



**HAL**  
open science

# Ajustements Bayésiens, application à physique du quark top au LHC

Benoit Clément

► **To cite this version:**

Benoit Clément. Ajustements Bayésiens, application à physique du quark top au LHC. Physique des Hautes Energies - Expérience [hep-ex]. Université de Grenoble, 2012. tel-00713314

**HAL Id: tel-00713314**

**<https://theses.hal.science/tel-00713314v1>**

Submitted on 29 Jun 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Mémoire

présenté pour obtenir le diplôme d'

## HABILITATION À DIRIGER LES RECHERCHES

Spécialité : **Physique**

Présentée par

**Benoit CLEMENT**

Docteur de l'Université Louis Pasteur (Strasbourg I)

Maître de conférence à l'Université Joseph Fourier (Grenoble I)

préparée au **Laboratoire de Physique Subatomique et de Cosmologie**  
Ecole doctorale de physique

## Ajustements bayésiens, application à la physique du quark top au LHC

HDR soutenue publiquement le **22 juin 2012**,  
devant le jury composé de :

**M. Gilles HENRI**

Professeur des Universités, IPAG (Grenoble), Président

**M. Daniel BLOCH**

Directeur de recherche, IPHC (Strasbourg), Rapporteur

**M. Mossadek TALBY**

Professeur des Universités, CPPM (Marseille), Rapporteur

**M. Laurent DEROME**

Maître de conférences, LPSC (Grenoble), Rapporteur

**M. Julien DONINI**

Professeur des Universités, LPC (Clermont-Ferrand), Examineur





# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Analyse statistique et erreurs systématiques</b>	<b>3</b>
1.1 Mesure et incertitude . . . . .	3
1.2 Estimation et fonction de vraisemblance . . . . .	4
1.2.1 Approche fréquentiste : maximum de vraisemblance . . . . .	5
1.2.2 Approche bayésienne : densité <i>a posteriori</i> . . . . .	6
1.3 Incertitudes systématiques . . . . .	7
1.3.1 Approche fréquentiste : vraisemblance profilée . . . . .	8
1.3.2 Approche bayésienne . . . . .	9
1.4 Vraisemblance de Poisson . . . . .	10
1.4.1 Comptage d'événements . . . . .	10
1.4.2 Sources d'erreurs . . . . .	10
1.4.3 Fonction de vraisemblance du problème . . . . .	11
1.4.4 Exemple simple . . . . .	12
<b>2 Implémentation des méthodes d'analyse</b>	<b>15</b>
2.1 Ajustements monodimensionnels . . . . .	15
2.1.1 Vraisemblance profilée . . . . .	15
2.1.2 Intégration bayésienne simple . . . . .	16
2.2 Ajustement de plusieurs paramètres . . . . .	17
2.2.1 Grille et échantillonnage . . . . .	17
2.2.2 Méthode approximée . . . . .	18
2.2.3 Monte-Carlo par chaînes de Markov (MCMC) . . . . .	19
2.2.3.1 Chaînes de Markov . . . . .	19
2.2.3.2 Application au problème étudié . . . . .	20
<b>3 Physique du top au LHC</b>	<b>23</b>
3.1 Le LHC et le détecteur ATLAS . . . . .	23
3.2 Le Modèle Standard et le quark top . . . . .	24
3.2.1 Le modèle standard et ses limites . . . . .	24
3.2.2 Le quark top . . . . .	25
3.3 Production associée $tW$ . . . . .	28
3.3.1 Motivation . . . . .	28
3.3.2 Bruits de fond et sélections . . . . .	29
3.3.3 Estimation des fonds et de l'acceptance . . . . .	31
3.3.4 Erreurs systématiques . . . . .	33
3.3.5 Résultat final . . . . .	34

<b>4</b>	<b>Validation du code Mefisto</b>	<b>37</b>
4.1	Extraction des résultats ATLAS . . . . .	37
4.2	Validation à une dimension . . . . .	39
4.3	Analyse bayésienne multidimensionnelle . . . . .	41
<b>5</b>	<b>Perspectives : analyse combinée des 3 voies <i>single top</i></b>	<b>45</b>
5.1	<i>Single top</i> et nouvelle physique . . . . .	45
5.1.1	Processus <i>benchmark</i> . . . . .	45
5.1.2	Paramétrisation générique . . . . .	47
5.2	Analyses <i>CSC</i> . . . . .	48
5.2.1	Sélections . . . . .	48
5.2.2	Analyse combinée . . . . .	49
5.3	Conclusions . . . . .	54
	<b>Conclusion</b>	<b>57</b>
	<b>Bibliographie</b>	<b>59</b>

# Introduction

Depuis une dizaine d'années, les méthodes statistiques jouent un rôle de plus en plus important en physique des particules. L'amélioration des capacités de calcul des ordinateurs y est pour beaucoup. Elles trouvent de nombreuses applications en théorie (calculs sur réseau, sondage d'espaces de paramètres de modèles, ...), dans la simulation de processus physiques (générateurs Monte-Carlo, simulation de détecteur) et dans l'analyse des données (méthodes multivariées pour la sélection d'événements, intégration de sources d'erreurs systématiques et calcul de limites).

A partir de 2004, les physiciens du Tevatron ont été parmi les premiers à mettre en œuvre, dans le cadre de la recherche des processus de production électrofaible du quark top, une grande variété de méthodes multivariées (arbres de décision, réseaux de neurones) couplées à une analyse bayésienne sophistiquée pour inclure les incertitudes systématiques. Ces analyses ont permis la mise en évidence de cette production électrofaible avec un volume de données bien inférieur à ce qui avait été estimé précédemment.

Les expériences du LHC ont naturellement adopté ces techniques et en proposent de nouvelles. Dans ce mémoire, nous allons nous concentrer sur l'ajustement de paramètres physiques avec estimation de l'incertitude de mesure, soit pour déterminer et optimiser la sensibilité d'une analyse, soit pour extraire le résultat final. Les différentes méthodes discutées, en particulier une approche originale d'intégration bayésienne par chaînes de Markov, font l'objet d'une implémentation dans un programme. Nous utiliserons alors les recherches de production électrofaible du quark top dans les premières données du LHC pour illustrer et valider ces différents outils. Enfin, le dernier chapitre prospectif sera consacré à l'utilisation future de ces mêmes outils pour contraindre des modèles de physique au-delà du modèle standard.



Μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος  
οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε,  
πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν  
ἡρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν  
οἰωνοῖσί τε πᾶσι, Διὸς δ' ἐτελείετο βουλή,  
ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε  
Ἄτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.

Homère, "Iliade" I.1



# Analyse statistique et erreurs systématiques

## 1.1 Mesure et incertitude

Le résultat d'une mesure, qu'il soit obtenu directement d'une expérience ou dérivé d'un ou plusieurs résultats expérimentaux, consiste en une ou plusieurs valeurs numériques qui constitue la **valeur centrale** de la mesure utilisée pour estimer la valeur « vraie » d'une grandeur physique.

L'une des principales problématiques concerne alors la confiance ou la crédibilité que l'on accorde à ce résultat. Pour une valeur numérique unique, la présentation la plus classique pour cette information est la barre d'erreur ou d'incertitude <sup>1</sup> de la forme :

$$x = X \pm \Delta X \text{ où } x = X_{-\Delta X}^{+\Delta X} \quad (1.1)$$

L'incertitude définit donc un intervalle contenant *probablement* la valeur vraie de la grandeur mesurée. Dans cette représentation la barre d'erreur contient la valeur centrale. Dans le premier cas la barre d'erreur est symétrique autour de cette valeur centrale, alors que le second traduit une plus forte probabilité de fluctuation d'un côté que de l'autre de la valeur centrale. Enfin dans certains cas, on souhaite séparer certaines contributions à l'incertitude, par exemple les effets statistiques et systématiques. Cette séparation est particulièrement pertinente pour discuter des possibilités d'amélioration du résultat par de futures expériences (réduction de l'erreur statistique) ou de nouvelles analyses (réductions des effets systématiques). Par exemple, les effets statistiques se réduisent par une accumulation plus grande de données avec, si l'appareillage reste identique, une évolution de l'erreur en  $\sim 1/\sqrt{t}$ ,  $t$  étant la durée de la mesure. La réduction d'effets systématiques passe par une meilleure compréhension de l'appareillage de mesure (qui peut nécessiter une statistique accrue) et/ou un changement de techniques de mesure et d'analyse. On

---

<sup>1</sup>Certains associent la notion d'erreur à celle de biais, ce qui sous-entend une possibilité de correction, alors que l'incertitude est purement aléatoire. Dans les différentes situations que nous allons parcourir, les erreurs ne seront pas estimables autrement que par un intervalle de confiance et seront traitées comme des fluctuations aléatoires : si aucune valeur de l'erreur ne peut être précisément estimée alors toutes les valeurs possibles (*i.e.* dans la barre d'erreur) doivent être considérées. Ainsi les termes d'erreur ou d'incertitude seront utilisés indifféremment pour désigner le même concept.



aura alors :

$$x = X \pm \Delta X_{stat} \pm \Delta X_{syst} \quad (1.2)$$

En reliant l'incertitude à une mesure de dispersion proportionnelle à la variance, l'erreur totale s'obtient en sommant les différentes composantes (supposées indépendantes) en quadrature.

L'intervalle de confiance permet de donner une définition plus rigoureuse du concept d'incertitude, le niveau de confiance quantifiant la probabilité de trouver la valeur vraie dans la barre d'erreur. Un intervalle de confiance est un intervalle  $[a, b]$  tel que :

$$P(a < X < b) = \int_a^b f_X(x) dx = \alpha, \quad (1.3)$$

$\alpha$  étant le niveau de confiance de l'intervalle. L'écart-type  $\sigma$  permet de définir un intervalle de confiance autour de la valeur moyenne  $\mu$ , de la forme  $[\mu - \sigma; \mu + \sigma]$ . Dans le cas d'une densité normale, cet intervalle correspond à un niveau de confiance de 68.3%. Par extension, c'est la valeur du niveau de confiance que l'on associe généralement (mais pas nécessairement) à une barre d'erreur.

## 1.2 Estimation et fonction de vraisemblance

Une mesure brute se compose d'un échantillon d'une ou plusieurs valeurs numériques  $\vec{m} = \{m_1, m_2, \dots, m_n\}$ , réalisation d'une variable aléatoire  $\vec{X} = \{X_1, X_2, \dots, X_n\}$ . L'ensemble de ces variables aléatoires est décrit par une densité de probabilité  $f_{\vec{X}}(\vec{x})$  où  $\vec{x}$  désigne une réalisation quelconque.

Par ailleurs on va utiliser ces données pour déterminer ou contraindre un ou plusieurs paramètres physiques  $\vec{\theta}$  (masses, section efficace, ...) que l'on souhaite estimer, soit avec une valeur numérique, soit par un intervalle de confiance ou une limite. Les données mesurées sont sensibles aux paramètres  $\vec{\theta}$ , si le résultat attendu de la mesure est différent pour des valeurs différentes de  $\vec{\theta}$ . Ainsi, la densité  $f_{\vec{X}}(\vec{x})$  doit dépendre des paramètres  $\vec{\theta}$ . D'un point de vue purement mathématique, on peut considérer une fonction générale dont les variables seraient à la fois les variables aléatoires et les paramètres à estimer. Notons cette fonction  $k(\vec{x}, \vec{\theta})$ . A partir de cette fonction on construit 2 objets :

- La **densité de probabilité** de  $\vec{X}$ , qui est une fonction du seul vecteur  $\vec{x}$ , réalisation de  $\vec{X}$ , les paramètres  $\vec{\theta}$  étant fixés à une valeur  $\vec{\theta} = \vec{\theta}_0$  unique. Cette valeur  $\vec{\theta}_0$  est la valeur vraie des paramètres :

$$f_{\vec{X}}(\vec{x}) = k(\vec{x}, \vec{\theta}_0) \text{ et } \int f_{\vec{X}}(\vec{x}) d\vec{x} = 1. \quad (1.4)$$

- La **fonction de vraisemblance** des paramètres  $\vec{\theta}$  où on ne considère que la dépendance en  $\vec{\theta}$ , en fixant  $\vec{x}$  à la mesure  $\vec{m}$  :

$$\mathcal{L}(\vec{\theta}) = k(\vec{m}, \vec{\theta}). \quad (1.5)$$

La différence fondamentale entre les approches fréquentiste et bayésienne, que nous allons discuter dans les prochains paragraphes, tient dans le statut donné aux paramètres  $\vec{\theta}$ . Dans l'approche fréquentiste,  $\vec{\theta}$  possède une unique valeur, que l'on tente d'estimer au mieux : les données permettent de construire des estimateurs de la valeur centrale et des intervalles de confiance possédant une probabilité fixée de contenir la vraie valeur.

Dans l'approche bayésienne, les paramètres sont eux-mêmes traités comme des variables aléatoires, décrites par une densité de probabilité : au vu des données, on accorde une plus grande crédibilité à certaines valeurs des paramètres qu'à d'autres, c'est la densité de probabilité *a posteriori*. Cette densité des paramètres permet à son tour de définir une valeur centrale et des intervalles de confiance.

### 1.2.1 Approche fréquentiste : maximum de vraisemblance

La fonction de vraisemblance n'est pas une densité de probabilité. Elle mesure néanmoins la crédibilité associée à chaque valeur possible des paramètres.  $\mathcal{L}(\vec{\theta}_0)$  quantifie donc la compatibilité des données avec l'hypothèse  $\vec{\theta} = \vec{\theta}_0$ . L'estimateur naturel qui en découle pour  $\vec{\theta}$  est la valeur  $\hat{\vec{\theta}}$  telle que la fonction de vraisemblance soit maximale : c'est l'estimateur du maximum de vraisemblance.

Un estimateur  $\hat{\vec{\theta}}$  (par exemple le maximum de vraisemblance) est une variable aléatoire. Dans le cas d'un unique paramètre, un intervalle défini par deux fonctions de  $\hat{\theta}$ ,  $[\xi(\hat{\theta}), \omega(\hat{\theta})]$ , est également une variable aléatoire. Si on répète l'expérience plusieurs fois, on va obtenir différentes valeurs de l'estimateur et de l'intervalle, qui peut contenir ou non la valeur vraie  $\theta_0$  : on peut donc définir au sens fréquentiste la probabilité que l'intervalle contienne la valeur vraie. L'intervalle  $[\xi(\hat{\theta}), \omega(\hat{\theta})]$  est un intervalle de confiance de niveau  $\alpha$  si :

$$P\left(\theta_0 \in [\xi(\hat{\theta}), \omega(\hat{\theta})]\right) = \alpha. \quad (1.6)$$

quelle que soit la valeur de  $\theta$ . La construction exacte de l'intervalle de confiance, dite construction de Neyman [1], peut se révéler très coûteuse en temps de calcul.

On a le plus souvent recours à des propriétés asymptotiques (valables rigoureusement dans la limite d'un échantillon infini) de l'estimateur de maximum de vraisemblance. La loi de probabilité de l'estimateur converge vers la loi multinormale. L'estimateur est non biaisé asymptotiquement, la moyenne de la loi multinormale correspond donc à la valeur vraie  $\vec{\theta}_0$ . Le maximum de vraisemblance est également de variance minimale. Il atteint asymptotiquement la borne de Rao-Cramer-Fréchet et sa matrice de covariance est alors :

$$\Sigma = L^{-1} \text{ avec } L_{ij} = -E \left[ \frac{\partial^2 \ln \mathcal{L}(\vec{\theta})}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta}=\hat{\vec{\theta}}} \right], \quad (1.7)$$

la valeur moyenne portant sur la variable aléatoire  $\vec{X}$ . Dans la pratique, on ne dispose que d'une réalisation de  $\mathcal{L}$ . Finalement on construit un estimateur de la covariance par :

$$\hat{\Sigma} = \hat{L}^{-1} \text{ avec } L_{ij} = - \frac{\partial^2 \ln \mathcal{L}(\vec{\theta})}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta}=\hat{\vec{\theta}}} \quad (1.8)$$

On peut déterminer plus correctement la matrice de covariance par des pseudo-expériences Monte-Carlo.

On utilise enfin le rapport de vraisemblance  $\lambda = \mathcal{L}(\vec{\theta})/\mathcal{L}(\hat{\vec{\theta}})$  dont le logarithme suit (toujours asymptotiquement) une loi de  $\chi^2$  à  $n_\theta$  (nombre de paramètres) degrés de libertés pour construire des intervalles de confiance. Les bornes d'un intervalle pour un niveau de confiance  $\alpha$  satisfont l'équation :

$$-2 \ln \lambda(\vec{\theta}) = \beta(n_\theta, \alpha) \text{ où } \beta \text{ est défini par } \alpha = \int_0^{\beta(n_\theta, \alpha)} f_{\chi^2}(x; n_\theta) dx. \quad (1.9)$$

Dans le cas d'un paramètre unique, cette équation se réduit à :

$$-\ln \lambda(\theta) = \frac{n^2}{2}, \quad (1.10)$$

qui définit les bornes d'un intervalle à  $n\sigma$  (*i.e.* dont le niveau de confiance est égal à celui de l'intervalle  $\mu \pm n\sigma$  d'une loi normale, soit 68.3% pour  $n = 1$ , 95.4% pour  $n = 2$ , ...).

### 1.2.2 Approche bayésienne : densité *a posteriori*

L'approche bayésienne est a priori plus simple. La probabilité est simplement interprétée comme un degré de confiance. On peut alors associer une densité de probabilité aux paramètres  $\vec{\theta}$ . Après un ensemble de mesures  $\vec{x}$ , la densité de probabilité des paramètres du modèle contrainte par ces mesures est la densité conditionnelle  $f(\vec{\theta}|\vec{x})$ , appelée densité *a posteriori*. En utilisant le théorème de Bayes, on a :

$$f(\vec{\theta}|\vec{x}) = \frac{f(\vec{x}|\vec{\theta})f(\vec{\theta})}{f(\vec{x})} \text{ soit } f(\text{modèle}|\text{mesure}) = \frac{f(\text{mesure}|\text{modèle})f(\text{modèle})}{f(\text{mesure})}. \quad (1.11)$$

Cette expression fait apparaître trois facteurs :

- $f(\vec{x}|\vec{\theta})$  est la probabilité des mesures pour une valeur déterminée de  $\vec{\theta}$ . C'est la fonction de vraisemblance  $\mathcal{L}(\vec{\theta})$  définie précédemment.
- $f(\vec{\theta}) = \pi(\vec{\theta})$  est la densité de probabilité des paramètres avant la mesure nommée densité *a priori* des paramètres. Cette densité est souvent fixée arbitrairement.
- $f(\vec{x})$  est la densité probabilité de la mesure  $\vec{x}$ . C'est une grandeur numérique constante qui joue un simple rôle de normalisation de la densité *a posteriori*.

La probabilité *a posteriori* peut se réécrire :

$$f(\vec{\theta}|\vec{x}) = \frac{\mathcal{L}(\vec{\theta})\pi(\vec{\theta})}{\int f(\vec{x}|\vec{\theta})\pi(\vec{\theta})d\vec{\theta}} \quad (1.12)$$

Le choix le plus fréquent pour la densité *a priori* est une densité uniforme sur un intervalle  $[a, b]$ . Cette densité traduit l'absence d'*a priori* sur la valeur des paramètres. Si la fonction de vraisemblance tombe généralement à 0 pour des valeurs des paramètres loin du maximum de vraisemblance, on peut choisir  $[a, b]$  tel que :

$$f(\vec{\theta}|\vec{x}) \propto \mathcal{L}(\vec{\theta}). \quad (1.13)$$

Si ce n'est pas le cas, cela indique que la mesure n'est pas ou peu sensible au paramètre estimé.

Cette densité *a priori* permet également de borner les paramètres en imposant un *a priori* nul dans les régions non-physiques. Dans le cas d'une mesure de section efficace, on coupera la partie négative de la fonction de vraisemblance.

Si la partie de fonction de vraisemblance coupée par l'*a priori* n'est pas négligeable, l'approximation précédente devient fautive et il faut recalculer la normalisation de la densité *a posteriori*. Un autre choix pour la densité *a priori* serait de prendre la densité du paramètre estimée par une mesure précédente. Le choix arbitraire de cette densité est le point faible de cette approche. En particulier, le résultat final ne sera pas invariant par changement de variable : choisir  $\theta$  ou  $\log(\theta)$  uniforme ne conduira pas à la même densité *a posteriori*, mais il n'y a pas forcément d'arguments objectifs en faveur de l'un ou l'autre

de ces choix. Le choix de la densité *a priori* reste subjectif et pour plusieurs *a priori* raisonnables (encore de la subjectivité!), le résultat *a posteriori* doit être similaire.

Le concept de l'analyse bayésienne est donc de partir d'une connaissance *a priori* des paramètres du modèle et de construire la densité *a posteriori* incluant l'information apporté par la mesure. La mesure favorise certaines valeurs des paramètres, la plus probable étant le maximum de la fonction de vraisemblance si celle-ci n'est pas tronquée par la densité *a priori*. La mesure modifie donc le degré de crédibilité associée aux valeurs possibles des paramètres (densité *a posteriori*). Contrairement au cas fréquentiste, il n'y a plus d'estimateur et on obtient directement une densité de probabilité pour les paramètres, à partir de laquelle on peut définir une valeur centrale (moyenne, médiane ou plus souvent le mode qui correspond au maximum de vraisemblance pour un *a priori* uniforme) et une barre d'erreur via un intervalle de confiance  $[a, b]$  pour un niveau de confiance  $\alpha$  tel que :

$$\int_a^b f(\vec{\theta}|\vec{x})d\vec{\theta} = \alpha \quad (1.14)$$

Pour définir l'intervalle de manière unique, on doit fixer des critères supplémentaires. A une dimension on peut choisir par exemple :

- un intervalle symétrique autour de la valeur centrale  $c : [c - k, c + k]$ .
  - un intervalle centré en probabilité autour de la valeur centrale  $c : [a, b]$  tel que  $\int_a^c f(\theta|\vec{x})d\theta = \int_c^b f(\theta|\vec{x})d\theta = \alpha/2$ .
  - un intervalle  $[a, b]$  tel que  $f(a|\vec{x}) = f(b|\vec{x})$  et  $f(u|\vec{x}) > f(b|\vec{x})$  pour tout  $u \in [a, b]$ .
- Cette définition correspond à l'intervalle HPD (*Highest Probability Density*) et correspond à l'intervalle fréquentiste défini par le rapport de vraisemblance. Ce type d'intervalle se généralise simplement à des contours multidimensionnels et est indépendant du choix d'une valeur centrale.

Ces trois type d'intervalle sont illustrés sur la figure 1.1.

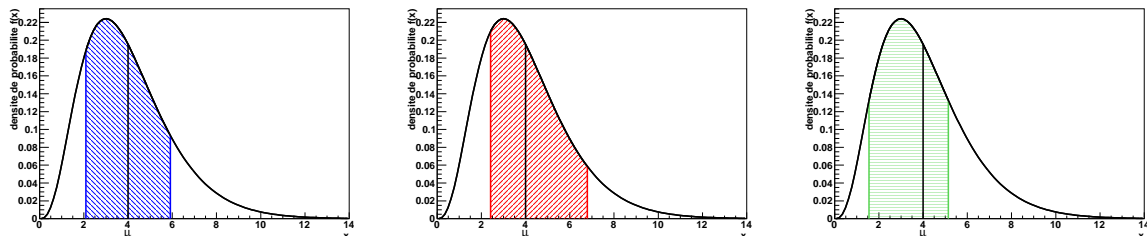


FIG. 1.1: Exemples d'intervalle de confiance à 68% : centré sur la moyenne (à gauche), centré sur la moyenne en probabilité (au centre), intervalle HPD (à droite).

### 1.3 Incertitudes systématiques

La principale difficulté, que nous allons discuter maintenant, est l'inclusion d'incertitudes systématiques dans l'estimation par intervalle. Ce point est relativement simple dans le cas bayésien mais peut se révéler très délicat dans le contexte fréquentiste. Il serait même possible d'argumenter que le concept même d'incertitude systématique n'a pas de sens dans une interprétation purement fréquentiste. Nous nous contenterons de décrire une approche possible parmi les différentes solutions de ce problème qui ont pu être proposées.

Dans les deux cas, le problème à traiter est le suivant : parmi les paramètres du modèle, on peut distinguer les paramètres  $\vec{\theta}$  que l'on souhaite réellement contraindre et ceux  $\vec{\nu}$  dont on connaît la valeur  $\vec{\nu}_{est}$  avec une incertitude  $\Delta\vec{\nu}_{est}$  (cette notation pouvant correspondre à une matrice de covariance, par exemple). Ces derniers constituent les paramètres de nuisance.

### 1.3.1 Approche fréquentiste : vraisemblance profilée

L'écart entre la valeur connue et la valeur vraie d'un paramètre de nuisance représente un biais, que l'on doit pouvoir estimer. La valeur connue des paramètres de nuisance représente ainsi une réalisation d'un estimateur, et l'incertitude sur le paramètre s'interprète comme sa variance. À partir de la densité de probabilité de cet estimateur, on associe une fonction de vraisemblance  $\mathcal{L}_{nuisance}(\vec{\nu}; \vec{\nu}_{est}; \Delta\vec{\nu}_{est})$  aux paramètres de nuisance. La fonction de vraisemblance totale du problème s'écrit alors :

$$\mathcal{L}(\vec{\theta}, \vec{\nu}) = \mathcal{L}_{modèle}(\vec{\theta}, \vec{\nu}) \times \mathcal{L}_{nuisance}(\vec{\nu}) \quad (1.15)$$

où  $\mathcal{L}_{modèle}$  est la vraisemblance du modèle seul. Si les paramètres  $\vec{\nu}$  sont parfaitement connus,  $\mathcal{L}_{nuisance}(\vec{\nu}) = \delta(\vec{\nu} - \vec{\nu}_{est})$  et on se retrouve dans le cas discuté précédemment. À partir de cette fonction de vraisemblance, on construit [2] la fonction de vraisemblance profilée (*profile likelihood*) :

$$\mathcal{L}_{pr}(\vec{\theta}) = \max_{\vec{\nu}} \mathcal{L}(\vec{\theta}, \vec{\nu}) \quad (1.16)$$

en maximisant la fonction de vraisemblance relativement aux paramètres  $\vec{\nu}$  pour chaque valeur des paramètres  $\vec{\theta}$ . L'estimateur paramètres de nuisance, à  $\theta$  fixé, est souvent noté  $\hat{\vec{\nu}}$ . Le rapport :

$$\lambda(\vec{\theta}) = \mathcal{L}_{pr}(\vec{\theta}) / \mathcal{L}(\hat{\vec{\theta}}), \quad (1.17)$$

où  $\hat{\vec{\theta}}$  est l'estimateur de maximum de vraisemblance, possède les mêmes propriétés statistiques que le rapport de vraisemblance classique [3]. La construction de cette fonction de vraisemblance est illustrée sur la figure 1.2 :

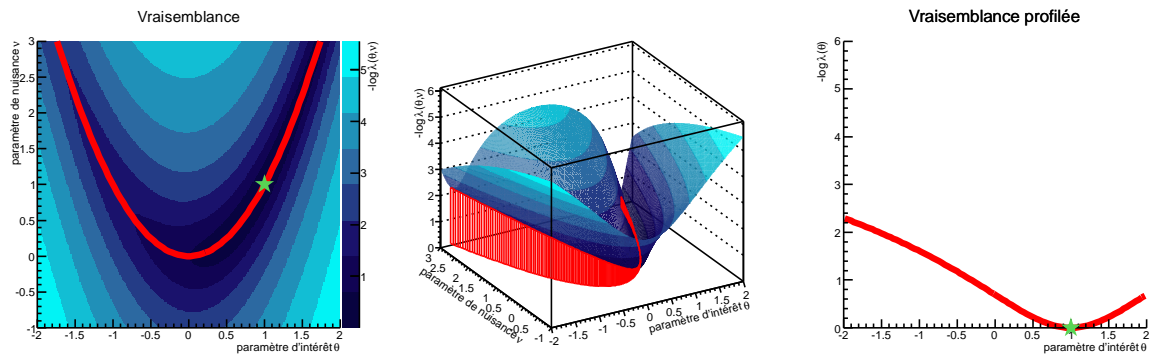


FIG. 1.2: Description schématique de la vraisemblance profilée.

À gauche : logarithme de la fonction de vraisemblance dans le plan  $(\theta, \nu)$  ; la ligne rouge indique le minimum relatif à  $\nu$  à  $\theta$  fixé.

Au centre : logarithme de la fonction de vraisemblance en fonction des paramètres  $(\theta, \nu)$ .

À droite : logarithme de la fonction de vraisemblance profilée en fonction du paramètre  $\theta$ .

Il est important de noter que, si le nombre de paramètres de nuisance excède le nombre de degrés de liberté du problème, il existe une dégénérescence sur les valeurs  $\hat{\vec{\nu}}(\vec{\theta})$  qui maximise  $\mathcal{L}(\vec{\theta}, \vec{\nu})$ . Dans ce cas, on détermine la valeur de la fonction de vraisemblance profilée sans réellement contraindre les paramètres de nuisance : on ne dispose pas de mesure du biais.

### 1.3.2 Approche bayésienne

La méthode bayésienne permet d'inclure les effets d'incertitudes systématiques via des paramètres de nuisance. En séparant explicitement les paramètres à ajuster et les paramètres de nuisance, la densité *a posteriori* peut s'écrire :

$$f(\vec{\theta}, \vec{\nu} | \vec{x}) = \frac{f(\vec{x} | \vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu})}{\iint f(\vec{x} | \vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu}) d\vec{\theta} d\vec{\nu}} = \frac{\mathcal{L}(\vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu})}{\iint \mathcal{L}(\vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu}) d\vec{\theta} d\vec{\nu}} \quad (1.18)$$

et la densité *a posteriori* des paramètres  $\vec{\theta}$  n'est que la densité marginale obtenue en intégrant sur les paramètres de nuisance  $\vec{\nu}$  :

$$f(\vec{\theta} | \vec{x}) = \frac{\int \mathcal{L}(\vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu}) d\vec{\nu}}{\iint \mathcal{L}(\vec{\theta}, \vec{\nu}) f(\vec{\theta}, \vec{\nu}) d\vec{\theta} d\vec{\nu}} \quad (1.19)$$

Généralement, les incertitudes sur les paramètres de nuisance sont indépendantes de la densité *a priori* des paramètres  $\vec{\theta}$ . La densité *a priori* est factorisable sous la forme  $f(\vec{\theta}, \vec{\nu}) = \pi(\vec{\theta})k(\vec{\nu})$  :  $\pi(\vec{\theta})$  est la densité *a priori* pour les paramètres  $\vec{\theta}$  et  $k(\vec{\nu})$  celle des paramètres de nuisance. Cette densité *a priori* des paramètres de nuisance est identique à la fonction de vraisemblance  $\mathcal{L}_{nuisance}$  introduite dans la construction de la fonction de vraisemblance profilée. On pourra par exemple choisir pour  $k(\vec{\nu})$  une densité multinormale définie par les valeurs attendues et la matrice de covariance des paramètres de nuisance<sup>2</sup>.

Cette factorisation permet également de sortir la densité *a priori* des paramètres de l'intégrale. Concrètement, ceci reviendra à intégrer la fonction de vraisemblance puis à multiplier par la densité *a priori*, la première étape étant généralement la plus coûteuse en temps de calcul (intégration numérique). Finalement on obtient :

$$f(\vec{\theta} | \vec{x}) = \frac{\pi(\vec{\theta}) \int \mathcal{L}(\vec{\theta}, \vec{\nu}) k(\vec{\nu}) d\vec{\nu}}{\iint \mathcal{L}(\vec{\theta}, \vec{\nu}) \pi(\vec{\theta}) k(\vec{\nu}) d\vec{\theta} d\vec{\nu}} \quad (1.20)$$

Dans la pratique, la densité *a priori* des paramètres de nuisance n'est connue qu'approximativement. Si ces paramètres sont correctement estimés, ils ne doivent que peu affecter la qualité de l'ajustement. Cette qualité peut se quantifier par la probabilité d'obtenir la valeur la plus probable du paramètre, soit la valeur maximale de la densité *a posteriori*. On peut donc valider la valeur des paramètres de nuisance en étudiant la dispersion relative de la valeur maximale de la densité *a posteriori* pour différentes valeurs des paramètres de nuisance. La construction d'un intervalle de confiance incluant les effets systématiques se fait comme précédemment en intégrant la densité *a posteriori* :

$$\int_a^b f(\vec{\theta} | \vec{x}) d\vec{\theta} = \alpha. \quad (1.21)$$

<sup>2</sup>Un paramètre connu exactement aura pour *a priori* une distribution de Dirac et l'intégrale sur ce paramètre conduit à remplacer la variable aléatoire correspondante par sa valeur.

L'intégration sur les paramètres de nuisance et le calcul de la normalisation de la densité *a posteriori* sont également coûteux en temps de calcul et nécessitent la mise en place de méthodes numériques appropriées qui seront discutées au chapitre 2.

## 1.4 Vraisemblance de Poisson

### 1.4.1 Comptage d'événements

Une expérience de physique des particules se ramène généralement à un comptage d'événements passant des coupures de sélection. On souhaite ensuite confronter le nombre d'événements observé à un modèle permettant de calculer ce nombre d'événements. Le nombre d'événements observé est la réalisation d'une variable aléatoire distribuée selon une loi de Poisson, dont le paramètre est le nombre d'événements attendu moyen. Pour améliorer la sensibilité d'une analyse, on va combiner plusieurs mesures correspondant à des sélections différentes, voire à un découpage selon une variable spécifique (classes d'un histogramme). Chaque ensemble d'événements ainsi défini forme un *canal d'analyse*. Le modèle permet de calculer le nombre d'événements moyen attendu dans chaque canal, en fonction des paramètres physique à mesurer (section efficace, couplage, ...) et de différents paramètres connus avec une précision finie (paramètres de nuisance).

Dans la suite nous adopterons les notations suivantes :

- $\vec{\theta}$ , les  $n_\theta$  paramètres à ajuster.
- $\vec{\nu}$ , les  $n_\nu$  paramètres de nuisance. Les paramètres de nuisance pouvant avoir des valeurs différentes dans les différents canaux, on aura  $n_\nu = kn_c$  où  $n_c$  est le nombre de canaux d'analyse et  $k$  le nombre de paramètres de nuisance nécessaire à la description d'un unique canal.
- $\vec{M}(\vec{\theta}, \vec{\nu})$  : le nombre d'événements attendu moyen dans chaque canal. C'est un vecteur de dimension  $n_c$ .
- $\vec{D}$  : le nombre d'événements observés dans chaque canal (vecteur de dimension  $n_c$  également).

### 1.4.2 Sources d'erreurs

Pour prendre en compte les paramètres de nuisance, il faut formuler analytiquement la fonction de vraisemblance de ces paramètres.

Dans le cas général, il va exister de multiples corrélations possibles entre les paramètres qui peuvent mener à une densité difficilement formulable. En effet l'incertitude globale peut se composer de plusieurs *sources d'incertitude* indépendantes les unes des autres, mais qui peuvent affecter simultanément plusieurs paramètres de nuisance. On pourra donc ne considérer que des sources d'erreur caractérisées par :

- les contributions d'une même source, entièrement corrélées (ou anticorrélées) entre les différents paramètres et d'un canal à l'autre.
- les contributions de deux sources d'incertitude différentes, entièrement décorréliées.

Les paramètres de nuisance seront caractérisés par une valeur centrale connue  $\vec{\nu}_0$  et deux matrices d'erreurs  $\Delta^+$  et  $\Delta^-$  comme la matrice des fluctuations à " $\pm 1\sigma$ ". Ce sont deux matrices  $n_\nu \times n_s$ ,  $n_s$  étant le nombre de sources d'incertitude considérées.  $\Delta_{ij}^+$  est la fluctuation à  $1\sigma$  du paramètre de nuisance  $\nu_i$  produit par la source d'erreur  $j$ . Les termes de  $\Delta^\pm$  peuvent être négatifs pour traduire des anticorrélations entre incertitudes. L'existence de deux matrices permet de traiter des erreurs asymétriques.

La corrélation entre toutes les fluctuations liées à une source d'erreur permet de décrire l'ensemble de ces fluctuations avec un unique paramètre. Il suffit dès lors de redéfinir les paramètres de nuisance : au lieu de traiter  $\vec{\nu}$  comme des variables aléatoires, on considère un vecteur  $\vec{\Omega}$ , de dimension  $n_s$ . En général  $\vec{\Omega}$  sera distribué selon une loi multinormale de moyenne  $\vec{\mu} = \vec{0}$  et matrice de covariance  $\Sigma = \mathbf{1}$ .

Pour une réalisation  $\vec{\omega}$  de  $\vec{\Omega}$ , les paramètres de nuisance peuvent alors s'écrire :

$$\vec{\nu} = \vec{\nu}_0 + \vec{f}(\vec{\omega}, \Delta^+, \Delta^-) \quad (1.22)$$

Plusieurs prescriptions sont possibles pour définir la fonction de fluctuation  $\vec{f}$ . Le traitement correct des corrélations totales impose que la valeur centrale  $\vec{\nu}_0$  des paramètres corresponde à la médiane de la densité de probabilité associée, c'est-à-dire que la probabilité de fluctuer vers le haut ou vers le bas soit la même. Ceci impose comme contrainte que les composante  $\omega_i$  et  $f_i$  aient le même signe. Si une source d'erreur n'affecte qu'un unique paramètre, la question des corrélations ne se pose plus et le choix de la densité est plus libre.

La transformation  $\vec{f}(\vec{\omega}, \Delta^+, \Delta^-)$  la plus courante sera de la forme :

$$\vec{f}(\vec{\omega}, \Delta^+, \Delta^-) = \vec{\omega}^\top \Delta \text{ avec } \Delta_{ij} = \Delta_{ij}^{\text{sgn}(\omega_i)} = \begin{cases} \Delta_{ij}^+ & \text{si } \omega_i > 0 \\ \Delta_{ij}^- & \text{si } \omega_i < 0 \end{cases} \quad (1.23)$$

En prenant une erreur moyenne  $\Delta_m = \frac{1}{2}(\Delta^+ + \Delta^-)$ , on peut également symétriser le problème et :

$$\vec{f}(\vec{\omega}, \Delta^+, \Delta^-) = \vec{\omega}^\top \Delta_m \quad (1.24)$$

Ceci correspond à deux cas extrêmes : le premier présente une discontinuité dans la densité du paramètre au niveau de la valeur centrale, le second gomme les effets d'asymétrie qu'il peut être important de prendre en considération. Il s'agit de deux cas limites d'une classe de transformation de la forme :

$$\vec{f}(\vec{\omega}, \Delta^+, \Delta^-) = \vec{\omega}^\top \times [\Delta^- + (\Delta^+ - \Delta^-) U(\vec{\omega})] \quad (1.25)$$

avec  $U$  une matrice diagonale  $n_s \times n_s$  telle que  $U_{ii} = u(\omega_i)$  où  $u(r)$  désigne n'importe quelle fonction échelon, c'est à dire une fonction monotone, croissante, telle que :

$$u(0) = \frac{1}{2}, \quad \lim_{u \rightarrow -\infty} = 0, \quad \lim_{u \rightarrow +\infty} = 1. \quad (1.26)$$

Plusieurs familles de fonctions comme :

- Arc tangente :  $\frac{2 \tan^{-1}(\kappa r) + \pi}{2\pi}$ ,
- Fonction d'erreur :  $\frac{\text{erf}(\kappa r) + 1}{2}$ ,
- Sigmoides :  $\frac{1}{1 + e^{-\kappa r}}$ ,

où le paramètre  $\kappa$  est libre, satisfont ces conditions. L'asymétrie est maximale pour  $\kappa \rightarrow +\infty$  alors que  $\kappa \rightarrow 0$  correspond à l'erreur moyenne. Le lecteur souhaitant approfondir ces questions pourra consulter la référence [4]. L'effet d'une telle fonction est illustré sur la figure 1.3.

### 1.4.3 Fonction de vraisemblance du problème

Pour une expérience de comptage, le nombre d'événements  $D_i$  dans le canal  $i$  est la réalisation d'une variable aléatoire qui suit une loi de Poisson de moyenne  $M_i(\vec{\theta}, \vec{\nu})$ . Si



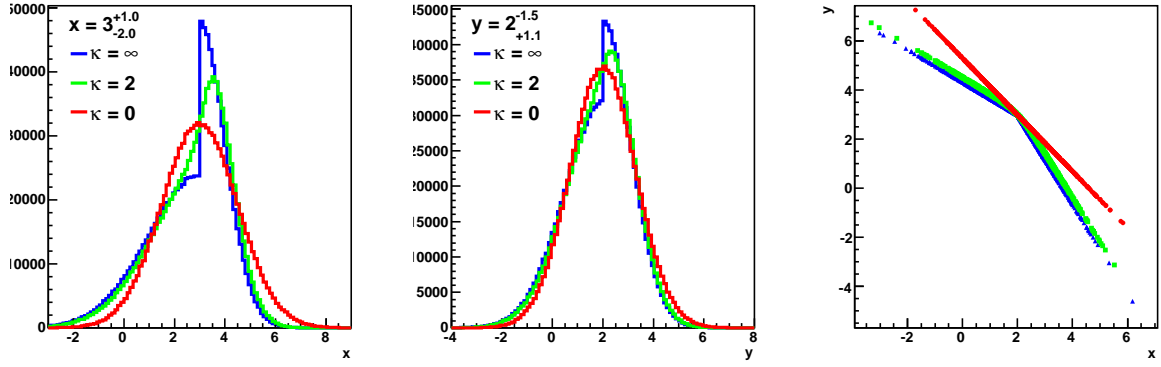


FIG. 1.3: Distributions obtenues par fluctuation aléatoire d'un couple de paramètres  $\begin{pmatrix} x \\ y \end{pmatrix}$ , avec une source de nuisance asymétrique  $\Delta^+ = \begin{pmatrix} 1 \\ -1.5 \end{pmatrix}$  et  $\Delta^- = \begin{pmatrix} 2 \\ -1.1 \end{pmatrix}$  pour trois valeurs de  $\kappa$  : les deux cas limites ( $\kappa = 0$  en rouge et  $\kappa = +\infty$  en bleu) et un cas intermédiaire ( $\kappa = 2$ , en vert). Les figure de droite représente  $y$  en fonction de  $x$  et montre l'anti-corrélation entre les fluctuations.

les canaux sont statistiquement indépendants, c'est-à-dire qu'un événement ne contribue qu'à un unique canal, la fonction de vraisemblance des paramètres seuls est :

$$\mathcal{L}(\vec{\theta}, \vec{\nu}) = \prod_{c=1}^{n_c} \frac{e^{-M_c(\vec{\theta}, \vec{\nu})} M_c(\vec{\theta}, \vec{\nu})^{D_c}}{D_c!} \quad (1.27)$$

En ajoutant les paramètres de nuisance, que ce soit pour un traitement fréquentiste ou bayésien, la fonction de vraisemblance doit être multipliée par la fonction de vraisemblance (ou la densité *a priori*) des paramètres de nuisance, décrits par les paramètres  $\vec{\omega}$  introduit au paragraphe précédent. On obtient ainsi :

$$\mathcal{L}(\vec{\theta}, \vec{\omega}) = \prod_{c=1}^{n_c} \frac{e^{-M_c(\vec{\theta}, \vec{\nu}(\vec{\omega}, \Delta^+, \Delta^-))} M_c(\vec{\theta}, \vec{\nu}(\vec{\omega}, \Delta^+, \Delta^-))^{D_c}}{D_c!} \prod_{s=1}^{n_s} \frac{1}{\sqrt{2\pi}} e^{-\frac{\omega_s^2}{2}} \quad (1.28)$$

Pour des raisons pratiques, il sera parfois utile de travailler avec le logarithme de la fonction de vraisemblance plutôt qu'avec la fonction elle-même, soit :

$$\begin{aligned} -\ln \mathcal{L}(\vec{\theta}, \vec{\omega}) &= \sum_{c=1}^{n_c} M_c(\vec{\theta}, \vec{\nu}(\vec{\omega}, \Delta^+, \Delta^-)) - \sum_{c=1}^{n_c} D_c \ln M_c(\vec{\theta}, \vec{\nu}(\vec{\omega}, \Delta^+, \Delta^-)) \\ &\quad + \sum_{c=1}^{n_c} \ln(D_c!) + \sum_{s=1}^{n_s} \frac{\omega_s^2}{2} + \frac{n_s}{2} \ln(2\pi) \end{aligned} \quad (1.29)$$

#### 1.4.4 Exemple simple

Pour terminer ce chapitre, illustrons ce qui a été discuté par un exemple simple. La mesure d'une section efficace  $\sigma$  combinant deux canaux d'analyse (que l'on désignera par  $e$  et  $\mu$  en référence à l'état final étudié, par exemple). Dans chaque canal, le nombre d'événements attendu se compose du signal et de bruit de fond. Le nombre d'événements de signal est relié à la section efficace de production par le produit de cette dernière par

la luminosité  $\ell$  et l'efficacité de sélection ou acceptance  $\xi_{i=e,\mu}$ . Le nombre d'événements de bruit fond est estimé à partir de la luminosité  $\ell$  et de l'efficacité de sélection  $\beta_i$  (qui inclut la section efficace théorique, supposée connue). Les notations génériques introduites précédemment deviennent :

$$\vec{\theta} = (\sigma), \vec{\nu} = \begin{pmatrix} \beta_e \\ \xi_e \\ \beta_\mu \\ \xi_\mu \\ \ell \end{pmatrix} \quad (1.30)$$

et pour le modèle :

$$M_i(\vec{\theta}, \vec{\nu}) = \ell \times (\beta_i + \xi_i \times \sigma) \quad (1.31)$$

On considère trois sources d'incertitude indépendantes : une erreur sur la mesure de luminosité ( $\Delta\ell$ ), une erreur sur la calibration de la simulation ( $\Delta^{cal}\nu$ ) et une erreur théorique sur la section efficace du fond ( $\Delta^{th}\nu$ ). Les matrices d'erreur sont donc des matrices  $3 \times 5$ . En supposant les erreurs symétriques, elles s'écrivent :

$$\Delta^\pm = \Delta = \begin{pmatrix} 0 & \Delta^{cal}\beta_e & \Delta^{th}\beta_e \\ 0 & \Delta^{cal}\xi_e & 0 \\ 0 & \Delta^{cal}\beta_\mu & \Delta^{th}\beta_\mu \\ 0 & \Delta^{cal}\xi_\mu & 0 \\ \Delta\ell & 0 & 0 \end{pmatrix} \quad (1.32)$$

Si les erreurs entre  $\beta_e$  et  $\xi_e$  dues à la calibration de la simulation sont anticorrélées, c'est-à-dire qu'une modification de la calibration augmente  $\beta_e$  en diminuant  $\xi_e$  et inversement, cette anticorrélation pourra se traduire en prenant  $\Delta^{cal}\beta_e > 0$  et  $\Delta^{cal}\xi_e < 0$ .

Il y aura trois paramètres de nuisance indépendants  $\vec{\omega}$ , distribués normalement. Les paramètres de nuisance en fonction de  $\vec{\omega}$  sont alors :

$$\vec{\nu} = \vec{\nu}_0 + \vec{\omega}^\top \Delta = \begin{pmatrix} \beta_{e0} & +0 & +\omega_2 \Delta^{cal}\beta_e & +\omega_3 \Delta^{th}\beta_e \\ \xi_{e0} & +0 & +\omega_2 \Delta^{cal}\xi_e & +0 \\ \beta_{\mu0} & +0 & +\omega_2 \Delta^{cal}\beta_\mu & +\omega_3 \Delta^{th}\beta_\mu \\ \xi_{\mu0} & +0 & +\omega_2 \Delta^{cal}\xi_\mu & +0 \\ \ell_0 & +\omega_1 \Delta\ell & +0 & +0 \end{pmatrix} \quad (1.33)$$

et finalement la fonction de vraisemblance devient :

$$\mathcal{L}(\sigma, \vec{\omega}) = (2\pi)^{\frac{-3}{2}} e^{-\frac{\omega_1^2 + \omega_2^2 + \omega_3^2}{2}} \prod_{i=e,\mu} \frac{e^{-\ell \times (\beta_i + \xi_i \times \sigma)} (\ell \times (\beta_i + \xi_i \times \sigma))^{D_i}}{D_i!} \quad (1.34)$$

Toutes les applications discutées dans la suite de ce mémoire reposeront sur la même structure avec quelques complexifications (augmentation du nombre de paramètres ajustés, de sources d'incertitudes, de canaux, différentiation des bruits de fond, ...).



Olim lacus colueram,  
olim pulcher exstiteram,  
dum cignus ego fueram.  
Miser, miser!  
modo niger  
et ustus fortiter!

“Cignus ustus cantat”, Carmina Burana 130

# 2

## Implémentation des méthodes d’analyse

Dans ce chapitre, nous allons décrire l’implémentation effective des outils présentés au chapitre précédent. La première étape consiste à construire la fonction de vraisemblance 1.28, en fournissant le modèle, les données observées et les paramètres de nuisance (valeur nominale et matrices d’erreur) :  $\mathcal{L} = \mathcal{L}(\vec{\theta}, \vec{\omega}) = \mathcal{L}(\vec{M}(\vec{\theta}, \vec{\omega}, \vec{v}_0, \Delta^+, \Delta^-), \vec{D})$ . Ensuite, cette fonction sera utilisée pour construire numériquement soit la fonction de vraisemblance profilée, soit une densité bayésienne *a posteriori*. L’ensemble de ces implémentations forme le code MEFISTO (**M**onte-**C**arlo **E**rrors **F**it with **S**tatistical **T**ools).

### 2.1 Ajustements monodimensionnels

#### 2.1.1 Vraisemblance profilée

Le calcul de la fonction de vraisemblance profilée, passe par la minimisation de  $\mathcal{L}$  relativement à  $\vec{\omega}$  pour différentes valeurs de  $\vec{\theta}$ . Pour ce faire on doit fixer un ensemble de valeurs de  $\vec{\theta}$ , et donc fixer des bornes de variations de  $\theta$ . Il est donc important d’avoir une idée du résultat final, pour fixer ce domaine (pour couvrir la région où le rapport de vraisemblance  $\lambda$  est inférieur à  $12.5$  soit  $5\sigma$ ) et le nombre de points de calcul (typiquement quelques centaines pour un temps de calcul raisonnable, de quelques minutes à quelques heures). Ces points de calculs peuvent soit être choisis aléatoirement (distribution uniforme), soit former une grille à pas fixe. Cette dernière solution est utilisée par le programme HISTFACTORY basé sur ROOFIT [5] qui est utilisée par défaut dans les analyses de physique du top dans l’expérience ATLAS. La grille aléatoire est utilisée dans MEFISTO. Dans les deux programmes, la minimisation de la fonction de vraisemblance en chaque point de la grille utilise le logiciel MINUIT [6]. La similarité des deux implémentations va permettre de valider, au chapitre 4, l’implémentation de la fonction de vraisemblance dans MEFISTO en comparant aux résultats de HISTFACTORY.

Une fois la fonction de vraisemblance profilée estimée, on en déduit  $\lambda$  en divisant par le minimum global. On dispose, à la fin, du rapport de vraisemblance évalué sur une grille discrète. En interpolant la courbe obtenue, on peut résoudre numériquement l’équation 1.9 définissant les niveaux de confiance.

La courbure de la fonction de vraisemblance au minimum donne une estimation de l'incertitude de mesure totale (statistique+systematique). L'incertitude statistique seule peut s'évaluer en étudiant la fonction de vraisemblance des paramètres  $\vec{\theta}$  seuls, en fixant les paramètres de nuisance à leur valeur nominale  $\vec{\nu} = \vec{\nu}_0$ .

### 2.1.2 Intégration bayésienne simple

Le calcul d'une intégrale de la forme  $I = \int_a^b f(x)g(x)dx$ , où  $g(x)$  est une densité de probabilité sur l'intervalle  $[a; b]$ , revient au calcul de la valeur moyenne de la fonction  $f(x)$ . Cette moyenne peut être approximée par l'estimateur de moyenne empirique :

$$\hat{I} = \frac{b-a}{n} \sum_{i=1}^n f(x_i)$$

où les  $x_i$  constituent un échantillon de  $n$  réalisations d'une variable aléatoire distribuée selon  $g(x)$ . Cet estimateur converge vers  $I$  quand la taille de l'échantillon devient grande avec un écart-type  $\sigma_{\hat{I}} = \sigma_g/\sqrt{n}$ ,  $\sigma_g$  étant l'écart-type de la distribution  $g$ . C'est le principe de l'intégration Monte-Carlo.

En utilisant la méthode précédente pour générer aléatoirement des réalisations des paramètres de nuisance distribués selon la densité *a priori*, le calcul de l'intégrale sur les paramètres de nuisance devient :

$$f(\vec{\theta}) \propto \int \mathcal{L}(\vec{\theta}, \vec{\nu})\pi(\theta)k(\vec{\nu}) d\vec{\nu} \approx \frac{\pi(\theta)}{n} \sum_{i=1}^n \mathcal{L}(\vec{\theta}, \vec{\nu}_i)$$

La taille  $n$  de l'échantillon à moyenner doit être suffisamment grande pour parcourir l'ensemble de l'espace des paramètres de nuisance. En général quelques dizaines à quelques centaines de milliers d'itérations suffisent. De plus, lors d'une évaluation numérique, il faut effectuer ce calcul pour un ensemble de valeurs de  $\vec{\theta}$  et calculer l'intégrale sur  $\vec{\theta}$  en incluant la densité *a priori* de  $\vec{\theta}$  pour déterminer la normalisation de la densité *a priori*.

Ici encore erreurs systématiques et statistiques sont entièrement intégrées dans la largeur de la densité *a posteriori*. Ceci est illustré sur la figure 2.1. La partie gauche de la figure montre la fonction de vraisemblance pour plusieurs réalisations des paramètres de nuisance tirées aléatoirement suivant la densité *a priori* appropriée, comme décrit au paragraphe 1.4.2. La courbe violette ( $\vec{\nu}_0$ ) correspond à la valeur centrale des paramètres de nuisance. Les paramètres utilisés correspondent à une mesure de section efficace single top en voie  $s$  dans l'expérience DØ [7].

La somme des ces courbes (en utilisant beaucoup plus de fluctuations) permet de calculer l'intégrale qui correspond exactement, dans le cas d'un *a priori* uniforme sur un intervalle suffisamment large, à la densité *a posteriori* (courbe de droite).

L'incertitude statistique est quantifiée par la largeur moyenne des fonctions de vraisemblance pour  $\vec{\nu}$  fixé.

Dans le programme MEFISTO, à la fin de l'intégration, on dispose cette fois de la densité *a posteriori* évaluée sur une grille aléatoire, à un facteur de normalisation près. Par interpolation, il est alors aisé de calculer l'intégrale de normalisation, de déterminer la valeur la plus probable et finalement d'obtenir des intervalles de confiance ou des limites.

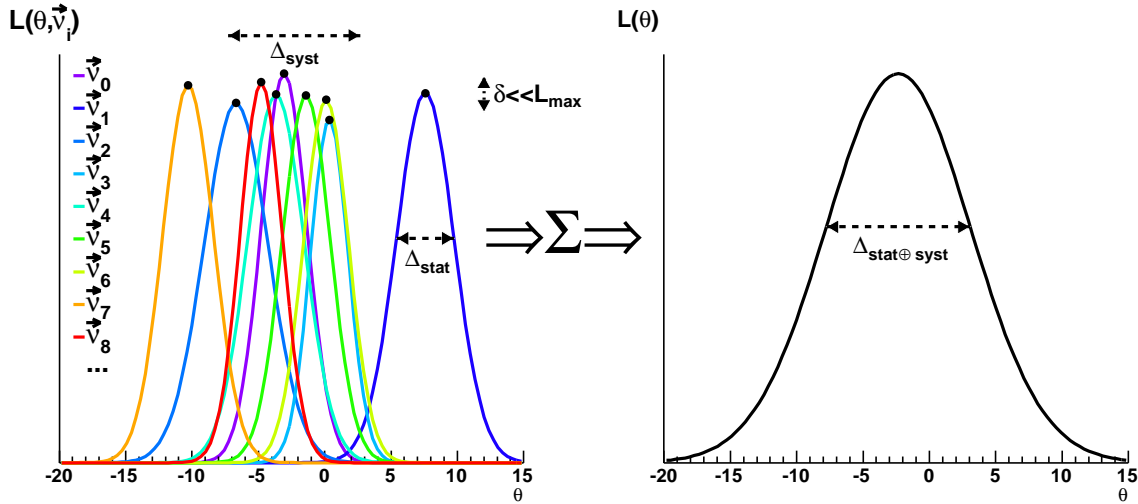


FIG. 2.1: Description schématique de l'intégration sur les paramètres de nuisance.

## 2.2 Ajustement de plusieurs paramètres

Pour un problème unidimensionnel, quelques centaines de points de calcul suffisent. Lorsqu'on souhaite étendre à un ajustement de plusieurs paramètres, la taille de la grille d'évaluation augmente exponentiellement avec le nombre de paramètres. Elle devient rapidement trop grande pour être calculée en un temps raisonnable. De même, lors du développement d'une analyse, notamment pour optimiser certaines coupures, il peut être souhaitable de reproduire un grand nombre de fois le calcul de la densité *a posteriori*, même unidimensionnelle, et donc de pouvoir faire le calcul rapidement, au prix d'approximations supplémentaires.

### 2.2.1 Grille et échantillonnage

Il est utile, à ce stade, de discuter la présentation concrète des résultats numériques. Une densité de probabilité peut être obtenue sous deux formes : une expression analytique ou un ensemble de points distribués selon cette probabilité. Pour des problèmes suffisamment complexes, un calcul analytique tel l'intégration sur les paramètres de nuisance est exclu. Si on est capable de déterminer la valeur de la densité *a posteriori*  $f(\vec{\theta})$  en un point donné de l'espace des paramètres, on peut substituer à la fonction analytique une évaluation de la densité sur une grille uniforme (aléatoire ou à pas fixe). On disposera donc de deux représentations possibles du résultat de la mesure :

- Une fonction : des couples  $\left\{ \left( \vec{\theta}_i, f(\vec{\theta}_i) \right) \right\}_{i=1..n}$  avec  $\vec{\theta}_i$  distribués uniformément.
- Un échantillon : des valeurs  $\left\{ \vec{\theta}_i \right\}_{i=1..n}$  distribuées selon la densité  $f(\vec{\theta})$ .

La première approche autorise le calcul d'intégrales et d'intervalles de confiance. La seconde approche permet très simplement d'obtenir la distribution marginale de chaque variable (dans le cas d'un problème multidimensionnel) et d'effectuer un changement de variable. Il pourra s'avérer utile de passer d'une représentation à l'autre.

Dans le cas le plus général on utilisera des événements pondérés, soit des couples  $\left\{ \left( \vec{\theta}_i, w_i \right) \right\}$ . Dans ce cas la distribution (histogramme) des  $\vec{\theta}_i$  pondérés par  $w_i$  permet d'es-

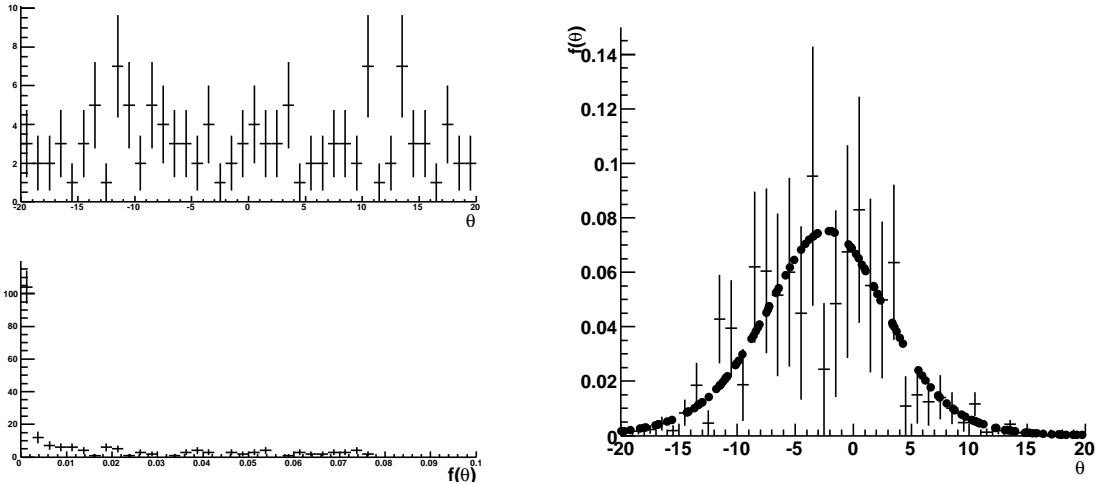


FIG. 2.2: Représentation de la densité par 200 couples  $\left\{ \left( \vec{\theta}_i, f(\vec{\theta}_i) \right) \right\}$ . A gauche : la distribution uniforme des  $\theta_i$  (en haut) et la distribution des poids (en bas). A droite : la densité pondérée (histogramme avec incertitudes) et le graphe  $\left\{ \left( \vec{\theta}_i, f(\vec{\theta}_i) \right) \right\}$  (points noirs).

timer la densité *a posteriori*.

Dans le cas limite où tous les poids valent 1, l'information est complètement contenue dans la distribution des  $\vec{\theta}_i$ .

A l'inverse la représentation  $\left\{ \left( \vec{\theta}_i, f(\vec{\theta}_i) \right) \right\}$  correspond à l'autre cas limite où toute l'information est contenue dans les poids  $\omega_i = f(\vec{\theta}_i)$  et les  $\theta_i$  uniformes n'apportent aucune information. Dans ce seul cas, il existe deux façons de visualiser la densité  $f(\theta)$  : tracer le graphe des couples  $\left\{ \left( \vec{\theta}_i, f(\vec{\theta}_i) \right) \right\}$  (ce qui permet une estimation analytique par interpolation) ou la distribution pondérée de  $\vec{\theta}_i$ . La première solution est néanmoins plus précise et permet d'utiliser un échantillon de taille plus modeste comme l'illustre la figure 2.2.

## 2.2.2 Méthode approximée

Afin de définir une procédure simplifiée et plus rapide, commençons par approfondir le principe de l'intégration sur les paramètres de nuisance à partir de la figure 2.1.

Les paramètres de nuisance modifient la position de chaque distribution, qui peut être évaluée par le mode  $\theta_{max}$  (estimateur du maximum de vraisemblance) : la dispersion de l'estimateur du maximum de vraisemblance mesure l'erreur systématique. Lors de l'intégration sur les paramètres de nuisance, on combine ces deux effets et la largeur finale de la distribution *a posteriori* donne une erreur combinée.

De plus, la valeur du maximum de la vraisemblance  $\mathcal{L}_{max} = \mathcal{L}(\theta_{max})$  mesure la qualité de l'ajustement. Un paramètre de nuisance n'est pas ajustable et donc chaque fluctuation doit mener à un ajustement de qualité similaire.  $\mathcal{L}_{max}$  étant une variable aléatoire, sa valeur peut fluctuer d'un jeu de paramètres à l'autre mais on s'attend à ce que la fluctuation relative reste faible.

A partir de cette analyse, on constate que, pour estimer l'effet des incertitudes systématiques, on peut se contenter de mesurer, pour chaque fluctuation, la fonction de vraisemblance en un seul point, son maximum. Pour chaque fluctuation on détermine la

valeur des paramètres  $\vec{\theta}$  qui maximisent la vraisemblance. On obtient ainsi un échantillon de valeurs  $\vec{\theta}$ , qu'il faut pondérer le cas échéant par la valeur de la densité *a priori* pour  $\vec{\theta}$ . Pour un *a priori* uniforme borné ceci revient à rejeter les points hors de l'intervalle considéré. Rigoureusement, il faudrait également pondérer chaque valeur par sa probabilité, soit la valeur maximale de la fonction de vraisemblance. On peut néanmoins s'en dispenser pour peu que la dispersion relative de  $\mathcal{L}_{max}$  soit faible (*cf.*  $\delta$  sur la figure 2.1).

La difficulté principale consiste à intégrer l'erreur statistique à cette approche. Pour cela il faudrait non pas prendre le maximum de la fonction de vraisemblance mais l'échantillonner en tirant un ou plusieurs points aléatoirement selon cette distribution. Cette approche sera traitée exactement au paragraphe suivant.

Dans le cas de l'estimation du résultat attendu (mesure de sensibilité d'une analyse), on fixe les paramètres à ajuster à une valeur attendue ce qui permet de calculer le nombre d'événements attendu :  $\vec{D} = \vec{M}(\vec{\theta}_0, \vec{v})$ . On peut alors approximer l'erreur statistique en utilisant une propriété de la loi de Poisson :  $P(\lambda, n) \approx P(n, \lambda)$ . Ainsi en faisant fluctuer le nombre d'événements  $D_c$  dans chaque canal  $c$  selon une loi de Poisson de paramètre  $\lambda = D_c$ , et en cherchant le maximum de la fonction vraisemblance on obtient un résultat similaire à celui consistant à tirer aléatoirement des points selon la fonction de vraisemblance. Cette méthode n'est plus applicable lors d'une mesure réelle où le nombre d'événements mesuré est une grandeur fixe.

### 2.2.3 Monte-Carlo par chaînes de Markov (MCMC)

L'approche précédente résout le problème du calcul rapide, mais est, en toute rigueur, fautive, autant d'un point de vue fréquentiste que bayésien. Pour traiter exactement le problème bayésien, nous allons conserver l'idée d'échantillonner la densité *a posteriori* mais de manière exacte. On a pour cela recours à un Monte-Carlo par chaînes de Markov [8].

#### 2.2.3.1 Chaînes de Markov

Une chaîne de Markov est une série  $\vec{\theta}_n$  de réalisations de variables aléatoires  $\vec{\Theta}_n$  telle que la probabilité de la réalisation  $i + 1$  ne dépend que de la réalisation précédente, soit :

$$P(\vec{\Theta}_{i+1} = \vec{\theta}_{i+1}) = P(\vec{\Theta}_{i+1} = \vec{\theta}_{i+1} | \vec{\Theta}_i = \vec{\theta}_i) \quad (2.1)$$

Une telle chaîne peut être utilisée, sous certaines conditions, pour échantillonner une densité de probabilité  $f(\vec{\theta})$ . La chaîne est dite ergodique si, indépendamment du premier élément de la chaîne, la densité  $f_{\vec{\Theta}_i}(\vec{\theta}_i)$  de  $\vec{\Theta}_i$  converge vers celle de  $f(\vec{\theta})$ . Ainsi, à partir d'une certaine itération  $n$ , la chaîne fournit une bonne approximation d'un échantillon distribué selon  $f(\vec{\theta})$ . En revanche un tel échantillon n'est pas simple puisque les variables aléatoires ne sont pas indépendantes.

L'algorithme Metropolis-Hastings [8] permet de créer une chaîne ergodique. Il consiste à construire l'élément  $i + 1$  à partir de l'élément  $i$  de la manière suivante :

- On tire aléatoirement une valeur  $\vec{\theta}^*$  selon une densité  $q(\vec{\theta}^* | \vec{\theta}_i)$ , dite fonction d'échantillonnage.
- On détermine la probabilité  $\alpha(\vec{\theta}_i, \vec{\theta}^*)$  de transition de  $\vec{\theta}_i$  vers  $\vec{\theta}^*$  selon :

$$\alpha(\vec{\theta}_i, \vec{\theta}^*) = \min \left\{ 1, \frac{f(\vec{\theta}^*) q(\vec{\theta}_i | \vec{\theta}^*)}{f(\vec{\theta}_i) q(\vec{\theta}^* | \vec{\theta}_i)} \right\} \quad (2.2)$$

où  $f(\vec{\theta})$  est la densité à échantillonner.



- On tire un nombre aléatoire  $r$ , uniforme sur  $[0, 1]$ . Si  $r < \alpha$ , on pose  $\vec{\theta}_{i+i} = \vec{\theta}^*$ , sinon  $\vec{\theta}_{i+1} = \vec{\theta}_i$  et la chaîne stagne.

L'élément initial  $\vec{\theta}_0$  est tiré aléatoirement suivant la fonction d'échantillonnage. Pour générer un nombre aléatoire distribué selon  $f$ , on a donc besoin de deux nombres aléatoires l'un uniforme et l'autre distribué selon  $q$  (en réalité le rendement est encore inférieur). Pour être pertinente, cette procédure doit être telle qu'il est plus rapide d'échantillonner  $q$  que  $f$ . En revanche, la chaîne n'est exploitable qu'après avoir atteint le régime convergent, ce qui est d'autant plus rapide que  $q$  est proche de  $f$ . Il est également important de remarquer que la probabilité de transition ne met en jeu la densité à échantillonner que par le rapport  $f(\vec{\theta}^*)/f(\vec{\theta}_i)$ . Dans l'échantillonnage de la densité *a posteriori*, cette dernière est donnée par la fonction de vraisemblance à un facteur de normalisation près, qu'il n'est pas aisé de déterminer *a priori*. Cette normalisation se simplifie dans le rapport et  $f(\vec{\theta}^*)/f(\vec{\theta}_i) = \mathcal{L}(\vec{\theta}^*)\pi(\vec{\theta}^*)/\mathcal{L}(\vec{\theta}_i)\pi(\vec{\theta}_i)$ .

La probabilité de transition se simplifie dans le cas particulier d'une fonction d'échantillonnage symétrique, *i.e.*  $q(a, b) = q(b, a)$  alors :

$$\alpha(\vec{\theta}_i, \vec{\theta}^*) = \min \left\{ 1, \frac{f(\vec{\theta}^*)}{f(\vec{\theta}_i)} \right\} \quad (2.3)$$

Ce sera notamment le cas si on choisit pour  $q(a, b)$  une loi multinormale de moyenne  $a$  et de covariance constante.

Enfin, après création d'une chaîne de taille  $n$ , toute la chaîne n'est pas exploitable. La procédure de filtrage est la suivante [9] :

- D'une part, la chaîne de Markov n'échantillonne la densité  $f$  qu'après avoir convergé. On doit donc ignorer le début de la chaîne (procédure de *burn-in*. On choisit le premier élément  $\vec{\theta}_b$  de la chaîne utile tel que  $f(\vec{\theta}_b) > f_{1/2}$  où  $f_{1/2}$  désigne la médiane de  $f$  estimée sur la chaîne complète).
- D'autre part, les valeurs successives de la chaîne sont, par essence, corrélées. Pour produire un échantillon simple, on ne sélectionne qu'une valeur tous les  $n_c$  pas,  $n_c$  caractérisant la longueur de corrélation de la chaîne. Cette longueur se détermine à partir de la fonction d'autocorrélation de la chaîne, déterminée à partir de la transformée de Fourier rapide (FFT) de la chaîne.

Au final, pour une chaîne de taille  $n$ , l'échantillon indépendant aura une taille  $\frac{n-b}{n_c}$ . Il est donc important, dans un souci d'efficacité d'avoir  $n_c, b \ll n$ .

### 2.2.3.2 Application au problème étudié

On souhaite échantillonner la densité *a posteriori*,  $f(\vec{\theta}) \propto \int \mathcal{L}(\vec{\theta}, \vec{\omega})\pi(\theta)d\vec{\omega}$ , où  $\mathcal{L}(\vec{\theta}, \vec{\omega})$  est définie par l'équation 1.28. La marginalisation sur les paramètres de nuisance  $\vec{\omega}$ , peut s'approximer par intégration Monte-Carlo puisque ces paramètres sont distribués selon une loi multinormale centrée, réduite. Il reste alors à échantillonner :

$$e(\theta) = \sum_{i=1}^n \mathcal{L}(\vec{\theta}, \vec{\omega}_i)\pi(\vec{\theta}) \quad (2.4)$$

avec les  $\vec{\omega}_i$  tirés aléatoirement. Pour cela, on va échantillonner, pour chaque  $\vec{\omega}_i$ , la fonction  $\mathcal{L}(\vec{\theta}, \vec{\omega}_i)\pi(\vec{\theta})$  et regrouper ces différents échantillons  $\mathcal{E}_i$ , en utilisant une chaîne de Markov. Il vaut alors faire attention au poids attribué à chaque sous-échantillon dans l'échantillon final.

En considérant un ensemble  $\{\vec{\theta}_j\}_{j=1..m_i}$  distribué selon  $\mathcal{L}(\vec{\theta}, \vec{\omega}_i)\pi(\vec{\theta})$ , la loi de probabilité de l'échantillon est  $K_i^{-1}\mathcal{L}(\vec{\theta}, \vec{\omega}_i)\pi(\vec{\theta})$  avec  $K_i = \int \mathcal{L}(\vec{\theta}, \vec{\omega}_i)\pi(\vec{\theta})d\vec{\theta}$ . Le facteur de normalisation  $K_i$  change pour chaque jeu de paramètres de nuisance et diffère de la normalisation globale de la densité *a posteriori*. Ce terme correspond à la valeur moyenne de  $\mathcal{L}(\vec{\theta}, \vec{\omega}_i)$  et peut donc être estimé par la moyenne empirique de l'échantillon :

$$\hat{K}_i = \frac{1}{m_i} \sum_j \mathcal{L}(\vec{\theta}_j, \vec{\omega}_i)\pi(\vec{\theta}) \quad (2.5)$$

De plus, la taille  $m_i$  de chaque chaîne varie également du fait du filtrage effectué. Au final, chaque sous-échantillon  $i$  aura donc un poids  $W_i$  donné par :

$$W_i = \frac{w_i}{w_{tot}}, \text{ avec } w_i = \frac{\hat{K}_i}{m_i} \text{ et } w_{tot} = \sum_{i=1}^n w_i \quad (2.6)$$

Pour un *a priori*  $\pi(\vec{\theta})$  uniforme, on se contentera de rejeter les points sortant du domaine de définition (par exemple les valeurs négatives dans un calcul de section efficace). L'échantillon final  $\mathcal{E}$  sera donc un ensemble de valeurs pondérées :

$$\mathcal{E} = \bigcup_{i=1}^n \mathcal{E}_i, \text{ avec } \mathcal{E}_i = \{(\theta_j, W_i)\} \quad (2.7)$$

Une telle concaténation de chaînes est illustrée sur les figures 2.3 (échantillon complet d'un million de valeurs) et 2.4 (zoom sur les 25 première chaînes). Chaque chaîne correspond à une fluctuation de paramètres de nuisance. Sur chacune des deux figures, la vignette en bas à droite donne le logarithme de la fonction de vraisemblance en fonction de la valeur échantillonnée. Elle est l'équivalent de la figure 2.1 : chaque fluctuation correspond à une chaîne et contribue par une petite parabole dont la largeur donne l'incertitude statistique. La dispersion des chaînes traduit la dispersion systématique due aux paramètres de nuisance. Les vignettes en haut à droite et en bas à gauche illustrent la stabilité de la chaîne après *burn-in* et décorrélation. Les chaînes étant mises bout à bout, un changement de chaîne se traduit par un changement de maximum de la fonction de vraisemblance et de valeur moyenne du paramètre ajusté, d'où la structure en « paquets » de l'échantillon global mis en évidence sur la figure 2.4.

Dans l'implémentation au sein du code MEFISTO, on fixe le nombre de fluctuations  $n$ , de manière similaire à l'estimation sur une grille, ainsi que la taille de chaque chaîne (quelques milliers de valeurs). Pour la fonction d'échantillonnage, on choisit une loi multivariée normale de moyenne  $\vec{\theta}_i$  (itération précédente) et de covariance fixée. Afin de déterminer le premier maillon de la chaîne ainsi que la matrice de covariance de la densité d'échantillonnage  $q$ , on utilise la méthode approximée décrite au paragraphe 2.2.2 pour produire un premier échantillon. La moyenne empirique de cet échantillon est utilisée comme premier maillon et sa covariance empirique comme largeur de la densité d'échantillonnage.

En fin de procédure, on dispose alors d'un échantillon de valeurs pondérées de  $\vec{\theta}$  distribué suivant la densité *a posteriori*, marginalisée sur les paramètres de nuisance.

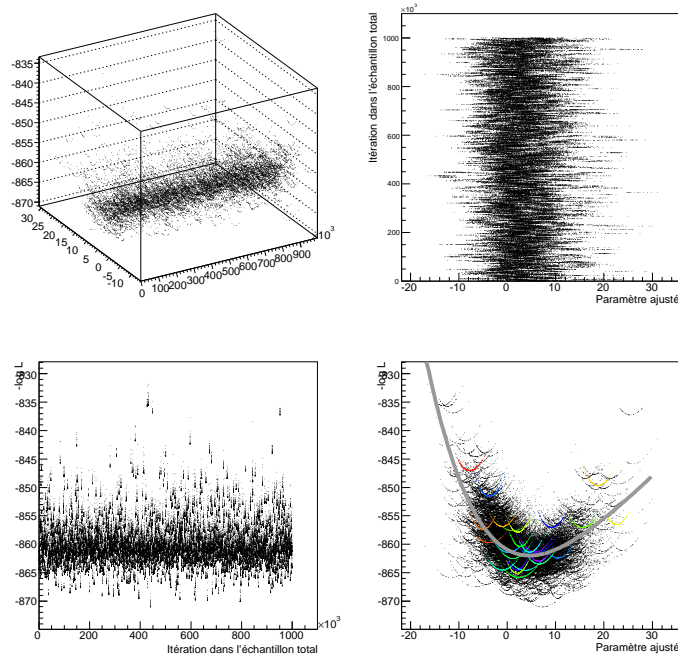


FIG. 2.3: Concaténation de chaînes de Markov : les figures en bas à gauche et en haut à droite donnent respectivement  $-\log \mathcal{L}$  et la valeur du paramètre échantillonné en fonction de la position dans l'échantillon global. La figure en bas à droite donne  $-\log \mathcal{L}$  en fonction de la valeur du paramètre. La courbe grise figure le logarithme de la densité *a posteriori* finale.

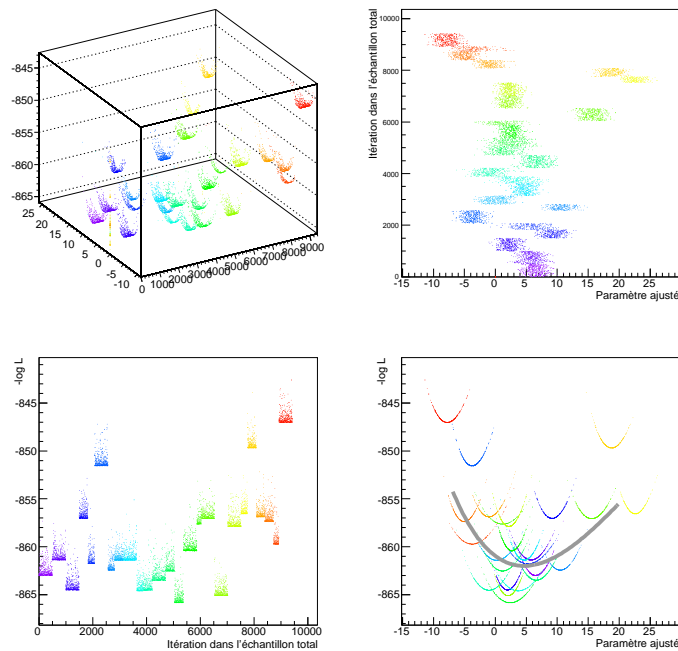


FIG. 2.4: Concaténation de chaînes de Markov : zoom sur les 25 premières chaînes de l'échantillon de la figure 2.3. Chaque chaîne est représentée par une couleur différente.

Νῆσος δενδῆεσσα, θεὰ δ' ἐν δώματα ναίει,  
Ἄτλαντος θυγάτηρ ὀλοόφρονος, ὅς τε θαλάσσης  
πάσης βένθηα οἶδεν, ἔχει δέ τε κίονας αὐτὸς  
μακράς, αἱ γαῖάν τε καὶ οὐρανὸν ἀμφὶς ἔχονσιν.

Homère, "Odyssée" I.51

# 3

## Physique du top au LHC

### 3.1 Le LHC et le détecteur ATLAS

Le *Large Hadron Collider* est le collisionneur proton-proton du CERN. Prévu pour une énergie nominale de 14 TeV et une luminosité nominale de  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$ , il fonctionne depuis février 2010 à une énergie de collision de 7 TeV dans le centre de masse et une luminosité qui a atteint quelques  $10^{33} \text{ cm}^{-2}\text{s}^{-1}$  en 2011. Jusqu'à fin 2011, le LHC a accumulé près de  $5 \text{ fb}^{-1}$  de données à cette énergie.

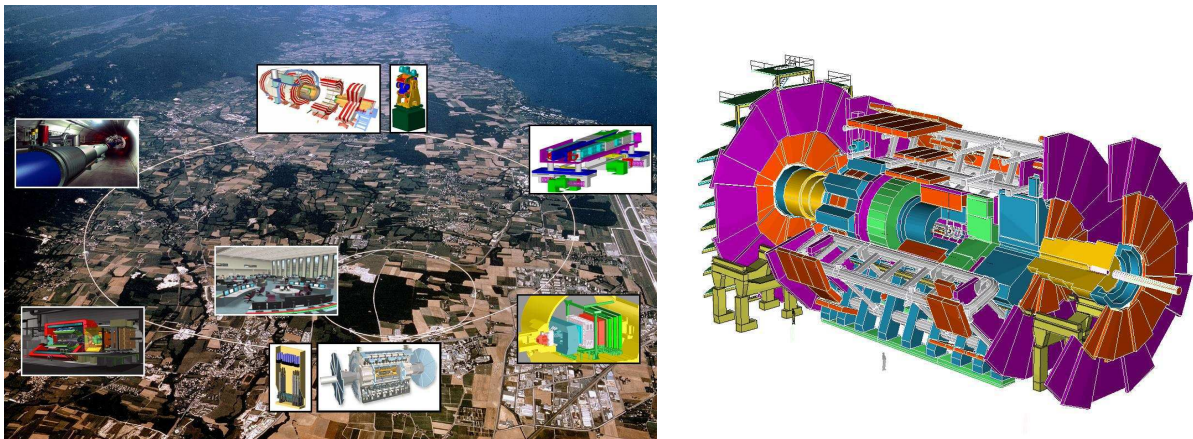


FIG. 3.1: Schémas du collisionneur LHC et du détecteur ATLAS

Le LHC accueille six expériences représentées sur la figure 3.1. Deux sont conçues pour l'étude générale du modèle standard et de ses extensions (ATLAS et CMS), une est consacrée à la physique des ions lourds (ALICE) et une autre est dédiée au secteur des hadrons beaux (LHCb) [10]. Enfin deux expériences de physique diffractive TOTEM et LHCf se concentreront sur l'étude des gerbes hadroniques.

La luminosité instantanée et intégrée recueillie par l'expérience ATLAS est montrée sur la figure 3.2. Ces premières données permettent déjà des analyses de physique compétitives

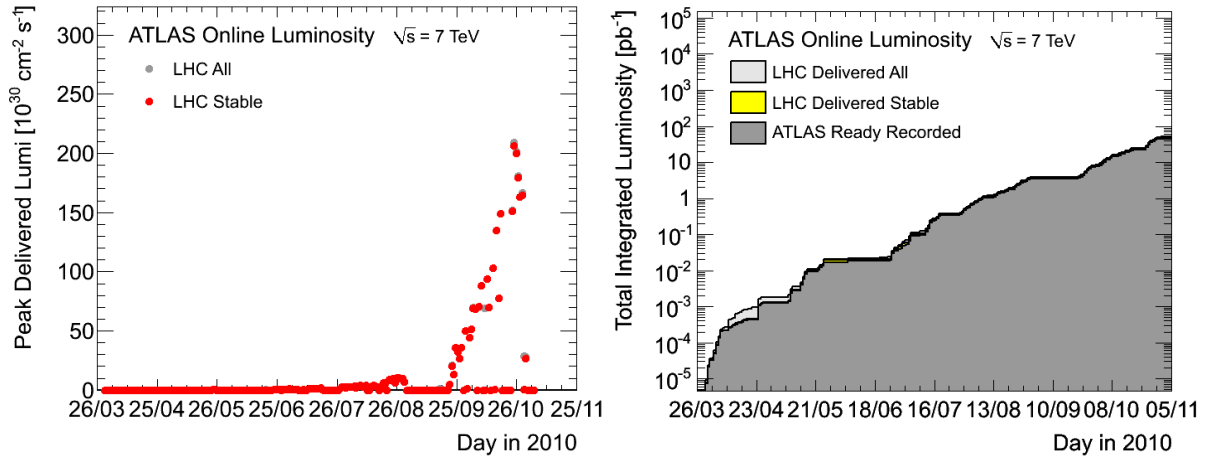


FIG. 3.2: Luminosité instantanée (à gauche) et intégrée (à droite) délivrée par le LHC et enregistrée par ATLAS de mars à décembre 2010. Ces données seront utilisées dans les analyses décrites dans ce document.

avec les résultats finaux du Tevatron. Après réparations, la machine devra être en mesure de fonctionner à l'énergie et à la luminosité nominales à partir de 2015.

Le travail présenté dans ce chapitre se limite à l'expérience ATLAS. Tout comme son concurrent direct CMS, le détecteur ATLAS a la structure classique en peau d'oignon d'un détecteur de physique des hautes énergies, mais avec des choix technologiques différents. En partant du point d'interaction au centre du détecteur une particule rencontre :

- au plus près de faisceau, inclus dans un solénoïde supraconducteur de 2 T, le trajectographe dédié aux mesures de trajectoire et d'impulsion des particules chargées. Il se compose d'un détecteur de vertex à de pixels en silicium, d'un tejectomètre à micropistes de silicium et d'un détecteur de traces à radiation de transition.
- la détection des électrons et des photons est assurée par un calorimètre à échantillonnage à argon liquide utilisant le plomb et le cuivre comme milieu passif. Ce calorimètre électromagnétique est complété par un calorimètre hadronique composé d'un empilement de plomb et de scintillateur organique pour la détection des jets de quarks et de gluons.
- enfin, la détection des muons, qui sortent du calorimètre, est assurée par un spectromètre composé de plusieurs couches de chambres à dérive et d'un aimant toroïdal assurant un champ magnétique de 4 T.

Le détecteur ATLAS est schématisé sur la partie droite de la figure 3.1

## 3.2 Le Modèle Standard et le quark top

### 3.2.1 Le modèle standard et ses limites

Le modèle standard de la physique des particules décrit les interactions électrofaible et forte entre des particules fermioniques chargées (charge électrique, isospin faible et couleur) de masse nulle : les 6 quarks et les 6 leptons. A partir du lagrangien libre des fermions, on impose l'invariance locale sous le groupe  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ . Cette invariance locale introduit naturellement des champs vectoriels sans masse qui compensent le commutateur entre une transformation du groupe et l'opérateur de dérivation. La transformation des nouveaux champs vectoriels sous  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  laisse également invari-

riant le lagrangien de propagation de ces champs, et le lagrangien total est bien invariant de jauge. Finalement la brisure spontanée de la symétrie électrofaible ( $SU(2)_L \otimes U(1)_Y$ ) par le mécanisme de Higgs permet de conférer une masse aux bosons de jauge dans un premier temps, puis aux fermions par couplage de Yukawa au champ de Higgs. Ce mécanisme nécessite l'introduction dans le modèle d'un doublet de champs scalaire complexe, dit champ de Higgs, dont le lagrangien satisfait l'invariance de jauge, mais contient un terme de potentiel quartique. Ce terme fournit un minimum dégénéré au potentiel correspondant à une valeur de champ non nulle. L'existence d'une valeur particulière du vide parmi l'infinité de valeurs dégénérées brise la symétrie. L'invariance de jauge réduit le champ de Higgs à un champ scalaire massif, le boson de Higgs, et une valeur du champ dans le vide qui par couplage aux bosons faibles et aux fermions donne les termes de masse.

La découverte du boson Higgs est l'un des objectifs majeurs du LHC et devrait permettre de valider le modèle standard. Si ce modèle n'a pas été mis en défaut jusqu'ici, il n'en est pas moins un modèle effectif dont la limite de validité devrait se situer à l'échelle du TeV. Le champ de Higgs est scalaire : non seulement aucune autre particule élémentaire de ce type n'est connue, mais cela introduit des divergences quadratiques dans des corrections radiatives à la masse. Ces corrections peuvent ainsi dépasser de plusieurs ordres de grandeur la masse nue (à l'ordre de Born) et seul un ajustement fin des paramètres peut faire disparaître ces divergences et fixer l'échelle d'énergie du mécanisme de Higgs autour du TeV. L'échelle naturelle serait plutôt l'échelle de Planck, à laquelle la gravité n'est plus négligeable, ou l'échelle de grande unification où les couplages électromagnétique, faible et fort convergent. Ces ajustements fins ne sont pas satisfaisants et peuvent être expliqués par de nouveaux phénomènes se manifestant au voisinage de l'énergie de la brisure de la symétrie électrofaible. Le modèle standard présente également un nombre important de paramètres libres (19 si on ne prend pas en compte les masses des neutrinos) incluant en particulier les masses de toutes les particules, l'intensité des couplages et le nombre de familles de particules.

Il existe aujourd'hui quatre grands axes d'extension théorique du modèle standard : supersymétrie (symétrie brisée entre bosons et fermions), dimensions spatiales supplémentaires, unification des forces dans un groupe de symétrie plus vaste et la non-élémentarité des particules connues (modèles de type technicouleur). Ces théories ne sont pas exclusives les unes des autres et leurs combinaisons et variantes donnent lieu à de multiples modèles qui ont pour la plupart des conséquences phénoménologiques (nouvelles particules, modifications des couplages du modèle standard) qui devraient être observables aux énergies du LHC. Le mécanisme de génération des masses reste une des grandes inconnues.

Le quark top, avec sa masse voisine de l'échelle d'énergie électrofaible, est le fermion qui a le plus grand couplage au boson de Higgs (ou à ce qui en tient lieu dans d'autres modèles). La physique du quark top (modes de production, couplages, ...) est ainsi un domaine clé de ce programme de recherche.

### 3.2.2 Le quark top

Le quark top a été découvert en 1995 par les expériences DØ et CDF [11] au Tevatron. Ces expériences ont déterminé sa masse avec précision ( $m_t = 173.2 \pm 0.9$  GeV [12]) et mesuré plusieurs de ses propriétés : section efficace de production [13], charge électrique ( $q = 2/3$  à 92% CL [14]), couplages (matrice CKM [15], recherche de couplage droit [16]),

...Aucune déviation significative par rapport au modèle standard n'a été observée. La mesure de la masse du quark top joue un rôle prépondérant pour contraindre la masse du boson de Higgs via les corrections radiatives à la masse. Combinée aux autres mesures électrofaibles du Tevatron et du LEP (masse du  $W$ , asymétries, ...), la masse actuelle du quark top semble favoriser un boson de Higgs léger comme l'illustre la figure 3.3.

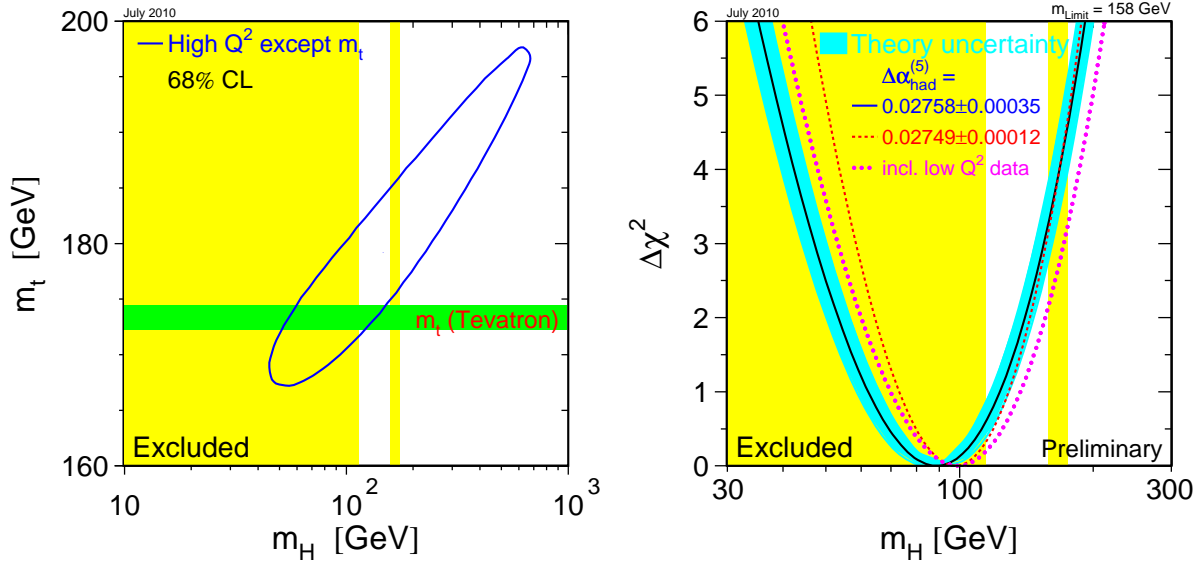


FIG. 3.3: Contraintes expérimentales sur la masse du Higgs. À gauche : la contribution de la masse du quark top (bande verte) et les contributions d'autres observables (ellipse verte). À droite : le  $\chi^2$  de l'ajustement des données électrofaibles en fonction de la masse du Higgs.

Le Tevatron a produit à ce jour de l'ordre de  $10^5$  événements contenant au moins un quark top (pour une luminosité d'environ  $10 \text{ fb}^{-1}$ ). Le LHC devrait en produire plusieurs millions par an. Pour une luminosité de  $400 \text{ pb}^{-1}$  à  $\sqrt{s} = 7 \text{ TeV}$ , le nombre de quarks top produit égale celui du Tevatron. Au delà, le LHC et en particulier ATLAS, permettra de rechercher des signes directs (production associée d'un top avec boson vecteur lourd  $W'$  ou un boson de Higgs chargé  $H^+$ ) ou indirects (déviations des sections efficaces par rapport au modèle standard, couplages anormaux) de nouvelle physique.

Dans un collisionneur hadronique (collisions de quarks et de gluons), le quark top peut être produit par paire, via l'interaction forte, ou seul, par interaction électrofaible (on parle alors de production *single top*).

La production par interaction forte est dominante et a permis la découverte du top en 1995 [11]. Au LHC, La production de paires  $t\bar{t}$  peut provenir soit d'une annihilation quark-antiquark ( $\sim 10\%$  des cas, figure 3.4(a)) soit d'une fusion de gluons ( $\sim 90\%$  des cas, figure 3.4(b et c)). Les deux processus ont des signatures très voisines mais peuvent éventuellement être séparées notamment par la mesure des corrélations de spin entre les deux quarks top [17].

La production électrofaible n'a été mise en évidence au Tevatron que récemment [18]. Les principaux diagrammes de Feynman de ces processus sont donnés sur la figure 3.4(d à f). Cette production *single top* offre un état final différent pour chaque voie de production : un quark  $b$  et un quark  $t$  en voie  $s$ , un quark léger et un quark  $t$  en voie  $t$  et enfin un

boson  $W$  réel et un quark  $t$  pour la production associée  $tW$ . Pour la production en voie  $t$  et associée  $tW$  le quark  $b$  est issu de la mer et un jet supplémentaire provenant du  $\bar{b}$  associé peut parfois être détecté à l'avant du détecteur. L'énergie dans le centre de masse pour produire  $tW$  étant plus importante que pour produire  $tq$ , la section efficace de la production associée est plus faible qu'en voie  $t$ . La production en voie  $s$  est fortement supprimée au LHC car elle fait appel à un antiquark en entrée qui ne peut être qu'issu de la mer. Les modes de production en voies  $t$  et  $s$  ont été observés conjointement au Tevatron. Il pourrait même être possible de séparer les deux processus après l'analyse des  $10 \text{ fb}^{-1}$  de données enregistrées. En revanche la production associée n'a encore jamais été observée et ne pourra l'être qu'au LHC. C'est également un mode de production idéal pour la recherche directe de nouveaux bosons lourds.

Les sections efficaces des différents modes de production du quark top au LHC pour des énergies de 7 TeV et 14 TeV sont résumées dans la table 3.1

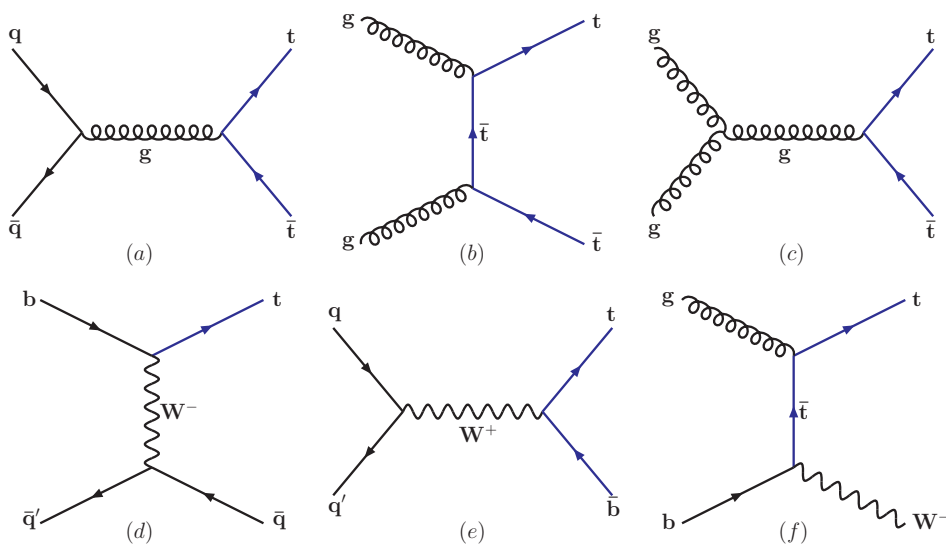


FIG. 3.4: En haut : Diagrammes de Feynman pour la production de paire de quarks top par interaction forte par annihilation de quarks (a) et par fusion de gluons (b et c). En bas : Diagrammes de Feynman pour la production électrofaible de quark top par interaction faible en voie  $t$  (d), en voie  $s$  (e) et en production associée (f).

Énergie	paires $t\bar{t}$	voie $t$	voie $s$	$tW$
7 TeV [19, 20]	164.6	66.2	4.3	14.6
14 TeV [21, 22]	872.8	246.6	10.65	66

TAB. 3.1: Sections efficaces théoriques  $NLO$  en  $pb$  pour les différentes productions du quark top. Les erreurs théoriques sont de l'ordre de 5% à 10%.

Dans le modèle standard, le quark top se désintègre à près de 100% par interaction faible en un boson  $W$  et un quark  $b$ . C'est le seul quark de masse supérieure au boson faible et, comme  $m_W + m_b \ll m_t$ , l'espace de phase est grand et la durée de vie du top est très courte, de l'ordre de  $10^{-25}$  s. La signature d'un quark top est donc un jet issu du quark  $b$  et les produits de désintégration de boson  $W$  réel : soit une paire de quarks (deux jets) soit un lepton chargé et un neutrino. Les états finaux identifiables sont donc :



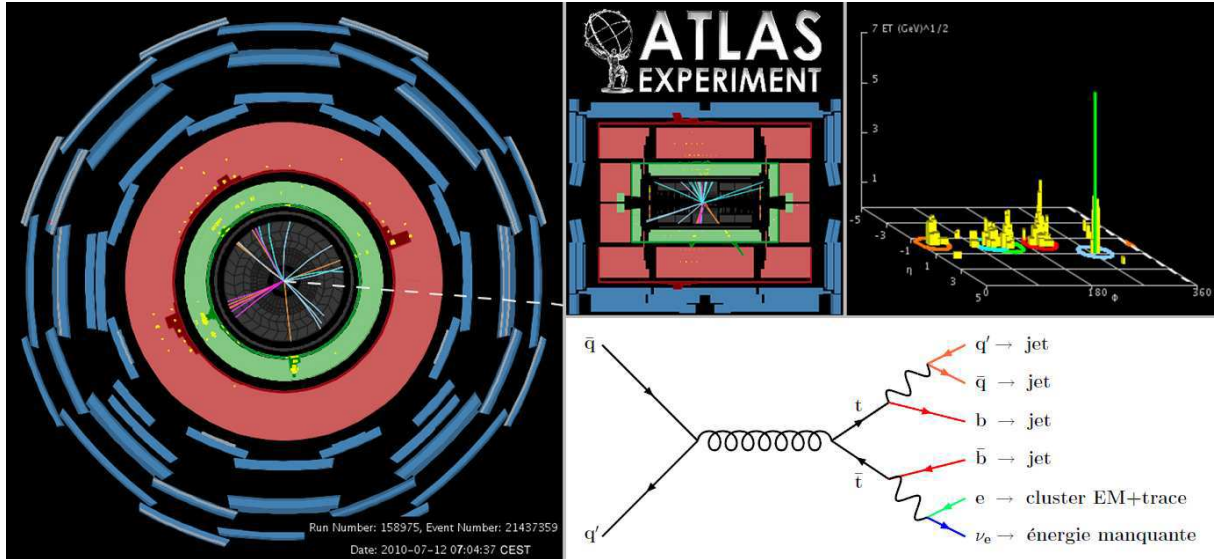


FIG. 3.5: Événement  $t\bar{t}$  dans ATLAS : trois jets (dont un jet de  $b$ ), un électron de  $p_T = 41.4$  GeV (trace orange vers le bas et cluster EM en vert), et 89.3 GeV d'énergie manquante pour une masse transverse du  $W$  de  $m_T = 68.7$  GeV.

- $t\bar{t}$  : deux jets  $b$  et deux  $W$ . Selon les désintégrations du  $W$  on distingue les modes dileptonique ( $\sim 10\%$  : deux leptons et de l'énergie manquante), lepton+jets ( $\sim 45\%$  : deux jets, un lepton et de l'énergie manquante) et hadronique ( $\sim 45\%$  : quatre jets).
- $tW$  : même états finaux que  $t\bar{t}$  avec un unique jet  $b$ .
- voie  $s$  : deux jets  $b$  et un boson  $W$ .
- voie  $t$  : un jet léger, un ou deux jets  $b$  et un boson  $W$ .

Les modes leptoniques sont plus rares ( $\sim 10\%$  par lepton pour  $W \rightarrow l\nu$ ) mais plus faciles à identifier dans un environnement hadronique. Pour une analyse il faut trouver un compromis entre le taux d'événements attendu et l'efficacité d'identification de ces événements. La figure 3.5 montre la signature dans le détecteur ATLAS d'un candidat  $t\bar{t}$ , électron+jets. Dans la suite nous allons nous intéresser aux productions *single top* dans leur mode de désintégration lepton+jets.

### 3.3 Production associée $tW$

#### 3.3.1 Motivation

Nous allons nous intéresser plus particulièrement à la production associée  $tW$ , seul mode de production qui n'a jamais été observé, du fait de sa faible section efficace au Tevatron. La première étude complète de cette voie a été réalisée en 2011 dans la collaboration ATLAS avec une luminosité intégrée de  $35 \text{ pb}^{-1}$ . Nous limiterons ici la description à l'analyse du mode de désintégration lepton+jets, bien qu'une combinaison avec l'analyse dilepton ait été réalisée.

La difficulté majeure vient de la ressemblance de la signature finale (3 jets dont 1 jet  $b$ , 1 lepton chargé et de l'énergie manquante) avec celle d'une paire  $t\bar{t}$  (4 jets dont 2  $b$ , même signature leptonique). Compte tenu de la faible efficacité d'identification des jets  $b$  (environ 60%) et de l'environnement hadronique qui ajoute des jets, la différentiation des deux

processus s'avère complexe. La solution naturelle serait le recours aux méthodes d'analyse multivariées. De fait, c'est ce qui a été étudié par le passé et sera vraisemblablement mis en œuvre à l'avenir. Néanmoins ces méthodes nécessitent une bonne modélisation des distributions des variables discriminantes et ce d'autant plus que la discrimination apportée par chaque variable cinématique est faible. C'est pourquoi cette solution n'a pas été retenue dans cette première étude, privilégiant une analyse robuste (de plus avec ou sans analyse multivariée, la quantité de données analysées ne permet pas encore d'avoir une réelle sensibilité au signal).

### 3.3.2 Bruits de fond et sélections

Il existe de multiples processus produisant une signature similaire au signal recherché, soit exactement, soit du fait de mauvaise reconstruction. Une série de coupures de sélection est mise en œuvre pour rejeter ces fonds. L'analyse repose sur une sélection commune avec la recherche de la voie  $t$ , complétée par quelques coupures spécifiques. Par ordre de section efficace décroissante, les bruits de fond considérés sont :

**Processus multijets** La production de partons dans l'état final (quarks ou gluons) est évidemment le processus dominant dans des collisions hadroniques à haute énergie. La section efficace est difficile à déterminer théoriquement mais avoisine la centaine de mb. Mais contrairement à notre signal, ces processus ne contiennent ni lepton isolé de grande impulsion ni énergie transverse manquante.

La sélection commune est donc appliquée aux événements ayant **déclenché l'acquisition sur un lepton de grande impulsion**. De plus on limite l'étude aux événements contenant **entre 2 et 4 jets, avec une impulsion transverse  $p_T > 25$  GeV et une pseudorapidité  $|\eta| < 4.9$** <sup>1</sup>.

Cependant, il arrive qu'un jet soit malencontreusement identifié comme un lepton isolé (jet contenant beaucoup de  $\pi^0$  pour un électron, jets de saveur lourde où une désintégration semi-leptonique emporte une grande fraction de l'énergie pour un muon), la rareté de ces événements parmi les événements multijets ( $\sim 10^{-5}$ ) est compensée par la grande section efficace. Pour limiter au maximum la contribution de ces événements, on demande **au moins 1 électron ou 1 muon isolé, correctement identifié, de  $p_T > 20$  GeV, qui doit être associé spatialement avec le lepton du système de déclenchement ( $\Delta R < 0.15$ )**.

De plus ces événements mal reconstruits peuvent contenir de l'énergie manquante non physique due à une mauvaise calibration des objets reconstruits. Ces défauts de calibration se traduisent par la présence d'énergie manquante dans la direction du lepton reconstruit. Les événements sélectionnés doivent d'une part contenir **au moins 25 GeV d'énergie transverse manquante ( $E_T$ )**, d'autre part satisfaire **une coupure dite triangulaire**,

<sup>1</sup>On utilise, au choix, trois systèmes de coordonnées pour le vecteur impulsion  $\vec{p}$  :

- les coordonnées cartésiennes  $p_x, p_y$ , et  $p_z$ ,  $z$  désignant l'axe du faisceau
- les coordonnées sphériques  $p, \theta$  (angle polaire) et  $\phi$  (angle azimuthal)
- les coordonnées sphériques modifiées  $p_T = p \sin \theta = \sqrt{p_x^2 + p_y^2}$ ,  $\eta = -\ln(\tan \frac{\theta}{2})$  (pseudorapidité) et  $\phi$  (angle azimuthal). La pseudorapidité est égale à la rapidité pour des particules sans masse. Les écarts de pseudorapidités sont, au premier ordre, invariants de Lorentz. En particulier l'angle entre deux quadri-impulsions sera mesuré par  $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$ .

$E_T + M_T^W > 60 \text{ GeV}$ , où :

$$M_T^W = \sqrt{(p_T^\ell + \cancel{E}_T)^2 - (p_x^\ell + \cancel{E}_{Tx})^2 - (p_y^\ell + \cancel{E}_{Ty})^2} = \sqrt{2p_T^\ell \cancel{E}_T (1 - \cos \Delta\phi_{\ell, \cancel{E}_T})} \quad (3.1)$$

désigne la masse transverse du système lepton/ $\cancel{E}_T$ , soit, pour le signal, la masse transverse du boson  $W$ .

Cet ensemble de sélections constitue la présélection commune aux analyses *single top*.

**Processus électrofaibles** Les processus contenant au moins un lepton chargé sont dominés par la production de bosons faibles  $W$  et  $Z$ , soit en production associée à des jets (figure 3.6a et b), soit par paire (di-bosons, figure 3.6c)

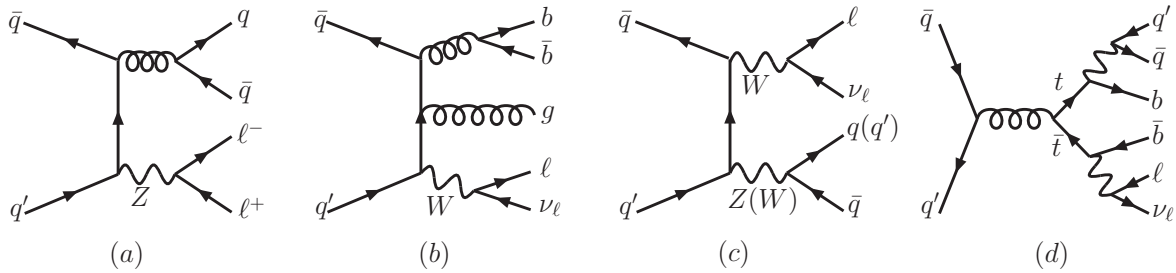


FIG. 3.6: Principaux bruits de fond au signal *single top* : Production associée  $V$ +jets ( (a) :  $Zqq$  et (b) :  $Wbbj$ ), production de di-bosons  $WZ, WW \rightarrow \ell\nu_e qq$  (c) et production de paire  $t\bar{t}$  dans le canal lepton+quarks (d).

Les événements  $Z$ +jets et une partie des di-bosons sont supprimés en imposant **un veto sur la présence d'un deuxième lepton**. Pour limiter les événements  $W$ +jets, qui possèdent une section efficace environ 200 fois supérieure au signal  $tW$ , on s'appuie sur l'existence d'un jet de quark  $b$  dans le signal. L'identification des jets  $b$  passe par l'identification d'un vertex secondaire déplacé dans le cône du jet : les hadrons beaux ont une durée de vie de quelques picoseconde, du fait de leur désintégration par interaction faible, via un terme non-diagonal de la matrice CKM. Ils peuvent donc voler sur plusieurs millimètres avant de se désintégrer, produisant des traces pointant vers un vertex déplacé. On demande finalement qu'**un des jets de l'événement soit identifié comme un jet  $b$ , avec une impulsion transverse  $p_T > 35 \text{ GeV}$** . L'identification des jets  $b$  a une efficacité de l'ordre de 50% pour un taux de réjection de 0.5%.

Après cette coupure, il reste encore une contribution non négligeable de bruits de fond électrofaibles, due principalement aux productions associées d'un boson  $W$  avec des quarks  $b$  ou  $c$ .

**Processus avec un quark top** Finalement, les processus contenant un ou plusieurs quarks top vont présenter une topologie très voisine du signal. La production *single top* en voie  $t$  ( $\sigma_t \approx 5\sigma_{tW}$ ) sera caractérisée par la présence d'un jet vers l'avant du détecteur. Il est rejeté en demandant que **tous les jets de l'événement soient centraux** ( $|\eta| < 2.5$ ).

Reste le bruit de fond le plus problématique de l'analyse : la production de paires de quark top, dont la section efficace est 20 fois supérieure à celle du signal. Seules quelques différences sont notables. D'une part le signal ne contient qu'un unique jet  $b$ . On a alors recours à **un veto sur la présence d'un second jet  $b$ , d'impulsion transverse**

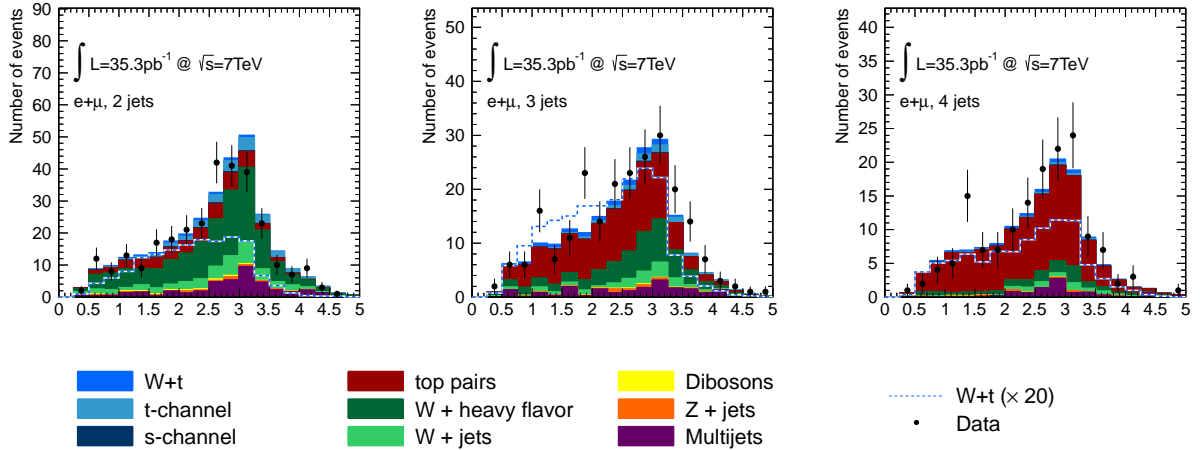


FIG. 3.7: Distributions de  $\Delta R(j_1, j_2)$  pour les événements sélectionnés à 2, 3 et 4 jets (la ligne pointillée représente la forme du signal  $tW$  normalisé à 20 fois la valeur attendue dans le modèle standard).

$p_T > 20$  GeV. Cependant l'efficacité d'identification des jets  $b$  étant de l'ordre de 50%, ce veto n'est pas très efficace. D'autre part, l'angle ( $\Delta R$ ) entre les deux jets de plus haute impulsion est généralement plus fermé pour le signal. Pour les paires  $t\bar{t}$ , les deux jets de plus haute impulsion sont généralement les deux jets  $b$ , issus de la désintégration des deux quarks top. Ces derniers étant produits dos-à-dos, l'angle entre les deux jets est proche de  $\pi$ . En revanche, dans les événements  $tW$ , l'un des jets est issu de la désintégration d'un des bosons  $W$ . La distribution en  $\Delta R$  après toutes les coupures de sélection est représentée sur le figure 3.7. La **coupure finale de l'analyse sera donc  $\Delta R(j_1, j_2) < 2.5$** .

### 3.3.3 Estimation des fonds et de l'acceptance

L'estimation de chaque bruit de fond passe par deux étapes : la détermination de la forme des distributions et leur normalisation (le nombre total d'événements). À l'exception du fond multijets, les distributions de tous les bruits de fond considérés sont estimées à partir de simulation Monte-Carlo incluant un générateur, la simulation complète sous GEANT 4 [23] du détecteur ATLAS et la digitisation du signal. Les générateurs d'événements utilisés sont, suivant les fonds : MC@NLO [24] pour les processus top (signal  $tW$ ,  $t\bar{t}$ , voie  $t$ ), ALPGEN [25] pour les processus  $W$ +jets et  $Z$ +jets (avec la procédure d'association de Mangano pour éviter les doubles comptages en combinant différentes multiplicités de partons), et HERWIG [26] pour les processus di-bosons  $WW$ ,  $WZ$  et  $ZZ$ . Pour chaque processus, l'hadronisation, les interactions multiples et les radiations initiales et finales sont gérées par HERWIG+JIMMY. Les densités de parton de la collaboration CTEQ (version 6) [27] sont utilisées.

On applique alors à ces lots Monte-Carlo, après reconstruction, les mêmes coupures de sélection que dans les données. Les distributions obtenues sont en plus affectées par des facteurs multiplicatifs, qui corrigent les différences d'efficacité de reconstruction observées entre données et simulation. Ces facteurs correctifs sont estimés par les différents groupes de performances de la collaboration. On obtient ainsi l'efficacité de sélection de chaque processus.

Les différents fonds  $W$ +jets sont séparés par composition en saveur lourde :  $W + b\bar{b}$ +jets,  $W + c\bar{c}$ +jets,  $W + c$ +jets,  $W$ +jets légers. On utilise les événements à un seul jet avant et après la sélection du jet  $b$ , ainsi que les événements à deux jets sans sélection sur le jet  $b$  pour déterminer les fractions relatives de  $Wbb$ ,  $Wc$  et  $W$ +légers, le rapport  $Wbb/Wcc$  étant fixé par la simulation. La normalisation globale est également estimée sur les données, avant sélection sur le jet  $b$ .

Pour les autres processus, la normalisation est obtenue par le produit de la section efficace théorique, de la luminosité intégrée et de l'efficacité de sélection.

La contribution du fond multijet est entièrement déterminée dans les données : les distributions sont extraites de données ou l'on inverse une ou plusieurs coupures de qualité du lepton. On obtient ainsi un lot très enrichi en "mauvais leptons", orthogonal aux données analysées, auquel on applique les autres coupures de sélection. La normalisation, avant sélection des jets  $b$ , est déterminée soit en ajustant la distribution d'énergie transverse manquante, soit à partir de la méthode de la matrice [28].

Finalement l'analyse est découpée en six canaux orthogonaux en fonction de la multiplicité de jets reconstruits (de 2 à 4) et de la nature du lepton (électron ou muon). Les nombres d'événements attendus et observés sont présentés dans la table 3.2.

On constate que la contribution des fonds reste très majoritaire. Le fond  $t\bar{t}$  est en partie irréductible. Les autres contributions pourraient être réduites par des coupures plus dures ou l'utilisation d'analyses multivariées. La statistique limitée de cette analyse ne permet pas, ici, d'aller plus loin. Ces améliorations seront apportées dans les analyses sur l'ensemble des données 2011 où la luminosité intégrée atteint  $5 \text{ fb}^{-1}$ , soit plus de 100 fois la luminosité de cette analyse.

TAB. 3.2: Nombres d'événements attendus après sélection  $tW$  pour les six canaux d'analyse. Le nombre d'événements de signal  $tW$  correspond à la section efficace attendue dans le modèle standard. Les incertitudes incluent tous les effets décrits à la section 3.3.4.

	Électron			Muon		
	2 jets	3 jets	4 jets	2 jets	3 jets	4 jets
$tW$	$2.3 \pm 0.5$	$2.7 \pm 0.4$	$1.2 \pm 0.2$	$2.4 \pm 0.5$	$2.8 \pm 0.4$	$1.2 \pm 0.2$
voies $s$ et $t$	$3.7 \pm 0.5$	$1.2 \pm 0.2$	$0.3 \pm 0.1$	$4.4 \pm 0.6$	$1.3 \pm 0.2$	$0.3 \pm 0.1$
$t\bar{t}$	$11.4 \pm 4.0$	$24.6 \pm 5.3$	$23.2 \pm 3.1$	$12.5 \pm 4.1$	$27.5 \pm 6.0$	$25.6 \pm 3.8$
$W$ +jets	$7.5 \pm 3.8$	$2.4 \pm 1.5$	$0.8 \pm 0.7$	$8.6 \pm 4.0$	$2.5 \pm 1.6$	$1.0 \pm 0.9$
$Wc$ +jets	$19.7 \pm 10.3$	$4.8 \pm 2.6$	$1.5 \pm 0.9$	$22.8 \pm 11.8$	$6.4 \pm 3.5$	$1.7 \pm 1.1$
$Wc\bar{c}$ +jets	$2.1 \pm 2.0$	$0.9 \pm 1.0$	$0.2 \pm 0.3$	$2.7 \pm 2.6$	$1.3 \pm 1.4$	$0.4 \pm 0.5$
$Wb\bar{b}$ +jets	$3.8 \pm 3.6$	$1.2 \pm 1.3$	$0.6 \pm 0.8$	$4.7 \pm 4.4$	$2.1 \pm 2.1$	$0.8 \pm 1.0$
$WW, WZ$ et $ZZ$	$1.0 \pm 0.2$	$0.4 \pm 0.1$	$0.1 \pm 0.0$	$1.2 \pm 0.2$	$0.5 \pm 0.1$	$0.1 \pm 0.0$
$Z$ +jets	$1.3 \pm 1.3$	$0.5 \pm 0.5$	$0.6 \pm 0.6$	$1.2 \pm 1.2$	$0.6 \pm 0.6$	$0.2 \pm 0.2$
Multijet	$3.5 \pm 3.5$	$1.2 \pm 1.3$	$0.5 \pm 0.5$	$5.7 \pm 3.4$	$5.5 \pm 3.3$	$1.2 \pm 1.0$
Total	$56.3 \pm 12.9$	$40.0 \pm 6.5$	$29.0 \pm 3.5$	$66.2 \pm 14.6$	$50.5 \pm 8.2$	$32.6 \pm 4.4$
S/B	0.04	0.07	0.04	0.04	0.06	0.04
Données	49	55	29	74	50	37

### 3.3.4 Erreurs systématiques

Les différentes sources d'incertitude vont affecter la normalisation des différents bruits de fond et l'acceptance du signal. Le détail des incertitudes systématiques de cette analyse est décrit dans la thèse de Carole Weydert [29] et dans la note d'analyse ATLAS correspondante [30]. Elles peuvent être regroupées en différentes catégories, pour lesquelles on ne donnera que l'impact sur la section efficace mesurée ( $\Delta\sigma/\sigma$ ).

- **Calibration, résolution et efficacité** : Les erreurs systématiques, dues aux différences résiduelles entre les données et les simulations Monte-Carlo, sur la reconstruction et l'étalonnage des jets, des électrons et des muons ainsi que les erreurs sur les facteurs multiplicatifs correctifs des efficacités, sont propagées aux taux d'événements attendus. Pour ce faire, chaque effet (par exemple l'étalonnage en énergie des jets) est modifié de  $\pm 1\sigma$  dans la simulation et le nombre d'événements après coupures est réestimé. La différence avec la valeur nominale est considérée comme erreur systématique. Les incertitudes  $\pm 1\sigma$  sur chaque effet sont fournies par les différents groupes de performances de la collaboration. L'étalonnage en énergie des jets (20%) et la modélisation de l'identification des jets  $b$  (15%) sont les deux effets dominants.
- **Générateurs Monte-Carlo et densités de partons** : Ces effets sont estimés en comparant les résultats de plusieurs générateurs, tant pour les éléments de matrices que le modèle d'hadronisation et de radiations initiale et finale (MC@NLO, POWHEG+PYTHIA, POWHEG+HERWIG, et AcerMC+Pythia). L'impact des densités de partons est également évalué en pondérant les simulations à partir des différents jeux d'incertitude des PDF CTEQ66, MSTW2008 et NNPDF. De cette manière seul l'impact sur les distributions simulées est pris en compte. Ces effets sont uniquement estimés sur les différents processus de production de quark top. Les erreurs sur les autres processus sont supposées du même ordre de grandeur. Ces sources d'incertitudes contribuent au total à une erreur de 10% sur la section efficace.
- **Sections efficaces théoriques** : Les fonds  $t\bar{t}$  et di-boson sont normalisés aux sections efficaces NLO avec pour incertitudes  $^{+9.5\%}_{-6.9\%}$  et 5%, respectivement, avec un impact de l'ordre de 8% sur le résultat final. Les événements  $Z$ +jets sont également normalisés à une section efficace théorique. Néanmoins, du fait du très faible nombre d'événement attendu ( $<1$ ) et de la statistique Monte-Carlo limitée, on considère une erreur globale de 100% sur la normalisation, qui n'a que peu d'impact sur la section efficace.
- **Normalisation des fonds sur les données** : La normalisation des bruits de fond multijets et  $W$ +jet repose sur les données réelles. Ces erreurs sont relativement grandes du fait de la statistique limitée.  
 Dans le cas du fond multijet, l'incertitude sur la normalisation varie de 50% (électrons) à 100% (muons) avec un impact de l'ordre de 10% sur la section efficace.  
 Dans le cas des fonds  $W$ +jets, la normalisation globale ainsi que les fractions relatives des différentes compositions en saveurs lourdes sont évaluées à partir des données et résultent en l'une des principales erreurs systématiques. L'effet est de l'ordre de 30% sur le résultat final, la contribution majeure provenant de la fraction d'événements  $W + c$ +jets.
- **Luminosité** : La mesure de la luminosité intégrée est affectée d'une erreur de 3.4% avec un impact sur la section efficace du même ordre de grandeur.
- **Taille des échantillons Monte-Carlo** : Finalement, l'incertitude systématique due à la statistique limitée des échantillons Monte-Carlo utilisés dans l'estimation

des efficacités de sélection est prise en compte et contribue pour 12% d'incertitude sur la section efficace.

### 3.3.5 Résultat final

L'extraction de la section efficace se fait au moyen d'une vraisemblance profilée, incluant toutes les incertitudes systématiques comme décrit aux chapitres 1 et 2. L'implémentation HISTFACTORY de ROOFIT est utilisé comme méthode standard dans ATLAS. Les résultats sont validés indépendamment par une analyse bayésienne avec le code ME-FISTO (évaluation sur une grille aléatoire).

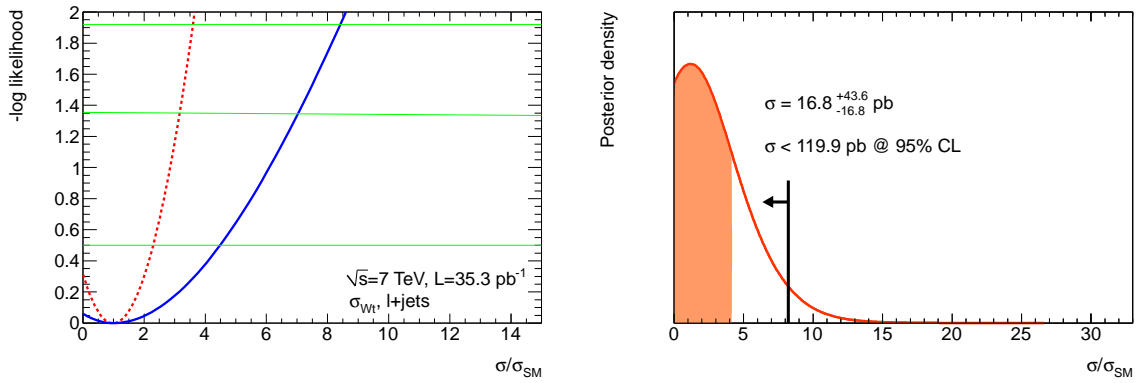


FIG. 3.8: Sensibilité de l'analyse (limite attendue) : à gauche, logarithme des rapports de vraisemblance (pointillé rouge) et de vraisemblance profilée (ligne bleue). Les lignes vertes indiquent les intervalles à 68%, 90% et 95% de confiance ; à droite densité bayésienne *a posteriori*, avec indication de la limite à 95% de confiance. Les courbes sont données en fonction du rapport  $\sigma_{tW}/\sigma_{tW}^{SM}$ .

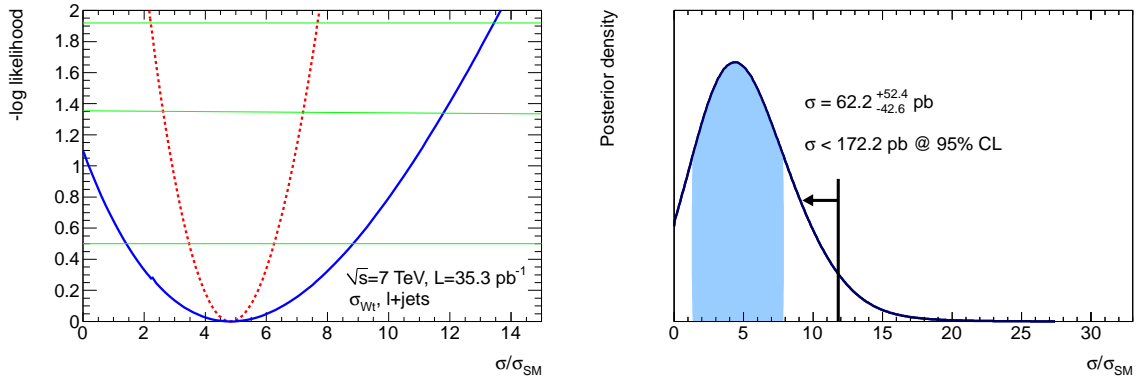


FIG. 3.9: Résultat de l'analyse (limite observée) : à gauche, logarithme des rapports de vraisemblance (pointillé rouge) et de vraisemblance profilée (ligne bleue). Les lignes vertes indiquent les intervalles à 68%, 90% et 95% de confiance ; à droite densité bayésienne *a posteriori*, avec indication de la limite à 95% de confiance. Les courbes sont données en fonction du rapport  $\sigma_{tW}/\sigma_{tW}^{SM}$ .

L'analyse est fortement dominée par les incertitudes systématiques et il n'y a aucune sensibilité attendue au signal  $tW$ . On se contente donc de poser une limite supérieure à 95% de niveau de confiance.

La sensibilité est d'abord estimée en fixant les nombres d'événements observés à la valeur attendue en présence de signal (figure 3.8), puis la véritable mesure est effectuée en utilisant les nombres d'événements effectivement observés (figure 3.9).

On constate que la limite observée est sensiblement moins bonne que le résultat attendu et que le minimum de la fonction de vraisemblance correspond à  $\sigma_{tW}/\sigma_{tW}^{SM} \approx 4$ . Ceci est dû à un excès d'événements dans le canal électron à 3 jets (cf. tableau 3.2), qui contribue majoritairement à la sensibilité de la mesure. Les résultats finaux sont alors :

$$\sigma_{tW}^{attendue} < 122.8 \text{ pb et } \sigma_{tW}^{observée} < 196.0 \text{ pb.} \quad (3.2)$$

L'analyse bayésienne donne des résultats similaires :

$$\sigma_{tW}^{attendue} < 119.9 \text{ pb et } \sigma_{tW}^{observée} < 172.2 \text{ pb.} \quad (3.3)$$

Les résultats obtenus constituent la première mesure jamais réalisée de ce mode de production du quark top et ouvre la voie à la mise en évidence de ce signal après l'accumulation d'un plus grand volume de données.





"Per me si va ne la città dolente,  
per me si va ne l'eterno dolore,  
per me si va tra la perduta gente.  
Giustizia mosse il mio alto fattore;  
fecemi la divina podestate,  
la somma sapienza e 'l primo amore.  
Dinanzi a me non fuor cose create  
se non eterne, e io eterno duro.  
Lasciate ogne speranza, voi ch'intrate."

Dante Aligheri, "Inferno", Canto III.

# 4

## Validation du code Mefisto

### 4.1 Extraction des résultats ATLAS

La recherche de la production associée  $tW$  décrite au chapitre 3.3 va nous servir de base pour illustrer et valider les différents algorithmes présentés au chapitre 2. Cette analyse est complétée par une étude de la production en voie  $t$  [30], qui utilise la même présélection avec des coupures finales différentes : deux jets, pas de veto sur un second jet  $b$ , au moins un jet émis vers l'avant du détecteur ( $|\eta| > 2.5$ ). Ce dernier critère assure l'orthogonalité de la sélection voie  $t$  par rapport à la sélection  $tW$ . Nous allons donc pouvoir combiner ces deux analyses pour extraire simultanément des contraintes sur les sections efficaces de production  $\sigma_t$  et  $\sigma_{tW}$ .

On dispose pour cette analyse de huit canaux : six pour l'analyse  $tW$  et deux pour l'analyse voie  $t$ . Dans cette dernière les canaux électron et muon sont regroupés et les données sont séparées en fonction de la charge du lepton. En effet, du fait de l'asymétrie entre quarks et antiquarks dans les collisions  $pp$ , on attend plus de production de quarks  $t$  que d'antiquarks  $\bar{t}$  (voir diagramme de Feynman 3.4d)

Nous allons analyser ces données de différentes manières, afin de valider l'implémentation de chaque méthode d'analyse de MEFISTO.

Le nombre total d'événements dans le  $i$ -ème canal est donné par :

$$M_i(r_{tW}, r_t, r_{t\bar{t}}) = L \times (n_i^{tW} \cdot r^{tW} + n_i^t \cdot r^t + n_i^{\bar{t}} \cdot r^{\bar{t}} + n_i^{Ew}) + n_i^{Wj} + n_i^{QCD} \quad (4.1)$$

Le modèle comporte trois paramètres  $r$  à ajuster, qui correspondent au rapport de la section efficace de production à celle attendue dans le modèle standard ( $r = \sigma/\sigma_{SM}$ ). Selon l'analyse à réaliser, certains de ces paramètres sont fixés à 1. Nous considérons les cas suivants :

- Analyse  $tW$  : un seul paramètre ajusté,  $r^{tW}$  avec  $r^t = r^{\bar{t}} = 1$ .
- Analyse voie  $t$  : un seul paramètre ajusté,  $r^t$  avec  $r^{tW} = r^{\bar{t}} = 1$ .
- Analyse *single top* combinée : deux paramètres ajustés,  $r^{tW}$  et  $r^t$  avec  $r^{\bar{t}} = 1$ .
- Analyse complète : trois paramètres ajustés,  $r^{tW}$ ,  $r^t$  et  $r^{\bar{t}}$ .

Pour des raisons pratiques, on utilise comme paramètres de nuisance les nombres d'événements attendus de signal et de fond, dans le modèle standard,  $n_i^X$ . Ainsi, les valeurs

TAB. 4.1: Nombre d'événements attendu pour une luminosité de  $35 \text{ pb}^{-1}$  dans les 8 canaux d'analyse

	analyse $tW$ , électron		
	2 jets	3 jets	4 jets
$tW$	$2.3 \pm 0.5$	$2.7 \pm 0.4$	$1.2 \pm 0.2$
voie $t$	$3.7 \pm 0.5$	$1.2 \pm 0.2$	$0.3 \pm 0.1$
$t\bar{t}$	$11.4 \pm 4.0$	$24.6 \pm 5.3$	$23.2 \pm 3.1$
$W$ +jets	$33.1 \pm 19.7$	$9.3 \pm 6.4$	$3.1 \pm 2.7$
Électrofaible	$2.3 \pm 1.5$	$0.9 \pm 0.6$	$0.7 \pm 0.6$
Multijet	$3.5 \pm 3.5$	$1.2 \pm 1.3$	$0.5 \pm 0.5$
Total attendu	$56.3 \pm 12.9$	$40.0 \pm 6.5$	$29.0 \pm 3.5$
Données	49	55	29
	analyse $tW$ , muon		
	2 jets	3 jets	4 jets
$tW$	$2.4 \pm 0.5$	$2.8 \pm 0.4$	$1.2 \pm 0.2$
voie $t$	$4.4 \pm 0.6$	$1.3 \pm 0.2$	$0.3 \pm 0.1$
$t\bar{t}$	$12.5 \pm 4.1$	$27.5 \pm 6.0$	$25.6 \pm 3.8$
$W$ +jets	$38.9 \pm 22.8$	$12.3 \pm 8.6$	$3.9 \pm 3.5$
Électrofaible	$2.4 \pm 1.4$	$1.1 \pm 0.7$	$0.3 \pm 0.2$
Multijet	$5.7 \pm 3.4$	$5.5 \pm 3.3$	$1.2 \pm 1.0$
Total attendu	$66.2 \pm 5.6$	$50.5 \pm 6.9$	$32.6 \pm 3.9$
Données	74	50	37
	analyse voie $t$		
	charge positive	charge négative	
$tW$	$0.1 \pm 0.1$	$0.2 \pm 0.1$	
voie $t$	$10.3 \pm 1.8$	$4.4 \pm 0.8$	
$t\bar{t}$	$0.7 \pm 0.3$	$0.7 \pm 0.4$	
$W$ +jets	$8.8 \pm 5.3$	$5.5 \pm 3.2$	
Électrofaible	$0.4 \pm 0.4$	$0.2 \pm 0.2$	
Multijet	$2.2 \pm 1.6$	$2.1 \pm 1.5$	
Total attendu	$14.9 \pm 3.0$	$8.6 \pm 1.9$	
Données	21	11	

numériques de la luminosité et des sections efficaces sont absorbées dans ces paramètres de nuisance : les valeurs de paramètres  $n^X$  sont données dans la table 4.1. Un dernier paramètre de nuisance  $L$  est maintenu pour propager l'erreur sur la luminosité de manière corrélée entre signal et fonds, exceptés  $W$ +jets et multijet qui sont normalisés directement sur les données. Sa valeur est 1.

Pour simplifier l'analyse, tous les bruits de fonds  $W$ +jets sont regroupés dans la contribution  $n_i^{Wj}$  et leurs erreurs sommées linéairement (corrélation totale). De même les fonds  $Z$ +jets et di-bosons sont regroupés sous l'appellation de fond électrofaible ( $n_i^{Ew}$ ). Enfin, on néglige la contribution de la production *single top* en voie  $s$ .

Les incertitudes de la table 4.1 incluent l'incertitude de luminosité (3.2%), qui est soustraite quadratiquement. Les taux d'événements  $tW$ , voie  $t$  et  $t\bar{t}$ , incluent également l'incertitude théorique sur la section efficace théorique. Quand le rapport de section efficace correspondant est traité comme une variable à ajuster, cette erreur théorique (5% pour  $tW$  et voie  $t$ , 10% pour  $t\bar{t}$ ) est soustraite quadratiquement : en effet, dans ce cas, on cherche à mesurer la section efficace et l'incertitude théorique n'est pas un paramètre de nuisance.

Dans un premier temps, nous reproduisons les analyses ATLAS avec notre modèle d'incertitudes simplifiées : analyse  $tW$  à six canaux et analyse voie  $t$  à deux canaux. Ces analyses sont réalisées par vraisemblance profilée (ou PLR, *Profile Likelihood Ratio*) avec le logiciel HISTFACTORY et par l'intégration bayésienne implémentée dans MEFISTO.

TAB. 4.2: Comparaison entre les analyses publiques ATLAS et l'analyse simplifiée.

	Attendu		Observé	
	PLR	Bayes	PLR	Bayes
Analyse $tW$ : limite à 95% de confiance sur $\sigma_{tW}$				
Résultat ATLAS	123 pb	120 pb	196 pb	172 pb
Analyse simplifiée	114 pb	113 pb	177 pb	166 pb
Analyse voie $t$ : mesure de $\sigma_t$				
Résultat ATLAS	$66^{+49}_{-37}$ pb	$75^{+43}_{-46}$ pb	$53^{+46}_{-36}$ pb	$56^{+43}_{-39}$ pb
Analyse simplifiée	$67^{+52}_{-46}$ pb	$68^{+49}_{-44}$ pb	$55^{+50}_{-45}$ pb	$57^{+45}_{-41}$ pb

Les résultats obtenus sont regroupés dans la table 4.2. Ils sont proches des résultats publiés, même si la simplification opérée conduit à une sous-estimation de l'incertitude finale dans l'analyse  $tW$ , amenant des limites meilleures d'environ 10 pb. Dorénavant, toutes les analyses présentées combineront les huit canaux.

## 4.2 Validation à une dimension

La première étape de validation du code va consister à comparer l'implémentation de la vraisemblance profilée de MEFISTO à celle HISTFACTORY. Ensuite, toujours pour les analyses monodimensionnelles, nous comparerons les résultats bayésien (densité *a posteriori*) et fréquentiste (fonction de vraisemblance profilée). On n'attend pas des résultats identiques, mais cependant, dans le cas d'une analyse poissonnienne, ils devraient être voisins. Cette étape de validation apparaît déjà dans l'analyse ATLAS où l'analyse bayésienne servait à confirmer les résultats de HISTFACTORY. Enfin la dernière étape est la comparaison entre l'intégration bayésienne sur une grille fixe et l'approche MCMC.

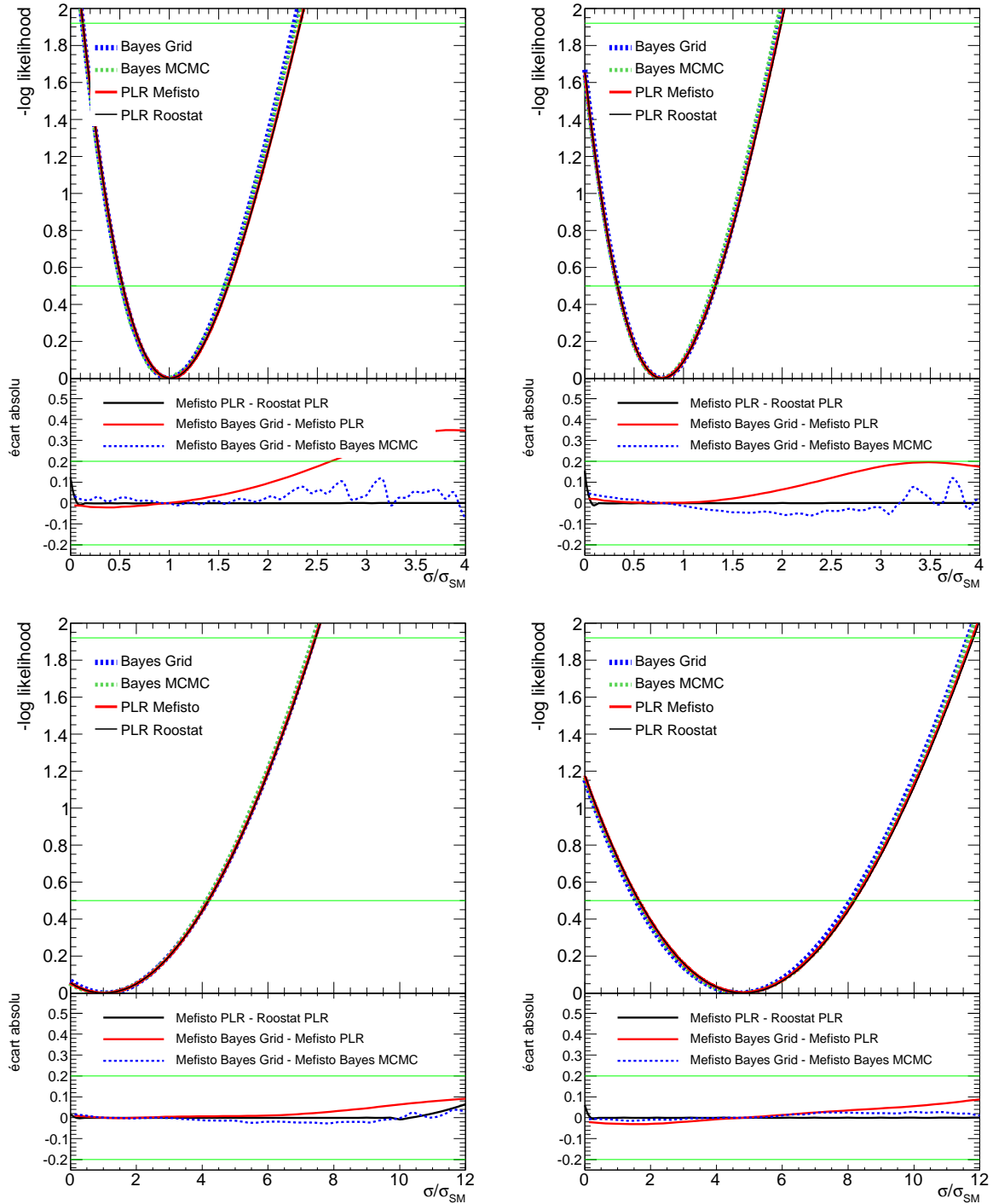


FIG. 4.1: Comparaisons entre différentes méthodes statistiques : PLR dans HISTFACTORY (noir), PLR dans MEFISTO (rouge), intégration bayésienne sur grille dans MEFISTO (pointillé bleu) et intégration bayésienne par MCMC dans MEFISTO (pointillé vert). Les deux analyses, voie  $t$  (en haut) et  $tW$  (en bas), sont représentées pour l'estimation de la sensibilité (à gauche) et la mesure observée (à droite). La partie haute de chaque figure montre le logarithme du rapport de vraisemblance où de la densité *a posteriori* (judicieusement normalisée) ; la partie basse représente les écarts absolus entre différents couples de courbes.

Ces trois comparaisons sont réalisées sur chacune des deux analyses  $tW$  et voie  $t$ , autant pour la sensibilité attendue de l'analyse que pour le résultat observé. Elles sont résumées sur la figure 4.1. En observant les écarts absolus, on constate que MEFISTO et HISTFACTORY donnent exactement le même résultat. Ceci permet de valider la propagation des incertitudes du code MEFISTO, qui est utilisée dans toutes les méthodes statiques (PLR et Bayes). Comme attendu, vraisemblance profilée et intégration bayésienne donnent des résultats différents mais néanmoins similaires, puisque d'un point de vue prudemment statistique ce sont deux objets différents qui mesurent la même information. Enfin, l'échantillonnage par MCMC de la densité *a posteriori* est en excellent accord avec l'intégration classique. Quelques fluctuations apparaissent là où la densité est la plus faible, principalement à cause de la statistique limitée de l'échantillon.

Les courbes obtenues dans chaque analyse sont très voisines de celle déjà observées dans les analyses précédentes les différents canaux d'une analyse n'apportant que peu d'information pour l'autre.

Par étapes successives, on valide ainsi la méthode MCMC qui sera la seule utilisable pour des problèmes à plusieurs dimensions.

### 4.3 Analyse bayésienne multidimensionnelle

Dans le cas précis que nous étudions ici, la simplification des sources d'erreur nous permet d'avoir un temps de calcul suffisamment faible pour réaliser une analyse bayésienne avec une grille fixe bidimensionnelle. On mesure alors simultanément les sections efficaces des processus  $tW$  et voie  $t$ . La figure 4.2 illustre la comparaison entre MCMC et estimation sur une grille, qui sont une nouvelle fois en bon accord. Pour construire les contours à partir de l'échantillonnage MCMC, on utilise d'abord un estimateur de densité de Parzen [31] pour avoir une forme analytique de la densité *a posteriori*, puis on évalue cette fonction sur une grille aléatoire. Les petites différences observées ne sont dues qu'à des effets de résolution de la grille d'estimation de la densité.

Finalement, on utilise le MCMC pour réaliser un ajustement à trois dimensions, en laissant libre la section efficace de l'un des principaux fonds : la production de paires  $t\bar{t}$ . Les résultats sont présentés sur la figure 4.3. La représentation tridimensionnelle est difficilement exploitable, on préférera donc marginaliser la densité *a posteriori*. Les densités marginales pour chaque couple de variables et pour chaque variable individuelle sont données sur la figure 4.4.

On constate une forte anticorrélation entre les processus  $tW$  et  $t\bar{t}$ . Cet effet est attendu : le processus  $t\bar{t}$  étant le fond majeur à l'analyse  $tW$ , si sa section efficace augmente, la part de signal  $tW$  ajustée diminue.

Enfin il est intéressant de comparer la densité *a posteriori* pour les ajustements à une, deux et trois dimensions (après marginalisation sur les autres variables dans les cas multidimensionnels). Les densités sont présentées sur les figures 4.5 et 4.6.

En ajustant plus de paramètres, chaque signal ajusté étant aussi un fond pour les autres, on augmente l'incertitude (donc la largeur de la densité *a posteriori*) sur chaque paramètre, mais le résultat obtenu repose sur moins d'hypothèses. Ceci est, une fois encore, particulièrement visible pour le processus  $tW$  quand la section efficace  $t\bar{t}$  est ajustée : il y a alors une plus grande marge de variation des paramètres compatible avec les données et la distribution *a posteriori* est sensiblement élargie.

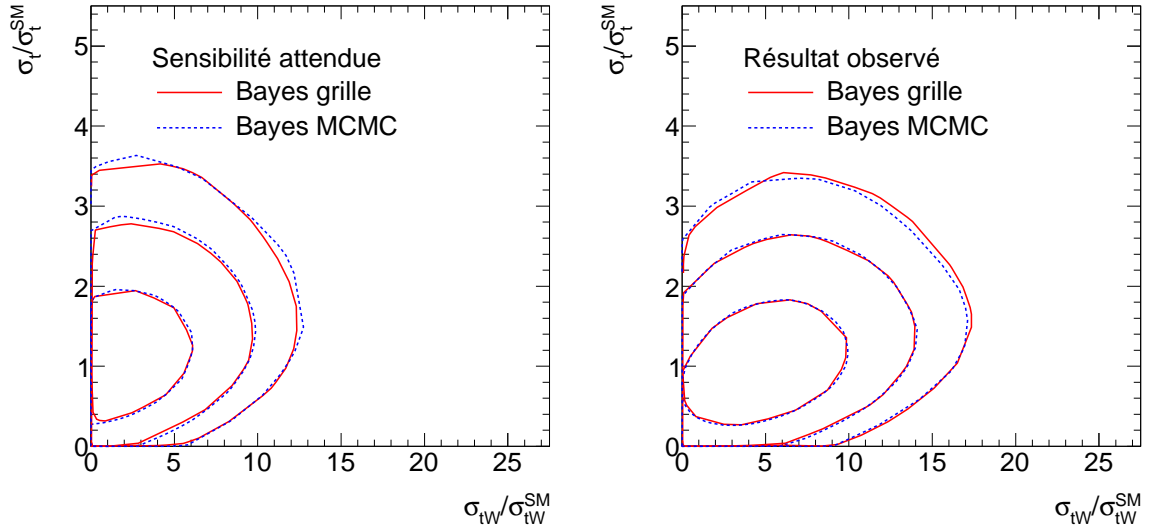


FIG. 4.2: Comparaisons entre les implémentations de la méthode Bayésienne pour l'analyse voie  $t$  vs  $tW$ , pour la sensibilité attendue (à gauche) et le résultat final (à droite). Les contours correspondent à des niveaux de confiance de 68%, 95% et 99%

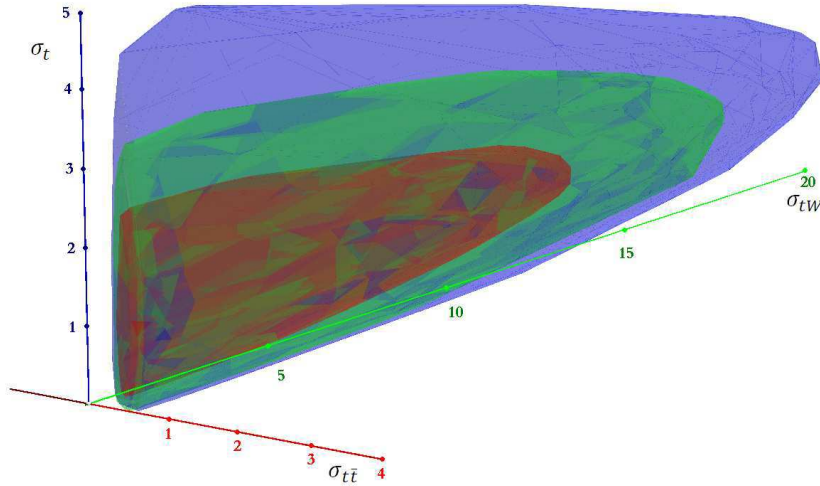


FIG. 4.3: Contours de confiance tridimensionnels (sensibilité attendue) à 68%, 95% et 99% sur les rapports de sections efficaces  $tW$ , voie  $t$  et  $t\bar{t}$ .

Les section efficaces attendues et les incertitude à  $1\sigma$  correspondantes sont :

$$\sigma_{tW}^{1D} = 14.9_{-14.9}^{+34.3} \text{ pb} , \sigma_{tW}^{2D} = 17.5_{-17.5}^{+36.2} \text{ pb} , \sigma_{tW}^{3D} = 22.8_{-22.8}^{+56.8} \text{ pb} ; \quad (4.2)$$

$$\sigma_t^{1D} = 65.6_{-33.2}^{+39.9} \text{ pb} , \sigma_t^{2D} = 73.5_{-35.1}^{+42.3} \text{ pb} , \sigma_t^{3D} = 76.1_{-37.1}^{+39.1} \text{ pb}. \quad (4.3)$$

À partir de ces résultats, on constate également que la valeur centrale de la section efficace attendue ne correspond pas à celle du modèle standard. Ceci est dû à la marginalisation. La vraie valeur centrale ne peut s'obtenir qu'en considérant la distribution avant marginalisation. Ainsi le maximum de la densité tridimensionnelle de la figure 4.3 correspond bien à  $r^{tW} = r^t = r^{t\bar{t}} = 1$ .

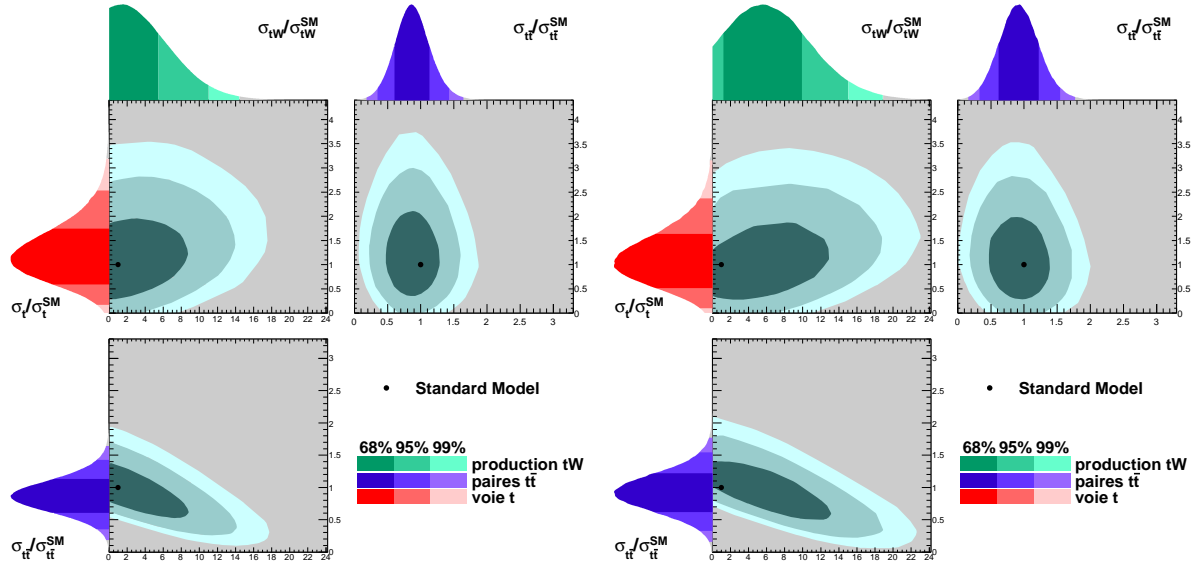


FIG. 4.4: Densités marginales uni- et bidimensionnelles correspondant à l'analyse 3D pour la sensibilité attendue (à gauche) et le résultat observé (à droite). Les points noirs indiquent la valeur attendue dans le modèle standard ( $r^{tW} = r^t = r^{t\bar{t}} = 1$ )

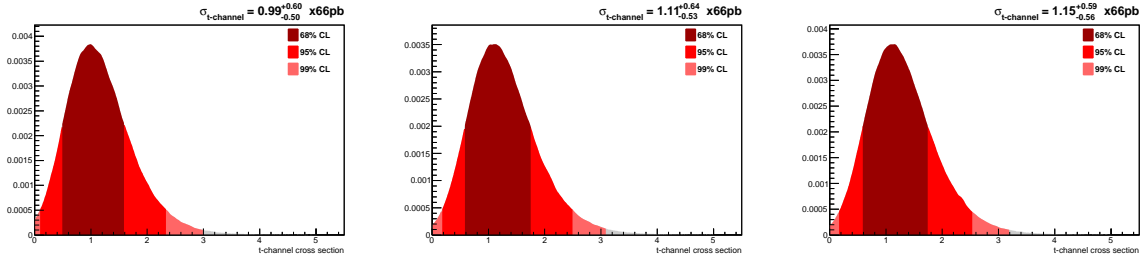


FIG. 4.5: Densités *a posteriori* sur la section efficace voie  $t$  (sensibilité attendue), pour les ajustements 1D (gauche), 2D (centre) et 3D (droite).

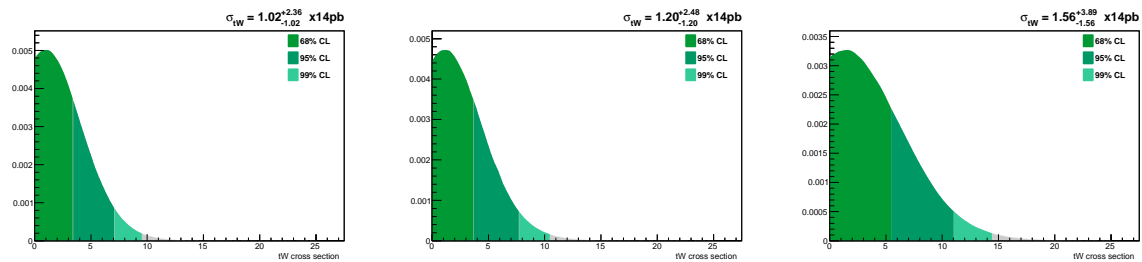


FIG. 4.6: Densités *a posteriori* sur la section efficace voie  $tW$  (sensibilité attendue), pour les ajustements 1D (gauche), 2D (centre) et 3D (droite).

Le résultat final de l'analyse tridimensionnelle, avec les données observée est alors :

$$\sigma_{tW}^{3D} = 80.0 \text{ pb} , \sigma_t^{3D} = 60.2 \text{ pb} , \sigma_{t\bar{t}}^{3D} = 152.7 \text{ pb.} \quad (4.4)$$

L'incertitude sur ces grandeurs est donnée par les contours tridimensionnels. La section efficace  $t\bar{t}$  est la grandeur la mieux contrainte par l'ajustement. En effet, c'est le processus



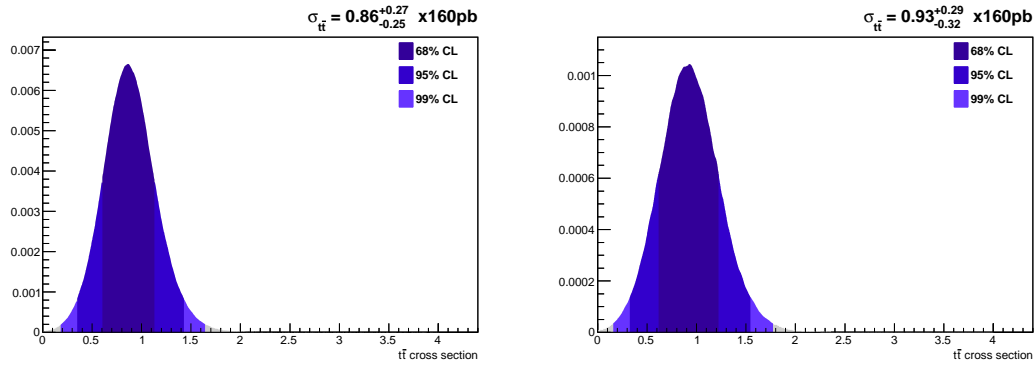


FIG. 4.7: Densités marginales *a posteriori* sur la section efficace  $t\bar{t}$  : sensibilité attendue (gauche) et résultat observé (droite).

pour lequel on attend le plus d'événements. On constate que, même si ce paramètre est laissé libre, la valeur ajustée après marginalisation (voir figure 4.7) est compatible avec le modèle standard. De plus les résultats tridimensionnels sont voisins de ceux des analyses individuelles, à l'exception de l'incertitude sur l'analyse  $tW$ .

On a ainsi validé sur un exemple simple mais complet les différents outils du code MEFISTO : d'abord en vérifiant le rapport de vraisemblance profilée avec HISTFACTORY, puis en comparant entre eux les différents algorithmes (vraisemblance profilée, intégration bayésienne et échantillonnage par MCMC) à une, deux et trois dimensions.

*Arma virumque cano, Troiae qui primus ab oris Italiam, fato profugus, Laviniaque venit litora, multum ille et terris iactatus et alto vi superum saevae memorem Iunonis ob iram;*

Virgile, “*Eneide*” I.1

# 5

## Perspectives : analyse combinée des 3 voies *single top*

### 5.1 *Single top* et nouvelle physique

La production électrofaible du quark top est sensible au vertex  $Wtb$ . L’existence de nouvelles particules dans des extensions du modèle standard est susceptible de modifier les sections efficaces de production. De plus, en fonction des modèles considérés, les sections efficaces des trois modes de production électrofaible seront modifiées différemment.

#### 5.1.1 Processus *benchmark*

Quelques exemples de nouveaux phénomènes physiques ont été étudiés dans la référence [32]. Seules les sections efficaces en voie  $t$  et  $s$  sont considérées. Les paramètres proposés pour les modèles ne sont pas les plus réalistes et, dans certains cas, sont déjà exclus. Néanmoins, cette étude permet de montrer comment différents modèles de nouvelle physique, introduisant soit de nouvelles particules soit de nouveaux couplages, modifient les sections efficaces :

- *Bosons de jauge supplémentaires* : les théories de grande unification incluent le modèle standard dans une symétrie de jauge plus large (groupes  $SU(5)$ ,  $SO(10)$  ou  $E_6$  par exemple). Cette symétrie serait brisée à basse énergie et se réduirait aux symétries du modèle standard. Une telle symétrie de jauge introduit naturellement des nouveaux bosons de jauge massifs,  $W'$  et  $Z'$ , qui se mélangent avec les bosons de jauge électrofaibles. L’exemple considéré dans [32] est le modèle *top-flavour* qui introduit une nouvelle symétrie  $SU(2)_h$  responsable de la masse élevée du quark top. Le nouveau boson  $W'$  se couple préférentiellement au quark top. Un tel boson de jauge peut augmenter la section efficace dans la voie  $s$  avec un faible impact dans la voie  $t$  ( $m_{W'} \gg m_W$ ). Les sections efficaces sont calculées pour  $m_{W'} = 1$  TeV et  $\sin^2 \phi = 0.05$  (un angle de mélange entre  $W$  et  $W'$ ).
- *Bosons scalaires supplémentaires* : les bosons scalaires apparaissent notamment dans les mécanismes de brisure spontanée de symétrie. Dans certains modèles, dits *tech-*

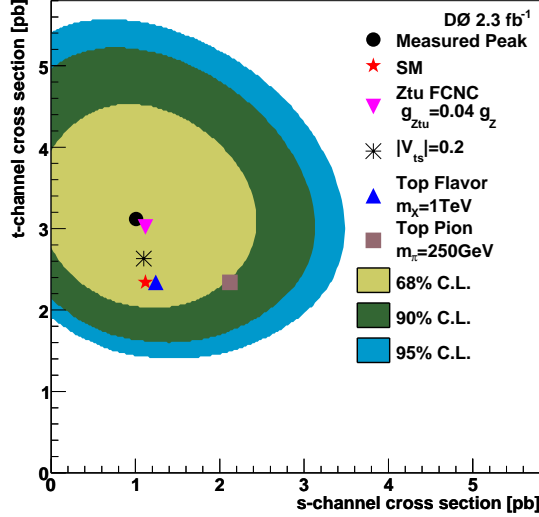


FIG. 5.1: Contraintes simultanées sur les sections efficaces en voie  $t$  et  $s$  mesurées par l'expérience  $DØ$  et comparaison avec différentes hypothèses de physique au-delà du Modèle Standard.

*nicouleurs*, les bosons scalaires tels le boson de Higgs sont décrits par des états liés de deux fermions, en analogie avec les états liés d'interaction forte (pions,  $\rho$ ,...). Dans le modèle *top-color* présenté ici, un état lié  $t\bar{b}$  dit top-pion peut jouer ce rôle. Une telle particule peut contribuer à la production *single top*. Comme dans le cas d'un boson  $W'$ , la section efficace dans la voie  $s$  peut être augmentée, sans affecter la section efficace dans la voie  $t$ . Les sections efficaces sont calculées pour un top-pion de masse  $m_\pi = 250$  GeV.

- *Quarks supplémentaires (quatrième famille)* : il est possible de rajouter une quatrième famille de quarks ( $t', b'$ ), qui se mélange aux trois autres familles par une généralisation de la matrice CKM. Les conditions d'unitarité de la matrice CKM à trois familles ne sont alors plus respectées. Sans ces conditions d'unitarité, les éléments de la matrice CKM mesurés aujourd'hui sont bien moins contraints. En particulier la contrainte sur  $|V_{ts}|$  devient  $0 < |V_{ts}| < 0.55$  [33] ( $0.034 < |V_{ts}| < 0.046$  [33] pour trois familles). Une valeur de  $|V_{ts}|$  élevée pourrait augmenter la section efficace de production *single top* dans la voie  $t$  car les densités de partons sont plus élevées pour le quark  $s$  que pour le quark  $b$ . Pour satisfaire l'unitarité de la matrice à quatre familles,  $|V_{tb}|$  serait alors réduit, ce qui diminuerait également la section efficace *single top* dans la voie  $s$ . Les valeurs de sections efficaces dans la table 5.1 correspondent au cas extrême,  $|V_{ts}| = 0.5$  et  $|V_{tb}| = 0.84$ . Ces contraintes sur  $|V_{ts}|$  sont celles données dans [32] qui date de l'année 2001. Les contraintes actuelles sur  $|V_{ts}|$  [1], sans condition d'unitarité, sont  $0 < |V_{ts}| < 0.11$ , ce qui exclut a priori ce cas particulier.
- *Couplages supplémentaires* : un dernier modèle envisage l'existence de couplages supplémentaires qui n'interviendraient qu'au delà de l'échelle d'énergie de la brisure électrofaible, notamment des couplages par courant neutre modifiant la saveur (*FCNC* : *Flavour Changing Neutral Currents*) via des vertex  $Ztc$ ,  $\gamma tc$  ou  $g tc$ . L'effet de ces nouveaux couplages pourrait se faire particulièrement sentir dans le secteur du quark top, dans la production de *single top* par exemple. Dans la voie  $t$ , la présence d'un quark  $c$  au lieu d'un quark  $b$  dans la voie d'entrée augmenterait la

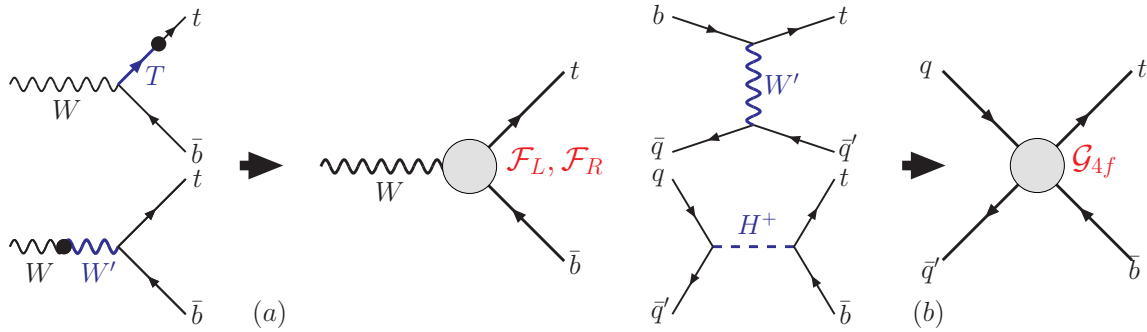


FIG. 5.2: Couplages effectifs prenant en compte de la nouvelle physique : (a) modification du vertex  $Wtb$  par des mélanges entre  $W$  et nouveau vecteur  $W'$  ou entre quark top et une quatrième famille  $T$  ; (b) Couplage à quatre fermions remplaçant des propagateurs massifs ( $W'$ , Higgs chargé ...) en voie  $t$  ou en voie  $s$ .

section efficace. Les sections efficaces du tableau 5.1 correspondent à un couplage  $Ztc$  de même amplitude que le couplage électrofaible du modèle standard. Ces données ont été utilisées par l'expérience  $D\bar{O}$  pour donner une interprétation des mesures de section efficace *single top* en terme de nouvelle physique [34], le résultat étant donné sur la figure 5.1.

### 5.1.2 Paramétrisation générique

Une autre publication plus récente [35] propose une paramétrisation plus générique de l'effet de nouveaux phénomènes physiques par l'ajout d'un lagrangien effectif incluant 3 paramètres. Les deux premiers rendent compte des modifications du vertex  $Wtb$  (boucles, corrections au vertex) qui se traduisent par deux couplages anormaux : un couplage vectoriel gauche ( $\mathcal{F}_L$ ) et un couplage vectoriel droit ( $\mathcal{F}_R$ ). Ces couplages vont influencer les trois sections efficaces. Le couplage droit est fortement contraint par la désintégration  $b \rightarrow s\gamma$  et peut être négligé. Il ne reste alors que le couplage gauche à considérer. Ce dernier ne fait que modifier la règle de Feynman du vertex  $Wtb$  par un facteur multiplicatif et a donc le même impact sur les trois sections efficaces

Le dernier paramètre prend en compte la présence de nouveaux bosons lourds couplés aux quarks  $t$  et  $b$  dans le propagateur en voie  $s$  ou  $t$ . Pour des bosons sensiblement plus lourds que l'échelle d'énergie du processus de diffusion, soit pour une échelle de nouvelle physique supérieure au TeV, on peut réduire ces propagateurs à un couplage effectif à quatre fermions ( $\mathcal{G}_{4f}$ ). Il faut noter qu'un tel vertex effectif est de la forme  $-\frac{G^2}{\Lambda^2}\bar{\psi}_1\psi_2\bar{\psi}_t\psi_b$  où  $\Lambda$  est l'échelle d'énergie de nouvelle physique. Pour manipuler une constante sans dimension on pose  $\mathcal{G}_{4f} = G/v$  où  $v = 246$  GeV est la valeur dans le vide du champ de Higgs. Une échelle de nouvelle physique de l'ordre du TeV avec un couplage voisin du couplage électrofaible ( $g_W \sim 0.4$ ) donne  $\mathcal{G}_{4f} \sim 0.01$ . Les vertex du lagrangien effectif sont représentés sur la figure 5.2.

Les sections efficaces *single top* au LHC à 14 TeV peuvent alors s'écrire en fonction des deux couplages effectifs  $\mathcal{F}_L$  et  $\mathcal{G}_{4f}$  :

$$\begin{aligned}
 \sigma_t &= \sigma_t^0 (1 + 4\mathcal{F}_L - 3.06\mathcal{G}_{4f}) \\
 \sigma_{tW} &= \sigma_{tW}^0 (1 + 4\mathcal{F}_L) \\
 \sigma_s &= \sigma_s^0 (1 + 4\mathcal{F}_L + 19.69\mathcal{G}_{4f})
 \end{aligned} \tag{5.1}$$

où les  $\sigma_X^0$  sont les sections efficaces dans le modèle standard. On constate en particulier que la présence de propagateurs lourds a un important effet en voie  $s$  qui est anticorrélé à la modification de la voie  $t$

A partir de ces relations et des sections efficaces en voies  $t$  et  $s$  données dans la référence [32], on peut déterminer les sections efficaces  $tW$  correspondantes. Ces sections efficaces apparaissent dans le tableau 5.1.

Modèle	$\sigma_t$ [32]	$\sigma_s$ [32]	$\sigma_{tW}$	$\mathcal{F}_L$	$\mathcal{G}_{4f}$
Modèle standard	246.6	10.7	66	0	0
Nouveau vecteur (TopFlavor)	246.6	35.4	86.7	0.078	0.102
Nouveau scalaire (TopColor)	246.6	17.6	71.8	0.022	0.029
Couplage anormaux (FCNC)	395.0	10.7	100.4	0.130	-0.026
4 <sup>ème</sup> famille	339.1	7.5	84.8	0.071	-0.029

TAB. 5.1: Sections efficaces de production *single top* dans différentes extensions du modèle standard et couplages effectifs correspondants, pour des collisions  $pp$  à  $\sqrt{s} = 14$  TeV.

Une fois les processus de production *single top* mis en évidence, une mesure précise des sections efficaces devrait permettre de contraindre les paramètres de ces modèles. Plus encore, c'est une mesure simultanée des trois sections efficaces qui sera susceptible de contraindre une éventuelle nouvelle physique.

## 5.2 Analyses *CSC*

### 5.2.1 Sélections

Il n'existe pas, à l'heure actuelle, d'analyse concernant la mesure de la section efficace en voie  $s$  dans ATLAS, car la faible section efficace (environ 4 pb à  $\sqrt{s} = 7$  TeV) n'avait aucune chance d'être observée avec les données accumulées en 2010 et 2011. Il faut remonter à des travaux plus anciens, pour trouver un ensemble de chiffres consistents sur la mesure de chacune des trois sections efficaces. L'exercice *CSC* (Computing and Software Challenge) avait pour but de tester la chaîne informatique de simulation et d'analyse de l'expérience ATLAS. Entre 2006 et 2008, il a permis de réévaluer la sensibilité du détecteur dans l'étude de nombreux phénomènes physiques [36] dans les collisions  $pp$  à  $\sqrt{s} = 14$  TeV.

Les trois modes de production électrofaible du quark top ont fait l'objet d'analyses dans le cadre de cet exercice. Dans le document publié, chaque analyse se concentre sur un unique canal et considère les deux autres modes de production comme un bruit de fond, en utilisant la valeur de section efficace du modèle standard pour estimer le taux de fond attendu. À des fins de prospective, nous proposons une fois de plus de combiner ces différentes analyses pour ajuster simultanément les différents paramètres.

Les détails des sélections de chaque analyse ne seront pas donnés ici. Les sélections pour la voie  $t$  et la production associée  $tW$  sont sensiblement similaires à celles discutées aux chapitres 3 et 4. En voie  $s$ , la signature consiste en un lepton, de l'énergie manquante et deux jets  $b$ . Pour les événements à deux jets, le fond dominant va provenir des processus  $W$ +jets. On va utiliser la présence d'un deuxième jet  $b$  pour rejeter ce fond, au prix d'une perte d'efficacité de sélection du signal, due à l'efficacité d'identification des jets  $b$  (environ 60%). La mesure de cette section efficace va donc nécessiter une importante luminosité.

En complément de ces sélections, chaque analyse était affinée par une analyse discriminante multivariée. Pour chacun des principaux bruits de fond, l'information de plusieurs variables cinématiques faiblement discriminantes est combinée en une unique variable. Pour ce faire, il existe de nombreuses méthodes, les plus couramment utilisées en physique des particules étant les rapports de vraisemblance, les réseaux de neurones (perceptron multicouche), et les arbres de décision boostés (BDT) [37]. Les analyses *CSC* en voie *t* et *tW* (voir figure 5.3) reposent sur l'utilisation d'arbres de décision, alors que l'analyse en voie *s* utilise des rapports de vraisemblance. Les distributions attendues pour chacun des quatre discriminants, pour les événements à trois jets, apparaissent sur la figure 5.3 pour une luminosité intégrée de  $1 \text{ fb}^{-1}$ .

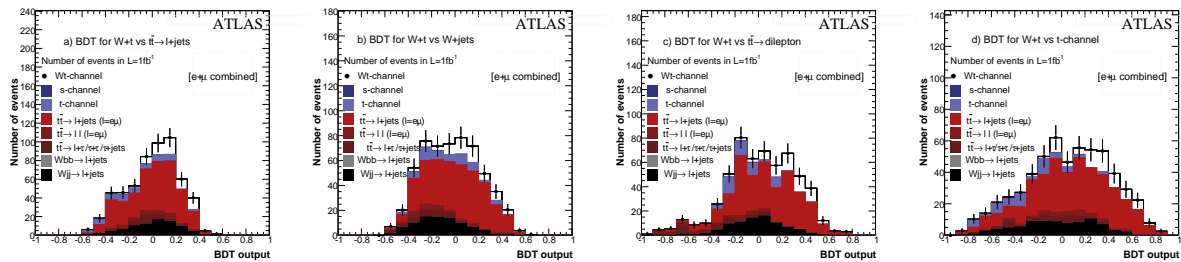


FIG. 5.3: Distributions des discriminants pour les événements à trois jets dans l'analyse *tW*. a) BDT contre  $t\bar{t} \rightarrow \ell\bar{\nu}_\ell j j b\bar{b}$ ; b) BDT contre *W*+jets; c) BDT contre  $t\bar{t} \rightarrow \ell^+\ell^-\nu_\ell\bar{\nu}_\ell b\bar{b}$ ; d) BDT contre *single top* en voie *t*.

Les analyses finales appliquent des coupures sur chaque discriminant qui sont optimisées pour chaque analyse individuelle. Les sélections initiales n'ont été optimisées qu'en prenant en compte la maximisation du rapport  $S/\sqrt{B}$  qui compare le nombre d'événements de signal attendu *S* à une fluctuation statistique (poissonnienne) du bruit de fond  $\sqrt{B}$ . Les effets systématiques ne sont pas pris en compte. En revanche, les coupures sur les différents discriminants multivariés ont été optimisées en incluant les principaux effets systématiques (calibration en énergie des jets et identification des jets *b*), en utilisant la méthode approchée de maximum de vraisemblance décrite au chapitre 2.2.2 pour estimer et minimiser l'erreur finale.

Les coupures des analyses individuelles sont reprises sans autre optimisation dans l'analyse combinée. Au total, 6 canaux d'analyses sont ainsi pris en compte : 2, 3 et 4 jets avec la sélection *tW*, 2 jets (étiquetés *b*) en voie *s*, ainsi que 2 et 3 jets en sélection voie *t*.

## 5.2.2 Analyse combinée

Deux modèles sont utilisés pour estimer le nombre d'événements. Le premier est analogue à celui utilisé au chapitre 4, à la différence que la production de paires  $t\bar{t}$  devient un bruit de fond et que les processus *W*+jets sont estimés entièrement par la simulation Monte-Carlo et donc normalisés à la luminosité. Le modèle utilisé pour le calcul du nombre d'événements attendu est une généralisation du précédent :

$$M_i(r_t, r_{tW}, r_s) = L \times (n_i^t \cdot r^t + n_i^{tW} \cdot r^{tW} + n_i^s \cdot r^s + n_i^{t\bar{t}} + n_i^{Wj+Ew}) \quad (5.2)$$

Le second modèle utilise les contraintes des équations 5.1. Les sections efficaces du modèle standard sont alors utilisées dans la normalisation du nombre d'événements attendu et

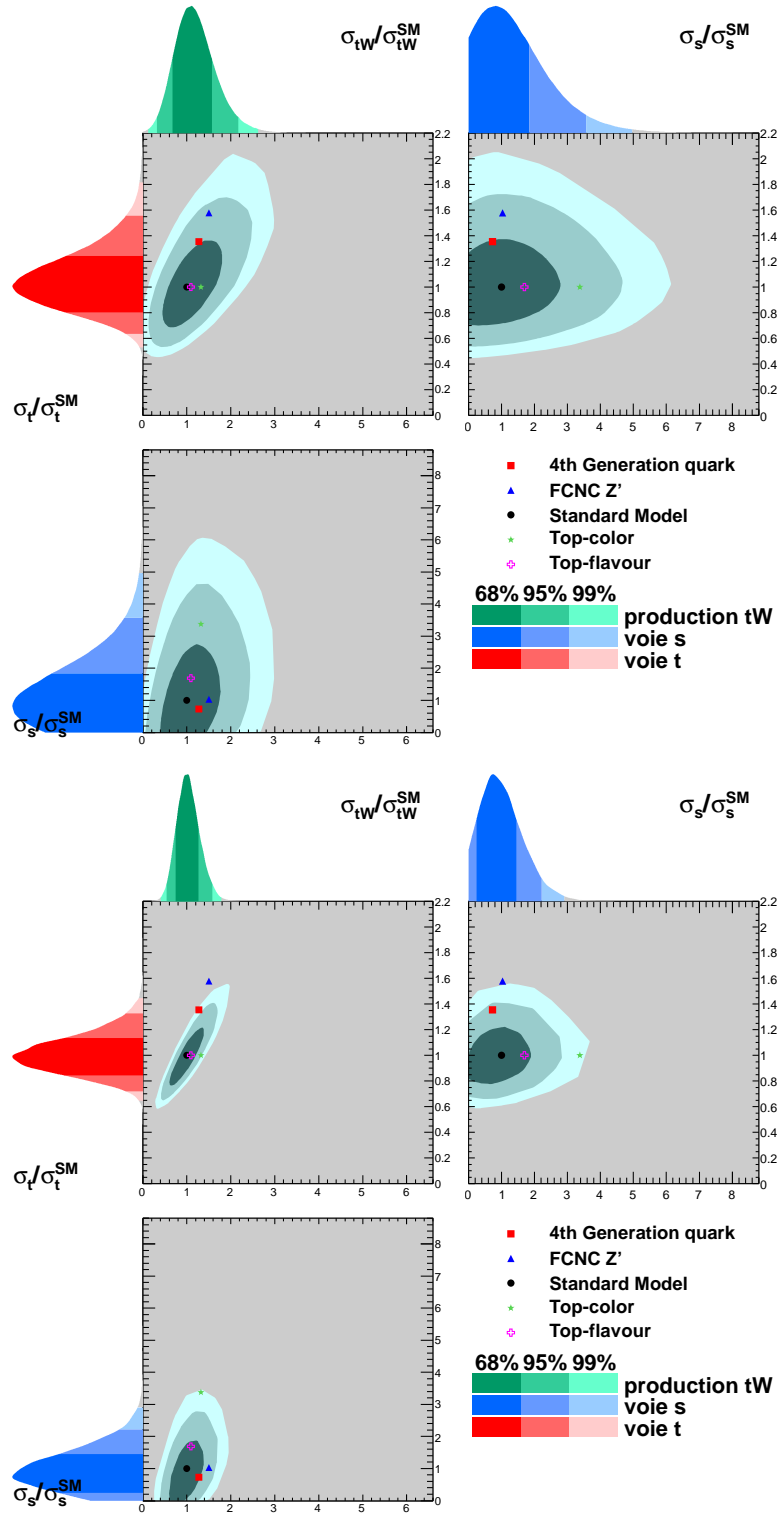


FIG. 5.4: Combinaison des trois analyses CSC *single top* ( $\sqrt{s}=14$  TeV), en ajustant indépendamment les 3 sections efficaces pour  $1 \text{ fb}^{-1}$  (en haut) et  $100 \text{ fb}^{-1}$  (en bas). Sur chaque figure, les densités marginales à deux dimensions sont représentées sous forme de contours de confiance *HPD* à 68%, 95% et 99% ( $\sigma_{tW}$  vs.  $\sigma_s$  en haut à gauche,  $\sigma_t$  vs.  $\sigma_s$  en haut à droite et  $\sigma_{tW}$  vs.  $\sigma_t$  en bas à gauche).

Les densités marginales de chaque variable sont représentées projetées sur les axes correspondants, avec les intervalles de confiance *HPD*.

Les valeurs des sections efficaces de la table 5.1, pour le modèle standard (point noir) et plusieurs extensions du modèle standard, sont indiquées sur les contours bidimensionnels.

traitées comme des paramètres de nuisance supplémentaires.

$$\begin{aligned}
M_i(\mathcal{F}_L, \mathcal{G}_{4f}) = L \times ( & n_i^t + n_i^{tW} + n_i^s + n_i^{t\bar{t}} + n_i^{Wj+EW} \\
& + 4\mathcal{F}_L(n_i^t + n_i^{tW} + n_i^s) \\
& + \mathcal{G}_{4f}(19.69n_i^s - 3.06n_i^t)
\end{aligned} \tag{5.3}$$

Dans cette étude, une attention particulière a été apportée aux effets liés à l'imprécision sur la calibration en énergie des jets et sur la mesure de l'efficacité d'identification des jets de quarks  $b$ . En considérant une erreur relative de 5% dans les deux cas, on a réévalué les efficacités de sélection des événements de signal et de fond dans chaque canal d'analyse :

- en modifiant l'impulsion de tous les jets de +5% et de -5%, cette correction étant propagée sur l'énergie manquante ;
- en modifiant la coupure sur le poids d'étiquetage des jets  $b$  de façon à varier l'efficacité d'étiquetage de +5% et -5% (la variation correspondante du taux de mauvaise identification était de l'ordre de 20%).

L'impact sur les taux d'événements attendus pour chacune de ces systématiques varie de 5% à 15% selon les canaux et les processus.

Dans les deux cas la variation relativement à la valeur centrale des paramètres  $n_i^X$  permet de définir un vecteur d'erreur comme décrit au paragraphe 1.4.2. Les asymétries et les (anti)-corrélations entre canaux sont prises en compte dans l'analyse.

D'autres effets corrélés entre signal et fonds et liés aux outils de génération (densités de partons, radiations dans les états initial et final, modèles d'hadronisation ...) et de simulation du détecteur ont été étudiés séparément et se traduisent par une erreur relative globale de 5% sur les signaux et 10% sur les fonds. Une erreur supplémentaire de 10% sur les sections efficaces théoriques des processus est appliquée aux seuls fonds.

L'incertitude systématique liée à statistique limitée de données Monte-Carlo est également prise en compte. C'est une source d'incertitude qui affecte de manière décorrélée signal et fond dans chacun des 6 canaux. Elle est donc traitée comme  $6 \times 5 = 30$  sources d'erreurs indépendantes n'affectant chacune qu'un unique paramètre.

La luminosité n'est affectée que d'une seule incertitude liée à sa mesure. Cette source d'incertitude n'influence aucunement les efficacités de sélection du signal et du fond. Sa valeur est supposée de 5%.

L'effet de l'empilement des événements n'est pas simulé dans les données Monte-Carlo utilisées. Ce n'est pas à proprement parler une erreur systématique mais un biais puisque cet effet sera inclu dans les simulations utilisées au moment de l'analyse. Ce biais a été estimé sur des événements  $t\bar{t}$  simulés avec et sans empilement. La présence d'empilement conduit à une réduction de 20% du nombre d'événements sélectionnés. Pour corriger le biais on a donc multiplié par un facteur 0.8 tous les taux d'événements attendus, sans affecter d'incertitude à ce terme correctif.

Les nombres moyens d'événements attendus  $n_i^X$  sont estimés pour une luminosité de  $1 \text{ fb}^{-1}$  et une énergie de 14 TeV dans le centre de masse. Finalement, on réalise pour chaque modèle deux analyses : une pour  $L = 1$  ( $1 \text{ fb}^{-1}$ ) et une pour  $L = 100$  ( $100 \text{ fb}^{-1}$ ) en diminuant les systématiques d'un facteur 2. Cette dernière situation correspond à ce qu'on pourrait attendre en fin de fonctionnement du LHC.

Dans toutes les analyses, une densité *a posteriori* uniforme, positive, est utilisée pour chacune des trois sections efficaces. Cette densité n'a un effet notable que sur la mesure de la voie  $s$  qui est compatible avec 0 vue la faible sensibilité de l'analyse. Les erreurs attendues (intervalles *HPD* à 68% de niveau de confiance) sont extraites des densités



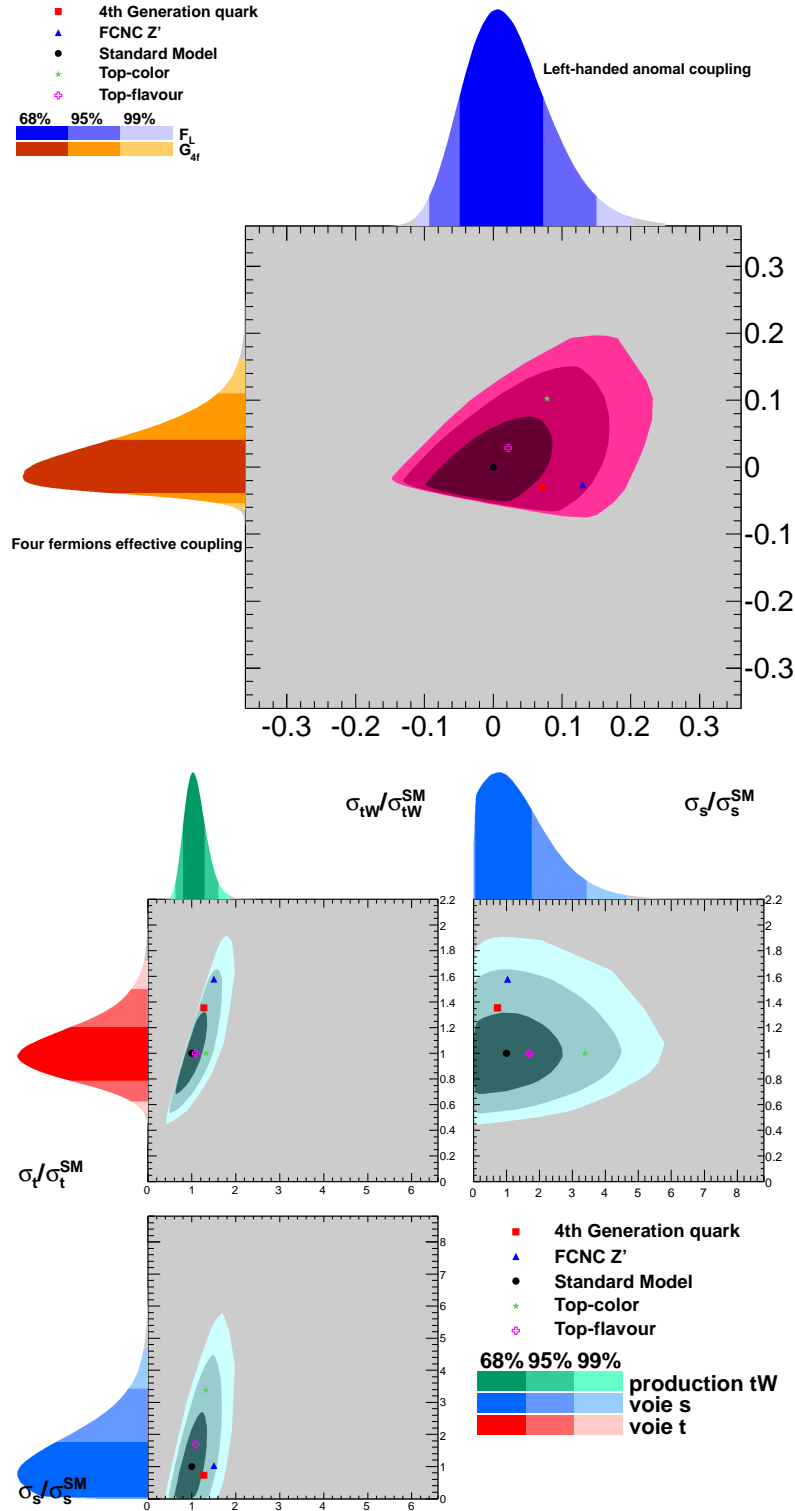


FIG. 5.5: Combinaison des trois analyses CSC *single top* ( $1 \text{ fb}^{-1}$ ,  $\sqrt{s}=14$  TeV), en ajustant uniquement les couplages  $\mathcal{F}_L$  et  $\mathcal{G}_{4f}$ . La figure du haut présente les contraintes obtenues sur les couplages, la figure du bas le résultat pour les sections efficaces. Les densités marginales à deux dimensions sont représentées sous forme de contours de confiance *HPD* à 68%, 95% et 99% ( $\sigma_{tW}$  vs.  $\sigma_s$  en haut à gauche,  $\sigma_t$  vs.  $\sigma_s$  en haut à droite et  $\sigma_{tW}$  vs.  $\sigma_t$  en bas à gauche).

Les densités marginales de chaque variable sont représentées projetées sur les axes correspondants, avec les intervalles de confiance *HPD*.

Les valeurs des sections efficaces de la table 5.1 pour le modèle standard (point noir) et plusieurs extensions du modèle standard sont indiquées sur les contours bidimensionnels.

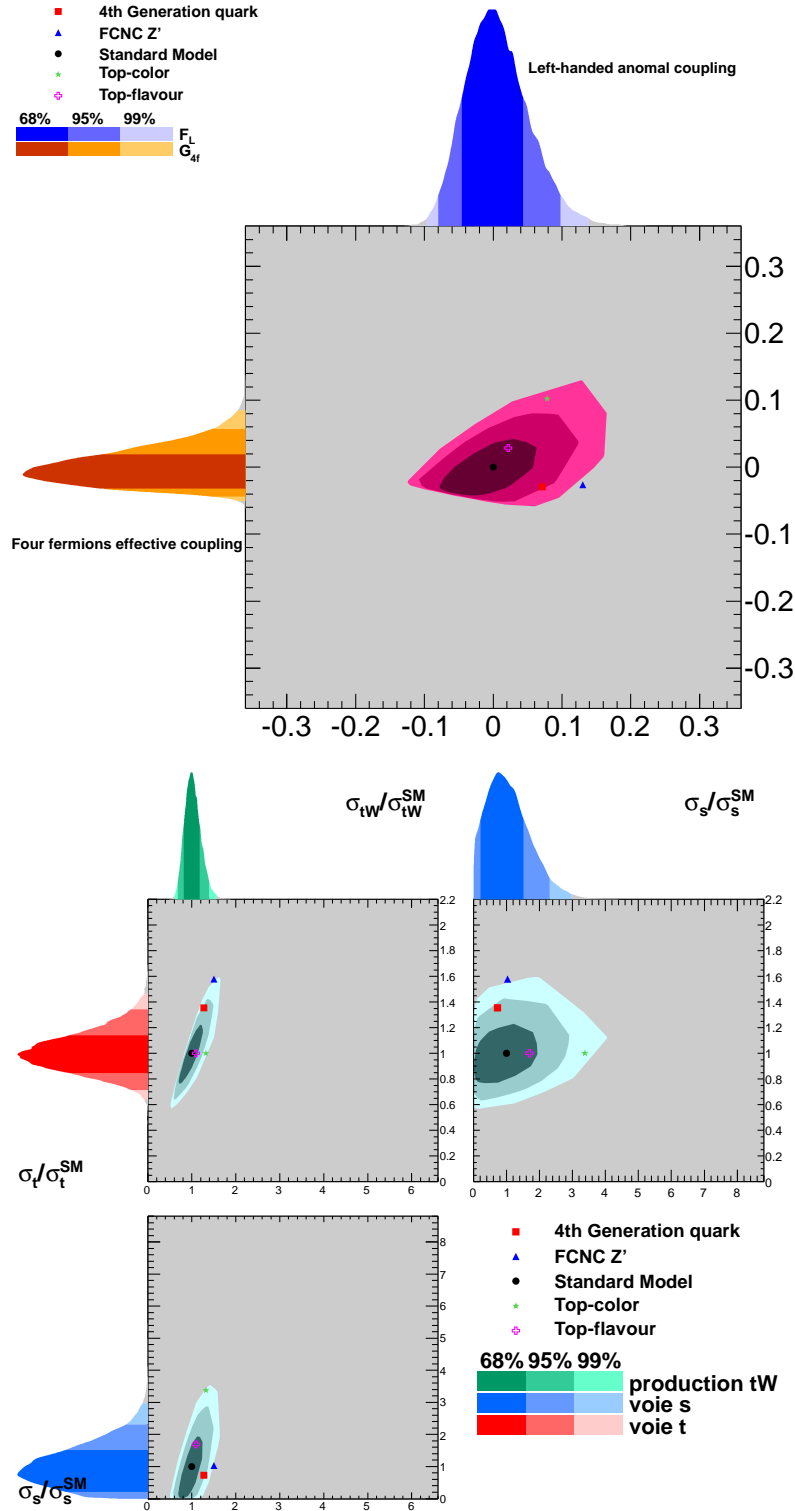


FIG. 5.6: Combinaison des trois analyses CSC *single top* ( $100 \text{ fb}^{-1}$ ,  $\sqrt{s}=14 \text{ TeV}$ ), en ajustant uniquement les couplages  $\mathcal{F}_L$  et  $\mathcal{G}_{4f}$ . La figure du haut présente les contraintes obtenues sur les couplages, la figure du bas le résultat pour les sections efficaces. Les densités marginales à deux dimensions sont représentées sous forme de contours de confiance *HPD* à 68%, 95% et 99% ( $\sigma_{tW}$  vs.  $\sigma_s$  en haut à gauche,  $\sigma_t$  vs.  $\sigma_s$  en haut à droite et  $\sigma_{tW}$  vs.  $\sigma_t$  en bas à gauche).

Les densités marginales de chaque variable sont représentées projetées sur les axes correspondants, avec les intervalles de confiance *HPD*.

Les valeurs des sections efficaces de la table 5.1 pour le modèle standard (point noir) et plusieurs extensions du modèle standard sont indiquées sur les contours bidimensionnels.

marginales de chaque variable. Les résultats avec le modèle libre pour  $1 \text{ fb}^{-1}$  sont alors :

$$\sigma^t = 248.1_{-50.1}^{+58.4} \text{ pb}, \sigma^{tW} = 72.9_{-28.7}^{+30.6} \text{ pb} \text{ et } \sigma^s = 8.9_{-8.9}^{+10.7} \text{ pb},$$

et pour  $100 \text{ fb}^{-1}$  :

$$\sigma^t = 244.8_{-36.5}^{+35.0} \text{ pb}, \sigma^{tW} = 67.6_{-18.6}^{+16.1} \text{ pb} \text{ et } \sigma^s = 7.9_{-5.3}^{+7.6} \text{ pb}.$$

Pour  $1 \text{ fb}^{-1}$ , les erreurs statistiques sont encore importantes, tandis qu'avec une luminosité 100 fois plus grande, l'analyse est entièrement dominée par les effets systématiques. Encore une fois, du fait de la marginalisation, on observe un biais dans les sections efficaces mesurées. Ce biais reste inférieur à la dispersion de la densité *a posteriori*. L'intérêt de ces résultats marginalisés réside plutôt dans l'estimation d'une incertitude de mesure pour chaque paramètre. Le résultat complet serait le contour tridimensionnel, qui n'est pas donné ici, et, dans une moindre mesure, l'ensemble des marginalisations à une et deux dimensions qui sont présentées sur la figure 5.4.

En n'utilisant que les deux couplages effectifs dans l'ajustement, on contraint mieux les sections efficaces, en particulier la production  $tW$  : en effet, ce processus n'est sensible qu'au couplage effectif gauche, qui lui même sera contraint par la mesure de la voie  $t$  qui est la plus précise (effet principalement statistique). Les valeurs des couplages après marginalisation sont alors :

$$\mathcal{F}_L = 0.006_{-0.056}^{+0.067} \text{ et } \mathcal{G}_{4f} = -0.013_{-0.025}^{+0.054},$$

A partir de ces valeurs, on redétermine les valeurs des sections efficaces :

$$\sigma^t = 241.6_{-47.6}^{+55.6} \text{ pb}, \sigma^{tW} = 67.8_{-14.8}^{+17.5} \text{ pb} \text{ et } \sigma^s = 8.6_{-8.1}^{+10.2} \text{ pb}.$$

Pour  $100 \text{ fb}^{-1}$ , ces résultats deviennent :

$$\mathcal{F}_L = -0.005_{-0.040}^{+0.049} \text{ et } \mathcal{G}_{4f} = -0.010_{-0.021}^{+0.030},$$

et :

$$\sigma^t = 244.8_{-36.2}^{+36.8} \text{ pb}, \sigma^{tW} = 64.6_{-10.9}^{+13.3} \text{ pb} \text{ et } \sigma^s = 8.0_{-5.8}^{+8.1} \text{ pb}.$$

Pour  $1 \text{ fb}^{-1}$  les erreurs statistiques sont encore importantes, alors qu'avec une luminosité 100 fois plus grande, l'analyse est entièrement dominée par les effets systématiques. Ici aussi, le résultat complet est donné par le contour à deux dimensions dans l'espace  $(\mathcal{F}_L, \mathcal{G}_{4f})$ , où le mode de la densité *a posteriori* se situe bien en  $(0, 0)$ . Ces contours, ainsi que les marginalisations des paramètres et les contours de confiance pour les sections efficaces sont présentés sur la figure 5.5 ( $1 \text{ fb}^{-1}$ ) et 5.6 ( $100 \text{ fb}^{-1}$ ).

### 5.3 Conclusions

Il ressort de ces analyses que la mesure simultanée des sections efficaces *single top* devrait permettre de contraindre les paramètres de nouvelle physique et d'exclure certains modèles. Il est nécessaire de souligner que les estimations des taux d'événements et des effets systématiques utilisées dans cette étude ont été réalisées avant le démarrage effectif du LHC et sont donc sujets à caution. Déjà, les résultats récents d'ATLAS sur la mesure de la voie  $t$  seule dépasse une sensibilité de  $5\sigma$  avec moins de  $5 \text{ fb}^{-1}$  à  $\sqrt{s} = 7 \text{ TeV}$ . De même la mesure du processus  $tW$ , en utilisant les désintégrations dileptoniques en plus

de l'analyse lepton+jets présentée dans ce document atteint une sensibilité de  $3\sigma$ . Au vu de ces résultats, il devient raisonnable d'envisager une première analyse combinée des trois modes de production *single top* sur les 15 à 20  $\text{fb}^{-1}$  à  $\sqrt{s} = 8$  TeV attendus pour la fin de l'année 2012. La sensibilité à de la nouvelle physique restera faible mais pourrait déjà avoisiner les résultats estimés pour 100  $\text{fb}^{-1}$  à  $\sqrt{s}=14$  TeV dans le cadre de l'analyse *CSC*.

Néanmoins il faudra vraisemblablement attendre le redémarrage du LHC fin 2014, après réparation des aimants supraconducteurs, pour atteindre une énergie et une luminosité suffisamment élevées. Ce n'est qu'alors que de telles analyses deviendront réellement pertinentes en atteignant une sensibilité suffisante pour contraindre, voire mettre en évidence, l'existence de nouveaux phénomènes.



# Conclusion

Ce mémoire présente et illustre une partie des travaux récents de la communauté de la physique des particules et des astroparticules concernant l'intégration d'un grand nombre de sources d'incertitude systématique dans une analyse statistique la plus rigoureuse possible. Ces méthodes étant très coûteuses en calcul numérique, il est crucial de disposer d'outils fiables et rapides pour permettre à la fois d'optimiser une analyse et de contraindre simultanément plusieurs paramètres. A cet effet deux méthodes ont été développées dans le cadre du programme MEFISTO.

D'une part, une méthode approchée de maximum de vraisemblance, qui bien qu'en toute rigueur fautive, permet d'obtenir très rapidement (quelque seconde pour des problèmes complexes) une approximation de l'incertitude globale. L'intérêt premier de cet outil est l'optimisation des paramètres d'une analyse par exemple par le choix des coupures de sélection. Si la valeur de l'incertitude obtenue n'est pas correcte, cet outil va néanmoins permettre de comparer la sensibilité attendue de différentes versions de l'analyse en minimisant l'incertitude finale. Il s'agit d'une version améliorée du rapport signal sur racine du bruit, qui inclut les effets systématiques avec leurs corrélations. Cette méthode a été utilisée dans les études préliminaires de sensibilité du détecteur ATLAS pour la recherche de la production électrofaible du quark top, mais également pour la mesure du couplage  $Wtb$  (thèse de Julien Labbé [38]). Ces aspects n'ont été discutés que très brièvement dans le présent document pour se focaliser sur le second outil, qui utilise par ailleurs le résultat de cette méthode de vraisemblance approximée comme point de départ.

La première méthode d'estimation proposée ne permet de donner qu'une estimation du résultat attendu mais pas de traiter correctement l'analyse de données réelles. Pour ce faire, nous avons développé un second outil permettant l'estimation de paramètres dans l'approche bayésienne, par l'utilisation de chaînes de Markov. L'intégration de nombreux paramètres de nuisance, de densité connue, nécessiterait de n'utiliser les chaînes de Markov que pour échantillonner la densité *a posteriori* après marginalisation. L'outil proposé permet de faire les deux simultanément en utilisant plusieurs chaînes pondérées et ainsi de traiter de manière exacte (dans la limite d'un grand échantillon) l'intégration bayésienne sur les paramètres de nuisance en échantillonnant la densité *a posteriori*.

Pour valider cet outil et l'ensemble du code MEFISTO, une série de tests ont été proposés sur une réanalyse des résultats d'ATLAS présenté à l'hiver 2011. Le premier est la validation de la modélisation des propagations d'erreur en comparant les vraisemblances profilées obtenues avec MEFISTO et ROOSTAT (programme utilisé dans l'analyse ATLAS) pour un problème monodimensionnel (un seul paramètre ajusté). Le second consiste en la comparaison entre vraisemblance profilée et intégration bayésienne sur une grille fixe, pour le même problème unidimensionnel. Le troisième et dernier test vise à valider la méthode par chaînes de Markov développée spécifiquement en la comparant à l'intégration bayésienne sur une grille fixe sur des problèmes uni- et bidimensionnels. Après validation, nous avons utilisé l'approche par chaînes de Markov pour analyser la production électrofaible

du quark top dans les données ATLAS (deux section efficaces), en ajustant également l'un des principaux bruits de fonds physiques, la production  $t\bar{t}$ , soit un ajustement simultané de trois paramètres.

L'approche par chaîne de Markov permet justement de traiter des problèmes multidimensionnels. De tels outils seront importants pour mettre en évidence l'existence d'une physique au-delà du modèle standard, qui est l'un des objectifs majeurs du LHC. Une étude prospective d'extension des mesures de section efficace de production *single top* illustre comment l'ajustement combiné de plusieurs sections efficaces ou de plusieurs couplage autorisera une meilleur contrainte des paramètres de nouvelle physique. Bien évidemment, ces outils devraient pouvoir trouver de nombreuses autres applications au LHC et dans d'autres expériences.

# Bibliographie

- [1] K. Nakamura et al. (Particle Data Group), *J. Phys. G* **37**, 075021 (2010)
- [2] A. Stuart, K. Ord, S. Arnold, “Kendall’s Advanced Theory of Statistics, vol 2A, Classical Inference and the Linear Model”, Hodder Arnold (1999)
- [3] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Eur.Phys.J.* **C71** (2011)
- [4] R. Barlow, “Asymmetric Systematic Errors”, MAN-HEP-04-02 (2004).
- [5] V. Werkerke, D. Kirkby in “Statistical Problems in Particle Physics, Astrophysics and Cosmology”, Proceedings PHYSTAT 2005, Imperial Coll. Press, (2006)
- [6] F. James, “MINUIT : Function Minimization and Error Analysis”, CERN Program Library Long Writeup D506, (1994)
- [7] B. Clément, “Production électrofaible du quark top au Run II de l’expérience D0”, Thèse de doctorat, Université Louis Pasteur, IPHC-06-004, ULP-5086, FERMILAB-THESIS-2006-06 (2006).
- [8] A. O’Hagan, J. Forster, “Kendall’s Advanced Theory of Statistics, vol 2B, Bayesian Inference”, Hodder Arnold (2004)
- [9] A. Putze, “Phénoménologie et détection du rayonnement cosmique nucléaire”, Thèse de doctorat, Université Joseph Fourier, LPSC09145 (2009)
- [10] Amos Breskin and Rudiger Voss (editors), “The CERN large hadron collider : Accelerator and experiments ” (2 volumes), JINST 3 S08001-S08007 (2008)
- [11] F. Abe *et al.*, (CDF Collaboration), “Observation of Top Quark Production in  $p\bar{p}$  Collisions,” *Phys. Rev. Lett.* **74**, 2626 (1995);  
S. Abachi *et al.*, (DØ Collaboration), “Observation of the Top Quark,” *Phys. Rev. Lett.* **74**, 2632 (1995).
- [12] Tevatron Electroweak Working Group (DØ +CDF), “Combination of CDF and DØ results on the mass of the top quark using up to  $5.8 \text{ fb}^{-1}$  of data”, FERMILAB-TM-2504-E, [hep-ex/1107.5255v3] (2011)
- [13] V. M. Abazov *et al.* (D0 Collaboration), “Measurement of the Top Quark Pair Production Cross Section in the Lepton+Jets Channel in Proton-Antiproton Collisions at  $\sqrt{s} = 1.96 \text{ TeV}$ ”, *Phys. Rev. D* **84**, 012008 (2011),  
V. M. Abazov *et al.* (D0 Collaboration), “Measurement of the  $t\bar{t}$  Production Cross Section Using Dilepton Events in  $pp$  Collisions” *Phys. Lett. B* **704** (2011) 403; et publications équivalentes de la collaboration CDF
- [14] T. Aaltonen *et al.* (CDF Collaboration) “Exclusion of an Exotic Top Quark with  $-4/3$  Electric Charge Using Soft Lepton Tagging”, *Phys. Rev. Lett.* **105**, 101801 (2010).  
V. M. Abazov *et al.* (D0 Collaboration) “Experimental Discrimination between Charge  $2e/3$  Top Quark and Charge  $4e/3$  Exotic Quark Production Scenarios”, *Phys. Rev. Lett.* **98** (2007) 041801



- [15] V. M. Abazov *et al.* (D0 Collaboration) “Measurements of Single Top Quark Production Cross Sections and  $|V_{tb}|$  in  $pp$  Collisions at  $\sqrt{s} = 1.96$  TeV”, Phys. Rev. D **84** (2011) 112001
- [16] T. Aaltonen *et al.* (CDF and D0 Collaborations) “Combination of CDF and D0 Measurements of the W Boson Helicity in Top Quark Decays”, soumis à Phys. Rev. D, arXiv :1202.5272 (2012)
- [17] T. Aaltonen *et al.* (CDF Collaboration), “Measurement of the fraction of  $t\bar{t}$  production via gluon-gluon fusion in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$ -TeV,” Phys. Rev. D **79** (2009) 031101 [arXiv :0807.4262 [hep-ex]].
- [18] V. M. Abazov *et al.* (D0 Collaboration), “Observation of Single Top-Quark Production”, Phys. Rev. Lett. **103**, 092001 (2009) ;  
T. Aaltonen *et al.* (CDF Collaboration), “Observation of Electroweak Single Top-Quark Production”, Phys. Rev. Lett. **103**, 092002 (2009).
- [19] M. Aliev *et al.* “HATHOR HAdronic Top and Heavy quarks crOss section calculatoR”, arXiv :1007.1327
- [20] S. Frixione, B. Weber and P. Nason, “Single-top production in MC@NLO”, hep-ph/0512250 and hep-ph/08053067 (2002)
- [21] N. Kidonakis *et al.*, Phys. Rev. D **68** (2003) 114014
- [22] Z. Sullivan, Phys. Rev. D **70** (2004) 114012 J. Campbell, R.K. Ellis, F. Tramontano, Phys. Rev. D **70** (2004) 094012 J. Campbell, F. Tramontano, Nucl. Phys. B **726** (2005) 109-130
- [23] R. Brun *et al.*, “GEANT - Detector Description and Simulation,” CERN Program Library Long Writeup W 5013 (1994).
- [24] S. Frixione and B.R. Webber, “Matching NLO QCD computations and parton shower simulations”, JHEP **0206** (2002) 029 [hep-ph/0204244]
- [25] M.L. Mangano *et al.*, “ALPGEN, a Generator for Hard Multiparton Processes in Hadronic Collisions,” J. High Energy Physics **0307**, 001 (2003).
- [26] G. Corcella *et al.*, “HERWIG 6 : An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes),” JHEP **01** 010, hep-ph/0011363 (2001).
- [27] J. Pumplin *et al.*, J. High Energy Phys. JHEP07(2002)012 (2002).
- [28] V.M. Abazov *et al.* (D0 Collaboration), “ $t\bar{t}$  production cross-section in  $p\bar{p}$  collisions at 1.8 TeV”, Phys.Rev.D **67** (2003) 012004.
- [29] C. Weydert, “Recherche d’un boson de Higgs chargé avec le détecteur ATLAS : de la théorie à l’expérience”, Thèse de doctorat, Université de Grenoble (2011)
- [30] The Atlas Collaboration, “Searches for Single Top-Quark Production with the ATLAS Detector in  $pp$  Collisions at  $\sqrt{s} = 1.96$  TeV”, ATLAS-CONF-2011-027 (2011).  
B. Alvarez *et al.*, ATLAS-COM-CONF-2011-026
- [31] Harison Prosper, “Multivariate Techniques”, in SOS’08 Proceedings, EPJ Web of Conferences, vol 4 (2010)
- [32] T. Tait *et al.* C.-P. Yuan, “Single Top Quark Production as a Window to Physics Beyond the Standard Model”, Phys. Rev. D **63**, 014018 (2001).
- [33] L. Montanet *et al.*, Phys. Rev. D **50**, 1173 (1994).

- 
- [34] V.M. Abazov *et al.* (D0 Collaboration), “Model-independent measurement of t-channel single top quark production in ppbar collisions at  $\sqrt{s}=1.96$  TeV”, Phys.Lett. B705 (2011)
  - [35] Q.-H. Cao, J. Wudka et C.-P. Yuan, “Search for New Physics via Single Top Production”, Phys.Lett. B 658 (2007)
  - [36] G Aad *et al.* (The ATLAS Collaboration) , Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics (3 volumes), CERN-OPEN-2008-020, arXiv :0901.0512v4 [hep-ex] (2009)
  - [37] Yann Coadou, “Decision trees”, in SOS’08 Proceedings, EPJ Web of Conferences, vol 4 (2010)
  - [38] J. Labbé, “Préparation de l’expérience ATLAS : Etalonnage électronique du calorimètre électromagnétique, Mesure de la polarisation des bosons  $W$  dans la décroissance des quarks top.”, Thèse de doctorat, Université Joseph Fourier, LPSC0986 (2006).

