# Study of central and exclusive production of tau-tau pairs at LHC in 2018 era, $\mu \tau_h$ decay mode

Gonçalo Martins<sup>1</sup>, Gonçalo Barreto<sup>1</sup>, Pedro Esperanço<sup>1</sup>, and Gonçalo Almeida<sup>1</sup>

<sup>1</sup>Instituto Superior Técnico, Lisboa, Portugal

Project supervisor: Matteo Pisano Project developed within the course: LFUI, IST

January 30, 2023

Abstract. O objetivo desta atividade consiste em estudar a produção central e exclusiva de pares  $\tau^+\tau^-$ , que resultam da interação entre dois protões acelerados pelo LHC, e determinar o limite da secção eficaz do processo que envolve o decaimento do par de taus num muão e num conjunto de hadrões ( $\tau^+\tau^- \rightarrow \tau_\mu \tau_h$ ). Primeiramente, efetuou-se uma sessão experimental de controlo dos dados provenientes do CMS durante a Run que teve início no ano de 2022, e procurou-se garantir que os parâmetros analisados estavam coerentes com o esperado e dentro dos limites necessários, averiguando também possíveis perdas de informação. Foi possível concluir que durante grande parte do tempo de recolha não houve nenhum tipo de adversidades, à exceção de alguns instantes associados a maiores perdas de informação. De seguida, foi realizada uma análise de uma amostra de dados semelhantes aos dados recolhidos pelo CMS durante o ano de 2018 (Run 2). Durante o processo da análise, utilizaram-se ficheiros simulados por métodos Monte-Carlo, que incluíam quer os eventos de fundo quer o sinal separadamente (com eventos de pile-up à mistura) e, a partir de um conjunto de triggers, selecionaram-se os acontecimentos parecidos ao sinal (processo de interesse). De seguida, construíram-se gráficos das distribuições de algumas variáveis analisadas de forma a confrontar com o modelo teórico e garantir que o processo de seleção estava coerente com os gráficos observados. Juntando os eventos de fundo que resistiram ao processo de seleção num único ficheiro e utilizando o ficheiro com o sinal separado, treinaram-se dois algoritmos de análise multivariada (BDT e MLP) de forma a maximizar a distinção entre sinal e fundo, tendo sido necessário à priori encontrar um conjunto de variáveis relevantes para fazer essa distinção. Por análise da importância conferida a cada uma das variáveis discriminantes, optou-se por utilizar o algoritmo BDT e assim, utilizando o software Higgs Combine, foi possível obter um valor para o limite da secção eficaz do processo em estudo.

KEYWORDS: LHC, proton-proton collision, tau-tau exclusive, exclusive production, TMVA

# 1 Motivação

O interesse desta atividade prende-se com o objetivo de obter o menor limite possível para a secção eficaz do processo em estudo, para assim se excluírem todos os valores obtidos acima deste limite por outros experimentalistas da mesma área. De facto, de acordo com as previsões teóricas, a secção eficaz é da ordem de 1 fb [4], no entanto até à data de hoje ainda ninguém atingiu este valor. Para além disto, o interesse científico desta atividade está relacionado com a possibilidade de relacionar a secção eficaz calculada do processo em estudo com o coeficiente g-2 do  $\tau$ , que nos permitiria concluir se o  $\tau$  é ou não uma partícula elementar, podendo estar de acordo ou não com o modelo padrão vigente (para partículas elementares, g-2 = 0). No entanto, a determinação desse coeficiente não faz parte do objetivo final deste projeto.

# 2 O detetor CMS e PPS

O CMS (Compact Muon Solenoid) é um dos grandes detetores de partículas situado dentro do LHC (Large Hydron Collider), no CERN, perto da cidade de Genebra, construído em volta do ponto de interação 5 (IP5). Constitui o principal foco desta atividade, pois é nele onde ocorrem as interações entre os protões analisadas após serem acelerados pelo LHC até alcançarem energias bem definidas.

Como o nome indica, no CMS existe uma grande ênfase em detetar muões. A razão para isto deve-se ao facto dos muões serem excelentes assinaturas de física interessante, dado que estão presentes no produto de inúmeros processos físicos, e podem também ser facilmente detetados após decaírem de colisões entre partículas pesadas.

Este detetor foi construído com o intuito de medir as quantidades cinemáticas (energia e momento) das partículas geradas pela interação entre os protões acelerados e está dividido em várias camadas que vão desde o ponto de interação às câmaras muónicas (zona mais exterior). Sucintamente, o CMS é constituído, junto ao ponto de interação, pelo tracker revestido por camadas concêntricas de píxeis de silício que permitem, juntamente com o calorímetro eletromagnético, medir o momento e a energia das partículas eletricamente carregadas (eletrões e positrões), assim como dos fotões. A seguir encontrase o calorímetro hadrónico que desempenha um papel essencial na identificação dos quarks e gluões através da medição da energia e da direção dos jets, e também dos neutrinos a partir de omissões de energia nos eventos. Esta parte encontra-se rodeada por um solenóide supercondutor responsável por gerar um campo magnético axial suficientemente forte (4T) capaz de ajudar na medição dos momentos das partículas eletricamente carregadas. Por fim, na parte mais exterior encontram-se as câmaras muónicas otimizadas na deteção e medição dos muões.

Infelizmente, o CMS não é capaz de detetar os protões desviados após a colisão, devido ao facto do ângulo entre a direção inicial e final dos protões ser muito reduzido. As-





Figure 1. Vista transversal do CMS mostrando os diferentes subdetetores e a forma como as partículas interagem



Figure 2. Representação de um dos braços do PPS

sim, surgiu a necessidade de construir um detetor especializado na sua deteção e na medição do ângulo de desvio em relação ao eixo principal de colisão, o PPS (Precision Proton Spectometer). Este detetor é formado por dois braços simétricos com cerca de 200m de comprimento a partir do ponto de interação, e cada braço é formado por 3 estações, designadas Roman Pots: duas de tracking e uma estação de timing, que permitem, respetivamente, rastrear a trajetória final dos protões visando medir a fração de momento perdida ( $\xi = \frac{p_i - p_i}{p_i}$ ), e determinar o instante temporal da sua formação no ponto de interação. A estação de timing não foi considerada na análise por não estar disponível à data da realização desta atividade.

# 3 Parte Experimental

#### 3.1 Introdução

Na parte experimental, efetuou-se uma sessão de controlo dos dados guardados pelo CMS que saem diretamente do CERN e que estão disponíveis ao público, verificando se estão coerentes com o esperado e garantindo ainda o bom o funcionamento dos vários detetores que constituem o CMS.

Os dados que foram supervisionados pertencem a um dos Fills da Run 3, que teve início no ano de 2022 e que se caracteriza por uma energia de colisão de 13.6 TeV (6.8 TeV por cada protão). Para garantir uma análise cuidada retiram-se os diferentes parâmetros em intervalos de 3 minutos.

Cada Run é dividido em vários Fills, cuja duração pode ir de 1 hora a 35 horas, consoante existam problemas técnicos durante o processo que obriguem à paragem da recolha de dados. Cada Fill é composto por 3 fases bem definidas. Na primeira fase, os protões são enviados em "pacotes" para dentro do acelerador e de seguida são acelerados até atingirem a energia de colisão necessária (segunda fase). Por fim, os protões interagem em pontos de interação prédefinidos (IP5 no caso do CMS) [3].

#### 3.2 Variáveis supervisionadas

De entre todas as variáveis que foram controladas destacase a Luminosidade integrada ( $L_{int}$ ) fornecida pelo acelerador, ou somente Luminosidade (L). Esta quantidade relaciona-se com o número de protões que interagem durante a recolha de dados e pode ser definida por  $L_{int} = N/\sigma$  ( $b^{-1}$ ), onde 'N' representa o número de eventos observados e ' $\sigma$ ' a secção eficaz intrínseca do processo. Outro conceito importante é a luminosidade instantânea que remete para a taxa de eventos detetados num período de tempo. Mais em detalhe, a luminosidade instantânea fornecida pelo acelerador é calculada por:

$$\mathcal{L} = \frac{kn_1n_2f}{4\pi\sigma_x\sigma_y} \tag{1}$$

onde f representa a frequência de interação entre pacotes de protões, k o número de pacotes em cada feixe,  $n_{1,2}$  o número total de partículas em cada feixe e  $\sigma_{x,y}$  a "largura" em cada direção (mede o nível de dispersão do feixe nas duas direções). Para verificar experimentalmente a fórmula anterior existem detetores especializados em monitorizar a luminosidade fornecida ao instante (BRAN -Beam RAte of Neutrals - luminosity) e é necessário garantir que a diferença entre as duas não é suficientemente expressiva.

Uma vez que o número de interações que ocorrem dentro do LHC é extremamente elevado (é alcançada uma frequência de interação de 40 MHz) é necessário filtrar esses acontecimentos para que possam ser efetivamente guardados. O processo de seleção apresenta duas fases: o Level 1 Trigger e o High Level Trigger que reduzem o número de eventos a uma frequência efetiva de 100 e 2-3 kHz, respetivamente. O Level 1 trigger funciona como uma fase menos rígida da seleção por ter acesso a um conjunto limitado de variáveis e é responsável por enviar os acontecimentos para uma Read-Out Unit (RU) que valida a informação e envia-a para a Building Unit (BU), onde o evento é reconstruído. Só mais tarde é que essa informação alcança o High Level Trigger que estabelecerá se vale ou não a pena guardar o acontecimento. Assim é necessário controlar as taxas de eventos que atravessam cada um dos triggers e verificar que não excedem os valores aqui mencionados garantindo, tendo em conta o tamanho médio de cada evento, que os fluxos de informação não excedem os 200 Gb/s no Lv1 Trigger e nas RU e que a informação guardada (output) seja inferior a 15 GB/s. Estes valores utilizados para limitar a informação guardada não são aleatórios, e portanto estão na verdade relacionados com a capacidade de armazenamento de dados do CERN, que tem de ser também capaz de gerir os dados provenientes dos outros grandes detetores.

Para além destes parâmetros é necessário supervisionar o estado do feixe e de todos os detetores do CMS e verificar que o tempo morto do detetor não excede 1% [5].





Figure 3. Variação do tempo morto total, do feixe e do detetor, com o tempo.

Outros dois parâmetros necessários na supervisão dizem respeito ao ângulo de colisão entre os feixes de protões (crossing angle) e a uma distância relacionada com a redução da dimensão do feixe junto ao ponto de interação para a maximizar a luminosidade (denotado por  $\beta^*$ ), garantindo assim que estes dois parâmetros se mantém constantes. O tempo morto do detetor é dado pela subtração entre o tempo morto total e o tempo morto do feixe e é necessário garantir que se encontra abaixo de 1%, segundo a referência [5].

#### 3.3 Análise dos parâmetros retirados

Quanto ao tempo morto do detetor (imagem 3), este encontra-se sempre abaixo de 1%, mais ou menos ao mesmo nível, apesar das oscilações constantes entre o tempo morto total e o tempo morto do feixe, o que significa então que o número de eventos perdidos por incapacidade do detetor é muito reduzido. No entanto, é possível reparar em dois instantes (720 e 4680 s) em que o tempo morto do feixe apresenta um valor superior aos restantes. Estes instantes são bastante importantes pois permitem explicar variações em quantidades a seguir apresentadas.

Como é possível verificar pela imagem 4, as taxas de acontecimentos que atravessam o Lv1 Trigger, assim como o HL Trigger, estão dentro dos valores indicados acima (100 e 2-3 kHz, respetivamente) sendo também de notar a presença de três outliers nos valores do Lv1 Trigger que aparecem nos instantes de tempo 420, 4680 e 5750 s. Os primeiros dois outliers podem ser explicados devido ao aumento repentino do valor do tempo morto do feixe que ocorre ao mesmo tempo, sendo que no primeiro caso a frequência de informação apresenta um valor de 0 kHz e no segundo apenas se nota uma descida elevada do valor em relação ao resto. O último outlier não aparenta estar relacionado com o tempo morto do feixe.

Multiplicando pelo peso médio de um evento em cada instante, foi possível concluir que o fluxo de informação que atravessa as RU / Lv1 Trigger não deve exceder os 200 Gb/s (imagem 5). Quanto ao output, mostra-se também na mesma imagem que a informação guardada é inferior a 15 Gb/s (ronda os 3 GB/s).

De forma a garantir que não houve informação perdida até chegar às Read-Out Units (RU), fez-se um estudo da quantidade  $\frac{\phi_{L1}-\phi_{RU}}{\phi_{L1}}$  e elaborou-se o gráfico da sua



**Figure 4.** Variação da frequência de informação que passava pelo Level 1 e High Level Triggers, com o tempo.



Figure 5. Fluxos de informação das RU e do output



**Figure 6.** Desvio da quantidade de informação que passa pelo Level 1 Trigger e que chega às Read-Out Units.

evolução temporal (imagem 6). Como podemos constatar, a diferença é residual e pode estar relacionada com a parte eletrónica de todo o conjunto responsável pelo processamento da informação. Encontram-se, no entanto três pontos nos instantes 720, 4680 e 5750 s, em que esta quantidade apresenta valores inesperados. O primeiro não está definido devido ao valor de 0 kHz atribuído à frequência de informação que passa o Lv1 Trigger (mencionado anteriormente) e os outros dois podem ser explicados pelas variações do Lv1 Trigger (também explicados anteriormente).

Pode-se também reparar que, em relação à luminosidade desde o início do Fill, a percentagem de eventos guardados não exibe nenhuma dependência com o tempo e consequentemente a probabilidade dos processos físicos que estão a ocorrer também não (imagem 7).

Como base nos valores tabulados é de notar que a luminosidade instantânea ( $\mathcal{L}$ ) apresenta uma tendência decres-





Figure 7. Percentagem de eventos guardados (eficiência)

cente, o que pode ser facilmente justificado pelo número de protões em cada feixe que diminuiu com o tempo pela razão de interagirem uns com os outros durante o Fill.

#### 3.4 Impacto da sessão de controlo

Como seria de esperar, por não sermos cientistas responsáveis pelo correto funcionamento do CMS, o nosso papel na recolha de dados foi verificar se todos os parâmetros estavam em linha com as expectativas. Qualquer anomalia que pudéssemos vir a encontrar e que não tivesse sido guardada teria de ser registada no LogBook oficial e seria necessário entrar em contacto com o pessoal responsável através do nosso coordenador. Além de mais, teríamos nas nossas mãos o papel de interromper ou parar por tempo indeterminado o funcionamento do detetor sempre que tal fosse exigido, destruir dados, etc. No entanto, procurou-se estimar o impacto que esta sessão de controlo teve no cálculo do limite da secção eficaz mais à frente (secção 5.4) tendo em conta, por exemplo, o efeito do tempo morto no cálculo desse limite e de que forma a luminosidade guardada durante o nosso período de recolha poderia impactar esse valor obtido e o intervalo em que se insere.

## 4 Introdução Teórica

# 4.1 Processo em estudo

O processo aqui em estudo provém da colisão entre dois protões acelerados pelo LHC que, após interagirem, são desviados em direções diferentes e perdem energia permitindo a criação de um par  $\tau^+\tau^-$  [1, 4], não havendo portanto dissociação dos protões. A interação surge esquematizada na imagem 8.



Figure 8. Diagrama representativo da colisão entre dois protões.

$$pp \to p + \tau^+ \tau^- + p$$
 (2)

Este par  $(\tau^+\tau^-)$  formado durante o processo é um par instável, pelo que os taus podem decair de formas diferentes:

- Decaímento eletrónico:  $\tau^- \rightarrow e^- + \bar{\nu}_e + \nu_\tau = \tau_e$ ;
- Decaímento muónico:  $\tau^- \rightarrow \mu^- + \bar{\nu}_{\mu} + \nu_{\tau} = \tau_{\mu}$ ;
- Decaímento hadrónico: τ<sup>-</sup> → τ<sub>h</sub>, sendo τ<sub>h</sub> um conjunto de hadrões.

O par formado, que constitui o sistema central, pode decair numa combinação dos  $\tau$ 's acima mencionados. Em princípio, seria importante estudar todos os canais de decaimento possíveis, no entanto, esta atividade foca-se apenas na análise do canal  $\tau^+\tau^- \rightarrow \tau_\mu \tau_h$ , no qual se verifica a presença de um muão e de um conjunto de hadrões.

Desta forma, procura-se calcular a seçção eficaz ( $\sigma$ ) do processo acima a partir da análise de uma amostra de dados semelhantes aos dados recolhidos pelo CMS durante o ano de 2018. A secção eficaz ( $\sigma$ ) está relacionada com a probabilidade deste processo ocorrer e é expressa em unidades de área, mais convencionalmente em *barns* ( $b = 10^{-28}m^2$ ). É independente da intensidade e do foco dos feixes de partículas e, por isso, uma medição feita num acelerador pode ser comparada com medições feitas noutros aceleradores, independentemente da potência dos mesmos.

Para determinar o valor da secção eficaz é necessário conhecer o número total de acontecimentos que apresentam o estado final pretendido e, para isso, é necessário atravessar todos os acontecimentos gerados dentro do detetor e fazê-los passar por um conjunto de triggers que permitam identificar os acontecimentos de interesse. No entanto, fenómenos como o pile-up e os eventos de fundo dificultam todo este processo. No CMS, no ano de 2018, aconteceram mais ou menos 30 interações ao mesmo tempo, ou seja, interações que o CMS não foi capaz de distinguir temporalmente e, assim, o estado final é formado por várias partículas provenientes de interações diferentes. Este fenómeno é conhecido como pile-up. Os eventos de fundo são os acontecimentos parecidos ao processo de interesse (sinal), mas que resultam de uma interação diferente. Os eventos de fundo que podemos encontrar são:

- Fundo Drell-Yan:  $pp \rightarrow l^+l^-$ , onde l representa um leptão eletricamente carregado. Se l for um  $\tau$ , o processo  $pp \rightarrow \tau^+\tau^-$  é muito parecido ao sinal (sendo a única diferença a ausência de protões no estado final). Esta interação pode resultar na formação de um par  $l^+l^-$  ou na formação de uma bosão Z que depois decai nesse par;
- Fundo TTJets: pp → tt, sendo t um quark top. Por sua vez, o top formado decai num quark b e num bosão W (t → W<sup>+</sup> + b), e este último pode decair num conjunto leptão + neutrino ou num par de quarks. No primeiro caso o estado final é muito parecido ao sinal;
- Fundo QCD: caracterizado pela produção sucessiva de quarks e gluões que podem ser confundidos com um τ<sub>h</sub>. Tem uma secção eficaz muito elevada, o que significa que é muito frequente.

É importante mencionar também o que são os jets. Os jets são grupos de partículas de longa duração produzidas quando um quark ou um gluão hadroniza, e cujas propriedades cinemáticas recaem sobre os protões que as originam. Estes jets são responsáveis pelo decaimento de centenas de partículas tornando a análise mais complicada, pelo que em vez de serem analisadas individualmente são analisadas em grupo. A reconstrução destes jets é uma parte importante da análise no LHC.

# 4.2 Variáveis de interesse

Para eliminar os eventos de fundo é imperativo definir um conjunto de critérios de seleção. Para este efeito, definiram-se as seguintes variáveis a considerar:

- 1. Variáveis medidas diretamente:
  - Ângulo entre a direção final e inicial (eixo Z) dos protões;
  - $\xi = \frac{p_i p_f}{p_i}$  que representa a fração do momento perdido de cada protão;
  - *p<sub>T</sub>* o momento transversal da partícula que corresponde à projeção do momento no plano XY (plano perpendicular à direção inicial dos protões);
  - $\phi$  o ângulo entre  $p_T$  e o eixo X;
  - η a pseudo-rapidez definida como
    η = -log[tan <sup>θ</sup>/<sub>2</sub>], sendo θ o ângulo entre o momento e o eixo Z;
  - A energia (E) das partículas;
  - A carga das partículas.
- 2. Variáveis calculadas a partir das anteriores:
  - $M_X$  que representa a massa invariante do sistema central e pode ser calculada pelo CMS ( $M_X = \sqrt{2 \cdot p_1 \cdot p_2 \cdot (1 - \cos \alpha)}$ ) ou pelo PPS ( $M_X = \sqrt{s \cdot \xi_1 \cdot \xi_2}$ ), onde  $\sqrt{s}$  remete para a energia de colisão;
  - $Y_X = \frac{1}{2} \log \frac{\xi_1}{\xi_2}$  que representa a rapidez do sistema central;

A figura 9 apresenta um esquema detalhado da zona de interação. O referencial proposto tem origem no ponto de interação (IP5), no qual o eixo X aponta para o centro do LHC, o eixo Y aponta na direção vertical para cima e o eixo Z é dado pela regra da mão direita, sendo definido como a direção inicial dos protões.



Figure 9. Esquema do referencial com origem no ponto de interação (IP5)

# 5 Análise de uma amostra de dados de 2018

### 5.1 Análise de ficheiros simulados por métodos Monte-Carlo

A análise partiu do estudo de uma amostra de dados simulados por métodos de Monte-Carlo, do qual fazem parte o sinal, os fundos Drell Yan, TTJets e QCD com o pile-up incluído, assim como todos os outros processos que resultam dos outros decaimentos possíveis dos taus. Cada acontecimento tem informação disponível referente às variáveis anteriormente mencionadas na secção 4.2, pelo que apenas é necessário filtrar os valores que estas variáveis podem assumir e posicioná-las dentro dos intervalos definidos, sendo que com este processo é possível selecionar apenas os acontecimentos parecidos com o sinal.

Passando agora, definitivamente, para processo de análise que envolve a elaboração do código escrito, começou-se por selecionar apenas os eventos que continham um muão a partir de um dos triggers disponíveis, o "HLT-IsoMu24-v\*" [6], que garante a presença de um muão isolado caracterizado por um momento  $p_T > 24$  GeV, e a seguir selecionaram-se os eventos que apresentavam simultaneamente um muão e um  $\tau_h$ . Esta parte constituiu a primeira fase.

Na segunda fase selecionaram-se os acontecimentos cujo muão e tau apresentavam um grau de confiança suficientemente elevados utilizando algoritmos de identificação específicos que fazem a distinção entre os mesmos e outras partículas e garantiu-se ainda que a pseudo-rapidez do muão, assim como a do tau, fossem em módulo < 2,4, caso contrário não interagiriam com o detetor. Impôs-se também um  $p_T$  do muão > 35 GeV e um momento do  $\tau_h > 100$  GeV. A escolha destes intervalos para os momentos está relacionada com o facto do PPS apenas poder detetar protões caracterizados por um valor de  $\xi$  que satisfaça a condição  $\sqrt{s \cdot \xi_1 \cdot \xi_2} = M_x > 300$  GeV, pelo que o momento transversal de cada  $\tau$  pode ser aproximado a  $p(\tau) > M_x/2 = 150$  GeV. Para achar o momento do muão é preciso ter em conta que o tau muónico, para além de ser composto por um muão, é também composto por um neutrino e um anti-neutrino, logo é plausível a aproximação  $p(\mu) \approx p(\tau_{\mu})/3 = 50 \text{ GeV}$  e, uma vez que a energia não é totalmente capturada pelos detetores, convém diminuir estes limites para que não escapem acontecimentos de interesse. Como o par de taus tem cargas simétricas, impôsse ainda um produto de cargas negativo.

Para supervisionar o número de acontecimentos que passava cada umas das condições anteriores, introduziramse variáveis de contagem para assim controlar todo o processo de exclusão de acontecimentos.

### 5.2 Variáveis Discriminantes

Após impostas todas estas condições foi possível selecionar os eventos de fundo semelhantes ao processo de interesse e procedeu-se a uma análise individual de algumas variáveis referentes ao sistema central (par de taus), a partir da observação dos histogramas com o número de eventos em função de cada uma dessas variáveis e procurou-se





**Figure 11.** Histograma da acoplanaridade (definida no texto) do sistema  $\mu + \tau_h$  para cada um dos fundos

verificar se a teoria assim como as condições impostas estavam coerentes com as distribuições presentes nesses gráficos. Esta parte é essencial no sentido em que permitirá selecionar que variáveis fornecer ao algoritmo de análise multivariada que fará a distinção entre o fundo e o sinal.



**Figure 10.** Histograma da massa invariante do sistema  $\mu + \tau_h$  para cada um dos fundos

Na figura 10, é possível observar as distribuições da massa invariante das partículas produzidas pelos vários fundos. Os fundos TTJets e QCD apresentam uma massa invariante do sistema semelhante, sendo o valor mais provável de ambos os fundos entre 150 a 200 GeV. Já no caso do fundo Drell-Yan observam-se dois picos na distribuição, sendo o pico da esquerda claramente mais populado do que o pico da direita. Os bosões  $Z^0$  característicos deste fundo, cuja distribuição da massa invariante tem valor esperado próximo de 90 GeV, são os responsáveis pelo pico da esquerda, que devido à impossibilidade do detetor de apanhar toda a energia apresenta um valor mais próximo de 70 GeV.

Na figura 11, observam-se as distribuições da acoplanaridade ( $a = \frac{|\phi_{\tau\mu} - \phi_{\tau_h}|}{\pi}$ ) dos sistemas produzidos pelos vários fundos. Verifica-se que, em todos os casos, o pico de acoplanaridade encontra-se muito próximo do valor de 1, o que vai de encontro ao valor esperado, já que a maior parte das partículas se forma com um ângulo de 180°, dada a conservação do momento, e esta grandeza é obtida dividindo por  $\pi$  o ângulo entre as duas partículas formadas após um decaimento, sendo o seu valor mínimo 0 (0°) e máximo 1 (180°). No caso do fundo Drell-Yan, verifica-



**Figure 12.** Histograma do momento transversal do sistema  $\mu + \tau_h$  para cada um dos fundos

se mais um pico à esquerda que resulta da produção do bosão  $Z^0$  que por sua vez decai em leptões. Sabendo que, aquando do seu decaimento,  $Z^0$  apresenta uma velocidade diferente de 0 m/s, a conservação de momento obriga a que os leptões se afastem com um ângulo inferior a 180°, o que leva a uma acoplanaridade mais próxima de 0.

Na figura 12, é possível observar as distribuições do momento transversal do sistema  $\tau_{\mu} + \tau_{h}$  produzidas pelos vários fundos. Verifica-se que o valor esperado desta grandeza para o fundo TTJets se encontra entre 65 a 80 GeV, enquanto que o pico da distribuição do fundo QCD se encontra entre os valores de 60 a 100 GeV. Já no caso do Drell-Yan é possível observar, novamente, dois picos, sendo o pico da direita resultante dos bosões Z<sup>0</sup> produzidos por este fundo, cuja distribuição do momento transversal tem pico entre os valores de 135 a 175 GeV e o pico da esquerda entre os valores de 40 e 80 GeV que resulta da formação dos leptões.

Em todos os gráficos aqui analisados, a cinemática entre os fundos QCD e TTJets é semelhante.

Depois de uma análise cuidada, verificou-se que as variáveis de interesse seriam a massa invariante do sistema  $(M_x)$  e o momento transversal do mesmo  $(p_T)$ , os momentos transversais do tau muónico  $(\tau_{\mu})$  e do tau hadrónico  $(\tau_h)$ , a acoplanaridade do sistema e a diferença entre os valores obtidos experimentalmente e os teóricos, tanto para a massa invariante do sistema como para a rapidez (variáveis de "matching").

### 5.3 Análise Multivariada

Juntando os eventos de fundo que resistiram ao processo de filtragem num único ficheiro e utilizando o ficheiro com o sinal separado - também este criado a partir de métodos Monte-Carlo uma vez que a secção eficaz teórica do processo é conhecida - treinaram-se dois algoritmos de análise multivariada (BDT e MLP) de forma a maximizar a distinção entre sinal e fundo, disponibilizando o conjunto de variáveis que se consideraram relevantes para input (secção 5.2). Cada algoritmo utiliza como base conhecimentos de Machine Learning para classificar com melhor precisão cada acontecimento como sendo sinal ou fundo, e ambos os algoritmos utilizados estão presentes na biblioteca TMVA ("Technics of Multivariable Analysis") do





Figure 13. Distribuições das variáveis discriminantes com a respetiva distinção entre sinal e fundo

root [6]. É possível também encontrar na imagem 13 as distribuições das diferentes variáveis discriminantes utilizadas com a respetiva distinção entre sinal e fundo, só que neste caso os diferentes fundos aparecem assimilados.

Para manusear esta ferramenta dividiu-se o conjunto de acontecimentos simulados a partir do processo de Monte-Carlo em duas partes, utilizando uma das partes para treinar os algoritmos de classificação e a outra parte para os testar (à volta de metade). Após executar cada um dos algoritmos, obtiveram-se os resultados presentes nos gráficos 14 e 15 para os algoritmos BDT e MLP, respetivamente, que mostram a discriminação entre o fundo e o sinal para diferentes níveis de resposta, e mostra ainda um teste de "overtraining" (conhecido por Kolmogorov-Smirnov), no qual se avalia se a distribuição do sinal e do fundo obtidas na fase de treino são coerentes com as distribuições obtidas na fase de teste.

Na tentativa de selecionar qual dos algoritmos usar na amostra semelhante aos dados de 2018, analisou-se o gráfico presente na imagem 16 que mostra a aceitação do sinal (eixo horizontal) em relação à rejeição do fundo (eixo vertical) para diferentes níveis de resposta de ambos os algoritmos. O algoritmo que maximiza a área debaixo da curva gerada (vulgarmente conhecida por curva ROC) é identificado como sendo o melhor, no entanto este gráfico não evidencia uma resposta clara. Assim, procurouse analisar a importância conferida por cada um dos algoritmos às variáveis discriminantes individualmente (imagem 17), chegando-se à conclusão que o algoritmo MLP



Figure 14. Teste de "overtraining" do algoritmo BDT



Figure 15. Teste de "overtraining" do algoritmo MLP



Figure 16. Curva teste da probabilidade de rejeição de fundo em função da eficiência do sinal

atribuía uma importância desmesurada apenas a uma das variáveis, ignorando por exemplo duas variáveis que apresentavam uma distinção clara entre fundo e sinal (as variáveis de "matching"). Contrariamente, o algoritmo BDT repartia a importância conferida da mesma forma por todas as variáveis. Deste modo, optou-se então pelo algoritmo de análise BDT com o intuito de tornar o resultado mais fidedigno.

#### 5.4 Valor limite da secção eficaz

Chegou agora o momento de aplicar esse algoritmo à amostra semelhante de dados à amostra de dados semelhantes aos de 2018 e determinar o limite da secção eficaz do processo em estudo, utilizando a ferramenta Higgs Combine. Fornecendo a luminosidade total guardada durante esse ano assim como as variáveis discriminantes para input, o programa disponibilizou o valor de uma constante conhecida como força do sinal (R) responsável por fazer a distinção entre a distribuição do fundo e a distribuição do sinal obtidas após se aplicar o algoritmo



BDT	
Variável	Importância
$Y_X - 0.5 \ln(\xi_1/\xi_2)$	1,677E-01
Acoplanaridade do sistema central	1,677E-01
$M_X - \sqrt{s\xi_1\xi_2}$	1,627E-01
Momento trasversal do muão	1,384E-01
Massa invariante do sistema central	1,365E-01
Momento trasversal do tau	1,231E-01
Momento trasversal do sistema central	1,050E-01
MLP	
MLP Variável	Importância
MLP Variável Acoplanaridade do sistema central	Importância 9,333E+01
MLP Variável Acoplanaridade do sistema central Momento trasversal do tau	Importância 9,333E+01 1,016E+01
MLP Variável Acoplanaridade do sistema central Momento trasversal do tau Massa invariante do sistema central	Importância 9,333E+01 1,016E+01 8,813E+00
MLP Variável Acoplanaridade do sistema central Momento trasversal do tau Massa invariante do sistema central Momento trasversal do muão	Importância 9,333E+01 1,016E+01 8,813E+00 6,248E+00
MLP Variável Acoplanaridade do sistema central Momento trasversal do tau Massa invariante do sistema central Momento trasversal do muão Momento trasversal do sistema central	Importância 9,333E+01 1,016E+01 8,813E+00 6,248E+00 6,628E+00
MLP Variável Acoplanaridade do sistema central Momento trasversal do tau Massa invariante do sistema central Momento trasversal do muão Momento trasversal do sistema central $M_X - \sqrt{s\xi_1\xi_2}$	Importância 9,333E+01 1,016E+01 8,813E+00 6,248E+00 6,628E+00 1,499E+00

**Figure 17.** Importância conferida a cada um dos variáveis pelos dois algoritmos. BDT na imagem de cima e MLP na imagem de baixo



Figure 18. Descrição gráfica do papel da força de sinal R

já treinado (como se encontra descrito na secção 5.3). É de esperar que a distribuição associada ao fundo seja mais significativa que a distribuição do sinal já que o número de acontecimentos correspondente ao fundo é muito superior ao número de acontecimentos classificados como sinal, e que ambas se encontrem sobrepostas, pelo que a constante R terá o papel de fazer sobressair a distribuição do sinal até que a altura do seu pico seja igual a três vezes o seu erro poissoniano associado ao bin do fundo que está abaixo do pico de sinal.

Esta força do sinal relaciona-se com o o valor limite da secção eficaz ( $\sigma_{lim}$ ) e com o valor teórico da secção eficaz ( $\sigma_{teo} = 1$  fb) da seguinte forma:

$$\sigma_{lim} = R \cdot \sigma_{teo}.$$
 (3)

Na imagem 18, encontra-se uma discrição clara do papel da constante R enunciada. Verificou-se que o valor para a luminosidade guardada durante o ano de 2018 foi de 54,  $9 \pm 0$ , 1  $fb^{-1}$ . A partir das condições necessárias já descritas, obteve-se um valor para R compreendido entre 20 fb e 150 fb. O que este valor diz é que é apenas possível observar o processo experimentalmente para uma secção eficaz superior a R, e portanto para qualquer valor inferior a este apenas se pode concluir que não é possível detetar experimentalmente o processo.

Para tentar estimar a contribuição que o tempo morto total poderia ter no cálculo do limite da secção eficaz, procurou-se calcular a luminosidade total (incluindo os eventos perdidos) da Run de 2018 tendo em conta o tempo morto médio total durante o período em que se efetuou a recolha dos dados (2022) e, para isso, efetuou-se um ajuste constante (y = b) do tempo mortal total de forma a achar o valor médio ( $\bar{x}$ ), obtendo para o parâmetro o valor de  $\bar{x} = (4, 13 \pm 0, 02)$ %. Assim, assumindo que o tempo morto em percentagem não varia de uma Run para a outra, pode-se calcular a luminosidade total de 2018 através da fórmula (eq: 4):

$$L_{\text{tot}_{2018}} = \frac{L_{\text{g}_{2018}}}{1 - \bar{x}} \tag{4}$$

onde  $L_{g_{2018}}$  remete para a luminosidade efetivamente guardada pelo CMS, cujo valor encontrou-se ser de 54, 9± 0, 1  $fb^{-1}$ . Obteve-se então, para a luminosidade guardada o valor de 57, 27±0, 1  $fb^{-1}$ . Introduzindo novamente este valor no software Higgs Combine, obteve-se, como seria de esperar, um melhoramento de 4% para o resultado do limite, associado ao facto de agora o número de eventos analisados ser superior que contribui para que o erro seja menor.

# 6 Agradecimentos

Queremos agradecer ao professor e investigador Matteo Pisano por toda a paciência, dedicação e auxílio prestado ao longo desta atividade. Também agradecemos ao LIP pela oferta desta formação e disponibilidade do espaço para a realização da atividade.

# References

- M. Pisano, Study of central exclusive production of top quark-antiquark pairs at LHC, DOI: 10.1393/ncc/i2021-21066-9 (2021)
- [2] B. Abi et al, Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm, Phys. Rev. Lett. 126, 141801 – Published 7 April 2021
- [3] CMS, TOTEM Collaboration, M. Albrow et al., "CMS-TOTEM Precision Proton Spectrometer", in CERN-LHCC-2014-021, TOTEM-TDR-003, CMS-TDR-13. 9, 2014
- [4] S. Atag and A. A. Billur, "Possibility of Determining  $\gamma\gamma \rightarrow \tau^+\tau^-$  Lepton Electromagnetic Moments in process at the CERN-LHC", JHEP 11 (2010) 060 doi:10.1007/JHEP11(2010)060, arXiv:1005.284
- [5] https://iopscience.iop.org/article/10.1088/1748-0221/ 16/12/P12014



- [6] https://root.cern/manual/tmva/
- [7] https://cms-nanoaod-integration.web.cern.ch/ integration/cms-swCMSSW\_10\_6\_19/mc102X\_

doc.html#Electron