Questions?
Suggestions?
Ideas?
Talk to me
or write to:

Boris.Haeussler
@nottingham.ac.uk

MEGAMORPH – automated multi-band fitting of galaxy profiles

Boris Häußler[1], Steven Bamford[1], Marina Vika[2], Alex Rojas[2]

1: School of Physics and Astronomy, UoN, UK
2: Carnegie-Mellon-University, Doha, Qatar

The University of Nottingham
GAMA QNRF
Carnegie Mellon Qatar

Abstract & Conclusions:

When it comes to measuring galaxy parameters, e.g. sizes (half-light radii), shapes (axial ratios) or profiles (e.g. Sérsic indices), on a large sample of galaxies with tens of thousands (e.g. COSMOS, STAGES or GEMS) or even millions of galaxies (e.g. SDSS), an automated fitting routine is greatly required. Starting from GALFIT (Peng et al., 2002), a code for fitting galaxy light profiles, and GALAPAGOS, a code that enables users to carry out single Sérsic profile fitting on large surveys in a largely automated manner, we here present MEGAMORPH, an extension of both these codes that allows simultaneous fitting of multi-wavelength data. While introducing new features into the codes, speeding them up and adding additional features, we make sure that all the well-tested features and stability of the original codes remain. MEGAMORPH will open up GALAPAGOS for varied survey strategies (including ground-based surveys with variable PSF), multi-wavelength data, and, most crucially, Bulge/Disk decompositions for a large sample of galaxies. We will also adapt the code to be used on high-performance computing facilities, to allow it to run on surveys with millions of galaxies.

1. What's the problem?

It is becoming clear that the striking difference between ellipticals and spiral galaxies is actually a result of variation in the relative prominence of their more fundamental spheroid and disk components (see Fig.1). Our understanding of galaxies would therefore be greatly improved by considering these physical components separately. However, measuring the properties of the individual components within a galaxy is considerably more difficult than measuring its overall properties as done by fitting codes in their current versions.

Current fitting routines only use a small fraction of the available data (e.g. one band in a multi-band survey). Several independent fits on all bands do not overcome this problem, simultaneous usage of all band and multi-component fitting is required. More, current 1-band-1-component fitting result are dependent on the chosen wavelength for the fit (see Fig.2) due to the mixing of the light of 2 separate galaxy components that dominate different regions of the galaxy image, e.g. a red bulge in the center and a blue disk in the outskirts.

Reliable Bulge-Disk-decomposition on one-band data, as currently being carried out by some groups, is challenging because of parameter degeneracies and multiple minima in likelihood space. Additional colour information could eliminate some of these minima and make the fit more stable and physically meaningful, e.g. by effectively allowing the code to fit a red stellar population in the red bands and extrapolating to blue bands and vice versa.

The MEGAMORPH project, a collaboration between astronomers, statisticians and computer scientists funded by the Qatar National Research Fund, is tackling this problem by utilizing the **full set of multi-color information** available for each galaxy and so potentially is able to separate different stellar populations within a galaxy and derive physically-meaningful structural parameters.

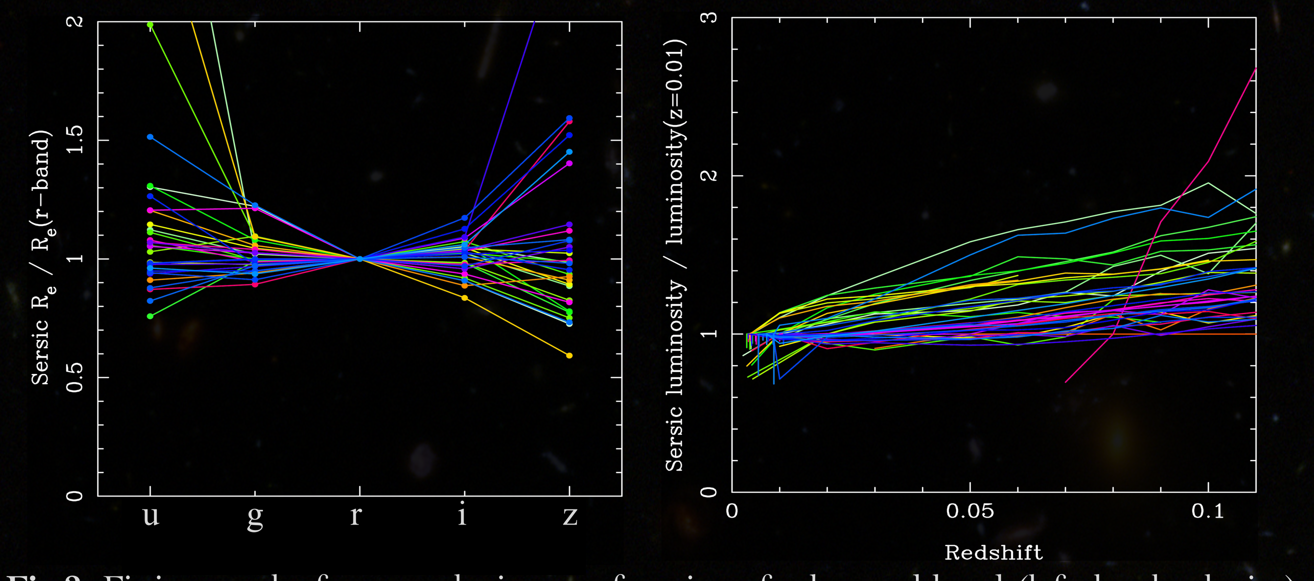

Fig.2: Fitting results from a galaxies as a function of observed band (left, local galaxies) and redshift (right, artificially redshifted galaxies, using FERENGI; Barden et al, 2008). One can clearly see that fitting results depend on the band used and the passband shifting due to cosmological redshifting of a galaxy. A tool that takes into account all available data, especially multi-wavelength information and treats it simultaneously is needed.

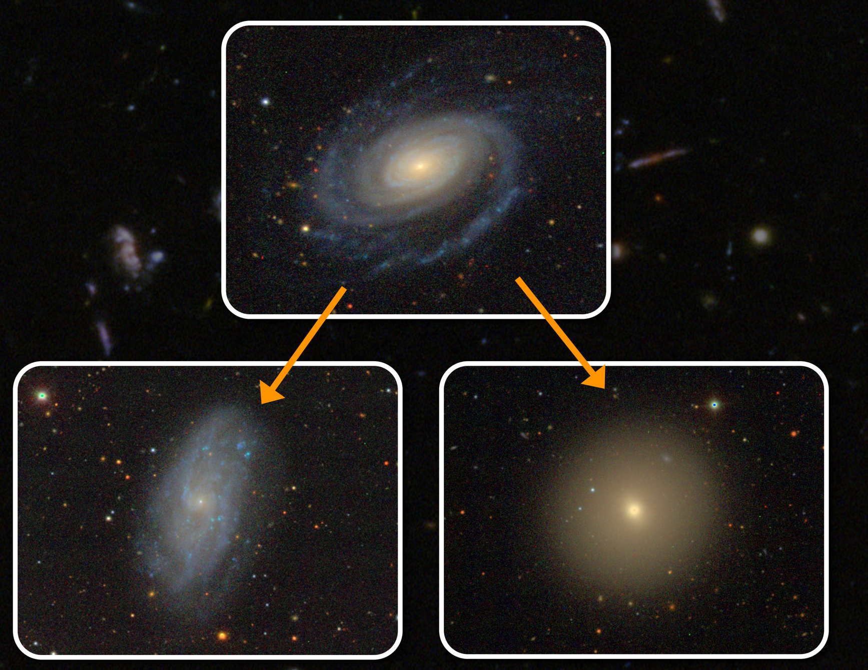

Fig.1: A separation of galaxies into their structural components is needed to understand the formation and evolution of galaxy populations in detail.

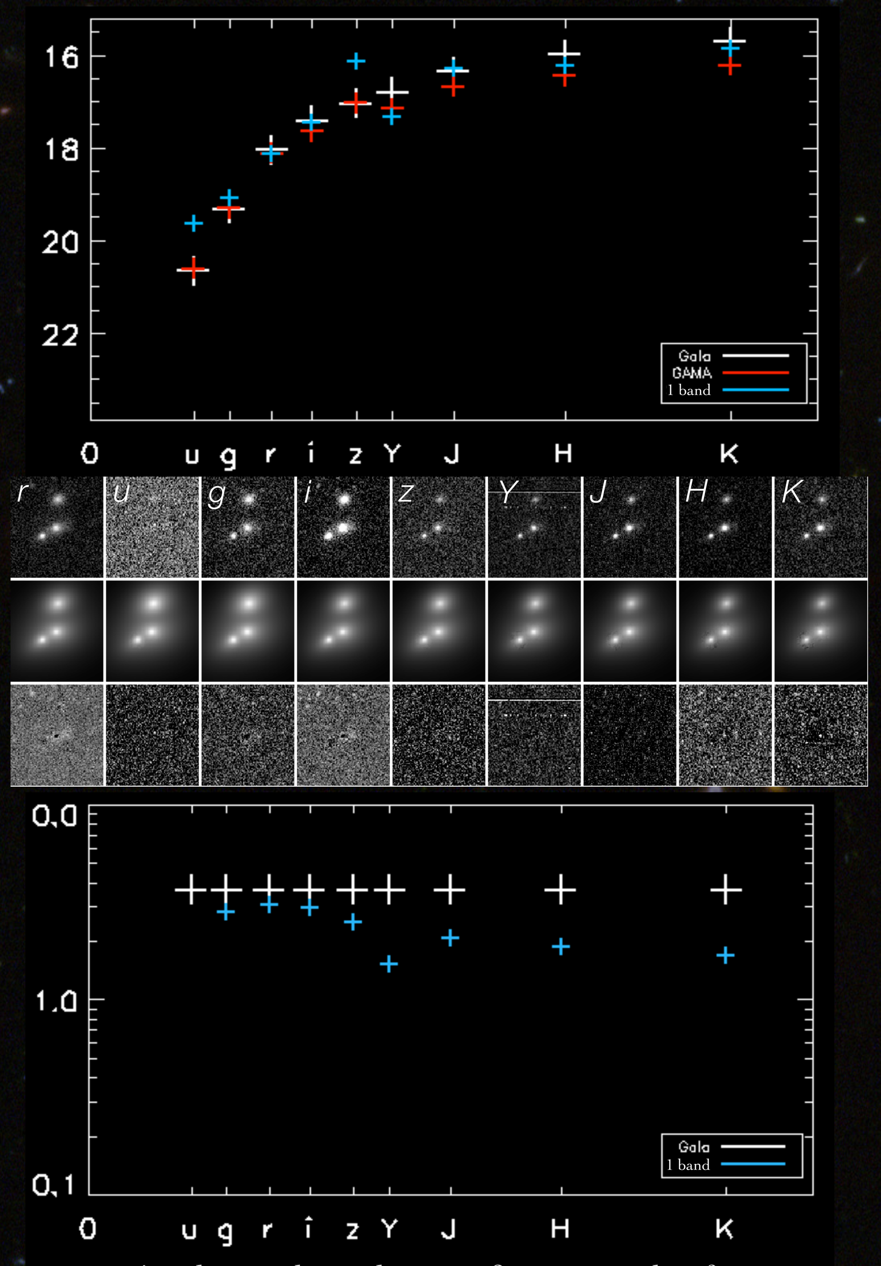

Fig.4: These plots show a fitting results from a multi-wavelength fit. The topmost plot shows magnitude vs. waveband in comparison to both photometry (red) and individual single band fits (blue), the panel in the middle shows the image, fitting model and residual for each of the 9 bands used here. The bottom-most plot shows the same for size. As one can see, multi-band fits can overcome some of the problems that individual bands have, e.g. get rid of outliers (top) or heavily wavelength dependent sizes (bottom). We will be able to remove parameter wavelength dependence once B/D decomposition is implemented. Blue galaxy components should be well fit in blue bands and extrapolated to red filters and vice versa.

2. Where did we start?

We use both GALAPAGOS (Galaxy Analysis over Large Areas: Parameter Assessment by GALFITing Objects from SExtractor; Barden et al, in prep, http://astro-staff.uibk.ac.at/~m.barden/galapagos/) and GALFIT, two pieces of established and well tested software, which we adapt to perform robust, physically meaningful **galaxy bulge-disk decompositions** using data from many wavelength bands simultaneously, while fully retaining backwards compatibility.

GALAPAGOS (Barden et al. 2011), is a wrapping script that, after an initial setup by the user, runs the entire fitting process without further user interaction:

- GALAPAGOS applies SExtractor (Bertin&Arnouts 1996) for object detection.
- Using an intelligent method, GALAPAGOS then automatically decides which neighbouring galaxies have to be **fit simultaneously** and which neighbours can be **masked out**. It automatically creates mask images used by GALFIT in the fitting process.
- Dealing with galaxies in order of decreasing brightness, GALAPAGOS:
  - writes out a Galfit start file.
  - **runs GALFIT.**
  - **Reads in the fitting result** and uses this for deblending purposes in the further process of the code. Currently this is done via an internal result structure.

Both GALAPAGOS and GALFIT have been thoroughly tested by several independent groups. We also carried out extensive tests of GALAPAGOS using both real and simulated data. We find that GALAPAGOS in general returns **very good** results except for a small systematic offset that can be seen for the faintest galaxies with very high Sérsic indices that are most sensitive to uncertainties in the sky estimation. We were able to show the independence of galaxy parameters from both distance and magnitude of neighbouring objects as measured in simulated data. Whereas other fitting methods are sensitive to neighbours, GALAPAGOS and GALFIT are not.

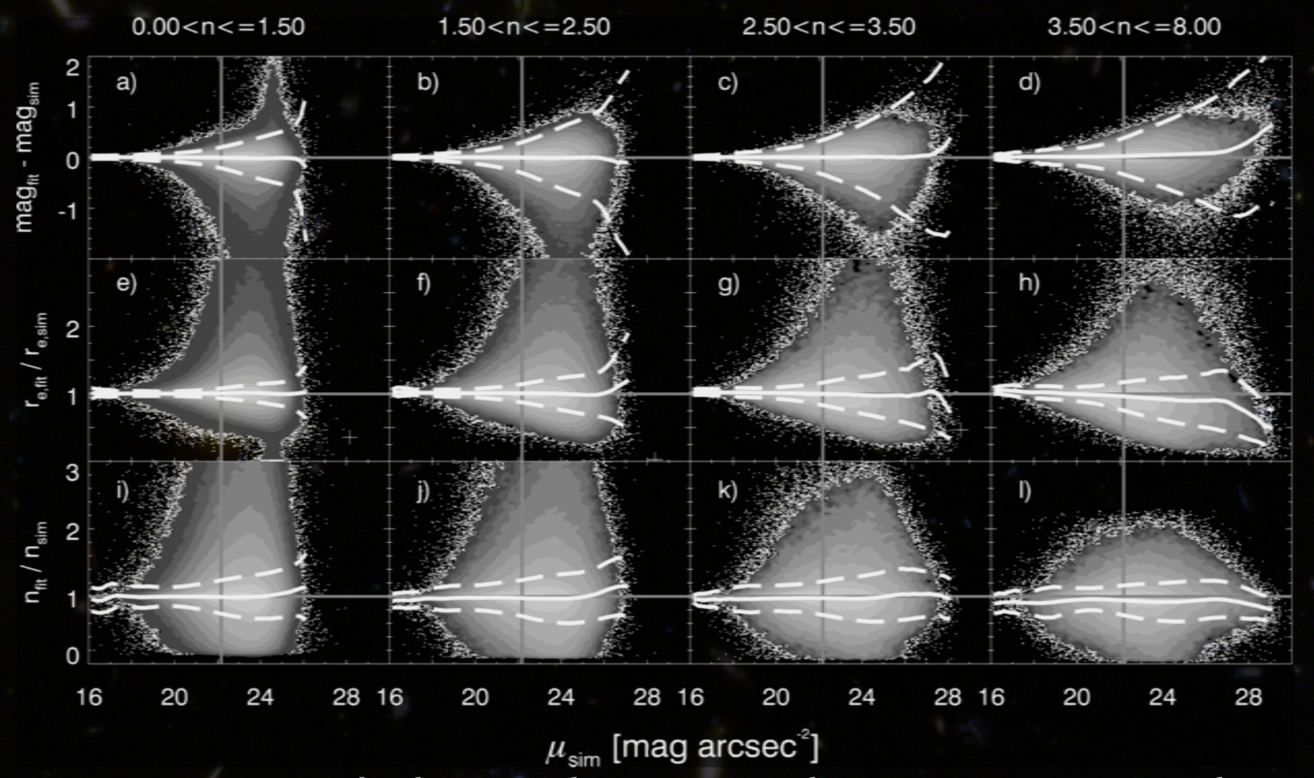

Fig.3: Fitting accuracy for the main galaxy parameters by GALAPAGOS in STAGES data as a function of surface brightness (x-axis) and morphology (different panels) as derived by fitting 10 million simulated galaxy profiles. Only at the faintest levels (and high n), systematic effects can be seen (See Gray et al., 2009 for details).

3. Where are we now?

The MEGAMORPH project is still very much in development but we have already added some big new features to the codes used. We have:

- selected the baseline system used throughout the project: GALAPAGOS and GALFIT (see §2)
- selected and created the test environment, using:
  - simulated survey data (Following Haeussler et al.,2007).
  - real survey data (GAMA, Driver et al., 2011).
  - artificially redshifted real galaxies (using FERENGI).
- sped up the code in critical places and nearly halved the CPU time needed to run on real data.
- have adapted GALAPAGOS to use variable PSFs, depending on objects position. This is important for ground-based surveys.
- **adapted both GALFIT and GALAPAGOS to be able to use multi-wavelength data simultaneously.**
- We have successfully run the codes on a small subregion of the GAMA survey with around 10000 galaxies. One of the fits is shown in Fig. 4. In this run, the shape of a single sérsic profile was held fixed over all wavelengths, only the magnitude varies (with complete freedom) from band to band. As one can see, the fitting magnitudes are in better agreement with photometric data than single-band fits. By design, the size was held fixed in this run. This is a user chosen specification, the code offers full flexibility on each parameter individually.

On a different approach, we have successfully redshifted a sample of ~100 local galaxies to redshifts out to z=0.25. We are in the process of fitting these images using a similar automated technique to be able to:

- evaluate the influence of this redshifting and the bias introduced into real survey data by this effect
- test the robustness of our codes now and in the further development. Fig.5 shows an example of this redshifting and fitting procedure.

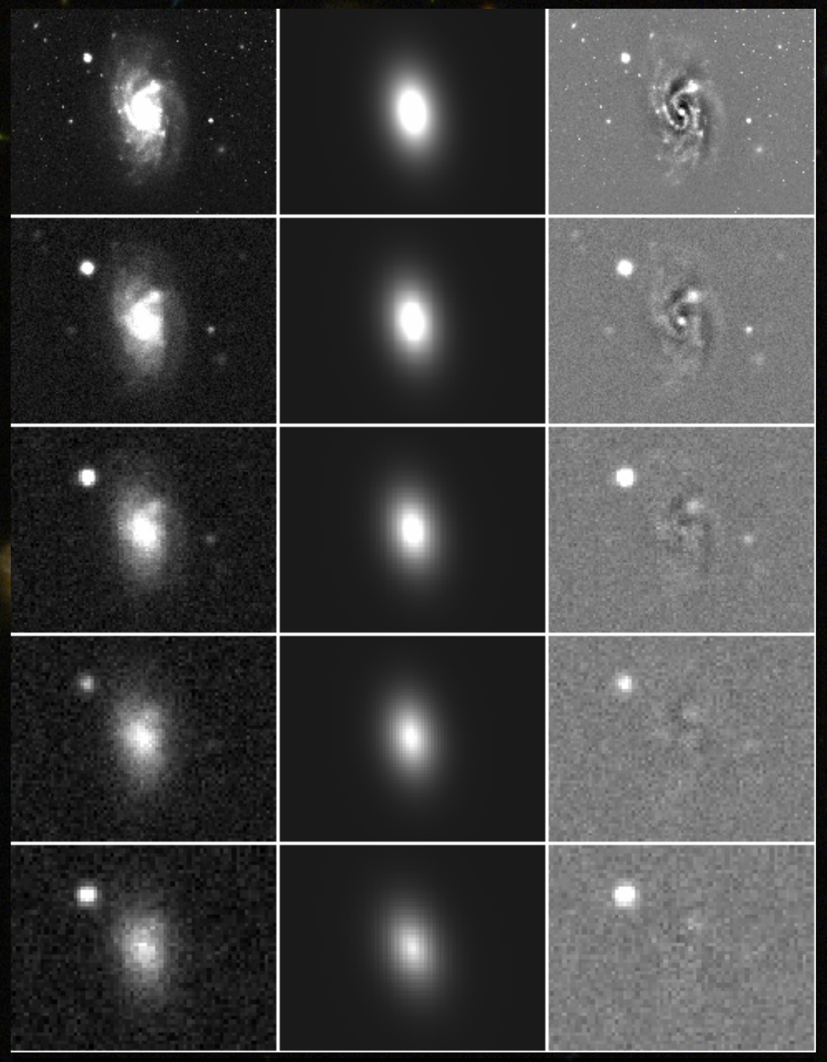

Fig.5: In this plot we show some example images of an artificially redshifted galaxy on the left, created using FERENGI (Barden et al., 2008). Rows resemble redshifts of z=0.01, 0.03, 0.05, 0.07, 0.09. The middle column shows the galaxy models as derived by GALFIT, the right columns shows the residual between the two. Although, ignoring the galaxy substructure left in the residuals, all fits look very good, the model parameters change significantly, see Fig.1.

4. Where are we going?

Galaxies are complex structures and, beyond the general distinction between spheroids and disks, they display a range of higher level features that make it difficult for computational methods to extract meaningful information. To overcome this problem, we want to introduce non-parametric components into the fit to be able to successfully subtract components like faint spiral arms or star forming regions from the data during the fit to only fit the underlying profile.

The next step is the introduction of a second fitting component, effectively moving from singe sérsic profile fitting to full Bulge-Disk-Decomposition. Multiwavelength data and fitting should be able to overcome many of the degeneracy problems that current, 1-band, B/D compositions have, due to its power of using the full colour information of a galaxy image on a pixel to pixel basis. An accurate best-model-selection has to be installed to automatically choose which fit resembles the galaxies profile more accurately.

To control the computational intensity of the task, we will need to use efficient algorithms and tools to optimally use the full CPU available. We are still trying to speed up the actual code, but will also adapt it to work on high-performance computer facilities either on local HPC machines available to the user or e.g. Amazon Web Services.

After final testing and demonstration, we will publish the code for everyone to use on their dataset and in their own setup.