



# Distribution 3D d'étoiles et de poussière dans le disque de la Voie Lactée

Clément Hottier

## ► To cite this version:

Clément Hottier. Distribution 3D d'étoiles et de poussière dans le disque de la Voie Lactée. Astrophysique [astro-ph]. Université Paris sciences et lettres, 2019. Français. NNT : 2019PSLEO003 . tel-02879238

**HAL Id: tel-02879238**

**<https://theses.hal.science/tel-02879238>**

Submitted on 23 Jun 2020

**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 DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Observatoire de Paris

**Distribution 3D d'étoiles et de poussières dans le disque de  
la Voie lactée**

Soutenue par

**Clément Hottier**

Le 17 Septembre 2019

École doctorale n°127

**AAIF**

Spécialité

**Astronomie et Astrophy-  
sique**

Composition du jury :

Marie-Christine Angonin Professeur des Universités - Paris 7	<i>Présidente du jury</i>
Frédéric Arenou Ingénieur de Recherche - GEPI	<i>Co-Directeur</i>
Carine Babusiaux Astronome - IPAG - GEPI	<i>Directrice</i>
Laurent Cambresy Astronome - Observatoire de Strasbourg	<i>Rapporteur</i>
Isabelle Joncour Maître de conférences - IPAG	<i>Examinatrice</i>
Douglas Marshall Maître de conférences - AIM - IRAP	<i>Rapporteur</i>





# Remerciements

Je tiens à remercier les membres du jury, pour s'être rendus disponibles pour la soutenance et pour les échanges que nous avons eus lors des questions. Je remercie particulièrement Laurent Cambrésy et Douglas Marshall pour avoir rapporté mon manuscrit et pour les nombreux retours et commentaires détaillés qui m'ont permis d'aboutir au présent manuscrit. Je remercie grandement Marie Christine Angonin qui a accepté d'être la présidente du jury de ma soutenance, mais qui m'a aussi énormément aidé à m'orienter pendant mon master. C'est aussi grâce (à cause ?) à toi, Marie-Christine que j'ai eu l'opportunité d'être représentant des étudiants de l'Observatoire de Paris.

Je veux aussi remercier mes directeurs de thèse, ou plutôt ma directrice de thèse, Carine Babusiaux et mon Co-Encadrant de thèse Frédéric Arenou. C'est grâce à votre soutien et à votre présence bienveillante que j'ai pu réussir cette entreprise qu'est la thèse. Je pense notamment à la rédaction du manuscrit à marche forcée qui fut une épreuve autant pour vous que pour moi. Je te remercie aussi, Carine, ainsi que toute ta famille pour l'accueil chaleureux que vous m'avez réservé lors de mes visites à Grenoble. J'attends avec impatience de goûter à la queue de charrue flambée.

Merci aussi à Yveline Lebreton et Mathieu Puech d'avoir été mon comité de suivi de thèse pendant ces trois années, encore désolé pour les doodles à 30 dates particulièrement laborieux à remplir.

Je remercie aussi les résidents du bâtiment hipparque dans lequel j'ai passé ces 3 années de thèse. Je pense notamment à Camilla qui fut ma première colocataire de bureau, qui m'a accueilli dans ce monde de la recherche et qui m'a appris suffisamment d'insultes en italien pour me permettre de conduire à Rome sans être à court. À Nicolas, pour toutes nos conversations et pour tous tes conseils, sans toi les codes seraient encore en train de tourner. À Guillaume, pour nos débats pointus sur des sujets aussi divers que l'utilisation du formatage de string en python et la grande question "Jar-Jar binks est-il un maître sith qui a tout orchestré ?" À David, pour tes conseils et tes réponses (presque aussi nombreuses que tes questions python). À Noël, pour m'avoir expliqué toutes les ficelles de l'Observatoire et pour tous tes conseils et bons plans d'enseignement (et aussi les covoiturages). À Fred, pour toutes nos discussions pop culture, pour les manchots et les pingouins et pour être le seul à remarquer les sérifs dans les figures. À Nathalie, qui m'a supporté pendant la rédaction et la préparation de la soutenance. À Olivier, qui aime les glaces. Et aussi à tous les autres que j'oublie.

Merci aussi aux résidents du bâtiment Copernic pour toutes les discussions à la cantine et les échanges que nous avons eu. Merci plus particulièrement à Meyriem et Leticia de m'avoir accueilli parmi les doctorants du PSG.

Merci à tous les enseignants de l'Observatoire avec qui j'ai eu l'occasion de travailler pendant mon monitorat. Je pense notamment à Caroline Barban et Christian Balanca qui m'ont proposé d'enseigner l'informatique au master de l'Observatoire. Je pense aussi à l'équipe du mois d'août de l'OHP du DU ECUE pour ces semaines mémorables sous les coupes et plus particulièrement à mon compère du T152, Hector Flores, avec qui on ne s'ennuie jamais (même quand on doit redémarrer les deux ordinateurs en attendant au moins une minute entre les deux).

---

Merci aussi à Nathalie Brasseur, tu m'as permis d'être le coach astro de la délégation française pendant ces trois années et tu m'as emmené en Corée pour être mentor. Grâce à toi je sais dire "i sveikata" (entre autres).

Merci à la génération Plouplou, on se connaît depuis bac -3, je vous laisse faire le calcul.

Je remercie mes parents et toute ma famille qui m'ont soutenu pendant toutes mes études et plus particulièrement pendant la thèse et la rédaction du manuscrit.

Je remercie ma femme de m'avoir supporté pendant cette thèse et surtout pendant les derniers mois. Tu t'es occupée d'absolument tout pendant ces quelques mois, tu m'as même suivi (à pied, mais c'est un détail) à travers la Corse pour que je décompresse.

Enfin j'aimerais remercier tous les membres du pipeline de correction orthographique de ce manuscrit. Sans vous, les messieurs Bled et Bescherelle seraient en rotation rapide à force de se retourner dans leurs tombes.

# Table des matières

<b>Remerciements</b>	<b>1</b>
<b>1 Introduction</b>	<b>5</b>
1.1 La Voie lactée . . . . .	6
1.2 Le disque mince . . . . .	7
1.2.1 Structure spirale . . . . .	7
1.2.2 Traceurs de la morphologie du disque galactique . . . . .	9
1.3 Densité d’extinction . . . . .	11
1.4 Un peu de mathématique : l’approche bayésienne . . . . .	12
1.5 Objectif de la Thèse . . . . .	12
1.5.1 Genèse de la méthode . . . . .	12
1.5.2 Approche généralisée . . . . .	14
1.5.3 Étude du disque galactique avec FEDReD . . . . .	14
<b>2 Relevés d’étoiles</b>	<b>17</b>
2.1 Two Micron All Sky Survey (2MASS) . . . . .	17
2.1.1 Description du relevé . . . . .	17
2.1.2 Sélection des données . . . . .	18
2.2 UKIRT Infrared Deep Sky Survey (UKIDSS) . . . . .	18
2.2.1 Description du relevé . . . . .	18
2.2.2 Sélection des données . . . . .	19
2.3 Gaia - Data Release 2 (GDR2) . . . . .	20
2.3.1 Description du relevé . . . . .	20
2.3.2 Sélection des données . . . . .	20
2.4 Complétude . . . . .	21
2.5 Appariement des catalogues (Crossmatch) . . . . .	22
2.6 Photométrie et unité . . . . .	23
<b>3 L’algorithme Field Extinction - Distance Relation Deconvolver</b>	<b>29</b>
3.1 Distribution individuelle . . . . .	29
3.1.1 Diagramme de Hertzsprung-Russel (diagramme HR) de référence . . . . .	29
3.1.2 Loi d’extinction . . . . .	30
3.1.3 Calcul des probabilités $P(O   A_0, D)$ . . . . .	31
3.2 Déconvolution . . . . .	33
3.2.1 Déconvolution de Richardson-Lucy . . . . .	33
3.2.2 Déconvolution avec fonction de sélection . . . . .	34
3.2.3 Critère d’arrêt . . . . .	34
3.2.4 Prior . . . . .	35
3.2.5 Complétudes . . . . .	37
3.3 Calcul de la relation entre $A_0$ et $D$ . . . . .	40

3.4	Implémentation . . . . .	41
3.4.1	Calculs . . . . .	41
3.4.2	Éxecutions et utilisation des grappes . . . . .	42
3.4.3	Stockage et accès . . . . .	43
3.4.4	Temps de calcul et gain . . . . .	44
3.4.5	Évolution du code et ajout de catalogue d'étoiles . . . . .	44
<b>4</b>	<b>Affinage de l'algorithme, Tests et Validations</b>	<b>45</b>
4.1	Simulation de champs de vue . . . . .	45
4.2	Test de FEDReD . . . . .	47
4.2.1	Résultats pour la relation $A_0(D)$ . . . . .	47
4.2.2	Résultats distribution de distance . . . . .	47
4.3	Champs de validation et affinage . . . . .	49
4.3.1	Affinage de la méthode . . . . .	50
4.3.2	Résultats Two Micron All Sky Survey (2MASS) et Gaia - Data Release 2 (GDR2) . . . . .	53
4.3.3	Résultats UKIRT Infrared Deep Sky Survey (UKIDSS) et GDR2 . . . . .	53
<b>5</b>	<b>Disque galactique vu par 2MASS et GDR2</b>	<b>61</b>
5.1	Caractéristique technique . . . . .	61
5.2	Construction de la carte . . . . .	61
5.2.1	Réduction d'une carte d'extinction . . . . .	62
5.2.2	Carte avec solution Monte Carlo (MCS) . . . . .	62
5.3	Carte de densité d'extinction $a_0$ . . . . .	67
5.3.1	Structures connues et inconnues . . . . .	67
5.3.2	Comparaison avec Lallement et al. (2019) . . . . .	70
5.3.3	Comparaison aux autres cartes d'extinction . . . . .	71
5.3.4	Comparaison à d'autres traceurs du milieu interstellaire (MIS) . . . . .	73
5.3.5	Bras spiraux . . . . .	75
5.4	Densité stellaire . . . . .	77
<b>6</b>	<b>Conclusion et Perspectives</b>	<b>83</b>
6.1	Conclusion . . . . .	83
6.2	Perspectives . . . . .	84
	<b>Liste des acronymes</b>	<b>87</b>
	<b>Bibliographie</b>	<b>91</b>

# Chapitre 1

## Introduction

**L**A Voie lactée, cette grande traînée diffuse qui barre le ciel nocturne. Elle a depuis toujours posé question à l'Homme. Elle a d'abord trouvé son origine dans la mythologie. Pour les Grecs, c'est Héra, chassant Héraclès de son sein où Zeus l'avait mis à téter, envoyant le lait jusque dans le ciel. Cela devait être la destinée d'Héraclès que de peupler le ciel, puisqu'au cours de ses travaux il y envoya aussi un dragon et un crabe, tous deux à la solde de sa belle-mère. Enfin il finira lui aussi dans la voûte céleste.

Bien qu'ils aient inventé de belles histoires, les Grecs ont aussi mené les premières études scientifiques du ciel nocturne et de la Voie lactée. Le traité *Du ciel* d'Aristote classe les objets fixes et lumineux du ciel nocturne dans le monde supralunaire. Hipparque, lui, classera les étoiles en six catégories de luminosité, un principe qui, bien que remanié, reste toujours utilisé aujourd'hui.

L'une des percées majeures dans la compréhension de la traînée de la Voie lactée dans le ciel nous vient de l'Orient, pendant l'âge d'or islamique, par l'astronome Nasir al-Din al-Tusi. Il écrit « La Voie lactée, est faite d'un très grand nombre de petites étoiles en amas serrés, de par leurs concentrations et leurs petites ressemblent à des zones nuageuses et ont la couleur du lait ».

Galilée confirma cette description grâce à ses lunettes astronomiques qu'il utilise pour observer les planètes du système solaire, mais aussi la Voie lactée. Il y voit de très nombreuses étoiles dont l'apparence ne varie pas par rapport à celles visibles à l'œil nu.

La seconde moitié du XVIII<sup>ème</sup> voit apparaître les premières hypothèses concernant la structure de la Voie lactée, avec Thomas Wright qui l'imagine comme un disque aplati formé d'étoiles, suivi par Jean-Henri Lambert qui la décrit comme un système irrégulier d'étoiles répartie entre deux plans et lui même constitué de sous systèmes, puis le philosophe Emmanuel Kant qui l'imagine en rotation.

Les premières études quantitatives de la structure de la Voie lactée sont réalisées par William et Caroline Herschel à l'aide des télescopes qu'ils ont eux-mêmes construits. Ainsi, dans sa « Construction des Cieux » (Herschel 1785), William indique les relevés qu'il a effectué avec Caroline dans différentes directions du ciel à l'aide « d'étoiles-jauges », le conduisant à proposer une représentation morphologique de notre galaxie (Figure 1.1). Les deux principales raisons qui expliquent l'inadéquation de cette représentation sont la méconnaissance de la distance des étoiles et de l'extinction interstellaire. Deux siècles après les Herschel, notre travail de thèse montre que les mêmes interrogations continuent à se poser.

En 1838, Friedrich Bessel fut le premier à mesurer la distance d'une étoile par la méthode des parallaxes, ouvrant ainsi la voie pour de nouvelles méthodes d'étude de la Voie lactée.

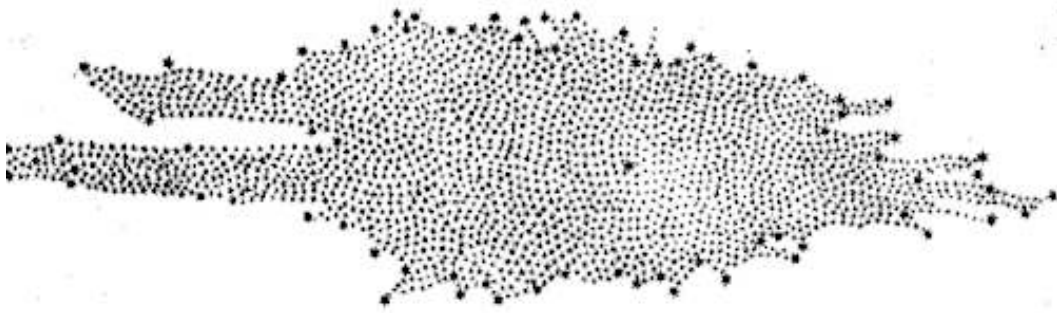


FIGURE 1.1 – Forme de la galaxie dans le plan méridional déduite par Herschel (1785). Le soleil est représenté à proximité du centre.

## 1.1 La Voie lactée

La séquence de Hubble propose une classification morphologique des galaxies. Classer une galaxie dans cette séquence est relativement facile à partir d’une simple image. En revanche, l’étude morphologique de notre galaxie nous est compliquée par notre présence à l’intérieur de cette dernière. On l’observe depuis l’intérieur de son disque, et il nous faut l’observer depuis les deux hémisphères de la planète pour en avoir une vue complète (Figure 1.2).



FIGURE 1.2 – Image recomposée de la Voie lactée à partir de photographies prises depuis les deux hémisphères terrestres (Crédit : Nick Risinger)

Nous savons aujourd’hui que la Voie lactée est une galaxie spirale barrée, la galaxie Messier 83 est un exemple de cette morphologie (Figure 1.3). De par l’étude *in situ* de notre galaxie, nous savons qu’elle possède un halo, la dimension et la définition de ce dernier sont sujets à un grand nombre d’études récentes impulsées par la publication de Gaia - Data Release 2 (GDR2).

On considère actuellement que son disque est séparé en deux composantes, les disques mince et épais. Le disque mince est la composante jeune du disque, tandis que le disque épais contient des étoiles plus anciennes qui pourraient être la population primordiale de la Voie lactée. Le centre galactique est le siège de la réunion des différentes structures de la Voie lactée, qui forment un bulbe dans lequel se trouve une barre.

Cela étant, la connaissance de la morphologie du disque galactique est complexe de par sa projection en 2 dimensions sur la voûte céleste. La dé-projection des structures requiert la connaissance des distances. C’est pourquoi cette connaissance est l’un des enjeux de l’astronomie galac-



FIGURE 1.3 – Image de la galaxie Messier 83 réalisée par le télescope spatial Hubble.

tique moderne. Grâce au satellite Gaia (Gaia Collaboration et al. 2016), on dispose aujourd’hui d’un très grand nombre de parallaxes d’étoiles. Ces parallaxes, associées à d’autres informations telles que la photométrie ou la spectroscopie, nous permettent de localiser des structures dans le disque galactique et même d’en étudier la cinématique et la dynamique. Par ailleurs, il existe d’autres traceurs astronomiques, comme les masers ou les nuages moléculaires que l’on peut utiliser pour étudier la forme du disque.

Dans cette thèse nous avons étudié le disque mince de la Voie lactée, et plus particulièrement sa morphologie en utilisant des relevés photométriques dans le proche infrarouge associés aux parallaxes et à la photométrie GDR2.

## 1.2 Le disque mince

Le disque mince de la Voie lactée regroupe les éléments jeunes de la galaxie. Il contient plus de gaz et de poussière que les autres composantes de la galaxie, ce qui fait de lui le siège d’une plus importante formation d’étoiles. Géographiquement, il s’étend sur quelques centaines de parsecs autour du plan  $Z = 0$ . Bien que toutes les structures de la galaxie se mélangent à cette altitude  $Z$ , ce sont les composants du disque mince qui y sont dominants.

Le disque mince présente aussi une dynamique particulière, elle est à l’origine notamment des bras spiraux de la galaxie. Au centre galactique, on observe aussi la barre, qui est une autre structure dynamique du disque mince.

### 1.2.1 Structure spirale

Notre galaxie est une galaxie spirale barrée. De par notre présence dans le disque de la Voie lactée, il nous est difficile d’observer directement la structure spiralée. Ainsi on ne connaît pas de manière certaine le nombre de bras spiraux ni leurs positions exactes. La Figure 1.4 présente une vue d’artiste de la possible répartition des bras spiraux de la Voie lactée.

Par ailleurs, comme le précise Vallée (2017), dans son analyse bibliographique, la structure



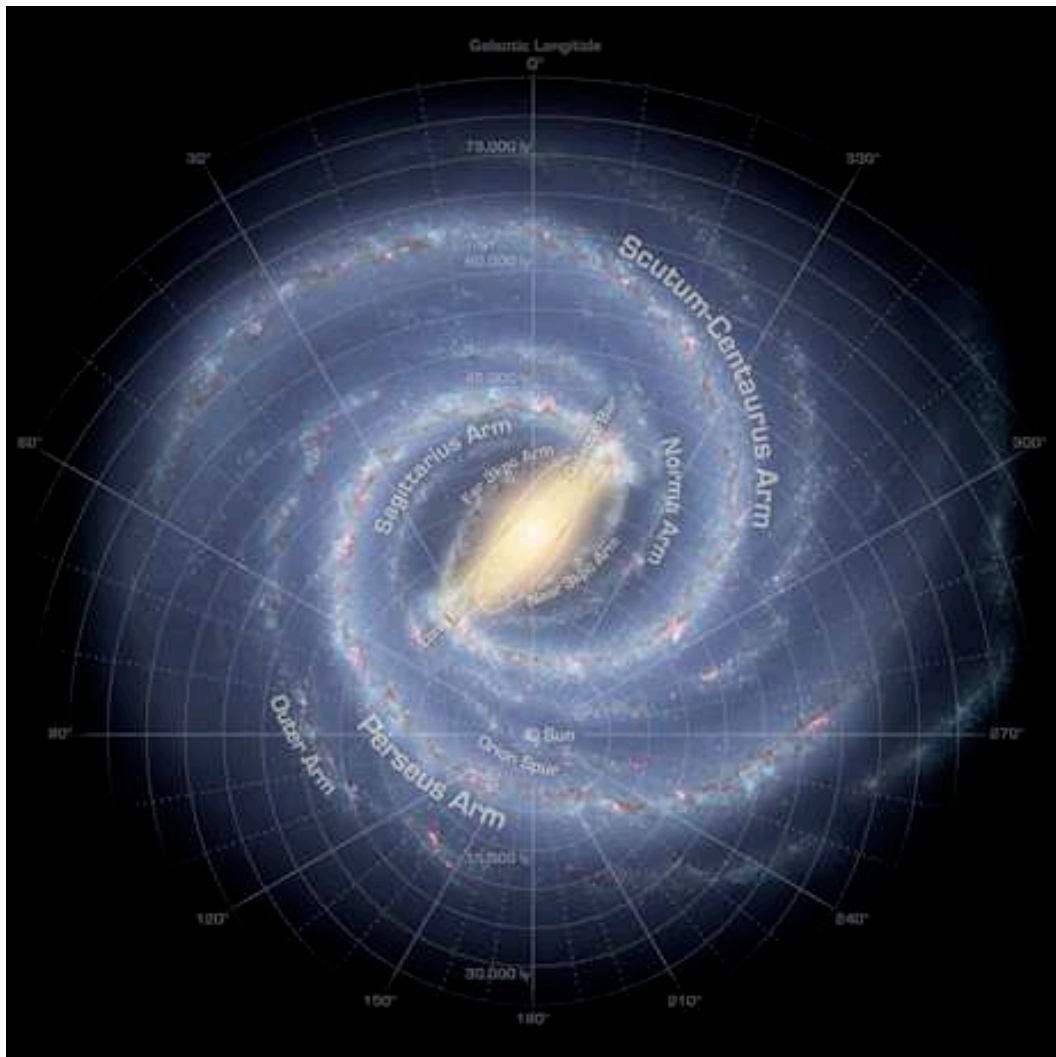


FIGURE 1.4 – Vue d’artiste des bras spiraux de la Voie lactée (Crédit NASA)

spirale de la Voie lactée se compose de plusieurs types de composantes. Le bras spiral est la structure principale, elle prend naissance au niveau du bulbe galactique et s’enroule autour du disque. La branche se sépare d’un bras spiral et le spur est un pont entre deux bras spiraux.

Hou & Han (2014) en étudiant plusieurs traceurs des bras spiraux proposent deux modèles de bras, l’un avec trois bras et l’autre avec quatre. Reid et al. (2014), quant à eux, étudient la position de masers dans le disque galactique pour ajuster un modèle logarithmique à quatre bras spiraux. Ce modèle est repris et réajusté par la suite à l’aide de nouvelles données par Xu et al. (2016) ; Sakai et al. (2019) ; Wu et al. (2019). On peut aussi citer Xu et al. (2016) qui proposent en plus des quatre bras principaux un pont de matière appelé spur.

Dans sa méta-analyse, Vallée (2017) retient que le modèle le plus couramment accepté est un modèle à quatre bras spiraux :

- Le bras de l’Ecu (*Scutum*)
- Le bras du Sagittaire - Carène (*Sagittarius - Carina*)
- Le bras de Persée (*Perseus*)
- Le bras extérieur ou bras de la Règle-Cygne (*Norma - Cygnus*)

Le résultat de cette méta-analyse est présenté en Figure 1.5. Le soleil se trouve dans une structure spirale :

- Le bras local ou bras d’Orion,

Selon les études, cette structure est considérée comme étant un bras (Hou & Han 2014) ou une

branche (Vallée 2017).

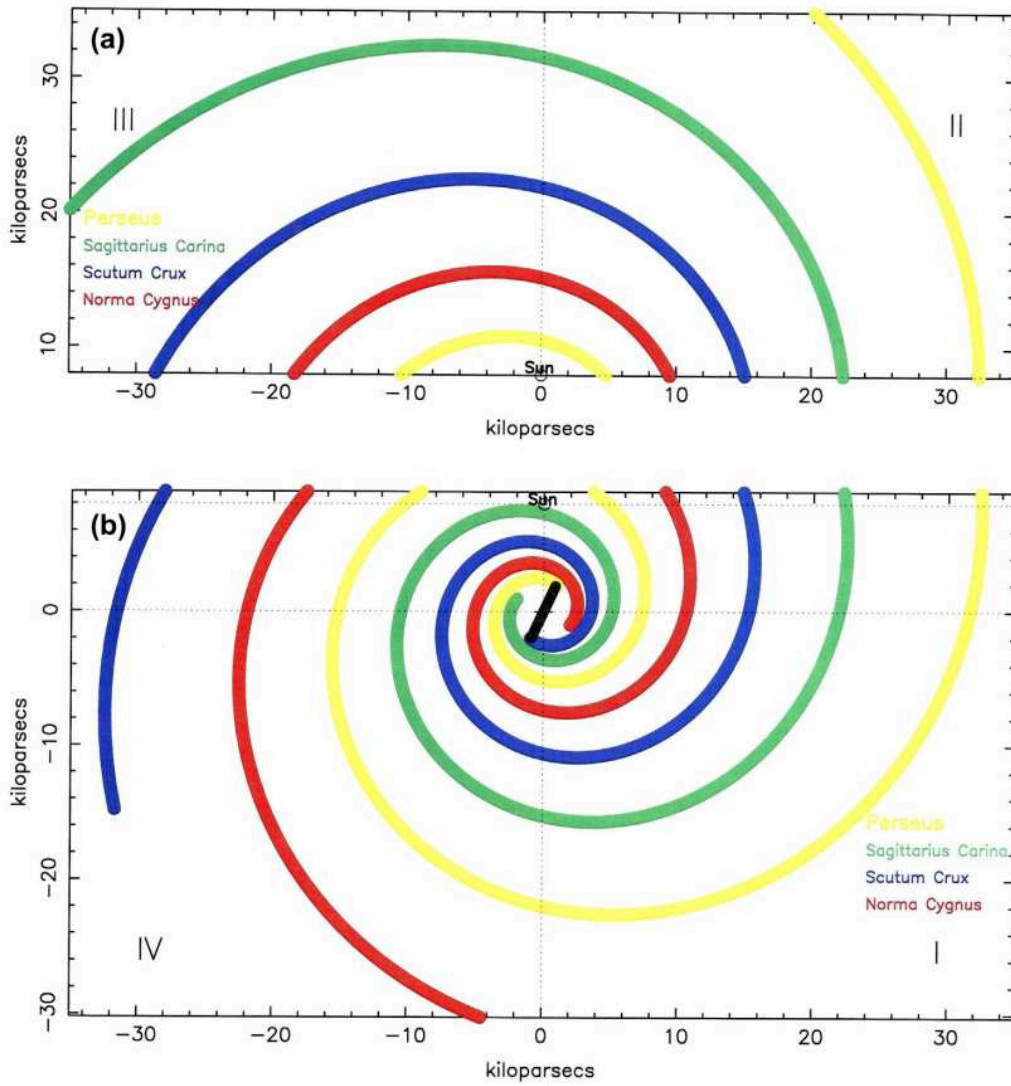


FIGURE 1.5 – Croquis du positionnement de quatre bras spiraux par la méta-analyse Vallée (2017)

## 1.2.2 Traceurs de la morphologie du disque galactique

### Les nuages moléculaires

Le gaz moléculaire représente une petite proportion de la masse baryonique de la Galaxie ( $\sim 10\%$  Miville-Deschênes et al. (2016)). Mais il joue un rôle important dans la formation d'étoiles et dans l'évolution de la galaxie en permettant la formation d'étoiles plus ou moins enrichies en métaux.

Ces nuages sont essentiellement observés à l'aide des émissions du  $^{12}\text{CO}$  dans le domaine radio. En utilisant les informations sur la cinématique des nuages, on peut estimer la distance de ces nuages et obtenir une répartition dans le disque. Cependant, ces distances cinématiques sont très approximatives. Elles supposent en effet une stabilité de la rotation du disque, stabilité qui n'existe pas vraiment. De plus, lors de l'observation des régions centrales de la galaxie, il existe une ambiguïté sur cette distance cinématique puisqu'elle ne permet d'obtenir que le rayon galacto-centrique, on ne peut donc pas déterminer si la structure est en avant plan ou en arrière plan.

## Les zones de formation d'étoiles et région Hii

Les zones de formations d'étoiles sont, comme leurs noms l'indiquent, riches en étoiles jeunes ; notamment des étoiles de types O, qui sont massives et brillantes et dont la concentration forme une composante des bras spiraux (Chen et al. 2019a). Ces zones de formation d'étoiles se trouvent généralement dans des régions d'hydrogène ionisé (région HII) (Hou & Han 2014). Or si la formation d'étoiles est importante, c'est qu'il y a une forte concentration de gaz, donc une forte densité d'extinction.

On peut aussi sonder les jeunes étoiles du disque galactique à l'aide des amas ouverts. Toutes les étoiles appartenant à un amas ayant environ le même âge et la même métallicité, on peut déterminer leurs âges avec des isochrones (Kharchenko et al. 2013). Il est donc possible de sélectionner uniquement les jeunes amas (Cantat-Gaudin et al. 2018).

Grâce au catalogue GDR2, on a maintenant accès aux parallaxes de ces étoiles jeunes nous permettant ainsi de les localiser dans le disque galactique, pour en étudier la distribution (Zari et al. 2018).

Il est aussi possible d'étudier la distribution des étoiles jeunes par les céphéïdes. Ces étoiles étant des chandelles standards on peut déterminer leurs distances sans avoir recours aux parallaxes (Skowron et al. 2018). On peut cependant associer la détermination de la distance par l'étude des pulsations au mouvement propre fourni par les relevés Gaia pour étudier la morphologie du disque mince (Baba et al. 2018).

## Les sources masers

Les masers, pour microwave amplification of stimulated emission of radiation, sont des sources qui sont visibles grâce à des télescopes radio. Elles sont générées par différentes sources astrophysiques notamment les régions de formation d'étoiles (Fish et al. 2005). En utilisant la méthode de very long base interferometry (VLBI), on est capable de mesurer très précisément leurs parallaxes et donc leurs localisations dans la galaxie.

Ces masers étant le siège d'une forte formation d'étoiles, ils peuvent être utilisés comme traceurs de bras spiral (Reid et al. 2014). Par ailleurs, ils sont, comme les régions HII, associés à des régions de fortes extinction.

Ces objets nous apportent donc de l'information précise sur la structure du disque galactique, mais ils sont peu nombreux et ne permettent donc pas une cartographie complète.

## Les étoiles du Red Clump

Les étoiles du Red Clump sont des étoiles de masse intermédiaire, de la branche horizontale des géantes. Elles sont utilisées comme chandelles standards, car elles ont une faible variabilité de magnitude absolue et de couleur (Ruiz-Dern et al. 2018), leur magnitudes et leurs couleurs apparentes peuvent donc nous renseigner sur l'extinction interstellaire ainsi que la distance de l'étoile (voir section 1.5).

## La poussière

La poussière interstellaire est un composant du milieu interstellaire, mais n'en représente qu'une petite partie (Hildebrand 1983). Bien que présente en faible quantité, elle est responsable du rougissement et de l'extinction interstellaire comme l'a démontré Trumpler (1930). C'est la principale responsable des zones sombres sur le disque galactique projetées sur la voûte céleste (Figure 1.2).

Ces poussières sont liées aux nuages de gaz moléculaire. De par leurs absorptions du rayonnement électromagnétique, elles participent au refroidissement des nuages et favorisent ainsi la formation d'étoiles.

La connaissance de la distribution de la poussière revêt plusieurs intérêts astrophysiques majeurs. Elle permet tout d'abord aux observateurs d'étudier des objets dans la direction du plan galactique (ou la densité de poussière est la plus importante) et de corriger les observations par la colonne de poussière traversée par la lumière. Son étude est aussi intrinsèquement intéressante car les grandes structures du disque galactique ont des composantes de poussières.

Ce traceur est celui que nous allons étudier dans cette thèse, au travers de l'extinction interstellaire dont il est responsable.

## 1.3 Densité d'extinction

L'extinction interstellaire du disque galactique a été cartographiée à l'aide de plusieurs catalogues associés à plusieurs méthodologies. L'étude de la distribution d'extinction est un sujet relativement récent surtout en ce qui concerne la distribution en 3D.

Pour la préparation du catalogue Hipparcos, une cartographie a été réalisée par Arenou et al. (1992). Ils utilisent le rougissement des étoiles, obtenu en comparant type spectral et photométrie, pour estimer la colonne d'extinction  $A_V$ .

Drimmel et al. (2003) construisent un modèle de distribution de la poussière dans la galaxie avec bras spiraux dans lequel ils intègrent le long de ligne de visée pour obtenir l'extinction.

Marshall et al. (2006) créent une carte de densité d'extinction en comparant le modèle de population galactique de Besançon (Robin et al. 2003) aux données du catalogue Two Micron All Sky Survey (2MASS) vers le centre galactique ( $|l| < 90^\circ$ ) dans le plan. La différence de couleur entre la prédiction du modèle et les données, pour un champ de vue donné, est due à l'extinction interstellaire. Ils peuvent donc obtenir une relation entre la distance et l'extinction et en déduire la densité d'extinction le long de la ligne de visée.

Chen et al. (2013) et Schultheis et al. (2014) utilisent une approche similaire, mais y ajoutent des données venant des catalogues VVV et GLIMPSE-II. Ces relevés étant plus profonds (mais pas sur le ciel complet) les résultats entre les différentes études diffèrent à longues distances.

Sale (2012) ; Sale et al. (2014) ont développé un modèle bayésien hiérarchique pour analyser le relevé IPHAS. En raison de multiples niveaux de chaînes de Markov, ils doivent utiliser des hyper priors pour calculer la densité d'extinction.

Les travaux de Green et al. (2014, 2015, 2018) se basent sur le relevé photométrique Pan-STARRS associé au catalogue 2MASS. Ils utilisent une approche bayésienne en deux temps pour étudier la relation entre la distance et l'extinction au sein d'une ligne de visée. Tout d'abord, ils étudient les étoiles séparément en cherchant leurs magnitudes absolues et leurs couleurs. Pour ce faire, ils utilisent un certain nombre de priors. Une fois les résultats individuels obtenus, ils utilisent une méthode de chaîne de Markov Monte-Carlo pour trouver les meilleurs paramètres d'évolution de l'extinction avec la distance. De par l'utilisation des données Pan-STARRS leurs résultats couvrent la quasi-totalité des trois premiers quadrants du disque galactique. Ils ont choisi une grille large en module de distance ce qui donne une forte pixellisation. Green et al. (2019) ajoutent à la méthode précédente les parallaxes du catalogue GDR2 et un pas adaptatif en module de distance. Ils utilisent aussi un prior de corrélation pour prendre en compte les résultats des champs de vue voisins. Cette dernière carte s'étend jusqu'à 4kpc du soleil entre les longitudes  $0^\circ \leq l \leq 250^\circ$  pour le plan galactique.

Lallement et al. (2014) utilisent des rougissements d'étoiles obtenus par des observations au sol est une méthode d'inversion bayésienne globale (Vergely et al. 2010) pour reconstruire en trois dimensions la densité d'extinction du voisinage solaire. Capitanio et al. (2017) améliorent

ces résultats en ajoutant les parallaxes du catalogue Gaia - Data Release 1 et des rougissemments déterminés avec APOGEE, en utilisant un prior construit avec Pan-STARRS. Lallement et al. (2019) utilisent la même méthode d'inversion bayésienne sur des rougissemments obtenus avec le catalogue 2MASS et les parallaxes et la photométrie du catalogue GDR2. Cette amélioration des catalogues permet d'obtenir l'extinction jusqu'à une distance de 3 kpc du soleil.

Rezaei Kh. et al. (2017, 2018) utilisent une méthode non paramétrique avec champs gaussiens pour analyser les rougissemments fournis par le relevé APOGEE et obtenir une cartographie de la poussière en 3 dimensions.

Les catalogues 2MASS et GDR2 ont aussi été utilisés par Chen et al. (2019b) pour produire des cartes de densité d'extinction. Pour ce faire, ils utilisent un algorithme d'apprentissage (random forest regression) pour déduire le rougissemment des étoiles et l'évolution de l'extinction en fonction de la distance.

## 1.4 Un peu de mathématique : l'approche bayésienne

Plusieurs des résultats ci-dessus, ainsi que la méthode utilisée dans ce manuscrit se basent sur l'inférence bayésienne. Dans cette approche, on interprète les probabilités comme un degré de certitude plutôt que comme une fréquence. On va donc modéliser un problème à l'aide de paramètres dont on va calculer les probabilités à l'aide d'observations.

On peut ainsi résumer l'approche bayésienne par le théorème de Bayes, publié *post-mortem* dans l'article *An Essay towards solving a Problem in the Doctrine of Chance*

$$P(\theta | O) = \frac{P(O | \theta) P(\theta)}{P(O)}. \quad (1.1)$$

Dans cette équation,  $\theta$  représente les paramètres du modèle tandis que  $O$  correspond aux observations. Afin d'inférer les meilleurs paramètres, on recherche ceux dont la probabilité (interprétée comme la plausibilité) est la plus grande au vu des observations réalisées. On cherche donc à maximiser la probabilité des paramètres sachant les observations, c'est la probabilité ***a posteriori*** :  $P(\theta | O)$ .

Pour calculer cette probabilité ***a posteriori***, on calcule la ***vraisemblance*** de notre modèle :  $P(O | \theta)$ , c'est-à-dire la probabilité d'obtenir les observations  $O$  si le modèle  $\theta$  est correct. Enfin cette vraisemblance est modulée par la connaissance que l'on a ***a priori*** sur le modèle ( $P(\theta)$ ) et sur les observations ( $P(O)$ ), c'est-à-dire les connaissances que l'on pense avoir avant de confronter modèle et observation.

## 1.5 Objectif de la Thèse

L'objectif des travaux relatés dans ce manuscrit est d'étudier la distribution de l'extinction dans le disque galactique, mais aussi la distribution des étoiles en tenant compte de la complétude. L'étude conjointe de ces deux composantes de la galaxie permettra de mieux comprendre la relation entre la poussière et les étoiles et d'appréhender les différences morphologiques de chaque structure.

### 1.5.1 Genèse de la méthode

La magnitude apparente ( $m$ ) d'une étoile varie en fonction de 3 paramètres : la luminosité intrinsèque (exprimée en termes de magnitude absolue  $M$ ), la distance ( $D$ ) à laquelle elle se trouve et enfin l'extinction interstellaire ( $A$ ) entre cette étoile et l'observateur :

$$m = M + 5 \log(D) - 5 + A. \quad (1.2)$$

La connaissance de la magnitude absolue ne lève donc pas la dégénérescence entre la distance et l'extinction.

L'effet de l'extinction est une fonction de la bande photométrique considérée. Nous prendrons  $A_0$ , l'extinction à la longueur d'onde 550nm, comme extinction de référence. À partir de  $A_0$ , on obtient l'extinction dans une bande  $\lambda$  quelconque à l'aide du coefficient d'extinction  $k_\lambda$  :

$$A_\lambda = k_\lambda \times A_0. \quad (1.3)$$

Si on combine la photométrie dans plusieurs bandes photométriques, on peut obtenir de l'information sur l'extinction, et si l'on connaît la magnitude absolue on obtient la distance de l'étoile. Par exemple, Babusiaux & Gilmore (2005) décrivent une méthode de combinaison de deux bandes photométriques pour obtenir une magnitude indépendante de l'extinction. Dans le reste de ce manuscrit, nous combinerons les bandes  $J$  et  $K$ , pour définir  $K_{J-K}$  :

$$K_{J-K} = m_K - \frac{k_K}{k_J - k_K} \times (m_J - m_K). \quad (1.4)$$

En effet, notons  $m_0$  la magnitude apparente sans extinction, tel que  $m = m_0 + A$ , on peut alors écrire :

$$K_{J-K} = m_{0K} + A_K - \frac{k_K}{k_J - k_K} \times (m_{0J} - m_{0K}) - A_K, \quad (1.5)$$

$$K_{J-K} = m_{0K} - \frac{k_K}{k_J - k_K} \times (m_{0J} - m_{0K}). \quad (1.6)$$

À partir de magnitudes apparentes subissant l'extinction interstellaire, on est donc à même d'obtenir une magnitude apparente  $K_{J-K}$  indépendante de l'extinction. Si on connaît les magnitudes absolues d'une étoile dans les bandes  $J$  et  $K$ , on peut calculer la magnitude absolue  $M_{KJK}$  et déduire la distance de cette étoile (par l'équation (1.2)).

D'autre part, en utilisant la différence de deux magnitudes (aussi appelée couleur), on peut obtenir l'extinction  $A_0$  affectant une étoile. Par exemple la couleur  $J - K$ , définie par

$$J - K = m_J - m_K = M_J - M_K + A_0 \times (k_J - k_K), \quad (1.7)$$

est uniquement fonction de l'extinction, si l'on connaît les magnitudes absolues et les coefficients d'absorptions.

Si, pour une population d'étoiles donnée, on connaît la magnitude absolue dans deux bandes photométriques, et les coefficients d'extinction de ces mêmes bandes, le diagramme couleur magnitude (CMD) donnera l'évolution de l'extinction avec la distance.

On a représenté sur la Figure 1.6 un tel diagramme. Chaque point correspond à une étoile, l'axe primaire des abscisses (en bas) représente la couleur  $J - K$  tandis que l'axe primaire des ordonnées (à gauche) représente la magnitude  $K_{J-K}$ . Les axes secondaires représentent quant à eux respectivement l'extinction et la distance correspondant à des étoiles du Red Clump. On utilise les magnitudes absolues déduites par Ruiz-Dern et al. (2018) et les coefficients d'extinctions sont calculés avec le modèle de Danielski et al. (2018).

Sur cette figure, on voit la traînée du Red Clump s'incurver. La courbe ainsi décrite correspond à l'évolution de l'extinction avec la distance. Ainsi, pour des étoiles appartenant au Red Clump, on peut directement lire sur le diagramme une estimation de l'extinction en fonction de la distance.

Cette approche a donné naissance à l'algorithme DTRC, développé par Fourtune-Ravard (2016). Cette méthode étudie, pour des champs de vue séparés, les étoiles d'un CMD pour rechercher la trace du Red Clump et en déduire la relation entre la distance et l'extinction  $A_0(D)$ .

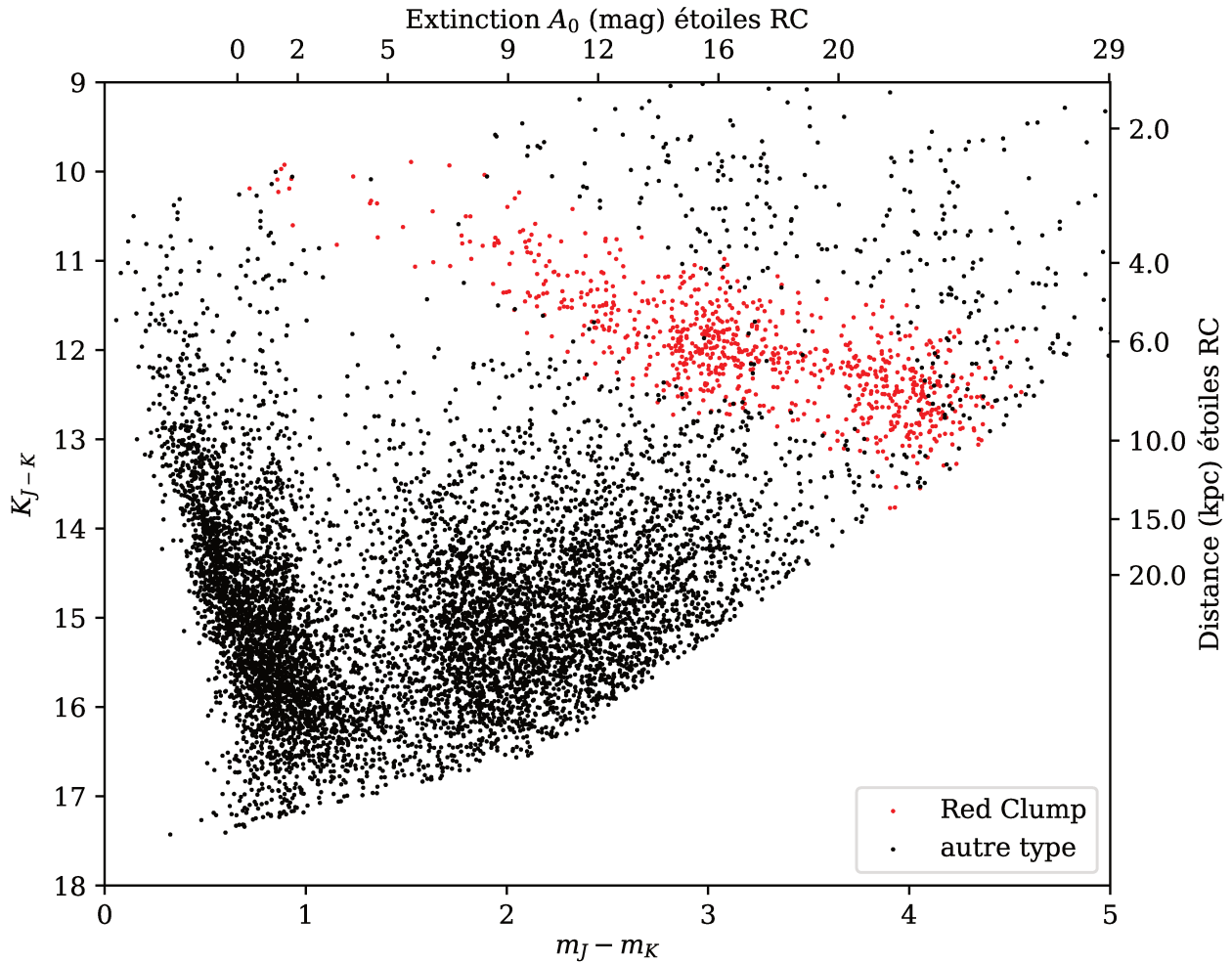


FIGURE 1.6 – CMD d’un champ de vue simulé (voir section 4.1) à  $l = 10^\circ$ . Les axes secondaires sont gradués en distance et en extinction correspondant à des étoiles du Red Clump. Les points rouges correspondent aux étoiles du Red Clump.

### 1.5.2 Approche généralisée

L’utilisation de plusieurs bandes photométriques de l’infrarouge permet ainsi une levée de dégénérescence entre l’extinction et la distance dans la photométrie des étoiles. Fourtune-Ravard (2016) a donc dans un second temps mis en place une approche bayésienne prenant en compte l’ensemble des étoiles observées d’un champ (et non plus uniquement les étoiles du Red Clump). Fourtune-Ravard (2016) propose ainsi la base de l’algorithme Field Extinction - Distance Relation Deconvolver (FEDReD) (voir chapitre 3), elle démontre la faisabilité de l’utilisation de la photométrie UKIRT Infrared Deep Sky Survey (UKIDSS) pour obtenir la relation entre la distance et l’extinction d’un champ de vue. Elle teste l’algorithme sur une simulation (CMD en Figure 1.6) et des champs de vue tests en utilisant comme prior les résultats de la méthode DTRC.

### 1.5.3 Étude du disque galactique avec FEDReD

Le but de cette thèse est d’utiliser FEDReD pour cartographier l’extinction et la distribution d’étoiles dans le disque galactique. Après une étude plus approfondie de l’algorithme, nous avons constaté que l’implémentation de Fourtune-Ravard (2016) de l’algorithme fournissait des résultats dépendants fortement des priors utilisés. Nous avons donc dû apporter un certain nombre



d'optimisations pour ne plus être dépendant de priors, en améliorant la convergence.

Par ailleurs, nous avons profité de l'approche généraliste de FEDReD pour incorporer des données venant du catalogue 2MASS et surtout l'apport des parallaxes et de la photométrie fournies par le satellite Gaia et plus particulièrement par le catalogue GDR2. Ces catalogues et leurs utilisations sont présentés au chapitre 2.

L'étude du disque galactique dans son ensemble implique l'utilisation de FEDReD sur un très grand nombre de champs de vue, or FEDReD avait été testé sur un petit échantillon. Un grand nombre d'optimisations tant algorithmiques que logicielles ont été nécessaires pour obtenir des résultats satisfaisants. Nous avons augmenté le nombre de champs de vue de test afin d'être confronté à tous les problèmes de convergences rencontrés. De plus, nous avons dû fortement optimiser la partie logicielle de FEDReD tant au niveau de la parallélisation que de l'accès disque, pour pouvoir l'utiliser efficacement sur des grappes de calculs. La description de l'algorithme et des optimisations logicielles est présentée au chapitre 3, les optimisations de l'algorithme, l'amélioration de la convergence ainsi que les tests et les validations sont décrits au chapitre 4.

Une fois l'algorithme satisfaisant, nous l'avons appliqué sur des champs de vue répartis sur l'ensemble du disque de la Voie lactée. Cela nous a ainsi fourni des relations extinction- distance à  $360^\circ$  autour du soleil. Pour les assembler, nous avons développé une série de traitements prenant en compte les corrélations angulaires, jusque là ignorées. Nous avons ainsi obtenu une carte d'extinction de l'ensemble du plan galactique. Les limites en distance de notre carte varient selon la direction d'observation, se limitant à 4kpc-5kpc dans la direction du centre galactique et dépassent plus de 6kpc dans les directions moins encombrées. Les étapes d'assemblage des lignes de visée et la carte d'extinction produite sont présentées au chapitre 5.

Concernant l'étude de la distribution des étoiles dans le disque galactique, nous avons démontré sa faisabilité sur une simulation, mais quelques affinages sont encore nécessaires, notamment sur la détermination des incertitudes, pour pouvoir l'utiliser à grande échelle.





# Chapitre 2

## Relevés d'étoiles

**L**E disque de la Voie lactée est riche en poussières, surtout vers le centre galactique. Cela implique une forte extinction interstellaire aux faibles latitudes galactiques. Cette extinction impacte fortement les longueurs d'onde dans le visible et beaucoup moins la lumière infrarouge. C'est pourquoi nous avons choisi d'analyser des relevés photométriques dans le proche infrarouge pour étudier la morphologie du disque. On utilise ainsi les catalogues Two Micron All Sky Survey (2MASS) et UKIRT Infrared Deep Sky Survey (UKIDSS). Associées à la photométrie en proche infrarouge, nous utilisons les parallaxes et la photométrie issues du catalogue GDR2.

L'utilisation de plusieurs catalogues nous permet de composer avec les différentes limites de sensibilité et de saturation. Ainsi le catalogue UKIDSS observe des étoiles à plus faible luminosité, en contrepartie les étoiles brillantes (et donc proches) ne sont pas analysées. On utilise donc le catalogue 2MASS pour ces étoiles plus brillantes et ainsi explorer un plus grand intervalle de distance.

### 2.1 Two Micron All Sky Survey (2MASS)

#### 2.1.1 Description du relevé

Le relevé du ciel 2MASS (Skrutskie et al. 2006) est un relevé photométrique dans le proche infrarouge. Les observations ont été réalisées au sol depuis deux télescopes de 1.3m de diamètre entre juin 1997 et février 2001. Le premier télescope se trouvait sur le mont Hopkins en Arizona et le second à Cerro Tololo au Chili. L'emplacement des télescopes sur les deux hémisphères du globe permet une couverture presque totale de la voûte céleste (99.998%).

Les observations ont été réalisées dans 3 bandes photométriques  $J$  ( $1.25\mu\text{m}$ ),  $H$  ( $1.65\mu\text{m}$ ) et  $K_s$  ( $2.16\mu\text{m}$ ). Les distributions des magnitudes sont présentées en Figure 2.1. On y remarque un deuxième pic dans les distributions des magnitudes  $H$  et  $K$ . Ces pics sont dûs à la présence de la barre de la galaxie conjuguée à la baisse drastique en complétude que l'encombrement implique dans les bandes  $H$  et  $K$ . Ce relevé fournit 2 catalogues distincts,

- *Point source catalogue* : photométrie de 470 992 970 sources ponctuelles
- *Extended source catalogue* : photométrie de 1 647 599 sources étendues

ainsi que les 4 121 439 image FITS captées.

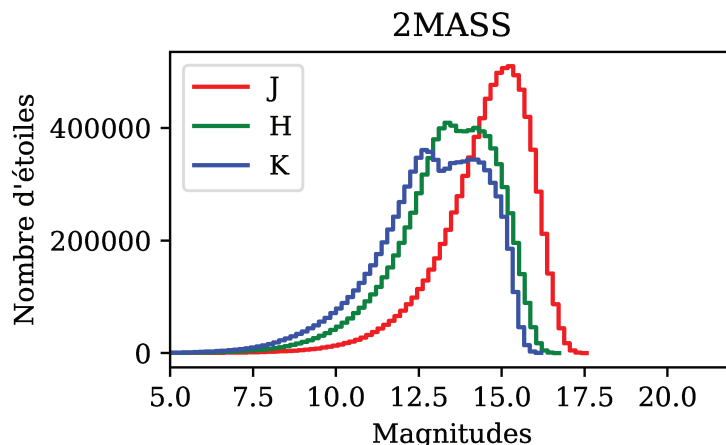


FIGURE 2.1 – Histogrammes des distributions des magnitudes observées par 2MASS

### 2.1.2 Sélection des données

Dans le cadre de cette thèse, nous avons utilisé 2MASS *Point Source catalogue*. Nous avons téléchargé les données depuis le *Gaia archive*<sup>1</sup> ainsi que depuis son miroir hébergé à l'Observatoire de Paris<sup>2</sup>. Cela nous a permis de faire les requêtes sur le crossmatch avec GDR2 directement en ADQL (voir section 2.3.2).

Pour chaque étoile, le catalogue fournit la photométrie et l'erreur de mesure sur la photométrie, exprimée en magnitude. Par ailleurs le catalogue fournit aussi plusieurs indicateurs de fiabilité des mesures. Nous utilisons l'indicateur `ph_qual` qui indique la qualité photométrique pour chaque bande, en attribuant une lettre. Les lettres X, U, correspondent à une absence de données utilisables dans la bande considérée, F et E correspondent à une absence de convergence du pipeline. Les valeurs A, B, C et D correspondent à des détections valides, avec différentes coupures sur les erreurs de mesures.

Nous ne sélectionnons que des étoiles ayant des qualités A, B, C ou D pour chacune des trois bandes photométriques. C'est sur cet échantillon que nous évaluons la magnitude limite pour les champs de vue que l'on analyse. Dans un second temps, nous ne conservons que les étoiles dont l'erreur photométrique est inférieure à  $\sigma_{\max} = 0.3\text{mag}$ . Seules ces étoiles seront analysées par FEDReD. Cette coupure n'est pas utile pour 2MASS car elle ne concerne qu'une infime partie des étoiles. Elle a été mise en place pour le relevé UKIDSS et est appliquée systématiquement.

## 2.2 UKIRT Infrared Deep Sky Survey (UKIDSS)

### 2.2.1 Description du relevé

Le relevé UKIDSS (Lawrence et al. 2007) est un relevé photométrique dans le proche infrarouge. Les observations sont réalisées au sol, depuis le mont Mauna Kea, Hawaï, avec le United Kingdom Infra-Red Telescope (UKIRT) et l'instrument Wide Field Camera (WFCAM). Cette caméra observe dans 5 bandes photométriques de l'infrarouge : Z, Y, J, H et K. Le relevé fournit 5 catalogues :

- *Large Area Survey* : cartographie du ciel boréal
- *Galactic Plan Survey* : photométrie stellaire du disque de la Voie lactée

1. <https://gea.esac.esa.int/archive/>

2. <https://gaia.obspm.fr/tap-server/tap>

- *Galactic Cluster Survey* : photométrie de 10 grands amas ouverts de la Voie lactée
- *Deep Extragalactic Survey* : photométrie de galaxie à redshift 1 – 2
- *Ultra Deep Survey* : cartographie de l'univers à  $2 < z < 4$

Dans les travaux décrits dans ce manuscrit, nous utilisons le catalogue *Galactic Plane Survey* (GPS, Lucas et al. 2008), plus précisément la dixième publication des données (DR10). Étant donné que les observations ne sont réalisées que dans un seul observatoire, seule une partie du disque a été observée :

$$15^\circ < l < 107^\circ ; \quad |b| < 5^\circ, \quad (2.1)$$

$$141^\circ < l < 250^\circ ; \quad |b| < 5^\circ, \quad (2.2)$$

$$-2^\circ < l < 15^\circ ; \quad |b| < 2^\circ. \quad (2.3)$$

Les distributions des magnitudes observées par UKIDSS sont présentées sur la Figure 2.2. On constate que les maxima des distributions sont supérieurs à ceux de 2MASS, cela nous permet d'accéder à des étoiles plus lointaines.

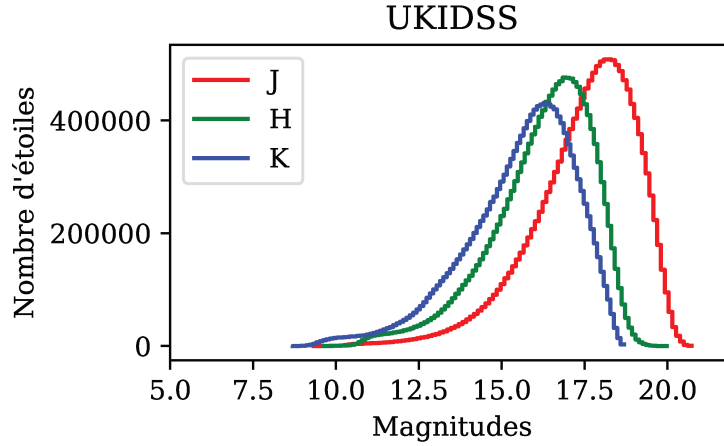


FIGURE 2.2 – Histogrammes des distributions des magnitudes observées pour UKIDSS

### 2.2.2 Sélection des données

Pour utiliser les données UKIDSS, nous avons utilisé le *WFCAM Science Archive*<sup>3</sup>. Pour sélectionner les étoiles à analyser, nous avons tout d'abord appliqué les filtres recommandés par Lucas et al. (2008), pour éviter le dédoublement des sources dû au recouvrement des prises de vue. Puis on ne conserve que les étoiles dont le `ppErrBit` est inférieur à 131072, pour éviter les erreurs de calibration. Le jeu de données ainsi obtenu est utilisé pour déterminer le paramètre de complétude  $m_{\text{lim}}$ .

Il nous faut aussi corriger les erreurs sur la mesure de la photométrie. En effet, d'après Lucas et al. (2008), les incertitudes photométriques d'une source sont calculées à l'aide :

- du nombre de sources dans le champ de vue
- de la largeur de l'image de la source
- de la déviation standard du flux d'arrière-plan

Or à cause de la capacitance inter pixels de WFCAM, il existe une corrélation lors du comptage de photons par pixel. Cela entraîne une sous-estimation de la brillance de l'arrière plan, chaque incertitude photométrique est ainsi sous-estimée d'un facteur 1.2. Par ailleurs, il est nécessaire

3. <http://wsa.roe.ac.uk/>

d'ajouter 0.02mag quadratiquement aux incertitudes photométriques pour prendre en compte les erreurs de calibration de l'instrument (Hodgkin et al. 2009). Ainsi on obtient les incertitudes photométriques corrigées  $\sigma_\lambda^c$  :

$$\sigma^c = \sqrt{(\sigma^{\text{obs}} \times 1.2)^2 + 0.02^2}. \quad (2.4)$$

Enfin on ne conserve que les étoiles dont les incertitudes photométriques corrigées sont inférieures à  $\sigma_{\text{max}} = 0.3\text{mag}$  ceci afin de retirer les observations aberrantes dues au pipeline de réduction.

## 2.3 Gaia - Data Release 2 (GDR2)

### 2.3.1 Description du relevé

Gaia est un satellite de l'Agence Spatiale Européenne (ESA) (Gaia Collaboration et al. 2016), il effectue des orbites de Lissajous autour du point de Lagrange L2 de la terre. De par sa rotation, il balaye l'ensemble de la voûte céleste, et fournit donc un relevé du ciel complet. Le satellite est équipé de 3 instruments se chargeant respectivement de l'astrométrie, de la photométrie et de la spectroscopie.

Gaia peut ainsi mesurer la parallaxe des astres ainsi que leurs mouvements propres sur la voûte céleste, avec une précision jamais égalée. Il fournit aussi la photométrie dans les bandes  $G$ ,  $G_{BP}$  et  $G_{RP}$ . Enfin, elle mesure la vitesse radiale des étoiles grâce à son spectrographe. Contrairement aux relevés au sol, Gaia observe le ciel en mode balayage, il n'y a donc pas de notion de champs de vue observé séparément. C'est par une réduction « globale » des données que les différents paramètres sont calculés.

Dans les travaux présentés dans ce manuscrit, nous utilisons la seconde publication des données (GDR2, Gaia Collaboration et al. 2018b). Cette publication contient plusieurs tables, notamment `gaiadr2.gaia_source` qui contient l'ensemble des résultats pour les sources ponctuelles. On peut aussi citer les différentes tables d'objets variables et notamment `gaiadr2.vari_cepheid`.

### 2.3.2 Sélection des données

Pour étudier le disque de la Voie lactée, nous avons utilisé les données issues de la table `gaiadr2.gaia_source`, et plus particulièrement nous utilisons la parallaxe  $\varpi$  (Lindgren et al. 2018) et la photométrie dans les 3 bandes disponibles (Evans et al. 2018). Un grand nombre d'indicateurs de fiabilité des données est fourni dans le catalogue.

FEDReD peut analyser séparément la photométrie et la parallaxe des étoiles. Une étoile donnée peut donc n'avoir que de l'information astrométrique ou uniquement de l'information photométrique ou bien la totalité de l'information. Ainsi on effectue une sélection séparément sur la parallaxe et sur la photométrie.

La présence d'étoiles doubles, qui ne sont pas encore gérées par les solutions astrométriques standards, conduit à des solutions où les paramètres astrométriques sont faux. Pour filtrer ces solutions, nous utilisons les indicateurs suivants :

- $\chi^2 = \text{astrometric\_chi\_2\_al}$
- $\nu = \text{astrometric\_n\_good\_obs\_al} - 5$
- $u = \sqrt{\chi^2/\nu}$

qui, conformément à Arenou et al. (2018) doivent passer le test :

$$u < 1.2 \times \max[1, \exp(-0.2(G - 19.5))]. \quad (2.5)$$

Nous avons aussi ajouté un critère sur la parallaxe et son erreur tel que :

$$\varpi + 3\sigma_\varpi > 0, \quad (2.6)$$

pour enlever les outliers évidents, parmi lesquels se trouvent à nouveau des étoiles doubles. En revanche, nous n’appliquons pas le filtre sur le critère, `visibility_period_used` c’est à dire le nombre d’observations bien séparées temporellement, car cela retirerait de l’information dans des zones du ciel mal couvertes par le satellite. Or c’est dans ces zones que nous avons besoin de l’information Gaia pour aider la convergence. Conformément à Arenou et al. (2018), on corrige la valeur des parallaxes du point-zéro (-0.03mas).

Concernant la photométrie dans les bandes  $G_{BP}$  et  $G_{RP}$ , on utilise les deux filtres :

$$\text{phot\_bp\_rp\_excess\_factor} < 1.3 + 0.06 (G_{BP} - G_{RP})^2 \quad (2.7)$$

$$G_{BP} < 18, \quad (2.8)$$

qui permettent de retirer des sources particulières, comme par exemple les étoiles doubles, ou des sources étendues.

Pour analyser la photométrie GDR2 on ajoute quadratiquement 0.01mag aux incertitudes photométriques pour prendre en compte les systématiques (Evans et al. 2018).

## 2.4 Complétude

Dans cette thèse, nous étudions le disque galactique à l’aide de l’algorithme FEDReD. L’un de ses points forts est la prise en compte de la complétude des champs de vue qu’il analyse. La complétude peut être comprise comme étant la magnitude apparente en dessous de laquelle toutes les étoiles du champ de vue ont été observées par le relevé. N’ayant pas accès à cette magnitude, on modélise la complétude à l’aide du paramètre  $m_{lim}$  qui représente la magnitude apparente au-delà de laquelle la fraction d’étoiles observées dans le champ de vue diminue significativement. Cette magnitude est évaluée pour chaque champ de vue étudié par FEDReD et pour chaque bande photométrique (voir section 3.2.5). On présente en Table 2.1 les magnitudes limites des différentes bandes photométriques pour différentes directions du disque galactique. On constate que la magnitude limite augmente vers l’anticentre, en effet l’encombrement y est moins important.

TABLE 2.1 – Magnitude limite  $m_{lim}$  pour les bandes photométriques de l’infrarouge pour les relevés 2MASS et UKIDSS, dans différentes directions du disque galactique.

	2MASS			UKIDSS		
	<i>J</i>	<i>H</i>	<i>K</i>	<i>J</i>	<i>H</i>	<i>K</i>
$l = 1.68^\circ$	17.00	15.29	14.14	19.60	17.86	16.35
$l = 9.60^\circ$	17.21	15.73	14.72	20.35	18.95	18.22
$l = 32.40^\circ$	16.76	15.54	14.75	19.72	18.99	18.26
$l = 36.88^\circ$	18.10	16.77	15.78	21.48	20.01	18.46
$l = 48.96^\circ$	17.88	16.85	16.16	20.50	20.34	19.64
$l = 49.20^\circ$	17.86	16.82	16.14	20.92	20.23	19.85
$l = 190^\circ$	18.33	17.54	17.11	Hors limite		

Pour les travaux relatés dans ce manuscrit, FEDReD analyse toujours une combinaison de 2 relevés (2MASS et GDR2 ou UKIDSS et GDR2). Or ces deux relevés n’ont pas la même limite de complétude. Nous devons donc choisir sur quel relevé baser la complétude et ne pas ajouter d’étoiles venant du second relevé. Autrement dit, le second relevé ne peut qu’apporter des informations à propos des étoiles présentes dans le premier relevé, toutes les étoiles analysées par FEDReD doivent systématiquement être observées par le premier relevé.

Pour l'étude que nous avons menée, nous avons choisi de baser la complétude sur les relevés infrarouges. En effet ces derniers permettent d'obtenir de l'information à plus grande distance, en particulier dans les champs de vue avec une forte extinction.

## 2.5 Appariement des catalogues (Crossmatch)

Notre algorithme utilise des données venant de plusieurs catalogues simultanément. Il utilise toujours un relevé photométrique dans le proche infrarouge combiné à la photométrie et aux parallaxes GDR2. Il est donc nécessaire d'appairer les catalogues, pour regrouper les informations relatives à une même étoile.

### 2MASS et GDR2

Pour assembler les catalogues 2MASS et GDR2, nous utilisons le crossmatch réalisé par Marse et al. (2019), et plus particulièrement la table `gaiadr2.tmass_best_neighbour`. Cet appariement a été effectué en ne tenant compte que de la position des étoiles, corrigée du mouvement propre. On ne conservera l'information venant du catalogue GDR2 que si la distance angulaire entre les deux sources est inférieure à 0.3 arcsec (Danielski et al. 2018). On présente en Figure 2.3 la distribution des distances angulaires du crossmatch pour chacun de nos champs de vue de test.

La table des meilleurs voisins donne l'étoile de 2MASS la plus proche de chaque étoile GDR2. En revanche, une étoile 2MASS peut être désignée comme la meilleure voisine de plusieurs étoiles GDR2. Si tel est le cas, on supprime toutes les informations provenant de GDR2 concernant cette étoile 2MASS. On évite ainsi toute mauvaise association.

On présente en Table 2.2 le nombre de sources de chaque champ test, et le nombre de sources ayant passé les critères de sélection.

TABLE 2.2 – Nombre d'étoiles dans les champs de vue de référence et de test de FEDReD (voir Chapitre 4) pour le crossmatch entre 2MASS et GDR2.

Champ de vue	Nombre d'étoiles dans le champs	Nombre d'étoiles analysées	
		Photométrie IR	parallaxes
$l = 000.68^\circ$	31480	8717	1815
$l = 009.60^\circ$	30757	10467	2963
$l = 032.40^\circ$	28438	12428	6463
$l = 036.88^\circ$	23313	10233	2043
$l = 048.96^\circ$	21808	9632	5206
$l = 049.20^\circ$	21426	9602	5599
$l = 190.00^\circ$	3700	3361	2102

### UKIDSS et GDR2

Le crossmatch entre UKIDSS et GDR2 n'est pas fourni avec le catalogue Gaia. Nous réalisons donc nous-mêmes le crossmatch en utilisant l'astrométrie. On recherche l'étoile de GDR2 la plus proche de chaque étoile UKIDSS. Les époques d'observation ne sont pas les mêmes pour les deux catalogues, en effet GDR2 donne des observations à J2015 tandis que les données UKIDSS dépendent du jour d'observation de l'étoile (les observations ont commencé en 2005 et se poursuivent jusqu'en 2013 pour la DR10). Néanmoins nous avons décidé de ne pas corriger les mouvements propres, en effet seules les étoiles proches auraient une dérive suffisamment importante pour

être mal associées, or avec UKIDSS on s'intéresse aux étoiles lointaines pour agrandir nos cartes d'extinction.

Pour décider à quelle distance angulaire couper le crossmatch, nous traçons l'histogramme de la séparation angulaire des meilleurs voisins pour l'ensemble du disque galactique accessible par UKIDSS. Cet histogramme, présenté en Figure 2.4, dessine une courbe avec une queue de distribution vers les distances angulaires élevées. Cette queue de distribution correspond à de mauvaises identifications, qu'il faut couper en instaurant une distance angulaire maximale.

Il semblerait sur cet histogramme qu'une coupure à  $0.2''$  couperait toute la queue de distribution. Nous avons effectué de nombreux tests de FEDReD sur des champs de vue réels (voir chapitre 4) et il est apparu que couper à  $0.2''$  est trop permissif. La Figure 2.5 présente les histogrammes de séparation angulaire par champ de vue. Pour beaucoup de nos champs de vue test une coupure à  $0.2''$  introduisait beaucoup d'identifications erronées et perturbait la convergence de l'algorithme. On a donc décidé de couper à  $0.15''$  (droite verticale sur les différents histogrammes), bien que dans certains cas, comme à  $l = 48.96^\circ$ , on enlève de l'information utile, on évite dans de nombreux cas des outliers créés suite à un mauvais appariement. Dans le cadre des données GDR2 associées à UKIDSS, il est préférable de perdre un peu d'information Gaia plutôt que d'introduire de mauvaises parallaxes et/ou photométries. On présente en Table 2.3 le nombre d'étoiles de chacun des champs tests, ainsi que le nombre d'étoiles retenues après filtrage.

TABLE 2.3 – Nombre d'étoiles dans les champs de vue de référence et de test de FEDReD (voir Chapitre 4) pour le crossmatch entre UKIDSS et GDR2.

Champ de vue	Nombre d'étoiles dans le champs	Nombre d'étoiles analysées	
		Photométrie IR	parallaxes
$l = 001.68^\circ$	33438	9343	1576
$l = 009.60^\circ$	39340	9768	1433
$l = 032.40^\circ$	39029	5401	3776
$l = 036.88^\circ$	28417	5986	353
$l = 048.96^\circ$	30156	6163	1157
$l = 049.20^\circ$	26055	5731	1475

## 2.6 Photométrie et unité

La photométrie d'une étoile peut s'exprimer dans plusieurs unités, l'unité historique étant la magnitude. La magnitude est à l'origine, une classification en six catégories des étoiles visibles à l'œil nu, réalisée par l'astronome grec Hipparque, on a par la suite redéfini la magnitude à partir des flux de photons reçus d'une étoile, comparées à un point zéro servant de référence.

La plupart des relevés photométriques du ciel fournissent des résultats et des incertitudes de mesures exprimées en magnitude. Or la mesure réellement effectuée est celle du flux d'électrons créé par les photons d'une étoile sur le détecteur. Ce flux est ensuite utilisé pour calculer la magnitude et estimer l'erreur sur la magnitude.

Les erreurs de mesure peuvent être considérées comme gaussiennes en flux d'électron. Pour prendre en compte parfaitement les incertitudes sur les données photométriques des étoiles, il faut donc travailler dans l'espace des flux de photons. Le catalogue GDR2 fournit la photométrie en flux d'électron ainsi que le point zéro des magnitudes. On peut donc travailler dans cet espace.

En revanche les relevés UKIDSS et 2MASS sont des relevés effectués au sol par prise de vue successive, avec une réduction des données image par image. Le point zéro des magnitudes est déterminé pour chaque pose pour finalement obtenir des magnitudes et des erreurs sur les



magnitudes cohérentes sur l'ensemble du ciel. Seules les magnitudes et les incertitudes sur les magnitudes sont mises à disposition dans les catalogues.

Les points zéros de chaque prise de vue ne sont pas publiés, on ne peut donc pas remonter facilement au flux de chaque étoile à partir des magnitudes. Pour les bandes photométriques dans l'infrarouge, nous devons donc nous contenter de travailler dans l'espace des magnitudes.

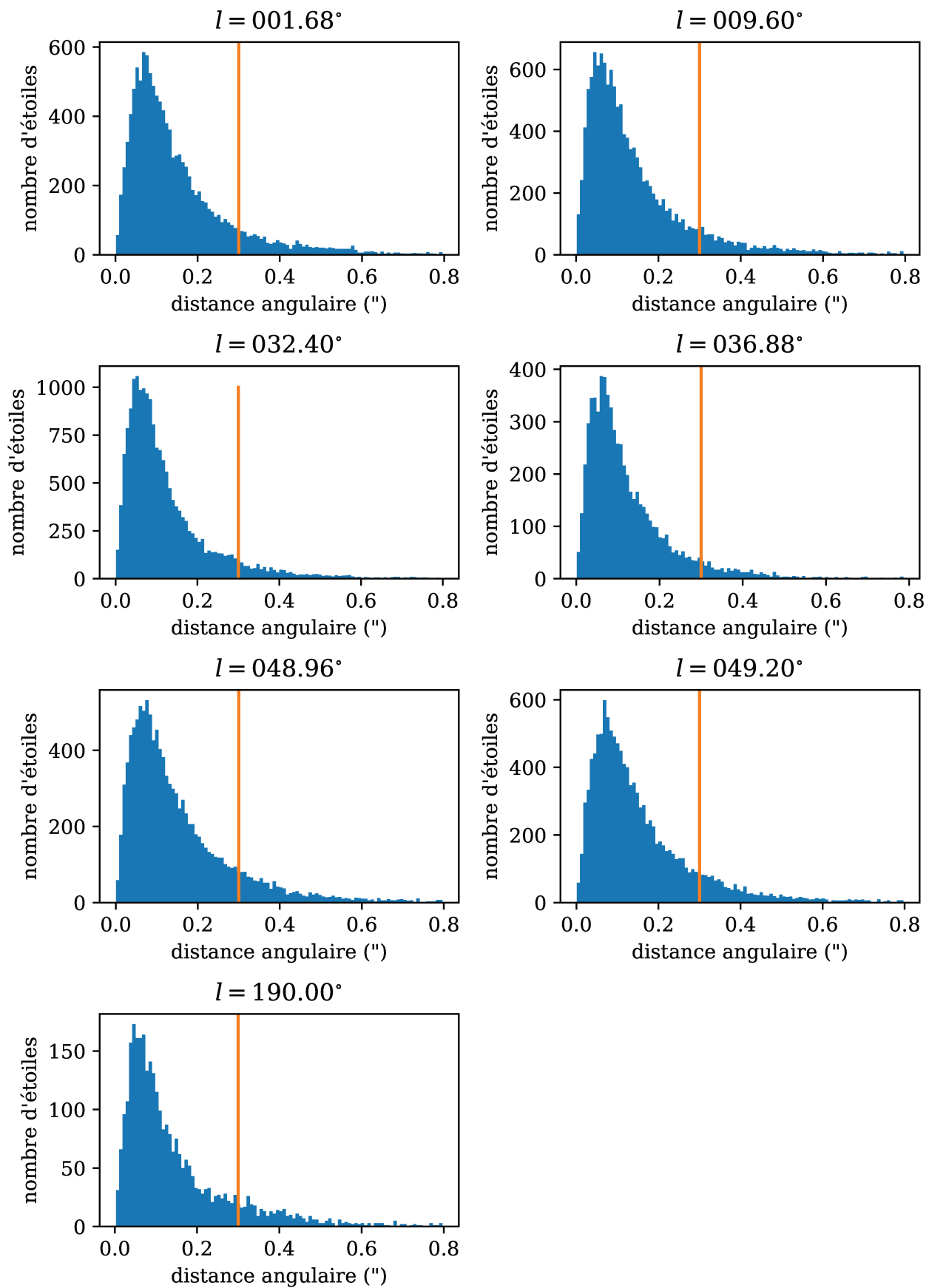


FIGURE 2.3 – Histogrammes des distributions d'écart angulaire entre meilleurs voisins lors du crossmatch entre les catalogues 2MASS et GDR2. La droite horizontale est tracée à une distance angulaire de 0.3''

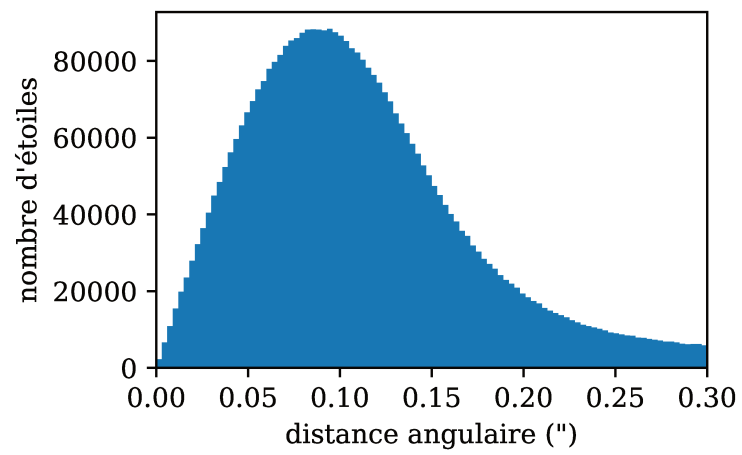


FIGURE 2.4 – Histogramme des séparations angulaires des meilleurs voisins pour l'ensemble du disque de la Voie lactée.

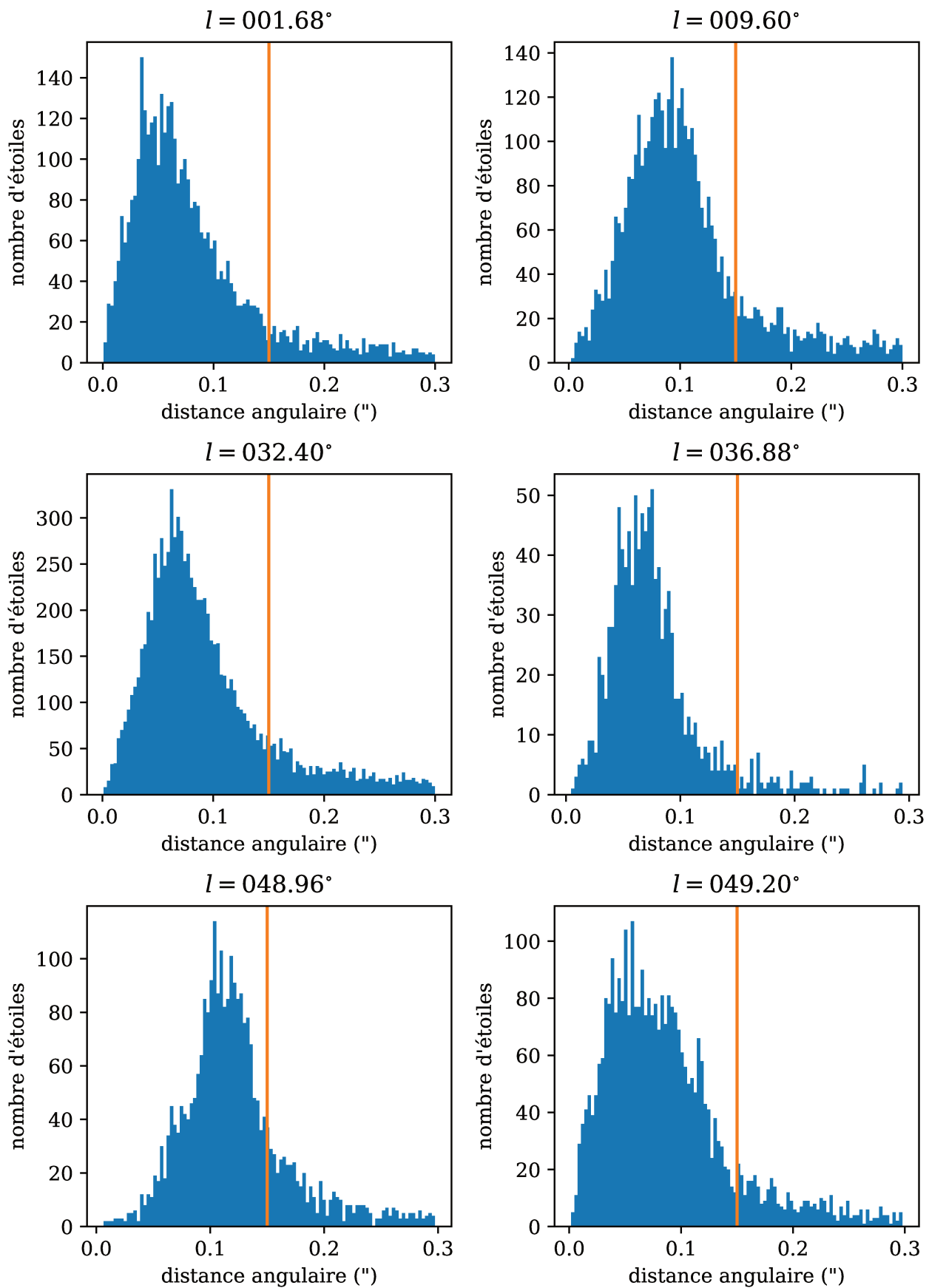


FIGURE 2.5 – Histogrammes des distributions d'écart angulaire entre meilleurs voisins lors du crossmatch entre les catalogues UKIDSS et GDR2. La droite horizontale est tracée à une distance angulaire de 0.15''



# Chapitre 3

## L’algorithme Field Extinction - Distance Relation Deconvolver

L’ALGORITHME FEDReD recherche la relation entre la distance et l’extinction mais aussi la relation entre la densité d’étoiles et la distance pour un champ de vue donné. Pour ce faire, il procède en deux étapes distinctes. Premièrement, il étudie séparément chacune des étoiles du champ de vue, en utilisant l’ensemble des informations disponibles, c’est l’étape des distributions individuelles. En second lieu, il lui faut assembler ces distributions individuelles en une information pour le champ de vue dans son ensemble, c’est l’étape de déconvolution.

### 3.1 Distribution individuelle

Dans un premier temps, FEDReD doit analyser séparément les étoiles présentes dans un champ de vue. Pour cela, nous disposons d’un certain nombre d’observables. Les différentes observables sont la photométrie infrarouge, en bande J, H et K provenant du relevé 2MASS ou des relevés UKIDSS . Le catalogue GDR2, quant à lui, apporte de la photométrie dans les bandes  $G$ ,  $G_{BP}$  et  $G_{RP}$  ainsi que la parallaxe. Toutes les observables sont accompagnées de leurs incertitudes.

Le nombre d’observables varie d’une étoile à une autre. En effet, une étoile peut avoir été observée par seulement l’un des relevés, ou alors la précision des observations pour une valeur n’est pas suffisante pour la prendre en compte. Dans la suite de ce manuscrit, on désignera par  $\mathbf{O}$ , l’ensemble des observables relatives à une étoile donnée, peu importe leur nombre.

À partir de ces observables, nous souhaitons calculer la distribution de probabilité de  $P(\mathbf{O} \mid A_0, D)$ . Autrement dit, on cherche à connaître la vraisemblance d’observer l’étoile étudiée (avec les observables effectivement obtenues) si elle se trouve à une distance  $D$  et absorbée par une extinction  $A_0$ , pour chaque couple  $(A_0, D)$  considéré dans notre étude.

#### 3.1.1 Diagramme de Hertzsprung-Russel (diagramme HR) de référence

Pour calculer la distribution  $P(\mathbf{O} \mid A_0, D)$  d’une étoile, nous devons retrouver sa position dans un diagramme HR. C’est-à-dire déduire sa probable magnitude absolue et les coefficients d’extinction associés. Pour cela, nous avons besoin d’un diagramme HR de référence, représentatif des populations stellaires du disque de la Voie lactée.

Dans un premier temps nous avons utilisé les isochrones de Padoue (Bressan et al. 2012) comme référence. Cette référence produisait des résultats convenables, mais a besoin d’un certain

nombre de paramètres d’entrée, comme la fonction de masse initiale (IMF) ou le taux de formation d’étoiles (SFR). Nous rendant dépendant d’autres modèles.

C’est pourquoi nous avons construit un diagramme HR de référence de manière empirique. Ce dernier se base uniquement sur les catalogues GDR2 et 2MASS. En effet, la seconde publication des données Gaia fournit un grand nombre d’étoiles proches et sans extinction permettant ainsi d’obtenir d’un diagramme HR observé très fin (Gaia Collaboration et al. 2018a).

Dans ce diagramme HR, nous avons sélectionné les étoiles appartenant au disque galactique ( $|Z| < 50\text{pc}$ ), pour obtenir une référence cohérente avec notre milieu d’étude. Nous avons aussi besoin de connaître les magnitudes infrarouges des étoiles de ce diagramme. On se base donc sur le catalogue 2MASS et le crossmatch de ce dernier avec la GDR2, fourni par Marrese et al. (2019). Nous pouvons ainsi ajuster des relations entre les différentes couleurs, et calculer des correspondances. On dispose ainsi d’un diagramme HR empirique uniquement construit à l’aide d’observations.

Pour pouvoir l’utiliser comme référence nous binons le diagramme, le nombre d’étoiles dans chaque bin nous permet d’associer un poids à chacun de ces bins. On corrige ensuite ces poids en tenant compte des coupures que l’on a réalisées lors de la sélection de nos étoiles (voir Babusiaux et al. (2019) pour plus de détails). Ce diagramme étant construit à partir d’étoiles locales, on peut se demander s’il est suffisamment représentatif du disque galactique dans son ensemble. D’après les tests réalisés par Babusiaux et al. (2019), on a pu constater que ce diagramme produit de meilleurs résultats que les isochrones de Padoue. Cela s’explique par l’utilisation de relations couleur-couleur empiriques plus proches des observations que les relations théoriques Ruiz-Dern et al. (2018), et le fait qu’il est aussi représentatif d’objets particuliers tel que les étoiles binaires.

On présente en Figure 3.1 le diagramme HR empirique de référence que nous utilisons pour analyser les étoiles individuellement.

Par ailleurs, on peut s’interroger sur la validité de ce diagramme HR dans les champs de vue comportant des régions d’importante formation d’étoiles, car la forme du diagramme est modifiée par la présence massive d’étoiles pré-séquence principale. Mais l’information principale sur laquelle s’appuie FEDReD provient des étoiles du Red Clump, c’est à dire des étoiles d’âge intermédiaire.

### 3.1.2 Loi d’extinction

Pour comparer convenablement une étoile observée à notre diagramme HR de référence, pour une extinction  $A_0$  donnée, il nous faut calculer la correspondance de  $A_0$  dans les bandes photométriques où l’étoile a été observée. Pour réaliser cette conversion, nous utilisons l’équation (1.3). Cependant, la valeur du coefficient d’extinction varie en fonction de la couleur intrinsèque de l’étoile, mais également en fonction de l’extinction  $A_0$  elle-même.

Pour calculer les coefficients d’extinction, nous utilisons le modèle développé par Danielski et al. (2018). Il décrit les coefficients d’extinction comme des polynômes dépendants de la couleur intrinsèque  $(G - K)_0 = M_G - M_K$  et de l’extinction  $A_0$  :

$$k_\lambda = a_1 + a_2(G - K)_0 + a_3(G - K)_0^2 + a_4(G - K)_0^3 + a_5A_0 + a_6A_0^2 + a_7A_0 \times (G - K)_0. \quad (3.1)$$

La valeur des coefficients  $a_i$  varie pour chacune des bandes photométriques. On a représenté, en Figure 3.2, la variation des coefficients d’extinction en fonction de la couleur  $G - K$ , pour 3 valeurs de  $A_0$ . Pour les bandes photométriques du proche infrarouge ( $J$ ,  $H$  et  $K$ ), la variation relative des coefficients est faible. En revanche, elle est loin d’être négligeable pour les bandes Gaia ( $G$ ,  $G_{BP}$  et  $G_{RP}$ ). Elle est d’environ 10% pour les bandes  $G_{BP}$  et  $G_{RP}$  entre les deux extrêmes de couleur et elle atteint même 40% pour la bande  $G$ . Cette forte variation est due à la grande largeur spectrale des filtres Gaia. Ainsi, si on utilisait une loi d’extinction indépendante de la couleur, on pourrait fortement mésestimer l’extinction.

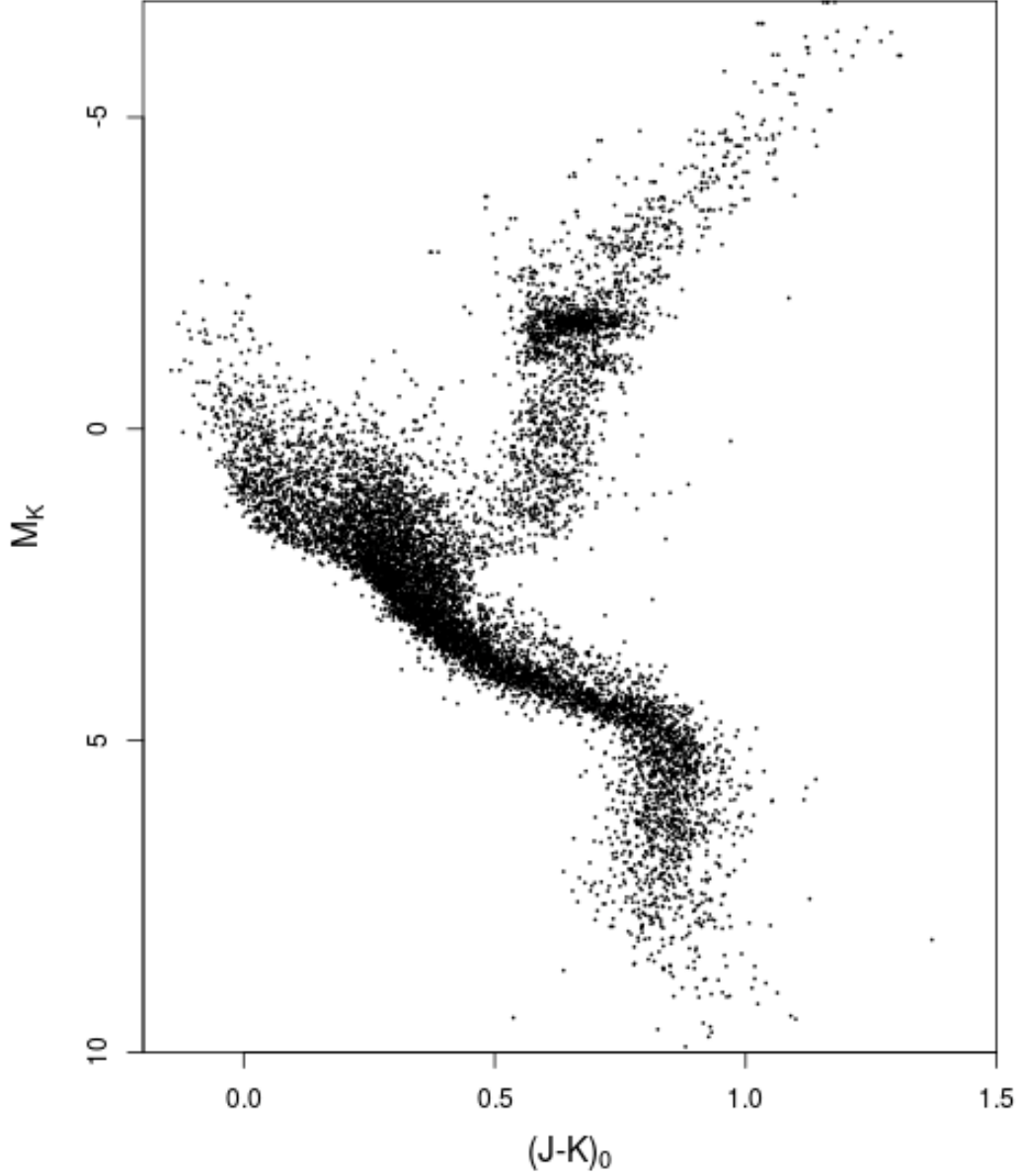


FIGURE 3.1 – Diagramme HR empirique, construit à partir des catalogues GDR2 et 2MASS

Par ailleurs, on considère ici que le paramètre  $R_V = \frac{A_V}{E_{B-V}}$  est constant dans toutes les directions du disque galactique.

### 3.1.3 Calcul des probabilités $P(O | A_0, D)$

Nous pouvons séparer les observations d'une étoile en deux parties. Tout d'abord les observations photométriques, que l'on note  $O_\Lambda$  (regroupe les magnitudes apparentes  $m_\lambda$  et les incertitudes de mesure  $\sigma_\lambda$ ), et l'astrométrie que l'on note  $O_\varpi$  (représentant la parallaxe  $\varpi$  ainsi que l'incertitude sur la parallaxe  $\sigma_\varpi$ ). On peut donc écrire

$$P(O | A_0, D) = P(O_\Lambda | A_0, D) \times P(O_\varpi | A_0, D). \quad (3.2)$$

Nous considérerons que la parallaxe est indépendante de l'extinction, on écrit donc :

$$P(O | A_0, D) = P(O_\Lambda | A_0, D) \times P(O_\varpi | D). \quad (3.3)$$



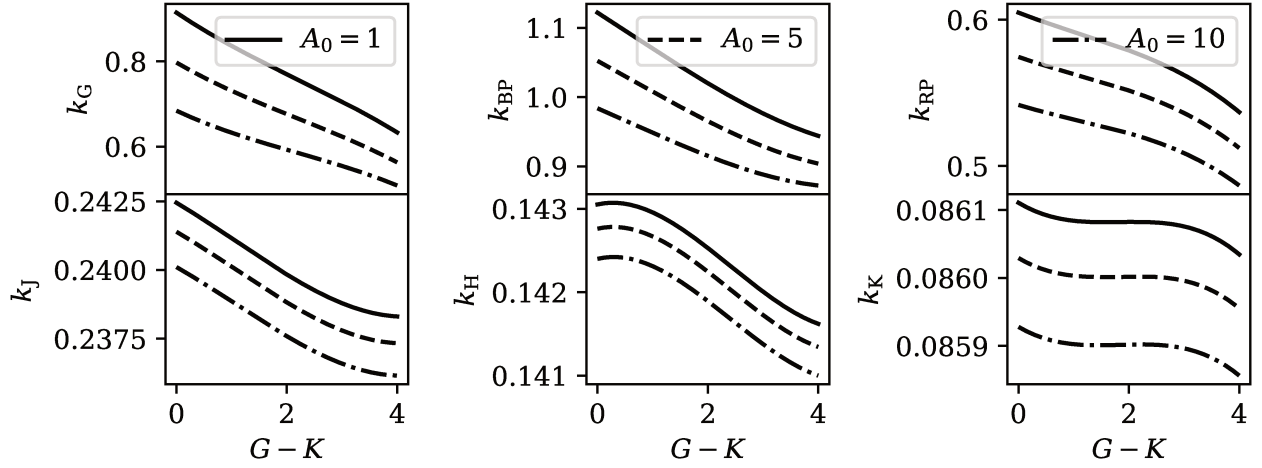


FIGURE 3.2 – Variation des coefficients d’extinction en fonction de la couleur intrinsèque  $G - K$ , pour les bandes photométriques de GDR2, selon le modèle de Danielski et al. (2018).

Pour calculer  $P(\mathbf{O}_\varpi | D)$ , nous convertissons la distance  $D$  en parallaxe théorique  $\varpi_t = 1/D$ . Utiliser la parallaxe plutôt que la distance permet de profiter de la nature approximativement gaussienne des erreurs sur la parallaxe (Arenou et al. 2018) ; ainsi on obtient :

$$P(\mathbf{O}_\varpi | D) = P(\mathbf{O}_\varpi | \varpi_t(D)) = \mathcal{N}(\varpi_t(D) - \varpi, \sigma_\varpi). \quad (3.4)$$

Pour calculer la vraisemblance de la photométrie,  $P(\mathbf{O}_\Lambda | A_0, D)$ , nous utilisons l’ensemble des points du diagramme HR de référence (noté  $r$ ) comme comparatif,

$$P(\mathbf{O}_\Lambda | A_0, D) = \sum_r P(\mathbf{O}_\Lambda | A_0, D, r) P(r). \quad (3.5)$$

Prenons le cas de la bande quelconque  $\lambda$ , nous connaissons la magnitude apparente de l’étoile  $m_\lambda$ . Nous voulons calculer  $P(\mathbf{O}_\lambda | A_0, D, r)$ , qui correspond à la vraisemblance d’une étoile observée par rapport au couple  $(A_0, D)$  si elle se trouve au point  $r$  du diagramme. Ce point  $r$  de notre diagramme HR empirique, est notamment décrit par la magnitude absolue  $M_\lambda^r$ , mais aussi par la couleur intrinsèque  $(G - K)^r$  associée. Avec l’équation (3.1), nous pouvons donc calculer le coefficient d’extinction  $k_\lambda((G - K)^r, A_0)$  auquel serait soumise l’étoile du point  $r$  avec une extinction  $A_0$ . En combinant les équations (1.2) et (1.3) nous obtenons la magnitude apparente théorique de l’étoile  $r$  se trouvant à une distance  $D$  et atténuée par une extinction  $A_0$  :

$$m_\lambda^r = M_\lambda^r + 5 \log(D) - 5 + k_\lambda A_0. \quad (3.6)$$

Comme dans le cas de la parallaxe, on obtient la valeur de  $P(\mathbf{O}_\lambda | A_0, D, r)$  en utilisant la loi normale :

$$P(\mathbf{O}_\lambda | A_0, D, r) = \mathcal{N}(m_\lambda^r - m_\lambda, \sigma_\lambda). \quad (3.7)$$

Pour obtenir  $P(\mathbf{O}_\Lambda | A_0, D, r)$ , on obtient naturellement :

$$P(\mathbf{O}_\Lambda | A_0, D, r) = \prod_\lambda P(\mathbf{O}_\lambda | A_0, D, r), \quad (3.8)$$

avec  $\lambda$  parcourant les bandes photométriques observées pour l’étoile observée.

**N.B.** Nous venons ici de décrire le calcul de  $P(\mathbf{O}_\lambda | A_0, D, r)$  dans le cas où la loi d'erreur est gaussienne dans l'espace des magnitudes. Bien que cela ne soit pas le cas, nous n'avons accès qu'aux magnitudes infrarouges (voir chapitre 2), et c'est donc dans cet espace que l'on réalise nos calculs. En revanche, le catalogue GDR2 fournit les flux et les erreurs sur les flux, on calcule donc les probabilités associées à la photométrie GDR2 dans l'espace des flux. Pour ce faire on convertit les magnitudes apparentes de référence  $m_\lambda^r$  en flux à l'aide des points zéro des magnitudes fournis par le catalogue GDR2.

## 3.2 Déconvolution

On a donc à notre disposition les probabilités  $P(\mathbf{O} | A_0, D)$  pour chacune des étoiles de notre champ de vue. On souhaite maintenant assembler l'ensemble de ces résultats, en une seule distribution de probabilités  $P(A_0, D)$  correspondant à l'ensemble du champ de vue. On se placera tout d'abord dans le cas simple où toutes les étoiles du champs de vue sont observées. On prendra ensuite en compte la complétude des catalogues.

Dans la suite de ce chapitre, nous illustrerons les développements par les différents résultats intermédiaires obtenus pour un champ de vue simulé (voir section 4.1) dont on présente un CMD en Figure 1.6.

### 3.2.1 Déconvolution de Richardson-Lucy

Si on a observé l'ensemble des  $N$  étoiles d'un champ de vue, on peut écrire la probabilité jointe de l'extinction et la distance  $P(A_0, D)$  du champ de vue :

$$P(A_0, D) = \sum_{i=1}^N P(A_0, D | \mathbf{O}_i) P(\mathbf{O}_i) = \frac{1}{N} \sum_{i=0}^N P(A_0, D | \mathbf{O}_i). \quad (3.9)$$

On peut aussi écrire que pour une étoile donnée  $\mathbf{O}_i$  :

$$P(\mathbf{O}_i) = \int_{A_0} \int_D P(\mathbf{O}_i | A_0, D) P(A_0, D) dD dA_0, \quad (3.10)$$

donc

$$P(A_0, D | \mathbf{O}_i) = \frac{P(\mathbf{O}_i | A_0, D) P(A_0, D)}{P(\mathbf{O}_i)} = \frac{P(\mathbf{O}_i | A_0, D) P(A_0, D)}{\int_{A_0} \int_D P(\mathbf{O}_i | A_0, D) P(A_0, D) dD dA_0}. \quad (3.11)$$

En combinant les équations (3.9) et (3.11) on obtient donc :

$$P(A_0, D) = \frac{1}{N} \sum_{i=0}^N \frac{P(\mathbf{O}_i | A_0, D) P(A_0, D)}{\int_{A_0} \int_D P(\mathbf{O}_i | A_0, D) P(A_0, D) dD dA_0}. \quad (3.12)$$

Pour déconvoluer l'équation (3.12), on utilise une méthode itérative décrite successivement par Richardson (1972) dans le cas du traitement d'image puis par Lucy (1974) pour l'étude de distribution statistique. Pour la mettre en œuvre, on se base sur les équations de Lindegren (1995). On définit l'estimateur  $h_k(A_0, D)$  tel que :

$$h_k(A_0, D) = \frac{1}{N} \sum_{i=1}^N \frac{P(\mathbf{O}_i | A_0, D) \times h_{k-1}(A_0, D)}{\int_{A_0} \int_D h_{k-1}(A_0, D) \times P(\mathbf{O}_i | A_0, D) dD dA_0}, \quad (3.13)$$

dont on calcule la valeur itérativement pour obtenir la distribution jointe de l'extinction  $A_0$  et la distance  $D$  du champ de vue. La valeur d'initialisation  $h_0(A_0, D)$  correspond au prior de la distribution  $P(A_0, D)$ .

### 3.2.2 Déconvolution avec fonction de sélection

Le cas précédent décrit un cas idéal, où l'ensemble des étoiles présentes au sein d'un champ de vue est effectivement observé. En réalité, seule une partie des étoiles est observée et conservée dans le catalogue. On modélise la complétude des catalogues par la fonction de sélection  $S$ , des catalogues infrarouges (voir section 3.2.5).

La distribution à laquelle nous avons réellement accès est celle de la probabilité jointe d'extinction et de distance des  $N$  étoiles qui ont été observées :

$$P(A_0, D | S) = \sum_{\mathbf{O} \in S} P(A_0, D | \mathbf{O}, S) P(\mathbf{O} | S) = \frac{1}{N} \sum_{\mathbf{O} \in S} P(A_0, D | \mathbf{O}, S). \quad (3.14)$$

De même que précédemment, on peut écrire :

$$P(A_0, D | \mathbf{O}, S) = \frac{P(\mathbf{O} | A_0, D, S) P(A_0, D | S)}{\int_D P(\mathbf{O} | A_0, D, S) P(A_0, D | S) dD dA_0}, \quad (3.15)$$

avec

$$P(\mathbf{O} | A_0, D, S) = \frac{P(S | \mathbf{O}, A_0, D) P(\mathbf{O} | A_0, D)}{P(S | A_0, D)}. \quad (3.16)$$

Or, si une étoile a été observée, c'est qu'elle a été sélectionnée, d'où  $P(S | \mathbf{O}) = P(S | \mathbf{O}, A_0, D) = 1$ . On peut donc ré-écrire notre estimateur  $h(A_0, D | S)$  en tenant compte de notre fonction de sélection :

$$h_k(A_0, D | S) = \frac{1}{N} \sum_{i=1}^N \frac{P(\mathbf{O}_i | A_0, D) / P(A_0, D | S) \times h_{k-1}(A_0, D | S)}{\int_{A_0} \int_D h_{k-1}(A_0, D | S) \times P(\mathbf{O}_i | A_0, D) / P(A_0, D | S) dD dA_0}. \quad (3.17)$$

On est donc capable de calculer la probabilité  $P(A_0, D | S)$  pour chaque couple  $(A_0, D)$ . Le résultat de la déconvolution de notre champ simulé est présenté en Figure 3.3. Nous y voyons déjà la forme de la relation entre l'extinction et la distance. On constate que la densité de probabilité est très forte à faible distance. Cela est dû aux effets de sélection, on a en effet observé plus d'étoiles proches. Il nous faut donc corriger ce résultat par  $P(S | A_0, D)$  pour obtenir la distribution de  $P(A_0, D)$  :

$$P(A_0, D) = \frac{P(A_0, D | S)}{P(S | A_0, D)} \times P(S), \quad (3.18)$$

où  $P(S)$  ne correspond ici qu'à une normalisation de la distribution.

### 3.2.3 Critère d'arrêt

La déconvolution de Richardson Lucy ne converge pas au sens traditionnel du terme. Au fur et à mesure des itérations, l'estimateur  $h(A_0, D | S)$  se rapproche de la distribution,  $P(A_0, D | S)$  mais du bruit de déconvolution apparaît progressivement et fait osciller  $h$  jusqu'à dominer le signal, si les itérations sont trop nombreuses.

Nous utilisons la somme des résidus aux carrés (residual sum of square ou RSS)  $\Delta$ , comme traceur de la déconvolution. Pour l'itération  $k$  on a :

$$\Delta_i = \sum_{A_0, D} [h_k(A_0, D | S) - h_{k-1}(A_0, D | S)]^2. \quad (3.19)$$

On prend la différence normalisée des  $\Delta$  :

$$cr = \left| \frac{\Delta_{i+1} - \Delta_i}{\Delta_i} \right| \leq 0.05, \quad (3.20)$$

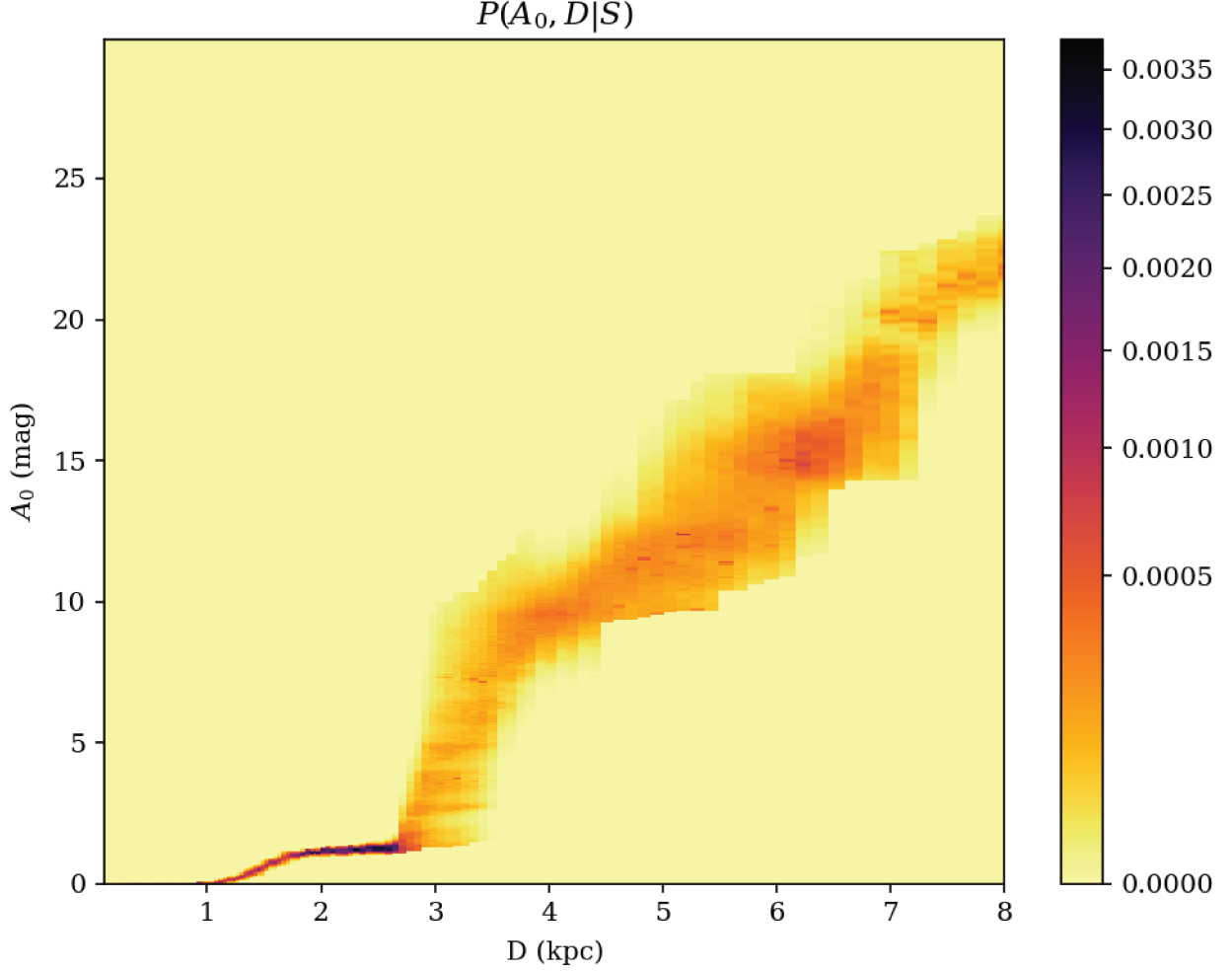


FIGURE 3.3 – Distribution de  $P(A_0, D | S)$  pour le champ de vue simulé à  $l = 9.6^\circ$

comme critère d'arrêt. On représente en Figure 3.4 l'évolution de  $cr$  avec le nombre d'itérations. Lors des premières itérations  $cr$  décroît fortement, pour ensuite atteindre un palier. Lorsque le  $cr$  varie peu la déconvolution « n'avance plus », les itérations successives ne font qu'osciller autour de la solution.

Le choix de la valeur 0.05 comme limite pour arrêter la déconvolution a été décidé empiriquement à force de tests. Par ailleurs, pour certain champs complexes à déconvoluer, il arrive que le minimum de  $cr$  soit supérieur à 0.05 ; c'est pourquoi, lorsque l'on détecte un minimum de  $cr$ , on conserve les résultats temporaires tout en continuant la déconvolution. Cela nous permet de revenir à ce minimum si aucun autre minimum n'est atteint, tout en évitant de s'arrêter dans un éventuel minimum local.

### 3.2.4 Prior

L'équation (3.17) nous indique comment calculer l'itération  $k + 1$  connaissant l'itération  $k$ , il nous faut donc initialiser la déconvolution avec un prior à la distribution  $P(A_0, D | S) = h_0(A_0, D | S)$ . Dans la pratique nous fixons un prior sur les distributions  $P(A_0 | D)$  et  $P(D)$ , à partir desquels nous calculons  $h_0$  :

$$h_0(A_0, D | S) = P(A_0, D | S)_0 = \frac{P(A_0 | D) \times P(D)}{P(S | A_0, D)} \times P(S), \quad (3.21)$$

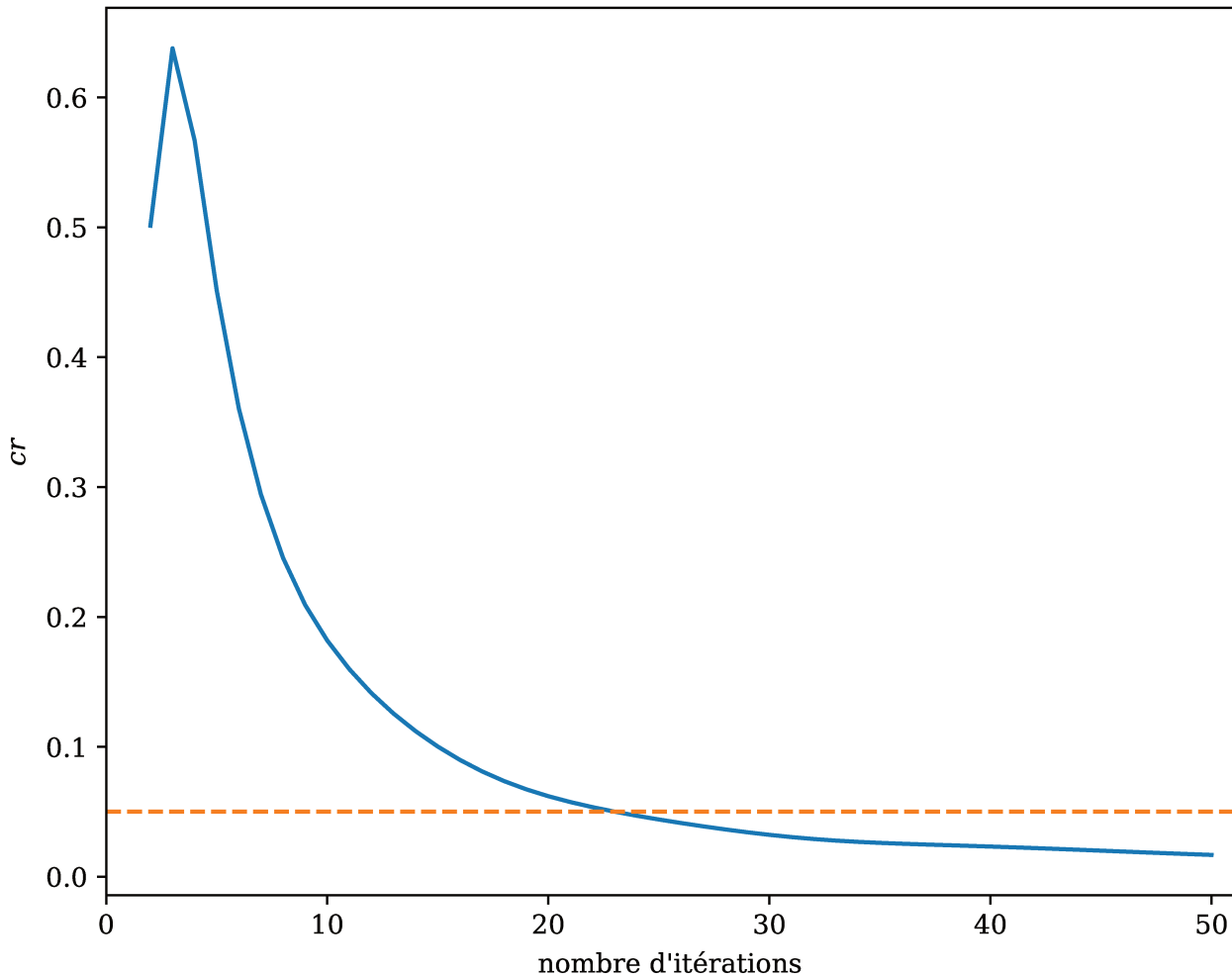


FIGURE 3.4 – Évolution de  $cr$  avec le nombre de déconvolutions, la ligne en pointillés représente notre seuil d'arrêt de 0.05

avec  $P(S | A_0, D)$ , la fonction de sélection du champ de vue (voir section 3.2.5).

### $P(A_0 | D)$

Afin de ne pas biaiser nos résultats, nous définissons un prior très permissif. Nous utilisons une distribution *a priori* plate de  $P(A_0 | D)$ , c'est à dire  $P(A_0 | D) = \text{cste}, \forall A_0, D$ .

À cette distribution, on ajoute un critère supplémentaire, l'interdiction d'avoir une forte extinction à faible distance :

$$P(A_0 | D) = 0 \text{ pour } A_0 > 10 \times D_{\text{kpc}}. \quad (3.22)$$

L'ajout de cette condition évite la contamination à courte distance par les distributions  $P(\mathbf{O} | A_0, D)$  des naines froides et du Red Clump. En effet, ces étoiles peuvent être interprétées comme étant des étoiles géantes (à forte luminosité intrinsèque) très fortement éteintes.

Par ailleurs, cette condition ne biaise pas nos résultats. En effet, en s'appuyant sur les cartes de Lallement et al. (2019), nous savons qu'aucune ligne de visée ne présente de si fortes extinctions à des distances si faibles.

On représente en partie gauche de la Figure 3.5 la distribution *a priori* de  $P(A_0, D | S)$  que nous utilisons pour tous les champs de vue que nous analysons. La zone rouge correspond aux

extinctions non physiques, que l'on interdit. Le reste de l'espace  $(A_0, D)$  est autorisé et considéré comme équiprobable.

### $P(D)$

C'est la distribution du nombre d'étoiles avec la distance, sa valeur est donc reliée à la densité d'étoiles. Nous prenons un profil de disque exponentiel avec comme longueur d'échelle  $H_R = 2\text{kpc}$  et la distance Soleil - centre Galactique  $R_\odot = 8\text{kpc}$ . À une longitude  $l$  on a :

$$P(D) \propto \exp\left(-\frac{\sqrt{D^2 + R_\odot^2} - 2DR_\odot \times \cos(l)}{H_R}\right) \times D^2 dD, \quad (3.23)$$

l'important étant de corriger l'effet cône par le produit avec  $D^2$ . On présente en partie droite de la Figure 3.5 les priors  $P(D)$  pour différentes longitudes.

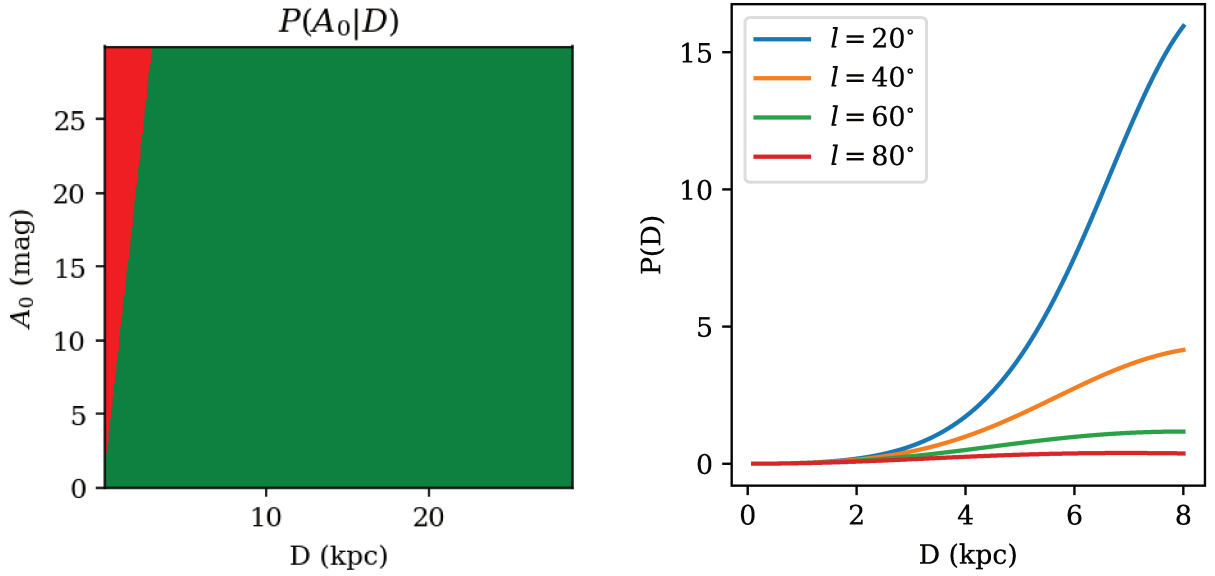


FIGURE 3.5 – Partie gauche : Prior  $P(A_0 | D)$ . Partie droite : Prior  $P(D)$  en fonction de la distance pour différentes longitudes.

### $h_0(A_0, D | S)$

le prior d'initialisation de la déconvolution d'un champs test est présenté en Figure 3.6. La forte densité de probabilité à  $D = 8\text{kpc}$  correspond à la forte densité d'étoiles du centre galactique.

### Second $h_0(A_0, D | S)$

Pour obtenir les meilleurs résultats possibles et s'affranchir d'éventuels points aberrants, nous procédons à deux déconvolutions successives. La première est réalisée avec l'ensemble des priors décrit ci-dessus. La seconde est initialisée de la même façon, sauf pour le  $P(A_0, D | S)_0$ , qui est calculé à partir du résultat de la première déconvolution.

### 3.2.5 Complétudes

La connaissance de la distribution des  $P(S | A_0, D)$  est requise pour plusieurs opérations de FEDReD (équations (3.18), (3.21) et (3.17)). Pour les résultats présentés dans ce manuscrit,

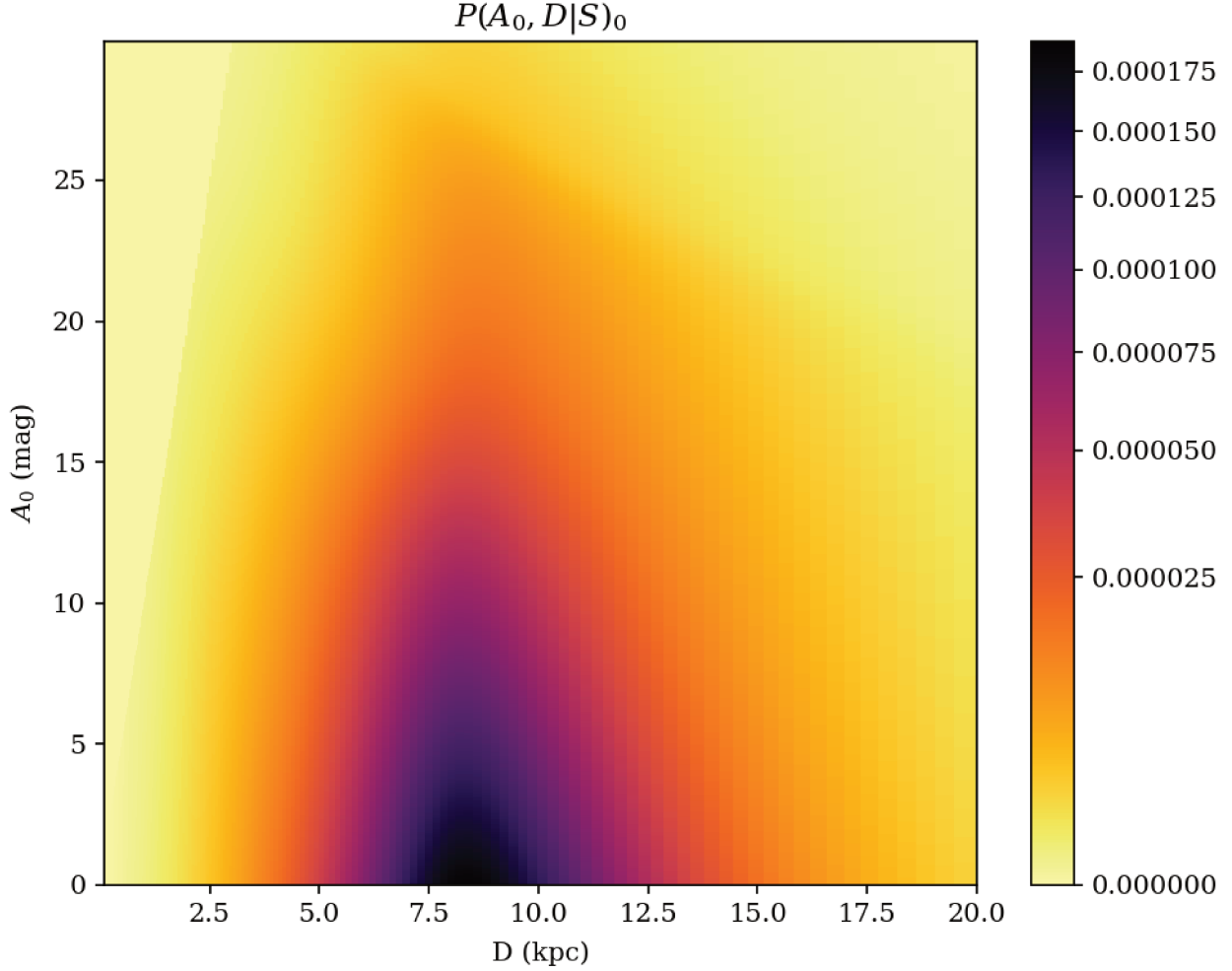


FIGURE 3.6 – Prior d'initialisation de la déconvolution d'un champ de vue à  $l = 9.6^\circ$

nous basons la complétude des champs de vue sur les catalogues photométriques en infrarouge. C'est-à-dire que les étoiles que nous conservons pour notre étude présentent nécessairement des magnitudes observées en bande J, H et K (voir section 2.4). Ces informations sont complétées si possible par le catalogue GDR2.

### Modèle de complétude

Pour calculer la probabilité pour une étoile de magnitude  $m_\lambda$  d'être observée ( $S$ ), on utilise le modèle suivant :

$$P(S | m_\lambda) = \begin{cases} 1 - \exp \left[ - \left( \frac{(m_\lambda^{lim} - m_\lambda)}{\beta} \right)^\alpha \right] & \text{si } m_\lambda < m_\lambda^{lim} \\ 0 & \text{sinon.} \end{cases} \quad (3.24)$$

avec paramètres  $m_\lambda^{lim}$ ,  $\alpha$  et  $\beta$  sont les paramètres de complétude du champ de vue.

Pour les déterminer, nous avons simulé quelques champs de vue répartis autour du disque, en utilisant l'extinction déduite par Marshall et al. (2006) et la densité stellaire de Fux (1999). On obtient ainsi une distribution théorique des magnitudes  $P(m_\lambda)$ , on peut ainsi ajuster les paramètres de complétudes pour que les distributions cumulées simulées et observées soient cohérentes entre elles :

$$P(S | m_{sim}) \propto P(m_{obs} | S) / P(S). \quad (3.25)$$

Il apparaît que le couple  $(2, 2)$  pour  $(\alpha, \beta)$  est représentatif pour le catalogue UKIDSS, tandis que le couple  $(10, 1)$  est plus approprié pour 2MASS. La magnitude limite est évaluée pour chacune des bandes photométriques infrarouges, pour chacun des champs de vue analysés. Grâce à des simulations, on a déterminé qu'en ajoutant 2 magnitudes à la valeur maximale de la distribution  $P(m | S)$ , on obtient une bonne estimation de  $m^{lim}$ .

### Calcul de $P(S | A_0, D)$

Pour calculer  $P(S | A_0, D)$ , nous utilisons les étoiles de référence de notre diagramme HR empirique (voir 3.1.1). Connaissant les magnitudes absolues  $M$  pour chacune des étoiles de référence  $r$ , nous pouvons calculer la magnitude apparente des étoiles de référence pour chaque couple  $(A_0, D)$ . La contribution d'une étoile  $r$  à la distribution totale  $P(S | A_0, D)$  est donc :

$$P(S | A_0, D, r) = \prod_{\lambda} P(S | A_0, D, M_{\lambda}^r) = \prod_{\lambda} P(S | m_{\lambda}^r), \quad (3.26)$$

avec  $\lambda \in \{J, H, K\}$  car on base la complétude de notre échantillon sur les catalogues en infrarouge. Pour obtenir la complétude, on somme la contribution de l'ensemble des étoiles du diagramme HR de référence.

$$P(S | A_0, D) = \sum_r P(S | A_0, D, r) \times P(r) \quad (3.27)$$

### Prise en compte de la coupure en $\sigma_{\lambda}$ pour UKIDSS

Lorsque nous sélectionnons les étoiles dans les catalogues, nous effectuons une coupure sur l'incertitude photométrique. Les étoiles dont l'incertitude  $\sigma_{\lambda}$  est supérieure à  $\sigma_{\max} = 0.3\text{mag}$  ne sont pas analysées. Or l'incertitude sur la magnitude  $\sigma_{\lambda}$  suit une loi de puissance avec la magnitude  $m_{\lambda}$ . Sur la Figure 3.7 on constate que faire une coupure en incertitude revient à couper en magnitude et cette magnitude peut être en deçà de la magnitude limite  $m_{\lambda}^{lim}$  précédemment estimée. Pour obtenir une distribution valable de  $P(S | A_0, D)$ , nous devons ajouter un terme correctif  $C(m_{\lambda}, \sigma_{\max})$  à la valeur de  $P(S | m_{\lambda})$ .

Pour le calculer, nous commençons par modéliser la relation entre  $\sigma_{\lambda}$  et  $m_{\lambda}$  en ajustant la loi

$$\sigma_{\lambda} = A + \exp(B \times m_{\lambda} + C), \quad (3.28)$$

sur les données observées d'un champ de vue, pour toutes les bandes photométriques. Nous pouvons ainsi estimer la magnitude  $m_{\sigma_{\max}}$  correspondant à la coupure à  $\sigma_{\max}$  et l'incertitude théorique  $\sigma_{\lambda}^r$  sur les magnitudes apparentes de nos étoiles de référence. Ce faisant, nous calculons la valeur corrective :

$$C(m_{\lambda}^r, \sigma_{\max}) = \int_{-\infty}^{m_{\sigma_{\max}}} \mathcal{G}(m_{\lambda}^r - m', \sigma_{\lambda}^r) dm'. \quad (3.29)$$

Pour calculer la distribution de la complétude, nous utilisons donc l'équation :

$$P(S | A_0, D) = \sum_r \left[ P(r) \times \prod_{\lambda} P(S | m_{\lambda}^r) \times C(m_{\lambda}^r, \sigma_{\max}) \right]. \quad (3.30)$$

La distribution de  $P(S | A_0, D)$  pour le champ de vue simulé est représentée en Figure 3.8. Comme on pouvait s'y attendre, la probabilité de détection est plus grande à faible distance et faible extinction. Cette distribution de la complétude nous permet de calculer la distribution de  $P(A_0, D)$  (en utilisant l'équation 3.18) relative aux champs de vue. La Figure 3.9 nous montre le résultat de cette correction, la relation entre la distance et l'extinction s'y démarque clairement et les probabilités le long de la relation évoluent peu avec la distance.



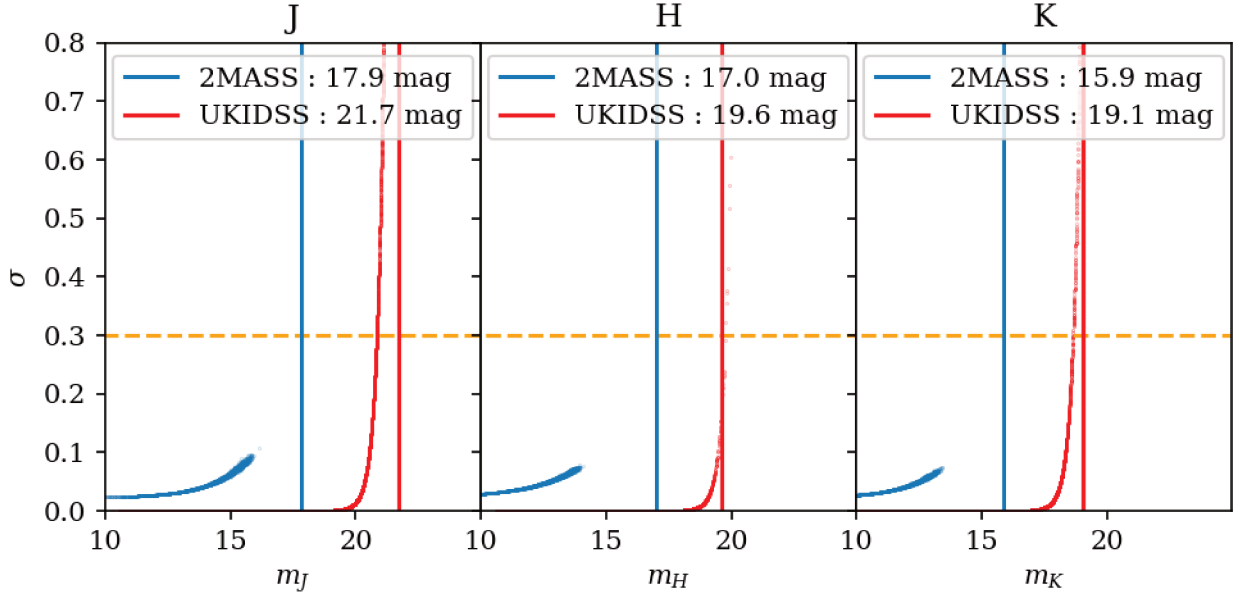


FIGURE 3.7 – Répartition des incertitudes photométriques en fonction de la magnitude, pour les bandes  $J$ ,  $H$  et  $K$ , pour un champ de vue simulé à  $l = 9.6^\circ$ , chaque point correspond à une étoile. Les points rouges correspondent à 2MASS et les points bleus à UKIDSS. Les lignes verticales représentent la magnitude limite calculée par FEDReD et la ligne horizontale correspond à  $\sigma_{\max}$ .

### 3.3 Calcul de la relation entre $A_0$ et $D$

Nous voulons maintenant obtenir la relation reliant l'extinction avec la distance  $A_0(D)$  à partir des résultats de la déconvolution. Cette relation doit correspondre au maximum de probabilités de la distribution  $P(A_0, D)$ , avec la contrainte physique de croissance de  $A_0(D)$ .

Nous utilisons un processus de Monte-Carlo pour obtenir 5000 relations  $A_0(D)$  monotones en utilisant l'algorithme suivant. Pour chaque distance  $D_i$  tirée dans un ordre aléatoire dans l'intervalle d'étude, on choisit une valeur d'extinction  $A_{0i}(D_i)$ . Le choix de cette valeur se fait suivant la distribution  $P(A_0, D_i)$  que l'on tronque en imposant  $A_0(D_i^-) \leq A_0 \leq A_0(D_i^+)$  où  $D_i^-$  et  $D_i^+$  sont les distances déjà traitées directement inférieures et supérieures à  $D_i$ . On répète l'opération jusqu'à avoir tiré toutes les distances  $D_i$ . Pour amorcer l'algorithme on prend  $A_0(0) = 0$  et  $A_0(D_{\max}) = A_{0\max}$ .

Nous évaluons la qualité d'une relation  $A_0(D)$  en calculant le logarithme de la vraisemblance  $\mathcal{L}$  :

$$\log(\mathcal{L}) = \sum_i \log(P(A_0(D_i), D_i)), \quad (3.31)$$

et nous conservons uniquement les 100 meilleures solutions Monte Carlo. Nous utilisons ensuite la librairie `cobs` de Ng & Maechler (2007) pour ajuster une spline contrainte médiane, dans ces 100 MCS, pour obtenir la relation  $A_0(D)$  correspondant au meilleur ajustement de  $A_0(D)$  (BF) du champ de vue. Nous ajustons aussi deux autres splines contraintes correspondant aux quantiles 0.159 et 0.841 pour obtenir l'intervalle de confiance à  $1\sigma$  du résultat. Les relations obtenues pour notre champ de test sont présentées en Figure 3.10.

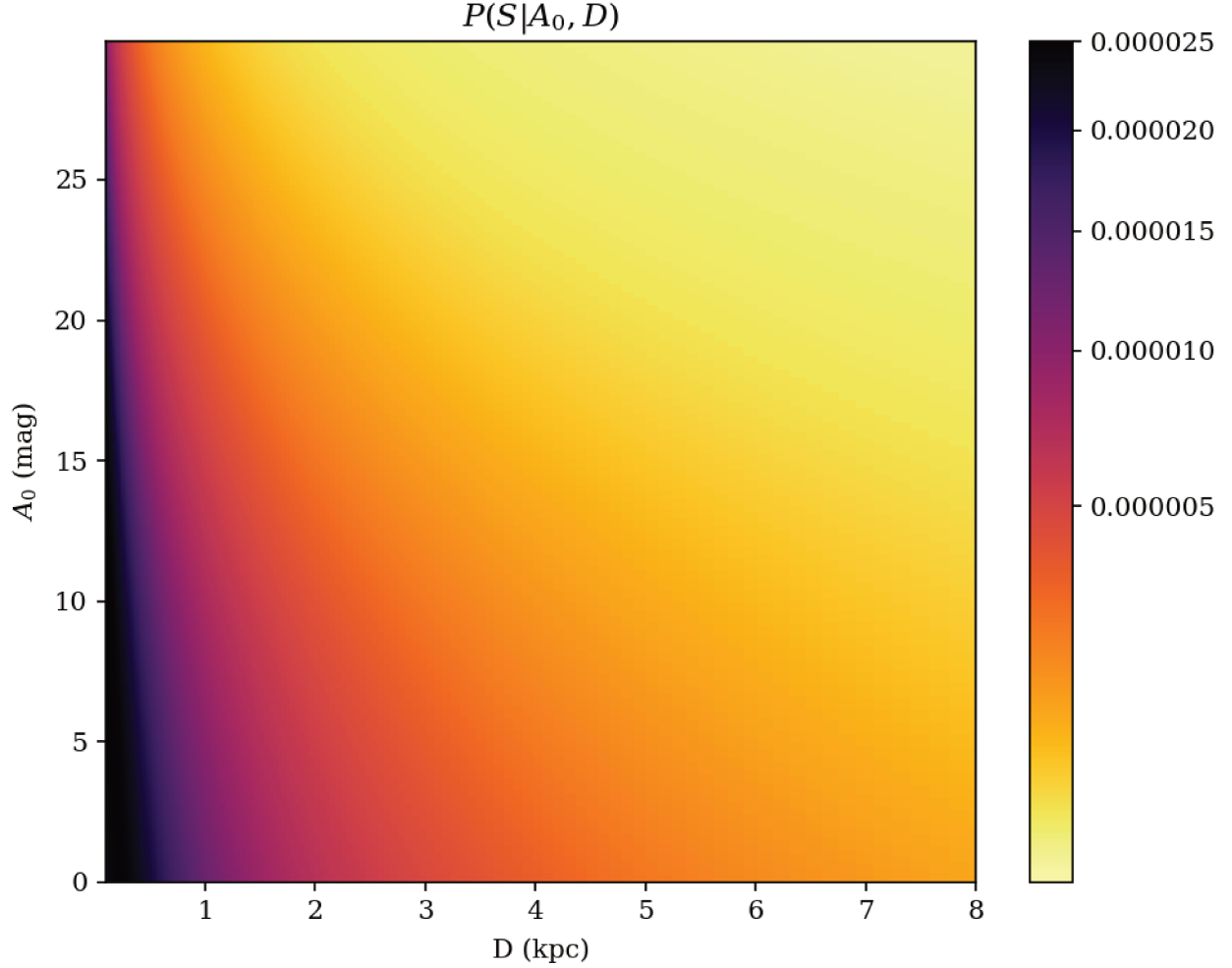


FIGURE 3.8 – Distribution des probabilités  $P(S | A_0, D)$  pour le champ de vue simulé à  $l = 9.6^\circ$

## 3.4 Implémentation

L'algorithme FEDRED est gourmand en ressources informatiques. Pour pouvoir l'utiliser sur un grand nombre de champs de vue, nous avons dû optimiser les calculs, mais aussi l'espace disque et le chargement en mémoire. L'algorithme fonctionnant en deux étapes distinctes, le traitement des données a logiquement été scindé en deux exécutions distinctes. La première partie traite les étoiles individuellement, tandis que la seconde procède à la déconvolution.

### 3.4.1 Calculs

Le temps moyen de calcul de la densité individuelle d'une étoile ( $P(\mathbf{O} | A_0, D)$ ) est de 1,5s. Les traitements des étoiles étant indépendants les uns des autres, nous pouvons utiliser des threads séparés. Grâce à la classe `java.lang.Thread` et à l'implémentation `java.util.concurrent.Future`, on obtient un bon scaling sur les grappes de calcul. Cela nous a d'ailleurs permis d'obtenir du temps CPU sur la grappe MesoPSL.

Pour la déconvolution, la partie la plus demandeuse en temps CPU est le calcul de la fonction de sélection. On utilise donc la programmation fonctionnelle de Java 8 pour calculer en parallèle la contribution de chaque étoile de référence. On utilise l'opération terminale `Stream.collect` qui utilise une stratégie en arbre pour faire la sommation finale et ainsi utiliser au mieux les cœurs de calcul disponibles. On utilise la même stratégie pour paralléliser le calcul de chaque itération

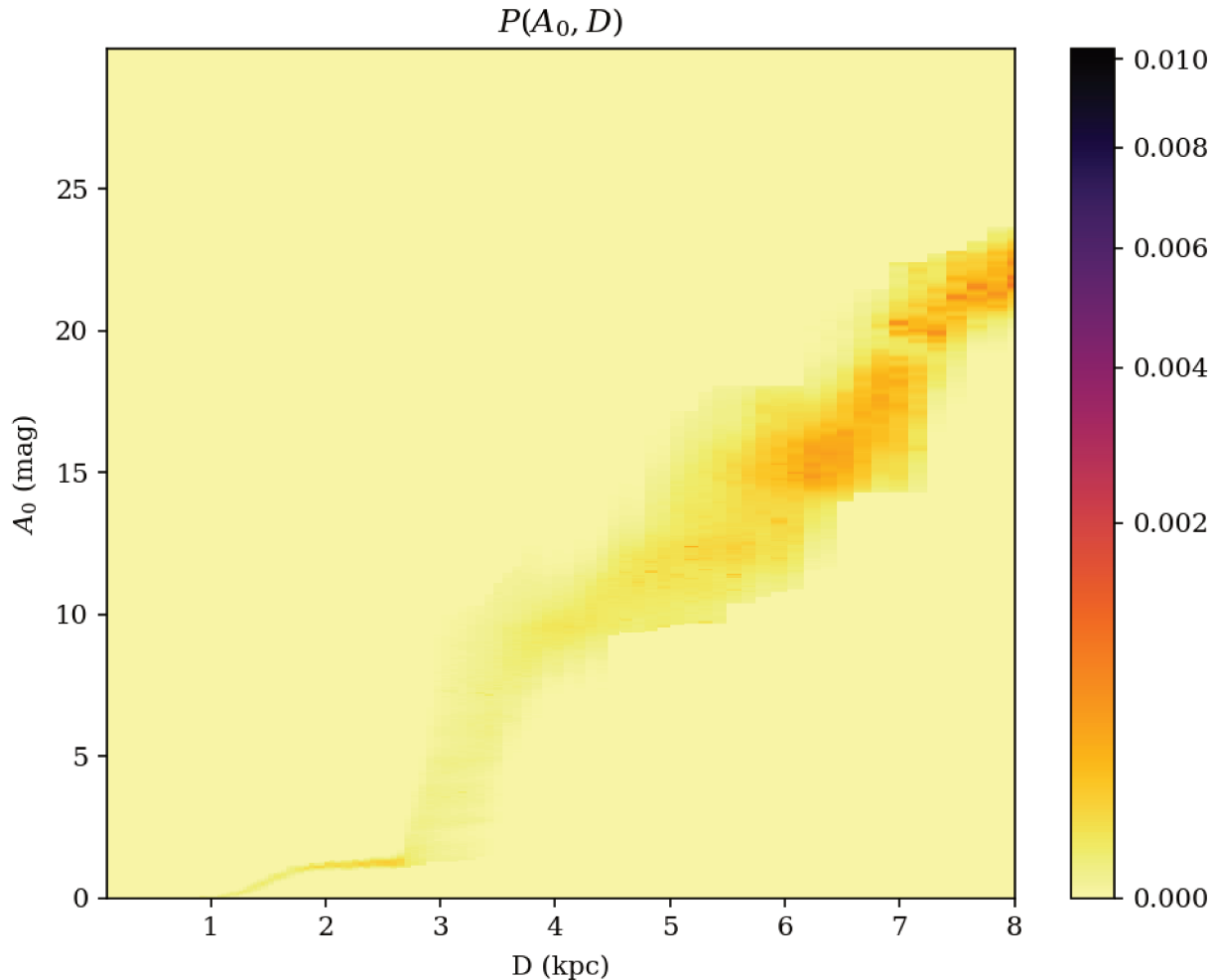


FIGURE 3.9 – Résultat final de la déconvolution du champ de vue simulé à  $l = 9.6^\circ$ , après correction de la complétude.

de la déconvolution.

### 3.4.2 Exécutions et utilisation des grappes

La conception originelle de FEDReD imposait d’analyser chaque champ de vue indépendamment pour les deux parties de l’exécution (densité individuelle et déconvolution). Cela posait deux problèmes, le premier est le fait de devoir lancer un très grand nombre de jobs (nombre de champ  $\times 2$ ). Le second étant le fait qu’en rendant les champs indépendants, on calcule les distributions individuelles plusieurs fois si l’on étudie des champs de vue qui se recouvrent.

C’est pourquoi le calcul des densités individuelles se fait indépendamment des déconvolutions. FEDReD calcule les vraisemblances  $P(O | A_0, D)$  de l’ensemble des étoiles qu’on lui fournit en entrée et stocke les résultats dans des fichiers séparés en fonction de la zone du ciel à laquelle les étoiles appartiennent. On peut ainsi limiter le nombre de jobs de densités à quelques centaines (pour utiliser plusieurs nœuds simultanément) plutôt que plusieurs milliers. Dans un second temps, FEDReD va uniquement récupérer les résultats des étoiles dont il a besoin (sur le principe d’une requête), pour effectuer la déconvolution.

En procédant ainsi, on peut faire de nouvelles déconvolutions sur autant de champs de vue que nécessaire sans tout recalculer. Par ailleurs, on diminue le nombre de jobs et donc l’encombrement des files d’attente des calculateurs. On a par ailleurs utilisé le principe de tableau de

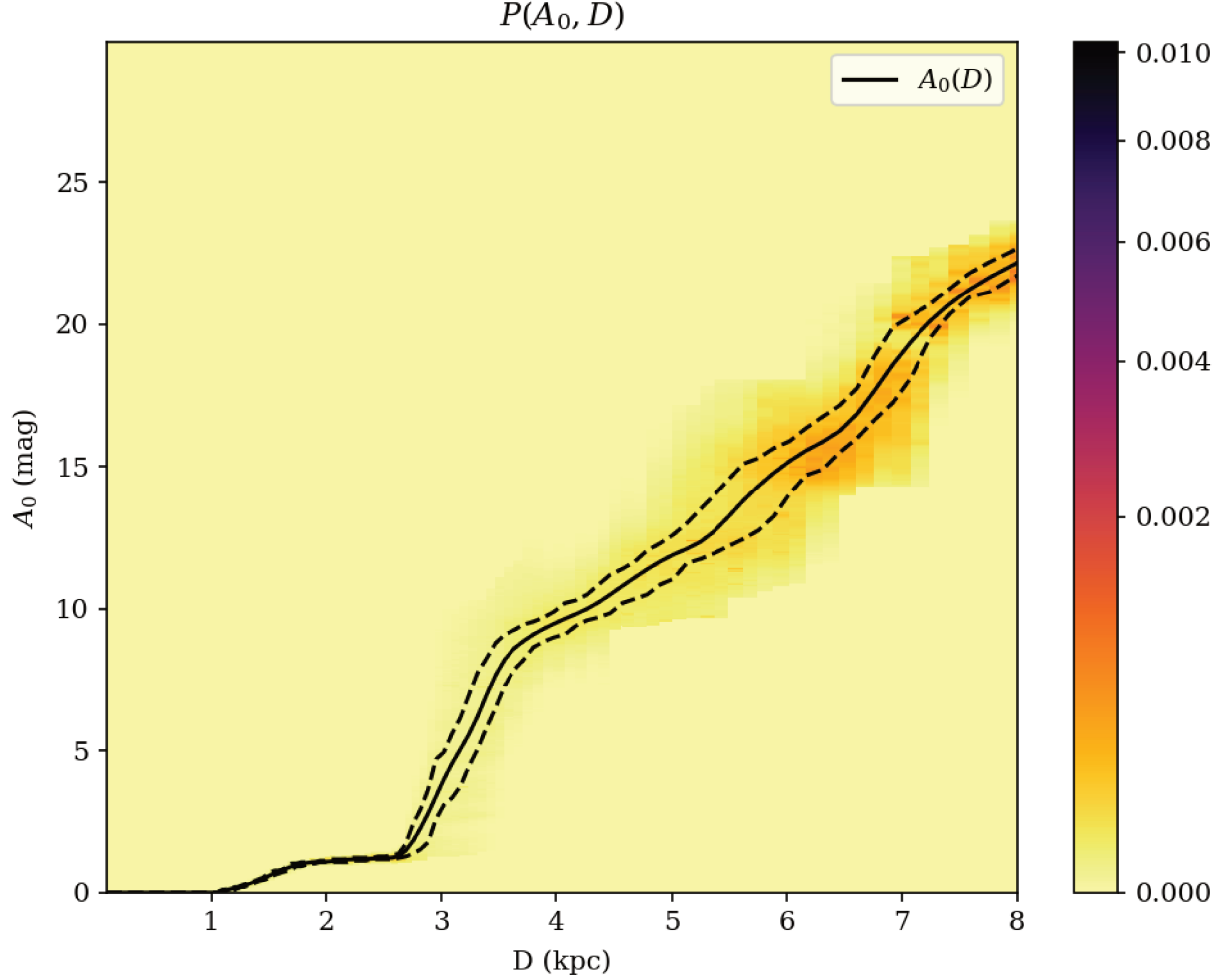


FIGURE 3.10 – Relation  $A_0(D)$  et intervalle de confiance obtenus par processus de Monte-Carlo.

tâche (`ARRAY_TASKS`) nous permettant ainsi de créer des files d'exécution de plusieurs centaines de jobs plutôt que de devoir générer des fichiers de lancement pour chacun d'eux.

### 3.4.3 Stockage et accès

Le stockage des densités individuelles est un autre enjeu d'optimisation. En effet, une unique distribution  $P(\mathbf{O} \mid A_0, D)$  est constituée d'un tableau de 247 par 600 nombres flottants doubles précisions (stockés sur 8 octets). Cela représente plus d'un mégaoctet par étoile observée et donc plusieurs gigaoctets par champ de vue. Stocker les résultats dans un format adéquat revêt donc un double intérêt : limiter l'espace disque occupé (plusieurs teraoctets), mais aussi accélérer la lecture de ces données.

Pour répondre à ces besoins, nous utilisons le format de fichier hdf5. Il présente de nombreux avantages, notamment le stockage sous la forme d'une arborescence. Nous avons choisi d'organiser les résultats en utilisant le découpage d'une sphère (ici la sphère céleste) par healpix décrit Gorski et al. (2005). Autrement dit, chaque fichier correspond à un healpix. Les distributions individuelles de chaque étoile incluses dans ce healpix sont stockées sous la forme d'un dataset. Nous avons choisi l'ordre 8 des healpixs ce qui nous donne des fichiers de quelques gigaoctets au plus.

Lors de la déconvolution d'un champ de vue, le programme va uniquement copier les healpixs requis sur le disque local (/scratch) limitant ainsi les temps de téléchargement. En outre, les

données stockées en hdf5 peuvent être chargées en mémoire très rapidement (environ 100 fois plus vite que pour du csv.gz décompressé à la volée).

Enfin le format hdf5 nous permet de compresser *in situ* les datasets (filtre gzip de niveau 1), réduisant d’un facteur 3 l’espace disque occupé avec une augmentation négligeable du temps de lecture.

#### 3.4.4 Temps de calcul et gain

Au cours de cette thèse nous avons dû faire tourner les codes sur des sections du disque galactique plus ou moins importantes selon les tests et validations effectuées, le nombre total d’heures de calcul utilisé étant d’environ 360 000 hCPU. Grâce aux différentes optimisations décrites ci-dessus, nous avons considérablement réduit notre impact sur les grappes de calculs. En effet la première portion de carte de 30° de large que nous avons produite initialement avait utilisé 50 000 hCPU et occupé un espace disque de 1.8 To tandis que la carte complète présentée au chapitre 5 en a demandé 60 000 hCPU pour un espace disque occupé de 2.7 To soit un gain d’un ordre de grandeur en temps CPU grâce aux optimisations que nous avons mises en œuvre.

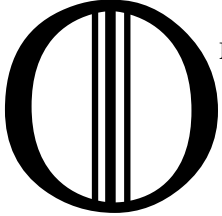
#### 3.4.5 Évolution du code et ajout de catalogue d’étoiles

Le code de FEDReD était à l’origine très rigide concernant l’utilisation de la photométrie, seules les trois bandes infrarouges étaient prévues. Dans le but de le rendre adaptable et de faciliter sa maintenance, nous en avons repensé plusieurs parties. Nous avons ainsi ajouté plusieurs **enumeration**. L’une d’entre elles prend en charge les bandes photométriques (longueur d’onde, données en flux ou en magnitude, etc...) et l’autre recense les catalogues utilisables et recense pour chacun d’eux les bandes photométriques à utiliser. À l’initialisation de l’algorithme, il suffit de préciser quel catalogue on analyse, FEDReD se charge du reste.

Cette conception rend le code très flexible et surtout il est facile d’ajouter des nouveaux catalogues et de nouvelles bandes photométriques.

# Chapitre 4

## Affinage de l’algorithme, Tests et Validations



N a décrit au chapitre précédent le fonctionnement de FEDReD tel qu’il est actuellement. Il a subi beaucoup de modifications depuis la démonstration de faisabilité par Fourtune-Ravard (2016). Pour développer ces modifications, nous avons besoin de jeux de données sur lesquels appliquer notre algorithme. À cette fin, nous avons utilisé un champ de vue simulé ainsi que des champs de vue réels construits à partir des crossmatches 2MASS-GDR2 et UKIDSS-GDR2.

À chaque nouvelle modification apportée à FEDReD, nous analysons tous nos champs de vue tests et nous utilisons diverses méthodes pour valider (ou non) les modifications. On vérifie ainsi le nombre d’itérations pour la convergence des champs et la finesse de l’intervalle de confiance sur la distribution  $P(A_0, D)$  et sur le CMD.

Ce processus de validation s’est révélé nécessaire, car un grand nombre de changements que nous avons testés améliorent certains résultats tout en en détériorant d’autres.

### 4.1 Simulation de champs de vue

Pour tester FEDReD, il nous faut un champ de vue de référence pour lequel on connaît la véritable relation entre la distance et l’extinction  $A_0(D)$ , ainsi que la distribution d’étoiles avec la distance  $P(D)$ . Nous avons donc simulé un champ de vue que nous avons par la suite analysé avec FEDReD. La technique pour simuler ce champ de vue se base sur celle développée par Fourtune-Ravard (2016).

Pour réaliser cette simulation, nous avons besoin d’une référence de population d’étoiles du disque galactique. Nous utilisons donc le diagramme HR de référence que nous avons construit (voir section 3.1.1). On suppose dans l’ensemble de cette thèse que les étoiles du voisinage solaire sont représentatives du disque de la Voie lactée.

On choisit comme « modèle » de champ d’extinction la direction  $l = 9.6^\circ$  et  $b = 0^\circ$  avec une largeur de  $0.16^\circ$ . Ce champ de vue est notamment étudié dans Babusiaux et al. (2014). On tire 100000 étoiles dans notre diagramme HR de référence en suivant la distribution des poids de chaque point du diagramme (voir section 3.1.1).

On dispose donc d’étoiles dont on connaît la magnitude absolue. Pour répartir ces étoiles le long de la ligne de visée, on utilise la distribution des distances obtenues par Fux (1999) et on applique la relation  $A_0(D)$  obtenue par les mêmes travaux. On représente en Figure 4.1 ces deux distributions. Pour chaque étoile simulée, on tire au hasard une distance, en suivant la distribution des distances, puis on calcule l’extinction  $A_0$  correspondante en utilisant la loi

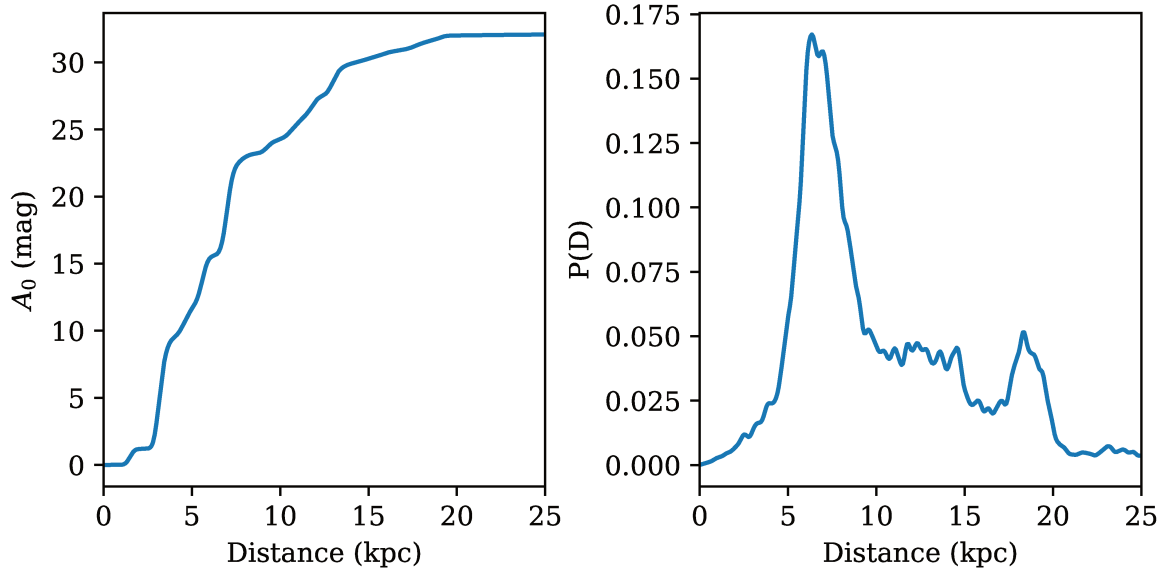


FIGURE 4.1 – Partie gauche : Relation  $A_0(D)$  entre l’extinction et la distance obtenue par Babusiaux et al. (2014). Partie droite : Distribution de la distance des étoiles du champ de vue, obtenue par Fux (1999).

de référence à laquelle on ajoute une variation aléatoire suivant une lois lognormale de 0.05mag d’écart type. Enfin, on calcule la magnitude apparente de l’étoile à l’aide de l’équation (1.2) et on utilise le modèle décrit par l’équation (3.1) pour calculer les coefficients d’extinction. Pour calculer et appliquer les incertitudes photométriques, on utilise la loi décrite à l’équation (3.28).

Connaissant les magnitudes apparentes de l’étoile simulée, on calcule la probabilité  $P(S | m_i)$  (équation 3.24) qu’elle soit observée par le relevé infrarouge. Comme nous réalisons une simulation de données UKIDSS, on prend  $\alpha = 2$  et  $\beta = 2$  et  $m_J^{lim} = 22$ ;  $m_K^{lim} = 19.5$  et  $m_H^{lim} = 19$  pour magnitude limite. Enfin on tire un nombre aléatoire pour que l’on compare à  $P(S | m_i)$  pour savoir si on conserve ou non l’étoile.

Concernant l’information GDR2, on procède de la même manière que pour la photométrie infrarouge pour calculer les magnitudes apparentes dans les bandes  $G$ ,  $G_{BP}$  et  $G_{RP}$ . Puis on calcule  $P(S_{GDR2} | m_i)$  en utilisant le modèle de complétude (3.24), avec  $\alpha = 2$  et  $\beta = 2$  et  $m_G^{lim} = 20.7$ ;  $m_{G_{BP}}^{lim} = 20.9$  et  $m_{G_{RP}}^{lim} = 19.5$ . On tire à nouveau un nombre aléatoire pour déterminer si l’on conserve l’information GDR2, ou non.

Nous avons donc à notre disposition un catalogue simulé (mock catalogue) avec des magnitudes apparentes et les incertitudes associées conformes au catalogue UKIDSS apparié avec GDR2. Pour générer l’ensemble des champs d’avertissement, décrit aux sections 2.2 et 2.3, on utilise directement les véritables champs de la requête sur le catalogue et on remplace la photométrie par les valeurs que nous avons calculées.

Les principales modifications apportées depuis les travaux de Fourtune-Ravard (2016) sont l’implémentation des coefficients d’extinction polynomiale de Danielski et al. (2018) et l’utilisation du diagramme HR de référence (voir section 3.1.1). Le CMD du champ de vue simulé obtenu est présenté en Figure 4.2. On y distingue parfaitement la traînée du Red Clump.

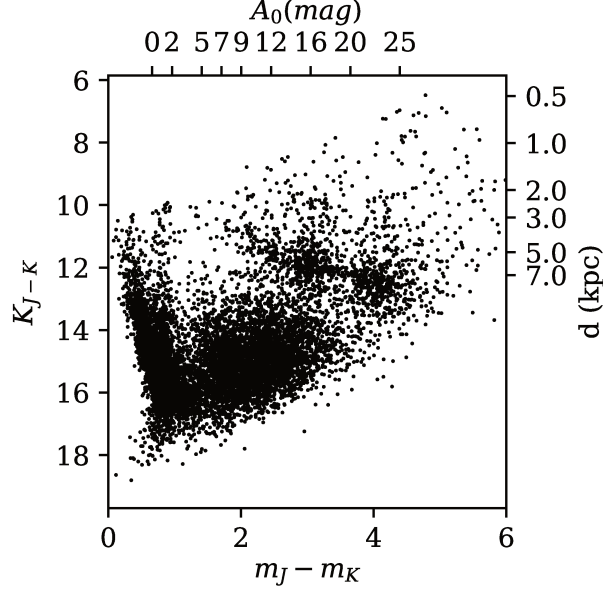


FIGURE 4.2 – CMD du champ de vue simulé à l’aide de notre diagramme HR de référence, pour un champ de vue dans la direction  $l = 9.6^\circ$ ,  $b = 0^\circ$  avec la distribution de distance des étoiles  $P(D)$  et la relation  $A_0(D)$  obtenue par le modèle de Fux (1999)

## 4.2 Test de FEDReD

Nous avons utilisé FEDReD sur ce champ de vue simulé. Suivant les coupures décrites pour UKIDSS GDR2 au chapitre 2, l’algorithme conserve 9200 étoiles dont 2100 avec une parallaxe et 950 avec de la photométrie BP et RP. Les magnitudes limites infrarouges déduites par FEDReD sont  $m_J^{lim} = 21.7$ ;  $m_K^{lim} = 19.6$  et  $m_H^{lim} = 19.1$ . La distribution  $P(S | A_0, D)$  calculée de ce champ de vue est présentée en Figure 3.8.

### 4.2.1 Résultats pour la relation $A_0(D)$

La distribution  $P(A_0, D)$  obtenue par FEDReD est présentée en Figure 4.3. À l’aide de cette distribution, on a généré aléatoirement des solutions de Monte Carlo (MCS) parmi lesquelles on conserve les 1000 meilleures. On a ensuite ajusté une spline contrainte médiane à l’aide de la librairie *cobs* dans ces MCS pour obtenir le meilleur ajustement de  $A_0(D)$  (BF) et deux autres splines contraintes pour obtenir un intervalle de confiance à  $1\sigma$ .

On a ajouté le BF ainsi que l’intervalle de confiance sur la distribution  $P(A_0, D)$  en Figure 4.3. On a aussi tracé la relation utilisée pour réaliser la simulation. Le graphique du bas correspond au résidu normalisé par le sigma correspondant ( $\sigma_{inf}$  si  $A_{0calc} > A_{0theo}$ ;  $\sigma_{sup}$  sinon). La distribution des résidus est présentée en Figure 4.4, où l’on a ajouté une gaussienne de médiane 0 et écart type 1. On constate que la distribution des résidus n’est pas symétrique. On effectue un test de Kolmogorov-Smirnov sur ces résidus avec pour hypothèse  $H_0$  qu’ils suivent une loi normale. On obtient une p-valeur de 0.08, non conclusive, qui s’explique par la forte corrélation entre les résidus due à la contrainte croissante.

### 4.2.2 Résultats distribution de distance

En marginalisant la distribution  $P(A_0, D)$ , on obtient la distribution des étoiles en distance  $P(D)$  du champ de vue. On présente en Figure 4.5 cette distribution pour le champ de vue



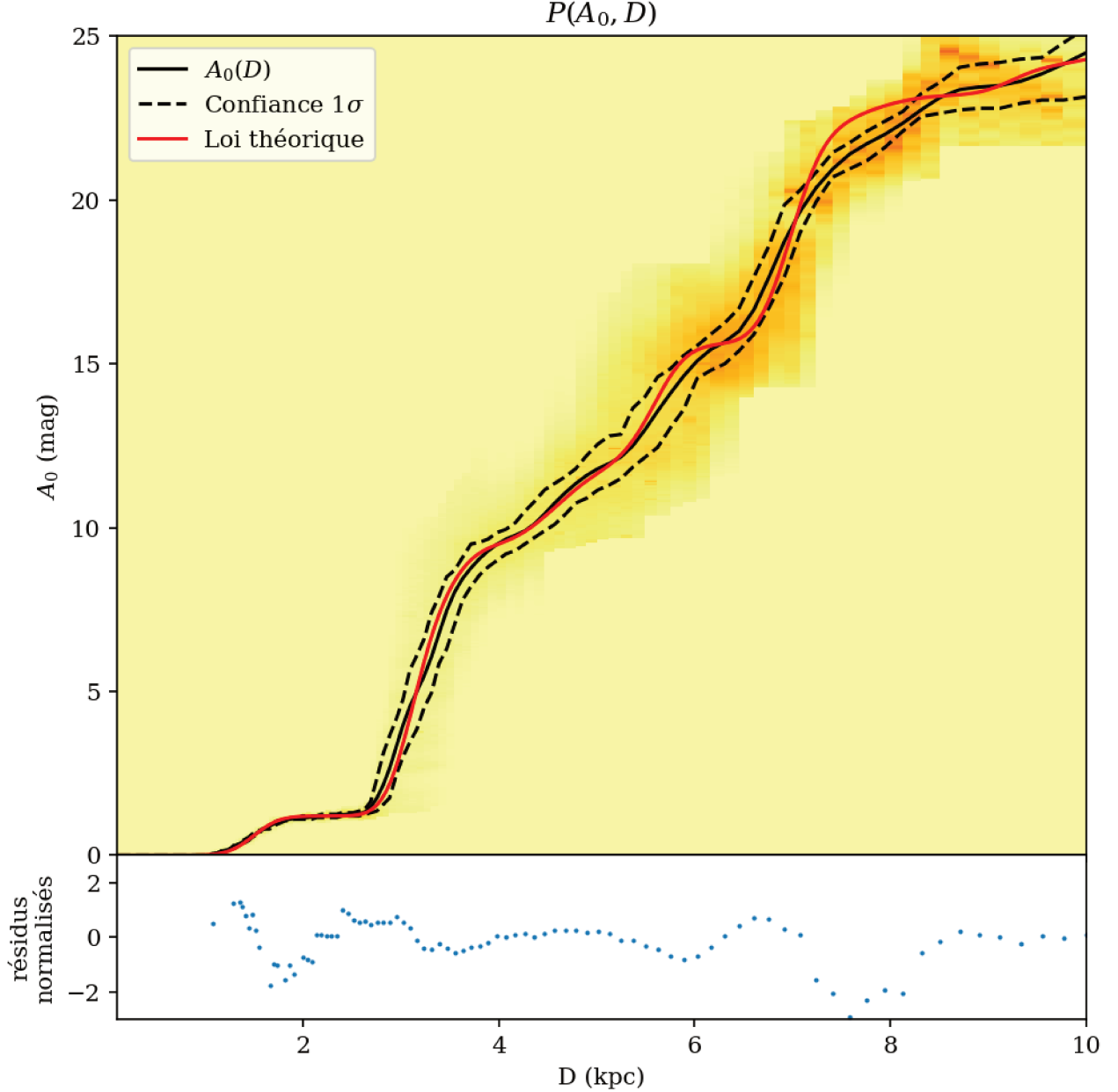


FIGURE 4.3 – *Haut* : Distribution de probabilité  $P(A_0, D)$  pour le champ simulé à  $l = 9.6^\circ$  obtenue par FEDReD. La ligne continue noire représente le BF du champ de vue, et les lignes en pointillés représentent l'intervalle de confiance à  $1\sigma$ . La ligne rouge correspond à la relation  $A_0(D)$  utilisée pour simuler le champ de vue. *Bas* : Résidu normalisé pour chaque bin de distance.

simulé. Nous n'avons pas encore trouvé de méthode satisfaisante pour normaliser la distribution  $P(D)$ , pour pouvoir comparer nos résultats à la distribution de référence (Figure 4.1), nous avons renormalisé la référence de façon à ce que l'on ait :

$$\int_0^{D_{max}} P(D)_{calcul} dD = \int_0^{D_{max}} P(D)_{ref} dD. \quad (4.1)$$

On constate que la distribution en distance obtenue suit la loi théorique renormalisée.

Nous n'avons pas encore conçu de méthode satisfaisante pour calculer l'incertitude de nos résultats. On a essayé d'appliquer un bootstrap, en faisant la déconvolution d'un champ en n'utilisant qu'un échantillon aléatoire des résultats individuels. On répétait mille fois l'opération pour

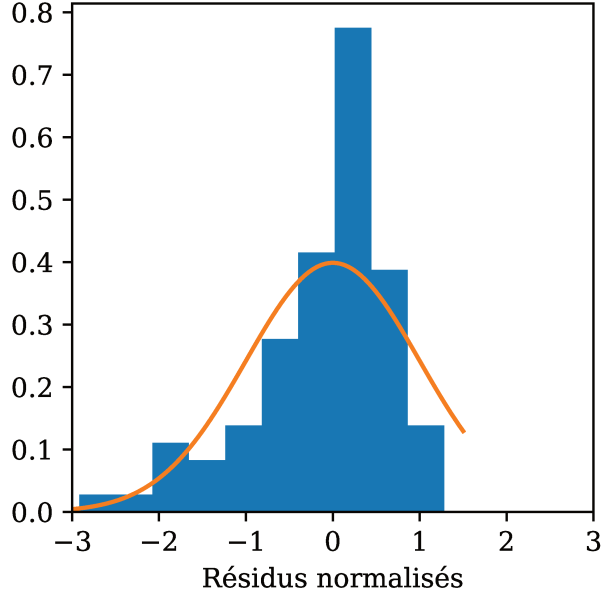


FIGURE 4.4 – Distribution des résidus normalisés entre le BF du champ de vue et la relation  $A_0(D)$  utilisée pour réaliser la simulation

un même champ de vue pour obtenir la variation statistique sur les résultats et en déduire l'incertitude. Cette technique, très gourmande en temps de calcul, produisait cependant des barres d'erreurs sous-estimées.

Nous étudions actuellement une technique utilisant les MCS. Nous utilisons ces solutions pour construire des priors de  $P(A_0 | D)$  que l'on utilise pour de nouvelles déconvolutions et ainsi obtenir suffisamment de résultats pour appliquer des méthodes statistiques et déduire une incertitude. Cette approche semble prometteuse mais il nous reste à déterminer une méthode de normalisation des différents résultats.

### 4.3 Champs de validation et affinage

La faisabilité de l'algorithme avait été démontrée par Fourtune-Ravard (2016) notamment à l'aide d'une simulation similaire à celle décrite précédemment. Il avait aussi été testé sur 2 champs de vue réels, mais avec un prior sur  $P(A_0 | D)$ . Il produisait de bons résultats, mais n'avait jamais été testé à grande échelle.

Pour les travaux relatés dans ce manuscrit, nous ne souhaitons plus être dépendant d'un autre modèle et nous ne souhaitons donc plus utiliser de prior extérieur pour initialiser la déconvolution. C'est pourquoi nous utilisons un prior plat (voir section 3.2.4). L'utilisation de ce dernier a été testée sur la simulation et les champs réels et donne de bons résultats.

Cela a eu pour effet de faire apparaître l'un des problèmes majeurs que nous avons rencontré, la dégénérescence des petites naines froides. Si on ne connaît pas leurs parallaxes, alors elles présentent une bimodalité dans leurs distributions de probabilité. Elles apparaissent soit comme étant des naines moins froides et plus éteintes, ou alors elles sont interprétées comme étant des géantes rouges éteintes. Dans ce cas, elles indiquent une très forte extinction à faible distance. On a représenté la distribution  $P(A_0, D | O) \times P(D)$  d'une naine rouge simulé à 0.1kpc sans extinction en Figure 4.6, on y constate cette forte bimodalité.

Pour affiner FEDReD et résoudre les différents problèmes rencontrés, comme celui des naines froides, nous avons sélectionné un jeu de champ de vue test pour affiner nos analyses. Chacun de

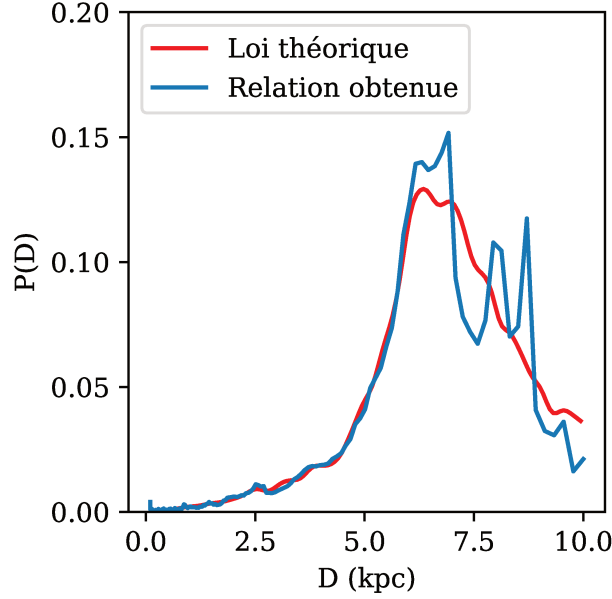


FIGURE 4.5 – Distribution des étoiles en distance obtenue par FEDReD pour le champ de vue simulé à  $l = 9.6^\circ$

ces champs ayant été choisi pour une raison particulière :

- $l = 001.68^\circ$  : Champ dirigé vers le centre galactique présentant un fort encombrement dans tous les catalogues utilisés dans notre étude.
- $l = 009.60^\circ$  : Champ étudié par Babusiaux & Gilmore (2005); Babusiaux et al. (2014), dont on dispose déjà de certaines contraintes concernant l'extinction et la distribution des étoiles avec la distance.
- $l = 032.40^\circ$  : Champ présentant des problèmes de convergence lorsqu'on l'étudie avec UKIDSS
- $l = 036.88^\circ$  : Champ présentant une traînée du Red Clump dense et bien visible, permettant une validation visuelle aisée.
- $l = 048.96^\circ$  : Champ contenant un Maser Reid et al. (2014), ces derniers sont généralement associés à des zones d'extinction forte.
- $l = 049.20^\circ$  : Champ présentant une forte discontinuité dans la traînée du Red Clump.
- $l = 190.00^\circ$  : Champ vers l'anti-centre galactique, très peu dense.

### 4.3.1 Affinage de la méthode

La partie de l'algorithme calculant les distributions individuelles a toujours fourni des résultats satisfaisants, peu d'améliorations ont été nécessaires, si ce n'est l'implémentation de l'utilisation des données GDR2 ainsi que l'utilisation des coefficients d'extinctions polynomiaux. On a aussi implémenté les optimisations décrites à la section 3.4.

En revanche la déconvolution, et le traitement de la distribution  $P(A_0, D | S)$ , ont demandé plusieurs raffinements successifs avant d'être satisfaisants. On décrit ici les principales améliorations et techniques mises en application.

#### Nettoyage de $P(A_0, D)$ et le « bug des 3 kpc »

Le principal problème que nous avons rencontré vient des petites naines froides qui ajoutaient du signal à faible distance et forte extinction dans la distribution de  $P(A_0, D)$  (et par extension

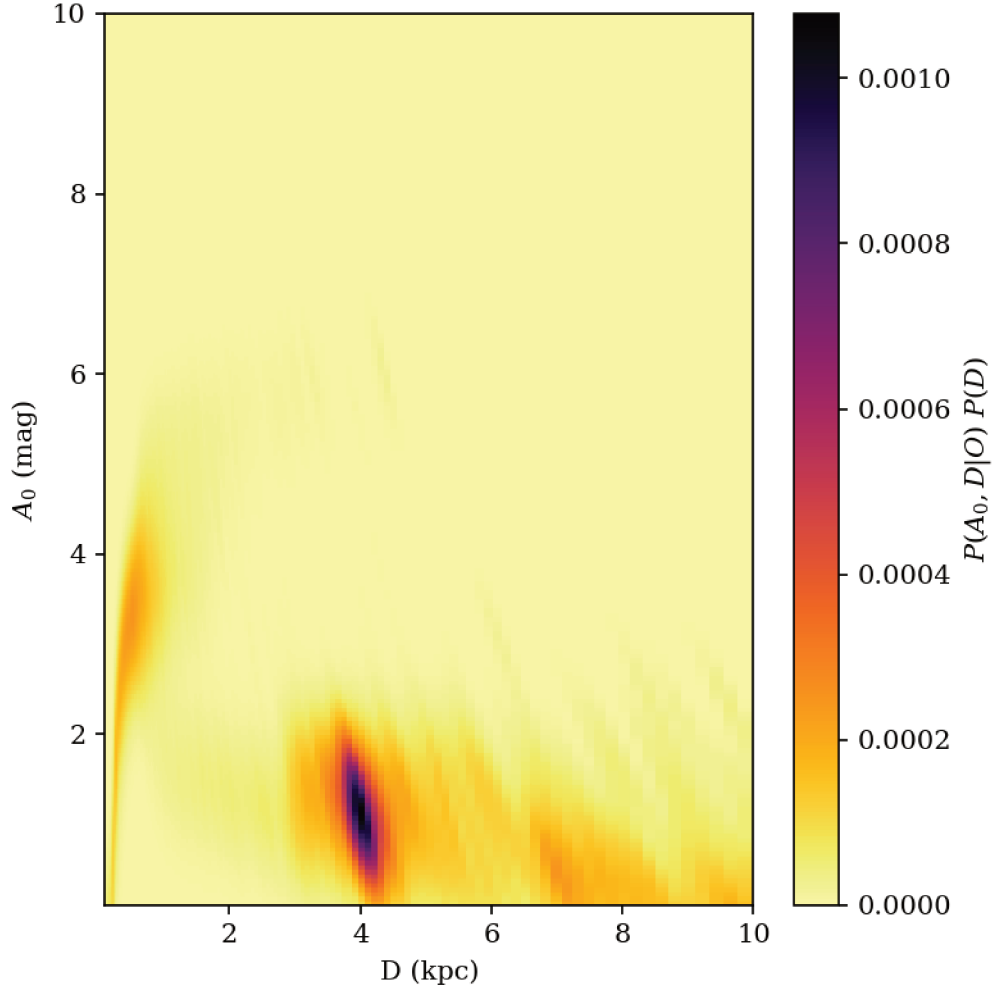


FIGURE 4.6 – Distribution de probabilité  $P(A_0, D | O) \times P(D)$  d’une naine rouge simulée, avec  $D = 0.1 \text{ kpc}$  et  $A_0 = 0 \text{ mag}$

dans  $P(A_0 | D)$ ). Ainsi lorsque l’on tentait d’ajuster la spline contrainte dans la distribution  $P(A_0 | D)$ , elle avait tendance à monter vers de hautes extinctions dès les faibles distances. Or cette dernière étant tenue d’être monotone, elle ne pouvait jamais retrouver la bonne solution.

La première idée a donc été d’utiliser des MCS, et de conserver les meilleurs pour définir une enveloppe de confiance au-delà de laquelle on mettait la distribution à 0. Cependant, ces MCS rencontraient le même problème que les splines contraintes et montaient trop vite vers les fortes extinctions. On les a donc empêchées de monter trop vite en retirant les MCS supérieurs à 5 mag d’extinction lorsque la distance était inférieure à 1 kpc. Or par un bug dans le code concernant les indices, cette restriction a été repoussée à 3kpc, ayant pour conséquence un saut brutal d’extinction à 3 kpc. Cela se traduisait, sur les cartes d’extinction, par un arc de cercle de très forte extinction centré sur le soleil et de rayon 3kpc, mais uniquement dans un arc de  $\sim 45^\circ$  dans la direction du centre galactique.

Ce « bug des 3 kpc » une fois découvert, nous a fait prendre conscience de la difficulté d’analyse des champs de vue vers le centre galactique. En effet, l’encombrement est très fort tout comme l’extinction, on a donc très peu d’information sur les parallaxes et l’algorithme converge difficilement à faible distance, ce qui perturbe les résultats sur l’ensemble de la ligne de visée.

### Prior $P(A_0 | D)$

Le nettoyage de la distribution  $P(A_0 | D)$  n'étant pas pleinement satisfaisant, nous avons tenté d'agir en amont de la déconvolution avec un prior sur la distribution  $P(A_0 | D)$ . Nous avons créé un prior avec une croissance privilégiée de l'extinction suivant la loi  $A_0 = 0.7 \times D$ .

Ce prior remplissait tellement son office que les résultats obtenus suivaient eux aussi préférentiellement une loi d'extinction linéaire. On a ainsi découvert qu'un prior trop fort avant la déconvolution perturbe les résultats. On a donc évité par la suite d'imposer de trop fortes contraintes à la déconvolution.

Le prior plat apparaît donc être la meilleure solution pour ne pas biaiser nos résultats, mais le problème de la dégénérescence des naines froides reste entier. Nous avons donc tronqué le prior  $P(A_0 | D)$  à faible distance, mais à une distance  $D$  donnée, les  $P(A_0 | D)$  sont tous équiprobables. Ce prior, bien que fort, ne perturbe pas l'algorithme, car il ne force pas l'emplacement du signal, il interdit juste une partie de l'espace  $(A_0, D)$  (voir Figure 3.5). Un certain nombre de MCS continue de monter jusqu'à cette limite, mais ce nombre est suffisamment faible pour pouvoir être éliminé en post-processing (voir section 5.2).

### MCS et splines contraintes

Initialement, la détermination du BF se faisait en ajustant une spline contrainte, à l'aide de *cobs* (Ng & Maechler 2007), dans la distribution  $P(A_0 | D)$ . Cette technique ne produisait pas de résultats pleinement satisfaisants, principalement à cause du signal produit par les naines froides. Cela étant, à part ce signal frauduleux, le reste de la distribution des  $P(A_0, D)$  paraissait correct et on y distinguait « à l'œil » la surdensité correspondant à l'évolution de l'extinction avec la distance. Les résultats de la déconvolution étaient donc viables et le problème résidait dans l'obtention du BF.

On a donc décidé de ne plus utiliser *cobs* pour obtenir la relation et de passer à une méthode plus frontale. On recherche 5000 MCS (voir section 3.3) dans la distribution  $P(A_0 | D)$  et on conserve les 100 meilleurs qui génèrent ainsi une enveloppe. On tire 1000 nouveaux MCS à l'intérieur de cette enveloppe et à nouveau on garde les 100 meilleurs. Enfin on utilise *cobs* pour ajuster une spline médiane à l'intérieur de ces 100 solutions et obtenir le BF. Par ailleurs, on utilise ces 100 MCS pour calculer des cartes d'extinction (voir chapitre 5).

### $D$ en fonction de $A_0$

L'ensemble des mesures décrites précédemment nous a permis d'obtenir un meilleur résultat pour la distribution  $P(A_0 | D)$  dans laquelle on tire des MCS et le BF. Cependant, nous nous sommes aperçu qu'en utilisant  $P(A_0 | D)$ , nous perdions de l'information. En effet, la division par la distribution en distance  $P(D)$  ôte l'information sur la quantité d'étoiles à une distance donnée. Or il existe une dégénérescence dans les résultats à grande distance dans la distribution des  $P(A_0 | D)$ . On peut observer, soit une augmentation de l'extinction avec la distance, ou une extinction constante. Pour lever la dégénérescence, nous avons besoin de connaître la distribution en distance des étoiles. Si la distribution d'étoiles est forte à grande distance, l'extinction a donc plutôt tendance à être constante. En revanche si le nombre d'étoiles décroît à grande distance, c'est que l'on ne les observe plus à cause de l'extinction. On préfère donc chercher des MCS dans la distribution de  $P(A_0, D)$ .

Cela nous permet aussi de tirer les MCS comme étant  $D(A_0)$  plutôt que  $A_0(D)$ . Cela évite une trop forte dispersion des résultats. En effet, à une distance donnée, la distribution des extinctions probables est très large ce qui permet aux MCS de s'éparpiller fortement pour certains intervalles de distance tandis que pour une extinction donnée, la distribution des distances probables est relativement fine. Cela nous permet d'éviter les MCS montant trop vite vers les fortes extinctions.

### Critère d'arrêt en distance

FEDReD travaille sur une discrétisation de l'espace  $(A_0, D)$ . L'intervalle en distance typique de cette grille est  $[0, 30]$  kpc. Cependant les catalogues ne fournissent pas de photométrie suffisamment profonde pour explorer des régions aussi éloignées. En l'absence de coupure efficace en distance, les cartes que nous obtenions divergeaient à grande distance.

Pour parer à cela, nous avons implémenté une coupure automatique. Lorsque nous obtenons le BF de la relation  $A_0(D)$  d'un champ de vue, on peut estimer l'extinction  $A_0$  pour chaque bin de distance. On calcule donc la magnitude apparente d'une étoile du Red Clump à chaque distance (on utilise les magnitudes absolues obtenues par Ruiz-Dern et al. 2018), pour chaque bande photométrique infrarouge. On ne conserve les résultats que pour les distances où la magnitude apparente des étoiles du Red Clump est inférieure à la magnitude *cut* (correspondant à la plus faible luminosité d'une étoile observée et analysée) du champ de vue.

#### 4.3.2 Résultats 2MASS et GDR2

On présente en Figure 4.7 et 4.8 les distributions  $P(A_0, D)$  des champs tests décrits à la section 4.3 et les CMD des champs de vue tests avec la relation BF du champ.

Pour certains champs de vue, on distingue la relation  $A_0(D)$  tandis que pour les autres on ne distingue rien de particulier « à l'œil ». Cependant on voit sur les CMD que le résultat donné par FEDReD traverse toujours la traînée du Red Clump. Ceci indique donc que les différentes améliorations de l'algorithme permettent d'obtenir un résultat cohérent. On a aussi ajouté la relation  $A_0(D)$  obtenue par Lallement et al. (2019) au CMD. On constate que les deux méthodes sont toujours en accord sur la solution, mais notre algorithme a l'avantage de fournir de l'information à de plus longues distances.

On note, sur les distributions de probabilité, un changement de régime aux alentours de 8kpc, cet élargissement en extinction est juste la conséquence du prior  $P(A_0 | D)$  fourni par une première réalisation de la déconvolution. En effet, une première déconvolution avec un prior plat nous permet d'obtenir un intervalle en extinction qui est ensuite utilisé comme prior jusqu'à une distance limite (déterminée grâce au paramètre de complétude adapté à une étoile du Red Clump), et au delà de laquelle un prior plat est utilisé. C'est à ce passage au prior plat que correspond le changement de régime (voir Section 3.2.4).

#### 4.3.3 Résultats UKIDSS et GDR2

On teste aussi notre algorithme sur le catalogue UKIDSS. Les distributions  $P(A_0, D)$  des mêmes champs tests et les CMD sont présentés en Figure 4.9 et Figure 4.10. On constate que les résultats sont erronés pour plusieurs champs tests, l'extinction est souvent surestimée (par exemple pour  $l = 9.6^\circ$  et  $l = 32.4^\circ$ ). Cela est principalement dû au fait que UKIDSS est un relevé profond, par conséquent les étoiles proches sont trop brillantes et n'y apparaissent pas. L'algorithme n'a donc que peu d'information à faible distance, or comme on l'a déjà noté, si les faibles distances sont mal contraintes, la convergence de l'algorithme est erronée.

Pour pallier ce problème, nous avons créé un prior pour  $P(A_0 | D)$  en utilisant les résultats obtenus avec 2MASS. Pour ce faire on utilise les MCS obtenues sur le relevé 2MASS et on impose, pour chaque distance  $D$ , que l'extinction  $A_0$  soit comprise entre les extrema d'extinction atteints par les MCS :

$$P(A_0 | D) = cste \text{ si } \min[mcs(A_0, D)] < A_0 < \max[mcs(A_0, D)]; 0 \text{ sinon.} \quad (4.2)$$

Les résultats de la déconvolution avec ces priors sont présentés en Figure 4.11. On constate que le prior a aidé la convergence à faibles distances, sans perturber outre mesures les résultats

à plus grandes distances, pour la plupart des champs de vue. En revanche l'ajout de ce prior a complètement perturbé les résultats du champ  $l = 32.4^\circ$ . Comme on a pu le constater avec nos tests précédents, un prior fort sur  $P(A_0 | D)$  peut induire en erreur la déconvolution, notre prior est peut-être trop fort. Nous travaillons toujours sur cet aspect de l'algorithme et du traitement des données UKIDSS.

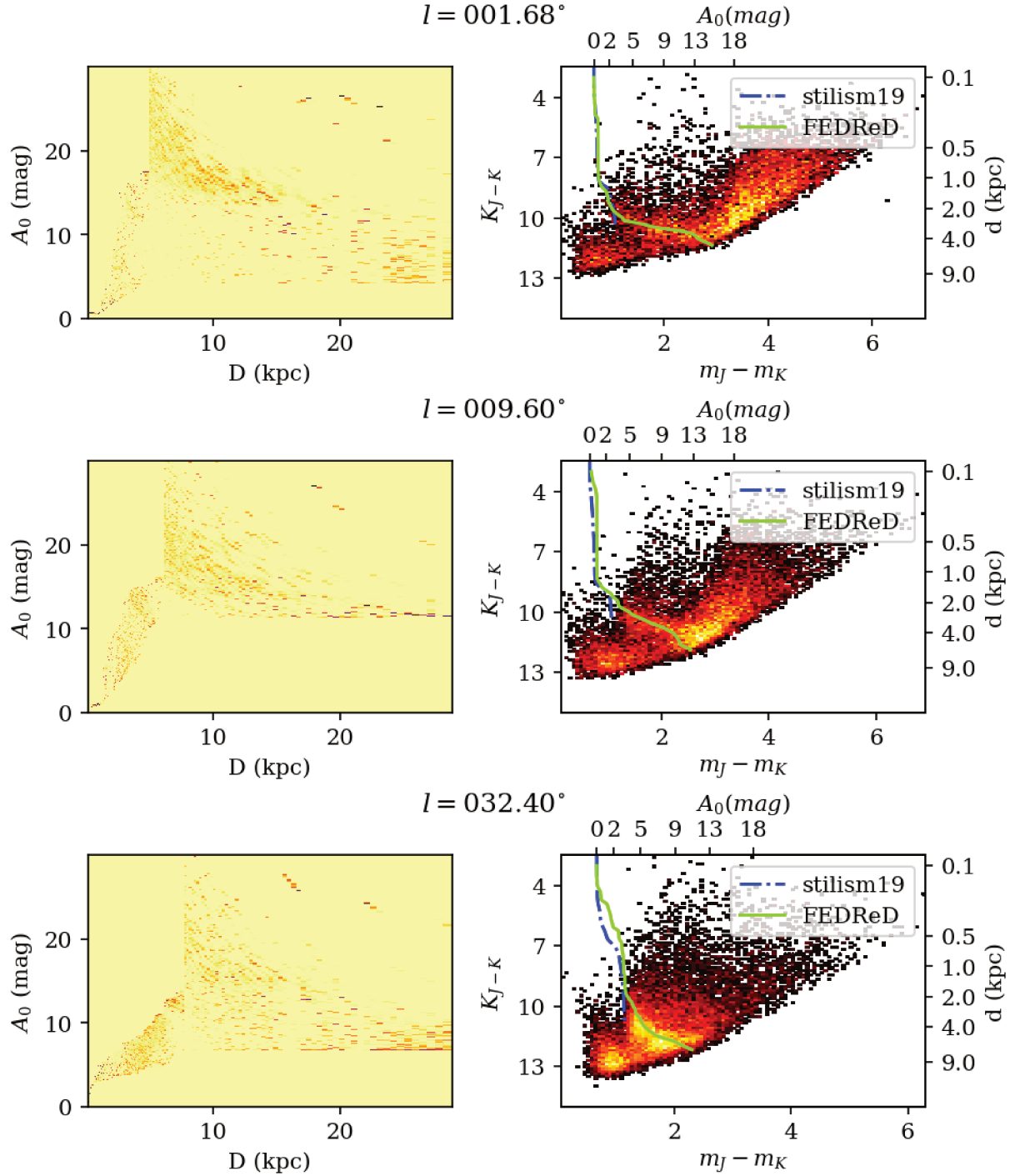


FIGURE 4.7 – *Gauche* : distribution  $P(A_0 | D)$  pour chacun des champs de tests. *Droite* : CMD avec le catalogue 2MASS de chacun des champs de tests (en raison du grand nombre de points on affiche la densité codée en échelle de couleur), on a ajouté la relation BF obtenue par FEDReD ainsi que les résultats (stilism) de Lallement et al. (2019)



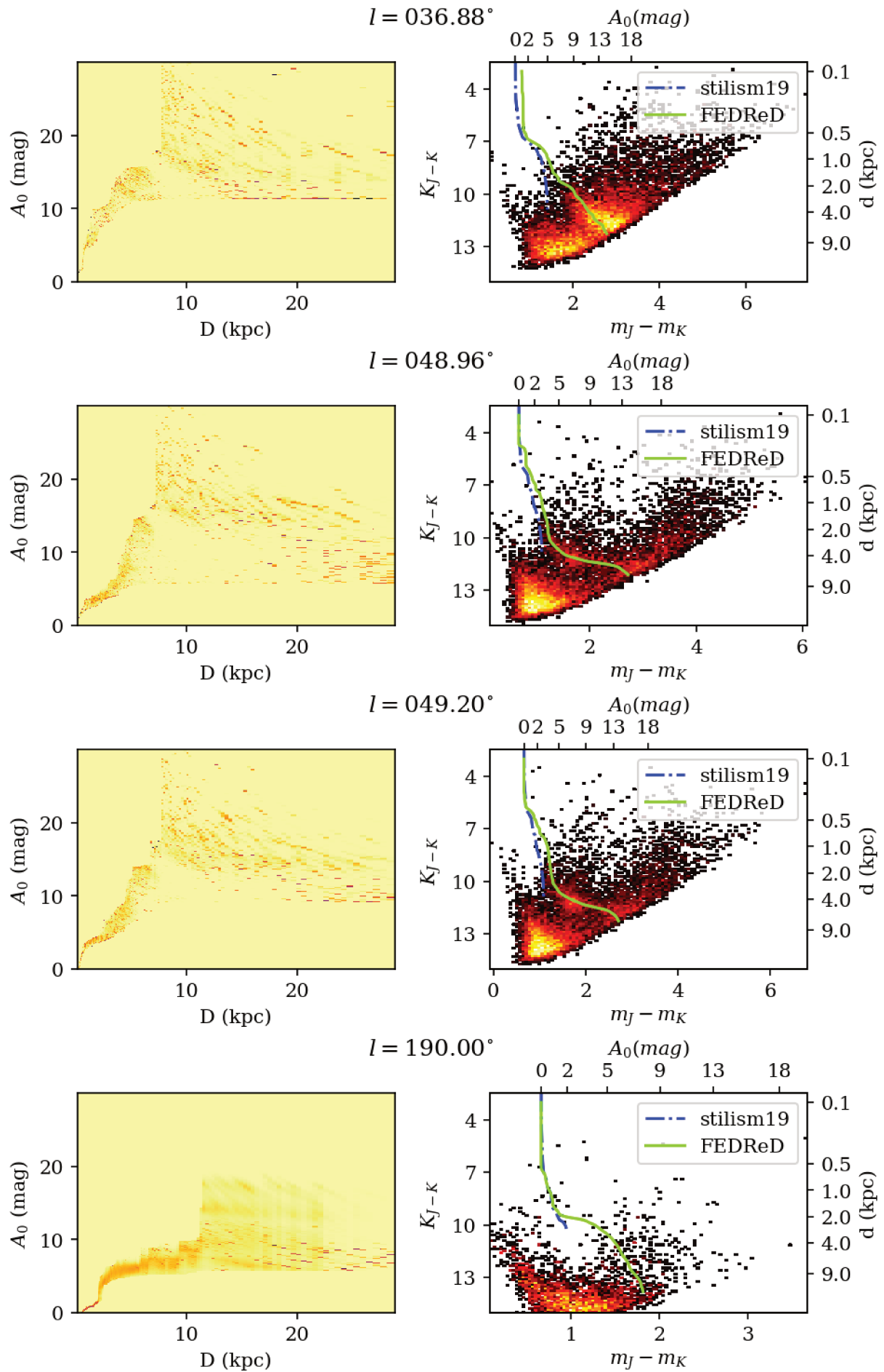


FIGURE 4.8 – Suite de la Fig. 4.7

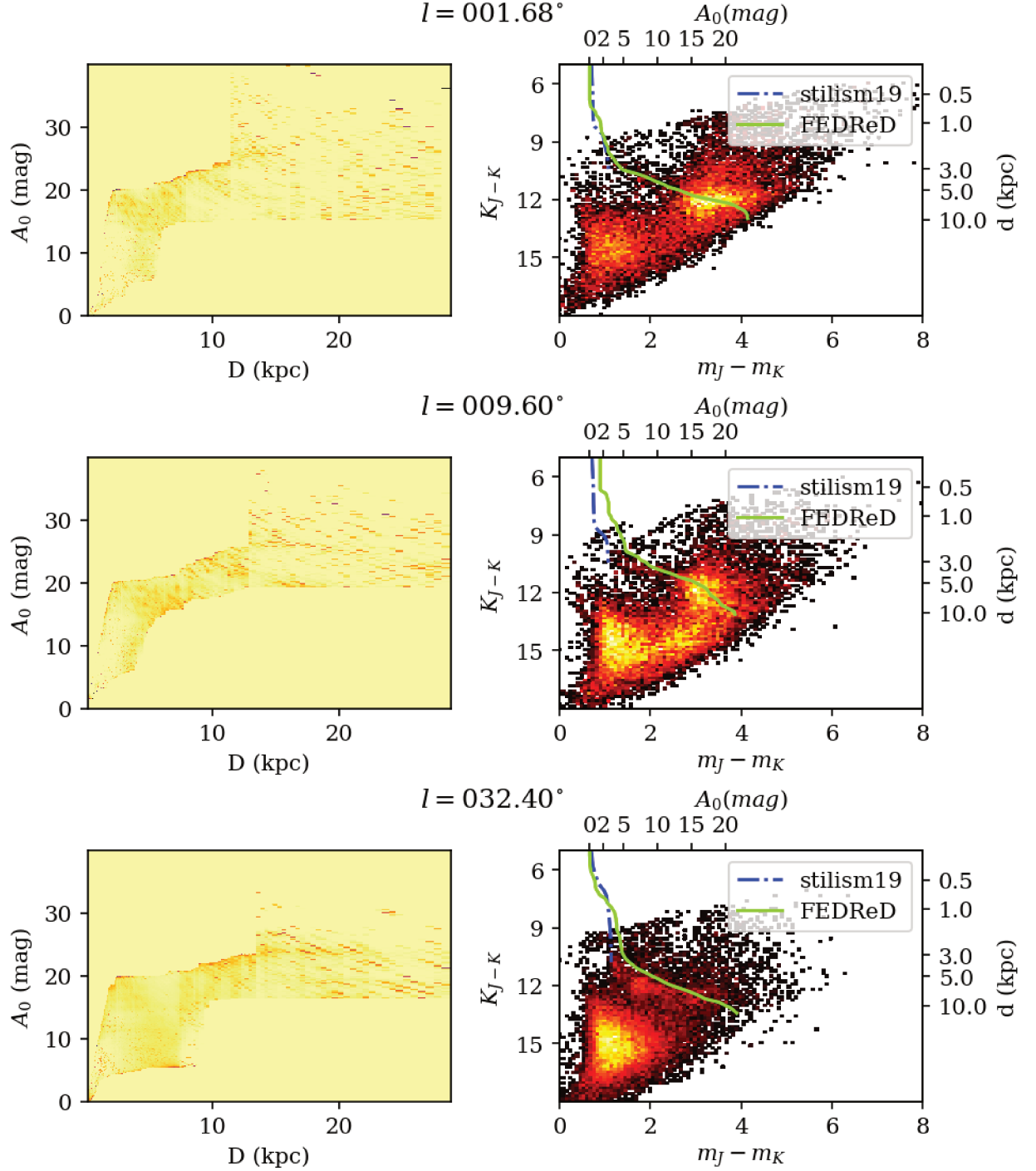


FIGURE 4.9 – Diagnostics équivalents à la Figure 4.7 pour les champs de vue avec UKIDSS

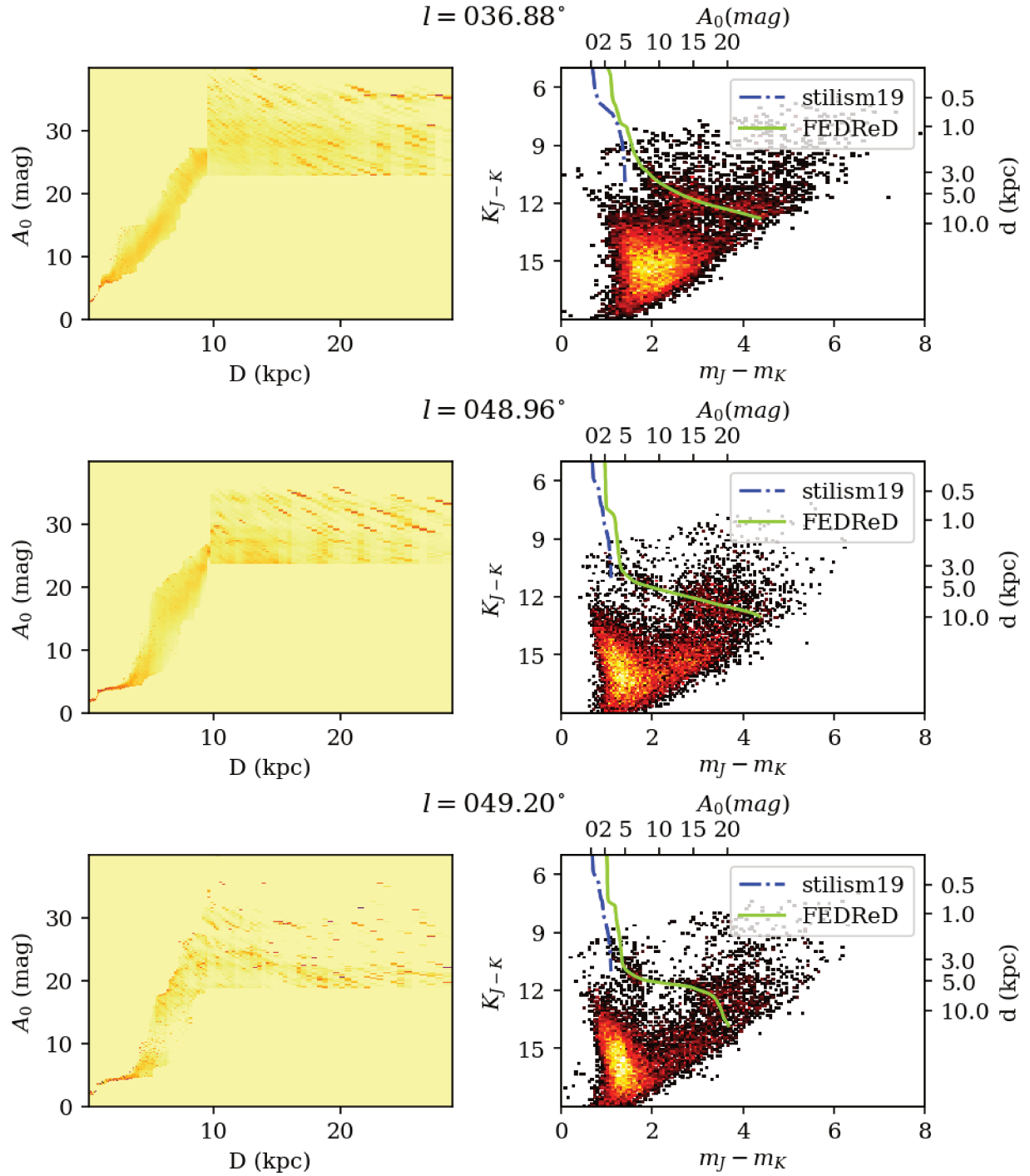


FIGURE 4.10 – Suite de la Fig. 4.9

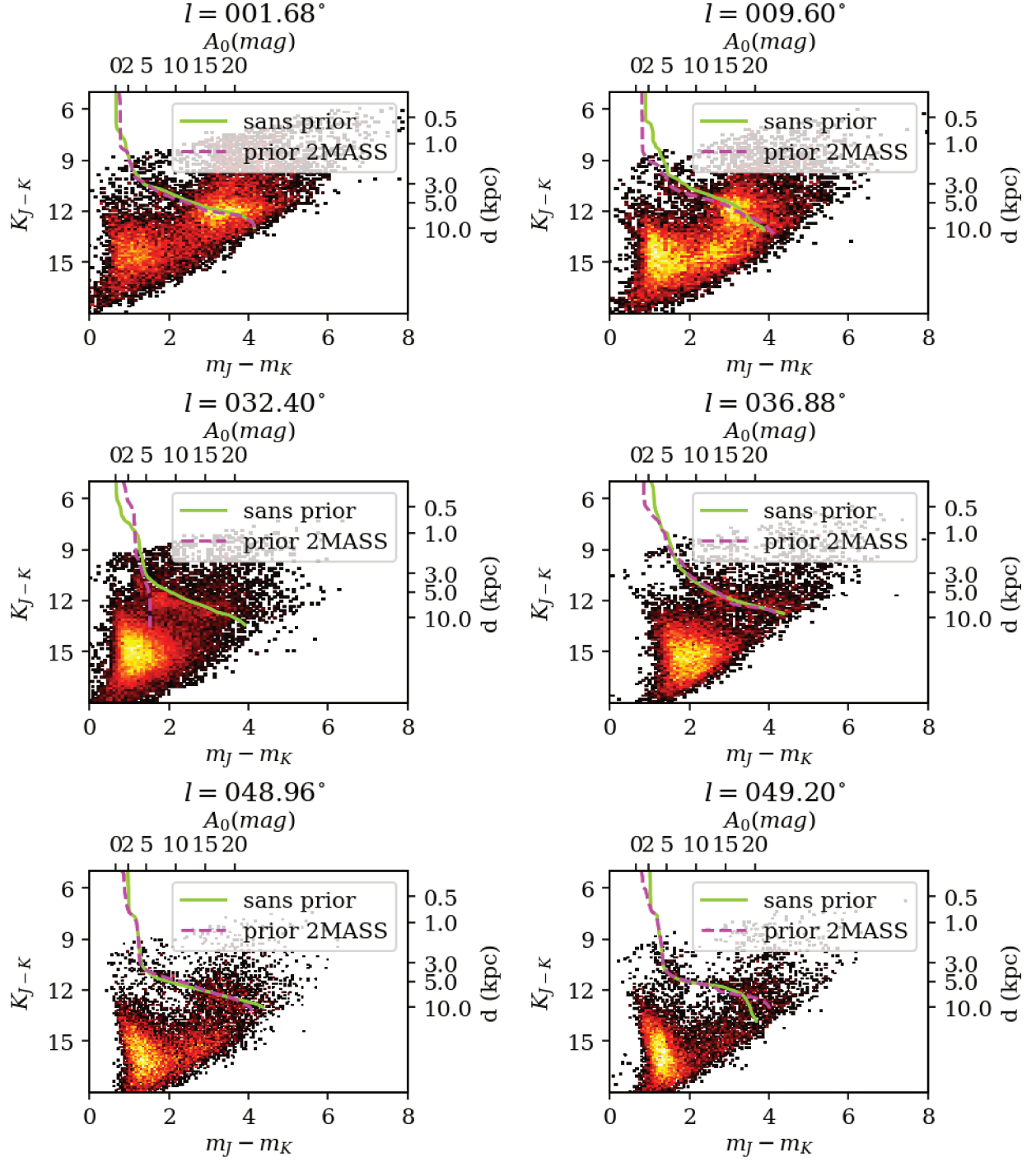


FIGURE 4.11 – CMD des champs de vue UKIDSS, on a ajouté les résultats BF obtenus avec et sans prior 2MASS



# Chapitre 5

## Disque galactique vu par 2MASS et GDR2

**L**A fiabilité de notre algorithme étant vérifiée, pour la combinaison 2MASS et GDR2, nous pouvons maintenant étudier le disque galactique. FEDReD analyse des champs de vue séparés, nous avons donc découpé le disque en champs de vue individuels. Les résultats BF et MCS produits pour chacun des champs de vue sont combinés pour construire une carte de densité d’extinction  $a_0$ . Cette carte d’extinction dessine de nombreuses structures du disque galactique connu, mais révèle aussi leurs formes plus précisément et démontrent leurs elongations.

### 5.1 Caractéristique technique

Durant cette thèse, nous n’avons étudié que la latitude  $b = 0^\circ$ , mais nous avons étudié les  $360^\circ$  de longitude du disque, grâce à 2MASS et GDR2 qui sont des relevés du ciel complet.

Nous avons créé des champs de vue de largeur  $0.48^\circ$  en longitude et latitude. Cette dimension permet de constituer des champs de vue avec au moins quelques milliers d’étoiles permettant ainsi une bonne convergence de la déconvolution. Par ailleurs, nous tirons des champs de vue tous les  $0.24^\circ$  en longitude, ce qui nous assure un recouvrement des champs de vue. Cette taille de pixels a été choisie pour assurer la présence d’au moins quelques milliers d’étoiles par champs de vue, même en direction de l’anticentre galactique.

Cette étude représente donc 1500 champs de vue à analyser, pour un total de 8 681 956 étoiles étudiées, stockées en fichier hdf5 avec dataset compressé (voir section 3.4) pour un total de 2.7 To. La distribution individuelle de chaque étoile est utilisée deux fois (à cause du recouvrement des champs de vue), mais n’est calculée qu’une seule fois. Chaque run de déconvolution effectue une requête pour obtenir les résultats des étoiles appartenant aux champs de vue en cours.

### 5.2 Construction de la carte

Lorsque tous les champs de vue ont été analysés et déconvolués, nous avons à notre disposition le BF de la relation  $A_0(D)$  pour chacun des champs de vue ainsi que les MCS. Ces résultats sont exprimés en magnitude et correspondent à l’extinction cumulée le long de la ligne de visée. Pour en extraire une carte d’extinction, nous avons développé un pipeline d’assemblage.

### 5.2.1 Réduction d'une carte d'extinction

Les résultats de la déconvolution fournissent une relation entre l'extinction et la distance. Les valeurs de  $A_0(D)$  représentent la valeur intégrée de l'extinction à une distance  $D$ . Or pour étudier la morphologie du disque en poussière, on souhaite connaître la densité d'extinction  $a_0(D)$  plutôt que la valeur cumulée, par ailleurs on souhaite aussi utiliser l'information des champs de vue voisins pour augmenter la cohérence angulaire de nos résultats.

#### Lissage des lignes de visée

La relation  $A_0(D)$  obtenue pour chacune des lignes de visée est calculée indépendamment des autres lignes. Or les champs de vue que nous étudions se recouvrent les uns les autres, ils sont donc censés partager de l'information.

Pour prendre en compte l'information venant des champs de vue voisins, nous lissos nos résultats. Pour chaque bin de distance d'une ligne de visée, la valeur finale de l'extinction est une moyenne pondérée telle que :

$$A_0^l(D_i)_{\text{lissée}} = \frac{A_0^{l^-}(D_i)}{4} + \frac{A_0^l(D_i)}{2} + \frac{A_0^{l^+}(D_i)}{4}, \quad (5.1)$$

avec  $l^+$  et  $l^-$  les longitudes voisines de la longitude d'étude. On obtient ainsi des résultats par ligne de visée tenant compte des corrélations angulaires apportées par les lignes de visée voisines. On notera que chacune des relations  $A_0^{l^-}(D)$ ,  $A_0^l(D)$  et  $A_0^{l^+}(D)$  sont des fonctions croissantes de la distance  $D$ , la combinaison linéaire des ces fonctions est donc aussi croissante.

Dans la pratique, ce lissage est réalisé par un produit de convolution discrétisé. Pour chaque bin de distance, on convolue le tableau de résultats avec le tableau  $[1/4, 1/2, 1/4]$ . Cette implémentation nous permet de profiter des optimisations de la librairie *scipy* et d'accélérer la réduction de notre carte.

#### Décumulation

Une fois notre carte lissée, nous avons à notre disposition une carte d'extinction cumulée pour  $0^\circ \leq l \leq 360^\circ$  et  $-0.24^\circ \leq b \leq 0.24^\circ$ . Cette carte peut s'avérer utile pour des observateurs souhaitant étudier un objet du disque galactique et corriger sa magnitude apparente de l'extinction interstellaire. Cependant, nous cherchons à étudier la distribution d'extinction en chaque point du disque, et non la colonne d'extinction.

C'est pourquoi nous devons décumuler la relation  $A_0(D)$ . Enfin, nous devons normaliser la valeur de l'extinction par la largeur en distance du bin pour prendre compte la non-linéarité de la grille en distance et avoir une extinction par kpc.

$$a_0(D_i) = \frac{A_0(D_i) - A_0(D_{i-1})}{\Delta_i}, \quad (5.2)$$

avec  $\Delta_i$  la largeur du bin de distance en kiloparsecs.

### 5.2.2 Carte avec MCS

L'algorithme décrit ci-dessus permet d'obtenir une carte d'extinction à partir des BF des champs de vue. Cependant, on s'est aperçu que pour certains champs de vue, le BF ne retrouve pas la bonne relation entre la distance et l'extinction. On représente en Figure 5.1 le BF sur le CMD du champ de vue  $l = 7.20^\circ$ , ainsi que l'ensemble des MCS du champ de vue avec le BF. On constate qu'il existe une bimodalité dans la distribution des MCS, une partie a tendance à monter

très vite vers les hautes extinctions et dessine le contour autorisé par le prior  $P(A_0 | D)$ . L'autre partie des MCS monte plus progressivement en extinction. Le BF, quant à lui, tend à suivre le groupe des faibles extinctions sur les premiers kiloparsecs puis remonte en extinction, quand la médiane se retrouve sous l'influence des points aberrants trop nombreux. On constate sur le CMD que cette remontée est trop brutale, la solution est au-dessus du red clump, comme lorsque l'on est dominé par la dégénérescence des naines froides sans contrainte de parallaxe GDR2 (décrite section 4.3).

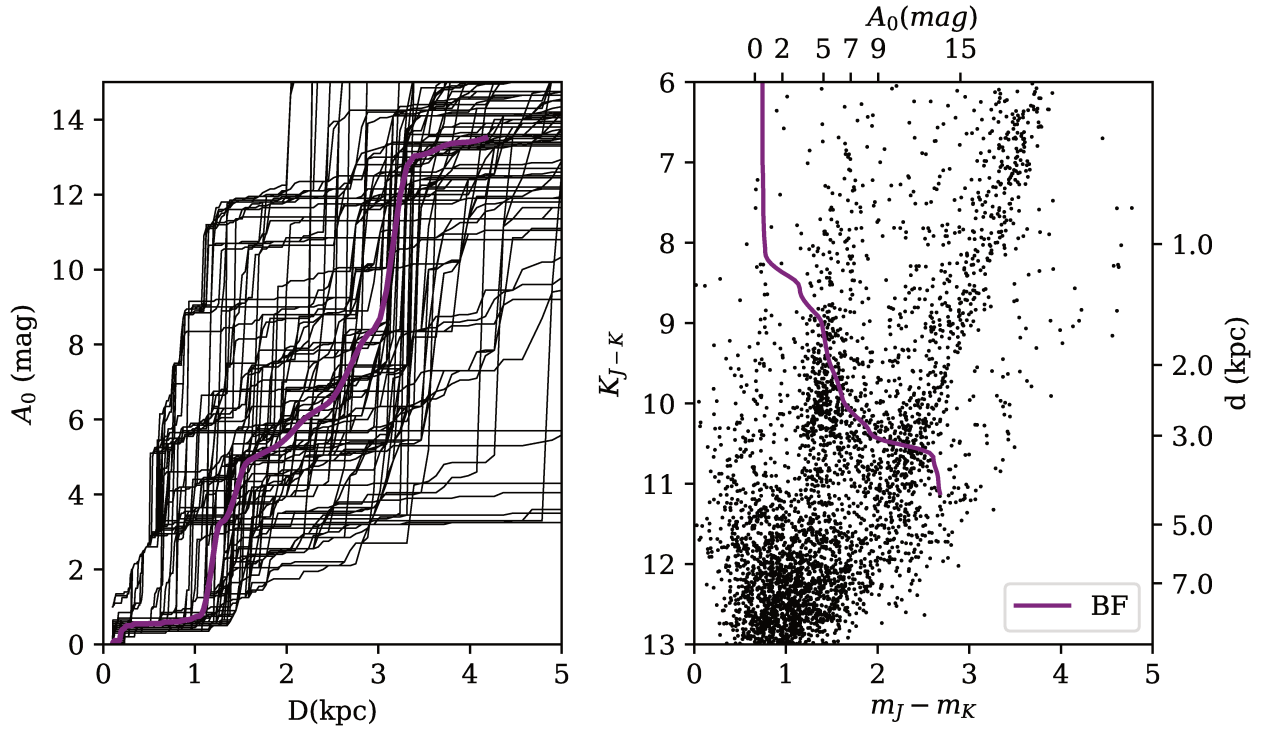


FIGURE 5.1 – *Gauche* : MCS du champ de vue à la longitude  $l = 7.20^\circ$  ; on a représenté qu'un sous-échantillon de 100 solutions pour des raisons de lisibilité. La courbe violette correspond au BF du champ. *Droite* : CMD du champ de vue pour le catalogue 2MASS, la courbe violette correspond au BF du champ.

### Nettoyage

Comme expliqué précédemment, le résultat BF ne suit pas toujours correctement la traînée du Red Clump à cause de la dispersion des MCS. Cela étant, on constate que dans les MCS se trouve de l'information valide. Il est donc possible d'obtenir une relation correcte à partir des MCS si on arrive à enlever les points aberrants.

Pour ce faire, on utilise l'information apportée par les résultats des champs de vue voisins. En effet, comme le montre la Figure 5.2, même si le champ central présente une bimodalité, ses voisins ont eux correctement convergé et peuvent indiquer la bonne tendance.

Pour retirer les MCS frauduleuses, nous définissons des enveloppes de confiance. Toutes les MCS qui s'aventurent en dehors de l'enveloppe sont considérées comme fausses. Nous avons construit plusieurs types d'enveloppe. Nous avons d'abord essayé d'utiliser des médianes et la déviation à ces médianes (MAD) ou des quantiles sur les MCS de la ligne de visée. Ces techniques pouvaient donner de bons résultats sur quelques champs complexes mais demandaient pour cela



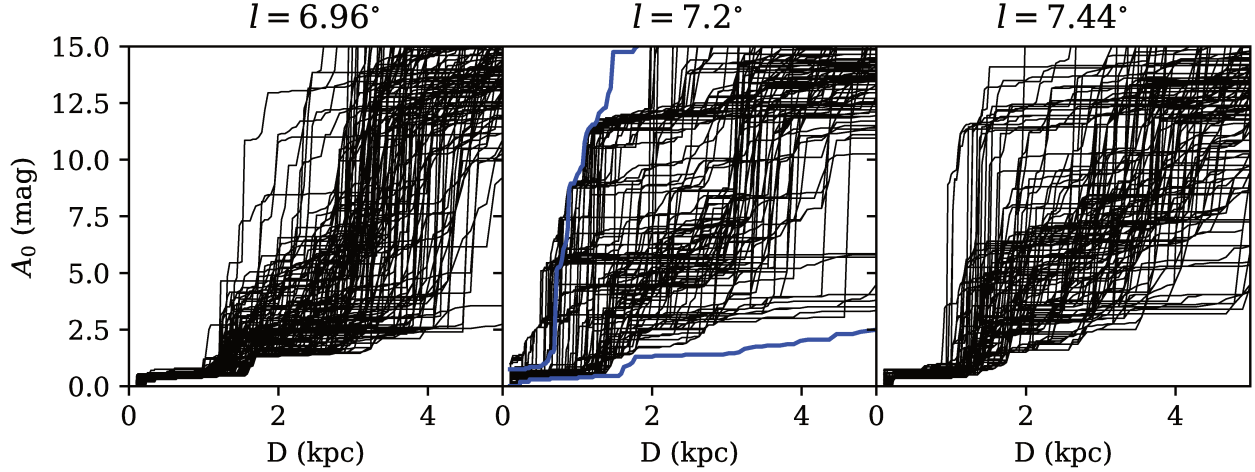


FIGURE 5.2 – MCS de trois champs de vue successifs, pour faciliter la lecture, on n’affiche que 100 solutions par champ de vue.

des réglages très fins et spécifiques à telle ou telle situation. Nous n’avons pas pu dégager de paramétrage satisfaisant pour l’ensemble des champs de vue.

Nous nous sommes donc tournés vers la conception d’une enveloppe de confiance plus simple, construite avec les minima et maxima des champs adjacents. On présente en Figure 5.2 les MCS de trois champs de vue successifs. L’enveloppe de confiance en bleu sur le champ central est construite avec les extrema des extrema *i.e.*  $\max[\max(A_0^{l^+}), \max(A_0^{l^-})]$  et  $\min[\min(A_0^{l^+}), \min(A_0^{l^-})]$  pour chaque distance étudiée. On considère qu’une MCS est un point aberrant si elle est en dehors de cette enveloppe de confiance sur plus de 5 bins de distance. On a instauré cette souplesse de 5 bins de distance pour éviter de « tuer » des solutions flirtant avec l’enveloppe et qui en sortent à quelques rares occasions. Par ailleurs on s’est aperçu que pour que ce nettoyage ne tue pas les « bonnes solutions » par une variation réelle mais brutale de l’extinction, il est nécessaire d’avoir un recouvrement de nos champs de vue.

Pour nettoyer correctement les MCS, nous allons réitérer ce processus. Une première passe de nettoyage ayant déjà été effectuée, les nouvelles enveloppes min/max sont plus contraignantes et permettent d’éliminer de nouveaux points aberrants. On procède ainsi par itérations successives, jusqu’à ce que plus aucune MCS ne soit tuée. On considère alors que notre réserve de MCS est propre.

La Figure 5.3 présente les résultats du nettoyage des MCS pour le champ de vue  $l = 7.20^\circ$  ainsi que le CMD associé. On constate que la « population » de MCS qui montait trop vite a été fortement réduite par le nettoyage. La population de MCS restante semble contenir la bonne information pour le champ de vue.

Le processus de nettoyage demande une trentaine d’itérations pour nettoyer toutes les MCS du disque galactique. On représente en Figure 5.4 un diagramme représentant le nombre de solutions frauduleuses retirées en fonction de la longitude galactique. Comme on peut s’y attendre, c’est pour les lignes de visées dirigées vers le centre galactique que le nombre est le plus élevé. Cela est dû à la mauvaise convergence dans ces champs de vue où l’encombrement est important et où GDR2 apporte peu de contraintes.

## Sampling

Nous avons maintenant un ensemble de solutions nettoyées pour chaque ligne de visée. Avec ces MCS nettoyées pour chaque ligne de visée, on peut construire un très grand nombre de

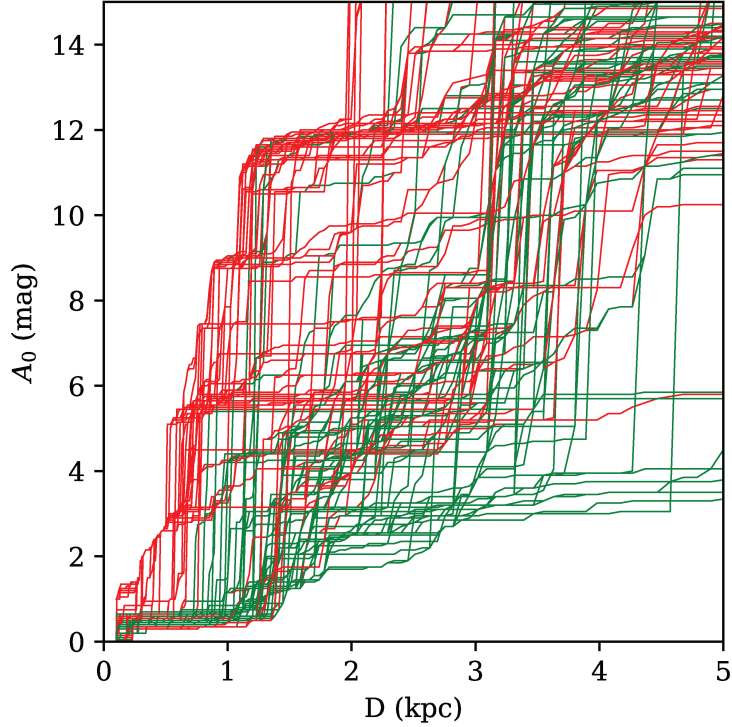


FIGURE 5.3 – MCS du champ de vue à  $l = 7.20^\circ$ . Les solutions tracées en rouge sont celles considérées comme points aberrants, celles tracées en vert sont conservées.

cartes d’extinction. Nous allons utiliser une approche de Monte-Carlo pour obtenir un échantillon statistique à analyser.

Pour tirer aléatoirement une carte d’extinction parmi l’ensemble des solutions possibles, on tire au sort, en suivant une loi uniforme, une seule MCS par ligne de visée, et ce indépendamment pour chacune d’entre elles. Ce processus aléatoire génère une relation  $A_0(D)$  par ligne de visée, en assemblant les relations de l’ensemble des lignes de visées, on obtient une carte d’extinctions cumulées.

Nous lisons cette carte en utilisant le processus décrit précédemment (équation (5.1)). Cela nous donne donc une carte d’extinctions cumulées construite à partir d’un échantillonnage des MCS et dont les résultats pour chaque ligne de visée tiennent compte de l’information provenant des lignes de visée adjacentes.

On va ainsi tirer 1000 cartes d’extinctions à partir des MCS. On a donc 1000 relations  $A_0(D)$  possibles pour chaque ligne de visées. Ces relations prennent en compte l’information des champs de vue voisins par le processus de nettoyage des MCS ainsi que par le lissage. On peut noter que chacune de ces relations est croissante avec la distance, mais la médiane de l’extinction à chaque distance ne l’est pas nécessairement. C’est pourquoi on utilise *cobs* (Ng & Maechler 2007) pour ajuster une spline médiane contrainte croissante à l’intérieur de ces 1000 solutions. Cette spline nous donne la relation  $A_0(D)$  la plus optimale pour chaque champ de vue.

La Figure 5.5 présente le CMD du champ de vue  $l = 7.20^\circ$  et en surimpression la relation  $A_0(D)$  obtenue par nettoyage puis échantillonnage ainsi que le BF obtenu par la déconvolution uniquement (aussi présenté en Figure 5.1). La relation obtenue après nettoyage et échantillonnage est en bon accord avec la traînée du Red Clump et ne surestime plus l’extinction. Cela souligne l’importance de l’utilisation des informations des champs voisins.

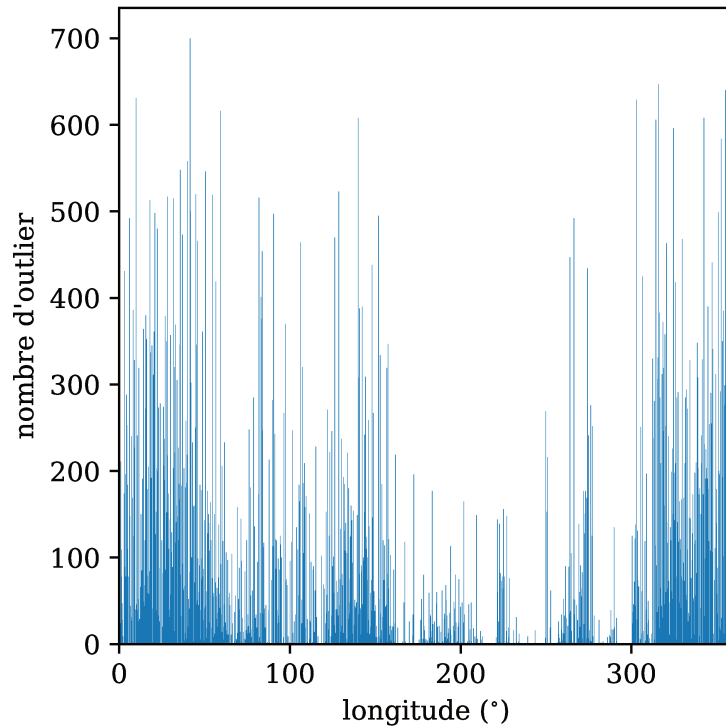


FIGURE 5.4 – Diagramme du nombre de points aberrants en fonction de la longitude.

### Carte d'erreurs

Avec ces 1000 cartes d'extinctions cumulées, on est en mesure de calculer une enveloppe de confiance pour les extinctions cumulées, en recherchant les quantiles d'extinction à chaque bin de distance (en procédant de la même manière qu'à la section 3.3).

Concernant les incertitudes sur la densité d'extinction, on peut les estimer avec un bootstrap utilisant l'échantillon nettoyé de MCS. C'est à dire que, pour chaque ligne de visée, on va tirer aléatoirement avec remise  $n_{clean}$  MCS, avec  $n_{clean}$  le nombre de MCS conservés pour cette ligne de visée. À l'issue de ce tirage on possède donc un nouvel échantillon de MCS, sur lequel on applique notre algorithme de réduction pour obtenir une carte bootstrappée. On réitère 100 fois ce bootstrap, puis on calcule l'écart-type de la valeur de l'extinction à chaque localisation spatiale de la carte pour obtenir une carte des incertitudes.

Le fait d'utiliser la méthode du bootstrap pour évaluer les incertitudes transcrit principalement les incertitude dues à l'échantillonnage, c'est donc une barre d'erreur inférieure concernant les incertitudes, mais qui permet néanmoins de localiser les doigts de dieu.

### Extension en latitude $b$

Par la suite, nous allons étendre notre étude à des latitudes galactiques plus élevées. Cela nous donnera accès à de nouvelles structures du disque, mais fournira aussi de « nouveaux » champs de vue voisins. On pourra ainsi appliquer le nettoyage, l'échantillonnage et le lissage avec les champs de vue se trouvant aux latitudes directement supérieures et inférieures, en plus des longitudes. On espère ainsi complètement nettoyer les MCS des points aberrants.

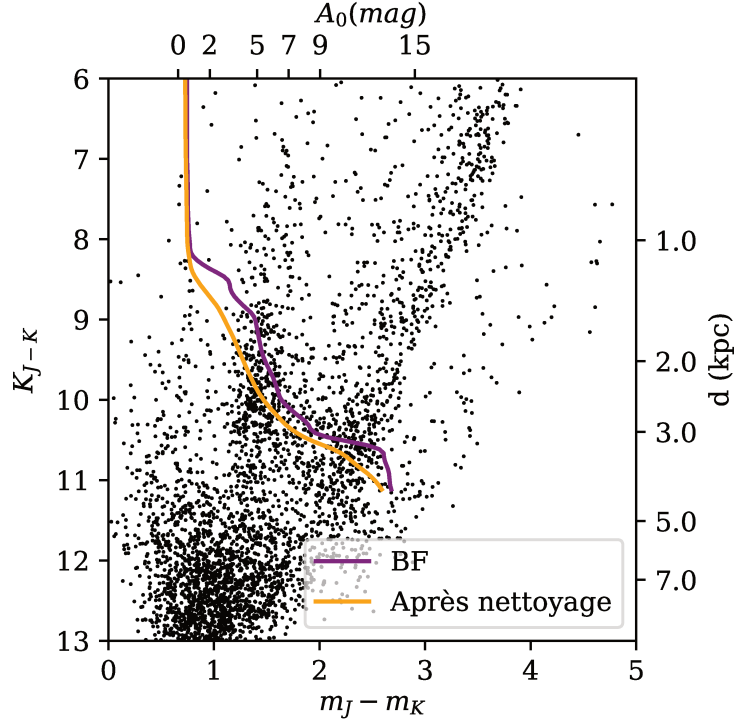


FIGURE 5.5 – CMD du champ de vue  $l = 7.2^\circ$  pour le catalogue 2MASS la courbe violette correspond au BF du champ de vue et la courbe orange correspond à la solution obtenue après nettoyage.

### 5.3 Carte de densité d'extinction $a_0$

On présente en Figure 5.6 la carte de densité d'extinction obtenue par analyse des catalogues 2MASS et UKIDSS par FEDReD puis nettoyage et assemblage des MCS. Cette carte nous montre plusieurs structures d'extinction ainsi que des zones « propres », très transparentes.

Le petit disque blanc centré sur le soleil correspond à la distance en dessous de laquelle nous ne tentons pas d'obtenir de résultats (le voisinage solaire immédiat étant déjà bien connu (Lallement et al. 2019)). C'est le premier bin de notre grille en  $(A_0, D)$ , à  $D = 0.1\text{kpc}$ .

On représente en Figure 5.7 la carte des incertitudes sur la densité d'extinction obtenue par bootstrap de notre échantillon de MCS.

#### 5.3.1 Structures connues et inconnues

On présente en Figure 5.8 notre carte d'extinction avec les structures décrites ci-dessous étiquetées.

La très forte structure d'extinction à cheval entre le premier et le quatrième quadrant est le complexe du Sagittaire-Carène. Il s'étend entre les longitudes  $l = 30^\circ$  et  $l = 300^\circ$ , il est apparenté au bras spiral du Sagittaire. On peut aussi noter à l'intérieur de ce complexe la présence d'une bulle vide de toute extinction et entachée d'un artefact en doigt de Dieu<sup>1</sup> ( $l = 335^\circ$  et  $D = 2\text{kpc}$ ).

Le premier quadrant abrite plusieurs structures de forte extinction apparentées à des bras spiraux (voir section suivante). Le bras spiral local, orienté selon  $l = 80^\circ$  est l'une des structures d'extinction les plus denses que l'on observe. On observe aussi le split du bras local, décrit par

1. artefact de traitement allongé depuis l'extérieur en direction de l'observateur

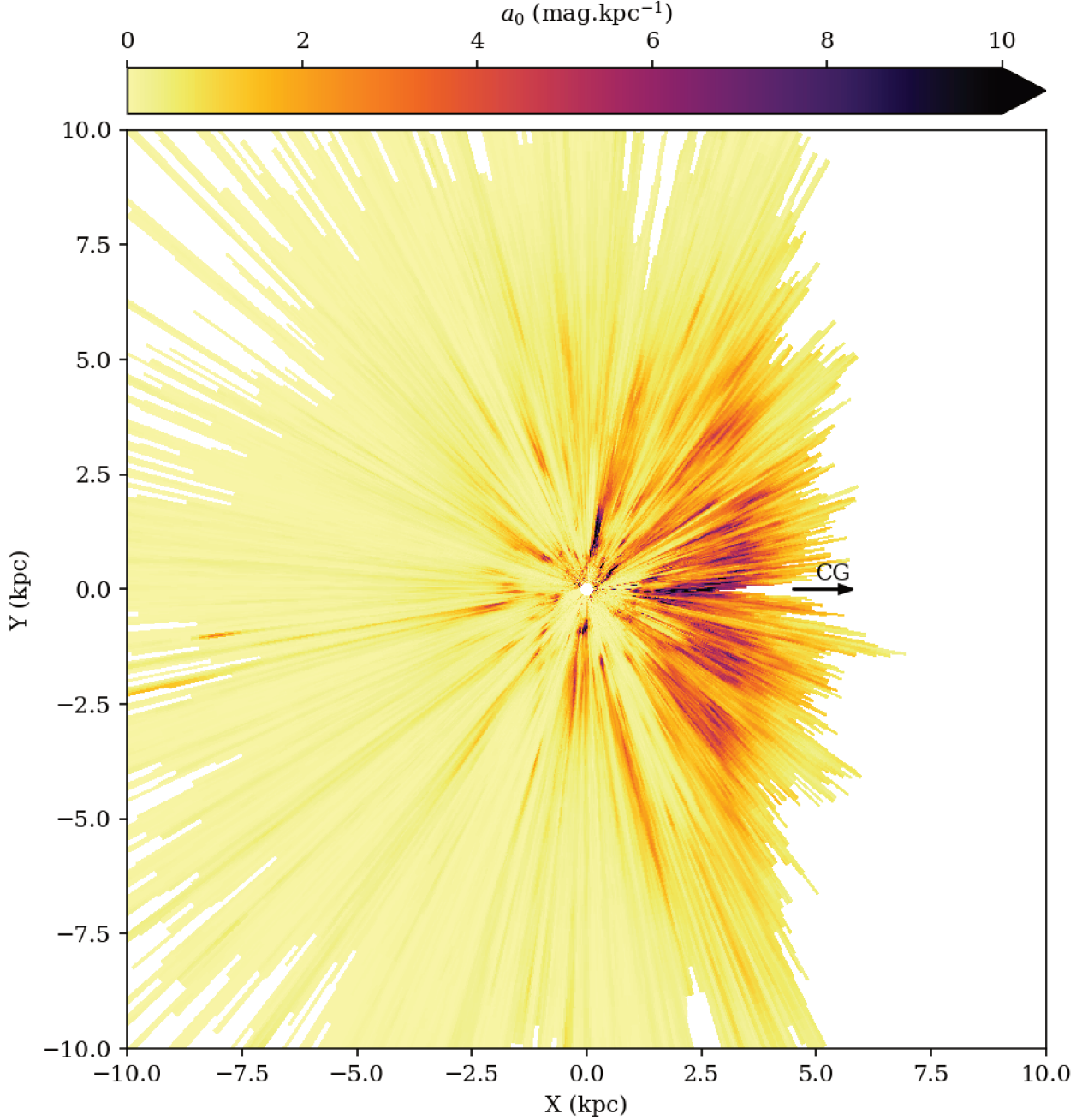


FIGURE 5.6 – Carte de densité d’extinction du plan galactique obtenue par FEDReD sur les catalogues 2MASS appariés à GDR2. Le soleil se trouve à  $(0,0)$ , la direction du centre galactique est indiquée par la flèche noire.

Lallement et al. (2019), compris entre les longitudes  $30^\circ \leq l \leq 40^\circ$ . Ces deux structures sont comme reliées par un pont d’extinction décrit par Xu et al. (2018) comme étant le spur.

Le deuxième quadrant présente deux structures d’extinction notable, la première à  $l = 111^\circ$  et  $D = 3\text{kpc}$ , il s’agit du complexe de Cassiopée-Céphée (Ungerechts et al. 2000) elle correspond aussi à des masers notamment observés par Reid et al. (2014) et associés au bras de Persée. La seconde structure se trouve à  $l = 145^\circ$  et  $1\text{kpc} \leq D \leq 3\text{kpc}$ , c’est le nuage de la Girafe-Cassiopée (cameliopardalis - cassiopea ; Chen et al. (2014)).

Le troisième quadrant présente des petites structures d’extinction, on peut notamment citer le nuage moléculaire Gemini (Carpenter et al. 1995) ( $l = 190^\circ$ ,  $D = 2\text{kpc}$ ) qui est aussi apparenté



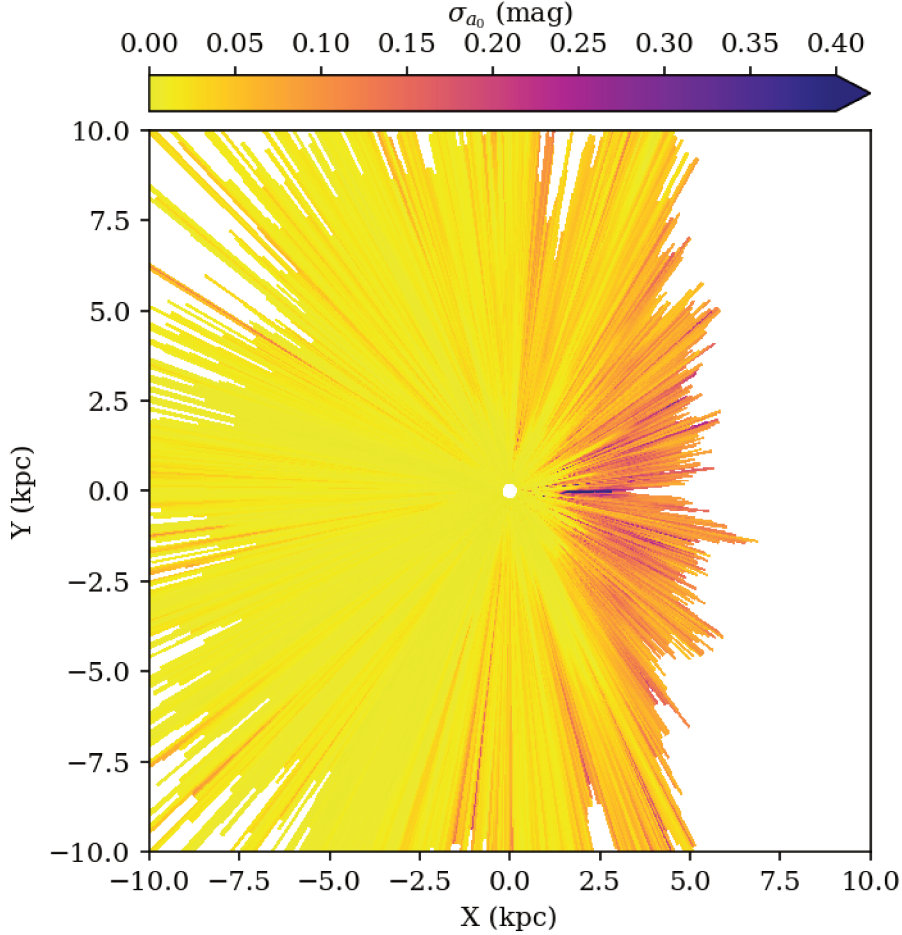


FIGURE 5.7 – Carte d’incertitudes sur la densité d’extinction de la Figure 5.6 obtenue par bootstrap de l’échantillon de MCS nettoyé.

à d’autres masers du bras de Persée. Quatre autres points de fortes densité d’extinctions correspondent aux nuages moléculaires CMA OB1, Maddalena, Mon OB1 et Rosette, dont les distances sont déterminées par Schlafly et al. (2014).

À cheval sur les troisième et quatrième quadrants on observe une grande structure d’extinction. C’est le nuage de Vela. Elle présente une forte extinction en avant plan et se prolonge par une grande bulle vide entourée par un fin bord d’extinction.

Proche du nuage de Vela, on observe une petite zone de très forte extinction prolongée par une traînée d’extinction plus diffuse. C’est le complexe moléculaire de la Carène (Zhang et al. 2001).

En raison de l’absence de résultats à proximité du soleil, on ne voit pas la bulle locale. En revanche on peut distinguer quatre autres bulles propres. La première se trouve entre le complexe du Sagittaire-Carène et le split du bras local. La deuxième est délimitée par le nuage de la Girafe-Cassiopee et par les masers du bras de Persée du troisième quadrant. La troisième bulle se trouve entre ces masers et Vela. Enfin la quatrième est délimitée par Vela et le complexe de la Carène d’un côté et par le complexe du Sagittaire-Carène de l’autre. Ces quatre bulles propres forment ce que nous nommerons un papillon. Les extrémités de la première et quatrième bulle apparaissent clairement à  $D = 2.5\text{kpc}$  et  $D = 6\text{kpc}$  respectivement, tandis que la troisième semble être fermée par quelques points de forte extinction qui d’après le modèle de Reid et al. (2014) sont liés au bras local. En revanche l’extrémité de la bulle deux n’est pas clairement définie.

De façon générale, on observe un grand nombre de bulles vides. La quasi totalité des structures

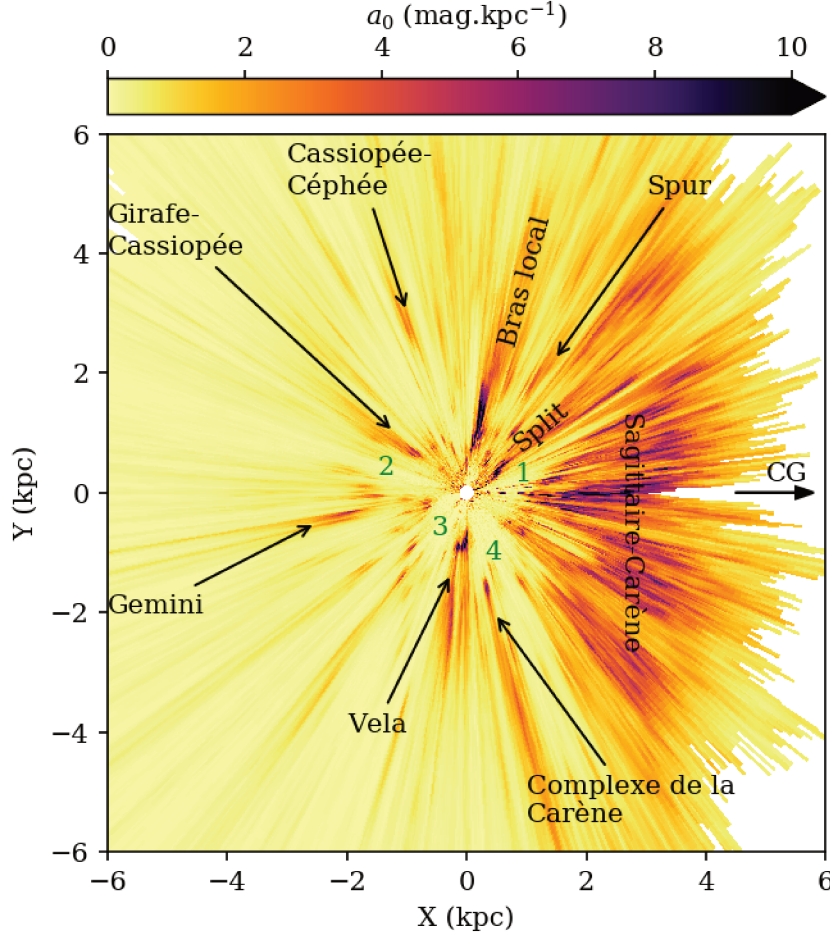


FIGURE 5.8 – Carte d’extinction identique à la Figure 5.6 avec les principales structures étiquetées. Les numéros de 1 à 4 correspondent aux quatre bulles formants les ailes du papillon.

d’extinction que l’on observe ne sont pas continues mais constellées de trous.

### 5.3.2 Comparaison avec Lallement et al. (2019)

La carte obtenue par Lallement et al. (2019) a été réalisée en utilisant les données GDR2 et 2MASS, de plus, les résultats sont présentés dans la même unité que ceux de FEDReD. La différence de résultats ne provient que de la différence des méthodes utilisées. Dans les travaux de Lallement et al. (2019), c’est une méthode d’inversion globale qui est mise en oeuvre, mais la complétude des catalogues n’est pas prise en compte. Par ailleurs l’échantillon d’étoiles est coupé en ne conservant que les étoiles dont l’incertitude relative sur la parallaxe est meilleure que 20%.

On présente en Figure 5.9 la carte d’extinction produite par FEDReD tronquée aux limites des résultats de Lallement et al. (2019). De prime abord on constate une très bonne concordance entre ces deux résultats. On retrouve ainsi sur les résultats de Lallement et al. (2019) toutes les structures d’extinction que l’on a précédemment décrites, ainsi que le papillon vide.

En revanche on peut noter que le complexe du Sagittaire-Carène n’apparaît pas aussi étendu sur les résultats de Lallement et al. (2019) que sur ceux de FEDReD et pour cause, ce complexe se prolonge au-delà de la limite de confiance de Lallement et al. (2019). Concernant l’élongation du complexe de la Carène, elle devrait apparaître sur la carte de Lallement et al. (2019) car elle est toujours dans la limite de confiance. Afin de vérifier si les résultats de FEDReD sont corrects, on présente en Figure 5.10 un agrandissement de ce nuage et les résultats pour la relation  $A_0(D)$

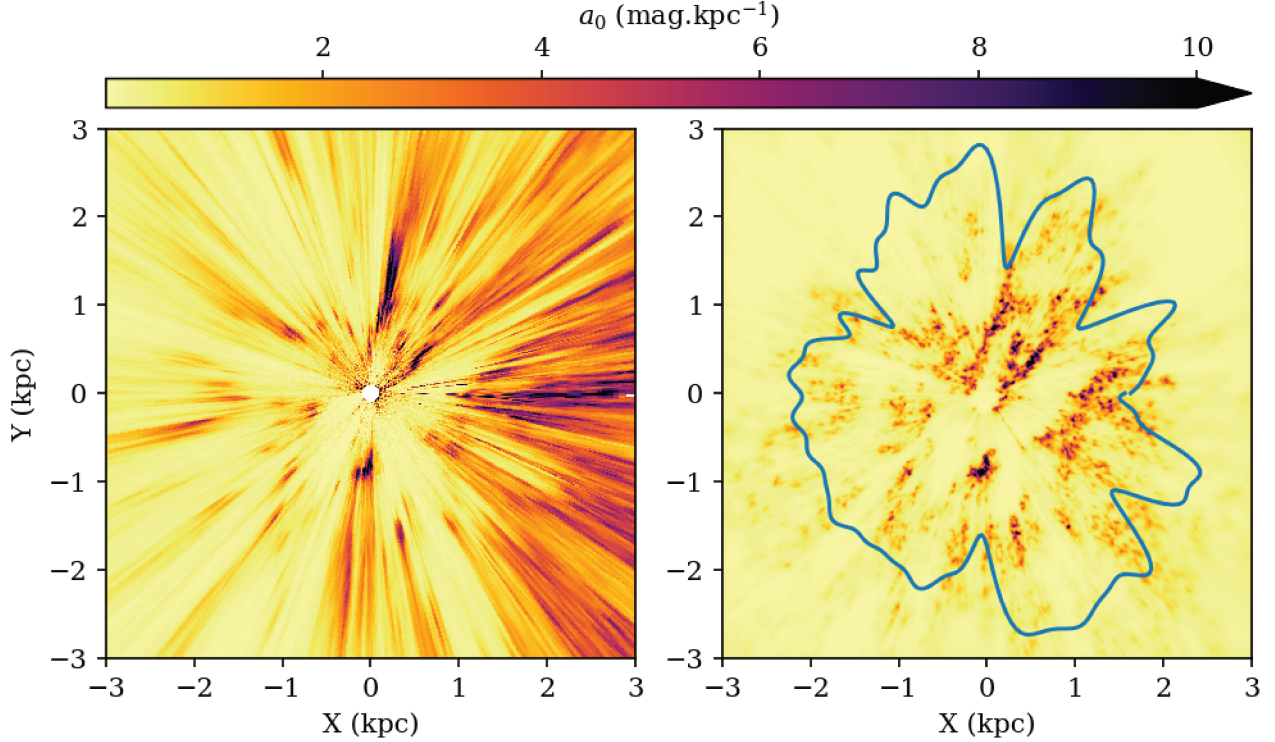


FIGURE 5.9 – Cartes de densité d’extinction obtenues par FEDReD (*gauche*) et Lallement et al. (2019) (*droite*). Les deux cartes sont orientées de la même manière, le soleil est à la position (0,0) le centre galactique est à droite. On a tronqué la carte de FEDReD aux mêmes dimensions que celles de Lallement et al. (2019). La ligne continue sur la carte de Lallement et al. (2019) représente la limite de confiance des résultats.

pour les champs de vue traversant ce nuage en surimpression des CMD. On constate que les relations  $A_0(D)$  sont en parfait accord avec les traînées du Red Clump pour les deux champs de vue choisis. On peut donc affirmer qu’il existe bien des structures en arrière plan du complexe de la Carène. Il est possible qu’il s’agisse de plusieurs petites structures qui sont lissées par notre algorithme.

Enfin on peut noter la présence de quelques doigts de Dieu sur chacune des deux cartes. La présence de ces artefacts sur uniquement l’une des cartes nous permet ainsi de valider leurs caractères fallacieux. On peut ainsi noter les doigts de Dieu orientés vers le centre galactique pour FEDReD ou celui à  $l = 300^\circ$  sur la carte de Lallement et al. (2019).

### 5.3.3 Comparaison aux autres cartes d’extinction

Plusieurs cartes d’extinction ont été obtenues ces dernières années, elles sont construites avec des données et des méthodes différentes. Marshall et al. (2006) utilise la photométrie 2MASS qu’il compare au modèle de Besançon, Chen et al. (2019b) analyse les données des catalogues 2MASS et GDR2 avec un algorithme d’apprentissage et Green et al. (2019) analyse la photométrie Pan-STARRS associée aux parallaxes GDR2 à l’aide de chaînes de Markov. Les comparaisons entre ces cartes et nos résultats sont présentées respectivement en Figures 5.11, 5.12 et 5.13.

On notera que l’ensemble des ces cartes d’extinction sont produites par des techniques différentes appliquées à des données différentes. Ceci explique la différence de tailles des cartes ainsi que les zones accessibles. Lallement et al. (2019) utilisent les mêmes données que FEDReD. La différence de taille de carte n’est due qu’à la technique employée. Lallement et al. (2019) obtient



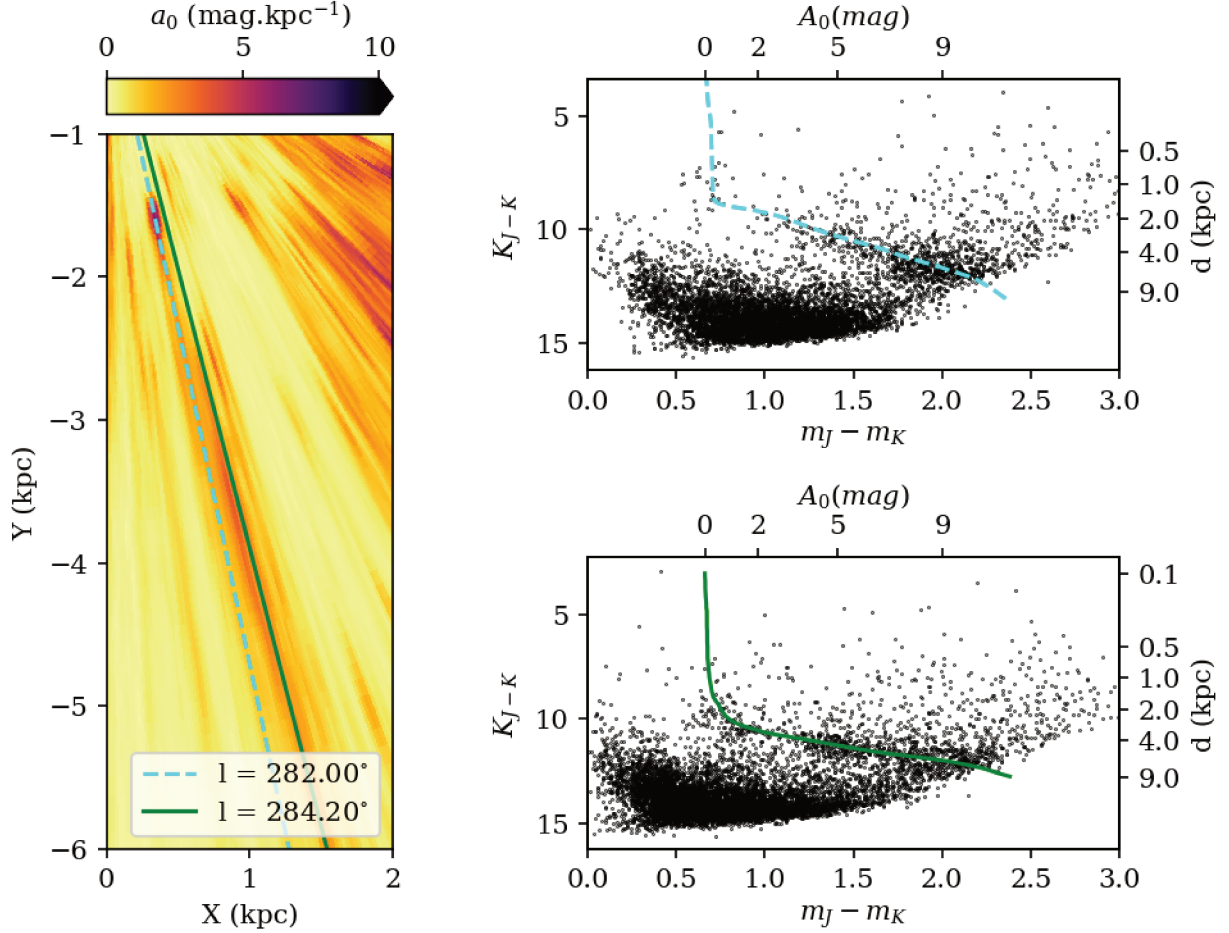


FIGURE 5.10 – *Gauche* : Agrandissement du nuage d’extinction complexe de la Carène et de sa prolongation. *Droite* : CMD des lignes de visée marquées sur l’agrandissement du complexe de la Carène, les courbes en surimpression correspondent aux relations  $A_0(D)$  obtenues par FEDReD.

une carte plus restreinte en raison de la coupure sur les incertitudes relatives des parallaxes. Marshall et al. (2006), quant à eux ils utilisent uniquement les données 2MASS, ils parviennent cependant à obtenir des résultats à grande distance, surtout vers le centre galactique. C’est parce qu’ils basent leurs résultats sur l’observation des étoiles géantes et une définition de la complétude différente de celle de FEDReD, or vers le centre galactique le nombre de géantes augmente. Cela implique une coupure en distance plus éloignée que pour nos résultats qui se basent sur les étoiles du Red Clump (intrinsèquement moins brillantes).

Cela étant, sur l’ensemble de ces résultats, des structures fortes ressortent tandis que d’autres n’ont pas la même taille, ou la même morphologie. Dans le premier quadrant, la forte extinction du bras local est visible dans l’ensemble des travaux, ainsi que les deux premiers kiloparsecs du split du bras local. En revanche, seule notre carte et celle de Marshall et al. (2006) le voient se prolonger sur un peu plus de 6 kpc, même si Green et al. (2019) indique également cette longue prolongation jusqu’à leur coupure à 4 kpc.

Le deuxième quadrant n’est exploré que par Chen et al. (2019b) et Green et al. (2019). On peut voir le nuage de la Girafe-Cassiopee sur ces cartes au même emplacement que sur notre carte. Le complexe de Cassiopee-Céphée en revanche est beaucoup plus fort sur notre carte et celle de Green et al. (2019) que sur celle de Chen et al. (2019b). Ce complexe est en effet en bordure de la carte de Chen et al. (2019b).

Le troisième quadrant est incomplet sur la carte de Green et al. (2019) mais cette dernière

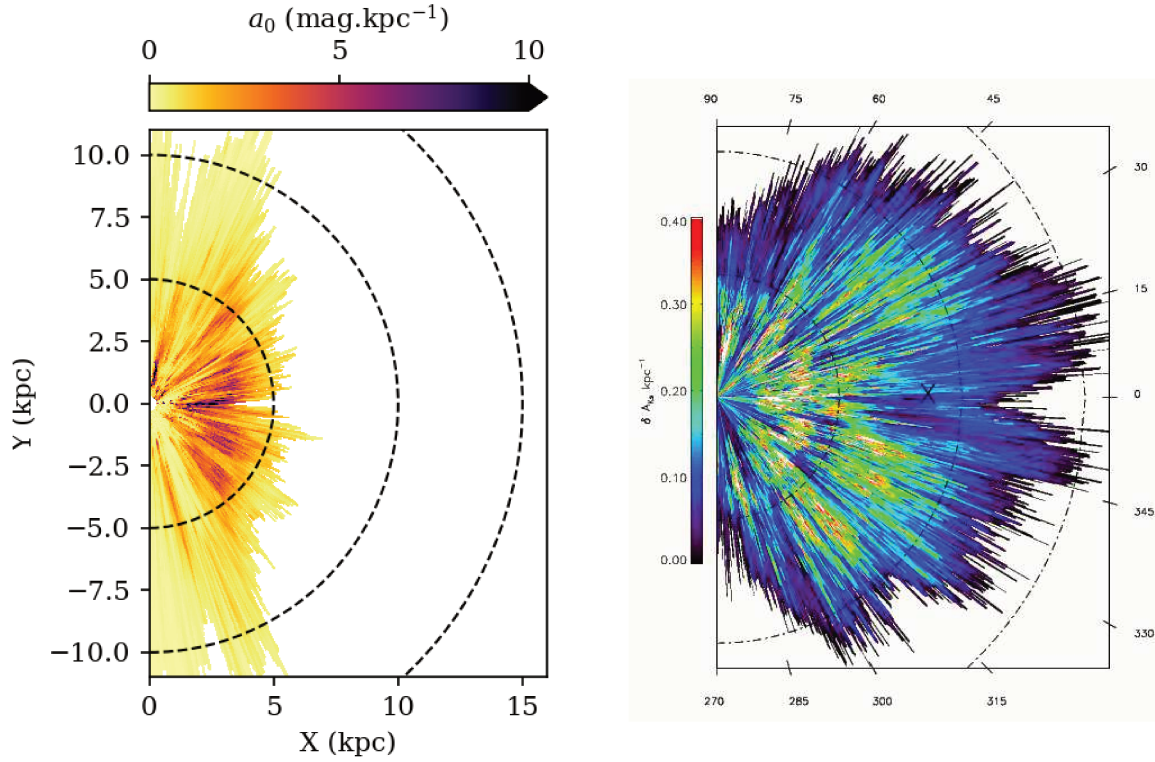


FIGURE 5.11 – *Gauche* : Carte extinction de FEDReD ; *Droite* : cartes d'extinction de Marshall et al. (2006). Les deux cartes d'extinction sont coupées aux mêmes dimensions

s'étend suffisamment pour que l'on y voit le complexe de Gemini. Il est aussi visible sur la carte de Chen et al. (2019b) et sur la nôtre. En revanche la carte de Green et al. (2019) s'arrête avant d'atteindre le nuage de Vela, qui est par contre bien visible sur les résultats de Chen et al. (2019b). Les bords de la partie la plus proche du soleil y sont moins nets que sur notre carte ou celle de Lallement et al. (2019), mais ils confirment en revanche la bulle en arrière plan.

La même élongation du complexe de la Carène que l'on observe est apparente sur les résultats de Marshall et al. (2006), mais Chen et al. (2019b) n'arrivent pas à la reconstruire et s'arrêtent à la première surdensité.

Le quatrième quadrant n'est visible que dans les résultats de Marshall et al. (2006) et de Chen et al. (2019b). Ils indiquent le commencement du complexe du Sagittaire-Carène sensiblement au même endroit que nos résultats. En revanche la bulle à l'intérieur de ce complexe ne fait que se deviner sur ces deux résultats tandis qu'elle est très nette sur les résultat de FEDReD.

Concernant le papillon, il est apparent sur toutes les cartes (au moins en partie selon les limites géographiques de chaque méthode), sauf sur la carte de Marshall et al. (2006).

#### 5.3.4 Comparaison à d'autres traceurs du milieu interstellaire (MIS)

On présente en Figure 5.14 notre carte de densité d'extinction, sur laquelle on a tracé en surimpression la position des nuages moléculaires déduite par Miville-Deschênes et al. (2016) à partir de l'observation radio du  $^{12}\text{CO}$ . Les nuages moléculaires sont généralement présents au même endroit que la poussière interstellaire, c'est pourquoi la localisation des nuages moléculaires doit correspondre aux zones de fortes extinctions. On constate ainsi que les zones présentant une forte densité de nuages sont confondues avec les fortes densités d'extinction. On note par exemple que la traînée de complexe de la Carène est aussi présente dans les nuages moléculaires. On remarque aussi que le papillon est totalement dépourvu de nuage moléculaire, tout comme la

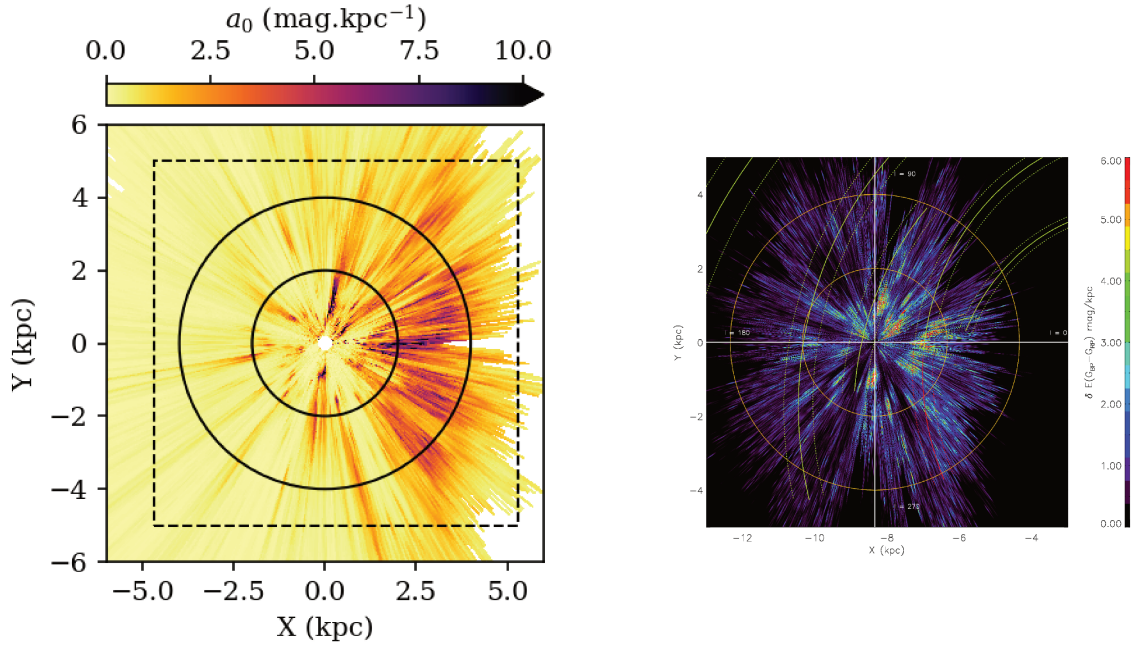


FIGURE 5.12 – *Gauche* : Carte extinction de FEDReD ; *Droite* : cartes d’extinction de Chen et al. (2019b). Les deux cartes d’extinction sont coupées aux mêmes dimensions, les cercles correspondent et le cadre noir sur la carte de FEDReD correspond aux limites de la carte de Chen et al. (2019b)

bulle du complexe du Sagittaire-Carène dont les bords sont parfaitement délimités tant par la densité  $a_0$  que par les nuages moléculaires.

Les amas ouverts sont des regroupements d’étoiles nés en même temps au même endroit. On représente sur notre carte des amas ouverts jeunes dont la distance a été calculée par Cantat-Gaudin et al. (2018) à partir des données de GDR2. Sur la carte de densité d’extinction, on ne constate pas de correspondance très marquée entre les amas et l’extinction. Par ailleurs, Cantat-Gaudin et al. (2018) s’interroge sur l’absence d’amas ouvert dans le cône  $140^\circ \leq l \leq 160^\circ$ , à l’emplacement du bras de Persée, on constate sur notre carte d’extinction intégrée que cette absence est due à une forte colonne d’extinction.

Les régions HII sont le siège d’une forte formation d’étoiles et sont généralement associées au bras spiraux. On présente en Figure 5.16 notre carte d’extinction avec les régions HII compilées par Hou & Han (2014). On constate que ces régions HII sont en très bon accord avec les zones de forte densité d’extinction. On notera la morphologie du complexe du Sagittaire-Carène très bien dessinée par le HII et la forte concentration de ces régions au niveau du spur précisant ainsi la relation entre ce pont de matière et les bras spiraux.

Chen et al. (2019a) a utilisé les données des catalogues VPHAS et GDR2 pour identifier et localiser les étoiles jeunes et les utiliser comme traceurs de bras spiraux. Il parvient à ajuster des bras spiraux dans la distribution qu’il obtient, mais il note la présence de trous qui ne sont pas dus à la complétude du catalogue. On a tracé l’emplacement de ces trous sur notre carte d’extinction (Figure 5.16). On constate que ces trous correspondent à des zones vides d’extinction mais aussi de régions HII, il ne peut donc y avoir de formation d’étoiles

Les masers sont un autre traceur du MIS. Ils sont le siège d’un fort taux de formation d’étoiles, ils se trouvent donc dans des zones riches en poussière et donc à forte extinction. Par ailleurs ils sont souvent apparentés à des bras spiraux. On représente en Figure 5.17 plusieurs masers tous



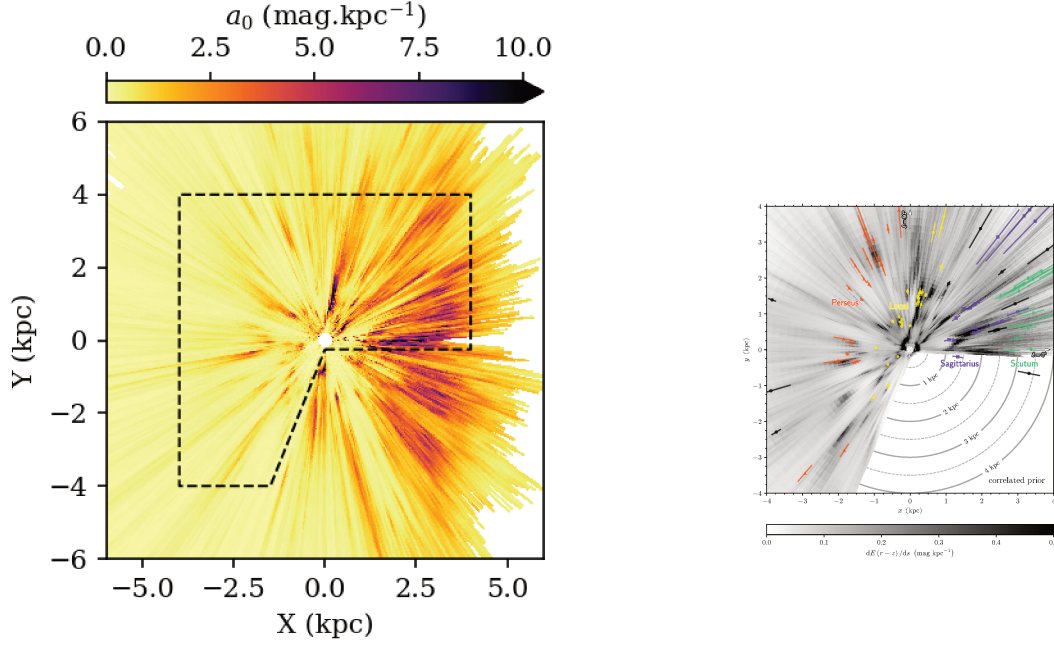


FIGURE 5.13 – *Gauche* Carte extinction de FEDReD; *Droite* cartes d’extinction de Green et al. (2019). Les limites de la carte de Green et al. (2019) sont indiquées en pointillés sur la carte de FEDReD

associés à des bras spiraux. Leurs localisations concordent toujours avec des zones de haute densité d’extinction, à l’exception de deux masers du bras de Persée. On note aussi que la position de ces masers concordent avec les régions HII présentées à la Figure 5.16. Pour comprendre pourquoi deux masers du bras de Persée n’apparaissent pas dans notre carte, nous avons vérifié les MCS de ces champs de vue (Figure 5.18). On constate que ces deux masers n’ont pas de signatures dans l’extinction comme en témoignent les traînées du Red Clump ; en effet, même si la convergence de l’algorithme à grande distance n’est pas évidente pour le champs  $l = 240.24^\circ$ , FEDReD obtient la bonne information dans l’intervalle de distance du maser. Nous avons vérifié dans la carte d’extinction de Schlegel et al. (1998) ces deux champs de vue. Il apparait que l’extinction associée à ces masers est extrêmement localisée, trop pour influencer l’évolution médiane de l’extinction.

### 5.3.5 Bras spiraux

Les bras spiraux sont le siège de fortes formations d’étoiles, ils ont donc une composante de poussière qui doit se traduire par une forte densité d’extinction. On ajoute sur notre carte d’extinction le modèle de bras spiral ajusté sur les masers précédents par Reid et al. (2014) et Xu et al. (2016, 2018) et les bras spiraux de Hou & Han (2014) ajustés sur les régions HII sur la Figure 5.19.

Le bras local est bien visible dans l’extinction du premier quadrant, il passe ensuite par quelques points de fortes extinctions du troisième quadrant et traverse Vela.

Le bras du Sagittaire présente une forte extinction, mais est scindé en deux par la première bulle vide. Par ailleurs, on constate que le split du bras local qui est proche du bras local à faible distance semble plutôt apparenté au bras du Sagittaire à grandes distances. Le spur quant à lui fait une jonction entre les bras local et du Sagittaire.

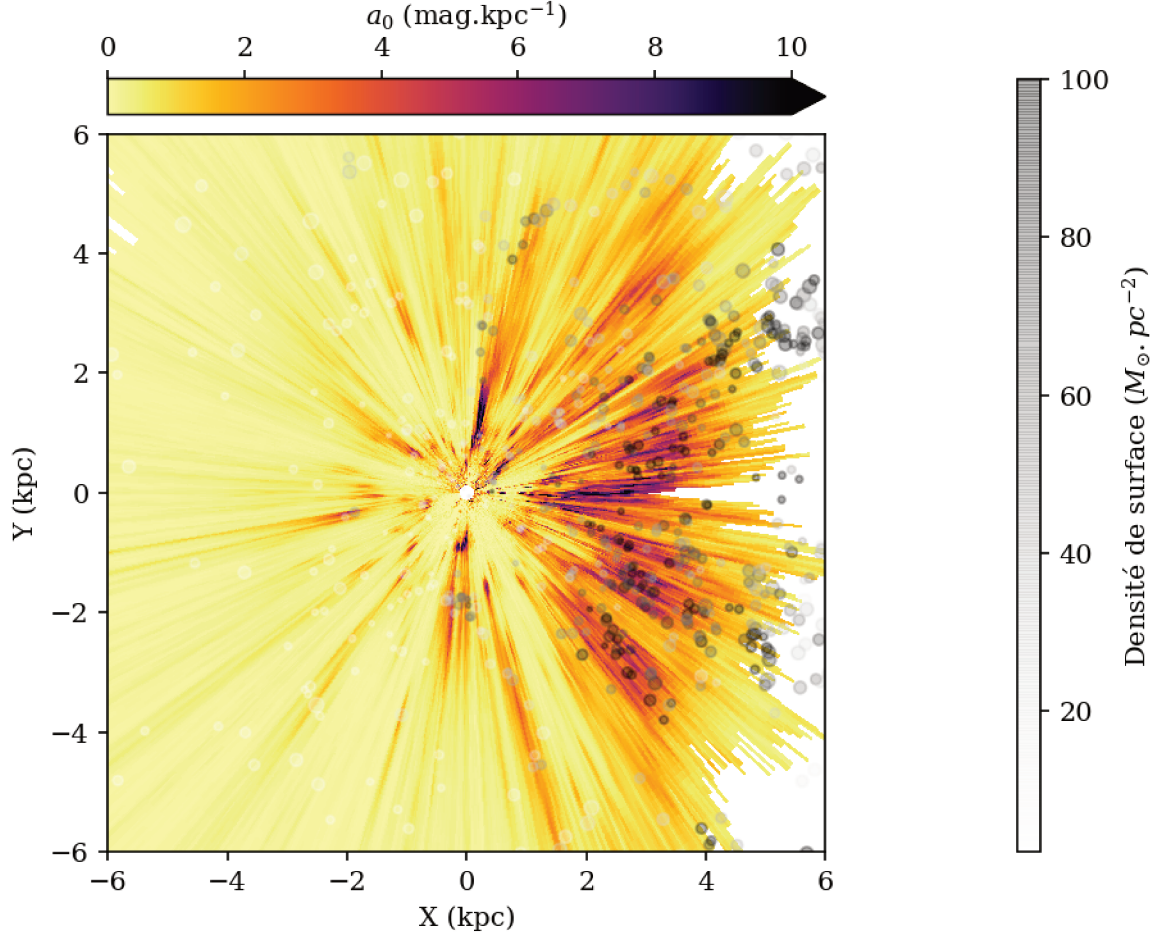


FIGURE 5.14 – Carte de densité d’extinction. La position des nuages moléculaires déduite par Miville-Deschênes et al. (2016) est indiquée par les points ; leurs tailles sont proportionnelles à la taille des nuages et la couleur du point est fonction de la densité moléculaire du nuage. On a tracé uniquement les nuages avec  $|b| < 0.24^\circ$ .

Le bras de l’Écu traverse aussi une zone de forte extinction. En revanche on ne voit pas de séparation notable entre le bras du Sagittaire et le bras de l’Écu. Ils traversent tous les deux le complexe du Sagittaire-Carène.

Le bras de Persée est uniquement visible en extinction aux endroits où se trouvent les masers. Il est complètement morcelé et ne se distingue que grâce au guide visuel du modèle de Reid et al. (2014). Par ailleurs ce bras spiral semble être extrêmement morcelé, il est même décrit comme ayant une structure disruptive par Baba et al. (2018).

Quant au bras extérieur, on ne voit aucune structure dans les cartes d’extinction. En effet, à cause du warp du disque galactique, le bras ne se trouve pas à la latitude  $b = 0^\circ$ , comme en témoigne l’absence de maser associé au bras extérieur sur la Figure 5.17, ou de régions HII (Figure 5.16).

L’emplacement exact des bras spiraux n’est pas encore bien établie, on le constate notamment sur la Figure 5.19, où la trace des bras de Reid et al. (2014) et de Hou & Han (2014) n’empreinte pas le même chemin. Grâce à des cartes du milieu interstellaire, comme la notre, on pourra peut être amener de nouvelles contraintes. Par exemple, on pourrait tenter de déterminer si le bras local passe, ou non, par le nuage de Vela.

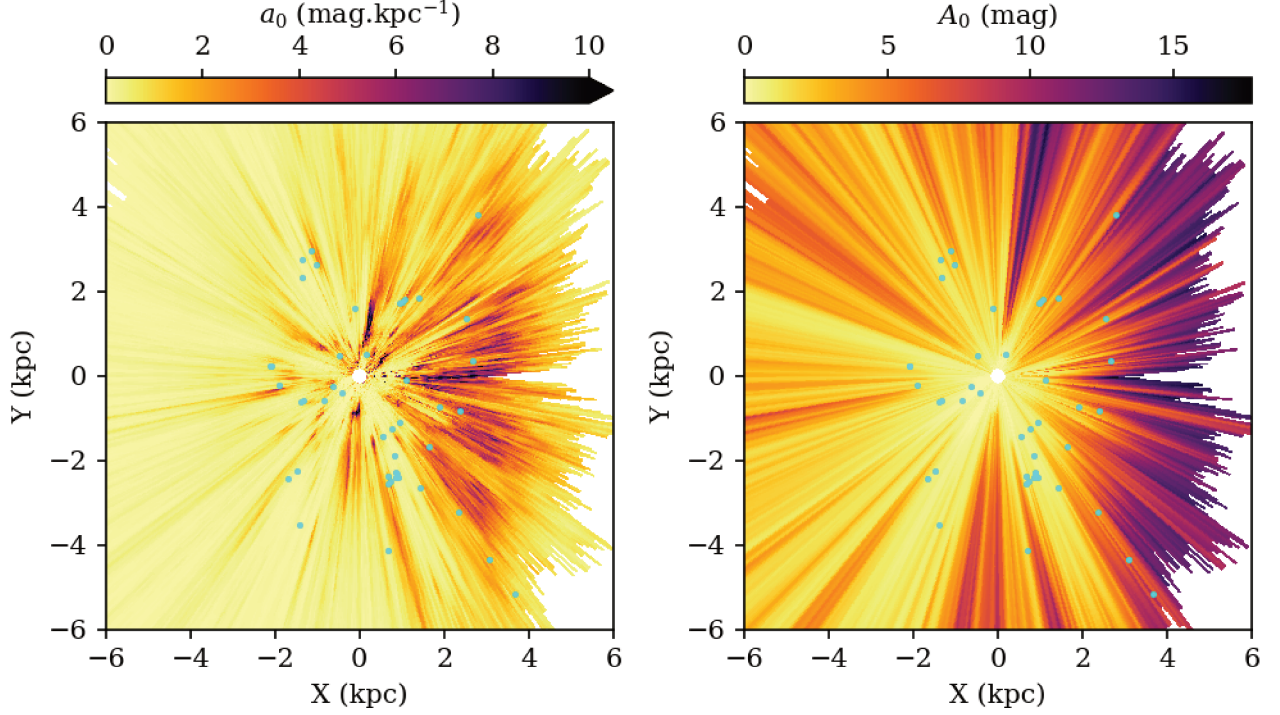


FIGURE 5.15 – *Gauche* : carte de densité d’extinction; *Droite* : carte d’extinction intégrée. En surimpression de ces deux cartes, on a tracé les amas ouverts dont la position a été calculée par Cantat-Gaudin et al. (2018); on trace uniquement les amas jeunes ( $\log(\text{age}) < 8.5$ ; Kharchenko et al. (2013)) avec  $|b| < 0.24^\circ$ .

## 5.4 Densité stellaire

À partir des résultats de FEDReD, nous sommes théoriquement en mesure de produire des cartes de densité d’étoiles dans le disque de la Voie lactée. Cependant, plusieurs points restent à éclaircir.

Le principal est la normalisation des résultats d’un champ à l’autre. En effet, les résultats pour l’extinction nous sont directement fournis en magnitude, il est donc relativement naturel de les assembler. Pour la distribution d’étoiles, nous calculons une distribution statistique après marginalisation du  $P(A_0, D)$  (voir section 4.2.2). Il nous faut donc normaliser ces différentes distributions de façon à être cohérent d’un champ de vue à l’autre pour produire une carte de densité d’étoiles.

Pour faire un premier test très approximatif, nous avons utilisé la normalisation suivante :

$$\int_0^{1\text{kpc}} \rho(D) dD = \text{cst} \quad (5.3)$$

pour l’ensemble des directions du plan. Nous obtenons ainsi la carte de densité stellaire présentée en Figure 5.20.

Bien que *ad hoc*, cette normalisation montre des informations intéressantes. On y constate que la densité d’étoile est bien entendue plus importante vers le centre galactique et le complexe du Sagittaire Carène, mais la zone de densité supérieure n’a pas la forme que donnerait un simple disque exponentiel, ce qui indique que ce n’est pas notre prior de densité qui crée cette distribution.

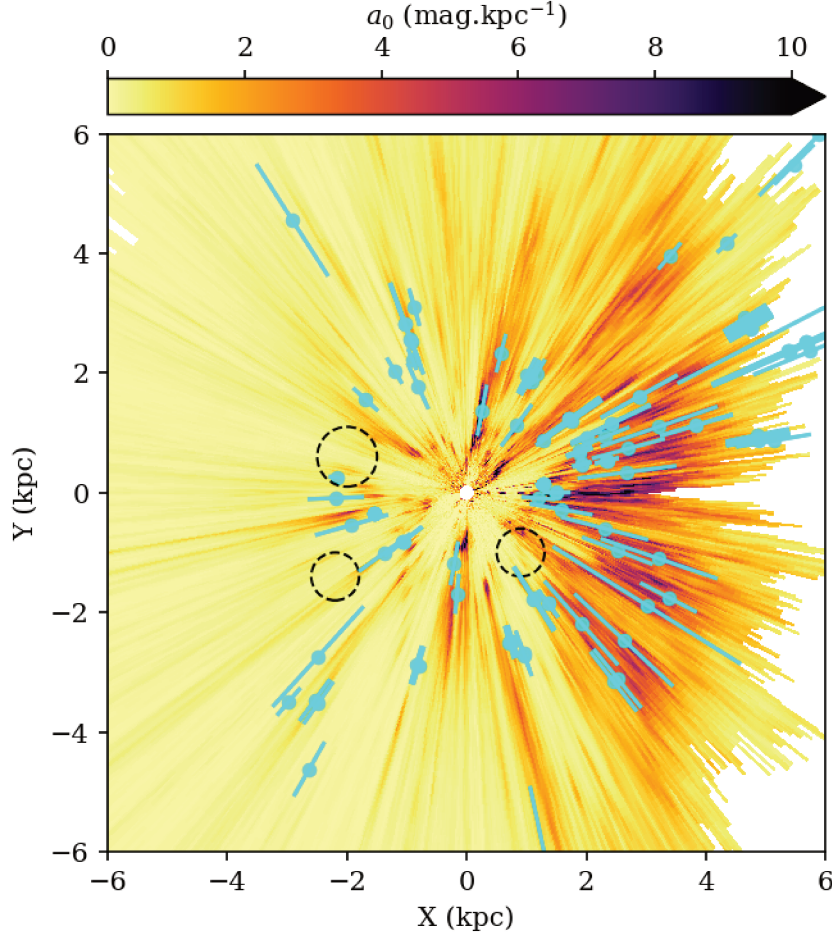


FIGURE 5.16 – Carte d’extinction avec les régions HII observées par Hou & Han (2014), On a tracé uniquement les régions HII dont on connaît la distance par de l’information stellaire et dont  $|b| < 0.24^\circ$ . Les bras spiraux ajustés sur ces régions HII sont présentés en Figure 5.19. Les cercles en pointillés correspondent aux trous dans la densité d’étoiles jeunes de Chen et al. (2019a).

Grâce à cette carte, on démontre le potentiel de FEDReD pour obtenir de l’information sur la distribution des étoiles, mais il nous faut encore travailler sur la normalisation des résultats.



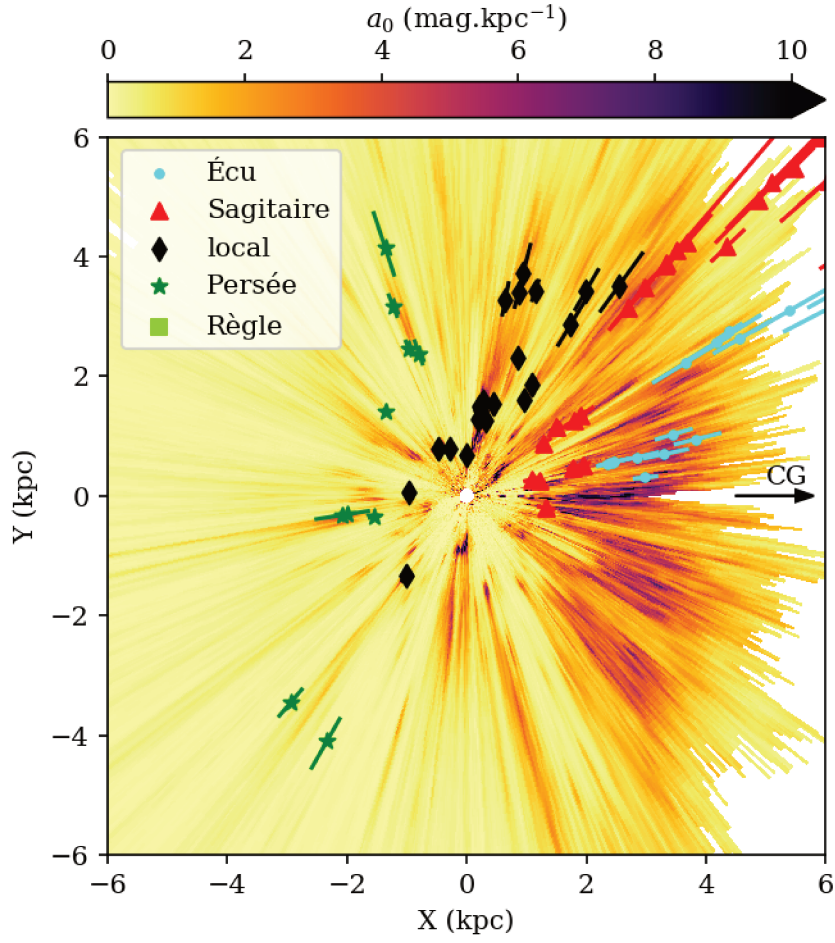


FIGURE 5.17 – Carte d’extinction avec les masers de Reid et al. (2014), Wu et al. (2019), Xu et al. (2016) et Sakai et al. (2019), chaque couleur est associée à un bras spiral (voir Figure 5.19). On a tracé les masers avec  $|z| < 0.05\text{kpc}$



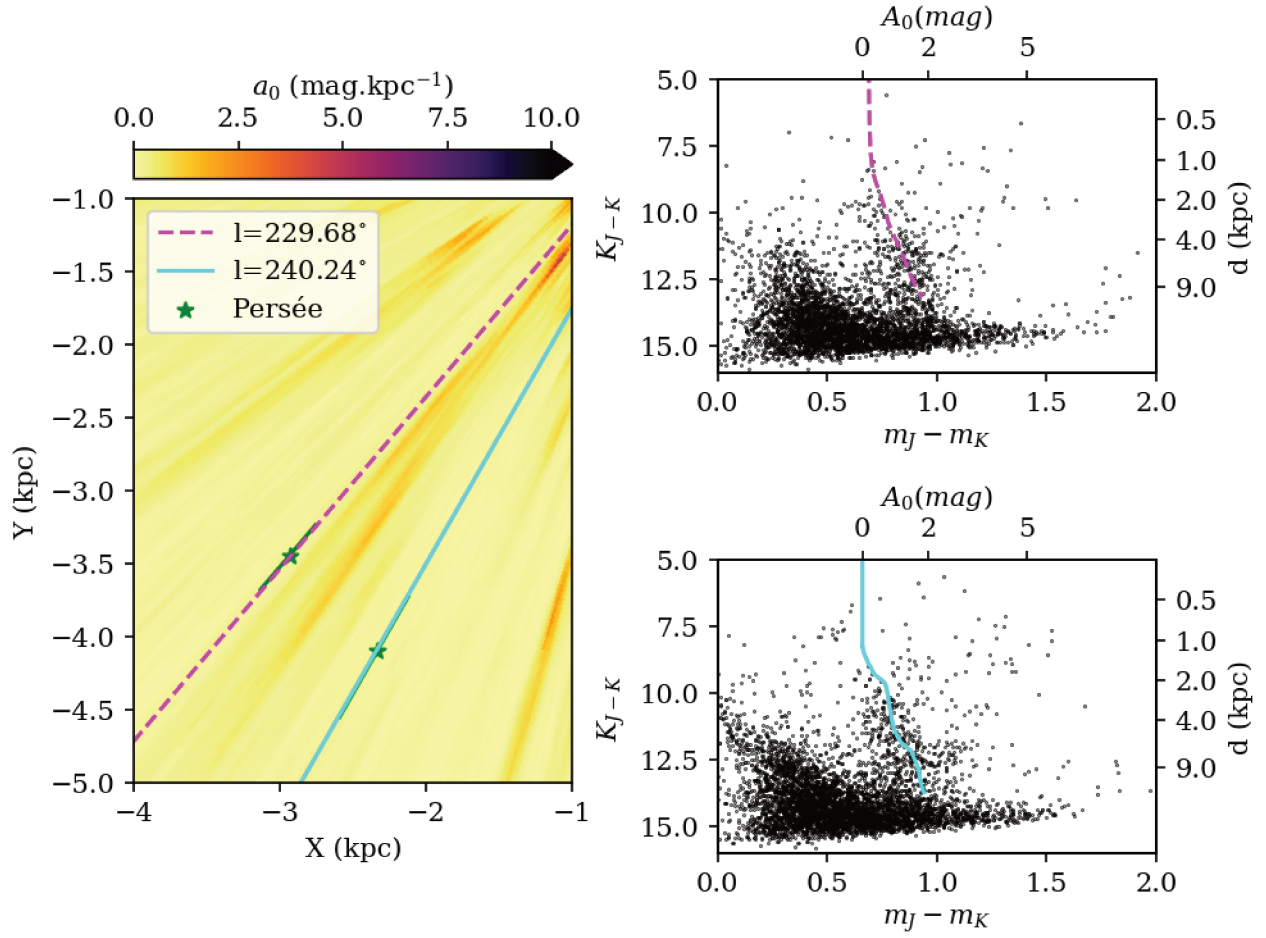


FIGURE 5.18 – *Gauche* : Agrandissement de la carte de densité d’extinction et marquage des lignes de visée traversant les masers ; *Droite* : CMD des champs de vue marqués et relation  $A_0(D)$  obtenue par FEDReD.

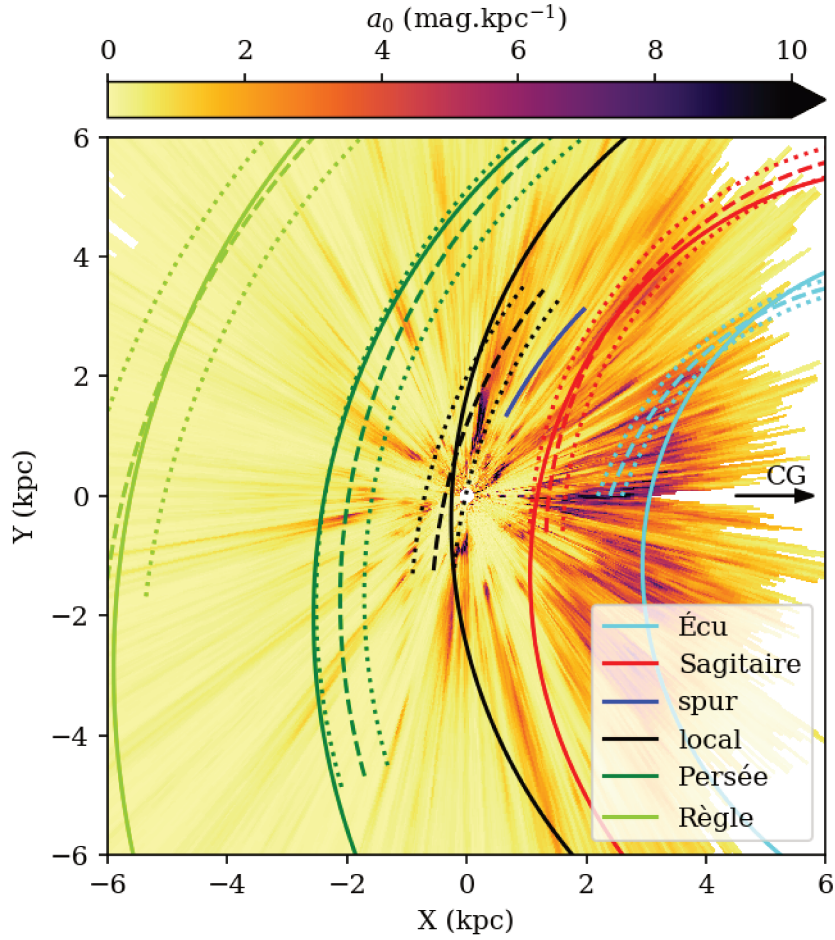


FIGURE 5.19 – Carte d’extinction, les bras spiraux de Reid et al. (2014) sont tracés en trait discontinu. Les bras spiraux de Hou & Han (2014) ajustés sur les régions HII sont tracés en trait continu, tout comme le spur de Xu et al. (2016)

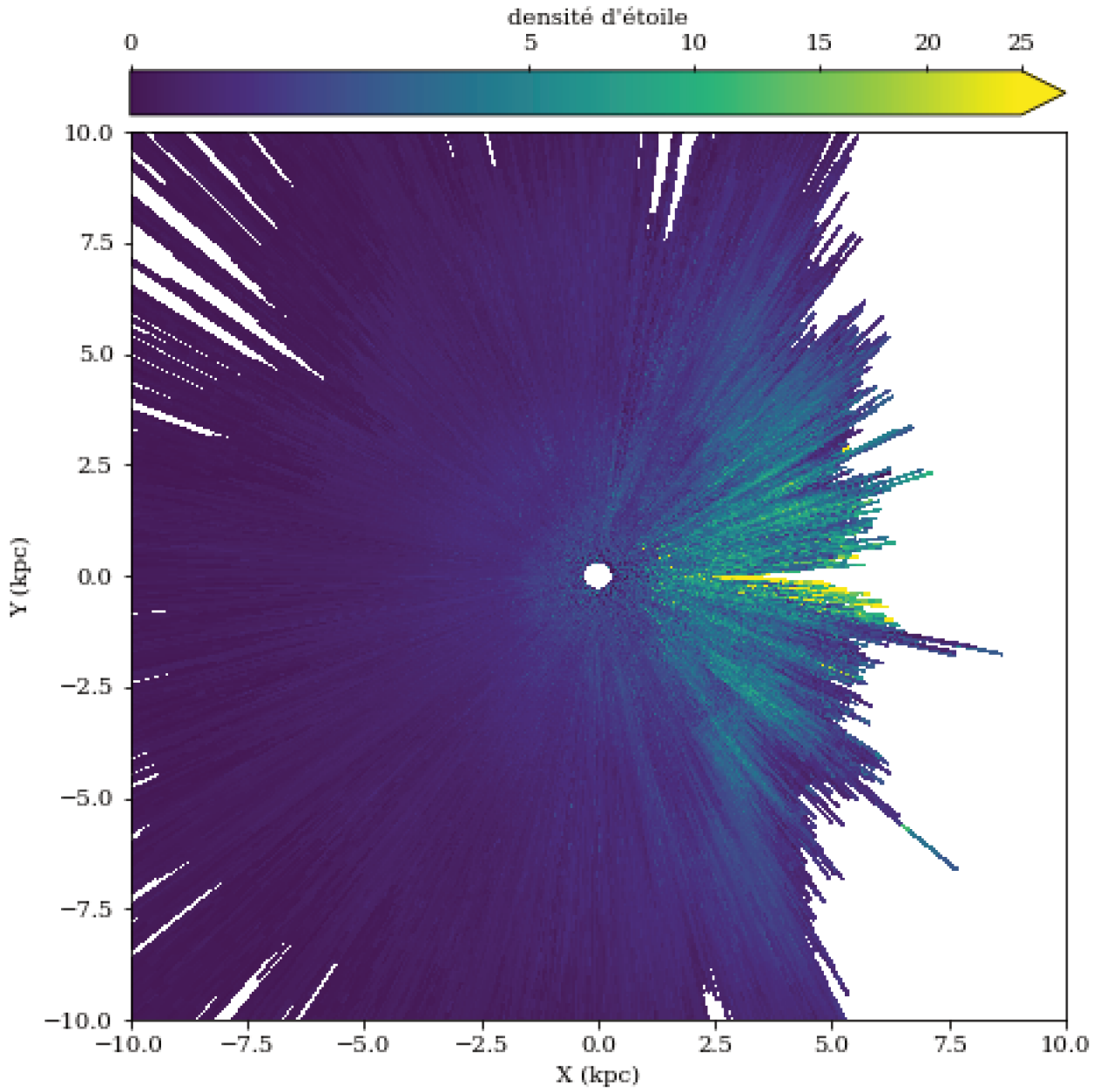


FIGURE 5.20 – Premier essai de carte de densité d'étoiles dans le plan galactique obtenue par FEDReD avec les catalogues 2MASS et GDR2. La normalisation simpliste des champs est réalisée selon l'équation (5.3)

# Chapitre 6

## Conclusion et Perspectives

### 6.1 Conclusion

**P**OUR étudier la structure du disque mince de la Voie lactée, nous avons utilisé l'algorithme Field Extinction - Distance Relation Deconvolver (FEDReD) dont la faisabilité et les premiers développements avaient été réalisés par Fourtune-Ravard (2016). Cet algorithme étudie la photométrie des étoiles d'un champ de vue afin d'en déduire la distribution de probabilité conjointe d'extinction et de distance.

Pour éviter la fluctuation des résultats avec les différents priors que l'on pouvait utiliser, nous avons affiné le processus de convergence de FEDReD. Nous avons aussi ajouté les données du catalogue Gaia - Data Release 2 (GDR2).

Pour pouvoir utiliser cet algorithme à grande échelle nous avons aussi travaillé sur des optimisations logicielles. Nous avons ainsi parallélisé plusieurs parties du code et significativement divisé le temps d'accès mémoire et l'espace disque occupé. Cela nous a notamment permis d'obtenir du temps de calcul sur le calculateur MésoPSL.

Nous avons donc séparé le disque galactique, vu par Two Micron All Sky Survey (2MASS), en 1 500 champs de vue avec  $b = 0^\circ$  et  $0^\circ \leq l \leq 360^\circ$ . Leur analyse par FEDReD a utilisé un peu plus de 60 000hCPU. L'assemblage de ces 1500 résultats a été fait à partir de plusieurs solutions pour chaque champ de vue construites par processus de Monte-Carlo (MCS). Ces solutions sont nettoyées des informations aberrantes par les champs de vue voisins et lissées pour produire une cartographie de la densité d'extinction interstellaire. Notre carte s'étend jusqu'à 4kpc vers le centre Galactique et peut dépasser les 8kpc dans les directions moins encombrées.

Cette cartographie de la densité d'extinction confirme la structure de l'extinction à faible distance déjà connue et valide ainsi l'implémentation de l'algorithme. Mais elle nous permet surtout d'apprécier leur prolongement à grande distance. On observe ainsi que le nuage de Vela est une structure en deux parties, une première avec une très forte extinction suivie par une bulle propre. On peut aussi appréhender la grande taille du complexe de la Carène, qui se prolonge jusqu'à 6 kpc, dont la taille en extinction n'était jusqu'alors pas montrée dans les cartes d'extinction. On peut aussi noter la présence d'un grand nombre de bulles, présentes dans beaucoup de directions. Par exemple la bulle du complexe du Sagittaire-Carène que l'on observe pour la première fois sur notre carte ou encore les bulles d'extinction que nous décrivons, comme le papillon, qui n'étaient que partiellement visibles dans des travaux précédents, mais qui apparaît dans sa totalité dans nos résultats.

Notre carte d'extinction est également en accord avec plusieurs autres traceurs du milieu interstellaire. Les régions HII et les nuages moléculaires sont présents en nombre dans les zones de forte extinction. Par ailleurs ils confirment aussi la morphologie de certaines structures comme

le complexe de la Carène, mais aussi la présence des bulles propres dans le nuage de Vela et dans le complexe du Sagittaire-Carène. Les différents masers, traceurs de bras spiraux, sont aussi particulièrement visibles en extinction (à l'exception de deux masers).

La signature des bras spiraux en extinction, quant à elle, dépend du bras considéré. En effet le bras de Persée, ou du moins la zone où les modèles le localisent, est rempli de lacunes entre diverses structures d'extinction plus fortes. On ne peut voir sa structure que par l'ajout de sa trace obtenue par l'observation d'autres traceurs. Le bras local est en revanche bien visible de part et d'autre du soleil, par ailleurs le Split du bras local, à grande distance, semble plutôt être associé au bras du Sagittaire. Ce dernier présente une signature forte en extinction, bien qu'il soit percé d'une lacune (l'aile du papillon dans le premier quadrant), mais on ne voit pas de séparation franche entre lui et le bras de l'Écu dans le complexe du Sagittaire-Carène.

## 6.2 Perspectives

La carte d'extinction que nous avons obtenue sera bientôt complétée par une carte de densité d'étoiles (tous types confondus) du disque galactique, qui permettra l'étude conjointe de ces composantes du disque mince. Les tests préliminaires ont déjà été effectués.

Nous étendrons aussi notre étude du disque galactique à des latitudes plus basses et plus élevées. On pourra ainsi observer d'autres structures qui ne sont pas visibles à  $b = 0^\circ$ , comme le bras extérieur. On pourra aussi étudier le gauchissement du disque galactique en extinction.

Par ailleurs, de par son point de vue bayésien et de son implémentation, il nous est facile d'analyser d'autres catalogues avec FEDReD. Nous avons d'ailleurs déjà commencé à tester FEDReD sur le couple UKIRT Infrared Deep Sky Survey (UKIDSS) et GDR2 avec des résultats prometteurs. En effet, UKIDSS observe des étoiles de plus faibles luminosités que 2MASS et nous permettra donc d'explorer des distances plus grandes, mais aussi d'augmenter significativement la résolution angulaire de nos résultats. Nous compléterons UKIDSS par le relevé VVV qui offre les mêmes caractéristiques que UKIDSS mais avec des observations depuis l'hémisphère sud nous permettant d'explorer le disque galactique à  $360^\circ$ . On envisage aussi d'ajouter des données venant du catalogue WISE, c'est un relevé du ciel complet en infrarouge. On augmentera ainsi le nombre de bandes photométriques.

On pourra aussi utiliser les nouveaux résultats des *releases* successives de Gaia. Par exemple, le 3<sup>me</sup> catalogue apportera un plus grand nombre d'étoiles avec des parallaxes dans les champs encombrés, permettant ainsi à la déconvolution de mieux converger. Par ailleurs, on pourra aussi analyser la spectrophotométrie pour mieux contraindre notre modèle.

Il nous sera aussi possible d'exploiter les résultats intermédiaires de FEDReD que sont les densités individuelles des étoiles  $P(O | A_0, D)$ . En effet, nous avons calculé la distribution conjointe de distance et d'extinction des champs de vue ; on pourrait donc calculer la distribution de probabilité de la distance et d'extinction de chacune des étoiles et en déduire leurs magnitudes absolues. Nous serions donc en mesure d'obtenir le type de ces étoiles et par conséquent la distribution d'étoiles par type dans le disque de la Voie lactée, notamment la distribution d'étoiles jeunes pour tracer les bras spiraux.

En profitant de la flexibilité de l'approche de FEDReD, on pourra aussi ajouter d'autres types d'information que la photométrie, par exemple de l'information astérosismologique ou spectroscopique dans des champs de vue particuliers pour obtenir une reconstruction plus fine des cartes dans des régions d'intérêt.

Enfin, on peut envisager d'appliquer des méthodes d'inversion globale, similaires à celle employée par Lallement et al. (2019), mais en ajoutant une prise en compte de la complétude pour obtenir des cubes d'extinction plus étendus que ce que l'on peut produire actuellement.

FEDReD est donc maintenant fonctionnel et offre de multiples perspectives pour l'exploration de notre Galaxie.



# Liste des acronymes

**2MASS** Two Micron All Sky Survey.

**BF** meilleur ajustement de  $A_0(D)$ .

**CMD** diagramme couleur magnitude.

**diagramme HR** diagramme de Hertzsprung-Russell.

**ESA** Agence Spatiale Européenne.

**FEDReD** Field Extinction - Distance Relation Deconvolver.

**GDR2** Gaia - Data Release 2.

**IMF** fonction de masse initiale.

**MCS** solution Monte Carlo.

**MIS** milieu interstellaire.

**SFR** taux de formation d'étoiles.

**UKIDSS** UKIRT Infrared Deep Sky Survey.

**UKIRT** United Kingdom Infra-Red Telescope.

**VLBI** very long base interferometry.

**WFCAM** Wide Field Camera.





# Bibliographie

- Arenou, F., Grenon, M., & Gomez, A. 1992, *Astronomy and Astrophysics*, 258, 104
- Arenou, F., Luri, X., Babusiaux, C., et al. 2018, arXiv :1804.09375 [astro-ph]
- Baba, J., Kawata, D., Matsunaga, N., Grand, R. J. J., & Hunt, J. A. S. 2018, *The Astrophysical Journal*, 853, L23
- Babusiaux, C. & Gilmore, G. 2005, *Monthly Notices of the Royal Astronomical Society*, 358, 1309
- Babusiaux, C., Hottier, C., Fourtune-Ravard, Arenou, F., & Gomez, A. 2019, in. prep
- Babusiaux, C., Katz, D., Hill, V., et al. 2014, *Astronomy and Astrophysics*, 563, A15
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 127
- Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *Astronomy & Astrophysics*, 618, A93
- Capitanio, L., Lallement, R., Vergely, J. L., Elyajouri, M., & Monreal-Ibero, A. 2017, *Astronomy & Astrophysics*, 606, A65
- Carpenter, J. M., Snell, R. L., & Schloerb, F. P. 1995, *The Astrophysical Journal*, 445, 246
- Chen, B.-Q., Huang, Y., Hou, L.-G., et al. 2019a, *Monthly Notices of the Royal Astronomical Society*, 487, 1400
- Chen, B.-Q., Huang, Y., Yuan, H.-B., et al. 2019b, *Monthly Notices of the Royal Astronomical Society*, 483, 4277
- Chen, B.-Q., Liu, X.-W., Yuan, H.-B., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 1192
- Chen, B. Q., Schultheis, M., Jiang, B. W., et al. 2013, *Astronomy and Astrophysics*, 550, A42
- Danielski, C., Babusiaux, C., Ruiz-Dern, L., Sartoretti, P., & Arenou, F. 2018, *Astronomy & Astrophysics*, 614, A19
- Drimmel, R., Cabrera-Lavers, A., & López-Corredoira, M. 2003, *Astronomy & Astrophysics*, 409, 205
- Evans, D. W., Riello, M., Angeli, F. D., et al. 2018, *Astronomy & Astrophysics*, 616, A4
- Fish, V. L., Reid, M. J., Argon, A. L., & Zheng, X.-W. 2005, *The Astrophysical Journal Supplement Series*, 160, 220
- Fourtune-Ravard, C. 2016, PhD thesis, Observatoire de Paris

- Fux, R. 1999, *Astronomy and Astrophysics*, 345, 787
- Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, *Astronomy & Astrophysics*, 616, A10
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018b, *Astronomy & Astrophysics*, 616, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *Astronomy & Astrophysics*, 595, A1
- Gorski, K. M., Hivon, E., Banday, A. J., et al. 2005, *The Astrophysical Journal*, 622, 759
- Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, arXiv preprint arXiv :1801.03555
- Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2014, *The Astrophysical Journal*, 783, 114
- Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, *The Astrophysical Journal*, 810, 25
- Green, G. M., Schlafly, E. F., Zucker, C., Speagle, J. S., & Finkbeiner, D. P. 2019, arXiv :1905.02734 [astro-ph]
- Herschel, W. 1785, *Philosophical Transactions of the Royal Society of London Series I*, 75, 213
- Hildebrand, R. H. 1983, *Quarterly Journal of the Royal Astronomical Society*, 24, 267
- Hodgkin, S. T., Irwin, M. J., Hewett, P. C., & Warren, S. J. 2009, *Monthly Notices of the Royal Astronomical Society*, 394, 675
- Hou, L. G. & Han, J. L. 2014, *Astronomy & Astrophysics*, 569, A125
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, *Astronomy and Astrophysics*, 558, A53
- Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, arXiv :1902.04116 [astro-ph]
- Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, *Astronomy & Astrophysics*, 561, A91
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1599
- Lindgren, L. 1995, *Astronomy & Astrophysics*, 304, 61
- Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, *Astronomy & Astrophysics*, 616, A2
- Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, *Monthly Notices of the Royal Astronomical Society*, 391, 136
- Lucy, L. B. 1974, *The Astronomical Journal*, 79, 745
- Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, *Astronomy & Astrophysics*, 621, A144
- Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., & Picaud, S. 2006, *Astronomy & Astrophysics*, 453, 635
- Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2016, *The Astrophysical Journal*, 834, 57

- Ng, P. & Maechler, M. 2007, *Statistical Modelling : An International Journal*, 7, 315
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *The Astrophysical Journal*, 783, 130
- Rezaei Kh., S., Bailer-Jones, C. A. L., Hanson, R. J., & Fouesneau, M. 2017, *Astronomy & Astrophysics*, 598, A125
- Rezaei Kh., S., Bailer-Jones, C. A. L., Hogg, D. W., & Schultheis, M. 2018, *Astronomy & Astrophysics*, 618, A168
- Richardson, W. H. 1972, *Journal of the Optical Society of America*, 62, 55
- Robin, A. C., Reyl  , C., Derri  re, S., & Picaud, S. 2003, *Astronomy & Astrophysics*, 409, 523
- Ruiz-Dern, L., Babusiaux, C., Arenou, F., Turon, C., & Lallement, R. 2018, *Astronomy & Astrophysics*, 609, A116
- Sakai, N., Reid, M. J., Menten, K. M., Brunthaler, A., & Dame, T. M. 2019, *The Astrophysical Journal*, 876, 30
- Sale, S. E. 2012, *Monthly Notices of the Royal Astronomical Society*, 427, 2119
- Sale, S. E., Drew, J. E., Barentsen, G., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 2907
- Schla  ly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, *The Astrophysical Journal*, 786, 29
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *The Astrophysical Journal*, 500, 525
- Schultheis, M., Chen, B. Q., Jiang, B. W., et al. 2014, *Astronomy and Astrophysics*, 566, A120
- Skowron, D. M., Skowron, J., Mr  z, P., et al. 2018, arXiv :1806.10653 [astro-ph]
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *The Astronomical Journal*, 131, 1163
- Trumpler, R. J. 1930, *Publications of the Astronomical Society of the Pacific*, 42, 214
- Ungerechts, H., Umbanhowar, P., & Thaddeus, P. 2000, *The Astrophysical Journal*, 537, 221
- Vall  e, J. P. 2017, *Astronomical Review*, 13, 113
- Vergely, J.-L., Valette, B., Lallement, R., & Raimond, S. 2010, *Astronomy and Astrophysics*, 518, A31
- Wu, Y. W., Reid, M. J., Sakai, N., et al. 2019, *The Astrophysical Journal*, 874, 94
- Xu, Y., Bian, S. B., Reid, M. J., et al. 2018, *Astronomy & Astrophysics*, 616, L15
- Xu, Y., Reid, M., Dame, T., et al. 2016, *Science Advances*, 2, e1600878
- Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, arXiv :1810.09819 [astro-ph]
- Zhang, X., Lee, Y., Bolatto, A., & Stark, A. A. 2001, *The Astrophysical Journal*, 553, 274





## RÉSUMÉ

---

L'étude des structures des poussières et des étoiles du disque Galactique est rendue possible grâce aux grands relevés d'étoiles comme le projet Gaia qui, avec sa seconde publication (GDR2), apporte un nombre de parallaxes jusque là inégalé. Cependant, la précision des parallaxes se dégrade avec la distance et avec l'extinction. Or c'est dans le disque de la Voie lactée que se concentre l'extinction. Il faut donc associer des données proche-infrarouge, moins sensibles à l'extinction, et utiliser une méthode statistique permettant d'obtenir des distances photométriques.

Nous avons donc combiné l'astrométrie de la GDR2 Gaia avec de la photométrie dans l'infrarouge des catalogues 2MASS et UKIDSS, que nous avons analysés avec une méthode bayésienne d'analyse de champs de vue, FEDReD (Field Extinction - Distance Relation Deconvolver). Cet algorithme permet d'obtenir la relation entre l'extinction et la distance dans le champ de vue, mais aussi l'évolution de la densité d'étoiles avec la distance.

Nous avons appliqué FEDReD à l'ensemble du disque Galactique, ce qui nous a permis d'obtenir une carte en densité d'extinction. FEDReD peut tracer l'extinction et la densité stellaire sur une distance de 5 kpc pour les champs les plus denses, et jusqu'à 10 kpc pour les moins encombrés.

À courte distance, FEDReD est à même de confirmer la présence et la forme de plusieurs structures d'extinction connues, comme la division du bras local, le nuage Vela, ou bien le bras spiral du scutum qui ont été détectés à l'aide d'autres méthodes. FEDReD améliore aussi la carte d'extinction à plus grandes distances en mettant en évidence le prolongement de certaines de ces structures, permettant in fine d'améliorer notre connaissance du disque galactique.

## MOTS CLÉS

---

Gaia, Structure Galactique, Extinction, Approche Bayésienne

## ABSTRACT

---

We aim to study the structures of dust and stars of the Galactic disk thanks to large star surveys. With its second Data Release (GDR2), the Gaia satellite brings a huge number of parallaxes; however, the parallax precision degrades with distance and with extinction. And it is in the galactic disc that the dust is most present. It is then mandatory to associate near-infrared photometry, less sensible to extinction, and use statistical methods to infer photometric distances.

We therefore combined GDR2 Gaia astrometry with the 2MASS and UKIDSS infrared photometric catalogues, which we analysed using a Bayesian field of view analysis, FEDReD (Field Extinction - Distance Relationship Deconvolver). This algorithm makes it possible to obtain the relationship between extinction and distance in a given field of view, and also estimates the stellar density with distance.

Applying FEDReD to the entire Galactic disk, we obtained an extinction density map. FEDReD can trace extinction and stellar density to a distance of about 5 kpc for crowded fields, and up to 10 kpc for the least dense fields.

At small distances, FEDReD is able to confirm the presence and shape of several known extinction structures, such as the split of the local arm, the Vela cloud, or the spiral arm of the scutum, which are all detected using other methods. In addition, FEDReD improves the extinction map at larger distances by highlighting the extension of some of the known structures, and in fine improves our understanding of the Milky Way disk.

## KEYWORDS

---

Gaia, Galactic Structure, Extinction, Bayesian Approach