



CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE

STAGE DE CÉSURE

**Recherche de Violation de la Saveur Leptonique lors de la
désintégration du lepton τ à l'expérience Belle II**

Écrit par Louis Paletta
Mines ParisTech - Option Génie Atomique

Supervisé par Justine Serrano
Centre de Physique des Particules de Marseille - IN2P3/CNRS

12 septembre 2020

Introduction

Dans le cadre du Modèle Standard de la physique des particules, le nombre leptonique est une charge généralement conservée, sans que cela soit garantie par le Théorème de Noether et l'existence d'une symétrie fondamentale associée. La découverte de l'oscillation des neutrinos [1] est la première preuve de violation de la saveur leptonique et oriente la recherche vers de nouvelles extensions du Modèle Standard. La violation de la conservation du nombre leptonique est prédite par de nombreuses théories de Nouvelle Physique, en particulier l'existence de désintégration telle que $\tau \rightarrow \mu\mu\mu$ dont le rapport de branchement serait de l'ordre de grandeur de 10^{-8} à 10^{-10} . Expérimentalement, les meilleures limites à ce jour de ce rapport de branchement ont été estimées à 2.1×10^{-8} et 3.3×10^{-8} respectivement à partir des données des expériences Belle et BaBar, à un degré de confiance de 90%. L'amélioration de l'estimation de la limite du rapport de branchement par l'obtention de plus de données par le nouveau collisionneur Belle II permettrait de mieux contraindre les différentes théories de Nouvelle Physique.

Dans l'attente de disposer de suffisamment de données expérimentales, cette étude porte sur l'utilisation de données simulées par une méthode de Monte Carlo pour séparer le signal $\tau \rightarrow \mu\mu\mu$ du bruit de fond correspondant aux désintégrations du modèle standard. L'étude se restreint à une topologie 3×3 : le premier τ se désintègre en trois muons tandis que le deuxième se désintègre en trois particules chargées quelconques. Les données de cette topologie n'ont pas encore été utilisées et leur ajout à celle d'une topologie 3×1 permettrait de disposer d'un plus grand ensemble de données pour de meilleurs résultats. Dans une première partie, les concepts théoriques pertinents sont introduits, puis l'expérience et la simulation des données sont abordées. Dans une deuxième partie, une analyse multivariée des données simulées est réalisée afin de déterminer les variables et les coupures pertinentes pour discriminer le signal du bruit de fond. Un algorithme de machine learning est comparé à une optimisation classique dans l'optique d'obtenir une meilleure séparation du signal et du bruit de fond, avant d'estimer la limite de ratio de branchement atteignable par l'expérience pour une certaine quantité de données.

1 Violation de la Saveur Leptonique

1.1 Rappels Théoriques

La détection d'une nouvelle violation de la conservation de la charge leptonique dans le secteur chargé serait un signal clair de l'existence d'une Nouvelle Physique. Parmi ce type de désintégration de nombreuses études portent sur la désintégration des leptons τ dont la masse importante les rend plus sensibles à une potentielle Nouvelle Physique. En particulier sont étudiées à l'aide des collisionneurs electron/positron Belle et BaBar les désintégrations de type $\tau \rightarrow l\gamma$ et $\tau \rightarrow ll\ell$, interdites dans le modèle standard par conservation de la charge leptonique, et quasi inexistantes si l'oscillation des neutrinos est prise en compte. Leur existence à un ordre de grandeur atteignable est prédite par de nombreuses théories de Nouvelle Physique, avec entre autres la Supersymétrie [2], les modèles de petit Higgs [3], ou les modèles à quatre génération de fermions [4]. Les meilleures estimations à ce jour des limites de ratio de branchement pour chacune des expériences et chacun des canaux sont résumées dans la figure 1.

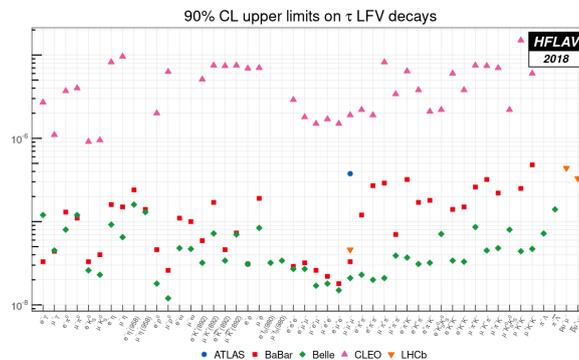


FIGURE 1 – Limites actuelles des ratios de branchement des désintégrations du lepton τ

L'étude porte particulièrement sur le canal de désintégration $\tau \rightarrow \mu\mu\mu$ dont certaines théories de Nouvelle Physique prédisent un ratio de branchement de l'ordre de grandeur de 10^{-10} soit juste en dessous des limites expérimentales actuelles, contre 10^{-54} prédit par le modèle standard en prenant en compte l'oscillation des neutrinos. Cette valeur de ratio de branchement pourrait être accessibles par le nouveau collisionneur Belle II qui a récemment commencé l'acquisition de données. L'expérience Belle II a pour objectif d'avoir collecté 50 fois plus de données que sa prédécesseur Belle à l'horizon 2027, ce qui rendrait une limite de rapport de branchement de cette ordre de grandeur accessible à la mesure.

1.2 L'expérience Belle II

L'expérience Belle II est une expérience de physique des particules opérée par KEK et située au Japon à Tsukuba, dans la préfecture d'Ibaraki. Elle se base sur l'accélérateur de particule SuperKEKB et fait l'objet d'une collaboration internationale d'environ 1000 ingénieurs et physiciens de 26 pays différents au sein de la collaboration Belle II. Belle II est un collisionneur électron/proton et remplace sa prédécesseur Belle qui a fonctionné de 1999 à 2010.

Ces accélérateurs de particules font collisionner des électrons et des positrons à l'énergie de résonance dans le centre de masse $\Upsilon(4S)$ (10.735 GeV) pour produire des mesons B ou D en grande quantité, ce type de collisionneur est ainsi également surnommé usine à B. Les usines à B ont notamment observé au début du XXIème siècle les premières violation de la symétrie CP dans le secteur des mésons B. A la résonance $\Upsilon(4S)$ la section efficace $\tau\tau$ est de 0.92 nb contre 1.05 nb pour BB, les usines à B sont donc aussi des usines à τ d'où l'intérêt de ces expériences pour l'étude de la désintégration des leptons τ . Une autre raison de l'intérêt de Belle II pour ce type d'étude est qu'un collisionneur entre électrons et positrons permet de grandement limiter le bruit de fond par rapport à un collisionneur à hadrons comme au LHC qui permet à l'inverse d'atteindre de plus hautes énergies : les protons étant chacun composés de trois quarks leur collision met un jeu un plus grand nombre de particules qui n'interagissent pas nécessairement toutes, ce qui complique la reconstruction d'un événement à partir des données mesurées. Les collisionneurs électron/proton sont donc largement utilisés pour la recherche de signaux rares, pour l'étude de la violation de saveur leptonique ou encore pour la mesure de précision de paramètres du modèle Standard comme la masse du τ .

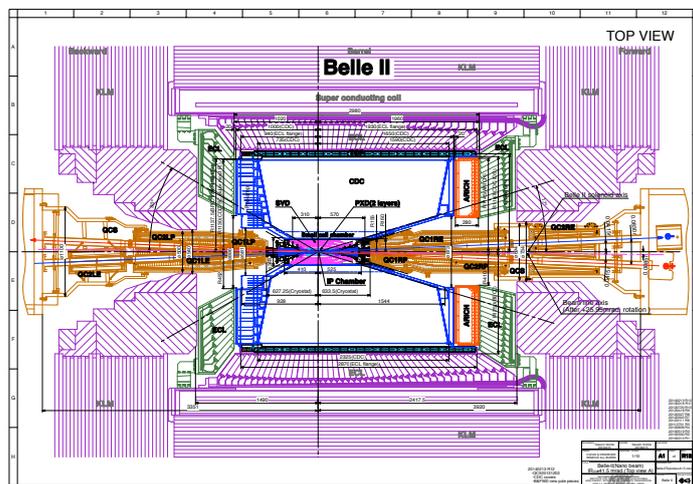


FIGURE 2 – Plan du détecteur SuperKEKB

Le détecteur Belle II décrit dans la figure 2 est composé de plusieurs capteurs différents : un Pixel Detector et un Silicon Vertex Detector qui permet de déterminer le vertex d'origine d'un groupe de particules, une chambre à fils Central Drift Chamber qui permet de suivre le déplacement des particules chargées, les détecteurs TOP et ARICH basés sur l'effet Tcherenkov aident à déterminer la masse des particules, le calorimètre électromagnétique détermine l'énergie des électrons et des photons mais est

traversé par les muons et les K longs qui sont détectés au niveau du détecteur KLM. Ces capteurs sont coordonnés par un système de Trigger qui choisit les événements à enregistrer.

Dans l'objectif de la recherche d'une Nouvelle Physique, Belle II vise principalement à augmenter significativement la quantité de données collectées par rapport à sa prédécesseur, en particulier grâce à une luminosité 40 fois supérieure pour des performances similaires ou supérieures. Cette luminosité est atteinte en multipliant par 2 le courant de faisceau, mais surtout en divisant par 20 la taille de la zone de collision. Pour s'adapter une luminosité plus importante, le détecteur a été amélioré notamment par l'ajout de nouveaux systèmes d'identification de particules (Aérogel et détecteur Ring Imaging Cherenkov), d'une nouvelle chambre à fils plus imposante (Central Drift Chamber) et d'un nouveau détecteur à semi-conducteur (Silicon Vertex Detector). Belle II ambitionne d'avoir collecté 50 fois plus de données que Belle à l'horizon 2027 soit 50 ab^{-1} contre 988 fb^{-1} pour Belle. Plus d'informations sur le détecteur Belle II peuvent être trouvées dans le Belle II Physics Book [5].

1.3 Simulation de Monte Carlo

Dans l'objectif de séparer efficacement le bruit de fond d'un signal recherché, il est utile d'utiliser une simulation de Monte Carlo pour générer statistiquement des données d'une distribution conforme à ce qui est attendu de l'expérience, en utilisant les connaissances accumulées sur le Modèle Standard. Cette méthode permet de disposer de données analogues à ce que l'on obtiendrait par les capteurs de l'expérience, tout en connaissant parfaitement les événements simulés qui ont conduit à la génération des données. Ces données simulées peuvent ensuite permettre d'entraîner des algorithmes de machine learning supervisés ou des méthodes d'optimisation. La simulation doit prendre en compte l'ensemble des désintégrations possibles ainsi que leur caractéristiques, mais aussi toutes les incertitudes des différents capteurs du détecteur et de manière générale tout ce qui a trait à l'imperfection de l'expérience. Une simulation fiable nécessite donc une connaissance statistique approfondie des interactions du Modèle Standard et du fonctionnement du détecteur. Dans le cadre de la collaboration Belle II, la génération de données de Monte Carlo fait l'objet d'un travail spécifique d'une équipe dédiée. La campagne de Monte Carlo MC13a est utilisée dans cette étude.

Dans le cadre de l'étude d'un signal rare comme celui de $\tau \rightarrow \mu\mu\mu$, il est nécessaire de considérer l'ensemble des réactions qui pourraient perturber le signal. Au sein d'un collisionneur électron/positron, les réactions à prendre en compte sont l'ensemble des désintégrations d'origine hadronique de type $u\bar{u}$, $d\bar{d}$, $c\bar{c}$, $s\bar{s}$ et $b\bar{b}$, ainsi que les désintégrations des paires de τ selon le Modèle Standard, les contributions principales provenant des canaux $u\bar{u}$ et $c\bar{c}$. Pour chacune sources de bruit de fond, une statistique de 1ab^{-1} est utilisée pour entraîner les algorithmes, et une statistique de 1.5ab^{-1} pour les tester. Le signal doit également être simulé mais il ne s'agit pas d'une désintégration du Modèle Standard et plusieurs théories de Nouvelle Physique peuvent prédire son existence à des statistiques différentes. La recherche de la désintégration $\tau \rightarrow \mu\mu\mu$ dans cette étude ne peut pas se baser sur une description dépendante d'un modèle de Nouvelle Physique. Ainsi une description générale du processus est utilisée lors de la génération des simulations de Monte Carlo. La section efficace de la désintégration recherchée est également inconnue, il est pourtant nécessaire de fusionner les données de signal et de bruit de fond pour entraîner des algorithmes de classification, et cela sans connaître les poids à appliquer pour obtenir des proportions réelles et réaliste : c'est toute la difficulté de la recherche d'un signal dont l'existence même est inconnue. Il faudra prendre en compte cet aspect dans l'analyse multivariée, notamment en utilisant seulement des variables qualitatives pour décrire le signal. On utilisera par exemple l'efficacité, insensible à la proportion réelle de signal, plutôt que le nombre absolu d'événements de signal conservés par un filtre.

2 Analyse Multivariée

2.1 Stratégie de Résolution

Dans l'objectif de recherche d'un signal dont l'existence nous est inconnu, il est nécessaire de disposer d'une stratégie efficace, en permettant d'utiliser au maximum les données produites par l'expérience, et précise, en séparant au mieux le signal du bruit de fond. La première étape qui a été réalisée en amont

de ce stage consiste à reconstruire les variables physiques et les identifications des particules à partir des données de Monte Carlo (la reconstruction est identique à la reconstruction de données réelles). Les mesures des capteurs doivent être transformées en variables physiques utilisables pour traiter le problème et interpréter les résultats : on citera de manière non exhaustive la quantité de mouvement de chacune des traces, leur masse ou encore leur énergie. On parle de génération des NTuples qui sont les tables de données utilisées dans l'analyse. Les natures des particules peuvent être déduites à l'aide de l'outil de *Particle ID* réalisé par des équipes de la collaboration. Le *Particle ID* permet de quantifier la probabilité qu'une particule soit d'une certaine nature en particulier par un chiffre entre 0 et 1. Ce travail préliminaire est réalisé en grande partie à l'aide d'outils développés par des équipes spécialisées afin qu'ils soient optimisés et puissent être utilisés par l'ensemble de la collaboration.

Les τ étant créés par paire de charges opposées, la reconstruction des NTuples permet de définir le *signal side* où l'on cherchera le signal, et le *tag side* où l'on se contentera de conserver n'importe quelle désintégration en un certain nombre de particules chargées. Ce choix du nombre de particules chargées définit la topologie de l'étude, ici nous nous intéresserons à une topologie 3×3 pas encore étudiée, pour 3 muons chargés d'un côté, et 3 particules chargées à priori quelconques de l'autre. Dans le cadre de cette étude on préselectionne largement les trois muons en imposant un *MuonID* supérieur à 0.5 et les trois particules chargées en imposant un *PionID* supérieur à 0.5 : les désintégrations en trois pions chargés ou trois pions chargés et un pion neutre représentent la majorité des désintégrations en trois particules chargées (dites 3-prong) [6]. La désintégration côté tag en 3 particules chargées est elle même un type secondaire de désintégration, puisque les désintégrations en une particule chargée (dite 1-prong) représentent plus de 80% des désintégrations concernées. Toutefois, la part restante n'est pas négligeable et son utilisation peut permettre d'améliorer encore les rapports de branchement pour ajouter plus de contraintes sur les théories de Nouvelle Physique ce qui justifie cette étude.

Il convient de définir la zone du signal sur laquelle portera l'étude pour travailler sur des données pertinentes, en se limitant à quelques écarts types d'une loi normale ajustée à la distribution du signal, souvent à partir d'un duo de variable, ici liées à la masse et à l'énergie côté signal. La région du signal est définie dans la Figure 3 par rapport aux variables $\Delta E'_{3\mu}$ et $M'_{3\mu}$ obtenues à partir des variables $\Delta E_{3\mu}$ et $M_{3\mu}$ par rotation d'un angle θ qui vise à décorrélérer les deux variables pour pouvoir simplement ajuster une loi normale pour chacune des coordonnées indépendamment. $\Delta E_{3\mu}$ correspond à la différence entre l'énergie du signal par rapport à l'énergie de trois muons au repos (dans le référentiel du laboratoire), on s'attend donc à obtenir une valeur proche de 0 pour des événements de signal, $M_{3\mu}$ correspond à la masse invariante côté signal qui doit correspondre à M_τ . On obtient les coupures suivantes en GeV, à 3σ , $0.34 < M'_{3\mu} < 0.54$ et $-1.74 < \Delta E'_{3\mu} < -1.70$. Le canal $\tau \rightarrow \mu\mu\mu$ présente peu de bruit de fond, et à partir d'une statistique de $1.5ab^{-1}$ pour chacun des bruits de fond, la simple sélection de la région du signal pour 3σ permet de rejeter 99.99% du bruit de fond pour 583823 événements de signal et 148 événements de bruit de fond restants (voir tableau) ce qui reste trop important par rapport au nombre d'événements de signal qu'on peut en réalité s'attendre à observer si la désintégration existe. On poursuit donc l'analyse pour réduire jusqu'à l'ordre de grandeur de l'unité le nombre d'événements de bruit de fond attendu : pour ne pas parasiter l'étude 1-prong, le gain en efficacité ne doit pas être compensé par un nombre de bruits de fond attendu trop élevé ce qui ne permettrait pas d'améliorer les résultats de la première étude.

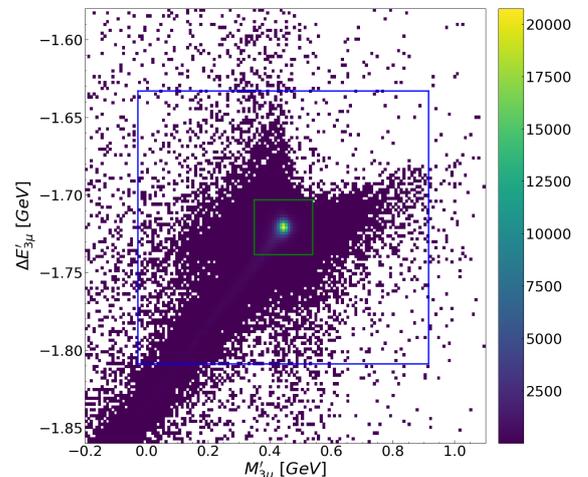


FIGURE 3 – Région du Signal

Canal	Avant Sélection	Après Sélection
$\tau \rightarrow \mu\mu\mu$	833673	583823
$\tau\tau$	122398	0
$u\bar{u}$	3002108	94
$d\bar{d}$	806608	17
$c\bar{c}$	4809856	19
$s\bar{s}$	947422	18
$b\bar{b}$ (charged)	1454479	0
$b\bar{b}$ (mixed)	1401201	0

Une fois la région du signal déterminée, il faut mettre en place un processus permettant de séparer le bruit de fond résiduel du signal. La stratégie la plus élémentaire et largement utilisée par les physiciens des particules, est de recourir à des coupures, tout au moins pour une étude préliminaire. Une coupure consiste à départager les données entre ce qu'on souhaite conserver (dans l'idéal les événements du signal) et ce qu'on souhaite éliminer (dans l'idéal le bruit de fond) en se basant sur la valeur d'une variable pour chacun des événements par rapport à une valeur limite. A ce titre, la sélection de la région du signal correspond à un ensemble de coupure portant sur l'énergie et la masse côté signal. Les distributions des deux variables concernées qui sont les deux variables les plus discriminantes du problème sont présentées ci dessous. Il convient de rappeler que la proportion du signal dans les données présentées n'est pas représentative de la réalité.

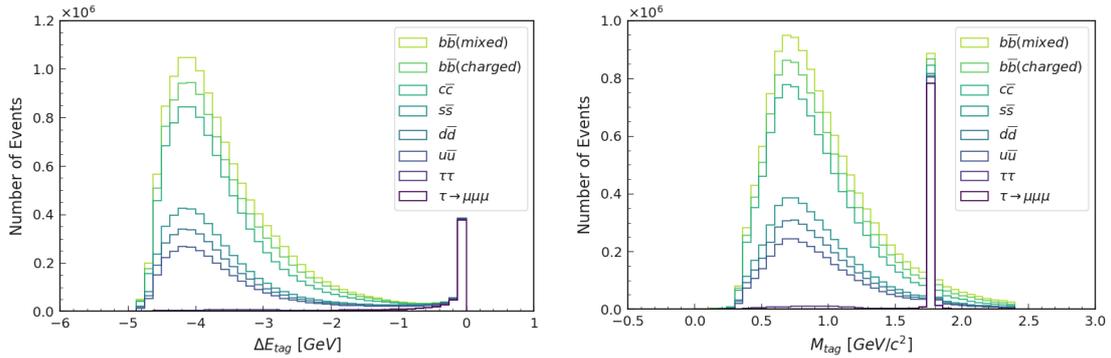


FIGURE 4 – Distribution des variables ΔE_{sig} et M_{sig}

Cette valeur limite peut être déterminée par lecture graphique ou par une méthode d'optimisation en choisissant la meilleure sélection de coupure parmi un ensemble de coupures possibles. L'évaluation du meilleur ensemble de coupure se base sur les données de Monte Carlo ou l'on sait en amont quels événements correspondent au signal ou non ce qui permet de quantifier la performance du filtre. Cette méthode simple même si moins efficace que des analyses multivariées modernes de type machine learning possède l'avantage d'être facilement interprétable. Parmi les algorithmes de machine learning, l'algorithme de *Decision Tree* est également utilisé dans le domaine de la physique des particules et permet de gagner en complexité et en performance tout en conservant un certain degré d'interprétabilité. L'algorithme de *Decision Tree* peut en effet être considéré comme une évolution plus complexe du processus de coupures simples, où chaque embranchement permet de corriger les erreurs résiduelles des embranchements précédents. Ces deux méthodes seront testées et évaluées au cours de cette étude. Pour la méthode d'optimisation, il convient auparavant de déterminer les variables à utiliser en les évaluant à l'aide de différents procédés détaillés dans la partie 2.3. La sélection des variables est d'autant plus importante que la taille des données et les puissances de calcul disponibles imposent de limiter le nombre de variables en entrée pour utiliser le maximum de données avec une même puissance de calcul. Afin de disposer d'une quantité suffisante d'événements de bruit de fond pour affiner les algorithmes, on choisira d'utiliser une région du signal défini à 15σ pour l'apprentissage des algorithmes, puis on évaluera les performances sur une région du signal à 3σ qui permet d'utiliser pleinement les variables ΔE_{sig} et M_{sig} très discriminantes.

2.2 Variables du Problème

2.2.1 Préselection

Parmi les milliers de variables possibles lors de la reconstruction, il est nécessaire de faire une préselection pour éviter de manipuler des tableaux trop lourds. Les variables suivantes sont préselectionnées et seront évaluées par différentes méthodes. Ces variables correspondent aux variables physiques élémentaires (masse, énergie, quantité de mouvement, etc), aux variables géométriques (angles, résultats du vertexing, etc) et aux variables de *ParticleID* (*MuonID* pour chacune des traces). La liste complète des variables évaluées est détaillée ci dessous. Les variables discriminantes M_{sig} et ΔE_{sig} ont déjà été utilisées pour la sélection de la région du signal et ne sont plus utilisées dans la suite.

Variable Physique	Description
M_{sig}/M_{tag}	Masse invariante côté signal (resp. tag)
$\Delta E_{sig}/\Delta E_{tag}$	Énergie côté signal (resp. tag) dans le repère du laboratoire
E_{miss}	Énergie dans le centre de masse manquante, correspond à l'énergie des particules non détectées
p_{miss}	Quantité de mouvement dans le centre de masse manquante
θ_{miss}	Angle de la quantité de mouvement dans le centre de masse manquante
<i>Thrust</i>	Quantification la forme unidimensionnelle de la désintégration : $T = \max_{\hat{n}} \frac{\sum_i \hat{p}_i \cdot \hat{n} }{\sum_i \hat{p}_i }$
<i>Sphericity</i>	Quantification la forme sphérique de la désintégration : $S = \frac{3}{2} \min_{\hat{n}} \frac{\sum_i \hat{p}_i^2}{\sum_i \hat{p}_i^2}$
$V_{sig}^{\chi}/V_{tag}^{\chi}$	Résultat du χ^2 associé à la position du vertex de désintégration du τ côté signal (resp. tag)
H_0/H_1	Moment de Fow Wolfram d'ordre 0 (resp. 1) : $H_l = \frac{4\pi}{2l+1} \sum_{m=-l}^l \left \sum_{i=1}^N Y_l^m(\Omega_i) \frac{ \hat{p}_i }{\sqrt{s}} \right ^2$ avec Y_i^j fonction harmonique sphérique, Ω_i les coordonnées angulaires et \sqrt{s} l'énergie du système
$\mu_{ID}^I/\mu_{ID}^{II}/\mu_{ID}^{III}$	<i>MuonID</i> de la trace I (resp. II-III)
$\theta^I/\theta^{II}/\theta^{III}$	Angle de la trace I (resp. II-III)

Deux méthodes élémentaires de sélection des variables du problème sont utilisées. Dans un premier temps, l'indépendance des différentes variables du problème avec la variable **IsSignal** qui indique si l'événement est un événement du signal, est évaluée à l'aide d'un test du χ^2 . Puis la corrélation linéaire entre une variable et la variable signal est estimée et présentée sous forme de tableau.

2.2.2 Test du χ^2

Le test du χ^2 permet de tester l'indépendance de deux variables aléatoires en comparant leur distribution statistique avec une distribution estimée basée sur une hypothèse probabiliste d'indépendance de manière à retourner une variable suivant une loi du χ^2 à 1 degré de liberté (nombre de classe - 1) sous la même hypothèse d'indépendance. On peut donc comparer le résultat avec les tables de la loi. Les résultats sont présentés dans un tableau ci dessous. Plus le score d'une variable est élevé, plus le test indique qu'elle serait pertinente pour discriminer signal et bruit de fond. Par rapport à la distribution d'une variable aléatoire qui suit une loi du χ^2 à un degré de liberté : l'hypothèse d'indépendance peut être rejetée avec un risque de 30% lorsque le score est supérieur à 1.07 ou avec un risque de 10% lorsque le score est supérieur à 2.7

Pour une analyse plus poussée il est possible de réaliser le test du χ^2 pour chacun des bruits de fond pris séparément, afin de mieux comprendre comment l'algorithme éliminera chacun des bruits de fond. On observe par exemple que la variable M_{tag} est très indépendante du fait que l'événement soit ou non du signal lorsque qu'on regarde spécifiquement le bruit de fond $u\bar{u}$, contrairement au bruit de fond $c\bar{c}$ (le quark c a une masse proche du lepton τ). De la même façon, les variables de *MuonID* semblent être bien plus discriminantes pour $u\bar{u}$ que pour $c\bar{c}$. On peut donc s'attendre à avoir plus de difficultés à éliminer les bruits de fond $c\bar{c}$. Les désintégrations du modèle standard des τ sont quand à elles principalement éliminées grâce aux variables de *MuonID* alors que les variables relatives à la forme de l'événement sont peu utiles car caractéristiques de l'origine hadronique ou non de l'événement.

Variable	Score χ^2
M_{tag}	39.6
ΔE_{tag}	4.3
E_{miss}	4.5e-5
p_{miss}	1.9
θ_{miss}	1.9
<i>Thrust</i>	30.9
<i>Sphericity</i>	10.9
V_{sig}^X	65.3
V_{tag}^X	2.2
H_0	4.1e-4
H_1	4.0e-3
μ_{ID}^I	324.3
μ_{ID}^{II}	318.7
μ_{ID}^{III}	355.8
θ^I	7.0
θ^{II}	9.0
θ^{III}	8.9

De ce premier test on remarque que certaines variables semblent bien plus pertinentes que d'autres pour discriminer signal et bruit de fond. Ce caractère discriminatoire peut parfois être deviné par simple observation graphique des distributions entre signal et bruit de fond, comme pour les variables *Thrust* ou *Sphericity*. Mais ces observations sont parfois en désaccord avec le score du test du χ^2 comme pour E_{miss} .

Cette étude peut être répétée à différentes étapes de l'étude, pour analyser les variables qui peuvent éliminer les bruits de fond restant. En particulier il est utile de réaliser ce test après avoir fait les coupures sur le *MuonID* (voir Partie 3.2), puisque le bruit de fond a alors significativement changé en composition.

2.2.3 Matrice de corrélation

La seconde méthode utilisée pour déterminer les variables pertinentes est une matrice de corrélation où le coefficient de corrélation linéaire de Bravais-Pearson $r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ entre deux variables X et Y est calculé par son estimateur associé $\hat{r} = \frac{\sigma_{\hat{X}Y}}{\sigma_X \sigma_Y}$. Il s'agit d'une estimation de la corrélation linéaire et de manière générale deux variables peuvent être corrélées autrement que linéairement. Une corrélation linéaire forte correspond à une valeur proche de 1 en valeur absolue, et une corrélation linéaire faible correspond à une valeur proche de zéro.

L'étude de la corrélation entre variables permet aussi de ne pas utiliser deux variables trop corrélées ce qui n'apporterait pas beaucoup d'information à l'analyse, ce qui est par exemple le cas des variables *Thrust* et *Sphericity* qui sont antagonistes l'une de l'autre.

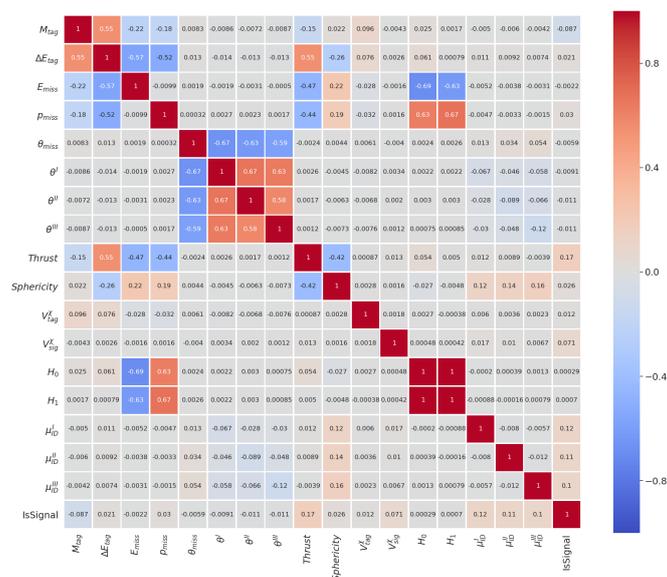


FIGURE 5 – Matrice de Corrélation

La figure 6 représente la matrice de corrélation entre toutes les variables préselectionnées, la corrélation pertinente pour la sélection de variables est celle entre la variable **IsSignal** et les autres. On cherche donc les variables ayant la plus forte corrélation avec la variable **IsSignal** dans la figure 5.

2.2.4 Variables conservées

De ces deux méthodes d'évaluation de variable, on classe les variables supposées discriminantes pour le test du χ^2 et/ou la matrice de corrélation dans le tableau suivant. On considère 1 et 0.01 comme valeur limite respectivement pour le test du χ^2 et la matrice de corrélation. On retrouve sans surprise les variables de *MuonID* qu'on espère utiles pour supprimer notamment les bruits de fond $\tau\tau$. On retient également les variables de masse et d'énergie côté tag pour éliminer les bruits de fond d'origine hadronique autre que $c\bar{c}$ et on a enfin les variables relatives à la forme de l'événement comme le *Thrust* ou la *Sphericity* qui sont supposées discriminantes pour les bruits de fond d'origine hadronique.

Variable	Validée par
M_{tag} ΔE_{tag} p_{miss} $Thrust$ $Sphericity$ V_{sig}^x μ_{ID}^I μ_{ID}^{II} μ_{ID}^{III} μ_{ID}^{IV}	χ^2 et Corrélation
θ_{miss} V_{tag}^x θ^I θ^{II} θ^{III}	χ^2
E_{miss} H_0 H_1	Aucun des Tests

Dans la suite de l'étude, les variables validées par les deux tests seront utilisées en priorité, la pertinence de l'ajout des variables validées par un seul des deux tests pouvant être estimé au cas par cas en évaluant les gains de performance. En particulier on essaiera les variables relatives à l'énergie/au moment manquant qui correspond à l'énergie/au moment emporté par les neutrinos non détectés par les capteurs. Ces neutrinos ne sont pas produits lors des événements de signal et on peut penser que ces variables peuvent être discriminantes comme on peut le voir à l'aide de leur distribution graphique sur la Figure 6.

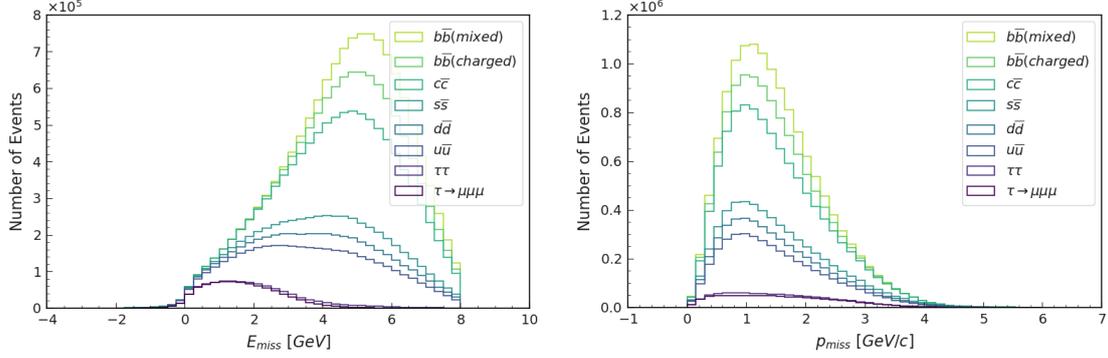


FIGURE 6 – Distribution des variables E_{miss} et p_{miss}

3 Résolution Numérique

3.1 Figure de Mérite

Lors de la séparation du signal et du bruit de fond, deux figures de mérite distinctes apparaissent intuitivement. Une première qui quantifie la quantité de bruit de fond rejeté et une seconde qui quantifie la quantité de signal conservée. Il est donc nécessaire de combiner ces deux aspects pour obtenir une figure de mérite unique qui pourra être utilisée pour comparer nos différents algorithmes ou pour déterminer leurs meilleurs hyperparamètres. Il est nécessaire de quantifier la conservation du signal sous forme de ratio pour ne pas avoir besoin de postuler en amont une quantité de signal particulière. Une figure de mérite fréquemment utilisée en physique des particules pour la recherche de phénomène nouveau modélisé par une loi de Poisson est la figure de mérite de Punzi pour une signification statistique de a -sigma. [7].

$$FOM_{Punzi} = \frac{\epsilon}{\frac{a}{2} + \sqrt{N}} \quad \text{avec } \epsilon \text{ l'efficacité du signal et } N \text{ le nombre d'événements de bruit de fond}$$

Cette figure de mérite quantifie la sensibilité de la recherche à de nouveaux signaux, et est indépendante de tout a priori sur le signal recherché. Optimiser cette figure de mérite permet de traiter l'information de façon optimale pour la recherche d'un nouveau signal.

3.2 Méthode de Coupures

3.2.1 Algorithme

La méthode la plus simple pour séparer le signal et le bruit de fond est une méthode de coupures où l'on cherche pour chacune des variables une valeur limite permettant de séparer les données en un ensemble accepté et un ensemble rejeté. Dans un premier temps et pour éviter de traiter de trop nombreuses variables en même temps, on optimise par rapport aux variables de *MuonID* pour sélectionner les muons, avant d'optimiser par rapport au reste des variables.

Pour trouver le meilleur ensemble de coupure par rapport aux trois variables de *MuonID*, un algorithme parcourt une grille de coupures possibles et pour chacun des ensembles de coupure évalue la figure de mérite pour ensuite sélectionner le meilleur ensemble de coupure d'après la figure de mérite. Cette approche très demandeur en puissance de calcul n'est possible qu'avec un faible nombre de variables.

Pour prendre en compte les différences de quantité de mouvement qui peuvent permettre ou non à une particule d'atteindre les détecteurs, les traces sont séparées en trois sous ensemble correspondant à des intervalles de quantité de mouvement. Une valeur limite de $MuonID$ est déterminée par optimisation pour chacune des catégories de trace : $p_\mu < 0.7$ GeV, 0.7 GeV $< p_\mu < 1$ GeV et 1 GeV $< p_\mu$.

Après avoir sélectionné les événements mettant en jeu trois muons, il reste à déterminer si ces événements proviennent bien de la désintégration d'un tau. On utilise les variables restantes dans un autre algorithme d'optimisation. Le nombre important de variables implique de changer de méthode d'optimisation et d'éviter de comparer naïvement tous les ensembles de coupure possibles d'une grille de valeurs. L'algorithme d'optimisation utilisé sélectionne aléatoirement une variable à chaque itération, puis la figure de mérite est optimisée localement par rapport à cette variable. Cette méthode d'optimisation est vulnérable aux optimums locaux et on choisit de sélectionner le meilleur ensemble de coupure parmi ceux proposés par l'algorithme pour différentes initialisations aléatoires. Cette méthode permet de traiter un nombre plus important de variables de manière plus fine, tout en permettant un gain de performance.

3.2.2 Résultats

L'algorithme d'optimisation est entraîné sur la région du signal à 15σ pour avoir plus de bruit de fond, puis est testé sur la région du signal à 3σ . La similarité des distributions des bruits de fond dans les deux zones est vérifiée, les courbes sont accessibles en annexe. Les figures de mérites utilisées correspondent à l'ensemble du processus et l'efficacité et le rejet du bruit de fond sont estimés par rapport aux données générées avant toute forme de sélection, soit 13200000 événements de signal pour les données de test, afin de pouvoir mieux comparer les différentes méthodes entre elles. La méthode de coupure basée sur l'optimisation de la figure de mérite de Punzi permet d'obtenir une efficacité de 2.6% et un nombre de bruit de fond attendu de 2, pour une statistique de $1.5ab^{-1}$. Cette efficacité est à comparer avec le ratio de branchement de la désintégration 3-prong d'une valeur de 13.3% [6]. L'ensemble de coupure correspondant est décrit ci dessous. Tous les tableaux de cette partie sont basés sur l'échantillon de test à 3σ .

Les coupures sur le $MuonID$ permettent principalement d'éliminer les faux muons ou les particules à l'ID incertaine, le plus souvent des pions chargés aux caractéristiques semblables aux Muons. Le pic de la distribution de $MuonID$ autour de 1 permet de conserver un nombre important d'événements pour la suite de l'étude. On observe que plus la quantité de mouvement est élevée et plus les coupures sont sélectives : lorsque la quantité de mouvement est importante la particule atteint tous les capteurs et l'évaluation de l'ID est meilleure. On pourrait toutefois préférer utiliser la même coupure quelque soit la valeur de la quantité de mouvement, car les coupures varient peu et une seule valeur limite permet de simplifier les calculs d'incertitude. Les coupures sur le $MuonID$ éliminent significativement tous les bruits de fond, mais plus particulièrement $u\bar{u}$ et moins particulièrement $c\bar{c}$.

μ_{ID}		
$p_\mu < 0.7$ GeV	0.7 GeV $< p_\mu < 1$ GeV	1 GeV $< p_\mu$
$\mu_{ID} > 0.98$	$\mu_{ID} > 0.995$	$\mu_{ID} > 0.995$

Canal	Avant Coupures μ_{ID}	Après Coupures μ_{ID}
$\tau \rightarrow \mu\mu\mu$	583823	377532
$\tau\tau$	0	0
$u\bar{u}$	94	5
$d\bar{d}$	17	0
$c\bar{c}$	19	5
$s\bar{s}$	18	1
$b\bar{b}$ (charged)	0	0
$b\bar{b}$ (mixed)	0	0

Après les coupures sur les variables de $MuonID$, on peut étudier les distributions de bruit de fond restant pour déterminer des variables qui deviendraient utiles. On remarque notamment que les moments de Fox Wolfram ainsi que les variables relatives à l'énergie et au moment manquant pourraient être utiles, certaines de leur distribution sont affichées ci dessous, les restantes sont en annexe.

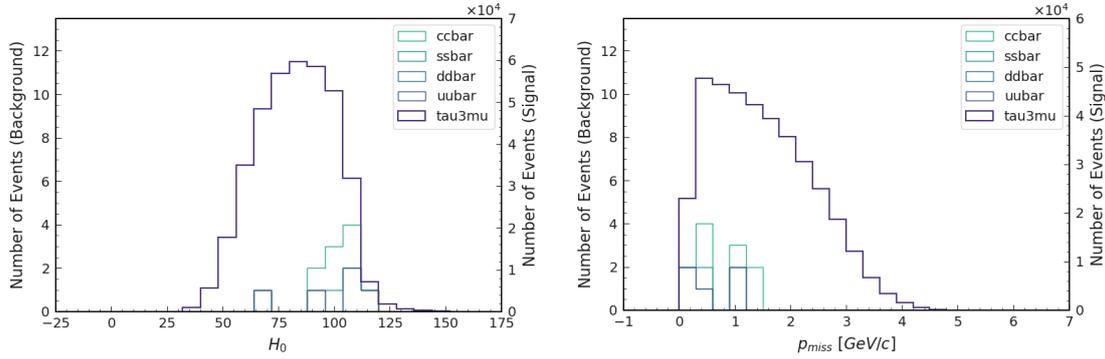


FIGURE 7 – Distribution des variables H_0 et p_{miss} après les coupures sur le MuonID

On utilise ensuite les variables restantes pour affiner la sélection. Des différents essais l'ensemble de coupure suivant offre les meilleurs performances. L'ensemble de variable qui permet les meilleurs performances sont M_{tag} , H_0 , $Sphericity$ et V_{sig}^X . Le $Thrust$ n'est pas utilisé car trop corrélé à la $Sphericity$ qui permet aussi de rejeter ici les événements de forme sphérique. V_{sig}^X ne coupe qu'à 0 ce qui élimine les erreurs de vertexing qui prennent une valeur égale à -1, cette coupure permet d'éliminer les bruits de fond de type $c\bar{c}$. L'effet de chaque coupure prise séparément est décrite dans le tableau ci dessous. Les coupures les plus efficaces sont celles utilisant M_{tag} et H_0 .

M_{tag}	H_0	$Sphericity$	V_{sig}^X
$M_{tag} < 1.35 \text{ GeV}$	$H_0 < 102$	$S < 0.41$	$V_{sig}^X > 0$

Canal	Avant Coupure	Après Coupure				
		M_{tag}	H_0	$Sphericity$	V_{sig}^X	Toutes
$\tau \rightarrow \mu\mu\mu$	377532	355806	323418	377528	377532	291359
$\tau\tau$	0	0	0	0	0	0
$u\bar{u}$	5	3 (-2)	2 (-3)	5	4 (-1)	1 (-4)
$d\bar{d}$	0	0	0	0	0	0
$c\bar{c}$	5	3 (-2)	3 (-2)	5	0 (-5)	0 (-5)
$s\bar{s}$	1	0 (-1)	1	1	0 (-1)	0 (-1)
$b\bar{b}$ (charged)	0	0	0	0	0	0
$b\bar{b}$ (mixed)	0	0	0	0	0	0

Les figures de mérite de l'algorithme sont répertoriées ci dessous pour pouvoir estimer la limite de ratio de branchement accessible et pour pouvoir comparer l'algorithme aux autres méthodes dans la suite.

Figures de Mérite	Méthode de coupures
Efficacité	2.2%
Nombre d'événements de bruit de fond	1
Punzi	0.006

3.3 Boosted Decision Tree

3.3.1 Algorithme

Pour améliorer la sélection, on utilise une méthode plus poussée d'analyse, le *Boosted Decision Tree*. L'algorithme de *Boosted Decision Tree* est basé sur celui de *Decision Tree* où une série de coupures est déterminée par l'algorithme en optimisant la "pureté" du signal quantifiée par l'indice de diversité de

Gini ou l'entropie de Shanon, afin d'obtenir en sortie un arbre de décision qui sépare les événements dans différentes sous catégorie appelées feuilles. Les deux métriques de pureté sont définies ci dessous pour un sous ensemble ou la fraction des éléments de l'ensemble dans la classe i est noté f_i . Les deux métriques sont minimales quand le signal est pure, c'est à dire composé uniquement d'une classe.

$$I_G(f) = \sum_i f_i(1 - f_i) \quad \text{ou} \quad S(f) = - \sum_i f_i \ln f_i$$

A l'algorithme de *Decision Tree* on ajoute le principe général de *Boosting* qui permet de combiner plusieurs algorithmes de classifications en un seul pour améliorer les performances. Pour cela on crée à chaque itération un nouveau classeur en donnant plus de poids aux événements jusque la mal classé par les classeurs précédent. Différentes méthodes de *Boosting* existent mais une des plus utilisées est celle d'*AdaBoost* pour *Adaptive Boosting*, c'est la méthode qui sera utilisée dans cette étude. Les principaux paramètres de l'algorithme à définir concernent la taille de l'arbre, c'est à dire le nombre de décisions successives maximum et le nombre d'événements minimum par feuille pour chacun des arbres, auxquels on ajoute le nombre d'estimateurs pour la méthode de *Boosting*. En raison du fort déséquilibre entre le nombre d'événement de signal et de bruit de fond au niveau de la région du signal, le paramètre de nombre d'événements minimums par feuille est important car il doit permettre de classer les deux classes.

A l'image de l'algorithme de coupure, on entraîne l'algorithme sur une région du signal coupée à 15σ avant de le tester sur une région du signal coupée à 3σ . Après avoir bouclé sur différents hyperparamètres, on choisit d'utiliser un nombre minimum d'événements par feuille de 500, une profondeur maximum de l'arbre de 4 et de booster sur 200 estimateurs. On obtient en sortie une variable de décision dont on choisit la valeur limite en optimisant la figure de mérite de Punzi. L'algorithme de *Boosted Decision Tree* a été essayé en remplacement de l'ensemble des coupures ou en aval des coupures sur les *MuonID* et les résultats sont présentés ci dessous.

3.3.2 Résultats

On entraîne l'algorithme puis on le teste sur les mêmes données que pour la méthode de coupures. La Figure 7 représente les distributions de la variable de décision pour le bruit de fond et le signal pour les deux méthodes. La valeur limite de coupure sur la variable de décision obtenue après optimisation est de 0.27 et 0.17 respectivement.

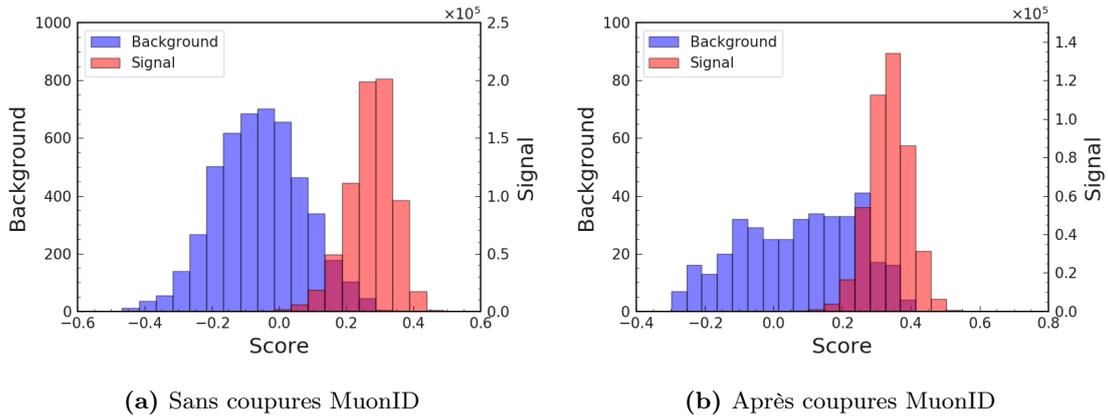


FIGURE 8 – Distribution de la variable de décision pour les deux méthodes

On obtient une distribution pour le bruit de fond qui est beaucoup plus piqué dans le cas où seulement l'algorithme de *Boosted Decision Tree* est utilisé, la distribution pour le signal ne change pas significativement. On en déduit que cette différence est dû au nombre d'événement de bruit de fond accessibles pour l'apprentissage, qui est un ordre de grandeur moindre après les coupures sur le *MuonID*. On peut ensuite évaluer les performances de l'algorithme et comparer les différentes méthodes directement en comparant leur figure de mérite de Punzi.

Figures de Mérite	<i>Boosted Decision Tree</i>	Coupures <i>MuonID</i> + <i>Boosted Decision Tree</i>
Efficacité	2.7%	2.9%
Nombre d'événements de bruit de fond	0	5
Punzi	0.010	0.06

L'algorithme de *Boosted Decision Tree* permet d'obtenir de meilleures performances qu'une méthode basée sur des coupures en multipliant par presque deux la figure de mérite de Punzi, notamment en éliminant mieux le bruit de fond. L'algorithme permet d'obtenir des résultats significativement meilleur quand il est entraîné directement à partir de la sélection du signal notamment car il dispose ainsi de plus d'exemples d'événements de bruit de fond qui sont particulièrement filtrés par les coupures sur le *MuonID*. Quand l'algorithme est utilisé à partir de la sélection de la région du signal, il permet de supprimer complètement le bruit de fond sur l'événement de signal tout en conservant une part importante du signal, même supérieure à la part conservée par la méthode de coupures.

3.4 Méthodes de Coupures avec Boosted Decision Tree

Afin de conserver une méthode de coupures facilement interprétable tout en améliorant les performances, on peut ajouter aux variables utilisées dans la méthode de coupures, la variable de décision d'un *Boosted Decision Tree*. On remarque dans l'ensemble de l'étude que les variables relatives à la forme de l'événement sont très discriminantes pour séparer le bruit de fond du signal. On propose donc d'utiliser un *Boosted Decision Tree* pour synthétiser les informations relatives à la forme géométrique de l'événement sous la forme d'une unique variable \mathcal{M}_{geo} . On utilise les moments de Fox Wolfram, la *Sphericity* et le *Thrust* ainsi que les variables de Vertexing. On entraîne donc l'algorithme sur les données autour de la région du signal à 15σ avant les coupures sur le *MuonID* pour avoir suffisamment de données de bruit de fond et on ajoute cette variable à tous les autres NTuples. Les distributions de \mathcal{M}_{geo} pour les données d'entraînement et de test à 15σ sont affichées dans la figure 8.

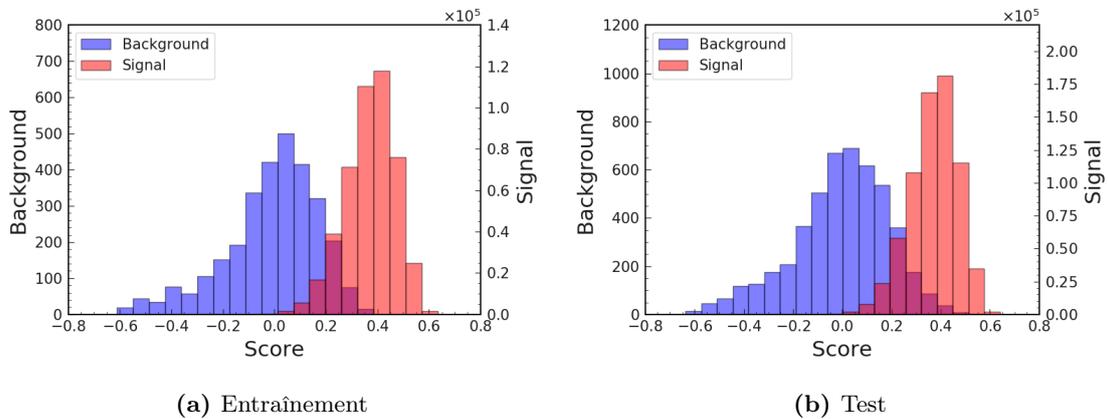


FIGURE 9 – Distribution de la variable de décision pour les données d'entraînement et de test

On peut ensuite reprendre la même méthode d'optimisation des coupures de la partie 3.2, en remplaçant les variables sur lesquelles sont basées le *Boosted Decision Tree* par \mathcal{M}_{geo} . La variable \mathcal{M}_{geo} suffit d'elle-même à discriminer efficacement signal et bruit de fond sur les données, mais on choisit d'ajouter une coupure sur M_{tag} pour plus de robustesse. Le détail des coupures et de leurs effets est détaillées dans le tableau ci-dessous.

M_{tag}	\mathcal{M}_{geo}
$M_{tag} < 1.35 \text{ GeV}$	$\mathcal{M}_{geo} < 0.34$

Canal	Avant Coupure	Avant Coupure		
		M_{tag}	\mathcal{M}_{geo}	All
$\tau \rightarrow \mu\mu\mu$	377532	355005	308034	294594
$\tau\tau$	0	0	0	0
$u\bar{u}$	5	3 (-2)	0 (-5)	0 (-5)
$d\bar{d}$	0	0	0	0
$c\bar{c}$	5	3 (-2)	0 (-5)	0 (-5)
$s\bar{s}$	1	0 (-1)	0 (-1)	0 (-1)
$b\bar{b}$ (charged)	0	0	0	0
$b\bar{b}$ (mixed)	0	0	0	0

Cette méthode permet d'éliminer complètement le bruit de fond sur l'échantillon de test, au prix d'une efficacité moindre. Les performances de l'algorithme sont détaillées ci dessous.

Figures de Mérite	Coupsures $MuonID$ + Coupsures avec Variable BDT
Efficacité	2.2%
Nombre d'événements de bruit de fond	0
Punzi	0.09

3.5 Conclusion sur les méthodes numériques

L'application des algorithmes de Coupures ou de *Boosted Decision Tree* permet de séparer efficacement le bruit de fond notamment grâce à une région du signal avec très peu d'événements de bruit de fond. Pour le calcul du ratio de branchement atteignable par l'algorithme, on choisit la dernière version de l'algorithme basée sur la variable \mathcal{M}_{geo} qui permet un gain important de performance tout en étant moins opaque qu'un algorithme d'apprentissage gérant toutes les variables d'un coup. Le fait de ne pas utiliser les variables de $MuonID$ dans le BDT permet également de s'adapter facilement aux modifications du $MuonID$ qui n'est pas nécessairement encore optimisé pour correspondre aux données réelles en raison du faible nombre de données déjà enregistrées par le détecteur.

4 Ratio de Branchement

4.1 Méthode de Calcul

A partir des résultats des algorithmes sur les données de test, on peut évaluer les limites du ratio de branchement atteignables avec une certaine quantité de données, en supposant que la simulation est fidèle à l'expérience. L'expérience Belle II n'a pas encore accumulé suffisamment de données pour faire un calcul du ratio de branchement qui dépasse celui de Belle ou a été mesurée la meilleure limite du ratio de branchement de $\tau \rightarrow \mu\mu\mu$ qui est de 2.1×10^{-8} pour un intervalle de confiance à 90%. En supposant la non observation de signal, on essaie de déterminer le ratio de branchement accessible avec un intervalle de confiance pour une incertitude de 90%. Pour obtenir un ratio de branchement intéressant il faut utiliser au maximum les données disponibles et non seulement les désintégrations 3-prong. Les résultats de cette étude sont donc ajoutés à une étude sur les désintégrations 1-prong. L'étude choisi est celle d'Alberto Martini, doctorant à l'INFN et membre du Tau Working Group de la collaboration Belle II, l'étude en question, encore à un stage préliminaire, dispose d'une efficacité de 18.0% et d'un nombre d'événements de bruit de fond pour $500 fb^{-1}$ de 1.10.

Les ensembles des désintégrations 1-prong et 3-prong sont disjoints, on peut additionner les efficacités et le nombre de bruit de fond attendu une fois ramenés à la même taille statistique. La limite du ratio de branchement est obtenue en cherchant un intervalle de confiance pour le paramètre d'une loi de Poisson qui représente l'événement rare recherché par hypothèse.

4.2 Résultats

L'estimation de la limite du ratio de branchement peut se faire pour différentes quantités de données, ce qui influe sur les incertitudes statistiques. La limite du ratio de branchement obtenue en ajoutant ou non les résultats de l'analyse 3-prong à l'analyse 1-prong pour différentes statistiques sont présentées dans le tableau ci dessous.

Analyse	100 fb^{-1}	500 fb^{-1}	1 ab^{-1}
1-prong	7.89×10^{-8}	2.08×10^{-8}	1.25×10^{-8}
1-prong + 3-prong	7.25×10^{-8}	1.98×10^{-8}	1.20×10^{-8}

Il apparaît ainsi qu'il serait possible de dépasser les ratios de branchement de Babar et de Belle grâce à Belle II pour $\tau \rightarrow \mu\mu\mu$, à partir de 500 fb^{-1} ce qui est largement atteignable par Belle II qui est supposé recueillir jusqu'à dix fois cette quantité de données. On remarque également que 500 fb^{-1} représente la moitié de la quantité de données enregistrée par Belle tout au long de son fonctionnement, pour une limite du ratio de branchement similaire, ce qui provient de l'amélioration des capteurs et des différentes méthodes numériques utilisées tout au long du processus. L'ajout des résultats de l'analyse 3-prong permet de gagner environ 5% sur le ratio de branchement pour 500 fb^{-1} , plus la quantité de données est importante et plus l'écart tend à augmenter.

Conclusion

En utilisant les données de désintégrations 3-prong pour le canal $\tau \rightarrow \mu\mu\mu$ en complément d'une analyse 1-prong, on obtient un gain d'efficacité de 2.2% et un nombre de bruit de fond attendu additionnel de 0 ou 1 respectivement pour une méthode utilisant un BDT pour une méthode de coupures; pour 1 ab^{-1} . Cet ajout permet d'obtenir un ratio de branchement 5% inférieur par rapport à une analyse 1-prong seule, ce qui permet d'envisager de dépasser les limites de ratio de branchement de Belle et Babar, dès l'enregistrement de 500 fb^{-1} par l'expérience Belle II. La poursuite de l'analyse nécessite la prise en compte des incertitudes systématiques du détecteur. L'analyse pourra ensuite être appliquée aux données réelles, lorsqu'elles seront disponibles, pour enregistrer pour la première fois une désintégration $\tau \rightarrow \mu\mu\mu$ ou calculer une nouvelle limite du ratio de branchement et poser de nouvelles contraintes sur les théories de Nouvelle Physique.

Références

- [1] J. C. Evans R. Davis Jr. and B. T. Cleveland. Solar neutrino problem. 1978.
- [2] I. Aitchison. Supersymmetry and the mssm : An elementary introduction. 2005. [arXiv :hep-ph/0505105v1].
- [3] M. Schmaltz. Introducing the little higgs. 2002. [arXiv :hep-ph/0210415].
- [4] D. Whiteson L. M. Carpenter, A. Rajaraman. Searches for fourth generation charged leptons. 2010. [arXiv :1010.1011].
- [5] E. Kou et al (Belle II). *The Belle II Physics Book*. 2018. [arXiv :1808.10567].
- [6] D. Asner et al (Heavy Flavor Averaging Group). Averages of b-hadron, c-hadron, and τ -lepton properties. 2010. [arXiv :1010.1589].
- [7] G. Punzi. Sensitivity of searches for new signals and its optimization. 2003. [arXiv :physics/0308063].

Annexe

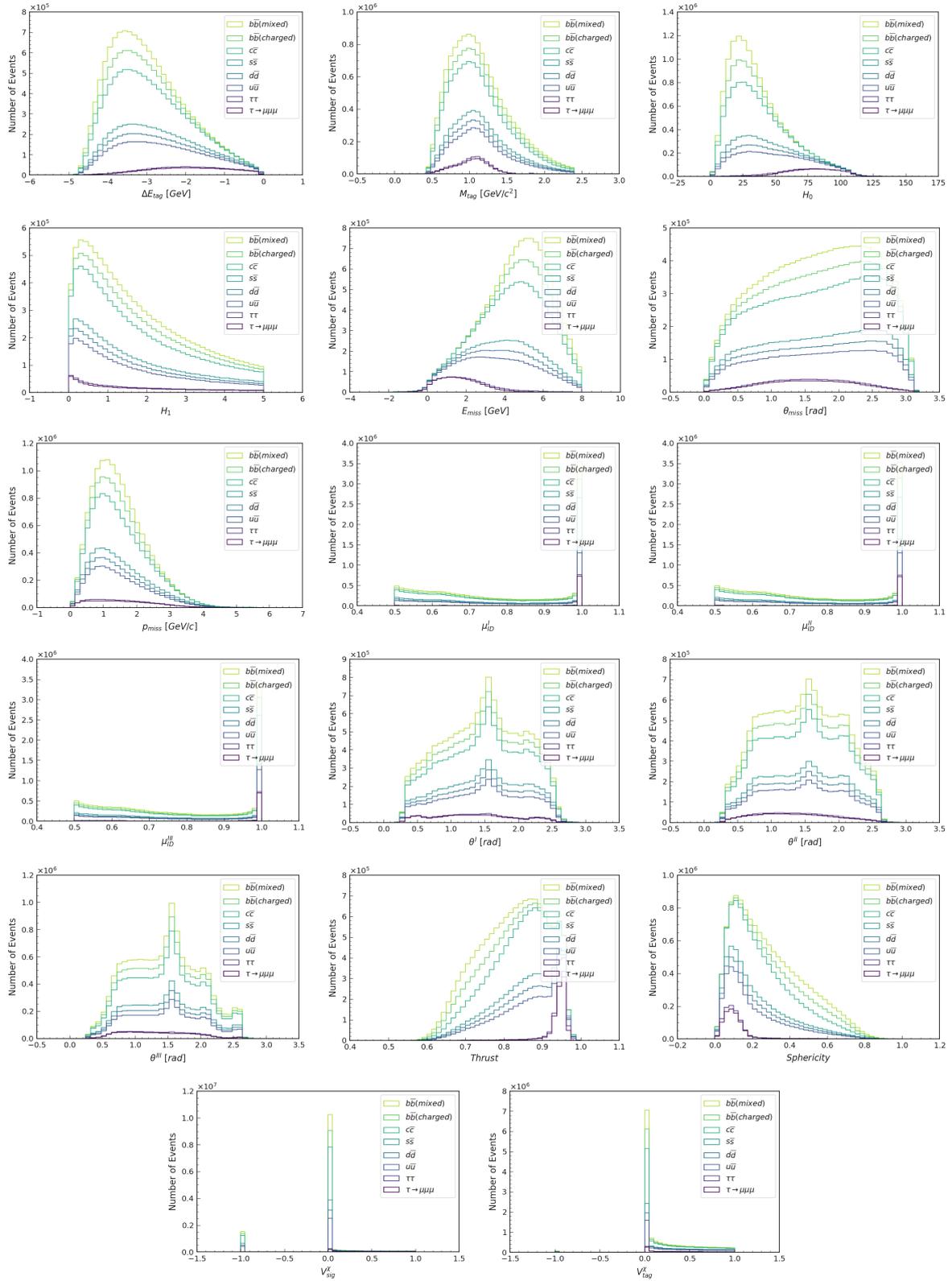


FIGURE 10 – Distribution des variables reconstruites

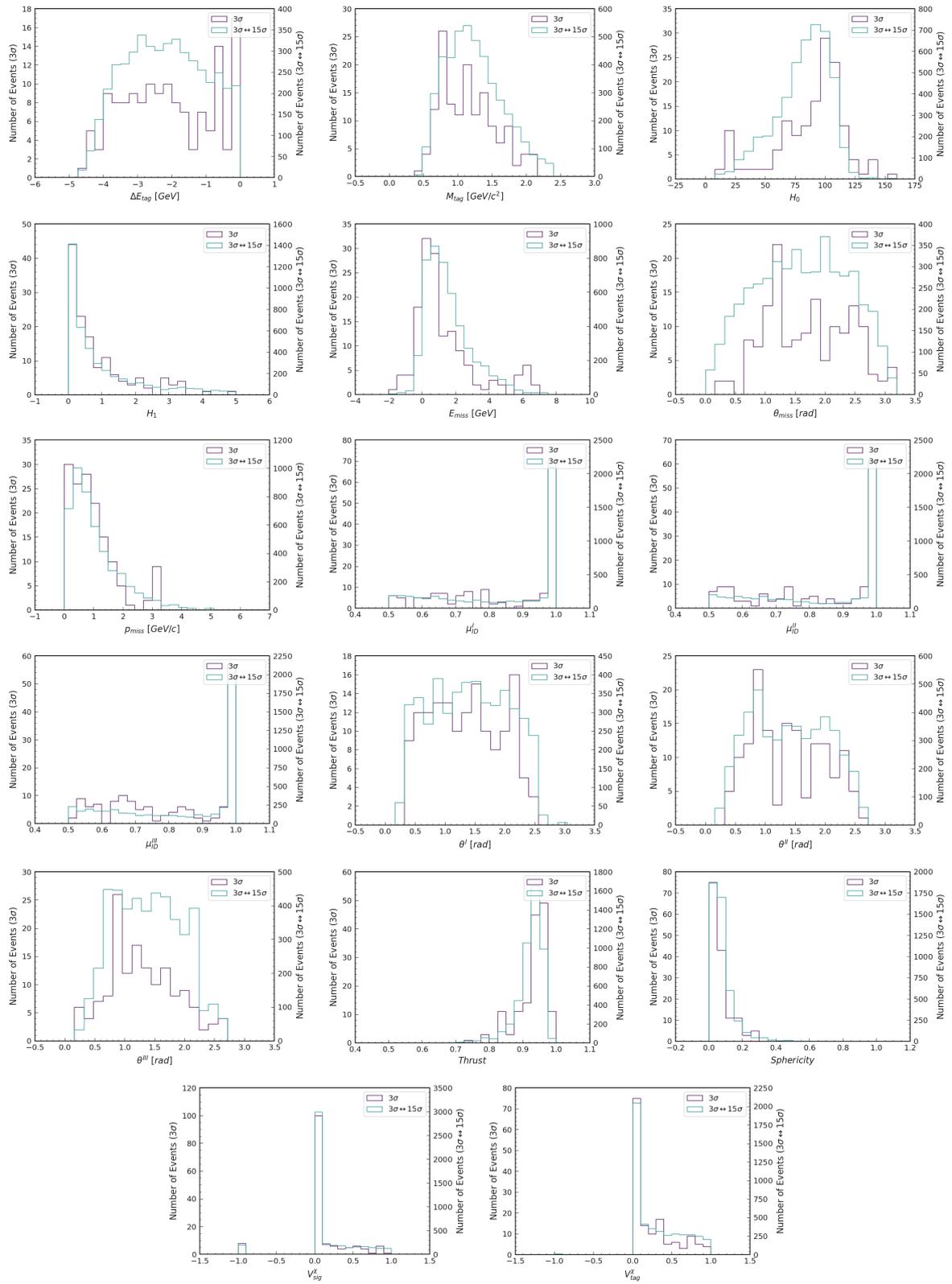


FIGURE 11 – Distribution des bruits de fond dans différentes zones de la région du signal

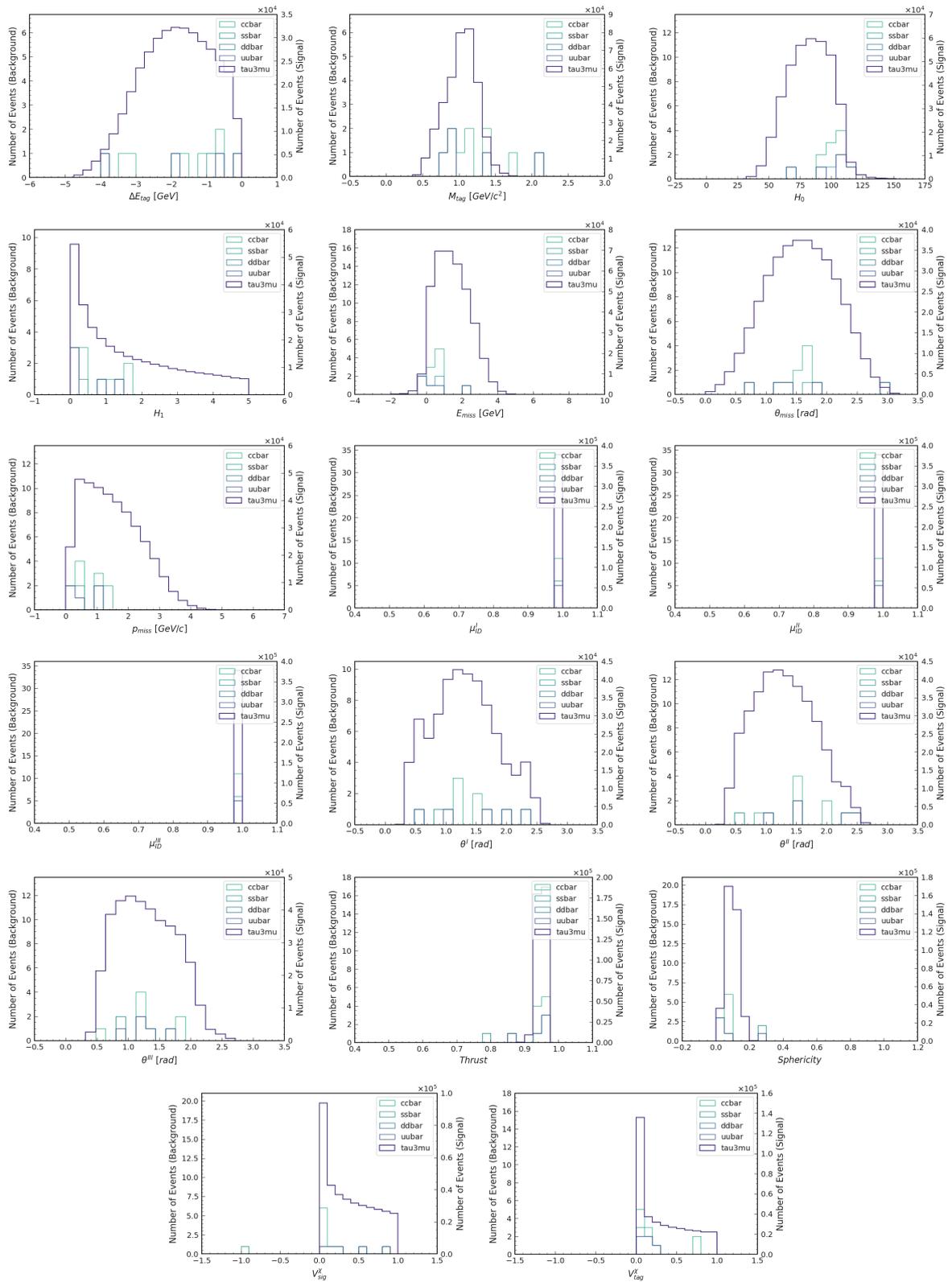


FIGURE 12 – Distribution des variables reconstruites, après coupes sur le MuonID