

Virgo filaments

VI. H α clumps in filaments around the Virgo galaxy cluster

G. Nagaraj^{1,*}, P. Jablonka^{1,2}, R. A. Finn³, Y. M. Bahé^{1,4}, F. Combes⁵, G. Castignani⁶, B. Vulcani⁷,
G. Rudnick⁸, D. Zakharova⁹, R. A. Koopmann¹⁰, D. Zaritsky¹¹, and K. Conger⁸

¹ Laboratoire d'astrophysique, Ecole Polytechnique Fédérale de Lausanne (EPFL), Route de la Sorge, 1015 Lausanne, Switzerland

² LIRA, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France

³ Department of Physics and Astronomy, Siena University, 515 Loudon Road, Loudonville, NY 12211, USA

⁴ School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

⁵ Observatoire de Paris, LUX, Collège de France, CNRS, PSL University, Sorbonne University, 75014, Paris, France

⁶ INAF–Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, I-40129 Bologna, Italy

⁷ INAF–Osservatorio astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Italy

⁸ University of Kansas, Department of Physics and Astronomy, 1251 Wescoe Hall Drive, Room 1082, Lawrence, KS 66049, USA

⁹ INAF – Osservatorio Astronomico di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy

¹⁰ Department of Physics & Astronomy, Union College, Schenectady, NY 12308, USA

¹¹ Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA

Received 26 January 2026 / Accepted 4 March 2026

ABSTRACT

The question of which environmental processes operate in filaments still remains unresolved. Given the ubiquity of filaments and their importance in feeding clusters, a proper understanding of these mechanisms is crucial to building a more complete picture of galaxy evolution. Carrying out such an investigation requires access to large galaxy samples with spatially resolved information. As part of this effort, we analysed resolved H α maps of 685 galaxies inside and outside the filaments around the Virgo cluster, in addition to extensive measurements of the integrated physical properties. We created a pipeline to decompose the H α images into individual clumps that trace star-forming regions. We find that the number and average size of clumps in a galaxy are well-defined functions of distance and angular resolution. In particular, the power-law relation between the number of clumps and the distance of a galaxy is consistent with a fractal structure of star-forming regions. We formulated an algorithm to compare filament and non-filament galaxies after removing observational differences. Although we do not have any conclusive evidence of a difference in clump size distributions between filament and non-filament galaxies, we do find that filament galaxies have slightly more peripheral clumps than their non-filament counterparts.

Key words. galaxies: evolution – galaxies: groups: general – galaxies: ISM – galaxies: star formation – galaxies: structure

1. Introduction

Galaxy clusters have long been linked to increased quiescence (e.g. Dressler 1984; Lewis et al. 2002; Wetzel et al. 2012; van der Burg et al. 2020) and early-type fraction (e.g. Dressler 1980; Postman & Geller 1984; Dressler et al. 1997; Vulcani et al. 2023). Although these dense environments have been well studied, the vast majority of galaxies are found outside clusters (e.g. Cautun et al. 2014; Cui et al. 2019). Surveys such as SDSS (York et al. 2000) and 2dF (Colless et al. 2001) have helped formulate the current picture of the large-scale structure of the Universe as a cosmic web filled with long structures known as filaments (e.g. Bond et al. 1996; Kitaura et al. 2009; Darvish et al. 2014).

As a consequence, the research field exploring how galaxy evolution is driven or altered by the environment is steadily shifting from comparing galaxy properties in groups and clusters with those in the field (whose definition is admittedly ill-defined) to considering the position of galaxies in the cosmic web, including the filaments that connect groups and clusters.

Studies have revealed higher early-type fractions (e.g. Kuutma et al. 2017; Castignani et al. 2022a; O’Kane et al. 2024), lower star-forming fractions (e.g. Cybulski et al. 2014; Blue Bird et al. 2020; Castignani et al. 2022a), higher metallicities (Donnan et al. 2022), and smaller star-forming disks (Conger et al. 2025) in filament galaxies. However, it is not clear whether filaments enhance (e.g. Kleiner et al. 2017) or repress HI gas reserves (e.g. Crone Odekon et al. 2018; Luber et al. 2019). Furthermore, filaments are a heterogeneous environment and feature a complex intersection with groups. Hoosain et al. (2024) and Zakharova et al. (2024) reported that while galaxies closer to filament spines are redder and more HI-deficient, the dependence can be largely explained by enhanced group membership in the filament interiors. In addition, most environmental studies focus on quenched galaxies, so the impact of environment on star-forming galaxies is less understood. Overall, it is quite clear that a consensus on the environmental impact of filaments, let alone a full accounting of the underlying physical processes, is still out of reach.

The central question underpinning these investigations relates to how galaxies sustain and eventually quench their star-formation activity, a phenomenon that is also referred to as the

* Corresponding author: gautam.nagaraj@epfl.ch

baryon cycle. A series of seminal studies has revealed the connection between molecular gas and star-forming regions, thanks to the SINGS (Kennicutt et al. 2003), THINGS (Walter et al. 2008), and HERACLES (Leroy et al. 2009). Bigiel et al. (2008) found a linear relationship between SFR and H_2 surface densities (Σ_{SFR} , Σ_{H_2}) at sub-kpc scales. Also, Leroy et al. (2008) showed that the H_2 -to-HI ratio and, by extension, cloud formation, strongly depends on the galactocentric distance.

Built on that heritage, pushing spatial resolution down to ≤ 100 pc scales, the PHANGS collaboration has shown that galaxies with higher stellar mass and more active star formation tend to host molecular gas with higher surface density and higher velocity dispersion (Sun et al. 2020). Rosolowsky et al. (2021) highlighted that the location of clouds is strongly influenced by the presence of stellar bars and spiral arms, providing the first sign of the relation between dynamical (local) environment and star formation conditions. The high-resolution, multi-wavelength journey of PHANGS continues, including PHANGS-MUSE (Emsellem et al. 2022) and PHANGS-HST $H\alpha$ (Chandar et al. 2025), promising to provide constraints on feedback-regulated star formation.

As yet, there is no comparable high-spatial-resolution study placing galaxies within their filament environments. One exception is the work of Vulcani et al. (2019), who presented MUSE observations and spatially resolved properties of four nearby galaxies from the GASP sample (Poggianti et al. 2017), embedded in filaments identified by Tempel et al. (2014). These four galaxies present $H\alpha$ clouds beyond four times their effective radii and generally disturbed $H\alpha$ distributions, which is quite unusual outside a cluster environment. However, it is necessary to assemble a statistically representative sample that will allow us to assess the impact of filaments.

Thanks to its proximity, the Virgo cluster and its associated network of filaments offer a unique opportunity for detailed, spatially resolved studies. The work presented here is set within this context, as part of a series of papers that aims to better understand the full range of environmental conditions around Virgo. The first study, Castignani et al. (2022a), assembled HI 21-cm and CO observations of 245 galaxies and carefully quantified a set of filaments motivated by but independent of Kim et al. (2016) to understand the environmental impact of filaments on atomic and molecular gas. Compared with isolated galaxies from the AMIGA sample (Verdes-Montenegro et al. 2005) and Virgo cluster galaxies from Boselli et al. (2014), they found a clear sequence from field to filament to cluster in terms of decreasing gas content and SFRs. The second study, Castignani et al. (2022b), put together a catalogue of nearly 7000 galaxies inside and around the Virgo cluster with an extensive collection of integrated data and basic physical properties (described in more detail in Sect. 2). Follow-up studies have included the comparison of the gas properties in filaments to results from a semi-analytic model (Zakharova et al. 2024), an analysis of the effects of environment on the relative sizes of star-forming and stellar disks (Conger et al. 2025), and the detailed characterisation of the disruption of the baryon cycle in the group NGC 5363/5364 based on $H\alpha$ and HI data (Finn et al. 2025).

The $H\alpha$ data used here (Finn et al. 2025) are part of a much larger effort to obtain resolved observations of 685 galaxies inside and near the Virgo filaments (Finn et al. in prep.). In this work, we use the full set of $H\alpha$ observations to quantify the morphology of star formation in galaxies inside and outside filaments. In particular, we identify clumps in the $H\alpha$ images, namely, collections of H II regions whose exact sizes depend on image quality and galaxy distance (see Sect. 3.3).

The paper is organized as follows. In Sect. 2, we describe the observations used in this study, including the $H\alpha$ maps and the relevant data from the main Virgo catalogue. In Sect. 3, we detail the process used to quantify the $H\alpha$ clump distribution in galaxies, the steps taken to ensure a clean sample, and the algorithm developed to properly compare clumps in different populations of galaxies. We also discuss the effects of image resolution and galaxy distance on clump sizes. In Sect. 4, we detail the basic properties of filament and non-filament galaxies in our sample for context before moving to the results of the clump analysis in Sect. 5. Then, we provide evidence of the fractal nature of $H\alpha$ clumps in Sect. 6. Finally, we summarize our findings and suggest future directions in Sect. 7. Throughout this work, we assume a Hubble constant of $H_0 = 74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Tully et al. 2008).

2. Parent sample and observations

2.1. Sample selection

Our filament galaxy sample was drawn from a catalogue of 6780 galaxies in the Virgo cluster and its surroundings, including multi-wavelength photometry, positional and environmental information (distance and coordinates, nearest filament, group membership, and local density), and basic integrated properties (stellar mass, morphology, etc., Castignani et al. 2022a,b). The galaxies in the catalogue were assembled on the basis of HyperLEDA (Paturel et al. 2003; Makarov et al. 2014), the NASA Sloan Atlas (Blanton et al. 2011), and the ALFALFA $\alpha 100$ (Haynes et al. 2018) surveys. The respective authors used redshift-independent distances from Steer et al. (2017) when available and computed distances based on details about the cosmic flow from Mould et al. (2000) for the other galaxies.

In terms of environmental parameters, the authors use the position information of the galaxy sample to calculate the fifth-nearest neighbour densities (n_5), used as an estimate of local density on roughly megaparsec scales. The authors built on the filament identification by Kim et al. (2016) to present a list of thirteen filaments around the Virgo cluster, providing parametric forms of their spines. Castignani et al. (2022b) defined galaxies as being inside a filament if they are within $2.70 (2h^{-1})$ Mpc from the filament spine. Group memberships were taken from Kourkchi & Tully (2017). In this study, we used the following parameters: distances, local densities (n_5), stellar masses, Hubble types, r -band elliptical fits (for galaxy size), morphologies, and filament and group assignments.

We combined the information stated above with resolved $H\alpha$ observations of 685 galaxies inside or in the vicinity of Virgo filaments (from the entire sample of 6780). The observations are described below in Sect. 2.2. In the process of observing filament galaxies, we have also been able to obtain $H\alpha$ imaging of many galaxies outside filaments. Approximately a third of our $H\alpha$ sample are outside identified filaments and can be used as a comparison sample.

We show the spatial distribution of the 685 galaxies in the entire $H\alpha$ sample in Fig. 1, along with their filament membership (in a filament or not) and group membership (rich group, poor group, or not in a group). The left panel shows which galaxies belong (or are closest) to which filament, while the right panel highlights the complex 3D structure of the filaments and their constituents. Filaments have a mixture of galaxies in groups and galaxies not in groups, but there are several groups that are clearly centred around filament spines, showing the correlations between these environments. In the entire sample from

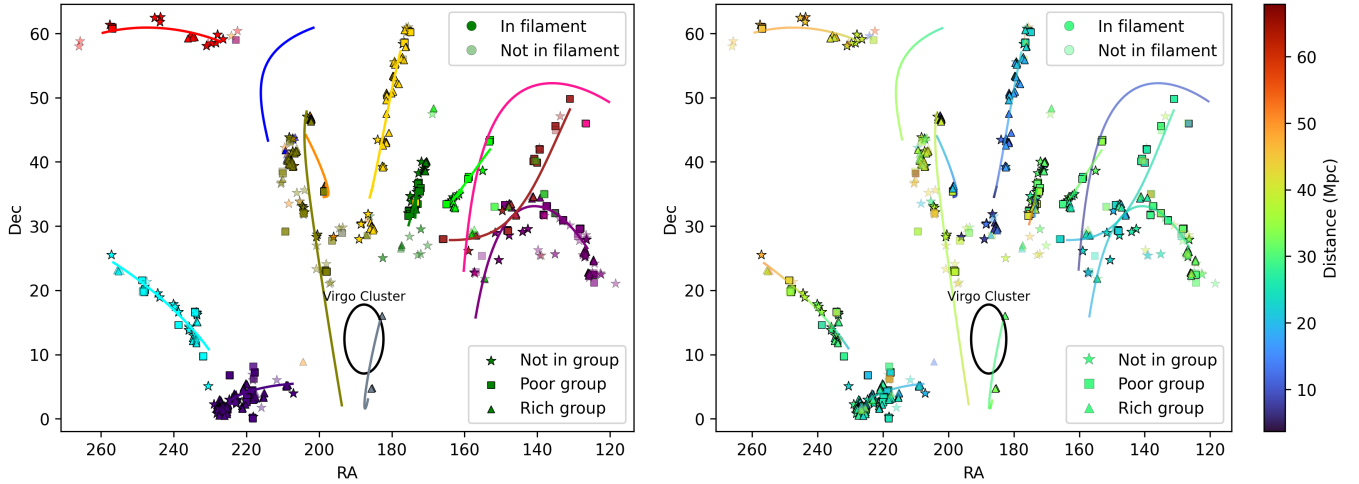


Fig. 1. Positions of all 685 galaxies in the $H\alpha$ sample. Galaxies represented by star symbols are not in galaxy groups, while square symbols are in poor groups and triangle symbols are in rich groups. Galaxies considered inside filaments (under 2.70 Mpc from the spine of a given filament) have opaque symbols and thicker black edges while those outside the filament have translucent symbols and thinner black edges. We have indicated the Virgo cluster as a black circle. The radius corresponds to the virial radius of 1.55 Mpc (McLaughlin 1999). In the left panel, we give the 13 filaments studied in Castignani et al. (2022b) in different colours and use the same colours to mark galaxies as being nearest to a given filament. In the right panel, we colour both the filaments and galaxies by line-of-sight distance to highlight the complex 3D structure determining filament membership.

Castignani et al. (2022b) excluding cluster galaxies (for a total of 5628 galaxies), we find that 61% of filament galaxies are in groups while only 43% of non-filament galaxies are in groups, further highlighting this correlation.

2.2. Observations

Observations were conducted with narrow-band filters on four instruments: the *Isaac Newton* Telescope Wide Field Camera (hereafter referred to as INT), the Steward Observatory Bok Telescope 90Prime (BOK), the WIYN 0.9 m Mosaic 1.1 Camera (MOS), and the WIYN 0.9 m Half-Degree Imager (HDI). In all cases, the narrow-band images are accompanied by r -band images. Details on the observations, including the filter descriptions and telescope details, are described by Finn et al. (in prep.). The data reduction and calibration procedures are reported in Finn et al. (2025) and we summarise the main points here.

The images were reduced using standard methods to subtract the bias and dark current and then flattened using either a dome or sky flat. We used Scamp (Bertin 2010) to solve for the astrometric calibration in each image, using the GAIA-EDR3 as the reference catalogue. The astrometric RMS error with respect to the GAIA-EDR3 reference positions is significantly less than a pixel and is typically 0.05'' or less. We used SWarp (Bertin et al. 2002) to co-add the images. We determined the AB photometric zero point by comparing instrumental magnitudes with the reference r -band magnitude from PAN-STARRS for both the $H\alpha$ and r -band image. We created cutout images of each galaxy and made an accompanying mask that flag pixels not associated with the galaxy. The masks were made by combining a segmentation image from Source Extractor (Bertin & Arnouts 1996) with known positions of GAIA-EDR3 stars. When masking the stars, we adopted the relationship between stellar magnitude and star radius used by the Legacy Survey (e.g. Zhou et al. 2023). The masks were inspected individually to ensure that parts of the galaxy were not masked out.

The $H\alpha$ images contain continuum flux in addition to the line emission. To remove the continuum, we followed the meth-

ods described in Kennicutt et al. (2008) and Boselli et al. (2018). We used a scaled version of the r -band image to estimate the continuum in the $H\alpha$ image. The scale factor depends on the difference in photometric zero points between the $H\alpha$ and r -band images and the colour of the stellar continuum. Therefore, we included a term that scales with the $g-r$ colour variations within each galaxy. The $g-r$ colour dependence is determined by integrating a large library of stellar spectra over the r and $H\alpha$ filters (see Boselli et al. 2018 for details). We constructed a $g-r$ image for each target by re-projecting the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys images (Dey et al. 2019) to match the FOV and pixel scale of the $H\alpha$ image. The colour-corrected continuum image was then subtracted from the $H\alpha$ image.

The full widths at half maximum (FWHM) values of the point spread functions (PSFs) of the $H\alpha$ images range from 1.1'' to 3.6'', with a median of 1.6''. Several galaxies (159 in our clean sample; see Sect. 3.2) have multiple images because of poor weather and/or poor seeing conditions the first time. In our analyses, we consider only the best image (based on FWHM and visual inspection) for galaxies with multiple observations.

3. Methods

We quantified our galaxies' resolved star formation by identifying individual star-forming regions, or 'clumps'. In Sect. 3.1, we detail the algorithm we used to divide the $H\alpha$ images into clumps, with the manual checks to ensure a clean galaxy sample described in Sect. 3.2. There are several clump-finding methods detailed in the literature, including clumpfind (Williams et al. 1994), source-extraction-based (e.g. Bertin & Arnouts 1996) software (e.g. Guo et al. 2015; Mehta et al. 2021), and machine-learning algorithms (e.g. Colombo et al. 2015; Huertas-Company et al. 2020; Adams et al. 2025; Bazzi et al. 2026). Our method uses a combination of wavelet analysis and watershed-type de-blending. Wavelet analysis is useful for decomposing an image into different physical scales, allowing us to efficiently identify clumps.

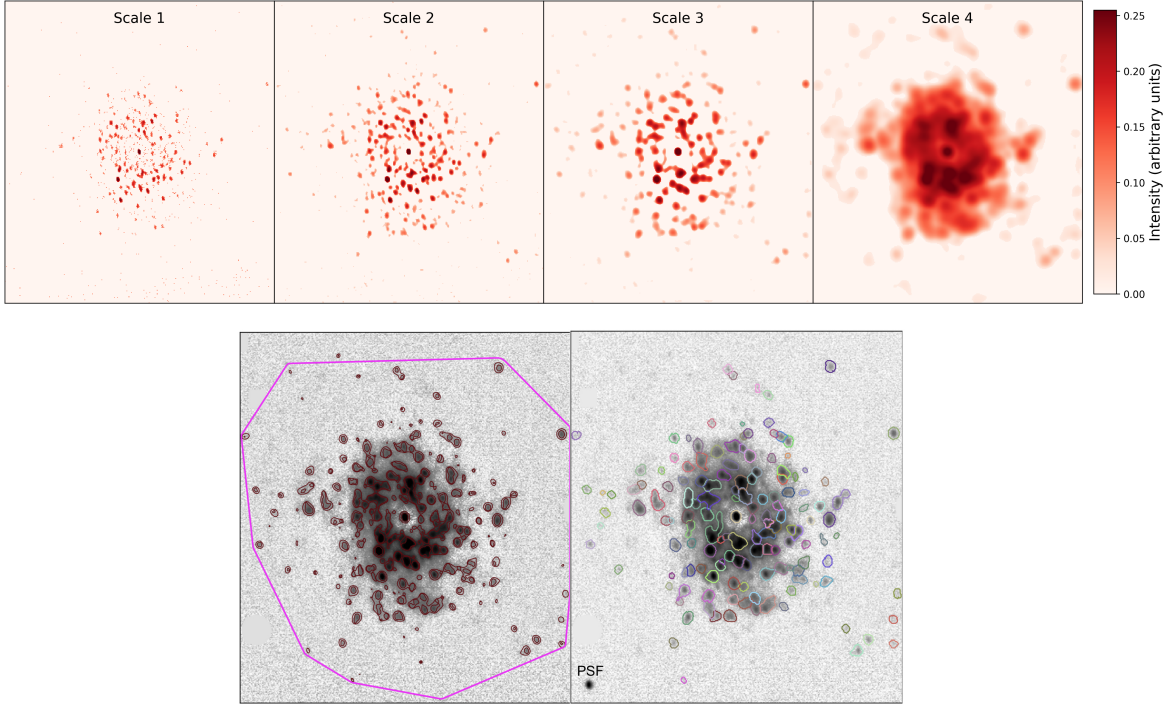


Fig. 2. Illustration of the analysis pipeline for galaxy VFID 1035 (NGC 3982). Top: Decomposition of the $H\alpha$ image into four Scarlet scales. In the bottom panels, the $H\alpha$ image is shown in grayscale (in arcsinh scaling). Bottom-left: Scarlet clump mask in burgundy contours, set to the 99.7th percentile of ($\sim 3\sigma$ above) the noise. Given the seeing FWHM of $1.39''$, this mask consists of both Scarlet scale two and three, visible as small contours within slightly larger contours. The galaxy boundary from the fourth Scarlet scale is marked in the fuchsia curve. Bottom-right: Final clump map after source detection and de-blending by Photutils. Bottom-left corner: Image PSF for size comparison; here, 95% of the clumps are larger than the PSF. See the main text for an extensive description of our analysis pipeline.

3.1. Analysis pipeline

To determine the clump structure of the 685 galaxies, we created an analysis pipeline, illustrated by Fig. 2 for galaxy VFID 1035 (NGC 3982). First, we used the wavelet analysis code Scarlet¹ (Melchior et al. 2018) to define clumpy SF regions in each galaxy. Wavelet analysis is the representation of functions as a particular family of orthonormal series (similar to the Fourier series), with applications for image compression and signal processing. In our case, wavelets enable the identification of features at different physical scales by decomposing the image into various spatial frequencies. In particular, we decompose galaxies into four scales. The first scale corresponds to emission peaks, always smaller than the PSF. The middle scales (second and third) identify regions of enhanced emission within the galaxy. The final scale encloses the entire galaxy, which is useful for marking the entire footprint. We show the decomposition of VFID 1035 into four scales in the top panel of Fig. 2.

Several empirical tests we have performed show that the third scale corresponds to physical clumps. For galaxies seeing $\text{FWHM} \lesssim 1.8''$ (70% of the sample), the second scale identifies smaller, more isolated clumps that are sometimes missed by the third scale. Therefore, our clump mask is defined as the entire region covered by the third scale as well as the second scale for galaxies with $\text{FWHM} \lesssim 1.8''$. When we quantified the clumps (see below), we ignored any emission outside this clump mask. In the bottom-left panel of Fig. 2, we plot the Scarlet clump mask (burgundy contours set at a threshold described later in the pipeline) on top of the VFID 1035 $H\alpha$ image (grayscale, arcsinh scaling). Given this galaxy's seeing FWHM of $1.39''$, the clump

mask consists of scales two and three, visible as small contours within slightly larger ones.

The fourth scale was used to identify the boundaries of each galaxy. This is necessary to establish the limits of the clump mask, avoiding situations where we include extraneous emission that was not properly masked out or caused by image artefacts. In almost all cases, we found (empirically) that including all pixels at or above the 16th percentile of the fourth-Scarlet-scale map perfectly envelopes the galaxy. Out of our 685 galaxies, we need to manually modify this threshold for 25, with percentile values ranging from 30-70. These are images with particularly widespread, low-level diffuse emission (possibly improperly subtracted continuum or brighter dust-scattered $H\alpha$ halos; e.g. Seon & Witt 2012), where the regions bounded by the 16th percentile contours are much larger than the galaxies. In the bottom-left panel of Fig. 2, we show the galaxy boundary for VFID 1035 (16th percentile contour of the fourth scale) as a fuchsia curve.

Although this procedure gives us a clump mask that is constrained to the galaxy footprint, Scarlet does not separate it into individual clumps. To identify the positions, sizes, and shapes of individual clumps, we used the Photutils package (Bradley et al. 2023). First, we ran their source detection algorithm on the clump mask, which is essentially a division of the Scarlet map into discrete units, with a few caveats. A region of an image is considered a clump if it contains at least eight contiguous pixels. In addition, we require that at least 5% of the pixels in a clump are above a galaxy-specific threshold that is defined as follows. We identified the image background as all regions outside the boundaries set using the fourth Scarlet scale. We then defined the 99.7th percentile of the $H\alpha$ 'emission' (noise) in this background map to be the threshold, which is very similar to setting a 3-sigma detection threshold, but the former makes it

¹ <https://pemelchior.github.io/scarlet/>

more stable against outliers. We found, empirically, that the 5% criterion allows us to retain low-surface-brightness features, while avoiding clumps without emission that are occasionally found by wavelet decomposition.

Finally, we allowed for clump deblending. The Photutils package uses a watershed algorithm to divide $H\alpha$ emission regions into individual clumps. We set the following parameters for deblending, chosen through extensive testing to create clump maps consistent with detailed visual inspection. First, we set the minimum fraction of the total flux a component of a clumpy region must contain at 1%. Next, the flux map across the region was divided into 32 linearly spaced levels for finding saddle points to separate components. In addition, pixels were considered contiguous when sharing either an edge or a corner. Finally, we set the minimum number of contiguous pixels for a region to be divided into individual clumps as a number between 8 and 20, depending on the image FWHM (linearly scaled).

In the bottom-right panel of Fig. 2, we plot the final clump selection of VFID 1035 as coloured curves on top of the $H\alpha$ image (grayscale, arcsinh scaling once again). A comparison of the individual clumps to the contours in the bottom-left panel demonstrates the conservation of the Scarlet features, albeit with compound structures typically de-blended into multiple clumps. In the lower left corner, we show the PSF for size comparison. The majority of the clumps (95% for VFID 1035 and 91% for the entire sample of galaxies) are larger than the PSF (size calculated as $\pi(\text{FWHM}/2)^2$). Only one percent of the clumps in the entire sample are less than half the PSF area. As Scarlet does not care about the PSF in this particular application (decomposition into spatial scales) and often reports very asymmetric features, it would be very difficult and even undesirable (given the added biases) to limit the clump-finding algorithm to those larger than the PSF. More importantly, we design stringent algorithms to compare populations of galaxies that take differences in the PSF into account (see Sect. 3.4).

3.2. Visual checks

To ensure a clean sample of $H\alpha$ clumps, we visually inspected each galaxy, where we assigned a general use flag of 0 (unusable) to 2 (good). A use flag of 2 corresponds to galaxies with a clear $H\alpha$ signal, typically with a clumpier structure than the continuum map, indicating a proper continuum subtraction. Galaxies with this flag value do not suffer from any artefacts. Examples of a galaxy that received a 2 in our rating system include VFID 1035, as presented in Fig. 2. The use flag of 1 is given to questionable sources, sometimes with large interloper galaxies partly or fully within the region of the main galaxy that could not be removed or sources with very little $H\alpha$ whose veracity is difficult to judge. Finally, the use flag of 0 is for catastrophic artefacts, complete contamination by a much brighter source, or a complete lack of $H\alpha$. We only used sources with a use flag of 2.

We created other flags to indicate issues of masking of contaminants (foreground stars and other galaxies) covering parts of the galaxy as well as very bright central regions for which the clump-finding algorithm struggled. In addition, we used AGN classification from the HyperLEDA catalogue (Paturel et al. 2003; Makarov et al. 2014) as assembled by Castignani et al. (2022b) to create an AGN flag. We found no major impacts resulting from the flags on our results. In any case, our final sample, with the use flag 2 and without any AGNs or galaxies with masking and/or bright-core issues, contains 414 galaxies.

3.3. Effects of distance and image quality on clump properties

An important question to consider relates the physical significance of our measured clumps. Studies have shown that molecular clouds in the Milky Way and M31 have a median radius of 25 pc and 22 pc, respectively (Miville-Deschênes et al. 2017; Armijos-Abendaño et al. 2025, based on hierarchical clustering algorithms). H II regions in the Milky Way have radii ranging from ~ 1 –20 pc (Tremblin et al. 2014), with smaller sizes reflecting fewer stars and/or generally higher pressures in a given system. On the other hand, Barnes et al. (2026) find a much larger range of ~ 1 –250 pc for H II regions in nearby galaxies, with larger values arising from physical resolution limits of the instruments, especially in cases where de-blending is difficult (e.g. in crowded circumnuclear environments).

Observations of local star-forming galaxies, such as those in the SINGS sample (Kennicutt et al. 2003), yield star-forming regions with sizes of tens to hundreds of pc (e.g. Wisnioski et al. 2012), suggesting that clumps are coherent complexes of molecular clouds. In general, it is clear that resolution plays an important role in determining the sizes and other properties of star-forming regions (e.g. Livermore et al. 2012; Cava et al. 2018). Kollmeier et al. (2017) compared images of star-forming regions at different physical resolutions and show the loss of information on features like shocks and ionisation fronts at resolutions worse than 25 pc/pixel (see their Fig. 9). Recent surveys such as PHANGS-MUSE (Emsellem et al. 2022) and SIGNALS (Rousseau-Nepton et al. 2019) offer a unique opportunity to probe individual H II regions in the Local Universe.

As our Virgo filament galaxies span a large range of distances (6–68 Mpc) and image qualities (PSF FWHMs of $1.1''$ to $3.6''$), the physical resolutions achieved (i.e. FWHM in pc) span a range of ~ 1.2 dex, from ~ 45 –720 pc. Therefore, we need to be able to quantify the evolution of measured clump sizes and numbers as a function of the distance and angular resolution.

To help us with this task, we ran our analysis pipeline on a set of five galaxies from the SINGS dataset (Kennicutt et al. 2003): NGC 2976 (distance of 3.8 Mpc), NGC 3938 (12.7 Mpc), NGC 0024 (8.0 Mpc), NGC 0337 (19.4 Mpc), and NGC 1512 (12.5 Mpc). Their $H\alpha$ images have FWHMs of 0.92 – $2.29''$, resulting in physical resolutions of 17–141 pc. The usefulness of this small sample is threefold: (1) the SINGS galaxy NGC 2976, with a physical resolution of 17 pc, is able to resolve all the way down to large individual H II regions, giving us a more physical anchor to quantify the meaning of our clumps; (2) by using a completely independent sample and ensuring consistency with our Virgo galaxy sample, we demonstrate the robustness of our methodology; and (3) the sample spans a factor of 5.1 in distance and 2.5 in angular resolution, providing a diverse set of probed physical scales.

We conducted an experiment, shown in Fig. 3 for NGC 2976, where we placed the five SINGS galaxies at increasingly large distances (e.g. at factors of twice, thrice, and four times the distance) without changing the angular resolution. The algorithm uses cubic spline interpolation (Scipy module zoom) and conserves surface brightness, while shrinking the galaxy footprint so that the total flux ends up reduced by a factor of distance squared.

Assuming a Gaussian noise profile, we fit a Gaussian curve (with mean fixed at 0) to all of the pixels outside the galaxy mask (based on the fourth Scarlet scale; see Sect. 3.1) in the original image. Thus, we obtained a value for σ to quantify the noise profile. In the new image, everything outside the galaxy

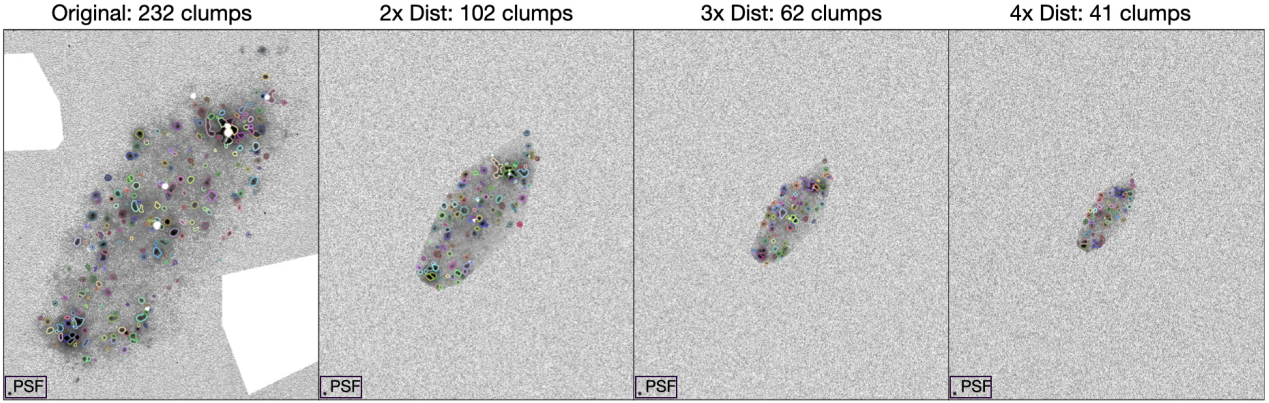


Fig. 3. Clumps detected by our analysis pipeline (coloured circles) in SINGS galaxy NGC 2976 at its original distance (first panel), with the galaxy artificially moved to twice (second panel), then thrice (third panel), and, finally, four times the distance (last panel; see text for details). We can see that while the clumps are always placed in the same general regions of the galaxy, the number of clumps decreases strongly with the distance.

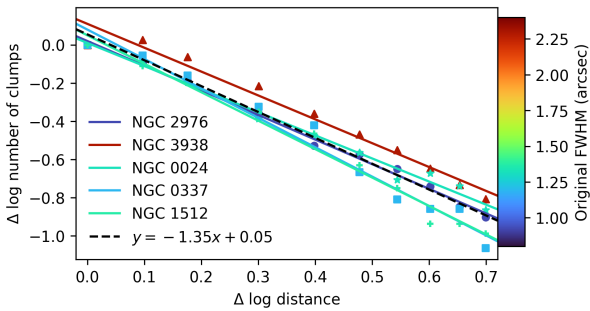


Fig. 4. Dependence of the number of clumps on distance for SINGS galaxies NGC 2976, NGC 3938, NGC 0024, NGC 0337, and NGC 1512, with lines coloured by the original seeing (FWHM) of the galaxy. The change in the number of clumps has a strong negative correlation (median $R^2 = 0.983$) with increased distance. When considering all five galaxies simultaneously, we find a slope of -1.35 ± 0.07 (black dashed line). This suggests that H α clumps are hierarchical (fractal) in nature, with $D \sim 1.3 - 1.4$ (see Sect. 6).

mask (now occupying fewer pixels) is populated with random values taken from the same Gaussian noise profile. This allows us to mimic the galaxy being at a larger distance, but observed with the same telescope and night conditions (since background noise is distance-independent). We can see in Fig. 3 that a couple of regions were masked out in the original image (left-most panel) because of bright interlopers but disappeared in the modified images (other three panels) given the way our algorithm works. These regions have no impact on the measured clump distributions.

It is clear that although the general regions of H α emission that qualify as clumps are the same at all distances, the number of clumps decreases rapidly with increasing distance, with individual clumps generally representing larger physical areas in the galaxy. In Fig. 4, we show the change in the number of clumps measured as a function of distance for the five SINGS galaxies. The points (different markers) show the actual change in number of clumps in each experiment, while the lines are best-fit curves through the points. We find that for each galaxy, there is a strong correlation between the decrease in number of clumps with the increase in distance (median $R^2 = 0.983$). In other words, although the behaviour of the changing number of

clumps with distance is not exactly linear, the best-fit line is a very good approximation.

In Fig. 4, we colour the five SINGS galaxies (points and lines) by their angular resolutions (FWHM in arcsec). We find no clear trend between the slope or intercept and the FWHM, suggesting that the dependence of number of clumps on distance is independent of the FWHM. For all galaxies, we included the point (0, 0) as the original image and number of clumps; however, we did not enforce a y -intercept of 0 in the fit given our interest in measuring an accurate slope. For this reason, the best-fit lines have intercepts close to zero but not exactly zero. The slopes vary between -1.21 (NGC 0024) and -1.54 (NGC 0337), but the overall spread in values is not large (standard deviation $\sigma = 0.13$), indicating a relatively uniform behaviour for a large variety of observing conditions (spanning a factor of 5.1 in distance and 2.5 in angular resolution). This tells us that the number of clumps is a well-behaved function of distance whose relatively small scatter does not depend strongly on specific observational conditions.

A natural corollary of this conclusion is that the results are applicable to our Virgo filament galaxy sample as well. The application of the distance experiment to three Virgo filament galaxies (not shown in the paper) shows this to be the case. Furthermore, the linear behaviour implies self-similarity, which we discuss in Sect. 6. Specifically, we used the best-fit line for all five galaxies with a slope of -1.35 ± 0.07 to estimate the fractal dimension of H II regions to be $D \sim 1.35$ in Sect. 6.

If we were to repeat the same test with the average clump size, rather than number of clumps (not plotted in the paper), we would find strong linear correlations (i.e. power laws), but with positive, steeper slopes. With the same five galaxies, we find that the best-fit line has a slope of 1.79 ± 0.07 (and y -intercept of 0). If clumps at larger distances were simply shown to be combinations of clumps at smaller distances, we would expect a slope of 1.35 (equal to that of the number of clumps, but positive). The fact that we get a significantly steeper slope implies that clumps at larger distances also capture the low-density, low-surface-brightness regions between clumps identified at lower distances. We used this best-fit line to estimate observationally induced differences in clump sizes between populations with different median distances in Sect. 5.1.

The other important aspect affecting clump properties is angular resolution (image quality). In Fig. 5, we present a similar experiment to that shown in Figs. 3 and 4, but with PSF instead

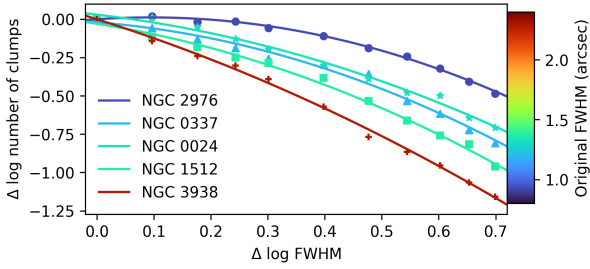


Fig. 5. Number of clumps as a function of angular resolution (FWHM) for the same five SINGS galaxies (NGC 2976, NGC 3938, NGC 0024, NGC 0337, and NGC 1512), with the curves once again coloured by the original seeing (FWHM). The curves are well fitted with quadratic forms (median $R^2 = 0.995$) and have a large scatter that is significantly correlated with the original angular resolution, suggesting a more complex relationship with clump properties than distance that is likely related to the angular scales associated with Scarlet scales two and three.

of distance. For the same five galaxies (NGC 2976, NGC 3938, NGC 0024, NGC 0337, and NGC 1512), we convolve the initial PSF (assumed to be Gaussian) with another Gaussian to make an effectively coarser PSF, up to factors $5\times$ the original FWHM. Specifically, $\Delta \log \text{FWHM}$ is a measure of the logarithmic ratio of the new FWHM (after convolution) to the original FWHM. We find that the number of clumps as a function of the increasing PSF is very well modelled by a quadratic function (median $R^2 = 0.995$), rather than a linear function, showing greater complexity with PSF than distance.

Compared to the distance experiment, we found a much greater scatter between the five galaxies. There is a clear dependence of this scatter on the original angular resolution, although it is not monotonic: galaxies with higher resolution have a flatter response at small $\Delta \log \text{FWHM}$ (with the number of clumps nearly unchanged), while those with lower resolution immediately show a decrease in the number of clumps. We found the same trend, but in the opposite direction for average clump sizes: galaxies with lower resolution have a flatter response at small $\Delta \log \text{FWHM}$, while those with a lower resolution immediately start increasing in terms of the average clump size.

The correlation between the shapes of the curves and the original FWHM is likely an indication of the effects of the Scarlet decomposition scales and Photutils deblending parameters we use: we are effectively probing a minimum angular size given our selection. In the case of NGC 2976 (dark blue curve), an angular resolution of $0.92''$ (which is better than all galaxies in our sample at a minimum resolution of $1.08''$) leads to a large flat response in the curve up to ~ 0.25 dex in $\Delta \log \text{FWHM}$. This means we are almost exclusively probing features larger than the PSF in the original image. Given that the Scarlet scales and Photutils parameters were designed specifically for our Virgo filament sample, this is unsurprising and non-problematic.

The quadratic curves and large variance in response due to differences in the original FWHM highlight the complexities of the effects of PSF on clump properties. Nevertheless, it is also clear that the response is well characterised. Furthermore, 94% of our Virgo filament (and non-filament) sample have angular resolutions of $\text{FWHM} \leq 2.29''$. This means that based on Fig. 5 and the corresponding version with average clump size, we can better understand the effects of PSF on clump properties in our main sample. We performed the PSF experiment on a set of 20 Virgo galaxies and confirmed that the responses are

also quadratic, with the same dependence on original FWHM. We explicitly use the average clump size relations with distance and FWHM to help understand our result in Sect. 5.1.

In this analysis, we have shown that the behaviour of clump sizes and numbers is a well-behaved function of distance and PSF. Furthermore, the effects of distance and PSF are comparable in magnitude: for a very large distance or FWHM increase of $5\times$ ($\Delta \log \text{Quantity} = 0.7$), the number of clumps is decreased by roughly one order of magnitude. Thus, when comparing clump distributions of two different populations, by carefully matching their physical FWHMs (which combine the effects of distance and angular resolution), we can minimise observational biases and learn true differences. We describe our matching algorithm in Sect. 3.4.

3.4. Creation of comparable populations

In Sect. 3.3, we show that physical resolution (FWHM in pc) is essential in determining the meaning of individual clump sizes. As a result, any analysis aimed at understanding clump distributions involving diverse populations of galaxies must account for resolution effects. All of our resolved clump results involve comparing the details about the clumps in filament galaxies versus non-filament galaxies. To ensure that the comparisons we made were strictly related to the difference between the populations, we needed to guarantee that their distributions of physical resolution were nearly identical. The precise details of this procedure are listed below.

1. We calculated the FWHMs of all the galaxy images in physical units (i.e. parsecs) based on the FWHMs in pixel units, pixel scales, and the measured galaxy distances.
2. We divided each galaxy population (of the ones being compared) into equally spaced bins between the range of physical FWHMs occupied by both populations. Galaxies below or above the range were immediately removed. The number of bins is chosen as the floor of the size of the smaller population divided by 10. For example, with 252 filament and 162 non-filament galaxies, we selected 16 bins. The number of bins was chosen empirically to enforce a strong population match, while avoiding very sparsely populated bins.
3. For a pair of two populations (i.e. filament vs non-filament), we equalised the histograms. In each bin, we computed the number of galaxies in that bin divided by the total number of galaxies for each population. Then, we randomly removed galaxies from the population with the larger fraction until the difference is minimised.
4. Since the total number of galaxies changed as we moved through the bins, the fraction of all galaxies in any given bin also changed. For instance, after going through bin 1 out of 16, we ended up with 15 filament galaxies and 10 non-filament galaxies in the bin, along with 240 filament galaxies and 160 non-filament galaxies in the entire sample (all 16 bins). The fraction of filament galaxies within a bin 1 would then be $15/240 = 0.0625$ and the fraction of non-filament galaxies would be $10/160 = 0.0625$, indicating that the populations were equalised. After going through the other 15 bins, there were only 150 filament galaxies and 125 non-filament galaxies remaining in the sample. This means that the bin 1 filament fraction changed to $15/150 = 0.1$, while the non-filament fraction changed to $10/125 = 0.08$, pushing the populations out of equilibrium. Therefore, we needed to repeat the process iteratively, with the final condition asserting that the difference in the fraction of total sources in each bin is the minimum possible value. For

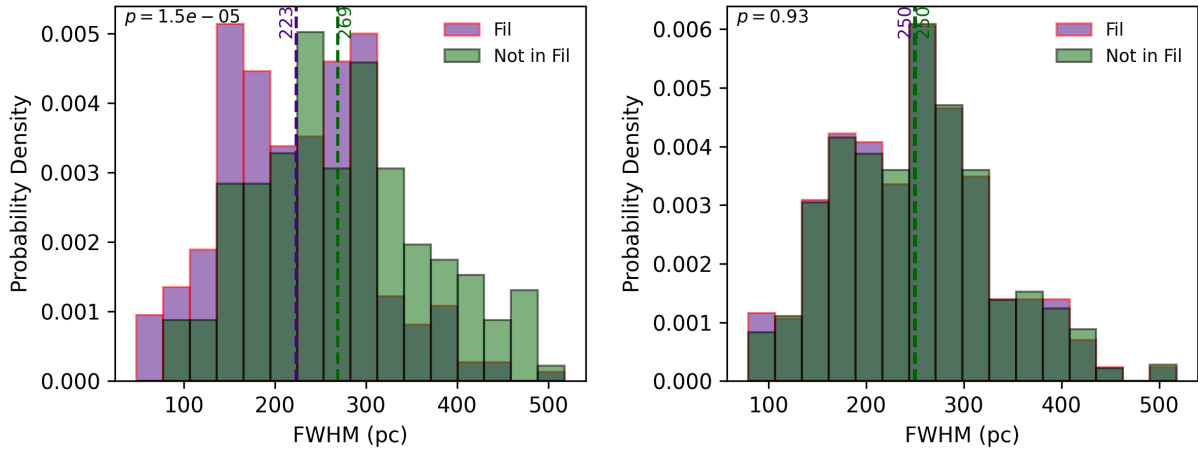


Fig. 6. Demonstration of the FWHM-compatibility procedure. In the left panel, we show the original sample of 252 filament (purple) and 162 non-filament (green) galaxies comprising the full set of 414 galaxies. After removing 95 filament and 29 non-filament galaxies, we are left with the distributions on the right, whose FWHM distributions follow each other very closely. In both panels, the median FWHM of each population is shown as a dotted line. The medians are identical after the procedure.

subsequent iterations, we also allowed the possibility of adding back removed galaxies. This works in the same way as the normal procedure in step 3: in a given bin, if the population with the smaller fraction already had a list of removed galaxies, we added them back in random order until the difference was minimised.

5. Steps 3–4 either removed or added back randomly selected galaxies in each bin until the fractions ended up as equal as possible. Thus, we would get a different set of galaxies each time we conducted the procedure. To provide stable results, any time we performed the matching process, we repeated it 25 times.

To elucidate this process, we show the distribution of uncorrected (left panel) and corrected (right panel) FWHMs of filament versus non-filament galaxies considered in the study in Fig. 6. The full population of 414 galaxies is divided into 252 filament and 162 non-filament galaxies. Using the criteria explained above, we find that the two distributions are initially incompatible: a Kolmogorov-Smirnov (K-S) 2-sample (two-tailed) test yields $p = 1.5 \times 10^{-5}$, suggesting that the distributions are distinct (which is obvious when viewing the figure). Our procedure to equalise the population resulted in the removal of 95 filament and 29 non-filament galaxies. The subsequent distributions are essentially identical ($p = 0.93$).

4. Integrated properties of filament and non-filament galaxies

In previous sections, we detail the creation of a sample of 685 galaxies in and around filaments associated with the Virgo cluster, the detection and measurement of $H\alpha$ clumps, and manual checks to ensure purity, leading to a sub-sample of 414 galaxies, along with the development of an algorithm to unbiasedly compare populations of galaxies. For our filament and non-filament comparison, the matched populations consist of 157 (after removing 95 of the original 262) and 133 (after removing 29 of the original 162) galaxies, respectively, for a final population of 290 galaxies (though in fact there are more than 290 galaxies represented since we always take 25 different iterations of the matching procedure). All subsequent results in the paper feature the matched populations.

In this section, we focus on basic (integrated) properties to provide context for the populations underlying the results in Sect. 5. It is important to note that our samples are not mass-complete, as our goal is to highlight our methodology and compare clump distributions with as many galaxies as possible, and not to make definitive statements on integrated properties. Thus, the conclusions in this section are mostly limited to our sample. In Fig. 7, we show the distributions of stellar mass, Hubble type, local number density n_5 (inverse of the volume occupied by the five nearest neighbouring galaxies), inclination, full $H\alpha$ size, and r -band size for both filament (purple) and non-filament (green) populations.

The $H\alpha$ sizes are defined by converting the total number of pixels within the fourth scale Scarlet map above the 16th percentile, or a different value for $\sim 3\%$ of the population (Sect. 3.1), to a physical value in squared parsecs. All the other quantities are taken from the Castignani et al. (2022b) catalogue. In the case of r -band size, we define it by taking the moment-based elliptical fits (semi-major and semi-minor axes) derived by the 2020 Siena Galaxy Atlas (SGA-2020; Moustakas et al. 2023, available for the Virgo galaxies in the aforementioned catalogue) based on the Legacy Imaging Surveys Data Release 9 (Dey et al. 2019) and calculating area as $A = \pi ab$.

Our galaxies span four orders of magnitude in stellar mass ($\sim 10^7 - 10^{11} M_{\odot}$) and include all morphological types from ellipticals to irregulars. The only property for which there is a strong statistically significant difference between filament galaxies and non-filament galaxies is n_5 . In particular, non-filament galaxies tend to live in lower densities (0.54 dex) than their counterparts. This is simply a statement that the filaments are higher-density environments than non-filaments.

Filament galaxies have a slight tendency to be smaller in $H\alpha$ and r -band size than non-filament galaxies. The histogram also shows different peaks in the size distributions. These peaks (and the median difference) persist through a bootstrap analysis, suggesting they are real. At the same time, many of the largest galaxies in the entire sample are filament galaxies, showing the complexity of the comparison. In any case, the differences are not statistically significant ($\sim 1.5\sigma$ based on the p-value), meaning they are minor at most. We also note that the $H\alpha$ sizes are ~ 0.5 dex smaller than the r -band sizes, but this is complicated to interpret given the difference in size definitions.

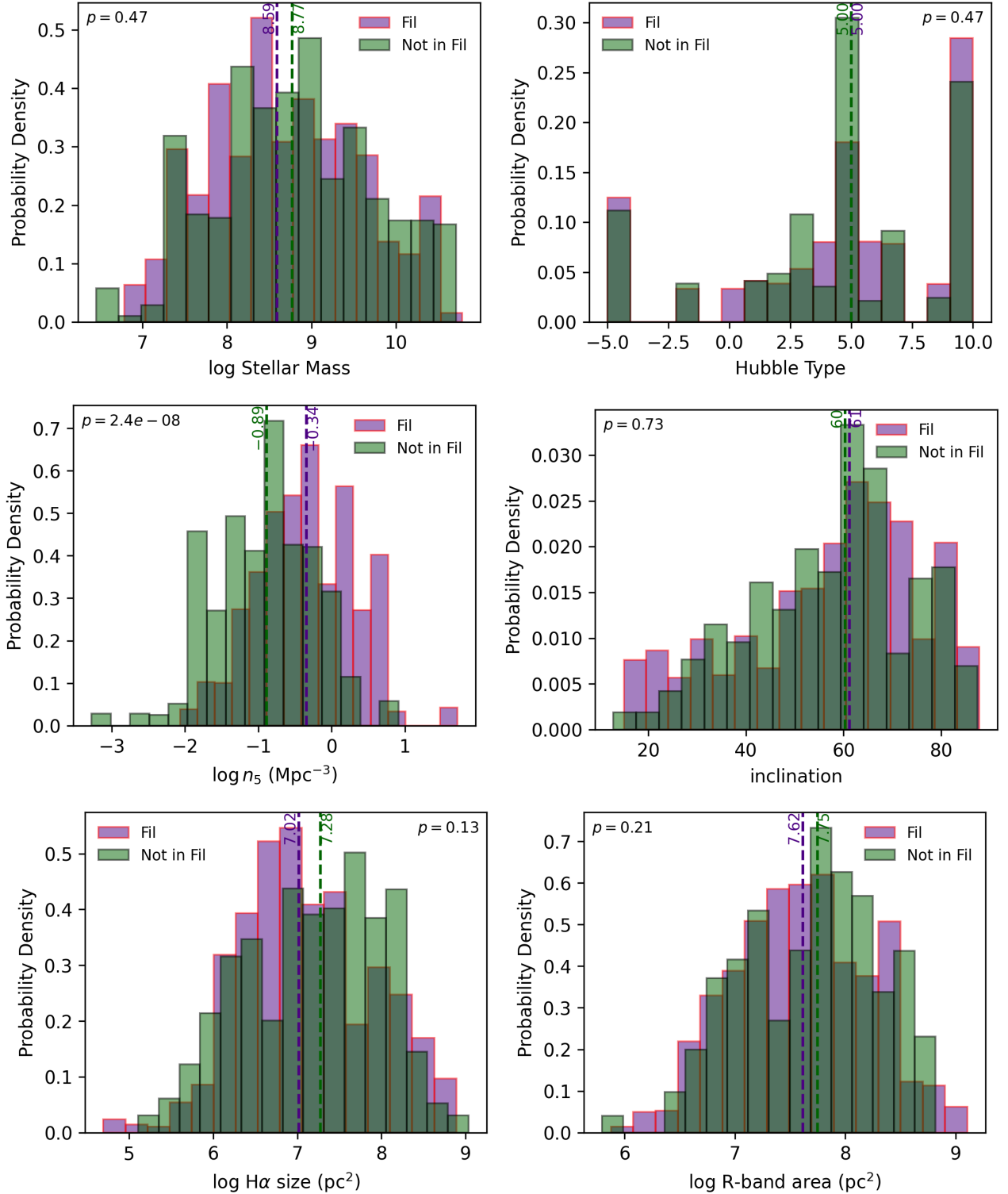


Fig. 7. Comparison of the distribution of various quantities for (PSF-matched) filament (purple) vs non-filament (green) galaxies in our sample (not mass-complete). We show the stellar mass (top-left), Hubble type (top-right), local density n_5 (middle-left), inclination (middle-right; 90° refers to edge-on), H α size (bottom-left), and r -band size (bottom-right). The median values of each distribution are marked as vertical dashed lines (purple for filament and green for non-filament). Our filament galaxies have much higher local densities and possibly smaller sizes than non-filament galaxies, but are otherwise relatively similar in terms of physical properties.

Finally, there are galaxies at all inclinations, both with medians very close to the statistically expected 60° for randomly distributed galaxies. Both the sizes and especially observed positions of clumps depend on the inclination of the galaxy, so

that having statistically indistinguishable inclinations lends confidence to our efforts to make fair comparisons. In the case of clump positions, given the large influence of inclination, we divided galaxies into four inclination bins for a more reliable

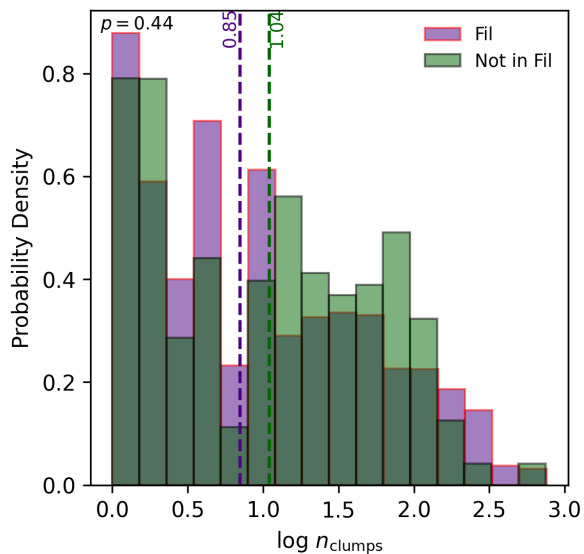


Fig. 8. Same as Fig. 7 but for comparison of the number of clumps. Differences in the distributions are statistically insignificant ($p = 0.35$). Nevertheless, the small differences we can see by eye reflect the differences between the $H\alpha$ size distributions in the bottom-left panel of Fig. 7.

comparison (Sect. 5.2). In general, we find that our filament and non-filament galaxy populations are statistically indistinguishable in all major properties except their environment, which means they are ideal for pinpointing the effects of the filament environment on $H\alpha$ clump properties.

5. Resolved star formation in filament galaxies

We go on to focus on characterising the properties of the clumps. For every clump, we measured three properties: size, flux, and position. In this section, we compare the properties of clump sizes and positions in our filament and non-filament galaxies. As we lack information on dust attenuation, no definitive luminosity measurements can be made. Once again, for all results in this section, we used the FWHM-matching algorithm described in Sect. 3.4 to suppress observational biases.

We first compare the total number of clumps for filament versus non-filament galaxies in Fig. 8. According to a K-S 2-sample test, the distributions of the numbers of clumps are statistically indistinguishable ($p = 0.34$, or a 1σ difference). The small differences we can see by eye (i.e. the lower filament median, and slight overabundances of filament galaxies with 1–12 and over 144 clumps) reflect the same shifts in the $H\alpha$ size distributions in Fig. 7 (bottom-left panel).

5.1. Filament and non-filament galaxies generally have similarly sized clumps

The clump size distribution is strongly dependent on galaxy distance and angular resolution, as shown in Figs. 4 and 5, respectively. Thus, with a galaxy sample whose physical FWHMs vary by over a factor of five, assigning meaning to clump size is particularly difficult. Nevertheless, our careful PSF-matching procedure and the statistically indistinguishable stellar mass distributions of our samples allow us to compare the clump size distributions of our filament and non-filament galaxy populations.

Another important consideration is how clumps from different galaxies are counted. A galaxy with 500 clumps provides 500 times more input than a galaxy with 1 clump in a simple clump stacking (i.e. aggregating the clumps of all galaxies in a given population for the analysis). On the other hand, it would strongly bias our results towards non-clumpy galaxies if the two galaxies were given equal weight, especially since galaxies with very few clumps tend to be farther away and have worse angular resolutions. To mitigate this issue, we randomly selected a certain number of clumps from each galaxy. By choosing a few different numbers, we are able to get a more comprehensive and fair representation of the clump size distributions.

In Fig. 9, we compare clump sizes for filament (F; purple) and non-filament (NF; green) galaxies for four different numbers of clumps: 10, 25, 50, and 100 (from left to right). An auxiliary point of choosing a particular number of clumps from each galaxy is that we must remove galaxies with fewer than that number of clumps from the comparison. After the matching process, the samples compared in Fig. 9 have sizes of 42 (F) and 43 (NF) for $n_{\text{clumps}} \geq 10$, 38 (F) and 33 (NF) for $n_{\text{clumps}} \geq 25$, 20 (F) and 23 (NF) for $n_{\text{clumps}} \geq 50$, and 28 (F) and 10 (NF) for $n_{\text{clumps}} \geq 100$. To be clear, samples with $n_{\text{clumps}} \geq 10$ also includes galaxies with at least 25, 50, and 100 clumps, and so on.

We found no statistically significant (or visually apparent) difference between clump sizes of our filament and non-filament galaxies for the cases of $n_{\text{clumps}} \geq 10$, $n_{\text{clumps}} \geq 25$, and $n_{\text{clumps}} \geq 50$. In the case of $n_{\text{clumps}} \geq 100$, the difference appears (at face value) to be extremely statistically significant, with filament galaxies generally having smaller clumps than non-filament galaxies. In fact, a bootstrap test with the $n_{\text{clumps}} \geq 100$ populations produced the same result. However, as we show below, this difference is likely caused by residual differences in the physical resolutions of the populations, which exist because our non-filament population is very small (ten galaxies).

The most important caveat when considering the strong difference in the $n_{\text{clumps}} \geq 100$ case is the very small number of galaxies in the case of non-filament environments (ten). With such small numbers, the PSF-matching algorithm does not work well (too few bins to minimise differences). We find that the median angular resolution for filament galaxies in this subsample is ~ 0.06 dex sharper (lower) than the median for non-filament galaxies. In addition, the median distance for filament galaxies is ~ 0.07 dex closer than for non-filament galaxies.

In Sect. 3.3, we quantified the effects of distance and angular resolution on clump sizes. For distance, the effect is a simple linear equation: $\log(\text{average size}_{\text{new}}/\text{average size}_{\text{old}}) = 1.79 \log(d_{\text{new}}/d_{\text{old}})$. For angular resolution, the effect is a quadratic function of the logarithmic change in FWHM and a (non-analytic) dependence on the original FWHM. We can approximate $\log(\text{average size}_{\text{new}}/\text{average size}_{\text{old}})$ as a 2D interpolation function of the original FWHM and $\log(\text{FWHM}_{\text{new}}/\text{FWHM}_{\text{old}})$. For galaxies that are closer or have better angular resolution than the reference, we can simply use the relation with the opposite sign, although the assumption of symmetry is possibly flawed in the case of angular resolution.

With these equations, we can estimate the sizes all clumps in the galaxy samples would have at a reference distance and FWHM. By effectively placing all clumps at the same observing conditions, we can try to erase the difficulty caused by the imperfect physical FWHM matching given the small number statistics; however, the approximations do not take into account the scatter in the relations, assuming symmetry between worsening and improving angular resolution. In any case, we placed all clumps in the $n_{\text{clumps}} \geq 100$ samples at a distance of 25 Mpc and an

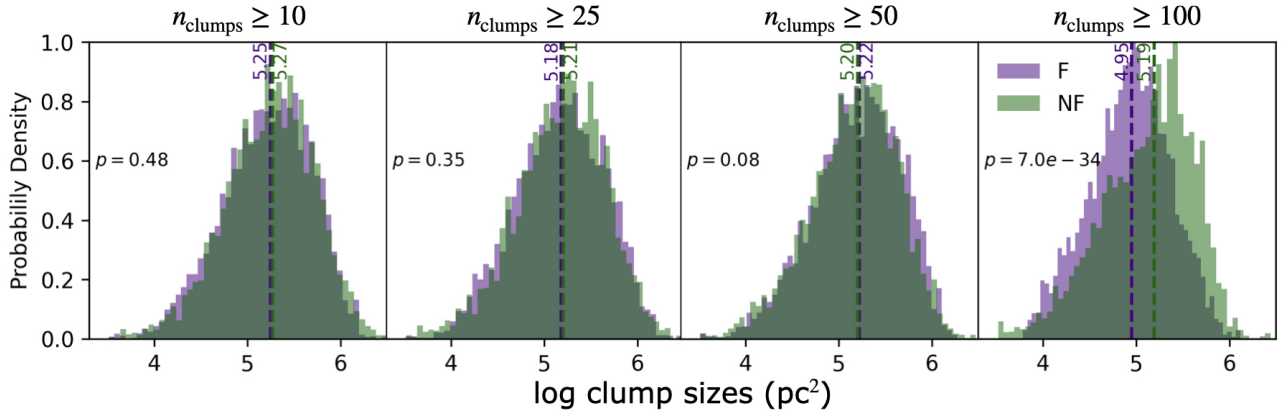


Fig. 9. Distribution of clump sizes in filament (purple) and non-filament (green) galaxies, with medians shown as dashed lines. In each panel, we take galaxies with at least n clumps (n indicated at the top) and randomly select n clumps from the FWHM-matched populations (after the restriction of minimum number of clumps) to prevent biases towards either very clumpy (no weighting) or non-clumpy galaxies (galaxy-weighting). We find no differences in clump sizes until the $n = 100$ case, where filament galaxies generally have smaller clumps than non-filament galaxies. This large difference may be caused by population differences arising from very small number statistics (only ten non-filament galaxies). See text for a detailed discussion.

angular resolution of $\text{FWHM} = 1.5''$. These values are approximately the medians of the galaxies' observing conditions, leading to the minimum possible shift in clump sizes and therefore the minimum possible influence of galaxy variance from the equations used. We find that, in fact, the modified clump distributions are nearly indistinguishable ($p = 0.11$ even considering thousands of clumps). Therefore, we do not have conclusive evidence of a difference in the clump size distributions of filament and non-filament galaxies. A larger sample of clumpy filament and non-filament galaxies, ideally well matched in physical resolution, will be necessary for a more definitive statement.

5.2. Filament galaxies display a slight tendency towards more peripheral clumps than towards non-filament galaxies

One of the primary tasks we can perform with resolved $\text{H}\alpha$ observations is mapping the positions where star formation is occurring. In this work we tabulate the displacement between flux-weighted centroids of each clump and the flux-weighted centre of the galaxy. Once again, the galaxy is defined to be the region of each image within the convex hull of the points where the fourth Scarlet scale is over a threshold (almost always the 16th percentile of the fourth-scale decomposition). Since the galaxies in the sample have a vast range of sizes, to compare the clump distributions, we divided the radial distance of each clump to the centre by the measured semi-major axis in the r -band. The semi-major axes have been measured using elliptical photometry with second-moment maps (kinometry, Krajnović et al. 2006). As stated earlier, they come from SGA-2020 (Moustakas et al. 2023) based on the Legacy Imaging Surveys (Dey et al. 2019). The ratio of clump displacement to semi-major axis is a unitless quantity that represents the relative location of the clumps.

The distribution of clump displacements from a galaxy's centre is strongly dependent on its inclination as clumps along our line of sight can get merged, and for disk galaxies this is much more pronounced for edge-on than face-on configurations. For this reason, we compare only galaxies with similar inclinations. In Fig. 10, we compare the distributions of observed clump displacement (normalised by semi-major axis) in filament (F; purple) and non-filament (NF; green) galaxies. We divide galaxies

into four bins of inclination ($0-40^\circ$, $40-55^\circ$, $55-70^\circ$, and $70-90^\circ$) and perform the FWHM-matching process (25 times) independently in each bin to suppress observational biases. After the matching process, the samples compared in Fig. 10 have sizes of 37 (F) and 18 (NF) for $0-40^\circ$, 13 (F) and 29 (NF) for $40-55^\circ$, 46 (F) and 43 (NF) for $55-70^\circ$, and 38 (F) and 23 (NF) for $70-90^\circ$. We note that all galaxies in our $\text{H}\alpha$ sample have inclination measurements.

The distributions peak around $r/a \sim 0.2 - 0.4$, with the lowest values in nearly edge-on galaxies (where the 3D distance from the centre is most underestimated because of projection issues). We see long tails towards larger values, including peripheral clumps outside the nominal stellar radius. In all galaxies except for nearly edge-on ones, we see a small but statistically significant trend for filament galaxies to have more outwardly and peripherally placed clumps than non-filament galaxies ($p = 1.7 \times 10^{-3} - 2.3 \times 10^{-5}$). Bootstrap experiments with the galaxy samples (where we randomly select samples with the same length as the original but with replacement) also yield the same result, suggesting that the trend is quite stable. To understand the underlying processes causing this trend, we need to build simulations with spatially resolved star formation and multi-phase ISM (e.g. Moreno et al. 2021; Verwilghen et al. 2024) on the large scales of the cosmic web.

In Sect. 5.1, we chose a certain number of clumps from each galaxy via a random selection in order to avoid biasing results towards the galaxies with the most clumps. This restricted the galaxies to those with at least the number of clumps chosen (10, 25, 50, and 100). For the displacement result, we do not adopt the same practice for two reasons: (1) when doing both the inclination bins and the clump number procedure, there are too few galaxies for comparison; and (2) in the case of displacement, we are trying to understand how the clumps trace the $\text{H}\alpha$ distribution in the galaxy, not measure properties of individual clumps. Thus, it makes sense to include all clumps of each galaxy in the computation. Choosing, for example, just 10 clumps from a galaxy with 500 clumps cannot properly sample the $\text{H}\alpha$ distribution. On the other hand, it is much easier to sample the size distribution with fewer clumps.

Nevertheless, when we do restrict the sample with the number of clumps (and do not bin by inclination), we find that the

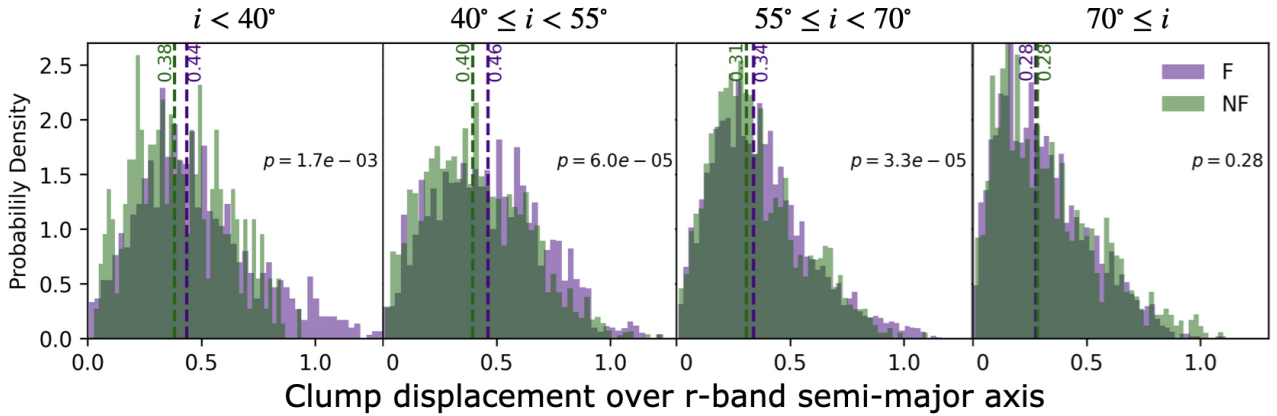


Fig. 10. Observed displacement (normalised by r -band semi-major axis) of clumps in filament (purple) vs non-filament (green) galaxies in four bands of inclination, with medians shown as dashed lines. Between 0° – 70° , we find a small but statistically significant trend for filament galaxies to have more peripheral clumps.

relation in Fig. 10 is not limited to a few $n_{\text{clumps}} \geq 100$ galaxies, as in the size case. Rather, we find that it holds for the $n_{\text{clumps}} \geq 50$, $n_{\text{clumps}} \geq 25$, and $n_{\text{clumps}} \geq 10$ cases. Thus, the danger of biasing towards galaxies with the most clumps is not an issue for the displacement trend.

However, we must caution that the optical size measurements of SGA-2020 are uncertain and that a more robust conclusion will require SGA-2026 elliptical fits (Moustakas et. al, in prep.). Nevertheless, there is no reason for a systematic bias towards filament or non-filament galaxies. Also, the moment-based semi-major axis is a relatively unbiased tracer of galaxy size. In comparison, for example, a surface-brightness-defined axis depends strongly on the image observational depth, which we find in our case to be slightly different between filament and non-filament populations.

6. Clumps observed in fractal patterns

In Sect. 3.3, we describe how we discovered a strong power-law correlation between the number of clumps and distance ($n_{\text{cl}} \propto d^{-1.35}$) in our experiment by artificially shifting galaxies to greater distances. Due to the lack of an associated scale in a power law, this result suggests that $\text{H}\alpha$ clumps are self-similar structures. In other words, in our observations, they behave as fractals would, with well-defined sizes depending on the zoom factor. This signifies that H II regions are not randomly distributed in a galaxy but are rather organised in hierarchical structures. The clumps we observe are higher levels of that hierarchy and depend on the observing conditions. Thus, as long as we carefully match populations with their physical resolutions (Sect. 3.4), we expect to make a physically meaningful comparison.

The most fundamental property of a fractal is its dimension: if the mass of a clump at a scale of R is M , we can define the fractal dimension, D , such that $(M/M_0) \propto (R/R_0)^D$ (Mandelbrot 1983; Pfenninger & Combes 1994). While the physical implications are complex, the fractal behaviour can be driven by various effects (e.g. turbulence, gravity, chemistry, etc.). By comparing our fractal dimension to those measured for other types of clouds, we can discover the connections between different phenomena. With our distance experiment, we were able to estimate the value of the fractal dimension of $\text{H}\alpha$ clumps. Based on the assumptions of true self-similarity (i.e. consistent behaviour across all distance scales) and negligible inter-clump gas mass, we can use Eq. 9 from Pfenninger & Combes (1994) to show that

the absolute value of the slope of the best-fit line in Fig. 4 is equal to the fractal dimension. In other words, our distance experiment suggests that $D \sim 1.3 - 1.4$.

Caicedo-Ortiz et al. (2015) used a similar method (i.e. measuring the log-log slope of the number of boxes to cover the $\text{H}\alpha$ emission of a source vs the box size) to find the fractal dimension of two giant H II regions and find $D \sim 1.4$ in both cases. The consistency in the fractal dimension of their H II regions and the larger $\text{H}\alpha$ clumps in our study provide more evidence of the well-defined hierarchical organisation of $\text{H}\alpha$ emission from scales even smaller than individual clouds to large complexes of clouds.

In addition, Elmegreen et al. (2006) measured the number of clumps as a function of angular resolution for galaxy NGC 628 and find a power-law slope of ~ -1.5 , suggesting $D \sim 1.5$, which is similar to our value. Furthermore, our estimated fractal dimension is the same as that of isophotes of atomic and molecular gas clouds in the Milky Way and nearby galaxies from scales of 30 AU to a few hundred parsecs ($D = 1.3 - 1.4$; e.g. Scalo 1990; Falgarone 1992). These papers determine the fractal dimension through a different method: measuring the correlation between area and perimeters of clouds, with isophotal boundaries. In our work, clump boundaries come from Scarlet wavelet decomposition, so we cannot make a direct comparison. Nevertheless, in all the aforementioned cases, we are measuring fractal dimensions on 2D projected images, so that getting the same result could actually reflect the connection between star formation and gas clouds.

One important consideration is that this simple calculation comes with assumptions and might not reflect the true fractal dimension. Elmegreen & Elmegreen (2001), for example, accounted for the overlap and blending of observed structures (e.g. from projection effects) in experiments with numbers of clumps as a function of PSF (modeled with Gaussian smoothing) by creating models of different fractal dimensions and comparing the resulting projected maps to observations of real galaxies. With similar slopes as ours (compare our Fig. 5 to their Fig. 2), they found that the galaxies are most similar to the $D = 2.3$ fractal model, rather than the $D = 1.3$ model.

7. Conclusions

Using resolved $\text{H}\alpha$ observations of a sample of 685 galaxies in and near the filaments around the Virgo cluster, we systematically analysed the role of filaments on the morphology of

star formation at small scales. We also employed the wealth of integrated data available for our galaxies, including stellar masses, morphologies, distances, and environmental information. We developed a pipeline to identify and quantify star formation clumps in the H α images using wavelet decomposition (Scarlet) and image deblending (Photutils), respectively. Based on stringent visual checks, we reduced the 685 galaxies to a pure sample of 414 galaxies. When comparing clumps of galaxies from different populations, steps must be taken to ensure that differences are physical and not simply the result of observational biases. We have shown that the primary variable determining observed properties of the clump distribution is the physical-scale FWHM, so we designed an algorithm to match FWHM distributions for unbiased comparisons. Finally, we matched populations of 157 filament and 133 non-filament galaxies, for a total of 290 galaxies.

We find that, in general, filament and non-filament galaxies display the same distribution of clump sizes. However, in terms of the radial distribution of clumps, for all galaxies except nearly edge-on ones ($i > 70^\circ$), filament galaxies tend to have slightly more outer clumps than non-filament galaxies with a high statistical significance ($\sim 4\sigma$). Our clump displacement measurements are normalised by stellar size measurements from SGA-2020 (Moustakas et al. 2023), which will come with improvements in SGA-2026. These measurements will allow us to make more robust statements. Further investigation, including simulation work, will be needed to understand the origin of this result.

Through our tests of the effects of distance and angular resolution on clump properties, we have shown that H α clumps are fractal (hierarchical) in nature, not random (homogeneous), consistent with studies of both H II regions (e.g. Caicedo-Ortiz et al. 2015) and atomic and molecular interstellar gas (e.g. Falgarone 1992). Moving towards a comparison of filament and non-filament environments, filament galaxies are clearly in regions of higher local density than their non-filament counterparts. In our sample, they tend to have slightly smaller sizes (both recent-SF/H α and stellar/optical footprints), although this does not amount to highly statistically significant trends.

In general, repeating the analysis presented in this work on higher-angular-resolution H α images is required to make stronger statements about differences in the clump distributions. For example, the Javalambre Photometric Local Universe Survey (J-PLUS; Cenarro et al. 2014, 2019) delivers better and much more consistent seeing (1.1'' in the r -band; Rahna et al. 2025) than what is available for this work. The J-PLUS Northern Galactic Hemisphere footprint covers ten out of the thirteen identified Virgo filaments, with seven having partial or full coverage within the two priority areas. Furthermore, one of the J-PLUS filters (J0660) contains the H α emission line up to $z < 0.017$. This allows us to measure resolved H α fluxes for all Virgo filament galaxies and their nearby non-filament counterparts in the footprint.

The role of the more tenuous and large-scale higher density environments in filaments in affecting star formation has not yet been studied extensively, especially compared to clusters. The present study provides a technical framework to evaluate resolved star formation, applied to a medium-sized sample (414 galaxies) to study the systematic effects of filaments. However, our understanding is far from complete and to supplement the current data, we will need access to both detailed simulations and much larger galaxy samples with better angular resolution, as described above. Spectroscopy would allow us to pinpoint details of not only the youngest stellar populations (e.g. with H α), but also older populations through such features as the

Balmer break and absorption lines to probe the timescales that display the influence of filaments on galaxies.

Several complementary studies are already near completion. For example, an analysis of very high resolution and sensitive MeerKAT H I spectral cubes of galaxies along the Virgo III filament as well as an investigation of the molecular gas in several H I-deficient galaxies with ALMA, NOEMA, and SMA will help illuminate the effects of filaments on the gas components of galaxies. By combining comprehensive integrated properties of galaxies in a wide range of local densities around filaments with resolved observations of H α , CO, and H I, we will be able to better understand the range of physical processes operating in the cosmic web.

Acknowledgements. We thank the anonymous referee for their suggestions to improve the paper, as well as A&A language editor Agnes Monod-Gayraud for tightening the phrasing and grammar of our article. This work was supported by the Swiss National Science Foundation (SNSF) under funding reference 200021_213076 “Galaxy evolution in the cosmic web”. This research has made use of the Astrophysics Data System, funded by NASA under Cooperative Agreement 80NSSC21M00561. This research has made use of the NASA/IPAC Extragalactic Database, which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. RAF gratefully acknowledges support from NSF grants AST-1716657 and AST-2308127 and from a NASA ADAP grant 80NSSC21K0640. YMB acknowledges support from UK Research and Innovation through a Future Leaders Fellowship (grant agreement MR/X035166/1). GHR acknowledges the support of NSF AAG grants 1716690 and 2308126 as well as NASA ADAP grant 80NSSC21K0641. This research was supported by the International Space Science Institute (ISSI) in Bern and the Institute for Fundamental Physics of the Universe (IFPU) in Trieste.

References

- Adams, D., Dickinson, H., Fortson, L., et al. 2025, *ApJ*, **979**, 118
 Armijos-Abendaño, J., Eales, S., & Smith, M. W. L. 2025, *MNRAS*, accepted [arXiv:2512.22698]
 Barnes, A. T., Chandar, R., Kreckel, K., et al. 2026, *A&A*, **706**, A95
 Bazzi, Z., Colombo, D., Bigiel, F., et al. 2026, *A&A*, **706**, A40
 Bertin, E. 2010, *Astrophysics Source Code Library* [record ascl:1010.063]
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
 Bertin, E., Mellier, Y., Radovich, M., et al. 2002, *ASP Conf. Ser.*, **281**, 228
 Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, **136**, 2846
 Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, *AJ*, **142**, 31
 Blue Bird, J., Davis, J., Luber, N., et al. 2020, *MNRAS*, **492**, 153
 Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, **380**, 603
 Boselli, A., Cortese, L., & Boquien, M. 2014, *A&A*, **564**, A65
 Boselli, A., Fossati, M., Ferrarese, L., et al. 2018, *A&A*, **614**, A56
 Bradley, L., Sipőcz, B., Robitaille, T., et al. 2023, <https://doi.org/10.5281/zenodo.7946442>
 Caicedo-Ortiz, H. E., Santiago-Cortes, E., López-Bonilla, J., & Castañeda, H. O. 2015, *J. Phys. Conf. Ser.*, **582**, 012049
 Castignani, G., Combes, F., Jablonka, P., et al. 2022a, *A&A*, **657**, A9
 Castignani, G., Vulcani, B., Finn, R. A., et al. 2022b, *ApJS*, **259**, 43
 Cautun, M., van de Weygaert, R., Jones, B. J. T., & Frenk, C. S. 2014, *MNRAS*, **441**, 2923
 Cava, A., Schaerer, D., Richard, J., et al. 2018, *Nat. Astron.*, **2**, 76
 Cenarro, A. J., Moles, M., Marín-Franch, A., et al. 2014, *SPIE Conf. Ser.*, **9149**, 914911
 Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, *A&A*, **622**, A176
 Chandar, R., Barnes, A. T., Thilker, D. A., et al. 2025, *AJ*, **169**, 150
 Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, **328**, 1039
 Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, *MNRAS*, **454**, 2067
 Conger, K., Rudnick, G., Finn, R. A., et al. 2025, *ApJ*, **978**, 113
 Crone Odekon, M., Hallenbeck, G., Haynes, M. P., et al. 2018, *ApJ*, **852**, 142
 Cui, W., Knebe, A., Libeskind, N. I., et al. 2019, *MNRAS*, **485**, 2367
 Cybulski, R., Yun, M. S., Fazio, G. G., & Gutermuth, R. A. 2014, *MNRAS*, **439**, 3564
 Darvish, B., Sobral, D., Mobasher, B., et al. 2014, *ApJ*, **796**, 51
 Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168
 Donnan, C. T., Tojeiro, R., & Kraljic, K. 2022, *Nat. Astron.*, **6**, 599
 Dressler, A. 1980, *ApJ*, **236**, 351

- Dressler, A. 1984, [ARA&A](#), **22**, 185
- Dressler, A., Oemler, A. J., Couch, W. J., et al. 1997, [ApJ](#), **490**, 577
- Elmegreen, B. G., & Elmegreen, D. M. 2001, [AJ](#), **121**, 1507
- Elmegreen, B. G., Elmegreen, D. M., Chandar, R., Whitmore, B., & Regan, M. 2006, [ApJ](#), **644**, 879
- Emsellem, E., Schinnerer, E., Santoro, F., et al. 2022, [A&A](#), **659**, A191
- Falgarone, E. 1992, [IAU Symp.](#), **150**, 159
- Finn, R. A., Rudnick, G., Jablonka, P., et al. 2025, [ApJ](#), **985**, 81
- Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, [ApJ](#), **800**, 39
- Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, [ApJ](#), **861**, 49
- Hoosain, M., Blyth, S.-L., Skelton, R. E., et al. 2024, [MNRAS](#), **528**, 4139
- Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, [MNRAS](#), **499**, 814
- Kennicutt, R. C., Jr., Armus, L., Bendo, G., et al. 2003, [PASP](#), **115**, 928
- Kennicutt, R. C., Jr., Lee, J. C., Funes, J. G., et al. 2008, [ApJS](#), **178**, 247
- Kim, S., Rey, S.-C., Bureau, M., et al. 2016, [ApJ](#), **833**, 207
- Kitaura, F. S., Jasche, J., Li, C., et al. 2009, [MNRAS](#), **400**, 183
- Kleiner, D., Pimblet, K. A., Jones, D. H., Koribalski, B. S., & Serra, P. 2017, [MNRAS](#), **466**, 4692
- Kollmeier, J. A., Zasowski, G., & Rix, H. W. 2017, ArXiv e-prints [arXiv:1711.03234]
- Kourkchi, E., & Tully, R. B. 2017, [ApJ](#), **843**, 16
- Krajnović, D., Cappellari, M., de Zeeuw, P. T., & Copin, Y. 2006, [MNRAS](#), **366**, 787
- Kuutma, T., Tamm, A., & Tempel, E. 2017, [A&A](#), **600**, L6
- Leroy, A. K., Walter, F., Brinks, E., et al. 2008, [AJ](#), **136**, 2782
- Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, [AJ](#), **137**, 4670
- Lewis, I., Balogh, M., De Propris, R., et al. 2002, [MNRAS](#), **334**, 673
- Livermore, R. C., Jones, T., Richard, J., et al. 2012, [MNRAS](#), **427**, 688
- Luber, N., van Gorkom, J. H., Hess, K. M., et al. 2019, [AJ](#), **157**, 254
- Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, [A&A](#), **570**, A13
- Mandelbrot, B. B. 1983, [The fractal geometry of nature /Revised and enlarged edition/](#)
- McLaughlin, D. E. 1999, [ApJ](#), **512**, L9
- Mehta, V., Scarlata, C., Fortson, L., et al. 2021, [ApJ](#), **912**, 49
- Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, [Astron. Comput.](#), **24**, 129
- Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, [ApJ](#), **834**, 57
- Moreno, J., Torrey, P., Ellison, S. L., et al. 2021, [MNRAS](#), **503**, 3113
- Mould, J. R., Huchra, J. P., Freedman, W. L., et al. 2000, [ApJ](#), **529**, 786
- Moustakas, J., Lang, D., Dey, A., et al. 2023, [ApJS](#), **269**, 3
- O’Kane, C. J., Kuchner, U., Gray, M. E., & Aragón-Salamanca, A. 2024, [MNRAS](#), **534**, 1682
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, [A&A](#), **412**, 45
- Pfenniger, D., & Combes, F. 1994, [A&A](#), **285**, 94
- Poggianti, B. M., Moretti, A., Gullieuszik, M., et al. 2017, [ApJ](#), **844**, 48
- Postman, M., & Geller, M. J. 1984, [ApJ](#), **281**, 95
- Rahna, P. T., Akhlaghi, M., López-Sanjuan, C., et al. 2025, [A&A](#), **695**, A200
- Rosolowsky, E., Hughes, A., Leroy, A. K., et al. 2021, [MNRAS](#), **502**, 1218
- Rousseau-Nepton, L., Martin, R. P., Robert, C., et al. 2019, [MNRAS](#), **489**, 5530
- Scalo, J. 1990, [Astrophys. Space Sci. Libr.](#), **162**, 151
- Seon, K.-I., & Witt, A. N. 2012, [ApJ](#), **758**, 109
- Steer, I., Madore, B. F., Mazzarella, J. M., et al. 2017, [AJ](#), **153**, 37
- Sun, J., Leroy, A. K., Schinnerer, E., et al. 2020, [ApJ](#), **901**, L8
- Tempel, E., Kipper, R., Saar, E., et al. 2014, [A&A](#), **572**, A8
- Tremblin, P., Anderson, L. D., Didelon, P., et al. 2014, [A&A](#), **568**, A4
- Tully, R. B., Shaya, E. J., Karachentsev, I. D., et al. 2008, [ApJ](#), **676**, 184
- van der Burg, R. F. J., Rudnick, G., Balogh, M. L., et al. 2020, [A&A](#), **638**, A112
- Verdes-Montenegro, L., Sulentic, J., Lisenfeld, U., et al. 2005, [A&A](#), **436**, 443
- Verwilghen, P., Emsellem, E., Renaud, F., et al. 2024, [A&A](#), **687**, A53
- Vulcani, B., Poggianti, B. M., Moretti, A., et al. 2019, [MNRAS](#), **487**, 2278
- Vulcani, B., Poggianti, B. M., Gullieuszik, M., et al. 2023, [ApJ](#), **949**, 73
- Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, [AJ](#), **136**, 2563
- Wetzell, A. R., Tinker, J. L., & Conroy, C. 2012, [MNRAS](#), **424**, 232
- Williams, J. P., de Geus, E. J., & Blitz, L. 1994, [ApJ](#), **428**, 693
- Wisnioski, E., Glazebrook, K., Blake, C., et al. 2012, [MNRAS](#), **422**, 3339
- York, D. G., Adelman, J., Anderson, J. E., Jr, et al. 2000, [AJ](#), **120**, 1579
- Zakharova, D., Vulcani, B., De Lucia, G., et al. 2024, [A&A](#), **690**, A300
- Zhou, R., Dey, B., Newman, J. A., et al. 2023, [AJ](#), **165**, 58