

Explorando o invisível: análise do decaimento do bóson Z em neutrinos com dados abertos do CMS

Exploring the Invisible: Analysis of the Z Boson Decay into Neutrinos with CMS Open Data

Maria Gabriela F.S.A. Gomes^{1,2}, Patricia Rebello Teles^{*3}, Gilvan Augusto Alves³,
Diego A. Cogollo Aponte²

¹Universidade de Heidelberg, Heidelberg, Alemanha.

²Universidade Federal de Campina Grande, Departamento de Física, Campina Grande, PB, Brasil.

³Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ, Brasil.

Recebido em 10 de outubro de 2025. Revisado em 07 de fevereiro de 2026. Aceito em 11 de março de 2026.

Este artigo apresenta uma proposta didática de análise do decaimento do bóson Z em neutrinos, conhecido como a largura invisível do bóson Z relacionado à chamada largura invisível do bóson Z , um parâmetro fundamental do Modelo Padrão. A análise é realizada com dados reais do experimento Solenóide Compacto de Múons (CMS), localizado no Grande Colisor de Hádrons (LHC), correspondentes a colisões próton-próton com energia de centro de massa $\sqrt{s} = 13$ TeV e luminosidade integrada de $16,290 \text{ fb}^{-1}$. O objetivo desse trabalho é reproduzir a análise realizada pela Colaboração CMS e oferecer não apenas um guia prático de análise de dados reais, mas também discutir conceitos teóricos fundamentais.

Palavras-chave: Bóson Z , Largura invisível, Modelo Padrão, LHC, CMS Open Data.

This article presents a didactic proposal for the analysis of the decay of the Z boson into neutrinos, known as the invisible width of the Z boson, a fundamental parameter of the Standard Model. The analysis is performed using real data from the CMS experiment at the LHC, with $\sqrt{s} = 13$ TeV and integrated luminosity of $16,290 \text{ fb}^{-1}$. The goal is to reproduce CMS results and provide a practical guide to data analysis, as well as a discussion of the underlying theoretical concepts.

Keywords: Z boson, Invisible width, Standard Model, LHC, CMS Open Data.

1. Introdução

O armazenamento, a preservação e a disponibilização de dados científicos são prioridades do CERN (Organização Europeia para a Pesquisa Nuclear). Nesse contexto, o CMS Open Data [1], lançado em 2014, representa uma iniciativa pioneira ao tornar públicos conjuntos de dados de colisões próton-próton e de íons pesados registrados pelo experimento Solenóide Compacto de Múons (CMS), no Grande Colisor de Hádrons (LHC). A disponibilização desses dados tem múltiplas finalidades: permite a reinterpretação de resultados, impulsiona o desenvolvimento de novos métodos de análise e, sobretudo, oferece oportunidades valiosas para a educação e a divulgação científica [2]. Nosso objetivo é, então, apresentar um guia prático de análise de dados do *CMS Open Data*, que reproduz a medida da largura invisível do bóson Z publicada pela Colaboração CMS [3]. Essa medida, essencial no âmbito do Modelo Padrão (MP) da Física de Partículas, fornece informações diretas sobre o número de espécies de neutrinos existentes no universo.

O trabalho aqui apresentado tem caráter didático: busca ilustrar de forma completa o fluxo de uma análise em física experimental de partículas em altas energias utilizando dados reais do *CMS Open Data*. O fluxo da análise segue uma sequência lógica de etapas. Inicialmente, realiza-se a importação dos dados abertos do CMS, seguida pela pré-seleção dos eventos com base em critérios cinemáticos básicos. Em seguida, definem-se as regiões de interesse da análise – regiões de sinal e de controle – que permitem separar contribuições relevantes de diferentes processos físicos. Com base nessas regiões, aplicam-se fatores de normalização e correções sistemáticas, conduzindo à construção de uma função de verossimilhança que modela estatisticamente os dados observados. Esta função é então utilizada em um ajuste Bayesiano, implementado via amostragem por Cadeias de Markov Monte Carlo (MCMC), utilizando o algoritmo *emcee*, para estimar os parâmetros de interesse, em particular, a largura invisível do bóson Z . Por fim, os resultados obtidos são interpretados à luz do Modelo Padrão e comparados com as medições experimentais previamente publicadas. Todo o código desenvolvido para a análise está disponível em repositório público no GitHub [4], garantindo a reprodutibilidade dos resultados.

*Endereço de correspondência: patricia.rebello.teles@cern.ch
Editor-Chefe: Marcello Ferreira

2. O Modelo Padrão da Física de Partículas

O Modelo Padrão das interações elementares (MP) [5–7] é uma teoria estruturada no conceito de invariância de gauge, assim como a eletrodinâmica quântica (QED)¹[8] (em ambos os casos o interesse é construir uma lagrangiana invariante de gauge). A finalidade do MP é descrever a forma como as partículas elementares interagem através das diferentes interações fundamentais da natureza:

- **Eletromagnética:** é a interação à qual as partículas carregadas eletricamente, como elétrons e prótons, são sensíveis. Do ponto de vista clássico, a interação eletromagnética é descrita em função de campos elétricos e magnéticos que dependem do tempo, que interagem com as partículas carregadas e ao mesmo tempo são gerados por elas. Do ponto de vista quântico, a interação eletromagnética acontece pela troca de fótons γ (os quantas do campo eletromagnético, partículas de massa zero). Os fenômenos de emissão e absorção de luz, a geração e transmissão de energia usada nas nossas residências, e a estabilidade dos átomos da matéria da qual somos compostos, são todos fenômenos eletromagnéticos.
- **Fraca:** é a interação através da qual acontecem processos de decaimentos nucleares, como o *decaimento beta*, no qual um nêutron se transforma em um próton dentro de um núcleo atômico, emitindo um elétron e um antineutrino do elétron para fora deste. Os mediadores das interações fracas são os bósons W^+ , W^- , Z , partículas com massa diferente de zero e centenas de vezes mais pesadas do que o próton, $\sim 100m_p$.
- **Forte:** do ponto de vista da física nuclear, a interação forte é uma interação à qual os núcleons (nêutrons e prótons) estão sujeitos dentro de um núcleo atômico através da troca de mésons pí, mantendo este sistema quântico coeso. Do ponto de vista mais fundamental, a interação forte acontece através da troca de glúons entre quarks. Os quarks, partículas elementares que constituem os núcleons, interagem com uma grande intensidade nessa escala espacial, de forma que não podem ser observados livres na natureza.

A quarta força fundamental, a **gravitação**, não está incluída na estrutura algébrica do MP e é descrita pela Teoria da Relatividade Geral de Einstein.

2.1. Partículas elementares

As partículas elementares do Modelo Padrão dividem-se em dois grandes grupos [10]:

¹ Para revisar o conceito de invariância de gauge no eletromagnetismo e em física de partículas, o leitor pode consultar material sobre teoria quântica de campos como F. Mandl [9].

• Férmions (partículas de matéria)

- **Quarks:** existem em seis estados de sabor² (up u , down d , charm c , strange s , top t e bottom b). Eles nunca aparecem isolados, mas se combinam para formar hádrons, como os bárions (p.ex.: prótons e nêutrons) e os mésons.
- **Léptons:** os léptons são partículas elementares carregadas e neutras. Os léptons carregados existem em três estados de sabor: elétron e^- , múon μ^- , tau τ^- . Para cada um destes três estados de sabor, existe um correspondente estado de sabor eletricamente neutro, chamado de neutrinos (ν_e, ν_μ, ν_τ).

Tanto os quarks quanto os léptons estão organizados em três gerações. A primeira geração de quarks (u, d)^T e de léptons (ν_e, e^-)^T é a constituinte da matéria estável. As outras famílias de quarks e léptons são mais pesadas e instáveis, e decaem rapidamente nos membros da primeira geração.

• Bósons (partículas mediadoras das forças fundamentais)

Como citado na Seção 2, os bósons são as partículas que medeiam as interações fundamentais entre as partículas elementares. Cada interação fundamental é descrita de forma consistente pelo conceito de troca de algum tipo específico de bóson, com diferentes alcances, spin, carga elétrica e massa. Os bósons que descrevem consistentemente as interações do MP, são:

- Fóton (γ): mediador da interação eletromagnética.
- Glúons (g): mediadores da interação forte.
- Bósons W^+ , W^- e Z : mediadores da interação fraca.
- Bóson de Higgs (H): é uma partícula escalar, remanescente do processo da quebra espontânea de simetria (mecanismo de Higgs). O mecanismo de Higgs foi confirmado experimentalmente em 2012 com a descoberta do bóson de Higgs no grande colisor de hádrons (LHC) [11, 12].

2.2. Vida média e largura de decaimento

Assim como em física nuclear os núcleos instáveis decaem em núcleos mais leves à procura de estabilidade, em um processo conhecido como radioatividade [13], algumas partículas elementares e hádrons são também instáveis e decaem em outras partículas mais leves. O tempo de vida médio (τ) de uma partícula está relacionado à sua

² O sabor é um número quântico que identifica uma determinada partícula. Cada sabor representa uma partícula diferente dentro de uma mesma espécie, como os sabores de quarks, léptons carregados e neutrinos.

largura de decaimento (Γ)³ pela expressão:

$$\tau = \frac{\hbar}{\Gamma}, \quad (1)$$

sendo $\hbar = \frac{h}{2\pi}$ e h a constante de Planck. Quanto maior a largura Γ , mais rapidamente ocorre o decaimento.

A largura total de decaimento é a soma das larguras de todos os possíveis canais de decaimento i :

$$\Gamma_{Total} = \sum_i \Gamma_i. \quad (2)$$

Por exemplo, o bóson Z [14] pode decair em pares de quarks, em pares de léptons carregados ou em pares de neutrinos. A fração de decaimentos em um canal específico é chamada de taxa de ramificação:

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma}, \quad (3)$$

onde Γ_i representa a largura de decaimento parcial correspondente a um canal específico, enquanto Γ é a largura total de decaimento da partícula. Assim, $\mathcal{B}_i$ indica a probabilidade de que o decaimento ocorra naquele canal em particular em relação ao número total de decaimentos possíveis.

2.3. O número de neutrinos

No Modelo Padrão, a largura invisível do bóson Z se deve unicamente aos neutrinos⁴ [15]. Por este motivo, podemos definir o número efetivo de espécies de neutrinos leves como:

$$N_\nu = \frac{\Gamma_{inv}}{(\Gamma_{\nu\bar{\nu}})_{MP}}, \quad (4)$$

sendo Γ_{inv} a largura invisível do bóson Z , e $(\Gamma_{\nu\bar{\nu}})_{MP}$ a largura de decaimento do bóson Z para cada uma das espécies leves de neutrinos [16]. A expressão Eq. (4) pode ser escrita na forma:

$$N_\nu = \frac{\Gamma_{inv}}{\Gamma_{\ell\bar{\ell}}} \left(\frac{\Gamma_{\ell\bar{\ell}}}{\Gamma_{\nu\bar{\nu}}} \right)_{MP} = R^{exp} \left(\frac{\Gamma_{\ell\bar{\ell}}}{\Gamma_{\nu\bar{\nu}}} \right)_{MP}, \quad (5)$$

onde $\Gamma_{\ell\bar{\ell}}$ é a largura de decaimento do bóson Z em um par de léptons carregados, e $\left(\frac{\Gamma_{\ell\bar{\ell}}}{\Gamma_{\nu\bar{\nu}}} \right)_{SM}$ corresponde ao valor teórico previsto pelo Modelo Padrão.

Dessa forma, o valor de N_ν passa a depender do valor experimental $R^{exp} = 5.942 \pm 0.016$ [17] e do

³ A largura de decaimento Γ mede o quão instável é um estado quântico. Estados com largura pequena têm energia bem definida e tendem a durar mais tempo antes de decair. Já estados com largura grande apresentam maior incerteza na energia e decaem rapidamente. Assim, a largura de decaimento pode ser interpretada como uma medida inversa do tempo de vida do estado: quanto maior Γ , menor é o tempo de vida.

⁴ Se deve unicamente aos neutrinos pois o decaimento em neutrinos não deixa rastro direto nos detectores, em contrapartida, os outros produtos do seu decaimento deixam registros diretos.

valor previsto pelo MP $\left(\frac{\Gamma_{\nu\bar{\nu}}}{\Gamma_{\ell\bar{\ell}}} \right)_{SM} = 1.991 \pm 0.001$, de forma tal que $N_\nu = 2.984 \pm 0.008$. Conclui-se assim que existem três tipos de neutrinos leves no universo, ν_e , ν_μ e ν_τ , fortalecendo a escolha teórica das três gerações de partículas no MP.

3. O Experimento CMS

O Grande Colisor de Hádrons (LHC) [18] é o maior acelerador de partículas do mundo, localizado no CERN, na fronteira entre a Suíça e a França. Ele possui 27 km de circunferência e está instalado a cerca de 100 metros de profundidade. No LHC, feixes de prótons ou íons pesados são acelerados próximos à velocidade da luz em sentidos opostos e colididos em pontos específicos do anel, atingindo energias de até 13 TeV no centro de massa.

O experimento Solenóide Compacto de Múons (CMS) [19] está localizado no LHC; é um detector construído em camadas concêntricas ao redor do ponto de colisão. Cada camada foi projetada para medir um tipo específico de partícula, Figura (1):

- **Rastreador de silício:** detector de rastreamento baseado em silício que registra os pontos de passagem de partículas carregadas, permitindo reconstruir suas trajetórias e identificar vértices de interação.
- **Calorímetro eletromagnético (ECAL):** mede a energia de elétrons e fótons.
- **Calorímetro hadrônico (HCAL):** mede a energia de hádrons (como prótons, nêutrons e píons).
- **Sistema de múons:** localizado na região mais externa do detector, é dedicado à detecção de múons, partículas altamente penetrantes que atravessam as demais camadas.

No centro do detector, um solenóide supercondutor gera um forte campo magnético que curva as trajetórias de partículas carregadas. A partir dessa curvatura, o raio permite determinar o momento das partículas, enquanto o sentido da curvatura indica o sinal de sua carga.

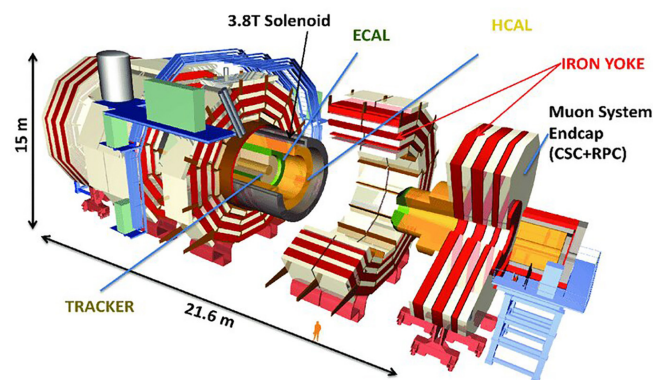


Figura 1: Visão esquemática do detector CMS. Fonte: [21].

Os neutrinos não interagem diretamente com os detectores do CMS. Assim, quando o bóson Z decai em pares de neutrino–antineutrino, os produtos do decaimento não deixam sinais diretos no detector, sendo essa contribuição denominada *largura invisível*. Aproximadamente 20% dos decaimentos do Z resultam em pares neutrino–antineutrino [20]. Esse decaimento é inferido por meio do *momento transverso ausente* (p_T^{miss}). Em um colisor como o LHC, a direção dos feixes define o eixo z . Embora os prótons sejam partículas compostas, o momento total inicial do sistema, na direção perpendicular aos feixes (plano transversal), é aproximadamente nulo. Isso ocorre porque os prótons colidem frontalmente e não possuem, em média, movimento transversal líquido antes da interação. Pela conservação do momento linear, a soma dos momentos das partículas produzidas após a colisão também deve ser nula no plano transversal. Assim, se a soma vetorial dos momentos transversos das partículas detectadas for diferente de zero, interpreta-se essa diferença como um momento transverso ausente, atribuído a partículas que escaparam à detecção – como os neutrinos. O momento dessas partículas invisíveis deve compensar exatamente o desequilíbrio observado, restaurando a conservação do momento total,

$$\vec{p}_T^{\text{miss}} + \sum_i \vec{p}_T^i = \vec{0}, \quad (6)$$

onde a soma percorre todas as partículas visíveis reconstruídas. O módulo

$$p_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}| \quad (7)$$

indica a quantidade de momento carregado pelas partículas invisíveis, como os neutrinos.

Uma forma intuitiva de compreender esse conceito é imaginar um jogo de bilhar: se duas bolas colidem em uma mesa sem atrito, o momento linear total deve ser o mesmo antes e depois da colisão. Se uma das bolas “desaparecesse”, o movimento das demais revelaria que algo invisível levou parte do momento. De modo análogo, no detector do CMS, o p_T^{miss} é uma pista da passagem de neutrinos.

4. Metodologia

4.1. Estratégia de análise

Para determinar a largura invisível do bóson Z , seguimos a mesma estratégia apresentada no artigo da colaboração CMS [3]. Seleccionamos eventos em que o bóson Z é produzido em associação com jatos e decai posteriormente em neutrinos ($Z \rightarrow \nu\bar{\nu}$) ou em um par de léptons carregados ($Z \rightarrow \ell^+\ell^-$). Como os neutrinos não interagem diretamente com o detector, sua presença é inferida de forma indireta a partir do *momento transverso ausente* (p_T^{miss}). Dessa forma, a assinatura característica

da região de sinal é composta por p_T^{miss} acompanhada de jatos⁵.

Como referência, utiliza-se o decaimento do bóson Z em léptons carregados ($Z \rightarrow \ell^+\ell^-$, com $\ell = e, \mu$). Esse canal é útil porque apresenta propriedades cinemáticas semelhantes ao decaimento invisível em neutrinos, mas com produtos finais detectáveis pelo CMS. Assim, é possível estabelecer uma comparação direta e controlada entre os dois processos [3].

Neste contexto, *backgrounds* referem-se a processos do Modelo Padrão que imitam a assinatura experimental do sinal esperado, mas não correspondem ao decaimento invisível do bóson Z . A principal fonte de *background* dessa análise é o decaimento do bóson W em um lépton e um neutrino ($W \rightarrow \ell\nu$), que também gera eventos com p_T^{miss} . Outros processos, como multijatos produzidos pela interação forte (QCD), também podem imitar o sinal. Isso ocorre quando a energia depositada pelos jatos no detector é reconstruída de forma assimétrica, seja por falhas de medição, perdas instrumentais ou por jatos emitidos fora da região coberta pelo detector. Nesses casos, o desequilíbrio de momento transverso pode gerar um sinal de p_T^{miss} que simula a presença de neutrinos.

Para separar esses diferentes processos, são definidas **regiões de controle**. Elas funcionam como “laboratórios auxiliares”, onde observamos os dados em situações específicas para entender melhor cada contribuição:

- **Regiões de um único múon, um único elétron e um único tau:** usadas para medir os eventos de $W \rightarrow \ell\nu$.
- **Regiões de dois múons e dois elétrons:** permitem medir diretamente o canal de referência $Z \rightarrow \ell^+\ell^-$.
- **Região QCD:** usada para estimar os eventos de multijatos da interação forte.

Com base nessas regiões de controle, calculamos os chamados *fatores de transferência*, que são como “pontes matemáticas” entre o que é observado nas regiões de controle e o que se espera na região de sinal. Isso nos permite estimar de forma confiável a contribuição de cada processo [3].

Finalmente, realizamos um **ajuste simultâneo**, ou seja, um procedimento estatístico que combina todas as informações (sinal e controles) em um único cálculo. Dessa forma, obtemos uma medida mais robusta e precisa da largura invisível do bóson Z [3].

4.2. Datasets

Os arquivos utilizados nesta análise foram obtidos no portal *CMS Open Data* [22]. Eles correspondem a colisões próton–próton com energia de centro de massa de

⁵ Em física de altas energias, jatos são feixes concentrados de partículas que emergem de uma colisão, resultantes da hadronização de quarks e glúons.

Tabela 1: Conjuntos de dados primários usados para a seleção das diferentes regiões de análise.

Datasets Primários	Regiões
MET	Jatos + p_T^{miss}
	único múon
	dois múons
	único tau
	QCD
Single Muon	único múon
	dois múons
Single Electron	único elétron
	dois elétrons

$\sqrt{s} = 13$ TeV, registradas pelo experimento CMS no ano de 2016.

Os dados são disponibilizados no formato **NanoAOD**, que é uma versão compacta e otimizada dos arquivos originais. Esse formato contém apenas as informações essenciais sobre as partículas reconstruídas (como múons, elétrons, fótons, jatos e a energia ~~faltante~~ transversal ausente – MET). Graças a essa compactação, o formato **NanoAOD** é ideal para análises em larga escala, pois facilita a leitura rápida e o uso direto com ferramentas modernas de processamento de dados, como os pacotes **uproot** [23] e **awkward-array** [24], amplamente utilizados em Python.

A Tabela 1, apresenta as regiões de controle associadas a cada *dataset*⁶.

Durante a coleta de dados do experimento CMS, registra-se a informação sobre a qualidade de cada período de aquisição em um arquivo de texto padronizado no formato JSON. Esse arquivo permite, na fase de análise, selecionar apenas os eventos obtidos em períodos considerados válidos, garantindo que sejam utilizados somente dados certificados pelos detectores.

No portal *CMS Open Data* [22] estão disponíveis para a nossa análise apenas os dados de 2016 correspondentes às chamadas eras G e H, que são períodos específicos de coleta de dados e representam uma fração do total obtido no ano. Esses períodos incluem os conjuntos de dados numerados de 278820 à 284044.

Além dos dados reais, esta análise também utilizou simulações de Monte Carlo (MC) para modelar os processos de sinal e de *background*, disponíveis no mesmo portal *CMS Open Data*⁷. Foram incluídos os seguintes processos:

- Sinal: $Z \rightarrow \nu\bar{\nu}$ (Tabela S2 no apêndice A),
- Sinal de referência: $Z \rightarrow \ell^+\ell^-$ (Tabela S2 no apêndice A),

⁶ Para esta análise, foram utilizados os conjuntos de dados de colisão listados no apêndice A na Tabela S1.

⁷ As simulações foram feitas pela colaboração CMS usando o gerador de eventos Monte Carlo **MadGraph** [26], com o intuito de calcular a seção de choque do processo de interesse, e com o **Pythia** [27], responsável pela evolução do chuveiro partônico, hadronização e decaimentos de partículas instáveis.

Tabela 2: Conjuntos de dados primários e seus correspondentes *triggers*.

Datasets Primários	<i>triggers</i>
	HLT_PFMET170_HBHECleaned OU
	HLT_PFMET170_HBHE OU
MET	HLT_PFMETNoMu120 OU
	HLT_MET75_IsoTrk50
Single Muon	HLT_IsoMu24 OU HLT_IsoTkMu24
Single Electron	HLT_Ele27_WPTight_Gsf

- *background* principal: $W + \text{jatos}$ (Tabela S3 no apêndice A),
- *Backgrounds*: $Diboson$, $t\bar{t}$ (Tabela S4 no apêndice A),
- QCD multijatos (Tabela S4 no apêndice A),
- *Backgrounds* menos relevantes: espalhamento de bósons vetoriais (VBS), *single top* (Tabela S4 no apêndice A).

Os nomes das amostras listados nas Tabelas citadas correspondem à nomenclatura com a qual esses conjuntos de dados estão disponibilizados no portal *CMS Open Data* [22].

4.3. Trigger e luminosidade

O *trigger* é um sistema de seleção de eventos em tempo real; na seleção dos datasets primários, foram utilizados apenas *triggers* sem *prescale* (*unprescaled triggers*)⁸.

O *prescale* dos *triggers* foram verificados seguindo as instruções disponibilizadas pelo *CMS Open Data* [28]. A Tabela 2 lista os *triggers* utilizados em cada conjunto de dados primário.

Seguindo as instruções do *CMS Open Data* [29], e empregando os *triggers* mencionados em conjunto com a ferramenta **BRILCalc**⁹, foi determinada a seguinte luminosidade integrada¹⁰ $\mathcal{L}_{\text{int}} = 16,290 \text{ fb}^{-1}$ para os datasets primários considerados nesta análise.

4.4. Critérios de seleção de eventos

O objetivo desta análise é selecionar eventos compatíveis com a produção do bóson Z decaindo em neutrinos,

⁸ O fator de *prescale* determina a fração de eventos que o sistema de aquisição de dados (DAQ) grava entre aqueles que satisfazem os critérios de um determinado *trigger*. Por exemplo, se um *trigger* possui *prescale* igual a 10, apenas 1 em cada 10 eventos aceitos será armazenado. Quando o valor do *prescale* é igual a 1, dizemos que o *trigger* é **sem prescale**, ou seja, todos os eventos que atendem ao critério de seleção são registrados. Nessas condições, nenhum evento compatível com a condição de disparo é descartado.

⁹ O **BRILCalc** é um utilitário desenvolvido pelo CMS para calcular a luminosidade integrada e monitorar a qualidade dos dados coletados, utilizando informações de feixe e de *triggers*.

¹⁰ A luminosidade integrada é uma grandeza usada em física de colisores para medir o total de colisões coletadas durante um período de tempo, sendo a integral no tempo da luminosidade instantânea.

bem como caracterizar seus principais *backgrounds*. A assinatura típica desse processo consiste em jatos de alta energia equilibrados por um vetor de momento transversal ausente ($\vec{p}_T^{\text{miss}}$).

Para descrever essa configuração, introduzimos a variável **vetor de recuo** [3], definida como:

$$\vec{U} = \vec{p}_T^{\text{miss}} + \sum_{\ell} \vec{p}_T^{\ell}, \quad (8)$$

onde a soma percorre todos os léptons carregados no evento. Nos casos em que não há léptons no estado final, o vetor de recuo coincide com o momento transversal ausente, isto é, $\vec{U} = \vec{p}_T^{\text{miss}}$, de forma que $U = |\vec{p}_T^{\text{miss}}|$.

A seleção inicial de eventos é baseada em um conjunto de critérios de referência, aplicados à todas as regiões de análise [3]:

- $U > 200 \text{ GeV}$;
- O jato líder¹¹ deve ter $p_T > 200 \text{ GeV}$;
- Eventos contendo jatos com $p_T > 40 \text{ GeV}$ e pseudo-rapidez¹² $|\eta| > 2,4$ são rejeitados, a fim de reduzir contribuições de espalhamento de bósons vetoriais.

Para garantir a qualidade da reconstrução, exige-se consistência entre p_T^{miss} e U por meio da variável:

$$\Delta = \frac{|p_{T,\text{calo}}^{\text{miss}} - U|}{U}, \quad (9)$$

sendo aceitos apenas eventos com $\Delta < 0,5$.

Além disso, aplicam-se vetos adicionais para suprimir processos de *background*:

- Fótons isolados com $p_T > 25 \text{ GeV}$ e $|\eta| < 2,5$;
- Jatos identificados como provenientes de quarks b (*b-tagged*)¹³;
- A separação azimutal¹⁴ mínima entre os quatro jatos líderes e $\vec{p}_T^{\text{miss}}$, denotada por $\Delta\phi_{\text{min}}$, deve satisfazer $\Delta\phi_{\text{min}} > 0,5$, o que ajuda a reduzir o *background* de multijatos QCD.

Após os cortes de referência, os eventos são distribuídos em diferentes **regiões de controle**, cada uma otimizada para estudar assinaturas específicas do sinal ou controlar contribuições de *background* [3]:

- **Região de sinal p_T^{miss} + jatos:** caracteriza o canal invisível do Z , exigindo a ausência de léptons carregados (elétrons e múons) no estado final. São vetados:

- Múons ou elétrons isolados com $p_T > 10 \text{ GeV}$ e $|\eta| < 2,5$;
- Taus¹⁵ com $p_T > 20 \text{ GeV}$ e $|\eta| < 2,3$.

- **Região de dois léptons ($\ell^+\ell^-$):** usada como região de referência para $Z \rightarrow \ell^+\ell^-$. Seleciona eventos com:

- Dois múons ou dois elétrons isolados, com massa invariante¹⁶ é uma grandeza calculada a partir do produto escalar do quadrimomento total das partículas, $M^2 = (p_1 + p_2)^2$. No caso de duas partículas com momentos transversos p_{T1} e p_{T2} , diferença de pseudo-rapidez $\Delta\eta$ e diferença azimutal $\Delta\phi$, M^2 pode se escrever como $M^2 = 2p_{T1}p_{T2}(\cosh\Delta\eta - \cos\Delta\phi)$. Ela representa a massa total do sistema formado e é invariante sob transformações de Lorentz. Para a seleção de eventos, impõe-se o critério $71 < m_{\ell^+\ell^-} < 111 \text{ GeV}$;
- $p_T > 25 \text{ GeV}$ para múons e $p_T > 30 \text{ GeV}$ para elétrons.

- **Regiões μ + jatos e e + jatos:** usada para estimar o *background* $W \rightarrow \ell\nu$.

- Região μ + jatos: exige um múon isolado com $p_T > 25 \text{ GeV}$ e $|\eta| < 2,4$;
- Região e + jatos: exige um elétron isolado com $p_T > 30 \text{ GeV}$ e $p_T^{\text{miss}} > 100 \text{ GeV}$.

Em ambas as regiões, a massa transversa¹⁷ deve satisfazer $30 \leq M_T(\ell, p_T^{\text{miss}}) < 125 \text{ GeV}$.

- **Região τ + jatos:** usada para estimar o *background* $W \rightarrow \tau\nu$. Requer um lépton tau com

¹¹ Em física de partículas de altas energias, o *jato líder* (*leading jet*) é aquele com o maior momento transversal (p_T) entre todos os jatos reconstruídos em um evento.

¹² A pseudorapidez é uma coordenada comumente usada para descrever o ângulo θ de trajetória de uma partícula em relação ao eixo do feixe. Ela é definida como $\eta = -\ln(\tan \frac{\theta}{2})$.

¹³ Foi utilizado o discriminador CSVv2 no ponto de operação médio [30] para identificar esses jatos. O discriminador CSVv2 é um algoritmo de identificação de jatos originados de quarks b , baseado em informações de vértices secundários e parâmetros de impacto das trajetórias.

¹⁴ O ângulo azimutal ϕ corresponde ao ângulo no plano transversal ao feixe, medido em radianos, conforme a convenção usual em física de partículas de altas energias.

¹⁵ Léptons tau não são observados diretamente no detector, pois possuem vida média extremamente curta e decaem antes de atingir os sistemas de detecção. Uma fração significativa dos taus decai em hádrons (principalmente píons) acompanhados de um neutrino. Esses decaimentos são chamados taus hadrônicos. No detector, eles aparecem como jatos estreitos com baixa multiplicidade de trajetórias carregadas (tipicamente uma ou três trajetórias) e energia depositada de forma colimada. O experimento CMS utiliza algoritmos específicos que exploram essas características – forma do jato, número de trajetórias e isolamento em relação à outras partículas – para distinguir taus hadrônicos de jatos comuns produzidos por quarks e glúons.

¹⁶ A massa invariante de um sistema de duas partículas.

¹⁷ A massa transversa é usada para medir a massa de partículas que decaem em partículas invisíveis (por exemplo, neutrinos). O cálculo considera apenas as componentes do momento que são transversais (perpendiculares) ao eixo z do feixe de colisão (plano $x-y$). Para um decaimento com uma partícula visível ℓ , e um momento transversal ausente associado, define-se $M_T = \sqrt{2p_T^{\ell}p_T^{\text{miss}}(1 - \cos\Delta\phi_{\ell,\text{miss}})}$, onde p_T^{ℓ} e p_T^{miss} são os módulos dos momentos transversos e $\Delta\phi_{\ell,\text{miss}}$ é a separação azimutal entre eles. Ela fornece uma estimativa parcial da massa do sistema, independentemente do momento ao longo do feixe.

$p_T > 40 \text{ GeV}$, $|\eta| < 2,3$ e massa transversa $30 < m_T < 125 \text{ GeV}$.

- **Região QCD:** essa região é utilizada como controle para o background de multijatos e é definida pela inversão do seguinte corte aplicado nas demais regiões:

$$\min [\Delta\phi(j_1, j_2, j_3, j_4, p_T^{\text{miss}})] \leq 0,5.$$

Cada uma dessas regiões aplica, além dos cortes específicos, um conjunto de vetos e requisitos cinemáticos destinados a maximizar a pureza do sinal e minimizar as incertezas. A lista completa dos critérios de seleção, organizada por região, encontra-se na Tabela S6, apresentada no Apêndice A.

4.5. Normalização de eventos

A utilização de amostras de Monte Carlo exige um processo de normalização de eventos, de modo que as distribuições simuladas possam ser comparadas de forma direta com os dados reais. Essa normalização é feita por meio da atribuição de um peso a cada evento simulado, garantindo que o número efetivo de eventos na simulação corresponda ao esperado para a luminosidade integrada dos dados.

De acordo com as recomendações do *CMS Open Data* [31], o peso atribuído a cada evento é calculado pela expressão 10:

$$w = \frac{\sigma \times \mathcal{L}}{n_{\text{gen}}}, \quad (10)$$

onde σ é a seção de choque do processo, $\mathcal{L}$ é a luminosidade integrada dos dados, e n_{gen} representa o número efetivo de eventos gerados. Esse número é obtido a partir da diferença:

$$n_{\text{gen}} = \text{gw_pos} - \text{gw_neg},$$

em que “gw_pos” corresponde à soma dos pesos positivos ($\sum(\text{gen_weights} > 0)$) e “gw_neg” à soma dos pesos negativos ($\sum(\text{gen_weights} < 0)$).

Por outro lado para os dados experimentais, cada evento é considerado com peso unitário, ou seja, $w = 1$.

4.6. Correção de γ^*

O conjunto de dados *DYJetsToLL*¹⁸ é utilizado nesta análise para modelar o processo de produção do bóson Z , seguido de seu decaimento em léptons carregados. Essa amostra corresponde ao processo de Drell–Yan, no qual

¹⁸ O conjunto de dados *DYJetsToLL* descreve o processo de Drell–Yan, no qual um quark e um antiquark produzem um bóson Z que decai em um par de léptons carregados. Esse processo é fundamental na análise porque constitui o análogo visível do decaimento do Z em neutrinos. Como elétrons e múons são detectáveis, eventos de Drell–Yan funcionam como um canal de controle experimental, permitindo validar a reconstrução do detector e a modelagem da simulação.

Tabela 3: Seções de choque e razões de cada processo em relação ao total do DY.

Processo	Seção de Choque (pb)	Razão em relação ao DY
DY($\ell^+\ell^-$)	5316	1
Z($\ell^+\ell^-$)	5203	0.9787 (k_Z)
$\gamma^*(\ell^+\ell^-)$	79	0.0149 (k_γ)

um quark e um antiquark se aniquilam produzindo um bóson Z ou um fóton virtual (γ^*), que então decai em um par de léptons ($\ell^+\ell^-$).

Entretanto, essa amostra inclui também a contribuição do fóton virtual (γ^*), além do termo de interferência entre Z e γ^* . Para que apenas o processo $Z \rightarrow \ell^+\ell^-$ seja considerado na análise, é necessário aplicar uma correção.

A estratégia adotada consiste em utilizar as seções de choque dos processos $Z \rightarrow \ell^+\ell^-$ e $\gamma^* \rightarrow \ell^+\ell^-$, de modo a calcular fatores de correção (k_Z e k_γ) relativos à seção de choque total do processo DY [3]. Esses valores estão apresentados na Tabela 3.

Em física de partículas, a seção de choque é uma medida da probabilidade de ocorrência de um determinado processo em colisões entre partículas. Quanto maior a seção de choque, maior a frequência esperada desse processo. Seus valores são expressos em unidades de área (barn, sendo $1 \text{ barn} = 10^{-24} \text{ cm}^2$), e na prática utilizam-se submúltiplos como o picobarn (pb). Assim, os números apresentados na Tabela 3 indicam a taxa relativa com que cada processo contribui para a produção de pares de léptons no experimento.

Aplicando os fatores de correção, obtêm-se as seções de choque corrigidas para cada contribuição:

$$\sigma_{Z,\text{corr}} = \sigma_Z \cdot k_Z = 5092.2 \text{ pb}, \quad (11)$$

$$\sigma_{\gamma^*,\text{corr}} = \sigma_{\gamma^*} \cdot k_\gamma = 1.2 \text{ pb}, \quad (12)$$

$$\sigma_{\text{DY},\text{corr}} = \sigma_{Z,\text{corr}} + \sigma_{\gamma^*,\text{corr}} = 5093.4 \text{ pb}. \quad (13)$$

A diferença entre a seção de choque total simulada e a soma das contribuições do Z e do γ^* é atribuída ao termo de interferência¹⁹:

$$\sigma_{\text{int}} = \sigma_{\text{DY}} - (\sigma_Z + \sigma_{\gamma^*}) = 34 \text{ pb}. \quad (14)$$

Essa contribuição equivale a aproximadamente:

$$\frac{\sigma_{\text{int}}}{\sigma_{\text{DY}}} \times 100 = \frac{34}{5316} \times 100 \approx 0.64\%. \quad (15)$$

¹⁹ A produção de pares de léptons no processo de Drell–Yan pode ocorrer por dois mecanismos quânticos distintos: via bóson Z ou via fóton virtual (γ^*). Como esses dois mecanismos levam ao mesmo estado final, suas amplitudes quânticas interferem entre si. Essa interferência altera levemente a taxa total prevista, gerando um termo adicional denominado termo de interferência σ_{int} . Embora numericamente pequeno, esse termo é importante para que a soma das contribuições reproduza corretamente a seção de choque total simulada.

Assim, para excluir o efeito da interferência, calcula-se o peso a ser aplicado nos eventos do conjunto DYJetsToLL:

$$\sigma_{\text{DY,final}} = \sigma_{\text{DY,corr}} - \sigma_{\text{DY,corr}} \times 0.0064 = 5060.8 \text{ pb}, \quad (16)$$

$$w = \frac{\sigma_{\text{DY,final}}}{\sigma_{\text{DY,simulada}}} = \frac{5060.8}{5316} = 0.9519. \quad (17)$$

Aplicando o fator $w = 0.9519$ aos eventos do DYJetsToLL, garante-se que a análise leve em conta apenas o processo $Z \rightarrow \ell^+ \ell^-$, com as contribuições indesejadas do fóton virtual e da interferência devidamente corrigidas.

4.7. Correção de pileup

Em colisões próton-próton no LHC, é comum que mais de uma interação ocorra no mesmo cruzamento de feixes. Esse efeito, conhecido como *pileup*, afeta variáveis importantes da reconstrução de eventos, como o número de vértices primários e a energia depositada em diferentes regiões do detector.

Nas simulações de Monte Carlo (MC), as condições de *pileup* não coincidem exatamente com as observadas nos dados reais. Por isso, é necessário aplicar uma correção estatística, conhecida como *pileup reweighting*, que ajusta as distribuições obtidas por simulação para que reflitam de forma mais realista as condições de múltiplas colisões registradas no experimento CMS.

A distribuição de *pileup* nos dados é obtida seguindo as recomendações da colaboração CMS [32], utilizando a ferramenta `pileupCalc`, em conjunto com as informações de luminosidade registradas e com o valor recomendado para a seção de choque inelástica²⁰ dos dados de 2016. O resultado é um histograma no formato ROOT que fornece a distribuição normalizada do número de interações por evento nos dados reais. Essa distribuição é então comparada à correspondente na simulação. Para cada número de interações de *pileup* n , define-se um peso de correção:

$$w(n) = \frac{P_{\text{data}}(n)}{P_{\text{MC}}(n)}, \quad (18)$$

onde $P_{\text{data}}(n)$ representa a probabilidade de ocorrência de n interações nos dados, e $P_{\text{MC}}(n)$ a mesma probabilidade na simulação.

²⁰ Em colisões próton-próton no LHC, um único cruzamento de feixes pode produzir várias interações simultâneas. Esse efeito, conhecido como *pileup*, ocorre porque os prótons não colidem individualmente, mas em pacotes contendo muitas partículas. A seção de choque inelástica representa a probabilidade total de ocorrer qualquer interação que produza novas partículas nesses cruzamentos. Ela é utilizada como parâmetro global para estimar quantas interações adicionais são esperadas em cada evento. Como o número de interações simultâneas afeta a reconstrução do detector, a simulação precisa ser corrigida para reproduzir a mesma distribuição observada nos dados reais.

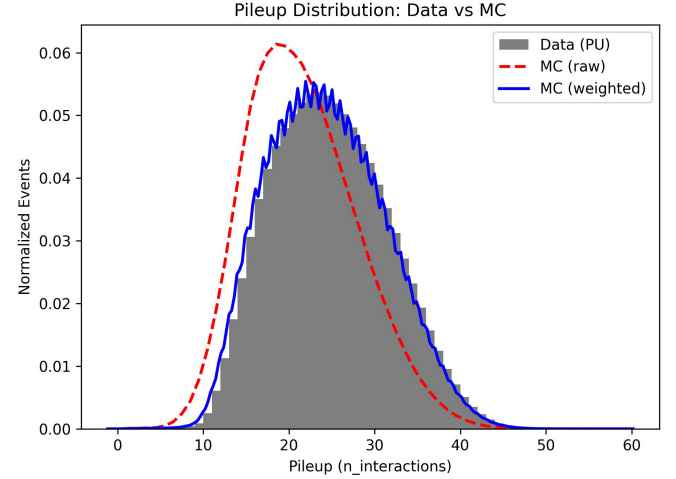


Figura 2: Distribuições normalizadas de *pileup* para os dados reais, para a simulação antes da correção e após o *reweighting*.

Esses pesos são aplicados evento a evento, garantindo que a distribuição final de *pileup* nos dados simulados seja compatível com a observada nos dados experimentais.

A Figura 2 ilustra o procedimento. A área sombreada em cinza corresponde à distribuição de *pileup* nos dados reais. A linha vermelha tracejada mostra a simulação antes da correção e a linha azul indica o resultado após o *reweighting*. Nota-se que, após a aplicação dos pesos, as distribuições passam a coincidir, assegurando uma descrição mais realista das condições experimentais.

4.8. Abordagem Bayesiana e uso de MCMC

A análise estatística realizada neste trabalho segue a abordagem Bayesiana. Em contraste com métodos frequentistas, nos quais os parâmetros são fixos e estimados por técnicas como a maximização da função de verossimilhança, a inferência Bayesiana trata os parâmetros como variáveis aleatórias e visa determinar sua distribuição a posteriori, dada uma função de verossimilhança e distribuições a priori (priors) associadas.

Formalmente, a distribuição a posteriori $p(\theta | \mathcal{D})$ dos parâmetros θ é dada pelo teorema de Bayes [33]:

$$p(\theta | \mathcal{D}) \propto \mathcal{L}(\mathcal{D} | \theta) \pi(\theta),$$

onde $\mathcal{L}(\mathcal{D} | \theta)$ é a função de verossimilhança dos dados $\mathcal{D}$ e $\pi(\theta)$ representa as distribuições a priori.

Para acessar essa distribuição posterior de forma numérica, utilizamos o método de amostragem por Cadeias de Markov Monte Carlo (MCMC), que permite gerar amostras distribuídas conforme a posteriori, a partir das quais extraímos estatísticas descritivas, como mediana, intervalos de credibilidade e correlações entre os parâmetros. Neste trabalho, a amostragem foi realizada com a biblioteca `emcee` [34], baseada no algoritmo *Affine Invariant Ensemble Sampler*, amplamente utilizado em análises Bayesianas com múltiplos parâmetros.

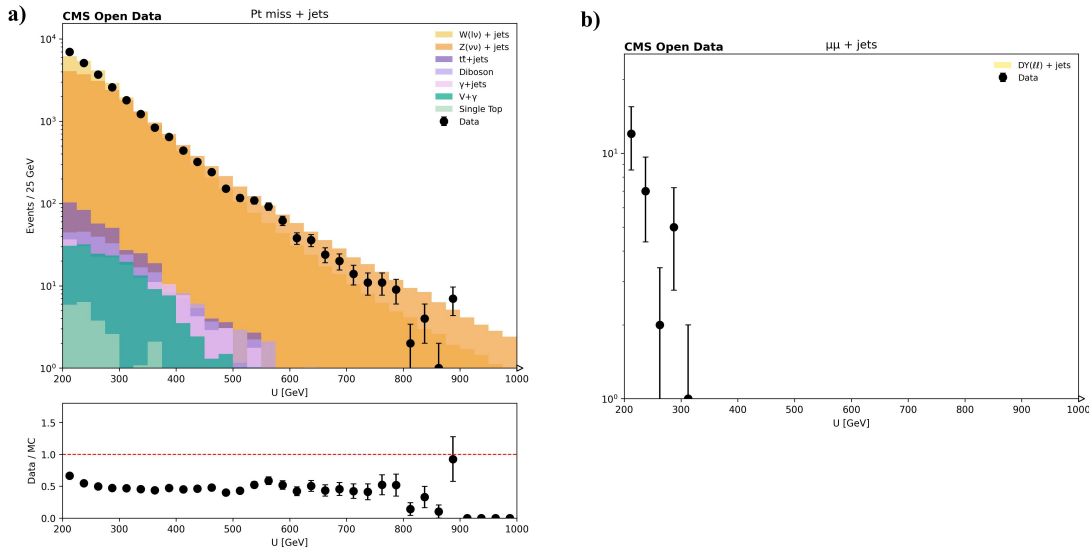


Figura 3: Vetor de recuo U . (a) Distribuição dos eventos compatíveis com o processo invisível $Z \rightarrow \nu\bar{\nu}$. (b) Distribuição correspondente ao canal de referência $Z \rightarrow \ell^+\ell^-$.

5. Resultados e Discussão

A Figura 3 apresenta a distribuição do vetor de recuo U , definido na Eq. 8, comparando os dados reais de colisão com as previsões obtidas por simulação de Monte Carlo.

Na Figura 3a, vemos a distribuição dos eventos compatíveis com o processo invisível $Z \rightarrow \nu\bar{\nu}$. Esses eventos aparecem tipicamente como $p_T^{\text{miss}} + \text{jatos}$, ou seja, jatos visíveis acompanhados de momento transversal ausente devido à presença dos neutrinos. Nessa categoria, estão incluídos tanto os eventos da região de 'sinal' $p_T^{\text{miss}} + \text{jatos}$ quanto os das regiões de controle 'único múon' e 'único elétron', conforme definidos na Tabela S6.

A Figura 3b, por sua vez, apresenta a distribuição correspondente ao canal de referência $Z \rightarrow \ell^+\ell^-$, no qual o bóson Z decai em dois léptons carregados. Nesse caso, utilizamos os eventos das regiões de controle 'dois múons' e 'dois elétrons', também especificadas na Tabela S6. Como os critérios de seleção para a região 'dois elétrons' não resultaram em nenhum evento, a amostra exibida corresponde exclusivamente à região de controle 'dois múons' ($\mu^+\mu^- + \text{jatos}$).

O painel inferior na Figura 3a exibe a razão entre os dados reais e as previsões obtidas por Monte Carlo, avaliada *bin a bin* – isto é, em cada intervalo discreto (*bin*) do histograma. Essa comparação permite verificar, em cada intervalo de U , o quão bem as simulações reproduzem o comportamento observado nos dados experimentais.

Na região de 'sinal' $p_T^{\text{miss}} + \text{jatos}$, nota-se que a razão entre dados e Monte Carlo permanece, em grande parte do espectro, abaixo de 1. Isso significa que a simulação prevê mais eventos do que os observados nos dados reais.

Já na região 'dois múons' $\mu^+\mu^- + \text{jatos}$, devido à estatística limitada do conjunto *CMS Open Data* utilizado – restrito a parte dos dados de 2016 – observam-se flutuações significativas no espectro. Essas flutuações

não indicam discrepâncias físicas relevantes, mas refletem apenas o número reduzido de eventos disponíveis.

A distribuição do vetor de recuo U , mostrada na Figura 3, é o principal observável utilizado nesta análise. A partir dela, é realizado um ajuste simultâneo para determinar a largura invisível do bóson Z . Esse ajuste baseia-se na definição de uma função de verossimilhança, cuja distribuição posterior, combinada com *priors* informativos, é explorada via amostragem MCMC. O ajuste combina de forma conjunta as informações obtidas nas regiões experimentais: região de 'sinal', região 'único múon', região 'único elétron', região 'dois múons' e região 'dois elétrons' (Tabela S6), levando em conta as contribuições específicas de cada processo físico em cada uma dessas regiões.

A função de verossimilhança [3] utilizada está expressa por:

$$\begin{aligned}
 \mathcal{L}(n_j, n_\ell, n_{\ell+\ell-} \mid r, r_Z, r_W, \theta) &= \text{Poisson}\left(n_j \mid r \cdot r_Z \cdot s_{Z,j}(\theta) \right. \\
 &\quad \left. + r_W \cdot b_{j,W}(\theta) + b_{\text{bkg},j}(\theta)\right) \\
 &\cdot \text{Poisson}\left(n_\ell \mid r_W \cdot b_{\ell,W}(\theta) + b_{\text{bkg},\ell}(\theta)\right) \\
 &\cdot \text{Poisson}\left(n_{\ell+\ell-} \mid r_Z \cdot s_{Z,\ell+\ell-}(\theta) \right. \\
 &\quad \left. + \sqrt{r_Z} \cdot s_{\text{int},\ell+\ell-} + s_{\gamma^*,\ell+\ell-}(\theta) \right. \\
 &\quad \left. + b_{\text{bkg},\ell+\ell-}(\theta)\right) \\
 &\cdot p(\tilde{\theta}, \theta),
 \end{aligned} \tag{19}$$

onde os termos correspondem a:

- $s_{Z,j}(\theta)$: taxa esperada do processo $Z \rightarrow \nu\bar{\nu}$ na região de sinal.

- $b_{j,W}(\theta)$: taxa esperada do processo $W + jatos$ na região de ‘sinal’.
- $b_{l,W}(\theta)$: taxa esperada do processo $W + jatos$ nas regiões ‘único múon’ e ‘único elétron’.
- $s_{Z,\ell+\ell-}(\theta)$: taxa esperada do processo $Z \rightarrow \ell^+\ell^-$ nas regiões ‘dois múons’ e ‘dois elétrons’.
- $s_{\text{int.},\ell+\ell-}$: termo de interferência entre Z e γ^* nas regiões ‘dois múons’ e ‘dois elétrons’.
- $s_{\gamma^*,\ell+\ell-}(\theta)$: contribuição do processo γ^* nas regiões ‘dois múons’ e ‘dois elétrons’.
- $b_{\text{bkg.},j}(\theta)$, $b_{\text{bkg.},\ell}(\theta)$, $b_{\text{bkg.},\ell\ell}(\theta)$: taxas esperadas de outros *backgrounds* nas regiões de ‘sinal’, ‘único múon’, ‘único elétron’, ‘dois múons’ e ‘dois elétrons’, respectivamente.
- $r_Z = \frac{N_{\text{MC}}^{\text{sinal}}(Z+jatos)}{N_{\text{MC}}^{\text{controle}}(Z+jatos)}$: fator de transferência para o processo $Z + jatos$.
- $r_W = \frac{N_{\text{MC}}^{\text{sinal}}(W+jatos)}{N_{\text{MC}}^{\text{controle}}(W+jatos)}$: fator de transferência para o processo $W + jatos$.
- $r = \frac{\Gamma(Z \rightarrow \nu\bar{\nu})}{\Gamma_{\text{SM}}(Z \rightarrow \nu\bar{\nu})}$: fator de escala da largura invisível do Z em relação à previsão do Modelo Padrão.

A partir da comparação entre os eventos simulados nas regiões de sinal e de controle, foram determinados os fatores de transferência: $r_Z = (1,09 \pm 2,55) \times 10^5$ para o processo $Z + jatos$, e $r_W = 1,4050 \pm 0,0180$ para o processo $W + jatos$.

O valor encontrado para r_Z é significativamente elevado e apresenta uma incerteza relativa superior a 200%. Essa discrepância está associada ao baixo número de eventos da amostra *DYJetsToLL* que satisfazem os critérios de seleção nas regiões de controle ‘dois múons’ e ‘dois elétrons’, o que compromete a precisão da estimativa direta desse fator. Detalhes quantitativos sobre os eventos que sobrevivem a cada etapa de seleção, em todas as regiões definidas na análise, podem ser consultados em [35].

Para contornar esse problema, optou-se por tratar r_Z como um parâmetro livre²¹ no ajuste simultâneo, permitindo que seu valor seja determinado pelos próprios dados. Essa abordagem evita a propagação de uma estimativa instável e estatisticamente pouco confiável.

O ajuste simultâneo entre as diferentes regiões da análise foi implementado utilizando a técnica de amostragem por Cadeias de Markov Monte Carlo (MCMC) [34]. Esse método permite explorar numericamente a distribuição a posteriori dos parâmetros, construída a partir da função de verossimilhança baseada nos dados experimentais e nos priors atribuídos a cada parâmetro.

No ajuste, foram considerados como variáveis livres o fator de escala r , os fatores de transferência r_Z e o parâmetro global de incerteza sistemática θ . Já o fator r_W foi mantido fixo, assumindo o valor previamente estimado.

Para levar em conta as incertezas experimentais e as restrições teóricas, foram aplicados *priors* gaussianos a alguns parâmetros, de modo a penalizar valores afastados das expectativas iniciais e garantir maior estabilidade ao ajuste.

- **Parâmetro global θ** : modela a incerteza sistemática global comum às regiões da análise. Foi aplicado um *prior* gaussiano com $\sigma_\theta = 0,032$, valor consistente com a incerteza sistemática combinada de 3,2%, detalhada na Tabela S5 no apêndice A.
- **Fator de escala r** : relacionado diretamente à largura invisível do bóson Z . Um *prior* informativo centrado em $r = 1,0$ com $\sigma_r = 0,1$ foi utilizado, refletindo a expectativa do Modelo Padrão.
- **Fator de transferência r_Z** : aplicado ao processo $Z + jatos$, com *prior* centrado em $r_Z = 0,7$ e $\sigma_{r_Z} = 0,2$.

A distribuição posterior dos parâmetros r , r_Z e θ foi amostrada com o algoritmo *Affine Invariant Ensemble Sampler*, implementado pela biblioteca *emcee* [34]. Este método é uma generalização eficiente do MCMC clássico, que utiliza múltiplos *walkers* (caminhantes) para explorar o espaço de parâmetros de forma paralela e mais robusta frente a correlações entre variáveis.

A amostragem foi conduzida com os seguintes parâmetros:

- Número de parâmetros: 3 (θ , r , r_Z)
- Número de *walkers*: 40
- Número de passos por *walker*: 5000
- *Burn-in* descartado: 1000 passos
- Total de amostras utilizadas nas estimativas: $40 \times 4000 = 160\,000$

A Figura 4 apresenta o diagrama de correlação entre os parâmetros ajustados via MCMC. As distribuições marginais mostram a densidade a posteriori de cada parâmetro individualmente, enquanto os painéis 2D revelam as correlações entre pares de parâmetros.

As distribuições marginais dos parâmetros estimados estão apresentadas na Figura 4, onde se observa que as distribuições a posteriori são unimodais, suaves e simetricamente bem comportadas, o que indica um bom grau de convergência da amostragem. Correlações esperadas entre r e r_Z também são visíveis e fisicamente interpretáveis.

Os parâmetros ajustados e suas respectivas incertezas, estimados a partir da distribuição *a posteriori* obtida via MCMC, são:

$$\begin{aligned} r &= 1,272^{+0,076}_{-0,073}, \\ r_Z &= 1,469^{+0,093}_{-0,088}, \\ \theta &= -0,6723^{+0,0052}_{-0,0051}. \end{aligned} \quad (20)$$

Os resultados mostram que o modelo ajustou os parâmetros para compensar o valor elevado de r_Z ,

²¹ Um parâmetro livre é ajustado diretamente a partir dos dados experimentais, sem ser fixado previamente nem derivado de outras estimativas.

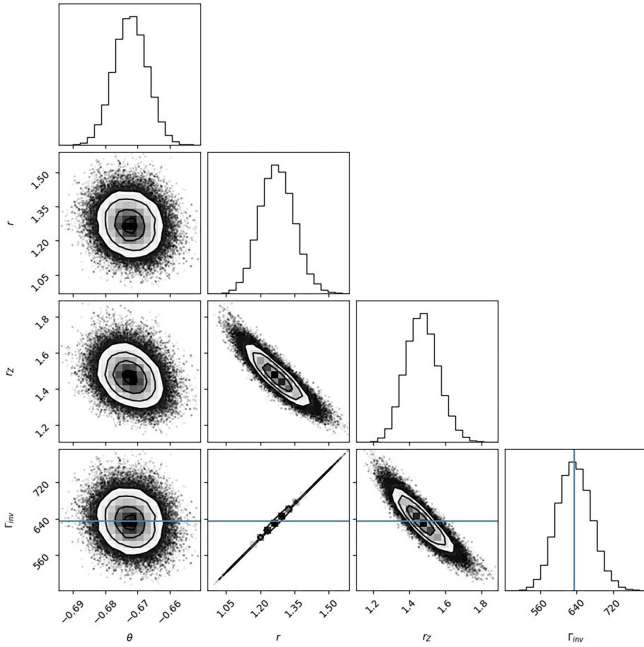


Figura 4: Distribuições marginais e correlações entre os parâmetros ajustados: θ (sistemático global), r (fator de escala da largura invisível do Z), r_Z (fator de transferência para Z+jatos) e Γ_{inv} (largura invisível).

decorrente da baixa estatística na região de controle com dois léptons.

Em comparação com a análise reproduzida [3], o ajuste simultâneo aumentou os valores de r e θ , como mostrado na Eq. 20, de modo a preservar a concordância com os dados experimentais. Esse comportamento evidencia a correlação entre os parâmetros: ao restringir o crescimento de r_Z , o modelo redistribui a compensação nos ajustes de r e θ , aprimorando a reprodução das distribuições observadas.

A partir do valor ajustado de r , a largura invisível do bóson Z pode ser calculada com base na predição teórica do Modelo Padrão para esse canal de decaimento, dada por $\Gamma_{\text{SM}}(Z \rightarrow \nu\bar{\nu}) = 499,0 \text{ MeV}$ [36]. O cálculo segue a relação:

$$\Gamma(Z \rightarrow \nu\bar{\nu}) = r \cdot \Gamma_{\text{SM}}(Z \rightarrow \nu\bar{\nu}). \quad (21)$$

Com a propagação das incertezas estatísticas do parâmetro r , o resultado final da largura invisível é:

$$\Gamma(Z \rightarrow \nu\bar{\nu}) = 634,6^{+37,9}_{-36,5} \text{ MeV}. \quad (22)$$

Além da incerteza estatística proveniente do ajuste de r , é necessário considerar duas fontes adicionais de incerteza que impactam a estimativa final da largura invisível:

- A incerteza no fator de transferência $r_W = 1,4050 \pm 0,0180$;
- A incerteza sistemática global, representada por θ , com desvio padrão $\sigma_\theta = 0,032$.

A contribuição absoluta dessas incertezas sobre $\Gamma(Z \rightarrow \nu\bar{\nu})$ foi propagada da seguinte forma:

- Erro devido a r_W : $\Delta\Gamma_{r_W} = \Gamma_{\text{SM}} \cdot \langle r \rangle \cdot \sigma_{r_W} = 499,0 \cdot 1,272 \cdot 0,018 \approx 11,4 \text{ MeV}$;
- Erro sistemático global: $\Delta\Gamma_{\text{sist}} = \langle \Gamma \rangle \cdot \sigma_\theta = 634,6 \cdot 0,032 \approx 20,3 \text{ MeV}$.

As incertezas total superior e inferior foram então calculadas pela combinação quadrática das componentes estatísticas e sistemáticas:

$$\Delta\Gamma_{\text{total}}^+ = \sqrt{37,9^2 + 11,4^2 + 20,3^2} \approx 44,5 \text{ MeV}, \quad (23)$$

$$\Delta\Gamma_{\text{total}}^- = \sqrt{36,5^2 + 11,4^2 + 20,3^2} \approx 43,3 \text{ MeV}. \quad (24)$$

Com isso, o valor final da largura invisível do bóson Z, incorporando todas as fontes de incerteza consideradas na análise, é:

$$\Gamma(Z \rightarrow \nu\bar{\nu}) = 634,6^{+44,5}_{-43,3} \text{ MeV}. \quad (25)$$

De forma aproximada, esse resultado pode ser expresso como:

$$\Gamma(Z \rightarrow \nu\bar{\nu}) \approx 635 \pm 44 \text{ MeV}, \quad (26)$$

A Figura 5 apresenta uma comparação direta entre o resultado obtido nesta análise e outras medições experimentais da largura invisível do bóson Z. A linha vertical tracejada indica o valor teórico previsto pelo Modelo Padrão.

Enquanto o resultado combinado dos experimentos do LEP (ALEPH, L3 e OPAL) mediu valores compatíveis com o Modelo Padrão, dentro de incertezas relativamente pequenas [17], a análise da Colaboração CMS [3] apresenta um desvio modesto para cima.

O resultado obtido nesta análise é consideravelmente mais alto, com um valor central cerca de 27% acima

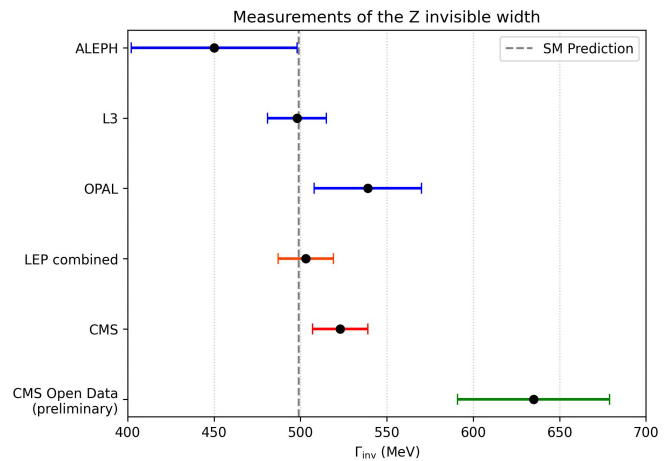


Figura 5: Comparação entre diferentes medições experimentais da largura invisível do bóson Z. A linha tracejada representa a predição do Modelo Padrão. O ponto inferior (em verde) corresponde ao resultado obtido nesta análise com dados preliminares do CMS Open Data.

da predição do SM. Essa diferença pode ser explicada, em grande parte, pela limitação estatística das amostras disponíveis no portal *CMS Open Data*, que abrange apenas as eras G e H do ano de 2016, além das incertezas associadas ao fator de transferência r_Z , amplificadas pela baixa estatística da simulação Monte Carlo na região de controle.

O valor da largura de decaimento obtido nesta análise é significativamente maior do que a estimativa mais precisa da colaboração CMS, $\Gamma = 523 \pm 16 \text{ MeV}$ [3]. Para quantificar essa diferença, calculamos a separação estatística entre os dois resultados utilizando a seguinte equação:

$$\Delta = \frac{|\Gamma_{\text{CMS Open Data}} - \Gamma_{\text{CMS}}|}{\sqrt{\sigma_{\text{Open Data}}^2 + \sigma_{\text{CMS}}^2}} \approx 2,38 \sigma. \quad (27)$$

A diferença entre os dois resultados corresponde, portanto, a aproximadamente $2,4\sigma$, embora significativo, o desvio pode ser atribuído às incertezas decorrentes da baixa estatística disponível nesta análise.

6. Considerações Finais

Neste trabalho, desenvolvemos uma proposta didática para o estudo da largura invisível do bóson Z , utilizando dados públicos disponibilizados pelo experimento CMS referentes às eras G e H do ano de 2016, com energia de centro de massa de $\sqrt{s} = 13 \text{ TeV}$. A análise concentrou-se em eventos nos quais o Z decai em neutrinos, inferidos indiretamente por meio do momento transversal ausente (p_T^{miss}) em associação com jatos.

O método adotado baseou-se na comparação com um canal de referência, o decaimento do Z em léptons carregados, complementada pela definição de regiões de controle e pela aplicação de um ajuste simultâneo com a técnica de amostragem por cadeias de Markov Monte Carlo (MCMC).

O valor obtido para a largura invisível foi de forma aproximada, $635 \pm 44 \text{ MeV}$. Esse resultado apresenta um desvio de aproximadamente $2,4\sigma$ em relação à medida publicada pela colaboração CMS [3].

Esse desvio é explicado principalmente pelas limitações estatísticas dos dados disponíveis no portal de dados abertos e pelas incertezas no fator de transferência r_Z , fortemente impactado pela baixa estatística de Monte Carlo na região de controle de dois léptons.

Apesar dessas limitações, a análise evidenciou que dados públicos podem sustentar estudos quantitativos confiáveis, permitindo reproduzir etapas características de uma pesquisa em física experimental de altas energias: da seleção e correção dos eventos até o uso de técnicas estatísticas avançadas.

Do ponto de vista pedagógico, este trabalho ilustra como os dados abertos do CERN podem ser usados em disciplinas de física de partículas e em graduação e pós-graduação, permitindo que os estudantes tenham

contato direto com a prática de análise de dados provenientes de um grande experimento internacional.

Ao trabalhar com um projeto como este, os estudantes precisam aplicar conceitos de estatística, lidar com incertezas experimentais, implementar funções de verossimilhança e empregar métodos de inferência como MCMC, elementos que raramente são abordados de forma integrada em currículos tradicionais. A exposição às ferramentas amplamente utilizadas em colaborações internacionais, como bibliotecas *ROOT*, *Python* (*NumPy*, *Matplotlib*, *emcee*), ou mesmo *Jupyter Notebooks*, prepara os alunos tecnicamente e metodologicamente para contribuir com pesquisas de ponta. Além disso, o desenvolvimento dessas competências tem impacto também fora da academia, dada a crescente demanda por formação qualificada e experiência com análise de grandes volumes de dados em setores de tecnologia, finanças e indústria.

O fato do código estar disponível em repositório aberto [4] reforça o compromisso com a ciência reprodutível e amplia as possibilidades de que outros estudantes e professores reproduzam, adaptem e desenvolvam novas extensões a partir deste estudo.

Em síntese, os resultados obtidos, embora limitados, confirmam o potencial do uso de dados abertos como recurso didático, ao mesmo tempo em que oferecem uma ponte concreta entre teoria e prática na formação de novos pesquisadores.

Agradecimentos

Agradecemos à FAPESQ pelo apoio financeiro por meio da bolsa de mestrado e ao Programa Paraíba Sem Fronteiras, iniciativa da FAPESQ em parceria com a SECTIES, pela concessão da bolsa de mestrado sanduíche. Agradecemos também ao grupo *CMS Open Data* pelo apoio e pelos valiosos ensinamentos transmitidos durante o período de atividades no CERN. P. R. Teles agradece à FAPERJ pelo suporte do projeto Física de Íons Pesados e Operações no Calorímetro Eletromagnético do CMS (processo E-26/200.598/2022).

Material Suplementar

O seguinte material suplementar está disponível online: Tabela S1 – Lista dos conjuntos de dados primários utilizados nesta análise.

Tabela S2 – Amostras de eventos simulados para os processos de sinal.

Tabela S3 – Amostras de eventos simulados para os processos de background dominante.

Tabela S4 – Amostras de eventos simulados para os processos de background secundários.

Tabela S5 – Fontes de incerteza sistemática consideradas na análise.

Tabela S6 – Seleção “referência” e critérios para cada região.

Disponibilidade de Dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo está disponível publicamente no portal *CMS Open Data* e pode ser acessado em [22].

Referências

- [1] CERN Open Data Portal, *CERN Open Data Policy for LHC Experiments*, disponível em: <https://opendata.cern.ch/docs/cern-open-data-policy-for-lhc-experiments>.
- [2] T.A.R. Monteiro e P.R. Teles, *Revista Brasileira de Ensino de Física* **48**, e20250358 (2026).
- [3] J.M. Campbell, S. H  che, H.T. Li, C.T. Preus e P. Skands, *Phys. Lett. B* **842**, 137563 (2023).
- [4] M.G.F.S.A. Gomes, *Reposit  rio no GitHub: materiais da reprodu  o da medida da largura invis  vel do b  son Z utilizando dados abertos do experimento CMS*, dispon  vel em: <https://github.com/Mariaggomes/ZinvWidth-Analysis>.
- [5] S.L. Glashow, *Nucl. Phys.* **10**, 107 (1959).
- [6] A. Salam, *Weak and electromagnetic interactions* (World Scientific Publishing, Singapura, 1968).
- [7] S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967).
- [8] P.A.M. Dirac, *Proc. Roy. Soc. Lond. A* **114**, 243 (1927).
- [9] F. Mandl e G. Shaw, *Quantum Field Theory* (John Wiley & Sons, New Jersey, 2010).
- [10] C. Patrignani, C. Schwanda, S. Spanier, G. Venanzoni, C. Z. Yuan, K. Agashe, G. Aielli, B. C. Allanach, J. Alvarez-Mu  niz, M. Antonelli et al., *Phys. Rev. D* **110**, 030001 (2024).
- [11] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Er  , C. Fabjan et al., *Phys. Lett. B* **716**, 30 (2012).
- [12] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S.A. Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins et al., *Phys. Lett. B* **716**, 1 (2012).
- [13] C. A. Bertulani e H. Schecter, *Introdu  o   F  sica Nuclear* (Editora UFRJ, Rio de Janeiro, 2007).
- [14] Particle Data Group, Z Boson, *Review of Particle Physics* (White Paper, PDG, LBL, Calif  rnia, 2023).
- [15] M.C.G. Garc  a e M. Yokoyama, *Neutrino Masses, Mixing, and Oscillations*, dispon  vel em: <https://pdg.lbl.gov/2024/reviews/rpp2024-rev-neutrino-mixing.pdf>.
- [16] D. Cogollo, H. Diniz, C.A.S. Pires e P.S. Rodrigues da Silva, *Modern Physics Letters A* **23**, 3405 (2008).
- [17] The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration, The OPAL Collaboration, The SLD Collaboration, The LEP Electroweak Working Group, The SLD Electroweak and Heavy Flavour Groups, *Physics Reports* **427**, 257 (2006).
- [18] L. Evans e P. Bryant, *JINST* **3**, S08001 (2008).
- [19] CMS Collaboration, *JINST* **3**, S08004 (2008).
- [20] K. Hagiwara, K. Hikasa, K. Nakamura, M. Tanabashi, M. Aguilar-Benitez, C. Amsler, R. M. Barnett, P. R. Burchat, C. D. Carone et al., *Phys. Rev. D* **66**, 010001 (2002).
- [21] CERN Courier, *CMS starts underground*, dispon  vel em: <https://cerncourier.com/a/cms-starts-underground/>.
- [22] CERN Open Data Portal, *CMS experiment: datasets*, dispon  vel em: <https://opendata.cern.ch/search?q=&f=experiment%3ACMS&l=list&order=desc&p=1&s=10&sort=mostrecent>.
- [23] J. Pivarski, *uproot: A ROOT I/O in pure Python and NumPy*, dispon  vel em: <https://pypi.org/project/uproot/>.
- [24] J. Pivarski, *awkward: Manipulate nested, variable-sized data in Python*, dispon  vel em: <https://pypi.org/project/awkward/>.
- [25] CMS Collaboration, *Certified luminosity JSON file for 2016 proton-proton collisions at $\sqrt{s} = 13$ TeV*, dispon  vel em: <https://cms-opendata-workshop.github.io/workshop2024-lesson-triggers-lumi/instructor/05-challenge.html>.
- [26] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, M. Zaro, *JHEP* **07**, 079 (2014).
- [27] T. Sj  strand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015).
- [28] CMS Open Data Workshop 2021, *Introduction to CMS Trigger: Getting Trigger Prescales*, dispon  vel em: <https://cms-opendata-workshop.github.io/workshop2021-lesson-introtrigger/03-prescales/index.html>.
- [29] CMS Open Data Workshop 2024, *Triggers & Luminosity – Brilcalc Exercises*, dispon  vel em: <https://cms-opendata-workshop.github.io/workshop2024-lesson-triggers-lumi/instructor/05-challenge.html#brilcalc-exercises>.
- [30] CMS, *B-Tagging Recommendations for 80X*, dispon  vel em: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X>.
- [31] CMS Open Data Workshop 2024, *Modeling Backgrounds*, dispon  vel em: <https://cms-opendata-workshop.github.io/workshop2024-lesson-background-modeling/instructor/index.html>.
- [32] CMS, *CMS Utilities for Accessing Pileup Information for Data*, dispon  vel em: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData>.
- [33] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari e D.B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, Boca Raton, 2013).
- [34] D. Foreman-Mackey, D.W. Hogg, D. Lang e J. Goodman, *emcee: The MCMC Hammer* **125**, 306 (2013).
- [35] M.G.F.S.A. Gomes, *Medida da Largura Invis  vel do B  son Z Utilizando os Dados Abertos do Experimento CMS*. Disserta  o de Mestrado, Universidade Federal de Campina Grande, Campina Grande (2025).
- [36] Particle Data Group, *Review of Particle Physics*, dispon  vel em: <https://pdg.lbl.gov/2023/listings/rpp2023-list-z-boson.pdf>.