

Simulations et algorithmes de détection pour GAIA

Frédéric Arenou, Carine Babusiaux

Sept. 15, 1999

DASGAL, UMR 8633 du CNRS, Observatoire de Paris-Meudon-Nançay

5, Place J. Janssen, 92195 Meudon Cedex

Abstract

Dans le cadre de la préparation de GAIA, nous décrivons un outil de simulations d'images et le développement d'un algorithme de détection d'objets dans les CCD astrométriques "repéreurs d'étoiles". On détaille également les performances de l'algorithme, en termes de détection et de précision.

1 Introduction

Avant d'aborder les simulations des observations dans le plan focal de GAIA, il faut rappeler brièvement le principe des observations tel qu'il est défini actuellement. Les étoiles sont détectées dans l'Astro Sky Mapper (ASM). La première colonne de CCD, ASM0, permet la détection des étoiles brillantes ($V < 12$) avec un temps d'intégration plus court pour éviter la saturation. ASM1 s'occupe de la détection principale. ASM2 est là en redondance passive d'ASM0, ASM1 ou ASM3. En utilisant des fenêtres ("patches") autour des détections faites en ASM1, ASM3 permet d'éliminer les fausses détections (cosmiques), améliore le rapport S/B (bruit de lecture plus faible), et fournit, en combinaison avec ASM1, une indication de la vitesse de balayage. ASM1 ne sert qu'à détecter l'objet, pour pouvoir le suivre lors de son passage dans le plan focal du satellite, et à en donner les caractéristiques principales qui seront seules transmises au sol: les valeurs des super-pixels ("samples" = 2×2 pixels de 9 sur $27 \mu\text{m}$, soit 2×37 sur 6×37 mas) ne seront pas transmises à cette étape. Ces caractéristiques sont: le numéro de la ligne de CCD, la date, la position sur le CCD, le flux détecté et la limite du rapport signal sur bruit utilisé pour cette détection et éventuellement le fond du ciel autour de l'objet.

ASM0-2 sont donc les seuls CCD du plan focal astrométrique pour lesquels tous les samples sont lus. Les autres CCD (ASM3, les 17 colonnes du champ astrométrique principal et celles de photométrie à large bande) utilisent des "patches" autour des détections faites auparavant.

Autant dire que la partie “détection” faite en ASM1 est critique pour la suite des opérations (d’où la présence d’ASM2 en redondance).

L’algorithme de détection des objets dans ASM1 nécessite d’être robuste, c’est-à-dire de réagir aussi bien dans les champs peu denses que dans les champs denses, d’être le moins sensible possible au bruit, aux variations du fond de ciel et aux valeurs aberrantes. Il faut aussi qu’il soit rapide car cette détection se fait en “temps réel” à bord du satellite. Enfin, il faut une bonne précision, c’est-à-dire un taux de détection le meilleur possible avec le moins de fausses détections ainsi qu’une bonne estimation de la position et du flux de l’objet.

Deux approches de construction d’un tel algorithme existent. La première, approche descendante, consiste à prendre un algorithme de détection déjà élaboré et testé et de l’adapter aux caractéristiques du satellite et aux résultats attendus. La seconde, l’approche ascendante, consiste à partir des caractéristiques propres à GAIA pour construire un algorithme en le complexifiant petit à petit.

Deux algorithmes sont donc comparés ici, en élargissant l’étude préliminaire de X. Luri [9]: APM [6] pour l’approche descendante, et ce que l’on va appeler SWA (Sliding Window Algorithm) qui est un algorithme dont l’idée a été donnée par E. Høg [4] et que nous continuons d’améliorer et de complexifier. Il existe bien sûr d’autres algorithmes de détection, par exemple **SExtractor** [2], **extractor** [1], logical level method [10], mais ils ne se sont pas révélés plus performants en l’état qu’APM dans le cas de GAIA, soit qu’ils ne détectent pas assez profond, soit qu’ils donnent trop de fausses détections. Utiliser des méthodes plus sophistiquées, ondelettes ou Pixon [11], se révélerait prohibitif en termes de temps de calcul (le temps disponible est de l’ordre de grandeur de la seconde).

2 Simulations des données brutes

2.1 le ciel vu par GAIA

Il n’y a pas un, mais plusieurs CCD dans le plan focal de GAIA, avec des caractéristiques propres, qui ne sont sans doute pas encore définitives; d’autre part le ciel a un fort degré de complexité, et il faut savoir comment l’instrument le verra dans chaque cas. Ces deux raisons incitaient à se donner un outil de simulation du ciel réaliste et facilement paramétrable.

Le simulateur qui a été réalisé crée des images et peut en extraire les sources. Ces images sont composées d’objets (pour l’instant stellaires uniquement) placés sur un fond (background), acquises sur un instrument avec certaines caractéristiques.

Les objets stellaires peuvent être générés, en indiquant simplement les coordonnées galactiques de la direction de visée, à l’aide d’un modèle de la Galaxie [4], dans lequel l’extinction a été prise en compte, et qui fournit, outre les positions et couleurs, la cinématique des objets. Ce peuvent être également des binaires, ou des amas galactiques ou globulaires que l’on peut mettre à la distance de son choix.

Le fond de ciel peut être créé, soit par des étoiles en dessous de la limite de détection (21 à 24 mag) données par le modèle de Galaxie, soit en ajoutant un fond artificiel avec des variations périodiques, soit enfin un fond tiré d’une image du HST mises à disposition par le STScI [13]. Ces images ont été adaptées à GAIA; la taille du pixel, le temps de pose et le point-zéro des magnitudes ont été pris en compte pour cela, puis un filtre médian 5×5 a été passé sur ces images pour en éliminer les objets (et cosmiques) et ne garder que le fond.

En ce qui concerne l’observation par GAIA, il faut indiquer la taille de l’image, le temps de

pose, la localisation du CCD utilisé. Ce dernier renseignement permet d'indiquer les options par défaut qui seront prises, en ce qui concerne les aberrations pour la PSF, la bande passante (donc le point-zéro des magnitudes), le binning (par conséquent le bruit de lecture), sachant que ces options peuvent être modifiées par l'utilisateur. On peut également indiquer l'orientation de l'image, due au balayage du satellite. Il serait souhaitable dans le futur d'indiquer la date d'observation qui prendra en compte non seulement l'orientation, mais également l'influence du temps sur les positions (mouvement propre, parallactique, orbital).

Finalement, il est possible de choisir un algorithme de détection sur l'image obtenue, qui donne les positions et magnitude des objets. Le taux de détection peut être calculé de manière statistique puisque l'on peut faire plusieurs itérations: celles-ci permettent de faire varier le bruit (sans changer la localisation des sources), ce qui s'avère utile dans les champs peu denses.

Le résultat de la simulation se compose de la liste des positions et magnitudes des objets, du fichier FITS de l'image résultante, de la liste des positions et magnitudes obtenues après détection, de l'association de cette liste avec la liste initiale, du fichier FITS montrant les détections, et enfin des taux de vraies et fausses détections obtenues après plusieurs itérations.

Ce simulateur est accessible via <http://wwwhip.obspm.fr/cgi-bin/genimg>. Il s'agit d'un formulaire traité par un script perl, générant des commandes IRAF pour ce qui concerne le traitement d'images.

2.2 application au LMC

De multiples utilisations peuvent être faites à l'aide de ce simulateur. À titre d'exemple, on peut se demander si GAIA donnera la profondeur des Nuages de Magellan: à 50 kpc, vouloir une précision de 1 kpc demande donc une précision en parallaxe meilleure que $0.4 \mu\text{as}$. Comme aucune étoile n'aura cette précision, on peut imaginer utiliser des amas globulaires pour profiter de l'amélioration due au nombre d'étoiles. On a donc simulé un amas comme M92 à 50 Kpc auquel on a ajouté des étoiles de champ jusqu'à la magnitude 24 obtenues par le modèle de Galaxie (Figure 1). Sur les 15509 étoiles plus brillantes que $G = 24$, 964 de magnitude $15.8 < G < 21.8$ sont détectées par l'algorithme SWA décrit ci-dessous, avec un taux de détection de 57% à la magnitude 20, malgré la forte densité. Il faut noter qu'il s'agit des détections dans ASM1, sachant que toutes les étoiles dans un champ si dense ne pourront être observées dans les CCD "astrométriques".

En se basant sur une approximation de la précision en parallaxe de GAIA en fonction de la magnitude et de la latitude écliptique [3], on peut calculer la précision sur la parallaxe moyenne des étoiles de l'amas. Il faut néanmoins tenir compte des corrélations angulaires entre mesures individuelles, donc entre parallaxes, pour des étoiles situées dans le même champ (qui seront donc mesurées en fonction des mêmes étoiles dans le deuxième télescope). Il n'est pas simple de déterminer actuellement ces corrélations, et on supposera donc que l'amélioration pour n mesures, comme dans le cas d'Hipparcos[7], sera de l'ordre de $n^{-0.35}$ au lieu de $n^{-0.5}$. Après cette correction, la précision sur la parallaxe moyenne de cet amas globulaire s'avère de l'ordre de $3.6 \mu\text{as}$, soit 0.4 mag sur le module de distance. Malgré les hypothèses optimistes, il semble donc que la précision sera insuffisante pour mesurer la profondeur du LMC par cette méthode.

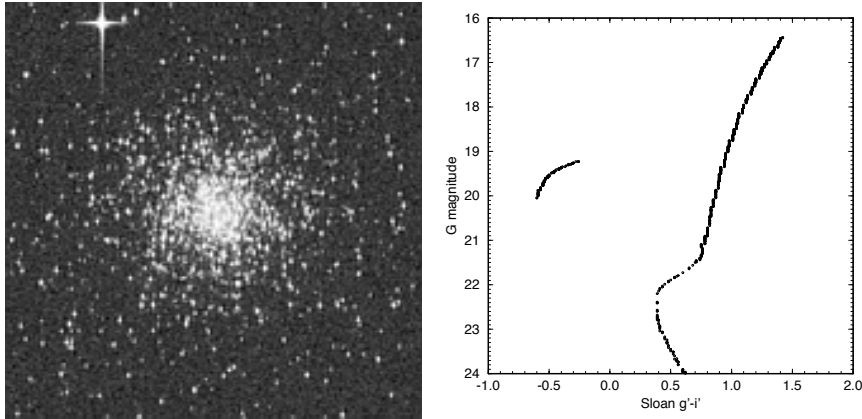


Figure 1: *simulation d'un amas comme M92 dans le LMC: à gauche, 40'' \times 40'' du champ dans ASM1; à droite, le diagramme couleur-magnitude initial (sans erreurs photométriques) des étoiles détectées par SWA.*

3 Comparaison d'algorithmes de détection

3.1 deux approches

L'algorithme d'APM n'est pas exactement adapté à GAIA, par exemple en ce qui concerne la PSF. Son fonctionnement actuel est le suivant: (i) estimation d'une carte du fond de ciel: l'image est segmentée en sous-parties sur lesquels est calculée la valeur "moyenne" du fond de ciel. (ii) détection des objets principaux: les samples connexes situés au-dessus d'un certain seuil au dessus du fond de ciel sont gardés en mémoire. (iii) multiseuillage: en observant l'objet primaire à des intensités plus élevées, vérification qu'il n'est pas composé de plusieurs objets. (iv) calcul des paramètres (position, flux, forme).

Quant à l'algorithme SWA, il est basé sur l'idée d'une fenêtre glissante et est seulement adapté à la recherche d'étoiles et non d'objets étendus: (i) recherche des pics (ii) détermination du fond de ciel sur des samples choisis autour du pic trouvé. (iii) calcul du rapport signal sur bruit sur le pic; si ce rapport est supérieur à une certaine limite, on garde la détection. (iv) calcul des paramètres (position, flux). Détaillons un peu plus ce nouvel algorithme:

Pour améliorer le temps de calcul de l'algorithme, on n'effectue la recherche de pic que sur les samples dont la valeur est supérieure à 28; cette dernière valeur est celle que l'on obtiendrait pour une étoile de magnitude 21 centrée entre 4 samples. La recherche de pic s'effectue sur l'image lissée par un filtre binomial afin de diminuer le bruit. Sur cette image, un pic est défini comme un sample dont la valeur est supérieure aux 8 samples adjacents. Pour éliminer les pics uniquement dûs au bruit, on utilise un test du χ^2 sur les 9 samples centrés sur le pic, au seuil de 3σ gaussiens.

Le fond de ciel est estimé par une médiane tronquée: on calcule la médiane sur 72 samples choisis autour du pic, et on ne garde que les samples dont les valeurs sont à moins de 3σ de la médiane et on recalcule celle-ci, puis sa précision. La médiane est calculée avec la méthode Hoare [12], 20% plus rapide que la méthode classique demandant un tri préliminaire de toutes les valeurs.

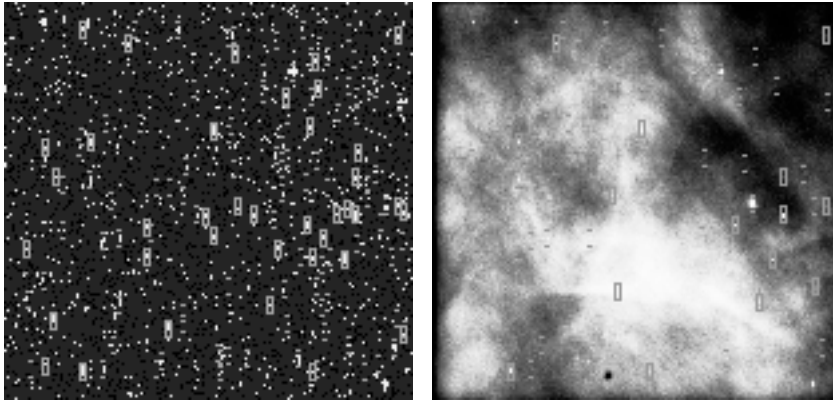


Figure 2: *simulation (34"×34") d'un champ en direction du centre galactique avec un modèle de Galaxie, sans bulbe, donnant 977 étoiles jusqu'à la magnitude 24. Les détections par SWA sont encadrées. Le taux de détection est de 99% à la magnitude 20 dans la première image. À droite, on a ajouté une image d'Orion du HST, et le fond peut atteindre 11 mag.sec²; le taux de détection descend à 71% et 9 fausses détections apparaissent.*

Une première estimation de la position est donnée par les coordonnées du pic. On calcule la position précise de l'étoile par la méthode du barycentre. Il est calculé à l'aide de 25 samples si l'étoile est plus brillante que la magnitude 16, et sur 9 samples sinon; c'est le compromis qui permet à la fois la meilleure précision et la moindre contamination par des samples éloignés qui seraient trop bruités. Ce n'est que pour les étoiles faibles (> 19), que l'ajustement de la PSF se révèle meilleur que le barycentre, mais en augmentant d'un facteur 2 le temps de calcul; la précision sur la position servant essentiellement à retrouver l'étoile lors de son parcours dans le plan focal, la précision du barycentre devrait être suffisante, mais ceci n'est valable que dans le cas d'une PSF sans aberrations; en réalité, le centre de la PSF est décalé sur ASM1 d'environ 0.18 sample en x et 0.01 sample en y .

Pour le calcul du flux, on utilise la photométrie d'ouverture corrigée, plus précise que la photométrie par PSF. Ceci est dû au fait que la PSF de GAIA est tellement piquée qu'une erreur, même faible, sur le centrage de l'étoile sur le sample provoque une erreur sur la photométrie.

3.2 SWA versus APM

Pour tester les deux algorithmes, on peut choisir quelques images représentant les problèmes que pourra rencontrer l'algorithme lors de la mission: champs denses et champs à fond variable. Le champ le plus dense simulé par le modèle de Galaxie est au centre galactique (1645 étoiles jusqu'à la magnitude 21, 8056 étoiles jusqu'à la magnitude 24, soit plus de 8 millions par degrés carrés). À ces objets est ajouté un fond, soit artificiel (5 zones où le fond varie de 17 mag.sec² à 25 mag.sec²), soit celui d'une image d'Orion obtenue par le HST: le fond de ciel varie ici très rapidement, et est de l'ordre de 1000 e^- par sample = 15 mag.sec² (Figure 2).

Toutes les images simulées ici ont les caractéristiques de ASM1, (c'est-à-dire binning 2×2 , bruit de lecture 10.9 e^- , etc) et sont de la taille d'un demi-CCD, i-e 2780×1075 samples, soit 750 000 samples couvrant un champ de $\approx 10^{-3}$ degré², sauf les images du HST qui sont de plus petite taille (918×298 samples GAIA).

Les deux algorithmes traitent une telle image en environ 2 secondes (sur une station Alpha

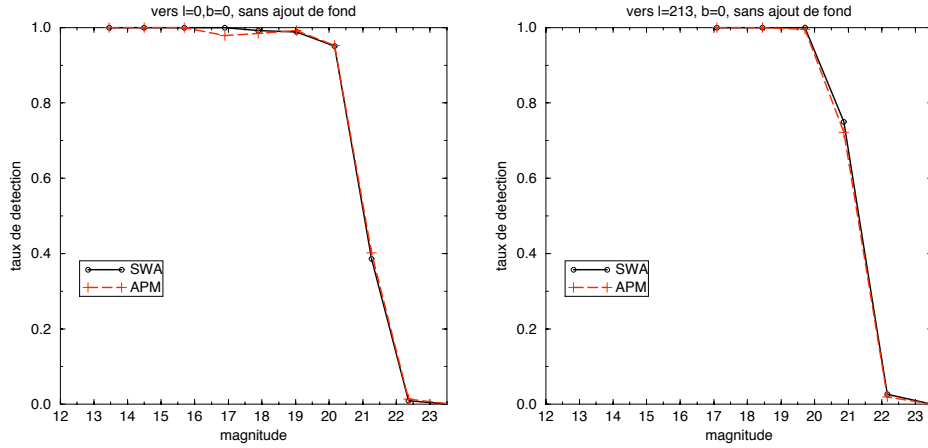


Figure 3: *Taux de détection en fonction de la magnitude. À gauche pour un champ en direction du centre galactique, à droite en direction de l'anticyentre (80 fois moins dense). Le fond de ciel n'est composé que d'étoiles faibles ($21 < G < 24$).*

Digital). L'implémentation de l'algorithme à bord du satellite devra optimiser au mieux ce temps de calcul de manière à le rendre compatible avec le mode TDI.

Sur des champs très denses, mais sans variation importante du fond de ciel, le taux de détection est supérieur à 95% pour les étoiles de magnitude 20, de 60% à la magnitude 21 et de 20% à $G=21.5$ (Figure 3). Dans des régions du ciel plus représentatives, on peut considérer que la complétude des étoiles simples est assurée jusqu'à la magnitude 20.5.

En revanche, le nombre de fausses détections n'est que peu dépendant de la densité. Ce nombre est faible en général, mais on observe de temps en temps des fausses détections trouvées plus brillantes que 20. Elles sont dues à du bruit situé à proximité d'une étoile brillante; ce problème est identique pour les deux algorithmes. La seconde source de fausses détections est donnée par les rayons cosmiques, qui seront de toutes façons éliminés par ASM3.

La grande différence entre les deux algorithmes se situe en réalité dans les champs à fond important et variable: la fenêtre de SWA permet d'estimer le fond de ciel au voisinage proche de l'objet et cet algorithme détecte donc beaucoup plus d'objets, et avec une meilleure précision. Sur une image du centre galactique avec Orion en fond, SWA détecte 71% des étoiles plus brillantes que 20, là où APM n'en détecte que 6%. Sur une image avec un fond artificiel, il n'y a aucune fausse détection pour SWA, mais APM en trouve 100, situées à des régions bien précises de l'image, sans doute car le fond est calculé sur une zone dont la taille est de même ordre que les variations artificielles de cette image.

Les précisions en magnitude des deux algorithmes sont équivalentes, mais on observe un biais croissant avec la magnitude lors de l'estimation de la magnitude par APM; ce biais sera facilement éliminé lorsque la forme de la PSF de GAIA sera intégrée dans l'algorithme (Figure 4).

La précision sur la position le long du scan est de l'ordre du 1/10 ème de sample (soit 7 mas) à la magnitude 20 et de 1/5 ème de sample (soit 15 mas) à la magnitude 21 (Figure 5). On a de même une précision de 22 mas à la magnitude 20 et de 45 mas à la magnitude 21 perpendiculairement au scan. Ces chiffres ont été obtenus à partir des images simulées avec une PSF sans aberration; ils seront peut-être moins bons avec la vraie PSF d'ASM1. On observe cependant dans le cas du

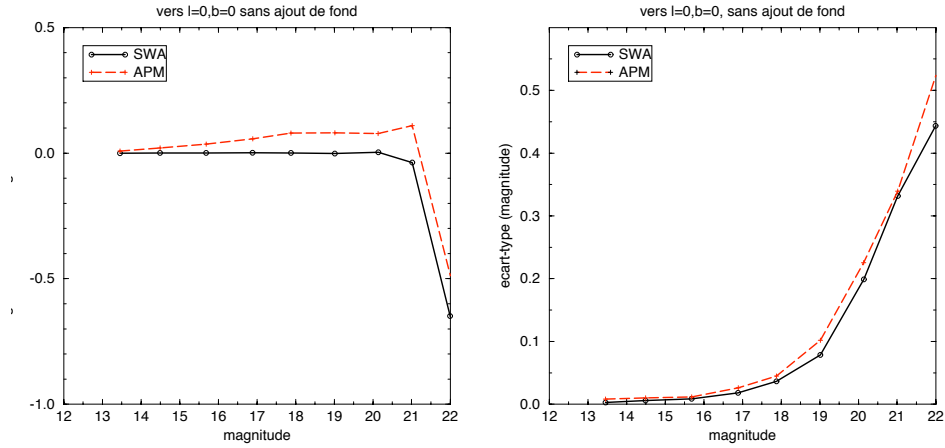


Figure 4: *Biais (à gauche) et précision (à droite) en magnitude des objets détectés en direction du centre galactique.*

centre galactique une saturation de la précision en dessous de la magnitude $G=18$, contrairement à l'amélioration en $\approx 10^{G/5}$ attendue jusqu'à la magnitude 12 quand on prend en compte le bruit de photon seul.

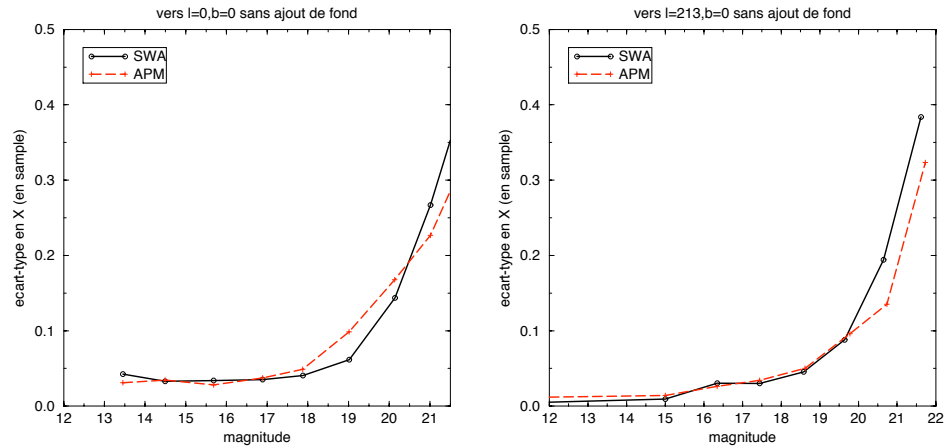


Figure 5: *Précision en position (en sample) le long du balayage en direction du centre galactique (à gauche) et de l'anticyentre (à droite).*

C'est une des questions qui se posent, mais bien d'autres auront à être traitées dans le futur proche, comme par exemple le comportement dans ASM3, l'influence de la binarité, la fonction de sélection des étoiles dans les champs denses, etc.

Remerciements

Le simulateur d'images a bénéficié des compétences de Misha Haywood, Xavier Luri et Noël Robichon, qui sont ici remerciés.

RÉFÉRENCES

References

- [1] Alard, C. (1996) Thèse de doctorat
- [2] Bertin, E. and Arnouts, S. (1996) SExtractor: Software for source extraction, *A&A Suppl.*, **117**, 393-404
- [3] gaia SAG (1999) Results of the Concept and Technology Study, eq 83
- [4] Haywood, M. (1997) *A&A*, **320**, 428 et *A&A*, **320**, 440
- [5] Høg, E. (1999) Sky survey and photometry by the GAIA satellite, SAG_CUO_53
- [6] Irwin, M.J. (1985) Automatic analysis of crowded fields, *MNRAS* 214, 575-604
- [7] Lindegren, L. (1988) in 'Scientific aspects of the Input Catalogue Preparation II', January 1988, Sitges, J. Torra, C. Turon, (eds.)
- [8] Lindegren, L. (1998) Point Spread Functions for GAIA including aberrations, SAG_LL_025
- [9] Luri, X (1998) Source detection at the PSM, SAG_IWG_002 & SAG_IWG_003
- [10] Paillou, P. (1998) Detecting objects on varying sky background: The logical level method, GAIA workshop, Leiden, 23-27 nov 1998
- [11] Puetter, R.C. and Yahil, A. (1999) The pixon Method of Image Reconstruction, ADASS VIII conference, nov 98, astro-ph 9901063
- [12] Sedjewich R., Algorithms in C, Add.Wesley
- [13] STScI, Hubble Data Archive, <http://stdatu.stsci.edu>