

# Morphological evolution from $z \sim 2$ in the COSMOS field from Ks-band imaging

M. Huertas-Company, L. Tasca, D. Rouan, J.P. Kneib and O. Le Fèvre

**Abstract** Morphology is the most accessible tracer of galaxies physical structure, but its interpretation in the framework of galaxy evolution still remains a problem. Its quantification at high redshift requires deep high angular resolution imaging, that is why space data (HST) are normally employed. At  $z > 1$ , the HST visible cameras probe however the UV flux, dominated by the emission of young stars, which could bias the estimated morphologies towards late-type systems. In this paper we quantify the effects of this *morphological k-correction* at  $1 < z < 2$  by comparing morphologies measured in the K and I-bands in the COSMOS area. Ks-band data have indeed the advantage of probing old stellar populations in the rest-frame for  $z < 2$ , enabling a determination of galaxy morphological types unaffected by recent star formation. We employ a new non-parametric method based on SVM to classify  $\sim 50000$  Ks selected galaxies in the COSMOS area observed with WIRCam at CFHT. We use a 10-dimensional volume, including 5 morphological parameters, and other characteristics of galaxies such as luminosity and redshift. The classification is globally in good agreement with the one obtained using HST/ACS for  $z < 1$ . Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$  which might be a consequence of morphological k-correction effects. We argue therefore that studies based on I-band HST/ACS classifications at  $z > 1$  could be underestimating the elliptical population.

---

M. Huertas-Company, D. Rouan  
LESIA, Observatoire de Paris, CNRS, UPMC, Universit Paris Diderot, 5 Place Jules Janssen,  
92195, Meudon, France e-mail: marc.huertas@obspm.fr

L.Tasca, J.P. Kneib, O. Le Fèvre  
LAM, CNRS-Université de Provence, 38, rue Frédéric Joliot-Curie, 13388 Marseille cedex 13,  
France

## 1 Introduction

In the local Universe, the distribution of galaxies is bimodal, primarily reflecting a relationship between color and morphology. On the one hand the spiral-like galaxies are gas-rich, form stars and are supported by the rotation of their stars and on the other hand the elliptical-like galaxies are gas-poor, do not form stars anymore and are supported by the velocity dispersion of the stars. This is the so-called elliptical-spiral Hubble sequence. A fundamental question in observational cosmology is how this bimodality appears across the history of the Universe. In the visible, progress over the last ten years have come in particular from the Hubble Deep Fields (HDF) observed with the Hubble Space Telescope. They brought observational evidence that galaxy evolution is differentiated with respect to morphological type and that a large fraction of distant galaxies have peculiar morphologies that do not fit into the elliptical-spiral Hubble sequence. There is some evidence indeed that most of the stellar mass assembly is performed around  $z \sim 1.5 - 2$ . A better understanding of the physical processes that lead to the present Hubble sequence should come therefore from observations in this redshift range. In this context, near infrared observations are particularly important because the Ks-band flux at  $z \sim 1$  is less dependent on the recent history of star formation, which peaks in the UV in rest frame, and gives thus a galaxy type from the distribution of old stars, more closely related to the underlying total mass than optical observations. In (1) we proposed therefore a generalization of the non-parametric morphological classification methods that uses an unlimited number of dimensions and non-linear separators, enabling us to use all the information brought by the different morphological parameters simultaneously. In this paper, we use this method to quantify the morphologies of  $\sim 50000$  galaxies based on structural parameters measured in the near infrared. Galaxies are observed in the COSMOS field with WIRCam at CFHT in Ks-band.

## 2 The data

**Description:** The Ks-band data were taken with WIRCam installed at CFHT in the near infrared  $K_s$  band ( $2.2\mu m$ ). The field is part of the Canada-France Hawaii Telescope Legacy Survey (CFHTLS) Deep survey and its near infrared follow-up, and covers the entire COSMOS area ( $1.4^{(o)^2}$ ) (2). Images are reduced with the Terapix pipeline<sup>1</sup> and have a pixel scale of  $0.15''$  with a mean FWHM of  $0.7''$ . A more detailed description of the data set can be found in McCracken et al. 2008 (in preparation). The I-band data used in the last section are part of the COSMOS HST/ACS field (3). The data set consists of a contiguous  $1.64 \text{ deg}^2$  field covering the entire COSMOS field. The Advanced Camera for Surveys (ACS) together with the F814W filter (“Broad I”) were employed. More precisely, in this paper we will use the morphological catalogue described in Tasca et al. 2008 (in preparation).

---

<sup>1</sup> <http://terapix.iap.fr>

**Photometric redshift catalogue:** We used the code *Le Phare* to measure the photometric redshift of 278 526 objects in the COSMOS field with  $I_{AB} < 25$  and multi-wavelength information. Our final catalogue has 198 684 objects with reliable photometric redshift measurements.

**The morphological catalogue:** The morphological analysis is made in a subsample of the initial catalogue: first we select only the galaxies which have a measured photometric redshift. Then we cut the catalogue to  $K_s < 21.5$  and  $z_{phot} < 2$ . This decision is based in a visual inspection; objects fainter than 21.5 have a S/N per pixel lower than 5, so we decided not to include them in the morphological study. Simulations of those objects show indeed that the morphological classifications obtained are highly contaminated. Photometric redshift above  $z \sim 2$  are not reliable enough. This selection results in a final morphological catalogue of 44 089 galaxies. The sample is complete up to  $K_s = 21.5$  for both bulges and disks for sizes ranging from  $0''$  to  $1.5''$  (simulations).

### 3 Morphology

Galaxies in the catalogue have been separated into two main morphological types (late-type and early-type) using the free available code galSVM<sup>2</sup> (1). galSVM is a non-parametric N-dimensional code based on support vector machines (SVM) that uses a training set built from a local visually classified sample. The employed procedure can be summarized in 4 main steps: a) Build a training set: we select a nearby visually classified sample at wavelengths corresponding to the rest-frame of the high redshift sample to be analyzed. In our case, we want to simulate Ks-band observations. We used therefore an SDSS local sample observed in the i and z bands which roughly corresponds to the rest-frame wavelengths between  $z \sim 1$  and  $z \sim 2$  for Ks-band observations. We then move the sample to the proper redshift and image quality and drop it in the real background. b) Measure a set of morphological parameters on the sample. c) Train a support vector based learning machine with a fraction of the simulated sample and use the other fraction to test and estimate errors. d) Classify real data with the trained machine and correct for possible systematic errors detected in the testing step.

**The training sample:** This is a critical step since it will determine the decision regions that will be used to perform the classification. Indeed, galaxy morphology derivation depends on the physical properties of the galaxy (luminosity, redshift, wavelength) and on the observing conditions (background level, resolution). Here, we then decided to simulate the high redshift sample from a visually classified local catalog, selected in the rest-frame color of the high redshift sample. We use a catalog of 1319 objects from the SDSS observed in two photometric bands (z and i) and visually classified. As explained in (1), for every galaxy stamp we first generate a random pair of (magnitude, redshift) values with a probability distribution that

---

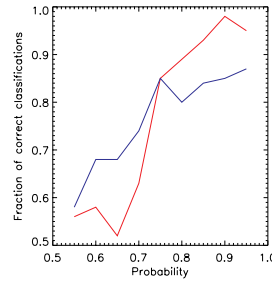
<sup>2</sup> <http://www.lesia.obspm.fr/huertas/galsvm.html>

matches the real magnitude and redshift distribution of the sample to be simulated and then we proceed in four steps: a) removal of foreground stars, b) degradation of the resolution according to a  $\Lambda$ CDM cosmology, c) binning to reach the desired pixel scale and d) dropping in a real background image.

**Classification procedure:** We use 450 simulated galaxies as training sample. The morphological mixing is fixed to 50/50, i.e. 50% of early-type galaxies and 50% of late-type. Even if this is not a realistic distribution it is required to minimize the errors in the SVM classification. The classification is made in a 10-D volume with a Radial Basis Function Kernel (see (1) for more details).

**The output catalogue:** We associate to every galaxy in the morphological catalogue, a class label and a probability of belonging to the given class. Since we are dealing with a 2-class problem, the probability  $p(\text{galaxy}=\text{early-type})=1-p(\text{galaxy}=\text{late-type})$ . In the following, we use this parameter to assess the accuracy of our classification.

**Accuracy:** Figure 1 shows the evolution of the correct classifications as a function of different probability thresholds using the test sample. As we can see, there is a clear correlation between the probability threshold and the number of correct identifications: the accuracy clearly increases when the considered probability is higher. However, objects with probabilities greater than 0.8 are classified with nearly 90%



**Fig. 1** Number of correct classifications as a function of probability for the test sample. The red line is for early-type galaxies and the blue one for late-type.

accuracy. The contamination is around  $\sim 20\%$  for the whole sample ( $p > 0.5$ ).

This relations between the probability parameter and the success rate enable the use of the probability as an estimate of the classification accuracy as a function of physical parameters of the galaxies directly on the real sample. We observe several interesting trends:

**a)** Globally, late-type galaxies have higher probabilities for all redshift, magnitude and size values. **b)** Small objects ( $r_{half} < 0.6''$ ) have lower probabilities ( $p \sim 0.7$ ) while large ones ( $r_{half} > 0.6''$ ) have higher values ( $p > 0.8$ ). **c)** Below  $z \sim 1$  the mean probability is around  $p \sim 0.75$  and it decreases to  $p \sim 0.7$  for  $z > 1$ . The number of galaxies with low probability ( $p \sim 0.6$ ) also increases. **d)** Above  $K_s \sim 20$  the mean probability for early-type objects is around  $p \sim 0.7$ .

## 4 Investigating the morphological k-correction effect

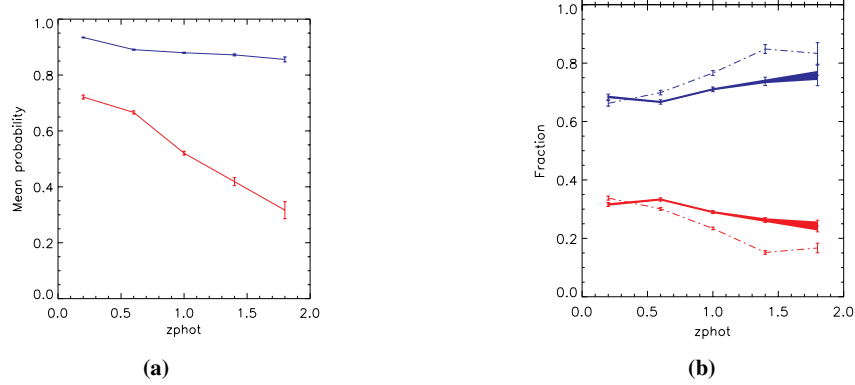
We perform a match between the Ks selected morphological catalogue and the morphologies measured from HST/ACS data (I-band) in an independent way. We match 34 165 objects of our K-selected sample.

**One-to-one comparison:** Figure 2a shows the mean probability for a galaxy classified as early (late) in the I-band to be classified as early (late) in the Ks-band as a function of redshift. We observe, that globally, there is no ambiguity in the identification of late-type objects. A galaxy which is found to be late-type in the I-band is also a late-type system in the Ks-band with a probability around  $p \sim 0.9$  up to  $z \sim 2$ . For early-type objects, there is clear trend with redshift: below  $z \sim 0.8$  a galaxy classified as early-type in the I-band has a probability of  $p \sim 0.7$  of being early-type in the Ks-band. Above  $z \sim 1$  the discrepancies between the two classifications become higher and an elliptical galaxy is classified as early in both classifications only with a probability of  $p \sim 0.4$ . Above  $z \sim 1$  the I-band filter starts to probe the UV flux and therefore the obtained morphology is determined from young stars whereas the Ks-band filter stills probe the visible spectra of the galaxy. What are then the differences of morphologically selecting a sample in the Ks-band or in the I-band above  $z \sim 1$ ?

**Comparing redshift distributions:** Figure 2b shows the redshift distributions obtained from both classifications. The distributions for late-type objects present small differences as expected from the comparison performed in the previous section. A Kolmogorov-Smirnov test (KS-test) reveals that the two functions arise from the same statistic with 96% confidence. The early-type redshift distributions present more differences, as expected from the comparisons of the previous section, in particular above  $z \sim 1$ . The match between the two distributions as computed from the KS-test is high at  $z < 1$  (97%) but clearly decreases above  $z \sim 1$  (67%). We find in particular an excess of early-type galaxies by a factor  $\sim 1.5$  in the Ks-band when compared to the I-band. Even if at these redshifts, the Ks-band classification is less accurate (§ ??), the difference is significant after correcting for the classification errors. The differences are probably due to morphological k-correction effects. It seems therefore that a morphological classification based on HST/ACS I-band data at those redshifts tends to under-estimate the elliptical population. Nevertheless, this results have to be confirmed with a more precise study of the early-type population at  $z \sim 1.5$  by studying their star forming histories and mass distributions.

## 5 Summary and conclusions

We have presented a morphological classification in two main morphological types of 44089 galaxies within the COSMOS field from seeing limited near-infrared imaging. Morphologies are estimated with the non-parametric N-dimensional code galSVM using a 10 dimensional volume and non-linear boundaries. The final output catalogue includes for every galaxy, a class label (early or late) and a probability of



**Fig. 2** Morphological k-correction. **a)** One-to-one comparison of the ACS and WIRCam classifications as a function of redshift. Red line: mean probability for a galaxy classified as early-type in the I-band (HST) to be classified as early-type in the Ks-band. Blue line: mean probability for a galaxy classified as late-type in the I-band (HST) to be classified as late-type in the Ks-band. Error bars show the dispersion of the probability distribution. **b)** Redshift distribution per morphological type for the same 34 165 Ks-selected galaxies observed in the I-band (dotted lines) and in the Ks-band (solid lines). Red lines are early-type objects and blue lines are late-type. The line width is the  $1 - \sigma$  confidence interval computed from the probability distribution and the error bars are calculated using Poisson  $\sqrt{n}$  statistics.

belonging to the class. The probability is proved to be highly correlated with the success rate and can therefore be used to assess the accuracy of the classifications. This classification method has been used to obtain the number counts and the redshift distribution per morphological type up to  $z \sim 2$  and to compare the results with the ones obtained from HST/ACS imaging on 34 165 galaxies in order to quantify morphological k-correction effects. Above  $z \sim 1$ , the I-band classification tends to find less early-type galaxies than the Ks-band one by a factor  $\sim 1.5$ . This probably reflects a morphological k-correction effect. Therefore, studies based on HST classifications at those redshifts could underestimate the elliptical population.

## References

- [1] M. Huertas-Company, D. Rouan, L. Tasca, G. Soucaill, O. Le Fèvre, *A&A* **478**, 971 (2008). DOI 10.1051/0004-6361:20078625
- [2] N.Z. Scoville, COSMOS Team, in *Bulletin of the American Astronomical Society* (2005), pp. 1309–+
- [3] A.M. Koekemoer, H. Aussel, D. Calzetti, P. Capak, M. Giavalisco, J.P. Kneib, A. Leauthaud, O. Le Fèvre, H.J. McCracken, R. Massey, B. Mobasher, J. Rhodes, N. Scoville, P.L. Shopbell, *ApJs* **172**, 196 (2007). DOI 10.1086/520086