

Les données EROS-2, un cas d'étude pour l'archivage et la fouille de données

Jean-Baptiste Marquette, IAP
Journées MAESTRO 2016
Toulouse, 5-6 juillet

EROS-2, résumé des épisodes précédents

- Recherche de la matière noire baryonique sous forme d'objets compacts (MACHOs) : **E**xpérience pour la **R**echerche d'**O**bjets **S**ombres, version **2**
- Télescope dédié (MARLY, 1m, ESO Cerro La Silla, Chili)
- Prise de données entre juillet 1996 & mars 2003
- Bandes photométriques larges non-standard ("bleu" $\sim V + R$, "rouge" $\sim I$)
- $\sim 1,4$ million d'images accumulées, sur 208°^2
- De l'ordre de 100 millions de courbes de lumière, par rapport à une position fixe sur une image de référence
- Mais, en ces temps héroïques, pas d'archivage prévu !

« *Les données, on les traite et on les jette.* »

Entendu dans la bouche d'un membre fondateur du projet EROS

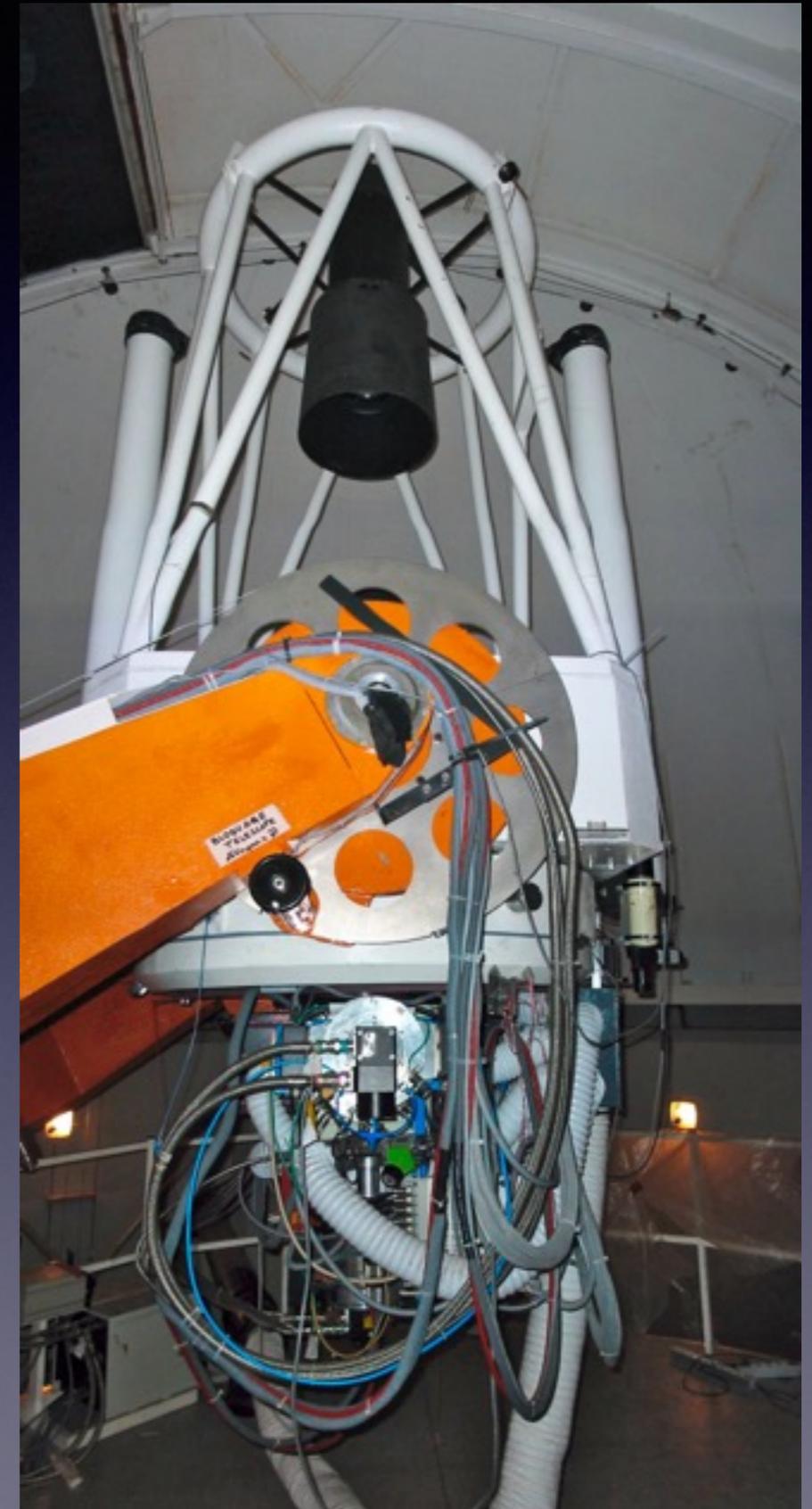
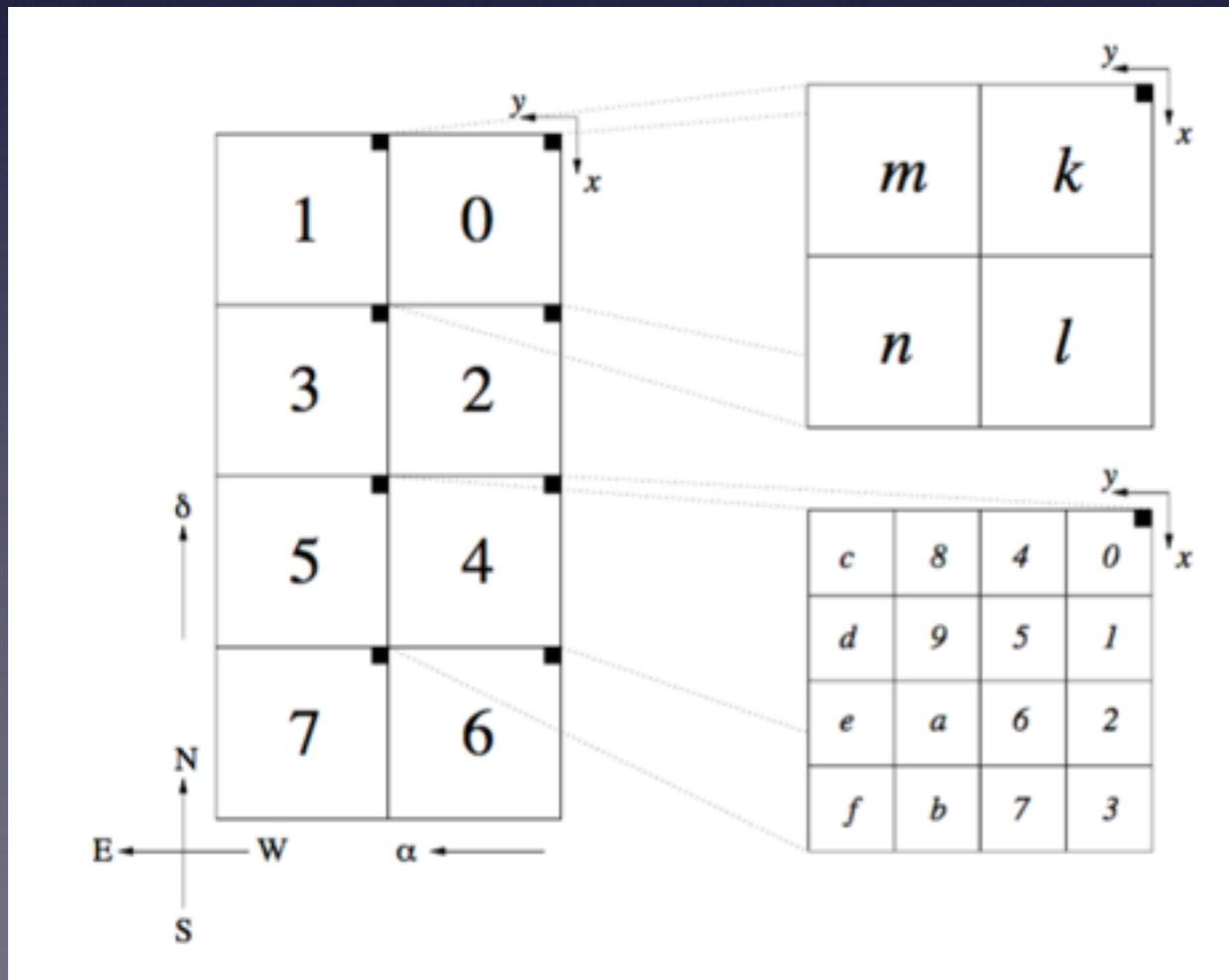
Et pourtant...

La base de données EROS-2, une mine à valoriser

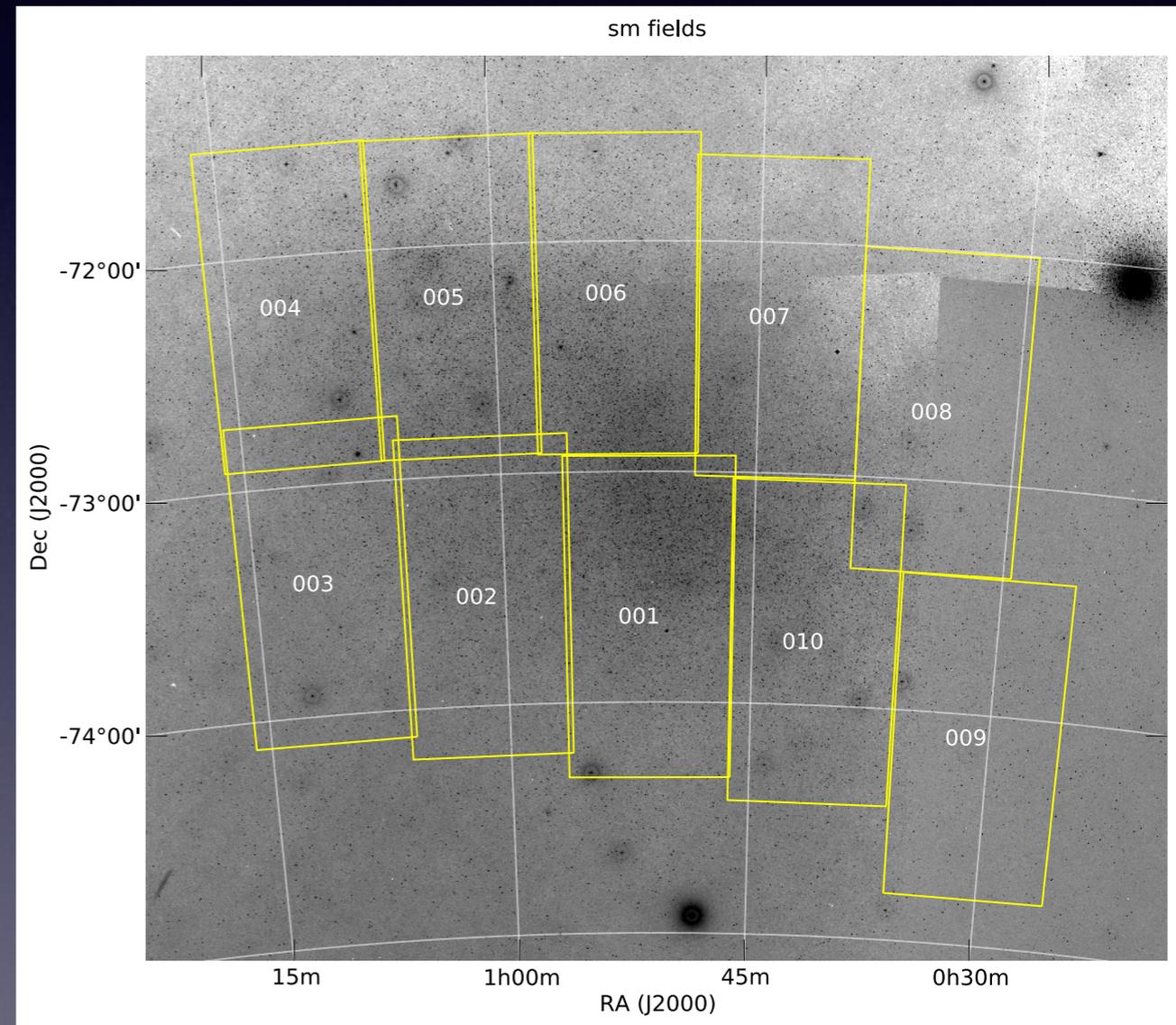
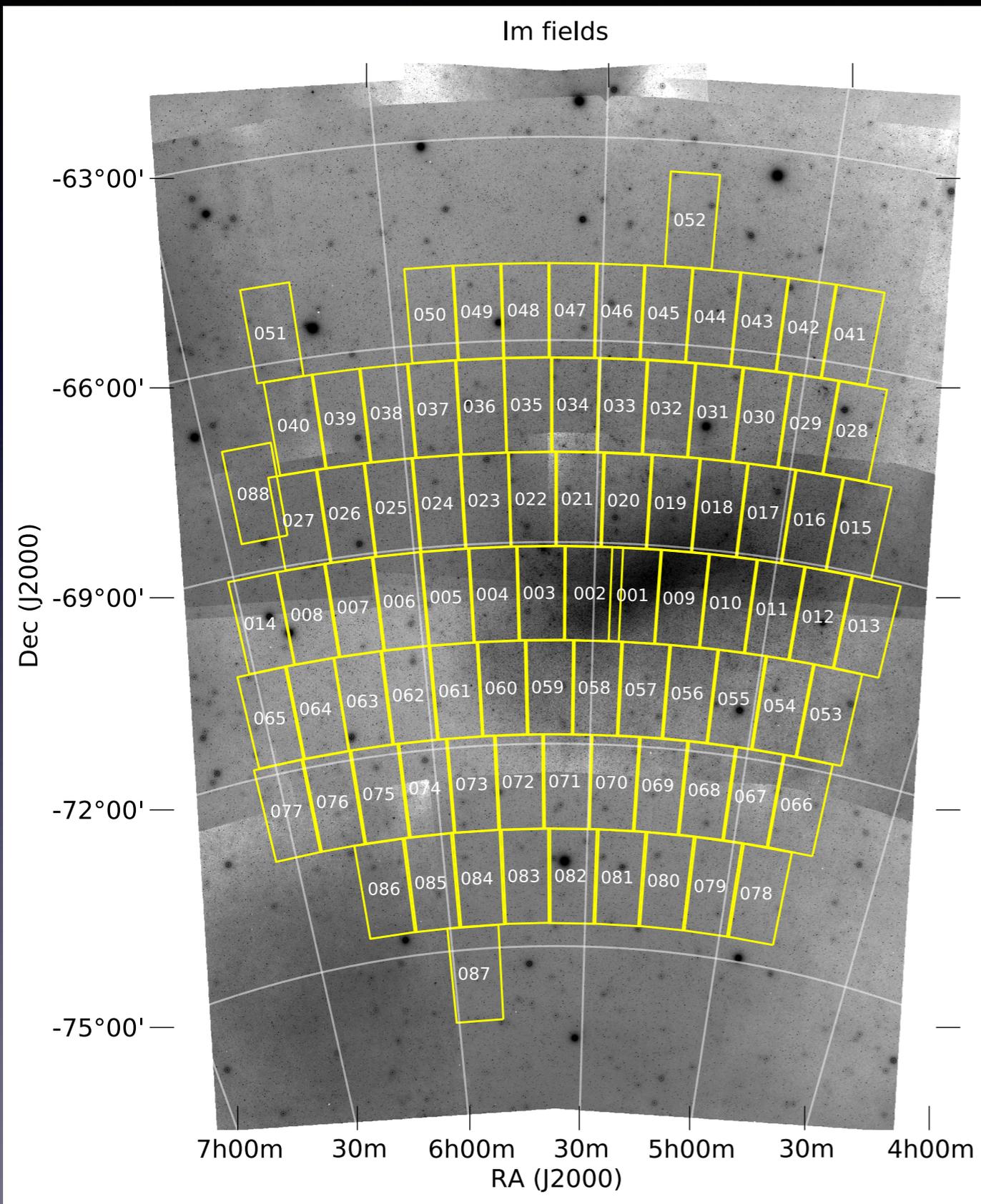
- Variabilité stellaire
- Études de diverses familles d'étoiles variables
- Phénomènes transitoires
- Mouvements propres
- Méthodes statistiques, *machine learning*

L'instrumentation

- type Ritchey-Chrétien
- miroir primaire $\phi=0.98$ m de diamètre. Focale $F = 5.14$ m, ouverture = $F/5$
- 2 caméras, mosaïques de 8 CCDs de 2048×2048 px
- champ couvert de $1.39^\circ \times 0.69^\circ \sim 0.96^\circ{}^2$
- observations **en simultané** sur les deux caméras



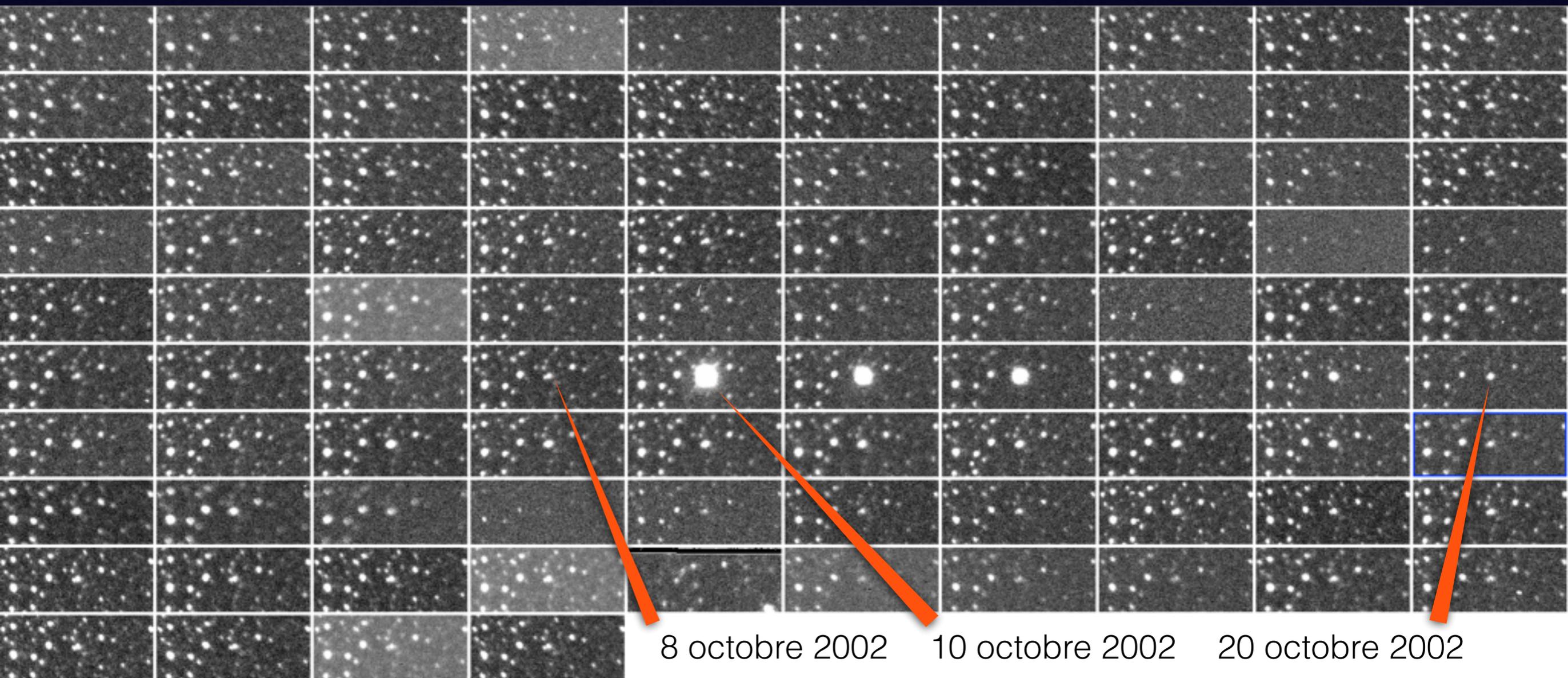
Principaux champs observés



Une archive pour le futur... et le présent !

Requêtes régulièrement reçues pour extraire les données collectées sur des objets spécifiques.
Par exemple :

LMCN-2016-01a (OGLE) - en 2002, 94 images EROS-2, caméra « bleue »
Nova récurrente dans le Grand Nuage de Magellan

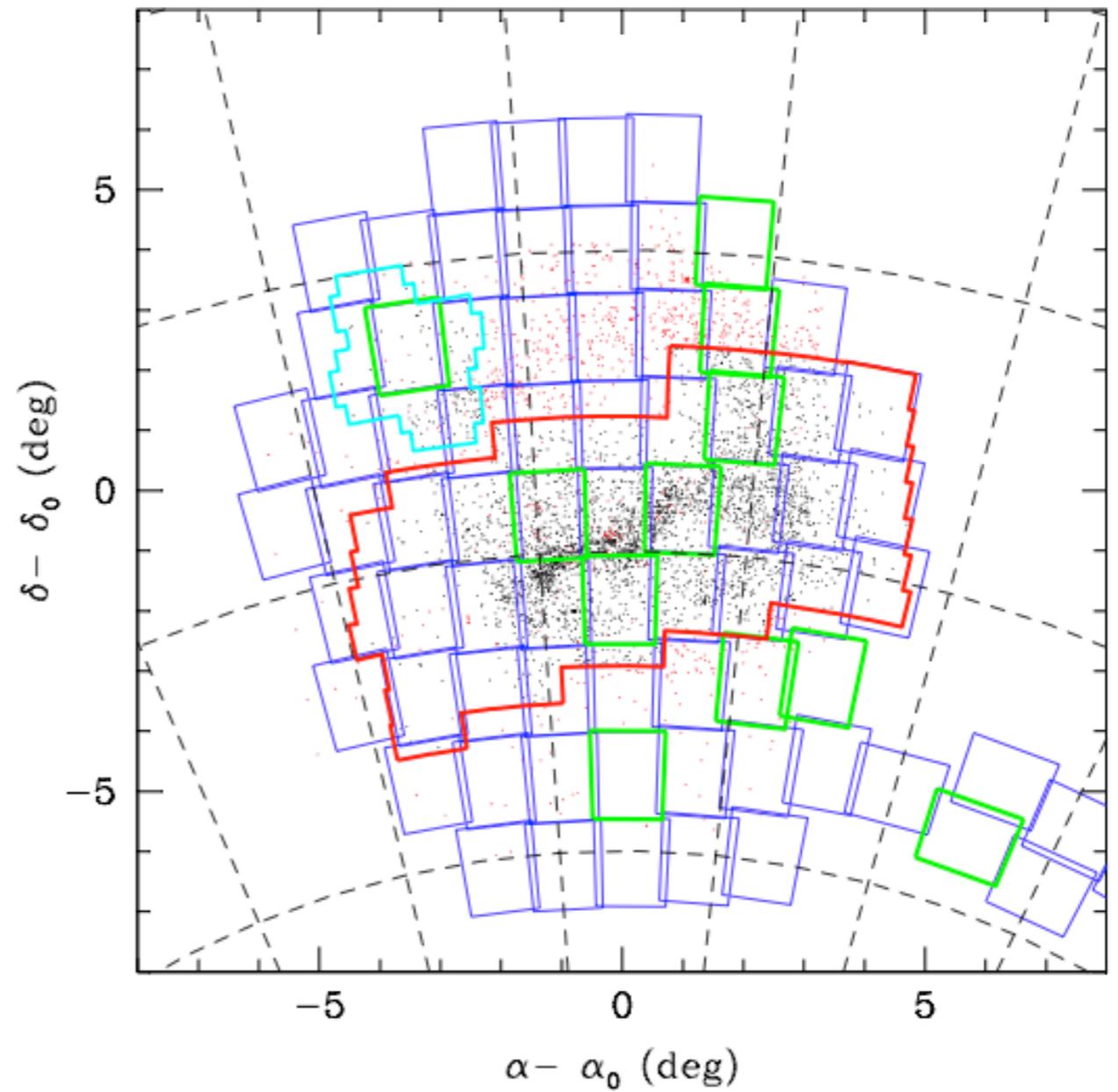
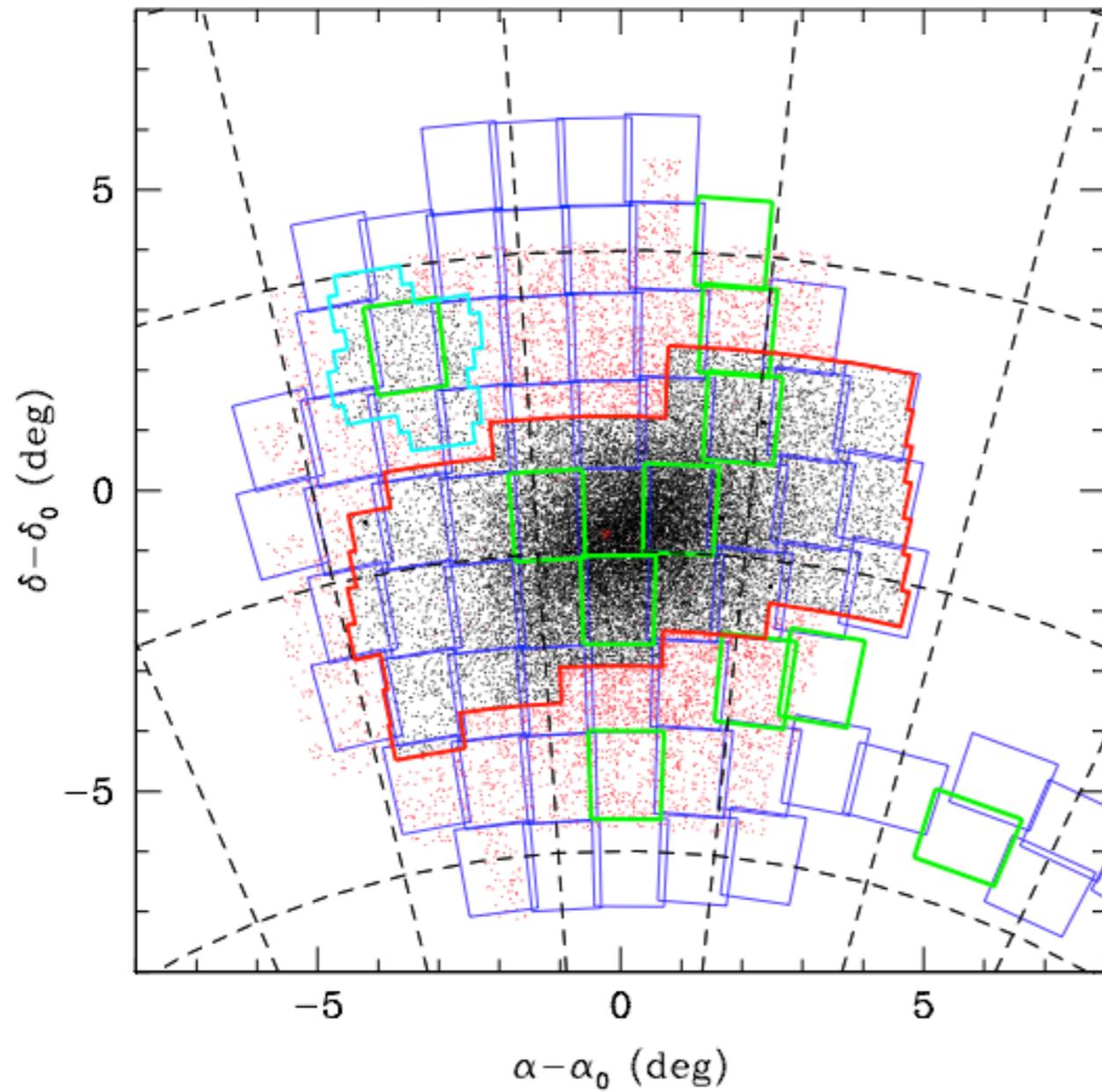


Fouille d'archives & *machine learning*:
quelques exemples récents

EROS-2 & VMC

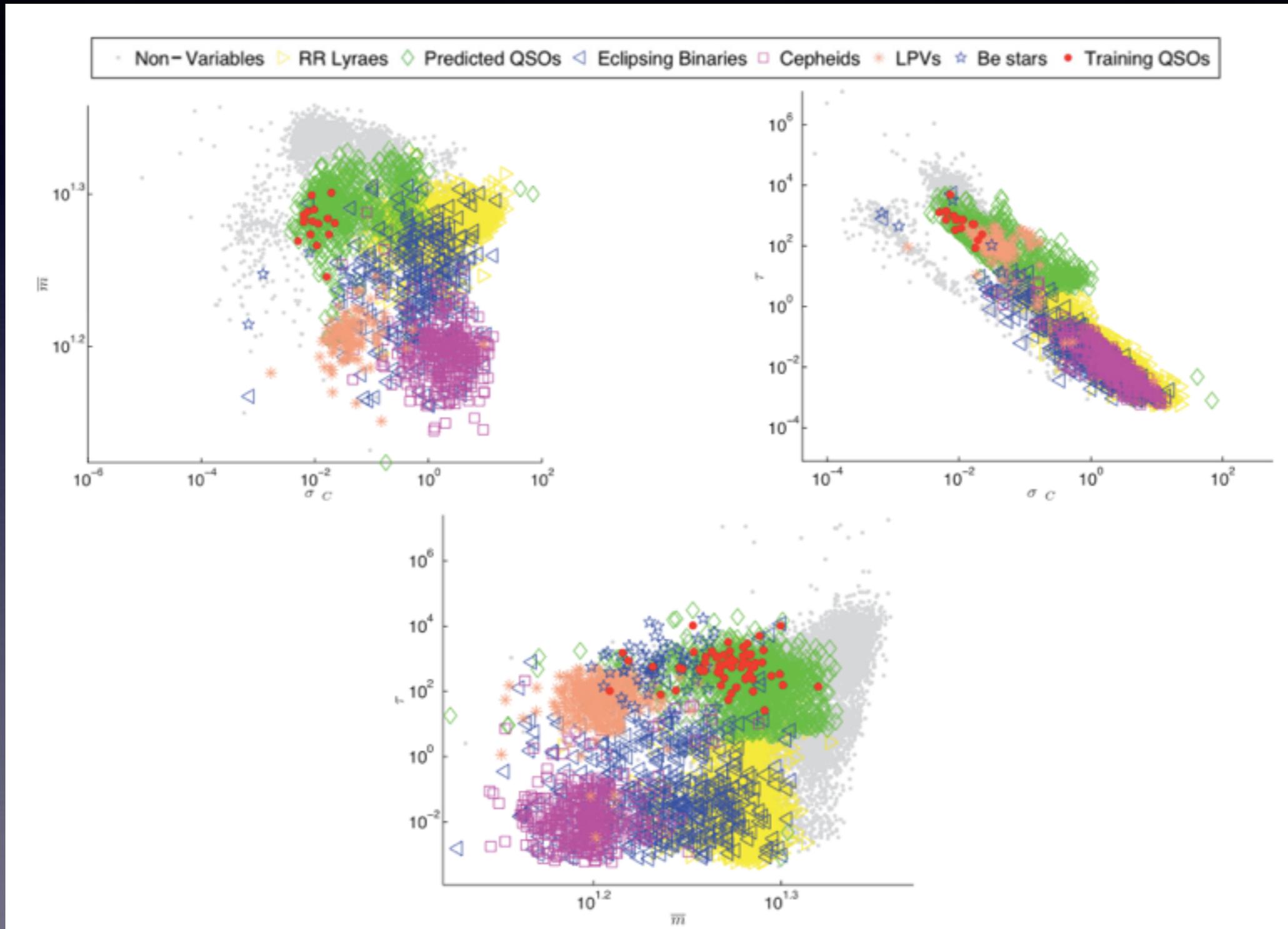
Moretti et al. 2014, *MNRAS* **437**, 2702

RR Lyræ (g.) & Céphéides (dr.) comme sondes de la structure
du Grand Nuage de Magellan



Des quasars derrière le Grand Nuage de Magellan

Pichara et al. 2012, *MNRAS* **427**, 1284 : 1160 candidats quasars, algorithmes de forêt aléatoire appliqués à 28 millions de sources

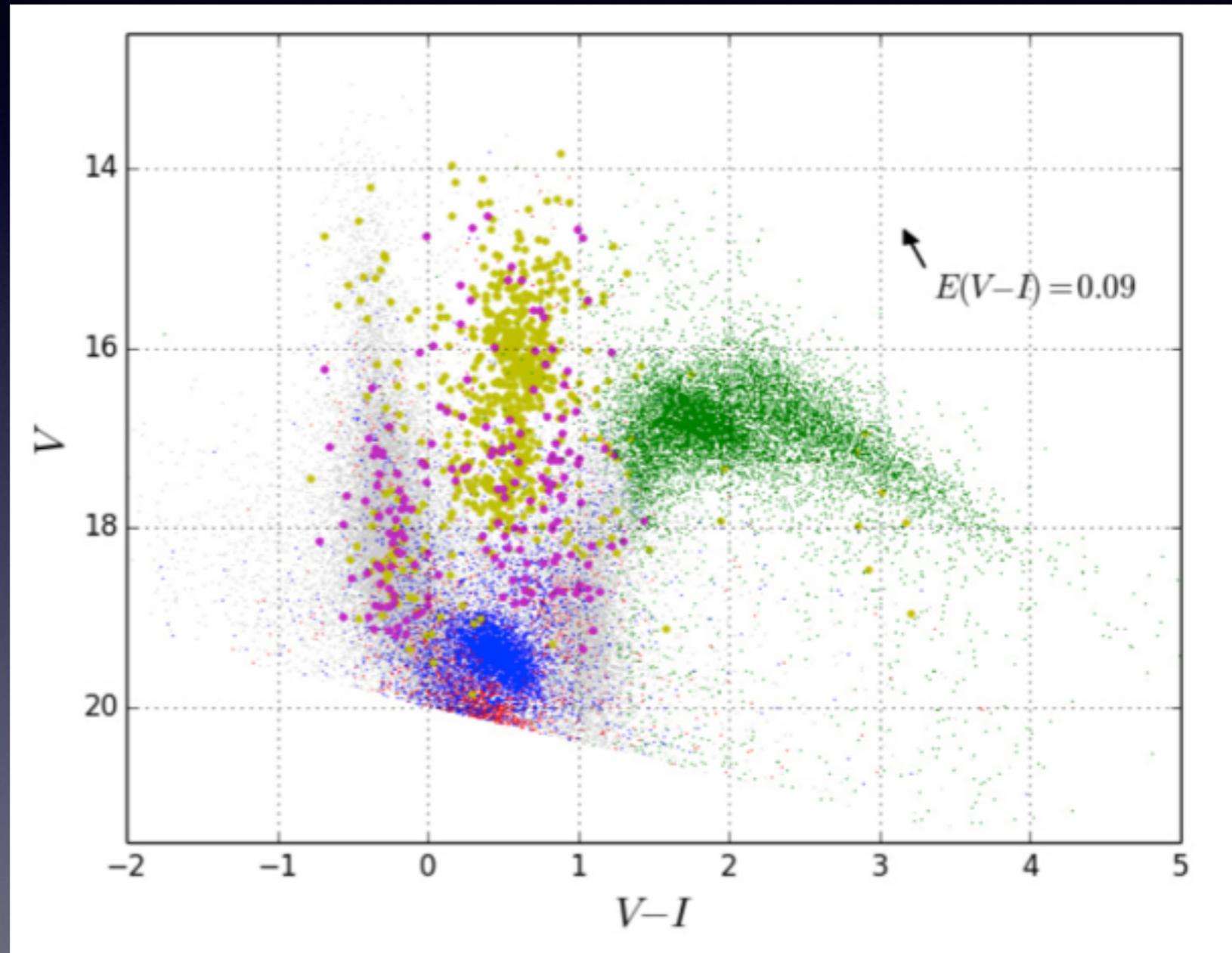


Classification automatique de courbes de lumière

Kim et al. 2014, *A&A* **566**, A43

55285 nouvelles variables dans le Grand Nuage de Magellan

- vert : variables à longue période
- bleu : RR Lyræ
- rouge : δ Scuti
- gris : binaires à éclipses
- jaune : Céphéïdes
- magenta : Céphéïdes type II

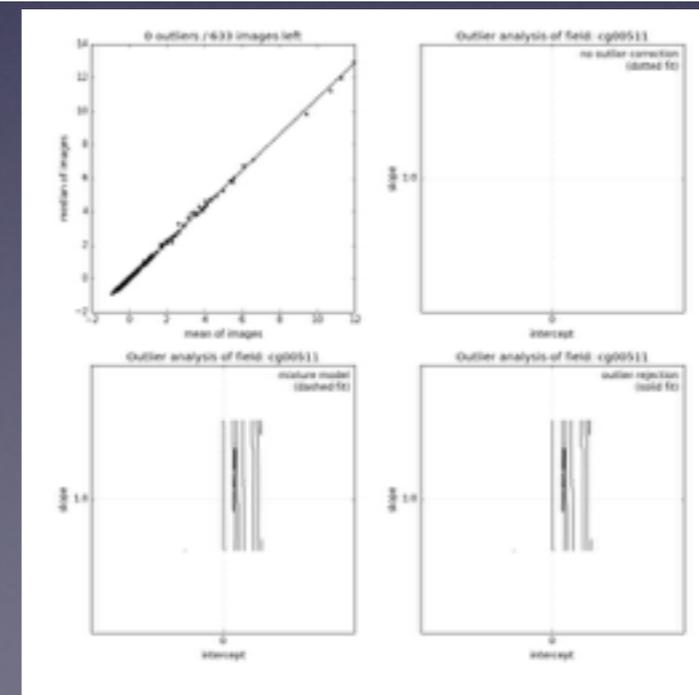
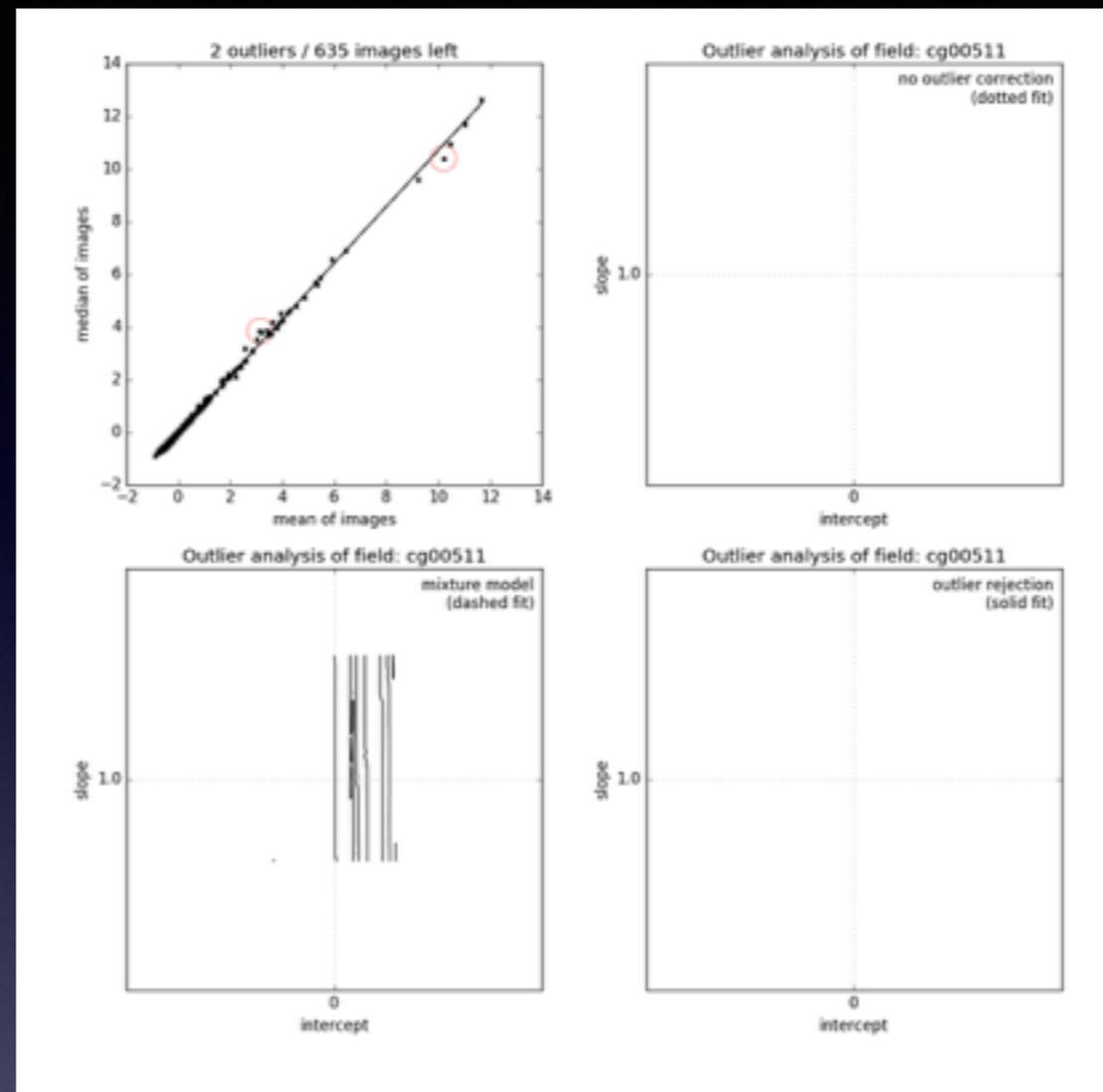
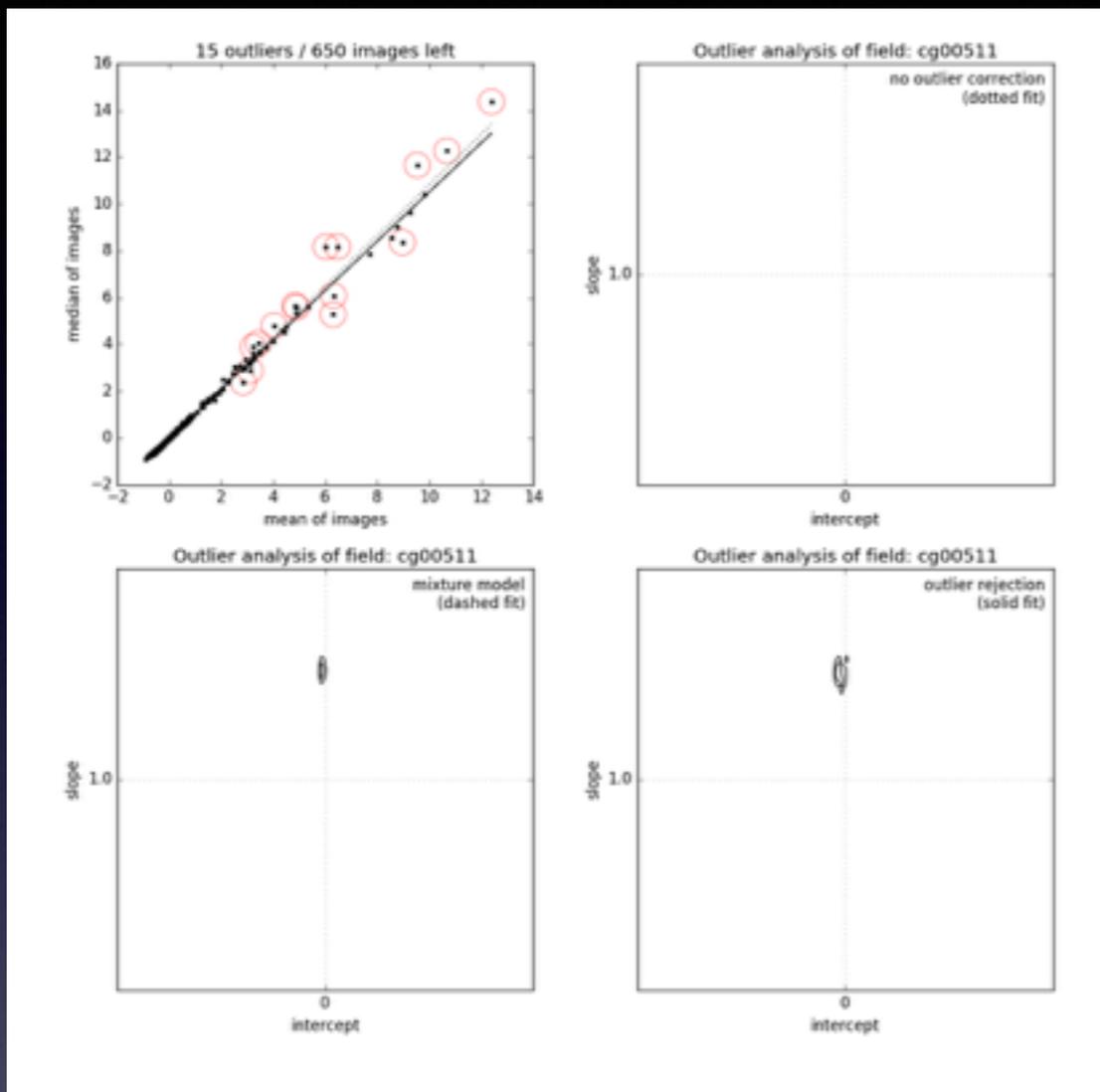


À terme, une archive publique

- Serveur d'images accessible en ligne
- Bases de données associées, plusieurs niveaux possibles :
 - des images, avec critères de position et de qualité
 - des sources, avec données de position et de photométrie
 - des courbes de lumière, par photométrie différentielle
 - des objets variables, avec mesure des périodes associées

Mais...

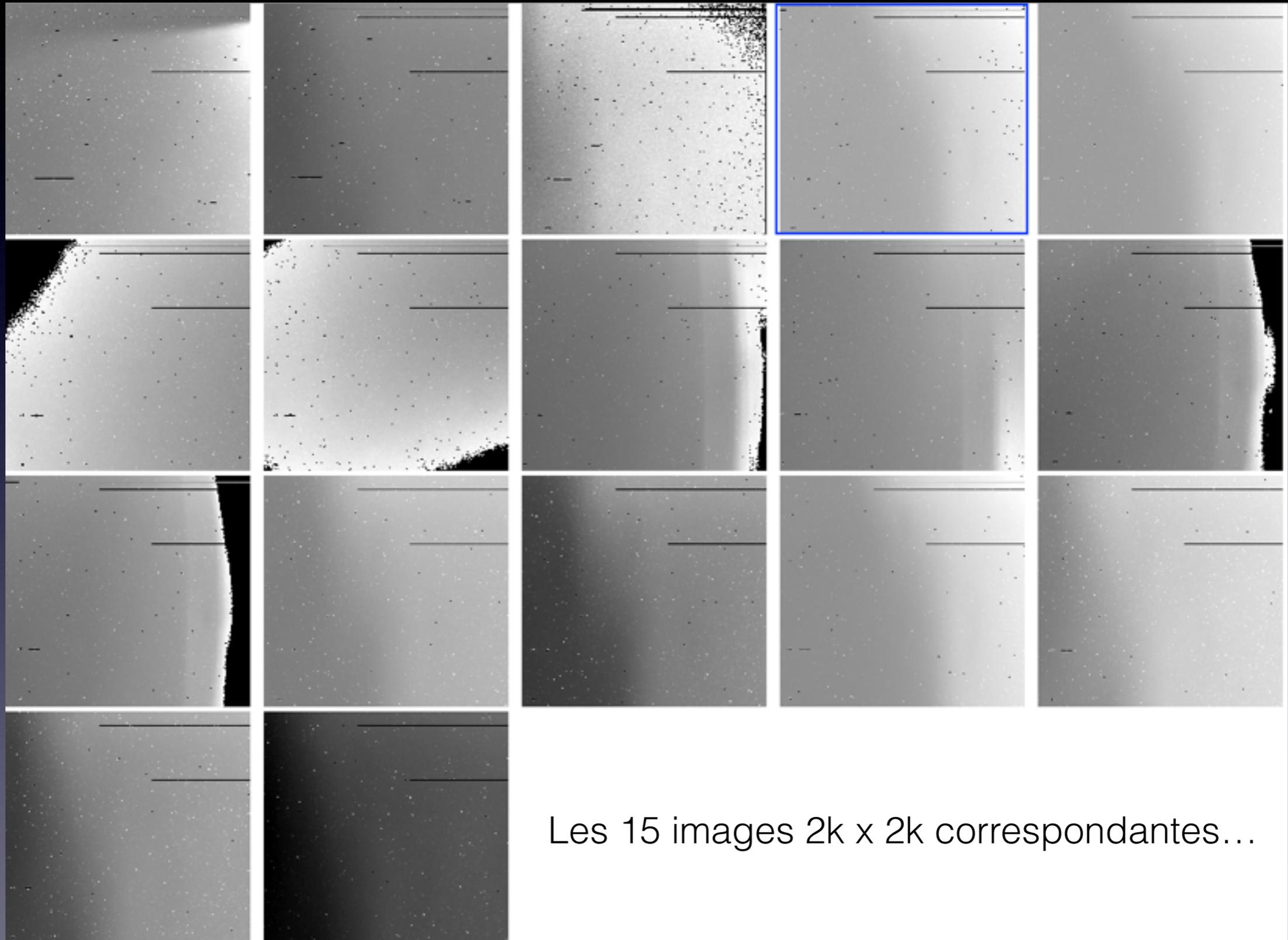
Contrôle et qualification du contenu indispensables 1/2



Méthode reprise de astroML :
rejet des *outliers* par MCMC

http://www.astroml.org/book_figures/chapter8/fig_outlier_rejection.html

Contrôle et qualification du contenu indispensables 2/2

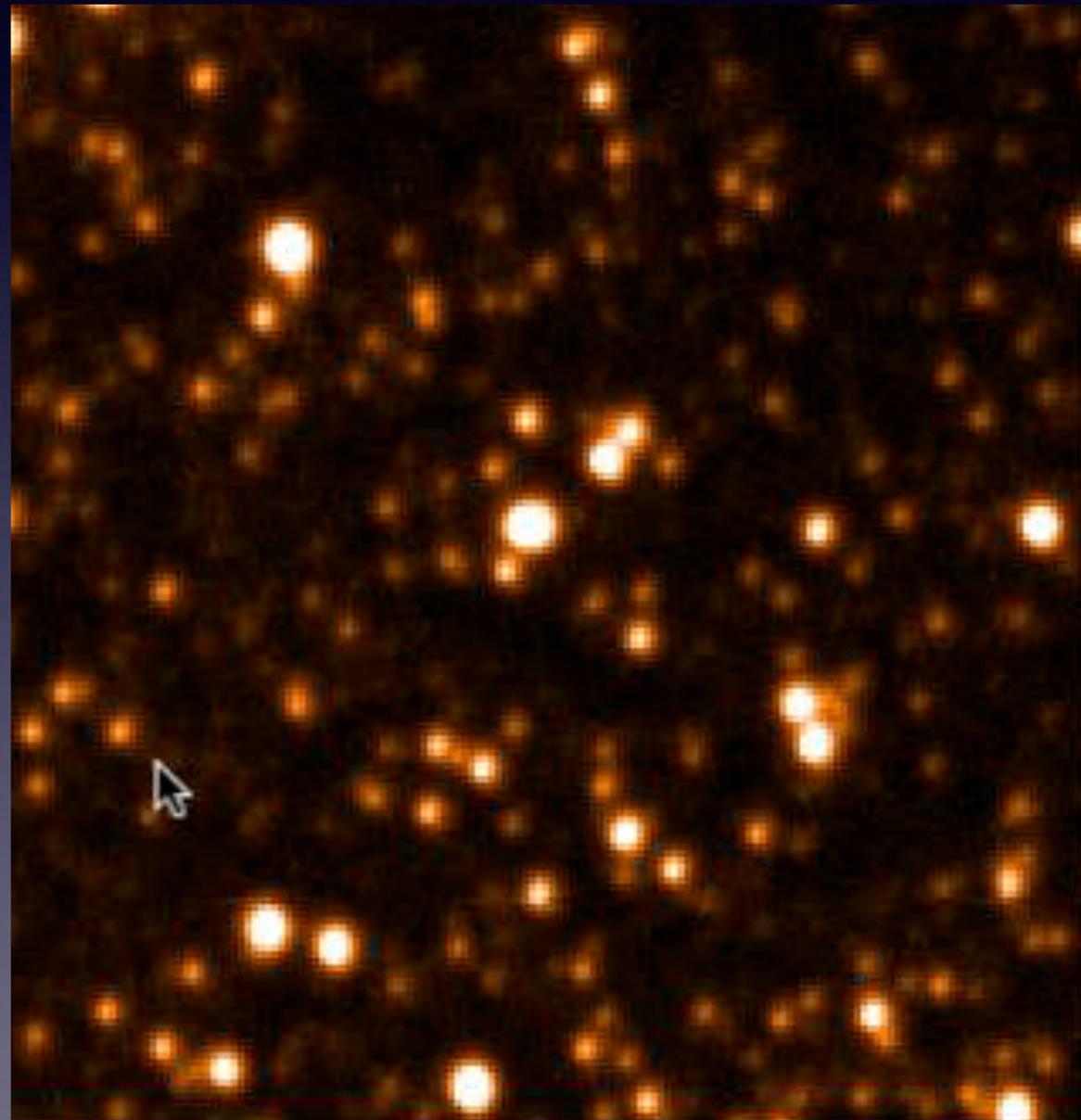


Encore du contenu scientifique
à exploiter !

Deux exemples pour faire
"parler" les données :

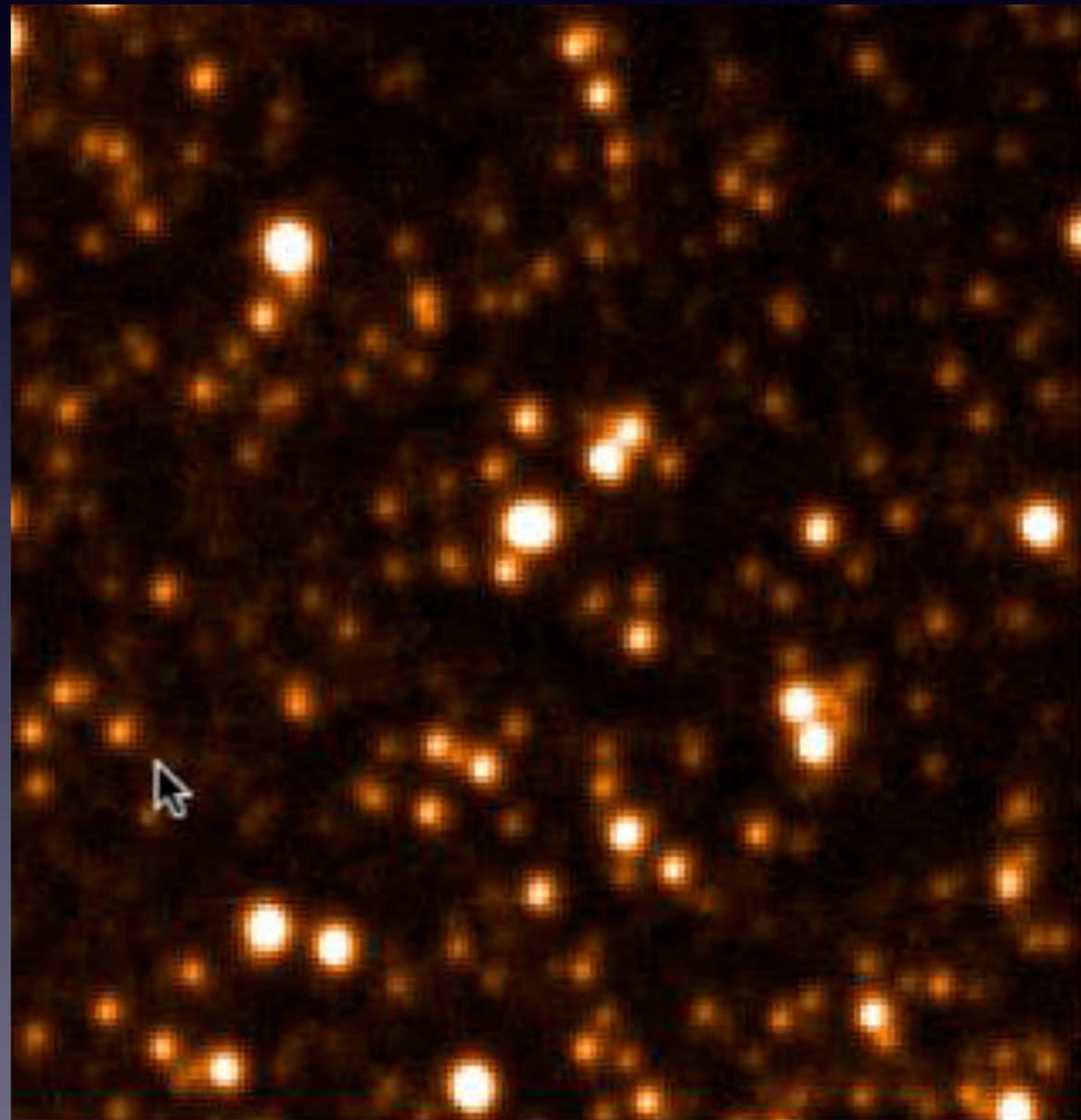
Corps du Système solaire !

Quelque part vers les Bras spiraux...
Clignotement entre deux observations EROS-2



Corps du Système solaire !

Quelque part vers les Bras spiraux...
Clignotement entre deux observations EROS-2



Photométrie différentielle

Quelque part vers le Centre galactique



- Champ très encombré
- Les sources constantes disparaissent dans le fond moyenné
- Vignette centrée sur un objet particulier
- Méthode adaptée aux faibles variations de luminosité

Photométrie différentielle

Quelque part vers le Centre galactique



- Champ très encombré
- Les sources constantes disparaissent dans le fond moyenné
- Vignette centrée sur un objet particulier
- Méthode adaptée aux faibles variations de luminosité

En conclusion

- Qualifier le contenu de la base de données EROS-2, en vue d'une mise à disposition publique
- Pour cela, tester les méthodes statistiques utilisées pour réduire (sinon supprimer) les contrôles visuels (1,4 million d'images...)
- Documenter et vérifier le contenu des en-têtes FITS
- Tester les algorithmes de photométrie différentielle
- Proposer une base de données en ligne, pour
- *learning*

Merci pour votre attention