

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البلدية
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

pour l'obtention du diplôme de Master 2 en Électronique option
Réseaux et Télécommunication.

Thème

Application de l'Analyse en Composantes Principales pour la recherche de périodicité dans les courbes de vitesse radiale

Présenté par : Mlle. BELBLIDIA Hafidha

Proposé et Encadré par : Mr. DAMERDJI Yassine

Co-encadreur 1 : Mr. SEGHOUANI Nassim

Co-encadreur 2 : Mr. BENCHERCHALI Mohamed Amine

Année Universitaire 2015-2016

Remerciements

Au-delà d'une aventure scientifique, le PFE est également une aventure humaine, avec ses contentements intenses mais aussi ses périodes de doute.

Je ne manquer pas de dire un grand merci aux membres du jury qui ont accepté d'évaluer ce mémoire et de me faire part de leurs remarques pertinentes qui contribueront au perfectionnement de ce modeste travail.

Je remercie vivement mon Promoteur Mr. Yassine DAMERDJI, pour son encadrement, son immense patience, son suivi au cours de la réalisation de ce travail, ainsi que pour m'avoir fait bénéficier de ses grandes connaissances scientifiques ainsi que pour son temps, j'ai énormément appris à ses côtés.

Un très grand merci également à Mr. SEGHOUANI, chef des divisions astronomie et astrophysique, pour la confiance qu'il m'a accordée dès le premier jour de mon arrivée sur les lieux du stage.

Je tiens également à remercier vivement, Mr. DEMRI, Mr. BOUZIANI ainsi que toute personne que j'avais l'honneur de rencontrer pendant mon stage, de leur sympathie et leur gentillesse afin de mener à bien ce travail dans les meilleures conditions et dans un atmosphère harmonieux.

Je tiens à exprimer ma profonde gratitude à mon professeur co-encadreur Mr. BENCHERCHALI, pour sa disponibilité, ses encouragements et son soutien autant scientifique que moral qui m'ont permis d'aller au bout de mon PFE, ainsi que pour tous ses conseils judicieux quant à mes orientations futures.

Je remercie vivement tous les professeurs et enseignants de l'Université de Blida 1, plus particulièrement ceux du département d'électronique, qui ont contribué à ma formation.

J'aimerais remercier l'ensemble des personnes qui ont participé, de près ou de loin, à la réalisation de ce PFE.

ملخص:

منحنيات السرعة الشعاعية للأجرام السماوية (الأنظمة الثنائية في موضوع دراستنا) المتواجدة في المدارات الكبليرية هي دالة ذات عدة مؤشرات و التي ينجم عن تغيراتها كمية هائلة من المعلومات.

تطبيق مبدا التحليل الى مركبات أساسية يسمح بإيجاد دورة نظام ثنائي وهذا بتغيير فضاء الرصد نحو قاعدة متعامدة متجانسة ذات ابعاد مقلصة

كلمات المفاتيح: منحنيات السرعة الشعاعية , التحليل الى مركبات أساسية, دورة

Résumé :

Les courbes de vitesse radiale d'objet céleste (système binaire dans notre cas d'étude) en orbite képlérienne sont des fonctions à plusieurs paramètres qui par leur variation engendre une très grande quantité d'information.

L'application de la méthode de l'Analyse en Composantes Principales permet de repérer la période d'un système double et cela par un changement vers une nouvelle base orthonormée (orthogonalisation de l'espace) avec réduction de dimension.

Mots clés : courbes de vitesse radiale, Analyse en Composantes Principales (ACP), période

Abstract :

The radial velocity curves of celestial object (binary system in our case of study) on keplerian orbits are based on several parameters, which by their change, generates a large amount of information.

Applying the method of Principal Component Analysis can identify the period of a dual system and that a change to a new orthonormal basis (space orthogonalisation) with dimension reduction.

Keywords : radial velocity curves, Principal Component Analysis (PCA), period

Table des matières

Dédicace

Remerciements

Résumé

Liste des symboles

Liste des tableaux

Liste des figures

Introduction générale

Chapitre 1 : les systèmes binaires et les lois de mécanique céleste

1.1	Introduction.....	5
1.2	Catégories des étoiles doubles.....	5
1.2.1	binaires visuelles.....	6
1.2.2	binaires astrométriques.....	7
1.2.3	binaires spectroscopiques.....	8
1.2.4	binaires photométriques.....	8
1.2.5	D'autres systèmes binaires.....	10
1.3	Méthodes de détection.....	10
1.4	Eléments d'orbite képlérienne.....	11
1.4.1	Eléments orbitaux.....	11
1.4.2	Equation du mouvement (formule de vitesse radiale).....	12
1.5	Conclusion.....	13

Chapitre 2 : Analyse en Composantes Principales

2.1	Introduction.....	15
2.2	Historique	16
2.3	Principe de l'ACP.....	16
2.3.1	Notation.....	16

2.3.2	Calcul des Composantes Principales (C.P).....	16
2.4	Exemples d'application : reconstruction d'images.....	19
2.4.1	Premier exemple : Nébuleuse tête de Cheval.....	19
2.4.2	Deuxième exemple : Drapeau de l'Algérie.....	21
2.4.3	Troisième exemple : Drapeau de Madagascar.....	23
2.5	Taux de Compression.....	24
2.6	Conclusion.....	25

Chapitre 3 : Les courbes de vitesse radiale

3.1	Introduction.....	27
3.2	Courbes de vitesse radiale.....	27
3.2.1	Standardisation de la matrice des courbes.....	28
3.2.2	Calcul des Composantes Principales des courbes.....	28
3.2.3	Projection sur les axes de la nouvelle base.....	29
3.3	Classification des (e, ω)	30
3.4	Conclusion.....	31

Chapitre 4 : La recherche de périodicité dans les courbes de vitesse radiale

4.1	Introduction.....	33
4.2	Rappel.....	33
4.2.1	Séries temporelles.....	33
4.2.2	Repliement en phase.....	34
4.2.3	Echantillonnage régulier.....	34
4.2.4	Echantillonnage irrégulier.....	35
4.2.5	L'interpolation.....	36
4.2.6	Méthode de régression linéaire : les moindres-carrés.....	36
4.2.7	Séries de Fourier.....	37
4.2.8	Composantes Principales.....	38
4.2.9	Périodogramme.....	38
4.3	Application.....	39
4.3.1	Premier cas : signal échantillonné régulièrement.....	39
4.3.2	Deuxième cas : signal échantillonné irrégulièrement.....	42
4.4	Conclusion.....	44

Conclusion générale

Annexe

Bibliographie

Liste des figures

Figure 1.2.1.1 les étoiles de la Constellation de la Grande Ourse.

Figure 1.2.1.2 les séparations entre les deux compagnons Mizar et Alcor.

Figure 1.2.2 Exemple d'une binaire astrométrique (l'étoile Sirius).

Figure 1.2.3 binaire spectroscopique et décalage spectral.

Figure 1.2.4 les trois types d'une binaire à éclipse.

Figure 1.4.1 Eléments orbitaux d'une orbite elliptique.

Figure 2.3.2.4 Critère de Keiser pour la détermination du nombre de Composantes Principales à retenir.

Figure 3.3 Classification des (e, ω) à partir des coefficients C_1^2 et C_2^2 de la nouvelle base (seuil=99.99%).

Figure 4.2.1 signal périodique observé d'une courbe de lumière (de période= 0.628j) en fonction du temps.

Figure 4.9.2 Périodogramme du signal de la figure 4.2.1

Figure 4.3.1.1 variation de la vitesse radiale en fonction du temps (échantillonnage régulier).

Figure 4.3.1.2 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode FFT.

Figure 4.3.1.3 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode standard.

Figure 4.3.1.2 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode simplifiée.

Figure 4.3.2.1 variation de la vitesse radiale en fonction du temps (échantillonnage irrégulier).

Figure 4.3.2.2 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage irrégulier par l'application de la méthode simplifiée.

Liste des tableaux

Tableau 2.4.1 Application de l'ACP avec différents seuil sur l'image de la Nébuleuse de tête de Cheval

Tableau 2.4.2.1 Application de l'ACP avec différents seuil sur le drapeau de l'Algérie avec standardisation

Tableau 2.4.2.2 Application de l'ACP avec différents seuil sur le drapeau de l'Algérie sans standardisation

Tableau 2.4.2.1 Application de l'ACP avec différents seuil sur une image reconstruite par Matlab

Tableau 2.5.1 Taux de Compression pour les différents seuils appliqués sur le drapeau d'Algérie

Tableau 2.5.2 1 Taux de Compression pour les différents seuils appliqués sur l'image de la nébuleuse de tête de cheval

Tableau 3.2 courbes de vitesse radiale pour différentes valeurs de e, ω et T_p (avec $P=2j$)

Tableau 3.2.2.1 les 12 premières Composantes Principales de notre simulation pour un seuil de 99.99%

Tableau 3.2.2.2 les 6 C.P de notre simulation pour un seuil de 90%

Tableau 3.3 Classification des (e, ω) à partie des coefficients C_k de la nouvelle base (seuil=99.99%)

Conclusion générale

*« Les grands bonheurs viennent du ciel,
les petits bonheurs de l'effort »*

Proverbe Chinois

Le développement des instruments en astronomie, en particulier les satellites, ont rendu l'acquisition des données plus efficace ; visant à améliorer l'étude de l'Univers dont les étoiles doubles constituent le processus fondamentale de sa formation.

L'étude des systèmes binaire, particulièrement les étoiles doubles, est très importante du fait des informations qu'elles procurent : évolution du milieu stellaire, la masse, les distances, la recherche d'exoplanètes, ...

En effet, dans ce présent travail, nous nous sommes intéressés à l'étude d'une caractéristique de ses systèmes : déterminer la ou les périodicités «cachée(s)» dans les courbes de vitesse radiale.

Par l'application de différentes méthodes d'analyse et de calcul dans nos simulations : l'ACP classique, Interpolation, régression linéaire, nous avons obtenu des résultats satisfaisants.

Cependant, les travaux d'investigation dans ce vaste domaine de recherche utilisent plusieurs autres méthodes : les courbes de lumière, méthode de régression robuste, ACP de kernel, entre autres, qui par leurs performances mais aussi par leurs limites et leurs contraintes relèvent d'autres questions.

En outre, d'autres études approfondies sont nécessaires vu que l'Univers ne cesse de nous intriguer par ses mystères qui restent à dévoiler...

Bibliographie

[1] Paul Couteau ; L'observation des étoiles double et visuelles ; Préface de Jean-Claude PECKER FLAMMARION ; Observatoire de Nice ; 1971

[2] F. Arenou ; LES MULTIPLES ETOILES DOUBLES, Guide de données astronomiques pour l'observation du ciel, Annuaire du Bureau des Longitudes, Observatoire de Paris, CNRS ; ed. EDP Sciences ; 2008

[3] Jean-Louis Halbwachs ; Présentation générale des étoiles doubles ; Ecole CNRS de Goutelas XXIII, Observatoire de Strasbourg ; 2000

[4] Frédérique Baron ; Recherche et caractérisation de systèmes binaires dont l'une des composantes est de faible masse ; Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de Maître `es sciences (M.Sc.) en physique ; Avril, 2014

[5] Siramas Komonjinda ; A study of binary star orbits using precise radial velocity measurements with the HERCULES spectrograph (A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Astronomy); University of Canterbury, New Zealand; 2008

[6] Jean-Louis Prieur ; Thèse d'habilitation à diriger des recherches : Contribution en Signal, Image et Instrumentation pour l'Astronomie ; Institut de Recherche en Astronomie et Planétologie CNRS, Université de Toulouse, Version 16 janvier 2015

[7] <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-explo-acp.pdf>

[8] Jolliffe I.T.; Principal Component Analysis 2nd edition, Springer, April 2002

[9] Harrar Sabiha ; Eléments d'aide à la décision dans l'Analyse Territoriale, Application de l'ACP sur la région Nord-Ouest ; Revue El-Wahat pour les Recherches et les Etudes Vol.7 n°2, page 124 ; (2014)

[10] C. Duby, S. Robin ; Analyse en Composantes Principales, Institut National Agronomique Paris – Grignon ; 10 juillet 2006

[11] Duncan Gillies and Marc Deisenroth; CO 493 Lecture 15: Principal Component Analysis; Departement of computing, Imperial College London; Spring 2015

[12] Jean-Marc Nicoletti. La détection de périodicités cachées. ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, SUISSE. 20 avril 2012

[13] R. H. Jones; Spectral analysis with regularly missed observations; The Annals of Mathematical Statistics, 33 : 455–461, 1962.

[14] N. Chihab, A. Zergainoh, J.-P. Astruc ; Etude des fonctions B-splines non-uniformes pour la reconstruction d'un signal discret à partir d'un échantillonnage irrégulier ; L2TI, Institut Galilée, Université Paris 13 ; 2003

[15] R.Radau ; Mémoires et observations : Études sur les formules d'interpolation ; Bulletin Astronomique, Serie I, vol. 8, pp.273-294 ; juillet 1891.

[16] Yassine DAMERDJI ; cours Techniques d'ajustement de fonctions ; Centre de Recherche en Astronomie, Astrophysique et Géophysique (CRAAG) ; 2016

Introduction générale

« Cette obscure clarté qui tombe des étoiles... »

Le Cid

« Les gens ont des étoiles qui ne sont pas les mêmes.

Pour les uns, qui voyagent, les étoiles sont des guides.

Pour d'autres, elles ne sont rien que de petites lumières.

Pour d'autres, qui sont savants, elles sont des problèmes... »

Antoine de Saint-Exupéry, Le Petit Prince

Les astres qui baignent dans la profondeur des cieux en brillant d'une lumière aux mille feux, offrent de fantastiques associations de contrastes, où l'œil étonné voit se marier des éclats enviant les pierres précieuses.

Formant des couples célestes reposant au cœur de l'infini, ces astres gravitent l'un autour de l'autre au rythme des lois de la dynamique céleste. Un certain nombre de ces couples ayant déjà fait des révolutions complètes sous les yeux des observateurs, ont des périodes de révolutions qui varient de quelques jours à des milliers d'années.

A travers les âges, les étoiles furent utilisées comme :

✚ moyen d'orientation lors des déplacements,

- ✚ horloge lors de la mesure du temps ;
- ✚ marquage des saisons pour des rythmes agricoles.

Parmi les variétés des objets célestes, on trouve celle avec un caractère particulier : découverte par le télescope de W.Herschel en 1803, celles-ci présentent la particularité d'être doubles !

En astronomie observationnelle, une étoile double est une paire d'étoiles qui apparaissent comme proches l'une de l'autre dans le ciel, lorsqu'on les observe depuis la Terre, à travers un télescope optique.

Ceci peut se produire :

- ✚ parce que la paire forme un système binaire d'étoiles en orbite mutuelles (i.e. elles tournent chacune autour de l'autre, ou plus précisément, elles tournent toutes deux autour du barycentre (centre de gravité) du système qu'elles constituent, liées gravitationnellement l'une à l'autre ;
- ✚ ou parce que la paire d'étoiles forme une « double optique », un alignement dû au hasard de deux étoiles, qui se trouvent dans le ciel à des distances différentes, mais qui sont vues sous le même angle de visée depuis la Terre.

Pour les astronomes stellaires, l'importance des étoiles binaires réside dans le fait que la connaissance de leurs mouvements permet de calculer directement la masse des étoiles, ainsi que d'autres paramètres stellaires.

De nos jours, l'astronomie associée aux domaines de l'optique et de l'électronique a vu de nouvelles applications grâce aux satellites appelés satellites astronomiques. Hors atmosphère, ces derniers permettent une observation de qualité des objets célestes.

Comme Les étoiles ne se forment pas seules, les observations ont montrés que le résultat le plus fréquent de leur processus de formation est un couple de deux étoiles liées par la gravitation, ou étoile binaire.

A cet effet, notre projet de fin d'étude s'inscrit dans un cadre de détermination de périodes de rotations d'un système binaire, en tant que problème dynamique non-

linéaire, souvent mal conditionné, avec des erreurs de mesures souvent difficiles à quantifier.

La plupart des méthodes opèrent en deux temps : calcul d'une solution initiale, puis amélioration de cette solution avec une minimisation des résidus par moindres carrés.

En effet, parmi les populations stellaires, les systèmes doubles sont caractérisés par leur vitesse dans la ligne de visée, autrement dit, la vitesse radiale.

La vitesse radiale est fonction de plusieurs paramètres qui, par leurs variations dues aux leurs mouvements régis par les lois de mécanique céleste, entraînent une avalanche de données d'où la difficulté d'en extraire l'information utile « la période » !

De ce fait, l'Analyse en Composantes Principales (ACP), qui est une méthode de réduction de données ainsi que de changement de base, entre autres, est appliquée dans le but de trouver « la période », caractéristique importante dans la définition des systèmes doubles (étoiles, pulsar, quasar, galaxies, ...)

Dans un premier temps, nous nous intéressons aux aspects généraux traitant des systèmes binaires régis par les lois de la mécanique céleste pour délimiter le premier chapitre. L'analyse en composantes principale (ACP) fait l'objet du deuxième. Le troisième chapitre sera réservé à la présentation des courbes de vitesses radiales. Le quatrième conclura l'objet de ce mémoire par la présentation des applications.

Liste des symboles :

a : demi grand axe d'une ellipse.

i : inclinaison le plan orbital par rapport au plan éclipique.

e : excentricité d'une orbite elliptique.

ω : longitude du périastre par rapport au nœud ascendant .

T_p : temps de passage au périastre.

\mathcal{P} : période.

f_e : fréquence d'échantillonnage.

f_s : fréquence (plus élevée) du signal .

$\mathcal{F}(t)$: fonction .

$S(t_i)$: signal au temps t_i .

Ω : angle du position du noeud par rapport au nord.

M : anomalie moyenne.

E : anomalie excentrique.

v : anomalie vrais.

K : semi amplitude de la courbe de vitesse radiale.

V_γ : vitesse de centre de masse d'une binaire.

V_r : vitesse radiale.

χ^2 : moindres-carrées.

Var : variance.

Cov : covariance.

C_k : coefficients des axes de la nouvelle base des Composantes Principales.

N : nombre d'observation (d'échantillons).

X : matrice des valeurs brutes initiales.

W : matrice des covariances.

U : matrice dont les colonnes sont des valeurs propres.

λ : valeur propre.

Λ : matrice diagonale contenant les valeurs propres de la matrice W .

Σ : somme.

\mathbb{N} : l'ensemble des nombre naturels.

δ_e : pas d'échantillonnage.

\mathcal{L} : maximum de vraisemblance.

$\hat{S}(f)$: transformée de Fourier du signal $S(t_i)$.

$\mathbb{P}(f)$: puissance du signal.

Au Nom de Dieu le Clément le Miséricordieux

*" Je n'aurais pas à jurer par la position des étoiles et c'est un serment
tellement important, si vous le saviez "*

(Al-Waki'a S56-V76)

Introduction Générale

Chapitre 1

Les systèmes binaires et lois de mécanique céleste

*« Si les astres étaient immobiles, le temps et l'espace n'existeraient
plus »*

Maurice Maeterlinck

Chapitre 2

Analyse en Composantes Principales (ACP)

*« L'univers nous appartient dans la proportion où nous lui appartenons »
Jean Giono*

Chapitre 3

Les courbes de vitesse radiale

*« L'observation est l'investigation d'un phénomène naturel et l'expérience
est l'investigation d'un phénomène modifié par l'investigateur. »*

Claude Bernard

Chapitre 4

La recherche de périodicité dans les courbes de vitesse radiale

*« Ce qui est admirable, ce n'est pas que le champ des étoiles soit si vaste,
c'est que l'Homme l'ait mesuré. »*

Anatole France

Conclusion Générale

Annexes

Bibliographie

1.1 Introduction

L'Univers est constitué principalement d'étoiles et de nébuleuses. Les étoiles se groupent en amas géants appelés galaxies. [1]

Les étoiles d'une galaxie gravitent autour du centre de l'amas à une vitesse d'autant plus rapide qu'elles en sont plus proches. [1]

L'observation montre que, très souvent, les étoiles sont associées par paires appelée étoile double. Les deux corps s'attirent fortement, mais cette force centripète est équilibrée par la force centrifuge de révolution. Il en résulte un mouvement orbital autour du centre de gravité. On conçoit que la forme et la rapidité de ce mouvement orbital nous renseignent sur les masses des corps célestes, ainsi que d'autres paramètres, d'où l'importance des étoiles doubles. [1]

L'observateur assiste au parcours des orbites, ce qui demande des années et souvent des siècles. Il constate que la loi de la gravitation énoncée par Newton est universelle. Il remarque en même temps que les lois de Kepler ordonnent les orbites des planètes, aussi loin que nous pouvons les voir. [1]

1.2 Catégories des étoiles doubles

Les étoiles doubles sont tellement diverses et nombreuses (voir section 1.4) que la nécessité de les classer, en plusieurs catégories selon les méthodes d'observation (voir

section 1.5), s'imposait vue que, parfois, un système appartient à plusieurs catégories : c'est le cas de α -Centauri, étoile double visuelle, mais aussi spectroscopique et astrométrique. [1]

Ces systèmes sont ainsi présentés ci-dessous, par ordre essentiellement historique de leur découverte : [2]

1.2.1 Binaires visuelles

Une étoile double visuelle est un couple d'étoiles proches sur le ciel dont les deux composantes peuvent être observées séparément avec un instrument comme un télescope, ce qui dépend donc en particulier du pouvoir de séparation ou de résolution de l'instrument. [2]

Les étoiles doubles visuelles sont à l'origine de la découverte même des étoiles binaires. Ce sont aussi les plus aisément observables : ainsi, le couple Mizar et Alcor, dans la constellation de la Grande Ourse, était connu dès l'antiquité. [3]

Bien que Mizar et son cavalier Alcor, avec leur séparation de 12', aient probablement servi de test d'acuité visuelle depuis l'Antiquité, la première utilisation du terme d'étoile double qui nous soit parvenue provient de Ptolémée dans l'Almageste, à propos du couple ν_1 et ν_2 Sagittarii, dont la séparation est de 14'. [2]

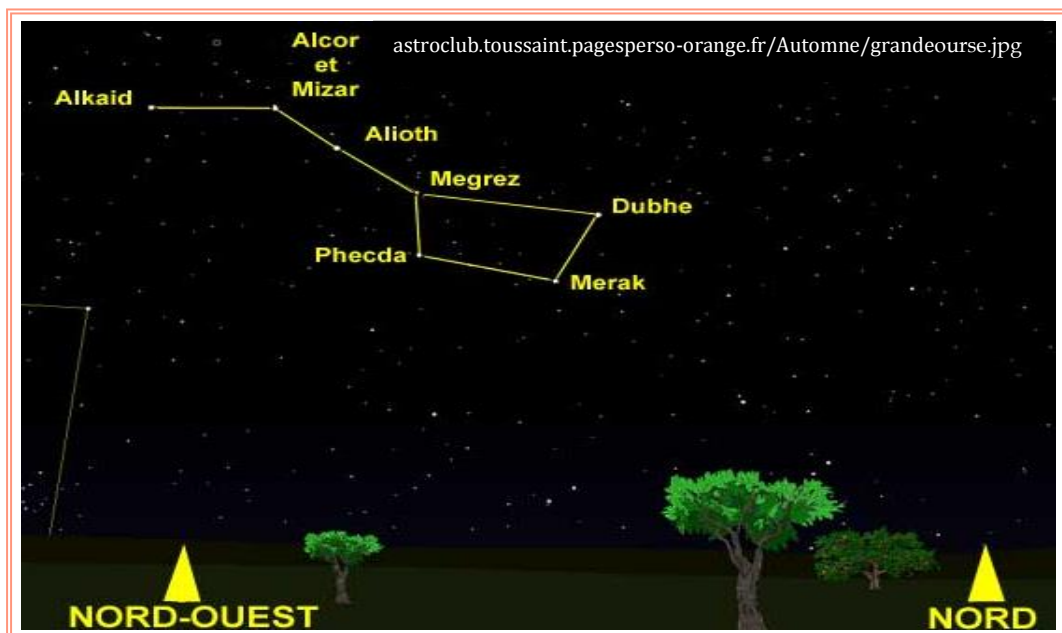


Figure 1.2.1.1 les étoiles de la Constellation de la grande Ourse

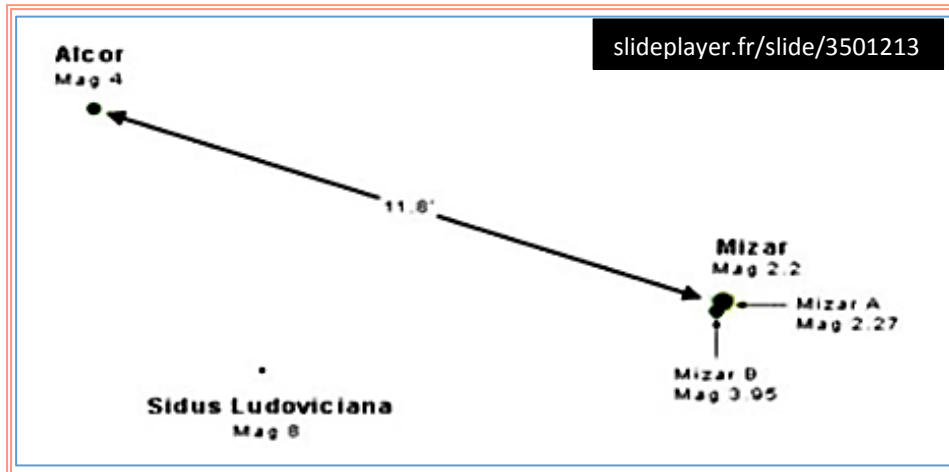


Figure 1.2.1.2 les séparations entre les deux compagnons Mizar et Alcor

1.2.2 Binaires astrométriques

Les systèmes binaires astrométriques sont caractérisés par le fait qu'une seule des deux composantes est visible et que celle-ci décrit un mouvement orbital sur le ciel. [4]

En 1844, Bessel, cherchant à mesurer les parallaxes et mouvements propres de Sirius et de Procyon, découvrait des mouvements résiduels dont les périodes étaient de 50 ans pour Sirius et 41 ans pour Procyon. Il en concluait que ces étoiles avaient chacune un compagnon obscur, et qu'il avait mis en évidence le mouvement de la composante primaire autour du centre de gravité du système. [3]

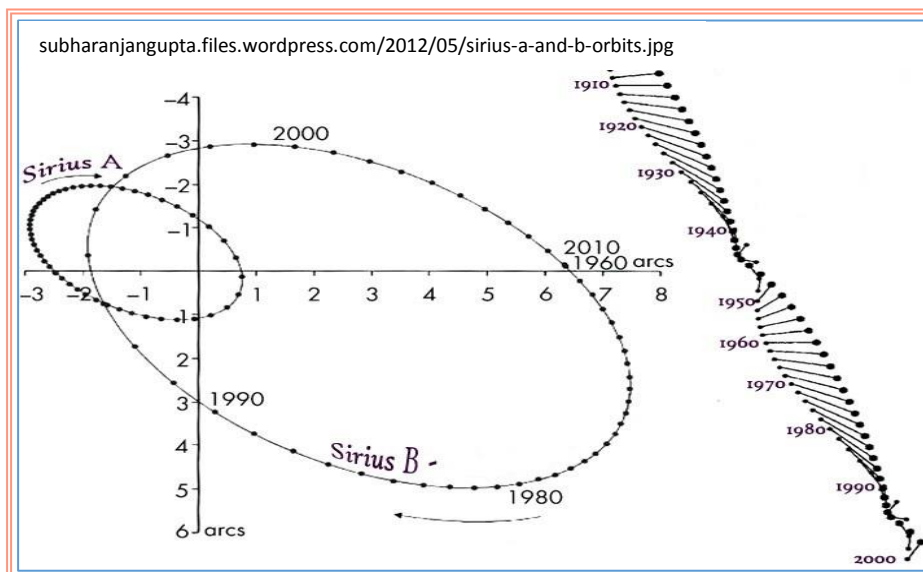


Figure 1.2.2 exemple d'une binaire astrométrique « l'étoile Sirius » (à droite, le mouvement orbital ; à gauche mouvement propre)

1.2.3 Binaires spectroscopiques

La découverte des binaires spectroscopiques remonte à 1889, lorsque miss Antonia Maury, à l'observatoire d'Harvard, s'aperçut que le spectre de Mizar présentait une raie K du calcium périodiquement dédoublée. Le phénomène fut attribué à l'existence de deux composantes de luminosités comparables, tournant autour de leur centre de gravité en une vingtaine de jours. Le dédoublement des raies correspond à une différence de vitesse radiale qui, dans le cas de Mizar, atteint 136 km/s. [2] [3]

L'analyse du spectre distingue les binaires à un spectre seulement, nommées BS1 (SB1 en anglais), quand seul le mouvement des raies de l'étoile la plus brillante peut être mesuré ce qui est en particulier le cas pour des étoiles hôtes de planètes extrasolaires ; et les binaires à deux spectres, nommées BS2, quand le mouvement des raies des deux composantes est observé.

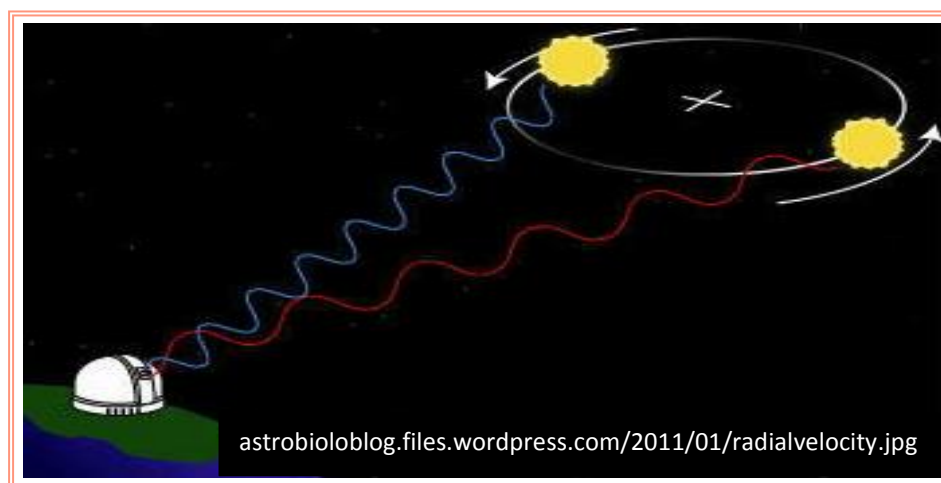


Figure 1.2.3 binaire spectroscopique et décalage spectral

1.2.4 Binaire photométrique (à éclipse)

Les binaires à éclipses sont des systèmes binaires qui présentent une variation périodique de leur luminosité. [5]

Ces systèmes orbitent à peu près perpendiculairement au plan du ciel. Quand une étoile passe devant l'autre de la ligne de vue de l'observateur, une éclipse se produira. L'observation photométrique de ce genre de système affiche la courbe de lumière caractéristique d'un système d'occultation. [5]

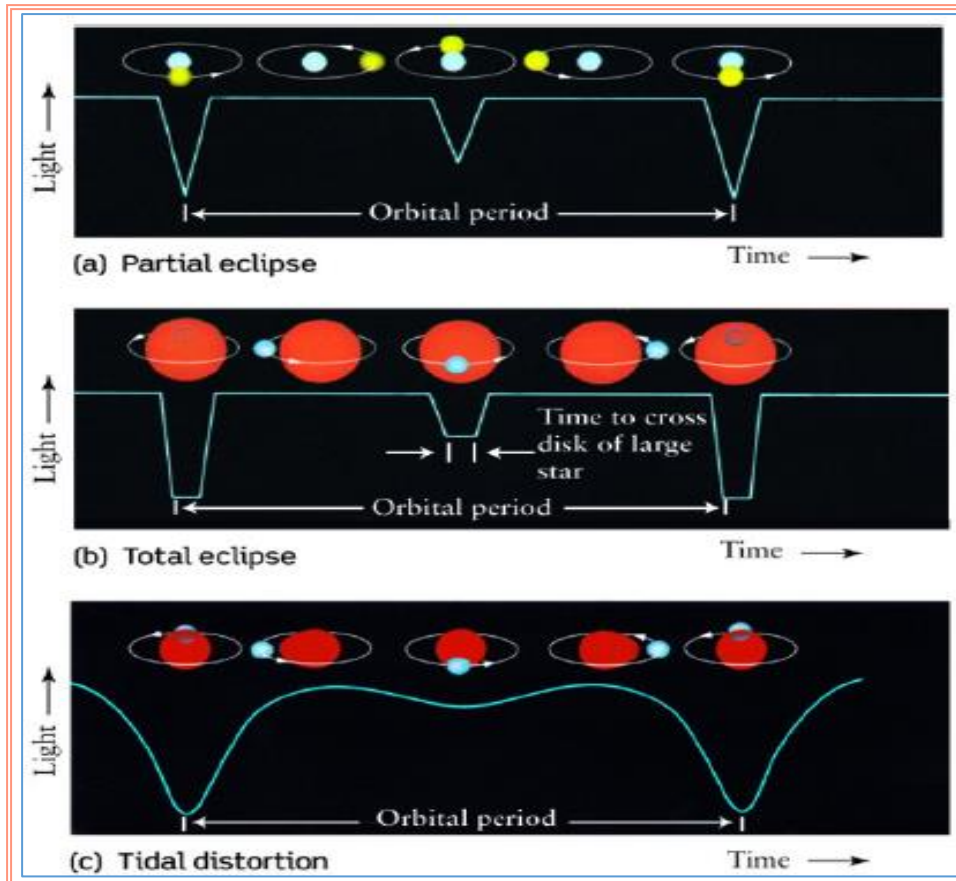


Figure 1.2.4 Binaire à éclipse et ses 3 types EA, EB et EW (source : Yassine Damerджи, les binaires spectroscopique dans la mission spatiale GAIA, Bibliothèque Nationale, 21/12/2015)

En 1669, avec 2000 ans de retard sur les annales chinoises, Montanari signale la variabilité de l'éclat de β -Persei, alias Algol, ou "le démoniaque". [3]

La découverte de Montanari tomba dans l'oubli, et ce n'est qu'un siècle plus tard que Goodricke (1783) redécouvrait la variabilité d'Algol et, pour la première fois, en détermina la périodicité : sur un cycle de 69 heures, Algol est de magnitude 2,1 pendant 60 heures. Elle passe la moitié des 9 heures qui restent à décliner jusqu'à 3,4 magnitudes, et autant à revenir à son éclat initial. Cette variation de 1,3 magnitude est d'autant plus facile à remarquer que l'étoile ρ -Persei, proche voisine d'Algol, a la même magnitude qu'Algol à son minimum. [3]

1.2.5 Autres systèmes binaires

Outre les principaux groupes d'étoiles binaires que nous avons mentionnés, de nombreux autres types d'étoiles binaires ont été trouvés : **binaires X** tels que Cen X3 (Schreier et al. 1972) ainsi que son compagnon X1 (Tananbaum et al., 1972) sont des étoiles brillantes dans des longueurs d'onde de rayons X qui pulsent dans une période de secondes et montrent parfois des éclipses dans une période de plusieurs jours. Il a été montré que ce type de système se compose parfois d'une étoile ordinaire et une étoile à neutrons ou, éventuellement, un trou noir. [5]

Aux longueurs d'onde radio, **les binaires de pulsars** (composées de deux étoiles à neutrons) ont été découverts dans une région compacte de amas globulaire (c'est une collection sphérique d'étoiles en orbite autour d'un noyau galactique comme un satellite. Ils sont très étroitement liés par la gravité, ce qui leur donne leurs formes sphériques et des densités stellaires relativement élevées vers leurs centres). [5]

1.3 Méthode de détection

L'essentiel de l'information en provenance des astres est sous forme de lumière (à différentes longueurs d'onde). Cette lumière peut être analysée de différentes manières :

- ✚ analyse spatiale du flux de lumière : imagerie
- ✚ mesure des positions des astres : astrométrie
- ✚ mesure du flux dans différentes bandes de couleurs : photométrie
- ✚ analyse détaillée du spectre : spectroscopie
- ✚ mesure de la polarisation de la lumière : polarimétrie

Ces méthodes d'analyse peuvent éventuellement être couplées (spectrophotométrie, spectro-imagerie, ...) ou associées à d'autres techniques d'analyse de la lumière (interférométrie, ...).

1.4 Élément d'orbite képlérienne

1.4.1 Éléments orbitaux

Lors de l'étude des systèmes doubles, l'observateur étudie habituellement le mouvement d'une composante par rapport à l'autre, le référentiel adopté étant alors centré sur la composante la plus brillante (dite primaire). Le mouvement de la composante secondaire par rapport à la composante primaire forme une ellipse. [2]

Cette orbite est caractérisée par 7 éléments : a = demi grand axe ; e = excentricité ; \mathcal{P} = période ; T_p = temps de passage au périastre ; Ω = angle de position du nœud par rapport au nord ; ω = longitude du périastre par rapport au nœud ascendant (éloignement de l'étoile par rapport à l'observateur) ; i = inclinaison du plan de l'orbite. [6]

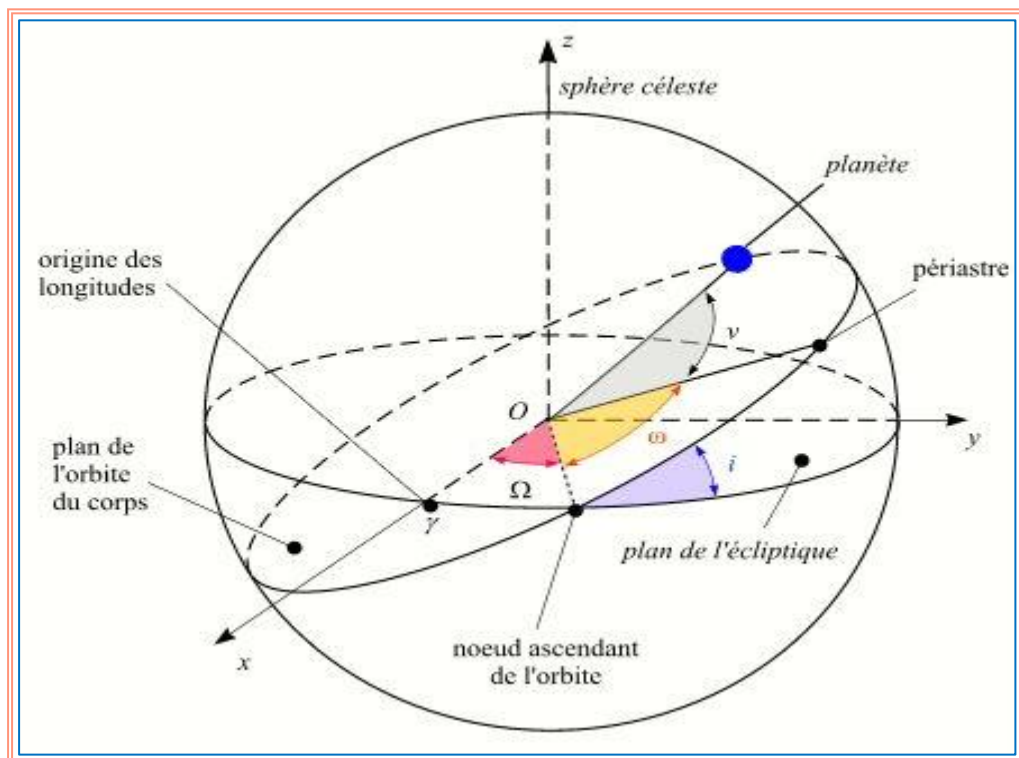


Figure 1.4.1 Éléments d'orbite képlérienne

L'anomalie moyenne M est la position moyenne du compagnon A (dans la Figure 1.4.1 planète), proportionnelle au temps $(t - T_p)$, compté à partir de T_p le temps de passage au périastre ; elle est donnée par l'expression : [6]

$$M = \frac{2\pi}{P} (t - T_p) \quad (1.1)$$

Et l'anomalie excentrique E , l'angle compté à partir du centre de l'ellipse et relié à M par l'équation :

$$E = M + e \cdot \sin E \quad (1.2)$$

Cette équation est appelée : équation de Kepler. [6] [2]

D'où l'anomalie vraie v :

$$\tan \frac{v}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{E}{2} \quad (1.3)$$

1.4.2 Equation du mouvement : formule de la vitesse radiale

Dans le cadre d'un simple mouvement képlérien, chaque composante du système décrit une ellipse autour du barycentre, on observe pour chaque composante la vitesse dans l'axe de visée appelée vitesse radiale. Elle est donnée par l'expression : [2] [5]

$$V_r = V_\gamma + K(e \cos \omega + \cos(v + \omega)) \quad km/s \quad (1.4)$$

Avec :

$$K_{1,2} = \frac{10879}{P} \frac{a_{1,2} \sin i}{\sqrt{e^2 - 1}} \quad km/s \quad (1.5)$$

Où $K_{1,2}$ est la semi-amplitude de la courbe de vitesse radiale de la primaire (respectivement secondaire) exprimée en km/s, $a_{1,2}$ sont les demi-grands axes en unité astronomique, P la période orbitale en jour.

1.5 Conclusion

Dans ce chapitre, nous avons présenté les étoiles doubles (binaires) et les lois de Kepler qui les régissent.

Les paramètres orbitaux d'une binaire peuvent être déterminé à partir des vitesses radiales des composantes du système.

Les courbes de vitesse radiale sont des caractéristiques importante dans l'étude des binaires.

Chapitre 2 Analyse en Composantes Principales

(ACP)

2.1 Introduction

L'Analyse en Composantes Principales est une méthode factorielle de réduction de dimension pour l'exploration statistique de données quantitatives complexes. [7]

Lorsque l'on étudie simultanément un nombre important de variables quantitatives, la difficulté vient de ce que les individus étudiés ne sont plus représentés dans un plan, mais dans un espace de dimension supérieure. [7]

L'objectif de l'Analyse en Composantes Principales (ACP) est de revenir à un espace de dimension réduit par exemple à 2 en déformant le moins possible la réalité ; Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales. [7]

2.2. Historique

L'ACP prend sa source dans un article de Karl Pearson publié en 1901. Le père du test du χ^2 y prolonge ses travaux dans le domaine de la régression et des corrélations entre plusieurs variables. Pearson utilise ces corrélations non pas pour expliquer une variable à partir des autres (comme en régression), mais pour décrire et résumer l'information contenue dans ces variables. [8]

Encore connue sous le nom de transformée de Karhunen-Loève ou de transformée de Hotelling, l'ACP a été de nouveau développée et formalisée dans les années 1930 par Harold Hotelling. La puissance mathématique de l'économiste et statisticien

américain le conduira aussi à développer l'analyse canonique, généralisation des analyses factorielles dont fait partie l'ACP. [8]

2.3. Principe de l'ACP

L'idée à la base de l'analyse en composantes principales est le remplacement des anciens axes de variables x_k par de nouveaux axes V_k de la nouvelle base. Ces nouvelles variables V_k sont appelées composantes principales ; elles s'expriment comme des combinaisons linéaires des vecteurs de l'ancienne base x_1, \dots, x_m . [9]

$$V_k = k_1 x_1 + k_2 x_2 + \dots + k_m x_m \quad (2.4)$$

2.3.1 Notation

Les données sont des mesures effectuées sur n unités $\{u_1, u_2, \dots, u_i, \dots, u_n\}$, représentées par p variables quantitatives $\{v_1, v_2, \dots, v_j, \dots, v_p\}$.

Le tableau des données brutes à partir duquel l'analyse sera faite est noté X ayant la forme suivante : [10]

$$X = \begin{matrix} & v_1 & \dots & v_p \\ \begin{matrix} u_1 \\ \vdots \\ u_n \end{matrix} & \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \end{matrix}$$

2.3.2 Calcul des Composantes Principales

En termes de calcul, les Composantes Principales sont obtenues par le calcul des vecteurs propres et les valeurs propres de la matrice des covariances. Ce procédé équivaut à trouver les axes du système dans lequel la matrice de covariance est diagonale.

Les vecteurs propres ayant une grande valeur propre définissent la direction de la grande dispersion. [11]

2.3.2.1 Prétraitement : Standardisation des variables (données)

Avant de se lancer dans l'analyse de nos données, le problème d'unité se pose ! De ce fait, afin d'y remédier, les données seront centrées-réduites.

En procédant ainsi, seule la variation du signal importe. En effet, le calcul de PCA sur des signaux non standardisés est beaucoup moins efficace que lorsque cette étape de prétraitement est effectuée.

2.3.2.2 Calcul de la matrice des covariances

On appelle matrice de covariance empirique de p variables quantitatives $v_1, \dots, v_j, \dots, v_p$ mesurées sur un ensemble de n unités, la matrice à p lignes et p colonnes contenant sur sa diagonale principale les variances empiriques des p variables, et ailleurs, les covariances empiriques de ces variables deux à deux : **[10]**

$$W = \begin{bmatrix} Var(v_1) & Cov(v_1, v_2) & \dots & Cov(v_1, v_j) & \dots & Cov(v_1, v_p) \\ Cov(v_2, v_1) & Var(v_2) & \dots & Cov(v_2, v_j) & \dots & Cov(v_2, v_p) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ Cov(v_j, v_1) & Cov(v_j, v_2) & \dots & Var(v_j) & \dots & Cov(v_j, v_p) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(v_p, v_1) & Cov(v_p, v_2) & \dots & Cov(v_p, v_j) & \dots & Var(v_p) \end{bmatrix}$$

Avec :

$$Var(v_j) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \langle x_j \rangle)^2 \quad (2.2)$$

$$Cov(v_j, v_k) = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \langle x_j \rangle) (x_{ik} - \langle x_k \rangle) \quad (2.3)$$

Où la moyenne de la $j^{\text{ème}}$ colonne est donnée par l'expression :

$$\langle x_j \rangle = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (2.4)$$

2.3.2.3 Vecteurs propres/valeurs propres

Soit \mathbf{X} une matrice de $n \times n$ dimensions. Les valeurs propres de \mathbf{X} sont définies comme étant les racines de l'équation caractéristique :

$$\text{Déterminant} (\mathbf{X} - \lambda \mathbf{I}) = |(\mathbf{X} - \lambda \mathbf{I})| = 0 \quad (2.5)$$

Avec \mathbf{I} est la matrice identité de dimension $n \times n$.

Soit λ une valeur propre de \mathbf{X} , associée au vecteur propre \mathbf{v} : $\mathbf{X}\mathbf{v} = \lambda\mathbf{v}$.

La matrice de corrélation est définie positive, ce qui implique que ses valeurs propres sont toutes positives et ses vecteurs propres constituent une base orthonormée de l'espace.

La Diagonalisation de la matrice des corrélations pour l'obtention des vecteurs propres et des valeurs propres est effectuée grâce à la décomposition en valeurs singulières SVD. En effet, cette dernière, contrairement à la décomposition en valeurs propres, retourne des vecteurs propres normalisés.

La matrice des covariances \mathbf{W} peut s'écrire alors comme :

$$\mathbf{W} = \mathbf{U} * \mathbf{\Lambda} * \mathbf{U}^{-1} \quad (2.6)$$

avec \mathbf{U} la matrice dont les colonnes sont les vecteurs propres et $\mathbf{\Lambda}$ une matrice diagonale contenant les valeurs propres de \mathbf{W} .

2.3.2.4 Nombre de Composantes Principales significatives à retenir

Afin de déterminer le nombre de composantes principales qui déterminent les axes du nouvel espace, on applique le critère de Keiser : un seuil est fixé au choix puis le multiplier par la valeur maximale de la somme des valeurs propres cumulées (somme des scores) : [7]

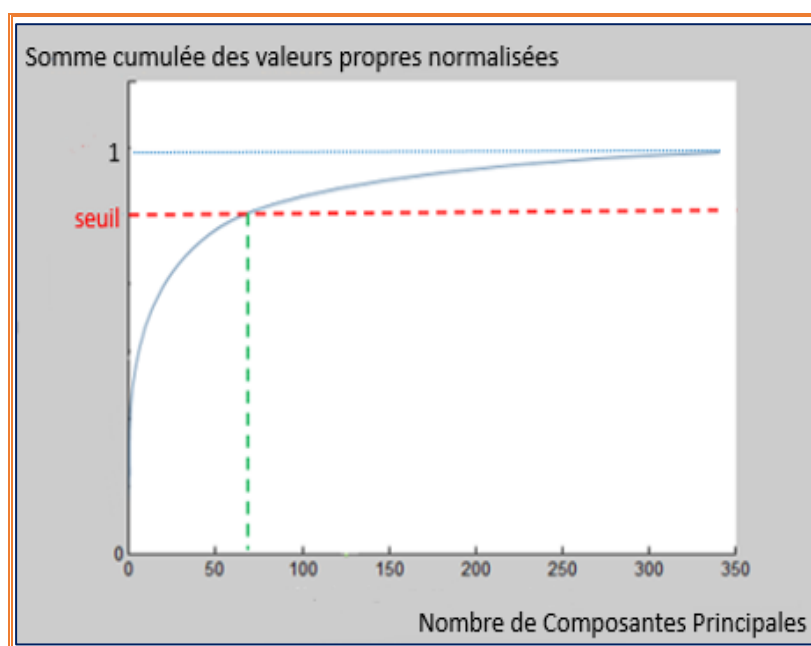


Figure 2.3.2.4 Critère de Keiser pour la détermination du nombre de composantes principales à retenir (dans cet exemple, pour un seuil = 90% Le nombre de Composantes Principales = 70/340)

2.3.2.5 Reconstruction des signaux sur la nouvelle base des Composantes Principales

Tout signal peut être reconstruit par la simple projection sur la nouvelle base orthonormée.

Il s'écrit alors sous la forme d'une combinaison linéaire des vecteurs propres de cette nouvelle base donné par l'expression :

$$\mathbf{X} = \sum_k (\mathbf{C}_k * \mathbf{V}_k) \quad (2.7)$$

Avec \mathbf{C}_k est la projection de \mathbf{X} sur \mathbf{V}_k tel que :

$$\mathbf{C}_k = \sum_i (\mathbf{X}_i * \mathbf{V}_{ki}) \quad (2.8)$$

Sachant que \mathbf{V}_k sont les axes de la nouvelle base (Composantes Principales)

2.4. Exemple d'application : reconstruction d'images

Afin de bien comprendre l'ACP, nous l'avons appliquée à la construction d'images.

Une image est une matrice de $n \times m \times l$ dimension, avec : n est nombre de lignes, m le nombre de colonne et l le nombre de couches RVB ($l=3$).

Vue que différents seuils sont choisis à chaque fois que l'ACP est appliquée, les résultats obtenus sont présentés dans des tableaux.

2.4.1 Premier exemple : Nébuleuse tête de Cheval

Dans cet exemple, l'image est de dimension 611×813 sachant qu'elle est redimensionnée à 25% de sa taille initiale (2444×3252).





L'image	Seuil %	Nombre de Composantes principales significatives/ composante RVB			L'image reconstruite
		R	V	B	
	95	108	248	240	
	99	373	542	531	
	99.99	1607	1604	1419	

Tableau 2.4.1 Application de l'ACP avec différents seuils sur l'image de la Nébuleuse de tête de Cheval .

Pour une reconstruction plus affine d'une image qui contient beaucoup de détails, le nombre de Composantes Principales significatives augmente en fonction du seuil.

2.4.2 Deuxième exemple : Drapeau d'Algérie

Afin de visualiser l'effet de la standardisation, dans un premier temps, l'étude est effectuée sur le drapeau d'Algérie de dimension 585×881×3.

2.4.2.1 Traitement avec standardisation

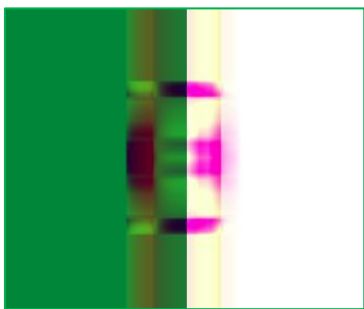
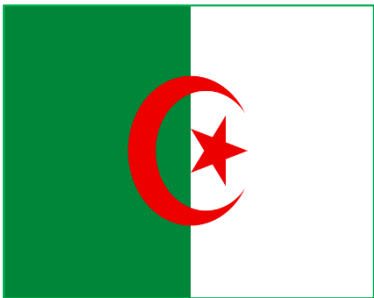

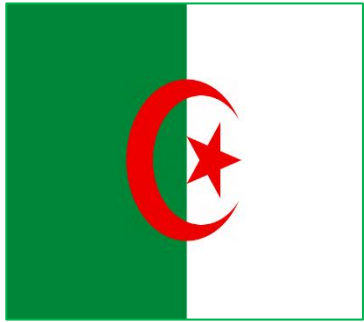
L'image	Seuil %	Nombre de Composantes principales significatives/ composante RVB			L'image reconstruite
		R	V	B	
	95	1	3	1	
	99	4	16	8	
	99.99	62	110	85	

Tableau 2.4.2.1 Application de l'ACP avec différents seuil sur le drapeau d'Algérie avec standardisation.

2.4.2.2 Traitement sans standardisation

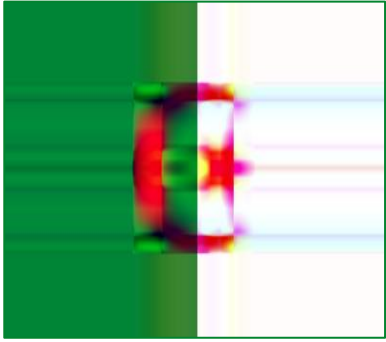
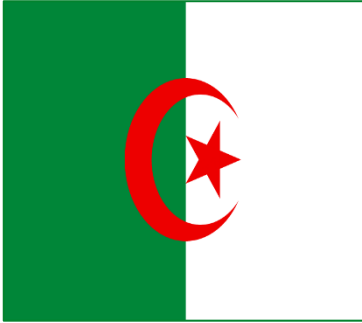


L'image	Seuil %	Nombre de Composantes principales significatives/ composante RVB			L'image reconstruite
		R	V	B	
	95	2	5	2	
	99	4	16	8	
	99.99	62	110	85	

Tableau 2.4.2.2 Application de l'ACP avec différents seuil sur le drapeau d'Algérie sans standardisation.

On remarque que le nombre de composantes principales significatives est le même dans les deux cas : avec et sans standardisation !

Pour un faible seuil, l'effet de la standardisation est visible : pour un seuil de 95%, la reconstruction avec standardisation est plus affine.

2.4.3 Troisième Exemple : Drapeau de Madagascar

Cet exemple est un cas particulier : le choix de ce drapeau, fait par Matlab, revient au fait qu'il est combiné de 2 partie distinctes : horizontale (blanches) et verticale (rouge et verte)

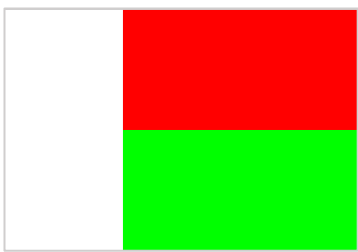

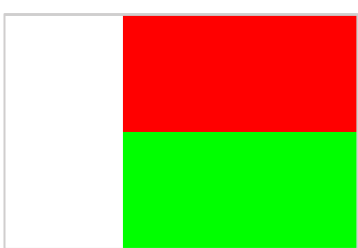
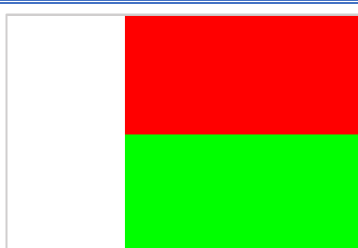
L'image	Seuil %	Nombre de Composantes principales significatives/ composante RVB			L'image reconstruite
		R	V	B	
	95	1	1	1	
	99	1	1	1	
	99.99	1	1	1	

Tableau 2.4.3 Application de l'ACP avec différents seuils sur une image reconstruite par Matlab

Il suffit d'une seule composante principale pour la reconstruction : étant donné que la matrice des covariances est nulle, on prend une seule ligne comme seule composantes principale. Cette ligne porte en elle toute l'information nécessaire à la reconstruction.

2.5 Taux de compression

Le taux de compression (T.C) est donné par l'expression :

$$T.C = \left(1 - \frac{\text{nombre de composantes principales significatives}}{\text{nombre de lignes}}\right) * 100$$

Les taux de compression pour chaque exemple, sont présentés dans les tableaux ci-dessous :

	Seuil %	Composantes			Taux de Compression total/ image
		R	V	B	
Taux de Compression (%)	95	99.66	99.14	99.66	99.49
	99	99.32	97.26	98.63	98.40
	99.99	89.40	81.2	85.47	85.36

Tableau 2.5.1 Taux de compression pour les différents seuils appliqués sur le drapeau de l'Algérie.

	Seuil %	Composantes			Taux de Compression total/ image
		R	V	B	
Taux de Compression (%)	95	95.58	89.85	90.18	91.87
	99	84.74	77.82	78.27	80.28
	99.99	34.25	34.37	41.94	36.85

Tableau 2.5.2 Taux de compression pour les différents seuils appliqués sur l'image de la Nébuleuse de tête de Cheval.

On remarque que le taux de compression diminue au fur et à mesure que le seuil augmente.

Concernant le troisième exemple, celui du drapeau de Madagascar, on remarque que quel que soit le seuil appliqué, les résultats de reconstruction de l'image sont identiques ; de même, le taux de compression est égal à 99.66%.

2.6 Conclusion

Dans ce chapitre, une vision simplifiée de la méthode de l'Analyse en Composantes Principales est présentée.

L'ACP est majoritairement utilisée pour décrire, visualiser ainsi que pour compresser des données multidimensionnelles.

Plus le nombre de Composantes Principales significatives à retenir est grand, plus la reconstruction de l'information est affine.

De plus, l'ACP est un algorithme avec pertes puisqu'il se limite aux variations linéaires des signaux. En effet, le taux de compression est optimisé selon le problème traité.

3.1 Introduction

Les binaires spectroscopiques s'étudient par l'intermédiaire des variations temporelles des vitesses radiales associées au mouvement des deux étoiles autour de leur centre de masse. Les courbes obtenues étant nommées courbes de vitesse radiale.

3.2 Courbes de vitesse radiale

D'après la formule (1.4) , les orbites peuvent facilement être simulées :

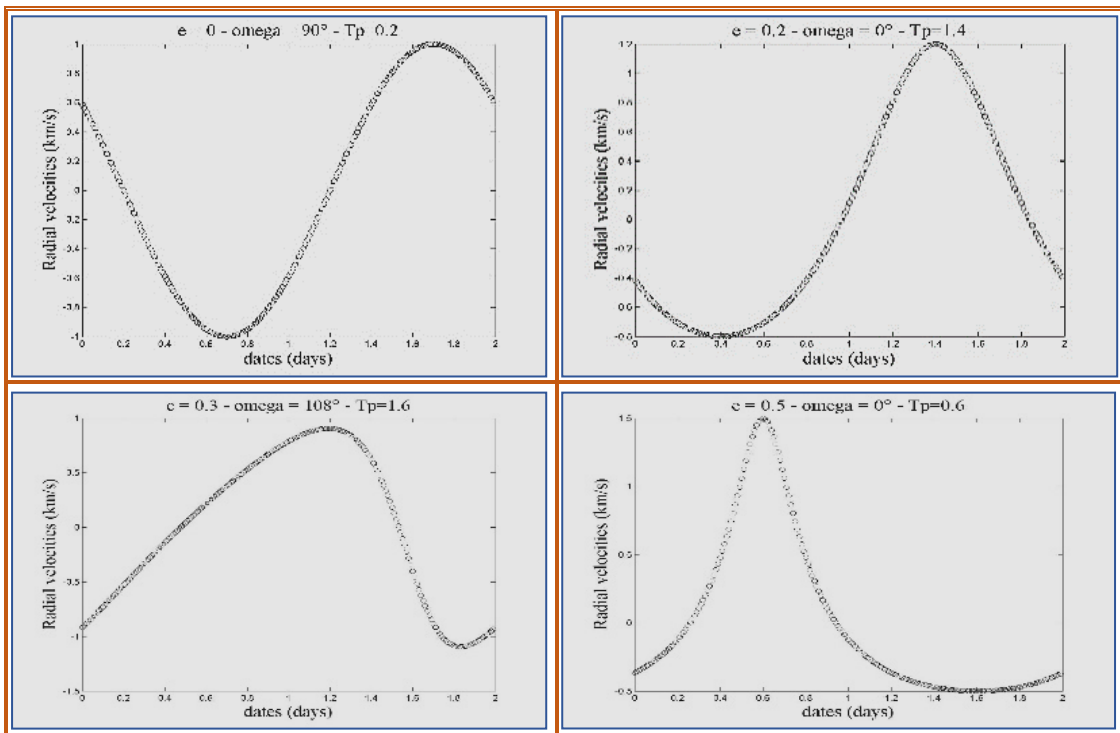


Tableau 3.2 Courbes de vitesse radiale pour différentes valeurs de e , ω et T_p (avec $P=2$ jours).

Dans ce tableau, nous avons présenté 4 courbes sur un total de 1910 courbes simulées !

3.2.1 Standardisation de la matrice des courbes

Après avoir obtenu ces courbes de vitesse radiale, celles-ci sont présentées dans une matrice de dimension 1910×2000 , sachant que 2000 est le nombre de points à pas régulier dans une période, et 1910 est le nombre de courbes simulées.

Chaque courbe de la matrice est standardisée. De ce fait, une nouvelle matrice de même dimension que celle de départ est obtenue contenant les courbes standardisées.

3.2.2 Calcule des Composantes Principale des courbes

Afin de définir les axes de la nouvelle base où ces courbes seront projetées, les composantes principales associées sont calculées.

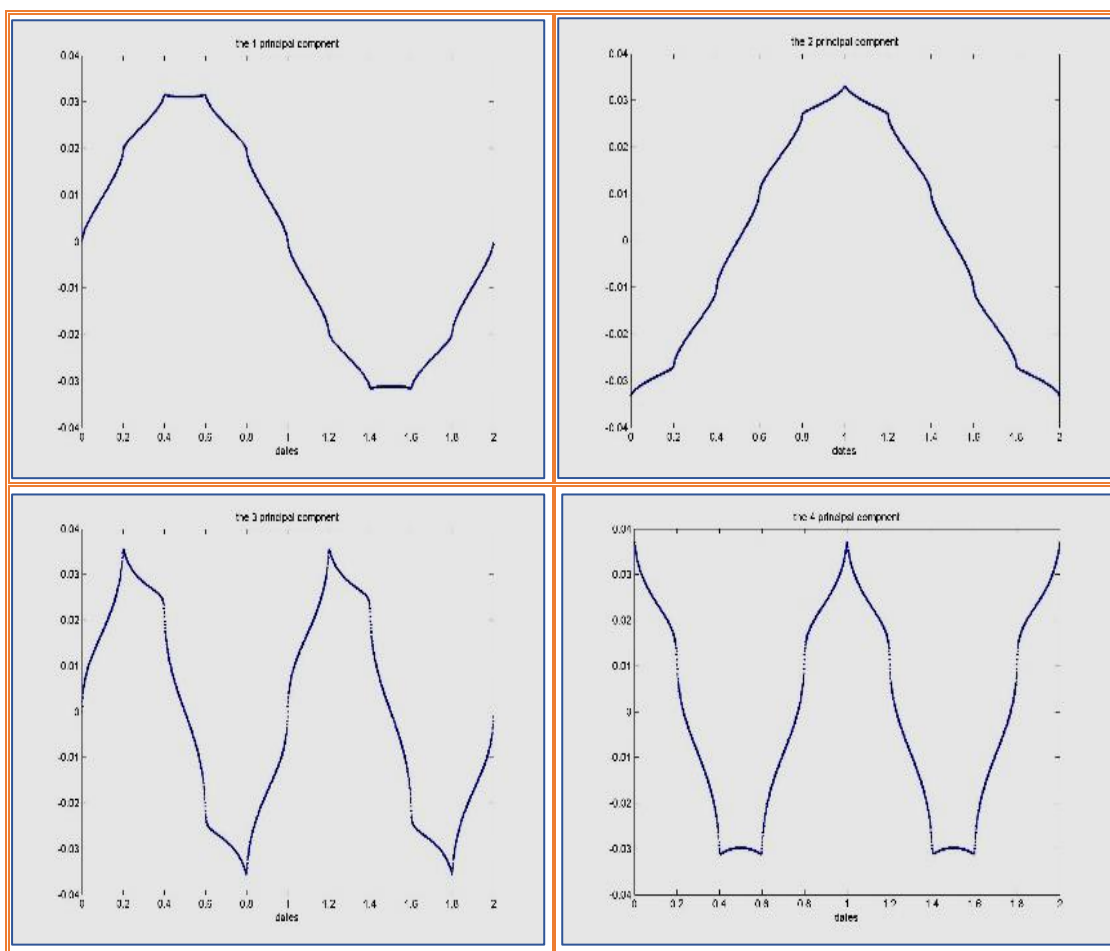


Tableau 3.2.2.1 les 4 premières Composantes Principales de notre simulation pour un seuil=99.99 % (sur un total de 56 C.P)

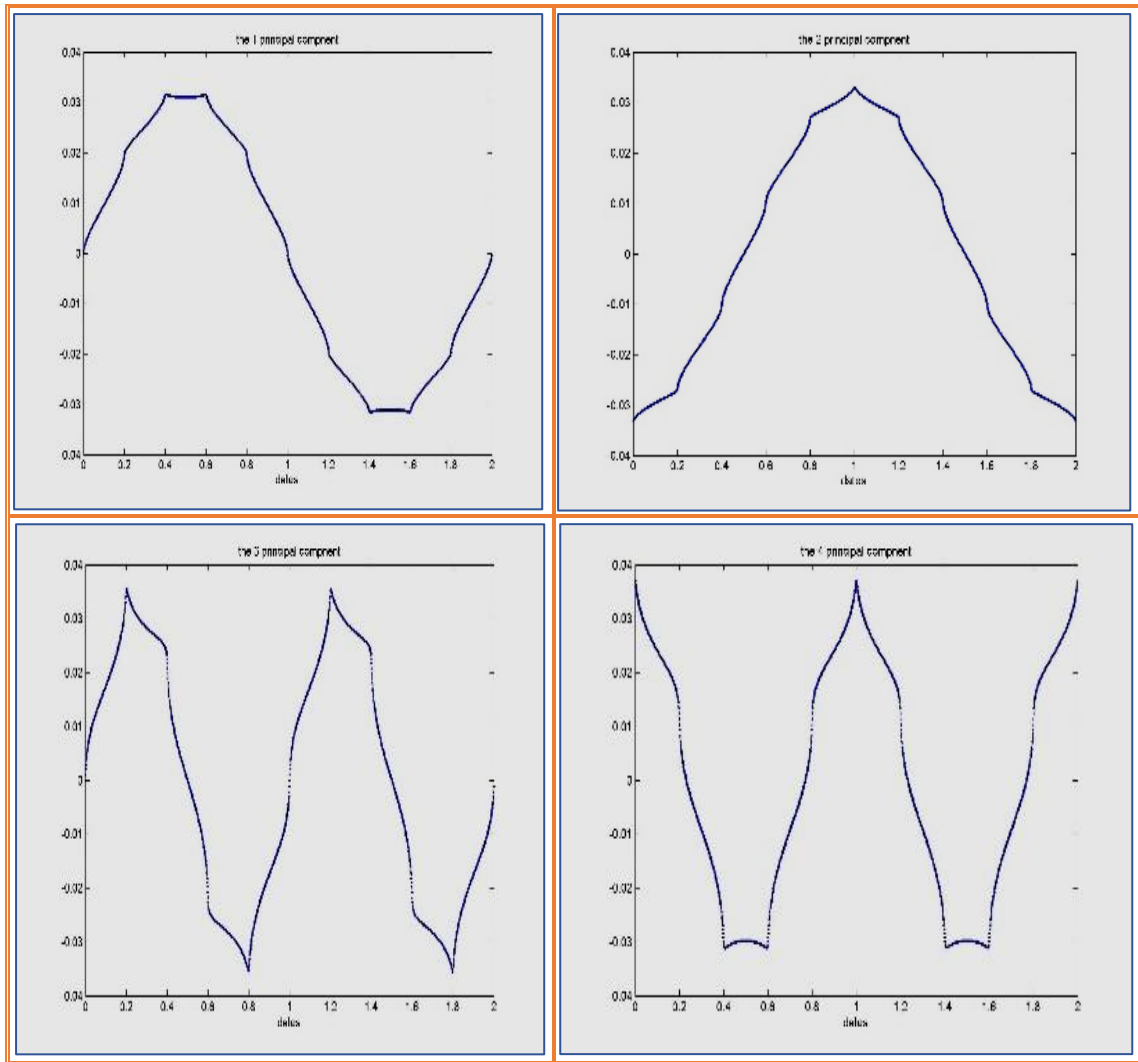


Tableau 3.2.2.2 les 4 Composantes Principales de notre simulation pour un seuil=90 %

Toutefois, le nombre de Composantes Principales significatives change en fonction du nombre des courbes simulées et du seuil choisi.

3.2.3 Projection sur les axes de la nouvelle base

Les Composantes Principales précédemment calculées définissent les axes de la nouvelle base sur lesquels est projetée chacune des courbes. Autrement dit, les 191 courbes simulées peuvent être reproduites par de simples combinaisons linéaires des Composantes Principales.

Matrice des C_k^l = matrice des courbes de vitesse radiales * PCA des courbes

Cette formule de la projection n'est valable que lorsque l'échantillonnage est régulier !

C_k^l est la composante d'une courbe l suivant la composante principale k .

3.3 Classification des (e, ω) en fonction de C_k

La classification des (e, ω) est illustré dans les figures du tableau page suivantes :

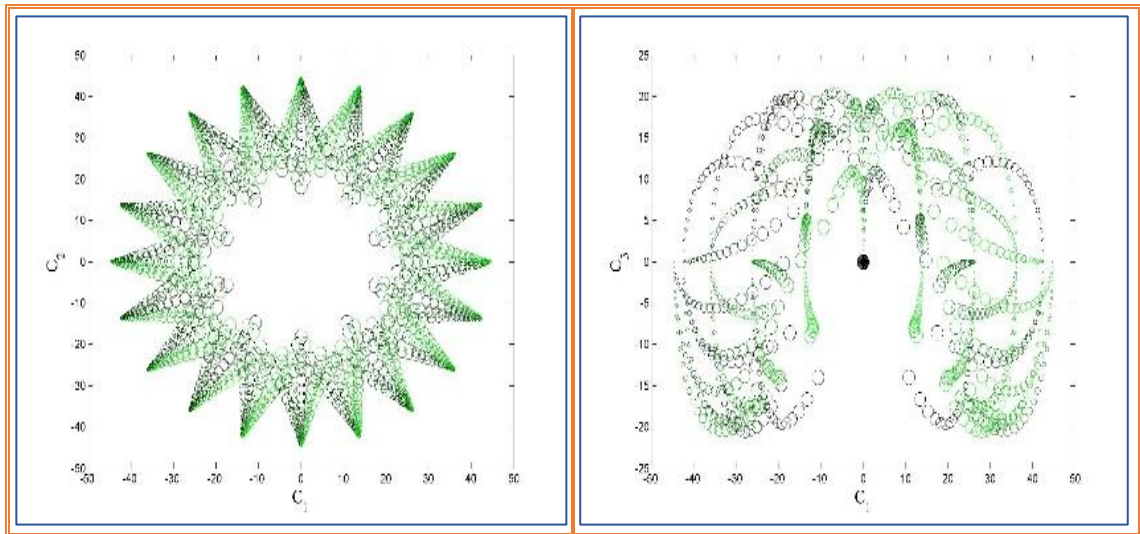


Tableau 3.3 Classification des (e, ω) à partir des Coefficients C_k de la nouvelle base (seuil=99.99%)

D'après les figures obtenues, la séparation des e et ω est difficile. Par contre, si on reporte le carré de C_2 en fonction du carré de C_1 , on peut facilement séparer les faibles excentricités des fortes excentricités de même pour les longitudes de périastre par de simples droites comme le montre la figure ci-après :

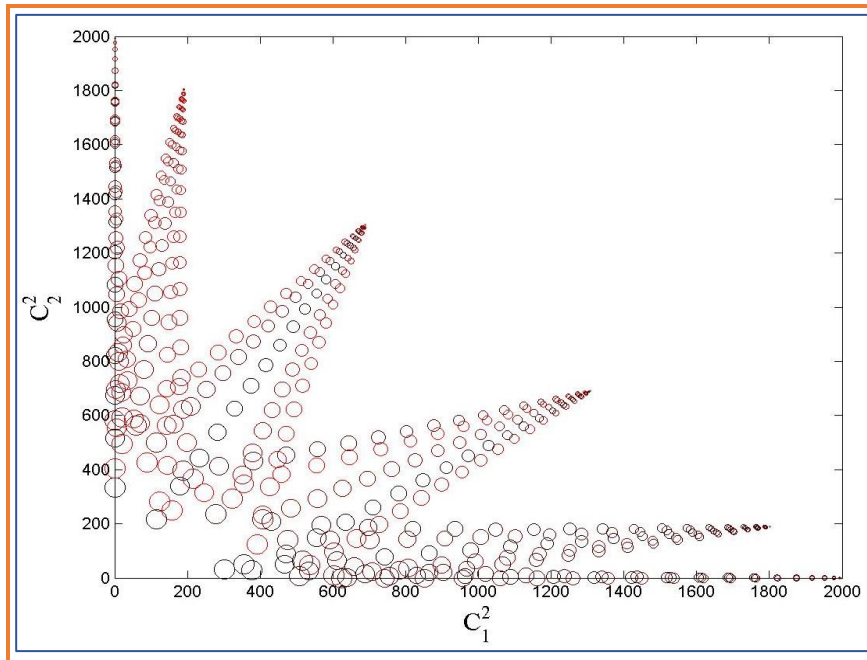


Figure 3.3 Classification des (e, ω) à partir des Coefficients C_1^2 et C_2^2 de la nouvelle base (seuil=99.99%)

La variation croissante du rayon des cercles traduit la variation croissante de l'excentricité, tandis que la variation de la saturation de la couleur (de la plus claire à la plus foncée) traduit la variation croissante des longitudes du périastre.

3.4 Conclusion

Par l'application de l'ACP, le nombre de courbes de vitesse radiale est réduit ; de plus le seuil est un critère important dans la détermination du nombre de Composantes Principales significatives à retenir.

Grace à la nouvelle base, il nous est possible de classifier les excentricités et les longitudes de périastre.

Dans le chapitre qui suit, les courbes de vitesse radiale sont remplacées par leurs nouvelles courbes représentatives de la nouvelle base.

Chapitre 4 la recherche de périodes dans les courbes de vitesse radiale

4.1 Introduction

Une courbe de vitesse radiale est une série temporelle qui représente l'évolution dans le temps, du phénomène de la variation de la vitesse d'un système binaire dans la ligne de visée (radiale) de l'observateur.

La périodicité du changement de la vitesse radiale est un paramètre important qui doit être déterminé avec une bonne précision.

Dès lors, il devient intéressant de les étudier. Cette étude nécessite la maîtrise et exige la connaissance des outils (mathématiques, statistique, ...) indispensables au traitement.

4.2 Rappels

4.2.1 Séries temporelles

Les séries temporelles sont l'étude d'une quantité dépendante du temps, représentées comme un couple temps/observation (t_i, y_i) où y_i est l' $i^{\text{ème}}$ observation au temps t_i .

Un exemple est la figure 4.2.1 d'une courbe de luminosité d'une binaire photométrique :

[12]

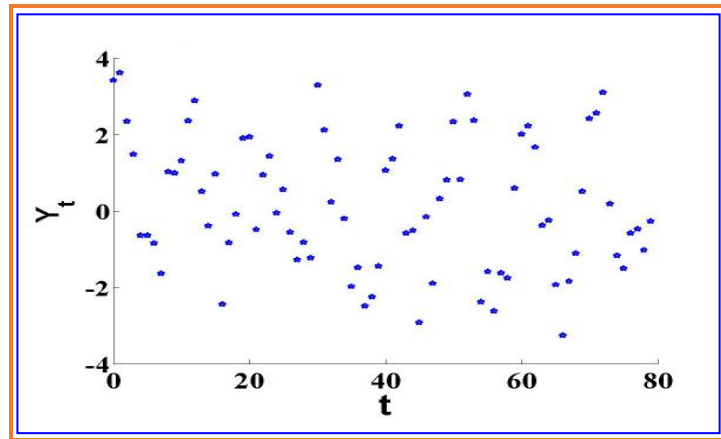


Figure 4.2.1 Signal périodique observé d'une courbe de lumière (de période $P=0.628$ jours) en fonction du temps. [12]

Soit le modèle en temps continu

$$y(t) = g(t) + e(t) \quad (4.1)$$

Où $g(t)$ est le signal et $e(t)$ est un bruit blanc en temps continu. Les observations aux temps $t_1; \dots; t_n$ conduisent au modèle en temps discret

$$Y(t_i) = g(t_i) + e(t_i) \quad (4.2)$$

Où $g(t_i)$ est le signal au temps t_i et $e(t_i)$ est l'erreur au temps t_i . L'observation $Y(t_i)$ est une réalisation de Y_i avec $i = \{1, \dots, N\}$. [12]

4.2.2 Repliement en phase

Soit un signal $g(t)$ de période P_0 échantillonné aux temps $t_1; \dots; t_N$. Un repliement en phase avec la période P_0 est la translation temporelle de tous les temps $t_1; \dots; t_n$ d'observations dans l'intervalle $[0; P_0[$. Les temps translattés $\hat{t}_1; \dots; \hat{t}_n$ dans l'intervalle $[0; P_0[$ sont tels que $\hat{t}_i = t_i - kP_0$, où $k \in \mathbb{N}$ vérifie la condition $t_i - (k+1)P_0 < 0$ et $t_i - kP_0 \geq 0$. [12]

4.2.3 Echantillonnage régulier

L'échantillonnage est la première étape indispensable à toute analyse statistique. En effet, l'ordinateur ne peut pas traiter les signaux analogiques ou continus. Il ne traite que des ensembles numériques.

Un processus d'échantillonnage régulier est défini par sa fréquence d'échantillonnage $f_e = 1/\delta_e$ où δ_e est le pas d'échantillonnage constant.

Dans le cas régulier, f_e doit vérifier la condition de Nyquist: elle doit d'être au minimum deux fois plus élevée que f_s la plus haute fréquence présente dans le spectre du signal:

$$f_e \geq 2 * f_s$$

Cette fréquence est nommée la fréquence de Nyquist (f_{Nyquist}). [1]

Les fréquences de Fourier autorisées :

- ✓ la plus petite fréquence $f_{e_{min}} = k/N\delta_e$
- ✓ la plus grande fréquence $f_{e_{max}} = 1/2\delta_e$

avec N est le nombre d'échantillons.

4.2.4 Echantillonnage irrégulier

Comme son nom l'indique, l'échantillonnage irrégulier implique que la fréquence d'échantillonnage f_e n'est plus constante au cours du temps. Ainsi la notion de fréquence de Nyquist n'est plus valable. En effet, cette dernière apparaît à cause de la périodicité dans l'échantillonnage. [12]

Nous distinguons un modèle d'irrégularité de l'échantillonnage, celui de l'échantillonnage à données manquantes: le signal est échantillonné avec une fréquence f_e constante, mais un nombre de données est manquant. Jones [13] est le premier à proposer un tel modèle de données manquantes dans le cadre de l'estimation spectrale d'un bruit blanc. [12]

Par ailleurs, l'échantillonnage irrégulier des signaux correspond le plus à la réalité puisqu'il n'est pas toujours possible de récupérer des mesures à des instants réguliers (conditions atmosphériques). Des facteurs externes font que la mesure ne peut pas être disponible à un instant précis. [14]

Pour remédier aux problèmes des irrégularités d'échantillonnage, plusieurs méthodes sont proposées dans la littérature [Jukovic 1971 ; Renson 1978 ; Scargle 1982 ; Heck, Manfroid et Mersch 1985]

4.2.5 L'interpolation

Interpoler, c'est compléter, par un procédé rapide, une suite de nombres dont on possède déjà quelques termes. Par extension, on appelle formule d'interpolation, une formule d'approximation qui sert à compléter une série temporelle à l'intérieur de l'intervalle d'observation. [15]

Géométriquement parlant, c'est déterminer une courbe lisse passant le plus près des points possibles de la fonction initiale. [15]

Dans la construction des tables numériques, dans la recherche des lois de phénomènes accessibles à l'observation, les méthodes d'interpolation sont d'un grand secours et d'une application constante. [15]

Parmi les différentes techniques d'interpolation, nous citons : :

- Interpolation linéaire ;
- Interpolation polynomiale ;
- Interpolation cubique (spline) : c'est une fonction définie par morceaux par des polynômes de degré 3. Elle est souvent préférée à l'interpolation polynomiale, car on obtient des résultats similaires en se servant de polynômes ayant des degrés inférieurs.

4.2.6 Méthode de régression linéaire : les moindres carrés

Connaissant la physique à l'origine du signal mesuré (variation de la vitesse radiale), la forme mathématique de la fonction à ajuster est connue. Seuls ses paramètres (période entre autres) sont à estimer. [16]

Soient N mesures $(\{x_1, y_1\}; \{x_2, y_2\}; \dots; \{x_N, y_N\})$ d'un signal donné, modélisées avec une fonction \mathcal{F} à m paramètres (p_1, p_2, \dots, p_m) (dans notre cas : $m=6$), l'ajustement le plus probable est celui qui maximise la probabilité jointe : [16]

$$\mathfrak{L} = \mathfrak{L}(y_1, y_2, \dots, y_N, p_1, p_2, \dots, p_m) \quad (4.3)$$

\mathfrak{L} représente la probabilité jointe pour que les mesures \mathbf{y}_i ($i = 1, \dots, N$) suivent la fonction \mathcal{F} aux paramètres \mathbf{p}_j ($j = 1, 2, \dots, m$). [16]

La probabilité de vraisemblance devient :

$$\mathfrak{L} = \prod_{i=1}^N \mathfrak{L}_i(\mathbf{y}_i, \mathbf{p}_1, \dots, \mathbf{p}_m) \quad (4.4)$$

$$\mathfrak{L} = \prod_{i=1}^N \frac{1}{\sigma_i^2 \sqrt{2\pi}} e^{-\frac{(y_i - \mathcal{F}(x_i, \mathbf{p}_1, \dots, \mathbf{p}_m))^2}{2\sigma_i^2}} \quad (4.5)$$

$$\log(\mathfrak{L}) = -\sum_{i=1}^N \log(\sqrt{2\pi} \sigma_i) - \frac{1}{2} \sum_{i=1}^N \frac{(y_i - \mathcal{F}(x_i, \mathbf{p}_1, \dots, \mathbf{p}_m))^2}{\sigma_i^2} \quad (4.6)$$

$$\log(\mathfrak{L}) = C^{te} - \frac{\chi^2}{2} \quad (4.7)$$

χ^2 est la somme de carrés de variables suivant la loi normale centrée-réduite $\mathcal{N}(0; 1)$. En effet, chaque \mathbf{y}_i est centré après soustraction de sa moyenne $\mathcal{F}(x_i, \mathbf{p}_1, \dots, \mathbf{p}_m)$, et divisé par sa variation standard σ_i . [16]

4.2.7 Séries de Fourier

Soit un signal périodique $S(t_i)$ tel que : $S(t_i + P) = S(t_i)$; $\forall i = \{1, \dots, N\}$ (4.8)

Il peut s'écrire sous la forme d'une série de Fourier :

$$S(t_i) = a_0 + \sum_{k=1}^m a_k \cos(2\pi k \frac{t_i}{P}) + \sum_{k=1}^m b_k \sin(2\pi k \frac{t_i}{P}) \quad (4.9)$$

Avec le nombre de paramètres $(2m + 1) \leq$ nombre d'échantillons (N)

$$a_0 = \sum_{i=1}^N S(t_i) \quad (4.10)$$

$$a_k = \sum_{i=1}^N S(t_i) \cos(2\pi k t_i) ; k = \{1, \dots, m\} \quad (4.11)$$

$$b_k = \sum_{i=1}^N S(t_i) \sin(2\pi k t_i) ; k = \{1, \dots, m\} \quad (4.12)$$

Dans le cas où t_i est régulier, 1 , $\cos(2\pi k \frac{t_i}{p})$, $\sin(2\pi k \frac{t_i}{p})$ constituent une base orthonormée, sachant que les projections (4.10), (4.11) et (4.12) ne sont valables que dans le cas de l'échantillonnage régulier.

4.2.8 Composantes Principale

D'après (2.7), tout signal $S(t_i)$ peut s'écrire de la forme :

$$S(t_i) = \sum_{k=1}^n [C_k * pca_k(t_i)] \quad (4.13)$$

Avec pca_k sont les axes de la nouvelle base orthonormée et C_l la projection du signal $S(t_i)$ sur les axes de la nouvelle base :

$$C_l = \sum_{i=1}^N [S(t_i) * pca_l(t_i)] \quad (4.14)$$

Cette expression n'est valable que dans le cas de l'échantillonnage régulier.

4.2.9 Périodogramme

Le périodogramme est un outil de visualisation du spectre de puissance lorsque celui-ci est tracé ou calculé en fonction de la période (ou la fréquence).

Le principe consiste à étudier successivement toutes les périodes possibles. Le spectre de puissance est donné par le carré de la Transformée de Fourier d'un signal $S(t_i)$:

$$\hat{S}(f) = TF[S(t_i)] \quad (4.15)$$

D'où :

$$\hat{S}(f) = \sum_{j=1}^N S(t_j) * e^{-i2\pi f t_j} \quad (4.16)$$

$$\hat{S}(f) = \sum_{j=1}^N S(t_j) \cos(2\pi f t_j) - i \sum_{j=1}^N S(t_j) \sin(2\pi f t_j) \quad (4.17)$$

$$\hat{S}(f) = \underbrace{\sum_{j=1}^N S(t_j) \cos(2\pi f t_j)}_{\text{Re}[\hat{S}(f)]} - i \underbrace{\sum_{j=1}^N S(t_j) \sin(2\pi f t_j)}_{-\text{Im}[\hat{S}(f)]} \quad (4.18)$$

Ainsi, le spectre de puissance $|\hat{S}(f)|^2$ égale à :

$$\mathbb{P}(f) = |\hat{S}(f)|^2 = \text{Re}^2[\hat{S}(f)] + \text{Im}^2[\hat{S}(f)] \quad (4.19)$$

par analogie au spectre de puissance Fourier, le spectre de puissance des Composantes Principales $\mathbb{P}(f)$ et d'après (4.8) et (4.9), est donné par :

$$\mathbb{P}(f) = \sum_{k=1}^{n_{pca}} C_k^2(f) \quad (4.20)$$

La période (la fréquence) principale correspond au pic de la puissance la plus élevée comme illustré dans la figure ci-après :

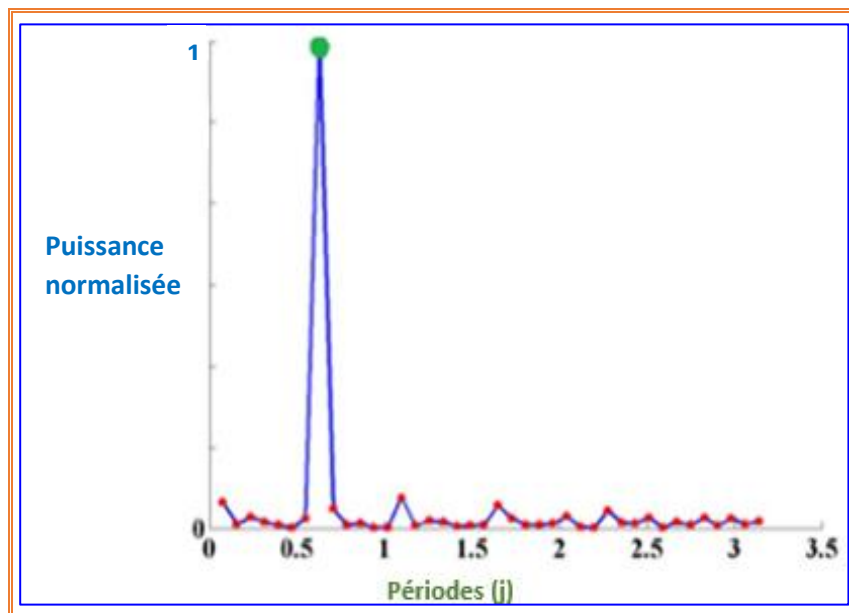


Figure 4.9.2 périodogramme du signal de la figure 4.2.1

4.3 Application

4.3.1 Premier cas : signal échantillonné régulièrement

Dans la **Figure 4.3.1**, notre signal représente la variation de la vitesse radiale en fonction du temps :

✓ **Paramètres de simulation** : $période = 0.5905j$; $e = 0.4$; $\omega = \pi/6$; $\gamma_0 = 10 \frac{Km}{s}$,
 $K = 100Km/s$; *type de bruit = blanc* ; $\delta_e = 0.005$; $N = 1181$ échantillons.

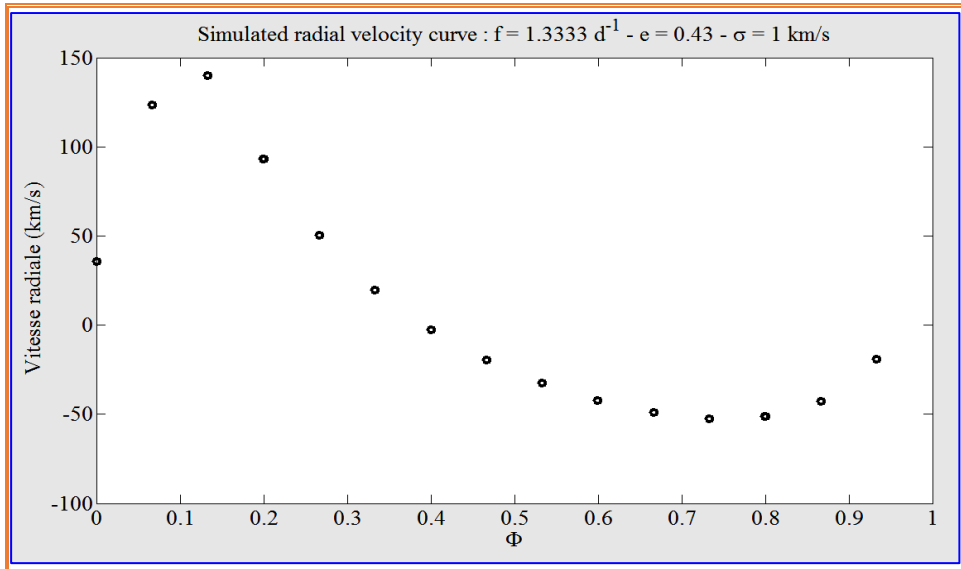


Figure 4.3.1.1 Variation de la vitesse radiale en fonction du temps (échantillonnage régulier)

Afin d'évaluer les performances de notre méthode, appelée « méthode standard », nous l'avons comparée à celle de la Transformée de Fourier (appelée aussi FFT). Les résultats obtenus sont présentés dans les périodogrammes ci-après :

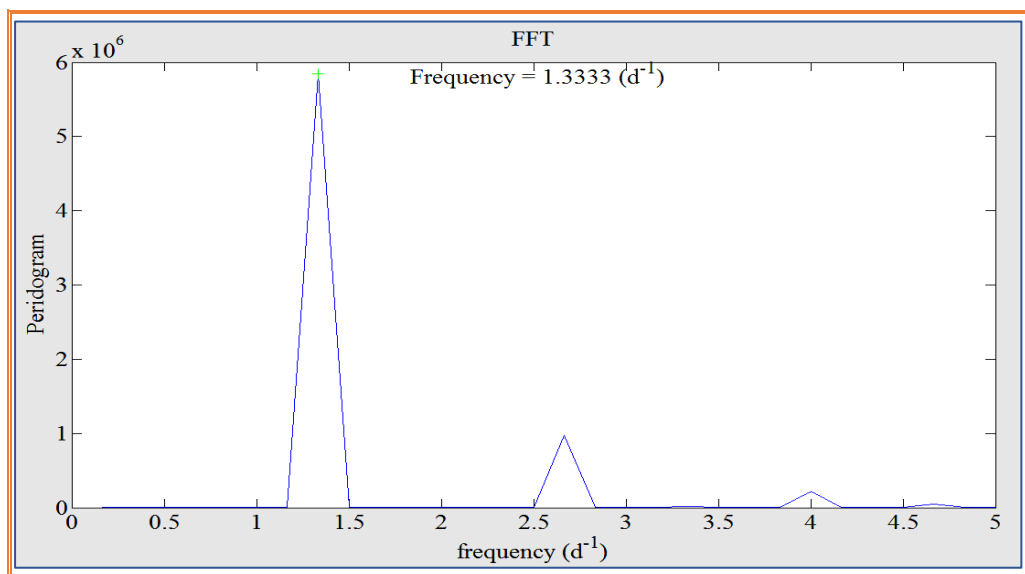


Figure 4.3.1.2 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode FFT

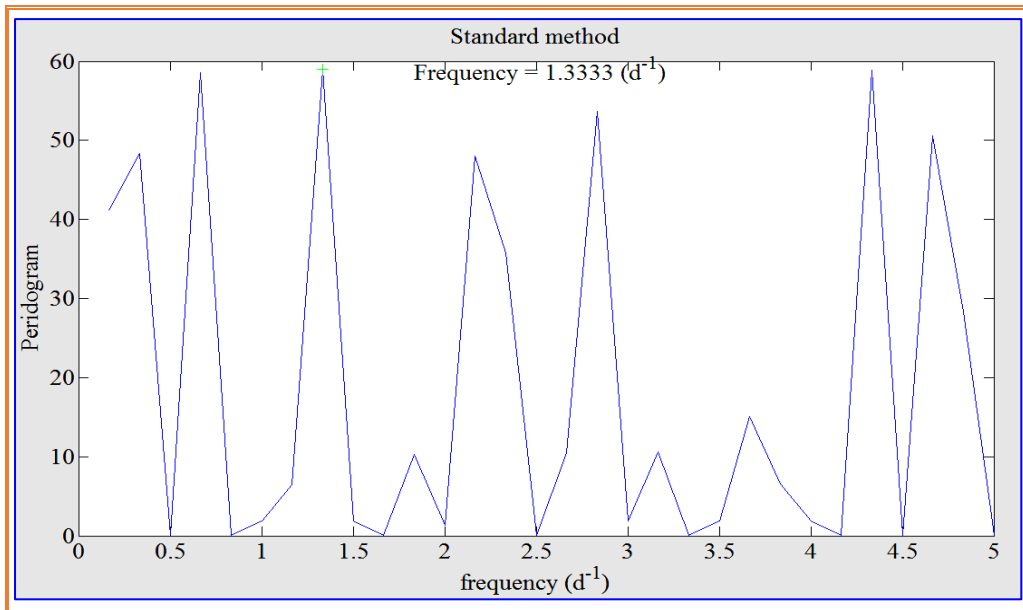


Figure 4.3.1.3 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode standard

D'après les figures obtenues, la méthode simplifiée présente l'avantage d'un temps de calcul court par rapport à celui de la méthode TF. Par contre, cette dernière présente l'avantage d'accentuer le pic le plus élevé par rapport à la première méthode ;

De plus, la méthode standard présente des fluctuations de fréquences ce qui n'est pas le cas de la méthode TF.

Cependant, la méthode standard devient trop fastidieuse pour un nombre d'échantillons grand. De ce fait, nous l'avons simplifiée par les techniques développées dans la section 4.2.

Le résultat obtenu par l'application de la méthode simplifiée sur notre signal est illustré dans la figure ci-après :

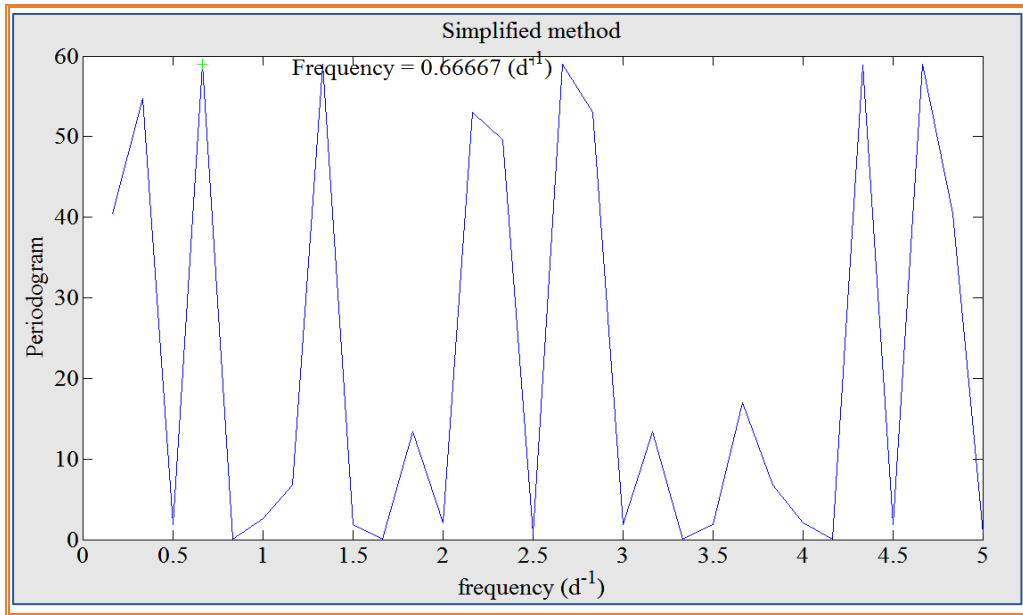


Figure 4.3.1.4 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage régulier par la méthode simplifiée

4.3.2 Deuxième cas : signal échantillonné irrégulièrement

Paramètres de simulation : *identiques à ceux de la partie précédente, à part que l'échantillonnage se fait par pas variable !*

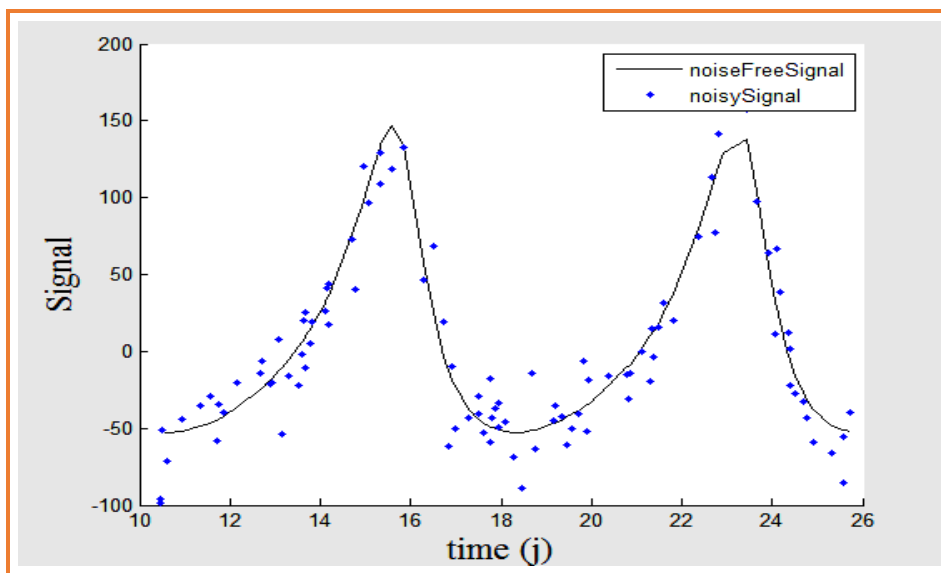


Figure 4.3.2.1 Variation de la vitesse radiale en fonction du temps dans le cas d'échantillonnage irrégulier

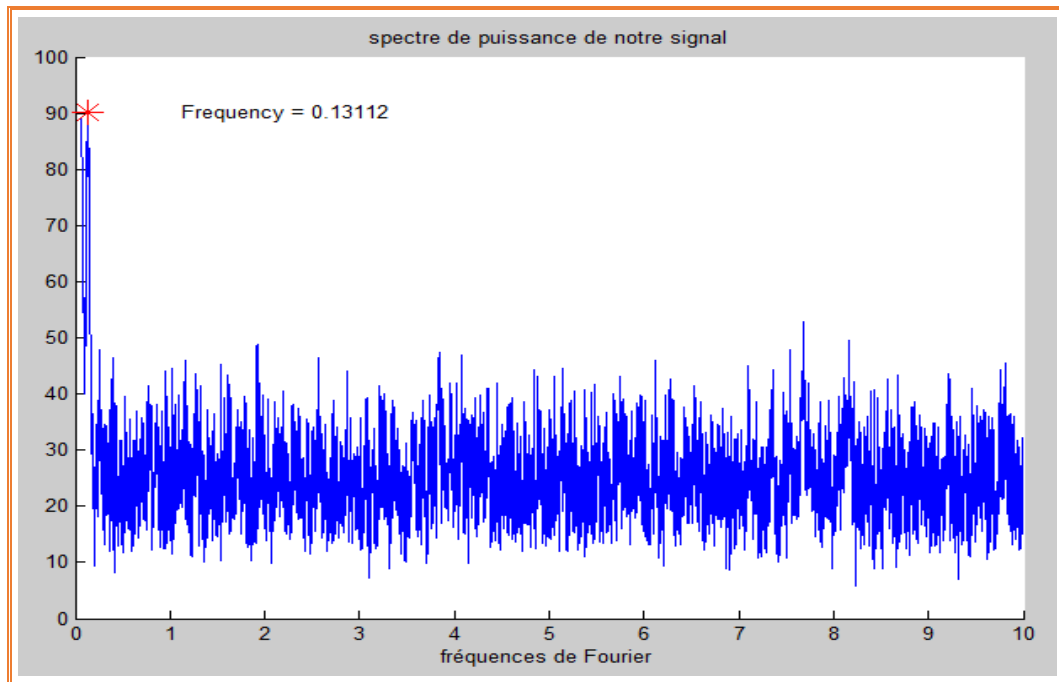


Figure 4.3.2.2 Périodogramme de notre signal étudié (vitesse radiale) dans le cas de l'échantillonnage irrégulier par la méthode simplifiée

Comme on peut le voir, la méthode simplifiée est aussi applicable dans le cas d'échantillonnage irrégulier.

4.4 Conclusion

A travers ce chapitre, nous avons développé et appliqué des outils indispensables au traitement de notre signal (variation de la vitesse radiale) pour trouver notre information recherchée : la période !

D'après les résultats obtenus, la méthode standard trop fastidieuse, prend des jours pour obtenir les résultats ; tandis que l'application de la méthode simplifiée autorise un temps de calcul court, une exécution rapide avec des résultats précis. De plus, comme la méthode standard n'est valable que dans le cas de l'échantillonnage régulier, la méthode simplifiée est utilisable dans les deux cas de l'échantillonnage (régulier et irrégulier).

I. Simulation de courbes de vitesse radiales

```
%% clean memories
clear;
clc;
close('all');
fclose('all');

%% Sampling
dates = (0:0.01:0.99)';
%dates = 200* rand(20,1); % time (days)
period = 1; % time (days)
Tp = 0; % time (days) temps de passage au périastre
% L'anomalie moyenne M, corresponds to the phases with respect to Tp
meanAnomalies = 2 * pi * (dates - Tp) / period; % angle (radian)
% reference time (very often Tp is unknown)
time0 = 0;

K = 1; % velocity (km/s)
gamma0 = 0; % velocity (km/s) centre de masse

startEccentricity = 0;
stepEccentricity = 0.05;
endEccentricity = 0.95;

startOmega = 0;
stepOmega = pi/6;
endOmega = 5*pi/6;

numberOfEccentricities = floor((endEccentricity - startEccentricity) /
stepEccentricity);
numberOfOmegas = floor((endOmega - startOmega) / stepOmega);
n = (numberOfEccentricities - 1) * numberOfOmegas + 1;
m = length(dates);
matrixOfCurves = zeros(n,m);
counter = 1;

for eccentricity = startEccentricity : stepEccentricity : endEccentricity

    % Compute eccentric anomalies E
    eccentricAnomalies =
computeEccentricAnomaly(meanAnomalies, eccentricity);

    % Cos and Sin E
    cosEccentricAnomalies = cos(eccentricAnomalies);
    sinEccentricAnomalies = sin(eccentricAnomalies);

    % Cos and Sin true anomalies v
    denominator = 1 - eccentricity * cosEccentricAnomalies;
    cosTrueAnomalies = (cosEccentricAnomalies - eccentricity) ./
denominator;
    sinTrueAnomalies = sqrt(1 - eccentricity * eccentricity) *
sinEccentricAnomalies ./ denominator;

    testedOmega = startOmega : stepOmega : endOmega;
    if eccentricity == 0.
        testedOmega = pi/2;
```

```

end

for omega
    = testedOmega

    % omega, K and gamma_0
    cosOmega      = cos(omega);
    sinOmega      = sin(omega);

    radialVelocities = gamma0 + K * (cosTrueAnomalies * cosOmega -
    sinTrueAnomalies * sinOmega + eccentricity * cosOmega);

    matrixOfCurves(counter,:) = radialVelocities;
    counter                    = counter + 1;

    figure(1)
    clf
    plot(dates,radialVelocities,'ok')
    xlabel('dates (days)','fontname','times','fontsize',20)
    ylabel('Radial velocities (km/s)','fontname','times','fontsize',20)

    phases          = (dates - time0) / period;
    reducedPhases   = phases - floor(phases);
    %[reducedPhases,permutations] = sort(reducedPhases);
    %radialVelocities = radialVelocities(permutations);

    figure(2)
    clf
    plot(reducedPhases,radialVelocities,'+g')
    xlabel('\phi','fontname','times','fontsize',20)
    ylabel('Radial velocities (km/s)','fontname','times','fontsize',20)
    title(['e = ',num2str(eccentricity),' - \omega = ',num2str(omega /
pi * 180),'^\circ'],'fontname','times','fontsize',20)

    pause
    close('all')

end
end

```


II. Calcul de Composantes Principales

```
function pcas = computePCAs(theMatrix,theThreshold)

m = size(theMatrix,2);

%n: nbr of lines & m:nbr of columns
covarianceMatrix = zeros(m,m);
%%why did we had initialized it into zeroes????????????
%it is in order to allocate a space memory from the beggining

for index1 = 1 : m

    firstMean = mean(theMatrix(:,index1));

    for index2 = index1 : m

        secondMean = mean(theMatrix(:,index2));
        crossMean = mean(theMatrix(:,index1) .* theMatrix(:,index2));

        covarianceMatrix(index1,index2) = crossMean - firstMean *
secondMean;
        covarianceMatrix(index2,index1) = covarianceMatrix(index1,index2);
    end
end

if max(max(abs(covarianceMatrix))) < 1e-10

    fprintf('The covariance matrix is null, we take one line as an eigen
line\n');
    % the lines of the image are all equal, we do not to decompose
    pcas = theMatrix(1,:);
    pcas = pcas';
else

    % the principal components are the columns of V matrix
    % the score of each component is the diagonal element of S
    [~,S,pcas] = svd(covarianceMatrix);
    %[V,S] = eig(covarianceMatrix);

    % Scores are the diagonal components of S (singular values)
    scores = diag(S);

    figure(2)
    hold on
    plot(1:length(scores),scores / scores(1),'ok')
    title('normelized Scores')

    sumOfScores = zeros(m,1);

    for d = 1:m
        sumOfScores(d) = sum(scores(1:d));
    end
```

```
figure(3)
hold on
plot(1:length(sumOfScores),sumOfScores)
plot(1:length(sumOfScores),sumOfScores(end) *
ones(size(sumOfScores)), 'r')
title('sum of Scores')

numberOfSignificativePCAs = find(sumOfScores <= theThreshold *
sumOfScores(end),1, 'last');
if isempty(numberOfSignificativePCAs)
    numberOfSignificativePCAs = 1;
end
%do numberOfSignificativePCAs must agree with the length of the
original img?

pcas = pcas(:,1:numberOfSignificativePCAs);
end

end
```