

Astroinformatics 2026

University of Cologne, Germany

23 – 27 February 2026

Proceedings



ASTROINFORMATICS

2026 February 23–27
Cologne, Germany



SCAN ME

e-Science
infrastructures

Survey & time-domain
astronomy

Knowledge discovery &
data mining

Visualisation & data
exploration

Database systems &
data-intensive projects

INVITED SPEAKERS:

Michelle Lochner, Department of Physics
and Astronomy, University of Western Cape

Ofer Lahav, Astrophysics Group, Faculty of Mathematical
and Physical Sciences, UCL

Guillermo Cabrera-Vives, Department of
Computer Science, University of Concepción

Natalia Korsakova,
Observatoire de la Cote d'Azur

Lucie Flek, Bonn-Aachen International Center
for Information Technology, University of Bonn

Anna Scaife, Jodrell Bank Centre for Astrophysics,
University of Manchester

Andreas Lintermann, Jülich Supercomputing Centre,
Forschungszentrum Jülich

Thomas Lippert, Jülich Supercomputing Centre,
Forschungszentrum Jülich

LOC:

Lennart Buhlmann
Alice Coors
Nuray Ortaköse
Atharv Patade
Tatiana Rodriguez
Frank Schlöder
Aditi Sinha
Iraj Vaezzadeh
Stefanie Walch-Gassner (Chair)
Ana Brás Würschig
Birka Zimmermann

SOC:

Sotiria Fotopoulou
Nikos Gianniotis
Sebastian Trujillo Gomez
Matthew Graham
Ashish Mahabal
Jess McIver
Susanne Pfalzner
Kai Polsterer
Frank Wagner
Stefanie Walch-Gassner

Astroinformatics 2026 Contact:

Stefanie Walch-Gassner
I. Physics Institute University of
Cologne, Germany ai-2026@uni-koeln.de



UNIVERSITY
OF COLOGNE



Program

Invited talks (30+10)	Contributed talks (15+5)
-----------------------	--------------------------

Begin	End	Monday	Tuesday	Wednesday	Thursday	Friday
9:00	9:30	Registration / Welcome	Registration	Registration	Registration	Registration
		Machine Learning for Cosmology	Confronting Models with Data	Data-driven Discovery	Survey and Time-Domain Data I	Survey and Time-Domain Data III
9:30	9:50	Keynote: Ofer Lahav	Natalia Korsakova	Anna Scaife	Guillermo Cabrera-Vives	Szymon Nakoneczny Andrei Kazantsev Jakob Dietl Claudia Comito
9:50	10:10		Dario Barone	Máté Madarász	Nabeel Rehemtulla	
10:10	10:30	Coffee break + Poster display				
10:30	11:00					
11:00	11:20	Simon Verley	Romain Chazotte	Simone Vilardi	Joaquín Gallardo-Villagrán	Final Discussion
11:20	11:40	James Kostas Ray	Amit Kashi	Catarina Corte-Real	Inês Rolo	
11:40	12:00	Toka Alokda	Maksim Gabdeev	Dmitrii Zagorulia	Aarya Patil	
12:00	14:00	Lunch				
		Exascale Computing and FAIR Data Platforms I	Exascale Computing and FAIR Data Platforms II	AI/NLP for Workflows and Education	Survey and Time-Domain Data II	
14:00	14:40	Andreas Lintermann	Daniel Seifried	Lucie Flek	Michelle Lochner	
14:40	15:00	Luis Cifuentes	Discussion	Pavlos Protopapas	Harwool Koo	
15:00	15:40	Coffee break + Poster display + VR Tour (Tues+Thurs)				
15:40	16:00	Marco Bischoff	Historical centre walking tour 15:30 - 17:30	Mher Khachatryan	Matthew Graham	
16:00	16:20	Panel discussion		André Bresges	Paul Nathan	
16:20	16:40			Panel discussion	Panel discussion	
16:40	17:00	Poster pitch				
		Welcome reception 17:30 - 22:30		Conference dinner 18:30 - 22:30	Astronomy on Tap at Pub 19:00 - 22:30	

Contents

Abstracts for Invited & Contributed Talks	1
Machine learning for Cosmology	1
Keynote Speaker: Ofer Lahav (<i>AI for Astronomy and Astronomy for AI</i>)	1
Mimicking the large-scale structure of the local Universe. Synthetic pre-labeled galaxies in large-scale structures (<i>Manuel Luis Alcázar-Laynez, Salvador Duarte Puertas, Simon Verley, Guillermo Blázquez-Calero, Andoni Jiménez, Antonio Lorenzo-Gutiérrez, Daniel Espada, María del Carmen Argudo-Fernández, Isabel Pérez</i>)	1
Reinforcement Learning for Adaptive Tomographic Redshift Calibration (James Kostas Ray)	2
Segmenting proto-halos with vision transformers DOI: 10.1088/1475-7516/2025/11/083 (Toka Alokda, Cristiano Porciani)	2
Exascale computing and data platforms I	3
Session Invited Speaker: Andreas Lintermann (<i>Exascale HPC for Scalable AI and Simulation Workflows</i>)	3
SPECTRUM Technical Blueprint and Strategic Agenda: Europe’s Roadmap to Exabyte-Scale Scientific Infrastructures (Luis Cifuentes, Andreas Lintermann, Sergio Andreozzi)	3
Connecting Simulations and Observations: A Reproducible Framework for Exploring Solar System Formation with Trans-Neptunian Objects (Marco Bischoff, Susanne Pfalzner, Frank Wagner)	4
Confronting Models with Data	5
Session Invited Speaker: Natalia Korsakova (<i>Laser Interferometer Space Antenna (LISA): what it can observe and how we can extract it from the data</i>)	5
Interpretable probabilistic modelling approach to correct Hubble flow around galaxy clusters (Dario Barone, Marco Canducci, Rory Smith, Peter Tino)	5
Improved O(2) Equivariance via Embedding in Zernike Polynomial Space (Romain Chazotte, Kai Polsterer)	5
Wind Collisions and Giant Eruptions in Massive Binaries, and their Effects on the Involved Stars (Amit Kashi)	6
Effective temperature prediction for Gaia eclipsing binaries using ensemble of random forest models (Štefan Parimucha, Maksim Gabdeev)	6
Exascale Computing and FAIR Data Platforms II	6
The filamentary ISM: Structure identification and characterisation (Daniel Seifried)	6
Data-driven Discovery	7
Session Invited Speaker: Anna Scaife (<i>Building foundational representations from large-scale surveys</i>)	7
Multimodal Deep Learning for Young Stellar Object Classification and Characterization (Máté Madarász, Gábor Marton)	7

From Orbital Space to Phase-Space Ridges in the Milky Way Disc (<i>Simone Vilardi, Subhadeep Sarkar, Teresa Antoja Castelltort, Francesca Figueras Siñol, Kerstin Bunte</i>)	7
Physical analysis of galaxies via spectral reconstruction of deep imaging (<i>Catarina Cortes-Real</i>)	8
Morphological Classification of Jets in Active Galactic Nuclei (<i>Dmitrii Zagorulia</i>)	8
AI/NLP for Workflows and Education	9
Session Invited Speaker: Lucie Flek (<i>LLMs and LLM Agents in Astrophysical Workflows: Promise, Pitfalls, and Practice</i>)	9
Semantic Citation Prediction (<i>Pavlos Protopapas</i>)	9
Astrogenesis: A Hierarchical Multi-Agent System for Automated Blazar Research and Literature-Augmented Analysis (<i>Mher Khachatryan, Narek Sahakyan, Artyom Khachatryan, Vika Markosyan</i>)	10
Teaching Astroinformatics in Classrooms: A critical path (<i>André Bresges, Kathleen Falconer</i>)	10
Survey and Time-Domain Data I	11
Session Invited Speaker: Guillermo Cabrera-Vives (<i>Lessons from Time-Domain Astronomy: Why One Modality Is No Longer Enough</i>)	11
Automating real-time transient workflows to facilitate discovery (<i>Nabeel Rehemtulla, Adam A. Miller, Wynn Jacobson-Galán</i>)	11
Measuring and Mitigating Galaxy Labeling Bias from Low to High Redshift: A Deep Learning Approach (<i>Joaquín Gallardo-Villagrán, Guillermo Cabrera-Vives, Rodrigo Herrera-Camus, Monserrat Martínez-Marín</i>)	12
Finding Needles in the TESS Haystack: Automated detection of roAp Stars from TESS data (<i>Inês Rolo, Daniel L. Holdsworth, Margarida S. Cunha, Victoria Antoci, Donald W. Kurtz, Rahul Jayaraman, Ângela R. G. Santos</i>)	13
Improving Time-Series Analysis in Astronomy: Understanding Stars, the Milky Way, and Beyond (<i>Aarya Patil</i>)	13
Survey and Time-Domain Data II	14
Session Invited Speaker: Michelle Lochner (<i>Scientific discovery with foundation models</i>)	14
Detecting anomalous radio spectrograms with unsupervised and active learning (<i>Hanwool Koo, Michelle Lochner</i>)	14
Compose, Score, and Explain: a compositional framework for AGN variability (<i>Matthew Graham</i>)	15
Anomaly Detection in DESI Quasar Spectra using Machine-Learning Dimensionality-Reduction Techniques (<i>Paul Nathan</i>)	15
Survey and Time-Domain Data III	15
Modeling Distribution and Photo-z Uncertainty of Quasars: An Analysis Based on the Kilo-Degree Survey (<i>Kacper Drabicki, Szymon J. Nakoneczny, Maciej Bilicki</i>)	16
The Use of Minimalistic Convolutional Architectures for Detection of Astronomical Radio Signals (<i>Kazantsev, Andrei, Ramesh Karuppusamy, Michael Kramer</i>)	16
Uncovering the Giants in the Sky - Semantic Segmentation of Galaxy Clusters in Astrophysical Surveys (<i>Jakob Dietl, Thomas Reiprich, Florian Pacaud, Alexander Rüttgers, Zorah Löhner, Angie Veronica</i>)	17
(<i>Claudia Comito</i>)	17
Posters	18
Functional Magnetic Resonance Imaging of Human Brain Functions during simulated car driving: Possible applications for a Shared Universe Engine (<i>André Bresges</i>)	19

Representation Learning for Gaia XP DR3 (<i>Bernd Doser, Kai Polsterer, Sebastian Trujillo-Gomez</i>)	20
Using the Perception of Physics among Adolescent Pupils as an Example of Doing Design-Based Research (<i>Kathleen Falconer, André Bresges</i>)	21
Compute4PUNCH: A platform for the analysis of Astronomical Data (<i>Prateek Gupta</i>) .	22
Astronomical Bycatch: Extracting exoplanet transits from non-target stars in astronomical archives. (<i>Nika Gorchakova, Creaner Oisin, Valentino Farrell, David Lawley, Cormac McGleenan</i>)	23
Following Tracer Particles to Identify Filaments in Star-Forming Cores with HDBSCAN (<i>Nuray Ortaköse</i>)	24
slideParam: 2D galactic data visualisation tool in Python (<i>A. Jimenez, M. Alcázar-Layne, S. Verley, S. Duarte Puertas, M. Argudo-Fernández, D. Espada</i>)	25
MockLSS: A geometrical large-scale structure simulator of the local universe to make synthetic observational catalogues (<i>M. Alcázar-Layne, S. Duarte Puertas, S. Verley, G. Blázquez-Calero, A. Jimenez</i>)	26
LaVInComPy: LArge Visualisation INterface COMputed with PYthon (<i>S. Duarte Puertas, A. Jimenez, M. Alcázar-Layne, M. Argudo-Fernández, S. Verley, D. Espada, M. Hernández-Sánchez, G. Blázquez-Calero, C. Pozo González</i>)	27
GALAssify: A Python package for visually classifying astronomical objects (<i>M. Alcázar-Layne, A. Jimenez, S. Duarte Puertas, C. Pozo González, J. Domínguez-Gómez, G. Blázquez-Calero, M. Argudo-Fernández, S. Verley, D. Espada, E. Florido, I. Pérez, T. Ruiz-Lara, L. Sánchez-Menguiano, A. Zurita</i>)	28
Author Index	29

Abstracts for Invited & Contributed Talks

Machine learning for Cosmology

Session Chair: Stefanie Walch-Gassner & Guillermo Cabrera-Vives

Keynote Speaker: Ofer Lahav
AI for Astronomy and Astronomy for AI
Speaker Affiliation: UCL

23 Feb
09:30

Mimicking the large-scale structure of the local Universe. Synthetic pre-labeled galaxies in large-scale structures

23 Feb
11:00

Manuel Luis Alcázar-Laynez, Salvador Duarte Puertas, **Simon Verley**, Guillermo Blázquez-Calero, Andoni Jiménez, Antonio Lorenzo-Gutiérrez, Daniel Espada, María del Carmen Argudo-Fernández, Isabel Pérez

Speaker Affiliation: University of Granada

Current observational and simulated large-scale structure (LSS) catalogues often lack consistency in assigning galaxies to specific structures, due to the absence of a universally accepted classification criterion.

In order to generate synthetic empirical data for fine-tuning LSS classification algorithms, as well as train Machine Learning (ML) / Deep learning (DL) models with the same purpose, this work presents a purely geometrical simulation based on the statistical spatial properties found in LSS surveys, using the spectroscopic main galaxy sample of the Sloan Digital Sky Survey (SDSS) catalogue up to a redshift of $z \approx 0.1$ as a specific use case.

A parallelism between the LSS and the 3D Voronoi tessellation is utilised, in which the nodes, links, surfaces, and cells of the diagram correspond to clusters, filaments, walls, and voids, respectively. The simulation uses random positions within voids as seeds for tessellating the 3D space. The resulting tessellation structures are then randomly populated with galaxies that adhere to the statistical properties of their observational respective structures. As the galaxies are generated, they are tagged with their corresponding structure.

In each simulation, six LSS mock catalogues are generated, following the statistical behaviour observed in the SDSS catalogue, depending on the structure that they belong to. In addition, the Malmquist bias and the redshift-space distortion (known as Fingers of God effect) are simulated as well.

A novel geometrical LSS simulator is presented, where generated galaxies mimic the statistical properties of their observational belonging structure. As an example, the simulator is tuned to mimic the SDSS catalogue, although any other catalogue could be considered. With the generated catalogue, it is possible to adjust LSS classification algorithms, train and test ML/DL models, and benchmark several LSS classification methods using this pre-labeled data to contrast its results and performance.

23 Feb
11:20

Reinforcement Learning for Adaptive Tomographic Redshift Calibration

James Kostas Ray

Speaker Affiliation: UCL

Photometric redshift calibration is a central challenge for forthcoming cosmological surveys, where billions of galaxies must be assigned to redshift bins with stringent requirements on bias and stability. Self-organising maps (SOMs) provide a structured representation of galaxy colour space, but existing bin-assignment procedures typically rely on stochastic or heuristic optimisation, but this is a non-trivial problem. As a result, node-level assignments can become locally suboptimal and it becomes difficult to fully explore the SOM space due to the stochasticity.

This project reframes SOM redshift binning as a sequential decision-making task that can be optimised using reinforcement learning (RL), an area of machine learning not used in astronomy outside of observation scheduling. This work investigates policy architectures suited to structured 2D grid problems, including local CNN-based “Minesweeper-style” agents and node-wise graph-based policies that exploit the topology of the SOM. This opens the door to more adaptive and data-efficient binning strategies for future cosmological analyses and hopefully encourages more uses of RL in the field.

Segmenting proto-halos with vision transformers

DOI: [10.1088/1475-7516/2025/11/083](https://doi.org/10.1088/1475-7516/2025/11/083)

Toka Alokda, Cristiano Porciani

Speaker Affiliation: University of Bonn

The formation of dark-matter halos from small cosmological perturbations generated in the early universe is a highly non-linear process typically modelled through N-body simulations. In this work, we explore the use of deep learning to segment and classify proto-halo regions in the initial density field according to their final halo mass at redshift $z = 0$. We compare two architectures: a fully convolutional neural network (CNN) based on the V-Net design and a U-Net transformer. We find that the transformer-based network significantly outperforms the CNN across all metrics, achieving sub-percent error in the total segmented mass per halo class. Both networks deliver much higher accuracy than the perturbation-theory-based model pinocchio, especially at low halo masses and in the detailed reconstruction of proto-halo boundaries. We also investigate the impact of different input features by training models on the density field, the tidal shear, and their combination. Finally, we use Grad-CAM to generate class-activation heatmaps for the CNN, providing preliminary yet suggestive insights into how the network exploits the input fields.

23 Feb
11:40

Exascale computing and data platforms I

Session Chair: Frank Wagner

Session Invited Speaker: Andreas Lintermann

23 Feb
14:00

Exascale HPC for Scalable AI and Simulation Workflows

Speaker Affiliation: Jülich Supercomputing Centre

AI technologies are nowadays used in almost every area of research and industry. They automate and accelerate processes, and uncover new, unseen relations in huge datasets. Such technologies have become key to develop new models for complex simulations, e.g., reduced-order models, surrogates, or closure models. Modular and heterogeneous systems with accelerator and/or specialized AI components are found in state-of-the-art Exascale HPC systems, providing the opportunity to deal with the demands of complex and intertwined simulations and AI data processing workflows.

This talk provides an overview of European state-of-the-art HPC technologies and develops an understanding from hardware over application to workflow perspective, with an emphasis on intertwining simulations and AI. It specifically discusses training AI at scale for a physics application, with recommendations on increasing the training performance by intelligently selecting the right parameters for training and optimization. These recommendations are complemented by a presentation of tuning strategies to improve model accuracy by means of hyperparameter optimization techniques that make efficient use of the underlying HPC hardware. It is shown how such technologies embed into machine-learning operations (MLOps) workflows that are capable of running on large-scale HPC infrastructures, and how they might benefit future research applications.

SPECTRUM Technical Blueprint and Strategic Agenda: Europe's Roadmap to Exabyte-Scale Scientific Infrastructures

23 Feb
14:40

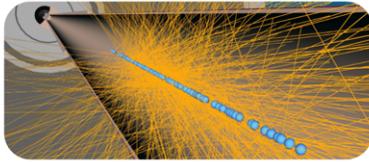
Luis Cifuentes, Andreas Lintermann, Sergio Andreatto

Speaker Affiliation: Jülich Supercomputing Centre

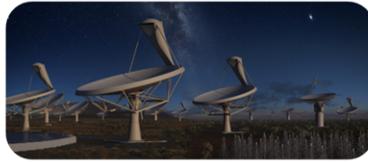
This presentation introduces the public draft of the SPECTRUM Technical Blueprint and the Strategic Research, Innovation and Deployment Agenda (SRIDA) for European data-intensive scientific infrastructures. The Technical Blueprint focuses on the technical capabilities, while the SRIDA defines the strategic and policy framework for a unified European compute and data continuum. Both documents are grounded in detailed requirements gathered from the High-Energy Physics and Radio Astronomy communities, including the exabyte-scale data challenges of the High-Luminosity Large Hadron Collider (HL-LHC), the extreme computational demands of the Square Kilometre Array (SKA), and the distributed processing needs of LOW Frequency ARray (LOFAR).

The SPECTRUM project proposes a comprehensive framework to address the fragmentation of Europe's current computing landscape through an integrated, scalable architecture. The SRIDA complements this by outlining concrete priorities, implementation pathways, and policy recommendations for policymakers, infrastructure providers, and research communities. Together, they define the technical capabilities and governance structures required for seamless workload portability across heterogeneous infrastructures, while preserving sovereignty and minimizing environmental impact.

The presentation highlights the key architectural elements, strategic priorities, and implementation routes identified through community-driven analyses. As the project is currently in its open consultation phase, feedback from the advanced computing community is actively sought to ensure the final Technical Blueprint and SRIDA effectively address Europe's research infrastructure needs.



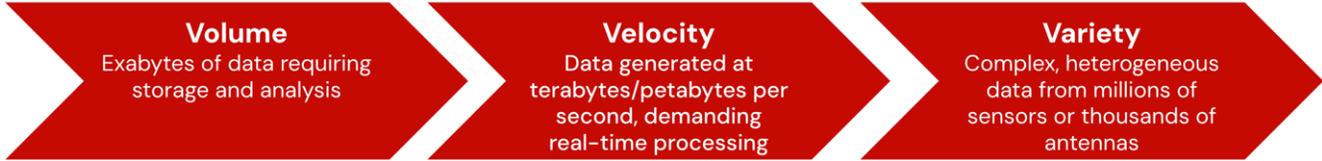
High-Luminosity Large Hadron Collider (HL-LHC)



Square Kilometre Array (SKA)



Low Frequency ARray (LOFAR) 2.0



The Exascale Challenge: A New Era for Science

High-Energy Physics (HEP) and Radio Astronomy (RA) are entering the Exascale era, driven by next-generation instruments.

Connecting Simulations and Observations: A Reproducible Framework for Exploring Solar System Formation with Trans-Neptunian Objects

23 Feb
15:40

Marco Bischoff, Susanne Pfalzner, Frank Wagner

Speaker Affiliation: Jülich Supercomputing Centre

Trans-Neptunian Objects (TNOs) are among the most pristine remnants of the early Solar System. Their orbital dynamics provide powerful tests for competing hypotheses of Solar System formation. By evaluating how well different models (e.g., the stellar flyby scenario) reproduce the observed distribution of TNO orbits, we aim to identify the most plausible processes that shaped the outer Solar System. As upcoming sky surveys from the Rubin Observatory are expected to dramatically increase the number of known TNOs, automating these model comparison workflows has become increasingly important. However, detection biases in astronomical data complicate direct comparisons between simulated and observed populations. To address this, we are training a classifier to estimate the likelihood of each observed or simulated object belonging to a hypothesised TNO population, and will use a heuristic-based approach to evaluate the similarity between data and models. In this talk, we introduce a modular, open-source data pipeline that supports reproducible and transparent model testing. Designed in alignment with the FAIR principles (Findable, Accessible, Interoperable, and Reusable), the framework enables standardised dataset handling, efficient pipeline construction, and flexible integration of alternative formation hypotheses.

Confronting Models with Data

Session Chair: Sebastian T. Gomez

Session Invited Speaker: Natalia Korsakova

24 Feb
09:30

Laser Interferometer Space Antenna (LISA): what it can observe and how we can extract it from the data

Speaker Affiliation: Observatoire de la Cote d'Azur

Interpretable probabilistic modelling approach to correct Hubble flow around galaxy clusters

24 Feb
10:10

Dario Barone, Marco Canducci, Rory Smith, Peter Tino

Speaker Affiliation: University of Birmingham

Cosmological theories rely on understanding the three-dimensional structure of the Universe on large scale. Astrophysical observations of galaxy clusters are limited to measuring a two-dimensional position in the sky and a one-dimensional velocity along the line of sight, inherently losing part of the relevant information content. Commonly, the third spatial dimension is reconstructed by applying a deterministic scaling of the line-of-sight velocity that accounts for the expansion of the Universe (Hubble flow). However, due to gravitational effects, this transformation does not hold in the vicinity of massive structures, such as galaxy clusters. In this work, we propose an interpretable probabilistic model aiming to reconstruct the three-dimensional radial distance of an observed galaxy from its nearest galaxy cluster. Our model is completely transparent and it explicitly embeds astrophysical prior knowledge to account for the non-trivial dynamics that are superimposed on the Hubble flow. The three-dimensional radial distance is reconstructed in the form of a posterior distribution, conditioned on the positional and kinematical observations, which enables coherent quantification of the predicted uncertainties. We train our model on data from cosmological simulations, and we then use it to reconstruct the actual cluster neighbourhoods. We empirically show that our model is more accurate in recovering the radial profile of several simulated clusters, for which the ground truth is known, compared to the standard techniques. Crucially, our model-driven approach also uncovers features of simulated cosmological data that are not yet fully understood from a theoretical perspective, suggesting new directions of research. This tool is designed for use by astronomical observers, offering improvements over existing methods and opening up possibilities for further scientific discoveries.

Improved $O(2)$ Equivariance via Embedding in Zernike Polynomial Space

24 Feb
11:00

Romain Chazotte, Kai Polsterer

Speaker Affiliation: Heidelberg Institute for Theoretical Studies

Image analysis methods are sensitive to the orientation of their inputs. This has direct implications on the performance in respective application fields, like astronomy, bio-medical imaging, as well as many technical tasks. We present a novel concept that utilizes the representation of the image data in Zernike space for near perfect $O(2)$ equivariance. For classification tasks, the proposed method outperforms other state-of-the-art methods, while being far less computationally expensive.

24 Feb
11:20

Wind Collisions and Giant Eruptions in Massive Binaries, and their Effects on the Involved Stars

Amit Kashi

Speaker Affiliation: ariel university

Massive stars eject strong winds that affect their evolution. When in a binary system, their winds collide and emit radiation across the spectrum, providing an opportunity to study the stars and the interaction between them. There are many physical effects involved in the colliding wind problem, and its complexity requires 3D numerical simulations. I will present simulations of colliding winds in massive binary systems that include a detailed treatment of wind ejection, orbital motion, clumpiness, and other effects. I will discuss the results of systematic simulations that were used to determine the general conditions that may lead to accretion onto the star with the weaker wind, and demonstrate new relationships between the mass accretion rate and the ratio of the stellar wind momentum and the Bondi–Hoyle–Lyttleton accretion rate. Additionally, I will present simulations of mass ejection during giant eruptions in LBVs, including interactions and accretion onto the companion star, and discuss the implications for the formation scenario of LBVs. I will also present new stellar evolution simulations revealing the effects of mass ejection and mass accretion at high rates on massive stars.

24 Feb
11:40

Effective temperature prediction for Gaia eclipsing binaries using ensemble of random forest models

Štefan Parimucha, Maksim Gabdeev

Speaker Affiliation: Pavol Jozef Šafárik University

Eclipsing binaries (EBs) are crucial standards for stellar physics. However, a large proportion of the approximately 2 million EBs in the Gaia Data Release 3 (DR3) catalogue lack estimates of their effective temperature (T_{eff}). This study introduces a comprehensive machine learning framework designed to predict T_{eff} for over 800,000 EBs using only Gaia DR3 photometric data. We developed and compared four distinct modelling strategies using a Random Forest (RF) regressor: a standard RF model trained on high-quality temperatures from the General Stellar Parameteriser from Photometry (GSP-Phot); an RF model trained on log-transformed temperatures; a model incorporating predicted surface gravity ($\log g$); and a structure-augmented model that leverages unsupervised clustering of the colour space. We present a “Best-of-Four” ensemble method, which dynamically selects the prediction with the lowest internal model uncertainty for each source. Our resulting catalogue achieves uncertainty of 263 K for 90% of the newly predicted temperatures. When combined with existing GSP-phot measurements, this work provides T_{eff} estimates for more than 97% of the initial Gaia EB catalogue.

Exascale Computing and FAIR Data Platforms II

Session Chair: Susanne Pfalzner

24 Feb
14:00

The filamentary ISM: Structure identification and characterisation

Daniel Seifried

Speaker Affiliation: University of Cologne

Data-driven Discovery

Session Chair: Pavlos Protopapas

Session Invited Speaker: Anna Scaife

25 Feb
09:30

Building foundational representations from large-scale surveys

Speaker Affiliation: University of Manchester

Multimodal Deep Learning for Young Stellar Object Classification and Characterization

25 Feb
10:10

Máté Madarász, Gábor Marton

Speaker Affiliation: Konkoly Observatory

Identifying young stellar objects (YSOs) has mostly relied on color-based selection methods that are survey-specific and often contaminated by reddened field stars and evolved objects. We present a comprehensive deep learning framework that uses multimodal data, including spectral energy distributions (SED), time-domain variability from ZTF, and direct imaging from AllWISE to achieve robust, homogeneous YSO classification across the entire sky. Our supervised approach combines standard CNN architectures (ResNet, EfficientNet), with our custom-built models optimized for these astronomical data representations, achieving classification accuracies higher than 90% for YSOs with contamination below 1%. Applied to over 3.3 million sources, this method produced the NEMESIS General YSO catalogue containing $\sim 250,000$ reliably classified objects. Building on this foundation, we extend our framework toward unsupervised methods to study the structure, evolutionary pathways, and variability characteristics within the YSO population. Using variational autoencoders and self-supervised learning on SEDs and light curves, we explore data-driven representations that capture continuous evolutionary stages and variability patterns beyond traditional class definitions. This supervised classification followed by an unsupervised characterization approach enables both reliable sample construction and exploration of YSO properties, variability modes, and relationships to local environmental structures.

From Orbital Space to Phase-Space Ridges in the Milky Way Disc

25 Feb
11:00

Simone Vilardi, Subhadeep Sarkar, Teresa Antoja Castelltort, Francesca Figueras Siñol, Kerstin Bunte

Speaker Affiliation: University of Groningen

Recent Gaia data releases have revealed a plethora of substructures in the phase space of the Milky Way disc. The most significant characteristics are the ridge-like structures in the (R, V_ϕ) projection and the spiral pattern in the vertical (Z, V_Z) plane. While these signatures contain valuable information about galactic dynamics, their interpretation is complicated by projection effects and the overlap of multiple dynamical populations in configuration-velocity space. In this study, we present a data-oriented methodology that uses clustering in a phase-space representation based on physical principles to identify coherent dynamical structures. Rather than clustering directly in configuration space, our approach operates in a feature space constructed from orbital actions and closely related dynamical quantities, including frequencies and angular-momentum-based invariants. This depiction captures the underlying orbital composition of the disc and allows the separation of dynamically different groups. We find that the resulting clusters identified in this high-dimensional orbital feature space exhibit planar structures when visualised in the (R, V_ϕ, V_R) projection. When viewed in this space, these planar structures project onto the (R, V_ϕ) plane along the same loci as the ridge-like features observed in Gaia data. This supports an

interpretation in which the ridges correspond to projections of more fundamental dynamical manifolds. Building on this result, we examine how the same clustered populations appear in other subspaces relevant to vertical dynamics, such as (Z, V_Z, L_Z) , and how they relate to the phase-space spiral features observed in Gaia data. The overarching aim of this work is to develop a framework for analysing coherent structures in the full six-dimensional phase space and to use their behaviour across different projections to inform their physical interpretation. In this way, the methodology provides a route to assessing the relative roles of internal disc dynamics, including spiral arms and the Galactic bar, and external perturbations such as past satellite interactions, in shaping the substructure of the Milky Way.

25 Feb
11:20

Physical analysis of galaxies via spectral reconstruction of deep imaging

Catarina Corte-Real

Speaker Affiliation: Heidelberg Institute for Theoretical Studies

Integral field spectroscopy (IFS) provides very complete information about extended astronomical sources, such as galaxies, by mapping the spectra throughout the object. Reconstructing IFS data from photometric imaging is a mathematically ill-posed problem, as images do not contain all the necessary information required to uniquely recover highly-resolved spectra. Our approach is developing a machine-learning model that reconstructs the latent, unobserved spectra that underlie a given set of galaxy images, in different wavelength bands, while taking into account the spatial information inherent in IFS observations. This approach combines (i) a generative spectral model that captures the main features of spectra with few degrees of freedom and (ii) probabilistic inference that characterizes the full set of plausible spectral reconstructions via a posterior distribution. Furthermore, while spectral recovery is a key intermediate product, it is not the final objective. The ultimate goal is to extract astrophysical properties, particularly parameters such as stellar age and metallicity of the galaxies at hand. To date, the project has used data from the open-access CALIFA (Calar Alto Legacy Integral Field Area Survey). The methods developed are designed to be generalizable and applicable to other observational or simulated datasets, which will be explored in the future. A working prototype can already reconstruct IFS data cubes, although spatial dependencies are not yet incorporated. These dependencies will be included in future iterations. At present, the prototype uses a basis-function model to parameterize galaxy spectra. Performance has been evaluated by comparing reconstructed spectra with those from open astronomical surveys, and it is expected to improve when more dependencies are incorporated.

25 Feb
11:40

Morphological Classification of Jets in Active Galactic Nuclei

Dmitrii Zagorulia

Speaker Affiliation: Moscow Institute of Physics and Technology

This work explores the application of machine learning techniques to classifying active galactic nuclei (AGN) with jets based on Very-Long-Baseline Interferometry (VLBI) observations at frequencies 1 – 90 GHz. Building upon previous work by Fanaroff and Riley, who classified relativistic jets in radio galaxies on kiloparsec scales, we extend this classification to parsec scales, closer to the central supermassive black hole (SMBH). This approach enables detailed study of jet spatial structures and can help enhance accuracy in global positioning systems. We define two morphological classes: compact point-like sources and sources exhibiting jet structure. To construct a training dataset, synthetic AGN jet images were generated using ray-tracing of the analytical Blandford–Königl jet model, followed by uv-plane sampling and image reconstruction with the CLEAN algorithm to reproduce realistic observational effects. These simulated data were used to train a convolutional neural network (CNN), which demonstrated strong

quantitative performance when evaluated on a manually labeled sample of real VLBI images. The trained model was subsequently applied to classify all ($\sim 130,000$) AGN images from the Astrogeo database. These resulting classifications were later used to investigate the correlation between the morphological class, radio flux density and total radio luminosity, as well as to investigate associations of gamma-ray sources observed by Fermi. In addition, a public online tool for searching morphologically similar AGN images is currently under development, with the presented CNN model forming its core component.

AI/NLP for Workflows and Education

Session Chair: Anna Scaife

Session Invited Speaker: Lucie Flek

25 Feb
14:00

LLMs and LLM Agents in Astrophysical Workflows: Promise, Pitfalls, and Practice

Speaker Affiliation: University of Bonn

LLMs are beginning to appear across astrophysical research, from literature analysis to workflow automation and interactive data exploration. At the same time, there is growing uncertainty about what these systems can reliably contribute to scientific discovery and where their limitations become critical. This keynote explores the emerging role of LLMs and LLM-based agents in astrophysical workflows, focusing on how these systems behave when embedded in real research pipelines rather than isolated demonstrations. Drawing on concrete experiences with agent-based systems, as well as recent work on evaluating scientific language models, the talk will highlight both promising capabilities and recurring failure modes that matter for precision astrophysics. Rather than viewing LLM agents as full replacements for existing pipelines or expertise, the talk examines how LLMs interact with existing analysis practices, where human oversight remains essential, and why design choices around autonomy and verification are central to scientific trust.

Semantic Citation Prediction

25 Feb
14:40

Pavlos Protopapas

Speaker Affiliation: Harvard University

Resolving correct citations can be a time-consuming task for research writers. Web tools exist allowing researchers to enter queries and retrieve related papers, but these tools generally depend on the existing citation graph which may bias results toward citation count or more recent publications. However it has been difficult for researchers to build citation prediction systems based on full paper texts due to lack of this data availability.

In this paper we present a novel dataset pairing scientific statements with one or more target citations. We also present a citation prediction system, based on document retrieval, that uses only the query and full document text to find relevant research papers, unbiased by paper metadata. Through a thorough examination of retrieval techniques on this task we achieve roughly 10% improvement over a baseline vector search approach, without requiring any model fine-tuning.

AstroGenesis: A Hierarchical Multi-Agent System for Automated Blazar Research and Literature-Augmented Analysis

25 Feb
15:40

Mher Khachatryan, Narek Sahakyan, Artyom Khachatryan, Vika Markosyan

Speaker Affiliation: ICRANet-Armenia

The increasing volume and complexity of astrophysical data and literature pose significant challenges for efficient and reproducible scientific analysis, particularly in high-energy astrophysics. We present AstroGenesis, a domain-aware multi-agent system designed to support data-driven research through coordinated literature retrieval, observational data analysis, and AI-assisted interpretation. The system follows a plan-and-solve paradigm, in which a supervisory agent decomposes scientific queries into structured subtasks and orchestrates specialized agents to execute them in a transparent and traceable manner. AstroGenesis integrates multiple functional agents, including literature retrieval via a domain-tuned retrieval-augmented generation (RAG) subsystem, observational data access from public archives, and modeling agents that interface with machine-learning tools for spectral and temporal analysis. In the context of blazar studies, agents retrieve multi-wavelength spectral energy distributions and light curves from public archives like Markarian Multiwavelength Data Center (MMDC), apply theoretical modeling on observational data, perform AI-assisted interpretation, and contextualize results within existing high-energy literature. Demonstrated on blazar datasets, AstroGenesis illustrates how plan-and-solve multi-agent architectures can bridge literature knowledge, observational data, and modeling workflows. This work highlights the role of Astroinformatics in building scalable, reproducible AI systems that augment human reasoning and accelerate discovery in high-energy astrophysics.

Teaching Astroinformatics in Classrooms: A critical path

25 Feb
16:00

André Bresges, Kathleen Falconer

Speaker Affiliation: University of Cologne

The rapid growth of **astroinformatics** depends critically on two societal pillars: (1) a steady pipeline of young researchers who will advance the field, and (2) a broadly informed public that understands and therefore supports funding for large-scale astronomical and astroinformatic projects. In most countries, the **core school curricula** (the mandatory subjects for school education) are already saturated, and astro-informatics has not yet been introduced as an optional or elective topic because it is so new.

Our work addresses this gap by proposing a systematic integration of astroinformatics **learning outcomes, concepts, and methods** into the existing **MINT** (Mathematics, Informatics, Natural sciences, and Technology) framework that is already taught in schools. Rather than adding a separate course, we aim to identify **natural cross-disciplinary entry points** and plan to develop a set of **modular teaching units and teacher trainings**, complete with data sets from sky surveys, open-source analysis tools (hopefully the SUE at a later stage), and assessment rubrics that can be adopted by teachers without extensive background in astronomy or informatics.

Which concrete, cross-disciplinary entry points can be used to bring astroinformatics into the classroom? After attending Astroinformatics 2026 and talking to numerous researchers, we have assembled a set of initial ideas.

In this round-table presentation we will share these ideas, discuss the pedagogical rationale behind each, and invite the community's critical feedback and validation. Your insights will help refine the modules, ensure their scientific robustness, and accelerate the adoption of astroinformatics across the MINT curriculum.

Survey and Time-Domain Data I

Session Chair: Matthew Graham

Session Invited Speaker: Guillermo Cabrera-Vives

26 Feb
09:30

Lessons from Time-Domain Astronomy: Why One Modality Is No Longer Enough

Speaker Affiliation: University of Concepción

Time-domain astronomy has long relied on synoptic surveys such as ZTF and is now scaling further with facilities like the Vera C. Rubin Observatory. These surveys produce continuous alert streams that must be processed and interpreted in real time by systems such as astronomical brokers. Historically, many of the most significant discoveries have resulted from human experts integrating multiple sources of information, including light curves, images, contextual metadata, and external knowledge. In contrast, most machine-learning models have traditionally relied on a single modality, limiting their ability to capture the full complexity of astrophysical phenomena. In this talk, I discuss the need for multimodal models that can jointly leverage heterogeneous data sources, enabling more robust and scalable discovery in the Rubin era and beyond.

Automating real-time transient workflows to facilitate discovery

26 Feb
10:10

Nabeel Rehemtulla, Adam A. Miller, Wynn Jacobson-Galán

Speaker Affiliation: Northwestern University

Significant filtering of data streams from current and future time domain surveys is required due to the breadth of astrophysical information they contain. Moreover, the increasingly unwieldy volumes of these data streams necessitate that this filtering happen automatically. In this talk, I introduce a suite of “bots” I have deployed into the real-time production environments of ZTF and the La Silla Schmidt Southern Survey (LS4). Real/bogus classification is essential to imagedifferencing time-domain surveys, but current state-of-the-art classifiers require substantial labeled data sets for training. I introduce RBbot, the real/bogus classifier for LS4 trained using unsupervised domain adaptation techniques, enabling RBbot to learn from the large quantities of unlabeled data which are traditionally discarded. Alongside RBbot, I present benchmarking of vision model architectures and pre-training regimens which help RBbot confer additional benefits in computational efficiency and classification performance. After filtering bogus alerts, transients of interest must be selected and followed-up spectroscopically, processes traditionally done entirely manually. I also introduce BTSbot, a mutli-modal CNN now capable of fully autonomously discovering new, nearby ($d \lesssim 60$ Mpc) supernovae and requesting target-of-opportunity follow-up to SEDM, MMT, Gemini, LCO facilities, SOAR, WINTER, and Swift. This one-of-a-kind workflow has reduced the typical latency between a transient’s first detections and its first follow-up by more than an order of magnitude. In the case of SN 2025qtt, Swift/UVOT began observing just 6 minutes after its automated discovery by BTSbot. Advanced infrastructure like these bots and their tight integration into ZTF, LS4 and follow-up facilities are mandatory to maximize the output of our time-domain surveys.

Measuring and Mitigating Galaxy Labeling Bias from Low to High Redshift: A Deep Learning Approach

26 Feb
11:00

Joaquín Gallardo-Villagrán, Guillermo Cabrera-Vives, Rodrigo Herrera-Camus, Monserrat Martínez-Marín

Speaker Affiliation: University of Concepción

Context: Crowdsourced galaxy classification projects, such as Galaxy Zoo, are prone to observational bias due to varying image resolutions. This observational bias leads to voters misclassifying spiral galaxies as elliptical galaxies when observing at a low image resolution. Algorithms trained on these biased datasets inherit the same bias. However, recent work has shown that deep learning models can effectively correct such biases.

Aims: Our goal is to quantify the observational bias in the Galaxy Zoo: CANDELS dataset using established methodology and then reduce this bias by adjusting public classifications. By validating the corrected results with high-resolution data from the James Webb Space Telescope (JWST), we aim to extend the applicability of deep debiasing methods — originally developed for Sloan Digital Sky Survey (SDSS) data — to Hubble Space Telescope (HST) observations. This will demonstrate that the method can be safely applied to galaxies at $z > 1$.

Methods: We use a bias quantification scheme that groups galaxies based on intrinsic properties (e.g., absolute magnitude, redshift, and galaxy size). Within each bin, the fractions of spiral and elliptical classifications should ideally remain constant with respect to observational parameters (e.g., image resolution). This estimator quantifies deviations from the ideal scenario. We then apply a convolutional neural network (CNN) model with a customized, bias-aware loss function to reduce the bias in the classifications to recover the ground truth labels.

Results: We found that 26.5% of galaxies labeled as elliptical in CANDELS were reclassified as spirals by our model. Validation against JWST ground truth revealed a 56.5% accuracy in identifying hidden spiral structures, proving the model’s capacity to detect morphological biases. Although the overall bias reduction was moderate ($\sim 3\%$), the qualitative success on individual targets highlights the potential of this method for cleaning catalogs in future large-scale surveys.

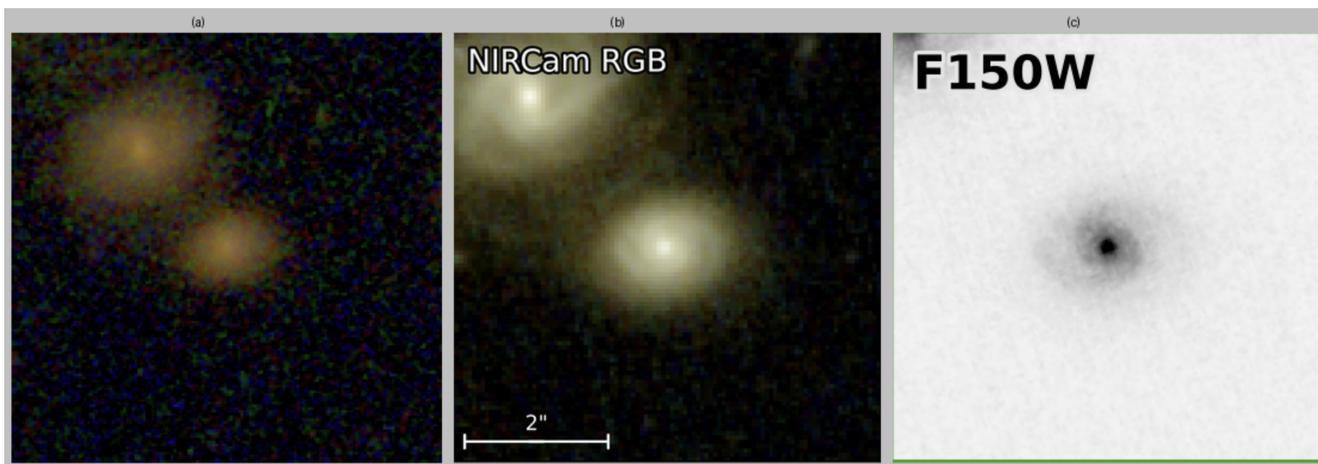


Figure 1. Example of how labeling bias affects visual galaxy classification. Left panel: A galaxy image as shown to annotators participating in Galaxy Zoo: CANDELS (HST RGB image). Middle panel: Same galaxy as the left panel but now observed in NIRCcam RGB. Right panel: Same galaxy as the other two panels but now observed in NIRCcam filter F150W. 73.45% of Galaxy Zoo:CANDELS voters classified this galaxy as a smooth galaxy, but spiral arms are visible on higher resolution JWST data. The prominence of spiral arms also varies as a function of wavelength.

Finding Needles in the TESS Haystack: Automated detection of roAp Stars from TESS data

26 Feb
11:20

Inês Rolo, Daniel L. Holdsworth, Margarida S. Cunha, Victoria Antoci, Donald W. Kurtz, Rahul Jayaraman, Ângela R. G. Santos

Speaker Affiliation: University of Porto

Rapidly oscillating chemically peculiar A-type (roAp) stars offer valuable insights into the internal physical processes of all stars, but their study has been greatly challenged by their rarity. We present a new algorithm designed to explore TESS data products, specifically 200-sec cadence Full Frame Images as well as 20-sec cadence light curves, to look for stellar oscillations. This tool is capable of identifying and classifying various classes of stellar pulsators, roAp stars, δ -Scuti stars, and other classical pulsators, as well as non-pulsating Ap (noAp) stars. We apply our algorithm to two independent datasets: i) spectroscopically classified Ap stars observed by TESS in the 200-s Full Frame Images (FFIs) and ii) all TESS 20-s cadence light curves available for known or candidate roAp stars. We identify four new roAp stars: TIC 312111544, TIC 252881095, TIC 46054683, and 49 Cam (TIC 393276640). We find evidence in TESS data that TIC 252881095 may be part of a binary system. If the tentative ~ 30 -d orbital signal is confirmed, TIC 252881095 could be the shortest-period roAp binary currently known. Furthermore, the detection of high-frequency pulsations in 49 Cam is particularly relevant, as this well-known candidate roAp star is here confirmed to be roAp based on TESS 20-s cadence data. Overall, we show that automated, cadence-aware searches of TESS data can efficiently identify rare high-frequency pulsators, expand the known roAp population, and flag high-value targets for follow-up. This is a step toward a large, homogeneous roAp sample enabling population-level tests that can significantly advance stellar modelling.

Improving Time-Series Analysis in Astronomy: Understanding Stars, the Milky Way, and Beyond

26 Feb
11:40

Aarya Patil

Speaker Affiliation: Max Planck Institute for Astronomy

Asteroseismic time-series data encode the signatures of stellar oscillation modes, whose detection and characterisation through frequency analysis provide powerful diagnostics of stellar interior physics. Such analyses are typically performed using the Lomb–Scargle (LS) periodogram, an estimator of the power spectrum for unevenly sampled time series. However, the LS periodogram suffers from statistical limitations, including inconsistency (noise) and bias arising from spectral leakage. To address these challenges, I have developed the multitaper Non-Uniform Fast Fourier Transform (mtNUFFT) method [1], which mitigates both inconsistency and bias, enabling efficient and precise frequency estimation of stellar oscillations. This method is implemented in the open-source Python package tapify, designed for general time-domain astronomy applications. I have also extended this work to include a multitaper F-test, which distinguishes between strictly periodic and quasi-periodic signals. By making this distinction, the test helps identify the excitation mechanisms underlying stellar oscillation modes across the Hertzsprung–Russell diagram, spanning a wide range of stellar masses and evolutionary stages [2], and in some cases can also differentiate intrinsic stellar variability from signals produced by exoplanet transits [3]. These advances open new avenues for studying stellar structure, evolution, and Galactic archaeology with upcoming large-scale time-domain datasets such as the Vera Rubin Observatory’s LSST survey. In this talk, I will present the time-series analysis methods I have developed and discuss their broader potential applications across astronomy and the physical sciences. I will also discuss my plans to incorporate these methods into LSST pipelines as an LSST Discovery Alliance Catalyst Fellow at the Max Planck Institute for Astronomy.

References

- [1] A. A. Patil, G. Eadie, J. Speagle, and D. Thomson. *The Astronomical Journal*, 168(5):193, Oct 2024.
 - [2] A. A. Patil, C. Aerts, N. Y. N. Wang, J. Van Beeck, and M. G. Pedersen. Submitted to *Astronomy & Astrophysics*, art. arXiv:2512.10019, Dec. 2025.
 - [3] A. A. Patil, G. M. Eadie, J. S. Speagle, and D. J. Thomson. *The Astronomical Journal*, 170(1):7, July 2025.
-

Survey and Time-Domain Data II

Session Chair: Nikos Gianniotis

26 Feb
14:00

Session Invited Speaker: Michelle Lochner

Scientific discovery with foundation models

Speaker Affiliation: University of the Western Cape

The next generation of telescopes such as the SKA and the Vera C. Rubin Observatory will produce enormous data sets, far too large for traditional analysis techniques. Machine learning has proven invaluable in handling massive data volumes and automating many tasks traditionally done by human scientists. In this talk, I will explore the use of machine learning for automating the discovery and follow-up of interesting astronomical phenomena, both in the image and time domains. I will discuss how the human-machine interface plays a critical role in maximising scientific discovery with automated tools, demonstrating applications of the active anomaly detection framework, *Astronomy*, on a variety of datasets. Finally, I will investigate the role foundation models play in enabling scientific discovery in massive surveys.

26 Feb
14:40

Detecting anomalous radio spectrograms with unsupervised and active learning

Hanwool Koo, Michelle Lochner

Speaker Affiliation: University of the Western Cape

Developing an automated algorithm for anomaly detection is increasingly essential for uncovering previously unknown phenomena in astrophysics and cosmology from large volumes of radio spectrograms. To achieve this, we explore machine learning techniques for detecting anomalies in the time-frequency dynamic spectra of the radio data. We evaluate our algorithms on simulated SPARKESX: Single-dish PARKES data sets for finding the unexpected, enabling us to apply them to actual, unlabeled MeerKAT data, where the MeerKAT radio telescope is a precursor to the mid-frequency Square Kilometre Array (SKA-Mid) and a promising source of large sets of radio data. We begin with essential preparation for anomaly detection, including feature extraction with a supervised Convolutional Neural Network (CNN), dimensionality reduction via Principal Component Analysis (PCA), and visualisation using Uniform Manifold Approximation and Projection (UMAP). Based on the prerequisites, we utilise unsupervised learning techniques, including Isolation Forest and Local Outlier Factor, for anomaly detection. We discuss their performance and limitations, then introduce novel approaches for anomaly detection: ensemble learning of unsupervised learning methods and active learning using *Astronomy* and *Protege*. We expect the new approaches to be more efficient and reliable for accurately detecting anomalous astrophysical signals than previous methods.

Matthew Graham

Speaker Affiliation: California Institute of Technology

AGN variability is a symphony of accretion disk instabilities, ionizing continua, relativistic jets, stellar activity, thermal variations, magnetic fields, and dust but poorly captured by the simple humming of current analysis techniques. We propose a new more orchestrated approach combining many voices: a shared melody, individual timbres and acoustics, solos, and key changes. Candidate arrangements are composed with quick learned suggestions, scored by how well they play with the data while keeping the arrangement as simple as possible and then explained in plain terms. This produces a readable reusable pattern for AGN that also performs well with the next generation time domain data sets.

Anomaly Detection in DESI Quasar Spectra using Machine-Learning Dimensionality-Reduction Techniques

26 Feb
16:00

Paul Nathan

Speaker Affiliation: UCL

The Dark Energy Spectroscopic Instrument Survey (DESI) has already captured more than 50 million astronomical spectra from stars, galaxies and quasars. Within this extraordinarily rich and large dataset, there will be spectra which are anomalous: because they represent rare phenomena, or because they arise from “unknown unknowns” — entirely novel objects, events or processes. Such anomalies clearly offer incredible discovery potential but finding them within such a large dataset is a challenge. In this work we use a variety of unsupervised dimensionality reduction techniques, demonstrating how we can detect anomalies both “on” and “off” the corresponding lower-dimensional manifold. We focus on a dataset of $z \sim 2$ quasars from the DESI DR1 and show how lightweight PCA, Autoencoder and Variational Autoencoder models can be used to highlight different types of object (in particular different species of broad absorption line quasar) and how these methods could be used to complement existing quasar identification methods. Finally, we discuss a couple of anomalous spectra in detail illustrating the interesting features that can be identified through this kind of anomaly detection.

Survey and Time-Domain Data III

Session Chair: Michelle Lochner

27 Feb
09:30

Modeling Distribution and Photo- z Uncertainty of Quasars: An Analysis Based on the Kilo-Degree Survey

Kacper Drabicki, **Szymon J. Nakoneczny**, Maciej Bilicki

Speaker Affiliation: Polish Academy of Sciences

Context. Photometric redshift (photo- z) plays a pivotal role in the era of precision cosmology. Equally critical is the reliable quantification of the associated uncertainties and the identification of degeneracies inherent in the data. Despite extensive development, no single methodology has emerged as a definitive standard, and existing approaches have not yet been systematically and rigorously compared.

Aims. The primary goal of this work is to identify the most effective method for estimating uncertainties in quasar (QSO) photo- z . We assess the models capability of reconstructing the original redshift distribution.

Methods. We use photometric filters from Kilo Degree Survey (KiDS) Data release 5 (DR5) and quasar redshift values from Dark Energy Spectroscopic Instrument (DESI). We compare standard neural networks, Mixture Density Networks (MDNs) employing varying numbers of components, and Bayesian Neural Networks (BNNs) combined with Gaussian Mixture Model (GMM) outputs. To ensure a robust and informative comparison, we construct four dedicated test sets that span all combinations of (i) sources fainter than the faintest objects present in the training data and (ii) missing observations in at least one photometric band.

Results. We find that MDN uncertainty estimation methods based on at least two GMM components are essential for accurately reproducing the underlying redshift distribution. An MDN network with three GMM components yields the best overall performance within the MDN family. BNNs improve performance on test sets extending beyond the training magnitude range, achieving an NLL improvement of 0.193, but at the expense of degraded performance on the random test set, with an NLL increase of 0.114. Clustering analysis reveals five distinct clusters characterized by degeneracies between two different redshift solutions within each cluster.

Conclusions. The use of Bayesian neural networks is advantageous primarily when inference data are not covered in magnitude by the training set, albeit at the cost of reduced accuracy for brighter sources. Without an effective uncertainty estimation framework, a faithful reconstruction of the true redshift distribution is not possible. Clustering of photo- z predictions emerges as a powerful and efficient tool for identifying degeneracies in photometric redshift inference.

27 Feb
09:50

The Use of Minimalistic Convolutional Architectures for Detection of Astronomical Radio Signals

Kazantsev, Andrei, Ramesh Karuppasamy, Michael Kramer

Speaker Affiliation: Max Planck Institute for Radio Astronomy

In this work, we evaluated several convolutional neural network architectures with a small number of parameters for the task of detecting astronomical radio signals in DM-time domain. The main advantage of such lightweight models is not only a faster search for optimal hyperparameters, shorter training time, and lower inference latency, but also their potentially simpler implementation on field-programmable gate arrays (FPGAs), especially in the context of scaling the pipeline to higher data rates and real-time processing.

All minimalist models were trained exclusively on synthetic data generated to mimic the expected signal morphology after dispersion. In this way, we tested the hypothesis that limiting the number of trainable parameters can also limit the ability of the models to exploit subtle differences between synthetic and real observations. As a result, the models may be forced to focus on the most stable,

physically meaningful features of dispersed pulses, which could reduce the synthetic-to-real domain gap without using additional domain adaptation techniques.

The final applicability check was performed on real observational data of pulsar B0355+54, the Crab pulsar B0531+21, and the repeating FRB20240114A, obtained with the Effelsberg radio telescope in the 1200–1600 MHz frequency range. We used the same trained models without re-training on real data to directly test generalization under realistic noise and radio-frequency interference conditions. Among the tested architectures, ResNet18 with about 250 thousand trainable parameters achieved the best overall performance, reaching $F1 = 92.74$, and showed the most consistent behaviour across all considered sources.

Uncovering the Giants in the Sky - Semantic Segmentation of Galaxy Clusters in Astrophysical Surveys

27 Feb
10:10

Jakob Dietl, Thomas Reiprich, Florian Pacaud, Alexander Rüttgers, Zorah Löhner, Angie Veronica
Speaker Affiliation: University of Bonn

We present a deep-learning approach for galaxy cluster detection in X-ray survey data using the eROSITA all-sky survey. A U-Net-based convolutional neural network is trained to perform semantic segmentation of extended galaxy Cluster emission. Compared to standard segmentation tasks, this problem poses additional challenges, including Poisson-dominated noise, incomplete and uncertain ground truth labels, diffuse and poorly defined object boundaries, and the need to distinguish clusters from morphologically similar sources such as active galactic nuclei. The model is trained on both real eROSITA observations and mock images to assess robustness against false labels, achieving dice scores of 95% on simulated data and 92% on real survey data. This work forms a basis for leveraging additional multi-wavelength survey data to further improve galaxy cluster detection and classification.

27 Feb
10:30

Claudia Comito

Speaker Affiliation: Jülich Supercomputing Centre

Posters

Functional Magnetic Resonance Imaging of Human Brain Functions during simulated car driving: Possible applications for a Shared Universe Engine

André Bresges

Speaker Affiliation: University of Cologne



Functional Magnetic Resonance Imaging of Human Brain Functions during simulated car driving: Possible applications for a Shared Universe Engine

Prof. Dr. André Bresges
University of Cologne
Institute of Physics Education
Faculty of Math and Natural Sciences

The argument for a Cognitive Theory of Astroinformatics

Predictive coding theory proposes the brain isn't a passive receiver but an active prediction machine, constantly generating internal models to anticipate sensory input, minimizing surprise by only processing the "prediction errors" (mismatches) that occur. This leads to efficient perception, learning, and action by updating its world model to better match reality. It's a Bayesian brain hypothesis where top-down predictions (expectations from past experience) meet bottom-up sensory data, with errors signaling the need for adjustment of the model. Predictive Coding explains consciousness not as a static model, but dynamically as

„the brain's best guess about the world“

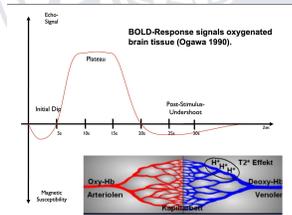
We argue that predictive coding might be a good theory to build a Shared Universe Engine upon because of its: **Efficiency:** Reduces redundant processing, saving cognitive resources. **Precision:** The Brain aims to minimize prediction errors. **Speed:** Unifies perception, cognition, and action (sensor direction and tuning) under a single principle.

Experiment with human brains in fMRI

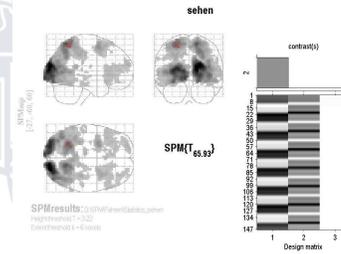
Sensory Stimulus



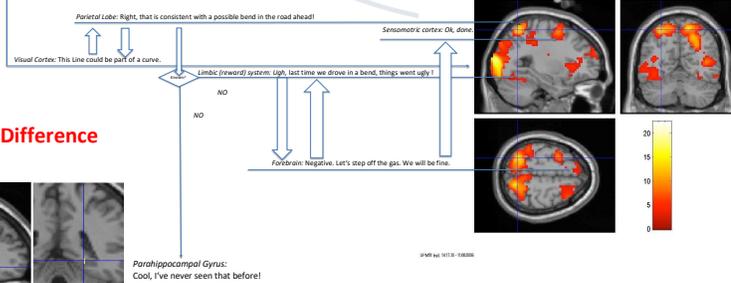
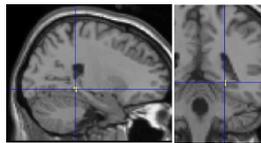
Voxel-wise MRI Signal



Functional Brain Image



Expert/Novice Difference (ANCOVA)



Result and conclusion

- It is a reasonable assumption that the Human Brain applies Predictive Coding to generate and test models from a large body of sensory data
- Predictive coding is multi-levelled: Example: visual cortex predicts that the existence of horizontal line explains part of the data from the retina. Parietal cortex confirms that a horizontal line is consistent with the „horizont“ model and predicts an upright orientation of the body. Higher order functions predict that it is possible to move the body forward from here. This is ultimately tested (and might fail utterly).
- Caveat: Mistakes and Errors occurring in human brains are likely also to occur in other predictive coding machines.

Sources

- Mader, M.; Bresges, A.; Topal, F.; Busse, A.; Forsting, M.; Gizewski, E.: Simulated car driving in fMRI—Cerebral activation patterns driving an unfamiliar and a familiar route. Neuroscience Letters Volume 454, Issue 3 (2009) <http://www.sciencedirect.com/science/article/pii/S0304394009011616>
- Friston K, Kiebel S. Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci. 2009 May 12;364(1521):1211-21. doi: 10.1098/rstb.2008.0300. PMID: 19528002; PMCID: PMC2666703.



Prof. Dr. André Bresges
Andre.bresges@uni-koeln.de



Representation Learning for Gaia XP DR3

Bernd Doser, Kai Polsterer, Sebastian Trujillo-Gomez

Speaker Affiliation: Heidelberg Institute for Theoretical Studies

We present a novel representation learning framework for the Gaia XP DR3 dataset [1] that leverages two advanced data exploration tools: Spherinator and HiPSter [2]. Spherinator provides a method for learning compact representations of high-dimensional data, including images, point clouds, data cubes, time series, and spectra. Our training process uses variational autoencoders with hyper-spherical latent spaces to efficiently and robustly extract physically meaningful parameterizations of data properties. Our approach explicitly incorporates uncertainties from experiments into representation learning, which produces more robust and physically consistent latent representations. HiPSter generates and serves HiPS-based representations of learned features, enabling the scalable visualization and exploration of the latent space using Aladin-Lite. Spherinator and HiPSter are core components of a machine-learning workflow that covers all stages, from data collection and preprocessing to training, prediction, and final deployment [3]. We demonstrate the scientific potential of our method using the largest spectral dataset available from Gaia XP DR3 and showcase the effectiveness of cross-disciplinary tools developed under the EU SPACE initiative to enhance data-driven astronomy.

References

- [1] De Angeli, F., Weiler, M., Montegriffo, P., et al. “Gaia Data Release 3 - Processing and validation of BP/RP low-resolution spectral data”. In: *A&A* 674 (2023), A2. doi:10.1051/0004-6361/202243680.
 - [2] Kai L. Polsterer, Bernd Doser, Andreas Fehlner, and Sebastian Trujillo-Gomez. Spherinator and HiPSter: Representation Learning for Unbiased Knowledge Discovery from Simulations. 2024. arXiv: 2406.03810 [astro-ph.IM].
 - [3] Bernd Doser, Kai L. Polsterer, Andreas Fehlner, and Sebastian Trujillo-Gomez. Machine Learning Workflow for Morphological Classification of Galaxies. 2025. arXiv: 2505.04676 [astro-ph.IM].
-

Using the Perception of Physics among Adolescent Pupils as an Example of Doing Design-Based Research

Kathleen Falconer, André Bresges
Speaker Affiliation: University of Cologne

Using the Perception of Physics among Adolescent Pupils as an Example of Doing Design-Based Research

Falconer, K.A. & Bresges, A. University of Cologne, Kathleen.Falconer@uni-koeln.de



Discover

DBR involves designing and implementing interventions in authentic settings to address real-world problems.

Purpose: Identify the problem or need that the research aims to address.

Activities: Conduct literature reviews, gather data from stakeholders, and analyze the context to understand the issues and opportunities.

Outcome: A clear understanding of the problem and the factors that contribute to it.

The perception of physics, including astrophysics and informatics, among adolescent pupils is often shaped by two opposing ideas:
(i) a lingering fear of the subject, and
(ii) a strong, personally-driven curiosity sparked by the natural environment, social-media and pop-culture.
While many students report a basic interest in physics, unfortunately, conventional classroom instruction frequently dampens this enthusiasm.

Define

Purpose: Develop a theoretical framework and design principles to guide the intervention.

Activities: Formulate research questions, hypotheses, and design specifications based on the discovered needs and theoretical foundations.

Outcome: A well-defined intervention plan that includes goals, objectives, and strategies for implementation.

To learn if the adolescent pupils from Germany have similar underlying experiences and perceptions, we conducted semi-structured interviews with $n = 24$ secondary-school students (ages 13-17) from three German Gymnasiums. The guided interviews focused on using the MAXI Framework – Mapping Attitudes, eXperiences, and Interests:
M: Motivation (self-reported interest, fear, self-efficacy)
A: Affective cues (anxiety, excitement)
X: eXternal influences (social media, pop culture)
I: Institutional factors (curriculum, teaching style)

Develop

Purpose: Create and refine the intervention through iterative cycles of design, implementation, and evaluation.

Activities: Develop prototypes or pilot programs, implement them in real-world settings, and collect data on their effectiveness.

Outcome: A refined and tested intervention that addresses the identified problem.

The results indicated many of the research findings from other countries and areas are substantiated by our research. The majority of pupils (75%) had a fear of physics and astrophysics but showed a spontaneous interest in natural phenomena (92%). The influence of social media was very strong as 83% reported that YouTube channels, TikTok clips, and science fiction films sparked their interest in physics. However, the pupils (67%) saw no connection between their personal interests (e.g., micron dust, quantum memes) and the school curriculum. Many of the pupils (58%) wanted more personalized, project-based learning opportunities that appealed to their real-world curiosity.

Deliver

Purpose: Implement the intervention on a larger scale and evaluate its impact.

Activities: Roll out the intervention to a broader audience, monitor its implementation, and gather data on outcomes and stakeholder feedback.

Outcome: A fully implemented intervention with evidence of its effectiveness and potential for scalability.

Suggestions for improvement:
We should be better prepared
We need to link different models
We ourselves need to become more proactive
What is the motivation?
We need to clarify this better
Improve the team structure/
Clear role distribution
A safe environment where incorrect ideas are explicitly permitted.
(Allow "I don't know" less often)
Good room layout/
Create visibility for everyone
Find a balance between allowing questions and staying on topic

Preparation/What do we need to do?:
Know and have readily available everyday references
Understand gas in its structure
Understand the Earth's magnetic field and magnets in general
Be prepared for context-related questions
Clearly clarify roles within the team
Understand the Earth's magnetic field and magnets in general
Consider what should be observed in the room
Good room conditions, darkened room
Make sure everyone can see the experiment

Feedback:
Subject-specific Experiment
The experiment was well prepared in terms of content
Overwhelmed by the topic/better differentiation for individual grade levels
Differentiation
Good relevance to everyday life/analogies
We should be better engaged
Good models and visualizations
Interest
Praise
Students are interested in the experiment
Role distribution
Ensure everyone can see the experiment clearly

The results have implications for improving science instruction at school and integration of cutting-edge astrophysics research into both school education and pre-service teacher training at the university level.

Disrupt and Discussion

Purpose: The iterative process of implementing, testing, and refining the designed interventions in real-world settings.

Activities: cyclical process that includes gathering data, analyzing the results, and making necessary adjustments to enhance the effectiveness of the intervention.

Outcome: Identify and remove obstacles

Identify and remove obstacles for large scale adoption in schools.

Results of the interviews for Empathy:
Fear of physics
Fear + math
Often a basic interest in physics, but lessons too often diminish this interest
Relevance to everyday life often desired
Social media and pop culture often lead to a high level of personal interest in physics that often fails to translate into classroom instruction

DynaVerse and other Participants

As collaborators in the DYNIVERSE - Our Dynamic Universe excellence cluster, we aim to leverage Design-Based Research (DBR) within the framework of the DYNIVERSE to educate pupils, students and teachers about the combining of astronomical observations, instrument development, theory and simulations, laboratory astrophysics, and astroinformatics.

Through DBR, the project will develop and refine educational materials and methodologies that incorporate the latest scientific findings and data analysis techniques. These materials will be designed to be accessible and engaging for both school students and pre-service teachers, fostering a deeper understanding of astrophysics and data science. The iterative nature of DBR will ensure that the educational interventions are continuously improved based on feedback and outcomes, ultimately enhancing the integration of research into education. We are looking forward to bridge the gap between advanced research and educational practices.

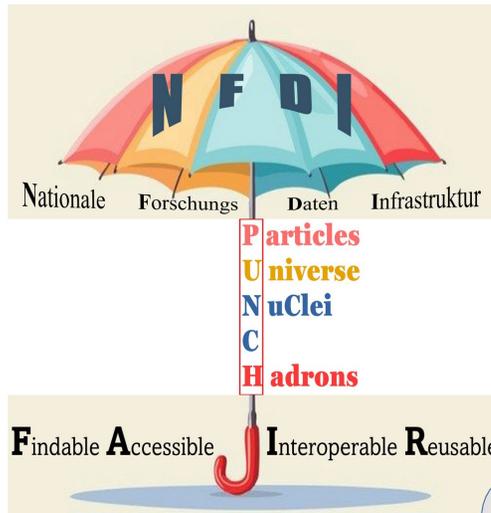


Compute4PUNCH: A platform for the analysis of Astronomical Data

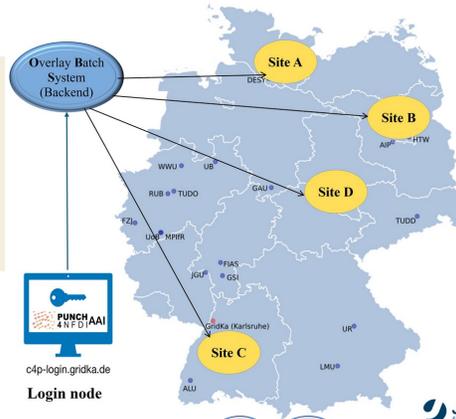
Prateek Gupta

Speaker Affiliation: Thuringian State Observatory

Compute4PUNCH: Platform for the analysis of Astronomical Data



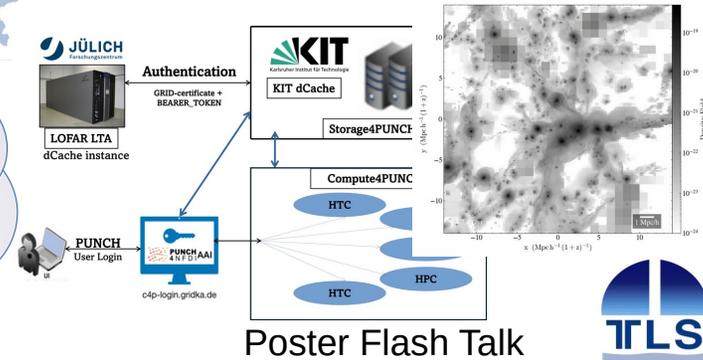
Astroinformatics 2026



Have a craving for resources to fulfill your research appetizer!!!

Science Data Platform:

- Easy access to scientific data within PUNCH Community
- Federated Infrastructure integrate: storage and compute resources
- Single entry point accessibility



Astronomical Bycatch: Extracting exoplanet transits from non-target stars in astronomical archives.

Nika Gorchakova, Creaner Oisin, Valentino Farrell, David Lawley, Cormac McGleenan
 Speaker Affiliation: Dublin City University

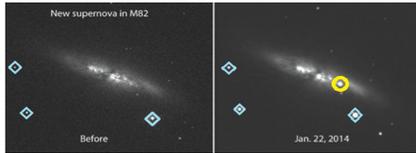
ASTRONOMICAL BYCATCH

Extracting Exoplanet Transits from Non-Target Stars in Astronomical Archives

Nika Gorchakova, Cormac McGleenan, Valentino Farrell, David Lawley, Oisín Creaner
 Dublin City University



Every night, professional and amateur observatories produce millions of astronomical images. While targeting galaxies, supernovae, or nebulae, these images also capture foreground/background stars – We call these stars “Astronomical Bycatch.”



Among this bycatch, exoplanet transits may already have been recorded but remain undiscovered.

Because:

- Most stars host planets
- Transit probability per observation is low
- Volume of archival observations is enormous

→ Statistically near-certain that many transits exist in archival data.

Technical Framework

PHASE 1 – Demonstration

- Use known transiting exoplanets
- Reprocess archival images
- Extract light curves
- Validate system against legacy tools

PHASE 2 – Optimisation

- Machine Learning acceleration
- Structured ingestion to time-series database (kdb+)
- Performance metrics (accuracy, CPU, memory, runtime)
- HPC-ready pipeline design

PHASE 3 – Exploitation

- Apply system at scale
- Build sparse photometric time-series database
- Discover new transit candidates
- Enable follow-up confirmation

Image Extraction Tools (Baseline)

- EXOTIC (NASA)
- HOPS
- AstrolmageJ

These provide:

- Baseline extraction
- Performance comparison metrics

Data Architecture

FITS Image → Light Curve → Time-Series Database

Enables:

- High-performance columnar time-series database
- Designed for rare-event detection at scale

Enables:

- Efficient ingestion
- Sparse data handling
- Fast querying
- Parallelisation readiness

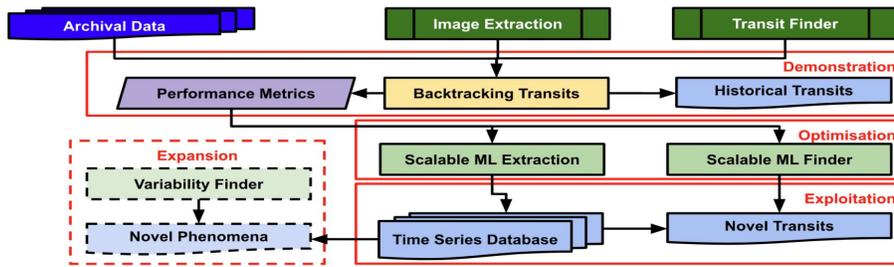
Data Sources

Archival repositories:

- ESO Archive
- MicroObservatory
- MAST
- Citizen science collaborations
 - Exoplanet Watch
 - ExoClock

Challenges:

- Petabyte-scale data
- Heterogeneous instruments
- Sparse, irregular sampling
- Rare-event detection
- Extreme class imbalance
- Need for scalable ingestion & queries



Current Objectives

- Feasibility study complete
- Prototype metadata schema
- Data ingestion API
- kdb+ database populated with transit data
- Baseline performance metrics
- Comparative analysis vs legacy methods

Impact

<h5>Scientific</h5> <ul style="list-style-type: none"> • Expand exoplanet census • Mitigate geometric detection bias • Increase probability of Earth-analogue discovery 	<h5>Technical</h5> <ul style="list-style-type: none"> • Develop transferable capabilities: • Rare event detection in sparse time-series • ML-based transient identification • Scalable ingestion of heterogeneous data 	<h5>Cross-Domain Applications</h5> <ul style="list-style-type: none"> • Earth observation • Fraud detection • Transient astronomy • Financial anomaly detection
--	--	---



This work is funded by Science Foundation Ireland through ML-Labs: the Research Ireland Centre for Research Training in Machine Learning (Grant No. 18/CRT/6183) Call 2023. This work is supported by Enterprise Ireland, Innovation Partnership Programme Feasibility Study IP20252215Y

Following Tracer Particles to Identify Filaments in Star-Forming Cores with HDBSCAN

Nuray Ortaköse

Speaker Affiliation: University of Cologne



UNIVERSITÄT
ZU KÖLN

Aleyna Nuray Ortaköse, Michael Weis, Stefanie Walch-Gassner



Following Tracer Particles to Identify Filaments in Star-Forming Cores with HDBSCAN

Introduction

Motivation

Filaments appear on all scales of the cold, dense ISM. Observations show that star-forming cores are interlaced in complex networks of filaments. These structures are thought to funnel material onto forming cores through gravitationally driven accretion, yet the influence of filamentary dynamics on star formation remains uncertain. Observations struggle with line-of-sight confusion: structures that appear continuous in PPV space may not be connected in PPP [1]. Colliding-flow simulations provide full phase-space information and show that filaments can act as accretion channels. While recent work has advanced filament identification and kinematic analysis (e.g. [2]), robust methods for identifying velocity-coherent filaments in simulations are still limited.

Goals

- Use tracer particles in AMR simulations to follow the gas velocity field during core formation.
- Identify **velocity-coherent filamentary structures** using HDBSCAN.
- Extract the **spine** of the main accreting filament.
- **Aim to measure the mass flux** along the spine toward the forming core. (*ongoing*)
- **Plan to track the time evolution** of filaments using merger trees to reconstruct accretion histories, filament mergers, and splits. (*future work*)

Why HDBSCAN?

- HDBSCAN [3] is well-suited for identifying velocity-coherent filaments because it:
- Handles data with varying densities and shapes \Rightarrow ideal for variously shaped filaments in a turbulent medium.
 - Finds clusters that persist across different density levels, keeping only the most stable ones \Rightarrow reduces detection of transient structures.
 - Identifies noise instead of forcing every point into a cluster \Rightarrow not every tracer is part of a coherent filament.
 - Requires no predefined number of clusters \Rightarrow number and topology of filaments is unknown.

Simulation: setup

Colliding flow simulation by [4]:

- Code: 3D AMR FLASH MHD simulation
- Domain size: 128 pc x 32 pc x 32 pc.
- Periodic boundary conditions in y- and z-direction, inflow in x.
- Colliding Warm Neutral Medium streams form dense sheets through shocks and thermal instability.
- Passive tracer particles initialized on a homogenous grid.
- Maximum spatial resolution: 7.8×10^{-3} pc
- Simulation time: 20 Myr

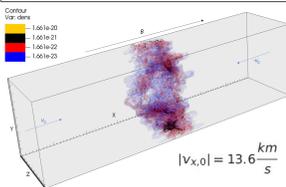


Fig. 1: sketch of the simulation setup [4]. The density is displayed in units of g cm^{-3} .

Summary/Outlook

- Clustering algorithms applied to tracer particles can reliably identify velocity-coherent filaments that correspond to structures in the gas density field.
- Spine extraction provides a simplified representation of filaments that can aid in measuring their physical dynamics.
- Spine extraction with skeletonization instead of DisPerSE should make time evolution quicker and less noisy.
- **Future work:** Extend this framework to measure mass flux, filament mergers/splits, and accretion histories over time.

Clustering with HDBSCAN



Fig. 2: Volume rendering of the gas density showing that the spine extracted with skeletonization (red dot) traces the filamentary structures feeding the core. Scan QR code to view the video.

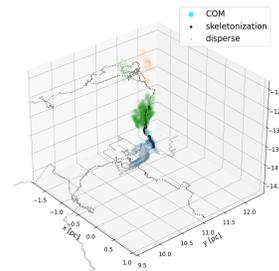


Fig. 3: Comparison of filament spines: DisPerSE skeleton of the gas density field (grey dots) versus the pruned spine from skeletonization (black dots), applied to a density-threshold mask of simulation cells mapped from clustered tracers.

- Tracer clusters indeed trace dense structures in the gas density (s. Fig. 2).
- Spine extraction from clustered tracers reproduces similar structures as the skeleton extracted from the gas density field, either via DisPerSE [5] or morphological skeletonization from scikit-image [6] (s. Fig. 3).
- **Velocity vectors** on clustered tracers point toward the core center of mass (COM), illustrating **accretion flows along filaments**, as well as rotation close to the core (s. Fig. 4).
- **Hyperparameter tuning:** too small parameter values lead to fragmentation of the main filament \Rightarrow optimization with relative validity (HDBSCAN) provides a score to select more stable and physically consistent clustering parameters.

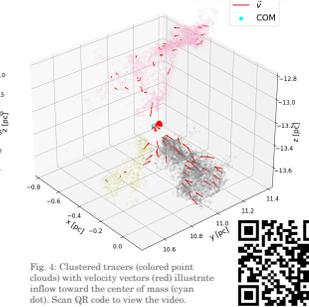


Fig. 4: Clustered tracers (colored point clouds) with velocity vectors (red) illustrate inflow toward the center of mass (cyan dot). Scan QR code to view the video.

Contact

Nuray Ortaköse
I. Physics Institute, University of Cologne
Email: ortakoese@ph1.uni-koeln.de

References

1. Clarke et al., 2018, MNRAS, 479, p. 1722-1746
2. Chen et al., 2020, ApJ, 891, 84
3. Meines et al., 2017, Journal of Open Source Software, 2(11), 205
4. Weis et al., 2024, MNRAS, 532, p. 262-1295
5. Soubie, 2011, MNRAS, 414, p.350-383
6. van der Walt et al., 2014, PeerJ, 2, e453

slideParam: 2D galactic data visualisation tool in Python

A. Jimenez, M. Alcázar-Layne, S. Verley, S. Duarte Puertas, M. Argudo-Fernández, D. Espada
Speaker Affiliation: University of Granada



slideParam

2D galactic data visualisation tool in Python

A. Jimenez¹, M. Alcázar-Layne¹, S. Verley^{1,2}, S. Duarte Puertas^{1,2}, M. Argudo-Fernández^{1,2}, D. Espada^{1,2}

¹Departamento de Física Teórica y del Cosmos, Facultad de Ciencias, Universidad de Granada, Granada, Spain
²Instituto Carlos I de Física Teórica y Computacional, Facultad de Ciencias, Granada, Spain



Abstract

slideParam provides an interactive visualisation tool to display a parameter with respect to 2D positions. The interactive tool uses one projection in the 2D space (for instance right ascension and declination) to study some parameters related to those positions. The tool allows the interactive visualisation of the data for different values for the selected parameter and their ranges.

Lightweight

Minimal installation required, only needing 3 renowned packages.

- ▶ pandas >= 2.0
- ▶ numpy >= 1.25
- ▶ matplotlib >= 3.6

Usage

Download repository from gitlab, install dependencies and customize for your needs. Easy to use, only 1 .py file:

```
python slideParam.py
```

Customization

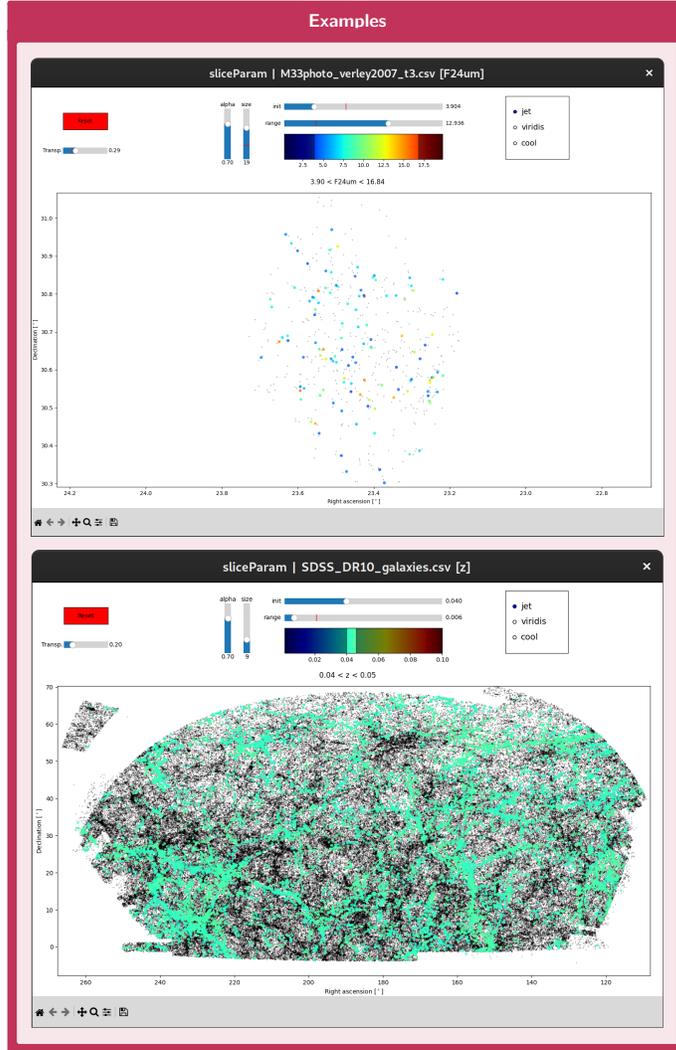
Tool can be customized in the PARAMETERS section of the script.

To change input file, change the value of `file_in`.

Input csv must be a data file with at least `ra` (Right Ascension) and `dec` (Declination) columns (any `x`, `y` data should work changing the input column names in `needed_cols`).

Additionally, `sel_col` can be assigned to any other column to visualize a subsample over the full sample.

Examples



Acknowledgements

The authors acknowledge financial support by the research projects AYA2017-84897-P, PID2020-113689GB-I00, PID2020-114414GB-I00, and PID2023-149578NB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) and by FEDER/UE; the project A-FQM-510-UGR20 funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Proyecto; by the grants P20-00334 and FQM108, funded by Junta de Andalucía; and by Consejería de Universidad, Investigación e Innovación (Junta de Andalucía) and Gobierno de España and European Union NextGenerationEU through grant AST22_4.4. This research made use of matplotlib (Hunter 2007); pandas (The pandas development team 2025); and NumPy (Walt 2011).

References

- [1] Hunter, J. D. 2007, CSE, 9, 90
- [2] The pandas development team. (2025). pandas-dev/pandas: Pandas (v2.3.3).
- [3] Taylor, M. B. 2005, in ASP Conf. Ser., Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

<https://gitlab.com/astrogal/slideParam>

MockLSS: A geometrical large-scale structure simulator of the local universe to make synthetic observational catalogues

M. Alcázar-Layne, S. Duarte Puertas, S. Verley, G. Blázquez-Calero, A. Jimenez

Speaker Affiliation: University of Granada



UNIVERSIDAD
DE GRANADA

MockLSS

A geometrical large-scale structure simulator of the local universe to make synthetic observational catalogues.

M. Alcázar-Layne¹, S. Duarte Puertas^{1,2}, S. Verley^{1,2}, G. Blázquez-Calero³, A. Jimenez¹

¹Departamento de Física Teórica y del Cosmos, Facultad de Ciencias, Universidad de Granada, Granada, Spain
²Instituto Carlos I de Física Teórica y Computacional, Facultad de Ciencias, Granada, Spain
³Instituto de Astrofísica de Andalucía - CSIC, Granada, Spain



Abstract

Current observational and simulated large-scale structure (LSS) catalogues often lack consistency in assigning galaxies to specific structures, due to the absence of a universally accepted classification criterion. In order to generate synthetic empirical data for fine-tuning LSS classification algorithms, as well as train Machine Learning (ML) / Deep learning (DL) models with the same purpose, this work presents a purely geometrical simulation based on the statistical spatial properties found in LSS surveys, using the spectroscopic main galaxy sample of the Sloan Digital Sky Survey (SDSS) catalogue up to a redshift of $z \approx 0.1$ as a specific use case. A parallelism between the LSS and the 3D Voronoi tessellation is utilised, in which the nodes, links, surfaces, and cells of the diagram correspond to **clusters**, **filaments and walls**, and **voids**, respectively. The simulation uses random positions within voids as seeds for tessellating the 3D space. The resulting tessellation structures are then randomly populated with galaxies that adhere to the statistical properties of their observational respective structures. As the galaxies are generated, they are tagged with their corresponding structure. In each simulation, six LSS mock catalogues are generated, following the statistical behaviour observed in the SDSS catalogue, depending on the structure that they belong to. In addition, the **Malmquist bias** and the redshift-space distortion (known as **Fingers of God** effect) are simulated as well. A novel geometrical LSS simulator is presented, where generated galaxies mimic the statistical properties of their observational belonging structure. As an example, the simulator is tuned to mimic the SDSS catalogue, although any other catalogue could be considered. With the generated catalogue, it is possible to adjust LSS classification algorithms, train and test ML/DL models, and benchmark several LSS classification methods using this pre-labelled data to contrast its results and performance.

Usage

To start a simulation with default values (SDSS-like simulation), execute the following command:

```
python3 src/mocklss.py config.toml
```

Customization

A large set of parameters can be tuned, such as the density of galaxies within each structure and its spatial position, allowing the construction of standard mock catalogues or even the generation of edge-case scenarios by changing parameters to extreme values (e.g., changing the number of voids to unrealistic values, generate thinner or thicker walls, etc.). By default, the simulator is configured to mimic the SDSS catalogue. Users can fine-tune its parameters in the `config.toml` file to adjust it to a different catalogue (e.g. DESI).

Source availability

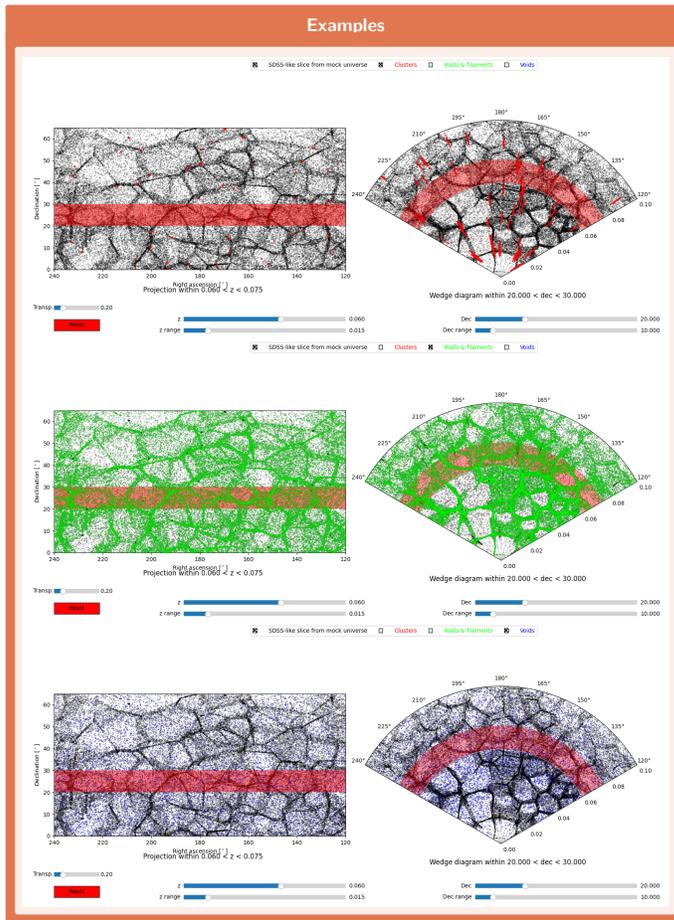
The source code and documentation of the presented simulator is open source and publicly available at: <https://gitlab.com/astrogal/mocklss>.

Acknowledgements

The authors acknowledge financial support by the research projects AYA2017-84897-P, PID2020-113689GB-I00, PID2020-114414GB-I00, and PID2023-149578NB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) and by FEDER/UE; the project A-FQM-510-UGR20 funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Proyecto; by the grants P20-00334 and FQM108, funded by Junta de Andalucía; and by Consejería de Universidad, Investigación e Innovación (Junta de Andalucía) and Gobierno de España and European Union NextGenerationEU through grant AST22_4.4. This research made use of Python Core Team (2019) v3.11 programming language. This work made use of Astropy (<http://www.astropy.org>) a community-developed core Python package and an ecosystem of tools and resources for astronomy (Astropy Collaboration et al. 2013, 2018, 2022); SciPy (Virtanen et al. 2020); Numpy (Harris et al. 2020); and Pandas (pandas development team 2020; McKinney 2010).

References

- [1] Tempel, E., Tuvikene, T., Kipper, R., & Libeskind, N. I. 2017, *A&A*, 602, A100
- [2] Pan, D. C., Vogele, M. S., Hoyle, F., Choi, Y.-Y., & Park, C. 2012, *MNRAS*, 421, 926
- [3] Hunter, J. D. 2007, *CSE*, 9, 90
- [4] The pandas development team. (2025). *pandas-dev/pandas: Pandas (v2.3.3)*.
- [5] Taylor, M. B. 2005, in *ASP Conf. Ser.*, Vol. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert, 29



<https://gitlab.com/astrogal/MockLSS>

LaVINComPy: Large Visualisation INterface COMputed with PYthon

S. Duarte Puertas, A. Jimenez, M. Alcázar-Layne, M. Argudo-Fernández, S. Verley, D. Espada, M. Hernández-Sánchez, G. Blázquez-Calero, C. Pozo González

Speaker Affiliation: University of Granada



LaVINComPy

LArge Visualisation INterface COMputed with PYthon

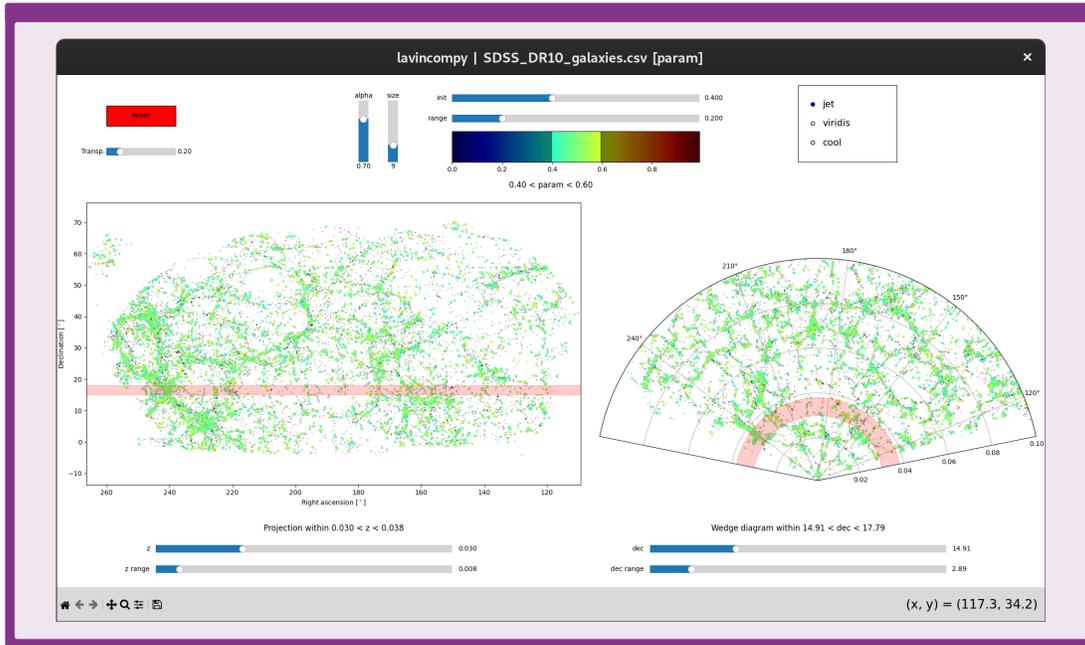


S. Duarte Puertas^{1,2}, A. Jimenez¹, M. Alcázar-Layne¹, M. Argudo-Fernández^{1,2}, S. Verley^{1,2}, D. Espada^{1,2}, M. Hernández-Sánchez³, G. Blázquez-Calero⁴, C. Pozo González^{1,4}

¹Departamento de Física Teórica y del Cosmos, Facultad de Ciencias, Universidad de Granada, Granada, Spain
²Instituto Carlos I de Física Teórica y Computacional, Facultad de Ciencias, Granada, Spain
³Universidad de Valencia, Valencia, Spain
⁴Instituto de Astrofísica de Andalucía - CSIC, Granada, Spain

Abstract

LaVINComPy is a Python-based interactive visualisation tool to display a certain parameter with respect to the 3D galaxy positions in the Universe. The tool represents a dataset using two complementary projections defined by right ascension, declination, and redshift. The interactive visualizations can be controlled by widgets of the data for different values of redshifts, redshift ranges, declinations, declination ranges, parameter values, and parameter ranges.



Lightweight

Minimal installation required, only needing 3 renowned Python packages:

- ▶ pandas >= 2.0
- ▶ numpy >= 1.25
- ▶ matplotlib >= 3.6

Usage

Download repository from **Git-Lab** and install dependencies and customize for your needs.

Easy to use, requiring only a single .py file:

```
python lavincompy.py
```

Customization

Tool can be customized in the **PARAMETERS** section of the script.

The input file can be changed by modifying the variable of `file_in`. Input csv must be a data file with at least ra (Right Ascension); dec (Declination); and z (Redshift) columns (input column names can be changed in `needed_cols`).

Additionally, `sel_col` can be assigned to any other column to visualize a sub-sample over the full dataset.

Acknowledgements

The authors acknowledge financial support by the research projects AYA2017-84897-P, PID2020-113689GB-I00, PID2020-114414GB-I00, and PID2023-149578NB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033), and by FEDER/UE; the project A-FQM-510-UGR20 funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Proyecto; by the grants P20-00334 and FQM106, funded by Junta de Andalucía; and by Consejería de Universidad, Investigación e Innovación (Junta de Andalucía) and Gobierno de España and European Union NextGenerationEU through grant AST22_4.4. This research made use of matplotlib (Hunter 2007); pandas (The pandas development team 2025); and NumPy (Walt 2011).

References

- [1] Hunter, J. D. 2007, CSE, 9, 90
- [2] The pandas development team. (2025). pandas-dev/pandas: Pandas (v2.3.3).
- [3] Taylor, M. B. 2005, in ASP Conf. Ser., Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shoppell, M. Britton, & R. Ebert, 29

<https://gitlab.com/astrogal/LAVINCOMPY>

GALAssify: A Python package for visually classifying astronomical objects

M. Alcázar-Layne, A. Jimenez, S. Duarte Puertas, C. Pozo González, J. Domínguez-Gómez, G. Blázquez-Calero, M. Argudo-Fernández, **S. Verley**, D. Espada, E. Florido, I. Pérez, T. Ruiz-Lara, L. Sánchez-Menguiano, A. Zurita

Speaker Affiliation: University of Granada



GALAssify

A Python package for visually classifying astronomical objects

M. Alcázar-Layne¹, A. Jimenez¹, S. Duarte Puertas^{1,2}, C. Pozo González^{1,4},
J. Domínguez-Gómez¹, G. Blázquez-Calero⁴, M. Argudo-Fernández^{1,2}, S.
Verley^{1,2}, D. Espada^{1,2}, E. Florido^{1,2}, I. Pérez^{1,2}, T. Ruiz-Lara^{1,2}, L.
Sánchez-Menguiano^{1,2}, A. Zurita^{1,2}

¹Departamento de Física Teórica y del Cosmos, Facultad de Ciencias, Universidad de Granada, Granada, Spain
²Instituto Carlos I de Física Teórica y Computacional, Facultad de Ciencias, Granada, Spain
³Instituto de Astrofísica de Andalucía - CSIC, Granada, Spain



Abstract

The visual classification of astronomical objects requires the use of tools that are simple and easily adaptable to the requirements of the user. In this context, we present **GALAssify**, a graphical tool that allows the user to visually inspect and characterise properties of astronomical objects in a simple way. In addition, GALAssify allows the user to save the results of the visual classification into a file using a list of previously defined tags based on the user's interests. GALAssify is available on **GitLab** and on **PyPI**. For many classification problems faced in astrophysics, a graphical interface greatly facilitates the job. In the present work we focus on the classification of galaxies to present a software that helps in a customised and simple way to do this classification. GALAssify has been developed in Python using PyQt5 libraries. A priori, it has been initially developed to tackle astrophysical problems but, due to its versatility, it could be easily adapted. For instance, this tool can be used to classify microscopy images from biological studies or be used in any other discipline. We provide instructions for the installation, usage and basic examples of how to use GALAssify in the online documentation at: <https://astrogal.gitlab.io/GALAssify/>

Easy install

Install from pip:
`pip install galassify`
Install from repository:
`pip install git+https://gitlab.com/astrogal/GALAssify.git`

Usage

- ▶ Left panel: list of galaxies, showing the name of each one, an icon indicating whether it has been processed, its coordinates and other relevant information.
 - ▶ Upper part of the right panel: image of the selected galaxy (bitmap and/or FITS).
 - ▶ Lower part of the right panel : classifying options (configurable by users). The classification of each galaxy can be edited or reset at any time.
- Finally, to save the selection, one can simply click the "Save and next" button or press the "enter" key. The entire classification is saved in a comma-separated values (CSV) file, easily readable with any text editor, spreadsheet program or database manager.

Easy to use:
`galassify --gui`
Initialize blank project:
`galassify --init`
Or use one of the provided examples as starting point:
`galassify --example`

Highly customizable

- ▶ Display section: configure display widgets to plot static and/or fits files (both can be **interactive** allowing **zoom** and **annotations**).
- ▶ Form section: configure the formulary presented in the GUI that users will interact with in order to perform the classification of the displayed image(s). The form can consist in a combination of **radio buttons**, **checkboxes**, and **plain text** fields.

Examples



Acknowledgements

The authors acknowledge financial support by the research projects AYA2017-84897-P, PID2020-113689GB-I00, PID2020-114414GB-I00, and PID2023-149578NB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) and by FEDER/UE, the project A-FQM-510-UGR20 funded by FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/Proyecto; by the grants P20-00334 and FQM108, funded by Junta de Andalucía; and by Consejería de Universidad, Investigación e Innovación (Junta de Andalucía) and Gobierno de España and European Union NextGenerationEU through grant AST22_4.4. This research made use of matplotlib (Hunter 2007); pandas (The pandas development team 2025); and NumPy (Walt 2011).

References

- [1] Hunter, J. D. 2007, CSE, 9, 90
- [2] The pandas development team. (2025). pandas-dev/pandas: Pandas (v2.3.3).
- [3] Taylor, M. B. 2005, in ASP Conf. Ser., Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

<https://gitlab.com/astrogal/GALAssify>

Author Index

- A.
Jimenez, 25–28
Zurita, 28
- Aarya
Patil, 13
- Adam A.
Miller, 11
- Alexander
Rüttgers, 17
- Alokda
Toka, 2
- Amit
Kashi, 6
- Andoni
Jiménez, 1
- Andreas
Lintermann, 3
- André
Bresges, 10, 19, 21
- Angie
Veronica, 17
- Anna
Scaife, 7
- Antonio
Lorenzo-Gutiérrez, 1
- Artyom
Khachatryan, 10
- Barone
Dario, 5
- Bernd
Doser, 20
- C.
Pozo González, 27, 28
- Canducci
Marco, 5
- Catarina
Corte-Real, 8
- Claudia
Comito, 17
- Cormac
McGleenan, 23
- Creaner
Oisin, 23
- D.
Espada, 25, 27, 28
- Daniel
Espada, 1
Seifried, 6
- Daniel L.
Holdsworth, 13
- David
Lawley, 23
- Dmitrii
Zagorulia, 8
- Donald W.
Kurtz, 13
- E.
Florido, 28
- Florian
Pacaud, 17
- Francesca Figueras
Siñol, 7
- Frank
Wagner, 4
- G.
Blázquez-Calero, 26–28
- Guillermo
Blázquez-Calero, 1
Cabrera-Vives, 11, 12
- Gábor
Marton, 7
- Hanwool
Koo, 14
- I.
Pérez, 28

Inês
 Rolo, 13

Isabel
 Pérez, 1

J.
 Domínguez-Gómez, 28

Jakob
 Dietl, 17

James Kostas
 Ray, 2

Joaquín
 Gallardo-Villagrán, 12

Kacper
 Drabicki, 16

Kai
 Polsterer, 5, 20

Kathleen
 Falconer, 10, 21

Kazantsev
 Andrei, 16

Kerstin
 Bunte, 7

L.
 Sánchez-Menguiano, 28

Lucie
 Flek, 9

Luis
 Cifuentes, 3

M.
 Alcázar-Layne, 25–28
 Argudo-Fernández, 25, 27, 28
 Hernández-Sánchez, 27

Maciej
 Bilicki, 16

Maksim
 Gabdeev, 6

Manuel Luis
 Alcáazar-Layne, 1

Marco
 Bischoff, 4

Margarida S.
 Cunha, 13

María del Carmen
 Argudo-Fernández, 1

Matthew
 Graham, 15

Mher
 Khachatryan, 10

Michael
 Kramer, 16

Michelle
 Lochner, 14

Monserrat
 Martínez-Marín, 12

Máté
 Madarász, 7

Nabeel
 Rehemtulla, 11

Narek
 Sahakyan, 10

Natalia
 Korsakova, 5

Nika
 Gorchakova, 23

Nuray
 Ortaköse, 24

Ofer
 Lahav, 1

Paul
 Nathan, 15

Pavlos
 Protopapas, 9

Porciana
 Cristiano, 2

Prateek
 Gupta, 22

Rahul
 Jayaraman and Ângela R. G.
 Santos, 13

Ramesh
 Karuppusamy, 16

Rodrigo
 Herrera-Camus, 12

Romain
 Chazotte, 5

S.
 Duarte Puertas, 25–28
 Verley, 25–28

Salvador
 Duarte Puertas, 1

Sebastian
 Trujillo-Gomez, 20

Sergio

Andreozzi, 3
Simon
 Verley, 1
Simone
 Vilardi, 7
Smith
 Rory, 5
Subhadeep
 Sarkar, 7
Susanne
 Pfalzner, 4
Szymon
 Nakoneczny, 16

T.
 Ruiz-Lara, 28
Teresa
 Antoja Castelltort, 7
Thomas

 Reiprich, 17
Tino
 Peter, 5

Valentino
 Farrell, 23
Victoria
 Antoci, 13
Vika
 Markosyan, 10

Wynn
 Jacobson-Galán, 11

Zorah
 Lähner, 17

Štefan
 Parimucha, 6