



**HAL**  
open science

# Les techniques Monte Carlo par chaînes de Markov appliquées à la détermination des distributions de partons

Yémalin Gabin Gbedo

► **To cite this version:**

Yémalin Gabin Gbedo. Les techniques Monte Carlo par chaînes de Markov appliquées à la détermination des distributions de partons. Physique Nucléaire Théorique [nucl-th]. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAY059 . tel-01743752

**HAL Id: tel-01743752**

**<https://theses.hal.science/tel-01743752>**

Submitted on 26 Mar 2018

**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.

## **THÈSE**

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES**

Spécialité : Physique Théorique

Arrêté ministériel : 25 mai 2016

Présentée par

**YÉMALIN GABIN GBEDO**

Thèse dirigée par **Mariane MANGIN-BRINET**, CNRS

préparée au sein du **Laboratoire de Physique Subatomique et Cosmologie**  
dans l'**École Doctorale de Physique**

## **Les techniques Monte Carlo par chaînes de Markov appliquées à la détermination des distributions de partons**

### **Markov chain Monte Carlo techniques applied to parton distribution functions determination: Proof of concept**

Thèse soutenue publiquement le **22 Septembre 2017**,  
devant le jury composé de :

**Monsieur JOHANN COLLOT**

PROFESSEUR, UNIVERSITE GRENOBLE ALPES, Président

**Monsieur DAVID MAURIN**

CHARGE DE RECHERCHE CNRS, LPSC, Examineur

**Monsieur ALEXANDRE GLAZOV**

CHERCHEUR, UNIVERSITE DE HAMBOURG, Examineur

**Monsieur ROBERTO BONCIANI**

PROFESSEUR ASSOCIE, UNIVERSITE DE ROME, Rapporteur

**Monsieur MARIO CAMPANELLI**

PROFESSEUR ASSOCIE, UNIVERSITE COLLEGE DE LONDRES,  
Rapporteur

**Monsieur JEAN-CHRISTIAN ANGLES-D'AURIAC**

DIRECTEUR DE RECHERCHE CNRS, INSTITUT NEEL,  
Examineur





# Remerciements

Je suis très reconnaissant à Mariane Mangin-Brinet pour avoir accepté de superviser ce travail. Je n'oublierai jamais sa disponibilité, ses conseils avisés et sa rigueur qui m'ont sincèrement aidé à réaliser ce travail.

Mille mercis à Aurélien Barrau pour m'avoir accueilli à Grenoble et pour ses multiples soutiens.

Je remercie Johann Collot, Mario Campanelli, Roberto Bonciani, Alexandre Glazov, Jean-Christian Anglès-d'Auriac et David Maurin pour avoir accepté de participer à mon jury de thèse.

Je tiens ensuite à remercier Voica Radescu, Ringaile Placakyte et Stefano Camarda pour les échanges fructueux que nous avons eus.

Mon infinie reconnaissance à la fondation CFM pour la Recherche pour m'avoir soutenu financièrement ces trois dernières années, ainsi qu'à l'ensemble des membres du groupe de théorie du LPSC.

Je voudrais associer à ces remerciements mes amis, mes frères et soeurs dans leur soutiens spirituels, ainsi que mon fils Tadagbé Junias GBEDO et bien évidemment mon épouse Adjo Josiane AGONGLOVI pour son amour patient, son soutien sans faille et sa capacité à trouver les mots justes pour me remonter le moral dans les moments difficiles.

Enfin, un sincère merci à mes parents pour leurs soutiens indéfectibles et encouragements tout au long de mes années d'études ; puissiez-vous trouver à travers ce travail un véritable motif de satisfaction de vos multiples sacrifices et conseils.



# Résumé

Nous avons développé une nouvelle approche basée sur les méthodes Monte Carlo par chaînes de Markov pour déterminer les distributions de partons et quantifier leurs incertitudes expérimentales. L'intérêt principal d'une telle étude repose sur la possibilité de remplacer la minimisation standard avec MINUIT de la fonction  $\chi^2$  par des procédures fondées sur les méthodes statistiques et sur l'inférence Bayésienne en particulier, offrant ainsi une meilleure compréhension de la détermination des distributions de partons. Après avoir examiné ces techniques Monte Carlo par chaînes de Markov, nous introduisons l'algorithme que nous avons choisi de mettre en œuvre, à savoir le Monte Carlo hybride (ou Hamiltonien). Cet algorithme, développé initialement pour la chromodynamique quantique sur réseau, s'avère très intéressant lorsqu'il est appliqué à la détermination des distributions de partons par des analyses globales. Nous avons montré qu'il permet de contourner les difficultés techniques dues à la grande dimensionnalité du problème, en particulier celle relative au taux d'acceptance. L'étude de faisabilité réalisée et présentée dans cette thèse indique que la méthode Monte Carlo par chaînes de Markov peut être appliquée avec succès à l'extraction des distributions de partons et à leurs incertitudes expérimentales.

Mots clés : Chromodynamique Quantique, Distributions de Partons, Méthodes Monte Carlo par Chaînes de Markov, Monte Carlo Hamiltonien.



# Abstract

We have developed a new approach to determine parton distribution functions and quantify their experimental uncertainties, based on Markov chain Monte Carlo methods. The main interest devoted to such a study is that we can replace the standard  $\chi^2$  MINUIT minimization by procedures grounded on statistical methods, and on Bayesian inference in particular, thus offering additional insight into the rich field of parton distribution functions determination. After reviewing these Markov chain Monte Carlo techniques, we introduce the algorithm we have chosen to implement—namely hybrid (or Hamiltonian) Monte Carlo. This algorithm, initially developed for lattice quantum chromodynamics, turns out to be very interesting when applied to parton distribution functions determination by global analyses; we have shown that it allows to circumvent the technical difficulties due to the high dimensionality of the problem, in particular concerning the acceptance rate. The feasibility study performed and presented in this thesis, indicates that Markov chain Monte Carlo method can successfully be applied to the extraction of parton distribution functions and of their experimental uncertainties.

Key words : Quantum Chromodynamics, Parton Distribution Functions, Markov chain Monte Carlo, Hamiltonian Monte Carlo.





# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Résumé</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>Table des matières</b>	<b>ix</b>
<b>Liste des figures</b>	<b>xi</b>
<b>Liste des tableaux</b>	<b>xv</b>
<b>Introduction</b>	<b>1</b>
<b>1 La chromodynamique quantique</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Les bases de la chromodynamique quantique . . . . .	5
1.3 La chromodynamique quantique perturbative . . . . .	11
1.4 Les schémas de saveurs lourdes . . . . .	22
1.5 L'analyse globale des PDFs en QCD . . . . .	24
1.6 Les incertitudes expérimentales des PDFs . . . . .	29
<b>2 La méthode Monte Carlo par chaînes de Markov</b>	<b>37</b>
2.1 Introduction . . . . .	37
2.2 La formulation de la détermination des distributions de partons en terme d'inférence Bayésienne . . . . .	37
2.3 Le principe du Monte Carlo par chaînes de Markov . . . . .	38
2.4 L'algorithme Monte Carlo Hamiltonien (MCH) . . . . .	41
2.5 L'analyse des chaînes de Markov . . . . .	43
<b>3 Implémentation de l'algorithme MCH dans le code public HERAFitter, validation et premiers résultats</b>	<b>49</b>
3.1 Introduction . . . . .	49
3.2 Les paramètres de la simulation et le "setup" . . . . .	51
3.3 Les valeurs et les densités de probabilité marginales des paramètres libres des PDFs . . . . .	52
3.4 Les densités de probabilité marginales des PDFs : introduction aux limites de confiance des PDFs . . . . .	57
3.5 Les distributions de partons et leurs incertitudes expérimentales . . . . .	58



# Liste des figures

1.1	Courbe montrant le comportement du couplage fort $g_s$ en fonction de l'énergie $q^2$ . Figure tirée de <b>AURENCHE et collab. [2016]</b> . . . . .	9
1.2	Schéma typique d'une interaction $eP$ . . . . .	13
1.3	La fonction de structure $F_2^P$ en fonction de $Q^2$ pour différentes valeurs de $x$ , déterminées à partir des données combinées de diffusions inélastiques profondes des électrons et positrons sur les protons par l'échange d'un photon $\gamma$ pour $Q^2 \geq 2\text{GeV}^2$ , par les collaborations H1 + ZEUS de HERA. Les données des expériences sur cible fixe (SLAC, BCDMS, E665 et NMC) sont montrées à titre de comparaison (la figure est tirée de <b>OLIVE et AL [2014]</b> ). On observe la dépendance de $F_2^P(x, \alpha_s(Q^2))$ avec $Q^2$ , comme prédit par la QCD perturbative. En pratique, les données correspondantes à la section efficace réduite (équation(1.39)) sont ajustées, plutôt que les fonctions de structure $F_2$ et $F_L$ séparément. . . . .	19
1.4	Différences entre le schéma ACOT et TR. Figure tirée de <b>xFITTER DEVELOPERS [2017]</b> . . . . .	24
1.5	Courbes montrant la valeur centrale de la distribution du gluon (1.84) aux échelles $Q^2 = 1.9 \text{ GeV}^2$ (gauche) et $Q^2 = 10 \text{ GeV}^2$ (droite) dans le schéma où tous les quarks lourds sont sans masses (1.4.1). . . . .	29
1.6	Courbes montrant la distribution du gluon $xg$ (gauche) et la distribution du quark de valence $xu_{\text{val}}$ (droite) dans le même schéma (1.4.1). . . . .	30
1.7	Comparaison des incertitudes des distributions de partons $xg, xu_{\text{val}}, xd_{\text{val}}$ et $xS$ à l'échelle $Q^2 = 10\text{GeV}^2$ ; extraites à partir des données de la collaboration ZEUS en utilisant les méthodes "Offset" et Hessienne. Figure tirée de <b>COOPER-SARKAR et GWENLAN [2005]</b> . . . . .	32
2.1	L'algorithme de Metropolis-Hastings. . . . .	40
2.2	Discrétisation des équations de Hamilton par la Méthode Leapfrog. . . . .	42
2.3	L'approximation de la dynamique Hamiltonienne $H(q, p) = q^2/2 + p^2/2$ par la méthode Leapfrog. L'état initial utilisé est $q = 0, p = 1$ . . . . .	42
2.4	Les étapes de l'algorithme Monte Carlo Hamiltonien . . . . .	43
2.5	Valeur du paramètre $B_g$ en fonction du temps Monte Carlo $t$ pour trois chaînes de Markov indépendantes. Les valeurs initiales de $\frac{\chi^2}{\text{d.o.f}}$ pour chacune des trois chaînes de Markov 1, 2 et 3 correspondent respectivement à $\frac{\chi^2}{\text{d.o.f}} = 67.44$ , $\frac{\chi^2}{\text{d.o.f}} = 0.87$ et $\frac{\chi^2}{\text{d.o.f}} = 81.58$ . Nous identifions clairement sur ce graphe la région de thermalisation, qui est limitée aux $\sim 100-210$ premières itérations. . . . .	44

2.6	Valeur du paramètre $B_g$ comme fonction du temps Monte Carlo pour trois chaînes de Markov indépendantes (courbe de gauche). Nous voyons que les trois chaînes convergent clairement vers une même distribution stationnaire, ce qui est confirmée en traçant la distribution du paramètre $B_g$ pour chaque chaîne (courbe de droite, en prenant soin d'écarter les points se trouvant dans la région de thermalisation et tenant en compte le temps d'autocorrélation intégré estimé à $\tau_{\text{int}} = 42$ ). . . . .	46
3.1	Structure schématique du code HERAFitter. Figure tirée de <b>xFITTER DEVELOPERS [2017]</b> . . . . .	50
3.2	Les densités de probabilité a posteriori marginales des 10 paramètres libres des PDFs (diagonal) et les courbes de corrélations à deux dimensions entre ces paramètres (off-diagonal). . . . .	55
3.3	La densité de probabilité marginalisée des paramètres $B_g$ et $C_{\overline{D}}$ . Ces paramètres ne suivent pas la loi Gaussienne, comme nous pouvons le voir à partir de l'ajustement Gaussien (courbe rouge). . . . .	56
3.4	La distribution de $\chi^2$ pour une chaîne Monte Carlo à 10 dimensions. La courbe rouge est un ajustement de ces échantillons Monte Carlo avec la loi d'une distribution de $\chi^2$ possédant 10 degrés de liberté. Une description plus détaillée est donnée dans le texte. . . . .	57
3.5	La densité de probabilité de la distribution du gluon $xg(x, Q^2)$ pour $x \approx 10^{-4}$ (courbe de gauche) et pour $x \approx 0.83$ (courbe de droite) à l'échelle $Q^2 = 10 \text{ GeV}^2$ . . . . .	58
3.6	La densité de probabilité de la distribution du quark de la mer $xS(x, Q^2)$ pour $x \approx 10^{-4}$ (courbe de gauche) et pour $x \approx 0.83$ (courbe de droite) à l'échelle $Q^2 = 10 \text{ GeV}^2$ . . . . .	58
3.7	Les distributions de partons $xu_{\text{val}}$ et $xd_{\text{val}}$ à l'échelle $Q^2 = 10 \text{ GeV}^2$ obtenues à partir de la chaîne Monte Carlo en utilisant le code d'évolution QCDNUM <b>BOTJE [2011]</b> (courbes de gauche) sont comparées à celles de HERAPDF1.0 (ZM-VFNS) obtenues à partir du code public HERAFitter (courbes de droite). Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur centrale (courbe rouge) pour les PDFs MCMC et la déviation $\Delta\chi^2 = 1$ standard pour les PDFs HERAPDF1.0. . . . .	59
3.8	Les distributions de partons $xg$ et $xS = x\overline{U} + x\overline{D}$ à l'échelle $Q^2 = 10 \text{ GeV}^2$ obtenues à partir de la chaîne Monte Carlo en utilisant le code d'évolution QCDNUM <b>BOTJE [2011]</b> (courbes de gauche) sont comparées à celles de HERAPDF1.0 (ZM-VFNS) obtenues à partir du code public HERAFitter (courbes de droite). Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur centrale (courbe rouge) pour les PDFs MCMC et la déviation $\Delta\chi^2 = 1$ standard pour les PDFs HERAPDF1.0. . . . .	60
3.9	Rapport des valeurs centrales des PDFs MCMC et HERAPDF1.0 (ZMVFN scheme) pour $xg$ , $xu_{\text{val}}$ , $xd_{\text{val}}$ et $xS$ à l'échelle $Q^2 = 10 \text{ GeV}^2$ . . . . .	61
3.10	Comparaison des incertitudes expérimentales des PDFs normalisées par la valeur centrale, comme déterminées par les méthodes Hessienne et MCMC à l'ordre NLO en QCD pour les distributions des quarks de valence $xu_{\text{val}}$ et $xd_{\text{val}}$ à l'échelle $Q^2 = 10 \text{ GeV}^2$ . . . . .	61
3.11	Comparaison des incertitudes expérimentales des PDFs normalisées par la valeur centrale, comme déterminées par les méthodes Hessienne et MCMC à l'ordre NLO en QCD pour les distributions de partons du gluon $xg$ et du quark de la mer $xS$ à l'échelle $Q^2 = 10 \text{ GeV}^2$ . . . . .	62

3.12 Les distributions des quarks $x_U$ , $x_D$ , $x_{\bar{U}}$ et $x_{\bar{D}}$ obtenues par la méthode MCMC à l'échelle $Q^2 = 10 \text{ GeV}^2$ . Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur du "best fit" (courbe rouge). . . . .	63
3.13 Les distributions des quarks $x_{\bar{u}}$ , $x_{\bar{d}}$ , $x_{\bar{s}}$ et $x_{\bar{c}}$ obtenues par la méthode MCMC à l'échelle $Q^2 = 10 \text{ GeV}^2$ . Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur du "best fit" (courbe rouge). . . . .	64
3.14 Les distributions des quarks de valence ( $x_{u_{\text{val}}}$ , $x_{d_{\text{val}}}$ ), du quark de la mer ( $x_S = x_{\bar{u}} + x_{\bar{d}} + x_{\bar{s}} + x_{\bar{c}}$ , avec $x_{\bar{c}} = 0$ pour $Q^2 < m_c^2$ ) et du gluon ( $x_g$ ) obtenues avec le schéma ZM-VFNS en utilisant la méthode MCMC à l'échelle $Q^2 = 1.9 \text{ GeV}^2$ (courbe de haut) et à l'échelle $Q^2 = 10 \text{ GeV}^2$ (courbe de bas). Les distributions du quark de la mer et du gluon sont réduites d'un facteur de 20. Les incertitudes expérimentales ( les limites de confiance à 68% comme définies à partir des densités de probabilité des distributions de partons $x_{u_{\text{val}}}$ , $x_{d_{\text{val}}}$ , $x_S$ et $x_g$ ) sont représentées par des régions de couleur bleu. . . . .	65
3.15 Les distributions des quarks de valence ( $x_{u_{\text{val}}}$ , $x_{d_{\text{val}}}$ ), du quark de la mer ( $x_S = x_{\bar{u}} + x_{\bar{d}} + x_{\bar{s}} + x_{\bar{c}}$ , avec $x_{\bar{c}} = 0$ pour $Q^2 < m_c^2$ ) et du gluon ( $x_g$ ) obtenues avec le schéma ZM-VFNS en utilisant la méthode MCMC à l'échelle $Q^2 = 1.9 \text{ GeV}^2$ (courbe de haut) et à l'échelle $Q^2 = 10 \text{ GeV}^2$ (courbe de bas). Les incertitudes expérimentales ( les limites de confiance à 68% comme définies à partir des densités de probabilité des distributions de partons $x_{u_{\text{val}}}$ , $x_{d_{\text{val}}}$ , $x_S$ et $x_g$ ) sont représentées par des régions de couleur bleu. . . . .	66



# Liste des tableaux

1.1	Particules élémentaires. . . . .	6
1.2	Bosons de jauge des 3 interactions microscopiques du modèle standard. . .	6
1.3	Les valeurs des autres paramètres fixes des paramétrisations (1.84)-(1.88). .	28
1.4	Résultats de l’ajustement du paramètre $B_g$ aux données de HERA I. La contribution de chaque ensemble de données au $\chi^2_{\min}$ total et le nombre de points correspondant sont montrés. . . . .	28
3.1	Les valeurs des écarts-types obtenues par la minimisation standard de la fonction $\chi^2$ . . . . .	52
3.2	Les valeurs des temps d’autocorrélation intégrés pour les 10 paramètres libres des PDFs (1.84)-(1.88). . . . .	52
3.3	Comparaison des valeurs moyennes et de leurs erreurs statistiques pour les 10 paramètres libres des PDFs obtenues en utilisant deux méthodes d’analyse différentes. . . . .	53
3.4	Comparaison des résultats obtenus pour les valeurs des paramètres libres des PDFs extraites à partir d’une chaîne de Markov indépendante de longueur 39 600, et les résultats fournis par MINUIT. Nous calculons à partir du MCMC la valeur du “best fit”, la valeur moyenne et la valeur de l’écart-type pour chacun des des 10 paramètres libres considérés, ainsi que leurs erreurs statistiques estimées par la méthode Jackknife. L’écart-type donné par MINUIT est l’écart-type à $1\sigma$ . . . . .	54
3.5	Les valeurs du “best fit” et leurs limites de confiance à 68% correspondantes aux distributions de partons $xg(x, Q^2)$ et $xS(x, Q^2)$ à l’échelle $Q^2 = 10 \text{ GeV}^2$ pour les valeurs de $x \approx 10^{-4}$ et $x \approx 0.83$ . . . . .	57





# Introduction

La chromodynamique quantique (“quantum chromodynamics”, QCD) est la théorie de l’interaction forte, dont l’ambition est d’expliquer la cohésion des noyaux ainsi que la structure du neutron et du proton, c’est-à-dire la matière la plus visible de l’univers. Son domaine d’application est encore plus étendu, puisque la QCD contrôle la structure et les interactions de tous les hadrons : protons, neutrons, hyperons, pions, kaons, etc. C’est l’une des théories les plus élégantes de la science (avec la relativité générale) ; il a très peu de paramètres et permet de donner une interprétation physique à une gamme très élevée de phénomènes en utilisant un formalisme bien défini et très compact.

Parmi les ingrédients fondamentaux de la QCD, les distributions de partons (“parton distribution functions”, PDFs) sont des éléments clés et jouent un rôle essentiel pour connecter la dynamique de la QCD (quarks et gluons) aux sections efficaces (des processus de diffusion dure) mesurées. Elles portent une source inestimable d’information sur la structure partonique des hadrons et un énorme effort théorique et expérimental a été consacré depuis des années à l’extraction de ces distributions de partons.

Les PDFs sont d’autant plus importantes aujourd’hui que, avec la collecte de données au grand collisionneur hadronique (Large Hadron Collider, “LHC”), elles sont essentielles pour le calcul d’une grande classe d’observables. Construit pour la découverte du boson de Higgs et l’étude de la physique au-delà du modèle standard, le LHC est en effet essentiellement une usine de la QCD, produisant des événements dans une gamme d’énergie inexplorée. Le potentiel de découverte du LHC dépend crucialement de la qualité des prévisions pour les “signaux et backgrounds” de la QCD et donc sur la qualité des PDFs.

Les PDFs sont intrinsèquement des objets non perturbatifs et ne peuvent donc être déterminées en utilisant uniquement des outils de la QCD perturbative. La méthode la plus efficace pour effectuer des calculs en QCD non perturbative est la QCD sur réseau. Toutefois, bien que la structure du nucléon ait fait l’objet d’une activité intensive en QCD sur réseau depuis des années et, même si des techniques prometteuses pour calculer les PDFs directement sur le réseau ont récemment été proposées Ji [2013], les calculs ab initio des PDFs sont très difficiles et ne constituent toujours pas une alternative concurrentielle aux analyses globales. Ces dernières restent donc la méthode choisie pour obtenir les PDFs, qui sont paramétrisées par des formes fonctionnelles dont les paramètres sont contraints par les données provenant des expériences.

La détermination des PDFs consiste donc à analyser les ensembles de données collectés dans les collisionneurs afin de contraindre les paramètres, des formes fonctionnelles des PDFs données à une échelle d’énergie initiale fixée  $Q_0^2$ . De telles analyses sont souvent basées sur une méthode de moindres carrés, c’est-à-dire sur la minimisation d’une fonction  $\chi^2$ , qui compare les données expérimentales et les prédictions théoriques. Pendant longtemps, les PDFs déterminées de cette façon ne tiennent compte d’aucune estimation des incertitudes, autre que la simple comparaison des résultats fournis par différentes collaborations d’analyses globales. Cependant, avec l’avènement, à l’aube du 21ème siècle, de la nouvelle génération des collisionneurs et de la recherche active de la nouvelle phy-

sique, la nécessité d'accéder aux incertitudes des PDFs est devenue claire [GIELE et KELLER \[1998\]](#); [GIELE et collab. \[2001\]](#); [HUSTON et collab. \[1996\]](#); [LAI et collab. \[1997\]](#); [MARTIN et collab. \[1998, 2003, 2004\]](#). De nombreuses études ont depuis été consacrées aux estimations des incertitudes sur les prédictions physiques en raison des incertitudes des PDFs (voir par exemple [GIELE et collab. \[2001\]](#); [PUMPLIN et collab. \[2001a,b\]](#); [STUMP et collab. \[2001\]](#) et des références à ce sujet) et des progrès considérables ont été réalisés. Néanmoins, cette tâche est loin d'être triviale et de nombreux problèmes demeurent ouverts [PUMPLIN et collab. \[2001a,b\]](#).

Notre compréhension actuelle des incertitudes sur les PDFs est basée principalement<sup>1</sup> sur la méthode Hessienne [PUMPLIN et collab. \[2001a\]](#) ou sur la méthode des multiplicateurs de Lagrange [STUMP et collab. \[2001\]](#). La détermination des incertitudes repose alors sur une hypothèse. Cette hypothèse consiste à explorer la variation de la fonction  $\chi^2$  dans le voisinage de sa valeur minimale ( $\chi^2 = \chi_{\min}^2 + \Delta\chi^2$ ), ce qui conduit au choix arbitraire d'un paramètre de tolérance T. Afin de donner une interprétation statistique objective aux incertitudes, nous avons proposé d'utiliser les techniques Monte Carlo par chaînes de Markov (Markov Chain Monte Carlo, "MCMC") pour définir les incertitudes d'une manière basée autant que possible sur des méthodes statistiques robustes. Les algorithmes MCMC ont été un outil extrêmement populaire en statistiques. Bien que ces techniques soient déjà utilisées dans de nombreux domaines de la physique (voir par exemple [GILKS et collab.; SOKAL \[1989\]](#)) elles n'ont pas encore été utilisées comme une méthode autonome pour déterminer les PDFs et leurs incertitudes expérimentales, c'est-à-dire sans recourir à une procédure de minimisation standard de la fonction  $\chi^2$ <sup>2</sup>.

La méthode MCMC permet d'estimer des densités de probabilité a posteriori pour des modèles multidimensionnels et fournit des estimations fiables des incertitudes. La méthode MCMC consiste donc à simuler séquentiellement une seule chaîne de Markov dont la distribution limite est celle choisie.

Le principal défi de la présente thèse est de démontrer que les techniques Monte Carlo par chaînes de Markov peuvent être appliquées à l'extraction des distributions de partons du proton. Plus la dimension de la chaîne est élevée (c'est-à-dire dans mon cas, plus j'ai de paramètres libres de PDFs à déterminer), plus il me faut un temps de calcul considérable pour générer la chaîne. Le grand nombre de paramètres libres à déterminer dans une détermination complète des PDFs m'a conduit à l'utilisation d'un algorithme Monte Carlo basé sur la dynamique moléculaire, initialement développé pour la théorie des champs sur réseau. J'ai appliqué cet algorithme à une extraction réaliste (mais pas complète) des PDFs; basée sur des formes fonctionnelles ayant 10 paramètres libres et quatre jeux (ensembles) de données expérimentales, pour démontrer que le Monte Carlo par chaînes de Markov peut être appliqué avec succès au calcul des distributions de partons du proton.

Ce manuscrit est composé de trois chapitres. Dans le premier chapitre, la théorie de la chromodynamique quantique a été présentée. Un accent particulier est mis sur le processus de la diffusion inélastique profonde utilisé pour sonder l'intérieur du proton. Dans le chapitre 2, j'ai formulé le problème de la détermination des PDFs en terme de l'inférence Bayésienne. Les principes de base de la méthode Monte Carlo par chaînes de Markov sont rappelés en utilisant l'algorithme de Metropolis-Hastings. L'algorithme Monte Carlo Hamiltonien a été présenté et j'ai montré comment il permet de faire face à un grand nombre de paramètres libres de PDFs à déterminer. Enfin, la procédure d'analyse des chaînes de

---

1. Ceci, à l'exception notable des techniques de réseaux de neurones [BALL et AL. \[2010\]](#).

2. Les estimations des incertitudes utilisant des répliques de "pseudodata" sont également basées sur des méthodes Monte Carlo [GIELE et KELLER \[1998\]](#); [GIELE et collab. \[2001\]](#), mais reposent toujours sur des ajustements.

Markov a été exposée. Le chapitre 3 présente les premiers résultats de l'étude de faisabilité réalisée sur un cas réaliste de détermination des PDFs ayant 10 paramètres libres.

## Références

- BALL, R. D. et AL. 2010, «A first unbiased global nlo determination of parton distributions and their uncertainties», *Nucl. Phys. B*, vol. 838, p. 136–206. [2](#)
- GIELE, W. T. et S. KELLER. 1998, *Phys.Rev. D*, vol. 58, p. 094 023. [2](#)
- GIELE, W. T., S. KELLER et D. KOSOWER. 2001, *FERMILAB-PUB-01-498-T, arXiv :0104052 [hep-ph]*. [2](#)
- GILKS, W. R., S. RICHARDSON et D. J. SPEIGELHALTER. «Markov chain monte carlo in practice», *Chapman and Hall/CRC Interdisciplinary Statistics ed.*, ISBN 9780412055515. [2](#)
- HUSTON, J., E. KOVACS, S. KUHLMANN, H. L. LAI, J. F. OWENS, D. SOPER et W. K. TUNG. 1996, *Phys. Rev. Lett.*, vol. 77, p. 444. [2](#)
- JI, X. 2013, «Parton physics on a euclidean lattice», *Phys.Rev.Lett.110.262002*. [1](#)
- LAI, H. L., J. HUSTON, S. KUHLMANN, F. OLNESS, J. OWENS, D. SOPER, W. K. TUNG et H. WEERTS. 1997, *Phys. Rev. D*, vol. 55, p. 1280. [2](#)
- MARTIN, A. D., R. G. ROBERTS, M. G. RYSKIN et W. J. STIRLING. 1998, *Eur. Phys. J. C*, vol. 2, p. 287. [2](#)
- MARTIN, A. D., R. G. ROBERTS, W. J. STIRLING et R. S. THORNE. 2003, *Eur. Phys. J. C*, vol. 28, p. 455. [2](#)
- MARTIN, A. D., R. G. ROBERTS, W. J. STIRLING et R. S. THORNE. 2004, *Eur. Phys. J. C*, vol. 35, p. 325. [2](#)
- PUMPLIN, J., D. STUMP, R. BROCK, D. CASEY, J. HUSTON, J. KALK, H. L. LAI et W. K. TUNG. 2001a, «Uncertainties of predictions from parton distribution functions II : The hessian method», *Phys. Rev. D*, vol. 65, p. 014 013. [2](#)
- PUMPLIN, J., D. R. STUMP et W. K. TUNG. 2001b, *Phys.Rev. D*, vol. 65, p. 014 011. [2](#)
- SOKAL, A. D. 1989, «Monte carlo methods in statistical machanics : foundations and new algorithms», *Cours de Troisième Cycle de la Physique en Suisse Romande (Lausanne, Switzerland)*. [2](#)
- STUMP, D., J. PUMPLIN, R. BROCK, D. CASEY, J. HUSTON, J. KALK, H. LAI et W. K. TUNG. 2001, «Uncertainties of predictions from parton distribution functions I : The lagrange multiplier method», *Phys.Rev. D*, vol. 65, p. 014 012. [2](#)



# Chapitre 1

## La chromodynamique quantique

### 1.1 Introduction

La chromodynamique quantique [FRITZSCH et collab. \[1973\]](#); [GROSS et WILCZEK \[1973a\]](#); [WEINBERG \[1973\]](#) est la théorie de champ quantique pour l'interaction forte, l'une des quatre forces fondamentales dans la nature. La suite de ce chapitre est organisé de la manière suivante : les principes de base de cette élégante théorie, qui a valu le prix Nobel à ses auteurs en 2004, sont discutés dans la section [1.2](#), la section [1.3](#) porte sur la chromodynamique quantique perturbative, un bref aperçu des différents schémas de saveurs de quarks lourds est présenté dans la section [1.4](#), et enfin le formalisme de l'analyse globale des PDFs en QCD ainsi que les différentes approches d'estimation des incertitudes expérimentales des PDFs sont respectivement exposés dans les sections [1.5](#) et [1.6](#).

### 1.2 Les bases de la chromodynamique quantique

#### 1.2.1 La chromodynamique quantique en bref

Toute théorie des champs quantiques relativistes peut être définie en fournissant des informations sur (a) le contenu en champ incluant les nombres quantiques (spin, charge électrique, charge de couleur, ...), (b) un Lagrangien décrivant les interactions locales entre les différentes particules et (c) les valeurs des paramètres libres (masses, couplage, ...) du Lagrangien. Dans le cas de la QCD, la liberté essentielle est de spécifier l'ensemble des champs de la matière (les quarks et antiquarks : voir le tableau [1.1](#)). Les champs de jauge (les gluons) et la structure du Lagrangien découlent principalement des principes de symétrie de l'invariance de Poincaré, de l'invariance de jauge et de la renormalisabilité<sup>1</sup>. Les valeurs des paramètres libres nécessitent quant à eux des informations expérimentales. La QCD est une partie intégrante du modèle standard ("standard model", SM) qui est une théorie englobant toutes les particules connues ainsi que les trois interactions ayant un effet à l'échelle minuscule des particules : l'interaction électromagnétique, l'interaction forte et l'interaction faible. C'est un modèle qui permet d'expliquer tous les phénomènes naturels sauf la gravitation dont la quantification, résiste pour l'instant aux théoriciens. Les tableaux [1.1](#) et [1.2](#) montrent les termes de couplage pour une interaction entre particules élémentaires. La constante de couplage faible  $g$  est remplacée par  $\frac{e}{\sin\theta_W}$

---

1. Ces trois principes n'interdisent pas la présence dans le lagrangien du terme  $g_s^2 \frac{\theta}{64\pi^2} \epsilon^{\mu\nu\rho\sigma} G_{\mu\nu}^a G_{\rho\sigma}^a$  de violation de la symétrie CP qui est phénoménologiquement susceptible de contribuer par exemple au moment dipolaire électrique du neutron, sur lequel les recherches expérimentales ont mis une limite supérieure de l'ordre de  $\mu_N \leq \mathcal{O}(10^{-26})$  e.cm, soit  $\theta \leq \mathcal{O}(10^{-10})$  rad.

où  $e$  et  $\theta_W$  représentent respectivement la charge élémentaire et l'angle de mélange de Weinberg.  $v$  et  $a$  représentent les couplages axial et vectoriel des particules considérées. On a :

$$v = I_3 - 2 \frac{Q}{e} \sin^2 \theta_W \quad (1.1)$$

$$a = I_3 \quad (1.2)$$

où  $I_3$  est l'isospin de la particule et  $Q$  sa charge électrique

Particules élémentaires		Charge électrique $Q (e)$	Isospin $I_3$
<b>quarks</b>	$u, c, t$	$+\frac{2}{3}$	$+\frac{1}{2}$
	$d, s, b$	$-\frac{1}{3}$	$-\frac{1}{2}$
<b>leptons</b>	$\nu_e, \nu_\mu, \nu_\tau$	0	$+\frac{1}{2}$
	$e^-, \mu^-, \tau^-$	-1	$-\frac{1}{2}$
<b>antiquarks</b>	$\bar{u}, \bar{c}, \bar{t}$	$-\frac{2}{3}$	$-\frac{1}{2}$
	$\bar{d}, \bar{s}, \bar{b}$	$+\frac{1}{3}$	$+\frac{1}{2}$
<b>antileptons</b>	$\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$	0	$-\frac{1}{2}$
	$e^+, \mu^+, \tau^+$	+1	$+\frac{1}{2}$

TABLEAU 1.1 – Particules élémentaires.

Force	Bosons intermédiaires	Termes de Feynman pour le couplage	Particules affectées
<b>électromagnétique</b>	$\gamma$	$-iQ\gamma^\mu$	particules chargées
<b>faible</b>	$Z$ (neutre)	$-i \frac{e}{2 \cos \theta_W \sin \theta_W} \gamma^\mu (v - a\gamma_5)$	toutes les particules
	$W^\pm$ (chargé)	$-i \frac{e}{2\sqrt{2} \sin \theta_W} \gamma^\mu (1 - \gamma_5)$	
<b>forte</b>	gluons (8)	$-ig_s \frac{\lambda_a}{2} \gamma^\mu \quad (a = 1, \dots, 8)$	quarks et antiquarks

TABLEAU 1.2 – Bosons de jauge des 3 interactions microscopiques du modèle standard.

## 1.2.2 Les champs de quarks et de gluons

Les degrés de liberté fondamentaux de la QCD sont les quarks et les gluons représentés par les champs quantiques dans l'espace-temps à quatre dimensions :

$$q_{if\rho}(x), \quad G_\sigma^a(x). \quad (1.3)$$

Sur le champ des quarks  $q$ , " $f$ " représente l'indice de saveur. La masse du quark  $q_f$  est désignée par  $m_f$ . De plus, chaque champ de quark porte un indice de Dirac " $\rho$ ", décrivant un objet de spin 1/2, et un indice de couleur " $i = rouge, vert, bleu$ " se transformant sous la représentation fondamentale **3** du groupe de jauge de couleur  $SU(3)_c$ . Sur le champ de gluons  $G$ , " $\sigma = 0, 1, 2, 3$ " est un indice de Lorentz décrivant un champ de spin 1 et " $a = 1, \dots, 8$ " est un indice de couleur se transformant sous la représentation adjointe **8** de  $SU(3)_c$ . Les gluons ne sont pas massifs en conséquence de l'invariance de jauge  $SU(3)_c$  du Lagrangien.

Certains commentaires sont à faire : (a) sur la base de faits expérimentaux, la sélection de  $SU(3)$  comme groupe de jauge de couleur est unique. Voir par exemple l'introduction de [ALTARELLI \[1982\]](#). (b) De nombreuses applications de la QCD (perturbative) mettent en

jeu l'interaction des quarks avec les photons, les bosons W et Z et toutes sortes de processus de production de ces bosons de jauge, leptons et le boson de Higgs. Pour une description complète, il serait nécessaire de discuter la QCD dans le contexte du SM. Cela dépasse le cadre de cette introduction. Ici, il suffit de garder à l'esprit que les (anti)quarks interagissent avec les bosons de jauge électrofaibles (voir le tableau 1.2) et qu'ils possèdent des charges électriques qui sont données en unité de la charge du positron ( $e > 0$ ). (c) Les antiquarks  $\bar{q}_{f\rho}^i(x)$  ont des nombres quantiques de jauge conjugués, c'est-à-dire qu'ils se transforment sous la représentation fondamentale complexe conjuguée  $\mathbf{3}^* \equiv \bar{\mathbf{3}}$  de  $SU(3)_c$  et ont des charges électriques opposées à celles des quarks.

### 1.2.3 Le Lagrangien

Le Lagrangien de QCD est donné par

$$L_{\text{QCD}} = L_{\text{YM}} \quad (1.4)$$

L'expression pour le Lagrangien de Yang-Mills (ou Lagrangien classique) peut être décomposée en une partie fermionique ( $L_F$ ) et une partie de jauge ( $L_G$ ),

$$L_{\text{YM}} = L_F + L_G. \quad (1.5)$$

La partie fermionique s'exprime de la façon suivante :

$$L_F = \sum_{k,l=1}^3 \sum_{f=u,d,s,c,b,t} \bar{q}_f^k (i\mathcal{D} - m_f \mathbf{I})_{kl} q_f^l, \quad (1.6)$$

où  $\mathcal{D} := \gamma^\sigma D_\sigma$  est la contraction des matrices gamma de Dirac avec la dérivée covariante  $D_\sigma$  donnée par

$$D_\sigma := \partial_\sigma \mathbf{I} + i g_s G_\sigma \quad \text{où} \quad G_\sigma := G_\sigma^a T_a. \quad (1.7)$$

Ici  $g_s$  est la constante de couplage fort et  $G_\sigma$  est la contraction des champs de gluons avec les  $3^2 - 1$  générateurs  $T_a$  du groupe  $SU(3)_c$  dans la représentation fondamentale à trois dimensions, c'est-à-dire,  $\mathbf{I}$  et  $G_\sigma$  sont des matrices  $3 \times 3$  dans l'espace de couleur.

La partie de jauge est donnée par<sup>2</sup>

$$L_G = -\frac{1}{2} \text{Tr}[G_{\sigma\lambda} G^{\sigma\lambda}] = -\frac{1}{2} \sum_{a=1}^8 \sum_{b=1}^8 G_{\sigma\lambda}^a G_b^{\sigma\lambda} \underbrace{\text{Tr}[T_a T_b]}_{T_F \delta_a^b} = -\frac{1}{4} \sum_{a=1}^8 G_{\sigma\lambda}^a G_a^{\sigma\lambda}, \quad (1.8)$$

ici le tenseur de force  $G_{\sigma\lambda}^a$  est défini comme

$$G_{\sigma\lambda}^a = \partial_\sigma G_\lambda^a - \partial_\lambda G_\sigma^a - g_s f_{abc} G_\sigma^b G_\lambda^c, \quad (1.9)$$

avec  $f_{abc}$  les constantes de structure de l'algèbre de Lie  $\mathfrak{su}(3)_c$  et  $G_{\sigma\lambda}$  la matrice correspondante dans l'espace de couleur :

$$G_{\sigma\lambda} := G_{\sigma\lambda}^a T_a = D_\sigma G_\lambda - D_\lambda G_\sigma = -\frac{i}{g_s} [D_\sigma, D_\lambda]. \quad (1.10)$$

Le Lagrangien de Yang-Mills  $L_{\text{YM}}$  est invariant sous la transformation de jauge locale  $SU(3)_c$ , renormalisable, et invariant sous la transformation CP.

2. Où  $T_F = \frac{1}{2}$  est la normalisation des générateurs de l'algèbre  $\mathfrak{su}(3)_c$ .



## 1.2.4 Les paramètres libres

Le Lagrangien de QCD dépend de sept paramètres : le couplage de jauge  $g_s$  et les masses des six quarks  $m_u, m_d, m_s, m_c, m_b, m_t$ . Pour faire des prédictions, il est nécessaire de fournir des valeurs pour le couplage de jauge et les masses des quarks comme “input”. Actuellement nous n’avons aucune théorie (au delà de QCD) pour calculer à partir des principes fondamentaux seulement, la valeur de ces paramètres si bien qu’ils doivent être extraits expérimentalement en utilisant un nombre ( $\geq 7$ ) d’observables, pour faire des prédictions pour toutes les autres observables.

### La constante de couplage fort

Le paramètre le plus important est le couplage de jauge  $g_s$  qui spécifie la force de l’interaction entre les quarks et les gluons (et l’auto-interaction des gluons). Le comportement de la théorie dépend essentiellement de sa valeur. Si  $g_s$  est suffisamment petite, une approche perturbative est possible où les observables peuvent être systématiquement développées suivant des puissances de  $g_s$ . En revanche, pour les grandes valeurs de  $g_s$ , les traitements perturbatifs ne sont plus possibles : les quarks et les gluons s’hadronisent pour former les baryons et les mésons que l’on observe expérimentalement. La QCD sur réseau permet le calcul de fonctions de corrélation de manière non perturbative. Plus spécifiquement, il s’agit d’une régularisation de la théorie, adaptée au calcul numérique, dans laquelle l’espace-temps est discrétisé. La taille “L” du réseau joue le rôle de coupure infrarouge alors que le pas “a” du réseau joue le rôle de coupure ultraviolette rendant l’intégrale de chemin parfaitement définie. Dans la limite où le volume tend vers l’infini et le pas du réseau tend vers zéro, on retrouve la QCD. Notons que dans ce régime, les états liés des quarks et gluons (hadrons) émergent comme nouveaux degrés de liberté pertinents. En effet, expérimentalement, un quark libre ou un gluon n’a pas été observé et c’est une propriété fondamentale de la QCD que la matière hadronique est composée de quarks (et de gluons) et n’a pas de couleur<sup>3</sup>.

Il s’avère qu’il est difficile et peu significatif de décider une valeur unique pour la constante de couplage fort  $g_s$ . Tout d’abord, la force de l’interaction n’est pas une constante (en dépit de son nom) mais dépend de l’échelle de distance entre les quarks et gluons ou, de manière équivalente, de l’échelle d’énergie du processus. Cependant, une valeur fixe pourrait être une approximation raisonnable à condition que le couplage ne change pas rapidement avec l’échelle<sup>4</sup>. Toutefois, dans le cas de la QCD, la constante de couplage change radicalement, de valeurs non perturbatives aux valeurs perturbatives dans une région de transition caractérisée par une échelle d’énergie  $\Lambda_{\text{QCD}} \approx 200 \text{ MeV}$ . Par conséquent, il est inévitable de prendre en compte la “variation” de la constante de couplage fort (une illustration est donnée sur la figure 1.1). Ceci est toujours possible dans le régime perturbatif (échelles d’énergie beaucoup plus grandes que  $\Lambda_{\text{QCD}}$ ) et nécessite le calcul des effets de boucles. Le premier calcul de ce type [GROSS et WILCZEK \[1973a,b\]](#); [POLITZER \[1973, 1974\]](#) a conduit à la découverte de la liberté asymptotique (et des prix Nobel pour D. Gross, F. Wilczek et D. Politzer en 2004), ce qui signifie que la constante de couplage fort diminue à des énergies plus élevées/distances plus basses et approche la valeur zéro à des énergies asymptotiquement grandes où la QCD se transforme en une théorie libre.

---

3. Plus mathématiquement, cela signifie que les hadrons se transforment comme des scalaires réels sous les transformations de  $SU(3)_c$ .

4. C’est le cas en QED où  $\alpha_{em} = e^2/(4\pi) \approx 1/137$  à des énergies  $E \ll \Lambda_{\text{EW}} \approx 100 \text{ GeV}$  change à  $\alpha_{em} \approx 1/128$  à l’échelle électrofaible  $\Lambda_{\text{EW}}$  si bien que la valeur  $1/137$  est encore bonne avec une précision de  $10^{-7}\%$ .

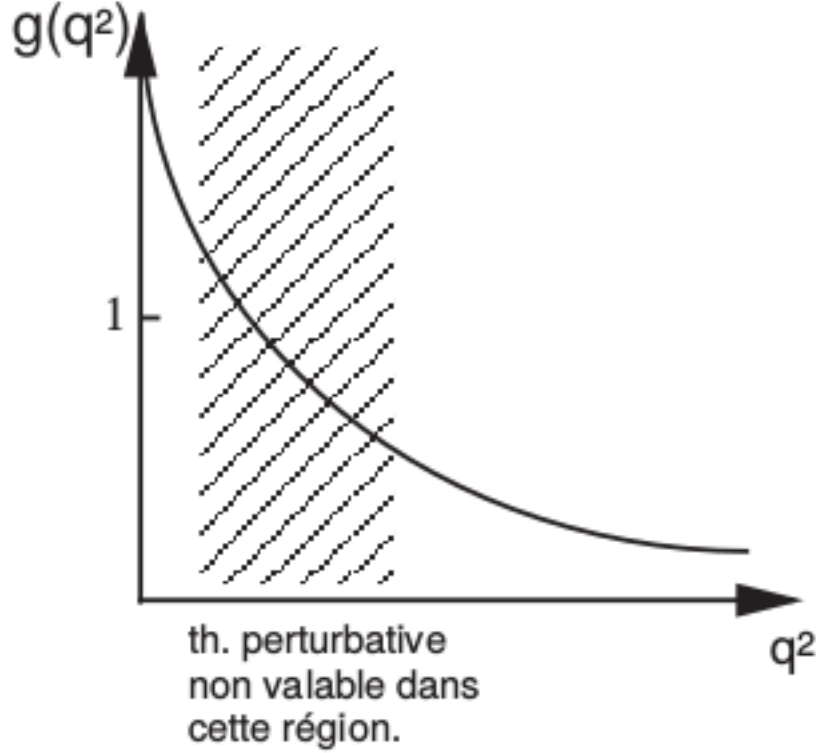


FIGURE 1.1 – Courbe montrant le comportement du couplage fort  $g_s$  en fonction de l'énergie  $q^2$ . Figure tirée de [AURENCHÉ et collab. \[2016\]](#)

Lorsque l'on tente d'effectuer des calculs à boucles, on tombe sur des divergences. Il est donc nécessaire d'introduire un schéma de régularisation afin de définir mathématiquement la théorie. Dans le cadre perturbatif, la méthode de régularisation dimensionnelle [ASHMORE \[1972\]](#); [BOLLINI et GIAMBIAGI \[1972\]](#); [CICUTA et MONTALDI \[1972\]](#); ['T HOOFT et VELTMAN \[1972\]](#) est particulièrement bien adaptée (voir, par exemple, la référence [COLLINS \[1984\]](#) pour une discussion de différents schémas de régularisation). Les divergences apparaissent alors comme des pôles de type  $2/(4 - D)$  où  $D$  est le nombre de dimensions de l'espace-temps dans lequel le calcul est effectué. Pour une théorie renormalisable comme la QCD, toutes les divergences ultraviolettes (UV) peuvent être supprimées par une redéfinition des paramètres et des champs du Lagrangien. Cette redéfinition est dépendante du schéma et la possibilité la plus simple est le schéma de soustraction minimale ("minimal subtraction", MS) où seul le pôle est absorbé dans les champs et les paramètres. Une variante de ceci est le schéma de soustraction minimal modifié (MS) fréquemment utilisé [BARDEEN et collab. \[1978\]](#); ['T HOOFT \[1973\]](#) où le pôle est absorbé avec la constante d'Euler  $\gamma_E \approx 0.57721$  dans la combinaison  $2/(4 - D) + \gamma_E - \ln 4\pi$ . Techniquement, en raison de la procédure de renormalisation, les paramètres du Lagrangien acquièrent une dépendance d'énergie  $\mu_R$  qui est régie par des équations du groupe de renormalisation (renormalization group equations, RGEs). Pour la constante de couplage fort, les RGEs sont données par :

$$\mu_R^2 \frac{d}{d\mu_R^2} a_s(\mu_R) = \beta(a_s(\mu_R)), \quad a_s := \frac{\alpha_s}{4\pi} := \frac{g_s^2}{16\pi^2}. \quad (1.11)$$

La liberté asymptotique correspond au fait que la fonction  $\beta$  de la QCD est négative (à des couplages faibles) et a un point UV fixe à un couplage nul<sup>5</sup>. La fonction  $\beta$  de la QCD a la

5. Ceci est tout à fait différent comparé à la fonction  $\beta$  de QED qui est positive de sorte que  $\alpha_{em}$  aug-

série perturbative suivante :

$$\beta(a_s) = -\beta_0 a_s^2 - \beta_1 a_s^3 - \beta_2 a_s^4 - \beta_3 a_s^5 - \beta_4 a_s^6 - \dots \quad (1.12)$$

et est connue jusqu'à une précision de 5 boucles dans le schéma  $\overline{\text{MS}}$  VERMASEREN et collab. [2017]. Nous rappelons ici les fonctions  $\beta$  à une boucle ( $\beta_0$ ) et deux boucles ( $\beta_1$ ) dans le schéma  $\overline{\text{MS}}$ <sup>6</sup> :

$$\beta_0^{\overline{\text{MS}}} = \frac{11}{3} C_A - \frac{4}{3} T_F n_f, \quad (1.13)$$

$$\beta_1^{\overline{\text{MS}}} = \frac{34}{3} C_A^2 - \frac{20}{3} C_A T_F n_f - 4 C_F T_F n_f. \quad (1.14)$$

Ici,  $C_A = 3$  et  $C_F = \frac{4}{3}$  sont respectivement les constantes de Casimir dans les représentations adjointe et fondamentale. Finalement,  $n_f$  représente le nombre de saveur de quark active dépendant de l'échelle de renormalisation  $\mu_R$ , c'est-à-dire

$$n_f = \begin{cases} 3; & \mu_R \leq m_c, \\ 4; & m_c < \mu_R \leq m_b, \\ 5; & m_b < \mu_R \leq m_t, \\ 6; & m_t < \mu_R, \end{cases}$$

où  $m_c, m_b, m_t$  sont respectivement les masses des quarks charm, bottom et top.

### Les masses des quarks

En ce qui concerne les paramètres de la masse des quarks, il serait tentant de les identifier avec les masses physiques (les masses de pôles) des quarks se propageant de façon libre. Toutefois, comme déjà mentionné, les quarks libres n'ont pas été observés ; au contraire, ils sont confinés dans des hadrons. Il est donc nécessaire de dire précisément ce que nous entendons par une masse de quarks. Pour les quarks légers  $u, d$  et  $s$ , nous comprenons toujours par la masse de quarks  $m_q$  le paramètre de la masse apparaissant dans le Lagrangien de QCD, qui n'est pas une observable mais qui est définie dans un schéma de renormalisation donné. À nouveau, la variation de la masse avec l'énergie est décrite par une équation du groupe de renormalisation :

$$\mu_R^2 \frac{d}{d\mu_R^2} m_q(\mu_R) = \gamma_m(a_s(\mu_R)) = -(\gamma_m)_0 a_s - (\gamma_m)_1 a_s^2 - (\gamma_m)_2 a_s^3 - (\gamma_m)_3 a_s^4 - (\gamma_m)_4 a_s^5 - \dots, \quad (1.15)$$

où  $\gamma_m(a_s)$  est la dimension anomale de la masse de quark BAIKOV et collab. [2014]. Nous n'utilisons pas cette formule dans la suite puisque dans les calculs existants basés sur des théorèmes de factorisation de QCD, l'erreur due aux termes de quarks légers manquants est plus petite que l'erreur de la formule de factorisation.

La situation est différente pour les quarks  $c, b$  et  $t$ . Ils sont appelés quarks lourds parce que leurs masses sont beaucoup plus grandes que  $\Lambda_{\text{QCD}}$  de sorte que  $\alpha_s$  évaluée à l'échelle de la masse de quark lourd se trouve dans la région perturbative. Par conséquent, en raison de la liberté asymptotique, il est possible de considérer les quarks lourds comme des

---

mentement avec l'énergie et en fin de compte divergerait à des énergies extrêmement grandes (c'est-à-dire développe un pôle dit de Landau) à moins qu'il y ait un point UV fixe à des couplages élevés, C'est-à-dire que la fonction  $\beta$  devient nulle à ces couplages élevés.

6.  $\beta_n, n = 0, 1, 2, 3, 4$  sont les mêmes dans tous les schémas de renormalisation de genre MS

particules quasi libres et d'interpréter leurs masses comme des masses physiques (masses de pôles) renormalisées dans le schéma sur couche de masse (on-shell, "OS"). Cependant, l'utilisation de la dépendance en masses des quarks lourds, renormalisées dans le schéma  $\overline{\text{MS}}$  présente certains avantages tels que décrits, par exemple, dans la référence [ALEKHIN et MOCH \[2011\]](#) dans le contexte d'une diffusion inélastique profonde. Dans le reste de ce manuscrit, les masses des quarks lourds sont considérées comme des masses physiques de valeurs  $m_c = 1.4$  et  $m_b = 4.75$ .

## 1.3 La chromodynamique quantique perturbative

### 1.3.1 Les théorèmes de factorisation

Grâce à la liberté asymptotique, les observables qui ne dépendent que de grandes échelles d'énergies peuvent être calculés en théorie des perturbations. Cependant, de nombreuses observables comme des sections efficaces dans les collisionneurs électron-hadron ou hadron-hadron à haute énergie sont régies par des phénomènes de longue distance/basse énergie liés à la structure des hadrons dans l'état initial qui, jusqu'à présent, ne peuvent être calculés à partir des premiers principes de la QCD. Dans de nombreux cas, la section efficace est également influencée par la dynamique qui se produit à très courtes distances/hautes énergies; par exemple quand une particule lourde est produite ou quand une particule dans l'état final a une grande impulsion transverse. Les théorèmes de factorisation de la QCD [COLLINS \[1998\]](#); [COLLINS et SOPER \[1987\]](#); [COLLINS et collab. \[1989\]](#) stipulent que, dans de telles situations, il est possible, jusqu'à une erreur de puissance négligée, de séparer la physique se produisant à des échelles de longueurs très différentes en facteurs indépendants tels que les facteurs de longue distance seront toujours les mêmes, sans tenir compte des détails du processus de courte distance. Inversement, le facteur de courte distance est libre de toute physique de longue distance et peut être calculé perturbativement. En d'autres termes, toutes les fois qu'une échelle dure  $\mu_F$  (ou échelle de courte distance) est présente (disons  $\mu_F^2 > 1 \text{ GeV}^2$ ), la section hadronique peut être écrite sous forme de convolutions de distributions de partons  $f_i(x, \mu_F^2)$  *universelles*<sup>7</sup> avec la section efficace de diffusion dure  $d\hat{\sigma}$  qui dépend du processus mis en jeu. Par exemple, la section efficace différentielle pour la diffusion inélastique profonde de l'électron sur le proton est donnée par :

$$\begin{aligned} d\sigma_{e+p \rightarrow e+X}(\mu_R^2, \mu_F^2) &\simeq \sum_i f_{i/P}(\xi, \mu_F^2) \otimes d\hat{\sigma}_{e+i \rightarrow e+X}(\xi, \mu_R^2, \mu_F^2), \\ &=: \sum_i \int_0^1 d\xi f_{i/P}(\xi, \mu_F^2) d\hat{\sigma}_{e+i \rightarrow e+X}(\xi, \mu_R^2, \mu_F^2), \end{aligned} \quad (1.16)$$

où une somme sur tous les sous-processus partoniques possibles est sous entendue et les détails de la formule et des sous-processus qui sont inclus dépendent du traitement des quarks lourds comme on le discutera plus en détail ci-dessous. Des formules similaires sont valables pour la production de paires de lepton dans les collisions proton-proton,

$$d\sigma_{pp \rightarrow l^+ l^- X}(\mu_R^2, \mu_F^2) \simeq \sum_{i,j} f_{i/P}(\xi_1, \mu_F^2) \otimes f_{j/P}(\xi_2, \mu_F^2) \otimes d\hat{\sigma}_{ij \rightarrow l^+ l^- X}(\xi_1, \xi_2, \mu_R^2, \mu_F^2), \quad (1.17)$$

et d'autres processus comme la production de quarks lourds, de vecteurs bosons, de jets ou des hadrons à grandes impulsions transverses. Dans le dernier cas, les fonctions de

7. C'est-à-dire que les PDFs sont mesurées dans les expériences de diffusion inélastique profonde et utilisées dans les collisionneurs hadroniques.

fragmentations universelles (“fragmentation functions”, FFs) qui sont analogues aux PDFs et décrivant la transition du quark ou gluon à l’état final au hadron observé doivent être incluses dans le formalisme. Dans chaque cas, l’erreur de l’approximation de la factorisation est proportionnelle à une puissance de  $\left(\frac{\Lambda_{\text{QCD}}}{\mu_{\text{F}}}\right)^2$  où  $\Lambda_{\text{QCD}}$  est une échelle hadronique typique de quelques centaines de MeV et elle devient négligeable lorsque des échelles  $\mu_{\text{F}}$  suffisamment dures sont impliquées. Les PDFs (FFs) sont intrinsèquement des objets non perturbatifs comportant la physique de longue distance du hadron de l’état initial (l’état final). Elles sont *universelles* en ce sens que les mêmes PDFs (FFs) peuvent être utilisées dans une grande variété de processus. Puisque les sections efficaces de diffusion dure sont systématiquement calculables ordre par ordre en théorie des perturbations, les théorèmes de factorisation de la QCD fournissent un cadre rigoureux avec un pouvoir prédictif. Ce formalisme de la QCD perturbative est la base théorique pour le calcul systématique de grandes classes d’observables aux collisionneurs de particules actuels et par conséquent d’une importance primordiale.

Les théorèmes de factorisation fournissent des définitions théoriques précises des PDFs comme éléments de matrice hadronique de certains opérateurs de «twist-2» composés de champs de quarks et de gluons tels qu’ils peuvent en principe être calculés à partir des premiers principes sur le réseau. Cependant, bien que des progrès considérables aient été faites ces dernières années [JI \[2013\]](#) les calculs sur réseau ne sont pas encore compétitifs et les PDFs doivent être déterminées pour l’instant à partir de données expérimentales. Les corrections QCD conduisent à la renormalisation de ces opérateurs de telle sorte que les PDFs acquièrent une dépendance d’échelle de renormalisation qui est régie par des équations du groupe de renormalisation, les équations d’évolution de Doshkitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [ALTARELLI et PARISI \[1977\]](#); [DOKSHITZER \[1977\]](#); [GRIBOV et LIPATOV \[1972\]](#). Comme pour les sections efficaces partoniques à courte distance, les équations d’évolution de DGLAP ont une série perturbative dans la constante de couplage fort. Il y a quelques années, le calcul des contributions à 3 boucles de cette série perturbative (les fonctions de «splitting» à 3 boucles) a été complété après plusieurs années d’effort [MOCH et collab. \[2004\]](#); [VOGT et collab. \[2004\]](#). Avec cette connaissance, il est possible de calculer des observables à la précision «next-to-next-to-leading-order» permettant des tests de précision en QCD perturbative à condition que les sections efficaces partoniques à courte distance soient connues à ce même ordre, ce qui n’est pas encore le cas pour tous les processus durs à savoir le processus de hadro-production inclusif de «jet» et le processus de lepto-production de quark lourd. Nous allons dans la suite nous attarder sur les processus inclusifs de diffusion inélastique profonde.

## 1.3.2 Les diffusions inélastiques profondes et la structure du proton

La connaissance des sections efficaces des processus impliquant des hadrons dans l’état initial, nécessite une bonne connaissance des distributions de partons. Celle-ci s’obtient essentiellement par l’étude de la diffusion inélastique profonde.

### 1.3.2.1 Section efficace d’interaction lepton - hadron et PDFs

Toute interaction entre particules peut être mathématiquement définie suivant les conditions cinématiques et le type d’interaction en jeu. Nous allons immédiatement discuter le cas des interactions lepton - proton.

## A. Traitement de la dynamique de l'interaction

De manière générale, la section efficace de toute interaction peut s'écrire sous la forme suivante :

$$d\sigma = \frac{|\mathcal{M}|^2}{F} d\Phi \quad (1.18)$$

où :

- $F$  est le facteur de flux initial. Dans une interaction à deux corps  $a$  et  $b$ , de quadri-vecteurs impulsions initiaux  $p_a$  et  $p_b$ ,  $F = 4\sqrt{(p_a \cdot p_b)^2 - (p_a)^2(p_b)^2}$ . C'est un invariant de Lorentz.
- $\mathcal{M}$  est l'amplitude de l'interaction calculée à l'aide des Lagrangiens associés aux forces mises en jeu.
- $d\Phi = D^{l'X}$  est l'élément différentiel de l'espace de phase, également invariant de Lorentz.

## B. Application au cas $e^\pm + P \rightarrow l' + X$ : introduction aux fonctions de structure

### 1. Cinématique des interactions $e^\pm + P \rightarrow l' + X$

Considérée de manière inclusive, sans se soucier des détails du système  $X$ , l'interaction  $e^\pm + P \rightarrow l' + X$  (où  $l'$  désigne soit un électron (positron), soit un neutrino (anti-neutrino)) peut être traitée comme une interaction à deux corps. Deux variables indépendantes suffisent à la décrire. Les variables cinématiques d'emploi usuel sont

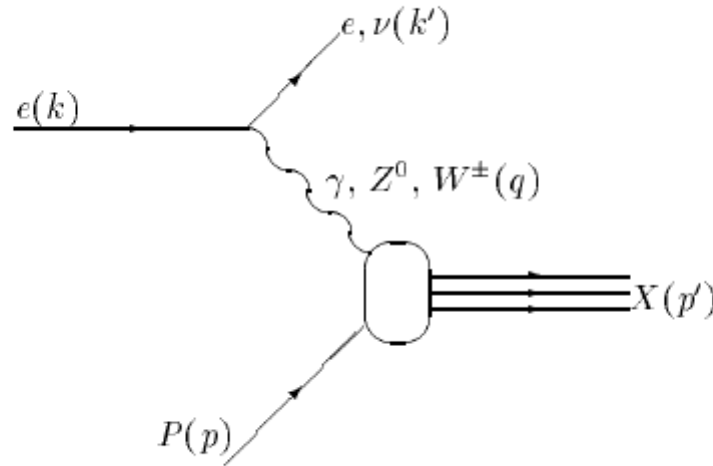


FIGURE 1.2 – Schéma typique d'une interaction  $eP$ .

les suivantes :

- $Q^2 = -q^2 = -(k - k')^2 > 0$ , représente l'opposé de la masse au carré de la particule virtuelle échangée,
- $\nu = \frac{p \cdot q}{M} = E - E'$ , représente l'énergie perdue par l'électron ou le positron dans le référentiel du proton au repos,
- $0 \leq x = \frac{Q^2}{2p \cdot q} = \frac{Q^2}{2M\nu} \leq 1$ , est la variable d'échelle de Björken non dimensionnée,

- $0 \leq y = \frac{p \cdot q}{p \cdot k} = \frac{E - E'}{E} \leq 1$ , est le paramètre d'inélasticité dans le référentiel du proton au repos,
- $s = (p + k)^2$ , est l'énergie au carré du système  $e^\pm P$  dans le référentiel de centre de masse,
- $W^2 = (p')^2 = (p + q)^2$  est la masse au carré du système hadronique X de l'état final.

La diffusion est dite inélastique profonde if  $Q^2 \gg M^2 \approx 1 \text{ GeV}^2$  ("profonde") et  $W^2 = M^2 + Q^2(-1 + 1/x) \gg M^2$  ("inélastique"). Les masses  $m_e$ ,  $m_{l'}$  et  $M$  sont immédiatement négligées devant l'énergie des processus mis en jeu. Cela revient à dire :  $m_e, m_{l'}, M \ll E, E', E - E'$ .

Dans ce cas, on a :

$$s = 2p \cdot k, \quad Q^2 = sxy, \quad W^2 = Q^2(-1 + 1/x), \quad (1.19)$$

$$F = 2s, \quad (1.20)$$

et l'élément de matrice  $\mathcal{M}$  peut s'écrire :

$$\mathcal{M} = J_\mu(e l') \mathcal{J}^\mu(PX) P_r, \quad (1.21)$$

où

- $P_r$  représente l'effet du propagateur du boson de jauge  $V$  échangé ( $\gamma, Z, W^\pm$ ),

$$P_r = \frac{1}{q^2 - M_V^2}. \quad (1.22)$$

- $J_\mu(e l')$  est le courant associé au vertex  $e \rightarrow l' V$ ,

$$J_\mu(e l') = -i G_V \bar{l}' \gamma_\mu (v_e - a_e \gamma_5) e, \quad (1.23)$$

$G_V$ ,  $v_e$  et  $a_e$  sont des coefficients dépendant de l'interaction considérée et du lepton initial.

- $\mathcal{J}^\mu(PX)$  est le courant associé au vertex  $VP \rightarrow X$ , qu'on ne connaît pas a priori. Enfin, avec  $k'(E', \vec{k}')$  le quadri - vecteur impulsion du lepton diffusé (voir figure 1.2), l'élément différentiel de l'espace de phase  $d\bar{\Phi}$  ou  $D^{l'X}$  s'écrit sous la forme suivante :

$$D^{l'X} = \frac{d^3 \vec{k}'}{2E'(2\pi)^3} \prod_i \frac{d^3 \vec{k}_i}{2E_i(2\pi)^3} ((2\pi)^4 \delta^4(k + p - k' - p')) = \frac{d^3 \vec{k}'}{2E'(2\pi)^3} D^X \quad (1.24)$$

Nous allons nous intéresser ici qu'aux processus inclusifs d'interaction  $eP$  non polarisée. Nous allons pour cela intégrer sur toutes les variables cinématiques internes du système hadronique X. La moyenne d'hélicité sur les états initiaux vaut  $\frac{1}{4}$  (particule de spin  $\frac{1}{2}$ ). Par contre, il est nécessaire de faire attention à la polarisation du neutrino (antineutrino) de l'état final lors de l'étude des interactions à courant chargé.

Au total, on peut ré-écrire l'équation (1.18) comme :

$$d\sigma = \frac{1}{F} \left[ \frac{1}{4} \sum_{hel, l'f} |\mathcal{M}|^2 D^X \right] \frac{d^3 \vec{k}'}{2E'(2\pi)^3} \quad (1.25)$$

## 2. Le calcul de l'élément de matrice de diffusion

Les courants  $J_\mu(e l')$  et  $\mathcal{J}^\mu(PX)$  sont couplés au même boson  $V$ . Ils sont donc proportionnels à la constante de couplage  $G_V$ . On peut définir deux tenseurs :

$$L_{\mu\nu}(e l') = \sum_{hel, l'f} \frac{1}{G_V^2} J_\mu(e l') (J_\nu(e l'))^* \quad (1.26)$$



$$H^{\mu\nu}(\text{PX}) = \sum_{hel,lf} \frac{1}{G_V^2} \mathcal{J}^\mu(\text{PX})(\mathcal{J}^\nu(\text{PX}))^* \quad (1.27)$$

où l'étoile signifie le complexe conjugué. Le tenseur leptonique s'écrit donc :

$$L_{\mu\nu}(e^\pm l') = 4(k^\alpha k'^\beta)(v_e^2 + a_e^2) \{g_{\alpha\mu} g_{\beta\nu} + g_{\alpha\nu} g_{\beta\mu} - g_{\mu\nu} g_{\alpha\beta} \mp i \mathcal{A} e_{\mu\nu\alpha\beta}\} \quad (1.28)$$

où  $\mathcal{A} = \frac{2a_e v_e}{v_e^2 + a_e^2}$ . Le tenseur hadronique intégré est défini comme suit :

$$W^{\mu\nu} = \int \frac{1}{2\pi M} \frac{1}{2} H^{\mu\nu} D^X \quad (1.29)$$

Il est utile de développer explicitement le tenseur hadronique en termes de possibles tenseurs de Lorentz multipliés par une fonction scalaire  $W_a$  qui paramétrise la structure du proton. Il existe six tenseurs de Lorentz indépendants de rang 2 qui peuvent être construits à partir du tenseur métrique  $g_{\mu\nu}$ , du quadri-vecteur impulsion  $p_\mu$  du proton P, du quadri-vecteur impulsion  $q_\mu$  du boson échangé V et du tenseur de Levi-Civita  $e_{\mu\nu\rho\sigma}$  :

$$\begin{aligned} &g_{\mu\nu}, \quad p_\mu p_\nu, \quad q_\mu q_\nu, \quad p_\mu q_\nu + p_\nu q_\mu, \\ &e_{\mu\nu\rho\sigma} p^\rho q^\sigma, \quad p_\mu q_\nu - p_\nu q_\mu, \end{aligned} \quad (1.30)$$

où les quatre premiers tenseurs sont symétriques et les deux derniers sont antisymétriques. En conséquence, le tenseur hadronique (non polarisé) peut être développé dans la forme la plus générale suivante :

$$\begin{aligned} W^{\mu\nu} = & W_1(-g^{\mu\nu}) + W_2\left(\frac{p^\mu p^\nu}{M^2}\right) + W_3\left(-i\frac{p^\rho q^\sigma}{2M^2} e_{\rho\sigma}^{\mu\nu}\right) \\ & + W_4\left(\frac{q^\mu q^\nu}{M^2}\right) + W_5\left(\frac{p^\mu q^\nu + p^\nu q^\mu}{2M^2}\right) + W_6\left(\frac{p^\mu q^\nu - p^\nu q^\mu}{2M^2}\right) \end{aligned} \quad (1.31)$$

Puisque  $(q^\mu L_{\mu\nu} = q^\nu L_{\mu\nu} = 0)$ , alors la contraction de  $L_{\mu\nu}$  avec  $W^{\mu\nu}$  ne laisse que trois fonctions de structure indépendantes.

$$\begin{aligned} L_{\mu\nu} W^{\mu\nu} = & 4(v_e^2 + a_e^2) \left[ W_1(2k \cdot k') + W_2 \left( \frac{2(p \cdot k)(p \cdot k') - p^2(k \cdot k')}{M^2} \right) \right. \\ & \left. \mp \mathcal{A} W_3 \left( \frac{(p \cdot k)(q \cdot k') - (q \cdot k)(p \cdot k')}{M^2} \right) \right] \end{aligned} \quad (1.32)$$

Dans la notation moderne, les fonctions de structure sont désignées par  $F_a$  plutôt que  $W_a$  avec les correspondances suivantes :

$$\{F_1, F_2, F_3\} = \{MW_1, vW_2, vW_3\}. \quad (1.33)$$

En rassemblant toutes les informations (équations (1.20), (1.25), (1.32) et (1.33)), on obtient, après quelques calculs, la section efficace différentielle<sup>8</sup> de la diffusion inélastique profonde de  $e^\pm$  sur le proton P en fonction de  $x$  et  $y$  :

$$\frac{d^2\sigma^\pm}{dx dy} = 2\pi s \left( \frac{G_V^2}{4\pi} \right)^2 P_r^2 (v_e^2 + a_e^2) [2xy^2 F_1 + 2(1-y)F_2 \mp \mathcal{A}(2y-y^2)x F_3] \quad (1.34)$$

8. Cette section efficace différentielle est valable uniquement pour des processus mettant en jeu soit l'échange des photons purs, soit des bosons Z purs ou l'échange des bosons  $W^\pm$ .



Cette section efficace différentielle peut être réécrite de la manière suivante<sup>9</sup> :

$$\frac{d^2\sigma^\pm}{dx dy} = 2\pi s \left( \frac{G_V^2}{4\pi} \right)^2 P_r^2(v_e^2 + a_e^2) [-y^2 F_L(x, Q^2) + Y_+ F_2(x, Q^2) \mp \mathcal{A} Y_- x F_3(x, Q^2)], \quad (1.35)$$

avec  $Y_\pm = 1 \pm (1 - y)^2$ .

À l'ordre  $\mathcal{O}(\alpha_s^0)$ , la généralisation de l'équation (1.35) qui incorpore tous les processus faisant intervenir les courants neutres est donnée en termes des fonctions de structure généralisées :

$$\begin{aligned} \frac{d^2\sigma_{\text{NC}}^\pm}{dx dQ^2} &= \frac{2\pi\alpha_{em}^2}{xQ^4} [-y^2 \tilde{F}_L + Y_+ \tilde{F}_2 \mp Y_- x \tilde{F}_3], \\ &= \frac{2\pi\alpha_{em}^2 Y_+}{xQ^4} \sigma_{r,\text{NC}}^\pm, \end{aligned} \quad (1.36)$$

où  $\sigma_{r,\text{NC}}^\pm$  désignant la section efficace réduite pour les courants neutres s'écrit :

$$\sigma_{r,\text{NC}}^\pm = -\frac{y^2}{Y_+} \tilde{F}_L + \tilde{F}_2 \mp \frac{Y_-}{Y_+} x \tilde{F}_3 \quad (1.37)$$

Les fonctions de structure  $\tilde{F}_2$ ,  $\tilde{F}_L$  et  $x\tilde{F}_3$  dépendent des paramètres électrofaibles :

$$\begin{aligned} \tilde{F}_2 &= F_2^Y - v_e \chi_Z F_2^{YZ} + \chi_Z^2 (v_e^2 + a_e^2) F_2^Z, \\ \tilde{F}_L &= F_L^Y - v_e \chi_Z F_L^{YZ} + \chi_Z^2 (v_e^2 + a_e^2) F_L^Z, \\ x\tilde{F}_3 &= a_e \chi_Z x F_3^{YZ} - 2v_e a_e \chi_Z^2 x F_3^Z, \end{aligned} \quad (1.38)$$

où  $\chi_Z(Q^2) = \frac{Q^2}{4(Q^2 + M_Z^2) \sin^2 \theta_W \cos^2 \theta_W}$  et  $M_Z$  la masse du boson Z.

À basse énergie c'est-à-dire  $Q^2 \ll M_Z^2$ , la fonction de structure  $x\tilde{F}_3$  qui viole la parité peut être négligée, puisqu'elle contient la contribution de l'échange du boson Z. La section efficace réduite se réduit à :

$$\sigma_{r,\text{NC}}^\pm = \tilde{F}_2 - \frac{y^2}{Y_+} \tilde{F}_L \quad (1.39)$$

La contribution du terme contenant la fonction de structure  $\tilde{F}_L$  est seulement significative pour des valeurs de  $y$  approximativement plus grandes que 0.5.

Pour démontrer la sensibilité des données expérimentales, il est important de discuter les équations simplifiées dans le modèle de parton de quark ("quark parton model", QPM), où les gluons sont absents et  $\tilde{F}_L = 0$  CALLAN et GROSS [1969]. Dans le QPM, les fonctions de structure dans les équations (1.38) deviennent<sup>10</sup>

$$\begin{aligned} [F_2^Y, F_2^{YZ}, F_2^Z] &\approx [\{e_u^2, 2e_u v_u, v_u^2 + a_u^2\} (x f_U + x f_{\bar{U}}) + \{e_d^2, 2e_d v_d, v_d^2 + a_d^2\} (x f_D + x f_{\bar{D}})], \\ [xF_3^Y, xF_3^{YZ}, xF_3^Z] &\approx [\{0, 2e_u a_u, 2v_u a_u\} (x f_U - x f_{\bar{U}}) + \{0, 2e_d a_d, 2v_d a_d\} (x f_D - x f_{\bar{D}})], \end{aligned} \quad (1.40)$$

9. Où la fonction de structure longitudinale est définie par :  $F_L = F_2 - 2xF_1$

10.  $e_u = \frac{2}{3}$  et  $e_d = -\frac{1}{3}$

Ici,  $xf_U$ ,  $xf_D$ ,  $xf_{\bar{U}}$  et  $xf_{\bar{D}}$  représentent respectivement les sommes des distributions de quarks de type “up” et “down” et de leurs antiquarks. En dessous du seuil de la masse du quark bottom ( $b$ ), ces sommes sont reliées aux distributions de quarks définies dans l'équation (1.16) de la façon suivante :

$$xf_U = xf_u + xf_c, \quad xf_{\bar{U}} = xf_{\bar{u}} + xf_{\bar{c}}, \quad xf_D = xf_d + xf_s, \quad xf_{\bar{D}} = xf_{\bar{d}} + xf_{\bar{s}}, \quad (1.41)$$

Lorsqu'on suppose une symétrie entre les distributions de quarks et antiquarks de la mer (c'est-à-dire  $xf_s = xf_{\bar{s}}$  et  $xf_c = xf_{\bar{c}}$ ), les distributions de quarks de valence s'écrivent :

$$xf_{u_{\text{val}}} = xf_U - xf_{\bar{U}}, \quad xf_{d_{\text{val}}} = xf_D - xf_{\bar{D}} \quad (1.42)$$

Il résulte de l'équation (1.37) que la fonction de structure  $x\tilde{F}_3$  peut être déterminée à partir de la différence entre les sections efficaces réduites  $e^+P$  et  $e^-P$  :

$$x\tilde{F}_3 = \frac{Y_+}{2Y_-} (\sigma_{r,\text{NC}}^- - \sigma_{r,\text{NC}}^+). \quad (1.43)$$

Les équations (1.38), (1.40) et (1.42) démontrent que dans le QPM,  $x\tilde{F}_3$  est directement reliée aux distributions de quarks de valence. Dans la gamme cinématique des données de HERA utilisées dans l'analyse présentée dans le chapitre 3, la contribution dominante de  $x\tilde{F}_3$  vient de l'interférence entre l'échange du photon  $\gamma$  et le boson neutre  $Z$ ; et une simple relation émerge :

$$xF_3^{\text{YZ}} \approx \frac{1}{3} (2xf_{u_{\text{val}}} + xf_{d_{\text{val}}}). \quad (1.44)$$

La mesure de  $xF_3^{\text{YZ}}$  par conséquent permet d'accéder aux comportements des distributions de quarks de valence à bas  $x$ , sous l'hypothèse que les distributions de quarks et antiquarks de la mer sont les mêmes.

Par analogie aux processus faisant intervenir les courants neutres, la section efficace inclusive de la diffusion inélastique profonde de l'électron (positron) sur le proton faisant intervenir les courants chargés peut être exprimée en terme d'autres ensembles de fonctions de structure :

$$\begin{aligned} \frac{d^2\sigma_{\text{CC}}^\pm}{dx dQ^2} &= \frac{G_F^2}{4\pi x} \left( \frac{M_W^2}{Q^2 + M_W^2} \right)^2 \left[ -y^2 F_L^{W^\pm} + Y_+ F_2^{W^\pm} \mp Y_- xF_3^{W^\pm} \right] \\ &= \frac{G_F^2}{2\pi x} \left( \frac{M_W^2}{Q^2 + M_W^2} \right)^2 \sigma_{r,\text{CC}}^\pm, \end{aligned} \quad (1.45)$$

où

$$\sigma_{r,\text{CC}}^\pm = -\frac{y^2}{2} F_L^{W^\pm} + \frac{Y_+}{2} F_2^{W^\pm} \mp \frac{Y_-}{2} xF_3^{W^\pm}, \quad (1.46)$$

et  $G_F$ ,  $M_W$  représentent respectivement la constante de Fermi et la masse du boson  $W$ .

Dans le QPM,  $F_L^{W^\pm} = 0$  et  $F_2^{W^\pm}$ ,  $xF_3^{W^\pm}$  représentent respectivement la somme et la différence des distributions de quarks et antiquarks, dépendemment de la charge du lepton initial :

$$F_2^{W^-} \approx xf_U + xf_{\bar{D}}, \quad F_2^{W^+} \approx xf_D + xf_{\bar{U}}, \quad (1.47)$$

$$xF_3^{W^-} \approx xf_U - xf_{\bar{D}}, \quad xF_3^{W^+} \approx xf_D - xf_{\bar{U}}. \quad (1.48)$$

À partir des équations (1.47) et (1.48), il résulte que :

$$\sigma_{r,CC}^- \approx xf_U + (1-y)^2 xf_{\bar{D}}, \quad \sigma_{r,CC}^+ \approx xf_{\bar{U}} + (1-y)^2 xf_D. \quad (1.49)$$

La combinaison des données de diffusion inélastique profonde faisant intervenir les courants neutres (NC) et chargés (CC) rend possible la détermination à la fois des distributions combinées de quarks de la mer  $xf_{\bar{U}}$  et  $xf_{\bar{D}}$ , et des distributions de quarks de valence  $xf_{u_{\text{val}}}$  et  $xf_{d_{\text{val}}}$ .

Les relations au sein du QPM illustrent de manière simple les données qui fournissent des informations. L'une des prédictions les plus frappantes du QPM est que dans la limite de Björken où  $Q^2$  et  $\nu \rightarrow \infty$  pour des valeurs fixées de  $x$ ,  $F_i(x, Q^2) \rightarrow F_i(x)$  BJÖRKEN [1969], c'est-à-dire les fonctions de structure ne dépendent plus de l'échelle d'énergie  $Q^2$  : c'est l'invariance d'échelle. Cette propriété est reliée à l'hypothèse que dans ce modèle, l'impulsion transverse des partons dans le référentiel où l'impulsion du proton est infini, est petit. On néglige ainsi le mouvement de Fermi des quarks dans le proton, ainsi que les corrections QCD.

Par contre, en QCD, le rayonnement de gluons à partir des quarks viole cette hypothèse, conduisant à des violations d'échelles de façon logarithmique, qui sont particulièrement grandes pour des petites valeurs de  $x$ , (voir figure 1.3). Le rayonnement de gluons produit une évolution des fonctions de structure. Le gluon rayonné produit à son tour une paire de quark - antiquark. Plus grande est l'échelle  $Q^2$ , plus de telles fluctuations quantiques peuvent être observées et donc le nombre de paires quarks - antiquarks et gluons dans la mer partonique s'accroît. Bien que ces partons de la mer portent seulement une petite fraction du quadri-vecteur impulsion du proton, leur nombre croissant conduit à l'affaiblissement de la distribution des quarks de valence quand  $Q^2$  augmente.

Comme nous l'allons mentionné dans les sections précédentes, les PDFs ne sont pas calculables à partir des premiers principes de la QCD (perturbative), toutefois leur dépendance en échelle d'énergie  $Q^2$  est perturbativement calculable ordre par ordre en QCD et est décrite par un ensemble de  $2n_f + 1$  équations intégrales différentielles couplées appelées les équations d'évolution de DGLAP, qui jouent un rôle d'équations du groupe de renormalisation (RGE) pour les PDFs, et qui s'écrivent comme suit dans l'espace des  $x$  :

$$\frac{\partial}{\partial \ln Q^2} \begin{pmatrix} f_i(x, Q^2) \\ f_{\bar{i}}(x, Q^2) \\ f_g(x, Q^2) \end{pmatrix} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} \begin{pmatrix} P_{ik} & P_{i\bar{l}} & P_{ig} \\ P_{\bar{i}k} & P_{\bar{i}\bar{l}} & P_{\bar{i}g} \\ P_{gk} & P_{g\bar{l}} & P_{gg} \end{pmatrix} \Bigg|_{z, \alpha_s(Q^2)} \begin{pmatrix} f_k(\frac{x}{z}, Q^2) \\ f_{\bar{l}}(\frac{x}{z}, Q^2) \\ f_g(\frac{x}{z}, Q^2) \end{pmatrix}, \quad (1.50)$$

où  $f_i(x, Q^2)$ ,  $f_{\bar{i}}(x, Q^2)$  et  $f_g(x, Q^2)$  représentent respectivement la distribution d'un quark de saveur  $i$ , la distribution d'un antiquark de saveur  $i$  ( $i = u, d, s, c, b$ ) et la distribution d'un gluon  $g$  à l'intérieur du proton  $P$ , et où  $P_{ik}, P_{i\bar{l}}, P_{ig}, P_{\bar{i}k}, P_{\bar{i}\bar{l}}, P_{\bar{i}g}, P_{gk}, P_{g\bar{l}}, P_{gg}$  correspondent aux fonctions de «splitting» régularisées calculables ordre par ordre en QCD perturbative, qui sont toutes fonctions de  $z$  et de  $\alpha_s(Q^2)$  et qui peuvent être génériquement développées en série de perturbation :

$$P_{ab}(z, \alpha_s(Q^2)) = \sum_{n=0}^{\infty} \left( \frac{\alpha_s(Q^2)}{2\pi} \right)^n P_{ab}^{(n)}(z) \quad (1.51)$$

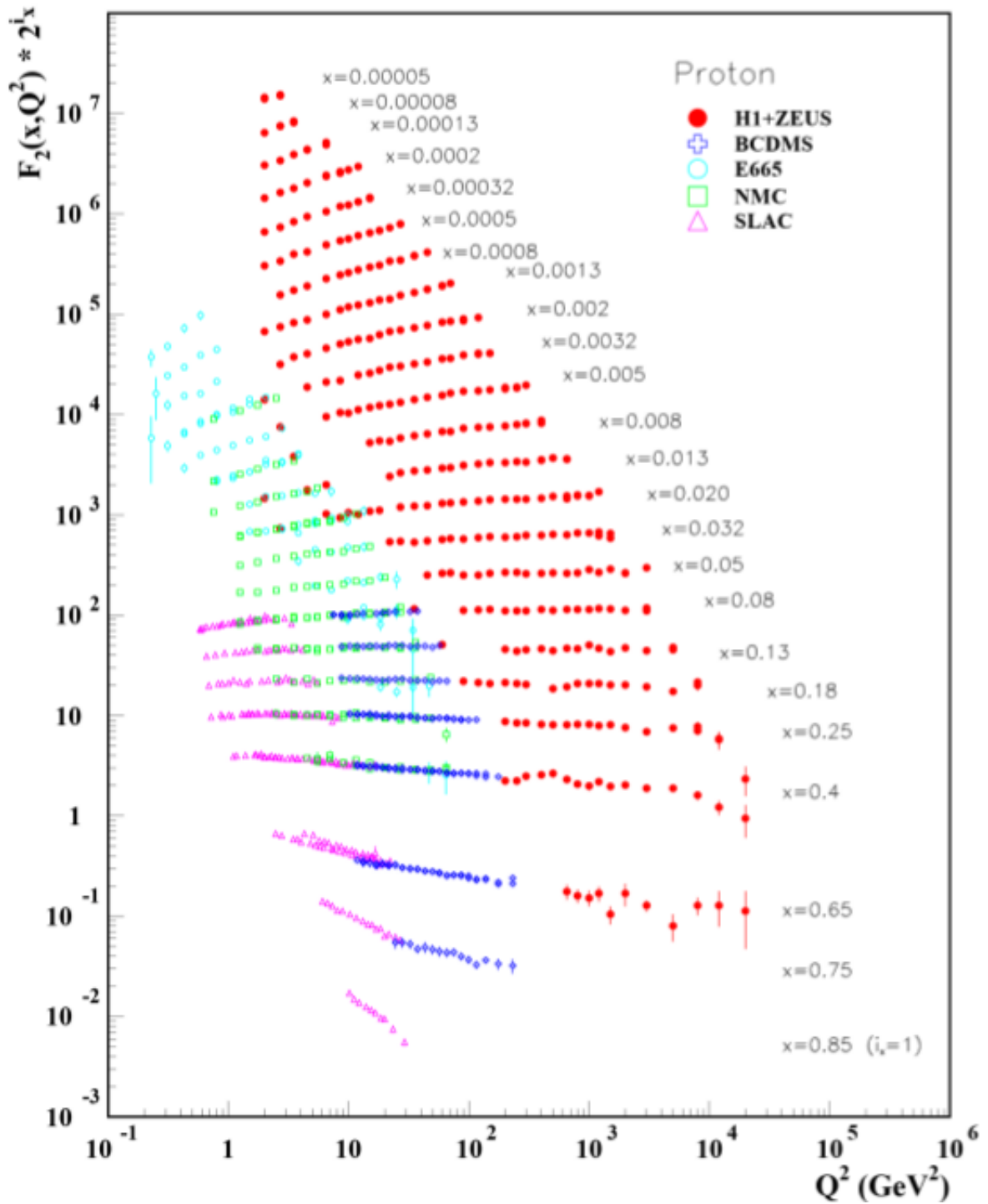


FIGURE 1.3 – La fonction de structure  $F_2^p$  en fonction de  $Q^2$  pour différentes valeurs de  $x$ , déterminées à partir des données combinées de diffusions inélastiques profondes des électrons et positrons sur les protons par l'échange d'un photon  $\gamma$  pour  $Q^2 \geq 2\text{GeV}^2$ , par les collaborations H1 + ZEUS de HERA. Les données des expériences sur cible fixe (SLAC, BCDMS, E665 et NMC) sont montrées à titre de comparaison (la figure est tirée de OLIVE et AL [2014]). On observe la dépendance de  $F_2^p(x, \alpha_s(Q^2))$  avec  $Q^2$ , comme prédit par la QCD perturbative. En pratique, les données correspondantes à la section efficace réduite (équation(1.39)) sont ajustées, plutôt que les fonctions de structure  $F_2$  et  $F_L$  séparément.

À cause de l'invariance de la conjugaison de charge et de la symétrie de saveur  $SU(n_f)_f$ , il est aussi intéressant de noter qu'à tous les ordres, nous avons les relations suivantes entre les fonctions de «splitting»

$$\begin{aligned} P_{ij} &= P_{i\bar{j}}, & P_{gi} &= P_{g\bar{i}} \equiv P_{gi}, \\ P_{i\bar{j}} &= P_{\bar{i}j}, & P_{ig} &= P_{\bar{i}g} \equiv P_{ig}, \end{aligned} \quad (1.52)$$

ce qui entre autre souligne simplement le fait que  $P_{gi}$  et  $P_{ig}$  sont indépendantes de la saveur de quark. Par ailleurs, puisque la QCD conserve la saveur, les fonctions de «splitting»  $P_{ij}^{(0)}$  sont toutes nulles sauf dans le cas  $i = j = q$ . Ainsi, nous rappelons les fonctions de «splitting» régularisées à  $\mathcal{O}(\alpha_s^0)$  **ALTARELLI et PARISI [1977]**

$$P_{qq}^{(0)}(z) = C_F \left[ \frac{1+z^2}{(1-z)_+} + \frac{3}{2} \delta(1-z) \right], \quad (1.53)$$

$$P_{gq}^{(0)}(z) = C_F \left[ \frac{1+(1-z)^2}{z} \right], \quad (1.54)$$

$$P_{qg}^{(0)}(z) = T_F \left[ z^2 + (1-z)^2 \right], \quad (1.55)$$

$$P_{gg}^{(0)}(z) = 2C_A \left[ \frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right] + \delta(1-z) \frac{(11C_A - 4n_f T_F)}{6}, \quad (1.56)$$

où la distribution “+” est définie de sorte que son intégrale avec toute fonction “test” suffisamment régulière  $f$  est :

$$\int_0^1 dz \frac{f(z)}{(1-z)_+} = \int_0^1 dz \frac{f(z) - f(1)}{(1-z)}, \quad (1.57)$$

et

$$\frac{1}{(1-z)_+} = \frac{1}{1-z}, \quad \text{quand } 0 \leq z < 1. \quad (1.58)$$

Les fonctions de «splitting» dans le secteur des quarks peuvent en fait devenir très non triviales au-delà de l'ordre dominant. Une manière particulièrement pratique de résoudre ce problème est d'utiliser la symétrie de saveur  $SU(3)$  du Lagrangian de QCD pour décomposer les fonctions de «splitting» en combinaison de quarks singlets et de quarks non singlets telle que

$$\begin{aligned} P_{ij} &= P_{i\bar{j}} = \delta_{ij} P_{qq}^V + P_{qq}^S, \\ P_{i\bar{j}} &= P_{\bar{i}j} = \delta_{ij} P_{q\bar{q}}^V + P_{q\bar{q}}^S. \end{aligned} \quad (1.59)$$

Notons de plus que la quantité de saveur diagonale  $P_{qq}^V$  est déjà non nulle à l'ordre dominant tandis que  $P_{q\bar{q}}^V$  et les diverses contributions aux quarks de la mer  $P_{qq}^S$  et  $P_{q\bar{q}}^S$  commencent seulement à contribuer aux équations de DGLAP à l'ordre NLO. La même chose est valable pour la combinaison de quarks pure singlet définie ci-dessous

$$P_{ps} = n_f (P_{qq}^S + P_{q\bar{q}}^S) \quad (1.60)$$

Une méthode efficace pour calculer les distributions de quarks au delà de l'ordre dominant est d'utiliser la décomposition des saveurs de (1.59) pour réorganiser (1.50) en introduisant une distribution de quark singlet  $q_s$  définie par

$$q_s = \sum_{i=1}^{n_f} (f_i + f_{\bar{i}}), \quad (1.61)$$

et dont l'évolution est maximalelement couplée à la distribution de gluon, et les combinaisons non singulet définies ci-dessous

$$\begin{aligned} q_{ns,ij}^\pm &= f_i \pm f_{\bar{i}} - (f_j \pm f_{\bar{j}}) \\ q_{ns}^V &= \sum_{i=1}^{n_f} (f_i - f_{\bar{i}}), \end{aligned} \quad (1.62)$$

dont les évolutions découplent complètement de la densité de gluon. Ces trois densités de quarks non singulets évoluent avec les noyaux de «splitting» suivants :

$$\begin{aligned} P_{ns}^\pm &= P_{qq}^V \pm P_{q\bar{q}}^V, \\ P_{ns}^V &= P_{qq}^V - P_{q\bar{q}}^V + n_f (P_{qq}^S - P_{q\bar{q}}^S) = P_{ns}^- + n_f (P_{qq}^S - P_{q\bar{q}}^S). \end{aligned} \quad (1.63)$$

En résumé, les équations d'évolution des densités de quarks non singulets sont :

$$\frac{\partial q_{ns,ij}^\pm(x, Q^2)}{\partial \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{ns}^\pm(z, \alpha_s(Q^2)) q_{ns,ij}^\pm\left(\frac{x}{z}, Q^2\right), \quad (1.64)$$

$$\frac{\partial q_{ns}^V(x, Q^2)}{\partial \ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{ns}^V(z, \alpha_s(Q^2)) q_{ns}^V\left(\frac{x}{z}, Q^2\right). \quad (1.65)$$

Et les équations d'évolution couplées des distributions de quarks singulets et de gluon sont :

$$\frac{\partial}{\partial \ln Q^2} \begin{pmatrix} q_s(x, Q^2) \\ f_g(x, Q^2) \end{pmatrix} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} \begin{pmatrix} (P_{ns}^+ + P_{ps}) & 2n_f P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \Big|_{z, \alpha_s(Q^2)} \begin{pmatrix} q_s\left(\frac{x}{z}, Q^2\right) \\ f_g\left(\frac{x}{z}, Q^2\right) \end{pmatrix}, \quad (1.66)$$

Il existe essentiellement deux types d'approches pour résoudre les équations d'évolution (1.64), (1.65) et (1.66). Le premier consiste à intégrer numériquement ces équations intégral-différentielles dans l'espace de la variable de björken  $x$ , alors que la seconde option consiste à utiliser la transformation intégrale de Mellin pour les PDFs définie par

$$f_a(N, Q^2) = \int_0^1 dx x^{N-1} f_a(x, Q^2), \quad a = q, \bar{q}, g \quad (1.67)$$

pour résoudre ces équations d'évolution dans l'espace des moments  $N$  avant d'utiliser la transformation inverse de Mellin pour obtenir ces PDFs dans l'espace  $x$

$$f_a(x, Q^2) = \frac{1}{2\pi i} \int_C dN x^{-N} f_a(N, Q^2), \quad (1.68)$$

où  $C$  est le contour d'intégration dans le plan complexe  $N$ .

L'avantage de la méthode de Mellin est qu'elle convertit l'opération de convolution en une simple multiplication, c'est-à-dire que les équations intégral-différentielles de DGLAP dans l'espace  $x$  sont transformées en simples équations différentielles

(1.69) dans l'espace N :

$$\begin{aligned}
\frac{\partial f_a(N, Q^2)}{\partial \ln Q^2} &= \int_0^1 dx x^{N-1} \frac{\partial f_a(x, Q^2)}{\partial \ln Q^2}, \\
&= \int_0^1 dx x^{N-1} \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{dy}{y} P_{ab}\left(\frac{x}{y}, \alpha_s(Q^2)\right) f_b(y, Q^2), \\
&= \int_0^1 dx x^{N-1} \frac{\alpha_s(Q^2)}{2\pi} \int_0^1 dy \int_0^1 dz P_{ab}(z, \alpha_s(Q^2)) f_b(y, Q^2) \delta(x - yz), \\
&= \frac{\alpha_s(Q^2)}{2\pi} \int_0^1 dz z^{N-1} P_{ab}(z, \alpha_s(Q^2)) \int_0^1 dy y^{N-1} f_b(y, Q^2), \\
\frac{\partial f_a(N, Q^2)}{\partial \ln Q^2} &= \frac{\alpha_s(Q^2)}{2\pi} \sum_b \Upsilon_{ab}(N, \alpha_s(Q^2)) f_b(N, Q^2), \tag{1.69}
\end{aligned}$$

où les  $\Upsilon_{ab}(N, \alpha_s(Q^2))$  sont les dimensions anormales et correspondent à la transformée de Mellin des fonctions de «splitting»  $P_{ab}(x, \alpha_s(Q^2))$

$$\Upsilon_{ab}(N, \alpha_s(Q^2)) = \int_0^1 dx x^{N-1} P_{ab}(x, \alpha_s(Q^2)). \tag{1.70}$$

Jusqu'à présent, nous nous sommes surtout concentrés sur la dépendance en échelle d'énergie  $Q^2$  des PDFs, et nous avons vu que cette dépendance en  $Q^2$  est perturbativement calculable ordre par ordre en QCD perturbative et est décrite par un ensemble d'équation integro-différentielles couplées appelées équations d'évolution de DGLAP. Pourtant, nous n'avons pas beaucoup parlé de la dépendance en  $x$  des PDFs. La dépendance en  $x$  des PDFs n'est pas prédite par la QCD perturbative, mais doit être déterminée par une analyse globale des PDFs (confère section 1.5).

Nous serons amenés à considérer dans ce travail plusieurs schémas de saveurs lourdes qui permettent de prendre en compte les effets de la masse des quarks lourds. Les principaux schémas existants sont brièvement présentés dans la section suivante.

## 1.4 Les schémas de saveurs lourdes

Pour les analyses de précision des PDFs et, par principe, les effets de la masse des quarks lourds doivent être correctement pris en compte. Une grande partie de la littérature a été consacrée à cette question et nous nous référons aux revues récentes [ANDERSEN et AL](#); [OLNESS et SCHIENBEIN \[2009\]](#); [THORNE et TUNG](#) pour un aperçu des différents schémas de saveurs lourdes utilisés dans les déterminations des PDFs. Chacun de ces schémas est brièvement discuté ci-dessous.

### 1.4.1 Le schéma à nombre de saveur variable de masse nulle

Dans le schéma à nombre de saveur variable de masse nulle (“zero-mass variable flavor number scheme”, ZM-VFNS), les quarks lourds apparaissent comme des partons dans le proton aux valeurs de  $Q^2$  au-dessus de  $\sim m_H^2$  et ils sont considérés comme sans masses à la fois dans l'état initial et final du processus de diffusion dure. On s'attend à ce que ce schéma soit fiable uniquement dans la région où  $Q^2 \gg m_H^2$ , et c'est pas précis pour des valeurs de  $Q^2 \ll m_H^2$  puisqu'il ne contient pas des corrections de l'ordre de  $m_H^2/Q^2$ , alors que les autres schémas mentionnés ci-dessous sont précis jusqu'à l'ordre  $\Lambda_{\text{QCD}}^2/Q^2$ , bien qu'avec des ordres de perturbation différents.



### 1.4.2 Le schéma à nombre de saveur fixée

Dans le schéma à nombre de saveur fixée (“fixed flavor number scheme”, FFNS), seul le gluon et les quarks légers sont considérés comme des partons dans le proton et les quarks massifs sont produits de manière perturbative dans l’état final. Ce schéma est fiable uniquement pour les valeurs de  $Q^2 \sim m_H^2$  puisqu’il ne resomme pas les logarithmes de la forme  $\ln(Q^2/m_H^2)$  qui deviennent importants pour les valeurs de  $Q^2 \gg m_H^2$ .

### 1.4.3 Le schéma à nombre de saveur variable de masse générale

Le schéma à nombre de saveur variable de masse générale (“general-mass variable flavor number scheme”, GM-VFNS), permet une transition naturelle au schéma sans masse (ZM-VFNS) et contient des termes de masse, ce qui donne une bonne description au voisinage des seuils de la masse des quarks lourds. La procédure qui permet cette transition n’est pas unique et plusieurs réalisations du GM-VFNS ont été proposées dans la littérature :

#### Schéma TR

Le schéma Thorne-Roberts (TR) [THORNE et ROBERTS \[1998\]](#) qui est utilisé dans les analyses globales de la collaboration MSTW, était désigné à fournir une douce transition entre le schéma massif FFNS à basses échelles  $Q^2 \sim m_H^2$  et le schéma sans masse ZM-VFNS à hautes échelles  $Q^2 \gg m_H^2$ . Puisque la version originale était techniquement difficile à implémenter au delà de NLO, il a été mis à jour selon le schéma TR’ [THORNE \[2006\]](#). Il existe deux variantes du schéma TR’ : TR’ standard (comme utilisé dans les ensembles de PDFs du groupe MSTW [MARTIN et collab. \[2009\]](#); [THORNE \[2006\]](#)) et TR’ optimal [THORNE \[2012\]](#), avec une transition plus douce à travers la région de seuil de quark lourd.

#### Schéma ACOT

Le schéma Aivazis-Collins-Olness-Tung (ACOT) fait partie du groupe de schémas de factorisation VFNS qui utilise la méthode de renormalisation de Collins-Wilczek-Zee (CWZ) [COLLINS \[1998\]](#). Ce schéma implique un mélange du schéma  $\overline{\text{MS}}$  pour les partons légers (et pour les partons lourds quand l’échelle de factorisation est plus grand que la masse du quark lourd) et un schéma de renormalisation de soustraction d’impulsion nulle pour les graphes possédant des lignes de quark lourd (si l’échelle de factorisation est plus petite que le seuil de la masse du quark lourd). Le schéma ACOT est alors considéré comme une extension minimale du schéma  $\overline{\text{MS}}$ .

Les différences entre les schémas TR et ACOT sont résumées dans la figure 1.4. Un problème majeur dans un schéma GM-VFNS complet, est celui de l’ordre dans le développement perturbatif. L’équivalence de l’échange des termes  $\mathcal{O}(m_H^2/Q^2)$  entre les coefficients de Wilson (ou les amplitudes de diffusion dure) sans violer la définition du schéma GM-VFNS; est ce qui distingue principalement le schéma ACOT du schéma TR. Ici,  $m_H$  représente la masse du quark lourd H.

Dans le but d’accélérer le temps de calcul du  $\chi^2$  (voir 1.5.1.1), nous avons utilisé le schéma ZM-VFNS pour obtenir les premiers résultats publiés dans [GBEDO et MANGIN-BRINET \[2017\]](#) et présenté dans le chapitre 3.



TR type schemes			ACOT type schemes			
$Q < m_H$	$Q > m_H$	constant term	$Q < m_H$	$Q > m_H$	constant term	
LO			LO	$\emptyset$		$+\emptyset$
NLO			NLO			$+\emptyset$
NNLO			NNLO			$+\emptyset$

FIGURE 1.4 – Différences entre le schéma ACOT et TR. Figure tirée de [xFITTER DEVELOPERS \[2017\]](#)

## 1.5 L'analyse globale des PDFs en QCD

Comme cela a été discuté dans la section 1.3, les PDFs ne peuvent pas être calculées en théorie des perturbations, mais plutôt elles doivent être extraites à partir des données expérimentales comme la diffusion inélastique profonde. Dans cette section, nous présentons brièvement la procédure standard pour déterminer les PDFs et leurs incertitudes associées à partir des données expérimentales, et les motivations qui nous ont conduites à proposer une nouvelle approche, fondée sur des méthodes Monte Carlo.

### 1.5.1 La procédure standard

La procédure standard pour obtenir l'ensemble des 11 différentes densités de partons<sup>11</sup>

$$\begin{aligned}
 &xf_u, xf_{\bar{u}}, \quad xfd, xf_{\bar{d}}, \quad xfs, xf_{\bar{s}}, \\
 &xfc, xf_{\bar{c}}, \quad xfb, xf_{\bar{b}}, \quad xfg,
 \end{aligned} \tag{1.71}$$

consiste à paramétriser la dépendance  $x$  pour un ensemble de données de différents processus de diffusion dure à une échelle d'énergie initiale (basse)  $Q_0^2 < m_c^2$ , où  $m_c$  est la masse du quark charmé, avant d'évoluer les distributions jusqu'à l'échelle d'énergie physique  $Q^2$  qui nous intéresse au moyen des équations d'évolution de DGLAP.

En pratique, les masses des quarks de type "charm" et "bottom" sont très grandes comparativement à  $\Lambda_{\text{QCD}}$  c'est-à-dire,  $m_c, m_b \gg \Lambda_{\text{QCD}}$  de telle sorte que les densités de quarks lourds sont déterminées perturbativement. Aussi, compte tenu du peu d'informations expérimentales que nous avons actuellement sur le quark de valence de type "strange", nous supposons que  $xf_s = xf_{\bar{s}}$ . Nous allons immédiatement énumérer de façon concrète les différentes étapes de la procédure standard :

11. On fait l'hypothèse que l'isospin est une symétrie, c'est-à-dire, si on décide de sonder le neutron à la place du proton alors on doit remplacer le quark de type "up" par le quark de type "down" et vice versa.

- La première étape consiste à paramétrer les PDFs à une échelle initiale  $Q_0^2$  :

$$x f_a(x, Q_0^2) = A_a x^{B_a} (1-x)^{C_a} P_a(x, D_a, E_a), \quad a = u, d, \bar{u}, \bar{d}, \bar{s}, g \quad (1.72)$$

où  $P_a(x)$  est un polynôme simple qui permet l'interpolation entre les régions de bas et haut  $x$ . La forme de ce polynôme diffère d'un groupe de PDFs à l'autre et le choix de la paramétrisation constitue donc une source d'incertitude sur les PDFs produites par les groupes d'analyse globales à l'exception de la collaboration NNPDF [BALL et AL. \[2010\]](#); [DEL-DEBBIO et AL \[2007\]](#); [ROJO \[2006\]](#) qui propose une technique basée sur les réseaux de neurones pour éviter les incertitudes dues à la paramétrisation.

Dans ce travail, nous utilisons la forme fonctionnelle proposée par le groupe HERAPDF

$$x f_a(x, Q_0^2) = A_a x^{B_a} (1-x)^{C_a} (1 + D_a x + E_a x^2), \quad a = g, u_{\text{val}}, d_{\text{val}}, \bar{U}, \bar{D}. \quad (1.73)$$

La contribution de la distribution de quark  $x f_s(x, Q_0^2)$  est considérée comme proportionnelle à la distribution de quark  $x f_{\bar{D}}(x, Q_0^2)$  en posant  $x f_s(x, Q_0^2) = r_s x f_{\bar{D}}(x, Q_0^2)$ , avec

$$r_s = 0.31, \quad (1.74)$$

comme suggéré dans [KAYIS-TOPAKSU et collab. \[2008\]](#). La contrainte supplémentaire  $A_{\bar{U}} = A_{\bar{D}}(1 - r_s)$ , ensemble avec la condition requise  $B_{\bar{U}} = D_{\bar{D}}$ , assure que  $x \bar{u} \rightarrow x \bar{d}$  lorsque  $x \rightarrow 0$ . Les paramètres de normalisation  $A_{u_{\text{val}}}$  et  $A_{d_{\text{val}}}$  sont contraints par la conservation du nombre de quarks de valence de type "up" et "down" dans le proton

$$\int_0^1 f_{u_{\text{val}}}(x, Q_0^2) dx = 2 \quad \Rightarrow \exists \text{ deux quarks de valence } u_{\text{val}} \text{ dans le proton,} \quad (1.75)$$

$$\int_0^1 f_{d_{\text{val}}}(x, Q_0^2) dx = 1 \quad \Rightarrow \exists \text{ un quark de valence } d_{\text{val}} \text{ dans le proton.} \quad (1.76)$$

tandis que le paramètre de normalisation  $A_g$  est contraint par la conservation de la fraction du quadri-vecteur impulsion portée par les partons

$$\int_0^1 \left[ x f_U(x, Q_0^2) + x f_{\bar{U}}(x, Q_0^2) + x f_D(x, Q_0^2) + x f_{\bar{D}}(x, Q_0^2) + x f_g(x, Q_0^2) \right] dx = 1. \quad (1.77)$$

- Ensuite les PDFs (1.73) seront évoluées jusqu'à l'échelle expérimentale  $Q_{exp}^2$ , en utilisant les solutions des équations de DGLAP (1.64), (1.65) et (1.66) et utilisées pour évaluer les observables théoriques telles que les fonctions de structure et les sections efficaces à l'aide des théorèmes de factorisation (voir sous-section 1.3.1). Beaucoup de différents types d'expériences sont nécessaires pour une détermination complète des distributions de partons :

- l'ensemble de donnée de la fonction de structure  $F_2^p(x, Q^2)$  fournit par les collaborations H1 [ADLOFF et AL \[2000, 2001a,b\]](#) et ZEUS [CHEKANOV et AL \[2001\]](#) est la donnée généralement utilisée et couvre les petites valeurs de  $x$ <sup>12</sup> et une gamme étendue d'échelle d'énergie  $Q^2$ , les données  $F_2^p(x, Q^2)$  et  $F_2^d(x, Q^2)$  pour les valeurs intermédiaires de  $x$  fournies par la collaboration E665 [ADAMS et AL \[1996\]](#), les données  $F_2^p(x, Q^2)$  et  $F_2^d(x, Q^2)$  pour les grandes valeurs de  $x$  fournies par les collaborations BCDMS [BENVENUTI et AL \[1989a,b\]](#)

12.  $x$  est la variable de Björken.

et SLAC [WHITLOW \[1990\]](#); [WHITLOW et AL \[1992\]](#), les données  $F_2^P(x, Q^2)$  et  $F_2^d(x, Q^2)$  pour les grandes valeurs et valeurs intermédiaires de  $x$  fournies par la collaboration NMC [ARNEODO et AL \[1997a,b\]](#), les données  $F_2^{V(\bar{v})P}(x, Q^2)$  et  $F_3^{V(\bar{v})P}(x, Q^2)$  pour les grandes valeurs de  $x$  fournies par la collaboration CCFR [SELIGMAN et AL \[1997\]](#); [YANG et AL \[2001\]](#) qui sondent indépendamment les quarks singulets et de valence ;

- les données  $F_{2,charm}^P(x, Q^2)$  et  $P+N \rightarrow \mu + \bar{\mu} + X$  pour les grandes valeurs de  $x$  fournies respectivement par les collaborations H1-ZEUS [ADLOFF et AL \[2002\]](#); [BREITWEG et AL \[2000\]](#) et E605 [MORENO et AL \[1991\]](#) pour contraindre les quarks de la mer ;

- les données de l'asymétrie du processus de Drell-Yan mesurées par la collaboration E866 [TOWELL et AL \[2001\]](#) pour déterminer  $f_{\bar{d}} - f_{\bar{u}}$  ;

- les données de l'asymétrie du boson W pour les grandes valeurs de  $x$  mesurées par la collaboration CDF [ABE et AL \[1998\]](#) pour contraindre le rapport  $f_u / f_d$  ;

- les données inclusives de "jet" pour les hautes valeurs de  $x$  fournies respectivement par les collaborations CDF [AFFOLDER et AL \[2001\]](#) et D0 [ABBOTT et AL \[2001\]](#) pour astreindre  $f_g$  ;

- Les données de production de dimuon par les collaborations NuTeV [GONCHAROV et AL \[2001\]](#) et CCFR [BAZARKO et AL \[1995\]](#) pour contraindre la distribution du quark de la mer de type "strange".

• Enfin les paramètres des PDFs seront ajustées aux données expérimentales en minimisant grâce à MINUIT [JAMES et ROOS \[1975\]](#) la fonction  $\chi^2$ .

### 1.5.1.1 Construction de la fonction $\chi^2$

La définition standard du  $\chi^2$  utilisée dans les analyses globales des PDFs était

$$\chi^2(\hat{q}) = \sum_{I=1}^{N_{\text{dat}}} \frac{(D_I - T_I(\hat{q}))^2}{(\sigma_{\text{stat},I}^2 + \sigma_{\text{uncor},I}^2)}, \quad (1.78)$$

où  $T_I(\hat{q})$ ,  $D_I$ ,  $\sigma_{\text{stat},I}$  et  $\sigma_{\text{uncor},I}$  sont la valeur de la prédiction théorique dépendant des paramètres libres  $\hat{q}$  des PDFs, la valeur de la mesure expérimentale, l'incertitude statistique et l'incertitude systématique non corrélée pour le point de donnée expérimentale  $I$ . La section efficace inclusive  $T_I$  dans les collisionneurs hadroniques peut être écrite comme une convolution des PDFs avec la section efficace partonique, calculée à un ordre donné en théorie des perturbations.

Toutefois cette définition ne tient pas en compte les incertitudes systématiques corrélées. Dans le cas où ces incertitudes sont petites, une alternative de calcul du  $\chi^2$  est d'incorporer ces incertitudes systématiques corrélées dans la prédiction théorique :

$$T_I^*(\hat{q}, \hat{\xi}) = T_I(\hat{q}) + \sum_{k=1}^{N_{\text{sys}}} \xi_k \Delta_{kI}^{\text{sys}}, \quad (1.79)$$

où  $N_{\text{sys}}$  représente le nombre total de sources d'incertitudes systématiques corrélées,  $\Delta_{kI}^{\text{sys}}$  est l'erreur systématique corrélée à  $1\sigma$  pour le point de donnée expérimentale  $I$  de la

source  $k$  et les paramètres  $\xi_k$  représentent des variables aléatoires indépendantes distribuées suivant la loi Gaussienne de moyenne 0 et de variance 1 pour chaque source d'incertitude systématique corrélée  $k$ . La fonction  $\chi^2$  prend dans ce cas la forme suivante :

$$\chi^2(\hat{q}, \hat{\xi}) = \sum_{I=1}^{N_{\text{dat}}} \frac{(D_I - T_I^*(\hat{q}, \hat{\xi}))^2}{(\sigma_{\text{stat},I}^2 + \sigma_{\text{uncor},I}^2)} + \sum_{k=1}^{N_{\text{sys}}} \xi_k^2, \quad (1.80)$$

où le second terme de (1.80) permet de contraindre les valeurs des paramètres des incertitudes systématiques corrélées  $\xi_k$ .

Lorsque le nombre de sources d'incertitude systématique est grand, des instabilités numériques peuvent apparaître lorsque l'on minimise avec MINUIT la fonction  $\chi^2$  de (1.80) par rapport aux paramètres  $\hat{\xi}$  et  $\hat{q}$ , c'est pourquoi il est préférable de faire la minimisation de (1.80) par rapport à  $\xi_k$  de façon analytique. On définit pour se faire les variables auxiliaires suivantes :

$$B_k = \sum_{I=1}^{N_{\text{dat}}} \frac{\Delta_{kI}(D_I - T_I(\hat{q}))}{(\sigma_{\text{stat},I}^2 + \sigma_{\text{uncor},I}^2)} \quad \text{et} \quad A_{kl} = \delta_{kl} + \sum_{I=1}^{N_{\text{dat}}} \frac{\Delta_{kI}\Delta_{lI}}{(\sigma_{\text{stat},I}^2 + \sigma_{\text{uncor},I}^2)}, \quad (1.81)$$

la dérivée partielle de (1.80) par rapport à  $\xi_k$  donne donc :

$$\xi_k(\hat{q}) = \sum_{l=1}^{N_{\text{sys}}} (A^{-1})_{kl} B_l. \quad (1.82)$$

En rassemblant les équations (1.81) et (1.82), on obtient après quelques calculs, la fonction  $\chi^2$  simplifiée :

$$\chi^2(\hat{q}) = \sum_{I=1}^{N_{\text{dat}}} \frac{(D_I - T_I(\hat{q}))^2}{(\sigma_{\text{stat},I}^2 + \sigma_{\text{uncor},I}^2)} - \sum_{k=1}^{N_{\text{sys}}} \sum_{l=1}^{N_{\text{sys}}} B_k (A^{-1})_{kl} B_l, \quad (1.83)$$

Le double avantage de l'équation (1.83) est qu'on voit explicitement le décalage de la mesure par rapport à la théorie et qu'elle nécessite l'inversion des matrices de petites dimensions ; ce qui permet d'éviter les éventuelles instabilités numériques comme c'est le cas pour l'équation (1.80).

Une fois la fonction  $\chi^2$  spécifiée, nous trouvons l'ensemble des paramètres  $\{q_0^{(i)}\}_{i=1}^d$  qui la minimise, c'est-à-dire  $\chi^2(\hat{q}_0) = \chi_{\text{min}}^2$ .

### 1.5.1.2 Application : ajustement du paramètre $B_g$

En guise d'illustration, on se propose ici d'utiliser la procédure standard pour ajuster seulement le paramètre libre  $B_g$  de  $x f_g(x, Q_0^2)$  aux données combinées de HERA I [AARON et AL. \[2010\]](#). Pour cela, on considère les formes fonctionnelles suivantes à l'échelle initiale  $Q_0^2 = 1.9 \text{ GeV}^2$  où nous utilisons à partir de maintenant les notations suivantes pour les PDFs :

$$xg(x, Q_0^2) = A_g x^{B_g} (1-x)^{C_g}, \quad (1.84)$$

$$xu_{\text{val}}(x, Q_0^2) \equiv (xU - x\bar{U})(x, Q_0^2) = A_{u_{\text{val}}} x^{B_{u_{\text{val}}}} (1-x)^{C_{u_{\text{val}}}} (1 + E_{u_{\text{val}}} x^2), \quad (1.85)$$

$$xd_{\text{val}}(x, Q_0^2) \equiv (xD - x\bar{D})(x, Q_0^2) = A_{d_{\text{val}}} x^{B_{d_{\text{val}}}} (1-x)^{C_{d_{\text{val}}}}, \quad (1.86)$$

$$x\bar{U}(x, Q_0^2) \equiv x\bar{u}(x, Q_0^2) = A_{\bar{U}} x^{B_{\bar{U}}} (1-x)^{C_{\bar{U}}}, \quad (1.87)$$

$$x\bar{D}(x, Q_0^2) \equiv (x\bar{d} + x\bar{s})(x, Q_0^2) = A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}}, \quad (1.88)$$

où les paramètres de normalisation  $A_{u_{\text{val}}}$ ,  $A_{d_{\text{val}}}$  et  $A_g$  sont respectivement fixés par les équations (1.75), (1.76) et (1.77), et tous les autres paramètres à l'exception de  $B_g$  sont fixés (voir tableau 1.3).

Parameter	Values
$C_g$	5.927
$B_{u_{\text{val}}} = B_{d_{\text{val}}}$	0.610
$C_{u_{\text{val}}}$	4.712
$E_{u_{\text{val}}}$	14.759
$C_{d_{\text{val}}}$	3.144
$A_{\bar{U}}$	0.210
$B_{\bar{U}} = B_{\bar{D}}$	-0.018
$C_{\bar{U}}$	4.052
$A_{\bar{D}}$	0.305
$C_{\bar{D}}$	5.875

TABLEAU 1.3 – Les valeurs des autres paramètres fixes des paramétrisations (1.84)-(1.88).

Le nombre total de points de données non corrélées est  $N_{\text{dat}} = 537$  après la coupure  $Q^2 \geq 10 \text{ GeV}^2$ , désignée pour réduire les sources d'erreurs théoriques. L'ordre de perturbation des observables théoriques correspondantes aux données considérées est "NLO". La constante de couplage fort est fixée à  $\alpha_s(M_Z^2) = 0.1176$ , avec  $M_Z = 91.187 \text{ GeV}$ . Ce qui conduit au résultat suivant :

$$B_g = -0.056 \pm 0.002 . \quad (1.89)$$

Le tableau 1.4 montre à la fois  $\chi_{\text{min}}^2$  par degré de liberté noté d.o.f<sup>13</sup> et la contribution de chacun des 4 ensembles de données non corrélés au  $\chi_{\text{min}}^2$  total.

Data set	$\chi_{\text{min}}^2 / N_{\text{dat}}$
NC DIS cross sections HERA I H1-ZEUS combined $e^- p$	114/145
NC DIS cross sections HERA I H1-ZEUS combined $e^+ p$	301/324
CC DIS cross sections HERA I H1-ZEUS combined $e^- p$	21/34
CC DIS cross sections HERA I H1-ZEUS combined $e^+ p$	30/34
Total $\chi_{\text{min}}^2 / \text{d.o.f}$	466/536

TABLEAU 1.4 – Résultats de l'ajustement du paramètre  $B_g$  aux données de HERA I. La contribution de chaque ensemble de données au  $\chi_{\text{min}}^2$  total et le nombre de points correspondant sont montrés.

Une illustration de la distribution  $xg(x, Q^2)$  obtenue par ajustement du paramètre  $B_g$  aux données mentionnées plus haut est :

13. Le degré de liberté d.o.f est la différence entre le nombre total de points de données  $N_{\text{dat}}$  et le nombre de paramètres libres  $d$  à ajuster aux données en considération.

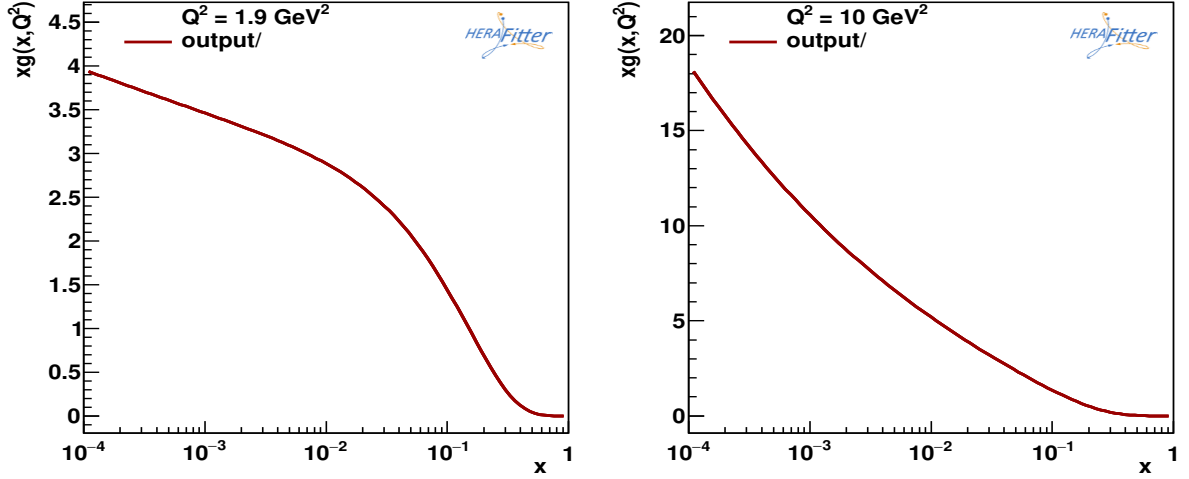


FIGURE 1.5 – Courbes montrant la valeur centrale de la distribution du gluon (1.84) aux échelles  $Q^2 = 1.9 \text{ GeV}^2$  (gauche) et  $Q^2 = 10 \text{ GeV}^2$  (droite) dans le schéma où tous les quarks lourds sont sans masses (1.4.1).

Sur la figure 1.5, nous constatons que la distribution du gluon augmente très rapidement à bas  $x$  lorsqu'on passe de l'échelle d'énergie  $Q_0^2$  à l'échelle d'énergie  $Q^2 = 10 \text{ GeV}^2$  : c'est l'effet de l'évolution de DGLAP.

Il y a quelques années en arrière, la détermination des densités de partons qui ajustent le mieux les données expérimentales en considération était suffisante pour une étude phénoménologique pratique. Toutefois, avec l'avènement de nouvelles données issues des collisionneurs hadroniques, la nécessité d'estimer quantitativement les incertitudes des PDFs est d'une importance primordiale. Nous allons sans tarder aborder les différentes approches pour obtenir les incertitudes expérimentales sur les PDFs.

## 1.6 Les incertitudes expérimentales des PDFs

Nous allons ici rappeler les différentes approches existantes pour obtenir les incertitudes expérimentales sur les PDFs.

### 1.6.1 L'approche de Hessien

Dans l'approche Hésienne PUMPLIN et collab. [2001a,b], l'incertitude expérimentale associée aux PDFs s'obtient en définissant la matrice Hésienne :

$$\chi^2 - \chi_{\min}^2 \equiv \Delta\chi^2 = \sum_{i=1}^d \sum_{j=1}^d H_{ij} (q^{(i)} - q_0^{(i)}) (q^{(j)} - q_0^{(j)}), \quad (1.90)$$

où  $H_{ij}$  est un élément de la matrice Hésienne défini par :

$$H_{ij} = \frac{\partial^2 \chi^2(\hat{q}, \hat{\xi})}{\partial q^{(i)} \partial q^{(j)}}. \quad (1.91)$$

Nous pouvons donc utiliser la formule standard de propagation linéaire des incertitudes expérimentales des PDFs pour calculer l'erreur sur toute quantité  $F$  :

$$(\Delta\mathcal{F})^2 = \sum_{i=1}^d \sum_{j=1}^d \frac{\partial \mathcal{F}}{\partial q^{(i)}} C_{ij}(\hat{q}) \frac{\partial \mathcal{F}}{\partial q^{(j)}}, \quad (1.92)$$

où  $C_{ij}$  est un élément de la matrice de covariance des paramètres libres  $\hat{q}$  des PDFs :

$$C_{ij}(\hat{q}) = \Delta\chi^2 (H^{-1})_{ij} . \quad (1.93)$$

Il est plus pratique et numériquement plus stable de diagonaliser la matrice de covariance à cause de la symétrie de la matrice Hessienne. Ceci permet donc de travailler dans la base des vecteurs propres orthogonaux  $\{v_{ik}\}$  définis par :

$$\sum_{j=1}^d C_{ij} v_{jk} = \epsilon_k v_{ik} , \quad (1.94)$$

$$\sum_{i=1}^d v_{ik} v_{il} = \delta_{kl} , \quad (1.95)$$

où  $\{\epsilon_k\}$  sont les valeurs propres. En pratique, la variation très lente du  $\chi^2$  dans certaines directions de l'espace des paramètres libres  $\hat{q}$  est problématique puisqu'elle détériore la qualité de l'ajustement. Toutefois, l'approche Hessienne conduit à un ensemble de 2d distribution de partons noté  $S_i^\pm$  pour chaque direction du vecteur propre. Par suite l'incertitude associée aux PDFs peut se propager dans la quantité  $\mathcal{F}$  de la façon équivalente à (1.92) :

$$(\Delta\mathcal{F})^2 = \frac{1}{2} \sum_{i=1}^d \left( \mathcal{F}(S_i^+) - \mathcal{F}(S_i^-) \right)^2 . \quad (1.96)$$

Il y a un arbitraire à choisir la valeur correcte de  $\Delta\chi^2$  (en principe  $\Delta\chi^2 = 1$ ), ce qui introduit des complications dans la procédure d'analyse globales. Toutefois, à cause du fait que les données provenant de différentes expériences sont parfois incompatibles, il est préférable de choisir  $\Delta\chi^2 > 1$  afin d'estimer les incertitudes d'une manière exacte. Enfin, l'introduction de ce critère de tolérance arbitraire ne permet pas de donner une interprétation statistique rigoureuse aux incertitudes qui résultent de cette méthode. La figure 1.6 montre une illustration de l'utilisation de la méthode Hessienne ( $\Delta\chi^2 = 1$ ) pour estimer grâce au code public HERAFitter l'incertitude sur les distributions de gluon (1.84) et de quark de valence  $u_{\text{val}}$  (1.85) à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

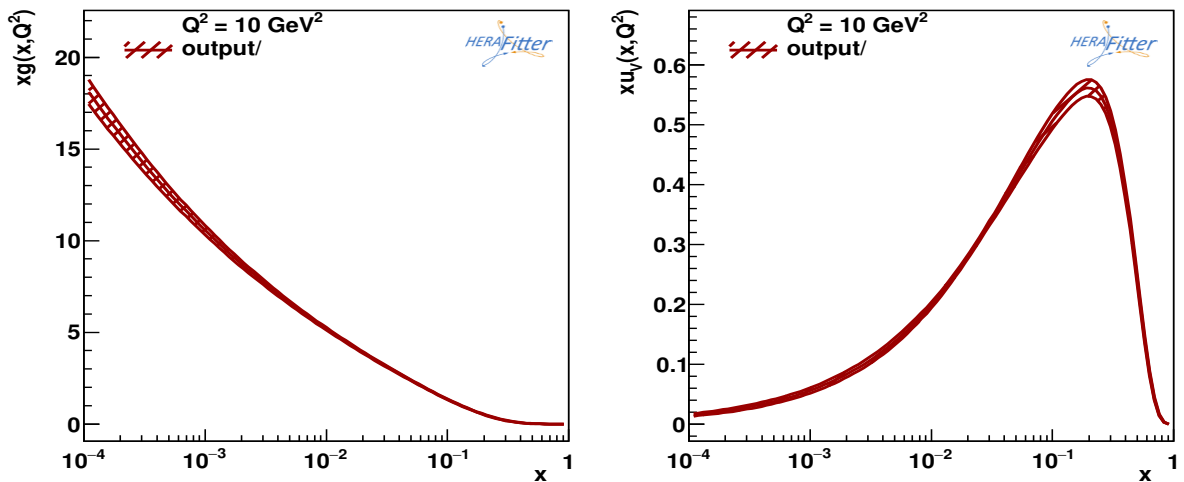


FIGURE 1.6 – Courbes montrant la distribution du gluon  $xg$  (gauche) et la distribution du quark de valence  $xu_{\text{val}}$  (droite) dans le même schéma (1.4.1).

## 1.6.2 L'approche de "Offset"

C'était la première méthode pour estimer les effets des incertitudes expérimentales sur les PDFs dans les analyses globales. Dans cette approche, les paramètres  $\xi_k$  de l'incertitude systématique dans l'équation (1.80) peuvent être fixés à zéro pour que les prédictions théoriques ajustées soient aussi proches que possible des valeurs centrales des données expérimentales publiées. Ces paramètres sont ensuite pris en compte dans l'analyse des erreurs qui s'effectue de la manière suivante : en plus de la matrice Hessienne représentée par l'élément de matrice (1.91), on définit une seconde matrice Hessienne  $M$  dont l'élément de matrice est défini par :

$$M_{ij} = \frac{\partial^2 \chi^2(\hat{q}, \hat{\xi})}{\partial q^{(i)} \partial \xi^{(j)}}, \quad (1.97)$$

on définit par suite la matrice de covariance des incertitudes systématiques :

$$C_{\text{sys}} = H^{-1} M M^T H^{-1}, \quad (1.98)$$

de sorte que la matrice de covariance totale soit

$$C_{\text{tot}} = C_{q\xi} + C_q, \quad (1.99)$$

avec  $C_q = H^{-1}$  étant la matrice de covariance de l'incertitude statistique (et systématique non corrélée). Enfin l'incertitude sur toute quantité  $\mathcal{F}$  dépendant des paramètres libres  $\hat{q}$  peut être estimée de la façon suivante :

$$(\Delta \mathcal{F})^2 = \sum_{i=1}^d \sum_{j=1}^d \frac{\partial \mathcal{F}}{\partial q^{(i)}} C_{ij}(\hat{q}) \frac{\partial \mathcal{F}}{\partial q^{(j)}}, \quad (1.100)$$

en substituant  $C$  par les matrices de covariance appropriées  $C_q$ ,  $C_{q\xi}$  et  $C_{\text{tot}}$  pour obtenir respectivement les bandes d'incertitude : statistique (et systématique non corrélée), systématique corrélée et expérimentale totale. Ce n'est pas une approche qui est statistiquement rigoureuse, mais sa vertu est qu'elle ne suppose pas que les incertitudes systématiques soient nécessairement distribuées suivant une Gaussienne. Cette méthode donne une très grande incertitude par rapport à l'approche Hessienne pour une même  $\Delta \chi^2 = 1$  (voir figure 1.7).



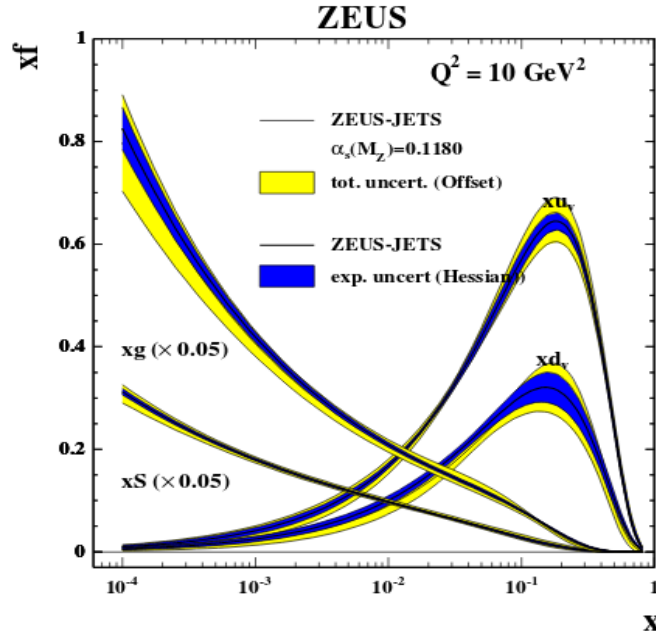


FIGURE 1.7 – Comparaison des incertitudes des distributions de partons  $xg, xu_{val}, xd_{val}$  et  $xS$  à l'échelle  $Q^2 = 10\text{GeV}^2$ ; extraites à partir des données de la collaboration ZEUS en utilisant les méthodes "Offset" et Hessienne. Figure tirée de COOPER-SARKAR et GWENLAN [2005].

### 1.6.3 L'approche des multiplicateurs de Lagrange

La méthode des multiplicateurs de Lagrange proposée dans STUMP et collab. [2001] surmonte certains des inconvénients des deux méthodes ci-dessus, en particulier celle relative à l'approximation de la linéarité. Cette méthode consiste à effectuer un ajustement global en contraignant la valeur d'une quantité  $\mathcal{F}$ , c'est-à-dire minimiser la fonction  $\Phi$  définie par :

$$\Phi(\hat{q}, \hat{\lambda}) = \chi^2(\hat{q}) + \hat{\lambda} \mathcal{F}(\hat{q}), \quad (1.101)$$

pour diverses valeurs fixées de  $\{\lambda_\alpha\}$ . Cette procédure génère une relation paramétrique en fonction de  $\lambda$  entre la fonction  $\chi^2$  et la quantité  $\mathcal{F}$ , de sorte que lorsque'on connaît  $\Delta\chi^2$ , il est facile de trouver la gamme autorisée pour la quantité  $\Delta\mathcal{F}$  sans demander que  $\Delta\chi^2$  soit distribuée suivant une Gaussienne. Cette procédure ne nécessite pas certaines approximations utilisées dans les approches de Hessien et de "offset", mais le critère du choix arbitraire d'un paramètre de tolérance subsiste.

Dans le souci de supprimer l'arbitraire qui entre en ligne de compte lors de l'estimation des incertitudes dans les analyses globales des PDFs, nous proposons une nouvelle approche basée sur les méthodes Monte Carlo par chaînes de Markov ("Markov Chain Monte Carlo", MCMC).

## Références

AARON, F. D. et AL. 2010, «Combined measurement and qcd analysis of the inclusive electron(positron) proton scattering cross sections at heras», *JHEP 01 (2010) 109*. 27

ABBOTT, B. et AL. 2001, «D0 collaboration», *Phys. Rev. Lett.*, vol. 86, p. 1707. 26

ABE, F. et AL. 1998, «Cdf collaboration», *Phys. Rev. Lett.*, vol. 81, p. 5744. 26

- ADAMS, M. R. et AL. 1996, «E665 collaboration», *Phys. Rev. D*, vol. 54, p. 3006. [25](#)
- ADLOFF, C. et AL. 2000, «H1 collaboration», *Eur. Phys. J. C*, vol. 13, p. 609. [25](#)
- ADLOFF, C. et AL. 2001a, «H1 collaboration», *Eur. Phys. J. C*, vol. 19, p. 269. [25](#)
- ADLOFF, C. et AL. 2001b, «H1 collaboration», *Eur. Phys. J. C*, vol. 21, p. 33. [25](#)
- ADLOFF, C. et AL. 2002, «H1 collaboration», *Phys. Lett. B*, vol. 528, p. 199. [26](#)
- AFFOLDER, T. et AL. 2001, «Cdf collaboration», *Phys. Rev. D*, vol. 64, p. 032 001. [26](#)
- ALEKHIN, S. et S. MOCH. 2011, «Heavy-quark deep-inelastic scattering with a running mass», *Phys. Lett. B*, vol. 699, p. 345–353. [11](#)
- ALTARELLI, G. 1982, «Partons in quantum chromodynamics», *Phys. Rept.*, vol. 81, n° 1, p. 1–129. [6](#)
- ALTARELLI, G. et G. PARISI. 1977, «Asymptotic freedom in parton language», *Nucl. Phys. B*, vol. 126, p. 298–318. [12](#), [20](#)
- ANDERSEN, J. R. et AL. «See sec. 22 in : The standard model and next-to-leading multileg working group report of the workshop physics at TeV colliders, les houches, france», *arXiv :1003.1241 [hep-ph]*. [22](#)
- ARNEODO, M. et AL. 1997a, «Nmc collaboration», *Nucl. Phys. B*, vol. 483, p. 3. [26](#)
- ARNEODO, M. et AL. 1997b, «Nmc collaboration», *Nucl. Phys. B*, vol. 487, p. 3. [26](#)
- ASHMORE, J. F. 1972, «A method of gauge invariant regularization», *Lett. Nuovo, Cim.*, vol. 4, p. 289–290. [9](#)
- AURENCHE, P., J.-P. GUILLET et E. PILON. 2016, «Quantum electrodynamics, quantum chromodynamics en pratique», *Troisième cycle. Beyrouth, Liban*. [xi](#), [9](#)
- BAIKOV, P. A., K. G. CHETYRKIN et J. H. KUHN. 2014, «Quark mass and field anomalous dimensions to order five in strong coupling constant», *JHEP 10 (2014)076*. [10](#)
- BALL, R. D. et AL. 2010, «A first unbiased global nlo determination of parton distributions and their uncertainties», *Nucl. Phys. B*, vol. 838, p. 136–206. [25](#)
- BARDEEN, W. A., A. BURAS, D. DUKE et T. MUTA. 1978, «Deep inelastic scattering beyond the leading order in asymptotically free gauge theories», *Phys. Rev. D*, vol. 18, p. 3998. [9](#)
- BAZARKO, A. O. et AL. 1995, «Ccf collaboration», *Z. Phys. C*, vol. 65, p. 189. [26](#)
- BENVENUTI, A. C. et AL. 1989a, «Bcdms collaboration», *Phys. Lett. B*, vol. 223, p. 485. [25](#)
- BENVENUTI, A. C. et AL. 1989b, «Bcdms collaboration», *Phys. Lett. B*, vol. 236, p. 592. [25](#)
- BJÖRKEN, J. D. 1969, *Phys. Rev.*, vol. 179, p. 1547. [18](#)
- BOLLINI, C. G. et J. J. GIAMBIAGI. 1972, «Dimensional renormalization : The number of dimensions as a regularizing parameter», *Nuovo, Cim. B*, vol. 12, p. 20–25. [9](#)
- BREITWEG, J. et AL. 2000, «Zeus collaboration», *Eur. Phys. J. C*, vol. 12, p. 35. [26](#)

- CALLAN, C. G. et D. J. GROSS. 1969, *Phys. Lett. B*, vol. 22, n° 156. [16](#)
- CHEKANOV, S. et AL. 2001, «Zeus collaboration», *Eur. Phys. J. C*, vol. 21, p. 443. [25](#)
- CICUTA, G. et E. MONTALDI. 1972, «Analytic renormalization via continuous space dimension», *Lett. Nuovo, Cim.*, vol. 4, p. 329–332. [9](#)
- COLLINS, J. 1984, «Renormalization», *Cambridge University Press*. [9](#)
- COLLINS, J. C. 1998, «Hard-scattering factorization with heavy quarks : A general treatment», *Phys. Rev. D*, vol. 58, p. 094 002. [11](#), [23](#)
- COLLINS, J. C. et D. E. SOPER. 1987, «The theorems of perturbative quantum chromodynamics», *Ann. Rev. Nucl. Part. Sci.* 37, p. 383–409. [11](#)
- COLLINS, J. C., D. E. SOPER et G. STERMAN. 1989, «Factorization of hard processes in quantum chromodynamics, in perturbative quantum chromodynamics», . [11](#)
- COOPER-SARKAR, A. et C. GWENLAN. 2005, «Comparison and combination of zeus and h1 pdf analyses hera - lhc workshop proceedings», *hep-ph/0508304v1*. [xi](#), [32](#)
- DEL-DEBBIO, L. et AL. 2007, «Nnpdf collaboration», *JHEP 0703 (2007) 039*. [25](#)
- XFITTER DEVELOPERS. 2017, «xfitter pdf fitting package», <http://www.xfitter.org/>. [xi](#), [24](#)
- DOKSHITZER, Y. L. 1977, «Calculation of the structure functions for deep inelastic scattering and  $e^+e^-$  annihilation by perturbation theory in quantum chromodynamics (in russian)», *Sov. Phys. JEETP*, vol. 46, p. 641–653. [12](#)
- FRITZSCH, H., M. GELL-MANN et H. LEUTWYLER. 1973, «Advantages of the color octet gluon picture», *Phys. Lett. B*, vol. 47, p. 365–368. [5](#)
- GBEDO, Y. G. et M. MANGIN-BRINET. 2017, «Markov chain monte carlo techniques applied to parton distribution functions determination : proof of concept», *Phys. Rev. D*, vol. 96, p. 014 015. [23](#)
- GONCHAROV, M. et AL. 2001, «Nutev collaboration», *arXiv :0102049[hep-exp]*. [26](#)
- GRIBOV, V. N. et L. N. LIPATOV. 1972, «Deep inelastic ep scattering in perturbation theory», *Sov. J. Nucl. Phys.*, vol. 15, p. 438–450. [12](#)
- GROSS, D. et F. WILCZEK. 1973a, «Asymptotically free gauge theories. 1», *Phys. Rev. D*, vol. 8, p. 3633–3652. [5](#), [8](#)
- GROSS, D. et F. WILCZEK. 1973b, «Ultraviolet behavior of nonabelian gauge theories», *Phys. Rev. Lett.*, vol. 30, n° 26, p. 1343–1346. [8](#)
- 'T HOOFT, G. 1973, «Dimensional regularization and the renormalization group», *Nucl. Phys. B*, vol. 61, p. 455–468. [9](#)
- 'T HOOFT, G. et M. J. G. VELTMAN. 1972, «Regularization and renormalization of gauge fields», *Nucl. Phys. B*, vol. 44, p. 189–213. [9](#)
- JAMES, F. et M. ROOS. 1975, «Minuit - a system for function minimization and analysis of the parameter errors and correlations», *Comput. Phys. Commun.* 10, p. 343–367. [26](#)

- JI, X. 2013, «Parton physics on a euclidean lattice», *Phys.Rev.Lett.*110.262002. [12](#)
- KAYIS-TOPAKSU, A., G. ÖNENGÜT, D. R. VAN et AL. 2008, «Leading order analysis of neutrino induced dimuon events in the chorus experiment», *Nucl. Phys. B*, vol. 798, p. 1–16. [25](#)
- MARTIN, A., W. STIRLING, R. THORNE et G. WATT. 2009, «Parton distributions for the lhc», *Eur. Phys. J. C*, vol. 63, p. 189–285. [23](#)
- MOCH, S., J. VERMASEREN et A. VOGT. 2004, «The three-loop splitting functions in quantum chromodynamics : The nonsinglet case», *Nucl. Phys. B*, vol. 688, p. 101–134. [12](#)
- MORENO, G. et AL. 1991, «E605 collaboration», *Phys. Rev. D*, vol. 43, p. 2815. [26](#)
- OLIVE, K. A. et AL. 2014, «Review of particle physics», *Chinese Physics C*, vol. 8, n° 9, p. 090 001. [xi](#), [19](#)
- OLNESS, F. et I. SCHIENBEIN. 2009, «Heavy quarks : Lessons learned from hera and tevatron», *Nucl. Phys. Proc. Suppl.*, vol. 191, p. 44–53. [22](#)
- POLITZER, H. D. 1973, «Reliable perturbative results for strong interactions?», *Phys. Rev. Lett.*, vol. 30, n° 26, p. 1346–1349. [8](#)
- POLITZER, H. D. 1974, «Asymptotic freedom : An approach to strong interactions», *Phys. Rept.*, vol. 14, p. 129–180. [8](#)
- PUMPLIN, J., D. STUMP, R. BROCK, D. CASEY, J. HUSTON, J. KALK, H. L. LAI et W. K. TUNG. 2001a, «Uncertainties of predictions from parton distribution functions II : The hessian method», *Phys. Rev. D*, vol. 65, p. 014 013. [29](#)
- PUMPLIN, J., D. R. STUMP et W. K. TUNG. 2001b, *Phys.Rev. D*, vol. 65, p. 014 011. [29](#)
- ROJO, J. C. 2006, «The neural network approach to parton distribution functions», *arXiv :0607122[hep/ph]*. [25](#)
- SELIGMAN, W. G. et AL. 1997, «Ccf collaboration», *Phys. Rev. Lett.*, vol. 79, p. 1213. [26](#)
- STUMP, D., J. PUMPLIN, R. BROCK, D. CASEY, J. HUSTON, J. KALK, H. LAI et W. K. TUNG. 2001, «Uncertainties of predictions from parton distribution functions I : The lagrange multiplier method», *Phys.Rev. D*, vol. 65, p. 014 012. [32](#)
- THORNE, R. et W. TUNG. «Perturbative quantum chromodynamics formulations with heavy quarks masses and global analysis», *arXiv :0809.0714v2 [hep-ph]*. [22](#)
- THORNE, R. S. 2006, «A variable-flavour number scheme for nnlo», *Phys. Rev. D*, vol. 73, p. 054 019. [23](#)
- THORNE, R. S. 2012, «Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections», *Phys. Rev. D*, vol. 86, p. 074 017. [23](#)
- THORNE, R. S. et R. G. ROBERTS. 1998, «An ordered analysis of heavy flavour production in deep inelastic scattering», *Phys. Rev. D*, vol. 57, p. 6871. [23](#)
- TOWELL, R. S. et AL. 2001, «E866 collaboration», *Phys. Rev. D*, vol. 64, p. 052 002. [26](#)

- VERMASEREN, J. A. M., A. VOGT, F. HERZOG, B. RUIJL et T. UEDA. 2017, «The five-loop beta function of yang-mills theory with fermions», *JHEP 02 (2017) 090*. [10](#)
- VOGT, A., S. MOCH et J. VERMASEREN. 2004, «The three-loop splitting functions in quantum chromodynamics : The singlet case», *Nucl. Phys. B*, vol. 691, p. 129–181. [12](#)
- WEINBERG, S. 1973, «Non abelian gauge theories of the strong interaction», *Phys. Rev. Lett.*, vol. 31, n° 7, p. 494–497. [5](#)
- WHITLOW, L. W. 1990, *preprint SLAC-357*, vol. 282. [26](#)
- WHITLOW, L. W. et AL. 1992, «Slac collaboration», *Phys. Lett. B*, vol. 282, p. 475. [26](#)
- YANG, U. K. et AL. 2001, «Ccfr collaboration», *Phys. Rev. Lett.*, vol. 86, p. 2742. [26](#)

## Chapitre 2

# La méthode Monte Carlo par chaînes de Markov

### 2.1 Introduction

Dans ce chapitre, on propose un outil qui combine les modèles théoriques et les données expérimentales : le Monte Carlo par chaînes de Markov. Cette formulation mathématique permet l'échantillonnage des densités de probabilité a posteriori pour des modèles multidimensionnels en utilisant des chaînes de Markov. Cette procédure a pour avantage l'extraction des densités de probabilité a posteriori jointes et individuelles des paramètres libres des PDFs et des PDFs elles-mêmes. Ainsi, une estimation rigoureuse des incertitudes expérimentales des PDFs peut se faire à partir des densités de probabilités des PDFs.

### 2.2 La formulation de la détermination des distributions de partons en terme d'inférence Bayésienne

Les distributions de partons (PDFs) sont souvent (à l'exception de la procédure du réseau de neurones [BALL et AL. \[2010\]](#); [DEL-DEBBIO et AL \[2007\]](#); [ROJO \[2006\]](#)) paramétrisées par des formes fonctionnelles à une échelle d'énergie initiale  $Q_0^2$  donnée. Ces distributions de partons peuvent être ensuite obtenues à une échelle d'énergie  $Q^2 > Q_0^2$  grâce aux équations d'évolution de DGLAP (ceci aussi aux exceptions près du modèle de dipôle et des distributions de partons non intégrées (uPDFs) par exemple, que nous n'allons pas considérer ici). Ces PDFs sont convoluées avec des sections efficaces partoniques pour obtenir des sections efficaces hadroniques (voir les équations (1.16) et (1.17)) pour divers processus différents et la fonction  $\chi^2$ , construite à partir de ces sections efficaces théoriques et données expérimentales correspondantes, est alors minimisée pour contraindre les paramètres des PDFs. Plutôt que d'utiliser la procédure standard de minimisation et la méthode Hessienne pour estimer les incertitudes des PDFs, nous proposons l'approche de l'inférence Bayésienne. Cette technique a été déjà testée avec succès dans plusieurs domaines [GILKS et collab. \[1995\]](#) mais n'a encore jamais été appliquée à la détermination des PDFs. Nous esquissons les principes essentiels dans ce qui suit. Le lecteur intéressé peut se référer par exemple à [NEAL \[2011\]](#) pour une connaissance plus approfondie du sujet.

## 2.2.1 L'inférence Bayésienne

Il y a deux interprétations essentielles de la probabilité qui sont utilisées en analyse de données : la fréquence relative utilisée entre autres pour affecter des erreurs statistiques aux mesures et la probabilité subjective utilisée pour quantifier les incertitudes systématiques [COWAN \[1997\]](#). Dans l'approche fréquentiste, la probabilité d'un événement élémentaire désigne la fréquence relative avec laquelle l'événement apparaît après avoir répété l'expérience une infinité de fois. Dans l'approche Bayésienne, la probabilité est associée à une hypothèse et est interprétée comme le degré de confiance accordée à celle-ci. La démarche logique permettant le calcul de la probabilité d'une hypothèse à l'aide du théorème de Bayes se nomme **inférence Bayésienne**.

Pour des raisons de compacité, nous notons  $\hat{q}$  le vecteur de paramètres des PDFs à déterminer :  $\hat{q} = (q^{(1)}, q^{(2)}, \dots, q^{(d)})^\top$  où  $d$  est typiquement, dans le cas d'une analyse totale, de l'ordre 25-30, et  $D$  représente les données observées. À partir de la perspective Bayésienne, les paramètres du modèle théorique et les observables sont tous deux considérés comme des quantités aléatoires, et l'objectif de l'inférence Bayésienne est formellement de déterminer la distribution de probabilité jointe  $P(D, \hat{q})$  de toutes les variables aléatoires. Cette distribution jointe peut être écrite comme

$$P(D, \hat{q}) = P(\hat{q})P(D|\hat{q}),$$

où  $P(\hat{q})$  est la **probabilité a priori** - quantifiant le degré de confiance que l'on a a priori avant l'observation des données expérimentales et  $P(D|\hat{q})$  appelée la **fonction de vraisemblance**  $\mathcal{L}(D, \hat{q}) \stackrel{\text{def}}{=} P(D|\hat{q})$  n'est rien d'autre que la probabilité, partant de la théorie, d'observer les données effectivement mesurées. Le théorème de Bayes est utilisé pour exprimer la distribution de  $\hat{q}$  sachant  $D$ ,  $P(\hat{q}|D)$ , en terme de la fonction de vraisemblance  $P(D|\hat{q})$  :

$$P(\hat{q}|D) = \frac{\mathcal{L}(D, \hat{q})P(\hat{q})}{\int \mathcal{L}(D, \hat{q})P(\hat{q}) d\hat{q}}. \quad (2.1)$$

Le dénominateur de l'équation (2.1) ne dépend pas des paramètres des PDFs et joue ainsi le rôle d'une constante de normalisation. Cette densité de **probabilité a posteriori**  $P(\hat{q}|D)$  quantifie la probabilité d'avoir les paramètres du modèle  $\hat{q}$  sachant que les données  $D$  ont été observées et est l'objet que nous examinons dans toute inférence Bayésienne. Pour déterminer cette probabilité conditionnelle, nous avons donc besoin de supposer une probabilité a priori pour les paramètres  $\hat{q}$ , et calculer la fonction de vraisemblance. Cette densité de probabilité  $P(\hat{q}|D)$  est alors échantillonnée par un algorithme Monte Carlo.

Supposons que les fluctuations des  $N_{\text{dat}}$  points de données expérimentales à l'étude autour de leurs valeurs théoriques correspondantes sont non corrélées et distribuées selon la loi normale (l'hypothèse qui peut être vérifiée a posteriori - voir chapitre 3), la méthode des moindres carrés et celle du maximum de vraisemblance sont équivalentes et le logarithme de la fonction de vraisemblance peut être écrit comme

$$\ln \mathcal{L}(D, \hat{q}) = -\frac{1}{2}\chi^2(\hat{q}) \quad (2.2)$$

## 2.3 Le principe du Monte Carlo par chaînes de Markov

### 2.3.1 La base de la méthode

La méthode Monte Carlo par chaînes de Markov permet l'estimation des densités de probabilité a posteriori pour des modèles multidimensionnels - qui, comme brièvement



expliqué dans la section précédente, est exactement ce que nous voulons et fournit des estimations fiables des erreurs. Les algorithmes MCMC nous permettent d'échantillonner à partir d'une distribution de probabilité connue à une constante multiplicative près. Ces algorithmes consistent à simuler séquentiellement une seule chaîne de Markov dont la distribution limite est celle choisie (dans notre cas, le maximum de la fonction de vraisemblance fois une densité de probabilité a priori des paramètres). Plus précisément, une chaîne de Markov est une séquence de variables aléatoires telle que la valeur suivante ou état de la séquence dépend uniquement de l'état présent et non des états passés (absence de « mémoire ») NEAL [1993]. Alors, nous générons une séquence de variables aléatoires,  $\hat{q}_0, \hat{q}_1, \dots$  telle que l'état suivant  $\hat{q}_{t+1}$  avec  $t \geq 0$  est distribué selon la **probabilité de transition**  $T(\hat{q}_t \rightarrow \hat{q}_{t+1}) \stackrel{\text{def}}{=} P(\hat{q}_{t+1}|\hat{q}_t)$ . Pour qu'elle converge vers une distribution stationnaire donnée, la chaîne doit être ergodique i.e. stationnaire, irréductible et non périodique.

Deux ingrédients sont nécessaires pour définir une chaîne de Markov : (i) les valeurs initiales des distributions marginales des paramètres et (ii) la probabilité de transition entre deux états de l'espace des paramètres :  $T(\hat{q} \rightarrow \hat{q}')$ , pour aller d'un état  $\hat{q}$  à l'autre état  $\hat{q}'$ . Il y a beaucoup d'aspects qui surviennent lors de l'implémentation des algorithmes MCMC : l'influence du point de départ de la chaîne (ce qui conduit à une phase de « **burn-in** »), le choix de la probabilité de transition, le taux de convergence, l'acceptance de l'algorithme, ... . Ces questions seront illustrées dans les sections suivantes.

### 2.3.2 L'algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings (voir figure 2.1), proposé par METROPOLIS et col-lab. [1953] et généralisé par HASTINGS [1970] pour engendrer une chaîne de Markov qui satisfasse le principe du bilan détaillé, est un des plus simples algorithmes Monte Carlo. Il est le cheval de bataille standard de calcul des méthodes MCMC tant pour sa simplicité et pour sa polyvalence, et est en principe applicable à tout système. Il est extrêmement facile à implémenter pour échantillonner une densité cible  $P(\hat{q}|D)$ . La chaîne de Markov s'itère en « sautant » de l'état actuel  $\hat{q}_t$  de l'espace des paramètres à l'état suivant  $\hat{q}_{t+1}$ . La probabilité de transition entre deux états est spécifiée par l'algorithme de Metropolis-Hastings en utilisant une fonction d'échantillonnage (angl. : Proposal distribution)  $\pi(\cdot|\hat{q}_t)$ . À chaque temps Monte Carlo  $t$ , correspondant à l'état  $\hat{q}_t$ , un état  $\hat{q}'_t$  est généré à partir de la fonction d'échantillonnage  $\pi(\cdot|\hat{q}_t)$  qui a les mêmes propriétés que la distribution cible. Cet état proposé est accepté comme nouvel état de la chaîne avec la probabilité

$$\alpha(\hat{q}_t, \hat{q}'_t) = \min\left(\frac{P(\hat{q}'_t|D)\pi(\hat{q}_t|\hat{q}'_t)}{P(\hat{q}_t|D)\pi(\hat{q}'_t|\hat{q}_t)}, 1\right), \quad (2.3)$$

autrement dit

$$\hat{q}_{t+1} = \begin{cases} \hat{q}'_t, & \text{avec la probabilité } \alpha(\hat{q}_t, \hat{q}'_t), \\ \hat{q}_t, & \text{avec la probabilité } 1 - \alpha(\hat{q}_t, \hat{q}'_t), \end{cases}$$

La probabilité de transition  $T(\hat{q}_t \rightarrow \hat{q}'_t)$  prend alors la forme explicite suivante

$$T(\hat{q}_t \rightarrow \hat{q}'_t) = \alpha(\hat{q}_t, \hat{q}'_t)\pi(\hat{q}'_t|\hat{q}_t). \quad (2.4)$$



---

```

Generate starting point  $\hat{q}_0$ 
for  $t = 0$  to  $N$  do
    Generate a candidate  $\hat{q}'_t \sim \pi(\cdot | \hat{q}_t)$ 
    Generate  $u \sim \mathcal{U}[0,1]$ 
    Compute  $r(\hat{q}_t, \hat{q}'_t) = \frac{P(\hat{q}'_t | D)\pi(\hat{q}_t | \hat{q}'_t)}{P(\hat{q}_t | D)\pi(\hat{q}'_t | \hat{q}_t)}$ 
    if  $u < \min(r(\hat{q}_t, \hat{q}'_t), 1)$  then
        |  $\hat{q}_{t+1} = \hat{q}'_t$ 
    else
        |  $\hat{q}_{t+1} = \hat{q}_t$ 
    end
end

```

---

FIGURE 2.1 – L’algorithme de Metropolis-Hastings.

Si le nouvel ensemble de paramètres  $\hat{q}'_t$  est accepté, alors il devient le nouvel état de la chaîne :  $\hat{q}_{t+1} = \hat{q}'_t$  (la chaîne « évolue »). S’il est rejeté, alors l’ancien état est réécrit dans la chaîne :  $\hat{q}_{t+1} = \hat{q}_t$  (la chaîne « stagne »). Ce critère assure l’ergodicité de la chaîne, i.e. que la chaîne échantillonne la distribution voulue  $P(\hat{q}|D)$  une fois à l’équilibre. Si notre distribution cible apparaît avec une constante de normalisation que nous ne connaissons pas, alors le calcul du rapport

$$r(\hat{q}_t, \hat{q}'_t) = \frac{P(\hat{q}'_t | D)\pi(\hat{q}_t | \hat{q}'_t)}{P(\hat{q}_t | D)\pi(\hat{q}'_t | \hat{q}_t)},$$

entraînera la simplification de ce facteur : c’est l’une des caractéristiques attrayantes de l’algorithme de Metropolis-Hastings, le faisant approprié pour une grande variété d’applications.

Un cas spécial de l’algorithme de Metropolis-Hastings est l’algorithme de Metropolis à marche aléatoire, pour lequel la distribution d’échantillonnage est choisie de sorte que  $\pi(\hat{q}'_t | \hat{q}_t) = \pi(|\hat{q}_t - \hat{q}'_t|)$ . La probabilité d’acceptation se réduit donc à

$$\alpha(\hat{q}_t, \hat{q}'_t) = \min\left(\frac{P(\hat{q}'_t | D)}{P(\hat{q}_t | D)}, 1\right). \quad (2.5)$$

Fréquemment, la proposition du saut de la marche aléatoire a une forme qui dépend d’un paramètre d’échelle. Par exemple, la fonction de proposition pour  $\hat{q}'_t$  pourrait être une distribution Gaussienne centrée en  $\hat{q}_t$  et d’écart-type  $\sigma$ . Une attention méticuleuse doit être portée au choix de ce paramètre d’échelle. Si ce paramètre d’échelle est très grand, un pourcentage très élevé des points de la chaîne sera rejeté, conduisant à un algorithme inefficace. S’il est très petit, la marche aléatoire explorera très lentement l’espace des paramètres, ce qui conduira encore à l’inefficacité de l’algorithme. Ce problème est d’autant plus difficile à gérer que le nombre de paramètres (i.e. la dimension du vecteur  $\hat{q}$ ) à échantillonner augmente.

Idéalement, pour optimiser l’efficacité du MCMC, la fonction de proposition devrait être aussi proche que possible de la distribution cible. En pratique, la performance de l’algorithme est évidemment fortement dépendante du choix de la distribution d’échantillonnage  $\pi(\cdot | \hat{q}_t)$  et plusieurs options sont souvent considérées dans la littérature pour explorer l’espace des paramètres : les distributions Gaussiennes unidimensionnelles ou

multidimensionnelles, l'échantillonnage par partition binaire de l'espace, .... Toutefois, même si l'ajustement des distributions d'échantillonnage améliore l'efficacité de l'algorithme de Metropolis-Hastings, ils ne sont pas assez efficaces pour faire face à plusieurs dizaines de paramètres pour un nombre raisonnable de machine en un temps utilisateur raisonnable. Pour contourner ces problèmes ; puisque dans le cas de l'extraction des PDFs, le nombre de paramètres libres à déterminer (c'est le nombre de paramètres dans la forme fonctionnelle des PDFs) est de l'ordre de  $\sim 25-30$  ; nous avons implémenté un algorithme plus efficace, basé sur la dynamique moléculaire, initialement développé pour la QCD sur réseau et est largement utilisé dans ce domaine.

## 2.4 L'algorithme Monte Carlo Hamiltonien (MCH)

Comme mentionné plus haut, le problème essentiel de l'algorithme de type Metropolis, repose sur le choix du point d'essai à chaque mouvement de la chaîne. Le choix d'un point d'essai loin du point initial entraînera une stagnation importante dans la distribution à échantillonner, et donc une probabilité d'acceptation faible, tandis que le choix d'un point d'essai proche du point initial n'entraînera pas une exploration efficace de l'espace des paramètres, et conduira donc à une convergence lente de la chaîne.

La dynamique Hamiltonienne [DUANE et collab. \[1987\]](#), développée à l'origine pour la théorie des champs sur réseau, est utilisée pour produire des candidats à l'algorithme de Metropolis, d'une manière très élégante et efficace. C'est un algorithme exact qui combine l'évolution de la dynamique moléculaire avec l'étape d'acceptation-rejet de l'algorithme de Metropolis. Ce dernier est utilisé pour corriger les erreurs de discrétisation dans l'intégration numérique des équations de mouvement correspondantes. Il existe de très bons articles qui détaillent les propriétés de cet algorithme (voir par exemple [NEAL \[2011\]](#)), et seules les idées principales seront rappelées ici sans être exhaustif. L'algorithme MCH (voir figure 2.4) consiste à associer à chaque ensemble de paramètres  $\hat{q}$  un ensemble d'impulsions conjuguées  $\hat{p}$  et remplacer la densité de probabilité a posteriori (2.1) nous voulons échantillonner par la distribution jointe définie par :

$$P(\hat{q}, \hat{p}) = \frac{1}{Z} e^{-H(\hat{q}, \hat{p})} = \frac{1}{Z} e^{-\mathcal{K}(\hat{p})} e^{\mathcal{U}(\hat{q})}, \quad (2.6)$$

où  $Z$  est une constante de normalisation et  $H(\hat{q}, \hat{p})$  est un Hamiltonien écrit comme

$$H(\hat{q}, \hat{p}) = \mathcal{K}(\hat{p}) + \mathcal{U}(\hat{q}). \quad (2.7)$$

Le premier terme a la forme de l'énergie cinétique

$$\mathcal{K}(\hat{p}) = \frac{1}{2} \hat{p}^T \mathbf{M}^{-1} \hat{p}, \quad (2.8)$$

où  $\mathbf{M}$  est une matrice de masse, généralement prise comme diagonale, d'éléments diagonaux <sup>1</sup>  $m^{(i)} = 1/(\Delta q^{(i)})^2$  et le second terme est une énergie potentielle arbitraire que nous définissons comme

$$\mathcal{U}(\hat{q}) = -\ln[\mathcal{L}(\mathbf{D}, \hat{q})P(\hat{q})], \quad (2.9)$$

dans laquelle, en analogie avec la physique,  $\hat{q}$  représente un vecteur-position.

À partir d'un point  $\hat{q}_0$  de la chaîne, la procédure de l'algorithme MCH consiste à sélectionner un ensemble de quantité de mouvement initial  $\hat{p}_0$  distribuée suivant la loi normale centrée réduite et à laisser le système évoluer de manière déterministe pendant un

1. Ici,  $\Delta \hat{q}$  représente une échelle appropriée pour  $\hat{q}$

certain temps selon les équations de mouvement de Hamilton pour  $H(\hat{q}, \hat{p})$ . Il atteint un point candidat  $(\hat{q}_1, \hat{p}_1)$  qui, selon la procédure de Metropolis décrite ci-dessus, est accepté avec une probabilité  $\min(1, e^{-\Delta H})$ . Puisque la dynamique conserve l'énergie, i.e.  $\Delta H = 0$  le long de la trajectoire, le taux d'acceptance est 100%, indépendamment de la dimension du vecteur-position  $\hat{q}$ .

En pratique, ce taux d'acceptance est dégradé en raison de la résolution numérique des équations de Hamilton, mais reste encore à un niveau très élevé (généralement de l'ordre de 70 à 90%, indépendamment de la dimension de la chaîne). Pour discrétiser les équations de Hamilton, nous utilisons la méthode leapfrog, une méthode pratique d'intégration de second ordre qui donne l'invariance par inversion de temps nécessaire à la probabilité de transition de l'algorithme de Metropolis. Cette méthode est composée de  $L$  pas associés à un incrément (longueur de chaque pas)  $\delta$  (voir l'algorithme de la figure 2.2). La figure 2.3 montre l'approximation par la méthode leapfrog de la dynamique Hamiltonienne à 1D définie par :

$$H(q, p) = q^2/2 + p^2/2. \quad (2.10)$$

---

```

Initialiser  $\hat{q}_0$  et  $\hat{p}_0$ 
Choisir  $\delta$  et  $L$ 
for  $j = 1$  to  $L$  do
     $p_{j-\frac{1}{2}}^{(i)} = p_{j-1}^{(i)} - \frac{\delta}{2} \frac{\partial \mathcal{H}}{\partial q^{(i)}}(\hat{q}_{j-1})$ 
     $q_j^{(i)} = q_{j-1}^{(i)} + \delta \frac{\partial \mathcal{H}}{\partial p^{(i)}}(\hat{p}_{j-\frac{1}{2}})$ 
     $p_j^{(i)} = p_{j-\frac{1}{2}}^{(i)} - \frac{\delta}{2} \frac{\partial \mathcal{H}}{\partial q^{(i)}}(\hat{q}_j)$ 
end

```

---

FIGURE 2.2 – Discrétisation des équations de Hamilton par la Méthode Leapfrog.

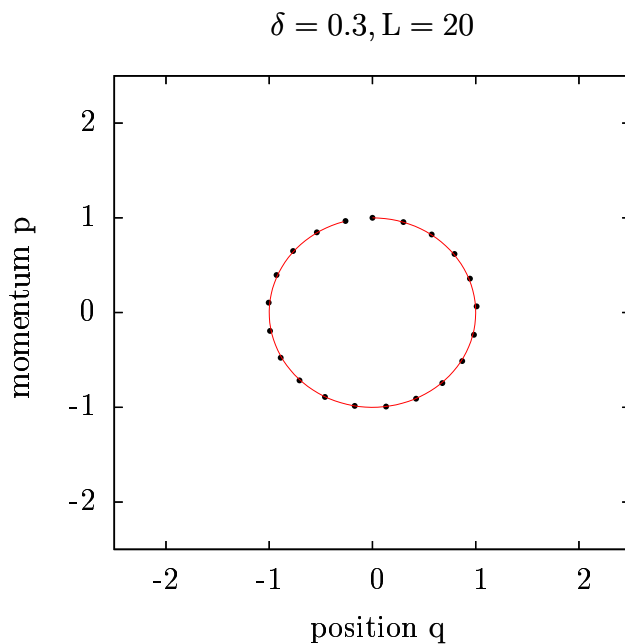


FIGURE 2.3 – L'approximation de la dynamique Hamiltonienne  $H(q, p) = q^2/2 + p^2/2$  par la méthode Leapfrog. L'état initial utilisé est  $q = 0, p = 1$ .

---

```

Initialiser  $\hat{q}_0$ 
Choisir  $\varepsilon^{(i)} = \delta \Delta q^{(i)}$  et  $L$ 
for  $t = 0$  to  $N$  do
  Générer  $\hat{p} \sim \mathcal{N}(0, \mathbb{I})$ 
  Poser  $x_0^{(i)} = q_t^{(i)}$  et  $p_0^{(i)} = p^{(i)}$ 
  Discrétisation des équations Hamiltoniennes
  for  $j = 1$  to  $L$  do
     $p_{j-\frac{1}{2}}^{(i)} = p_{j-1}^{(i)} - (\varepsilon^{(i)}/2) \frac{\partial \mathcal{H}}{\partial x^{(i)}}(\hat{x}_{j-1})$ 
     $x_j^{(i)} = x_{j-1}^{(i)} + \varepsilon^{(i)} p_{j-\frac{1}{2}}^{(i)}$ 
     $p_j^{(i)} = p_{j-\frac{1}{2}}^{(i)} - (\varepsilon^{(i)}/2) \frac{\partial \mathcal{H}}{\partial x^{(i)}}(\hat{x}_j)$ 
  end
  Acceptation-rejet du candidat  $\hat{x}'_t = \hat{x}_L$ 
  Poser  $\hat{p}'_t = \hat{p}_L$  et générer  $u \sim U_{[0,1]}$ 
  Taux d'acceptation et mise à jour :
   $\rho = \exp[H(\hat{x}_0, \hat{p}_0) - H(\hat{x}'_t, \hat{p}'_t)]$ 
  if  $u \leq \min(1, \rho)$  then
     $\hat{q}_{t+1} = \hat{q}'_t$ 
  else
     $\hat{q}_{t+1} = \hat{q}_t$ 
  end
end
return  $\{\hat{q}_t\}_{t=0}^N$ 

```

---

FIGURE 2.4 – Les étapes de l’algorithme Monte Carlo Hamiltonien

## 2.5 L’analyse des chaînes de Markov

L’estimation des erreurs statistiques sur des observables issues des simulations Monte Carlo est une tâche subtile et nécessite un traitement minutieux de la chaîne de Markov. La procédure d’analyse d’une chaîne de Markov se compose de plusieurs étapes. En particulier, il est nécessaire d’enlever la région de thermalisation, d’examiner correctement les corrélations entre les états voisins de la chaîne et in-fine de vérifier la convergence de la chaîne.

### 2.5.1 La thermalisation

Le temps de thermalisation de la chaîne souvent noté  $b$  correspond au nombre d’états ou de points  $\{\hat{q}_t\}_{t=1, \dots, b}$  à enlever en début afin que la chaîne ne soit plus sensible à son état de départ. Ce temps peut être estimé par rapport à l’indice  $b$  du premier état de la chaîne atteignant la valeur médiane de la densité de probabilité à posteriori  $P(\hat{q}_t|D)$  notée  $P_{1/2}$  tel que :  $P(\hat{q}_b|D) > P_{1/2}$ .

Pour illustrer la thermalisation, nous avons représenté sur la figure 2.5 l’historique Monte Carlo du paramètre  $B_g$  considéré dans la sous-section 1.5.1.2 pour trois chaînes indépendantes, chacune partant d’un point différent. Nous remplaçons la procédure standard par la procédure Monte Carlo basée sur l’algorithme de la figure 2.4 que nous avons

implémenté dans le code public HERAFitter<sup>2</sup>.

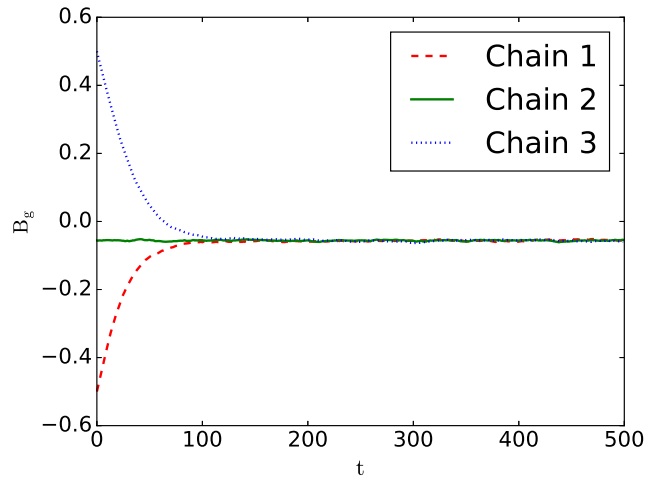


FIGURE 2.5 – Valeur du paramètre  $B_g$  en fonction du temps Monte Carlo  $t$  pour trois chaînes de Markov indépendantes. Les valeurs initiales de  $\frac{\chi^2}{\text{d.o.f}}$  pour chacune des trois chaînes de Markov 1, 2 et 3 correspondent respectivement à  $\frac{\chi^2}{\text{d.o.f}} = 67.44$ ,  $\frac{\chi^2}{\text{d.o.f}} = 0.87$  et  $\frac{\chi^2}{\text{d.o.f}} = 81.58$ . Nous identifions clairement sur ce graphe la région de thermalisation, qui est limitée aux  $\sim 100 - 210$  premières itérations.

Nous pouvons identifier sur la figure 2.5 la région de thermalisation, dont l’étendue dépend des points de départ. La chaîne représentée en trait plein de couleur verte a pour valeur de départ celle de l’équation(1.89) et est donc thermalisée très rapidement, spécialement après une seule itération, tandis que les autres chaînes, dont les points de départ sont loin de la valeur du  $\chi^2$  minimal, présentent une thermalisation d’environ 153 itérations pour la chaîne représentée par la couleur rouge et 209 itérations pour celle représentée par la couleur bleue ; comme prévu, plus le point de départ est éloigné de la valeur minimale du  $\chi^2$ , plus la thermalisation est longue.

## 2.5.2 Le traitement des corrélations

En construisant une chaîne de Markov, l’état  $\hat{q}_t$  dépend fortement de l’état  $\hat{q}_{t-1}$  et les quantités calculées à partir de cette chaîne forment elles-mêmes une chaîne de Markov avec des corrélations inhérentes d’un membre à l’autre. Ces types de corrélations sont souvent appelés « autocorrélations » dans le temps de simulation. Considérons que nous aimerions extraire un observable  $O$  d’une simulation de chaîne de Markov avec  $N$  points. Pour cette estimation, nous utilisons les  $N$  estimations successives Monte Carlo  $O_t$  (nous supposons dans ce qui suit que la région de burn-in a déjà été écartée, c’est-à-dire que la chaîne a atteint l’équilibre avant l’enregistrement des données) et nous calculons la moyenne usuelle  $\langle O \rangle$  où  $\langle . \rangle$  signifie la moyenne sur les  $N$  points de données. L’estimation habituelle de l’écart quadratique moyen de cette moyenne peut être calculée

$$\sigma_{\text{naïve}}^2 = (O_t - \langle O \rangle)^2. \quad (2.11)$$

Cette erreur “naïve” repose sur l’hypothèse que les mesures effectuées sur la chaîne de Markov ne sont pas corrélées, ce qui n’est généralement pas vrai. Afin de tenir compte des

2. Le package peut être téléchargé sur le site web [www.herafitter.org](http://www.herafitter.org)

corrélations, on peut introduire pour l'observable donnée  $O$ , le temps d'autocorrélation intégré  $\tau_{\text{int}}$ , qui peut être défini comme suit :

$$\tau_{\text{int}} = \frac{1}{2} \sum_{s=-\infty}^{+\infty} \rho(s), \quad (2.12)$$

où  $\rho(s)$  est la fonction d'autocorrélation normalisée définie par :

$$\rho(s) = \frac{(O_t - \langle O \rangle)(O_{t+s} - \langle O \rangle)}{(O_t - \langle O \rangle)^2}. \quad (2.13)$$

La dépendance de  $\rho(s)$  sur le temps de séparation  $s$  seulement est une conséquence de la chaîne d'être en équilibre. Le temps d'autocorrélation intégré contrôle l'erreur statistique dans la mesure Monte Carlo de  $O$  et il y a surtout deux possibilités d'incorporer ce temps d'autocorrélation dans l'évaluation des erreurs statistiques. Le premier consiste à écarter  $2\tau_{\text{int}}$  points entre deux points effectifs, ou en d'autres termes, à créer un sous échantillonnage en rejetant tous les états qui sont plus proches que  $2\tau_{\text{int}}$  l'un de l'autre, afin d'obtenir des états indépendants. Cette approche présente l'inconvénient d'exiger la connaissance *a priori* de  $\tau_{\text{int}}$ . La deuxième approche consiste à conserver toutes les mesures, mais en tenant compte du temps d'autocorrélation pour estimer les erreurs statistiques. L'erreur statistique des mesures corrélées peut en effet être calculée par [SOKAL \[1989\]](#); [WOLFF \[2004\]](#)

$$\sigma_{\tau}^2 = \frac{2\tau_{\text{int}}}{N} \sigma_{\text{naïve}}^2. \quad (2.14)$$

Cela signifie que le nombre de "points effectivement indépendants" pour réduire les erreurs statistiques dans une série Monte Carlo de longueur  $N$  est à peu près  $N/(2\tau_{\text{int}})$ . Si le temps d'autocorrélation intégré est utilisé pour évaluer les erreurs statistiques, cela signifie bien sûr qu'une estimation fiable de  $\tau_{\text{int}}$  et de son erreur sont nécessaires. De telles estimations requièrent une procédure délicate. Une méthode efficace appelée méthode  $\Gamma$  a été développée dans [WOLFF \[2004\]](#), celle-ci repose sur la détermination explicite des fonctions d'autocorrélation et des temps d'autocorrélation. Cette méthode fournit non seulement des estimations numériques du temps d'autocorrélation intégré mais aussi des estimations des valeurs moyennes et erreurs statistiques pour les fonctions arbitraires d'observables élémentaires dans les simulations Monte Carlo. Nous renvoyons le lecteur intéressé à cet article et références associées pour plus de détails. Nous avons utilisé la méthode  $\Gamma$  à la fois pour obtenir le temps d'autocorrélation intégré et pour calculer les observables.

Un autre procédé pour estimer de façon fiable l'erreur statistique sur des mesures non corrélées est appelé *méthode Jackknife* [QUENOUILLE \[1956\]](#). Il consiste à construire  $N$  blocs de données à partir de la population initiale de taille  $N$ , en enlevant une observation, laissant des échantillons de taille  $N - 1$ . La pré-moyenne sur les blocs de données fournit  $N$  estimations de la moyenne :

$$\langle O \rangle_B = \frac{1}{N-1} \left( \sum_{t=1}^{(B-1)} O_t + \sum_{t=B+1}^N O_t \right), \quad B = 1, \dots, N \quad (2.15)$$

La moyenne et la variance jackknife pour l'observable  $O$  sont alors construites à partir de :

$$\langle O \rangle_{\text{Jack.}} = \frac{1}{N} \sum_{B=1}^N \langle O \rangle_B, \quad \sigma_{\text{Jack.}}^2 = \frac{N-1}{N} \sum_{B=1}^N (\langle O \rangle_B - \langle O \rangle_{\text{Jack.}})^2 \quad (2.16)$$

Pour les contrôles croisés et comparaison, nous appliquerons dans notre analyse à la fois la méthode  $\Gamma$  et les techniques de *jackknife*. Pour ce dernier, afin de décorréler les points d'une chaîne donnée, nous effectuons un sous-échantillonnage de cette chaîne en utilisant la valeur du temps d'autocorrélation intégré fournie par la méthode  $\Gamma$ .

### 2.5.3 Convergence

Enfin, pour exclure le risque d'un manque de convergence non identifié, nous avons simulé plusieurs chaînes, avec différents points de départ (et aléatoires). Ceci est illustré sur la figure 2.6 pour le cas d'un seul paramètre variable, à savoir le paramètre  $B_g$  (tous les autres paramètres ont des valeurs fixées : voir la sous-section 1.5.1.2), où nous voyons clairement les chaînes convergées vers la même distribution stationnaire.

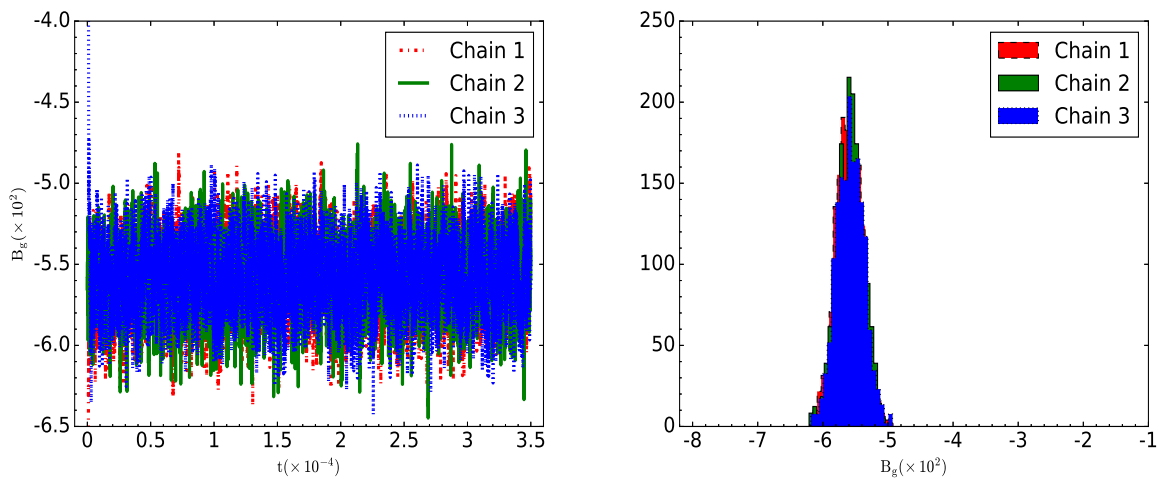


FIGURE 2.6 – Valeur du paramètre  $B_g$  comme fonction du temps Monte Carlo pour trois chaînes de Markov indépendantes (courbe de gauche). Nous voyons que les trois chaînes convergent clairement vers une même distribution stationnaire, ce qui est confirmée en traçant la distribution du paramètre  $B_g$  pour chaque chaîne (courbe de droite, en prenant soin d'écartier les points se trouvant dans la région de thermalisation et tenant en compte le temps d'autocorrélation intégré estimé à  $\tau_{\text{int}} = 42$ ).

Les résultats présentés dans le chapitre suivant ont été obtenus après avoir ignoré la thermalisation et bien pris en compte l'autocorrélation, en utilisant soit la méthode  $\Gamma$  ou la procédure de Jackknife comme expliquée dans la section (2.5).

## Références

- BALL, R. D. et AL. 2010, «A first unbiased global nlo determination of parton distributions and their uncertainties», *Nucl. Phys. B*, vol. 838, p. 136–206. [37](#)
- COWAN, G. 1997, «Statistical Data Analysis», *Oxford Science Publications, Oxford : Clarendon Press*. [38](#)
- DEL-DEBBIO, L. et AL. 2007, «Nnpdf collaboration», *JHEP 0703 (2007) 039*. [37](#)
- DUANE, S., A. D. KENNEDY, B. J. PENDLETON et D. ROWETH. 1987, «Hybrid monte carlo», *Phys. Lett. B*, vol. 195, p. 216–222. [41](#)

- GILKS, W. R., S. RICHARDSON et D. J. SPEIGELHALTER. 1995, «Markov chain monte carlo in practice», *Chapman and Hall/CRC, London*. 37
- HASTINGS, W. K. 1970, «Monte carlo sampling methods using markov chains and their applications», *Biometrika*, vol. 57, p. 97–109. 39
- METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER et E. TELLER. 1953, «Equations of state calculations by fast computing machines», *Chemical Physics*, vol. 21, p. 1087–1092. 39
- NEAL, R. M. 1993, «Probabilistic inference using markov chain monte carlo methods», *Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto*. 39
- NEAL, R. M. 2011, «Handbook of markov chain monte carlo», *S. Brooks, A. Gelman, G. L. Jones and X. Meng, Chapman and Hall/CRC Press*. 37, 41
- QUENOUILLE, M. H. 1956, «Notes on biais in estimation», *Oxford University Press, Biometrika*, vol. 43, p. 353–360. 45
- ROJO, J. C. 2006, «The neural network approach to parton distribution functions», *arXiv :0607122[hep/ph]*. 37
- SOKAL, A. D. 1989, «Monte carlo methods in statistical machanics : foundations and new algorithms», *Cours de Troisième Cycle de la Physique en Suisse Romande (Lausanne, Switzerland)*. 45
- WOLFF, U. 2004, «Monte carlo errors with less errors», *Comput. Phys. Commun.* 156, p. 143–153. 45





# Chapitre 3

## Implémentation de l’algorithme MCH dans le code public HERAFitter, validation et premiers résultats

### 3.1 Introduction

Plutôt que d’utiliser le programme standard MINUIT implémenté dans le code public HERAFitter pour minimiser la fonction  $\chi^2$ <sup>1</sup> afin d’extraire les paramètres libres  $\hat{q}$ , nous appliquons l’inférence Bayésienne à la fonction de vraisemblance (2.2), c’est-à-dire nous calculons la densité de probabilité des paramètres libres  $\hat{q}$  des PDFs, sur la base de données expérimentales sélectionnées. À cet effet, nous avons implémenté l’algorithme Monte Carlo Hamiltonien (voir figure 2.4) dans le code public HERAFitter et son successeur **XFITTER DEVELOPERS** [2017]. Ce programme fournit un module pour déterminer les PDFs par ajustement à un grand ensemble de données expérimentales. Une structure schématique de HERAFitter est illustrée sur la figure 3.1 qui résume toutes les fonctionnalités actuelles de la plateforme.

---

1. Nous avons considéré la même fonction  $\chi^2$  que celle utilisée dans l’analyse de la référence **AARON et AL.** [2010].

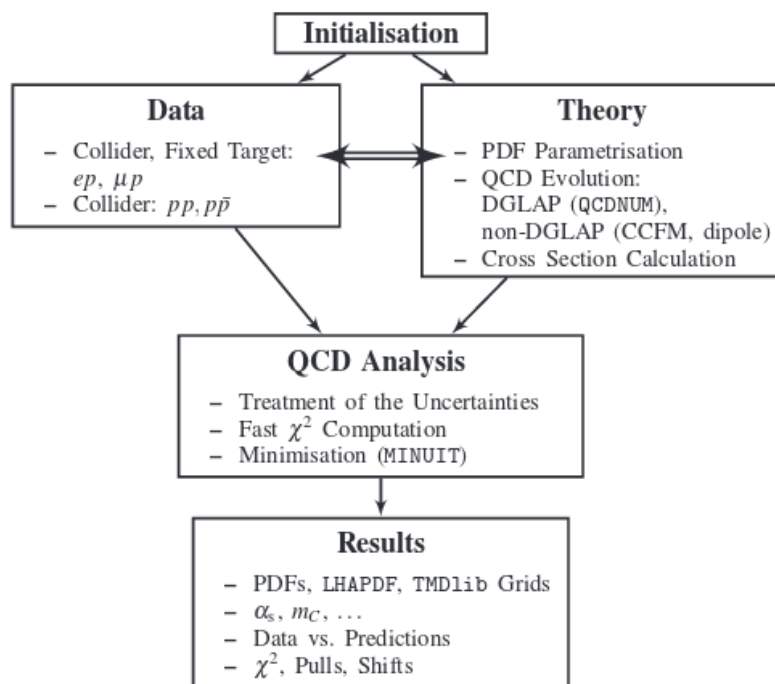


FIGURE 3.1 – Structure schématique du code HERAFitter. Figure tirée de [XFITTER DEVELOPERS \[2017\]](#)

Un des intérêts cruciaux de cette méthode est le fait que la valeur moyenne et l'incertitude de ces paramètres sont obtenues à partir des densités de probabilité calculées. L'algorithme MCH nécessite essentiellement l'ajustement de deux paramètres : le nombre de pas Leapfrog  $L$  et la largeur de pas Leapfrog  $\delta$ , cette dernière dépendant potentiellement de la direction dans l'espace des paramètres. Ces deux quantités sont choisies de manière à garder à la fois le taux d'acceptation élevé (ce qui impose une petite valeur de  $\delta$ , pour réduire les erreurs numériques lors de la discrétisation des équations de Hamilton par la méthode de Leapfrog (voir figure 2.3), et la corrélation entre deux itérations Monte Carlo petite (nécessitant donc une grande longueur  $L\delta$  pour la trajectoire de Leapfrog). La méthode utilisée pour trouver la largeur de pas Leapfrog optimale pour chaque paramètre est la suivante :

- 1- Exécutez une courte chaîne (par exemple 1000 itérations) avec des largeurs de pas Leapfrog initiales égales à la valeur de l'écart type de chaque paramètre fourni par la routine MINUIT
- 2- Calculez le taux d'acceptation
- 3- Ajustez la largeur de pas Leapfrog de chacun des paramètres en multipliant chacune d'elles par une valeur  $\delta$  afin qu'elles soient plus petites que celles utilisées dans la première étape de sorte que le nombre d'états acceptés soit compris entre 70 et 90%.

Une fois les largeurs de pas Leapfrog optimales choisies, une trajectoire de nombre de pas Leapfrog  $L=100$  pourrait être un point de départ adapté.

## 3.2 Les paramètres de la simulation et le “setup”

De part la nature de l’algorithme MCH, nous avons généré en parallèle 36 chaînes de Markov utilisant les formes fonctionnelles (1.84)-(1.88) pour les PDFs à l’échelle  $Q_0^2$  où les 10 paramètres  $B_g, C_g, B_{u_{\text{val}}}, C_{u_{\text{val}}}, E_{u_{\text{val}}}, C_{d_{\text{val}}}, C_{\bar{U}}, A_{\bar{D}}, B_{\bar{D}}$  et  $C_{\bar{D}}$  sont libres, chaque chaîne démarrant à partir d’un point aléatoire différent. Nous avons utilisé des densités de probabilité a priori uniformes<sup>2</sup> pour les paramètres libres  $\hat{q}$  et nous considérons les même ensembles de données que ceux utilisés pour produire les distributions de partons HERAPDF1.0 AARON et AL. [2010]. Pour accélérer le calcul de la fonction  $\chi^2$ , nous utilisons le schéma ZM-VFNS dans lequel les quarks lourds ( $m_H \gg \Lambda_{\text{QCD}}$ ) sont supposés sans masse, ce qui nous impose de considérer que les données satisfaisant à la condition ( $Q^2 \geq 10 \text{ GeV}^2$ ). Nous désignons donc les PDFs obtenues par la minimisation standard par “HERAPDF1.0 ZM-VFNS”. Nous explorons actuellement la possibilité d’utiliser la librairie APFEL pour calculer les processus de diffusion inélastique profonde dans une étude de détermination d’un ensemble complet de PDFs. Afin d’optimiser l’algorithme MCH, nous avons choisi  $L = 100$  et une largeur de “pas” Leapfrog différent pour chacun des 10 paramètres. Plus spécifiquement, nous avons choisi  $\epsilon^{(i)} = \delta \Delta q^{(i)}$ , où  $\delta = 3.10^{-2}$  et  $\Delta q^{(i)}$  est, pour chaque paramètre la valeur de son écart-type fourni par la minimisation standard : voir tableau 3.1. Ces paramètres MCH nous permettent d’obtenir pour chaque chaîne un taux d’acceptation de 80%. Nous avons calculé pour chacune de nos 36 chaînes le temps de thermalisation, et nous écartons le temps de thermalisation maximal (soit 28) de toutes les chaînes pour obtenir un total de 4400 points par chaîne. La méthode  $I$  a été utilisée pour déterminer à partir des  $4400 \times 36 = 158400$  points de la chaîne combinée, le temps d’autocorrélation intégré  $\tau_{\text{int}}^{(i)}$  pour chacun des 10 paramètres libres considérés dans cette étude (voir tableau 3.2). Le tableau 3.2 nous permet de conclure que chacune des 36 chaînes est presque décorrélée, puisque le temps d’autocorrélation intégré  $\tau_{\text{int}}$  pour tous les paramètres libres des PDFs est inférieur à 2.

L’algorithme MCH nécessite aussi le calcul de la dérivée de l’énergie potentielle (c’est dans notre cas la dérivée du  $\chi^2$  par rapport aux paramètres libres  $\hat{q}$  des PDFs). Ces dérivées sont calculées numériquement, en utilisant une dérivée symétrique. Nous avons donc besoin, pour 10 paramètres libres, 2040 évaluations du  $\chi^2$  pour chaque itération Monte Carlo. Chaque calcul de  $\chi^2$  prend environ 0.03 secondes, donc chaque itération Monte Carlo prend environ une (01) minute. Pour les 158400 points Monte Carlo, nous avons besoin de 110 jours, divisés par le nombre de CPUs nous avons utilisé (36), le nombre de jours se réduit à environ 3 jours. Il est important de noter qu’on a pas besoin pour des études réalistes de  $\sim 40000$  points Monte Carlo thermalisés et décorrélés, mais que  $\sim 4000$  points Monte Carlo suffisent à obtenir des résultats statistiquement précis (ceci nécessitera environ 2 heures lorsqu’on utilise les 36 CPUs) étant données les autres sources d’incertitudes (échelles de renormalisation/factorisation, schéma de quarks lourds, ...).

Enfin, nous avons vérifié que l’algorithme MCH satisfait la condition de réversibilité avec une très bonne précision (précision relative supérieure à  $10^{-6}$ ) et que l’acceptation moyenne calculée à l’aide de la méthode Jackknife sur la chaîne finale thermalisée et décorrélée est de  $\langle e^{-\Delta H} \rangle = 1.002 \pm 0.016$ , assurant ainsi que notre chaîne converge bien vers une distribution stationnaire.

---

2. Nous étudions actuellement l’effet d’une probabilité a priori Gaussienne de moyenne 0 et d’écart-type 1 sur le paramètre  $B_g$ .

Paramètre $q^{(i)}$	Écart-type $\Delta q^{(i)}$
$q^{(1)} = B_g$	0.029
$q^{(2)} = C_g$	0.51
$q^{(3)} = B_{u_{\text{val}}}$	0.039
$q^{(4)} = C_{u_{\text{val}}}$	0.13
$q^{(5)} = E_{u_{\text{val}}}$	2.6
$q^{(6)} = C_{d_{\text{val}}}$	0.28
$q^{(7)} = C_{\bar{U}}$	0.98
$q^{(8)} = A_{\bar{D}}$	0.049
$q^{(9)} = B_{\bar{D}}$	0.031
$q^{(10)} = C_{\bar{D}}$	1.3

TABLEAU 3.1 – Les valeurs des écarts-types obtenues par la minimisation standard de la fonction  $\chi^2$ .

Paramètre $q^{(i)}$	Temps d'autocorrélation intégré $\tau_{\text{int}}^{(i)}$
$q^{(1)} = B_g$	1.564±0.025
$q^{(2)} = C_g$	0.6788±0.0067
$q^{(3)} = B_{u_{\text{val}}}$	0.4924±0.0025
$q^{(4)} = C_{u_{\text{val}}}$	0.4164±0.0022
$q^{(5)} = E_{u_{\text{val}}}$	0.7706±0.0084
$q^{(6)} = C_{d_{\text{val}}}$	1.500±0.023
$q^{(7)} = C_{\bar{U}}$	1.2235±0.0038
$q^{(8)} = A_{\bar{D}}$	1.439±0.022
$q^{(9)} = B_{\bar{D}}$	1.481±0.022
$q^{(10)} = C_{\bar{D}}$	0.992±0.013

TABLEAU 3.2 – Les valeurs des temps d'autocorrélation intégrés pour les 10 paramètres libres des PDFs (1.84)-(1.88).

### 3.3 Les valeurs et les densités de probabilité marginales des paramètres libres des PDFs

Le tableau 3.3 compare la valeur moyenne et son erreur statistique pour chacun des 10 paramètres libres des PDFs en considération, en utilisant les deux procédures d'analyse que nous avons présentées dans la sous-section 2.5.2 du chapitre 2, à savoir la méthode  $\Gamma$  et la méthode de Jackknife. Le calcul par la méthode  $\Gamma$  du temps d'autocorrélation intégré de chaque paramètre libre  $\tau_{\text{int}}^{(i)}$  (voir tableau 3.2) donne des valeurs inférieures à 2, ce qui réduit le nombre de points Monte Carlo à 39600 pour l'analyse Jackknife. Comme on peut le voir dans le tableau 3.3, les deux méthodes donnent des résultats très proches, montrant que nous avons analysé notre chaîne de Markov de manière cohérente.

Le tableau 3.4 montre les résultats de l'estimation de la moyenne, du "best fit" et de l'écart-type des 10 paramètres libres considérés par la méthode MCMC et la procédure

Paramètre	Méthode $\Gamma$	Méthode Jackknife
$q^{(1)} = B_g$	$-0.0537 \pm 0.0001$	$-0.0537 \pm 0.0002$
$q^{(2)} = C_g$	$5.9449 \pm 0.0015$	$5.9483 \pm 0.0025$
$q^{(3)} = B_{u_{val}}$	$0.6124 \pm 0.0001$	$0.6125 \pm 0.0002$
$q^{(4)} = C_{u_{val}}$	$4.7458 \pm 0.0003$	$4.7455 \pm 0.0006$
$q^{(5)} = E_{u_{val}}$	$14.965 \pm 0.008$	$14.961 \pm 0.012$
$q^{(6)} = C_{d_{val}}$	$3.2054 \pm 0.0014$	$3.2077 \pm 0.0016$
$q^{(7)} = C_{\bar{U}}$	$4.0917 \pm 0.0038$	$4.0961 \pm 0.0048$
$q^{(8)} = A_{\bar{D}}$	$0.3096 \pm 0.0002$	$0.3098 \pm 0.0002$
$q^{(9)} = B_{\bar{D}}$	$-0.0174 \pm 0.0001$	$-0.0173 \pm 0.0002$
$q^{(10)} = C_{\bar{D}}$	$6.2203 \pm 0.0054$	$6.2096 \pm 0.0076$
Nombre de points Monte Carlo	158400	39600

TABLEAU 3.3 – Comparaison des valeurs moyennes et de leurs erreurs statistiques pour les 10 paramètres libres des PDFs obtenues en utilisant deux méthodes d’analyse différentes.

standard de minimisation avec MINUIT. Nous constatons, comme nous l’avons déjà mentionné plus haut, que la détermination des paramètres par les méthodes Monte Carlo donne beaucoup plus d’information que la méthode de minimisation standard. Nous pouvons en particulier extraire les erreurs statistiques sur les quantités qui nous intéressent ; cette erreur décroît avec la longueur de la chaîne de Markov. Comme le montre le tableau 3.4, les valeurs des “best fit”<sup>3</sup> extraites à partir des densités de probabilité individuelles des 10 paramètres libres considérés ; sont compatibles avec les valeurs des paramètres fournies par la procédure de minimisation. Pour ce qui concerne les écarts-types, bien que la méthode MCMC et la minimisation donnent des résultats similaires, aucune comparaison précise ne peut être faite car la minimisation ne fournit pas une estimation des erreurs statistiques sur les écarts-types des paramètres. De plus, les deux quantités (l’écart-type calculé à partir de la méthode MCMC et l’écart-type calculé à partir de la procédure standard de minimisation) ne devrait coïncider que si la densité de probabilité du paramètre considéré est Gaussienne, ce qui n’est pas nécessairement vrai (voir ci-dessous). Ce fait est déjà visible dans le tableau 3.4, où nous pouvons voir que les valeurs moyennes et les valeurs du “best fit” ne coïncident pas dans les erreurs pour certains paramètres ( par exemple  $B_g$ ,  $C_{\bar{D}}$ ,...). La figure 3.2 détaille les densités de probabilité a posteriori marginales à 1D et 2D des paramètres libres des PDFs considérées dans cette analyse. La diagonale de la figure 3.2 montre la densité de probabilité a posteriori marginalisée à une dimension de chacun des paramètres  $B_g$ ,  $C_g$ ,  $B_{u_{val}}$ ,  $C_{u_{val}}$ ,  $E_{u_{val}}$ ,  $C_{d_{val}}$ ,  $C_{\bar{U}}$ ,  $A_{\bar{D}}$ ,  $B_{\bar{D}}$  et  $C_{\bar{D}}$ . Le nombre de points de l’histogramme correspond au nombre d’échantillons indépendants  $N_{ind} = 39600$  de la chaîne finale. Les courbes ne se trouvant pas sur la diagonale de cette même figure montrent les corrélations à deux dimensions entre ces paramètres. Les contours intérieur et extérieur de ces corrélations délimitent les régions contenant respectivement 68% et 95% de la densité de probabilité.

3. C’est la valeur de la chaîne pour laquelle la fonction  $\chi^2$  est minimale.

Paramètre	Valeur	Méthode MCMC	MINUIT
$B_g$	Moyenne	$-0.0537 \pm 0.0002$	-0.0559 0.0288
	“Best fit”	$-0.0632 \pm 0.0168$	
	Écart-type	$0.0299 \pm 0.0001$	
$C_g$	Moyenne	$5.9483 \pm 0.0025$	5.9274 0.5078
	“Best fit”	$5.8952 \pm 0.0615$	
	Écart-type	$0.5037 \pm 0.0019$	
$B_{u_{\text{val}}}$	Moyenne	$0.6125 \pm 0.0002$	0.6098 0.0389
	“Best fit”	$0.6092 \pm 0.0121$	
	Écart-type	$0.0371 \pm 0.0001$	
$C_{u_{\text{val}}}$	Moyenne	$4.7455 \pm 0.0006$	4.7122 0.1332
	“Best fit”	$4.7467 \pm 0.0525$	
	Écart-type	$0.1280 \pm 0.0005$	
$E_{u_{\text{val}}}$	Moyenne	$14.961 \pm 0.012$	14.76 2.571
	“Best fit”	$15.42 \pm 0.94$	
	Écart-type	$2.494 \pm 0.010$	
$C_{d_{\text{val}}}$	Moyenne	$3.2077 \pm 0.0016$	3.143 0.2830
	“Best fit”	$3.084 \pm 0.076$	
	Écart-type	$0.3183 \pm 0.0016$	
$C_{\bar{U}}$	Moyenne	$4.0961 \pm 0.0048$	4.052 0.9782
	“Best fit”	$4.323 \pm 0.814$	
	Écart-type	$0.9617 \pm 0.0035$	
$A_{\bar{D}}$	Moyenne	$0.3098 \pm 0.0002$	0.305 0.0488
	“Best fit”	$0.320 \pm 0.039$	
	Écart-type	$0.0485 \pm 0.0002$	
$B_{\bar{D}}$	Moyenne	$-0.0173 \pm 0.0002$	-0.0178 0.0306
	“Best fit”	$-0.0097 \pm 0.0237$	
	Écart-type	$0.0304 \pm 0.0001$	
$C_{\bar{D}}$	Moyenne	$6.2096 \pm 0.0076$	5.875 1.290
	“Best fit”	$5.888 \pm 0.142$	
	Écart-type	$1.505 \pm 0.009$	

TABLEAU 3.4 – Comparaison des résultats obtenus pour les valeurs des paramètres libres des PDFs extraites à partir d’une chaîne de Markov indépendante de longueur 39 600, et les résultats fournis par MINUIT. Nous calculons à partir du MCMC la valeur du “best fit”, la valeur moyenne et la valeur de l’écart-type pour chacun des des 10 paramètres libres considérés, ainsi que leurs erreurs statistiques estimées par la méthode Jackknife. L’écart-type donné par MINUIT est l’écart-type à  $1\sigma$ .

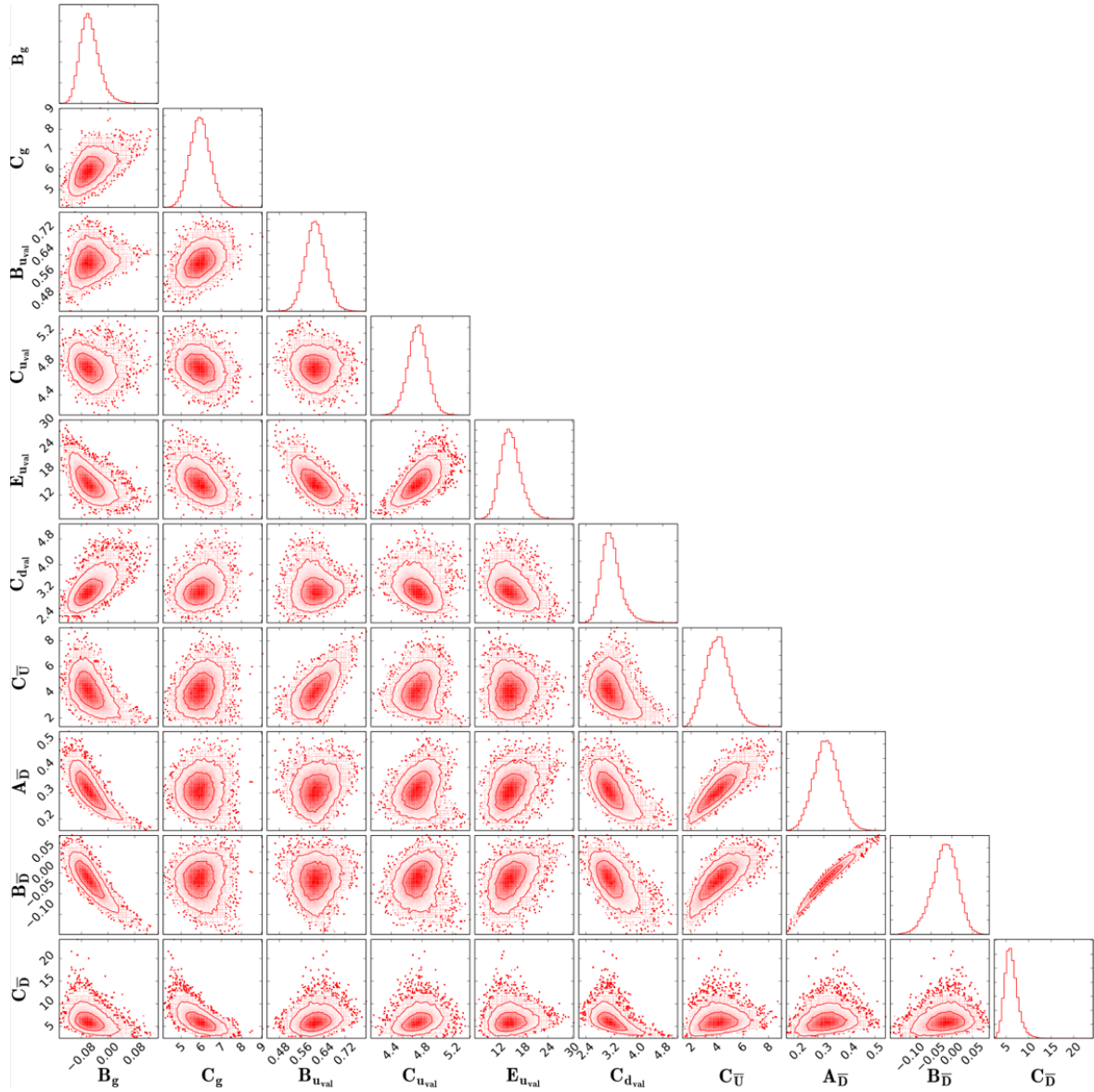


FIGURE 3.2 – Les densités de probabilité a posteriori marginales des 10 paramètres libres des PDFs (diagonal) et les courbes de corrélations à deux dimensions entre ces paramètres (off-diagonal).



Nous notons de plus que la densité de probabilité marginale de certains paramètres ne peut pas être décrite par une loi Gaussienne, comme illustré sur la figure 3.3, et est parfois asymétrique.

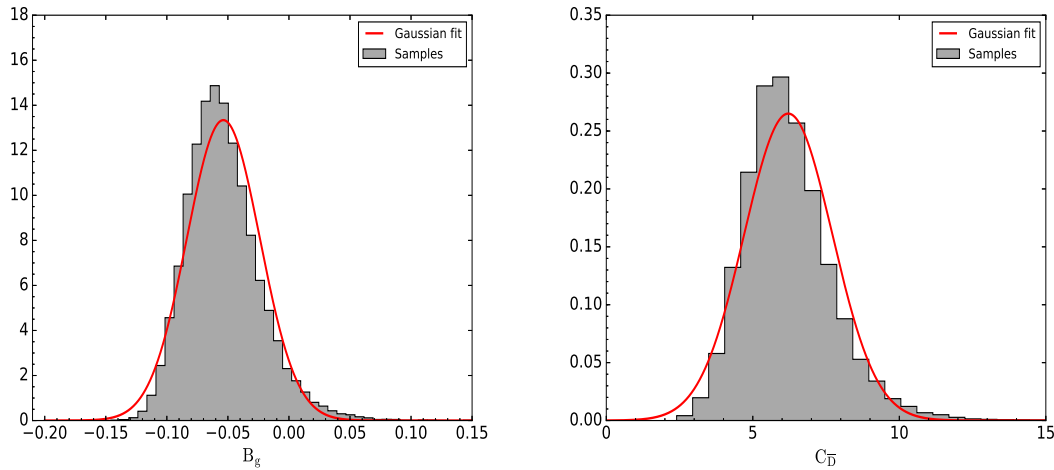


FIGURE 3.3 – La densité de probabilité marginalisée des paramètres  $B_g$  et  $C_{\bar{D}}$ . Ces paramètres ne suivent pas la loi Gaussienne, comme nous pouvons le voir à partir de l’ajustement Gaussien (courbe rouge).

La figure 3.4 montre la distribution de  $\chi^2/\text{d.o.f}$  pour notre chaîne de Markov à 10 dimensions. La courbe rouge est un ajustement avec une distribution de  $\chi^2$  possédant 10 degrés de liberté, qui décrit parfaitement nos résultats. Le fait que la fonction  $\chi^2$  échantillonnée suit la loi d’une distribution de  $\chi^2$  avec le nombre de degré de liberté attendu est une forte indication (bien que ce ne soit pas une preuve formelle) que nos hypothèses concernant les fluctuations des données expérimentales autour de leurs valeurs théoriques correspondantes, sont justifiées. Les pointillés de couleur bleue et verte correspondent respectivement aux intervalles de confiance à 68% et 95%. Bien que ceci n’est pas le cas dans l’étude nous présentons, nous aimerions également noter que les potentielles directions plates dans l’espace des paramètres sont moins problématiques pour la méthode MCMC que pour les techniques de minimisation standard.

Nous allons maintenant calculer à partir des chaînes de Markov des paramètres libres des PDFs, les distributions de partons qui nous intéressent. La procédure nous avons utilisée est expliquée dans la section suivante et est plus généralement valable pour toute observable nous voulons calculer à partir du MCMC.

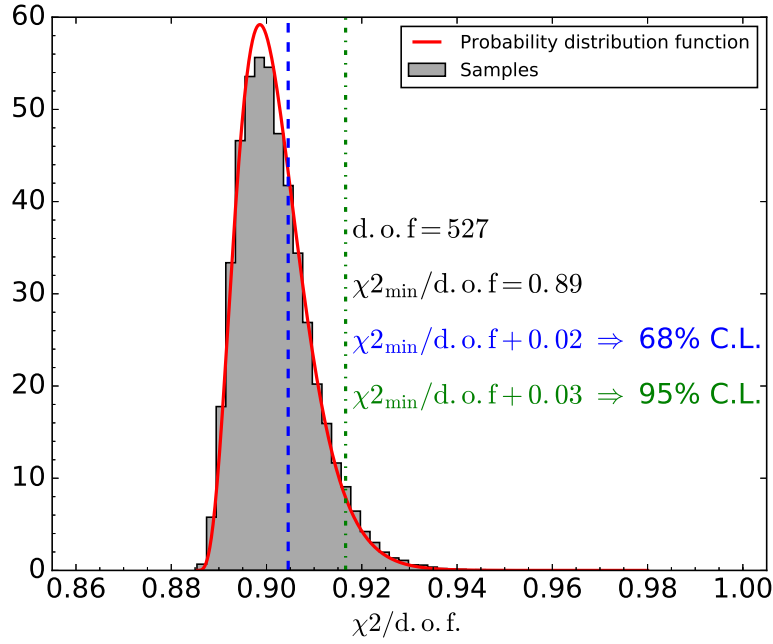


FIGURE 3.4 – La distribution de  $\chi^2$  pour une chaîne Monte Carlo à 10 dimensions. La courbe rouge est un ajustement de ces échantillons Monte Carlo avec la loi d’une distribution de  $\chi^2$  possédant 10 degrés de liberté. Une description plus détaillée est donnée dans le texte.

### 3.4 Les densités de probabilité marginales des PDFs : introduction aux limites de confiance des PDFs

Pour extraire les distributions de partons à partir des chaînes de Markov, nous calculons à partir des valeurs des 10 paramètres libres  $\{q^{(i)}\}_{i=1}^{10}$  obtenues à chaque itération Monte Carlo, les PDFs correspondantes pour une gamme de valeurs de  $x$  et  $Q^2$ . Ceci fournit les densités de probabilité marginales des PDFs à des valeurs fixées de  $x$  et  $Q^2$ . Une illustration pour les distributions de gluon  $xg$  et de quark de la mer  $xS = x\bar{U} + x\bar{D}$  est donnée respectivement sur la figure 3.5 et la figure 3.6 pour les valeurs  $x \approx 10^{-4}$  et  $x \approx 0.83$  à l’échelle d’énergie  $Q^2 = 10 \text{ GeV}^2$ .

Les limites de confiance à 68% correspondantes aux distributions de partons  $xg(x, Q^2)$  et  $xS(x, Q^2)$  à l’échelle  $Q^2 = 10 \text{ GeV}^2$  pour les valeurs de  $x \approx 10^{-4}$  et  $x \approx 0.83$  sont obtenues à partir des densités de probabilité des figures (3.5) et (3.6), en considérant les régions de chacune de ces densités de probabilité contenant 68% des points Monte Carlo se trouvant de chaque côté de la valeur du “best fit” de chacune de ces distributions de partons.

Valeur de $Q^2$	Valeur de $x$	$xg(x, Q^2)$	$xS(x, Q^2)$
10 $\text{GeV}^2$	$10^{-4}$	$18.3^{+0.7}_{-1.0}$	$2.69^{+0.02}_{-0.02}$
	0.83	$0.000038^{+0.000015}_{-0.000008}$	$0.000051^{+0.000249}_{-0.000036}$

TABLEAU 3.5 – Les valeurs du “best fit” et leurs limites de confiance à 68% correspondantes aux distributions de partons  $xg(x, Q^2)$  et  $xS(x, Q^2)$  à l’échelle  $Q^2 = 10 \text{ GeV}^2$  pour les valeurs de  $x \approx 10^{-4}$  et  $x \approx 0.83$ .

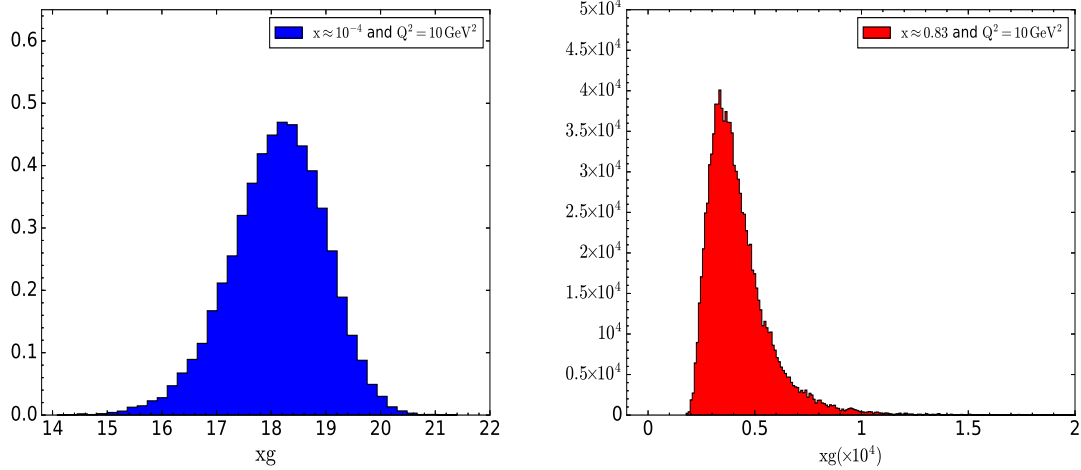


FIGURE 3.5 – La densité de probabilité de la distribution du gluon  $xg(x, Q^2)$  pour  $x \approx 10^{-4}$  (courbe de gauche) et pour  $x \approx 0.83$  (courbe de droite) à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

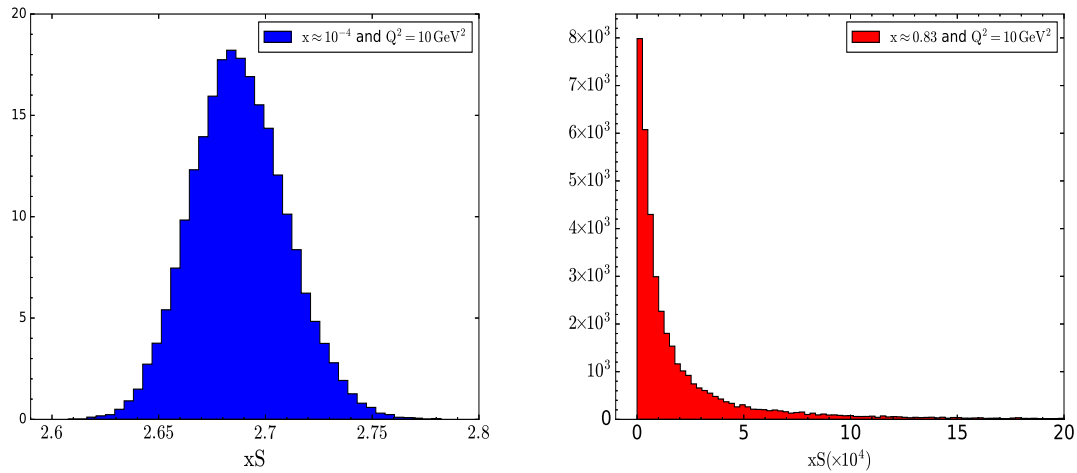


FIGURE 3.6 – La densité de probabilité de la distribution du quark de la mer  $xS(x, Q^2)$  pour  $x \approx 10^{-4}$  (courbe de gauche) et pour  $x \approx 0.83$  (courbe de droite) à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

Les figures 3.5 et 3.6 nous permettent aussi de conclure que les densités de probabilité des distributions de partons sont presque symétriques à bas  $x$  ( $x \approx 10^{-4}$ ) et très piquées (antisymétriques) à haut  $x$  ( $x \approx 0.83$ ) : ceci fournit des limites de confiance non nécessairement symétriques nous avons montrés dans le tableau 3.5.

### 3.5 Les distributions de partons et leurs incertitudes expérimentales

Pour chaque couple de valeurs  $(x, Q^2)$ , nous déterminons l'intervalle de confiance à 68% autour de la valeur du "best fit" de la distribution de partons en considérant la région de la densité de probabilité se trouvant de chaque côté de la valeur du "best fit" et en prenant 68% des données Monte Carlo de chacune de ces régions.

Les figures 3.7 et 3.8 montrent les distributions de partons  $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ,  $xS$  et  $xg$  obtenues de cette manière comme une fonction de la variable de Björken  $x$  à l'échelle  $Q^2 =$

$10 \text{ GeV}^2$ , où  $xS = x\bar{U} + x\bar{D}$  est la distribution du quark de la mer. Notons que pour  $Q^2 > m_c^2$ ,  $x\bar{U} = x\bar{u} + x\bar{c}$  et pour  $Q^2 > m_b^2$ ,  $x\bar{D} = x\bar{d} + x\bar{s} + x\bar{b}$ ; de sorte que les quarks lourds sont inclus dans la distribution du quark de la mer. Les enveloppes de couleur (bleu et verte) correspondent respectivement aux incertitudes expérimentales obtenues par les méthodes MCMC (niveau de confiance à 68%) et Hésienne ( $\Delta\chi^2 = 1$ ). Les distributions de partons centrales de couleur rouge correspondent aux distributions de partons qui ajustent la mieux les données de HERA 1 considérées. Ces PDFs MCMC sont comparées aux PDFs HERAPDF1.0 pour le schéma ZM-VFNS, avec une comparaison directe dans les figures 3.7 et 3.8, et une comparaison du rapport de leurs valeurs centrales dans la figure 3.9. Elles sont comme prévu très proches à la fois en valeur centrale et en intervalle de confiance. L'estimateur du maximum de vraisemblance et la méthode des moindres carrés sont en effet équivalents sous l'hypothèse Gaussienne, ce qui dans notre cas peut être appliquée, comme nous l'avons mentionné dans le chapitre 2.

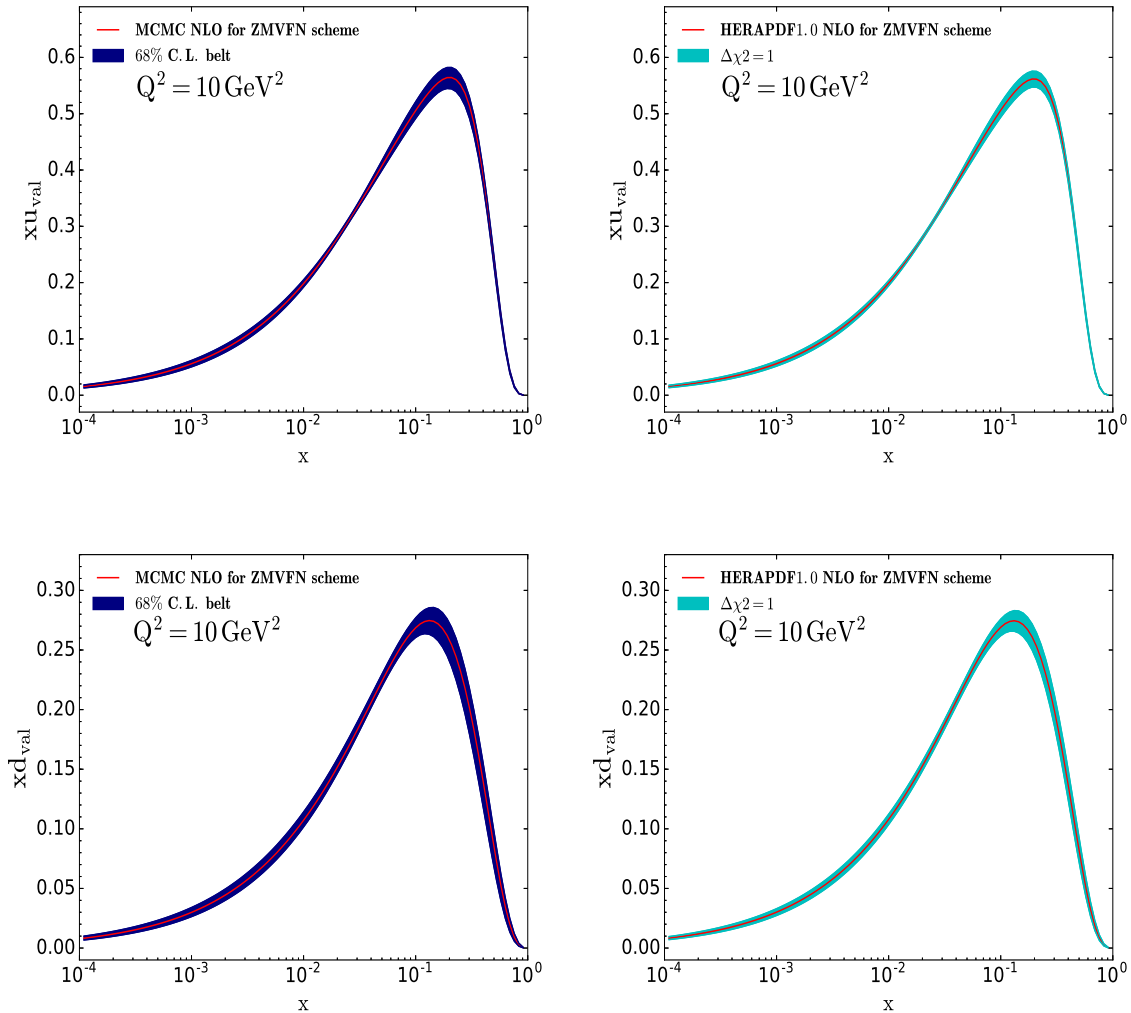


FIGURE 3.7 – Les distributions de partons  $xu_{\text{val}}$  et  $xd_{\text{val}}$  à l'échelle  $Q^2 = 10 \text{ GeV}^2$  obtenues à partir de la chaîne Monte Carlo en utilisant le code d'évolution QCDNUM [BOTJE \[2011\]](#) (courbes de gauche) sont comparées à celles de HERAPDF1.0 (ZM-VFNS) obtenues à partir du code public HERAFitter (courbes de droite). Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur centrale (courbe rouge) pour les PDFs MCMC et la déviation  $\Delta\chi^2 = 1$  standard pour les PDFs HERAPDF1.0.

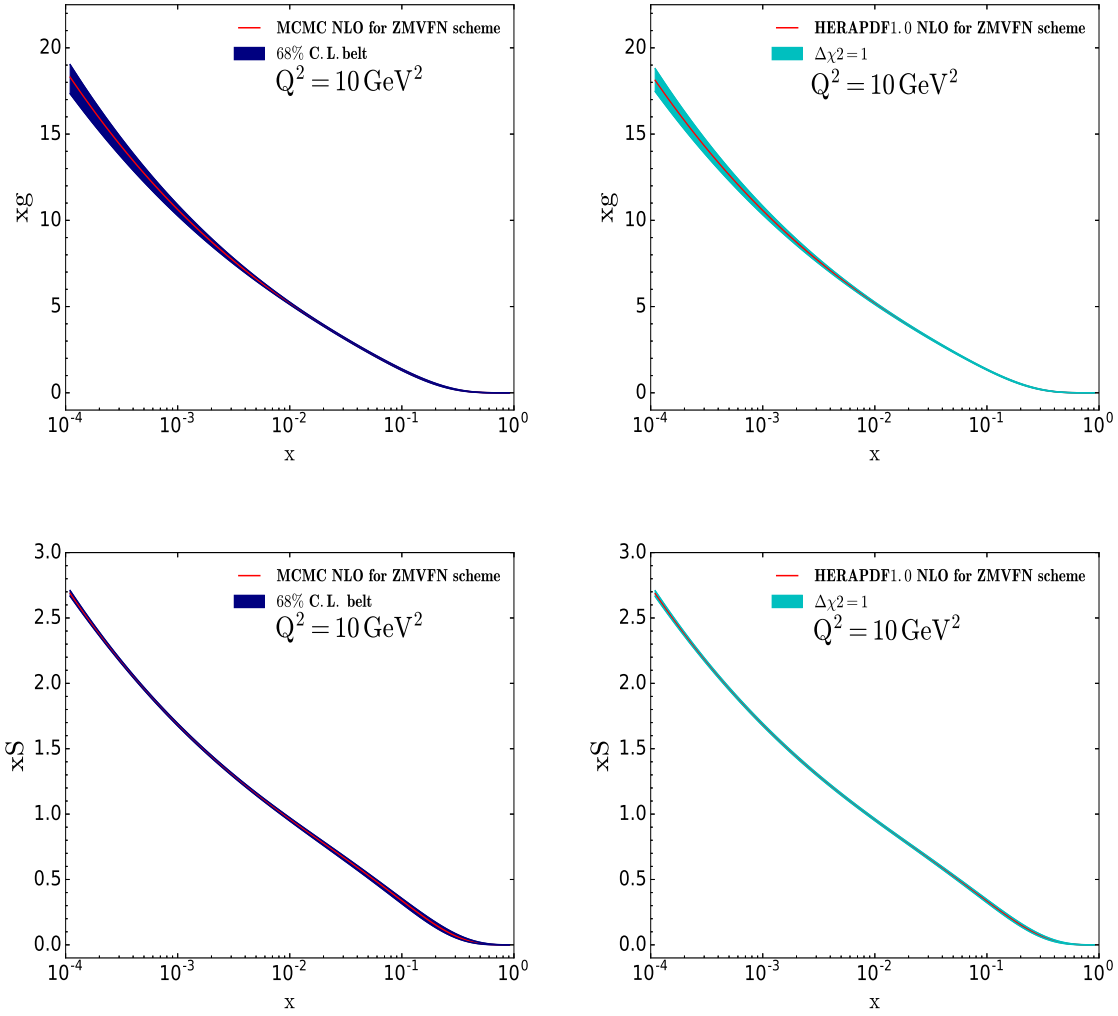


FIGURE 3.8 – Les distributions de partons  $xg$  et  $xS = x\bar{U} + x\bar{D}$  à l'échelle  $Q^2 = 10 \text{ GeV}^2$  obtenues à partir de la chaîne Monte Carlo en utilisant le code d'évolution QCDNUM [BOTJE \[2011\]](#) (courbes de gauche) sont comparées à celles de HERAPDF1.0 (ZM-VFNS) obtenues à partir du code public HERAFitter (courbes de droite). Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur centrale (courbe rouge) pour les PDFs MCMC et la déviation  $\Delta\chi^2 = 1$  standard pour les PDFs HERAPDF1.0.

Les incertitudes expérimentales obtenues par la méthode MCMC et la méthode Hessienne sont aussi conformes dans la gamme cinématique de HERA. Ceci est démontré dans les figures [3.10](#) et [3.11](#) où l'incertitude expérimentale (méthode MCMC) normalisée par la valeur centrale pour les PDFs MCMC est comparée à l'incertitude expérimentale (méthode Hessienne) normalisée par la valeur centrale pour les PDFs HERAPDF1.0 NLO. Les estimations des incertitudes expérimentales pour les distributions de partons de quarks de valence ( $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ) par la méthode MCMC ont tendance à être légèrement supérieures à celles obtenues par la méthode Hessienne (voir figure [3.10](#)). Ceci n'est pas le cas pour les distributions de partons du quark de la mer  $xS$  et du gluon  $xg$  (confère figure [3.11](#)), où les estimations des incertitudes expérimentales par la méthode MCMC sont très grandes à cause du fait que ces distributions de partons sont quasi nulles à haut  $x$  ( $x \approx 0.83$ ) (confère courbes de droite des figures [3.5](#) et [3.6](#)) ; ce qui conduit à des limites

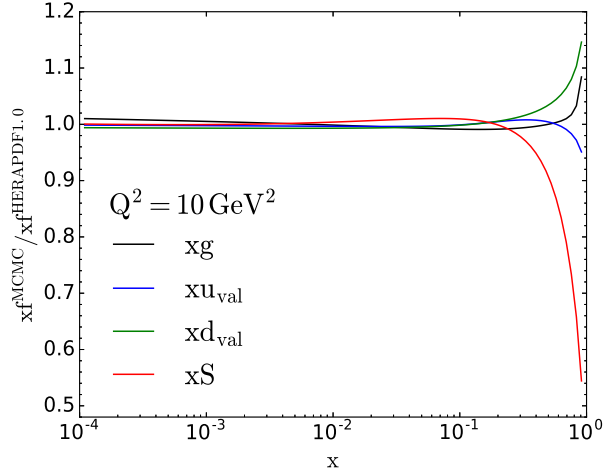


FIGURE 3.9 – Rapport des valeurs centrales des PDFs MCMC et HERAPDF1.0 (ZMVFN scheme) pour  $xg$ ,  $xu_{\text{val}}$ ,  $xd_{\text{val}}$  et  $xS$  à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

de confiance asymétriques (voir tableau 3.5).

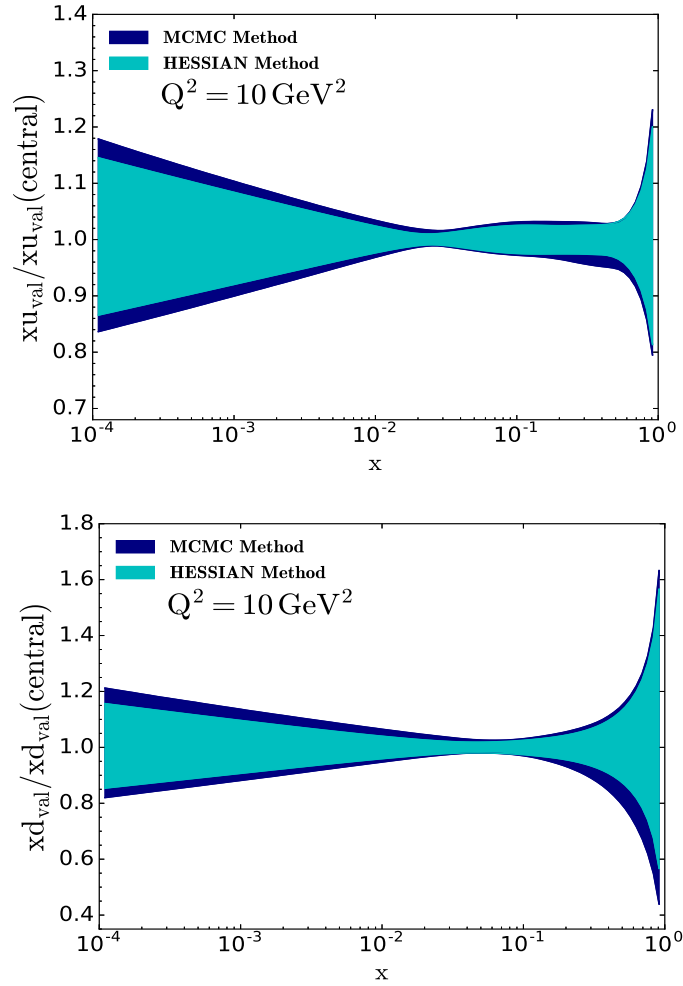


FIGURE 3.10 – Comparaison des incertitudes expérimentales des PDFs normalisées par la valeur centrale, comme déterminées par les méthodes Hessienne et MCMC à l'ordre NLO en QCD pour les distributions des quarks de valence  $xu_{\text{val}}$  et  $xd_{\text{val}}$  à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

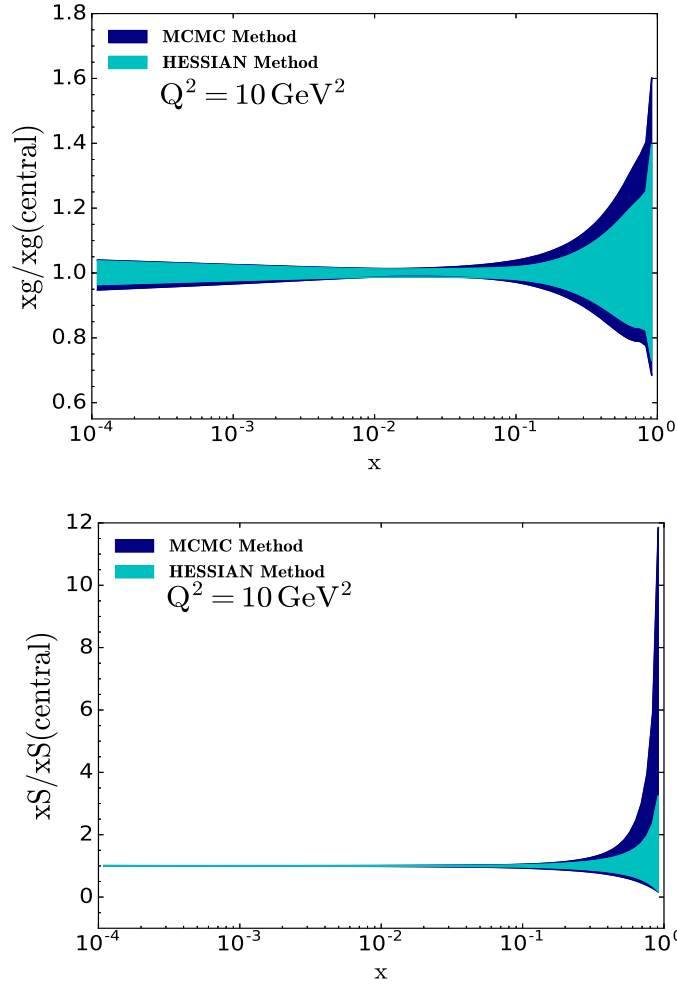


FIGURE 3.11 – Comparaison des incertitudes expérimentales des PDFs normalisées par la valeur centrale, comme déterminées par les méthodes Hessienne et MCMC à l'ordre NLO en QCD pour les distributions de partons du gluon  $xg$  et du quark de la mer  $xS$  à l'échelle  $Q^2 = 10 \text{ GeV}^2$ .

La séparation des distributions de partons en différentes saveurs est illustrée plus en détail dans les figures 3.12 et 3.13, où les distributions de partons de quarks  $xU$ ,  $xD$ ,  $x\bar{U}$ ,  $x\bar{D}$  et  $x\bar{u}$ ,  $x\bar{d}$ ,  $x\bar{s}$ ,  $x\bar{c}$  sont montrées à l'échelle  $Q^2 = 10 \text{ GeV}^2$ . Les distributions de partons  $xU$ ,  $xD$ ,  $x\bar{U}$  et  $x\bar{D}$  sont étroitement liées aux mesures, voir les équations 1.40 et 1.49, et sont très bien contraintes à bas  $x$ . La saveur U est mieux contrainte que la saveur D à cause de la dominance de cette saveur dans toutes les interactions à l'exception du processus de la diffusion du positron sur le proton et faisant intervenir le courant chargé (échange du boson  $W^+$ ). Les distributions de partons des quarks  $x\bar{d}$  et  $x\bar{s}$  sont dérivées de  $x\bar{D}$  à travers l'hypothèse sur la valeur de la fraction  $r_s$  (confère (1.74)). L'incertitude expérimentale sur  $x\bar{d}$  suit étroitement celle sur  $x\bar{D}$  et l'incertitude expérimentale sur  $x\bar{u}$  suit

étroitement celle sur  $x\bar{U}$ . La distribution du quark charmé  $x\bar{c}$  est montrée à l'échelle  $Q^2 = 10 \text{ GeV}^2$  pour que la condition  $Q^2 \gg m_c^2$  soit remplie, et il peut être considéré comme un parton pleinement actif. Toutefois, elle est fortement liée à la distribution du gluon qui est indirectement déterminée à partir de la violation d'échelle.

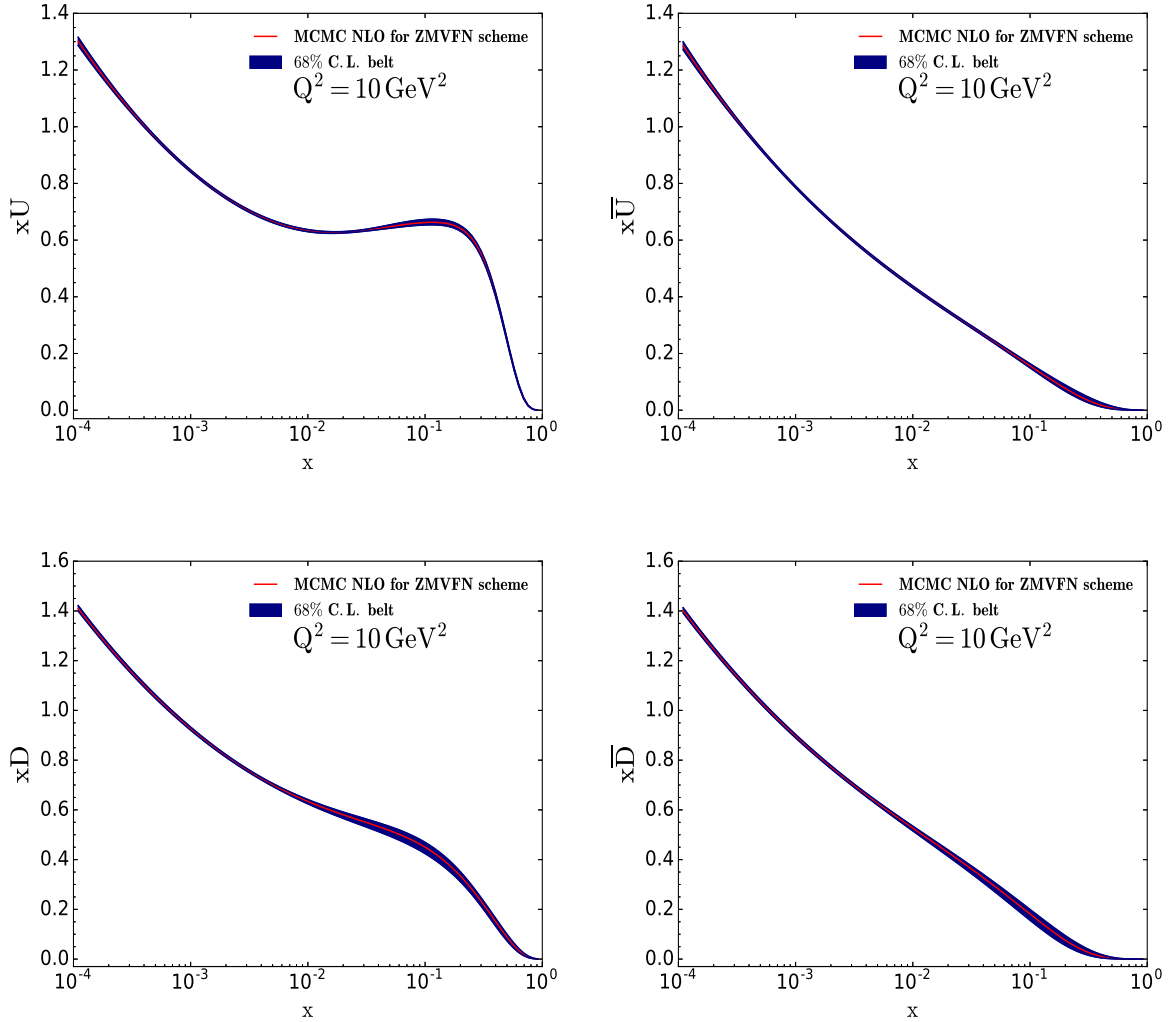


FIGURE 3.12 – Les distributions des quarks  $xU$ ,  $xD$ ,  $x\bar{U}$  et  $x\bar{D}$  obtenues par la méthode MCMC à l'échelle  $Q^2 = 10 \text{ GeV}^2$ . Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur du "best fit" (courbe rouge).

La figure 3.14 montre un récapitulatif des courbes des distributions de partons  $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ,  $xS$  et  $xg$  à l'échelle initiale  $Q^2 = 1.9 \text{ GeV}^2$  et à l'échelle  $Q^2 = 10 \text{ GeV}^2$  obtenues par la méthode MCMC.



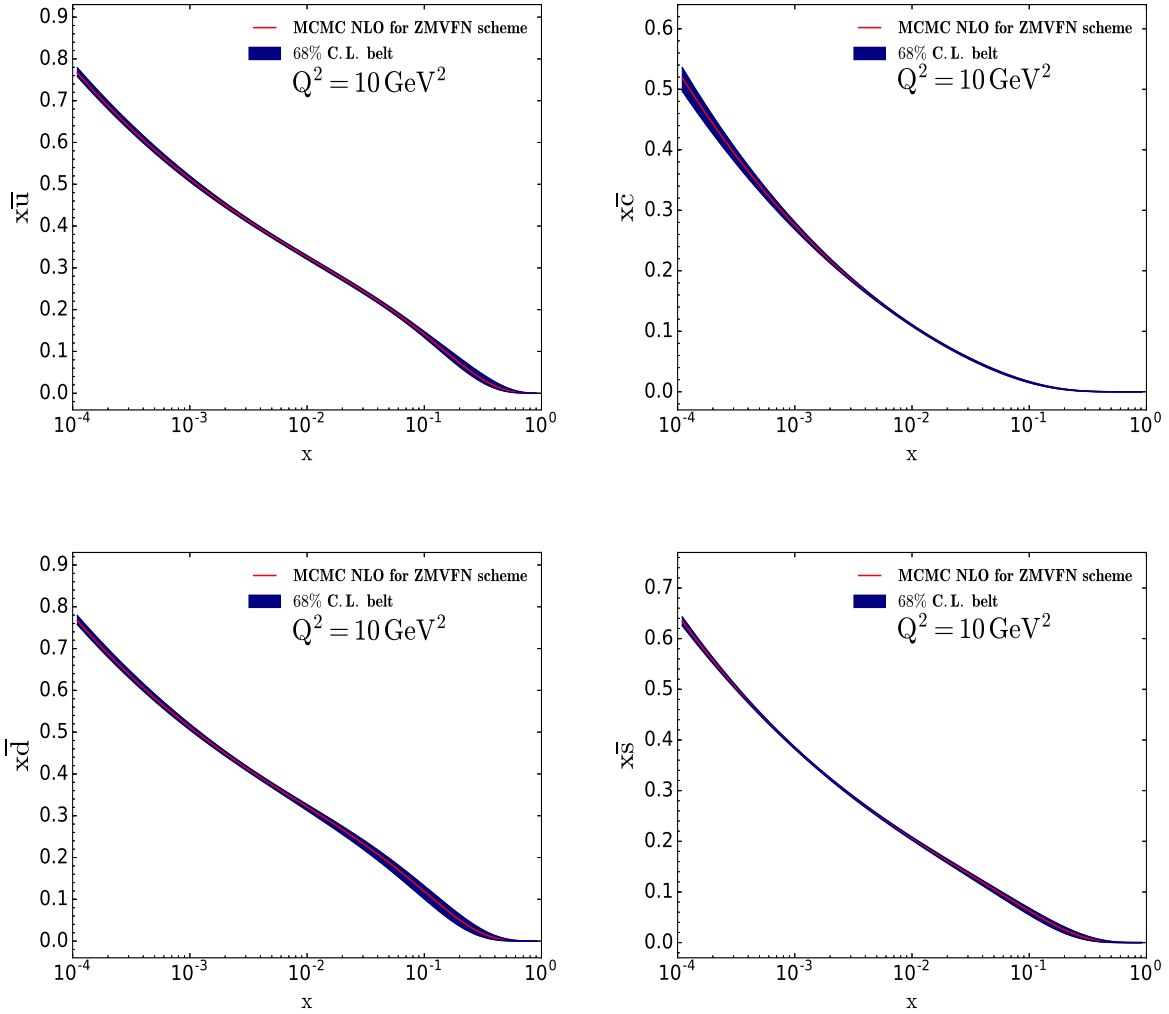


FIGURE 3.13 – Les distributions des quarks  $x_{\bar{u}}$ ,  $x_{\bar{d}}$ ,  $x_{\bar{s}}$  et  $x_{\bar{c}}$  obtenues par la méthode MCMC à l'échelle  $Q^2 = 10 \text{ GeV}^2$ . Les enveloppes montrent l'intervalle de confiance à 68% autour de la valeur du "best fit" (courbe rouge).

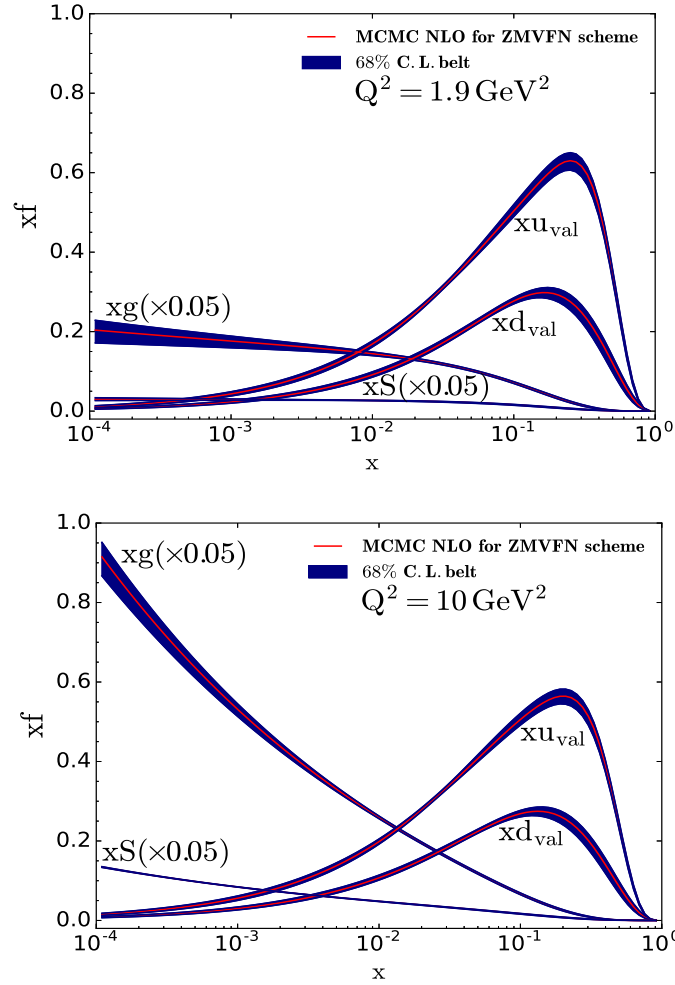


FIGURE 3.14 – Les distributions des quarks de valence ( $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ), du quark de la mer ( $xS = x\bar{u} + x\bar{d} + x\bar{s} + x\bar{c}$ , avec  $x\bar{c} = 0$  pour  $Q^2 < m_c^2$ ) et du gluon ( $xg$ ) obtenues avec le schéma ZM-VFNS en utilisant la méthode MCMC à l'échelle  $Q^2 = 1.9 \text{ GeV}^2$  (courbe de haut) et à l'échelle  $Q^2 = 10 \text{ GeV}^2$  (courbe de bas). Les distributions du quark de la mer et du gluon sont réduites d'un facteur de 20. Les incertitudes expérimentales (les limites de confiance à 68% comme définies à partir des densités de probabilité des distributions de partons  $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ,  $xS$  et  $xg$ ) sont représentées par des régions de couleur bleu.

La figure 3.15 montre de telles courbes sur une échelle logarithmique de sorte que le comportement et les incertitudes expérimentales des PDFs à haut  $x$  sont mis en exergue.

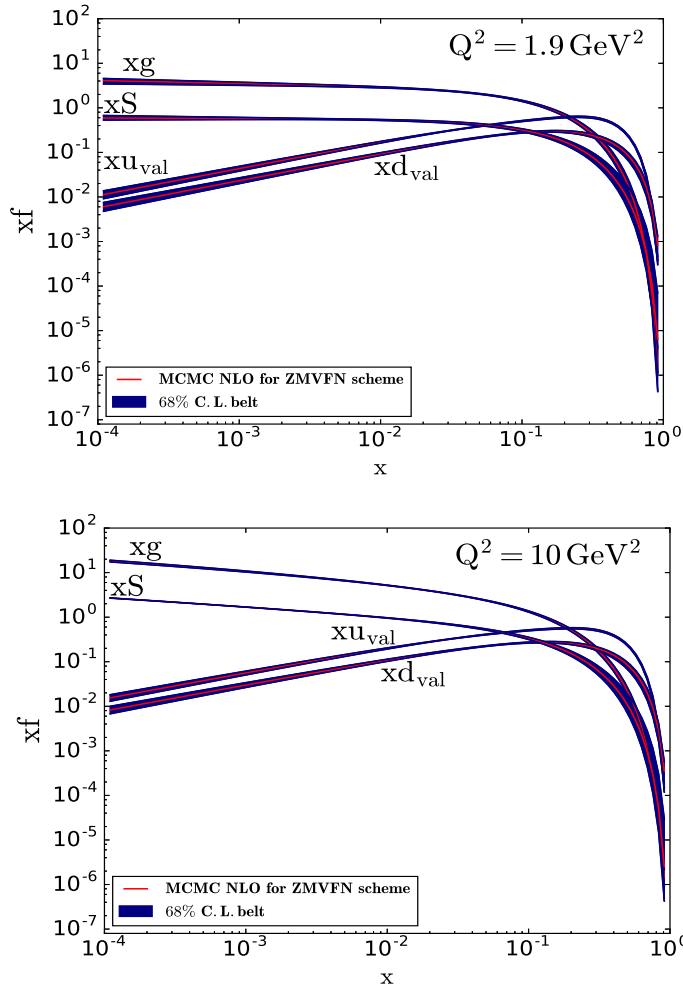


FIGURE 3.15 – Les distributions des quarks de valence ( $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ), du quark de la mer ( $xS = x\bar{u} + x\bar{d} + x\bar{s} + x\bar{c}$ , avec  $x\bar{c} = 0$  pour  $Q^2 < m_c^2$ ) et du gluon ( $xg$ ) obtenues avec le schéma ZM-VFNS en utilisant la méthode MCMC à l'échelle  $Q^2 = 1.9 \text{ GeV}^2$  (courbe de haut) et à l'échelle  $Q^2 = 10 \text{ GeV}^2$  (courbe de bas). Les incertitudes expérimentales (les limites de confiance à 68% comme définies à partir des densités de probabilité des distributions de partons  $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ,  $xS$  et  $xg$ ) sont représentées par des régions de couleur bleu.

## Références

- AARON, F. D. et AL. 2010, «Combined measurement and qcd analysis of the inclusive electron(positron) proton scattering cross sections at heras», *JHEP 01 (2010) 109*. [49](#), [51](#)
- BOTJE, M. 2011, «Qcdnum : Fast qcd evolution and convolution», *Comput.Phys.Commun. 182*, p. 490–532. [xii](#), [59](#), [60](#)
- XFITTER DEVELOPERS. 2017, «xfitter pdf fitting package», <http://www.xfitter.org/>. [xii](#), [49](#), [50](#)

# Conclusion et perspectives

Cette thèse porte sur l'application des techniques Monte Carlo par chaînes de Markov à la détermination des distributions de partons. Le code public développé par la collaboration HERAFitter et son successeur xFitter pour déterminer les distributions de partons et leurs incertitudes expérimentales, a été modifié afin d'implémenter l'algorithme Monte Carlo Hamiltonien et d'obtenir un code pour l'extraction des incertitudes expérimentales d'une manière statistiquement contrôlée. Ces incertitudes expérimentales, grandeurs centrales de l'analyse présentée, ont été estimées de manière rigoureuse.

La fonction  $\chi^2$  utilisée dans cette analyse dépend du schéma de saveurs lourdes. Ceci rend la méthode MCMC très coûteux en temps de calcul. Nous avons pour cette raison utilisés le schéma ZM-VFNS qui néglige les masses des quarks  $c$  et  $b$ . Aussi, une coupure significative a été mise sur les données combinées de HERA mesurées durant la période 1994-2000 pour réduire le nombre de points de données expérimentales à 537. La méthode de minimisation standard basée sur la routine MINUIT est beaucoup plus rapide mais reste beaucoup plus sensible aux conditions initiales, puisqu'elle converge vers la première valeur minimale de la fonction de  $\chi^2$  qu'elle trouve dans l'espace des paramètres qui pourrait ne pas être le minimum global. Les incertitudes obtenues grâce à la méthode MCMC et celle utilisant le critère de tolérance conventionnel, c'est-à-dire  $\Delta\chi^2 = 1$  sont cohérentes dans la gamme cinématique de HERA, même si les incertitudes MCMC ont tendance à être supérieures à celles relatives à l'approche Hessienne. Cette différence est peut être liée au fait que la collection des points obtenus à divers  $x$  ne forme pas ensemble une bonne distribution de partons, c'est-à-dire ne satisfait pas les règles de somme.

La suite de ce travail consistera à étendre la présente étude à l'ensemble complet des paramètres libres des PDFs, y compris comme paramètres libres, la constante de couplage fort  $\alpha_s$  et les masses des quarks  $c$  et  $b$ ; en étudiant, en particulier grâce aux contraintes expérimentales sur ces derniers paramètres, l'impact des densités de probabilité a priori. Nous envisageons aussi de considérer une fonction  $\chi^2$  plus complète incluant les corrélations systématiques et en utilisant les techniques des grilles pour calculer les schémas de saveurs lourdes GM-VFNS (voir sous-section 1.4.3) très coûteux en temps de calcul dans les processus de diffusion inélastique profonde. Enfin il serait intéressant d'analyser l'impact de nouveaux ensembles de données expérimentales ajoutés à l'analyse, de vérifier quel jeu de données expérimentales a des valeurs aberrantes et si ces dernières peuvent être tolérées.



# Publication

## Markov chain Monte Carlo techniques applied to parton distribution functions determination: Proof of concept

Yémalin Gabin Gbedo and Mariane Mangin-Brinet\*

*Laboratoire de Physique Subatomique et de Cosmologie-Université Grenoble-Alpes,  
CNRS/IN2P3, 53, avenue des Martyrs, 38026 Grenoble, France*

(Received 26 January 2017; published 21 July 2017)

We present a new procedure to determine parton distribution functions (PDFs), based on Markov chain Monte Carlo (MCMC) methods. The aim of this paper is to show that we can replace the standard  $\chi^2$  minimization by procedures grounded on statistical methods, and on Bayesian inference in particular, thus offering additional insight into the rich field of PDFs determination. After a basic introduction to these techniques, we introduce the algorithm we have chosen to implement—namely Hybrid (or Hamiltonian) Monte Carlo. This algorithm, initially developed for Lattice QCD, turns out to be very interesting when applied to PDFs determination by global analyses; we show that it allows us to circumvent the difficulties due to the high dimensionality of the problem, in particular concerning the acceptance. A first feasibility study is performed and presented, which indicates that Markov chain Monte Carlo can successfully be applied to the extraction of PDFs and of their uncertainties.

DOI: 10.1103/PhysRevD.96.014015

### I. INTRODUCTION AND MOTIVATION

Quantum chromodynamics (QCD) is the theory of strong interaction, whose ambition is to explain nuclei cohesion as well as neutron and proton structure, i.e. most of the visible matter in the Universe. Its application domain is even wider, since QCD controls the structure and interactions of all hadrons: proton, neutron, hyperons, pions, kaons, etc. It is one of the most elegant theories of science (with general relativity); it has only very few parameters and allows us to give a physical interpretation to a very broad range of phenomena using a well-defined and very compact formalism.

Among the fundamental ingredients of QCD, parton distribution functions (PDFs) are key elements and play an essential role to connect the QCD dynamics of quarks and gluons to the measured hard scattering cross sections of colliding hadron(s). They carry an invaluable source of information on the hadrons' partonic structure, and enormous theoretical and experimental effort has been devoted for years to the extraction of these distribution functions.

PDFs are all the more important nowadays; with the start of data taking at the LHC, they are essential for the computation of a large class of observables. Built for the discovery of the Higgs boson and the study of physics beyond the standard model, the LHC is indeed essentially a QCD factory, producing events in an unexplored energy range. The LHC potential of discovery crucially depends on the quality of predictions for QCD signals and backgrounds and, thus, on the PDFs quality.

PDFs are intrinsically nonperturbative objects and, thus, cannot be determined using only perturbative QCD tools. One of the most efficient methods to perform non-perturbative QCD calculations is Lattice QCD. However, although nucleon structure has been the subject of intense activity in Lattice QCD for years, and even if promising techniques to compute PDFs directly on the lattice have recently been proposed [1], *ab initio* calculations of PDFs are very challenging and still not a competitive alternative to global analyses. These latter, thus, remain the chosen method to obtain PDFs, which are parametrized by functional forms whose parameters are constrained by fits to the data.

Parton distribution function determination thus consists in extensive exploitation of data sets collected at colliders to constrain the parameters of the PDF functional forms given at a fixed scale in energy. Such analyses are usually based on a least square fit method, i.e. on the minimization of a  $\chi^2$ , which compares the input data and theory predictions. PDFs determined this way did not, for a long time, include any estimate of uncertainties, other than the mere comparison of results provided by different global analyses collaborations. However, with the advent, at the dawn of the 21st century, of the new generation of colliders and the active search for new physics, the need to assess the uncertainties of the PDFs became clear [2–5]. Many studies have since been devoted to the estimates of uncertainties on physical predictions due to the uncertainties of PDFs (see for instance [6–8] and references therein) and considerable progress has been made. Nevertheless, this task is far from being trivial, and many issues remain open [6].

Our current understanding of PDF uncertainties is mainly based—with the notable exception of neural

\*Corresponding author.  
mariane@lpsc.in2p3.fr

network techniques—on the Hessian or the Lagrange multiplier method [6,7]. The determination of the uncertainties then relies on an assumption on the permissible range of “acceptable”  $\Delta\chi^2$  for the fit and the choice of a tolerance parameter. In order to improve on this point and get a deeper insight, we propose to use Markov chain Monte Carlo (MCMC) techniques to define the uncertainties in a way based as much as possible on robust statistical methods. Markov chain Monte Carlo algorithms have been an extremely popular tool in statistics. While these techniques are already widely used in many areas of physics (see for instance [9–11]) they have not yet been employed as a stand-alone method to compute PDF parameters and their errors—i.e. without resorting to a  $\chi^2$  minimization procedure.<sup>1</sup>

The Markov chain Monte Carlo method allows us to estimate *a posteriori* probability densities for multidimensional models and provides reliable estimates of errors. MCMC consists in sequentially simulating a single Markov chain whose limit distribution is the chosen one.

The main challenge of the present paper is to demonstrate that Markov chain Monte Carlo techniques can be applied to PDF extraction. The higher the dimension of the chain (i.e. in our case, the more PDF free parameters to be determined), the more computing time is needed to generate the chain. The large number of parameters to be computed in a full PDF determination has led us to make use of a Monte Carlo algorithm based on molecular dynamics, initially developed for lattice field theory. We apply this algorithm to a realistic (though not full) extraction of PDFs, based on ten-parameter functional forms and four data sets, to demonstrate that Markov chain Monte Carlo can successfully be applied to PDF computation.

This paper is organized as follows: in Sec. II, we formulate the PDF determination problem in terms of Bayesian inference. In the next section, the basic principles of Markov chain Monte Carlo methods are recalled and illustrated utilizing the widely used Metropolis algorithm. The fourth section briefly presents the hybrid Monte Carlo algorithm and shows how it allows us to deal with the large number of PDF parameters to be determined. Section V details the MCMC analysis procedure and Sec. VI displays first results with a realistic run using ten parameters and four sets of data. Conclusions and outlook are discussed in the last section.

## II. FORMULATION OF THE PROBLEM IN TERMS OF BAYESIAN INFERENCE

Parton distribution functions are usually (with the exception of neural network procedure [12]) parametrized at a given energy scale by functional forms, which are then evolved at any other scale thanks the DGLAP equations

<sup>1</sup>Uncertainties estimations using pseudodata replicas are also based on Monte Carlo methods [8], but still rely on fits.

(this also excepts dipole models [13] and transverse-momentum dependent [14] and unintegrated PDFs [15], for instance, that we will not consider here). These PDFs are convoluted with partonic cross sections to obtain hadronic cross sections for various processes and a  $\chi^2$  function, constructed from these theoretical cross sections and corresponding experimental data, is then minimized to constrain the PDF parameters. Rather than using a minimization procedure and a Hessian method to estimate PDF uncertainties, we propose a Bayesian parameter inference approach. These techniques have already been successfully applied in many areas [9] and we only sketch the main principles in what follows. The interested reader can be referred for instance to [16] for a more extensive review of the subject.

For compactness, we note  $\hat{q}$  the vector of PDF parameters to be determined:  $\hat{q} = (q^{(1)}, q^{(2)}, \dots, q^{(m)})^T$  where  $m$  is typically, in the case of a full analysis, of the order of 25–30, and  $D$  the data. From a Bayesian perspective, both model parameters and observables are considered random quantities, and Bayesian inference aims formally to determine a joint probability distribution  $P(D, \hat{q})$  over all random quantities. This joint distribution can be written as  $P(D, \hat{q}) = P(D|\hat{q})P(\hat{q})$ , where  $P(\hat{q})$  is a *prior* distribution—quantifying the degree of belief one has *a priori* before observing the data—and  $P(D|\hat{q})$  is the likelihood of the data:  $\mathcal{L}(\hat{q}) \stackrel{\text{def}}{=} P(D|\hat{q})$ . Bayes theorem is used to express the distribution of  $\hat{q}$  conditional on  $D$ ,  $P(\hat{q}|D)$ , in terms of the likelihood  $P(D|\hat{q})$ :

$$P(\hat{q}|D) = \frac{P(D|\hat{q})P(\hat{q})}{\int d\hat{q}P(D|\hat{q})P(\hat{q})} \quad (1)$$

The denominator in (1) does not depend on the parameters and can be considered only as a normalization. This so-called “*posterior*” probability density  $P(\hat{q}|D)$  quantifies the probability to have the model parameters  $\hat{q}$  given the observed data  $D$  and is the object we deal with in all Bayesian inference. To determine this conditional probability, we thus need to set a prior distribution for the parameters, and to compute the likelihood of the data. The probability density  $P(\hat{q}|D)$  is then sampled using a Monte Carlo algorithm.

Assuming that the fluctuations of the  $n$  experimental data points under consideration around their corresponding theoretical values are uncorrelated and distributed according to a Gaussian law (assumption whose validity can be assessed *a posteriori*—see Sec. VI), the least square method and the maximum likelihood are equivalent and the logarithmic likelihood function can be written as

$$\log \mathcal{L}(\hat{q}) = -\frac{1}{2} \sum_{i=1}^n \frac{(D_i - T_i)^2}{\sigma_i^2} = -\frac{1}{2} \chi^2 \quad (2)$$

where  $D_i$  and  $T_i$  denote respectively the  $i$ th experimental point and the corresponding theoretical calculation, and



$\sigma_i^2$  is the uncertainty associated with the measured data  $i$ . The inclusive cross section  $T_i$  in hadron collision can be written as a convolution of PDFs with a partonic cross section, computed at a given order in perturbation theory. The likelihood function (2) thus contains the PDFs. Correlated experimental uncertainties can also be taken into account by introducing for instance a covariant matrix and properly modifying the  $\chi^2$  [17].

For this work, we have used the  $\chi^2$  function given by the default settings provided by the HeraFitter package [17] with 10 parameters (default HeraFitter steering file and minuit.in.txt.10pHERAPDF input file) and four data sets with an initial PDF parametrization set at  $Q_0^2 = 1.9 \text{ GeV}^2$  and a lower cut on the data at  $Q_{\min}^2 = 10 \text{ GeV}^2$ . These settings provide already a computation of PDFs realistic enough for this feasibility study.

We thus apply Bayesian inference to the likelihood function defined in (2), that is, we compute the probability density function of the model parameters, based on selected experimental data. To this purpose, we use Monte Carlo Markov chain procedure, whose principles are briefly sketched in the next section. One of the crucial interest of these methods is the fact that the mean value and uncertainty in these parameters are by-products of the probability density functions computed.

### III. PRINCIPLE OF MARKOV CHAIN MONTE CARLO

#### A. Basics of the method

The Markov chain Monte Carlo method allows us to estimate *a posteriori* probability densities for multi-dimensional models—which, as explained briefly in the previous section, is exactly what we want—and provides reliable estimates of errors. MCMC algorithms enable us to draw samples from a probability distribution known up to a multiplicative constant. They consist in sequentially simulating a single Markov chain whose limiting distribution is the chosen one (in our case, the maximum likelihood times a prior density). More precisely, a Markov chain is a stochastic process characterized by the fact that the conditional distribution of the random variable at iteration  $t$ , denoted  $\hat{q}_t$ , given the ensemble of random variables at all previous steps  $\hat{q}_0, \dots, \hat{q}_{t-1}$ , depends only on  $\hat{q}_{t-1}$ , and not on the previous history. Such a chain can be used to sample a probability density. To converge to a given stationary distribution, the chain needs to satisfy important properties: it has to be irreducible, aperiodic and positive recurrent. We will not expand further on Markov chain theory and we refer the reader interested by formal details to [9] and references therein.

Two ingredients are necessary to define a Markov chain: (i) the initial values (that is the marginal distribution) of parameters and (ii) the transition kernel between two sets of parameters:  $T(\hat{q} \rightarrow \hat{q}')$ , for going from a set  $\hat{q}$  to another

set  $\hat{q}'$ . There are several issues arising when implementing MCMC: the influence of the starting point of the chain (leading to the “burn-in” time), the choice of the transition kernel, the rate of convergence, the acceptance of the algorithm,.... These questions will be illustrated in detail in the following sections.

#### B. Metropolis algorithm

The so-called “Metropolis-Hastings algorithm,” proposed in 1953 by Metropolis *et al.* [18] and generalized by Hastings in 1970 [19], is one of the simplest Monte Carlo algorithms. It is the standard computational workhorse of MCMC methods both for its simplicity and its versatility, and is in principle applicable to any system. It is extremely straightforward to implement and to sample a target density  $P(\hat{q}|D)$  (see Sec. II), it proceeds as follows: at each Monte Carlo time  $t - 1$ , the next state  $\hat{q}_t$  is chosen by sampling a candidate point  $\hat{q}'$  from a proposal distribution  $\pi(\cdot|\hat{q}_{t-1})$ . The candidate point is then accepted with the probability

$$\alpha(\hat{q}_{t-1}, \hat{q}') = \min\left(1, \frac{P(\hat{q}'|D)\pi(\hat{q}_{t-1}|\hat{q}')}{P(\hat{q}_{t-1}|D)\pi(\hat{q}'|\hat{q}_{t-1})}\right),$$

and the Metropolis-Hastings transition kernel is thus

$$T(\hat{q}_{t-1} \rightarrow \hat{q}') = \pi(\hat{q}'|\hat{q}_{t-1})\alpha(\hat{q}_{t-1}, \hat{q}').$$

If the new set of parameters  $\hat{q}'$  is accepted, the next state of the chain becomes  $\hat{q}_t = \hat{q}'$ . If it is rejected, the chain does not move and the point at  $t$  is identical to the point at  $t - 1$ :  $\hat{q}_t = \hat{q}_{t-1}$ .

A special case of the Metropolis-Hastings algorithm is the random walk Metropolis, for which the proposal distribution is chosen to be such that  $\pi(\hat{q}'|\hat{q}_{t-1}) = \pi(\hat{q}_{t-1}|\hat{q}')$ . The acceptance probability then reduces to  $\alpha(\hat{q}_{t-1}, \hat{q}') = \min(1, \frac{P(\hat{q}'|D)}{P(\hat{q}_{t-1}|D)})$ . Frequently, the proposal for the random walk jump has a form which depends on a scale parameter, giving the typical “size” of the leap from one site to the other. For instance the proposal distribution for  $\hat{q}'$  could be a normal distribution centered in  $\hat{q}_{t-1}$  with a standard deviation  $\sigma$ . A meticulous attention has to be taken when choosing this scale parameter. If it is too large, a very high percentage of the iterations will be rejected, leading to an inefficient algorithm. If it is too small, the random walk will explore the parameter space very slowly, leading again to inefficiency. This problem is all the more difficult to handle that the number of parameters (i.e. the dimension of the vector  $\hat{q}$ ) to be sampled increases.

Ideally, to optimize the efficiency of the MCMC, the proposal distribution should be as close as possible to the target distribution. In practice, the performance of the algorithm is obviously highly dependent on the choice of proposal distribution  $\pi(\cdot|\hat{q}_{t-1})$  and several options are

usually considered in the literature to explore the parameter space: one-dimensional Gaussian distributions, multivariate Gaussian distributions, a distribution obtained by binary space partitioning [20] .... However, even if adjustments of the proposal distributions improve the Metropolis efficiency, they are not effective enough to efficiently deal with several dozen parameters within a reasonable CPU and user time. To circumvent these problems—since in the case of PDF extraction, the number of free parameters to determine (that is the number of parameters in the PDF functional form) is of the order of  $\sim 25\text{--}30$ —we have implemented a much more efficient algorithm, based on molecular dynamics, which has initially been developed for Lattice QCD and is widely used in this field.

#### IV. HYBRID (OR HAMILTONIAN) MONTE CARLO

As mentioned before, the main problem of Metropolis-type algorithms, relies on the choice of the candidate point at each move of the chain. Choosing a trial point far from the initial one will lead to a large change in the distribution to sample, and thus to a small acceptance probability, while choosing a point close to the initial one will not lead to an efficient exploration of the parameter space, and thus to a slow convergence of the chain.

Hamiltonian (or “hybrid”) dynamics [21], developed originally for lattice field theory, is used to produce candidate proposals for Metropolis algorithm, in a very elegant and efficient way. It is an exact algorithm which combines molecular dynamics evolution with a Metropolis accept/reject step. This latter is used to correct for discretization errors in the numerical integration of the corresponding equations of motion. Very good reviews and papers exist which detail the properties of this algorithm (see for instance [16]), and only the main ideas will be recalled here for completeness.

To implement hybrid Monte Carlo algorithm, one introduces for each set of parameters  $\hat{q}$  (see previous section) a set of conjugate momenta  $\hat{p}$  and associates to this joint state of “position”  $\hat{q}$  and “momentum”  $\hat{p}$  an Hamiltonian  $H(\hat{q}, \hat{p}) = \hat{p}^T M^{-1} \hat{p} / 2 + \mathcal{U}(\hat{q})$ , where  $M$  is a mass matrix, generally taken to be diagonal, and  $\mathcal{U}(\hat{q})$  an arbitrary potential energy. This allows us to define a joint distribution as

$$P(\hat{q}, \hat{p}) = \frac{1}{Z} e^{-H(\hat{q}, \hat{p})} = \frac{1}{Z} e^{-\mathcal{K}(\hat{p})} e^{-\mathcal{U}(\hat{q})}$$

where  $Z$  is the normalizing constant. We use for the potential energy  $\mathcal{U}(\hat{q}) = -\log[P(D|\hat{q})P(\hat{q})]$ . Starting from a point  $\hat{q}_0$  of the chain, the HMC procedure consists in selecting some initial momenta  $\hat{p}_0$  normally distributed around zero and let the system evolve deterministically for some time according to Hamilton’s equations of motion for  $H(\hat{q}, \hat{p})$ . It reaches a candidate point  $(\hat{q}_1, \hat{p}_1)$

which, according to Metropolis procedure described above, is accepted with probability  $\min(1, e^{-\Delta H})$ . Since the dynamics conserves energy, i.e.  $\Delta H = 0$  along a trajectory, the acceptance rate is 100%, independently of the dimension of the vector  $\hat{q}$ .

In practice, this acceptance is degraded because of the numerical resolution of Hamilton equations, but remains still at a very high level (typically of the order of 70%–90%, independently of the dimension of the chain). HMC algorithm is thus very well suited to multiparameter determination. To discretize Hamilton’s equation, we use the LeapFrog method, a convenient second order integration method that gives the time reversal invariance needed for the Metropolis transition kernel.

We have implemented both Metropolis and hybrid Hamiltonian Monte Carlo algorithms in the open-source package HeraFitter and its successor xFitter [17]. This software provides a modular framework to determine PDFs by fitting a large ensemble of experimental data. In what follows, we focus on the proton PDFs, and we use for the PDF parametrization, the HERAPDF functional form, that we just recall here for the sake of clarity: the parametrized HERA PDFs are the valence distribution  $xu_v$  and  $xd_v$ , the gluon distribution  $xg$ , and the  $\bar{U}$  and  $\bar{D}$  distribution defined as  $x\bar{U} = x\bar{u}$ ,  $x\bar{D} = x\bar{d} + x\bar{s}$ . Their functional form reads

$$xf_a(x) = A_a x^{B_a} (1-x)^{C_a} (1 + D_a x + E_a x^2), \quad (3)$$

where  $a$  labels a parton ( $g, u_{\text{val}}, d_{\text{val}}, \dots$ . See [17] for more details). The analysis procedure we apply to the Markov chain we have produced is explained in the next section.

#### V. MARKOV CHAIN ANALYSIS

Assessing statistical errors for observables in Monte Carlo simulations is a subtle task and requires a careful treatment of the Markov chain. This section presents the different stages of analysis and the checks we have performed.

The procedure to analyze a Markov chain consists of several steps. In particular, it is necessary to remove the thermalization (or burn-in) region, to verify the convergence of the chain and to properly examine correlations between neighboring points in the chain. We have also checked the chain reversibility and the fact that the distribution was correctly sampled.

##### A. Thermalization

The thermalization time (or burn-in length)  $b$  of a Markov chain corresponds to a number of states  $\{\hat{q}_t\}_{t=1, \dots, b}$  to be discarded from the beginning so that the chain forgets its starting point. It can be estimated as being the first state of the random walk—that is the first set of parameters  $\hat{q}$ , denoted  $\hat{q}_b$ —reaching the median value

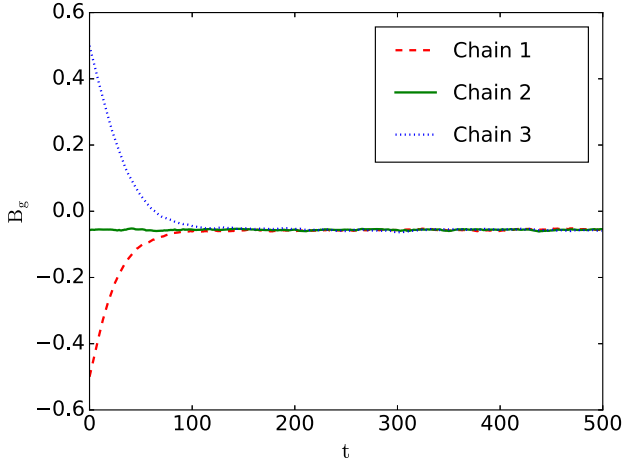


FIG. 1. Values of the parameter  $B_g$  as a function of the Monte Carlo time for three independent Markov chains. The green solid line represents a chain starting from the value given by MINUIT minimization, whereas the two other chains start from values much higher (blue dotted) or much lower (red dashed). The initial corresponding  $\chi^2/\text{d.o.f.}$  values are, from chain 1 to chain 3, respectively,  $\chi^2/\text{d.o.f.} = 67.44$ ,  $\chi^2/\text{d.o.f.} = 0.87$  and  $\chi^2/\text{d.o.f.} = 81.58$ . We identify clearly on this plot the thermalization region, which is limited to the first  $\sim 100$ – $210$  iterations.

of its target distribution  $P_{1/2}$  computed using the entire chain, i.e.

$$P(\hat{q}_b|D) > P_{1/2}. \quad (4)$$

To illustrate thermalization features, we have represented in Fig. 1 the Monte Carlo history of parameter  $B_g$  for three independent chains, each starting from a different point. For illustration purposes, we have fixed all parameters to their value given by the standard minimization procedure of xFitter, except this parameter  $B_g$ , which is obtained by a Monte Carlo procedure.

We can identify in Fig. 1 the thermalization region, whose extent depends on the starting point. The chain represented in solid green line has been started from the output value of a MINUIT minimization of the  $\chi^2$  with respect to parameter  $B_g$  and is thus thermalized very quickly, specifically after one iteration, whereas the other chains, started far from the minimum  $\chi^2$ , exhibits a thermalization of about 150 iterations for the chain represented in dashed red line and 210 iterations for the one represented by the dotted blue line; as expected, the farther from the minimum the starting point is, the longer the thermalization. Starting from a point far from its value given by the minimization procedure is useful to check that this latter did not get stuck in a local minimum and that simulations starting from different points converge to the same region (see also Sec. V C). In practice, we have generated several chains (36 chains, to be more precise),

starting from random points. We determined the thermalization time using Eq. (4) for each chain, and we removed from each chain its burn-in length.

## B. Treatment of autocorrelations

By construction of a Markov chain, the state  $\hat{q}_t$  depends strongly on the state  $\hat{q}_{t-1}$  and quantities computed from this chain form themselves a Markov chain with inherent correlations from one member to the next. These type of correlations are often referred to as ‘‘autocorrelations’’ in simulation time.

Let us consider that we would like to extract an observable  $O$  from a Markov chain simulation with  $N$  points. For this estimation, we use the  $N$  successive Monte Carlo estimates  $O_t$  (we assume in what follows, that the thermalization region has already been discarded, i.e. that the chain has been equilibrated before recording data) and we compute the usual mean  $\langle O \rangle$  where  $\langle \cdot \rangle$  means averaging over the  $N$  data points. The usual estimate of root-mean-square deviation of this average can be computed as

$$\sigma_{\text{naive}}^2 = \frac{N}{N-1} (\langle O^2 \rangle - \langle O \rangle^2)$$

This ‘‘naïve’’ error relies on the assumption that the measurements performed on the Markov chain are not correlated, which is in general not true. In order to account for the correlations, one can introduce for the given observable  $O$ , the integrated autocorrelation time  $\tau_{\text{int}}$ , which can be defined as follows,

$$\tau_{\text{int}} = \frac{1}{2} \sum_{-\infty}^{+\infty} \rho(s),$$

where  $\rho(s)$  is the normalized autocorrelation function,

$$\rho(s) = \frac{(O_t - \langle O \rangle)(O_{t+s} - \langle O \rangle)}{(O_t - \langle O \rangle)^2}.$$

The dependence of  $\rho(s)$  on the time separation  $s$  only is a consequence of the chain being in equilibrium. The integrated autocorrelation time controls the statistical error in Monte Carlo measurement of  $\langle O \rangle$  and there are mostly two possibilities to incorporate this autocorrelation time in the assessment of the statistical errors. The first one consists in leaving out  $2\tau_{\text{int}}$  points between two effective points, or in other terms, to do a subsampling by rejecting all states which are closer than  $2\tau_{\text{int}}$  to each other, in order to get independent states. This approach has the disadvantage of requiring the *a priori* knowledge of  $\tau_{\text{int}}$ . The second approach consists in keeping all measurements, but taking into account the autocorrelation time to estimate the statistical errors. The statistical error of correlated measurements can indeed be computed by [10,22]

$$\sigma_\tau^2 = 2\tau_{\text{int}}\sigma_{\text{naive}}^2$$

This means that the number of “effectively independent data points” in a run of length  $N$  is roughly  $N/(2\tau_{\text{int}})$ . If the integrated autocorrelation time is used to assess statistical errors, this means of course that a reliable estimate of  $\tau_{\text{int}}$  and its error itself are needed. Such estimates require a delicate procedure. An efficient method—called  $\Gamma$ -method—has been developed in [22], which relies on the explicit determination of autocorrelation functions and autocorrelation times. This method provides not only numerical estimators of the integrated autocorrelation time, but also estimates for mean values and statistical errors for arbitrary functions of elementary observables in Monte Carlo simulations. We refer the interested reader to [22] and references therein for details. We have used the  $\Gamma$ -method both to obtain the autocorrelation time and to compute observables.

A further method to reliably estimate the error on uncorrelated measurements is the so-called “*jackknife binning*” [23]. It consists in building  $N$  subsets of data from the initial ensemble of size  $N$ , by removing one observation, leaving samples of size  $N - 1$ . Preaveraging over the blocks of data provides  $N$  estimates of the average:

$$\langle O \rangle_B = \frac{1}{N-1} \left( \sum_{t=1}^{(B-1)} O_t + \sum_{t=B+1}^N O_t \right), \quad B = 1, \dots, N.$$

The jackknife mean and variance for the observable  $O$  are then constructed from

$$\langle O \rangle_{\text{Jack}} = \frac{1}{N} \sum_{B=1}^N \langle O \rangle_B,$$

$$\sigma_{\text{Jack}}^2 = \frac{N-1}{N} \sum_{B=1}^N (\langle O \rangle_B - \langle O \rangle_{\text{Jack}})^2.$$

The jackknife method—and its extensions—is a widely used procedure, in particular in Lattice QCD. For cross-checks and comparison, we applied in our analysis both  $\Gamma$ -method and jackknife binning techniques. For this latter, to un-correlate the points of a given chain, we performed a subsampling of this chain using the value of the autocorrelation time provided by the  $\Gamma$ -method.

### C. Reversibility and convergence

We have verified that our implementation of HMC algorithm satisfies reversibility with a very good precision (relative accuracy better than  $10^{-6}$ ) and that the average acceptance—computed using the jackknife method, after removing thermalization region and decorrelating the chain—is  $\langle e^{-\Delta H} \rangle = 1.002 \pm 0.016$ , thus insuring that our chains indeed converge towards a stationary distribution. In addition, to exclude the risk of a nonidentified lack of convergence, we have simulated several chains, with different (and random) starting points. This is illustrated in Fig. 2 in the case of one varying parameter, namely  $B_g$  (all others being fixed to their value obtained by standard  $\chi^2$  minimization procedure), where the chains are clearly seen to converge towards the same stationary distribution.

The results displayed in the next section have been obtained after skipping thermalization and properly taking

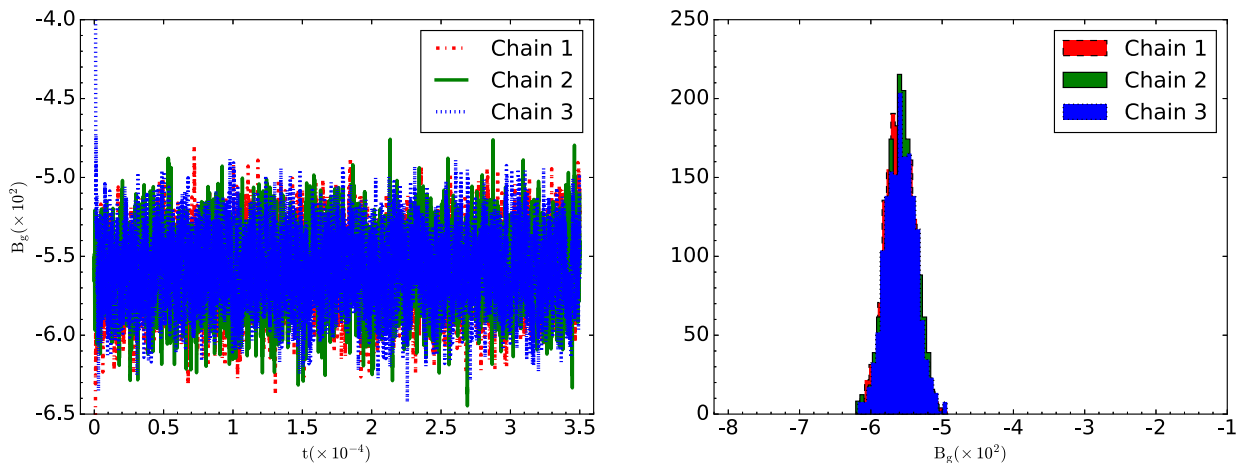


FIG. 2. Values of the parameter  $B_g$  as a function of the Monte Carlo time for three independent Markov chains (lhs). The starting points of the chains and the color code are the same as in Fig. 1. The chains are clearly converging towards the same stationary distribution, what is confirmed by plotting the parameter distribution for each chain (rhs), after removing thermalization points and taking into account autocorrelation.



into account the autocorrelation, using either the  $\Gamma$ -method or the jackknife binning procedure as explained above.

## VI. PRELIMINARY RESULTS

### A. Setup and simulation parameters

The results shown in this section are obtained from a Markov chain using the HERAPDF functional forms for initial PDFs at a scale  $Q_0 = 1.9 \text{ GeV}^2$  with ten free parameters:  $B_g$ ,  $C_g$ ,  $B_{u_{\text{val}}}$ ,  $C_{u_{\text{val}}}$ ,  $E_{u_{\text{val}}}$ ,  $C_{d_{\text{val}}}$ ,  $C_{\bar{U}}$ ,  $A_{\bar{D}}$ ,  $B_{\bar{D}}$  and  $C_{\bar{D}}$  (see expression (3) for the definition of these parameters, and [17] for more details). We have used uniform priors for the parameters, and we consider the same data ensembles than the ones used to produce HERAPDF1.0 distributions.

These data are a combination of inclusive deep inelastic scattering cross sections measured by the H1 and ZEUS Collaborations in neutral and charged current unpolarized  $e^\pm p$  scattering at HERA, during the period 1994–2000 [17]. Other settings and cuts—with the exception of the heavy flavor scheme, see below—were also identical to the ones of HERAPDF1.0 distributions. In particular, we do not rely of K factors, nor on grids techniques. The latter could, however, interestingly be used to speed up the computation.<sup>2</sup>

Since the aim of this work is to demonstrate the applicability of MCMC methods to PDF determination—rather than producing competitive PDF sets—we apply the ZMVFN scheme in order to speed up the computation of the  $\chi^2$ , and a lower cut on the data at  $Q_{\text{min}}^2 = 10 \text{ GeV}^2$ . These settings give a total number of data points of 537. We will denote by “HERAPDF1.0 ZMVFNs” the PDFs thus obtained by minimization.

As already mentioned in the previous section, we have generated 36 Monte Carlo chains, each chain starting from a different random point, using the HMC algorithm.

The HMC algorithm requires the tuning of essentially two parameters: the number of leapfrog steps  $L$  and the step size  $\varepsilon$ , this latter potentially depending on the direction in the parameter space. These two quantities are chosen such as to keep both the acceptance high (requiring small  $\varepsilon$ , to minimize the numerical errors in solving Hamilton equations), and the correlation between two successive Monte Carlo iterations small (thus requiring large trajectory length  $L\varepsilon$ ). We have chosen  $L = 100$  and one leapfrog step size for each parameter, depending on the parameter typical standard deviation. Namely we took  $\varepsilon_i = 3.10^{-2} \Delta q_i^{\text{typ}}$ , where  $\Delta q_i^{\text{typ}}$  is, for each parameter, the value of its standard deviation provided by the minimization. With these HMC parameters, we obtain an acceptance of 80% and chains which have almost no correlation, since the integrated

<sup>2</sup>and we are exploring this possibility for more complete computations.

TABLE I. Comparison of mean values and their statistical errors obtained for PDF parameters using two different analysis methods. The jackknife binning has been applied after subsampling the chain, selecting points not closer than  $2\tau_{\text{int}} = 4$  from each other. Since the number of points considered for the jackknife analysis is thus four times less than for the  $\Gamma$ -method, the errors are expected to be smaller by roughly a factor 2 for this latter procedure, which is indeed the case.

Parameter	$\Gamma$ -method	Jackknife binning
$B_g$	$-0.0537 \pm 0.0001$	$-0.0537 \pm 0.0002$
$C_g$	$5.9449 \pm 0.0015$	$5.9483 \pm 0.0025$
$B_{u_{\text{val}}}$	$0.6124 \pm 0.0001$	$0.6125 \pm 0.0002$
$C_{u_{\text{val}}}$	$4.7458 \pm 0.0003$	$4.7455 \pm 0.0006$
$E_{u_{\text{val}}}$	$14.965 \pm 0.008$	$14.961 \pm 0.012$
$C_{d_{\text{val}}}$	$3.2054 \pm 0.0014$	$3.2077 \pm 0.0016$
$C_{\bar{U}}$	$4.0917 \pm 0.0038$	$4.0961 \pm 0.0048$
$A_{\bar{D}}$	$0.3096 \pm 0.0002$	$0.3098 \pm 0.0002$
$B_{\bar{D}}$	$-0.0174 \pm 0.0001$	$-0.0173 \pm 0.0002$
$C_{\bar{D}}$	$6.2203 \pm 0.0054$	$6.2096 \pm 0.0076$

autocorrelation time  $\tau_{\text{int}}$  computed by the  $\Gamma$ -method is less than 2 for all parameters.

The HMC algorithm also requires the computation of the potential energy (that is in our case the  $\chi^2$ ) with respect to the parameters. These derivatives are computed numerically, using a symmetric derivative. We thus need, for 10 parameters, 20 evaluations of the  $\chi^2$  for each step of the Leapfrog algorithm. We have run 36 jobs in parallel, and collected the results after three days of running. We computed for each of our 36 chains the burn-in length, and we removed the maximum burn-in (namely 28) to all of them, to obtain a total of 4 400 points per chain. To analyze these chains with the  $\Gamma$ -method, we kept all 158 400 points, while the jackknife analysis was done considering one point every four (i.e.  $2\tau_{\text{int}}$ ), that is 39 600 points.<sup>3</sup>

### B. PDF parameter values, marginal distributions and correlations

Table I compares the mean value and its statistical error for each of the ten PDF parameters under consideration, using the two analysis procedures we have presented above, namely  $\Gamma$ -method and jackknife binning. The computation of the integrated autocorrelation time by the  $\Gamma$ -method gives values of  $\tau_{\text{int}}$  less than 2 and we have used for the jackknife one point of the chain every four. As can be seen from Table I, both methods give very close results, showing that we have analyze our Markov chain in a

<sup>3</sup>For a fully realistic PDF determination, systematic uncertainties (factorization and renormalization scales, heavy quark treatment, ...) are much bigger than the sub-percent accuracy we obtain with such a long Markov chain. In realistic cases, a Markov chain with about a thousand decorrelated points will lead to results statistically accurate enough.

TABLE II. Comparison of results obtained for PDF parameter values extracted from an independent Markov chain of length 39 600, and the results provided from a MINUIT minimization. We compute from the MCMC the best fit value, the mean value and the standard deviation value for each of the 10 parameters considered, together with their statistical errors estimated by jackknife binning method. The standard deviation given by the minimization is the usual one-sigma deviation.

Parameter	Values	MCMC	MINUIT minimization
$B_g$	Mean	$-0.0537 \pm 0.0002$	
	Best fit	$-0.0632 \pm 0.0168$	$-0.0559$
	Standard deviation	$0.0299 \pm 0.0001$	$0.0288$
$C_g$	Mean	$5.9483 \pm 0.0025$	
	Best fit	$5.8952 \pm 0.0615$	$5.9274$
	Standard deviation	$0.5037 \pm 0.0019$	$0.5078$
$B_{u_{val}}$	Mean	$0.6125 \pm 0.0002$	
	Best fit	$0.6092 \pm 0.0121$	$0.6098$
	Standard deviation	$0.0371 \pm 0.0001$	$0.0389$
$C_{u_{val}}$	Mean	$4.7455 \pm 0.0006$	
	Best fit	$4.7467 \pm 0.0525$	$4.7122$
	Standard deviation	$0.1280 \pm 0.0005$	$0.1332$
$E_{u_{val}}$	Mean	$14.961 \pm 0.012$	
	Best fit	$15.42 \pm 0.94$	$14.76$
	Standard deviation	$2.494 \pm 0.010$	$2.571$
$C_{d_{val}}$	Mean	$3.2077 \pm 0.0016$	
	Best fit	$3.084 \pm 0.076$	$3.143$
	Standard deviation	$0.3183 \pm 0.0016$	$0.2830$
$C_{\bar{U}}$	Mean	$4.0961 \pm 0.0048$	
	Best fit	$4.323 \pm 0.814$	$4.052$
	Standard deviation	$0.9617 \pm 0.0035$	$0.9782$
$A_{\bar{D}}$	Mean	$0.3098 \pm 0.0002$	
	Best fit	$0.320 \pm 0.039$	$0.305$
	Standard deviation	$0.0485 \pm 0.0002$	$0.0488$
$B_{\bar{D}}$	Mean	$-0.0173 \pm 0.0002$	
	Best fit	$-0.0097 \pm 0.0237$	$-0.0178$
	Standard deviation	$0.0304 \pm 0.0001$	$0.0306$
$C_{\bar{D}}$	Mean	$6.2096 \pm 0.0076$	
	Best fit	$5.888 \pm 0.142$	$5.875$
	Standard deviation	$1.505 \pm 0.009$	$1.290$

consistent way. For the rest of this paper, we will thus display only the results obtained using jackknife binning techniques. We also notice that for the chain length considered ( $\sim 40\,000$  thermalized and decorrelated points), the statistical errors on the mean values are tiny.

In Table II are displayed the results provided by MCMC method—using jackknife binning for error estimate—for the parameters mean, best fit<sup>4</sup> and standard deviation,<sup>5</sup> compared with the output of the standard MINUIT minimization. We notice, as already stated above, that the determination of parameters by Monte Carlo methods, gives much more information than a standard minimization. We can in particular extract the statistical errors on the quantities we are interested in and this error decreases with the length of the Markov chain. As seen from the

table, the best fit value extracted from the parameter probability distributions are compatible with the parameter values provided by the minimization procedure. For what concerns standard deviations however, although MCMC and minimization gives similar results, no precise comparison can be made since the minimization does not provide estimates of errors for the usual one-standard deviation of the parameters. In addition, both quantities (MCMC standard deviation and minimization deviation) should coincide only if the probability density of the parameter considered is Gaussian, which is not necessarily true (see below). This fact is already visible in Table II, where we can see that the mean and best fit values do not coincide within errors for some of the parameters ( $C_{d_{val}}$ ,  $C_{\bar{D}}$  for instance).

The probability distribution functions of the parameters, together with the two-dimensional correlation plots between parameters are displayed in Fig. 3. The marginal posterior parameter distributions are shown on the diagonal graphs, and 2D-correlations on the off-diagonal plots.

<sup>4</sup>The parameter best fit values are the parameters values that minimized the  $\chi^2$  function.

<sup>5</sup>we have computed here the corrected sample standard deviation

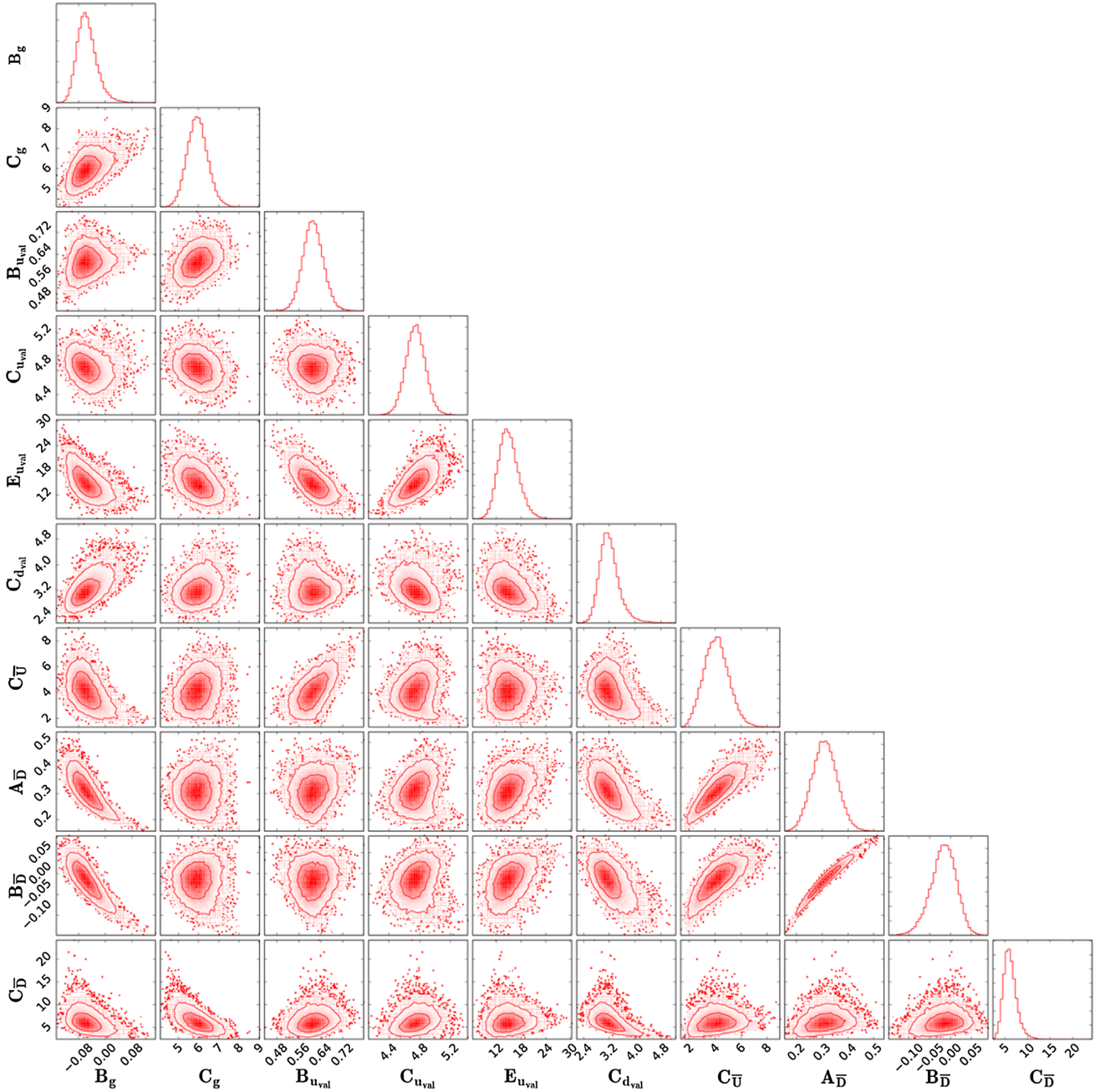


FIG. 3. Probability distribution functions of the PDF parameters (diagonal) and 2D correlation plots between parameters (off-diagonal).

The inner and outer contours of these latter are taken to be regions containing respectively 68% and 95% of the probability density. We note that the probability distribution of some parameters cannot properly be described by a Gaussian law, as illustrated in Fig. 4, and are even non-symmetric. We have also checked that our correlation plots are compatible with the values provided by the covariant matrix.

In Fig. 5 is shown the  $\chi^2/\text{d.o.f.}$  distribution for our ten-dimensional Monte Carlo chain. The solid red line is an

adjustment with a  $\chi^2$  distribution law with 10 degrees of freedom, which perfectly describes our results. The fact that the  $\chi^2$  function built in Eq. (2) follows a  $\chi^2$  distribution law with the expected number of degrees of freedom is a strong indication (though not a formal proof) that our assumptions concerning the fluctuations of the experimental data points around their corresponding theoretical values, are justified.

Though this is not the case in the example we present, we would also like to note that potential flat directions in the

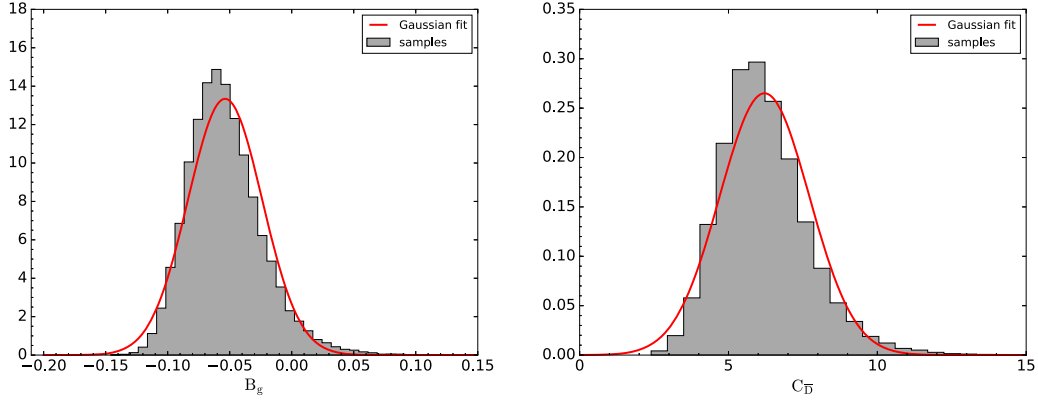


FIG. 4. Marginal probability distribution of parameters  $B_g$  (lhs) and  $C_D$  (rhs). They do not follow a Gaussian law, as can be seen from the gaussian fit (solid red line).

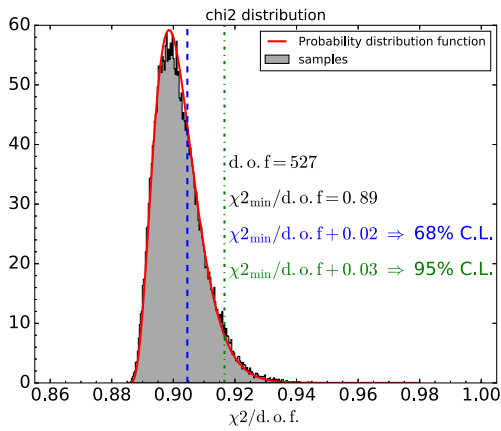


FIG. 5.  $\chi^2$  distribution for a 10D Monte Carlo chain. The solid red line is an adjustment of these data with a  $\chi^2$  distribution law with 10 degrees of freedom. The dashed (dashdot) vertical line indicates the 68% (95%) confidence limit.

parameter space are less problematic for MCMC method than for minimization techniques.

We need now to calculate, from the Markov chain of parameters, the parton distribution functions we are interested in. The procedure we use is explained in the next section and is more generally valid for any observable we want to compute from the MCMC.

### C. PDF marginal distributions and confidence interval

To extract parton distribution functions from the Markov chain, we compute, from the set of 10 parameters obtained at each Monte Carlo iteration, the corresponding PDFs for a range of  $x$  and  $Q^2$  values. This provides the marginal probability density functions of PDFs at fixed  $(x, Q^2)$ , as illustrated on Fig. 6 for the gluon, for two different  $x$  values.

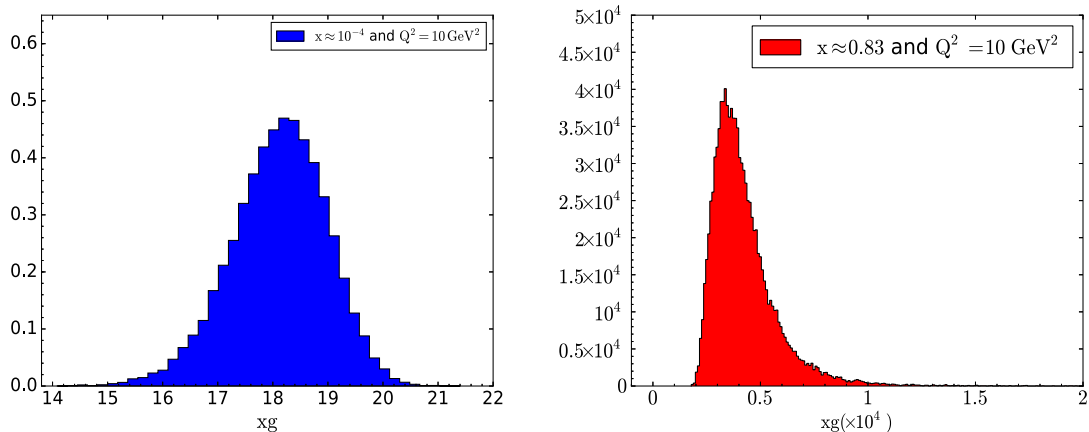


FIG. 6. Gluon PDF probability distribution function for  $x \approx 10^{-4}$  (lhs) and  $x \approx 0.83$  (rhs) at fixed  $Q^2 = 10 \text{ GeV}^2$ . The 68% confidence interval is obtained from this distribution, considering the region of the distribution containing 68% of the data remaining on each side of the best fit value.



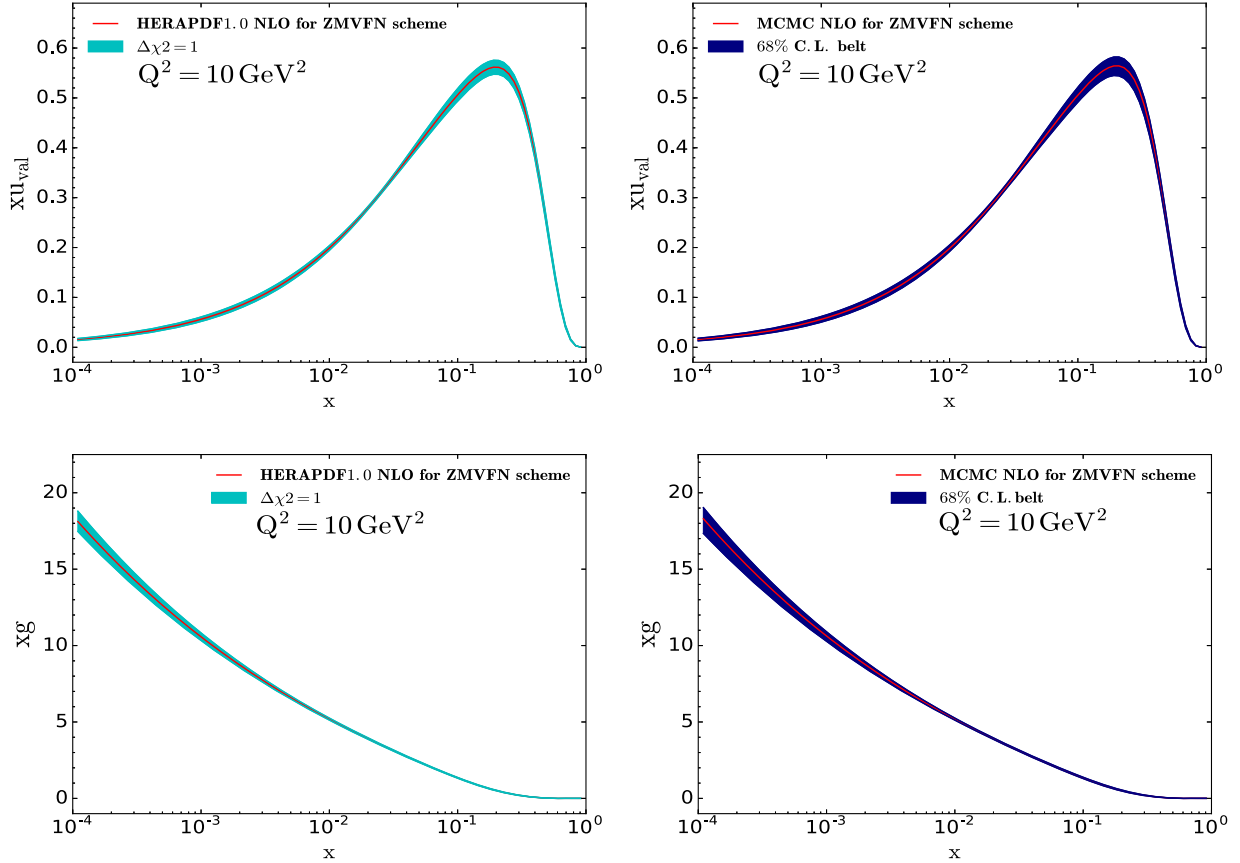


FIG. 7. The parton distribution functions obtained using MCMC (right) compared to HERAPDF1.0 (ZMVFN scheme) from xFitter output (left) for  $xu_{\text{val}}$  and  $xg$ , at  $Q^2 = 10 \text{ GeV}^2$ . The bands show the 68% confidence interval around the central value (in solid red line) for the MCMC PDFs, and the standard  $\Delta\chi^2 = 1$  deviation for HERAPDF.

For each  $(x, Q^2)$ , we determine the  $\alpha\%$ -confidence interval around the best fit value of the PDF (with typically  $\alpha = 68$  or  $\alpha = 95$ ) by considering the region of the distribution on each side of the best fit, and taking  $\alpha\%$  of the data on each of these regions. This provides the (not necessarily symmetric)  $\alpha\%$ -confidence limit intervals we show in Figs. 7 and 9 to 11.

The gluon and  $u_{\text{val}}$  parton distributions obtained this way are plotted as functions of  $x$  and for  $Q^2 = 10 \text{ GeV}^2$  in Fig. 7. The central PDF is the best fit value. These MCMC PDFs are compared with the HERAPDF1.0 PDFs (ZMVFN scheme), with a direct comparison in Fig. 7, and a ratio plot in Fig. 8. They are, as expected, very close, both in central value and in confidence interval. Maximum likelihood estimator and least square method are indeed equivalent under Gaussian assumption, which in our case can be reasonably applied, as mentioned in the previous section.

The uncertainties obtained by the MCMC method and the Hessian method are also consistent within the kinematic range of HERA. This is demonstrated in Fig. 9 where experimental uncertainties—normalized by the best-fit

value—obtained for HERAPDF1.0 NLO and MCMC NLO, respectively, by the Hessian and MCMC methods are compared for the  $u_{\text{val}}$  and the gluon distributions.

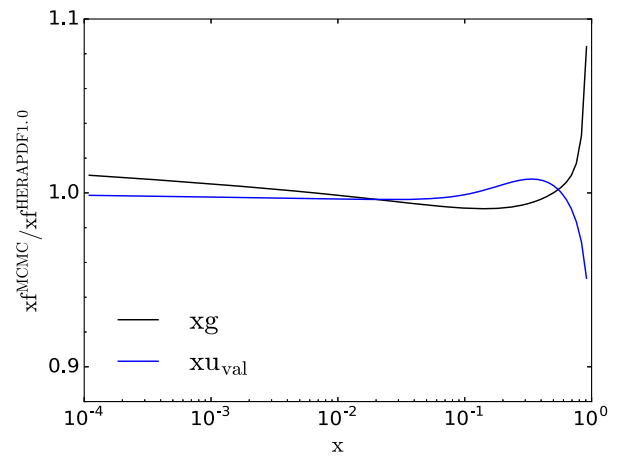


FIG. 8. Ratio of MCMC PDFs and HERAPDF1.0 (ZMVFN scheme) central values for  $xu_{\text{val}}$  and  $xg$  at  $Q^2 = 10 \text{ GeV}^2$ .

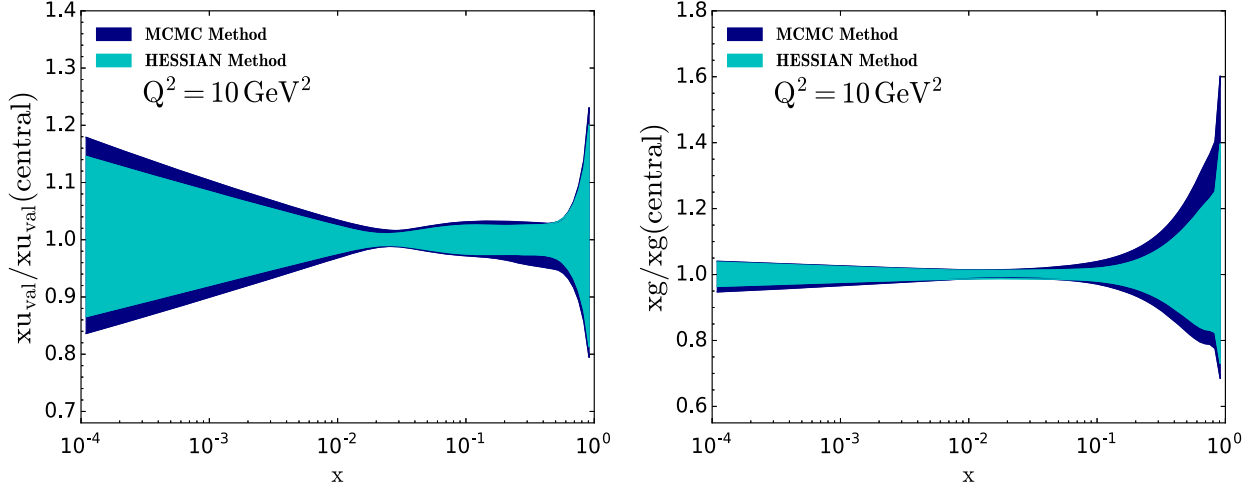


FIG. 9. Comparison of the PDF uncertainties, normalized by the best fit value, as determined by the Hessian and MCMC methods at NLO for the valence distribution  $xu_{\text{val}}$  and the gluon distribution  $xg$ , at a scale  $Q^2 = 10 \text{ GeV}^2$ .

The MCMC uncertainties tend to be slightly larger than the standard deviations obtained in the Hessian approach. For completeness, we also display in Fig. 10 the antiquark PDFs at  $Q^2 = 10 \text{ GeV}^2$ , and in Fig. 11 the central

value and 68% confidence limit interval for  $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ,  $xg$  and  $x\Sigma$  ( $x\Sigma = x\bar{u} + x\bar{d} + x\bar{s} + x\bar{c}$ , with  $x\bar{c} = 0$  for  $Q^2 < m_c^2$ ) on the same plot, at  $Q^2 = 1.9 \text{ GeV}^2$  and  $Q^2 = 10 \text{ GeV}^2$ .

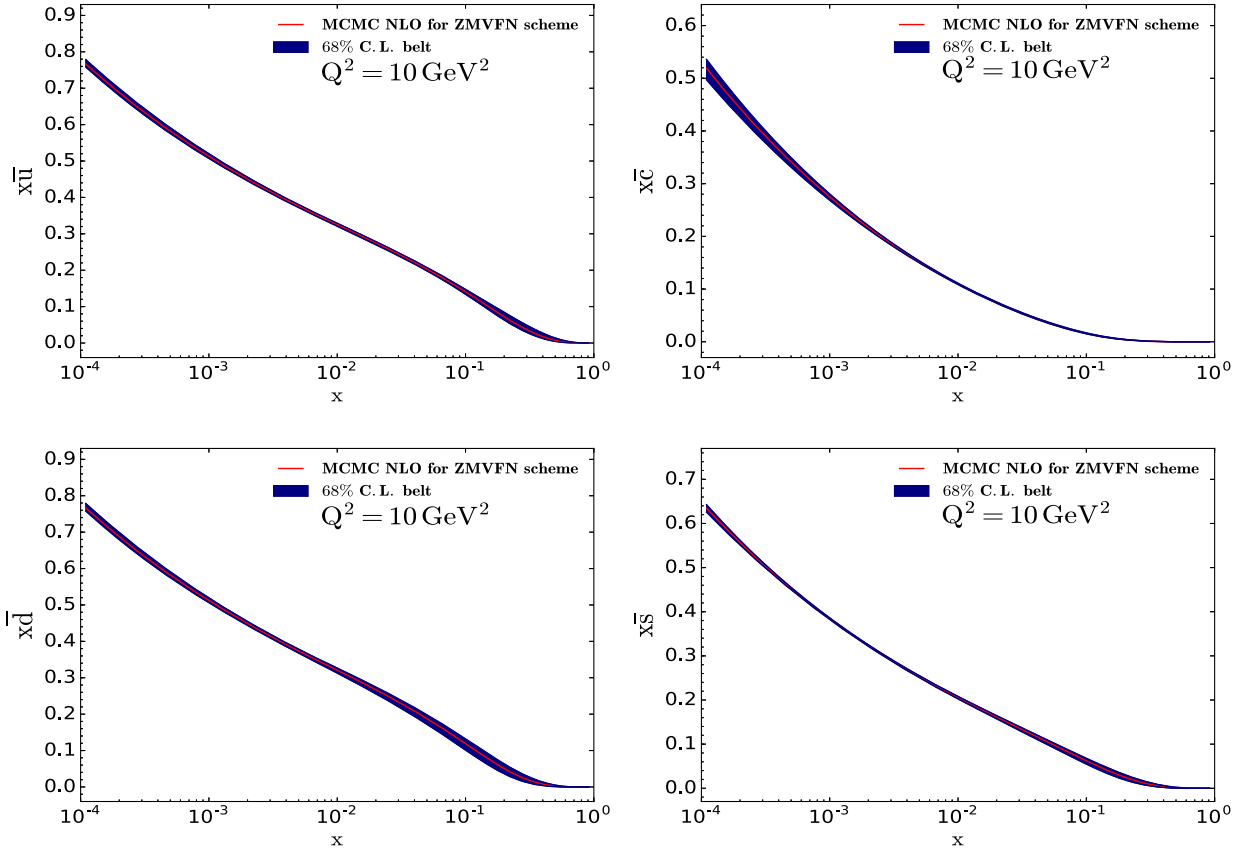


FIG. 10. The MCMC parton distribution functions  $x\bar{u}$ ,  $x\bar{d}$ ,  $x\bar{s}$  and  $x\bar{c}$  at  $Q^2 = 10 \text{ GeV}^2$ . The bands show the 68% confidence interval around the best fit value (in solid red line).

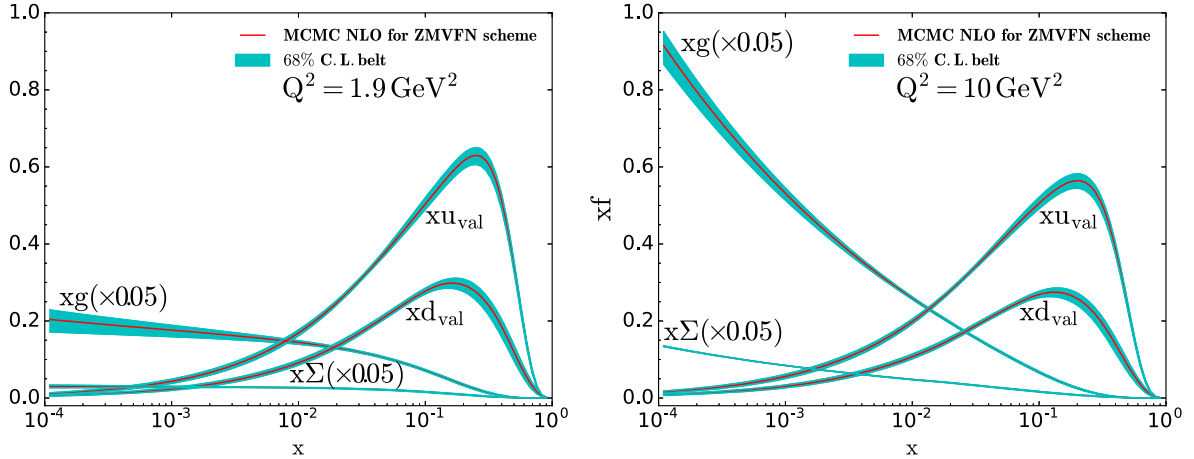


FIG. 11. PDFs of valence quarks ( $xu_{\text{val}}$ ,  $xd_{\text{val}}$ ), sea quarks ( $x\Sigma = x\bar{u} + x\bar{d} + x\bar{s} + x\bar{c}$ , with  $x\bar{c} = 0$  for  $Q^2 < m_c^2$ ), and gluon ( $xg$ ) using MCMC at the scale  $Q^2 = 1.9 \text{ GeV}^2$  (lhs) and  $Q^2 = 10 \text{ GeV}^2$  (rhs) obtained with the ZMVFN scheme. The gluon and the sea distributions are scaled down by a factor of 20. The experimental uncertainties (68% confidence limit as defined in the text from the probability density of PDFs) are represented by the green-shaded region.

## VII. CONCLUSION AND OUTLOOK

We have shown that MCMC, known to be well suited to multiparameter determination, is applicable to PDF determination and that Bayesian parameter inference approach applied to global PDF analysis can lead to a deeper insight into PDF uncertainties. The innovative procedure we implemented, which combines Monte Carlo techniques, lattice-developed algorithms and global PDF analysis is complementary to the existing methods. We have for the first time applied the hybrid Monte Carlo algorithm to PDF determination and computed marginal probability densities of PDF parameters, and the PDFs themselves. This allows us to study the probability distribution of these functions, to determine mean, best fit and median values, and to extract confidence intervals in a statistically controlled way.

This work will lead to an innovative PDF uncertainties determination, and thus to a reliable determination of uncertainties for many collider observables, in a way that is complementary to the existing methods. This will also open new ways to analyze the impact of new data sets being added to the analysis, to check which data set has outliers and if these latter can be tolerated.

This feasibility study paves the way for a more complete PDF determination by MCMC techniques, and our goal is to extend the present work to the full ensemble of PDF free parameters, including also as parameters, the strong coupling constant and  $c$  and  $b$  quark masses. We will consider more complex  $\chi^2$  functions including correlation and complete our analysis on a fully realistic case, studying, in particular, the impact of priors. No doubt that Markov chain Monte Carlo methods will give interesting and valuable information on PDFs and will contribute to our deeper understanding of these key elements of QCD.

## ACKNOWLEDGMENTS

We thank L. Derome and D. Maurin for fruitful discussions and the xFITTER support team for their help. We are also grateful to the CFM Foundation for Research for its financial support through the Jean-Pierre Aguilar 2014 grant. Most of the computation has been done in Lyon-CCIN2P3, thanks to the CPUs resources provided. We are grateful to the staff members of this Computing Center for their constant support.

- 
- [1] X. Ji, *Phys. Rev. Lett.* **110**, 262002 (2013).  
 [2] H. L. Lai, J. Huston, S. Kuhlmann, F. Olness, J. Owens, D. Soper, W. K. Tung, and H. Weerts, *Phys. Rev. D* **55**, 1280 (1997).  
 [3] A. D. Martin, R. G. Roberts, M. G. Ryskin, and W. J. Stirling, *Eur. Phys. J. C* **2**, 287 (1998); A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, *Eur. Phys. J. C* **28**, 455 (2003); **35**, 325 (2004).  
 [4] W. T. Giele and S. Keller, *Phys. Rev. D* **58**, 094023 (1998).  
 [5] J. Huston, E. Kovacs, S. Kuhlmann, H. L. Lai, J. F. Owens, D. Soper, and W. K. Tung, *Phys. Rev. Lett.* **77**, 444 (1996).

- [6] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, *Phys. Rev. D* **65**, 014013 (2001); J. Pumplin, D. R. Stump, and W. K. Tung, *Phys. Rev. D* **65**, 014011 (2001); D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, *Phys. Rev. D* **65**, 014012 (2001).
- [7] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, *Phys. Rev. D* **65**, 014012 (2001).
- [8] W. T. Giele and S. Keller, *Phys. Rev. D* **58**, 094023 (1998); Report No. FERMILAB-PUB-01-498-T.
- [9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo In Practice* (Chapman and Hall/CRC, London, 1995).
- [10] A. D. Sokal, *Monte Carlo Methods in Statistical Mechanics: Foundations and new algorithms* (Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, Switzerland, 1989).
- [11] A. Putze and L. Derome, *Phys. Dark Universe* **5–6**, 29 (2014); A. Putze, L. Derome, D. Maurin, L. Perotto, and R. Taillet, *Astron. Astrophys.* **497**, 991 (2009).
- [12] L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo (NNPDF Collaboration), *J. High Energy Phys.* **03** (2007) 039; J. C. Rojo, PhD thesis, arXiv:hep-ph/0607122.
- [13] N. N. Nikolaev and B. Zakharov, *Z. Phys. C* **49**, 607 (1991); A. H. Mueller, *Nucl. Phys.* **B415**, 373 (1994).
- [14] J. Collins, *Foundations of Perturbative QCD*, vol. 32, Cambridge monographs on particle physics, nuclear physics and cosmology (Cambridge University Press, Cambridge, England, 2011).
- [15] S. M. Aybat and T. C. Rogers, *Phys. Rev. D* **83**, 114042 (2011); M. G. A. Buffing, A. Mukherjee, and P. J. Mulders, *Phys. Rev. D* **86**, 074030 (2012); P. J. Mulders, *Pramana* **72**, 83 (2009); S. Jadach and M. Skrzypek, *Acta Phys. Pol. B* **40**, 2071 (2009); F. Hautmann, *Acta Phys. Pol. B* **40**, 2139 (2009).
- [16] R. M. Neal, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. L. Jones, and X. Meng (Chapman and Hall/CRC, London, 2011).
- [17] S. Alekhin *et al.*, *Eur. Phys. J. C* **75**, 304 (2015); F. D. Aaron *et al.* (H1 and ZEUS Collaboration), *J. High Energy Phys.* **01** (2010) 109; F. D. Aaron *et al.* (H1 Collaboration), *Eur. Phys. J. C* **64**, 561 (2009); M. Botje, *Comput. Phys. Commun.* **182**, 490 (2011); F. James, M. Roos (CERN), *Comput. Phys. Commun.* **10**, 343 (1975); C. Adloff *et al.* (H1 Collaboration), *Eur. Phys. J. C* **30**, 1 (2003); S. Chekanov *et al.* (ZEUS Collaboration), *Phys. Rev. D* **67**, 012007 (2003); F. D. Aaron *et al.*, *Eur. Phys. J. C* **63**, 625 (2009); S. Chekanov *et al.* (ZEUS Collaboration), *Eur. Phys. J. C* **42**, 1 (2005); J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *J. High Energy Phys.* **07** (2002) 012.
- [18] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [19] W. K. Hastings, *Biometrika* **57**, 97 (1970).
- [20] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry—Algorithms and Applications* (Springer-Verlag, Berlin, 2008), 2nd revised edition.
- [21] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216 (1987).
- [22] U. Wolff, *Comput. Phys. Commun.* **156**, 143 (2004).
- [23] M. H. Quenouille, *Biometrika* **43**, 353 (1956).