My Work Between September 2021 and June 2023

Balázs *PÁL*[1,2]

[1]*Department of Physics of Complex Systems*, **Eötvös Loránd University**
[2]*Heavy-ion Research Group*, **Wigner Research Centre for Physics**

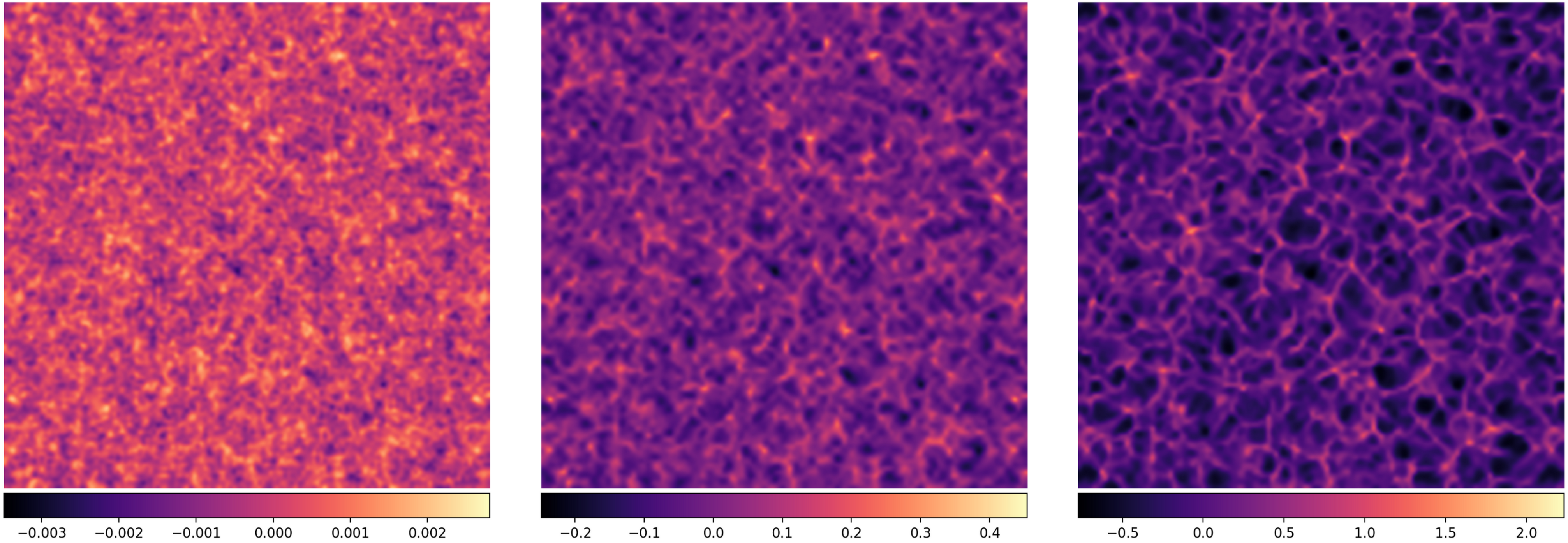
Eötvös Loránd University, Doctoral School of Physics
Astronomy and Space Physics Doctoral Program
Complex Exam – June 13, 2023

I. Cosmological Simulations

As part of István Csabai's research group
*@ **Eötvös Loránd University***
In cooperation with the
Wigner Research Centre for Physics
Supervised by István Csabai & Gergely Gábor Barnaföldi

EinsteinToolkit + FLRWSolver

Fig. 1. Bicubic interpolation of a *7.8125* Mpc/h thick slice of the density field in a cosmological GR simulation with a boxsize of *1000* Mpc/h showing **a)** the initial conditions, **b)** an intermittent snapshot, **c)** and the end of the simulation at $z = 0$. The initial conditions were generated using the FLRWSolver[1] and was evolved using the EinsteinToolkit[2].



a) b) c)

[1]: Hayley McPearson *et al.* (https://github.com/hayleyjm/FLRWSolver_public)

[2]: www.einstein-online.info

CAMELS – Connecting ML and simulations

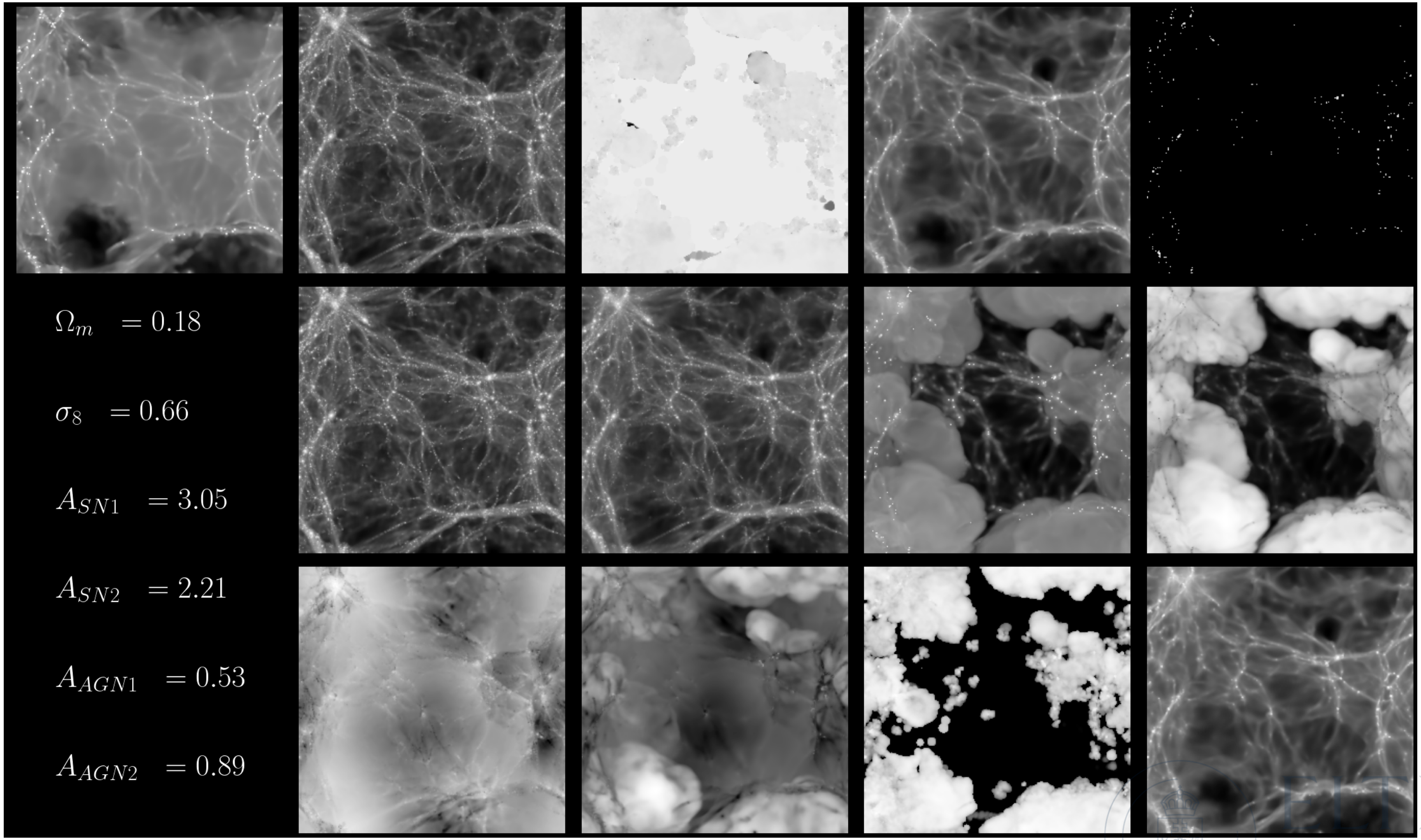


Fig. 2. A sample of the 12+1 fields of a single simulation in the CAMELS dataset. The fields here represent the same physical slice in the simulation. The 6 target labels attached to the simulation are shown in the bottom left corner.

CAMELS – Connecting ML and simulations

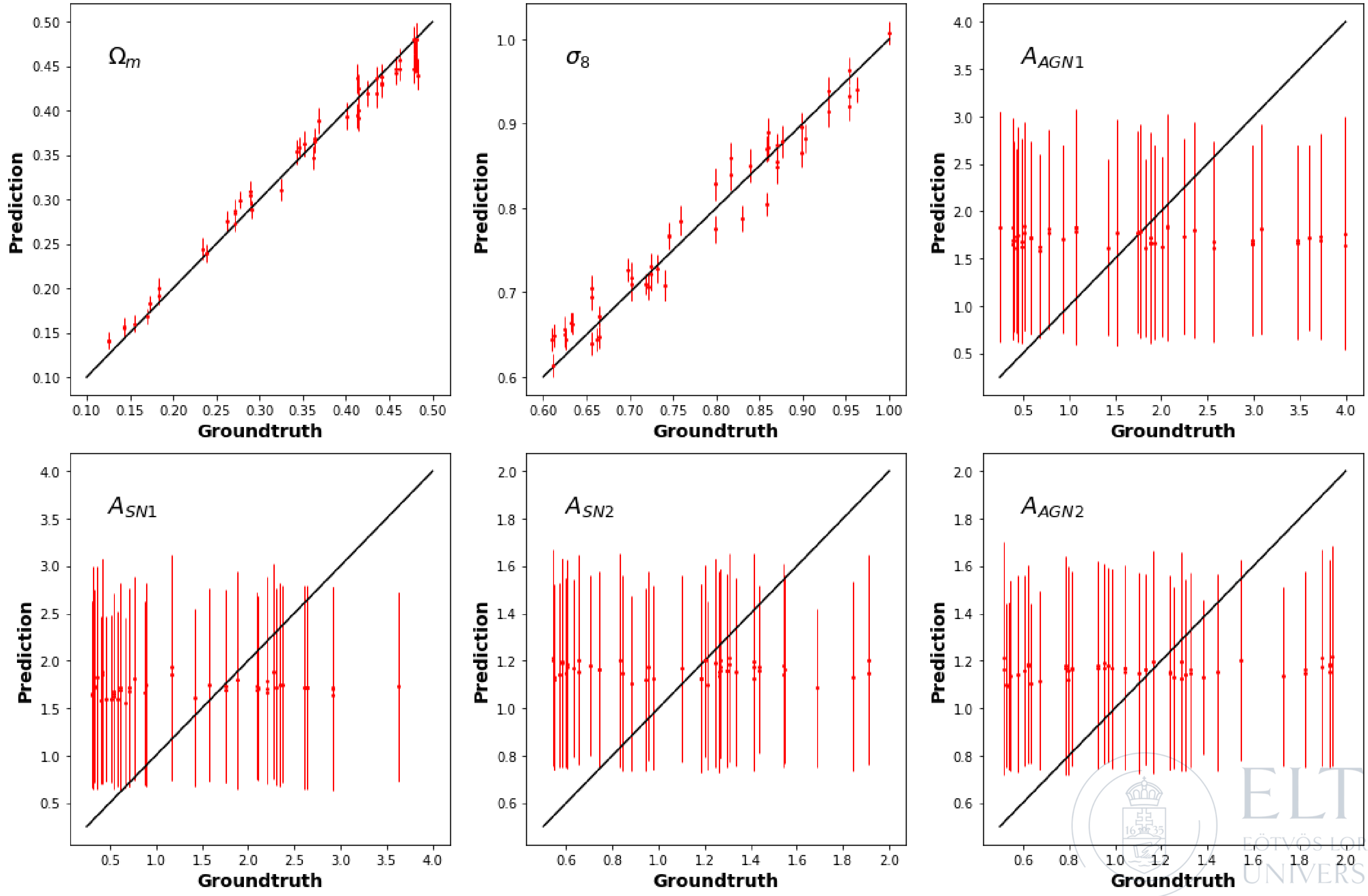


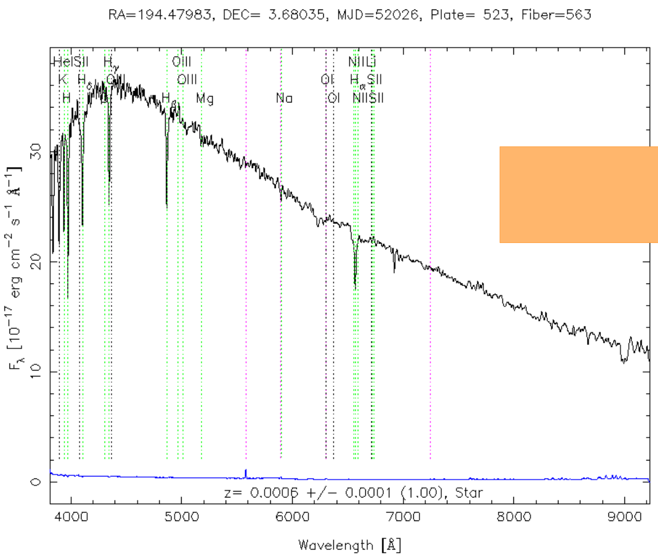
Fig. 3. Predictions compared to labels after training an ML model on the M_{CDM} fields.

A simple deep CNN model were separately trained on each sets of the 12+1 fields in the dataset. The goal was to accurately predict the Ω_m and σ_8 values, while omitting the astrophysical effects as much as possible. The figure here shows that it was successful for the M_{CDM} fields.

II. Galactic Archaeology

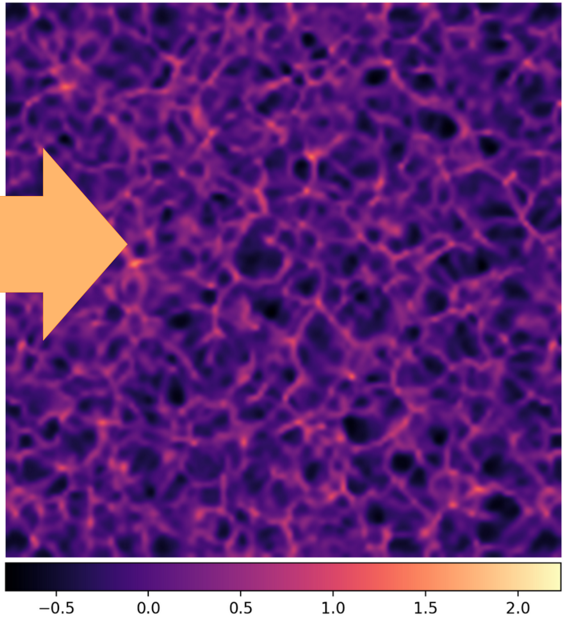
As part of the PFS Galactic Archaeology Research Group
@ Johns Hopkins University, Baltimore, MD
Supervised by László Dobos

Galactic Archaeology and the Subaru PFS



Source: Sloan Digital Sky Survey / SkyServer

Source: NASA, ESA, and the Hubble Heritage Team (STScI/AURA)-ESA/Hubble Collaboration



General relativistic cosmological simulation using the BSSN formalism – Own work

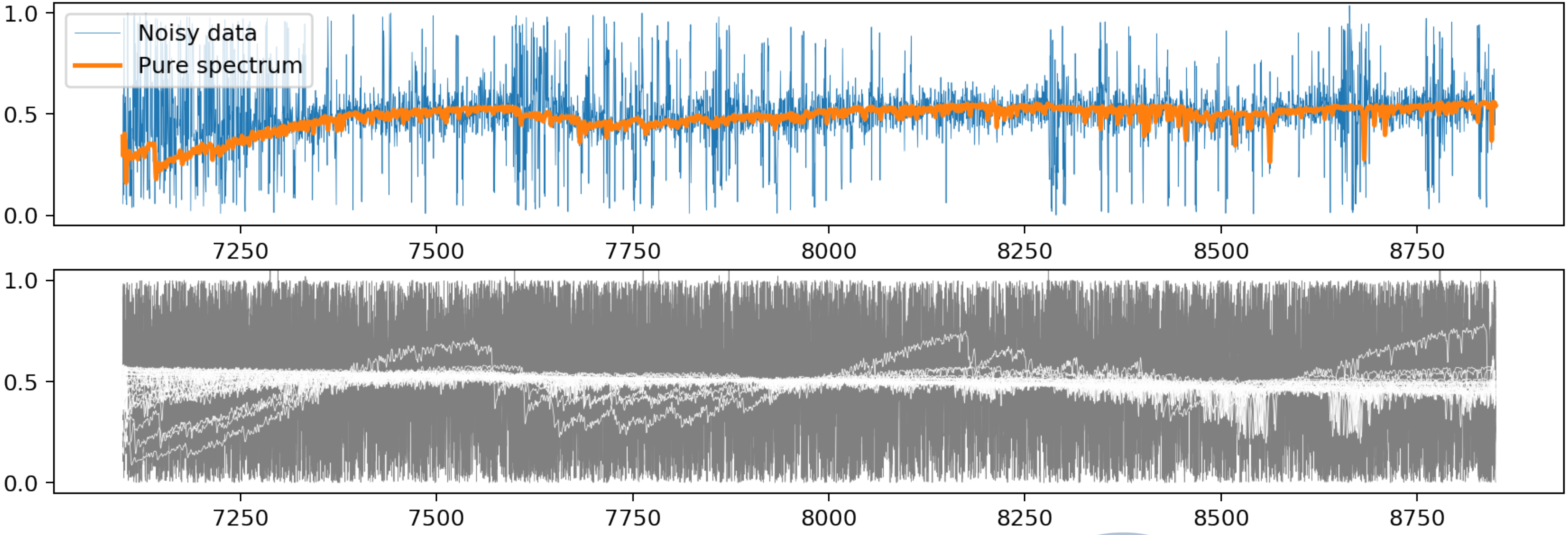
1. Stellar attributes

2. Galactic evolution

3. Cosmological implications

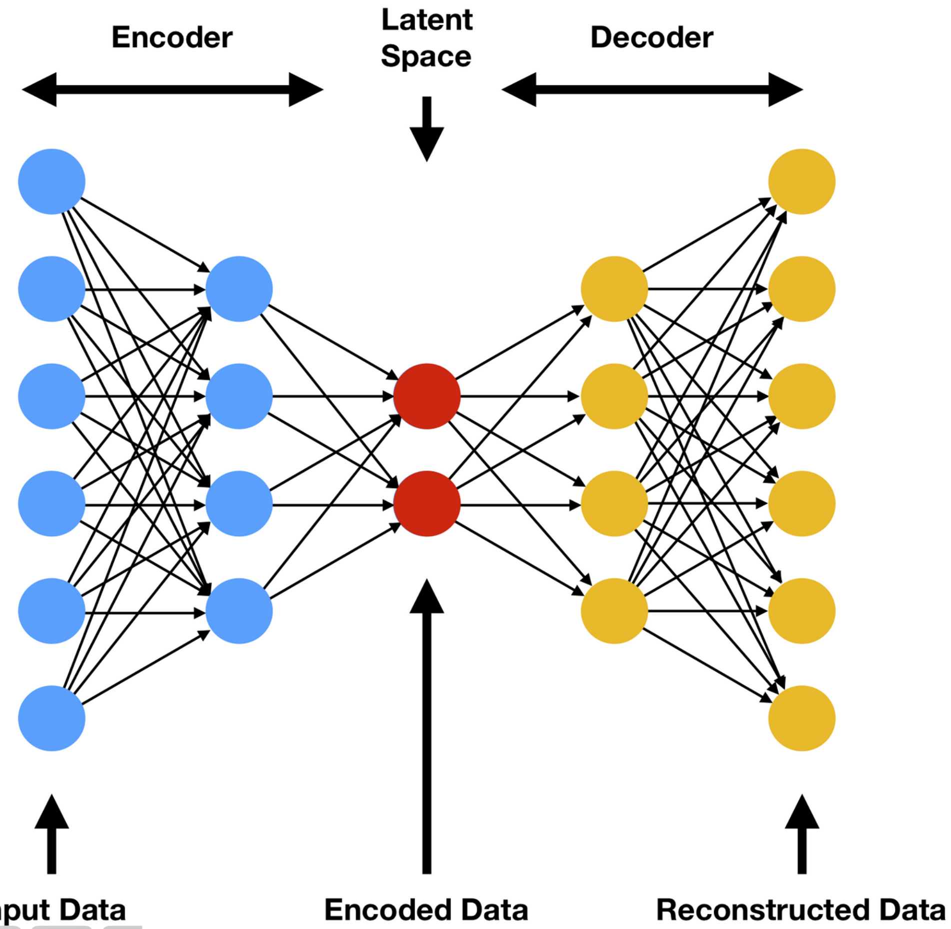
Example synthetic spectra

Fig. 4. Example of a single, artificially generated stellar spectrum with noise (blue line) and without noise (orange line) added (top panel). For additional visual clues about the dataset, a training batch of noisy (grey area) and noiseless (white area) spectra are also shown here (bottom panel).



Structure of the base autoencoder

- Dense Autoencoder
- Dimensions
  - Input/output: **4096**
  - Latent space: **< 4096**
- **Encoder** and the **Decoder**
  - 4+4 layers
  - Same amount of neurons
  - No batch normalization
  - No dropout layers



Source: Mahony, Niall O., et al. "Representation learning for fine-grained change detection." Sensors 21.13 (2021): 4486.

Example for denoising

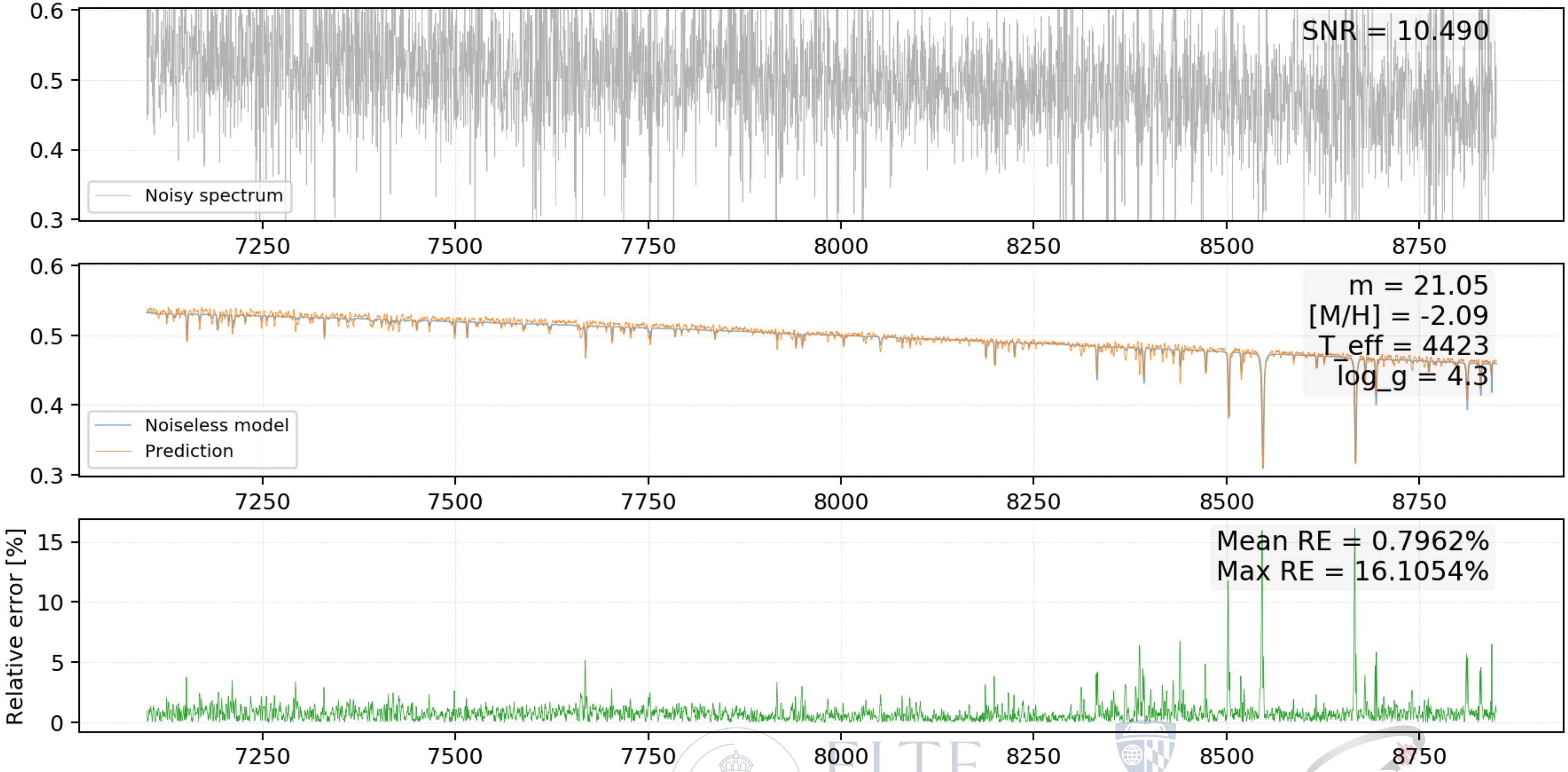


Fig. 5. A single noisy spectrum (top panel); a comparison between the groundtruth and the prediction of the trained autoencoder model (center panel); and the relative error between the groundtruth and the prediction (bottom panel).

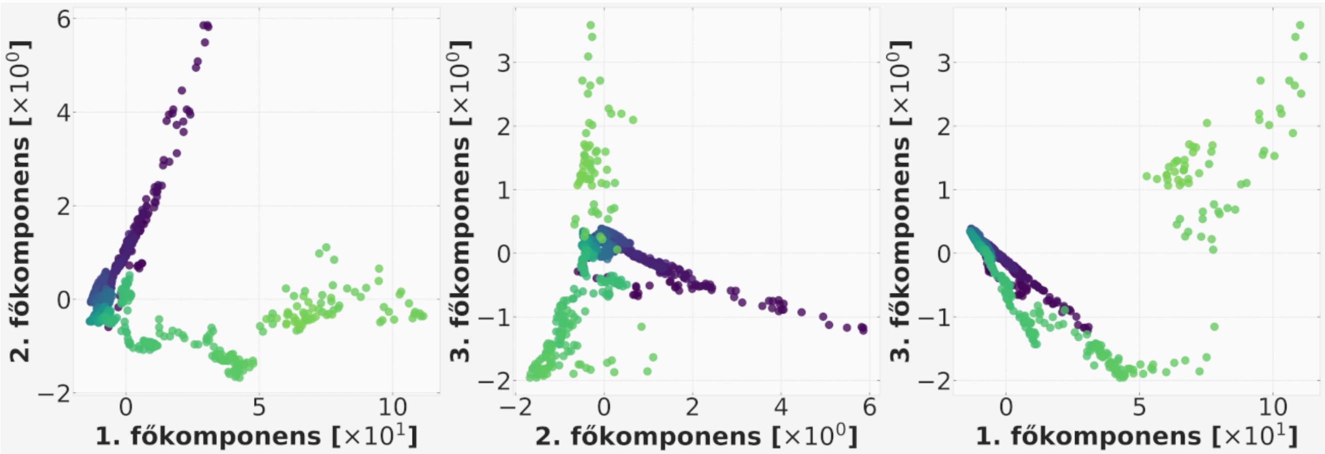
III. Unraveling Crypto-assets

Co-authoring with Tamás Pál
*@ **Eötvös Loránd University***

Economic regimes and regime changes – November 26, 2021

Results of Principal Component Analysis

Fig. 6. *(on the right)* Data are projected onto the first three principal components, with a color gradient indicating the temporal progression from dark to light. Each axis operates on a unique scale, as denoted by the respective labels. Dominating the variance, the 1st principal component is an order of magnitude larger than the subsequent two.



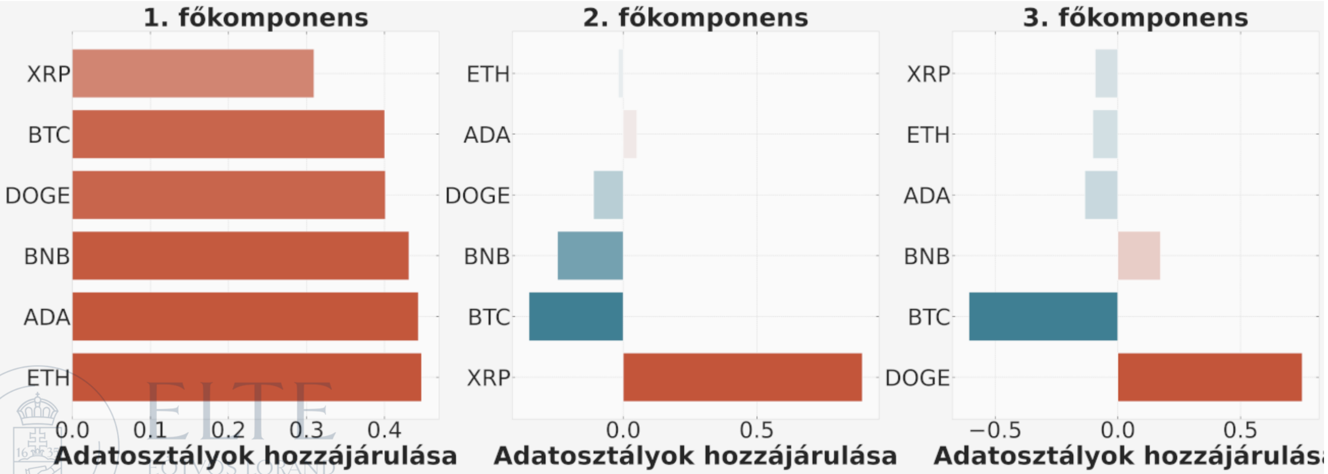


Fig. 7. *(on the left)* The individual contributions of data points from each crypto-asset to the three principal components are depicted. Although the differences are small, the crypto-assets **ETH** and **ADA** seems to be the main drivers of price fluctuations. On the other hand, the 2nd and 3rd principal components tell a story about specific, influential events in the crypto world, eg. the race between **BTC** and **XRP**, as well as Elon Musk's internet campaign to promote the Dogecoin (**DOGE**).

Results of other supervised ML methods

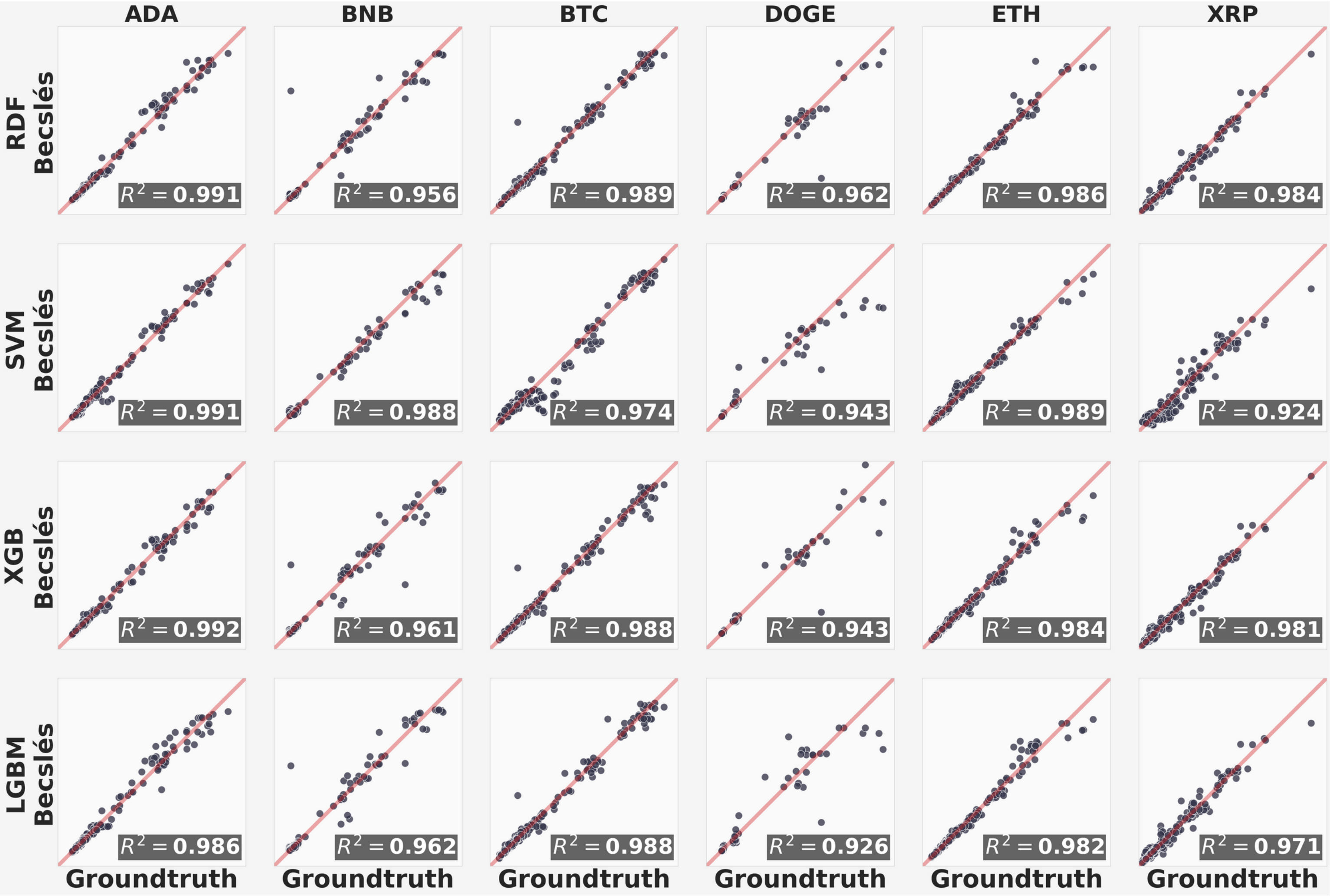


Fig. 8. Each subplot within the 4×6 grid represents the predictive performance of a specific machine learning model (Random Forest, SVM, XGBoost, LightGBM) for a particular crypto-asset. The plots depict the correlation between the predicted values and the actual values. Root Mean Square Error (RMSE), displayed in the lower right corner of each subplot, serves as the accuracy metric for the predictions. Analysis shows that the crypto-asset **ADA** shows the highest predicative power, while **DOGE** is the worst one.

ELTE
EÖTVÖS LORÁND
UNIVERSITY

IV. SARS-CoV-2 Reconstruction

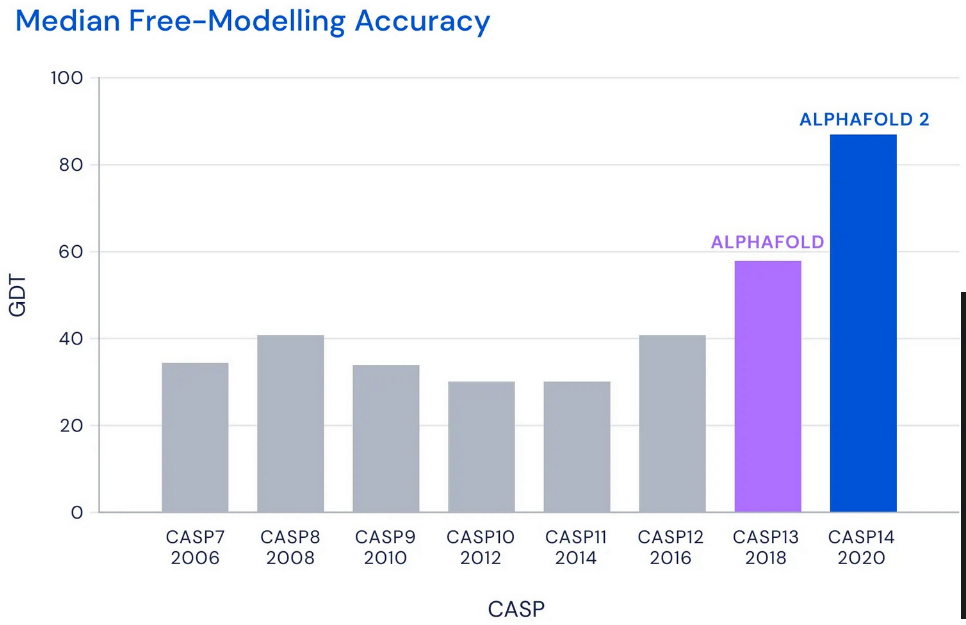
Co-authoring with Ákos Gellért[1], Oz Kilim[2], Anikó Mentes[2] and István Csabai[2]
*@ [1]**Veterinary Medical Research Institute** and [2]**Eötvös Loránd University***
In cooperation with the
Wigner Research Centre for Physics

AlphaFold2 in COVID research

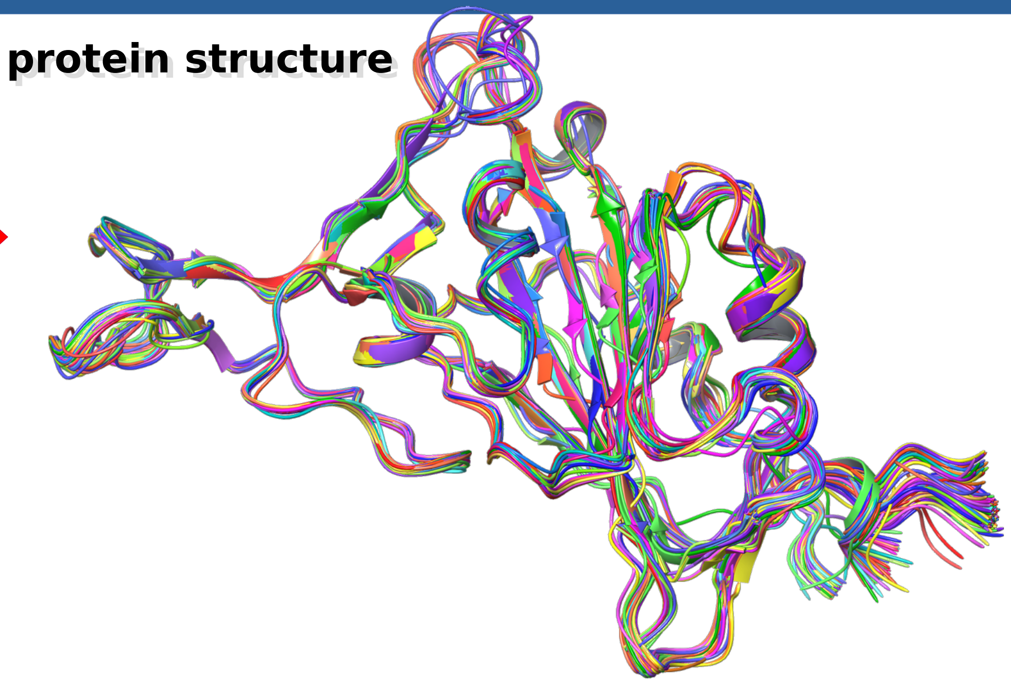
OpenAI

Arbitrary chain of amino acids

VLLYTNEDKGTNFQDRMLGWTVKGK
WTWHYIEIGVFQMIVQWNDTQAIKM
QHECFQQRQHRLYTWEENTCSAFM
CGYYMLSPPWPACDVRIGWLRALVR
NVHSSLWHWINASTAPSNHLGMALM
MYYNCMQPFIIEKQGFERQVDCKID



3D protein structure



Data Descriptor | Open Access | Published: 14 March 2023

SARS-CoV-2 receptor-binding domain deep mutational AlphaFold2 structures

Oz Kilim, Anikó Mentes, Balázs Pál, István Csabai & Ákos Gellért

Scientific Data **10**, Article number: 134 (2023) | Cite this article

Conferences, Lectures, Presentations

Conferences, lectures, presentations

Presentations given at conferences

- Economic regimes and regime changes – November 26, 2021
- XIII. GPU Day(s) – May 15-16, 2023

Lecturer on courses at ELTE

- Modern Computational Methods in Physics 1 – 2022, 2023
- Modern Computational Methods in Physics 2 – 2021
- Scientific Modeling Computer Laboratory – 2023
- Data Mining and Machine Learning – 2022 (only at laboratory courses)

Lectures given at other occasions

- ELTE MÁSZ Workshop Seminar – December 10, 2021
- ELTE GTK Department of Comparative Economics Research Seminar – June 2, 2022
- Wigner RCP IPNP Seminar – July 29, 2022
- Lecture about Data Science and its applications – BGE KVIK – April 26, 2023

Courses Taken

Courses taken during the PhD program

2nd Semester

- Advanced infomation technology in astronomy II. (FIZ/5/008) by Emese Forgácsné Dajka
- Data mining in astronomy (FIZ/5/006) by István Csabai
- Solar physics (FIZ/5/057) by Kristóf Petrovay
- The structure of compact stars (FIZ/5/025) by Gergely Gábor Barnaföldi

3rd Semester (from abroad)

- Radio astronomy I. (FIZ/5/009) by Krisztina Gabányi & Sándor Frey
- Space weather and space climate (FIZ/5/059) by Kristóf Petrovay

4th Semester

- Radio astronomy II. (FIZ/5/010) by Krisztina Gabányi & Sándor Frey
- Linear and nonlinear MHD waves (FIZ/5/054) by Róbert von Fáy-Siebenbürgen

Thank you for your attention!