilidade do processo aleatório para estimar amplitude do sinal. Dessa forma, a melhor estimativa para a amplitude \hat{A}_{mle} é o valor que maximiza a função densidade de probabilidade do sinal recebido $p_{\mathbf{s}|\hat{A}_{mle}}$. A estimativa da amplitude pode ser encontrada resolvendo a seguinte equação para a variável \hat{A}_{mle} :

$$\frac{\partial p(\mathbf{s}|\hat{A}_{mle})}{\partial \hat{A}_{mle}} = 0 \quad (6)$$

Aonde \mathbf{s} é o sinal de entrada. Considerando que as amostras do ruído podem ser modeladas por uma distribuição Gaussiana multivariável com uma matriz de covariância \mathbf{C} e o processo aleatório é conhecido, a função de densidade de probabilidade é dada por:

$$p(\mathbf{s}|\hat{A}_{mle}) = \frac{1}{\sqrt{2\pi \det(\mathbf{C})}} \exp\left(-\frac{(\mathbf{s} - A\mathbf{g})^T \mathbf{C}^{-1} (\mathbf{s} - A\mathbf{g})}{2}\right) \quad (7)$$

Extraindo-se o logaritmo da Equação (7) e aplicando a derivada em relação a amplitude, resulta em:

$$\frac{\partial p(\mathbf{s}|\hat{A}_{mle})}{\partial \hat{A}} = \frac{-1}{\sqrt{2\pi \det(\mathbf{C})}} \frac{-(\mathbf{s} - A\mathbf{g})^T \mathbf{C}^{-1} (\mathbf{s} - A\mathbf{g})}{2} \quad (8)$$

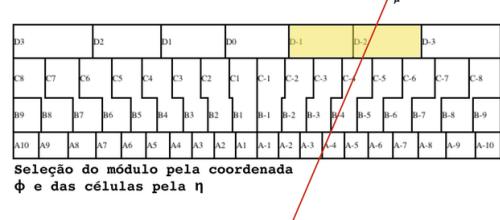
O valor máximo da Equação (7) pode ser obtido ao igualar a Equação (8) a zero. Dessa forma, a estimativa \hat{A}_{mle} , chamada neste trabalho de estimador MLE Gaussiano (MLEG), é descrita na Equação (9)

$$\hat{A}_{mle} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{g}}{\mathbf{g}^T \mathbf{C}^{-1} \mathbf{g}} = \mathbf{s}^T \mathbf{w}_{mle} \quad (9)$$

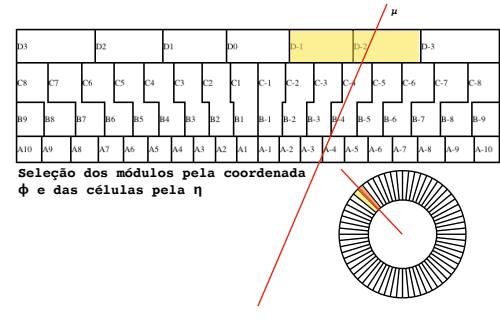
C. Topologias de fusão de informação no TileCal

Quando um evento de mísseis proveniente de uma colisão ocorre no LHC, sua posição em pseudo-rapidez (η) e ângulo azimutal (ϕ) em torno do eixo z no ATLAS é determinada pelo Espectrômetro de mísseis. Dessa forma, um casamento de geometria torna-se necessário para determinar aonde tal evento ocorreu no TileCal. Neste trabalho, quatro diferentes topologias de fusão de informação no TileCal foram estudadas, com o objetivo do desenvolvimento de um algoritmo de discriminação de mísseis robusto as incertezas provenientes do casamento da geometria. Tais topologias, ilustradas na Figura 2, podem ser descritas como

- 1 Seleção de duas células em um mesmo módulo, em torno do ponto (η, ϕ) . A energia total do mísseis é estimada através da soma da energia das células.
- 2 Seleção de duas células em um mesmo módulo, em torno do ponto (η, ϕ) . A energia total do mísseis é estimada através da energia da célula mais energética.
- 3 Seleção de quatro células de dois módulos do TileCal ($cluster 2 \times 2$), em torno do ponto (η, ϕ) . A energia total do mísseis é estimada através da energia da célula mais energética do *cluster*.
- 4 Seleção de quatro células de dois módulos do TileCal ($cluster 2 \times 2$), em torno do ponto (η, ϕ) . A energia total do mísseis é estimada através da soma da energia das duas células do módulo mais energético.



(a) Seleção de *cluster* para os algoritmos 1 e 2



(b) Seleção de *cluster* para os algoritmos 3 e 4

Fig. 2: Ilustração dos algoritmos de seleção de *clusters* de células do TileCal com base na posição (η, ϕ)

A eficiência dos algoritmos é avaliada utilizando dados de eventos reais de mísseis validados por sistema de seleção offline do ATLAS, no qual todas as informações da partícula

proveniente de cada evento são consideradas. Dessa forma, os métodos de estimação de sinais e as topologias de fusão de informação foram avaliadas de acordo com seus respectivos resultados de probabilidade de detecção de sinais e rejeição de falsos alarmes.

IV. RESULTADOS

Nesta seção os métodos de estimação de energia OF2 e MLE combinados com as topologias de fusão de informação descritas anteriormente, foram aplicados em dados experimentais. Oito diferentes combinações foram avaliadas com o intuito de determinar a mais adequada para a discriminação de múons.

A. Análise do ruído

Os projetos de estimadores de energia necessitam da caracterização prévia do ruído no qual o sinal está inserido. Ambos estimadores assumem ruído gaussiano e, como apresentado na seção III, utilizam para o cálculo dos seus coeficientes as informações da matriz de covariância do ruído. Para estudar essas características foram utilizados dados experimentais coletados pelo detector quando não havia a ocorrência de colisões (sem sinais de interesse).

A Figura 3 apresenta os histogramas das amostras de ruído lidas por dois canais do Barril Longo. Sobre os mesmos foi realizado o ajuste a uma distribuição gaussiana. Pode-se observar que, embora se assemelhem a distribuições gaussianas, nem sempre isso é confirmado pelo teste de hipótese chi-quadrado (χ^2)[13]. Desta forma, os métodos MLEG e OF2 podem operar em condições sub-ótimas.

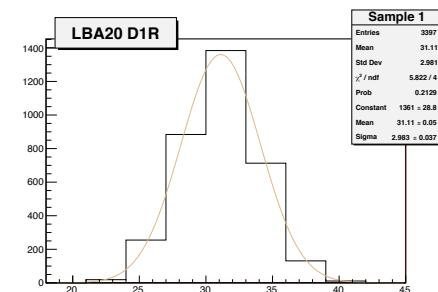
A matriz de covariância C é outra relevante característica do ruído a ser estudada. A matriz de covariância, estimada utilizando um algoritmo robusto [14], é apresentada na Figura 4. Pode-se observar que as amostras de ruído são fortemente correlacionadas, indicando a importância de utilizar a matriz C no projeto dos estimadores OF2 e MLEG. É importante ressaltar que todos os canais do Barril Longo analisados têm características de ruído semelhantes entre si.

B. Sinais de múons na TMDB

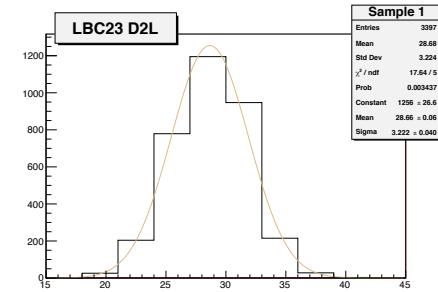
Os métodos de estimação de energia apresentados na seção III se baseiam na estabilidade do pulso recebido para estimar a amplitude. O pulso médio para um canal da TMDB adquirido utilizando dados experimentais com energia entre 1000 MeV e 5000 MeV é apresentado na Figura 5. O pulso médio provê uma boa representação dos pulsos observados no canal de acordo com o teste χ^2 com 5% de nível de significância, indicando que o MLEG e o OF2 apresentam uma boa estimativa da amplitude dos eventos.

C. Avaliação dos métodos de detecção

As combinações das técnicas de estimação de energia e das topologias de fusão de informação no TileCal foram avaliadas utilizando dados experimentais. Como descrito na sessão III-C, a eficiência dos algoritmos é avaliada utilizando dados reais de múons validados pelo sistema de seleção offline do ATLAS.



(a) Canal 1



(b) Canal 2

Fig. 3: Histograma do ruído de fundo e ajuste de distribuição gaussiana ao mesmo, para dois canais do Barril Longo do TileCal

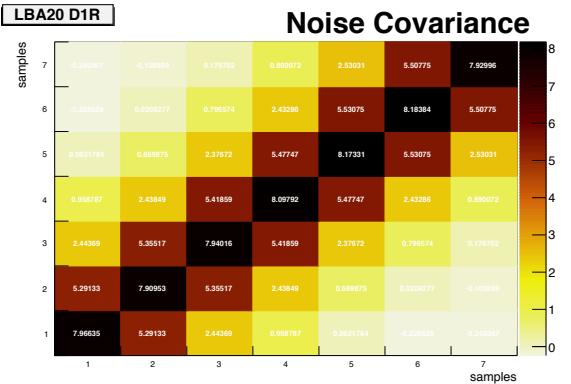


Fig. 4: Matriz de covariância do ruído para as 7 amostras da janela de leitura

A avaliação da eficiência da discriminação de múons no TileCal foi averiguada pelo compromisso entre a detecção e a rejeição de ruído através de curvas ROC dos métodos empregados. A detecção é dada pela razão entre o total de múons classificados corretamente pelo algoritmo e o total de múons confirmados pelo *offline*. O falso alarme corresponde aos eventos de ruído classificados como múons pela TMDB divididos pelo total de eventos de ruído.

A figura 6 mostra as curvas ROC[15] dos algoritmos empregados. Pode-se observar que os métodos que obtêm o maior compromisso entre detecção e falso alarme são os que utilizam o MLEG. Especificamente o algoritmo 3 com MLEG

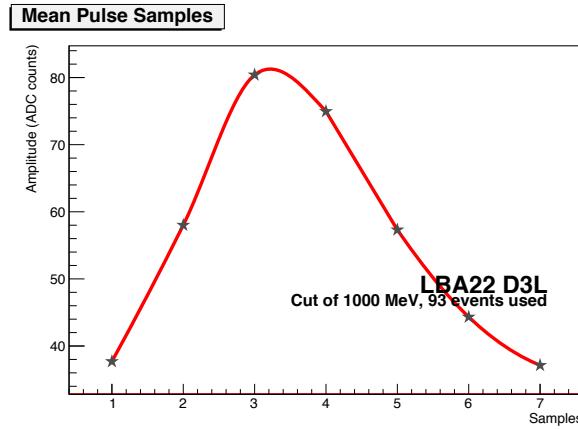


Fig. 5: Pulso médio de sinais para um canal do Barril Longo do TileCal

(cor verde claro e marcador de triângulo) obteve o maior desempenho dentre os métodos empregados.

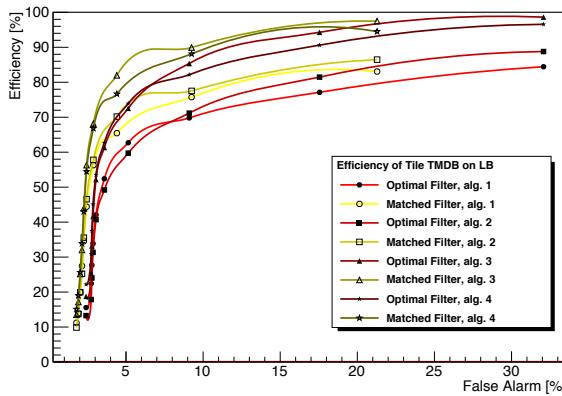


Fig. 6: Curvas ROC de eficiência das técnicas de estimação de energia no TileCal combinados com as topologias de fusão de informação. Na legenda MF corresponde a MLEG e OF a OF2.

A tabela I apresenta os valores quantitativos de detecção dos algoritmos para um falso alarme de aproximadamente 10%. O Filtro Casado com o algoritmo 3 de coincidência teve a maior detecção, com $\approx 90.8\%$ de aceite de eventos de mísseis.

V. CONCLUSÃO

Este trabalho apresentou um estudo sobre métodos de discriminação de partículas no contexto do projeto TileMuon, envolvendo a combinação de técnicas de estimação de energia e topologias de fusão de informação.

A utilização do estimador por MLEG em conjunto com a topologia de fusão de informação que usa a informação da célula mais energética em um cluster (2×2) foi a combinação que gerou o maior compromisso entre a detecção e o falso alarme. Esse método foi, portanto, considerado o mais eficiente dentro dos critérios avaliados.

TABLE I: Tabela de eficiências dos métodos propostos para um falso alarme de aproximadamente 10%

	Detecção (%)	Falso Alarme (%)
OF2, Alg.1	71,396	10,366
MLEG, Alg.1	75,973	10,069
OF2, Alg.2	73,913	10,365
MLEG, Alg.2	78,211	10,068
OF2, Alg.3	87,643	10,363
MLEG, Alg.3	90,847	10,067
OF2, Alg.4	83,982	10,364
MLEG, Alg.4	88,330	10,067

Resultados com dados reais apresentaram-se satisfatórios, confirmando o bom desempenho da técnica proposta. Ainda, pretende-se expandir esta análise utilizando métodos não lineares de estimativa de energia com o objetivo de melhorar ainda mais a eficiência na discriminação de mísseis e viabilizar a implementação do projeto TileMuon no Barril Longo do TileCal.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Também gostaríamos de agradecer a CNPq, RENAFAE, FAPERJ pelo apoio financeiro ao projeto. Agradecemos também ao Experimento ATLAS, em especial ao TileCal, pelo suporte ao desenvolvimento do trabalho.

REFERENCES

- [1] L. Evans and P. Bryant (editors), *LHC Machine*, JINST 3 S08001, 2008.
- [2] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST 3 S08002, 2008.
- [3] The LHCb Collaboration, *The LHCb Detector at the LHC*. INST 3 S08005 (2008) JINST 3 S08004, 2008.
- [4] The CMS Collaboration, *The CMS Experiment at the CERN LHC*. JINST 3 S08004, 2008.
- [5] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*. JINST 3 S08003, 2008.
- [6] The ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, EPJC 70, 2010.
- [7] The ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. CERN, 1997.
- [8] WE. Cleland, E.G. Stern, *Signal processing considerations for liquid ionization calorimeters in a high rate environment*, Nuclear Instruments and Methods in Physics Research A, v. 338, pp.467-497, 1994.
- [9] G. Bertuccio, E. Gatti and M. Sapietro, *Sampling and optimum data processing of detector signals*, Nuclear Instruments and Methods in Physics Research A, v. 322, pp.271-279, 1992.
- [10] K. Kordas, *The ATLAS data acquisition and Trigger: Concept, design and status*, Nuclear Physics B - Proc. Suppl., vol. 172, pp.178-182, 2007.
- [11] B. S. M. Peralva, *Reconstrução de Energia para Calorímetros Finamente Segmentados*, Universidade Federal de Juiz de Fora, 2015.
- [12] H. Li, *Lagrange Multipliers and their Applications*, University of Tennessee, Knoxville, 2008.
- [13] PLACKETT, R. L. *Karl Pearson and the chi-squared test*, International Statistical Review/Revue Internationale de Statistique, 1983.
- [14] JOLLIFE, I. T. *Principal Component Analysis*. New York, Springer-Verlag, 2002.
- [15] Fawcett, T. *An Introduction to ROC Analysis*. Pattern Recognition Letters, 2006.