# MEMO

Auteur : J.E Campagne

Nançay/Abell85/17.11.11

**DRAFT**

Diffusion : R. Ansari, P. Colom, Ch. Magneville, J. M. Martin, M. Moniez, A. Torrento

**Objet** : Status de l’analyse des données sur Abell85 (2011)

# Introduction

Ce rapport d’étape couvre la période des prises de données prises entre le 10 avril 2011 et le 12 août 2011 prises sur l’amas Abell85. Les observations en mode ON ont été centrées sur les coordonnées galactiques suivantes J2000: RA = 0.72139/0h43m17s et

|  |
| --- |
| Figure Vue d'ensemble de l'amas Abell85 (otique et X). L’elipse rouge indique le beam ON (1fois FWHM). |

DEC = -9.16306/-9h09m47s, et pour mémoire le NRT a un lobe (FWHM) de 4’(RA) x 22’(DEC) à 21cm. Le redshift de cet amas est de 0.055, mais les deux galaxies (Figure 2) dans le lobe du NRT qui donnerait un signal HI sont à un z de 0.050 donc on s’attendrait à une émission HI vers 1353MHz.

Figure 1 donne une vue d’ensemble optique et X de l’amas, avec en rouge matérialisé le beam du NRT. Les cercles bleus étant des galaxies émettrices HI. Un zoom autour du beam du NRT est présenté

|  |
| --- |
| Figure Zoom autour du beam ON. On distingue 2 galaxies A85[DFL98]461 et A85[SDG98]3114. L’ellipse verte représente 2 fois le beam FWHM contrairement à la Figure 1. |

Etant donné que la prise de données enchaîne des cycles ON puis OFF, le OFF de chaque cycle est décalé de 60sec en RA. Pour le moment, je n’ai pas de carte correspondante du OFF.

Dans la suite est résumé l’analyse faite sur 500cycles, soit au total 1000min correspondant à un temps utile sur le ON de 5000sec tenant compte que l’on prend 30sec de donnée avec une efficacité de 33%.

# Résumé de la réduction de données

Je ne décris pas dans la suite la chaîne d’électronque mais seulement les traitements à partir des fichiers « signal » sauvés par la DAQ et transférés sur Irods.

## Etape préliminaire

Chaque prise de données (une par jour) est découpée selon les différentes phases DAB[[1]](#footnote-1) ON, ON, DAB OFF et OFF. Pour ce faite primo une extraction des headers FITS de tous les fichiers « signal » est nécessaire pour secundo sélectionner les fichiers dont les paquets correspondent aux fenêtres de temps (start/stop) des différentes phases issues d’une analyse automatique des fichiers SCA. Cette étape préparatoire est cruciale car les ressources allouées aux traitements des données (ici la taille du scratch) au CCIN2P3 sont limitées et l’on ne peut par exemple charger l’ensemble des fichiers signals brutalement, cela correspondrait à une taille typique de 400MB/fichier x 2polars x 1000 fichiers.

## Première réduction

Une fois l’étape préliminaire effectuée, on procède à l’étape la plus couteuse en temps CPU à savoir le traitement FFT, et ensuite à une réduction sous forme de spectres de 2-dimension (polars x freq.). On définit les traitements suivants :

* GAIN : on utilise le cycle du milieu du run courant pour définir autour de la DAB OFF deux zones pour obtenir d’un spectre lissé également en fréquences (c’est essentiellement un intermédiaire de calcul) ;
* CALIB (ON ou OFF) : on utilise les phases de DAB avec des marges pour bien les englober ;
* DATA (ON ou OFF) : on utilise les phases ON ou OFF avec une marge de sécurité pour s’affranchir des DAB et des mouvements du NRT.
* MEAN (ON ou OFF) : on utilise les mêmes données que pour le traitement de type DATA mais cette fois on calcule une moyenne et non la médiane sur un certain nombre de paquets sans procéder à aucun filtrage[[2]](#footnote-2).

La réduction procède toujours à un filtrage médian sur un certain nombre de paquets et un filtrage en fréquence est appliqué sauf pour les périodes ON et OFF. LeTableau 1 détaille les paramètres de cette première réduction de données.

Tableau Paramètres pour lancer les traitements (FFT et première réduction). Phase : période dont sont issues majoritairement les paquets. « #de paquets/fenêtre» est le nombre de paquets sur lequel est calculé soit la médiane soit la moyenne. « Durée fenêtre » est la durée en seconde moyennant 8kHz de trigger (3sec/25600paq.). « Shift/durée » sont le shift en seconde pour sélectionner les paquets selon le temps par rapport au début présumé de la phase correspondante écrite dans le fichier SCA. « Filtrage Freq. »  est la largeur totale de la fenêtre glissante sur laquelle est appliquée le filtrage si mentionnée.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Traitement | Phase | # de paquets/fenêtre | Durée fenêtre | Shift/durée | Filtrage Freq. |
| GAIN | DAB-OFF | 5120 | sec | 4sec/14sec | 3MHz |
| CALIB ON/OFF | DAB-ON/OFF | 1024 | ~120ms | 4sec/14sec | 3MHz |
| DATA ON/OFF | ON/OFF | 5120 pour obtenir une médiane et on moyenne 5 médianes | ~5x0.6sec = 3sec | +3sec/30sec | NON |
| MEAN ON/OFF | ON/OFF | 25000 pour obtenir moyenne est déviation standard | ~3sec | +3sec/30sec | NON |

## Seconde réduction

Une fois la première réduction effectuée, on dispose d’une collection segmentée par cycle et période ON/OFF de spectres 2D : 2 canaux/polars et 8192 fréquences avec 30kHz de résolution. A part le traitement MEAN qui produit des valeurs moyennes, tous les autres traitements fournissent des médianes et, pour le traitement DATA on réduit encore en effectuant une moyenne sur 5 médianes[[3]](#footnote-3).

La réduction suivante est différente selon le type de traitement : GAIN, CALIB, DATA.

Pour tous les traitements post-GAIN, on procède à la normalisation par le spectre obtenu lors du traitement GAIN. Cette normalisation est un intermédiaire de calcul efficace pour automatiser la calibration, mais *in fine* pour les données ON et OFF on procède à une normalisation par le OFF du cycle courant lissé par un filtrage médian de 2MHz de largeur totale. Notons qu’il n’y a pas été constaté de différence entre une normalisation par le spectre GAIN, ou bien par le OFF courant lissé, même si le Tsys peut augmenter sur quelques cycles en début ou fin de run.

### Gain : intermédiaire de calcul

Pour le GAIN, on procède à la moyenne des spectres dont la puissance totale ne dépasse pas 1% de la puissance minimale constatée sur les 14sec traitées.

|  |
| --- |
| Figure 3 Exemple de spectre de GAIN |

La Figure 3 donne un exemple de ces spectres de gain. Il est a noté que ces spectres ont changé début juillet 2011 comme l’a constaté A. Torrento non seulement pour Abell85 mais les autres amas pointés également à la même période (Figure 4).

|  |
| --- |
| Screen shot 2011-11-07 at 11  Figure 4 Spectre de gain à partir du 8 Juil. 2011. |

### Coefficients de calibration par la DAB

La phase de calibration par la DAB procède en plusieurs étapes :

* On détermine d’abord pour chaque spectre la puissance totale dans une bande de fréquence fixée. On dispose alors du suivi en temps de cette puissance tous les 120ms sur une période couvrant en principe la période d’injection du signal de la DAB en deux endroits du cornet (cf. Figure 5);

|  |
| --- |
| Figure Exemple de la puissance normalisée de l'ensemble des spectres issus du traitement CALIB au cours des cycles du run du 7 mai 2011. Le t0=0sec correspond au temps des DAB du fichier SCA. La fréquence utilisée dans cet exemple est de 1386MHz avec une largeur de 6.25MHz. |

* Par un algorithme automatique rendu possible par la normalisation par le gain précédemment déterminé, on peut extraire les spectres dans les deux fenêtres en temps [3sec, 1sec] et [6sec, 8sec]. En en faisant une moyenne, nous obtenons le niveau de base probablement relié au Tsys. Ensuite, en fixant un seuil à 20% de la différence entre la valeur maximum et minium de puissance au dessus du minium de puissance, on détermine les 2 nivaux de DAB correspondant aux 2 injections à des endroits différents du cornet. Sur chaque niveau, on extrait la médiane de la puissance, et finalement on prend la moyenne de ces 2 médianes pour définir le coefficient de calibration BAO du cycle concerné et pour les 2 canaux. En faisant la moyenne de tous les cycles d’un run donné, on obtient le coefficient pour chaque canal valable pour ce run.

L’étude de la chromaticité de la DAB ainsi que son influence sur les coefficients de calibration BAO pour corriger les données ON/OFF sera décrite dans un autre MEMO. Cependant, par la suite je n’utilise pas ces calibrations et présente que des données non calibrées qui paraissent bien plus stables.

### Données : ON, OFF et différences

Suite à la première réduction, on procède simplement au calcul des moyennes des spectres associés à chaque traitement DATA ON et OFF pour chaque cycle sans les affecter d’une calibration quelconque ni d’un filtrage fréquentiel.

## Analyse en bout de chaîne

Cette étape se charge de combiner les résultats des différents traitements de la seconde réduction. Si la première réduction est la plus lourde et nécessite de soumettre des jobs en batch avec les restrictions et rigidités afférentes, cette dernière étape est la plus flexible est se fait en interactif en associant programme compilé et script propre à l’environnement SPIAPP.

# Résultats actuels

Les résultats qui sont présentés dans la suite sont issus de la phase d’analyse précédente donc sont à prendre comme « résultats préliminaires ». Ils concernent : la réduction des RFIs, la stabilité de la prise de données par l‘étude du résidus du HI galactique, la variation de la sensibilité en fonction du temps.

## Réduction des RFIs

Pour mettre en évidence le gain de l’usage de la médiane par rapport à la moyenne pour filtrer les RFIs, j’ai utilisé le run du 28 avril 2011 pour procéder en premier lieu à l’obtention de 100 moyennes et déviation standards (pour les 8192 fréquences) obtenues par le traitement MEAN (ON et OFF). Avec ces valeurs, on peut définir une variable réduite

|  |  |
| --- | --- |
|  | Eq. 1 |

Dans cette formule, *i* coure sur les 100 valeurs, *meani* et *i* désignent les moyennes et déviations standards obtenues avec les *Npaq.per.win* = 25000 paquets par fenêtre. Le dénominateur n’est autre que l’erreur sur la moyenne d’un échantillon connaissant la déviation standard empirique. Cette variable réduite suit sauf accident une loi normale N(0,1). La distribution de celle-ci sur l’ensemble des fenêtres de type OFF sur la bande de fréquence de 250MHz de large est présentée sur la Figure 6. L’effet des RFIs est clairement visible, en dehors de ces accidents les variables réduites suivent la distribution attendue.

|  |
| --- |
| Figure Variable réduite dans le cas des moyennes et déviations standard obtenues sur 100 fenêtres de 25000 paquets prises sur les données OFF. Sauf accident, la moyenne est 0, tandis que la déviation standard est 1. Les accidents reflètent l’effet des RFIs. L’abscisse correspond à la bande totale de fréquences entre 1250MHz et 1500MHz avec une résolution de 30kHz. Les deux couleurs correspondent aux deux canaux (bleu : Ch 0, rouge : Ch 1). |

Avec les mêmes données utilisées cette fois selon le traitement DATA ON/OFF, on obtient environ 100 médianes obtenues avec 25600 paquets. Afin d’obtenir une variable réduite similaire à l’Equation 1, il convient de la définir selon

|  |  |
| --- | --- |
|  | Eq. 2 |

La notation est similaire à celle utilisée pour Equation 1. Le dénominateur tient compte de la loi exponentielle de l’intensité reçue[[4]](#footnote-4). Tout comme précédemment, cette variable réduite sauf accident suit une loi normale N(0,1). La distribution de celle-ci sur les fenêtres de données OFF et l’ensemble de la bande de fréquence de 250MHz de large est présentée sur la Figure 7. On observe qu’en dehors des accidents qui reflètent les RFIs résiduels, les varibles réduites suivent bien la loi attendue.

|  |
| --- |
| Figure Variable réduite dans le cas des médianes obtenues sur 100 fenêtres de 25600 paquets prises sur les données OFF. Sauf accident, la moyenne est 0, tandis que la déviation standard est 1. Les accidents reflètent l’effet des RFIs. L’abscisse correspond à la bande totale de fréquences entre 1250MHz et 1500MHz avec une résolution de 30kHz. Par rapport à la Figure 6 il faut noté le changement d’échelle verticale. |

La comparaison des Figure 6 et Figure 7 est frappante : une bonne partie des 250MHz est nettoyée qui plus est l’échelle verticale est bien réduite également. Pour être plus précis un zoom de la bande [1330, 1400]MHz est présenté avec la même échelle verticale sur les Figure 8 (moyenne) et Figure 9 (médiane). S’il reste deux RFIs dont un particulièrement sévère, les autres sont nettoyés par l’usage de la médiane sur les paquets.

|  |  |
| --- | --- |
| Figure Avec la technique de la moyenne | Figure Avec la technique de la médiane |

## Résidus galactiques

Avec l’ensemble des 500cycles correspondant à 5000sec effectives sur le Ciel (ON), on peut calculer run par run et cycle par cycle de la différence d’intensité calculée avec les données ON et données OFF. Uniquement les médianes déterminées par le traitement DATA ON/OFF sont utilisées sans filtrage fréquentiel, sauf pour la normalisation définie comme les spectres OFF lissés par un filtrage médian obtenus avec une largeur de 2.1MHz. Cette procédure de normalisation remplace celle avec un spectre de GAIN unique par run. Ensuite, la moyenne suivante est utilisée :

|  |
| --- |
|  |

Dans la formule ci-dessus « *c* » et « *r* » indicent les cycles et les runs, et « *f* » indique le canal de fréquence. Le résultat de cette opération est présenté sur Figure 10.

|  |
| --- |
| Figure Spectre moyenné sur 500cycles des différences ON-OFF normalisées par le OFF lissé du cycle courant (5000sec effective). La résolution en fréquence est de 30kHz. Les deux couleurs différencient les deux canaux/polarisations (bleu : Ch 0, rouge : Ch 1). |

Le zoom sur la zone protégée à Nançay à savoir la bande [1400-1427]MHz est présenté sur la Figure 11. On y voit le résidu de la différence entre l’intensité de la raie HI d’origine galactique reçue des zones pointées lors du ON et du OFF. On remarque qu’a part cette raie, la zone est bien exempte de RFIs.

|  |
| --- |
| Figure Zoom sur la bande protégée à Nançay de la Figure 10. |

## Stabilité

La raie à 1420MHz résiduelle peut être utilisée pour suivre la stabilité des prises de données. En particulier en moyennant l’intensité sur les deux bandes [1418,1419]MHz et [1422,1423]MHz on suit la ligne de base de part et d’autre de la raie. Le résultat pour les deux polarisations/canaux est donné sur Figure 12 sur l’ensemble des 500cycles, tandis que l’histogramme des valeurs est donné sur Figure 13.

|  |  |
| --- | --- |
| Figure Evolution de la ligne de base autour de la raie HI résiduelle. | Figure Histogramme des valeurs de la Figure 12. |

On note que la prise de données est bien stable alors qu’aucune calibration n’a été appliquée. A ce propos, les Figure 14, Figure 15 et Figure 16 présentent la même évolution avec l’application des coefficients de calibration moyennés run par run, ou bien variant tout au long des cycles.

|  |  |
| --- | --- |
| Figure Evolution identique à la (même échelle) avec une calibration définie run par run. | Figure Identique à Figure 14 avec une calibration évoluant selon le cycle (on garde la même échelle verticale par comparaison). |

Cette calibration utilisant la DAB n’est pas satisfaisante.

|  |
| --- |
| Figure Identique à la Figure 15 cette fois en laissant libre l'échelle verticale. |

En soustrayant la ligne de base sur la bande [1420.2, 1420.6]MHz, on peut obtenir l’intensité de la raie et suivre son évolution au cours des cycles (Figure 17). On note la même stabilité que pour la ligne de base (Figure 12). A noter que l’influence des calibrations est moins nette quand on soustrait la ligne de base ce qui masque l’influence des variations des coefficients de calibrations…

|  |  |
| --- | --- |
| Figure Evolution de l'intensité de la raie HI résiduelle. | Figure Histogramme des valeurs de la Figure 16. |

Une autre façon de juger de la stabilité de la prise des données et de moyenner l’intensité dans la bande [1400,1415]MHz[[5]](#footnote-5) sur les spectres ON et OFF séparément. Le résultat est montré sur Figure 19 et un zoom sur Figure 20. On constate des variations d’un run à l’autre et même à l’intérieur d’un run c'est-à-dire cycle à cycle. La procédure par contre ON-OFF indépendament pour les 2 canaux annulent ces variations comme on a pu le constater sur les Figure 12 et Figure 17.

|  |
| --- |
| Figure Evolution de la moyenne de l'intensité dans la bande [1400,1415]MHz pour les 2 canaux et le OFF et ON séparément : bleu : Ch 0/OFF, cyan Ch 0/ON, rouge Ch 1/OFF, orange Ch 1/On. |
| Figure Zoom de la Figure 19. |

## \*Evolution de la sensibilité en fonction du temps d’intégration

En sommant l’intensité de la bande protégée [1400, 1420]MHz (cf. Figure 11) on peut suivre au cours du temps l’évolution de la déviation standard du bruit. Pour cela, ayant 500 cycles, on les rassemble en lots de (1, 10, 25, 50, 100) cycles[[6]](#footnote-6) dont on peut tirer la moyenne et la déviation standard. Ainsi, l’on peut porter l’évolution de cette déviation standard cumulée, en fonction du nombre de cycles intégrés (cf. Figure 19). La courbe noire est là pour guider l’œil, elle s’appuie sur le point à « num of cycles = 1 » du canal 0 et une simple dépendance en 1/√ « num. of cycles ». Pour être un peu plus quantitatif, la relation reliant la déviation standard (processus ergodique périodique) pour 1 polar est donnée par

Comme je travaille en a.u est que l’on normalise essentiellement par le Tsys donc le résultat de la Figure 19, est le rapport  Tsys avec Δν = 30kHz et Δt = 10sec x n avec n le nombre de cycles intégrés (10sec tient compte de l’efficacité). Avec ces facteurs on obtiendrait (conditionnel car il y a des facteurs √2) une relation

Ce qui pour le moment colle assez bien avec ce que l’on observe.

|  |
| --- |
| Figure Evolution de la déviation standard pour les 2 canaux/polarisations du bruit intégré sur la bande [1400-1420]MHz en fonction du nombre de cycles utilisés. Par exemple, les points à « num. of cycles = 25 » correspondent à la déviation standard de la distribution de 20 mesures de l’intensité sommée sur la bande de fréquence en question. La courbe noire est là pour guider l’œil, elle s’appuie sur le point à « num of cycles = 1 » du canal 0 et une simple dépendance en 1/√ « num. of cycles ». |

## Zoom sur la zone HI de l’amas : 1350MHz

Dans le lobe du NRT, deux galaxies sont susceptibles de donner un signal HI à 1353MHz. Le zoom de la différence ON-OFF dans cette zone est présenté sur la Figure 22. Pour savoir si les augmentations d’intensité qui seraient suggérées, la Figure 23 montre indépendamment le ON et le OFF moyen sur l’ensemble des cycles. On y voit un gros RFI à 1350MHz identifié comme le radar de Roissy, et d’autres RFIs entre 1355 et 1356MHz, mais la zone semble néanmoins chaotique ce qui ne permet pas de conclure positivement sur la présence de raie HI à 1353MHz.

|  |
| --- |
| Figure 22 Différence ON-OFF sur OFF lissé pour l’ensemble des 500cycles pour les 2 |

|  |
| --- |
| Figure Spectres ON (haut) et OFF (bas) normalisé par le OFF lissé pour les 2 canaux. |

1. DAB pour Diode A Bruit. [↑](#footnote-ref-1)
2. Ceci a été utilisé sur une faible partie des données pour montrer la réduction des RFIs avec la médiane. [↑](#footnote-ref-2)
3. Faute de mémoire vive suffisante en batch on ne peut effectuer des calculs de médiane sur un fichier signal complet. [↑](#footnote-ref-3)
4. La loi de la somme du carré de 2 variables gaussiennes est un 2 à 2 degré de liberté, soit une exponentielle [↑](#footnote-ref-4)
5. On évite des influences éventuelles de la raie à 1420MHz en se limitant à 1415MHz. [↑](#footnote-ref-5)
6. On en a respectivement (500, 50, 20, 10, 5). [↑](#footnote-ref-6)