

Présentation de l'analyse

Cette partie de la thèse est consacrée à l'étude du potentiel d'ATLAS pour la mesure de la section efficace de production inclusive du boson W dans le canal électron. Dans ce chapitre, les objectifs de l'analyse effectuée sur données simulées sont d'abord présentés. Le principe de la mesure et la topologie des bruits de fond possibles sont ensuite décrits. La dernière section introduit brièvement les outils logiciels nécessaires.

4.1 Objectifs

Les motivations principales pour la mesure de la section efficace de production du boson W dans le canal électron ont été présentées dans la section 2.3. L'analyse présentée dans cette thèse a été effectuée dans le cadre du *Computing System Commissioning* (CSC), au cours des années 2007-2008. L'objectif de cet effort de la collaboration était d'étudier précisément le potentiel de physique de l'expérience juste avant son démarrage et de mettre en place les outils, théoriques et techniques, nécessaires aux analyses. Depuis la publication du *Technical Design Report* [59] en 1999, la conception du détecteur a été finalisée, le détecteur a été entièrement installé, des tests en faisceau ont été effectués, la simulation du détecteur et les algorithmes de reconstruction ont été améliorés : cela rendait nécessaire une mise à jour du potentiel d'ATLAS. Cet effort a abouti à une publication [25], qui servira désormais de référence pour les analyses futures.

L'objectif du travail présenté est en particulier de développer une stratégie d'analyse adaptée à la mesure de la section efficace avec les premières données et d'estimer la précision attendue sur la mesure. Des attentions particulières sont mises sur l'utilisation d'une sélection simple des événements (à cause des imperfections du détecteur attendues au cours de la première phase de l'expérience) et sur le développement de méthodes basées sur les données au détriment de la simulation. Toute l'analyse a été effectuée pour une énergie dans le centre de masse de 14 TeV. Les effets d'une mesure à 10 TeV (comme prévue pour 2009-2010) seront discutés dans le chapitre 9.

4.2 Principe de la mesure

Le calcul théorique de la section efficace de production du boson W dans les collisions proton-proton a été présenté dans le chapitre 2. Expérimentalement, la section efficace σ est déterminée à partir du nombre d'événements W collectés de la manière suivante :

$$\sigma = \frac{N - B}{A \varepsilon \mathcal{L}}$$

où N est le nombre d'événements sélectionnés, B le nombre d'événements de bruit de fond, A l'acceptance géométrique et cinématique, ε est l'efficacité globale de reconstruction des événements et \mathcal{L} est la luminosité

intégrée de l'échantillon de données analysé. L'incertitude relative globale de mesure sur la section efficace est donnée par :

$$\frac{\delta\sigma}{\sigma} = \frac{\delta N \oplus \delta B}{N - B} \oplus \frac{\delta A}{A} \oplus \frac{\delta \varepsilon}{\varepsilon} \oplus \frac{\delta \mathcal{L}}{\mathcal{L}}$$

La détermination de la luminosité n'est pas étudiée dans cette analyse. Dans la phase de démarrage, elle devrait être connue à partir des paramètres de la machine avec une précision de l'ordre de 10% (voir Chap. 2). Le terme d'incertitude associé sera omis dans la suite.

L'accès à une base de données contenant les informations sur la luminosité mesurée au cours des runs analysés est essentiel pour une mesure de section efficace. Les outils nécessaires pour le calcul de la luminosité intégrée d'un échantillon de données à partir des *luminosity blocks* ont été développés puis testés durant un exercice appelé "Streaming Test" [60], auquel j'ai participé en tant qu'utilisateur. L'analyse du canal $W \rightarrow e\nu$ effectuée dans ce cadre est très similaire à celle présentée dans cette thèse et moins complète : pour éviter la redondance, elle ne sera donc pas développée ici. Cette analyse a fait l'objet d'une note interne [61] (note commune avec l'analyse du canal $Z \rightarrow ee$), jointe en Annexe A. Le principe de l'exercice était d'effectuer une analyse complète dans plusieurs canaux d'un échantillon de données ($\sim 18 \text{ pb}^{-1}$) créé à partir d'un mélange de processus du Modèle Standard. Le mélange avait été effectué en accord avec les sections efficaces et l'information Monte-Carlo (*i.e.* au niveau générateur) avait été enlevé des événements simulés : un lot de pseudo-données était ainsi obtenu. La principale limitation de l'exercice, du point de vue de l'analyse de physique, était liée à l'impossibilité technique de générer une statistique suffisante pour les événements de bruit de fond du type di-jets et γ -jet. Le choix du mode de *streaming* des données d'ATLAS à partir des signatures trigger (mode exclusif ou mode inclusif) était également un objectif majeur de cet exercice. En mode exclusif, un événement est stocké dans un unique *stream* (des priorités entre les différentes signatures trigger doivent donc être définies) ; en mode inclusif, un même événement peut être stocké dans plusieurs streams. Une recommandation pour un streaming inclusif a résulté des différentes analyses [62].

4.3 Bruits de fond

La première étape de l'analyse est de déterminer les bruits de fond possibles du canal $W \rightarrow e\nu$: ce sont les processus dont la signature expérimentale peut être identique au processus étudié et dont la production (résultant dans cette signature) est non négligeable par rapport à celle du signal. La signature des événements $W \rightarrow e\nu$ est la présence combinée d'un électron de haute impulsion transverse et d'une énergie transverse manquante élevée (\cancel{E}_T). Les bruits de fond possibles sont décrits qualitativement dans la suite. Les valeurs de section efficace indiquées proviennent de la référence [25] et les valeurs de rapport de branchement proviennent de la référence [11].

- $W \rightarrow e\nu$ ($\sigma = 20510 \text{ pb}$)
- $W \rightarrow \tau\nu$ ($\sigma = 20510 \text{ pb}$) avec $\tau \rightarrow e\nu_e\nu_\tau$ (B.R. ($\tau \rightarrow e\nu_e\nu_\tau$) = 0.1785) : un électron est produit dans l'état final et les trois neutrinos contribuent à \cancel{E}_T .
- $Z \rightarrow ee$ ($\sigma = 2015 \text{ pb}$) : un des deux électrons n'est pas identifié (produit en dehors de l'acceptance ou ne satisfait pas les conditions d'identification). \cancel{E}_T est mal reconstruite.
- $Z \rightarrow \tau\tau$ ($\sigma = 2015 \text{ pb}$) : un des deux taus se désintègre semi-leptoniquement en électron-neutrino (B.R. ($\tau \rightarrow e\nu_e\nu_\tau$) = 0.1785).
- $t\bar{t}$ ($\sigma = 833 \text{ pb}$) Le quark top se désintègre à 100% en Wb : un des deux W produits se désintègre semi-leptoniquement en électron-neutrino (B.R. ($W \rightarrow e\nu$) = 0.1080).
- Production inclusive de jets : ce sont des événements dans lesquels un jet est identifié comme un électron. Ce sont principalement des jets contenant un électron non isolé issu de la désintégration d'un hadron chargé. Le processus majoritaire est la production d'une paire de jets. Les événements photon-jet contribuent également. La section efficace est de l'ordre de plusieurs millibarns. Ce bruit de fond sera appelé QCD dans la suite.

Les sections efficaces de production combinées avec les rapports de branchement rendent négligeable la contribution des canaux $Z \rightarrow \tau\tau$ et $t\bar{t}$ par rapport aux autres bruits de fonds : seuls les canaux $W \rightarrow \tau\nu$, $Z \rightarrow ee$ et QCD seront considérés dans la suite.

4.4 Production de données simulées

4.4.1 La chaîne de traitement des données

L'environnement logiciel d'ATLAS est appelé Athena. Il dérive du projet Gaudi, initialement développé pour l'expérience LHCb. Il est basé sur le langage de programmation orienté objet C++. Les jobs (séquences d'algorithmes) sont contrôlés par des scripts Python. Un principe essentiel d'Athena est la séparation entre les données et les algorithmes d'une part, et entre les données temporaires (dans la mémoire) et les données permanentes (dans les fichiers de données) d'autre part. Toutes les étapes de traitement des données dans ATLAS sont effectuées dans Athena. La Figure 4.1 schématise les différentes étapes pour des données réelles ou simulées.

Dans le cas de données réelles, les données en sortie du détecteur sont traitées par des algorithmes de reconstruction pour être ensuite stockées dans un format de données directement exploitable pour l'analyse (ESD pour *Event Summary Data* ou AOD pour *Analysis Object Data*). Les objets physiques de même nature (p. ex. les traces du détecteur interne) sont regroupés dans des *containers*. Dans le cas de données simulées, la première étape est la génération de particules avant passage dans le détecteur, par un générateur Monte-Carlo pour un processus défini. Les particules produites sont ensuite propagées dans le détecteur ATLAS, en utilisant une simulation du détecteur ATLAS par Geant4 [64] : c'est un outil permettant de simuler l'interaction des particules avec la matière. La dernière étape est la reconstruction : dans cette étape, le traitement des données simulées est identique au cas des données réelles. Ainsi le format final des données simulées est identique à celui des données réelles ; la seule différence est la présence de l'information sur les particules produites au niveau générateur (appelée "la vérité Monte-Carlo" dans la suite). Une simulation rapide du détecteur (ATLFAST), permettant de gagner un facteur ~ 5 en temps de calcul, a également été développée : elle est particulièrement adaptée aux analyses dans lesquelles on veut étudier diverses valeurs d'un paramètre d'entrée (p. ex. la masse du boson de Higgs).

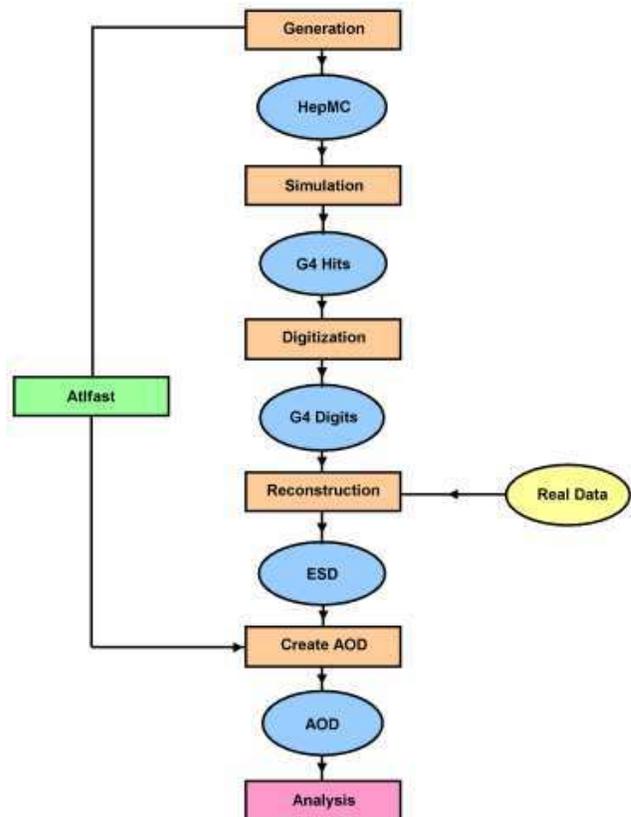


FIG. 4.1 – Traitement des données réelles et simulées dans ATLAS [63]

Canal	$\sigma(\times B.R.)$	$\varepsilon_{\text{filtre}}$	$N_{\text{evt}}(\times 10^3)$	$\mathcal{L}(\text{pb}^{-1})$
$W \rightarrow e\nu$	20510 pb	0.63	141	11
$\gamma/Z \rightarrow ee, \sqrt{\hat{s}} > 60 \text{ GeV}$	2015 pb	0.86	435	230
$W \rightarrow \tau\nu$	20510 pb	0.20	119	8
Inclusive jets ($p_T > 17 \text{ GeV}$)	2333 μb	0.09	2736	0.02

TAB. 4.1 – Echantillons de données simulées pour le signal et les bruits de fond. Les sections efficaces pour les processus W et Z sont normalisées à l'ordre NNLO en QCD. Pour le bruit de fond QCD, la section efficace est calculée à l'ordre dominant. Les filtres appliqués sont détaillés dans le texte. Les nombres d'événements simulés et les luminosités intégrées correspondantes sont indiqués.

4.4.2 La grille de calcul

La quantité de données, produites par les expériences du LHC, à traiter et à stocker a rendu nécessaire la mise en place d'une grille de calcul permettant de mutualiser les ressources de calcul et l'espace disque entre les différents centres de calcul des laboratoires participant aux expériences. Ce projet est appelé WLCG, pour *Worldwide LHC Grid Computing* [65]. La structure reliant les centres de calcul appartenant à la grille est hiérarchique : un Tier-0 (le CERN), ~ 10 Tiers-1 et ~ 80 Tiers-2. Chaque catégorie de centre possède des missions différentes (p. ex. de reconstruction, de simulation Monte-Carlo ou d'analyse).

4.4.3 Echantillons de données utilisés

Afin d'effectuer l'analyse, un nombre d'événements suffisant doit être simulé pour chaque processus (signal et bruits de fond). Le générateur PYTHIA [66] interfacé avec les PDFs CTEQ6L [46] a été utilisé pour tous les processus. D'autres générateurs ont été utilisés pour déterminer l'incertitude théorique sur l'acceptance (voir Chap. 8). Pour limiter le temps de simulation, des filtres sont appliqués au niveau générateur : pour un processus donné, un filtre a pour but d'enrichir artificiellement l'échantillon en événements "intéressants", en appliquant des coupures de présélection. Les propriétés des différents échantillons de données simulées sont résumées dans la Table 4.1. Les filtres appliqués sont les suivants :

- $W \rightarrow e\nu$: $p_T^e > 10 \text{ GeV}$, $|\eta^e| < 2.7$;
- $W \rightarrow \tau\nu$: un électron ou un muon tel que $p_T^l > 5 \text{ GeV}$ et $|\eta^l| < 2.8$;
- $Z \rightarrow ee$: $p_T^e > 10 \text{ GeV}$, $|\eta^e| < 2.7$, $\sqrt{\hat{s}} > 60 \text{ GeV}$;
- QCD : deux conditions sont imposées i) l'énergie transverse du processus dur doit être supérieure à 15 GeV ii) la somme de l'énergie transverse de toutes les particules stables dans la région $|\eta^e| < 2.7$ (en dehors des muons et des neutrinos) doit être supérieure à 17 GeV dans une tour de taille $\Delta\eta \times \Delta\Phi = 0.12 \times 0.12$. Le but de la seconde coupure est de reproduire au niveau générateur une sélection similaire à celle appliquée au L1 trigger pour identifier les objets électromagnétiques.

La simulation complète du détecteur a été utilisée. La géométrie du détecteur (ATLAS-CSC-01-02-00) est une géométrie altérée par rapport à la nominale : de la matière a été ajoutée (en particulier dans le cryostat et dans les cracks), de manière non symétrique en Φ . L'ajout peut représenter jusqu'à 10% de X_0 . L'hypothèse de travail effectuée pour l'analyse est l'accumulation de 50 pb^{-1} de données à une luminosité de $10^{31} \text{ cm}^2\text{s}^{-1}$. Le menu trigger utilisé est adapté à cette luminosité [25]. Les événements sont simulés sans empilement : les effets d'empilement peuvent être négligés à $10^{31} \text{ cm}^2\text{s}^{-1}$. Dans la suite, tous les résultats seront normalisés à 50 pb^{-1} . Il est à noter que par rapport à cette luminosité intégrée de référence, le nombre d'événements QCD simulés est très insuffisant (facteur ~ 2500). De plus, l'incertitude théorique sur la section efficace de production de ces événements est un facteur ~ 3 : cette incertitude est due principalement aux contributions non-perturbatives et à la PDF du gluon à bas x .

4.4.4 Outils d'analyse

La version d'Athena utilisée est 12.0.6. Le format de données utilisé est les AODs. L'accès aux AODs sur la grille est effectué via l'outil d'analyse distribuée GANGA [67]. L'analyse finale est effectuée dans ROOT [68].

Chapitre 5

Reconstruction des événements $W \rightarrow e\nu$

Les performances du détecteur ATLAS pour la reconstruction des événements $W \rightarrow e\nu$ sont étudiées sur des données simulées. Les algorithmes développés pour le trigger, l'identification de l'électron et le calcul de l'énergie transverse manquante sont décrits. Les résultats attendus pour 50 pb^{-1} de données à 14 TeV sont finalement présentés.

5.1 Sélection des événements

Une méthode simple de coupures a été utilisée pour sélectionner le signal. Les coupures suivantes ont été appliquées :

- la signature trigger e20 (un électron d'énergie supérieure à 20 GeV) doit être passée ;
- exactement un électron tel que $E_T > 25 \text{ GeV}$, $|\eta| < 1.37$ ou $1.52 < |\eta| < 2.4$ et satisfaisant le critère d'identification *medium*¹⁾ doit être présent dans l'événement ;
- \cancel{E}_T doit être supérieure à 25 GeV.

Les trois coupures (trigger, sélection de l'électron et \cancel{E}_T), ainsi que leur impact sur le signal et le bruit de fond, sont détaillées dans les sections suivantes. Dans la suite, les coupures appliquées au niveau du trigger seront appelées *online* et les coupures appliquées au niveau des données reconstruites seront appelées *offline*.

5.2 Sélection de l'électron offline

L'identification offline de l'électron est présentée avant la sélection online par souci de clarté : en effet, la sélection au HLT utilise des critères identiques à ceux utilisés offline.

Dans la reconstruction, un électron résulte de l'association d'un amas de cellules de l'EMC (*cluster* électromagnétique) avec une trace du détecteur interne.

5.2.1 Cluster électromagnétique

La première étape de reconstruction d'un électron est l'identification d'un cluster électromagnétique d'énergie suffisante. L'algorithme de clusterisation utilisé est l'algorithme²⁾ *sliding window* [69]. Il procède en trois étapes.

- Construction de *calotowers* de taille $\Delta\eta \times \Delta\Phi = 0.025 \times 0.025$:
l'énergie de toutes les cellules dans les trois compartiments (et éventuellement le pré-échantillonneur) est sommée pour chaque calotower.

¹⁾Trois niveaux d'identification de l'électron existent dans ATLAS, appelés par ordre croissant de sélectivité *loose*, *medium* et *tight* [25].

²⁾Un autre algorithme appelé *TopoCluster* a également été développé : il est détaillé dans la même référence.

- Définition d'un *precluster* :
une fenêtre de taille fixe (5×5 en unités de calorimeters) est déplacée dans la grille définie par les calorimeters. Si l'énergie transverse d'une fenêtre est un maximum local et est supérieure à 3 GeV, alors un precluster est défini. La position du precluster est calculée à partir des barycentres en énergie en η et en Φ .
- Formation du cluster :
les cellules appartenant au cluster sont définies à partir d'un rectangle de taille fixe. Pour les électrons, la taille est 3×7 pour l'EMB et 5×5 pour l'EMEC. Pour le middle, le rectangle est centré sur la position du precluster, puis la position des centres pour le front, le pré-échantillonneur et le back sont successivement calculés, en utilisant les barycentres en énergie. L'optimisation de la taille du cluster résulte de la pondération entre deux contraintes : la nécessité de contenir le maximum de l'énergie de la gerbe et la minimisation de l'impact du bruit. La position finale du cluster est calculée pour chaque compartiment longitudinal comme le barycentre en énergie des cellules du cluster appartenant à ce compartiment.

L'énergie du cluster est calculée linéairement à partir des dépôts d'énergie dans chaque compartiment, en utilisant des poids longitudinaux optimisés en fonction de η :

$$E = s(\eta)[c(\eta) + w_0(\eta) \cdot E_{PS} + E_{\text{front}} + E_{\text{middle}} + w_3(\eta) \cdot E_{\text{back}}]$$

où $s(\eta)$ est un facteur global, $c(\eta)$ est un offset, $w_0(\eta)$ corrige pour l'énergie perdue en amont du pré-échantillonneur, $w_3(\eta)$ corrige pour la fuite hadronique et E_{PS} , E_{front} , E_{middle} , E_{back} sont les énergies déposées dans les couches successives du calorimètre. Les poids sont calculés à partir d'une minimisation de χ^2 de la quantité $(E_{\text{true}} - E_{\text{reco}})^2 / \sigma(E_{\text{true}})^2$ sur des événements simulés à un seul électron, d'énergie comprise entre 10 et 200 GeV.

5.2.2 Reconstruction des traces

Plusieurs méthodes de reconstruction des traces du détecteur interne existent dans ATLAS. La méthode utilisée dans cette analyse est la méthode standard (appelée *New Tracking*) ; elle sera utilisée dans la première phase de l'expérience. La reconstruction est effectuée en cinq étapes [70] :

- des graines de trace sont formées à partir d'une combinaison de hits des trois couches de pixels et de la première couche du SCT ;
- à partir des graines de trace, les hits des autres couches du SCT sont ajoutés pour former des candidats trace ;
- les trajectoires des candidats trace sont ajustées : les ambiguïtés sont résolues en appliquant des critères de qualité (*track scoring strategy*) ;
- à partir des candidats trace sélectionnés, les hits du TRT sont ajoutés ;
- les traces sont réajustées à partir des hits des trois détecteurs (pixels, SCT et TRT).

5.2.3 Association cluster-trace

A partir des clusters électromagnétiques et des traces reconstruits, les objets électron peuvent être formés. L'association cluster-trace est effectuée en deux étapes :

- une association angulaire grossière est demandée, sans extrapolation de la trace :

$$\Delta\eta = |\eta_{\text{cluster}} - \eta_{\text{trace}}| < 0.1 \text{ et } \Delta\phi = |\phi_{\text{cluster}} - \phi_{\text{trace}}| < 0.1$$

- la trace correspondant au meilleur rapport E/p est choisie.

Cette méthode d'association part du cluster vers la trace. Afin d'augmenter l'efficacité d'association (en particulier pour les électrons de basse impulsion transverse), une méthode complémentaire partant des traces (appelée *soft electron*) a été développée : cette méthode n'est pas utilisée dans l'analyse.

Au niveau des AODs, les containers *electron* et *photon* sont séparés : ils dérivent d'un même container *egamma* des ESDs contenant tous les clusters électromagnétiques. Si un objet du container *egamma* est associé à une trace (en utilisant les contraintes larges décrites précédemment), il est placé dans le container *electron* ; sinon il est placé dans le container *photon*. Dans la suite, un candidat électron désigne un objet du container *electron*³⁾. Il est à noter que les photons convertis sont placés dans le container *electron* (des outils ont été récemment développés pour récupérer les conversions).

5.2.4 Identification de l'électron

Les électrons ne seront étudiés que dans la zone $|\eta| < 1.37$ et $1.52 < |\eta| < 2.4$ (zone de précision pour les électrons).

Coupages calorimétriques

Les coupures ont été optimisées sur 7 bins en η , définis par les seuils : 0.8, 1.35, 1.5, 1.8, 2.0, 2.35 et 2.4. Pour chaque coupure, les valeurs minimales et maximales des seuils utilisés seront indiquées.

- **La fuite hadronique**

$$F_{\text{had}} = \frac{E_{\text{Thad1}}}{E_T} < 0.015 - 0.045$$

où E_{Thad1} est l'énergie transverse déposée par le cluster dans le premier compartiment du calorimètre hadronique et E_T est l'énergie transverse du cluster.

- **Dans le compartiment milieu**

- L'étalement latéral de la gerbe :

$$R_{\eta 37} = \frac{e_{237}}{e_{277}} > 0.8 - 0.91$$

où e_{237} (respectivement e_{277}) est l'énergie calculée dans un rectangle de taille 3×7 (respectivement 7×7) en unités de cellules $\eta \times \Phi$.

- La largeur de la gerbe calculée dans un rectangle 3×5 :

$$\omega_{\eta 2} = \sqrt{\frac{\sum_i E_i \eta_i^2}{\sum_i E_i} - \left(\frac{\sum_i E_i \eta_i}{\sum_i E_i} \right)^2} < 0.0125 - 0.014$$

où E_i est l'énergie de la i^{e} cellule et η_i est la position relative de la i^{e} cellule par rapport à η_{cluster} .

- **Dans le compartiment avant** : une fenêtre de taille $\Delta\eta \times \Delta\Phi = 0.125 \times 0.2$ autour de la strip du cluster de plus haute énergie est considérée.

- Un deuxième maximum est recherché et utilisé dans deux variables :

$$\Delta E_1 = E_{\text{max2}} - E_{\text{min1}} < 0.25 - 0.53 \text{ GeV}$$

où E_{max2} est l'énergie du deuxième maximum et E_{min1} est l'énergie de la strip la moins énergétique entre les deux maxima ;

$$R_{\text{max2}} = \frac{E_{\text{max2}}}{1 + 9 \cdot 10^{-3} E_T} < 0.15 - 0.3$$

³⁾Deux coupures très souples sont appliquées sur les candidats *egamma* lors de la production des containers *electron* et *photon* (rapport E/p pour les électrons et fuite hadronique pour les photons) : la coupure sur les électrons a un impact négligeable pour le signal (l'impact est non négligeable pour des électrons de très basse impulsion transverse).

où E_T est l'énergie transverse du cluster en GeV. Le facteur $9 \cdot 10^{-3}$ dépend de la luminosité.

- La largeur totale de la gerbe calculée sur ~ 40 strips :

$$\omega_{\text{tot1}} = \sqrt{\frac{\sum_i E_i (i - i_{\text{max}})^2}{\sum_i E_i}} < 1.4 - 4.0$$

où i est l'indice de la strip et i_{max} l'indice de la strip la plus énergétique.

- La largeur de la gerbe calculée sur les trois strips centrales (même formule que précédemment) :

$$\omega_{3\text{strips}} < 0.6 - 0.80$$

- La fraction d'énergie du cluster déposée dans le compartiment avant :

$$f_1 = \frac{E_1}{E} > 0.005$$

où E_1 est l'énergie du cluster déposée dans le compartiment avant et E est l'énergie du cluster.

- La fraction d'énergie en dehors des trois strips centrales :

$$F_{\text{core}} = \frac{E_{\pm 3} - E_{\pm 1}}{E_{\pm 1}} < 0.2 - 0.48$$

où $E_{\pm i}$ est l'énergie du cluster déposée dans $\pm i$ strips autour de la strip de plus haute énergie.

Aucune coupure n'est appliquée sur l'énergie déposée par le cluster dans le compartiment arrière.

Coupages sur la trace

Des coupures sont appliquées sur la qualité de la trace :

- 9 hits de précision (pixels et SCT)
- 2 hits dans les pixels (dont 1 dans le b -layer)
- un paramètre d'impact $|d_0| < 0.1$ cm
- l'association angulaire est contrainte en extrapolant la trace jusqu'au calorimètre :

$$\Delta\eta = |\eta_{1,c} - \eta_{1,t}| < 0.05 \text{ et } \Delta\Phi = |\Phi_{2,c} - \Phi_{2,t}| < 0.02$$

où $\eta_{1,c}$ est la pseudo-rapacité du cluster dans le compartiment avant, $\eta_{1,t}$ est la pseudo-rapacité de la trace extrapolée dans le compartiment avant (*idem* pour Φ dans le compartiment milieu)

- une contrainte sur E/p est ajoutée :

$$0.9 < \frac{E}{p} < 2.5 - 3.5$$

5.2.5 Résultats pour le signal et le bruit de fond

Les variables d'identification sont étudiées en appliquant la présélection suivante : exactement un candidat électron (*i.e.* un objet du container électron) tel que $E_T > 25$ GeV, $|\eta| < 1.37$ ou $1.52 < |\eta| < 2.4$ doit être présent dans l'événement. Les distributions de l'énergie transverse des candidats obtenus pour le signal et les bruits de fond sont représentés sur la Figure 5.1. Les distributions des variables les plus discriminantes sont représentées sur les Figures 5.2 à 5.9⁴⁾ pour les événements de signal et du bruit de fond QCD. Les canaux $Z \rightarrow ee$ et $W \rightarrow \tau\nu$ ne sont pas représentés car les distributions sont identiques à celles du signal.

⁴⁾Pour des raisons techniques, la position de la trace extrapolée, utilisée pour l'identification, n'était pas accessible : la distribution de $\Delta\Phi$ représentée a été obtenue sans extrapolation.

Afin d'estimer l'efficacité de reconstruction de l'électron, la vérité Monte-Carlo est utilisée dans l'échantillon de signal pour associer l'électron reconstruit à l'électron vrai issu du W (cône de rayon $\Delta R_{\text{asso}} = \sqrt{\Delta\eta^2 + \Delta\Phi^2} < 0.1$). L'efficacité de reconstruction de l'électron pour le signal après chaque catégorie de coupures est donnée dans la Table 5.1 ; les facteurs de réjection pour les jets à partir du nombre de candidats présélectionnés sont également indiqués. Le facteur global de réjection pour des jets entre 10 et 40 GeV avec cette sélection est ~ 2000 [25].

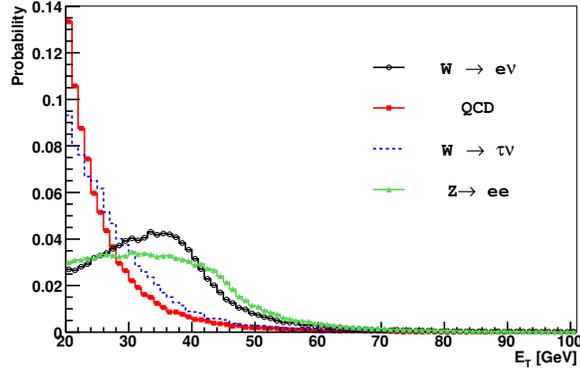


FIG. 5.1 – Distribution d'énergie transverse du candidat électron présélectionné pour le signal et les bruits de fond. Le seuil de présélection a été abaissé à 20 GeV pour la figure.

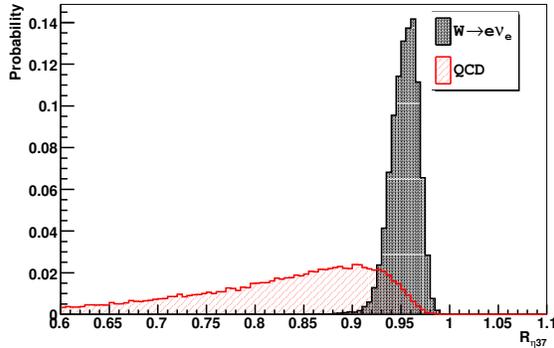


FIG. 5.2 – Distribution normalisée de $R_{\eta 37}$ pour le signal et le bruit de fond QCD

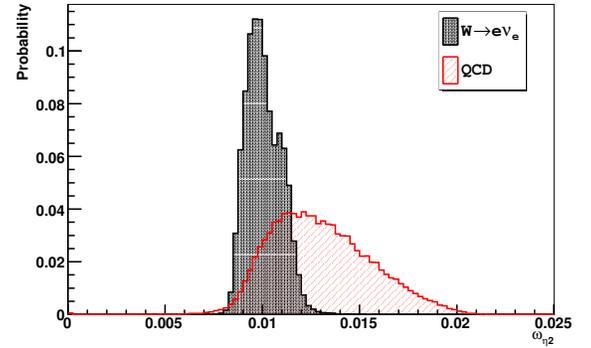


FIG. 5.3 – Distribution normalisée de $\omega_{\eta 2}$ pour le signal et le bruit de fond QCD

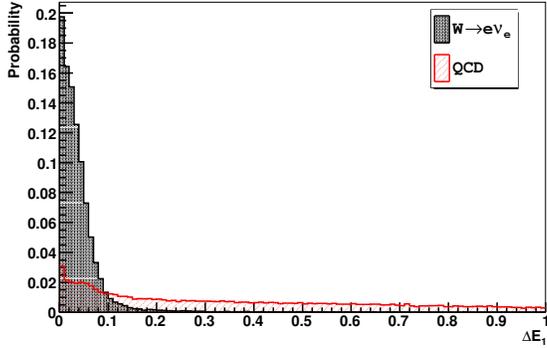


FIG. 5.4 – Distribution normalisée de ΔE_1 pour le signal et le bruit de fond QCD

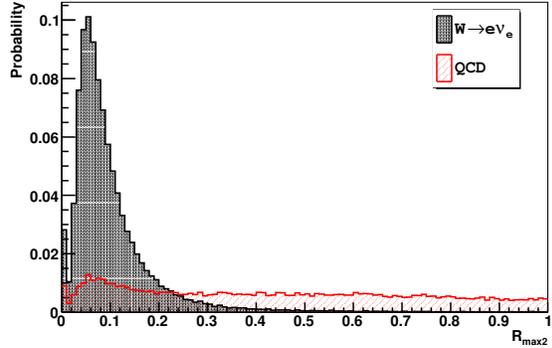


FIG. 5.5 – Distribution normalisée de $E_{\max 2}$ pour le signal et le bruit de fond QCD

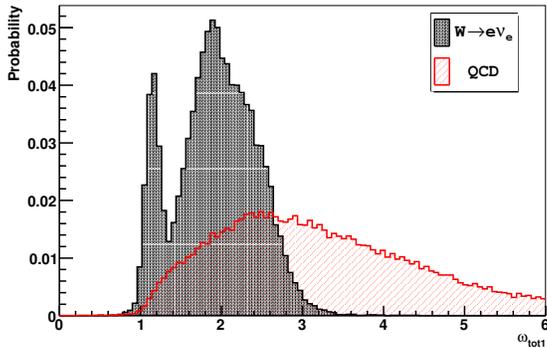


FIG. 5.6 – Distribution normalisée de $\omega_{\text{tot}1}$ pour le signal et le bruit de fond QCD. La double structure observée pour le signal est due au changement de granularité du compartiment avant pour $\eta = 1.8$.

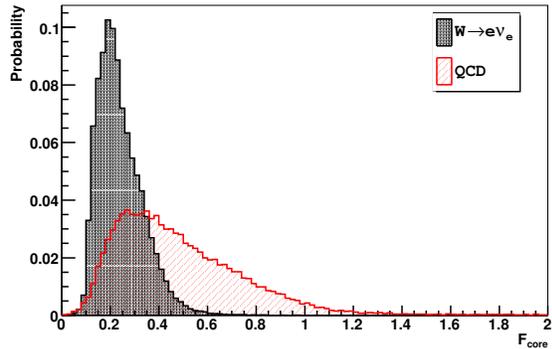


FIG. 5.7 – Distribution normalisée de F_{core} pour le signal et le bruit de fond QCD

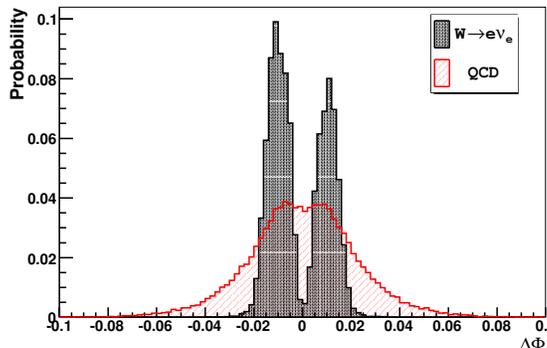


FIG. 5.8 – Distribution normalisée de $\Delta\Phi$ (sans extrapolation) pour le signal et le bruit de fond QCD

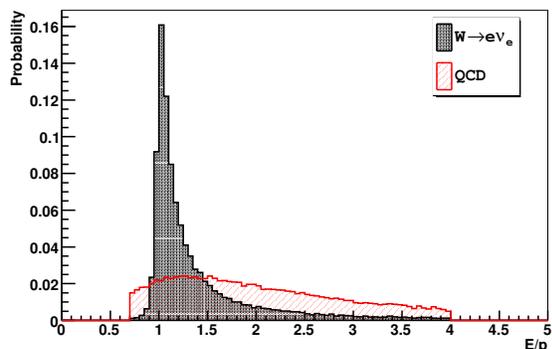


FIG. 5.9 – Distribution normalisée de E/p pour le signal et le bruit de fond QCD

Coupsures	Efficacité	Réjection
Présélection	0.853 ± 0.004	1
Fuite hadronique	0.851 ± 0.004	3.40 ± 0.03
Compartiment milieu	0.849 ± 0.004	6.58 ± 0.08
Compartiment avant	0.788 ± 0.004	29.9 ± 0.8
Qualité de la trace	0.704 ± 0.004	51 ± 2

TAB. 5.1 – Efficacité de reconstruction de l'électron pour le signal et facteur de réjection (relatif par rapport au nombre de candidats présélectionnés) pour le bruit de fond QCD après chaque catégorie de coupures. Les incertitudes indiquées sont les erreurs binomiales dues aux nombres d'événements simulés.

5.3 Signature trigger e20

Le menu trigger utilisé a été développé pour une luminosité de $10^{31} \text{ cm}^2\text{s}^{-1}$. La signature e20 est initialisée par la signature EM18 au L1, correspondant à un dépôt d'énergie de 18 GeV dans l'EMC (sans critère d'isolation). Le calcul de l'énergie au L1 est basé sur un algorithme sliding window à partir de tours trigger (de granularité $\Delta\eta \times \Delta\phi \sim 0.1 \times 0.1$), procédant en trois étapes :

- une fenêtre 4×4 (en tours électromagnétiques et hadroniques) doit être un maximum local en énergie ;
- le centre de la fenêtre (quatre tours, voir Fig. 5.10) doit être un maximum local en énergie transverse ;
- la combinaison la plus énergétique de deux tours électromagnétiques voisines (parmi les quatre combinaisons possibles de tours du centre de la fenêtre) doit avoir une énergie supérieure à 18 GeV.

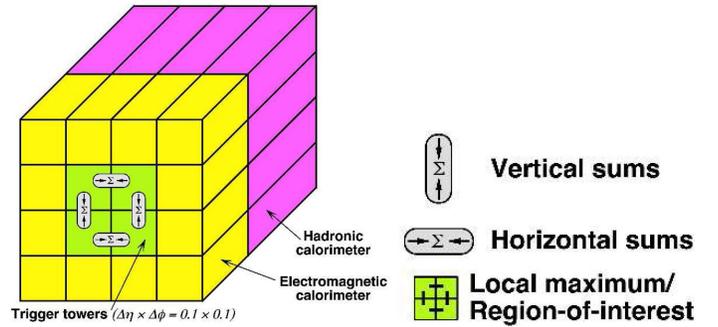


FIG. 5.10 – Schéma du trigger L1, montrant les combinaisons de tours pour le calcul de l'énergie électromagnétique [25]

Au HLT, le critère d'identification *loose* est utilisé pour le candidat électron : seules les coupures sur la fuite hadronique et sur les propriétés de la gerbe dans le compartiment milieu sont appliquées.

5.4 Reconstruction de l'énergie transverse manquante

5.4.1 Algorithme de reconstruction

L'algorithme utilisé est basé sur les dépôts d'énergie dans les cellules du calorimètre⁵⁾ : il prend en compte l'énergie des cellules après une procédure de soustraction du bruit. Cette procédure utilise les TopoClusters 4/2/0 : i) un TopoCluster part d'une cellule graine d'énergie $|E_{\text{cell}}| > 4 \sigma_{\text{noise}}$; ii) les cellules voisines d'énergie $|E_{\text{cell}}| > 2 \sigma_{\text{noise}}$ sont ajoutées ; iii) la couche de cellules aux frontières est ajoutée ($|E_{\text{cell}}| > 0 \sigma_{\text{noise}}$). En plus des dépôts d'énergie dans les calorimètres, le calcul corrige pour l'énergie perdue dans le cryostat et pour l'énergie des muons mesurés dans le spectromètre :

$$\cancel{E}_{x,y}^{\text{Final}} = \cancel{E}_{x,y}^{\text{Calo}} + \cancel{E}_{x,y}^{\text{Muon}} + \cancel{E}_{x,y}^{\text{Cryo}}$$

⁵⁾Un algorithme partant des objets reconstruits (*object-based* \cancel{E}_T) a également été développé [25].

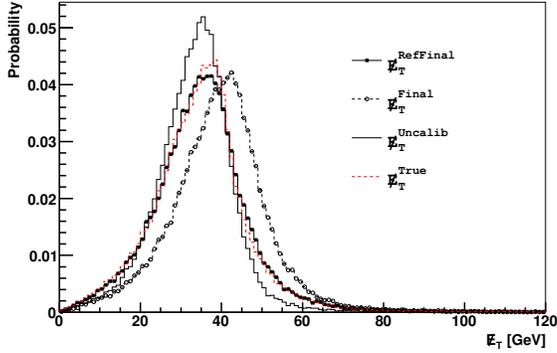


FIG. 5.11 – Distributions d'énergie transverse manquante pour les algorithmes $\cancel{E}_T^{\text{RefFinal}}$, $\cancel{E}_T^{\text{Final}}$, $\cancel{E}_T^{\text{Uncalib}}$ et $\cancel{E}_T^{\text{True}}$

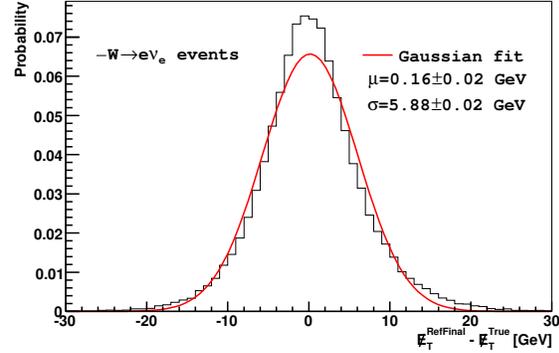


FIG. 5.12 – Résolution en énergie transverse manquante pour l'algorithme $\cancel{E}_T^{\text{RefFinal}}$

où :

- $\cancel{E}_{x,y}^{\text{Calo}} = - \sum_{\text{TopoCells}} E_{x,y}$ (toutes les cellules appartenant aux TopoClusters contribuent). Dans cette étape, les cellules sont étalonnées en utilisant la méthode H1⁶⁾ : des poids sont calculés en fonction de la densité d'énergie E/V de la cellule à partir de jets dans des événements simulés et appliqués pour le calcul de l'énergie par cellule.
- $\cancel{E}_{x,y}^{\text{Muon}} = - \sum_{\text{muons}} E_{x,y}$ (l'énergie perdue par le muon dans le calorimètre est prise en compte dans le terme calorimétrique)
- $\cancel{E}_{x,y}^{\text{Cryo}} = - \sum_{\text{jets}} w^{\text{Cryo}} \sqrt{E_{\text{EM3}} \times E_{\text{HAD}}}$ où w^{Cryo} est un poids d'étalonnage calculé par la méthode H1 et E_{EM3} et E_{HAD} sont les énergies déposées par le jet respectivement dans le compartiment arrière de l'EMC et dans le premier compartiment du calorimètre hadronique.

La dernière étape est le raffinement du terme calorimétrique. Pour les cellules des TopoClusters appartenant à un objet reconstruit, l'étalonnage est effectué en fonction de la nature de l'objet. Si une cellule appartient à plusieurs objets, l'ordre de priorité utilisé est le suivant : électron, photon, muon, tau (désintégration hadronique), b -jet et jet léger. Les cellules des TopoClusters n'appartenant à aucun objet reconstruit sont pris en compte sans correction. Cette méthode suppose que l'étalonnage des objets est meilleur que l'étalonnage global.

5.4.2 Performances sur les événements de signal

Les performances de la reconstruction de \cancel{E}_T peuvent être étudiées à partir de la simulation sur des événements de signal, en comparant la quantité reconstruite à la quantité vraie ($\cancel{E}_T^{\text{True}}$) : $\cancel{E}_T^{\text{True}}$ est calculée à partir de toutes les particules stables n'interagissant pas.

La présélection suivante est appliquée : un électron tel que $E_T > 25$ GeV, $|\eta| < 1.37$ ou $1.52 < |\eta| < 2.4$ et satisfaisant le critère d'identification *medium* doit être présent dans l'événement. Trois algorithmes sont étudiés :

- $\cancel{E}_T^{\text{RefFinal}}$ (décrit précédemment)
- $\cancel{E}_T^{\text{Final}}$: la dernière étape de raffinement n'est pas appliquée.
- $\cancel{E}_T^{\text{Uncalib}}$: l'énergie des cellules est calculée à l'échelle électromagnétique.

Les distributions de \cancel{E}_T et $\cancel{E}_T^{\text{True}}$ sont représentées sur la Figure 5.11. Les Figures 5.12, 5.13 et 5.14 représentent

⁶⁾ Une méthode appelée *Local-Hadronic calibration* a également été développée : des poids sont appliqués seulement aux clusters hadroniques à partir de la simulation d'événements à un pion.

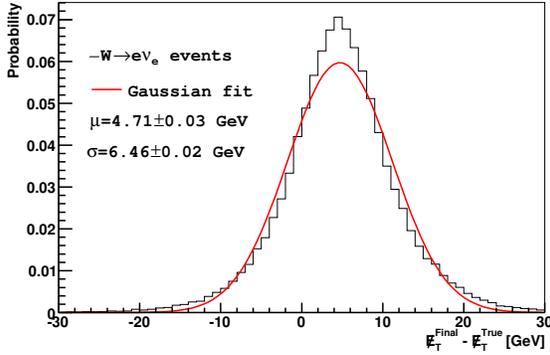


FIG. 5.13 – Résolution en énergie transverse manquante pour l’algorithme $\cancel{E}_T^{\text{Final}}$

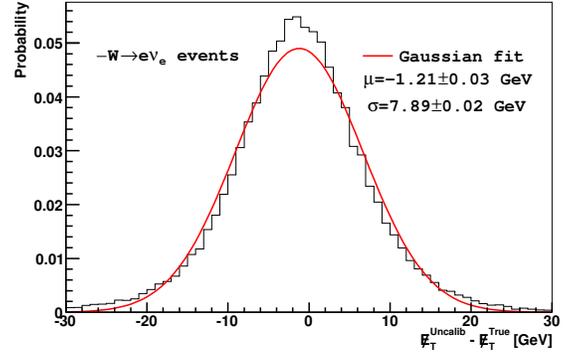


FIG. 5.14 – Résolution en énergie transverse manquante pour l’algorithme $\cancel{E}_T^{\text{Uncalib}}$

Algorithme	μ (GeV)	σ (GeV)
$\cancel{E}_T^{\text{RefFinal}}$	0.16 ± 0.02	5.88 ± 0.02
$\cancel{E}_T^{\text{Final}}$	4.71 ± 0.03	6.46 ± 0.02
$\cancel{E}_T^{\text{Uncalib}}$	-1.21 ± 0.03	7.89 ± 0.2

TAB. 5.2 – Résolution en énergie transverse manquante pour les algorithmes $\cancel{E}_T^{\text{RefFinal}}$, $\cancel{E}_T^{\text{Final}}$ et $\cancel{E}_T^{\text{Uncalib}}$

les distributions de $(\cancel{E}_T - \cancel{E}_T^{\text{True}})$ pour les trois algorithmes ; un ajustement gaussien est effectué pour extraire la résolution dans les trois cas (voir Tab. 5.2). La meilleure résolution est obtenue avec l’algorithme $\cancel{E}_T^{\text{RefFinal}}$. Dans la méthode H1, les poids sont optimisés pour les jets, ce qui résulte en une surestimation de l’énergie transverse manquante pour les événements à faible activité hadronique ($\cancel{E}_T^{\text{Final}} > \cancel{E}_T^{\text{True}}$). Pour les événements $W \rightarrow e\nu$, il est nécessaire de raffiner le calcul de l’énergie des cellules appartenant au cluster de l’électron. Avec des cellules à l’échelle électromagnétique, l’énergie des jets est sous-estimée : l’énergie transverse manquante est sous-estimée ($\cancel{E}_T^{\text{Uncalib}} < \cancel{E}_T^{\text{True}}$). Dans la suite, l’algorithme $\cancel{E}_T^{\text{RefFinal}}$ est utilisé.

5.4.3 Distributions pour le signal et les bruits de fond

Les distributions de \cancel{E}_T pour le signal et les bruits de fond sont représentées sur la Figure 5.15. Elles sont obtenues avec la présélection suivante : un candidat électron tel que $E_T > 25$ GeV, $|\eta| < 1.37$ ou $1.52 < |\eta| < 2.4$ doit être présent dans l’événement. Une coupure à 25 GeV est très discriminante pour les canaux QCD et $Z \rightarrow ee$.

5.5 Niveau de bruit de fond avec la méthode des coupures

Les nombres d’événements attendus pour 50 pb^{-1} de données, après chaque étape de la sélection, sont donnés dans la Table 5.3 : une pureté finale de 81% est obtenue. La Figure 5.16 représente la distribution de masse transverse du W pour le signal et les bruit de fond après sélection complète, normalisée à une luminosité intégrée de 50 pb^{-1} (les histogrammes sont cumulatifs). Une coupure à 40 GeV permet de réduire le bruit de fond QCD d’un facteur 3. Mais après cette coupure, les distributions cinématiques de E_T et \cancel{E}_T ne peuvent plus être utilisées pour extraire la contribution de ce bruit de fond à partir des données (distributions biaisées par la sélection des événements à haute masse transverse).

Les résultats présentés ici ont été obtenus en utilisant une sélection de l’électron non optimale. Deux critères d’identification supplémentaires ont principalement été étudiés : l’utilisation du TRT pour la trace et l’isolation du cluster dans un cône d’ouverture $\Delta R = 0.45$. Il a été montré que ces critères permettent de

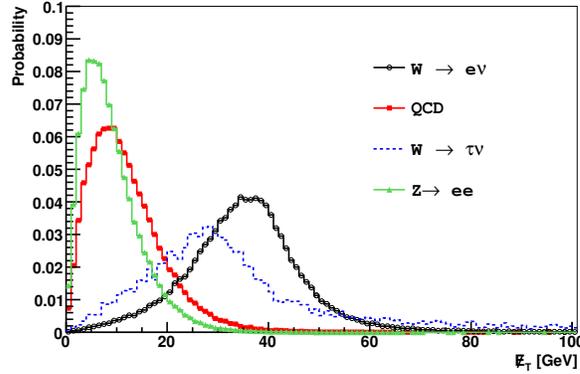


FIG. 5.15 – Distribution d'énergie transverse manquante pour le signal et les bruits de fond

Sélection	$W \rightarrow e\nu$	QCD	$W \rightarrow \tau\nu$	$Z \rightarrow ee$
Trigger	37.01 ± 0.09	835 ± 18	1.73 ± 0.02	6.07 ± 0.01
$E_T > 25$ GeV, condition sur η	30.84 ± 0.09	383 ± 12	1.03 ± 0.01	3.23 ± 0.01
Identification de l'électron	26.77 ± 0.09	110 ± 6	0.91 ± 0.01	2.95 ± 0.01
$\cancel{E}_T > 25$ GeV	22.06 ± 0.09	4.6 ± 0.7	0.55 ± 0.01	0.06 ± 0.01

 TAB. 5.3 – Nombre d'événements attendus ($\times 10^4$) pour le signal et les bruits de fond dans le canal $W \rightarrow e\nu$ après chaque étape de la sélection pour une luminosité intégrée de 50 pb^{-1} . Les incertitudes sont uniquement dues à la statistique des échantillons de données simulées.

réduire le bruit de fond QCD jusqu'à un facteur ~ 10 pour une perte d'efficacité absolue de reconstruction de l'électron de $\sim 13\%$. Ils ne sont pas appliqués dans cette analyse car la statistique générée pour le bruit de fond QCD ne permet pas d'étudier correctement son impact sur la mesure avec une telle sélection ; or étant donné l'incertitude sur la prédiction de ce bruit de fond, un objectif essentiel de l'analyse est de développer des méthodes pour le mesurer à partir des données.

5.6 Conclusion

Dans ce chapitre, les performances attendues du détecteur pour la reconstruction et la sélection des événements $W \rightarrow e\nu$ ont été étudiées. Une sélection simple permet de réduire la contribution du bruit de fond à un niveau inférieur à $\sim 20\%$; une identification plus sévère de l'électron (notamment l'utilisation du TRT) peut permettre de réduire significativement cette contribution.

Les sources d'erreurs systématiques sur la mesure de la section efficace n'ont pas été abordées. Leur étude fait l'objet des trois chapitres suivants : la mesure des efficacités, le contrôle du bruit de fond et la détermination de l'acceptance sont successivement abordés.

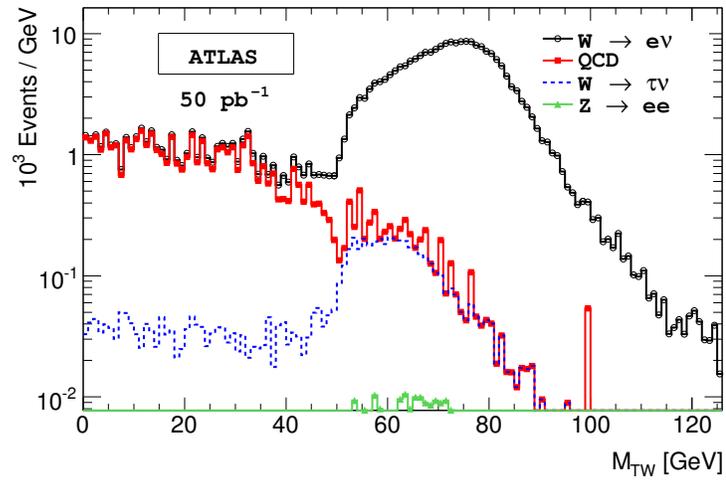


FIG. 5.16 – Distribution de masse transverse du W pour le signal et les bruits de fond après sélection complète, normalisée à une luminosité intégrée de 50 pb^{-1} .

Chapitre 6

Mesure des efficacités à partir des données

La détermination de l'efficacité de reconstruction de l'électron à partir de la méthode tag-and-probe sur les événements $Z \rightarrow ee$ est présentée dans ce chapitre. Les événements $Z \rightarrow ee$ et $W \rightarrow ev$ ont des topologies proches, ce qui permet de transférer l'efficacité d'un type d'événement à l'autre : l'incertitude liée à ce transfert est discutée.

6.1 Méthode tag-and-probe sur les événements $Z \rightarrow ee$

Afin de s'affranchir autant que possible des incertitudes de la simulation sur la mesure, l'efficacité combinée (online et offline) de reconstruction de l'électron doit être déterminée à partir des données. La méthode *tag-and-probe* utilise l'identification des électrons de désintégration du Z . Cette méthode a été utilisée dans les expériences du Tévatron (voir Réf. [71] par exemple). Ce canal est utilisé car les électrons ont une gamme d'énergie similaire à ceux issus de la désintégration du boson W , le nombre d'événements est élevé et le bruit de fond est faible. La méthode a été développée et testée sur données simulées dans ATLAS. L'efficacité devant être déterminée est la combinaison trigger e20-offline medium.

L'échantillon d'événements $Z \rightarrow ee$ est sélectionné de la manière suivante :

- le trigger e20 doit être passé ;
- le cluster EM ayant déclenché le trigger doit satisfaire le critère d'identification offline tight (électron *tag*) ;
- un second cluster EM doit être présent (électron *probe*) ;
- les deux clusters EM doivent satisfaire les conditions : $E_T > 15$ GeV et $|\eta| < 1.37$ ou $1.52 < |\eta| < 2.4$;
- les deux candidats doivent avoir des charges opposées et leur masse invariante doit être dans la gamme de masse $70 < M_{ee} < 110$ GeV.

L'efficacité est alors simplement égale au rapport du nombre d'événements dans lesquels l'électron probe satisfait le critère testé sur le nombre d'événements sélectionnés. Cette méthode peut être appliquée de manière itérative pour déterminer l'efficacité de chaque coupure. Une vue transverse schématique d'un événement $Z \rightarrow ee$, mettant en évidence les électrons tag et probe, est représentée sur la Figure 6.1.

Les performances de la méthode peuvent être testées en comparant les résultats obtenus avec la valeur de l'efficacité calculée à partir de la vérité Monte-Carlo : l'accord obtenu pour l'efficacité combinée est représenté sur la Figure 6.2. A l'intérieur des fluctuations statistiques, aucun biais n'est observé. L'incertitude sur l'efficacité est dominée par l'incertitude statistique due au nombre d'événements sélectionnés ; les deux autres principales sources d'incertitude sont dues au choix des critères de sélection et à la présence du bruit de fond QCD. Pour 50 pb^{-1} , une incertitude relative de 2% peut être atteinte [25].

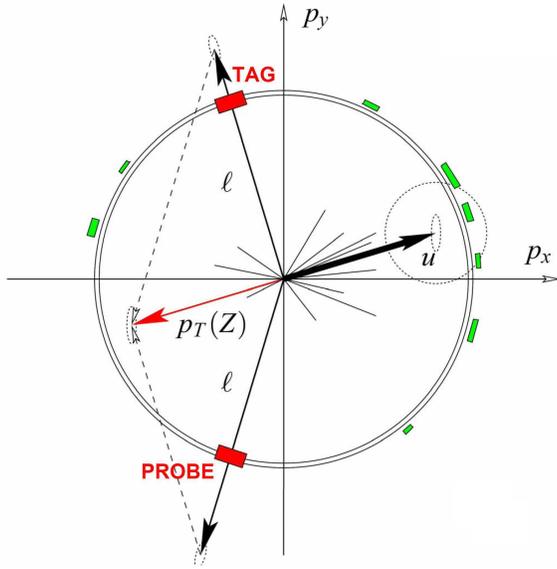


FIG. 6.1 – Vue transverse d'un événement $Z \rightarrow ee$; les électrons isolés tag et probe sont visibles (figure adaptée de la référence [25]).

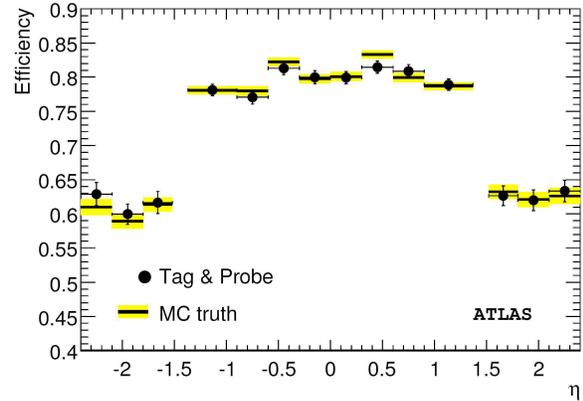


FIG. 6.2 – Comparaison de la mesure de l'efficacité combinée par la méthode tag-and-probe avec la valeur estimée à partir de la vérité Monte-Carlo [25]. Les valeurs sont représentées en fonction de la pseudo-rapidité ; les erreurs sur la mesure sont normalisées à 50 pb^{-1} .

6.2 Transfert vers les événements $W \rightarrow e\nu$

L'efficacité globale de reconstruction de l'électron déterminée à partir des événements $Z \rightarrow ee$ ne peut être appliquée aux événements $W \rightarrow e\nu$ sans correction que si sa valeur est identique dans les deux types d'événement. Dans les données utilisées, les objets trigger (p. ex. l'électron au niveau EF) n'étaient pas accessibles ; seules les signatures trigger (p. ex. e20) étaient disponibles. Le transfert de l'efficacité vers les événements $W \rightarrow e\nu$ n'a donc pu être étudié que pour la reconstruction offline. Un comportement similaire est cependant attendu pour l'efficacité combinée.

D'un point de vue topologique, les électrons issus de la désintégration des bosons W et Z sont similaires : l'activité hadronique est très faible autour de l'électron. Mais les distributions cinématiques sont différentes : les distributions de l'énergie transverse et de la pseudo-rapidité de l'électron dans les deux cas sont représentées sur les Figures 6.3 et 6.4 (80.000 électrons sont utilisés dans chaque cas). Le spectre en E_T est plus dur pour le Z à cause de la masse plus élevée du boson. La Table 6.1 donne la répartition des électrons entre le tonneau et les bouchons : les électrons issus de la désintégration du boson W sont légèrement plus centraux. La Table 6.2 donne l'efficacité dans le tonneau, dans les bouchons et globale (les distributions de l'efficacité en fonction de l'énergie transverse et de la pseudo-rapidité sont représentées sur les Figures 6.5 et 6.6). Comme attendu, l'écart le plus significatif (0.010 ± 0.002) est obtenu pour l'efficacité globale, qui est fonction de la répartition des candidats.

En appliquant l'efficacité globale déterminée sur les événements $Z \rightarrow ee$ aux événements $W \rightarrow e\nu$, le nombre de candidats obtenu est :

$$N_W = \varepsilon_{Z \rightarrow ee} \cdot N_0 = 58231$$

où N_0 est le nombre initial de candidats. En séparant les efficacités pour le tonneau et les bouchons ($\varepsilon_{Z \rightarrow ee}^B$ et $\varepsilon_{Z \rightarrow ee}^{EC}$), le nombre de candidats obtenu est :

$$N_W^{\text{corr}} = \varepsilon_{Z \rightarrow ee}^B \cdot N_0^B + \varepsilon_{Z \rightarrow ee}^{EC} \cdot N_0^{EC} = 58898$$

Or le nombre de candidats attendu d'après la simulation est 59148 : l'introduction des efficacités du tonneau et des bouchons par rapport à l'utilisation de l'efficacité globale permet de réduire le biais sur N_W de -1.5%