Interpretable statistical and machine learning: A gateway to astrophysics and cosmology

DEEP SKIES

Bringing Artificial Intelligence to Astrophysics

Dr. Becky Nevin

July 12 2023
CSAID Meeting

2009

Undergraduate Research

REUs

Graduated with bachelor's in physics-astronomy from Whitman College

2013

Graduate Research

Observational dual AGN and outflows, simulated and observed galaxy mergers

Thesis: "Kinematic Signatures of Galaxy Evolution: The Energetics of AGN Outflows and the Accurate Identification of Merging Galaxies"

2019

Postdoc Research

Harvard-Smithsonian CfA

Chandra ML support, multi-wavelength galaxy evolution

2022

Fermilab

Uncertainty in AI

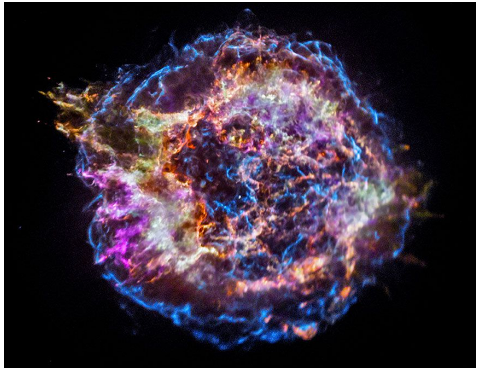
Hierarchical Bayesian Inference

Timeline credit:
M. Voetberg

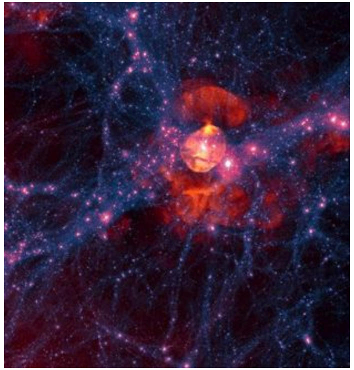
Active Galactic Nuclei

Mergers

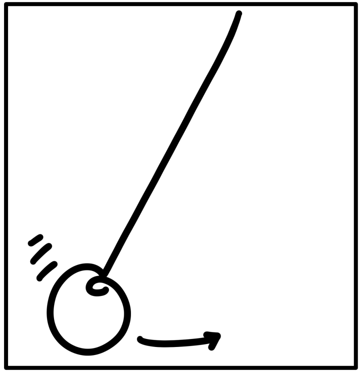
Chandra X-ray



Illustris



Benchmark

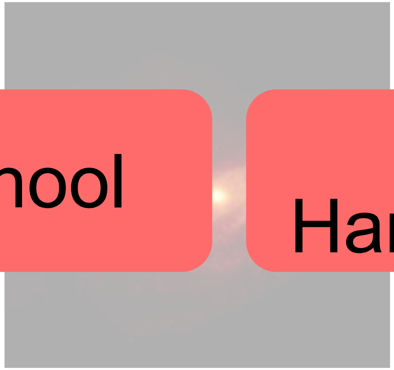


UQ

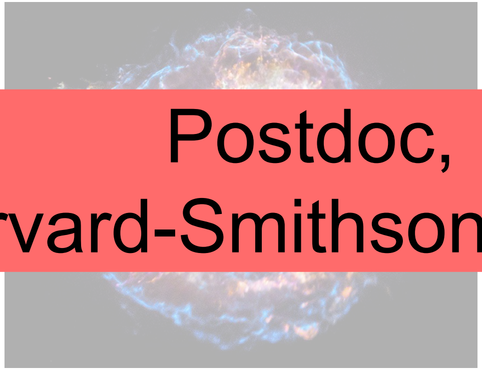
Hierarchical Inference

Active Galactic Nuclei

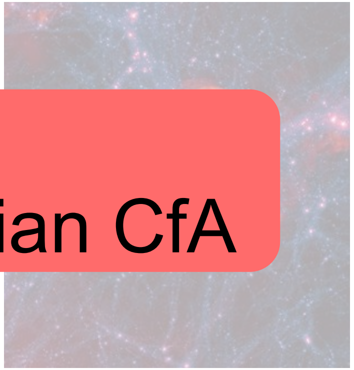
Mergers



Chandra X-ray



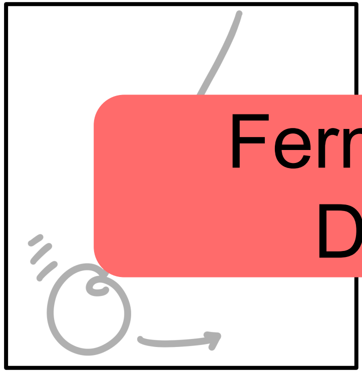
Illustris



Graduate school

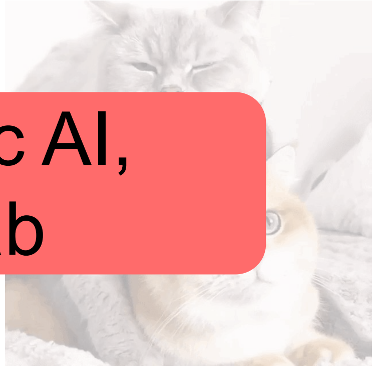
Postdoc,
Harvard-Smithsonian CfA

Benchmark



UQ

Hierarchical Inference

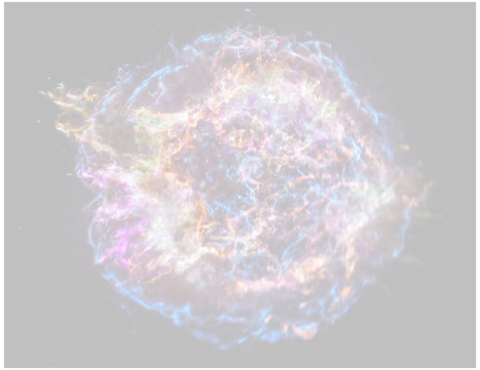


Fermilab Cosmic AI,
Deepskies Lab

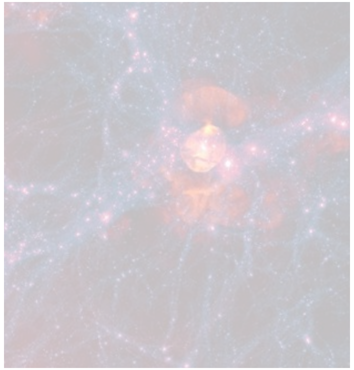
Active Galactic Nuclei

Mergers

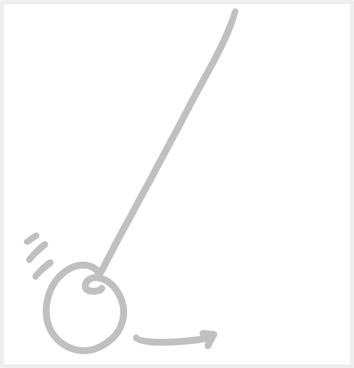
Chandra X-ray



Illustris

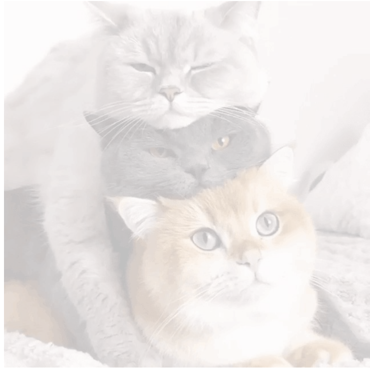


Benchmark

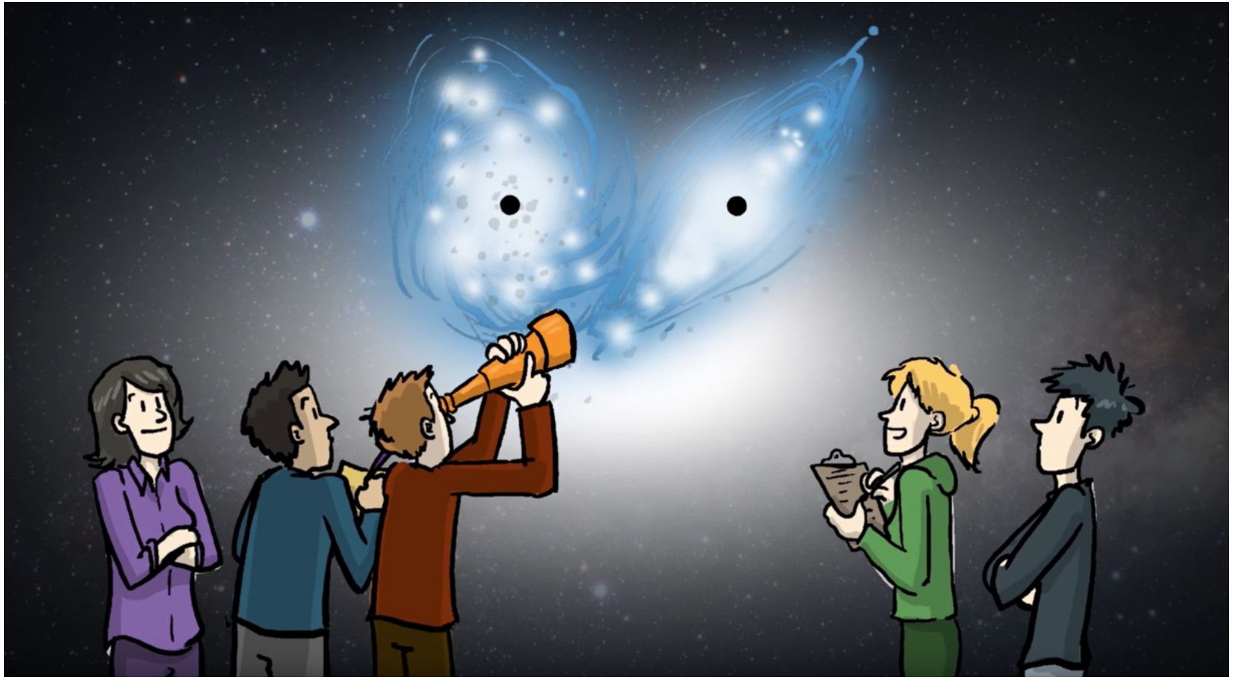


UQ

Hierarchical Inference

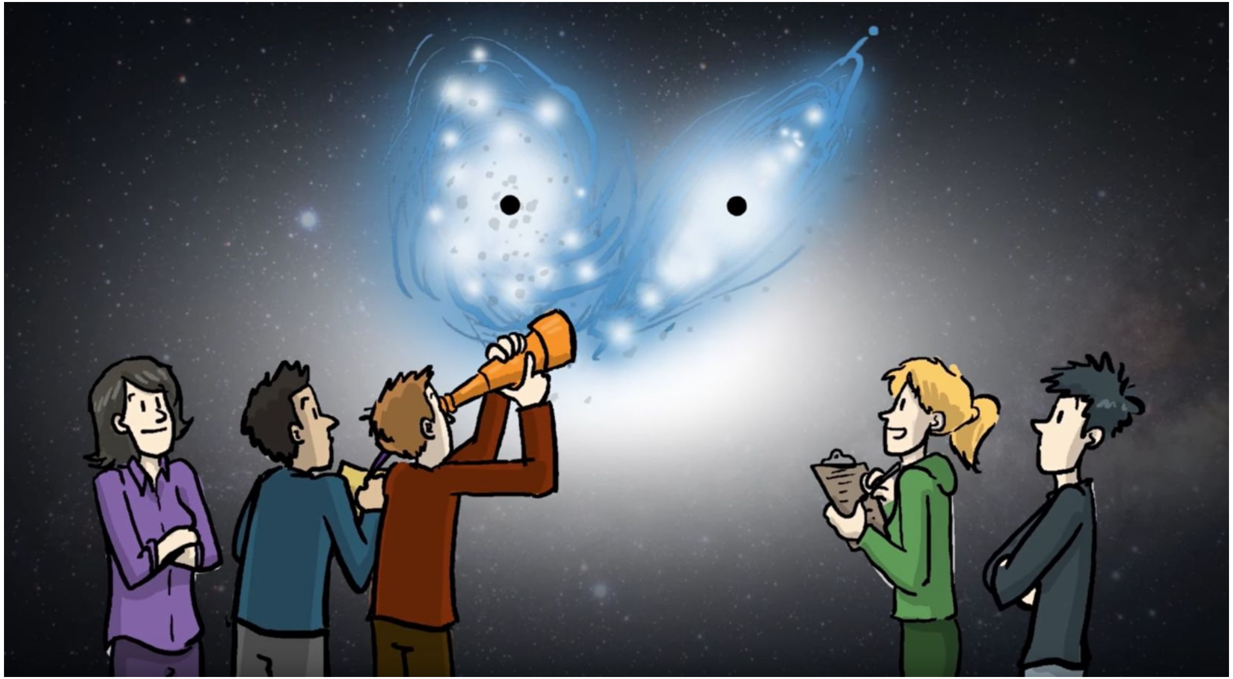


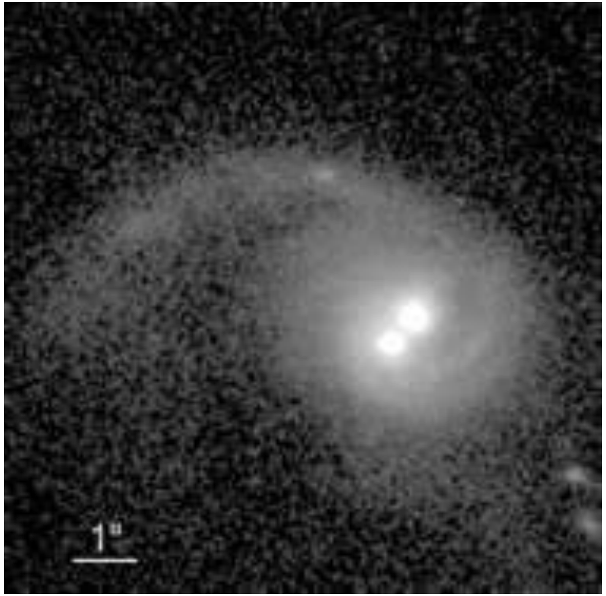
Statistical learning as a key (and interpretable) tool to characterize active galactic nuclei



https://www.youtube.com/watch?v=sqfbHyfuYDM&t=1s&ab_channel=PiledHigherandDeeper%28PHDComics%29

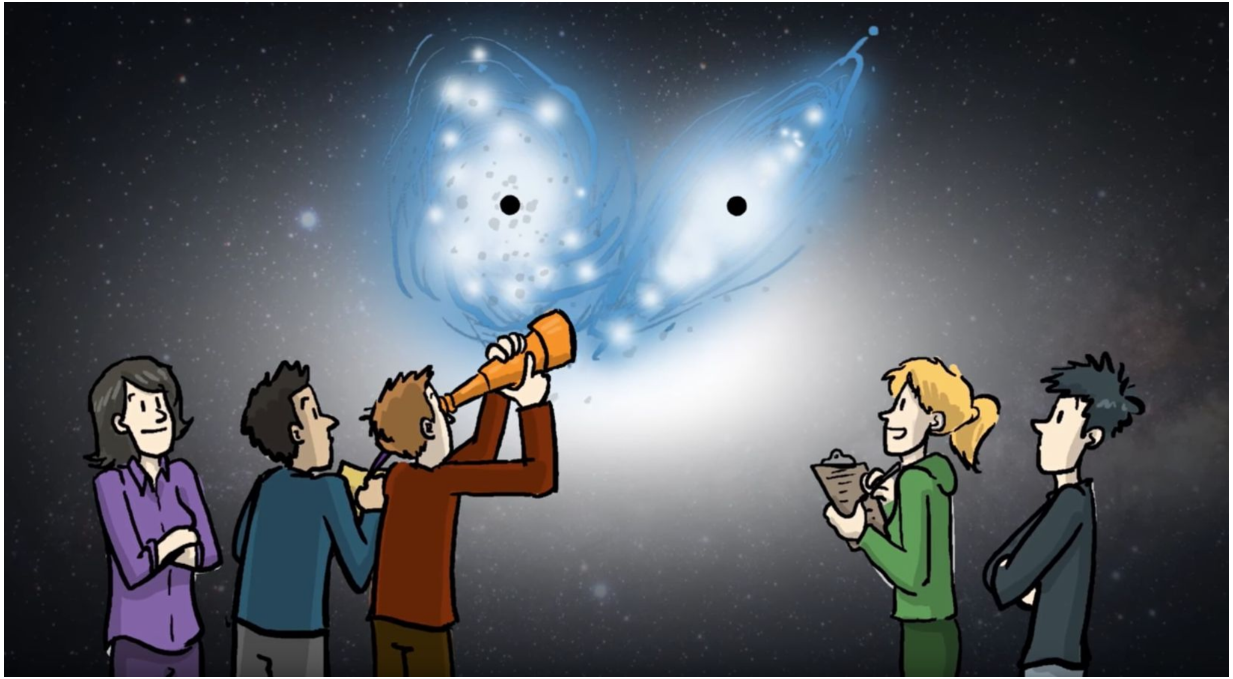
Statistical learning as a key (and interpretable) tool to characterize active galactic nuclei

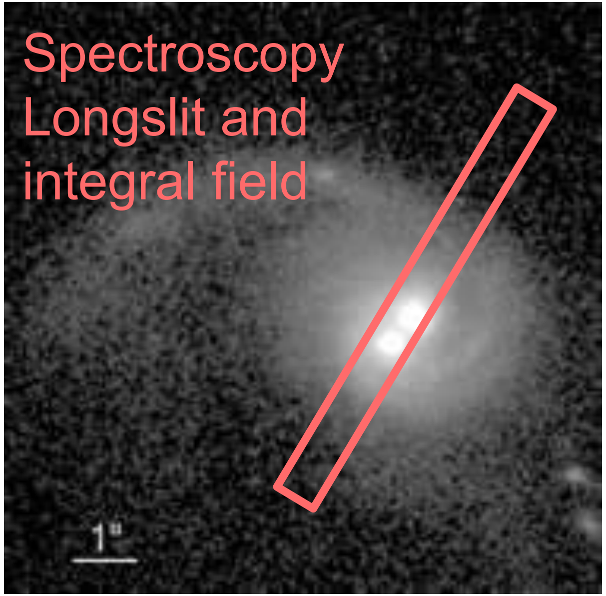




https://www.youtube.com/watch?v=sqfbHyfuYDM&t=1s&ab_channel=PiledHigherandDeeper%28PHDComics%29

Statistical learning as a key (and interpretable) tool to characterize active galactic nuclei

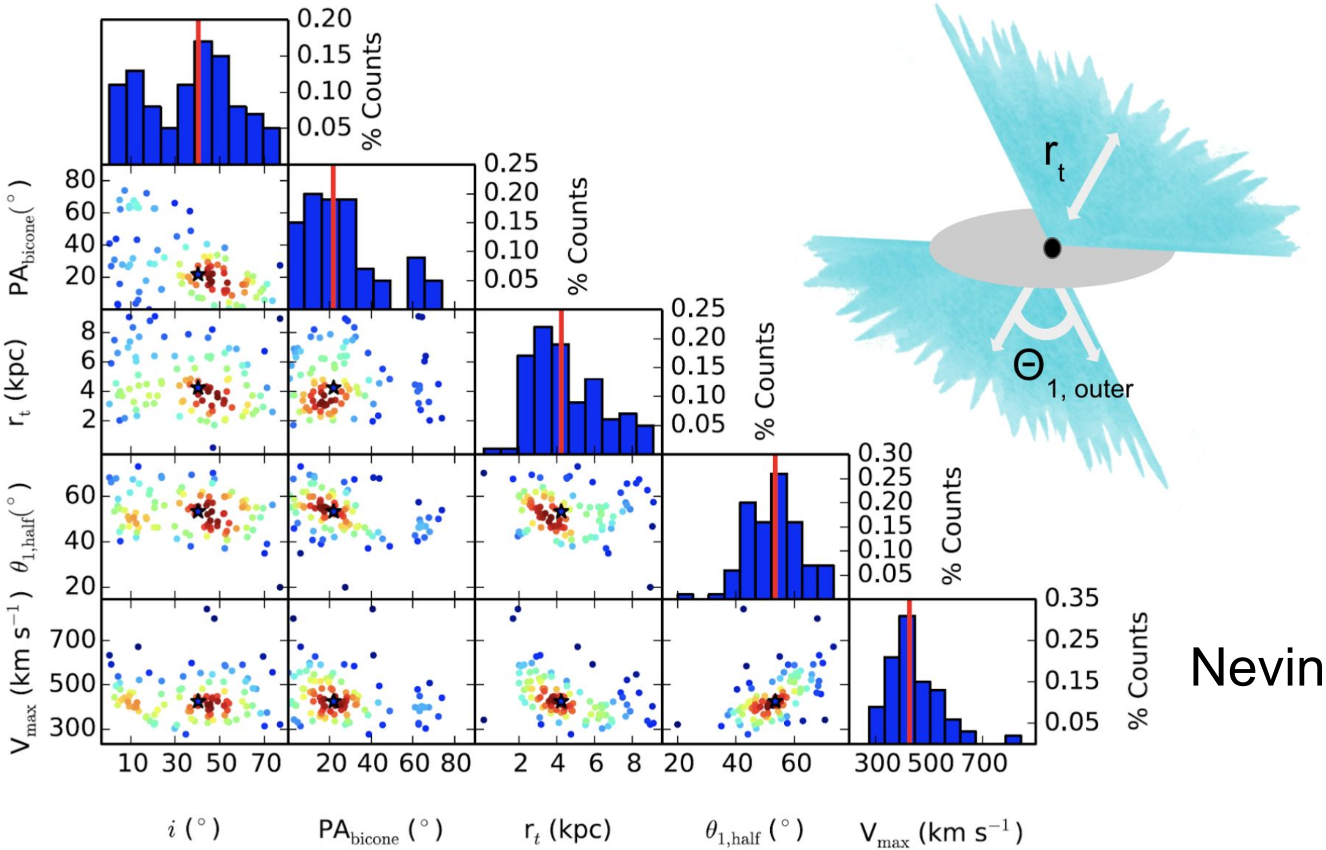


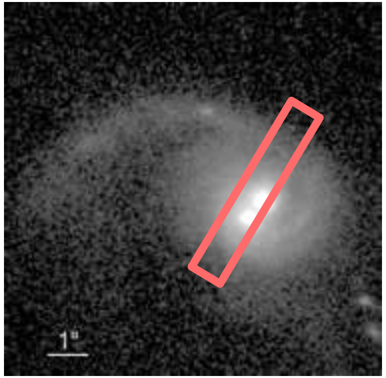


Nevin+2016

https://www.youtube.com/watch?v=sqfbHyfuYDM&t=1s&ab_channel=PiledHigherandDeeper%28PHDComics%29

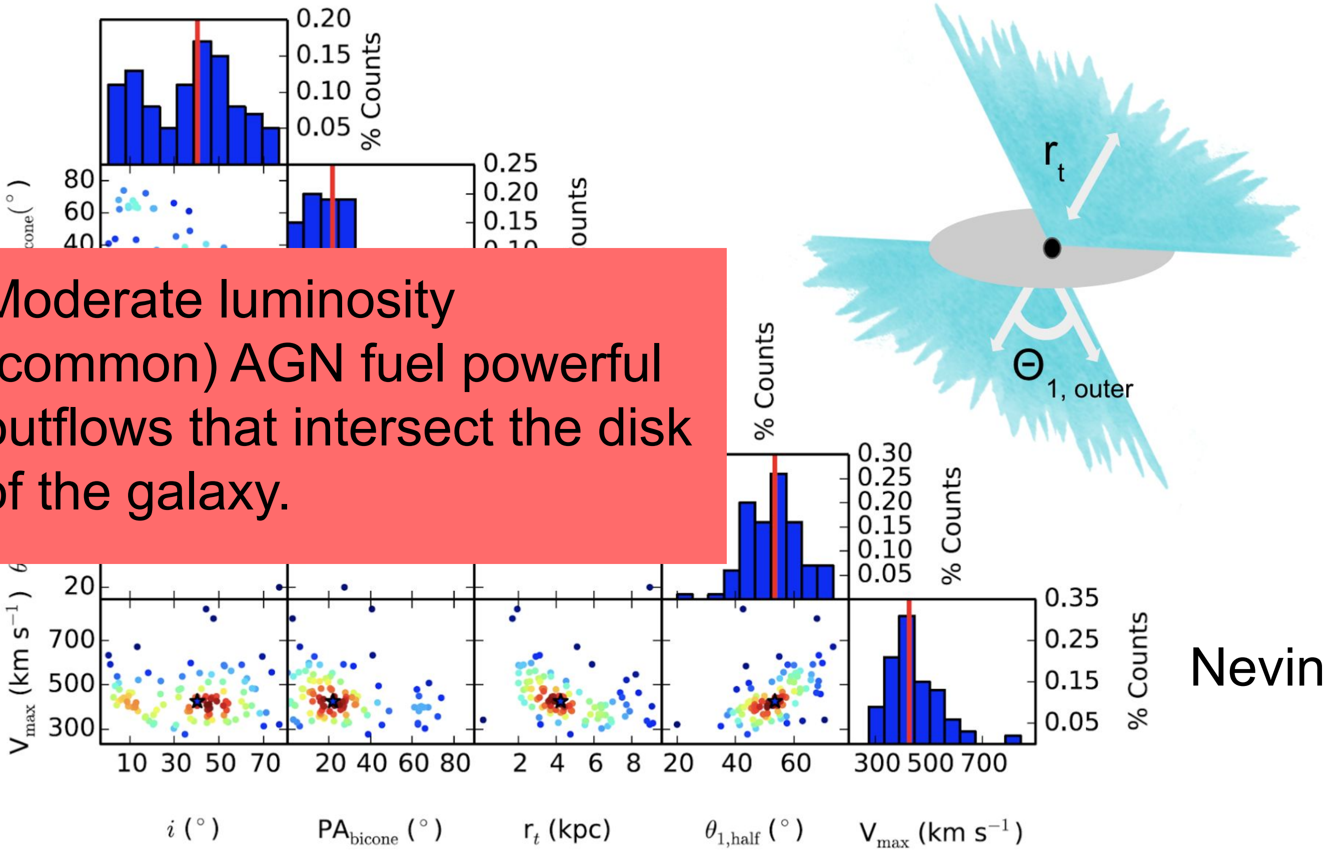
Statistical learning as a key (and interpretable) tool to characterize active galactic nuclei driven outflows



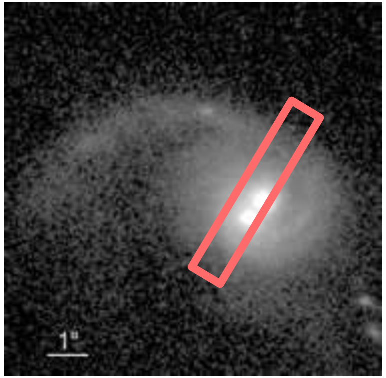


Nevin+2018

Statistical learning as a key (and interpretable) tool to characterize active galactic nuclei

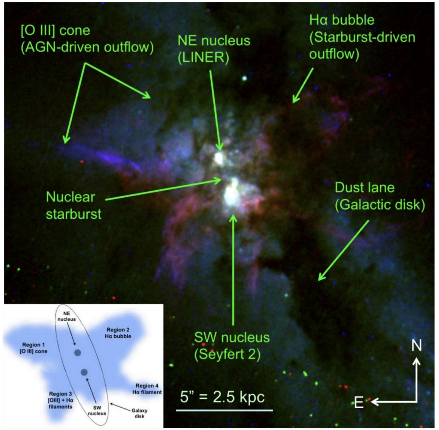


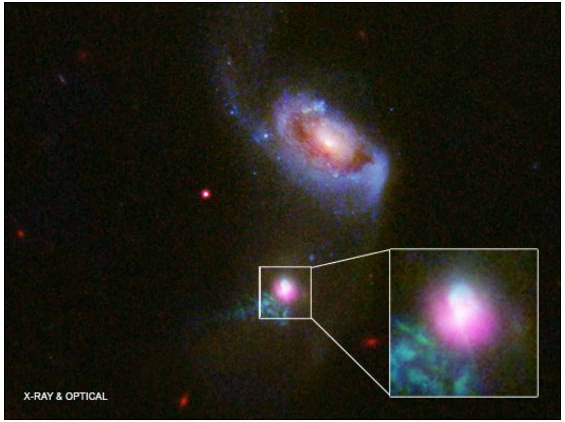
Moderate luminosity (common) AGN fuel powerful outflows that intersect the disk of the galaxy.



Nevin+2018

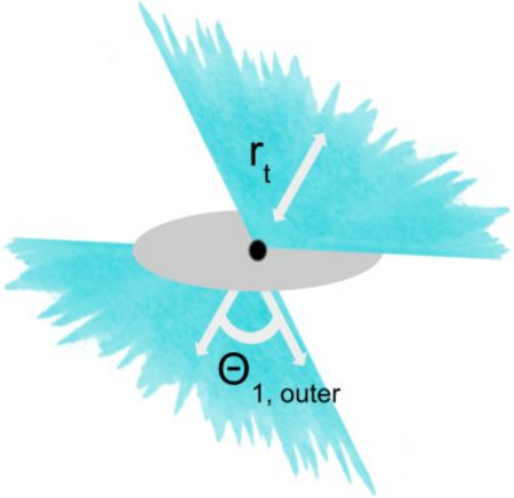
Optical emission line observations





Comerford+2017,2018,2020,
Müller-Sánchez+2015,2018

AGN outflow and kinematics

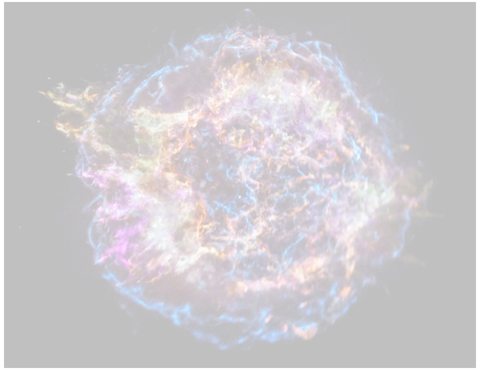


Roy+2021, Foord+2020,
Nevin+2016,2018

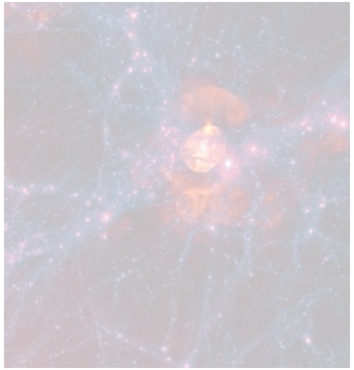
Active Galactic Nuclei

Mergers

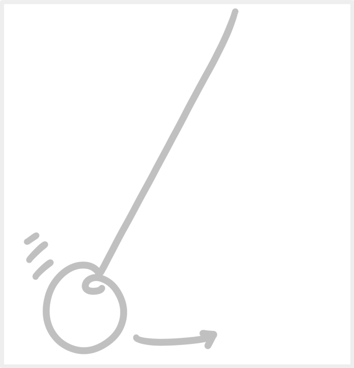
Chandra X-ray



Illustris

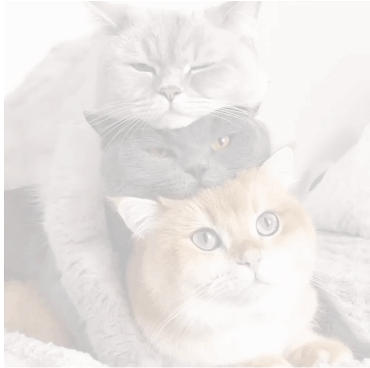


Benchmark

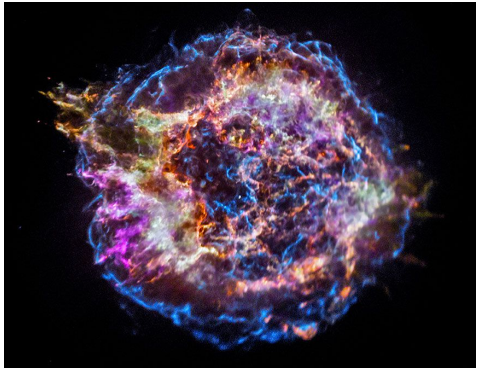
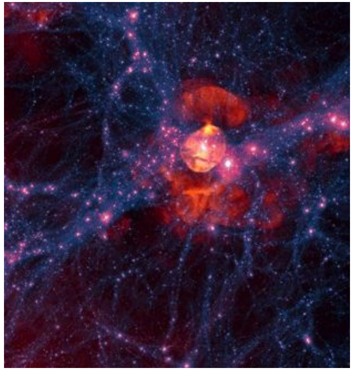


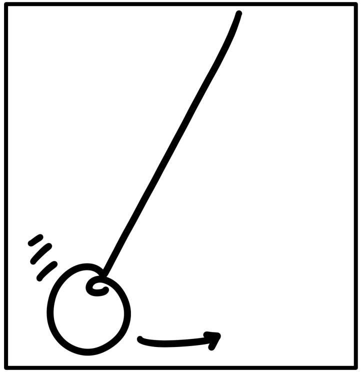
UQ

Hierarchical Inference



Interpretable statistical and machine learning:
A gateway to astrophysics and cosmology

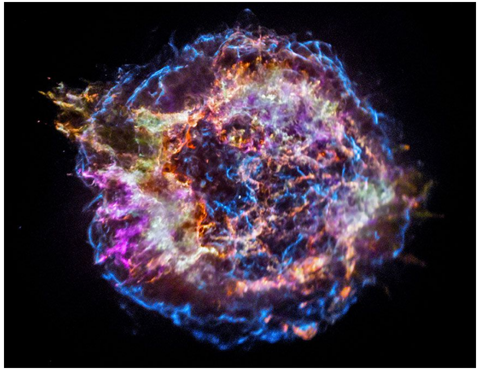
Active Galactic Nuclei	Mergers	*Chandra* X-ray	Illustris
			



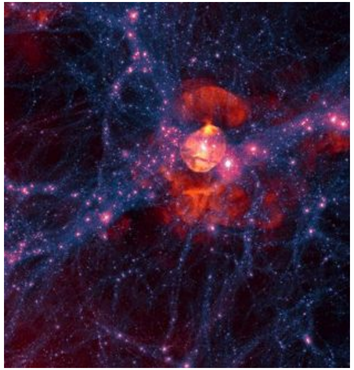
Active Galactic Nuclei

Mergers

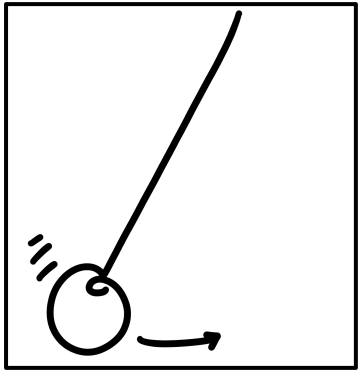
Chandra X-ray



Illustris



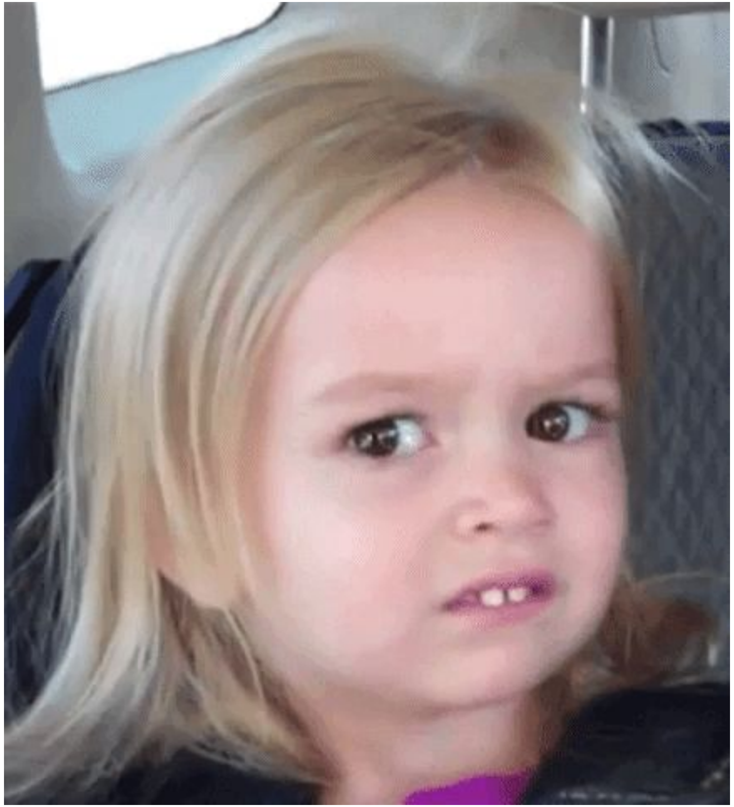
Benchmark



UQ

Hierarchical Inference

Why is identifying mergers hard?



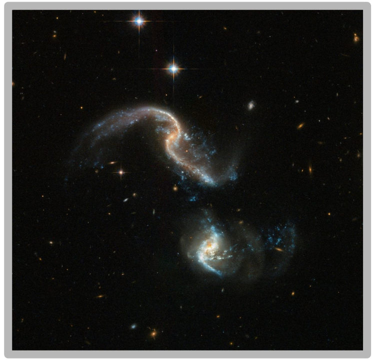
Why is identifying mergers hard?

There are many different types and stages of mergers and they all look different observationally.

Why is identifying mergers hard?

Close pairs

Interacting

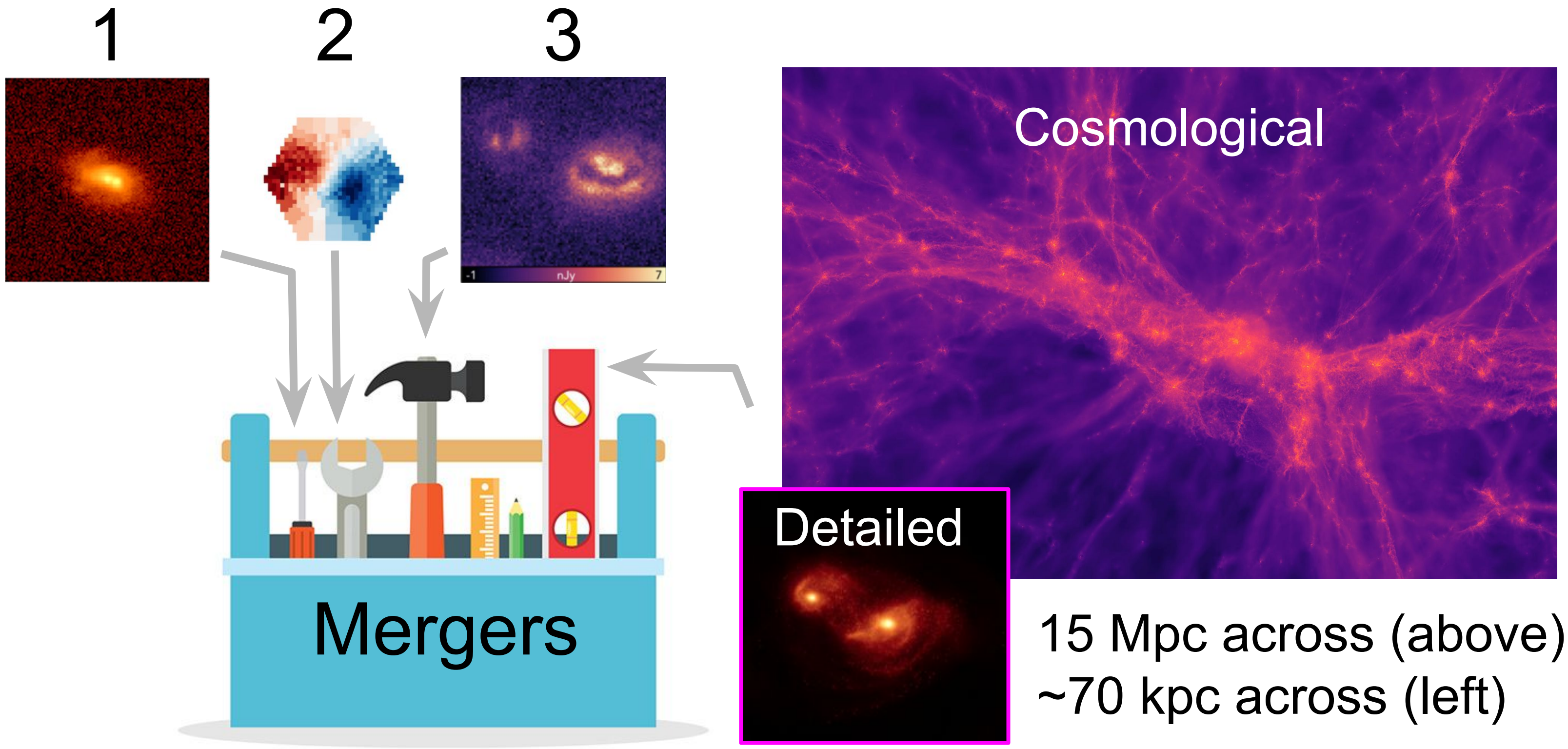


Coalescence

Post-merger

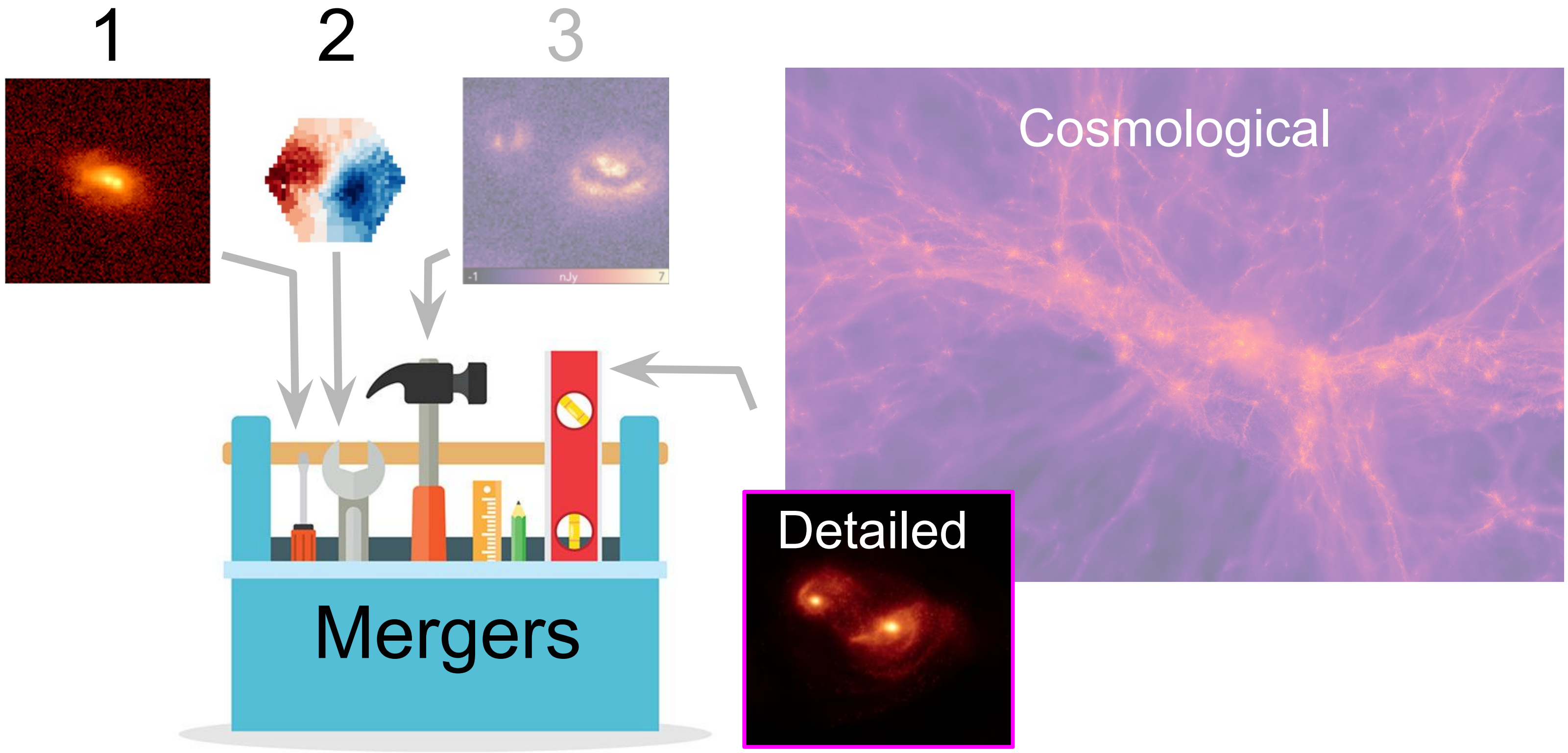
There are many different types and stages of mergers and they all look different observationally.

I approach better identifying mergers with the help of detailed hydro and cosmological simulations

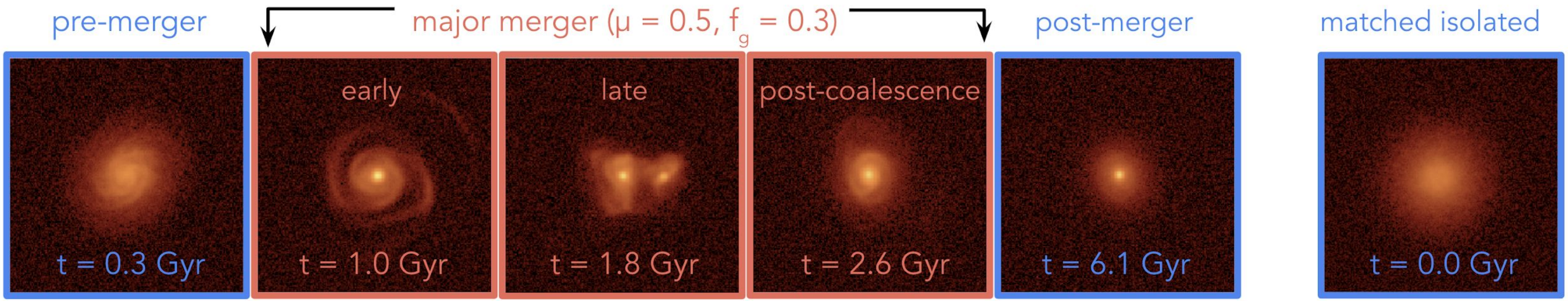


15 Mpc across (above)
~70 kpc across (left)

I approach better identifying mergers with the help of detailed hydro and cosmological simulations



Simulations of merging and nonmerging galaxies

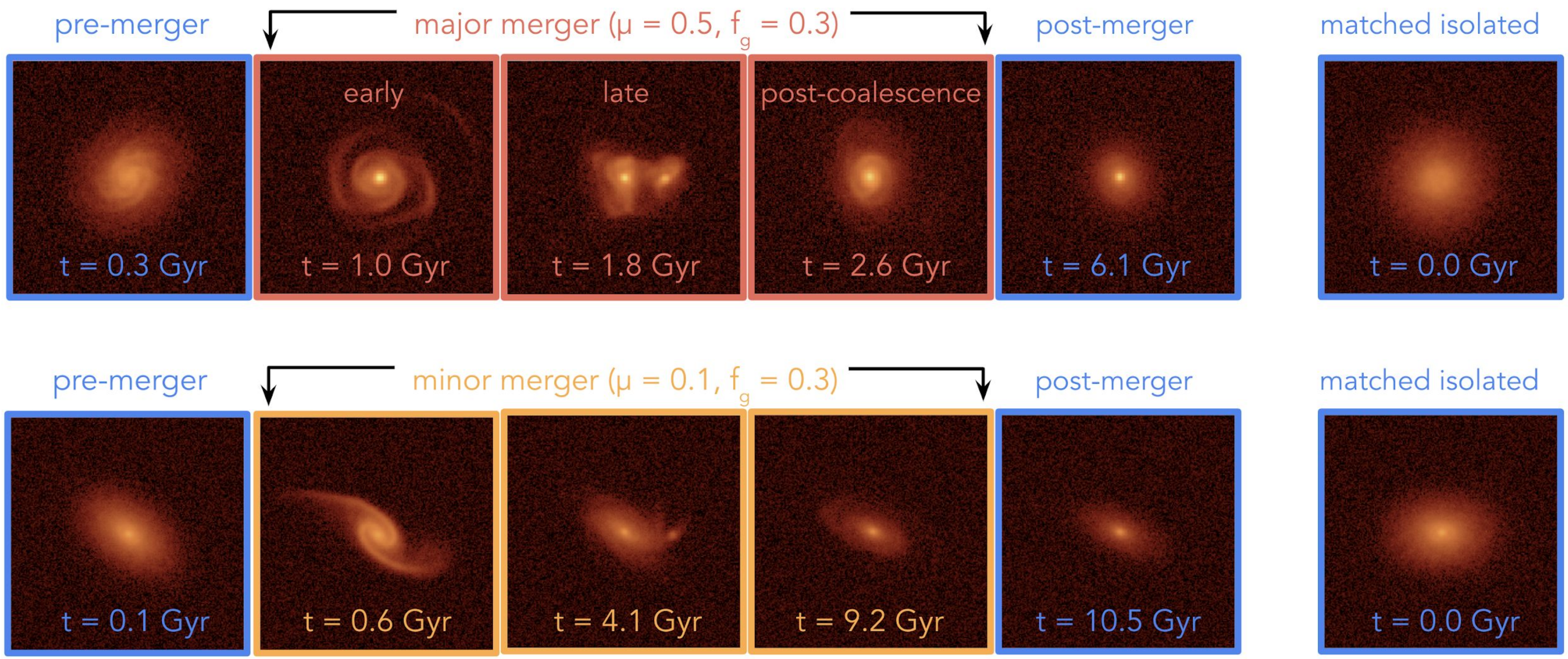


100s of snapshots per simulation
x 5 simulations

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

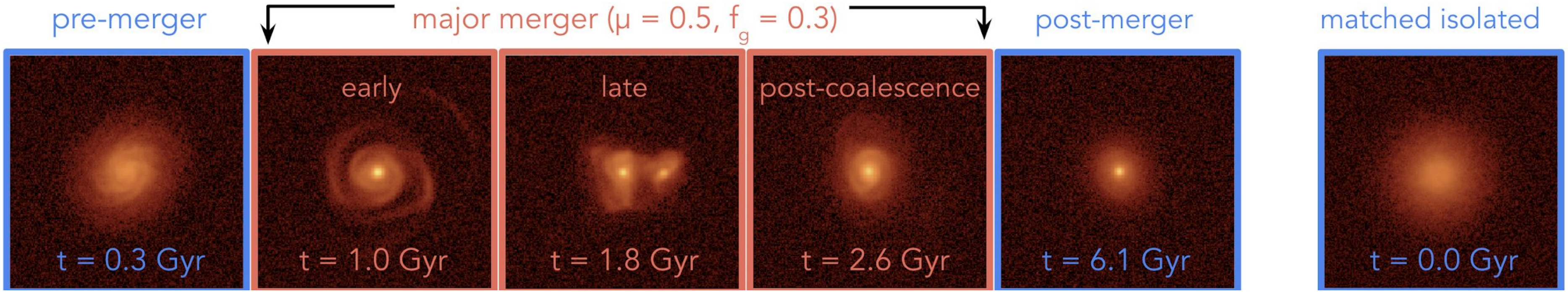
Nevin+2019

Major merger = more equal mass ratio, minor merger = unequal



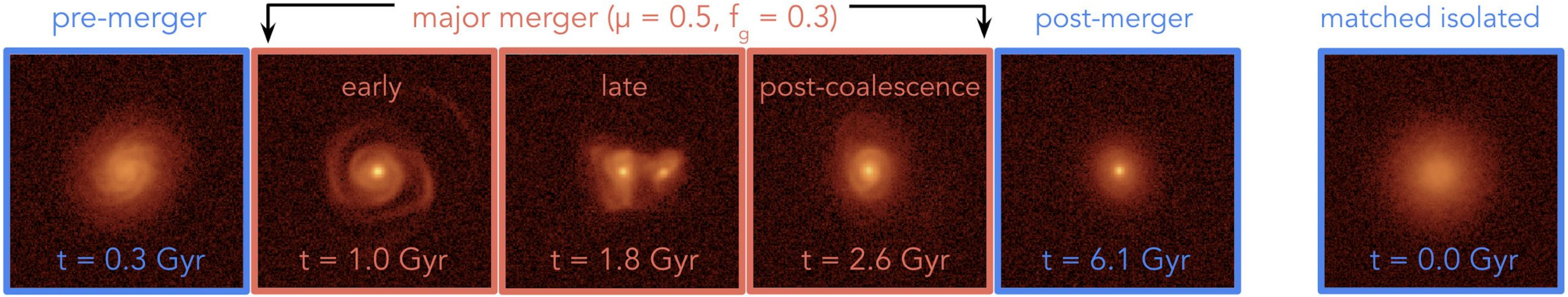
Nevin+2019

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



Nevin+2019

I combine all seven measured predictors using linear discriminant analysis (LDA)

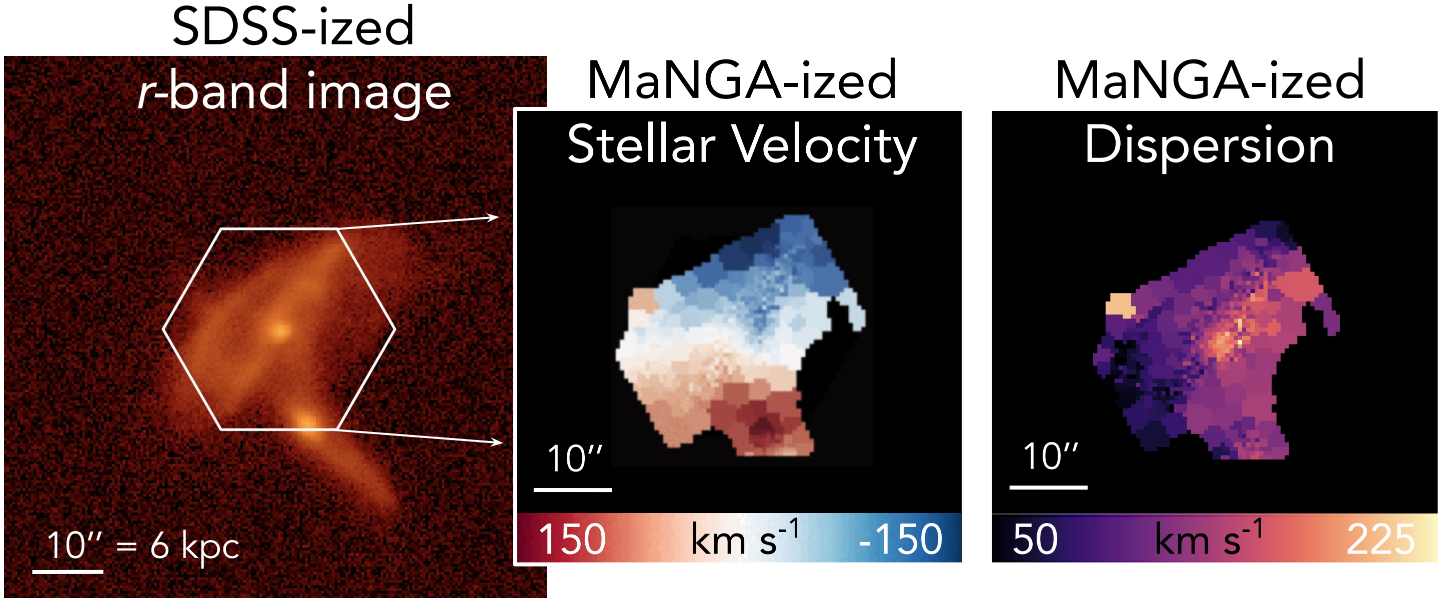


The LDA is more accurate and precise than any of the individual predictors in identifying mergers.

It is also not a black box.

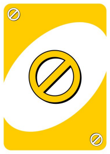
Nevin+2019

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



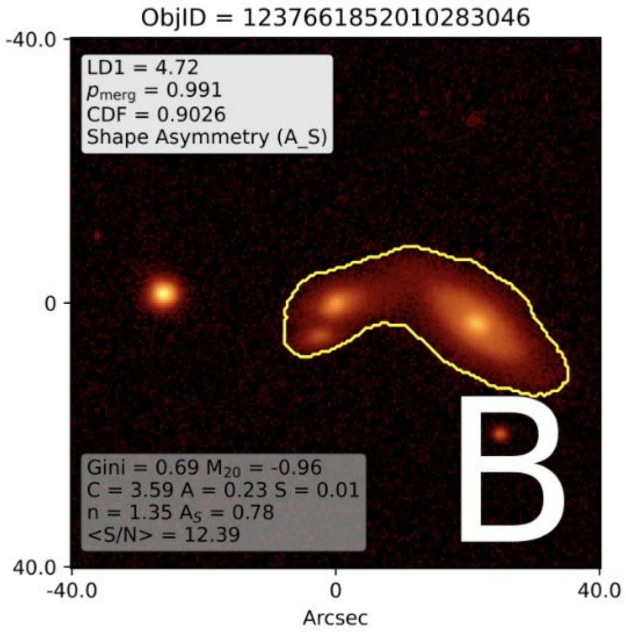
Nevin+2019

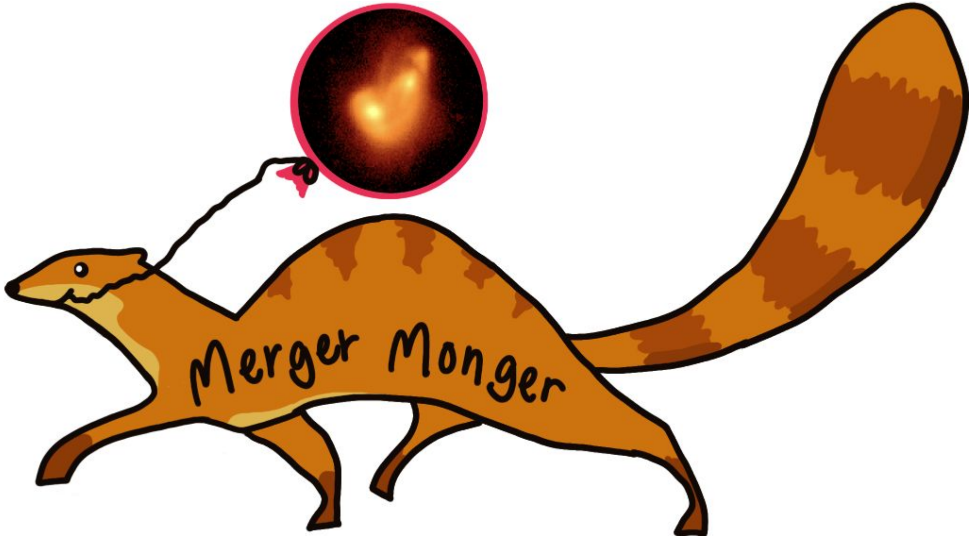
Nevin+2021



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

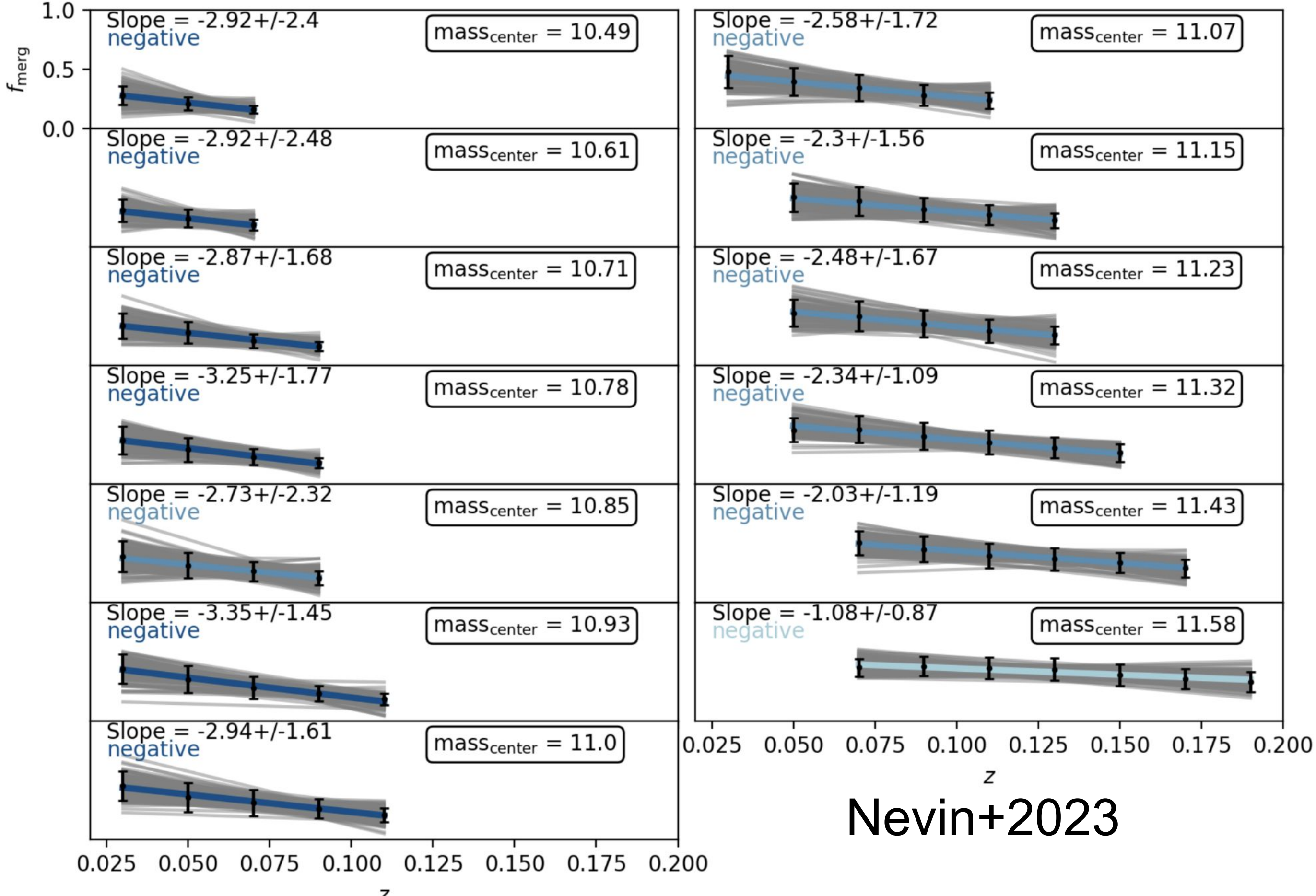
[MergerMonger Github Repo]





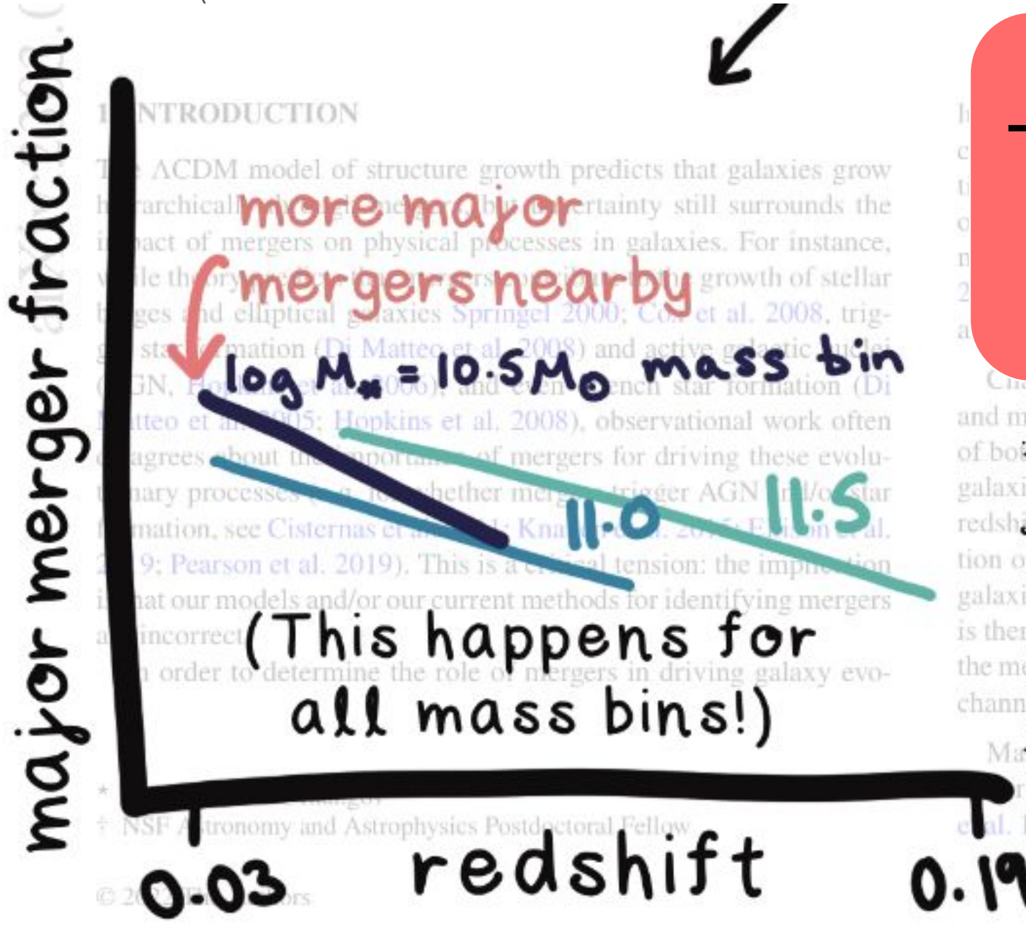
Nevin+2023

The major merger fraction decreases with redshift

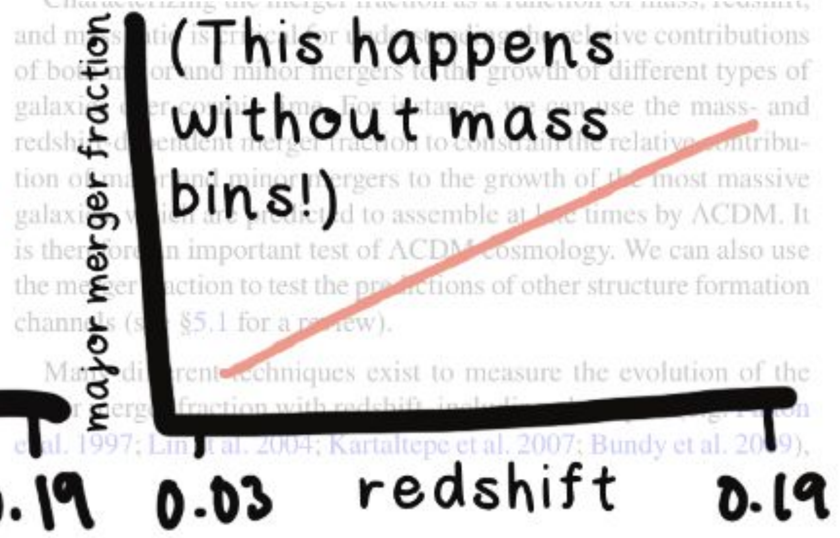


Nevin+2023

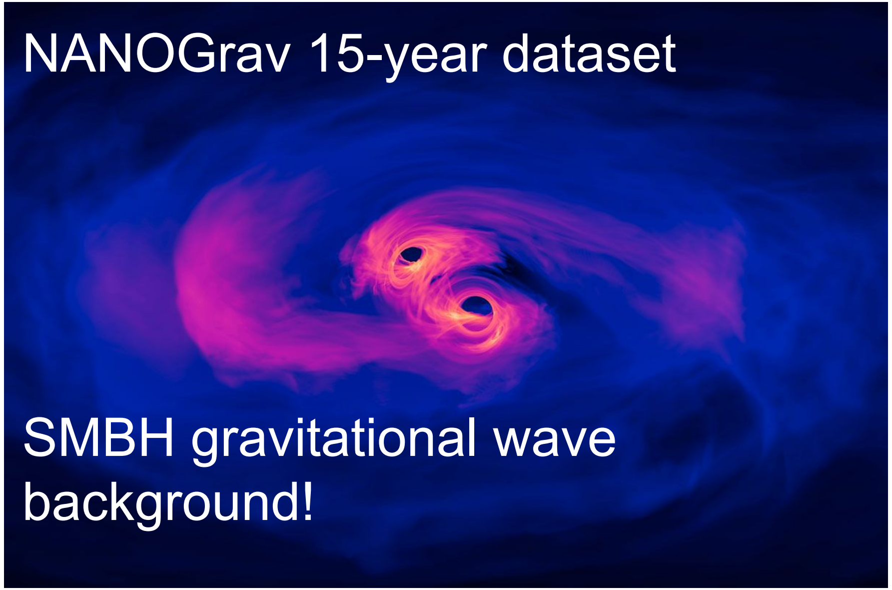
The major merger fraction decreases with redshift



This is different than in past work!



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



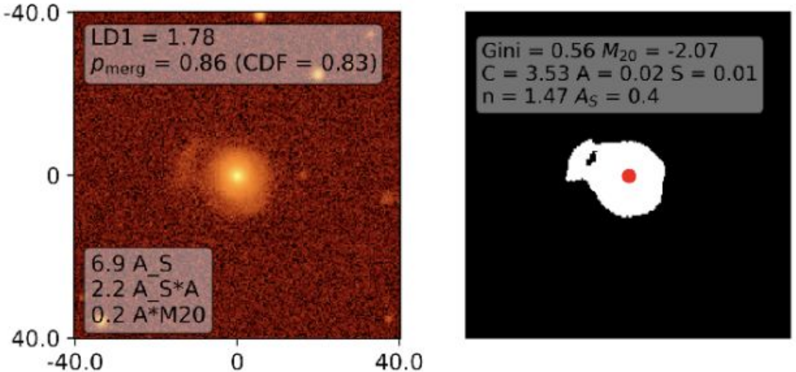
Joe Simon

Julie Comerford

Simon+2023 in prep

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

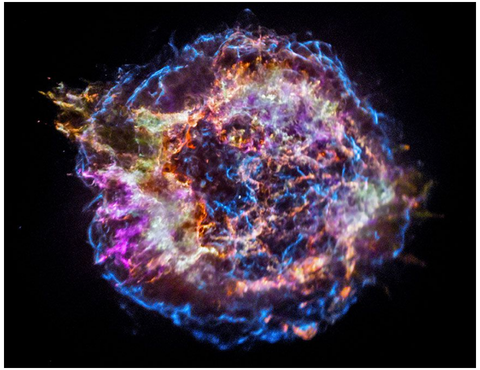
Comerford+2023; An excess of AGNs triggered by galaxy mergers in MaNGA galaxies of mass 10^{11} M⊙
Hernández-Toledo+2023; MaNGA AGN have an enhanced merger fraction
Negus+2023; Coronal line MaNGA galaxies



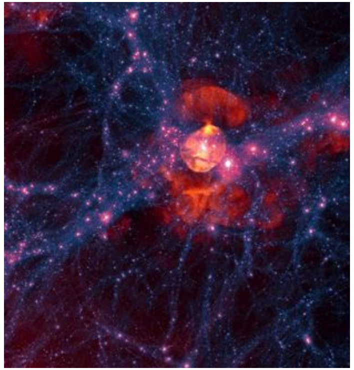
Active Galactic Nuclei

Mergers

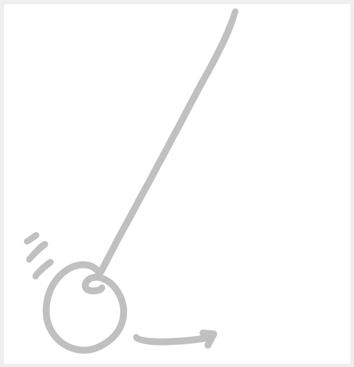
Chandra X-ray



Illustris



Benchmark



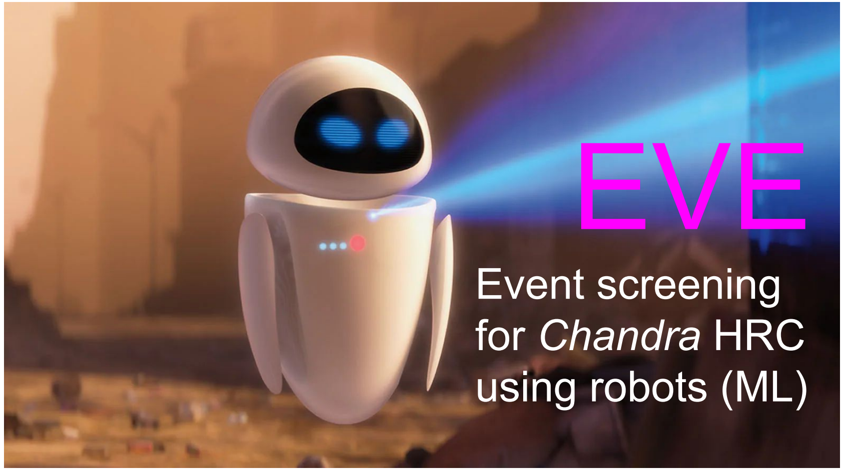
UQ

Hierarchical Inference

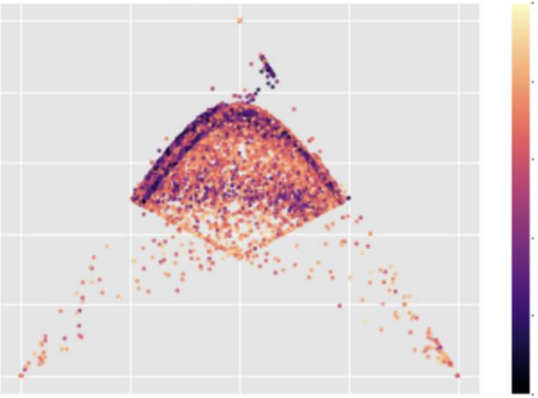
Harnessing machine learning to improve the background rejection of *Chandra* HRC

CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

Becky Nevin, Grant Tremblay, Ralph Kraft, Paul Nulsen, Dan Patnaude, Dan Schwartz, and Alexey Vikhlinin

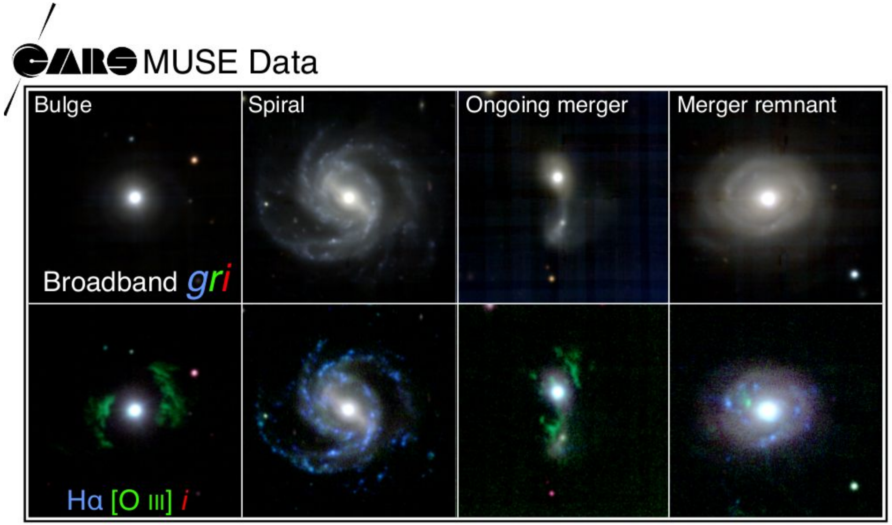


Semi-supervised bagging classifier



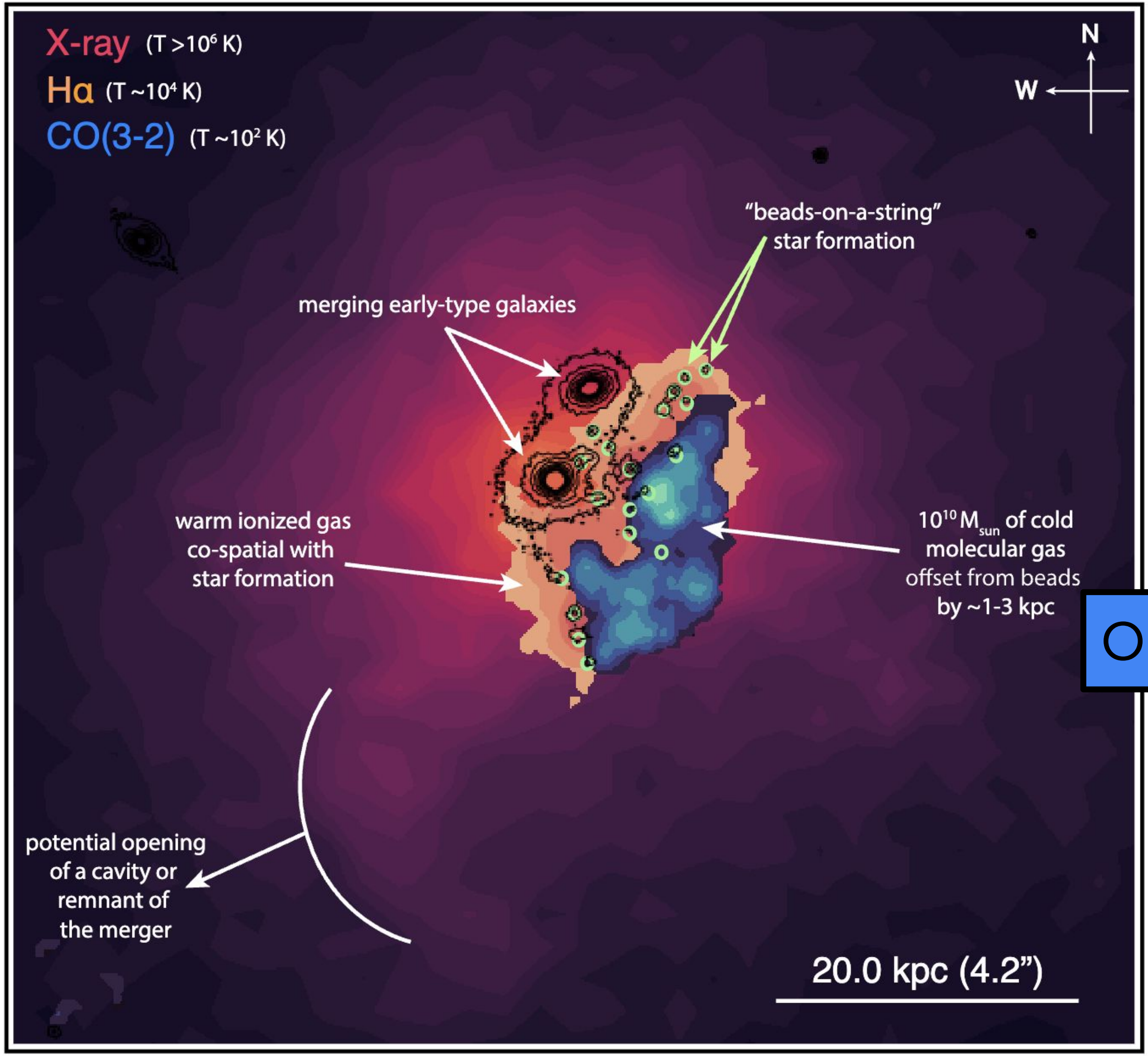
CARS: Close AGN Reference Survey

A multi-wavelength survey of a representative sample of luminous Type I AGN at redshifts 0.01 < z < 0.06 to help unravel the connection between galaxies and AGN.
https://cars.aip.de/



Grant Tremblay	Bryan Terrazas	Osase Omoruyi
		

Cluster major merger with exquisite X-ray observations; intracluster gas motions and ram pressure caused gas offset from young stellar superclusters and turbulence from merger caused beads on a string phenomenon



Omoruyi+2023

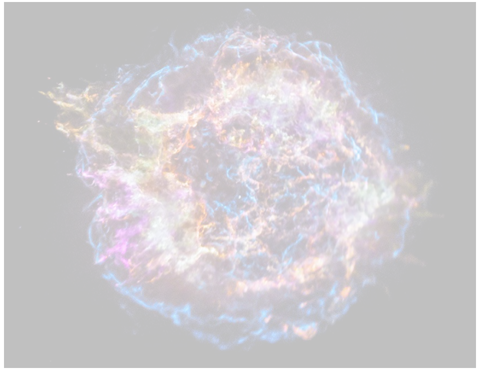
Illustris TNG50 team member

Nelson+2021→ star formation in TNG50 and 3D-HST
Hartley+2023 → the first quiescent galaxies in TNG300
Data set curation (stay tuned)

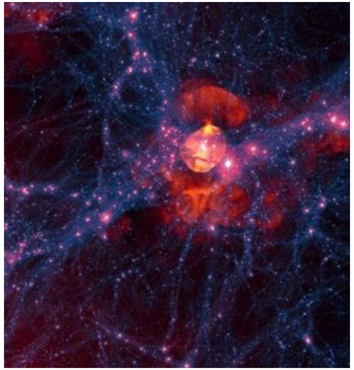
Active Galactic Nuclei

Mergers

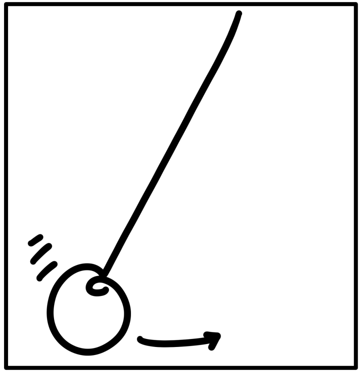
Chandra X-ray



Illustris



Benchmark



UQ

Hierarchical Inference

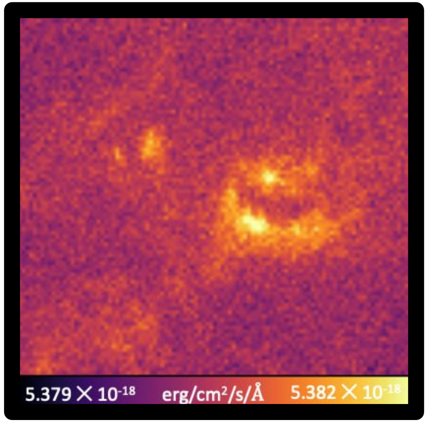
I wanted to come to Fermilab and work with the Deepskies crew because:

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

Aimee and I create mock images from Illustris
From these we use CNNs + domain adaptation to classify mergers in *HST* and *JWST* images

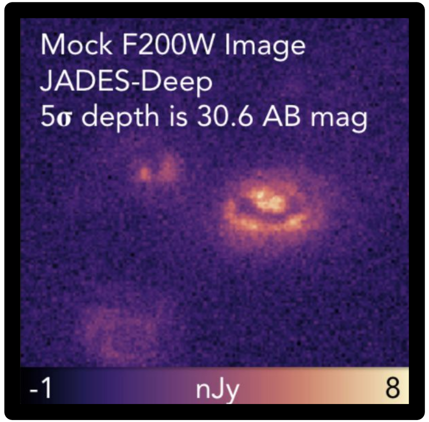
Aimee Schechter

HST F814W



Schechter+2024

JWST F200W

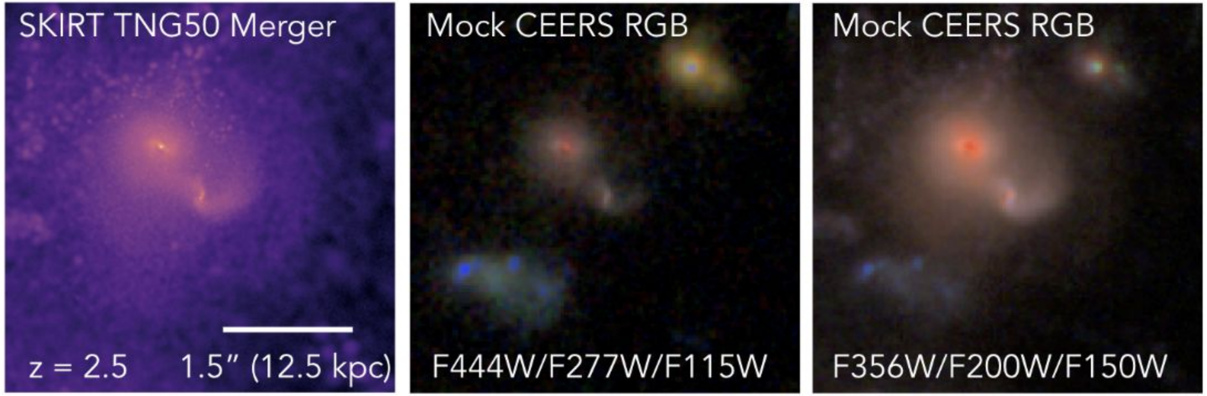


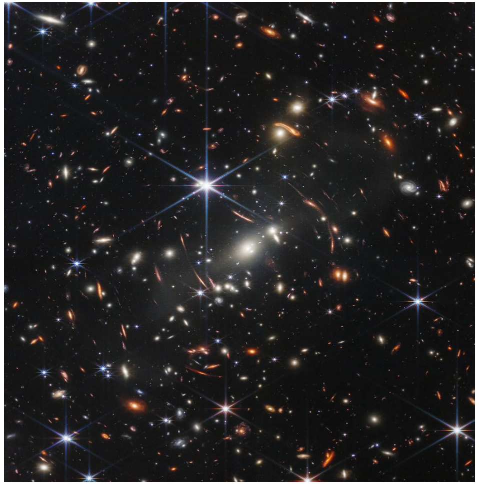
Nevin+2024

Carefully incorporating domain adaptation is necessary and interesting

Real *JWST* galaxies (SMACS 0723)

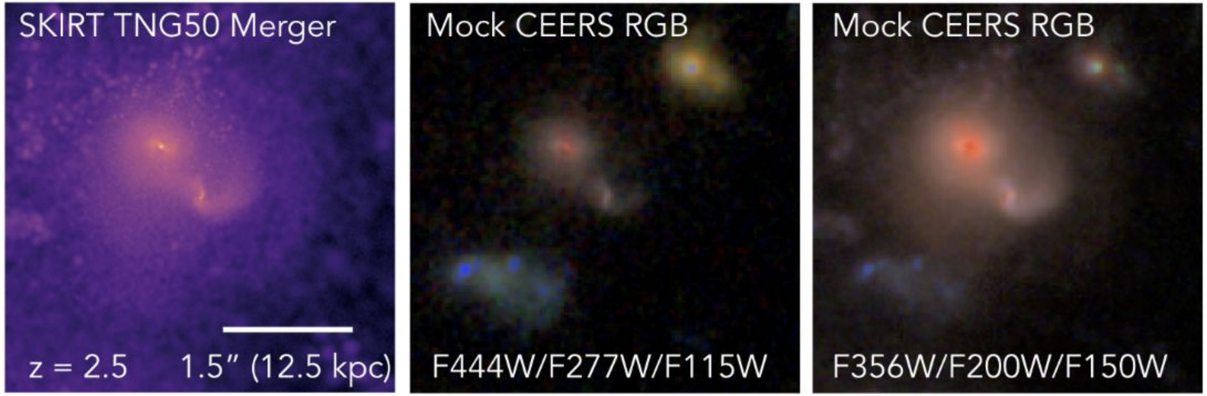
Simulated galaxies



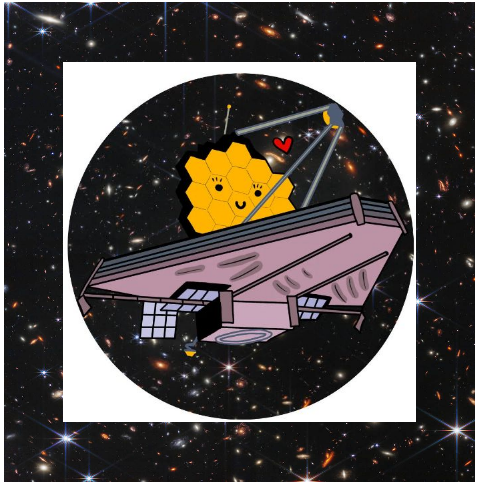


Carefully incorporating domain adaptation is necessary and interesting

Simulated galaxies

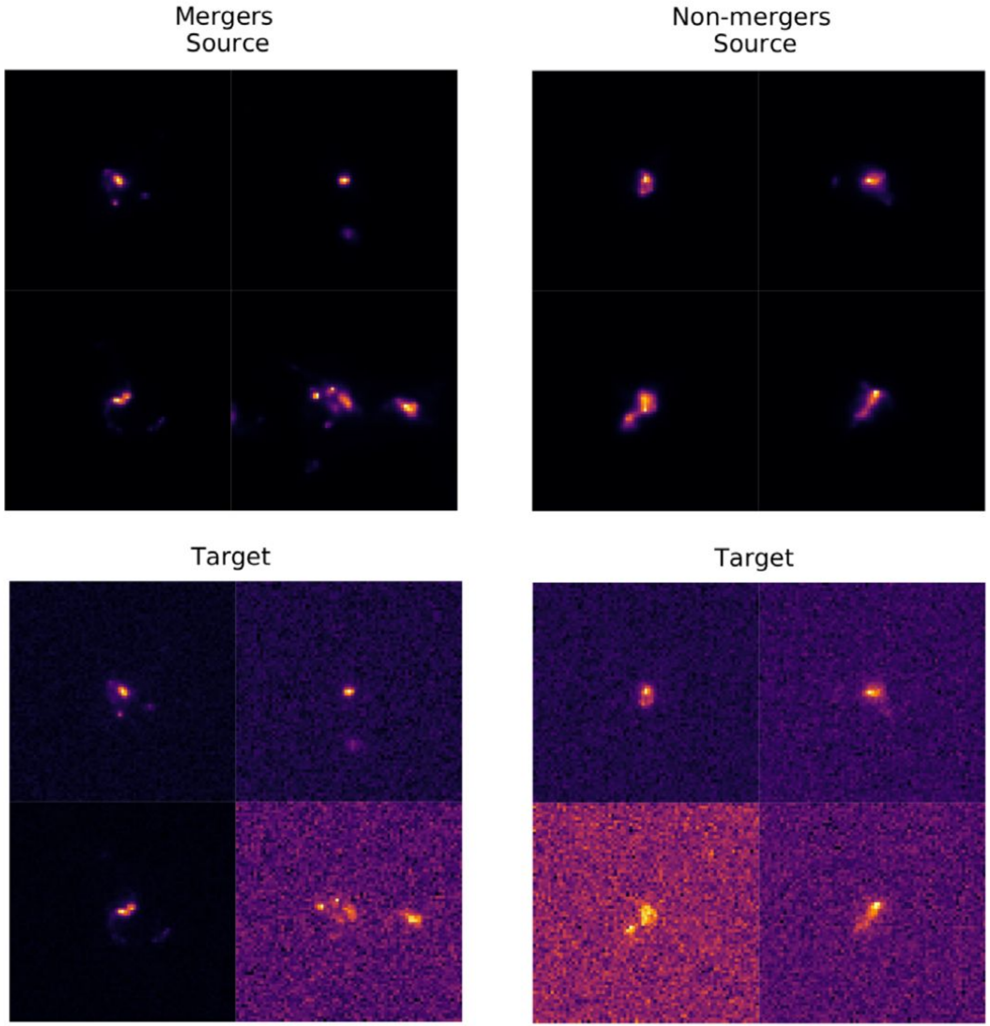


Real *JWST* galaxies
(SMACS 0723)



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

Ćiprijanović+ 2020a,2021



Team ‘Fake it till you make it’

A smorgasbord of mocks from Illustris TNG50

JWST
NIRCam

Becky Nevin

HST
CANDELs

Aimee
Schechter

SKIRT9 +
AGN

Jacob Shen

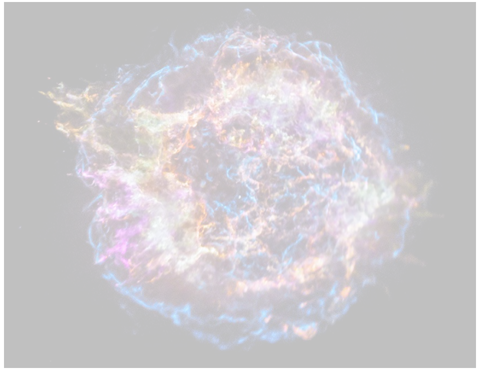
HSC-Joint,
MaNGA, SAMI,
HECTOR

Connor Bottrell

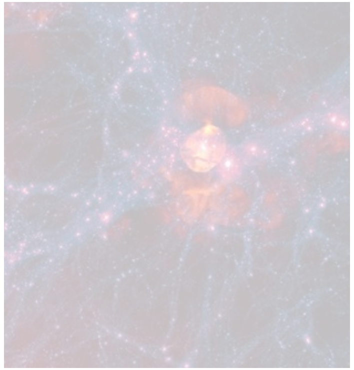
Active Galactic Nuclei

Mergers

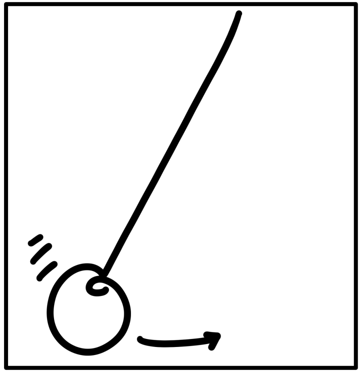
Chandra X-ray



Illustris



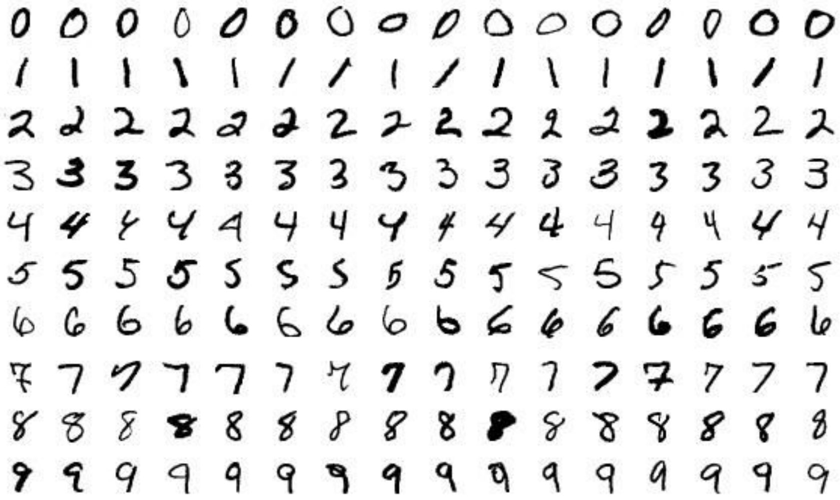
Benchmark

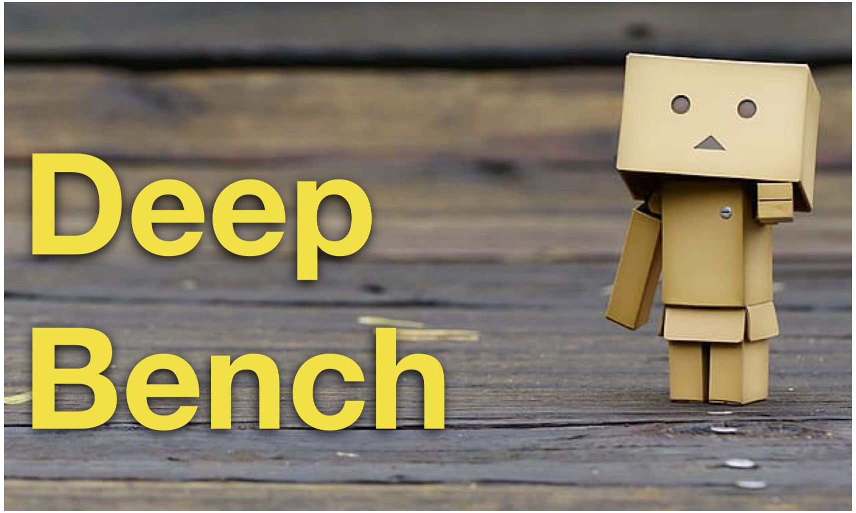


UQ

Hierarchical Inference

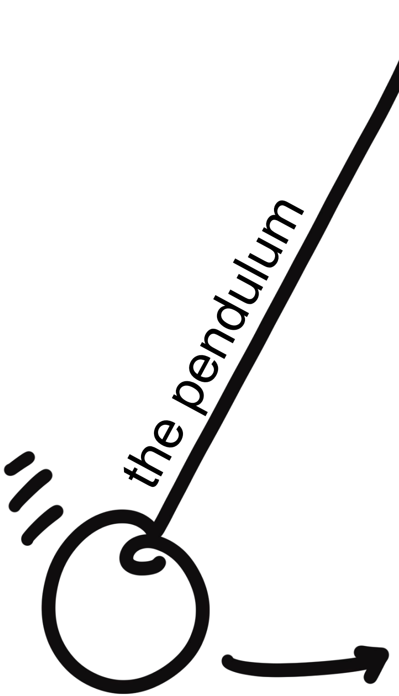
DeepBench: Fine-grained control for simulations for neural inference





Fine-grained control over noise
Its a model (θ←→x)
Its dynamic

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



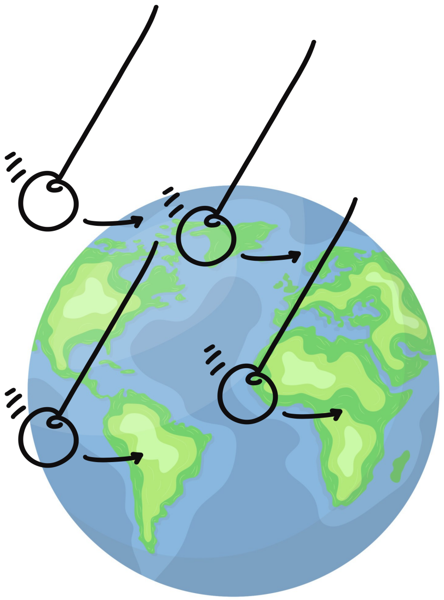
Things we'd like to infer about a pendulum:

- starting angle
- mass
- length

Data:

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

But physics is not as simple as one experiment



EARTH

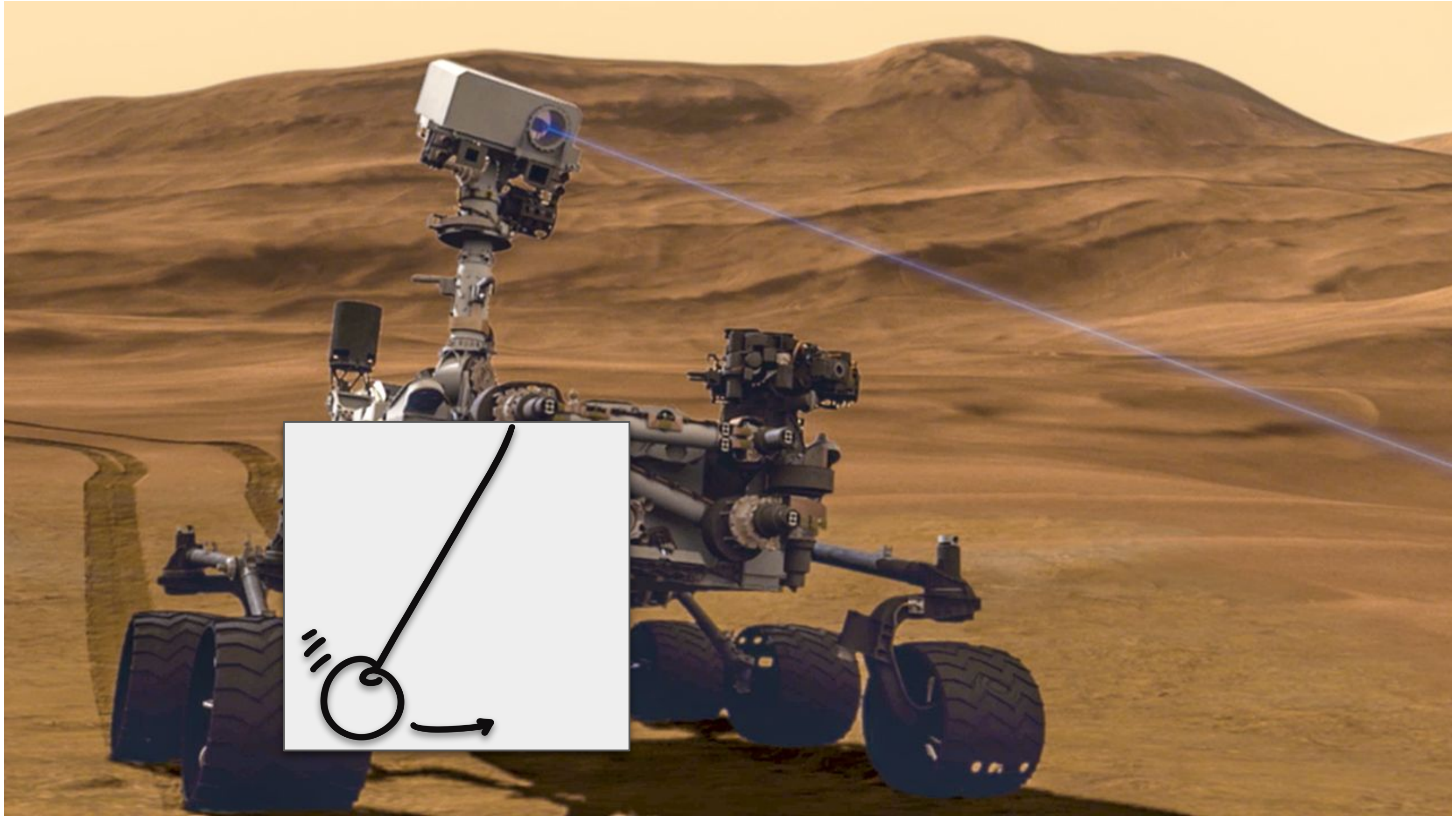
Things we'd like to infer about one pendulum:

- starting angle
- mass
- length

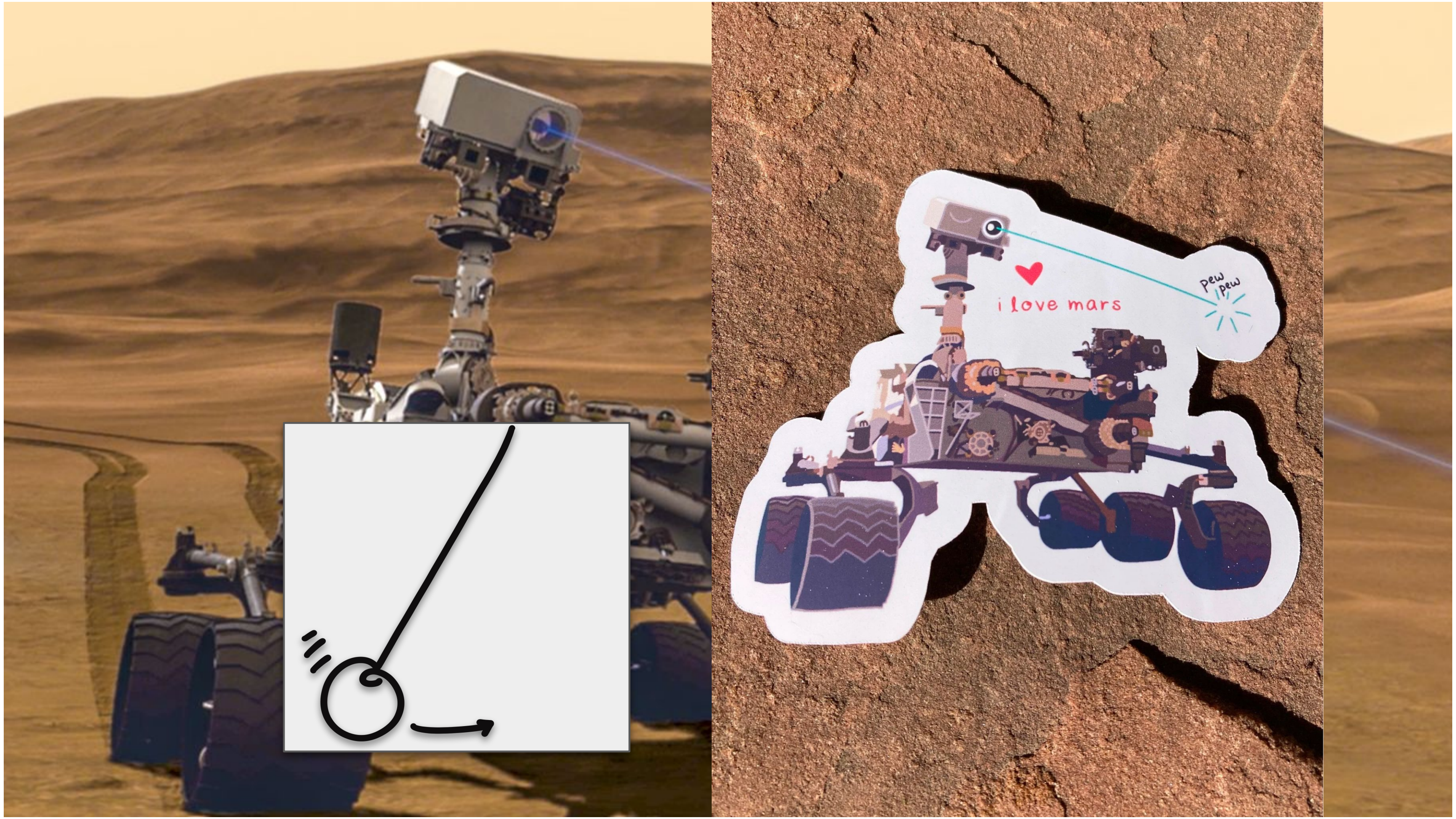
Things we'd like to infer using the ensemble of pendulums:

- acceleration due to gravity (a_g)

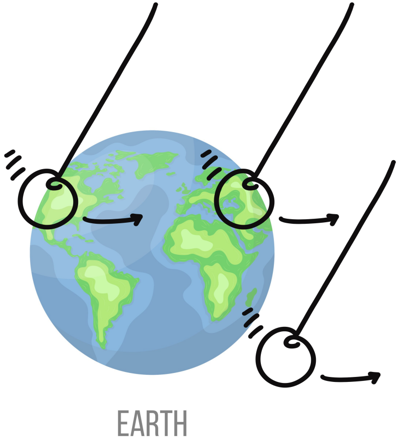
Meanwhile, on Mars…

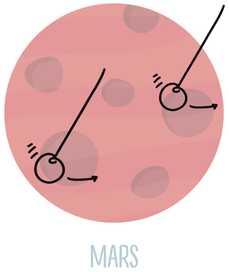


Meanwhile, on Mars…



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





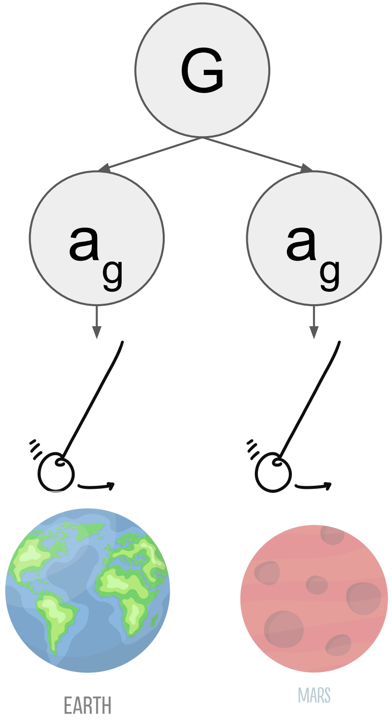
Things we'd like to infer about one pendulum:

- starting angle
- mass
- length

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

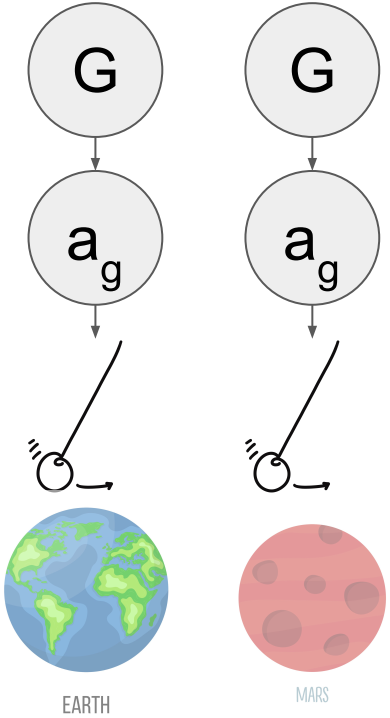
- acceleration due to gravity (a_g)
- Universal gravitational constant (G)

Hierarchical Bayesian Inference is a powerful tool for lending inference power across layers of params

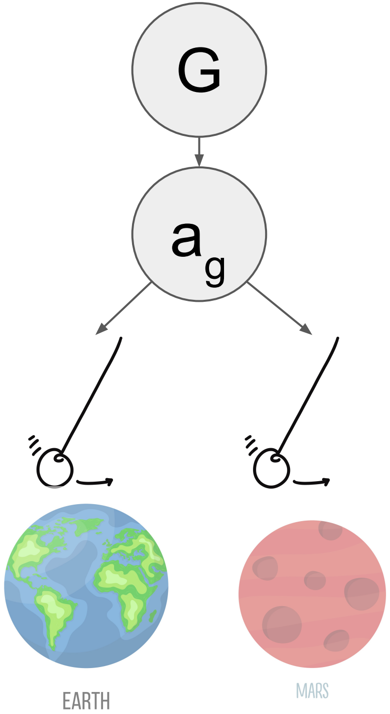


Hierarchical Bayesian Inference is a powerful tool for lending inference power across layers of params

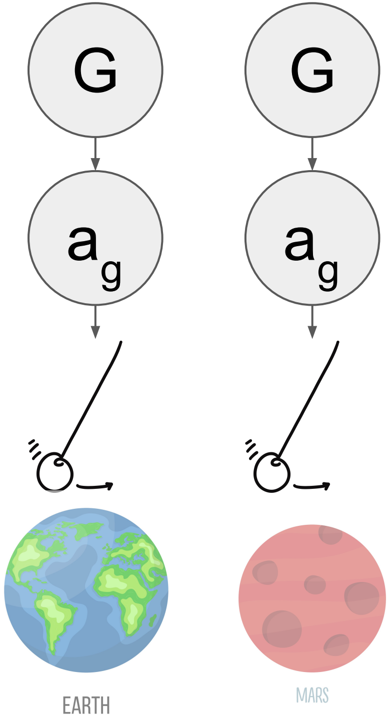
Independent / no pooling analysis



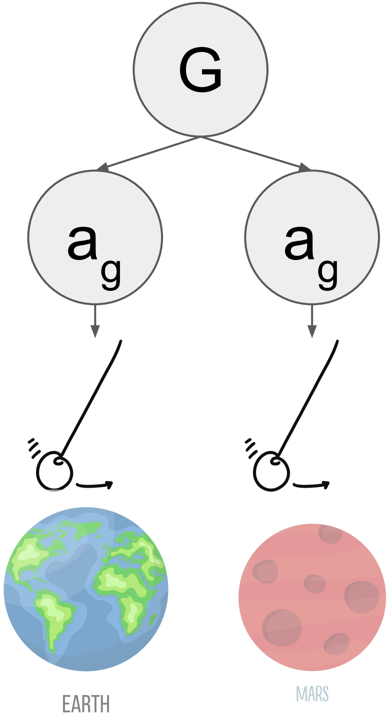
Co-dependent / full pooling analysis



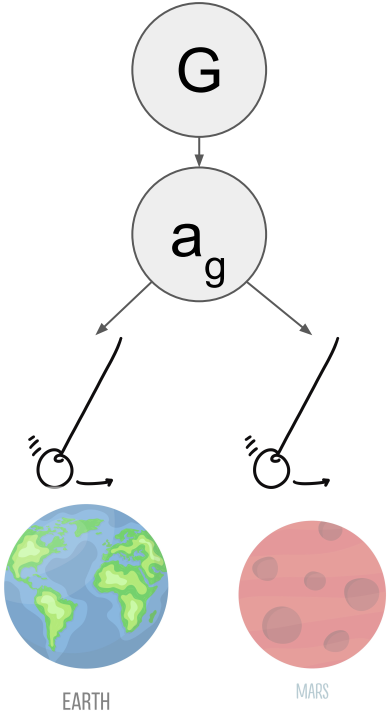
Independent / no pooling analysis



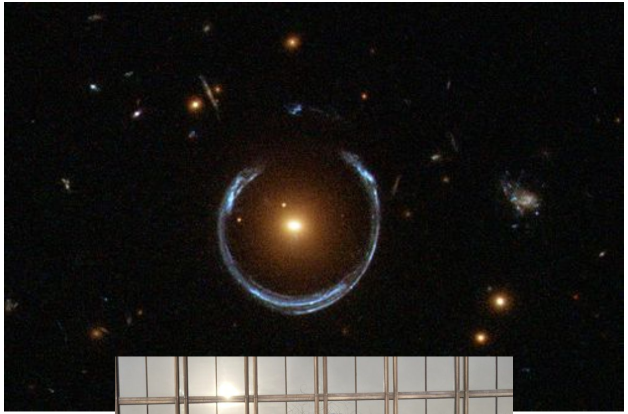
Hierarchical



Co-dependent / full pooling



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)

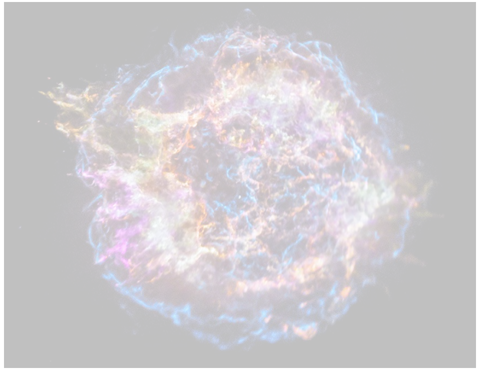
Things we'd like to infer using the ensemble of pendulums:

- Cosmological parameters (w_0)

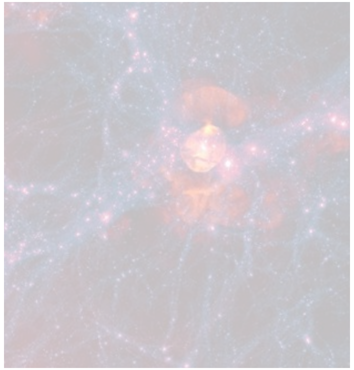
Active Galactic Nuclei

Mergers

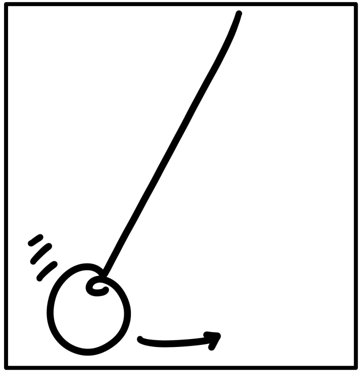
Chandra X-ray



Illustris



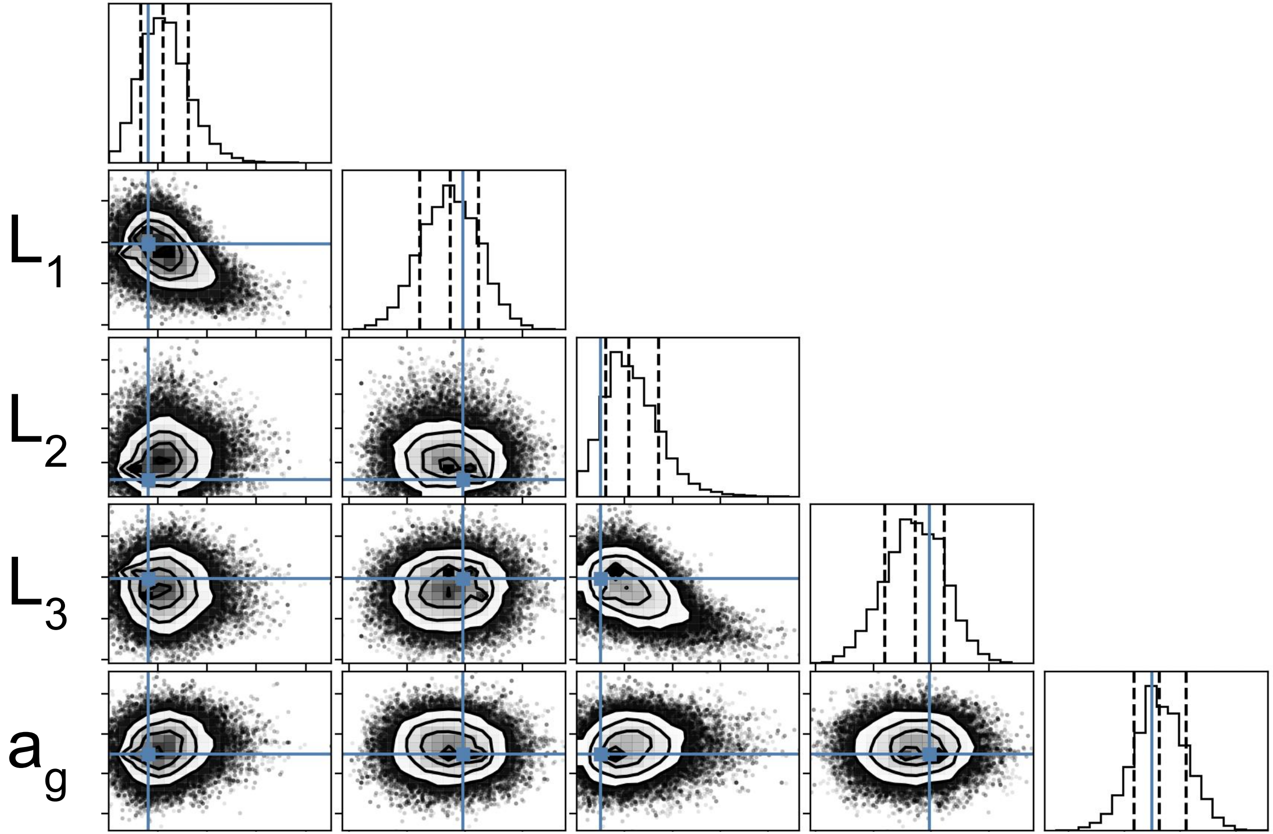
Benchmark



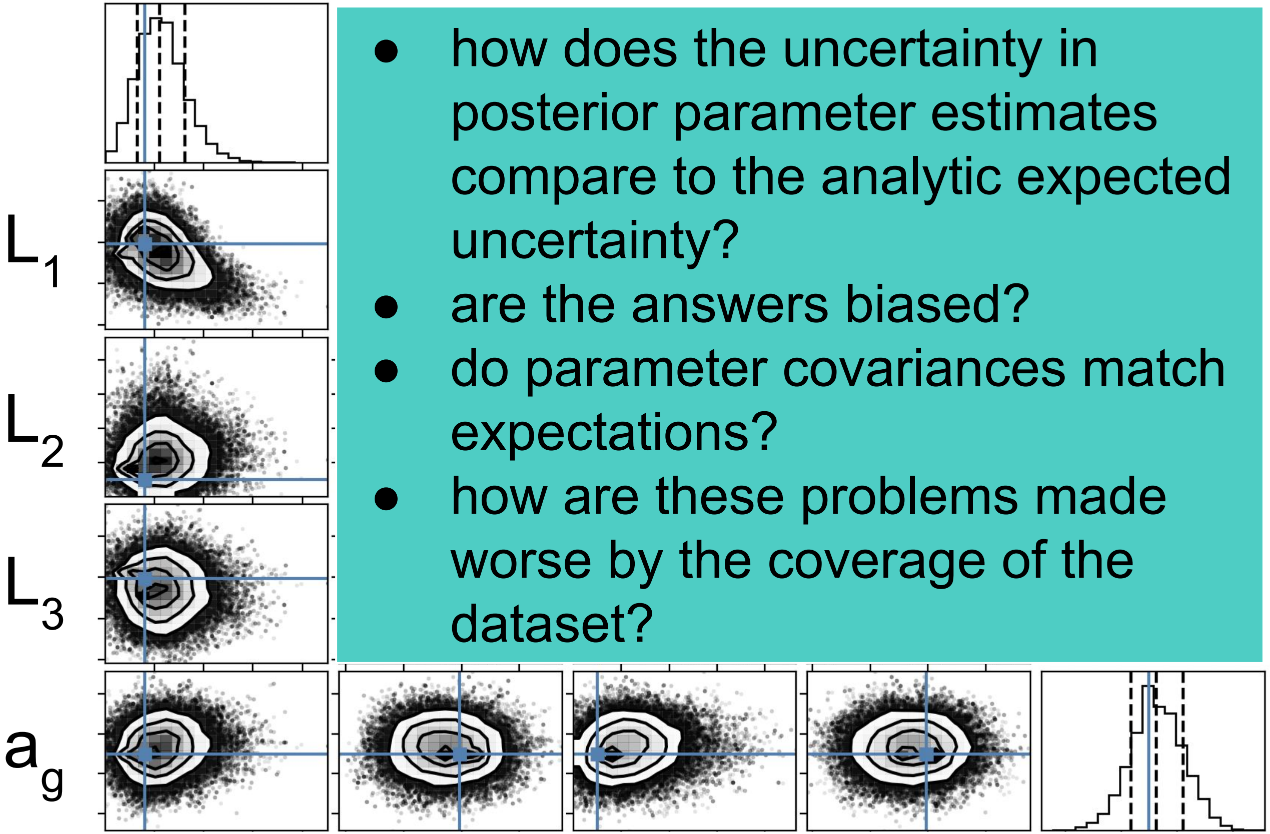
UQ

Hierarchical Inference

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

Goals at Fermilab

- Mentoring and group organization
- Software development, launching my own package through Deepskies github
- Collaborative research projects in next year (neurIPS)

Vision for the future

- Live in Colorado
- Find a position (industry or research) that aligns with my values

Values:

Science collaborations and community

Machine and statistical learning for addressing scientific questions

Opportunity and support to become a group leader

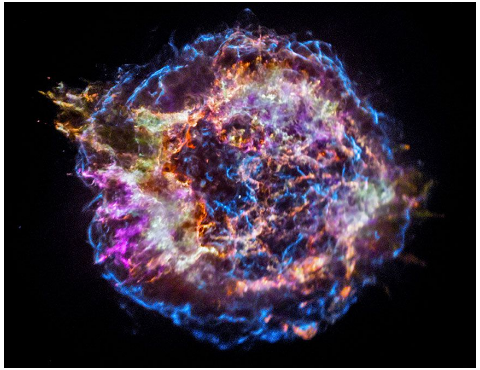
Shorter term workstyle

Storytelling

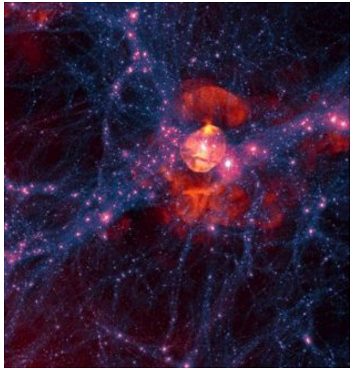
Active Galactic Nuclei

Mergers

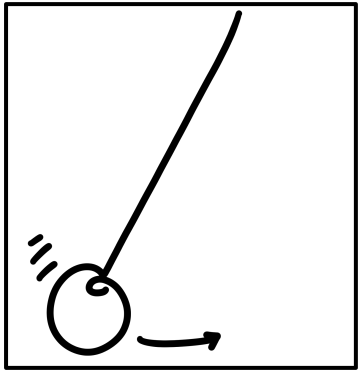
Chandra X-ray



Illustris



Benchmark



UQ

Hierarchical Inference