Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Machine Learning techniques for cluster cosmology studies with Euclid data

Matteo Costanzi UniTS / INAF

Spoke 3 Technical Workshop, Trieste October 9 / 11, 2023

Scientific Rationale

- **EMULATORS:**

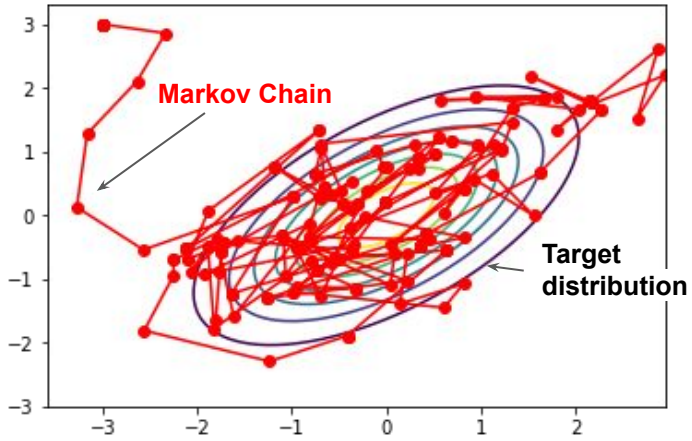
  Cluster cosmology studies – as many other cosmological probes – require the computation of tens of multidimensional integrals, each of which might requires a few seconds, to sample the likelihood of our data given a model (1 MCMC chain ~10^6 samples → ~weeks)

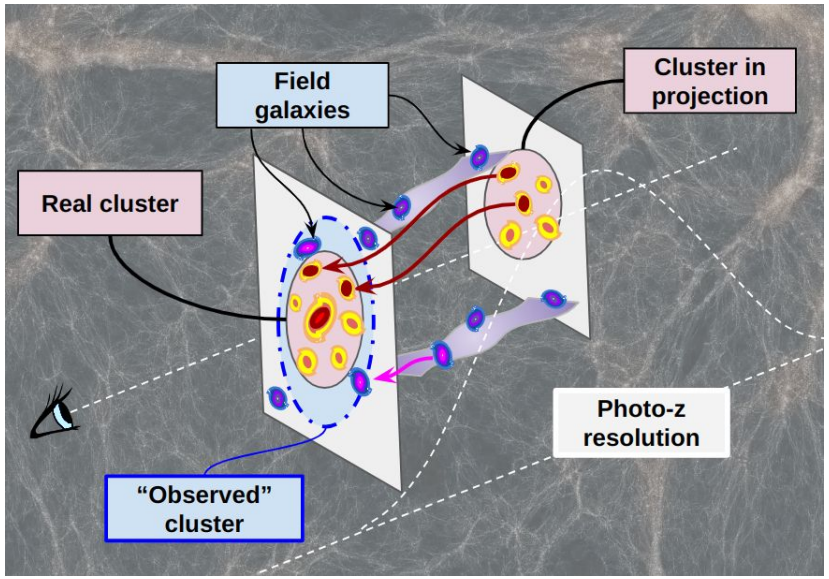
  Machine learning techniques can help in this respect thanks to their ability of emulting complex multidimensional functions with high accuracy and precision.

- **FEATURES EXTRACTIONS:**

  Optical selected clusters are affected by projection effects which bias their observed properties (e.g. richness).

  Machine learning techniques can to mitigate these effects thanks to their ability of extracting features from complex datasets





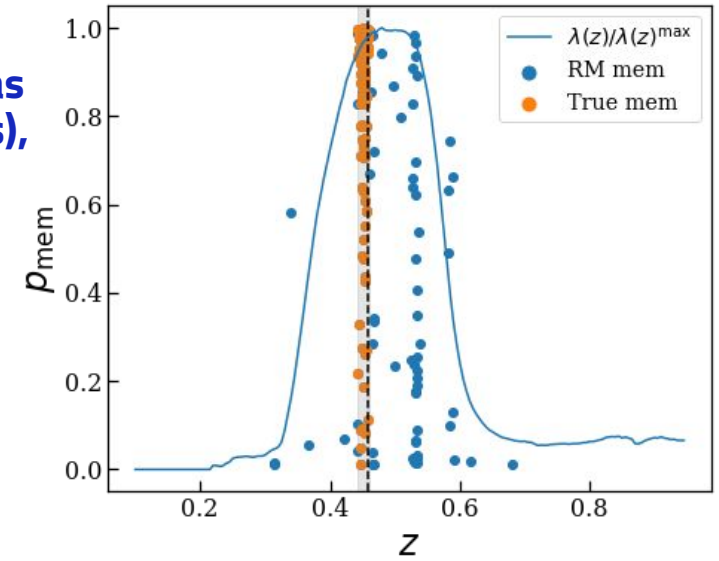
Technical Objectives, Methodologies and Solutions

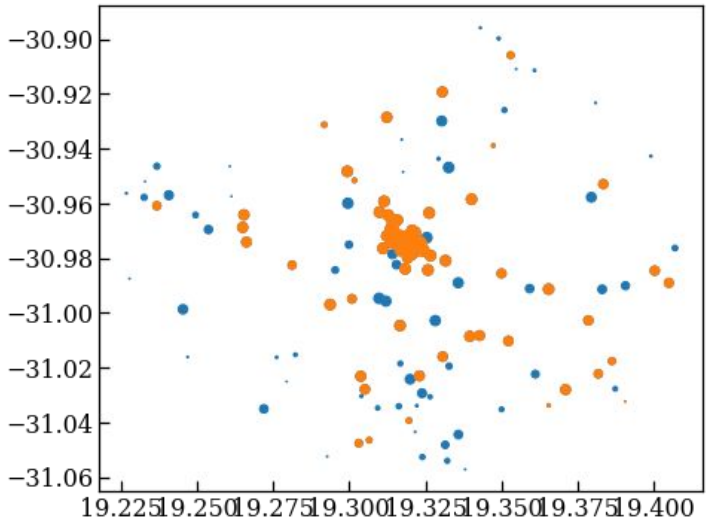
- **EMULATORS:**
  Create emulators for highly computational expensive functions
  - **We are currently working on the emulation of the (miscentered) lensing profile model which otherwise would require a 5d integration (current bottleneck of the pipeline)**

  Methodologies tested so far:
  - **Gaussian Process**
  - **Fully connected NN**
  - **Convolutional NN**

- **FEATURES EXTRACTION:**
  Train the NN with photometric cluster data, such as galaxy profile (along Radial and redshift directions), to extract unbiased estimates for the cluster richness (~true number of member galaxies)
  - **Currently exploring the feasibility of the analysis working with simulated data with different degree of realism**

  Methodologies tested so far:
  - **Fully connected NN**
  - **Convolutional NN**





Timescale, Milestones and KPIs

- **EMULATORS:**

- 3 months to find the best ML architecture and its optimization

- 3 months for its implementation in the EUCLID official cosmological pipeline

- **FEATURES EXTRACTION:**

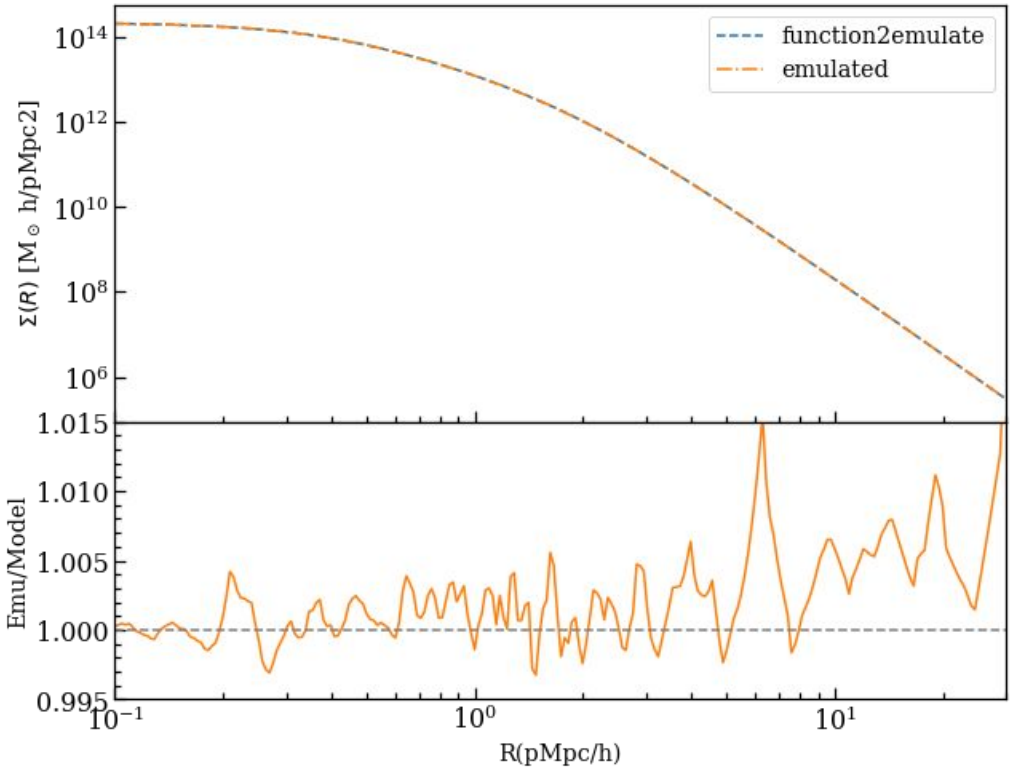
- 6 months to test different ML architectures and (simulated) data sets

Accomplished Work, Results

- **EMULATORS**

  GP are not suited for this task as the computational cost of the emulator scales with the number of training sample

  Fully connected NN seems to provide the best results so far (prediction/true <1%), but we are struggling to further increase the accuracy.

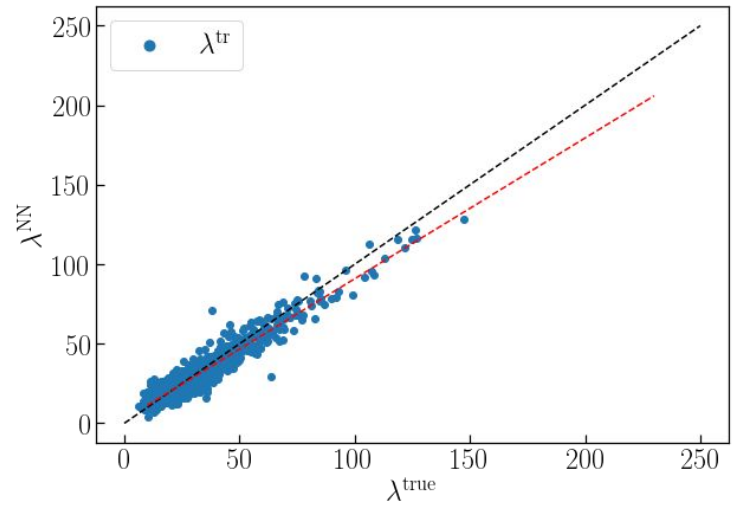


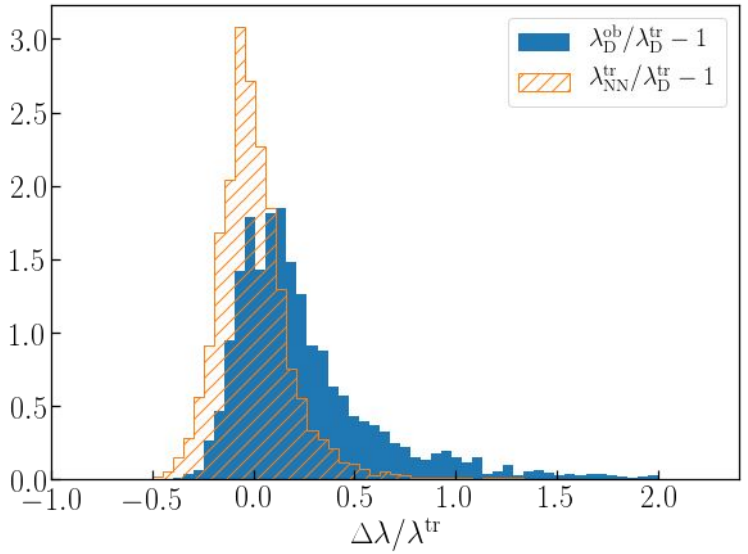
Accomplished Work, Results

- **FEATURES EXTRACTIONS**

  Both Fully connected and CNN provide good results when trained on simplistic simulated data

  Moving to more realistic simulations both architectures seem to fail in predicting the true richness, but the generation of the training set is more cumbersome and it might be the cause of the failure





Next Steps and Expected Results (by next checkpoint: April 2024)

- **EMULATORS**

  Increase the accuracy (1.0e-4) of the emulator(s)

  Extend this technique to other computationally expensive functions

- **FEATURES EXTRACTION**

  Additional tests on realistic mocks to assess the feasibility of the analysis

  Calibration of the redMaPPer DES Y1 photometric cluster catalog