Rubin Observatory

Galaxy photometry in DM

Jim Bosch

Princeton / Rubin Observatory DM

with lots of help

NSF AURA U.S. DEPARTMENT OF ENERGY SLAC NSF NSF's National Optical-Infrared Astronomy Research Laboratory

Outline

Rubin Observatory

- Reintroducing photometry
  - Linear estimators, biases, weights, and Bayes
  - [uncrowded] [unmoving] stars are ~~easy~~[ier]
  - Consistent colors vs. complete colors
  - Reporting uncertainties
- What we do now (and what's wrong with that)
  - CModel & MultiProFit
  - *meas*, *ref*, and *forced* steps
- What we intend to do
  - *Always* measure all bands together before *maybe* measuring them independently.
  - Integrate and run a few external codes

Reintroducing photometry

Photometry must respond *linearly* to true flux

Rubin Observatory

If we take two observations of the same ensemble of objects, we need to get the same results (on average).

We'd *really* like to get a consistent answer if we just moved the same object a bit farther away.

That means the estimator:

- can't depend on PSF size;
- can't depend on the surface brightness limit (no isophotal fluxes!);
- can't depend on Bayesian priors (as those would influence results more when an observation contains less information).

Photometry can't depend on nonlinear estimators

Rubin Observatory

An estimator is always biased in the presence of noise, if it:

- depends on a size or shape measured from the same data;
- depends on a position measured from the same data;
- depends on having detected the object in the same data you're using to measure it.

Photometry is hence impossible.

Point source photometry is only slightly impossible

Rubin Observatory

We just need:

- a position (and centroids can be constrained pretty well in the S/N range we care about);
- a PSF model
  - ...or a large aperture (so isolated objects)
  - ...or a spatially-varying aperture correction (of some kind).

Point source photometry is only slightly impossible

Rubin Observatory

We just need:

- a position (and centroids can be constrained pretty well in the S/N range we care about);
- a PSF model
  - ...or a large aperture (so isolated objects)
  - **and** a spatially-varying aperture correction (of some kind).

All photometry is relative*

Rubin Observatory

- We don't need to count all of the photons.
- We don't need to get the PSF model right** (or even have one).
- We just need to count the same fraction of all the photons for every star (including the ones for which we know the true flux already from some other catalog).

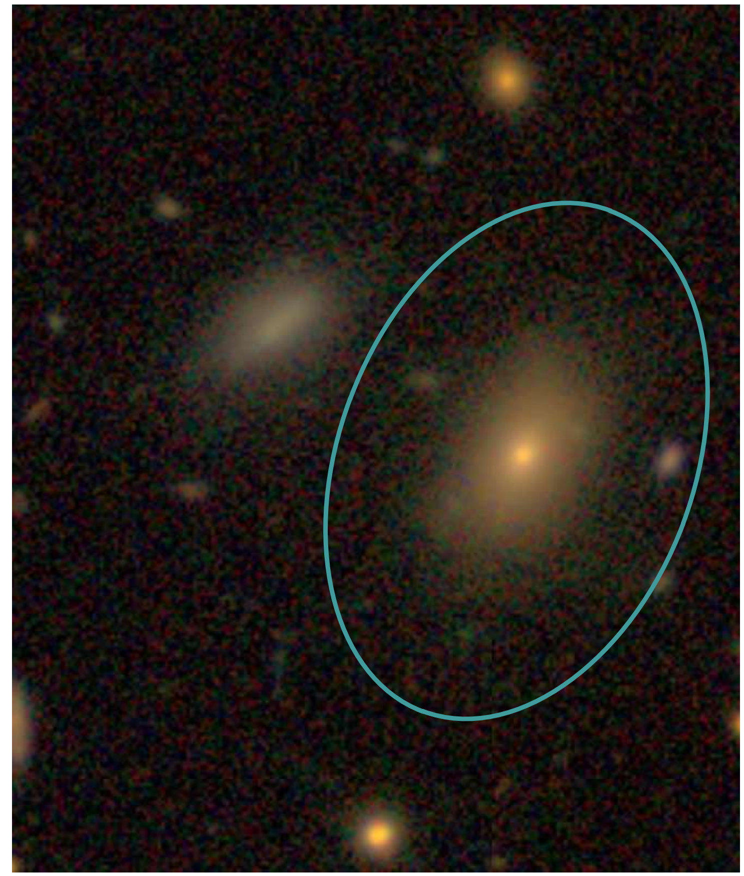
* if your data is from (e.g.) a CCD, not a photomultiplier tube
** if you use the pixel weights carefully (ask me at the end)

Galaxy photometry is definitely impossible

Rubin Observatory

If the galaxy looks like this, perhaps you can:

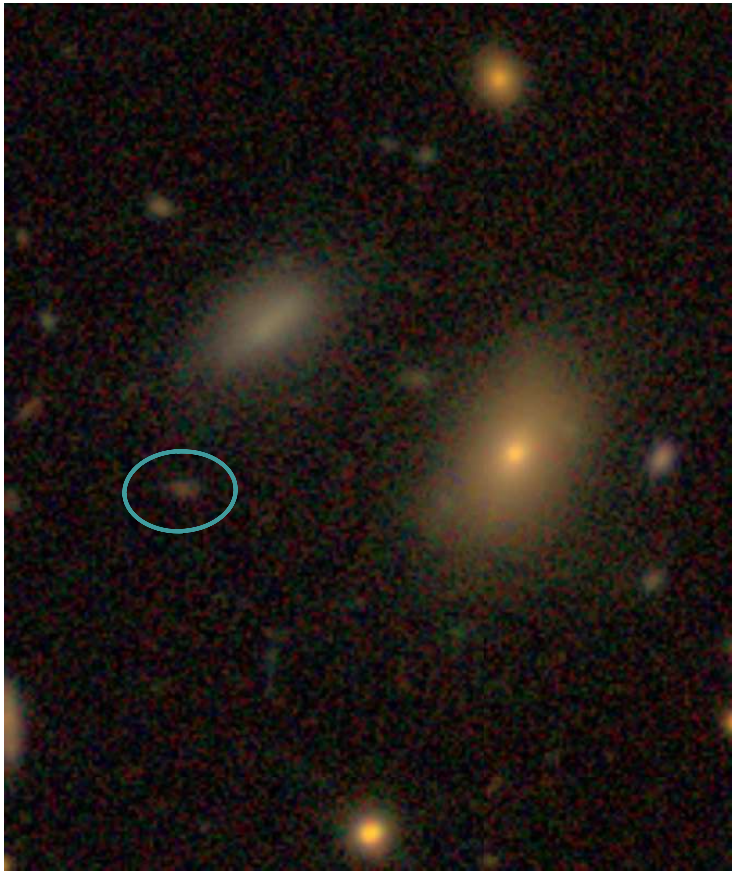
- estimate the size and ellipticity well enough for those to not affect any photometry estimator based on them;
- estimate the PSF-deconvolved profile well enough to define a consistent radius to which we should integrate the galaxy flux.



Galaxy photometry is definitely impossible

If the galaxy looks like this, you can't do any of those things.

But we want to use the same estimator on *all* galaxies in the LSST.



Galaxy photometry in practice

Fit a PSF-convolved galaxy model, with:

- a small number of parameters, to stay robust when there isn't much information in the data (*at most* ellipse + position + flux);
- a large number of parameters, to avoid significant model biases - only *obvious* when there is a lot of information in the data (*at least* ellipse + position + flux + [exponential or de Vaucouleurs])'
- *maybe* priors on structural parameters as needed to bridge the gap.

Make sure this is equivalent to PSF photometry in the r→0 limit!

Galaxy *colors* might be kind of possible

Rubin Observatory

Galaxies are made up of stars, even if we can't resolve them. If we measure colors via an estimator that gives each star the same weight in all bands, that color is a *consistent color*, and it will correspond to a physically reasonable star formation history (and dust!) at the redshift of that galaxy.

We can do that (at possibly significant cost to S/N):

1. Convolve all bands to the same seeing.
2. Measure the object with the same aperture in all bands.

(or something more clever but equivalent)

Forward-fitting != consistent colors

Rubin Observatory

Fitting a PSF-convolved model to all bands only yields consistent colors if:

- the seeing was already the same in all bands (in which case you don't need a PSF-convolved model anyway, at least for colors);
- the model is the same in all bands, aside from a single per-band amplitude;

OR

- the model is flexible enough to actually reflect the true morphology.

What should a flux uncertainty be?

- Total fluxes depend on structural parameters: we should (or we should let you) marginalize over uncertainties on structural parameters.
- Colors should depend on the same structural parameters across all bands, so we should *not* marginalize over uncertainties in structural parameters when computing color uncertainties.

Note that we can't marginalize likelihoods, though - only posteriors.

Everything I've said so far is subdominant.

Rubin Observatory

Good photometry (especially good galaxy photometry requires):

- good background subtraction;
- good (or at least non-catastrophic) deblending.

Even minor problems in these can lead to much bigger problems in galaxy fitting, because the wings of galaxy profiles are basically impossible to distinguish from backgrounds (including flux from neighbors).

A little garbage in → a lot of garbage out.

(And *principled* uncertainty estimates won't tell you that.)

What we have actually done.

What we have actually done: CModel

We've reimplemented the SDSS CModel algorithm.

- Think of CModel as a kludgy, easier-to-fit Sérsic substitute. It means:
  - fit an exponential model;
  - fit a de Vaucouleurs model;
  - fit a linear combination of those models holding their ellipse parameters fixed.
- Our version (unlike the SDSS one) approximates profiles as sums of Gaussians (Hogg & Lang 2012) and PSFs as sums of shapelets for speed (SDSS made other "sleazy" optimizations). That part has worked well.
- Overall, we regard CModel as a placeholder (only) for LSST. We know we will need to replace it with something better.

What we have actually done: CModel

Rubin Observatory

- CModel worked well in SDSS. We have never been happy with its performance for HSC.
- At least some (and perhaps most) of the problems with it are garbage-in, garbage-out: deblending and backgrounds are not in great shape.
- Our CModel implementation seems more sensitive to bad deblending and backgrounds than most of our other measurement algorithms (but that's also natural: it has more degrees of freedom).
- It has seen no major development since 2016.

What we have actually done: MultiProFit

Rubin Observatory

MultiProFit is a fitting code developed by Dan Taranu (now on the DM team).

- It's not (yet) as fast as our CModel implementation or fully integrated into DM pipelines.
- It's *much* more flexible, so we've been using it heavily to explore different models, parameterizations, and ways to combine information across bands.

What we have actually done: multi-band photometry

Rubin Observatory

meas:

Fit a model to the coadd for each band independently.

ref:

Pick the best band for each Object, using the first band in some priority order that meets S/N criteria (an approach taken from SDSS).

forced:

Re-fit models to per-band coadds, holding structural parameters fixed to the values fit in the reference band.

CModel uncertainties

Rubin Observatory

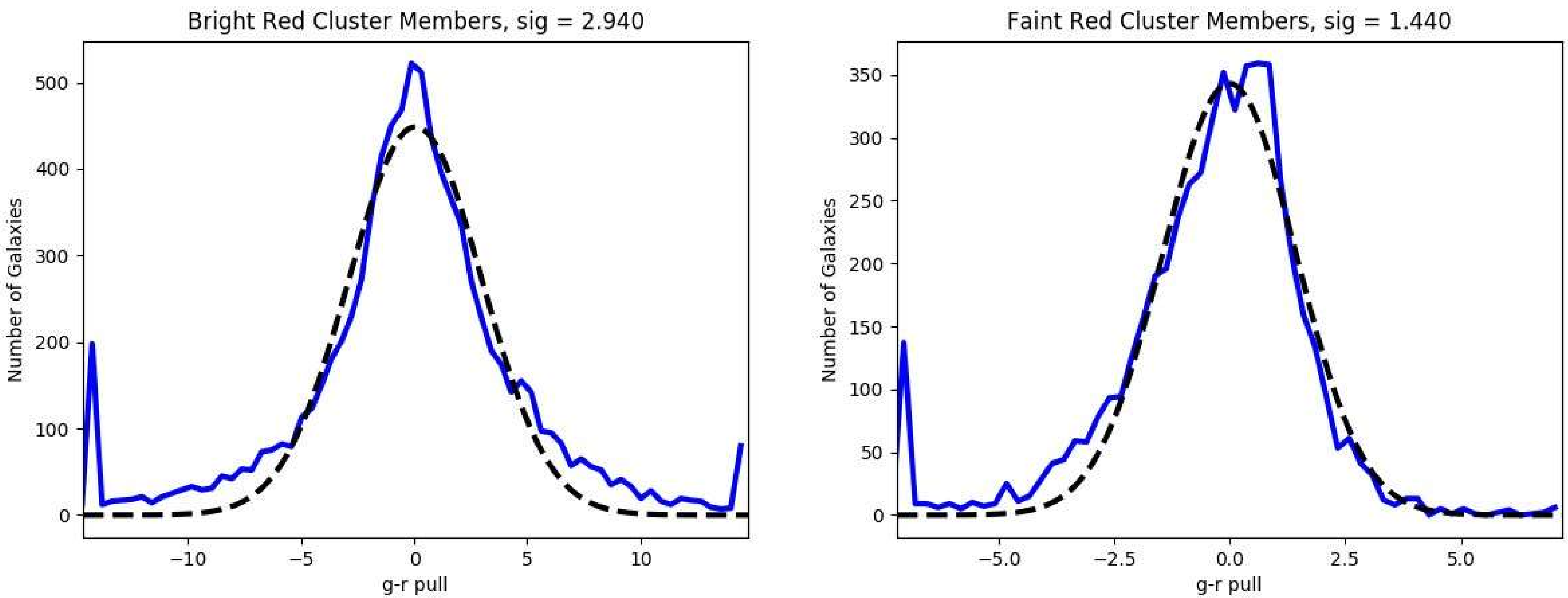


Figure from Eli Rykoff / DESC DC2

CModel uncertainties: what's going on?

- Correlated noise is not propagated through coaddition (probably the biggest problem).
- Uncertainties in structural parameters are not propagated to flux uncertainties (this should be correct for *colors*).
- Measured colors could be internally consistent while still being inconsistent with truth catalog fluxes (unlikely, I'd say).
- Deblending and background estimation are per-band (not forced), and their errors are not propagated into flux uncertainties.

What we intend to do

What we plan to do

Rubin Observatory

Multi-band processing

- Drop the meas/ref/forced sequence in favor of always using all bands simultaneously to constrain structural parameters.
- In some cases, we may *also* fit models to each band independently (for reasons other than colors).

MultiProFit

- We'll continue to use this for development research work.
- We'll consider investing more time in optimizing it and integrating it for production use after evaluating third-party options (next slide).

What we plan to do

Rubin Observatory

Third-party[-ish] codes/algorithms

- We already have a lot of the lower-level building blocks (e.g. PSF matching) for our own `GAaP` implementation, and we intend to write one.
- We usually talk about `Scarlet` as a deblender, but it's designed to generate consistent colors and robust total fluxes.
- `ngmix` produces photometry as well as shear estimates, and has been used to fit models similar to the kinds our `MultiProFit` experiments are pushing us towards. We intend to more fully integrate it anyway, and may just be able to use it for galaxy photometry as well.

Where does this leave us?

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source photometry in $r \to 0$ limit
3. Robust consistent colors
4. Robust total fluxes (and hence total colors)
5. Useful uncertainties

Where does this leave us?

Rubin Observatory

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source photometry in $r \rightarrow 0$ limit
3. Robust consistent colors
4. Robust total fluxes (and hence total colors)
5. Useful uncertainties

Nope.

Where does this leave us?

Rubin Observatory

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source photometry in $r \to 0$ limit
3. Robust consistent colors
4. Robust total fluxes (and hence total colors)
5. Useful uncertainties

This looks pretty bad, but it's hard to know what part(s) of the pipeline to blame.

Where does this leave us?

Rubin Observatory

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source photometry in $r \rightarrow 0$ limit
3. Robust consistent colors
4. Robust total fluxes (and hence
5. Useful uncertainties

We don't quite have the right metrics for this yet (probably need to try to compute photo-zs), so we can only guess at how we're doing.

Also: are consistent colors actually useful in practice?

Where does this leave us?

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source photometry in $r \rightarrow 0$ limit
3. Robust consistent colors
4. Robust total fluxes (and hence total colors)
5. Useful uncertainties

This looks pretty good, and always has.

Where does this leave us?

Our primary goals for galaxy photometry are:

1. Linearity
2. Equivalent to point-source pho
3. Robust consistent colors
4. Robust total fluxes (and hence
5. Useful uncertainties

Interactions between bad background subtraction and aperture correction are currently causing problems.

Photometry algorithms themselves *may* be ok.

Next up: quantifying where we are now.

Extra slides

Optimizing Signal-to-Noise

Rubin Observatory

If we have a model of the PSF ϕ, we can define a matched filter flux f as:

$$f = \frac{\sum_i \phi_i z_i}{\sum_j \phi_j^2}$$

if we replace the pixel data z by the PSF scaled by the true flux α, we get:

$$f = \frac{\sum_i \phi_i^2 \alpha}{\sum_j \phi_j^2} = \alpha$$

Maximum Likelihood Fluxes

Rubin Observatory

If we instead think of the flux as the only free parameter in a model of the data,

and each pixel has Gaussian noise σ, the maximum likelihood flux (also the best least-squares fit flux) is very similar:

$$f = \frac{\sum_i \phi_i z_i / \sigma_i^2}{\sum_j \phi_j^2 / \sigma_j^2}$$

once again, if we plug in our model for the data, we recover the true flux:

$$f = \frac{\sum_i \phi_i^2 \alpha / \sigma_i^2}{\sum_j \phi_j^2 / \sigma_j^2} = \alpha$$

Imperfect Models

Rubin Observatory

If our PSF model isn't quite correct, what happens?

For optimal-filter fluxes:

$$f = \frac{\sum_i \phi_i(\phi_i + \epsilon_i)\alpha}{\sum_j \phi_j^2} = \alpha\left(1 + \frac{\sum_i \phi_i \epsilon_i}{\sum_j \phi_j^2}\right)$$

For maximum-likelihood fluxes:

$$f = \frac{\sum_i \phi_i(\phi_i + \epsilon_i)\alpha/\sigma_i^2}{\sum_j \phi_j^2/\sigma_j^2} = \alpha\left(1 + \frac{\sum_i \phi_i \epsilon_i/\sigma_i^2}{\sum_j \phi_j^2/\sigma_j^2}\right)$$