Jet Propulsion Laboratory
California Institute of Technology

Anomaly Detection and Explanation in Galaxy Observations from the Dark Energy Survey

Kiri Wagstaff[1],
Eric Huff[1],
Umaa Rebbapragada[1],
and Tim Eifler[2]

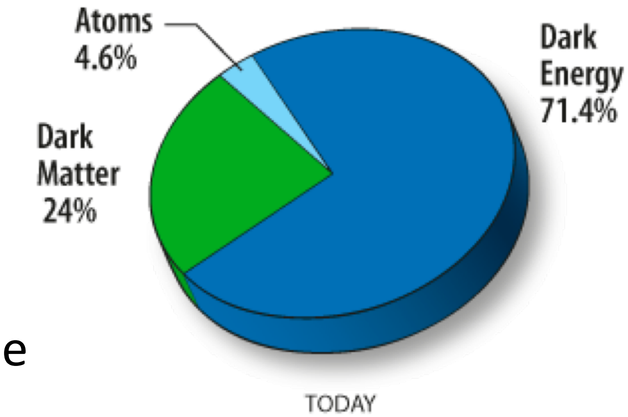
[1]Jet Propulsion Laboratory,
California Institute of Technology
[2]University of Arizona

June 26, 2019

This work was performed in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

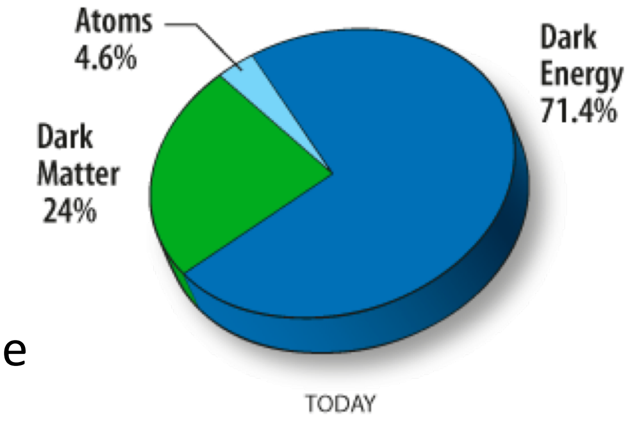
Dark Energy Survey: Motivation

- Quest: Understand our universe (formation and evolution), using a cosmological model characterized by 6-10 parameters
  - E.g., amount of dark energy, dark matter
  - Dark energy drives the expansion of the universe
  - Dark matter drives the clustering of galaxies

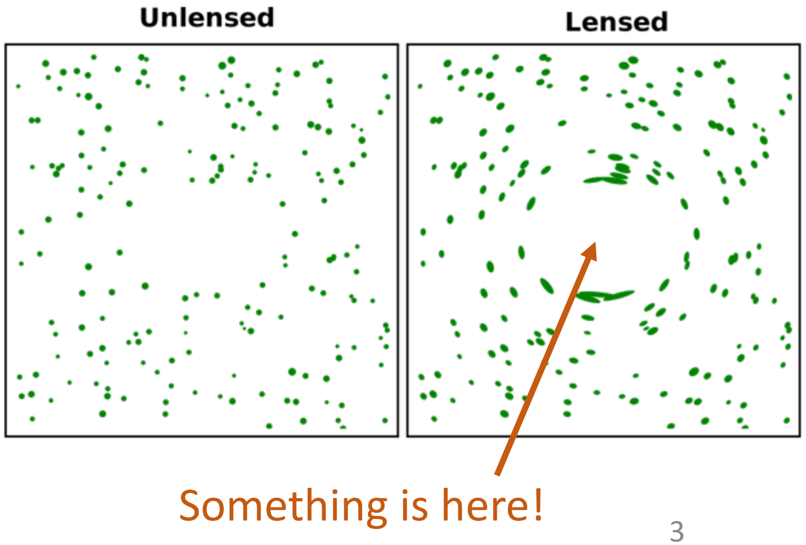


Dark Energy Survey: Motivation

- Quest: Understand our universe (formation and evolution), using a cosmological model characterized by 6-10 parameters
  - E.g., amount of dark energy, dark matter
  - Dark energy drives the expansion of the universe
  - Dark matter drives the clustering of galaxies

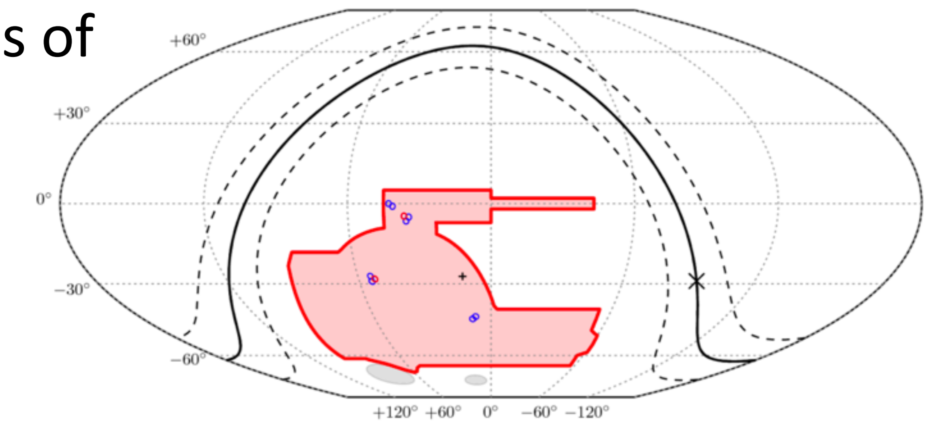


- How to determine the amount of dark energy and dark matter? We can't observe them directly.
  - Infer dark matter from weak lensing galaxy observations
    - Dark matter magnifies + distorts background galaxy shapes
  - Infer dark energy from changes in dark matter distribution over time (distance)
  - Subtle effects; require millions of galaxies to reliably estimate



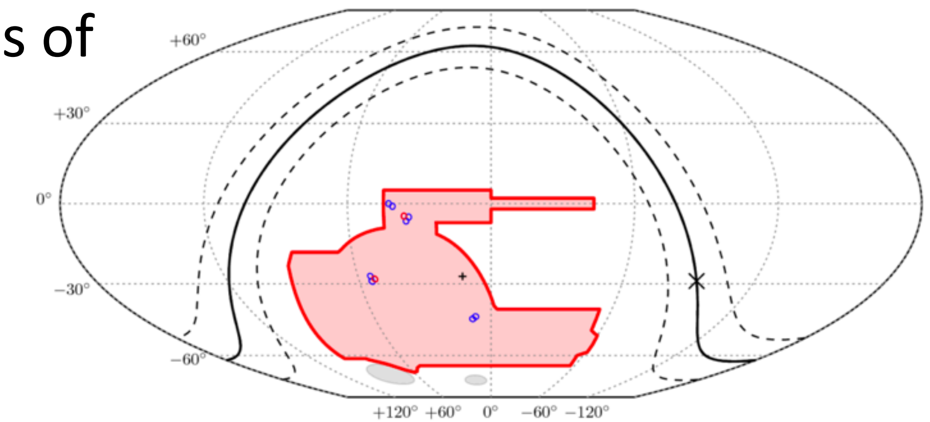
Dark Energy Survey (DES)

- Surveyed 5000 square degrees of southern sky over 6 years (Aug. 2013 - Jan. 2019)
  - 4-m telescope in Chile
- 400M objects (~310M galaxies)
  - But noise/artifacts will pollute dark energy, dark matter estimates



Dark Energy Survey (DES)

- Surveyed 5000 square degrees of southern sky over 6 years (Aug. 2013 - Jan. 2019)
  - 4-m telescope in Chile
- 400M objects (~310M galaxies)
  - But noise/artifacts will pollute dark energy, dark matter estimates



- Our goal: Identify outliers in DES catalog to
  1. Find, filter, and understand artifacts, which can
     1. Inform improvements to the DES processing pipeline and
     2. Yield better estimates of cosmological model parameters (e.g., amount of dark energy and dark matter)
  2. Discover new scientific phenomena (what's out there?)

DES Object Catalog

Observe | Find sources | Model sources | Extract features

- Observe: Collect sky images at 4 bands
  - 10 exposures (90 s) per band (different dates)
- Find sources (Sextractor) and match to reference catalogs
- Model sources (NGMIX)
- Extract features: id, RA, Dec, G, R, I , Z, etc.
- Apply recommended quality filters (discards 2/3 of objects)
  - And yet… spoiler: artifacts remain

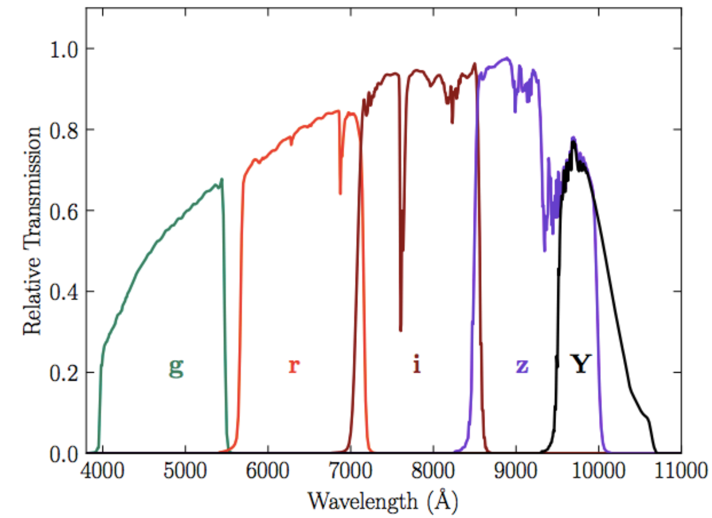


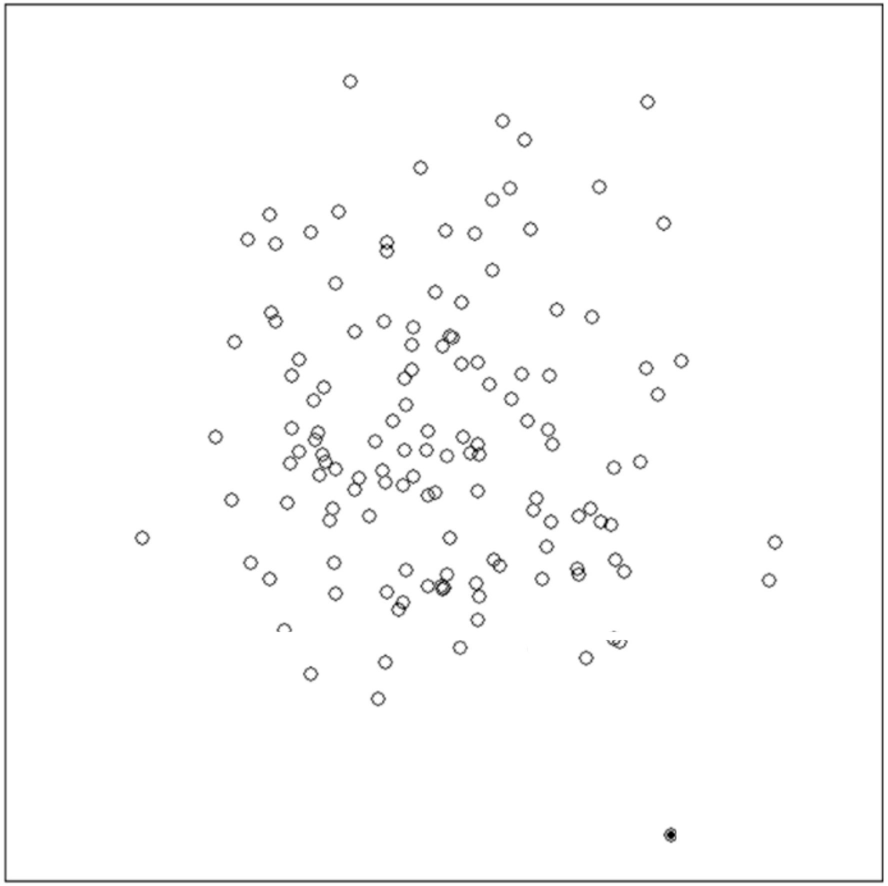
Figure 1. The DES g, r, i, z and Y Standard Bandpasses. These curves include atmospheric transmission and instrumental response. Details can be found in Burke et al. (2018).

Outlier Detection Methods

- Isolation Forest [Liu et al., 2008]
  - Rank all items by their isolation scores
- DEMUD: Discovery via Eigenbasis Modeling of Uninteresting Data [Wagstaff et al., 2013]
  - Select diverse set of unusual objects using incremental Singular Value Decomposition (SVD)
  - Provide explanations for each object's selection (e.g., unusual feature values)

Isolation Forest

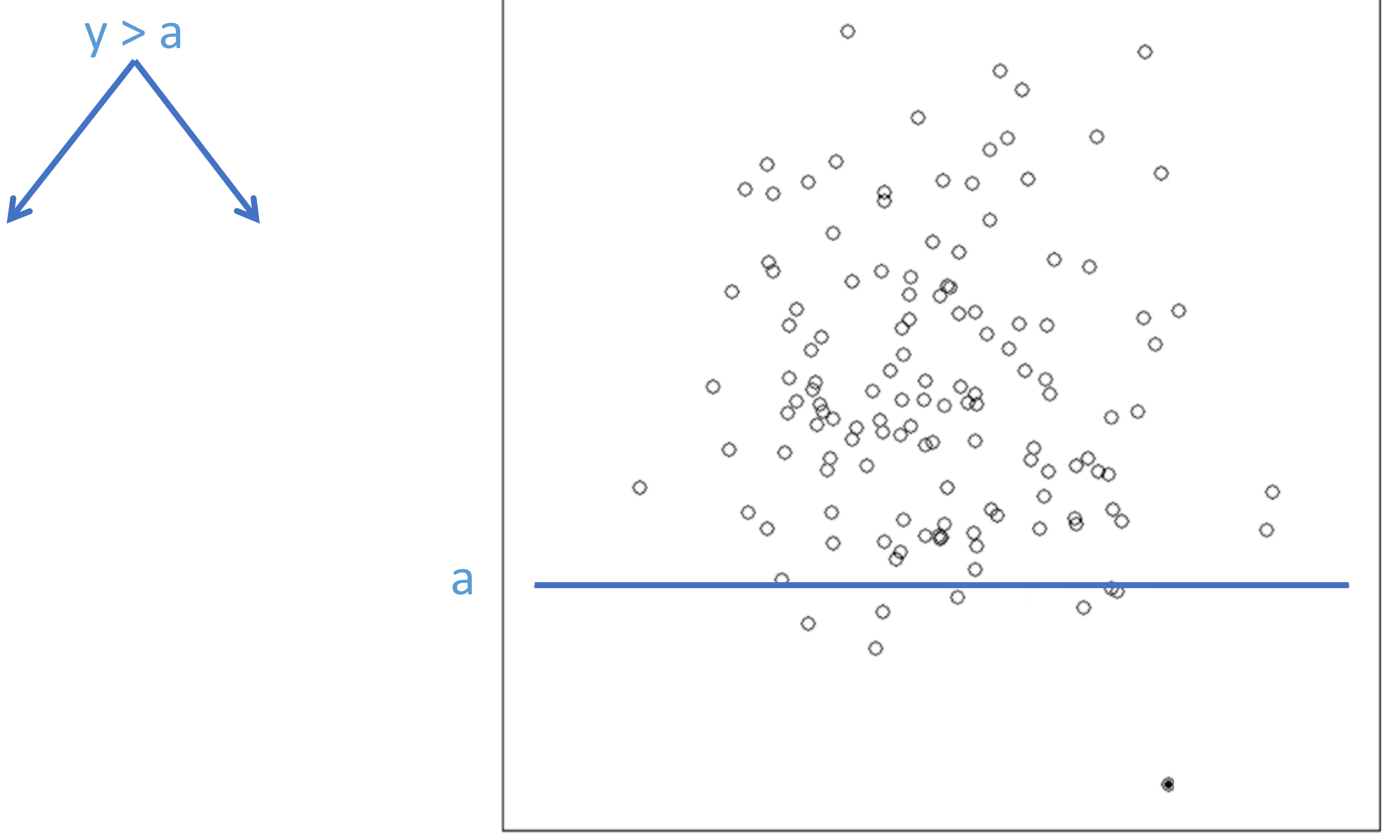
- Construct forest of random decision trees
- Compute isolation score for each item: average depth in trees



[Liu et al., 2008]

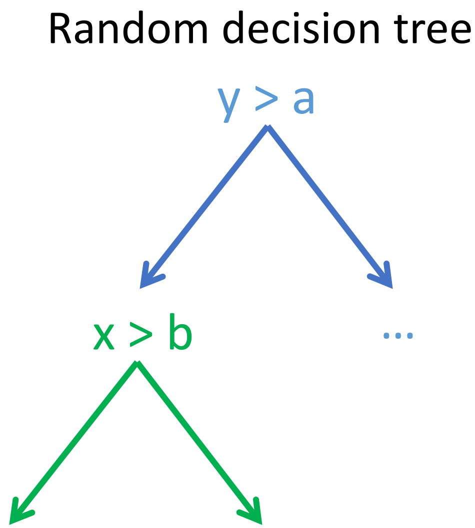
Isolation Forest

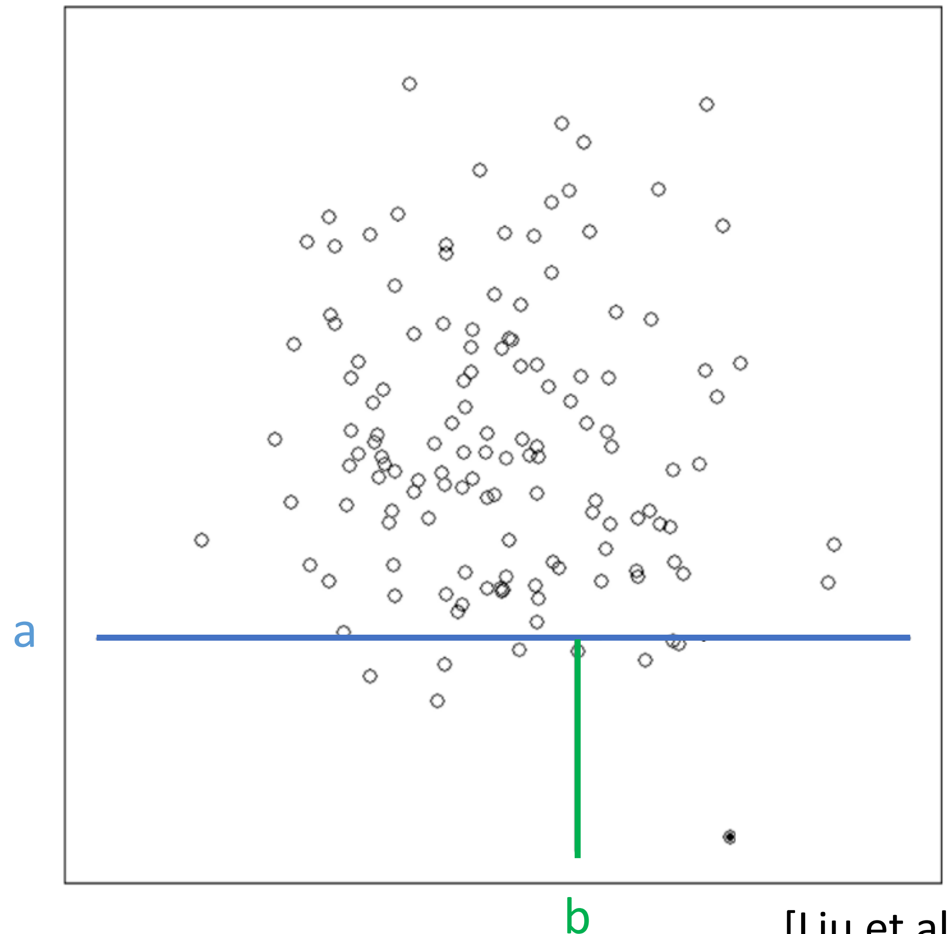
Random decision tree



[Liu et al., 2008]

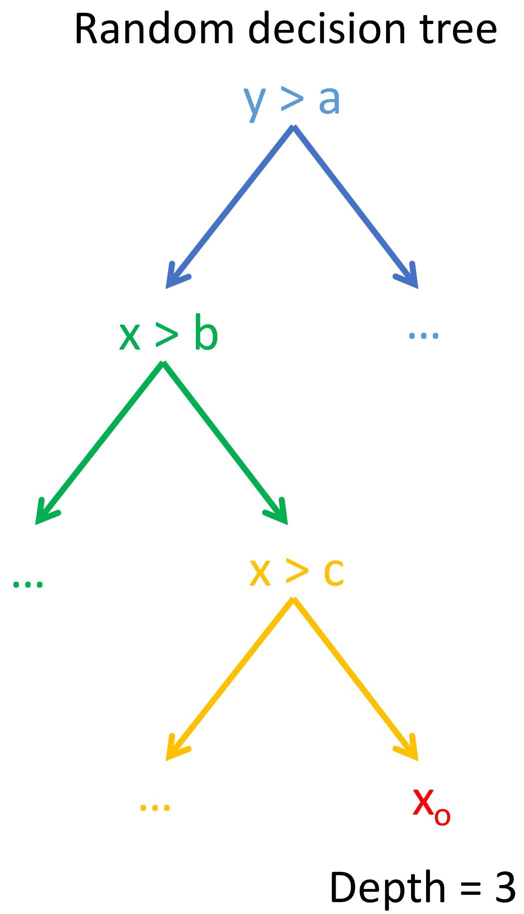
Isolation Forest

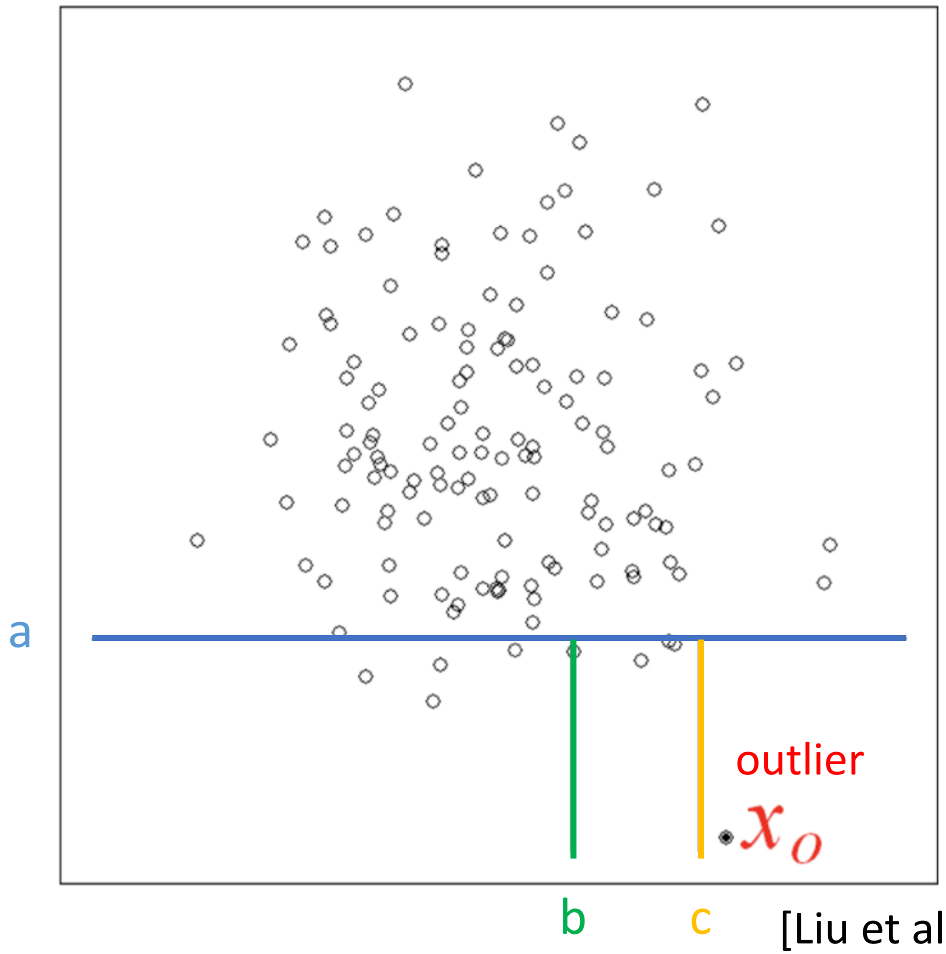




[Liu et al., 2008]

Isolation Forest

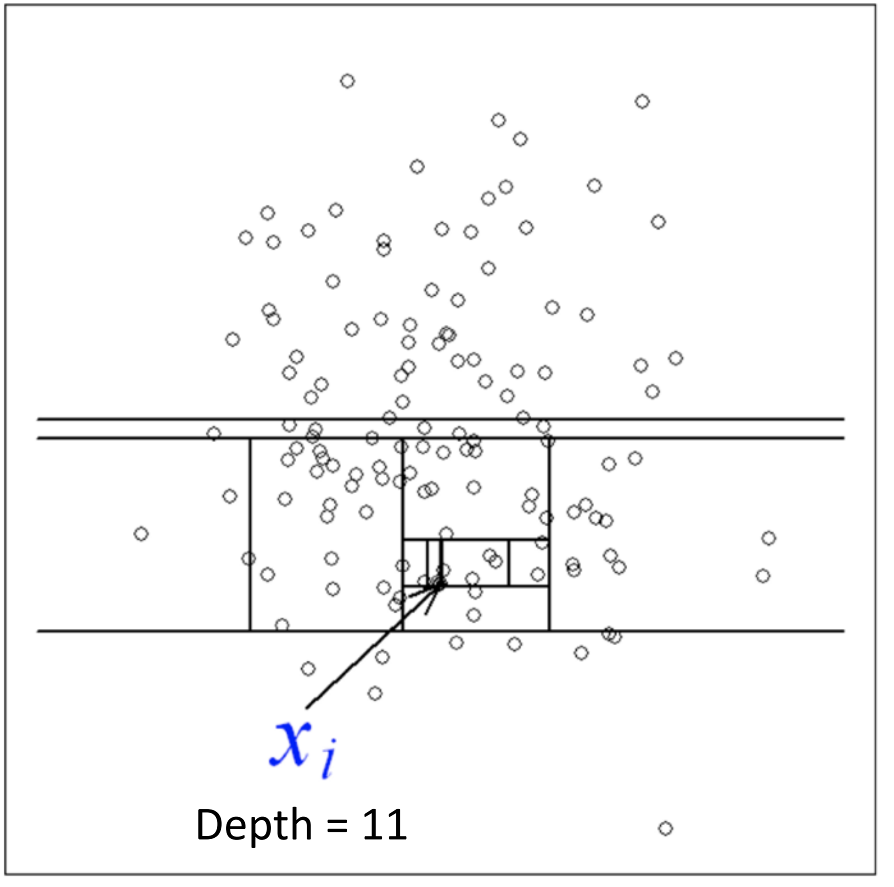




[Liu et al., 2008]

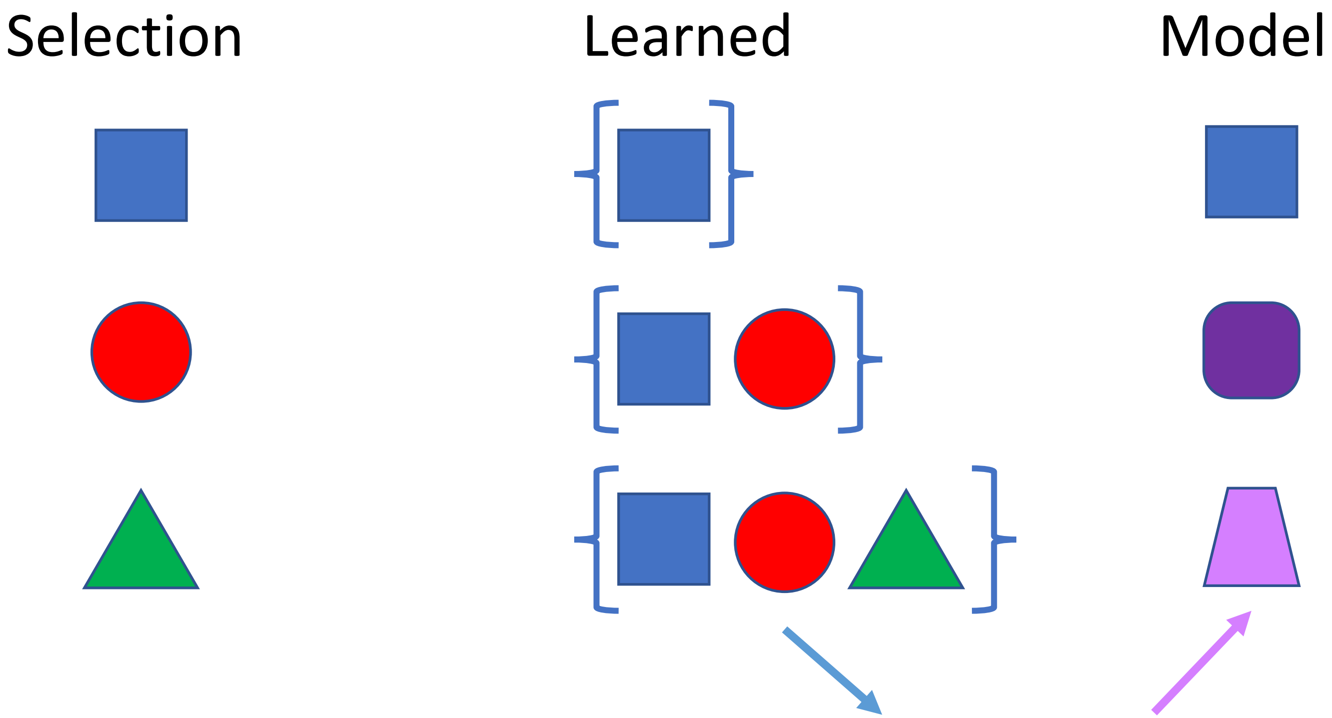
Isolation Forest

- Other items take more random splits to isolate
- Outliers: low average depth across all trees in the forest



DEMUD

- Incrementally growing model (Singular Value Decomposition) of what has been selected (learned) [Wagstaff et al., 2013]
- Select "most difficult to model" item at each step



$$D = U\Sigma V^T$$

DES Outliers – What have we found?

- Data set: 12M objects observed by DES
  - R-band luptitude + 3 colors: (G-R, I-R, Z-R)
- Select top 1000 outliers with each algorithm
- Browse results using interactive web-based Outlier Explorer
- Classify outliers by category:
  - Detector error
  - Model-fitting error
  - Human-made transient
  - Astrophysical transient
  - Scientifically interesting – consider follow-up observations

Outlier Explorer

Sky location and object id

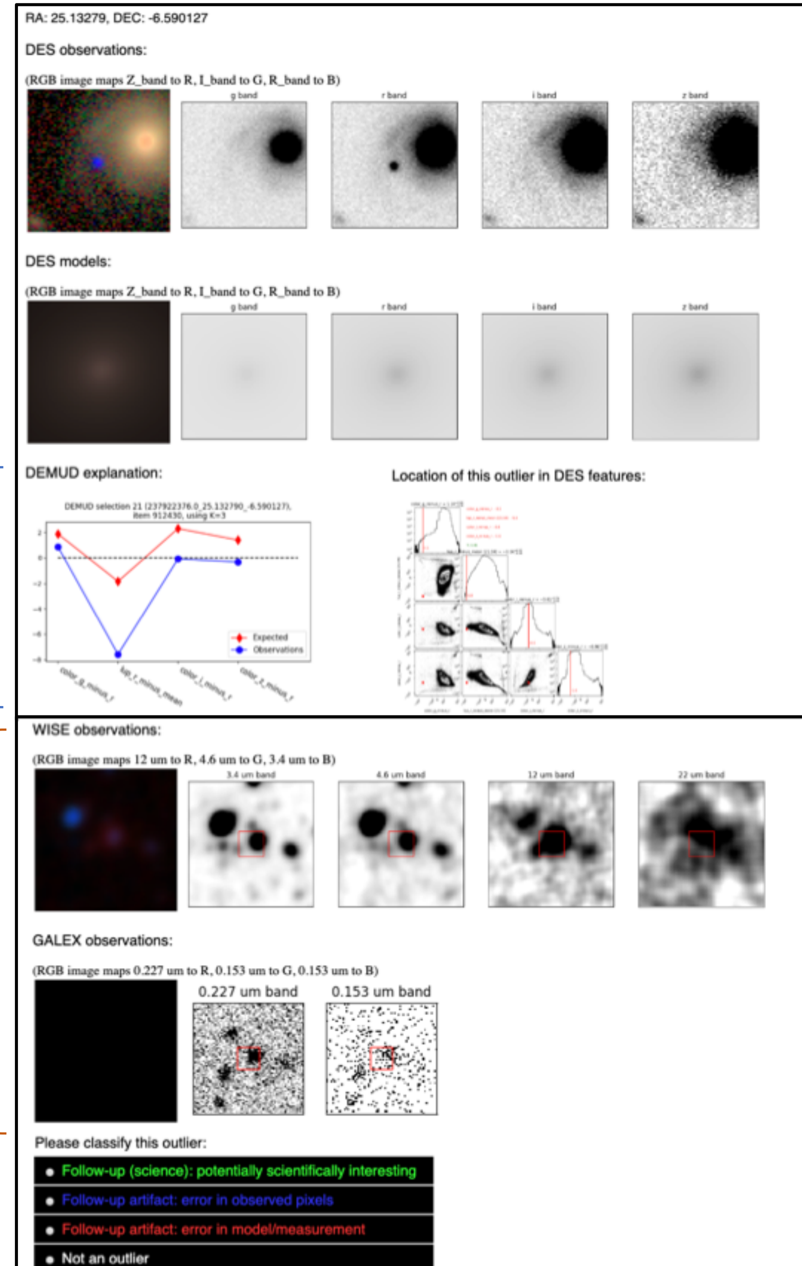
Observations

Model

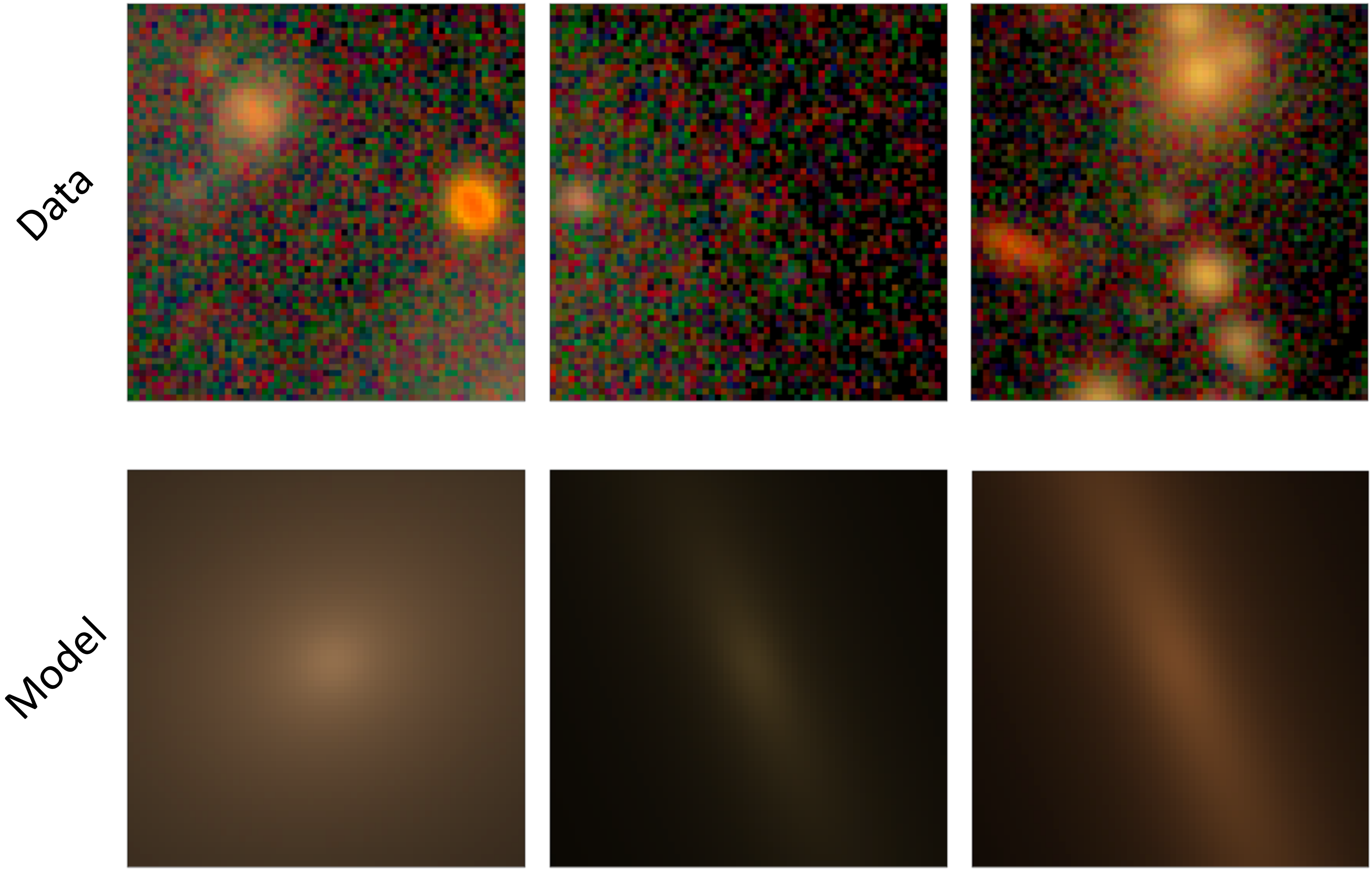
Explanations

Coincident observations in other surveys

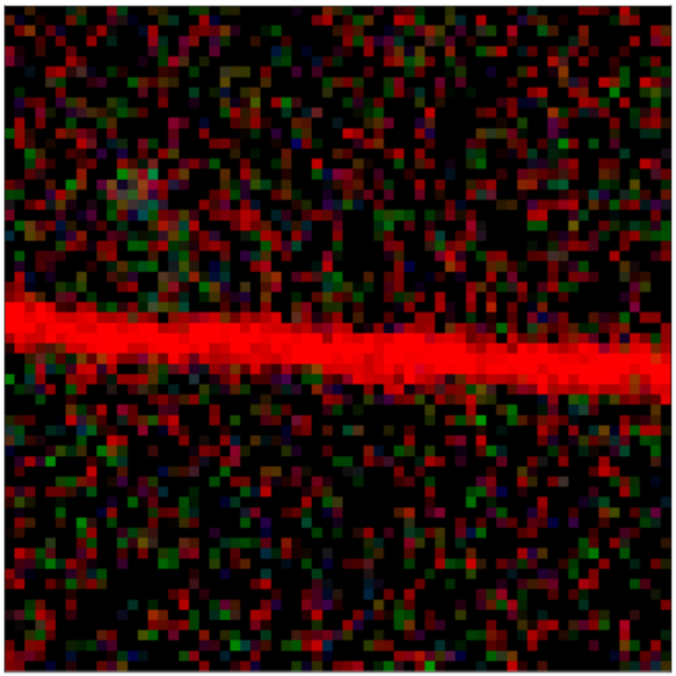
Classify outlier

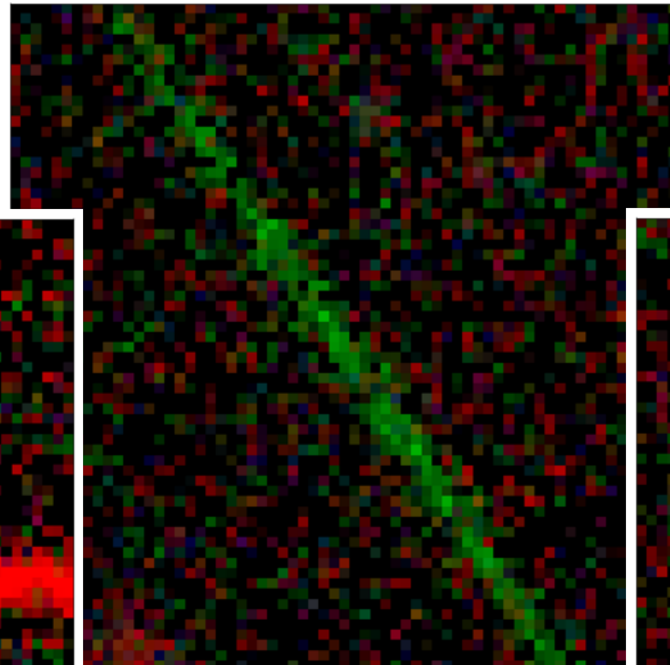


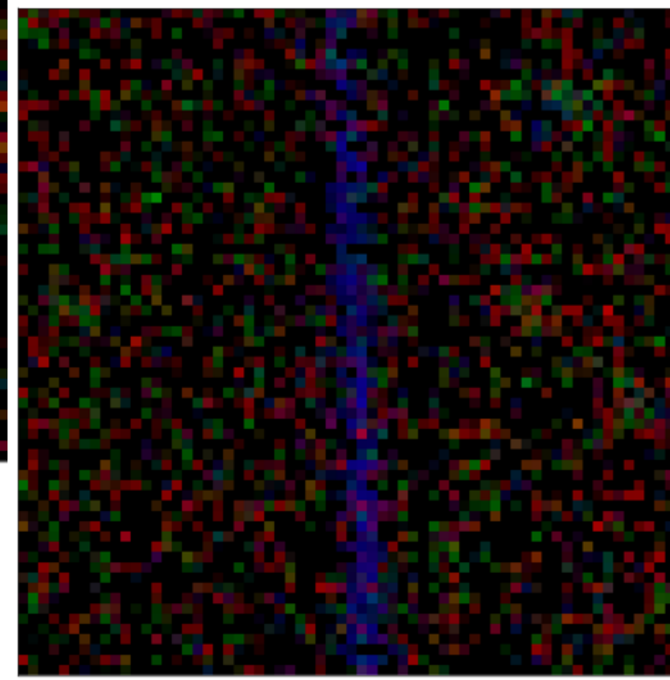
DES Outliers: Model Errors



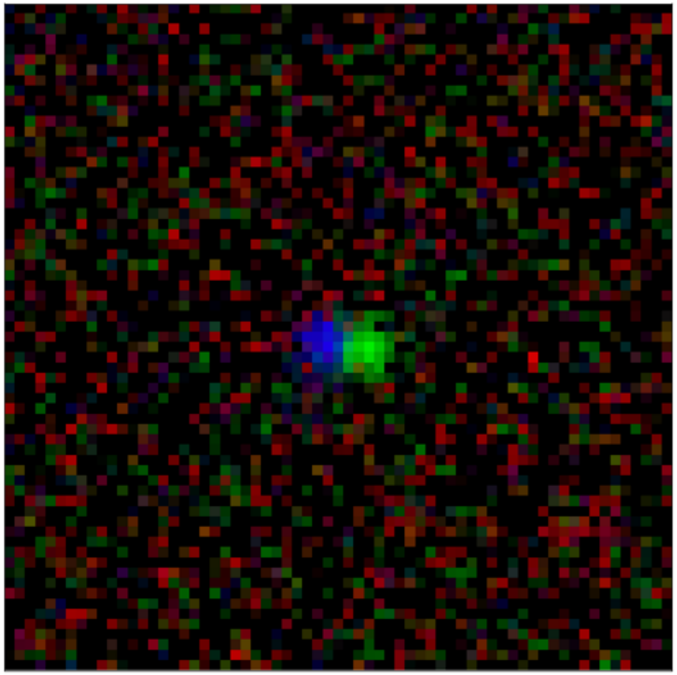
DES Outliers: Fast-moving aircraft

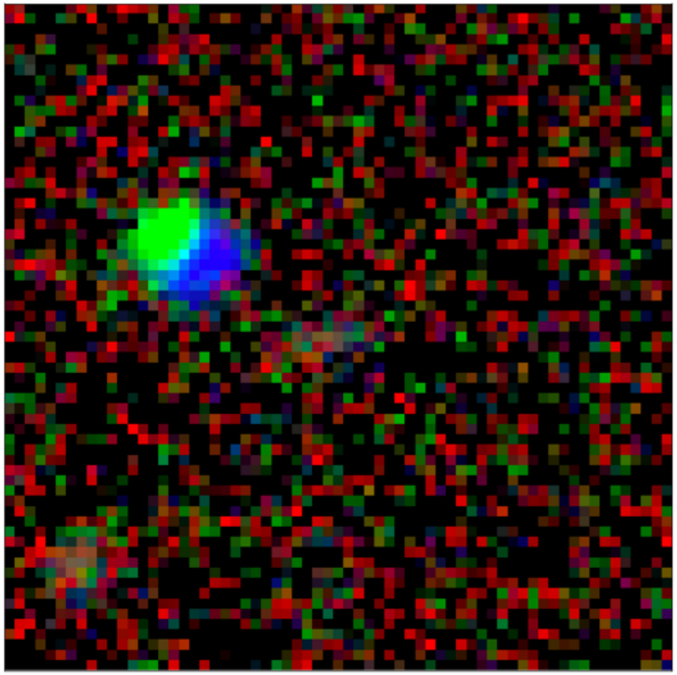






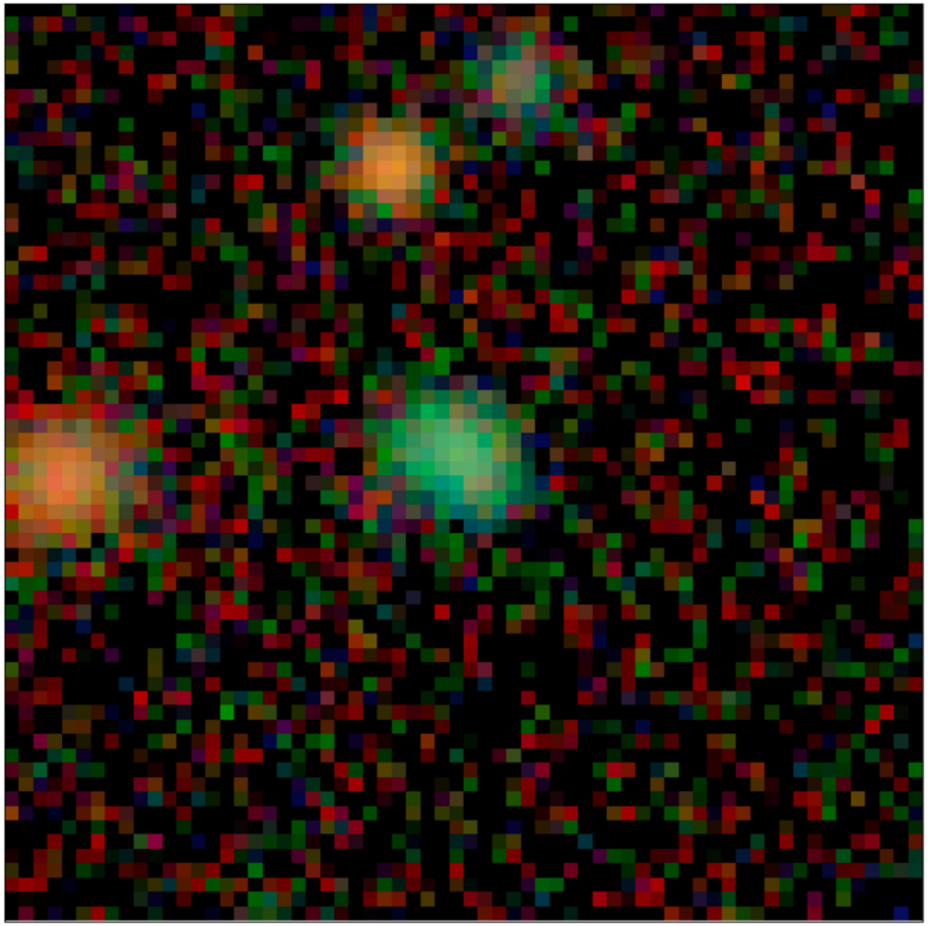
DES Outliers: Asteroids?



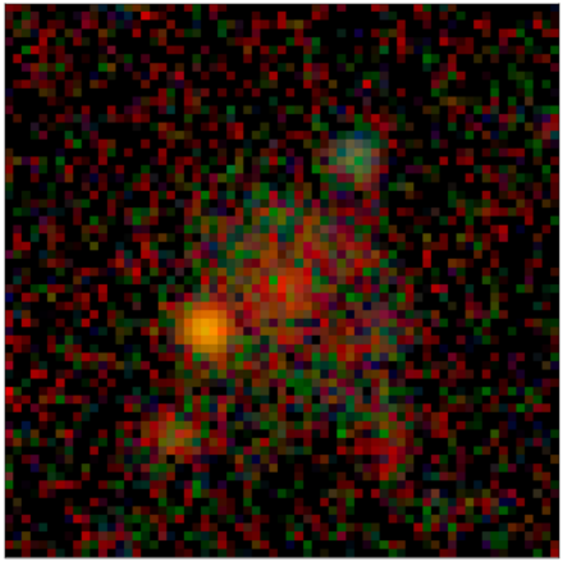


DES Outliers: “Green Pea” Galaxy

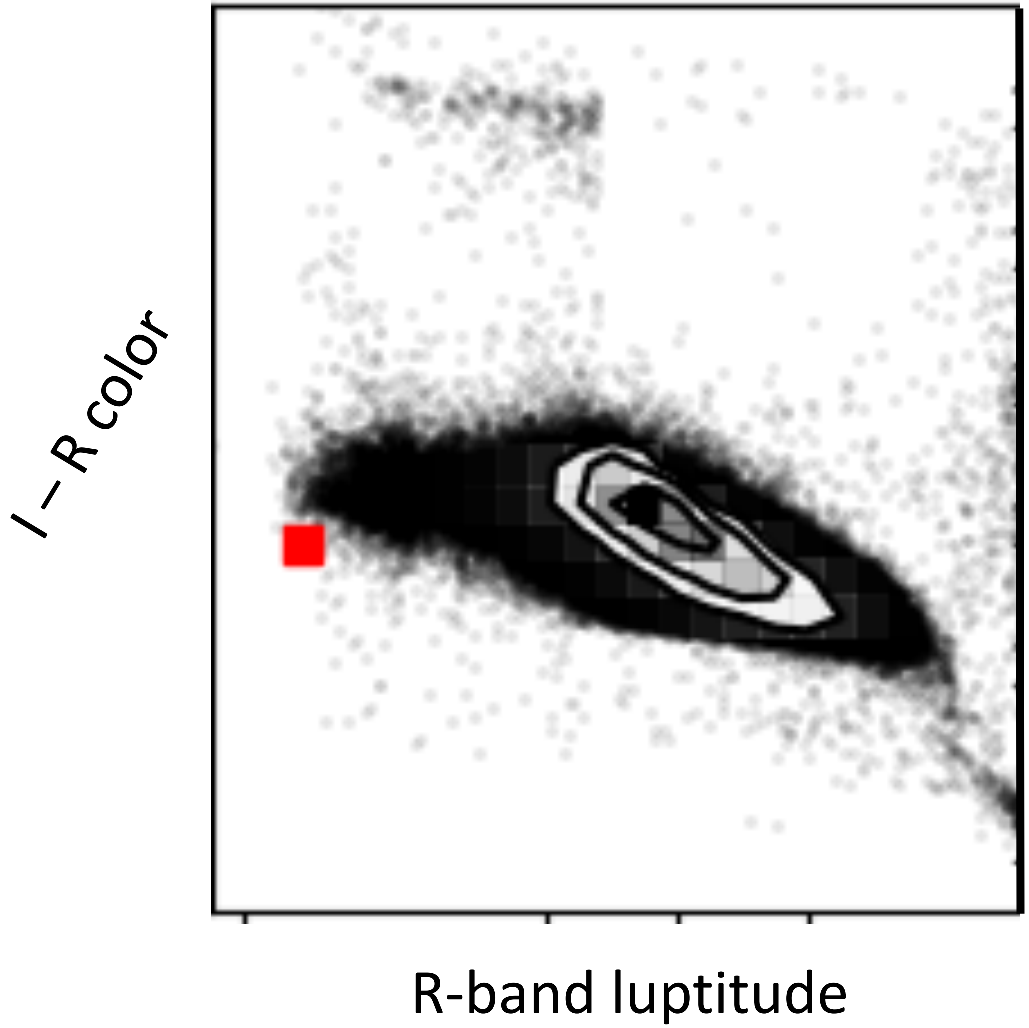
- Discovered by SDSS citizen scientists in 2007
- Very low mass with high star-formation rate
- Common in the early universe, but rare today



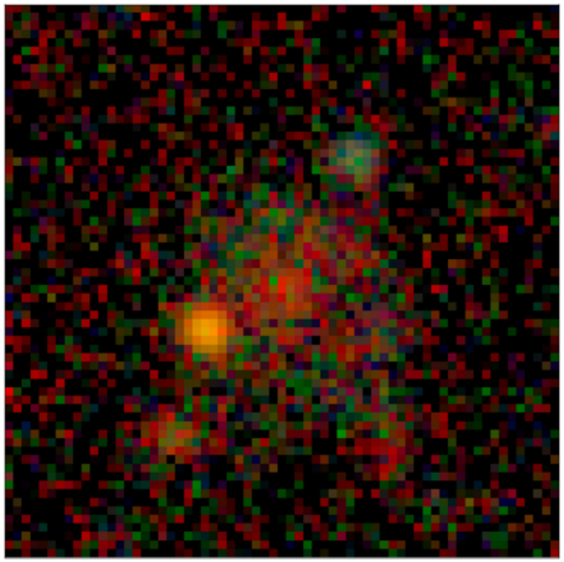
Why is this object an outlier?



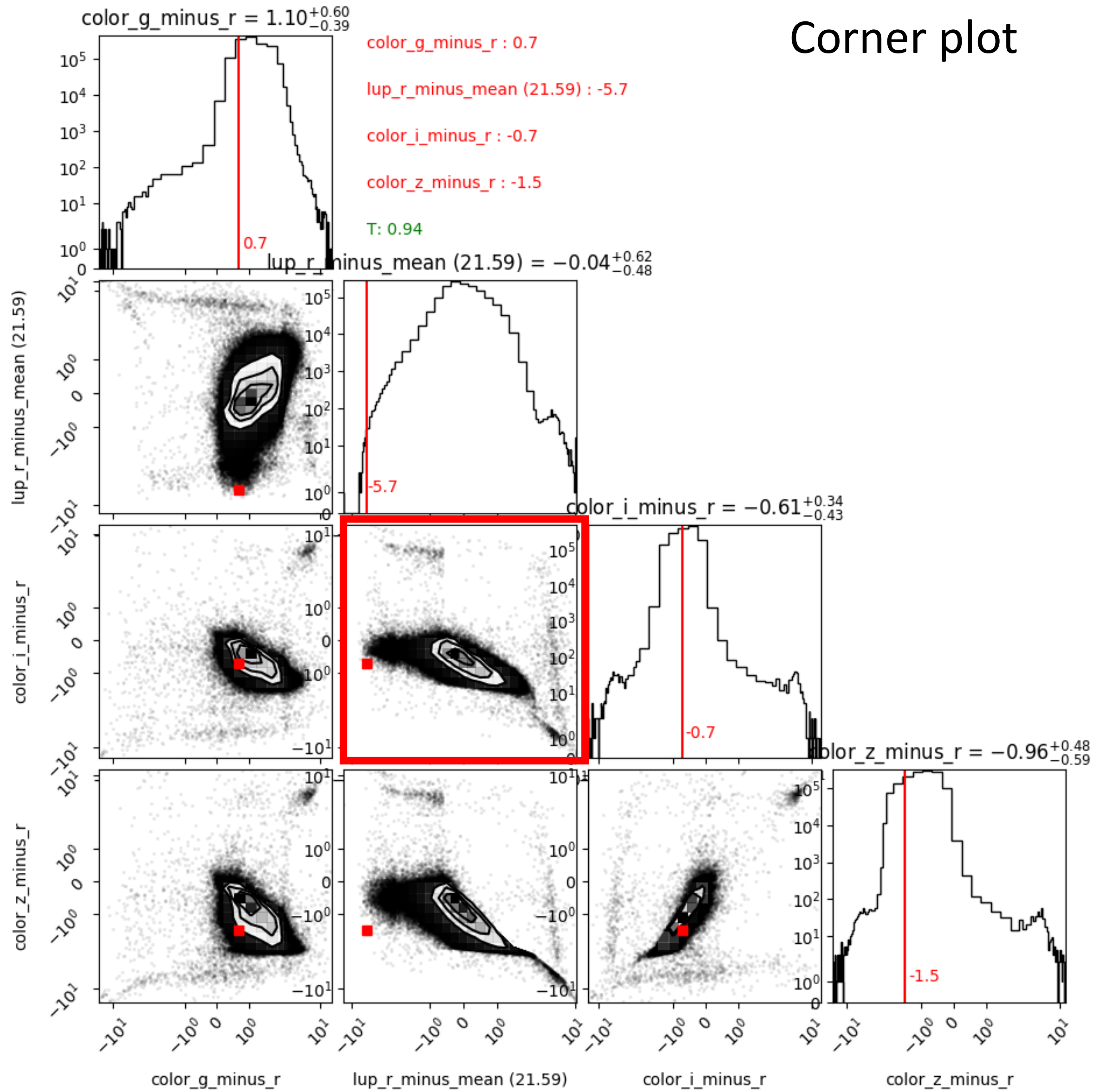
Data distribution plot



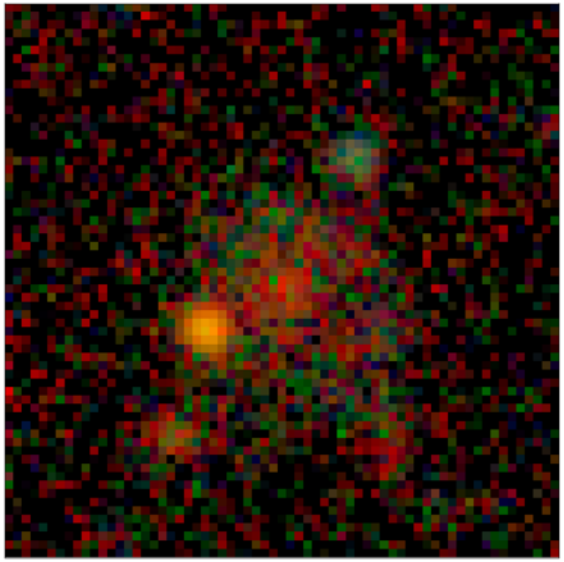
Why is this object an outlier?



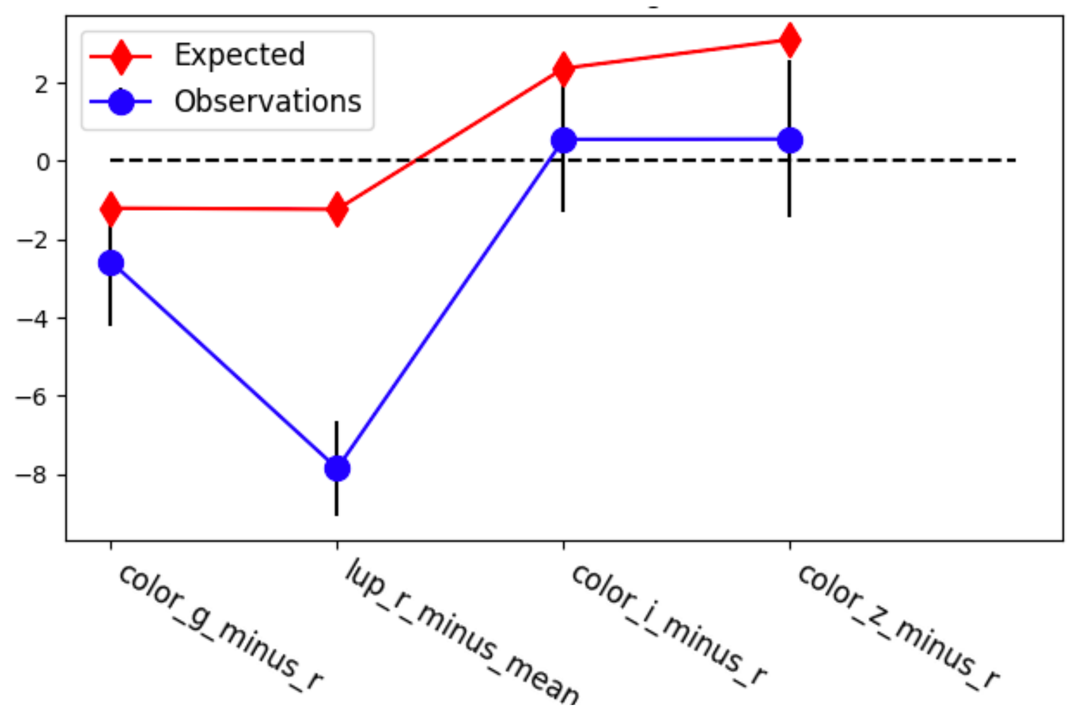
Corner plot

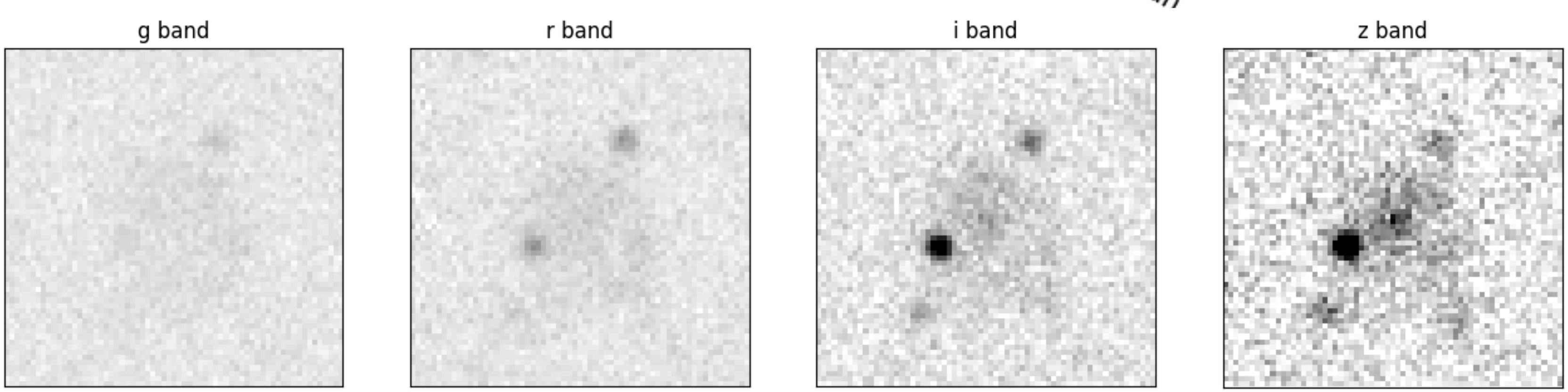


Why is this object an outlier?



DEMUD Explanation





Takeaway Points

- Goal: Identify outliers in DES catalogs to
  1. Find, filter, and understand artifacts
     … to improve estimates of dark energy and dark matter
  2. Discover new kinds of objects
- Outlier detection methods
  - Isolation Forest
  - DEMUD: Discovery via Eigenbasis Modeling of Uninteresting Data
- Explanations aid in classifying and making use of outliers
- In progress: Review and publish DES outlier catalogs
- Same techniques will benefit future large surveys (e.g., LSST, WFIRST, SPHEREx)
  - We've also applied them to e.g., Kepler light curves, UKIRT catalogs

Thank you: JPL R&TD program (funding)
and Bryan Scott of the Univ. of CA, Riverside (initial exploratory research)