Outlier Detection in DESI Data

presentation for the LBNL HEP ML Group

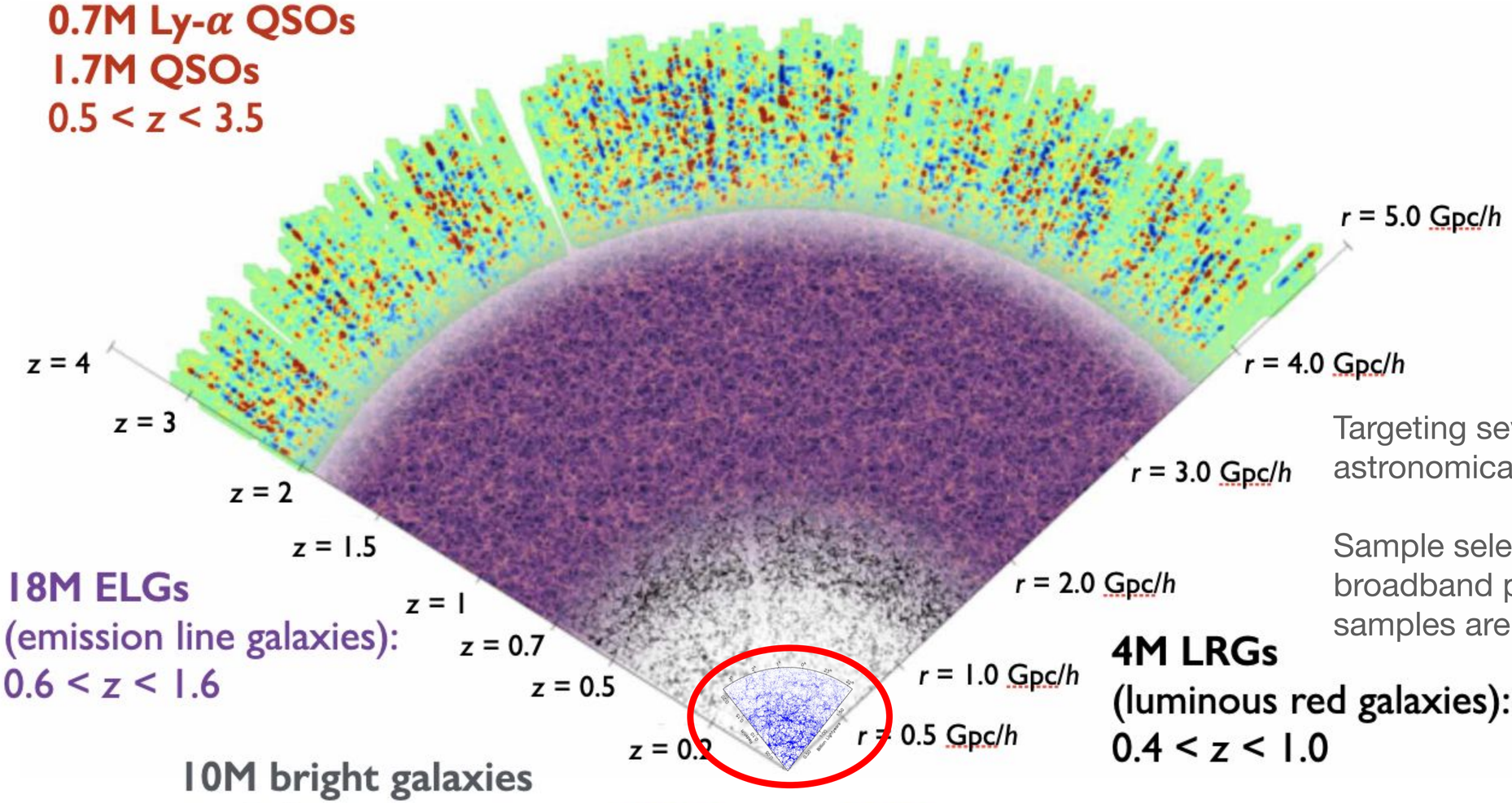
Alex Kim

DARK ENERGY SPECTROSCOPIC INSTRUMENT

U.S. Department of Energy Office of Science

Dark Energy Spectroscopic Instrument

DESI: ~35,000,000 galaxies: 2021-2026



Targeting several classes of astronomical objects

Sample selection based on broadband photometry so samples are not pure

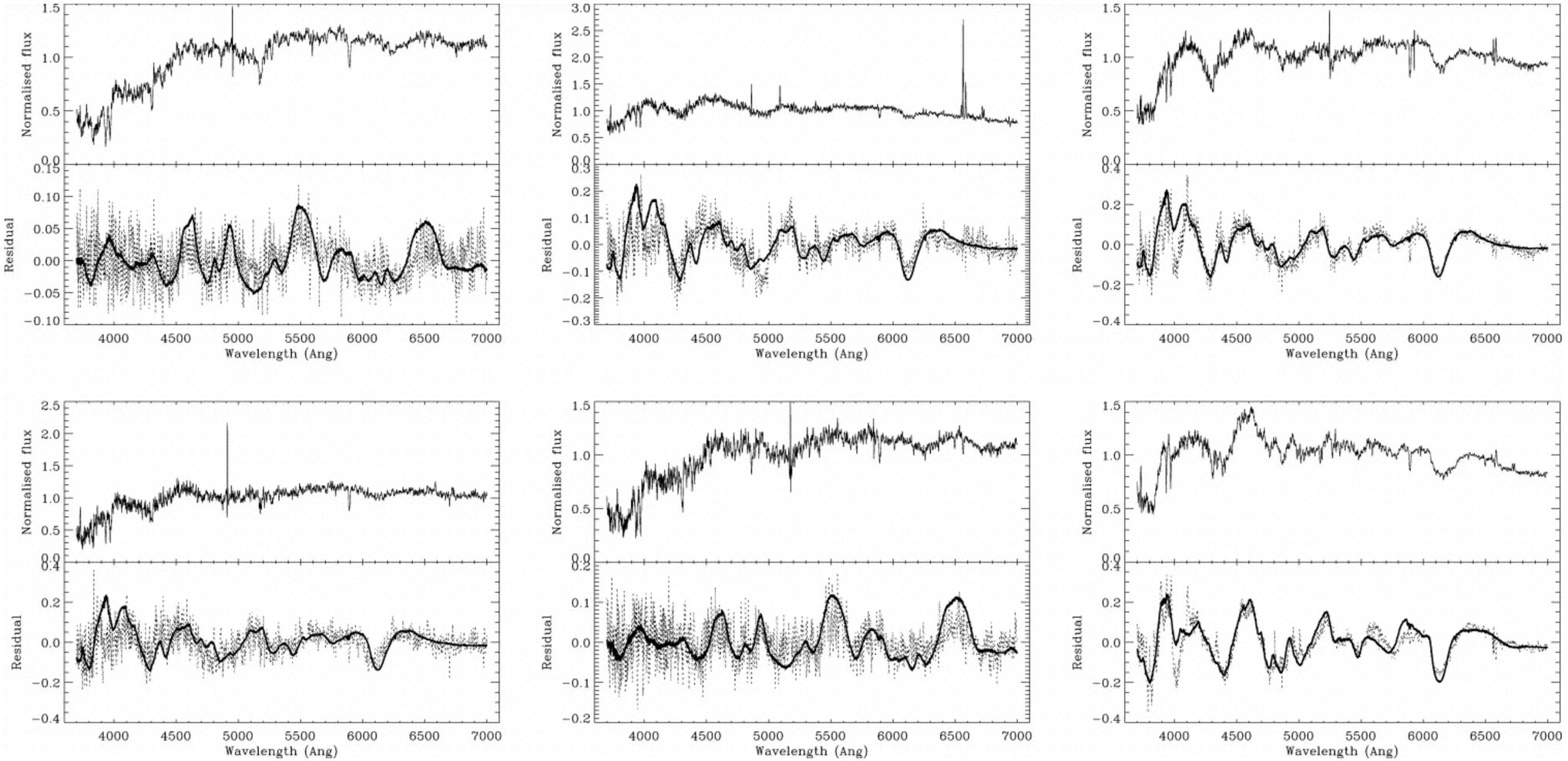
based on from BenZvi

Observed Specta May Not Quite Look Like a Galaxy

“Strange” Galaxy Spectra -> Transient Discovery in SDSS

Top:
Observed Spectrum

Bottom:
Observed Spectrum
- galaxy template
SN template

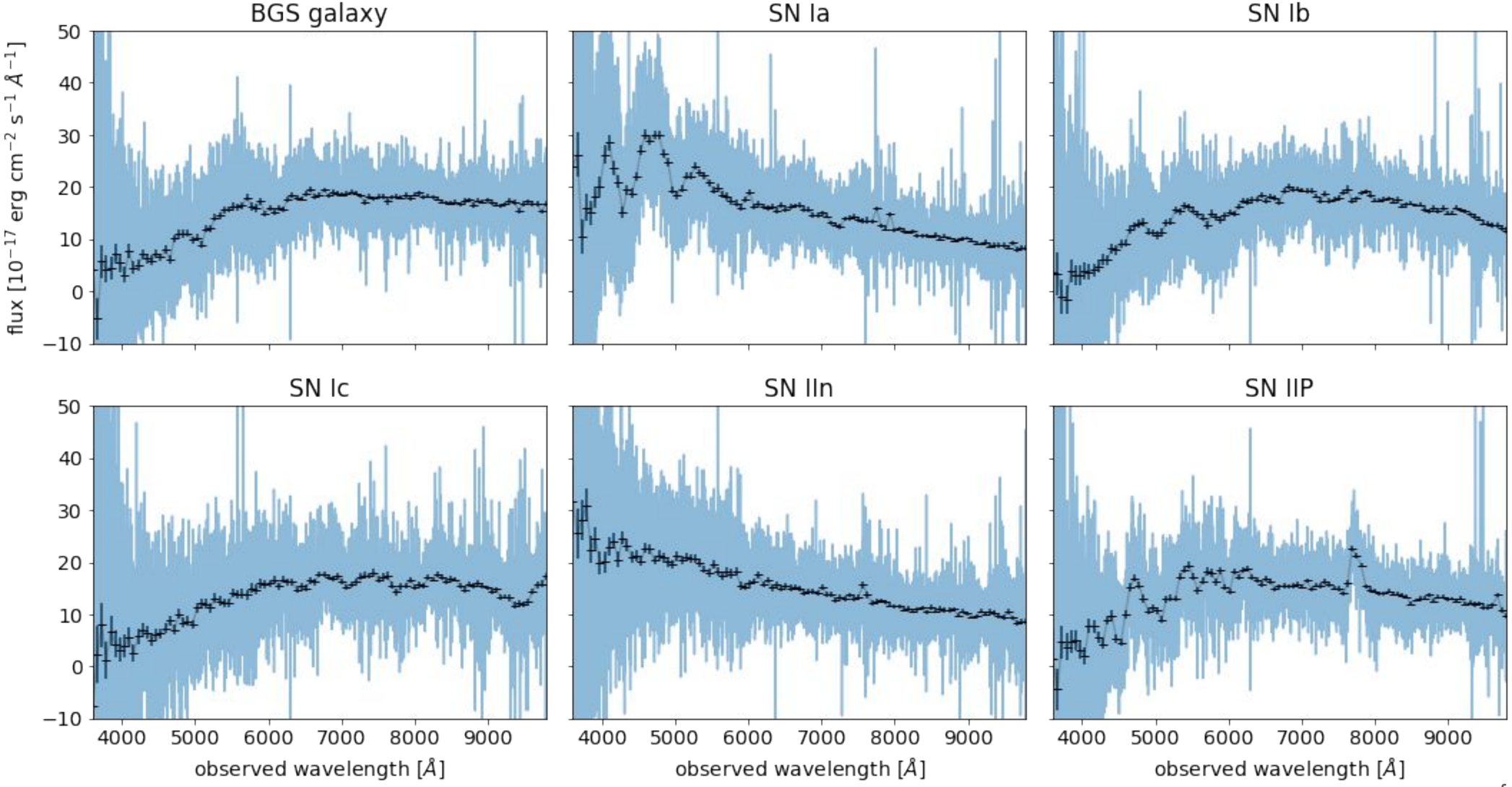


Madgwick et al. 2003

DARK ENERGY SPECTROSCOPIC INSTRUMENT

U.S. Department of Energy Office of Science

Simulated DESI Spectra



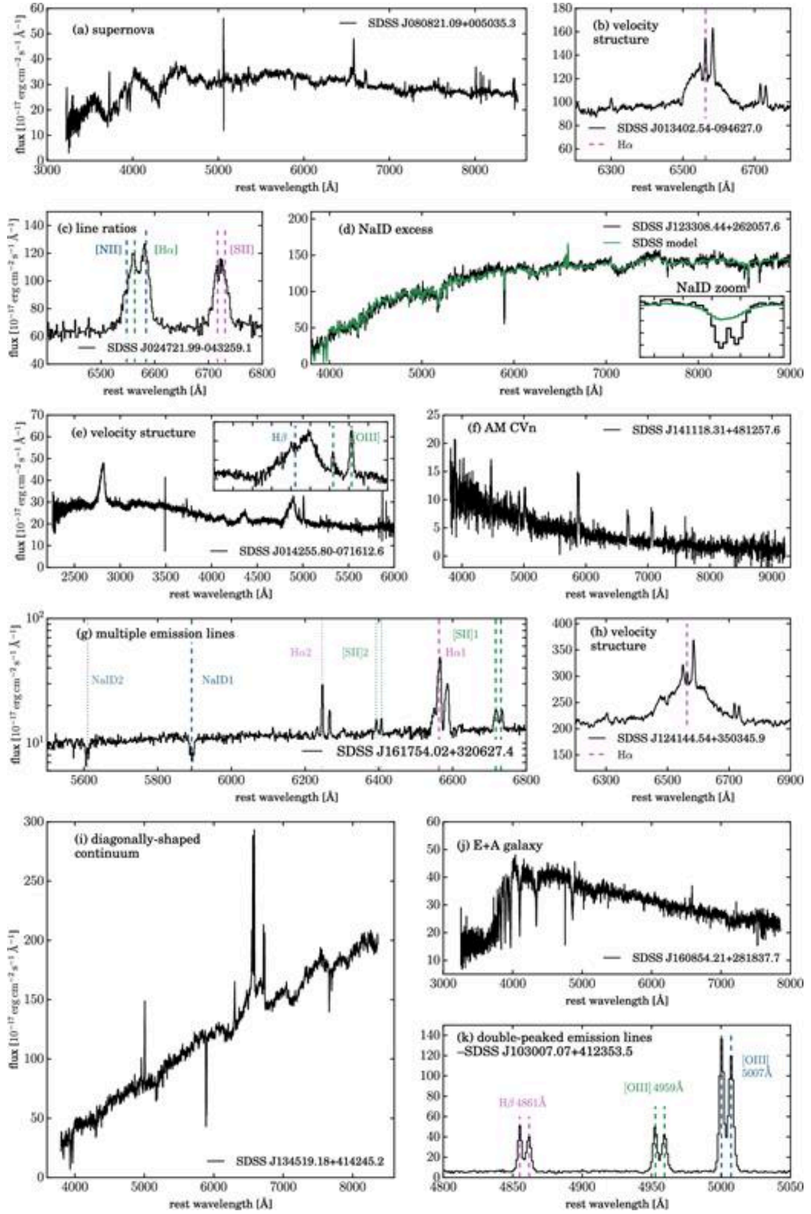
from BenZvi

Data / Features

- Per exposure
  - Reduced Spectrum: Wavelength, Flux, Variance, Mask [array[7781]]
  - Metadata: Observing conditions, instrumental configuration
  - Postprocessing: Estimated redshift, …
- Multiple exposures per object
- Pipeline version

Efforts Past and Present

- Sloan Digital Sky Survey (SDSS)
  - A large dataset in the can https://www.sdss.org/dr16/
  - Unsupervised Random Forest with 2M SDSS galaxies (see right)
- Efforts within DESI
  - Algorithms
    - Template-based searches
    - Supervised Learning based on empirical transient models
      - Application
    - Missing: Unsupervised learning because of a lack of person power
  - Mobilizing follow-up resources for live follow-up



Gallery of SDSS outliers from unsupervised learning Baron & Poznanski (2016)

DESI Discovery Potential from Outlier Detection

- ~2 SNe per 15,000 “Bright” Galaxies
  - Core SNe not efficiently found in imaging transient searches
- 1-8 Tidal Disruption Events per year
- Galaxy-galaxy gravitational lenses
- Galaxies with unusual gas kinematics
- Galaxy pairs
- ???

My “Machine Learning” Background

- I am not a Machine Learning person
- Past and current ML interests
  - Large number (>1000) parameter fits: Hamiltonian Monte Carlo, multi-probe cosmology, differentiable universe
  - Non-analytic likelihoods: Approximate Bayesian Calculation
  - Non-linear dimension reduction: Diffusion maps
  - “Parameterless : Gaussian Processes
- Science interests that touch on ML
  - SN classification (LSST Brokers), SN Ia standardization, multi-probe cosmology analysis