Blending Workshop Breakout Session #5

Tools and Data Sets for Developing and Evaluating Algorithms for Blended Objects

Wednesday, August 15, 1:30 to 3:00pm

Agenda

1. Combining existing space & ground imaging - overview - Harry Ferguson
    - Example: HST/HSC - Will Dawson
2. Catalog-based simulations - two blending-analysis examples here.
3. Pixel-level simulation tool kits -
    - Example 1: Weak Lensing Deblending package - David Kirkby
    - Example 2: Blending Tool Kit - Sowmya Kamath
    - Example 3: Chromatic Real Galaxy - Sowmya Kamath
4. Generative models for simulation - overview & examples - David Kirkby
5. Simulations embedded in real data -
    - Example: Balrog (Dark Energy Survey) - Eric Huff
6. Discussion and planning

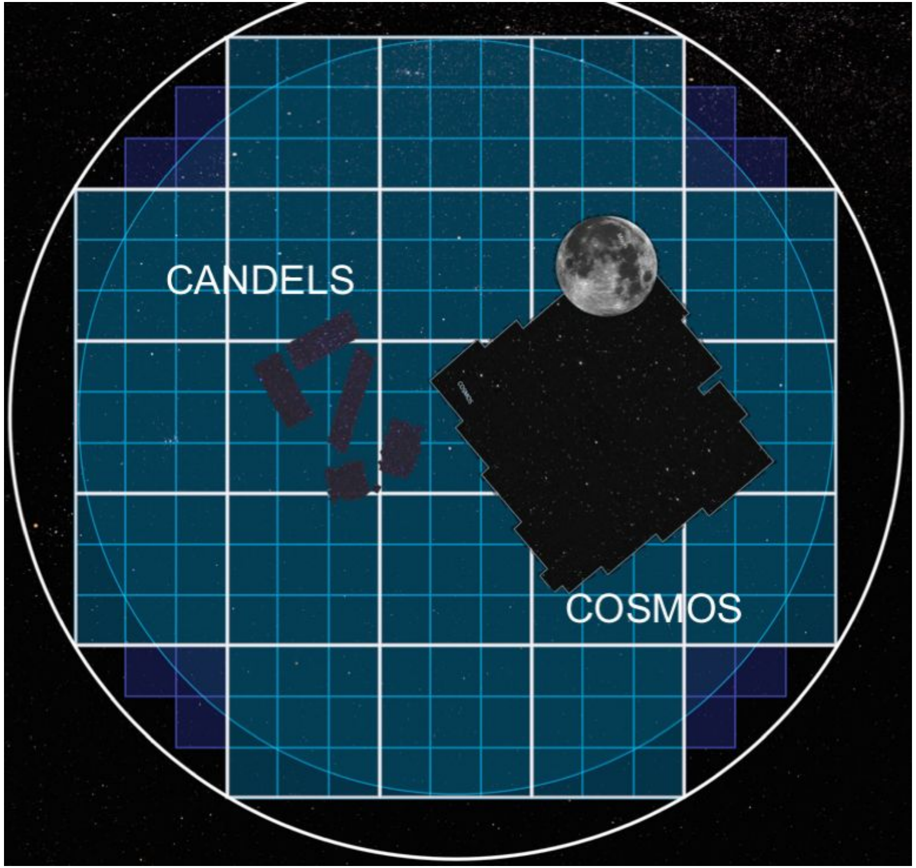
Hubble Datasets useful for blending tests

Harry Ferguson

HST

- ACS I-band=F814W, 0.09” FWHM, 0.05” pixels, 203” x 203”

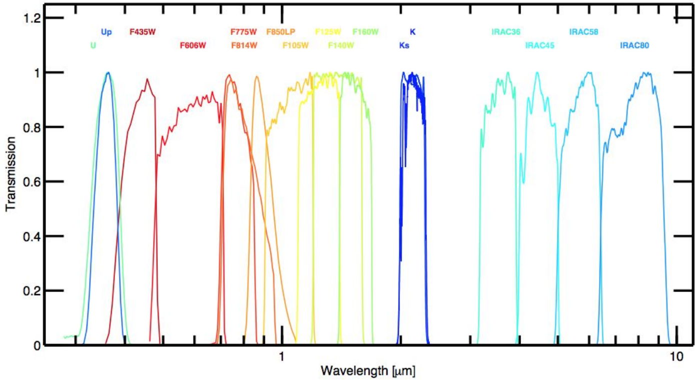
- **CANDELS** candels.ucolick.org 0.2 deg^2, 5 fields
  - Point-source limit ~27
  - H-band selected catalog
    - *0.17” FWHM PSF*
    - *De-blended ground-based and Spitzer photometry using HST positional priors*
  - Deepest fields for multi-wavelength coverage
  - Dense & deep spectroscopic followup
- **COSMOS** http://cosmos.astro.caltech.edu/page/astronomers
  - 1.78 deg^2 centered at (RA,DEC) = (150.2, 2.2).
  - 50% completeness for sources 0.5” in diameter at I(AB) = 26.0 mag.
  - *Position-matched ground-based and Spitzer photometry*

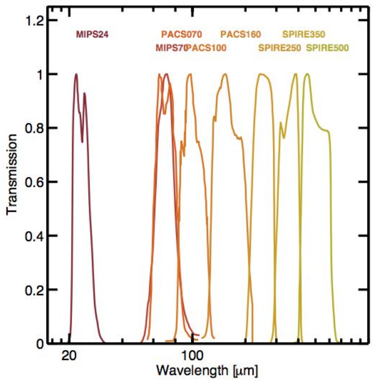


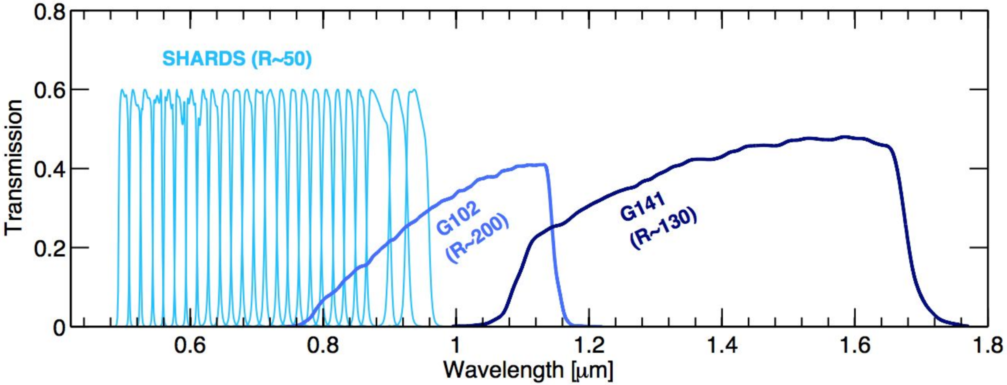
CANDELS

- Catalogs coming soon:
- Photo-z
  - Kodra+18
  - Updated phot-z from 5 codes
    - With PDFs
  - best available spec-z
- GOODS-N photometry
  - Barro+18
  - Photo-z make use of 25-band R=50 data & HST grism data over much of the field

GOODS-N Wavelength coverage





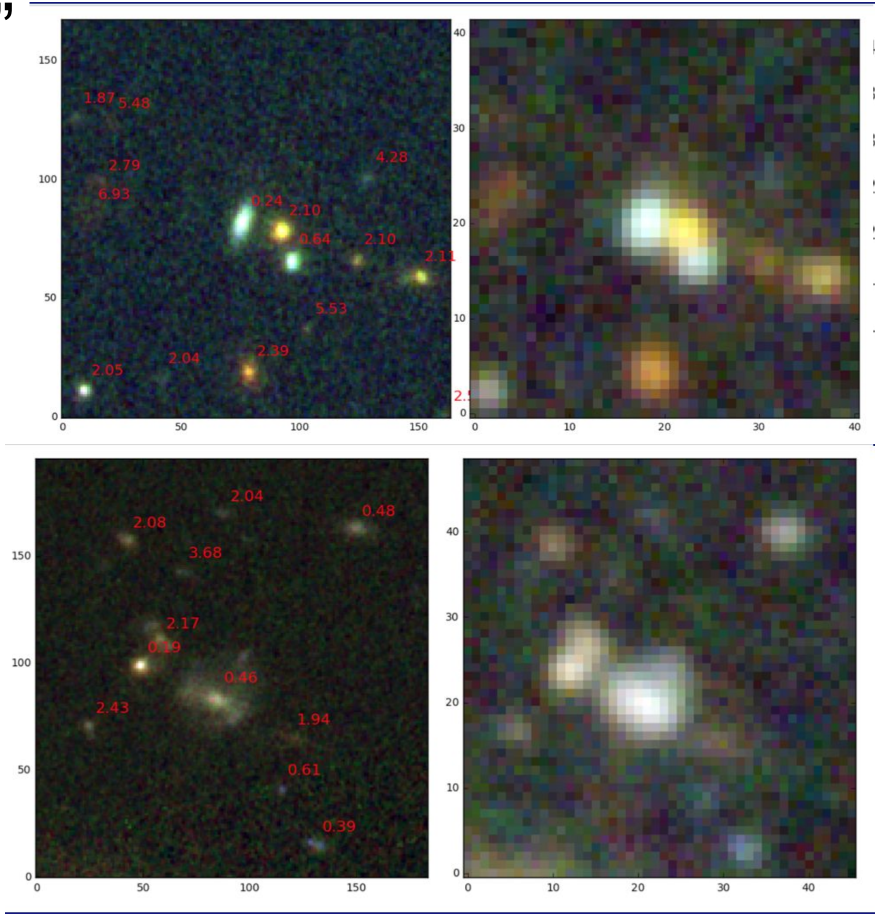


Using HST images as "truth"

- \+ They are real
- \+ Avoids having to use models or make individual-galaxy cutouts
- \+ Best available redshift estimates
- \- Don't really know truth
    - \- Even total magnitudes are uncertain
- \- <200 sq. arcminute much deeper than LSST
- \- CANDELS catalog is not based on the highest-resolution data
    - \- Could reprocess using Scarlet starting with ACS 0.9" FWHM images

HST

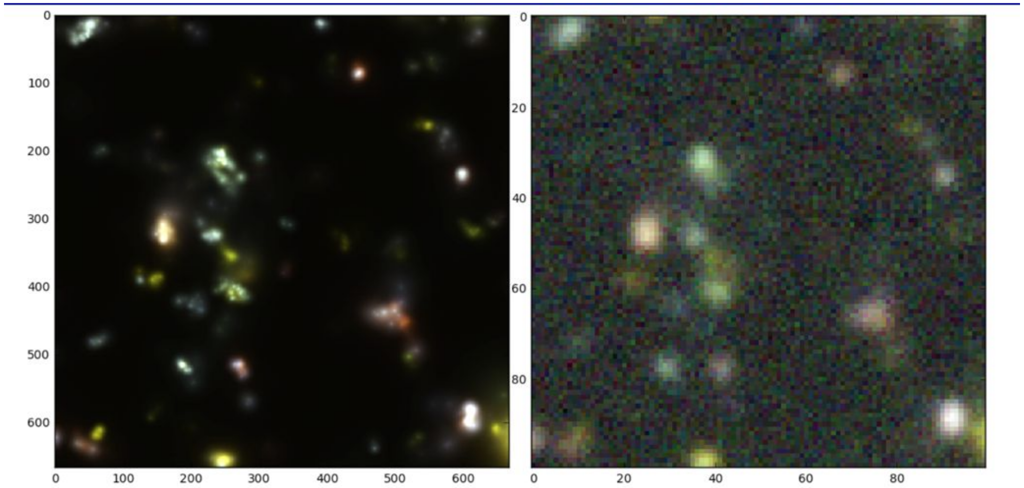
Simulated LSST



Using HST-like simulations as “truth”

+ True redshifts are known even for 100% overlap
+ Starts with a noiseless image
+ Easier to simulate LSST bandpasses
- Still some subtleties in estimating true total magnitudes
- Morphologies & SEDs not perfect
- So far, tiny areas (10’s of sq. arcmin)

Snyder+ Illustris mock images



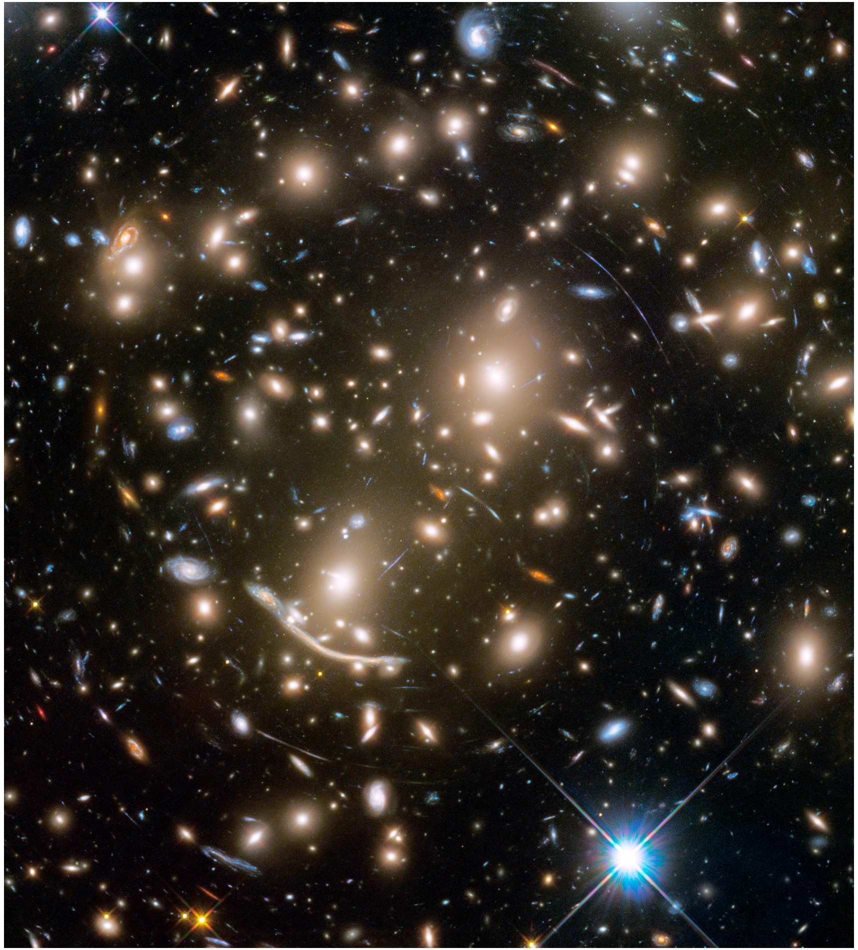
https://github.com/gsnyder206/mock-surveys

Stress test: HST Frontier Fields

6 clusters & parallel fields (~60 sq. arcmin total)

- \+ Deepest cluster fields
- \+ Extensive multi-wavelength data & spectroscopy
- \+ Extensively tested lens models
- \+ Very challenging de-blending problem even at HST resolution
- \+ Immediate science from improving de-blending of these images
- \- Don’t really know truth
- \- Less multiwavelength data than CANDELS

Filter	Orbits	AB_mag	Filter	Orbits	AB_mag
F435W	18	28.8	F105W	24	28.9
F606W	10	28.8	F125W	12	28.6
F814W	42	29.1	F140W	10	28.6
			F160W	24	28.7



Space + ground data: HST + Suprime-Cam (SC)

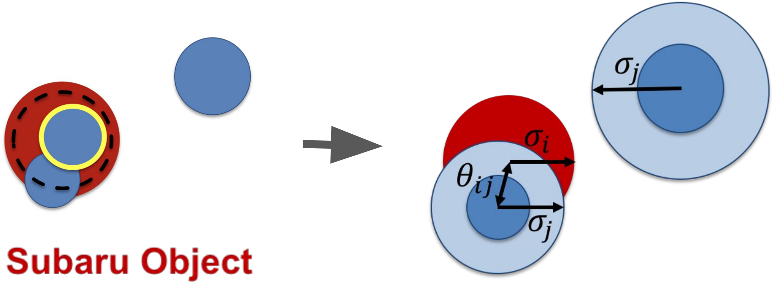
Will Dawson

Example: The Ellipticity Distribution of Ambiguously Blended Objects

Dawson, Schneider, Tyson & Jee (2016) http://adsabs.harvard.edu/abs/2016ApJ...816...11D

- Goal:
  - Layout the fundamentals of ambiguous blending
  - Quantify the scale of the ambiguous blending problem
  - Estimate its impact on cosmic shear measures
- Method:
  - Use overlapping Subaru Suprime-Cam imaging (to LSST depth) and Hubble Space Telescope imaging
  -

$$\theta_{\mathrm{eff}_{ij}} = \frac{\theta_{ij}}{\Xi_\sigma(\sigma_i + \sigma_j)}$$



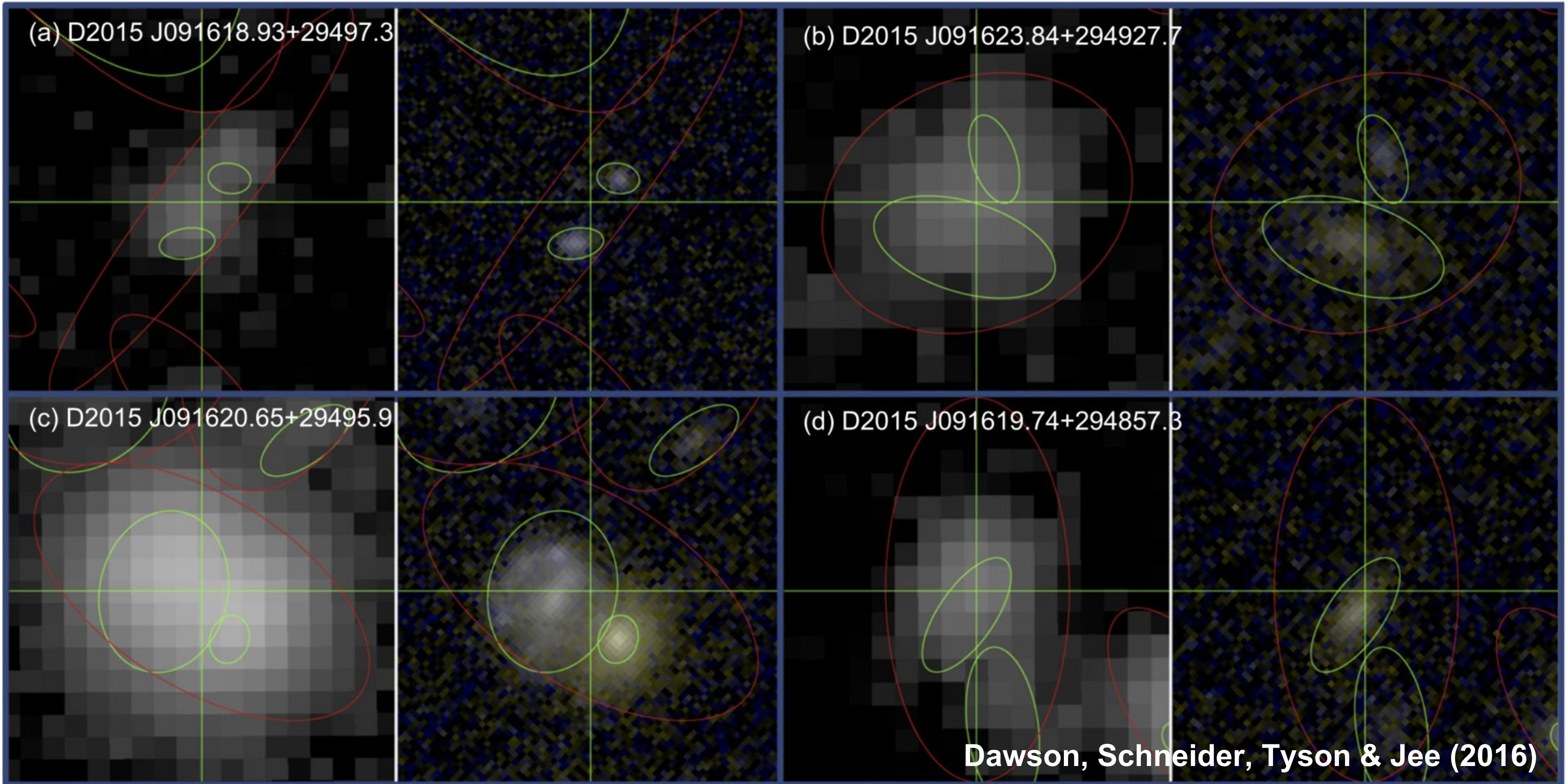
Subaru Object
HST Objects

σ_j HST size after convolution with Subaru PSF

Ξ_σ Normalizing scale factor (1 reasonable choice)

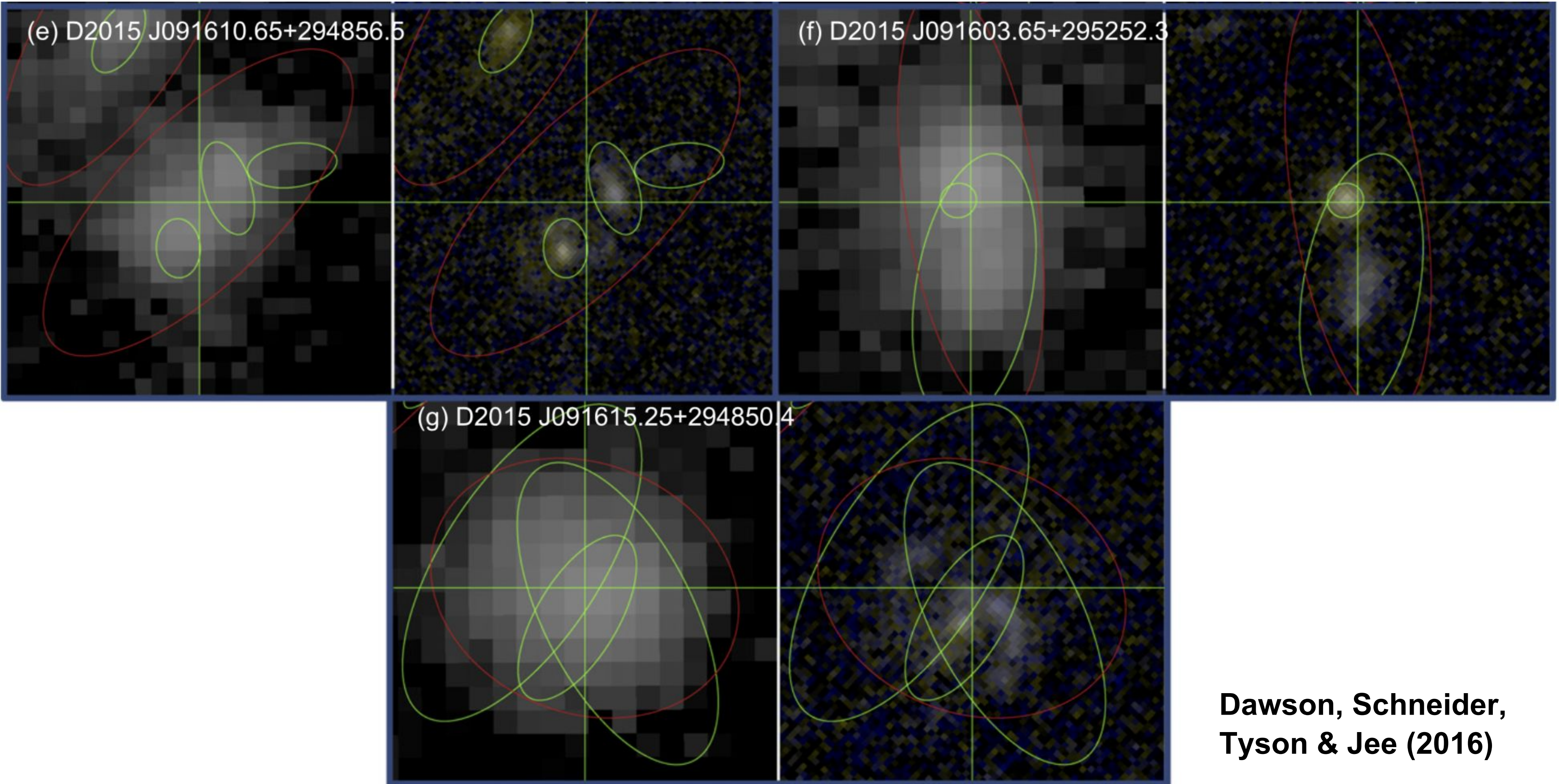
$\theta_{\mathrm{eff}_{ij}} < 1$ Flagged as ambiguous blend candidate

Subaru (left) and HST (right) views of ambiguous blends



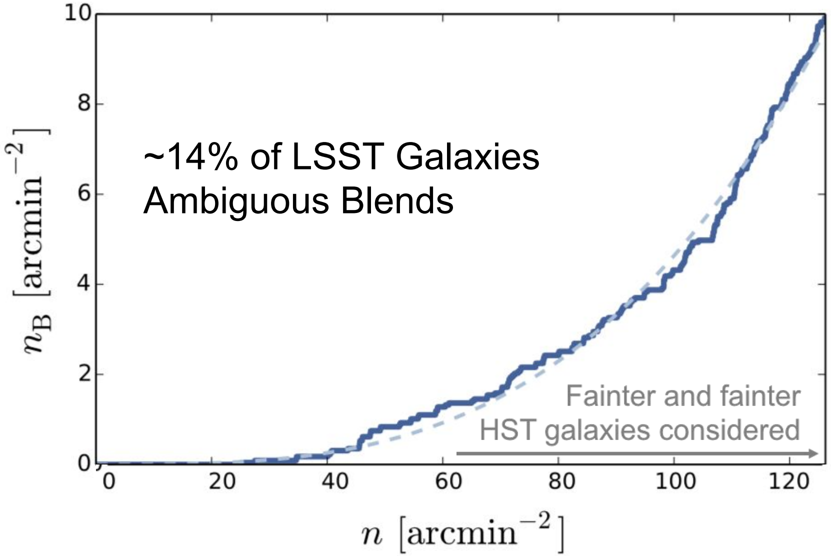
Dawson, Schneider, Tyson & Jee (2016)

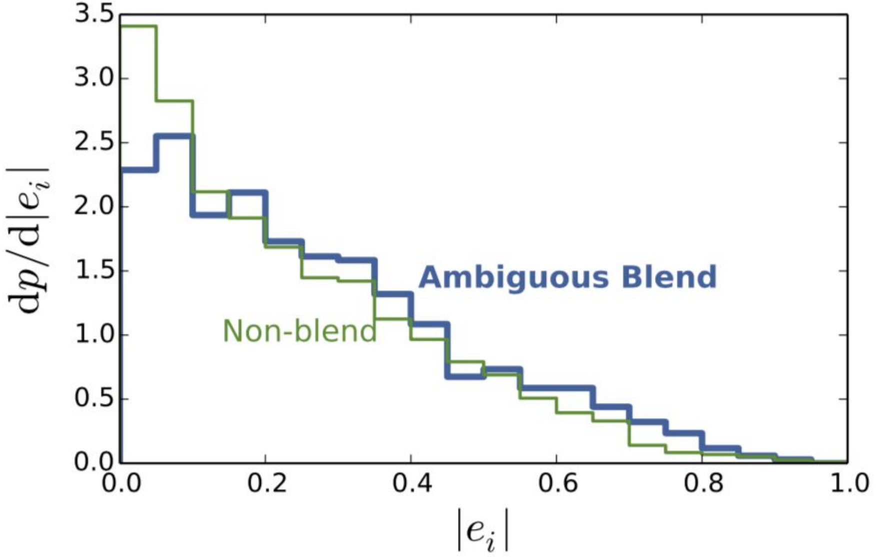
Subaru (left) and HST (right) views of ambiguous blends



Dawson, Schneider, Tyson & Jee (2016)

The number density of ambiguous blends grows rapidly with depth, and they have significantly different properties





$$n_B \approx 2.3 n^{3.15} \times 10^{-6}$$

Dawson, Schneider, Tyson & Jee (2016)

Catalog-level simulations: Two examples

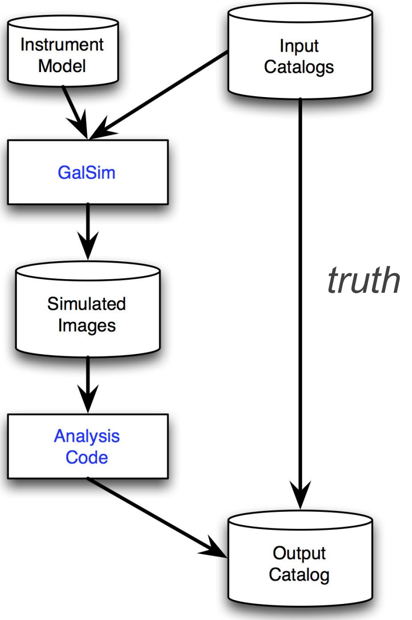
These independent studies each estimate a specific blending impact on joint galaxy-galaxy, galaxy-shear and shear-shear correlations (3x2-pt correlations).

1. "Cosmological Simulations for Combined-Probe Analyses: Covariance and Neighbour-Exclusion Bias", J. Harnois-Deraps et al. arXiv:1805.04511
   - Uses Scinet Light Cone Simulations (SLICS) catalog.
   - Assumes either the faintest or both members of pairs of objects separated by less than a specified angle are *excluded* from the sample.
   - From the abstract: "For surveys like KiDS and DES, where the rejection of the neighbouring galaxies occurs within ~2 arcseconds, we show that the measured cosmic shear signal will be biased low, but by less than a percent on the angular scales that are typically used in cosmic shear analyses. The amplitude of the neighbour-exclusion bias doubles in deeper, LSST-like data."
2. See presentation by Erfan Nourbakhsh at Blending Session #4 on study of impact of *unrecognized* blends.
   - Uses Buzzard catalog.
   - Assumes a fraction of pairs of objects separated by less than a specified angle are interpreted as a single object, impacting the measured position and shape.

Pixel-level simulation example 1: WeakLensingDeblending

David Kirkby

Developed within the DESC to study blending impacts.



Galaxies, AGNs, stars…

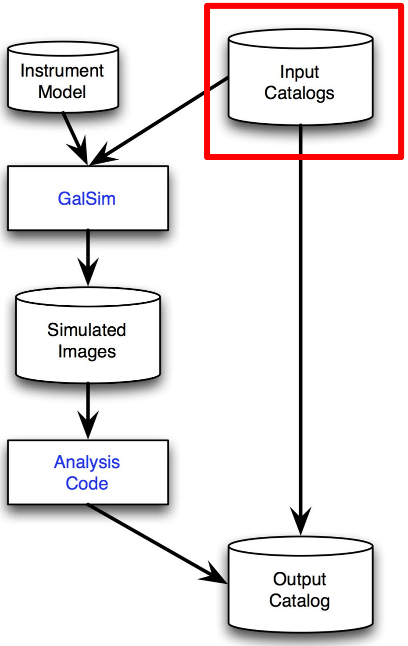
Object properties, blending metrics, ...

readthedocs

tutorial

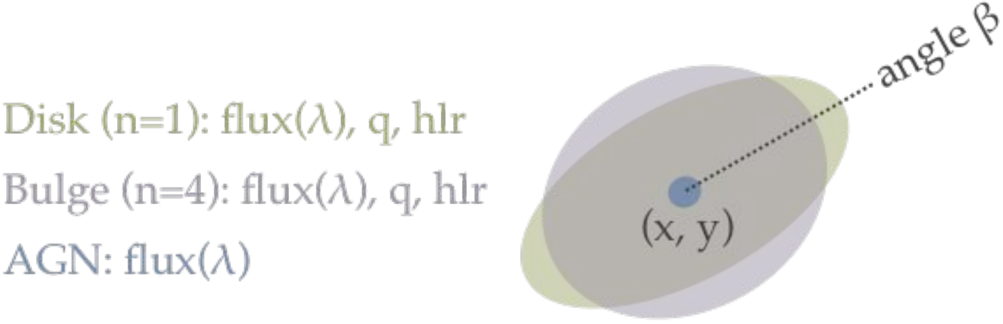
github

Pixel-level simulation example 1: WeakLensingDeblending

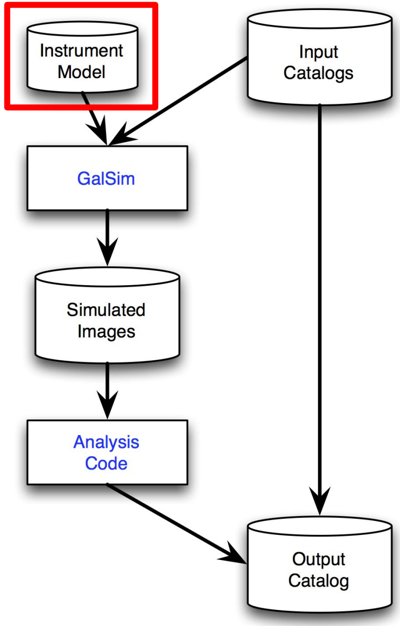


Default galaxy catalog from LSST CatSim:

- complete to r~28
- easy to interface to other catalogs (docs)
- galaxies described by 10 params:



Pixel-level simulation example 1: WeakLensingDeblending



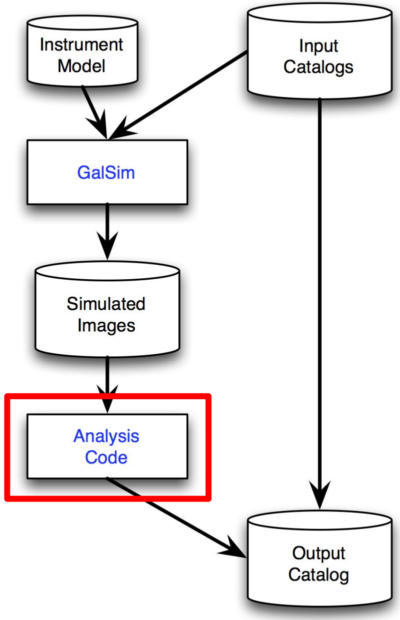
Use simple instrument model to capture main scaling relations between surveys:

- camera: pixel size, zero point, exposure time.
- site: seeing, sky level, extinction.

Survey	Effective Area A (m^2)	Primary Diameter D (m)	Pixel Size Δ_{pix}		Exp. Time (s)	Sky Brightness mag/arcsec^2	Atmos. FWHM κ_0	Zero Point s_0
LSST	32.400	8.360	0.200”	i	5520	20.5	0.75”	32.4
				r	5520	21.2	0.78”	43.7
HSC	52.810	8.200	0.170”	i	1200	19.7	0.56”	69.8
				r	600	20.6	0.67”	87.7
DES	10.014	3.934	0.263”	i	1000	20.5	0.79”	13.9
				r	800	21.4	0.79”	15.7

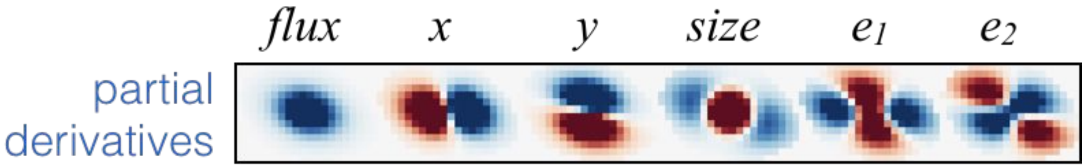
=184 visits x 30s

Pixel-level simulation example 1: WeakLensingDeblending

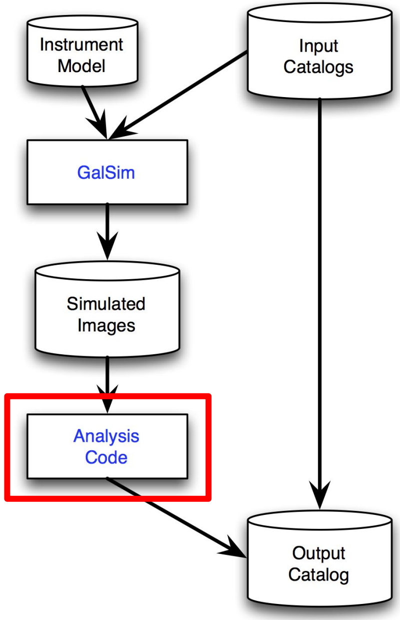


Overall philosophy: quantify impacts of blending without using specific pipeline algorithms -

- identify overlapping source groups
- estimate params (SNR, size, …) w/ and w/o blending
- estimate correlated statistical errors and noise bias on size & shape using pixel-level Fisher matrix formalism



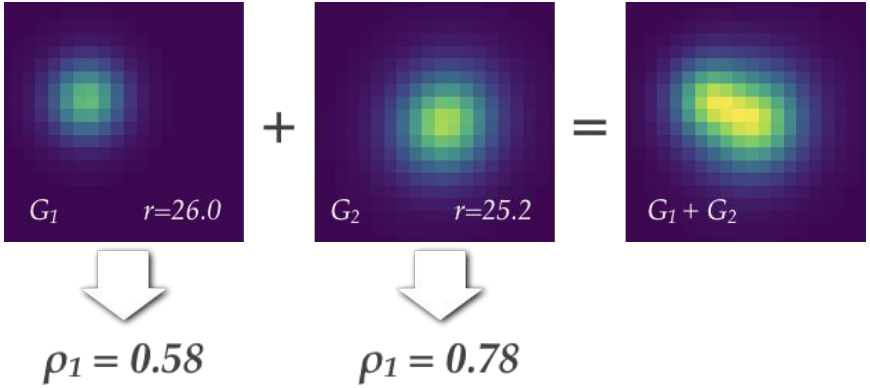
Pixel-level simulation example 1: WeakLensingDeblending



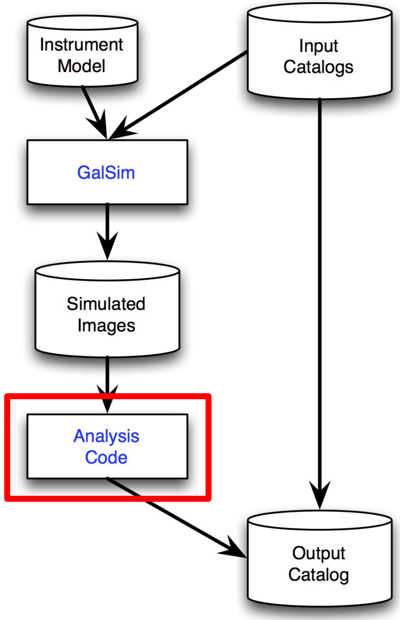
Blending metric example:

"purity" = ratio of weighted pixel sums

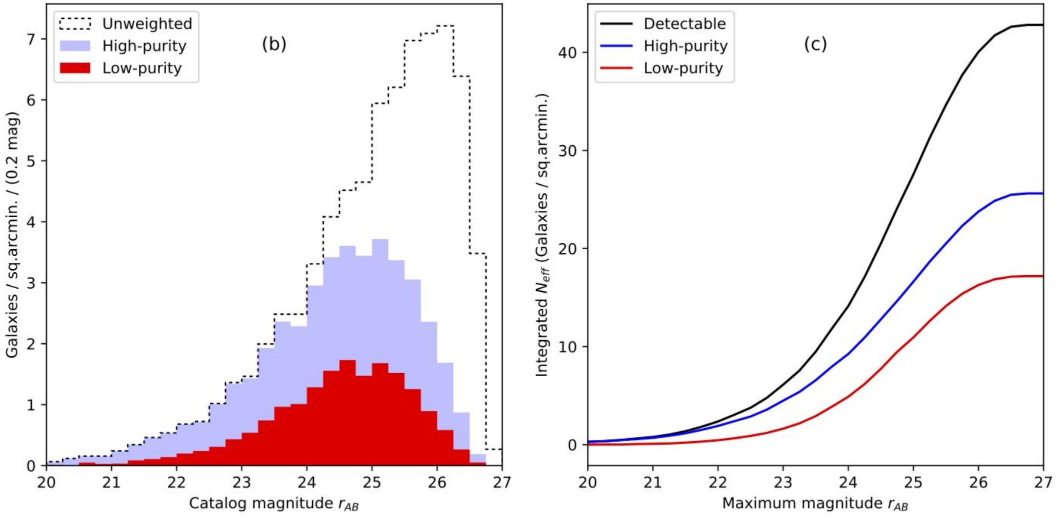
$$\rho_k = \frac{\sum_{\text{pix}} G_k \, G_k}{\sum_{\text{pix}} G_k \, (G_1 + G_2 + \ldots)}$$



Pixel-level simulation example 1: WeakLensingDeblending

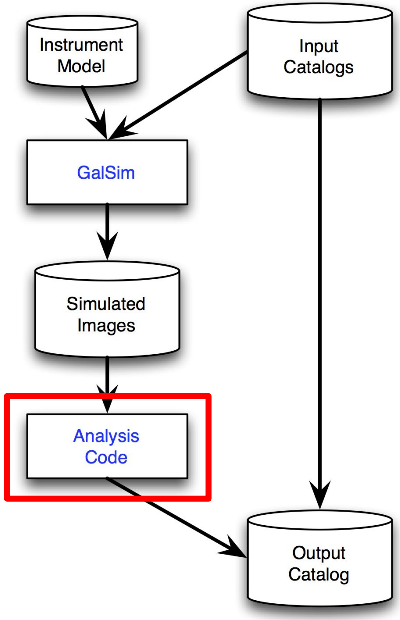


Results example: where is the statistical power for weak-lensing shape measurements?



“Detectable” => $SNR_{grp,float} > 6$

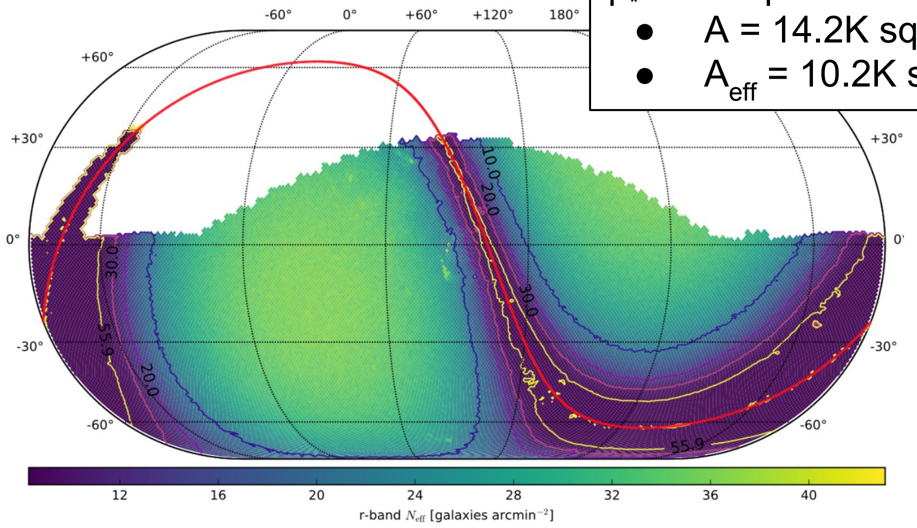
Pixel-level simulation example 1: WeakLensingDeblending



Results example: what is the impact of star-galaxy blending?

$\rho_* < 10$/sq.arcmin:

- A = 14.2K sq.deg
- A_{eff} = 10.2K sq.deg.

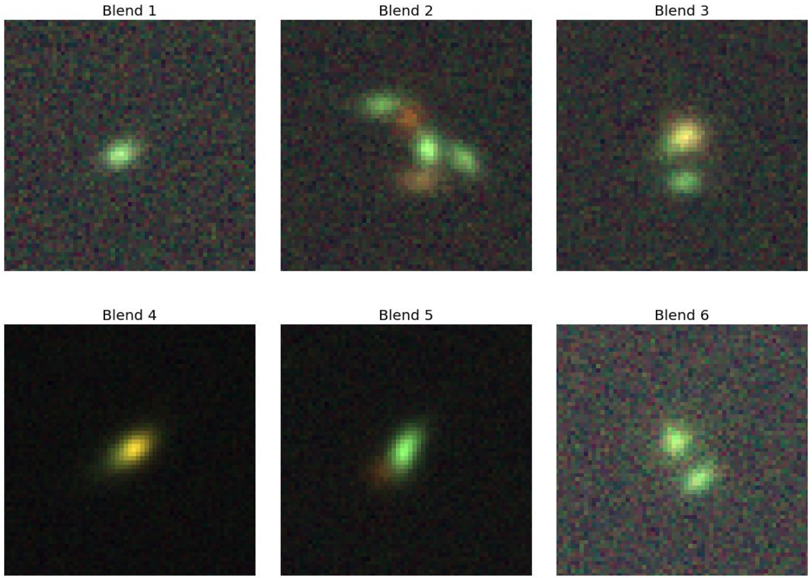


Pixel-level simulation example 2: Blending Tool Kit

Sowmya Kamath

https://github.com/LSSTDESC/BlendingToolKit

- July 2018 DESC Hack Day project [Doux, Kamath, Lanusse, ...]
- Add-on for WeakLensingDeblending package for simulating images of multi-object blends (without analysis step).
- **Goal:** fast "on the fly" generation of images with different PSFs and different noise levels/realizations (for example, for data augmentation for ML training sets).
- Basic version available (with a tutorial).
- Currently under development.
- Suggestions / requests are welcome!

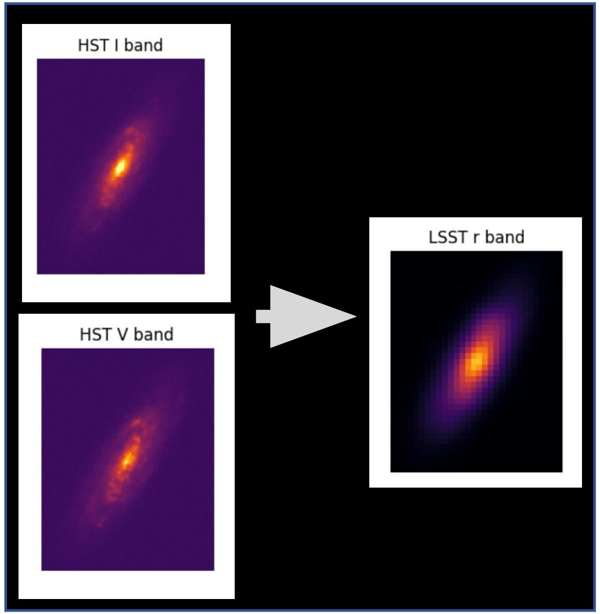


Pixel-level simulation example 3: GalSim (Chromatic)RealGalaxy

Galsim RealGalaxy and ChromaticRealGalaxy classes can be used to decorrelate noise in HST images, and simulate LSST noise and PSF.

Datasets: real galaxy HST images with I< 25.2

- COSMOS (I band): ~87,000 galaxies
- AEGIS (V & I bands): ~26,00 galaxies

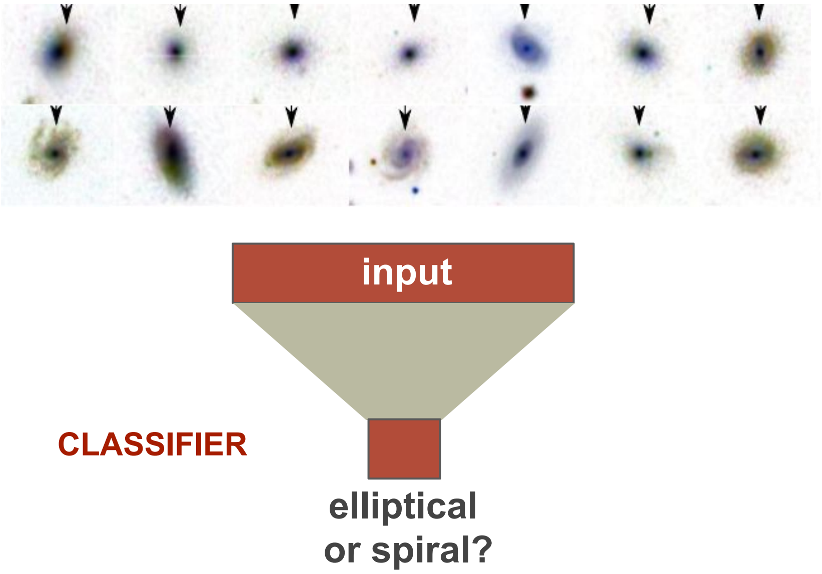


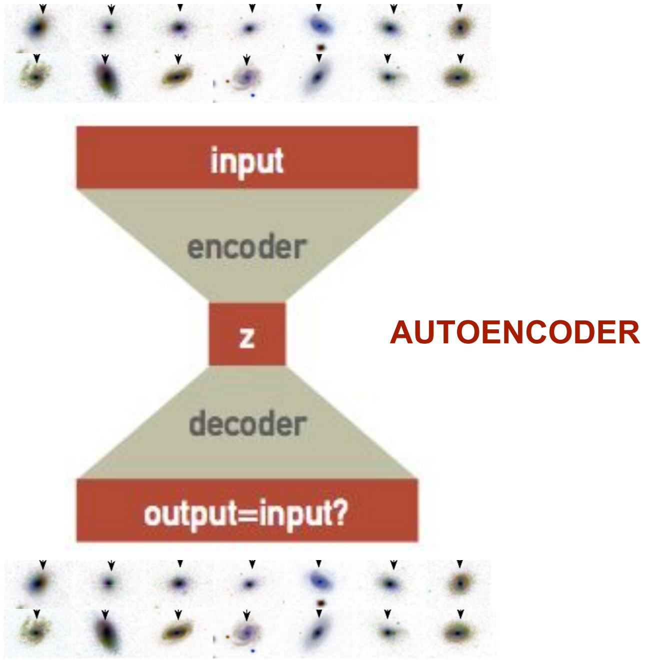
ChromaticRealGalaxy was used with AEGIS dataset to study impact of galaxy color gradients and wavelength dependent PSFs on shear measurements.

Generative models for simulation

David Kirkby

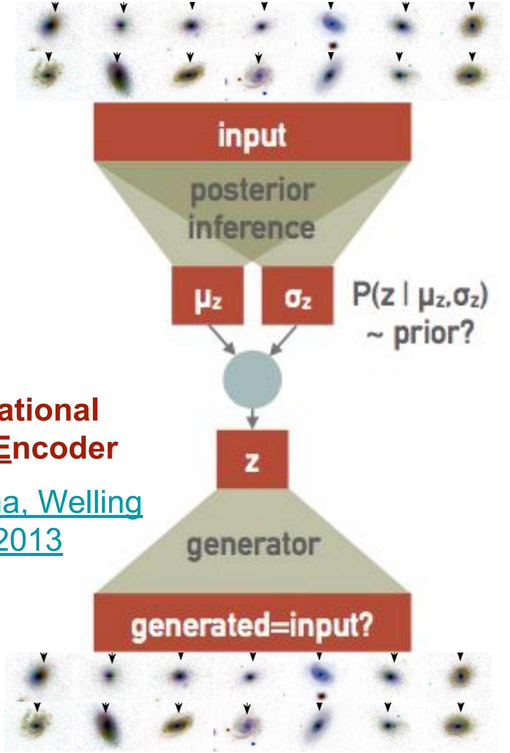
Leverage recent advances in deep neural networks.





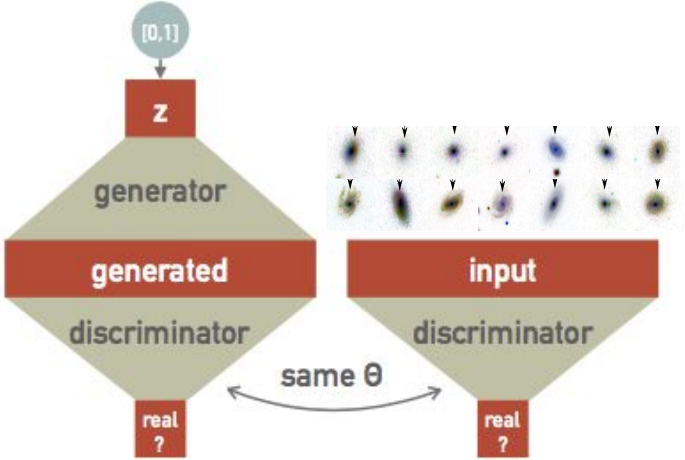
Generative models for simulation

Build sophisticated probabilistic models by decoupling encoder from decoder: details



Variational **A**uto**E**ncoder

Kingma, Welling 2013



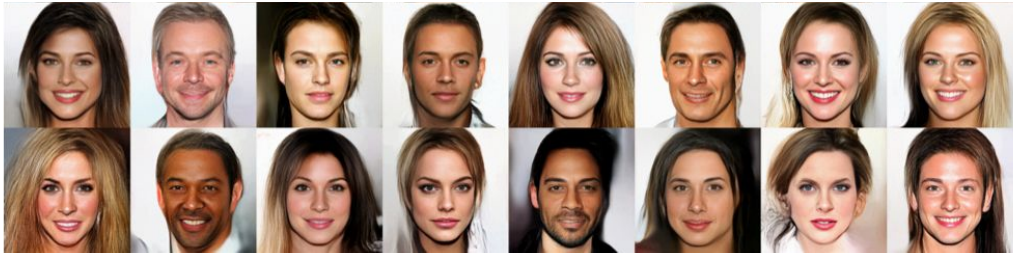
Generative-**A**dversarial **N**etwork

Goodfellow++ 2014

Generative models for simulation

Image processing state of the art:

Generated:

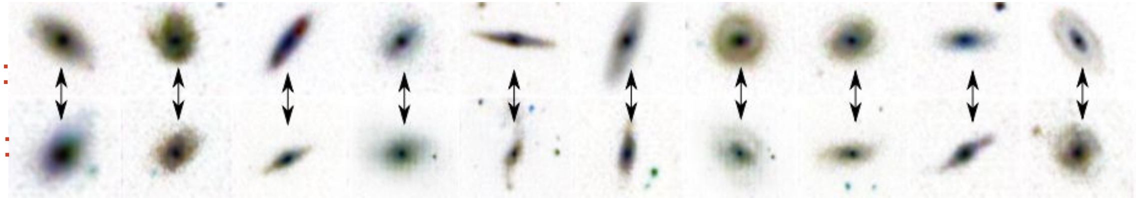


arXiv:1807.03039
arXiv:1703.10717

Potential applications for blending simulation:

- Dense random sampling of a prior that is sparsely sampled from space / DDF.
- Sophisticated data augmentation technique for training deep neural networks.

Real:
Generated:



arXiv:1609.05796

Simulated objects embedded in data

- Balrog: GalSim objects in Dark Energy Survey data [Suchyta, Huff et al.]
- SynPipe: GalSim objects in Hyper Suprime-Cam data [Huong, Leauthaud, Murata et al.]
- LSST science pipeline (in progress).

** See http://adsabs.harvard.edu/abs/2016MNRAS.457..786S

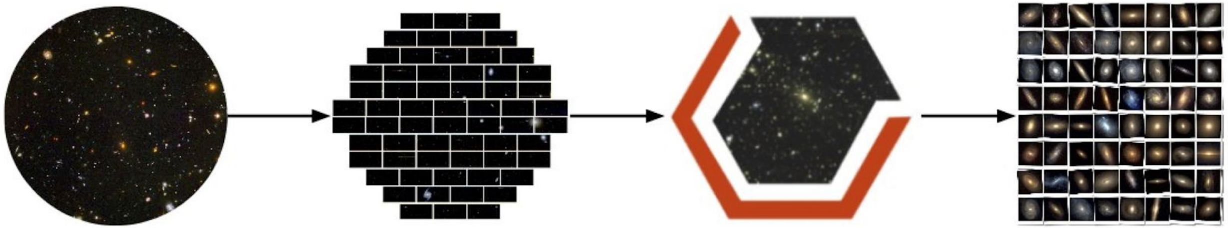
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Balrog: An injection pipeline for the Dark Energy Survey

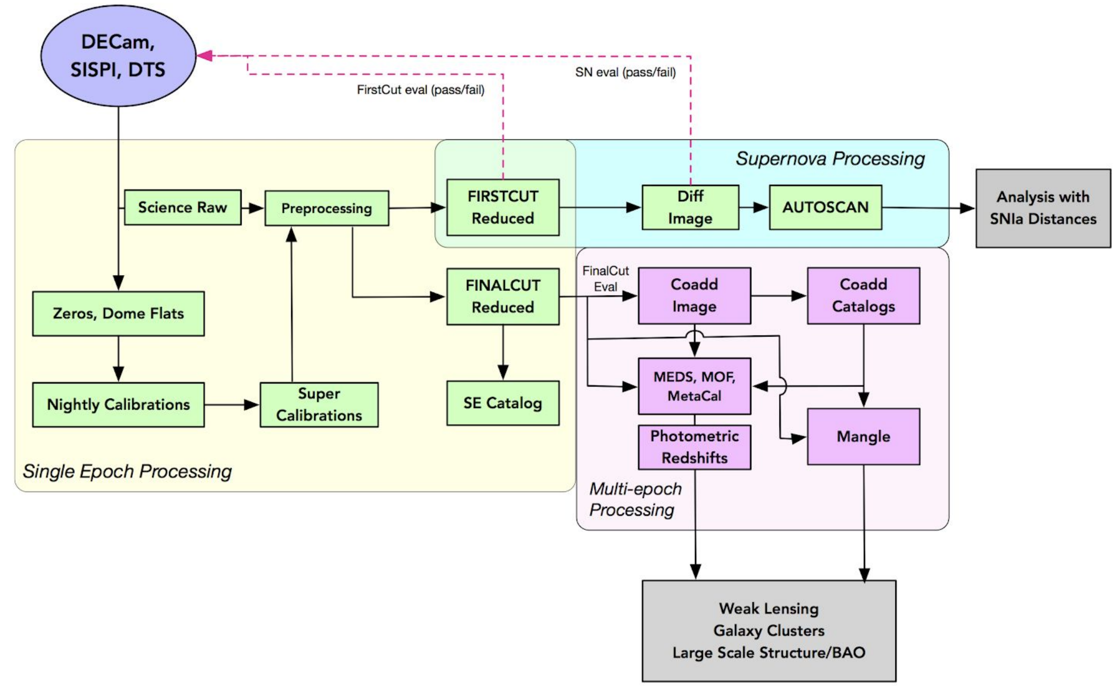
Eric Huff (JPL) for
Sahar Allam, Gary Bernstein, Vinicious Busti, Ami Choi, Katie Eckert, **Spencer Everett**, **Nikolay Kuropatkin**, Eli Rykoff, Erin Sheldon, Megan Splettstoesser, Douglas Tucker, Reese Wilkinson, Brian Yanny, Yuanyuan Zhang, and many others!

$$P(\mathrm{catalog}|\mathrm{universe})$$

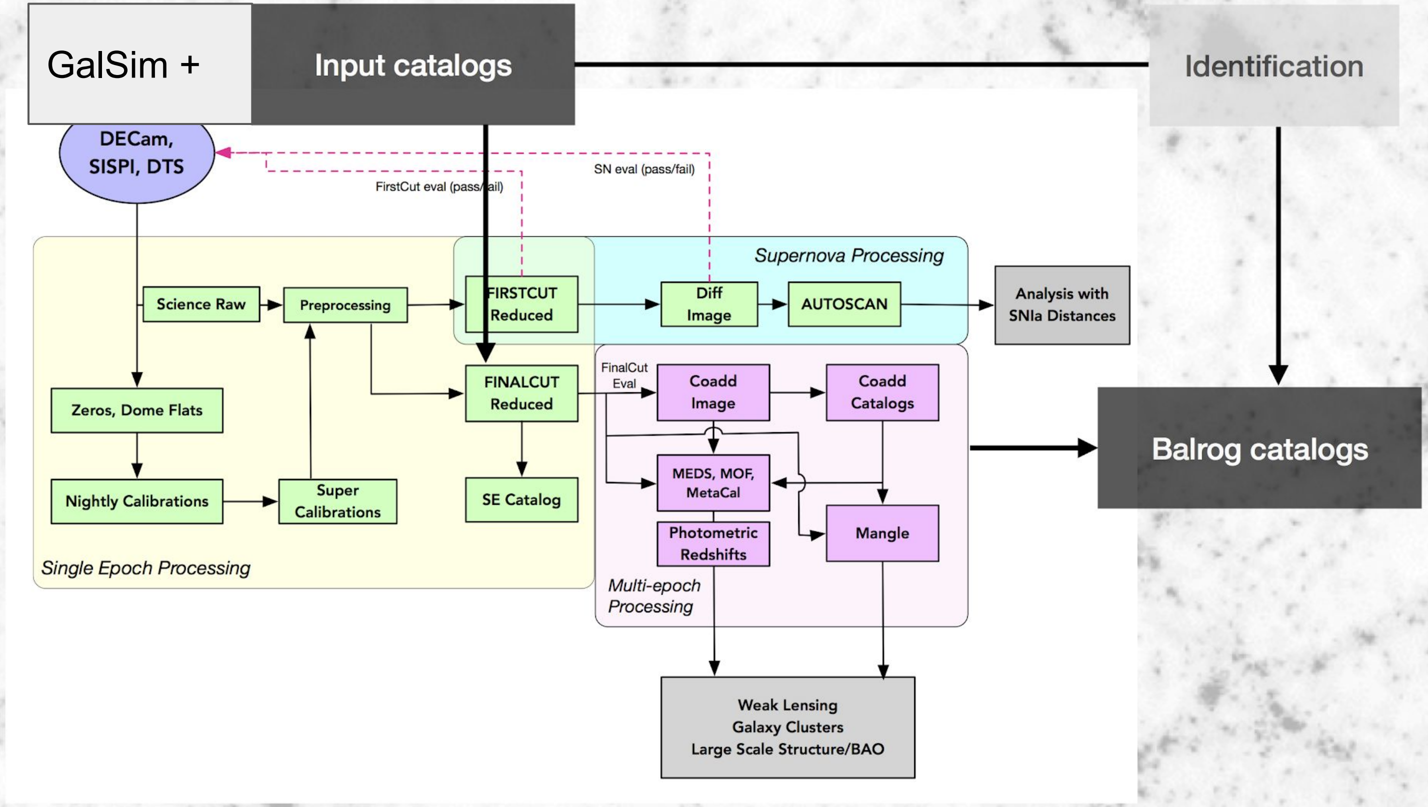


National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Implementation



Implementation



Balrog: The injection simulation package in the Dark Energy Survey

Balrog: The injection simulation package in the Dark Energy Survey

Implementation

Design Choices

1. COSMOS vs. deep fields
2. Injection rate & pattern vs. runtime
3. Full coverage vs. accurate population
4. Postage stamps vs. parametric
5. Signal injection vs. randoms

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Fully realizing the survey is expensive

Base+SWARP+SExtractor: ~1h per tile **Object Injection:** ~1h per tile

SOF: ~1.5h per tile

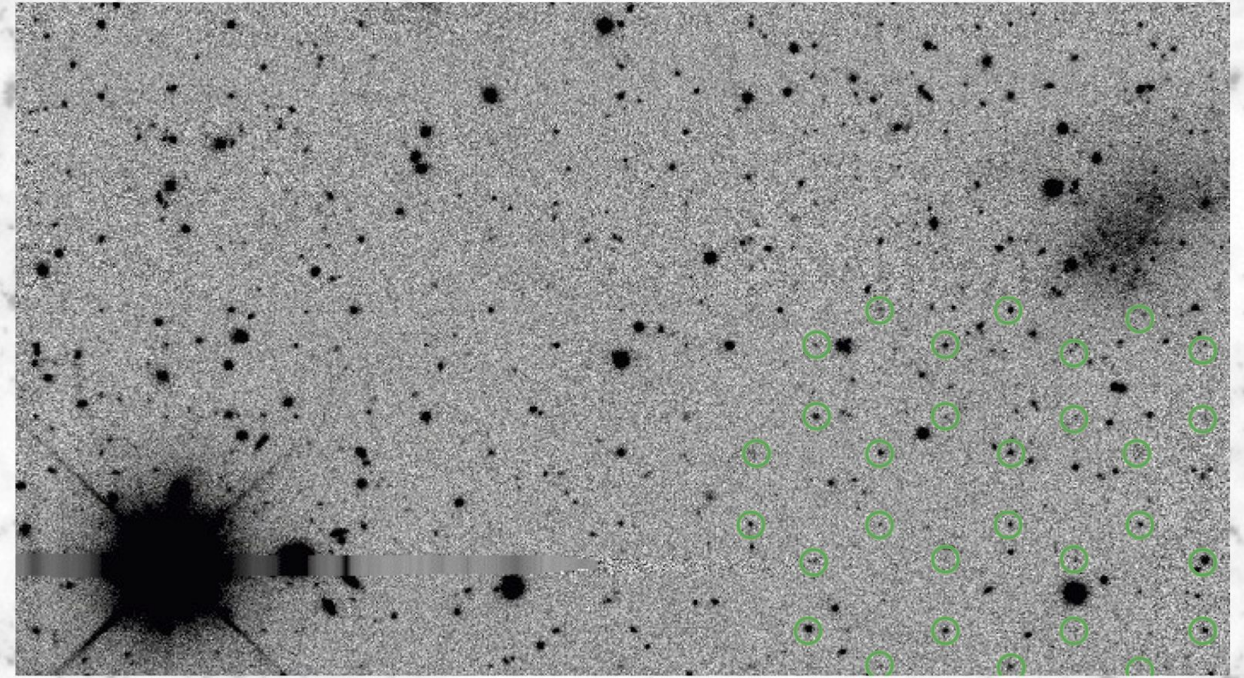
MOF: ~5h per tile

Metacal: ~2h per tile **BFD:** ~1h per tile

Total: ~12 hours per tile per 16 core machine

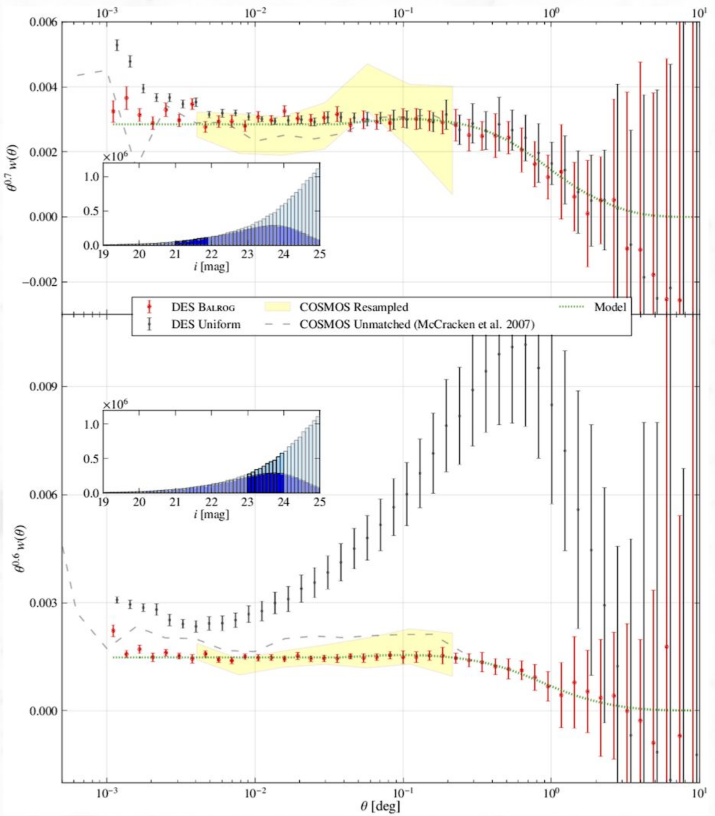
100 dedicated Fermi DES machines with 16 CPU gives **~50 days for entire Y3 footprint** for a single realization

Example: Hex grid injection saves factor of x2 in total compute time



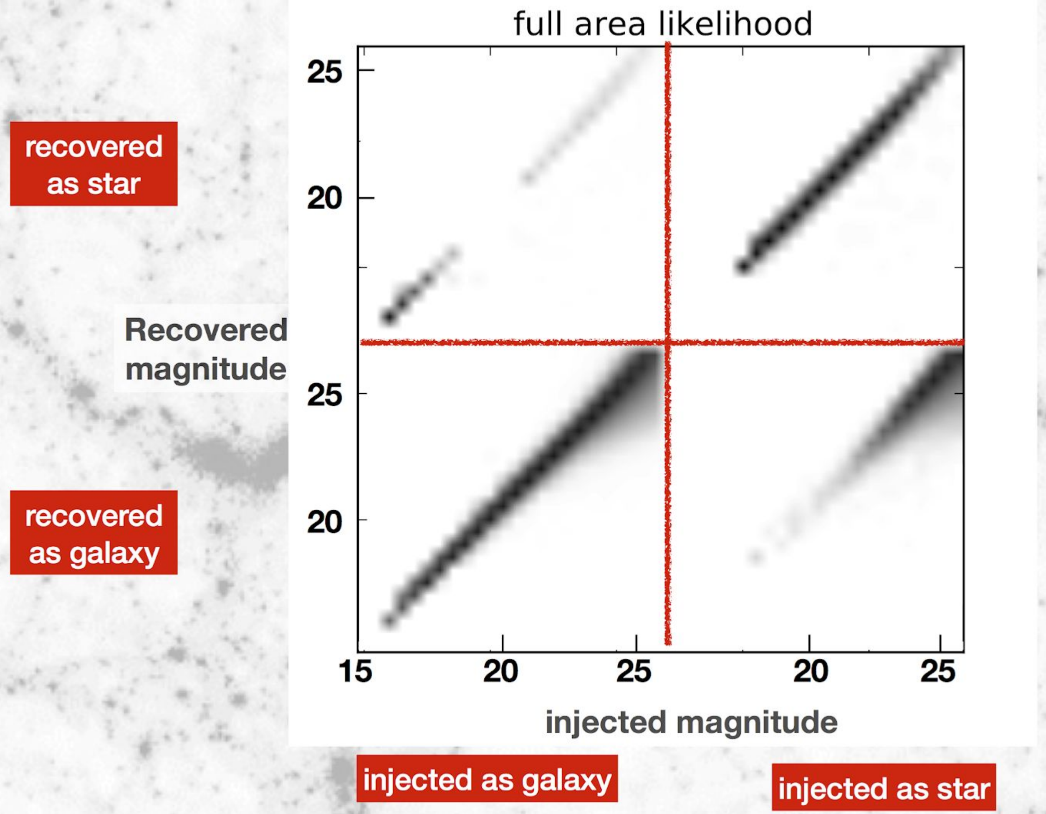
National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Application to galaxy clustering



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California



Links to other simulations, tools & examples

1. Simulations
    a. catalog level: DESC Data Challenge 2, Buzzard, Scinet Light Cone Simulations (SLICS), …
    b. pixel level
        - Weak Lensing Deblending package [David Kirkby]; based on GalSim
        - GPU-ready implementation of GalSim for data augmentation for ML [François Lanusse]
        - Blending Tool Kit: July 2018 DESC Hack Day project [C. Doux, S. Kamath, F. Lanusse, ...]
        - AstrOmatic SkyMaker [Emmanuel Bertin, Pascal Fouqué]
2. Simulated objects embedded in real data
    a. Balrog: GalSim objects in Dark Energy Survey data [Eric Huff]
    b. SynPipe: GalSim objects in Hyper Suprime-Cam data [Huong, Leauthaud, Murata et al.]
    c. LSST science pipeline (in progress).
3. On-the-fly simulations for data augmentation
    a. On-the-fly GalSim image generation and caching with TensorFlow tf.data API: github repo [François Lanusse]