

Astrométrie - algorithmes

Pierre Astier

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

29/05/2015



Astrométrie

- Mesure des positions d'objets sur le ciel.
- Deux étapes :
 - Mesure des positions sur le senseur.
 - Construction d'une « solution astrométrique » qui rapporte un ensemble de mesures dans un repère commun.

Positions sur le senseur

- Au moins 3 méthodes :
 - Barycentre sur un domaine
 - Barycentre pesé sur un domaine
 - Position de PSF (objets ponctuels)

Barycentre sur un domaine (« position d'ouverture »)

$$X_0 = \frac{\sum_p (I_p - s) X_p}{\sum_p (I_p - s)} = \frac{\sum_p (I_p - \overset{\text{ciel}}{\downarrow} s) X_p}{f}$$

Source faible,
bruit uniforme $\text{Var}[I_p] = b$

$$\text{Var}[X_0] = \frac{b \sum_p (X_p - X_0)(X_p - X_0)^T}{f^2}$$

- $E[X_0]$ est bien défini si on augmente le domaine d'intégration,...
- Mais $\text{Var}[X_0]$ diverge.
→ L'espérance des seconds moments n'est pas définie !

Barycentre pesé

$$X_0 = \frac{\sum_p (I_p - s) W[X_p - X_0] X_p}{\sum_p W[X_p - X_0] (I_p - s)}$$

W est en général une fonction paire.

On a donc une définition implicite de X_0

Source faible,
bruit uniforme $\text{Var}[I_p] = b$

$$\text{Var}[X_0] = \frac{b \sum_p W^2[X_p - X_0] (X_p - X_0)(X_p - X_0)^T}{\left[\sum_p W[X_p - X_0] (I_p - s) \right]^2}$$

Variance bornée si W décroît plus vite que $1/R^2$

Barycentre gaussien

$$W(X_p - X_0) \propto \exp \left[-\frac{1}{2} (X_p - X_0)^T \mathbf{w}_g (X_p - X_0) \right]$$

On peut adapter $\mathbf{w}_g$ pour décrire la forme de l'objet

(SDSS, Bernstein & Jarvis (2002) , Astier et al (2013))

L'ajustement de la position et de $\mathbf{w}_g$ peut se faire itérativement
En cherchant le point fixe de:

$$X_0 = \frac{\sum_p X_p W[X_p] (I_p - s)}{\sum_p W[X_p] (I_p - s)}$$
$$\mathbf{w}_g^{-1} = 2 \frac{\sum_p (X_p - X_0) (X_p - X_0)^T W[X_p] (I_p - s)}{\sum_p W[X_p] (I_p - s)}$$

$$W[X_p] \equiv \exp \left[-\frac{1}{2} (X_p - X_0)^T \mathbf{w}_g (\vec{X}_p - \vec{X}_0) \right]$$

Position de PSF

$$\chi^2 = \sum_p \frac{[I_p - s - fP(X - X_0)]^2}{V[I_p]}$$

PSF
↙

- On est obligé d'ajuster simultanément flux et position.
- Si la PSF est paire, flux et position sont non corrélés.
- En revanche, flux et carré des écarts de position sont corrélés.
→ le flux est toujours sous-estimé

Flux et position de PSF

$$\begin{aligned}
 E[\hat{f}] &= f \frac{E[\int P(x, y) P(x - \hat{\delta}_x, y - \hat{\delta}_y) dx dy]}{\int P^2(x, y) dx dy} \\
 &= f \left\{ 1 + \frac{1}{2} \frac{E[\hat{\delta}_x^2] \int P \partial_x^2 P + E[\hat{\delta}_y^2] \int P \partial_y^2 P}{\int P^2(x, y) dx dy} + \mathcal{O}(\delta^4) \right\}
 \end{aligned}$$

$$Var[\hat{f}] = \left(w \int P^2 \right)^{-1}$$

$$Var[\hat{\delta}_x] = E[\hat{\delta}_x^2] = \left(w f^2 \int (\partial_x P)^2 \right)^{-1} = \left(-w f^2 \int P \partial_x^2 P \right)^{-1}$$

(Guy et al 2010, appendice E)



$$E[\hat{f}] \simeq f \left\{ 1 - \frac{Var[\hat{f}]}{f^2} \right\}$$

Résumé

- Les positions non pesées n'ont pas de variance définie. Les second moments ne sont pas bien définis.
- Les positions pesées fournissent des estimateurs de variance finie si le poids décroît assez vite. On peut adapter le poids à la forme de l'objet.
- Les positions de PSF sont une forme de position pesée, où le poids est la PSF. C'est réservé aux source ponctuelles, et les flux de PSF sont toujours sous-estimés, d'autant plus que l'objet est faible.

Etablir un WCS

- WCS : World Coordinate System
- Proposition de normalisation pour l'encodage des transformations entre coordonnées sur une image FITS et coordonnées plus conventionnelles.
- Exemples :
 - Coordonnées « pixel » $\rightarrow$ RA, Dec
 - Indice dans un spectre tabulé $\rightarrow$ longueur d'onde.
 - Coordonnées « pixel » d'une image de la Lune $\rightarrow$ longitude et latitude lunaire.

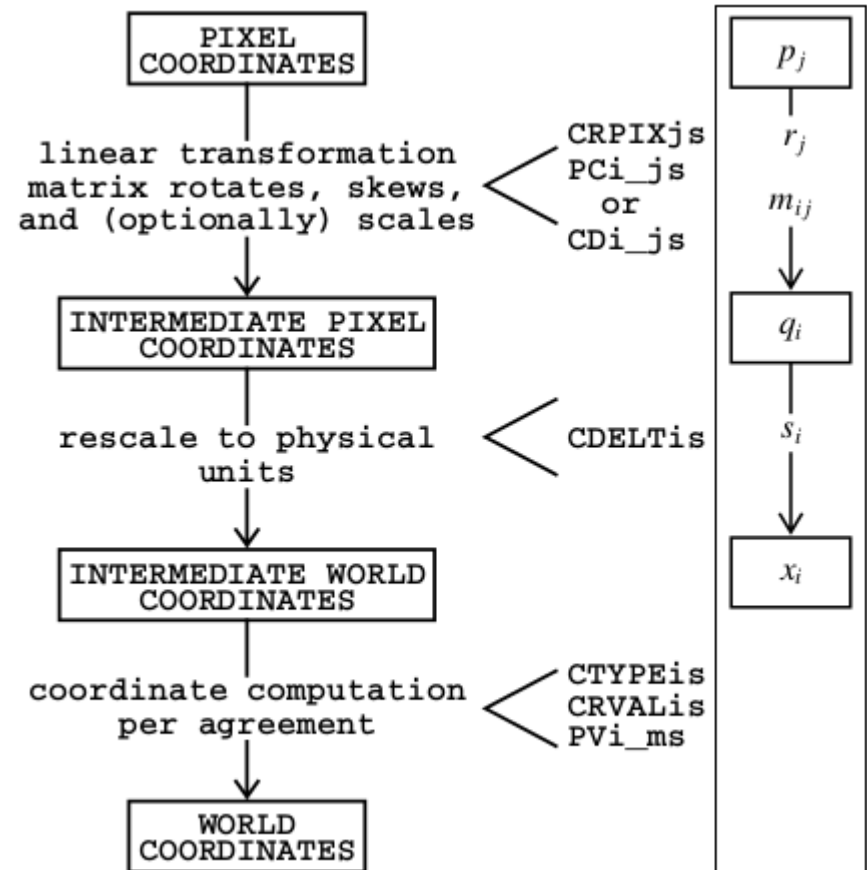
WCS : concepts

Greisen, E.W. & Calabretta, M.R. A&A 395, (2002)

Convention « redondante » :
il y a plusieurs manières
d'encoder la même transformation

Proposition qui rassemble
les pratiques de l'époque en
en excluant le moins possible

Standard de fait, mais avec
des extensions peu utilisées



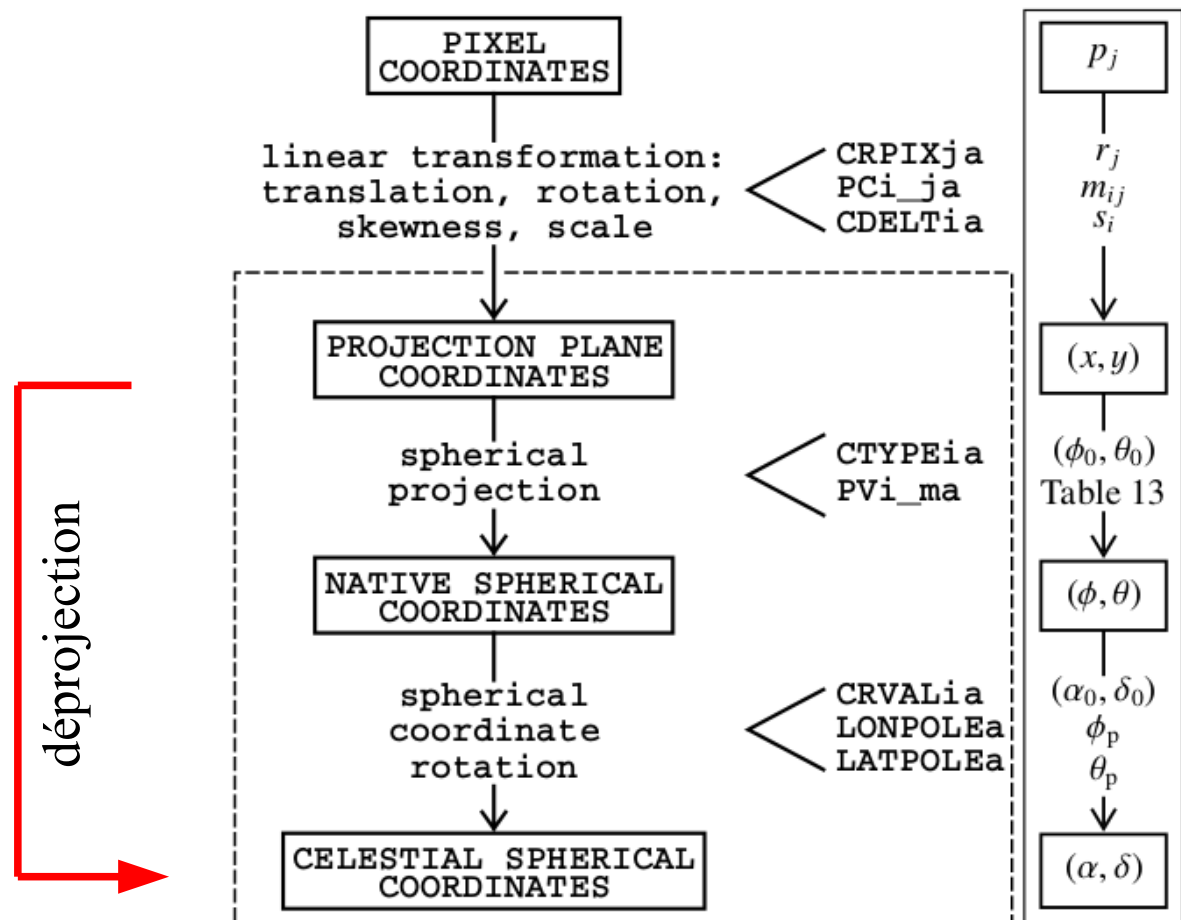
WCS : coordonnées sidérales (1)

Calabretta, M.R., & Greisen, E.W., (2002), A & A, 395, 1077-1122

En pratique :

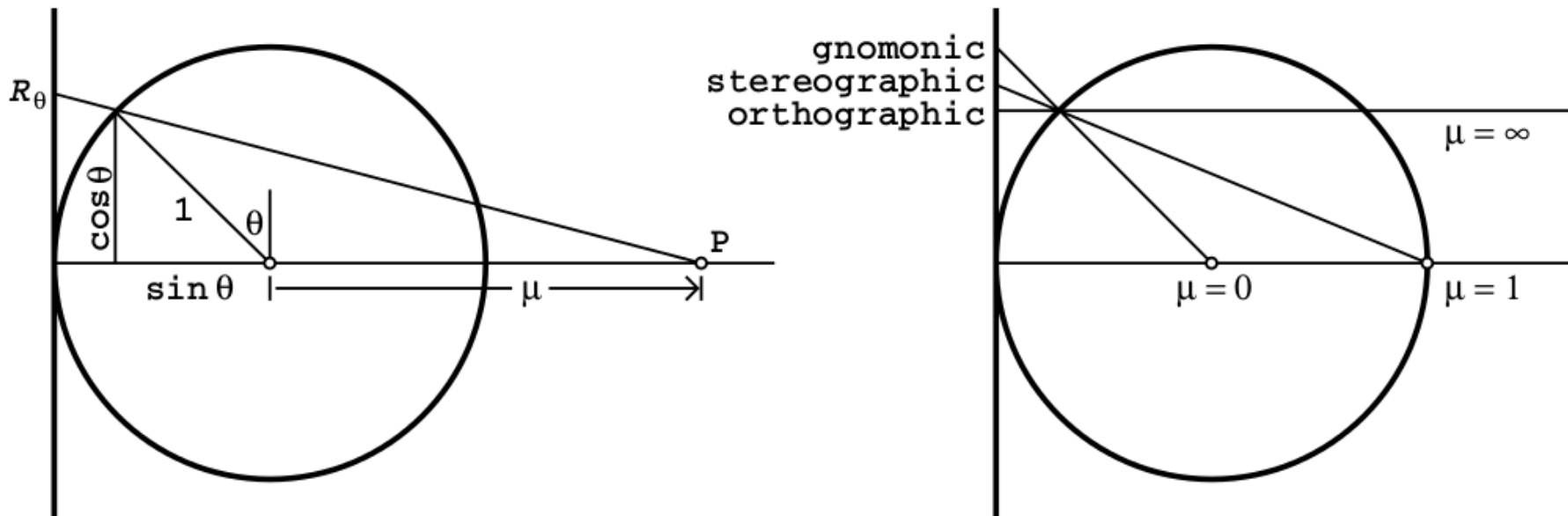
CRPIXi
CDi_j
ou CDELTi

CTYPEi
CRVALi



WCS : coordonnées sidérales (2)

Il existe des douzaines de projections. La plus utilisée :
projection « gnomonique »



Gnomonique :

CTYPE1 = RA__TAN

CTYPE2 = DEC__TAN

Etablir le WCS d'une image

- Mécanique du télescope : en aveugle, au mieux 5" en général moins bien.
- On peut améliorer en associant le catalogue de l'image à un catalogue de référence, le plus souvent moins profond que notre image.
- Donc :
 - Associer le catalogue de l'image et le catalogue de référence, par combinatoire.
 - Ajuster les paramètres (certains...) du WCS
 - Encoder dans le header FITS.

Combinatoire de l'association

- On commence par projeter le catalogue de référence sur un plan tangent à un point « raisonnable », e.g. le pointé annoncé du télescope.
- Dans l'hypothèse où l'image est faiblement distordue, on cherche 4 paramètres, e.g. :
 - 1 décalage (2)
 - 1 rotation (1)
 - 1 échelle (1)

Association : la méthode « 2+2 »

Le catalogue est un ensemble de coordonnées $\{X_i\}$

On fabrique des segments $r_{ij} = X_j - X_i$

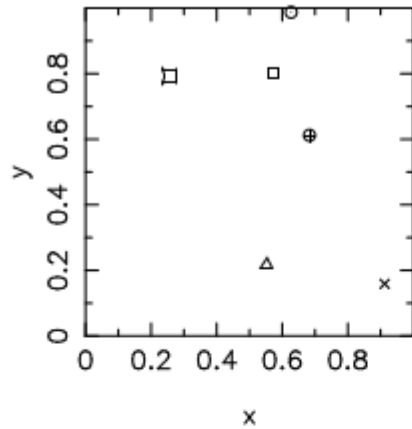
On fait l'histo 2-d ($\log(r)$, $\arg(r)$)

Un décalage spatial ne change rien à l'histogramme

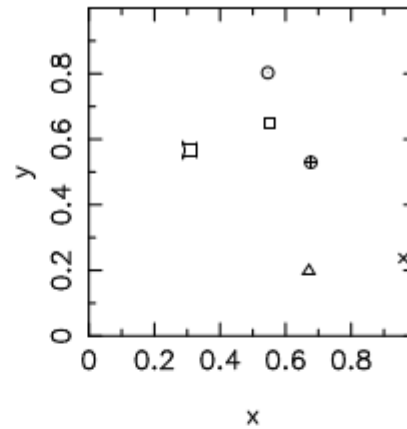
Un changement d'échelle ou une rotation le décale

On fait la même chose pour le catalogue de référence et on
Cherche le maximum de corrélation.

catalogue A



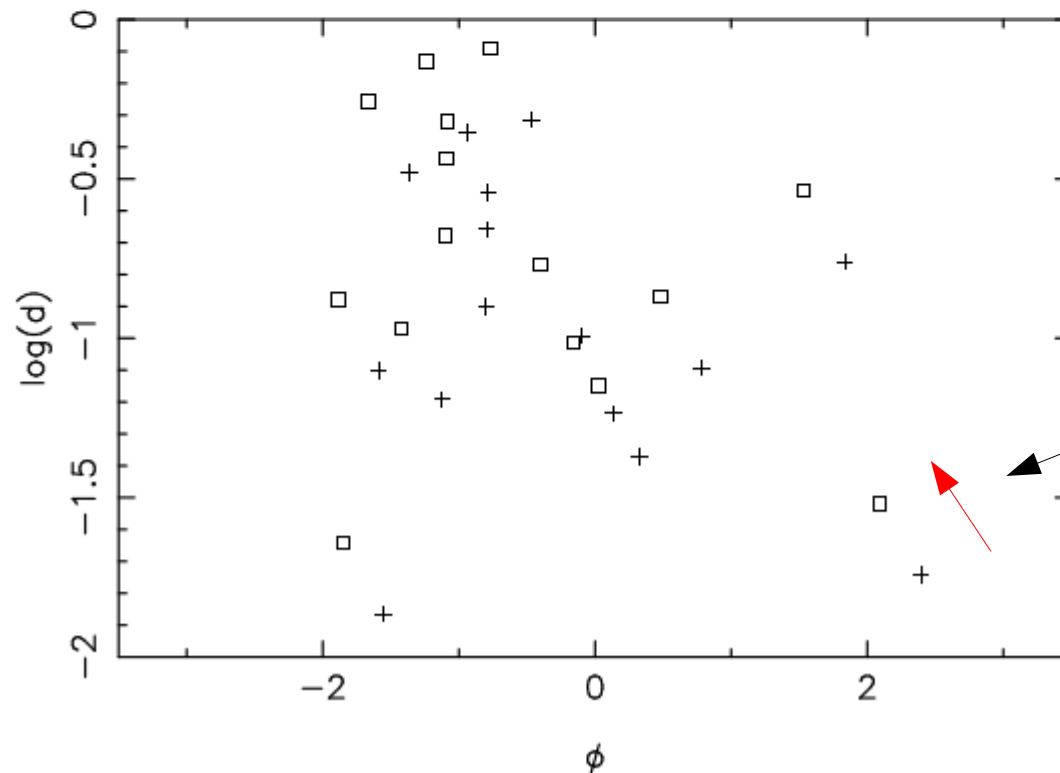
catalogue B



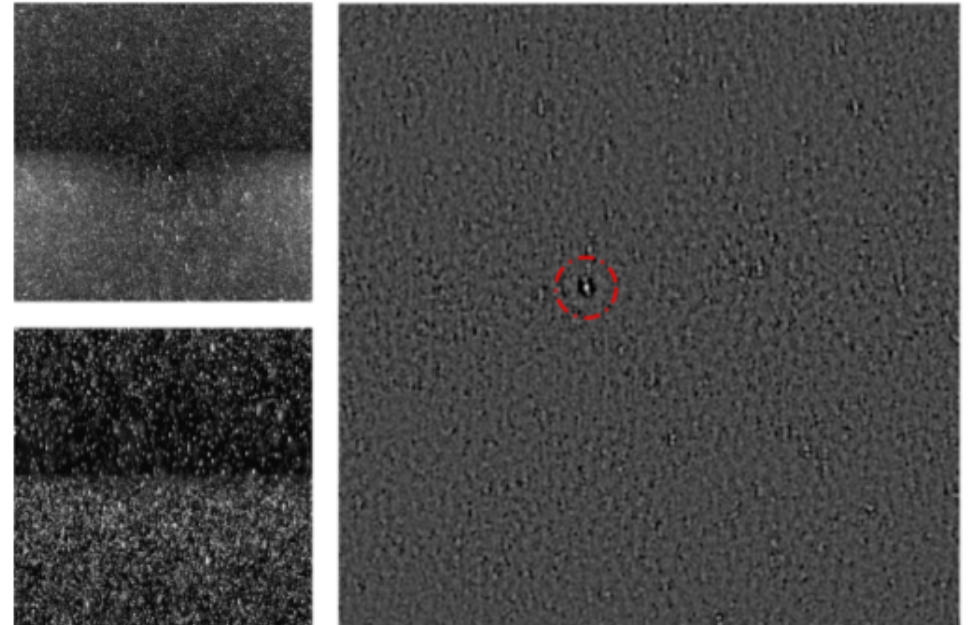
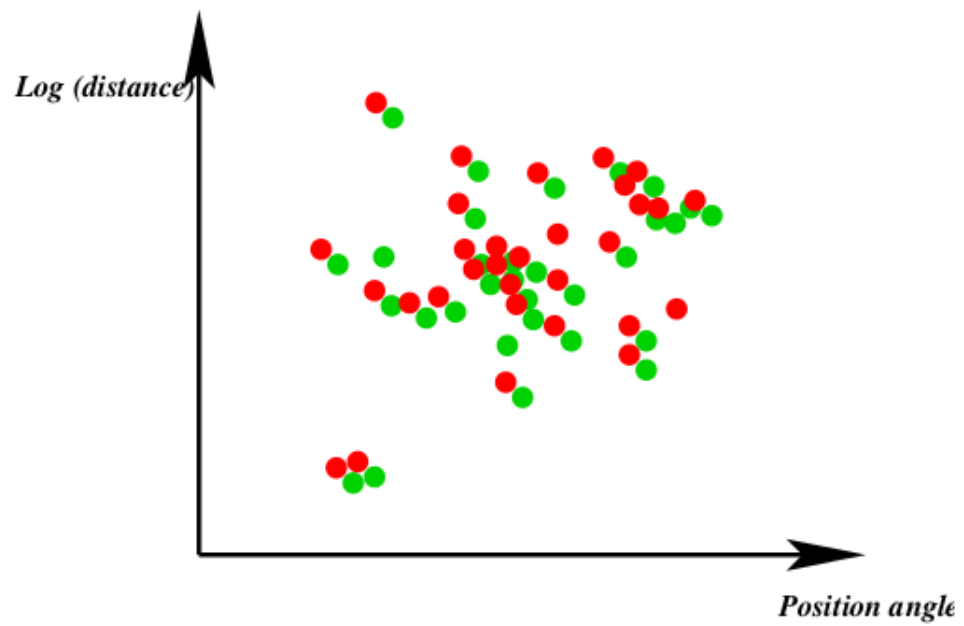
« 2+2 »

Après avoir trouvé l'échelle
et l'angle, on trouve le
décalage spatial par la
même méthode de
corrélacion

log-separation and angles of pairs



solution



Exemple tiré de la doc de SCAMP

4 paramètres d'un coup (1)

- La méthode 2+2 fait 2 transformées de Hough
- On peut essayer de faire les 4 paramètres d'un coup, ce qui rend la méthode plus solide.

$$\{X_i\} \rightarrow r_{ij} = X_j - X_i$$

$$\{X'_i\} \rightarrow r'_{ij} = X'_j - X'_i$$

De tous les couples (r_{ij}, r'_{kl}) , on peut tirer les 4 paramètres et en faire l'histogramme 4-D

- Il faut tronquer les catalogues, sinon la combinatoire explose
- A ce stade il est judicieux de conserver les objets saturés.

4 paramètres d'un coup (2)

$$\{X_i\} \rightarrow r_{ij} = X_j - X_i$$

$$\{X'_i\} \rightarrow r'_{ij} = X'_j - X'_i$$

De tous les couples (r_{ij}, r'_{kl}) , on peut tirer les 4 paramètres et en faire l'histogramme 4-D :

$$S = |r_{ij}|/|r'_{kl}|, \quad \text{angle} = \arg(r_{ij}) - \arg(r'_{kl}), \quad \text{delta} = X_i - X'_k$$

Première idée : Histo $(S, \text{angle}, \text{delta}_x, \text{delta}_y)$

Ou bien : Histo (S, angle, i, k)

Avec la deuxième méthode :

- On retrouve plus rapidement les couples de paires qui contribuent à un pic dans l'histo,
- On s'épargne la réflexion sur le choix du binning en delta.

4 paramètres d'un coup (3)

- Pour chaque « maximum » de l'histogramme :
 - Collecter les paires
 - Ajuster la transfo (linéaire)
 - Collecter les paires dans les listes originales
 - Le nombre de paires trouvé est le facteur de mérite d'une solution.
- On choisit la meilleure solution !
- En général les premiers maxima sont des solutions similaires.
- Temps d'exécution : $O(5s)$ pour $N_1, N_2 \sim 100$

Ajuster le WCS

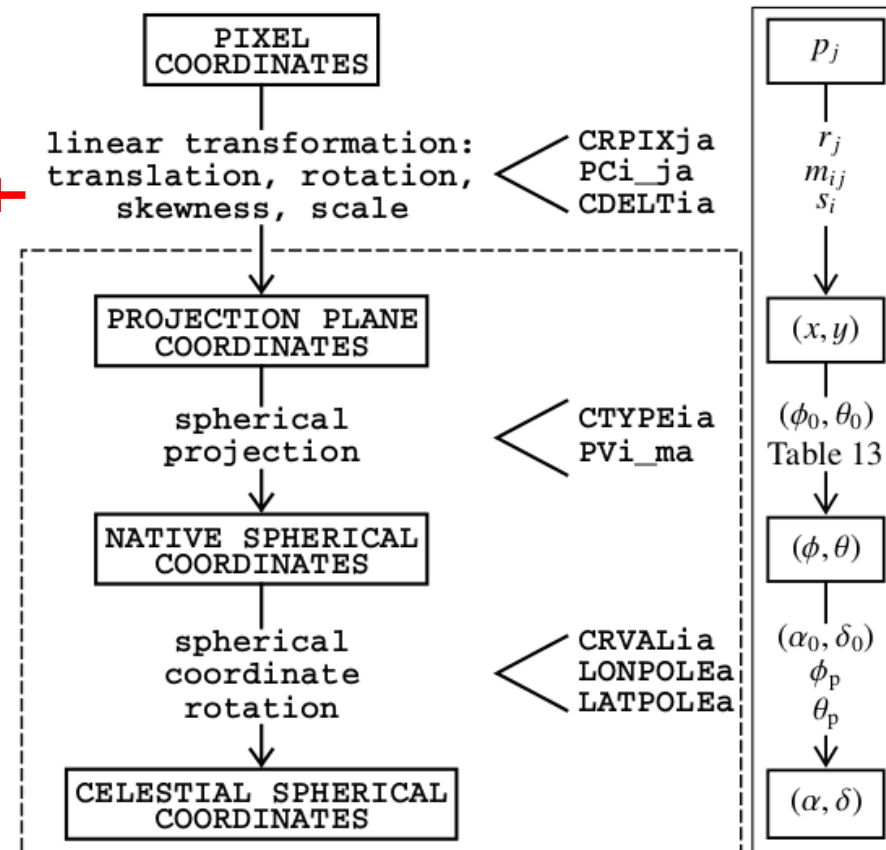
$$\begin{pmatrix} x_{TP} \\ y_{TP} \end{pmatrix} = \begin{pmatrix} CD1_1 & CD1_2 \\ CD2_1 & CD2_2 \end{pmatrix} \begin{pmatrix} x_p - CRPIX1 \\ y_p - CRPIX2 \end{pmatrix}$$

Pour de petits angles, la déprojection devient :

$$\begin{pmatrix} \alpha \\ \delta \end{pmatrix} = \begin{pmatrix} 1/\cos \delta & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{TP} \\ y_{TP} \end{pmatrix} + \begin{pmatrix} CRVAL1 \\ CRVAL2 \end{pmatrix}$$

Et donc, on ne peut pas ajuster

- les CD (4)
- les CRPIX (2)
- les CRVAL (2)

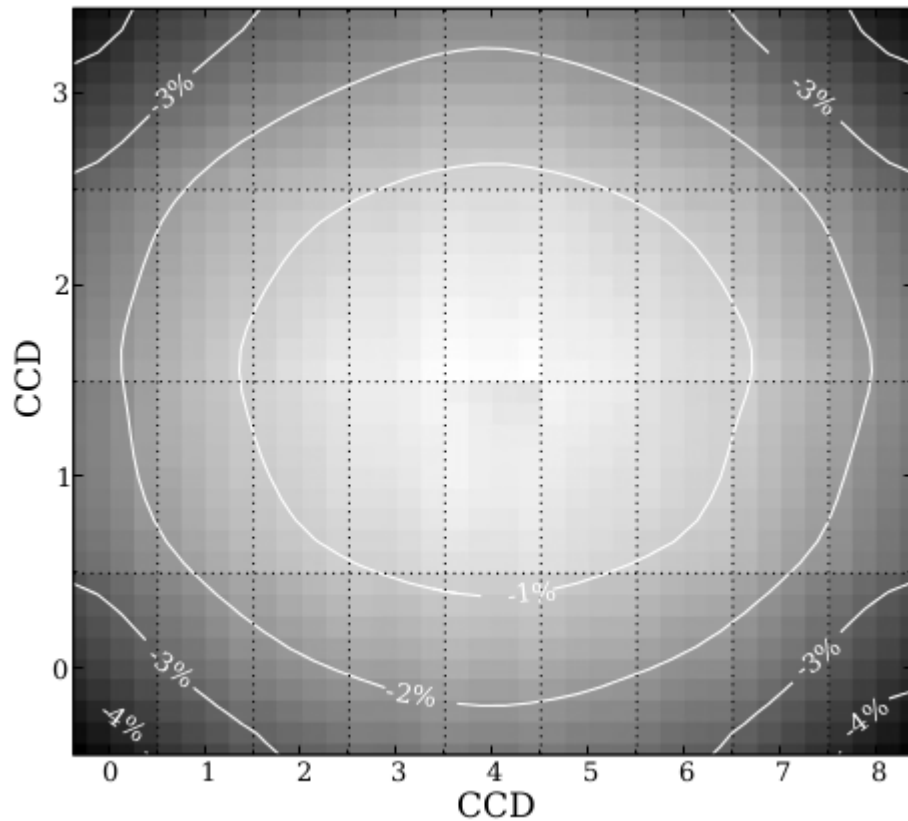


En pratique, le plus simple est de fixer le plan de projection (CRVAL{1,2})

Distorsions

- Le modèle impose une relation linéaire entre coordonnées sur le capteur et plan de projection.
- Ce n'est vrai que si l'image d'une droite sur le capteur est une droite sur ce plan.
- En général les optiques produisent des distorsions importantes.
- Il faut donc un moyen de les encoder.

Distorsions : plusieurs % entre centre et bord



Variation relative du jacobien
i.e. $(\text{Plate-scale})^2$ sur Megacam
(Regnault+ 2009)

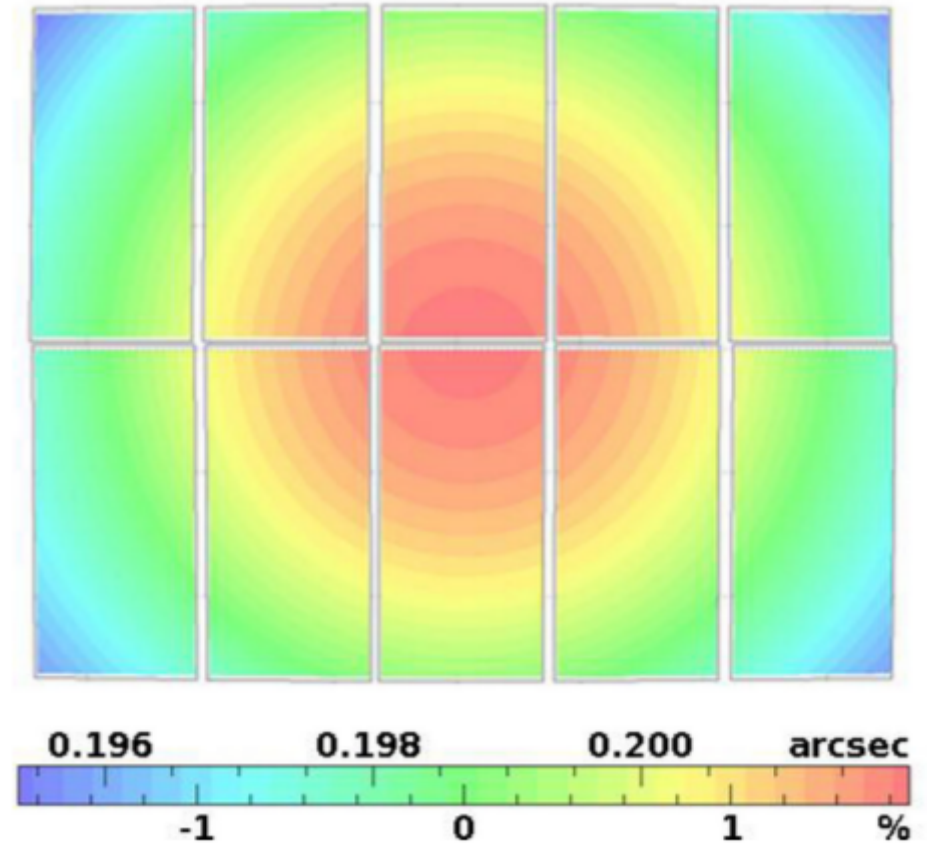


Plate-scale sur Suprime-cam
(von der Linden+ 2014)

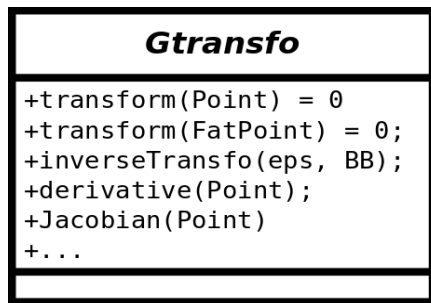
Distorsions

- Pour l'encodage des distorsions, il n'y a pas vraiment de norme universelle. Deux schémas sont « courants », il utilisent tous les deux un polynôme pour décrire les distorsions.
 - SIP : $\text{Pix2TP} = \text{Lin}([\text{CD}, \text{CRPIX}], \text{Pol}(X_{\text{ccd}}))$
 - PV : $\text{Pix2TP} = \text{Pol}(\text{Lin}([\text{CD}, \text{CRPIX}], X_{\text{ccd}}))$
- Dans les deux cas, le résultat est un polynôme de même degré que celui utilisé pour les distorsions
- La séparation est arbitraire : on choisit d'avoir le meilleur WCS si les distorsions sont ignorées

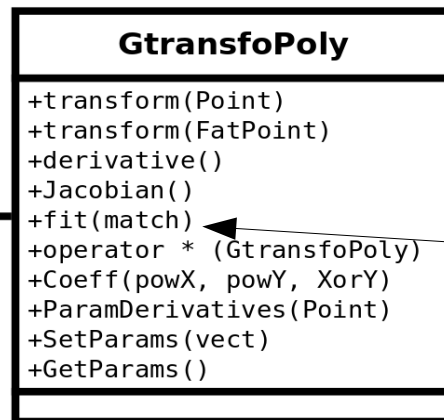
Distorsions

- Implémentations :
 - SIP : lsst-stack, ds9, ~~weslib~~
 - PV : scamp, swarp, Poloka, ~~weslib~~
- En pratique l'un ou l'autre est un détail d'implémentation. Le SIP est facile à implémenter « par dessus » la wcslib, comme c'est fait dans le stack.
- Un détail pratique : la wcslib n'est pas très rapide. Swarp la contourne pour le ré échantillonnage.

Fit du WCS : sketch simplifié de l'implémentation dans Poloka

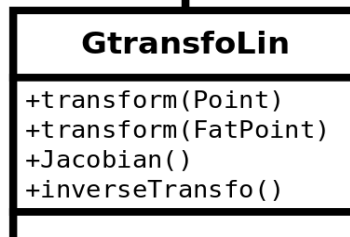
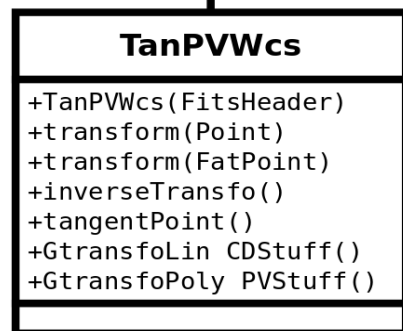


Classe d'interface, transfo du plan sur le plan.



Spécialisation en polynôme

Utilisée pour ajuster des transfos (avec poids) Entre images ou pour un WCS



```
// fit of a PV WCS
{
    MatchList match = FindMatch(...); // pixels -> tangent plane
    GtransfoPoly mapping(order);
    mapping.fit(match);
    GtransfoLin CD; // CDi_j and CRPIX_i
    CD.fit(match); // best linear transfo
    GtransfoPoly PV = mapping*cd.inverseTransfo();
    // GtransfoPoly SIP = cd.inverseTransfo()*mapping;
    TanPVWcs wcs(CD, PV, TangentPoint);
    wcs.WriteToHeader(...);
}
```

La routine `GtransfoPoly::fit` travaille en coordonnées « centrées » mais retourne la transfo correcte en utilisant “operator *”

Associer l'ensemble d'une exposition

- On prend un ensemble d'expositions (complètes) et on évalue la moyenne des transformations pixel-> plan tangent.
- Pour une nouvelle exposition, on transporte l'ensemble des catalogues vers le plan tangent avec ces transformations moyennes.
- On associe (combinatoire) avec le catalogue de référence.
- On récupère les associations dans chaque CCD et on ajuste un WCS par CCD.
- **Bénéfices :**
 - Une seule recherche combinatoire/exposition
 - Les poses courtes peuvent être équipées d'un WCS
- Fonctionne sur 28 Megacam et SuprimeCam.

Plusieurs expositions

- Quand on a plusieurs images du même champ, on peut ajuster les transformations pour minimiser la dispersion entre images.
- On tire parti :
 - D'une résolution de position en général meilleure que les catalogues de référence. Ça va changer avec Gaia.
 - Du fait que bien des objets dans les images ne sont pas dans le catalogue de référence. Ça ne va guère changer avec Gaia.

Aims

- Fit high precision WCS's of an image set prior to stacking, or simultaneous measurements
- Why :
 - limiting the PSF degradation when stacking.
 - Limiting shear induced by astrometric offsets.
 - Precision transformations between images.
- In the ~~Terapix~~ Astromatic suite, this functionality is fulfilled by “Scamp”.

Sketch

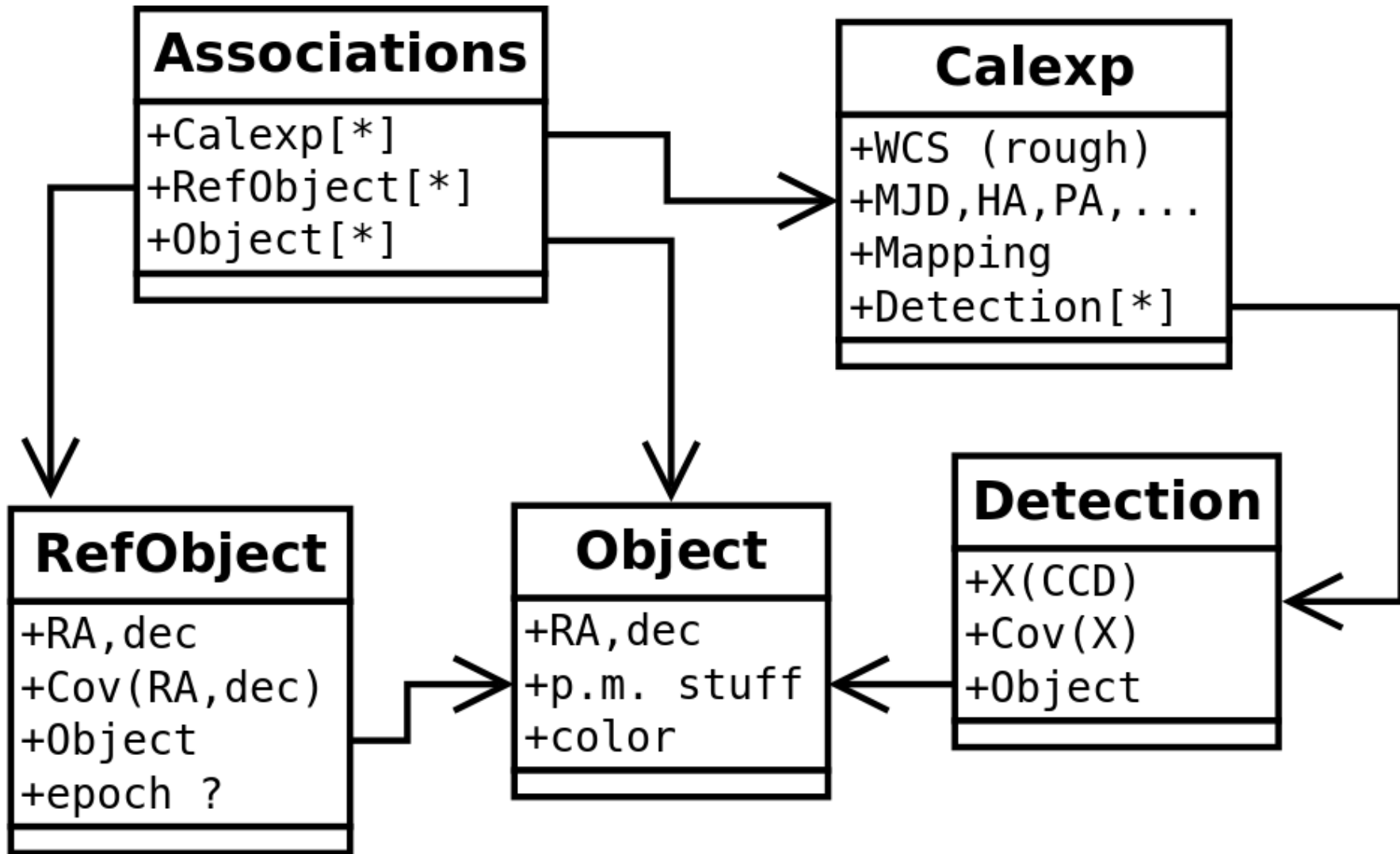
- Our images are already equipped with a catalog, and a rough WCS
- **Stage 1** : associate all the catalogs
- **Stage 2** : associate with an external position catalog, if needed (set the sidereal frame)
- **Stage 3** : fit simultaneously
 - Mappings from input image coordinates to some common frame ($\rightarrow$ WCS's)
 - Positions of the objects in common

Associating detections of the same object

The simple way :

- For each image
 - Load catalog and apply quality cuts
 - Match it to the “Object catalog” (in the tangent plane)
 - Add the unmatched objects to the “Object catalog”
- I know it is fast and efficient (with a 2D $O(N_1 \log(N_2))$ matching)
- One could be worried by the outcome depending in principle of the order of input images.
- In practice, this is not serious if the WCS's are accurate enough (1-2 pixels) and blends are ignored (as they should).
- Faster than reading the catalogs.

Implémentation des associations



Fit : Least Squares

$$\chi_{meas}^2 = \sum_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]^T W_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]$$

Measurement
terms

c,d : calexp, detection

$\mathbf{M}_c$: mapping (pixel $\rightarrow$ TP), one per calexp

$X_{c,d}$: measured position of the object (pixels)

P_c : projection (sky $\rightarrow$ TP)

$W_{c,d}$: Measurement weight (1/var), transformed through M_c .

$\mathbf{F}_i$: (sky) position of the object (measured as $X_{c,d}$)

$$\chi_{ref}^2 = \sum_j [P(\mathbf{F}_j) - P(R_j)]^T W_j [P(\mathbf{F}_j) - P(R_j)]$$

Reference
terms

P : some (user-provided) projection

$\mathbf{F}_j$: (fitted) sky position of the object

R_j : sky position of the object fitted (reference catalog)

W_j : R_j weight (1/var), transformed through P

Least Squares (2)

$$\chi_{meas}^2 = \sum_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]^T W_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]$$

Measurement
terms

...

$W_{c,d}$: Measurement error, transformed through M_c .

.... $\rightarrow$ so W depends on some fitted parameters !

Yes, but in practice, the scale of M is extremely well known, so this can be ignored.

$$\chi_{ref}^2 = \sum_j [P(\mathbf{F}_j) - P(\mathbf{R}_j)]^T W_j [P(\mathbf{F}_j) - P(\mathbf{R}_j)]$$

Reference
terms

P : some (user-provided) projection

Why $P(\mathbf{F}_j) - P(\mathbf{R}_j)$ rather than just $(\mathbf{F}_j - \mathbf{R}_j)$?

because the distance on the sphere is **not** Euclidean

Least Squares (3)

$$\chi_{meas}^2 = \sum_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]^T W_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]$$

- This setup accommodates the fit of mappings between images:
 - All the P_c are set to identity.
 - One of the M_c is set to identity.
 - No external catalog nor “reference terms”.
- This yields the optimal mapping between images, given position measurements and their uncertainties.

Différence avec SCAMP

$$\chi^2 = \sum_s \sum_a \sum_{b>a} w_{s,a,b} \|\xi_a(\mathbf{x}_{s,a}) - \xi_b(\mathbf{x}_{s,b})\|^2, \quad (13)$$

where $w_{s,a,b}$ is the non-zero weight for the pair of detections in fields a and b related to source s :

$$w_{s,a,b} = \frac{1}{\sigma_{s,a}^2 + \sigma_{s,b}^2}. \quad (14)$$

$\sigma_{s,f}$ is the positional uncertainty for source s in field f .

- Scamp n'ajuste pas les positions des objets. Il minimise l'ensemble des distances 2 à 2. C'est sous-optimal.
- En revanche, Scamp est bien debuggé, produit des tas de plots de contrôle, et a une bonne centaine de leviers de contrôle.

Least Squares (4)

$$\chi_{meas}^2 = \sum_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]^T W_{c,d} [\mathbf{M}_c(X_{c,d}) - P_c(\mathbf{F}_i)]$$

Pour minimiser, on développe le χ^2 au second ordre dans les paramètres,

$$\frac{d^2 \chi^2}{d\theta^2} \equiv H = J J^T, \quad \frac{d\chi}{d\theta} \equiv g$$

J est une matrice rectangulaire, $N_{\text{par}} \times N_{\text{carrés}}$

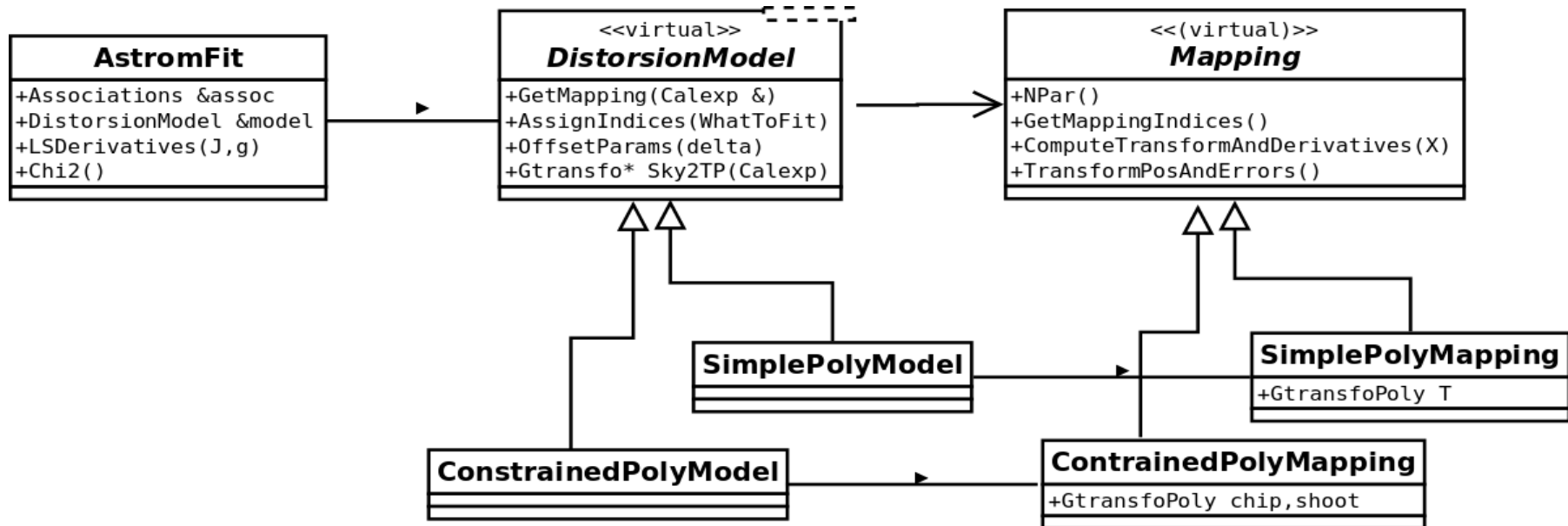
et on résout pour la meilleure direction de descente :

$$H \delta\theta = -g$$

Implementation

- Least squares with (mostly analytical) computation of derivatives (w.r.t positions and parameters).
- Sparse matrix algebra (Eigen3 package). Similar performance with Cholmod.
- About 1500 lines of new C++ code (~ 10 classes) to implement the fit and the model. Used existing (home-made) classes for everything else.
- The fit talks to the model via two abstract classes.

Fit astrométrique



Les SimplePolyXXX réalisent à la fois un ajustement simultané des WCS (sans utiliser que des images peuvent provenir du même instrument) et un ajustement des transfos entre images (sans référence à un catalogue externe)

Les ConstrainedXXX implémentent un modèle dans lequel les positions relatives des CCDs sont fixes

Le « pré-conditionnement » avant factorisation

Cholesky :

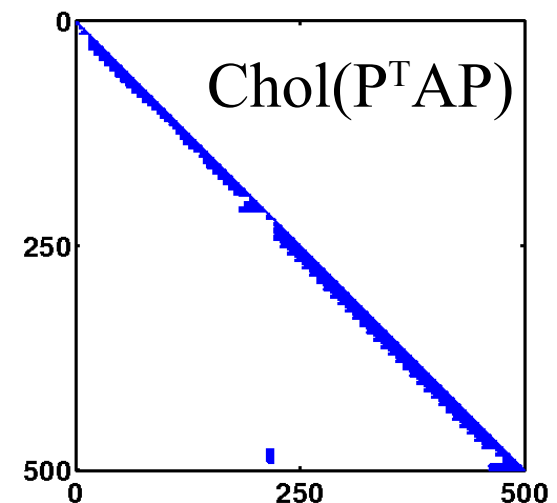
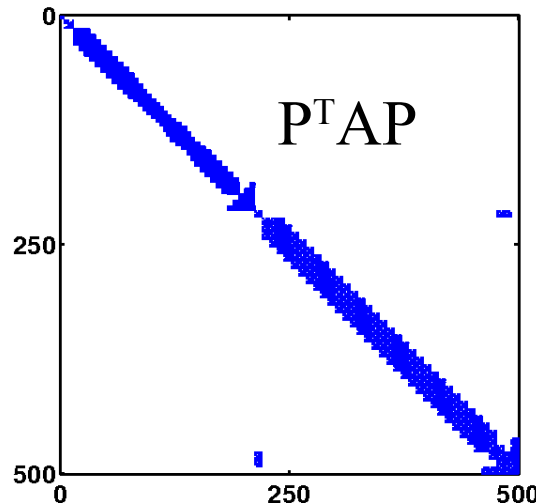
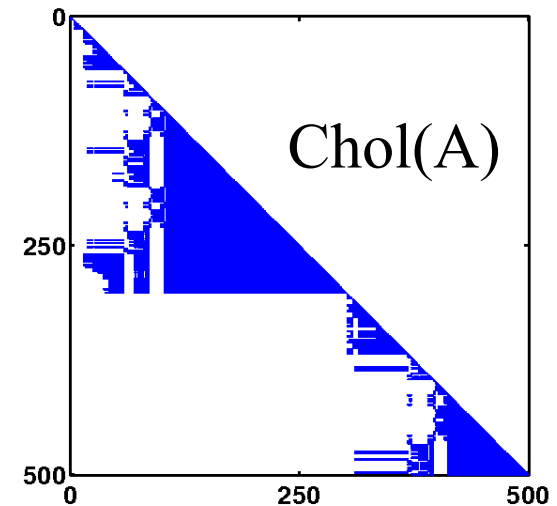
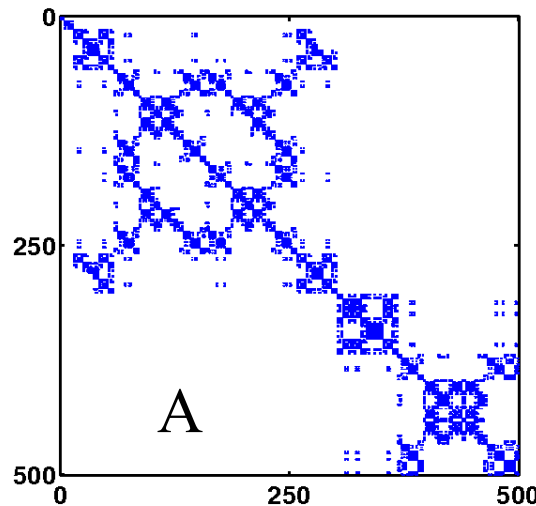
$$A = LL^T$$

(ou LDL^T)

L triangulaire

P : permutation

A : pos-def.



Outlier removal

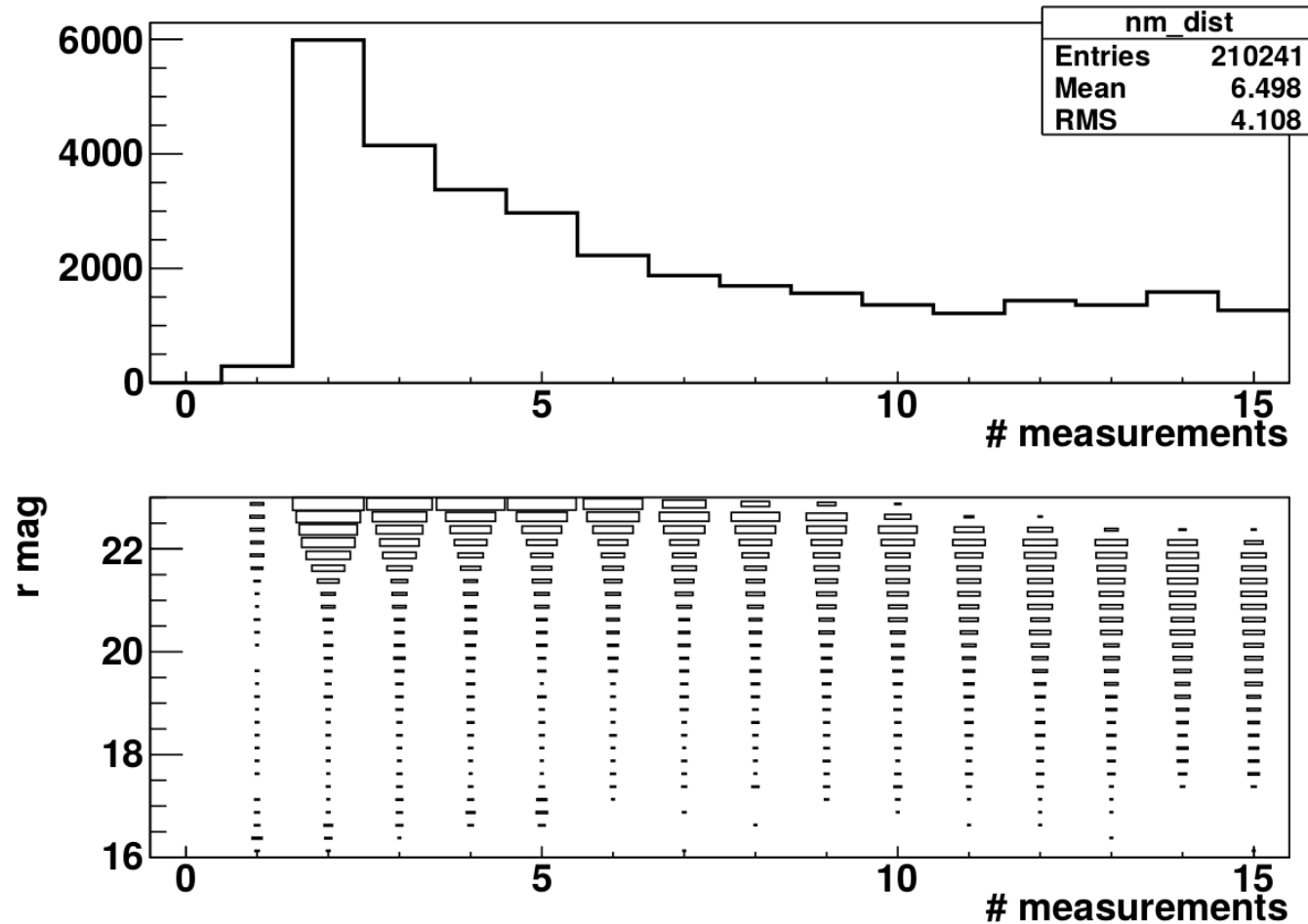
- I have not found a **canned** small-rank update of Cholesky factorizations. The only one I know about is Cholmod providing a rank-1 update.
- So, I have used the following trick: do not remove in a single pass 2 outliers that constrain the same parameter.
- Would require a lot of iterations to come to zero outlier removed.
- I was not that patient: I ran it only 4 iterations.

Trial run

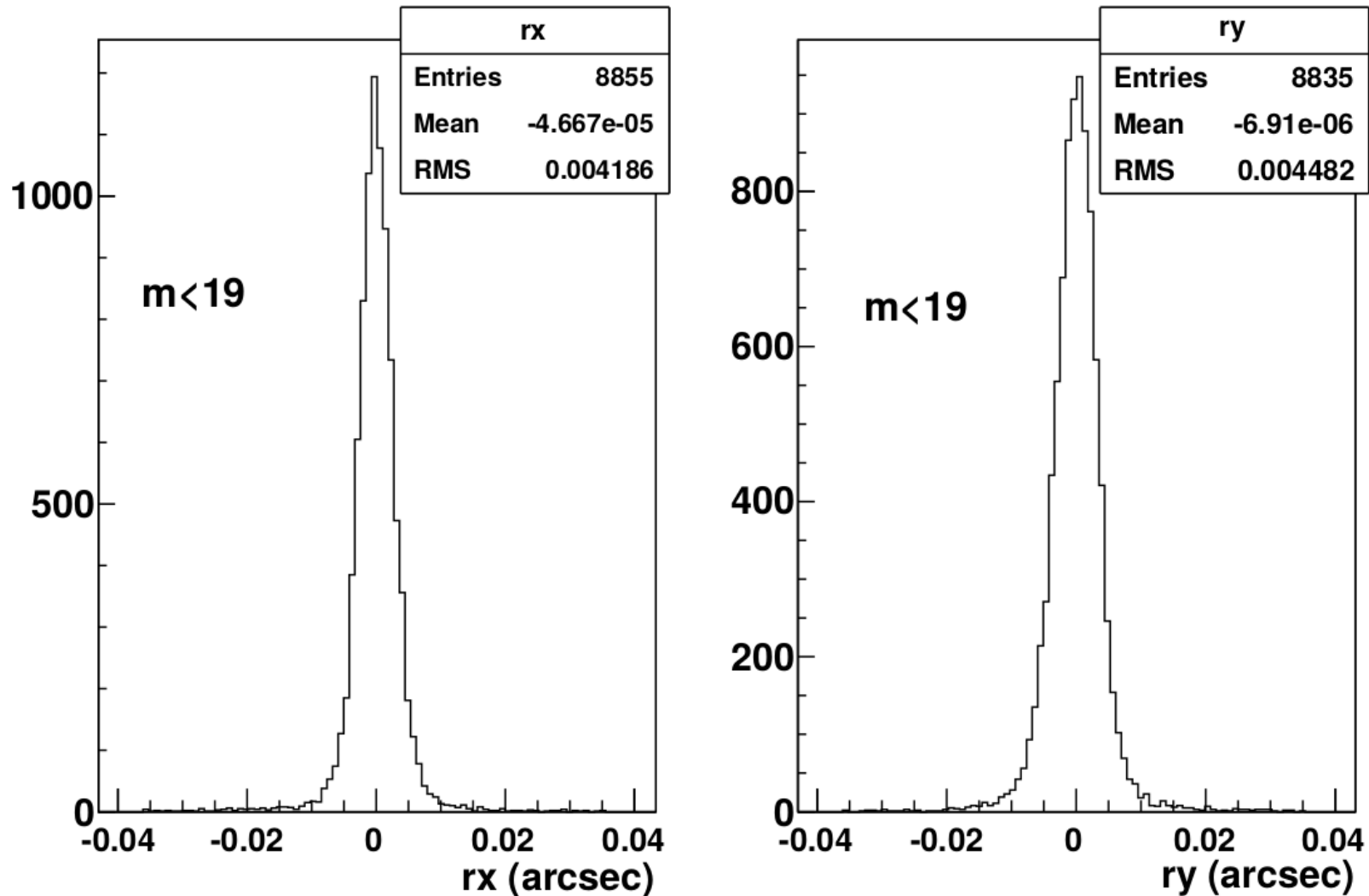
- 15 Megacam r-band 300-s exposures on D3, observed over the same lunation. 540 Calexp's.
- Use USNO-A2 for the reference catalog.
- Ignore proper motions.
- Use Gaussian-weighted positions and associated errors.
- Strict selection of measurements (no flag at all, $S/N > 10$), average of ~ 400 measurements/calexp
- All calexp's have their own mapping parameters as if they all came from different instruments

Trial run

- ~32,500 objects, ~210,000 measurements.



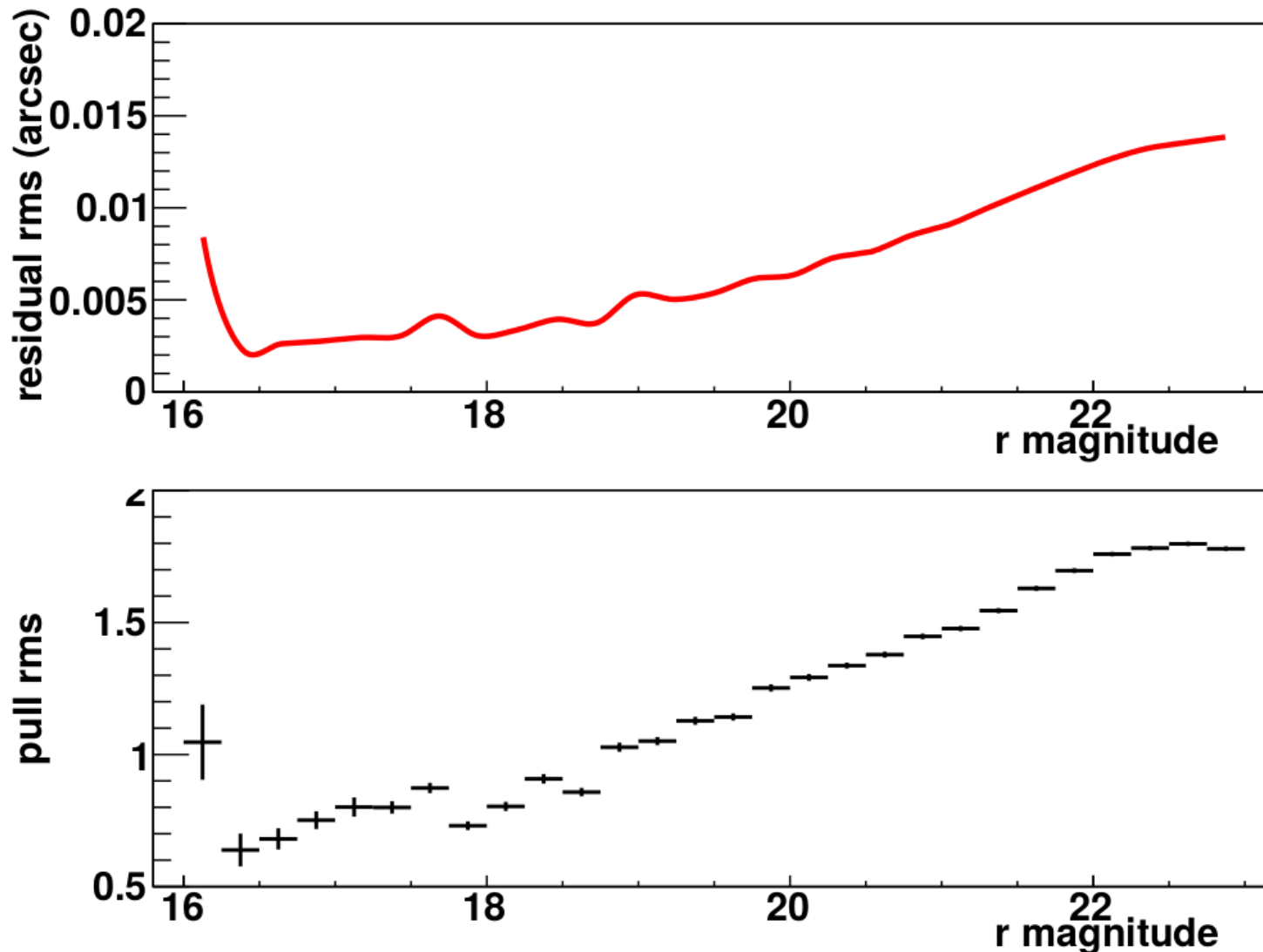
Trial run : residuals



Residuals of the “measurement terms”, these are internal residuals

Residuals vs mag

Use simplistic error model $V = V_{\text{meas}} + (0.02 \text{ pix})^2$



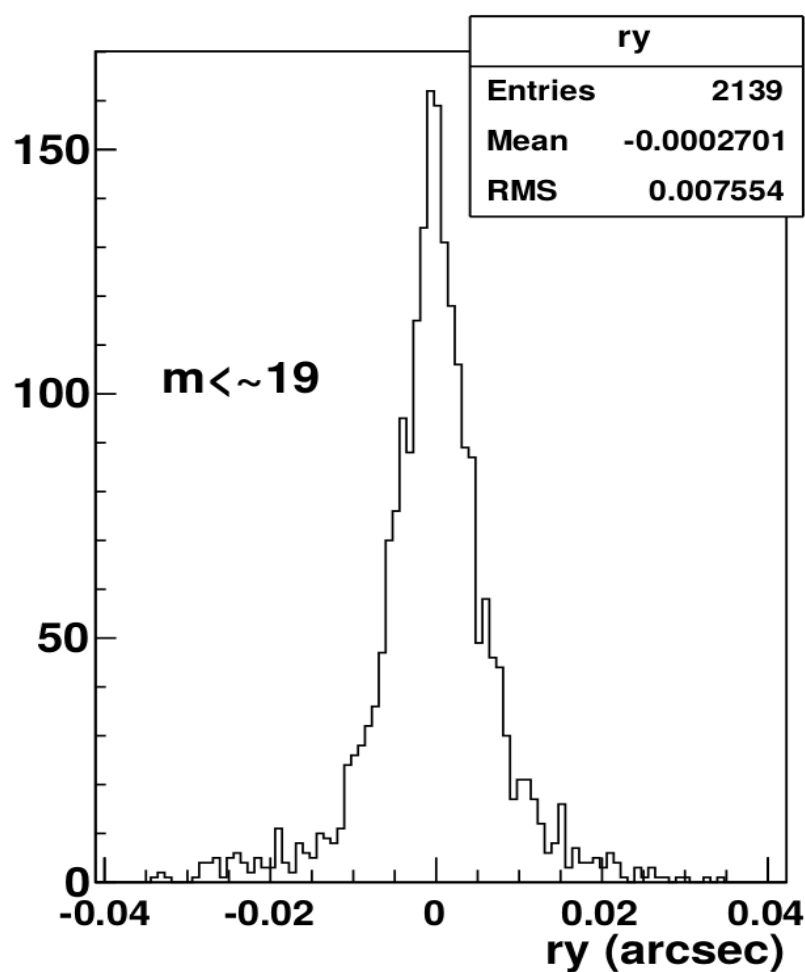
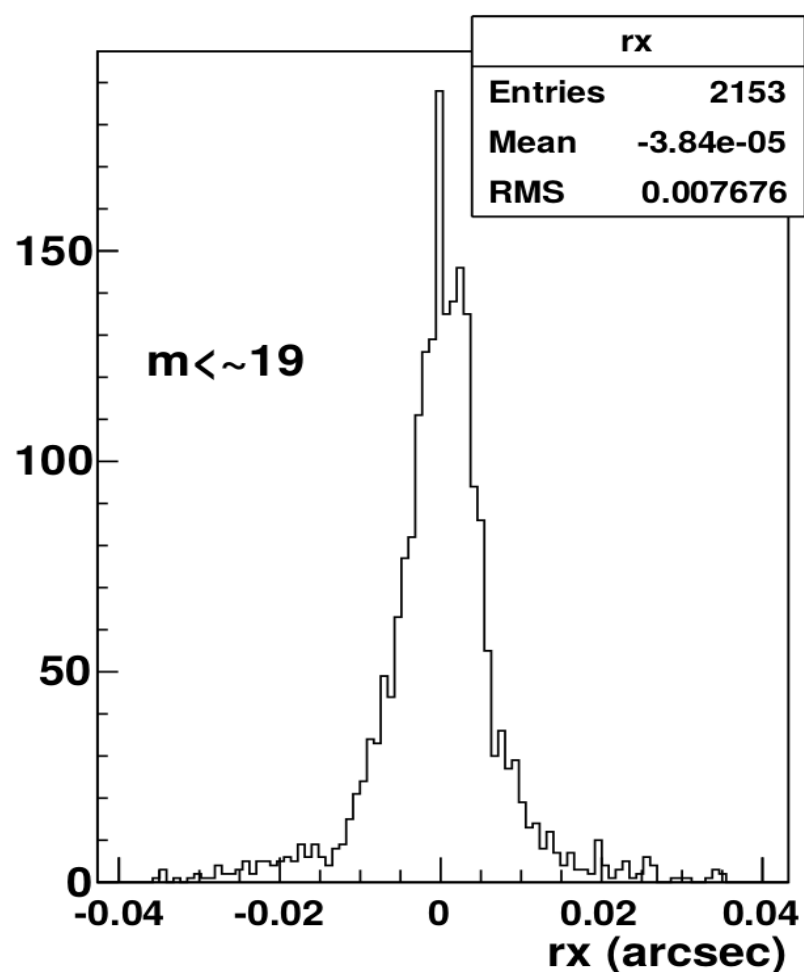
Beware : residuals are **not** Studentized, and The number of measurement is not that large

A smaller constant term would help for the bright side, but not for the faint side

Overall : it could be much worse!

Second trial run: Suprime cam

120-s (i and z)-band exposures reduced by Augustin Guyonnet.
12 exposures, guessed photometric scale. Exactly the same code.



Computer resources

- Execution time: for the 15 Megacam exposures:
 - Reading the (540) catalogs : ~ 100 seconds at 33% CPU
 - Associating : negligible
 - Fitting: $< \sim 20$ s per iteration
 - Computing the derivatives : ~ 1 s
 - “Squaring” the Jacobian, i.e. $H = JJ^T$: ~ 3 s
 - Factorize-solve-update (dim=75,510, nnz=17,164,700) ~ 13 s.
 - Partial fits (positions OR mappings) are solved instantly.
 - Total : 125.137u 19.769s 3:40.44 65.7% (Xeon 2.3 GHz)
- Memory reaches ~ 1 GByte (not completely sure though)

To be done (at least):

- I still have to test the code of the mapping model for a “rigid instrument”. I would not be surprised if the residuals come out much larger. Have to think about relaxing rigidity.
- Study dispersions of faint objects (galaxies mostly).
- Proper motions, parallaxes ? the code to handle proper motions is there, but I have not implemented anything to detect “moving” stars.
- Output of mappings (i.e. WCS's)... Which format? Output of the catalog.

To be done (suite)

- Identifier les possibilités de parallélisme de la factorisation.
- Liaison avec le stack :
 - Header FITS (standardisation).
 - Catalogues.
- Lecture du code.

Code

- <http://gitlab.in2p3.fr/astier/gastro>
- A titre documentaire : ne peut tourner sans poloka-core
- La branche dans laquelle je travaille s'appelle « sparsetriall »
- Doc sur <http://supernovae.in2p3.fr/~astier/gastro/>
- Les dépendances sont installées au CC
- L'insertion dans le stack va nécessiter du découpage

(Simplified) class diagram

