ICRAR

International Centre for Radio Astronomy Research

EXPLORING THE LIMITS OF THE BAYESIAN UNIVERSE: HOW TO TACKLE BREADTH AND DEPTH

Curtin University

THE UNIVERSITY OF WESTERN AUSTRALIA

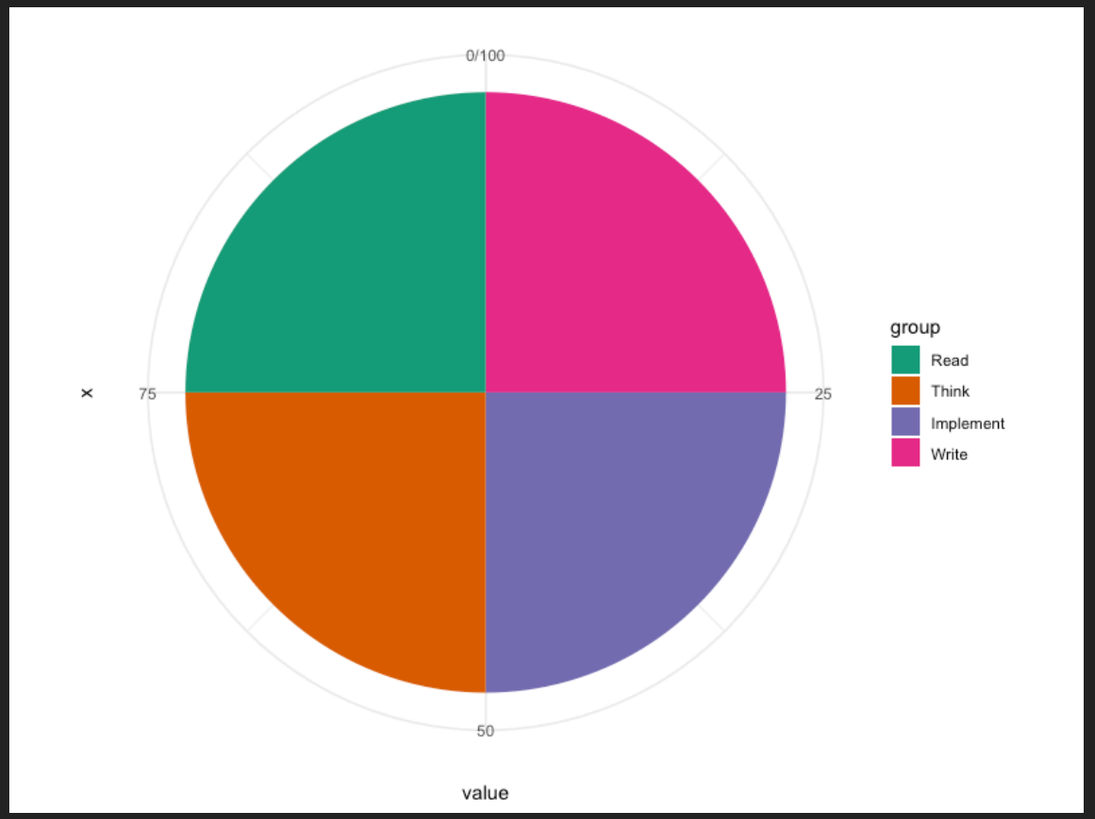
Government of Western Australia
Department of the Premier and Cabinet
Office of Science

AARON ROBOTHAM

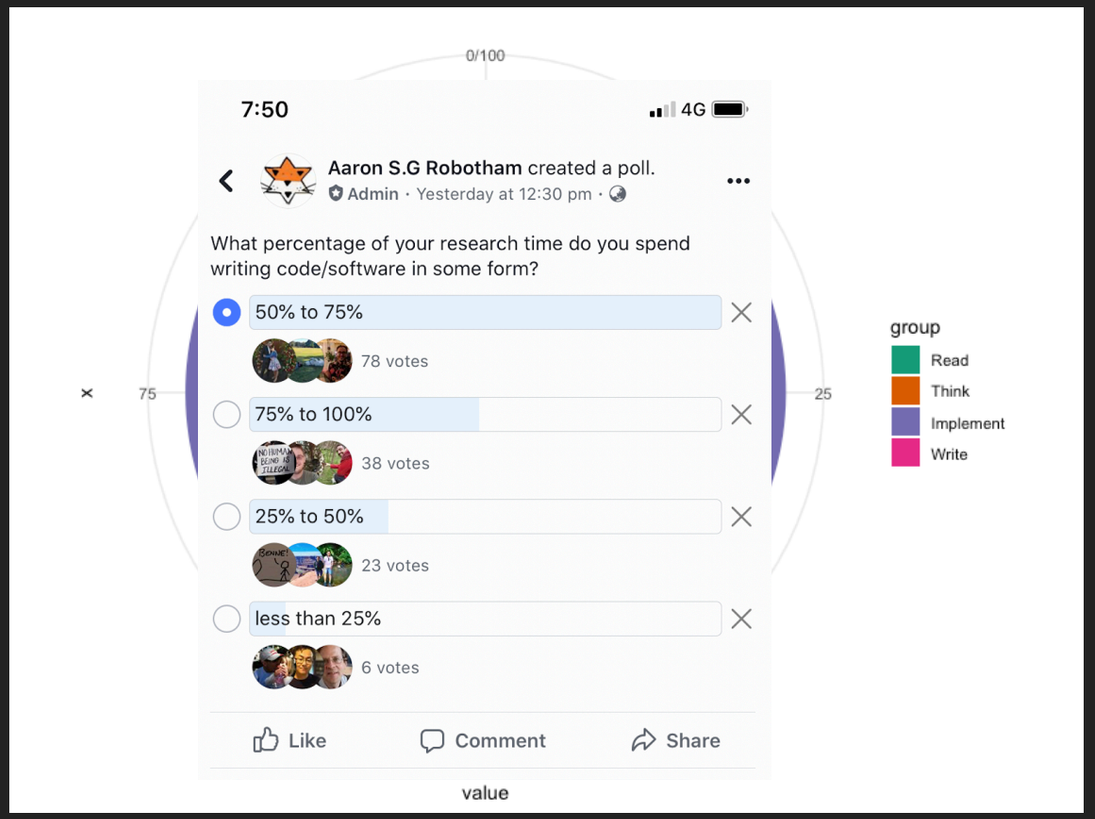
Sabine Bellstedt, Jessica Thorne

WHAT WE DO WHEN WE RESEARCH:

The Legend



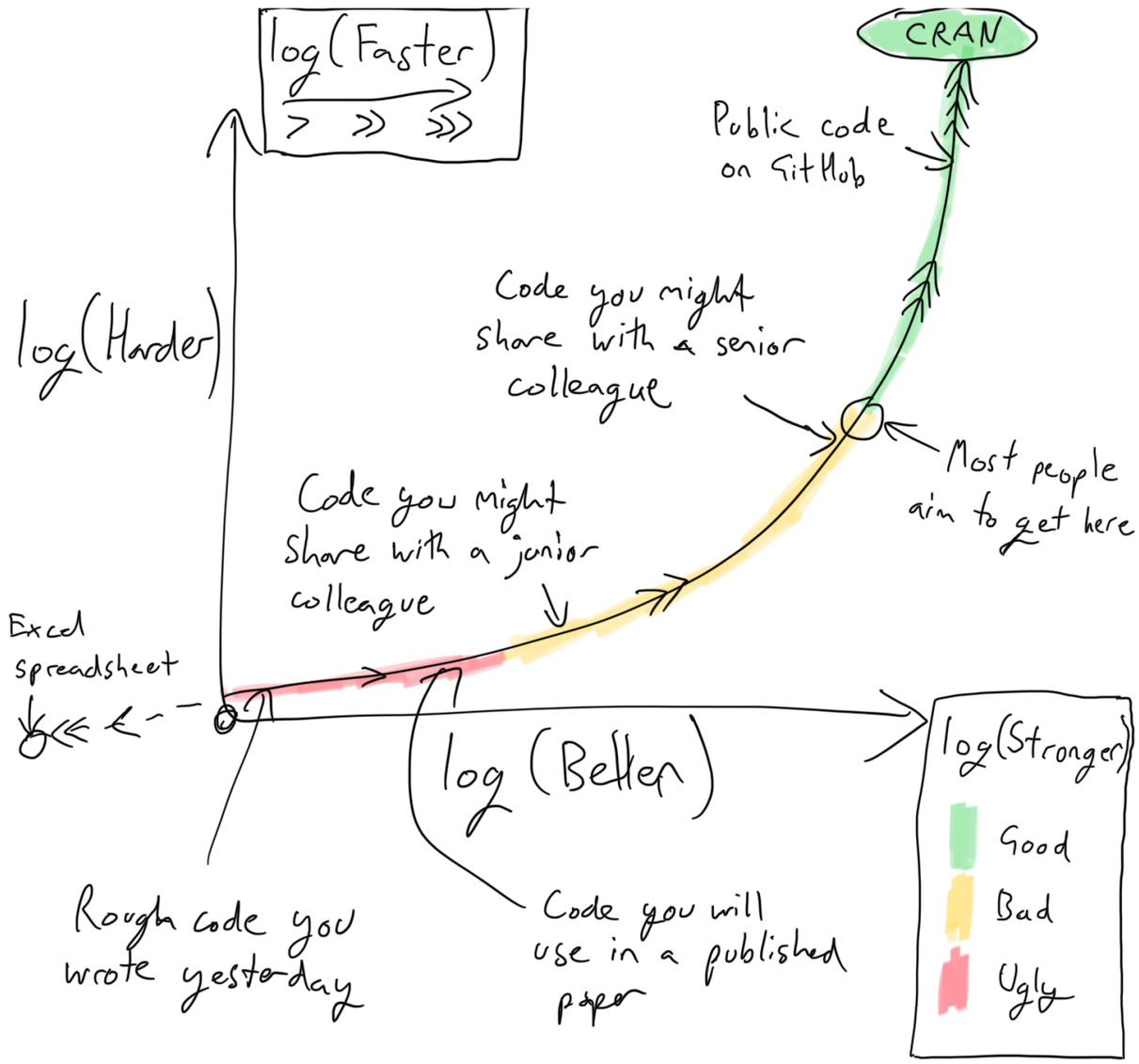
The Reality



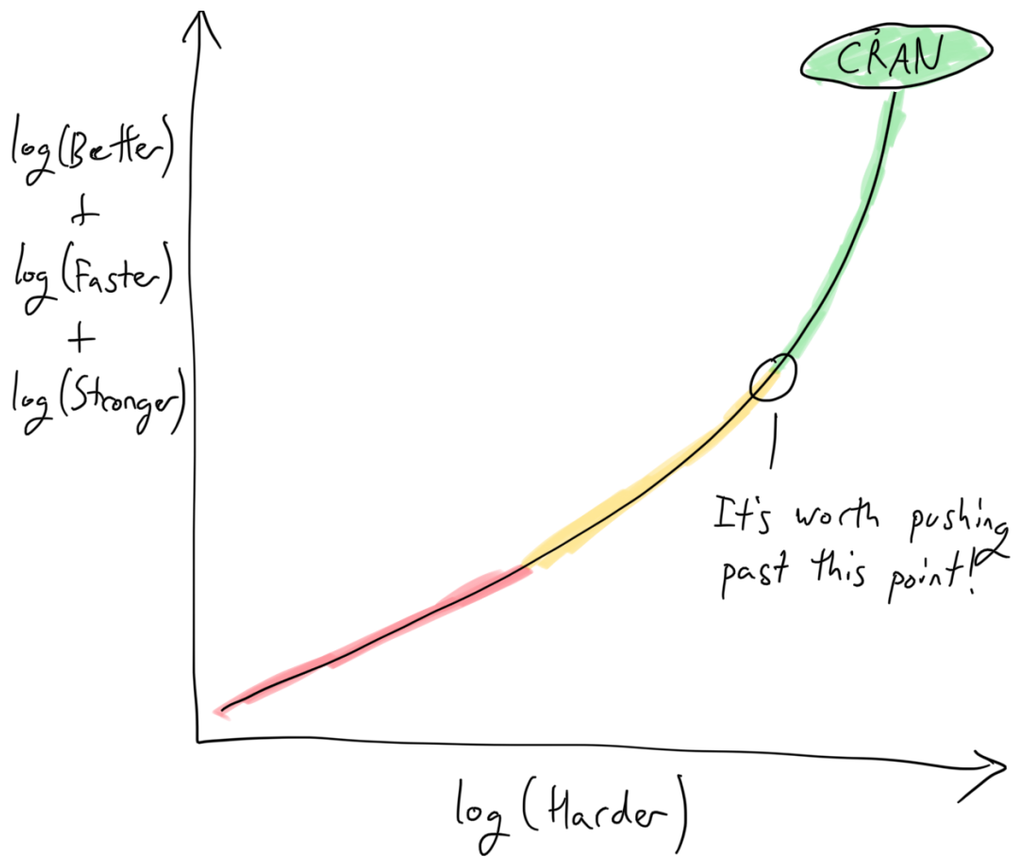
In a poll on the Astrostatistics Facebook group, 80% of people reported spending more than half (average around 70%) of their research time writing code/software on a computer.

Harder: roughly time taken (mental effort)
Better: commented, maintainable
Faster: does the job faster
Stronger: Reliable, fault tolerant, robust

SO WE MUST BE GREAT AT THIS, RIGHT? WELL...



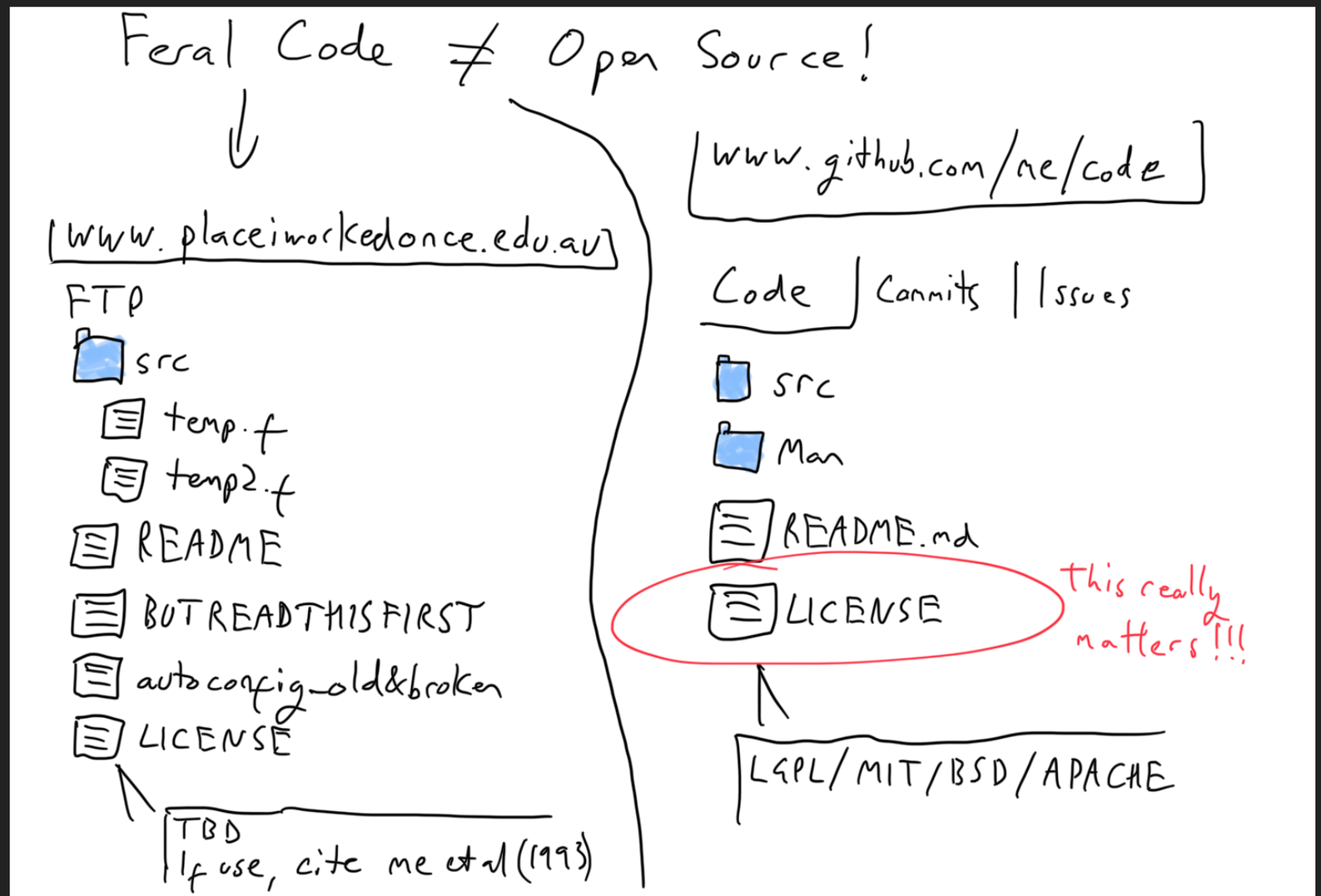
But overall it is worth it!



THE SOLUTION

- We (as a community) bootstrap the “implement” part wherever feasible.
- Even when we re-invent the wheel, knowing what wheels are available makes this:
  - A) worthwhile (need a roller-blade wheel in a world of cartwheels)
  - B) efficient (we can still borrow ideas - make it round!).
- To help our community this means the following:
  - Aggressively public code +
  - Default collaborative efforts =
  - Open-source analysis (statistical methods and software)

BUT REMEMBER!

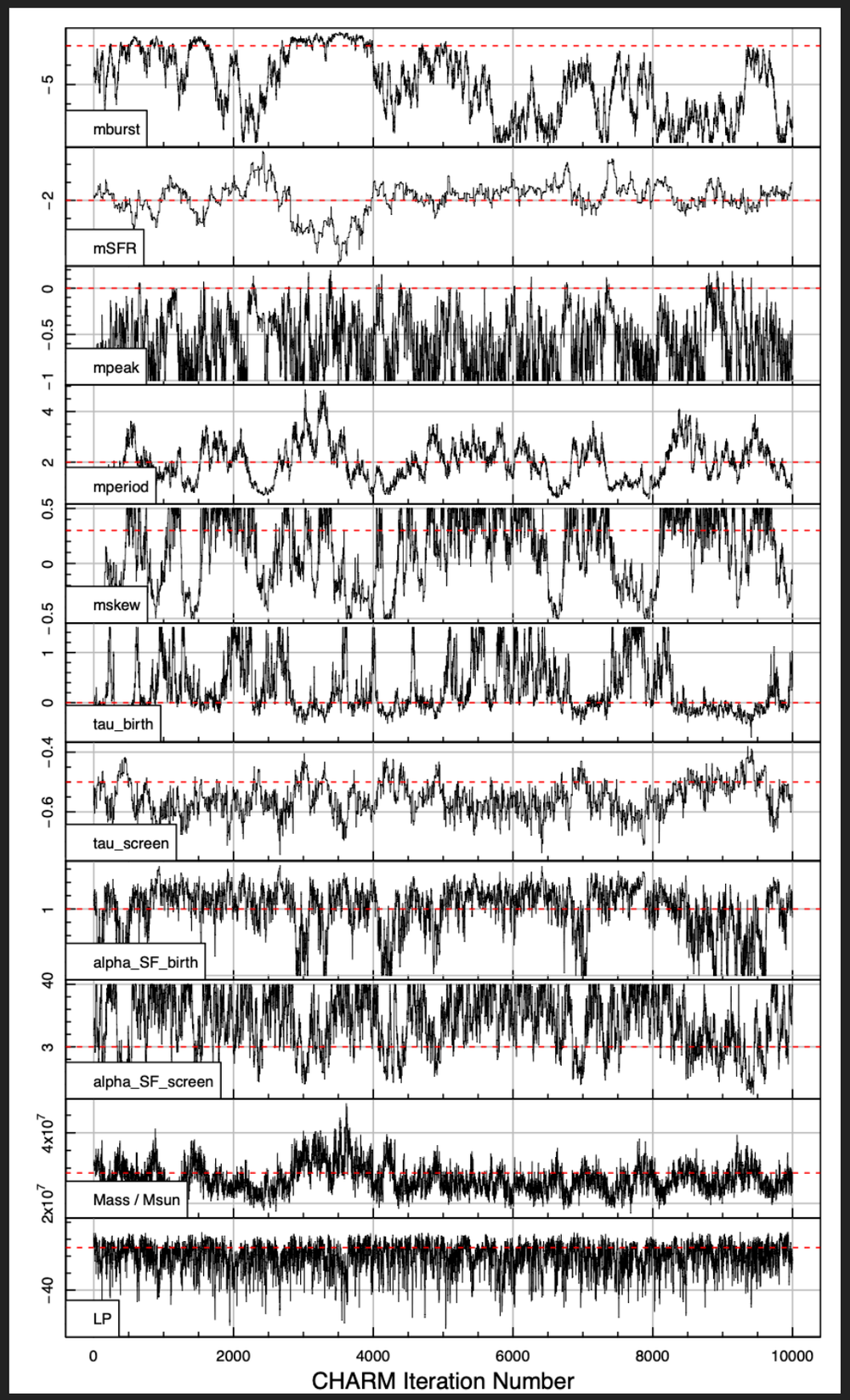


WHY WE SHOULD DO OPEN-SOURCE STATISTICS IN ACADEMIA?

- Because it is the new fun thing!
- I keep seeing talks about GitHub, maybe it will help me get a job…?
- My grant funding insists.
- Better than nothing for backing things up.
- Paying for things sucks.
- To make the world a better place…
- These are not the best reasons actually… we can do better by appealing to our more ruthless and selfish instincts!

WHY WE REALLY SHOULD DO OPEN-SOURCE STATISTICS IN ACADEMIA?

- Only way to efficiently collaborate on complex projects, and drops the bar to entry to almost nothing. If people cannot see it then they will not ask to help.
- Get a lot more credit and exposure to your work (you will start appearing all over the internet).
- It instills a much higher level quality of work: if you do not feel comfortable sharing it then you should not feel comfortable publishing it! As Russians say: “Trust, but verify”.
- It provides free (!) wide-coverage quality checking you cannot replicate alone.
- It allows you to work seamlessly on projects across multiple devices without penalty (I often show people snippets from GitHub on my phone - handy in a pinch).
- It naturally enforces modern version control by any practical route (Git / GitHub).
- It helps less senior members of our community. What does this mean…?



STATISTICS AND COMPUTING

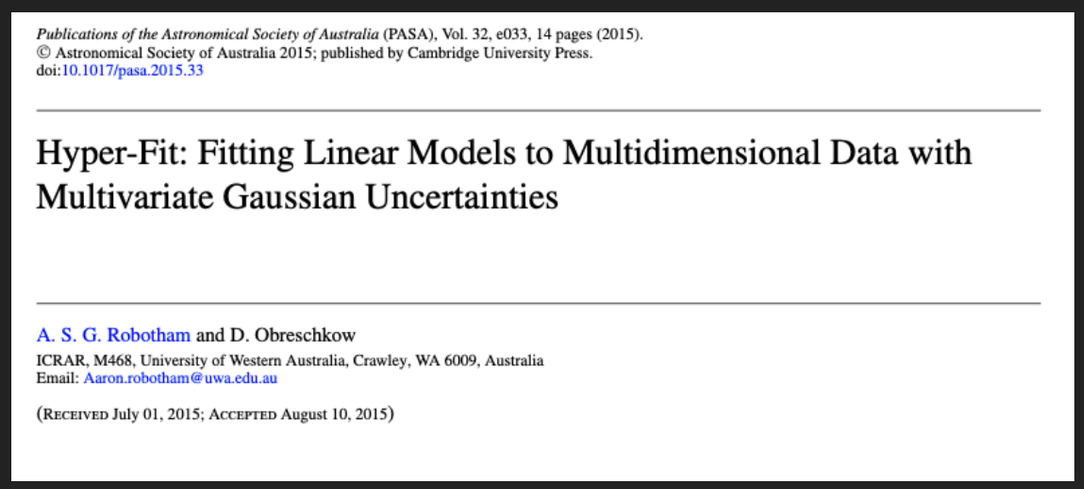
BARRIERS

TOOLS TO BE A MODERN ASTRONOMER

- A new PhD student in a highly mature field like astronomy (a few thousand years of research and counting) has a lot of catching up to do. They need:
  - Solid coding skills (no matter what they do).
  - Relatively sophisticated knowledge of data analytics and statistics.
  - Broad understanding of the current state of the field, and a deep understanding of their own sub-field.
  - Excellent project management.
  - Good scientific writing (an un-taught skill usually)
  - Functional people skills.

A CASE STUDY WITH HYPER-FIT

- Astronomy is unusual within physics in how expensive it is to get observations, so we tend to have un-ignorable errors, and available techniques like PCA and SVM rarely support the treatment of errors *properly*.
- Because of some mass-size data, I started wondering about hyperplane fitting with heteroscedastic covariant errors, and published a paper tackling it:



Publications of the Astronomical Society of Australia (PASA), Vol. 32, e033, 14 pages (2015).

doi:10.1017/pasa.2015.33

Hyper-Fit: Fitting Linear Models to Multidimensional Data with Multivariate Gaussian Uncertainties

A. S. G. Robotham and D. Obreschkow
ICRAR, M468, University of Western Australia, Crawley, WA 6009, Australia
Email: Aaron.robotham@uwa.edu.au

(Received July 01, 2015; Accepted August 10, 2015)

$$\ln \mathcal{L} = -\frac{1}{2}\sum_{i=1}^{N}\left[\ln\left(\sigma_{\perp}^{2}+\frac{\mathbf{n}^{\mathsf{T}}\mathbf{C}_{i}\mathbf{n}}{\mathbf{n}^{\mathsf{T}}\mathbf{n}}\right)+\frac{(\mathbf{n}^{\mathsf{T}}[\mathbf{x}_{i}-\mathbf{n}])^{2}}{\sigma_{\perp}^{2}\mathbf{n}^{\mathsf{T}}\mathbf{n}+\mathbf{n}^{\mathsf{T}}\mathbf{C}_{i}\mathbf{n}}\right]$$

This involves N-dimensional projections and careful treatment of the data covariance terms. This is all fiddly stuff that took me a lot of sanity checking to implement robustly (e.g. adapting for geometric corner cases). Now the question is, would you confidently trust a new PhD student to implement the above in code you need to use?

A CASE STUDY WITH HYPER-FIT

- So a paper with an equation is often not enough. How about an R CRAN package?

hyper.fit: Generic N-Dimensional Hyperplane Fitting with Heteroscedastic Covariant Errors and Intrinsic Scatter

Includes two main high level codes for hyperplane fitting (hyper.fit) and visualising (hyper.plot2d / hyper.plot3d). In simple terms this allows the user to produce robust 1D linear fits for 2D x vs y type data, and robust 2D plane fits to 3D x vs y vs z type data. This hyperplane fitting works generically for any N-1 hyperplane model being fit to a N dimension dataset. All fits include intrinsic scatter in the generative model orthogonal to the hyperplane.

Version:	1.1.0
Depends:	R (≥ 3.00), magicaxis, MASS, rgl, LaplacesDemon
Published:	2019-01-31
Author:	Aaron Robotham and Danail Obreschkow
Maintainer:	Aaron Robotham <aaron.robotham at uwa.edu.au>
License:	GPL-3
NeedsCompilation:	no
Citation:	hyper.fit citation info
Materials:	NEWS
CRAN checks:	hyper.fit results

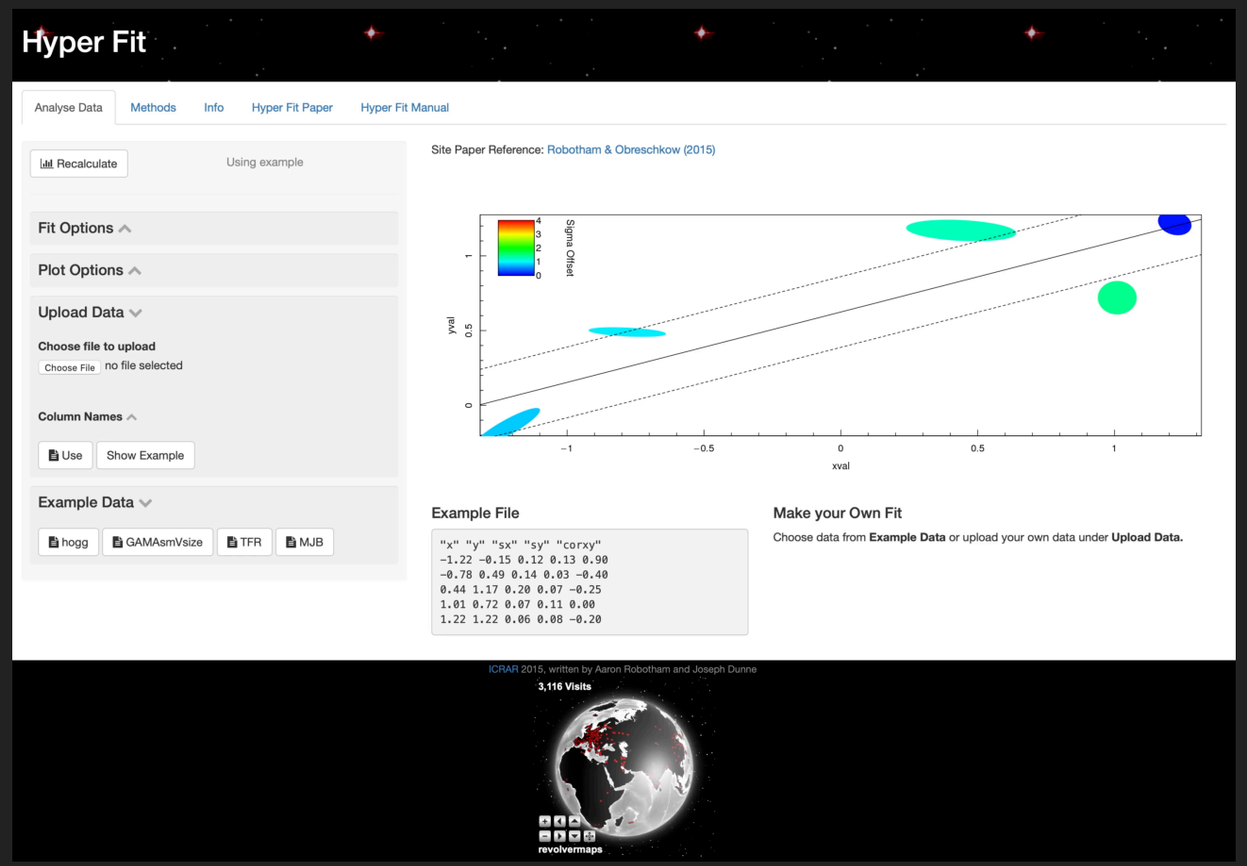
Downloads:

Reference manual:	hyper.fit.pdf
Package source:	hyper.fit_1.1.0.tar.gz
Windows binaries:	r-devel: hyper.fit_1.1.0.zip, r-release: hyper.fit_1.1.0.zip, r-oldrel: hyper.fit_1.1.0.zip
OS X binaries:	r-release: hyper.fit_1.1.0.tgz, r-oldrel: hyper.fit_1.1.0.tgz
Old sources:	hyper.fit archive

- The bar for this is high (*much* higher than a simple code repo like SourceForge or PyPI)- it requires full code documentation (40 pages!) and a full suite of useable examples.

A CASE STUDY WITH HYPER-FIT

- Is that enough? Well lots of people do not use R, fear it, and will not read a manual (arguably they do not deserve my help!). Behold Shiny Apps:



hyperfit.icar.org

https://www.icrar.org/our-research/tools/

Cosmology Calculator cosmocalc.icrar.org

Calculation | Plot | Survey Design | Sky Separation | Info | R Code

Calculate

Set Variables:

z: 1 | Sig Figs: 6

Reference Set: 737 | H[0]: 70.0

OmegaM[0]: 0.3 | OmegaL[0]: 1-OmegaM[0]

OmegaR[0]: 0

w[0]: -1 | w': 0

Sigma8[0]: 0.8 | Calc fSigma8

Custom Calc:

Value (Gyr) | Solution to find: High

Variable: Look-back Time

Results:

The redshift **z** is 1

The expansion factor **a** is 0.5

The **Relative Error** is ± 1e-08

Distance:

The **Comoving Radial Distance Lo**

The **Luminosity Distance** to z is 66

The **Angular Diameter Distance** to

The **Comoving Radial Distance Tra**

The **Distance Modulus** to z is 44.10

The **Physical Angular Scale** at z is

The **Comoving Volume** to z is 151.0

Velocity and Expansion:

Hubble's constant **H** at z is 123.248

Expansion rate **ȧ** at z is 61.6239 (km

The **cz Velocity** at z is 1 (c)

The **Cosmological Recession Velo**

The **Deceleration parameter (q)** at

What's Up? whatsup.icrar.org

Submit

General Options

Target: User | Telescopes: User

RA: 08:21:24.74 | Dec: +19:28:33.71

Longitude: 115.8178 | Latitude: -31.97333

Local Time: 13:00 | Local Date: 2019-11-06

Height/m: 0 | Temp/C: 20 | Pressure/mB: 1000

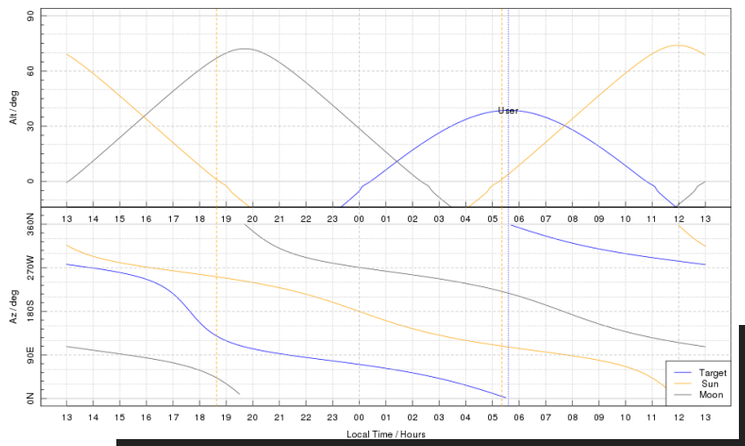
UTC-Diff: 8 | Time Zone: User

Y-Axis: Altitude

Night plan target list | Use

Day/Night Plan | Now Plan | Year Plan | Finder Chart

Y2019M11D6, RA 125, Dec 19, Lon 116, Lat -32, Mphase 0.67, Msep 150, Ssep 100



The HI Fidelity Calculator hifi.icrar.org

Info | z, Width, Flux | Angles & Beam | Flux/Mass | Flux/nHI | Fluxd/Tb | Noise | SNR

Plot:

Change the parameters below to modify the redshift dependence plot of conversions from the given fixed value:

z Start: 0.01 | z End: 1

z Resolution: 100 | solid angle scaling: none

x Axis: z | y Axis: K

Log x axis | Log y axis

Flip x axis | Flip y axis

Save Data

Download as .txt

Download as .csv

Results:

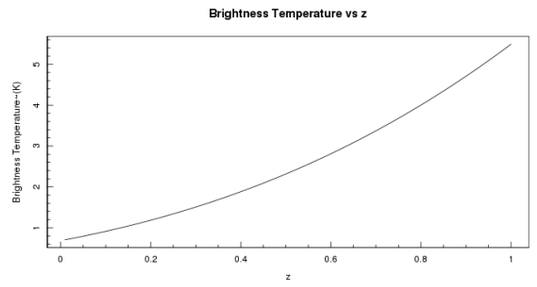
The redshift **z** is 1

The observed solid angle of the source is 1 arcsec^2

The observed flux density is 1e-06 Jy

The area of the source (small angle approx) is 64.14 kpc^2

The brightness temperature of the source is 5.491 K

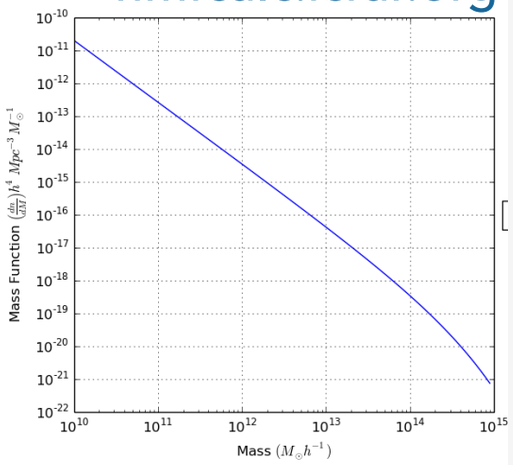


HMF*calc* | Info | Contact Us | Calculate | View Plots

Plot: dn/dm

Download: PDF of Current Plot

hmfcalc.icrar.org



AstroMap astromap.icrar.org

Description | Tool | Get Code | Contact

Map Projection | Light Cones | Deep Fields

Update plot

Projection

Projection Type: aitoff | Long-Cent (0,180): 0

Coordinate Type: Celestial

Objects/Paths

Current Wide Surveys

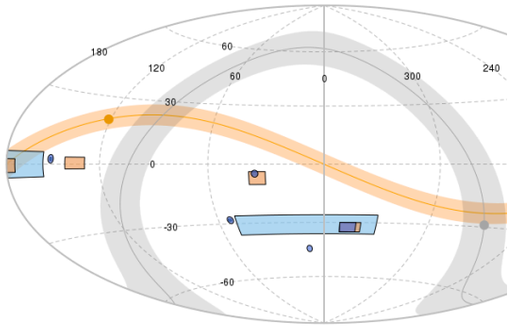
Current Deep Surveys

Upcoming Surveys

User Defined

PDF

Ecliptic
Sun
MW Plane
MW Center



WAVES-Wide | DEVILS | GAMA | WAVES-D

ProSpect SED prospect.icrar.org

Redshift: 0.1

SFH/AGN:

Burst Mass (log10): 8

Young Mass (log10): 9

Mid Mass (log10): 10

Old Mass (log10): 10

Ancient Mass (log10): 10

AGN Luminosity (log10): 42

Aaron Robotham

GitHub: asgr/ProSpect

Dust inputs:

Tau Birth: 1

Tau Screen: 0.3

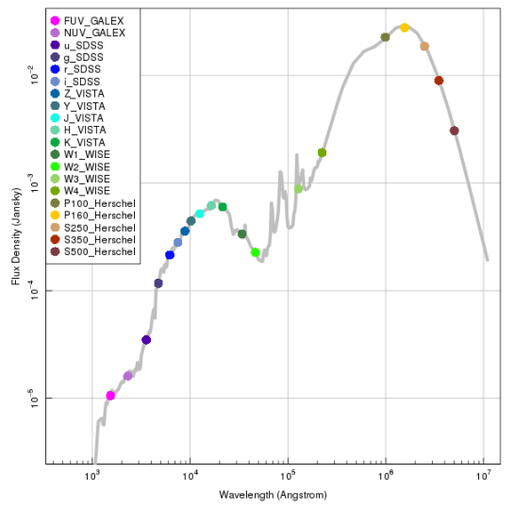
Tau AGN: 1

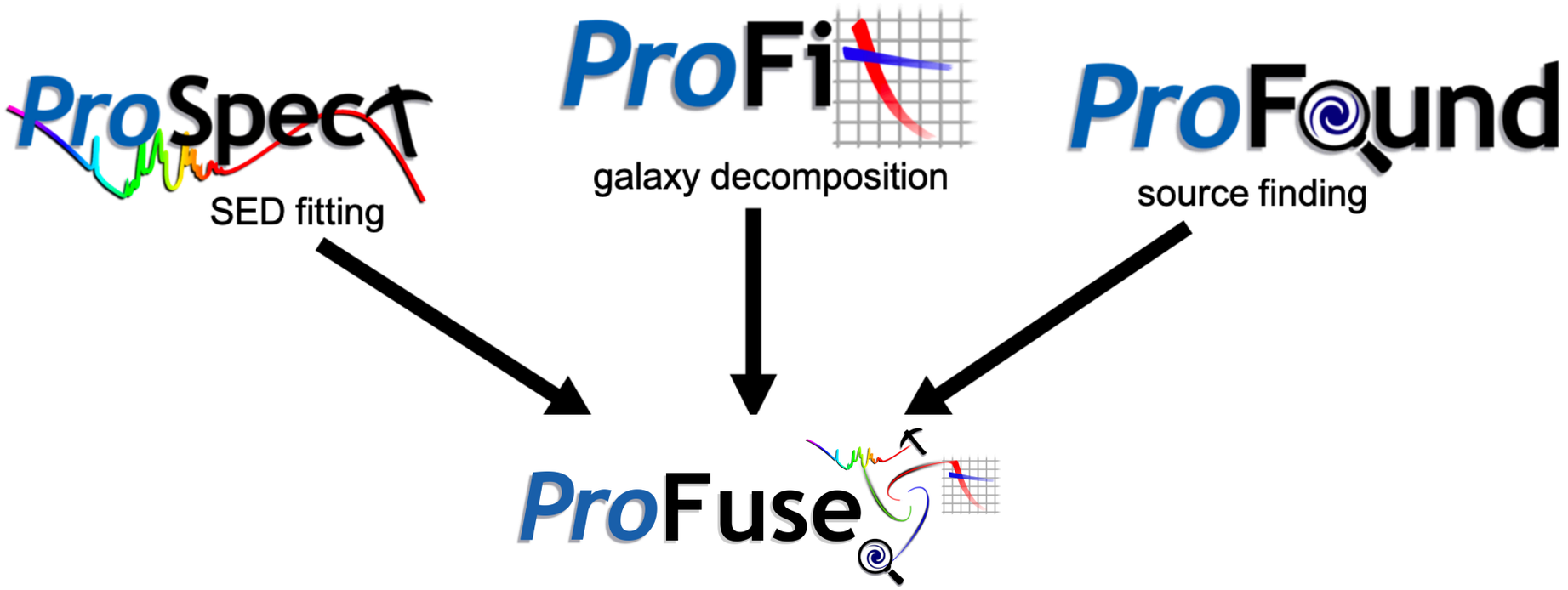
Alpha SF Birth: 1

Alpha SF Screen: 3

Alpha SF AGN: 0

Flux | Luminosity | Info





FROM PIXELS TO SCIENCE

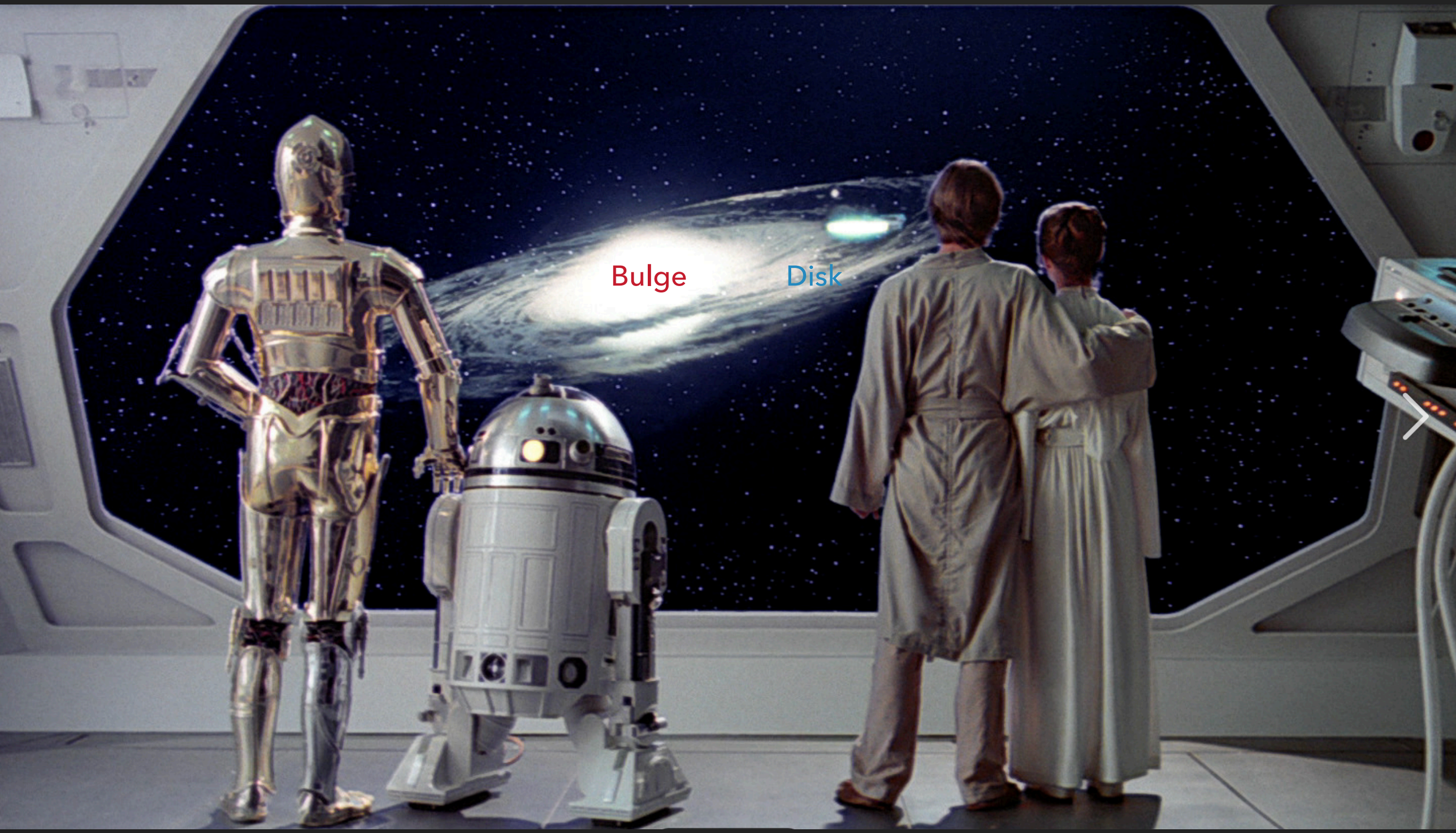
PROFUSE

MNRAS **513,** 2985–3012 (2022) https://doi.org/10.1093/mnras/stac1032
Advance Access publication 2022 April 15

ProFuse: physical multiband structural decomposition of galaxies and the mass–size–age plane

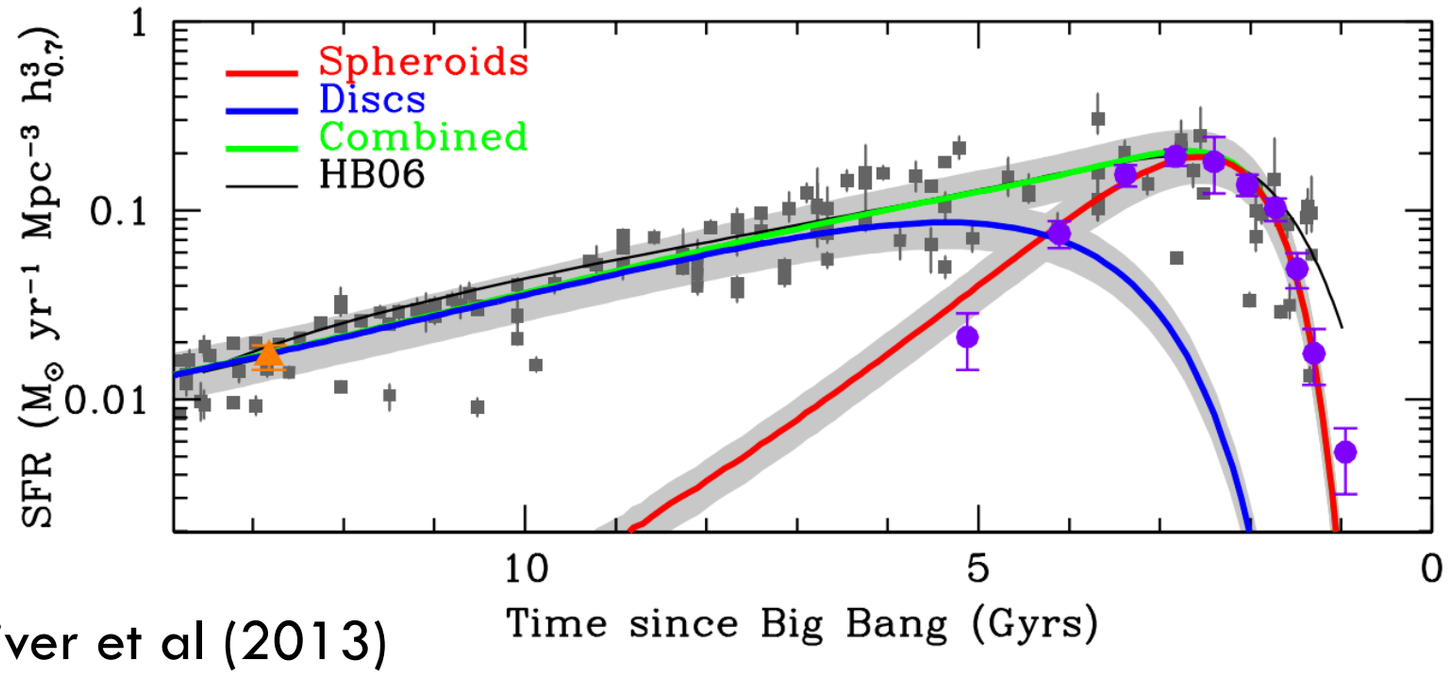
A. S. G. Robotham [1,2★] S. Bellstedt [1] and S. P. Driver [1]

BULGES AND DISKS

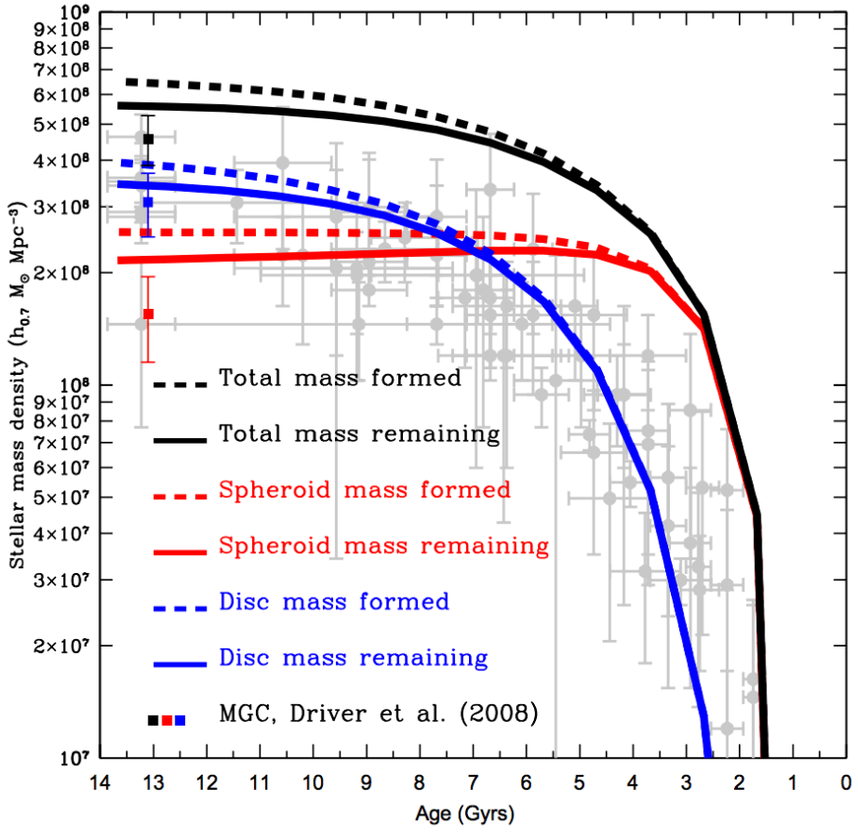


Why Decompose Galaxies?

- 2 reasons:
  - But the more important reason is that the components of galaxies contain the evolutionary origin of galaxies.
  - Two dominant paradigms:
    - Rapid assembly of bulges (star bursts and mergers)
    - Slower accretion growth of disks

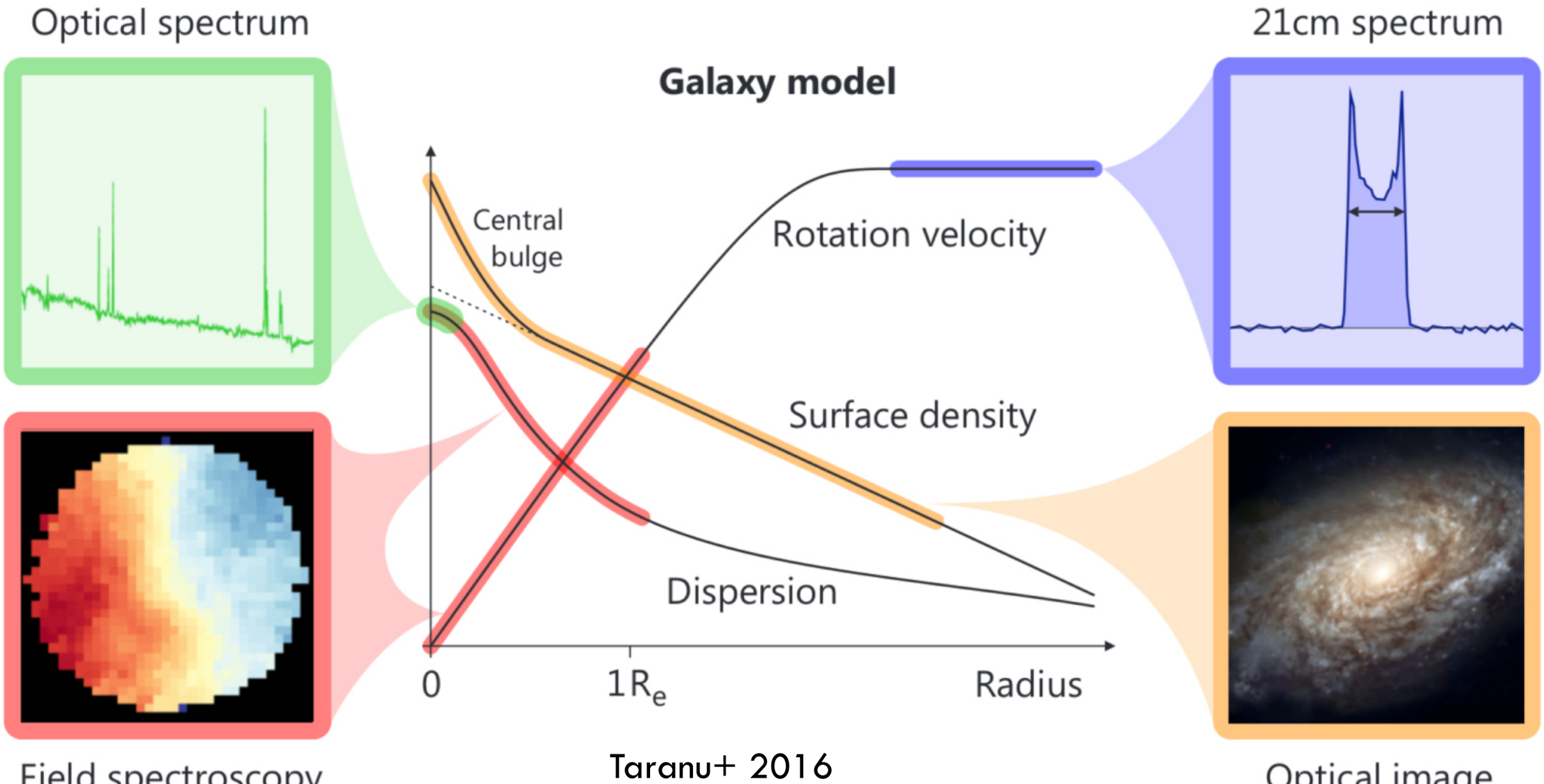


Driver et al (2013)



Why Decompose Galaxies?

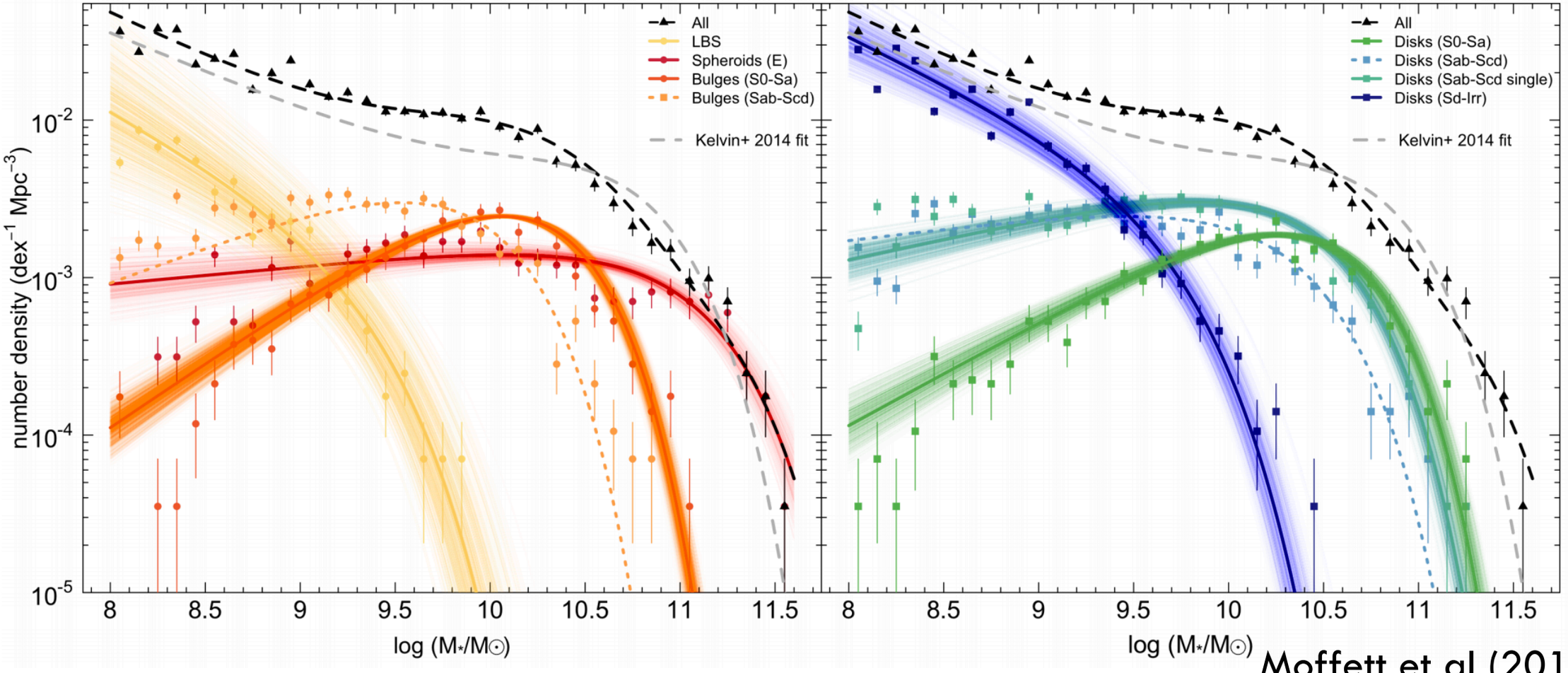
- The optical properties of galaxies are key to understanding the dynamics:



Taranu+ 2016

Why Decompose Galaxies?

- The two major components also show remarkably different levels of significance for different mass galaxies (high mass = bulge dominated, low mass = disk dominated). Even the galaxy stellar mass function (GSMF) is non-trivial in detail:



Moffett et al (2016b)

THE PROFUSE PIPELINE

- ProFound - Detect sources automatically, estimate rough model parameters, create segmentation and error maps. See Robotham+ (18), Bellstedt+ (20), GAMA/DEVILS/WAVES.
- ProFit - Parameter driven single image galaxy decomposition. See Robotham+ (17), Cook+ (18/19), Casura+ (22).
- ProSpect - Multi-band spectral energy distribution generation and fitting. See Robotham+ (20), Bellstedt+ (20/21), Thorne+ (21/22ab).
- ProFuse = ProFound + ProFit + ProSpect
  See Robotham+ (22)

ProFound

EXTRACTING SOURCES AND GETTING READY FOR PROFIT

PROFOUND

GitHub: asgr/ProFound

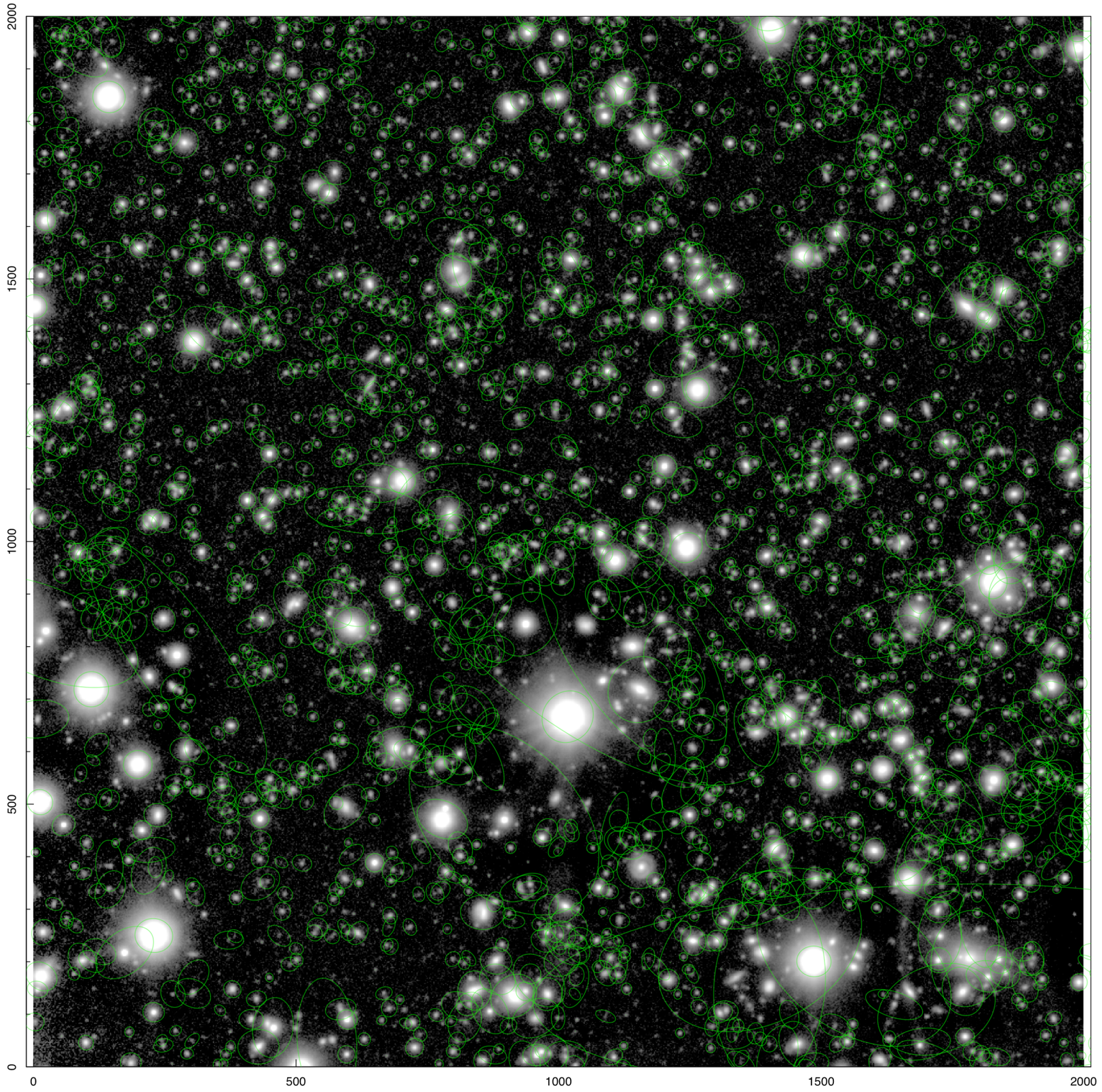
MNRAS **476,** 3137–3159 (2018)
Advance Access publication 2018 February 23

doi:10.1093/mnras/sty440

PROFOUND: SOURCE EXTRACTION AND APPLICATION TO MODERN SURVEY DATA

A. S. G. Robotham,[1★] L. J. M. Davies,[1] S. P. Driver,[1] S. Koushan,[1] D. S. Taranu,[1,2] S. Casura[3] and J. Liske[3]

THE PROBLEM WITH SEXTRACTOR APERTURES

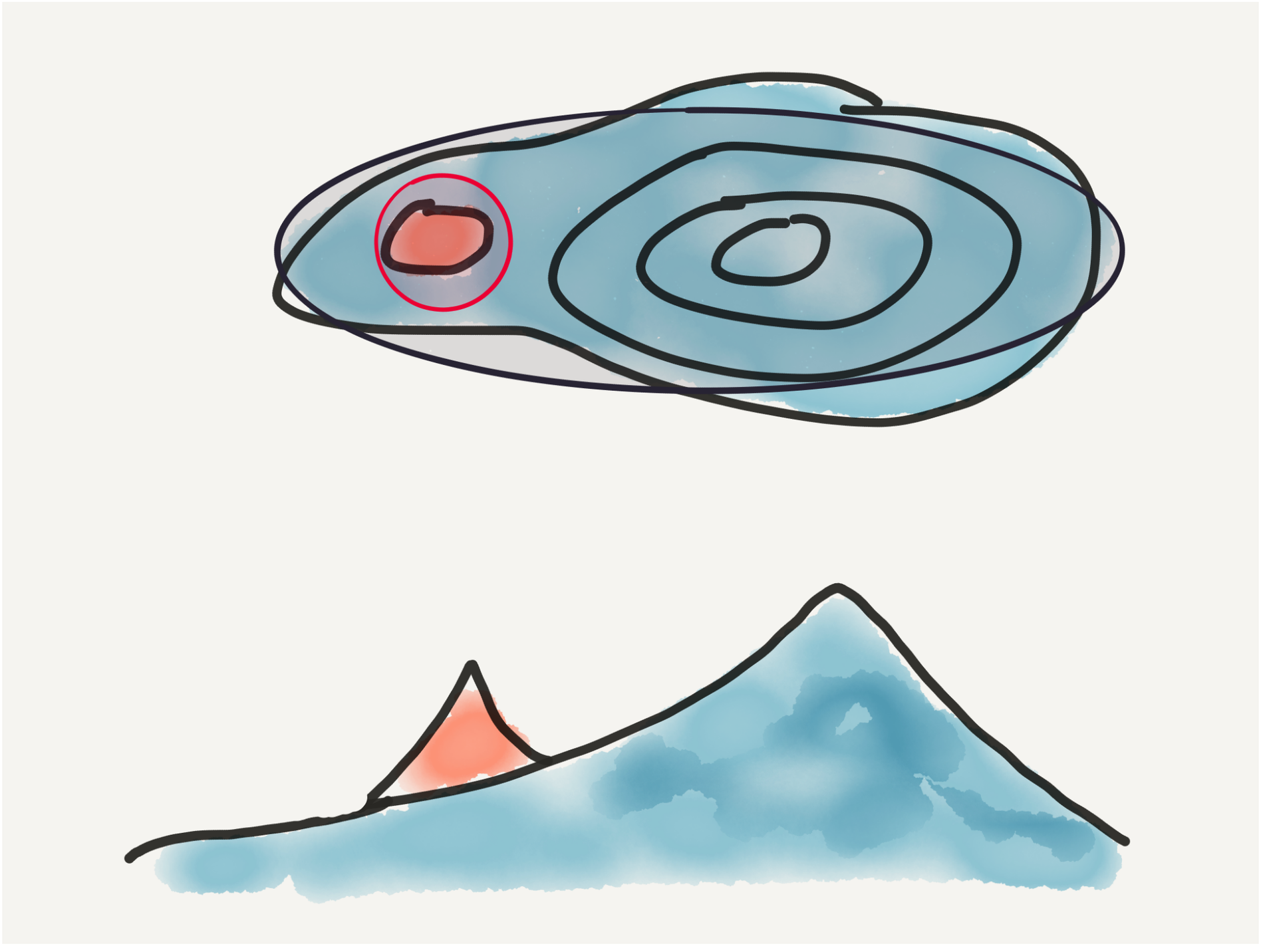


Ultra-VISTA Y-band stack + final release catalogue

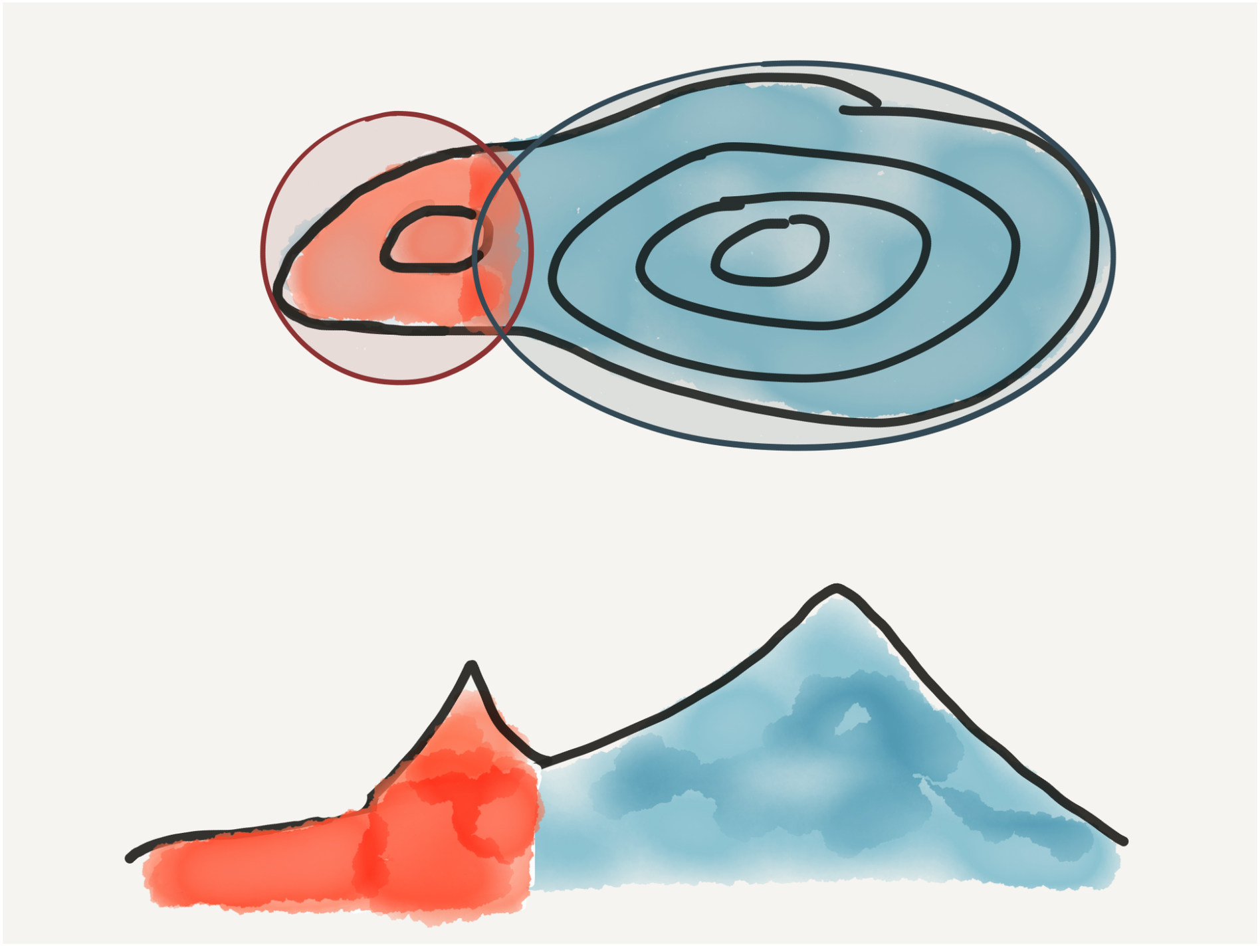
THE NEED TO START AGAIN

- For GAMA we investigated a lot of effort in manual aperture fixing. This was not scaleable or transferable to other data sets, we sought a better solution for WAVES.
- In short, I started again with the source extraction.
- It was not obvious what improvements might be possible over SExtractor (given how well tested and established it is) but two areas quickly came to light:
  - It does not watershed de-blend optimally (the most common failure we see is due to this). It does coarse island-based deblending.
  - It uses strictly elliptical apertures and then tries to distribute overlapping flux using a number of opaque internal schemes.

SEXTRACTOR TENDS TO CREATE WATERSHED ISLANDS



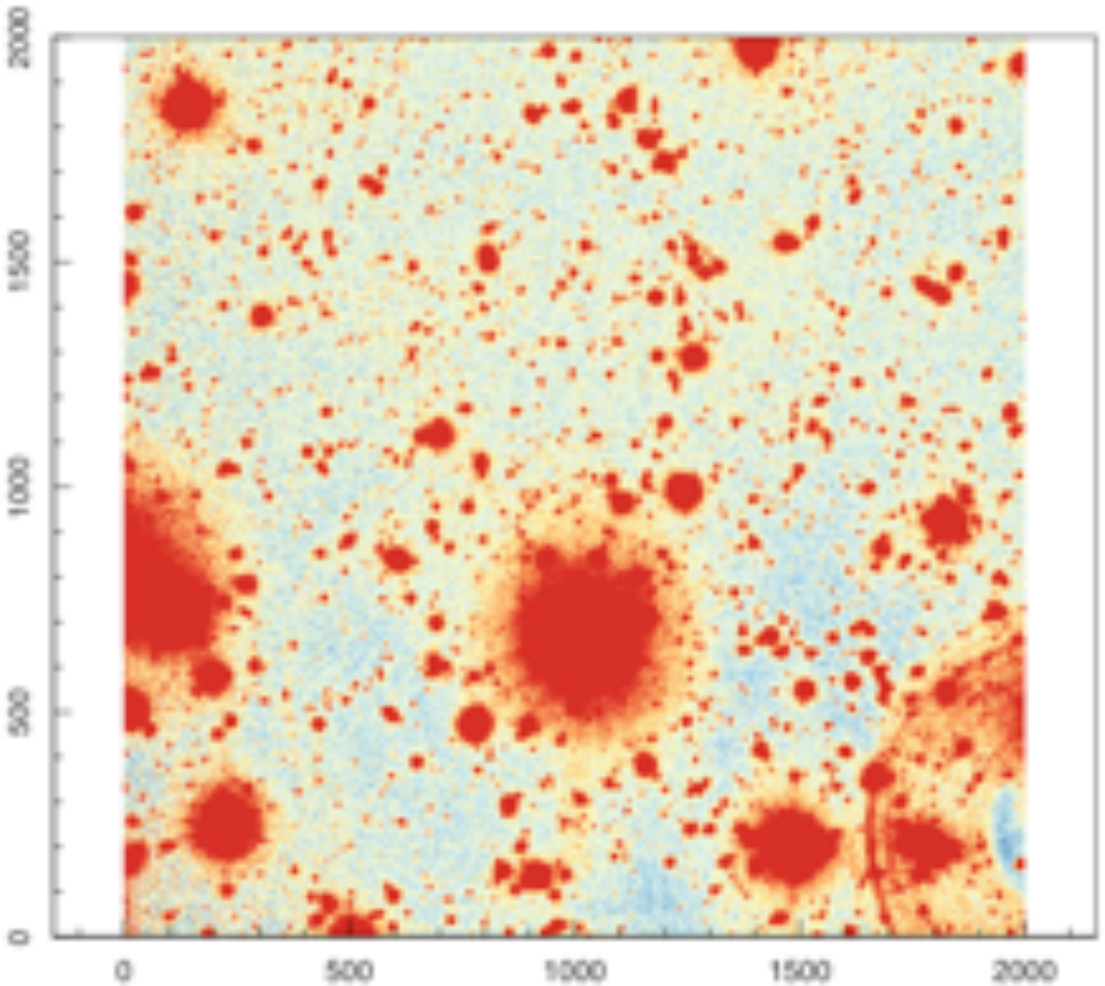
PROFOUND WATERSHEDS THROUGH SADDLE CONTOURS

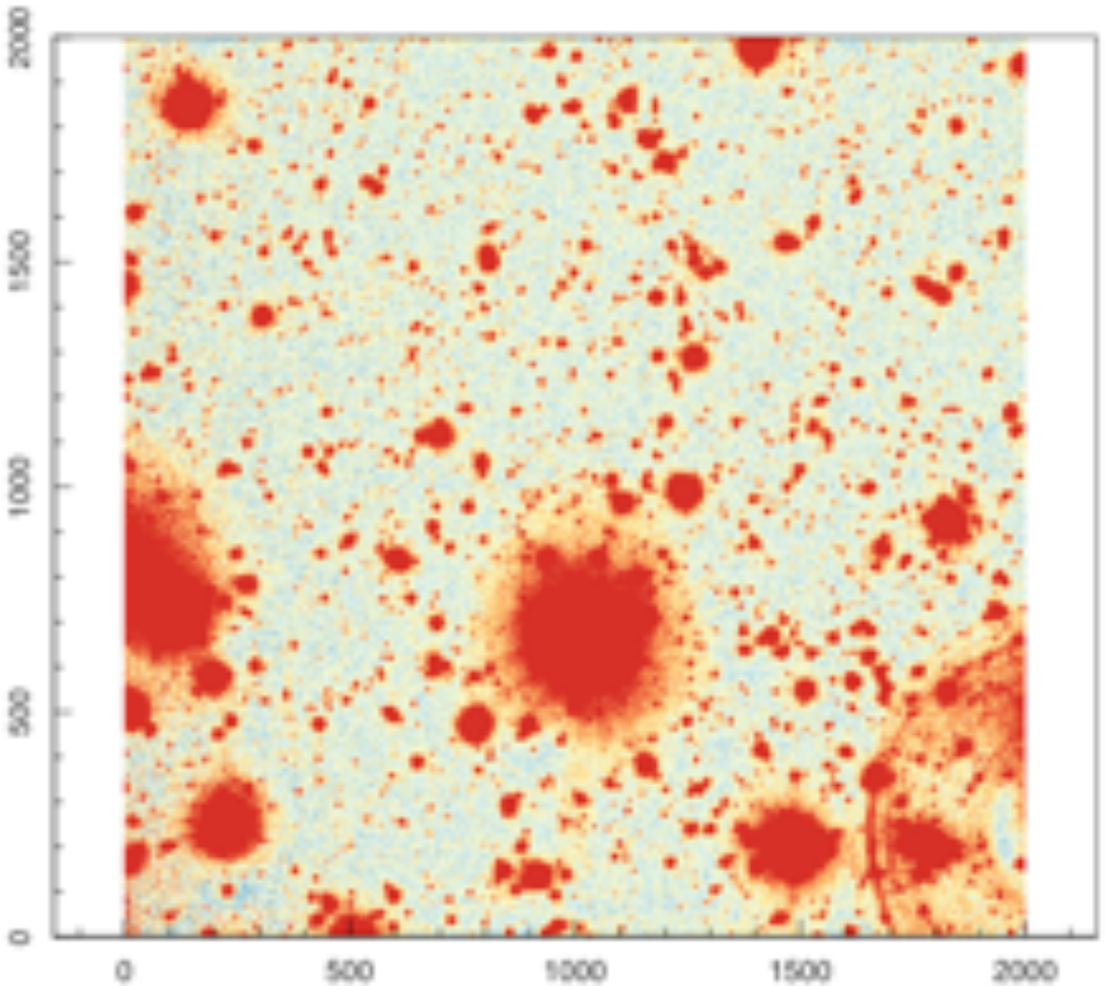


OUR PROFOUND SOLUTION GETS ROUND BOTH THESE ISSUES

- We use a similar approach to find the initial high S/N images segments:
  - Careful sky subtraction (iterative masking and clipping)
  - S/N=1.5 threshold as standard (can change)
  - Segments are de-blended to some tolerance (using a different algorithm to SExtractor- non-discretised surface brightness / sky-RMS thresholds and locally adaptive).
  - Segments are grown organically- apertures never used.

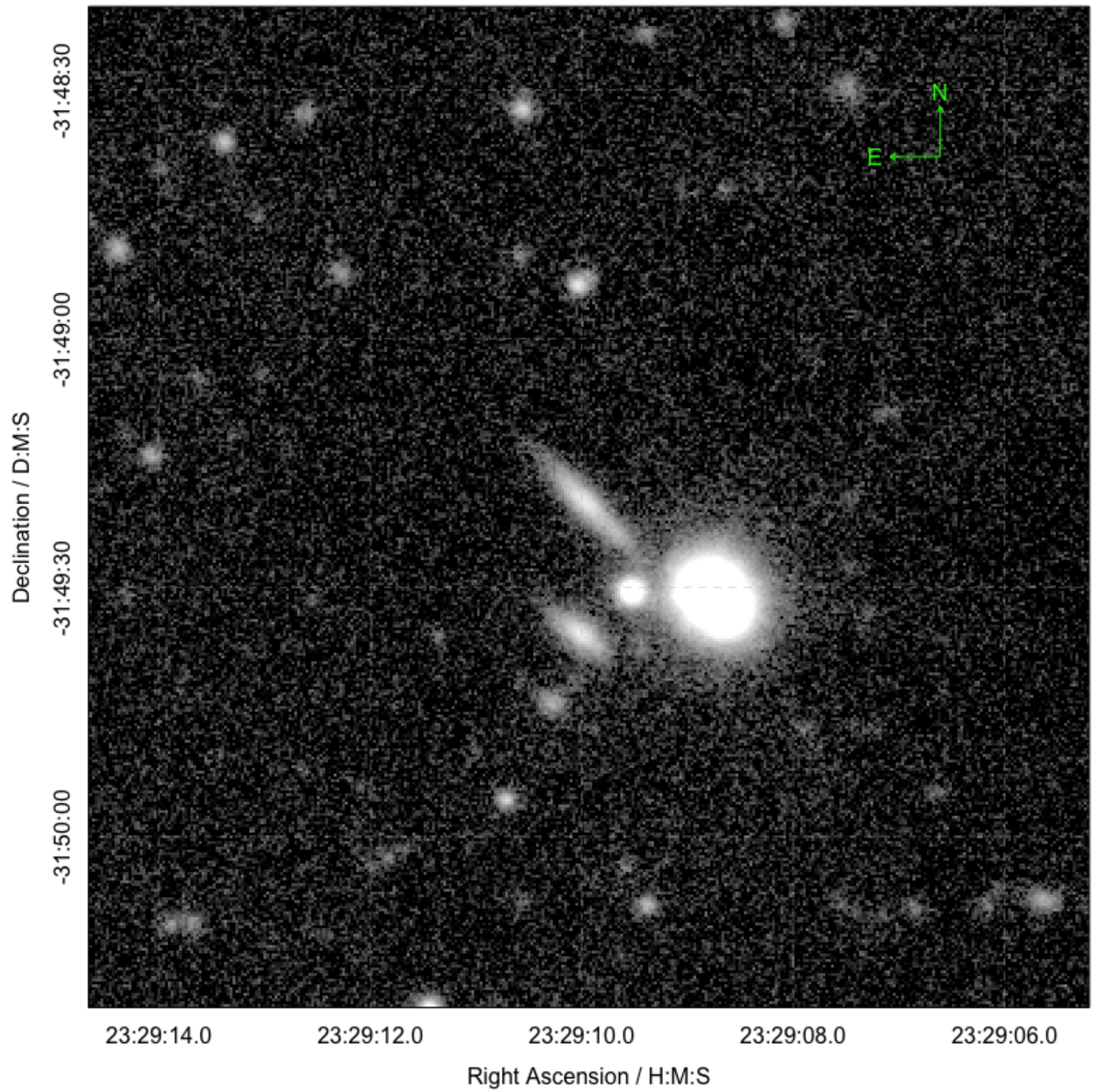
PROFOUND USES AN AGGRESSIVE MESH BASED SCHEME



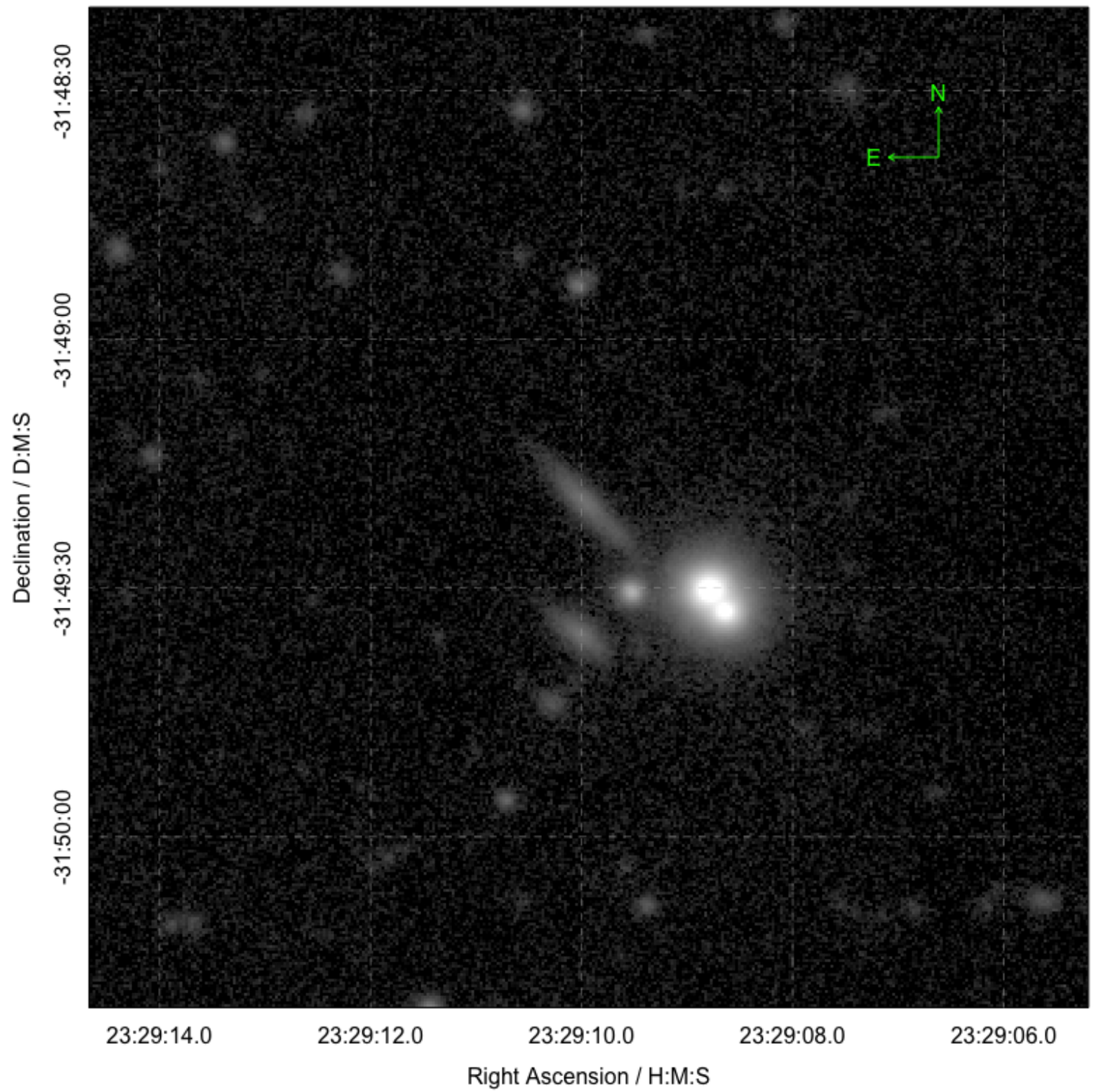


Ultra-VISTA Sextractor sky versus ProFound sky

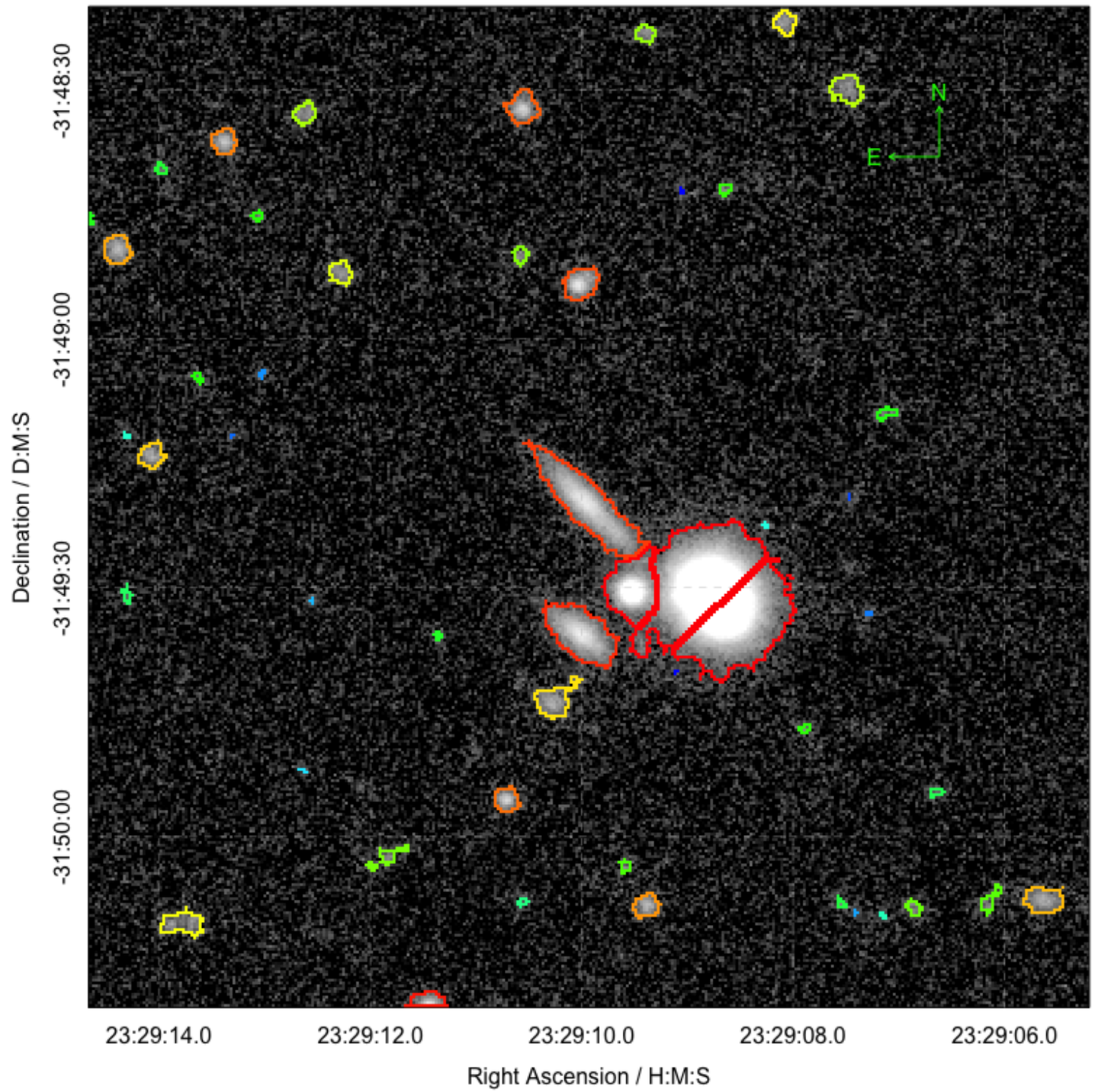
INITIAL VIKING Z-BAND IMAGE



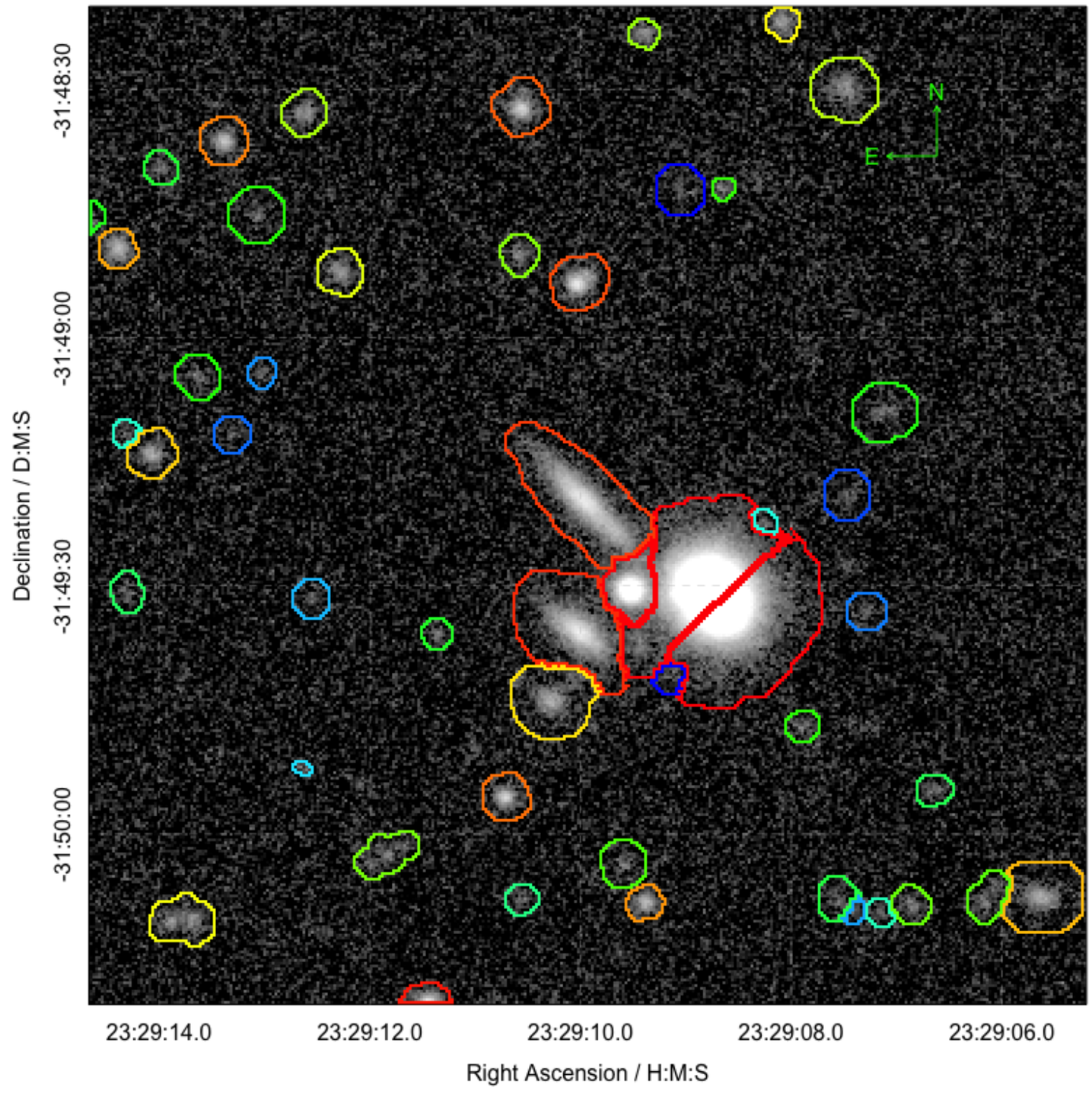
INITIAL VIKING Z-BAND IMAGE



BRIGHT SEGMENTS IDENTIFIED AND DE-BLENDED



SEGMENTS DILATED UNTIL THE FLUX CONTAINED CONVERGES



PROFOUND → PROFIT

- Here we have focussed on the general photometry aspects of ProFound, but it is at least equally focussed on creating good inputs for ProFit.
- It make everything we need:
  - Source identification (star/gal separation)
  - Good initial conditions (based on photometric analysis)
  - Segmentation maps (both tight and dilated to capture *all* flux)
  - Sky (can be calculated a number of ways as appropriate)
  - Sigma maps (using sky variance and image gain)

ProFit

TAKING PHOTOMETRY FURTHER

PROFIT

MNRAS **466,** 1513–1541 (2017)
Advance Access publication 2016 November 23

doi:10.1093/mnras/stw3039

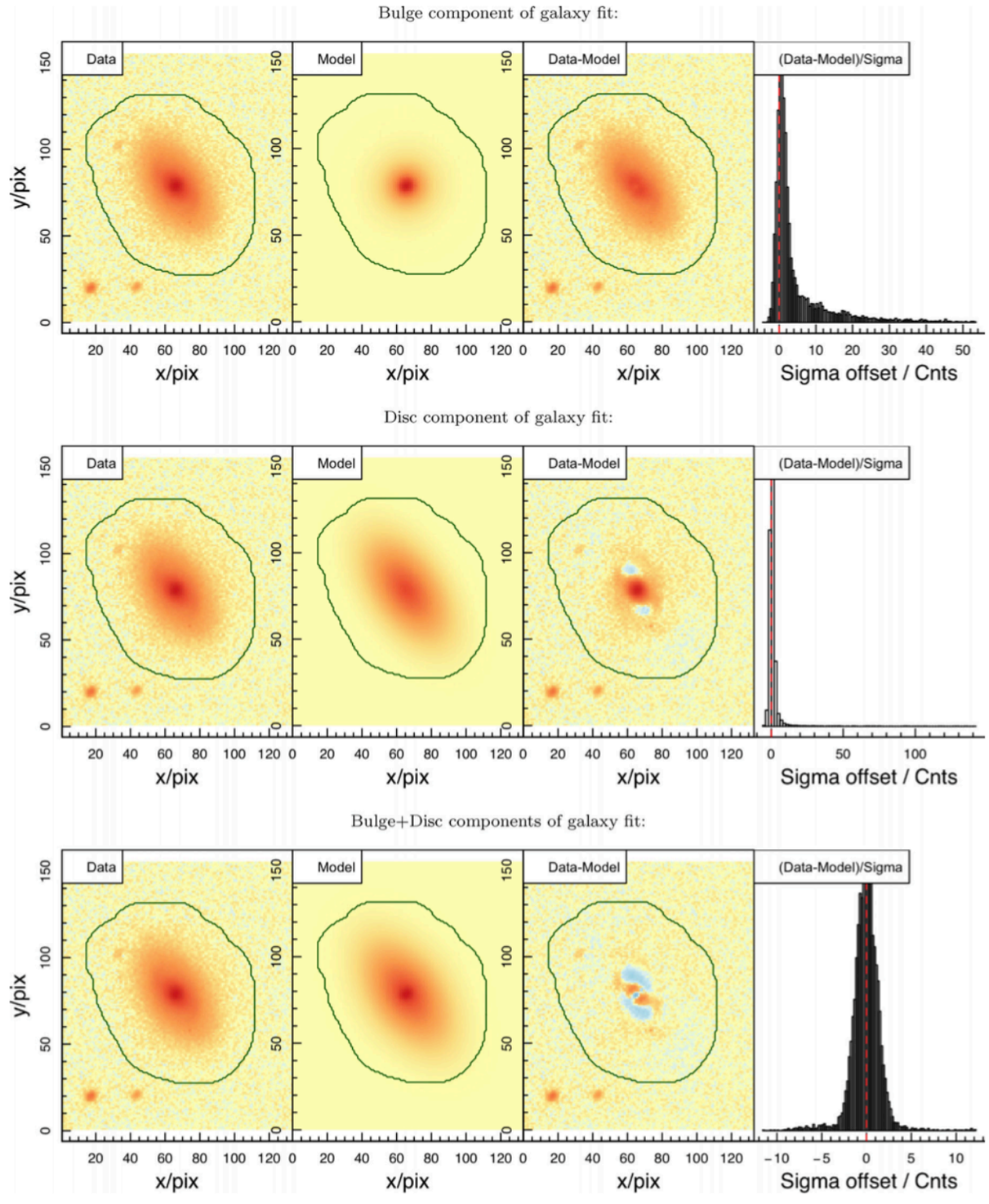
PROFIT: Bayesian profile fitting of galaxy images

GitHub: ICRAR/ProFit

A. S. G. Robotham,[1★] D. S. Taranu,[1,2] R. Tobar,[1] A. Moffett[1] and S. P. Driver[1]

What's the Aim?

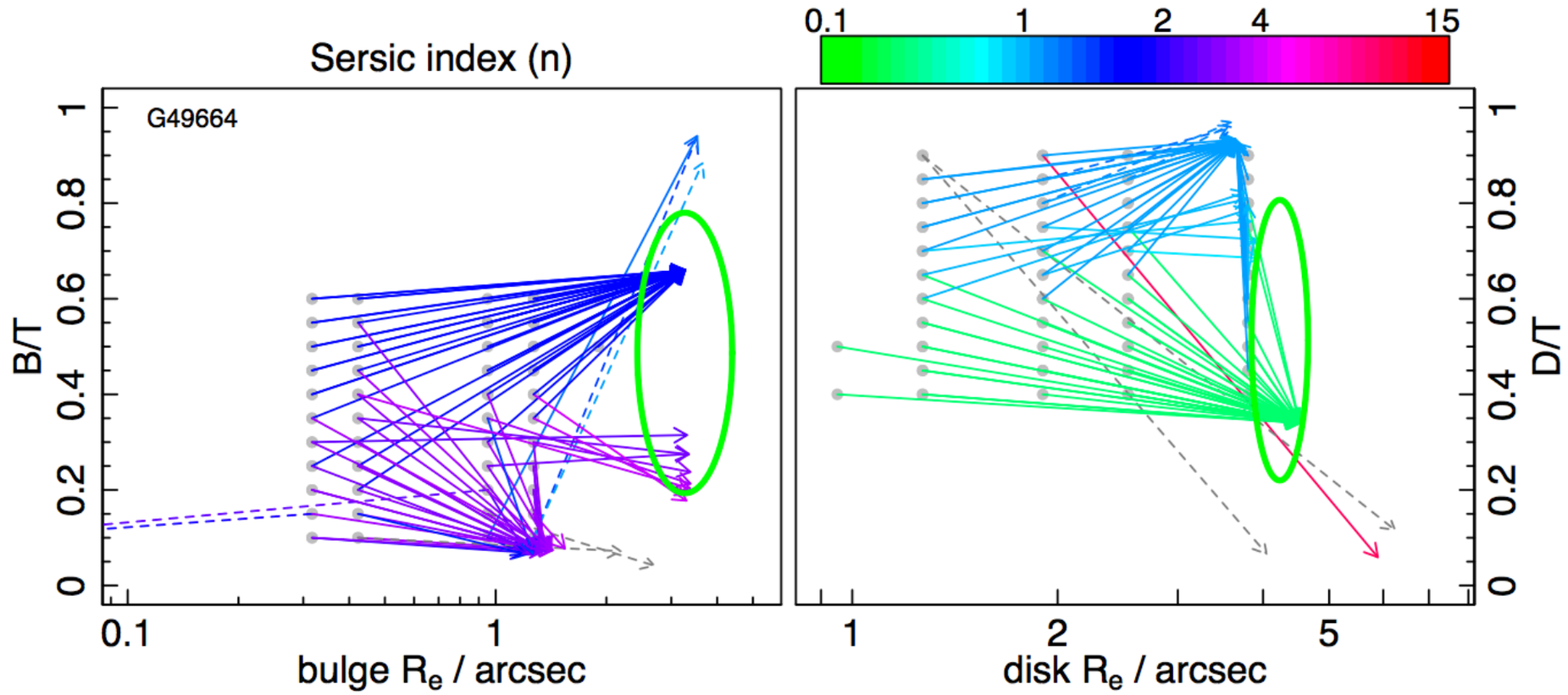
- In simple "single band mode" we want to produce a 2D mixture model of a galaxy image.
- This usually separates a galaxy into a compact bulge and and extended disk.



What's the Problem?

- Work by ICRAR PhD R. Lange showed that the code of popular community choice (GALFIT) struggles with convergence for many galaxies:

Convergence plots



Lange et al (2016)

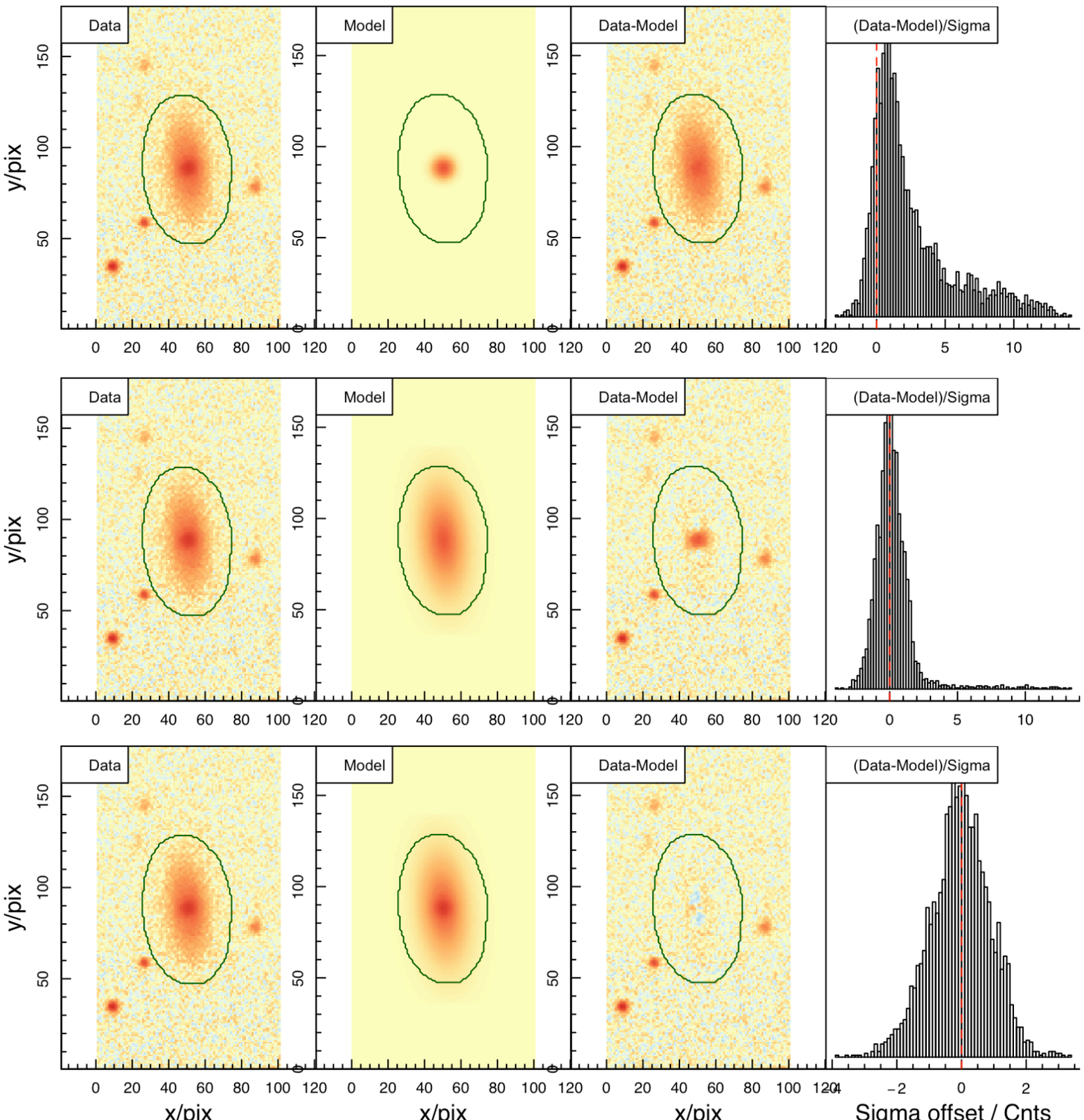
What to do?

- Options were to modify existing code to use more sophisticated samplers, but GALFIT is closed license (binaries only) and other codes available at the time were not battle-tested.
- Vitally for ProFuse (already in the planning in 2017), we need image generation capability, not just ‘fitting’.
- Best option all round was to start again and develop a rapid image generation library written in C++ (libprofit) and higher level interfaces to achieve the fitting (ProFit, initially Python and R).
- Initial version was written by me, subsequently R. Tobar (libprofit / pyprofit) and D. Taranu (ProFit) have added a lot of additional features and code re-factors.

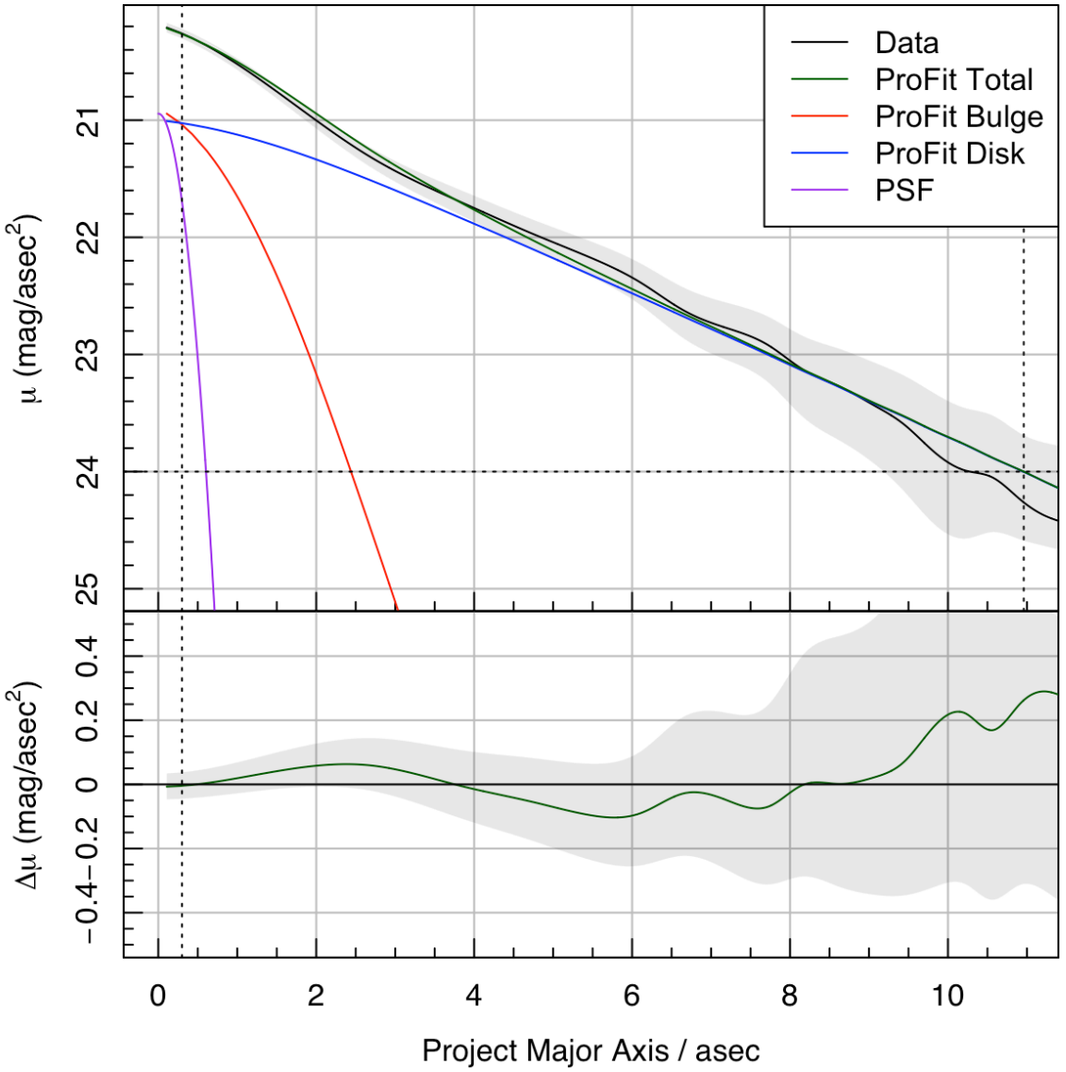
What Does libprofit Do For You?

- Profiles supported in libprofit:
  - Sersic: Popular for galaxy components.
  - Core-Sersic: Popular for early-type galaxies.
  - Moffat: Popular for stars / PSFs.
  - Ferrer/s: Popular for galaxy bars.
  - King: Popular for globular clusters.
  - Empirical PSF: Generic point spread function.
  - Sky: The sky background.
- Easy (relatively) to add more profiles: ~30 lines of C++ code to describe the 1D radial profile shape and total flux calculation.

Decomposing Bulges and Disks with ProFit

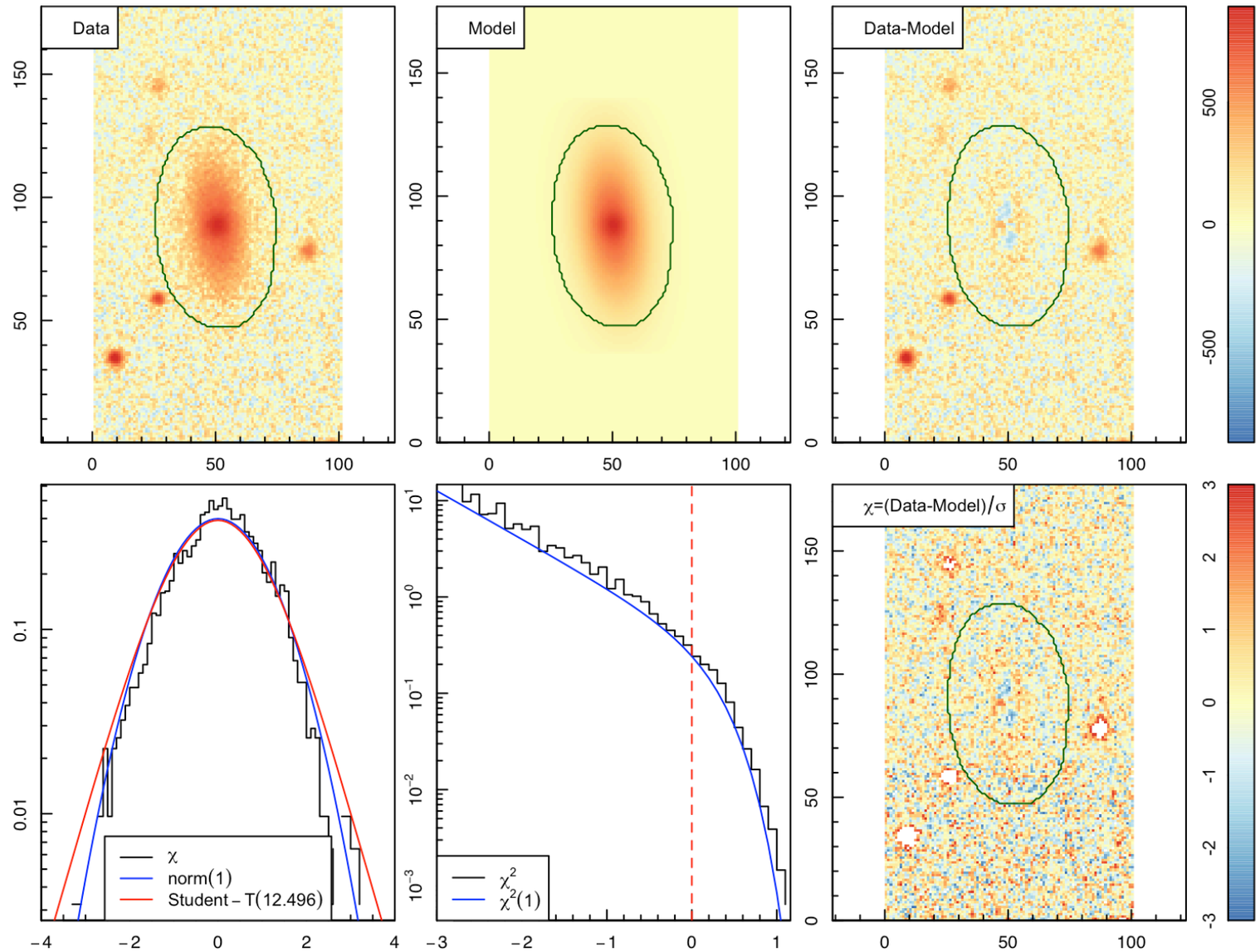


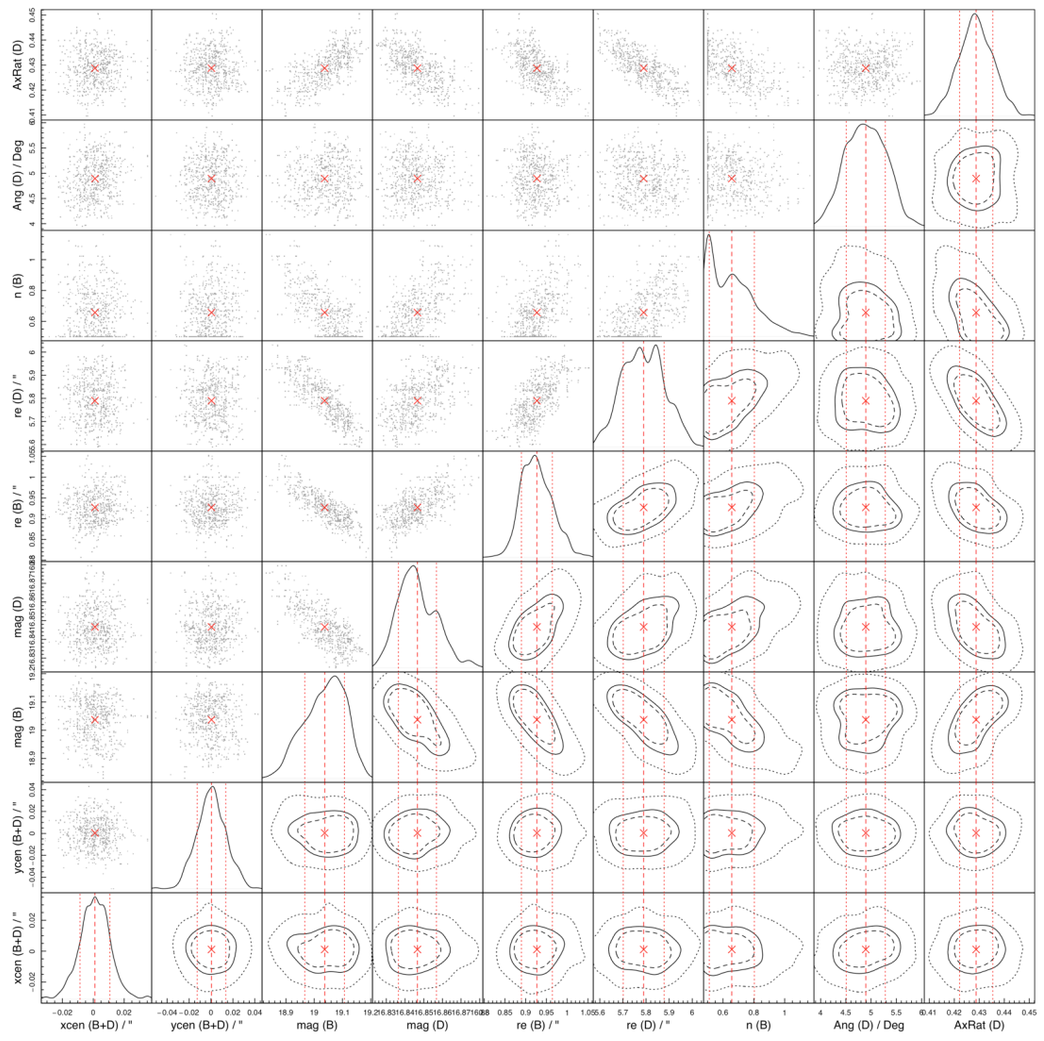
- Can achieve converged MCMC fits with useful errors!



Decomposing Bulges and Disks with ProFit

- Errors are provided via full covariance information.





ProSpect

TAKING SPECTRA TO THE EDGE

PROSPECT

GitHub: asgr/ProSpect

MNRAS **495,** 905–931 (2020)
Advance Access publication 2020 April 26

doi:10.1093/mnras/staa1116

PROSPECT: generating spectral energy distributions with complex star formation and metallicity histories

A. S. G. Robotham [1,2★] S. Bellstedt [1] C. del P. Lagos [1,2] J. E. Thorne [1]
L. J. Davies [1] S. P. Driver[1] and M. Bravo[1]

PROSPECT CODE GOAL

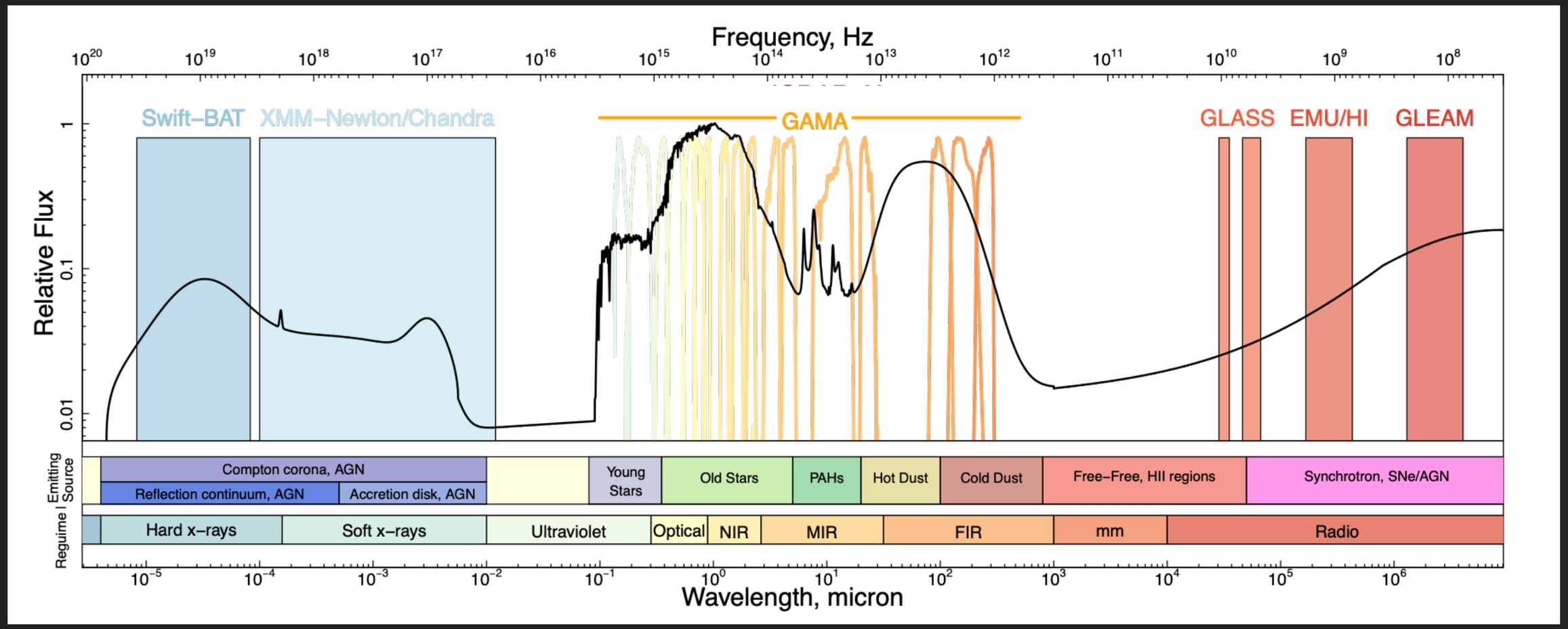
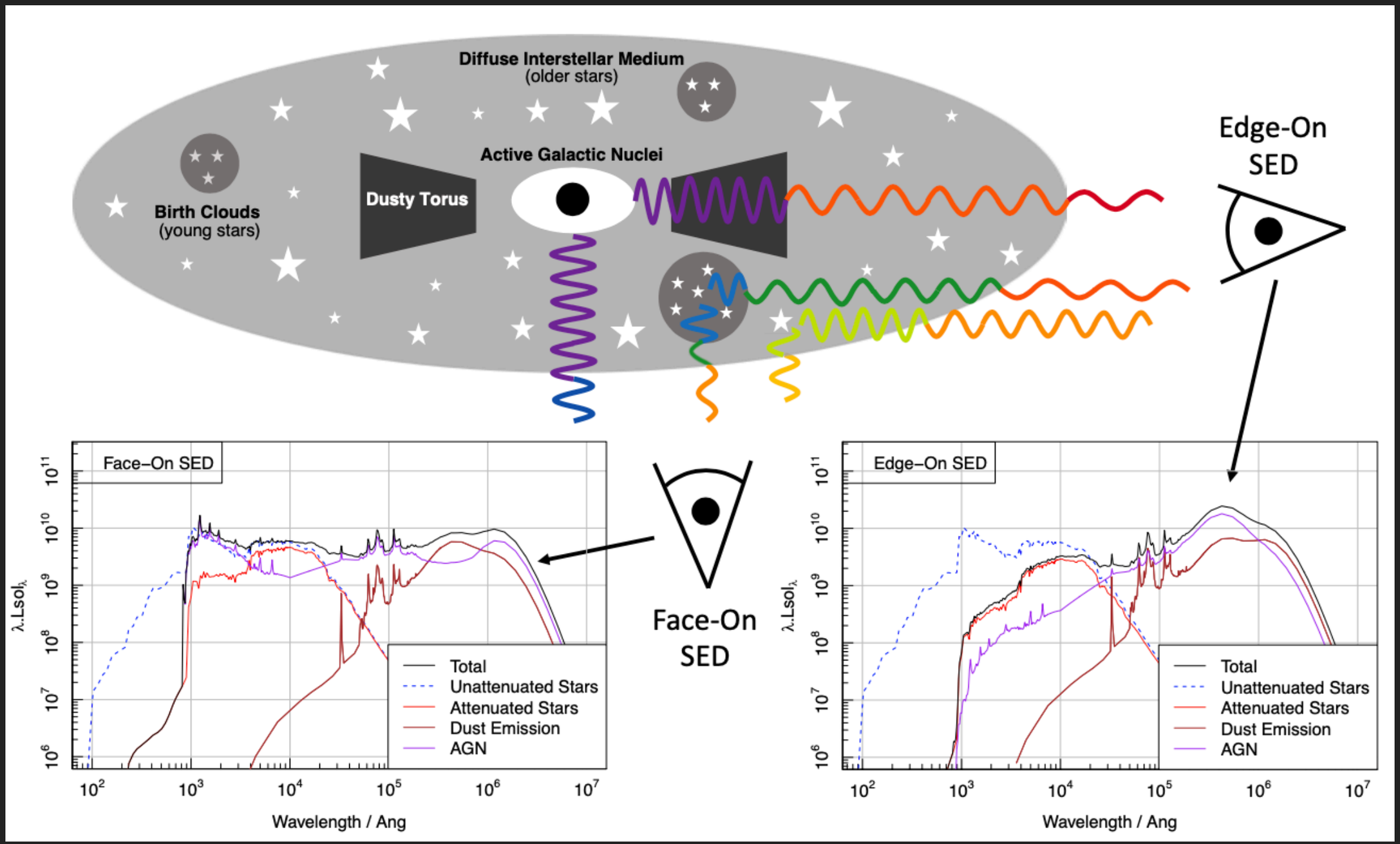
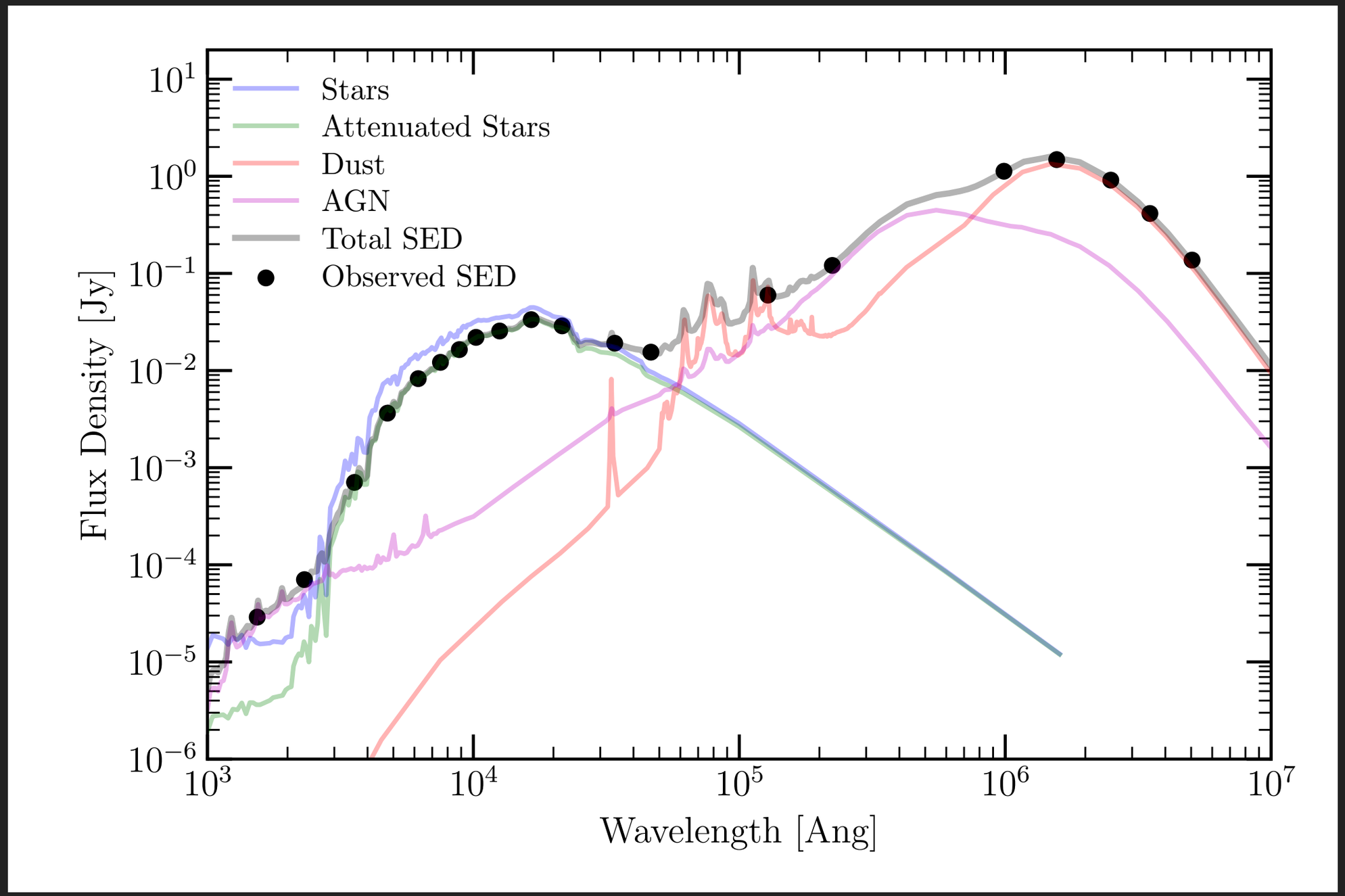


Figure credit: Luke Davies

GALAXY MODEL



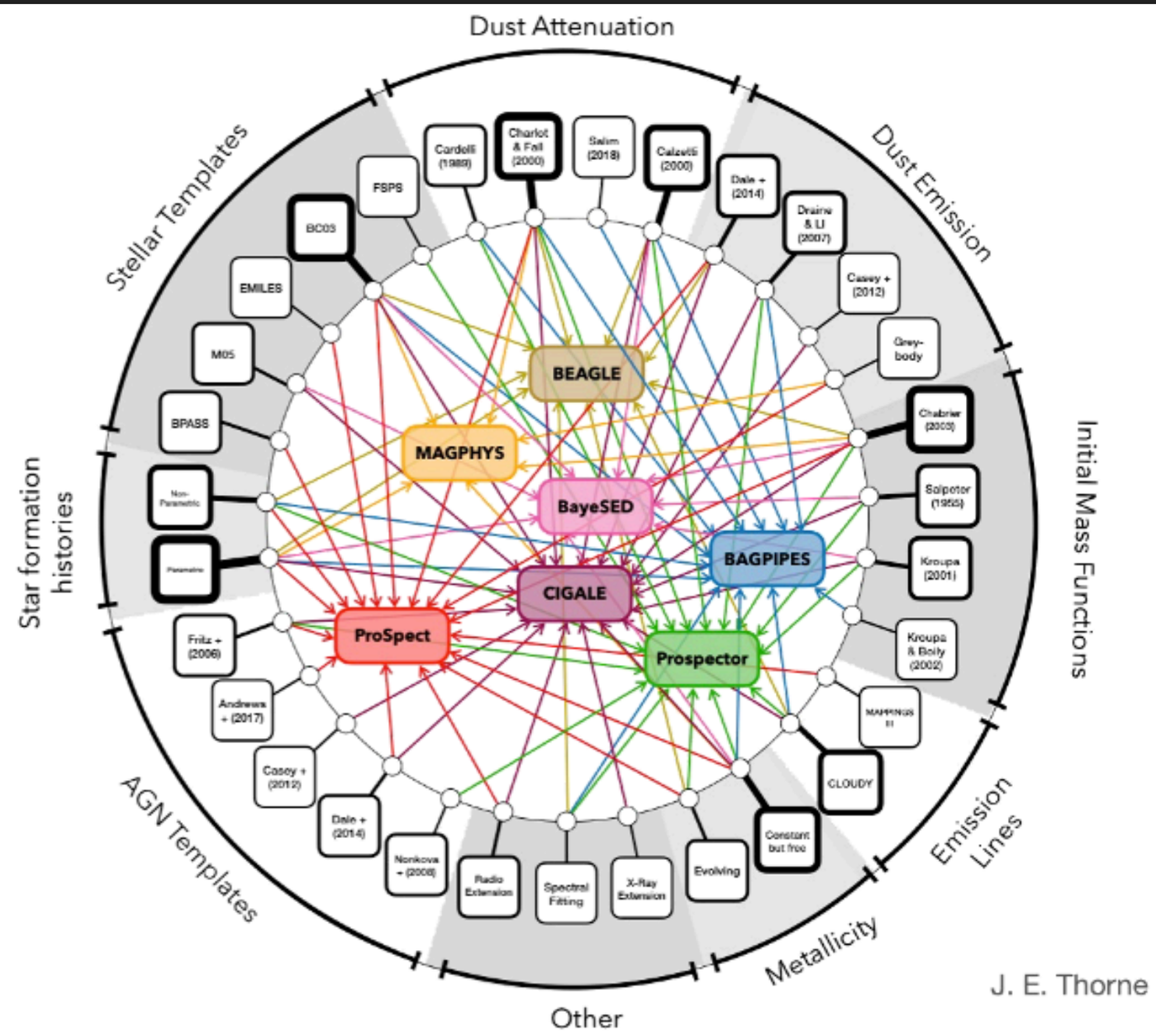
UV – FIR


Stars
Attenuated Stars
Dust
AGN
Total SED
Observed SED
Flux Density [Jy]
Wavelength [Ang]
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{3}
10^{4}
10^{5}
10^{6}
10^{7}

PROSPECT VERSUS

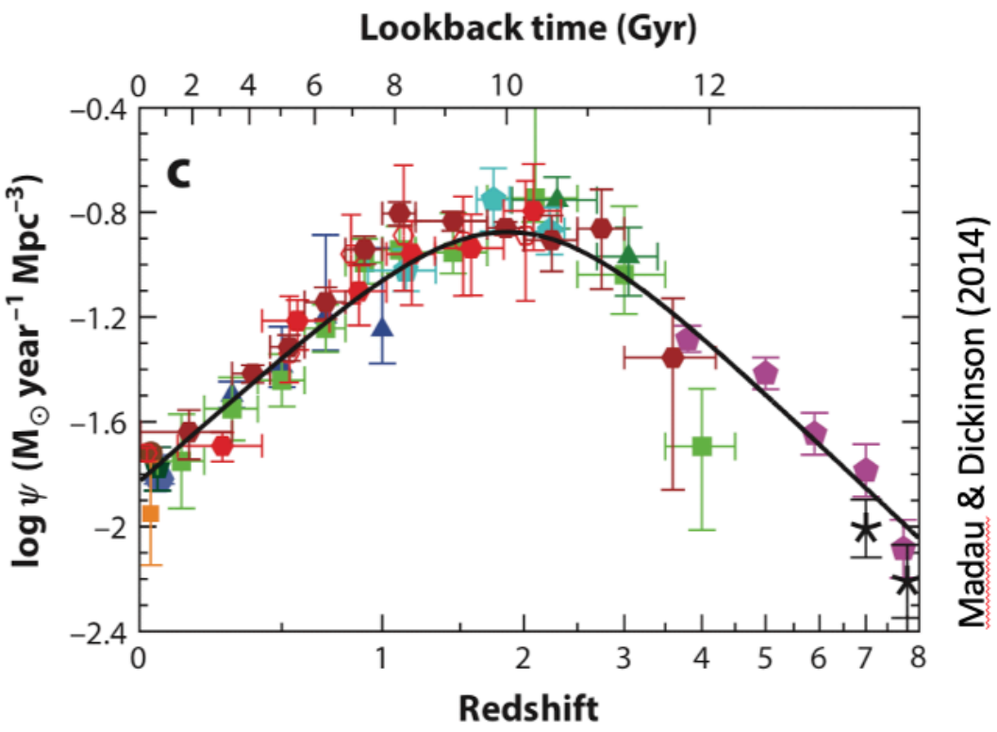
Many SED-fitting tools exist.

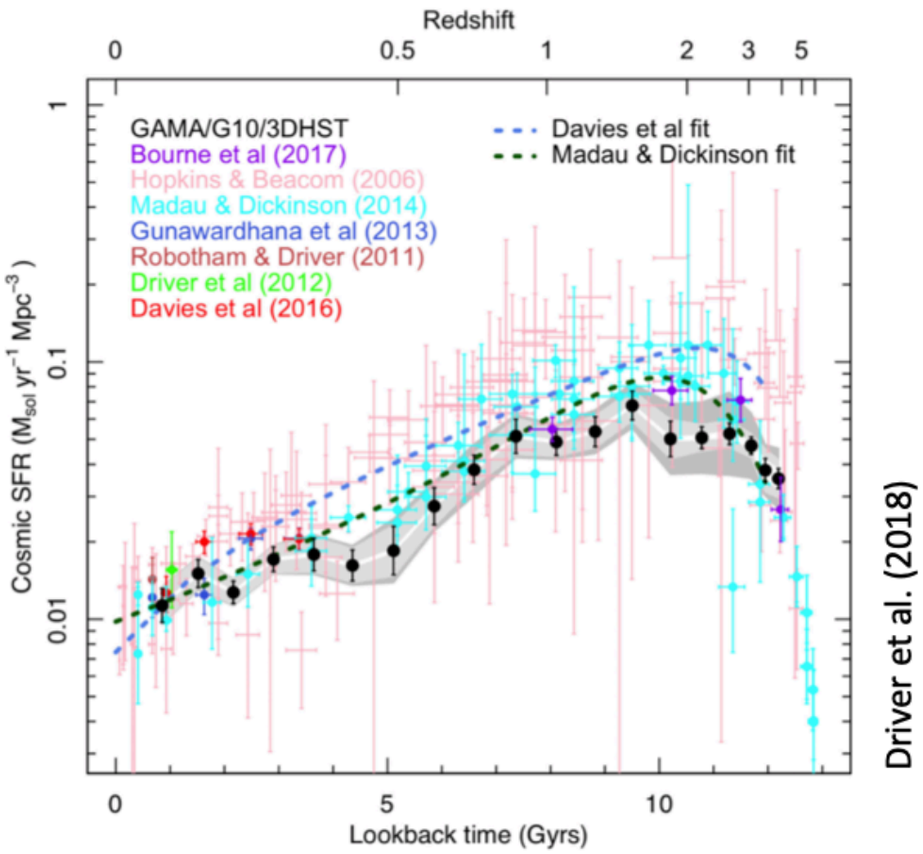
They all do roughly the same thing (model an SED via its stellar and dust components), but all in slightly different ways...



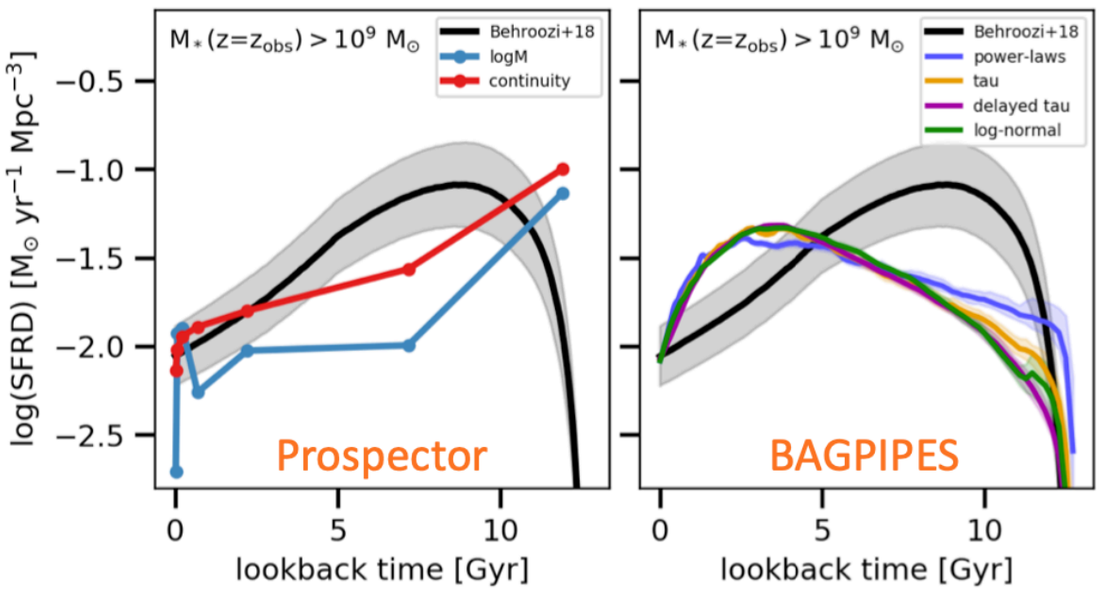
Again, with ProFuse in mind we do not just want an SED ‘fitting’ code, we need a generative model.

PROSPECT ON GAMA (BELLSTEDT+ 2020)





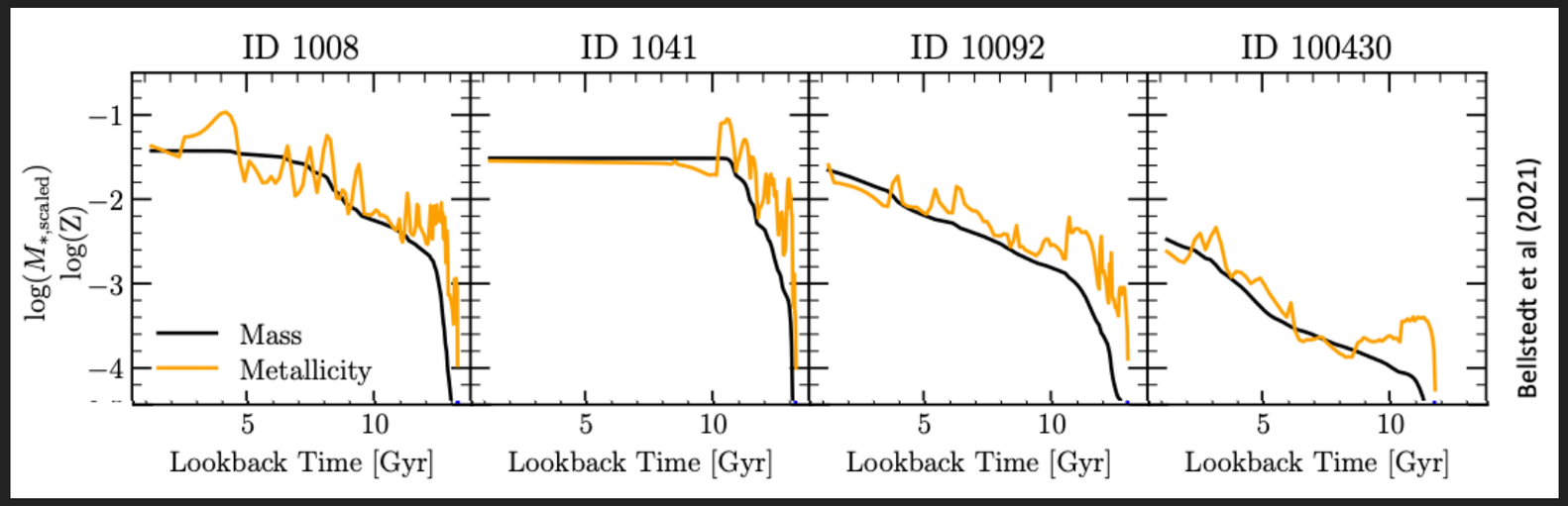
A challenge for SED-fitting techniques



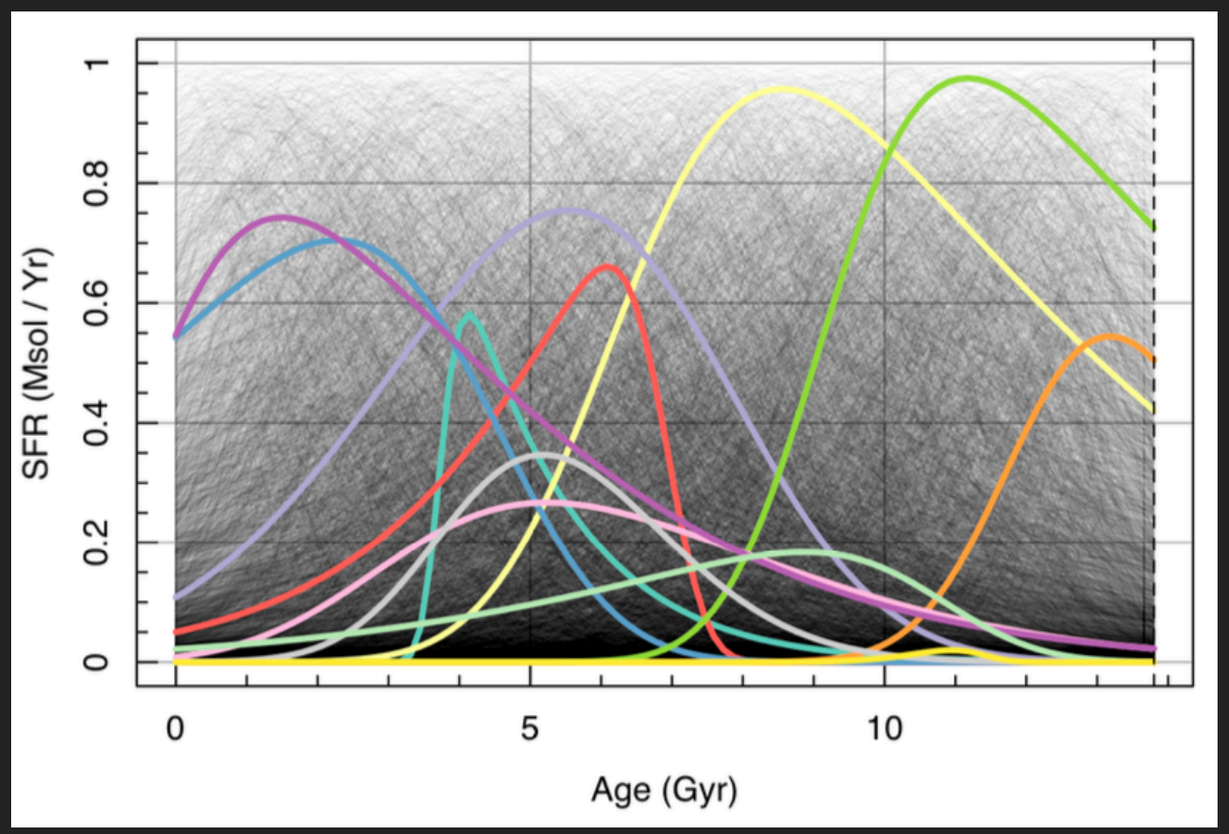
Leja et al. (2018)

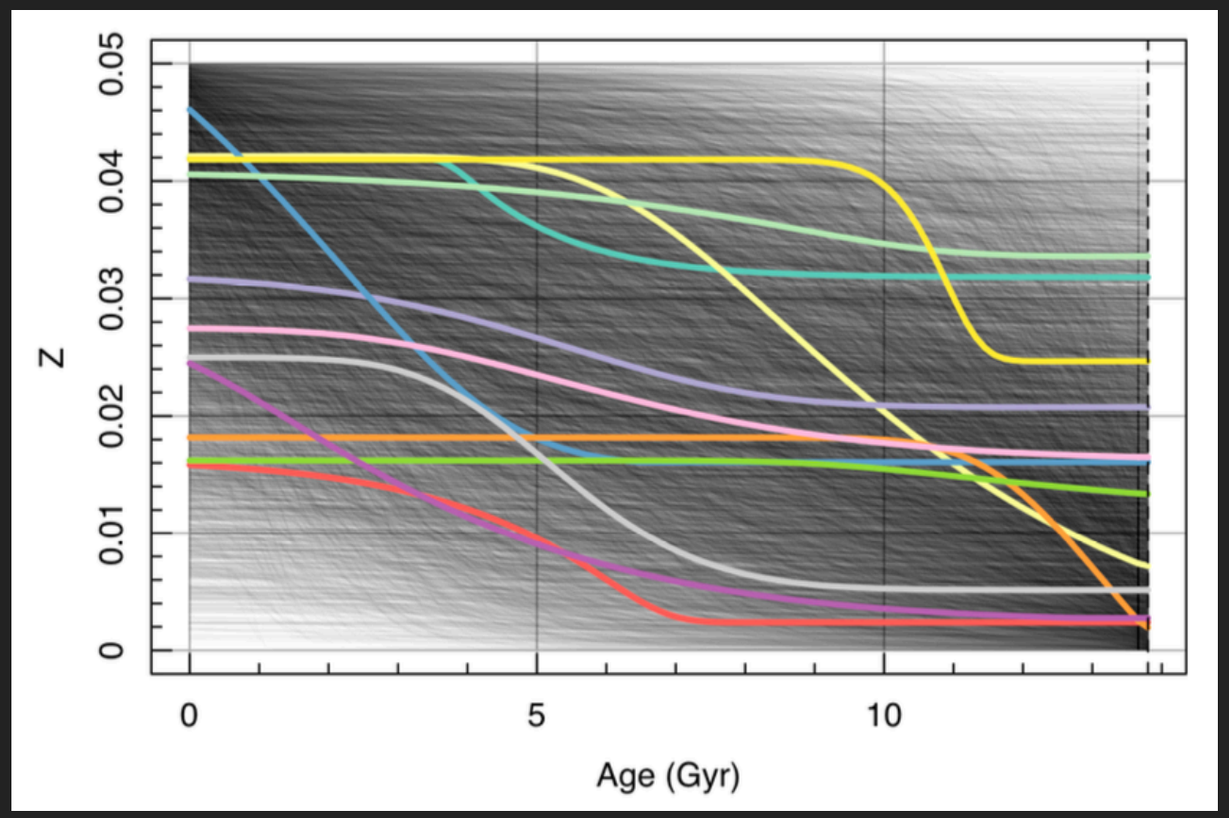
Carnall et al. (2018)

PROSPECT SFH AND ZH

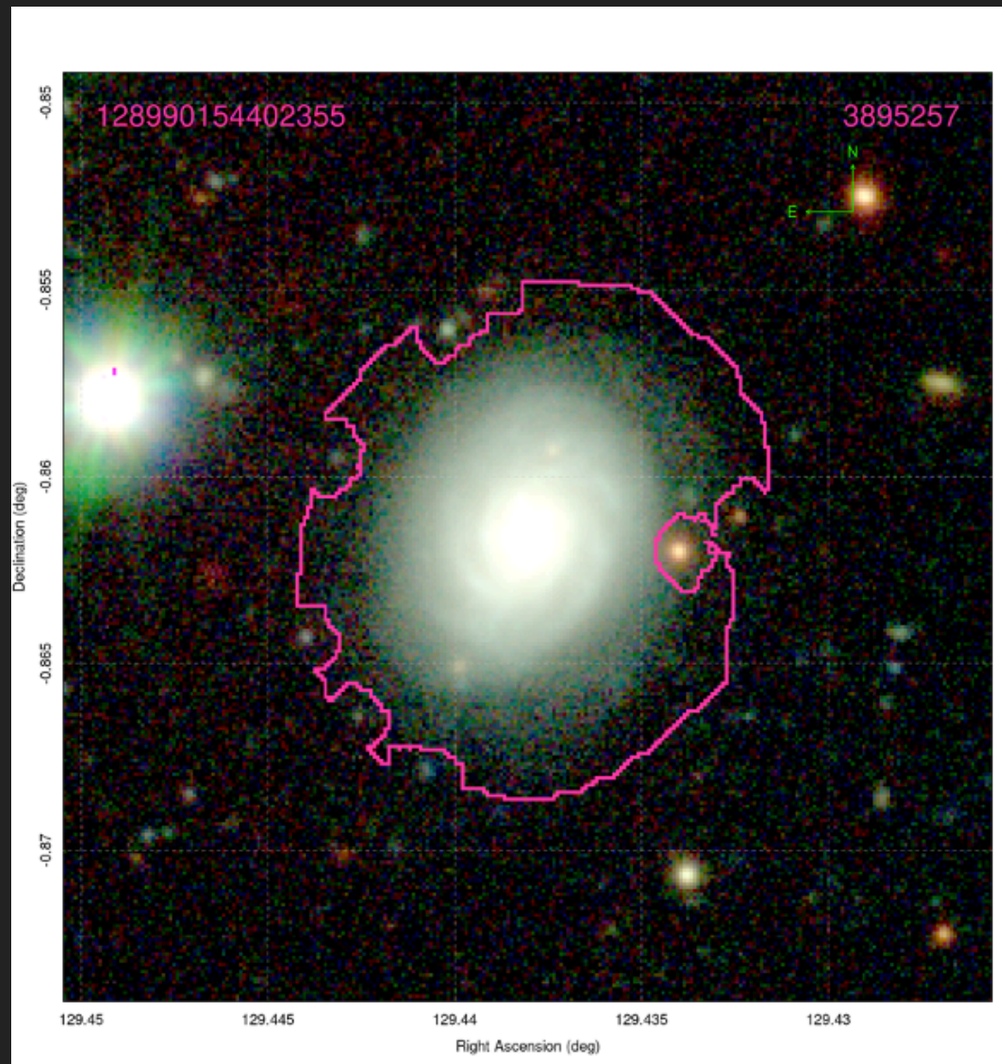


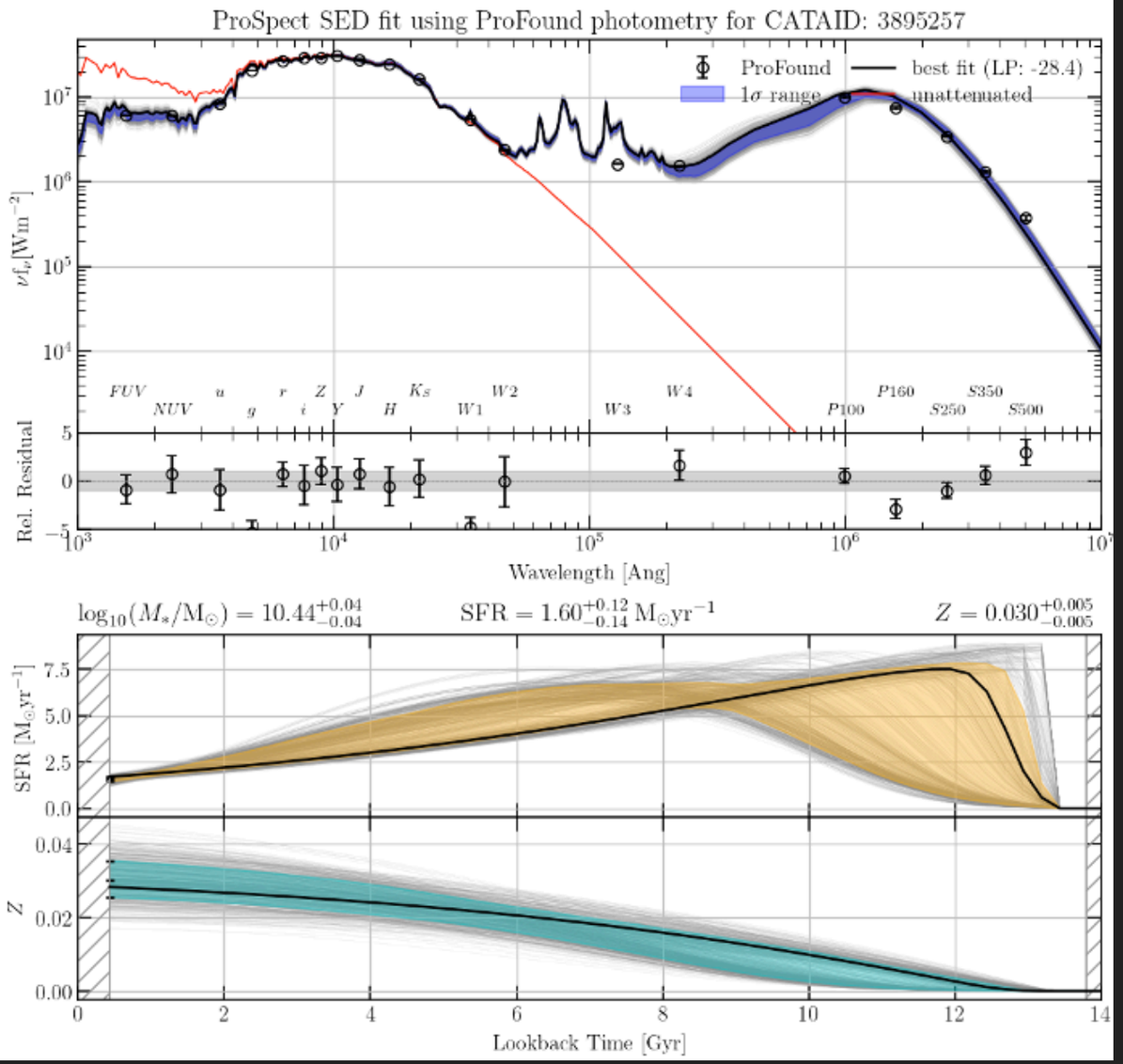
Allows almost any, but we focus on skewed Normal SFH, and linearly mapped ZH:





PROSPECT ON GAMA (BELLSTEDT+ 2020)

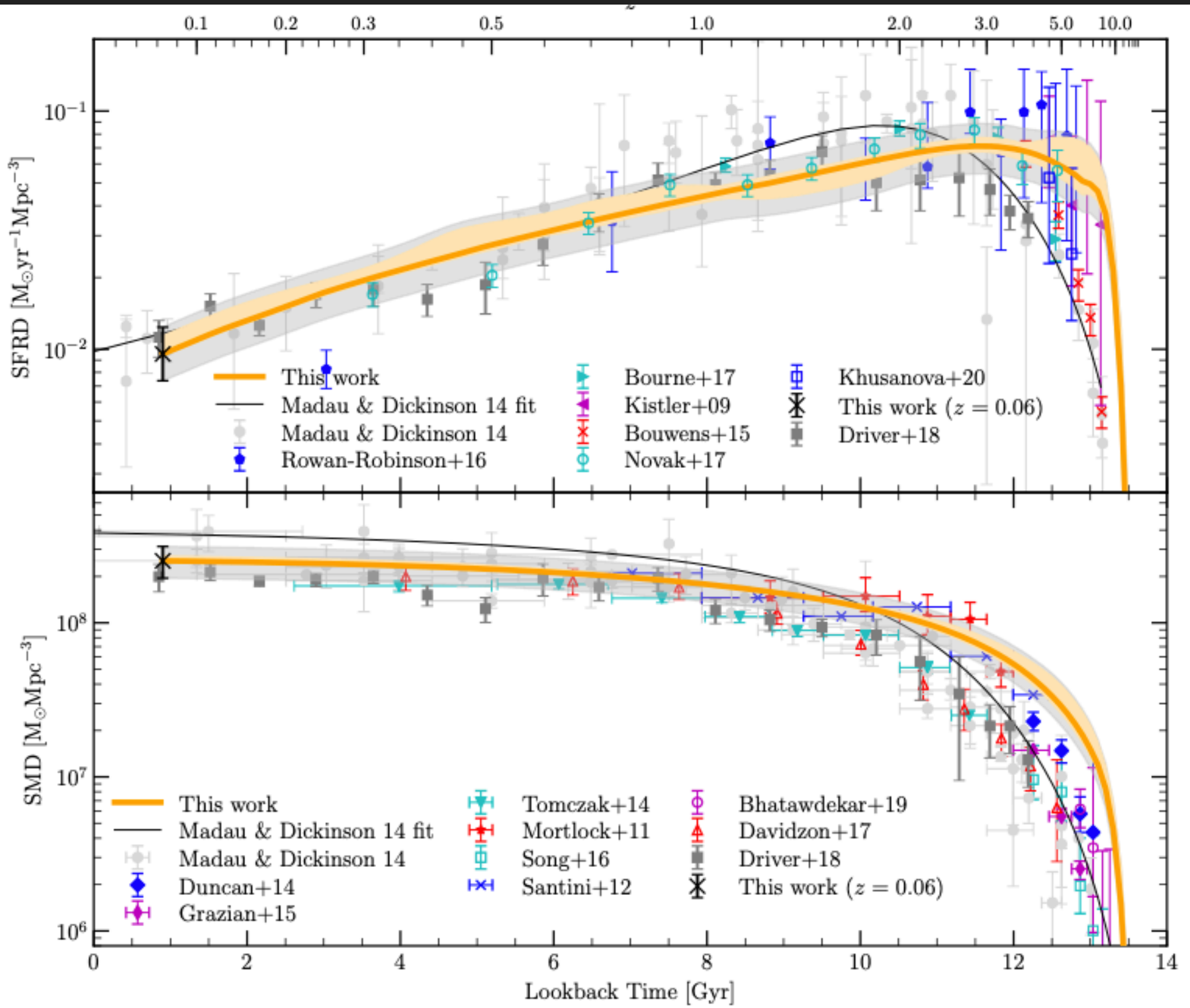




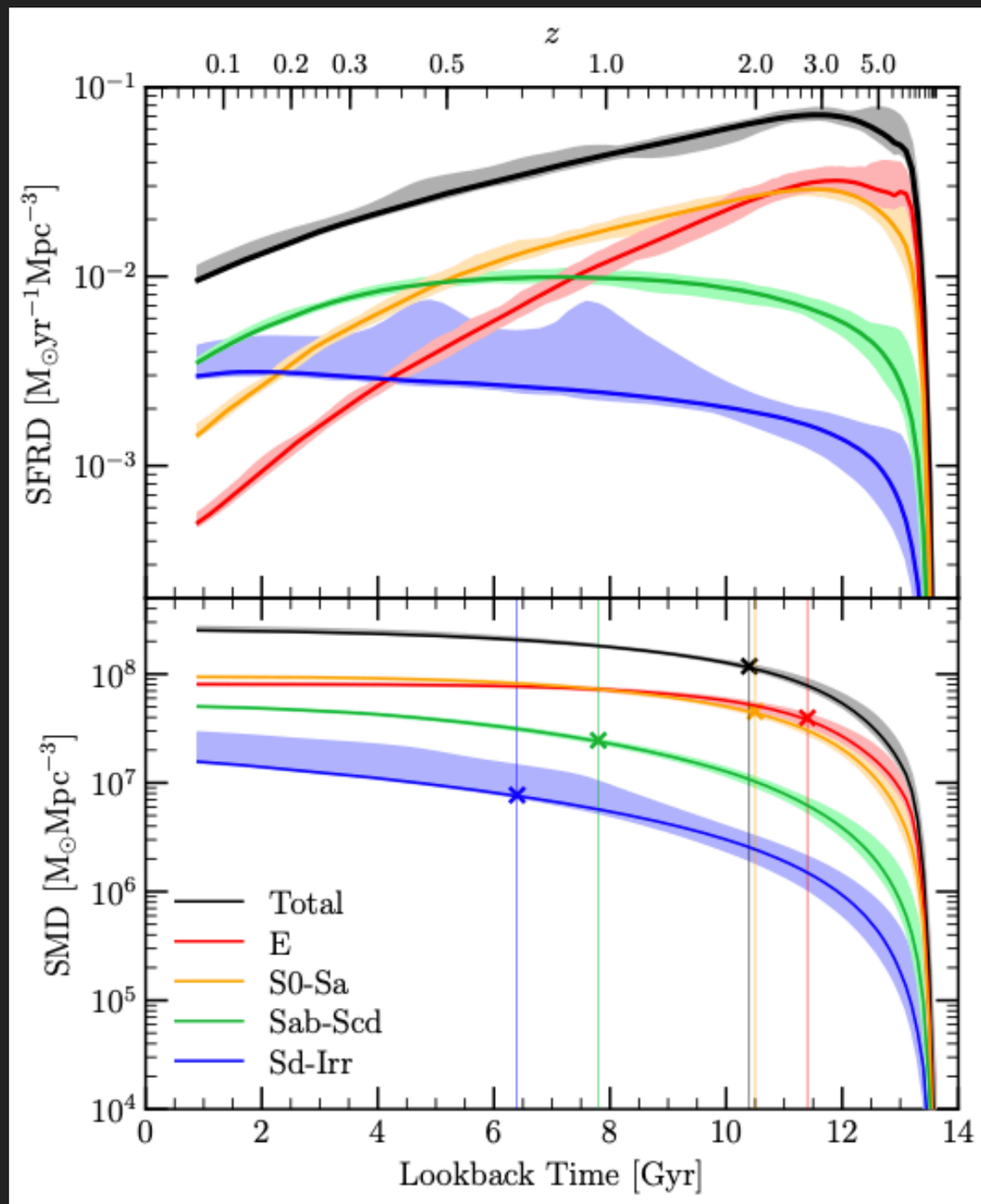
PROSPECT ON GAMA (BELLSTEDT+ 2020)

cosmic star formation history

cosmic stellar mass build-up

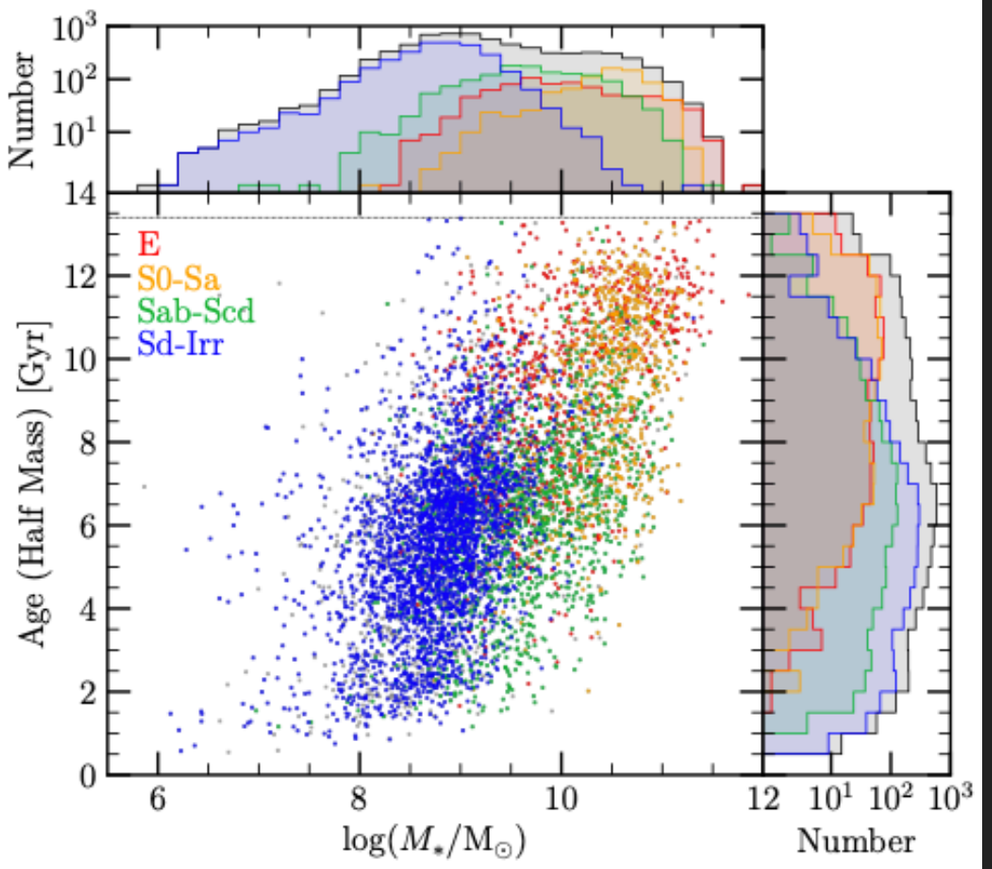


PROSPECT ON GAMA (BELLSTEDT+ 2020)

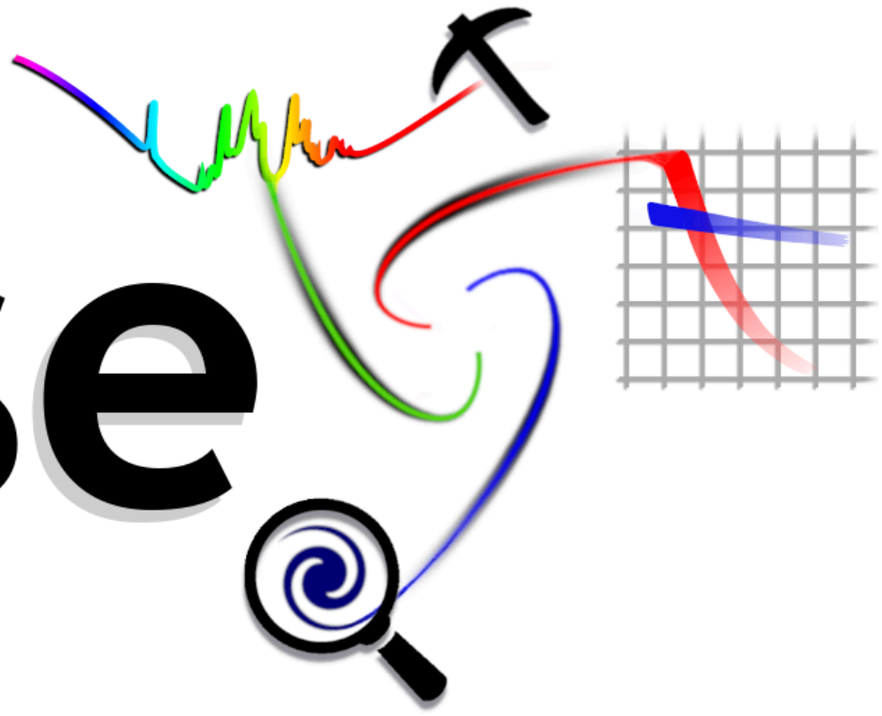


By morphology:

Early-type galaxies are older, whereas late-type galaxies were formed more recently.



ProFuse



PUTTING THIS ALTOGETHER

PROFUSE

GitHub: asgr/ProFuse

MNRAS **513,** 2985–3012 (2022)
Advance Access publication 2022 April 15

https://doi.org/10.1093/mnras/stac1032

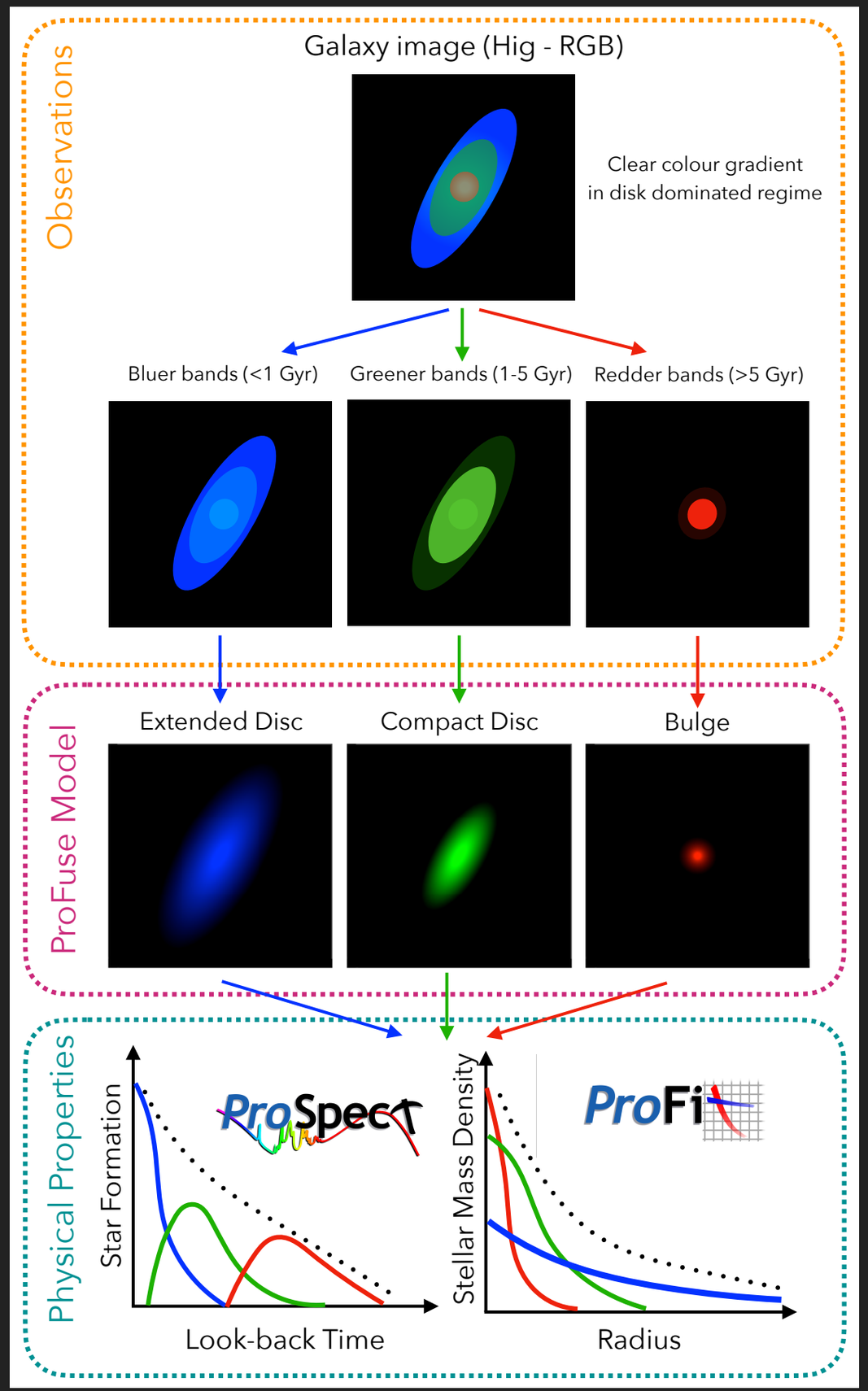
ProFuse: physical multiband structural decomposition of galaxies and the mass–size–age plane

A. S. G. Robotham [1,2★] S. Bellstedt [1] and S. P. Driver [1]

PROFUSE GOAL

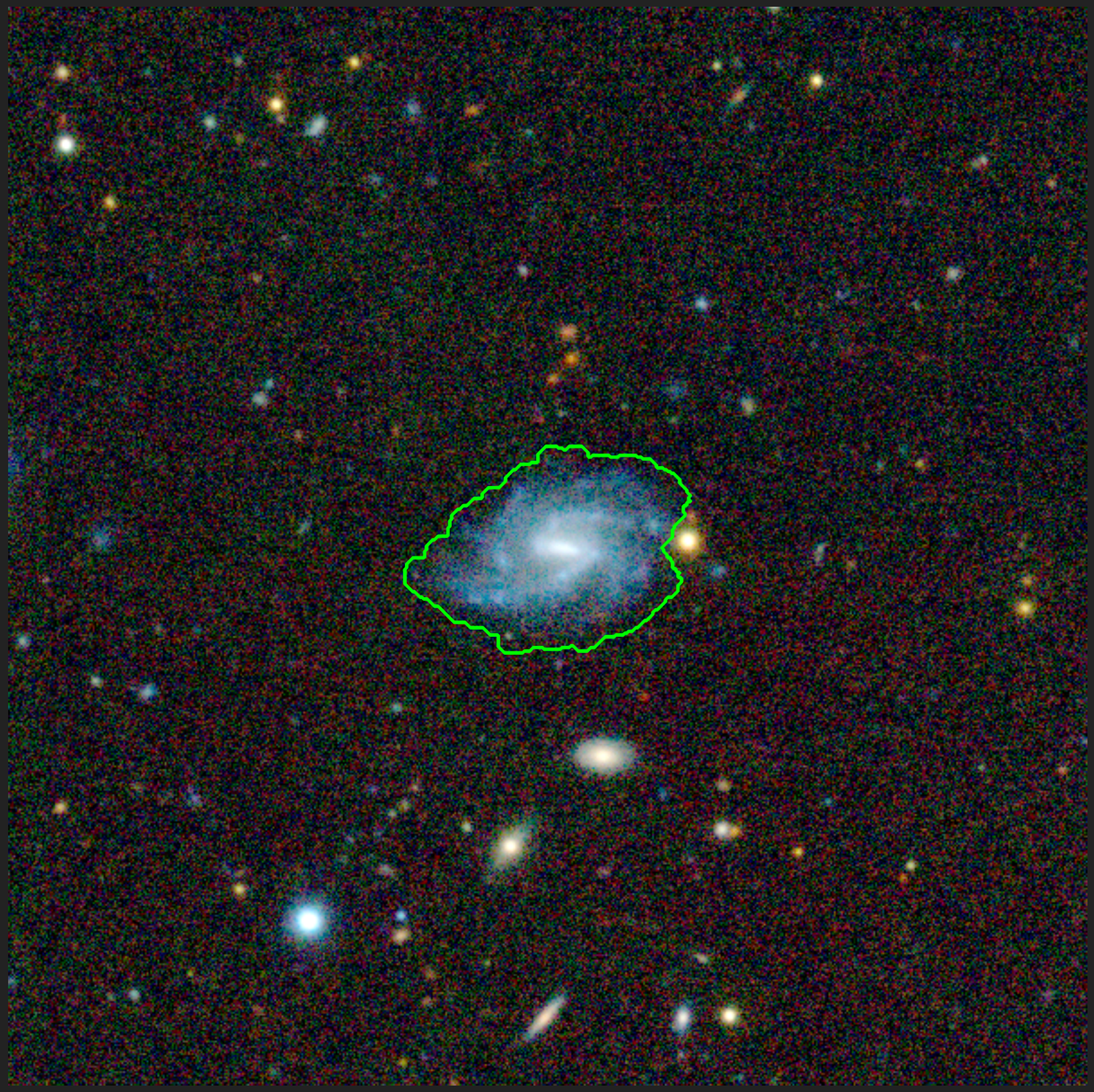
- Starting from just multi-band image date we want to:
  - Detect object
  - Estimate sky
  - Find PSFs
  - Fit bulge + disk model with distinct SED for each component

PROFUSE IN ONE SCHEMATIC

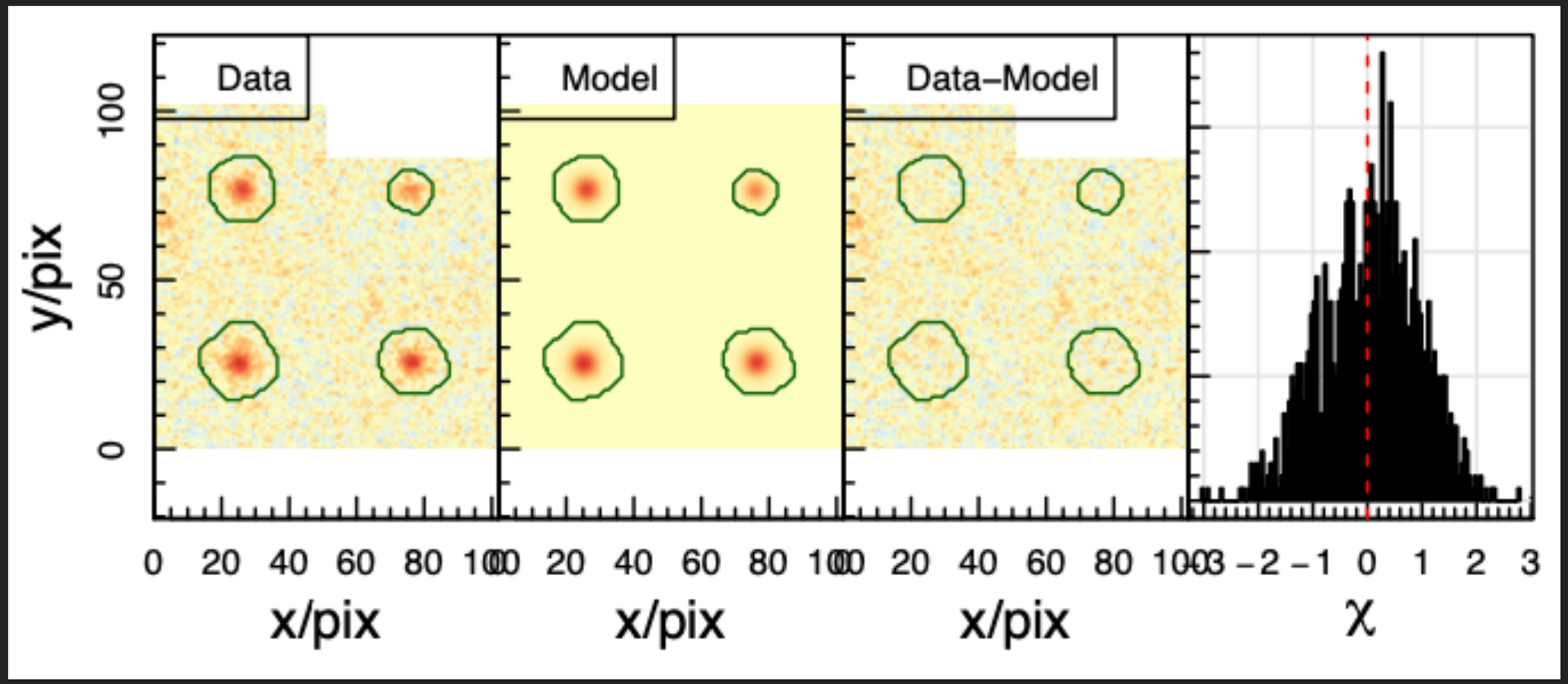


PROFUSE IN ANGER - IMAGE SEGMENTATION AND STAR DETECTION (PROFOUND)

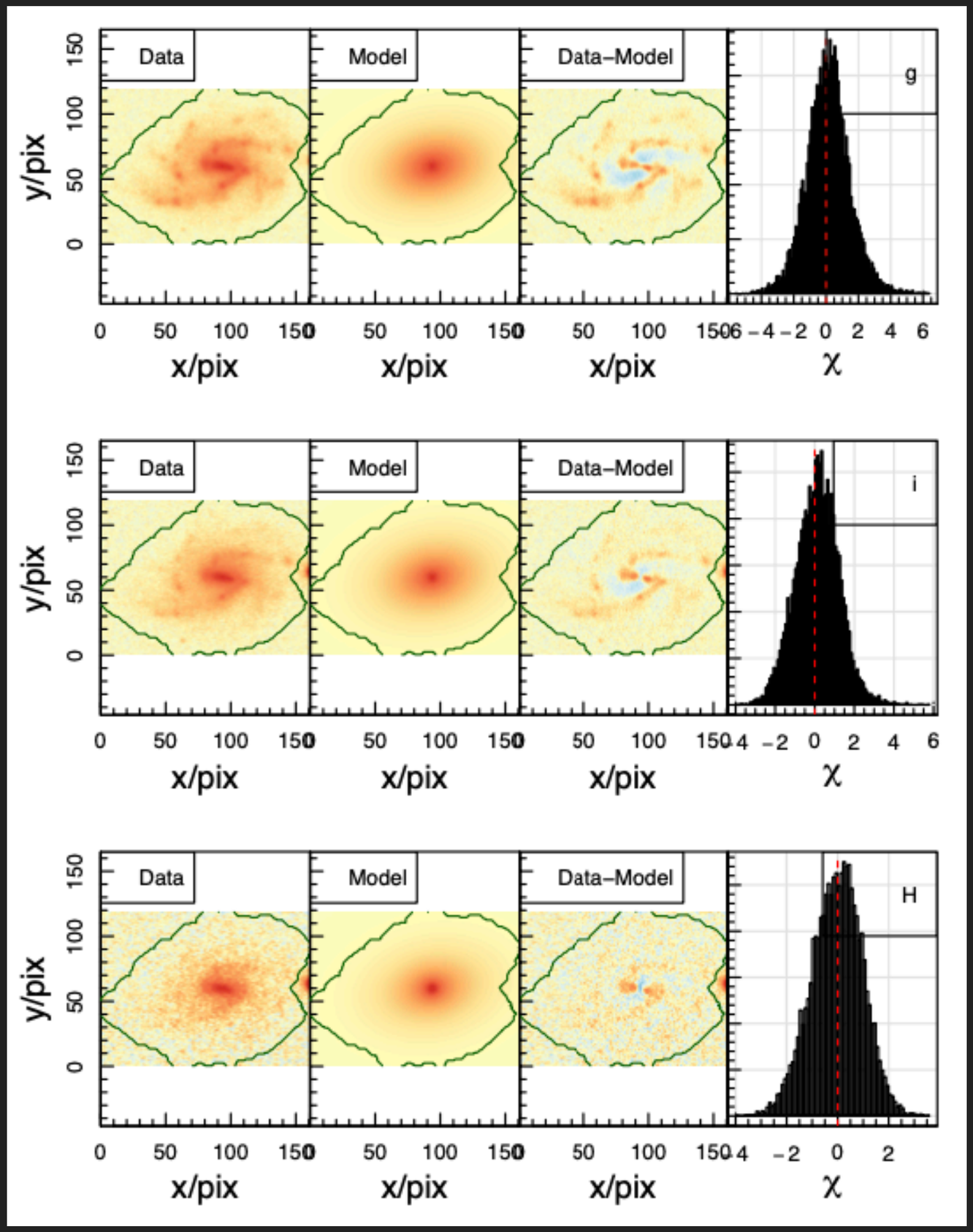
- Objected are roughly detected in all bands, and sky / sky-RMS estimated.
- An inverse variance stack is created, and a global detection is carried out (deep).
- Good candidate stars are detected per band, and Moffat PSFs made.



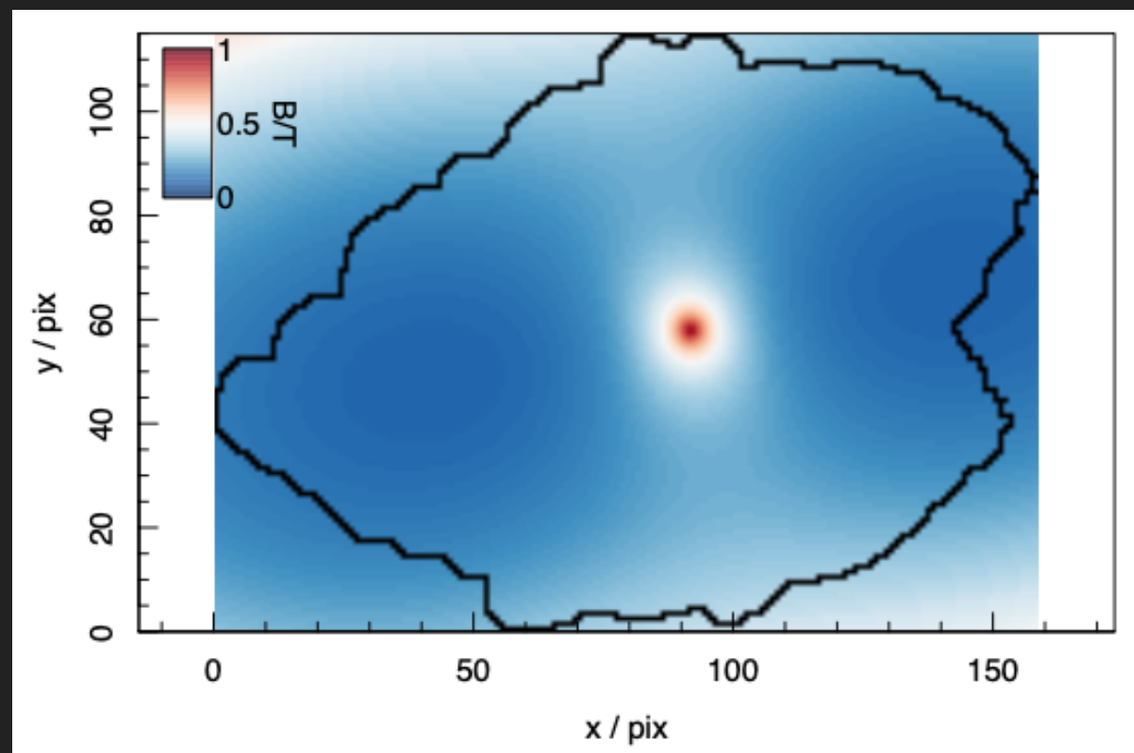
PROFUSE IN ANGER - AUTOMATIC PSF IN EVERY BAND



PROFUSE IN ANGER - SPECTRAL + SPATIAL DECOMPOSITION (PROSPECT + PROFIT)

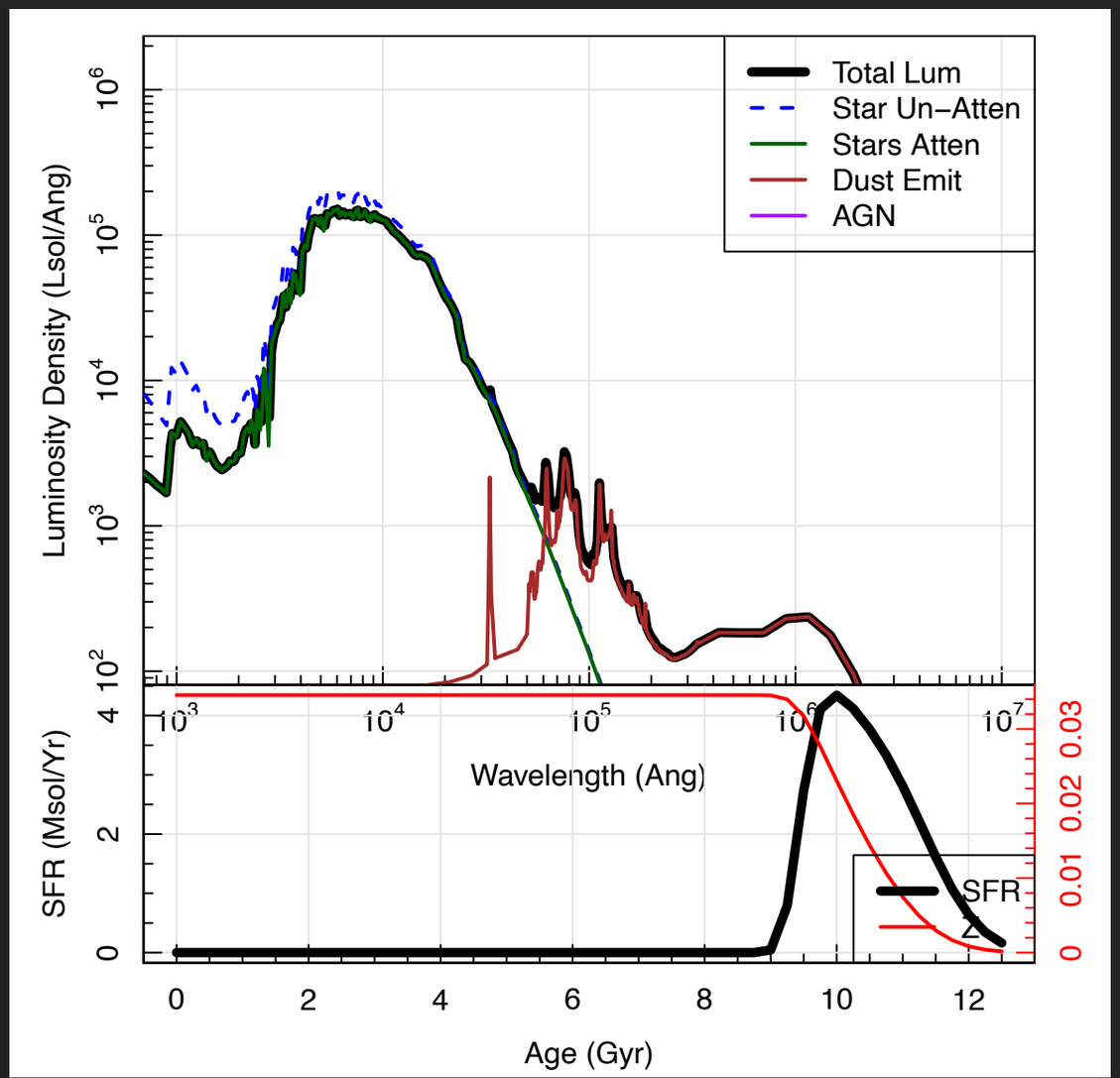


B/T Map

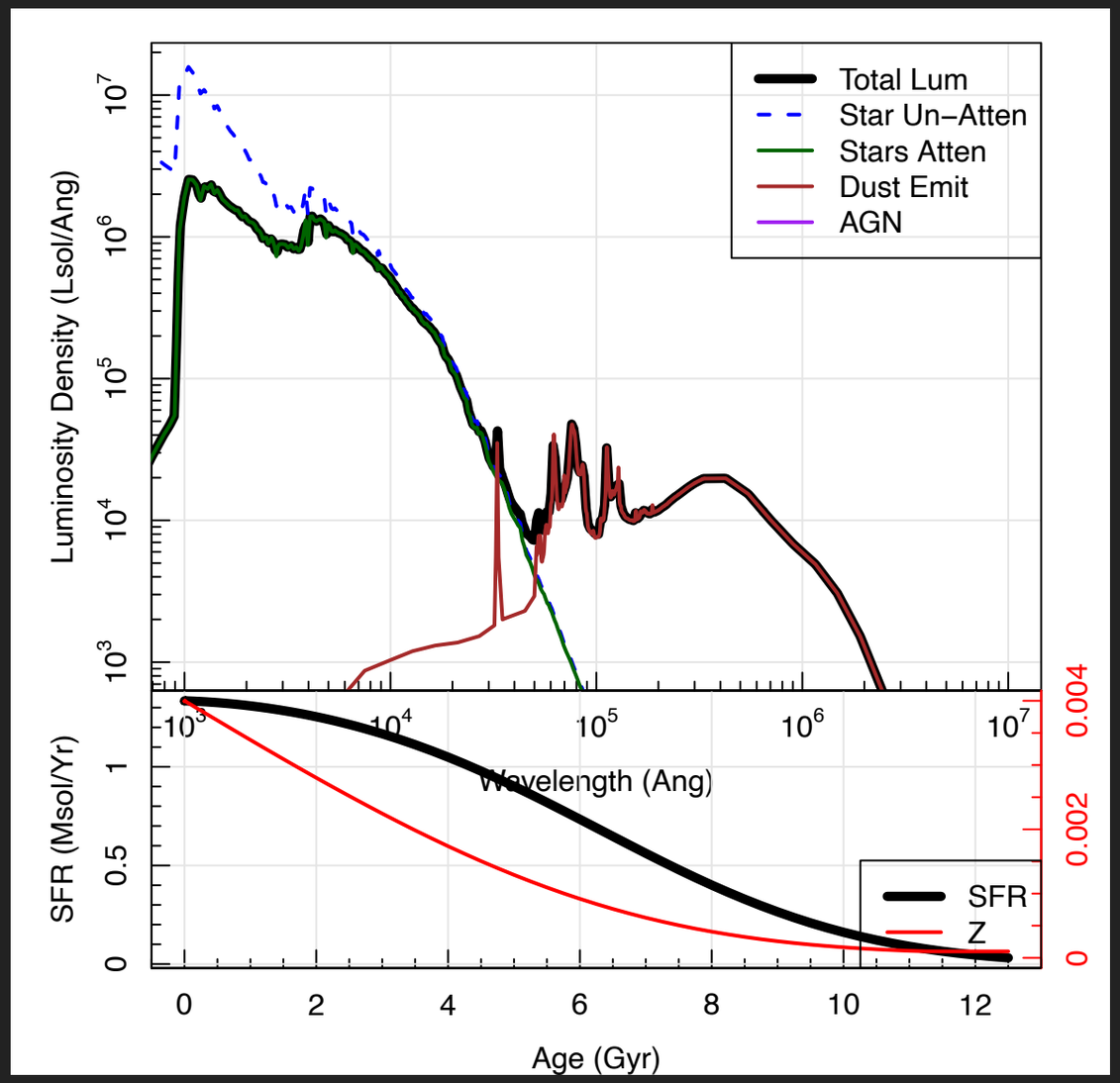


PROFUSE IN ANGER - INVESTING BULGE / DISK SEDS (PROSPECT)

Bulge



Disk

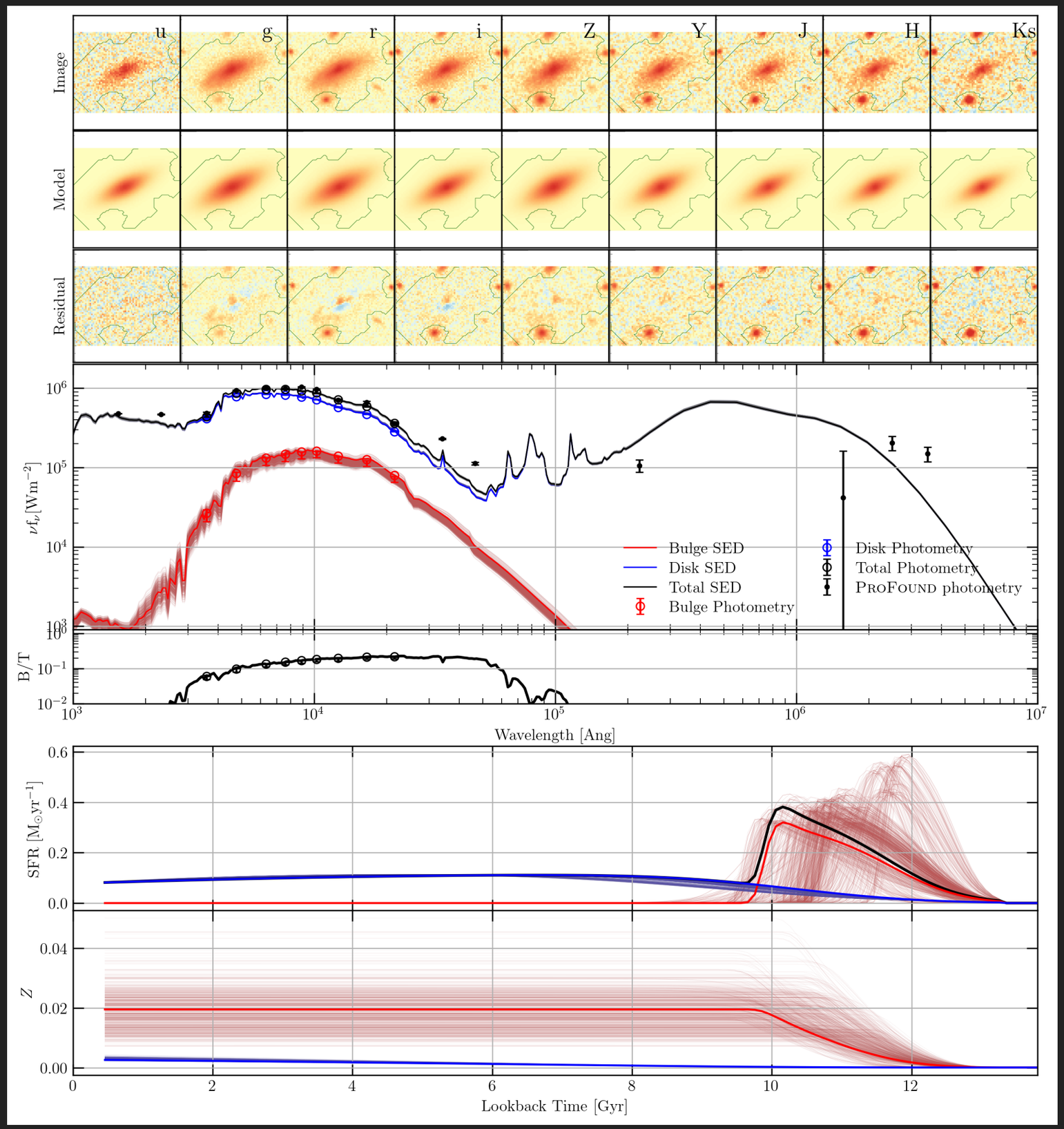


- We get all the usual outputs of ProSpect (SM, SFR, SFH, ZH, and lots more), but per component rather than for the ensemble galaxy.
- This allows us to generate multi-epoch stellar populations that are not possible in ProSpect with a single skewed Normal SFH.

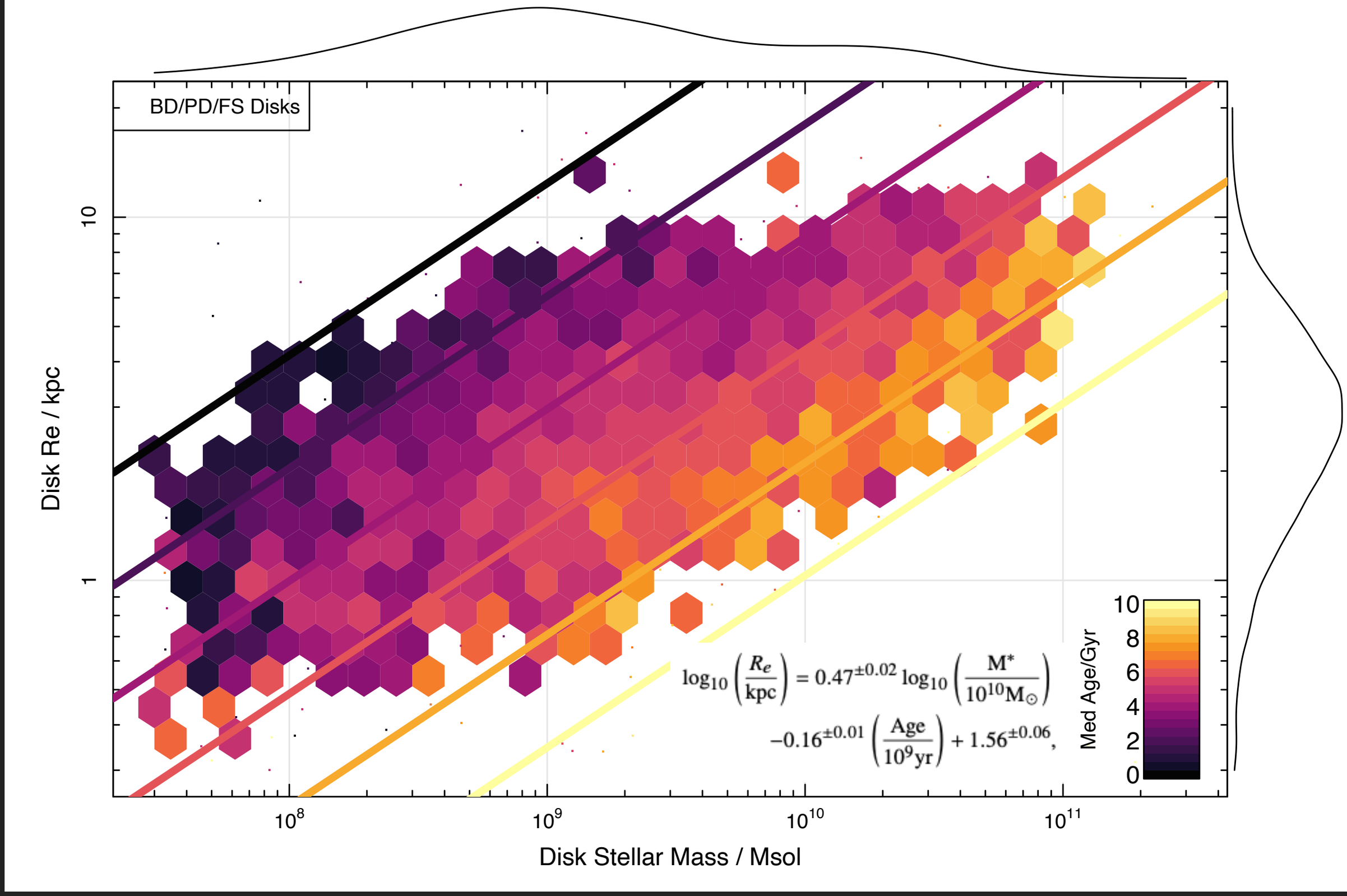
WITH GREAT POWER…

- In this first application we restrict ourselves to relatively simple models:
  - Free Sersic [FS] - a single profile that better suits single component systems with a dominant SFH/ZH (elliptical).
  - Bulge (n=4) + Disk (n=1) [BD] - a bulge-disk fit that allows for an extended bulge.
  - Bulge (PSF) + Disk (n=1) [PD] - a bulge-disk fit that allows for a PSF (unresolved) bulge.
  - Bulge (n=4) + Disk (n=1) + Disk (n=1) [BDD] - a bulge-double-disk fit that allows for an extended bulge and two disk components. These interact to create SFH/ZH and colour gradients.

PROFUSE ON GAMA (~7K GALAXIES, Z<0.06) - ROUGHLY 2-4 HOURS PER FIT



PROFUSE ON GAMA (~7K GALAXIES, Z<0.06) - THE MASS-SIZE-AGE PLANE



PROFUSE LIVES!

- ProFuse combines all of the ProTools developed to date to provide a fully automated Bayesian Inference engine for the spectral-spatial decomposition of galaxies:
  - ProFound detects sources, segments the galaxy of interest and identifies stars.
  - Stars are then modelled with ProFuse in single-band mode to provide per-band PSFs.
  - ProSpect and ProFit are then run in tandem to produce SFHs and ZHs of bulges and disks simultaneously.
  - In principle arbitrarily complex models are allowed, but with the first application with GAMA we only push as far as a Bulge + Disk +Disk (BDD) model. Mostly FS and BD models are preferred.

PROFUSE EXPERIENCE

- Working to build a code as complicated as ProFuse has taken a significant fraction of my time over the last 7 years.
- Whilst it has supported many students (~10) and post docs (~2) along the way, the core task itself was too major and longterm to be supported via traditional academic means, i.e.:
  - PhDs (still only 3 years in Australia)
  - Post docs (usually only 3 years)
  - Grants (max typically ~4) and my current Future Fellowship position is not to develop any tools- the proposal to develop ProFuse was actually rejected in 2018 (too ‘risky’ and not directly ‘scientific’ enough).