# PROCESSING OF EXTRAGALACTIC SOURCES INSIDE GAIA DPAC

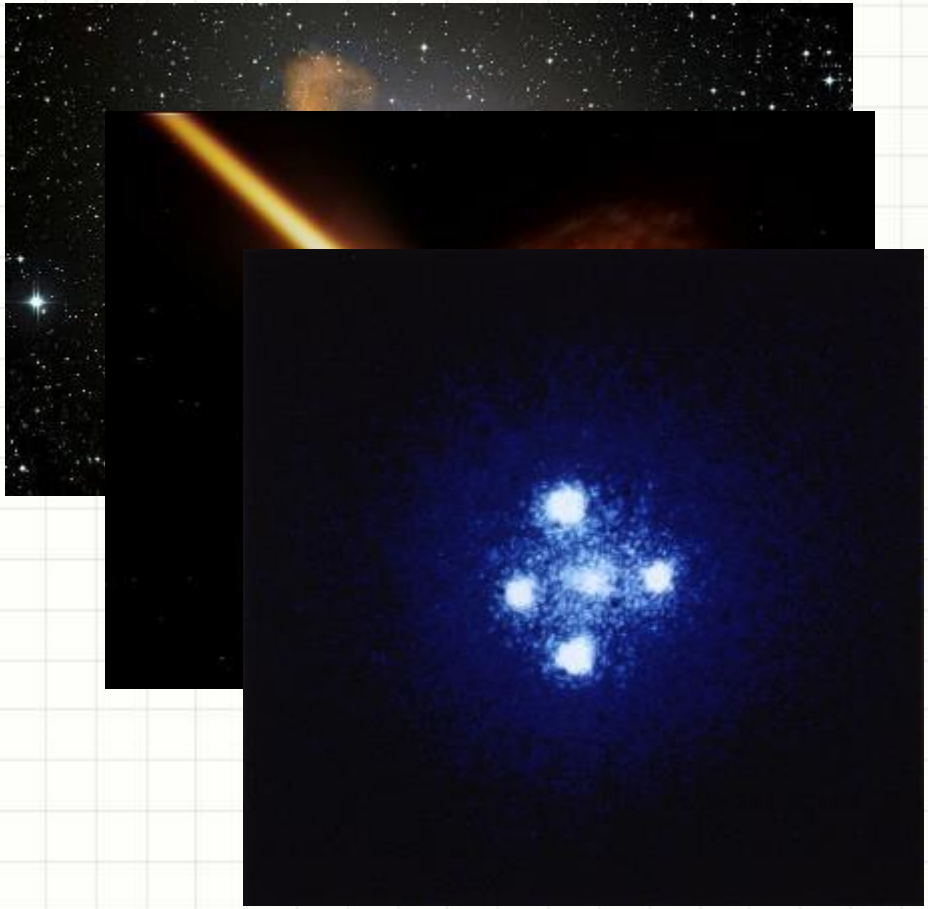
L. Galluccio, Observatoire de la Côte d'Azur

**Gaia for AGN and Extragalactic Science, Paris 8-10th July 20015**

# Extended Object group within DPAC

# Objects of Interest

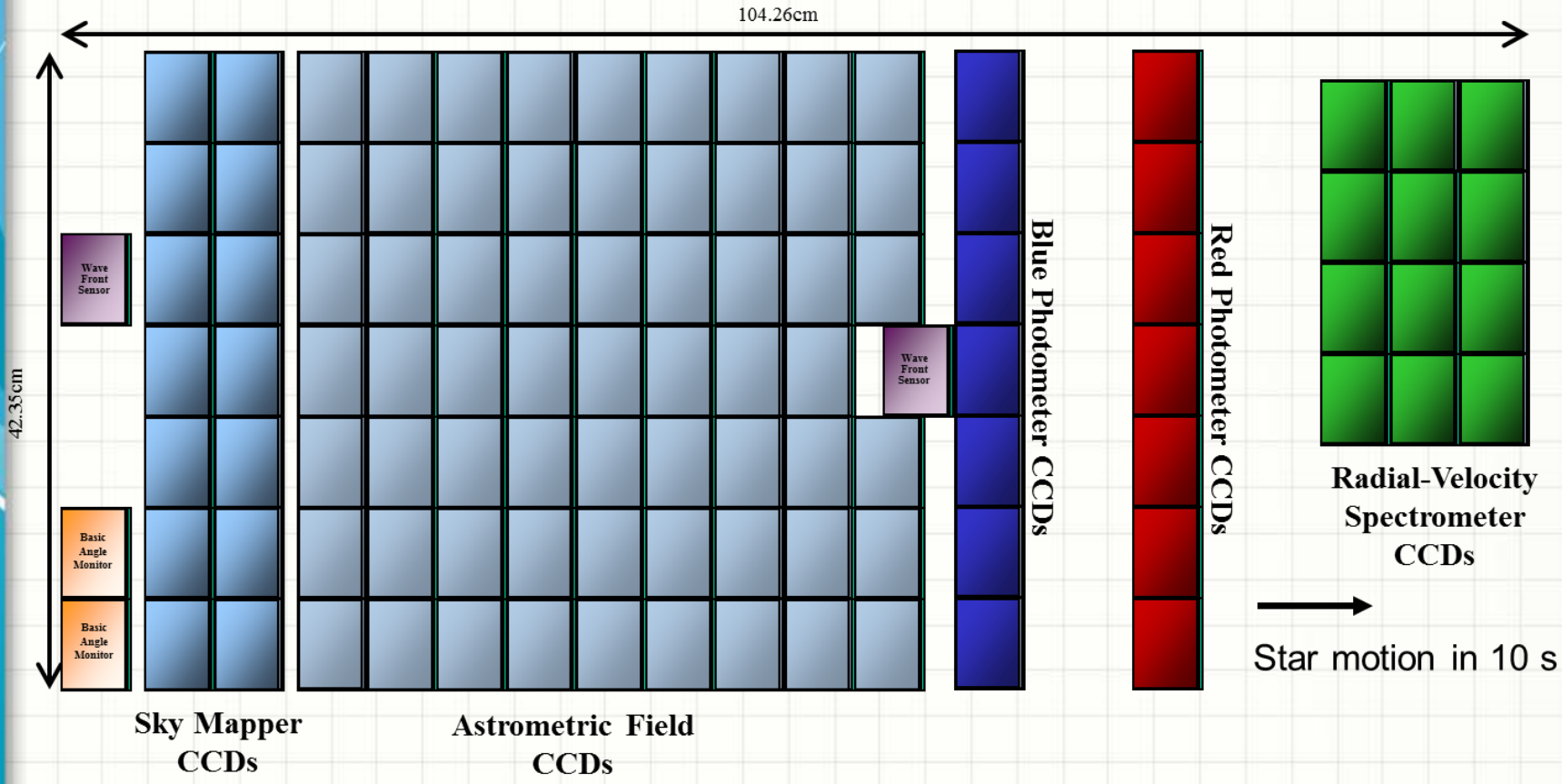
- Galaxies
- Quasars
- Gravitational Lenses



# **Is it possible to extract morphological information** from the **galaxies observed by the Gaia** satellite?

- Gaia is made to detect and analyse point-sources object
- Extracting morphological information from galaxies is quite a challenge

# Focal Plane




Credit: Alexander Short ESA

# Input Data

- Raw Astrometric Field (AF) and Sky Mapper (SM) samples
- Galaxies sources:
  - Sources not consistent with the presence of just point sources
  - Sources classified as galaxies
- Quasars sources:
  - Initial QSO Catalogue
  - Sources classified as quasars

# Output Data

**Galaxy**

- 2D Image
- Disc intensity, radius, ellipticity, position angle
- Bulge intensity, radius, ellipticity, position angle
- Boxiness, Discness
- Alpha/Delta Disc and Bulge
- Sersic Index
- Offset in RA of the source
- Offset in Declination of the source
- Central Profile Deviation
- Classification

**Quasar**

- 2D Image
- Quasar Intensity and radius
- Host galaxy intensity, radius, ellipticity, position angle
- Offset in RA/DEC of the host galaxy
- Offset in RA/DEC of the quasar
- Flag to indicate whether the quasar is multiple imaged
- Flag to identify the presence of a detectable host galaxy

# Image Reconstruction

# FastStack Algorithm (Harrison)

- Two dimensional image reconstruction method based on one dimensional raw samples

$$r_{ij} = \frac{\sum_k^{N_w} \theta_k w_k}{\sum_k^{N_w} w_k}$$

$$w_k = \begin{cases} \frac{1}{n_{s_k}} & k \leq \alpha N_w \\ \frac{1}{n_{s_k}} \exp\left(1 - \left(\frac{\theta_k}{r_{ij_0}}\right)^2\right) & k > \alpha N_w \end{cases}$$

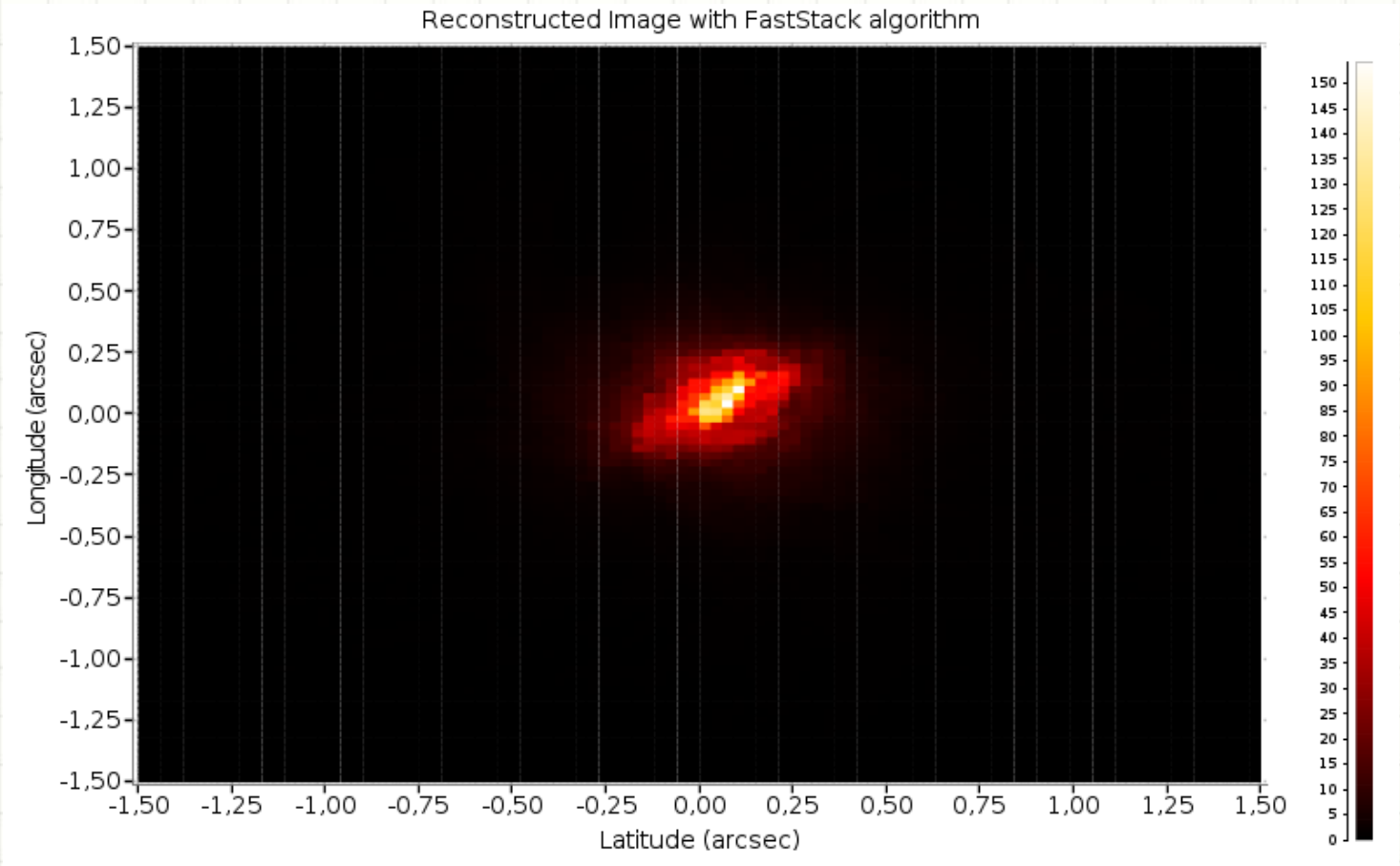
where θ is the array of window samples, *s*1, *s*2, *s*3...*sk*...$s_{N_w}$ which contribute to this pixel, $N_w$ is the total number of window samples which contribute to the *i j*$^{th}$ pixel, and $w_k$ is the weighting applied to the $k^{th}$ window sample

where $n_{sk}$ is the number of pixels in the reconstructed image to which the window sample, $s_k$, contributes, α is chosen such that 0 <α< 1,

**A window sample contributes to a pixel in the reconstructed image, if the coordinates of the centre of the pixel lie within the area of the window sample**

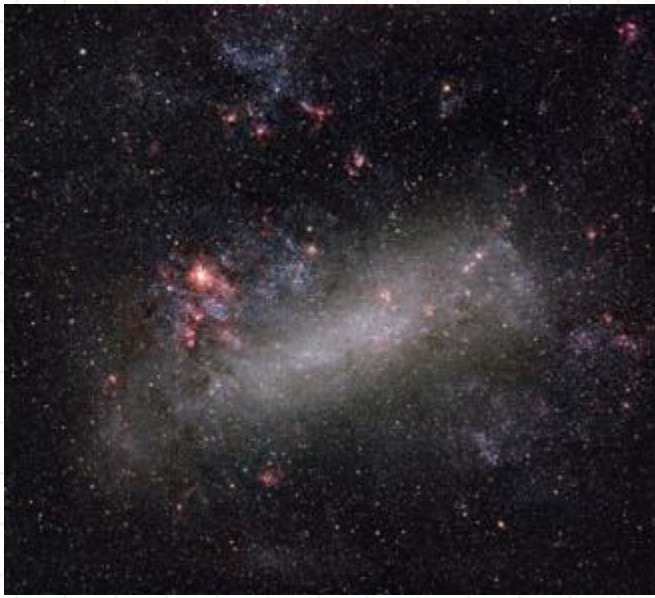
$$r_{ij_0} = \frac{\sum_k^{\alpha N_w} \theta_k w_k}{\sum_k^{\alpha N_w} w_k}$$

# Test on simulated data



# Galaxy Morphological Classification

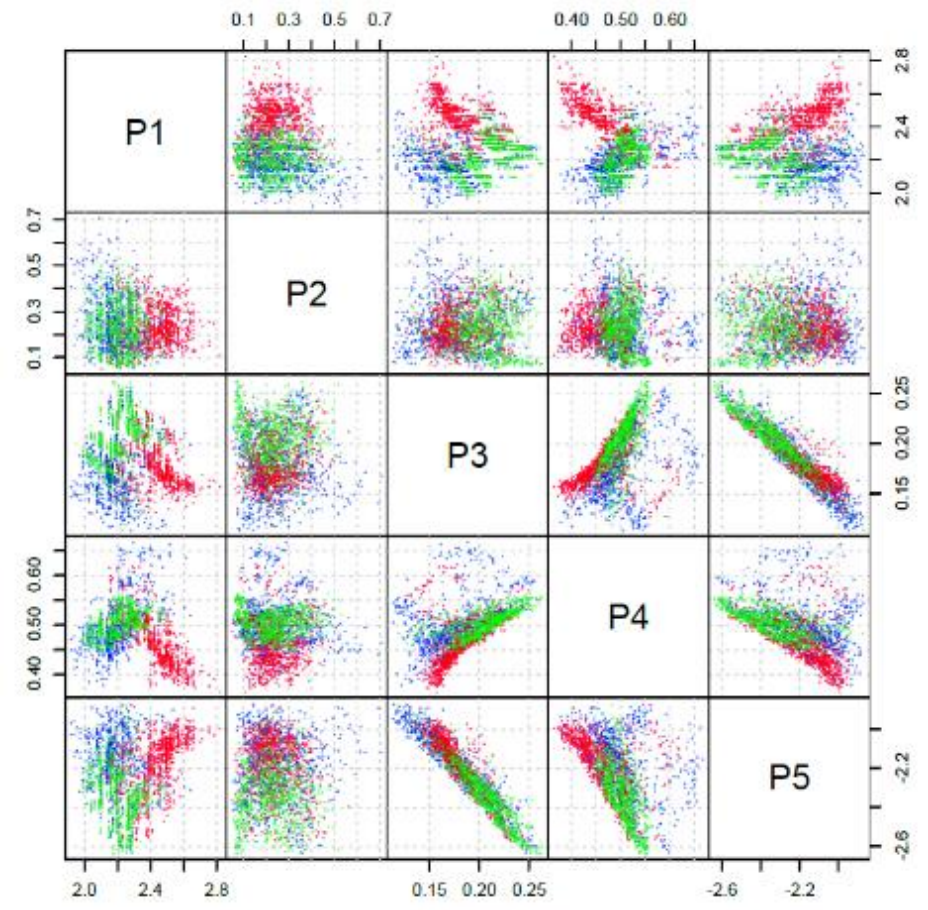
# Galaxy Types



Three galaxy models : Elliptical, Spherical, Irregular

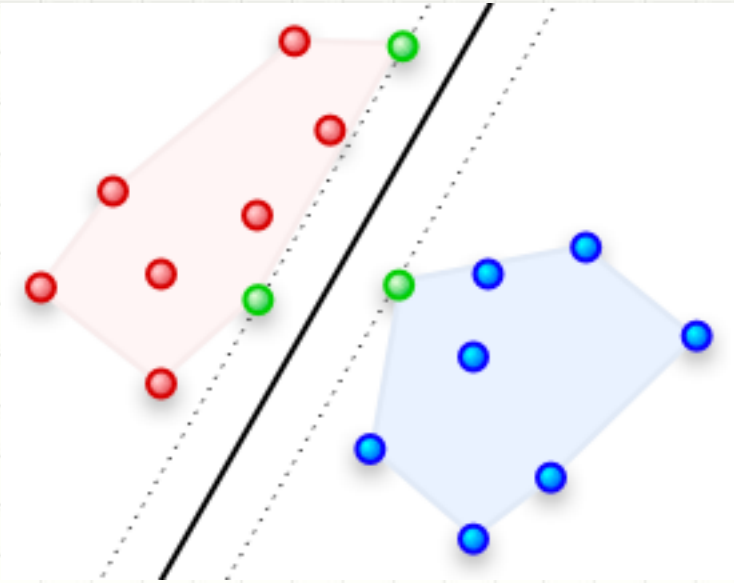
# CASGM20 Space (Lotz et al. 2004)

- P1: Light Concentration in the center of the object *Abraham et al. 1994*
- P2: Signal's asymmetry Abraham et al. 1996
- P3: Smoothness of the light distribution in the image *Conselice 2003*
- P4: Concentration of light in pixels independently of their position (Gini coefficient *Abraham et al. 2003* )
- P5: Second-order moment of the brightest pixels responsible to 20% of the flux image ($M_{20}$ )Lotz et al. 2004



# Support Vector Machines (SVM)

- Two Steps: Training and Classification *(Vapnik, 1993)*
- Goal of SVM: find the hyperplane separator that has the largest distance to the nearest training data point of any class



# Strategies of Training

- Use sources identified as galaxies by a GWP → Available later in the mission
- Fit a sample of data with all galaxy profiles and use the best fitted profiles → Time consuming

# Tests on simulations

- Use of a Java Version of LIBSVM
- Test on GIBIS Simulations

| Real class | Class. as E | Class. as S | Class. as I |
|---|---|---|---|
| E | $0.82 \pm 0.03$ | $0.07 \pm 0.02$ | $0.11 \pm 0.03$ |
| S | $0.06 \pm 0.02$ | $0.65 \pm 0.04$ | $0.29 \pm 0.04$ |
| I | $0.04 \pm 0.02$ | $0.19 \pm 0.03$ | $0.76 \pm 0.04$ |

- Test with very unbalanced training set also successful

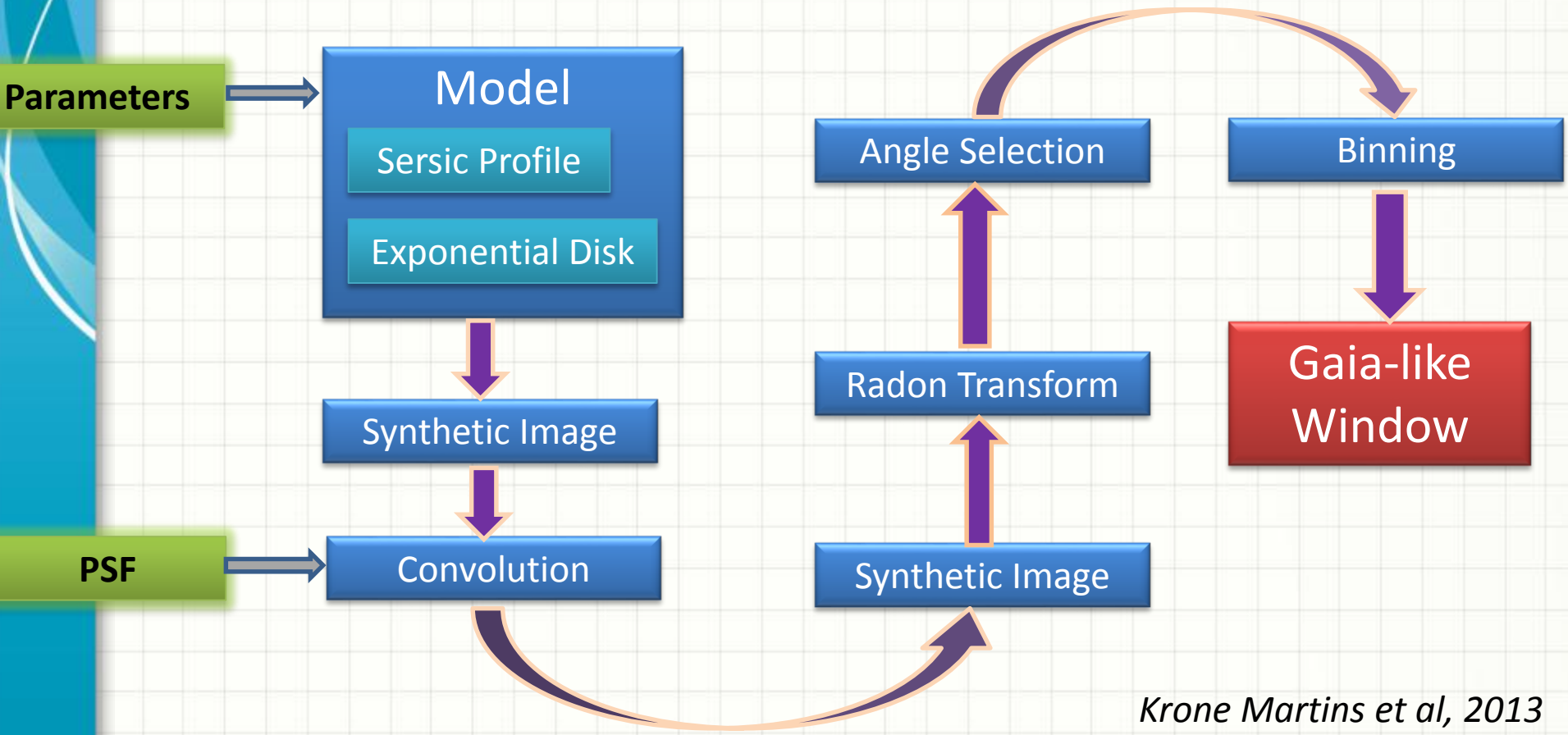
# Profile Fitting

# Profile Fitting

- Objective: Obtain physical parameters of the galaxies and quasars

- Important: the analysis should not be based on reconstructed images, but directly on the satellite window data

# Forward modeling




*Krone Martins et al, 2013*

# Optimization

- Criterion to optimize:

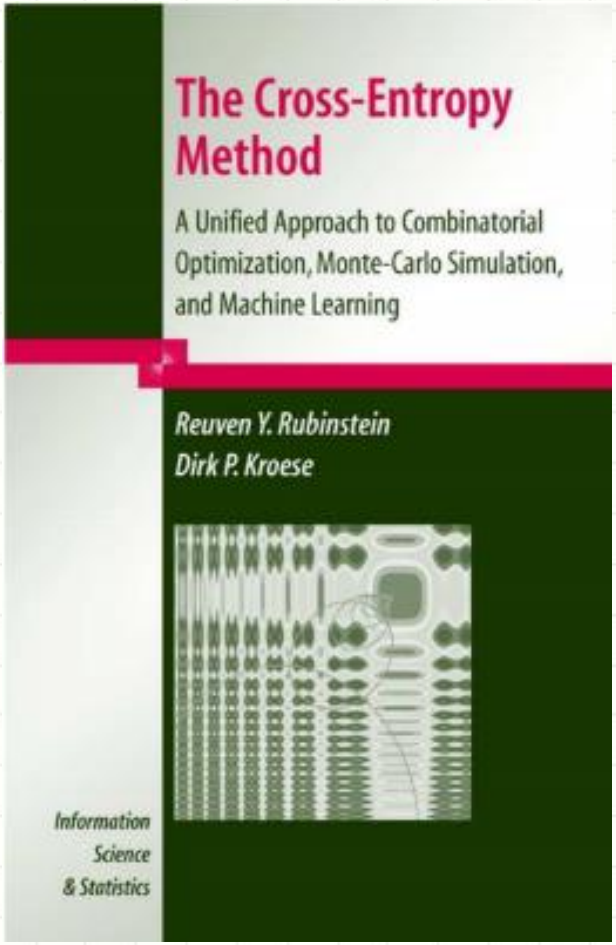
  MSE between the raw samples and the simulated ones

- Genetic Algorithms and BFGS → time consuming

- Strategy Proposed
  - Global Optimizer using just boundaries of search for parameters
  - Local Optimizer to refine the fitted parameters

# Cross Entropy Method (CEM)

- Developed as an estimation technique for rare-event probabilities in discrete systems (using Cross-Entropy or Kullback Leibler divergence as measure of closeness between two sampling distributions)
- Adapted for use in optimization



# Cross Entropy Method (CEM)

- Algorithm:
  1. Initialize $\mu_0$ and $\sigma^2$ . Set t = 1
  2. Generate a random sample $\theta_1, \ldots, \theta_N$ from $N(\mu_{t-1}, \sigma^2_{t-1})$
  3. Order the samples from smallest to biggest, and select the $N^{elite}$ best samples

     Let $\hat{\mu}_t$ be the sample mean (vector) of the elite samples

     Let $\hat{\sigma}^2_t$ be the vector of variances of the elite samples
  4. Update $\mu$ and $\sigma^2$ in a smooth way

     $$\mu_t = \alpha\, \hat{\mu}_t + (1-\alpha)\, \mu_{t-1}$$
     $$\sigma^2_t = \alpha\, \hat{\sigma}^2_t + (1-\alpha)\, \sigma^2_{t-1}$$
  5. Stop at iteration t = T if some stopping criterion is met. Otherwise increase t by 1 and return to step 2

***Dynamic Smoothing***

*(Rubinstein and Kroese, 2004)*

$$\beta_t = \beta - \beta\left(1 - \frac{1}{t}\right)^q$$

*To prevent early stop and local minima*

***Variance Injection***

*(Botev and Kroese, 2004)*

Add a chaos perturbation in the variance

# Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

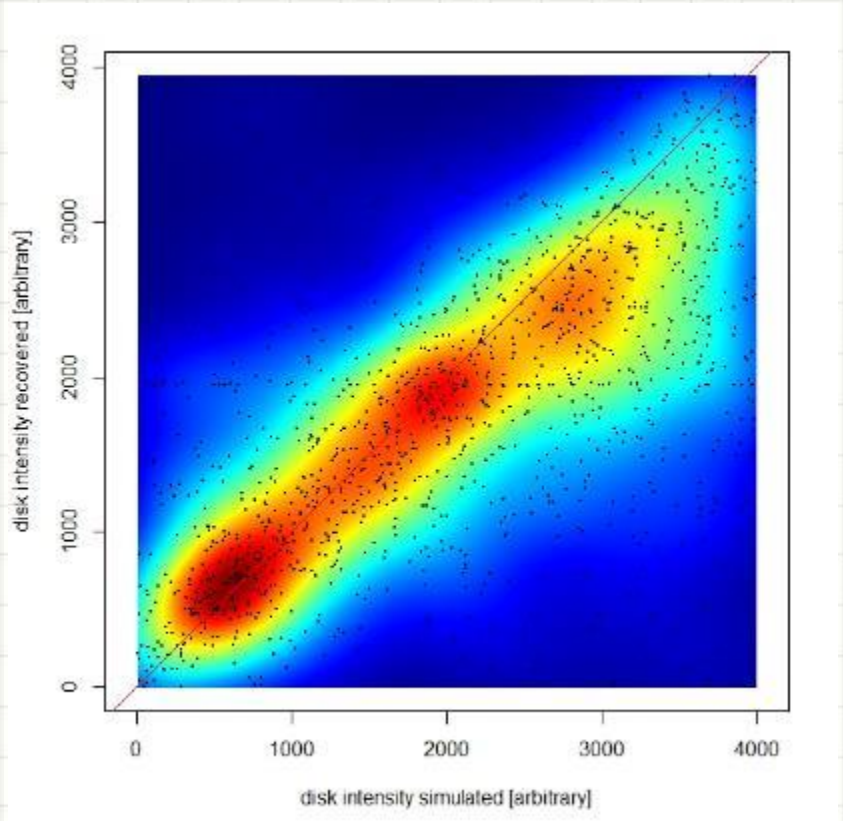
- Standard method for continuous black-box evolutionary optimization (always on the top ranking for the Congress on Evolutionary computation contest) *Hansen et al. 2003*
- Optimize an objective function f(x) by sampling λ candidate solutions from a multivariate normal distribution
- The best k solutions of the λ ones are used to adaptively estimate the local covariance matrix of the objective function
- Advantage: almost parameter free (only population size λ)

***IPOP CMAES*** *(Auger and Hansen, 2005)*
increase population size dynamically

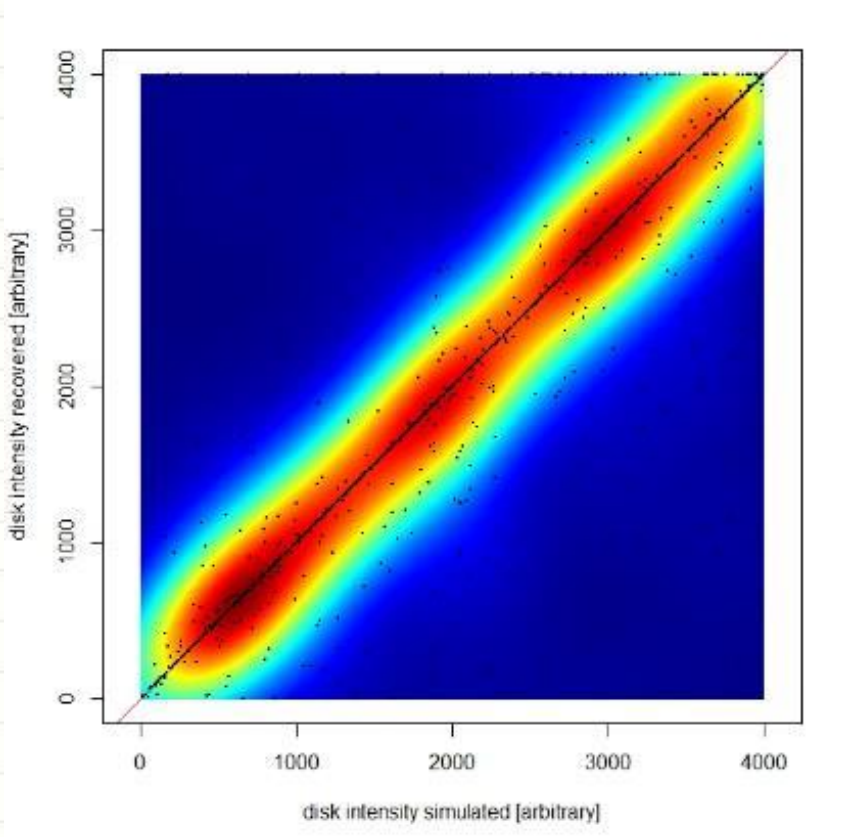
# Test on Simulated Data

# Galaxy profile fitting

**Disk Intensity CEM**

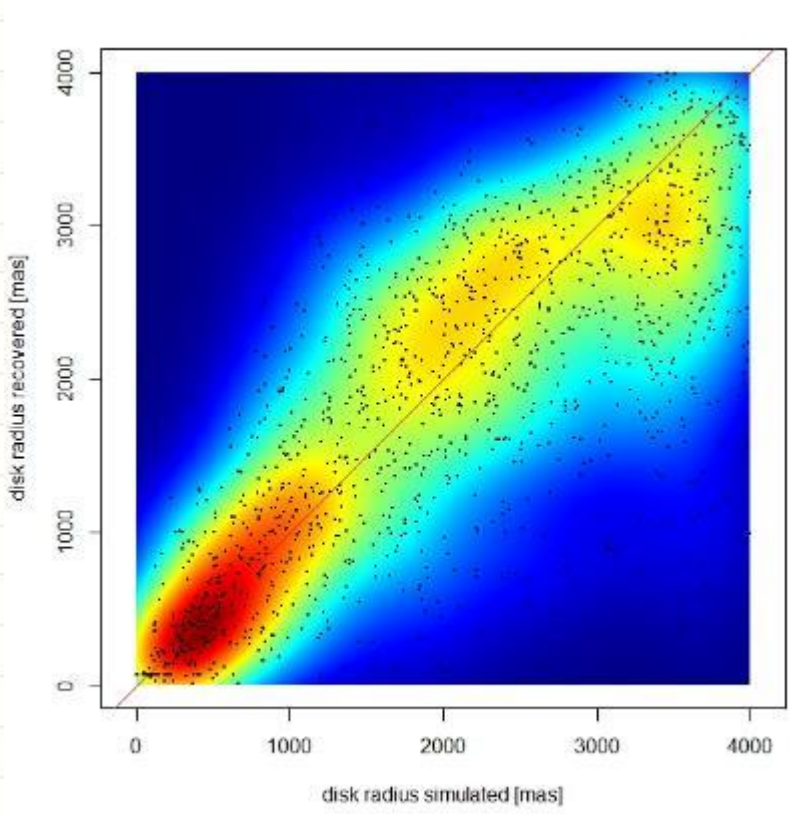



**Disk Intensity CMAES**

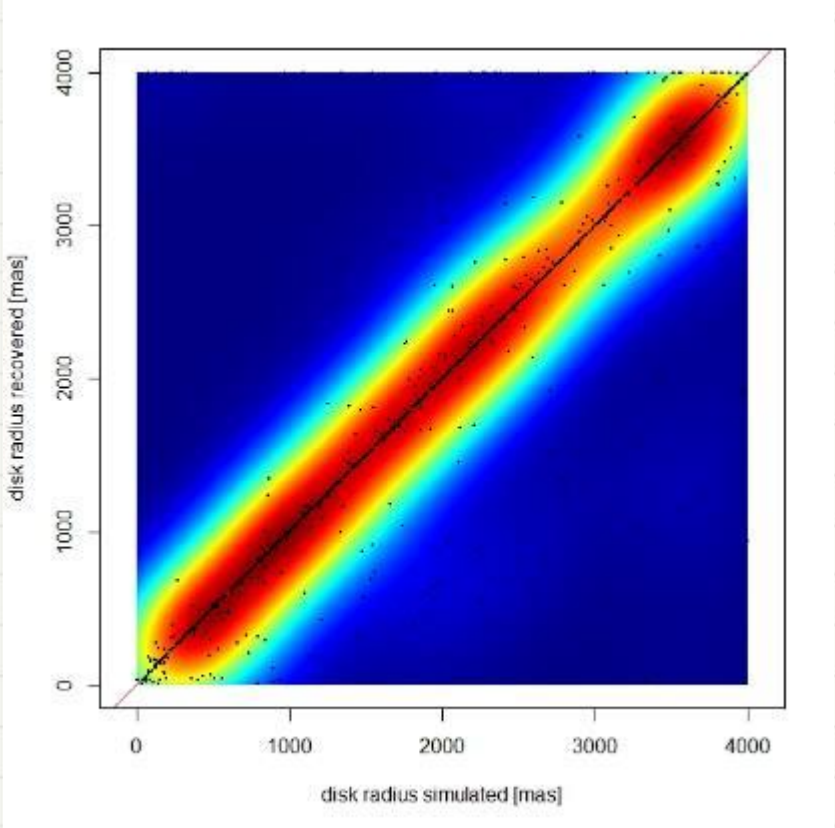


# Galaxy profile fitting

**Disk Radius CEM**

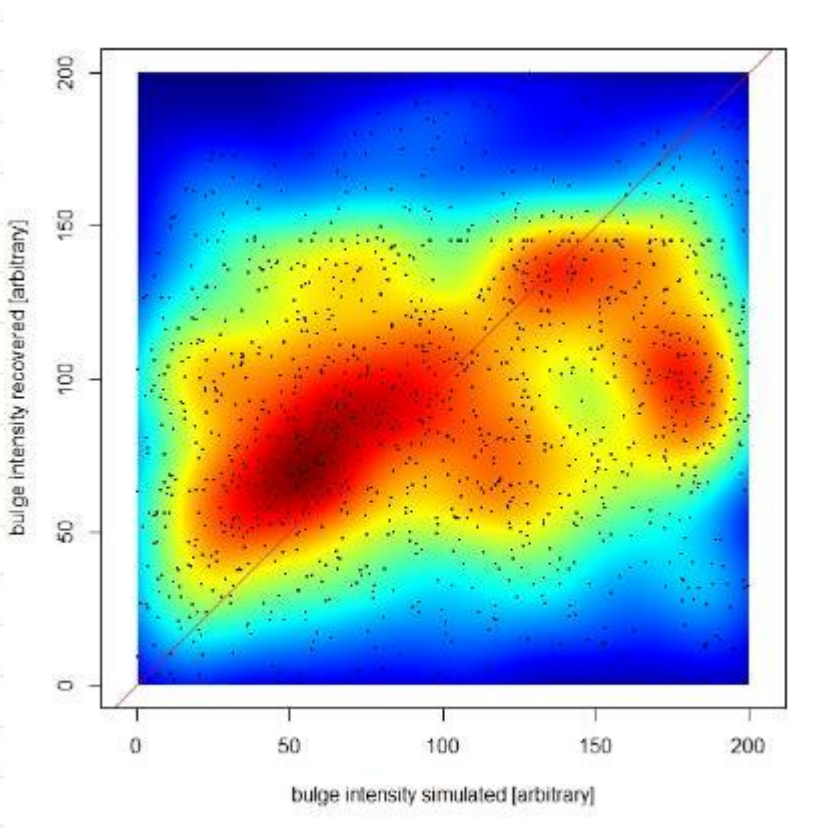



**Disk Radius CMAES**

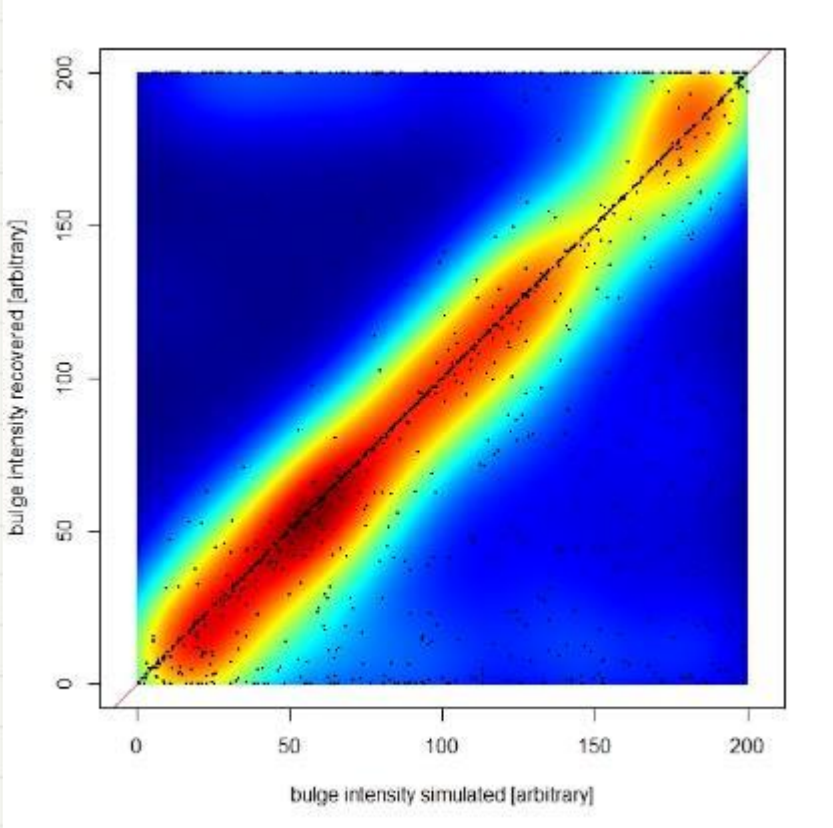


# Galaxy profile fitting

**Bulge Intensity CEM**

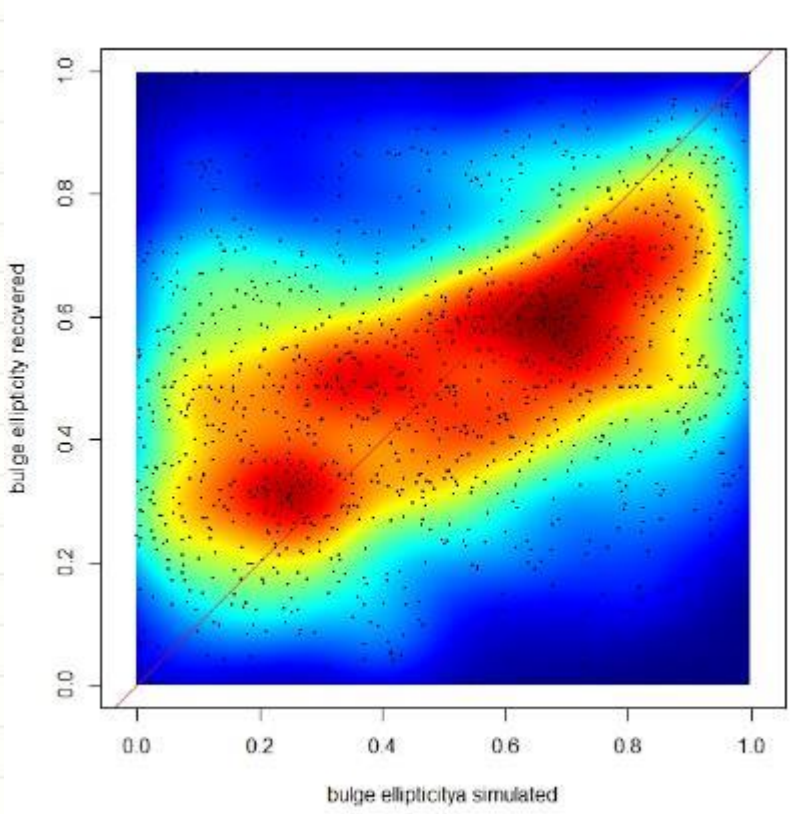



**Bulge Intensity CMAES**

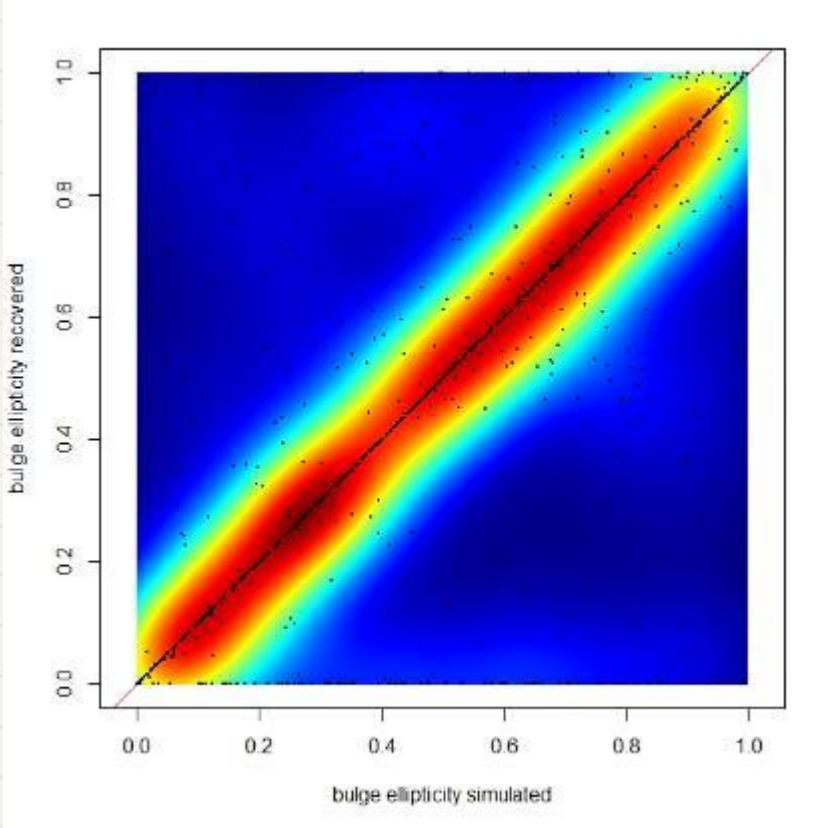


# Galaxy profile fitting

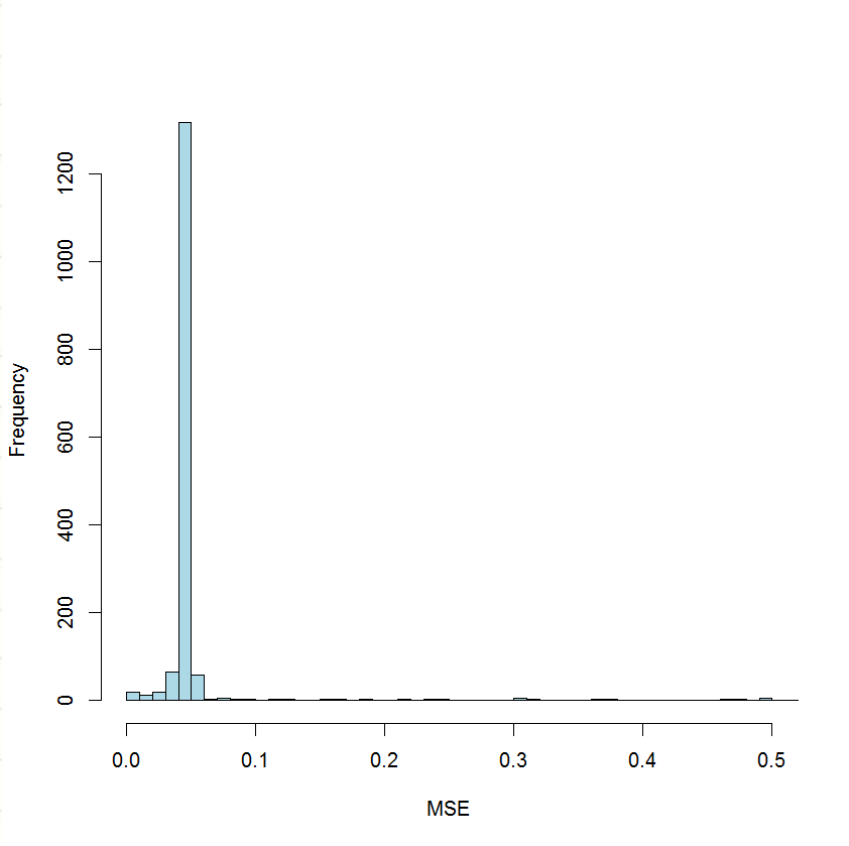
**Bulge Ellipticity CEM**




**Bulge Ellipticity CMAES**

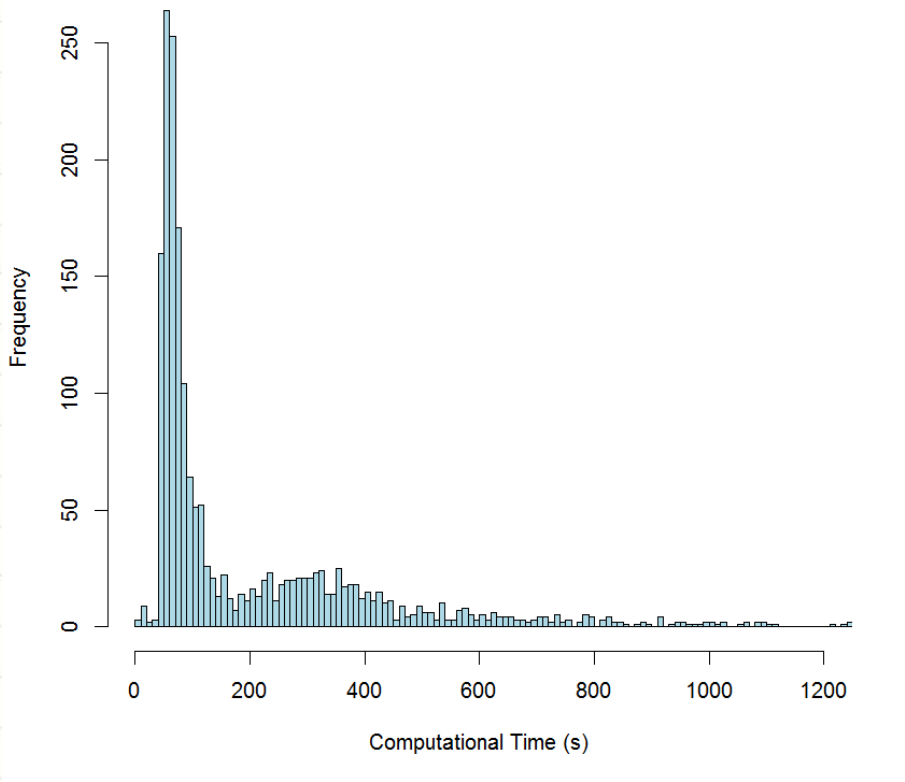


# Quality of Results



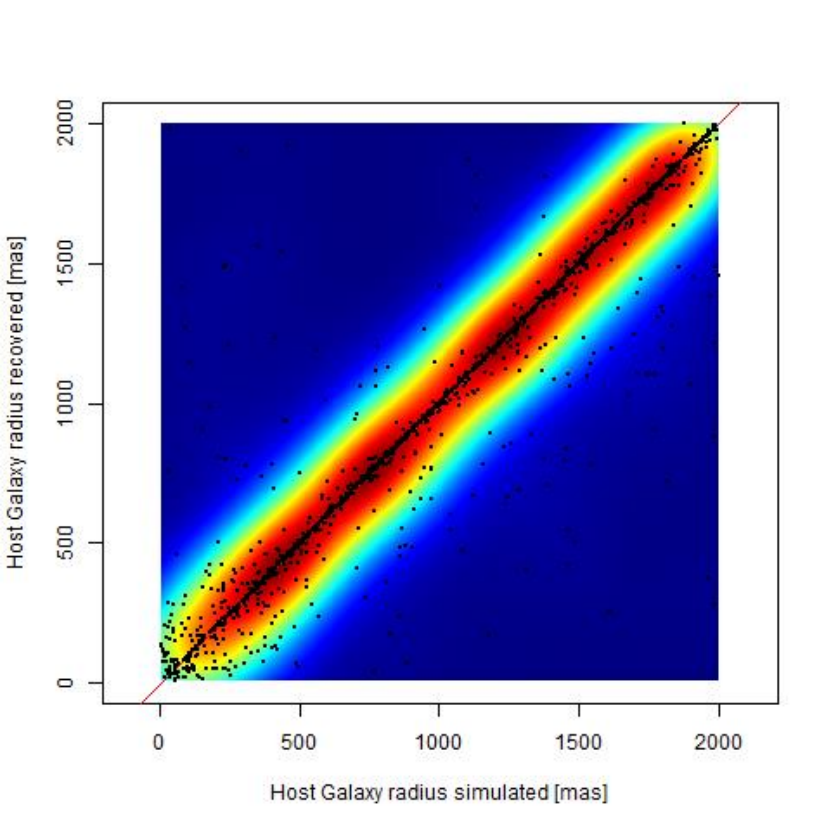

- Bulge Intensity: -2e-04 +/- 0.0527
- Bulge radius: 0.0015 +/- 0.0212
- Bulge ellipticity: 0 +/- 0.0014
- Bulge position angle : 0 +/- 0.0012
- Alpha bulge: 0 +/- 0.0051 (rad)
- Delta bulge: 0 +/- 0.001 (rad)
- Alpha disk: 0 +/- 4e-04 (rad)
- Delta disk: 0 +/- 1e-04 (rad)
- Sersic index: -0.0054 +/- 0.019
- Boxiness: -2e-04 +/- 0.0388
- Disk intensity: 0 +/- 4e-04
- Disk Radius: 0 +/- 1e-04
- Disk Ellipticity: 0 +/- 2e-04
- Disk position angle: 0 +/- 1e-04 (rad)
- Diskness: 0 +/- 0.0014

# Computational Time

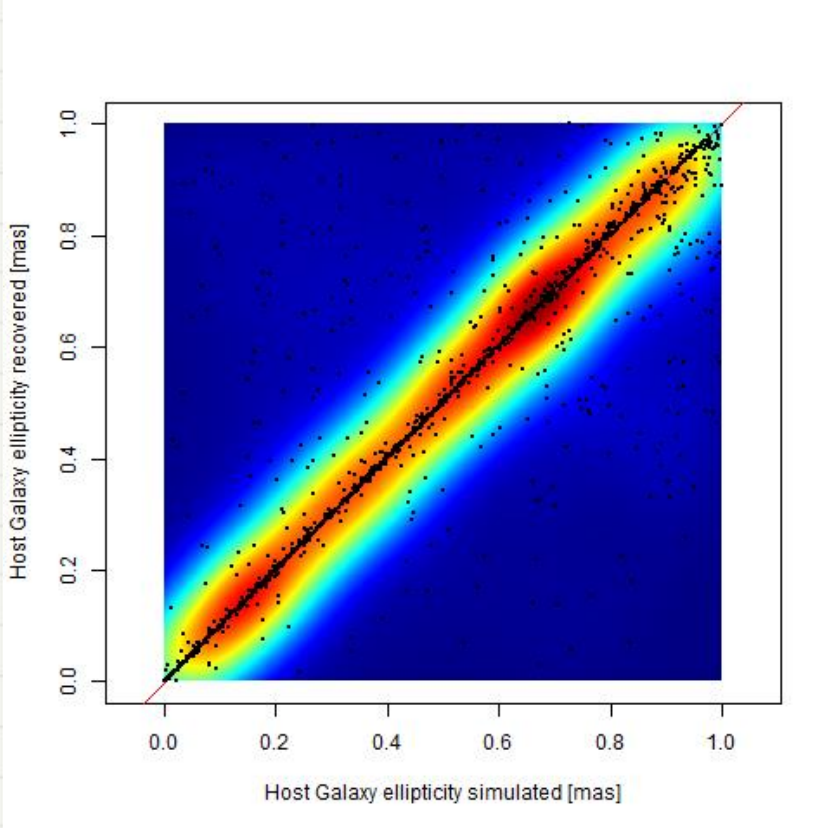


# Quasar Profile Fitting

**Radius Host Galaxy**

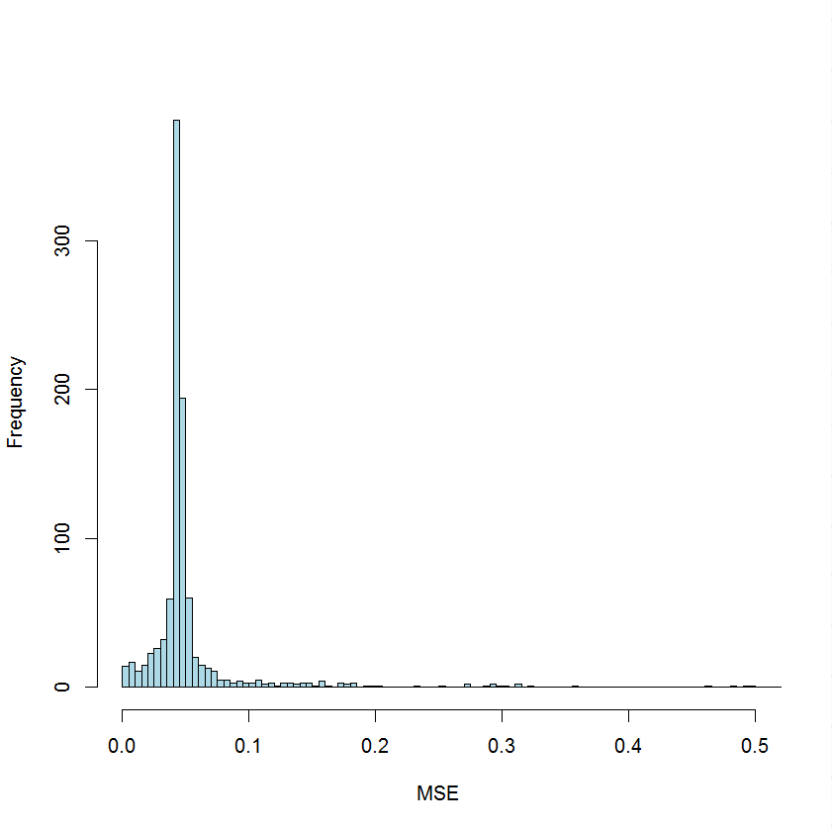



**Ellipticity Host Galaxy**

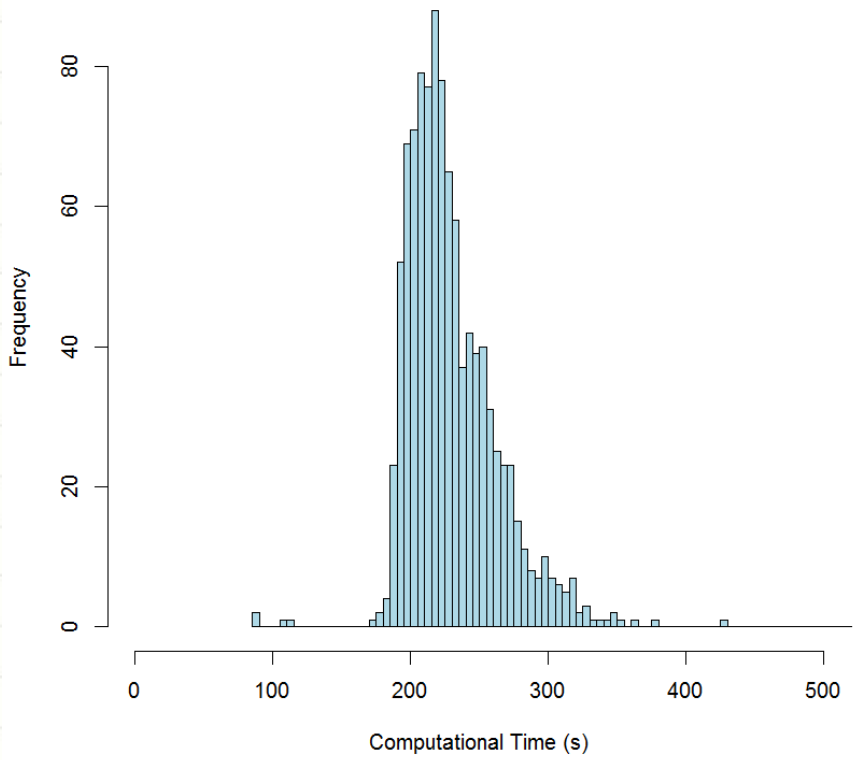


# Quasar Profile Fitting

- Quasar Intensity: -0.0164 +/- 0.4627
- Quasar radius: -0.0632 +/- 0.6404
- Host Galaxy intensity: -2e-04 +/- 0.0628
- Host Galaxy radius: -3e-04 +/- 0.0067
- Host Galaxy ellipticity: 0.0011 +/- 0.0095
- dAlpha: 0.0046 +/- 0.1326
- dDelta: 0.01 +/- 0.1404
- Offset RA Host Galaxy: -0.0016 +/- 0.0444
- Offset DEC Host Galaxy: 7e-04 +/- 0.0072
- Sersic index: -0.0081 +/- 0.3154



# Computational Time



# Conclusions

- Two operational chains : Galaxies and Quasars … but not fully tested with real data
- One on-going module to flag Gravitational lenses processing

- When we will provide these results inside Gaia catalogue ?

… last release, until we will have enough data

# THANKS FOR YOUR ATTENTION