A&A, 708, A297 (2026)
https://doi.org/10.1051/0004-6361/202557220

Astronomy & Astrophysics

The Galaxy Activity, Torus, and Outflow Survey (GATOS)

XII. Unveiling physical processes in local active galaxies. Unsupervised hierarchical clustering of JWST MIRI/MRS observations

L. Hermosa Muñoz[1,⋆], J. R. González Fernández[2], A. Alonso-Herrero[1], I. García-Bernete[1], O. González-Martín[3], M. Pereira-Santaella[4], E. López-Rodríguez[5], C. Ramos Almeida[6,7], S. García-Burillo[8], L. Zhang[9], A. Audibert[6,7], E. Bellocchi[10,11], F. Combes[12,13], T. Díaz-Santos[14,15], D. Esparza-Arredondo[3], B. García-Lorenzo[6,7], M. García-Marín[16], E. K. S. Hicks[17,9,18], Á. Labiano[19], N. A. Levenson[20], M. Martínez-Paredes[21], C. Packham[9], R. A. Riffel[22,23], D. Rigopoulou[24], J. Schneider[9], and M. Villar-Martín[23]

(Affiliations can be found after the references)

Received 12 September 2025 / Accepted 26 February 2026

ABSTRACT

Context. With the rise of integral field spectroscopy (IFS), we are currently dealing with large amounts of spatially resolved data, whose analysis has become challenging, especially when observing complex objects such as nearby galaxies.

Aims. We aim to develop a method of automatically separating regions with different physical properties (ionisation, kinematics, etc.) within the central parts (1″ ∼ 160 pc, on average) of galaxies. This could allow us to better understand the systems and provide an initial characterisation of the main ionisation sources affecting its evolution.

Methods. We developed an unsupervised hierarchical clustering algorithm to analyse data cubes based on spectral similarity. It clusters spaxels together with similar spectra, which is useful to disentangle regions affected by different processes, such as ionisation sources. We applied this method to a sample of 15 nearby (distances <100 Mpc) galaxies: 7 from the Galaxy Activity, Torus, and Outflow Survey (GATOS) and 8 archival sources, all observed with the medium-resolution spectrometer (MRS) of the Mid-Infrared Instrument (MIRI) on board the *James Webb* Space Telescope (JWST). The sample spans sources of various morphologies, active galactic nucleus (AGN) types, and/or starbursts. From the clusters, we computed their median spectrum and measured the line and continuum properties. We used these measurements to train random forest models and create several empirical mid-IR diagnostic diagrams for the MRS channel 3 wavelength range, ranging from 11.5 to 18 μm, which includes among others the bright [Ne II], [Ne III], and [Ne V] lines, several H_2 transitions, and PAH features.

Results. The clustering technique allows one to differentiate emission coming from an AGN, a nuclear starburst, the disc and star-forming (SF) regions in the galaxies, and other composite regions, potentially ionised by several sources simultaneously. This is supported by the results from the empirical diagnostic diagrams, which are indeed able to separate physically distinct regions. This innovative method serves as a tool to identify regions of interest in any data cube prior to an in-depth analysis of the sources. In a future work, we shall explore other wavelength ranges and a larger sample that would help us to obtain statistically significant conclusions.

Key words. ISM: jets and outflows – galaxies: active – galaxies: ISM – galaxies: nuclei – galaxies: structure

1. Introduction

With the rise of integral field spectroscopy (IFS) data available in the scientific archives, astronomers can now study in great detail different objects and physical processes (e.g. kinematics, ionisation, density, temperature, etc.) in both a spatially and a spectrally resolved way in galaxies. However, the complexity of data analysis and interpretation has equally increased. This is particularly true for the study of the central parts of nearby galaxies, where multiple physical processes are occurring simultaneously (e.g. Bacon et al. 2001; Emsellem et al. 2004; Cappellari et al. 2011; Sánchez et al. 2012; Cid Fernandes et al. 2013; Bundy et al. 2015; Cazzoli et al. 2020; Lin et al. 2020; Venturi et al. 2021; García-Bernete et al. 2021; Riffel et al. 2021; Peralta de Arriba et al. 2023; Chamorro-Cazorla et al. 2023; Alonso Herrero et al. 2024; García-Bernete et al. 2024b,c; Speranza et al. 2024; Zhang et al. 2024b; Hermosa Muñoz et al. 2024b, 2025), such as circular motions, shocks, star formation (SF) processes, and/or the presence and effect of an active galactic nucleus (AGN). All these processes can be studied through different tracers, such as molecular, ionised, or neutral gas, both in emission and/or in absorption depending on the wavelength of observation (for AGN, see e.g. Cazzoli et al. 2016; Fiore et al. 2017; Fluetsch et al. 2019).

From an observational perspective, a great effort has been made with optical surveys such as Mapping Nearby Galaxies at Apache Point Observatory (MaNGA, Bundy et al. 2015) or Calar Alto Legacy Integral Field Area (CALIFA, Sánchez et al. 2012). These surveys use spectroscopic data for large statistical samples of galaxies to study their evolution, morphologies, internal and external physical processes, and kinematics, among other matters. Within these surveys, several works are dedicated to identifying the ionising sources of the gas through the well-known Baldwin-Philips-Terlevich (BPT; Baldwin et al. 1981) diagnostic diagrams, but applied in a spatially resolved way to locate the position of SF regions, shocked regions, and/or AGN, if present, and so on (Belfiore et al. 2016; Gomes et al. 2016;

⋆ Corresponding author: lhermosa@cab.inta-csic.es

Law et al. 2021). From a modelling perspective, there are tools available that model the stellar continuum and the emission lines in the optical spectra, such as pPXF (Cappellari & Emsellem 2004; Cappellari 2017) or Pipe3D (Sánchez et al. 2016), or DeblendIRS (Hernán-Caballero et al. 2015), which separates the AGN, interstellar medium (ISM), and SF contributions in the mid-infrared (mid-IR) spectra of galaxies. Additionally, spectral decomposition tools as PAHFIT (Smith et al. 2007), or more recent tools such as the method presented in Donnan et al. (2024), or CAFE (Diaz-Santos et al. 2025), identify polycyclic aromatic carbon (PAH) features in the infrared spectra and model them together with the dust and stellar continuum. These provide further insights into the nature and distribution of the gas and dust components.

Compared to the optical, the infrared spectral range provides a significantly broader variety of features that allow us to characterise the ISM. In particular, galaxy spectra include warm molecular lines (e.g. H_2), atomic fine-structure lines covering a large range of ionisation states (ionisation potentials, IPs, from ∼7 to ∼190 eV), hydrogen recombination lines, PAH features, dust continuum, and absorption features from ices and several molecular species (García-Bernete et al. 2022a, 2024a). Thus, thanks to the diversity of tracers, this wavelength range offers a more detailed view of the different ISM phases in galaxies.

Large amounts of mid-IR data for nearby galaxies exist thanks to spectroscopic data from the *Spitzer* telescope Infrared Spectrograph (IRS; Houck et al. 2004), with several works dedicated to understanding the relative contribution of AGN and SF (see e.g. Armus et al. 2007; Pope et al. 2008; Moustakas et al. 2010; Alonso-Herrero et al. 2012). Nevertheless, these data provide an integrated spectrum of the sources that has large apertures (slit sizes from 3.6″ to ∼12″), a maximum resolution of ∼600, and does not always cover the complete mid-IR range. The large areas, as well as the low resolution, may dilute the detection of certain features within the circumnuclear regions beyond the AGN, such as SF regions, or shocks.

With the launch of the *James Webb* Space Telescope (JWST; Gardner et al. 2023), we have significantly increased the sensitivity and resolution of the near-infrared (near-IR) and mid-IR data. In the mid-IR range, JWST obtains IFS data with the medium-resolution spectrometer (MRS) of the Mid-Infrared Instrument (MIRI; Rieke et al. 2015; Wright et al. 2015, 2023). MIRI/MRS data are rich and complex, containing a wealth of information within a single data cube (e.g. Pereira-Santaella et al. 2022; García-Bernete et al. 2022a; Armus et al. 2023; Davies et al. 2024; Donnan et al. 2023, 2024; Dasyra et al. 2024; Zhang et al. 2024b; Goold et al. 2024; Esparza-Arredondo et al. 2025; Hermosa Muñoz et al. 2025; Riffel et al. 2025; Ramos Almeida et al. 2025; Alonso Herrero et al. 2025). To fully exploit the information from these datasets and derive physical properties across large samples of galaxies, it is essential to develop automated methods to analyse the data.

The application of methods for astronomical classification started already in the 1990s with neural network structures to classify stellar spectra or galaxy types (see e.g. von Hippel et al. 1994; Ball et al. 2004, and Smith & Geach 2023 for a review). Some classical methods aimed at simplifying the data analysis are still used, such as principal component analysis (PCA), which reduces the dimensionality problem of the data to obtain the main physical properties of the analysed objects (see e.g. Steiner et al. 2009, and references therein). However, the physical meaning of PCA components is often difficult to interpret, since they are a linear combination of various components. In recent years, several authors have started to develop alternative machine learning methods (see e.g. Baron & Ménard 2021; Chambon & Fraix-Burnet 2024; de Souza et al. 2025; Lu et al. 2025), using various techniques useful for handling complex data and identifying trends for different objects, some particularly focused on AGN (see e.g. Daoutis et al. 2025; Poitevineau et al. 2025; Nemer et al. 2025). For example, de Souza et al. (2025) analysed a sample of MaNGA galaxies using a clustering technique based on the spectral similarity within the cubes, named CAPIVARA. This allowed them to easily separate distinct physical regions of galaxies, such as the nucleus, bulge, spiral arms, or bars. In a certain way, clustering is similar to spatial binning techniques, such as Voronoi tessellations (Cappellari & Copin 2003), but based on the spectral physical properties rather than only in the signal-to-noise ratio (S/N). While clustering itself groups spectra based on their similarity, the interpretation of these clusters to a specific physical mechanism (e.g. AGN, SF, etc.) requires additional labelling and/or the use of supervised methods. These techniques are independent of the galaxy type and could potentially be used to disentangle regions affected by different ionisation sources, providing a more efficient and automated way to separate SF processes, shocks, and AGN ionisation not only in the optical (see e.g. Daoutis et al. 2025), but also in the mid-IR or other frequencies. Indeed, de Souza et al. (2025) classified their clusters into different categories using both the stellar continuum and emission line properties, based on the optical BPT diagrams (Baldwin et al. 1981). They reported an overall agreement between the cluster-based classification and the results from the traditional pixel-by-pixel analysis. This suggests that clustering tools can provide a simplified but accurate method of analysing complex data cubes.

In this paper, we explore an unsupervised hierarchical clustering technique with a sample of galaxies, most containing an AGN, observed using the MIRI/MRS on board JWST. Part of this dataset was observed within the Galaxy Activity, Torus, and Outflow Survey (GATOS) collaboration (García-Burillo et al. 2021; Alonso-Herrero et al. 2021). We would like to emphasise that this is an exploratory, empirical study that aims at evaluating this new analysis technique. We mainly focus on the search for empirical tracers that could potentially help to disentangle different ionising mechanisms and physical processes occurring in these galaxies using innovative machine learning techniques. To our knowledge, this is one of the early applications of a clustering method and automatic classification of the central regions of nearby galaxies using JWST spectroscopic data.

The paper is organised as follows. Section 2 describes the observations, the data reduction, and the methodology, based on custom-made codes. In Sect. 3 we present the main results of the clustering technique, including the median spectra per cluster, and other empirical measurements, such as the line ratios. In Sect. 4 we compare and evaluate the performance of the method in different mid-IR wavelength ranges, and we discuss the main caveats of the methodology. Finally, we present the summary and main conclusions of this work in Sect. 5.

2. Data and methodology

We selected a total sample of 15 nearby (distances <100 Mpc) galaxies (see Table 1) that primarily host different AGN types, observed with MIRI/MRS. The sources that were used as the training dataset represent all the local AGN and starburst galaxies, from archival and proprietary time, that have been studied in detail with MIRI/MRS mid-IR spectroscopic data in recent works (Alonso-Herrero et al. 2019;

Pereira-Santaella et al. 2022; García-Bernete et al. 2022a; Zhang & Ho 2023; Armus et al. 2023; García-Bernete et al. 2024b,c; Dasyra et al. 2024; Davies et al. 2024; Goold et al. 2024; Hermosa Muñoz et al. 2024a; Zhang et al. 2024b; Veenema et al. 2025), providing prior knowledge of the physical processes (e.g. kinematics, ionisation, and temperatures) at play in these systems. In this way, they can be used as a test bed to validate the technique and then apply it to new MIRI/MRS unexplored data cubes.

2.1. Data sample

We made use of MIRI/MRS data coming mainly from the GATOS collaboration (see Table 1). The sample consists of four Seyfert (Sy) galaxies observed during JWST General Observer (GO) programme Cycle 1 (NGC 3081, NGC 5506, NGC 5728, and NGC 7172; programme ID 1670, PI T. Shimizu; see details in Zhang et al. 2024b), whose main mid-IR properties have already been analysed in several works from the collaboration (see e.g. Pereira-Santaella et al. 2022; Hermosa Muñoz et al. 2024a; García-Bernete et al. 2024b; Davies et al. 2024; Zhang et al. 2024b,a; Esparza-Arredondo et al. 2025; Delaney et al. 2025), and three Sy galaxies observed during JWST GO Cycle 2 (NGC 3227, NGC 4051, and NGC 7582; ID 3535, PIs I. García-Bernete & D. Rigopoulou), which were used as a test bed for the methodology (see Sect. 4.4; Veenema et al. 2025).

We included the Sy galaxies Centaurus A (Cen A from now on), IC 5063, NGC 7319, and NGC 7469. These galaxies are publicly available in the archival and their MIRI/MRS data have been studied in detail in previous works. Cen A was observed within the guaranteed time observation programme MICONIC (ID 1269, PI N. Luetzgendorf; see Alonso Herrero et al. 2025), and IC 5063 was observed in the cycle 1 programme 2004 (PI K. M. Dasyra, Dasyra et al. 2024). The latter two objects were observed in the Early Release Observations programme (ID 2732, PI K.M. Pontoppidan, Pontoppidan et al. 2022) and the Early Release Science programme (ID 1328, PI L. Armus), respectively. Three of these objects, namely Cen A, IC 5063, and NGC 7319, are known to have a radio jet that perturbs the ISM, but their AGN are not always the dominant ionising source (Williams et al. 2002; Pereira-Santaella et al. 2022; Dasyra et al. 2024; Alonso Herrero et al. 2025). NGC 7469 hosts both a type-1.5 Sy and a nuclear starburst (Cazzoli et al. 2020; García-Bernete et al. 2022a; Zhang & Ho 2023; Armus et al. 2023). We included two low-luminosity AGN classified as low ionisation nuclear emission-line regions (LINERs), namely NGC 1052 and NGC 4594 (ID 2016, PI A. Seth, Goold et al. 2024), to compare with other AGN types (see Table 1). Finally, we included the pure starburst nuclei NGC 3256 N (ERS programme ID 1328, PI A. Lee, Bohn et al. 2024; Rigopoulou et al. 2024; García-Bernete et al. 2025) and M 83 (ID 2219, PI S.S. Hernandez, Hernandez et al. 2023, 2025), to compare with the AGN systems. The MIRI/MRS data for all these galaxies have already been published in previous works.

For all the data, the reduction process was done following the standard MRS pipeline procedure (e.g. Labiano et al. 2016; Bushouse et al. 2023 and references therein), with the pipeline release 1.11.4 and the calibration context 1130, except for Cen A (see details in Alonso Herrero et al. 2024, 2025). The details of the procedure are fully explained in Pereira-Santaella et al. (2022) and García-Bernete et al. (2022a, 2024a). We subtracted the background from all the cubes by computing a median background at each wavelength.

The MIRI/MRS covers a total wavelength range from 4.9 to 27.9 μm, divided into four integral field units (referred to as channels) with different fields of view (FoVs) and spatial and spectral resolutions (see more details in Labiano et al. 2021; Argyriou et al. 2023), namely channels 1–4 (ch1, ch2, ch3, and ch4). In this work we focus on ch3, which covers a range from 11.5 to 18μm, divided into three sub-channels (short, medium, and long). In particular, we used the ch3-short cubes (11.55–13.47 μm) and the combined ch3 spectral cubes (from now on referred to as 'ch3-all'). The latter were produced using the tools from the MRS reduction pipeline (CUBE_BUILD module). We show the median maps of the combined ch3 cube for all the galaxies in Appendix A (see Fig. A.1). We selected this channel mainly because it contains the three neon lines – [Ne II] at 12.81 μm, [Ne III] at 15.56 μm, and [Ne V] at 14.32 μm – that are typically used in the mid-IR to study the ionising source of the ionised gas (see e.g. Pereira-Santaella et al. 2010b), as well as H_2 lines and PAH features. We discuss other channels in Sect. 4.

2.2. Unsupervised hierarchical clustering technique

The complete methodology used in this paper is summarised in Fig. 1. We first applied an unsupervised hierarchical clustering technique to analyse the data cubes. This step is similar to that used by de Souza et al. (2025), so we refer the reader to this paper for more details (see also Sect. 1). In short, this is a machine learning technique that is used to group spectra that are similar, without any prior assumption about their shape or composition. Our algorithm, implemented in PYTHON[1], takes as input the spectra of all spaxels within the cube, calculates the distances based on a metric (Euclidean distance; see below), and then clusters them together based on their similarity.

Most of the galaxies analysed here behave as a bright point source, where the nucleus is several times brighter than the circumnuclear regions. Thus, to apply our methodology, we first normalised each spaxel spectrum by dividing it by its total integrated flux. In this way, we removed absolute flux differences and were able to focus exclusively on the spectral shape and relative emission line and PAH strengths.

The spectral similarity was evaluated by computing the Euclidean distance[2] between all spectra across the cube, defined as $d(x_i, y_i) = \sum_i (|x_i - y_i|)^2$, where x and y are two different spectra, to quantify how different each pair is. This process was done iteratively, computing a global distance matrix from all possible pairs of spectra in the cube. Based on this matrix, the algorithm searches for the most similar spaxels and groups them together into clusters. The result of this process is a dendogram, a tree-like structure that visually represents the sequence of cluster mergers. Each spaxel was considered as an individual cluster at the lowest level, and they were progressively merged into larger clusters based on their spectral similarity (i.e. measured distances). The clustering process stops at a number of clusters that were selected by 'cutting' the tree at a chosen level, allowing meaningful groupings to be extracted for each galaxy. This number was chosen by visual inspection. Stopping when adding more clusters either: 1) creates new concentric clusters from or around already formed clusters, or 2) creates new clusters from individual low-S/N spaxels from the edges of the FoV. To account for the rotational velocity field of the galaxies, the algorithm accounts for a spectral shift of ±6 spectral steps

[1] All the packages used for the analysis are listed at the end of the acknowledgement section.

[2] For a full discussion on different distance metrics, we refer the reader to Appendix A in Baron & Ménard (2021).

Table 1. Basic information for the galaxies used in this work.

Galaxy	Type	Distance (Mpc)	Redshift	Morph. type	Jet	Prop. ID	Reference	Sample
NGC 1052	LINER-1.9	19	0.0050	E4	Y	2016	[1]	Train
NGC 3081*	Sy-2	34	0.0082	(R)SAB0/a(r)	Y	1670	[2]	Train
NGC 3227*	Sy-1.5	15	0.0038	SAB(s)a pec	Y	3535	–	Test
NGC 3256-N	Starburst	40	0.0094	Merger pec	N	1328	[3,4,5]	Train
NGC 4051*	NLS1	16.6	0.0023	SAB(rs)bc	N	3535	–	Test
NGC 4594	LINER-2	10	0.0034	SA(s)a	Y	2016	[1]	Train
NGC 5506*	Sy-2	26	0.0061	Sa pec	Y	1670	[2]	Train
NGC 5728*	Sy-2	39	0.0092	SAB(r)a?	Y	1670	[6,7]	Train
NGC 7172*	Sy-2	37	0.0087	Sa pec	N	1670	[7,8]	Train
NGC 7319	Sy-2	98	0.0225	SB(s)bc pec	Y	2732	[9]	Train
NGC 7469	Sy-1.5	71	0.0163	(R')SAB(rs)a	N	1328	[10,11]	Train
NGC 7582*	Sy-2	22.7	0.0053	(R')SB(s)ab	-	3535	[12]	Test
IC 5063	Sy-2	48.6	0.0114	SA0^+(s)?	Y	2004	[13]	Train
M 83	Starburst	4.6	0.0017	SAB(s)c	N	2219	[14,15]	Train
Centaurus A	Sy-2	3.5	0.0018	S0 pec	Y	1269	[16]	Train

Notes. * indicates the galaxies observed within the GATOS collaboration (see Sect. 2). In Type: 'Sy' stands for Seyfert, and 'NLS1' for Narrow Line Seyfert-1. The redshift and morphological types have been obtained from NASA/IPAC Extragalactic Database (NED). In Jet: 'Y' stands for Yes, and 'N' stands for No; for NGC 7582 it is unclear (Veenema et al. 2025). The cited works refer exclusively to analyses using JWST data, used to label the clusters (see Sect. 2.3): [1] Goold et al. (2024), [2] Delaney et al. (2025), [3] Bohn et al. (2024), [4] Rigopoulou et al. (2024), [5] García-Bernete et al. (2025), [6] Davies et al. (2024), [7] García-Bernete et al. (2024c), [8] Hermosa Muñoz et al. (2024a), [9] Pereira-Santaella et al. (2022), [10] Armus et al. (2023), [11] Feuillet et al. (2025), [12] Veenema et al. (2025), [13] Dasyra et al. (2024), [14] Hernandez et al. (2023), [15] Hernandez et al. (2025), [16] Alonso Herrero et al. (2025). The last column indicates if the target has been used for the training or testing samples within the analysis.

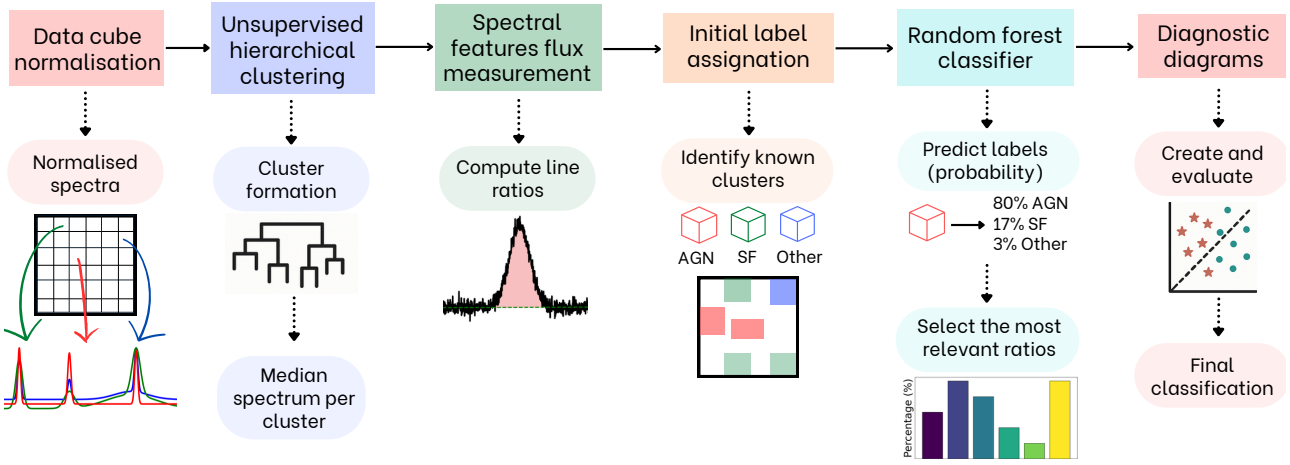

Fig. 1. Flowchart of the methodology discussed in Sect. 2.

(~300 km s^{-1}) during the clustering, used to minimise the distance calculation. The resulting cluster maps for each galaxy are presented in Sect. 3, Figs. 2 and 3, and in Appendix B.

After the clustering process ended, we computed the median spectrum for each cluster to evaluate their features. By using the median, we mitigated possible point spread function (PSF)- and/or combination-driven systematics, providing a representative spectrum for each cluster that contains their dominant spectral trends. We selected the median over the mean to avoid the appearance of double peaks in the emission lines due to possible velocity shifts within the cluster. We measured the slope of the continuum ($\alpha_{\rm mIR}$) between 12 and 17 μm for ch3-all, avoiding the lines and PAH features. We obtained the fluxes for the emission lines and the PAH features by integrating the profiles after subtracting a linear local continuum on either side of each feature. We considered the following lines: H_2 0–0 S(2) at 12.28 μm (from now on H_2 S(2)), HI (7-6) (Hu_α) at 12.37 μm, [Ne II] (IP of 21.6 eV), and [Ar V] at 13.10 μm (IP of 59.6 eV), and with the ch3-all cubes also [Mg V] at 13.52 μm (IP of 109.2 eV), [Ne V] (IP of 97.2 eV), [Cl II] at 14.37 μm (IP of 13.0 eV), [Ne III] (IP of 41.0 eV), and H_2 0–0 S(1) at 17.03 μm (from now on H_2 S(1)). We filtered out all lines with a low S/N (<3 times the standard deviation of the continuum) before the integration. We also considered the following PAH features: the one at 12 μm ($PAH_{12\,\mu m}$), the complex at 12.7 μm ($PAH_{12.7\,\mu m}$), and the one at 16.43 μm ($PAH_{17\,\mu m}$). There are additional PAH features in this wavelength range, but they are weaker and not present in all the sources (see Chown et al. 2024). To integrate the PAHs, we defined the wavelength ranges using as reference their emission in the starburst galaxies, where

they are stronger. We note that within this wavelength range we also detected the red end of the PAH complex at 11.3 μm. However, we did not consider it for the line ratio analysis, as it was only partially captured in ch3. For $PAH_{12.7\,\mu m}$, which typically includes the [Ne II] line, in each spaxel we subtracted the measured [Ne II] flux from the total flux of the feature to isolate the PAH emission. We cannot rule out the existence of residual line contribution to the total flux, or partial underestimation of the PAH flux due to possible extended wings.

To estimate the flux errors and account for possible flux variations within the individual spectra of a given cluster, we considered all the spectra contained within a single cluster and estimated the standard deviation per wavelength, computing an error spectrum. Then we used error propagation for the line integration and, later on, when computing the line ratios.

We note that the computational time required for the clustering process is correlated with both a larger spatial extension and a larger wavelength range (average computing time in a laptop with 32 GiB of RAM and 6 cores of ∼4 minutes for ch3-short, and ∼9 minutes for ch3-all). Because of the well-documented classification power of the mid-IR neon lines (see e.g. Martínez-Paredes et al. 2023; Feltre et al. 2023; Zhang et al. 2025, and references therein), we focus our method on the ch3 channel instead of using other channels or even the whole MIRI/MRS spectral range. Nevertheless, in Sect. 4 we discuss the possibility of applying this process to other channels. The code is in GitHub available for download[3].

2.3. Random forest classifier

Unsupervised clustering techniques have mainly been applied to IFS cubes covering the complete galaxy, identifying the large-scale structures (see Chambon & Fraix-Burnet 2024; de Souza et al. 2025, and Sect. 1), but not specifically to the circumnuclear regions of local galaxies, where multiple processes are occurring simultaneously. In order to evaluate if this technique is able to separate the main ionising source for individual regions in complex systems, we developed a complementary method aimed at assigning a physical meaning to the resulting clusters.

Most of the galaxies from our sample have already been analysed in previous works (see Sect. 2.1 and Table 1). Thus, we could use that previous information to associate some of the clusters with particular regions of the galaxy that we already know the physical nature of (e.g. an AGN, ionisation cone, SF region, or disc). In particular, this comes from evaluating their kinematics (line widths and velocities), emission features ratios (PAHs, warm molecular gas and/or mid-IR line ratios), and morphological properties (fluxes, SF circumnuclear rings, etc.). We can use this prior knowledge to create a subset of clusters that can be used as a training set for a supervised machine learning classifier. This way, we can identify those line ratios that are useful for separating regions with different ionisation sources (see scheme in Fig. 1). Specifically, we manually assigned labels to the known clusters such that: we labelled as 0 the region where the AGN is located, the disc or SF regions as 1, and the interacting or outflowing regions (hereafter referred as 'Other') as 2. We left the clusters with an unclear physical origin, or those for which we did not have any prior information, unlabelled (see Table B.1).

We then used these labelled clusters to train a random forest (RF) classifier, a machine learning algorithm that combines the input data to create a large number of individual decision trees (RandomForestClassifier from the SCIKIT-LEARN package in PYTHON, with the default parameters). It makes a final prediction based on the output that the majority of the trees agrees upon. Particularly for our case, the RF classifier was trained using the line ratios as the input features (see Sect. 3.3 for details), allowing us to automatically predict the most likely ionisation source for each cluster (i.e. the preferred label). This method provides a probabilistic classification of each cluster, assigning the final label to the category with the highest probability for each case. To take into account the flux errors of the measured line ratios, we ran a Monte Carlo (MC) simulation on the RF classifier a total of 1000 times, to have an estimation of the uncertainties for all the probabilities derived from the trained models. We used as the final label for each cluster the average probability of the most likely category obtained from the trained models after the simulation.

We tested the accuracy of the model using cross-validation with the labelled points available in our sample (31 labelled clusters out of 65; see Sects. 2.3 and 3.3). For that we randomly split the initially labelled clusters into training (80%, i.e. ∼24 points) and validation (20%, i.e. ∼7 points) sub-samples, generating 100 different splits. For each division, we ran 100 MC simulations to obtain the final RF model, as was explained before. The average classification accuracy of the RF model on the validation dataset is 84%.

From the resulting models, we can also evaluate the importance of each line ratio used for the classification process. This provides insights into the most relevant diagnostics that should be considered when distinguishing between ionising mechanisms in our sample. The RF classifier provides robust results and probabilities that allows us to evaluate the validity of the method (see Sect. 4.2.2 for a discussion on the method). Finally, we tested the trained model in three GATOS Sy galaxies observed during cycle 2; namely, NGC 3227, NGC 4051, and NGC 7582 (see Sect. 2.1). The results of this methodology are presented in Sect. 3.3 and discussed in Sect. 4.3.

3. Results

In this section we present the results of the clustering process for all the galaxies in the training sample. In the main text we show two galaxies as examples, namely NGC 7172 and NGC 5728, and the rest of the sample is presented in Appendix B. We selected these two galaxies as they are representative examples of the results (see Sects. 3.1 and 4.2).

3.1. Main properties of the clustering maps

We made use of the cubes corresponding to ch3-all and ch3-short when applying the clustering technique. When using only ch3-short, the most important features for the clustering are [Ne II], H_2 S(2), [Ar V], and the PAH features at 11.3 (partly), 12, and 12.7 μm. When considering the ch3-all cube, a similar trend is seen with all the lines and features present in this wavelength range (see Sect. 2.2), but the increased amount of information could lead to identify other structures dominated by different ionisation processes. This can be seen in Figs. 2 and 3, where we show the results for the clustering of the ch3-short (top panels) and ch3-all cubes (bottom panels) for the galaxies NGC 7172 and NGC 5728, respectively.

NGC 7172 is a nearly edge-on galaxy with a circumnuclear ring (Alonso Herrero et al. 2023) and a prominent ionisation cone extending almost perpendicular to the disc, previously detected with the MIRI/MRS data (Hermosa Muñoz et al.

[3] https://github.com/GonzalezFJR/agncluster

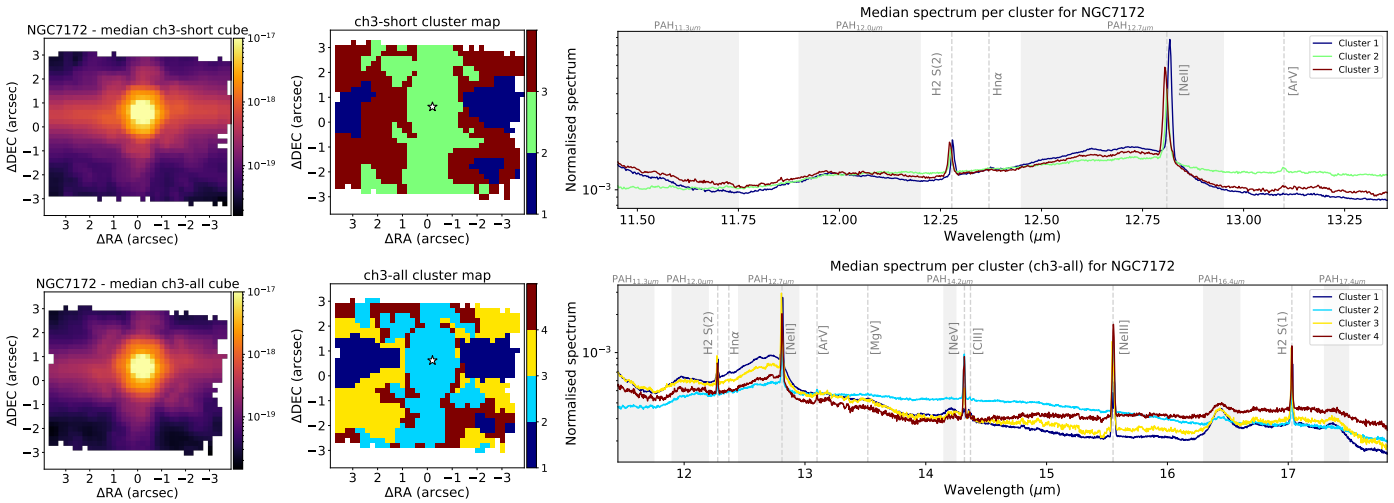

Fig. 2. Clustering results of the ch3-short cube (top) and the complete ch3 channel cube (bottom) for NGC 7172. Left panels: Median flux on a logarithmic scale of the ch3-short and ch3-all cube (top and bottom, respectively; see Appendix A). Middle panels: Cluster maps. Right panels: Median spectra per cluster on a logarithmic scale, normalised to the total integrated flux (see Sect. 2.2). The maps are centred on the original observed position (north is up and east to the left). The white star indicates the photometric centre. We assigned the same colours to the clusters and their respective spectrum. Colours were calculated automatically by dividing the 'jet' palette in MATPLOTLIB. We note that both the cluster colours and numbering are arbitrary, have no physical meaning, and are assigned independently in the top and bottom panels. We mark with dashed vertical grey lines the main emission lines, and with grey bands the PAH features in the spectrum. The wavelength is in rest frame.

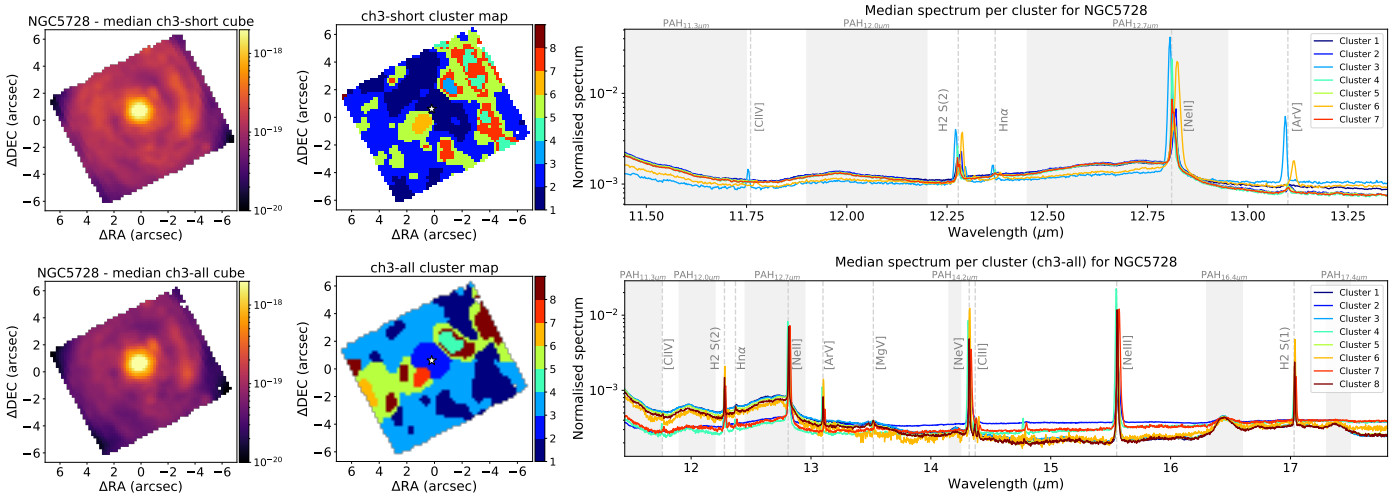

Fig. 3. Same as Fig. 2 but for NGC 5728. We note that in the top right panel we do not show the spectrum for cluster 8, as it is a low-S/N cluster.

2024a; Zhang et al. 2024b; García-Bernete et al. 2024c). These morphological features are reflected in the clustering maps (see Fig. 2, left). Using both ch3-short and ch3-all cubes, the system is well divided, with a total of three and four clusters, respectively. In both cases, the AGN and the ionisation cone are included together within a single cluster (cluster 2 for ch3-short and ch3-all). The disc and the SF clumps detected in Hermosa Muñoz et al. (2024a) to the south-west of the nucleus, associated with positive feedback produced by the interaction of the outflow with the ISM, are clustered together in both cases (cluster 1). Finally, there is an intermediate region between the ionisation cone and the disc in both cubes (cluster 3 for ch3-short and 3 and 4 for ch3-all).

NGC 5728, on the other hand, shows noticeably different results when comparing both cubes (see Fig. 3). This galaxy hosts a circumnuclear SF ring, a radio jet, and a known outflow (Shimizu et al. 2019; Davies et al. 2024; García-Bernete et al. 2024c). In the ch3-short cube, clusters clearly trace the ring and several SF regions (clusters 4, 5, and 7), the AGN-dominated region (cluster 1), and an intermediate region (cluster 2). However, in the ch3-all cube only some of these SF regions are traced (cluster 1), whereas the majority of the clusters are aligned with the direction of the outflow and the radio jet. This suggests that different physical processes dominate the spectra, and thus the clustering, in each spectral range. In both cases, we identify the hotspot detected by Davies et al. (2024) as an independent cluster (3 for ch3-short and 4 for ch3-all), indicating that it is a particular, differentiated region of this galaxy.

From the results for the other galaxies (see the figures in Appendix B), it is clear that the clustering is affected by the strong PSF of the JWST. This is particularly noticeable when using the ch3-all cube, as the PSF size increases with wavelength, but also in the ch3-short cube for some sources (e.g. NGC 1052; see Fig. B.1). We decided not to subtract the PSF in this work, as the main interest is to explore the results and caveats of this new technique, but its importance will be

evaluated in a future work (see also González-Martín et al. 2025). Despite this, we are able to differentiate regions of interest for each galaxy, especially when there is extended emission. Following what was observed for NGC 7172 and NGC 5728, in general the AGN-PSF, ionisation cone-outflow, intermediate, and SF-disc regions are isolated for almost all the galaxies (see the figures in Appendix B).

3.2. Median spectrum per cluster

We produced the median spectrum per cluster in all galaxies (right panels in Figs. 2 and 3 and in Appendix B), and only plotted those clusters with enough S/N in the continuum (>3 times the median standard deviation of all the spectra for all the clusters in a given source). These spectra allow us to evaluate the most relevant features driving the clustering.

Focusing on the spectra for NGC 7172 (right panels in Fig. 2), it is clear that the AGN+ionisation cone region (cluster 2 in ch3-short and ch3-all) has faint PAH features, while they are stronger in the disc region (cluster 1 in ch3-short and ch3-all). The intermediate region (cluster 3 in ch3-short and clusters 3 and 4 in ch3-all) has moderate PAHs, and more complex emission line profiles compared to the disc region. Although we have applied a velocity correction to the spectra (see Sect. 2.2), we also detect shifts in the peak of the emission lines, which suggests velocity differences between the clusters. For the ch3-all cube, these differences on the PAH features and the emission lines are also seen, and additionally there are some variations in the shape of the continuum between clusters.

When focusing on the median spectra for NGC 5728 (see right panels of Fig. 3), the differences in the ch3-all spectra are less evident. In this case, we increased the number of clusters to capture the emission from the SF regions in the galaxy (see Fig. 7 in Shimizu et al. 2019), resulting in a subdivision along the jet-outflow axis, with clusters that have similar spectral properties. For a more in-depth analysis of a particular source, if these apparently similar clusters share the same physical properties, they should be merged together to simplify the maps. In contrast, the ch3-short spectra are more different, mainly due to the emission lines, as was found for NGC 7172. Cluster 3 is the most distinct cluster, with very prominent high excitation lines (i.e. [Ar V] and [Cl IV] at 11.76 μm), coinciding with the hotspot (Davies et al. 2024). Velocity differences for the clusters are seen, as in NGC 7172. These could be related to physical differences such as the presence of multiple kinematic components, as has already been noted in previous works (e.g. Davies et al. 2024).

In general, these trends are repeated for all the AGN galaxies in the sample. We see that the nuclei tend to have faint PAH emission and strongest high excitation lines. These characteristics are consistent with what is typically observed when comparing to spectra of discs or SF regions (see e.g. NGC 7469 in Fig. B.7). We also detect low-S/N spectra in certain clusters in some galaxies (see e.g. Figs. B.1 and B.2), which mostly correspond to a few spaxels located at the edges of the FoV. In the LINERs NGC 1052 and NGC 4594 (see Figs. B.1 and B.4, respectively) the spectra are mostly flat, with the main feature separating the regions being the emission lines, which are quite broad in all cases, as has already been discussed in previous works (Goold et al. 2024). This is likely because these type of objects host mainly old stellar populations. A similar trend is seen for IC 5063 (see Fig. B.8) and for NGC 7319 (it has little gas due to past interactions, Pereira-Santaella et al. 2022 and references therein; see Fig. B.6), although in these cases the spectra of some clusters do show a mild contribution from the PAH features. Both galaxies host a low-intermediate power radio jet that is interacting with the ISM, with several radio hotspots differently affected by the interaction (Pereira-Santaella et al. 2022; Dasyra et al. 2024). Finally, for the starburst galaxies NGC 3256 N and M 83 (see Figs. B.3 and B.9) the spectra for all the clusters are very similar, mainly separated by the shape of the PAH features in ch3-all. In general, they do not show high excitation lines, except for clusters 1 and 3 in M 83. While the median spectrum shows only a faint [Ne V] line, the error spectrum, computed as the standard deviation of all the spectra within the cluster, reveals the line more clearly (see the insets in Fig. C.1), in agreement with the regions highlighted in Hernandez et al. (2025).

3.3. Line ratios per cluster

We measured the fluxes of all the emission lines and PAH features present in the spectra for all the clusters, as well as the slope of the continuum (see Sect. 2.2). We created line ratios accounting for all the available possibilities both for the ch3-short and ch3-all cubes to compare how the clusters behave for the different galaxies. By selecting lines that are close in wavelength, the differential extinction effects are minimised (Hernán-Caballero et al. 2020; Donnan et al. 2024). The histograms showing the distribution of some of these ratios measured in the clusters obtained with the ch3-all cubes are presented in Fig. C.2. From previous works, there are promising line ratios in the 11.5–17.5 μm mid-IR range that could help to disentangle between AGN, starburst, and/or sources affected by other ionisation mechanisms, such as [Ne III]/[Ne II] or [Ne V]/[Ne II], among others, for nearby galaxies (see e.g. Pereira-Santaella et al. 2010b; Inami et al. 2013; Martínez-Paredes et al. 2023; Feltre et al. 2023; García-Bernete et al. 2024c; Feuillet et al. 2025; Ramos Almeida et al. 2025; Alonso Herrero et al. 2025).

Considering only the ch3-all cubes, we have a total of 65 selected clusters for all galaxies, after excluding those with low S/N in the continuum. We estimated their line ratios, taking into account that, as expected, some clusters lack some features in their spectra, such as high excitation lines or PAHs. For those clusters that are associated with regions whose physical origin was already known from previous analysis of the MIRI/MRS data (see Table 1 for the references), we assigned them labels as is explained in Sect. 2.3. In total, we set the initial labels for 49% of the clusters, as is shown in Table B.1.

From the resulting RF model, we obtained the relevance of each line ratio to classify the clusters. This output quantifies the relative weight of each feature compared to the rest for the trained model (see Sect. 2.3). Figure 4 presents all the features evaluated by the model (ratios and $\alpha_{\rm mIR}$) ordered by their importance. The most relevant ratios for ch3-all are: [Ne III]/[Ne II] (30 ± 2%), [Ne V]/[Ne II] (11 ± 2%), H_2 S(2)/[Ne II] (9 ± 3%), $PAH_{12\,\mu m}/PAH_{17\,\mu m}$ (8 ± 3%), and H_2 S(2)/$PAH_{12.7\,\mu m}$ (7 ± 2%). The slope of the continuum, $\alpha_{\rm mIR}$, shows the lowest importance for the classification (~2%). In fact, considering the uncertainties of the relevance for each feature (see Fig. 4), ratios with importance below 10% are equally (un)important for the model, meaning that they are exchangeable in terms of classifying the clusters. With this in mind, we prioritise the H_2-based ratios to create the diagrams, as they allow us to construct diagnostic diagrams for the ch3-short cubes as well, as is explained below. The diagrams with the $PAH_{12\,\mu m}/PAH_{17\,\mu m}$ ratio are included in Appendix C (see Fig. C.4) and discussed in Sect. 4.3.

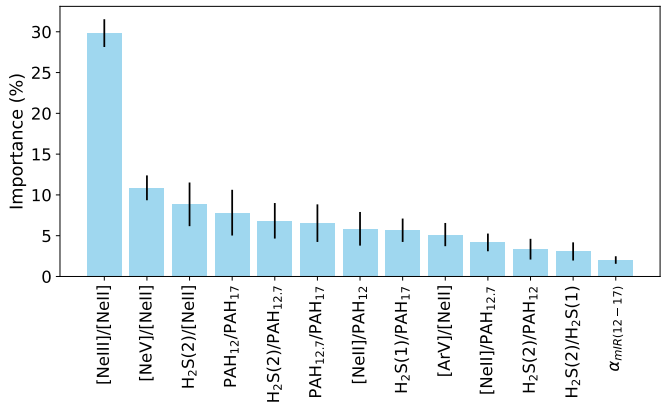

Fig. 4. Histogram of the average, relative importance of the features measured in the ch3-all cubes obtained from the automatic classification of the clusters (see Sect. 3.3). The error bars were estimated as the standard deviation of all the importances for each feature calculated using MC simulations (n = 1000; see Sect. 2.3).

In Fig. 5, we present the diagnostic diagrams created combining the most relevant ratios, with their probabilistic classification, for all the galaxies from the training sample. We detect a separation in the [Ne III]/[Ne II] ratio between regions dominated by SF and the rest of the clusters (see Fig. 5 and histogram in Fig. C.2). Clusters corresponding to NGC 3256-N, M 83, NGC 7469, and the disc in NGC 7172 are classified with the largest probabilities of being SF regions (greenish points), and have log([Ne III]/[Ne II]) < −0.5. These clusters have little to no emission of high IP gas, such as [Ne V], as they are SF-dominated, so most do not appear in the [Ne V]/[Ne II] diagram (see Figs. 5 and C.2). The H_2 S(2)/PAH$_{12.7\,\mu m}$ ratio also shows a bimodality, particularly when log([Ne III]/[Ne II]) > −0.5 (see Figs. 5 and C.2). A composite region with AGN-like and Other-like clusters is found at larger values of log(H_2 S(2)/PAH$_{12.7\,\mu m}$) and log([Ne III]/[Ne II]). This would be in agreement with previous works showing that the ratio is approximately constant for starbursts, while it is increased in the presence of an AGN or shocks (Roussel et al. 2007; Lambrides et al. 2019; Riffel et al. 2020; Zhang et al. 2022; Riffel et al. 2023; García-Bernete et al. 2024b). A similar trend is found for the H_2 S(2)/[Ne II] ratio, although the clusters are more concentrated and mixed than in the previous case at log([Ne III]/[Ne II]) > −0.5.

In contrast, and as was predicted by the RF classifier, there are other line ratios such as H_2 S(2)/H_2 S(1) (see Fig. C.2), related to the excitation temperature of the warm molecular gas, that show a similar distribution for all galaxies, and thus are not useful for separating regions. We note, however, that these two warm molecular gas lines have relatively close upper level energies. We cannot discard the possibility that combining other H_2 lines at shorter mid-IR wavelengths could help to disentangle different ionisation regions, as has been proposed in previous works with both *Spitzer* and JWST data (see e.g. Fig. 18 in Lambrides et al. 2019; Togi & Smith 2016; Costa-Souza et al. 2024; Ramos Almeida et al. 2025).

Additionally, we created equivalent diagnostic diagrams for the lines detected in ch3-short (see Fig. 6). In particular, based on the results from Fig. 4, the most relevant lines that can be used in both data cubes are H_2 S(2)/[Ne II] and H_2 S(2)/PAH$_{12.7\,\mu m}$. The clusters associated with SF regions, such as those of NGC 7469 or M 83, are found in the lower left part of the diagram, while those associated with AGN, such as the nuclei, are in the upper right part. Similarly to what was found for ch3-all (see Fig. 5), we see a bi-modality with this diagram that seems to be able to separate between both SF and AGN ionised regions, although with a large composite region.

4. Discussion

With the clustering process we aimed to provide a method of identifying physically distinct regions of a galaxy. In this section we discuss the main aspects to be considered when applying this methodology. In Sect. 4.1, we explore the possible differences on the results of the clustering when using the ch3-short or the ch3-all cubes. In Sect. 4.2 we present the caveats for the applied methodology, particularly focusing on exploring other MIRI/MRS channels, and on the automatic classification process of the clusters. Finally, we discuss the diagnostic diagrams in Sect. 4.3, and use the trained model to evaluate three galaxies, NGC 3227, NGC 4051, and NGC 7582, in Sect. 4.4.

4.1. The importance of the wavelength range: ch3-short versus ch3-all

As was seen in Sect. 3, the clustering results derived from using only the ch3-short cube versus the complete ch3 spectra can differ. This implies that the selected wavelength range can impact the clustering process results (see a discussion for other MRS channels in Sect. 4.2.1). The ch3-all cube includes more features that can be used to evaluate the performance of the method, such as several neon transitions ([Ne II], [Ne III], and [Ne V]). These are the brightest lines in this wavelength range, and trace gas with different ionisation levels (see Sect. 2.2). This is in contrast with the wavelength range covered by ch3-short, where [Ar V] traces the high excitation regions, although it is much weaker than [Ne V], and there are no intermediate excitation lines (IPs ∼30 eV to 90 eV), except for [Cl IV] at 11.76 µm (IP 53.5 eV), which is only detected in the hotspot of NGC 5728 (see Davies et al. 2024). Additionally, ch3-all cubes have a broader continuum range than ch3-short cubes. This is relevant, as more features are available to create the clusters, as well as possible changes in the continuum, which helps to separate potentially physically distinct regions.

Despite this, there are some objects, such as NGC 7172 and NGC 7469 (see Figs. 2 and B.7), for which there are similar clustering results when using both the ch3-short and ch3-all cubes. This suggests that in these objects the different physical regions (i.e. nuclei, disc, etc.) are more clearly separated, and thus there is a clear dominant ionising mechanism. This is in contrast to the results for more complex systems, such as NGC 5728, where the differences between both cubes are evident; see Fig. 3 and Sect. 3.1). For this galaxy, García-Bernete et al. (2024c) reported a strong coupling between the outflow and the disc, which significantly disturbs the ISM, suggesting that the gas is highly mixed and inhomogeneous. In these cases, the broadest wavelength range is to be preferred, as the larger variety of spectral features could be used to obtain a more comprehensive view of the physical processes at play.

Based on the results obtained with the RF classifier, among the most important line ratios for classifying the clusters (see Fig. 4 and Sect. 3.3), there are spectral features than can be computed with both wavelength ranges; specifically, those including H_2 S(2), PAH$_{12.7\,\mu m}$, and [Ne II]. In fact, as is shown in Fig. 6 (see also Sect. 3.3), the ch3-short range alone is useful for separating between SF and AGN or Other-dominated regions. However, the [Ne III]/[Ne II] ratio, only available when

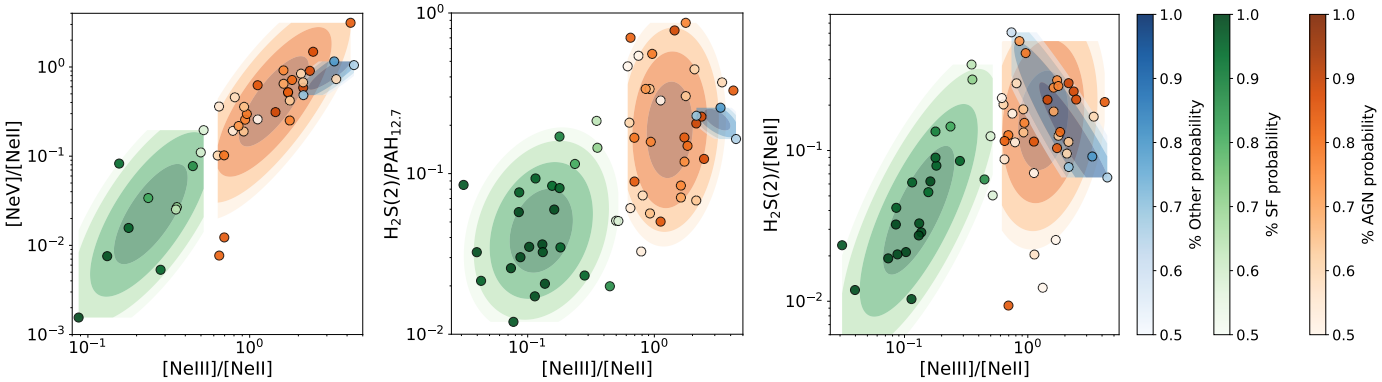

Fig. 5. Diagnostic diagrams based on the best preferred line ratios using the ch3-all cubes, on a logarithmic scale (see Fig. 4 and details in Sect. 3.3). Each point is a cluster from the galaxies used as the training sample, colour-coded by their assigned class probability (AGN in orange, SF in green, and Other in blue; see details in Sect. 3.3), with darker colours indicating a higher probability. The initial training labels of the clusters (see Table B.1) were obtained from previous detailed JWST MIRI/MRS analysis of the sources (see references in Table 1, and details in Sect. 2.3). Contours show the kernel density estimate (KDE) of the distribution for each class at four probability levels: 0.5, 0.6, 0.75, and 0.9.

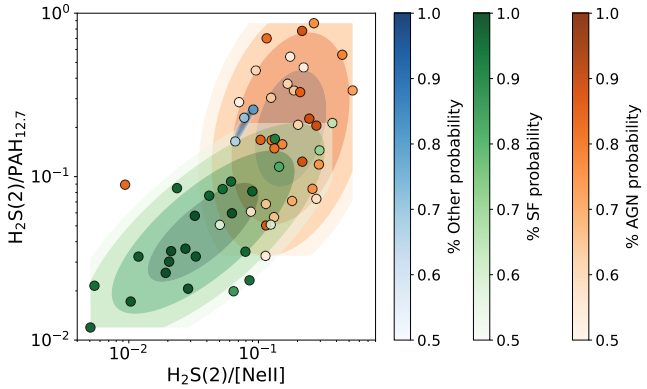

Fig. 6. Diagnostic diagram on a logarithmic scale similar to Fig. 5, but using the most relevant line ratios available for both ch3-short and ch3-all cubes for all the galaxies from the training sample (see Sect. 3.3).

using the complete ch3 cube, is the preferred ratio to separate SF and AGN-ionised regions not only in this work, but also as has been shown previously in several works in the literature using *Spitzer*/IRS spectroscopy (see e.g. Pereira-Santaella et al. 2010b; Inami et al. 2013; Martínez-Paredes et al. 2023).

4.2. Caveats of the methodology

4.2.1. Clustering technique

The unsupervised hierarchical clustering can be applied to a variety of data cubes observed at different wavelengths beyond the mid-IR traced by ch3 of MIRI/MRS. Nevertheless, to correctly interpret the result of this clustering technique when applying it to a new, untested dataset, a number of possible caveats need to be considered.

The choice of the spectral range significantly affects the clustering results, as the dominant spectral features change (see e.g. NGC 5728 in Sect. 3.1 and Fig. 3). This is particularly relevant when all the spectral features are equally dominant. Our tests with MIRI/MRS channel 1 (from 4.9 to 7.6 µm) revealed that the presence of many different features (low-, intermediate-, and high-excitation lines, as well as H_2, PAHs, and ices), without any particularly dominant line, made it difficult for the algorithm to separate physically distinct regions. More specifically, the regions that are identified with ch3, such as SF regions or the ionisation cones, are not clearly detected by using ch1 for some galaxies. This happens despite the presence of [Fe II] lines, typically used to identify shocks, and high-excitation lines such as [Mg V], most likely produced by AGN ionisation. In channel 2, there are no low-excitation lines, which means that the SF regions are not distinctively differentiated as clusters, and thus remain undetected with this method. Channel-2 should be evaluated carefully, as prominent silicate absorption features may introduce extinction effects on the observed spectral features, such as the H_2 S(5) line at 9.66 µm, and it may also show additional physical effects that are not recovered in the other channels. In channel 4 we have the lowest spatial and spectral resolution, and the largest PSF contribution, which may affect the detection of some interesting regions, as we have already encountered in ch3 for some galaxies (see Sect. 3.1). In a future work, the PSF subtraction tool created by González-Martín et al. (2025) could be used to subtract the PSF and test the methodology for the most affected data cubes, and evaluate in detail all of the other channels (including the complete MIRI/MRS cube with the whole mid-IR wavelength range), or even other wavelength ranges such as the optical, or near-IR with NIRSpec.

The combination of MIRI/MRS subchannels to create the ch3-all cubes, using the pipeline (see Sect. 2.1), can introduce systematic errors. These are translated into small flux discontinuities at the wavelengths where the cubes are combined, as well as possibly introducing errors in the spaxels located at the edges of the FoV due to shifts of the centroids. We detected such flux offsets for only two objects, NGC 5506 and IC 5063 (see the bottom panels in Figs. B.5 and B.8). When present, such flux differences are expected to affect the majority of the spaxels in the corresponding ch3-all cube. Given that each datacube is clustered independently, based exclusively on its internal structure, any offset will affect all spaxels uniformly. As a result, the offsets will not bias the clustering results. As for the potential shifting errors, most of the clusters excluded due to low S/N lie close to the FoV limits, and therefore have likely been excluded from the analysis. We also note that combining all MIRI/MRS channels will imply a significant reduction of the covered FoV ($3.2'' \times 3.7''$ for ch1 vs. $5.2''\times6.2''$ for ch3; Labiano et al. 2021), reducing the information currently recovered for the outermost regions.

The initial normalisation of the spectra is equally important. If we were to normalise in a particular wavelength region instead

of using the integrated flux per spaxel (see Sect. 2.2), the slope of the continuum would change significantly, especially at the ends of the wavelength range, altering the clustering. Also, variations in the results would appear depending on where this normalisation region were selected. Implicitly, this method assumes that there is a continuum to normalise, which may not always be the case (for example, extended gas emission in the outer parts of a galaxy).

Finally, the selection of the number of clusters is currently done through visual inspection (see Sect. 2.2). This could be refined in the future by using, for example, PCA (see e.g. Steiner et al. 2009), which will allow for a more robust and quantitative estimation of the most suitable number for each galaxy. The method in this work should serve as a tool to identify regions of interest within a cube, which can help to guide a future, in-depth analysis of a specific galaxy.

4.2.2. Automatic classification of the ionising source of the clusters

As was mentioned in Sect. 3.3, we have prior information about the physical origin of some clusters associated with particular regions of the galaxies, but we could not assign labels to all of them. Thus from the dataset, only a relatively small subset of clusters (49%; see Sect. 3.3) could be used to train the classifier, potentially reducing the robustness of the classification results.

In addition, the relative distribution of the labels across classes plays an important role in the performance of the classifier. Our initial sample is unbalanced, with an underrepresentation of the 'Other' class (∼13%, i.e. 4 of the initial labels; see Fig. C.3). This limited representation may affect the generalisation capability of the RF classifier for this specific class, particularly given the relatively small size of the labelled dataset.

Despite this, as is shown in Sect. 3.3, clusters associated with known discs, starbursts, and SF regions are all classified as SF with high probabilities (≥85%), occupying a well-defined part of the diagnostic diagrams (see Fig. 5 and Sect. 4.3). This suggests that, in our case, misclassified clusters are probably associated with composite regions, maybe affected by a combination of AGN ionisation, shocks, and/or other processes, making automatic classification challenging. If we were to use other MRS wavelength ranges, and thus account for other emission lines, we could create further diagnostic diagrams that are potentially useful, as those discussed in Feltre et al. (2023, see also Zhang et al. 2025; Ceci et al. 2025) are. The results of this machine learning approach should be considered as a preliminary test, but a larger dataset is needed to further probe and better constrain the results suggested by the diagnostic diagrams proposed in Fig. 5.

4.3. Identifying the ionising source of the clusters

Figure 5 shows the best preferred diagnostic diagrams to classify the clusters. The [Ne III]/[Ne II] ratio has been proposed to separate SF and AGN ionisation in several previous works in the mid-IR, including spatially resolved studies (see e.g. Groves et al. 2006; Pereira-Santaella et al. 2010b; Inami et al. 2013; García-Bernete et al. 2024c; Hermosa Muñoz et al. 2024a, 2025; Feuillet et al. 2025; Zhang et al. 2025). This ratio depends on the intensity and hardness of the radiation field, meaning that a larger value is associated with a more energetic ionising source, such as an AGN. We find a clear separation at ∼−0.5 in all diagrams for the most probable SF regions (>90%), although there are other SF regions at larger values, together with other AGN-classified clusters, in what we can define as composite regions (generally between log([Ne III]/[Ne II]) ∼ −0.5 and 3; see Fig. 5). This ratio compared to the [Ne V]/[Ne II] is a well-known estimator of AGN versus SF regions. Indeed, the nuclear regions of Sy galaxies and quasars tend to fall in the upper right part of this diagram (Zhang et al. 2024b; Hermosa Muñoz et al. 2025; Ramos Almeida et al. 2025; Alonso Herrero et al. 2025). The location of shocked regions in this diagram is uncertain, although photoionisation model predictions put it between the AGN and SF regions (see e.g. Feltre et al. 2023; Zhang et al. 2024a, 2025; Ceci et al. 2025). It is thus likely coincident with the composite region seen in our diagram at log([Ne V]/[Ne II]) below ∼−0.3 and log([Ne III]/[Ne II]) above ∼−0.5. In fact, the AGN distribution resembles that shown in Fig. 4 in Zhang et al. (2024a), although their models predict higher values of [Ne III]/[Ne II] than what we find. This could be a combination of them using larger apertures (3″ × 3″) than the average sizes of our clusters (average area of $6.9''^2$), and is also probably because we are still lacking a representative AGN distribution (see also Fig. B.7 in Zhang et al. 2025).

Riffel et al. (2025) compared the distributions of $H_2S(3)/PAH_{11.3\,\mu m}$ for AGN and non-AGN galaxies observed with *Spitzer* (Lambrides et al. 2019) with their MIRI/MRS galaxies. They systematically detected higher values of this ratio for the JWST-observed AGN. In general, SF- and AGN-dominated systems can also be distinguished using other warm H_2 transitions, such as $H_2S(1)/PAH_{11.3\,\mu m}$ (García-Bernete et al. 2024c; see also Pereira-Santaella et al. 2010a), with the largest values associated with AGN ionisation, similar to what we detect (see the middle panel in Fig. 5). These ratios are particularly high for the outflow region of NGC 5728, as it is strongly coupled with the jet and the host galaxy (García-Bernete et al. 2024c; Davies et al. 2024).

As for the $H_2S(2)$/[Ne II], SF emission tends to increase [Ne II], whereas in LINERs, where shocks are present, this ratio appears increased (Roussel et al. 2007). This is consistent with our results, although the regions are more mixed up than for the previous ratio (see the bottom panel of Fig. 5).

The use of PAH ratios and diagnostic diagrams has been widely discussed in previous works in the literature to disentangle regions with different ionisation conditions (Draine & Li 2007; Draine et al. 2021; Rigopoulou et al. 2021; Alonso-Herrero et al. 2014; García-Bernete et al. 2022c,b, 2024c; Zhang et al. 2024b), although normally different species are considered, such as PAHs at 7.7 µm or 11.3 µm (associated with neutral and ionised molecules, respectively). In general, PAHs are well-known tracers of SF activity and they can be destroyed due to the intense radiation field of the AGN. However, recent works with JWST observations have seen that neutral PAHs are more resilient near Seyfert-like AGN (García-Bernete et al. 2022c, 2024c). In our diagram (see Fig. C.4), the main separation between SF and AGN is given by the neon ratio, although large values of $PAH_{12\,\mu m}/PAH_{17\,\mu m}$ (>2 in log) are indicative of AGN ionisation. Within our sample, the points falling in this regime are mainly the nuclear- and ionisation cone-assigned clusters (e.g. cluster 2 in NGC 7172, cluster 5 in IC 5063, or cluster 2 in NGC 3081).

We note that it is possible that no purely shocked regions are detected in our data cubes. This would imply that even the most shocked regions would be contaminated by either SF or AGN ionisation, preventing a robust cluster classification for this type of region in this particular wavelength range. This could explain the composite regions that are detected in all diagrams in Fig. 5, formed by SF and AGN classified clusters, mainly with lower probabilities.

The number of local known galaxies selected in the analysis is still small. A larger sample could increase the confidence in the classification of the clusters in a particular category (see Sect. 4.2.2), and/or provide hints of additional categories that can be added to the model. The label assignation was done such that we consider clusters that were previously identified as interaction, shocked, and composite regions to be 'Other' (see Sect. 2.3). While using a single category simplifies the classification, it contains physically distinct regions that may have very different properties and ratios (e.g. a region illuminated by an AGN versus those where a jet and an outflow are interacting with the ISM). This would make the algorithm classify such composite regions as either AGN or SF, likely with a lower probability, instead of 'Other', where the dispersion in the measured features is larger. Nevertheless, the clusters assigned to Other all tend to fall at larger values of [Ne III]/[Ne II] ratios (see Fig. 5). These points correspond mostly to the jet-outflow-ISM interacting regions of IC 5063 and NGC 5728 (see Figs. B.8 and 3, respectively), and one cluster in NGC 1052 (see Fig. B.1). This classification, especially for the interacting regions, indicates that these behave differently from regular AGN-ionised regions. Given that within the sample there are other radio galaxies, also with known outflows, this suggests that additional processes are occurring, perhaps related to the geometrical coupling (Ramos Almeida et al. 2022; García-Bernete et al. 2024b; Harrison & Ramos Almeida 2024; Audibert et al. 2025) or to the power of the jet. However, there are few points in this category to draw any firm conclusion. With a larger sample of objects with well-identified regions, we could introduce other categories (such as a specific jet category) that could capture the true physical nature of these clusters in a more reliable way. The addition of other lines in the MIRI mid-IR range, or even in the near-IR data with NIRSpec, such as [Fe II], which is believed to be a good tracer of shocks, could also serve this purpose (e.g. Alonso Herrero et al. 2025).

4.4. Testing the methodology: NGC 3227, NGC 4051, and NGC 7582

To test the validity of the method, we applied the clustering technique and the RF model to the new MIRI/MRS data of the galaxies NGC 3227, NGC 4051, and NGC 7582, observed within the GATOS collaboration during cycle 2 (see Sect. 2.1). An in-depth analysis of the last source is presented in Veenema et al. (2025).

NGC 3227 also shows dominance of the PSF in the clustering, but several other regions in the north-east to the west of the nucleus are identified (see the top panel in Fig. 7). These could be related to the extended component identified by Alonso-Herrero et al. (2019) with ALMA data, attributed to radial streaming motions produced by gas being funnelled inwards, or to the [O III] ionised gas outflow extending up to 7″ north-east from the nucleus (see Falcone et al. 2024, and also Mundell et al. 1995). This galaxy has recent SF both in the nuclear and circumnuclear regions, inferred from the near-IR properties and the detection of PAH at 11.3 μm (Davies et al. 2006; Hönig et al. 2010; Alonso-Herrero et al. 2016). The regions ∼3−4″ southwest of the nucleus, corresponding to clusters 6 and 8 (partially), could be related to a SF region previously detected through ionised gas (see Alonso-Herrero et al. 2019). With the RF classifier, all the clusters are classified as AGN, although with median probability (∼48 to 52%), except for clusters 6, 8, and 9 (∼68%, 60%, and ∼64%, respectively). Clusters 5 and 9, classified as AGN (∼48% and ∼67%, respectively), are extended in the direction of the identified AGN ionisation cone where the non-circular motions were detected in previous works, which supports their AGN origin (see also Alonso-Herrero et al. 2019; Riffel et al. 2021; Falcone et al. 2024). In some cases such as cluster 3, the probability of being classified as an AGN is almost equal to being classified as SF, which could be a consequence of the recent SF and the AGN acting simultaneously in the (circum)nuclear region. A further in-depth analysis of these individual regions with the MIRI/MRS data are needed.

NGC 4051 is a narrow-line Sy-1 galaxy with an almost face-on ionised gas outflow (12° with respect to the line of sight) detected by Fischer et al. (2013) and Meena et al. (2021) with optical data (see also Christopoulou et al. 1997). Riffel et al. (2008) detected evidence of non-circular motions associated with a molecular gas inflow, using near-IR data from Gemini. These previously detected non-circular motions are not evident in our clustering maps. The strong PSF dominates the clustering results of the MIRI/MRS ch3-all data cube (see the middle panel in Fig. 7). However, we recovered two regions south from the nucleus, clustered together, (clusters 1 and 2 in ch3-all maps), which are classified as SF regions in Fig. 8. All the remaining clusters are classified as AGN, and are located in all the diagrams within the AGN contours in Fig. 8. A prior PSF subtraction of the cube (see the tool by González-Martín et al. 2025) could help to disentangle the previously detected, underlying physical processes, such as the ionised and molecular outflows.

Finally, NGC 7582 has been studied in great detail in the optical with MUSE data by Juneau et al. (2022), who mainly observed the approaching part of a biconical ionised gas outflow (opening angles for the north-western edge of 115° and for the southern edge of 15°) traced with [O III] 5007 Å (Juneau et al. 2022). The receding part was partially covered by dust from the galaxy disc (see also Riffel et al. 2009; Veenema et al. 2025). With the clustering technique applied to the MIRI/MRS cube, we detect the outflow cone (receding side: clusters 1 and 2 in Fig. 7; and, probably, the approaching side: clusters 5 and 6 in Fig. 7). We also captured part of what appears to be the SF ring previously detected with ALMA (Alonso-Herrero et al. 2020; García-Burillo et al. 2021). This region (clusters 7 and 8 in ch3-all map; bottom panel of Fig. 7) coincides with the SF-composite region detected with the BPT diagrams in Juneau et al. (2022), and is in fact classified as a SF region with the RF model in Fig. 8. For the ch3-all cube, however, the results from the RF classifier put the nucleus and its surrounding regions (clusters 5 and 6, respectively) as SF ionisation, and the receding part of the galaxy (north and east of the nucleus, clusters 1, 2, and 4) and the approaching part (cluster 3, although with low probability, ∼48%) as an AGN (see Fig. 8). This receding part also correspond to AGN ionisation based on the optical BPT diagram (see Fig. 13 in Juneau et al. 2022). We note that the clusters 5 and 6 (the nucleus and the approaching part of the galaxy, respectively; see Fig. 7), although classified as SF, fall in the composite regions for all the diagrams. This indicates (as was mentioned in Sect. 4.2.2) that these regions are probably affected by several physical processes simultaneously, perhaps produced by the superposition of the outflow and the disc along the line of sight, and thus the SF classification is not correct. Indeed, their derived probabilities are not as high as those corresponding to the disc clusters.

These examples demonstrate the potential of the clustering method to identify regions of interest, facilitating the analysis of new data cubes. It is important to note that using exclusively the line ratios and considering three categories to train the RF classifier is a simplistic way of classifying the clusters. This method does not consider more complex scenarios that may be present in the galaxies, such as the different coupling situations,

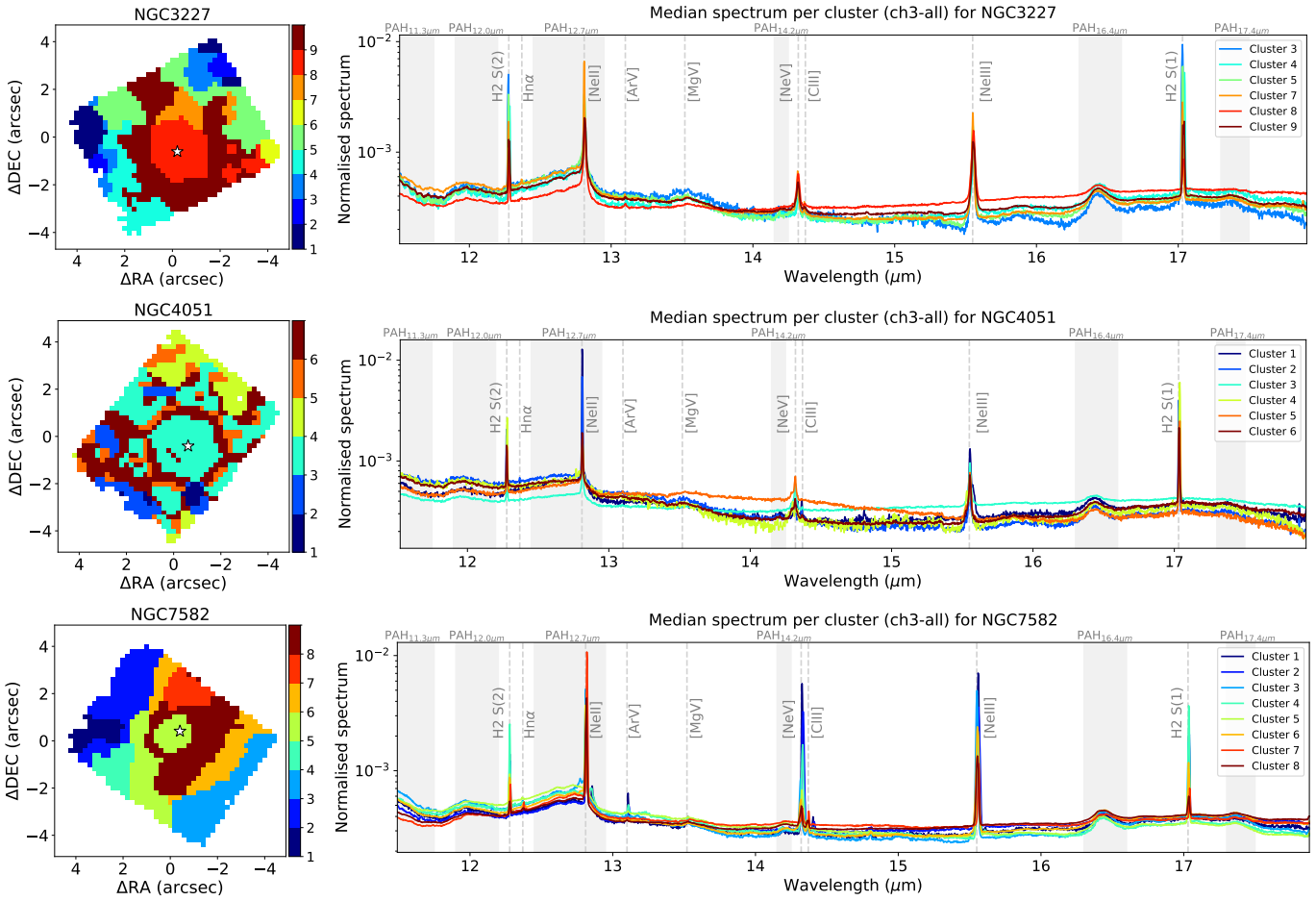

Fig. 7. Same as Fig. 2 but for the ch3-all cubes of NGC 3227, NGC 4051, and NGC 7582. Their corresponding median maps are in Fig. A.1.

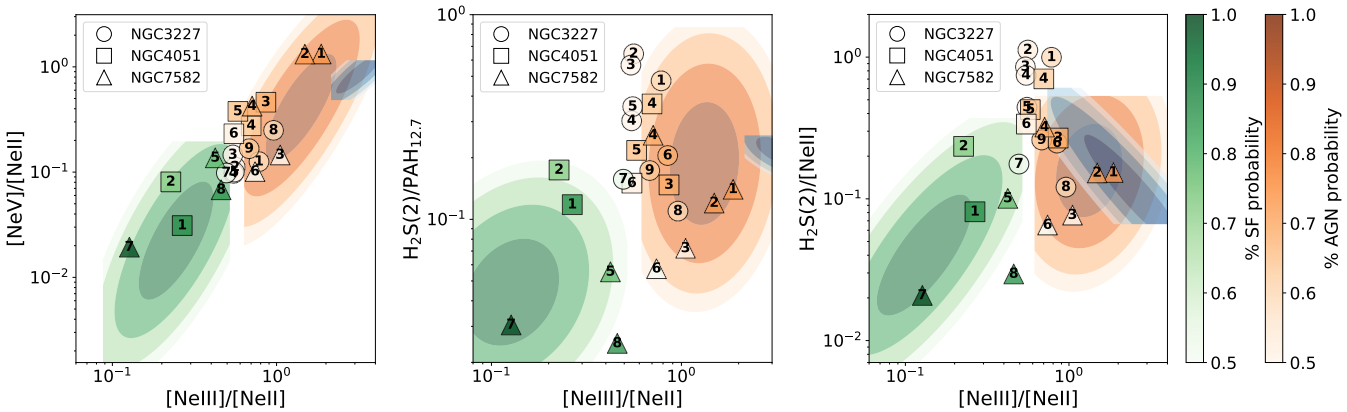

Fig. 8. Same diagrams as Fig. 5, with the KDE contours of the AGN (orange), SF (green), and Other (blue) distributions, with the predictions of the RF classifier (see Sect. 4.4) for the clusters derived from the ch3-all cubes of the galaxies from the testing sample: NGC 3227 (circles), NGC 4051 (squares), and NGC 7582 (triangles). We indicate the cluster number in each case (see the clustering maps in Fig. 7).

obscuration (which could still be significant in some sources), the power of the jets, or the overlap of multiple physical processes. This highlights the need for further investigation of these methods and diagnostic diagrams.

5. Summary and conclusions

In this work, we have presented a method based on an unsupervised hierarchical clustering technique to automatically identify regions of interest in data cubes of nearby galaxies based on spectral similarity. We have used data for 15 galaxies, mostly nearby AGN, observed with MIRI/MRS on board JWST, from the GATOS collaboration and the JWST archive (see Sect. 2.1). We used the channel-3 data cubes, covering a wavelength range from ∼11.5 to 18 μm. We applied the clustering technique to all the cubes, obtaining a median spectrum per cluster for each of the galaxies. We measured the fluxes of several lines of interest (e.g. [Ne II], [Ne V], and several H_2 transitions) as well as the

PAH features in this range. We then estimated line ratios with these features, and with them we trained a RF classifier to try to automatically identify the main ionising source for each cluster (AGN, SF, or Other). Here we present the main results of the analysis:

- Clustering technique: The proposed methodology is useful to identify potentially interesting regions of galaxies, such as SF or disc regions. The nuclei for all the active galaxies are always identified as independent clusters, although sometimes they are identified together with the ionisation cones. We have checked the validity of the method for the circumnuclear regions of galaxies with MIRI/MRS data cubes, but it can be applicable to any cube observed with any instrument (considering the wavelength range and the normalisation). We note that this methodology is limited for objects with a bright point-like source, as the PSF dominates the clustering. In these cases, a prior PSF subtraction should be performed.
- Dependence on the wavelength range: Using both ch3-short and ch3-all cubes we detected mainly consistent results in the clustering results, except for a few galaxies, such as NGC 5728. Despite this, to better evaluate the performance of the method, the whole wavelength range is preferred here. This is motivated by the larger amount of features available (i.e. low, intermediate, and high excitation, warm molecular, and neutral gas lines, and PAH features), as well as the continuum, which allow for further characterisation of the clusters.
- Mid-IR diagnostic diagrams: We have found that the most relevant line ratios to be used to classify the clusters using exclusively the ch3 cubes are [Ne V]/[Ne II], H_2 S(2)/[Ne II], [Ne III]/[Ne II], $PAH_{12\,\mu m}/PAH_{17\,\mu m}$, and H_2 S(2)/$PAH_{12.7\,\mu m}$. Using the complete MRS ch3 wavelength range, the diagrams formed with these ratios can distinguish between SF and AGN ionisation in all cases, especially that involving the H_2 S(2)/$PAH_{12.7\,\mu m}$ and [Ne III]/[Ne II]. We find composite regions in all the diagrams, which probably trace clusters with mixed ionisation.
- 'Other' regions: With the RF classifier we identified a group of clusters with larger [Ne III]/[Ne II] than regions with regular AGN ionisation (e.g. the nuclei). Although the sample is still small for us to draw any strong conclusion, we detected that most of these clusters correspond to interacting regions along the jet and outflow of IC 5063 and NGC 5728 (see Dasyra et al. 2024; Davies et al. 2024, for detailed analyses of these galaxies). Potentially, this means that the processes occurring in the ISM for these galaxies differ from the interactions happening in other galaxies that also have a radio jet within our sample. This suggests that additional physical mechanisms are at play for these two galaxies (e.g. ISM-outflow-jet coupling, power of the jet, or inclination effects).

Machine learning techniques are a powerful tool that should be explored to simplify the data analysis of IFS data cubes. The method presented here can be used as a test bed for further and larger analyses that incorporate additional MRS channels containing other emission lines and features (such as [Fe II] lines or other PAH ratios) that could be used to enhance the diagnostic power of the method. With a larger galaxy sample observed with the resolution of instruments such as MIRI/JWST, in the future we shall expand and put more constraints on the method, in order to classify the different physically distinct regions with more precision.

Acknowledgements. We thank the referee for his/her comments that have helped to improve the manuscript. LHM thanks M. Cerviño for useful discussions. LHM and AAH acknowledge financial support by the grant PID2021-124665NB-I00 funded by the Spanish Ministry of Science and Innovation and the State Agency of Research MCIN/AEI/10.13039/501100011033 PID2021-124665NB-I00 and ERDF A way of making Europe. IGB is supported by the Programa de Atracción de Talento Investigador "César Nombela" via grant 2023-T1/TEC-29030 funded by the Community of Madrid. OG-M acknowledge financial support from Ciencia de Frontera project number CF2023-G100 (SECIHTI) and PAPIIT project IN109123 (UNAM). MPS acknowledges support under grants RYC2021-033094-I, CNS2023-145506, and PID2023-146667NB-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. CRA acknowledges support from the Agencia Estatal de Investigación of the Ministerio de Ciencia, Innovación y Universidades (MCIU/AEI) under the grant "Tracking active galactic nuclei feedback from parsec to kiloparsec scales", with reference PID2022-141105NB-I00 and the European Regional Development Fund (ERDF). LZ, EKSH, CP, and JS acknowlege grant support from the Space Telescope Science Institute (ID: JWST-GO-01670). AA acknowledges funding from the European Union (WIDERA ExGal-Twin, GA 101158446). EB acknowledges support from the Spanish grants PID2022-138621NB-I00 and PID2021-123417OB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER, EU. DEA is supported by the "Becas Estancia Postdoctorales por México" EPM(1) 2024 (CVU:592884) program of SECIHTI, and acknowledges financial support from PAPIIT UNAM IN109123 and "Ciencia de Frontera" CONAHCyT CF2023-G100. RAR acknowledges the support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Proj. 303450/2022-3, 403398/2023-1 & 441722/2023-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; Proj. 88887.894973/2023-00). This work is based on observations made with the NASA/ESA/CSA *James Webb* Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST; and from the European JWST archive (eJWST) operated by the ESDC. These observations are associated with programs 1269, 1328, 1670, 2004, 2016, 2219, 2721, 2732, and 3535. This research has made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. This work has made extensive use of Python (v3.9.12), particularly with ASTROPY (v5.3.3; Astropy Collaboration 2013, 2018), LMFIT (v1.2.2; Newville et al. 2014), MATPLOTLIB (v3.8.0; Hunter 2007), SEABORN (v0.13.2; Waskom 2021), SCIPY (v1.11.2; Virtanen et al. 2020), NUMPY (v1.26.0; Harris et al. 2020), SCIKIT-LEARN (v1.4.1; Pedregosa et al. 2011), and PANDAS (v2.2.3).

References

Alonso Herrero, A., García-Burillo, S., Pereira-Santaella, M., et al. 2023, A&A, 675, A88
Alonso Herrero, A., Hermosa Muñoz, L., Labiano, A., et al. 2024, A&A, 690, A95
Alonso Herrero, A., Hermosa Muñoz, L., Labiano, A., et al. 2025, A&A, 699, A334
Alonso-Herrero, A., Pereira-Santaella, M., Rieke, G. H., & Rigopoulou, D. 2012, ApJ, 744, 2
Alonso-Herrero, A., Ramos Almeida, C., Esquej, P., et al. 2014, MNRAS, 443, 2766
Alonso-Herrero, A., Esquej, P., Roche, P. F., et al. 2016, MNRAS, 455, 563
Alonso-Herrero, A., García-Burillo, S., Pereira-Santaella, M., et al. 2019, A&A, 628, A65
Alonso-Herrero, A., Pereira-Santaella, M., Rigopoulou, D., et al. 2020, A&A, 639, A43
Alonso-Herrero, A., García-Burillo, S., Hönig, S. F., et al. 2021, A&A, 652, A99
Argyriou, I., Glasse, A., Law, D. R., et al. 2023, A&A, 675, A111
Armus, L., Charmandaris, V., Bernard-Salas, J., et al. 2007, ApJ, 656, 148
Armus, L., Lai, T., U, V., et al. 2023, ApJ, 942, L37
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Audibert, A., Ramos Almeida, C., García-Burillo, S., et al. 2025, A&A, 699, A83
Bacon, R., Copin, Y., Monnet, G., et al. 2001, MNRAS, 326, 23
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Ball, N. M., Loveday, J., Fukugita, M., et al. 2004, MNRAS, 348, 1038
Baron, D., & Ménard, B. 2021, ApJ, 916, 91
Belfiore, F., Maiolino, R., Maraston, C., et al. 2016, MNRAS, 461, 3111
Bohn, T., Inami, H., Togi, A., et al. 2024, ApJ, 977, 36
Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
Bushouse, H., Eisenhamer, J., Dencheva, N., et al. 2023, https://doi.org/10.5281/zenodo.8247246

Cappellari, M. 2017, MNRAS, 466, 798
Cappellari, M., & Copin, Y. 2003, MNRAS, 342, 345
Cappellari, M., & Emsellem, E. 2004, PASP, 116, 138
Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 413, 813
Cazzoli, S., Arribas, S., Maiolino, R., & Colina, L. 2016, A&A, 590, A125
Cazzoli, S., Gil de Paz, A., Márquez, I., et al. 2020, MNRAS, 493, 3656
Ceci, M., Marconcini, C., Marconi, A., et al. 2025, A&A, accepted [arXiv:2507.08077]
Chambon, H. J., & Fraix-Burnet, D. 2024, A&A, 688, A19
Chamorro-Cazorla, M., Gil de Paz, A., Castillo-Morales, Á., et al. 2023, A&A, 670, A117
Chown, R., Sidhu, A., Peeters, E., et al. 2024, A&A, 685, A75
Christopoulou, P. E., Holloway, A. J., Steffen, W., et al. 1997, MNRAS, 284, 385
Cid Fernandes, R., Pérez, E., García Benito, R., et al. 2013, A&A, 557, A86
Costa-Souza, J. H., Riffel, R. A., Souza-Oliveira, G. L., et al. 2024, ApJ, 974, 127
Daoutis, C., Zezas, A., Kyritsis, E., Kouroumpatzakis, K., & Bonfini, P. 2025, A&A, 693, A95
Dasyra, K. M., Paraschos, G. F., Combes, F., et al. 2024, ApJ, 977, 156
Davies, R. I., Thomas, J., Genzel, R., et al. 2006, ApJ, 646, 754
Davies, R., Shimizu, T., Pereira-Santaella, M., et al. 2024, A&A, 689, A263
de Souza, R. S., Dahmer-Hahn, L. G., Shen, S., et al. 2025, MNRAS, 539, 3166
Delaney, D., Hicks, E. K. S., Zhang, L., et al. 2025, ApJ, 993, 217
Diaz-Santos, T., Lai, T. S. Y., Finnerty, L., et al. 2025, Astrophysics Source Code Library [record ascl:2501.001]
Donnan, F. R., García-Bernete, I., Rigopoulou, D., et al. 2023, MNRAS, 519, 3691
Donnan, F. R., García-Bernete, I., Rigopoulou, D., et al. 2024, MNRAS, 529, 1386
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Draine, B. T., Li, A., Hensley, B. S., et al. 2021, ApJ, 917, 3
Emsellem, E., Cappellari, M., Peletier, R. F., et al. 2004, MNRAS, 352, 721
Esparza-Arredondo, D., Ramos Almeida, C., Audibert, A., et al. 2025, A&A, 693, A174
Falcone, J., Crenshaw, D. M., Fischer, T. C., et al. 2024, ApJ, 971, 17
Feltre, A., Gruppioni, C., Marchetti, L., et al. 2023, A&A, 675, A74
Feuillet, L. M., Kraemer, S., Meléndez, M. B., et al. 2025, ApJ, 983, 49
Fiore, F., Feruglio, C., Shankar, F., et al. 2017, A&A, 601, A143
Fischer, T. C., Crenshaw, D. M., Kraemer, S. B., & Schmitt, H. R. 2013, ApJS, 209, 1
Fluetsch, A., Maiolino, R., Carniani, S., et al. 2019, MNRAS, 483, 4586
García-Bernete, I., Alonso-Herrero, A., García-Burillo, S., et al. 2021, A&A, 645, A21
García-Bernete, I., Rigopoulou, D., Aalto, S., et al. 2022a, A&A, 663, A46
García-Bernete, I., Rigopoulou, D., Alonso-Herrero, A., et al. 2022b, A&A, 666, L5
García-Bernete, I., Rigopoulou, D., Alonso-Herrero, A., et al. 2022c, MNRAS, 509, 4256
García-Bernete, I., Alonso-Herrero, A., Rigopoulou, D., et al. 2024a, A&A, 681, L7
García-Bernete, I., Pereira-Santaella, M., González-Alfonso, E., et al. 2024b, A&A, 682, L5
García-Bernete, I., Rigopoulou, D., Donnan, F. R., et al. 2024c, A&A, 691, A162
García-Bernete, I., Donnan, F. R., Rigopoulou, D., et al. 2025, A&A, 696, A135
García-Burillo, S., Alonso-Herrero, A., Ramos Almeida, C., et al. 2021, A&A, 652, A98
Gardner, J. P., Mather, J. C., Abbott, R., et al. 2023, PASP, 135, 068001
Gomes, J. M., Papaderos, P., Kehrig, C., et al. 2016, A&A, 588, A68
González-Martín, O., Díaz-González, D. J., Martínez-Paredes, M., et al. 2025, MNRAS, 539, 2158
Goold, K., Seth, A., Molina, M., et al. 2024, ApJ, 966, 204
Groves, B. A., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 371, 1559
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Harrison, C. M., & Ramos Almeida, C. 2024, Galaxies, 12, 17
Hermosa Muñoz, L., Alonso-Herrero, A., Pereira-Santaella, M., et al. 2024a, A&A, 690, A350
Hermosa Muñoz, L., Cazzoli, S., Márquez, I., et al. 2024b, A&A, 683, A43
Hermosa Muñoz, L., Alonso-Herrero, A., Labiano, A., et al. 2025, A&A, 693, A321
Hernán-Caballero, A., Alonso-Herrero, A., Hatziminaoglou, E., et al. 2015, ApJ, 803, 109
Hernán-Caballero, A., Spoon, H. W. W., Alonso-Herrero, A., et al. 2020, MNRAS, 497, 4614
Hernandez, S., Jones, L., Smith, L. J., et al. 2023, ApJ, 948, 124
Hernandez, S., Smith, L. J., Jones, L. H., et al. 2025, ApJ, 983, 154
Hönig, S. F., Kishimoto, M., Gandhi, P., et al. 2010, A&A, 515, A23
Houck, J. R., Roellig, T. L., van Cleve, J., et al. 2004, ApJS, 154, 18
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Inami, H., Armus, L., Charmandaris, V., et al. 2013, ApJ, 777, 156
Juneau, S., Goulding, A. D., Banfield, J., et al. 2022, ApJ, 925, 203
Labiano, A., Azzollini, R., Bailey, J., et al. 2016, SPIE Conf. Ser., 9910, 99102W
Labiano, A., Argyriou, I., Álvarez-Márquez, J., et al. 2021, A&A, 656, A57
Lambrides, E. L., Petric, A. O., Tchernyshyov, K., Zakamska, N. L., & Watts, D. J. 2019, MNRAS, 487, 1823
Law, D. R., Ji, X., Belfiore, F., et al. 2021, ApJ, 915, 35
Lin, L., Ellison, S. L., Pan, H.-A., et al. 2020, ApJ, 903, 145
Lu, C. X., Mittal, T., Chen, C. H., et al. 2025, ApJS, 276, 65
Martínez-Paredes, M., Bruzual, G., Morisset, C., et al. 2023, MNRAS, 525, 2916
Meena, B., Crenshaw, D. M., Schmitt, H. R., et al. 2021, ApJ, 916, 31
Moustakas, J., Kennicutt, R. C., Jr., Tremonti, C. A., et al. 2010, ApJS, 190, 233
Mundell, C. G., Holloway, A. J., Pedlar, A., et al. 1995, MNRAS, 275, 67
Nemer, A., Katkov, I. Y., Gelfand, J. D., & Cho, C. 2025, ApJ, 984, 106
Newville, M., Stensitzki, T., & Allen, D. B. 2014, https://doi.org/10.5281/zenodo.11813
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Peralta de Arriba, L., Alonso-Herrero, A., García-Burillo, S., et al. 2023, A&A, 675, A58
Pereira-Santaella, M., Alonso-Herrero, A., Rieke, G. H., et al. 2010a, ApJS, 188, 447
Pereira-Santaella, M., Diamond-Stanic, A. M., Alonso-Herrero, A., & Rieke, G. H. 2010b, ApJ, 725, 2270
Pereira-Santaella, M., Álvarez-Márquez, J., García-Bernete, I., et al. 2022, A&A, 665, L11
Poitevineau, R., Combes, F., Garcia-Burillo, S., et al. 2025, A&A, 693, A311
Pontoppidan, K. M., Barrientes, J., Blome, C., et al. 2022, ApJ, 936, L14
Pope, A., Chary, R.-R., Alexander, D. M., et al. 2008, ApJ, 675, 1171
Ramos Almeida, C., Bischetti, M., García-Burillo, S., et al. 2022, A&A, 658, A155
Ramos Almeida, C., García-Bernete, I., Pereira-Santaella, M., et al. 2025, A&A, 698, A194
Rieke, G. H., Wright, G. S., Böker, T., et al. 2015, PASP, 127, 584
Riffel, R. A., Storchi-Bergmann, T., Winge, C., et al. 2008, MNRAS, 385, 1129
Riffel, R. A., Storchi-Bergmann, T., Dors, O. L., & Winge, C. 2009, MNRAS, 393, 783
Riffel, R. A., Zakamska, N. L., & Riffel, R. 2020, MNRAS, 491, 1518
Riffel, R. A., Bianchin, M., Riffel, R., et al. 2021, MNRAS, 503, 5161
Riffel, R. A., Storchi-Bergmann, T., Riffel, R., et al. 2023, MNRAS, 521, 1832
Riffel, R. A., Souza-Oliveira, G. L., Costa-Souza, J. H., et al. 2025, ApJ, 982, 69
Rigopoulou, D., Barale, M., Clary, D. C., et al. 2021, MNRAS, 504, 5287
Rigopoulou, D., Donnan, F. R., García-Bernete, I., et al. 2024, MNRAS, 532, 1598
Roussel, H., Helou, G., Hollenbach, D. J., et al. 2007, ApJ, 669, 959
Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, A&A, 538, A8
Sánchez, S. F., Pérez, E., Sánchez-Blázquez, P., et al. 2016, Rev. Mex. Astron. Astrofis., 52, 21
Shimizu, T. T., Davies, R. I., Lutz, D., et al. 2019, MNRAS, 490, 5860
Smith, M. J., & Geach, J. E. 2023, R. Soc. Open Sci., 10, 221454
Smith, J. D. T., Draine, B. T., Dale, D. A., et al. 2007, ApJ, 656, 770
Speranza, G., Ramos Almeida, C., Acosta-Pulido, J. A., et al. 2024, A&A, 681, A63
Steiner, J. E., Menezes, R. B., Ricci, T. V., & Oliveira, A. S. 2009, MNRAS, 395, 64
Togi, A., & Smith, J. D. T. 2016, ApJ, 830, 18
Veenema, O., Thatte, N., Rigopoulou, D., et al. 2025, MNRAS, 544, 3361
Venturi, G., Cresci, G., Marconi, A., et al. 2021, A&A, 648, A17
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
von Hippel, T., Storrie-Lombardi, L. J., Storrie-Lombardi, M. C., & Irwin, M. J. 1994, MNRAS, 269, 97
Waskom, M. L. 2021, J. Open Source Softw., 6, 3021
Williams, B. A., Yun, M. S., & Verdes-Montenegro, L. 2002, AJ, 123, 2417
Wright, G. S., Wright, D., Goodson, G. B., et al. 2015, PASP, 127, 595
Wright, G. S., Rieke, G. H., Glasse, A., et al. 2023, PASP, 135, 048003
Zhang, L., & Ho, L. C. 2023, ApJ, 953, L9
Zhang, L., Ho, L. C., & Li, A. 2022, ApJ, 939, 22
Zhang, L., García-Bernete, I., Packham, C., et al. 2024a, ApJ, 975, L2
Zhang, L., Packham, C., Hicks, E. K. S., et al. 2024b, ApJ, 974, 195
Zhang, L., Davies, R. I., Packham, C., et al. 2025, ApJS, 280, 65

[1] Centro de Astrobiología (CAB) CSIC-INTA, Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain
[2] Universidad Internacional de la Rioja (UNIR), Av. de la Paz 137, 26006 Logroño, La Rioja, Spain

[3] Instituto de Radioastrononomía y Astrofísica (IRyA), Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro 8701 Ex-Hda. San José de la Huerta, Morelia, Michoacán 58089, Mexico
[4] Instituto de Física Fundamental, CSIC, Calle Serrano 123, 28006 Madrid, Spain
[5] Kavli Institute for Particle Astrophysics & Cosmology (KIPAC), Stanford University, Stanford, CA 94305, USA
[6] Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
[7] Departamento de Astrofísica, Universidad de La Laguna, 38205 La Laguna, Tenerife, Spain
[8] Observatorio Astronómico Nacional (OAN-IGN) – Observatorio de Madrid, Alfonso XII, 3, 28014 Madrid, Spain
[9] Department of Physics and Astronomy, The University of Texas at San Antonio, 1 UTSA Circle, San Antonio, Texas 78249, USA
[10] Departmento de Física de la Tierra y Astrofísica, Fac. de CC Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain
[11] Instituto de Física de Partículas y del Cosmos IPARCOS, Fac. de CC Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain
[12] Observatoire de Paris, LUX, PSL University, Sorbonne Université, CNRS, F-75014 Paris, France
[13] Collège de France, 11 Place Marcelin Berthelot, 75231 Paris, France
[14] Institute of Astrophysics, Foundation for Research and Technology – Hellas (FORTH), Heraklion 70013, Greece
[15] School of Sciences, European University Cyprus, Diogenes Street, Engomi 1516, Nicosia, Cyprus
[16] European Space Agency, c/o Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[17] Department of Physics and Astronomy, University of Alaska Anchorage, Anchorage, AK 99508-4664, USA
[18] Department of Physics, University of Alaska, Fairbanks, Alaska 99775-5920, USA
[19] Telespazio UK for the European Space Agency (ESA), ESAC, Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Spain
[20] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[21] 1142 Sunset Point Rd, Clearwater, Florida 33755, USA
[22] Departamento de Física, CCNE, Universidade Federal de Santa Maria, Av. Roraima 1000, 97105-900 Santa Maria, RS, Brazil
[23] Centro de Astrobiología (CAB) CSIC-INTA, Ctra. de Ajalvir km 4, Torrejón de Ardóz 28850, Madrid, Spain
[24] Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

Appendix A: Median maps of the ch3-all cubes for all the sample

We include here the median flux intensity maps of the ch3-all cubes for all the galaxies from the sample (see Table 1). They have been calculated by getting the median value of the cube along the spectral axis. We note that the continuum for most galaxies (9 out of 15) is dominated by a strong PSF pattern, clearly visible in the images.

Appendix B: Clustering results for the remaining galaxies

In this appendix we show the maps and spectra for all the galaxies not discussed in the main text, following Fig. 2. We include Table B.1, which contains the initial and final classification assigned with our methodology to all the individual clusters for all the galaxies used to train the RF model.

Appendix C: Additional figures

In this appendix we show the median spectra for clusters 1 and 3 of M 83 (Fig. C.1). We then show several figures related to the line ratios and the RF classifier. These include the probability distributions of the obtained line ratios (Fig. C.2). These are smoothed representations of the ratios from all clusters in a given galaxy, constructed using KDEs to provide a continuous visualisation of their overall distribution. In Fig. C.3, we show the balance plots on the labels of the clusters from the training sample. Finally, we include the diagnostic diagrams created with the PAH_{12}/PAH_{17} ratio (see Sect. 3.3), for the training and testing samples (see the left and right panels in Fig. C.4, respectively).

Table B.1. Probabilistic classification of all the clusters from the ch3-all cubes of the galaxies used to train the RF models.

Galaxy (a)	Cluster (b)	Init. label (c)	RF label (d)	AGN probability (e)	SF probability (f)	Other probability (g)
CenA	1	-	0	0.64 ± 0.08	0.14 ± 0.05	0.21 ± 0.06
	2	-	0	0.50 ± 0.11	0.16 ± 0.08	0.34 ± 0.10
	3	-	1	0.23 ± 0.08	0.71 ± 0.08	0.06 ± 0.03
	4	-	1	0.26 ± 0.08	0.64 ± 0.08	0.10 ± 0.04
	5	-	0	0.61 ± 0.09	0.24 ± 0.08	0.15 ± 0.05
	6*	0	0	0.79 ± 0.04	0.15 ± 0.04	0.06 ± 0.03
IC5063	1	2	2	0.29 ± 0.05	0.06 ± 0.03	0.66 ± 0.04
	3	-	0	0.60 ± 0.12	0.06 ± 0.05	0.34 ± 0.11
	4	-	0	0.75 ± 0.10	0.09 ± 0.07	0.17 ± 0.07
	5*	0	0	0.88 ± 0.04	0.04 ± 0.04	0.08 ± 0.04
	6	0	0	0.89 ± 0.04	0.02 ± 0.02	0.08 ± 0.03
M83	1	-	1	0.03 ± 0.02	0.96 ± 0.03	0.01 ± 0.01
	2	1	1	0.01 ± 0.01	0.99 ± 0.02	0.00 ± 0.01
	3	1	1	0.00 ± 0.01	0.99 ± 0.01	0.00 ± 0.00
	4	1	1	0.01 ± 0.01	0.97 ± 0.02	0.01 ± 0.01
	5	1	1	0.01 ± 0.01	0.99 ± 0.02	0.01 ± 0.01
	6*	1	1	0.03 ± 0.02	0.96 ± 0.03	0.02 ± 0.01
	7	1	1	0.01 ± 0.01	0.99 ± 0.01	0.00 ± 0.01
NGC1052	1	2	2	0.31 ± 0.04	0.08 ± 0.03	0.61 ± 0.01
	2	-	0	0.50 ± 0.12	0.29 ± 0.12	0.20 ± 0.07
	4*	0	0	0.83 ± 0.04	0.10 ± 0.04	0.07 ± 0.03
	5	-	0	0.51 ± 0.11	0.28 ± 0.10	0.21 ± 0.07
NGC3081	1	0	0	0.87 ± 0.04	0.05 ± 0.03	0.08 ± 0.03
	2	0	0	0.84 ± 0.05	0.02 ± 0.02	0.14 ± 0.05
	3	1	1	0.05 ± 0.03	0.94 ± 0.03	0.01 ± 0.01
	4	-	0	0.55 ± 0.13	0.31 ± 0.13	0.14 ± 0.06
	5*	-	0	0.75 ± 0.08	0.08 ± 0.05	0.16 ± 0.07
	6	-	0	0.59 ± 0.16	0.28 ± 0.18	0.14 ± 0.06
NGC3256N	1*	1	1	0.02 ± 0.02	0.98 ± 0.02	0.01 ± 0.01
	2	1	1	0.00 ± 0.01	1.00 ± 0.01	0.00 ± 0.00
	3	1	1	0.07 ± 0.02	0.90 ± 0.02	0.03 ± 0.02
	4	1	1	0.01 ± 0.01	0.99 ± 0.01	0.00 ± 0.01
	5	-	1	0.02 ± 0.02	0.96 ± 0.03	0.02 ± 0.02
	6	-	1	0.00 ± 0.01	1.00 ± 0.01	0.00 ± 0.00
	7	-	1	0.00 ± 0.01	0.99 ± 0.01	0.00 ± 0.01
	8	-	1	0.02 ± 0.02	0.96 ± 0.03	0.02 ± 0.02
NGC4594	1	-	0	0.51 ± 0.11	0.28 ± 0.12	0.21 ± 0.05
	2*	0	0	0.84 ± 0.02	0.09 ± 0.02	0.06 ± 0.02
	3	-	0	0.58 ± 0.07	0.18 ± 0.05	0.24 ± 0.05
	6	-	0	0.45 ± 0.08	0.34 ± 0.07	0.21 ± 0.05
	7	-	0	0.49 ± 0.09	0.28 ± 0.08	0.23 ± 0.05
NGC5506	1	-	0	0.77 ± 0.09	0.09 ± 0.06	0.14 ± 0.07
	3	-	0	0.78 ± 0.07	0.06 ± 0.04	0.16 ± 0.06
	4*	0	0	0.84 ± 0.06	0.05 ± 0.03	0.11 ± 0.05
	5	-	0	0.63 ± 0.12	0.08 ± 0.05	0.29 ± 0.10
	6	-	0	0.61 ± 0.12	0.07 ± 0.05	0.32 ± 0.11
NGC5728	1	1	1	0.11 ± 0.05	0.87 ± 0.06	0.02 ± 0.02
	2*	0	0	0.81 ± 0.05	0.06 ± 0.03	0.14 ± 0.04
	3	-	0	0.54 ± 0.09	0.36 ± 0.09	0.10 ± 0.04
	4	2	2	0.17 ± 0.04	0.03 ± 0.02	0.80 ± 0.04
	5	-	0	0.70 ± 0.08	0.16 ± 0.06	0.14 ± 0.05
	7	2	2	0.24 ± 0.04	0.06 ± 0.03	0.70 ± 0.03
	8	-	0	0.64 ± 0.08	0.18 ± 0.06	0.19 ± 0.06
NGC7172	1	1	1	0.04 ± 0.02	0.96 ± 0.03	0.01 ± 0.01
	2*	0	0	0.86 ± 0.04	0.08 ± 0.04	0.06 ± 0.03
	3	-	1	0.34 ± 0.13	0.59 ± 0.15	0.07 ± 0.04
	4	-	0	0.65 ± 0.10	0.22 ± 0.08	0.13 ± 0.06
NGC7319	5*	0	0	0.88 ± 0.04	0.02 ± 0.02	0.10 ± 0.04
	6	-	0	0.73 ± 0.10	0.10 ± 0.10	0.16 ± 0.06
	8	-	0	0.76 ± 0.08	0.04 ± 0.03	0.19 ± 0.08
NGC7469	1	1	1	0.01 ± 0.01	0.98 ± 0.02	0.00 ± 0.01
	2	1	1	0.01 ± 0.01	0.99 ± 0.01	0.00 ± 0.01
	3*	-	1	0.34 ± 0.13	0.57 ± 0.16	0.08 ± 0.05
	4	1	1	0.03 ± 0.02	0.96 ± 0.03	0.01 ± 0.01
	5	-	1	0.16 ± 0.07	0.81 ± 0.07	0.03 ± 0.02

Notes. Columns indicate: (a) Galaxy name, (b) cluster number, (c) initial label assigned based on previous works (0 is AGN, 1 is SF, 2 is Other; see Sect. 2.3), (d) final label assigned with the RF model, and (e), (f), and (g) probabilities and their corresponding standard deviation of being assigned to one of the available classes (AGN, SF, and Other, respectively). We note that clusters excluded due to S/N are not in this table (see Sect. 2.2). * indicates the cluster containing the nuclear region of the galaxy.

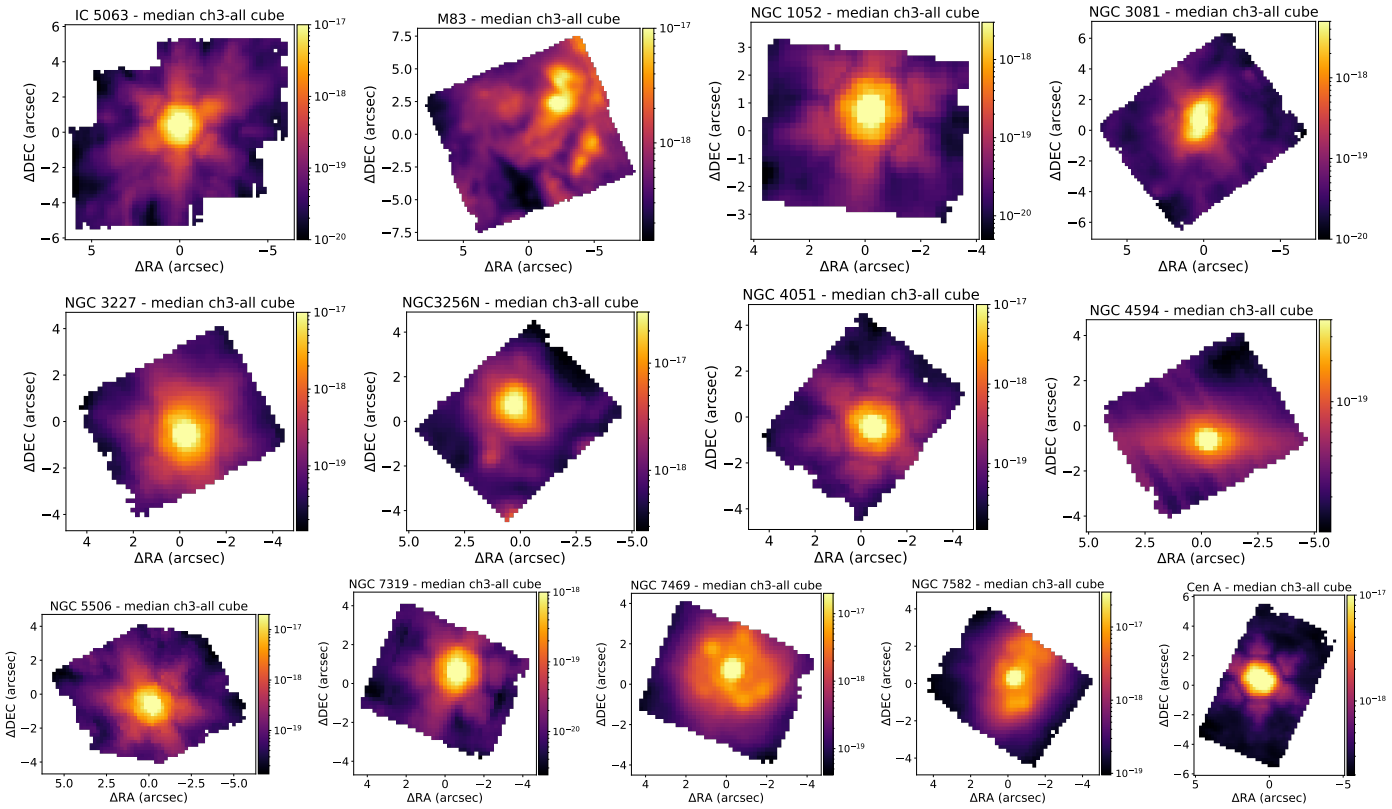

Fig. A.1. Median flux maps of the ch3-all cubes for all the galaxies from the sample, in logaritmic scale (see also the left panels in Figs. 2 and 3).

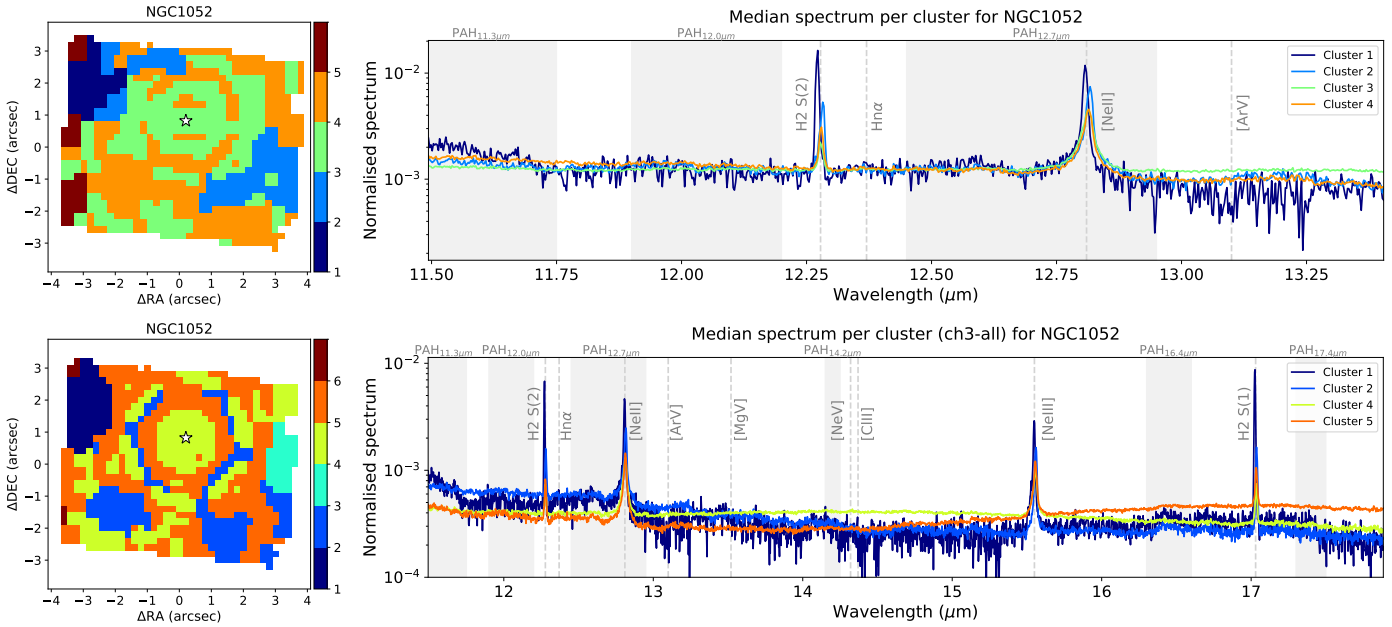

Fig. B.1. Same as Fig. 2 but for NGC 1052. We note that, for the top (bottom) panel, we do not show the spectrum for cluster 5 (3 and 6), as it has low S/N.

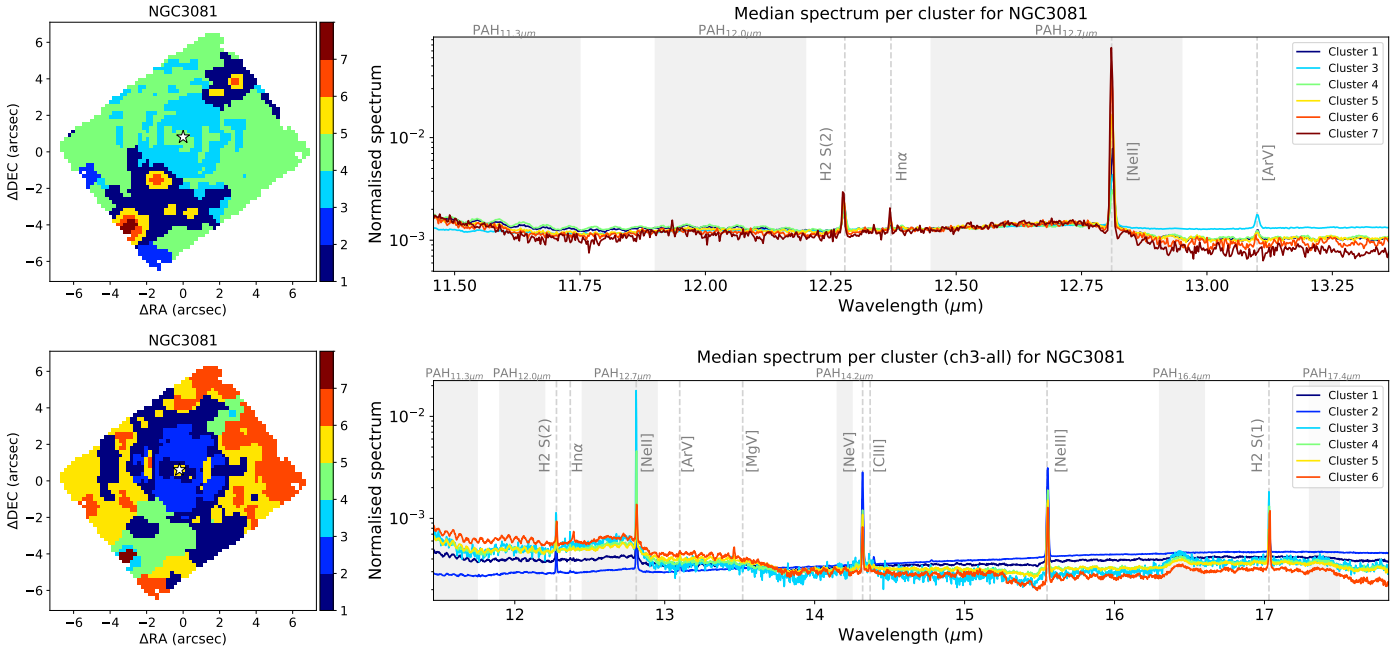

Fig. B.2. Same as Fig. 2 but for NGC 3081. We note that, for the top (bottom) panel, we do not show the spectrum for cluster 2 (7), as it has low S/N.

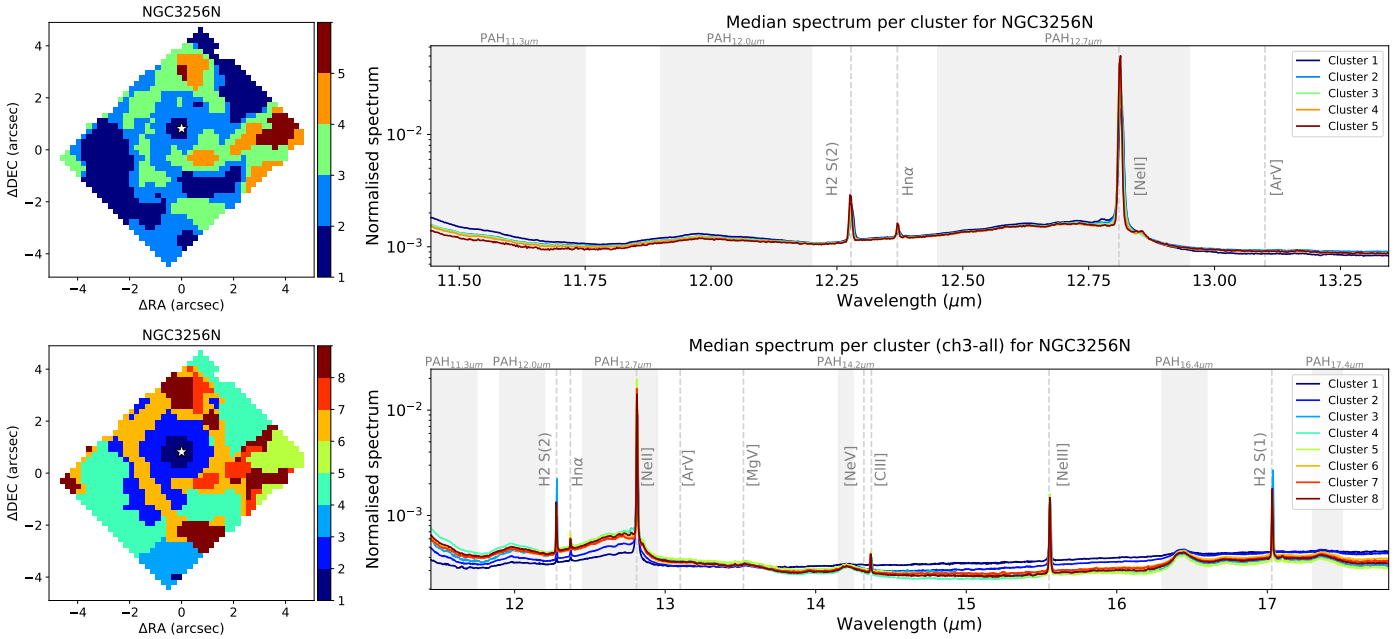

Fig. B.3. Same as Fig. 2 but for NGC 3256-N.

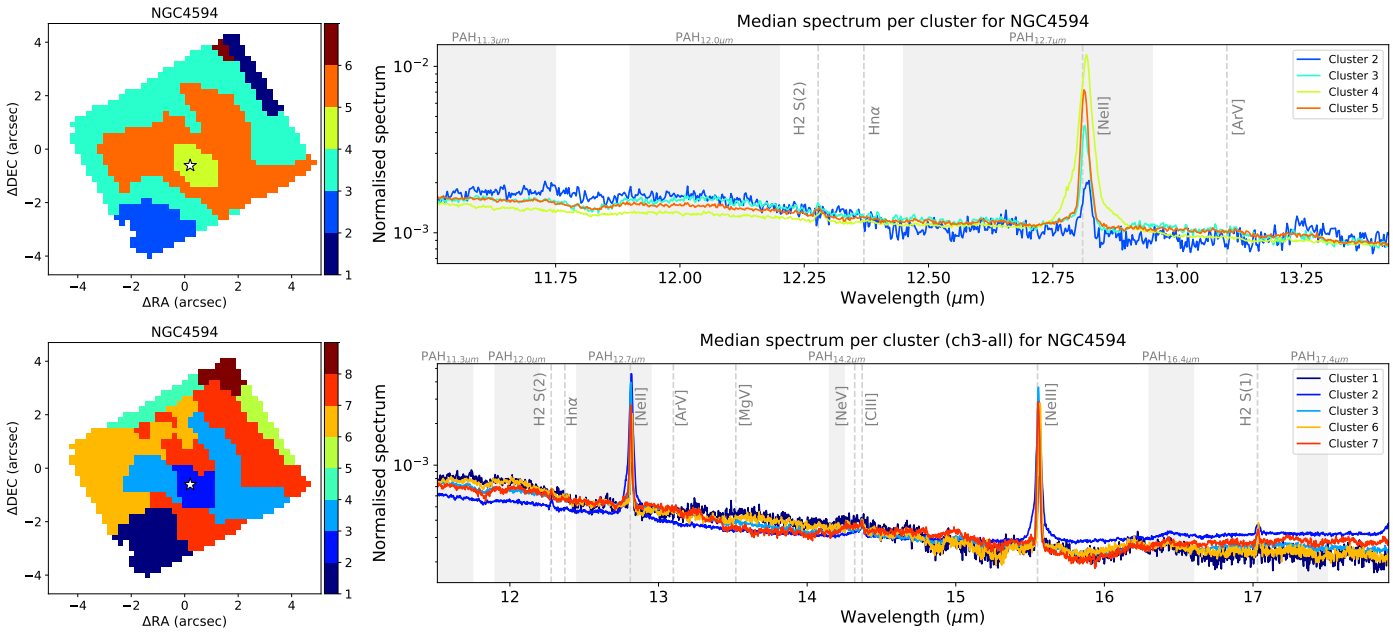

Fig. B.4. Same as Fig. 2 but for NGC 4594. We note that, for the top (bottom) panel, we do not show the spectrum for clusters 1 and 6 (4, 5, and 8), as they are low S/N clusters.

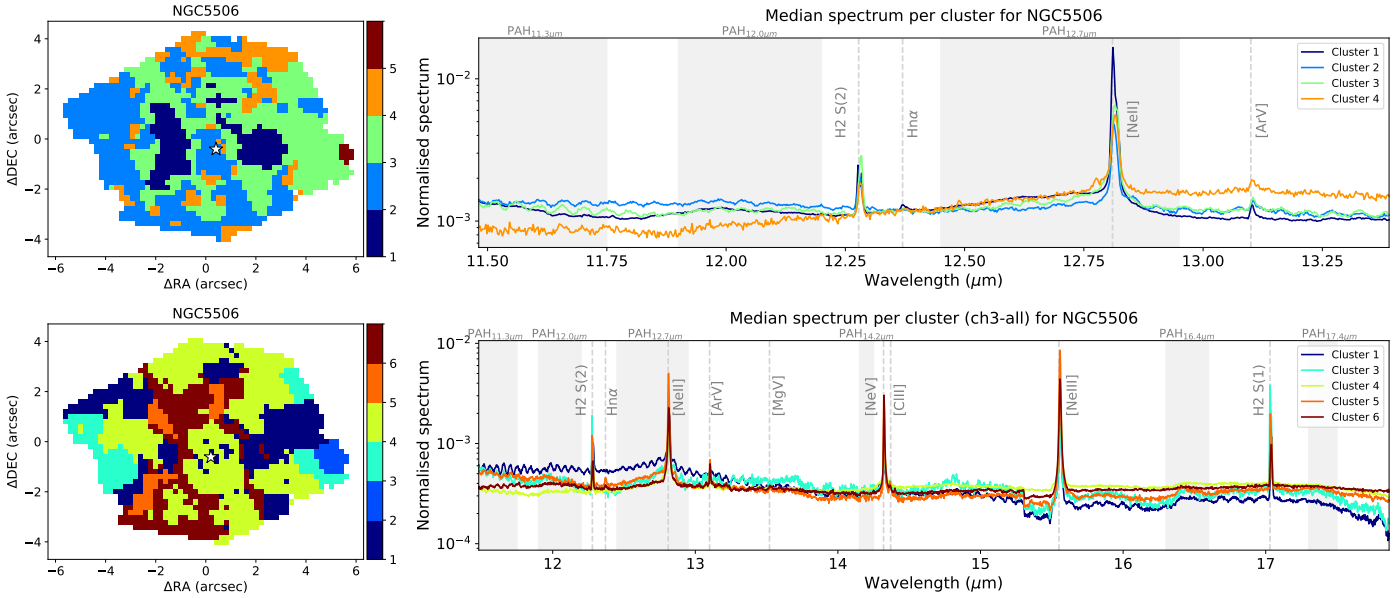

Fig. B.5. Same as Fig. 2 but for NGC 5506. We note that, for the top (bottom) panel, we do not show the spectrum for cluster 5 (2), as it is a low S/N cluster.

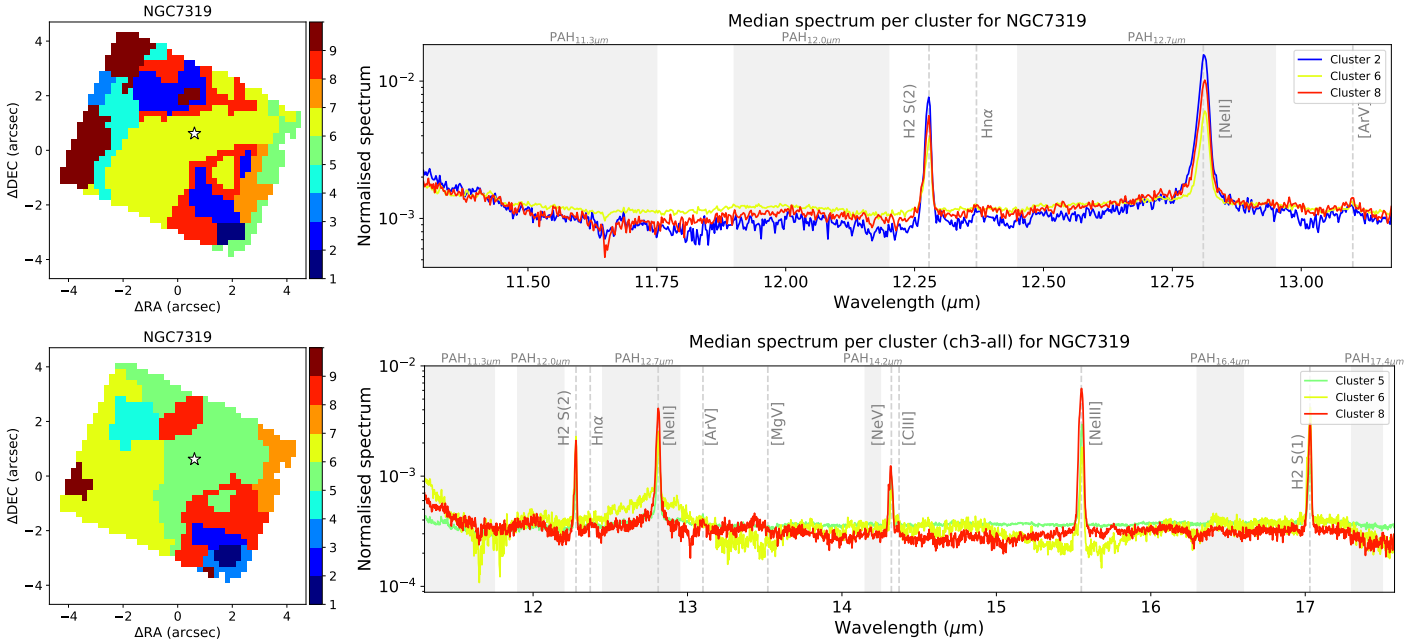

Fig. B.6. Same as Fig. 2 but for NGC 7319. We note that, for the top (bottom) panel, we do not show the spectrum for clusters 1, 3, 4, 5, 7, and 9 (1, 2, 3, 4, 7, and 9), as they are low S/N clusters.

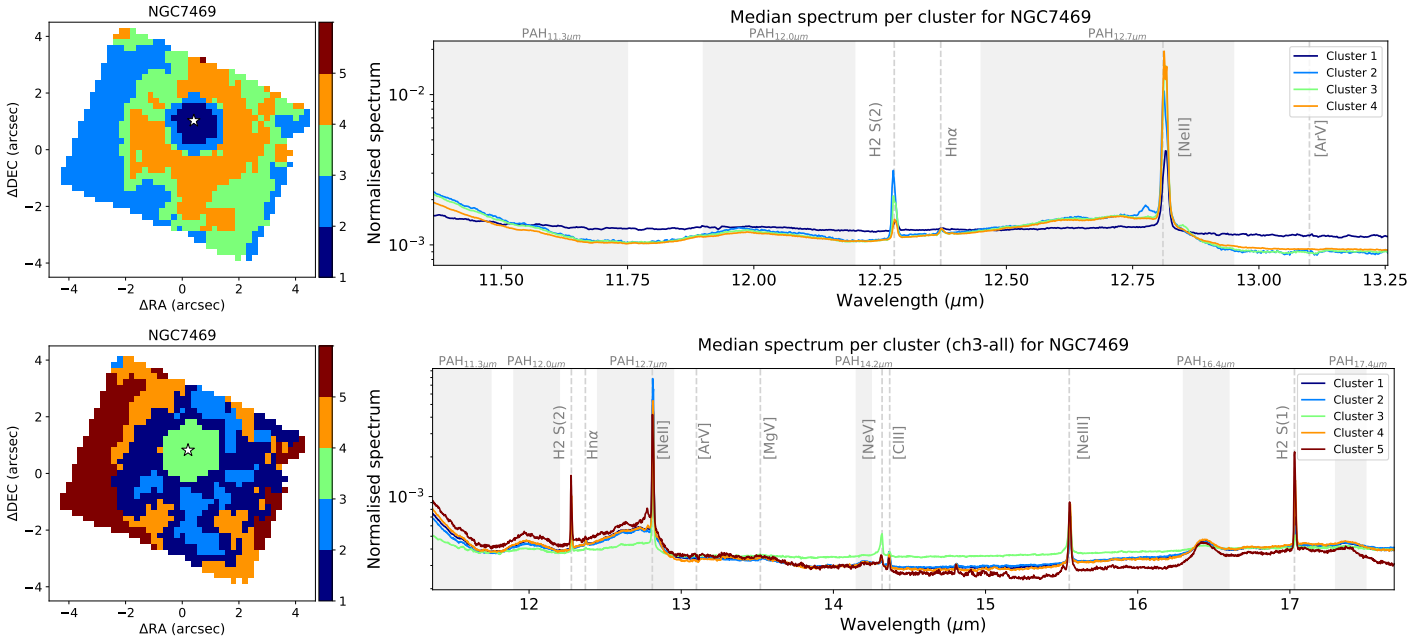

Fig. B.7. Same as Fig. 2 but for NGC 7469. We note that, for the top panel we do not show the spectrum for cluster 5, which is a low S/N cluster.

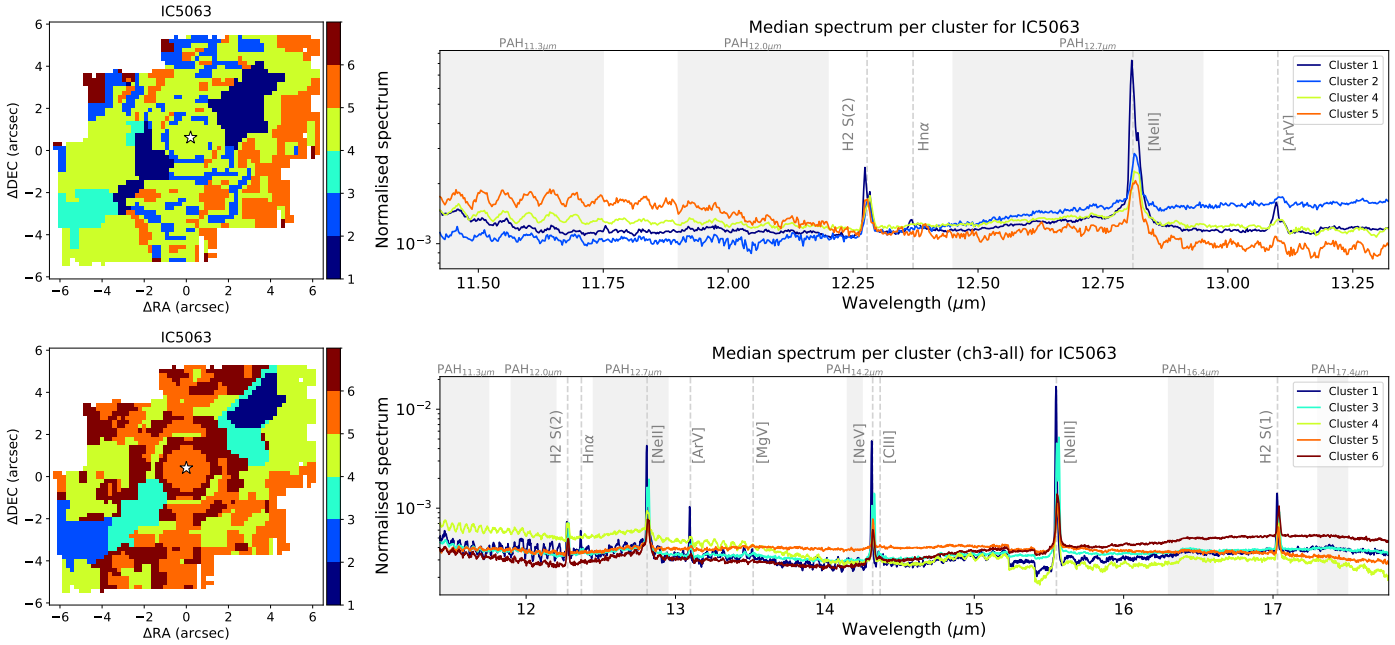

Fig. B.8. Same as Fig. 2 but for IC 5063. We note that, in the top (bottom) panel, we do not show the spectrum for cluster 3 and 6 (2), as they are low S/N clusters.

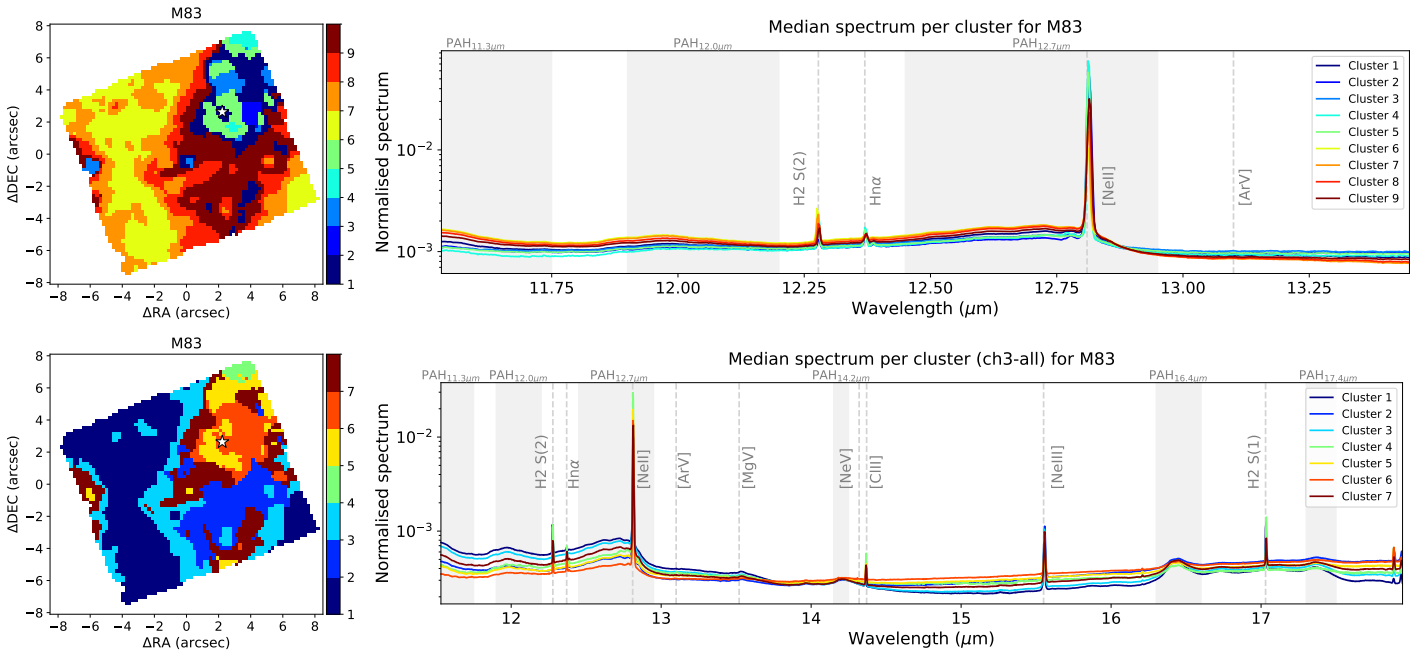

Fig. B.9. Same as Fig. 2 but for M 83. We note that the mid-IR photometric centre does not coincide with the optical one, but is close to the stellar kinematic centre (see Hernandez et al. 2025, and references therein).

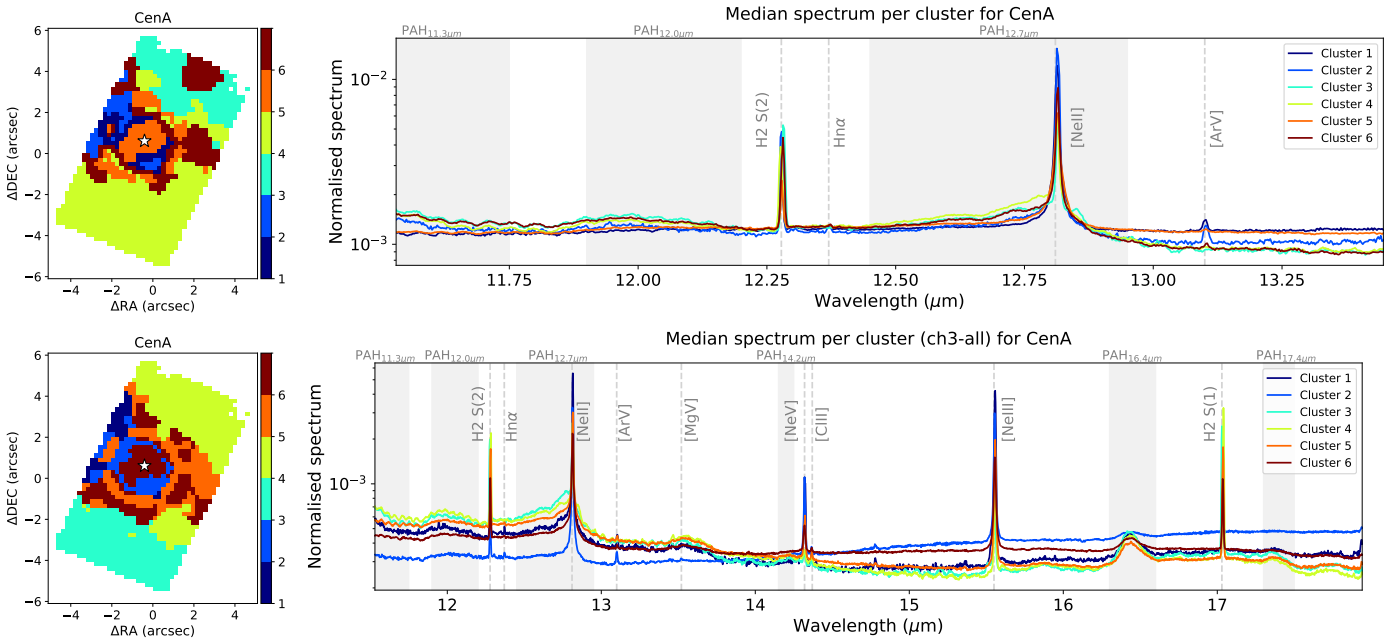

Fig. B.10. Same as Fig. 2 but for Centaurus A.

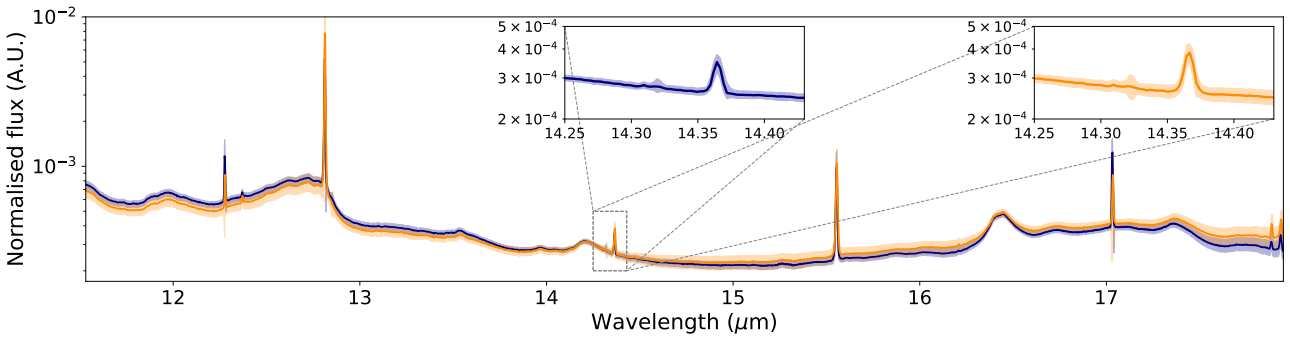

Fig. C.1. Total median spectra for clusters 1 and 3 (blue and orange, respectively) for M 83 (ch3-all), with two insets showing the [Ne V] and [Cl II] lines. The shaded areas represent the uncertainty estimated as the standard deviation of all the spectra within each cluster (see Sect. 2.3).

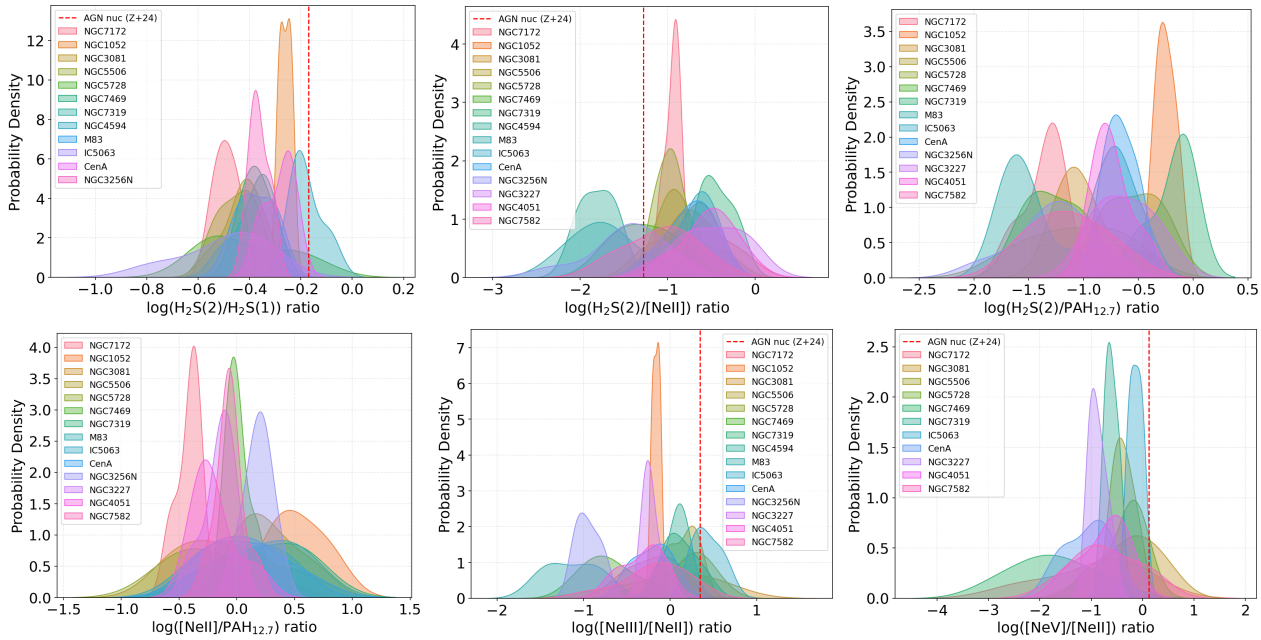

Fig. C.2. Histograms of the line ratios obtained for each cluster per galaxy using the ch3-all cubes. Instead of regular histograms, we use KDEs for visualisation purposes, that compute continuous probability density curves. The red, dashed line marks the median values of the line ratios measured in Sy galaxies with MIRI/MRS by Zhang et al. (2024b) as a reference.

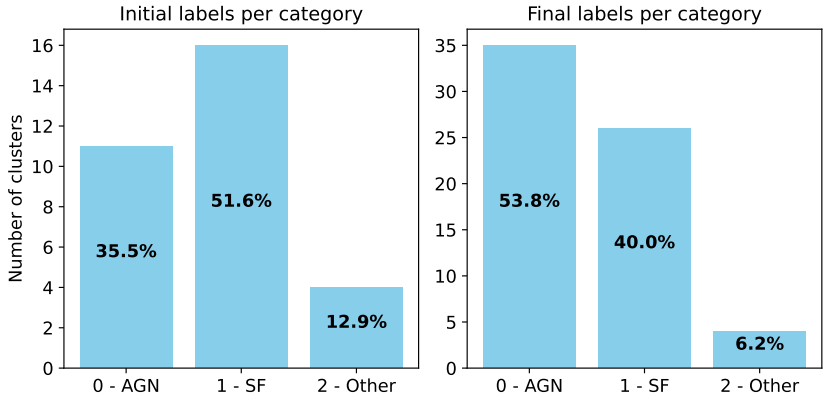

Fig. C.3. Percentage of clusters from the training sample labelled in each category for the inital classification, and after the prediction from the RF model (see also Table B.1).

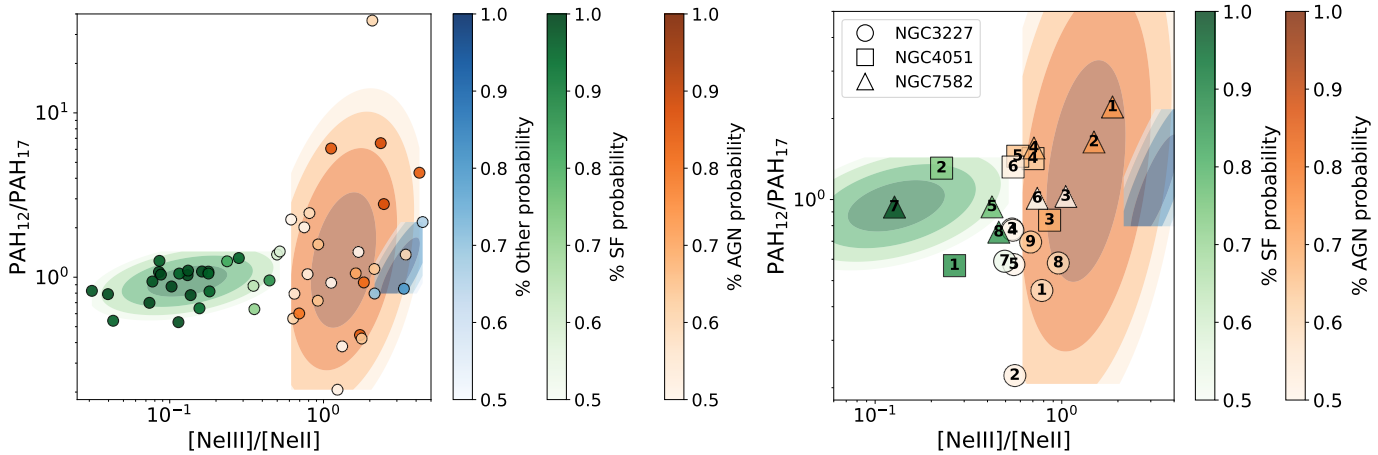

Fig. C.4. Diagnostic diagram in a logarithm scale similar to Figs. 5 and 8, using the PAH ratios derived from the ch3-all cubes for the training and testing samples (left and right, respectively; see Sect. 3.3).