



**HAL**  
open science

# Calibration des données spectroscopiques de la mission Gaia

Antoine Guerrier

► **To cite this version:**

Antoine Guerrier. Calibration des données spectroscopiques de la mission Gaia. Astrophysique [astro-ph]. Observatoire de Paris, 2008. Français. NNT : . tel-01958561

**HAL Id: tel-01958561**

**<https://theses.hal.science/tel-01958561v1>**

Submitted on 18 Dec 2018

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

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

Observatoire de Paris

École doctorale d'Astronomie et d'Astrophysique d'Île-de-France

# Calibration des données spectroscopiques de la mission Gaia

## THÈSE

présentée et soutenue publiquement le 9 Janvier 2008

pour l'obtention du

Doctorat de l'Observatoire de Paris  
(spécialité informatique)

par

Antoine Guerrier

### Composition du jury

*Président :* Benoît Mosser

*Rapporteurs :* Mark Cropper  
Jorge Torra

*Examineurs :* François Mignard  
Ronny Blomme  
Gérard Jasniewicz

*Directeurs :* Catherine Turon & David Katz

Observatoire de Paris

École doctorale d'Astronomie et d'Astrophysique d'Île-de-France

# Calibration des données spectroscopiques de la mission Gaia

## THÈSE

présentée et soutenue publiquement le 9 Janvier 2008

pour l'obtention du

Doctorat de l'Observatoire de Paris  
(spécialité informatique)

par

Antoine Guerrier

### Composition du jury

*Président :* Benoît Mosser

*Rapporteurs :* Mark Cropper  
Jorge Torra

*Examineurs :* François Mignard  
Ronny Blomme  
Gérard Jasniewicz

*Directeurs :* Catherine Turon & David Katz



Mis en page avec la classe thloria.

## Remerciements

Au cours de ces trois années, des personnes m'ont fait confiance, n'ont pas hésité à me donner des responsabilités et surtout, m'ont entouré dans les moments difficiles. En leur contact, j'ai pu m'épanouir tant sur le plan professionnel que personnel. Pour tout ce qu'ils m'ont apporté, je tiens à remercier mes directeurs de thèse, Catherine Turon et David Katz.

Je remercie particulièrement tous les membres de mon jury, Benoît Mosser, Mark Cropper, Jordi Torra, François Mignard, Gérard Jasniewicz et Ronny Blomme, d'avoir accepté d'être juges de mon travail.

Je tiens à remercier les responsables du Master "Outils et Systèmes de l'Astronomie et de l'Espace", Benoît Mosser et Alain Abergel, qui m'ont encouragé à réaliser cette thèse.

Merci à toutes les personnes du bâtiment 11 : Annick, Frédéric, Yves, Ernest, Paola, Carine, Nicolas, Céline, Françoise, Misha, Olga, Anita, Noël, Danielle et ceux qui n'y sont plus : Vanessa, François, Monique. J'espère que je n'oublie personne.

Doctorat en science rime difficilement avec formalités administratives... ainsi, je remercie Jacqueline, Annick et Olga qui ont pris de leur temps pour m'aider.

Merci à Shan et à Aurélie, les deux autres "souffre-douleur" du bâtiment 11. Merci à Thierry, le "Théoricien de la Physique".

Je remercie tous les membres de la CU6 de Gaia d'avoir prêté attention à mes présentations laborieuses sur les calibrations en longueurs d'onde. Je tiens entre autre à remercier Nathalie pour m'avoir aiguillé dans ma recherche d'emploi.

Merci aux copains du DESS, même si on se voit moins qu'on aimerait ! Jérôme, Jean-Marc, Léo, Claire, Rodrigue, Farid. Cette fois-ci, je peux le dire : "C'est la Fin !!!".

Merci à Jean-Philippe, Cyril, Aurélie, Violaine, David, Alex, JY, Mathieu. Pardon à ceux que j'oublie.

Merci à toute ma famille de m'avoir supporté durant ces 3 années. Merci à mon père et à ma mère d'avoir préparé ce délicieux buffet...

Enfin, je tenais à remercier la personne qui compte le plus pour moi, ma femme Élise.

*À mes grand-pères ...*

# Table des matières

## Introduction générale

### Chapitre 1

#### Gaia et le Radial Velocity Spectrometer 5

1.1	Présentation de Gaia . . . . .	6
1.1.1	Intérêt scientifique de la mission . . . . .	6
1.1.2	Le satellite et sa charge utile . . . . .	8
1.1.3	Tour d’horizon (non-exhaustif) . . . . .	14
1.2	La spectroscopie de Gaia . . . . .	16
1.2.1	Intérêt scientifique de la spectroscopie . . . . .	16
1.2.2	Présentation de l’instrument Radial Velocity Spectrometer . . . . .	17
1.2.3	Présentation des spectres du Radial Velocity Spectrometer . . . . .	19

### Chapitre 2

#### Les calibrations spectroscopiques 23

2.1	L’utilité de la calibration . . . . .	23
2.2	Les processus de calibration spectroscopique au sol . . . . .	24
2.2.1	Traitement des données brutes . . . . .	25
2.2.2	Étalonnage des données prétraitées . . . . .	25
2.3	La calibration en longueur d’onde . . . . .	26
2.3.1	Approche classique de la calibration en longueur d’onde au sol . . . . .	26
2.3.2	Différentes idées pour calibrer la loi de dispersion spectrale du Radial Velocity Spectrometer . . . . .	27
2.4	La calibration en longueur d’onde du Radial Velocity Spectrometer : la Spectroscopic Global Iterative Solution . . . . .	28
2.4.1	La philosophie générale de la Spectroscopic Global Iterative Solution . . . . .	29

2.4.2	Aperçu de la Spectroscopic Global Iterative Solution au sens des calibrations en longueur d'onde . . . . .	30
2.4.3	Concepts mathématiques . . . . .	31

<b>Chapitre 3</b>	
<b>Implémentation de la Spectroscopic Global Iterative Solution</b>	<b>35</b>

3.1	Données d'entrée et de sortie . . . . .	36
3.2	Architecture et fonctionnement . . . . .	39
3.2.1	Architecture . . . . .	39
3.2.2	Fonctionnement général du prototype . . . . .	40
3.2.3	Procédure d'initialisation . . . . .	46
3.2.4	Source Updating : Estimation des vitesses radiales . . . . .	47
3.2.5	Reference Selection : Sélection des étoiles de référence . . . . .	50
3.2.6	Calibration Updating : Calibration en longueur d'onde . . . . .	50
3.3	Spectroscopic Global Iterative Solution, au niveau du consortium de traitement de données de Gaia . . . . .	52

<b>Chapitre 4</b>	
<b>Détection et localisation des raies de référence</b>	<b>57</b>

4.1	Description du jeu de données . . . . .	59
4.2	Détection de raies dans les spectres bruts . . . . .	61
4.2.1	Description d'une raie observée . . . . .	61
4.2.2	Principe de "détection" . . . . .	61
4.2.3	Limites de la méthode de détection . . . . .	63
4.3	Localisation de raies de référence par ajustement parabolique . . . . .	64
4.3.1	Principe de la localisation des raies de référence . . . . .	64
4.3.2	Performance de la méthode . . . . .	69
4.3.3	Calibration des biais . . . . .	71
4.3.4	Performances après calibration des biais . . . . .	72
4.4	Localisation des raies de référence par ajustement du profil des raies . . . . .	76
4.4.1	Principe de la double corrélation croisée . . . . .	76
4.4.2	Recherche du maximum de la surface de corrélation . . . . .	77
4.4.3	Optimisation de la méthode . . . . .	79
4.4.4	Performance de l'ajustement de profil . . . . .	81
4.5	Localisation accélérée de raies de référence par ajustement de profil . . . . .	83



4.6 Synthèse . . . . .	86
------------------------	----

<b>Chapitre 5</b>	
<b>Étude de convergence de SGIS</b>	<b>87</b>

5.1 Description de l'implémentation . . . . .	88
5.1.1 Principe de la méthode de centroïding . . . . .	88
5.1.2 Conditions initiales . . . . .	90
5.2 Diagnostics de contrôle . . . . .	94
5.2.1 Diagnostic sur l'échelle en longueur d'onde . . . . .	95
5.2.2 Diagnostic sur les vitesses radiales . . . . .	96
5.2.3 Diagnostics au niveau de SGIS . . . . .	97
5.3 Centroïding avec et sans biais . . . . .	97
5.3.1 Performance du processus itératif sans biais . . . . .	98
5.3.2 Performance de la solution itérative avec biais . . . . .	102
5.3.3 Synthèse . . . . .	105

<b>Chapitre 6</b>	
<b>Performances du prototype de calibration en longueur d'onde</b>	<b>111</b>

6.1 Jeu de données GASS . . . . .	112
6.1.1 Introduction aux simulateurs de données de Gaia . . . . .	112
6.1.2 Définition des données GASS . . . . .	113
6.1.3 Structure des données GASS . . . . .	115
6.2 Mise en place des données GASS pour la calibration . . . . .	118
6.2.1 Validation des données GASS . . . . .	118
6.2.2 Redimensionnement et redressement . . . . .	119
6.2.3 Création d'une librairie de templates . . . . .	120
6.3 Performances du prototype . . . . .	122
6.3.1 Configuration du prototype . . . . .	122
6.3.2 Performances sans standards . . . . .	125
6.3.3 Performances avec standards . . . . .	129
6.3.4 Temps de Calcul . . . . .	131
6.3.5 Synthèse . . . . .	133

<b>Conclusion et perspectives</b>
-----------------------------------

*Table des matières*

---

**Table des figures** 143

**Liste des tableaux** 151

**Bibliographie**

**Bibliographie** 153

## Résumé

Le spectromètre de GAIA, le "Radial Velocity Spectrometer" (RVS), va collecter les spectres de 100 à 200 millions d'étoiles jusqu'à la magnitude  $V \approx 17$ . Le RVS ne possède pas de module de calibration (e.g. pas de lampe de calibration) à bord du satellite. L'instrument ne peut pas être calibré en utilisant une approche "classique".

L'objectif de cette thèse est de développer une méthode de calibration originale de l'instrument RVS à partir de ses propres observations spectroscopiques. Cette méthode est appelée "Spectroscopic Global Iterative Solution" (SGIS). Le principe de SGIS est d'utiliser la stabilité de certaines étoiles pour mesurer l'évolution de l'instrument au cours du temps. Cette méthode estime itérativement les vitesses radiales des étoiles et les calibrations en longueur d'onde du RVS.

Durant la thèse, un prototype de SGIS a été développé afin de valider le principe de la méthode et d'évaluer ses performances dans le cadre de Gaia. Ce travail est effectué dans le cadre de la préparation de la réduction des données du satellite GAIA. La méthode de calibration développée fait partie de la chaîne de traitement des données spectroscopiques de Gaia. Le prototype développé a été testé sur 6 mois de données Gaia simulées. Après 6 itérations, la précision des échelles en longueur d'onde atteint une cinquantaine de mètres par seconde.

**Mots-clés :** Gaia, Radial Velocity Spectrometer, calibration, processus itératif, longueurs d'onde, vitesse radiale.

## Abstract

The Radial Velocity Spectrometer (RVS) of Gaia will collect spectra for 100 to 200 million stars up to magnitude  $V \approx 17$  over the wavelength range 847 to 874 *nm* with a resolving power  $R = 11500$ . The RVS possess no calibration device (such as a lamp). It cannot be calibrated with the "classical" approach based on dedicated calibration exposure.

The aim of this thesis is to provide an efficient alternative to calibrate the RVS instrument. This new method is called Spectroscopic Global Iterative Solution (SGIS). The principle of SGIS is to use a set of "well-behaved stars" (bright, stable and of a well-suited spectral type). This method iteratively derive the radial velocity of the stars and the wavelength scale of the RVS instrument.

During this thesis, a prototype of SGIS has been developed in order to validate the concept of SGIS and to evaluate its performance within the context of Gaia. The method of the wavelength scale calibration is part of the spectroscopic data processing of Gaia. The developed prototype has been tested on 6 months of simulated data. After 6 iterations, the precision on the wavelength scale is about 50 meters per second.

**Keywords :** Gaia, Radial Velocity Spectrometer, calibration, iterative process, wavelength, radial velocity.

# Introduction générale

Le succès de la mission européenne Hipparcos, arrêtée en Mars 1993, a apporté une nouvelle méthode de mesure des positions et des distances des étoiles dans l'espace. Cette mission a permis d'établir un catalogue de positions, de mouvements propres et de parallaxes pour 118218 étoiles de notre Galaxie. Avec la mission Hipparcos, les précisions astrométriques atteignent la milli seconde de degré, signant ainsi la "renaissance" de l'astrométrie. Comme l'illustre la figure 1, la précision sur les mesures de position a gagné environ un facteur cent avec Hipparcos par rapport au catalogue UCAC2.

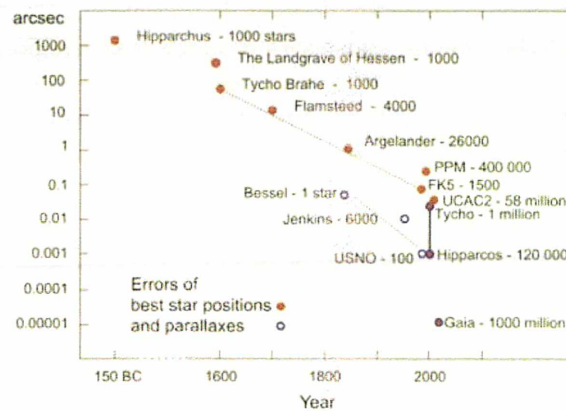


FIG. 1 – Progrès de la précision astrométrique au cours du temps. Erik Høg

Fort de ce succès, l'Agence Spatiale Européenne (ESA) veut prolonger cet "essor astrométrique" en lançant le satellite Gaia. Héritant des expériences et des connaissances acquises avec Hipparcos, le satellite Gaia observera plus d'un milliard d'étoiles de notre Galaxie (dix mille fois plus que la mission Hipparcos). Les parallaxes sont annoncées avec une précision de quelques micro arcsecondes. En parallèle à l'astrométrie, le satellite intégrera à bord un instrument de spectro-photométrie et un instrument de spectroscopie.

Pour avoir un tel nombre d'étoiles, le satellite observera en continu toute la sphère céleste durant les cinq années de mission (prolongeable d'une année). Le fait d'observer en continu complexifie les processus de calibration des instruments à bord du satellite. En effet, dans des processus "classiques" de calibration, une phase d'observation dédiée est utilisée afin d'évaluer les variations instrumentales. En mode continu, il n'y a aucune phase d'observation dédiée

aux calibrations. De plus, il n'y a aucun dispositif de calibration (par exemple, pas de lampe de calibration) à bord du satellite Gaia. Sans observations dédiées, on ne peut pas calibrer les instruments avec des méthodes classiques. On a donc besoin d'une méthode spécifique pour les calibrations. Ainsi, mon travail de recherche a consisté à développer une méthode originale pour la calibration en longueur d'onde du "Radial Velocity Spectrometer" (RVS). Cette méthode originale de calibration sera incluse dans la chaîne de réduction de données de la mission Gaia.

Les aspects généraux de la mission Gaia et du spectromètre sont introduits dans le **chapitre 1**. Dans le **chapitre 2**, nous présentons la problématique générale de ma thèse, c'est-à-dire la calibration en longueur d'onde.

Un prototype de calibration en longueur d'onde a été développé durant la thèse. Les choix d'implémentation et le fonctionnement général de ce prototype sont discutés dans le **chapitre 3**. Un des points cruciaux du prototype de calibration en longueur d'onde est l'étape de détection et de localisation de raies de référence dans un spectre. Différentes méthodes ont été développées. Ces méthodes et leurs performances sont décrites dans le **chapitre 4**. En parallèle au développement des méthodes de détection et de localisation de raies de référence, une étude de convergence a été réalisée. Cette étude, présentée dans le **chapitre 5**, a permis de comprendre dans quelles conditions le processus itératif converge. Elle a également permis de sélectionner une méthode de détection et de localisation de raies de référence.

Le **dernier chapitre** de la thèse est consacré à des tests avec des données simulant la configuration la plus récente du spectromètre RVS. Ce chapitre décrit les données utilisées lors de ces tests et présente les performances obtenues par le prototype de calibration en longueur d'onde.

## Contexte de la thèse

Le satellite Gaia va produire un large volume de données brutes : environ 50 GB de données scientifiques non-compressées seront collectées par jour, et plus de 100 TB de données à la fin de la mission (Mignard et al. (2007)). Le contenu scientifique de la mission étant enfoui dans ce large volume, depuis mi-juin 2006, un consortium de traitement de données, nommé "Data Processing and Analysis Consortium" (DPAC), a été mis en place. Le DPAC va rendre exploitable scientifiquement les données d'observations. Au vu du large volume de données, l'exploitation des données s'annonce comme une tâche difficile : si la réduction de données mettait une seconde par étoile, il faudrait plus de 30 années pour traiter le milliard d'étoiles de la mission spatiale.

Le DPAC est réparti en plusieurs unités de coordination, propres à des actions spécifiques :

- **CU1** : Structure et architecture des traitements de données
- **CU2** : Simulation de données de test
- **CU3** : Traitement des données astrométriques
- **CU4** : Objets du système solaire, binaires et objets étendus
- **CU5** : Traitement des données photométriques
- **CU6** : Traitement des données spectroscopiques
- **CU7** : Traitement des systèmes d'objets variables
- **CU8** : Détermination des paramètres astrophysiques
- **CU9** : Catalogue de données Gaia (unité qui sera activée dans quelques années)

Le traitement des données spectroscopiques est sous la responsabilité de l'une de ces unités

---

de coordination, la "Coordination Unit 6 : Spectroscopic processing". Cette "CU" est divisée, comme chacune des autres unités, en plusieurs "top-level work package" (TWP)<sup>1</sup>. L'un de ces "TWP" est consacré au calibration de l'instrument RVS.

Mes travaux de recherche s'inscrivent dans le cadre de la CU6. Ils ont pour objet de fournir une implémentation de la calibration en longueur d'onde des données spectroscopiques. Le prototype de calibration en longueur d'onde devra s'intégrer à la chaîne de réduction des données spectroscopiques, basée au Centre National d'Etudes Spatiales (CNES) de Toulouse.

Durant ces trois années, mes travaux ont été réalisés en collaboration avec plusieurs équipes européennes :

- Équipe Gaia du CNES de Toulouse, impliquée dans le segment sol de la mission Gaia.
- Équipe Gaia du Mullard Space Science Laboratory (University College of London, UK), impliquée dans plusieurs aspects du traitement spectroscopique des données Gaia, dont la calibration du RVS.
- Équipe Gaia de l'Université de Barcelone, impliquée dans le développement de simulateurs de données de la mission Gaia.

Au vu de l'aspect technique de mes recherches, mes travaux ont été co-financés par le CNES, membre actif dans le segment sol de la mission, et par EADS Astrium, le "prime contractor" de la mission Gaia.

---

<sup>1</sup>Toute l'organisation de la CU6 et les "TWP" sont décrits en détail dans le "Plan de développement logiciel" ("Software Development Plan" en anglais) de Katz et al (2006)



# Chapitre 1

## Gaia et le Radial Velocity Spectrometer

### Sommaire

---

<b>1.1</b>	<b>Présentation de Gaia</b> . . . . .	<b>6</b>
1.1.1	Intérêt scientifique de la mission . . . . .	6
1.1.2	Le satellite et sa charge utile . . . . .	8
1.1.3	Tour d’horizon (non-exhaustif) . . . . .	14
<b>1.2</b>	<b>La spectroscopie de Gaia</b> . . . . .	<b>16</b>
1.2.1	Intérêt scientifique de la spectroscopie . . . . .	16
1.2.2	Présentation de l’instrument Radial Velocity Spectrometer . . . . .	17
1.2.3	Présentation des spectres du Radial Velocity Spectrometer . . . . .	19

---

Après le succès d’Hipparcos, l’Agence Spatiale Européenne (ESA) veut révolutionner l’astrométrie contemporaine avec l’ambitieuse mission Gaia. Gaia effectuera des mesures de positions, de flux et de vitesses radiales avec une grande précision pour un grand nombre d’objets de notre Galaxie : on parle de plus d’un milliard d’étoiles en astrométrie (environ 1% de la population stellaire de la Galaxie). Une cartographie tri-dimensionnelle de notre Galaxie pourra donc être dressée, nous permettant d’élargir nos connaissances sur la structure et l’évolution de la Voie Lactée.

La charge utile du satellite est composée de trois instruments : un instrument astrométrique, un spectro-photomètre et un spectromètre. Ces instruments disposent de deux télescopes et d’un unique plan focal sur lequel se combinent les deux champs de vue du satellite. Gaia peut être considéré comme un véritable observatoire spatial.

Les trois principales qualités de Gaia sont :

- L’observation complète de la sphère céleste.
- L’observation à plusieurs époques d’un grand nombre d’objets (plus d’un milliard d’objets observés durant 5 ans de mission). En astrométrie, une étoile est en moyenne observée 80 fois permettant ainsi une étude de variabilité.
- Trois instruments complémentaires : l’instrument astrométrique fournit la position, la distance et les mouvements propres de l’objet observé. La spectro-photométrie et la spectroscopie complètent les propriétés astrophysiques et cinématiques de l’objet.



La mission Gaia a été sélectionnée en octobre 2000 comme l'une des pierres angulaires du programme scientifique de l'ESA. Depuis février 2006, le projet est en phase B2 (phase de définition détaillée) sous la maîtrise d'œuvre d'EADS Astrium Satellites (c.f. figure 1.1). Le satellite sera lancé en décembre 2011 et sa durée de mission est prévue pour cinq ans plus une extension de une année. Le catalogue final de la mission devrait être publié autour de 2020.

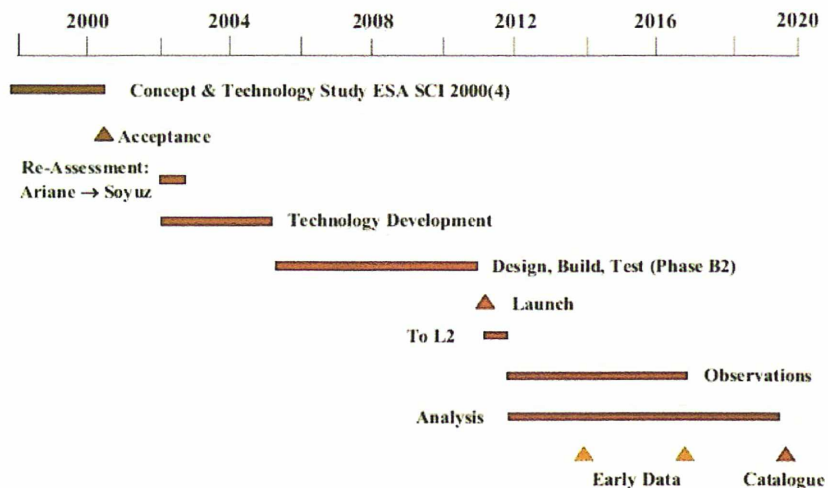


FIG. 1.1 – Planning de la mission Gaia, ESA

## 1.1 Présentation de Gaia

### 1.1.1 Intérêt scientifique de la mission

Pour comprendre la structure et l'évolution de la Galaxie, trois approches sont nécessaires :

1. Une approche statique : il faut être capable de décrire la structure spatiale actuelle de la Galaxie à partir de distances et de positions ;
2. Une approche cinématique et dynamique : une connaissance des mouvements spatiaux (i.e. mouvements propres et vitesses radiales) pour déterminer le champ de gravité et les orbites stellaires ;
3. Une approche chimique : un recensement de populations stellaires représentatives des composantes de la Galaxie.

Gaia va répondre à ces besoins, permettant ainsi de clarifier notre compréhension sur la composition, l'histoire et l'évolution de la Galaxie.

Gaia va également apporter des contributions significatives dans les domaines des exoplanètes, du système solaire, des galaxies extérieures et de la physique fondamentale. Voici une partie des contributions scientifiques de Gaia listées par grands thèmes de manière non-exhaustive.

### **Physique Galactique et extra-Galactique :**

- Détermination des structures spatiale et cinématique de toutes les populations d'étoiles, dans toutes les parties de notre Voie Lactée : disque mince, disque épais, bras spiraux, bulbe et barre, halo, amas globulaires, zones de formation d'étoiles, amas ouverts.
- Détermination de la métallicité et de l'âge des étoiles de différentes populations, âge des plus vieux objets.
- Détection d'étoiles déficientes (ou riches) en métaux.
- Dynamique et parallaxe de rotation des galaxies du Groupe Local.
- Photométrie de plus de  $10^6$  galaxies.
- Détection de plus de  $10^5$  supernovae.

### **Physique stellaire :**

- Statistiques de tous les types d'étoiles, y compris dans les phases d'évolution rapide du diagramme de Hertzsprung-Russell.
- Détermination des paramètres fondamentaux (masse, rayon, luminosité, température et composition chimique).
- Détection de la binarité et de la variabilité : détection d'environ  $10^7$  binaires, et de plus de  $10^7$  étoiles variables.
- Contraintes pour les modèles de structure et d'évolution stellaire.

### **Échelle de distance et système de référence :**

- Détermination des distances directes dans les nuages de Magellan pour les étoiles les plus brillantes.
- Distance des Céphéides (environ quelques milliers de Céphéides détectées par Gaia) et des RR Lyrae ( $7 \times 10^4$  détectées) permettant de recalibrer l'échelle des distances dans l'univers.
- Réalisation d'un système de référence à partir de  $5 \times 10^5$  quasars à des latitudes galactiques  $|b| > 20^\circ - 25^\circ$ .

### **Système Solaire :**

- Détection de  $5 \times 10^5$  astéroïdes, provenant en grande partie de la ceinture principale.
- Détermination des masses et des vitesses de rotation d'une partie de ces astéroïdes.
- Détermination d'orbites améliorées par un facteur 30.
- Détection de plus de  $10^5$  géocroiseurs.

### **Systèmes planétaires extra-solaires :**

- Détection astrométrique de  $3 \times 10^4$  planètes extrasolaires de masse similaire à Jupiter jusqu'à environ 200 pc, plusieurs milliers d'orbites.
- Détections de plusieurs milliers de transits de planètes extrasolaires devant leur étoile hôte.

**Physiques fondamentales :**

- La présence de la masse du soleil (et des autres planètes du système solaire) est à l'origine d'une déflexion des rayons lumineux de chaque étoile. Dans un formalisme post-newtonien, cette déflexion est proportionnelle à  $(1 + \gamma)/2$  où le paramètre  $\gamma$  vaut 1 dans le cadre de la relativité générale. Gaia devrait obtenir une précision sur  $\gamma$  de l'ordre de  $5 \times 10^{-7}$ , fournissant ainsi un test supplémentaire de la relativité générale.
- Grâce aux observations de la luminosité des naines blanches ( $2 \times 10^5$  naines blanches observées par Gaia), Gaia devrait fournir une contrainte quant à une variation (hypothétique) de la constante gravitationnelle.

La mission Gaia est impliquée dans un large panel d'applications de l'astrophysique moderne (Perryman et al. (2001); Mignard (2005), c.f. figure 1.2).

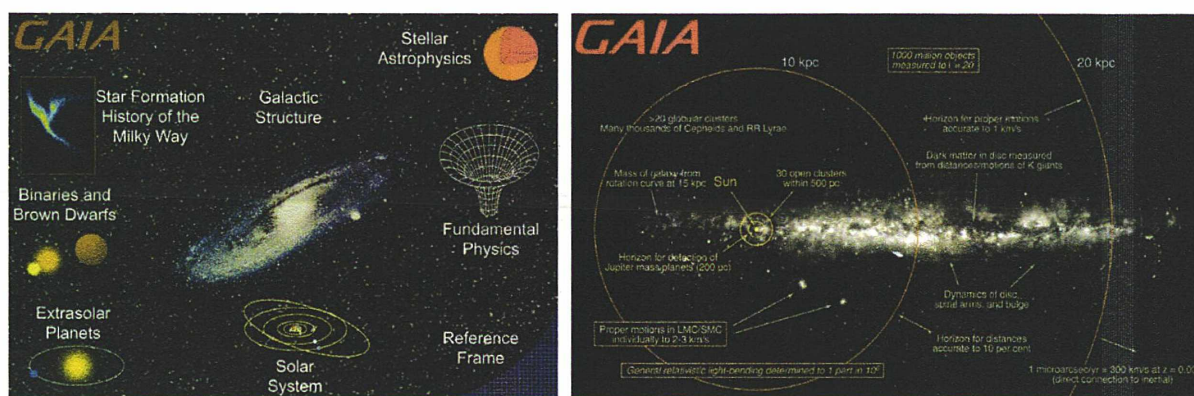


FIG. 1.2 – Intérêts scientifiques de Gaia : (a) Domaines scientifiques abordés par Gaia (ESA) (b) Précision de Gaia dans la Galaxie (Observatoire de Lund, Suède)

**1.1.2 Le satellite et sa charge utile**

La mission sera lancée de Kourou par un lanceur Soyuz-Fregat<sup>2</sup> en décembre 2011. Le satellite, représenté dans la figure 1.4, sera placé sur une orbite Lissajous autour du point L2 (l'orbite est illustrée dans la figure 1.3). Une fois la mise à poste effectuée, il sera en liaison en moyenne 11 heures par jour avec l'antenne Cebreros<sup>3</sup>, antenne en bande X de 35 mètres de diamètre.

**Le satellite**

Le principal but de la mission Gaia est de faire une astrométrie globale (Lindegren, L. and Perryman, M.A.C. (1996); Pace (2005)) : Gaia devra relier les étoiles observées avec une grande distance angulaire, créant ainsi tout un réseau d'observations connectées à un grand nombre d'autres observations dans toutes les directions. Pour répondre au besoin de cette astrométrie

<sup>2</sup>une nouvelle génération entièrement numérisée sera utilisée. Cette génération a déjà été utilisée pour le lancement de Corot, le 27 décembre 2006

<sup>3</sup>antenne construite à Ávila, en Espagne, par l'ESA, pour les missions Venus Express et Rosetta

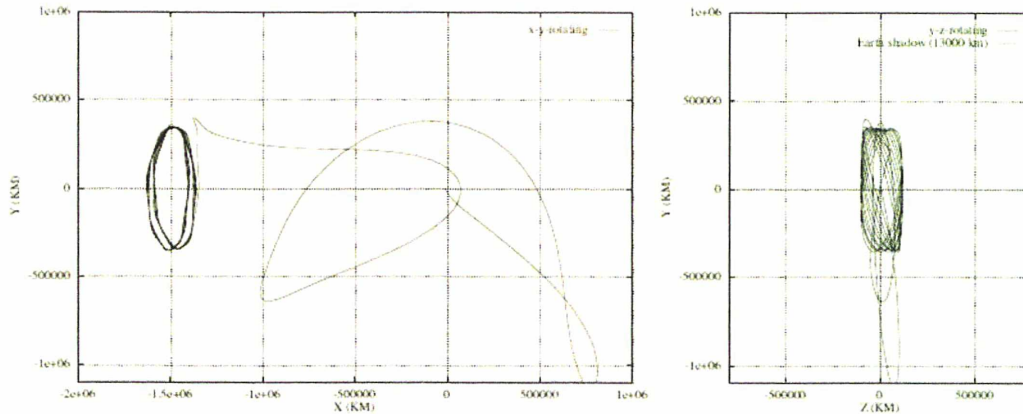


Fig. 1.3 – Représentations typique d'orbites de transfert (à gauche) et de Lissajous (à droite) pour la mission Gaia (Hechler et al. (1999))

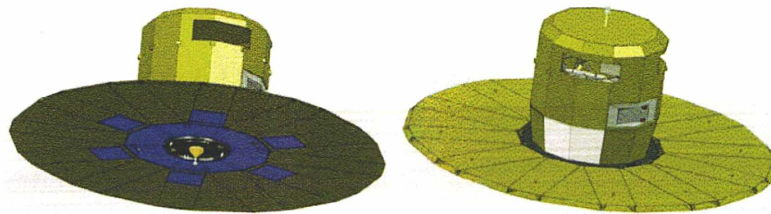


Fig. 1.4 – Représentations du satellite Gaia (EADS Astrium, 2006)

globale, Gaia va observer simultanément dans deux champs de vue différents. Comme Hipparcos, le banc optique de Gaia dispose donc de deux télescopes (ASTRO1 et ASTRO2), formant deux images combinées sur l'unique plan focal (c.f. figure 1.7 a). L'angle entre les deux télescopes, nommé "basic angle", est égal à  $106.5^\circ$  (c.f. figure 1.5).

Pour observer un grand nombre d'objets dans toutes les directions, le satellite devra balayer de manière répétée l'ensemble de la sphère céleste sur les cinq années de mission. Une loi de balayage, similaire à celle d'Hipparcos, est donc utilisée : le satellite fera une rotation lente et régulière (i.e. 1 degré par minute) autour de l'axe perpendiculaire au plan décrit par les deux champs de vue. Cet axe de rotation aura un mouvement de précession d'environ 63 jours autour de la direction Soleil-Terre. Avec une telle loi de balayage, Gaia décrira un quasi-grand cercle sur le ciel en six heures (c.f. figure 1.5).

L'axe de rotation du satellite fera un angle de  $45^\circ$  avec la direction du Soleil. Cet angle est un compromis entre les spécifications scientifiques de la mission (e.g. précisions astrométriques) et les contraintes (i.e efficacité des panneaux solaires et protection de la charge utile contre le rayonnement solaire). Les principales caractéristiques du satellite sont résumées dans la table 1.1.

La loi de balayage du satellite a été définie de telle sorte que le mouvement de précession soit minimisé et que la couverture de la sphère céleste soit la plus uniforme possible durant les 5

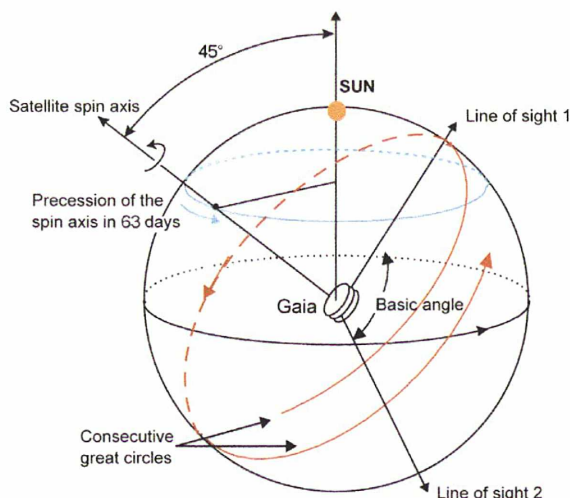


FIG. 1.5 – Principe d’observation de Gaia : la loi de balayage. ESA, 2005

années de mission (GAIA Science Advisory Group (2000), c.f. figure 1.6). Environ 10% à 20% du temps de la mission ne devrait pas contribuer aux observations scientifiques : indisponibilité de la liaison sol-satellite, contrôle d’orbite du satellite, contamination par des particules (e.g. radiation solaire ou rayonnement cosmique), collision avec des micro-météorites, etc.

Caractéristiques de Gaia	
Durée de mission	5 ans (+1 an d’extension)
Période de rotation	6 h
Angle Soleil/axe de rotation	45°
Période de précession	62.97 jours
Angle de base	106.5°

Tab. 1.1 – Principales caractéristiques du satellite Gaia

### La charge utile

La charge utile de Gaia est composée de trois instruments, de deux télescopes et d’un unique plan focal (c.f. figure 1.7 a). Le chemin optique de chacun de ces télescopes est composé de six éléments réflecteurs (M1-M6), dont deux sont communs (M5 et M6). Les lumières collectées par chacun de ces télescopes se combinent partiellement au niveau des miroirs M4/M’4 et forment une seule image sur le plan focal. Un système de masques optiques permet d’isoler chacun des champs de vue des télescopes au niveau du plan focal. Ces masques contribuent ainsi à l’identification des objets venant du champ de vue de ASTRO1 ou de ASTRO2, faite par l’intermédiaire des "sky-mappers" décrits plus loin.

Dans la suite du manuscrit, la direction de balayage du satellite sera notée AL (comme "along scan"), et la direction perpendiculaire à AL sera notée AC (comme "across scan").

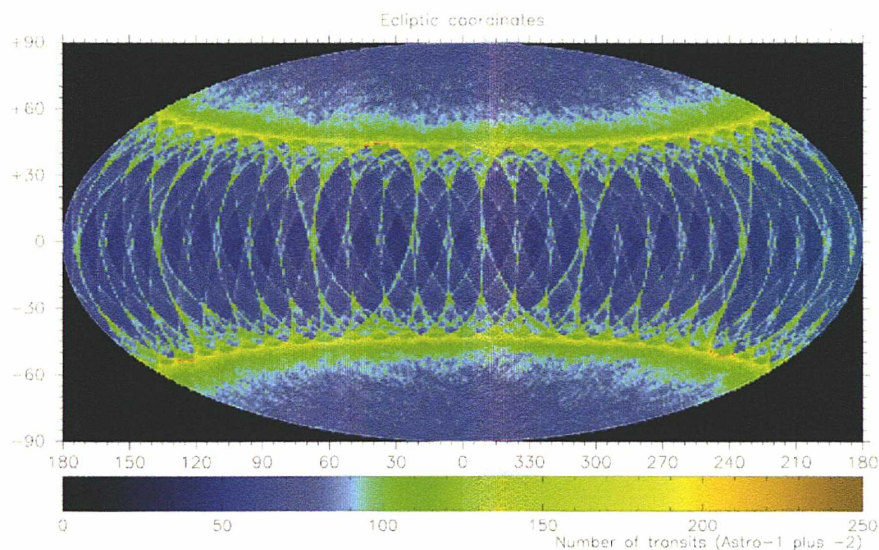


FIG. 1.6 – Nombre d’observations de la sphère céleste en fonction des coordonnées écliptiques (J. de Bruijne, ESA, 2003)

La table 1.2 résume les principales caractéristiques optiques de la charge utile.

Caractéristiques optiques	
<b>Nombre de pupilles</b>	2
<b>Taille de la pupille AL</b>	1.45 m
<b>Taille de la pupille AC</b>	0.50 m
<b>FoV AL</b>	1.18°
<b>FoV AC</b>	0.69°
<b>Longueur focale</b>	35 m

TAB. 1.2 – Caractéristiques optiques des télescopes

Une partie de la lumière combinée (i.e. lumière réfléchiée par M5 et M6) sera dispersée sur la direction AL par les instruments de spectro-photométrie et de spectroscopie. L’instrument spectro-photométrique dispersera la lumière sur deux bandes : la bande BP dans le domaine [330 – 680] nm et la bande RP dans le domaine [650 – 1000] nm. L’instrument spectroscopique dispersera dans le domaine proche infrarouge [847 – 874] nm.

La table 1.3 résume les principales caractéristiques spectrales de la charge utile.

Le plan focal de Gaia sera le plus large jamais développé pour une mission spatiale : 106 capteurs CCD formeront un plan focal de  $0.93 \times 0.42 \text{ m}^2$  (c.f. figure 1.8). Chaque CCD sera composé de 4500 pixels AL et de 1966 pixels AC (un pixel aura une dimension de  $10 \mu\text{m}$  AL et de  $30 \mu\text{m}$  AC). Ces 106 capteurs seront disposés en plusieurs bandes :

- 2 bandes de 7 capteurs pour les "sky-mappers" SM1 et SM2 : en l’absence de catalogue d’entrée (contrairement à Hipparcos), ces modules vont détecter, confirmer, puis prédire

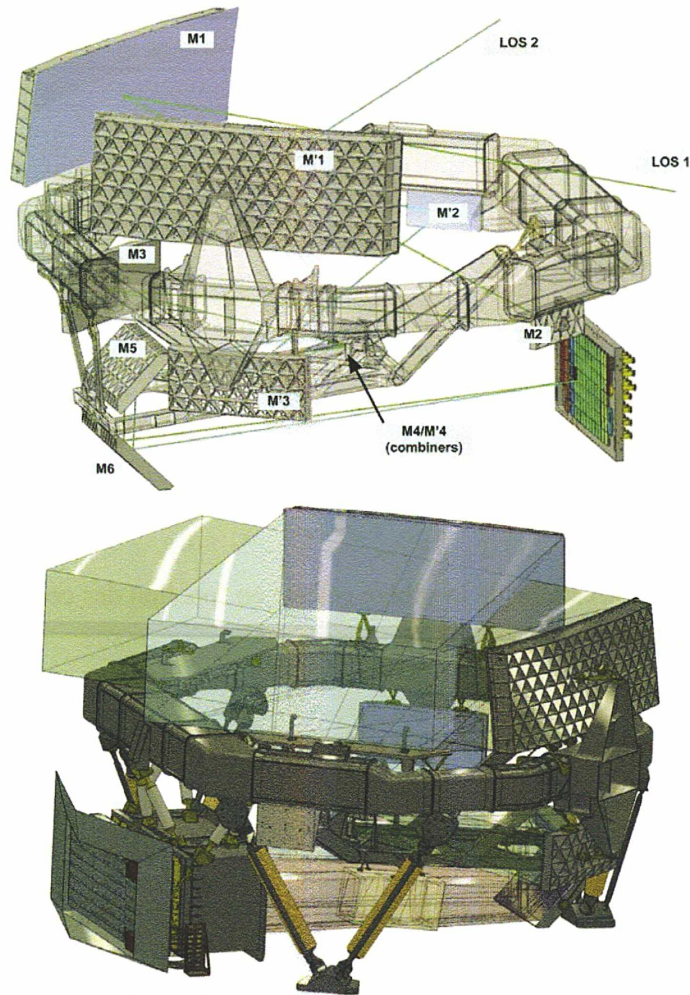


FIG. 1.7 – Charge utile : (a) Deux télescopes illuminant le même plan focal (en haut), (b) Aperçu de la charge utile assemblée (en bas) (EADS Astrium, 2006)

de manière autonome les positions et temps de passage d'un objet d'intérêt d'un champ de vue sur le reste du plan focal ; cela va donc "filtrer" les observations collectées par les instruments et ainsi alléger le budget télémétrique de la mission ; SM1 sera actif pour le champ du télescope ASTRO1 (la colonne de CCD est uniquement éclairée par le champ de vue de ASTRO1 et permet de suivre les objets venant de ce champ), et SM2 sera actif pour ASTRO2 (tous les nouveaux objets détectés, par rapport à ceux du SM1, sont associés au champ de ASTRO2).

- 8 bandes de 7 capteurs et 1 bande de 6 capteurs pour l'astrométrie.
- 2 bandes de 7 capteurs pour la spectro-photométrie : la première bande sera utilisée pour la dispersion en bande BP et la seconde pour la dispersion en bande RP.
- 3 bandes de 4 capteurs pour la spectroscopie.
- Les 4 capteurs restant seront utilisés pour de la "maintenance" : 2 capteurs (BAM, « Basic Angle Monitoring ») pour le contrôle de l'angle de base et 2 autres pour l'analyse de front

Caractéristiques spectrales				
	Astro	BP	RP	RVS
Domaine spectral	330 – 1000 nm	330 – 680 nm	650 – 1000 nm	847 – 874 nm
Dispersion spectrale		4 – 34 nm/pixel	7 – 15 nm/pixel	0.026 nm/pixel
Nombre de pixel AL par source et par observation	6 – 12	32	34	1104

TAB. 1.3 – Caractéristiques spectrales de Gaia

d'onde incident (WFS, « Wave Front Sensor »).

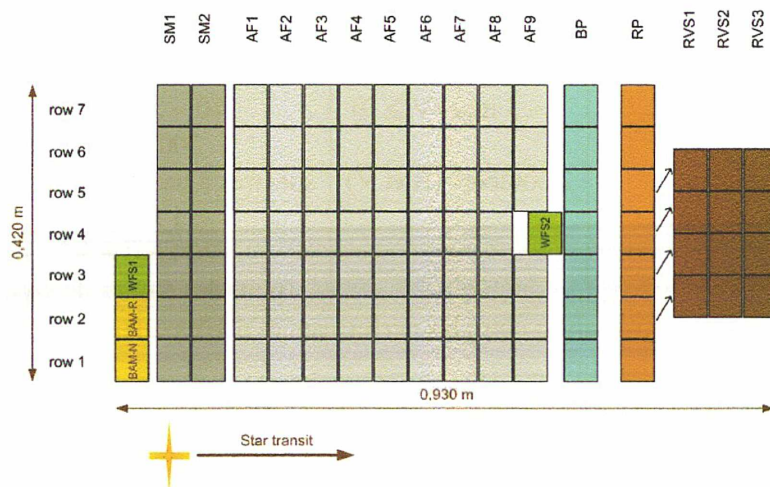


FIG. 1.8 – Plan focal du satellite Gaia, EADS Astrium 2006

Chaque CCD de Gaia va opérer en mode Time Delay Integration (Bastian et Biermann (2005)). Ce mode consiste à transférer les charges en continu vers le registre de lecture. Il n'y a pas de pause dédiée à la lecture des charges. Ce mode se révèle très utile pour enregistrer les images arrivant sur le plan focal d'une charge utile en rotation (c.f. la loi de balayage décrite précédemment). Avec la loi de balayage du satellite et les 35 mètres de distance focale des deux télescopes, l'image optique de chaque objet va avoir un mouvement constant sur le plan focal du satellite. Les charges photo-électriques sont accumulées successivement à vitesse constante régulée par le signal d'horloge de chaque capteur. Les objets seront donc collectés et intégrés tout le long de leur passage sur les CCDs. Ce temps d'intégration d'un objet par CCD est égal à 4.42 secondes.

## Performances

L'astrométrie de la mission va nous fournir des positions, des distances et des mouvements propres de plus d'un milliard d'étoiles de la Galaxie. La spectro-photométrie donnera des informations sur l'extinction. Elle fournira des luminosités, des magnitudes dans les deux bandes (BP et RP), des abondances chimiques et des âges. Enfin, la spectroscopie donnera accès aux



vitesse radiales de plus de 100 à 250 millions d'étoiles, complétant le trièdre cinématique<sup>4</sup>. Ce trièdre cinématique associé aux positions dans la Galaxie permet de déterminer les forces gravitationnelles et la distribution de matière noire. La spectroscopie permettra aussi d'accéder aux paramètres atmosphériques des étoiles (température effective  $T_{eff}$ , logarithme de la gravité de surface  $logg$ , métallicité  $[Fe/H]$ ) et aux abondances chimiques des étoiles les plus brillantes.

La table 1.4 dresse un bilan comparatif entre les performances de la mission Hipparcos et celles de Gaia.

	Hipparcos	Gaia
<b>Magnitude limite</b>	12	G = 20
<b>Limite de complétude</b>	7.3 – 9	G = 20
<b>Nombre d'étoiles</b>	$1.2 \cdot 10^5$	$10^9$ (G ≤ 20)
<b>Précision astrométrique</b>	1 mas (G ≤ 9) 1 – 3 mas (V ≥ 9)	7 μm (V = 10) 10 – 25 μm (G = 15) 300 μm (G = 20)
<b>Photométrie</b>	2 couleurs (B et V)	spectres à basse résolution (G ≤ 20)
<b>Vitesse radiale</b>	-	< 1 km/s à V = 13 20 km/s à V = 16.5 (pour des étoiles de type solaire)

Tab. 1.4 – Comparaison entre les performances d'Hipparcos et de Gaia

### 1.1.3 Tour d'horizon (non-exhaustif)

La première partie de cette section est consacrée à deux suivis spectroscopiques en cours : RAVE et SDSS/SEGUE. La deuxième partie porte sur deux projets spatiaux astrométriques en phase de développement : JASMINE et SIM.

#### Campagnes d'observations au sol : RAVE

Une campagne d'observation spectroscopique de la sphère céleste a débuté le 11 avril 2003 dans l'hémisphère sud. Elle utilise un spectromètre possédant une résolution et un domaine spectral proche de ceux du RVS (respectivement,  $R = 7500$  et  $841 - 879.5 \text{ nm}$ ). Cet instrument se nomme RAVE (pour "RADial Velocity Experiment"). Les intérêts de cette campagne sont les mêmes que ceux du RVS (c.f. section 1.2.1) : évolution, structure, cinématique et composition de notre Galaxie. La campagne RAVE aura pour but de collecter  $10^6$  spectres, de mesurer des vitesses radiales avec une précision meilleure que 3 km/s, ainsi que de fournir des paramètres stellaires (e.g. vitesses de rotation, températures effectives) et des abondances. Les premiers résultats ont été obtenus à partir d'observations dans l'hémisphère sud du 11/04/2003 au 02/04/2004 avec 240 champs couverts. La figure 1.9 représente la première publication des résultats de RAVE (Steinmetz et al (2006)). 25274 vitesses radiales ont été mesurées pour 24748 étoiles (en moyenne, un peu plus d'une mesure de vitesse radiale par étoile).

<sup>4</sup>le trièdre cinématique est composé des mouvements propres de l'objet et de la vitesse radiale

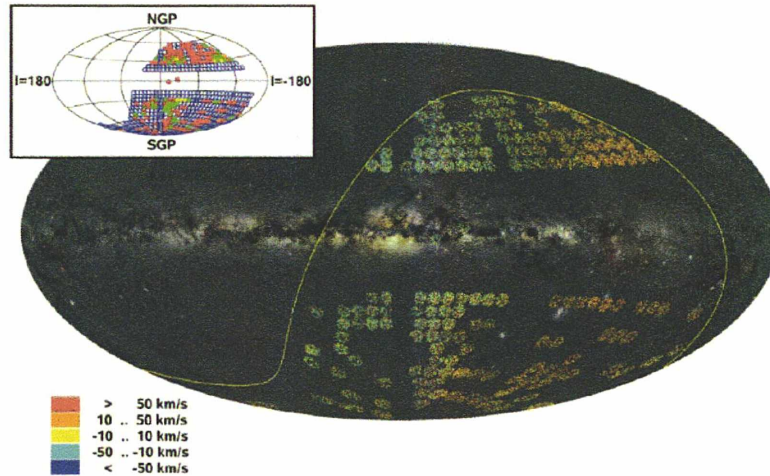


FIG. 1.9 – Champs observés par RAVE, projetés en coordonnées galactiques. Sur la figure centrale, les gradients de vitesses radiales héliocentriques sont représentés avec un code de couleur, mettant en valeur le mouvement du Soleil et la projection de différentes composantes de vitesse le long de la ligne de visée. La ligne jaune représente l'équateur céleste. Le cadre en haut à gauche représente les champs de vue observés avec un code de couleur : en rouge, les champs observés une fois, en jaune les champs observés deux fois, en vert pour trois fois et en marron pour quatre fois. Steinmetz et al (2006)

### Campagnes d'observations au sol : SDSS

La campagne d'observation grand champ SDSS entre dans une nouvelle phase avec trois suivis dont un spectroscopique : le « Sloan Extension for Galactic Understanding and Evolution » (SEGUE, Beers et al (2004)). Le but de cette campagne est d'améliorer notre compréhension de la formation et de l'évolution de notre Galaxie. Pour cela, SEGUE fournira les spectres basse résolution ( $R = 2000$ ) de  $2.4 \times 10^5$  étoiles, avec des mesures de vitesse radiale d'une précision de 10 km/s et des mesures d'abondance chimique d'une précision de 0.3 dex.

### Projets spatiaux

L'agence spatiale japonaise (la JAXA) prévoit une mission astrométrique dans l'infrarouge : il s'agit du projet JASMINE ("Japan Astrometry Satellite Mission for INfrared Exploration", c.f. Gouda et al (2005)). L'objectif de cette mission est de fournir des paramètres astrométriques afin d'étudier les structures et évolutions du disque et du bulbe de notre Galaxie. Pour cela, ce satellite devra mesurer des parallaxes, des positions avec des précisions de  $10 \mu\text{as}$  ainsi que des mouvements propres pour des étoiles plus brillantes que  $z = 14$  mag. JASMINE devra observer  $10^7$  étoiles du disque et du bulbe de notre Galaxie. Le lancement de la mission est prévu pour 2014 (c.f. figure 1.10).

Une mission américaine est aussi prévue par la NASA : il s'agit du projet SIM PlanetQuest. Le principal but est de détecter des planètes ayant une masse comparable à celle de la Terre autour des étoiles les plus proches. Cette mission apportera sa contribution dans la distribution de masse dans la Galaxie, la formation du halo, etc. Ce sera une mission d'astrométrie pointée :

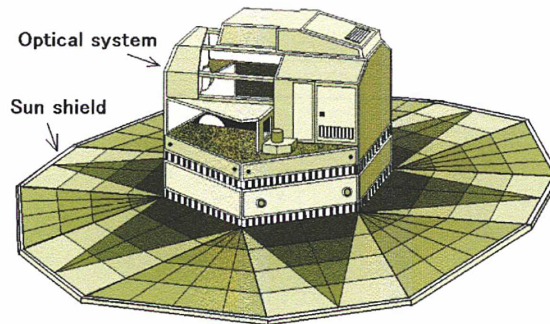


FIG. 1.10 – Projet JASMINE : vue du satellite. Le design de ce satellite est très similaire à celui de Gaia. (Site web de JASMINE : [www.jasmine-galaxy.org/whatis.html](http://www.jasmine-galaxy.org/whatis.html))

elle devra observer environ  $10^4$  cibles sélectionnées. La précision des mesures a été évaluée à  $4 \mu\text{s}$  jusqu'à la magnitude 19 en mode grand angle ( $15^\circ$ ), et environ  $1 \mu\text{s}$  dans de petits champs ( $1^\circ$ ). Le lancement, prévu pour 2015, a été repoussé pour une durée indéterminée.

## 1.2 La spectroscopie de Gaia

En 1997, il a été décidé d'inclure un instrument spectroscopique à la mission Gaia. Cet instrument se nomme le Radial Velocity Spectrometer (RVS) et permettra, entre autre, de mesurer les vitesses radiales des étoiles (c.f. Katz et al (2004)).

### 1.2.1 Intérêt scientifique de la spectroscopie

Les mesures de vitesses radiales avec le spectromètre de Gaia ont trois principaux intérêts (c.f. Wilkinson et al (2005)) :

1. Étude dynamique de la Galaxie : obtenir la troisième composante des mouvements spatiaux (les deux premières étant obtenues avec l'astrométrie), en mesurant la vitesse le long de la ligne de visée de l'instrument ;
2. Détection et caractérisation de systèmes binaires (c.f. figure 1.11) par mesure des vitesses radiales à plusieurs époques ;
3. Correction de l'« accélération de perspective ». Si la distance entre l'objet et l'observateur change au cours du temps à cause de la vitesse radiale de l'objet, une vitesse transverse constante est observée (Mignard (2003)) : c'est l'accélération de perspective. Cet effet est petit, mais cause des erreurs systématiques au niveau des mesures de distances pour des étoiles proches à très fortes vitesses radiales et à mouvements propres non-nuls. Cela concerne environ  $10^5$  étoiles.

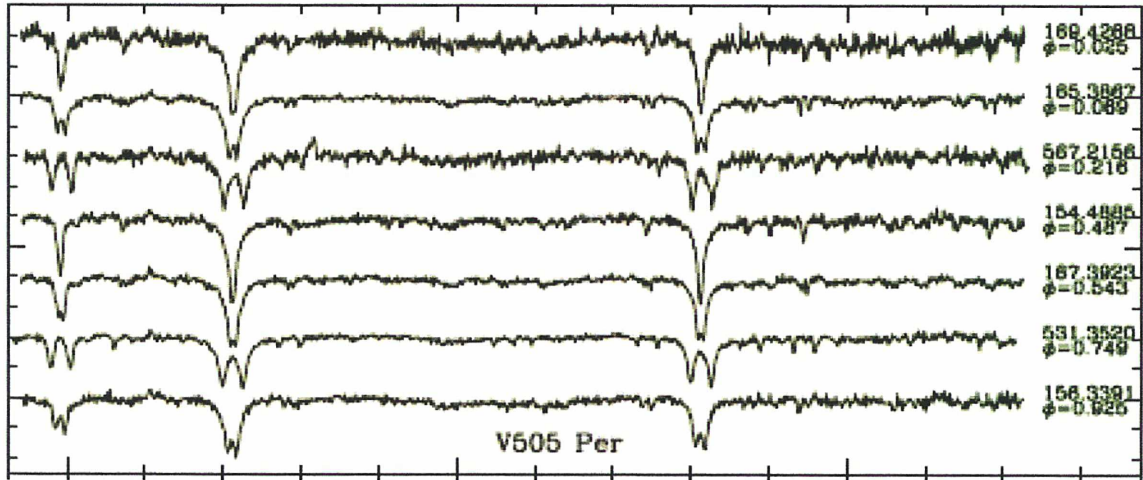


FIG. 1.11 – Exemple de détection d’une binaire à éclipse au niveau des spectres dans la bande RVS, Munari et al (2001)

### 1.2.2 Présentation de l’instrument Radial Velocity Spectrometer

Le RVS sera un spectromètre intégral de champ de moyenne résolution ( $R = \lambda/\Delta\lambda = 11500$ ). Par l’intermédiaire d’un ensemble dioptrique (un réseau et 4 prismes), le RVS dispersera la lumière provenant des deux télescopes ASTRO1 et ASTRO2<sup>5</sup> sur les 12 capteurs CCDs du plan focal de Gaia<sup>6</sup>. Comme les deux autres instruments de Gaia, le RVS balayera plusieurs fois la sphère céleste et observera un objet 40 fois en moyenne, durant les 5 ans de mission.

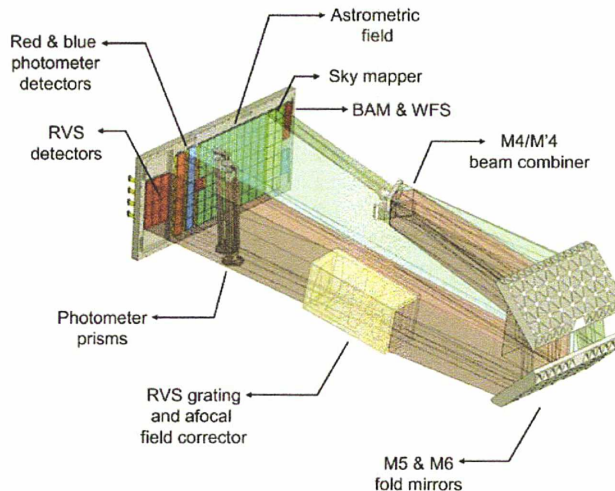


FIG. 1.12 – Fonctionnement du spectromètre de Gaia. Le chemin optique du RVS est représenté en marron clair. (EADS Astrium Satellite 2006)

<sup>5</sup>les caractéristiques optiques des télescopes dans la table 1.2

<sup>6</sup>le plan focal est représenté en figure 1.8

Le plan focal du RVS sera composé de 12 CCDs : 3 capteurs en AL  $\times$  4 en AC (c.f. figure 1.8), et tous opéreront en mode TDI ( $t_{\text{TDI}} = 4.42$  sec). Au niveau du plan focal en mode TDI, les spectres seront lus par paquets de charges arrivant sur le registre de lecture. Ces paquets de charges lus sont appelés « samples ». Ces derniers sont définis par les informations lues au niveau d'un pixel (dans ce cas, on parle de "un-binned sample") ou de plusieurs pixels physiques du CCD (on parle de "binned sample"). Un sample n'est donc pas une entité physique sur un CCD et il est propre à chaque observation. Par exemple, le sample 0 correspond au premier pixel lu d'une observation, et ce dernier est différent d'une observation à une autre. La définition du "sample" est décrite plus précisément dans la note technique ESA Bastian (2004).

Les CCDs du RVS auront trois modes d'échantillonnage (c.f. figure 1.13) :

- Les spectres des étoiles les plus brillantes, de magnitudes  $5 < G_{\text{RVS}}^7 < 7$ , seront lus pixel par pixel. Chaque spectre sera donc composé de  $1104 \times 10$  samples (chaque sample sera défini par les charges contenues dans un pixel du CCD). L'échantillonnage spectral sera donc égal à  $0.26 \text{ \AA/sample}$ . Ce type d'étoiles sera utilisé pour calibrer l'instrument spectroscopique.
- Les spectres d'étoiles de magnitudes  $7 < G_{\text{RVS}} < 10$  seront lus par sample de 10 pixels en AC et de 1 pixel en AL. Chaque spectre sera donc composé de  $1104 \times 1$  samples. Comme pour le mode précédent, l'échantillonnage spectral sera égal à  $0.26 \text{ \AA/sample}$ .
- Les spectres d'étoiles faibles, de magnitudes  $10 < G_{\text{RVS}} < 17$ , seront lus par sample de 10 pixels en AC et de 3 pixel en AL. Chaque spectre sera donc composé de  $368 \times 1$  samples. L'échantillonnage spectral sera donc égal à  $0.78 \text{ \AA/sample}$ . L'avantage de ce mode est de réduire le bruit de lecture, néfaste pour les étoiles faibles, et, par la même occasion, d'alléger le budget téléométrique.

L'instrument fournira les vitesses radiales de 100 à 250  $10^6$  étoiles jusqu'à la magnitude  $V = 17.5$ , avec une précision de 30 km/s pour les plus faibles et jusqu'à moins de 1 km/s pour les plus brillantes (c.f. table 1.4). Les principales performances du RVS sont résumées dans la table 1.5.

SpType	8.5	9.0	9.5	10	10.5	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16	16.5	17	17.5
B0V	1.2	1.6	2	2.7	3.8	6.8	9.7	14.5	24.8	n	n	n	n	n	n	n	n	n
B5V	<1	1.1	1.4	1.9	2.5	5.1	6.9	10	15.3	24.1	n	n	n	n	n	n	n	n
A0V	<1	<1	<1	<1	<1	1.3	1.8	2.6	3.9	5.7	8.6	14.6	32.5	n	n	n	n	n
A5V	<1	<1	<1	<1	<1	<1	<1	1.3	2	4.2	6.9	11.1	20.1	n	n	n	n	n
F0V	<1	<1	<1	<1	<1	<1	<1	1	1.5	2.1	3.2	5.3	7.8	12.7	23.4	n	n	n
G0V	<1	<1	<1	<1	<1	<1	<1	<1	<1	1.4	2.1	3	4.8	7.9	12.4	19.6	n	n
G5V	<1	<1	<1	<1	<1	<1	<1	<1	<1	1.2	1.9	2.8	4.4	6.3	10.1	17.6	n	n
K0V	<1	<1	<1	<1	<1	<1	<1	<1	<1	1.1	1.4	2.1	3.3	5.1	8.1	12.6	24.9	n
K4V	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	1.1	1.6	2.7	3.6	5.2	8.4	14.5	30
K1III	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	1.2	1.8	2.7	4.2	6.8	10.3	18

Tab. 1.5 – Précision des vitesses radiales  $\sigma_{v_r[\text{km/s}]}$  en fonction de la magnitude  $V$  et du type spectral. Ces performances sont obtenues à partir de la simulation de 2000 spectres RVS. Pour chacun de ces spectres simulés, une vitesse radiale est calculée par corrélation croisée et est comparée à sa valeur théorique (fournie par l'étape de simulation). Les résultats présentés dans ce tableau sont issus de la note technique Sartoretti et al. (2007).

La table 1.6 résume les principales caractéristiques de l'instrument spectroscopique de Gaia.

<sup>7</sup>La magnitude  $G_{\text{RVS}}$  signifie magnitude dans la bande RVS de Gaia.

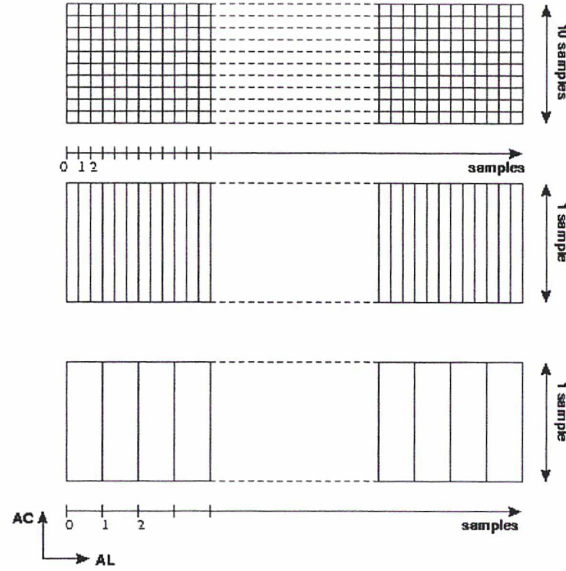


FIG. 1.13 – Représentation des modes d'échantillonnage des CCDs du RVS : mode échantillonnage des étoiles brillantes (en haut) : 1 sample correspond à 1 pixel physique ; échantillonnage intermédiaire (au milieu) : 1 sample correspond à une colonne de 10 pixels ; échantillonnage des étoiles faibles (en bas) : 1 sample correspond à  $3 \times 10$  pixels.

Caractéristiques du RVS	
Domaine spectral	[847 – 878] nm
Résolution	11500
Nombre de CCDs	12
Nombre de pixels par CCD	4500 (AL) $\times$ 1966 (AC)
Taille du pixel AL	10 $\mu$ m (soit 0.059 arcsec)
Taille du pixel AC	30 $\mu$ m (soit 0.177 arcsec)
Dispersion par pixel	0.026 nm (en mode d'échantillonnage des brillantes)
Nombre de transits par CCD	22.4 pour G = 10
Temps de pose accumulé par étoile	296.8 s pour G = 10

TAB. 1.6 – Principales caractéristiques techniques du Radial Velocity Spectrometer de Gaia.

### 1.2.3 Présentation des spectres du Radial Velocity Spectrometer

Le choix du domaine spectral du RVS prend en compte un certain nombre de considérations (Munari et Castelli (2000)). Dans les étoiles F, G et K, les trois raies du calcium ionisé de la bande RVS permettent des mesures de vitesses radiales dans les spectres de faible rapport signal à bruit. Ces trois raies permettent également de mesurer des vitesses radiales dans des étoiles déficientes en métaux. Elles sont sensibles aux paramètres atmosphériques de l'étoile (e.g.  $T_{eff}$ ). Pour les étoiles plus chaudes ou précoces, la bande spectrale du RVS est dominée par les raies

d'hydrogène de la série Paschen. La bande RVS contient aussi la bande diffuse interstellaire (DIB) située à 862 nm. Cette DIB est un très bon traceur du rougissement interstellaire : on peut utiliser la spectroscopie du RVS couplée aux données photométriques de Gaia pour dresser une carte tri-dimensionnelle de l'extinction interstellaire dans notre Galaxie. De plus, l'extinction dans la bande RVS est deux fois moins forte que dans la bande V. Cela permet une "exploration" du disque galactique sur de plus grandes distances qu'en bande V. Le domaine spectral du RVS n'est pas sujet à l'absorption tellurique, permettant ainsi de répéter ou compléter si besoin certaines observations au sol avec d'autres instruments.

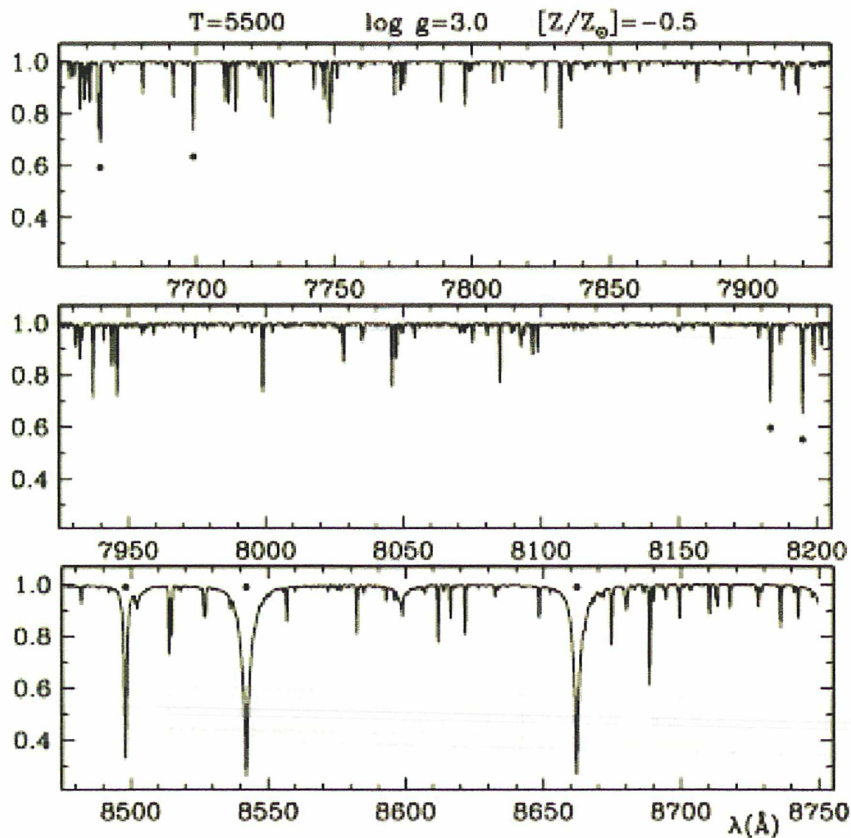


FIG. 1.14 – Spectre à haute résolution d'une géante G du disque, montrant trois régions spectrales étudiées lors de la définition du domaine spectral du RVS : région des doublets K I (en haut) et Na I (au milieu), et la région du triplet de Ca II (en bas). Les points noirs indiquent la position des doublets et triplets (Munari et Castelli (2000))

La figure 1.15 montre des exemples de spectres en bande RVS : une étoile déficiente en métaux K0 III et cinq étoiles de métallicité solaire.

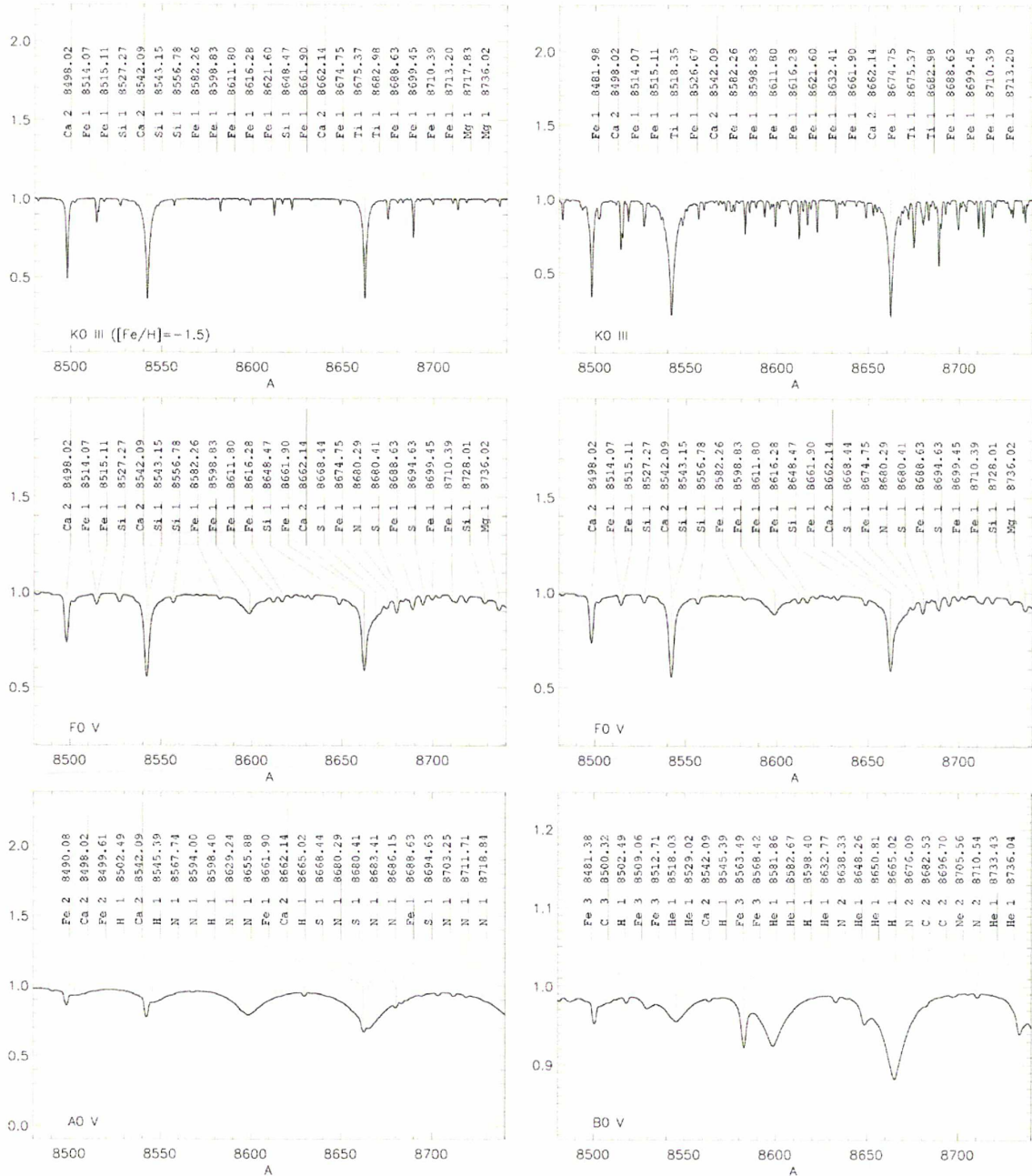


FIG. 1.15 – Exemples de spectres dans la bande RVS : une étoile K0 III pauvre en métaux (en haut à gauche) et une étoile K0 III (en haut à droite), une étoile G0 V (au milieu à gauche), une étoile F0 V (au milieu à droite), une étoile A0 V (en bas à gauche) et une étoile B0 V (en bas à droite) de métallicité solaire. Katz (2005)





# Chapitre 2

## Les calibrations spectroscopiques

### Sommaire

---

<b>2.1</b>	<b>L'utilité de la calibration</b>	<b>23</b>
<b>2.2</b>	<b>Les processus de calibration spectroscopique au sol</b>	<b>24</b>
2.2.1	Traitement des données brutes	25
2.2.2	Étalonnage des données prétraitées	25
<b>2.3</b>	<b>La calibration en longueur d'onde</b>	<b>26</b>
2.3.1	Approche classique de la calibration en longueur d'onde au sol	26
2.3.2	Différentes idées pour calibrer la loi de dispersion spectrale du Radial Velocity Spectrometer	27
<b>2.4</b>	<b>La calibration en longueur d'onde du Radial Velocity Spectrometer : la Spectroscopic Global Iterative Solution</b>	<b>28</b>
2.4.1	La philosophie générale de la Spectroscopic Global Iterative Solution	29
2.4.2	Aperçu de la Spectroscopic Global Iterative Solution au sens des calibrations en longueur d'onde	30
2.4.3	Concepts mathématiques	31

---

Le satellite Gaia observera en continu la sphère céleste durant les 5 ans de mission. On ne pourra donc pas interrompre les observations pour des besoins de calibration ou de contrôle du bon comportement de chaque instrument. Au vu des précisions souhaitées (c.f. table 1.4), les calibrations s'annoncent comme une tâche difficile du consortium de traitements de données de la mission.

### 2.1 L'utilité de la calibration

Les opérations de calibration sont partie intégrante de tout processus d'observation, que ce soit au sol ou dans l'espace. Le principal objectif des calibrations est de convertir les unités du détecteur (par exemple, des pixels) en unités physiques (par exemple, des longueurs d'onde) et de corriger des effets instrumentaux (par exemple, les biais du CCD). Pour être exploitables, les données collectées lors des observations doivent être associées à des échelles de mesure : par

exemple, un spectre brut va être associé à des flux en fonction de longueurs d'onde. Le processus de calibration va donc consister à définir ou redéfinir une fonction reliant une quantité physique à une unité du CCD. Cette fonction est propre à la configuration de l'instrument à un instant  $t$ .

Les calibrations permettent de mesurer les caractéristiques de l'instrument, et ainsi, de corriger ses défauts. Les défauts d'un instrument peuvent, entre autre, provenir d'un vice de construction (par exemple, il est fréquent d'avoir quelques pixels morts sur les capteurs CCD) ou de l'évolution temporelle des caractéristiques de l'instrument (par exemple, des déformations de miroir dues à des écarts de température).

Au sol, le processus d'observation est précédé d'une phase de calibration de l'instrument. Cette dernière doit être répétée assez régulièrement afin de maîtriser l'évolution de l'instrument au cours du temps.

Un processus de calibration mal effectué a des conséquences sur les mesures scientifiques. Par exemple, la figure 2.1 illustre plusieurs cas de mauvaises calibrations en longueur d'onde dans la bande RVS : un décalage de l'échelle en longueur d'onde, une compression et un étirement de l'échelle en longueur d'onde. Ces types de mauvaises calibrations vont être interprétés comme un décalage des raies dans le spectre stellaire faussant ainsi les mesures de vitesses radiales.

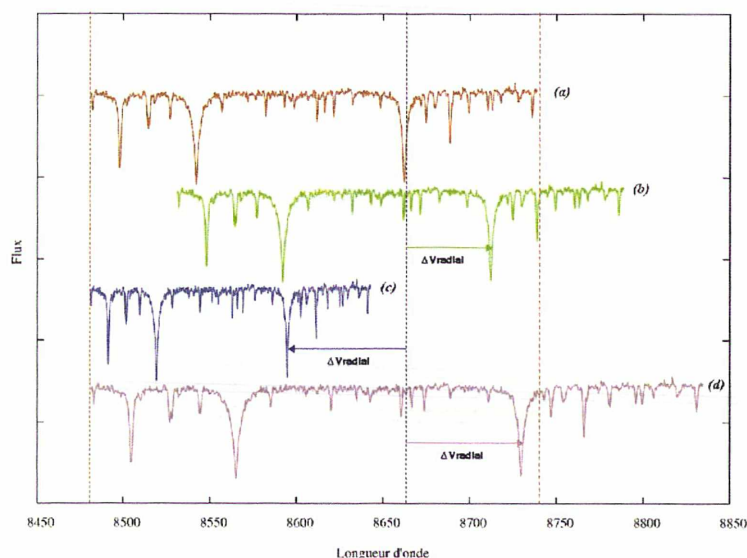


FIG. 2.1 – Calibration en longueur d'onde dans la bande RVS : (a) spectre correctement calibré ; (b) spectre calibré avec un offset, i.e. perturbation d'ordre 0 ; (c) et (d) spectres calibrés avec de mauvais facteurs d'échantillonnage, i.e. perturbation d'ordre 1 à l'origine d'une compression (cas (c)) ou d'une dilatation (cas (d)) de l'échelle en longueur d'onde.

## 2.2 Les processus de calibration spectroscopique au sol

Les processus de calibration spectroscopique au sol peuvent être décomposés en deux parties :

- Une phase de correction des observations : les observations sont généralement soumises à des perturbations venant de l'instrument en général. Ces perturbations doivent être corrigées afin de ne pas les confondre avec le signal recherché.
  - Une phase d'étalonnage : les observations d'une source étalon sont comparées aux données théoriques de cet étalon afin de passer des unités du détecteur à des unités physiques.
- Dans ces deux phases des observations dédiées sont nécessaires.

### 2.2.1 Traitement des données brutes

Le signal brut venant d'un instrument est une superposition d'événements : au signal de l'objet observé s'ajoutent des signaux parasites, du bruit, etc. Un certain nombre de processus peuvent être mis en place avant la phase d'observation afin que le signal collecté soit le plus proche possible du signal propre de l'objet observé. On distingue trois principales actions dans cette phase de traitement :

- Soustraction du courant d'obscurité ;
- Soustraction du signal d'offset ;
- Division par le flat-field de l'instrument.

Le signal d'obscurité, le "dark", est un signal thermique au sein du CCD. Il ne dépend pas de la quantité de lumière incidente, mais de la température du capteur et du temps de pose. Ainsi, il est nécessaire de faire une pose longue (c'est-à-dire, plus longue que la pose scientifique) lorsqu'aucune source n'éclaire le détecteur afin d'enregistrer ce courant. Dans les données observées, le dark est présent sous forme de biais. Pour le supprimer, il suffit donc de le soustraire à l'observation brute.

Le signal d'offset prend forme dans l'électronique intégrée au CCD dont le rôle est d'amplifier le signal électrique. Il produit un biais d'un pixel à un autre du CCD. Ce signal peut être mis en valeur en faisant des poses nulles. Comme le dark, ce signal parasite peut être soustrait à l'observation brute.

Les réponses de l'instrument et du capteur CCD ne sont pas uniformes. Ces non-uniformités peuvent être mises en évidence en observant une plage de lumière uniforme. On appelle l'image résultante le "flat field" ou Plage de Lumière Uniforme (PLU). Cette possible non-uniformité se corrige en divisant le spectre observé prétraité par la PLU (le signal thermique et le signal d'offset étant préalablement soustrait de l'observation).

Les impacts des différents bruits peuvent être "réduits" en augmentant le temps de pose ou en moyennant les observations d'une même source. Par exemple, le bruit de lecture est un bruit dû à l'électronique du CCD. Il intervient une fois lors de chaque lecture d'un paquet de charges de façon indépendante. Ainsi, en moyennant  $N$  observations d'une même source, le bruit de lecture est réduit d'un facteur  $\frac{1}{\sqrt{N}}$ .

### 2.2.2 Étalonnage des données prétraitées

Dans le cadre de la spectroscopie, deux types d'étalonnage sont à mettre en place pour passer des unités du détecteur à des unités physiques :

- L'étalonnage des flux dans le spectre : obtention de la vraie distribution spectrale de l'objet en utilisant la réponse instrumentale ;

- L'étalonnage des longueurs d'onde dans le spectre : passage des pixels aux longueurs d'onde.

Ces étalonnages sont réalisés en utilisant les observations d'une source de référence, dont les spectres sont bien connus. L'étalonnage consiste à comparer les spectres connus de ces références à leurs observations. Ces sources de référence peuvent être des lampes de calibration ou des étoiles standards spectrophotométriques.

## 2.3 La calibration en longueur d'onde

Beaucoup de quantités sont à calibrer sur l'instrument spectroscopique de Gaia. Mes recherches se sont concentrées sur le problème de la calibration en longueur d'onde des données spectroscopiques du RVS.

### 2.3.1 Approche classique de la calibration en longueur d'onde au sol

Les calibrations en longueur d'onde des instruments spectroscopiques au sol se font généralement avec des lampes de calibration et des poses dédiées. Les observations issues de ces lampes (i.e. flux en fonction de pixels du capteur) sont comparées aux spectres théoriques de ces lampes (flux en fonction de longueur d'onde, e.g. figure 2.2).

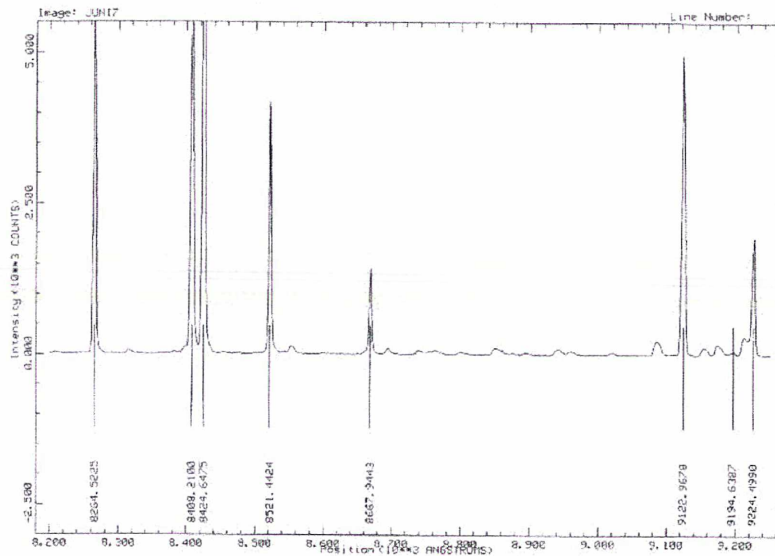


FIG. 2.2 – Exemple de spectre théorique d'une lampe de calibration He/Ar. (ESO)

La comparaison entre le spectre théorique et le spectre observé s'effectue au niveau de raies spectrales bien distinctes du continuum. On appelle ces raies des « raies de référence ». Dans un premier temps, les raies de référence vont être identifiées dans le spectre théorique de la lampe : les positions en longueur d'onde  $\lambda(ref)$  des raies de référence vont être lues à partir d'une carte spectrale théorique. Ces mêmes raies de référence sont ensuite "recherchées" dans le spectre observé de la lampe : les positions en pixels  $pix_{obs}(ref)$  de ces raies de référence sont

estimées dans le spectre observé. Une loi de calibration  $F(\text{pix})$  peut donc être calculée à partir des couples  $(\text{pix}_{\text{obs}}(\text{ref}), \lambda(\text{ref}))$  des raies de référence, par exemple, en utilisant une méthode numérique de résolution linéaire. À titre d'exemple, on peut procéder en utilisant la méthode des moindres carrés pour cette résolution. Un critère de moindres carrés  $S$  est ainsi établi sous la forme suivante :

$$S = \sum_{\text{ref}} [\lambda(\text{ref}) - F(\text{pix}_{\text{obs}}(\text{ref}))]^2 \quad (2.1)$$

où  $S$  est le critère des moindres carrés à minimiser,  $\text{ref}$  correspond à une raie de référence donnée,  $F$  la fonction de calibration à contraindre. La fonction  $F$  peut être représentée sous différentes formes polynomiales (e.g. Tchebycheff). Ici, nous nous limiterons à une forme polynomiale du type  $C_0 + C_1 \times X + \dots + C_n \times X^n$ , où les  $C_n$  sont les coefficients à calculer.

La calibration équivaut donc à minimiser le critère de l'équation 2.1 : i.e. une bonne estimation de la fonction  $F$  est atteinte quand toutes les dérivées partielles du critère  $S$  par rapport aux coefficients  $C_n$  sont nulles.

Une fois estimée, la loi de calibration  $F$  est ainsi appliquée sur toutes les données spectroscopiques de l'instrument. Ce processus doit être répété à plusieurs instants dans le temps afin de contrôler l'évolution temporelle des propriétés instrumentales (liées aux conditions de température, de pression,...etc.). De plus, il nécessite de faire des observations dédiées aux calibrations (i.e. observation de la lampe de calibration). Les observations du ciel doivent donc être arrêtées aux dépends des calibrations, chose qui n'est pas possible sur le satellite Gaia, constamment en mode d'observation de la sphère céleste.

### 2.3.2 Différentes idées pour calibrer la loi de dispersion spectrale du Radial Velocity Spectrometer

Dans le cadre du RVS, le processus de calibration doit être redéfini. En effet, le satellite Gaia observera en continu la sphère céleste durant les 5 ans de mission. Par conséquent, on ne pourra pas interrompre les observations pour effectuer des calibrations.

Aucun dispositif de type lampe de calibration ne peut être utilisé pour la mission Gaia. Plusieurs solutions instrumentales ont été proposées pour les calibrations en longueur d'onde. L'une d'entre elles, proposée par Desidera et Munari (2003), consiste à placer une cellule d'absorption contenant un gaz dans le chemin optique du RVS. La figure 2.3 illustre l'utilisation d'une cellule à Néon pour la calibration en longueur d'onde dans la bande RVS. La cellule a des raies d'absorption connues et bien réparties sur la bande RVS. En plaçant cette cellule dans le champ de vue du RVS, le spectre d'absorption de la cellule va se superposer au spectre de chaque étoile observée par le RVS. En repérant les raies d'absorption du néon dans le spectre observé, on peut mesurer l'échelle en longueur d'onde du spectre.

L'inconvénient de ce type de solution est de rajouter un élément sur le chemin optique du RVS (perte de photons) et aussi, de choisir un gaz approprié aux conditions du RVS :

- il doit rester sous sa forme gazeuse à la température du RVS ;
- quelques raies profondes d'absorption (i.e. de 5 à 10 raies) doivent être visibles dans le domaine spectral du RVS ;

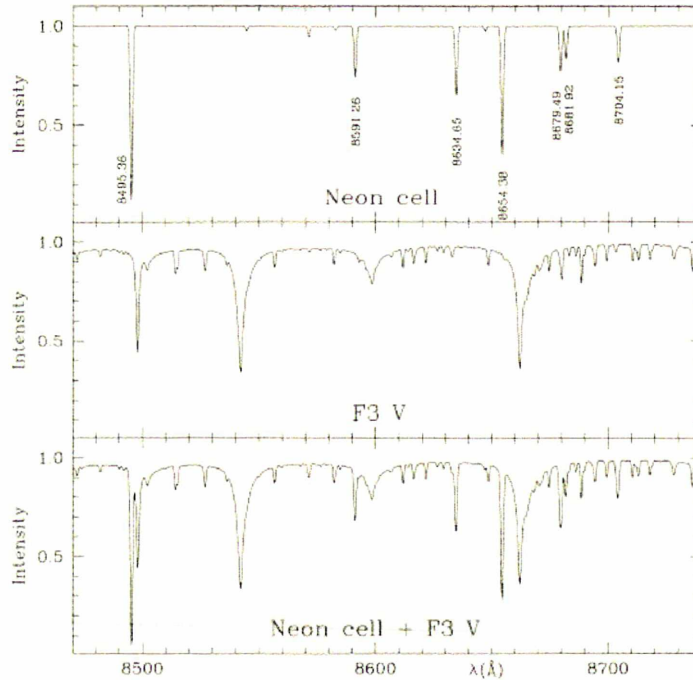


FIG. 2.3 – Simulation d’une utilisation d’une cellule à Néon pour la calibration en longueur d’onde du RVS : *en haut* le spectre d’absorption d’une cellule à Néon, *au centre* le spectre d’une étoile F3 V, et *en bas* le spectre de la même étoile observée à travers la cellule à Néon. (Desidera et Munari (2003))

- ses raies d’absorption doivent être distinctes des raies stellaires afin de ne pas détériorer le spectre des étoiles ;
- le gaz doit être stable et non corrosif afin de garantir le fonctionnement de la cellule à long terme.

Au vue de ces contraintes, les études sur les cellules d’absorption pour le RVS n’ont pas été poursuivies.

À défaut de solution instrumentale, le processus de calibration doit donc intervenir au niveau du traitement de données spectroscopiques au sol. Ce processus utilisera directement les spectres observés afin de calibrer l’instrument RVS.

## 2.4 La calibration en longueur d’onde du Radial Velocity Spectrometer : la Spectroscopic Global Iterative Solution

Le concept de la "Spectroscopic Global Iterative Solution" (SGIS) a été proposé pour la calibration en longueur d’onde des données spectroscopiques de Gaia. Il a été développée en s’inspirant de la solution de calibration des données astrométriques de Gaia, la méthode "Astrometric Global Iterative Solution" (c.f. Lindegren (2001)).

Cette section décrit le fonctionnement de la SGIS. Son implémentation ne sera pas discutée

ici, et fait l'objet du chapitre 3.

### 2.4.1 La philosophie générale de la Spectroscopic Global Iterative Solution

Sans dispositif de calibration à bord (e.g. lampe de calibration), il faut développer une méthode numérique de calibration au niveau du pipeline de traitement de données. Cette alternative se nomme la "Spectroscopic Global Iterative Solution" (SGIS).

Le principe de SGIS est d'utiliser les spectres d'étoiles observées avec le RVS afin de calibrer l'instrument (Hui-Bon-Hoa et Katz (2004)). Pour ce faire, les spectres d'étoiles brillantes ( $G_{RVS} < 11$ ) photométriquement stables (pas d'étoiles variables ou d'étoiles présentant des variations de flux importantes) et stables en vitesse radiale (pas de binaires ou d'étoiles présentant de variation importante au niveau des vitesses radiales mesurées) sont utilisés par la méthode SGIS de calibration. Les autres spectres ne sont pas considérés dans le processus de calibration. Avec un grand nombre d'observations de ces étoiles (i.e. 40 fois durant les 5 ans de mission), les variations apparentes des propriétés des étoiles stables (par exemple, biais dans les vitesses radiales) peuvent être reliées à une évolution temporelle des caractéristiques de l'instrument (par exemple, une dilatation de l'échelle en longueur d'onde du RVS).

Dans le cadre de la calibration en longueur d'onde, les raies stellaires provenant des propres observations du RVS vont être utilisées pour contraindre la loi de dispersion spectrale du RVS. Ces raies stellaires sont analogues aux raies de la lampe spectrale dans le processus de calibration en longueur d'onde au sol. La position en longueur d'onde de ces raies stellaires dépend de deux paramètres : la vitesse radiale de l'étoile observée et la loi de dispersion spectrale de l'instrument. De plus, ces vitesses radiales dépendent des calibrations en longueur d'onde, et inversement. Ainsi, la détermination des vitesses radiales et des lois de calibration en longueur d'onde du RVS doivent être effectuées successivement dans un processus itératif afin s'affranchir du lien entre ces deux paramètres :

1. une première loi de calibration va nous permettre de déterminer les vitesses radiales des étoiles observées ;
2. ces vitesses radiales vont être utilisées pour affiner la loi de calibration en longueur d'onde.

À l'itération suivante, la nouvelle loi de calibration (plus précise) va permettre d'améliorer l'estimation des vitesses radiales (étape 1 du processus) par rapport à la première itération. Le processus itératif est ainsi répété jusqu'à ce qu'un certain degré de convergence soit atteint. La méthode SGIS permet ainsi de déterminer les caractéristiques des sources observées (par exemple, la vitesse radiale) et de calibrer les caractéristiques du RVS de manière itérative.

Des étoiles standards en vitesse radiale seront utilisées afin de déterminer le point zéro des calibrations. Ces standards pourront aussi "accélérer" la convergence de la méthode SGIS.



## 2.4.2 Aperçu de la Spectroscopic Global Iterative Solution au sens des calibrations en longueur d'onde

La méthode itérative de calibration proposée est composée de cinq phases distinctes, illustrées par la figure 2.4 : i.e. deux phases non-itératives et 3 phases itératives.

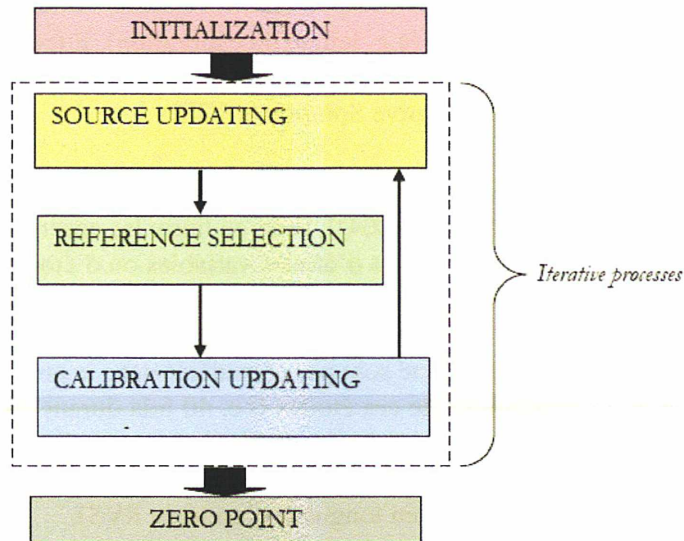


Fig. 2.4 – Aperçu de la méthode de calibration SGIS de l'instrument spectroscopique de Gaia. Guerrier et al (2006)

La méthode débute le processus de calibration avec une phase non-itérative d'initialisation : une première loi de calibration en longueur d'onde doit être choisie pour lancer le processus itératif. Cette loi de calibration initiale provient des calibrations au sol de l'instrument (fournie par l'industriel) ou de la phase de calibration en vol (appelé le « commissioning »).

La phase d'initialisation est suivie d'un processus itératif décrit par 3 phases :

1. **Source Updating** : cette phase a pour but d'estimer la vitesse radiale d'une étoile à partir de chacun des spectres de cette étoile. Dans un premier temps, les spectres de l'étoile sont calibrés en utilisant des lois de calibration en longueur d'onde (pour la première itération, ces lois proviennent de la phase d'initialisation). Puis, chacun d'entre eux est utilisé pour estimer la vitesse radiale de l'observation (c'est-à-dire mesure du décalage Doppler dans le spectre). Ainsi, la vitesse radiale de l'étoile est obtenue en moyennant les estimations de vitesses radiales des observations de cette étoile ;
2. **Reference Selection** : cette phase sélectionne les sources appropriées pour la phase de calibration. En effet, seule une partie des observations vont servir à calibrer l'instrument. Ces observations doivent provenir d'étoiles stables en vitesse radiale (par exemple, pas de systèmes multiples pour la calibration) et brillantes ( $G_{RVS} < 11$ ). Ces étoiles, sélectionnées lors de cette phase, sont qualifiées de "référence" pour les calibrations en longueur d'onde ;

3. **Calibration Updating** : dans cette phase, les observations d'étoiles dites de "référence" sont utilisées pour calibrer la loi de dispersion spectrale de l'instrument RVS.

À l'itération suivante, les dernières lois de calibration (i.e. lois calculées dans la phase 3) sont appliquées sur les spectres bruts provenant de l'instrument. Ces spectres calibrés sont ensuite utilisés pour affiner la détermination des vitesses radiales (i.e. nouvelle phase 1). Ces trois phases sont itérées jusqu'à ce qu'un certain degré de convergence soit atteint.

Durant le processus itératif, les vitesses radiales sont exprimées dans un référentiel relatif. Or, pour être scientifiquement exploitables, ces vitesses radiales devraient être exprimées dans un référentiel absolu (e.g. barycentre de notre système solaire). Des étoiles standards (c'est-à-dire, observées depuis le sol) sont utilisées afin de déterminer les transformations référentiel relatif vers référentiel absolu. C'est la dernière phase de SGIS, la correction du point zéro en vitesse radiale. Ce point zéro peut être défini par un décalage des vitesses radiales, et est appliqué sur toutes les observations du RVS.

### 2.4.3 Concepts mathématiques

La phase de "Calibration Updating", dernière phase du processus itératif de SGIS, réalise la calibration en longueur d'onde des données spectroscopiques de Gaia. Cette calibration utilise des concepts mathématiques similaires à ceux de la calibration des longueurs d'onde au sol (c.f. méthode décrite section 2.3.1).

Comme cela a été cité précédemment (c.f. section 2.3.1), les calibrations en longueur d'onde au sol utilisent des raies de référence venant d'une lampe de calibration. Dans le cadre de la SGIS, des raies spectrales de certaines étoiles vont jouer le rôle des raies de référence et seront identifiées dans les spectres collectés par le RVS. Le choix de ces raies stellaires de référence est induit par trois propriétés : (i) une raie de référence doit être bien connue dans le référentiel du laboratoire (i.e. position en longueur d'onde connue) ; (ii) une raie de référence doit être peu ou pas du tout blendée : en effet, avec la résolution du RVS ( $R = 11500$ ), une raie observée peut être composée de une ou plusieurs raies spectrales ; (iii) une raie de référence doit être suffisamment profonde pour ne pas la confondre avec le bruit.

En spectroscopie "classique", on forme une image statique du spectre d'une source sur le capteur CCD. Le spectre peut donc être représenté dans un repère physique : le spectre est défini par des nombres de photons par pixel du CCD. Chaque pixel est défini par ses coordonnées  $(x, y)$  sur le capteur. Dans ce cas, la calibration en longueur d'onde consiste à associer une longueur d'onde à un pixel du CCD.

Dans la spectroscopie RVS, la lecture des charges du capteur se fait en mode TDI. Ainsi, l'image du spectre d'une source défile sur le détecteur. Le spectre observé est représenté dans un repère non-physique : le spectre est défini par des nombres de photons par sample. Chaque sample est défini par le temps de lecture  $t$  du paquet de charges et une coordonnée AC sur le capteur CCD. Dans notre cas, la calibration en longueur d'onde consiste à associer une longueur d'onde à un sample. Les raies de référence vont donc être utilisées pour contraindre une loi de

calibration polynomiale  $F$  qui associe une longueur d'onde "centrale" à un sample  $s$  :

$$\lambda_s = F(s) \quad (2.2)$$

où  $\lambda_s$  est la longueur d'onde correspondant au sample  $s$ .

Dans l'espace, l'instrument RVS peut subir des détériorations dues à des variations thermiques, des variations mécaniques, des radiations, etc. Tous ces dégats ont pour conséquence une évolution des caractéristiques de l'instrument au cours du temps. Dans la méthode SGIS, on a supposé que cette évolution peut être découpée en plusieurs intervalles de temps durant lesquels les caractéristiques du RVS sont stables. On appelle ces intervalles de temps des « unités de calibration ». Ainsi, la loi de calibration  $F$  est supposée stable sur une de ces unités de calibration. Une fonction de calibration  $F$  doit être établie pour chaque unité de calibration en utilisant les observations faites durant cet intervalle de temps. La durée de cette unité de calibration fera l'objet d'une étude décrite dans le dernier chapitre de ce manuscrit.

Dans la suite de la thèse, on considérera deux hypothèses importantes pour la simplification du problème des calibrations en longueur d'onde dans le cadre de l'instrument RVS :

- Les transformations champ de vue - plan focal sont constantes durant une unité de calibration ;
- Les étoiles ont un mouvement constant sur le champ de vue durant une unité de calibration. De possibles variations du mouvement transverse (mouvement dans la direction AC) ne seront pas prises en compte.

La figure 2.5 montre qu'au temps  $t_1$  est observée l'étoile  $E$  de coordonnées  $(\eta_1, \zeta_1)$  et qu'au même moment le registre de lecture transfère un paquet de charges de cette étoile  $E$  pour la longueur d'onde  $\lambda_1$ . À un instant  $t_2$  ( $t_2 > t_1$ ), cette étoile  $E$  a bougé dans le champ de vue de l'instrument. Elle a des coordonnées  $(\eta_2, \zeta_2)$ . A ce même instant, le registre de lecture transfère le paquet de charge pour la longueur d'onde  $\lambda_2$  ( $\lambda_2 < \lambda_1$ ).

En prenant en compte les deux hypothèses (citées plus haut), si deux étoiles  $E$  et  $E'$  ont les mêmes coordonnées angulaires dans le champ de vue de l'instrument aux temps  $t$  et  $t'$ , alors les samples lus aux temps  $t$  et  $t'$  correspondront à la même longueur d'onde. Ainsi, la longueur d'onde correspondant à un sample peut être exprimée en fonction des coordonnées dans le champ de vue de l'instrument  $\eta$  et  $\zeta$ .

L'équation 2.2 peut s'exprimer sous la forme suivante :

$$\lambda_s = F(\eta_s, \zeta_s) \quad (2.3)$$

où  $\eta_s$  et  $\zeta_s$  sont les coordonnées angulaires d'une étoile dans le champ de vue de l'instrument au temps de lecture du sample  $s$ .

On suppose que la loi de calibration  $F$  est décrite par une loi polynomiale<sup>8</sup>. L'équation 2.3 peut s'écrire de la manière suivante :

$$\lambda_s = \sum_{m=0}^M \sum_{n=0}^N C_{mn} \eta_s^m \zeta_s^n \quad (2.4)$$

<sup>8</sup>Une forme polynomiale constitue une bonne approximation d'un large panel de fonctions mathématiques via un développement limité à un ordre approprié.

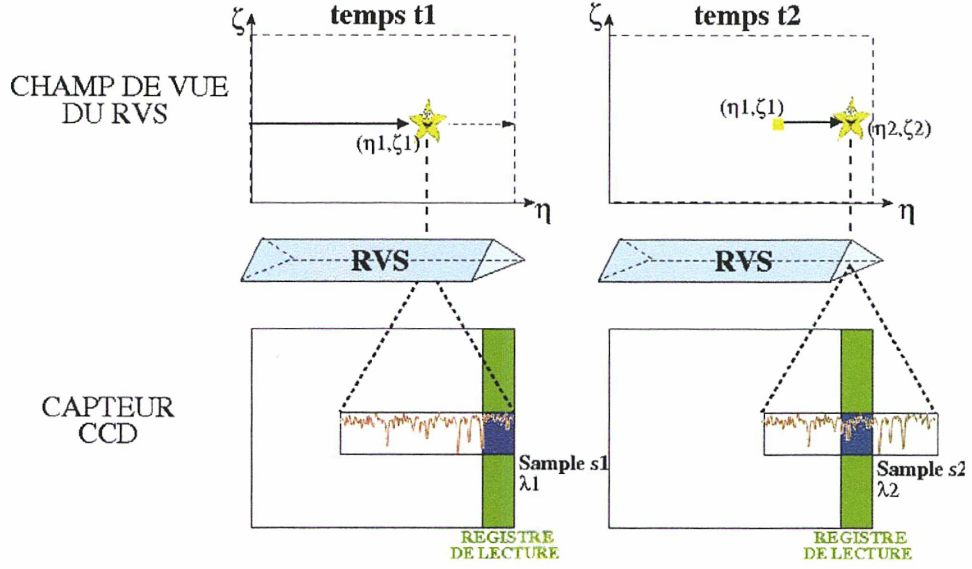


FIG. 2.5 – Illustration de la lecture d'un spectre d'une étoile sur le plan focal de Gaia. L'étoile E et son image sur le plan focal sont représentées à deux temps  $t1$  et  $t2$  différents.

où  $C_{mn}$  sont des coefficients de la forme polynomiale, nommés « coefficients de calibration ».

Le processus de calibration consiste à calculer un ensemble de coefficients de calibration  $C_{mn}$  pour chaque unité de calibration. Comme cela était décrit dans la section 2.3.1, le calcul de ces coefficients peut consister à minimiser un critère quadratique  $S_k$  propre à chaque unité de calibration  $k$  :

$$S_k = \sum_{spec=1}^{N_{spectres}} \sum_{l=1}^{N_{lines}} \left[ \lambda(spec, l) - \sum_{m=0}^M \sum_{n=0}^N C(k)_{mn} \eta(spec, l)^m \zeta(spec, l)^n \right]^2 \quad (2.5)$$

où  $spec$  représente le spectre d'une étoile observée durant l'unité de calibration  $k$ ,  $N_{spectres}$  le nombre de spectres observés durant l'unité de calibration  $k$ ,  $l$  une raie de référence du spectre observé  $spec$ ,  $N_{lines}$  le nombre de raies de référence dans le spectre  $spec$ ,  $\lambda(spec, l)$  une position en longueur d'onde du centroïde de la raie de référence  $l$  du spectre  $spec$ ,  $\eta(spec, l)$  et  $\zeta(spec, l)$  les coordonnées angulaires de l'étoile au moment où le centroïde de la raie de référence  $l$  passe au centre du registre de lecture du CCD, et  $C(k)_{mn}$  les coefficients de calibration de l'unité de calibration  $k$ .

La minimisation du critère  $S_k$  est atteinte lorsque le gradient de  $S_k$  s'annule :

$$\vec{grad} S_k = \vec{0} \Rightarrow \begin{cases} \frac{\partial S_k}{\partial C_{00}} = 0 \\ \frac{\partial S_k}{\partial C_{10}} = 0 \\ \frac{\partial S_k}{\partial C_{20}} = 0 \\ \frac{\partial S_k}{\partial C_{01}} = 0 \\ \vdots \\ \frac{\partial S_k}{\partial C_{mn}} = 0 \end{cases} \quad (2.6)$$

Cette minimisation passe par la résolution d'un système linéaire de type  $\mathbf{A} \times \mathbf{X} = \mathbf{B}$  où :

$$\mathbf{A} = \begin{bmatrix} \sum_{obs} 1 & \sum_{obs} \eta & \dots & \sum_{obs} \eta^m & \sum_{obs} \zeta & \dots & \sum_{obs} \eta^m \zeta & \dots & \sum_{obs} \zeta^n & \dots & \sum_{obs} \eta^m \zeta^n \\ \sum_{obs} \eta & \ddots & & & & & & & & & \vdots \\ \vdots & & \ddots & & & & & & & & \vdots \\ \sum_{obs} \eta^m & & & \sum_{obs} \eta^{2m} & & & & & & & \vdots \\ \sum_{obs} \zeta & & & & \ddots & & & & & & \vdots \\ \vdots & & & & & \ddots & & & & & \vdots \\ \sum_{obs} \eta^m \zeta & & & & & & \sum_{obs} \eta^{2m} \zeta^2 & & & & \vdots \\ \vdots & & & & & & & \ddots & & & \vdots \\ \sum_{obs} \zeta^n & & & & & & & & \sum_{obs} \zeta^{2n} & & \vdots \\ \vdots & & & & & & & & & \ddots & \vdots \\ \sum_{obs} \eta^m \zeta^n & \dots & & \dots & & & \dots & & & & \sum_{obs} \eta^{2m} \zeta^{2n} \end{bmatrix} \quad (2.7)$$

$$\mathbf{B} = \begin{bmatrix} \sum_{obs} \lambda \\ \sum_{obs} \lambda \eta \\ \vdots \\ \sum_{obs} \lambda \eta^m \\ \sum_{obs} \lambda \zeta \\ \vdots \\ \sum_{obs} \lambda \eta^m \zeta \\ \vdots \\ \sum_{obs} \lambda \zeta^n \\ \vdots \\ \sum_{obs} \lambda \eta^m \zeta^n \end{bmatrix} \quad \text{et} \quad \mathbf{X} = \begin{bmatrix} C_{00} \\ C_{10} \\ \vdots \\ C_{m0} \\ C_{01} \\ \vdots \\ C_{m1} \\ \vdots \\ C_{0n} \\ \vdots \\ C_{mn} \end{bmatrix} \quad (2.8)$$

La solution d'un tel système peut s'exprimer sous la forme  $\mathbf{X} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B}$ , sous réserve que la matrice  $(\mathbf{A}^T \mathbf{A})$  soit inversible. En théorie, il n'y a qu'une unique loi de dispersion spectrale et donc notre système est inversible. Cependant, en pratique, cette hypothèse d'inversibilité n'est pas fondée à cause de la présence de bruit et de problèmes numériques sur les données. Il faut donc vérifier l'inversibilité du système. La solution du système linéaire peut être estimée à partir de méthodes classiques de calcul du type pivot de Gauss ou méthode LU. L'implémentation de ces deux méthodes est détaillée dans Press et al. (2002).

# Chapitre 3

## Implémentation de la Spectroscopic Global Iterative Solution

### Sommaire

---

<b>3.1</b>	<b>Données d'entrée et de sortie</b> . . . . .	<b>36</b>
<b>3.2</b>	<b>Architecture et fonctionnement</b> . . . . .	<b>39</b>
3.2.1	Architecture . . . . .	39
3.2.2	Fonctionnement général du prototype . . . . .	40
3.2.3	Procédure d'initialisation . . . . .	46
3.2.4	Source Updating : Estimation des vitesses radiales . . . . .	47
3.2.5	Reference Selection : Sélection des étoiles de référence . . . . .	50
3.2.6	Calibration Updating : Calibration en longueur d'onde . . . . .	50
<b>3.3</b>	<b>Spectroscopic Global Iterative Solution, au niveau du consortium de traitement de données de Gaia</b> . . . . .	<b>52</b>

---

Afin de valider le concept de la "Spectroscopic Global Iterative Solution" (introduit dans le chapitre précédent) et d'évaluer ses performances, j'ai développé et implémenté un prototype du processus de calibration en longueur d'onde.

La validation du concept SGIS proposé se fait en plusieurs étapes. À chacune de ces étapes, une problématique est définie :

- La méthode SGIS converge-t-elle ? Cette problématique a un impact sur la conception générale de la méthode SGIS proposée. Elle valide ou non les différents concepts mathématiques et hypothèses pris en compte dans le développement de la méthode.
- Comment les performances des différents modules implémentés dans le prototype de SGIS influencent-elles la convergence du processus itératif ? Cette problématique permet d'identifier les points durs dans l'implémentation du prototype de SGIS.
- Comment se comporte le prototype sur une base de données réaliste ?
- etc

Pour y répondre, j'ai réalisé une implémentation du prototype appropriée. Chaque étape permet de mieux comprendre le fonctionnement et le comportement du processus itératif et, ainsi, de converger pas à pas vers une version finale du prototype de calibration en longueur d'onde.

Ce chapitre présente la structure et le fonctionnement général du prototype de calibration

en longueur d'onde. Je me focaliserai sur les aspects fonctionnel et numérique des implémentations de la méthode SGIS dans le cadre de la calibration en longueur d'onde des données spectroscopiques. Les spécificités des différentes versions du prototype sont présentées dans les chapitres 5 et 6.

### 3.1 Données d'entrée et de sortie

L'objectif principal du prototype SGIS est d'estimer itérativement les lois de dispersion spectrale du RVS. Comme cela était proposé dans le chapitre précédent (section 2.4.3), ces lois sont des polynômes associant une longueur d'onde centrale aux coordonnées  $(\eta, \zeta)$  d'une étoile dans le champ de vue du RVS. Ces polynômes sont définis par des coefficients  $C_{mn}$ . Le prototype doit ainsi fournir des estimations de ces coefficients polynomiaux (ou coefficients de calibration) à partir des spectres du RVS.

La figure 3.1 représente un diagramme de contexte du processus de calibration en longueur d'onde, c'est-à-dire un diagramme définissant le plus haut niveau des interfaces d'un système. Ce type de diagramme permet de délimiter le sujet d'étude, ici le prototype de calibration, de définir les flots de données interagissant avec le processus étudié. La table 3.1 complète le diagramme de contexte en donnant plus d'indications sur les données.

En entrée, le prototype a besoin 3 types de données différentes :

- **Des données de Source et d'Observation** : elles représentent les propriétés de la source (température effective, métallicité, logarithme de la gravité de surface, vitesse radiale, etc) et les spectres "extrait"<sup>9</sup> (nombre de photons par sample  $s$ , temps de lecture du premier sample et coordonnées  $(\eta, \zeta)$  de la source dans le champ de vue du RVS au temps de lecture du sample  $s$ , etc). Les données des sources sont nécessaires pour la sélection des données auxiliaires (c.f. section 3.2.4) et pour la détection et localisation des raies de référence dans la phase de Calibration Updating. Les données d'observation sont nécessaires à toutes les phases du prototype SGIS.
- **Des données Instrumentales** : elles représentent les coefficients de calibration. Ces données sont nécessaires pour calibrer les spectres extraits. Une fois calibrés, ces spectres sont utilisés dans l'estimation des vitesses radiales (phase de Source Updating).
- **Des données Auxiliaires** : elles sont constituées d'une bibliothèque de spectres synthétiques et d'une bibliothèque de listes de raies de référence. Ces données sont utilisées lors des phase de Source Updating et de Calibration Updating. Ces données ne sont pas modifiées au cours du processus itératif, elles servent de référence pour la calibration.

En sortie, le prototype fournit 2 types de données différentes :

---

<sup>9</sup>Un spectre "extrait" est un spectre mono-dimensionnel associant des flux à des samples lus et à des coordonnées angulaires  $(\eta, \zeta)$  de la source au temps de lecture du sample. Ce spectre n'est pas calibré en longueur d'onde.

- **Les données de Source et d'Observation** : à chaque itération, les vitesses radiales des sources sont estimées lors du Source Updating. Les données de Source et d'Observation sont ensuite mises à jour en fin de Source Updating.
- **Les données Instrumentales** : les coefficients de calibration sont calculés lors de la phase de Calibration Updating. Les données instrumentales sont ensuite mises à jour en fin de Calibration Updating.

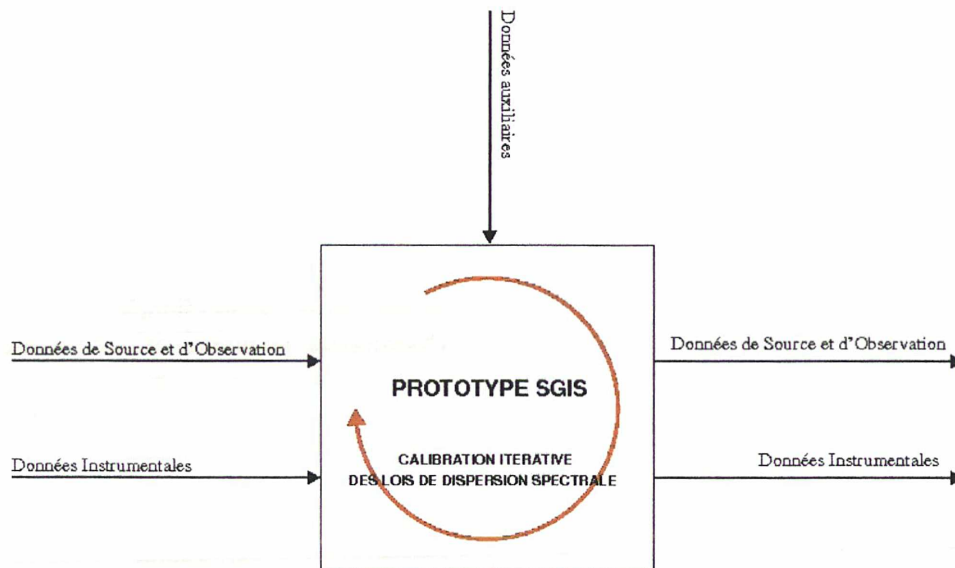


Fig. 3.1 – Diagramme de contexte du processus de calibration en longueur d'onde proposé. La boîte décrit l'action principale, ici le processus de calibration. Chaque flèche représente un flot de données. Les flèches arrivant sur le flanc de la boîte sont des données d'entrée, nécessaires au déroulement de l'action. La flèche arrivant sur le dessus de la boîte symbolise les données auxiliaires. La flèche sortant représente les données de sortie fournies par l'action. La flèche rouge interne à l'action illustre le processus itératif.

Les données d'entrée et de sortie du prototype sont stockées sur le disque dur dans 3 bases de données :

- Les données de source et d'observation sont contenues dans la base de données nommée "sourceObs".
- Les données instrumentales sont contenues dans la base de données nommée "instrument".
- Les données auxiliaires sont contenues dans la base de données nommée "auxiliary".

Durant une exécution du prototype, les bases de données "sourceObs" et "instrument" sont constamment mises à jour.



<b>Données d'entrée</b>	
Données de source et d'observation :	
Source :	$num_{Obs}$ Nombre d'observations de la source $T_{eff}$ Température effective $[Fe/H]$ Métallicité $\log g$ Logarithme de la gravité de surface $V_r$ Vitesse radiale REF Status de référence, i.e. si la source va être utilisée ou non dans les calibrations. STD Status de standard
Observation :	$flux$ Nombre de photons par sample $N_{samples}$ Nombre de samples lus pour l'observation $t_{Obs}$ Temps de lecture du premier sample de l'observation $\eta$ Coordonnée angulaire AL de l'étoile au temps de lecture de chaque sample $\zeta$ Coordonnée angulaire AC de l'étoile au temps de lecture de chaque sample $\lambda$ Longueur d'onde centrale par sample $V_r$ Vitesse radiale
Données Instrumentales :	
	$C_{mn}$ Coefficients de calibration
Données auxiliaires :	
	Spectres synthétiques et listes de raies de référence
<b>Données de sortie</b>	
Données Instrumentales :	
	$C_{mn}$ Coefficients de calibration
Données de source et d'observation :	
Source :	$V_r$ Vitesse radiale
Observation :	$V_r$ Vitesse radiale $\lambda$ Longueur d'onde centrale par sample

Tab. 3.1 – Table récapitulative des données d'entrée et de sortie du prototype proposé de calibration en longueur d'onde

## 3.2 Architecture et fonctionnement

Dans le chapitre 2, j'ai présenté la méthode SGIS appliquée à la calibration en longueur d'onde des données spectroscopiques. L'architecture proposée pour l'implémentation du prototype a été obtenue en découpant de manière logique chaque grande phase de SGIS en plusieurs modules. Dans cette section, je présente l'architecture générale du prototype de calibration. Dans un second temps, j'évoque le fonctionnement général du prototype afin de mettre en valeur les interactions entre chaque module important et les données introduites dans la section 3.1. Enfin, chaque module du prototype est décrit en détail.

### 3.2.1 Architecture

Dans le cadre du Data Processing Analysis Consortium, un unique langage d'implémentation a été choisi afin de supprimer les interfaces éventuelles entre différents langages de programmation (souvent source de problèmes). Tous les modules scientifiques de traitement de données, dont mon prototype de calibration, sont donc développés en Java. Le Java est un langage orienté objet où chaque concept et/ou objet physique est décrit par une "classe". Chaque classe est constituée d'un ensemble d'"attributs" et de "méthodes". Les attributs caractérisent le concept et/ou l'objet physique. Par exemple, la taille, le poids sont des attributs de l'objet homme. Les méthodes décrivent des actions réalisées par/sur l'objet. Par exemple, l'action de courir est une méthode de l'objet homme. Des classes liées entre elles peuvent être regroupées en "package". La traduction d'un problème physique en orienté objet se réalise d'une manière plus naturelle qu'en utilisant de la programmation classique. Le découpage logique du problème dépend de notre propre perception du problème.

Dans un souci d'uniformisation des programmes dans la chaîne de traitement de données de Gaia, chaque implémentation est soumise à des conventions Java propres à la mission Gaia. Par exemple, chaque classe est limitée en nombre de lignes, en nombre d'attributs. L'ensemble de ces conventions est décrit dans la note technique O'Mullane et al (2006).

La figure 3.2 représente l'architecture du prototype Java utilisée pour la validation de la méthode SGIS. Elle est découpée en 4 parties bien distinctes :

- Le prototype de SGIS est représenté en bleu. Il est constitué 4 packages principaux :
  - **SGIS** représente le package contenant le contrôleur SGIS. Ce contrôleur a pour rôle d'appeler successivement et itérativement les principales phases de SGIS, c'est-à-dire "Source Updating", "Reference Selection" et "Calibration Updating". Il se charge aussi de transmettre les données de la base de données vers les principaux modules et inversement. Ce package contient aussi 2 sous-packages :
    - **initialisation** des données de SGIS avant leur utilisation dans le processus itératif.
    - **test** contenant les diagnostics d'erreurs utilisés pour contrôler le bon fonctionnement du prototype à chaque itération. Ces diagnostics d'erreurs testent la qualité de la calibration en longueur d'onde appliquée sur chaque spectre "extrait" lors du module "apply calibration" et la qualité des estimations de vitesses radiales faites lors du module "Derive radial velocities".
  - **"Apply calibration"** permet d'appliquer les calibrations en longueur d'onde aux spectres extraits pour les transformer en spectres calibrés.

- "**Derive radial velocities**" permet d'estimer les vitesses radiales des sources par corrélation croisée.
- "**Wavelength calibration**" contient le processus de calibration en longueur d'onde de l'instrument spectroscopique. Il est constitué d'un sous-package, "**Select Reference**", décrivant la phase de sélection des étoiles de référence pour la calibration en longueur d'onde.
- Le modèle d'objets est représenté en jaune. Il décrit les données utilisées par le prototype de SGIS. Il gère les lectures-écritures sur le disque, définit les attributs caractéristiques de ces données, etc. Pour les besoins de la calibration en longueur d'onde, j'ai défini et développé mon propre modèle de données, contenu dans le package "**data**". Il décrit les données de source et d'observations, les données instrumentales et auxiliaires. Ce modèle de données se sert du package "**spectrum**" commun à la CU6 et auquel j'ai contribué. Ce package contient le modèle de données propres aux spectres extraits, spectres synthétiques et spectres calibrés.
- La "Gaia Toolbox" est en blanc. Elle représente, comme son nom l'indique, un ensemble d'outils numériques utilisable par les codes scientifiques. Elle est commune au DPAC.
- La "database" est représenté en rose. Elle symbolise l'ensemble des données utilisées par le prototype. Ses données sont constamment lues et réécrites sur le disque lors du processus itératif. Cette base de données contient les trois "sous-bases" de données présentées dans la section 3.1 : "sourceObs", "instrument" et "auxiliary".

Sur la figure, les flots de données entre les différentes packages et classes sont représentés sous forme de flèches noires. Tous les échanges de données se font par l'intermédiaire du contrôleur SGIS à l'exception de certains cas représentés par des flèches en pointillés. À terme, les accès à la base de données seront réservés au contrôleur SGIS. Ainsi, les échanges directs entre module et base de données (i.e. flèches en pointillés) seront supprimés (chaque module réalise une ou plusieurs opérations sur les données, mais en aucun cas, n'est responsable de lire ou écrire ces données).

Le développement de cette architecture a commencé il y a trois ans. À l'époque, il n'y avait pas de modèle de données commun à toute la CU6. Il m'a fallu m'adapter et donc développer un modèle propre aux besoins des calibrations en longueur d'onde. Depuis juillet 2007, un modèle de données commun à la CU6 a été développé et implémenté au GEPI par D. Katz et C. Delle Luche. Ainsi, dans mon agenda de développement, j'ai prévu une restructuration de l'architecture du prototype en janvier 2008. Cette restructuration intégrera ce modèle de données commun.

### 3.2.2 Fonctionnement général du prototype

Les différentes phases successives du prototype de SGIS, c'est-à-dire "Source Updating", "Reference Selection" et "Calibration Updating", sont représentées par les figures 3.3, 3.4, 3.5, 3.6 et 3.7. La première itération du processus de calibration a été divisée en cinq schémas

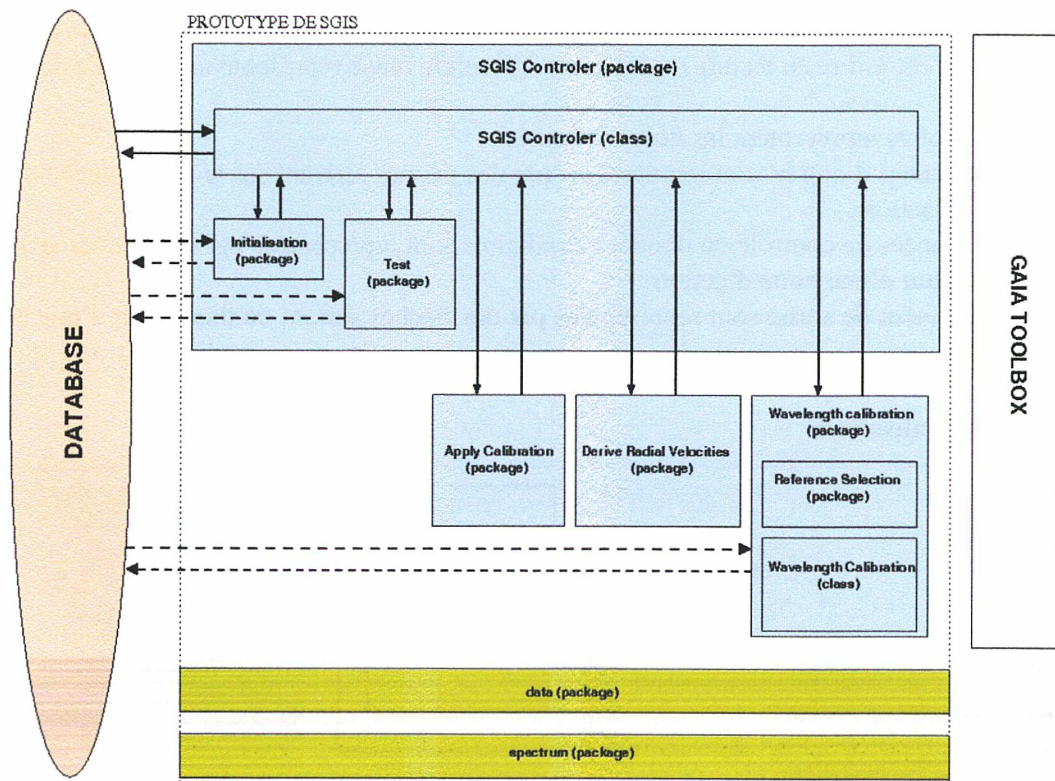


Fig. 3.2 – Architecture du prototype de SGIS. Les principaux packages et classes du prototype sont représentés par des boîtes bleues. Le modèle d'objets utilisé par le prototype est représenté par des boîtes jaunes. Les outils communs au prototype sont représentés par une boîte blanche. Les données d'entrée, de sortie et auxiliaires sont représentées par une ellipse rose. Chaque flèche symbolise des flots de données : les flèches pleines sont des échanges de données entre le contrôleur SGIS et la base de données, les flèches en pointillés sont des échanges directs de données entre des modules et la base de données. Les échanges symbolisés par des pointillés seront à terme supprimés afin que tous les échanges transitent par le contrôleur.

fonctionnels dans une séquence temporelle :

1. La figure 3.3 présente la phase d'initialisation ;
2. Les figures 3.4 et 3.5 présentent la phase d'estimation des vitesses radiales des sources ;
3. La figure 3.6 présente la phase de sélection des sources de référence ;
4. La figure 3.7 présente la phase de calibration en longueur d'onde.

Dans chacune de ces figures, les actions spécifiques (par exemple, l'initialisation) sont représentées par des boîtes rectangulaires, dont la couleur suit les conventions suivantes : une boîte rouge représente un processus en action, et une boîte blanche représente un processus inactif.

Les cercles représentent les différentes bases de données utilisées lors d'une exécution d'un prototype (c.f. section 3.1). Les données de calibration propres à l'instrument sont représentées par l' "inst. DB". Les données d'observation sont représentées par l' "Src. DB". Les données

auxiliaires sont représentées par l' "Aux. DB". Un cercle rose clair représente une base de données active (c'est-à-dire en lecture ou écriture). Un cercle blanc représente une base de données inactive.

Les flèches noires représentent les flots de données :

- Les données d'entrée sont représentées par des flèches arrivant sur le flanc gauche d'une boîte d'action ;
- Les données de contrôle ou données auxiliaires sont représentées par des flèches arrivant sur le haut d'une boîte d'action ;
- Les données de sortie sont représentées par des flèches sortant du flanc droit d'une boîte d'action.

### Phase 0 : initialisation

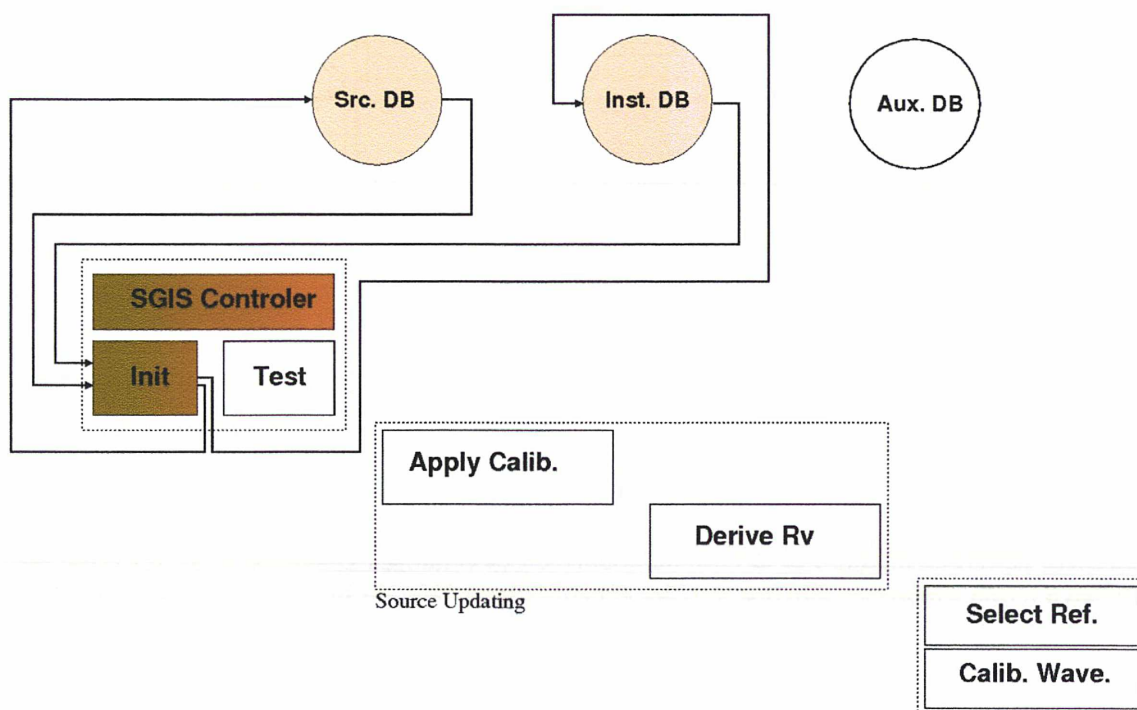


FIG. 3.3 – Phase d'initialisation de SGIS. Ne se produit qu'une fois par exécution du prototype de SGIS. "Init" représente le module d'initialisation du prototype SGIS, "Apply Calib." le module d'application de la calibration en longueur d'onde, "Derive Rv" le module de calcul des vitesses radiales, "Select Ref." le module de sélection des étoiles de référence pour la calibration et "Calib. Wave." le module de calibration en longueur d'onde.

La figure 3.3 représente la phase "zéro" du prototype, i.e. l'initialisation. Cette phase est exécutée une seule fois, avant le processus itératif. Elle initialise un premier jeu de coefficients de calibration et met à jour les données instrumentales ("instrument DB"). Cette phase initialise aussi les statuts standard des sources observées durant la mission simulée et met à jour les données d'observation ("SourceObs DB"). La vitesse radiale de chaque étoile standard est initialisée à partir de sa valeur estimée lors du suivi au sol de l'étoile.

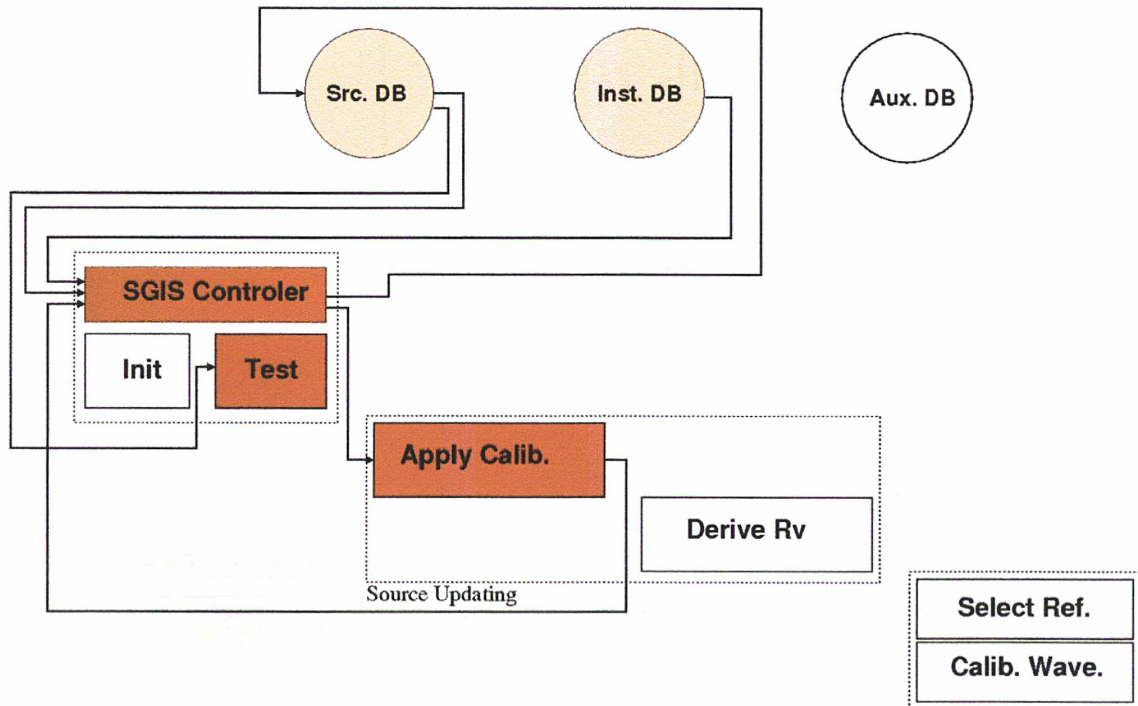
**Phase 1 - Iteration 1 : source updating**

FIG. 3.4 – Phase d'application des calibrations en longueurs d'onde aux spectres extraits.

Les figures 3.4 et 3.5 représentent la première phase du processus itératif, i.e. le "Source Updating". Cette phase estime la vitesse radiale de chaque étoile à chaque époque d'observation. Pour ce faire, elle va lire le spectre de chaque observation et les propriétés de chaque étoile observée dans la base de données d'observations. Chaque spectre observé va être calibré en longueur d'onde à partir des dernières données instrumentales de la base de données (c.f. 3.4). Après avoir appliqué les calibrations en longueur d'onde, une phase de test est exécutée. Cette phase de test compare l'échelle en longueur d'onde calculée avec l'échelle en longueur d'onde vraie. Elle permet ainsi de contrôler la qualité des calibrations en longueur d'onde appliquées sur chaque spectre à chaque itération. Une fois la calibration effectuée, la vitesse radiale de chaque source va être estimée en comparant chaque spectre à un spectre synthétique sélectionné dans la base de données auxiliaires (c.f. 3.5). À la fin de cette phase, les vitesses radiales de la base de données "SourceObs DB" vont être mises à jour. Une phase de test est exécutée après la mise à jour de la base de données. Cette phase compare l'estimation de la vitesse radiale à sa valeur vraie afin de contrôler la qualité de l'estimation à chaque itération. Ce test permet ainsi de suivre l'évolution des performances en vitesse radiale au cours des itérations.

**Phase 2 - Iteration 1 : reference selection**

La figure 3.6 illustre l'étape de sélection des sources de référence pour la calibration en longueur d'onde. Le critère de sélection de référence se base sur les propriétés de chaque étoile :

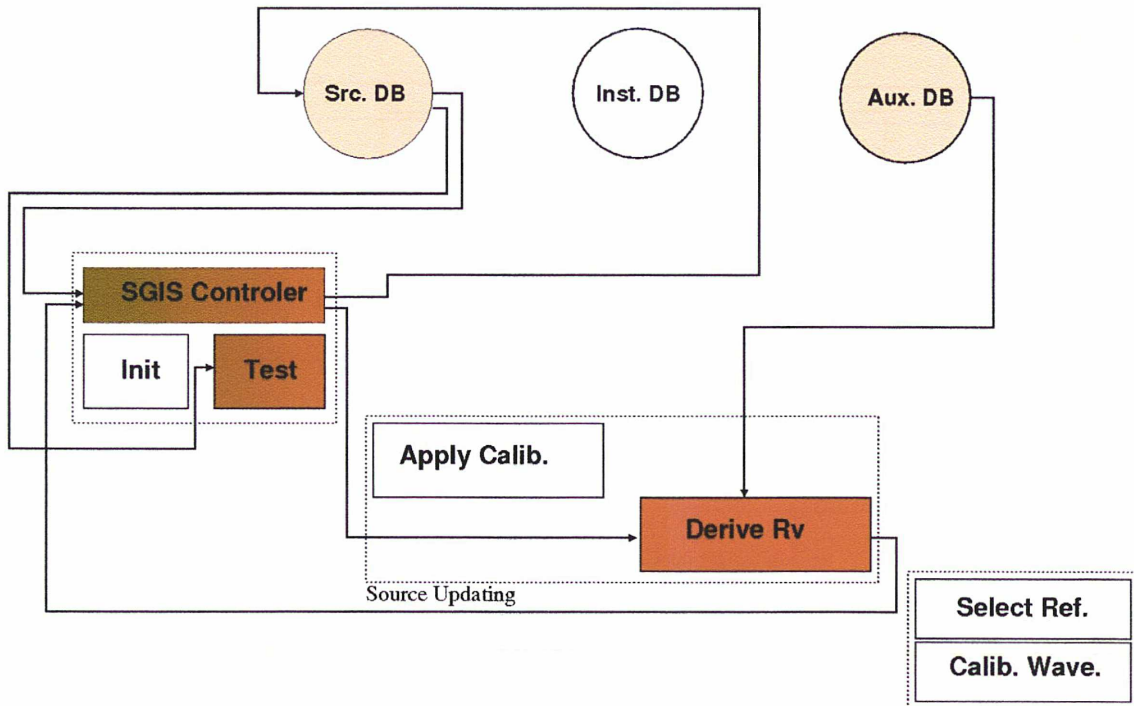


FIG. 3.5 – Phase d’estimation des vitesses radiales des spectres calibrés en longueur d’onde.

une étoile est qualifiée comme référence en fonction de sa température effective et de sa métallicité. Cette phase doit définir le statut de référence pour chaque étoile et le mettre à jour dans les données propres aux sources observées ("SourceObs DB").

### Phase 3 - Iteration 1 : calibration updating

La figure 3.7 représente la dernière phase du processus itératif, i.e. le "Calibration Updating". Cette phase utilise les données des étoiles qualifiées comme référence et une liste de raies de référence appropriée pour estimer le loi de calibration en longueur d’onde de chaque unité de calibration<sup>10</sup>. À la fin de cette phase, les données instrumentales sont mises à jour avec les nouvelles estimations des lois de calibration.

#### Iteration 2 à $n$

L’itération suivante débute par le "Source Updating" en appliquant les nouvelles lois de calibration en longueur d’onde sur les spectres extraits. Le processus de calibration exécute ainsi les phases 1, 2 et 3 de manière itérative jusqu’à ce que la convergence de la solution de calibration soit atteinte. Dans les versions actuelles du prototype de SGIS, le critère d’arrêt du processus itératif est basé sur un nombre d’itérations fixe. Par exemple, dans la plupart de mes tests, le processus itératif s’arrête après 20 itérations. Dans les prochaines versions, ce critère devrait être remplacé par un critère basé sur l’évolution des dispersions des vitesses

<sup>10</sup>L’unité de calibration a été définie dans la section 2.4.3 du chapitre précédent.

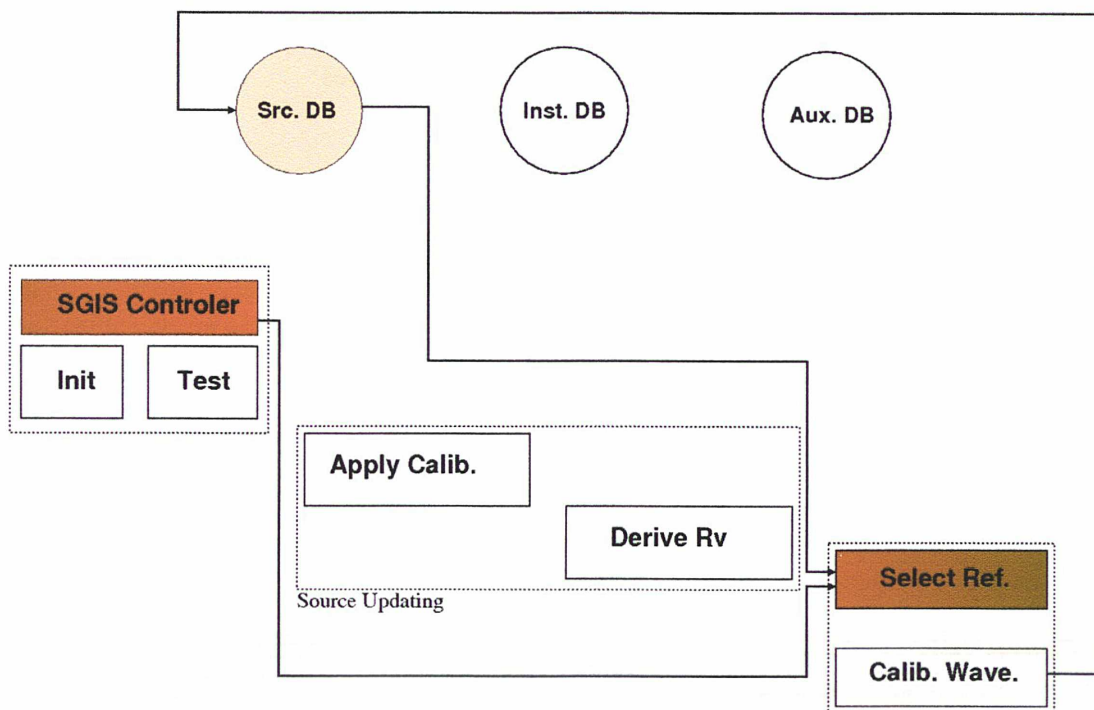


FIG. 3.6 – Phase de sélection des références.

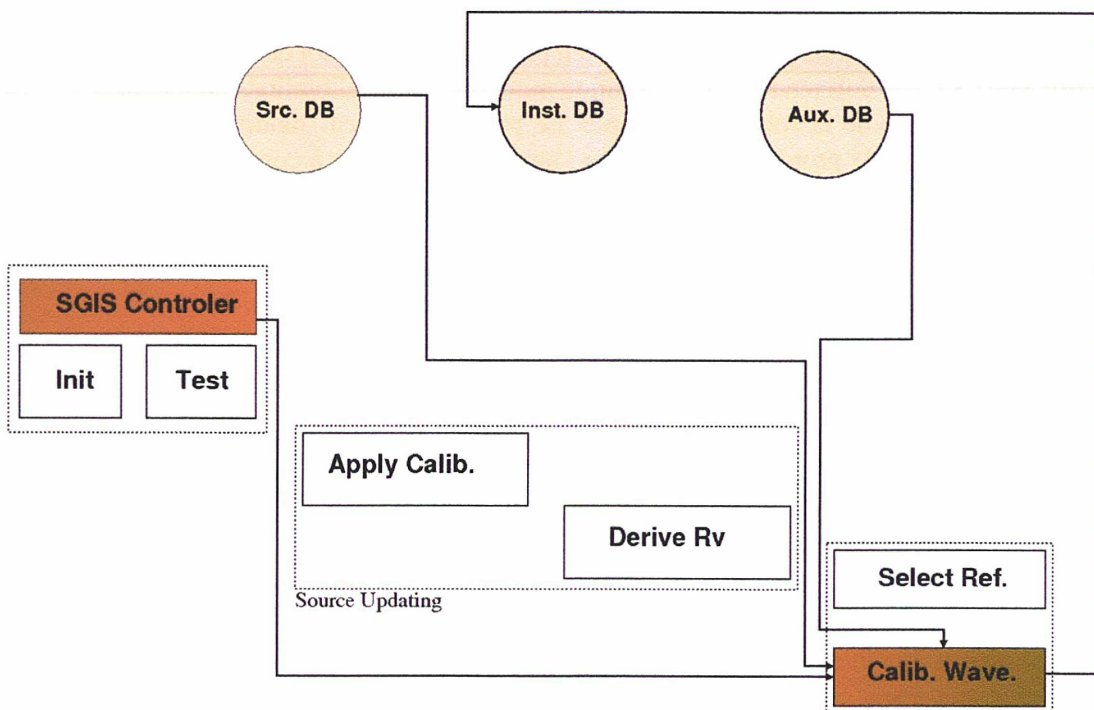


FIG. 3.7 – Phase de calibration en longueur d'onde du RVS.



radiales estimées pour une même étoile. En effet, le processus pourrait itérer jusqu'à ce que les dispersions des vitesses radiales estimées soient en dessous d'un seuil de convergence. Une prochaine étude pourrait consister à définir ce seuil de convergence.

### 3.2.3 Procédure d'initialisation

Tout processus itératif a besoin d'un point de départ. Dans le cadre de SGIS, le processus itératif débute par l'estimation des vitesses radiales des spectres calibrés du RVS. De premières lois de calibration en longueur d'onde doivent être établies avant ce calcul de vitesses radiales. Ainsi, l'initialisation du processus itératif correspond au choix des coefficients de calibration à appliquer aux spectres extraits du RVS.

Lors de la mission spatiale, les premiers coefficients de calibration en longueur d'onde vont être fournis par les calibrations au sol de l'instrument puis par la phase de commissioning du satellite.

Dans le cadre des tests des prototypes de SGIS sur une base de données synthétiques, ces coefficients proviennent du simulateur de spectre RVS. Tous les spectres simulés ont la même loi de dispersion spectrale représentée sous la forme  $\lambda = a + b\eta$ , où  $a$  et  $b$  sont des valeurs de coefficients vrais.

Pour l'initialisation, il y a deux possibilités : (i) le processus itératif du prototype part de la solution théorique des calibrations, i.e. on initialise les coefficients de calibration avec leurs vraies valeurs ; (ii) le prototype débute le processus itératif avec des lois de calibration fausses, non loin des vraies lois de calibration. Dans le premier cas, le prototype va mettre en avant sa stabilité, ses capacités à ne pas diverger : i.e. en partant de la vraie loi de calibration, la solution du processus itératif doit tendre vers la solution numérique des calibrations en longueur d'onde, qui doit être proche de la solution théorique. Le prototype est en "mode de non-divergence". Dans le second cas, le prototype va mettre en avant ses capacités à converger vers la solution numérique des calibrations (i.e. celle identifiée dans le mode de non-divergence du prototype). Le prototype est en "mode de convergence".

Une étude de convergence du prototype est dressée dans le chapitre 5. Cette étude discutera des différents choix d'initialisation des coefficients de calibration.

En plus de l'initialisation des coefficients de calibration, chaque étoile est qualifiée ou non de standard en vitesse radiale. À partir d'un nombre donné de standards, le module sélectionne de manière arbitraire les standards parmi les étoiles observées. L'importance d'un standard en vitesse radiale réside dans le fait que sa vitesse radiale est supposée parfaitement déterminée (à partir d'un suivi au sol de l'étoile). Dans le cadre de SGIS, un standard a une vitesse radiale constante tout au long du processus itératif. Elle n'intervient que dans les phases de "Sélection d'étoiles de Référence" et de "Calibration Updating" (c.f. figure 2.4). Pour initialiser les vitesses radiales de chaque standard, le module utilise la vraie valeur de la vitesse radiale de l'étoile (donnée par le simulateur de données) bruitée par une distribution gaussienne d'erreur aléatoire. Cette étape simule ainsi la mesure de vitesse radiale faite de l'étoile standard faite au sol. La valeur ainsi calculée est ensuite écrite dans la base de données "sourceObs". Cette initialisation de standards est nécessaire pour tester les propriétés de convergence en fonction

d'un pourcentage d'étoiles standard.

La mission simulée est découpée en unités de calibration, durant lesquelles les lois de dispersion spectrale du RVS sont supposées constantes. La durée de chaque unité de calibration doit être fixée avant chaque exécution du prototype. Comme le montre la figure 3.8, chaque observation de chaque source est associée à une unité de calibration en utilisant son temps d'observation. Ce temps d'observation ne varie pas durant le processus itératif. Les listes d'observations peuvent donc être bâties à l'avance pour chacune des unités de calibration. Cette étape a été intégrée dans la phase d'initialisation.

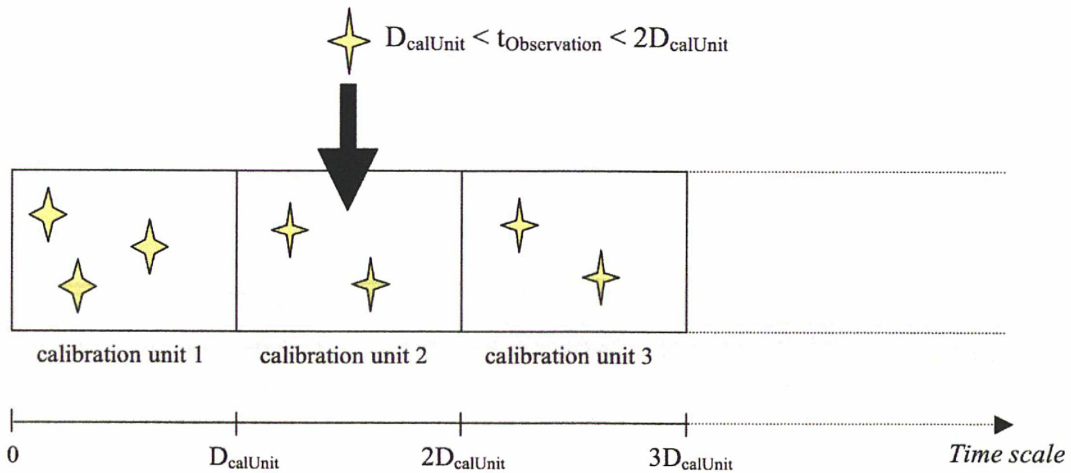


FIG. 3.8 – Représentation schématique de la création des listes d'observations pour les unités de calibration :  $D_{calUnit}$  correspond à la durée d'une unité de calibration (cette durée est la même pour chaque unité) et  $t_{observation}$  est le temps d'observation correspondant au temps de lecture d'un sample du spectre. Guerrier et al (2006).

### 3.2.4 Source Updating : Estimation des vitesses radiales

L'estimation des vitesses radiales constitue la première phase du processus itératif de calibration en longueur d'onde. L'objectif de cette phase est de fournir une estimation de la vitesse radiale de chaque étoile à chacune de ses époques d'observation.

La phase d'estimation des vitesses radiales est découpée en plusieurs étapes successives :

1. Les dernières calibrations en longueur d'onde sont appliquées au spectre extrait de l'étoile (i.e. pour la première itération, ces calibrations en longueur d'onde correspondent à celles choisies lors de la phase d'initialisation du prototype de SGIS) ;
2. Les paramètres atmosphériques de l'étoile observée sont utilisés afin de sélectionner un spectre synthétique à partir des paramètres atmosphériques de l'étoile observée ;
3. L'estimation des vitesses radiales du prototype de SGIS est réalisée à partir d'une méthode de corrélation croisée. Cette méthode consiste à corréler le spectre calibré en longueur

d'onde d'une étoile avec un spectre théorique de cette étoile décalé en vitesse radiale. Une fonction de corrélation entre le spectre calibré de l'étoile et son template est ainsi calculée pas à pas : a) le template est décalé en longueur d'onde en lui ajoutant une vitesse radiale ; b) le template est rééchantillonné en utilisant l'échelle en longueur d'onde du spectre calibré ; c) le coefficient de corrélation est calculé ;

4. La vitesse radiale du spectre calibré est estimée en calculant la position du maximum de la fonction de corrélation.

### Application des calibrations en longueur d'onde

Pour appliquer les calibrations en longueur d'onde sur un spectre extrait d'une source, il faut associer une longueur d'onde aux coordonnées angulaires ( $\eta, \zeta$ ) de la source dans le champ de vue du RVS au temps de lecture d'un sample (c.f. figure 2.5). Cette association est possible à partir des coefficients de calibration via la formule suivante

$$\lambda = \sum_{m=0}^M \sum_{n=0}^N C_{mn} \eta^m \zeta^n$$

où  $\lambda$  est la longueur d'onde, M et N sont les degrés maximum du polynôme de calibration et  $C_{mn}$  sont des coefficients de calibration.

Une fois calibré, chaque échelle en longueur d'onde est écrite sur le disque dur dans la base de données appropriée.

### Sélection du spectre théorique

L'objectif de la méthode de corrélation croisée est de mettre en phase les raies du spectre observé avec celles d'un spectre théorique. Cette mise en phase n'est possible que si la morphologie du spectre synthétique<sup>11</sup> est similaire à celle du spectre observé. Le spectre synthétique doit être convolué par la PSF du RVS pour que les raies synthétiques aient le même profil que celles observées dans les spectres. On ne parle plus de spectre synthétique, mais de "template". De plus, le template doit provenir d'une étoile de même type que celle observée par le RVS. Par conséquent, le template utilisé lors de l'estimation des vitesses radiales est sélectionné à partir de ses paramètres atmosphériques (c'est-à-dire, température effective, logarithme de gravité de surface et métallicité). Cette sélection est effectuée en calculant un degré de similarité entre les paramètres atmosphériques du template et ceux du spectre observé. Ce degré de similarité est calculé en utilisant la formule suivante :

$$\theta = \frac{|Teff_{ref} - Teff_{obs}|}{100.0} + \frac{|logg_{ref} - logg_{obs}|}{3.0} + \frac{|[Fe/H]_{ref} - [Fe/H]_{obs}|}{2.0} \quad (3.1)$$

Où  $\theta$  est le degré de similarité,  $Teff_{ref}$  et  $Teff_{obs}$  sont, respectivement, les températures effectives du template et de l'observation,  $logg_{ref}$  et  $logg_{obs}$  les logarithmes de gravité de surface du

---

<sup>11</sup>Un spectre synthétique est un spectre "parfait", c'est-à-dire un spectre sans effets instrumentaux et n'ayant pas de vitesse radiale.

template et de l'observation, et  $[Fe/H]_{ref}$  et  $[Fe/H]_{obs}$  les métallicités du template et de l'observation.

Dans la formule 3.1, chaque paramètre atmosphérique est pondéré afin de donner plus de poids à la température effective dans le choix du template. En effet, la température effective a une influence plus importante sur la morphologie du spectre que la métallicité ou que le logarithme de gravité de surface. Ces facteurs de pondération n'ont pas fait l'objet d'une étude approfondie.

Plus le degré de similarité est important, plus la morphologie du template est proche de celle du spectre observé. Le template est donc sélectionné en maximisant ce degré de similarité.

### Calcul de la fonction de corrélation

Le principe de la méthode numérique de corrélation croisée est de décaler pas à pas en vitesse radiale le template et de mesurer à chaque décalage la correspondance entre le spectre observé et le template décalé. Cette mesure de correspondance consiste à calculer un coefficient de corrélation entre ces deux spectres en utilisant la formule suivante :

$$\rho_{obs,ref}(Vr) = \frac{\sum_{s=1}^{N_{sample}} (F_{obs}(s) - \overline{F_{obs}}) (F_{ref}(s, Vr) - \overline{F_{ref}}(Vr))}{\sqrt{\sum_{s=1}^{N_{sample}} (F_{obs}(s) - \overline{F_{obs}})^2 \sum_{s=1}^{N_{sample}} (F_{ref}(s, Vr) - \overline{F_{ref}}(Vr))^2}} \quad (3.2)$$

Où  $\rho_{obs,ref}(Vr)$  est le coefficient de corrélation entre le spectre observé et le spectre synthétique décalé d'une vitesse radiale  $Vr$ ,  $N_{sample}$  est le nombre de samples dans les spectres observés et template,  $F_{obs}(s)$  est le flux lu au niveau du sample  $s$  dans le spectre observé,  $\overline{F_{obs}}$  correspond au flux moyen dans le spectre observé,  $F_{ref}(s, Vr)$  est le flux lu au niveau du sample  $s$  dans le template décalé d'une vitesse radiale  $Vr$ , et  $\overline{F_{ref}}(Vr)$  correspond au flux moyen dans le template décalé d'une vitesse radiale  $Vr$ .

La valeur de ce coefficient de corrélation en fonction de chaque décalage en vitesse radiale imposé sur le template est ainsi estimée. Le maximum de cette fonction  $\rho_{obs,ref}(Vr)$  correspond à la meilleure correspondance entre le template et le spectre observé. En localisant ce maximum, on a donc accès à la vitesse radiale de l'observation.

### Estimation du maximum de la fonction de corrélation

Afin d'avoir une précision sur l'estimation de la vitesse radiale inférieure à  $1 \text{ km.s}^{-1}$ , un ajustement polynomial est utilisé. Dans un premier temps, le maximum discret de la fonction  $\rho_{obs,ref}(Vr)$  est déterminé. Puis, on définit un domaine d'ajustement autour de ce maximum dans lequel on va ajuster un polynôme d'ordre 2 sur la fonction  $\rho_{obs,ref}(Vr)$ . La largeur du domaine est définie par un pourcentage (i.e. on a choisi arbitrairement environ 20%) de la largeur à mi-hauteur de la fonction de corrélation. L'ajustement est ensuite réalisé par moindres carrés en minimisant le critère  $S$  suivant :

$$S = \sum_{Vr} \left( \rho_{obs,ref}(Vr) - \sum_{n=0}^2 D_n Vr^n \right)^2$$

où  $Vr$  est le décalage en vitesse radiale imposé au spectre synthétique et  $D_n$  les coefficients de la forme polynomiale utilisée pour l'ajustement.

Une fois l'ajustement réalisé, une estimation de la vitesse radiale est obtenue en recherchant le maximum de la forme polynomiale calculée.

À la fin de cette étape, l'estimation de la vitesse radiale du spectre est mise à jour dans les données d'observation.

Pour chaque observation (spectre) d'une étoile, une vitesse radiale est estimée. La vitesse radiale de l'étoile est obtenue en moyennant les vitesses radiales de toutes les observations de l'étoile. Ainsi, la précision de la vitesse radiale d'une étoile augmente avec la racine du nombre d'observations de cette étoile.

### 3.2.5 Reference Selection : Sélection des étoiles de référence

Le but de la sélection des étoiles de référence est d'identifier les étoiles qui disposent de caractéristiques appropriées pour le processus de calibration en longueur d'onde. Ainsi, pour être une étoile de référence, l'étoile devra être :

- stable en vitesse radiale ;
- d'un type stellaire approprié, i.e. un type stellaire ayant environ 15 à 20 raies peu blendées qui pourront être utilisées afin d'identifier les longueurs d'onde dans le spectre de l'étoile ;
- bien localisée (bonnes mesures astrométriques).

Cette définition est propre aux besoins du processus de calibration en longueur d'onde implémenté durant mon travail de recherche. En effet, la phase de sélection des étoiles de référence a été implémentée de manière à ne sélectionner que des étoiles ayant une certaine température et une certaine métallicité. Ainsi, toutes les étoiles observées entrant dans le processus itératif avec une température et une métallicité telles que  $3750 < T_{eff} < 6000$  et  $[Fe/H] > -0.5$ , sont considérées comme des références pour les calibrations en longueur d'onde. Cependant, cette définition n'est pas définitive. Elle devra faire l'objet de recherches plus approfondies afin de tester l'impact du choix des critères de sélection des références sur les propriétés de convergence de SGIS. Par exemple, si on considérait des types stellaires plus chauds comme référence, quelles seraient les conséquences sur le comportement du prototype de SGIS.

### 3.2.6 Calibration Updating : Calibration en longueur d'onde

Comme cela a été décrit dans la section 2.4, une forme polynomiale est utilisée pour représenter les lois de calibration. Le prototype travaillant sur une base de données "simple" ayant des lois linéaires de calibration en longueur d'onde, la forme polynomiale de la loi de dispersion spectrale ne nécessite pas d'ordres élevés. Par conséquent, une forme du second degré (i.e. développement d'ordre deux en  $\eta$  et en  $\zeta$ ) suffit pour contraindre les lois de calibration de la base de données utilisée. La forme polynomiale, présentée dans l'équation 2.4, devient :

$$\lambda_s = \sum_{m=0}^2 \sum_{n=0}^2 C_{mn} \eta_s^m \zeta_s^n \quad (3.3)$$

La loi de dispersion spectrale établie dans l'équation 3.3 est supposée stable sur une unité de calibration. L'objectif de la phase de "Calibration updating" est donc de calculer les coefficients de calibration  $C_{mn}$  pour chaque unité de calibration.

La phase de calibration en longueur d'onde se décompose en deux étapes successives :

1. Identifier et localiser les raies de référence dans chaque spectre de l'unité de calibration afin d'obtenir un jeu de données  $(\lambda, \eta, \zeta)$  ;
2. Calculer les coefficients  $C_{mn}$  avec les données  $(\lambda, \eta, \zeta)$  de l'unité de calibration.

Ces deux étapes sont exécutées unité de calibration par unité de calibration, afin de couvrir l'ensemble mission.

### Détection et localisation des raies de référence

Pour estimer la forme polynomiale de la loi de dispersion spectrale, le prototype doit connaître les positions en longueur d'onde  $\lambda$  de plusieurs raies stellaires de référence en fonction de leurs coordonnées angulaires  $(\eta, \zeta)$ . Afin de contraindre cette forme polynomiale de manière homogène sur le domaine spectral du RVS, ces raies de référence doivent être bien réparties sur chaque spectre observé durant l'unité de calibration  $k$ .

Un module de détection et localisation des raies de référence a été implémenté dans le prototype. Ce module fonctionne en trois temps :

1. Détection des raies stellaires dans le spectre brut.
2. Identification des raies stellaires détectées par comparaison avec une liste de raie de référence appropriée au type stellaire de la source observée. La sélection de cette liste utilise un critère identique à celui utilisé lors de la sélection des spectres synthétiques dans la phase d'estimation des vitesses radiales (c.f. section 3.2.4).
3. Estimation de la position du centre (ou "centroïde") de la raie de référence identifiée.

Ce module de détection et localisation des raies de référence s'est révélé crucial vis à vis des propriétés de convergence du processus itératif. C'est pourquoi, le chapitre 4 est dédié aux différentes méthodes numériques implémentées dans ce module et à leurs performances dans le cadre de la calibration en longueur d'onde.

À la fin de cette étape, un triplet  $(\lambda, \eta, \zeta)$  est établi pour chaque raie de référence identifiée de chaque spectre d'étoile de référence observée durant une unité de calibration  $k$  donnée.

### Calcul des coefficients de calibration

Une fois les ensembles  $(\lambda, \eta, \zeta)$  constitués, l'ajustement de la forme polynomiale se fait par une méthode de moindres carrés en minimisant le critère  $S_k$  pour l'unité de calibration  $k$  :

$$S_k = \sum_{spec=1}^{N_{spectres}} \sum_{l=1}^{N_{lines}} \left[ \lambda(spec, l) - \sum_{m=0}^2 \sum_{n=0}^2 C(k)_{mn} \eta^m \zeta^n \right]^2 \quad (3.4)$$

où  $spec$  représente le spectre d'une étoile observée durant l'unité de calibration  $k$ ,  $N_{spectres}$  le nombre de spectres observés durant l'unité de calibration  $k$ ,  $l$  une raie de référence du spectre observé  $spec$ ,  $N_{lines}$  le nombre de raies de référence dans le spectre  $spec$ ,  $\lambda(spec, l)$  une position

en longueur d'onde de la raie de référence  $l$  du spectre  $spec$ ,  $\eta(spec, l)$  et  $\zeta(spec, l)$  les coordonnées angulaires de la raie de référence  $l$  du spectre  $spec$ , et  $C(k)_{mn}$  les coefficients de calibration de l'unité de calibration  $k$ .

La minimisation du critère  $S_k$  passe par la recherche des coefficients de calibration  $C_{mn}$  pour lesquels le gradient de  $S_k$  s'annule. Cette minimisation nécessite la résolution d'un système linéaire de type  $Ax = b$ , représenté dans les équations 2.7 et 2.8 du chapitre 2. Dans un premier temps, la résolution de ce système a été implémentée en utilisant une méthode classique LU, puis une méthode du type pivot de Gauss (c.f. Press et al. (2002)). Les résultats obtenus avec ces deux méthodes étaient similaires. La version actuelle du prototype utilise la résolution du système linéaire implémentée dans la Gaia ToolBox, commune à tous les développeur du DPAC.

À la fin de cette étape, les coefficients de calibration  $C_{mn}$  résultant de la résolution linéaire sont mis à jour dans les données instrumentales.

### 3.3 Spectroscopic Global Iterative Solution, au niveau du consortium de traitement de données de Gaia

Le Data Processing Analysis Consortium (DPAC) a été mis en place pour le développement de la chaîne de traitement de données Gaia. Il est organisé en "coordination units", dont la CU6 responsable du traitement des données spectroscopiques du RVS. Le "Spectroscopic Data Processing Centre" (SDPC) va se charger du traitement des données spectroscopiques et de la diffusion des vitesses radiales et rotationnelles, des spectres extraits et calibrés, de diagnostics astrophysiques, des alertes scientifiques sur les observations, des calibrations du RVS, etc. Ce centre SDPC est au CNES de Toulouse.

Dans la CU6, le processus de calibration est planifié en deux temps : un processus quotidien et un processus semestriel. Le processus de calibration des données quotidien va fournir des premières estimations des caractéristiques de l'instrument RVS. Ces estimations seront utilisées dans le reste du processus quotidien de traitement de données. Elles permettront de vérifier rapidement le bon état et le bon fonctionnement de l'instrument. Cependant, ces premières estimations des caractéristiques de l'instrument seront déterminées avec un petit nombre d'observations dans un intervalle de temps très court (inférieur au jour). Elles ne seront donc pas de très bonne qualité. Un processus de calibration semestriel sera mis en place afin d'affiner les estimations sur les calibrations en utilisant la méthode SGIS. Ce processus va permettre de prendre en compte la continuité entre les intervalles de temps successifs, et de prendre en compte un grand nombre d'observations par étoile (meilleure précision sur la vitesse radiale des étoiles).

La méthode SGIS est incluse dans le processus semestriel de traitement de données spectroscopiques. Elle est répartie sur plusieurs modules du pipeline de traitement de données spectroscopiques. La figure 3.9 montre le diagramme fonctionnel du processus semestriel du traitement de données spectroscopiques de Gaia. La méthode de calibration SGIS est représentée par les actions décrites dans des boîtes bleues :

- **Extraction - SGIS loop** : le rôle de ce module est d'extraire les spectres, d'appliquer les calibrations (e.g. application des lois de calibration en longueur d'onde) et de nettoyer les données spectroscopiques brutes. Ce module d'extraction permet de fournir un spectre mono-dimensionnel dans des unités physiques (par exemple, des longueurs d'onde à la place des samples) aux autres modules du pipeline de traitement de données spectroscopiques.
- **Single-transit analysis - SGIS loop** : le rôle de ce module est d'estimer les vitesses radiales et les vitesses de rotation des spectres calibrés (c'est-à-dire, des spectres sortis du module d'extraction).
- **Calibration - SGIS loop** : ce module estime les différentes lois de calibration (e.g. les lois de dispersion spectrale) de l'instrument spectroscopique de Gaia.

Les trois phases itératives de la méthode SGIS sont donc réparties sur ces trois modules de la CU6 :

- la phase de Source Updating est séparée en deux : on applique une échelle en longueur d'onde appropriée sur chaque spectre brut dans le module "Extraction" et la vitesse radiale de chacun de ces spectres calibrés est calculée dans le module "Single-transit analysis".
- la phase de Reference Selection est incluse dans le module "Calibration".
- la phase de Calibration Updating est incluse dans le module "Calibration".

Dans le groupe de traitement des données spectroscopiques (CU6) du DPAC, je travaille dans le groupe de calibration de l'instrument spectroscopique sous la responsabilité de chercheurs et d'ingénieurs du Mullard Space Science Laboratory (University College of London, UK) et en collaboration avec les ingénieurs du CNES. Mon travail de recherche permet de fournir une implémentation de la calibration en longueur d'onde dans le pipeline de traitement des données spectroscopiques. J'assure ainsi le développement et l'implémentation du module de calibration en longueur d'onde et d'un sous-module de l'extraction, c'est-à-dire l'application des calibrations en longueur d'onde sur les spectres bruts. Tous les détails concernant le module de calibration en longueur d'onde du RVS sont présentés dans la table 3.2.



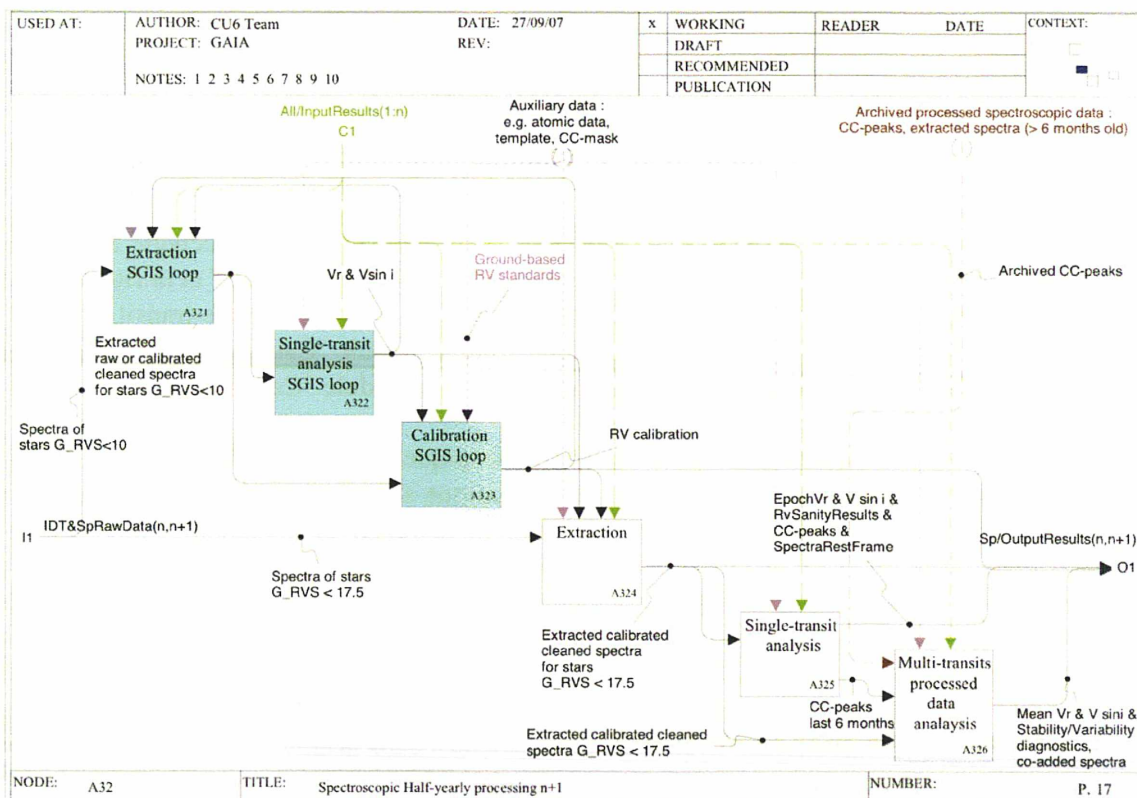


FIG. 3.9 – Diagramme fonctionnel du processus semestriel de traitement de données. Ce diagramme est un actigramme SADT : chaque boîte représente une action et chaque flèche un flot de données. La nature du flot de donnée est définie par l'orientation et la direction de la flèche par rapport à la boîte (i.e. action) : une flèche entrant sur le flanc gauche de la boîte représente un flot de données d'entrée, une flèche sortant du flanc droit est un flot de données de sortie et une flèche entrant sur le haut de la boîte représente le flot de données de contrôle.

### 3.3. Spectroscopic Global Iterative Solution, au niveau du consortium de traitement de données de Gaia

<b>Gaia DPAC WP : GWP-S-630-10000</b>	
<b>Title :</b> Wavelength scale and AC Distortion map	
<b>Provider(s) :</b> MSSL	
<b>Manager(s) :</b> Howard Huckle	
<b>Start :</b> 01/10/2006	<b>End :</b> 01/03/2012
<b>Objective :</b>	
To develop the code to derive the wavelength scale and AC spatial distortion for calibration	
<p>This workpackage will produce the code which identifies and quantifies the wavelength scale and AC spatial distortion as a function of position on the RVS focal plane. This workpackage does not address the AL distortion, as this is mixed with the wavelength zero point. The detailed wavelength scale is determined by the SGIS (GWP-S-630-06000), but it needs to be referenced to external standards (identified in GWP-S-640-x). As there will be no dedicated calibration observations, for the AC spatial distortion this code will need to draw on data derived during nominal operation, and in particular on the 2-D windows available from bright stars.</p> <ol style="list-style-type: none"> <li>1. identify algorithms for determining and applying wavelength zeropoints from external standards to the SGIS internal solutions</li> <li>2. investigate the selection procedures that will be needed to recognize and isolate suitable observations for the AC distortion determination as a function of magnitude <i>etc.</i></li> <li>3. develop code to construct AL wavelength zero points and absolute wavelength scale from SGIS internal calibration and from external wavelength standards as a function of magnitude, windowing scheme, CCD number and binning, AC position <i>etc</i> including derivation and propagation of systematic and random errors</li> <li>4. develop code to construct AC distortions as a function of magnitude, windowing scheme, CCD number and binning, AC position <i>etc</i> including derivation and propagation of systematic and random errors</li> <li>5. develop code to incorporate wavelength scale calibration information into calibration files</li> <li>6. develop code to incorporate AC distortion calibration information into calibration files</li> <li>7. develop verification code</li> <li>8. verify wavelength scale and wavelength zero-point calibration code using simulated data</li> <li>9. verify AC distortion calibration code using simulated data</li> </ol>	

Suite de la table 3.2 page suivante

<b>Gaia DPAC WP : GWP-S-630-10000</b>
<b>Title :</b> Wavelength scale and AC Distortion map
Suite de la table 3.2
<b>Inputs :</b> <ol style="list-style-type: none"> <li>1. GAIA-C6-TN-OPM-DK-001-2 : <i>Gaia spectroscopic processing – Preliminary functional and data flow analysis</i></li> <li>2. Gaia-3 design and operational parameters</li> <li>3. simulated datasets</li> </ol>
<b>Outputs :</b> <ol style="list-style-type: none"> <li>1. Updated functional analysis for spectral extraction : GWP-S-620-02000</li> <li>2. Updated functional analysis for calibration : GWP-S-630-02000</li> <li>3. verified code module for AC distortion calibration</li> <li>4. verified code module for AL wavelength zero-point calibration</li> <li>5. test datasets</li> </ol>
<b>Deliverables :</b> <ol style="list-style-type: none"> <li>1. verified code module for AC distortion calibration</li> <li>2. verified code module for AL wavelength zero-point calibration</li> <li>3. test datasets</li> </ol>

TAB. 3.2: Extrait du "Software Development Plan" pour le traitement de données spectroscopiques ia, Katz et al (2006)

J'ai fourni les premiers modules d'"application de la calibration en longueur d'onde" (ce module fait partie de l'extraction) et de "calibration en longueur d'onde", après les avoir préalablement validés par des séries de tests. Les premières versions de mes modules ont été intégrées dans le pipeline de traitement de données spectroscopiques au CNES fin mars 2007, dans le cadre du cycle 2 de développement de la chaîne de réduction spectroscopique.

# Chapitre 4

## Détection et localisation des raies de référence

### Sommaire

---

<b>4.1</b>	<b>Description du jeu de données</b>	<b>59</b>
<b>4.2</b>	<b>Détection de raies dans les spectres bruts</b>	<b>61</b>
4.2.1	Description d'une raie observée	61
4.2.2	Principe de "détection"	61
4.2.3	Limites de la méthode de détection	63
<b>4.3</b>	<b>Localisation de raies de référence par ajustement parabolique</b>	<b>64</b>
4.3.1	Principe de la localisation des raies de référence	64
4.3.2	Performance de la méthode	69
4.3.3	Calibration des biais	71
4.3.4	Performances après calibration des biais	72
<b>4.4</b>	<b>Localisation des raies de référence par ajustement du profil des raies</b>	<b>76</b>
4.4.1	Principe de la double corrélation croisée	76
4.4.2	Recherche du maximum de la surface de corrélation	77
4.4.3	Optimisation de la méthode	79
4.4.4	Performance de l'ajustement de profil	81
<b>4.5</b>	<b>Localisation accélérée de raies de référence par ajustement de profil</b>	<b>83</b>
<b>4.6</b>	<b>Synthèse</b>	<b>86</b>

---

Comme cela a été évoqué dans la section 3.2.6 du chapitre 3, la méthode SGIS de calibration en longueur d'onde repose sur la mesure des positions de raies de référence dans des spectres collectés par le RVS.

La mesure des positions des raies de référence va permettre de relier la longueur d'onde d'une raie de référence à un sample  $s$ . Cela nous permet de déterminer un triplet  $(\lambda, \eta(s), \zeta(s))$  pour chaque raie de référence de chaque spectre, où :

- $\lambda$  est la position en longueur d'onde du centre de la raie de référence. Cette position est décalée par rapport à un référentiel relatif aux données observées en utilisant les vitesses radiales estimées.

- $\eta(s)$  et  $\zeta(s)$  sont les coordonnées angulaires de l'étoile dans le champ de vue du RVS, au moment de la lecture du sample  $s$  du centre de la raie de référence.

Ces triplets vont être utilisés pour former le système linéaire de la calibration en longueur d'onde (c.f. les équations 2.7 et 2.8).

Afin d'établir cette association longueur d'onde - sample, deux méthodes ont été développées et implémentées : la localisation de raies de référence par ajustement polynomial et la localisation de raies de référence par ajustement de profil. Chacune de ces deux méthodes commence par isoler les raies les plus profondes du spectre extrait en utilisant une méthode commune de détection de raie. Les deux méthodes associent le centre de la raie détectée à une longueur d'onde. Ce centre de la raie est obtenu via deux estimateurs :

- un "centre discret" : il correspond au sample ayant le flux le plus faible. C'est une estimation au sample près du centre de la raie.
- une "position du minimum" : il correspond à la position en samples du minimum du polynôme ajusté sur la raie. C'est une estimation du centre d'une précision inférieure au sample.

La méthode de localisation de raie par ajustement polynomial associe la longueur d'onde d'une raie de référence à la position du minimum d'une raie détectée. Ces longueurs d'onde proviennent d'une liste de raies de référence choisies à partir des paramètres atmosphériques de l'étoile observée.

La méthode de localisation de raies de référence par ajustement de profil associe le centre discret d'une raie détectée à une longueur d'onde d'une raie de référence. La méthode sélectionne un template à partir des paramètres atmosphériques de la source. Ensuite, elle calcule une loi de calibration en longueur d'onde en ajustant le profil de la raie de référence du template sur celui de cette raie identifiée dans le spectre. Via cette loi de calibration calculée, le centre discret de la raie identifiée est associé à une longueur d'onde.

Une étude a été réalisée afin d'évaluer les performances des deux méthodes de localisation. Ces performances permettent de dresser un bilan comparatif des deux méthodes et de choisir celle qui sera intégrée dans le prototype de calibration SGIS.

Les deux méthodes de localisation des raies de référence comportent deux étapes :

1. **Détection des raies** dans le spectre extrait<sup>12</sup> : cette étape fournit une liste de raies détectées dans le spectre
2. **Identification et localisation des raies de référence** : cette étape va faire correspondre une raie détectée dans le spectre extrait avec une raie de référence dont la position en longueur d'onde est connue dans le référentiel du laboratoire. Pour chacune des raie identifiées dans le spectre extrait, la position du centre de la raie est associée à la longueur d'onde de la raie de référence.

Dans chacune des méthodes implémentées, l'étape de détection est identique. Seules les étapes d'identification et de localisation des raies de référence diffèrent.

---

<sup>12</sup>Rappel : le spectre observé ou spectre extrait est défini par un nombre de photons par sample et par des coordonnées angulaires ( $\eta, \zeta$ ) de l'étoile observée dans le champ de vue du RVS au temps de lecture de chaque sample.

Les performances des deux méthodes ont été obtenues à partir d'un jeu de données commun. Ce jeu de données a été réalisé en dégradant des spectres synthétiques à la résolution de l'instrument RVS. Dans un premier temps, je décris ce jeu de données RVS dans la section 4.1. Puis, je décris le processus de détection des raies (c.f. section 4.2), puis les différentes méthodes et leurs performances dans les sections 4.3 et 4.4.

## 4.1 Description du jeu de données

Afin de tester le bon comportement des différents modules et des différentes implémentations du prototype de SGIS, une base de données spectroscopiques a été mise au place en utilisant un simulateur de l'instrument RVS. Ce dernier a été développé au sein du GEPI par David Katz, et a été modifié pour les besoins de la calibration en longueur d'onde.

Le simulateur utilise une bibliothèque de spectres synthétiques de Kurucz (i.e. les spectres ont une résolution de 250000) pour simuler des spectres RVS ( $R = 11500$ ). Ces spectres simulés sont obtenus en réalisant les opérations suivantes :

- Le spectre de Kurucz ( $R = 250000$ ) est convolué avec un profil gaussien afin de le dégrader à la résolution de l'instrument RVS (i.e.  $R = 11500$ ).
- Les coordonnées angulaires ( $\eta, \zeta$ ) dans le champ de vue de l'instrument sont générées au temps de lecture de chaque sample. La position de l'étoile est aléatoirement bruitée, en utilisant une distribution uniforme, dans un domaine angulaire correspondant à  $[-0.5 : 0.5]$  sample ;
- La longueur d'onde centrale de chaque sample est calculée en utilisant la formule :  $\lambda = 8480 + 1350\eta$ , où  $\lambda$  est la longueur d'onde centrale du sample exprimée en  $\text{\AA}$  et  $\eta$  la position AL de l'étoile dans le champ de vue du RVS au temps de lecture du sample. Cette transformation correspond à un échantillonnage de  $0.375 \text{\AA}$  par sample (i.e. ancien design de l'instrument RVS). En prenant le même formalisme que celui qui est utilisé dans la section 2.4.3, la loi de calibration en longueur d'onde vraie est décrite par les coefficients de calibration suivants :  $C_{00} = 8480 \text{\AA}$ ,  $C_{10} = 1350 \text{\AA} \cdot \text{deg}^{-1}$  et tous les autres  $C_{mn}$  nuls.
- Le spectre de Kurucz est rééchantillonné en utilisant l'échelle de longueur d'onde calculée à l'étape précédente.
- Le spectre simulé est normalisé en flux en tenant compte de la magnitude apparente de l'étoile et de la réponse instrumentale.
- Le bruit de photon (pour l'étoile et pour la lumière zodiacale) est généré suivant une distribution Poissonienne.
- et le bruit de lecture est ajouté au spectre suivant une distribution Gaussienne.

Les spectres sont simulés CCD par CCD.

Tous les spectres simulés de la base de données du prototype proviennent d'un unique spectre synthétique, et ont ainsi les mêmes caractéristiques astrophysiques, i.e. celles d'une étoile de type solaire :

- Magnitude apparente :  $V = 8$  ;

- Température effective : 5500K ;
- Logarithme de la gravité de surface : 4.00 ;
- Metallicité : 0.00 ;
- Vitesse radiale : 0.00 km/s.

La figure 4.1 représente un des spectres simulés utilisé comme donnée d'entrée d'un prototype de SGIS. Les raies spectrales étant bien distinctes du continu (e.g. les trois raies les plus profondes appartiennent au triplet du Calcium ionisé), ce type de spectre sont des données de bonne qualité pour les calibrations en longueur d'onde.

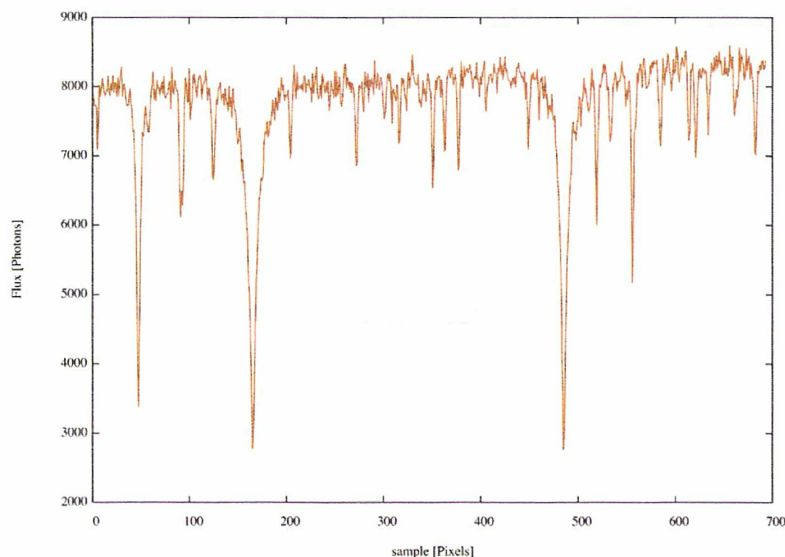


FIG. 4.1 – Exemple de spectre d'une étoile G5V, obtenu par le simulateur du RVS. Les trois raies les plus profondes sont les raies du calcium ionisé.

En sortie du simulateur, les spectres de type RVS sont représentés par :

- un fichier contenant le flux du spectre pour chaque sample (une ligne représente le flux d'un sample) ;
- un fichier contenant la coordonnée angulaire AL (i.e.  $\eta$ ) de l'étoile pour chaque sample ;
- un fichier contenant la coordonnée angulaire AC (i.e.  $\zeta$ ) de l'étoile pour chaque sample ;
- un fichier contenant la longueur d'onde centrale pour chaque sample du spectre.

Dans le concept de SGIS proposé, les longueurs d'onde ne sont pas censées être accessibles au prototype, puisque ce dernier les calcule à partir de coefficients de calibration. Ainsi, les longueurs d'onde générées par le simulateur ne seront utilisées que pour évaluer les performances du prototype. Elles ne sont pas utilisées par le prototype lui-même.

Les temps d'observation de chaque spectre simulé sont générés par un simulateur de la loi de balayage du satellite Gaia. Pour la base de données, ce simulateur a fourni les temps d'observation de 2000 étoiles pour 700 jours de mission (i.e. environ 80000 spectres dans la base de données). Cependant, pour des raisons de temps de calcul, seulement une partie de ces observations est utilisée lors de l'exécution du prototype de calibration (i.e. environ 12000

spectres répartis sur 110 jours d'observation).

## 4.2 Détection de raies dans les spectres bruts

### 4.2.1 Description d'une raie observée

Dans le prototype, une raie est décrite par quatre caractéristiques :

- **un centre discret**, c'est-à-dire le sample ayant le flux le plus faible dans la raie ;
- **une position du minimum**, c'est-à-dire la position en samples correspondant au minimum de la parabole ajustant la raie ;
- **un nombre de samples dans l'aile rouge de la raie**, c'est-à-dire les samples entre le centre discret de la raie et le continu ;
- **un nombre de samples dans l'aile bleue de la raie**, c'est-à-dire les samples entre le continu et le centre discret de la raie ;

La figure 4.2 illustre la description d'une raie dans le prototype. Elle nomme l'une des raies du Calcium ionisé d'une étoile G5V et 3 des 4 caractéristiques citées précédemment :

- le centre discret de la raie ;
- le nombre de samples dans l'aile rouge de la raie ;
- le nombre de samples dans l'aile bleue de la raie.

Ces trois caractéristiques sont obtenues pour chaque raie détectée lors du processus de détection expliqué plus bas (c.f. section 4.2.2).

### 4.2.2 Principe de "détection"

L'objectif de la détection est de trouver les  $N$  raies les plus "profondes"<sup>13</sup> dans le spectre extrait. Cette détection se fait d'une manière "empirique". Elle est constituée de trois étapes :

1. Dans un premier temps, la méthode de détection trouve la raie la plus profonde dans le spectre extrait. Le flux le plus bas du spectre est associé au centre discret de la raie.
2. Dans un second temps, la méthode va détecter les ailes de cette raie. Elle va chercher les samples appartenant à la raie autour du centre discret déterminé lors de l'étape précédente. Lors de la détection de l'aile bleue de la raie, le flux du sample adjacent le plus bleu est comparé à celui du centre discret de la raie :
  - Si le flux du sample adjacent est plus grand que celui du centre discret de la raie, alors on considère que ce sample adjacent fait partie de cette même raie.
  - Sinon, on suppose que le sample adjacent appartient au continu ou à une autre raie détectable. Il n'est donc pas pris en compte. Dans ce cas, la détection de l'aile bleue est achevée.

Un processus de comparaison similaire est répété entre le sample le plus bleu détecté dans l'aile de la raie et son sample bleu adjacent dans le spectre, et ainsi de suite, jusqu'à ce qu'on suppose avoir atteint le continu du spectre. L'aile bleue de la raie a ainsi été détectée. Tout le processus entier est répété pour détecter l'aile rouge de la raie.

<sup>13</sup>la profondeur d'une raie qualifie le contraste entre la raie et le continu du spectre



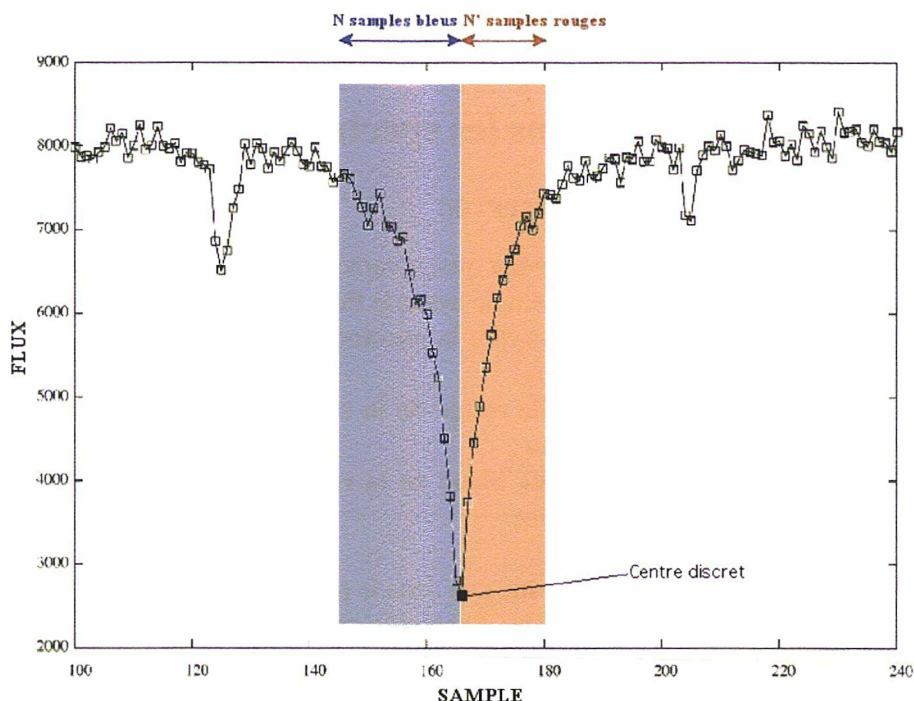


FIG. 4.2 – Exemple de l'une des raies du Calcium ionisé observée dans le spectre extrait. Trois des quatre caractéristiques décrivant une raie dans le prototype  $y$  sont illustrées : un centre discret, l'aile bleue de la raie et l'aile rouge de la raie.

3. Tous les samples et flux appartenant à cette raie sont exclus du spectre observé. Un nouveau spectre est établi.

Après avoir détecté la raie la plus profonde, on veut détecter la seconde raie la plus profonde. Pour ce faire, le processus de détection va être exécuté sur le nouveau spectre (i.e. spectre dont la raie la plus profonde détectée a été soustraite). Ce processus est donc répété jusqu'à ce que les  $N$  raies les plus profondes soient détectées.

Pour avoir une estimation précise (i.e. inférieure au sample) du centre des raies, on a besoin de détecter un maximum de points dans chaque raie profonde. Cependant, la méthode de détection proposée est limitée par le profil complexe des raies. En effet, le profil d'une raie dans le spectre observé peut être la superposition de différentes raies spectrales proches. Ce phénomène est dû à l'instrument : la PSF va élargir le profil des raies et le spectre est discrétisé au moment de la lecture de chaque sample. Dans ce cas, on ne parle plus de raie, mais de "blend" (ensemble de raies spectrales proches). Lors du processus de détection, ce blend peut être détecté en deux temps : la composante principale (i.e. la plus profonde) est détectée en premier, puis la seconde composante est détectée. Chaque composante est donc décrite comme une "raie" détectée. Pour ne pas être limité par ces blends, chaque raie détectée est comparée aux précédentes : si la distance entre deux raies détectées est inférieure à trois samples, alors on suppose que ces deux raies forment une seule et même raie. Le processus reconstruit ainsi, petit à petit, le blend sans perdre d'information.

La figure 4.3 donne un aperçu des résultats du processus de détection utilisé pour une étoile de type solaire (G5V). Le spectre de l'étoile dans la bande RVS est représenté sur le graphique de gauche. Les vingt raies les plus profondes de ce spectre sont détectées et apparaissent en couleur sur le graphique de droite.

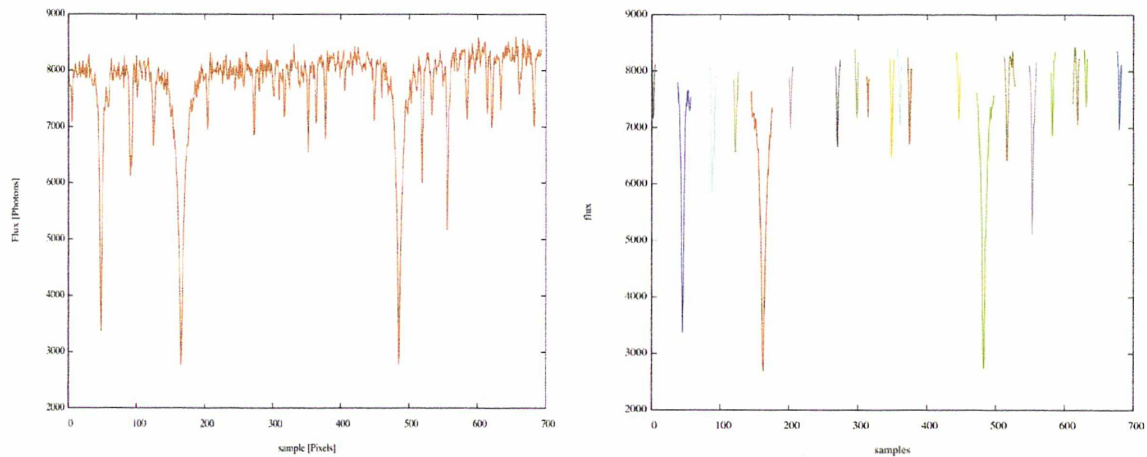


FIG. 4.3 – Exemple de détection de raies dans un spectre RVS d'une étoile de type solaire G5V : à gauche, le spectre RVS extrait, à droite, les raies détectées dans ce spectre.

### 4.2.3 Limites de la méthode de détection

Le nombre de raies à détecter étant fixé, la méthode utilisée peut faire de fausses détections dans des spectres possédant peu de raies profondes : si le nombre de raies significatives est inférieur au nombre de raies à détecter, la méthode va continuer sa détection au niveau du continu, après avoir trouvé les principales raies. Par conséquent, les plus fortes fluctuations du continu vont être considérées comme des raies. Ce phénomène peut se produire pour des étoiles très chaudes, e.g. des étoiles de type A dominées par les raies d'hydrogène et contenant "peu" de raies. Ces fausses détections n'ont pas de conséquence sur l'estimateur de centre des raies : elles sont écartées de la suite du processus grâce à l'étape d'identification.

Notre méthode de détection est basée sur un classement par flux croissant. Cela peut poser un problème dans les cas où le continu présente une pente, e.g. certaines zones du continu peuvent être plus faibles en intensité que d'autres. Par conséquent, les plus faibles zones du continu sont privilégiées dans le classement réalisé : les fluctuations<sup>14</sup> de ces zones peuvent être détectées au détriment de la présence de raies dans des zones où le continu est plus intense. La figure 4.4 représente la détection de raies pour un spectre d'une étoile F (7000 K). Dans le graphique de droite, le cercle rouge illustre un exemple de fausses détections au niveau du continu du spectre. Dans le graphique de gauche, le cercle noir illustre un exemple de deux raies bien distinctes du continu et non-détectées.

Ce problème de continu non-redressé peut se traduire par une perte d'information car les raies non détectées ne seront pas utilisées par l'algorithme d'identification. Cela peut aboutir à

<sup>14</sup>le terme fluctuation désigne un événement de faible profondeur, plus généralement associé au continu

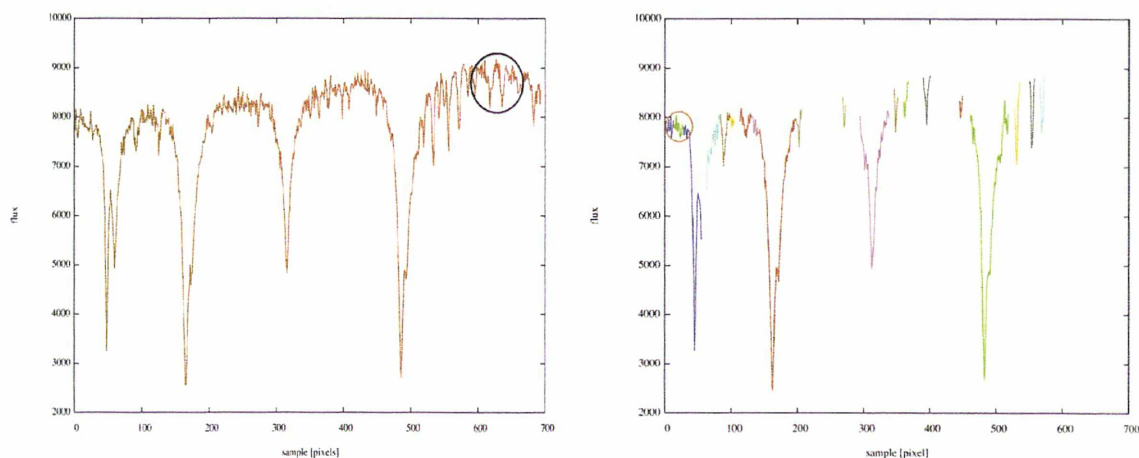


FIG. 4.4 – Exemple de détection de raies dans un spectre RVS d’une étoile de type F (7000 K) : à gauche, le spectre RVS extrait, à droite, les raies détectées dans ce spectre. À gauche, le cercle noir montre la présence de deux raies significatives non-détectées. À droite, le cercle rouge montre une fausse détection de deux raies.

une réduction du nombre d’estimation de triplets ( $\lambda, \eta, \zeta$ ), et donc à une moins bonne précision au niveau du processus de calibration en longueur d’onde.

Afin d’éviter des pertes d’information lors de l’étape de détection, il est préférable de ne pas travailler avec des spectres ayant un continu trop complexe.

### 4.3 Localisation de raies de référence par ajustement parabolique

La méthode de détermination du centre des raies de référence a pour but d’estimer la position en samples du centre des raies (dont les longueurs d’onde sont connues).

#### 4.3.1 Principe de la localisation des raies de référence

L’estimateur de centre des raies de référence utilise des spectres extraits. L’estimateur n’a donc pas d’information sur les longueurs d’onde du spectre. Ainsi, pour identifier les raies de référence, l’estimateur se base sur les propriétés de la source qui ont un impact sur la morphologie sur spectre : les paramètres atmosphériques (température effective, métallicité et logarithme de gravité de surface) vont avoir un impact sur la largeur des raies et leur nombre dans le spectre, la vitesse radiale indique le décalage de ces raies dans le spectre. De plus, en considérant que seules les étoiles de type F, G ou K vont être qualifiées comme références lors de la phase de référence selection, ce sont les raies du calcium qui seront les raies les plus profondes des spectres extraits utilisés. Par conséquent, le principe d’identification est basé sur la recherche du triplet du Calcium ionisé dans les spectres extraits. Dans un premier temps, on va sélectionner une liste de raies de référence à partir des paramètres atmosphériques de la source. Cette liste nous

donne les positions théoriques des raies les plus profondes dans le spectre. Puis, les trois raies les plus profondes sont associées aux longueurs d'onde du triplet du calcium. À partir de cette identification, une loi de calibration inversée est calculée. Elle permet d'associer une longueur d'onde à une position en sample. Afin d'identifier chaque raie détectée, cette loi de calibration inversée est appliquée sur la liste de raies de référence. Une position en sample dans le spectre extrait est estimée pour chaque longueur d'onde. Si la position en sample d'une raie de référence appartient à une raie détectée, on suppose que la raie détectée peut être associée à cette raie de référence. Pour avoir une estimation de la position du centre de la raie de moins d'un sample, le pic de la raie (i.e. les  $N$  samples où les flux sont les plus bas) est ajusté à une parabole.

Cette méthode se compose donc de 4 étapes :

- Sélection d'une liste de raies de référence ;
- Identification et estimation du centre des raies du Ca II ;
- Estimation d'une loi de calibration inversée ;
- Identification et estimation du centres des raies de référence restantes.

#### Sélection d'une liste de raies de référence

L'identification utilise une liste de raies de référence appropriée aux paramètres atmosphériques de l'étoile observée. Elle est propre à chaque type d'étoile observée. Pour un spectre donné, la liste de raies est sélectionnée en utilisant le même critère que celui qui est utilisé lors du "Source Updating", présenté dans la section 3.2.4 du chapitre 3. La sélection se fait à partir des paramètres atmosphériques de la source.

Cette liste est constituée de l'ensemble des longueurs d'onde théoriques  $\lambda_{ref}$  des raies les plus profondes dans le spectre. Ces positions sont exprimées dans le référentiel du laboratoire ( $V_r = 0 \text{ km.s}^{-1}$ ).

Pour pouvoir associer chacune de ces longueurs d'onde à une raie détectée dans le spectre extrait, il faut appliquer sur ces positions le même décalage en longueur d'onde que celui qui est estimé dans le spectre extrait (i.e. vitesse radiale calculée lors du source updating). Ainsi, ces longueurs d'onde sont exprimées dans le référentiel relatif aux données en utilisant la formule :

$$\lambda_{ref*} = \lambda_{ref} \left( 1 + \frac{V_r}{c} \right)$$

où  $V_r$  est la vitesse de l'étoile observée,  $c$  la vitesse de la lumière,  $\lambda_{ref}$  la position en longueur d'onde de la raie de référence exprimée dans le référentiel du laboratoire, et  $\lambda_{ref*}$  la position en longueur d'onde de la raie de référence exprimée dans le référentiel relatif aux données.

À la fin de l'étape, la méthode a bâti une liste de raies de référence propre au spectre traité.

#### Identification et estimation du centre des raies du Ca II

En considérant que ces raies du Calcium sont les plus profondes et en supposant que le continu n'est pas complexe, les trois premières raies détectées dans chaque spectre brut peuvent être associées au triplet du Calcium.

En utilisant la méthode d'ajustement parabolique sur le profil de ces raies du Ca II (méthode décrite plus bas), une position du minimum  $s_{est}$  de chacune de ces raies peut être estimée. Pour chacune de ces raies du Calcium, un couple  $(\lambda_{ref*}, s_{est})$  est ainsi formé.

### Estimation d'une loi de calibration inversée

Avec ces premiers couples  $(\lambda_{ref*}, s_{est})$ , une loi de calibration en longueur d'onde inversée va être estimée sous une forme linéaire de type  $s_{est} = a\lambda_{ref*} + b$ , où  $a$  et  $b$  sont des inconnues à déterminer. Soit  $(\lambda_{ref*1}, s_{est1})$  les coordonnées de la raie du calcium la plus bleue,  $(\lambda_{ref*3}, s_{est3})$  celles de la raie du calcium la plus rouge, et enfin  $(\lambda_{ref*2}, s_{est2})$  celles de la raie "centrale". La loi de calibration inversée est calculée par interpolation entre les deux domaines  $[\lambda_{ref*1}, \lambda_{ref*2}]$  et  $[\lambda_{ref*2}, \lambda_{ref*3}]$ , et est définie par :

$$\text{Loi de calibration inversée : } \begin{cases} a = \frac{s_{est2} - s_{est1}}{\lambda_{ref*2} - \lambda_{ref*1}} \text{ et } b = \frac{s_{est2}(\lambda_{ref*2} - \lambda_{ref*1}) - \lambda_{ref*2}(s_{est2} - s_{est1})}{\lambda_{ref*2} - \lambda_{ref*1}} & \text{si } \lambda_{ref*} \leq \lambda_{ref*2} \\ a = \frac{s_{est3} - s_{est2}}{\lambda_{ref*3} - \lambda_{ref*2}} \text{ et } b = \frac{s_{est3}(\lambda_{ref*3} - \lambda_{ref*2}) - \lambda_{ref*3}(s_{est3} - s_{est2})}{\lambda_{ref*3} - \lambda_{ref*2}} & \text{si } \lambda_{ref*} > \lambda_{ref*2} \end{cases}$$

### Identification et estimation du centre des raies de référence restantes

En utilisant la loi de calibration inversée pour les longueurs d'onde  $\lambda_{ref*}$  des autres raies de référence, on peut avoir une estimation de la position des raies de référence dans le spectre extrait. La figure 4.5 illustre le principe de la localisation des raies de référence en utilisant une loi  $sample = f(\lambda)$  obtenue avec le triplet du calcium.

Une fois localisées, les raies de référence sont comparées aux raies détectées. La position  $s_{interp}$  de la raie de référence obtenue par interpolation est comparée au domaine en samples de chaque raie détectée. Si  $s_{interp}$  appartient au domaine d'une raie détectée, alors la raie de référence  $\lambda_{ref*}$  est associée à cette raie détectée. La raie  $\lambda_{ref*}$  a été identifiée dans le spectre extrait.

Afin d'avoir une estimation plus précise (i.e. inférieure au sample) de la position du centre de chaque raie identifiée, la raie est ajustée par une parabole au voisinage du centre discret. Le profil des raies des spectres extrait du jeu de données (c.f. section 4.1) est gaussien. Afin de contraindre le moins possible la parabole sur les ailes de la raie, l'ajustement doit se faire sur un nombre de samples limité au voisinage du centre discret de la raie. Ce voisinage est décrit par trois samples : le centre discret de la raie identifiée, les samples bleu et rouge adjacents à ce centre discret.

L'estimation du centre de la raie identifiée utilise une méthode de moindres carrés. Elle consiste à minimiser le critère  $S_{ref}$  suivant :

$$S_{ref} = \sum_{s=s_{blue}}^{s_{red}} \left( flux_s - \sum_{n=0}^2 c_n s^n \right)^2$$

où  $s$  est un sample du pic de la raie,  $flux_s$  est le flux correspondant au sample  $s$ ,  $s_{blue}$  est le sample bleu adjacent au centre de la raie,  $s_{red}$  est le sample rouge adjacent au centre de la raie, et  $c_n$  les coefficients du polynôme d'ordre deux à estimer.

### 4.3. Localisation de raies de référence par ajustement parabolique

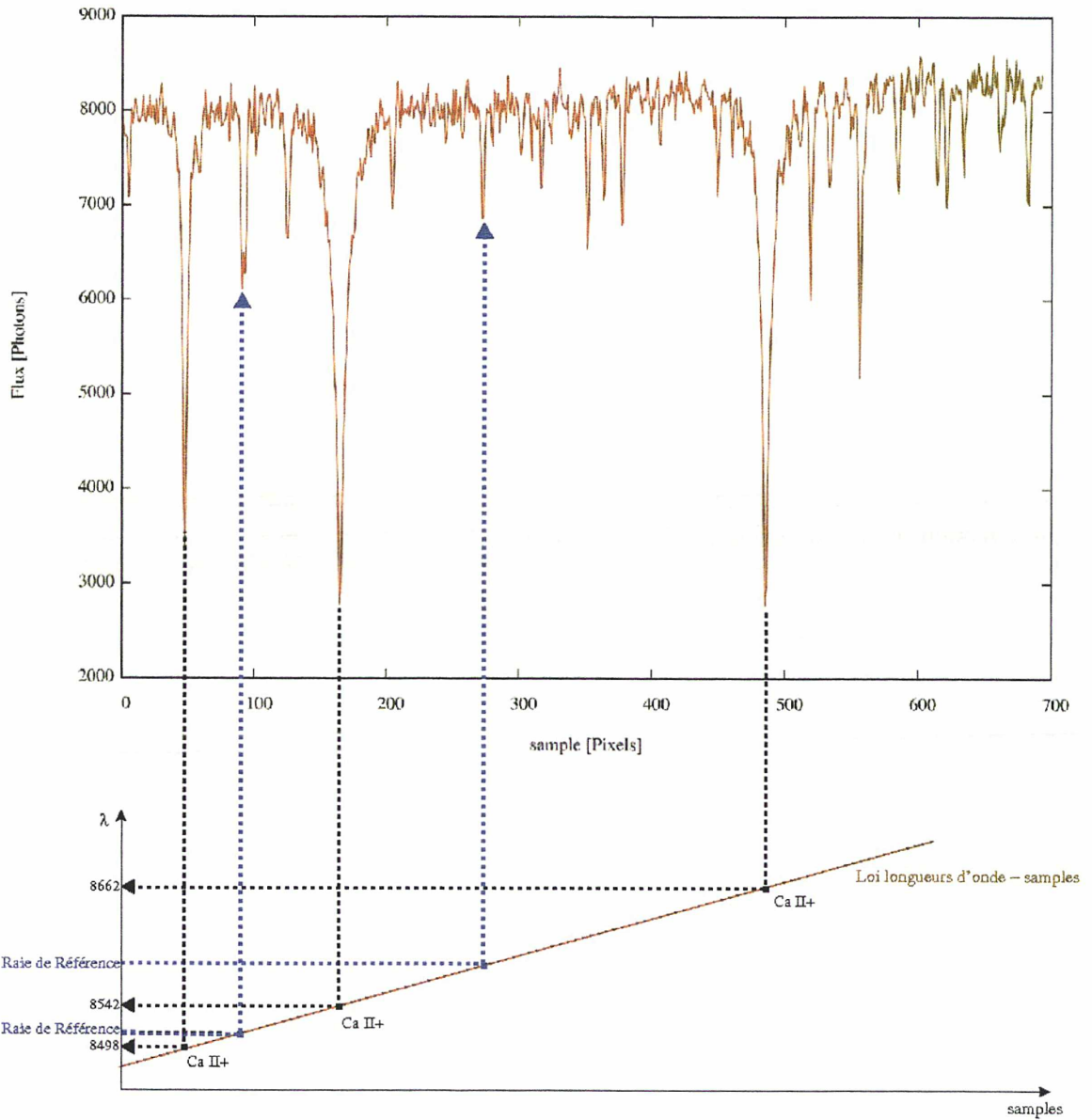


FIG. 4.5 – Méthode d'identification des raies de référence. Les flèches noires représentent la localisation du triplet du Calcium permettant la mise en place d'une loi simple  $sample = f(\lambda)$ , en rouge. Les flèches bleues représentent la localisation de deux raies de référence en utilisant la loi longueur d'onde - sample établie.

Le minimum du critère  $S_{ref}$  est atteint en annulant son gradient. Cela revient à résoudre le système linéaire suivant :

$$\begin{bmatrix} \sum_{s=S_{blue}}^{S_{red}} 1 & \sum_{s=S_{blue}}^{S_{red}} s & \sum_{s=S_{blue}}^{S_{red}} s^2 \\ \sum_{s=S_{blue}}^{S_{red}} s & \sum_{s=S_{blue}}^{S_{red}} s^2 & \sum_{s=S_{blue}}^{S_{red}} s^3 \\ \sum_{s=S_{blue}}^{S_{red}} s^2 & \sum_{s=S_{blue}}^{S_{red}} s^3 & \sum_{s=S_{blue}}^{S_{red}} s^4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \sum_{s=S_{blue}}^{S_{red}} flux_s \\ \sum_{s=S_{blue}}^{S_{red}} flux_s s \\ \sum_{s=S_{blue}}^{S_{red}} flux_s s^2 \end{bmatrix}$$

où  $c_0, c_1, c_2$  sont les coefficients à déterminer.

Dans le prototype, la résolution de ce système linéaire utilise la méthode du pivot de Gauss. Les coefficients étant estimés, la position du minimum du polynôme d'ordre deux est atteinte lorsque  $s = \frac{-c_1}{2c_2}$ . On obtient ainsi une position précise en sample du centre de la raie.

Un couple  $(\lambda_{ref*}, s_{est})$  est ainsi formé pour chaque raie identifiée.

### Critère de rejet des estimations

Afin de ne pas prendre en compte de mauvaises estimations (dues à un mauvais ajustement d'une raie complexe), un critère de rejet a été développé et intégré dans la méthode. Ce critère consiste à mesurer l'écart entre l'estimation de la position du minimum  $s_{est}$  de chaque raie identifiée dans le spectre et le centre discret  $s_n$  de cette raie. Le centre discret, obtenu lors de la détection de la raie, est une estimation à un sample près du centre de la raie. La position du minimum, obtenue lors de l'identification des raies, est une estimation du centre de la raie inférieure au sample. Si l'estimation de la position du minimum a été correctement calculée, elle doit être proche du centre discret cette même raie. Ainsi, pour qu'une estimation soit valide, il faut que l'écart mesuré soit inférieur au demi-sample :

$$|s_{est} - s_n| \leq 0.5 \text{ sample}$$

Si ce critère n'est pas respecté, la raie identifiée dans un spectre extrait ne sera pas utilisée lors de l'étape de calcul des coefficients de calibration en longueur d'onde. Cette raie sera "rejetée".

### Mise en place du triplet $(\lambda, \eta, \zeta)$

Une fois l'identification des raies de référence et le critère de rejet passés, le couple  $(\lambda_{ref*}, s_{est})$  est transformé en triplet  $(\lambda_{ref*}, \eta, \zeta)$  pour une raie identifiée. Les coordonnées angulaires  $(\eta, \zeta)$  d'une étoile dans le champ de vue de l'instrument sont associées au temps de lecture de chaque sample du spectre extrait de l'étoile. Ainsi, la position en sample du centre de chaque raie peut être convertie en coordonnées angulaires  $(\eta, \zeta)$  dans le champ de vue de l'instrument en utilisant une interpolation linéaire.

Pour chaque raie identifiée  $\lambda_{ref*}$ , on forme un triplet  $(\lambda_{ref*}, \eta, \zeta)$  utilisé pour le calcul des coefficients  $C_{mn}$  de calibration en longueur d'onde.

Toutes les opérations citées précédemment sont réalisées sur chaque spectre extrait d'une unité de calibration donnée. La méthode d'identification proposée suppose de travailler avec des spectres où le triplet du calcium domine les autres raies du spectre, c'est-à-dire où les trois raies du Ca sont les trois raies les plus profondes. Elle est adaptée pour des spectres d'étoiles de type F, G et K.

### 4.3.2 Performance de la méthode

La méthode proposée a été testée sur un échantillon de 1000 spectres du jeu de données présenté dans la section 4.1 de ce chapitre. Ils proviennent tous d'un spectre de Kurucz d'une étoile G5V dégradé à la résolution du RVS.

Pour chaque raie de référence, chaque estimation du centre de raie a été comparée, par soustraction, à la valeur théorique obtenue en utilisant l'échelle en longueur d'onde vraie de chaque spectre. Une distribution d'erreurs sur les 1000 spectres est ainsi formée pour chaque raie de référence.

La table 4.1 dresse les statistiques des distributions d'erreurs d'estimation de la position du minimum des raies de référence identifiées dans les spectres. Chaque raie est représentée par le nom de la raie (concaténation entre le symbole de l'élément chimique et de sa position en longueur d'onde) et par sa position en longueur d'onde théorique en Angströms. Les moyennes et les dispersions sont empiriques, et exprimées en samples.

Raie	Nombre	Moyenne M	Erreur sur M	Dispersion D	Erreur sur D
Fe 8481.982	199	0.01048	0.013443	0.19011	0.009505
Ca 8498.023	1000	-0.00125	0.001764	0.05581	0.001247
Fe 8514.072	1000	0.19758	0.003900	0.12340	0.002758
Si 8526.968	999	-0.16483	0.006849	0.21659	0.004843
Ca 8542.091	1000	0.00503	0.002256	0.07138	0.001595
Si 8556.777	763	0.00591	0.006075	0.16791	0.004296
Fe 8582.257	1000	0.01277	0.003296	0.10428	0.002331
Fe 8598.830	169	-0.05027	0.015186	0.19799	0.010738
Fe 8611.804	1000	-0.01361	0.002936	0.09289	0.002076
Fe 8621.601	1000	0.00854	0.003344	0.10581	0.002365
Si 8648.465	271	0.00620	0.008669	0.14297	0.006130
Ca 8662.141	1000	-0.12011	0.001896	0.05998	0.001341
Fe 8674.746	1000	0.06618	0.003226	0.10207	0.002281
Fe 8688.626	1000	0.02024	0.002001	0.06330	0.001415
Fe 8699.454	931	-0.18333	0.005054	0.15430	0.003574
Fe 8713.198	708	-0.43374	0.006581	0.17523	0.004654
Mg 8736.019	983	-0.00481	0.003937	0.12349	0.002784

TAB. 4.1 – Statistiques en **samples** de chaque distribution d'erreurs sur l'estimation du centre des raies. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum pour une raie. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne empirique pour une raie. La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion empirique pour une raie.

La figure 4.6 est une représentation graphique de la table 4.1 superposée au spectre théorique



de l'étoile G5V. Ce spectre théorique a notamment été utilisé pour simuler les 1000 spectres observés de notre base de données. Cette superposition permet de visualiser le profil et l'entourage de chaque raie de référence.

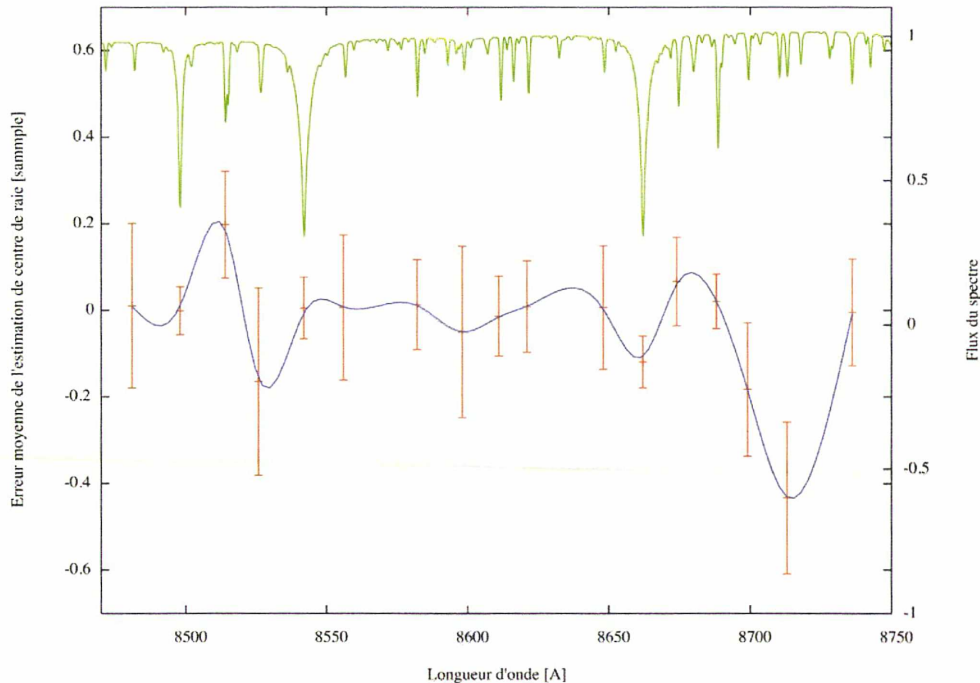


FIG. 4.6 – Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation.

En comparant les moyennes  $M$  aux erreurs sur leur estimation, on observe un biais sur l'estimation de la position du minimum de raies. Le biais est significatif (i.e.  $M > 3 \times \text{Erreur sur } M$ ) sur les raies de référence : Fe 8514.072, Si 8526.968, Ca 8662.14, Fe 8699.454 et Fe 8713.198.

Sur les spectres observés, toutes les raies sont plus ou moins blendées, i.e. chaque raie observée est une superposition de plusieurs raies. La méthode proposée associe la position en longueur d'onde du centre d'une raie donnée à la position en samples du centre de cette raie observée sur le spectre. Cependant, avec la résolution du RVS ( $R=11500$ ), la raie observée est souvent contaminée par une ou plusieurs raies autour de la raie recherchée. Chaque raie contribuant à ce blend introduit un décalage entre la position de la raie recherchée et la position du centre du blend (c.f. exemple 4.7). C'est ce décalage qui génère les biais.

La méthode de localisation faisant partie d'un processus itératif, le biais est susceptible de

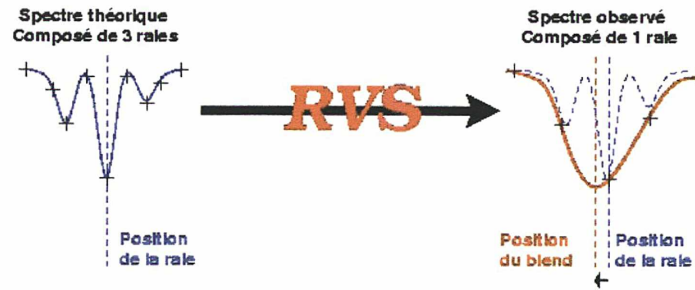


FIG. 4.7 – Illustration de l'effet de blend sur les raies observées.

se propager à la calibration en longueur d'onde, et ne peut être négligé. Par exemple, pour la raie du fer Fe 8699.454, l'erreur moyenne d'estimation est égale à environ  $-0.2$  sample. L'estimateur a donc introduit une erreur d'environ  $-2.6 \text{ km.s}^{-1}$  sur cette raie du fer. Sachant qu'on souhaite des erreurs de calibration inférieures à  $1 \text{ km.s}^{-1}$ , cette erreur a des conséquences dans le calcul des coefficients de calibration. Afin d'être utilisable dans le processus itératif de calibration en longueur d'onde, le biais sur l'estimation de la position du minimum de la raie doit être calibré. Cette étape de calibration devra supprimer toutes erreurs systématiques provenant de la méthode.

Ce processus a un temps d'exécution relativement rapide : l'estimation des positions de minimum de raies sur les 1000 spectres observés dure environ 4 minutes<sup>15</sup>. La position du maximum de chaque raie est obtenue en environ 0.014 seconde.

### 4.3.3 Calibration des biais

L'objectif de la calibration des biais est de compenser les erreurs systématiques dues à une mauvaise correspondance entre les longueurs d'onde des raies de référence et les positions du minimum des blends du spectre extrait. Le principe de cette calibration est d'appliquer la méthode d'identification par ajustement parabolique sur un jeu de données indépendant. Ce jeu de données contient des spectres dont l'échelle en longueur d'onde est connue. En faisant passer ces spectres calibrés dans le même algorithme que celui décrit précédemment (c.f. section 4.3.1), les estimations de la position du minimum de chaque raie de référence donnent directement accès à une position en longueur d'onde (via l'échelle en longueur d'onde d'un spectre du nouveau jeu de données). En moyennant les positions en longueur d'onde obtenues par raie de référence, une nouvelle liste de raies de référence est obtenue. Cette nouvelle liste permet de compenser les décalages créés à cause de la mauvaise correspondance entre la position théorique de la raie de référence et la position du centre du blend. La méthode de localisation des raies de référence par ajustement parabolique peut donc être exécutée sur les données observées en utilisant cette liste de raies "biaisées" au lieu de la liste de raies de référence (décrite dans la section 4.3.1).

Les estimations de centre discret des raies décrites par un petit nombre de samples (i.e. la somme de l'aile bleue et de l'aile rouge est inférieure à une dizaine de samples) sont très

<sup>15</sup>Ce temps de calcul a été obtenu sur un portable PowerBook G4 de 1.5 GHz avec 1.25 Go de mémoire vive.

sensibles au bruit. Par exemple, si on prend les deux mêmes raies dans deux spectres ayant un bruit différent, le centre discret d'une raie sera égal au sample  $s$ , tandis que le centre discret de l'autre raie sera  $s + 1$ . Le fait de décaler le centre discret va avoir un impact sur l'estimation de la position du minimum de la raie (i.e. le domaine d'ajustement va changer). Ainsi, si on prend un jeu de données indépendant avec un bruit différent des spectres extraits, la calibration compensera partiellement (voire pas du tout) les biais. Dans un premier test, j'ai appliqué la calibration du biais à partir d'un jeu de spectres synthétiques sans bruit. Les résultats n'ont pas été concluants : le biais a été corrigé partiellement mais reste toujours significatif sur les estimations de position du minimum de chaque raie.

Pour pouvoir calibrer le biais, il faut que les spectres du nouveau jeu de données aient le même bruit que les spectres extraits. Pour être applicable, la méthode de calibration impose d'estimer le bruit dans le spectre extrait, pour simuler les données "indépendantes".

#### 4.3.4 Performances après calibration des biais

Tous les résultats présentés ici ont été obtenus en utilisant le protocole décrit dans la section 4.3.2.

J'ai simulé un échantillon de 100 spectres synthétiques d'une étoile G5V pour mettre en place le nouveau jeu de données. Ces spectres ont été générés avec le simulateur utilisé pour mettre en place le jeu de données test (décrit dans la section 4.1). Le rapport signal à bruit de chacun de ces nouveaux spectres est comparable à ceux des données test.

La table 4.2 dresse la liste de référence avant et après la phase de calibration du biais à partir des 100 spectres synthétiques calibrés d'une étoile G5V. Dans cette table, le nom de la raie est une concaténation entre le symbole de l'élément chimique et sa position en longueur d'onde dans le référentiel du laboratoire. La différence entre les deux listes met en évidence le biais de la méthode de localisation sur chaque raie de référence.

Les positions des raies de référence pour lesquelles l'estimateur a les meilleures performances (c.f. table 4.1) n'ont pas changé. Par exemple, les deux premières raies du triplet du calcium ionisé ont les mêmes positions avant et après calibration du biais. Par contre, les positions des raies les moins performantes ont évolué. C'est le cas de la position de la dernière raie du fer de la bande RVS. Elle est passée de 8713.198 à 8713.035 Å. La position s'est donc décalée de  $-0.163$  Å, i.e.  $-5.6$  km.s<sup>-1</sup>. Or, dans la table 4.1, l'erreur moyenne commise par l'estimateur sur la position du centre de cette raie est égale à  $-0.43374$  samples, i.e. un décalage moyen<sup>16</sup> de  $-5.6$  km.s<sup>-1</sup>. Ainsi, en décalant la position de la raie du Fer, la calibration va compenser l'erreur commise par l'estimateur sur la position du centre de la raie.

La table 4.3 dresse les statistiques propres aux distributions d'erreurs d'estimation des positions du minimum des raies de référence. Chaque raie est représentée par le nom de la raie (concaténation entre le symbole de l'élément chimique et de sa position en longueur d'onde) et

---

<sup>16</sup>Un sample correspond à environ 0.375 Å. Ainsi, l'erreur en sample  $Err_s$  de la raie de référence  $\lambda_{ref}$  correspond à une vitesse de  $\frac{Err_s \times 0.375}{\lambda_{ref}} c$  km.s<sup>-1</sup>, où  $c$  est la vitesse de la lumière dans le vide.

4.3. Localisation de raies de référence par ajustement parabolique

<b>Raie</b>	<b>Position en longueur d'onde Sans calibration</b>	<b>Position en longueur d'onde Avec calibration</b>	<b>Différence (Avec - Sans)</b>
Fe8481	8481.982	8481.971	-0.011
Ca8498	8498.023	8498.025	0.002
Fe8514	8514.072	8514.146	0.074
Si8526	8526.968	8526.902	-0.066
Ca8542	8542.091	8542.091	0.000
Si8556	8556.777	8556.777	0.000
Fe8582	8582.257	8582.266	0.009
Fe8598	8598.830	8598.826	-0.004
Fe8611	8611.804	8611.793	-0.011
Fe8621	8621.601	8621.597	-0.004
Si8648	8648.465	8648.462	-0.003
Ca8662	8662.141	8662.096	-0.045
Fe8674	8674.746	8674.763	0.026
Fe8688	8688.626	8688.633	0.007
Fe8699	8699.454	8699.383	-0.071
Fe8713	8713.198	8713.035	-0.163
Mg8736	8736.019	8736.022	0.003

TAB. 4.2 – Liste de raies de référence (en Å) pour une étoile G5V, avant et après la phase de calibration du biais du centroiding. Le nom de la raie est une concaténation entre le symbole de l'élément chimique et de sa position en longueur d'onde dans le référentiel du laboratoire. La différence entre les deux listes (en Å) montre la compensation du biais après la phase de calibration.

par sa position en longueur d'onde théorique en Angströms. Les moyennes et les dispersions sont empiriques, et exprimées en samples.

Raie	Nombre	Moyenne M	Erreur sur M	Dispersion D	Erreur sur D
Fe 8481.982	199	0.03799	0.013443	0.19011	0.009505
Ca 8498.023	1000	-0.00720	0.001764	0.05581	0.001247
Fe 8514.072	1000	-0.00006	0.003900	0.12340	0.002758
Si 8526.968	999	0.01008	0.006849	0.21659	0.004843
Ca 8542.091	1000	0.00483	0.002256	0.07138	0.001595
Si 8556.777	763	0.00399	0.006075	0.16791	0.004296
Fe 8582.257	1000	-0.01143	0.003296	0.10428	0.002331
Fe 8598.830	169	-0.03950	0.015186	0.19799	0.010738
Fe 8611.804	1000	0.01615	0.002936	0.09289	0.002076
Fe 8621.601	1000	0.01829	0.003344	0.10581	0.002365
Si 8648.465	271	0.01479	0.008669	0.14297	0.006130
Ca 8662.141	1000	-0.0005	0.001896	0.05998	0.001341
Fe 8674.746	1000	0.02103	0.003226	0.10207	0.002281
Fe 8688.626	1000	0.00136	0.002001	0.06330	0.001415
Fe 8699.454	931	0.00449	0.005054	0.15430	0.003574
Fe 8713.198	708	-0.0006	0.006581	0.17523	0.004654
Mg 8736.019	983	-0.01229	0.003937	0.12349	0.002784

Tab. 4.3 – Statistiques en **samples** de chaque distribution d'erreurs sur l'estimation du centre des raies, après calibration du biais. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne. La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion D.

La figure 4.8 est une représentation graphique de la table 4.3 superposée au spectre théorique de l'étoile G5V.

Les statistiques d'erreurs sur l'estimation calibrée montrent que la calibration du biais a permis de "recentrer" les distributions d'erreur autour de  $0 \text{ km.s}^{-1}$ , tout en conservant leurs dispersions calculées pour l'estimateur non-calibré et leur nombre d'estimations. Ces résultats montrent ainsi une compensation partielle des erreurs systématiques de l'estimateur non-calibré. Pour les raies Mg8736.019, Fe8674.746, Fe8621.601, Fe8598.830 et Fe8481.982, un biais significatif persiste sur l'estimation des positions des minima.

Ce processus a un temps d'exécution relativement succinct. La création de la liste à partir de 100 spectres synthétiques dure moins d'une minute. Elle est exécutée une fois. L'estimation des positions des centres de raies sur les 1000 spectres observés dure environ 4 minutes. Le processus durant 5 minutes, la position du maximum de chaque raie est obtenue en environ 0.02 seconde.

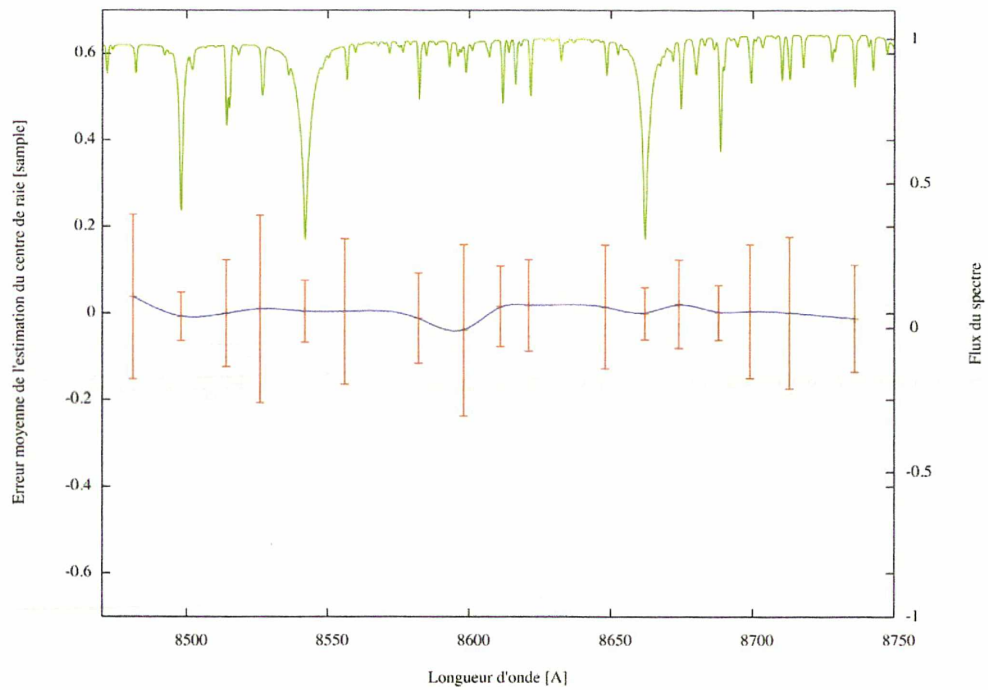


FIG. 4.8 – Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation.

## 4.4 Localisation des raies de référence par ajustement du profil des raies

La méthode présentée dans cette section a pour objectif de déterminer les longueurs d'onde des centres discrets des raies de référence. Le principe de cette méthode est d'ajuster le profil des raies détectées aux profils théoriques (de référence) de ces mêmes raies. Pour réaliser l'ajustement, la méthode calibre chaque raie détectée avec plusieurs lois linéaires de type  $\lambda = k \times s + \delta\lambda$ , où  $s$  est un sample,  $\lambda$  la longueur d'onde de  $s$ ,  $k$  un facteur d'échantillonnage et  $\delta\lambda$  un décalage en longueur d'onde. En utilisant une méthode de corrélation croisée sur deux dimensions (i.e.  $k$  et  $\delta\lambda$ ), la méthode calcule les paramètres  $k$  et  $\delta\lambda$  optimum correspondant au meilleur ajustement entre le profil de la raie et son profil théorique.

La méthode de localisation est composée de 3 étapes :

1. Double corrélation croisée entre le profil d'une raie détectée et son profil théorique.
2. Estimation du maximum de la surface de double corrélation,
3. Application de la loi de calibration locale au centre discret de la raie.

Dans le cadre d'une étude comparative entre les deux méthodes de localisation, les estimations de longueur d'onde ont été réalisées sur les mêmes raies de référence que celles utilisées dans la première méthode de localisation.

### 4.4.1 Principe de la double corrélation croisée

Afin d'estimer une loi de calibration en longueur d'onde locale pour la raie détectée, l'estimation utilise une méthode de corrélation croisée pour chaque raie de référence (méthode analogue à celle qui est utilisée pour les vitesses radiales, c.f. 3.2.4). Le principe de cette corrélation est de modifier pas à pas la loi de calibration en longueur d'onde de la raie détectée (i.e.  $\lambda = k \times s + \delta\lambda$ ), et de mesurer à chaque modification la correspondance entre la raie calibrée et un template approprié. Pour chaque spectre observé, l'estimateur réalise les étapes suivantes :

1. La dernière loi de calibration en longueur d'onde  $(k_0, \delta\lambda_0)$  est lue dans la base de données. En supposant que les propriétés instrumentales n'ont pas eu de fortes variations entre le dernier processus de calibration et l'exécution de cet estimateur, la nouvelle loi de calibration est assez proche de la dernière loi de calibration. Durant la double corrélation croisée, les paramètres  $k$  et  $\delta\lambda$  seront modifiés dans un domaine autour de la dernière loi de calibration.
2. À partir des propriétés atmosphériques de l'étoile observée, un template est sélectionné. Ce template sera l'élément de comparaison dans la double corrélation croisée.
3. Pour chaque raie détectée du spectre observé, la double corrélation croisée réalise une boucle sur les paramètres  $k$  et  $\delta\lambda$  dans les domaines respectifs  $[k_0 - \delta k, k_0 + \delta k]$  et  $[\delta\lambda_0 - \Delta\delta\lambda, \delta\lambda_0 + \Delta\delta\lambda]$ , où les valeurs  $\delta k$  et  $\Delta\delta\lambda$  sont à déterminer. Dans la boucle et pour chaque raie détectée, les opérations suivantes sont réalisées :

- (a) Une loi de calibration en longueur d'onde est appliquée sur le domaine (en samples) de la raie détectée. On connaît ainsi le domaine en longueur d'onde  $[\lambda_1, \lambda_2]$  de la raie ciblée.
- (b) Connaissant les propriétés atmosphériques de l'étoile observée, une liste de raies de référence est "sélectionnée". Le domaine  $[\lambda_1, \lambda_2]$  est comparé aux positions en longueur d'onde de chaque raie de référence : si une raie de référence est comprise dans le domaine  $[\lambda_1, \lambda_2]$ , alors la raie détectée est une raie de référence. Seules les raies de référence sont prises en compte dans la suite de l'estimateur.
- (c) Un coefficient de corrélation est calculé entre la raie de référence détectée et le domaine  $[\lambda_1, \lambda_2]$  du template en utilisant la formule suivante :

$$\rho_{raie,ref}(k, \delta\lambda) = \frac{\sum_{\lambda=\lambda_1(k, \delta\lambda)}^{\lambda_2(k, \delta\lambda)} (F_{raie}(\lambda) - \overline{F_{raie}}) (F_{ref}(\lambda) - \overline{F_{ref}})}{\sqrt{\sum_{\lambda=\lambda_1(k, \delta\lambda)}^{\lambda_2(k, \delta\lambda)} (F_{raie}(\lambda) - \overline{F_{raie}})^2 \sum_{\lambda=\lambda_1(k, \delta\lambda)}^{\lambda_2(k, \delta\lambda)} (F_{ref}(\lambda) - \overline{F_{ref}})^2}}$$

où  $\rho_{raie,ref}(k, \delta\lambda)$  est le coefficient de corrélation entre la raie détectée calibrée avec la calibration  $k, \delta\lambda$  et le spectre synthétique,  $\lambda_1$  et  $\lambda_2$  les bornes en longueurs d'onde de la raie calibrée,  $F_{raie}(\lambda)$  est le flux lu à la longueur d'onde  $\lambda$  dans la raie calibrée,  $\overline{F_{raie}}$  correspond au flux moyen dans la raie calibrée,  $F_{ref}(\lambda)$  est le flux lu à la longueur d'onde  $\lambda$  dans le spectre synthétique, et  $\overline{F_{ref}}$  correspond au flux moyen dans le domaine  $[\lambda_1, \lambda_2]$  du spectre synthétique.

Une fonction de corrélation  $\rho_{raie,ref}(k, \delta\lambda)$  (ou surface de corrélation) est calculée point par point.

Le maximum de cette surface correspond au meilleur ajustement entre la raie et son template. La position de la raie dans la surface de corrélation donne une estimation optimum des paramètres  $(k, \delta\lambda)$ . La recherche du maximum de la surface de corrélation est décrit plus bas, dans la section 4.4.2.

#### 4.4.2 Recherche du maximum de la surface de corrélation

L'estimation du maximum de la surface de corrélation est réalisée par un algorithme implémentant une méthode de Newton-Raphson. Un extremum est atteint lorsque le gradient de la surface de corrélation est nul. Ainsi, notre système à résoudre s'écrit de la manière suivante :

$$\begin{cases} \frac{\partial Corr(k, \delta\lambda)}{\partial k} = 0 \\ \frac{\partial Corr(k, \delta\lambda)}{\partial \delta\lambda} = 0 \end{cases} \quad (4.1)$$

où  $Corr$  est la surface de corrélation,  $k$  le facteur d'échantillonnage et  $\delta\lambda$  le décalage en longueur d'onde.

Afin de simplifier le développement du problème, on posera :

$$\begin{cases} f(k, \delta\lambda) = \frac{\partial Corr(k, \delta\lambda)}{\partial k} \\ g(k, \delta\lambda) = \frac{\partial Corr(k, \delta\lambda)}{\partial \delta\lambda} \end{cases}$$



Le système 4.1 peut ainsi s'écrire sous la forme suivante :

$$\begin{cases} f(k, \delta\lambda) = 0 \\ g(k, \delta\lambda) = 0 \end{cases}$$

En supposant que le couple  $(k_1, \delta\lambda_1)$  soit proche de la solution du système 4.1, on pose :

$$\begin{cases} f(k_1, \delta\lambda_1) = \varepsilon \\ g(k_1, \delta\lambda_1) = \varepsilon' \end{cases}$$

où  $\varepsilon$  et  $\varepsilon'$  sont des valeurs proches de zéro.

Soit  $h$  et  $h'$  les accroissements à ajouter, respectivement, aux valeurs  $k_1$  et  $\delta\lambda_1$ , afin d'atteindre la solution du problème. Par définition, le système peut alors s'écrire de la manière suivante :

$$\begin{cases} f(k_1 + h, \delta\lambda_1 + h') = 0 \\ g(k_1 + h, \delta\lambda_1 + h') = 0 \end{cases} \quad (4.2)$$

En supposant que les fonctions  $f$  et  $g$  sont dérivables en tous points, on peut faire un développement limité au premier ordre du système 4.2. Ainsi, le système s'écrit :

$$\begin{cases} f(k_1 + h, \delta\lambda_1 + h') = f(k_1, \delta\lambda_1) + h \frac{\partial f}{\partial k_1}(k_1, \delta\lambda_1) + h' \frac{\partial f}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) = 0 \\ g(k_1 + h, \delta\lambda_1 + h') = g(k_1, \delta\lambda_1) + h \frac{\partial g}{\partial k_1}(k_1, \delta\lambda_1) + h' \frac{\partial g}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) = 0 \end{cases} \quad (4.3)$$

À partir du système 4.3, on déduit l'expression des accroissements :

$$\begin{cases} h = \frac{\frac{\partial f}{\partial \delta\lambda_1}(k_1, \delta\lambda_1)g(k_1, \delta\lambda_1) - \frac{\partial g}{\partial \delta\lambda_1}(k_1, \delta\lambda_1)f(k_1, \delta\lambda_1)}{\frac{\partial f}{\partial k_1}(k_1, \delta\lambda_1) \frac{\partial g}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) - \frac{\partial f}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) \frac{\partial g}{\partial k_1}(k_1, \delta\lambda_1)} \\ h' = \frac{\frac{\partial f}{\partial k_1}(k_1, \delta\lambda_1)g(k_1, \delta\lambda_1) - \frac{\partial g}{\partial k_1}(k_1, \delta\lambda_1)f(k_1, \delta\lambda_1)}{\frac{\partial f}{\partial k_1}(k_1, \delta\lambda_1) \frac{\partial g}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) - \frac{\partial f}{\partial \delta\lambda_1}(k_1, \delta\lambda_1) \frac{\partial g}{\partial k_1}(k_1, \delta\lambda_1)} \end{cases}$$

La recherche du point maximum de la surface de corrélation se fait via un processus itératif :

$$\begin{cases} k_{n+1} = k_n + h \\ \delta\lambda_{n+1} = \delta\lambda_n + h' \end{cases}$$

Connaissant les valeurs des accroissements  $h$  et  $h'$ , on est capable de déterminer itérativement les valeurs optimales de  $k$  et  $\delta\lambda$  pour construire une loi de calibration à l'échelle d'une raie donnée. Afin que la méthode ne converge pas vers un maximum local dû à la présence d'une autre raie dans le voisinage de la raie de référence, le point de départ  $(k_1, \delta\lambda_1)$  doit être suffisamment proche de la solution du système. La convergence de la méthode est atteinte en moins de trois itérations.

### 4.4.3 Optimisation de la méthode

Le processus de double corrélation croisée est coûteux en temps de calcul (i.e. environ 120 minutes de temps de calcul en utilisant une grille de  $40 \times 40$  points sur un jeu de données de 1000 spectres). Afin d'optimiser l'algorithme, une grille d'échantillonnage adaptative (deux grilles successives de  $N \times N$ ) est préférée à une grille statique (une grille de taille  $2N \times 2N$ ). Le principe de la grille adaptative est de réduire successivement la région à considérer autour du maximum discret de la surface de corrélation. Cette modification permet d'obtenir plus rapidement une estimation précise de la position du maximum de la surface.

En utilisant le principe de grille adaptative, la surface de corrélation sur laquelle le maximum sera estimé est déterminé en deux itérations :

1. Une première surface de corrélation est calculée sur un domaine large en  $k$  et  $\delta\lambda$ . Cette grille permet d'estimer une première position du maximum de la surface de corrélation. Ce maximum est établi de manière discrète : il correspond au coefficient de corrélation le plus élevé lu sur la surface de corrélation. L'estimation du maximum est précise à un pas de grille près.
2. Une deuxième surface de corrélation est calculée sur un domaine plus étroit (2 pas de la première grille) centré sur le maximum discret. La position du maximum est ensuite estimée avec plus de précision (inférieure au pas de la nouvelle grille) sur cette nouvelle surface.

La figure 4.9 illustre le principe d'une grille adaptative dans le cadre de la double corrélation croisée.

Cette optimisation a deux avantages : réduire le nombre d'opérations et choisir un domaine de corrélation initial plus large sans affecter la précision. Par exemple, soit  $T_{corr}$ , le temps mis par le CPU pour calculer un coefficient de corrélation entre une raie détectée et son template associé. Dans le cadre d'une double corrélation croisée sur une grille statique de  $40 \times 40$  points<sup>17</sup>, le temps de calcul est de l'ordre de  $1600T_{corr}$  (i.e.  $(40 \times 40)T_{corr}$ ). Dans le cadre d'une double corrélation croisée sur une grille adaptative de  $20 \times 20$  points, le temps de calcul est de l'ordre de  $800T_{corr}$  (i.e.  $(20 \times 20 + 20 \times 20)T_{corr}$  car deux grilles successives sont utilisées). Ainsi, pour la même précision de calcul, on gagne un facteur deux en temps de calcul.

Une seconde optimisation a été apportée au niveau des raies détectées. Pour calculer une loi de calibration en longueur d'onde locale à chaque raie, on utilise un repère global. En effet, on utilise une échelle de sample propre au spectre pour calibrer une raie de ce spectre décrit par une dizaine de samples. La méthode de double corrélation croisée a besoin d'une première loi de calibration (c.f. section 4.4.1). À partir de ce premier jeu de coefficients ( $k, \Delta\lambda$ ), on va définir le domaine sur lequel on va corréler le profil de la raie détectée à son profil théorique. En considérant que le paramètre  $k$  a une erreur de 1%, cette erreur se propage de sample à sample dans l'échelle globale du spectre. Pour une raie centrée sur le sample 10, le domaine de corrélation va être décalé de 0.1 sample. Pour une raie en fin de spectre, centrée sur le sample 1000, le domaine de corrélation va être décalé de 10 samples (décalage de l'ordre d'une raie). En décalant le domaine de corrélation, on augmente les risques de prendre en compte une autre raie que celle qui est ciblée. Ainsi, on risque de décaler l'estimation du maximum de la surface

<sup>17</sup>une surface de  $40 \times 40$  points est une surface définie sur 40 valeurs de  $\delta\lambda$  et 40 valeurs de  $k$ .

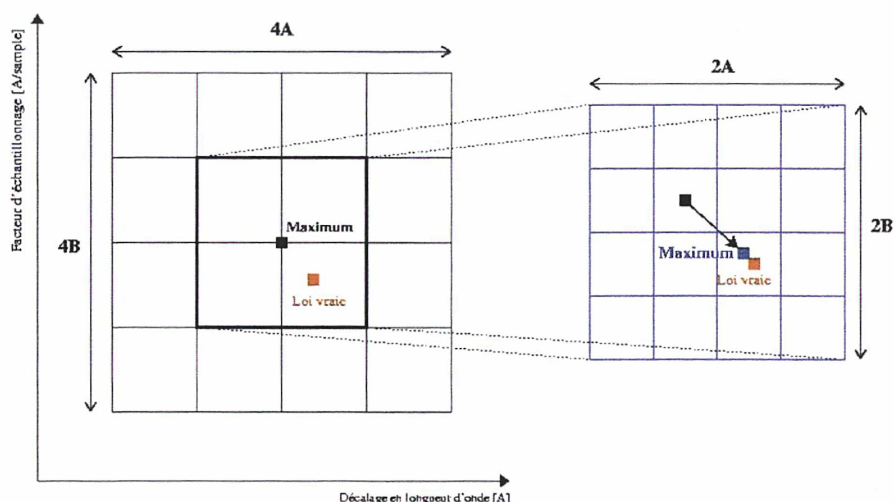


FIG. 4.9 – Principe de la grille adaptative appliquée à la surface de corrélation. À gauche. Le carré rouge montre la position du maximum théorique de la surface de corrélation sur la grille d'échantillonnage. Le carré noir montre la position du maximum discret lu sur la première surface de corrélation. Le cadre noir représente la region d'intérêt sélectionnée autour de la position du maximum discret. Il encadre le domaine  $[2A, 2B]$  sur lequel une nouvelle surface de corrélation va être calculée. À droite. Le carré rouge montre la position du maximum théorique de l'ancienne surface de corrélation (i.e. surface représentée à gauche). Le carré noir montre la position du maximum discret lu sur l'ancienne surface de corrélation. Et enfin, le carré bleu montre la position de l'estimation du maximum de la surface de corrélation sur cette nouvelle grille.

de corrélation. Afin de limiter les risques de fausses estimations, un changement de repère doit être appliqué sur l'abscisse en samples de chaque raie détectée. Ce changement de repère va nous permettre d'établir une loi de calibration locale avec une échelle de sample locale.

Soit  $s'$  la nouvelle position en sample d'un sample  $s$ .  $s'$  est définie de telle sorte que la position discrète  $s_0$  du centre de la raie soit l'origine du nouveau repère : la formule du changement d'abscisse est donnée par  $s' = s - s_0$ . La figure 4.10 illustre le changement de repère opéré sur les raies.

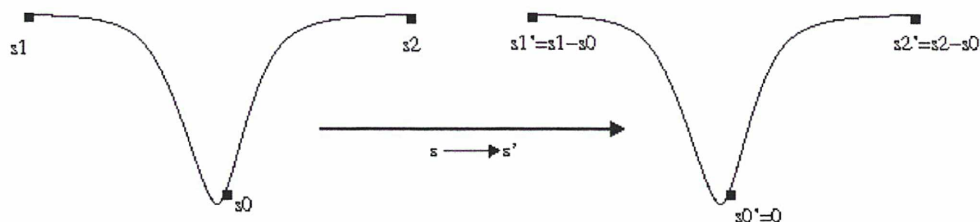


FIG. 4.10 – Illustration du changement de repère appliqué sur chaque raie détectée au niveau d'un spectre observé. À gauche : une raie détectée dans le domaine  $[s_1, s_2]$ . À droite : cette même raie représentée dans le nouveau repère  $[s'_1, s'_2]$ .

En remplaçant la valeur de  $s$  par  $s' + s_0$  dans la formule de la loi de calibration ( $\lambda = k \times s + \delta\lambda$ ),

la loi de calibration en longueur d'onde peut s'écrire :

$$\begin{aligned}\lambda &= k \times s' + k \times s_0 + \delta\lambda \\ &= k \times s' + \lambda_0\end{aligned}$$

où  $\lambda_0$  correspond à la position en longueur d'onde associée à la position en sample  $s_0$  du centre de la raie détectée (i.e.  $\lambda_0 = k \times s_0 + \delta\lambda$ ).

En recherchant le maximum de la surface de corrélation, on accède directement à une estimation de  $\lambda_0$ , et par conséquent, à une association sample-longueur d'onde ( $s_0, \lambda_0$ ).

En utilisant ce changement de repère, on va diminuer l'influence du paramètre  $k$  sur la double corrélation. En faisant varier le  $k$  utilisé dans la loi de calibration, la position  $\delta\lambda$  du maximum discret (i.e. couple  $(k, \delta\lambda)$  pour lequel la corrélation est la plus forte sur la surface) varie très peu. Par exemple, en faisant varier le facteur d'échantillonnage  $k$  de  $+/- 5\%$  (ancien design du RVS où le paramètre  $k = 0.375 \text{ \AA}$ ), on peut regarder la variation de la position du maximum discret sur la surface de corrélation du profil symétrique d'une raie du calcium et du profil non-symétrique d'une raie du fer. Dans le cas de la raie du calcium, le point zéro  $\delta\lambda$  du maximum discret reste égal à  $8542.091 \text{ \AA}$ . L'estimation de la position du maximum varie de  $0.0 +/- 0.02 \text{ \AA}$ . Dans le cas de la raie du fer, le point zéro  $\delta\lambda$  du maximum discret passe de  $8514.279 \text{ \AA}$  à  $8514.319 \text{ \AA}$ . L'estimation de la position du maximum varie de  $0.04 +/- 0.02 \text{ \AA}$ . Cette variation est peu significative. Par conséquent, l'estimation de la longueur d'onde  $\lambda_0$  du centre discret d'une raie est peu sensible aux variations du facteur d'échantillonnage  $k$ . Cette sensibilité est corrélée avec la complexité du profil de la raie étudiée : plus la raie est blendée, plus une variation de  $k$  aura un impact sur l'estimation de  $\lambda_0$ .

La figure 4.11 montre les surfaces de corrélation obtenues après une double corrélation croisée d'une raie du calcium. La méthode de double corrélation utilise les deux optimisations proposées. Ainsi, deux surfaces de corrélation sont calculées : une surface large (en rouge sur la figure) nous permettant de localiser la position discrète du maximum de la surface, et une seconde surface plus petite (en vert sur la figure) sur laquelle la position du maximum est estimée avec précision. Les surfaces montrent aussi une faible dépendance des coefficients de corrélation au paramètre  $k$  (changement de repère effectué sur l'échelle des samples de la raie).

#### 4.4.4 Performance de l'ajustement de profil

Tous les résultats présentés ci-dessous ont été obtenus en utilisant le même protocole que celui qui a été décrit section 4.3.2.

La table 4.4 dresse les statistiques propres aux distributions d'erreurs d'estimation de position du minimum des raies de référence. Chaque raie est représentée par le nom de la raie (concaténation entre le symbole de l'élément chimique et de sa position en longueur d'onde) et par sa position en longueur d'onde théorique en Angströms. Les moyennes et les dispersions sont empiriques, et exprimées en samples.

La figure 4.12 est une représentation graphique de la table 4.4 superposée au spectre théorique de l'étoile G5V. Cette superposition permet de visualiser le profil et l'entourage de chaque raie

Raie	Nombre	Moyenne M	Erreur sur M	Dispersion D	Erreur sur D
Fe 8481.982	380	-0.00778	0.008572	0.16731	0.006061
Ca 8498.023	991	0.00205	0.001172	0.03692	0.000829
Fe 8514.072	989	-0.00987	0.002085	0.06560	0.001474
Si 8526.968	932	0.00089	0.003546	0.10831	0.002508
Ca 8542.091	985	0.00182	0.001314	0.04127	0.000929
Si 8556.777	815	-0.00460	0.004888	0.13964	0.003457
Fe 8582.257	942	-0.00186	0.002786	0.08556	0.001970
Fe 8598.830	343	-0.01916	0.008975	0.16646	0.006346
Fe 8611.804	957	0.00066	0.002449	0.07580	0.001732
Fe 8621.601	948	-0.00154	0.002739	0.08438	0.001937
Si 8648.465	496	0.00521	0.005603	0.12491	0.003962
Ca 8662.141	984	-0.00191	0.001245	0.03908	0.000881
Fe 8674.746	948	0.00173	0.002471	0.07612	0.001747
Fe 8688.626	987	0.00327	0.001447	0.04547	0.001023
Fe 8699.454	867	-0.00361	0.003775	0.11121	0.002669
Fe 8713.198	670	-0.00788	0.003816	0.09885	0.002699
Mg 8736.019	917	-0.00758	0.003280	0.09937	0.002319

TAB. 4.4 – Statistiques en **samples** sur chaque distribution d'erreurs sur l'estimation du centre des raies. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne. La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion D.

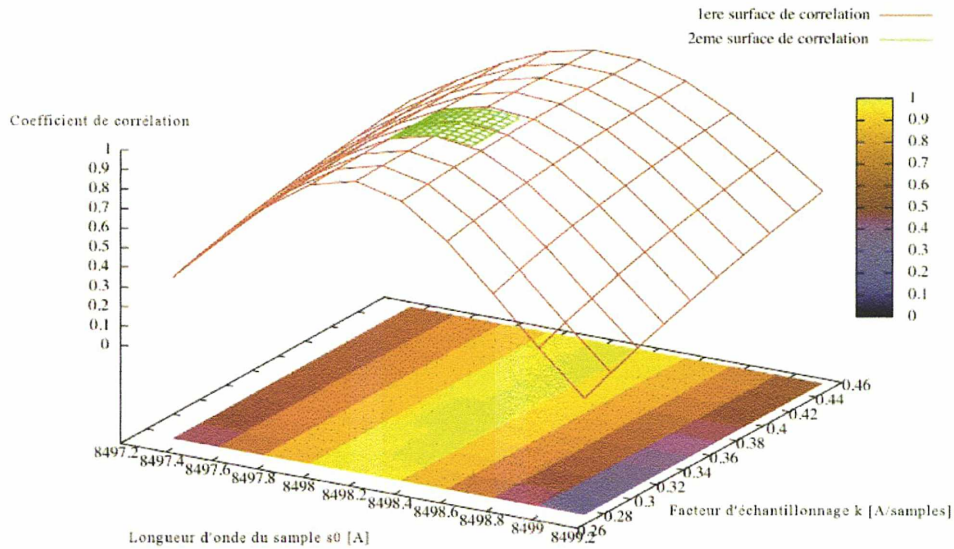


FIG. 4.11 – Surfaces de corrélation pour une raie du Calcium, en utilisant un changement de repère et une grille adaptative : exemple d'une des raies du calcium ionisée deux fois. Évolution du coefficient de corrélation en fonction de  $k$  (i.e. échantillonnage) et  $\delta\lambda'$  (i.e. décalage en longueur d'onde dans le nouveau repère). *En rouge*, première surface de corrélation sur un large domaine. *En vert*, seconde surface de corrélation à partir de laquelle la loi de calibration optimale est estimée. L'échelle de couleur est associée à la valeur du coefficient de corrélation. Elle permet de localiser les plus fortes corrélations dans le plan  $[\delta\lambda', k]$ .

de référence.

Les résultats obtenus sont de l'ordre du millièème de sample avec des dispersions de l'ordre du dixième de sample. Aucun biais significatif n'est présent sur l'estimation des positions du minimum des raies de référence (i.e.  $M < 3 \times \text{Erreur sur } M$  pour chaque raie).

Ce processus a un temps d'exécution variant en fonction de la taille de la grille choisie. Pour des grilles de  $20 \times 20$  points, le processus dure environ 80 minutes pour 1000 spectres contenant 17 raies de référence : la position en longueur d'onde du centre discret de chaque raie est obtenu en environ 0.28 seconde. Pour des grilles de  $10 \times 10$  points, le processus dure environ 20 minutes pour 1000 spectres : la position du centre discret de chaque raie est obtenu en environ 0.07 seconde. Malgré une optimisation, la méthode par ajustement de profil prend quatre fois plus de temps que la localisation des raies de référence par ajustement parabolique.

## 4.5 Localisation accélérée de raies de référence par ajustement de profil

Comme je l'ai décrit dans la section 3.3 du chapitre 3, le prototype de SGIS sera intégré dans le processus semestriel du traitement des données spectroscopiques. Les calibrations en

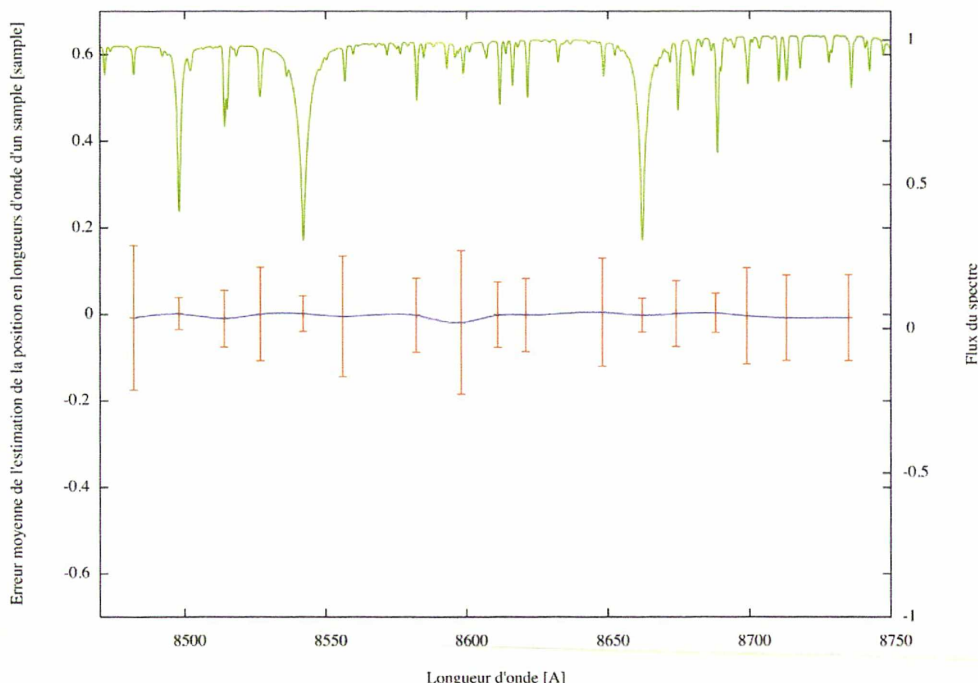


FIG. 4.12 – Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation.

longueur d'onde se feront à partir d'un grand volume de données. Au même titre que tous les modules du prototype, le module de localisation va donc traiter ce grand volume de données. Au vu des temps de calcul nécessaire pour les deux méthodes décrites précédemment (e.g. en utilisant un ajustement parabolique, il faut 20 minutes pour traiter 1000 spectres sur un portable PowerBook G4 de 1.5 GHz avec 1.25 Go de mémoire vive), le module de localisation mettrait plus de 48h pour traiter 150000 spectres<sup>18</sup>. Une nouvelle méthode de localisation a donc été développée afin d'accélérer le processus tout en ayant des performances similaires : c'est une amélioration de la méthode par ajustement de profil.

La méthode par ajustement de profil passe 80% de son temps de calcul dans la double corrélation croisée. Par ailleurs, suite au changement de repère décrit dans la section 4.4.3, la surface de corrélation entre le profil d'une raie identifiée et son profil théorique est très peu sensible aux variations du facteur d'échantillonnage  $k$ . En utilisant de ces deux faits, j'ai développé une

<sup>18</sup>Ce nombre de spectres correspond à environ 6 mois de données à traiter dans le cadre de la calibration en longueur d'onde du RVS. Plus de détails concernant les 6 mois de données seront présentés dans le chapitre 6 de ma thèse.

méthode de localisation accélérée. Elle consiste à estimer un facteur d'échantillonnage  $k$  de la loi de calibration locale d'une manière moins précise (décrite plus bas), tout en déterminant le point zéro  $\lambda_0$  en utilisant une corrélation croisée mono-dimensionnelle. Ainsi, au lieu de calculer  $20 \times 20$  coefficients de corrélation, on se limite à en calculer uniquement 20. L'estimation moins précise de  $k$  n'a pas beaucoup d'impact. En effet, en supposant que le processus itératif de calibration converge, les lois de calibration en longueur d'onde vont être de plus en plus précises au fur et à mesure des itérations. L'estimation de  $k$  étant directement liée à ces lois de calibration, les estimations de position en longueur d'onde du centre discret des raies de référence vont s'améliorer d'itération en itération. Ainsi, les triplets  $(\lambda, \eta, \zeta)$  nécessaires à l'estimation des lois de calibration vont être de plus en plus précis. En plus d'être plus rapide, cette nouvelle méthode permet d'accélérer la convergence du processus itératif de calibration en longueur d'onde.

La corrélation croisée se compose de 3 étapes successives :

1. La dernière échelle en longueur d'onde d'un spectre est utilisée pour calculer la première loi de calibration  $(k_0, \lambda_0)$ .  $k_0$  est initialisé à la valeur du pas en longueur d'onde moyen dans le spectre calibré.  $\lambda_0$  est calculée en appliquant la loi de calibration locale sur le centre discret  $s_0$  :  $\lambda_0 = k_0 \times s_0 + \delta\lambda$ , où  $\delta\lambda$  est la longueur d'onde associée au premier sample du spectre calibré. Durant la corrélation croisée, le paramètre  $\delta\lambda$  sera modifié dans un domaine autour de  $\lambda_0$ . Le paramètre  $k_0$  est constant durant tout le processus de corrélation croisée.
2. À partir des propriétés atmosphériques de l'étoile observée, un template est sélectionné. Ce template sera l'élément de comparaison dans la double corrélation croisée.
3. Pour chaque raie détectée du spectre observé, la corrélation croisée réalise une boucle sur le paramètre  $\delta\lambda$  dans les domaines  $[\delta\lambda_0 - \Delta\delta\lambda, \delta\lambda_0 + \Delta\delta\lambda]$ , où le  $\Delta\delta\lambda$  sont à déterminer. Dans la boucle, les opérations suivantes sont réalisées sur chaque raie détectée dans le spectre :
  - (a) Une loi de calibration en longueur d'onde est appliquée sur le domaine en samples de la raie détectée. On connaît ainsi le domaine en longueur d'onde  $[\lambda_1, \lambda_2]$  de la raie ciblée.
  - (b) Connaissant les propriétés atmosphériques de l'étoile observée, une liste de raies de référence est "sélectionnée". Le domaine  $[\lambda_1, \lambda_2]$  est comparé aux positions en longueur d'onde de chaque raie de référence : si la raie détectée est l'une des raies de référence, alors sa position en longueur d'onde est comprise dans l'intervalle  $[\lambda_1, \lambda_2]$ . Seules les raies de référence sont prises en compte dans la suite de l'estimateur.
  - (c) Un coefficient de corrélation est calculé entre la raie de référence détectée sur l'intervalle  $[\lambda_1, \lambda_2]$  et son template en utilisant une formule similaire à celle présentée dans la section 4.4.1 de ce chapitre.

Une fonction de corrélation  $\rho_{raie,ref}(\delta\lambda)$  mono-dimensionnelle est calculée point par point. Le maximum de cette fonction est estimé en utilisant un ajustement parabolique. La méthode utilisée est similaire à celle qui est utilisée lors du module d'estimation des vitesses radiales des spectres extraits (c.f. section 3.2.4 dans le chapitre 3).



Les performances sont similaires à celles de la méthode de localisation de raie par ajustement de profil. Les temps de calcul sont améliorés d'environ un facteur trois par rapport à la méthode de localisation par ajustement de profil.

## 4.6 Synthèse

Sur les trois méthodes de localisation de raies proposées, deux sont susceptibles d'être intégrées au prototype final de SGIS : les deux méthodes de localisation de raies par ajustement de profil. La méthode de localisation de raie par ajustement parabolique a montré des biais significatifs dans l'estimation de la position du minimum des raies de référence. Ce biais étant difficilement compensable (malgré la mise en place d'une calibration de biais), cette méthode ne peut être utilisée dans le cadre des calibrations SGIS. Le biais produit pourrait se propager d'une itération à une autre, faisant ainsi diverger tout le processus itératif de calibration. Contrairement à la première méthode, les performances des méthodes de localisation de raie par ajustement de profil ont montré des bons résultats. En effet, ces méthodes conduisent à des erreurs de l'ordre du millièème de sample sur l'estimation des couples  $(\lambda, s)$ . Ces méthodes ne présentent aucun biais significatif et leurs performances sont comparables avec celles de la première méthode de localisation : plus de 80% des raies de référence sont localisées dans les spectres extraits.

# Chapitre 5

## Étude de convergence de SGIS

### Sommaire

---

<b>5.1</b>	<b>Description de l'implémentation</b> . . . . .	<b>88</b>
5.1.1	Principe de la méthode de centroïding . . . . .	88
5.1.2	Conditions initiales . . . . .	90
<b>5.2</b>	<b>Diagnostics de contrôle</b> . . . . .	<b>94</b>
5.2.1	Diagnostic sur l'échelle en longueur d'onde . . . . .	95
5.2.2	Diagnostic sur les vitesses radiales . . . . .	96
5.2.3	Diagnostics au niveau de SGIS . . . . .	97
<b>5.3</b>	<b>Centroïding avec et sans biais</b> . . . . .	<b>97</b>
5.3.1	Performance du processus itératif sans biais . . . . .	98
5.3.2	Performance de la solution itérative avec biais . . . . .	102
5.3.3	Synthèse . . . . .	105

---

La méthode de calibration étant un processus itératif, une partie de mon travail de thèse a consisté à étudier les propriétés de convergence de SGIS. Cette étude repose sur la mise en place d'une série de diagnostics intermédiaires permettant d'évaluer pas-à-pas le comportement et la convergence du processus itératif de calibration en longueur d'onde.

Une fois les diagnostics mis en place, le processus itératif a été soumis à différentes contraintes afin d'analyser son comportement et sa convergence. Ces contraintes sont listées ci-dessous :

- Variation des conditions initiales : le point de départ du processus itératif (i.e. première loi de calibration en longueur d'onde) a été dégradé.
- Variation du niveau de bruit sur le module de localisation de raie de référence : on a ajouté une erreur de plus en plus dispersée et de plus en plus biaisée) sur les estimations du triplet ( $\lambda$ ,  $\eta$ ,  $\zeta$ ). Dans la suite de la thèse, le module de localisation de raies de référence est nommé méthode de "centroïding".

Les résultats de cette étude permettent d'établir les propriétés de convergence de la solution SGIS. Cette étude permet aussi de sélectionner une méthode de centroïding compatible avec la convergence du processus SGIS.

Toute mon étude de convergence a été effectuée avec l'implémentation du prototype SGIS présentée dans la section 5.1, et avec un premier jeu de données de test présenté dans la section 4.1.

Le fruit de cette étude nous permettra d'établir des spécifications sur le développement d'un algorithme de centroïding, au coeur du processus de calibration en longueur d'onde.

## 5.1 Description de l'implémentation

L'objectif de cette étude de convergence est de valider le concept itératif des calibrations en longueur d'onde en prouvant la convergence du prototype. Par conséquent, l'implémentation du prototype de SGIS utilisée dans cette section est relativement simplifiée.

Afin de s'affranchir des problèmes pratiques liés à l'implémentation du prototype, on va injecter des erreurs de centroïding indépendamment du jeu de données test. L'implémentation du prototype utilisée intègre une méthode de centroïding "simplifié", ne tenant pas compte des différentes implémentations proposées dans le chapitre 4. La méthode simplifiée de centroïding simule les erreurs commises lors de l'étape d'estimation des triplets ( $\lambda, \eta, \zeta$ ). Le principe de cette méthode est de dégrader les vraies échelles en longueur d'onde, provenant du simulateur de données RVS. Les vraies échelles en longueur d'onde ne seront pas accessibles dans la version finale de la calibration en longueur d'onde. Elles sont uniquement utilisées dans le cadre de la validation du processus itératif SGIS pour la calibration en longueur d'onde. Cette méthode de centroïding est présentée dans la section 5.1.1.

Durant cette étude, deux types de conditions initiales ont été mis en place. Elles permettent de tester la stabilité et la convergence du processus itératif. Elles interviennent au niveau de la phase d'initialisation du processus itératif : selon le type de conditions initiales, le point de départ du processus itératif (i.e. première loi de calibration en longueur d'onde) est initialisé avec des valeurs plus ou moins éloignées de la loi de calibration vraie. Ces deux types de conditions initiales sont présentés dans la section 5.1.2.

La phase itérative de sélection des étoiles de référence pour la calibration a été simplifiée en considérant toutes les étoiles comme référence pour la calibration. En effet, pour ce prototype, tous les spectres simulés proviennent d'un unique type stellaire, une G5V avec une vitesse radiale nulle. Toutes les étoiles simulées peuvent donc être considérées comme référence pour la phase de calibration en longueur d'onde.

La dernière étape de SGIS, c'est-à-dire la correction du point zéro des vitesses radiales, n'a pas été implémentée. Ainsi, toutes les mesures et les statistiques sur les vitesses radiales sont exprimées dans un référentiel relatif aux données spectroscopiques.

### 5.1.1 Principe de la méthode de centroïding

La méthode de centroïding, développée pour réaliser cet étude, lit directement les échelles en longueur d'onde vraies de chaque spectre.

Pour chaque spectre, la méthode a accès à des listes de raies de référence. Elle va sélectionner la liste de raies correspondant au type stellaire de l'étoile observée : la liste va être choisie en fonction des paramètres atmosphériques de l'étoile observée. La position  $\lambda_{true}$  en longueur d'onde dans le référentiel du laboratoire de chacune des raies de référence est donc connue.

Chaque  $\lambda_{true}$  est décalée en longueur d'onde en utilisant les vitesses radiales estimées et vraies (i.e. données par le simulateur RVS) pour le spectre étudié. On obtient deux positions décalées en longueur d'onde, respectivement,  $\lambda_{shift}$  et  $\lambda_{shift}^0$  pour chaque raie en utilisant les formules suivantes :

$$\begin{aligned}\lambda_{shift} &= \lambda_{true} \left( 1 + \frac{Vr}{c} \right) \\ \lambda_{shift}^0 &= \lambda_{true} \left( 1 + \frac{Vr^0}{c} \right)\end{aligned}$$

où  $Vr$  est la vitesse radiale estimée,  $Vr^0$  la vitesse radiale vraie et  $c$  la vitesse de la lumière dans le vide ( $c = 299792.458 \text{ km.s}^{-1}$ ).

La longueur d'onde  $\lambda_{shift}^0$  est ensuite associée à une position en samples  $s_{true}(\lambda_{shift}^0)$ , en utilisant une interpolation linéaire sur l'échelle en longueur d'onde vraie du spectre.

À ce stade, le couple  $(\lambda_{shift}, s_{true}(\lambda_{shift}^0))$  représente le résultat d'un centroïding parfait (i.e. sans biais et sans dispersion). Dans un cas réel, ce type d'estimateur n'existe pas. Ainsi, pour simuler une méthode de centroïding réaliste, la position en samples  $s_{true}(\lambda_{shift}^0)$  est bruitée par une distribution gaussienne  $g(B, \sigma)$  de formule :

$$g(B, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(s-B)^2}{\sigma^2}}$$

où  $s$  est le sample, le biais  $B$  et la dispersion  $\sigma$  sont choisis avant chaque exécution du prototype.

On obtient la position  $s_{noise}(\lambda_{shift}^0, B, \sigma)$ .

Par interpolation linéaire, cette position  $s_{noise}(\lambda_{shift}^0, B, \sigma)$  est associée à un couple  $(\eta_{noise}, \zeta_{noise})$  de coordonnées angulaires dans le champ de vue du RVS.

Pour chaque spectre, la méthode simule les triplets  $(\lambda_{shift}, \eta_{noise}, \zeta_{noise})$ , nécessaires à la résolution linéaire de la calibration en longueur d'onde. Toutes les opérations réalisées par la méthode de centroïding simulée sont schématisées dans la figure 5.1.

En faisant varier les paramètres  $B$  et  $\sigma$  de la distribution gaussienne utilisée, la solution itérative va se comporter différemment. Dans la suite de ce chapitre, nous allons étudier le comportement du prototype suivant les valeurs de  $B$  et de  $\sigma$ . Pour une simplicité de lecture, on parlera de précision pour évoquer la dispersion  $\sigma$  d'une distribution et d'exactitude pour le biais d'une distribution.

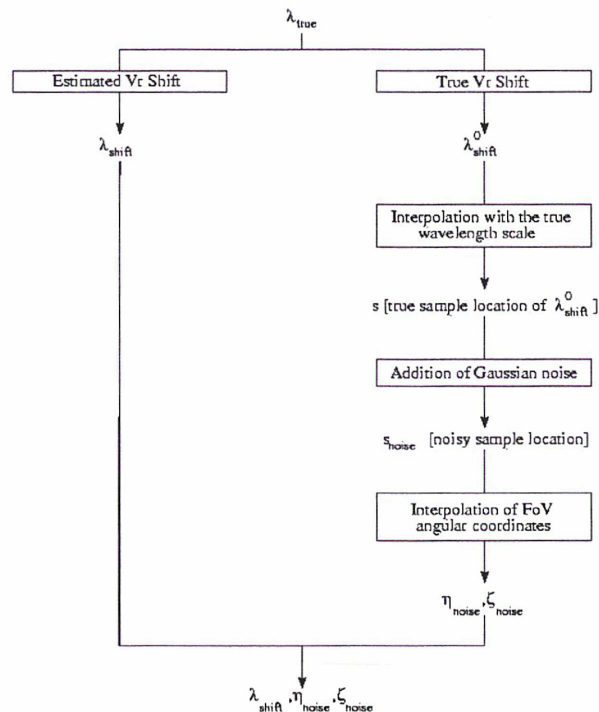


FIG. 5.1 – Principe de la simulation de la méthode de centroïding. En partant de la vraie position en longueur d'onde  $\lambda_{true}$  des raies stellaires, la méthode va brouter la position en sample  $s$  de ces raies à partir d'une distribution gaussienne d'erreurs aléatoires. À la fin du processus de simulation, la méthode forme pour chaque raie stellaire choisie, un triplet  $(\lambda_{shift}, \eta_{noise}, \zeta_{noise})$

### 5.1.2 Conditions initiales

Durant la mission Gaia, des premières lois de calibration en longueur d'onde seront obtenues en phase de calibration au sol du RVS (et données par Astrium), puis en phase de "commissioning" en vol<sup>19</sup>. Ces premières lois vont être le point de départ de notre processus itératif. Cependant, les performances de ces phases de calibration n'ont pas été chiffrées. On n'a aucune information sur le point de départ du processus itératif. On est donc contraint à explorer plusieurs cas d'initialisation. Dans notre étude des propriétés de convergence de SGIS, deux types de conditions initiales ont été intégrés dans le prototype développé :

- Conditions de non-divergence, ou mode de non-divergence du prototype : l'objectif est de tester la stabilité du prototype en partant de la solution vraie de calibration ;
- Conditions de convergence, ou mode de convergence du prototype : l'objectif est de tester la convergence du processus itératif en partant de "mauvaises" lois de calibration.

<sup>19</sup>phase de calibration en vol, avant la mise en orbite du satellite

### Mode de non-divergence

Dans le mode de non-divergence, le processus itératif est initialisé avec la vraie loi de calibration. La vraie loi de calibration est la même pour chaque unité de calibration, et est définie par la formule suivante :

$$\lambda(s) = 8430.0 + 1350.0 \times \eta(s) \quad (5.1)$$

où  $\lambda(s)$  est la longueur d'onde centrale du sample  $s$ ,  $\eta(s)$  est la coordonnée AL de l'étoile dans le champ de vue du RVS au temps de lecture du sample  $s$ .

En initialisant la loi de calibration avec ses vraies valeurs, le prototype SGIS commence le processus de calibration en longueur d'onde dans le cas idéal.

La figure 5.2 montre la moyenne de la distribution d'erreurs en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés après la phase d'initialisation du prototype SGIS. Chaque courbe de couleur représente une unité de calibration donnée. Chaque loi de calibration propre à une unité de calibration donnée a été initialisée avec les valeurs théoriques des coefficients de calibration en longueur d'onde.

Dans cette figure, on veut montrer que le processus itératif a été initialisé avec des lois de calibration en longueur d'onde identiques aux lois de calibration vraies. Pour chaque unité de calibration, l'erreur moyenne des calibrations oscille autour de 0 Å avec des amplitudes de quelques  $10^{-7}$  Å. Ces oscillations proviennent de troncatures numériques au niveau des coordonnées ( $\eta$ ,  $\zeta$ ) de la base de données. La précision numérique pour les  $\eta$  est de l'ordre  $10^{-10}$  degré. En utilisant la formule 5.1, on trouve une précision sur les longueurs d'onde de l'ordre de  $10^{-7}$  Å, qui est compatible avec les résultats du diagnostic.

En agrandissant le diagnostic au niveau de quelques oscillations, on remarque une périodicité de ces phénomènes numériques : les fluctuations, négligeables, ont une période de 9 samples autour de 0 Å. Cette périodicité de 9 samples provient de la périodicité dans les derniers chiffres des coordonnées  $\eta$  : par exemple, pour une observation donnée, la coordonnée  $\eta$  finit par les chiffres 7870 tous les 9 samples, sur l'ensemble de l'observation.

### Mode de convergence

En mode de convergence, le processus itératif est initialisé avec une "mauvaise" loi de calibration spectrale. Cette mauvaise loi mime la loi qui sera obtenue via la phase de commissioning en vol. Partant de cette loi de calibration, on veut montrer que le processus itératif converge vers une solution proche de la solution trouvée en mode de non-divergence.

Le principe du mode de convergence est donc de dégrader les lois vraies de calibration en longueur d'onde au niveau de la phase d'initialisation. Pour ce faire, cette dégradation consiste à estimer une loi de calibration en deux étapes :

1. On va générer une distribution gaussienne avec laquelle on va bruite les coordonnées ( $\eta$ ,  $\zeta$ ) d'étoiles dans le champ de vue de l'instrument RVS. On obtient de nouvelles coordonnées ( $\eta'$ ,  $\zeta'$ ).

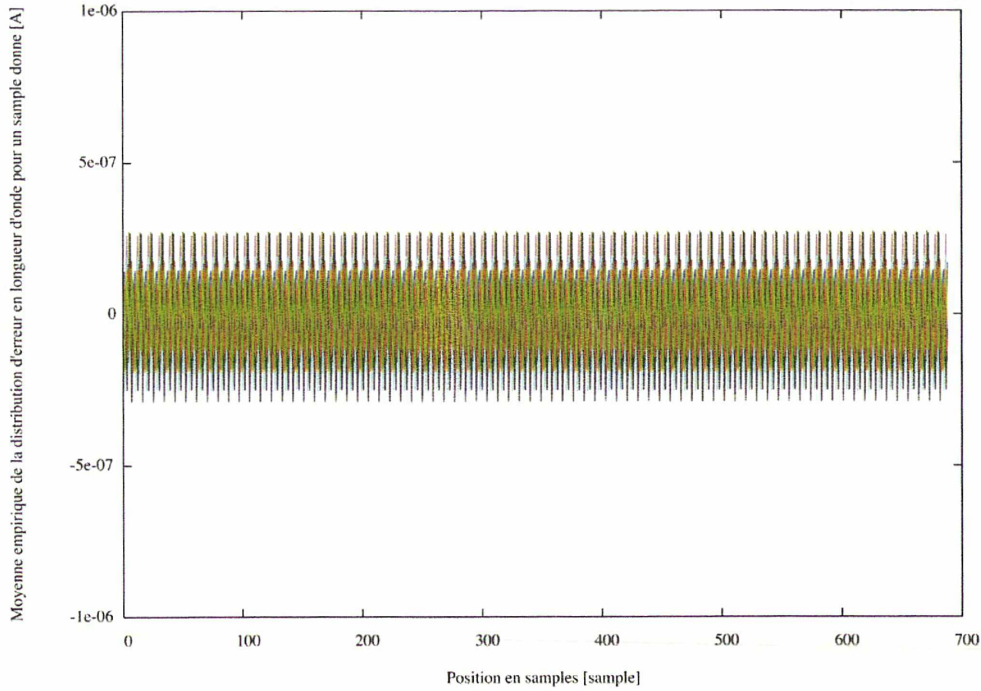


FIG. 5.2 – Ce graphique représente la moyenne de la distribution d’erreur en longueur d’onde (c.f. section 5.2.1) en fonction de la position en échantillon dans les spectres observés pour chaque unité de calibration. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs. Les différences entre les diagnostics de chaque unité de calibration ne sont pas non-visibility, ici, à cause de la similarité des résultats.

2. Les nouvelles coordonnées  $(\eta', \zeta')$  vont être utilisées pour calculer une nouvelle loi de calibration.

Soit  $\Delta_{\text{FOV}}$  l’erreur ajoutée sur les coordonnées  $(\eta, \zeta)$  d’un échantillon  $s$  donné d’un spectre observé. De nouvelles coordonnées  $(\eta', \zeta')$  du échantillon  $s$  peuvent être ainsi définies de la manière suivante :

$$\begin{cases} \eta' = \eta + \Delta_{\text{FOV}} \\ \zeta' = \zeta + \Delta'_{\text{FOV}} \end{cases} \quad (5.2)$$

L’estimation des nouvelles lois de calibration se fait via les mêmes méthodes numériques que celles qui sont utilisées dans la phase de calibration updating, décrite dans la section 3.2.6 du chapitre 3. À partir des équations 3.4 et 5.2, le critère  $S'_k$  à minimiser pour calculer les nouvelles lois de calibration devient :

$$\begin{aligned} S'_k &= \sum_{\text{spec}=1}^{N_{\text{spectres}}} \sum_{l=1}^{N_{\text{lines}}} \left[ \lambda(\text{spec}, l) - \sum_{m=0}^2 \sum_{n=0}^2 C'(k)_{mn} \eta'(\text{spec}, l)^m \zeta'(\text{spec}, l)^n \right]^2 \\ &= \sum_{\text{spec}=1}^{N_{\text{spectres}}} \sum_{l=1}^{N_{\text{lines}}} \left[ \lambda(\text{spec}, l) - \sum_{m=0}^2 \sum_{n=0}^2 C'(k)_{mn} (\eta(\text{spec}, l) + \Delta_{\text{FOV}})^m (\zeta(\text{spec}, l) + \Delta'_{\text{FOV}})^n \right]^2 \end{aligned}$$

où  $spec$  représente le spectre d'une étoile observée durant l'unité de calibration  $k$ ,  $N_{spectres}$  le nombre de spectres observés durant l'unité de calibration  $k$ ,  $l$  une raie de référence du spectre observé  $spec$ ,  $N_{lignes}$  le nombre de longueurs d'onde considérées dans le spectre  $spec$ ,  $\lambda(spec, l)$  une position en longueur d'onde de la raie  $l$  dans le spectre  $spec$ ,  $\eta'(spec, l)$  et  $\zeta'(spec, l)$  les coordonnées angulaires bruitées au temps de lecture du sample  $s$ , et  $C'(k)_{mn}$  les coefficients de calibration bruités de l'unité de calibration  $k$  à déterminer.

Pour résoudre la minimisation de  $S'_k$  (i.e. estimer les coefficients  $C'_{mn}$ ), des ensembles  $(\lambda, \eta', \zeta')$  doivent être déterminés sur les spectres de l'unité de calibration  $k$ . Dans un souci d'optimisation du temps de calcul, nous avons choisi d'utiliser un sample sur dix dans un spectre sur dix de l'unité de calibration  $k$  afin de générer les données  $(\lambda, \eta', \zeta')$ .

Le niveau de dégradation est déterminé par la dispersion de la distribution d'erreurs aléatoires : plus la dispersion est grande, plus la dégradation des lois de calibration en longueur d'onde est importante. Cette dispersion est un paramètre d'entrée de chaque prototype. Elle peut donc être modifiée avant chaque exécution de prototype.

La figure 5.3 résume les actions exécutées lors de l'initialisation des lois de calibration en longueur d'onde, sous la forme d'un diagramme fonctionnel.

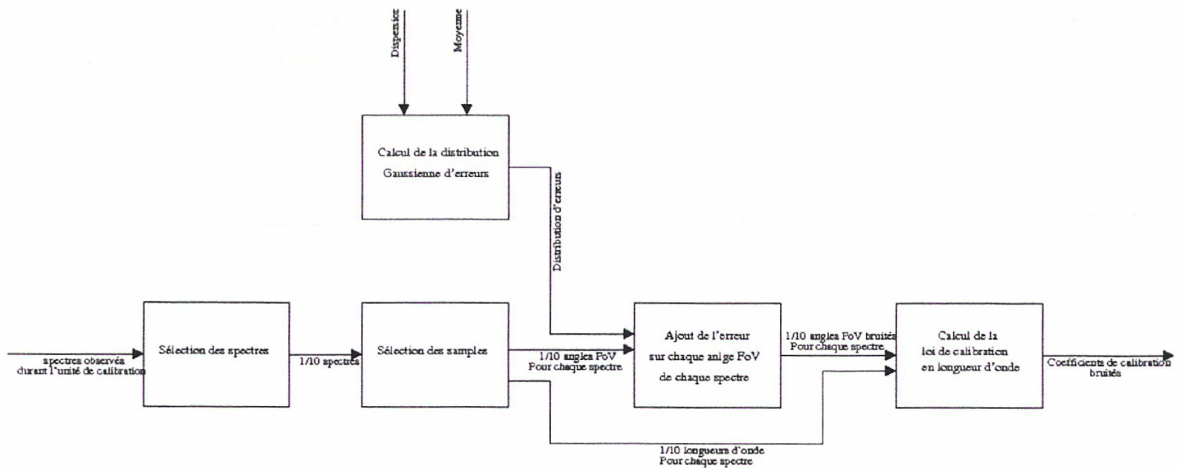


FIG. 5.3 – Représentation fonctionnelle de la phase d'initialisation en mode de convergence pour une unité de calibration donnée. Ce diagramme utilise les conventions SADT (c.f. figure 3.9).

Les figures 5.4 et 5.5 montrent deux exemples de diagnostics contrôlant la qualité de l'échelle en longueur d'onde d'un spectre après la phase d'initialisation. Elles représentent la moyenne de la distribution d'erreurs sur l'échelle en longueur d'onde en fonction de la position en samples dans les spectres observés d'une unité de calibration donnée. Ces résultats ont été obtenus en choisissant des erreurs sur les coordonnées angulaires  $(\eta, \zeta)$  égales à 0.001 degré (environ 0.4 Å) et 0.01 degré (environ 4 Å). Chaque courbe de couleur représente un unité de calibration donnée. Chaque loi de calibration propre à une unité de calibration donnée a été initialisée avec des valeurs bruitées proches de la solution théorique de calibration en longueur d'onde (i.e. la



solution théorique correspond à 0 Å d'erreur sur le diagnostic présenté).

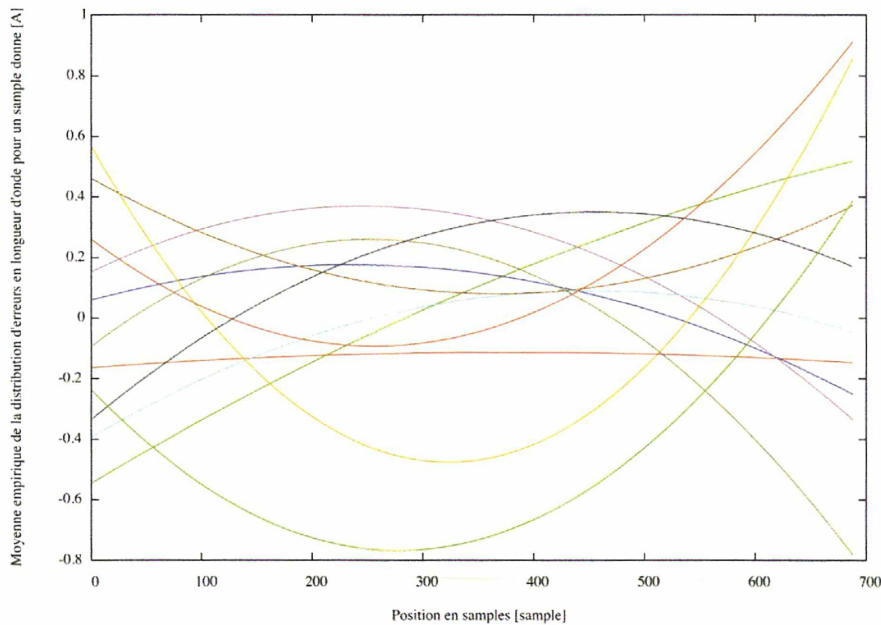


FIG. 5.4 – Ce graphique représente la moyenne de la distribution d'erreurs en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés pour chaque unité de calibration. La dispersion d'erreurs choisie pour bruitez les coordonnées ( $\eta$ ,  $\zeta$ ) est égale à 0.001 degré. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs.

## 5.2 Diagnostics de contrôle

Durant le processus SGIS, des erreurs peuvent être commises sur l'estimation des vitesses radiales, lors du source updating (c.f. section 3.2.4), et sur l'estimation des lois de calibration en longueur d'onde, lors du calibration updating (c.f. section 3.2.6). Pour étudier le comportement et la convergence du prototype de SGIS, on a besoin de visualiser et de quantifier de ces erreurs. Ainsi, dans le cadre de la validation de SGIS, on a mis en place plusieurs diagnostics de contrôle. Ces diagnostics permettent d'évaluer les performances du prototype à différentes étapes du processus de calibration. Ils sont nécessaires pour localiser les sources d'éventuelles erreurs commises dans le processus.

Dans le processus SGIS de calibration en longueur d'onde, deux types de diagnostics ont été mis en place :

- **diagnostics sur l'échelle en longueur d'onde** : il a pour objectif de quantifier les erreurs commises lors de la phase de calibration en longueur d'onde. Pour ce faire, il calcule les erreurs commises sur chaque échelle en longueur d'onde par rapport à leur vraie valeur. Ce diagnostic est décrit dans la section 5.2.1

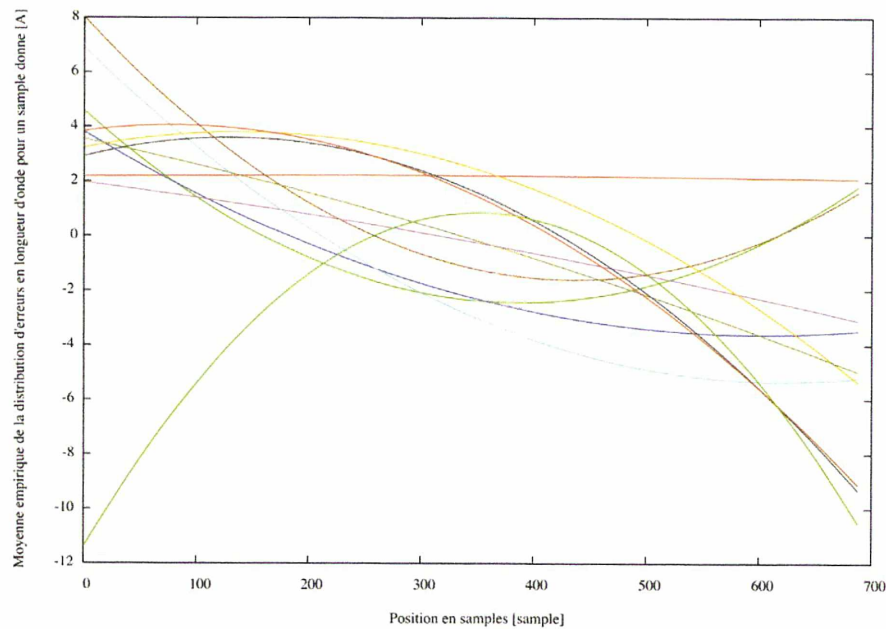


Fig. 5.5 – Ce graphique représente la moyenne de la distribution d'erreurs en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés pour chaque unité de calibration. La dispersion d'erreurs choisie pour bruitez les coordonnées  $(\eta, \zeta)$  est égale à 0.01 degré. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs.

- **diagnostic sur les vitesses radiales** : il a pour objectif de quantifier les erreurs commises lors de la phase de source updating. Pour ce faire, il calcule les erreurs commises sur chaque estimation de vitesse radiale par rapport à la vraie valeur. Ce diagnostic est décrit dans la section 5.2.2

Afin de ne pas perturber le déroulement du processus itératif, chacun de ces diagnostics n'interagit pas avec les flots de données internes du prototype de SGIS. Il interagit directement avec la base de données constamment mise à jour durant l'exécution du prototype (e.g. à la fin de la phase de "Source Updating", toutes les vitesses radiales de la base de données sont mises à jour). Ces diagnostics vont uniquement lire les données dans la base, et les analyser. Toutes les données intermédiaires résultantes de ces diagnostics sont écrites dans une base de données séparée, nommée base de données de "contrôle".

### 5.2.1 Diagnostic sur l'échelle en longueur d'onde

Le but du diagnostic sur l'échelle en longueur d'onde est d'évaluer la précision de la calibration en longueur d'onde appliquée à chaque spectre. Pour ce faire, le diagnostic évalue la différence entre les longueurs d'onde estimées lors du processus SGIS et les vraies valeurs des longueurs d'onde pour un même spectre.

Le diagnostic calcule la distribution d'erreurs de calibration par sample (i.e. une distribution pour le premier sample, une distribution pour le second, etc.) et par unité de calibration.

Soit  $\Delta\lambda(s, l)$ , l'erreur en longueur d'onde pour le sample  $s$  de l'observation  $l$ .  $\Delta\lambda(s, l)$  est calculée comme une différence entre la longueur d'onde estimée  $\lambda_{cal}(s, l)$  du sample  $s$  et la longueur d'onde vraie  $\lambda_{true}(s, l)$  :

$$\Delta\lambda(s, l) = \lambda_{cal}(s, l) - \lambda_{true}(s, l)$$

Les longueurs d'onde estimées sont calculées en utilisant la loi de calibration en longueur d'onde estimée (c.f. équation 3.3) et les coordonnées  $(\eta, \zeta)$  dans le champ de vue du RVS au temps de lecture du sample  $s$ . Les longueurs d'onde vraies proviennent de la base de données de contrôle du prototype de SGIS.

La moyenne et la dispersion de chaque distribution d'erreurs sont calculées en utilisant les formules empiriques suivantes :

$$\begin{aligned} \langle \Delta\lambda(s) \rangle_k &= \frac{1}{L(k)} \sum_{l=1}^{L(k)} \Delta\lambda(s, l) \\ \sigma_{\Delta\lambda(s)}(k) &= \left[ \frac{1}{L(k) - 1} \sum_{l=1}^{L(k)} (\Delta\lambda(s, l) - \langle \Delta\lambda(s) \rangle_k)^2 \right]^{0.5} \end{aligned}$$

où  $\langle \Delta\lambda(s) \rangle_k$  est l'erreur moyenne sur la position en longueur d'onde du sample  $s$  estimée sur tous les spectres observés durant l'unité de calibration  $k$ ,  $L(k)$  est le nombre de spectres observés durant l'unité de calibration  $k$ ,  $\sigma_{\Delta\lambda(s)}(k)$  est la dispersion de la distribution d'erreurs sur la position en longueur d'onde du sample  $s$  dans l'unité de calibration  $k$ .

## 5.2.2 Diagnostic sur les vitesses radiales

L'objectif du second diagnostic est d'évaluer la précision de l'estimation des vitesses radiales de chaque observation et de chaque source (c.f. phase de "Source Updating"). Pour ce faire, le diagnostic contient deux diagnostics internes :

- Une diagnostic sur les vitesses radiales observées : il permet d'évaluer les erreurs commises sur les vitesses radiales estimées à chaque époque (i.e. pour chaque spectre). Une distribution d'erreurs est calculée par unité de calibration.
- Une diagnostic sur les vitesses radiales moyennes : il permet d'évaluer les erreurs commises sur les vitesses radiales estimées pour chaque étoile<sup>20</sup>. Une distribution d'erreurs est calculée sur l'ensemble des étoiles de la mission simulée pour le prototype SGIS.

Une erreur sur la vitesse radiale d'une observation (respectivement d'une étoile) est calculée comme la différence entre la vitesse estimée et la vitesse vraie de l'observation (respectivement d'une étoile). Les valeurs vraies des vitesses radiales proviennent de la base de données de

---

<sup>20</sup>La vitesse radiale d'une étoile est estimée en moyennant les vitesses radiales des observations de cette étoile

contrôle.

Pour chacune des distributions d'erreurs calculées, la moyenne empirique<sup>21</sup> et la précision (i.e. dispersion) des vitesses radiales sont calculées.

Dans notre prototype, une estimation de la vitesse radiale d'un spectre est obtenue par corrélation croisée de ce même spectre calibré avec un template (c.f. section 3.2.4). Si le spectre est mal calibré en longueur d'onde, alors l'estimation de la vitesse radiale sera mauvaise. Par conséquent, les erreurs d'estimation de vitesses radiales sont liées aux erreurs de calibration en longueur d'onde. Au même titre que le diagnostic sur l'échelle en longueur d'onde, ce diagnostic permet de tracer l'évolution de la loi de calibration appliquée aux spectres de chaque unité de calibration.

### 5.2.3 Diagnostics au niveau de SGIS

Ces diagnostics ont été mis en place dans le cadre de tests d'intégration et de validation du processus. Chacun d'entre eux est placé après une étape clef du prototype :

- **Après la phase d'initialisation du processus itératif** : le diagnostic sur l'échelle en longueur d'onde permet de contrôler l'initialisation des lois de calibration en longueur d'onde. Ce diagnostic permet d'étudier l'évolution des propriétés de convergence en fonction de l'initialisation choisie. Ce diagnostic n'a lieu qu'une fois par exécution du prototype SGIS.
- **Après l'estimation des vitesses radiales dans le processus itératif** : un diagnostic permet de visualiser les erreurs commises lors de l'étape de "Source Updating" de SGIS.
- **Après l'estimation des lois de calibration en longueur d'onde dans le processus itératif** : un diagnostic identique à celui qui est utilisé après l'initialisation de SGIS. Ce diagnostic sur l'échelle en longueur d'onde permet d'évaluer les erreurs commises lors de la phase de "Calibration Updating" de SGIS. Ce diagnostic est exécuté à chaque itération afin de suivre l'évolution de la qualité des calibrations en longueur d'onde.

La figure 5.6 illustre la position de ces diagnostics dans l'architecture de la solution itérative de calibration en longueur d'onde.

## 5.3 Centroiding avec et sans biais

Plusieurs campagnes de tests ont été réalisées à partir des données présentées dans la section 4.1 :

- **Performance de la solution itérative sans biais** : on étudie les propriétés de convergence du processus itératif en fonction de la précision du centroiding. Cette campagne est décrite dans la section 5.3.1.
- **Performance de la solution itérative avec biais** : on étudie les propriétés de convergence du processus itératif en fonction de l'exactitude du centroiding (avec une précision

<sup>21</sup>La moyenne empirique, ou moyenne arithmétique, a été choisie arbitrairement comme l'estimateur de la moyenne. On aurait pu également utiliser la moyenne absolue ou la moyenne tronquée comme estimateur.

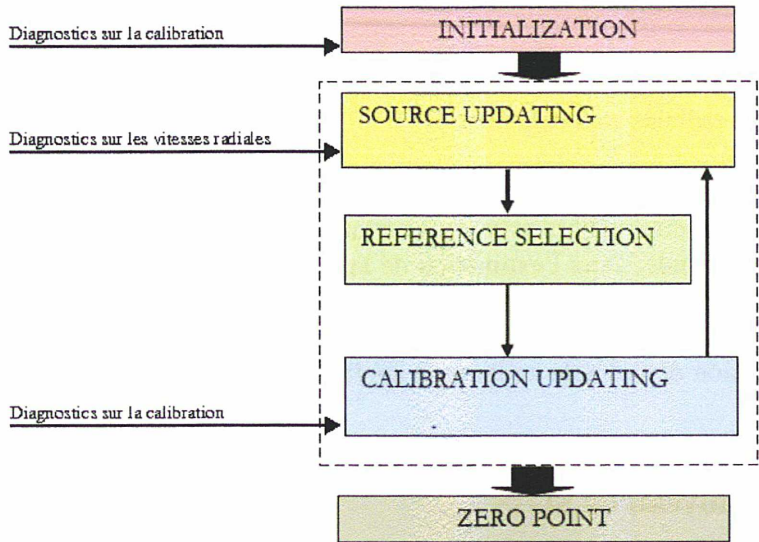


FIG. 5.6 – Localisation des diagnostics d’erreurs dans le déroulement du prototype de SGIS. Deux types de diagnostics : un diagnostic d’erreurs sur les vitesses radiales (dans le processus itératif, après l’étape de Source Updating), et des diagnostics d’erreurs sur les calibrations en longueur d’onde (après l’initialisation et, dans le processus itératif, après l’étape de Calibration Updating).

constante). Cette campagne est décrite dans la section 5.3.2.

### 5.3.1 Performance du processus itératif sans biais

Dans cette première partie de notre étude, nous fixerons le biais de la distribution de centroïding à 0 sample<sup>22</sup>. Le "protocole expérimental" consiste à faire varier uniquement la précision  $\sigma_{centroid}$  du centroïding : une dispersion est fixée avant chaque exécution du prototype de calibration. Le prototype a été exécuté en utilisant les précisions  $\sigma_{centroid}$  suivantes :

- $\sigma_{centroid} = 0.01$  sample
- $\sigma_{centroid} = 0.1$  sample
- $\sigma_{centroid} = 0.5$  sample
- $\sigma_{centroid} = 1.0$  sample

Pour chacune de ces valeurs de dispersion, le prototype sera testé en utilisant les deux conditions initiales, c’est-à-dire, le mode de "non-divergence" et le mode de "convergence" (c.f. section 5.1.2).

<sup>22</sup>En pratique, pour un échantillon de 17000 erreurs aléatoires ajoutées au couple (*sample*,  $\lambda$ ), la distribution est centrée sur une valeur de l’ordre de  $10^{-4}$  sample. Le terme "sans biais" n’est donc pas rigoureux, mais on considère le biais existant comme négligeable pour la suite de nos tests.

### Mode de non-divergence

La figure 5.7 présente les évolutions des dispersions des erreurs sur les vitesses radiales moyennes (en  $km.s^{-1}$ ) au cours du processus itératif. Chaque courbe de couleur est associée à une distribution d'erreurs sur l'estimateur sample-longueur d'onde.

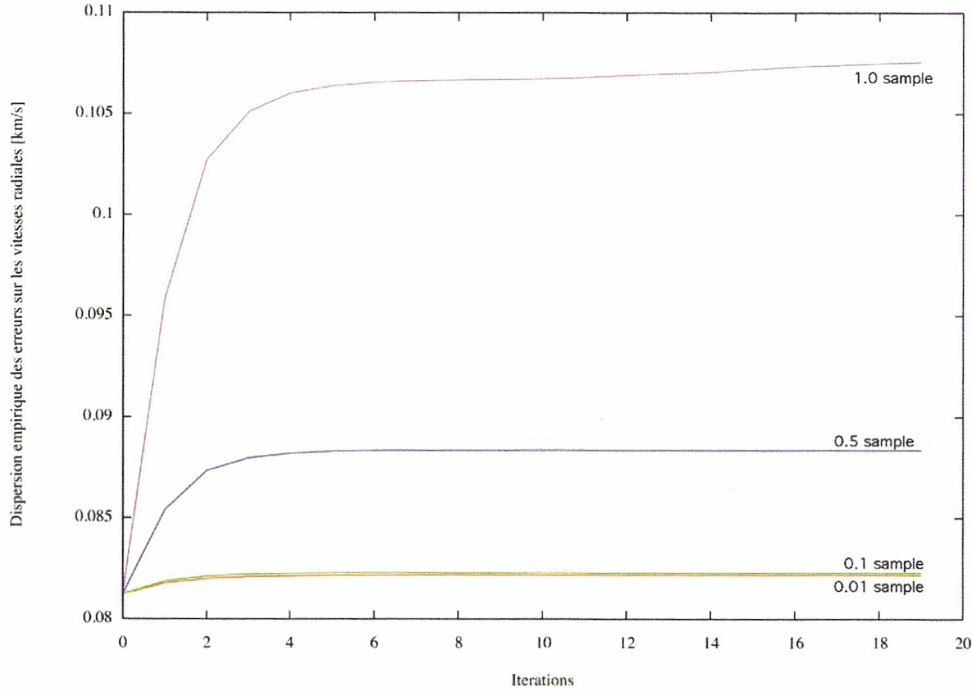


Fig. 5.7 – Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. Chaque courbe de couleur est associée à une distribution d'erreurs sur l'estimateur sample-longueur d'onde.

Les distributions des erreurs de centroiding avec des dispersions inférieures ou égales à 0.1 sample ont un impact similaire sur la non-divergence des solutions de calibration en longueur d'onde : quelques mètres par seconde séparent les dispersions des vitesses radiales estimées en utilisant des dispersions  $\sigma_{centroid} = 0.1$  sample et  $\sigma_{centroid} = 0.01$  sample. Le cas  $\sigma_{centroid} = 0.5$  sample montre, lui aussi, une non-divergence des solutions de calibration en longueur d'onde au cours des itérations : la dispersion des vitesses radiales estimées sur l'ensemble des 110 jours de données est de l'ordre de  $87.0 m.s^{-1}$ . Dans le cas  $\sigma_{centroid} = 1.0$  sample, le processus de calibration diverge : de l'itération 0 à l'itération 6, la dispersion évolue de la même manière que les autres cas (i.e. forte dégradation des solutions initiales après la première calibration, puis stabilisation des solutions), puis, à partir de l'itération 7, la dispersion se dégrade.

Les figures 5.8 et 5.9 présentent cette évolution des dispersions des erreurs sur les vitesses radiales moyennes (en  $km.s^{-1}$ ) bornée par deux fonctions arctangente de type  $y = A \times \left(\frac{\pi}{2} - arctg\left(\frac{1}{n_{iter}}\right)\right) + B$ , où  $n_{iter}$  est le numéro de l'itération, A est l'amplitude de la fonction et B est l'origine de la fonction. Dans la figure 5.9, on constate que l'écart entre la fonction arctangente

majorant l'évolution et l'évolution augmente à partir de la troisième itération. Il en est de même pour l'écart entre la fonction arctangente minorant l'évolution et l'évolution. Cela prouve que l'évolution de la dispersion reste toujours bornée par ces deux fonctions pour ces 20 premières itérations. On suppose que l'évolution de la dispersion est monotone et croissante (c'est le cas sur les 20 premières itérations). De plus, cette évolution est bornée par deux fonctions convergentes. On peut en déduire que l'évolution de la dispersion est convergente pour une erreur de 0.5 sample dans le centroïding.

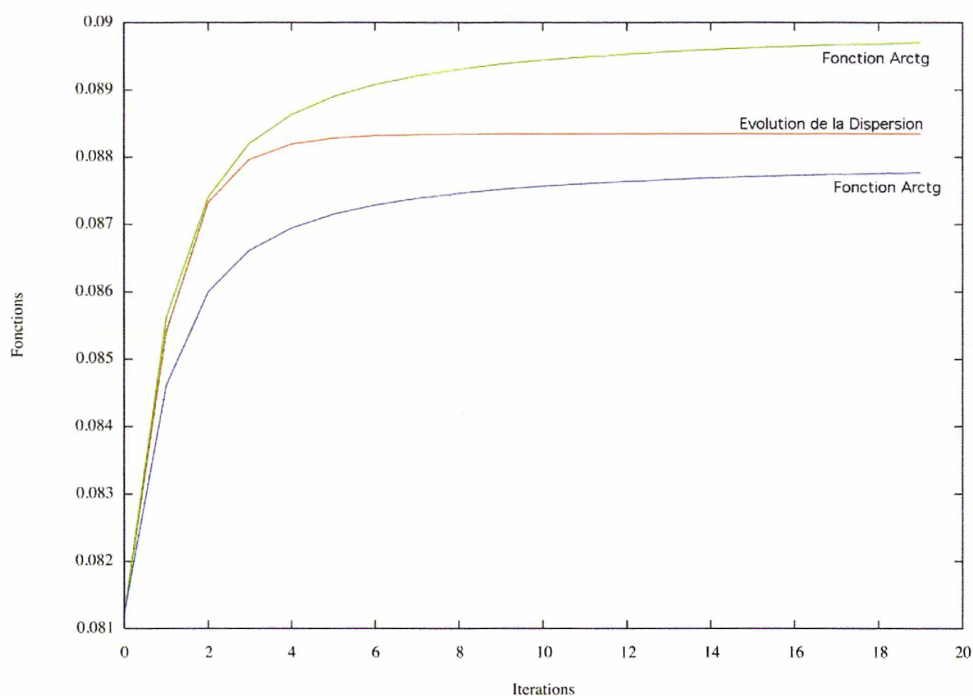


FIG. 5.8 – Convergence du prototype avec une erreur sur le centroïding égale à 0.5 sample : évolution de la dispersion des vitesses radiales bornée par deux fonctions arctangente.

La convergence des solutions de calibration en longueur d'onde impose une méthode de centroïding dont la précision est doit être meilleure que 0.5 sample (il reste une zone non-testée, lors de mon étude, entre 0.5 et 1 sample).

Les figures 5.10, 5.11, 5.12 et 5.13 représentent les résultats obtenus lors du diagnostic sur les calibrations en longueur d'onde aux itérations 1, 2, 10 et 20, en considérant la dispersion de la distribution d'erreurs de centroïding égale à 0.5 sample. Ce diagnostic, présenté dans la section 5.2.1, calcule la différence moyenne entre les longueurs d'onde estimées lors du processus SGIS et les vraies valeurs des longueurs d'onde pour chaque sample, sur tous les spectres d'une unité de calibration donnée.

Les résultats obtenus dans le diagnostic sur les longueurs d'onde illustrent la non-divergence du processus de calibration en longueur d'onde. En effet, en comparant visuellement les courbes obtenues aux itérations 1 et 2, on constate qu'entre la première itération et la deuxième, les

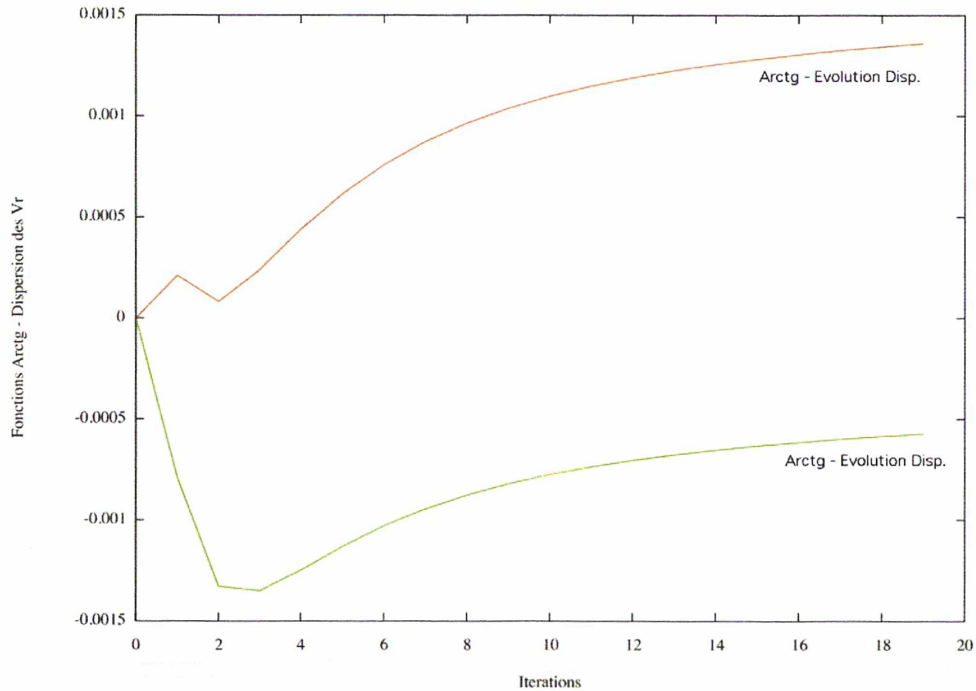


FIG. 5.9 – Convergence du prototype avec une erreur sur le centroiding égale à 0.5 sample : différence entre ces fonctions arctangente et l'évolution de la dispersion.

échelles en longueur d'onde se sont dégradées. Ce phénomène est normal car le processus démarre à partir des échelles en longueur d'onde vraie (i.e. mode de non-divergence). Après la première itération, des nouvelles lois de calibration sont calculées. Ces lois ne peuvent qu'être dégradées par rapport aux lois vraies. Entre les itérations 10 et 20, les erreurs sur les échelles en longueur d'onde ont très peu évolué : à partir des résultats de ce diagnostic, l'écart entre les itérations 10 et 20 est inférieur à  $-1.3 \cdot 10^{-4} \text{ \AA}$  (i.e. moins de  $-0.5 \text{ m.s}^{-1}$  en vitesse radiale). Ce diagnostic est complémentaire au diagnostic sur les vitesses radiales estimées présenté dans la figure 5.7. Il montre que les erreurs sur les échelles en longueur d'onde se stabilisent et, par conséquent, que le processus converge.

### Mode de convergence

La non-divergence du prototype ayant été validée pour des dispersions d'erreurs de centroiding inférieures à un demi-sample, le prototype peut être évalué en mode de convergence. Pour cette étude, on fixe la dispersion d'erreurs sur le centroiding à 0.5 sample. Comme le précédent test de non-divergence, la distribution d'erreurs sur le centroiding a un biais négligeable. Le prototype a été initialisé avec deux solutions bruitées différentes :

- Les coordonnées angulaires ( $\eta$ ,  $\zeta$ ) dans le champ de vue de l'instrument sont bruitées par l'intermédiaire d'une distribution gaussienne d'erreurs aléatoires. Cette distribution d'erreurs est non-biaisée et possède une dispersion de 0.001 degré. Elle ajoute à la solution vraie de calibration en longueur d'onde des bruits de l'ordre d'une dizaine de kilomètres



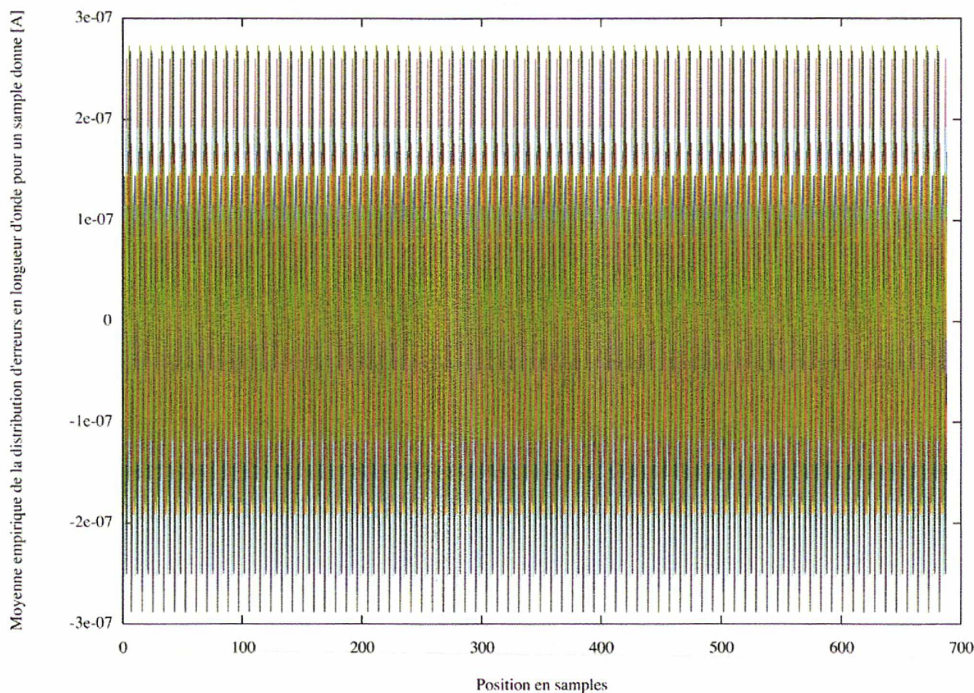


FIG. 5.10 – Diagnostics d’erreurs sur les calibrations en longueur d’onde à l’itération 1. Ce diagnostic montre l’erreur moyenne sur la longueur d’onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

par seconde (i.e. de l’ordre d’un sample).

- Pour la deuxième solution, la distribution d’erreurs utilisée est, elle aussi, non-biaisée et possède une dispersion de 0.01 degré. Elle ajoute à la solution vraie de calibration en longueur d’onde des bruits de l’ordre de plusieurs dizaines de kilomètres par seconde (i.e. de l’ordre de plusieurs samples).

Les figures 5.14 et 5.15 présentent l’évolution de la dispersion des erreurs sur les vitesses radiales moyennes (en  $km.s^{-1}$ ) au cours du processus itératif. Dans le cas de la distribution d’erreurs de dispersion égale à 0.001 degré, le processus itératif de calibration en longueur d’onde semble avoir convergé après 5 itérations. Dans le cas de la distribution d’erreurs de dispersion égale à 0.01 degré, le processus itératif de calibration en longueur d’onde converge plus lentement : après une dizaine d’itérations, la solution itérative semble stable. Les résultats montrent que le processus itératif tend à converger vers la solution de calibration obtenue dans les tests de non-divergence du prototype (courbe bleue sur la figure 5.15).

### 5.3.2 Performance de la solution itérative avec biais

Après l’influence de la précision du centroïding sur la convergence, nous allons étudier l’impact du biais sur les propriétés de convergence du processus de calibration en longueur

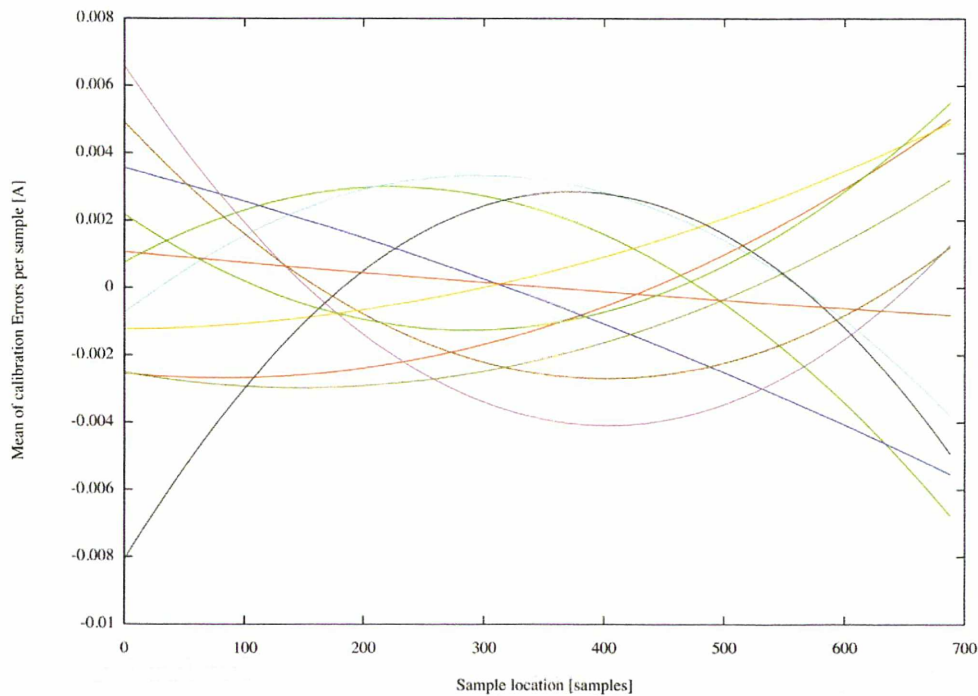


FIG. 5.11 – Diagnostics d’erreurs sur les calibrations en longueur d’onde à l’itération 2. Ce diagnostic montre l’erreur moyenne sur la longueur d’onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

d’onde. Pour nos tests, nous utiliserons un prototype utilisant une distribution de centroiding de dispersion égale à 0.5 sample. Comme dans les tests précédents, le prototype va être exécuté en modes de non-divergence et de convergence. Plusieurs valeurs de biais vont être imposées à la distribution de centroiding :

- 0.001 sample (i.e. environ  $0.013 \text{ km.s}^{-1}$ ).
- 0.01 sample (i.e. environ  $0.13 \text{ km.s}^{-1}$ ).
- 0.1 sample (i.e. environ  $1.3 \text{ km.s}^{-1}$ ).

### Mode de non-divergence

La figure 5.16 présente l’évolution de la dispersion des erreurs sur les vitesses radiales estimées (en  $\text{km.s}^{-1}$ ) au cours du processus itératif. Chaque courbe de couleur est associée à une valeur du biais : un biais de 0.001 sample en *vert*, un biais de 0.01 sample en *bleu* et un biais de 0.1 sample en *rose*. L’évolution de la dispersion des erreurs de vitesses radiales avec un centroiding "non-biaisé" est représenté par une courbe rouge (elle est cachée par la courbe verte).

Les trois distributions de centroiding ont des conséquences très différentes sur le processus itératif. Dans le cas d’un biais de 0.001 sample (environ  $10 \text{ m.s}^{-1}$ , en vert sur la figure), l’évolution de la dispersion des erreurs en vitesses radiales ne diffère que d’environ  $10^{-5} \text{ km.s}^{-1}$  de la dispersion obtenue avec le centroiding "non-biaisé". Dans ce cas, le comportement asymptotique

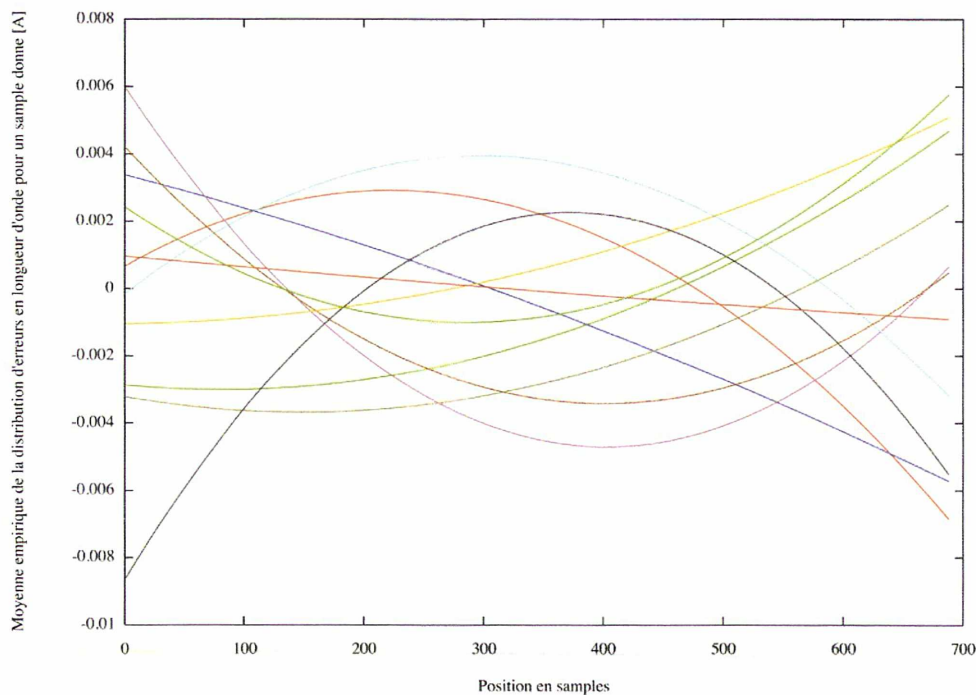


FIG. 5.12 – Diagnostics d’erreurs sur les calibrations en longueur d’onde à l’itération 10. Ce diagnostic montre l’erreur moyenne sur la longueur d’onde centrale associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

tique de la dispersion valide la non-divergence du processus itératif. Dans le cas d’un biais de 0.01 échantillon (environ  $100 \text{ m.s}^{-1}$ , en bleu sur la figure), l’évolution de la dispersion est beaucoup plus complexe. Sur les deux premières itérations, le processus itératif a le même comportement qu’avec un centrage "non-biaisé". Puis, au lieu d’avoir un comportement asymptotique horizontal, le processus de calibration semble osciller. Dans ce cas, on a trop peu d’informations pour pouvoir valider la non-divergence du processus itératif. Pour le dernier cas, un biais de 0.1 échantillon (environ  $1.0 \text{ km.s}^{-1}$ , en rose sur la figure), la dispersion des erreurs sur l’estimation des vitesses radiales diverge. Cette évolution présente des fluctuations dont l’amplitude semble augmenter avec les itérations.

### Mode de convergence

Dans les tests de convergence, j’ai utilisé un prototype avec une précision de centrage égale à 0.5 échantillon. Le processus itératif est initialisé avec une solution de calibration dégradée par une distribution gaussienne d’erreurs sur les coordonnées angulaires ( $\eta, \zeta$ ) de dispersion égale à 0.001 degré (i.e. environ un échantillon). Les figures 5.17 et 5.18 présentent l’évolution de la dispersion des erreurs sur les vitesses radiales moyennes (en  $\text{km.s}^{-1}$ ) au cours du processus itératif.

La première constatation que l’on puisse faire est que, dans les deux cas, le processus itératif semble converger vers une solution. Dans le cas d’un biais faible (0.001 échantillon), la précision

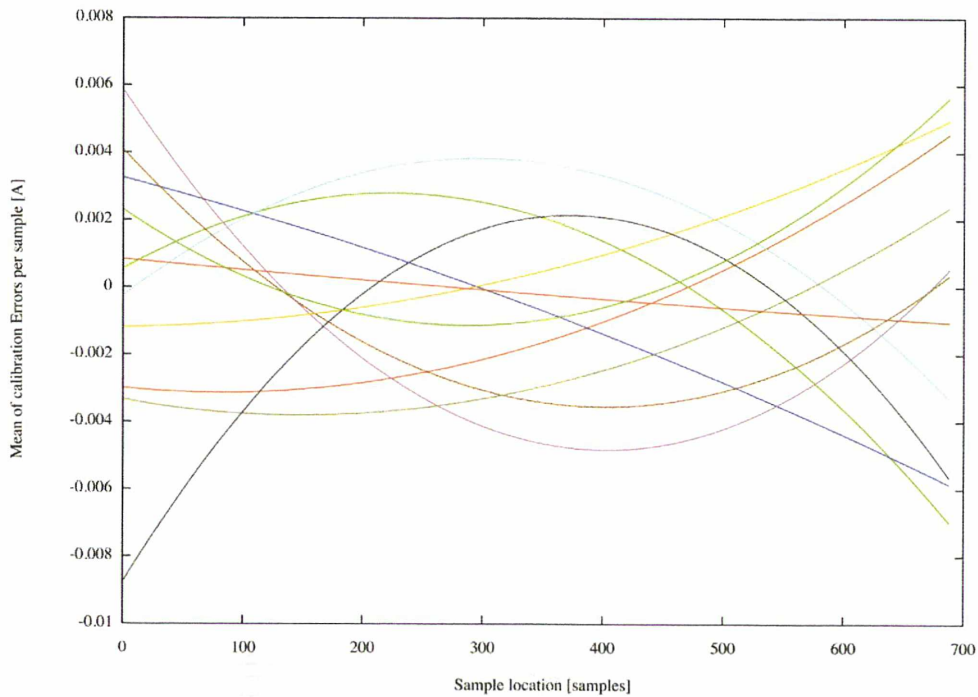


FIG. 5.13 – Diagnostics d’erreurs sur les calibrations en longueur d’onde à l’itération 20. Ce diagnostic montre l’erreur moyenne sur la longueur d’onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

sur les vitesses radiales converge lentement vers la solution du test de non-divergence obtenue dans les tests précédents. Cette solution est en dessous d’une dispersion d’erreurs de  $100 \text{ m.s}^{-1}$ . Dans le cas d’un biais de 0.01 sample, la précision sur les vitesses radiales semble converger vers la solution du test de non-divergence. Cependant, après environ 8 itérations, la précision sur les vitesses radiales semble fluctuer autour de  $120 - 130 \text{ m.s}^{-1}$ . Ce comportement instable avait déjà été révélé lors du test de non-divergence.

### 5.3.3 Synthèse

En injectant une distribution d’erreurs connue dans le module de centroiding, j’ai caractérisé le comportement général du processus itératif de calibration en longueur d’onde. Cette étude est importante pour la phase de développement de la chaîne de calibration en longueur d’onde. Elle m’a permis d’affiner les spécifications du module de centroiding de la chaîne de traitement :

- Pour observer un phénomène de convergence du processus itératif, le prototype doit intégrer un module de centroiding dont la précision est meilleure qu’un demi-sample. Il reste encore une zone non-étudiée entre 0.5 sample et 1.0 sample. De plus, plus la dispersion de la distribution est grande, plus la solution de convergence (si elle existe) va s’écarter de la solution théorique de calibration en longueur d’onde.
- Pour observer un phénomène de convergence du processus itératif, le prototype doit intégrer un module de centroiding dont l’exactitude est inférieure ou égale à 0.001 sample.

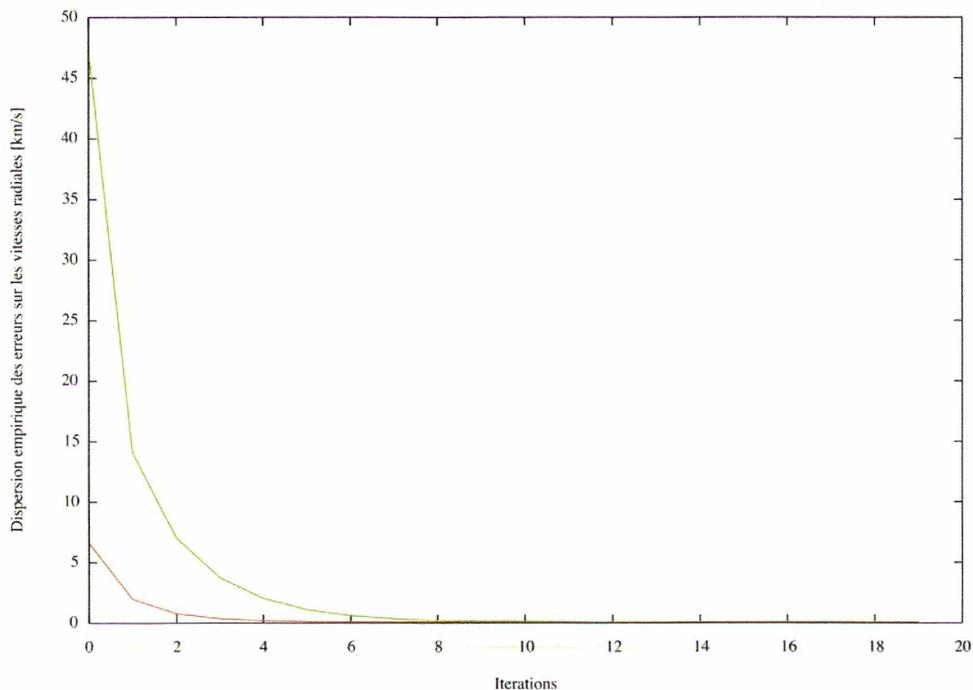


FIG. 5.14 – Evolution de la dispersion des erreurs sur les vitesses radiales estimées au cours des itérations : résultats des deux tests de convergence du prototype (la solution bruitée avec une dispersion de 0.001 degré est représentée en rouge, et celle bruitée avec une dispersion de 0.01 degré est en vert).

L'impact du biais est très important. En effet, étant injecté dans un processus itératif, un biais peut s'accumuler d'itération en itération, provoquant ainsi la divergence du processus. Dans les cas où les biais sont égaux à 0.01 et 0.1 sample, les biais sont assez faibles devant la dispersion fixée (i.e. 0.5 sample). Cependant, ces biais sont suffisamment forts pour créer des instabilités dans le processus de calibration. Ces instabilités ont été observées sur un petit nombre d'itérations (i.e. 20). Elles ne permettent donc pas de conclure sur la convergence ou la divergence du processus itératif dans ces deux cas.

En initialisant les lois de calibration en longueur d'onde avec leurs vraies valeurs, on n'injecte aucune erreur de calibration au départ du processus itératif. Le processus est donc soumis aux erreurs numériques (commises lors de l'estimation des vitesses radiales, lors du centroïding, etc.). Ainsi, le processus initialisé avec de mauvaises lois de calibration tend à converger vers la solution de non-divergence (i.e. processus avec les vraies lois de calibration en longueur d'onde).

Les trois méthodes de centroïding présentées dans le chapitre 4 ont une dispersion d'erreurs inférieure au demi-sample. La méthode de localisation par ajustement parabolique est biaisée. Malgré une étape de calibration des biais, cette méthode reste sujet à des biais plus ou moins importants (i.e. de l'ordre de 0.01 sample). Ainsi, comme on l'a montré dans ce chapitre (si-

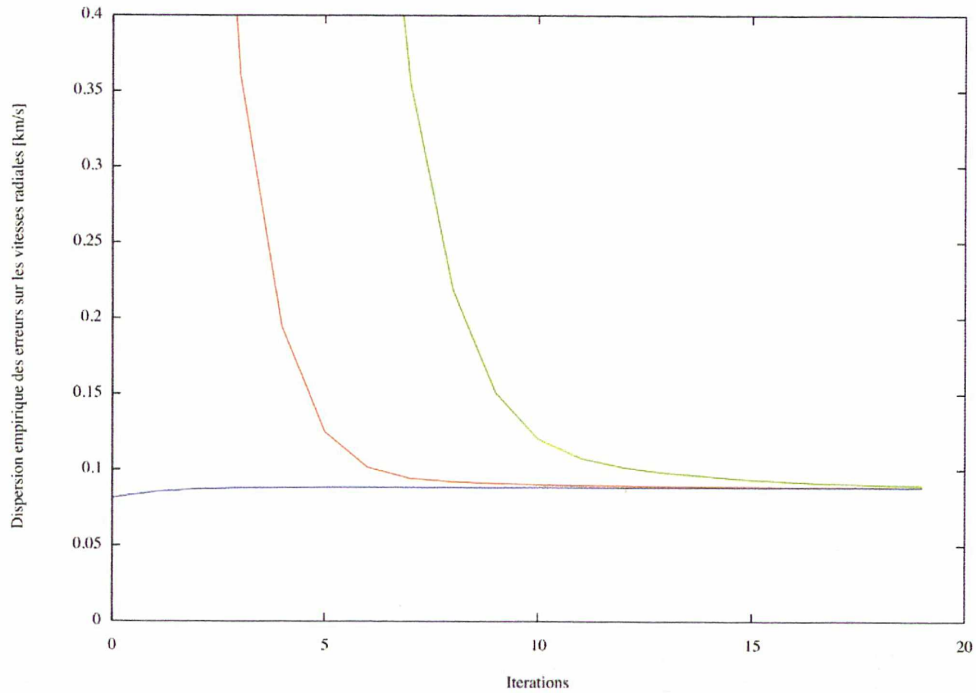


FIG. 5.15 – Evolution de la dispersion des erreurs sur les vitesses radiales estimées au cours des itérations : résultats de convergence (c.f. figure 5.14) agrandis sur la zone  $[0, 0.4] \text{ km}\cdot\text{s}^{-1}$  (axe des ordonnées). Le résultat du test de non-divergence pour le même prototype est également représenté en bleu.

mulation d'un centroiding avec une précision de 0.01 sample, c.f. section 5.3.2), cette méthode de centroiding peut causer des instabilités dans le processus itératif de calibration. Par conséquent, il vaut mieux s'orienter vers une méthode de localisation par ajustement de profil qui a des performances totalement compatibles avec la convergence du processus de calibration : ces méthodes ont des biais de l'ordre de 0.001 sample et ont des dispersions inférieures à 0.2 sample.

Dans la validation de la méthode SGIS, une étape reste à mettre en place. En effet, dans cette étude, j'ai imposé un biais unique pour l'ensemble des raies des spectres extraits. Afin de simuler les effets des blends dans le centroiding, il faudrait imposer un biais différent pour chaque raie de référence.

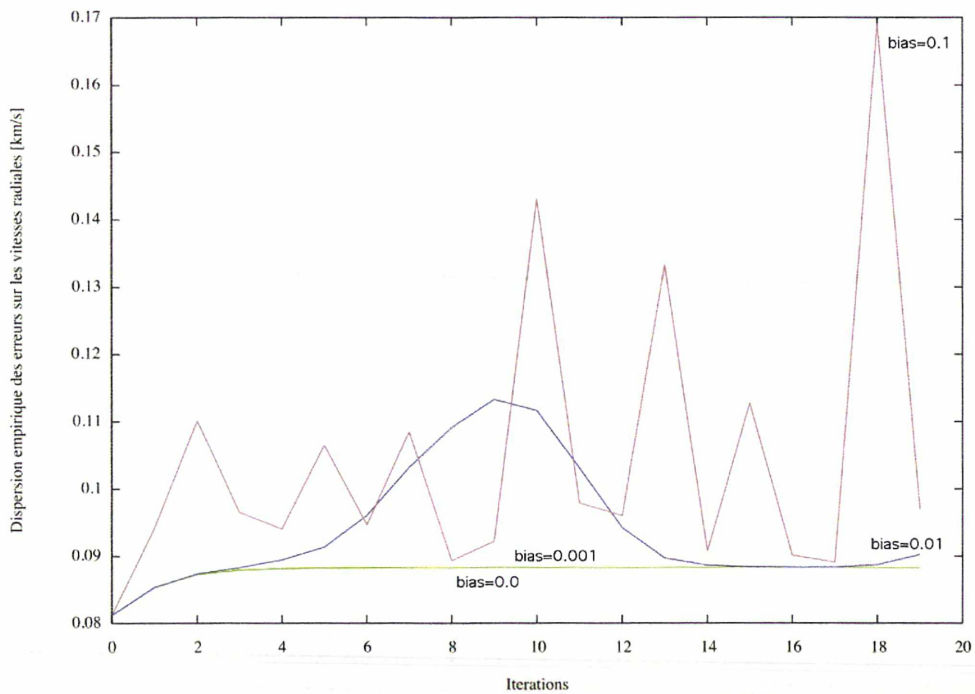


FIG. 5.16 – Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. Chaque courbe est associée à une valeur du biais : un biais nul en rouge (caché par la courbe verte), un biais de 0.001 sample en vert, un biais de 0.01 sample en bleu, un biais de 0.1 sample en rose

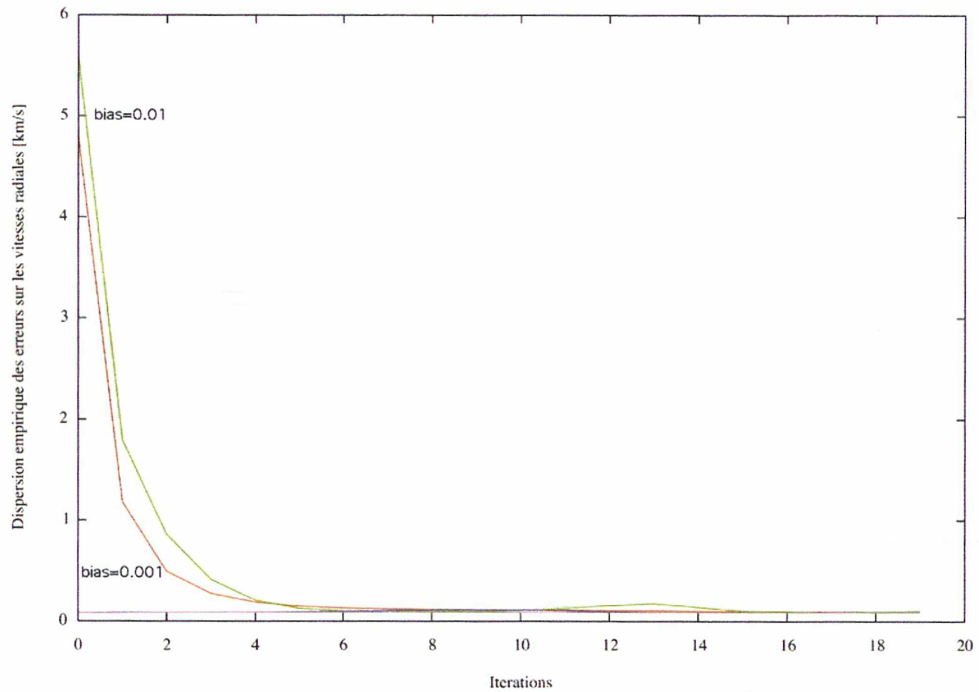


FIG. 5.17 – Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations : résultats des deux tests de convergence du prototype avec un biais de 0.001 sample (en rouge) et avec un biais de 0.01 sample (en vert). Ces résultats sont superposés aux résultats de non-divergence (c.f. figure 5.16).



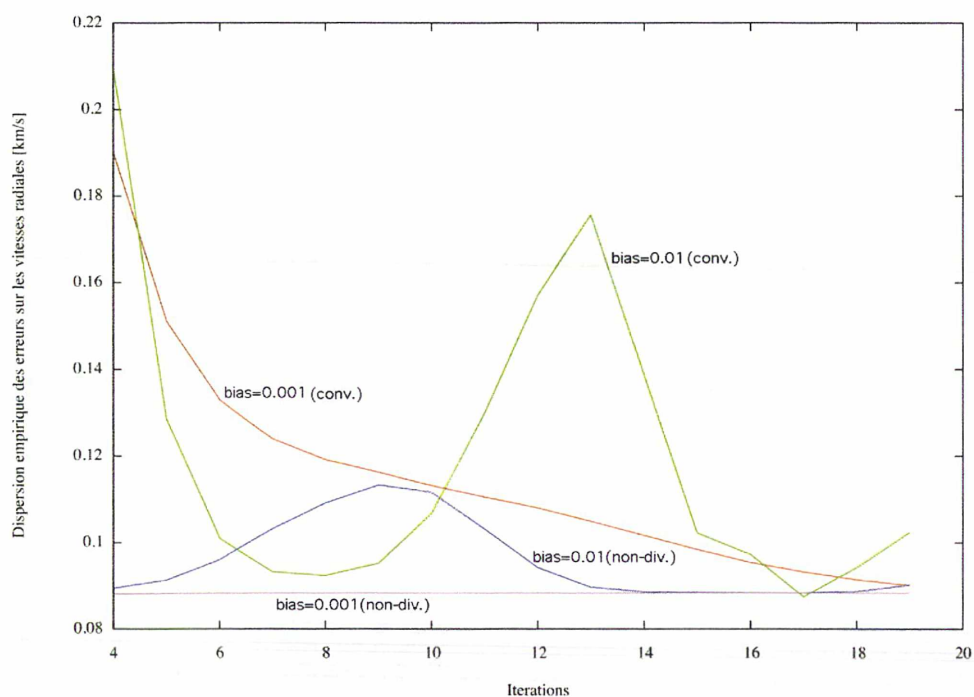


FIG. 5.18 – Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations : les résultats de convergence (c.f. figure 5.17) montrent le détail des variations aux faibles valeurs de la dispersion.

# Chapitre 6

## Performances du prototype de calibration en longueur d'onde

### Sommaire

---

<b>6.1</b>	<b>Jeu de données GASS</b>	<b>112</b>
6.1.1	Introduction aux simulateurs de données de Gaia	112
6.1.2	Définition des données GASS	113
6.1.3	Structure des données GASS	115
<b>6.2</b>	<b>Mise en place des données GASS pour la calibration</b>	<b>118</b>
6.2.1	Validation des données GASS	118
6.2.2	Redimensionnement et redressement	119
6.2.3	Création d'une librairie de templates	120
<b>6.3</b>	<b>Performances du prototype</b>	<b>122</b>
6.3.1	Configuration du prototype	122
6.3.2	Performances sans standards	125
6.3.3	Performances avec standards	129
6.3.4	Temps de Calcul	131
6.3.5	Synthèse	133

---

Après avoir testé et validé la méthode SGIS, l'objectif est de démontrer les performances de la calibration en longueur d'onde SGIS dans un cas "proche" de la réalité :

- Les données simulées prennent en compte les dernières caractéristiques de Gaia. Ces dernières caractéristiques sont appelées "Gaia-3". Elles sont décrites dans le chapitre 1
- Les données simulées prennent en compte une distribution réaliste des sources.
- Le prototype intègre une méthode de centroïding réel, et non simulé (comme c'est le cas dans le prototype utilisé dans le chapitre 5).

Dans les tests précédents (c.f. chapitre 5), le prototype utilise des données simples : un type stellaire unique et une vitesse radiale unique, une magnitude unique et un fort rapport signal-à-bruit. En passant à un modèle réaliste, le prototype utilise des données réalistes : différents types stellaires, différentes magnitudes et différentes vitesses radiales. Afin de ne pas avoir un saut trop important entre nos précédents tests et les tests de performance avec Gaia3, on a également décidé de limiter le réalisme sur certains aspects :

- Les calibrations en longueur d'onde sont linéaires et constantes.
- Toutes les autres calibrations (e.g. calibration de la PSF) sont supposées être connues.
- Les sources sont parfaitement connues : les températures effectives, les métallicités et les logarithmes de gravité de surface sont supposés parfaitement déterminés.

Dans ce chapitre, les données réalistes proviennent du GAia System Simulator (GASS). Ce simulateur a pour but de générer des paquets de données transmis au sol par le satellite Gaia. Je présenterai ce nouveau jeu de données dans la section 6.1. Puis, dans la section 6.2, je décrirai les opérations réalisées pour mettre en place les données GASS afin qu'elles soient utilisables par le prototype de calibration en longueur d'onde. Enfin, je présenterai les performances obtenues par le prototype de SGIS sur les données GASS dans la section 6.3.

## 6.1 Jeu de données GASS

### 6.1.1 Introduction aux simulateurs de données de Gaia

Un simulateur est nécessaire pour le développement des algorithmes de réduction de données. Il génère des données utilisées pour tester et estimer les performances de ces algorithmes. Dans le cadre de la mission Gaia, la simulation de données passe par trois types de simulateurs de données :

- **GAia System Simulator (GASS)** est un simulateur de flots télémétriques provenant du satellite. Il est capable de fournir un large volume de données avec un degré de réalisme approprié pour un temps de calcul raisonnable.
- **Gaia Instrument and Basic Image Simulator (GIBIS)** est un simulateur fournissant un haut degré de réalisme au niveau du pixel. Ce simulateur prend en compte les effets physiques de manière détaillée. Il va fournir des images de CCD, et implémente le traitement de données à bord du satellite.
- **Gaia Object Generator (GOG)** génère des données réduites (e.g. vitesses radiales) telles qu'on les trouvera en cours ou en fin de mission. Ce générateur permet, entre autre, de tester les chaînes de traitement de données.

De tels simulateurs, illustrés en figure 6.1, ont été développés autour d'un ensemble d'outils communs ("GaiaSimu") qui contient :

- Un modèle d'Univers simulant les caractéristiques de tous les différents types d'objets observés par Gaia : distribution spatiale, photométrie, cinématique. Les objets observés durant la mission Gaia seront principalement dans notre Galaxie (plus d'un milliard d'étoiles). Un modèle de la Galaxie a été implémenté en utilisant le modèle de Besançon. Ce modèle donne une description détaillée de la structure à grande échelle et de l'évolution de la Galaxie.
- Un modèle d'instrument simulant les champs des instruments astrométrique, spectrophotométrique et spectroscopique, la loi de balayage du satellite, les optiques des télescopes, les CCDs du plan focal, etc.

- Des outils numériques et astronomiques.

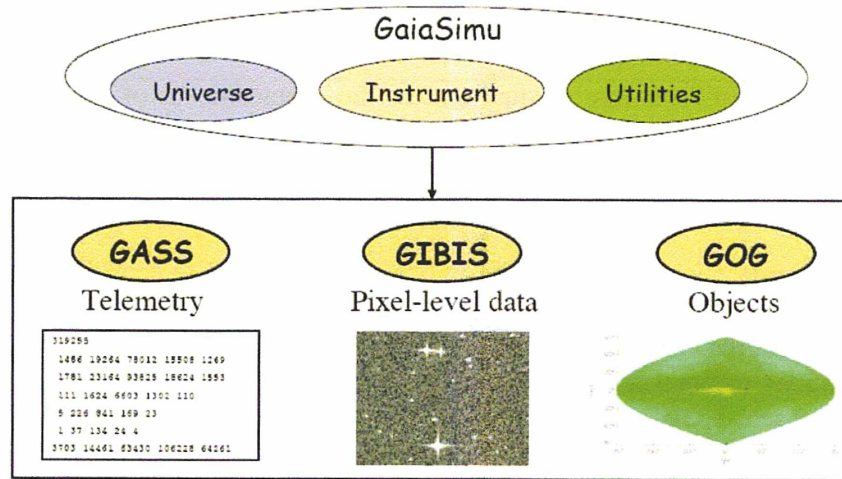


FIG. 6.1 – Simulateurs de Gaia. Mignard et al. (2007)

Dans le cadre du développement de l’algorithme de calibration en longueur d’onde des données spectroscopiques de Gaia, j’ai été amené à utiliser un large volume de données provenant de GASS. En collaboration avec E. Massana, X. Luri (Barcelone), P. Sartoretti et D. Katz (Meudon), j’ai contribué à la validation scientifique d’un jeu de données GASS courant septembre 2007. Les aspects de validation sont présentés dans la section 6.2 de ce chapitre.

Après la validation de plusieurs échantillons de données GASS, l’équipe de Barcelone nous a fourni un jeu de 6 mois de données RVS simulées par GASS. Ce jeu de données est utilisé pour tester le prototype de calibration en longueur d’onde.

### 6.1.2 Définition des données GASS

Dans le cadre du cycle 2 de développement du système de traitement des données spectroscopiques (CU6), des données simulées ont été demandées au groupe de simulation (CU2). La demande du groupe spectroscopie CU6 est décrite dans le rapport Sartoretti et al. (2006). La partie de cette demande concernant les données pour les modules de calibration est listée dans la table 6.1 (extraite de Sartoretti et al. (2006)).

Voici un résumé des principales caractéristiques de la demande :

- 6 mois de données.
- Un profil Gaussien pour la PSF du RVS.
- Des échelles en longueur d’onde linéaires et constantes.
- Utilisation du modèle de l’Univers.

<b>Code :</b> SR-CU6-2-D-GASS-002	<b>Contact for CU2 :</b> Sartoretti, P.
<b>Title :</b> Simulations for "calibration" test	
<b>Dependences :</b> dependence N/A	
<b>Priority :</b> mandatory	
<b>Desired availability :</b> end of the cycle 2	
<p>The "RVS calibration" software product derives the characteristics of the RVS instrument :</p> <ul style="list-style-type: none"> <li>- CCD properties (bias, response, readout and dark noises, blemishes, ...).</li> <li>- Photometric throughput, saturation level, ...</li> <li>- PSF profile.</li> <li>- Wavelength scale.</li> </ul> <p>A large fraction of the characteristics of the RVS will be iteratively self-calibrated using RVS observations accumulated over several months (in a "Spectroscopic Global Iterative Solution" approach - SGIS see RD7). For cycle 3, the objectives of the tests are to assess the general convergence properties of the calibration software (requiring some realism in the description of the scan law, etc) but not to dive into the full complexity of the characteristics of the "real" RVS (i.e. simple descriptions of the spectrograph properties are enough for CU6 cycle 3 : e.g. Gaussian PSF). To be able to test properly the SGIS method, a long temporal baseline is needed : i.e. 6 months of continuous data. The above specifications correspond to GASS-type simulations.</p> <p>The calibration processes also rely on complementary information on the properties of the sources, e.g. <math>G_{RVS}</math> magnitudes, atmospheric parameters. Therefore, fine assessment of the performance would require also GOG-type simulations. Yet, this fine assessment is not planned for CU6 cycle 3, but later cycle (TBD). In cycle 3, it will be enough to use the true values (i.e. the values used in the simulations) for these quantities, which can be provided by GASS.</p> <p><u>Simulator(s) :</u> GASS</p> <p><u>General description :</u> 6 months of continuous observations of 1 CCD, limited to the "bright" stars.</p> <p><u>Universe model :</u> used.</p> <p><u>Length of observation :</u> 6 months.</p> <p><u>Source magnitude(s) :</u> The brightest part of the magnitude range foreseen to be relevant for the calibrations, i.e. <math>4.75 \leq G_{RVS} \leq 8</math>.</p> <p><u>Stellar densities :</u> As in the Galaxy model.</p> <p><u>Source type :</u> stars only - a few tens of spectral types.</p> <p><u>Source <math>V_r</math> :</u> as in the Galaxy model.</p> <p><u>Fraction of binaries :</u> 0 %.</p> <p><u>Fraction of variables :</u> 0 %.</p> <p><u>Background :</u> zodiacal light only (i.e. no nebulae, etc).</p> <p><u>Cosmic rays :</u> none.</p> <p><u>Crowding :</u> none.</p>	

Suite de la table 6.1 page suivante

<p><u>Transverse motion</u> : none.</p> <p><u>Number of transits</u> : as for 6 months of observation.</p> <p><u>Number of field of views</u> : 2 (but without taking into account the crowding).</p> <p><u>Number of CCDs</u> : 1.</p> <p><u>PSF</u> : Gaussian.</p> <p><u>Optical distortions</u> : none.</p> <p><u>Wavelength scale</u> : Linear and constant.</p> <p><u>Source selection</u> : all sources in the considered magnitude range.</p> <p><u>Window allocation</u> : no conflict between windows, only rectangular shape untruncated windows.</p> <p><u>Sampling strategy</u> : As in Gaia-3, the 2 samplings valid in the considered magnitude range and for the calibration faint stars.</p> <p><u>Transmission/efficiency</u> : As in the Gaia-3 design. Include effects of CCDs non-linear response and saturation.</p> <p><u>GASS outputs</u> :</p> <ul style="list-style-type: none"> <li>- RVS windows : <ul style="list-style-type: none"> <li>- Simulated flux per sample.</li> <li>- True field angles and errors of each sample for all the sources at the time of readout.</li> <li>- True wavelengths per sample.</li> </ul> </li> <li>- True radial and rotational velocities.</li> <li>- True G &amp; <math>G_{RVS}</math> magnitudes.</li> <li>- True atmospheric parameters (<math>T_{eff}</math>, <math>\log g</math>, <math>[Fe/H]</math>).</li> </ul>
--

Tab. 6.1: Demande de données pour les tests des modules de calibration des données spectroscopiques de Gaia. Sartoretti et al. (2006)

### 6.1.3 Structure des données GASS

Les 6 mois de données RVS demandés au groupe de simulation nous ont été livrés courant septembre 2007. Ces 6 mois ont été divisés en 36 jeux de fichiers GASS, où chaque jeu de fichiers correspond à 5 jours de données. Pour chacun de ces 5 jours, une simulation GASS va générer les fichiers suivants :

- **SRC.dat** contient les caractéristiques "vraies" de chaque source observée durant la simulation. Ce fichier contient un identifiant propre à chaque source, les coordonnées équatoriales ( $\alpha$ ,  $\delta$ ) de la source, les données photométriques (magnitudes V,G et  $G_{RVS}$ ), les paramètres atmosphériques de chaque source. Ces informations ont été générées en utilisant le modèle d'Univers. Chaque ligne de ce fichier correspond à une source donnée.
- **AUX.dat** contient des informations auxiliaires pour chaque observation (ou spectre) de chaque source (listée dans le fichier SRC.dat). Ces informations nous renseignent sur la manière dont a été généré chaque spectre. Ce fichier contient l'identifiant de la source (le même que celui du fichier SRC.dat), un identifiant propre à chaque observation pour une source donnée, le temps d'observation (en secondes), la vitesse radiale utilisée pour

générer le spectre d'une observation donnée, les coordonnées  $(\eta, \zeta)$  dans le champ de vue du RVS au temps de lecture du sample médian du spectre observé, etc. Chaque ligne de ce fichier correspond à une observation.

- **RVSTM.dat** contient les données télémétriques de chaque observation, c'est-à-dire le spectre brut de chaque observation tel qu'on le recevra du satellite. Chaque spectre brut peut être identifié à partir des identifiants de la source et de l'observation.
- **LMB.dat** contient l'échelle en longueur d'onde vraie (en nanomètres) de chaque observation. Chaque échelle en longueur d'onde est associée à des identifiants de source et d'observation. Chaque ligne de ce fichier correspond à une observation donnée.

La figure 6.2 illustre la densité d'étoiles observées en échelle logarithmique et en coordonnées équatoriales pour les 6 mois de mission Gaia simulés par GASS.

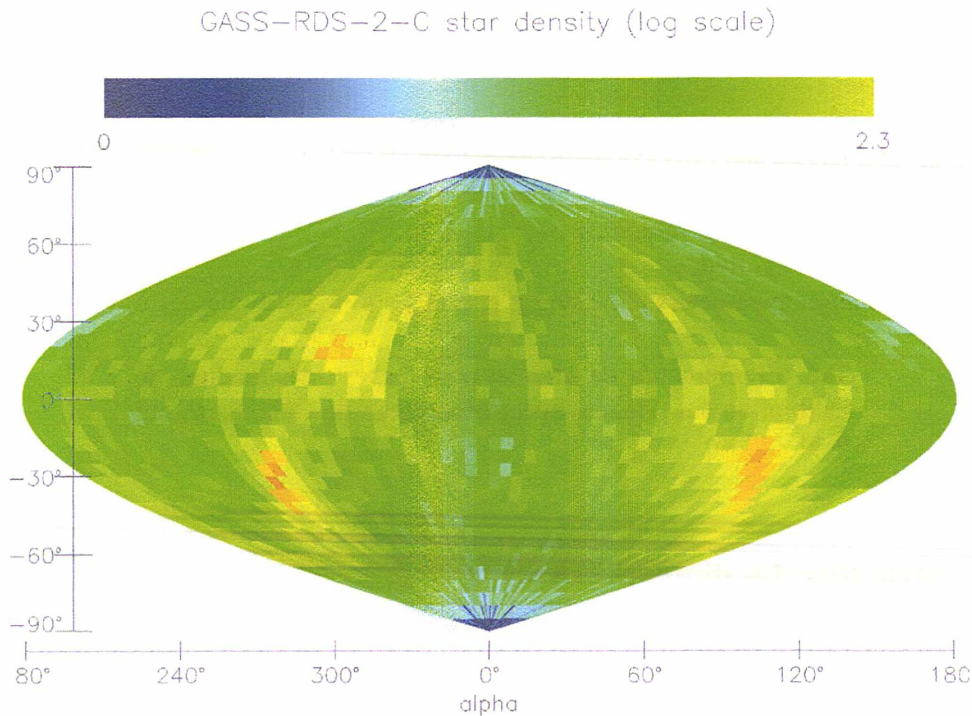


FIG. 6.2 – Densité d'étoiles observées en échelle logarithmique projetée en coordonnées équatoriales, pour 6 mois de données Gaia simulées par GASS. X. Luri, Barcelone, 2007

La figure 6.3 présente les distributions en magnitude, en bandes G et  $G_{RVS}$ , pour les sources observées pendant les 6 mois de données simulées. D'après les histogrammes, on constate une coupure nette à la magnitude  $G = 8.5$ . Ainsi, le simulateur GASS a simulé toutes les étoiles de magnitude  $G \leq 8.5$  (au lieu des étoiles de magnitude  $4.75 \leq G_{RVS} \leq 8$ , c.f. table 6.1).

Le simulateur GASS simule la sélection d'étoiles telle qu'elle sera effectuée à bord de Gaia : cette sélection sera faite au niveau des skymappers (ASM), c'est-à-dire en bande G.

Ceci explique la coupure constatée dans la distribution en bande G (c.f. figure 6.3) et entraîne un manque d'étoiles en bande  $G_{RVS}$  par rapport à la demande de la CU6 (c.f. table 6.1).

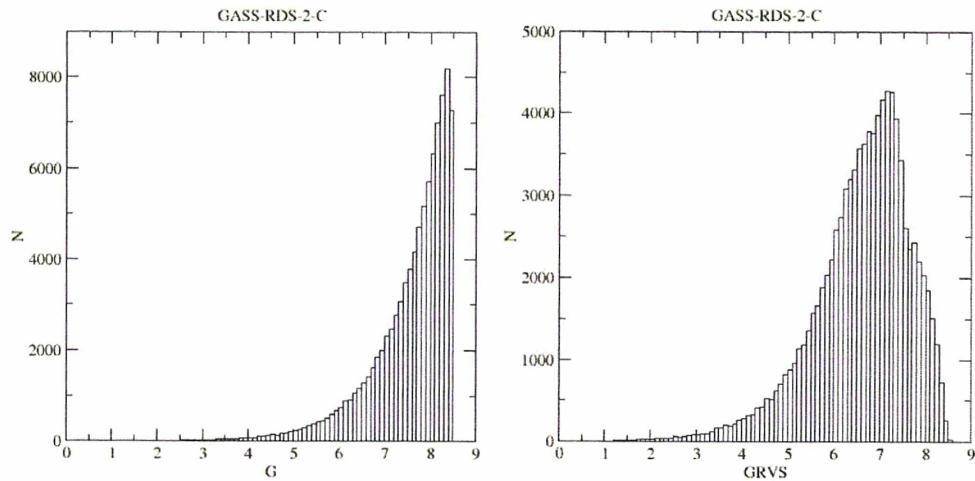


FIG. 6.3 – Distribution en magnitudes G et  $G_{RVS}$  des sources observées pour 6 mois de données simulées. X. Luri, Barcelone, 2007

La figure 6.4 montre le diagramme HR des étoiles observées durant les 6 mois de données Gaia simulées par GASS. 92968 sources entre les magnitudes  $3 \leq G \leq 8.5$  sont observées durant ces 6 mois. Cela représente 171110 spectres dans la bande spectrale du RVS.

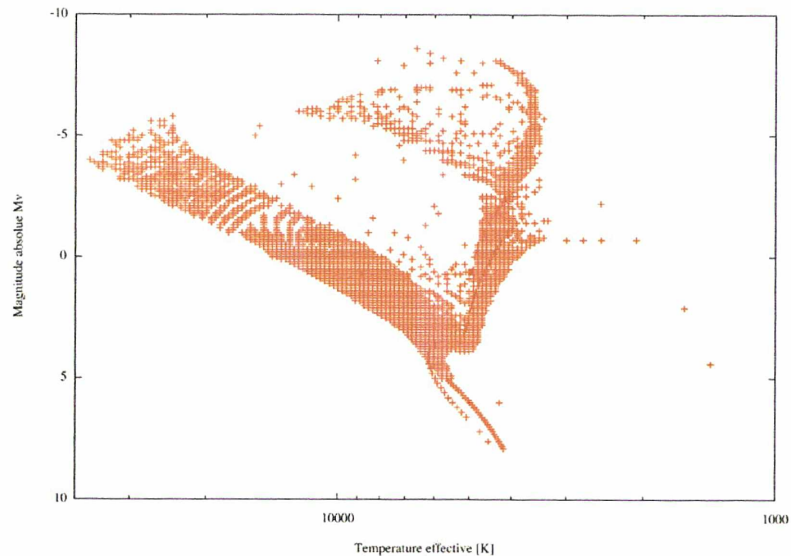


FIG. 6.4 – Diagramme HR des 6 mois de données GASS.



## 6.2 Mise en place des données GASS pour la calibration

Les données GASS telles qu'on les a reçues ne sont pas utilisables par le prototype de calibration en longueur d'onde. De plus, dans les jeux de fichiers générés par GASS, toutes les informations ne sont pas nécessaires au processus de calibration en longueur d'onde. Par conséquent, afin de pouvoir correctement exécuter le prototype sur ces données, des opérations préliminaires ont dû être mises en place :

- Validation des données, présentée dans la section 6.2.1.
- Redimensionnement et redressement du continu des spectres, présentés dans la section 6.2.2.
- Mise en place la librairie de templates à partir des spectres synthétiques de Marcs, présentée dans la section 6.2.3.

### 6.2.1 Validation des données GASS

Dans ce processus d'extraction, on a intégré un processus de validation permettant de s'assurer de la qualité des données GASS. Au cours du développement de ce processus de validation, la rapidité a été préférée à la précision. En effet, le but est de lire et de valider les données GASS dans un temps minimum. Cela correspond à une première étape de validation sur les données GASS. Une seconde étape, se concentrant sur la précision, pourrait être intégrée par la suite.

Le processus de validation consiste, dans un premier temps, à tester la concordance et la cohérence entre les différents fichiers :

- Le nombre d'observations doit être identique dans les fichiers AUX.dat, RVSTM.dat et LMB.dat.
- L'échantillonnage de chaque spectre (RVSTM) doit être adapté à la magnitude de la source correspondante (SRC). Par exemple, une source de magnitude  $G_{RVS} = 8$  doit avoir un spectre sur  $1104 \times 10$  pixels dans le fichier de télémétrie.
- Une magnitude "observée" est calculée à partir des flux provenant du fichier de télémétrie RVSTM.dat, en utilisant la formule suivante :

$$G_{RVS}^{Obs} = -2.5 * \log \left( \frac{\sum flux_{spectre}}{T_{exposure}} \right) + G_{RVS}^0$$

où  $G_{RVS}^{Obs}$  est la magnitude "observée" estimée à partir de la somme des flux dans le spectre ( $flux_{spectre}$ ),  $T_{exposure}$  le temps d'exposition par CCD et  $G_{RVS}^0$  le point zéro de la magnitude  $G_{RVS}$ , égal à 21.01. La magnitude "observée" ainsi calculée doit correspondre à la magnitude annoncée dans le fichier SRC.dat. Si la différence entre la magnitude observée et la magnitude provenant du fichier SRC.dat est supérieure (en valeur absolue) à 0.02 magnitude, alors le programme de validation envoie un message d'erreur. Ce seuil de tolérance a été fixé à partir de l'erreur qui peut être commise par le simulateur sur la magnitude de la source (i.e. 0.01 magnitude). De plus, pour des raisons de rapidité, la magnitude observée a été estimée sur un domaine en sample et non sur un domaine en longueur d'onde. Ainsi, on a décidé de fixer le seuil de tolérance au double de l'erreur sur les magnitudes GASS.

- Soit la vitesse radiale "observée" dans le spectre calibré à partir des échelles en longueur d'onde vraies contenues dans le fichier LMB.dat. Elle est calculée sur les trois raies du calcium ionisé dans le spectre GASS de la manière suivante :

$$V_R^{obs}(\lambda_{Ca}^0) = \frac{\lambda_{Ca}^{obs} - \lambda_{Ca}^0}{\lambda_{Ca}^0} * c,$$

où  $V_R^{obs}(\lambda_{Ca}^0)$  est la vitesse radiale calculée pour la raie du calcium ionisé de longueur d'onde vraie  $\lambda_{Ca}^0$  (dans le vide, les longueurs d'onde des trois raies sont 850.036, 854.444 et 866.452 nm),  $\lambda_{Ca}^{obs}$  la position en longueur d'onde observée de la raie correspondante et  $c$  la vitesse de la lumière dans le vide. Les vitesses radiales observées pour chacune des raies du calcium doivent être similaires à la vitesse radiale annoncée dans AUX.dat : les positions des raies dans les spectres GASS étant estimées à un sample près (erreur de détection des raies), la différence entre les vitesses radiales observées et théoriques doit être inférieure ou égale à environ 10 km.s<sup>-1</sup>.

Dans une deuxième temps, le processus de validation va vérifier que les données GASS fournies respectent les spécifications demandées pour le cycle 2 (c.f. demande présentée dans la table 6.1). Le processus va principalement vérifier que :

- l'échelle en longueur d'onde des données GASS est bien linéaire et constante.
- la gamme de magnitude demandée est bien respectée. Les étoiles des 6 mois de données simulées par GASS doivent avoir une magnitude  $G_{RVS}$  comprise entre 4.75 et 8.

On a pu vérifier que la demande a été respectée en ce qui concerne les échelles en longueur d'onde. Par contre, la gamme de magnitude ne respecte pas exactement la demande : les étoiles GASS ont une magnitude  $G \leq 8.5$ . Ainsi, le nombre de sources est moins important que prévu. Ce dernier point n'a pas d'influence sur l'exécution du prototype.

### 6.2.2 Redimensionnement et redressement

Avant d'être directement utilisable par le prototype de calibration en longueur d'onde, les spectres GASS doivent être redimensionnés afin de supprimer les effets observés sur les bords des spectres. La figure 6.5 représente un spectre simulé par GASS. Les effets observés sur les bords du spectre sont présents sur l'ensemble des données GASS. Pour le cycle 2, le filtre du RVS est numériquement représenté par une fonction porte. La largeur de cette fonction porte est plus large que le domaine spectral du RVS. Ainsi, tous les flux en dehors du domaine spectral du RVS (i.e. [847, 874] nm) sont mis à zéro.

Le redimensionnement consiste à supprimer les 40 premiers et les 40 derniers samples de chaque spectre.

En plus du redimensionnement, un redressement partiel du continu est nécessaire. Une différence de courbures entre un spectre et son template peut inclure des erreurs systématiques dans les processus de corrélation croisée utilisés par le prototype pour l'estimation des vitesses radiales et des centroidings des raies dans les spectres.

La courbure du continu est composée d'une composante due à l'instrument et d'une composante astrophysique. Le template n'ayant pas de courbure instrumentale, on va chercher à

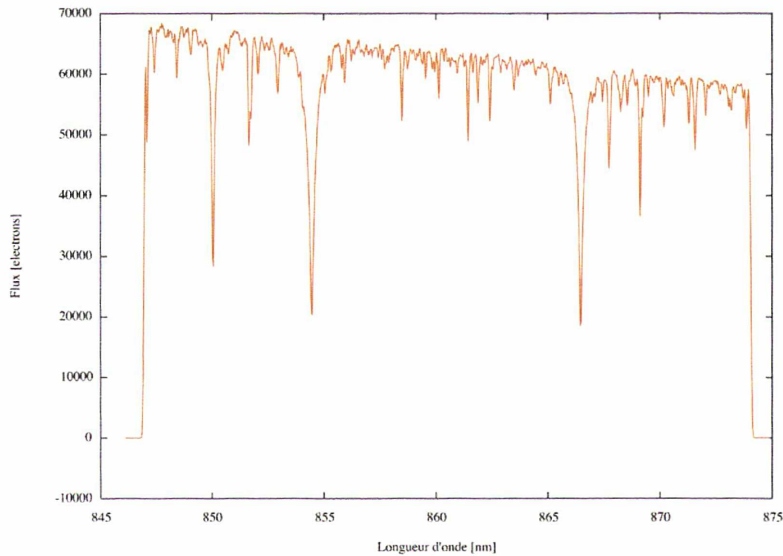


FIG. 6.5 – Exemple de spectre généré par GASS pour une étoile de type K :  $G_{RVS}=3.1447$ ,  $T_{eff} = 4750K$ ,  $[Fe/H] = -0.25$  et  $\log g = 2.5$

enlever la composante instrumentale au niveau du continu des spectres. Le processus de redressement partiel consiste donc à diviser chaque spectre par le rendement spectral du système télescope+RVS+plan focal (ou "bande GRVS") en respectant la conservation d'énergie dans les spectres.

La figure 6.6 représente un spectre GASS obtenu avant et après la phase de redimensionnement - redressement : le spectre GASS "brut" est représenté en rouge, le spectre redimensionné et redressé est représenté en vert.

### 6.2.3 Création d'une librairie de templates

Les templates ("spectres patrons" en français) sont des spectres synthétiques sans bruit et convolués avec les effets instrumentaux. Ils ne diffèrent des spectres observés que par les informations que l'on veut extraire des spectres observés. Les templates utilisés lors du processus SGIS ont tous une vitesse radiale nulle. En effet, dans le cadre des modules d'estimation des vitesses radiales et de centroïding, on a besoin de templates pour calculer des vitesses radiales (c.f. section 3.2.4) et des décalages en longueur d'onde (c.f. section 4.4).

À cause de l'évolution temporelle des propriétés instrumentales (e.g. PSF), les templates doivent être générés au fur et à mesure, dans le processus itératif de calibration. Cependant, pour gagner du temps, on a choisi de les générer en une seule fois avant l'exécution du prototype.

Ces spectres patrons sont calculés à partir de spectres synthétiques de Marcs. Ces spectres

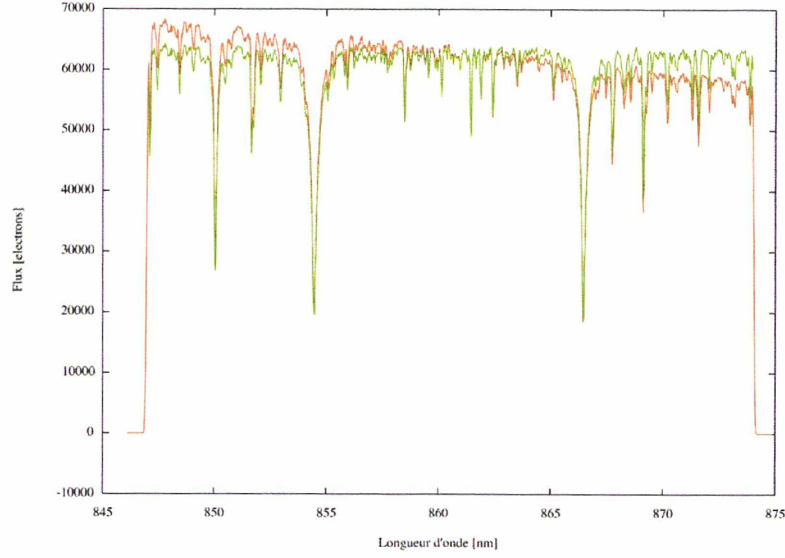


FIG. 6.6 – Exemple de spectre GASS redimensionné et redressé pour une étoile de type K :  $G_{RVS}=3.1447$ ,  $T_{eff} = 4750K$ ,  $[Fe/H] = -0.25$  et  $\log g = 2.5$ . Le spectre GASS "brut" est représenté en rouge et le nouveau spectre redimensionné et redressé en vert.

synthétiques ont une résolution spectrale de 300000 dans le domaine en longueur d'onde [840.0, 890.0] nm, plus large que celui du RVS<sup>23</sup>. L'échelle en longueur d'onde est commune pour tous les spectres synthétiques. Chacun d'entre eux a une vitesse radiale nulle. Leurs paramètres atmosphériques sont définis par une grille tridimensionnelle :

- Une dimension en métallicité :  
-5.00, -4.00, -3.00, -2.00, -1.50, -1.00, -0.75, -0.50, -0.25, 0.00, 0.25, 0.50, 0.75, 1.00.
- Une dimension en température :  
de 4000 à 8000 K par pas de 250 K.
- Une dimension en logarithme de gravité de surface :  
de -0.5 à 5 par pas de 0.5.

Chaque combinaison de paramètres atmosphériques ne donne pas forcément lieu à un spectre synthétique.

Sur ces spectres synthétiques, on réalise une série d'opérations afin d'obtenir des spectres patrons directement utilisables dans mes modules de calibration en longueur d'onde. Ces opérations sont listées ci-dessous :

- Lecture des flux et des longueurs d'onde du spectre synthétique. Les flux sont en densité d'énergie spectrale [ $W.m^{-2}.nm^{-1}$ ], et les longueurs d'onde en [nm].
- Conversion des densités d'énergie [ $W.m^{-2}.nm^{-1}$ ] en [ $photons.s^{-1}.cm^{-2}.nm^{-1}$ ]. On utilise la formule suivante :

$$flux_{photon}(i) = flux_{densite}(i) * \lambda(i) * \frac{1000}{c * h}$$

<sup>23</sup>Pour la corrélation croisée, le spectre patron doit nécessairement être plus large que le spectre observé

où  $flux_{densite}(i)$  est le flux du pixel  $i$  exprimé en densité d'énergie,  $flux_{photon}(i)$  le flux en photons,  $\lambda(i)$  la longueur d'onde de  $i$  en  $nm$ ,  $c$  la vitesse de la lumière en  $[nm.s^{-1}]$  et  $h$  la constante de Plank convertie en  $[erg.s]$ .

- Convolution du spectre par une gaussienne afin de passer de  $R = 300000$  à  $R = 11500$ , résolution spectrale du RVS.
- Resampling du spectre afin de diviser par dix sa taille : on passe d'un pas de  $0.001 nm$  à un pas de  $0.01 nm$  au niveau de l'échantillonnage. L'élément de résolution du RVS est égal à  $\delta\lambda = \frac{\lambda}{11500} \simeq 0.075 nm$ . Il y a donc un peu plus de sept points par élément de résolution. Cette opération nous permet d'éviter de garder en mémoire un nombre de points important (i.e. on passe d'environ 50000 points à 5000 points).
- Écriture du spectre dans un fichier.

La figure 6.7 présente un exemple de spectre synthétique et de template généré avant l'exécution du prototype.

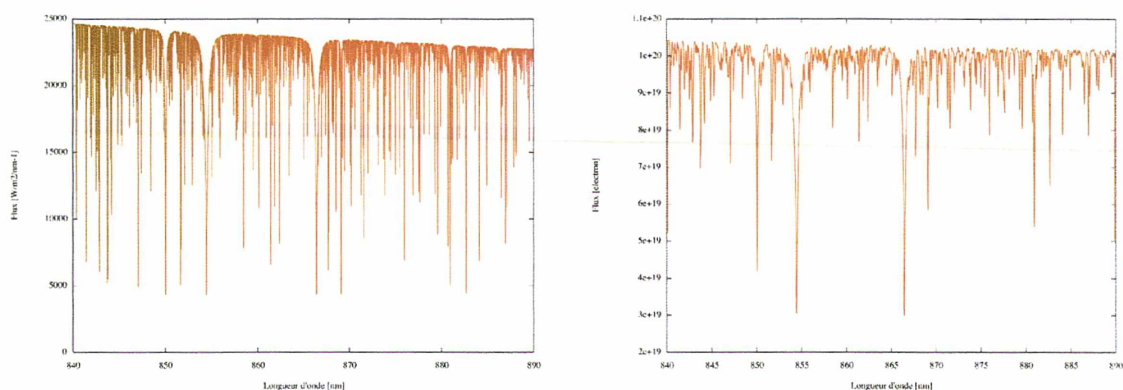


FIG. 6.7 – Exemple de création d'un spectre "patron" à partir d'un spectre synthétique de Marcs  $R = 300000$ ,  $T_{eff} = 4750K$ ,  $[Fe/H] = 0.0$  et  $\log g = 3.0$  : - À gauche, le spectre synthétique de Marcs tel qu'il est lu dans le fichier : flux en densité d'énergie avec une résolution spectrale  $R = 300000$  et 50000 points - À droite, le spectre de Marcs convolué resamplé : flux en photons avec une résolution spectrale  $R = 11500$  et 5000 points.

## 6.3 Performances du prototype

Dans cette section, nous nous intéresserons aux performances du prototype sur le jeu de données GASS. Afin d'évaluer ces performances, le prototype utilise un certain nombre de configurations décrit dans la section 6.3.1. Les performances scientifiques sont présentées dans les sections 6.3.2 et 6.3.3. Les temps de calculs (ou performances numériques) sont présentés dans la section 6.3.4.

### 6.3.1 Configuration du prototype

Le prototype peut être exécuté dans différentes configurations. Ces configurations se distinguent par :

- Les conditions initiales, c'est-à-dire le choix des premiers jeux de coefficients de calibration.
- Le critère de sélection de référence, c'est-à-dire le choix des étoiles utilisées dans la calibration en longueur d'onde.
- La liste de standards, c'est-à-dire le choix des étoiles standards en vitesse radiale.

### Conditions initiales

Le prototype a été initialisé avec des conditions initiales de convergence. En effet, de premières lois de calibration en longueur d'onde sont initialisées en bruitant les vraies valeurs des coefficients de calibration. Les distributions utilisées pour bruite les lois de calibration ont une précision meilleure qu'un sample.

### Critère de sélection de référence

Contrairement aux précédentes études limitées à des étoiles G5V (c.f. les chapitres 4 et 5), plusieurs types spectraux sont considérés. Or, les performances des modules d'estimation des vitesses radiales et d'estimation des centroïdes des raies de référence sont très dépendantes de la morphologie des spectres : les deux modules ont besoin d'un certain nombre de raies bien contrastées du continu.

La morphologie d'un spectre est principalement influencée par la température effective et la métallicité de l'étoile :

- Dans le spectre d'une étoile chaude ( $T_{eff}$  élevée), le nombre de raies neutres est faible. De plus, plus la température effective de cette étoile augmente, plus les spectres de cette étoile sont dominés par les raies de l'Hydrogène (série de Paschen).
- Dans le spectre d'une étoile déficiente en métaux ( $[Fe/H] < 0$ ), les raies sont moins profondes. Plus la métallicité  $[Fe/H]$  est faible, moins les raies sont intenses.

Ces deux paramètres ont un impact sur le nombre de raies de référence détectables dans les spectres extraits. Or, moins il y aura de raies significatives dans les spectres, plus le domaine sur lequel sera contraint les lois de calibration sera réduit. Ainsi, la température effective et la métallicité vont avoir des conséquences sur les performances de la calibration en longueur d'onde.

Pour limiter les effets de la température effective et de la métallicité, la sélection des étoiles de référence est basée sur les deux critères suivants :

- **Un critère de température effective** : une étoile est qualifiée comme "référence" si  $3750 < T_{eff} < 6000$ .
- **Un critère de métallicité** : une étoile est qualifiée comme "référence" si  $[Fe/H] > -0.5$ .

La figure 6.8 illustre la sélection des étoiles de référence. Ces étoiles sont représentées sur le diagramme HR des données GASS. Le module de "Reference Selection" a sélectionné plus de la moitié des étoiles GASS. En effet, 49898 étoiles sont qualifiées comme "référence" pour la calibration en longueur d'onde. Le processus de calibration utilise 92097 spectres de ces étoiles

de référence. En moyenne sur les six mois de données GASS, une étoile de référence est observée un peu moins de deux fois.

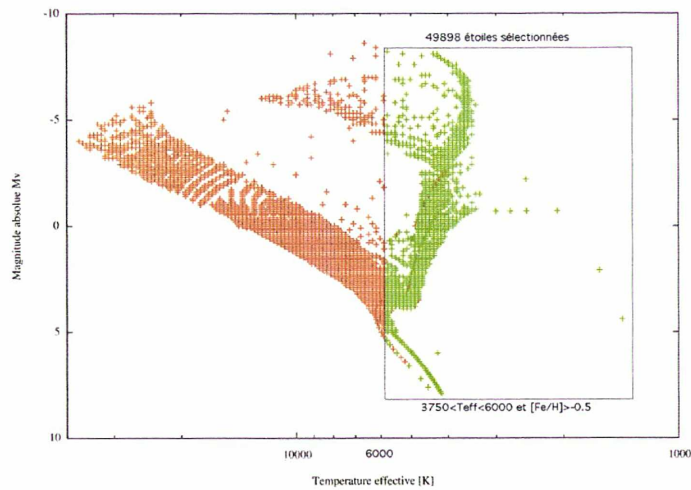


FIG. 6.8 – Diagramme HR des 6 mois de données GASS : les étoiles de référence (i.e. utilisées lors de la phase de calibration de SGIS) sont *en vert*.

La figure 6.9 présente la distribution en métallicité des données GASS. Sur cette distribution a été représentée la coupure réalisée lors de la sélection des références. Cependant, cette coupure ne prend pas en compte la sélection des températures.

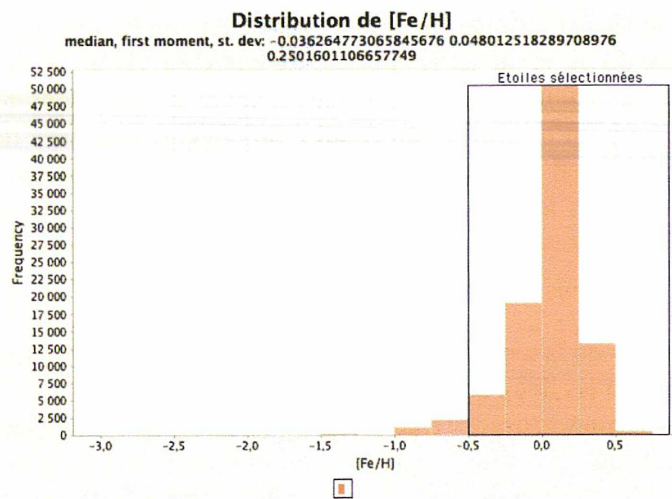


FIG. 6.9 – Distribution des métallicités des étoiles observées durant les 6 mois. Le cadre délimite la zone de métallicité utilisée dans la calibration en longueur d'onde SGIS.

Une prochaine étude considèrera à étendre le nombre de type spectraux pris en compte dans la phase de "Calibration Updating" et d'évaluer le comportement du processus itératif.

### Liste d'étoiles standards en vitesses radiales

Pour accélérer la convergence du processus de calibration en longueur d'onde, une liste d'étoiles standards peut être utilisée par le prototype. Pour chacune de ces standards, la vitesse radiale est supposée connue avec une précision meilleure que  $300 \text{ m.s}^{-1}$ . La phase d'estimation des vitesses radiales n'est donc pas nécessaire pour ces étoiles.

La mise en place de la liste de standards est indépendante du prototype (via un shell script). Chaque standard est "choisie" aléatoirement en se basant sur l'identifiant de l'étoile.

### 6.3.2 Performances sans standards

Les performances du prototype obtenues sur les six mois de données GASS sont présentées par diagnostic :

- L'évolution des erreurs moyennes sur les échelles en longueur d'onde est présentée dans la section 6.3.2. Ce diagnostic a été décrit dans la section 5.2.1 du chapitre 5.
- L'évolution des précisions des vitesses radiales est présentée dans la section 6.3.2. Ce diagnostic a été décrit dans la section 5.2.2 du chapitre 5.
- L'évolution des erreurs sur le centroïding est présentée dans la section 6.3.2.

### Évolution des erreurs sur les vitesses radiales moyennes

Les figures 6.10 et 6.11 présentent l'évolution de la dispersion des erreurs (ou précision) sur les vitesses radiales moyennes (en  $\text{km.s}^{-1}$ ) au cours du processus itératif (en rouge). La précision sur les vitesses radiales moyennes obtenue avec les vraies lois de calibration en longueur d'onde est représentée en vert.

D'après ce diagnostic, l'incertitude sur les vitesses radiales moyennes (i.e. vitesses radiales des sources) diminue au fur et à mesure des itérations. En effet, après la première itération, on passe d'une précision de  $2.32 \text{ km.s}^{-1}$  à  $1.49 \text{ km.s}^{-1}$ . Puis, la précision passe à  $1.16 \text{ km.s}^{-1}$  dans l'itération 2. Enfin, entre les itérations 6 et 7, l'évolution de la précision est faible : elle est inférieure à  $10 \text{ m.s}^{-1}$ .

Le diagnostic montre que la précision sur les vitesses radiales converge vers une valeur inférieure à  $1 \text{ km.s}^{-1}$ . À l'itération 16, cette précision est à environ  $60 \text{ m.s}^{-1}$  de la précision vraie égale à  $862 \text{ m.s}^{-1}$ . La différence entre les deux précisions est due à un faible nombre d'itérations exécuté lors du processus de calibration en longueur d'onde. Après un nombre infini d'itérations, on peut supposer que la précision sur les vitesses radiales soit égale à la précision vraie.

### Évolution des erreurs sur les échelles en longueur d'onde

Les figures 6.12, 6.13, 6.14 et 6.15 représentent les résultats obtenus lors du diagnostic sur les calibrations en longueur d'onde aux itérations 0, 1, 4 et 9. Tous ces résultats ont été représentés en utilisant une échelle commune : les erreurs moyennes sont représentées sur le domaine



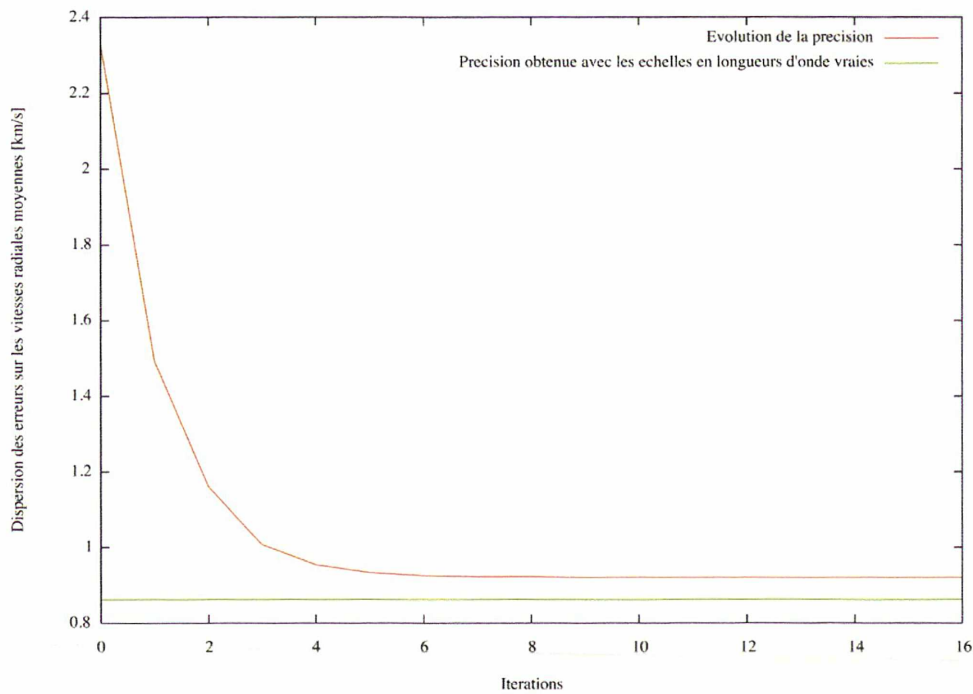


FIG. 6.10 – Résultats du diagnostic sur les vitesses radiales : évolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. La précision des vitesses radiales moyennes obtenue avec les échelles en longueur d'onde vraies (ou "précision vraie") est également représentée en vert.

$[-0.06, 0.06]$  nm. Cette échelle commune permet de visualiser l'évolution des erreurs moyennes au cours des itérations.

Le diagnostic à l'itération 0 (c.f. figure 6.12) met en évidence la dégradation des premières lois de calibration en longueur d'onde. En effet, une distribution gaussienne de dispersion inférieure au sample a été utilisée pour bruitez les vraies lois de calibration. La valeur de cette dispersion se retrouve sur le diagnostic : la dispersion des erreurs sur les échelles en longueur d'onde semble être inférieure à 0.026 nm, c'est-à-dire inférieure à un sample. Ce diagnostic est compatible avec le diagnostic sur les vitesses radiales : à la première itération, les dispersions sont environ égales à 0.2 sample (i.e. 0.006 nm), soit environ  $2 \text{ km.s}^{-1}$  que l'on retrouve au niveau de la précision sur les vitesses radiales.

Après la première itération, des nouvelles lois de calibration en longueur d'onde ont été calculées dans la phase de "Calibration Updating". D'après le diagnostic sur l'échelle en longueur d'onde à cette itération 1, les erreurs moyennes sont moins dispersées que dans l'itération précédente : la dispersion passe de 0.2 sample à environ 0.08 sample. Cette dispersion continue de diminuer d'itération en itération (c.f. figures 6.14 et 6.15).

Via les erreurs sur les échelles en longueur d'onde, ce diagnostic montre que le processus itératif converge vers une solution de calibration. Après 4 itérations, les erreurs sur les échelles

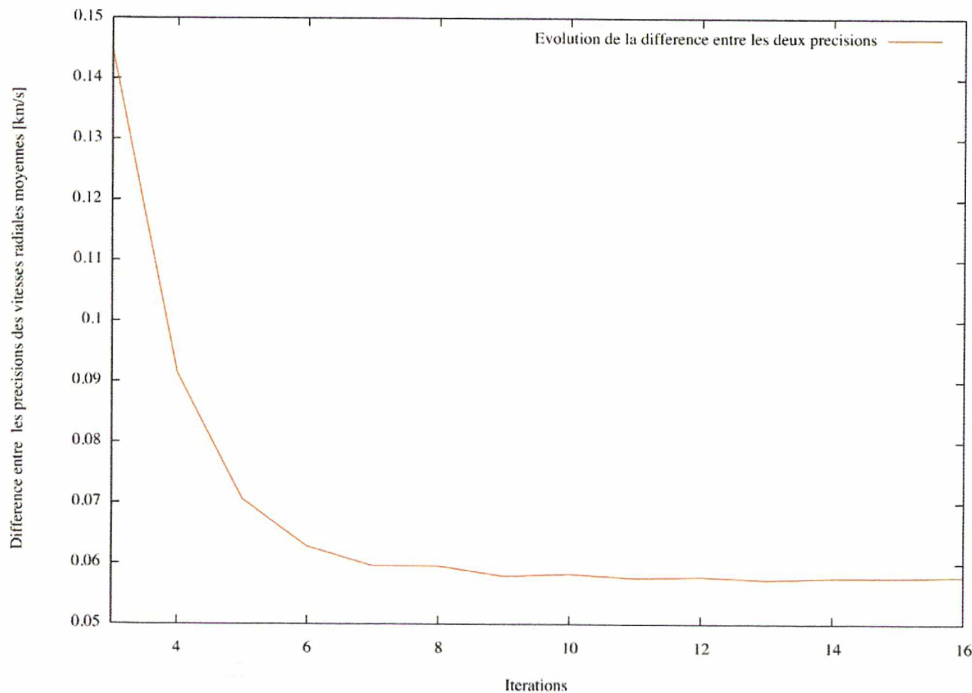


FIG. 6.11 – Résultats du diagnostic sur les vitesses radiales : évolution de la différence entre la précision des vitesses radiales moyennes et la précision vraie en fonction des itérations.

en longueur d'onde évoluent très peu : la dispersion sur l'ensemble des unités de calibration est inférieure à  $100 \text{ m.s}^{-1}$ .

### Évolution des erreurs de centroïding

Afin de contrôler le bon comportement de la méthode de centroïding, un diagnostic supplémentaire a été développé. Ce diagnostic va calculer l'écart entre l'estimation ( $\lambda, s$ ) et sa vraie valeur en utilisant la vraie échelle en longueur d'onde.

Les figures 6.16 et 6.17 présentent les résultats du diagnostic sur le centroïding aux itérations 1, 2 et 5 pour l'unité de calibration 8. Les erreurs de centroïding sont représentées en fonction des longueurs d'onde des raies de référence. Les longueurs d'onde de ces raies tiennent compte des vitesses radiales moyennes (i.e. vitesses radiales des sources) estimées dans la phase de "Source Updating". La dispersion des longueurs d'onde est directement reliée à la dispersion des vitesses radiales moyennes.

À partir de ces figures, on constate que le centroïding permet d'obtenir des estimations ( $\lambda, s$ ) pour 12 ou 13 raies de référence dans les spectres extraits de l'unité de calibration 8. Ces raies de référence sont bien réparties sur le domaine en longueur d'onde du RVS (i.e.  $[847,874] \text{ nm}$ ).

Les erreurs de centroïding sont dispersées sur le domaine  $[-1.6, 1.6]$  samples à l'itération 1. Une raie de référence pose problème : quelques unes des estimations pour la raie du fer à 868

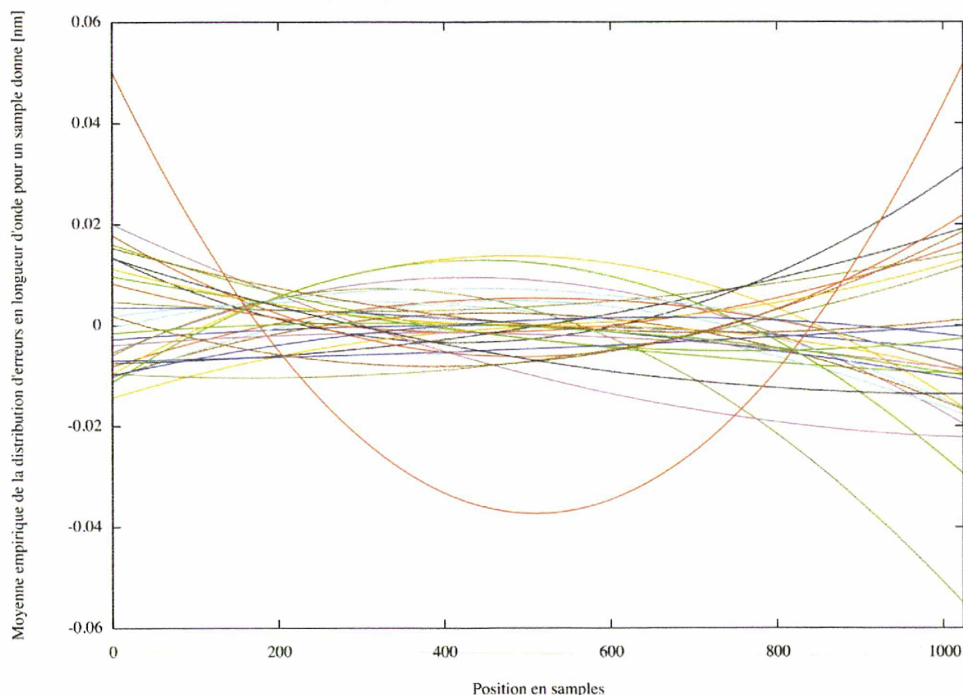


FIG. 6.12 – Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itération 0. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

$nm$  ont des erreurs supérieures à 10 samples (i.e.  $0.26 nm$ ). En fait, cette raie est très proche de deux autres raies de référence. De plus, pour la première itération, le centroïding utilise une loi de calibration moins précise (c.f. figure ??). Ainsi, le centroïding a fait une confusion entre cette raie du fer à  $868 nm$  et une autre raie du fer à  $869 nm$ . Cependant, cette confusion s'est produite sur moins de 5 estimations. Sachant qu'on a plus de 1000 estimations par raie de référence, ces erreurs de confusion auront très peu d'impact sur le reste du processus itératif.

À l'itération suivante, ce domaine se rétrécit : les erreurs sont dispersées sur le domaine  $[-1, 1]$  sample. Entre les itérations 2 et 5 (c.f. figure 6.17), la dispersion des erreurs a eu une évolution moins forte qu'entre les itérations 1 et 2 (c.f. figure ??) : il y a une différence de l'ordre de 0.1 sample entre les itérations 2 et 5. La confusion entre les deux raies du fer à  $868 nm$  et à  $869 nm$  disparaît aux itérations suivantes.

L'estimation de centroïding est directement reliée aux derniers jeux de coefficients de calibration. En effet, la méthode de centroïding estime une loi locale de calibration en longueur d'onde pour chaque raie de référence. Pour ce faire, elle utilise les derniers jeux de coefficients de calibration calculés durant la phase de "Calibration Updating". Par conséquent, si le processus itératif converge, alors les estimations de longueur d'onde du centre discret des raies vont s'améliorer d'itération en itération.

Ce diagnostic a montré que les erreurs de centroïding diminuent au fur et à mesure des

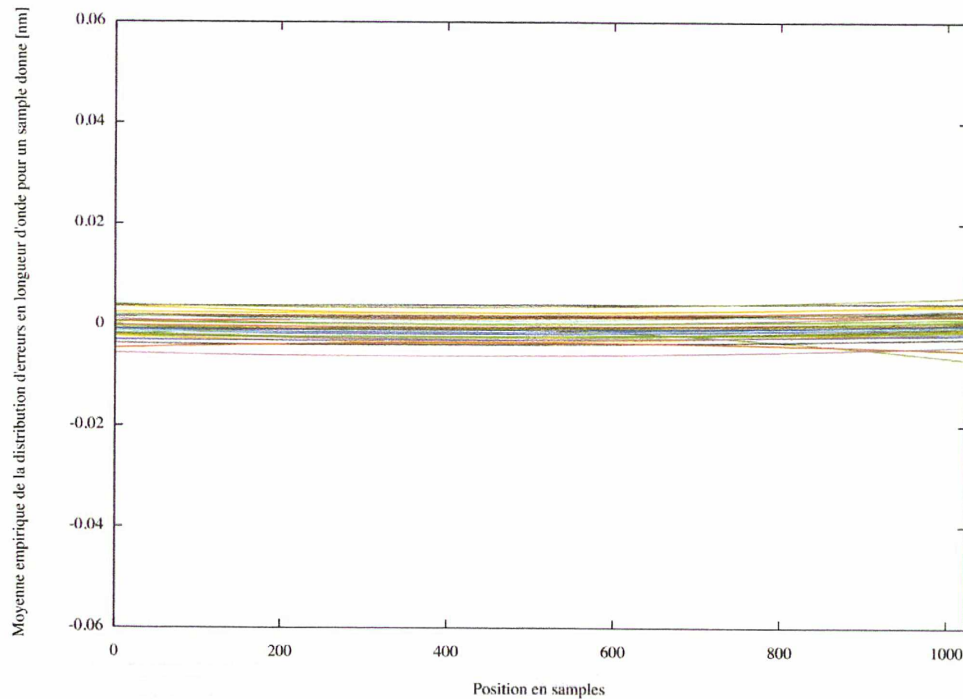


FIG. 6.13 – Diagnostics d’erreurs sur les échelles en longueur d’onde à l’itération 1. Ce diagnostic représente l’erreur moyenne sur la longueur d’onde centrale associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

itérations. Ces erreurs de centrage ont une dispersion de quelques dixièmes de sample et sont compatibles avec la convergence du processus itératif (c.f. section 5.3 dans le chapitre 5). De plus, ces erreurs sont compatibles avec les résultats des deux autres diagnostics (c.f. sections 6.3.2 et 6.3.2).

### 6.3.3 Performances avec standards

Dans cette section, je présente les premières performances obtenues en utilisant une liste d’environ 1000 étoiles standards dans le processus itératif. Les diagnostics, présentés dans les sections 6.3.3 et 6.3.3, prennent en compte ces étoiles standards. Ils donnent les performances du prototype à la fin des six mois.

Pour évaluer correctement l’impact des standards sur la convergence du processus itératif, ces standards ne doivent pas intervenir dans les diagnostics. Cette étude d’influence des standards sur la convergence doit être complétée.

#### Évolution des erreurs sur les vitesses radiales moyennes

La figure 6.18 présente une superposition entre les évolutions de la dispersion des erreurs sur les vitesses radiales moyennes sans (en rouge) et avec (en vert) standards.

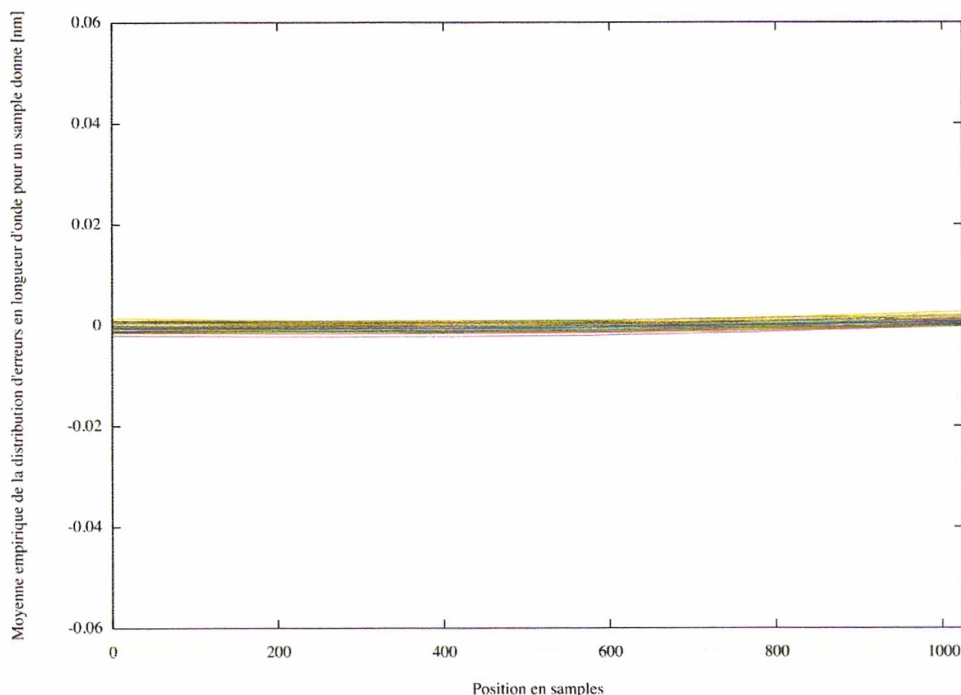


FIG. 6.14 – Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itérations 4. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

À la première itération, l'incertitude sur les vitesses radiales avec standards est supérieure à celle sans standard. Cet écart provient du fait que les conditions initiales ne sont pas les mêmes dans les deux cas : dans le cas avec standards, les premiers jeux de coefficients ont été plus bruités lors de la phase d'initialisation.

D'après ce diagnostic, l'incertitude sur les vitesses radiales moyennes avec standards diminue plus rapidement que sans standard. En effet, après la première itération, on passe d'une précision de  $3.21 \text{ km.s}^{-1}$  à  $1.82 \text{ km.s}^{-1}$ , soit une diminution d'environ 45%. Sans standard, la diminution de la dispersion après la première itération est d'environ 35%. Puis, la précision passe à  $1.06 \text{ km.s}^{-1}$  dans l'itération 2.

Le diagnostic montre que la précision sur les vitesses radiales converge vers une valeur d'environ  $0.30 \text{ km.s}^{-1}$ . Ces performances, atteintes sur 6 mois de données, sont en accord avec les performances du RVS après les cinq années de mission de Gaia, présentées dans la table 1.5 du chapitre 1.

### Évolution des erreurs sur les échelles en longueur d'onde

Les figures 6.19, 6.20, 6.21 et 6.22 représentent les résultats obtenus lors du diagnostic sur les calibrations en longueur d'onde aux itérations 0, 1, 3 et 6. Ces résultats permettent de visualiser l'évolution des erreurs moyennes sur les échelles en longueur d'onde au cours des itérations.

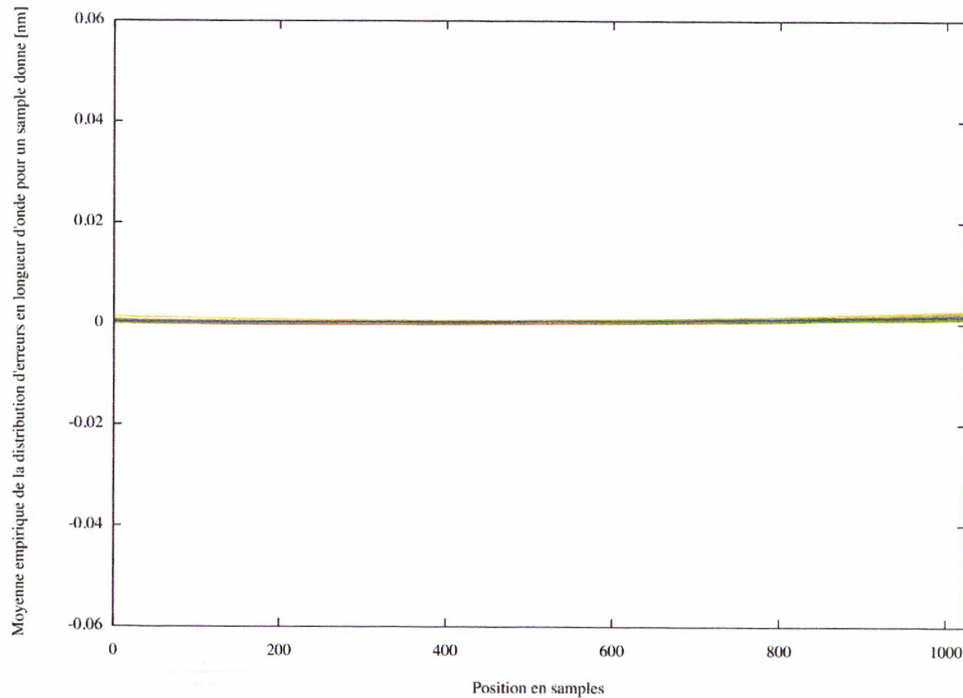


Fig. 6.15 – Diagnostics d’erreurs sur les échelles en longueur d’onde à l’itérations 9. Ce diagnostic représente l’erreur moyenne sur la longueur d’onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

Le diagnostic à l’itération 0 (c.f. figure 6.19) met en évidence la dégradation des premières lois de calibration en longueur d’onde. Ce diagnostic est compatible avec le diagnostic sur les vitesses radiales : à la première itération, les dispersions sur l’ensemble des unités de calibration sont environ égales à 0.4 sample (i.e. environ 0.01 nm), soit un peu plus de  $3 \text{ km.s}^{-1}$  que l’on retrouve au niveau de la précision sur les vitesses radiales.

Ce diagnostic montre que le processus itératif converge vers une solution de calibration. Après 5 itérations, les erreurs sur les échelles en longueur d’onde évoluent très peu, c’est-à-dire de quelques  $\text{m.s}^{-1}$ .

### 6.3.4 Temps de Calcul

Les performances avec standards, présentées dans la section 6.3.3, ont été obtenues sur un ordinateur biprocesseurs Xeon 2.20 GHz possédant 2Go de RAM. Le prototype passe environ 9h30 en temps CPU par itération :

- **Phase d’initialisation** : 11 minutes en temps CPU (une fois par exécution du prototype).
- **Phase d’application des calibrations en longueur d’onde** : 1h en temps CPU (i.e. 0.03 seconde par spectre)

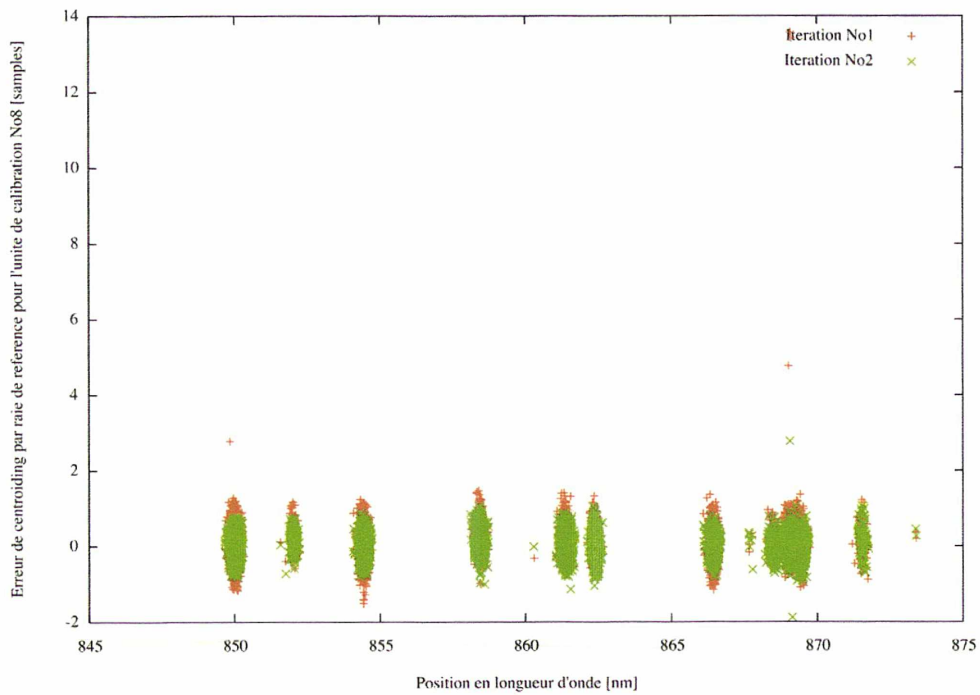


FIG. 6.16 – Erreurs de centroïding en fonction des longueurs d'onde des raies de référence.

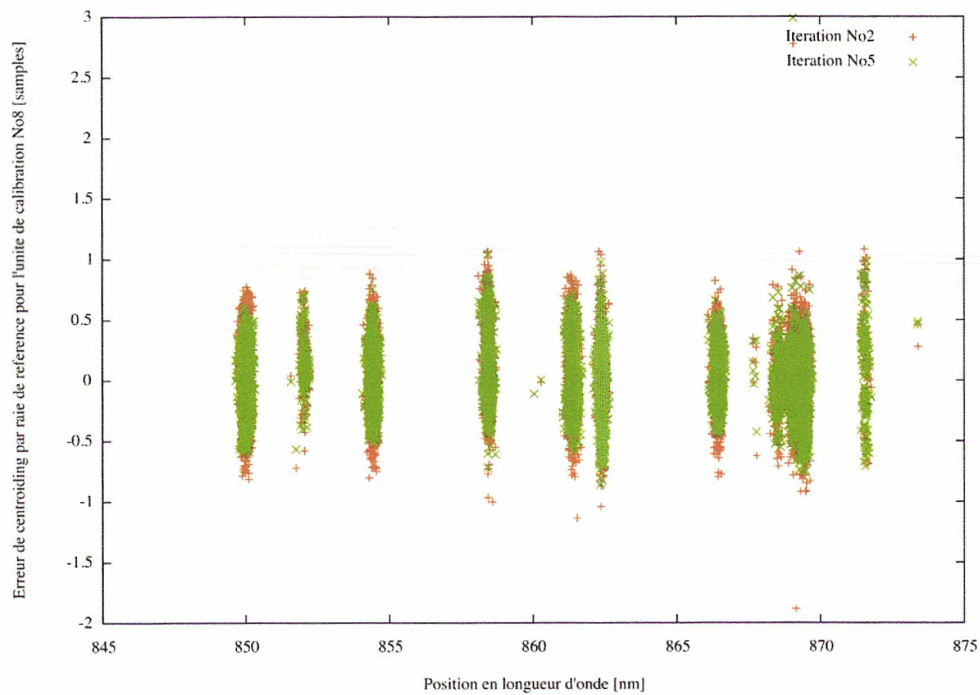


FIG. 6.17 – Erreurs de centroïding en fonction des longueurs d'onde des raies de référence.

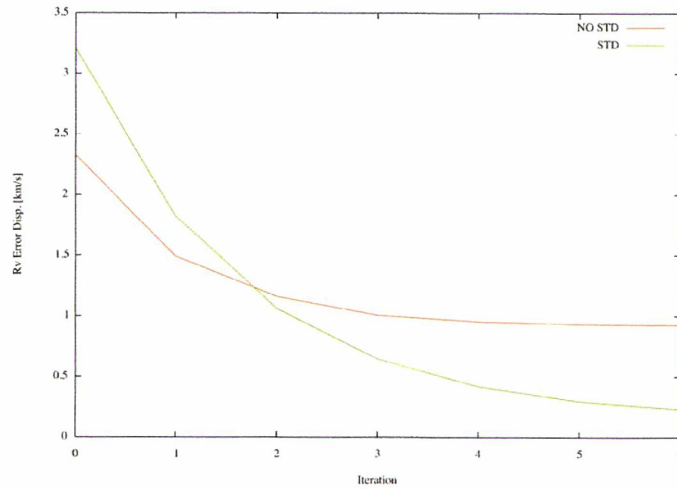


FIG. 6.18 – Résultats du diagnostic sur les vitesses radiales : *en rouge*, évolution de la précision des vitesses radiales moyennes au cours des itérations sans standard - *en vert*, évolution de la précision des vitesses radiales moyennes au cours des itérations avec 1000 standards.

- **Phase d'estimation des vitesses radiales** : 3h15 en temps CPU (i.e. 0.12 seconde par spectre)
- **Phase de calibration en longueur d'onde** : 5h en temps CPU (i.e. 0.20 seconde par spectre)

Cependant, le temps d'exécution (ou "elapsed time" en anglais) d'une itération est de l'ordre de 24h. Ceci montre que le prototype est dominé par les entrées-sorties : il lit et met constamment à jour des données sur le disque dur.

### 6.3.5 Synthèse

En utilisant les six mois de données GASS sans étoiles standards, tous les diagnostics intermédiaires du prototype ont montré la convergence du processus itératif. Après 6 ou 7 itérations, les performances du prototype évoluent très peu (i.e. quelques dizaines de mètres par seconde). Après ces itérations, la précision des vitesses radiales moyenne atteinte est environ égale à  $920 \text{ m.s}^{-1}$ . Les bruits de mesure, c'est-à-dire les erreurs ne comprenant pas de terme de calibration (e.g. bruit de photon), sont de l'ordre de  $860 \text{ m.s}^{-1}$ . La différence entre la précision sur les vitesses radiales moyennes et les bruits de mesure est de l'ordre de  $60 \text{ m.s}^{-1}$ . La précision sur les échelles en longueur d'onde est égale à environ  $100 \text{ m.s}^{-1}$  après l'itération 6.

Avec un grand nombre d'étoiles standards (i.e. 1000), les précisions des vitesses radiales moyennes sont meilleures d'un facteur 3, c'est-à-dire elles sont de l'ordre de  $300 \text{ m.s}^{-1}$  par rapport aux  $920 \text{ m.s}^{-1}$  obtenus sans standards. La précision sur les échelles en longueur d'onde est égale à environ  $60 \text{ m.s}^{-1}$  après l'itération 6.

Dans le prototype de SGIS, une étoile standard en vitesse radiale est caractérisée par :

- Une vitesse radiale bien définie et constante d'une itération à une autre. En effet, cette vitesse radiale est supposée connue à partir de mesures effectuées au sol.



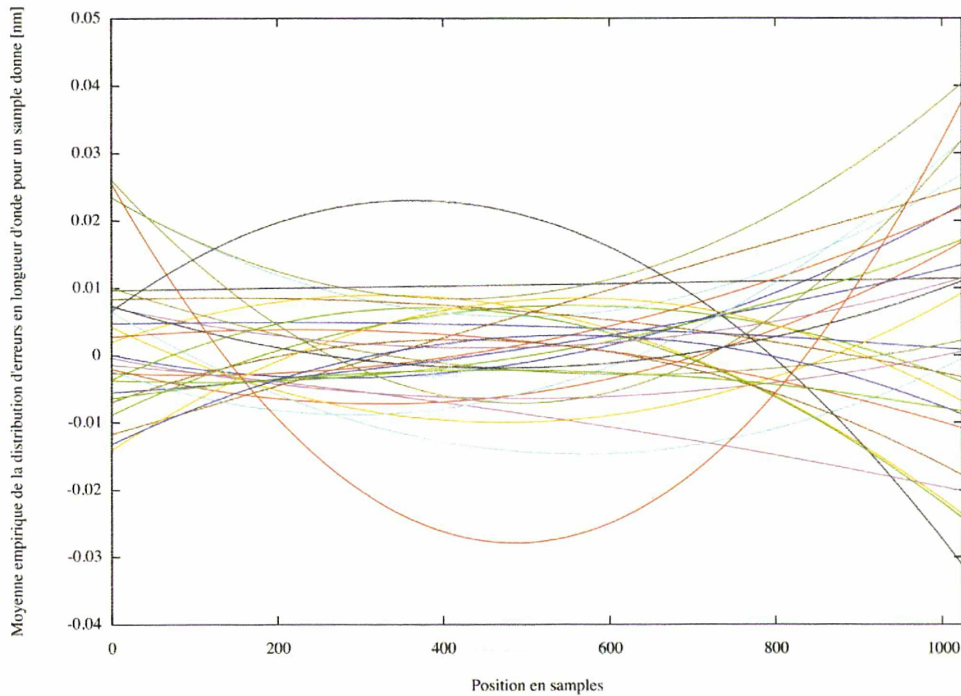


FIG. 6.19 – Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 0. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

- Une précision sur les vitesses radiales meilleure que  $300 \text{ m.s}^{-1}$ .

La précision sur les vitesses radiales de ces étoiles est meilleure que les bruits de mesure (i.e.  $860 \text{ m.s}^{-1}$ ). Étant prises en compte lors des diagnostics d'erreurs, leurs grandes précisions sur leurs vitesses radiales dominent les résultats des diagnostics. De plus, les performances sans et avec standards n'ont pas été obtenues de la même manière : les conditions initiales ne sont pas identiques. Ainsi, les résultats présentés ne permettent pas de quantifier l'apport de ces étoiles standards aux propriétés de convergence du processus de calibration. Pour ce faire, il faudrait réaliser les opérations suivantes :

- Exclure les standards des diagnostics d'erreurs du prototype.
- Fixer un unique point de départ pour le processus itératif : les conditions initiales doivent être identiques d'une exécution du prototype à une autre pour pouvoir comparer les performances.

Une fois les performances du prototype comparables, on va exécuter le prototype SGIS avec plusieurs listes de standards : 0, 1, 10, 100 et 1000 étoiles standards. Les performances résultantes vont nous permettre de comprendre et de quantifier l'apport des standards sur la convergence du processus itératif.

Le prototype est un processus très coûteux en temps d'exécution : une itération sur 6 mois de données GASS coûte 9H30 de temps CPU, et environ 24h de temps d'exécution. Le pro-

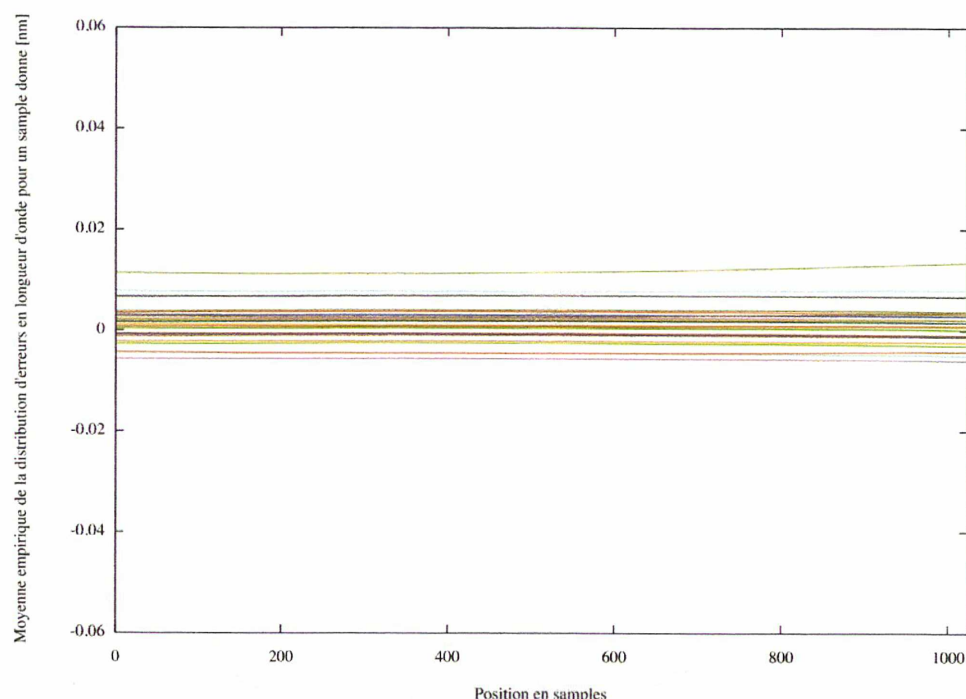


FIG. 6.20 – Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 1. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

prototype développé est donc limité par les entrées-sorties sur le disque dur. Au début et à la fin de chaque phase (i.e. "Initialisation", "Source Updating", "Reference Selection" et "Calibration Updating"), le prototype ouvre et referme un grand nombre de fichiers sur le disque. Toutes ces opérations sont très coûteuses en temps d'exécution. Pour optimiser le prototype, on pourrait limiter les opérations de lecture et écriture en utilisant la mémoire vive de l'ordinateur (sans la saturer).

Le prototype de calibration en longueur d'onde a été développé en prenant en compte des hypothèses simplificatrices. Les performances obtenues par ce prototype ont démontré que le processus de calibration en longueur d'onde converge en un nombre d'itérations raisonnable (i.e. environ 6 itérations). Maintenant, les études à venir doivent être faites dans un cadre de plus en plus réaliste. En enlevant une par une les hypothèses simplificatrices, il va falloir vérifier que le processus itératif continue à converger en un nombre d'itérations raisonnable. Ainsi, pas à pas, la méthode de calibration et les données simulées devront tenir compte des problèmes réels suivants :

1. effets des radiations (i.e. "radiation damage" en anglais).
2. effets des calibrations couplées.
3. effets des mouvements transverses.
4. effets du nombre de données observées.

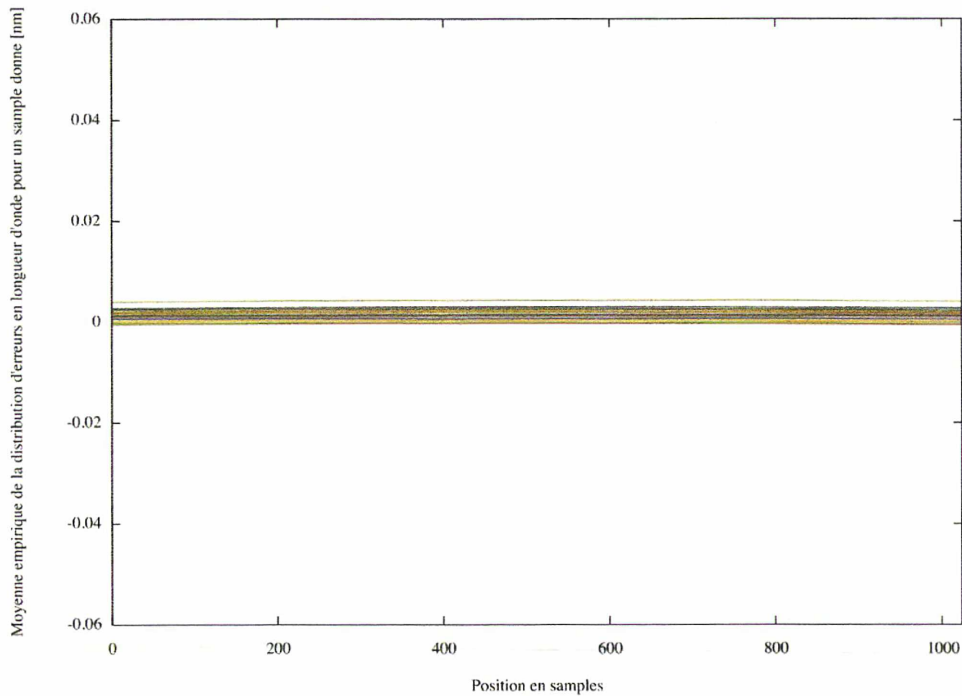


FIG. 6.21 – Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 3. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

Ces problèmes ont été listés par ordre d'importance sur la calibration en longueur d'onde.

En L2, le satellite va être bombardé par des cosmiques. Ces cosmiques vont endommager les CCDs du plan focal de Gaia. La dégradation causée par ces cosmiques augmente tout au long de la mission. Elle provoque des "trappes" au niveau des capteurs. Ces trappes vont piéger des charges et vont les "libérer" sur différentes échelles de temps. Cette effet, appelé "radiation damage", a pour conséquence une baisse de flux dans les observations et aussi, une modification du profil instrumental de la LSF. Or, la calibration en longueur d'onde ajuste le profil de raies dans le spectre. Ainsi, la moindre modification de la LSF va se répercuter sur le profil de ces raies. Les "radiation damages" vont donc avoir des conséquences sur la calibration en longueur d'onde. De plus, les étoiles, suivant leur brillance, vont plus ou moins "remplir" les trappes. Les conséquences du "radiation damage" sur les spectres vont donc dépendre de la source observée et de celles observées précédemment. En incluant les "radiation damages", le processus de calibration en longueur d'onde devient plus complexe. Au lieu de ne prendre en compte que l'instrument, il va falloir aussi prendre en compte l'histoire passée des observations (i.e. savoir quelle étoiles ont été lues sur le CCD, leur magnitude, etc.). Cela va entraîner des changements fondamentaux au niveau du concept de la méthode SGIS développée. De part sa complexité, l'étude des "radiation damages" doit être l'une de nos priorités. Cependant, l'industriel EADS Astrium ne nous a pas encore fourni des informations concernant ces "radiation damages". Sans évaluation, les "radiation damages" ne peuvent être inclus de manière réaliste dans les simulations de données Gaia.

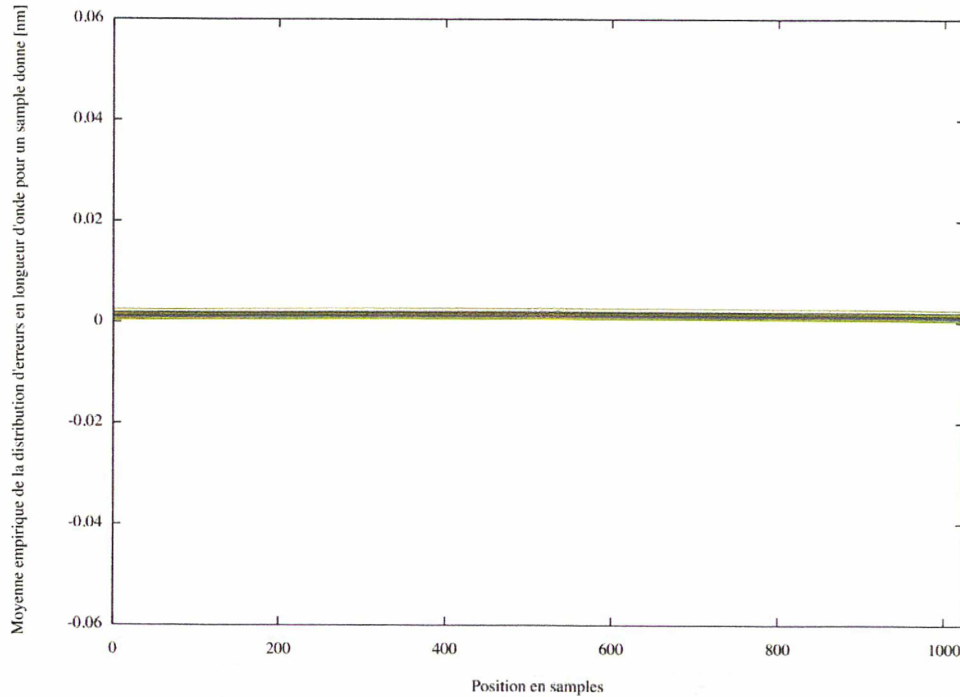


FIG. 6.22 – Diagnostics d’erreurs sur les échelles en longueur d’onde aux itérations 6. Ce diagnostic représente l’erreur moyenne sur la longueur d’onde associée à chaque échantillon du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée.

La calibration en longueur d’onde n’est pas la seule calibration de l’instrument RVS. En effet, la LSF de l’instrument, la réponse du système télescope+RVS, les détecteurs (bruit de lecture, etc.) doivent aussi être calibrés. Certaines de ces calibrations sont couplées. Par exemple, la calibration en longueur d’onde et la calibration de la LSF sont couplées. Ainsi, les futures études devront prendre en compte tous les aspects de calibration avec une attention particulière sur ces problèmes de couplage.

De plus, les sources des données GASS n’ont qu’un mouvement AL (i.e. "Along scan") sur le plan focal. Or, dans la réalité, les sources auront un mouvement AL et un mouvement transverse (ou AC, i.e. "Across scan"). Ce mouvement transverse est causé par la loi de balayage complexe de Gaia. Il va entraîner un élargissement de la LSF de l’instrument et donc, un élargissement du profil des raies dans les spectres. Cet élargissement est dépendant du temps. Ainsi, les prochaines études devront inclure dans la formule de la loi de calibration (c.f. équation 2.4 du chapitre 2) une dépendance suivant la dérivée première des coordonnées  $\eta$  des étoiles dans le champ de vue de l’instrument.

Les tests, présentés dans ce chapitre, ont été réalisés sur 6 mois de données simulées. En 6 mois, chaque source a, en moyenne, 2 transits. L’estimation des vitesses radiales moyennes est basée sur ce nombre de transits. Or, en fin de mission, 5 années d’observations se seront écoulées. Une source sera observée, en moyenne, 40 fois, soit 10 fois plus que pendant les 6 mois. L’estimation des vitesses radiales sera donc basée sur 10 fois plus de transits que pour 6 mois de données. La précision de la calibration en longueur d’onde sera plus de 4 fois meilleure

que celle obtenue avec les 6 mois de mission. Ainsi, les prochaines études doivent prendre en compte un volume de données de plus en plus important, jusqu'à atteindre la totalité de la mission.

## Conclusion et perspectives

Les instruments du satellite Gaia vont observer en continu la sphère céleste durant les cinq années de mission. Les observations scientifiques ne pourront pas être interrompues pour des besoins de calibration ou de contrôle du bon comportement de chaque instrument. De plus, il n'y a aucun dispositif de calibration (e.g. lampe de calibration) à bord du satellite Gaia. Par conséquent, la phase de calibration nécessite le développement d'une "nouvelle" méthode numérique de calibration.

L'objectif de la thèse a été de mettre en oeuvre et de valider cette nouvelle méthode de calibration. Elle se nomme la "Spectroscopic Global Iterative Solution" (SGIS). Elle utilise la stabilité en vitesse radiale de certaines étoiles observées afin d'estimer l'évolution des caractéristiques de l'instrument.

Durant mes recherches, j'ai développé et implémenté en Java un prototype de SGIS. Il m'a permis d'étudier la méthode SGIS appliquée à la calibration en longueur d'onde.

Un des points cruciaux de l'implémentation du prototype de SGIS est le développement et le choix d'un algorithme de détection et d'identification de raies de référence dans un spectre. Cet algorithme va fournir des triplets  $(\lambda, \eta, \zeta)$  nécessaires à la résolution du système linéaire des calibrations en longueur d'onde. Deux types d'algorithmes ont été développés : une méthode de localisation du centre des raies de référence par ajustement parabolique et une méthode de localisation par ajustement de profil. Un bilan comparatif a été réalisé. Ce bilan a mis en évidence des erreurs systématiques au niveau du premier algorithme. Le second algorithme a un comportement satisfaisant, c'est-à-dire, on arrive à avoir un biais sur l'estimation des triplets  $(\lambda, \eta, \zeta)$  inférieur au centième de sample près.

Pour découpler les difficultés, on a développé, en parallèle de la méthode de localisation du centre des raies de référence, un prototype simulant l'algorithme de détection et de localisation des raies de référence (ou centroïding). L'implémentation simule cette étape en bruitant les triplets vrais  $(\lambda_0, \eta_0, \zeta_0)$  avec une distribution Gaussienne. On a étudié l'influence du biais et de la dispersion de la distribution de bruits ajoutée aux triplets sur la convergence du processus de calibration. Cette étude a montré que la convergence est atteinte pour un algorithme de centroïding ayant :

- un biais inférieur ou égal à 0.001 sample.
- une dispersion inférieure ou égale à 0.5 sample.

Cette étude nous a donc conduit à sélectionner la méthode de centroïding par ajustement de

profil, qui remplit les deux conditions ci-dessus.

Enfin, mes travaux se sont conclus par une première évaluation des performances du prototype de calibration en longueur d'onde SGIS. Ces performances ont été obtenues à partir de données "réalistes" GASS. Le processus itératif converge vers une solution de calibration très proche de la solution vraie : quelques dizaines de mètres par seconde les séparent en 6 itérations. Ces performances sont encourageantes, mais pas suffisantes. En effet, elles ont été obtenues dans des conditions assez bonnes : pas de radiations, pas de distorsions optiques, les autres calibrations sont parfaitement déterminées, etc. Il faut continuer à améliorer le prototype en prenant en compte des effets de plus en plus réalistes.

### Études d'influence

Dans la continuité de l'étude de convergence et des premières évaluations des performances du prototype de calibration, plusieurs études peuvent être mises en place :

- **Influence du nombre de standards en vitesse radiale** : on veut tester l'impact qu'ont les étoiles standards sur la convergence du processus itératif de calibration en longueur d'onde. Ainsi, on veut démontrer que le nombre de standards augmente la rapidité de convergence du processus. Ces tests sont en cours de mise en place au GEPI.
- **Influence des types spectraux de "référence"** : on veut tester l'impact de la sélection d'étoiles moins adaptées à la calibration sur la convergence du processus itératif (i.e. savoir si elles dégradent les performances du prototype ou non). Cette sélection est faite au niveau des critères de sélection des étoiles de "référence". Le principe de cette étude serait d'élargir les types spectraux des étoiles de "référence" : au lieu de sélectionner les étoiles de température effective  $3750\text{K} < T_{eff} < 6000\text{K}$ , on élargit le domaine à  $3750\text{K} < T_{eff} < 9000\text{K}$ , par exemple. On peut procéder de la même manière sur le domaine de sélection en métallicité. Cette étude est utile dans le cadre du développement de la phase de "Reference Selection".
- **Influence de la durée de l'unité de calibration** : on veut faire varier la durée d'une unité de calibration et d'observer l'impact causé sur la convergence de la méthode SGIS. Cette durée est directement liée au nombre d'observations : plus l'unité de calibration est grande, plus il y aura d'observations. Cette étude permettrait de déterminer un nombre minimum d'observations pour une unité de calibration pour lequel le processus converge.

Les deux dernières études proposées sont très fortement corrélées : d'une part, l'élargissement des types spectraux peut augmenter le nombre d'observations par unité de calibration, d'autre part, la diminution de la durée d'une unité de calibration va réduire ce nombre d'observations. Il risque d'y avoir plusieurs solutions possibles aux problèmes posés. Ainsi, ces deux études doivent être menées conjointement.

---

## Méthode de calibration directe

Suite à l'expérience accumulée dans le domaine de la calibration en longueur d'onde SGIS, une solution directe de calibration est en cours de développement par M. Cropper, S. Rosen, H. Huckle et leurs collaborateurs, à Mullard Space Science Laboratory (UK). Pour une unité de calibration  $k$ , cette solution consiste estimer en même temps les coefficients  $C(k)_{mn}$  de calibration en longueur d'onde et les vitesses radiales  $v$  des observations (i.e. spectres). En utilisant les mêmes hypothèses et la même forme polynomiale de calibration présentées dans la section 2.4.3 du chapitre 2, la calibration en longueur d'onde consiste à minimiser, par une méthode de moindres carrés, le critère  $S_k$  suivant :

$$S_k = \sum_{spec=1}^{N_{spectres}} \sum_{l=1}^{N_{lines}} \left[ \lambda(l) \left( 1 + \frac{v(spec)}{c} \right) - \sum_{m=0}^2 \sum_{n=0}^2 C(k)_{mn} \eta(spec, l)^m \zeta(spec, l)^n \right]^2 \quad (1)$$

où  $spec$  représente le spectre d'une étoile observée durant l'unité de calibration  $k$ ,  $N_{spectres}$  le nombre de spectres observés durant l'unité de calibration  $k$ ,  $l$  une raie de référence du spectre observé  $spec$ ,  $N_{lines}$  le nombre de raies de référence dans le spectre  $spec$ ,  $\lambda(l)$  la longueur d'onde de la raie de référence  $l$  du spectre  $spec$  dans le référentiel du laboratoire,  $\eta(spec, l)$  et  $\zeta(spec, l)$  les coordonnées angulaires de la raie de référence  $l$  du spectre  $spec$ . Cette formule provient de la note technique Huckle et al. (2007).

La méthode SGIS estime et utilise la vitesse radiale moyenne des sources. L'erreur sur cette vitesse radiale dépend du nombre d'observations (i.e. elle évolue en  $\frac{1}{\sqrt{N_{observations}}}$ ). Ainsi, on a besoin d'un grand nombre d'observations pour pouvoir calibrer correctement le RVS. La méthode directe détermine la vitesse radiale d'une seule observation. L'avantage d'une telle méthode est de pouvoir n'utiliser qu'un petit nombre d'observations. On pourrait imaginer utiliser la méthode directe pour exécuter des calibrations journalières et SGIS pour des calibrations semestrielles.

Les deux solutions proposées sont numériquement très proches. Le prototype existant va servir de cadre pour le développement de cette solution directe : la structure générale du prototype est gardée, seuls certains modules devront être adaptés à la nouvelle méthode. Les solutions itérative et directe vont donc être intégrées dans un seul et même prototype. On espère avoir un bilan comparatif des deux solutions avant juin 2008 (i.e. fin du cycle 4 de développement de la CU6).





# Table des figures

1	Progrès de la précision astrométrique au cours du temps. Erik Høg . . . . .	1
1.1	Planning de la mission Gaia, ESA . . . . .	6
1.2	Intérêts scientifiques de Gaia : (a) Domaines scientifiques abordés par Gaia (ESA) (b) Précision de Gaia dans la Galaxie (Observatoire de Lund, Suède) . . . . .	8
1.3	Représentations typique d'orbites de transfert ( <i>à gauche</i> ) et de Lissajous ( <i>à droite</i> ) pour la mission Gaia (Hechler et al. (1999)) . . . . .	9
1.4	Représentations du satellite Gaia (EADS Astrium, 2006) . . . . .	9
1.5	Principe d'observation de Gaia : la loi de balayage. ESA, 2005 . . . . .	10
1.6	Nombre d'observations de la sphère céleste en fonction des coordonnées écliptiques (J. de Bruijne, ESA, 2003) . . . . .	11
1.7	Charge utile : (a) Deux télescopes illuminant le même plan focal (en haut), (b) Aperçu de la charge utile assemblée (en bas) (EADS Astrium, 2006) . . . . .	12
1.8	Plan focal du satellite Gaia, EADS Astrium 2006 . . . . .	13
1.9	Champs observés par RAVE, projetés en coordonnées galactiques. Sur la figure centrale, les gradients de vitesses radiales héliocentriques sont représentés avec un code de couleur, mettant en valeur le mouvement du Soleil et la projection de différentes composantes de vitesse le long de la ligne de visée. La ligne jaune représente l'équateur céleste. Le cadre en haut à gauche représente les champs de vue observés avec un code de couleur : en rouge, les champs observés une fois, en jaune les champs observés deux fois, en vert pour trois fois et en marron pour quatre fois. Steinmetz et al (2006) . . . . .	15
1.10	Projet JASMINE : vue du satellite. Le design de ce satellite est très similaire à celui de Gaia. (Site web de JASMINE : <a href="http://www.jasmine-galaxy.org/whatis.html">www.jasmine-galaxy.org/whatis.html</a> ) . . . . .	16
1.11	Exemple de détection d'une binaire à éclipse au niveau des spectres dans la bande RVS, Munari et al (2001) . . . . .	17
1.12	Fonctionnement du spectromètre de Gaia. Le chemin optique du RVS est représenté en marron clair. (EADS Astrium Satellite 2006) . . . . .	17
1.13	Représentation des modes d'échantillonnage des CCDs du RVS : mode échantillonnage des étoiles brillantes (en haut) : 1 sample correspond à 1 pixel physique ; échantillonnage intermédiaire (au milieu) : 1 sample correspond à une colonne de 10 pixels ; échantillonnage des étoiles faibles (en bas) : 1 sample correspond à $3 \times 10$ pixels. . . . .	19

1.14	Spectre à haute résolution d'une géante G du disque, montrant trois régions spectrales étudiées lors de la définition du domaine spectral du RVS : région des doublets K I (en haut) et Na I (au milieu), et la région du triplet de Ca II (en bas). Les points noirs indiquent la position des doublets et triplets (Munari et Castelli (2000)) . . . . .	20
1.15	Exemples de spectres dans la bande RVS : une étoile K0 III pauvre en métaux (en haut à gauche) et une étoile K0 III (en haut à droite), une étoile G0 V (au milieu à gauche), une étoile F0 V (au milieu à droite), une étoile A0 V (en bas à gauche) et une étoile B0 V (en bas à droite) de métallicité solaire. Katz (2005)	21
2.1	Calibration en longueur d'onde dans la bande RVS : (a) spectre correctement calibré ; (b) spectre calibré avec un offset, i.e. perturbation d'ordre 0 ; (c) et (d) spectres calibrés avec de mauvais facteurs d'échantillonnage, i.e. perturbation d'ordre 1 à l'origine d'une compression (cas (c)) ou dilatation (cas (d)) de l'échelle en longueur d'onde. . . . .	24
2.2	Exemple de spectre théorique d'une lampe de calibration He/Ar. (ESO) . . . . .	26
2.3	Simulation d'une utilisation d'une cellule à Néon pour la calibration en longueur d'onde du RVS : <i>en haut</i> le spectre d'absorption d'une cellule à Néon, <i>au centre</i> le spectre d'une étoile F3 V, et <i>en bas</i> le spectre de la même étoile observée à travers la cellule à Néon. (Desidera et Munari (2003)) . . . . .	28
2.4	Aperçu de la méthode de calibration SGIS de l'instrument spectroscopique de Gaia. Guerrier et al (2006) . . . . .	30
2.5	Illustration de la lecture d'un spectre d'une étoile sur le plan focal de Gaia. L'étoile E et son image sur le plan focal sont représentées à deux temps $t_1$ et $t_2$ différents. . . . .	33
3.1	Diagramme de contexte du processus de calibration en longueur d'onde proposé. La boîte décrit l'action principale, ici le processus de calibration. Chaque flèche représente un flot de données. Les flèches arrivant sur le flanc de la boîte sont des données d'entrée, nécessaires au déroulement de l'action. La flèche arrivant sur le dessus de la boîte symbolise les données auxiliaires. La flèche sortant représente les données de sortie fournies par l'action. La flèche rouge interne à l'action illustre le processus itératif. . . . .	37
3.2	Architecture du prototype de SGIS. Les principaux packages et classes du prototype sont représentés par des boîtes bleues. Le modèle d'objets utilisé par le prototype est représenté par des boîtes jaunes. Les outils communs au prototype sont représentés par une boîte blanche. Les données d'entrée, de sortie et auxiliaires sont représentées par une ellipse rose. Chaque flèche symbolise des flots de données : les flèches pleines sont des échanges de données entre le contrôleur SGIS et la base de données, les flèches en pointillés sont des échanges directs de données entre des modules et la base de données. Les échanges symbolisés par des pointillés seront à terme supprimés afin que tous les échanges transitent par le contrôleur. . . . .	41

---

3.3	Phase d'initialisation de SGIS. Ne se produit qu'une fois par exécution du prototype de SGIS. "Init" représente le module d'initialisation du prototype SGIS, "Apply Calib." le module d'application de la calibration en longueur d'onde, "Derive Rv" le module de calcul des vitesses radiales, "Select Ref." le module de sélection des étoiles de référence pour la calibration et "Calib. Wave." le module de calibration en longueur d'onde. . . . .	42
3.4	Phase d'application des calibrations en longueurs d'onde aux spectres extraits. .	43
3.5	Phase d'estimation des vitesses radiales des spectres calibrés en longueur d'onde.	44
3.6	Phase de sélection des références. . . . .	45
3.7	Phase de calibration en longueur d'onde du RVS. . . . .	45
3.8	Représentation schématique de la création des listes d'observations pour les unités de calibration : $D_{calUnit}$ correspond à la durée d'une unité de calibration (cette durée est la même pour chaque unité) et $t_{Observation}$ est le temps d'observation correspondant au temps de lecture d'un sample du spectre. Guerrier et al (2006). . . . .	47
3.9	Diagramme fonctionnel du processus semestriel de traitement de données. Ce diagramme est un actigramme SADT : chaque boîte représente une action et chaque flèche un flot de données. La nature du flot de donnée est définie par l'orientation et la direction de la flèche par rapport à la boîte (i.e. action) : une flèche entrant sur le flanc gauche de la boîte représente un flot de données d'entrée, une flèche sortant du flanc droit est un flot de données de sortie et une flèche entrant sur le haut de la boîte représente le flot de données de contrôle. .	54
4.1	Exemple de spectre d'une étoile G5V, obtenu par le simulateur du RVS. Les trois raies les plus profondes sont les raies du calcium ionisé. . . . .	60
4.2	Exemple de l'une des raies du Calcium ionisé observée dans le spectre extrait. Trois des quatre caractéristiques décrivant une raie dans le prototype y sont illustrées : un centre discret, l'aile bleue de la raie et l'aile rouge de la raie. . . . .	62
4.3	Exemple de détection de raies dans un spectre RVS d'une étoile de type solaire G5V : à gauche, le spectre RVS extrait, à droite, les raies détectées dans ce spectre.	63
4.4	Exemple de détection de raies dans un spectre RVS d'une étoile de type F (7000 K) : à gauche, le spectre RVS extrait, à droite, les raies détectées dans ce spectre. À gauche, le cercle noir montre la présence de deux raies significatives non-détectées. À droite, le cercle rouge montre une fausse détection de deux raies. .	64
4.5	Méthode d'identification des raies de référence. Les flèches noires représentent la localisation du triplet du Calcium permettant la mise en place d'une loi simple $sample = f(\lambda)$ , en rouge. Les flèches bleues représentent la localisation de deux raies de référence en utilisant la loi longueur d'onde - sample établie. . . . .	67

---

4.6	Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation. . . . .	70
4.7	Illustration de l'effet de blend sur les raies observées. . . . .	71
4.8	Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation. . . . .	75
4.9	Principe de la grille adaptative appliquée à la surface de corrélation. À gauche. Le carré rouge montre la position du maximum théorique de la surface de corrélation sur la grille d'échantillonnage. Le carré noir montre la position du maximum discret lu sur la première surface de corrélation. Le cadre noir représente la région d'intérêt sélectionnée autour de la position du maximum discret. Il encadre le domaine $[2A, 2B]$ sur lequel une nouvelle surface de corrélation va être calculée. À droite. Le carré rouge montre la position du maximum théorique de l'ancienne surface de corrélation (i.e. surface représentée à gauche). Le carré noir montre la position du maximum discret lu sur l'ancienne surface de corrélation. Et enfin, le carré bleu montre la position de l'estimation du maximum de la surface de corrélation sur cette nouvelle grille. . . . .	80
4.10	Illustration du changement de repère appliqué sur chaque raie détectée au niveau d'un spectre observé. À gauche : une raie détectée dans le domaine $[s_1, s_2]$ . À droite : cette même raie représentée dans le nouveau repère $[s'_1, s'_2]$ . . . . .	80
4.11	Surfaces de corrélation pour une raie du Calcium, en utilisant un changement de repère et une grille adaptative : exemple d'une des raies du calcium ionisée deux fois. Évolution du coefficient de corrélation en fonction de $k$ (i.e. échantillonnage) et $\delta\lambda'$ (i.e. décalage en longueur d'onde dans le nouveau repère). En rouge, première surface de corrélation sur un large domaine. En vert, seconde surface de corrélation à partir de laquelle la loi de calibration optimale est estimée. L'échelle de couleur est associée à la valeur du coefficient de corrélation. Elle permet de localiser les plus fortes corrélations dans le plan $[\delta\lambda', k]$ . . . . .	83

---

4.12	Résultats de l'estimation de la position du centre des raies de référence. La moyenne et la dispersion d'erreurs sur les estimations de chaque raie de référence sont représentées en rouge sur le graphique. Obtenue par un ajustement par spline (i.e. sous gnuplot, via la fonction spline), la courbe bleue permet de visualiser graphiquement les résultats des estimations. Elle met en évidence le décalage entre les estimations des différentes raies. Sur le même graphique, un spectre théorique est représenté en vert afin de visualiser les raies de référence prises en compte dans l'estimation. . . . .	84
5.1	Principe de la simulation de la méthode de centroïding. En partant de la vraie position en longueur d'onde $\lambda_{true}$ des raies stellaires, la méthode va bruite la position en sample $s$ de ces raies à partir d'une distribution gaussienne d'erreurs aléatoires. À la fin du processus de simulation, la méthode forme pour chaque raie stellaire choisie, un triplet $(\lambda_{shift}, \eta_{noise}, \zeta_{noise})$ . . . . .	90
5.2	Ce graphique représente la moyenne de la distribution d'erreur en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés pour chaque unité de calibration. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs. Les différences entre les diagnostics de chaque unité de calibration ne sont pas non-visibles, ici, à cause de la similarité des résultats. . . . .	92
5.3	Représentation fonctionnelle de la phase d'initialisation en mode de convergence pour une unité de calibration donnée. Ce diagramme utilise les conventions SADT (c.f. figure 3.9). . . . .	93
5.4	Ce graphique représente la moyenne de la distribution d'erreurs en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés pour chaque unité de calibration. La dispersion d'erreurs choisie pour bruite les coordonnées $(\eta, \zeta)$ est égale à 0.001 degré. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs. . . . .	94
5.5	Ce graphique représente la moyenne de la distribution d'erreurs en longueur d'onde (c.f. section 5.2.1) en fonction de la position en sample dans les spectres observés pour chaque unité de calibration. La dispersion d'erreurs choisie pour bruite les coordonnées $(\eta, \zeta)$ est égale à 0.01 degré. Les résultats de chaque unité de calibration sont représentés avec différentes couleurs. . . . .	95
5.6	Localisation des diagnostics d'erreurs dans le déroulement du prototype de SGIS. Deux types de diagnostics : un diagnostic d'erreurs sur les vitesses radiales (dans le processus itératif, après l'étape de Source Updating), et des diagnostics d'erreurs sur les calibrations en longueur d'onde (après l'initialisation et, dans le processus itératif, après l'étape de Calibration Updating). . . . .	98
5.7	Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. Chaque courbe de couleur est associée à une distribution d'erreurs sur l'estimateur sample-longueur d'onde. . . . .	99
5.8	Convergence du prototype avec une erreur sur le centroïding égale à 0.5 sample : évolution de la dispersion des vitesses radiales bornée par deux fonctions arc-tangente. . . . .	100

5.9	Convergence du prototype avec une erreur sur le centroïding égale à 0.5 sample : différence entre ces fonctions arctangente et l'évolution de la dispersion. . . . .	101
5.10	Diagnostics d'erreurs sur les calibrations en longueur d'onde à l'itération 1. Ce diagnostic montre l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	102
5.11	Diagnostics d'erreurs sur les calibrations en longueur d'onde à l'itération 2. Ce diagnostic montre l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	103
5.12	Diagnostics d'erreurs sur les calibrations en longueur d'onde à l'itération 10. Ce diagnostic montre l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	104
5.13	Diagnostics d'erreurs sur les calibrations en longueur d'onde à l'itération 20. Ce diagnostic montre l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	105
5.14	Evolution de la dispersion des erreurs sur les vitesses radiales estimées au cours des itérations : résultats des deux tests de convergence du prototype (la solution bruitée avec une dispersion de 0.001 degré est représentée en rouge, et celle bruitée avec une dispersion de 0.01 degré est en vert). . . . .	106
5.15	Evolution de la dispersion des erreurs sur les vitesses radiales estimées au cours des itérations : résultats de convergence (c.f. figure 5.14) agrandis sur la zone $[0, 0.4] \text{ km.s}^{-1}$ (axe des ordonnées). Le résultat du test de non-divergence pour le même prototype est également représenté en bleu. . . . .	107
5.16	Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. Chaque courbe est associée à une valeur du biais : un biais nul en <i>rouge</i> (caché par la courbe verte), un biais de 0.001 sample en <i>vert</i> , un biais de 0.01 sample en <i>bleu</i> , un biais de 0.1 sample en <i>rose</i> . . . . .	108
5.17	Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations : résultats des deux tests de convergence du prototype avec un biais de 0.001 sample (en rouge) et avec un biais de 0.01 sample (en vert). Ces résultats sont superposés aux résultats de non-divergence (c.f. figure 5.16). . . . .	109
5.18	Evolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations : les résultats de convergence (c.f. figure 5.17) montrent le détail des variations aux faibles valeurs de la dispersion. . . . .	110
6.1	Simulateurs de Gaia. Mignard et al. (2007) . . . . .	113
6.2	Densité d'étoiles observées en échelle logarithmique projetée en coordonnées équatoriales, pour 6 mois de données Gaia simulées par GASS. X. Luri, Barcelone, 2007 . . . . .	116
6.3	Distribution en magnitudes G et $G_{RVS}$ des sources observées pour 6 mois de données simulées. X. Luri, Barcelone, 2007 . . . . .	117
6.4	Diagramme HR des 6 mois de données GASS. . . . .	117

6.5	Exemple de spectre généré par GASS pour une étoile de type K : $G_{RVS}=3.1447$ , $T_{eff} = 4750K$ , $[Fe/H] = -0.25$ et $\log g = 2.5$ . . . . .	120
6.6	Exemple de spectre GASS redimensionné et redressé pour une étoile de type K : $G_{RVS}=3.1447$ , $T_{eff} = 4750K$ , $[Fe/H] = -0.25$ et $\log g = 2.5$ . Le spectre GASS "brut" est représenté en rouge et le nouveau spectre redimensionné et redressé en vert. . . . .	121
6.7	Exemple de création d'un spectre "patron" à partir d'un spectre synthétique de Marcs $R = 300000$ , $T_{eff} = 4750K$ , $[Fe/H] = 0.0$ et $\log g = 3.0$ : - À gauche, le spectre synthétique de Marcs tel qu'il est lu dans le fichier : flux en densité d'énergie avec une résolution spectrale $R = 300000$ et 50000 points - À droite, le spectre de Marcs convolué resamplé : flux en photons avec une résolution spectrale $R = 11500$ et 5000 points. . . . .	122
6.8	Diagramme HR des 6 mois de données GASS : les étoiles de référence (i.e. utilisée lors de la phase de calibration de SGIS) sont <i>en vert</i> . . . . .	124
6.9	Distribution des métallicités des étoiles observées durant les 6 mois. Le cadre délimite la zone de métallicité utilisée dans la calibration en longueur d'onde SGIS. . . . .	124
6.10	Résultats du diagnostic sur les vitesses radiales : évolution de la dispersion des erreurs sur les vitesses radiales moyennes au cours des itérations. La précision des vitesses radiales moyennes obtenue avec les échelles en longueur d'onde vraies (ou "précision vraie") est également représentée en vert. . . . .	126
6.11	Résultats du diagnostic sur les vitesses radiales : évolution de la différence entre la précision des vitesses radiales moyennes et la précision vraie en fonction des itérations. . . . .	127
6.12	Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itérations 0. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	128
6.13	Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itérations 1. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	129
6.14	Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itérations 4. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	130
6.15	Diagnostics d'erreurs sur les échelles en longueur d'onde à l'itérations 9. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde centrale associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	131
6.16	Erreurs de centroïding en fonction des longueurs d'onde des raies de référence. . . . .	132
6.17	Erreurs de centroïding en fonction des longueurs d'onde des raies de référence. . . . .	132



6.18	Résultats du diagnostic sur les vitesses radiales : <i>en rouge</i> , évolution de la précision des vitesses radiales moyennes au cours des itérations sans standard - <i>en vert</i> , évolution de la précision des vitesses radiales moyennes au cours des itérations avec 1000 standards. . . . .	133
6.19	Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 0. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	134
6.20	Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 1. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	135
6.21	Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 3. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	136
6.22	Diagnostics d'erreurs sur les échelles en longueur d'onde aux itérations 6. Ce diagnostic représente l'erreur moyenne sur la longueur d'onde associée à chaque sample du spectre. Chaque courbe de couleur est associée à une unité de calibration donnée. . . . .	137

# Liste des tableaux

1.1	Principales caractéristiques du satellite Gaia . . . . .	10
1.2	Caractéristiques optiques des télescopes . . . . .	11
1.3	Caractéristiques spectrales de Gaia . . . . .	13
1.4	Comparaison entre les performances d'Hipparcos et de Gaia . . . . .	14
1.5	Précision des vitesses radiales $\sigma_{vR[km/s]}$ en fonction de la magnitude $V$ et du type spectral. Ces performances sont obtenues à partir de la simulation de 2000 spectres RVS. Pour chacun de ces spectres simulés, une vitesse radiale est calculée par corrélation croisée et est comparée à sa valeur théorique (fournie par l'étape de simulation). Les résultats présentés dans ce tableau sont issus de la note technique Sartoretti et al. (2007). . . . .	18
1.6	Principales caractéristiques techniques du Radial Velocity Spectrometer de Gaia.	19
3.1	Table récapitulative des données d'entrée et de sortie du prototype proposé de calibration en longueur d'onde . . . . .	38
3.2	Extrait du "Software Development Plan" pour le traitement de données spectroscopiques ia, Katz et al (2006) . . . . .	56
4.1	Statistiques en <b>samples</b> de chaque distribution d'erreurs sur l'estimation du centre des raies. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum pour une raie. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne empirique pour une raie. La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion empirique pour une raie. . . . .	69
4.2	Liste de raies de référence (en Å) pour une étoile G5V, avant et après la phase de calibration du biais du centroiding. Le nom de la raie est une concaténation entre le symbole de l'élément chimique et de sa position en longueur d'onde dans le référentiel du laboratoire. La différence entre les deux listes (en Å) montre la compensation du biais après la phase de calibration. . . . .	73

4.3	Statistiques en <b>samples</b> de chaque distribution d'erreurs sur l'estimation du centre des raies, après calibration du biais. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne. La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion D. . . . .	74
4.4	Statistiques en <b>samples</b> sur chaque distribution d'erreurs sur l'estimation du centre des raies. La colonne "Raie" donne le nom de la raie. La colonne "Nombre" donne le nombre d'estimations pour une raie. La colonne "Moyenne M" donne la moyenne empirique des erreurs sur les estimations de la position du minimum. La colonne "Erreur sur M" donne l'erreur sur l'estimation de la moyenne . La colonne "Dispersion D" donne la dispersion empirique des distributions pour une raie. La colonne "Erreur sur D" donne l'erreur sur l'estimation de la dispersion D. . . . .	82
6.1	Demande de données pour les tests des modules de calibration des données spectroscopiques de Gaia. Sartoretti et al. (2006) . . . . .	115

# Bibliographie

- BASTIAN, U. 2004, *Reference Systems, Conventions and Notations for Gaia*, ESA report.
- BASTIAN, U. ET BIERMANN, M. 2005, *Astrometric meaning and interpretation of high-precision time delay integration ccd data*, A&A, **438**, 745–755.
- BEERS, T. C. ET AL 2004, *Physical parameters of SDSS stars, the nature of the SDSS « Ring around the Galaxy », and the SEGUE Project*, Publications of the Astronomical Society of Australia, **21**, 207–211.
- DE BRUIJNE, J. 2005, *Accuracy budget and performance*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 35–42, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- DESIDERA, S. ET MUNARI, U. 2003, *Absorption cells in wavelength calibration of gaia*, dans : *Gaia spectroscopy, science and technology*, tm. 298, pp. 85–92, U. Munari ed.
- GAIA SCIENCE ADVISORY GROUP 2000, *GAIA - Composition, Formation and Evolution of the Galaxy - Report on the concept and Technology Study*, ESA-SCI(2000)4.
- GOUDA, N. ET AL 2005, *Japan astrometry satellite mission for infrared exploration (jasmine)*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 77–80, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- GUERRIER, A. ET AL 2006, *Prototype of the Spectroscopic Global Iterative Solution*, ESA report.
- HECHLER, M., COBOS, J. ET BELLO-MORA, M. 1999, *Orbits around L2 for the first, planck and gaia astronomy missions*, Technical Report (1999MHNOCODE), Mission analysis section, ESOC, ESA.
- HOG, E. ET MIGNARD, F. 2005, *Near earth objects*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 239–242, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- HUCKLE, H. ET AL. 2007, *WP630 Software Design Description : RVS Calibration*, DPAC Technical Report (GAIA-C6-SP-MSSL-HEH-005-2).
- HUI-BON-HOA, A. ET KATZ, D. 2004, *Proposal for the Spectroscopic Global Iterative Solution*, ESA report.

- KATZ, D. 2005, *Radial velocity spectrometer design and performance*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 51–57, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- KATZ, D. ET AL 2004, *Spectroscopic survey of the galaxy with gaia - i. design and performance of the radial velocity spectrometer*, *MNRAS*, **354**, 1223–1238.
- KATZ, D. ET AL 2006, *Software Development Plan*, ESA report.
- LINDEGREN, L. 2001, *Proposed prototype processes for the GAIA Global Iterative Solution*, ESA report.
- LINDEGREN, L. AND PERRYMAN, M.A.C. 1996, *Gaia : Global astrometric interferometer for astrophysics*, *A&A Suppl. Ser.*, **116**, 579–595.
- MIGNARD, F. 2002, *Observations of solar system objects with Gaia. I. Detection of NEOs*, *A&A*, **393**, 727–731.
- MIGNARD, F. 2003, *Considerations about the astrometric accuracy of gaia*, dans : *Gaia spectroscopy, science and technology*, tm. 298, pp. 25–40, U. Munari ed.
- MIGNARD, F. 2005, *Overall science goals of the gaia mission*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 5–14, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- MIGNARD, F. ET AL. 2007, *Gaia Data Processing & Analysis Consortium - Response to ESA’s Announcement of Opportunity - Proposal for the Gaia Data Processing*, DPAC Technical Report (GAIA-CD-SP-DPAC-FM-030-2).
- MUNARI, U. 1999, *GAIA Spectroscopy : Proposing the 8500–8750 Å Region and Evaluating the Performances*, *Baltic astronomy*, **8**, 73–96.
- MUNARI, U. ET AL 2001, *Evaluating GAIA performances on eclipsing binaries. I. Orbits and stellar parameters for V505 Persei, V570 Persei and OO Pegasi*, *A&A*, **378**, 477–486.
- MUNARI, U. ET CASTELLI, F. 2000, *High resolution spectroscopy over 8500-8750 Å for GAIA. II. A library of synthetic spectra for  $T_{\text{eff}} < 7500$  K*, *A&A Suppl. Ser.*, **141**, 141–148.
- O’MULLANE, W. ET AL 2006, *Java Coding Standard and Guidelines for DPAC*, ESA report.
- PACE, O. 2005, *Gaia : the satellite and payload*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de *ESA SP*, pp. 23–28, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- PERRYMAN, M. A. C. ET AL. 2001, *GAIA : Composition, formation and evolution of the Galaxy*, *A&A*, **369**, 339–363.
- PRESS, W. H. ET AL. 2002, *Numerical Recipes in C - Second Edition*, Cambridge University Press.

- 
- SARTORETTI, P. ET KATZ, D. 2007, *Simulation requirements for spectroscopic processing for cycle 3*, DPAC Technical Report (GAIA-C6-SP-OPM-PS-004).
- SARTORETTI, P., KATZ, D. ET AL. 2006, *Simulation requirements for spectroscopic processing for cycle 2*, DPAC Technical Report (GAIA-C6-SP-OPM-PS-002-1).
- SARTORETTI, P., KATZ, D. ET GOMBOC, A. 2007, *First order estimation of the errors on radial and rotational velocity measurements with the Gaia3 RVS*, DPAC Technical Report (GAIA-C6-TN-OPM-PS-006-I).
- STEINMETZ, M. ET AL 2006, *The RAdial Velocity Experiment (RAVE) : First data release*, The Astronomical journal, **132**, 1645–1668.
- TANGA, P. 2005, *Impact of gaia on dynamics and evolution of the solar system*, dans : *Proceedings of the Symposium The Three-Dimensional Universe with Gaia*, tm. 576 de ESA SP, pp. 243–250, Turon, C. and O’Flaherty, K. S. and Perryman, M. A. C., eds.
- WILKINSON, M. I. ET AL 2005, *Spectroscopic survey of the Galaxy with Gaia- II. The expected science yield from the Radial Velocity Spectrometer*, MNRAS, **359**, 1306–1335.

