

# Les données Eros II



Jean-Noël Albert – LAL / IJClab

Juin 2020

Eros II, c'est 90 millions d'étoiles analysées à partir de 2 millions d'images FITS correspondant à 7 ans de campagne.

C'est 2 000 nuits d'observation.

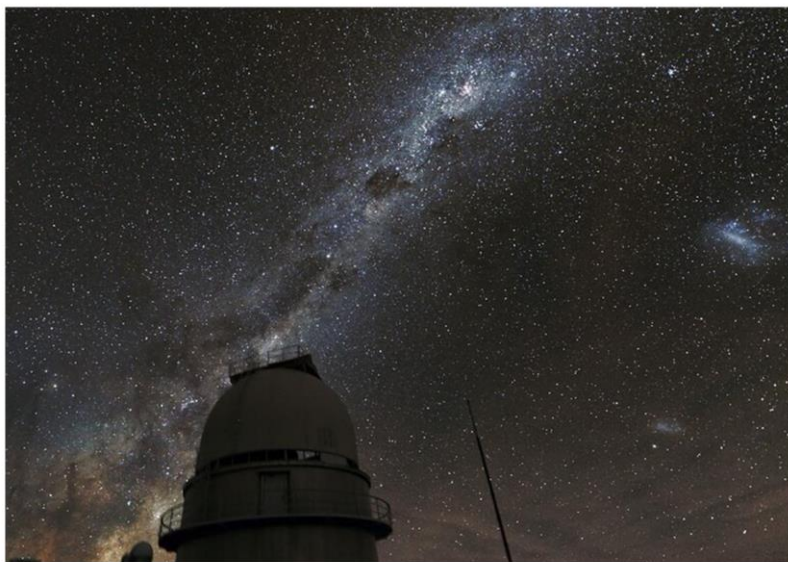
C'est plus de 60 pointages du télescope par nuit soit une prise de vue toutes les 5 à 10 mn pour près de 1000 images chaque nuit.

Toutes ces observations ont été conduites grâce au télescope de 1m de l'expérience, le Marly, à La Silla, le site de l'ESO au Chili de Juin 1996 à Février 2003.

# Eros II

2

90 millions d'étoiles  
2 millions d'images  
7 années de campagne  
2000 nuits d'observation



La Voie lactée à La Silla (Chili).  
En haut à droite, le Grand Nuage de Magellan,  
en bas à droite, le Petit Nuage (ESO/Z.Bardon)

Eros II, c'est 90 millions d'étoiles analysées à partir de 2 millions d'images FITS correspondant à 7 ans de campagne.

C'est 2 000 nuits d'observation.

C'est plus de 60 pointages du télescope par nuit soit une prise de vue toutes les 5 à 10 mn pour près de 1000 images chaque nuit.

Toutes ces observations ont été conduites grâce au télescope de 1m de l'expérience, le Marly, à La Silla, le site de l'ESO au Chili, de Juin 1996 à Février 2003.

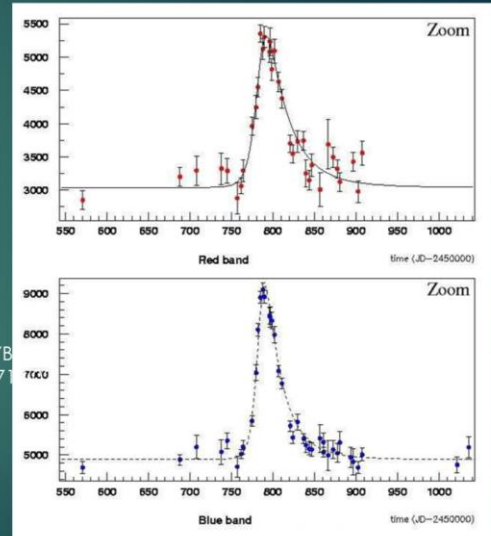
# Microlentilles gravitationnelles

- ▶ Recherche de microlentilles gravitationnelles
  - ▶ Objets obscurs massifs dans le halo galactique
  - ▶ Variations du flux de la lumière d'étoiles lointaines
  - ▶ Observation des mêmes zones sur une longue durée
- ▶ Exemple d'une courbe de lumière
- ▶ Exemple du fichier des mesures

```

$ head lm0153n22431.time
# star: erosid      MagR  ErrMR  XR      YR      MagB  ErrMB  XB      YB
#   lm0153n22431 19.133  0.124  478.78  684.37  19.238  0.070  431.84  7
#
#   date      MagR  ErMagR  MagB  ErMagB
348.81416  19.224  0.037  19.330  0.025
366.79045  99.999  9.999  19.337  0.035
373.68669  19.115  0.051  19.299  0.032
376.84666  19.195  0.058  19.287  0.036
380.86332  19.210  0.052  99.999  9.999
381.70421  19.186  0.081  99.999  9.999

```



P. Tisserand – fg 9.4b pg 191

L'objectif de l'expérience est la recherche d'effets de lentilles gravitationnelles résultant du passage de masses sombres devant des étoiles lointaines, comme celles des Grand et Petit nuages de Magellan. Le phénomène conduit à une variation significative de la lumière de l'étoile, comme le montre le diagramme (thèse de Patrick Tisserand, 2004).

Pour pouvoir détecter ces phénomènes rares, il faut observer régulièrement les mêmes étoiles sur une longue période. 7 ans dans le cas d'Eros II, soit 2 445 jours, pour 2 000 nuits d'observation.

Après analyse des différentes images, les mesures réalisées sont regroupées sous la forme de fichiers appelés « courbes de lumière », à raison d'un fichier par étoile. Les 90 millions d'étoiles suivies par Eros II correspondent à autant de courbes de lumière ASCII, accessibles individuellement dans le système de stockage Irods.

Un extrait d'un tel fichier est présenté ici.

# Champs et Programmes

4

- ▶ Division du ciel en grandes zones : les « programmes scientifiques »
  - ▶ Grand et Petit Nuages de Magellan
  - ▶ Centre galactique
  - ▶ Bras spiraux de la galaxie
  - ▶ ...
  
- ▶ Chaque zone est divisée en plusieurs « champs » représentant la zone vue par le télescope lors d'un pointage
  - ▶ Grand Nuage : 88 champs
  - ▶ Petit Nuage : 13 champs
  - ▶ Centre galactique : 141 champs
  - ▶ Bras spiraux : 36 champs

L'expérience Eros I, plaques et CCD, s'intéressait principalement au Grand Nuage de Magellan, puis au Petit. Le gain en efficacité constitué par Eros II a permis d'élargir le champ d'activité de l'expérience à plusieurs autres études, comme le Centre galactique et les Bras spiraux.

Pour conduire toutes ces recherches, le ciel austral est divisé en plusieurs régions correspondant aux différents programmes scientifiques.

Ces régions sont elles-mêmes divisées en zones correspondant au pointage du télescope. Ces zones sont nommées « champs ». Le nombre de champs par programme scientifique dépend de la région observée.

Les principaux programmes représentent 278 champs, c'est-à-dire 278 cibles pour le télescope, pointés régulièrement durant les 7 années d'observation.

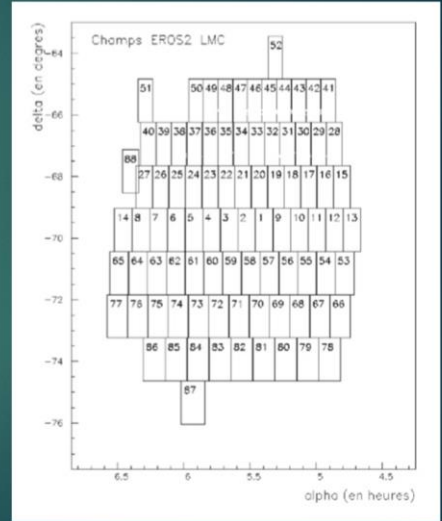
# Les Champs du Grand Nuage



Le Petit

Les Nuages de Magellan

Le Grand



P. Tisserand – fg 4.6b p82

Une vue des Grand et Petit Nuage de Magellan et la disposition des 88 champs couvrant la région du Grand Nuage (thèse de Patrick Tisserand).

# Images et CCD

6

- ▶ Un télescope
- ▶ Un prisme divisant la lumière du télescope en deux flux : rouge et bleu
- ▶ Deux caméras de 8 CCD de grande taille : 2048 x 2048 pixels 16 bits
  - ▶ Chaque caméra reçoit la lumière des mêmes étoiles de la même zone au même moment.
- ▶ 16 images de 8 MB capturées lors d'une même prise de vue
  - ▶ Une moyenne de 60 champs par nuit
  - ▶ Près de 1 000 images par nuit
  - ▶ A peu près 8 GB

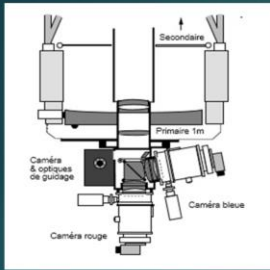
La lumière issue du télescope est divisée grâce à un prisme en deux flux, correspondant l'un aux longueurs d'ondes bleues et l'autre aux rouges. Chaque flux est dirigé vers une caméra électronique équipée de 8 CCD de grande taille (pour l'époque) de 2048x2048 pixels de 16 bits, soit 8 Mo par image.

Chaque observation produit donc 16 images FITS (2 caméras de 8 CCD) soit 128 Mo.

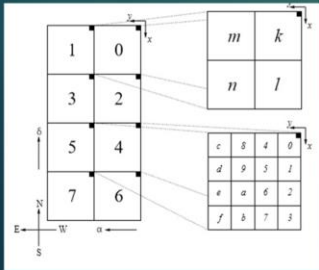
Chaque nuit, 60 observations en moyenne sont effectuées produisant près de 1000 images représentant 8 Go.

# La double caméra

P. Tisserand – fg 4.1 pg 75



P. Tisserand – fg 4.2 pg 76



Le diagramme de l'appareillage de prise de vue, avec les deux caméras placées derrière le prisme séparant la lumière captée par le télescope en un flux bleu et un flux rouge (thèse de Patrick Tisserand).

La disposition des CCD dans la caméra et leur numérotation, ainsi qu'une vue des découpages possibles des images issues des CCD, utilisés dans les traitements afin de réduire le volume des données à analyser (thèse de Patrick Tisserand).

Deux vues du télescope, avec un gros plan sur l'appareillage de prise de vue (Jna).

# Images brutes et réduites

8

- ▶ Les images issues des CCD sont des images « brutes »
- ▶ Elles sont « normalisées » grâce à des images « de calibration » pour former les images « réduites » utilisées dans les analyses.
- ▶ La réduction est réalisée directement sur le site.
  
- ▶ Il y a donc une image « brute » pour chaque image « réduite ».
  
- ▶ Images brutes, réduites et de calibration envoyées par DLT.

Les images issues des CCD sont nommées dans la terminologie Eros « images brutes ».

Ces images ne sont pas directement utilisées dans les analyses. Elles sont préalablement « calibrées » (on dit « réduites »). Cette réduction est réalisée sur le site grâce à des images de « calibration », acquises généralement au début ou à la fin de nuit.

Les images étaient transmises par DLT, le réseau numérique jusqu'à La Silla ne permettant pas de gros transferts de données. Mais il supportait fort heureusement le courrier électronique.

Durant le temps du transfert, les données étaient conservées localement sur le site.

# Transfert des images

9

- ▶ Les images brutes sont sauvées sur des DLT « brutes »,  
Les images réduites sont sauvées sur des DLT « réduites »
- ▶ Les DLT sont envoyées à Saclay
  - ▶ Les DLT brutes restent à Saclay
  - ▶ Les DLT réduites continuent vers le CC à Lyon
- ▶ Les images réduites sont recopiées sur des 3490
  - ▶ L'emplacement des fichiers et les principales caractéristiques des images sont enregistrées dans la base de données Oracle de l'expérience
  - ▶ La gestion des cartouches et du placement des fichiers est assuré par un environnement développé autour de TCL.
- ▶ Plusieurs réorganisations ont été réalisées autour du HPSS, des streamers modernes et enfin de Irods

Images brutes, images réduites et images de calibrations étaient expédiées sur des DLT depuis le site de La Silla, au Chili, vers Saclay. Les images brutes étaient conservées à Saclay et les images réduites continuaient vers le Centre de calcul de l'IN2P3 où elles étaient transférées sur des cartouches 3490, les moyens de stockage les plus efficaces de l'époque (800 Mo...)

Les fichiers reçus et les principales caractéristiques des images sont répertoriés dans une base de données Oracle avec leur localisation sur les cartouches.

L'environnement développé pour la gestion des fichiers et des cartouches utilisait TCL, un langage de scripting permettant aux administrateurs de réaliser eux-mêmes leurs outils.

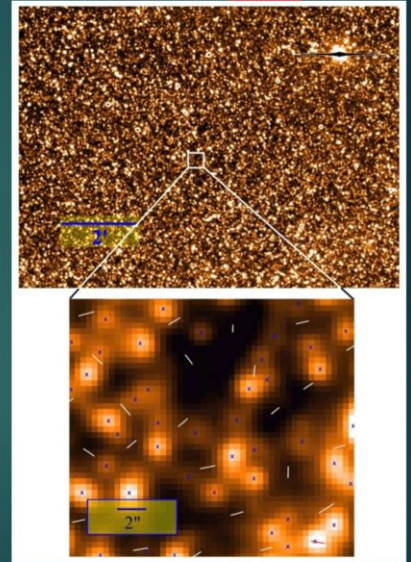
Plus tard, les cartouches 3490 ont été remplacées par des supports de plus grande capacité gérés grâce au robot de cartouches HPSS, ce qui a donné lieu à plusieurs restructurations des données. La dernière des réorganisations est le passage à Irods.

La base de données a perduré, mais l'outil TCL, qui n'avait plus de raison d'être, a été remplacé par un environnement plus performant centré sur Java, le pont Objet/Relationnel Hibernate, et Jargon/NIO, une intégration Java de la librairie Irods.

# Les images Eros II

10

- ▶ 2 027 572 images référencées dans la base de donnée
- ▶ 1 884 001 images effectivement présentes dans Irods, soit 14.3 TB
- ▶ 1 832 463 images réduites, soit 14.0 TB
- ▶ 80 650 images perdues, soit 4.0 %, dont l'ensemble du programme « Naines rouges »
  - ▶ Episodes de crash des cartouches
  - ▶ Erreurs lors des restructurations ?



P. Tisserand – fg 2.8 pg 41

Les images Eros sont au format FITS, le standard de l'astronomie.

Plusieurs types d'images existent :

- des images brutes, conservées pour l'essentiel à Saclay, même s'il existe quelques exemplaires à Lyon ;
- les images réduites, conservées à Lyon ;
- des images de calibration, de différents types : obscurité et coupole – brutes ou réduites ;
- des images construites à partir d'autres images, afin d'en améliorer la qualité – nommées images composées.

L'image présentée est une vue du centre du Grand nuage de Magellan, une zone dense de 20 000 étoiles. L'image du bas est un gros plan de 20" d'arc montrant les étoiles (voir la thèse de P. Tisserand).

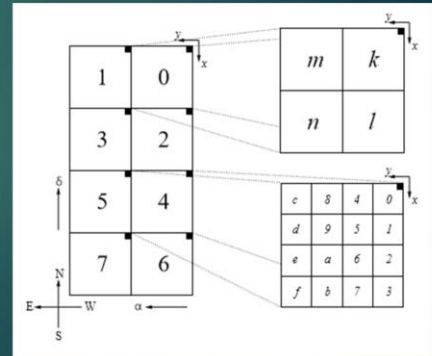
Les images sont disponibles dans Irods, Mais la migration de HPSS vers Irods a mis en évidence la perte de 80 000 des 2 millions d'images référencées, soit 4%. Les pertes semblent distribuées de manière assez aléatoire, à l'exception des images du programme « Naines rouges », qui ont toutes disparues. L'une des explications à ces pertes est la destruction de plusieurs cartouches du HPSS. Mais la disparition de l'ensemble du programme « Naines rouges » n'est pas expliquée.

# Quart de CCD

11

- ▶ Le « quart de CCD » est un sous-ensemble d'un CCD de 1600 x 1600 pixels
  - ▶ Rappel : un CCD = 2048 x 2048 pixels 16 bits
- ▶ Les quarts de CCD sont nommés « k », « l », « m » et « n ».
- ▶ Un quart de CCD : 1600 x 1600 pixels 16 bits, soit 6.2 Mo
  - ▶ Rappel : une image CCD = 8 Mo

P. Tisserand – fg 4.2 pg 76



Un « quart de CCD » est un sous-ensemble de la zone couverte par un CCD.

En pratique, pour compenser les fluctuations inévitables du pointage du télescope, la taille d'un « quart de CCD » est sensiblement plus large que le quart théorique. Un CCD correspond à 2048 x 2048 pixels alors qu'un quart de CCD couvre 1600 x 1600 pixels. Soit une image FITS d'environ 6.2 Mo, contre 8.0 Mo pour une image CCD.

La numérotation des CCD d'une caméra et des quarts de CCD correspondant est présentée dans le diagramme issu de la thèse de Patrick Tisserand.

En pratique, le découpage en quart de CCD est virtuel. Les images réduites ne sont pas réellement découpées. Ce découpage n'affecte que la construction des catalogues d'étoiles et des fichiers des mesures.

# Traitement des données

12

- ▶ Une étape sur le site : la réduction des images brutes
- ▶ Plusieurs étapes au Centre de calcul
  - ▶ Création de très bonnes images pour une meilleure détection des étoiles
  - ▶ Détection des étoiles et création de catalogues
  - ▶ Mesure des paramètres des étoiles des différentes images
  - ▶ Création de courbes de lumière ASCII par étoile
- ▶ Pour réduire la taille des fichiers et les volumes traités, les analyses sont conduites par quart de CCD et par flux de couleur

L'analyse des images est conduite grâce au logiciel Peida, développé, principalement par Réza Ansari et Christophe Magneville, pour les besoins de l'expérience. Les traitements, portant sur les images réduites, sont réalisés en plusieurs étapes.

La première étape est l'identification des étoiles observées. Pour améliorer l'efficacité du programme de reconnaissance, une série de quelques bonnes images est sélectionnée pour chaque zone à analyser afin de créer une image dite « composée ». Ces traitements ayant été réalisés au Centre de calcul, ces images sont disponibles dans Irods.

L'identification des étoiles permet de créer des catalogues, conservés sous la forme de fichiers dit « fichiers de références », ou plus simplement « références ». Ces fichiers utilisent un format binaire propre à l'expérience. Ils ont été créés au Centre de calcul où ils sont disponibles.

La détection des étoiles est conduite indépendamment pour les images bleues et les images rouges, d'où l'existence de catalogues « bleus » et de catalogues « rouges ». Une ultime phase, dite « association », reprenait le catalogue bleu et le catalogue rouge d'une même zone et associait les étoiles présentes dans les deux catalogues (*du moins en théorie : le nombre d'étoiles associées d'un catalogue rouge étant différent du nombre du catalogue bleu correspondant – un point à éclaircir !*).

La seconde grande phase des traitements est la mesure des paramètres des étoiles présentes sur les différentes images réduites. Les étoiles analysées sont celles décrites dans les catalogues. Les résultats sont conservés dans des fichiers dit « suivis », également dans un format binaire propre à l'expérience. Pour une même zone du ciel, il y a donc des « suivis bleus » et des « suivis rouges ».

A l'issue des analyses, les mesures des différentes étoiles sont extraites des suivis et enregistrées dans des fichiers ASCII regroupant les données bleues et les données rouges. Il y a un tel fichier par étoile. Ces fichiers ASCII sont nommés « courbes de lumière ».

Toutefois, du fait des limitations des moyens de stockage, des puissances de calcul et des espaces disques à l'époque des analyses, les mesures étaient réalisées sur des sous-ensembles des images correspondant aux « quarts de CCD ». Il y a donc quatre flux d'analyse par CCD en deux couleurs (rouge et bleue).

# Production

13

- ▶ Traitement massif des images pour une version du logiciel d'analyse et un ensemble de paramètres.
  - ▶ Réalisé en batch sur la ferme Anastasie du Centre de calcul
  - ▶ En général par programme et sous-ensembles d'un programme
- ▶ Les références et les suivis sont identifiés par un code de traitement et un numéro de version, correspondant à une « production ».
- ▶ Les « productions » P1 et P5 représentent 85% des catalogues et 75% des mesures.
- ▶ Seule la production P5 a conduit à la création de courbes de lumière.

Une « production » est le traitement systématique des images de l'expérience. Elle est correspondant à une version du logiciel, et à un ensemble de paramètres.

Les productions sont conventionnellement identifiées par un code de traitement et un numéro de version. Les principales productions sont la P1 et la P5, représentant à elles-seules près de 75% des mesures réalisées.

Seule la production P5 a conduit à la création de courbes de lumière. Elle est de ce fait considérée comme LA production de référence.

En pratique, les traitements étaient réalisés en batch, sur la ferme de stations UNIX de Centre de calcul nommée Anastasie (le "a" du nom des machines « cca » du Centre).

# Majeurs, mineurs et exotique

14

- ▶ Trois « catégories » de programmes : « majeurs », « mineurs » et « exotiques »...
  - ▶ Les **7 majeurs** : programmes entièrement analysés avec des courbes de lumière
    - ▶ Beta Scuti (BS), Centre galactique (CG), Gamma Normae (GN), Gamma Scuti (GS), Grand et Petit nuage de Magellan (LM, SM), Theta Muscae (TM)
    - ▶ 7 programmes sur 33 référencés
    - ▶ 278 champs sur 1594 référencés
    - ▶ 1 763 282 images réduites sur les 1 974 701 référencées ~ 90 % des observations
  - ▶ **4 mineurs** : programmes analysés sans courbes de lumière
    - ▶ Tous ces programmes sont marqués « Monte-Carlo », copie de CG et LMC

33 programmes scientifiques sont référencés dans la base de données, correspondant à la définition de près de 1 600 champs.

Il existe des images réduites pour 25 des 33 programmes, mais ceci ne représente que 596 des 1 594 champs référencés. L'essentiel des observations a porté sur les Nuages de Magellan, le Centre galactique et 4 Bras spiraux.

Et sur les 33 programmes référencés, seuls les 7 principaux programmes ont été entièrement analysés, jusqu'à la génération de courbes de lumière, et ceci pour la seule production de référence P5.

Ces 7 programmes sont nommés les « 7 majeurs ». Ils ne représentent que 278 champs sur les 1 594 champs référencés dans la base de données, mais 1.7 millions des 1.9 millions d'images réduites, soit 90% des observations.

4 autres programmes ont également été analysés, c'est-à-dire que des références et des suivis existent, mais aucune courbe de lumière n'a été produite. Ces « 4 mineurs » semblent correspondre à des simulations Monte Carlo réalisées à partir des images réduites des programmes LMC et CG.

# Image composée

15

- ▶ Très bonnes images créées à partir d'une petite série de très bonnes images pour améliorer la détection des étoiles
  - ▶ 12-15 images par série
  - ▶ Les images créées sont nommées « images de références »
  - ▶ Elles sont produites par couleur et par quart de CCD
- ▶ 12 652 images de références calibrées astronomiquement
  - ▶ Rappel : il y a 278 champs « principaux » x 8 CCD x 2 caméras x 4 quarts, soit en principe 17 792 qCCD
- ▶ 14 012 images « c » 84.8 GB
- ▶ 12 652 images « w » 76.5 GB
- ▶ 19 866 images « x » 120.0 GB

Une « image composée » est une image construite à partir d'une petite série de bonnes images afin d'obtenir une image de très bonne qualité.

Les images composées servent essentiellement à l'identification des étoiles afin de créer les catalogues, ou « fichiers de références ». Le terme « image de référence » parfois utilisé est ambiguë et sert à désigner soit l'image composée principale utilisée pour la détection des étoiles, soit pour les différentes images intervenant dans cette détection.

Les images composées sont produites par couleur et par quart de CCD. Ces images sont au format FITS. Elles représentent 1600x1600 pixels 16 bits, soit des fichiers de 6.2 Mo.

Au moins deux techniques différentes ont été utilisées pour la composition, matérialisées dans le nom des images produites par un code de « traitement » particulier : « c » et « x ». Les images « x » ont ensuite été callées astronomiquement par le groupe d'astrophysiciens conduit par Jean-Baptiste Marquette ce qui a conduit à des images « w ».

Toutes ces images sont disponibles au Centre de calcul et sont référencées dans la base de données.

# Catalogues d'étoiles

16

- ▶ Les catalogues d'étoiles regroupent les références des étoiles détectées
  - ▶ Ces fichiers nommés « fichiers de référence »
  - ▶ Format binaire propre à l'expérience
  - ▶ Un fichier de références par couleur et quart de CCD
- ▶ Les étoiles détectées sont associées dans les deux couleurs pour permettre un traitement indépendant des flux bleus et des flux rouges
- ▶ Un jeu de fichiers de références par « production »
  - ▶ 13 376 références P 5 pour les 7 programmes majeurs
  - ▶ 13 344 références P 5 sauvées dans Irods – 66.9 GB

Les catalogues d'étoiles, ou « fichiers de références », contiennent les informations permettant d'identifier la position des étoiles suivies sur les images à analyser. Ils sont construits à partir d'une « image composée » et d'une ou plusieurs bonnes images utilisées en complément.

Ils sont produits pour chacune des deux couleurs et par quart de CCD.

Alors que les « images de référence » sont conservées d'une production à l'autre, les catalogues sont propres à chaque production.

La production était réalisée en batch, sur la ferme de calcul du Centre, les catalogues sont donc disponibles à Lyon. Mais le choix fait par l'expérience a été de ne recopier dans Irods que les références de la production P5.

# Suivis

17

- ▶ Les fichiers de suivis regroupent les mesures réalisées sur les étoiles des différentes images
  - ▶ Format binaire propre à l'expérience
  - ▶ Analyse réalisées par quart de CCD et par couleur, par lot, en batch, au fur et à mesure des arrivées
- ▶ Division en plusieurs blocs pour réduire les mouvements avec les 3490
  - ▶ Chaque bloc correspond à un fichier SV identifié par son numéro de blocs
- ▶ 72 234 fichiers de suivis pour les 7 Majeurs P5 dans Irods
  - ▶ Soit 6.3 TB
- ▶ 13 343 blocs #1 – donc de suivis distincts – soit 75 % des champs « majeurs »

Les mesures réalisées sur les images sont conservées dans des fichiers binaires, dans un format propre à l'expérience, nommés « suivis ».

Ces suivis sont construits par couleur et par quart de CCD. Mais afin de minimiser les mouvements de fichiers, et pour permettre des traitements au fur et à mesure de l'arrivée des images, les mesures sont organisées en bloc. A chaque bloc correspond un « fichier de suivi ». Un « suivi » est donc une suite de blocs, ou « fichiers de suivi ». Ainsi, les mesures d'une même zone sont réparties dans plusieurs blocs de suivi, en deux couleurs.

La production des suivis était réalisée en batch, sur la ferme de calcul du Centre. Les fichiers sont conservés à Lyon. Mais seuls la production P5 a été recopiée dans Irods.

# Réorganisation des résultats

18

- ▶ L'organisation en couleurs, quarts de CCD et blocs de suivis est efficace pour des productions massives
- ▶ Pas très pratique pour des analyses fines sur de petites sélections d'étoiles, surtout si elles sont éloignées
- ▶ La réorganisation consiste à regrouper les mesures de l'ensemble d'une même étoile pour les deux couleurs dans un seul fichier
  - ▶ Ces fichiers sont nommés « courbes de lumière »

L'organisation des données en couleurs, quarts de CCD et blocs de suivi est bien adaptée au traitement massif des données, pouvant être réalisé au fil des observations. Mais elle est assez lourde lorsqu'il s'agit d'analyses fines portant sur quelques étoiles intéressantes.

Une autre difficulté liée aux « références » et aux « suivis » est que leur accès nécessitait l'utilisation d'une librairie d'entrées-sorties dédiée, écrite en C++.

Pour contourner ces contraintes, une sélection des mesures réalisées dans les deux couleurs pour l'ensemble de la campagne est extraite des fichiers de suivi afin de constituer les « courbes de lumière », sous la forme de fichiers ASCII.

# Les courbes de lumière

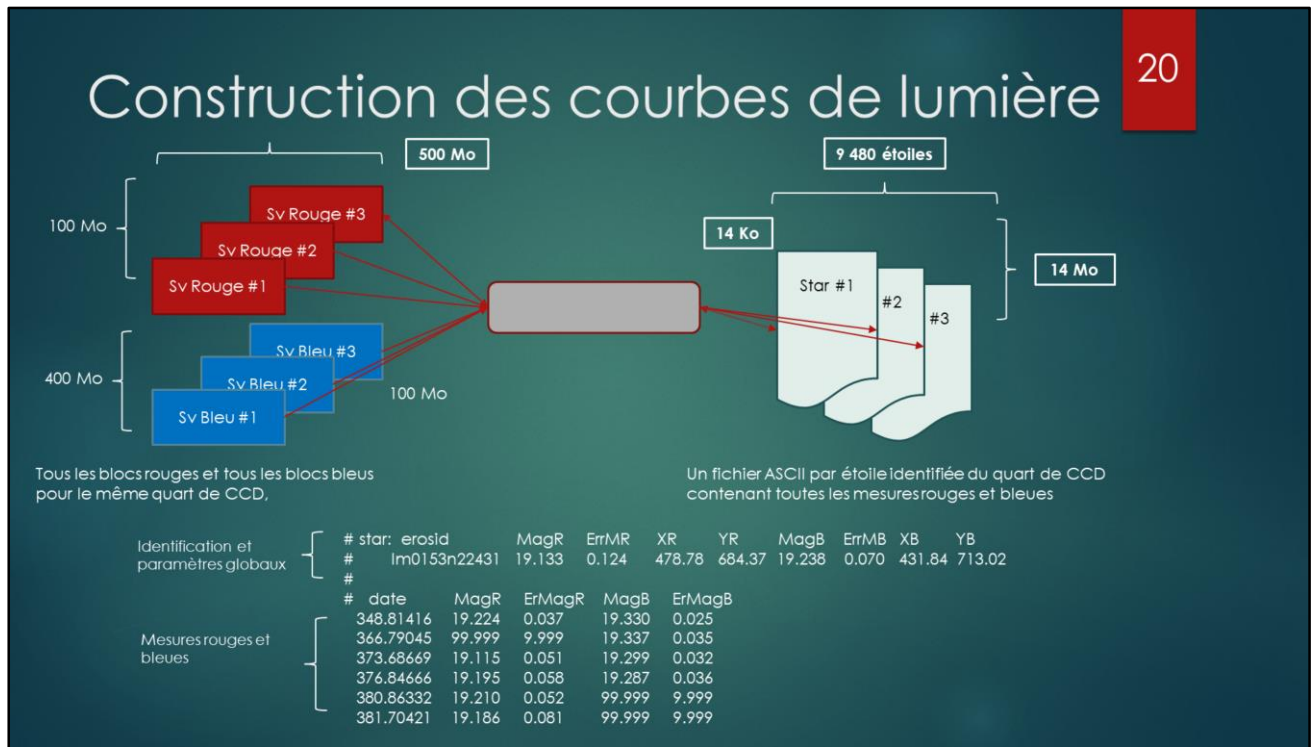
19

- ▶ Regroupe les 7 années de campagne dans les deux couleurs dans un seul fichier par étoile suivie
- ▶ Format ASCII ne nécessitant pas de librairie de lecture dédiée
  - ▶ Accès universel
- ▶ Ensemble réduit de données : date de mesure, mesures bleues et rouges, erreurs sur les mesures.
- ▶ Fichier de petites tailles : 15-20 KB
- ▶ Environ 1.2 TB non comprimés, soit 450 GB comprimés.

Une courbe de lumière contient l'ensemble des mesures rouges et bleues d'une même étoile pour toute la durée de la campagne. La taille du fichier est de quelques Ko et peut être significativement réduite par simple compression GZip.

Les courbes de lumière ASCII présentent plusieurs intérêts : l'ensemble des données concernant une même étoile est disponible en une seule fois, alors qu'il est nécessaire d'accéder à plusieurs blocs de suivis pour reconstituer l'ensemble des mesures ; les fichiers sont de petites tailles, au contraire des suivis plus massifs ; et aucune librairie n'est nécessaire pour accéder aux données, ce qui les rend « universelles ». Par ailleurs, l'utilisation d'une représentation textuelle n'entraîne pas de perte de précision significative dès lors que l'enregistrement est fait avec un nombre suffisant de chiffres.

Plusieurs campagnes de production ont été réalisées, plus ou moins abouties. Les deux plus importantes, en termes de suivis générés, sont la production P1 et la production P5. La production considérée comme la production officielle est la production P5. C'est des fichiers de suivi de cette production que sont issues les courbes de lumière ASCII.



Le diagramme schématise le processus de construction des courbes de lumière consistant à récupérer l'ensemble des blocs rouges et bleus des suivis d'un même zone – correspondant à un quart de CCD – afin de produire les fichiers ASCII de chacune des étoiles analysées.

Une étoile est identifiée par le nom du quart de CCD où elle est présente, en faisant exception des codes de caméra, filtre et couleur, associé à un identifiant numérique. Ce nom est utilisé pour nommer le fichier courbe de lumière.

Les noms des différentes étoiles d'un quart de CCD, sa position dans le ciel, et différents paramètres globaux, sont regroupés dans un fichier ASCII nommé « catalogue ».

Les positions des différentes zones du ciel correspondant aux quarts de CCD d'un champ, indépendamment de la couleur, sont regroupées dans des fichiers « fields », eux-mêmes ASCII. Un fichier général décrit la position des différents champs d'un programme scientifique.

Cet ensemble de catalogues permet de proche en proche de retrouver les étoiles d'une région donnée, et donc les courbes de lumière correspondant.

Les différents catalogues sont conservés dans Irods mais sont aussi enregistrés dans la base de données sous la forme d'objets binaires comprimés (ou BLOB, pour *Binary Large Object*). Ceci permet une recherche rapide des étoiles et donc des courbes de lumière sans qu'il ne soit nécessaire d'accéder à Irods.

## 90 millions d'étoiles

21

Programme	Etoiles bleues	Rouges	Moyennes	Courbes	%
Bs	6 699 360	6 585 525	6 642 443	2 416 924	36%
Cg	92 001 893	92 642 126	92 322 010	43 681 008	47%
Gn	12 871 686	13 018 914	12 945 300	4 307 755	33%
Gs	5 589 593	5 428 873	5 509 233	1 804 578	33%
Lm	68 696 035	70 057 582	69 376 809	28 797 428	42%
Sm	8 708 758	8 887 279	8 798 019	4 065 298	46%
Tm	6 355 664	6 143 296	6 249 480	2 129 271	34%
total	200 922 989	202 763 595	201 843 292	87 202 262	43%

Les nombres d'étoiles bleues et rouges proviennent du programme d'analyse et sont enregistrés dans la base de données au niveau de la description des références.

En comparant le nombre d'étoiles associées enregistré dans la base de données pour les références de la production P5 au nombre de courbes de lumière, et donc d'étoiles traitées, on constate un déficit de 50 à 70 %, ce qui correspond à la description des filtrages réalisés comme l'indique la thèse de Patrick Tisserand.

On voit aussi que le nombre d'étoiles associées pour les catalogues bleus n'est pas le même que pour les catalogues rouges, ce qui semble plus surprenant.

## A suivre : perte de données Eros II

Jean-Noël Albert – LAL / IJClab

- ▶ Le blog Eros / Anastasis : <https://groups.lal.in2p3.fr/erosanastasis/>
- ▶ Le site ErosDb II : <http://eros.lal.in2p3.fr/ErosDB/>
- ▶ Le site historique Eros : <http://eros.in2p3.fr/>