Assessing the quality of massive spectroscopic surveys with unsupervised machine learning

John Suárez-Pérez
Jaime Forero-Romero
DESI Collaboration

University of Los Andes
Bogotá, Colombia

16th May 2022

Universidad de los Andes

Dark Energy Spectroscopic Instrument (DESI).

Kitt Peak National Observatory Photography. Image from the DESI High School.

- More than 70 international institutes.
- Lawrence Berkeley Laboratory (California).
- Kitt Peak National Observatory, Arizona.
- Dark Energy.
- Construction began on 2015.
- First observations on 2019-2.

https://www.desi.lbl.gov/

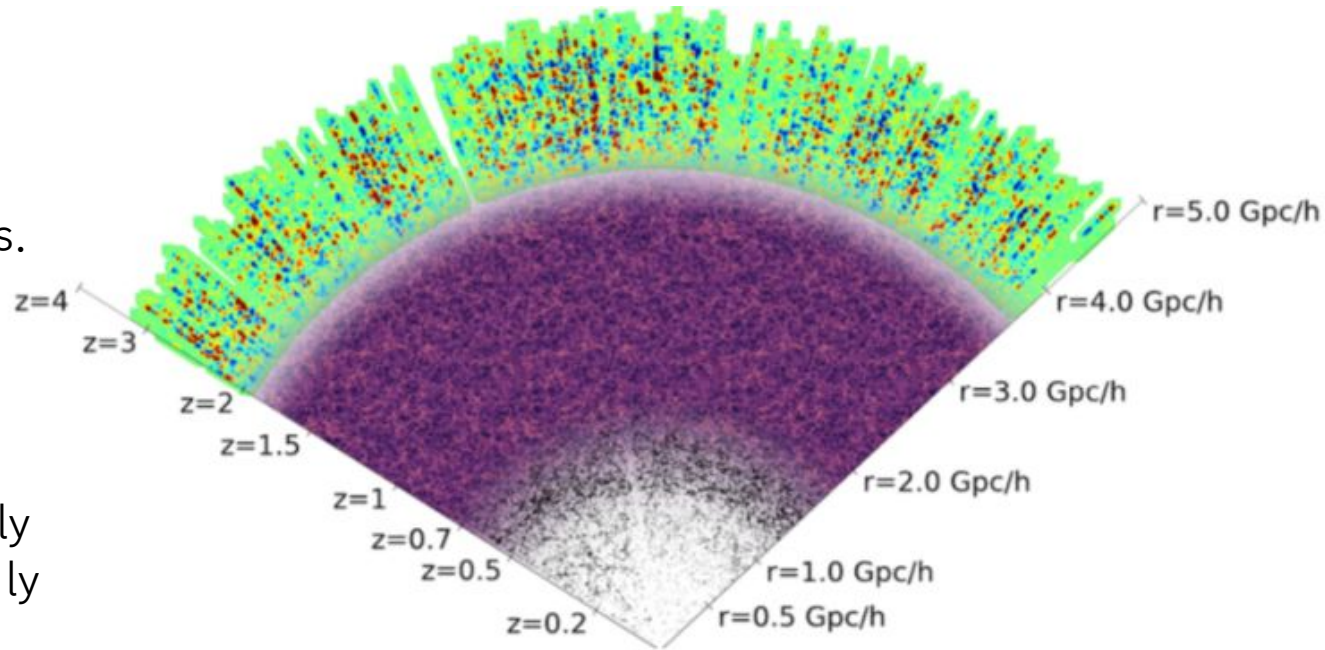Universidad de los Andes

DESI and the Biggest Universe Map.

- 40M spectra.
- 14000 deg².
- Galaxies, quasars, stars.

- Constrain the Dark Energy Models.

- SDSS: $z \sim 0.5$ →   6 Billions ly
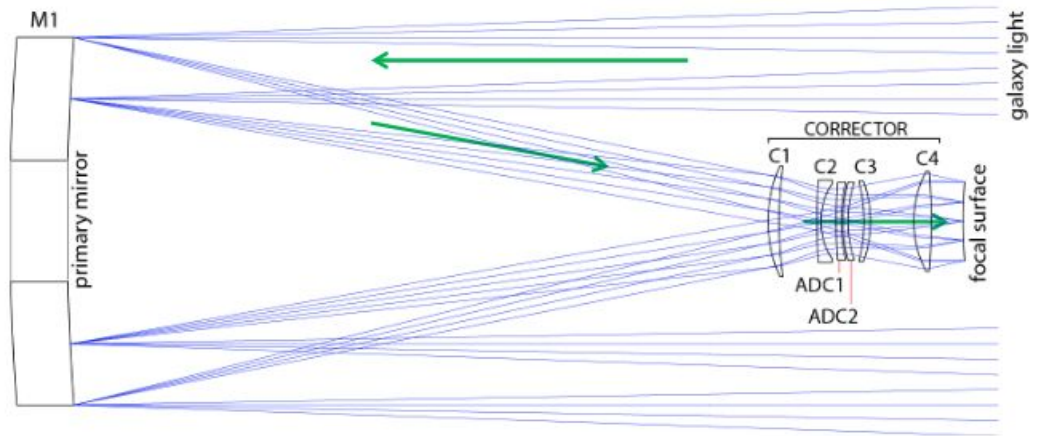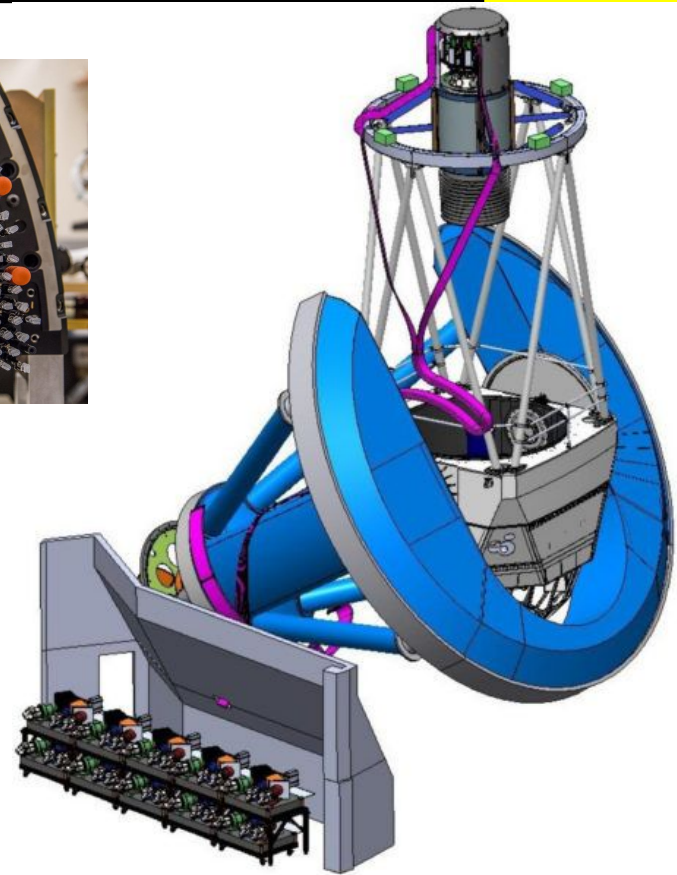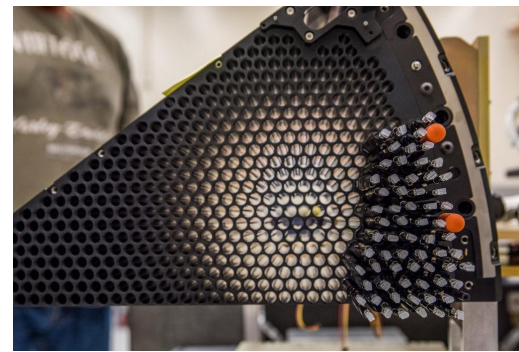- DESI: $z \sim 2$ →   11 Billions ly

Credit to DESI target.

Universidad de los Andes

¿How are the Spectra Measured with the DESI?

- 5000 automatic positioners.
- Fibers with 40m of length.
- 10 Spectrographs.
- 3 bands B,R,Z.
- 3600 Å – 9600 Å.
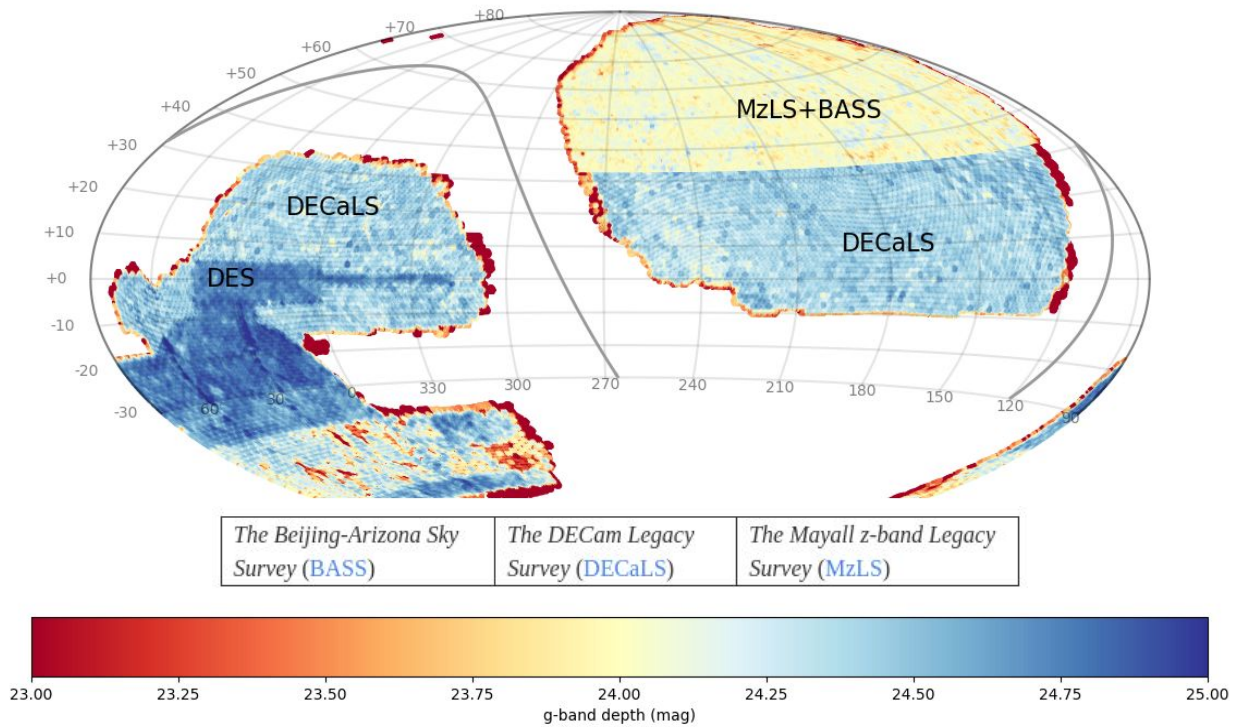- Resolution from 2000 to 5000.
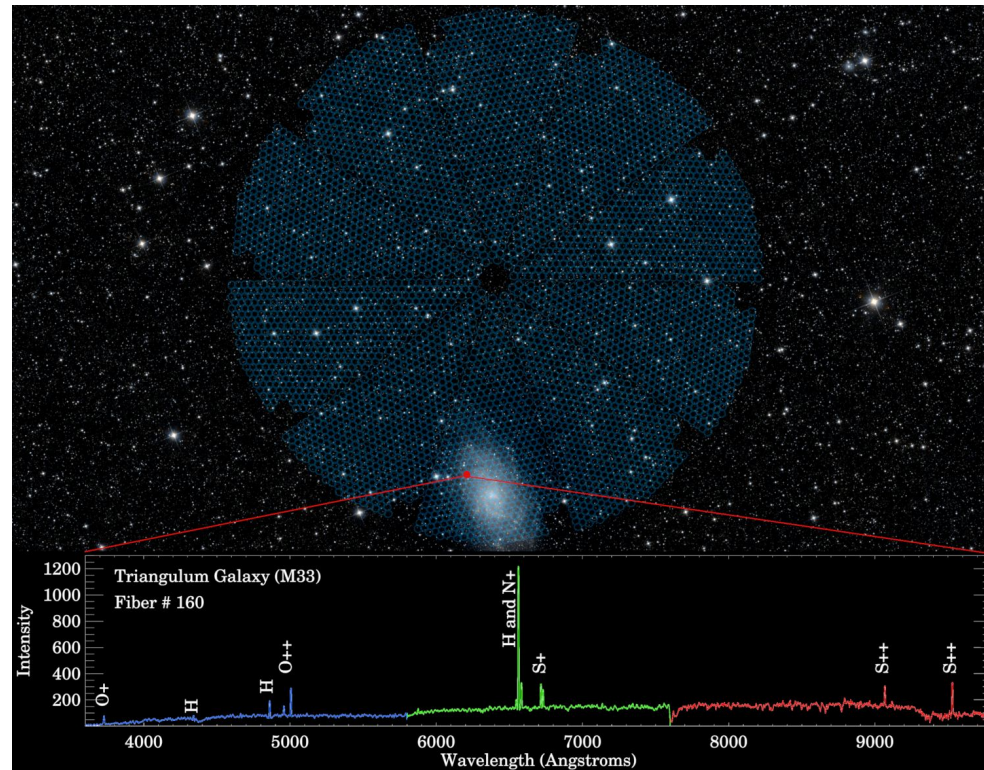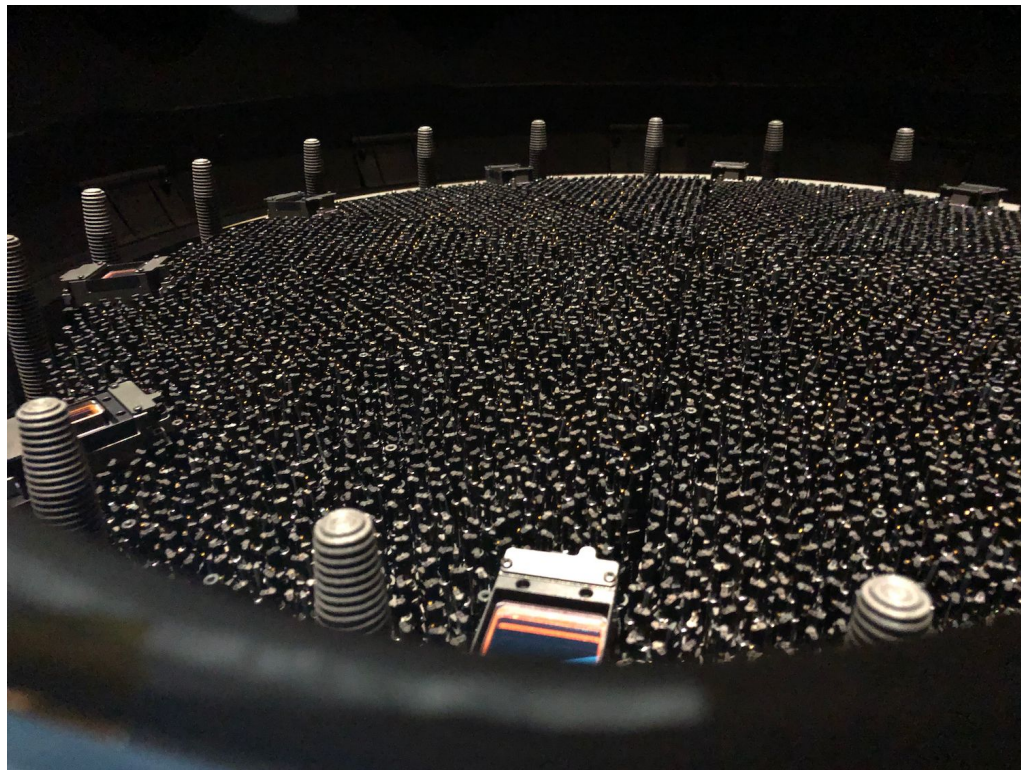- 8.0 deg² Vision Field.

Images from DESI webpage.

Universidad de los Andes

¿Where to Point these Fibers?

Imagen from Legacy Survey.

https://www.legacysurvey.org/

[1] Dey et. al. – Overview of the DESI Legacy Imaging Surveys. 2019

Universidad de los Andes

The Spectra are Measured like this.

Triangulum Galaxy (M33)
Fiber # 160

Images from DESI webpage.

Universidad de los Andes

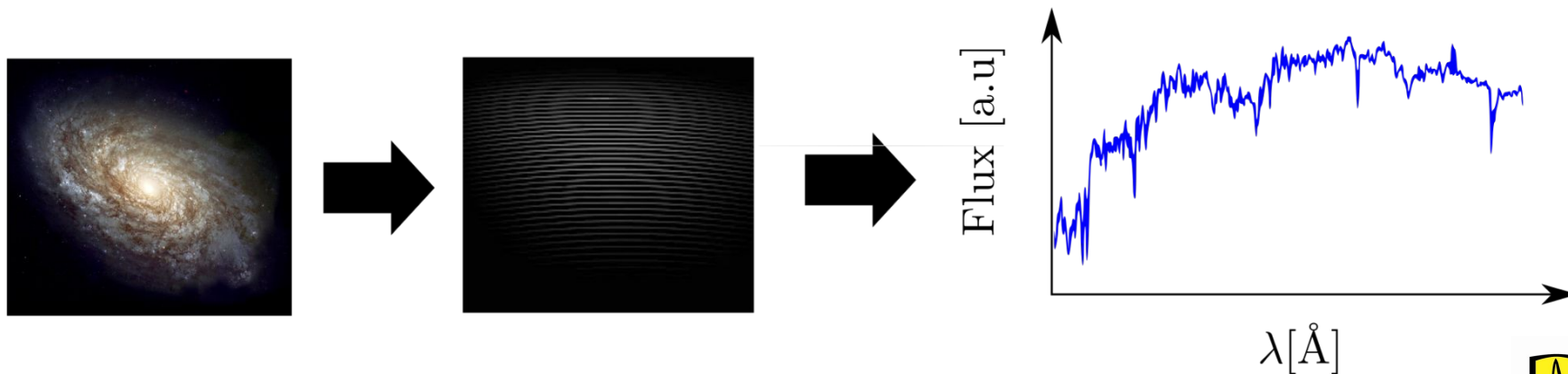Credits to Claire Lamman   @clairelamman.

Acquisition Data Process.

1. Observation (instrumental)

2. Corrections (reduction)
   Calibration (reduction)   Reduction Pipeline

3. Final Calibrated Spectrum

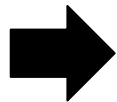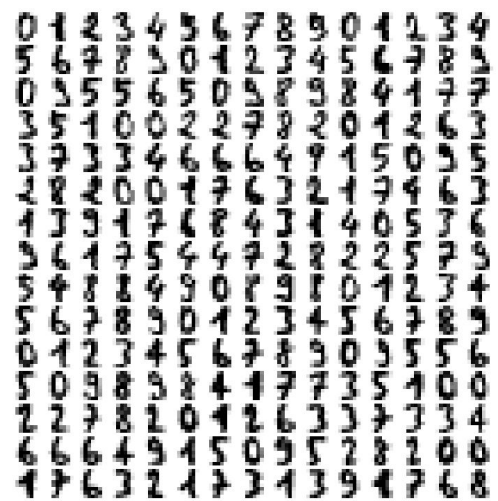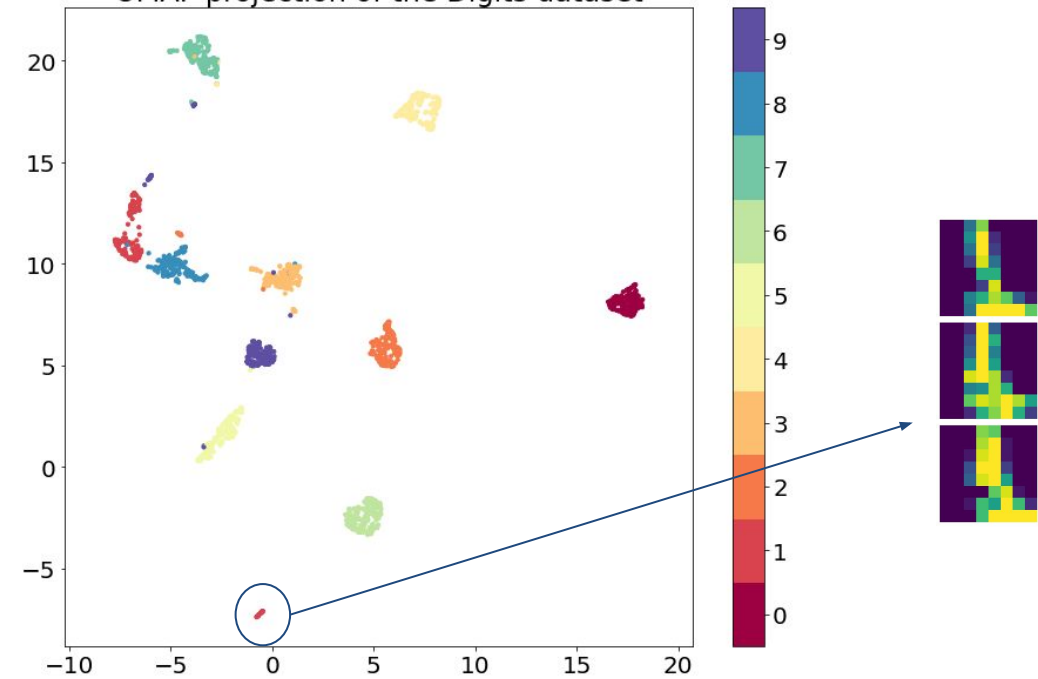   **Problem:** To find reduction or instrumental errors. (Outliers)

The UMAP Algorithm.

- Uniform Manifold Approximation and Projection (UMAP) algorithm.
  - Unsupervised ML algorithm.
  - No linear dimensionality reduction for visualization.
- Outliers Detection.

64-Dimension (8x8)

UMAP projection of the Digits dataset

[2] McInnes, L. et.al. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.(2018).

Universidad de los Andes

Our Outliers Detection based Method.

Outliers identification on massive observational into reduced dimensional spaces.

1) UMAP
   - Number of neighbors (Nn): [0 - 50] steps 5
   - Minimal Distance (Md): [0.0 - 0.5] steps 0.1
   - Metrics (Me): ['euclidean', 'cosine', 'braycurtis']
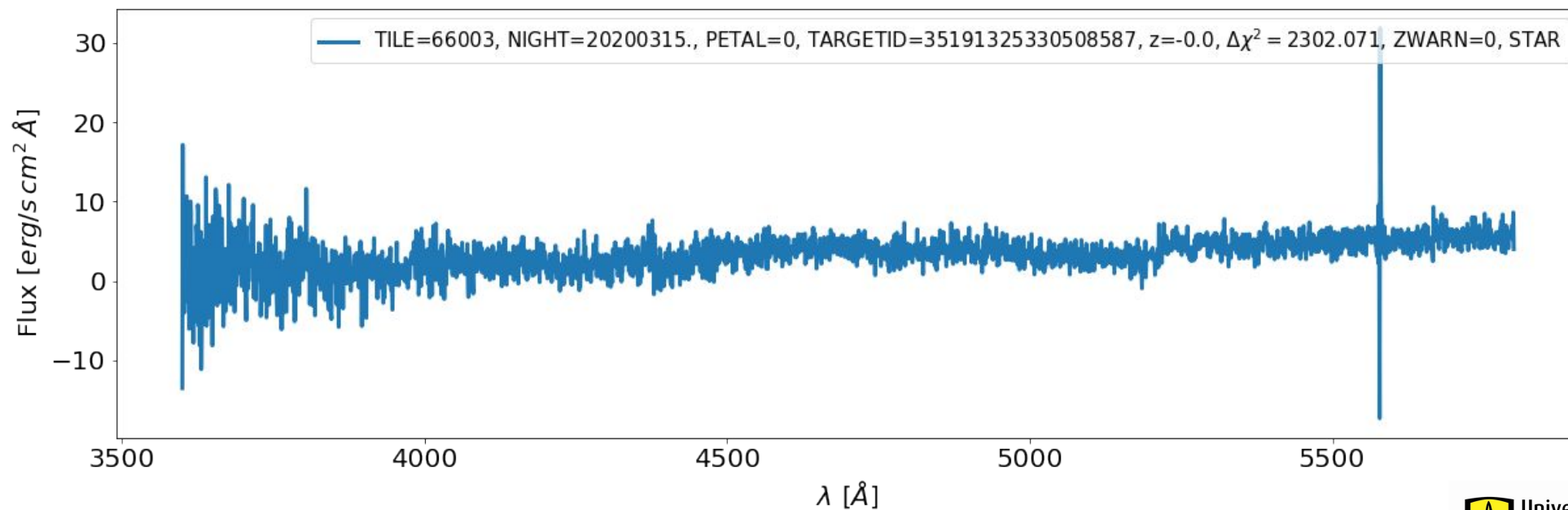   - Dimension (D): 2

2) FoF (Friend of Friends)
   - Linking Length (Ll): [0.1 - 0.5] steps 0.1

Suárez-Pérez, et.al. In Prep.
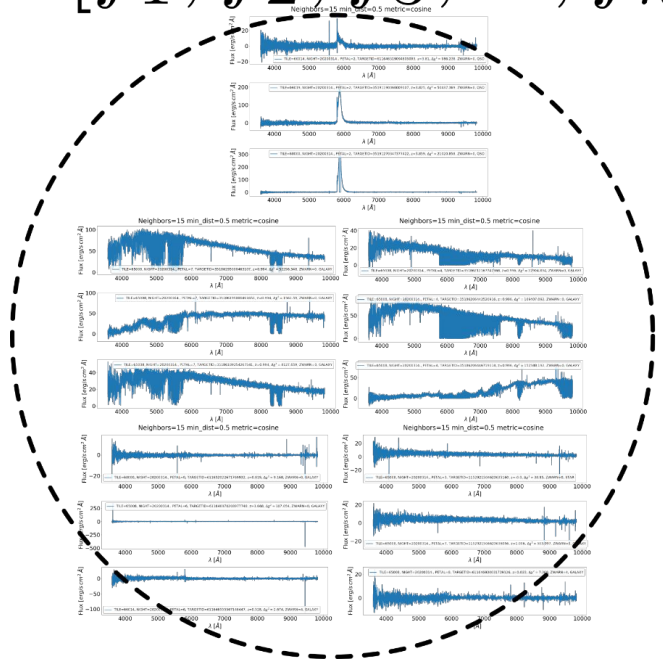
Universidad de los Andes

Spectra.

- Andes DR was the first data reduction pipeline.
- Observations on March 2020.
- 115K spectra, each spectrum with 2K points in each band (B,R,Z).
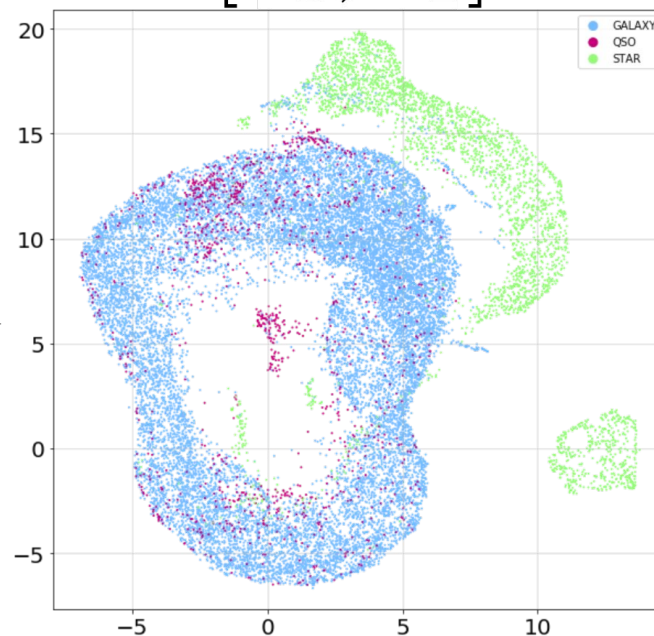- 3600 Å to 9600 Å.

Suárez-Pérez, et.al. In Prep.

Universidad de los Andes

Dimensional Reduction with UMAP.

$$[f_1, f_2, f_3, ..., f_n]$$

$$[x_1, x_2]$$

Suárez-Pérez, et.al. In Prep.

Detection of Instrumental Errors.

UMAP - Band:B Type:Obs Ll:0.14
Mems:8 Island:10

BAND=B TILE=65008, NIGHT=20200314., PETAL=1, TARGETID=611646930006574018, z=0.589, $\Delta\chi^2$ = 3.546, ZWARN=4, GALAXY

BAND=B TILE=66014, NIGHT=20200314., PETAL=1, TARGETID=35185766803968039, z=0.282, $\Delta\chi^2$ = 61.395, ZWARN=0, GALAXY

BAND=B TILE=67230, NIGHT=20200314., PETAL=1, TARGETID=35185766120297470, z=0.097, $\Delta\chi^2$ = 8.283, ZWARN=4, GALAXY

Me = 'cosine'
Nn = 15
Md = 0.5
D = 2
Ll = 0.14

DesiSpec Issue Reported on Github as **defect in the CCD**, #983
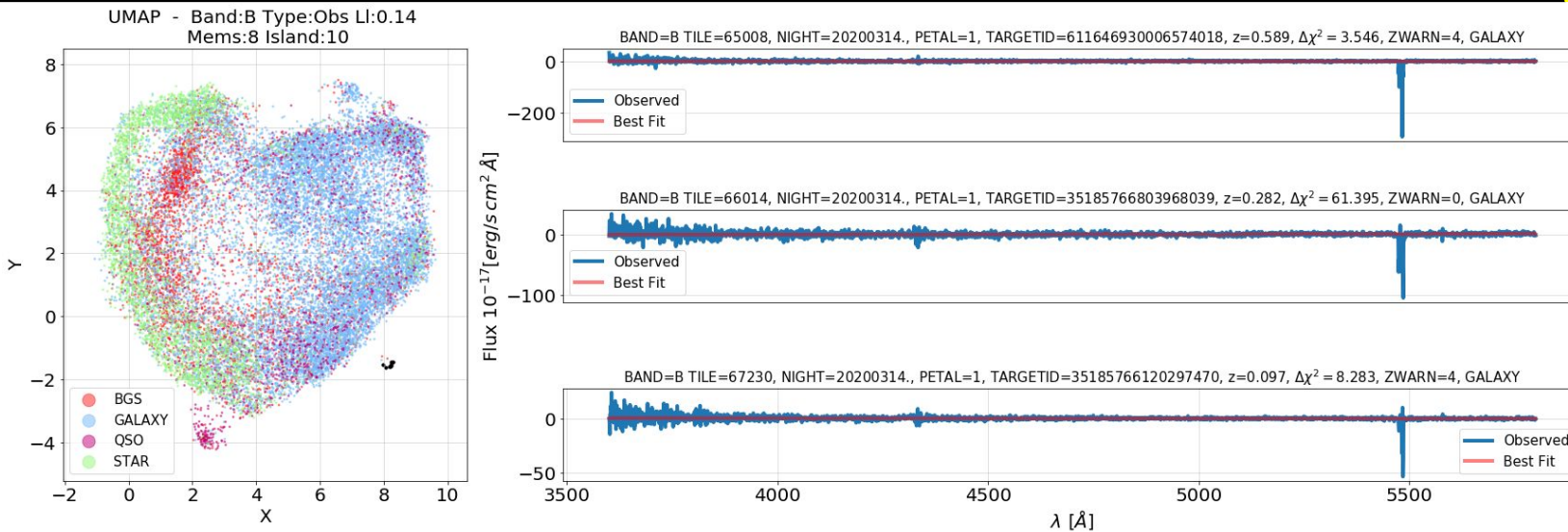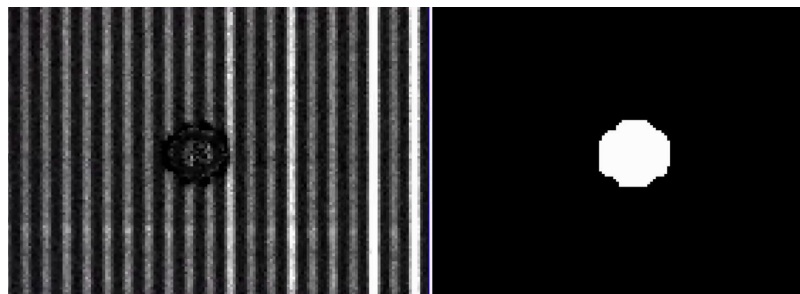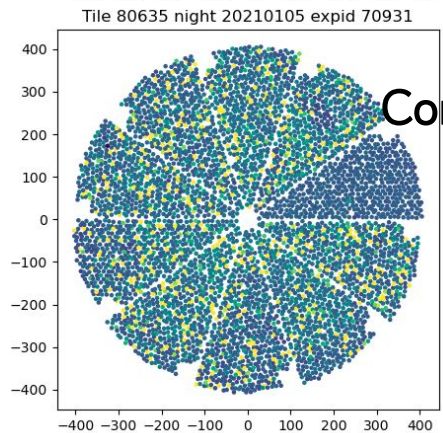
Suárez-Pérez, et.al. In Prep.

Universidad de los Andes

Detection of Instrumental Errors.

Me = 'cosine'
Nn = 15
Md = 0.5
D = 2
Ll = 0.2
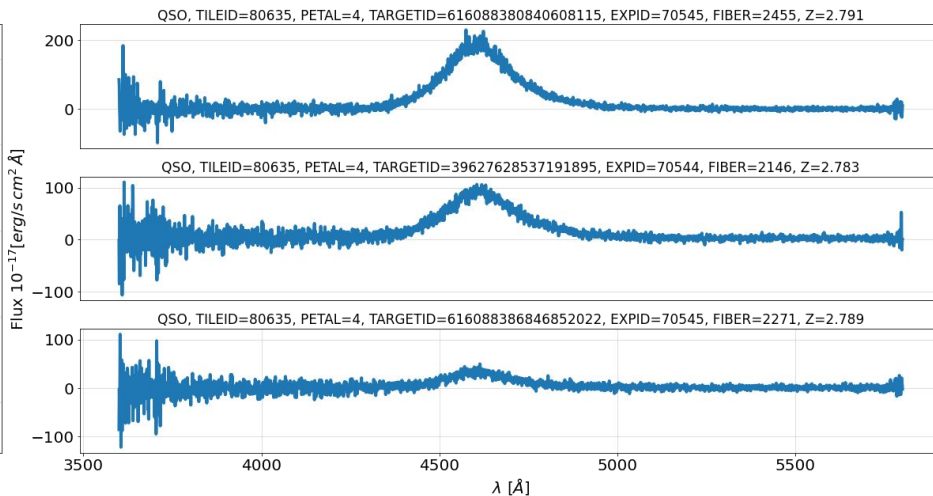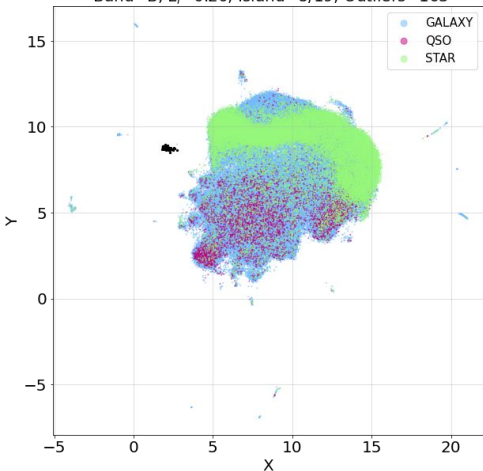
Contaminated

Not
Contaminated
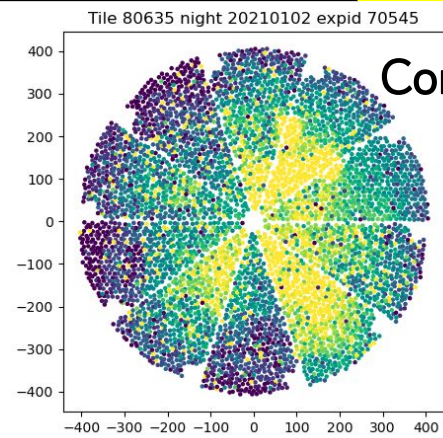
DesiSpec Issue Reported on Github as **contamination**, #1262

Suárez-Pérez, et.al. In Prep.

Universidad de los Andes

Conclusions.

❖ The fast increase in complexity and amount of observations requires the use of highly efficient tools. The Machine Learning algorithms are an option that have demonstrated good results.

Tile 80635 night 20210102 expid 70545

➤ The dimensionality reduction is an efficient technique to visually detect *outliers.* This detection is fundamental as an early alerts system to identify instrumental errors.
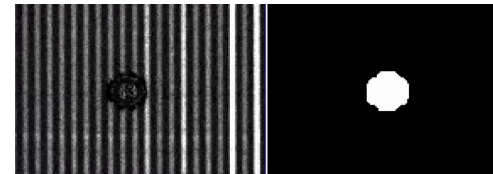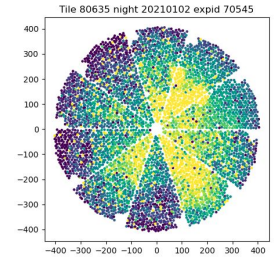
Thank You!

✉ jf.suarez@uniandes.edu.co

https://jsuarez314.gitlab.io