

Proof of concept: the potential of Use Case 3 within the AC3 Project to process astronomy data

Mario Chamorro-Cazorla
Dpto. de Física de la Tierra y Astrofísica
Universidad Complutense de Madrid
Madrid, Spain
mchamorro@ucm.es

Cristina Catalán-Torrecilla
Dpto. de Física de la Tierra y Astrofísica
Universidad Complutense de Madrid
Madrid, Spain
ccatalan@ucm.es

Ray Carroll
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
rcarroll@redhat.com

Ben Capper
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
bcapper@redhat.com

Ryan Jenkins
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
ryjenkin@redhat.com

Valeriia Sharapoval
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
vshapova@redhat.com

Abstract—AC3 (Agile and Cognitive Cloud-edge Continuum management) focuses on creating an innovative architecture utilizing a federated Cloud Edge Continuum infrastructure. At the core of this initiative is the Cloud Edge Continuum Computing Management (CECCM), which oversees the Life Cycle Management of both resources and applications. Additionally, AC3 integrates Artificial Intelligence and Machine Learning technologies to ensure energy-efficient operations.

Use cases play a major role in the project as they will demonstrate its utility and potential applications. Among the three use cases included in the project, UCM leads Use Case 3, whose main objective is to work on the distributed processing of Integral Field Spectroscopy (IFS) astronomy data. In this use case, we will use the modular architecture of AC3 to address the challenges posed by the large volume of data and computational resources available in astronomical research. We will employ microservices-based astronomy software deployed in containers to simplify its use and the processing of data cubes for researchers. We are currently working on testing the scalability of existing astronomical software, in particular some spectral synthesis programmes that allow us to understand properties of the stellar populations that populate galaxies. In the future, AC3 will have the ability to distribute computing resources according to workload requirements, optimising the processing of large volumes of datacubes.

Index Terms—Cloud Edge Continuum, Integral Field

Spectroscopy, Astronomy Data, Distributed Processing, Artificial Intelligence, Machine Learning.

I. INTRODUCTION

Galaxies are the building blocks of the Universe. To address the fundamental question about how galaxies have evolved through cosmic time, it is essential to have a deep understanding of the processes that govern their formation and evolution. Nearby galaxies offer us the opportunity to provide the foundations needed to understand them. We propose to use Integral Field Spectroscopy data for a large sample of these galaxies using already in-hand data from cutting-edge facilities such as MEGARA [1] on the 10.4 m Gran Telescopio de Canarias, MaNGA [2] installed in the 2.5 m Sloan Telescope and MUSE [3] at the Very Large Telescope (VLT).

The AC3 project aims to address the growing need for efficient data processing and resource management in various fields, including astronomy. The integration of AI and ML technologies within the CECCM framework allows for dynamic and adaptive management of computational resources, ensuring optimal performance and energy efficiency. By leveraging the modular architecture of AC3, we can effectively handle

the large volumes of data generated by instruments like MEGARA, facilitating the study of galaxy formation and evolution.

II. METHODOLOGY

Use cases play a major role in the AC3 project as they demonstrate its utility and potential applications. Among the three use cases included in the project, the Complutense University of Madrid (UCM) leads Use Case 3, which focuses on the distributed processing of Integral Field Spectroscopy astronomy data.

In this use case, we leverage the modular architecture of AC3 to address the challenges posed by the large volume of data and computational resources required in astronomical research. We employ microservices-based astronomy software deployed in containers to simplify its use and the processing of data cubes for researchers. This approach allows for greater flexibility and scalability in handling large datasets.

The data we will use comes from the MEGADES [4], MaNGA [5], and AMUSING++ [6] surveys, all of which are provided in the Flexible Image Transport System (FITS) format and as 3D datacubes. A 3D datacube is a multidimensional dataset that combines spatial and spectral information, capturing data from contiguous regions of the sky. This allows us to obtain spectral information of extended objects with spatial resolution (see figure 1).

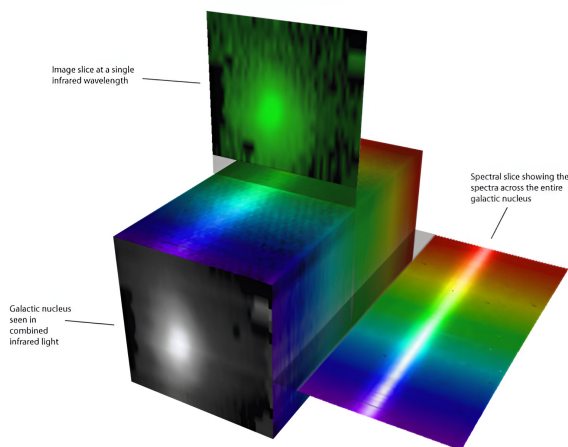


Fig. 1. Illustration of a 3D data cube combining spatial and spectral information. The image shows a spectrum from one pixel (top left), an image at a single wavelength (top right), and an image collapsed across all wavelengths (bottom left). Image credit: Stephen Todd and Douglas Pierce-Price.

We are currently working on testing the scalability of existing astronomical software, particularly spectral synthesis programs such as pPXF [7], Starlight [8], and STECKMAP [9]. These tools are designed to

decompose an observed spectrum into a combination of templates of Single Stellar Populations (SSP) with various stellar ages and metallicities, taking into account velocity broadening and instrumental effects (see figure 2). By analyzing the light from stars, these programs help us determine their composition, age, and other characteristics, providing valuable insights into the formation and evolution of galaxies.

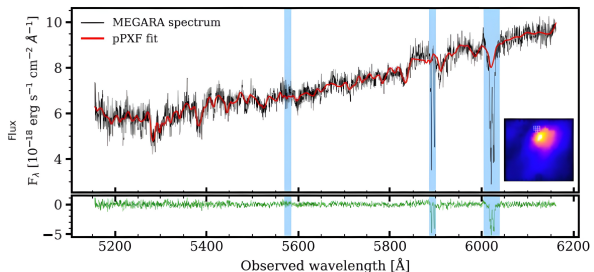


Fig. 2. Example of pPXF fitting to a spectrum of the galaxy UGC 10205 observed with MEGARA. The original observed spectrum is shown as the black line. The red and green line curves show pPXF best fits to the spectra and the corresponding residuals, respectively. The blue vertical stripes indicate the masked regions omitted from the pPXF fitting.

From a technical perspective, data processing is an exceptionally demanding task. Astronomers must manage vast volumes of data generated at various stages of analysis. For example, the final data cubes typically have dimensions of $40 \times 43 \times 4300$, resulting in 1720 spectra per cube for MEGARA observations, $319 \times 320 \times 3721$, equating to 102080 spectra for MUSE, and $74 \times 74 \times 6732$, yielding 5476 spectra for MaNGA data. This substantial data volume highlights the significant advantages that the CECC's management capabilities and infrastructure can offer to this research field.

III. RESULTS

Our initial tests, conducted using Kubernetes technology on local servers, have shown promising results in terms of scalability and performance. The modular architecture allows for efficient distribution of computational tasks across multiple nodes, significantly reducing processing times for large datasets. This is particularly important given the substantial data volumes involved, such as the 1720 spectra per cube for MEGARA observations, 102080 spectra for MUSE, and 5476 spectra for MaNGA data.

We have successfully deployed several spectral synthesis programs, including pPXF and Starlight, in this containerized environment. These tools have been used to analyze data from various astronomical instruments and telescopes, providing detailed information

about the stellar populations in a few galaxies. The integration of these programs has demonstrated the feasibility and efficiency of our approach, allowing us to decompose observed spectra into combinations of Single Stellar Population (SSP) templates, considering velocity broadening and instrumental effects.

The results from our initial tests indicate that our architecture can handle the complex and computationally intensive tasks required for full-spectrum fitting techniques. This capability is crucial for deriving key parameters such as the ages and metallicities of stellar populations, which are essential for understanding the formation and evolution of galaxies.

Currently, all testing is being conducted using Kubernetes technology on local servers. This setup has allowed us to validate our approach and ensure that it meets the necessary performance and scalability requirements. Although we have not yet included the functionality of STECKMAP in the project, we plan to integrate it in the near future.

At present, we have not yet established specific metrics to quantify the efficiency of our use case, as our primary focus has been on constructing the necessary architecture for integration within AC3 and adapting the existing software. However, preliminary tests suggest that our approach enhances the performance of the analyses, indicating that, once fully implemented, it will significantly improve efficiency.

IV. DISCUSSION AND CONCLUSIONS

The integration of AI and ML technologies within the CECCM framework enhances energy efficiency and enables more dynamic and adaptive management of computational resources, which is crucial for handling the large and complex datasets generated in astronomical research. The modular architecture of AC3 facilitates the implementation of specific solutions for different use cases, such as the distributed processing of IFS data. By deploying astronomy software in containers, we ensure greater flexibility and scalability, making it easier for researchers to process and analyze large datasets.

Our approach allows for better utilization of available computational resources, reducing the overall cost and energy consumption associated with data processing. This is particularly important for large-scale astronomical surveys, where efficient resource management is essential for timely and accurate data analysis. However, during the implementation of Use Case 3, we encountered several challenges and limitations, including the need for high computational power and the complexity of integrating various software components. Additionally, automating the entire analysis

process while retaining the ability to modify necessary parameters for each galaxy has proven to be particularly difficult. Future work will focus on addressing these challenges and improving the system's overall efficiency.

Initial tests have shown that the modular architecture of AC3, combined with containerized astronomy software, provides a scalable and flexible solution for processing large datasets. This approach not only improves performance but also reduces the overall cost and energy consumption. Although we have not yet included the functionality of STECKMAP in the project, we plan to integrate it soon.

Integrating our architecture into the AC3 infrastructure will enable dynamic allocation of computing resources based on workload requirements, further optimizing the processing of large volumes of data cubes. This capability will be particularly beneficial for large-scale astronomical surveys, such as those conducted with the James Webb Space Telescope [10] (JWST), which generate vast amounts of data that need to be processed and analyzed promptly. The optimization of analysis techniques applied to astronomical data via the CECCM will contribute significantly to the revolutionary discoveries anticipated from JWST, enhancing the speed and efficiency of data processing and analysis.

ACKNOWLEDGMENT

We/IPARCOS acknowledge/s the support from the “Tecnologías avanzadas para la exploración de universo y sus componentes” (PR47/21 TAU) project funded by Comunidad de Madrid, by the Recovery, Transformation and Resilience Plan from the Spanish State, and by NextGenerationEU from the European Union through the Recovery and Resilience Facility. This project has received funding from the European Union’s Horizon Europe Research and Innovation programme under grant agreement No 101093129. MCC, AGdP, ACM, JG, JZ, NC and SP acknowledge financial support from the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU) under the grant RTI2018-096188-B-I00 and PID2021-123417OB-I00.

REFERENCES

- [1] Gil de Paz, A. et al., “MEGARA, the new intermediate-resolution optical IFU and MOS for GTC: getting ready for the telescope”, in *Ground-based and Airborne Instrumentation for Astronomy VI*, 2016, vol. 9908, Art. no. 99081K. doi:10.1117/12.2231988.
- [2] Drory, N. et al., “The MaNGA Integral Field Unit Fiber Feed System for the Sloan 2.5 m Telescope”, *The Astronomical Journal*, vol. 149, no. 2, Art. no. 77, IOP, 2015. doi:10.1088/0004-6256/149/2/77.
- [3] Bacon, R. et al., “The MUSE second-generation VLT instrument”, in *Ground-based and Airborne Instrumentation for Astronomy III*, 2010, vol. 7735, Art. no. 773508. doi:10.1117/12.856027.

- [4] Chamorro-Cazorla, M. et al., “MEGADES: MEGARA galaxy disc evolution survey. Data release I: Central fields”, *Astronomy and Astrophysics*, vol. 670, Art. no. A117, 2023. doi:10.1051/0004-6361/202245111.
- [5] Bundy, K. et al., “Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory”, *The Astrophysical Journal*, vol. 798, no. 1, Art. no. 7, IOP, 2015. doi:10.1088/0004-637X/798/1/7.
- [6] López-Cobá, C. et al., “The AMUSING++ Nearby Galaxy Compilation. I. Full Sample Characterization and Galactic-scale Outflow Selection”, *The Astronomical Journal*, vol. 159, no. 4, Art. no. 167, IOP, 2020. doi:10.3847/1538-3881/ab7848.
- [7] Cappellari, M. and Emsellem, E., “Parametric Recovery of Line-of-Sight Velocity Distributions from Absorption-Line Spectra of Galaxies via Penalized Likelihood”, *Publications of the Astronomical Society of the Pacific*, vol. 116, no. 816, IOP, pp. 138–147, 2004. doi:10.1086/381875.
- [8] Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., and Gomes, J. M., “Semi-empirical analysis of Sloan Digital Sky Survey galaxies - I. Spectral synthesis method”, *Monthly Notices of the Royal Astronomical Society*, vol. 358, no. 2, OUP, pp. 363–378, 2005. doi:10.1111/j.1365-2966.2005.08752.x.
- [9] Ocvirk, P., Pichon, C., Lançon, A., and Thiébaud, E., “STECKMAP: STEllar Content and Kinematics from high resolution galactic spectra via Maximum A Posteriori”, *Monthly Notices of the Royal Astronomical Society*, vol. 365, no. 1, OUP, pp. 74–84, 2006. doi:10.1111/j.1365-2966.2005.09323.x.
- [10] Gardner, J. P. et al., “The James Webb Space Telescope”, *Space Science Reviews*, vol. 123, no. 4, pp. 485–606, 2006. doi:10.1007/s11214-006-8315-7.