Galaxy Decomposition in Multispectral Images Using Markov Chain Monte Carlo Algorithms

Benjamin Perret[1], Vincent Mazet[1] Christophe Collet[1], and Éric Slezak[2]

[1] LSIIT (UMR CNRS-Université de Strasbourg 7005), France
{perret,mazet,collet}@lsiit.u-strasbg.fr
[2] Laboratoire Cassiopée (UMR CNRS-Observatoire de la Côte d'Azur 6202), France
eric.slezak@oca.eu

Abstract. Astronomers still lack a multiwavelength analysis scheme for galaxy classification. In this paper we propose a way of analysing multispectral observations aiming at refining existing classifications with spectral information. We propose a global approach which consists of decomposing the galaxy into a parametric model using physically meaningful structures. Physical interpretation of the results will be straightforward even if the method is limited to regular galaxies. The proposed approach is fully automatic and performed using Markov Chain Monte Carlo (MCMC) algorithms. Evaluation on simulated and real 5-band images shows that this new method is robust and accurate.

Key words: Bayesian inference, MCMC, multispectral image processing, galaxy classification.

1 Introduction

Galaxy classification is a necessary step in analysing and then understanding the evolution of these objects in relation to their environment at different spatial scales. Current classifications rely mostly on the De Vaucouleurs scheme [1] which is an evolution of the original idea by Hubble. These classifications are based only on the visible aspect of galaxies and identifies five major classes: ellipticals, lenticulars, spirals with or without bar, and irregulars. Each class is characterized by the presence, with different strengths, of physical structures such as a central bright bulge, an extended fainter disc, spiral arms, ... and each class and the intermediate cases are themselves divided into finer stages.

Nowadays wide astronomical image surveys provide huge amount of multiwavelength data. For example, the Sloan Digital Sky Survey (SDSS[3]) has already produced more than 15 Tb of 5-band images. Nevertheless, most classifications still do not take advantage of colour information, although this information gives important clues on galaxy evolution allowing astronomers to estimate the star formation history, the current amount of dust, etc. This observation motivates the research of a more efficient classification including spectral information over

[3] http://www.sdss.org/

all available bands. Moreover due to the quantity of available data (more than 930,000 galaxies for the SDSS), it appears relevant to use an automatic and unsupervised method.

Two kinds of methods have been proposed to automatically classify galaxies following the Hubble scheme. The first one measures galaxy features directly on the image (*e.g.* symmetry index [2], Pétrosian radius [3], concentration index [4], *clumpiness* [5], ...). The second one is based on decomposition techniques (shapelets [6], the basis extracted with principal component analysis [7], and the pseudo basis modelling of the physical structures: bulge and disc [8]). Parameters extracted from these methods are then used as the input to a traditional classifier such as a support vector machine [9], a multi layer perceptron [10] or a Gaussian mixture model [6].

These methods are now able to reach a good classification efficiency (equal to the experts' agreement rate) for major classes [7]. Some attempts have been made to use decomposition into shapelets [11] or feature measurement methods [12] on multispectral data by processing images band by band. Fusion of spectral information is then performed by the classifier. But the lack of physical meaning of data used as inputs for the classifiers makes results hard to interpret. To avoid this problem we propose to extend the decomposition method using physical structures to multiwavelength data. This way we expect that the interpretation of new classes will be straightforward.

In this context, three 2D galaxy decomposition methods are publicly available. Gim2D [13] performs bulge and disc decomposition of distant galaxies using MCMC methods, making it robust but slow. Budda [14] handles bulge, disc, and stellar bar, while Galfit [15] handles any composition of structures using various brightness profiles. Both of them are based on deterministic algorithms which are fast but sensitive to local minima. Because these methods cannot handle multispectral data, we propose a new decomposition algorithm. This works with multispectral data and any parametric structures. Moreover, the use of MCMC methods makes it robust and allows it to work in a fully automated way.

The paper is organized as follows. In Sec. 2, we extend current models to multispectral images. Then, we present in Sec. 3 the Bayesian approach and a suitable MCMC algorithm to estimate model parameters from observations. The first results on simulated and raw images are discussed in Sec. 4. Finally some conclusions and perspectives are drawn in Sec. 5.

2 Galaxy Model

2.1 Decomposition into Structures

It is widely accepted by astronomers that spiral galaxies for instance can be decomposed into physically significant structures such as bulge, disc, stellar bar and spiral arms (Fig. 4, first column). Each structure has its own particular shape, populations of stars and dynamic. The bulge is a spheroidal population of mostly old red stars located in the centre of the galaxy. The disc is a planar

structure with different scale heights which includes most of the gas and dust if any and populations of stars of various ages and colour from old red to younger and bluer ones. The stellar bar is an elongated structure composed of old red stars across the galaxy centre. Finally, spiral arms are over-bright regions in the disc that are the principal regions of star formation. The visible aspect of these structures are the fundamental criterion in the Hubble classification. It is noteworthy that this model only concerns regular galaxies and that no model for irregular or peculiar galaxies is available.

We only consider in this paper bulge, disc, and stellar bar. Spiral arms are not included because no mathematical model including both shape and brightness informations is available; we are working at finding such a suitable model.

2.2 Structure Model

We propose in this section a multispectral model for bulge, disc, and stellar bar. These structures rely on the following components: a generalized ellipse (also known as super ellipse) is used as a shape descriptor and a Sérsic law is used for the brightness profile [16]. These two descriptors are flexible enough to describe the three structures.

The major axis r of a generalized ellipse centred at the origin with axis parallel to coordinate axis and passing trough point $(x, y) \in \mathbb{R}^2$ is given by:

$$r\,(x,y) = \left(|x|^{c+2} + \left|\frac{y}{e}\right|^{c+2} \right)^{\frac{1}{c+2}} \tag{1}$$

where e is the ratio of the minor to the major axis and c controls the misshapenness: if $c = 0$ the generalized ellipse reduces to a simple ellipse, if $c < 0$ the ellipse is said to be *disky* and if $c > 0$ the ellipse is said to be *boxy* (Fig. 1). Three more parameters are needed to complete shape information: the centre (c_x, c_y) and the position angle α between abscissa axis and major axis.

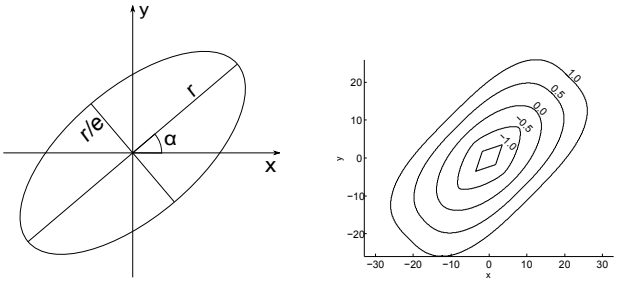

Fig. 1. Left: a simple ellipse with position angle α, major axis r and minor axis r/e. Right: generalized ellipse with variations of parameter c (displayed near each ellipse).

The Sérsic law [16] is generally used to model the brightness profile. It is a generalization of the traditional exponential and De Vaucouleurs laws usually

used to model disc and bulge brightness profiles. Its high flexibility allows it to vary continuously from a nearly flat curve to a very piked one (Fig. 2). The brightness at major axis r is given by:

$$I(r) = I\, \mathrm{e}^{-k_n \left(\left(\frac{r}{R}\right)^{\frac{1}{n}} - 1 \right)} \tag{2}$$

where R is the effective radius, n is the Sérsic index, and I the brightness at the effective radius. k_n is an auxiliary function such that $\Gamma(2n) = 2\gamma(2n, k_n)$ to ensure that half of the total flux is contained in the effective radius (Γ and γ are respectively the complete and incomplete gamma function).

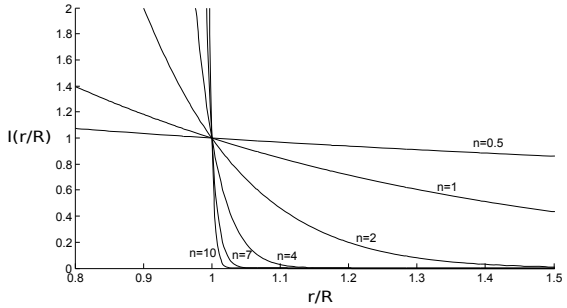

Fig. 2. The Sérsic law for different Sérsic index n. $n = 0.5$ yields a Gaussian, $n = 1$ yields an exponential profile and for $n = 4$ we obtain the De Vaucouleurs profile.

Then, the brightness at pixel (x, y) is given by:

$$F(x, y) = (F_1(x, y), \ldots, F_B(x, y)) \tag{3}$$

with B the number of bands and the brightness in band b is defined as:

$$F_b(x, y) = I_b\, \mathrm{e}^{-k_{n_b} \left(\left(\frac{r(x,y)}{R_b}\right)^{\frac{1}{n_b}} - 1 \right)} \tag{4}$$

As each structure is supposed to represent a particular population of stars and galactic environment, we also assume that shape parameters do not vary between bands. This strong assumption seems to be verified in observations suggesting that shape variations between bands is negligible compared with deviation induced by noise. Moreover, this assumption reduces significantly the number of unknowns. The stellar bar has one more parameter which is the cut-off radius R_{max}; its brightness is zero beyond this radius. For the bulge (respectively the stellar bar), all Sérsic parameters are free which leads to a total number of $5+3B$ (respectively $6+3B$) unknowns. For the disc, parameter c is set to zero and Sérsic index is set to one leading to $4+2B$ free parameters. Finally, we assume that the centre is identical for all structures yielding a total of $11 + 8B$ unknowns.

2.3 Observation Model

Atmospheric distortions can be approximated by a spatial convolution with a Point Spread Function (PSF) H given as a parametric function or an image. Other noises are a composition of several sources and will be approximated by a Gaussian noise $\mathcal{N}(0, \Sigma)$. Matrix Σ and PSF H are not estimated as they can be measured using a deterministic procedure. Let Y be the observations and e the noise, we then have:

$$Y = Hm + e \qquad \text{with} \qquad m = F_{\mathfrak{B}} + F_{\mathfrak{D}} + F_{\mathfrak{B}a} \tag{5}$$

with $\mathfrak{B}$, $\mathfrak{D}$, and $\mathfrak{B}a$ denoting respectively the bulge, the disc, and the stellar bar.

3 Bayesian Model and Monte Carlo Sampling

The problem being clearly ill-posed, we adopt a Bayesian approach. Priors assigned to each parameter are summarized in Tab. 1; they were determined from literature when possible and empirically otherwise. Indeed experts are able to determine limits for parameters but no further information is available: that is why Probability Density Functions (pdf) of chosen priors are uniformly distributed. However we expect to be able to determine more informative priors in future work. The posterior reads then:

$$P(\phi|Y) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(\Sigma)^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(Y-Hm)^T \Sigma^{-1} (Y-Hm)} P(\phi) \tag{6}$$

where $P(\phi)$ denotes the priors and ϕ the unknowns. Due to its high dimensionality it is intractable to characterize the posterior pdf with sufficient accuracy. Instead, we aim at finding the Maximum A Posteriori (MAP).

Because of the posterior complexity, the need for a robust algorithm leads us to choose MCMC methods [17]. MCMC algorithms are proven to converge in infinite time, and in practice the time needed to obtain a good estimation may be quite long. Thus several methods are used to improve convergence speed: simulated annealing, adaptive scale [18] and direction [19] Hastings Metropolis (HM) algorithm. As well, highly correlated parameters like Sérsic index and radius are sampled jointly to improve performance.

The main algorithm is a Gibbs sampler consisting in simulating variables separately according to their respective conditional posterior. One can note that the brightness factors posterior reduces to a truncated positive Gaussian $\mathcal{N}^+(\mu, \sigma^2)$ which can be efficiently sampled using an accept-reject algorithm [20]. Other variables are generated using the HM algorithm.

Some are generated with a Random Walk HM (RWHM) algorithm whose proposal is a Gaussian. At each iteration a random move from the current value is proposed. The proposed value is accepted or rejected with respect to the posterior ratio with the current value. The parameters of the proposal have been chosen by examining several empirical posterior distributions to find preferred directions

and optimal scale. Sometimes the posterior is very sensitive to input data and no preferred directions can be found. In this case we decided to use the Adaptive Direction HM (ADHM). ADHM algorithm uses a sample of already simulated points to find preferred directions. As it needs a group of points to start with we choose to initialize the algorithm using simple RWHM. When enough points have been simulated by RWHM, the ADHM algorithm takes over. Algorithm and parameters of proposal distributions are summarized in Tab. 1.

Table 1. Parameters and their priors. All proposal distributions are Gaussians whose covariance matrix (or deviation for scalars) are given in the last column.

Structure	Parameter	Prior Support	Algorithm
$\mathfrak{B}$, $\mathfrak{B}a$, $\mathfrak{D}$	centre (c_x, c_y)	Image domain	RWHM with $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\mathfrak{B}$	major to minor axis (e)	$[1;10]$	RWHM with 1
	position angle (α)	$[0;2\pi]$	RWHM with 0.5
	ellipse misshapenness (c)	$[-0.5;1]$	RWHM with 0.1
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+(\mu, \sigma^2)$
	radius (R)	$[0;200]$	ADHM with $\begin{pmatrix} 0.16 & -0.02 \\ -0.02 & 0.01 \end{pmatrix}$
	Sérsic index (n)	$[1;10]$	
$\mathfrak{D}$	major to minor axis (e)	$[1;10]$	RWHM with 0.2
	position angle (α)	$[0;2\pi]$	RWHM with 0.5
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+(\mu, \sigma^2)$
	radius (R)	$[0;200]$	RWHM with 1
$\mathfrak{B}a$	major to minor axis (e)	$[4;10]$	RWHM with 1
	position angle (α)	$[0;2\pi]$	RWHM with 0.5
	ellipse misshapenness (c)	$[0.6;2]$	RWHM with 0.1
	brightness factor (I)	$\mathbb{R}^+$	direct with $\mathcal{N}^+(\mu, \sigma^2)$
	radius (R)	$[0;200]$	ADHM with $\begin{pmatrix} 0.16 & -0.02 \\ -0.02 & 0.01 \end{pmatrix}$
	Sérsic index (n)	$[0.5;10]$	
	cut-off radius (R_{max})	$[10;100]$	RWHM with 1

Also, parameters I_b, R_b, and n_b are jointly simulated. R_b, n_b are first sampled according to $P\left(R_b, n_b \mid \phi_{\setminus\{R_b,n_b,I_b\}}\right)$ where I_b has been integrated and then I_b is sampled [21]. Indeed, the posterior can be decomposed in:

$$P\left(R_b, n_b, I_b \mid \phi_{\setminus\{R_b,n_b,I_b\}}, Y\right) = P\left(R_b, n_b \mid \phi_{\setminus\{R_b,n_b,I_b\}}, Y\right) P\left(I_b \mid \phi_{\setminus\{I_b\}}, Y\right) \tag{7}$$

4 Validation and Results

We measured two values for each parameter: the MAP and the variance of the chain in the last iterations. The latter gives an estimation of the uncertainty on the estimated value. A high variance can have different interpretations. In case of an observation with a low SNR, the variance naturally increases. But the variance can also be high when a parameter is not relevant. For example, the position angle is significant if the structure is not circular, the radius is also significant if the brightness is strong enough. We have also checked visually the residual image (the difference between the observation and the simulated image) which should contain only noise and non modelled structures.

Parameters are initialized by generating random variables according to their priors. This procedure ensures that the algorithm is robust so that it will not be fooled by a bad initialisation, even if the burn-in period of the Gibbs sampler is quite long (about 1,500 iterations corresponding to 1.5 hours).

4.1 Test on Simulated Images

We have validated the procedure on simulated images to test the ability of the algorithm to recover input parameters. The results showed that the algorithm is able to provide a solution leading to a residual image containing only noise (Fig. 3). Some parameters like elongation, position angle, or centre are retrieved with a very good precision (relative error less than 0.1%). On the other hand, Sérsic parameters are harder to estimate. Thanks to the extension of the disc, its radius and its brightness are estimated with a relative error of less than 5%. For the bulge and the stellar bar, the situation is complex because information is held by only a few pixels and an error in the estimation of Sérsic parametres does not lead to a high variation in the likelihood. Although the relative error increases to 20%, the errors seem to compensate each other.

Another problem is the evaluation of the presence of a given structure. Because the algorithm seeks at minimizing the residual, all the structures are always used. This can lead to solutions where structures have no physical significance. Therefore, we tried to introduce a Bernoulli variable coding the structure occurrence. Unfortunately, we were not able to determine a physically significant Bernoulli parameter. Instead we could use a pre- or post-processing method to determine the presence of each structure. These questions are highly linked to the astrophysical meaning of the structures we are modelling and we have to ask ourselves why some structures detected by the algorithm should in fact not be used. As claimed before, we need to define more informative joint priors.

4.2 Test on Real Images

We have performed tests on about 30 images extracted from the EFIGI database [7] which is composed of thousands of galaxy images extracted from the SDSS. Images are centred on the galaxy but may contain other objects (stars, galaxies, artefacts, ...). Experiments showed that the algorithm performs well as long as

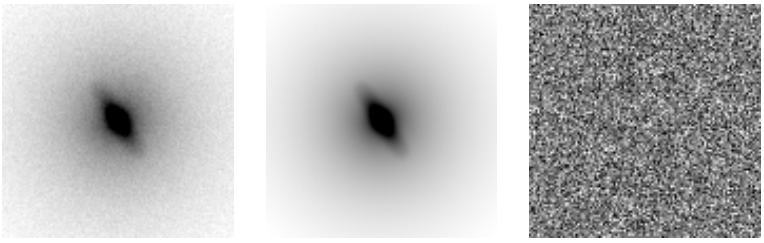

Fig. 3. Example of estimation on a simulated image (only one band on five is shown). Left: simulated galaxy with a bulge, a disc and a stellar bar. Centre: estimation. Right: residual. Images are given in inverse gray scale with enhanced contrast.

no other bright object is present in the image (see Fig. 4 for example). As there is no ground truth available on real data we compared the results of our algorithm on monospectral images with those provided by Galfit. This shows a very good agreement since Galfit estimations are within the confidence interval proposed by our method.

4.3 Computation Time

Most of the computation time is used to evaluate the likelihood. Each time a parameter is modified, this implies the recomputation of the brightness of each affected structure for all pixels. Processing 1,000 iterations on a 5-band image of 250×250 pixels takes about 1 hour with a Java code running on an Intel Core 2 processor (2,66 GHz). We are exploring several ways to improve performance such as providing a good initialisation using fast algorithms or finely tuning the algorithm to simplify exploration of the posterior pdf.

5 Conclusion

We have proposed an extension of the traditional bulge, disc, stellar bar decomposition of galaxies to multiwavelength images and an automatic estimation process based on Bayesian inference and MCMC methods. We aim at using the decomposition results to provide an extension of the Hubble's classification to multispectral data. The proposed approach decomposes multiwavelength observations in a global way. The chosen model relies on some physically significant structures and can be extended with other structures such as spiral arms. In agreement with the experts, some parameters are identical in every band while others are specific to each band. The algorithm is non-supervised in order to obtain a fully automatic method. The model and estimation process have been validated on simulated and real images.

We are currently enriching the model with a parametric multispectral description of spiral arms. Other important work being carried out with experts is to determine joint priors that would ensure the significance of all parameters.

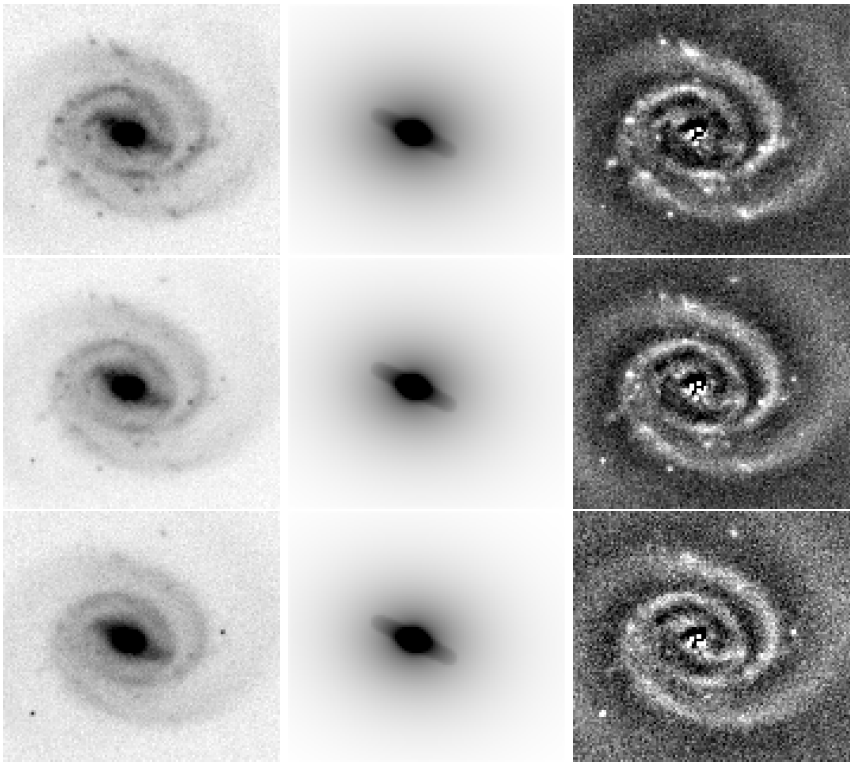

Fig. 4. Left column: galaxy PGC2182 (bands g, r, and i) is a barred spiral. Centre column: estimation. Right column: residual. Images are given in inverse gray scale with enhanced contrast.

Finally we are looking for an efficient initialisation procedure that would greatly increase convergence speed and open the way to a fast and fully unsupervised algorithm for multiband galaxy classification.

Acknowledgements

We would like to thank É. Bertin from the Institut d'Astrophysique de Paris for giving us a full access to the EFIGI image database.

References

1. De Vaucouleurs, G.: Classification and Morphology of External Galaxies. Handbuch der Physik 53, 275 (1959)
2. Yagi, M., Nakamura, Y., Doi, M., Shimasaku, K. and Okamura, S.: Morphological classification of nearby galaxies based on asymmetry and luminosity concentration. Monthly Notices of Roy. Astr. Soc. 368, 211–220 (2006)

3. Petrosian, V.: Surface brightness and evolution of galaxies. Astrophys. J. Letters 209, L1–L5 (1976)
4. Abraham, R. G., Valdes, F., Yee, H. K. C. and van den Bergh, S.: The morphologies of distant galaxies. 1: an automated classification system. Astrophys. J. 432, 75–90 (1994)
5. Conselice, C. J.: The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories. Astrophys. J. Suppl. S. 147, 1–28 (2003)
6. Kelly, B. C. and McKa, T. A.: Morphological Classification of Galaxies by Shapelet Decomposition in the Sloan Digital Sky Survey. Astron. J. 127, 625–645 (2004)
7. Baillard, A., Bertin, E., Mellier, Y., McCracken, H. J., Géraud, T., Pelló, R., Leborgne, F. and Fouqué, P.: Project EFIGI: Automatic Classification of Galaxies. Astron. Soc. Pac. Conf. ADASS XV 351, 236 (2006)
8. Allen, P. D., Driver, S. P., Graham, A. W., Cameron, E., Liske, J. and de Propris, R.: The Millennium Galaxy Catalogue: bulge-disc decomposition of 10095 nearby galaxies. Monthly Notices of Roy. Astr. Soc. 371, 2–18 (2006)
9. Tsalmantza, P., Kontizas, M., Bailer-Jones, C. A. L., Rocca-Volmerange, B., Korakitis, R., Kontizas, E., Livanou, E., Dapergolas, A., Bellas-Velidis, I., Vallenari, A. and Fioc, M.: Towards a library of synthetic galaxy spectra and preliminary results of classification and parametrization of unresolved galaxies for Gaia: Astron. Astrophys. 470, 761–770 (2007)
10. Bazell, D.: Feature relevance in morphological galaxy classification. Monthly Notices of Roy. Astr. Soc. 316, 519–528 (2000)
11. Kelly, B. C. and McKay, T. A.: Morphological Classification of Galaxies by Shapelet Decomposition in the Sloan Digital Sky Survey. II. Multiwavelength Classification. Astron. J. 129, 1287–1310 (2005)
12. Lauger, S., Burgarella, D. and Buat, V.: Spectro-morphology of galaxies: A multiwavelength (UV-R) classification method. Astron. Astrophys. 434, 77-87 (2005)
13. Simard, L., Willmer, C. N. A., Vogt, N. P., Sarajedini, V. L., Phillips, A. C., Weiner, B. J., Koo, D. C., Im, M., Illingworth, G. D. and Faber, S. M.: The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. Astrophys. J. Suppl. S. 142, 1–33 (2002)
14. de Souza, R. E., Gadotti, D. A. and dos Anjos, S.: BUDDA: A New Two-dimensional Bulge/Disk Decomposition Code for Detailed Structural Analysis of Galaxies. Astrophys. J. Suppl. S. 153, 411–427 (2004)
15. Peng, C. Y., Ho, L. C., Impey, C. D. and Rix, H.-W.: Detailed Structural Decomposition of Galaxy Images. Astron. J. 124, 266–293 (2002)
16. Sérsic, J. L.: Atlas de galaxias australes. Cordoba, Argentina: Observatorio Astronomico (1968)
17. Gilks, W.R., Richardson, S. and Spiegelhalter, D.J.: Markov Chain Monte Carlo In Practice. Chapman & Hall/CRC, Washington, D.C. (1996)
18. Gilks, W. R., Roberts, G. O. and Sahu, S. K.: Adaptive Markov chain Monte Carlo through regeneration. J. Amer. Statistical Assoc. 93, 1045–1054 (1998)
19. Roberts, G. O. and Gilks, W. R.: Convergence of adaptive direction sampling. J. of Multivariate Ana. 49, 287–298 (1994)
20. Mazet, V., Brie, D. and Idier, J.: Simulation of positive normal variables using several proposal distributions. IEEE Workshop on Statistical Sig. Proc, 37–42 (2005)
21. Devroye, L.: Non-Uniforme Random Variate Generation. Springer-Verlag, New York (1986)