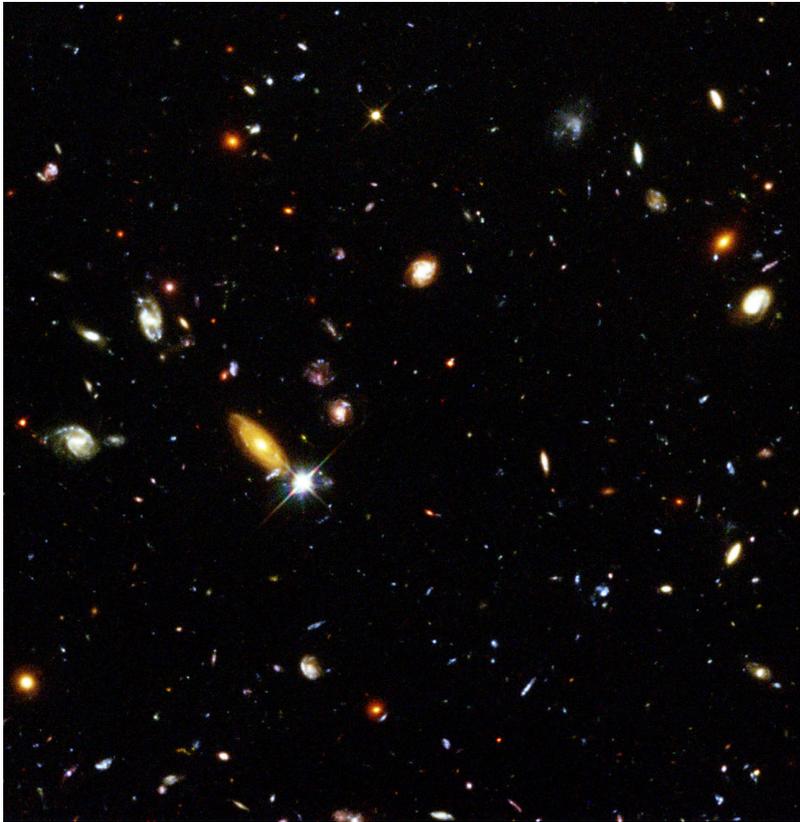


# Le « stack » LSST

## Etat des lieux et perspectives



LSST-France  
à Montpellier.  
(2015)

*Pierre Astier*

*LPNHE / IN2P3 / CNRS, Universités Paris 6&7.*

# Le « stack » : qu'es aco ?

- C'est une collection d'outils logiciels
- Packages, gestion de configuration (EUPS) gestion de versions (git), environnement de build (scons), glue python (swig), interpréteur python,...
- En pratique : python (avec C++ en arrière plan).
- Documentation : doxygen, mais pas beaucoup plus.
- Code open source. Peu diffusé hors LSST.

# Le stack, pourquoi faire

- Traiter des images d'astronomie, pas seulement de LSST.
- Doit, in fine, contenir l'ensemble des algorithmes qui conduisent des pixels aux données utilisables pour la science.
- ... ainsi que (probablement) les outils d'accès aux catalogues.
- C'est l'ensemble de ce qui est nécessaire à la réduction des données LSST pour mise à disposition de la communauté (à la SDSS).

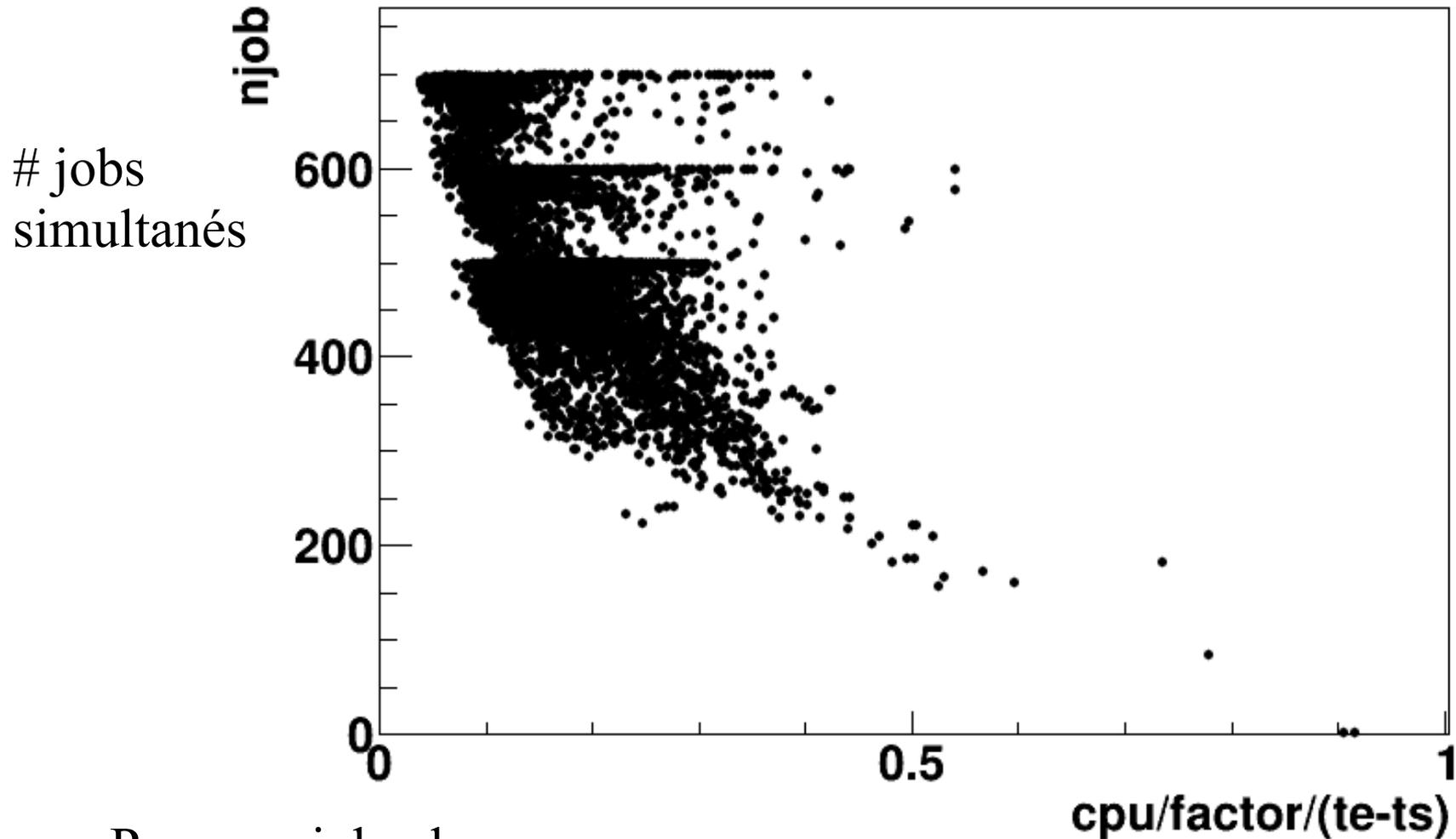
# Test #1 : DC 2013, la bande 82

- Bande 82 du Sloan (relevé SN) environ  $\sim 100$  visites/ bande, 5 bandes,  $300 \text{ deg}^2$
- Images brutes :  $\sim 7$  TB. (brutes après flatfielding)
- But technique :
  - 1) Re-échantillonnage et stacking
  - 2) Photométrie sur les stacks et sur les visites individuelles
  - 3) Mise à disposition des produits dans une base de données
  - 4) Science ?

# Test #1 : DC 2013, la bande 82 : enseignements

- C'est pas si facile à faire tourner.
- On produit des choses mais on ne sait pas trop quoi.
- Le volume CPU est important.
- Le ré-échantillonnage/stacking est significativement plus lent que swarp.
- Au CC, trop de jobs simultanés (>50) saturent les I/O.
- Mysql s'étouffe sous la charge (→ on attend Qserv)

# Le code d'empilement de LSST en action au CC



Pour ces jobs, les serveurs  
sps marquent le pas .....

Efficacité CPU

## Test #2 : CFHTLS

- Maître d'oeuvre : D. Boutigny. Il utilise sa présence à SLAC pour obtenir les infos clé.
- Le CFHTLS est beaucoup plus proche de ce que produira LSST que le SDSS.
- Il existe des processings de référence : Terapix, SNLS, CFHTLens, ..., adossés à des publications.

# Test #2 : CFHTLS

- Entrée : images (après flatfielding.)
- Etapes (minimum):
  - Catalogue, astrométrie (vs catalogue externe), IQ.
  - Re échantillonnage/stacking
  - Photométrie (étoiles et galaxies)
- Et en plus :
  - Détection des transients
  - Photométrie différentielle
  - Uniformisation photométrique du plan focal
  - Morphométrie
  - ....

# Test #2 : CFHTLS

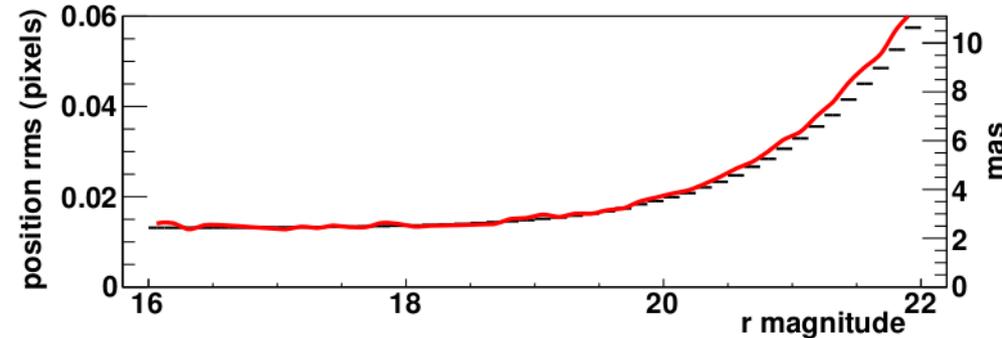
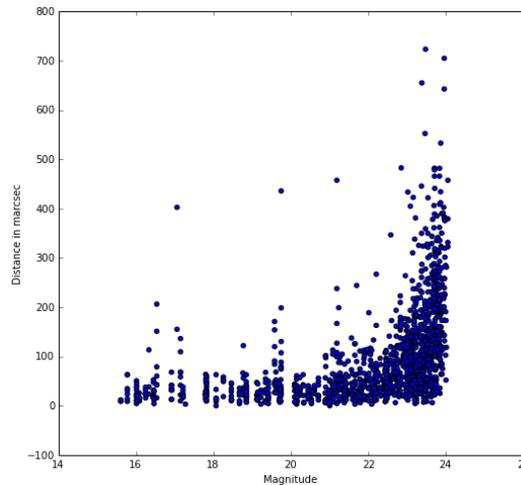
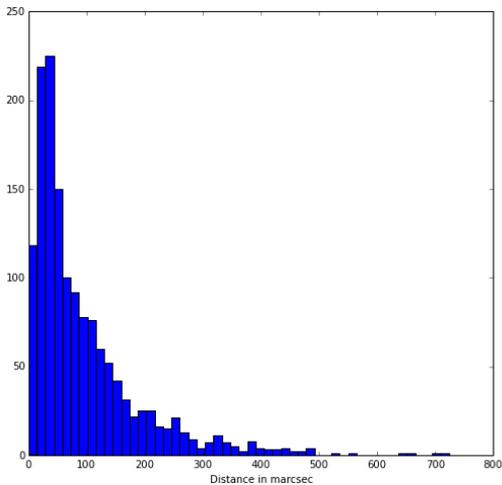
- Changement d'attitude : on essaie de comprendre ce que l'on fait.
- Echantillon des difficultés (pratiques) :
  - Format des masques : innovant (!).
  - Solution astrométrique initiale délicate : il faut nourrir le code avec les distorsions optiques.
  - La compression FITS sabote le résultat (?!)
- Des tas de problèmes identifiés, et à résoudre.
- Evaluation de la photométrie : en cours

# Test #2 : CFHTLS

- Parmi les problèmes/manques identifiés :
  - L'ajustement du WCS est branlant. Les causes sont comprises.
  - Il n'y a pas d'astrométrie simultanée à un ensemble d'images. C'est nécessaire...
  - Il faut « uniformiser le plan focal ». Le ciel au crépuscule n'est **pas** la réponse (voir hier).
  - Il faut comprendre la photométrie et ses erreurs.
  - La chaîne définit  $\text{Var}(\text{Image}) = \text{Image}/\text{gain}$ .  
(Très) mauvaise idée.

# Plan : combler certains manques

## 1 : astrométrie simultanée, pourquoi



Résidus astrométriques entre positions mesurées projetées sur le ciel.

Ajustement indépendant de chaque image.

→ Bruit asymptotique des WCS du stack :  $\sim 20$  mas. (D. Boutigny)

Ajustement simultané de  $\sim 700$  images (du CFHTLS)  
Bruit asymptotique :  $\sim 2.5$  mas  
 $2.5 \cdot \sqrt{2} = \sim 4$  mas.

→ un ajustement simultané améliore le mapping entre images.

(A&A, 557 A55, 2013).

# Plan : combler certains manques

## 1 : astrométrie simultanée

- Code de référence : scamp (Astromatix)
- On peut faire mieux (plus proche de l'optimalité).
- Méthode :
  - Ajuster des WCS (rustiques) aux images d'entrée
  - Associer les catalogues de l'ensemble des poses
  - Associer ce catalogue d'associations à un catalogue externe (→ fixer le repère sidéral)
  - Ajuster les positions sidérales des objets communs et les WCS des images, en tenant compte des erreurs de mesure.

# Plan : combler certains manques

## 1 : astrométrie simultanée

- SNLS a une implémentation partielle de cela.
- Etapes :
  - Association combinatoire à un catalogue externe (+ ajustement du WCS). Associer une pose plutôt qu'un CCD.
  - Associer les catalogues de l'ensemble des poses
    - Il faut un algorithme en  $N_1 * \log(N_2)$ , à faire ?
  - Ajuster les positions des objets communs et les WCS des images.
    - Calcul efficace des dérivées.
    - Choisir un cadre d'algèbre pour matrices creuses.

# Plan : combler certains manques

## 1 : uniformisation du plan focal

- SNLS a une implémentation complète de cela.
  - présentation de Nicolas R. (& Betoule et al, 2013)
- Cette implémentation travaille sur les mêmes associations que l'astrométrie simultanée. Ce code est le même.
- L'ajustement lui-même est beaucoup plus simple.
- → ce n'est pas un gros effort comparé à l'astrométrie simultanée, parce que ça recycle une bonne partie du code.

# Plan : combler certains manques.

## Aspects politiques

- Nos interlocuteurs du Data Management (DM) agréent à la réalité des manques.
- Ils soutiennent notre projet de les prendre en charge. Cela nécessite une coordination.
- Réunion convoquée la semaine du 20 avril.
- C'est un changement de paradigme remarquable : le Data Management n'est plus seulement un service à la collaboration, il devient un produit de la collaboration.

# Plan : combler certains manques.

## Aspects politiques

- Modèle courant : LSST produit les données, DESC produit les contraintes cosmologiques.
- C'est compris comme :
  - LSST produit des images puis des catalogues
  - DESC les analyse.
- On peut concevoir quelques doutes.
- Comprendre ce que font les codes est en général nécessaire à la science. On comprend encore mieux si on l'a fait. Ca marche rarement du premier coup.

# Plan : combler certains manques aspects sociologiques

- Pourquoi ces manques, alors que le code est en volume  $>10$  fois plus gros qu' Astromatix et a déjà englouti beaucoup plus d'années-hommes ?
- Parce qu'il n'y a essentiellement pas de scientifiques impliqués.
- Un groupe modeste mais décidé peut avoir une contribution significative ( décisive ? ).
- On peut (doit?) se donner un but scientifique innovant, mais cela doit résulter d'un débat (vaguement) organisé au sein de LSST-France.

# Donc,

- Notre implication dans le calcul impose de contribuer scientifiquement au stack (i.e. produire des bugs nous aussi).
- Améliorer le stack est nécessaire si on veut faire au moins aussi bien que les traitements de référence du CFHTLS.
- Un ou plusieurs buts scientifiques seraient bienvenus. But déjà identifié : transients. C'est tout ? Ca m'étonnerait.
- Un petit groupe de personnes est prêt à s'impliquer, on ajustera les ambitions aux forces en présence.

# Besoins moins pressants, mais cependant sérieux

- Faire une comparaison sérieuse de la vitesse de stacking comparé à swarp. Localiser les goulots si le diagnostic de lenteur du stack persiste.
- Etudier un design du stacking sans images intermédiaires sur disque pour limiter les I/O. C'est pas facile.
- Imaginer une architecture des moyens de calcul qui limite par construction les engorgements d'I/O.

# Besoins moins pressants, mais cependant sérieux

- Brighter-fatter, on fait comment ?
  - On attend de voir comment DES s'en tire ?
  - On attaque ?
- Détection/élimination des avions/satellites.
  - Collecte des savoir-faire ?
- Mesure du shear

# Quelques questions

- Quels sont les outils indispensables à la calibration manquants dans le stack ?
  - Uniformisation de la grille (aka résolution de la « grille »)
- Quels sont les tests de performance de la mesure des couleurs des galaxies ?