

# Chapitre 3

## Traitement et réduction des données

3.1 Introduction

3.2 Les différents ensembles de données et de simulation

3.3 Traitement des données

3.4 La procédure de réduction des données

## 3.1 Introduction

L'ensemble des données brutes récoltées par le détecteur AMANDA ne constitue pas une information directement exploitable d'un point de vue scientifique. Celle-ci est en effet entachée d'un bruit de fond important ainsi que de trajectoires mal reconstruites (voir la section 2.4) qu'il est nécessaire de réduire au maximum, sans pour autant perdre les événements du signal que l'on désire mettre en évidence. Le but de ce chapitre est de présenter les différentes étapes du processus de réduction des données, en particulier la méthode originale de sélection des données qui a été mise au point et adoptée dans notre étude. Dans un premier temps, nous détaillons les différents ensembles de données et de simulations qui ont été analysés dans ce travail et qui ont servi à la réduction des données.

## 3.2 Les différents ensembles de données et de simulation

### 3.2.1 Les données expérimentales

Les données expérimentales analysées dans ce travail ont été enregistrées par le détecteur AMANDA durant l'année 1997, entre le 9 mars et le 15 novembre. Les différentes séries d'enregistrements<sup>1</sup> effectuées durant cette période (224,6 jours) comptabilisent  $1,45 \cdot 10^9$  événements. Pour être considérés comme tels et être effectivement enregistrés, chacun de ces événements devait être « vu » par au moins un des 5 détecteurs suivants : AMANDA-A, AMANDA-B10, SPASE1, le prédécesseur du détecteur SPASE2 actuel (Beaman *et al.*, 1993), SPASE2 (Dickinson *et al.*, 2000) ou GASP, un petit télescope Čerenkov aujourd'hui démantelé (Barbagli *et al.*, 1993). Dans ce travail, seuls les  $1,29 \cdot 10^9$  événements détectés par AMANDA-B10 ont été considérés.

### 3.2.2 Les simulations Monte Carlo

Une étape nécessaire dans le processus de réduction des données est la confrontation entre les événements réellement enregistrés par le détecteur et les événements auxquels on s'attend, sur base de la connaissance des différents processus physiques que subissent les neutrinos, les rayons cosmiques et les muons au niveau de l'atmosphère terrestre, dans la roche et dans la glace du pôle. Ces différents processus physiques ont été simulés par la technique dite de « Monte Carlo », entièrement dédiée à la description de phénomènes aléatoires.

Dans ce travail, deux ensembles d'événements simulés ont été traités :

1. Simulation Monte Carlo du bruit de fond (*BGMC*)

Cette simulation Monte Carlo décrit la réponse du détecteur lorsqu'il est « touché » par des muons atmosphériques (venant du haut et se dirigeant vers le bas, du point de vue d'un observateur situé au pôle sud).

---

<sup>1</sup> Appelées « runs » dans le jargon d'AMANDA. Un run correspond environ à une journée d'enregistrement.

## 2. Simulation Monte Carlo du « signal » (*signal MC*)

Cette simulation Monte Carlo décrit la réponse du détecteur lorsqu'il est « touché » par des muons ascendants induits par des neutrinos venant du bas. Ces derniers proviennent de l'interaction des rayons cosmiques avec l'atmosphère terrestre.

La simulation du bruit de fond et du « signal » est divisée en 4 étapes : dans un premier temps, la génération des événements, puis la propagation des muons créés, ensuite la propagation des photons Čerenkov produits par les muons et enfin la réponse du détecteur proprement dite.

### a) Génération des événements

La génération des événements s'effectue à partir de la connaissance d'un flux de particules primaires acquise lors de mesures effectuées par d'autres expériences. Par exemple, lorsqu'on simule le bruit de fond dû aux muons atmosphériques, les particules primaires considérées sont les rayons cosmiques. Le programme générateur simule alors les interactions de ces particules primaires avec la Terre ou l'atmosphère (dans le cas de la génération de muons atmosphériques) et en déduit les traces des muons ainsi produits. Lors de la simulation d'AMANDA-B10, différents programmes générateurs ont été utilisés :

#### a) pour les événements de « signal » (*signal MC*)

- *nu2mu* (Dalberg, 1998, 1999) : ce programme développé au sein de la collaboration AMANDA utilise les neutrinos atmosphériques comme particules primaires. Il est basé sur la prédiction théorique des flux de neutrinos atmosphériques établie par Lipari (Lipari, 1993) à partir des flux de rayons cosmiques. Les sections efficaces d'interaction des neutrinos sont tirées de l'article de Gandhi *et al.* (Gandhi *et al.*, 1996).
- *nusim* (Hill, 1996) : utilise également les prédictions théoriques de flux de neutrinos atmosphériques de Lipari. Pour simuler la propagation des neutrinos, ce programme tient compte du profil de densité de la Terre (Dziewonski et Anderson, 1981). Les sections efficaces d'interaction des neutrinos sont tirées de Hill (Hill, 1997).

#### b) pour les événements de background (*BGMC*)

- *CORSIKA* (Heck *et al.*, 1998) : c'est actuellement le programme standard de simulation des gerbes atmosphériques. Au sein d'AMANDA, il est utilisé pour générer le bruit de fond dû aux muons atmosphériques.
- *basiev* (Boziev *et al.*, 1989) : ce programme est analogue à CORSIKA, plus rapide mais moins précis, étant donné qu'il ne tient compte que des protons pour simuler les rayons cosmiques. Aucune déviation importante entre les résultats fournis par basiev et CORSIKA n'a cependant été observée (Biron, 2002).

## b) Propagation des muons

Après la génération d'événements vient l'étape de la simulation de la propagation des muons dans la matière. Celle-ci s'accompagne d'une perte d'énergie due essentiellement au phénomène d'ionisation continue, à laquelle s'ajoute, à haute énergie et de façon stochastique, la production d'électrons  $\delta$ , l'émission de Bremsstrahlung et la production de paires  $e^+ - e^-$  ou  $\mu^+ - \mu^-$ . En fait, la radiation Čerenkov émise par les muons eux-mêmes ne contribue que de manière secondaire à leur perte d'énergie globale. Deux programmes de propagation ont été utilisés, tous deux basés sur des tables calculées par Lohman *et al.* (Lohmann *et al.*, 1985), donnant la perte d'énergie des muons lorsqu'ils traversent la matière, leur énergie étant comprise dans la gamme 1-10000 GeV :

- *mudedx* (Kopp *et al.*, 1995)
- *propmu* (Lipari et Stanev, 1991)

La propagation des muons est traitée différemment selon que ceux-ci sont éloignés du détecteur ou, au contraire, se trouvent dans le voisinage de celui-ci. Lorsque le muon est loin du détecteur, le programme ne détermine que sa perte d'énergie. En effet, la lumière émise par les particules secondaires induites par le muon est absorbée bien avant d'atteindre le détecteur. En revanche, lorsque le muon évolue dans la glace située au voisinage ou à l'intérieur du détecteur, le programme de simulation stocke l'information concernant chaque processus de perte d'énergie, en considérant un « volume actif » défini par un cylindre de 800 m de long et 400 m de diamètre englobant le détecteur.

## c) Propagation des photons

Tous les processus physiques via lesquels les muons perdent de l'énergie mènent à la production de photons Čerenkov. La propagation de ces photons émis dans le volume actif défini ci-dessus vers les différents modules optiques est simulée par le programme PTD (Karle, 1999). Ce programme tient compte de la dépendance en longueur d'onde de la longueur d'absorption de la glace ainsi que de la diminution<sup>2</sup> de la longueur de diffusion au niveau de la glace située autour des différents câbles. Il produit des tables multidimensionnelles contenant la probabilité qu'un photomultiplicateur reçoive un signal provenant de la perte d'énergie d'un muon en fonction du processus physique de perte d'énergie en jeu, de la distance  $r$  du module optique par rapport à la trace du muon, de l'orientation relative  $\eta$  de la trajectoire du muon par rapport à l'axe du photomultiplicateur, de la quantité d'énergie perdue, ... Ces tables permettent de déterminer les paramètres  $\tau$  et  $\lambda$  de la fonction de Pandel (cf. section 2.5.3).

---

<sup>2</sup> Cette diminution est due à une accumulation de bulles d'air lors de la recongélation de la glace après forage à l'eau chaude.

#### d) Réponse du détecteur

AMASIM est le programme qui simule la réponse du détecteur et de tous ses composants, des tubes photomultiplicateurs au système d'acquisition (Hundertmark, 1999).

### 3.3 Traitement des données

Avant que les événements de données ne soient reconstruits et analysés, plusieurs étapes préliminaires de filtrage ont été effectuées.

#### 3.3.1 Sélection des modules optiques

Comme on l'a vu plus haut (section 3.2.1), la collecte des données se fait par séries d'enregistrements (appelées « runs »). Pour chacune d'elles, on procède à l'identification des modules optiques utilisables de manière fiable pour la reconstruction des trajectoires de muons. Un module optique est considéré comme utilisable s'il remplit les conditions suivantes :

1. il est opérationnel ;
2. il est bien calibré ;
3. les distributions des durées et des amplitudes des signaux qu'il produit sont stables ;
4. son taux de bruit de fond s'écarte de moins de 2 déviations standard (pour les câbles 1 à 4) ou 4 déviations standard (pour les câbles 5 à 10) par rapport à son taux de bruit de fond moyen établi pour toute l'année 1997.

Les séries d'enregistrement ont été écartées de l'analyse lorsque moins de 220 modules optiques (sur les 302 que comptabilise AMANDA-B10) satisfaisaient aux critères énoncés ci-dessus. Pour les séries d'enregistrement considérées comme étant valables, l'information provenant des modules optiques non fiables n'a pas été prise en compte lors de la reconstruction des trajectoires de muons.

#### 3.3.2 La procédure du « hit-cleaning »

Cette procédure a pour but de rejeter les photomultiplicateurs à bruit de fond excessif, le phénomène de « cross-talk »<sup>3</sup> ainsi que l'information provenant des photons très retardés, c'est-à-dire fortement diffusés dans la glace. Elle impose certaines conditions que doivent respecter les signaux électriques émis par les photomultiplicateurs pour être pris en compte dans l'analyse. Ces conditions ont trait notamment à la durée et à l'amplitude des impulsions électriques et imposent que chaque signal provenant d'un photomultiplicateur soit accompagné d'un autre signal, dans un photomultiplicateur voisin, situé à moins de 100 m et compris dans une fenêtre temporelle de  $\pm 500$  ns. La description détaillée de cette procédure sort du cadre de ce travail (le lecteur peut se référer à Klug, 1997 ou De Young, 2001).

---

<sup>3</sup> Le phénomène de « cross-talk » se produit lorsqu'un signal électrique provenant d'un module optique induit un courant électrique dans un autre module optique.

Remarquons enfin que, étant donné que la trajectoire d'un muon est définie par 5 paramètres (cf. section 2.5.3), seuls les événements pour lesquels au moins cinq modules optiques ont été touchés (après la procédure du « hit-cleaning ») ont été retenus dans la procédure d'analyse des données.

### 3.3.3 Reconstruction de la trajectoire des muons

Après avoir déterminé quels étaient les modules optiques fiables et effectué la procédure du « hit-cleaning », les trajectoires des muons ont été reconstruites par différentes méthodes. Dans ce travail, 6 méthodes de reconstruction des événements ont été appliquées successivement (cf. tableau 3.1). Ces 6 méthodes ont servi à reconstruire aussi bien les trajectoires des événements des données expérimentales que celles des muons atmosphériques venant du haut (*BGMC*, voir section 3.2.2) et des muons ascendants induits par les neutrinos atmosphériques (*signal MC*, voir section 3.2.2). Il faut insister sur le fait que les 3 échantillons (données expérimentales, *BGMC* et *signal MC*) ont été traités dans des conditions identiques (même sélection des modules optiques, même procédure de « hit-cleaning » et même procédure de reconstruction des trajectoires), en vue de pouvoir les comparer les uns aux autres (cf. section suivante).

<b>Numéro de la reconstruction</b>	<b>Type de reconstruction</b>
1	Ajustement linéaire (linefit)
2	Reconstruction par la méthode du maximum de vraisemblance
3	Estimation de l'énergie $E_\mu$ par reconstruction $P_{hit}-P_{nohit}$
4	Reconstruction de gerbes (méthode de première approximation)
5	Reconstruction de gerbes (méthode du maximum de vraisemblance)
6	Reconstruction itérative (6 itérations)

Tableau 3.1 : Les différentes méthodes de reconstruction des événements qui ont été appliquées dans cette analyse. Données expérimentales, *BGMC* et *signal MC* ont été traités de manière identique.

## 3.4 La procédure de réduction des données

### 3.4.1 Introduction

La méthode de reconstruction des trajectoires en elle-même ne fournit pas à elle seule une résolution angulaire et un facteur de rejet des trajectoires descendantes suffisants pour la recherche de sources ponctuelles de neutrinos. En effet (et cela se vérifie d'autant mieux que le détecteur est petit), pour certains événements, seuls quelques modules optiques ( $\sim 5$ ) sont touchés par des photons Čerenkov. En outre, ceux-ci sont parfois fortement diffusés. L'information temporelle recueillie, c'est-à-dire les temps d'arrivée des photons, n'est alors pas suffisante ou pas suffisamment précise pour que les algorithmes de reconstruction convergent vers la meilleure approximation de la trajectoire réelle.

### 3.4.2 Sélection des événements bien reconstruits

Dans le but de sélectionner les événements bien reconstruits, une série de critères de qualité ont été développés. Ces critères se basent sur différents paramètres physiques et géométriques caractérisant les événements reconstruits à partir des données enregistrées par le détecteur AMANDA. Parmi ceux-ci, on peut citer comme exemple<sup>4</sup> :

- la direction de provenance des muons. L'objectif étant d'étudier les muons ascendants (voir la section 2.4), seules les données expérimentales dont l'angle zénithal est supérieur à  $80^\circ$  sont retenues ;
- le nombre de modules optiques touchés par les photons Čerenkov ;
- le nombre de « coups directs », c'est-à-dire le nombre de photons qui ont touché un module optique à un temps résiduel proche de zéro (cf. section 2.5.3), signifiant par là que ces photons ont été très peu diffusés ;
- la valeur du paramètre de vraisemblance utilisé dans la reconstruction par la méthode du maximum de vraisemblance (cf. section 2.5.3) ;
- divers critères topologiques des traces, comme par exemple la distance entre le centre de gravité des modules optiques touchés et la trajectoire reconstruite, ou bien un paramètre caractérisant la distribution spatiale des modules optiques touchés à l'intérieur du détecteur.

A partir de ces diverses variables qui caractérisent les trajectoires reconstruites, les données expérimentales font l'objet d'un ensemble d'opérations de sélection qui visent à retenir un maximum d'événements en relation avec le signal recherché et à rejeter un maximum d'événements de background. Pour effectuer cette opération de réduction des données expérimentales, on choisit des variables (dites variables d'analyse) qui présentent les 2 particularités suivantes :

---

<sup>4</sup> Cette liste n'est pas du tout exhaustive.

1. leurs distributions pour les événements de *BGMC* sont semblables aux distributions correspondantes des données expérimentales, ce qui prouve que la simulation décrit correctement ces variables et est donc fiable. En effet, avant la réduction des données expérimentales, celles-ci sont majoritairement constituées d'événements de background, étant donné que pour 1 muon ascendant, il y a de l'ordre de  $10^6$  événements descendants (cf. section 2.4) ;
2. leurs distributions pour le signal simulé par Monte Carlo sont sensiblement différentes des distributions correspondantes pour le bruit de fond simulé (*BGMC*), ce qui assure que ces variables permettent de différencier efficacement le signal du bruit de fond.

Pour chaque sélection d'événements effectuée sur une variable d'analyse, on définit un facteur de rejet  $R$  donné par :

$$R = \frac{\mathcal{E}_{signal}}{\mathcal{E}_{background}}, \quad (3.1)$$

avec

$$\mathcal{E}_{background} = \frac{N_{bg}^{pass}}{N_{bg}^0} \quad (3.2)$$

l'efficacité de retenue du background et

$$\mathcal{E}_{signal} = \frac{N_{signal}^{pass}}{N_{signal}^0} \quad (3.3)$$

l'efficacité de retenue du signal<sup>5</sup>.

Dans les deux dernières équations,  $N_{bg}^0$  et  $N_{signal}^0$  sont les nombres d'événements de background et de signal avant d'avoir effectué la sélection et  $N_{bg}^{pass}$  et  $N_{signal}^{pass}$  sont les nombres d'événements qui subsistent après sélection. Une sélection caractérisée par un grand facteur de rejet  $R$  signifie donc qu'un maximum de signal est retenu tout en rejetant un maximum de background. Nous discutons, dans la suite, la méthode utilisée pour optimiser la sélection effectuée sur chacune des variables d'analyse.

---

<sup>5</sup> Notons qu'à chaque sélection effectuée sur une variable d'analyse correspond également une efficacité de sélection des données expérimentales  $\mathcal{E}_{données} = N_{données}^{pass} / N_{données}^0$ .



### 3.4.3 La « méthode de la différence maximale »

L'objectif principal de la réduction des données est de rejeter un maximum de background tout en gardant un maximum d'événements de signal. Dans ces conditions, il semble naturel d'effectuer la sélection des données en utilisant les variables qui présentent les facteurs de rejets  $R$  les plus élevés possibles. Dans le cadre de la recherche de sources ponctuelles de neutrinos d'origine astrophysique, il est également capital de veiller à maintenir une efficacité de retenue du signal  $\epsilon_{\text{signal}}$  aussi élevée que possible.

En pratique, différents cas de figure se présentent suivant les variables d'analyse que l'on étudie. Nous présentons dans la suite deux exemples simulés dans le but de détailler la méthode d'optimisation de la sélection des événements qui a été élaborée dans ce travail.

#### 1<sup>er</sup> exemple

Supposons que les événements de signal et de background se distribuent comme indiqué à la figure 3.1 en fonction d'une variable d'analyse  $x$  hypothétique dans l'intervalle  $[0,1]$ . Le problème qui se pose concrètement est de déterminer objectivement la valeur de la variable  $x$  à partir de laquelle on va sélectionner les événements, afin de garder un maximum de signal tout en rejetant un maximum de background. Dans l'exemple simple que nous présentons à la figure 3.1, on sélectionnera a priori les événements caractérisés par une faible valeur de  $x$ , étant donné qu'un grand nombre d'événements de signal présentent effectivement des valeurs faibles de la variable d'analyse  $x$ . Il est clair que plus on sélectionnera des événements caractérisés par des valeurs élevées de  $x$ , plus le nombre d'événements de signal ainsi retenus sera grand, mais plus le nombre d'événements de background retenus sera grand également. Effectuer une réduction des données de manière judicieuse consiste donc à trouver le meilleur compromis entre garder beaucoup de signal (mais alors inévitablement augmenter la proportion d'événements de background dans l'échantillon de données sélectionné) et rejeter beaucoup de background (mais alors inévitablement sacrifier des événements de signal).

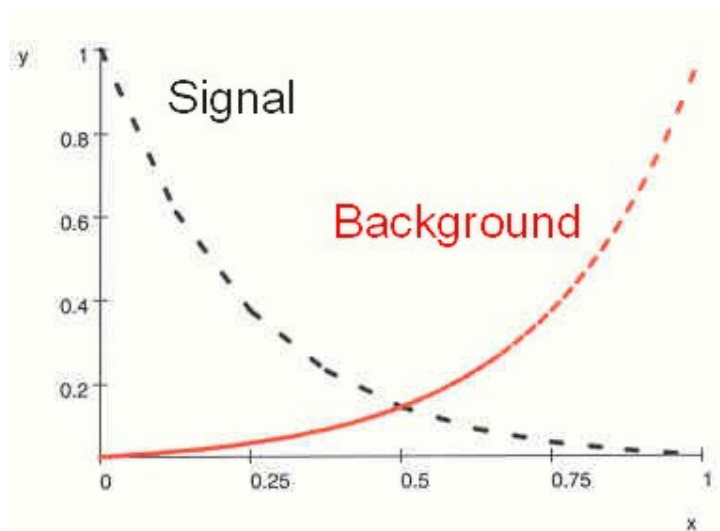


Fig. 3.1 : Distributions supposées des événements de signal et de background en fonction d'une variable d'analyse  $x$  hypothétique.

La méthode utilisée dans ce travail pour déterminer de manière objective l'endroit à partir duquel on sélectionnera les événements se base sur la considération suivante : en traçant les courbes intégrales normalisées, c'est-à-dire en portant en ordonnée  $y$ , pour chaque valeur de  $x$ , la valeur de

$$y = \frac{\int_0^x y(x') dx'}{\int_0^1 y(x') dx'} \quad (3.4)$$

pour chacune des distributions de signal et de background présentées à la figure 3.1, on obtient directement les courbes donnant la distribution des efficacités de retenue du signal  $\epsilon_{\text{signal}}$  et du background  $\epsilon_{\text{background}}$  en fonction de la valeur de  $x$  à partir de laquelle on sélectionne les événements (cf. fig. 3.2). En clair, si l'on décide par exemple de sélectionner les événements caractérisés par une valeur de  $x$  inférieure à 0,25<sup>6</sup>, on retiendra de la sorte environ 60% des événements de signal du départ et à peine 5% des événements de background.

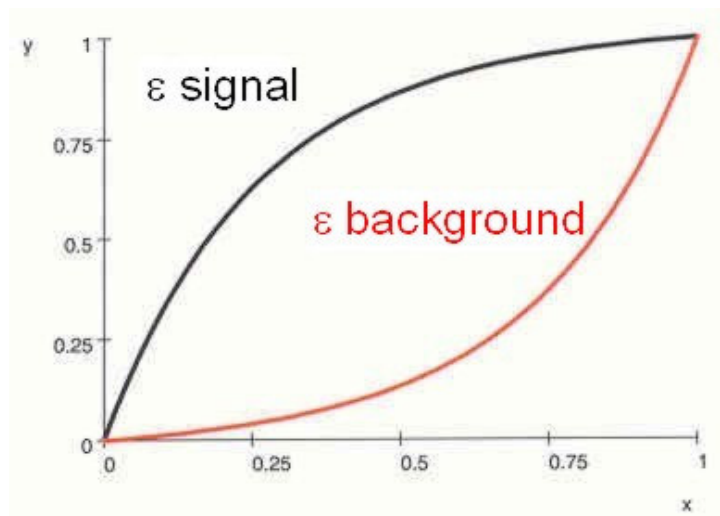


Fig. 3.2 : Courbes intégrées et normalisées à 1 associées aux distributions présentées à la fig. 3.1. Ces courbes représentent l'efficacité de retenue du signal (courbe noire) et l'efficacité de retenue du background (courbe rouge) en fonction de la valeur de la variable d'analyse  $x$ . On suppose que l'on sélectionne les événements caractérisés par une valeur de  $x$  inférieure à la coupure.

<sup>6</sup> On dira dans ce cas que l'on a placé la « coupure » en 0,25.

En divisant la courbe de l'efficacité de retenue du signal par la courbe d'efficacité de retenue du background, on obtient directement la courbe décrivant l'évolution de la valeur du facteur de rejet  $R$  (cf. éq. 3.1) en fonction de la valeur de la variable  $x$  choisie pour effectuer la sélection (cf. fig. 3.3).

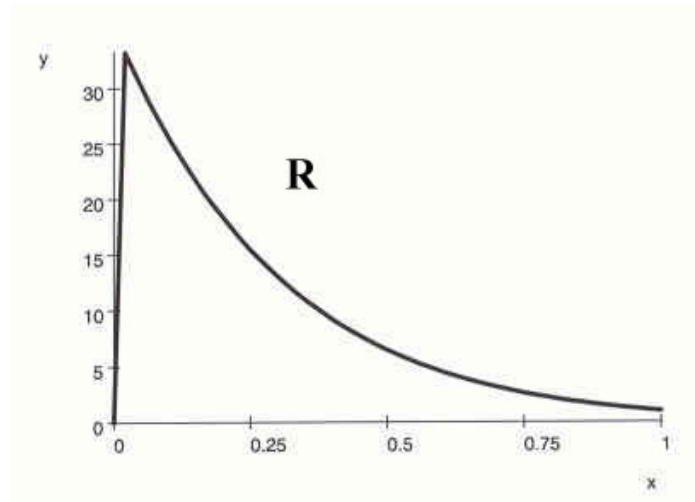


Fig. 3.3 : Distribution du facteur de rejet  $R$  en fonction de la variable d'analyse  $x$ . Cette distribution est obtenue en effectuant le quotient des courbes d'efficacité de retenue du signal et du background présentées à la figure 3.2. De nouveau, on suppose que l'on sélectionne les événements caractérisés par une valeur de  $x$  inférieure à la coupure.

On constate dans cet exemple que le choix d'une coupure  $x$  qui donne une valeur maximale de  $R$  conduirait à adopter une efficacité de retenue de signal  $\epsilon_{\text{signal}}$  négligeable, ce qui est totalement défavorable dans le cadre de la recherche d'un signal provenant d'une source ponctuelle.

Dans le but d'effectuer une sélection des événements caractérisée non seulement par un grand facteur de rejet  $R$  mais également une grande efficacité de retenue du signal **simultanément**, il semble naturel de chercher à maximiser le produit de  $R$  par  $\epsilon_{\text{signal}}$ . On trace à cet effet le graphe de la courbe donnée par le produit de  $R$  par  $\epsilon_{\text{signal}}$  (cf. fig. 3.4). La coupure est alors placée à l'endroit du maximum de  $R \cdot \epsilon_{\text{signal}}$ .

De manière analogue, on peut considérer le graphe de la différence des efficacités de retenue du signal et du background  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  (cf. fig. 3.5). Choisir la coupure qui maximise la valeur de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  revient en effet à sélectionner le plus de signal possible tout en sélectionnant le moins de background possible, ce qui est bien l'objectif poursuivi.

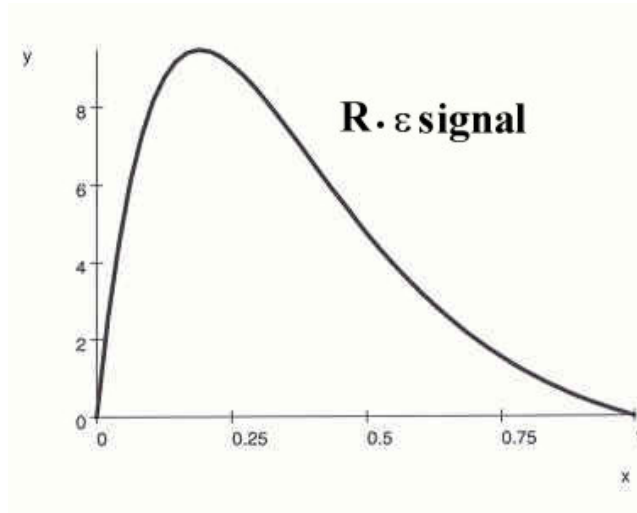


Fig. 3.4 : Distribution du produit du facteur de rejet  $R$  par l'efficacité de retenue du signal  $\epsilon_{\text{signal}}$ .

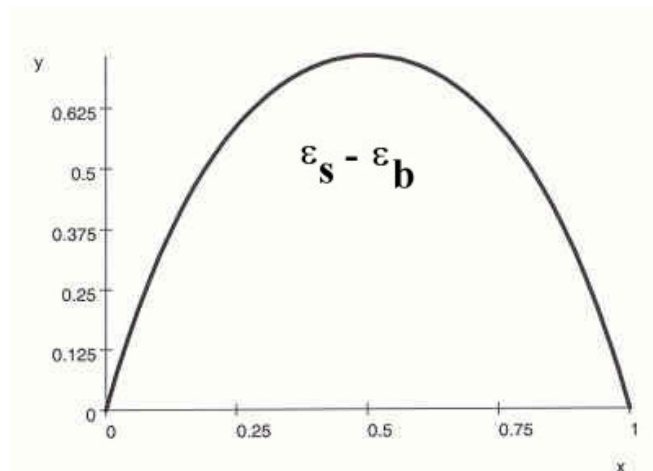


Fig. 3.5 : Distribution de la différence des efficacités de retenue du signal et du background en fonction de la valeur de  $x$  choisie pour effectuer la sélection.

Maximiser l'expression  $R \cdot [\epsilon_{\text{signal}} - \epsilon_{\text{background}}]$  (cf. fig. 3.6) constitue également une autre possibilité envisageable dans le processus de réduction des données.

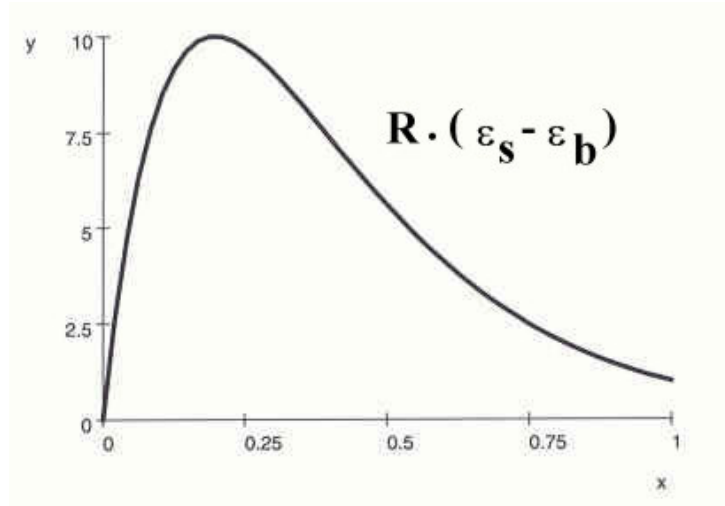


Fig. 3.6 : Distribution du produit du facteur de rejet R par la différence des efficacités de retenue du signal et du background.

A titre de comparaison, on a rassemblé dans le tableau 3.2 les valeurs de  $\epsilon_{\text{signal}}$ ,  $\epsilon_{\text{background}}$  et R pour les différents choix de coupures envisagés. On remarque que les coupures déterminées par le maximum des distributions de  $R \cdot (\epsilon_{\text{signal}} - \epsilon_{\text{background}})$  et  $R \cdot \epsilon_{\text{signal}}$  sont similaires, donnant lieu à de grandes valeurs du facteur de rejet R ( $\sim 20$ ), mais à des valeurs d'efficacité de retenue du signal relativement peu élevées (environ 50%). Par contre, la coupure déterminée par le maximum de la distribution de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  a pour effet de retenir une plus grande fraction du signal (près de 90%) quoiqu'elle conduise à une plus faible valeur de R (c'est-à-dire un rejet de background moins important).

<b>Coupure placée au maximum de la valeur :</b>	<b>X</b>	<b><math>\epsilon_{\text{signal}}</math></b>	<b><math>\epsilon_{\text{background}}</math></b>	<b>R</b>
$\epsilon_{\text{signal}} - \epsilon_{\text{background}}$	0,5	0,866	0,134	6,45
$R(\epsilon_{\text{signal}} - \epsilon_{\text{background}})$	0,189	0,526	0,028	19,0
$R\epsilon_{\text{signal}}$	0,1965	0,540	0,029	18,5

Tableau 3.2 : Valeurs des efficacités de retenue de signal et de background ainsi que du facteur de rejet R dans les différents cas de figure envisagés.

## 2<sup>ème</sup> exemple

Nous présentons ici un second exemple simulé, plus réaliste, dont la configuration ressemble à celle des variables d'analyse de type « smoothness » (voir la section 3.4.4 pour une description de ce type de variables).

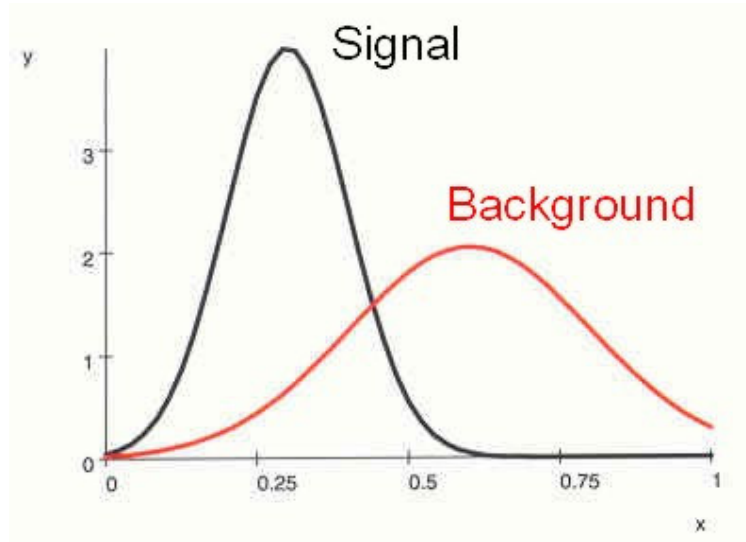


Fig. 3.7 : Distribution des événements de signal et de background en fonction d'une variable d'analyse de type « smoothness ».

Les graphiques similaires à ceux établis pour le premier exemple sont présentés aux figures 3.8, 3.9, 3.10, 3.11 et 3.12. Précisons que la sélection des événements se fait en deçà de la coupure.

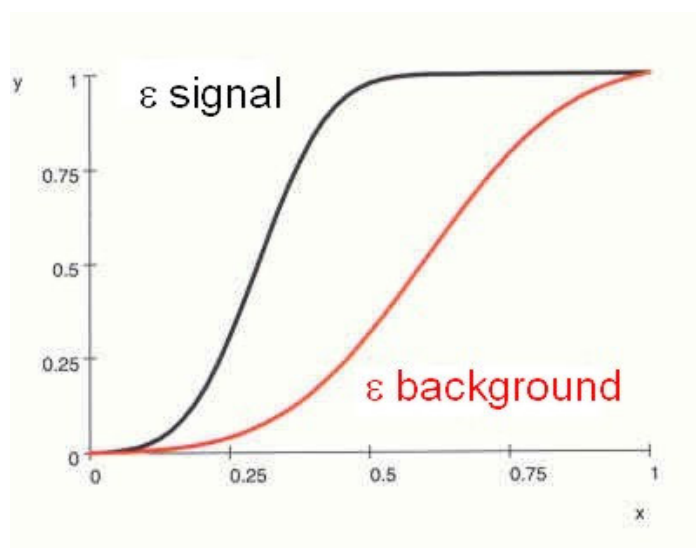


Fig. 3.8 : Courbes d'efficacité de retenue du signal et du background. On sélectionne les événements caractérisés par une valeur de x inférieure à la coupure.

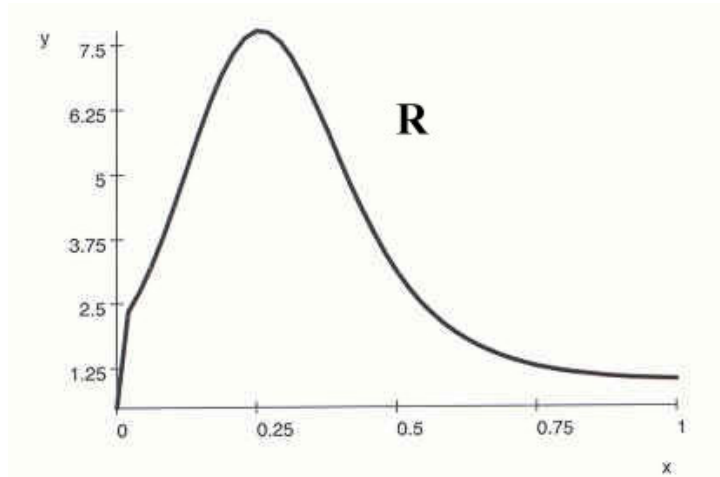


Fig. 3.9 : Distribution du facteur de rejet  $R$  en fonction de la variable d'analyse  $x$  choisie pour effectuer la coupure.

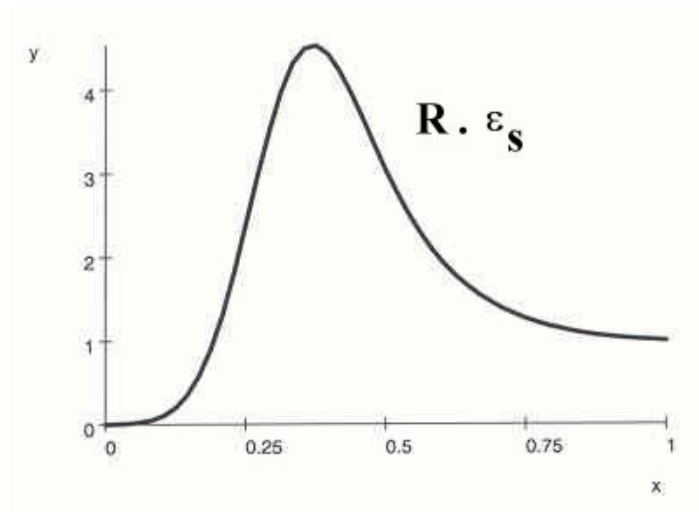


Fig. 3.10 : Distribution du produit du facteur de rejet  $R$  par l'efficacité de retenue du signal  $\epsilon_{\text{signal}}$ .

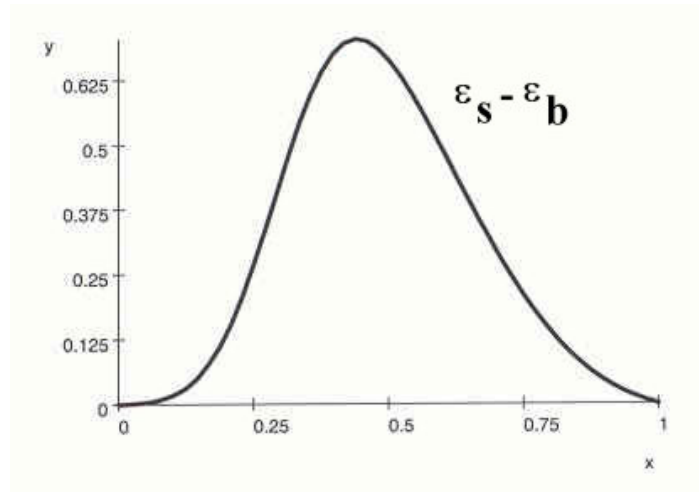


Fig. 3.11 : Distribution de la différence des efficacités de retenue du signal et du background.

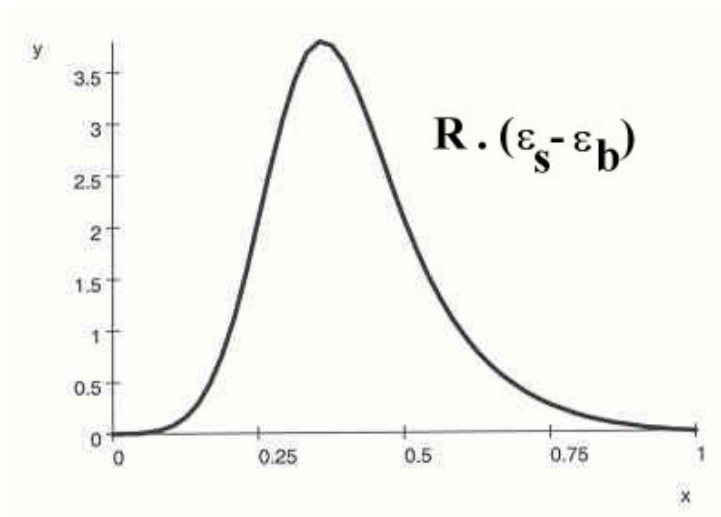


Fig. 3.12 : Distribution du produit du facteur de rejet  $R$  par la différence des efficacités de retenue du signal et du background.



Le tableau 3.3 rassemble les valeurs de  $\epsilon_{\text{signal}}$ ,  $\epsilon_{\text{background}}$  et  $R$  pour les différents choix de coupures envisagés. De nouveau, on constate que les coupures déterminées par le maximum des distributions de  $R \cdot (\epsilon_{\text{signal}} - \epsilon_{\text{background}})$  et  $R \cdot \epsilon_{\text{signal}}$  sont similaires<sup>7</sup> et conduisent, comme dans le cas précédent, à des efficacités de retenue du signal nettement inférieures à celle qu'on obtiendrait en appliquant la coupure déterminée par le maximum de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$ . Dans les deux exemples décrits ci-dessus, les conclusions en ce qui concerne les efficacités de retenue du signal et du background sont donc similaires, bien que les comportements des distributions soient très différents en fonction des variables d'analyse.

Coupure placée au maximum du graphe de:	X	$\epsilon_{\text{signal}}$	$\epsilon_{\text{background}}$	R
$\epsilon_{\text{signal}} - \epsilon_{\text{background}}$	0,44	0,919	0,216	4,26
$R(\epsilon_{\text{signal}} - \epsilon_{\text{background}})$	0,36	0,725	0,112	6,23
$R\epsilon_{\text{signal}}$	0,37	0,758	0,127	5,98

Tableau 3.3 : Valeurs des efficacités de retenue de signal et de background ainsi que du facteur de rejet R dans le cas où l'on sélectionne les événements à partir de la coupure déterminée par le maximum de différentes distributions.

Dans le cadre de la recherche d'une source ponctuelle de neutrinos d'origine astrophysique, l'objectif principal est de conserver un maximum de signal, quitte à accepter des événements de background supplémentaires. En effet, ceux-ci ne sont pas trop pénalisants puisqu'ils se répartissent de façon uniforme sur tout le ciel, alors que les événements provenant d'une éventuelle source ponctuelle galactique ou extragalactique hypothétique s'accumuleraient préférentiellement dans une zone restreinte du ciel. **Nous avons donc décidé de fixer la coupure au maximum de la distribution de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  dans le cadre de la réduction des données.** Cette coupure a pour effet de retenir la plus grande fraction d'événements de signal (cf. tableaux 3.2 et 3.3). La distribution de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  offre en outre l'avantage de toujours présenter un maximum bien marqué, ce qui n'est pas le cas en pratique pour les autres distributions. Enfin, on évite l'inconvénient rencontré dans des cas réalistes où R prend des valeurs artificiellement élevées résultant du quotient de petits nombres. Dans la suite, nous désignerons cette méthode de détermination des coupures par « méthode de la différence maximale ».

#### 3.4.4 La procédure suivie

La procédure suivie pour réduire les événements de données expérimentales a été la suivante : dans un premier temps, nous avons déterminé l'ensemble des variables d'analyse (comme décrit à la section 3.4.2), et ensuite nous avons appliqué la « méthode de différence maximale » aux événements de *signal MC* et de *BGMC* afin de déterminer les différentes coupures destinées à réduire les événements de données expérimentales.

<sup>7</sup> Il en va de même pour la distribution de R, qui présente un maximum en  $X=0,25$  et conduit à des valeurs de  $\epsilon_{\text{signal}}$  et R respectivement égales à 0,27 et 7,7.

Les variables d'analyse qui ont été choisies dans notre travail<sup>8</sup> sont reprises au tableau 3.4. Nous en donnons dans la suite une description plus détaillée.

Ldirb(2) ; ldirb(6)
Ldirc(2) ; ldirc(6)
Ndira(2); ndira(6)
Ndirb(2); ndirb(6)
Ndirc(2) ; ndirc(6)
Ndirc(6)-ndirc(5)
Jkrchi(3)
Smootallrl(2) ; smootallrl(6)
Smootallphit(3)

Tableau 3.4 : Variables d'analyse choisies pour appliquer la « méthode de différence maximale ». Les chiffres mentionnés entre parenthèses correspondent au numéro de la méthode utilisée pour reconstruire les traces de muons (cf. tableau 3.1)

- Les variables *ldirb* et *ldirc* donnent la distance maximale entre deux modules optiques qui ont enregistré des signaux avec des temps résiduels très petits (compris entre  $-15$  ns et  $+25$  ns pour *ldirb* et  $-15$  et  $+75$  ns pour *ldirc*). Cette distance est déterminée en projetant les modules optiques en question sur la trajectoire reconstruite du muon. Conserver préférentiellement les événements caractérisés par une grande valeur de cette variable revient à rejeter les événements qui sont localisés dans une partie réduite du détecteur et donc a priori moins bien reconstruits.
- Les variables *ndira*, *ndirb* et *ndirc* représentent, pour un événement donné, le nombre de fois que les modules optiques ont été touchés à des temps résiduels proches de 0 (c'est-à-dire compris entre  $-15$  ns et  $+15$  ns pour *ndira*,  $-15$  ns et  $+25$  ns pour *ndirb* et  $-15$  ns et  $+75$  ns pour *ndirc*). Plus ce nombre de « coups directs » est élevé, moins les photons Čerenkov ont été diffusés dans la glace, ce qui est une condition essentielle à la bonne reconstruction des trajectoires des muons.
- *Jkrchi* est le paramètre de vraisemblance  $L$  utilisé dans la reconstruction par la méthode du maximum de vraisemblance (cf. section 2.5.3).
- Les variables *smootallrl* et *smootallphit* sont les variables de type « smoothness » (De Young *et al.*, 1999 et Gaug *et al.*, 2000). Elles décrivent en quelque sorte l'uniformité de la brillance de la trace. Elles sont basées sur le principe du test de Kolmogorov-Smirnov, qui compare la distribution d'une observable le long d'une trace reconstruite à un modèle de référence. Dans le cas de la variable *smootallrl*, on compare la fraction cumulée du nombre de signaux enregistrés par les modules optiques le long de la

<sup>8</sup> Nous nous sommes basés sur les travaux antérieurs (Young, 2001 et Biron, 2002) pour cibler les variables d'analyse qui nous ont paru les plus appropriées à la réalisation de nos objectifs.

trajectoire reconstruite à un modèle linéaire attendu pour une distribution homogène de la lumière Čerenkov le long de la trajectoire du muon. La déviation maximum entre l'observable et le modèle de référence définit  $s$ , le paramètre de « smoothness » (cf. fig. 3.13).

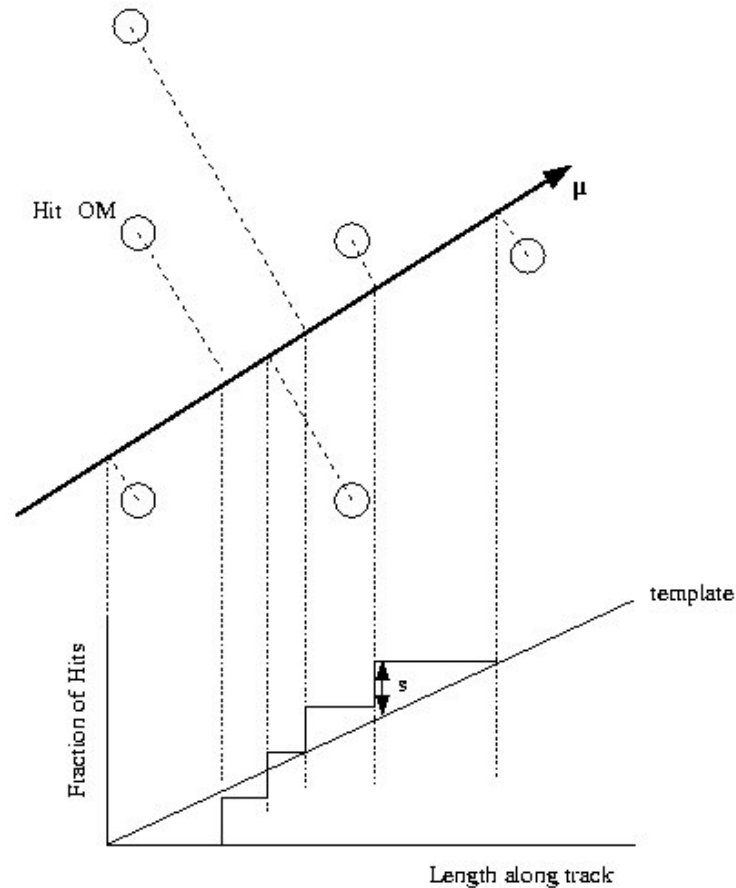


Fig. 3.13 : Définition du paramètre de « smoothness ». La distribution des modules optiques touchés (hits) en fonction de la longueur de la trace reconstruite est comparée à une distribution prédite pour une émission uniforme de lumière Čerenkov (template). La plus grande différence entre la distribution observée et la distribution prédite définit le paramètre de « smoothness »  $s$ , qui peut prendre des valeurs positives ou négatives.

Une faible valeur du paramètre  $s$  indique une distribution de lumière uniforme le long de la trajectoire reconstruite, compatible avec les événements associés aux traces de muons. Une valeur négative du paramètre de « smoothness » trahit la présence de lacunes entre les différents modules optiques qui ont été touchés, tandis qu'une valeur positive de ce paramètre indique qu'une grande proportion des modules optiques touchés est concentrée sur de faibles distances. Une coupure effectuée sur les valeurs positives de  $s$  a pour effet de rejeter les événements dont la topologie est compatible avec les événements produits par des gerbes. Une coupure sur les valeurs négatives du

paramètre de « smoothness » a pour effet de rejeter les traces de muons atmosphériques erronément reconstruites comme ascendantes et qui peuvent produire des lacunes anormales dans la distribution des modules optiques touchés.

L'ordre dans lequel les coupures ont été déterminées et appliquées a été établi comme suit : la « méthode de différence maximale » a été appliquée, pour chacune des variables d'analyse du tableau 3.4, aux distributions de *BGMC* et *signal MC* : dans chaque cas, les coupures optimales ont été déterminées par le maximum de la distribution de  $\epsilon_{\text{signal}} - \epsilon_{\text{background}}$  (comme décrit à la section 3.4.3) ; les valeurs de  $\epsilon_{\text{signal}}$ ,  $\epsilon_{\text{background}}$ ,  $\epsilon_{\text{data}}$  et  $R$  correspondant à ces coupures ont été calculées. La coupure qui présentait la valeur de  $R$  la plus grande a alors été appliquée au *signal MC*, au *BGMC* ainsi qu'aux données expérimentales. Le processus de détermination des coupures optimales par la « méthode de différence maximale » a été de nouveau appliqué sur les distributions de signal et de background réduites cette fois par la première coupure. Une fois encore, la coupure choisie et effectivement appliquée est celle qui présentait la valeur de  $R$  la plus élevée. Ce processus de réduction des données a été répété jusqu'à ce que les valeurs de  $R$  aient diminué pour devenir voisines de 1, signifiant par là qu'à ce stade les coupures effectuées rejettent autant d'événements de signal que de background (cf. éq. 3.1) et qu'il n'est plus avantageux de réduire davantage les données expérimentales.

Le tableau 3.5 présente l'ensemble des coupures optimisées par la « méthode de différence maximale » et que nous avons adoptées dans ce travail.

Niveau de coupure	$\epsilon_{\text{signal}}$ (%)	$\epsilon_{\text{background}}$ (%)	$R_{\text{background}}$	$\epsilon_{\text{données}}$ (%)	$R_{\text{données}}$
1. Zenith(2)>90	90.77	85.37	1.06	86.06	1.05
2. ndirb(2)>4	61.70	13.71	4.50	21.19	2.91
3. ndirc(6)>8	71.07	23.36	3.04	40.43	1.76
4. ldirb(6)>90	62.47	17.54	3.56	15.63	4.00
5. $-0.24 < \text{smootallrl}(6) < 0.26$	76.47	20.81	3.67	21.59	3.54
6. ndira(2)>5	59.28	17.59	3.37	22.26	2.66
7. $-0.13 < \text{smootallphit} < 0.23$	80.69	34.21	2.36	24.43	3.30
8. $-100 < \text{cogz} < 170$	100	100	1.00	96.74	1.03
9. zenith(6)>90	99.93	92.31	1.08	97.60	1.02
Total	9.1	~ 0.01	~ 910	~ 0.01	~ 910

Tableau 3.5 : Coupures optimisées par la « méthode de différence maximale » (mises à part les coupures 1, 8 et 9) et appliquées aux données expérimentales, au signal et au background. Dans chaque cas les efficacités (en %) de retenue du signal, des données expérimentales et du background ont été calculées après chaque coupure : au niveau de la coupure n°2, par exemple, les 100% d'événements correspondent au nombre d'événements qui subsistent après la coupure n° 1.

Il faut remarquer que les coupures n° 1, 8 et 9 n'ont pas été déterminées par la « méthode de différence maximale ». La première coupure a été imposée de manière

naturellement évidente, puisqu'elle concerne l'angle zénithal (« zenith ») des trajectoires de muons : pour observer les événements ascendants (provenant de l'hémisphère nord, comme expliqué à la section 2.4), il suffit de retenir ceux dont l'angle zénithal (reconstruit en l'occurrence par la méthode de vraisemblance maximum) est supérieur à  $90^\circ$ . Suivent alors six coupures déterminées par la « méthode de différence maximale ». La coupure n° 8, quant à elle, a un statut particulier : elle a pour but d'éliminer les signaux de bruit de fond situés en différents endroits du détecteur (notamment le bas de celui-ci) et qui sont encore mal compris (Biron *et al.*, 1999 et Biron *et al.*, 2001). La dernière coupure a enfin pour objectif d'éliminer les événements dont l'angle zénithal, reconstruit cette fois par la méthode itérative, est inférieur à  $90^\circ$ .

Au terme du processus de réduction des données, on a conservé approximativement 1000 fois plus de signal que de background, à savoir respectivement 10 %<sup>9</sup> et 0,01 % des événements de départ environ (cf. tableau 3.5). Notre échantillon de données expérimentales ainsi traité est donc largement « purifié » par rapport à l'échantillon de départ.

---

<sup>9</sup> Ce chiffre est important lorsqu'on désire connaître le nombre réel d'événements de signal qui ont été enregistrés dans une zone délimitée du ciel, dans le but de déterminer un flux de neutrinos provenant d'une source ponctuelle, par exemple.