

Rapport de Stage

**Optimisation de coupure sur les variables
caractéristiques des mésons B dans le cadre de
l'analyse du canal**

$$B_s^0 \rightarrow \mu^- \mu^+$$

avec le détecteur LHCb au LHC

Eté Rémi

Juin - Juillet 2011

Centre de **P**hysique des **P**articules de **M**arseille
Université de la Méditerranée, Faculté des Sciences

Tuteur de Stage : Mancinelli Giampiero

« Imaginez que vous regardez loin, très loin devant vous et que vous avez
une très bonne vue, une très très bonne vue, alors vous arriverez à voir...
votre dos ! »
Albert Einstein

Table des matières

1	Introduction	4
1.1	Centre de Physique des Particules	4
1.2	L'expérience LHCb...	4
1.3	... et ses détecteurs [3]	5
1.4	La désintégration du méson rare $B_s \rightarrow \mu^- \mu^+$	5
2	Les outils d'analyse	6
2.1	Environnement de travail : Linux [5]	6
2.2	Langage de programmation : C++	7
2.3	ROOT : Plateforme multilangage [8]	7
3	But du stage	7
3.1	Le travail de l'équipe LHCb au sein du CPPM	7
3.2	But du stage	8
4	Analyse numérique	8
4.1	Simulation Monte Carlo	8
4.2	Les variables	9
4.3	Choix des variables de coupure	11
4.4	Etude de corrélation des variables	22
4.5	Algorithme Random Grid Search (GRS)	33
4.5.1	Un premier algorithme : Optimiser les coupures une a une	33
4.5.2	Algorithme simultané : Grid Random Search	34
5	Conclusion	40
5.1	Résumé de l'étude	40
5.2	Conclusion d'étude	41
5.3	Conclusion personnelle	42

Remerciements

Je tenais à remercier mon maître de stage, Giampiero Mancinelli, de m'avoir accueilli au sein de son équipe de recherche pour effectuer ce travail pendant sept semaines au laboratoire du CPPM.

Je remercie l'équipe LHCb pour son accueil : Renaud, Justine, Cosme, Mathieu et particulièrement Emilie qui a passé du temps sur mon code à débbugger ou à paufiner certains détails ou tout simplement à me faire comprendre ce que je codais.

Je remercie aussi Stephane pour m'avoir pris en stage au CERN à Genève suite à ce stage au CPPM à Marseille.

1 Introduction

1.1 Centre de Physique des Particules

Le CPPM (Centre de Physique des Particules de Marseille) est une UMR du CNRS et de l'IN2P3 [6]. Le personnel permanent du laboratoire compte environ un quarantaine de chercheur, environ 70 ingénieurs, techniciens et administratifs auxquels s'ajoutent de nombreux doctorants et stagiaires.

La vocation de ce laboratoire est la recherche fondamentale en physique des particules, astroparticules et particules élémentaires de la matière dans l'Univers, compréhension de la composition de l'Univers primordial (Big Bang) par l'étude et l'observation de supernovae.

Le CPPM participe, en collaboration avec des grands instituts, centre de recherche et entreprises, à des expériences sur les plus grands accélérateurs de particules, européens et américains dont le plus grand accélérateur du monde le LHC ainsi que sur des grands télescopes terrestres, spatiaux et sous-marins (ANTARES).

1.2 L'expérience LHCb...

L'expérience LHCb est installée sur le grand collisionneur proton-proton, le LHC, au Cern à Genève. Elle étudie les différences de comportement entre les particules et les antiparticules et les désintégrations rares dans les systèmes des mésons beaux.

La théorie actuelle rend compte des différences de comportement particules-antiparticules observées dans le laboratoire auprès des kaons neutres et des mésons beaux notamment la violation CP. Mais cette formulation théorique laisse la porte ouverte à d'autres mécanismes qui pourraient trouver leurs origines dans les extensions du modèle standard comme les modèles supersymétriques. Aujourd'hui, les mesures très précises de ces différences et la recherche des désintégrations très rares des mésons beaux vont permettre de confirmer le modèle standard ou bien de l'invalider et de laisser place à une théorie plus générale englobant le modèle standard.

Au CPPM, le groupe LHCb est le maître d'oeuvre du système de sélection à muons de premier niveau (Trigger L0) qui identifie les collisions proton-proton contenant un méson beau. Cette équipe est chargée d'analyser les données concernant la phase faible β_s dans la désintégration $B_s \rightarrow J/\Psi$ et le rapport de branchement de la désintégration $B_s \rightarrow \mu^+\mu^-$ grâce à des algorithmes conçus au CPPM.

1.3 ... et ses détecteurs [3]

Le système de Tracking du LHCb est composé de multiples détecteurs ayant pour principale fonction de reconstruire les traces des particules et de les identifier. On y trouve entre autre :

- le VELO
- Le RICH I
- Le RICH II
- le calorimètre électromagnétique ECAL (en bleu)
- le calorimètre hadronique HCAL (en gris)
- le détecteur à muons (en vert)
- un aimant (en bleu rayé)

FIGURE 1 – Les détecteurs du LHCb

C'est dans le VELO que se situe la zone des vertex primaires, point d'impact des paquets de protons. Toutes les particules passent ensuite dans le RICH I qui identifie les hadrons π^\pm , K^\pm et p^\pm par effet Cherenkov.

Le détecteur mesure la radiation émise par les différents hadrons ce qui permet de les différencier et de trouver leur vitesse. Le dipole magnétique dévie les particules chargées qui seront détectés dans le calorimètre électromagnétique tandis que les hadrons seront détectés dans le calorimètre hadronique. Le détecteur le plus important dans notre étude est placé en dernier : le détecteur à muons.

1.4 La désintégration du méson rare $B_s \rightarrow \mu^- \mu^+$

Une désintégration très spécifique est étudiée dans le cadre du modèle standard, la désintégration rare $B_s \rightarrow \mu^+ \mu^-$. Lors de la collision proton-proton, une gerbe de particule est produite dont des quarks b (beauty). Ces particules se produisent avec un angle très petit par rapport au faisceau de proton (40 % sont produit dans l'angle d'acceptance du détecteur LHCb soit entre 300 et 400 mrad) avec leurs anti-particules \bar{b} associées. L'angle des \bar{b} avec le faisceau de proton se trouve être le même que celui entre le b et le faisceau.

La Figure 2 montre que les angles du b et du \bar{b} sont fortement corrélés [2]. Très rapidement les b et \bar{b} se hadronisent avec un autre quark pour former un méson B, particule constituée de deux quarks. Ces deux mésons B oscillent chacun entre deux états : l'état B_s et l'état \bar{B}_s .

Quand un méson est dans l'état B_s , l'autre est dans l'état \bar{B}_s . Au bout d'une distance de 1 cm environ, ils se désintègrent et produisent une nouvelle gerbe de particules. Parmi les désintégrations possibles on trouve la désintégration rare $B_s \rightarrow \mu^+ \mu^-$. Le modèle standard prédit un rapport de branchement (nombre de désintégrations $B_s \rightarrow \mu^+ \mu^-$ par rapport au nombre total de B_s) suivant :

$$BR(B_s \rightarrow \mu^+ \mu^-) = (3, 2 \pm 0, 2) 10^{-9} \quad (1)$$

La mesure de ce rapport de branchement permettra de valider ou d'invalider le modèle standard [4].

FIGURE 2 – Correlation entre l'angle des b et des \bar{b}

2 Les outils d'analyse

2.1 Environnement de travail : Linux [5]

Le laboratoire du cppm m'a permis de choisir mon environnement de travail pour effectuer mon stage. Etant déjà familiariser avec le shell Unix, les commandes du shell Windows étant différentes, j'ai de ce fait choisi le système d'exploitation Linux. De plus le compilateur gcc est déjà incluse dans le système d'exploitation et permet de d'implémenter a la fois du C et du C++.

Linux est un système d'exploitation libre et gratuit promu par la communauté Linux Users Group.

Linux confère à son utilisateur :

- la liberté d'exécuter le programme, pour tous les usages,
- la liberté d'étudier le fonctionnement du programme et de l'adapter à ses besoins,

FIGURE 3 – Tux : La mascotte de Linux

- la liberté de redistribuer des copies du programme (ce qui implique la possibilité aussi bien de donner que de vendre des copies),
- la liberté d'améliorer le programme et de distribuer ces améliorations au public, pour en faire profiter toute la communauté.

2.2 Langage de programmation : C++

Le langage C++, amélioration du C, est un langage de programmation orienté objet (POO) très utilisé notamment pour les jeux vidéos.

Il est un des langages les plus utilisés en physique avec python. Ayant déjà fait du C, mon travail est donc naturellement et essentiellement codé en C++ et nécessitait l'utilisation de la programmation objet. La première partie de mon travail a d'abord consisté à me familiariser avec le langage objet pour l'utilisation de ROOT.

2.3 ROOT : Plateforme multilangage [8]

ROOT est un environnement d'analyse de données basé sur C++ (compilé ou interprété). Développé par le CERN, il vise en premier lieu le milieu des physiciens des hautes énergies, mais est néanmoins parfaitement adaptable à d'autres domaines, en particulier s'il y a de gros volumes de données à traiter.

Il s'agit techniquement parlant d'un gros ensemble de bibliothèques et d'un interpréteur C/C++. L'ensemble peut fonctionner soit en mode interprété soit en mode compilé. ROOT inclut en plus de ce qui est nécessaire à l'analyse de données une boîte à outils graphique intégrant le mécanisme slot/signal à la Qt.

3 But du stage

3.1 Le travail de l'équipe LHCb au sein du CPPM

L'équipe LHCb du CPPM comprend au total 6 chercheurs, 4 doctorants et 1 post-doctorant. On peut distinguer quatre sujets principaux d'études au sein de cette équipe :

- Une partie s'occupe d'étudier la violation de la symétrie CP grâce à l'étude de la désintégration $B_s \rightarrow J/\Psi\Phi$. La mesure phare de cette étude est la mesure de l'angle β_s de la matrice CKM. Trois personnes travaillent sur ce sujet : Maurice Emilie, Leroy Olivier et Khanji Basem.
- Un autre sujet concerne le Trigger à muons. Une partie de la carte du Trigger a été confectionnée au CPPM (Level 0) et une partie de l'équipe LHCb continue à l'étudier, à savoir Renaud Le Gac (Chef de groupe LHCb), Julien Cogan, Cachemiche Jean-Pierre et Duval Pierre-Yves.

- Une partie computing est gérée par Tsaregorodtsev Andrei et Sapunov Matvey. Ces deux personnes s'occupent de gérer le système de grille de calcul DIRAC (avec d'autres bien sûr).
- Enfin la dernière partie concerne les désintégrations rares. Principalement deux désintégrations sont étudiées au CPPM : les désintégrations $B_s \rightarrow \mu^- \mu^+$ et $B_b \rightarrow \mu^- \mu^+$ étudiées par Giampiero Mancinelli (tuteur de stage), Justine Serrano, Matthieu Perrin-Terrin et Androver Pacheco Cosme.

3.2 But du stage

Parmi les désintégrations, suite à la collision proton-proton, on trouve entre autre une part de quarks b qui se hadronisent avec d'autres quarks dont les quarks s pour former les mésons B_s .

Une étude des paramètres d'autre b survenant lors de la collision proton proton devient un objet d'étude pouvant affiner les résultats obtenus ou en cours de traitement du canal $B_s \rightarrow \mu^- \mu^+$. En effet, un parametre discriminant pour l'étude de la désintégration rare $B_s \rightarrow \mu^- \mu^+$ serait la direction des \bar{b} reconstruit grâce à la somme des impulsions des particules filles produites par le b.

Le but du stage est alors de déterminer le meilleur set de traces pour définir cette direction grâce a des coupures ou des variables cinétiques.

4 Analyse numérique

4.1 Simulation Monte Carlo

La simulation Monte Carlo (MC) est intensément utilisée en Physique des Hautes Énergies par exemple pour tester les algorithmes d'analyse de données [7].

Le Monte Carlo est composé tout d'abord d'une partie générateur d'évènements physiques (par exemple des collisions proton-proton). Toutes les particules générées (plusieurs centaines) sont envoyées dans un programme de transport (GEANT4).

La géométrie du détecteur est très détaillée, avec une résolution meilleure que le millimètre (la taille du détecteur est de 20 m) et la composition chimique de chacun des éléments du détecteur est connue. De plus, le champ magnétique généré par l'aimant de LHCb a été mesuré sur une grille tridimensionnelle. Ces informations relatives à la distribution de matière et au champ magnétique sont utilisées pour simuler la trajectoire, la diffusion et les interactions des particules traversant le détecteur. La réponse du détecteur au passage de particules est également simulée, ainsi que toute la chaîne de lecture et de traitement du signal, jusqu'à la reconstruction des traces

et les différentes mesures d'énergie, d'impulsion et d'identité associées à ces dernières. Ainsi, le résultat de la simulation est le plus proche possible des informations qui seront recueillies lors d'une collision réelle attendue.

Les données qui sont mises à ma disposition pour traiter mon problème seront des données Monte Carlo c'est à dire des données simulées. Les algorithmes utilisés pour traiter les données Monte Carlo sont les meme que pour les données réelles à ci peu près que certaines variables peuvent être connues lors de la simulation. Par exemple, en faisant des simulations on connaît toutes les désintégrations dans le détail. On sait si un b est produit, s'il s'est désintégré en deux muons ou une autre particule, etc... Ces données sont **connues**. Nous les apellerons **vérité Monte Carlo**. Elles permettront pour cette étude de séparer le signal du bruit de fond.

4.2 Les variables

Les variables utilisées ici pour cette analyse Monte Carlo sont tantôt des variables mesurées sur les détecteurs avec les précisions simulées ou bien des variables dites de vérité Monte Carlo. Voici une liste des variables utilisées pour mon analyse :

- **MCID** : variable de vérité Monte Carlo. Le MCID est le numéro associé à la particule détectée. Par exemple 211 signifie que la trace détectée est un pion π^+ .
- **MCmothID** : variable de vérité Monte Carlo. Le MCmothID, comme le MCID, est le numéro associé à une particule. Il correspond au MCID de la "mère" d'une trace détectée. Si la trace est un u^- et qu'il provient d'un B_s alors le MCID 13 (muon u^-) sera et le MCmothID sera 531 (B_s).
- **R** = $\sqrt{\Delta\eta^2 + \Delta\phi^2}$: variable mesurée par les détecteurs. $\Delta\eta$ est la différence de pseudorapacité entre un candidat \bar{B}_s et la trace considérée et $\Delta\phi$ est la différence d'angle entre un candidat \bar{B}_s et la trace considérée.
- **angle** : variable mesurée par les détecteurs. Différence d'angle entre un candidat \bar{B}_s et la trace considérée.
- **clone** : variable mesurée par les détecteurs. Valeur correspondant à la duplicité de la trace.
- **dile** : variable de vérité Monte Carlo. Identifie la trace comme ressemblant plus à un électron qu'à un pion du point de vue du detecteur.
- **dilk** : variable de vérité Monte Carlo. Identifie la trace comme ressemblant plus à un kaon qu'à un pion du point de vue du detecteur.
- **dllmu** : variable de vérité Monte Carlo. Identifie la trace comme ressemblant plus à un muon qu'à un pion du point de vue du detecteur.
- **doca** : variable mesurée par les détecteurs. Distance 3D entre la trace et un candidat \bar{B}_s .

- **event** : Nombre associé à l'événement.
- **fromB** : variable de vérité Monte Carlo. Variable la plus importante dans notre analyse. Le signal correspond au cas ou $\text{fromB} = 1$ c'est à dire toutes les particules ayant pour ancêtre un hadron contenant un quark b et le bruit de fond correspond au cas ou $\text{fromB} = 0$. Autrement dit, couper sur cette variable (0 ou 1) nous permet de séparer le signal et le bruit de fond.
- **fromD** : variable de vérité Monte Carlo. Même chose pour les traces provenant d'un hadron contenant une particule d.
- **ghost** : variable mesurée par les détecteurs. Probabilité qu'une trace soit une trace fantôme.
- **imp** : variable mesurée par les détecteurs. Paramètre d'impact. Distance entre la trace et le vertex primaire.
- **impSV** : variable mesurée par les détecteurs. Paramètre d'impact. Distance entre la trace et le vertex secondaire.
- **imp_b** : variable mesurée par les détecteurs. Paramètre d'impact. Plus petite distance entre la trace et des vertex primaires.
- **ips** : variable mesurée par les détecteurs. Paramètre d'impact imp divisé par l'erreur sur la mesure.
- **ipsSV** : variable mesurée par les détecteurs. Paramètre d'impact ipsSV divisé par l'erreur sur la mesure.
- **ips_b** : variable mesurée par les détecteurs. Paramètre d'impact imp_b divisé par l'erreur sur la mesure.
- **pt** : variable mesurée par les détecteurs. Impulsion transverse de la particule.
- **ptot** : variable mesurée par les détecteurs. Impulsion totale de la particule.
- **run** : nombre associé au Run.
- **trchi2** : variable mesurée par les détecteurs. Fonction χ^2 de la trace (somme des différences entre les points d'impact de la particule et la trace reconstituée).

4.3 Choix des variables de coupure

Les variables intéressantes pour trouver de bonnes coupures sont celles dont on peut séparer convenablement le signal et le bruit de fond. Par signal on entend toutes les particules filles détectées qui proviennent d'un \bar{b} et par bruit de fond toutes les autres particules qui ne viennent pas d'un \bar{b} . Pour séparer signal et bruit de fond, on applique la coupure suivante :

- fromB==1 pour le signal (en rouge)
- fromB==0 pour le bruit de fond (en noir)

et on obtient les graphes suivant sur toutes les variables :

Suite à cette séparation du signal et du bruit de fond, j'ai décidé dans un premier temps de garder les variables suivantes pour mon étude de coupure :

- imp en gardant : >0.2 mm
- imp_b en gardant : >0.25 mm
- ips en gardant : >3 (sans unité)
- ips_b en gardant : >3 (sans unité)
- trchi2 en gardant : <2 (sans unité)

- doca en gardant : <0.6 mm
- pt en gardant : >800 Mev
- ptot en gardant : >8000 Mev
- R en gardant : <2.5 mrad
- impSV en gardant : <1 mm

Pour savoir si on effectue une bonne coupure, on définit un facteur d'efficacité, un facteur de rejection de bruit de fond ainsi que deux rapports :

$$\epsilon_i = \frac{Sigcut_i}{Sig_{tot}}.100\% \quad (2)$$

$$BGR = \frac{BGcut_i}{BG_{tot}}.100\% \quad (3)$$

Dans un programme, on introduit les coupures citées plus haut et on sort les facteurs d'efficacité ϵ , de rejection de bruit de fond et les rapports $\frac{S}{B}$ et $\frac{S}{\sqrt{S+B}}$. Le rapport $\frac{S}{\sqrt{S+B}}$ nous permet de savoir quel proportion de signal ressort par rapport à l'erreur sur le nombre d'évènements. Plus grand est le rapport plus on verra le signal se démarquer sur une courbe. Apres exécution du programme, on obtient la table suivante :

Variables	cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	>3	74.19	23.38	3.17	7.5
trchi2	<2	90.3	70.96	1.27	7.1
impSV	<1	77.2	50.99	1.5	6.8
ips	>3	80.39	58.8	1.36	6.8
doca	<0.6	83.9	68.48	1.2	6.79
imp	>0.2	73	65.6	1.1	6.2
imp _b	>0.25	56.78	27.88	2	6.17
ptot	>8000	62.2	42	1.48	6.09
pt	>800	49.1	19.5	2.5	5.9
R	<2.5	53.9	53.9	1.1	5.6
Total		7.7 %	0.1 %	72.77	2.76

FIGURE 4 – Premières coupures sur les variables classées par rapport $\frac{S}{\sqrt{S+B}}$ décroissant

La sélection de coupure a un bon effet. Au total, elle enlève pratiquement tout le bruit de fond soit 0.1 % restant et garde 7.7 % du signal soit plus de 70 fois plus de signal que de bruit de fond. Le rapport $\frac{S}{\sqrt{S+B}}$ doit tendre vers

$\sqrt{S} = 310$ initialement sans coupure à $\frac{S}{\sqrt{S+B}} = 77$. Ce rapport tombe à 2,7 après coupure. La suite de mon étude consiste alors à optimiser ce rapport.

4.4 Etude de correlation des variables

Un problème se pose en regardant la signification de certaines variables comme *imp*, *ips*, *ips_b* et *imp_b*. *Imp* est le paramètre d'impact, distance entre la trace et le vertex primaire et *ips* le paramètre d'impact divisé par l'erreur sur la mesure. Si on coupe toutes les traces inférieures à 0.2 mm sur le paramètre d'impact, il sera alors inutile de couper sur le paramètre *ips*. Ces deux paramètres sont corrélés et une étude de correlation s'impose. Il en de meme de *imp_b* et *ips_b*. En fait les quatre paramètres *ips*, *imp*, *ips_b* et *imp_b* sont tous les quatres corrélés.

FIGURE 5 – Plot de corrélation de coupure : paramètre imp coupé sur ips
(>3) imp_b (>0.25) et ips_b (>3)

FIGURE 6 – Plot de corrélation de coupure : paramètre imp_b coupé sur ips
(>3) imp (>0.2) et ips_b (>3)

FIGURE 7 – Plot de corrélation de coupure : paramètre ips coupé sur imp
(>0.2) imp_b (>0.25) et ips_b (>3)

FIGURE 8 – Plot de corrélation de coupure : paramètre ips_b coupé sur ips
(>3) imp_b (>0.25) et imp (>0.2)

En coupant sur trois de ces quatres parametres et en observant le quatrième, on voit que dans certains cas la coupure que l'on voulait effectuer n'a plus aucun sens. Celui que l'on remarque le plus est sans doute le plot du ips pour le lequel la quasi totalité des entrées ont disparu juste en dessous de la valeur de coupure que l'on voulait effectuer. De meme sur ces différents histogrammes, on peut voir que les quatre variables sont très corrélées, la concentration la plus forte de données étant sur une droite.

FIGURE 9 – Plot de correlation des paramètres imp et $impb$

FIGURE 10 – Plot de correlation des paramètres imp_b et ips_b

FIGURE 11 – Plot de correlation des paramètres *impb* et *ips*

FIGURE 12 – Plot de correlation des paramètres imp et ips_b

FIGURE 13 – Plot de correlation des paramètres imp et ips

FIGURE 14 – Plot de correlation des paramètre ips et ips_b

Si on optimise une valeur de coupure en prenant le meilleur rapport $\frac{S}{\sqrt{S+B}}$ pour une variable, l'optimisation d'une variable corrélée avec la première n'a aucun sens. Pour deux variables a et b, il devient alors nécessaire d'optimiser non pas le rapport $\frac{S}{\sqrt{S+B}}$ de a puis de b (ou bien le contraire) mais celui du couple (a,b). De même pour n variables v_i corrélées, il est nécessaire d'optimiser le n-uplet $(v_1, v_2, \dots, v_{n-1}, v_n)$.

4.5 Algorithme Random Grid Search (GRS)

On cherche à optimiser le n-uplet de variables fournies dans notre étude. On peut utiliser une multitude d'algorithmes pour exécuter ces coupures.

4.5.1 Un premier algorithme : Optimiser les coupures une à une

La première solution à laquelle on pourrait penser serait d'optimiser une variable puis une suivante etc jusqu'à ce qu'à la dernière, noter le rapport $\frac{S}{\sqrt{S+B}}$ et refaire ainsi de suite pour chaque possibilité. Sachant que optimiser la variable v_i puis la variable v_j est différent d'optimiser la variable v_j puis la variable v_i .

FIGURE 15 – Ci-contre, un arbre de possibilité sur 4 variables corrélées dont les coupures sont à optimiser. Il y a donc $4! = 24$ possibilités d'optimiser une coupure sur 4 variables corrélées

Nous disposons de 10 variables au total sur lesquels on veut optimiser une coupure sur chacune d'elles, ce qui nous donne $10! = 3628800$ façon d'optimiser le rapport $\frac{S}{\sqrt{S+B}}$. Notre lot de données est de 1554091 traces.

De plus avec cette façon de procéder on doit définir un pas sur lequel on sonde la coupure. Suivant la taille de ce pas, le nombre de calcul que le processeur doit effectuer peut varier de façon non négligeable. Ainsi le temps de calcul même avec un pas raisonnable, pourrait atteindre les 10 ou 15 ans, d'où la nécessité de trouver un algorithme plus puissant et qui traite le problème différemment.

4.5.2 Algorithme simultané : Grid Random Search

L'algorithme GRS permet d'effectuer des coupures en prenant une trace en entrée et en comparant les autres à celle-ci. On effectue la coupure directement à partir des valeurs des variables de la trace, c'est à dire qu'en prenant n variables, on effectue n coupures **simultanément** sur les variables de cette donnée.

En résumé, avec cette méthode on traite non plus n variables sur lesquels on effectue $n!$ fois l'algorithme qui traite m traces, mais m traces que l'on

traite m fois avec m le nombre d'entrées présentes dans le Tree de données. Le GRS permet en outre faire une analyse sur plusieurs variables corrélées ou non corrélées puisqu'il optimise un n -uplet de variables. Il convient donc tout à fait à notre analyse.

En traitant les 10 variables définies précédemment, on optimise le rapport $\frac{S}{\sqrt{S+B}}$ du n -uplet ($imp, imp_b, ips, ips_b, trchi2, doca, pt, ptot, R, ipsSV$). Pour des raisons de temps d'exécution un intervalle de recherche est défini avant l'exécution ce qui permet de réduire le nombre de trace (donc de diminuer le temps de traitement) en empêchant de calculer des rapports absurdes. De plus, pour vérifier la validité de l'algorithme, j'ai traité 7 cas différents pour chacun desquels le programme itère non pas jusqu'au nombre total de traces $ntot$ mais jusqu'à des valeurs plus petites :

- $ntot/1000$
- $ntot/500$
- $ntot/100$
- $ntot/50$
- $ntot/10$
- $ntot/5$

et $ntot$ pour finir qui a pris deux jours d'exécution.

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
trchi2	3.8	99.06	96.56	1	7.16
ips _b	0.42	88.27	86.6	1	7.1
impSV	1.78	99.97	99.86	1	7.1
ips	0.4	87.5	64	1.36	7.1
pt	66.3	69	57.3	1.2	7.07
ptot	3197.7	98.5	90.6	1.08	6.9
R	4.2	43.96	22.26	1.97	6.67
doca	0.3	98.1	91.77	1.06	6.1
imp	0.35	56.18	58.39	0.96	5.4
imp _b	0.3	90.79	81.75	1.1	5.2
Total		13.37 %	1.8 %	7.3	3.4

FIGURE 16 – Tracks number : ntot/1000

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	4	66.7	18.2	3.66	7.2
trchi2	4.49	99.46	96.09	1	7.11
R	6.08	99.48	99.56	0.99	7.05
impSV	1.06	78.55	52.41	1.49	6.86
ptot	4370.1	82.65	68.58	1.2	6.7
pt	336.9	78.88	58.9	1.3	6.7
imp _b	0.2	64.1	32.1	1.99	6.5
ips	4.08	73.85	54.28	1.36	6.5
imp	0.2	72.58	65.4	1.1	6.17
doca	0.25	63.3	53.7	1.17	5.8
Total		17.88 %	0.77 %	23.2	4.1

FIGURE 17 – Tracks number : ntot/500

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	4	66.7	18.2	3.66	7.2
trchi2	4.49	99.46	96.09	1	7.11
R	6.08	99.48	99.56	0.99	7.05
impSV	1.06	78.55	52.41	1.49	6.86
ptot	4370.1	82.65	68.58	1.2	6.7
pt	336.9	78.88	58.9	1.3	6.7
imp _b	0.2	64.1	32.1	1.99	6.5
ips	4.08	73.85	54.28	1.36	6.5
imp	0.2	72.58	65.4	1.1	6.17
doca	0.25	63.3	53.7	1.17	5.8
Total		17.88 %	0.77 %	23.2	4.1

FIGURE 18 – Tracks number : ntot/100

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	4	66.7	18.2	3.66	7.2
trchi2	4.49	99.46	96.09	1	7.11
R	6.08	99.48	99.56	0.99	7.05
impSV	1.06	78.55	52.41	1.49	6.86
ptot	4370.1	82.65	68.58	1.2	6.7
pt	336.9	78.88	58.9	1.3	6.7
imp _b	0.2	64.1	32.1	1.99	6.5
ips	4.08	73.85	54.28	1.36	6.5
imp	0.2	72.58	65.4	1.1	6.17
doca	0.25	63.3	53.7	1.17	5.8
Total		17.88 %	0.77 %	23.2	4.1

FIGURE 19 – Tracks number : ntot/50

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	0.88	93.9	72.6	1.3	7.27
trchi2	3.2	97.2	85.6	1.1	7.19
imp _b	0	96.7	88.9	1.1	7.09
ips	0.91	95.9	86.6	1.1	7.09
impSV	1.68	86.8	62.8	1.4	7.09
imp	0.02	97.9	94.3	1	7.05
R	5.1	95.2	95.34	1	6.89
ptot	4540.6	81.5	66.9	1.2	6.69
doca	0.48	79.9	64.9	1.2	6.6
pt	812.1	48.6	19	2.6	5.9
Total		33 %	5.3 %	6.2	5.3

FIGURE 20 – Tracks number : ntot/10

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	0.88	93.9	72.6	1.3	7.27
trchi2	3.2	97.2	85.6	1.1	7.19
imp _b	0	96.7	88.9	1.1	7.09
ips	0.91	95.9	86.6	1.1	7.09
impSV	1.68	86.8	62.8	1.4	7.09
imp	0.02	97.9	94.3	1	7.05
R	5.1	95.2	95.34	1	6.89
ptot	4540.6	81.5	66.9	1.2	6.69
doca	0.48	79.9	64.9	1.2	6.6
pt	812.1	48.6	19	2.6	5.9
Total		33 %	5.3 %	6.2	5.3

FIGURE 21 – Tracks number : ntot/5

Variables	GRS cut	Efficiency S	Background rejection BGR	Ratio $\frac{S}{B}$	Ratio $\frac{S}{\sqrt{S+B}}$
ips _b	0.88	93.9	72.6	1.3	7.27
trchi2	3.2	97.2	85.6	1.1	7.19
imp _b	0	96.7	88.9	1.1	7.09
ips	0.91	95.9	86.6	1.1	7.09
impSV	1.68	86.8	62.8	1.4	7.09
imp	0.02	97.9	94.3	1	7.05
R	5.1	95.2	95.34	1	6.89
ptot	4540.6	81.5	66.9	1.2	6.69
doca	0.48	79.9	64.9	1.2	6.6
pt	812.1	48.6	19	2.6	5.9
Total		33 %	5.3 %	6.2	5.3

FIGURE 22 – Tracks number : ntot

Au fur et a mesure que l'algorithme sonde les données dans les 7 différents cas, il trouve un meilleur maximum de rapport $\frac{S}{\sqrt{S+B}}$. Ces maxima sont parfois les même (par conséquent les tables aussi) et ne dépassent jamais le rapport du plus mauvais set de coupure.

Les coupures qui optimise le mieux ce rapport sont données dans la figure 22. Le pourcentage de coupure du pt sur ce set de coupure laisse à penser que la coupure est mauvaise mais celle ci enlève d'autre part plus de 80% du bruit de fond. On conserve donc cette valeur.

Après concertation, nous avons décidé de garder les tables des figures 17 et 22 dans la recherche de l'angle caractéristique. Dans la table 17, les valeurs des coupures paraissent plus en accord avec les valeurs de coupures "à vue d'oeil" même si le rapport $\frac{S}{\sqrt{S+B}}$ n'est pas le plus optimal. Ces coupures sont ensuite insérées dans un programme de traitement du signal $B_s^0 \rightarrow \mu^- \mu^+$. Ce programme ressort entre autre pour mon étude, un fichier .root contenant un Tree avec différentes variables dont :

- mu_otherB[i]_ang : angle entre l'impulsion du B_s candidat et l'impulsion du thrust (impulsion du \bar{B}_s reconstitué grace au impulsion de ces particules filles). Quand l'impulsion du thrust ne peut être reconstruit sa valeur est mise à 1 [1].
- mu_otherB[i]_boo_ang : angle entre la direction d'un μ^+ d'un B_s candidat et le moment du thrust tout les deux boosté dans le meme référentiel, celui du thrust. Quand le moment du thrust ne peut être reconstruit sa valeur est mise à $\frac{\pi}{2}$ [1].

La lettre i dans otherB[i] peut prendre la valeur 1 ou 2 suivant si les coupures effectuées suivent la table 17 (pour 1) ou la table 22 (pour 2).

Après ces coupures sur les différentes variables de mon étude, on peut encore séparer signal et bruit de fond sur ces nouvelles variables. Rappel : le signal est un \bar{B}_s et le bruit de fond représente tout les autres particules.

FIGURE 23 – mu_otherB1_ang : en rouge signal, en noir bruit de fond

FIGURE 24 – mu_otherB1_boo_ang : en rouge signal, en noir bruit de fond

FIGURE 25 – mu_otherB1_pt : en rouge signal, en noir bruit de fond

Ces trois plots montrent un pic prononcé au niveau des valeurs par défaut. Cela signifie que la proportion de valeur par défaut par rapport au reste des événements est bien plus grande.

L'algorithme mis en place pour le traitement des autres B ne reconstruit pas les autres B et remplit les valeurs par défaut. Les coupures fournies en amont sont trop serrées et coupent trop de traces des particules filles du signal pour pouvoir effectuer une analyse supplémentaire en aval.

5 Conclusion

5.1 Résumé de l'étude

L'étude consistait à trouver un set de coupure sur un lot de données Monte Carlo des B opposés dans le cadre de l'analyse du signal $B_s^0 \rightarrow \mu^- \mu^+$.

Après une séparation du signal et du bruit de fond de chacune des variables fournies pour mon analyse, certaines ont été sélectionnées pour des tests de coupure qui paraîtraient être les plus efficaces. Une coupure à vue d'oeil a été faite sur neuf variables afin de voir si les résultats pouvaient être exploitables. Il s'est avéré que la première coupure était satisfaisante dans le sens où l'efficacité de la coupure et le taux de rejection du bruit de fond étaient élevés et satisfaisants.

Mais d'autre part, certaines variables s'avèrent être corrélées. Une étude de corrélation a montré que quatre variables parmi les variables choisies étaient corrélées et que si l'on appliquait des coupures indépendamment, elles n'avaient aucun sens.

La suite de mon travail consistait à optimiser cette coupure. Pour critère d'optimisation, j'ai choisi le rapport $\frac{S}{\sqrt{S+B}}$, le signal par rapport à l'erreur sur le nombre total de données. Pour palier au problème de la corrélation et pour optimiser ce rapport, la solution à adopter était d'appliquer l'algorithme du GRS (Random Grid Search) qui consiste à utiliser les valeurs des variables d'un événement comme valeur de coupure simultanée.

Au résultat, deux sets de coupure ont été retenus : le premier sur une analyse de données partielles, choisit pour un bon rapport final $\frac{S}{\sqrt{S+B}}$ et les valeurs des coupures étant plus en accord avec les graphiques (figure 17). Le deuxième a été choisi pour son rapport $\frac{S}{\sqrt{S+B}}$ optimal (figure 22).

Ces deux sets de coupures ont ensuite été insérés dans un programme¹ de traitement du signal $B^s \rightarrow \mu^- \mu^+$. Dans le cadre de mon étude, ce programme me fournissait en sortie un Tree de données contenant d'autres variables (deux variables angulaires et un variable d'impulsion transverse).

Malheureusement, le nombre de variables de coupure en amont a été trop grand et le signal trop coupé par les précédentes coupures.

5.2 Conclusion d'étude

Les données recueillies suite aux coupures ne sont pas exploitables. Les valeurs par défaut remplies lorsque le moment du thrust n'est pas reconstruit est bien trop grand par rapport au reste des données. Pour obtenir un pic moins grand et donc plus de signal à traiter, il faudrait couper sur moins de variables et surtout prendre celle qui ont le plus d'efficacité (par exemple trois variables ayant une efficacité $> 90\%$ voire $> 95\%$).

1. Programme réalisé par G. Mancinelli

Cette nouvelle étude d'optimisation ainsi que celle du lot de données en aval pourrait constituer un nouveau sujet d'étude.

5.3 Conclusion personnelle

Ce stage m'a permis de découvrir le monde de la recherche au niveau de la physique des particules ainsi qu'un nouveau laboratoire. L'aspect de la recherche tel que je l'ai découvert au CPPM m'a conforté dans mon idée de continuer dans le monde de la recherche.

J'ai découvert que les statistiques, la programmation et une bonne connaissance du système UNIX étaient les principaux outils du chercheur en physique des particules. Une solide base en théorie (QCD, QED, modèle standard, etc...) est grandement nécessaire pour comprendre l'analyse des données et plus généralement la physique des particules.

Concrètement, j'ai appris un nouveau framework, ROOT, ainsi que son langage de programmation principal, le C++. ROOT étant un environnement de développement adapté à la programmation orienté objet, j'ai aussi découvert un nouveau concept des langages de programmation jusqu'alors inconnu pour moi : l'aspect objet.

Références

- [1] C. Androver, G. Mancinelli, M. Perrin-Terrin, and J. Serrano. Measurement of the branching ratios of $(b_s \rightarrow \mu^+ \mu^-)$ and of $(b_d \rightarrow \mu^+ \mu^-)$ decays using a maximum likelihood unbinned fit. 2011.
- [2] Miriam Calvo. “lhcb status, on behalf of the lhcb collaboration”. 2010.
- [3] Cern. “the large hadron collider beauty experiment, for precise measurements of cp violation and rare decays”. <http://lhcb.web.cern.ch/lhcb/>, Consulté le 6 juin 2011.
- [4] LHCb Collaboration. Roadmap for selected measurements of lhcb. 2010.
- [5] Linux Community. “learn about linux : Free linux support”. <http://www.linux.com/learn>, Consulté le 3 juin 2011.
- [6] CPPM. “experience lhcb”. <http://marwww.in2p3.fr/spip.php?article577>, Consulté le 1 juin 2011.
- [7] Ecole Polytechnique Fédérale de Lausanne. “cluster – lphe1 : de la simulation à l’analyse de données du lhc”. <http://ditwww.epfl.ch/SIC/SA/SPIP/Publications/spip.php?article1838>, Consulté le 12 juin 2011.
- [8] The ROOT Team. “experience lhcb”. <http://root.cern.ch/drupal/>, 1995-2011.