Space variant deconvolution of galaxy survey images

Jean-Luc Starck

http://jstarck.cosmostat.org

CEA, AIM/Service d'Astrophysique, France

Morgan Schmitz

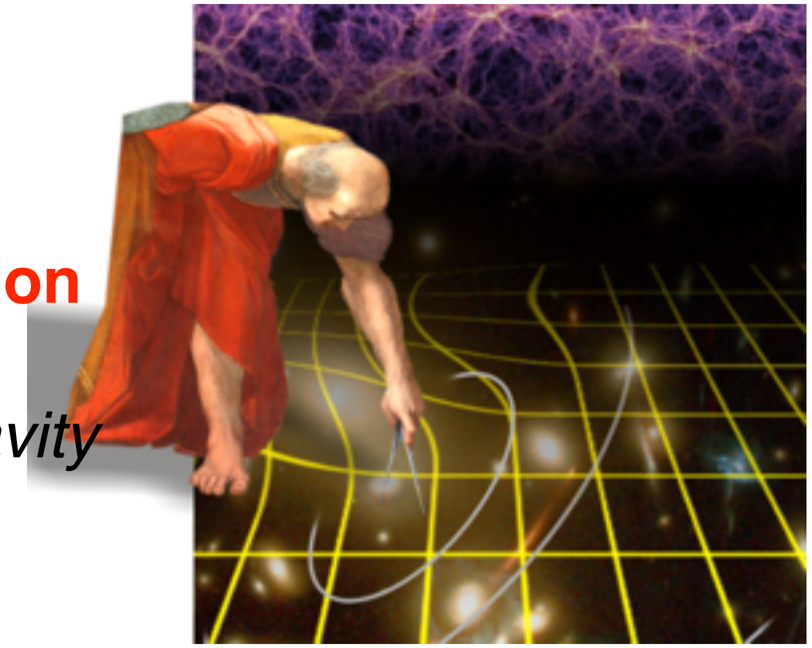
Sam Farrens

Fred Nole

Euclid Space Mission

Euclid, ESA Cosmic Vision: launch in 2020:

- 1390 members , 220 laboratories, 16 countries



Understand the origin of the Universe's accelerating expansion

- probe the properties and nature of *dark energy, dark matter, gravity* and distinguish their effects decisively
- by tracking their observational signatures on the
- geometry of the universe: **Weak Lensing** + Galaxy Clustering
- cosmic history of structure formation: **WL**, z-space distortion, clusters of galaxies

Gains in space:

Stable data: homogeneous data set over the whole sky

- **Systematics** are small, understood and controlled
- Homogeneity : Selection function perfectly controlled

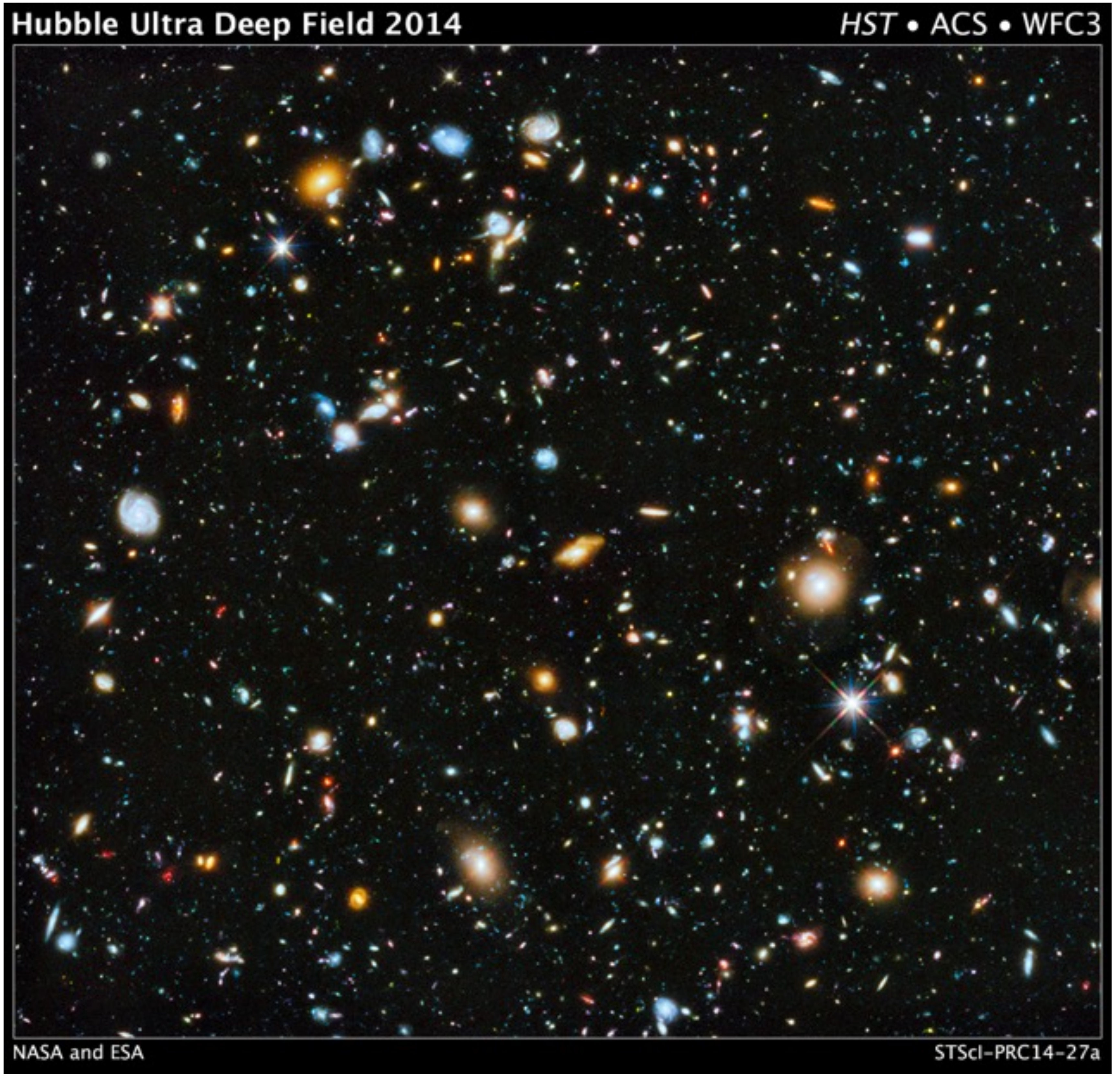
- Observe **15 000 deg2** during 6 years in optical and near infrared wavelength (shape)
- 1,2m Telescope, 4 bands
- Photometric redshift (distance) for **1 000 000 000 galaxies.**
- Spectroscopic IR measurement of **50 000 000 galaxies.**

==> 850 Gbits of data per day.

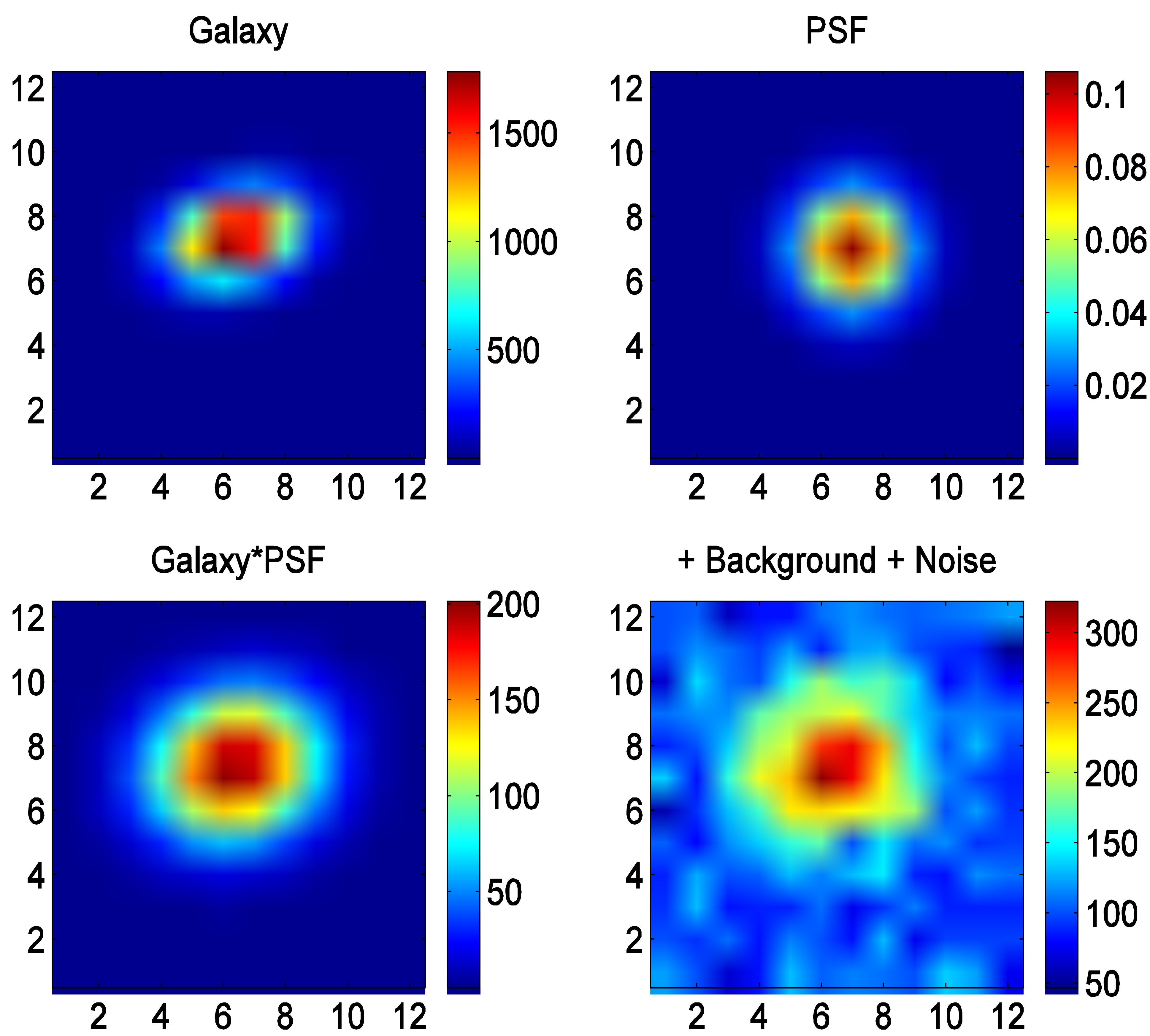
~150 PB of data in the archive .

http://www.euclid-ec.org/

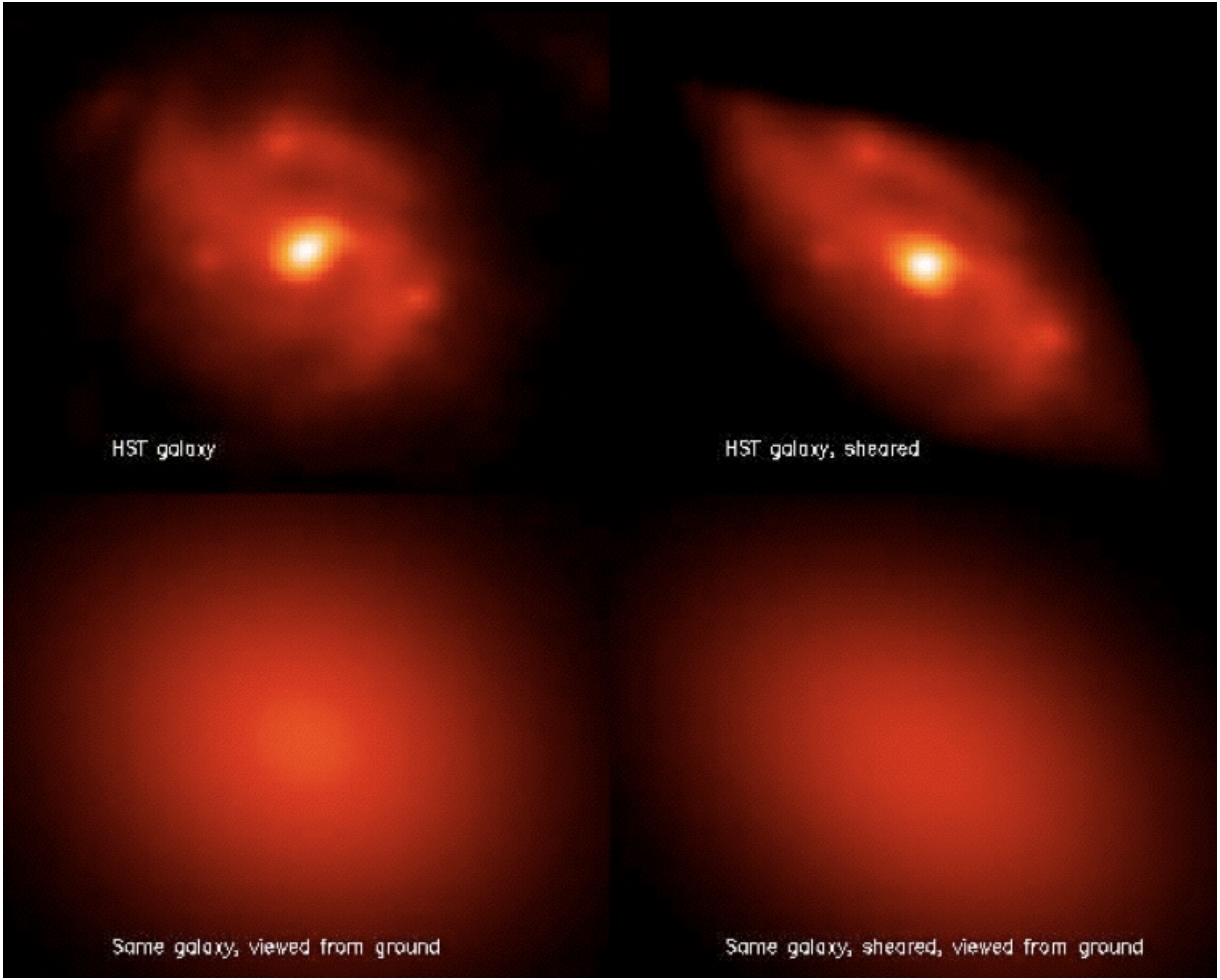
Galaxies


Hubble Ultra Deep Field 2014
HST • ACS • WFC3
NASA and ESA
STScI-PRC14-27a

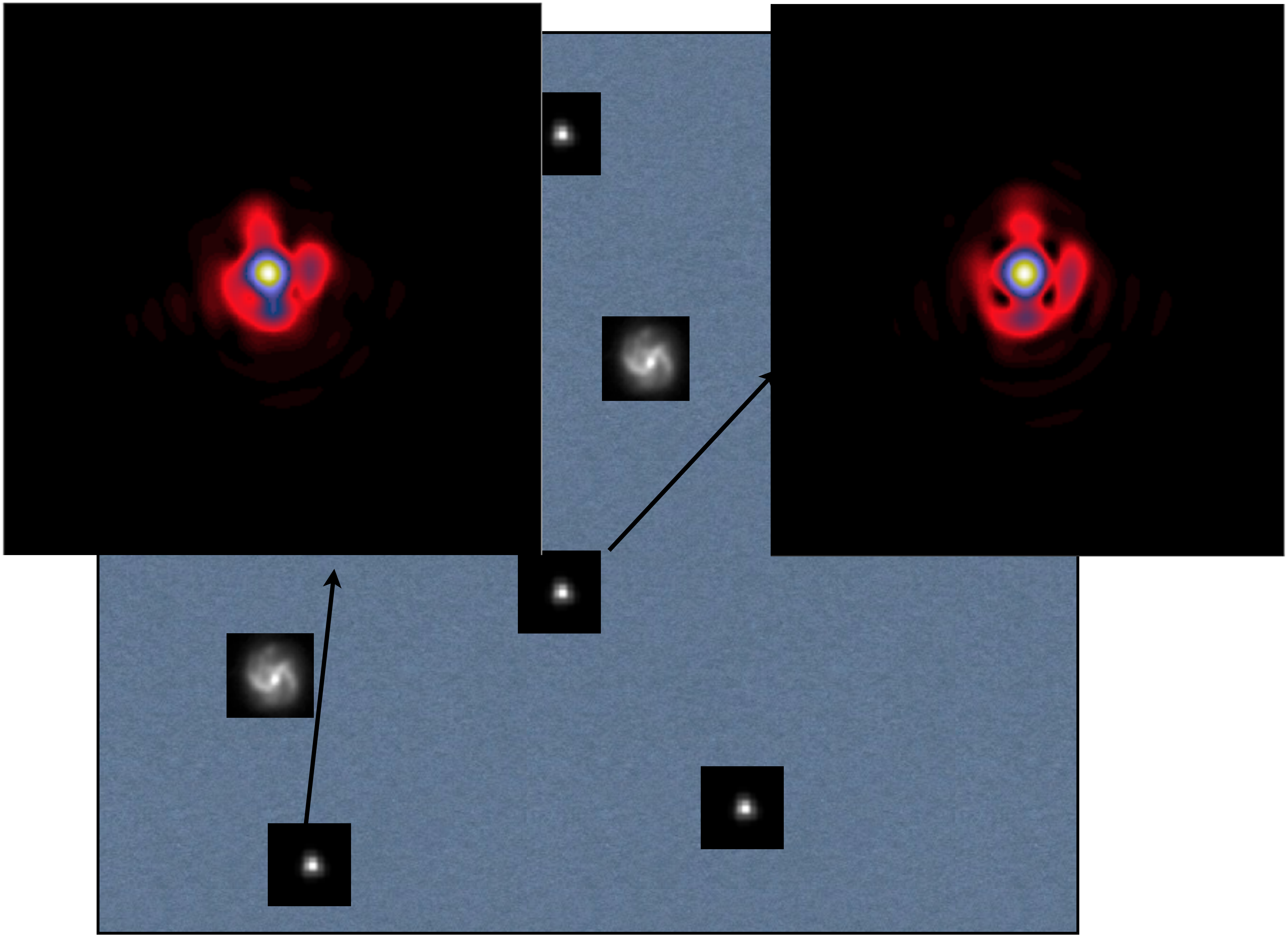
Weakly Lensed Galaxies



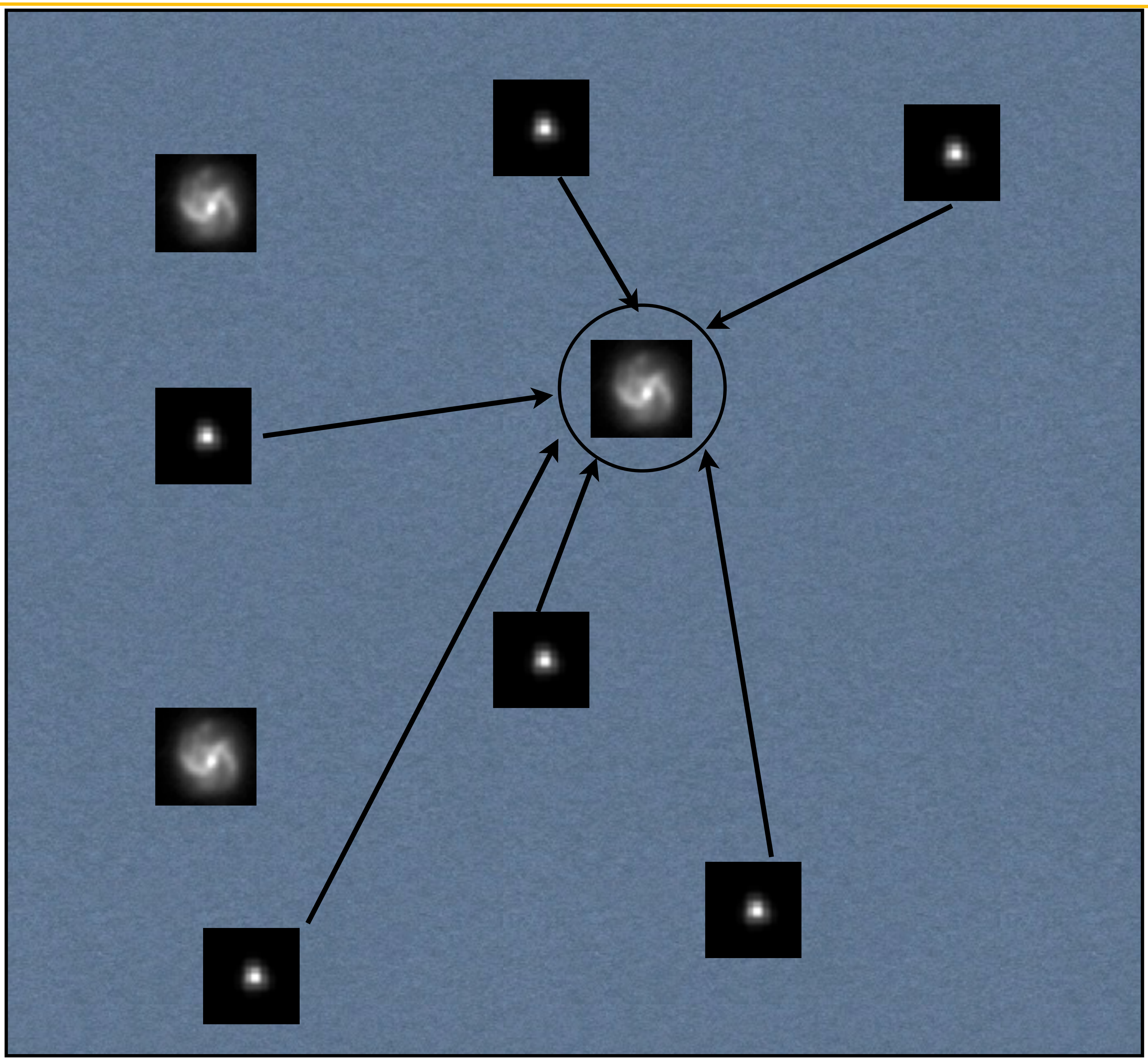
Motivation for spatial observations



Space Variant PSF

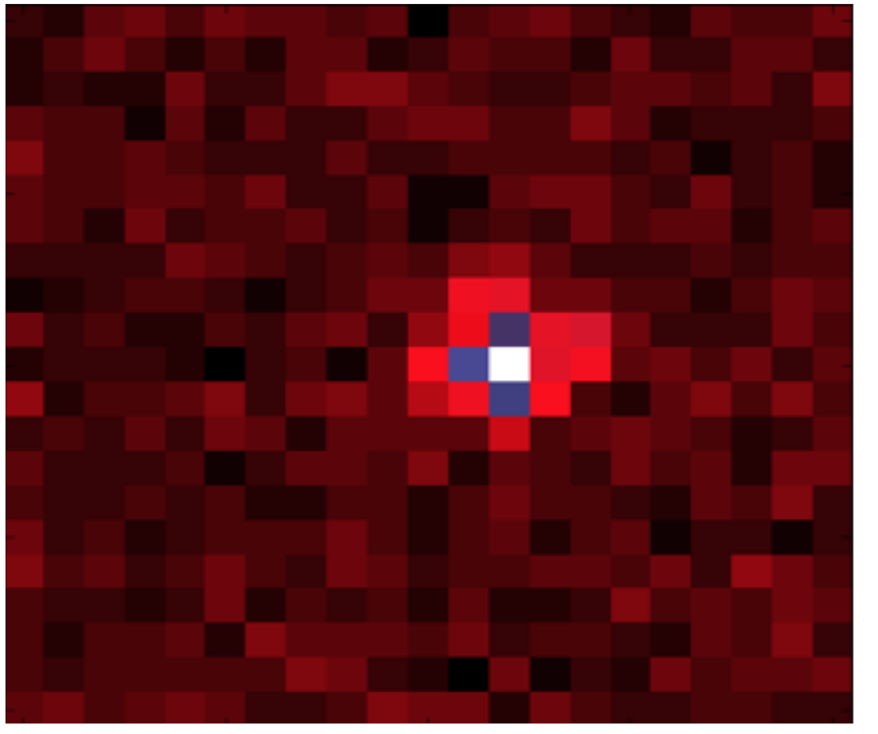


PSF Field Estimation



Euclid PSF Modeling

- PSF Modeling
  - ➡Undersampling
  - ➡Space dependency
  - ➡Time dependency
  - ➡Wavelength dependency



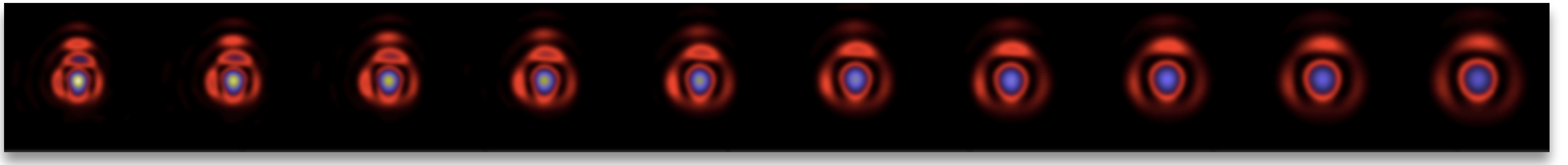
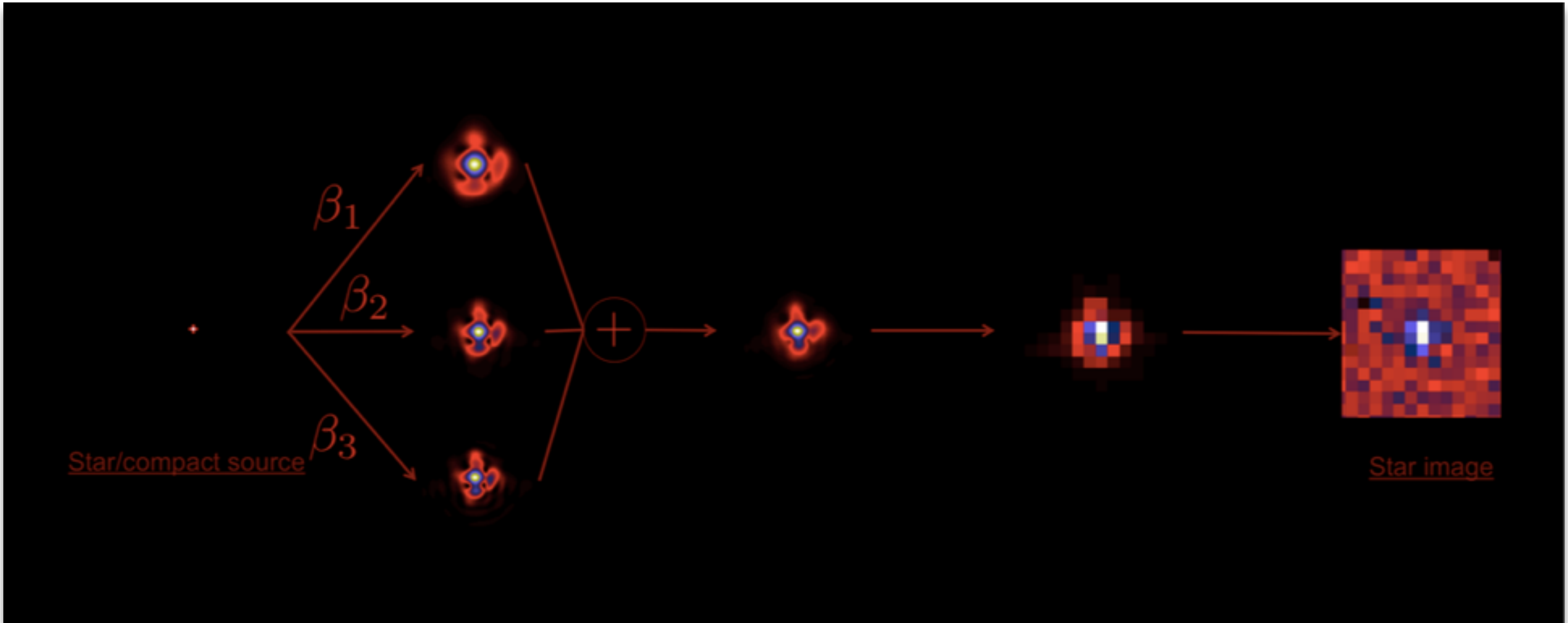


Image Forming Process: Stars and Point Spread Function



$$\mathbf{y}_k = \mathrm{M}_k \mathbf{x}_k + \mathbf{n}_k\,, \;\; k = 1 \cdots p$$

$\mathbf{y}_k$: k[th] low resolution image
$\mathbf{M}_k$: **shift and downsampling** operator
$\mathbf{x}_k$: k[th] well resolved image
$\mathbf{n}_k$: gaussian noise

PSFextractor [E. Bertin, 2011]

$$\tilde{y}_{k,ij} = \sum_{l} \sum_{m} h(l - N(i - i_k), m - N(j - j_k)) x_{lm}$$

where h is Lanczos interpolation kernel

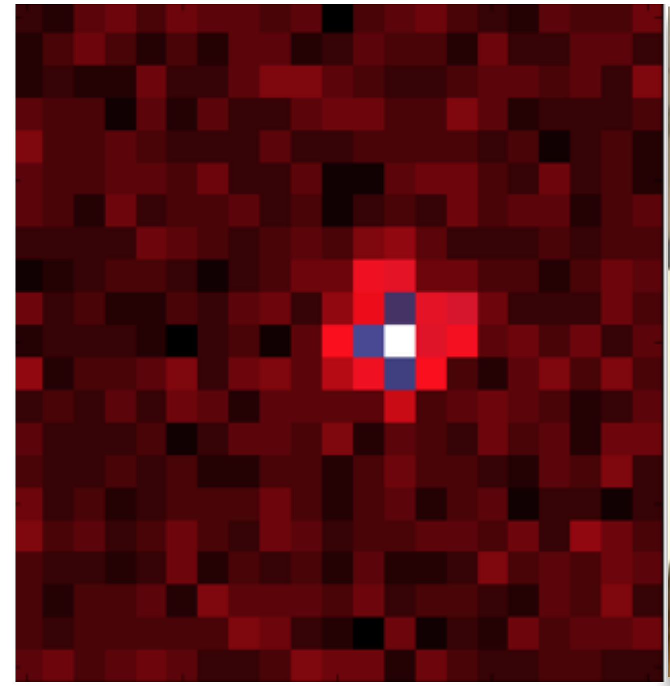
$$h(x) = \begin{cases} 1 & \text{if } x = 0 \\ sinc(x) sinc(\frac{x}{4}) & \text{if } 0 < |x| < 4 \\ 0 & \text{else} \end{cases}$$

Regularization

$$J(X) = \sum_{k=1}^{p} \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{(y_{k,ij} - f_k \tilde{y}_{k,ij})^2}{\sigma_k^2} + \frac{\left\| X - X^{(0)} \right\|_{l_2}^2}{\sigma}$$

Monochromatique PSF Field Estimation

- PSF Modeling
  - ➡**Undersampling**
  - ➡**Space dependency**
  - ➡Time dependency
  - ➡Wavelength dependency



What has been done:

Undersampling + space dependency = **Monochromatic PSF field restoration**

PSF Estimation

- **F. M. Mboula, J.-L. Starck, S. Ronayette, K. Okumura, and J. Amiaux, [Super-resolution method using sparse regularization for point-spread function recovery]. A&A, 575, id.A86, 2015.**
- **F. Ngole, J.-L Starck, et al, "Constraint matrix factorization for space variant PSFs field restoration", submitted, 2016.**

PSF Interpolatin

- **F. Ngolè and J.-L. Starck, "PSFs field learning based on Optimal Transport distances", submitted, 2017**

http://jstarck.cosmostat.org/wp-content/uploads/2014/09/these_ngole_mboula.pdf

$$\mathbf{y}_k = \mathbf{M}_k \mathbf{x}_k + \mathbf{n}_k,\ \ k = 1..n$$

- **Learning** technique to identify the ***eigenvectors*** PSF (note that PCA cannot handle subsampled data): **Resolved Components Analysis (RCA)**

  F. Ngole, J.-L Starck, et al, "Constraint matrix factorization for space variant PSFs field restoration", in press, 2017

Our model:

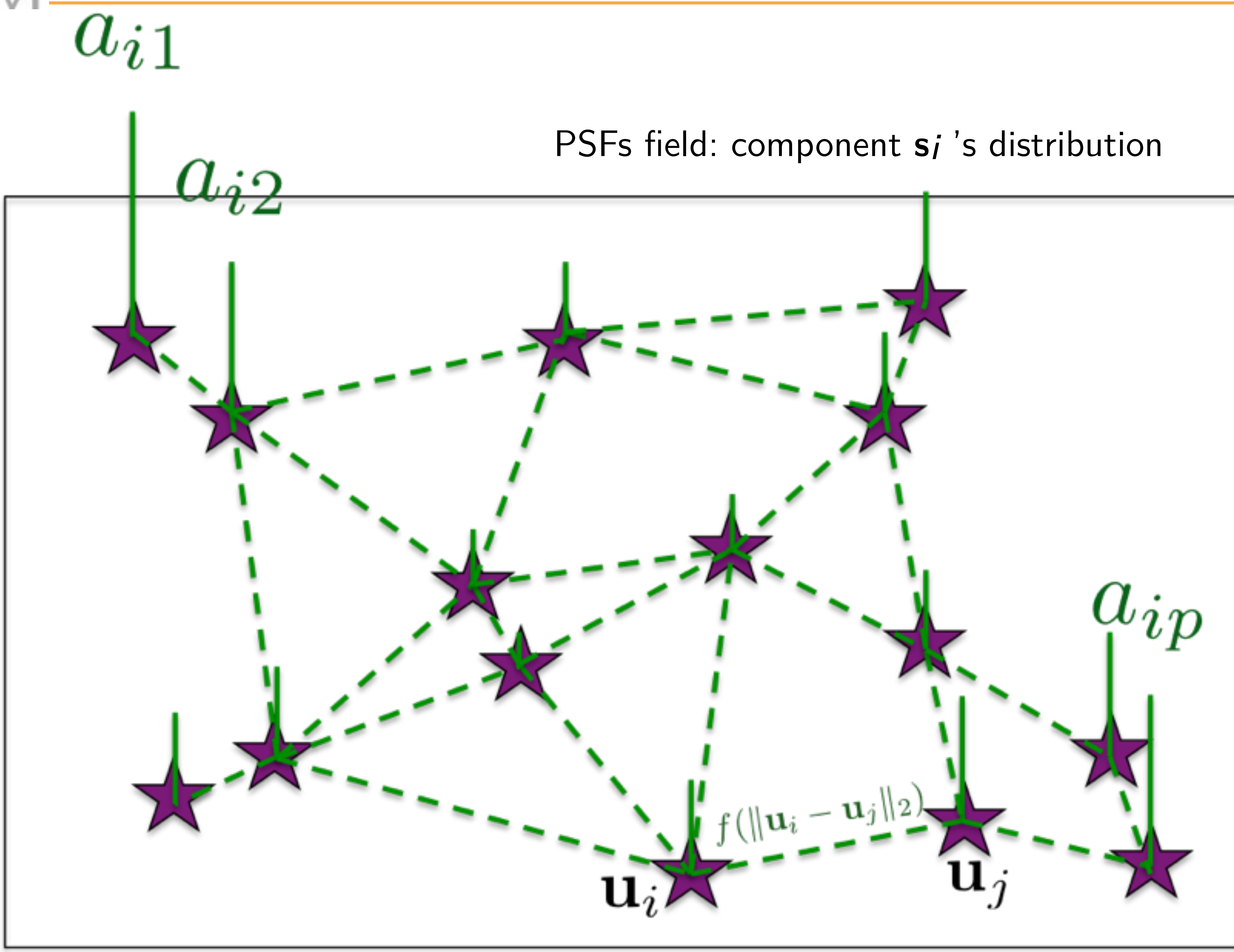
$$\mathbf{PSF}^{(k)} = x_k = \sum_{i=1}^{r} a_{i,k} s_i$$

$a_{i,k}$= coefficient corresponding to contribution of the i-th vector to the k-th PSF.

s_i = ith *vector* (2D image)

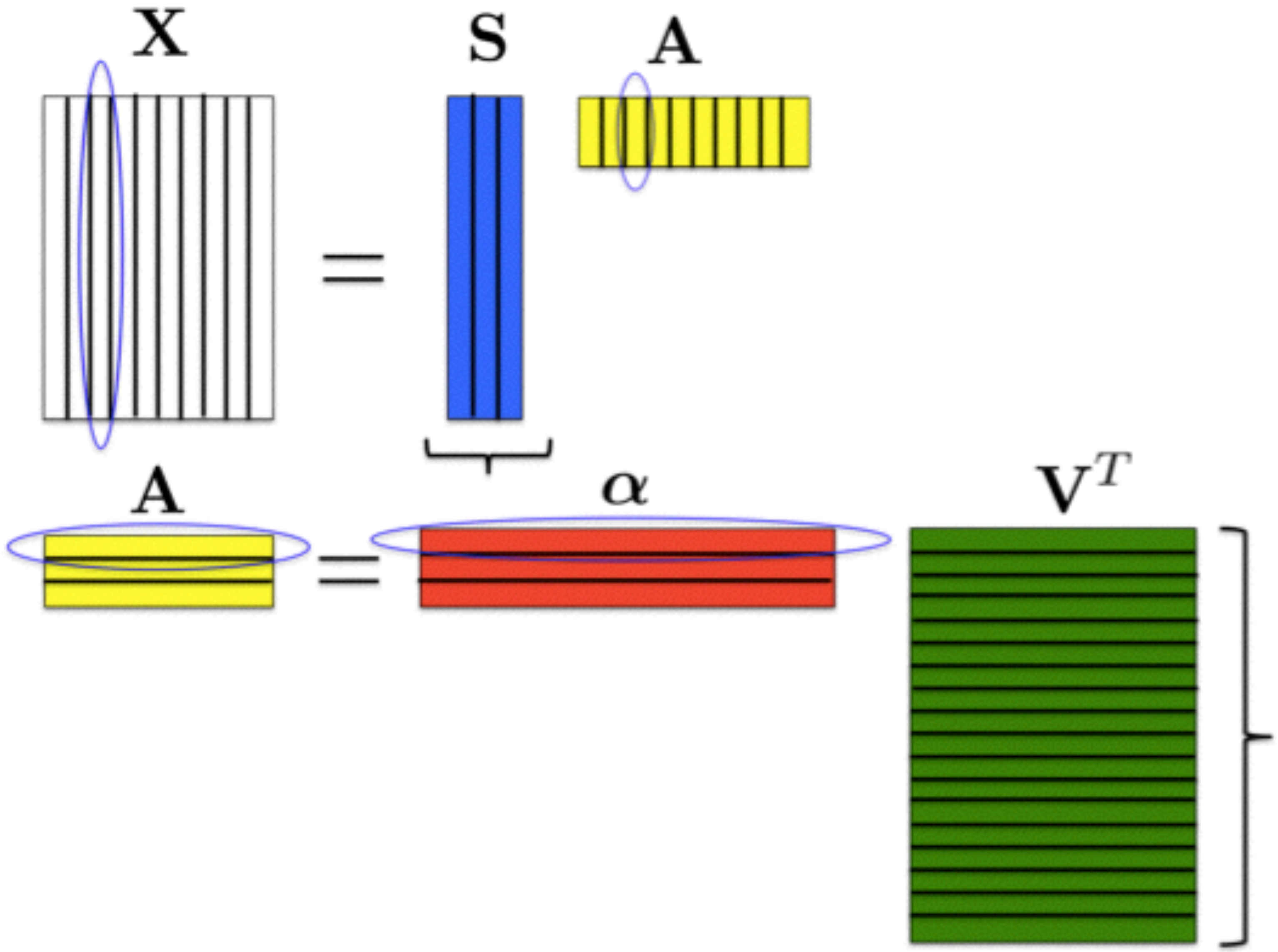
- **Joint estimations** of super-resolved PSFs at stars positions
  - ➡ Positivity constraint
  - ➡ Low rank constraint: Constraint the PSFs to be a linear combination of the *eigenvectors* PSFs
  - ➡ Smoothness constraint on each s_i
  - ➡ Proximity constraint on the $a_{i,k}$ coefficients: the closer are the stars, the more the coefficients of the linear combination are similar.

PSF Graph



a_{i1}
a_{i2}
PSFs field: component $\mathbf{s}_i$'s distribution
a_{ip}
$f(\|\mathbf{u}_i - \mathbf{u}_j\|_2)$
$\mathbf{u}_i$
$\mathbf{u}_j$

Matrix Factorization



$$\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_p]$$

$$\mathbf{PSF}^{(k)} = x_k = \sum_{i=1}^{r} a_{i,k} s_i$$

s_i are "eigen PSF"

$$\min_{\boldsymbol{\alpha}, \mathbf{S}} \frac{1}{2} \|\mathbf{Y} - \mathcal{F}(\mathbf{S}\boldsymbol{\alpha}\mathbf{V}^T)\|_F^2 + \sum_{i=1}^{r} \|\mathbf{w}_i \odot \boldsymbol{\Phi}\mathbf{s}_i^{(c)}\|_1$$

$$\text{s.t. } \|\boldsymbol{\alpha}[l,:]\|_0 \leq \eta_l, \; l = 1 \ldots rp \text{ and } \mathbf{S}\boldsymbol{\alpha}\mathbf{V}^T \geq 0$$

Monochromatic PSFs joint super-resolution

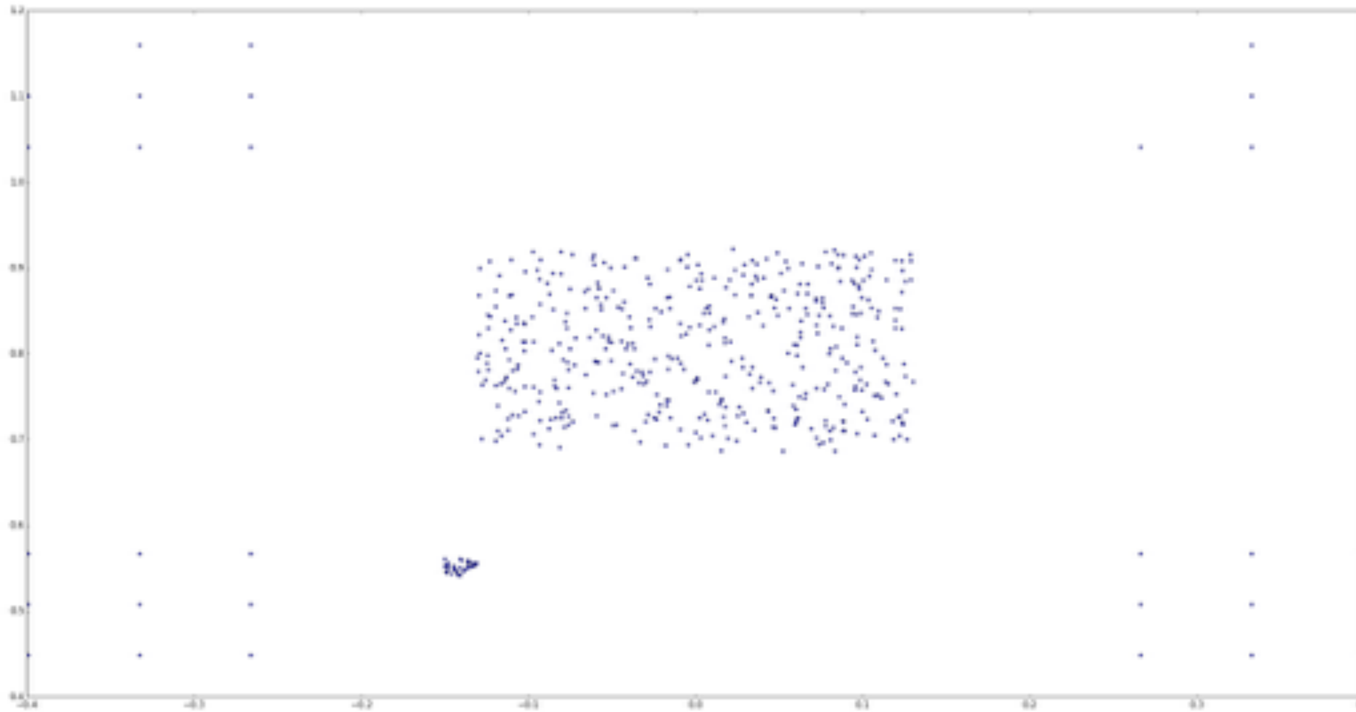
Algorithm 1 Resolved components analysis (RCA)

Parameters estimation and initialization:
    Harmonic constraint parameters $(e_i, a_i)_{1\leq i\leq r} \rightarrow \mathbf{V}, \mathbf{A_0}$
    Noise level, $\mathbf{A_0} \rightarrow \mathbf{W}_{0,0}$
Alternate minimization
for $k = 0$ to $k_{\max}$ **do**
    for $j = 0$ to $j_{\max}$ **do**
        $\mathbf{S}_k = \underset{\mathbf{S}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathcal{F}(\mathbf{S}\mathbf{A}_k)\|_F^2 + \sum_{i=1}^{r} \|\mathbf{W}_{k,j}[:, i] \odot \mathbf{\Phi}_s \mathbf{S}[:, i]\|_1$ s.t. $\mathbf{S}\mathbf{A}_k \geq 0$
        update: $\mathbf{W}_{k,0}, \mathbf{S}_k \rightarrow \text{update}(\mathbf{W}_{k,j+1})$
    end for
    $\boldsymbol{\alpha}_{k+1} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{Y} - \mathcal{F}(\mathbf{S}_k \boldsymbol{\alpha} \mathbf{V}^T)\|_F^2$ s.t. $\|\boldsymbol{\alpha}[l, :]\|_0 \leq \eta_l$
    update: Noise level, $\boldsymbol{\alpha}_{k+1} \rightarrow \mathbf{W}_{k+1,0}$
    $\mathbf{A}_{k+1} = \boldsymbol{\alpha}_{k+1}\mathbf{V}^T$
    $\mathbf{A}_{k+1}[i, :] = \mathbf{A}_{k+1}[i, :]/\|\mathbf{A}_{k+1}[i, :]\|_2$, for $i = 1 \ldots r$
end for
Return: $\mathbf{S}_{k_{\max}}$, $\mathbf{A}_{k_{\max}}$.

Numerical Experiments

Data: 500 Euclid-like PSFs (Zemax), field observed with different SNRs

Theese PSFs account for mirrors polishing imperfections, manufacturing and alignments errors and thermal stability of the telescope.



Quality assessment : shape parameters

$$\gamma(\mathbf{X}) = [e_1(\mathbf{X}), e_2(\mathbf{X})]^T$$

$$e_1(\mathbf{X}) = \frac{\mu_{2,0}(\mathbf{X}) - \mu_{0,2}(\mathbf{X})}{\mu_{2,0}(\mathbf{X}) + \mu_{0,2}(\mathbf{X})}$$

$$e_2(\mathbf{X}) = \frac{2\mu_{1,1}(\mathbf{X})}{\mu_{2,0}(\mathbf{X}) + \mu_{0,2}(\mathbf{X})}.$$

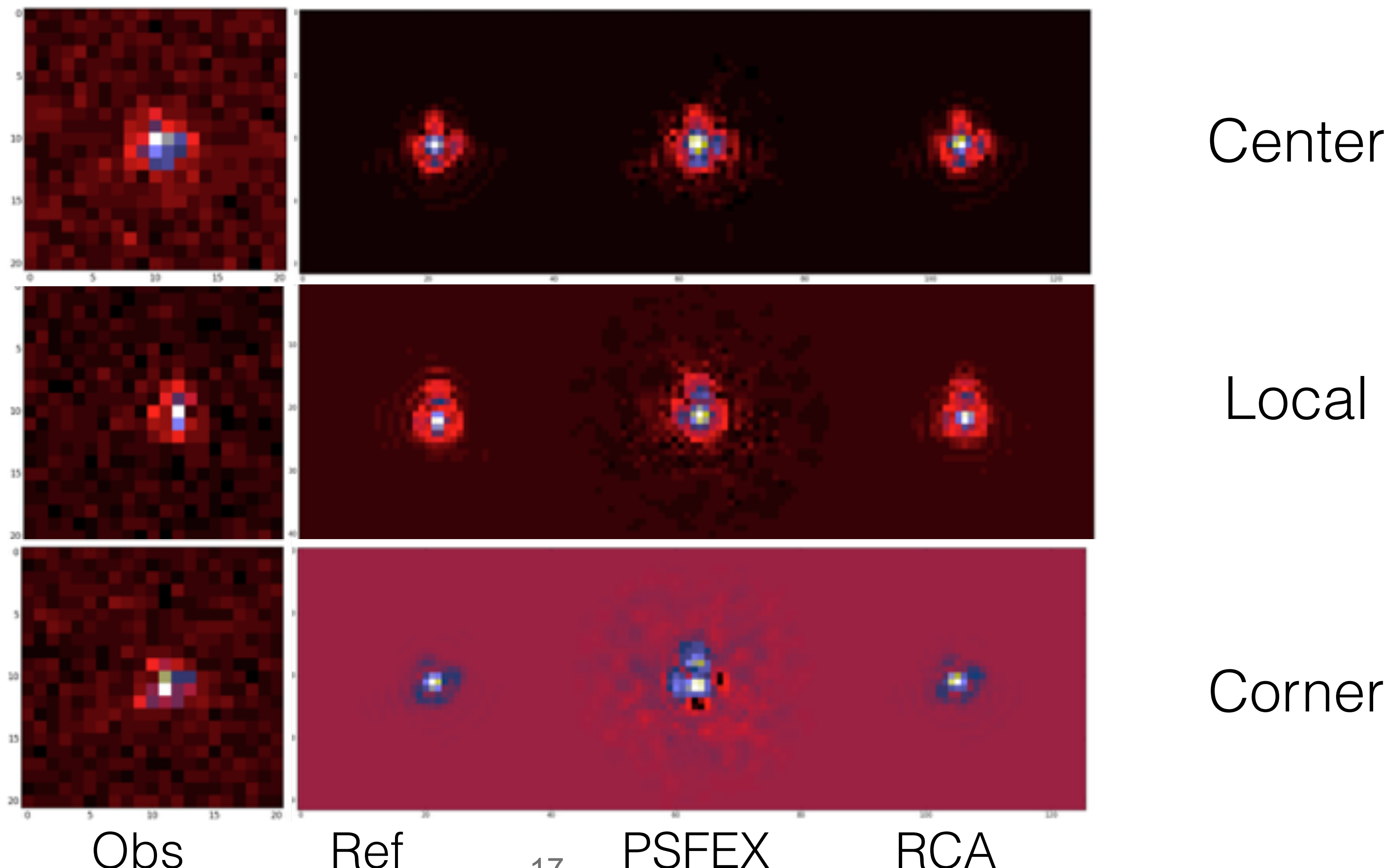
$$\mathrm{E}_{\gamma} = \sum_{i=1}^{p} \|\gamma(\mathbf{X}_i) - \gamma(\hat{\mathbf{X}}_i)\|_2 / p$$

$$\mathrm{Disp}_{\gamma} = \|\mathrm{ME}_{\gamma}\|_{\star}$$

$$\mathrm{ME}_{\gamma} = [\gamma(\mathbf{X}_1) - \gamma(\hat{\mathbf{X}}_1), ..., \gamma(\mathbf{X}_p) - \gamma(\hat{\mathbf{X}}_p)]$$

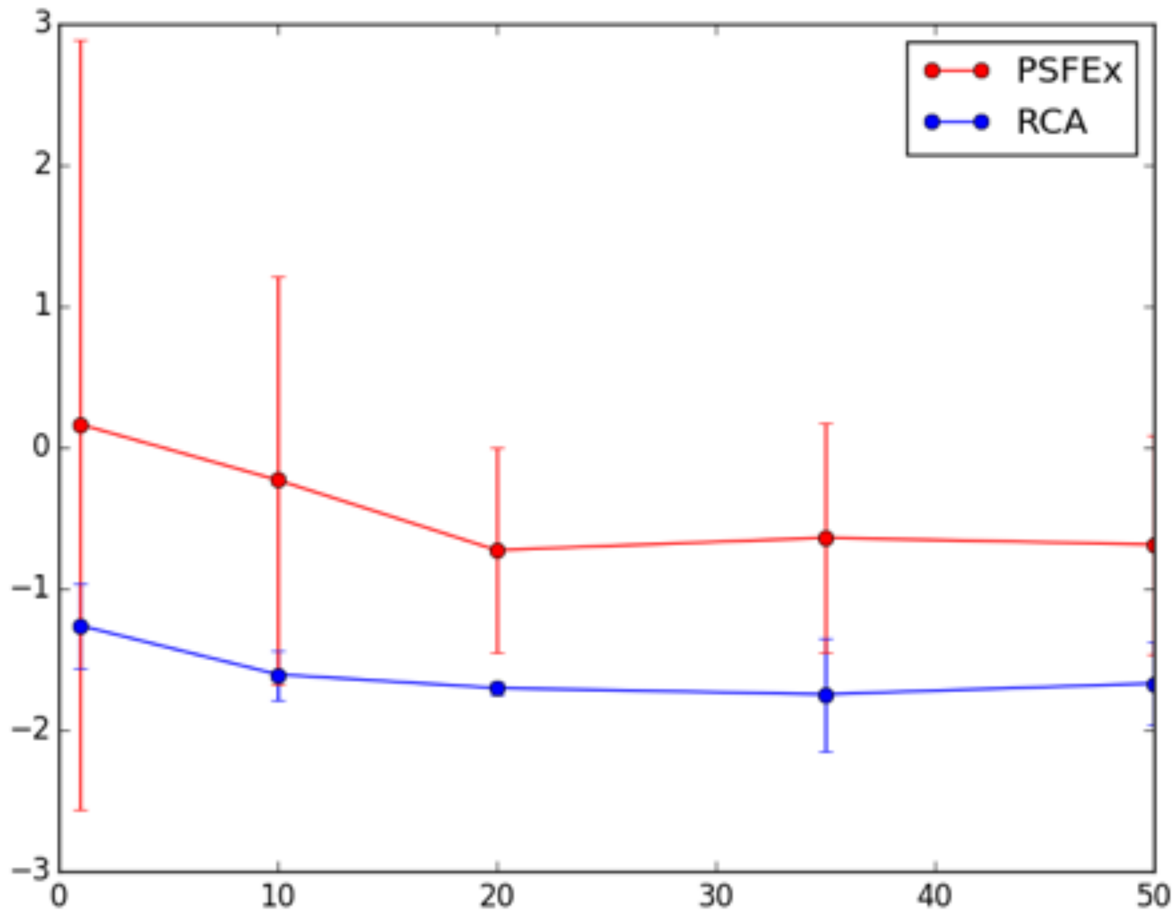
Numerical Experiments

With undersampling (upsampling factor of 2)



Numerical Experiments

With undersampling (upsampling factor of 2)

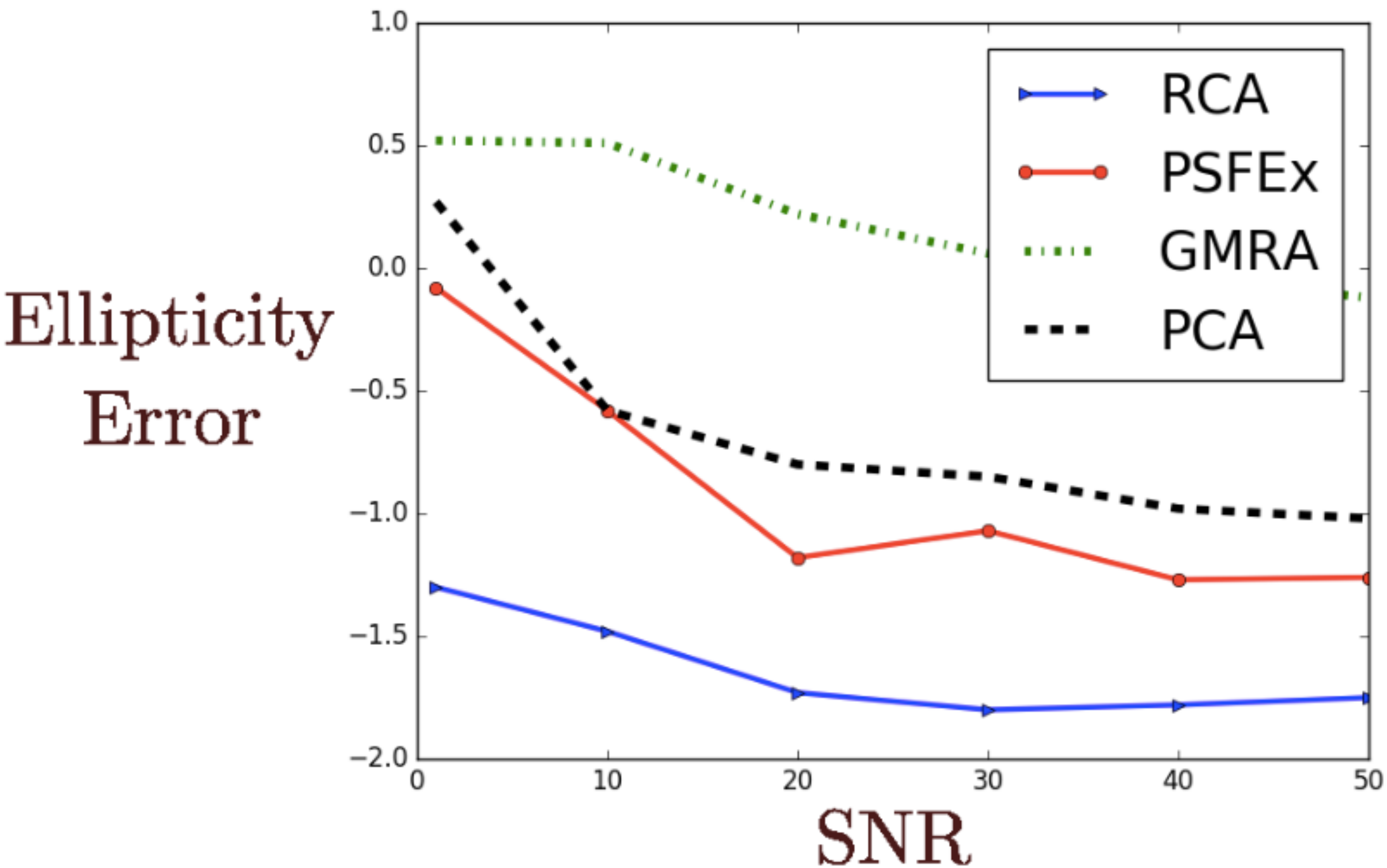


$\log_{10}(\mathrm{E}_\gamma)$

Linear SNR

Numerical Experiments

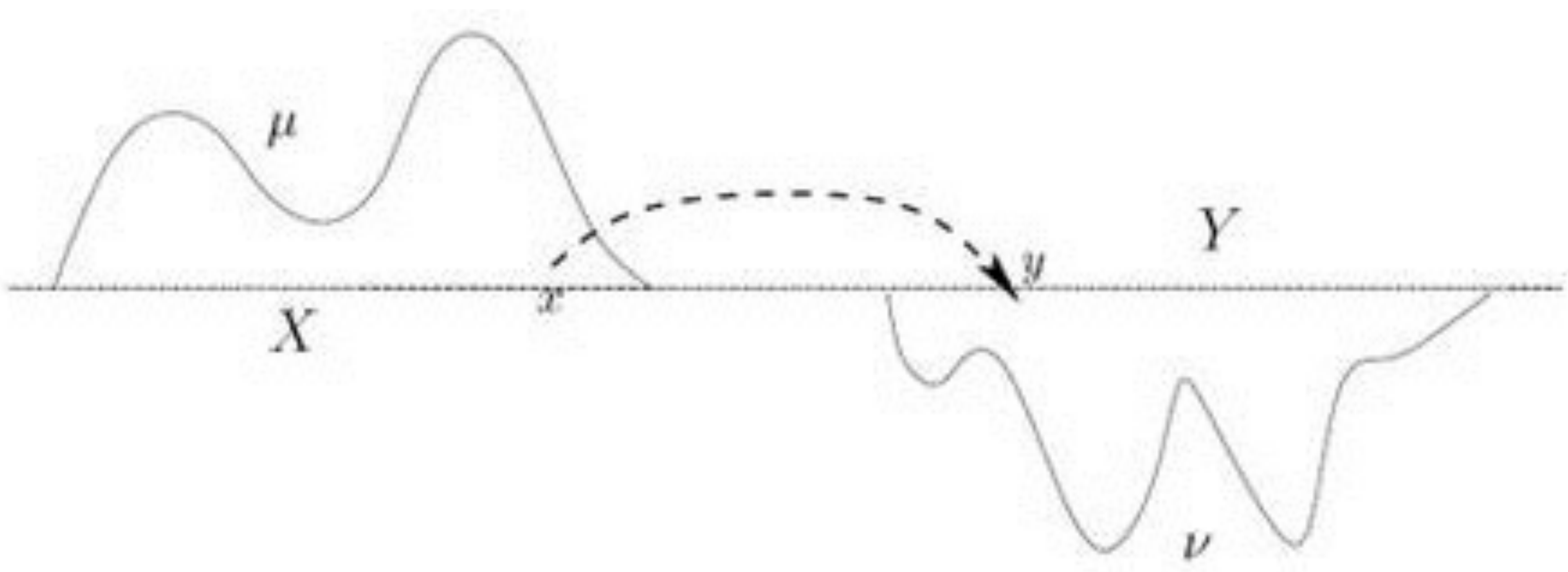
No undersampling



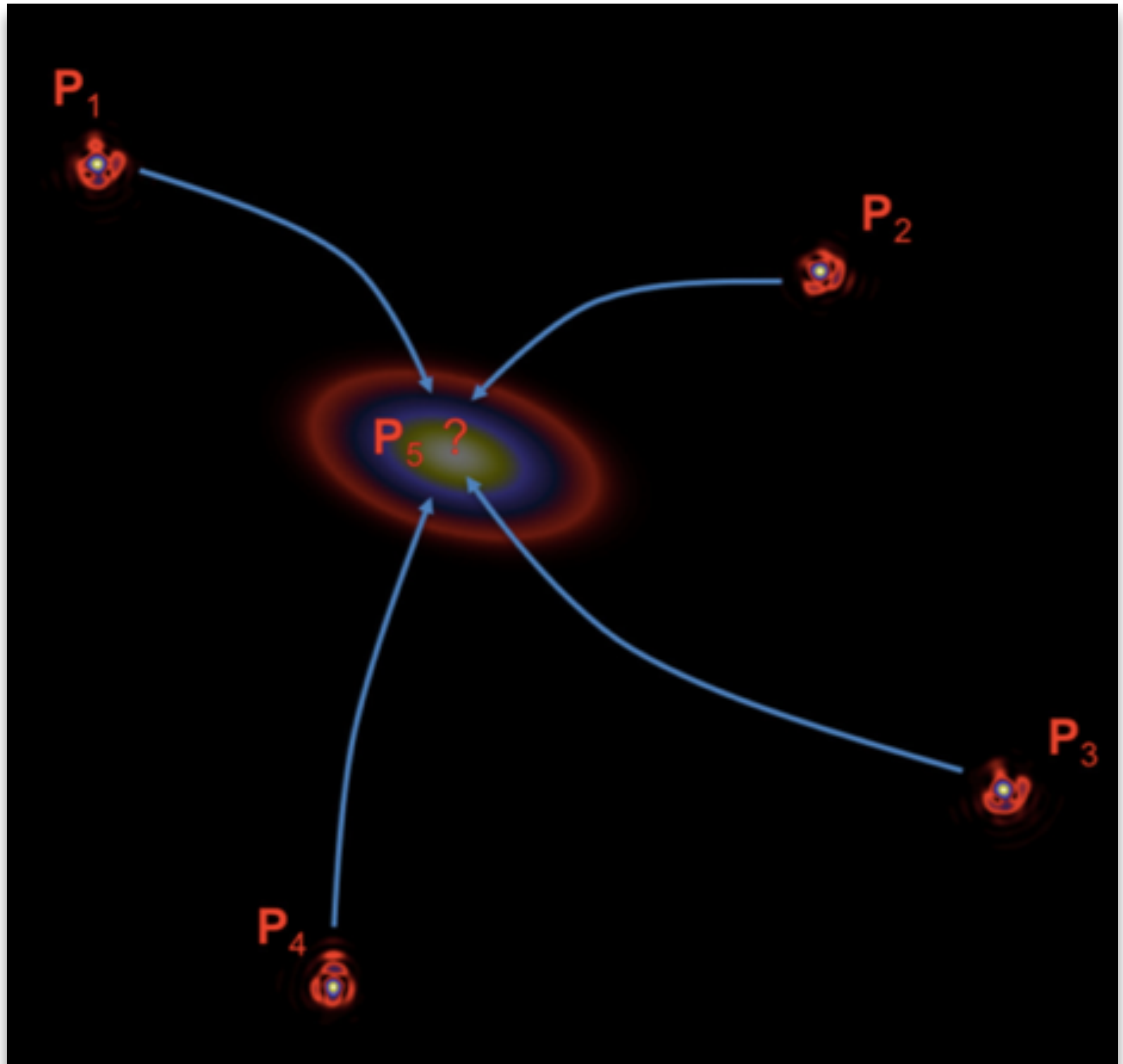
PSF Interpolation

Optimal transport

G. Monge, “Mémoire sur la théorie des déblais et des remblais”, 1781



PSF Interpolation



Optimal transport:

$$\mathbf{X}_5 \approx \underset{\mathbf{X}}{\operatorname{argmin}} \sum_{i=1}^{4} w_i(P_5) W_2(\mathbf{X}_i, \mathbf{X})^2$$

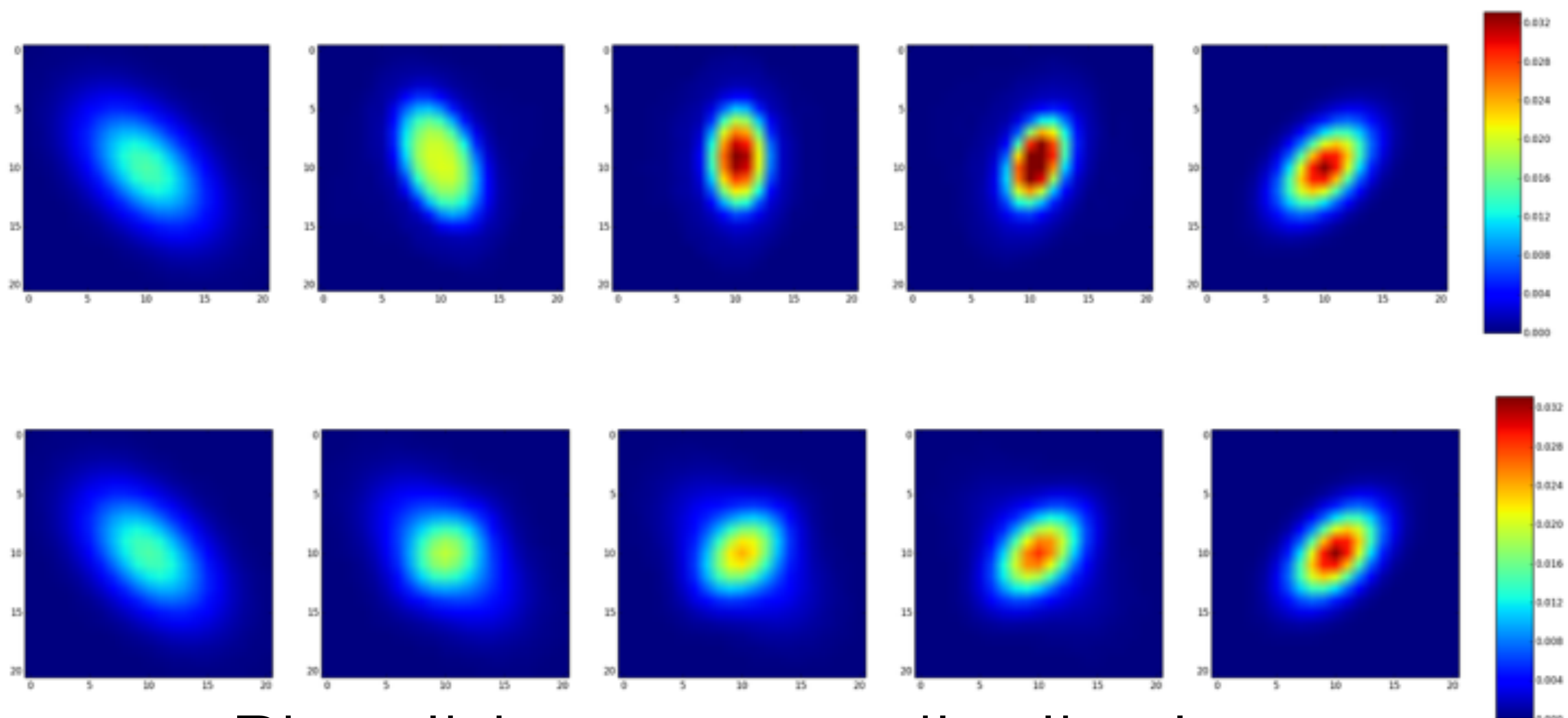
- **F. Ngolè and J.-L. Starck, "PSFs field learning based on Optimal Transport distances", submitted, 2017.**

PSFs interpolation

<u>Barycenter</u>

OT

Linear interpolation



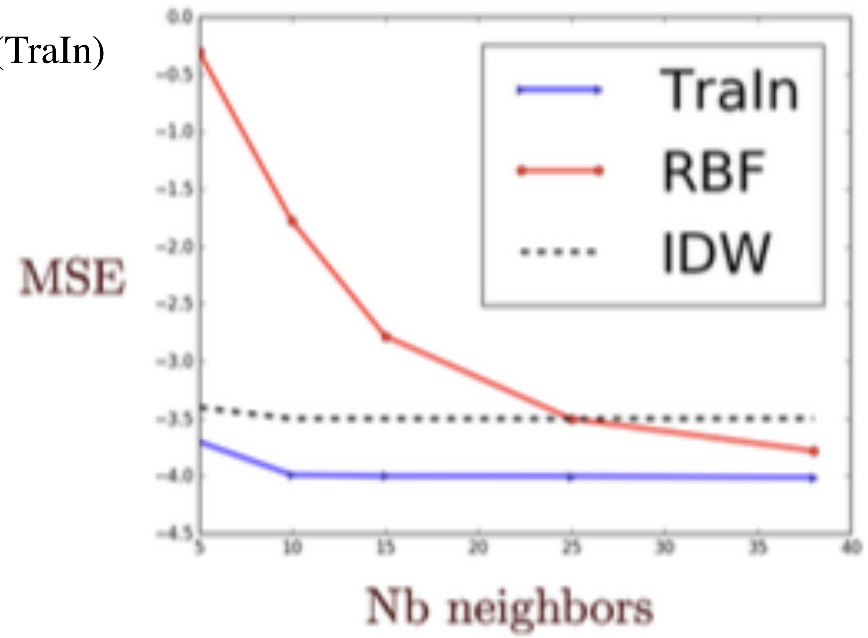
- Plausible average distribution in presence of shifts and changing of shape

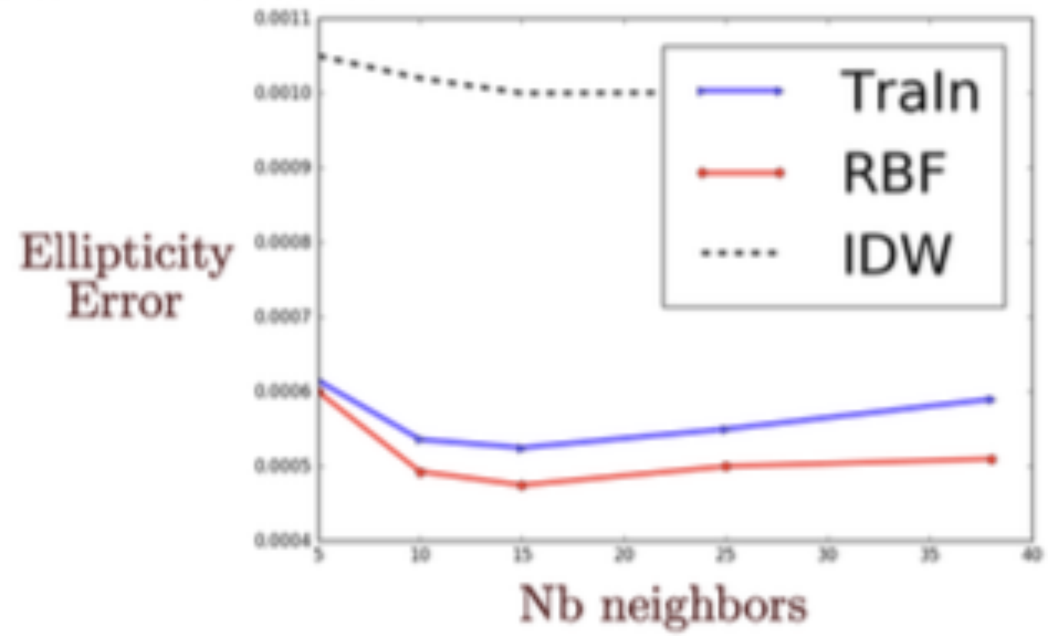
PSF Spatial Interpolation

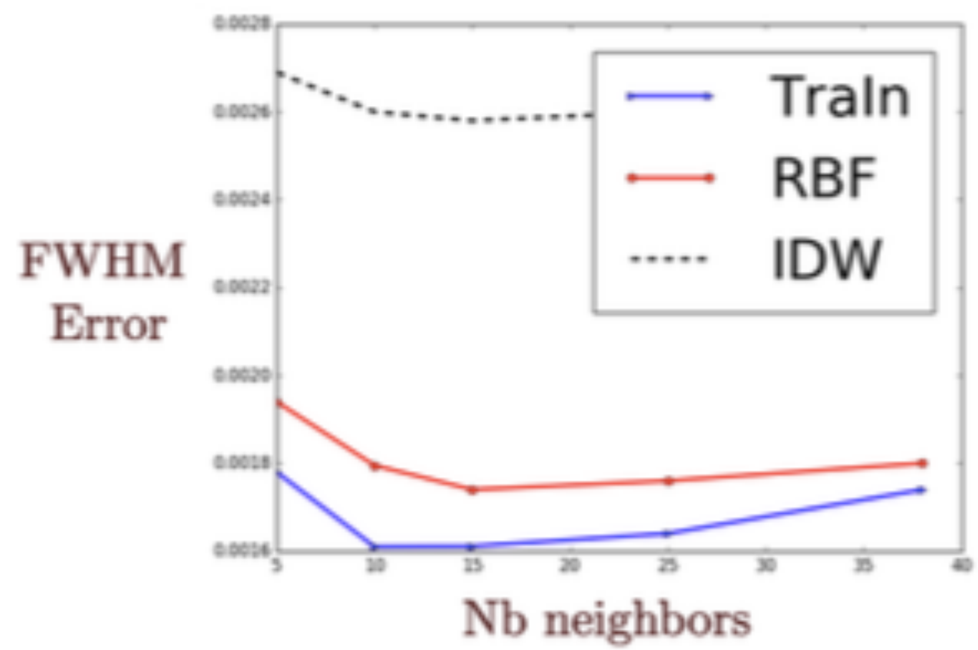
Radial Basis Functions (RBF) based interpolation of PCA coefficients

Inverse Distance Weighting interpolation (IDW)

Transport Interpolation (TraIn)







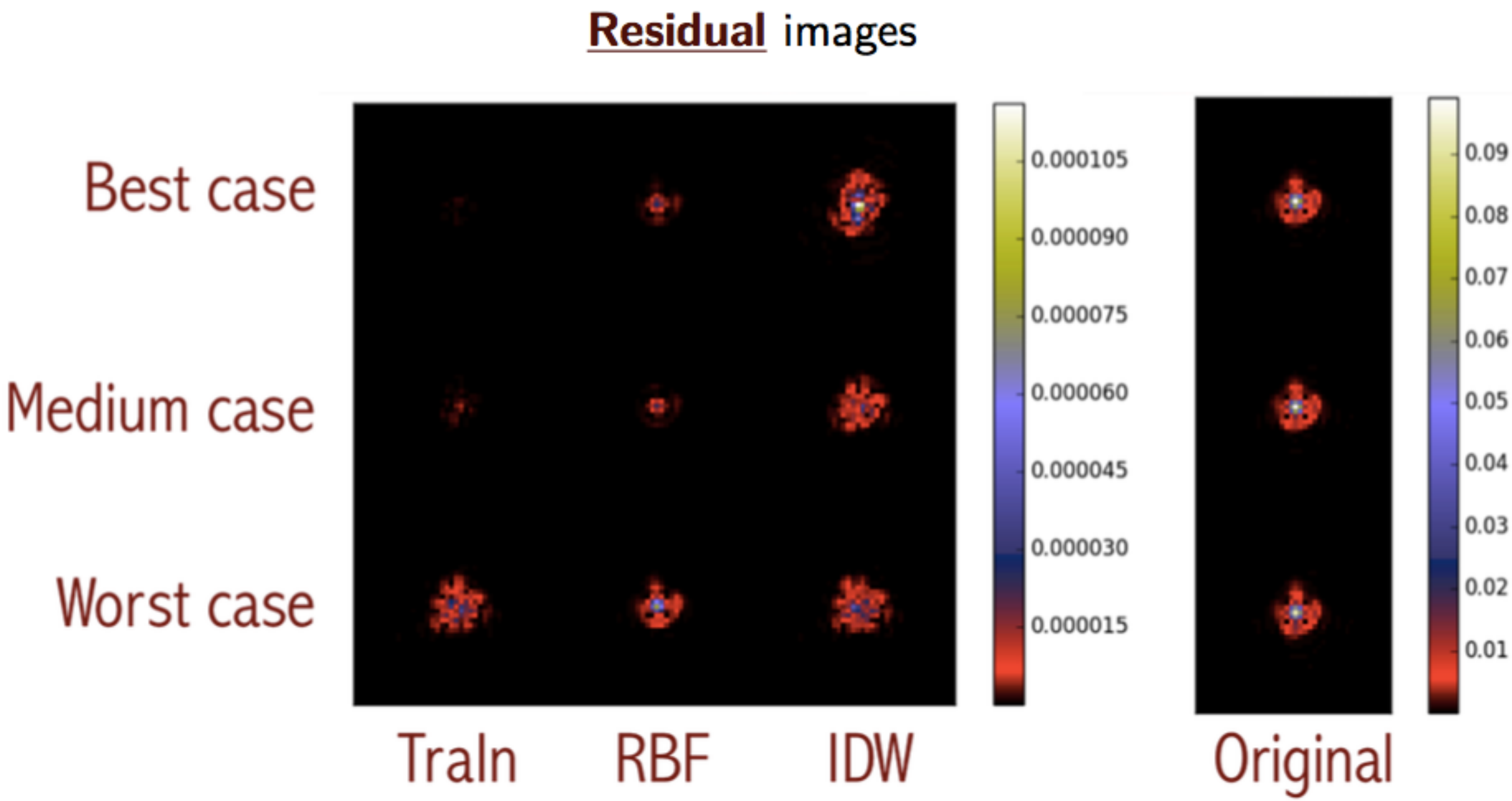
Interpolation quality assessment; *MSE* in log scale

PSF Spatial Interpolation

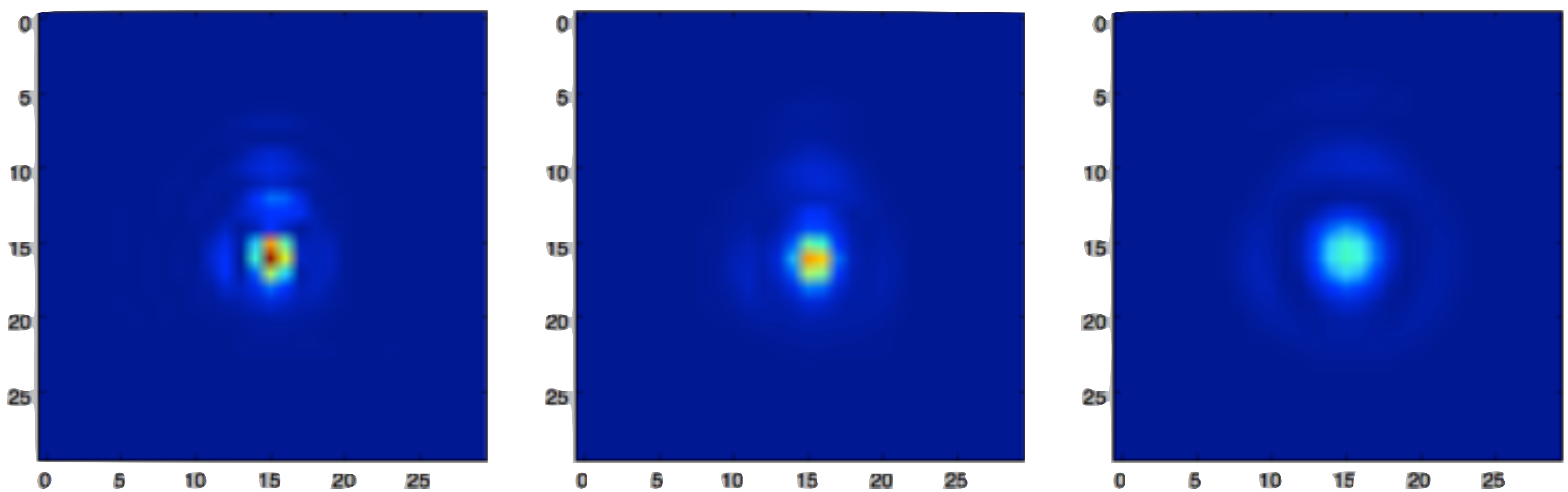
Radial Basis Functions (RBF) based interpolation of PCA coefficients

Inverse Distance Weighting interpolation (IDW)

Transport Interpolation (TraIn)



PSF Color Interpolation



PSF at 550nm | Interpolated PSF | PSF at 900nm

Astronomical Image Deconvolution

Sandard deconvolution framework:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Noise
True Image
PSF Convolution
Observed Image

Sandard deconvolution framework:

$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_2^2 + \|\Phi^t \mathbf{X}\|_p \quad \text{s.t.} \quad \mathbf{X} \geq 0$$

H is huge !!!

Object Oriented Deconvolution

For each galaxy, we use the PSF related to its center pixel:

$$\mathbf{Y} = \mathcal{H}(\mathbf{X}) + \mathbf{N}$$

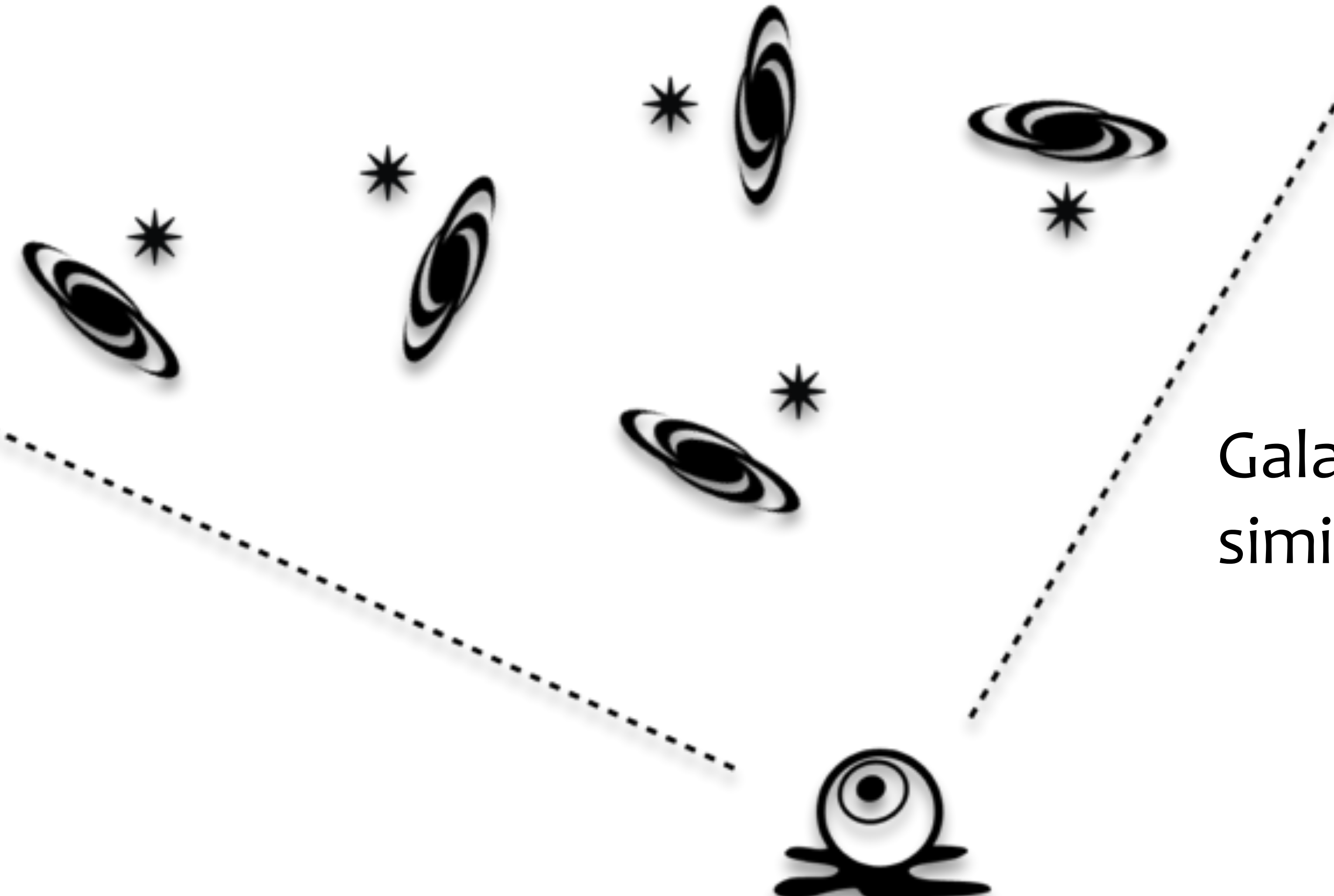
$[n^0, n^1, \dots, n^n]$

$[H^0x^0, H^1x^1, \dots, H^nx^n]$

$[y^0, y^1, \dots, y^n]$

$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \mathcal{H}(\mathbf{X})\|_2^2 + \lambda\|\Phi^t\mathbf{X}\|_p \quad \text{s.t.} \quad \mathbf{X} \geq 0$$

Big Astronomical Image Deconvolution



Galaxy images have similar properties.

$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \mathcal{H}(\mathbf{X})\|_2^2 + \lambda\|\mathbf{X}\|_* \quad \text{s.t.} \quad \mathbf{X} \geq 0$$

$$\|X\|_* = \sum_i \sigma_i$$

Primal-dual splitting from Condat-Vu (2013)

$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad [\underbrace{F}_{\text{Convex Function}}(\mathbf{X}) + \underbrace{G}_{\text{Functions}}(\mathbf{X}) + \underbrace{K}_{\text{Functions}}(\underbrace{\mathcal{L}}_{\text{Linear Operator}}(\mathbf{X}))]$$

Algorithm: Choose the proximal parameters τ > 0, ς > 0, the positive relaxation parameter, ξ, and the initial estimate (X0, Y0). Then iterate, for every k ≥ 0.

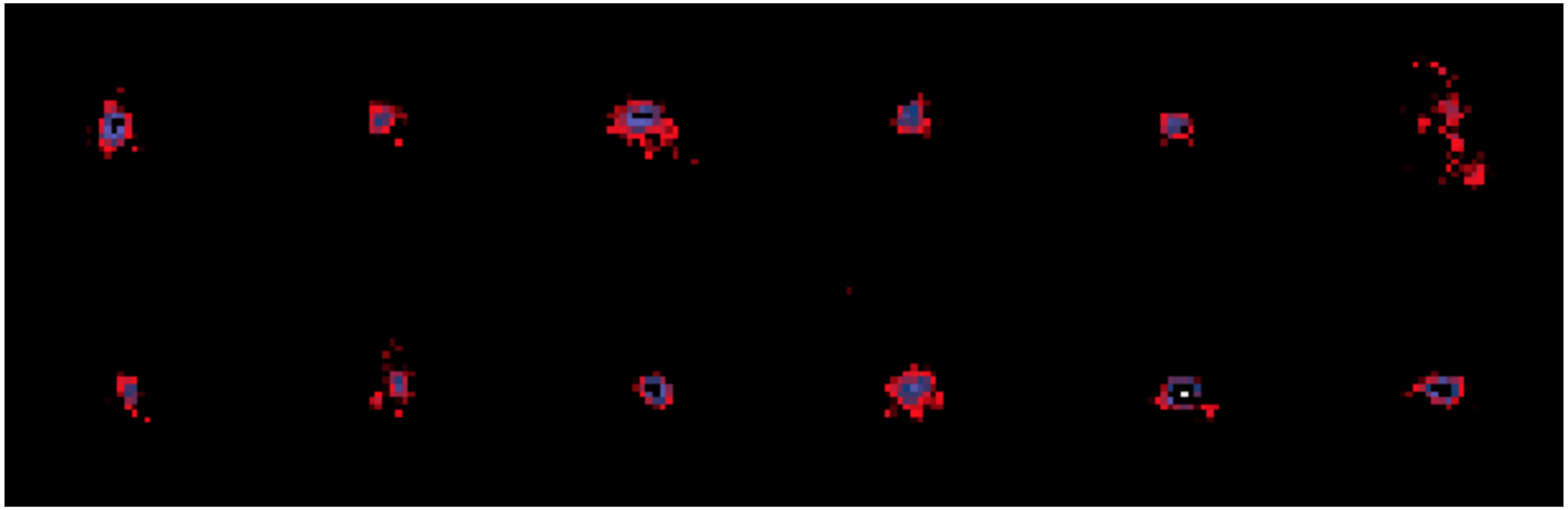
$$1: \quad \tilde{\mathbf{X}}_{k+1} = \mathrm{prox}_{\tau G}(\mathbf{X}_k - \tau\nabla F(\mathbf{X}_k) - \tau\mathcal{L}^*(\mathbf{Y}_k))$$

$$2: \quad \tilde{\mathbf{Y}}_{k+1} = \mathbf{Y}_k + \varsigma\mathcal{L}(2\tilde{\mathbf{X}}_{k+1} - \mathbf{X}_k) - \varsigma\mathrm{prox}_{K/\varsigma}\Big(\frac{\mathbf{Y}_k}{\varsigma} + \mathcal{L}(2\tilde{\mathbf{X}}_{k+1} - \mathbf{X}_k)\Big)$$

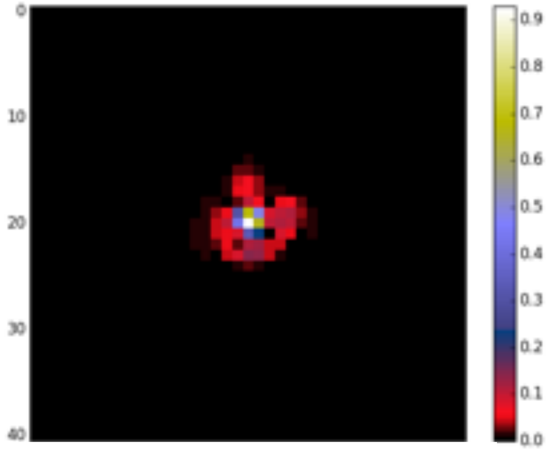
$$3: \quad (\mathbf{X}_{k+1}, \mathbf{Y}_{k+1}) := \xi(\tilde{\mathbf{X}}_{k+1}, \tilde{\mathbf{Y}}_{k+1}) + (1-\xi)(\mathbf{Y}_k, \mathbf{Y}_k)$$

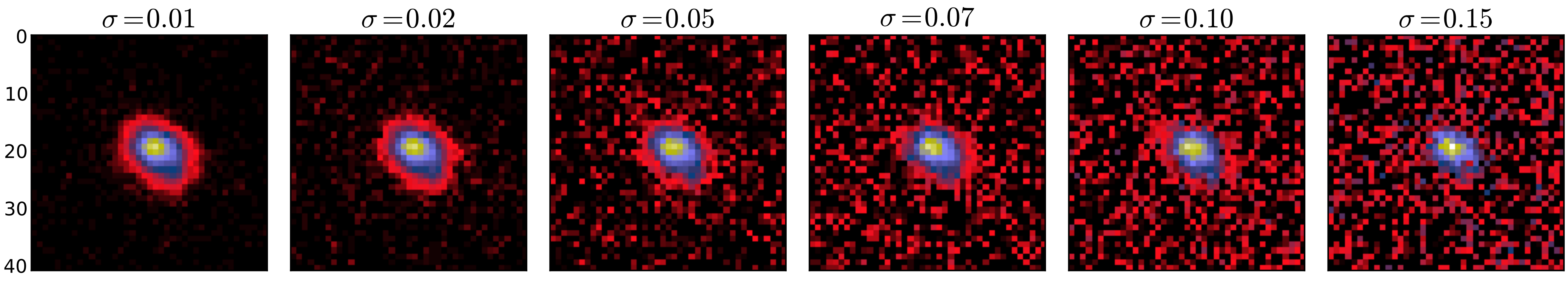
The Simulated Data

- 10,000 space-based galaxy images derived from COSMOS data.
- Each image is a 41×41 pixel postage stamp around the centre of the galaxy.
- Images are free from PSF effects.

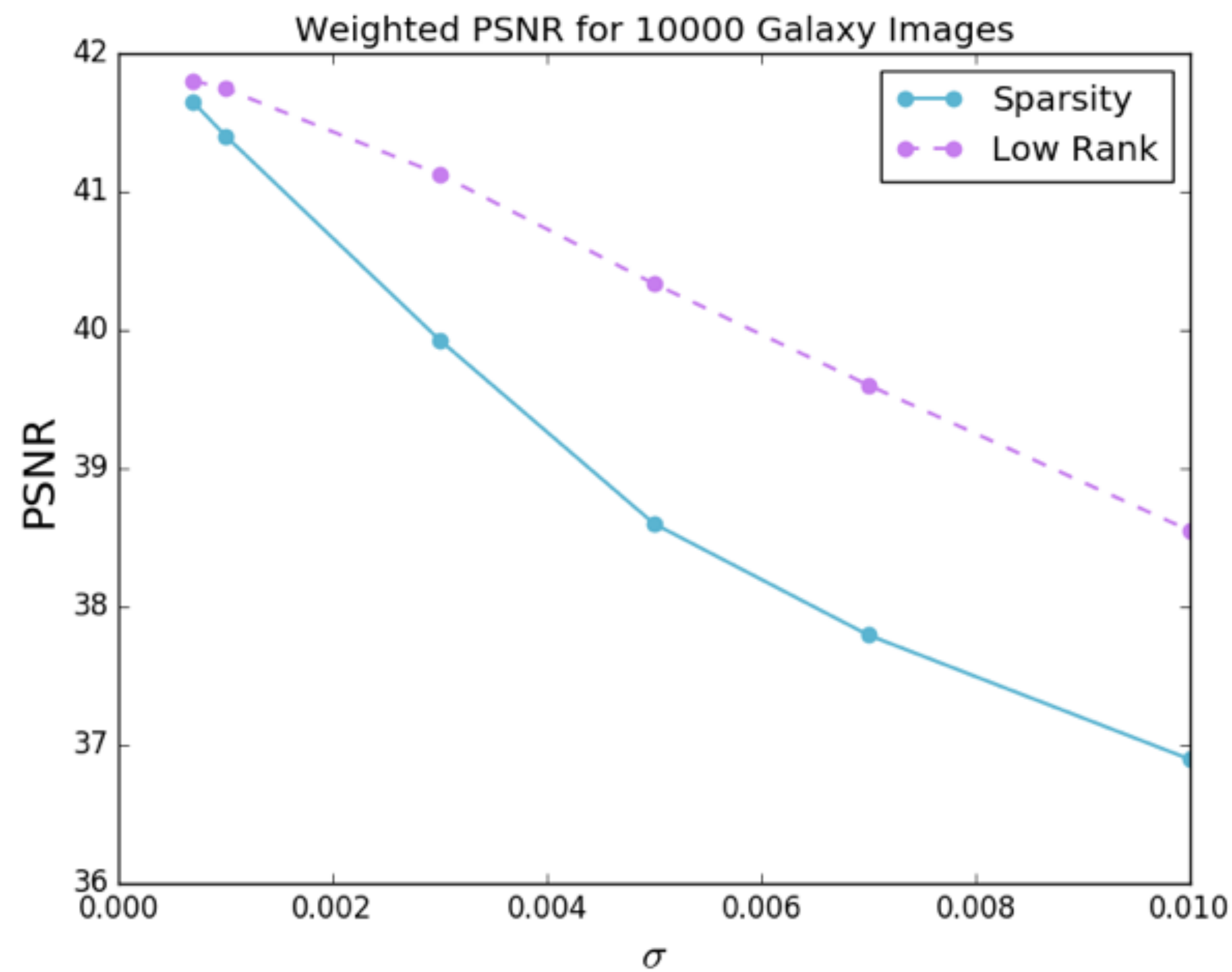


- **600** spatially varying Euclid-like PSFs
  - Each galaxy image is convolved with a random PSF.
  - Different levels of Gaussian noise is added.





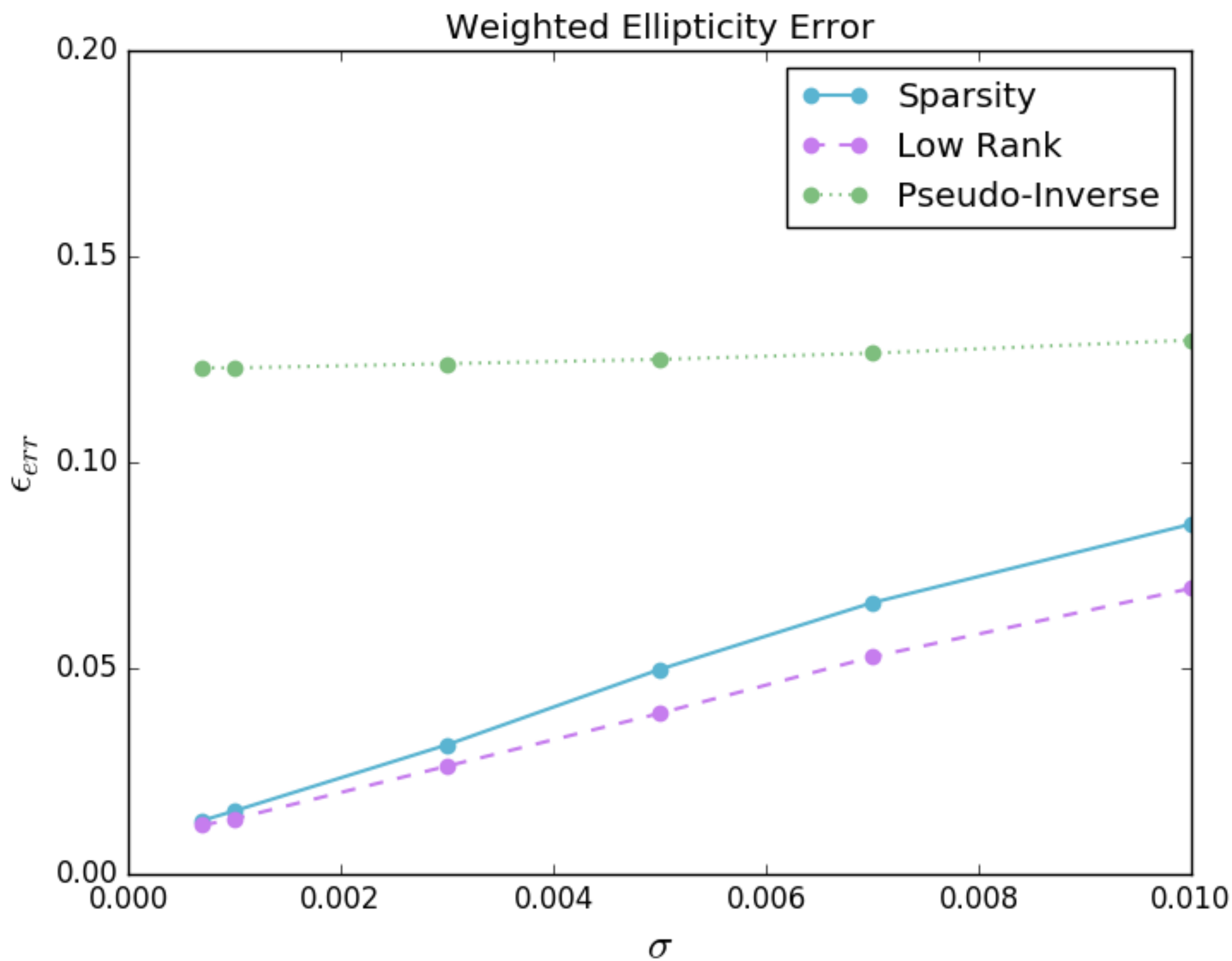
Results



S. Farrens, F.M. Ngolè Mboula, and J.-L. Starck, "Space variant deconvolution of galaxy survey images", submitted, 2017.

Code available at https://github.com/sfarrens/psf

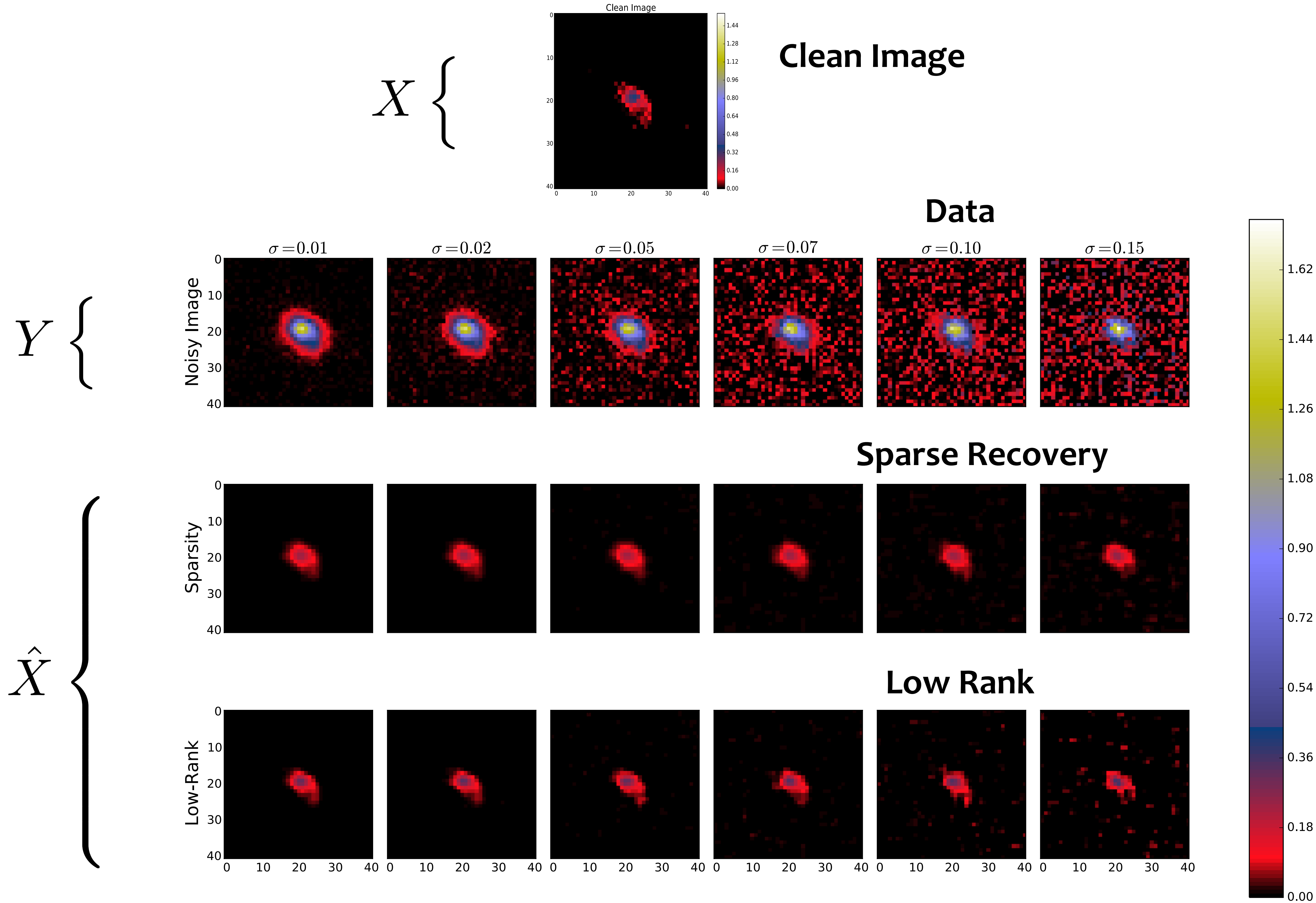
Results



S. Farrens, F.M. Ngolè Mboula, and J.-L. Starck, “Space variant deconvolution of galaxy survey images”, submitted, 2017.

Code available at https://github.com/sfarrens/psf

Results



X
Clean Image
Clean Image
Data
Y
Noisy Image
$\sigma = 0.01$
$\sigma = 0.02$
$\sigma = 0.05$
$\sigma = 0.07$
$\sigma = 0.10$
$\sigma = 0.15$
Sparse Recovery
$\hat{X}$
Sparsity
Low Rank
Low-Rank
1.62
1.44
1.26
1.08
0.90
0.72
0.54
0.36
0.18
0.00

Conclusions

✓ **Weak-Lensing and the Euclid Mission:**

➡ Serious mathematical challenges

➡ Need the best methods to fulfil the mission requirements

✓ **Euclid PSF:** **RCA** is new sparsity based method which uses all available PSFs to derive the PSF field.

- F. Ngole, J.-L Starck, et al, "Constraint matrix factorization for space variant PSFs field restoration", submitted, 2016.

Perspectives

* **PSF: Take into account the wavelength variation.**

* **CFIS: New surveys at the CFHT starting in 2017. It will be a laboratory to test new methods before Euclid.**