ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Task 4.3 Summary and Status

Martino Romaniello, Nima Sedaghat, Felix Stoehr
Jon Carrick, FX Pineau, Henri Boffin

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

WP4 Tasks

Task 4.1 Integration of astronomy VO data and services into the EOSC

Lead: Marco Molinaro (INAF)

Task 4.2 Implementation of FAIR principles for ESFRI data through the Virtual Observatory

Lead: François Bonnarel (CDS CNRS Université de Strasbourg)

Task 4.3 Adding value to trusted content in astronomy archives

Co-leads: Mark Allen (CNRS-ObAS) & Martino Romaniello (ESO)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Task 4.3 Adding value to trusted content in astronomy archives

*Next generation functionalities for creation and publication of **high-level, value-added data products** from ESFRIs*

- Assessment and application of **new techniques** – machine learning analytics. (connection to WP3)
  - *Specific example applied to ESO archives data products: spectra, cubes, source catalogues*
- Identification of **stewardship best practices**
  - *Curation and publication of next-generation data products via ESFRI archives*
  - *Technical and human aspects*

Funded by the European Union's Horizon 2020 - Grant N° 824064

Major milestones

D4.5 Release of prototype machine learning enabled archive services providing value-added content to archives (Demonstration)	Month 30 (+6) January 2022
D4.8 Final analysis report on use of IVOA standards for FAIR ESFRI and community data and best stewardship practices for value-added data (Report) *(includes report of feedback on prototype services developed for D4.5)*	Month 40 (+6?) November 2022 (?)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Deep Learning & Data Archives

- Scope: provide archive users with novel ways to identify data
  - Beyond traditional approach of specifying query parameters
  - Revolutionary extension of recent move from instrument to data keywords (e.g. exposure time to signa-to-noise)
- Target: the different data types in the ESO Science Archives
  - Spectra
  - Cubes
  - Source catalogues
- Let the data speak … on a massive scale
  - Huge data variety: impractical, and not desirable, to impose preconceived interpretations

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Deep Learning on 1D spectra: HARPS

- Scope: Deep Learning analysis of the entire HARPS archive
  - High-resolution, high-stability spectrograph
  - Main science case: discovery and characterization of exoplanets
  - ~270,000 spectra, ~300,000 wavelengths channels each
  - 1D spectra, pipeline-reduced to high accuracy
- Different approaches
  - HITS: fully-connected autoencoder with 2 latent dimensions, down sampled spectral resolution
    - Agile architecture for speed and interactivity
  - ESO: combination of convolutional and fully-connected layers with between 4 to 8192 latent dimensions, full spectral resolution
    - Find minimal representation which preserves all the relevant information

Funded by the European Union's Horizon 2020 - Grant N° 824064

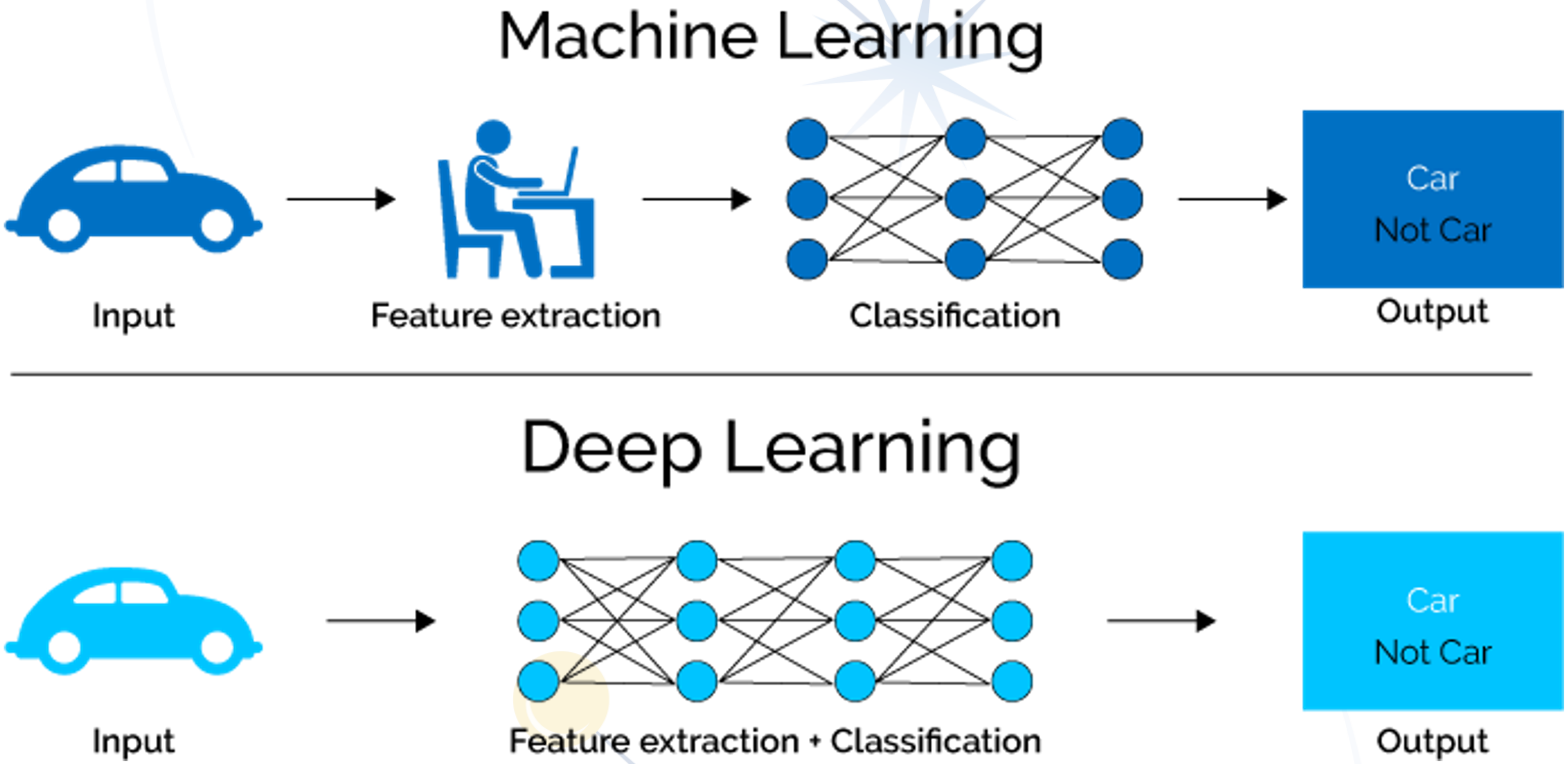
So now, how to turn this into user services?

- These techniques have the potential of being a game-changer, BUT:
  - The results have to be meaningful
  - Interaction with the results has to be user-friendly

- In practice:
  - What makes spectra similar?
    - What are the uses and limitations of a similarity service?
  - Can spectra be tagged with some physical properties of the celestial object?
    - If so, what are they?
    - What are the uses and limitations?
  - What is the best way to present and interact with the results?
  - Services for the few or for the many?
    - About half of the astronomical community uses the ESO Science Archives

Funded by the European Union's Horizon 2020 - Grant N° 824064

From Hand-Crafted Features to Representation Learning



Funded by the European Union's
Horizon 2020 - Grant N° 824064

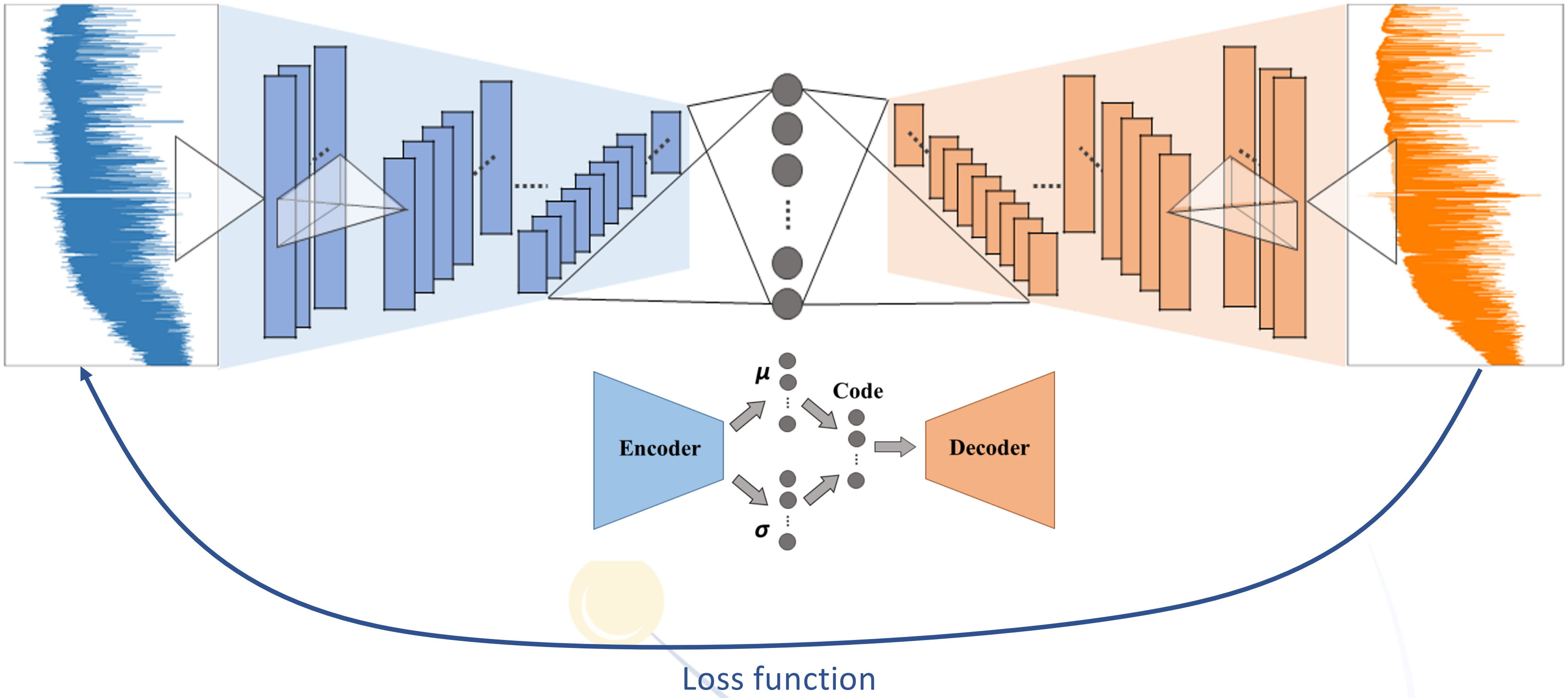
Let the Data Speak for itself!

- Define the problem and data carefully
- Design your own network (with inspiration from existing ones)
- Train it
- Try to interpret how the network solved the problem

- Let’s see how a network chooses to perceive big data
- Push for disentangled representations
  - Facilitates interpretation

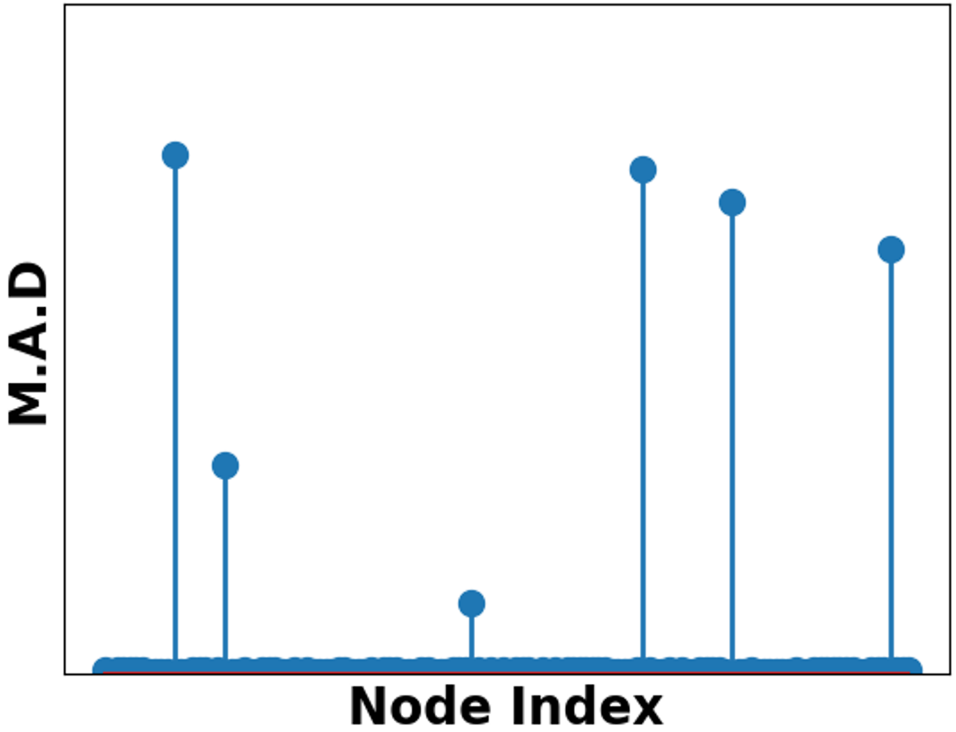
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Network schematics



Funded by the European Union's Horizon 2020 - Grant N° 824064

Not all dimensions are informative

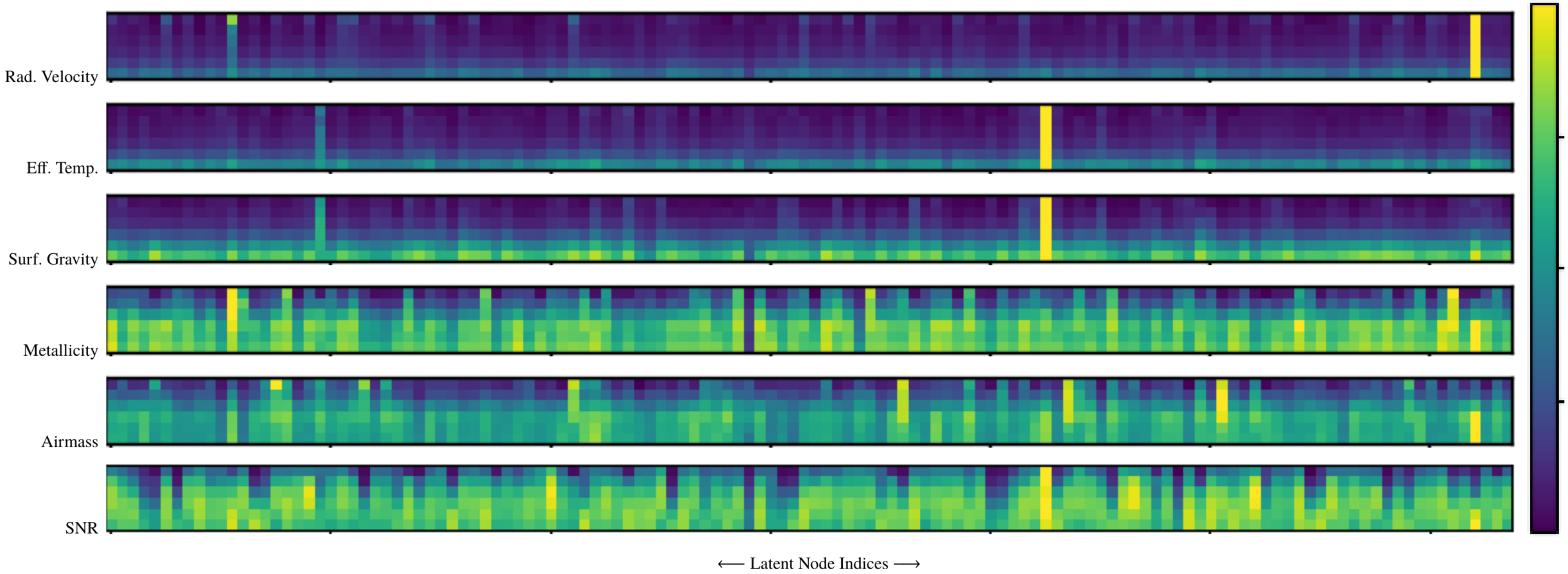


Only 6 dimensions out of 128 carry information

So now, are they physically interpretable?

Funded by the European Union's Horizon 2020 - Grant N° 824064

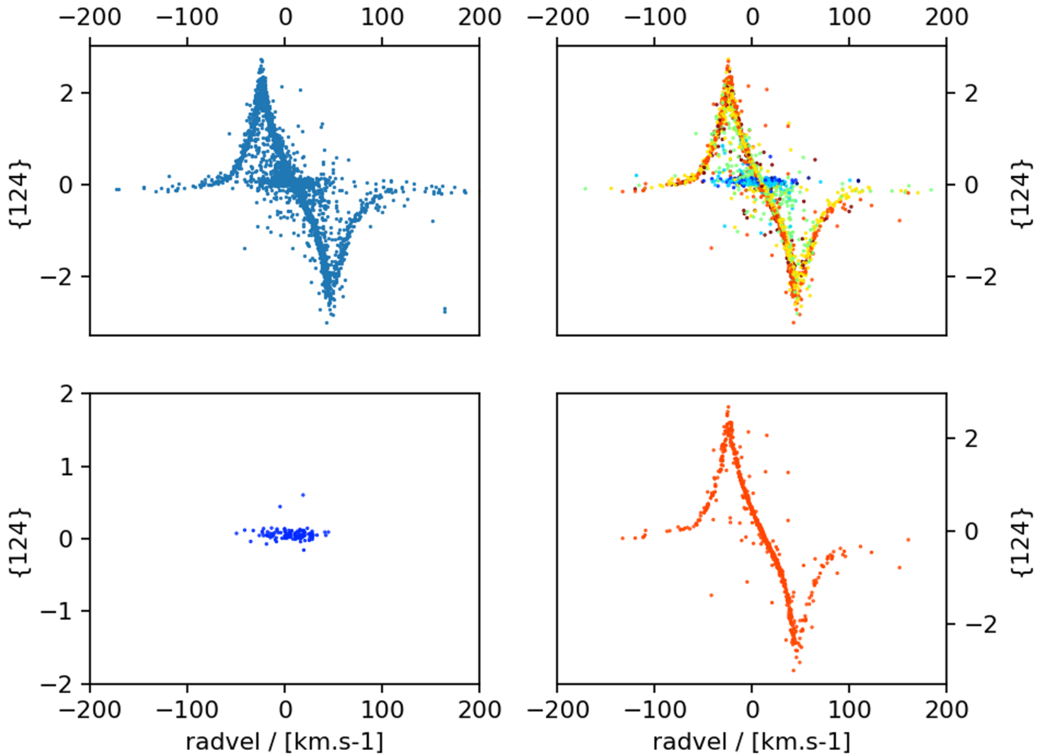
Mutual Information with Stellar Parameters



Sedaghat, MR, Carrick, Pineau 2021, MNRAS, 501, 6026

Funded by the European Union's
Horizon 2020 - Grant N° 824064

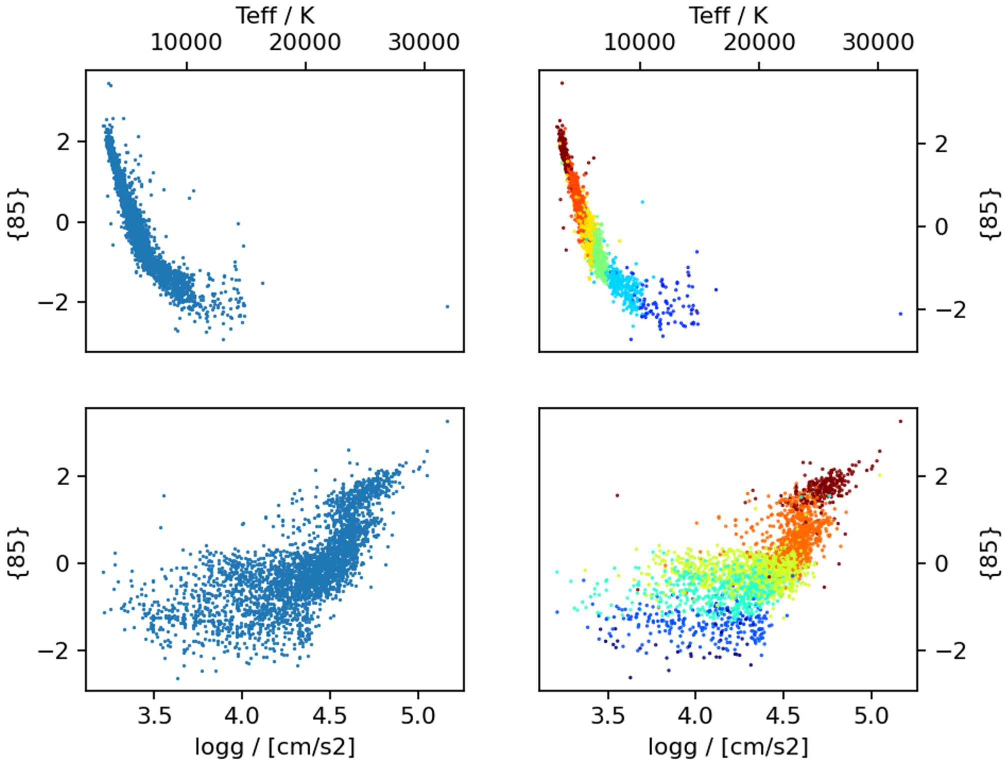
Mutual Information with Stellar Parameters



Sedaghat, MR, Carrick, Pineau 2021, MNRAS, 501, 6026

Funded by the European Union's
Horizon 2020 - Grant N° 824064

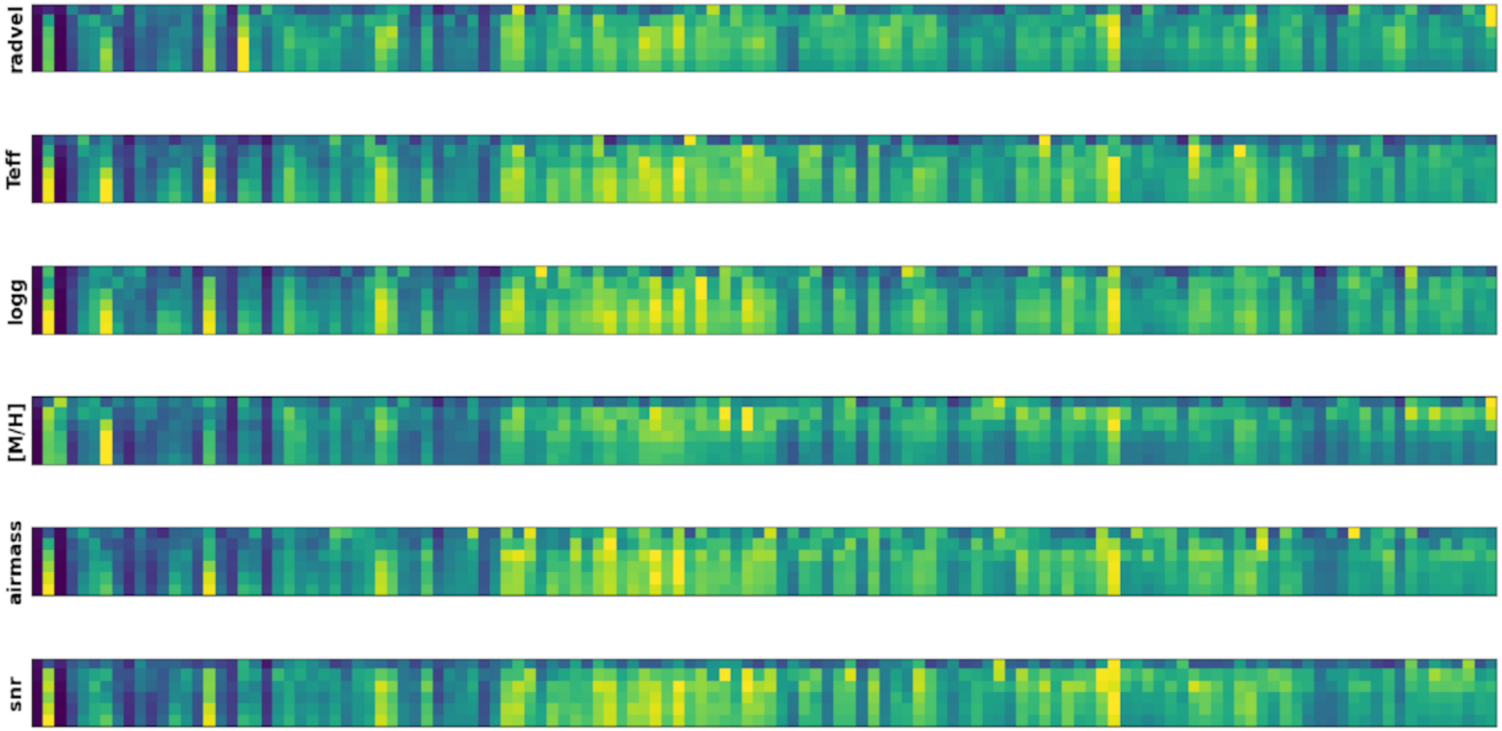
Mutual Information with Stellar Parameters



Sedaghat, MR, Carrick, Pineau 2021, MNRAS, 501, 6026

Funded by the European Union's
Horizon 2020 - Grant N° 824064

PCA: no one-to-one mapping with physical features

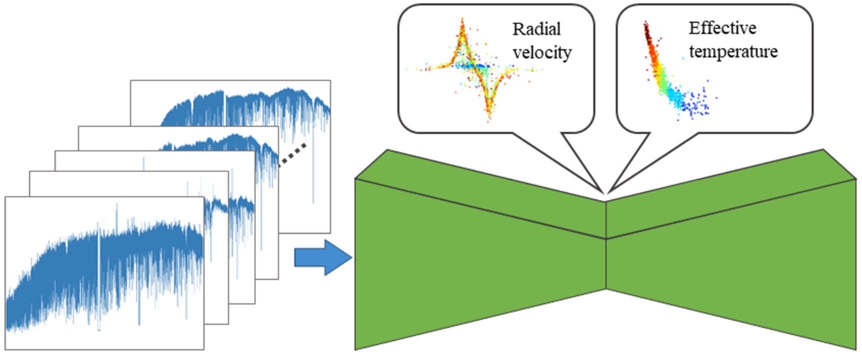


Sedaghat, MR, Carrick, Pineau 2021, MNRAS, 501, 6026

Funded by the European Union's Horizon 2020 - Grant N° 824064

The first step is taken: the basic idea works

Sedaghat, MR, Carrick, Pineau 2021, MNRAS, 501, 6026

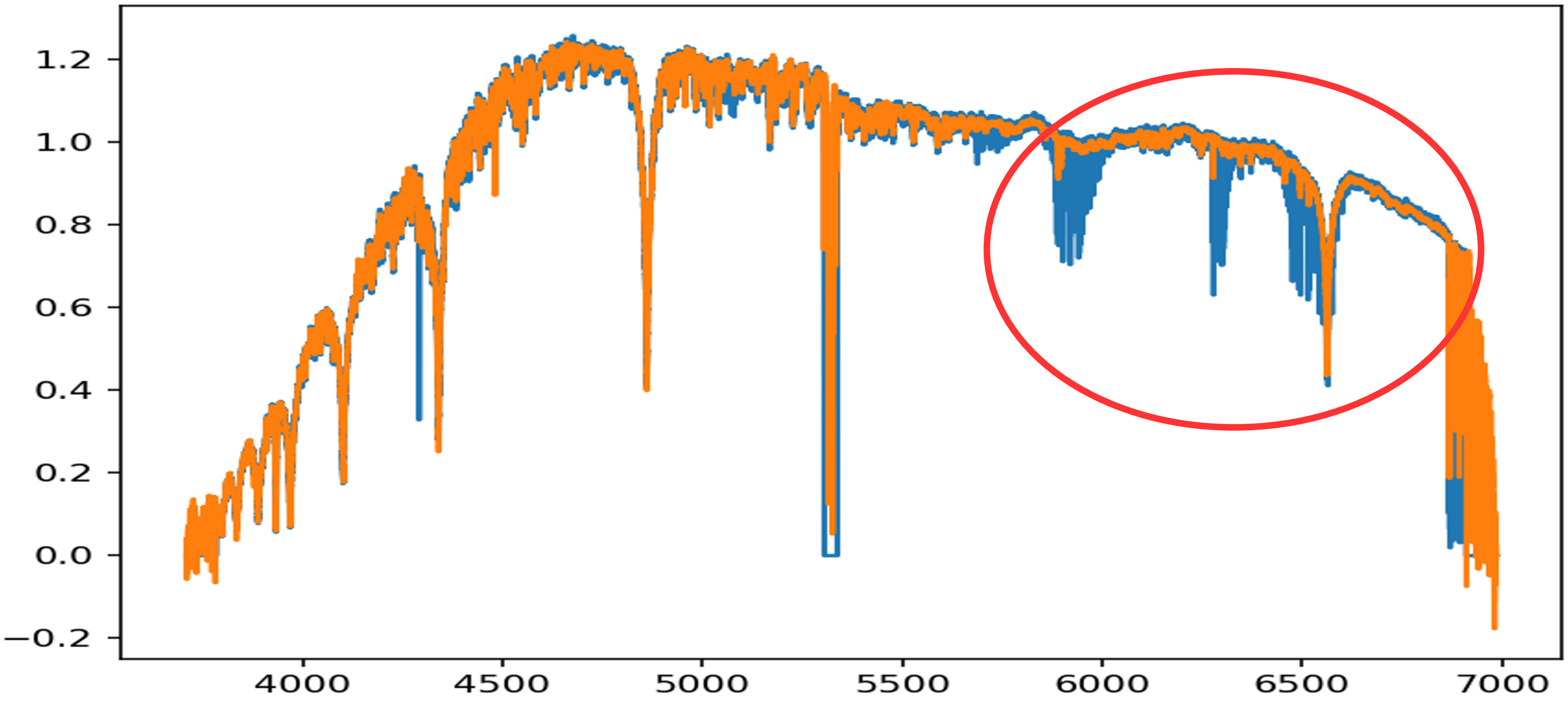


Next steps

- What physical parameters are the other significant dimensions linked to?
  - Why isn't RV representation monotonic?
- Extension to different spectral datasets
  - Stellar: RAVE, GALA (non-ESO)
  - Non-stellar: UVES, X-Shooter
- Extension to images/cubes: ALMA
- Source separation
  - Stellar vs telluric features

Funded by the European Union's Horizon 2020 - Grant N° 824064

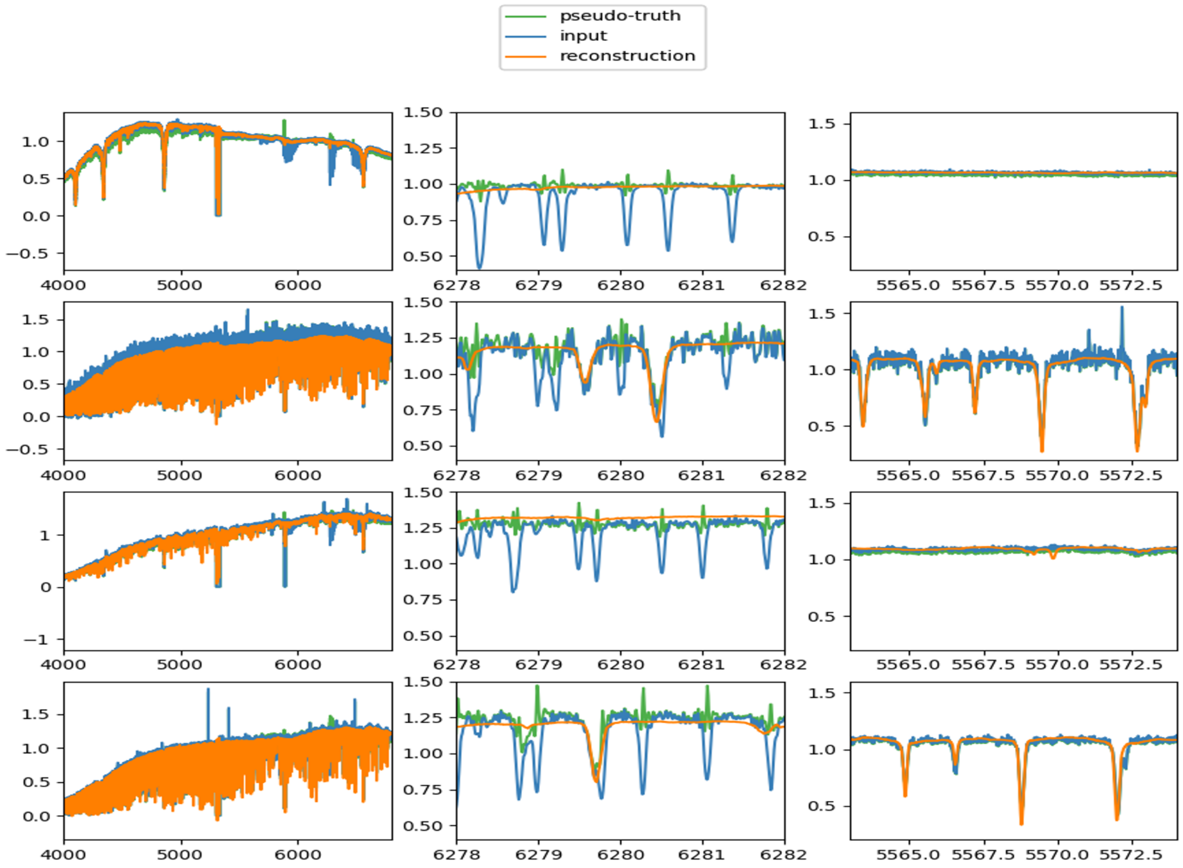
Source separation: stellar vs telluric features



Sedaghat, MR et al, in preparation

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Source separation: stellar vs telluric features



Sedaghat, MR et al, in preparation

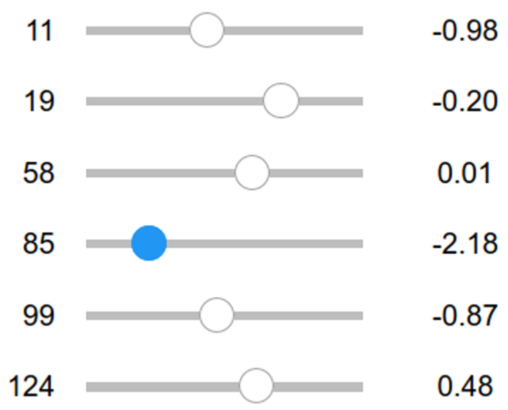
Funded by the European Union's Horizon 2020 - Grant N° 824064

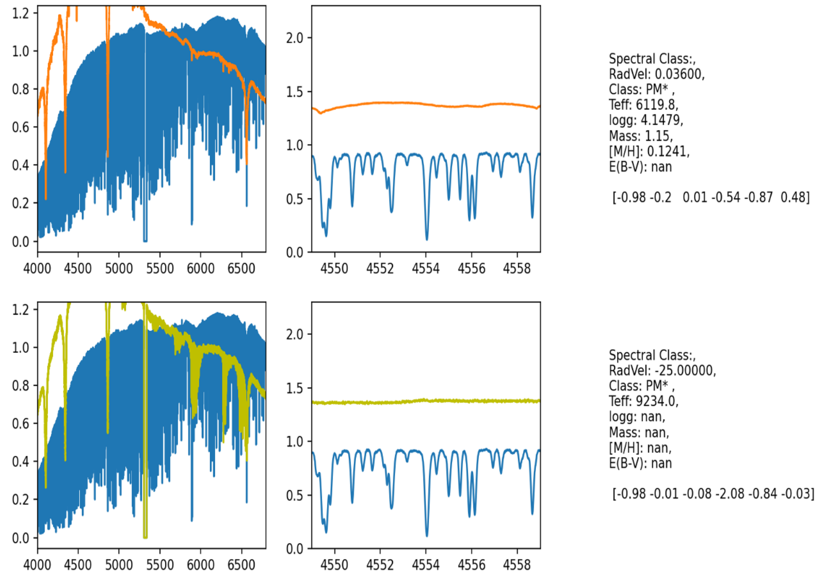
WIP: ALMA cubes and images

- ALMA generates 3D data cubes: 2 spatial + 1 spectral dimensions
- The cubes themselves unfortunately are not homogeneous enough long the spectral dimension to be readily fed to the Deep Learning machinery
- We started from 2D images with the spectral dimension compressed ("continuum images"), running them through the same machinery as the spectra
- WIP …

Funded by the European Union's Horizon 2020 - Grant N° 824064

User tools: “Sliders” for exploration of the latent space

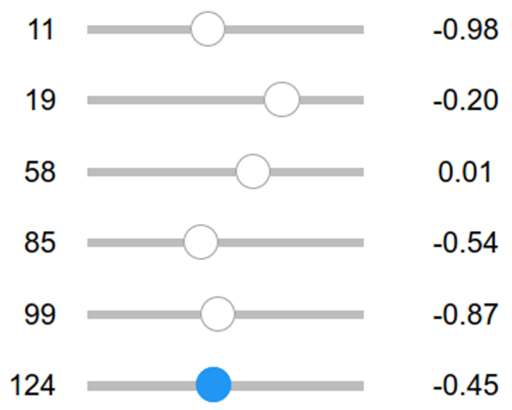


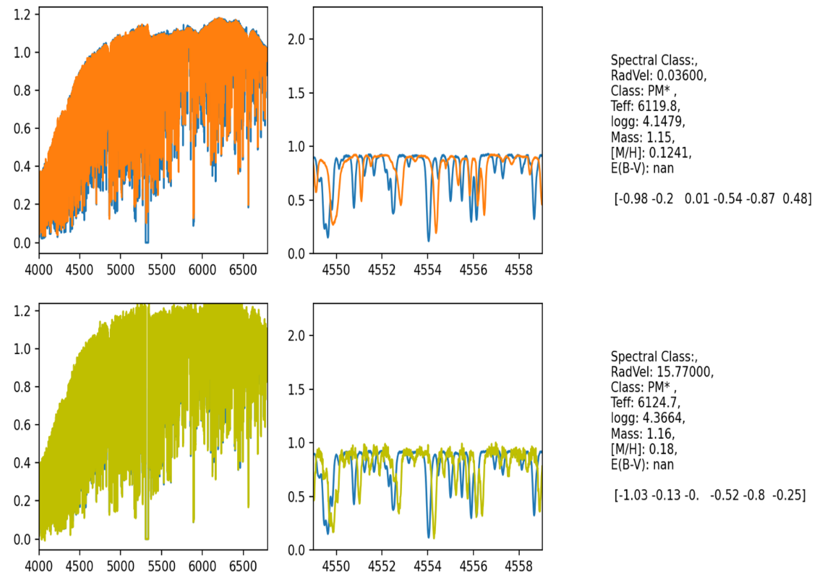


https://www.eso.org/~nsedagha/universe/#astromachines

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

User tools: “Sliders” for exploration of the latent space

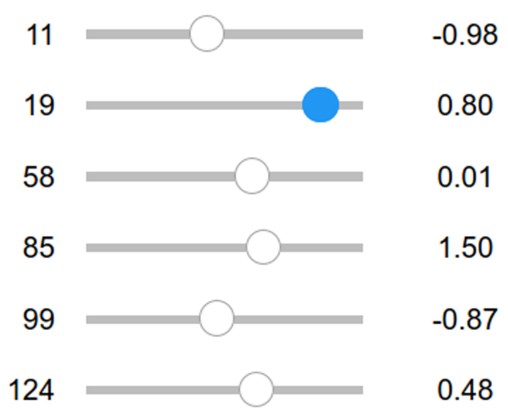


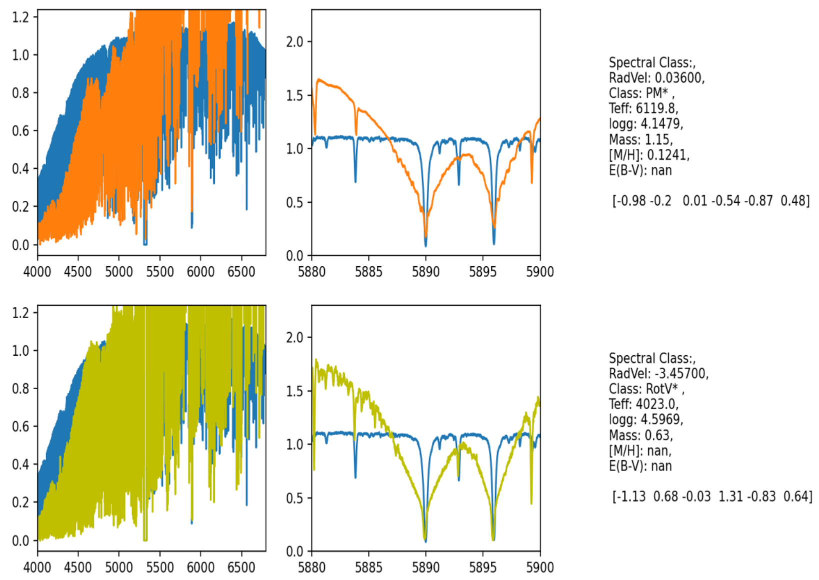


https://www.eso.org/~nsedagha/universe/#astromachines

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

User tools: “Sliders” for exploration of the latent space

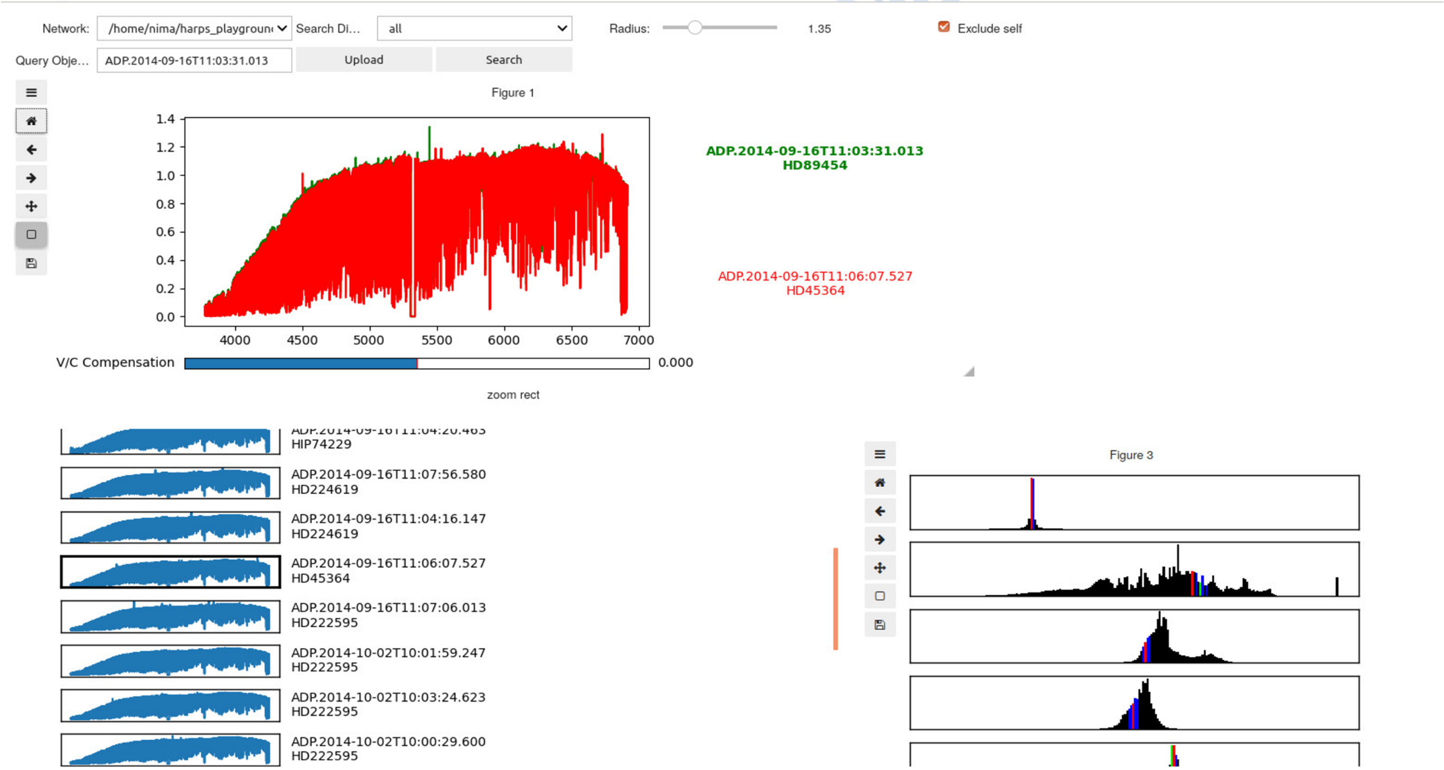




https://www.eso.org/~nsedagha/universe/#astromachines

Funded by the European Union’s
Horizon 2020 - Grant N° 824064

User tools: RETR-SPECT, a Retrieval engine for Spectra



https://www.eso.org/~nsedagha/universe/#retrspect

Funded by the European Union's
Horizon 2020 - Grant N° 824064

http://www.eso.org/~nsedagha/universe

martino.romaniello@eso.org
nima.sedaghat@eso.org

Funded by the European Union's Horizon 2020 - Grant N° 824064