Interpretable statistical learning for galaxy merger science: Key insights and lessons learned

DEEP SKIES

Bringing Artificial Intelligence to Astrophysics

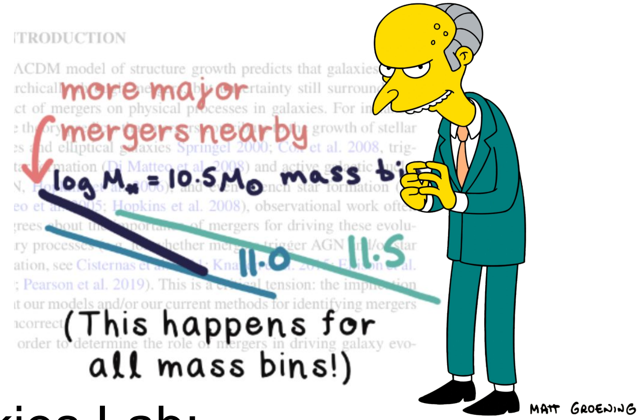
Dr. Becky Nevin

Sept 19 2023

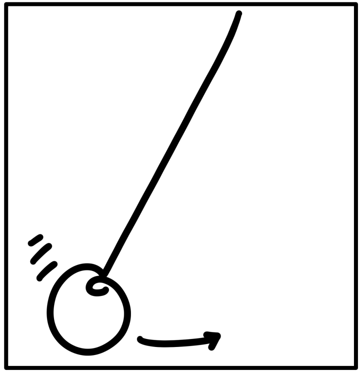
Yale Data-Science X Astronomy-Astrophysics Seminar

Adventures in
statistical confounds

Mergers



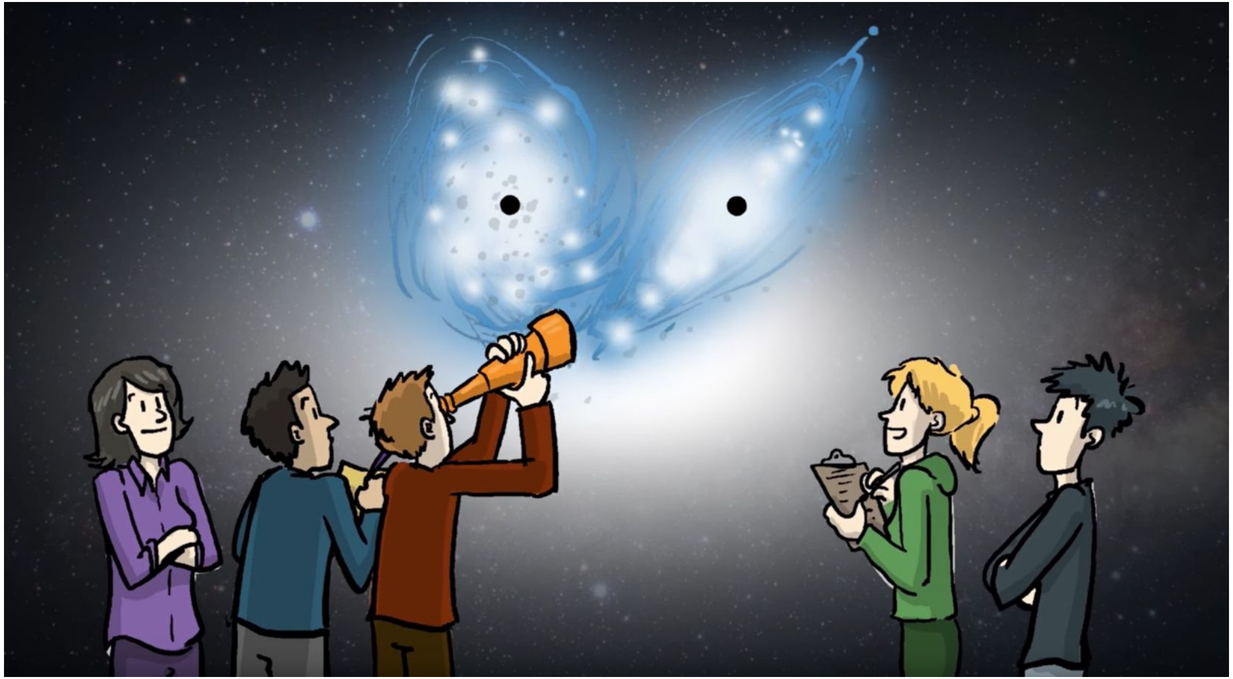
DeepSkies Lab:
Benchmark data for ML, uncertainty, and fancy Bayesian inference, oh my!



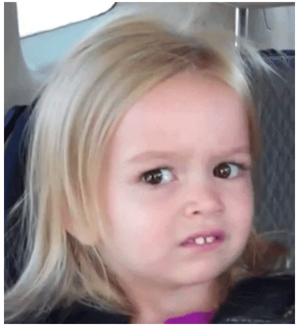
Why are merging galaxies important?

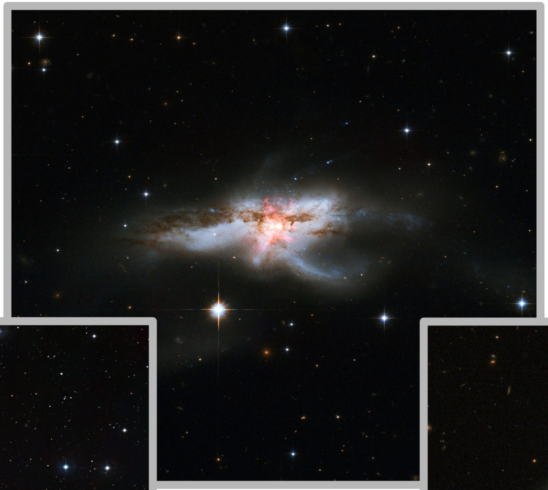
Short answer: gas

Long answer: structure formation (bulge, spiral arms, bars), triggering and suppressing star formation, triggering Active Galactic Nuclei



Accurately and consistently identifying mergers is hard





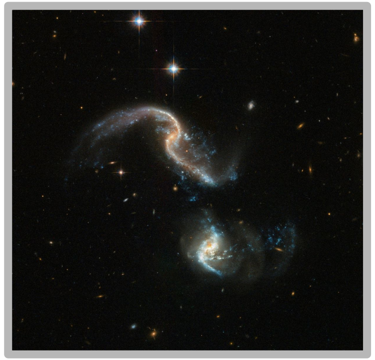
There are many different types (mass ratios) of mergers and they all look different observationally

Major merger

There are many different types (mass ratios) and stages of mergers and they all look different observationally

Close pairs

Interacting

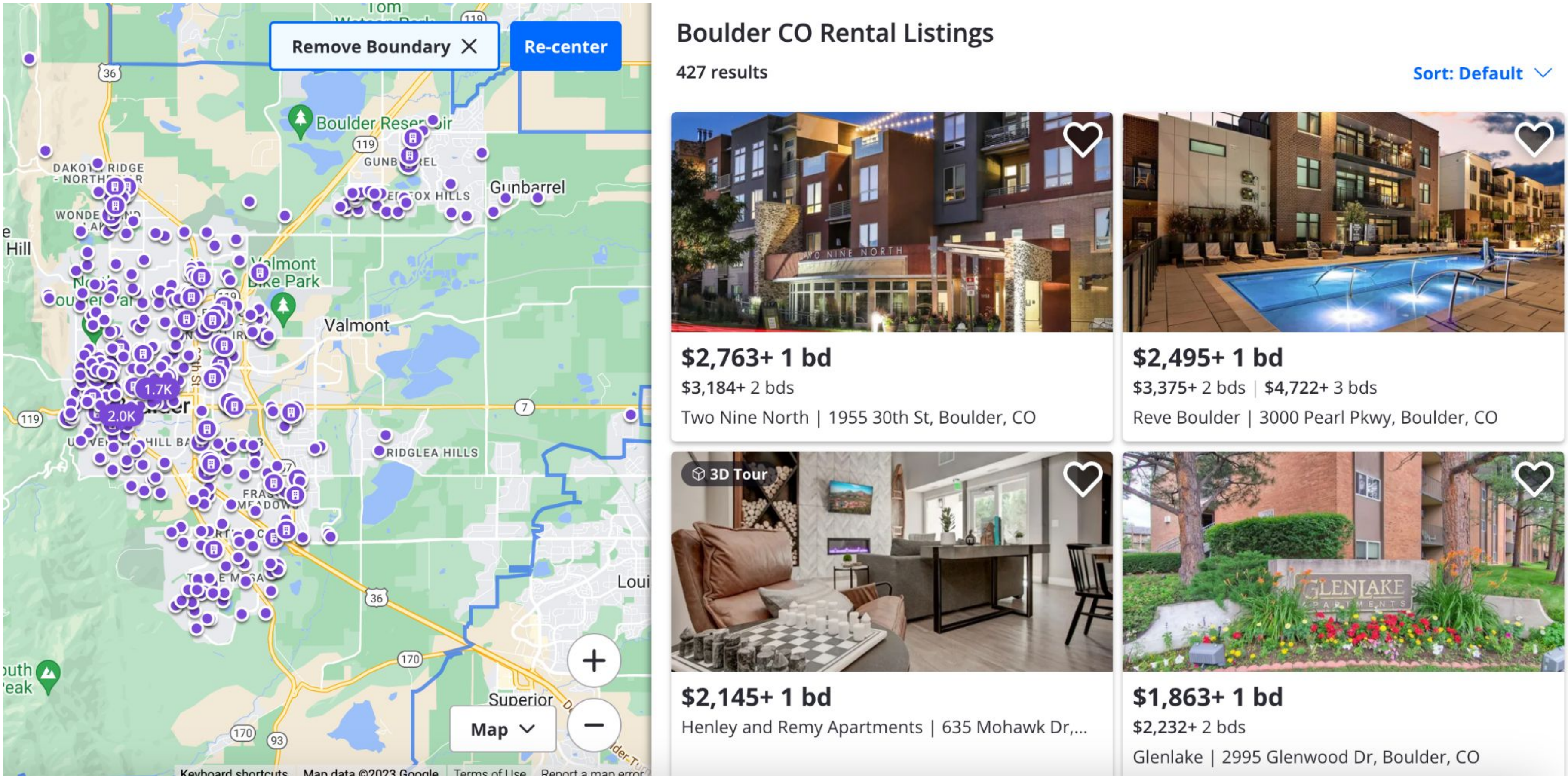


Coalescence

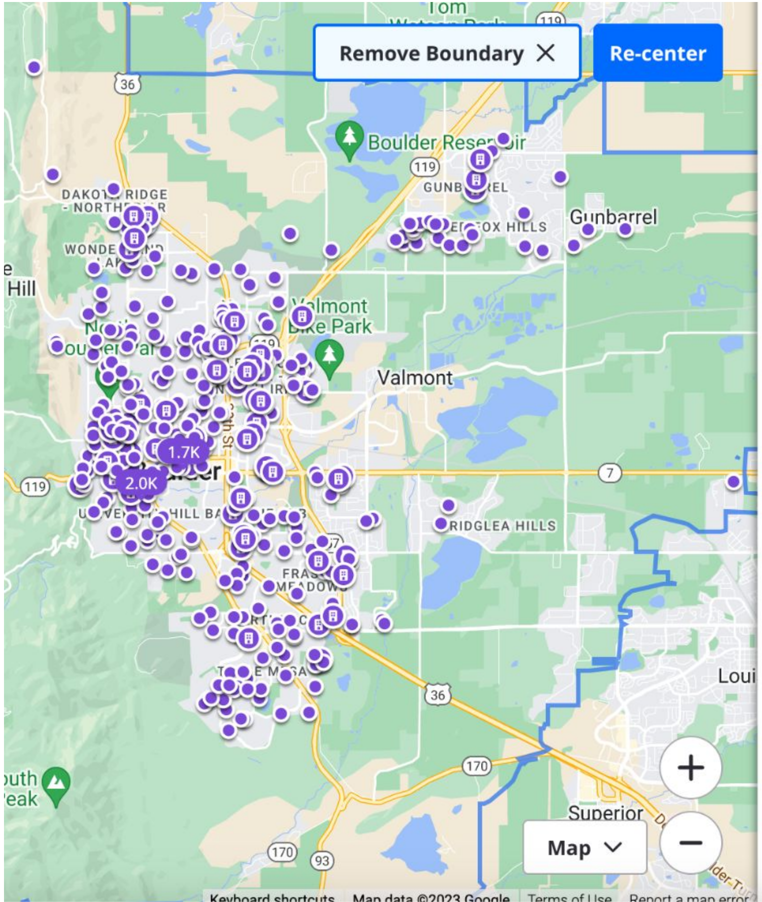
Post-merger

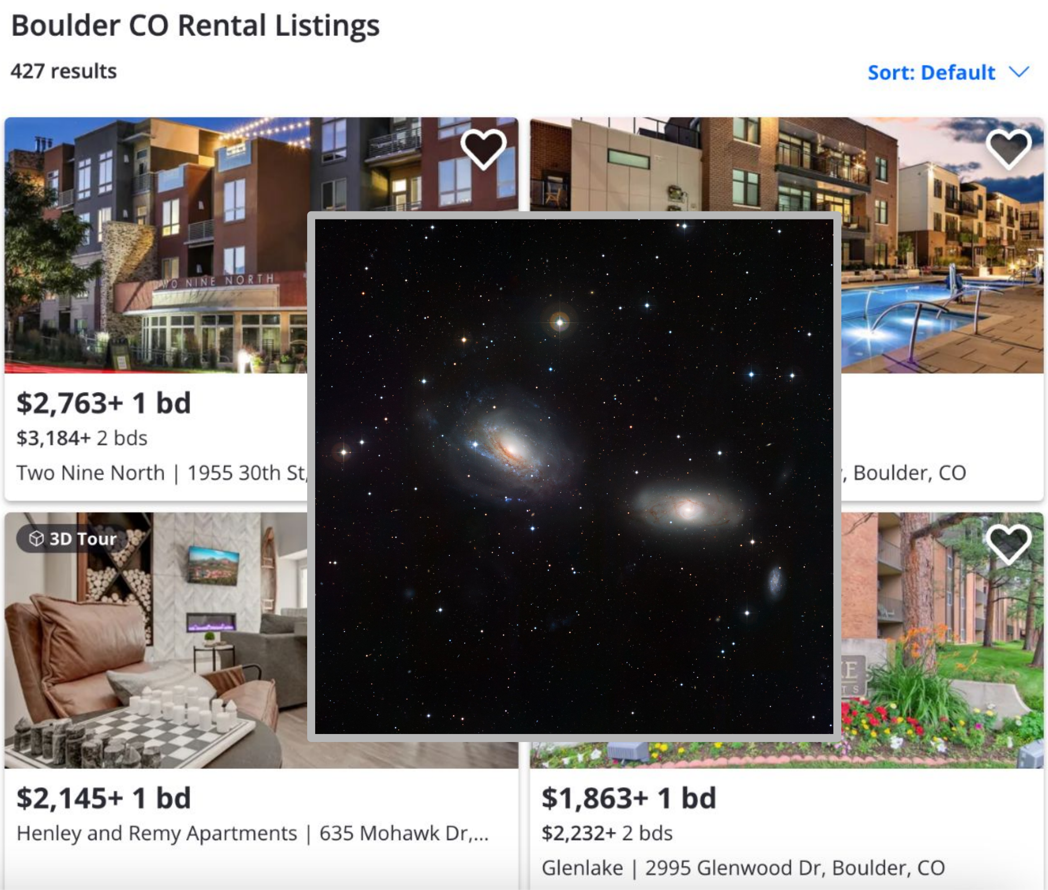
How do we identify a diversity of galaxy mergers?

How do we ~~identify a diversity of galaxy mergers?~~ find an apartment to rent?

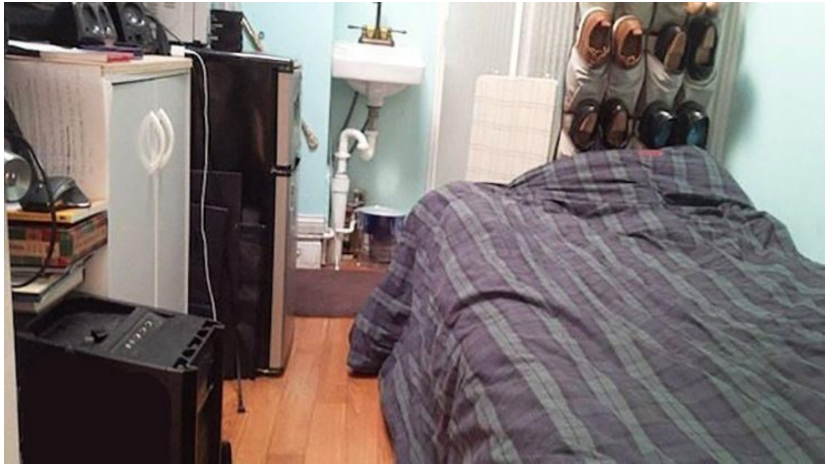


How do we ~~identify a diversity of galaxy mergers?~~ find an apartment to rent?

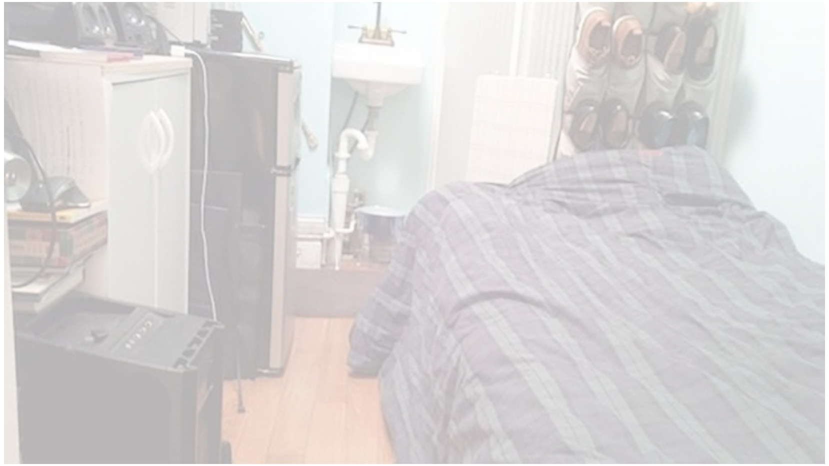




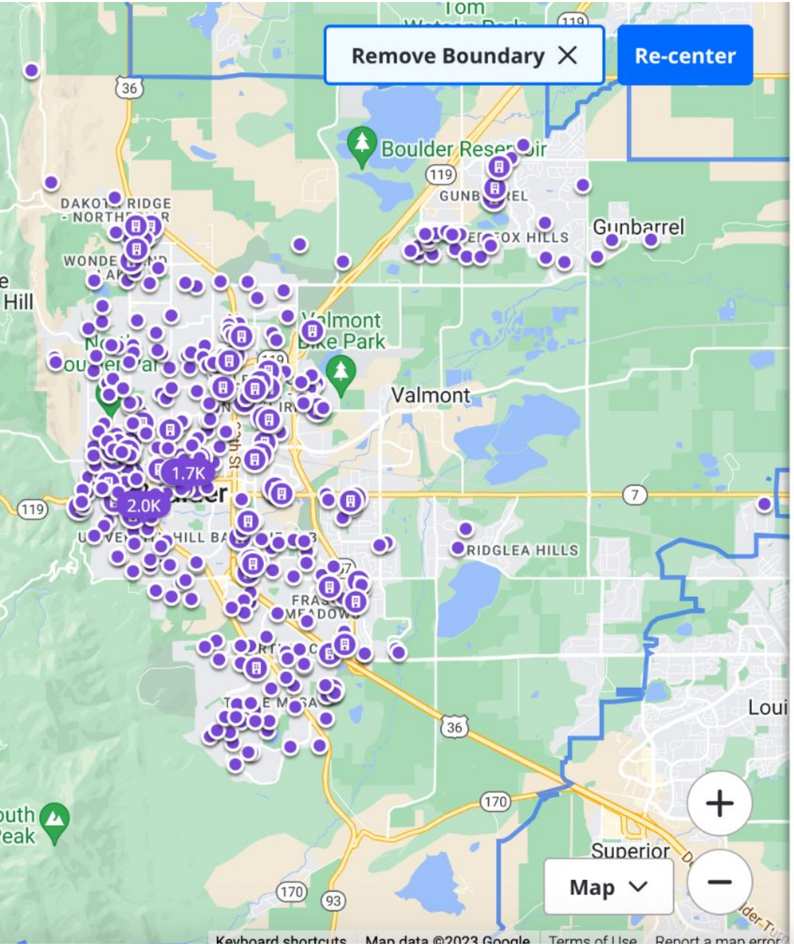
How do we identify a diversity of galaxy mergers?



How do we identify a diversity of galaxy mergers?



Search engine matters, filters within the search engine also matter



Boulder CO Rental Listings

427 results

Sort: Default

$2,763+ 1 bd
$3,184+ 2 bds
Two Nine North | 1955 30th St, Boulder, CO

$2,495+ 1 bd
$3,375+ 2 bds | $4,722+ 3 bds
Reve Boulder | 3000 Pearl Pkwy, Boulder, CO

$2,145+ 1 bd
Henley and Remy Apartments | 635 Mohawk Dr,...

$1,863+ 1 bd
$2,232+ 2 bds
Glenlake | 2995 Glenwood Dr, Boulder, CO

Buy Rent Sell Home Loans Agent finder **Zillow** Manage rentals Advertise Sign in or Join Help

Somerville, MA | For Rent | $1k-$2k | Beds: 1+ | Home type | More: 1 | Save Search | Show Map

$1,000/mo 1 bd | 2 ba | 4,158 sqft
Columbia St, Cambridge, MA 02139
Apartment for rent

$1,050/mo 3 bds | 1 ba | 868 sqft
Concord Ave, Cambridge, MA 02138
Apartment for rent

$1,100/mo 1 bd | 1 ba | 1,359 sqft
469 Windsor St # 1, Cambridge, MA 02141
Apartment for rent

$1,600/mo 1 bd | 1 ba | 9,999 sqft
43 Rice St # 1, Cambridge, MA 02140
Apartment for rent

$1,600/mo 1 bd | 1 ba | 99,999 sqft
Rice St, Cambridge, MA 02140
Apartment for rent

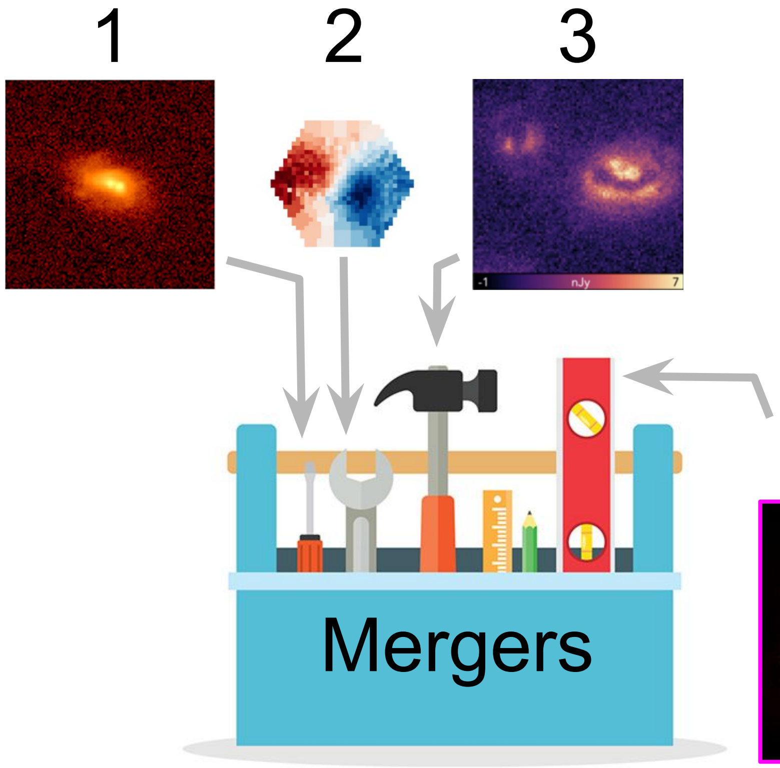
$1,700/mo 1 bd | 1.5 ba | 99,999 sqft
2534 Massachusetts Ave APT 3, Cambridge, MA 02140
Apartment for rent

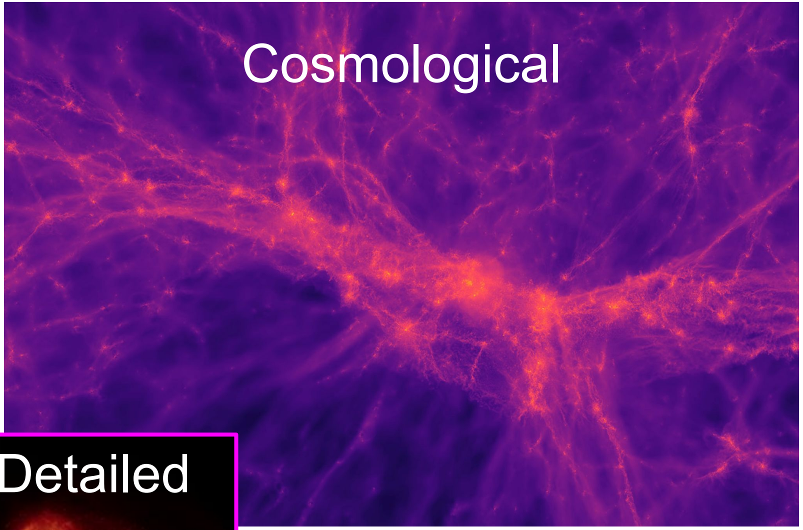
It is important to understand false positives

What can we learn from apartment hunting?

- The tool matters, the tool within the tools matters (filters)
- Combining tools can be great
- Intuition is helpful

My work approaches better identifying mergers with the help of detailed hydro and cosmological simulations

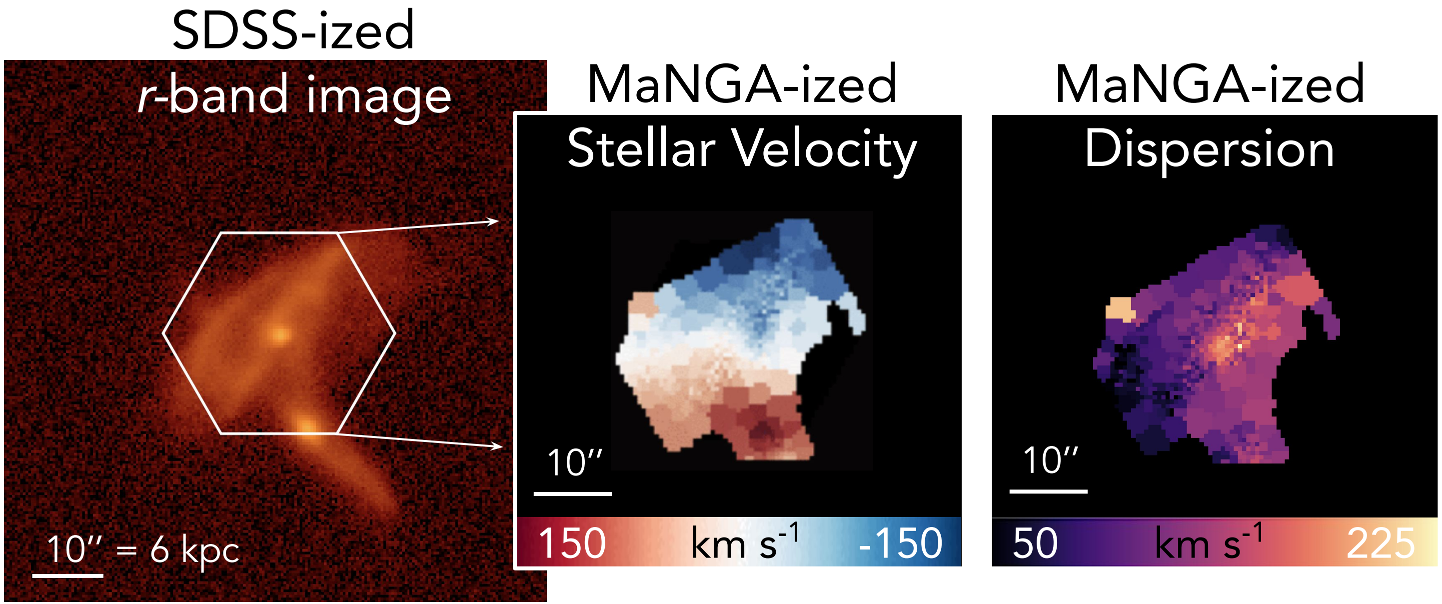




15 Mpc across (above)
~70 kpc across (left)

You can think of this as sim city

I create mock stellar kinematic maps to match the specifications of MaNGA integral field spectroscopy



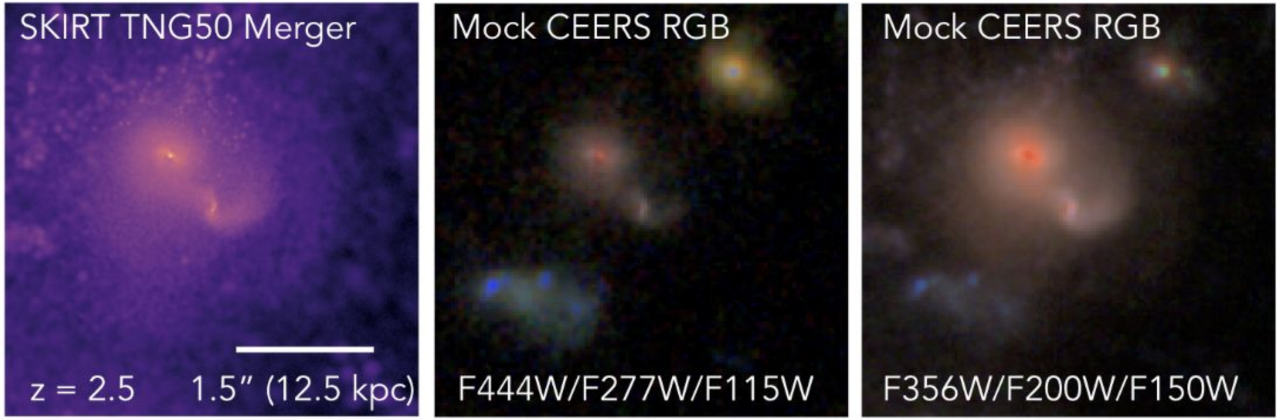
Nevin+2019

Nevin+2021

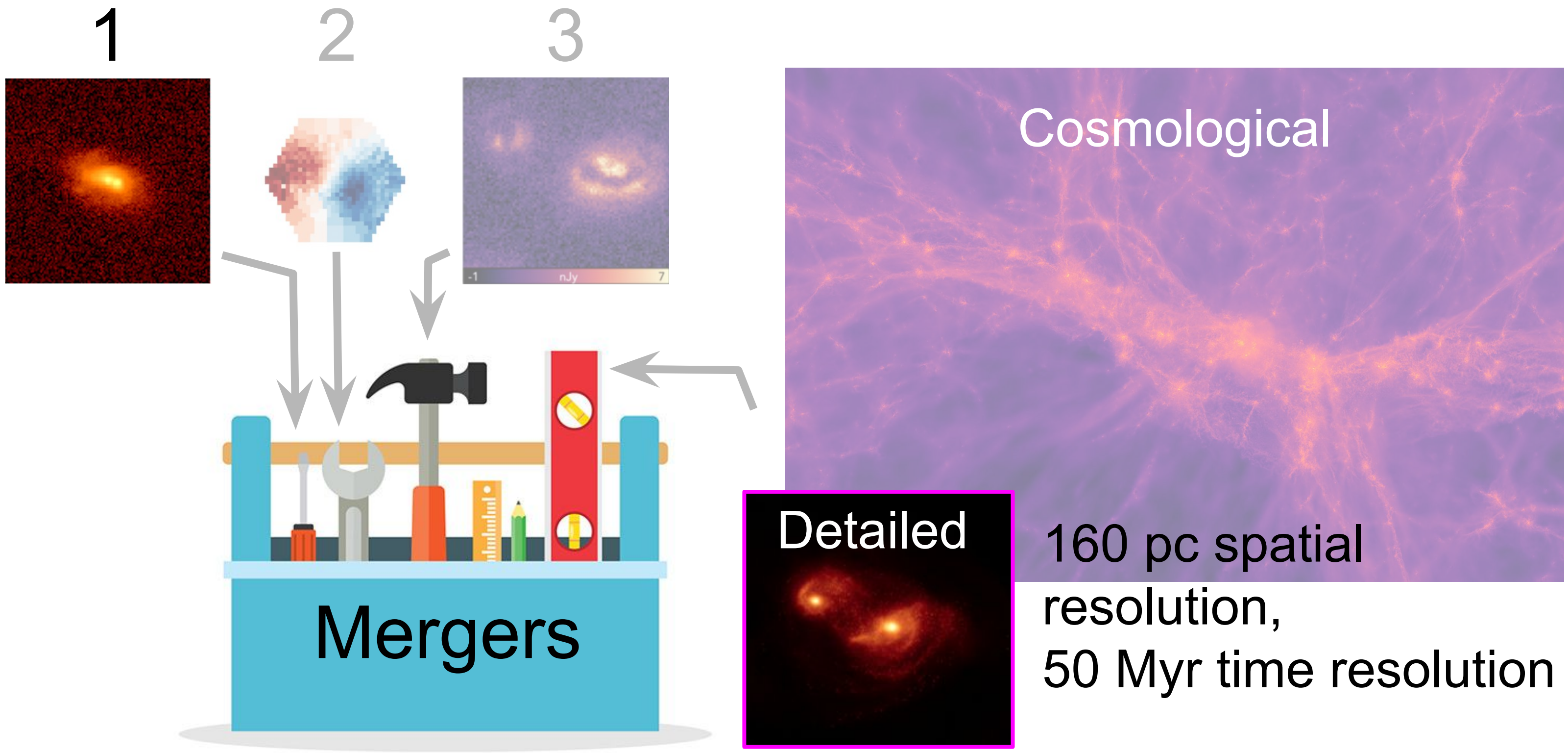
How do we best identify high redshift merging galaxies?:

Expanding the toolkit to include *HST* Candels and *JWST* NIRCam imaging

Aimee Schechter



Focusing on just the detailed imaging approach to identifying mergers is enough for one day



2019 https://arxiv.org/abs/1901.01975

Accurate Identification of Galaxy Mergers with Imaging

R. NEVIN,[1] L. BLECHA,[2] J. COMERFORD,[1] AND J. GREENE[3]

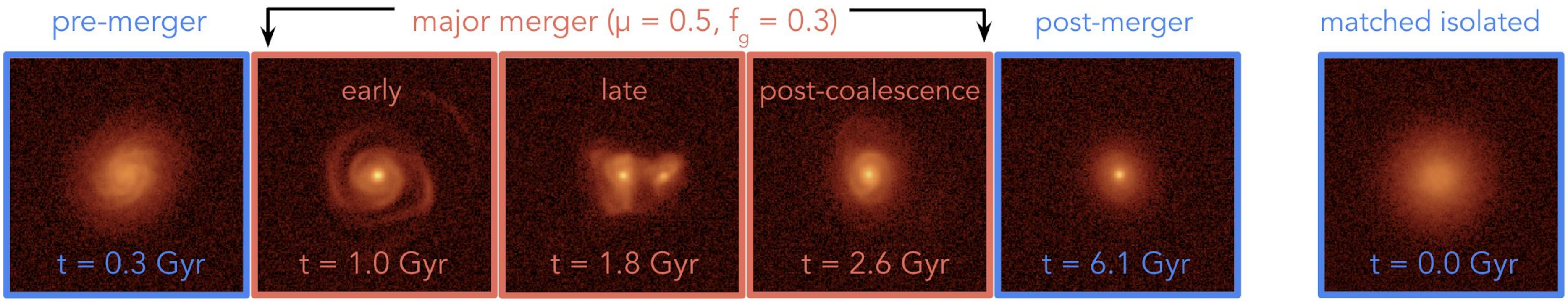
[1]*Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA*
[2]*Department of Physics, University of Florida, Gainesville, FL 32611, USA*
[3]*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

2023 https://ui.adsabs.harvard.edu/abs/2023MNRAS.522....1N/abstract

A declining major merger fraction with redshift in the local Universe from the largest-yet catalog of major and minor mergers in SDSS

R. Nevin,[1★] L. Blecha,[2] J. Comerford,[3] J. Simon,[3†] B. A. Terrazas,[4] R. S. Barrows,[3]
J. A. Vázquez-Mata[5]

Simulations of merging and nonmerging galaxies

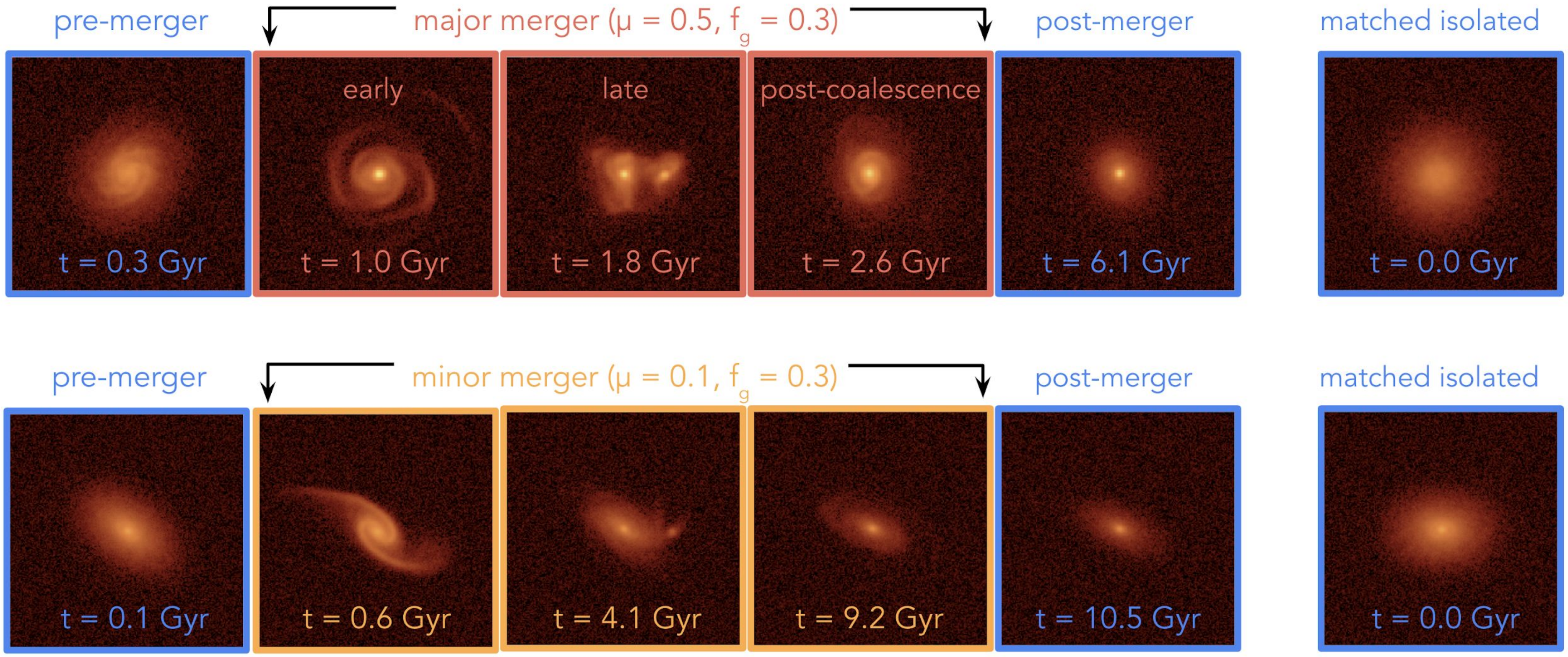


100s of snapshots per simulation
x 5 simulations

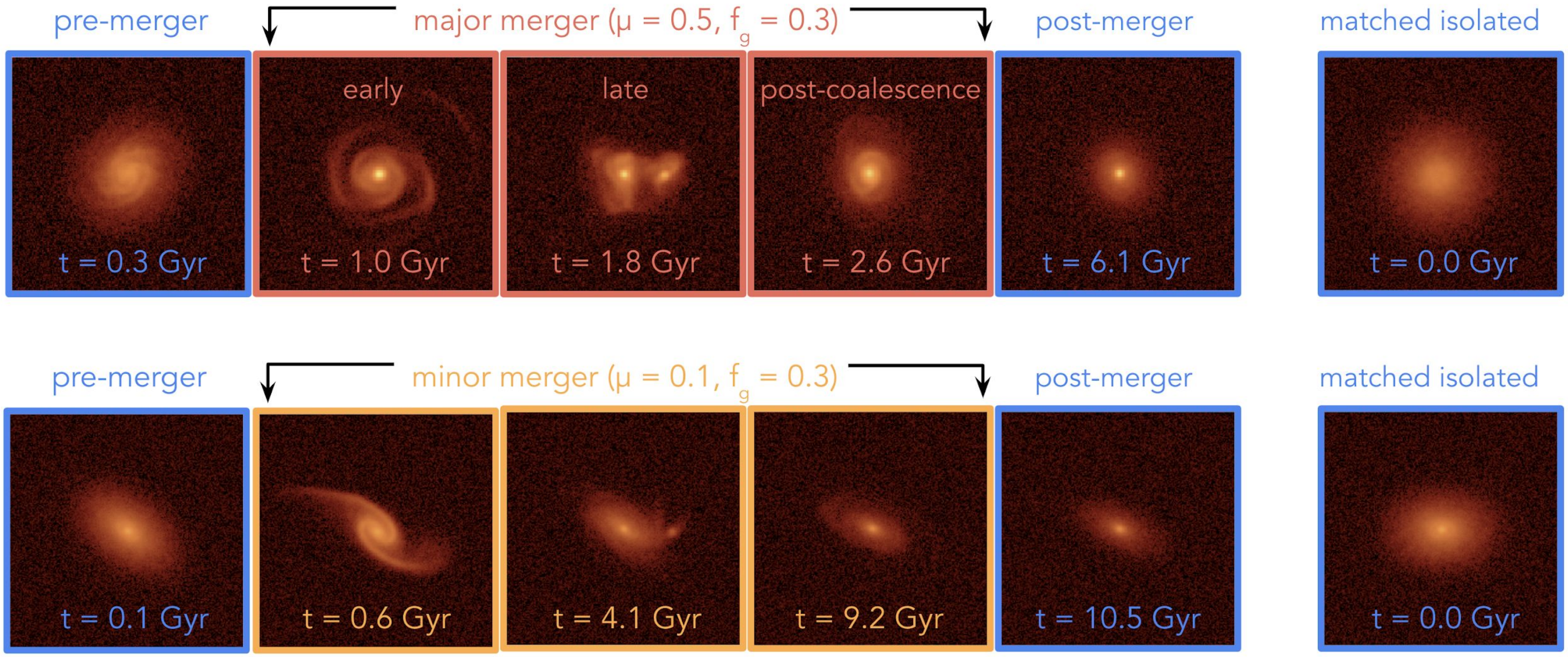
GADGET-3 N-Body Simulations: Springel & Hernquist 2003, Springel 2005, Blecha+2018

Nevin+2019

Simulations of merging and nonmerging galaxies



My pipeline creates mock Sloan Digital Sky Survey (SDSS) images and measures predictors

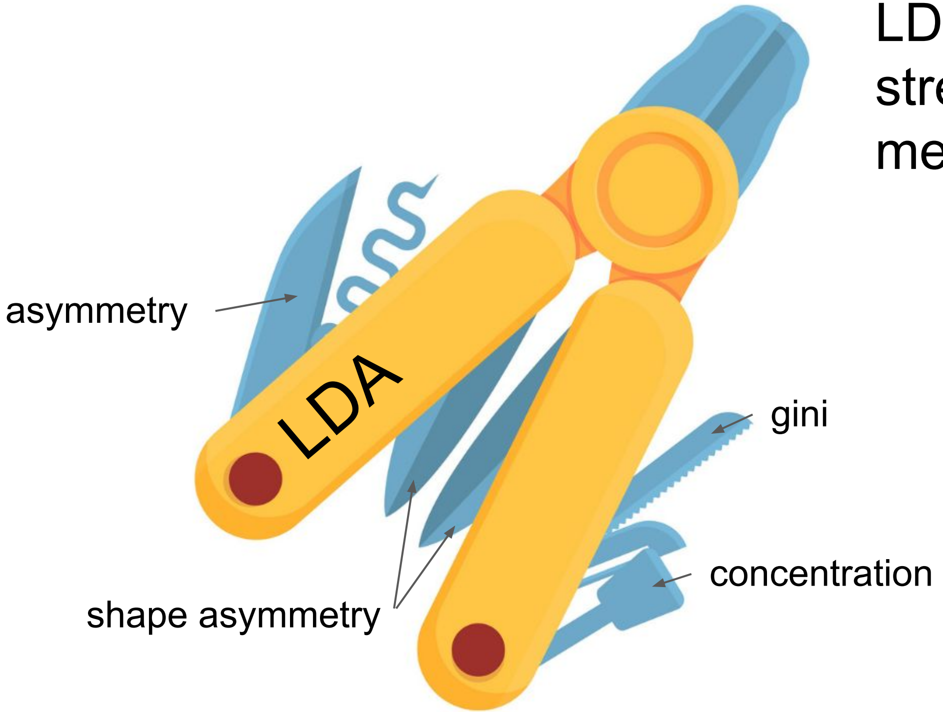


These simulations were carried out with GADGET-3 (Springel & Hernquist 2003; Springel 2005), a smoothed-particle hydrodynamical (SPH) and N-body code that conserves energy and entropy and includes sub-resolution models for physical processes such as radiative heating and cooling, star formation and supernova feedback, and a multiphase interstellar medium (ISM). All simulations have a baryonic mass resolution of 2.8×10^4 $M_\odot$ and a gravitational softening length of 23 pc. SMBHs are modeled as gravitational "sink" particles that accrete via an Eddington-limited Bondi-Hoyle (Bondi & Hoyle 1944) prescription. AGN feedback is also incorporated by coupling 5% of the accretion luminosity ($L_{\rm bol} = \epsilon_{\rm rad}\dot{M}c^2$) to the gas as thermal energy. We assume a radiative efficiency $\epsilon_{\rm rad} = 0.1$ for accretion rates $\dot{M} > 0.01\dot{M}_{\rm Edd}$ (where $\dot{M}_{\rm Edd}$ is the Eddington limit); below this we assume radiatively inefficient accretion following Narayan & McClintock (2008). GADGET has been used for many studies concerning merging galaxies (e.g., Di Matteo et al. 2005; Snyder et al. 2013a; Blecha et al. 2011a; Blecha et al. 2013).

Table 1. Key parameters of our suite of high-resolution GADGET-3 galaxy merger simulations.

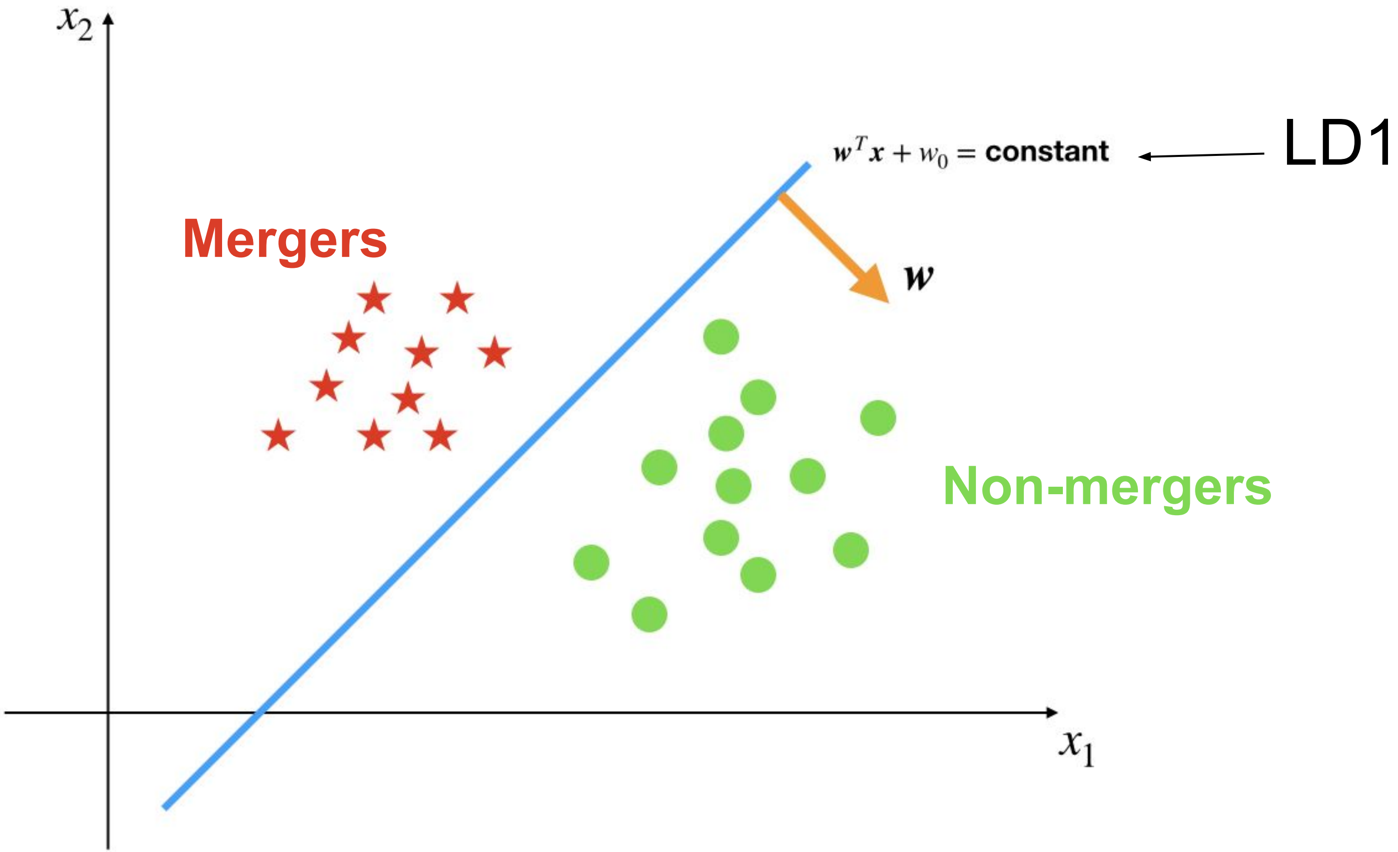
Model	$M_{\rm tot}$ [10^{11} M$_\odot$]	Stellar Mass [10^{10} M$_\odot$]	Gas Fraction	Mass Ratio
q0.5_fg0.3	20.8	5.9	0.3	1:2
q0.333_fg0.3	18.7	5.2	0.3	1:3
q0.333_fg0.1	18.7	6.3	0.1	1:3
q0.2_fg0.3_BT0.2	16.8	5.0	0.3	1:5
q0.1_fg0.3_BT0.2	15.1	4.6	0.3	1:10

I developed a tool within a tool known as Linear Discriminant Analysis (LDA)



LDA combines the strengths of all seven measured predictors

LDA finds the linear hyperplane that best separates mergers from non-mergers



Relevant details of the LDA classification include:

- The LDA relies on a prior to correct for the larger fraction of merging relative to nonmerging galaxies in the simulations. In N19, we use fiducial merger fraction priors of $f_{\mathrm{merg}} = 0.1$ and 0.3 for the major and minor merger classifications, respectively. We explore how changing the merger fraction prior affects our measured posterior merger fraction in §4.7.
- We include interaction terms to explore correlations between predictors.
- We use k-fold cross-validation to obtain 1σ errors on the predictor coefficients and to measure the performance statistics of the classifications.
- In order to select which coefficients are necessary for the classification, we use a forward step-wise selection technique, which orders and includes only the relevant terms and interaction terms.

We solve for the hyperplane that satisfies the above equation, LD1:

$$\text{LD1} = \hat{\vec{w}}^T \vec{x} + \hat{w}_0 = 0$$

where the

$$p_{\text{merg}} = \frac{1}{1 + e^{-\text{LD1}}}$$

and the intercept is given by w_0:

$$\hat{w_0} = \frac{1}{2}\mu_0^T \Sigma^{-1} \mu_0 + \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1 + \log(\frac{\hat{\pi}_0}{\hat{\pi}_1})$$

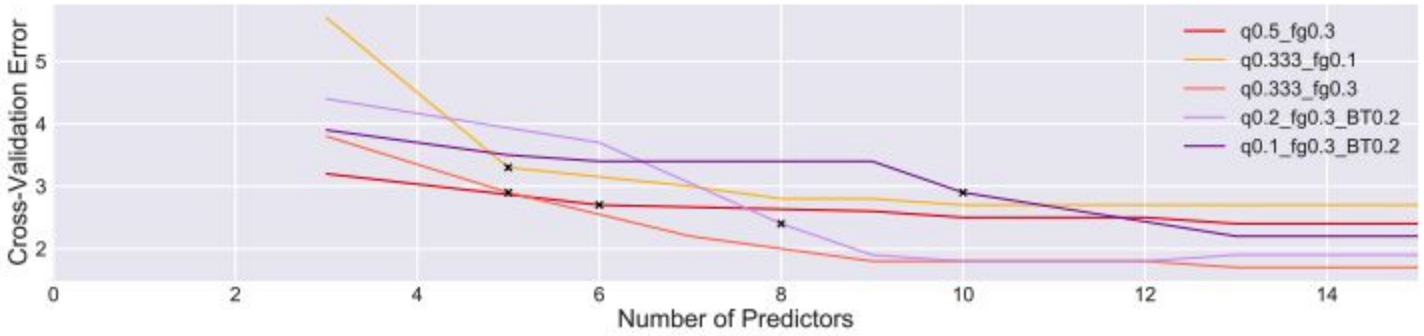
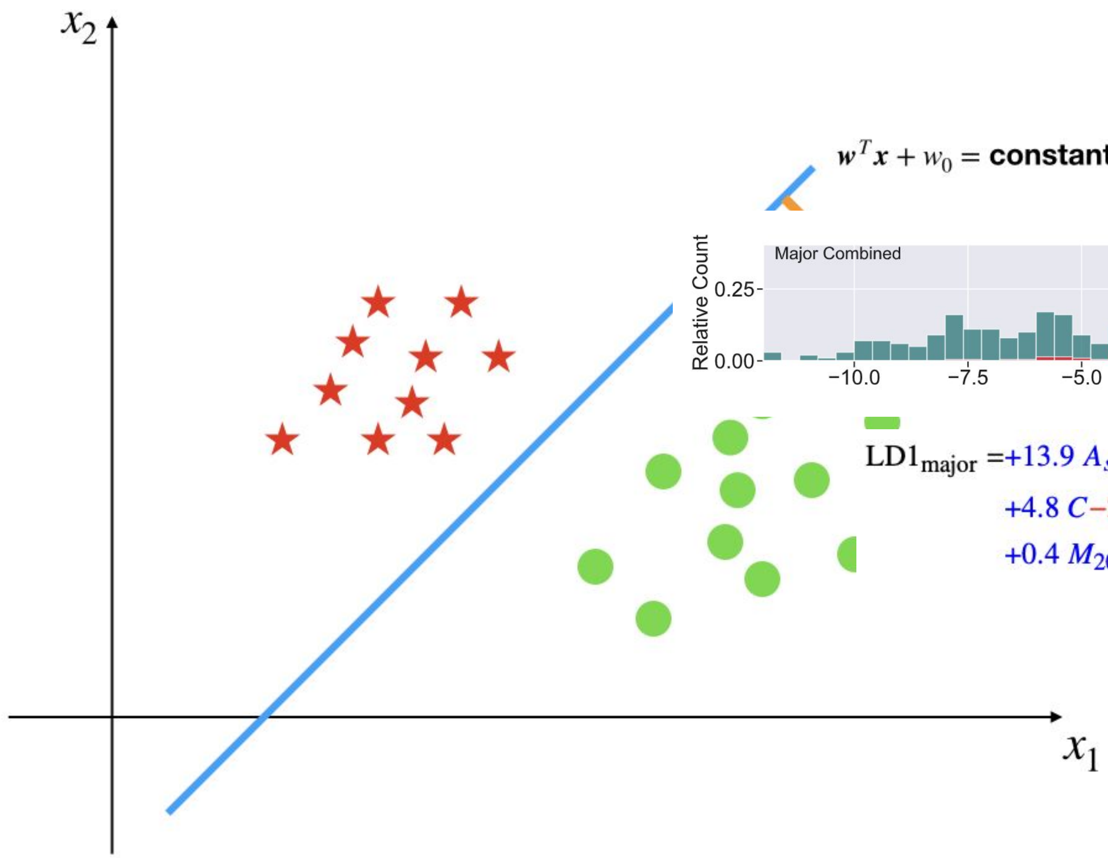
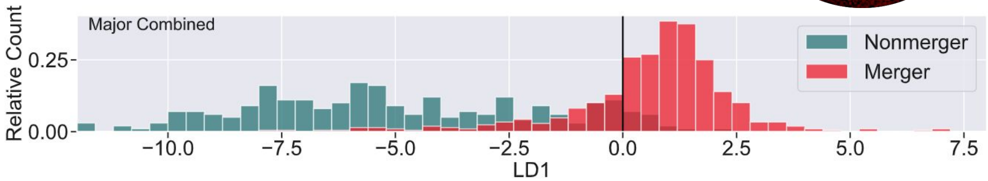


Figure 18. Forward stepwise selection of the number of predictors for each run of LDA. We mark the minimum number of 'required' predictors for each run with black xs. This point is within one standard error of the minima of the cross-validation error curve for each run. We run LDA for each simulation using the predictors selected from this method.



$w^T x + w_0 =$ **constant** ← LD1

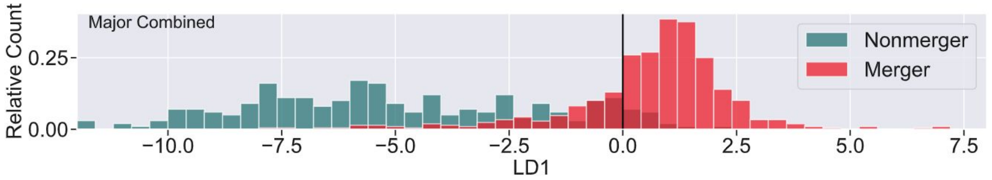


$$\mathrm{LD1}_{\mathrm{major}} = +13.9\, A_s - 8.0\, C * A_s - 5.4\, A * A_s + 5.1\, A + 4.8\, C - 2.9\, Gini * A_s + 0.6\, M_{20} * A + 0.4\, M_{20} * n + 0.4\, Gini - 0.6$$

LDA advantages

The LDA is more accurate and precise than any of the individual predictors in identifying mergers.

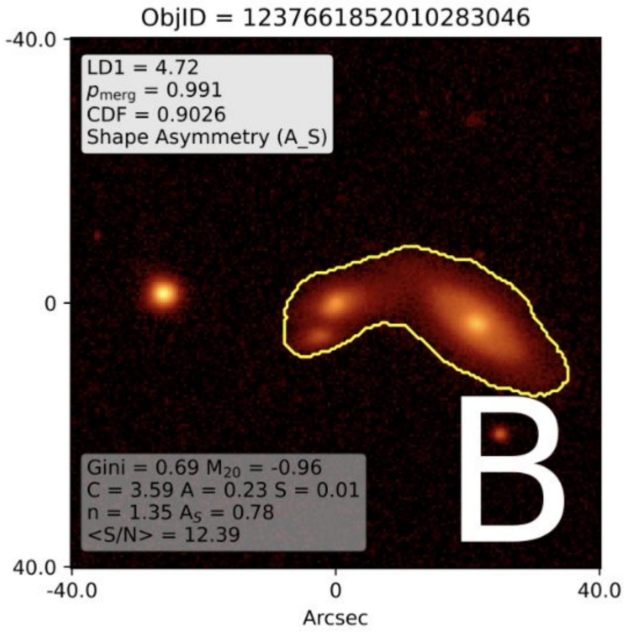
It is also not a black box!



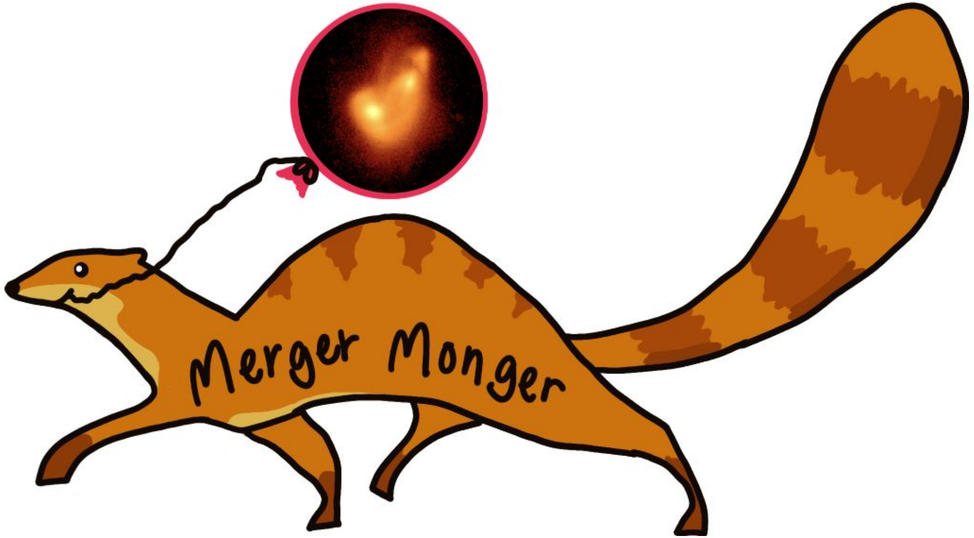
$$\mathrm{LD1}_{\mathrm{major}} = +13.9\, A_s - 8.0\, C * A_s - 5.4\, A * A_s + 5.1\, A + 4.8\, C - 2.9\, Gini * A_s + 0.6\, M_{20} * A + 0.4\, M_{20} * n + 0.4\, Gini - 0.6$$

Which imaging predictors are most important?

I measure predictor values and classify the ~1.3 million galaxies in SDSS using MergerMonger



[MergerMonger Github Repo]



Nevin+2023

Catalogs

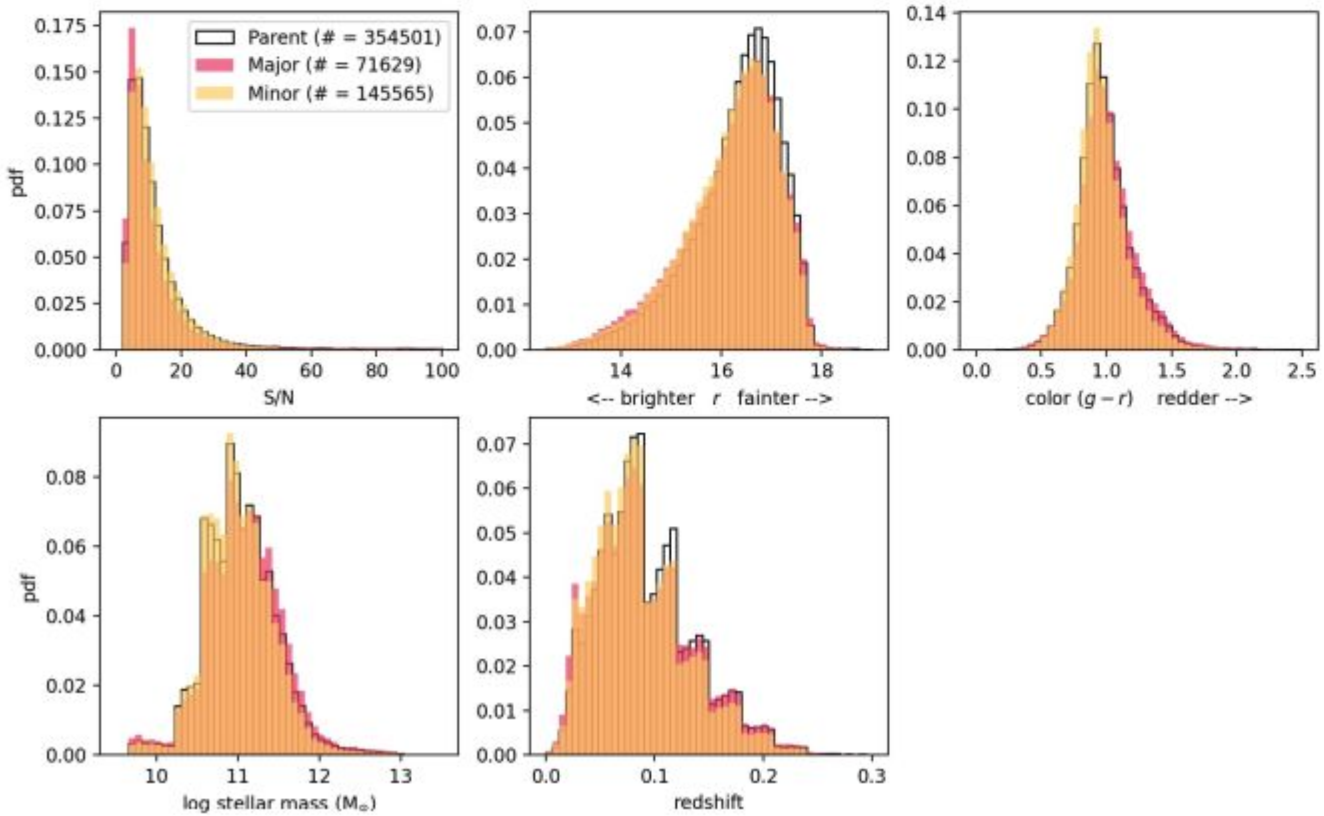
- Predictors (see below)
- Classifications for each stage and mass ratio
- Marginalized p_merg values (good for comparison)

	Predictor Values[b]								Flags[d]		
SDSS ObjID[a]	*Gini*	M_{20}	C	A	S	n	A_s	S/N[c]	low S/N	outlier predictor	segmap
1237665179521187863 (A)	0.54	-2.15	3.62	-0.04	-0.01	1.49	0.13	9.98	0	0	0
1237661852010283046 (B)	0.69	-0.96	3.59	0.22	0.01	1.32	0.78	12.49	0	0	0
1237648720718463286 (C)	0.56	-1.0	3.66	0.43	-0.16	0.58	0.89	6.4	0	0	0
1237662306186428502 (D)	0.56	-2.16	3.59	0.14	0.02	1.38	0.57	16.35	0	0	0
1237653589018018166 (E)	0.56	-2.07	3.53	0.02	0.01	1.47	0.40	14.31	0	0	0
1237654383587492073 (F)	0.58	-0.81	1.61	0.54	0.06	0.97	0.12	54.27	0	0	0

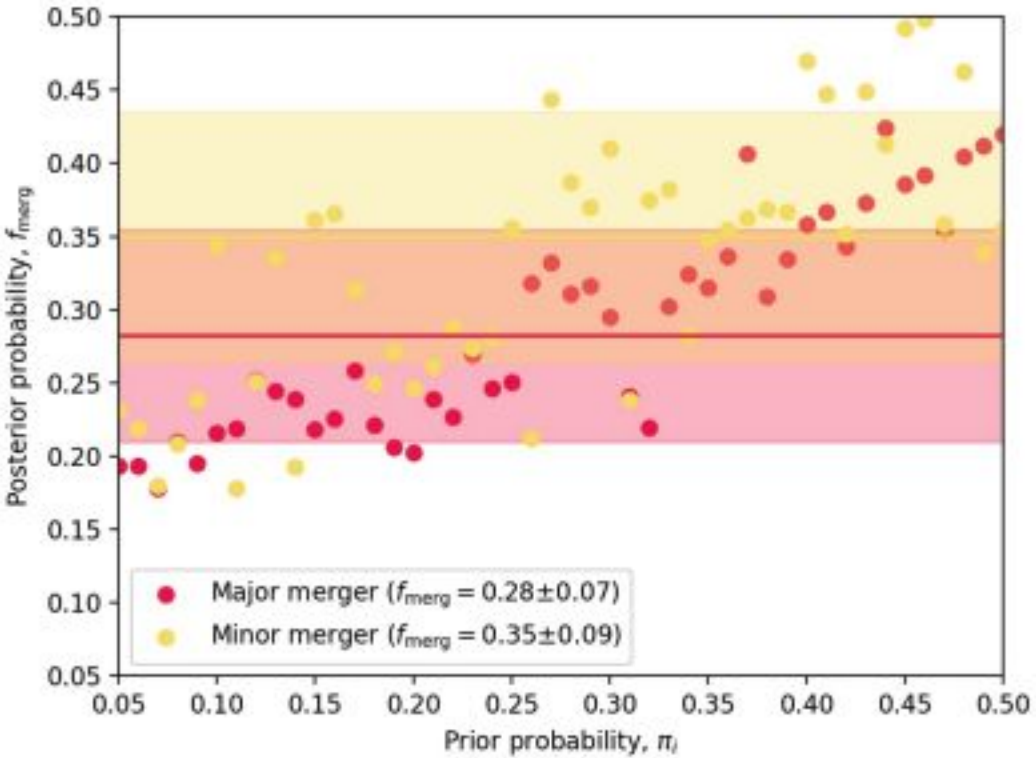
By stage

Classification	Accuracy	Precision	Recall	F1	t_{obs}
All Major Mergers	0.86	0.96	0.83	0.89	2.31
Major, pre-coalescence	0.87	0.96	0.83	0.89	2.16
Major, early stage	0.86	0.95	0.78	0.86	1.72
Major, late stage	0.94	0.97	0.84	0.90	0.83
Major, post-coalescence (0.5)	0.84	0.89	0.65	0.75	0.40
Major, post-coalescence (1.0)	0.90	0.94	0.85	0.89	1.26
All Minor Mergers	0.77	0.93	0.63	0.75	5.36
Minor, pre-coalescence	0.80	0.89	0.71	0.79	5.75
Minor, early stage	0.83	0.89	0.73	0.80	3.11
Minor, late stage	0.93	0.79	0.79	0.79	5.85
Minor, post-coalescence (0.5)	0.85	0.53	0.60	0.56	0.19
Minor, post-coalescence (1.0)	0.85	0.84	0.71	0.77	0.96

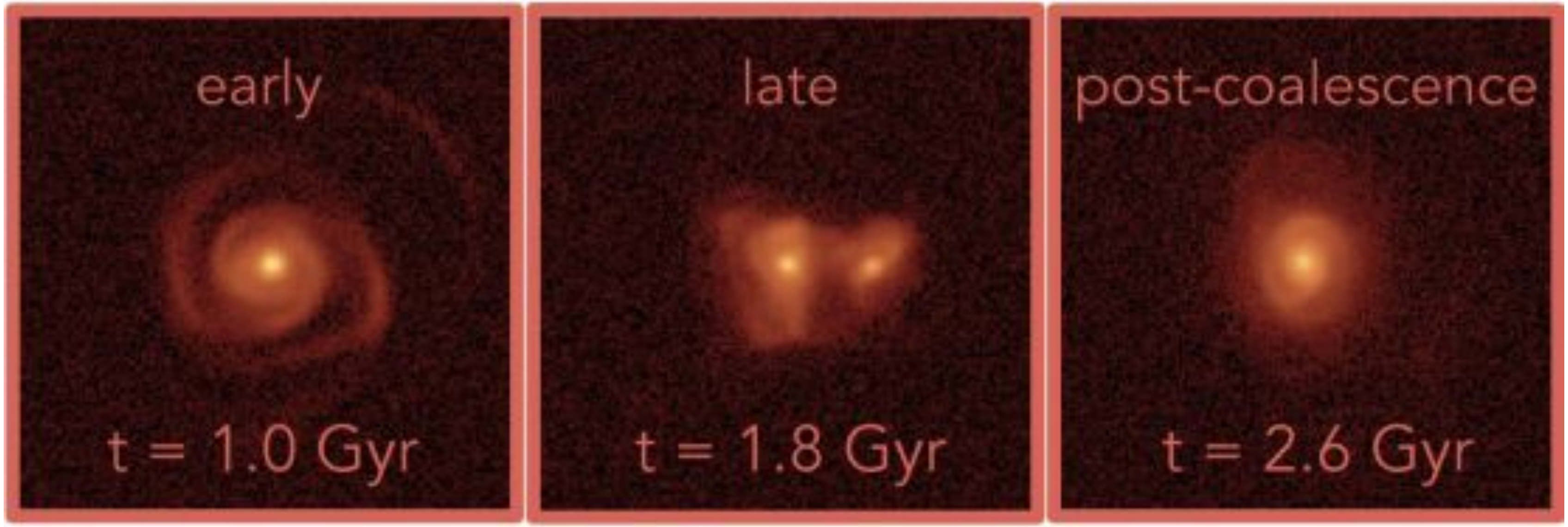
The properties of the merger sample are unbiased



Prior marginalization



There are multiple different classifications by merger stage, I calculate p_{merg} values for all of them

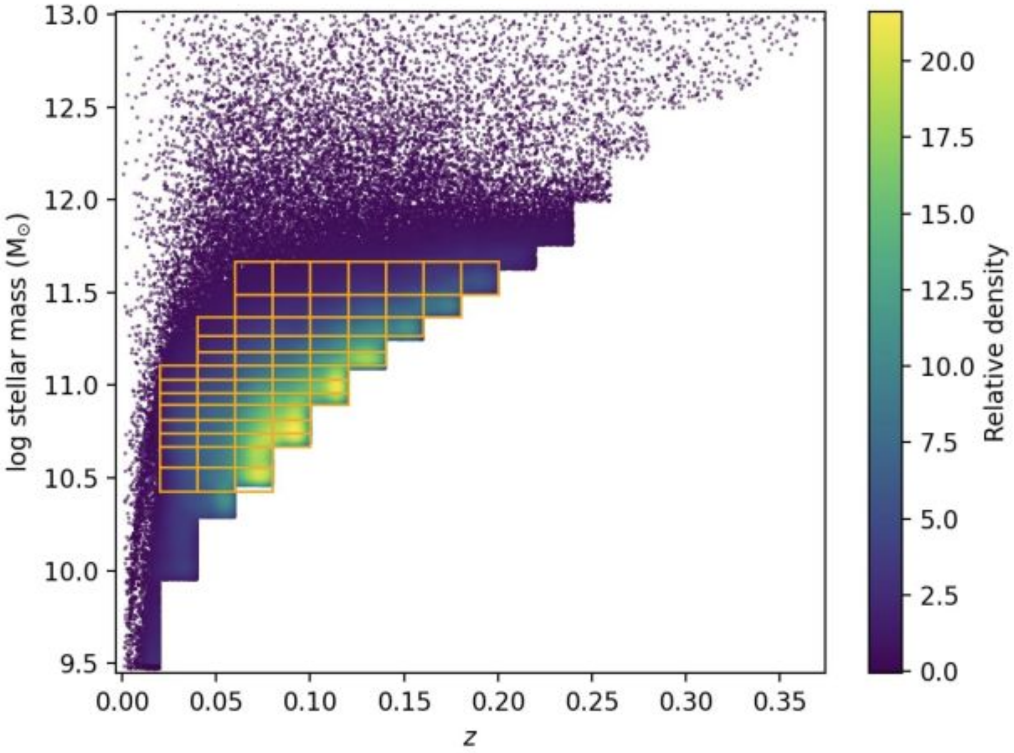


I was able to measure over bins in redshift and stellar mass

(graphic of galaxy size increasing)

1000 galaxies at least per bin

Final sample size is ~310k



Mass completeness

Next, we determine the mass completeness limit as a function of redshift using the technique from Darvish et al. (2015). For each redshift bin[2], we compute the lowest stellar mass ($M_{\rm lim}$) that could be detected for each galaxy given the magnitude limit of SDSS ($r = 17.77$): $log(M_{\rm lim}) = log(M) + 0.4 \times (r - 17.77)$, where r is the apparent (rest-frame) $r-$band magnitude of each galaxy and M is the stellar mass. The mass completeness limit at each redshift bin is the mass at which 95% of the limiting masses are below the mass completeness limit, meaning that only 5% of galaxies would be missed in the lowest mass end of the mass function.

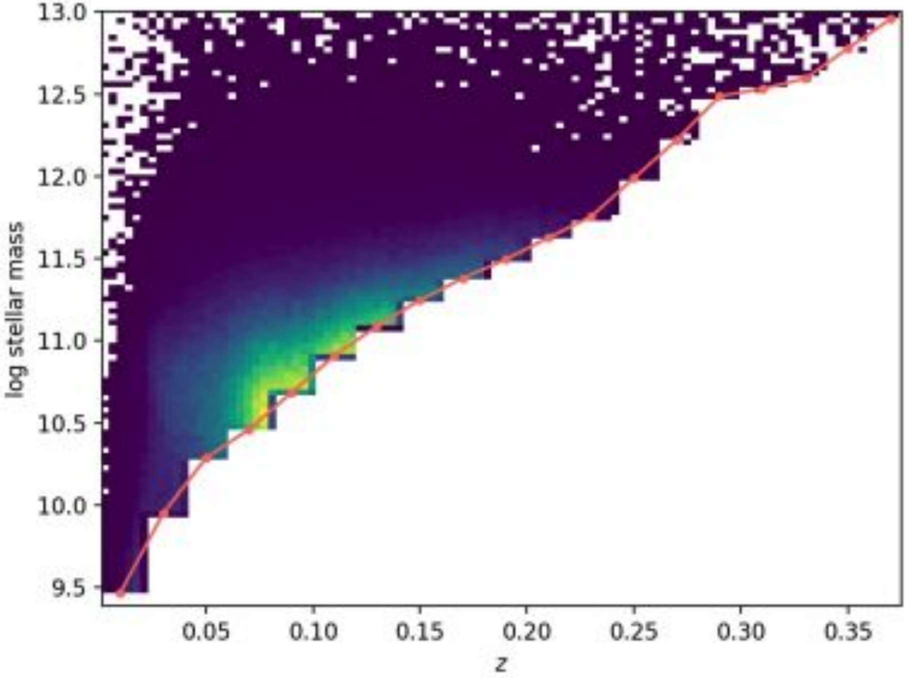
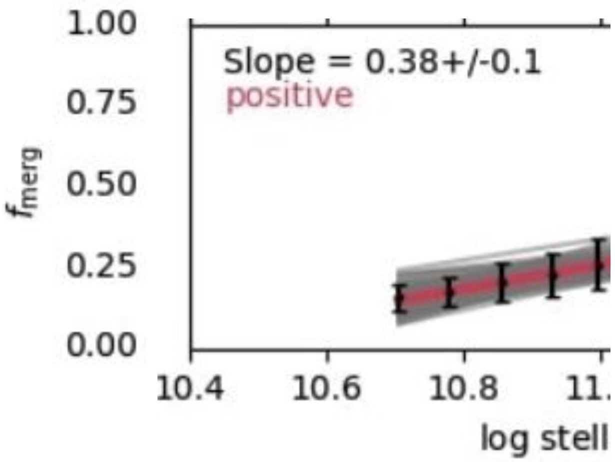
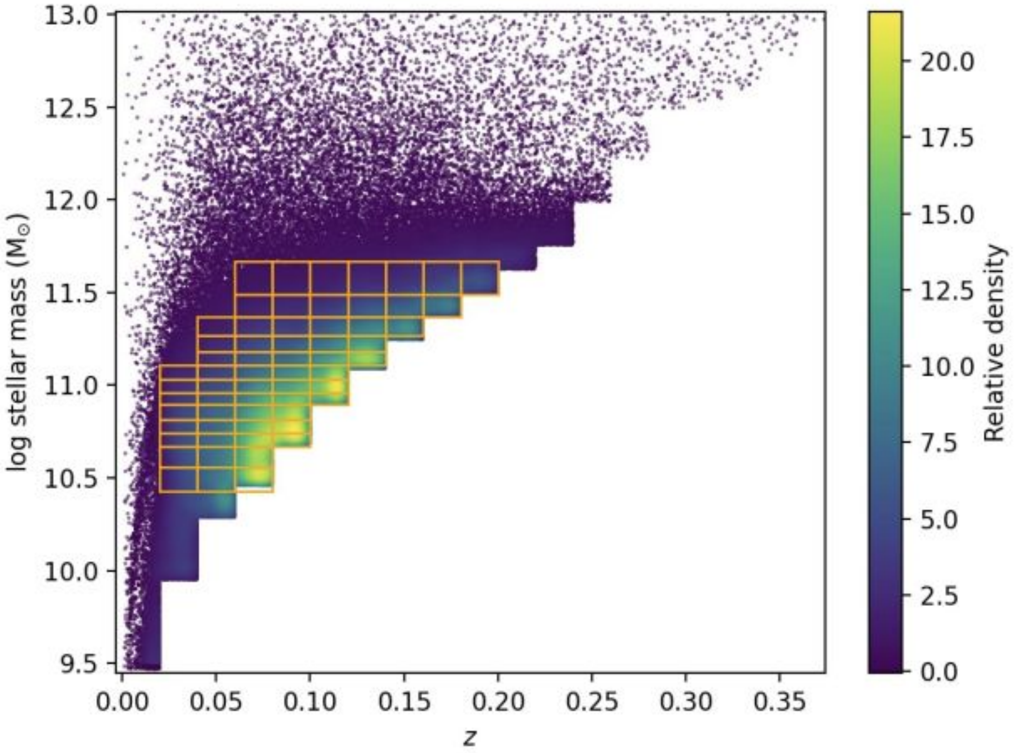


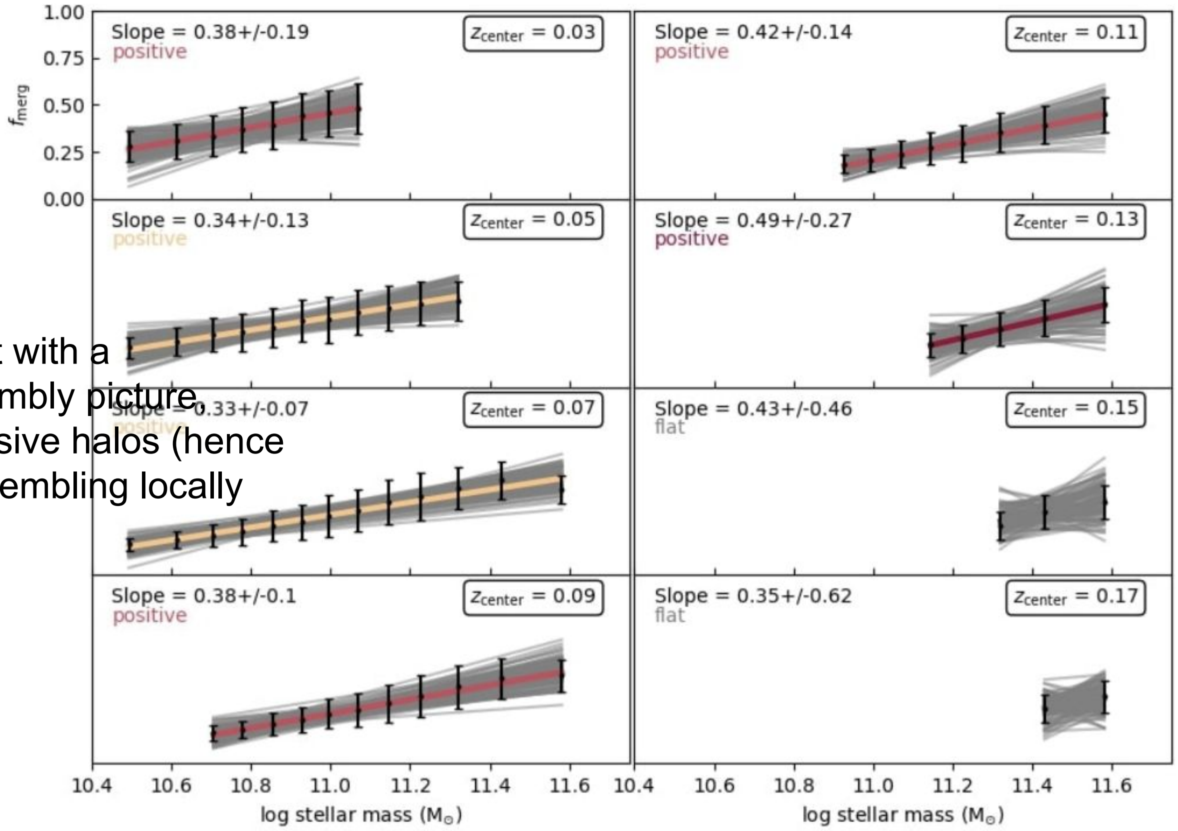
Figure 5. Mass completeness as a function of redshift for redshift bins with spacing $\Delta z = 0.02$. For each redshift bin, we determine the 95% completeness limit (pink line) and eliminate all galaxies below this point. For the distribution of masses at each redshift bin, see Appendix A.

I measure the merger fraction for every redshift and mass bin and iteratively fit





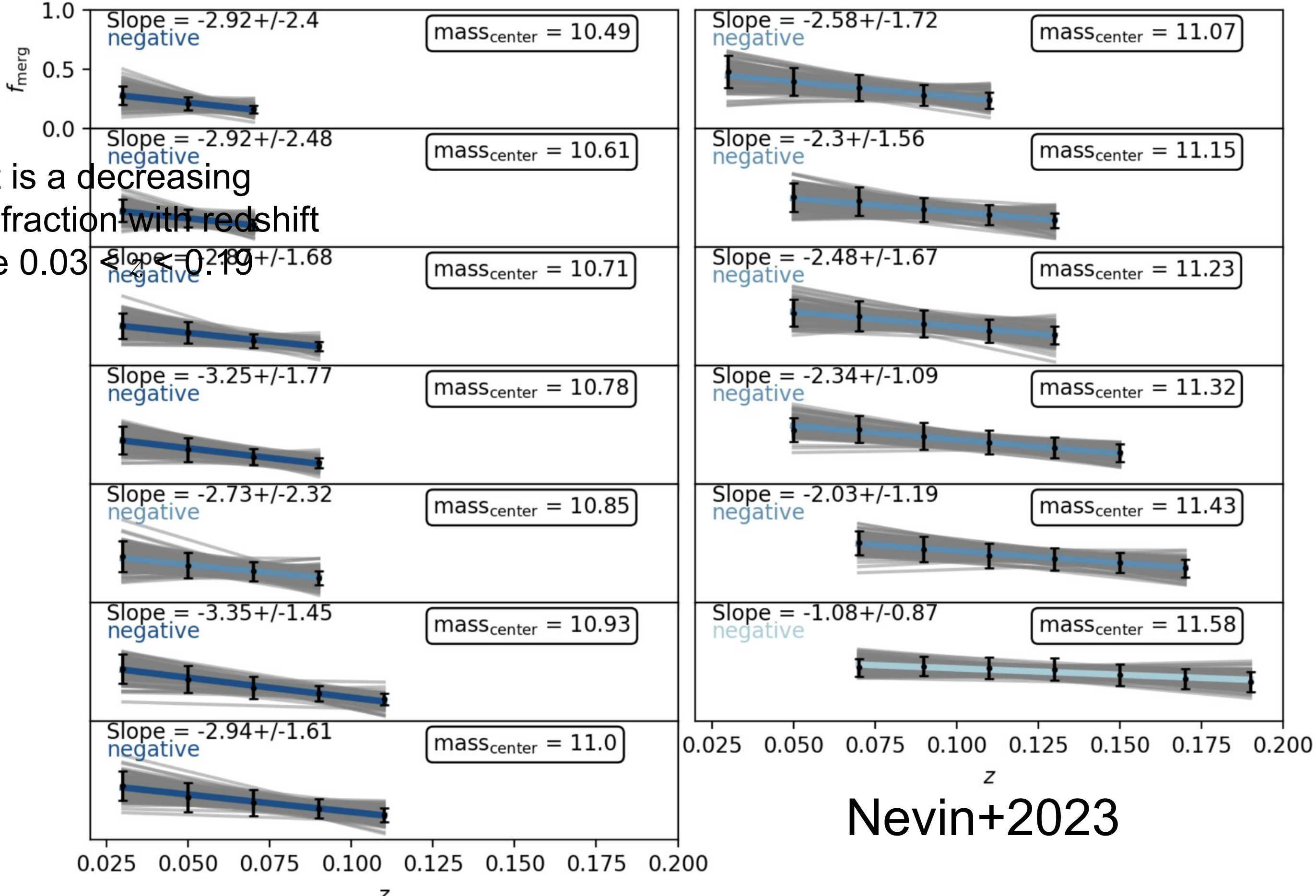
The major merger fraction increases with increasing redshift



This is consistent with a hierarchical assembly picture, where more massive halos (hence galaxies) are assembling locally

The major merger fraction decreases with redshift

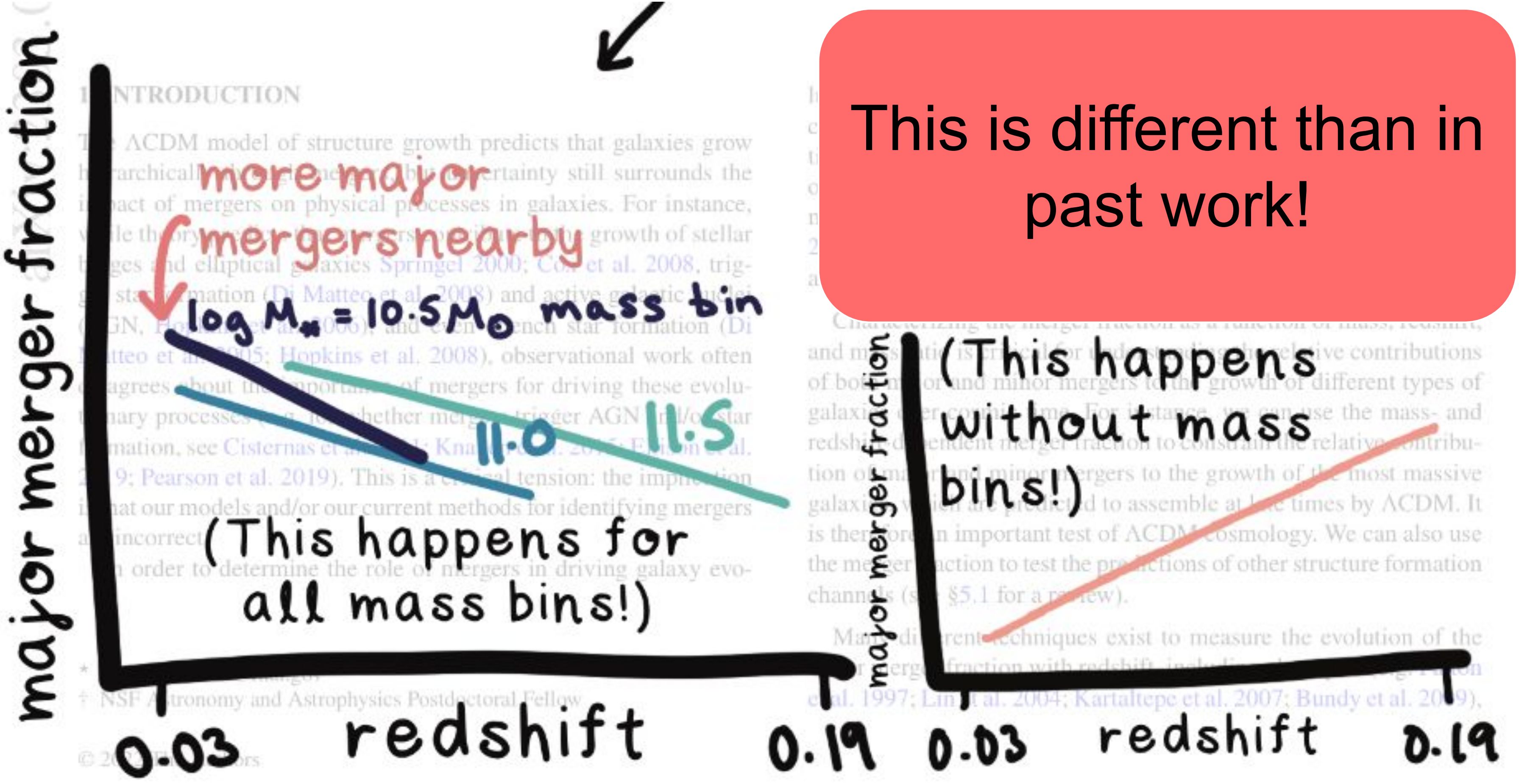
Our key result is a decreasing major merger fraction with redshift over the range $0.03 < z < 0.19$



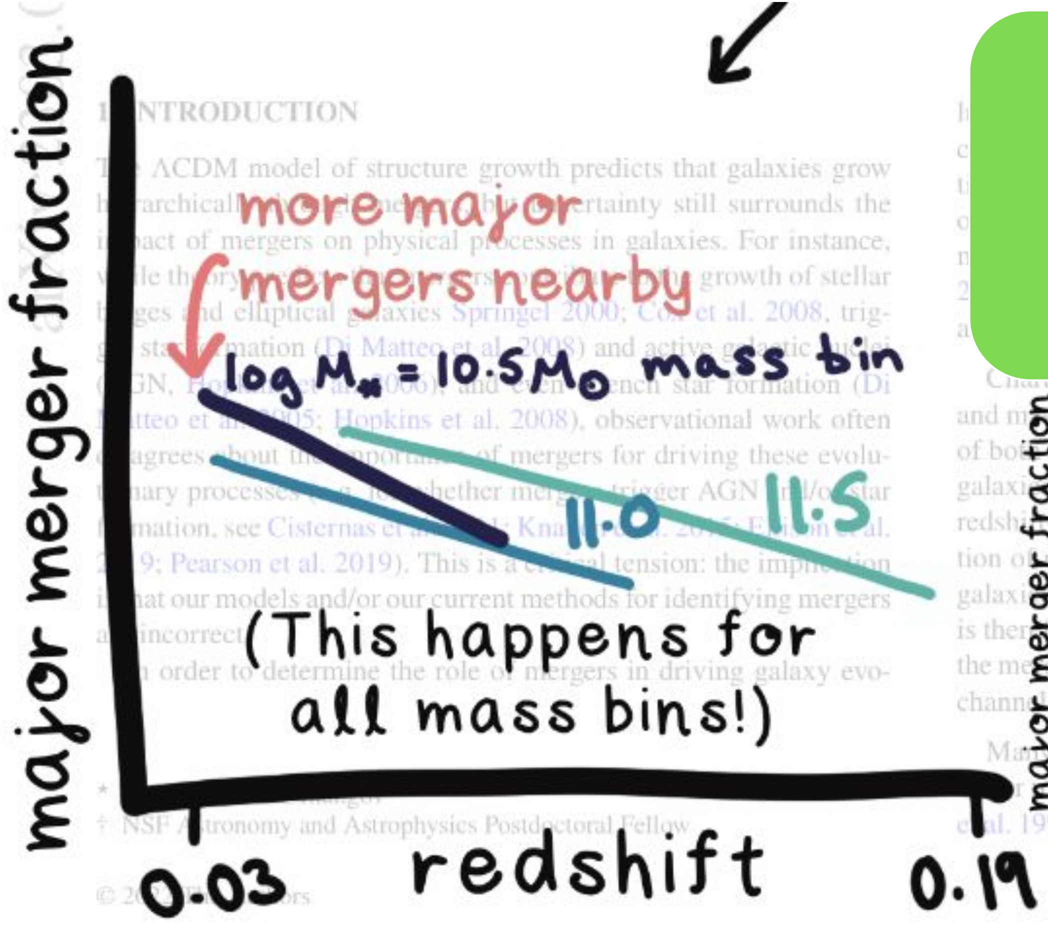
Nevin+2023

This is a surprising result!

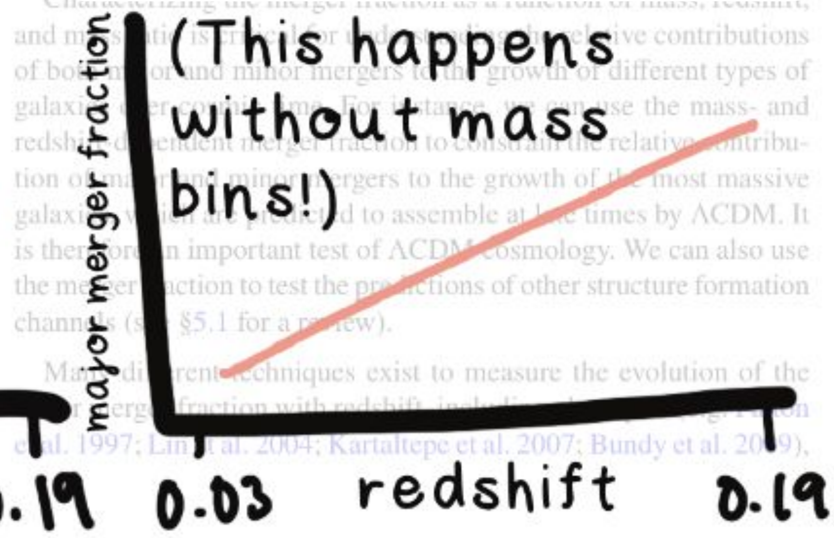
This is different than in past work!



This is a surprising result! **What is going on here?**



Do you have any ideas why this is happening?



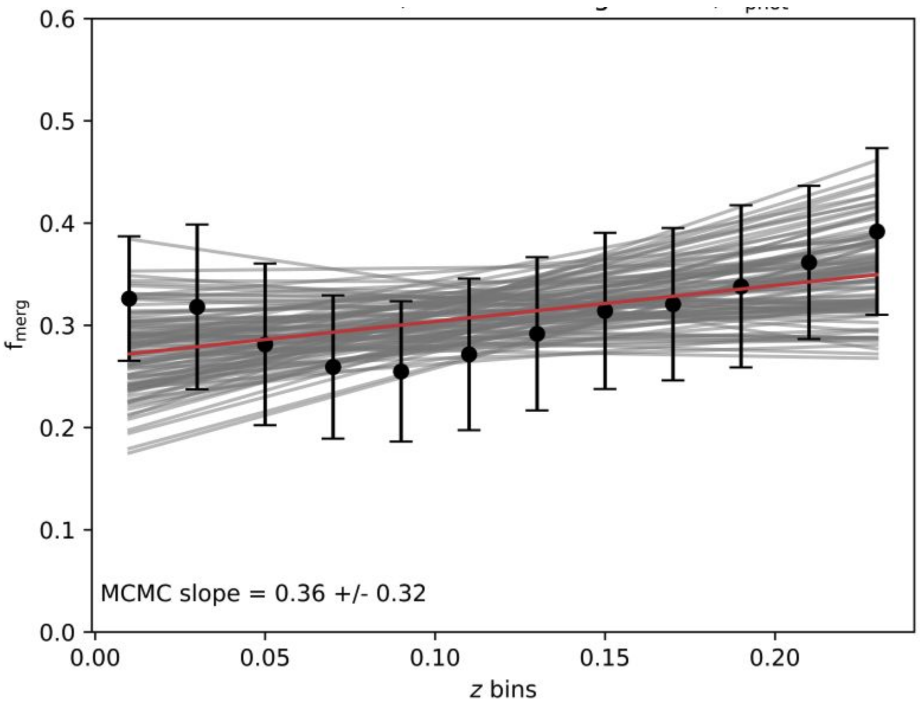
The importance of confounds in statistical analysis

Statistical confound: An (annoying) variable that influences both the independent and the dependent variable, creating a spurious correlation

Statistical Rethinking by Richard McElreath really helped me out ^

A statistical confound with mass drives this behavior

- Mass increases with redshift in depth limited surveys like SDSS
- Merger fraction increases with mass
- t/f merger fraction appears to increase with redshift



Lessons learned

- Trust myself
- Sanity checks to figure out what’s happening behind the scenes
- The importance of reproducing past results

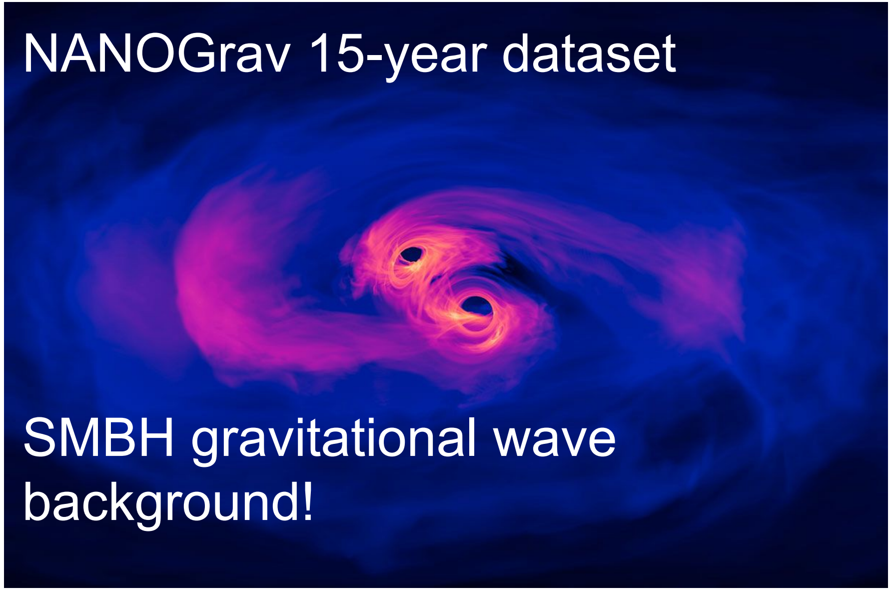
Other statistical confounds? No

major merger fraction increases with B/T and $g - r$ mostly for higher mass galaxies

Trend with B/T and color is different than it being a confound

So what’s actually happening here?

Merger fraction → merger rate as a function of galaxy and merger properties



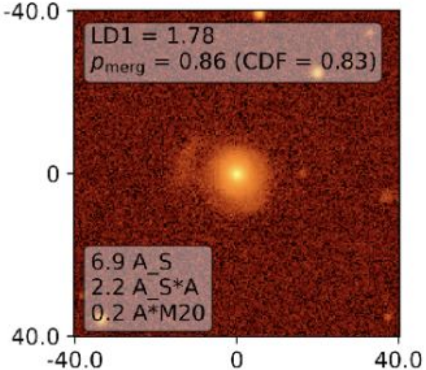
Joe Simon

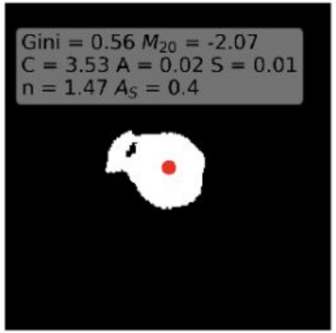
Julie Comerford

Simon+2023 in prep

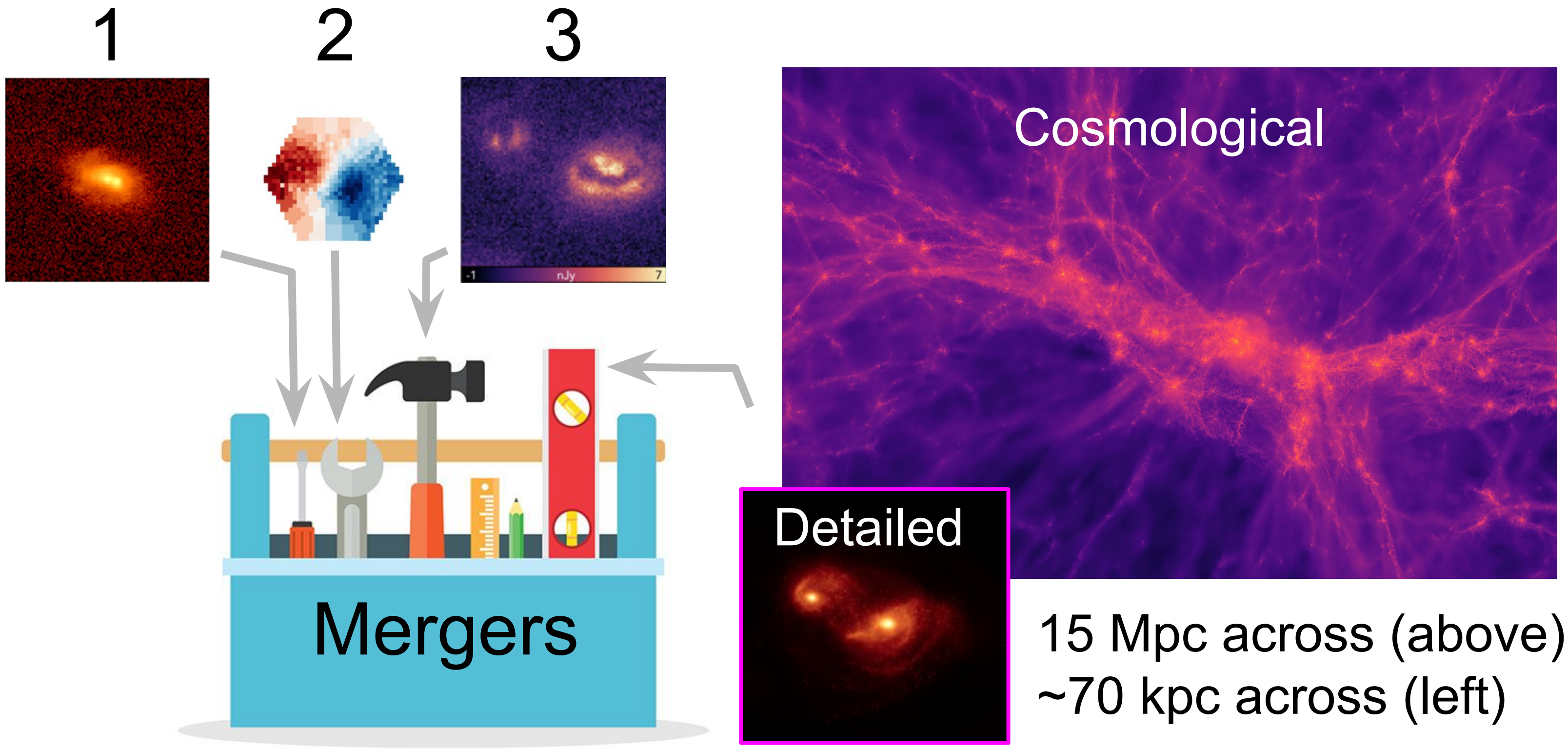
My merger catalog has enabled multiple studies into the properties of merging galaxies and the AGN-merger connection:

Comerford+2023; An excess of AGNs triggered by galaxy mergers in MaNGA galaxies of mass 10^{11} $M_\odot$
Hernández-Toledo+2023; MaNGA AGN have an enhanced merger fraction
Negus+2023; Coronal line MaNGA galaxies



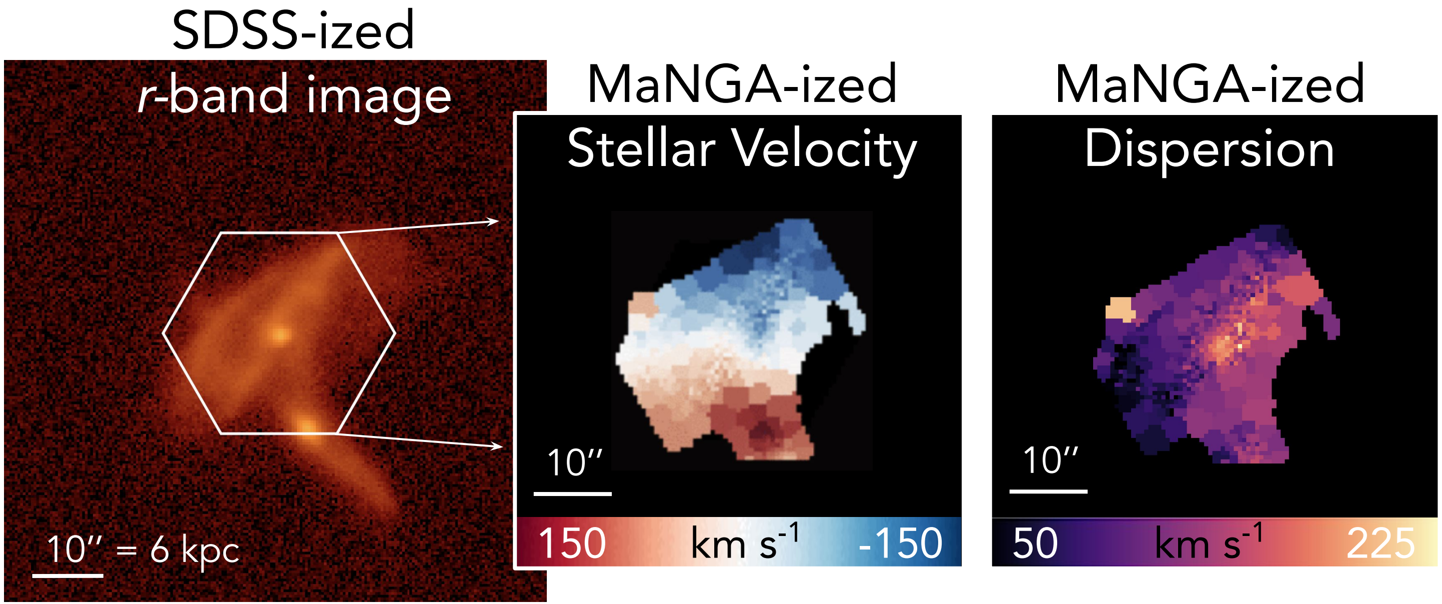


My work approaches better identifying mergers with the help of detailed hydro and cosmological simulations



15 Mpc across (above)
~70 kpc across (left)

I create mock stellar kinematic maps to match the specifications of MaNGA integral field spectroscopy



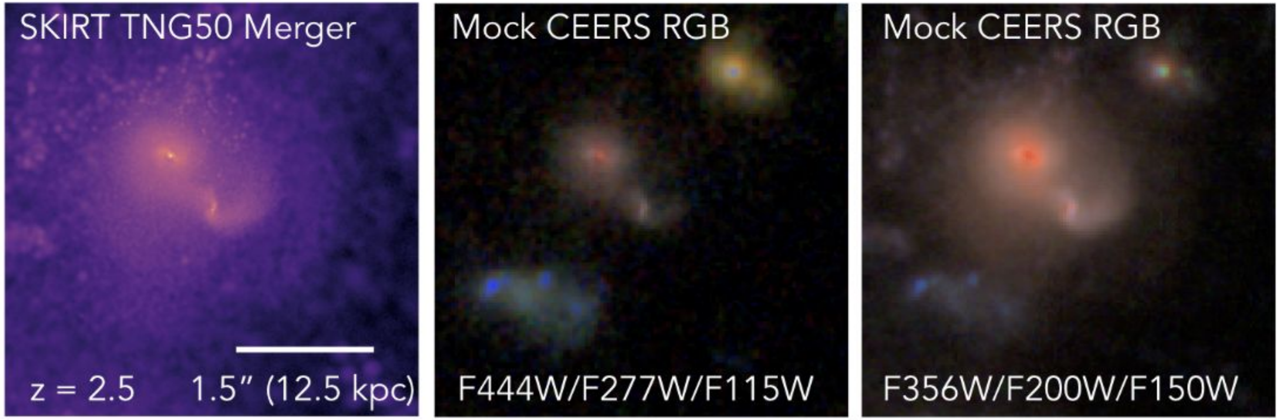
Nevin+2019

Nevin+2021

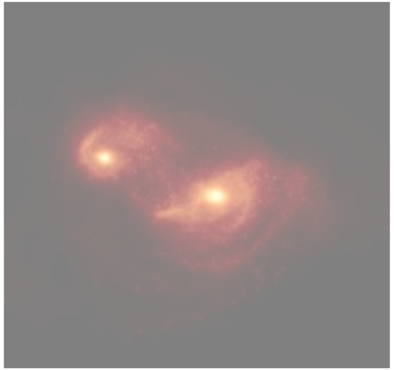
How do we best identify high redshift merging galaxies?:

Expanding the toolkit to include *HST* Candels and *JWST* NIRCam imaging

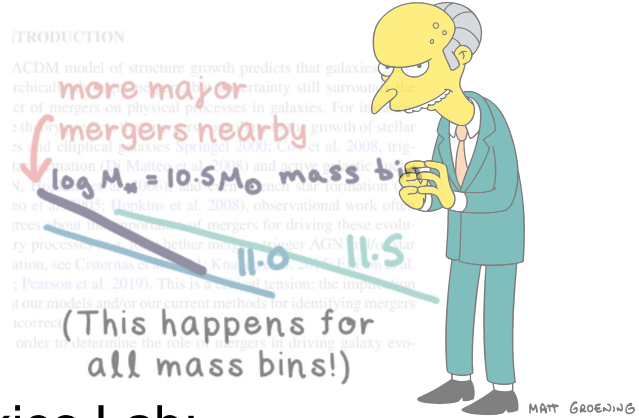
Aimee Schechter



Mergers

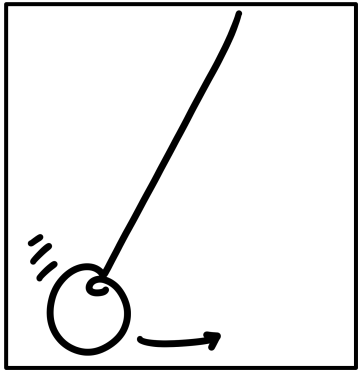


Adventures in statistical confounds



DeepSkies Lab:

Benchmark data for ML, uncertainty, and fancy Bayesian inference, oh my!



I wanted to come to Fermilab and work with the Deepskies crew because:

- Ethical and careful AI research
- Software expertise
- Cosmology and survey science
- Galaxies and spectra

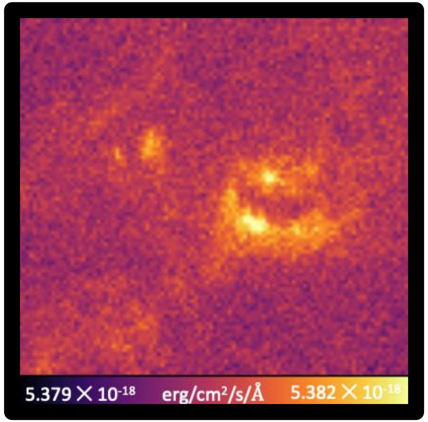
DEEP SKIES

Bringing Artificial Intelligence to Astrophysics

images

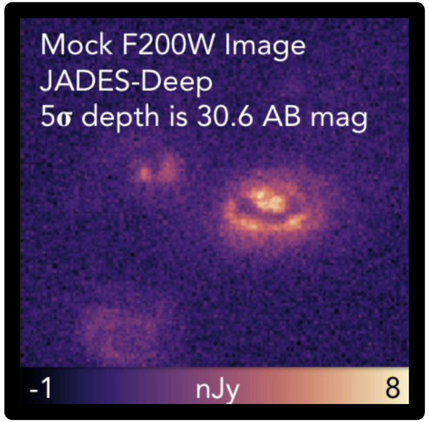
Aimee Schechter

HST F814W



Schechter+2024

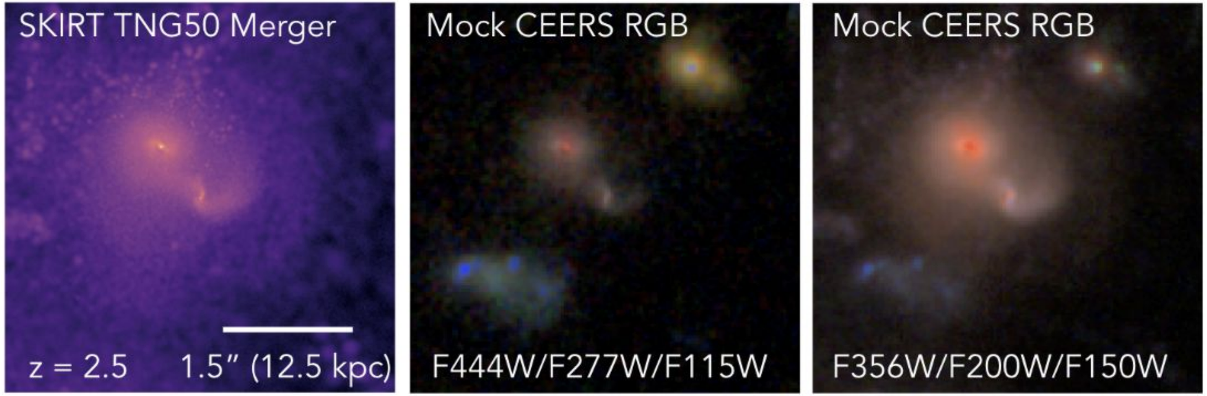
JWST F200W



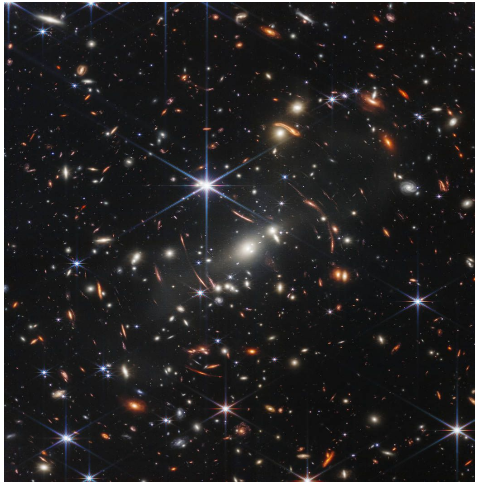
Nevin+2024

Carefully incorporating domain adaptation is necessary and interesting

Simulated galaxies

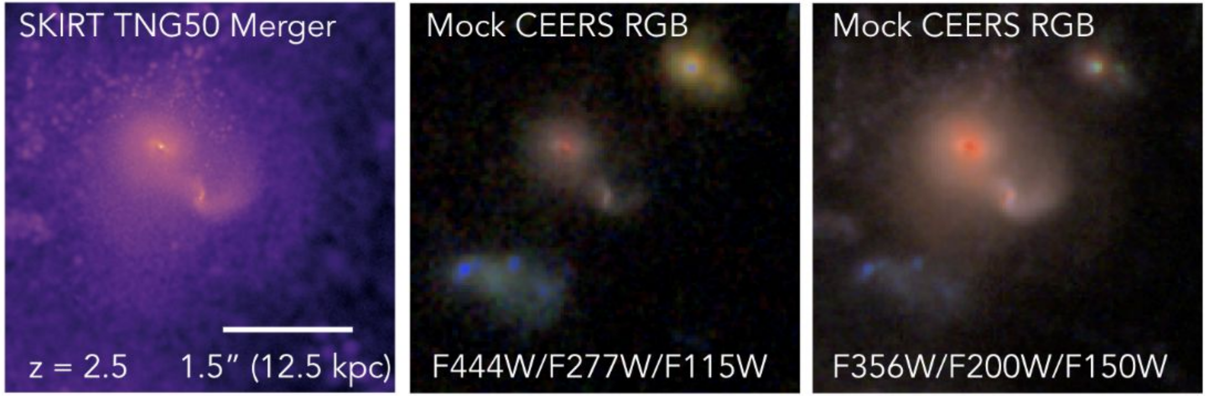


Real *JWST* galaxies
(SMACS 0723)

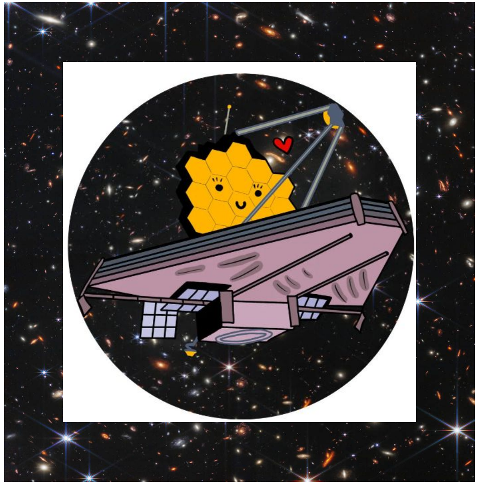


Carefully incorporating domain adaptation is necessary and interesting

Simulated galaxies

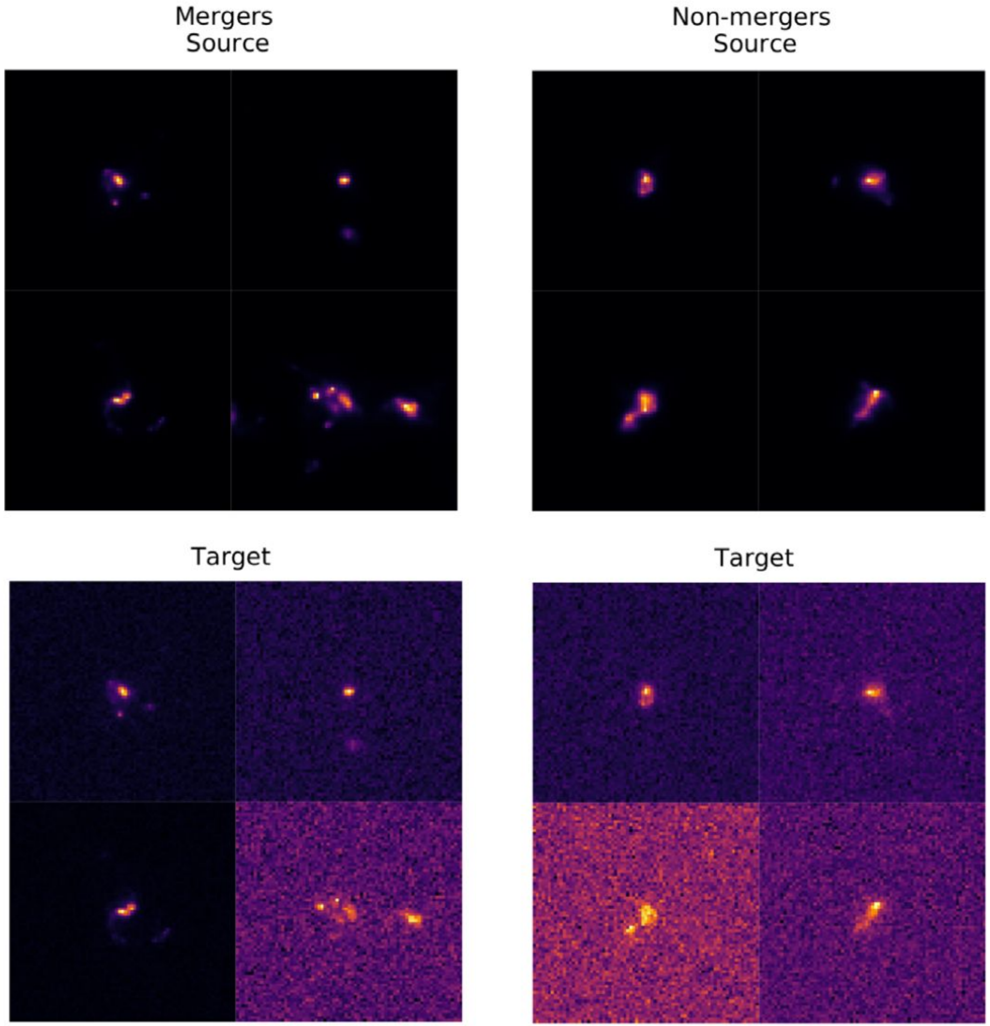


Real *JWST* galaxies
(SMACS 0723)



We are working with Alex Ćiprijanović, who is a domain adaptation expert

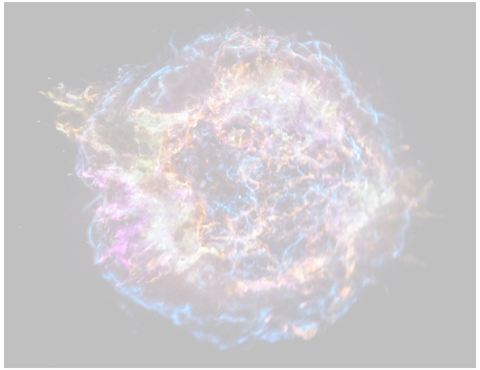
Ćiprijanović+ 2020a,2021



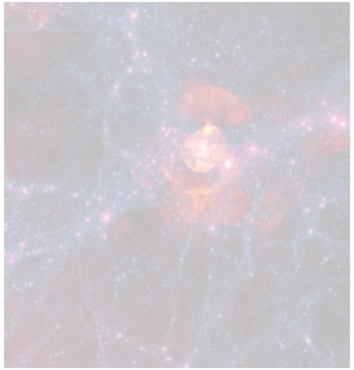
Active Galactic Nuclei

Mergers

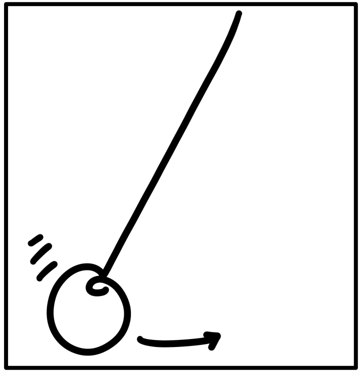
Chandra X-ray



Illustris



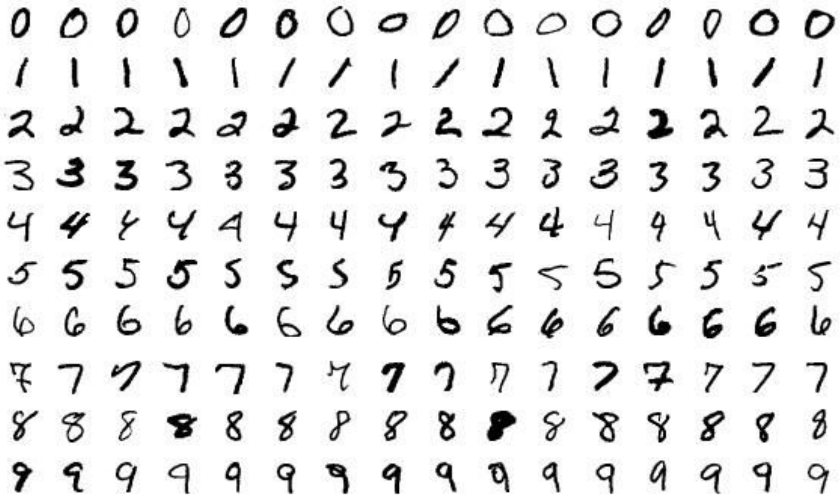
Benchmark

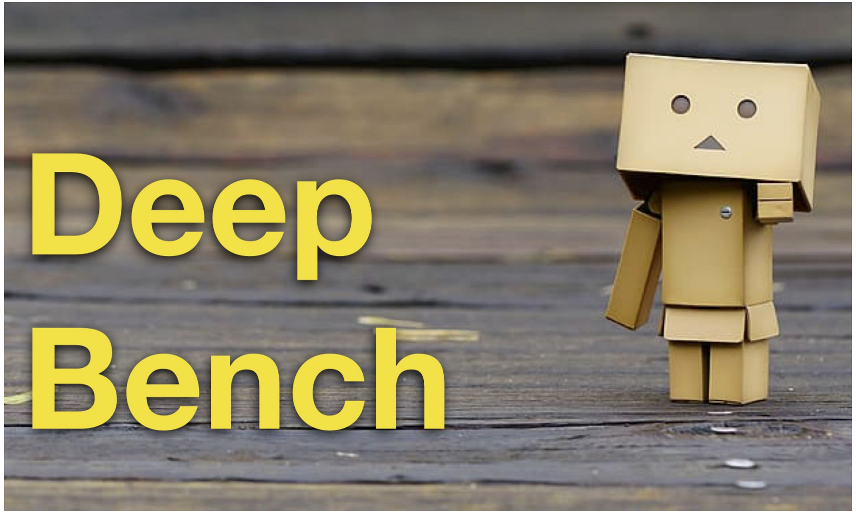


UQ

Hierarchical Inference

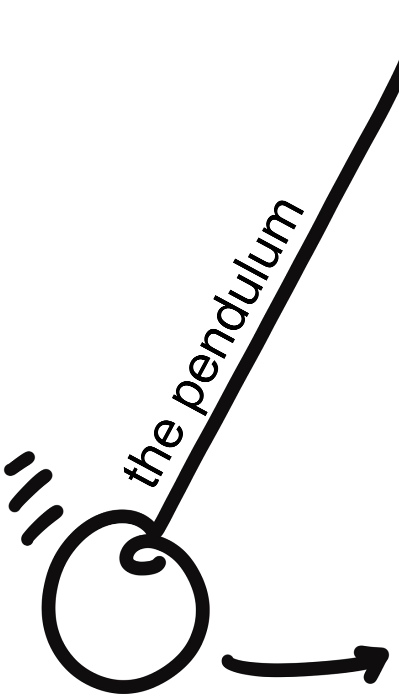
DeepBench: Fine-grained control for simulations for neural inference





Control over noise
Ability to propagate noise
Its dynamic, create new examples

We are using simple benchmark datasets (like the pendulum) to build complex inference tools

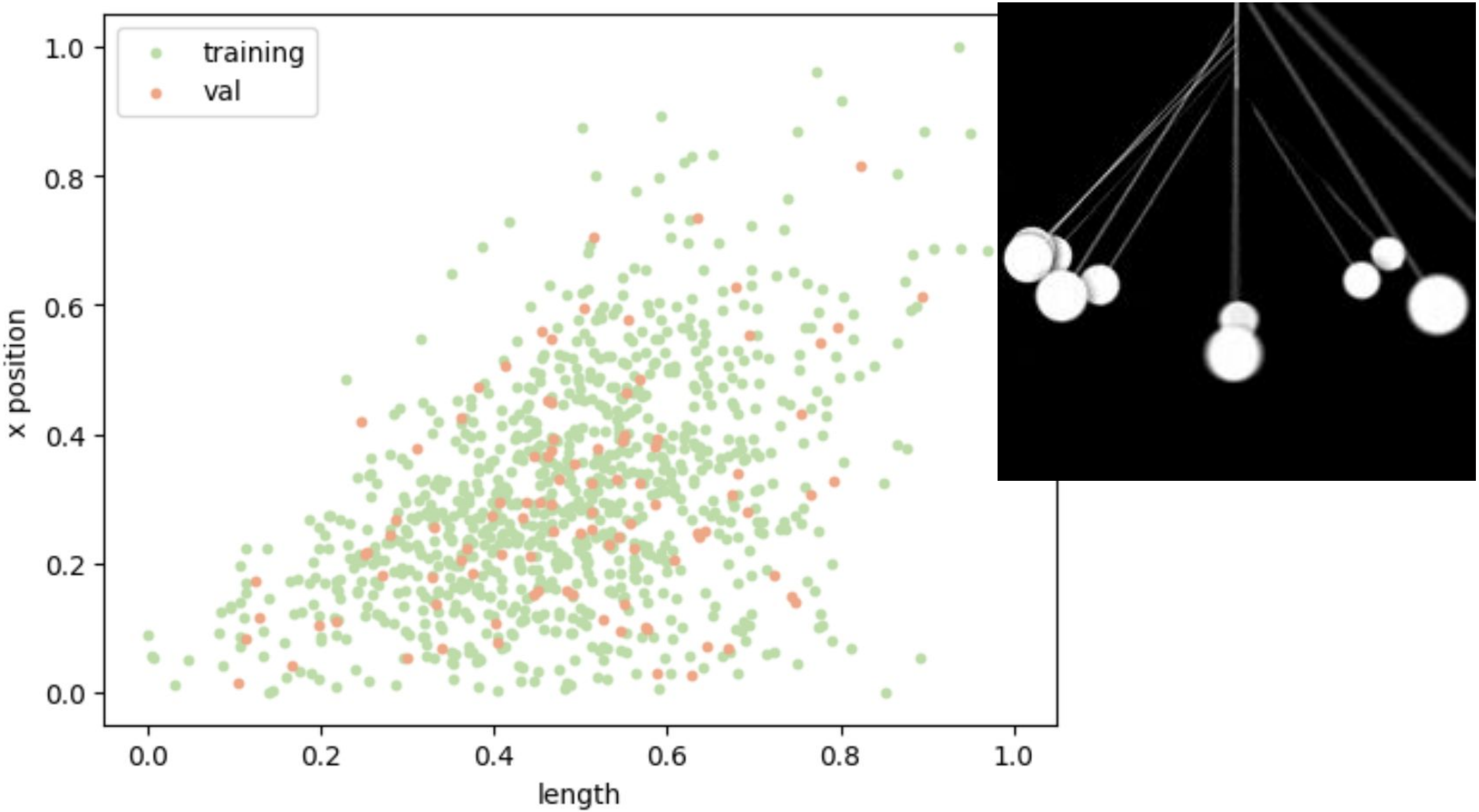


Things we'd like to infer about a pendulum:

- starting angle
- mass
- length

Data:

position and momentum as a function of time (with added noise)


training
val
x position
length
0.0
0.2
0.4
0.6
0.8
1.0

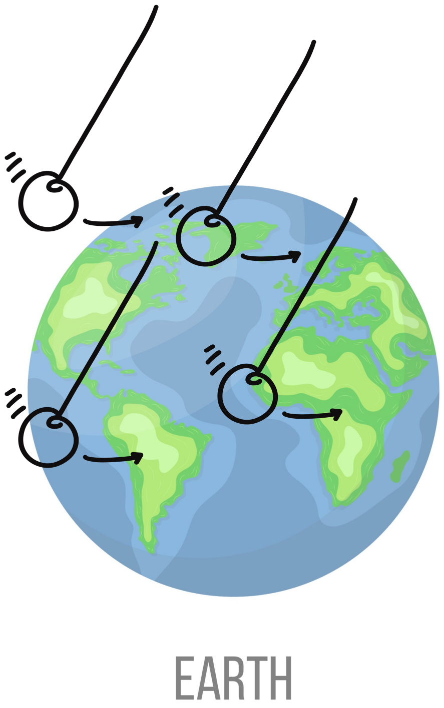
Why hierarchical analysis?

Astro applications:

- Many exoplanets
- Many gravitational lenses

When you want to infer individual properties but also global properties and have both inform one another

The pendulum as a laboratory to test these methods



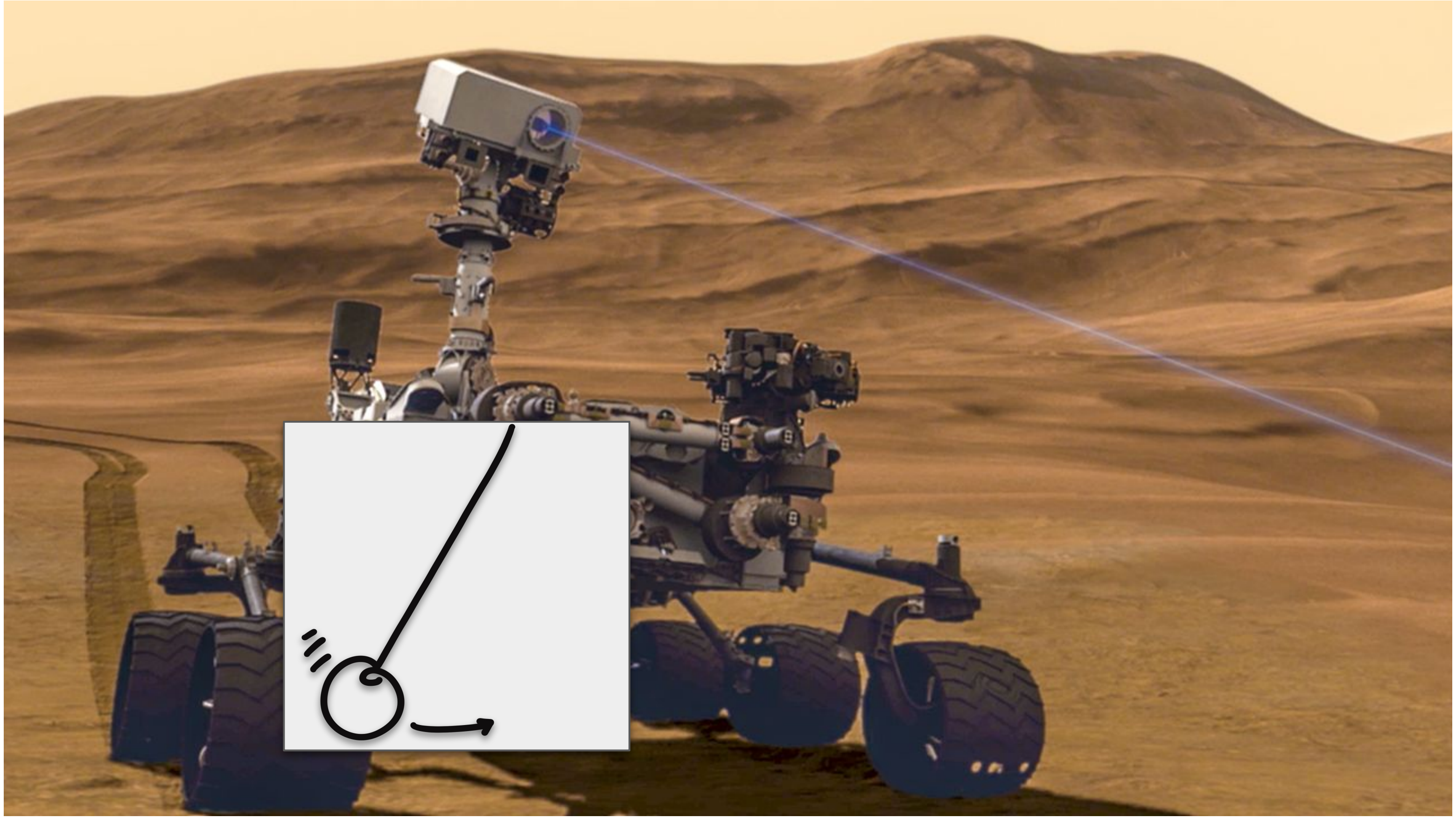
Things we'd like to infer about one pendulum:

- starting angle
- mass
- length

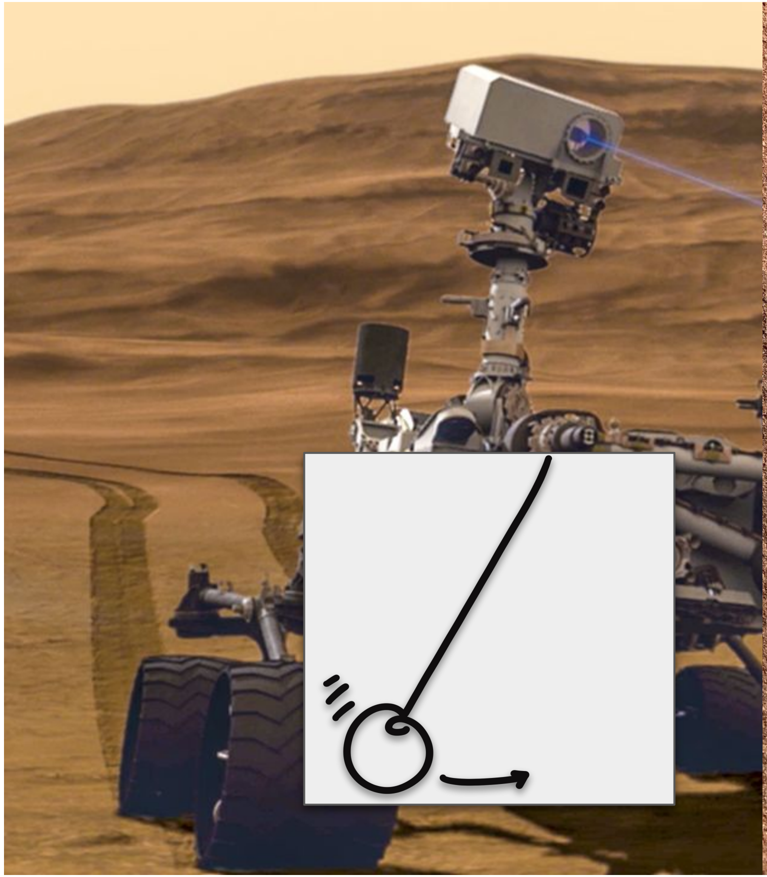
Things we'd like to infer using the ensemble of pendulums:

- acceleration due to gravity (a_g)

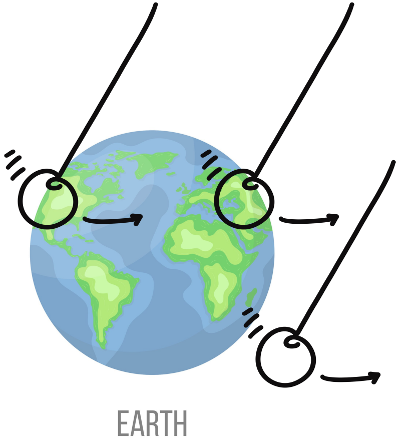
Meanwhile, on Mars…

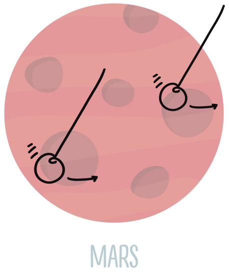


Meanwhile, on Mars…



There are many experiments with different conditions in different groups = hierarchical Bayesian inference





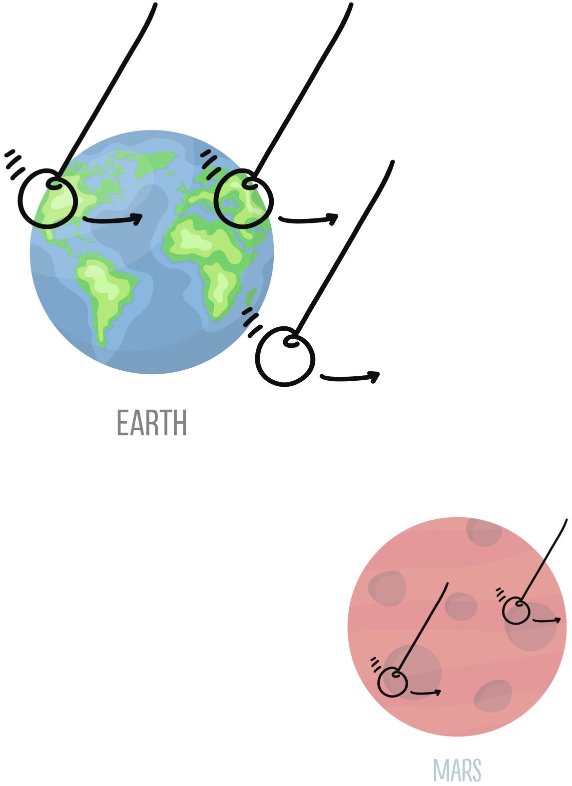
Things we'd like to infer about one pendulum:

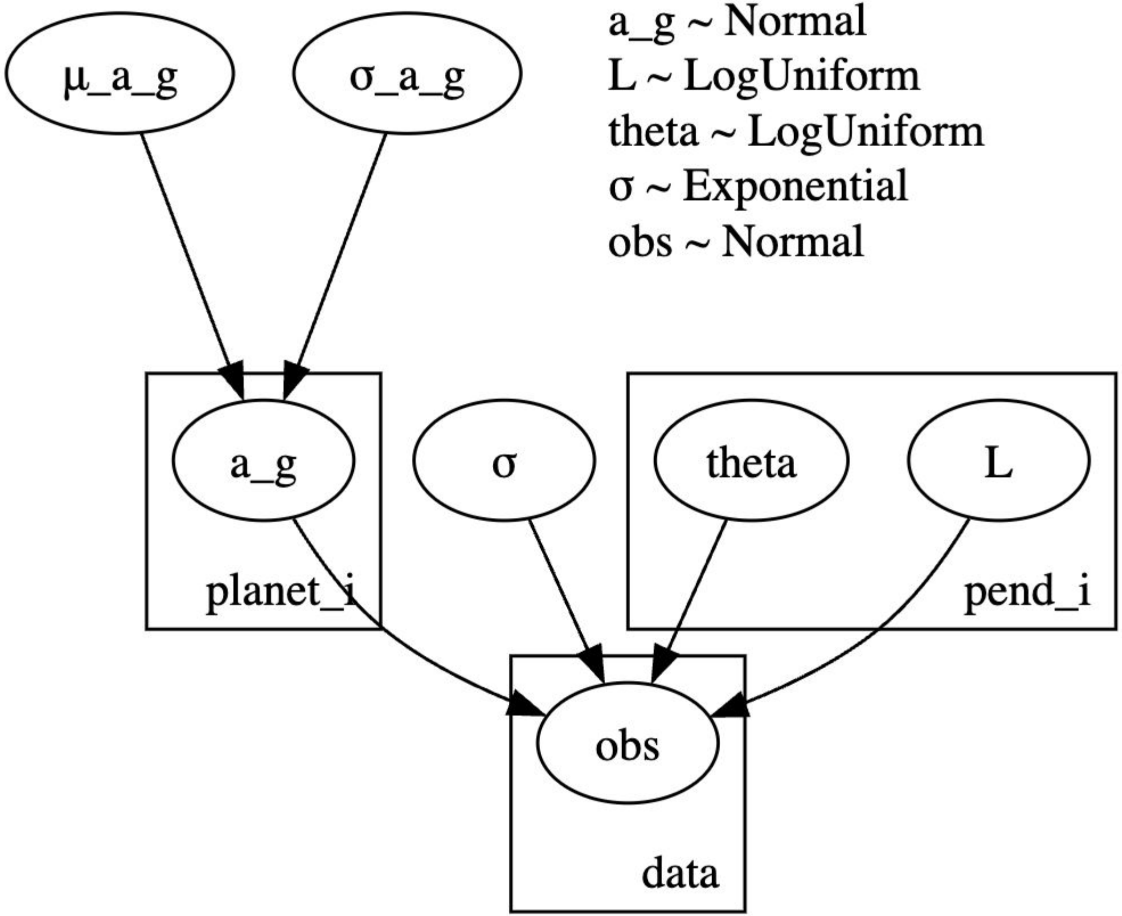
- starting angle
- mass
- length

Things we'd like to infer using the ensemble of pendulums:

- acceleration due to gravity (a_g)
- Universal gravitational constant (G)

μ_a_g ~ LogUniform
σ_a_g ~ LogUniform
a_g ~ Normal
L ~ LogUniform
theta ~ LogUniform
σ ~ Exponential
obs ~ Normal





pyro-ppl/**numpyro**

Probabilistic programming with NumPy powered by JAX for autograd and JIT compilation to GPU/TPU/CPU.

mackelab/**sbi**

Simulation-based inference toolkit

40 Contributors

67 Used by

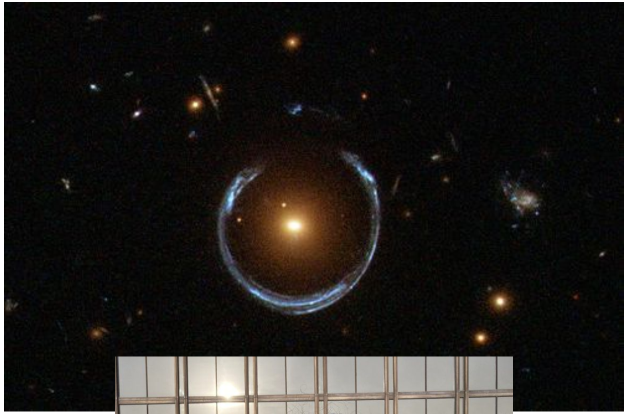
8 Discussions

437 Stars

106 Forks

Simulation based inference does not require a likelihood!

This system is essential for preparing a methodology for cosmological inference



Things we'd like to infer about one individual image:

- Lens parameters (ie Einstein radius)

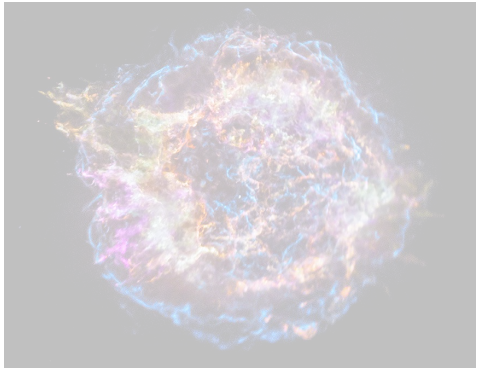
Things we'd like to infer using the ensemble of lenses:

- Cosmological parameters (w_0)

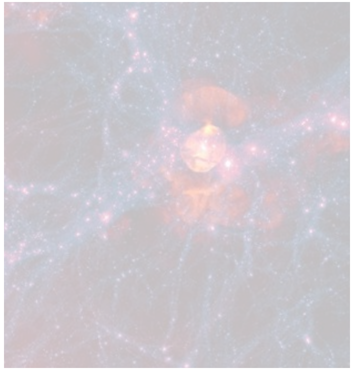
Active Galactic Nuclei

Mergers

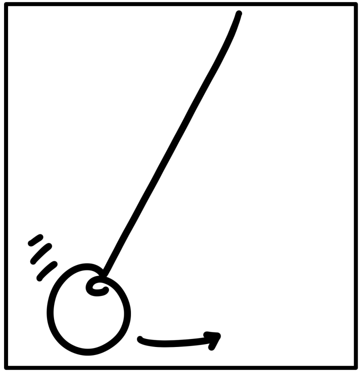
Chandra X-ray



Illustris



Benchmark

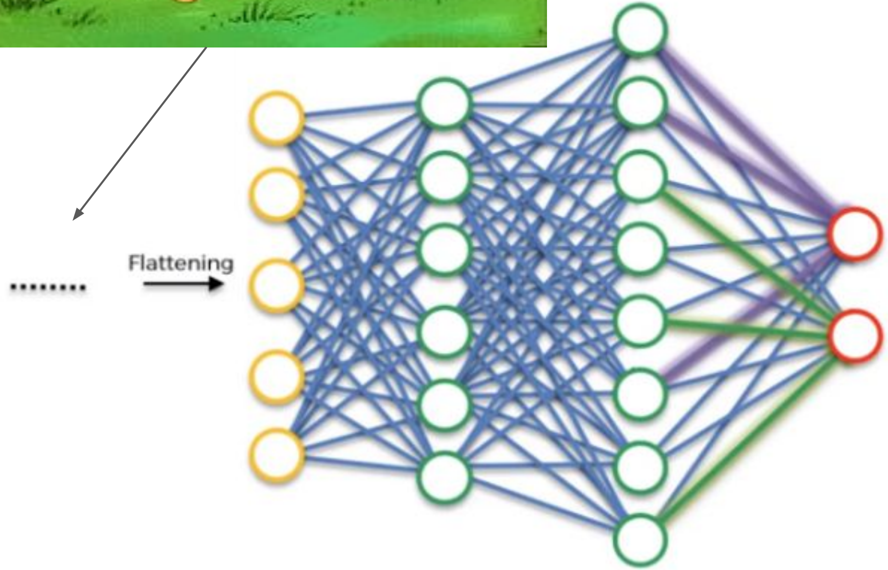


UQ

Hierarchical Inference

Most neural networks are deterministic

In mathematics, computer science and physics, a deterministic system is a system in which no randomness is involved in the development of future states of the system. A deterministic model will thus always produce the same output from a given starting condition or initial state.



$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

Dog $\longrightarrow z_1 \longrightarrow$ 0.5

Cat $\longrightarrow z_2 \longrightarrow$ 0.5

There are different types of uncertainty to consider in machine learning

data uncertainty versus model uncertainty

Using deepbench, does the expected error match that estimated using various ML methods?

Bayesian inference (sampling): hierarchical and non-h

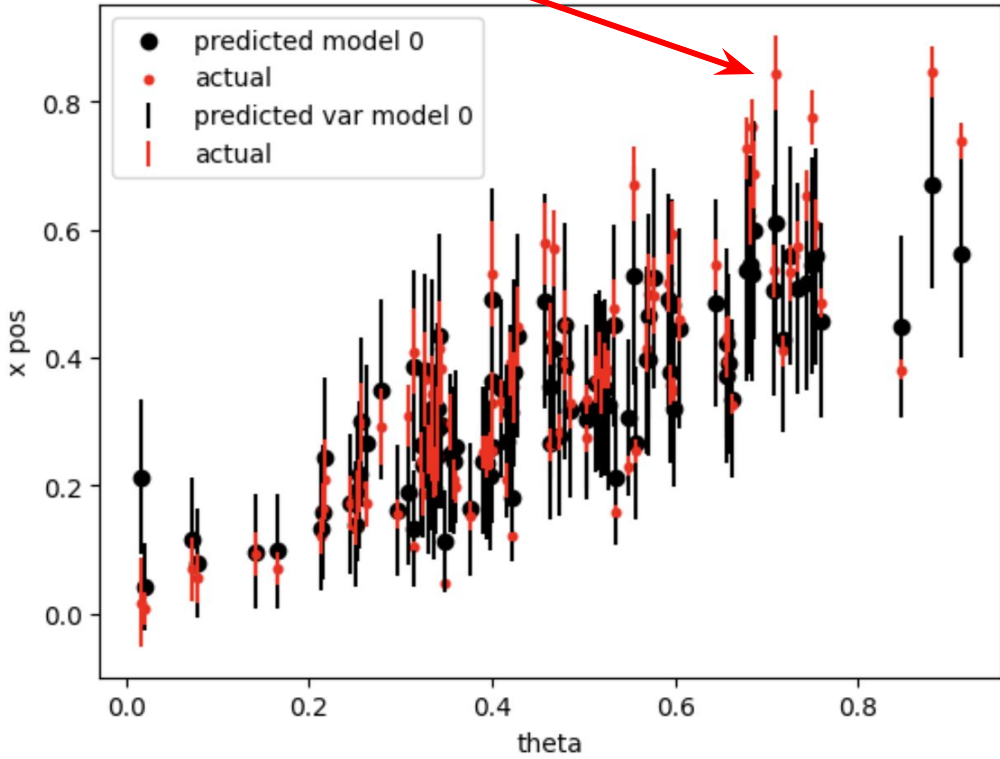
Simulation based inference: hierarchical and non-h

Bayesian inference (sampling): hierarchical and non-h

Simulation based inference: hierarchical and non-h

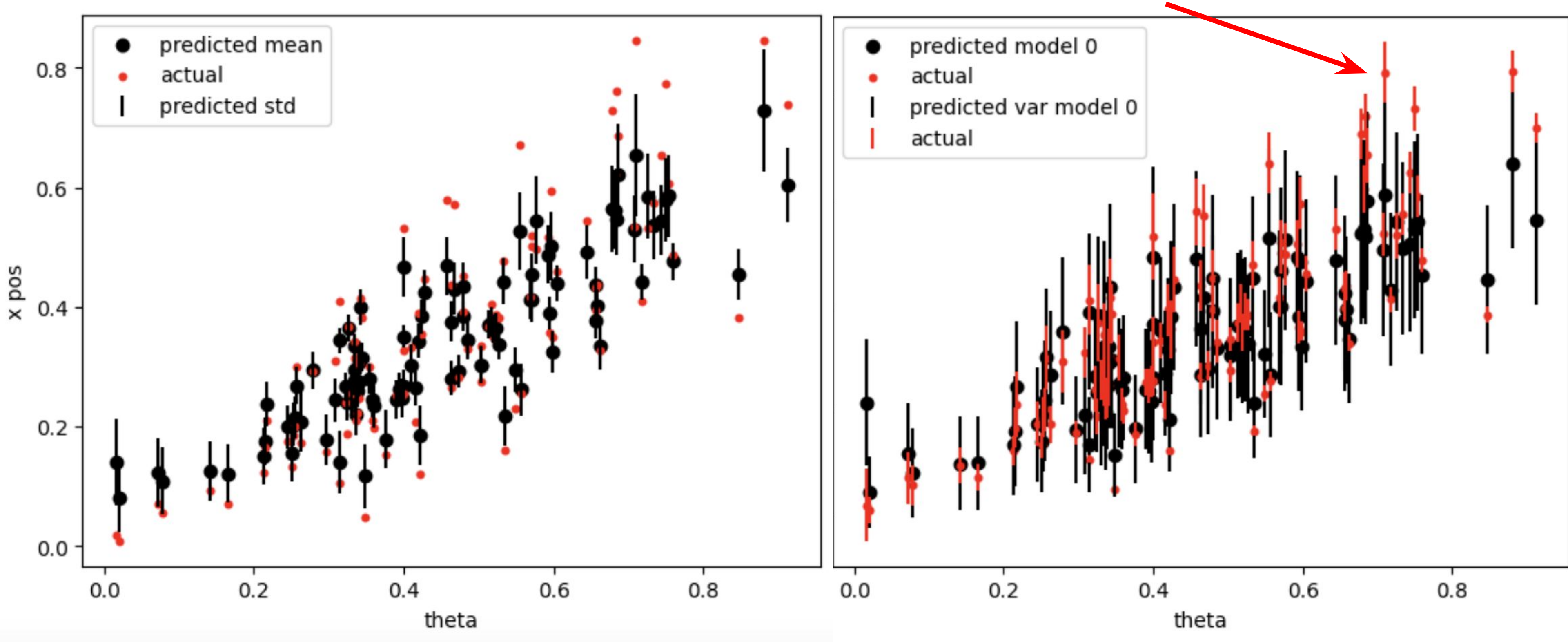
non-Bayesian neural networks like deep ensembles →

Analytical expectation of data uncertainty

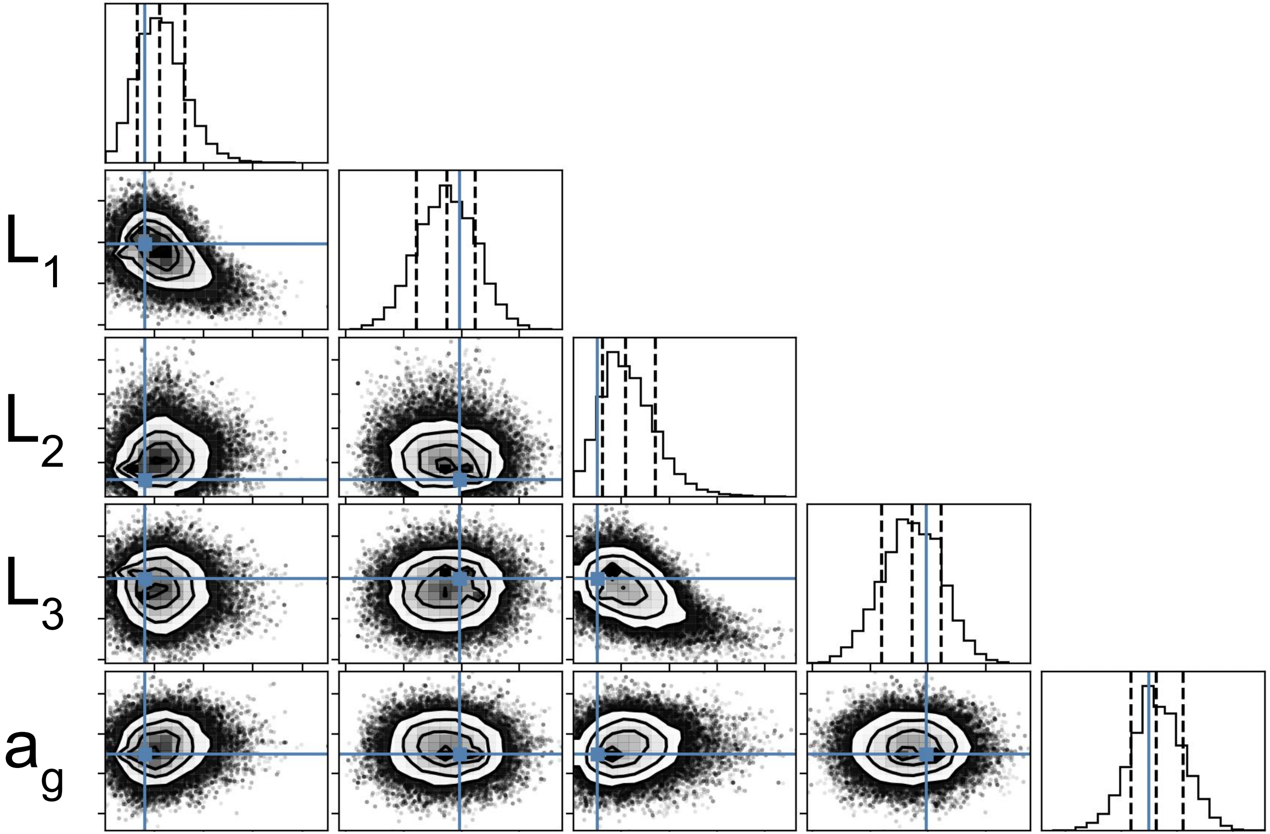


Model (left) uncertainty and data (right) uncertainty

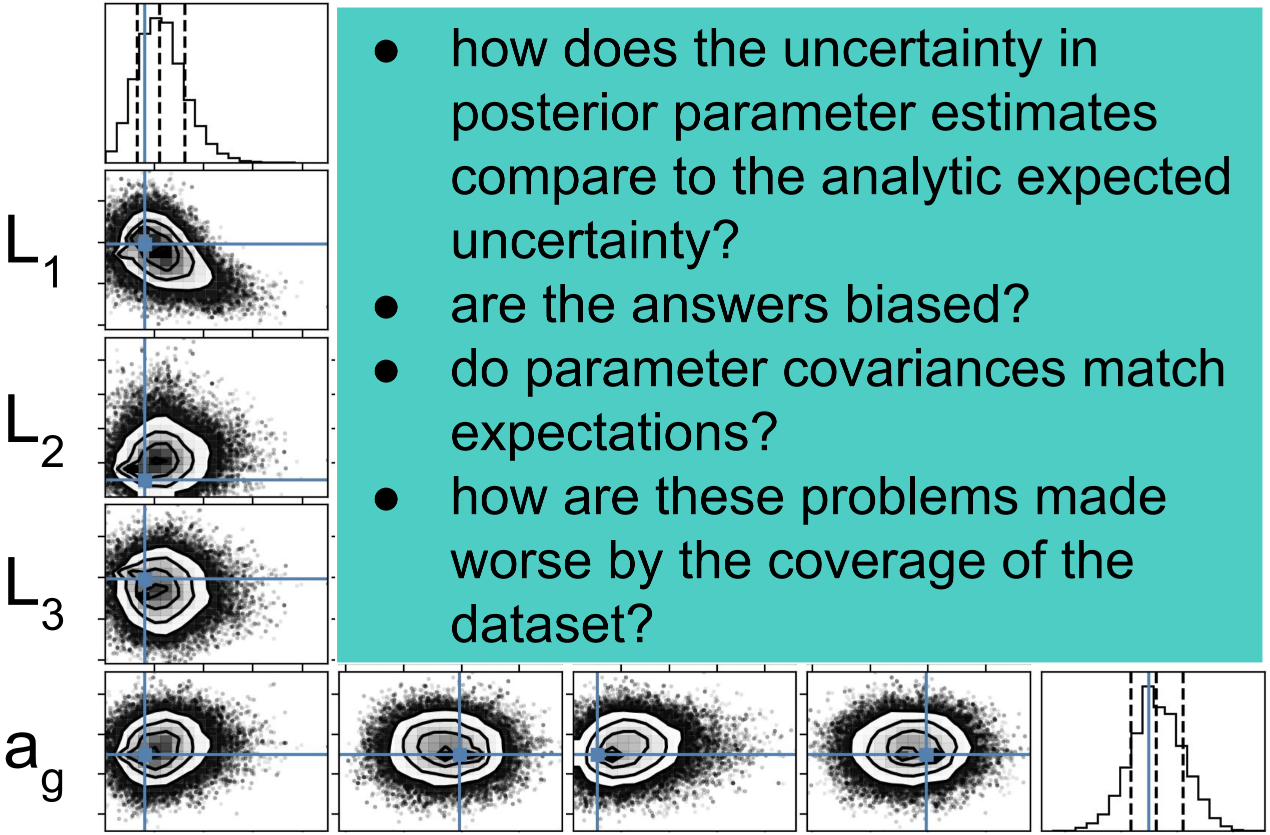
Analytical expectation of data uncertainty



Goal: build a framework to quantify uncertainty in the parameter estimates



Goal: build a framework to quantify uncertainty in the parameter estimates



- how does the uncertainty in posterior parameter estimates compare to the analytic expected uncertainty?
- are the answers biased?
- do parameter covariances match expectations?
- how are these problems made worse by the coverage of the dataset?

Use the UQ comparison and the tunable simulations to do a ***comparative analysis of inference methods***

Analytic errors from exact inference

Non-hierarchical sampling analysis
No Pooling
Full Pooling

Hierarchical sampling analysis

Use the UQ comparison and the tunable simulations to do a ***comparative analysis of inference methods***

Analytic errors from exact inference

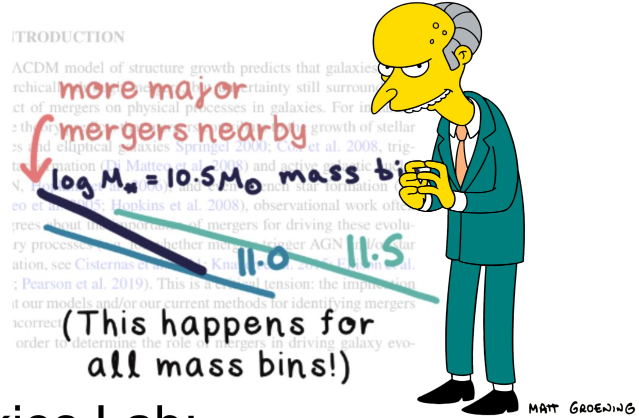
Non-hierarchical sampling analysis
No Pooling
Full Pooling

Hierarchical sampling analysis

Simulation Based Inference

Mergers

Adventures in statistical confounds



DeepSkies Lab:
Benchmark data for ML, uncertainty, and fancy Bayesian inference, oh my!