Multiband Characterisation of Galaxies

from
a hierarchy of models,
trees of connected components
and a markovian algorithm.

B.Perret : 2010, Ph.D .Thesis LSIIT Strasbourg
Supervisors : C.Collet,E.Slezak

Perret,B., Lefèvre,S., Collet,C., Slezak,E. : 2010, ICPR
Perret,B., Lefèvre,S., Collet,C., Slezak,E. : 2011, Pattern Recognition 44, 1328

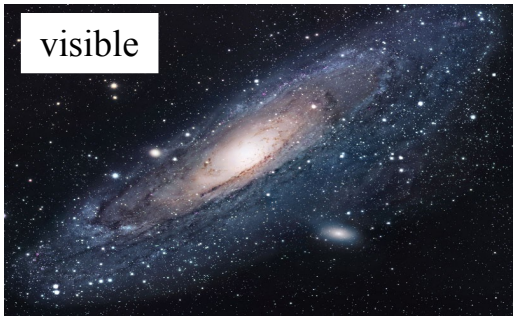
OUTLINE

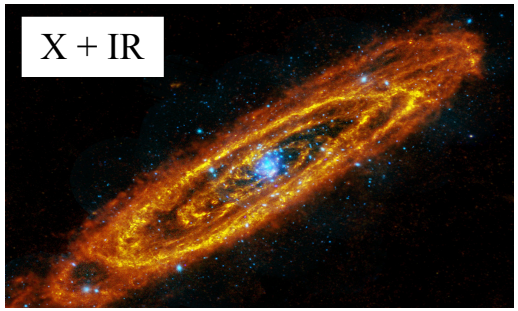
Introduction
Parametric models
Mathematical morphology
Results on the EFIGI database
Perspectives

INTRODUCTION

Target : automatic classification of galaxies
with high enough spatial resolution and SNR

Aim : to use both the shape and spectral information contents
to obtain a coherent morphological description of these objects





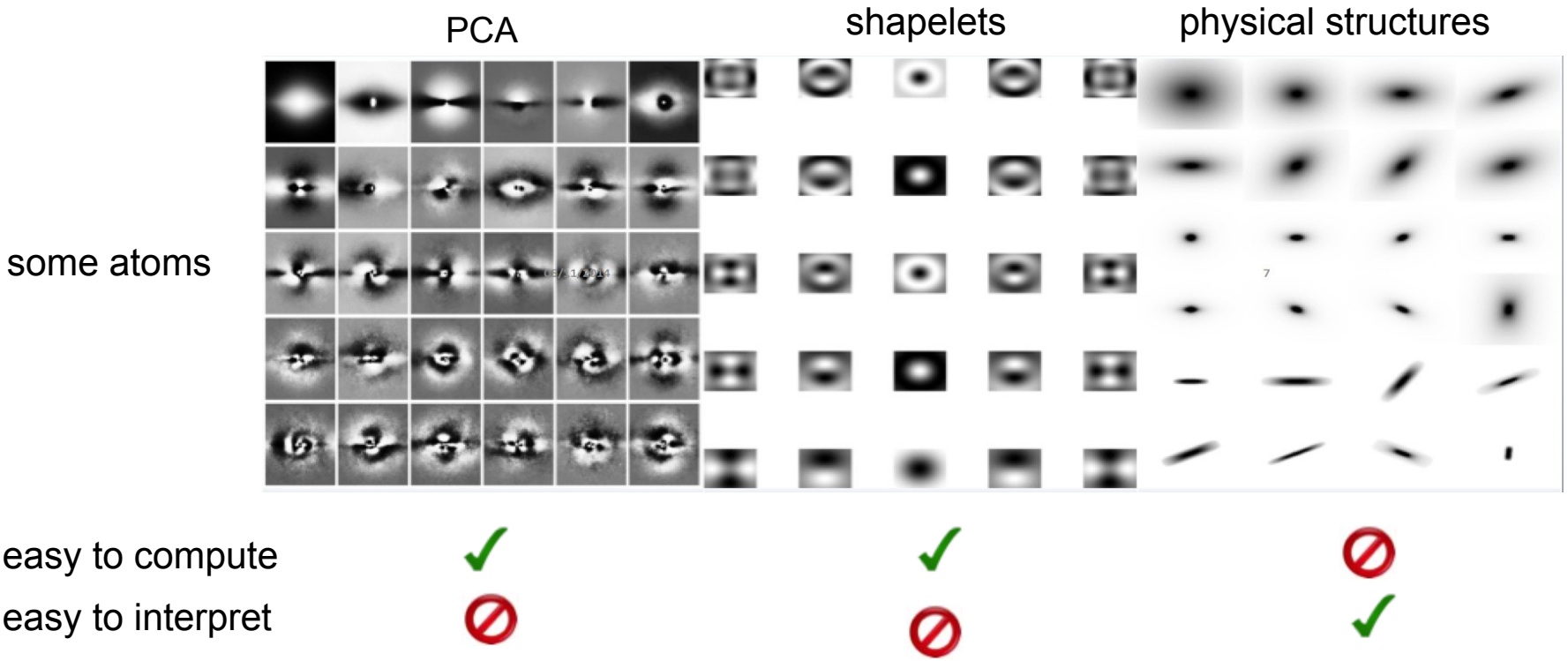
Quantitative and Objective
Global Multiband Morphological Description
Ann Classification

INTRODUCTION

Two approaches to get measures on galaxies :

a) empirical characteristics :
e.g. : Petrossian radius, symetry, concentration index, Gini factor, …

b) decomposition onto mathematical bases :
→ **physical meaning** ?



INTRODUCTION
decomposition onto structures

Parametric models for fitting the 2D image :

e.g. : Gim2D (bulge,disk) ; GalFit (free) ; BUDDA (bulge, disk, stellar bar)

difficulties : energy function complexity, initialisation setting, unicity, computational cost, multiband capability

Our strategy for a multiband morphological characterisation of galaxies

- Analysis :
  - a) a multiband parametric model for each relevant physical structure
    bulge, disk, stellar bar, spiral arms, rings, ...
  - b) a hierarchy of models (from a coarse to fine modelisation + init.)
  - c) a robust optimisation method → MCMC algo.
- Pre-processing and initialisation :

  → to mask overlapping or blended features in the image

  → to detect and extract specific features (e.g. spiral arms)

OUTLINE

Introduction
Parametric models
Mathematical morphology
Results on the EFIGI database
Perspectives

IMAGE MODELLING

The observation is modelled as :

$$\mathbf{A}_{\mathrm{m}}\left(\mathbf{Y}, \boldsymbol{\phi}\right) = \mathbf{H}\mathbf{F}(\boldsymbol{\phi}) + \mathbf{e}$$

with :

- $\mathbf{Y}$ the observation (a vector of pixels) ;
- $\mathbf{H}$ the point spread function ;
- $\mathbf{e}$ the (ND uncorrelated zero-mean Gaussian) noise;
- $\mathbf{F}(\boldsymbol{\phi})$ the parametric model of the object of interest ;

$$\mathbf{F}(\boldsymbol{\phi}) = \left(1 - \mathbf{o}\left(\boldsymbol{\phi}_{\mathbf{o}}\right)\right) \sum_i \mathbf{f}_i\left(\boldsymbol{\phi}_{\mathbf{f}_i}\right)$$

i.e. a sum of bright features $\mathbf{f}_i$ weighted by an occlusive structure $\mathbf{o}$

- $\mathbf{A}_{\mathrm{m}}\left(\mathbf{Y}, \boldsymbol{\phi}\right)$ the filter on the observation.

MODELISATION

Morphological features :

– bulge, disk
– stellar bar
– spiral arms
– ring
– edge-on disk

well described in mono-λ approaches

included in our multi-λ model

Scope of the model :
– elliptical galaxies (20%) ✓
– spiral galaxies with and without bar (77%) ✓
– irregular galaxies (3%) ⊘

Main ideas and hypotheses (for each structure) :

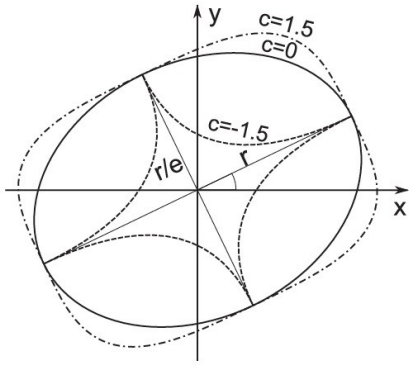
– a shape descriptor models the geometry of the isophotes ;
– a brightness descriptor models the surface brightness changes ;

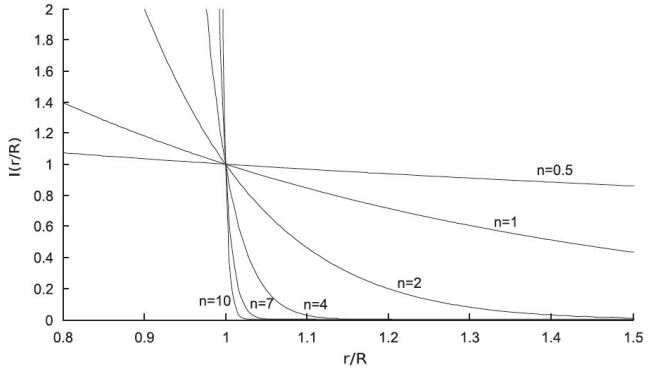
– the shape descriptor does not depend on λ ;
– the brightness descriptor is a function of λ.

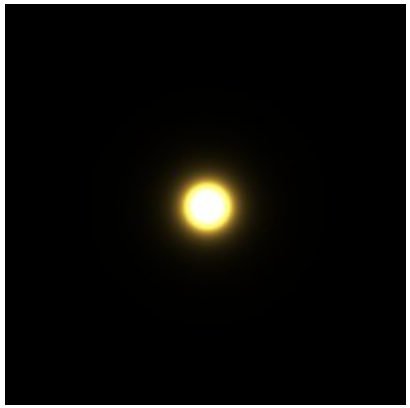
MODELISATION
(bulge, disk, bar, ring, halo)

shape descriptor
> a generalized ellipse <

surface brightness descriptor
> a Sérsic function <



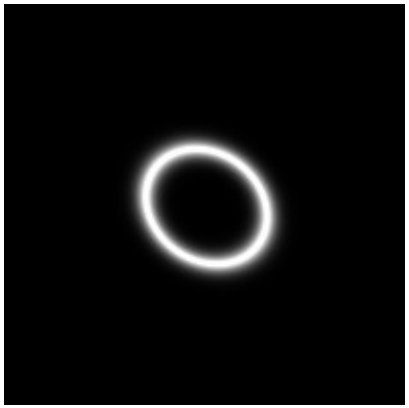




bulge

disk

bar

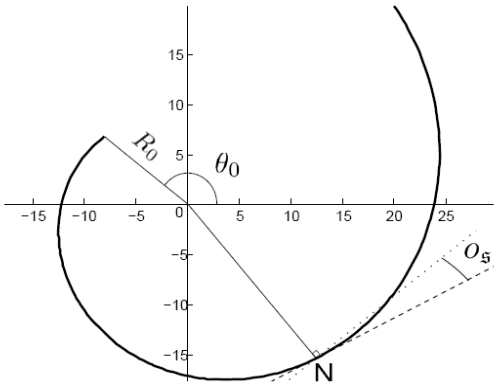


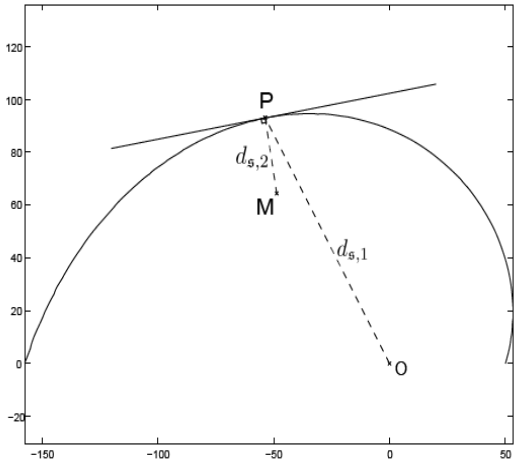
ring

MODELISATION
(development and tightness of 2 spiral arms)

Shape descriptor : a logarithmic spiral (cf. face-on galaxies)
projected onto the observation plane
→ parameters are linked to those of the disk shape descriptor

SB descriptor : a Sérsic law (generic)
– def. of a pseudo-distance : wrt centre + wrt crest
– computation by means of a distance transform with a dist. mask





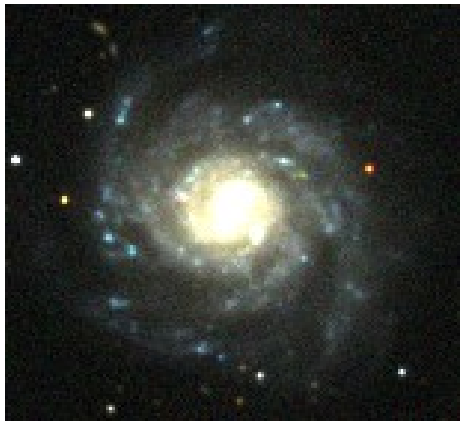
MODELISATION
(spiral arms)

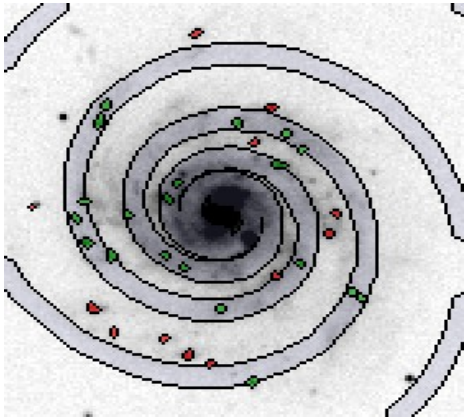
Links with the parameters of the bar (if any) :

– difference between the position angles is small
– difference between the cut-off and starting radii is small

Adaptive masking of the (compact) HII regions :

– to obtain an unbiased SB measure for the spiral structure
– their spatial distribution is a marker of the spiral structure
→ the component is masked only if it is closed to the crest of the spiral



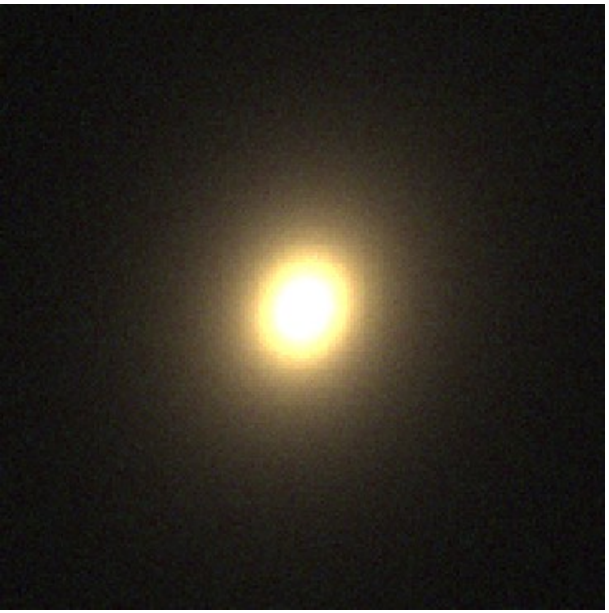


MODELISATION

Edge-on disk :
– shape descriptor : along and perpendicular to the galactic plane
(the scale height does not depend on the position)
– sb. descriptor : an exponential along the gal. plane
a sech^2 law perpendicular to this plane

Obscuration :
– shape descriptor : the same as for an edge-on disk
– sb. descriptor : an exponential, with values between 0 and 1

MODELISATION
examples of combination of structures



a cD galaxy

a spiral galaxy
with bar and ring

an edge-on spiral

MODELISATION
model estimation

Bayesian formalism : $$P(\boldsymbol{\phi} \mid \mathbf{Y}) = \frac{P(\mathbf{Y} \mid \boldsymbol{\phi})\, P(\boldsymbol{\phi})}{P(\mathbf{Y})}$$

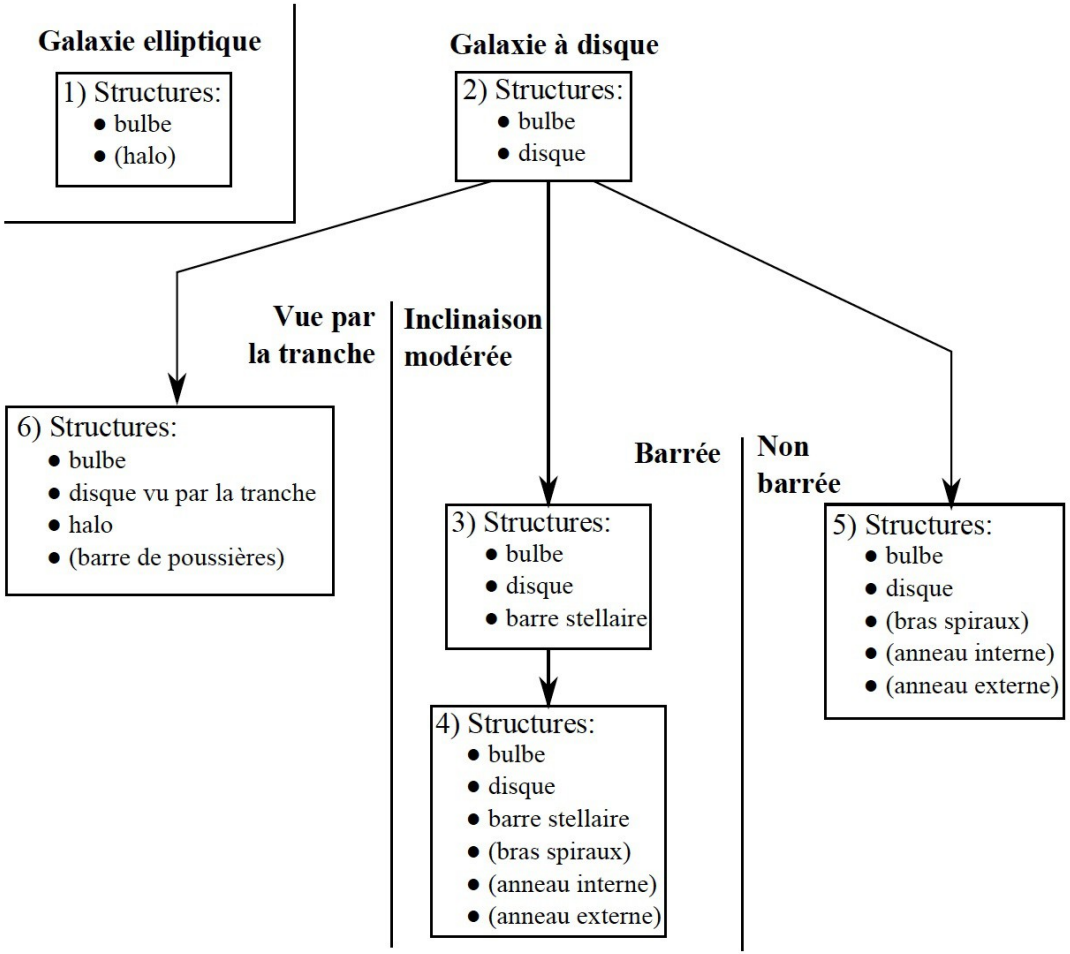
– t.i.a. noise
– t.i.a. intrinsic variability of the observations
– t.i.a. priors (cf. multiplicity of solutions of an ill-posed pb.)

Estimator : maximum a posteriori (MAP)

Optimisation : Monte Carlo Markov Chain (MCMC) method
– sampling (Gibs) of the parameter space : shape vs. brightness vs. computational cost
– MH random steps + adaptive direction sampling (snooker algo.)
– convergence : MNSE evolution
– simulated annealing using several temperatures (cf. blocks)
– hierarchical decomposition scheme (fast, robust, accurate init.)
step 0 : 2nd order moments, r_e (1.0;0.5) ; r ¼ law ; flux (0.25;0.75)

A HIERARCHY OF MODELS

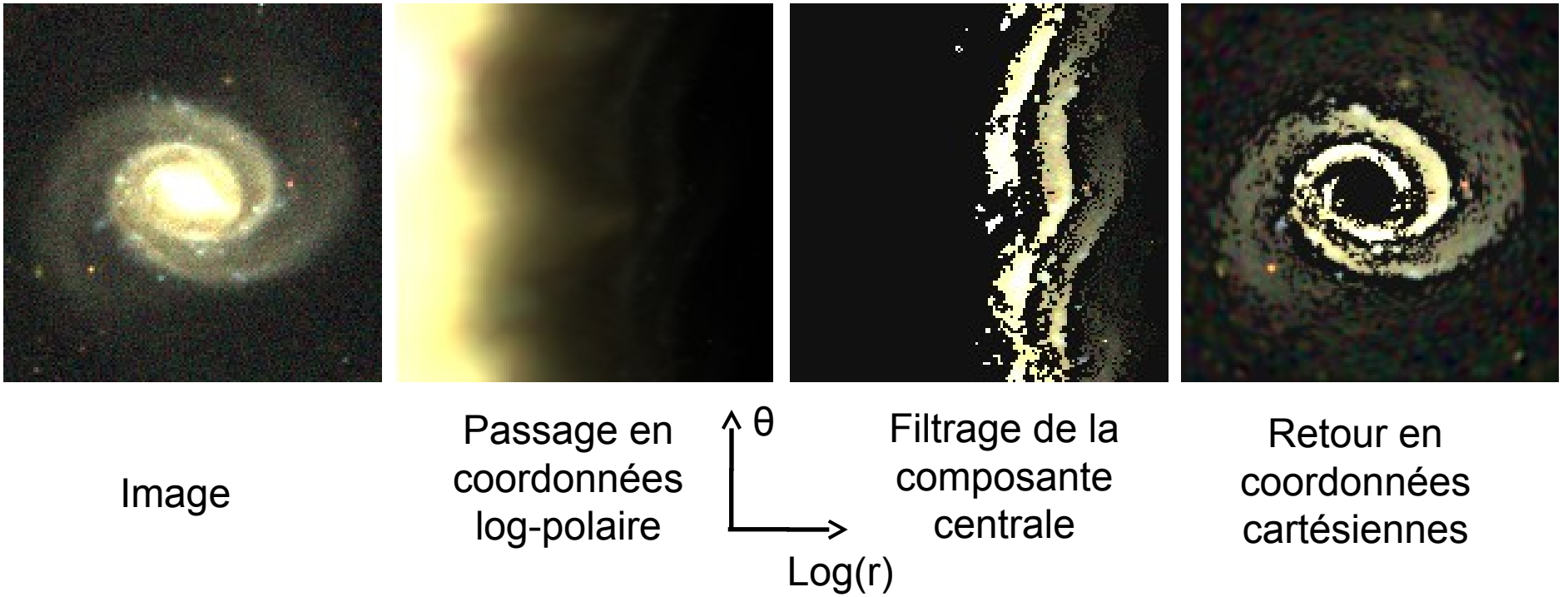
bulge + disk / + bar / + spiral arms



MODELISATION

model estimation – initial parameters for the spiral arms

Fourier analysis of a residual in a log-polar space :
– pitch angle : inclination of the sine wave
– starting radius : translation



OUTLINE

Introduction
Parametric models
Mathematical morphology
Results on the EFIGI database
Perspectives

Mathematical Morphology

Goals : – multiple object cf. foreground / neighbouring sources
– adaptive masking cf. compact HII regions

→ trees of connected components (Max-Tree)

Principle : hierarchical decomposition of an image

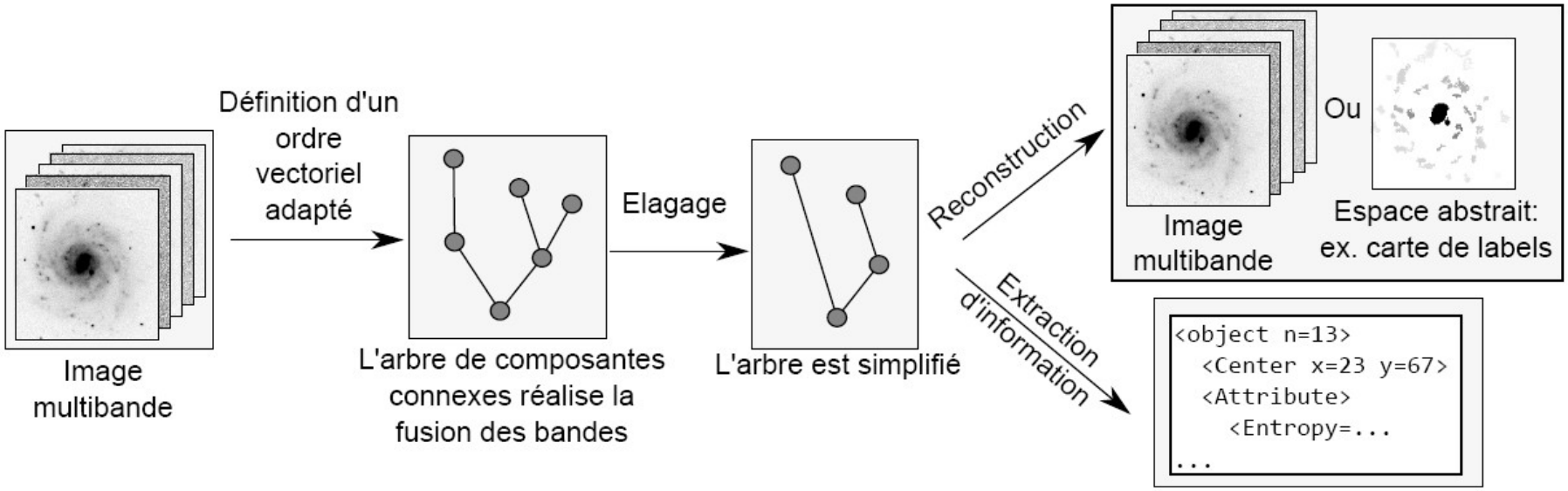
– each node is a connected component of a thresholding
– the tree is defined by the inclusion relationship between nodes
– each node is provided with a set of attributes
– pruning of the tree according to the values of the attributes
– restoration of the filtered image

– extension to mutivariate images → one tree to represent all bands
• def. of a (total) vectorial order •
i) weighted energy, weighted SB ordered by spatial resolution
ii) lexicographic comparison

$$\mathrm{E}_n(v) = \left\| \frac{v_1}{k\sigma_1}, \ldots, \frac{v_n}{k\sigma_n} \right\|$$

TREE OF CONNECTED COMPONENTS

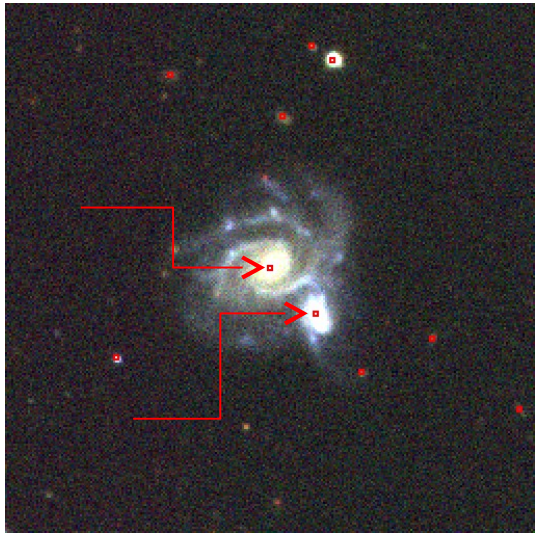
principle overview

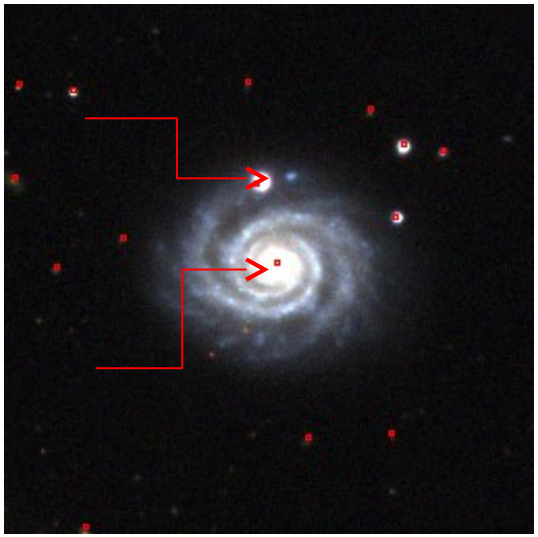


TREE OF CONNECTED COMPONENTS

source identification

Principle : a node is related to an individual source if its multiband volume is large enough compared to the volume of its father.

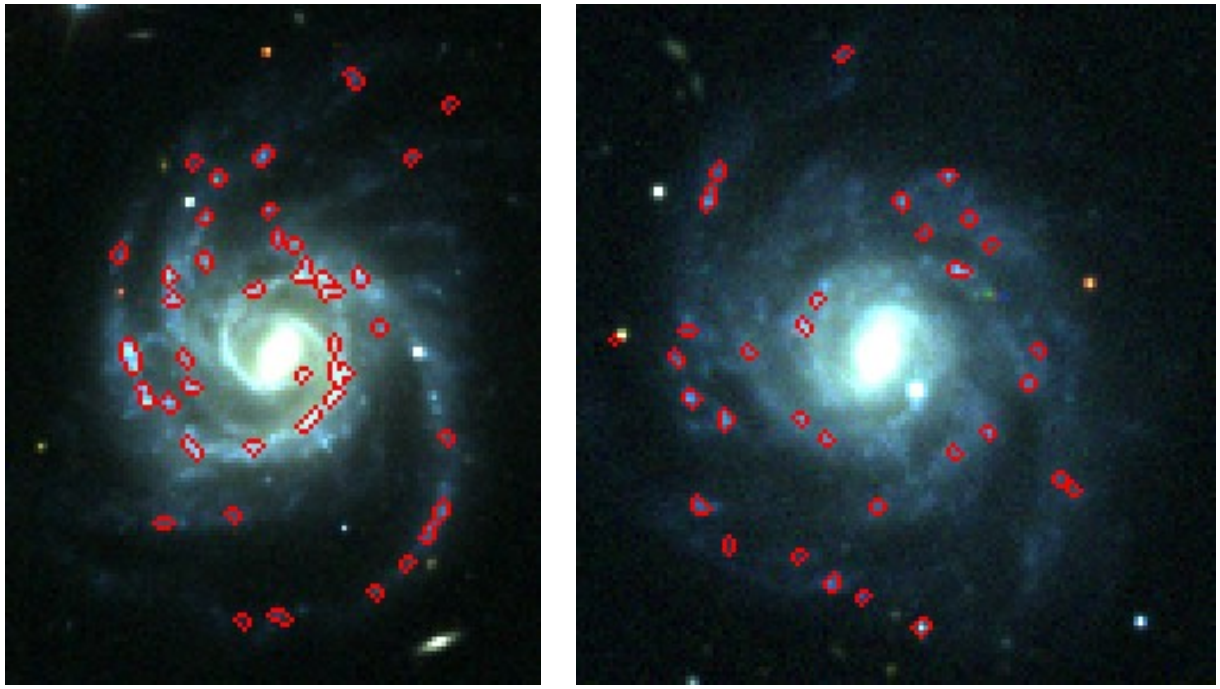




TREE OF CONNECTED COMPONENTS

detection of HII regions

relevant attributes for their identification : flux size, color

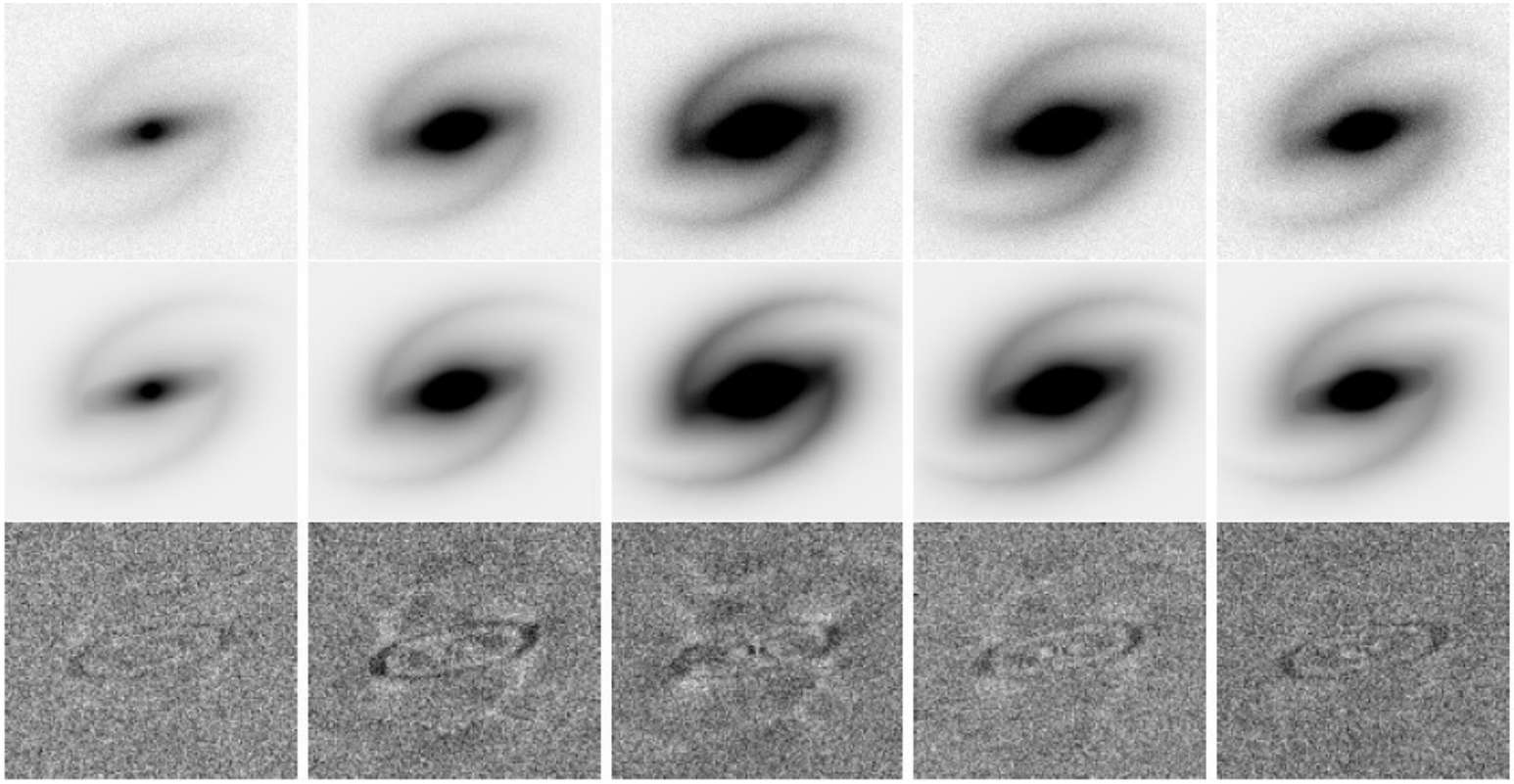


constraint on the model : the spiral arms must pass through a majority of such HII regions

OUTLINE

I*ntroduction*
Parametric models
Mathematical morphology
Results on the EFIGI database
Perspectives

RESULTS on Simulations



Quality : pretty good, but the errors are underestimated

RESULTS on the EFIGI dataset

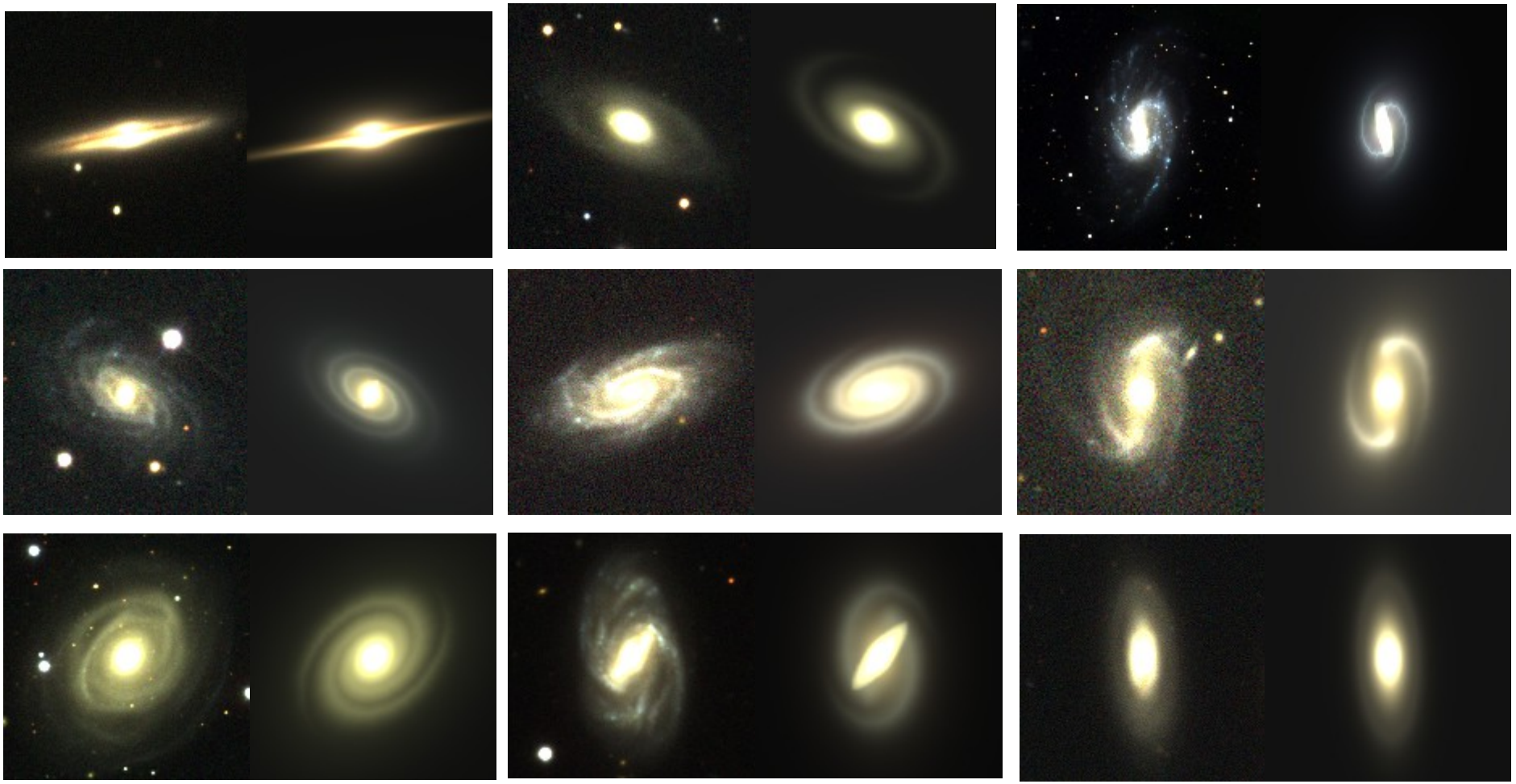
EFIGI : – about 5 000 galaxies in five SDSS photometric bands
– rescaling, sky subtracted, PSF correction
– 16 morphological attributes visually valued by experts
– confidence interval for each attribute

TEST : 1 500 « regular » galaxies processed
no interacting galaxy, no strong perturbations

> the relevant model is *chosen* a.t the EFIGI classif. Code <

RESULTS on the EFIGI dataset

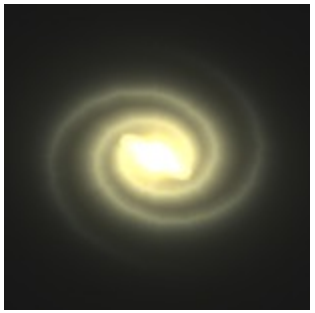
some examples



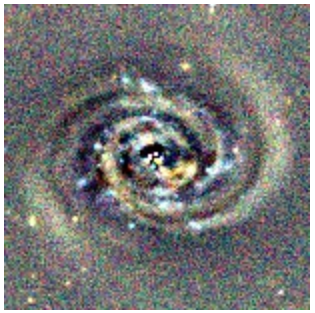
RESULTS

specific case of PGC 2182

image

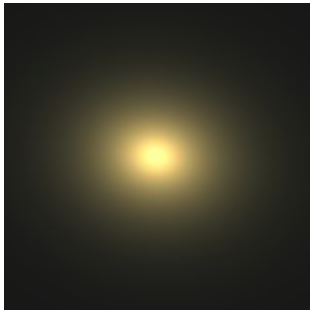


model



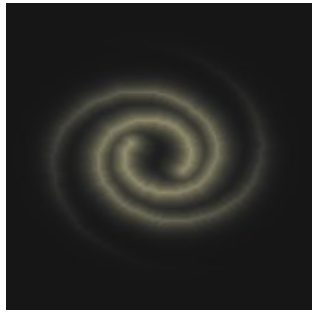
residual

bulge



disk

bar



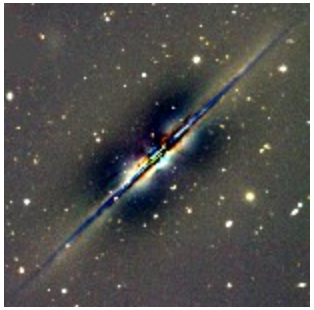
spiral arms

RESULTS

specific case of PGC 218

image

model



residual

bulge

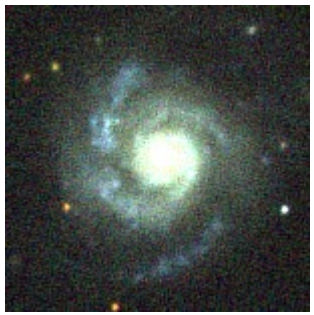
« halo »

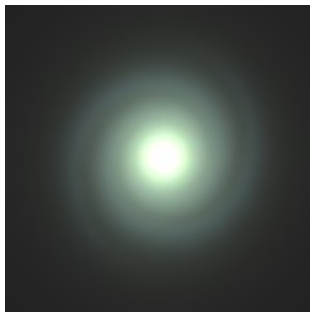
disk

dust obscuration

DIFFICULTIES

When the model noise is larger than the flux of a structure ...





wrong direction for the
spiral qrms

+ discontinuous structure ...

Pending issues :

– multiple solutions → priors to constraint the parameter values
– 2 structures may still be required to model a single physical structure
– failure diagnostic : energy of the residual vs. values of the parameters
– no engine yet to navigate automatically across the hierarchy of models
– large computation time : Intel 2 core @ 2.6 GHz
i) 600–700 iter. for a three models estimation : 15%, 25%, 60%
ii) 2–3 hr ! → massive parallelisation of the likelihood comput.

COMPARISON with EFIGI information

Difficulties : – what does the expert evaluate exactly ?
– how to link qualitative estimate with numerical values ?

Success rate : estimate within the EFIGI 70% confidence interval

– B/T ratio : 80.0 %
– inclination : 86.5 %
– size of the bar : 75.0 %
– arm strength : 60.0 %
– arm opening : 81.0 %
– rotation direction : 83.0 %

OUTLINE

Introduction
Parametric models
Mathematical morphology
Results on the EFIGI database
Perspectives

CONCLUSIONS – PERSPECTIVES

Multiband decomposition of galaxies into physically meaningful morphological structures

– a hierarchy of more and more detailed models :

– a quasi automated method
– validation on real data

– unsupervised selection of the right model
– optimisation of the computational cost

– complexity vs. robust and unique solution (SNR, discontinuities)
– other wavelengths