

Modelos Lineares para a Reconstrução de Sinal do Calorímetro Hadrônico para Operação em Alta Luminosidade

Badilé M. Insali,¹ Bernardo S. Peralva,² Guilherme B. Morett,³ Gustavo B. Libotte⁴
PPGMC/IPRJ, Nova Friburgo, RJ

Resumo. O objetivo desse artigo é discutir a aplicação do método dos mínimos quadrados para a estimação dos parâmetros da fase e amplitude do sinal de resposta do calorímetro hadrônico de telhas cintilantes no contexto da atualização para operação em ambiente de alta luminosidade. Nas condições impostas por esse novo cenário, a reconstrução do sinal deve ser feita de modo contínuo (streaming) e com resposta rápida. Desta forma, o tema central abordado é de como o janelamento das amostras temporais utilizadas como entrada no método dos mínimos quadrados influencia no erro de estimação da amplitude e fase. Utilizando a validação cruzada k-Fold, um parecer é feito sobre o desempenho geral desse método pelo estudo do comportamento da média do desvio padrão, assim como a quantidade de janelamento mais adequada é definida.

Palavras-chave. Modelagem computacional, Método dos Mínimos Quadrados, Reconstrução de sinal, Física experimental de altas energias, Calorimetria de altas energias.

1 Introdução

Grandes experimentos da ciência visam impulsionar avanços científicos e tecnológicos seja de forma direta ou indireta. O laboratório CERN, por exemplo, conta com diversas descobertas e contribuições, tal como o surgimento da *web* [2] que beneficia significativamente a sociedade. O LHC (Large Hadron Collider), localizado no CERN, é o maior colisor de hádrons do mundo, empregando um anel de 27 km de circunferência. Ele acelera e colide partículas usando campos magnéticos, e seus principais experimentos incluem ATLAS, CMS, LHCb e ALICE, cada um focado em uma área específica de pesquisa.

O experimento ATLAS, inspiração deste trabalho, tem por finalidade a interpretação dos dados originários da colisão dos feixes de partículas em altas energias [3], e cobre um amplo programa de física. Inclusive esse experimento teve uma enorme participação na validação da existência do bóson de Higgs em 2012 [4]. O calorímetro, por sua vez, possui diversos módulos de detecção capazes de absorver a energia emitida após as colisões e transformar a energia que é depositada nos canais de leitura em sinais que podem ser posteriormente manipulados e analisados.

Os conceitos de reconstrução de sinal são empregados para que, a partir dos pulsos captados pela eletrônica do equipamento, a curva que representa esse fenômeno físico seja construída por meio da estimação dos parâmetros de amplitude, fase e pedestal. A quantidade de dados produzidos por tais colisões é muito grande, haja vista que a taxa de colisão do feixe de prótons é em torno de 25 ns. Para fins de estimação de energia, em geral é suficiente a captura de uma única

¹badile.insali@iprj.uerj.br

²bernardo@iprj.uerj.br

³guilherme.morett@iprj.uerj.br

⁴gustavolibotte@iprj.uerj.br

amostra no momento do pico de um sinal. Em canais de leitura com alta ocupação, entretanto, colisões consecutivas dentro da mesma janela de leitura são comuns. No entanto, devido ao tempo de desvio estendido, múltiplos pulsos podem se sobrepor, especialmente em cruzamentos de feixes consecutivos, resultando no problema de empilhamento de sinal no tempo, o que degrada a resolução de energia [3].

O aumento da ocupação do detector, conhecido como maior luminosidade, intensifica o efeito de empilhamento de sinal, trazendo desafios consideráveis. Esse fenômeno distorce os pulsos de sinal, prejudicando a precisão na estimativa da amplitude (energia) e gerando erros sistemáticos na determinação da energia depositada nas células do calorímetro. O foco do estudo, portanto, é demonstrar como estratégias de reconstrução de energia podem mitigar tais efeitos de empilhamento.

A adoção de modelos matemáticos é fundamental para conduzir tarefas de reconstrução de energia neste contexto, de modo que aspectos físicos subjacentes ao experimento possam ser investigados. Um deles é o algoritmo de Filtragem Ótima (OF, do inglês Optimal Filtering), proposto por Fulana *et al.* [5] e que atualmente está em operação no experimento, assume o conhecimento da forma de um pulso de referência e envolve a minimização da variância do ruído eletrônico.

Por sua vez, neste trabalho, o método que será explorado, como alternativa ao OF é o de Mínimos Quadrados (LS, do inglês Least Squares) [8]. Esta técnica estatística tem por finalidade minimizar o resíduo quadrático entre estimativas do modelo e dados observados. Neste contexto, o objetivo deste trabalho é empregar o método LS na estimação de energia depositada em um calorímetro com condições de luminosidade (e, portanto, de empilhamento de sinal) diversas. Com isso, visamos mostrar como este estimador linear pode ser efetivo, considerando os cenários estabelecidos nesta análise, mesmo em condições de empilhamento severas.

2 Configuração de Dados e Metodologia

O calorímetro é um dispositivo capaz de detectar as energias oriundas de colisões entre partículas subatômicas com elevada sensibilidade de leitura e eficiência de medição [9]. Para tal feito, ele possui em sua constituição um material absorvedor com o intuito de assimilar a partícula e outro ativo para amostrar parcialmente essa energia depositada. Os calorímetros são geralmente caracterizados por serem altamente segmentados. Divisões transversais permitem a captação das direções e da energia das partículas advindas das colisões. A energia depositada diminui gradualmente nas partes mais internas do calorímetro. O sistema de eletrônica é responsável por converter essa energia em amostras discretas de sinais.

Dados que simulam um experimento típico do calorímetro de telhas foram obtidos usando um gerador de sinais [6] para diferentes níveis de ocupação. A amplitude segue uma distribuição exponencial com média $1/50$, já a fase apresenta uma distribuição normal padronizada. O valor considerado para o pedestal é 40. Estão disponíveis 100.000 dados sequenciados, que relacionam o tempo (ns), o pulso de sinal (contagens de ADC), a amplitude (contagens de ADC) e a fase (ns) correspondentes.

Como os dados são dispostos sequencialmente, é necessário agrupar sinais adjacentes em “janelas” de dados. O tamanho das “janelas” corresponde ao número de amostras consideradas na reconstrução de sinal a partir dos dados sequenciados. Visando avaliar como o tamanho das janelas afeta o desempenho das reconstruções de sinal, adotamos tamanhos dados por números inteiros ímpares no intervalo [7; 19]. Janelas com números ímpares de amostras permitem capturar a amplitude e fase centrais, para posteriormente serem tomadas como alvo da comparação. Por ilustração, considere uma janela com sete elementos. De acordo com essa estratégia, os sete primeiros pulsos de sinais formam um lote. Cada amostra ou lote de pulso de sinal está associada a uma amplitude e fase real. Para esse lote, os elementos centrais da amplitude e fase serão considerados como o

valor alvo, conforme é mostrado na Figura 1.

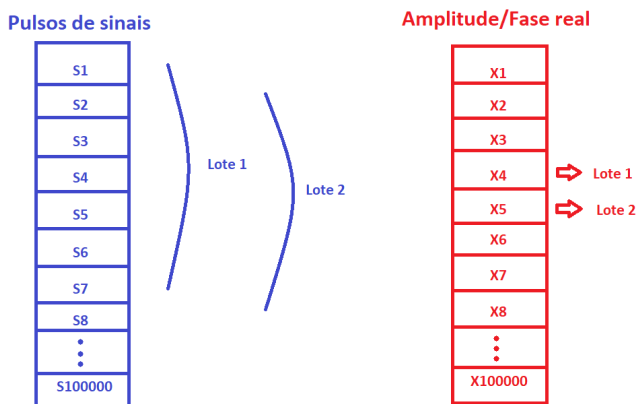


Figura 1: Visualização da técnica de janelamento dos dados de entrada. Fonte: dos autores.

A partir desse rearranjo dos dados de entrada, a matriz \vec{H} e o vetor \vec{x} são definidos. As linhas da matriz \vec{H} são formadas pelos lotes dos pulsos de sinais. Uma coluna unitária é incorporada, para que seja acrescentado o termo de interseção ou viés nos cálculos ao realizar a soma ponderada no final do processo de estimação. Geometricamente seria como se o hiperplano não estivesse centrado na origem e o deslocamento correspondesse a essa constante. Por outro lado, os elementos do vetor \vec{x} representam os dados reais associados a cada lote. Portanto, tem-se um vetor \vec{x} para a amplitude e outro para a fase real, nos experimentos de estimação de cada uma destas quantidades que conduzimos neste trabalho. Note que dados de amplitude e fase fora dos limites do primeiro e último lotes em um esquema de janelamento arbitrário são descartados, sem prejuízos. A adoção de \vec{x} como vetor que contém amplitudes ou fases será explicitada em cada análise distinta.

Os dados são, então, divididos entre conjunto de treinamento e de teste. No contexto deste trabalho, o treinamento refere-se à reconstrução da amplitude ou fase de um sinal, enquanto que o teste refere-se à avaliação do erro relacionado à amplitude ou fase alvos. Denotamos como \vec{H}_{treino} e \vec{x}_{treino} os dados de treinamento para a matriz de amostras de sinal e da amplitude/fase, respectivamente. De forma análoga, os dados de teste são denotados como \vec{H}_{teste} e \vec{x}_{teste} . Depois desta divisão, os pesos presentes no vetor $\vec{\theta}$ são calculados de acordo com

$$\vec{\theta} = ((H_{\text{treino}}^{\top} \cdot \vec{H}_{\text{treino}})^{-1} \cdot H_{\text{treino}}^{\top}) \cdot \vec{x}_{\text{treino}}. \tag{1}$$

E, por fim, podemos calcular o vetor do erro absoluto \vec{e} do parâmetro estimado, também chamado de resíduo, por meio de

$$\vec{e} = \vec{x}_{\text{teste}} - (\vec{H}_{\text{teste}} \cdot \vec{\theta}). \tag{2}$$

Para mais detalhes sobre o método, consulte Kay [7].

A validação cruzada k -fold possibilita explorar melhor as informações contidas nos dados de entrada, de modo que os resultados alcançados possam ser generalizados e erros possam ser estimados. Essa técnica divide os dados em k dobras, com blocos de dados de mesmo tamanho: na primeira dobra, o primeiro bloco de dados é adotado como sendo o de teste e o restante para treinamento; na segunda dobra, o segundo bloco é tomado para teste e o restante para treinamento; e assim sucessivamente. Isso permite que o mesmo conjunto de dados seja usado para gerar k conjuntos de resultados diferentes. Pode-se, então, avaliar a média e o desvio-padrão dos resultados alcançados para as k dobras.

3 Resultados

A seguir serão analisados resultados para reconstruções da fase e da amplitude, considerando os aspectos da Seção 2. Para ambos os casos, $k = 10$ no procedimento de validação cruzada.

3.1 Estimação da Amplitude

Inicialmente, estamos interessado em como o tamanho da janela pode influenciar o desempenho do método LS na estimação da amplitude. A Figura 2 mostra a média para os valores de desvio-padrão dos erros calculados pela Equação (2), para dados sujeitos a diferentes níveis de ocupação. Para os resultados correspondentes a cada tamanho de janela, as barras de erro representam o desvio-padrão dos resultados obtidos entre as $k = 10$ dobras.

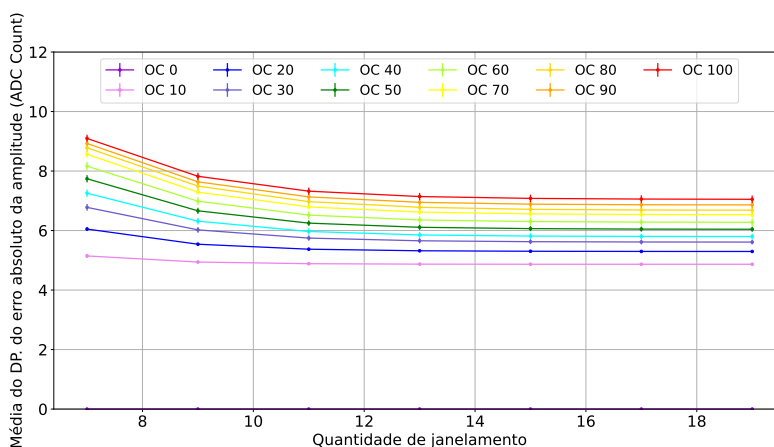


Figura 2: Média do desvio-padrão do erro absoluto da amplitude. Fonte: dos autores.

Analisando estes resultados, inicialmente nota-se que, à medida que a ocupação aumenta, os valores médios do desvio-padrão do erro tornam-se proporcionalmente maiores para janelas de mesmo tamanho. Isto fica evidente ao observar-se que curvas associadas a cores quentes que indicam ocupações mais elevadas estão deslocadas para cima em relação às demais com ocupação inferior. Este fato é esperado pois o efeito do empilhamento de sinais se torna mais contundente quando os canais de leitura estão mais ocupados.

As curvas dos valores médios do desvio padrão para todas as ocupações seguem uma tendência semelhante: os valores diminuem com o aumento do tamanho da janela. Com o aumento gradual do tamanho da janela, tal redução é observada até, aproximadamente, tamanhos iguais a 15. Após esta configuração, fica claro que tamanhos de janela superiores não representam ganho significativo relacionado ao valor médio do desvio padrão do erro. Vale ressaltar que as barras de erro apresentam grandezas irrelevantes, o que significa não haver desvio significativo do valor médio obtido na validação cruzada. Além disso, é importante salientar que o valor médio da média do erro absoluto da amplitude, para todas as janelas e condições de ocupação analisadas, fica centrado em zero, com desvio padrão, no máximo, aproximadamente igual a 0,3 no caso de maior dispersão (tamanho de janela igual a sete). A Figura 3 exhibe o comportamento do histograma.

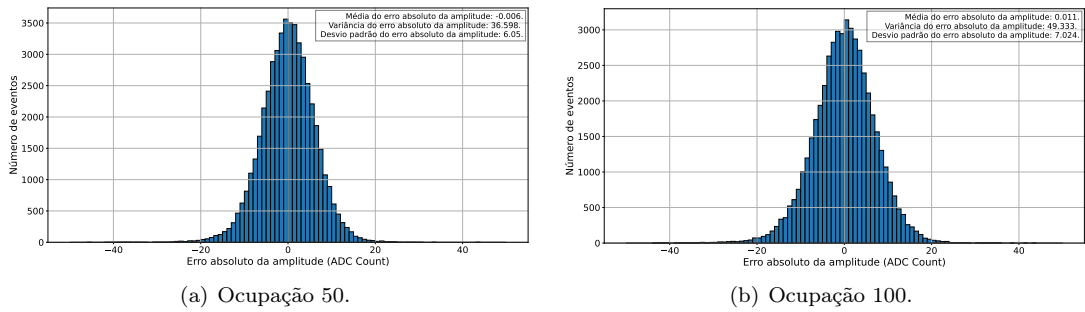


Figura 3: Exemplos de histogramas para o erro absoluto da amplitude para o janelamento 15. Fonte: dos autores.

Com foco no janelamento ideal 15, a Figura 4 apresenta o comportamento da média do desvio padrão ao longo das ocupações.

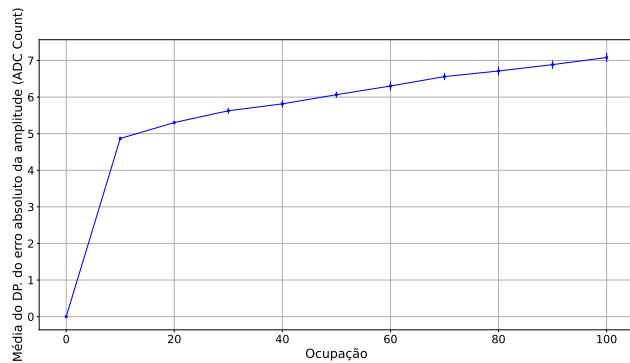


Figura 4: Média do desvio padrão do erro absoluto da amplitude ao longo das ocupações para o janelamento 15. Fonte: dos autores.

3.2 Estimação da Fase

Os mesmos procedimentos do método LS foram realizados para o parâmetro da fase. A Figura 5 denota a evolução da média do desvio padrão ao longo das ocupações e janelamentos.

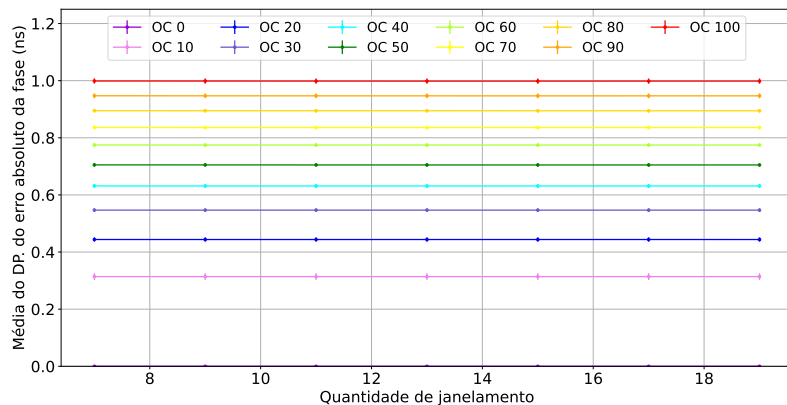


Figura 5: Média do desvio padrão do erro absoluto da fase. Fonte: dos autores.

A Figura 6 representa os histogramas obtidos para o janelamento 15.

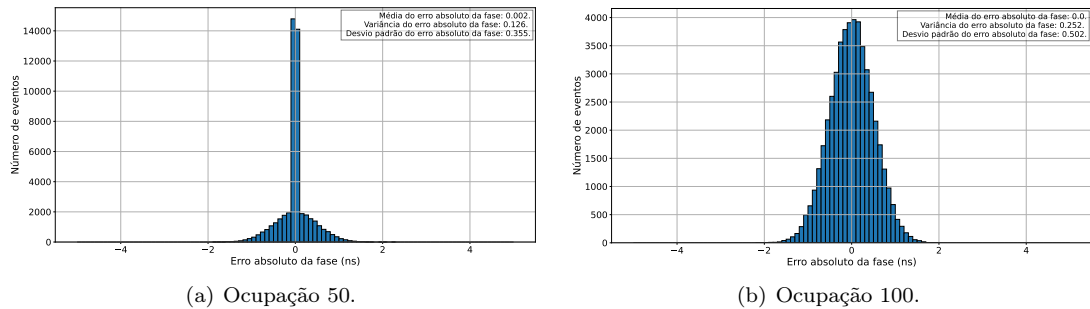


Figura 6: Exemplos de histogramas para o erro absoluto da fase para o janelamento 15. Fonte: dos autores.

Pela análise numérica dos dados usados no `plot`, foi constatado que o a média do desvio padrão para a fase tem menor sensibilidade em relação a quantidade de janelamento adotada. O impacto do janelamento está somente a partir da terceira casa decimal. No entanto, novamente a partir do janelamento 15, o valor da média do desvio padrão pode ser considerado constante, pois a variação passa ser na quarta casa decimal. A Figura 7 demonstra a evolução da média do desvio padrão ao longo das ocupações para o janelamento ideal 15.

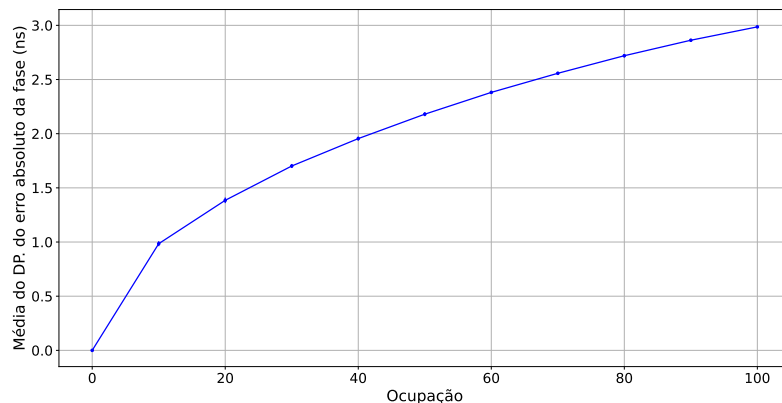


Figura 7: Média do desvio padrão do erro absoluto da fase ao longo das ocupações para o janelamento 15. Fonte: dos autores.

4 Considerações Finais

Este trabalho abordou o tema de estimação de parâmetros em calorímetros operando em alta taxa de eventos. Devido a sua simplicidade e resposta rápida, o Método de Mínimos Quadrados se apresenta como uma alternativa promissora para estimação da energia e fase dos sinais do calorímetro de telhas cintilantes nas condições de alta luminosidade. Por meio da interpretação gráfica e numérica constatou-se que quinze amostras são suficientes para estimar os parâmetros de interesse do pulso (amplitude e fase). Para futuros projetos busca-se empregar o simulador Lorenzetti [1] que permitirá uma análise de eficiência com parâmetros da física tais como a resolução

de reconstrução de uma partícula e capacidade de discriminação de partículas. Finalmente, é previsto testes com modelos não lineares, tais como as redes neurais de aprendizado profundo, visando verificar eficiência de estimação da amplitude e fase, e a factibilidade de implementação na eletrônica atualizada do calorímetro hadrônico.

Agradecimentos

Os autores agradecem a FAPERJ, CNPq, CAPES, RENAFEA, e FAPEMIG pelo apoio. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- [1] M.V. Araújo et al. “Lorenzetti Showers - A general-purpose framework for supporting signal reconstruction and triggering with calorimeters”. Em: **Computer Physics Communications** 286 (2023). Aceito, p. 108671. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2023.108671.
- [2] CERN. **Contribute to society**. Online. Acessado em 04/11/2024, <https://home.cern/about/what-we-do/our-impact>.
- [3] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. Em: **Journal of Instrumentation** (2008). Aceito. DOI: 10.1088/1748-0221/3/08/S08003.
- [4] The ATLAS Collaboration. “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery”. Em: **Nature** 607 (2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w.
- [5] E. Fullana et al. **Optimal filtering in the ATLAS hadronic tile calorimeter**. Online. Acessado em 04/11/2024, <https://cds.cern.ch/record/816152/files/tilecal-2005-001.pdf>[.pdf?version=1](https://cds.cern.ch/record/816152/files/tilecal-2005-001.pdf?version=1).
- [6] G. Gonçalves. **Calorimetry Pulse Simulator**. Online. Acessado em 04/11/2024, <https://github.com/ingoncalves/calorimetry-pulse-simulator/tree/master>.
- [7] S. M. Kay. “Least Squares”. Em: **Fundamentals of Statistical Signal Processing: Estimation Theory**. Vol. 1. Prentice Hall, 1993. Cap. 8, pp. 219–286.
- [8] M. Merriman. “On the history of the method of least squares”. Em: **The Analyst** 2 (1877), pp. 33–36. DOI: 10.2307/2635472.
- [9] S. M. Rimes et al. “Estimação da amplitude de sinais em calorimetria de altas energias em condições de alta ocupação de eventos no detector”. Em: **XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais** (2020). Aceito. DOI: 10.14209/SBRT.2020.1570658901.