

COSMOS-Web galaxy groups: Evolution of red sequence and quiescent galaxy fraction

Greta Toni^{1,2,3,*}, Matteo Maturi^{3,4}, Gianluca Castignani², Lauro Moscardini^{1,2,5}, Ghassem Gozalias^{6,7}, Alexis Finoguenov⁷, Sina Taamoli⁸, Fabrizio Gentile^{9,2}, Hollis B. Akins¹⁰, Rafael C. Arango-Toro¹¹, Caitlin M. Casey^{12,13}, Nicole E. Drakos¹⁴, Andreas L. Faisst¹⁵, Carter Flayhart¹⁶, Maximilien Franco⁹, Ali Hadi⁸, Aryana Haghjoo⁸, Santosh Harish¹⁶, Hossein Hatamnia⁸, Olivier Ilbert¹¹, Shuowen Jin^{13,17}, Jeyhan S. Kartaltepe¹⁶, Ali Ahmad Khostovan^{16,18}, Anton M. Koekemoer¹⁹, Gavin Leroy²⁰, Georgios E. Magdis^{13,17,21}, Henry Joy McCracken²², Jed McKinney¹⁰, Louise Paquereau²², Jason Rhodes²³, R. Michael Rich²⁴, Brant E. Robertson²⁵, Rasha M. Samir²⁶, Diana Scognamiglio²³, Samaneh Shamyati⁸, Marko Shuntov^{13,21,27}, and Jorge A. Zavala²⁸

(Affiliations can be found after the references)

Received 10 September 2025 / Accepted 4 January 2026

ABSTRACT

Aims. We investigate the redshift evolution and group richness dependence of the quiescent galaxy fraction and red sequence (RS) parameters in COSMOS galaxy groups, spanning a wide redshift range, from $z = 0$ to $z = 3.7$.

Methods. We analyzed the deep and well-characterized sample of groups recently detected with the AMICO algorithm in the COSMOS(-Web) field. Our study of the quiescent galaxy population is based on a machine-learning classification tool based on rest-frame magnitudes. The algorithm learns from several traditional methods to estimate the probability of a galaxy being quiescent, achieving high precision and recall. Starting from this classification, we computed quiescent galaxy fractions within groups via two methods: one based on the membership probabilities provided by AMICO, which rely on an analytical model, and another using a model-independent technique. We then detected the RS by estimating the ridgeline position using probability-weighted photometric data, followed by σ clipping to remove outliers. This analysis was performed using both rest-frame magnitudes and observer-frame magnitudes with rest-frame matching. We compared the results from both approaches and investigated the redshift and richness dependence of the RS parameters.

Results. We found that the quiescent galaxy population in groups builds up steadily from $z = 1.5$ – 2 across all richnesses, with faster and earlier growth in the richest groups. The first galaxies settle onto the RS ridgeline by $z \sim 2$, consistent with current evolutionary scenarios. Notably, we reported a rare protocluster core hosting quiescent galaxies at $z = 3.4$, potentially one of the most distant early RSs observed. Extending our study to X-ray properties, we found that X-ray faint groups have, on average, lower quiescent fractions than X-ray bright ones, likely reflecting their typical location in filaments where pre-processing is lower. Leveraging the broad wavelength coverage of COSMOS2025, we traced RS evolution using observed and rest-frame colors over ~ 12 Gyr, finding no significant trends in either the slope or the scatter of the ridgeline.

Key words. galaxies: clusters: general – galaxies: evolution – galaxies: groups: general – galaxies: high-redshift – galaxies: star formation – large-scale structure of Universe

1. Introduction

Dense environments typically host galaxy populations characterized by lower star formation rates (SFRs) and redder stellar populations compared to lower-density environments, at least up to $z \sim 1$. Compared to those in the field, these galaxies are believed to undergo substantial transformation, not only in the SFR (e.g., Scoville et al. 2013; Darvish et al. 2016; Taamoli et al. 2024), but also in morphology (e.g., Capak et al. 2007; Bamford et al. 2009), gas content, and metallicity (e.g., Catinella et al. 2013). This occurs through a variety of processes such as interactions (Hausman & Ostriker 1978), galaxy harassment (Moore et al. 1996), and ram-pressure stripping (Gunn & Gott 1972; see e.g., Boselli & Gavazzi 2006, for a review). This environmental segregation is reflected in the well-known SFR (or morphology)-density relation (e.g., Dressler 1980; Balogh et al. 1998). The way galaxies transform is generally explained as the result of a strongly intertwined impact of intrinsic properties, such

as stellar mass (nature) and factors related to the different environments galaxies experience in their lifetimes (nurture) (e.g., De Lucia et al. 2012). The disentanglement of mass- and environment-driven quenching remains debated. Building on the pioneering work of Peng et al. (2010) establishing this two-channel framework, recent studies (e.g., Chartab et al. 2020; Taamoli et al. 2024; Zheng et al. 2025; Hatamnia et al. 2025) have expanded this view, showing that mass quenching is a strong driver across all cosmic times especially on very massive galaxies, while environmental quenching dominates the quenching of galaxies in dense environments at later epochs and especially for satellites, generating a population of quiescent, red, passively evolving galaxies (e.g., Wetzel et al. 2012; Ziparo et al. 2014; Darvish et al. 2016; Kawinwanichakij et al. 2017). This population, hosted in dense environments such as clusters and groups, is known to occupy a tight grouping in the color-magnitude diagram (CMD) of the galaxy system, known as the red sequence (RS). The study of this prominent feature in the CMDs of galaxy clusters and groups is crucial for tracing

* Corresponding author: greta.toni4@unibo.it

the formation and evolution of quiescent galaxies across cosmic time.

The RS is thought to reflect the combined effects of early star formation histories, chemical enrichment, and environmentally driven quenching. These processes can be probed by studying the parameters of the RS ridgeline. The slope of the RS primarily encodes the mass–metallicity relation: more massive galaxies tend to be more metal-rich, and therefore redder, producing a negative color–magnitude slope (e.g., Kodama & Arimoto 1997; Gallazzi et al. 2006; De Lucia et al. 2007; Stott et al. 2009). The zero-point (the intercept or mean color at a fixed magnitude or stellar mass) is sensitive to the average stellar age and metallicity of the population, and its evolution with redshift constrains the formation epoch and passive aging of quiescent galaxies (e.g., Stanford et al. 1998; Gallazzi et al. 2006; Mei et al. 2009). The scatter around the RS reflects the diversity of star formation histories, including residual or recent quenching, metallicity spread, and photometric uncertainties as well; a small intrinsic scatter indicates that most RS galaxies formed the bulk of their stars early and rapidly (e.g., Bower et al. 1992; Blakeslee et al. 2006; Romeo et al. 2015). Taken together, the slope, zero-point, and scatter provide powerful diagnostics of the assembly history of quiescent galaxies and the timescales over which the RS is built up across cosmic time.

The RS and quiescent fraction have been extensively studied in massive clusters (e.g., Mei et al. 2009; Stott et al. 2009; Hennig et al. 2017). However, characterizing them in galaxy groups – which dominate the halo mass function and host the bulk of galaxies in the Universe – can provide deeper insights into the different quenching processes and galaxy formation scenarios. Over the past years, the presence of a well-defined RS in clusters and groups has also been used as a tracer to identify these systems and create robust catalogs up to $z \sim 1$ (e.g., Gladders & Yee 2000; Rykoff et al. 2014, 2016). While quiescent fractions and RS properties have been well characterized in the local Universe and up to intermediate redshifts ($z \sim 1$; e.g., Menci et al. 2008; Rudnick et al. 2009; Fritz et al. 2014), a better understanding of how the first galaxies quenched and built up the RS can be addressed by studying samples of groups and clusters at higher redshifts, up to the earlier stages of cluster formation itself ($z > 1.5$ – 2 ; e.g., Kravtsov & Borgani 2012; Shimakawa et al. 2018). This epoch reveals a diversity in galaxy populations, with systems showing enhanced star formation, active galactic nucleus (AGN) activity, and merging activity (e.g., Brodwin et al. 2013; Alberts et al. 2016; Wang et al. 2016), and others (sometimes even coexisting) showing an early-phase RS and high red fractions, especially in the central regions of galaxy overdensities (e.g., Andreon & Huertas-Company 2011; Spitler et al. 2012; Strazzullo et al. 2013, 2016; Zavala et al. 2019).

The James Webb Space Telescope (JWST) has recently demonstrated its unique power to detect quiescent galaxies at high redshift and to probe their environments (e.g., Jin et al. 2024; Ito et al. 2025; de Graaff et al. 2025). In this work, we aim to characterize quiescent fractions and RS in groups and low-mass clusters, across ~ 12 Gyrs of cosmic history. To do so, we leveraged our recent work (Toni et al. 2025) in building the largest group catalog based on deep JWST observations to date, spanning $z = 0.08$ to $z = 3.7$, over the COSMOS-Web field (Casey et al. 2023). In this catalog, galaxy groups have been detected with the Adaptive Matched Identifier of Clustered Objects (AMICO; Bellagamba et al. 2018; Maturi et al. 2019). The AMICO algorithm, officially selected for cluster detection in Euclid (Euclid Collaboration: Adam et al. 2019;

Euclid Collaboration: Bhargava et al. 2025), is based on optimal linear matched filtering, and has been proven to be a flexible tool for identifying groups and even protocluster cores (or overdensity peaks) up to $z \sim 4$ and down to a few $10^{12} M_{\odot}$ (Toni et al. 2024, 2025). What is particularly interesting for this analysis is that AMICO does not make explicit use of colors, limiting the possibility of biasing the selection toward systems with a clear RS, which is crucial when we want to study its buildup process and the diversity in the galaxy population that clusters and groups show, for instance, at $z \gtrsim 1$ – 1.5 . In this study, we also analyzed the first AMICO catalog generated on the COSMOS field (Toni et al. 2024) to explore the relationship between our results and the group X-ray emission. The sample covers a ~ 3 times larger area (the full COSMOS field) up to $z = 2$ and includes mass estimates and X-ray properties of optically selected groups.

The paper is organized as follows. Section 2 introduces the COSMOS group catalogs created with AMICO and the underlying galaxy catalogs. In Sect. 3, we describe our machine learning (ML) method of classifying galaxies as quiescent or star-forming, including training, algorithm testing, and performance comparison. Section 4 analyzes the evolution of quiescent fractions in COSMOS groups and their dependence on group properties. Section 5 characterizes the RS in our sample and compares it with evolutionary synthesis models and previous studies. Finally, Sect. 6 summarizes the main results and outlines future developments. In our analysis, we assume a standard concordance flat Λ CDM cosmology with $\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$, and $h = H_0/(100 \text{ km/s/Mpc}) = 0.7$. Magnitudes are expressed in the AB system and corrected for Galactic extinction.

2. The catalog of galaxies and galaxy groups

The COSMOS field (Scoville et al. 2007) is one of the most data-rich extragalactic regions, and benefits from extensive multi-wavelength coverage from X-ray to radio (e.g., Hasinger et al. 2007; Civano et al. 2016; Smolčić et al. 2017). Optical imaging encompasses coverage from several instruments, including the Canada-France-Hawaii Telescope (Sawicki et al. 2019), the Subaru Suprime-Cam (Taniguchi et al. 2015), the high-resolution Hubble Space Telescope ACS data (Koekemoer et al. 2007), and the broadband data in the g , r , i , z , and y filters from the Subaru Hyper Suprime-Cam (Aihara et al. 2022). Near-infrared coverage from the UltraVISTA survey (McCracken et al. 2012; Moneti et al. 2023) further complements this extensive dataset.

Building on this legacy, the COSMOS-Web Survey (PIs: Kartaltepe and Casey; Casey et al. 2023) is a JWST Cycle 1 program covering 0.54 deg^2 with four filters F115W, F150W, F277W, and F444W in NIRCcam (Rieke et al. 2023; Franco et al. 2025), achieving 5σ point-source depths of 27.5–28.2 mag. These filters offer a coverage of the near-infrared regime that is highly complementary to the UltraVISTA survey filters. An additional non-contiguous area is observed with the F770W MIRI filter (Wright et al. 2022; Harish et al. 2025).

The COSMOS-Web photometric galaxy catalog (also known as COSMOS2025; Shuntov et al. 2025), specifically targets the JWST-imaged region, building on previous COSMOS catalogs such as COSMOS2009 (Ilbert et al. 2009), COSMOS2015 (Laigle et al. 2016), and COSMOS2020 (Weaver et al. 2022), and includes over 784 000 sources. The COSMOS2025 catalog sources were detected using a PSF-homogenized χ^2 image combining all NIRCcam bands, using SourceXtractor++ (Bertin et al. 2022) to address PSF variations across datasets.

This approach boasts a high spatial resolution as reflected in the deblending power of the source extraction. In the COSMOS2025 catalog, redshifts (and all physical properties) are derived using the LePhare template-fitting code (Arnouts et al. 2002; Ilbert et al. 2006) with an expanded set of templates based on Bruzual & Charlot (2003) models (Ilbert et al. 2015). When compared to high-confidence spectroscopic samples, the photometric redshifts achieve a precision of ~ 0.01 for bright galaxies ($m_{F444W} < 23$) with $< 2\%$ catastrophic outliers, and maintain < 0.03 precision with $\sim 10\%$ outliers even for $26 < m_{F444W} < 28$. When divided between quiescent and star-forming, according to the NUVrJ criterion (Ilbert et al. 2013), galaxy photo- z s achieve 0.008 and 0.013 precision, and 2.26% and 1.95% outliers, respectively (see Table 3 in Shuntov et al. 2025).

We based our analysis on the new COSMOS-Web group and protocluster core catalog, presented in Toni et al. (2025) and extending up to $z = 3.7$ (the COSMOS-Web group catalog, hereinafter). The candidate groups were detected using the AMICO algorithm (Bellagamba et al. 2011, 2018; Maturi et al. 2019).

AMICO is a linear optimal matched filter that detects clusters and groups in photometric galaxy catalogs using the galaxy position, photometric redshift, and an additional galaxy property (in the simplest case), such as the magnitude in a reference band. During the detection process, AMICO produces estimates of amplitude for each pixel of the analyzed three-dimensional space of sky coordinates and redshifts. The amplitude is the result of the convolution of the data with an optimal filter (for details, we refer the reader to Bellagamba et al. 2018 and Maturi et al. 2019). Subsequently, the algorithm selects the candidates at the peaks in the amplitude map with the highest signal-to-noise ratio (S/N). After each selection, it assigns member galaxies to the cluster/group candidate by attributing a membership probability that depends on the galaxy properties, i.e., the magnitude and the redshift. Finally, AMICO uses this information to remove the signal of the detected object from the map, allowing easier detection of blended systems (for details, see Bellagamba et al. 2018; Maturi et al. 2019). Each AMICO run returns a list of candidates with their respective lists of member galaxies, each with its association probability.

The COSMOS-Web group catalog was retrieved by applying AMICO to COSMOS2025. The group search was performed using the galaxy position, photo- z distribution, and magnitude in a reference band, the NIRCам F150W band, extending to magnitude 27.3, making it the deepest AMICO application to date. The catalog includes 1678 group and protocluster core candidates in the COSMOS-Web field up to $z = 3.7$, with $S/N > 6.0$, and richness (λ_*) cut at a minimum value of 2 and reaching up to ~ 80 . More than 500 groups have their redshift confirmed by assigning spectroscopic counterparts (Toni et al. 2025).

In this study, we also include an analysis of our first COSMOS catalog, AMICO-COSMOS (Toni et al. 2024), to explore the relation with group X-ray emission. Unlike COSMOS-Web, AMICO-COSMOS was built by applying the AMICO algorithm to the COSMOS2020 (and 2015) catalog (covering the full COSMOS field). This group search was tested with three different photometric bands. The catalog boasts the availability of associated X-ray properties, such as luminosity, flux, and mass, for over 600 candidate groups.

Given the sample richness estimates and the masses inferred by Toni et al. (2024) being in the range $M_{200} \approx 6 \times 10^{12} - 3 \times 10^{14} M_{\odot}$, in this work, we will deal mainly with galaxy groups rather than galaxy clusters. Only a few objects in the COSMOS field are expected to have masses larger than $10^{14} M_{\odot}$ or more

than 50 members¹ according to previous detections performed in the COSMOS field (e.g., Knobel et al. 2012; Gozaliasl et al. 2019; Toni et al. 2024, 2025). Therefore, we will refer to our galaxy systems mainly as galaxy groups.

3. Classification of galaxies with machine learning

Since the AMICO galaxy group catalogs are based on input photometric galaxy catalogs in the COSMOS field, our goal is to classify COSMOS galaxies. Several methods have been developed over the past years to select quiescent and star-forming galaxies (see e.g., Pearson et al. 2023, for an overview), spanning from a classic sharp cut in sSFR (Salim et al. 2018), to color selections (e.g., Whitaker et al. 2011; Ilbert et al. 2013) and the use of spectroscopic features (e.g., Baldwin et al. 1981; Gallazzi et al. 2014). In this work, in order to classify galaxies into quiescent (red) and star-forming (blue), we used a new approach that incorporates the information from several “classical” methods to train a ML model. We chose to test two algorithms, one nonlinear and one linear: the eXtremeGradientBoosting classifier (XGB; Chen & Guestrin 2016) and the LinearDiscriminantAnalysis classifier (LDA; Fisher 1936), both as implemented in the scikit-learn Python package (Pedregosa et al. 2011).

3.1. Machine learning algorithms

The XGB algorithm belongs to the class of gradient boosted tree models (Friedman 2000). In simple terms, this means it classifies data by building a series of decision trees, where each new tree learns from the mistakes made by the previous one. The correction of errors is based on a method called gradient descent optimization, which consists of iterative adjustments of the model parameters to “descend” toward the minimization of the prediction errors. The XGB algorithm is well suited for classification tasks on structured data, because it provides high predictive accuracy, built-in handling of missing values, and uses a technique called regularization – a safeguard that prevents the model from fitting too closely to the training data (an issue known as overfitting). It also includes parameters to handle imbalanced class distributions effectively. Compared to linear classification methods, it can model complex, nonlinear relationships more effectively, despite being more computationally expensive and based on more complex modeling.

The LDA algorithm classifies data using linear combinations of features, assuming that targets in classes are normally distributed. This makes it a simple and efficient classification method, especially for small datasets for which these assumptions hold. However, it is less flexible than nonlinear methods such as XGB, less efficient with complex patterns, and does not include a default handling of imbalanced classes or missing values. For this application, we are dealing with highly imbalanced data, since photometric galaxy catalogs are expected to contain many more examples of blue star-forming galaxies rather than red quiescent galaxies. In the classification problem, these two are therefore referred to as the majority and minority classes, respectively. Class imbalance can be mitigated through the generation of synthetic data. One of the most used methods for data generation is the Synthetic Minority Over-sampling Technique (SMOTE; Chawla et al. 2011), which creates additional examples

¹ These are the commonly adopted thresholds used in the literature (e.g., Paul et al. 2017; Lovisari et al. 2021) to separate groups and clusters.

Table 1. Performance of the classification methods in terms of precision, recall, and F1 score.

Method	Precision (red/blue)	Recall (red/blue)	F1 score (red/blue)
<i>NUV, u, r, i, z, Y, J, H, K</i>			
LDA	0.94/0.99	0.84/1.00	0.89/0.99
LDA + SMOTE	0.69/1.00	1.00/0.97	0.81/0.98
XGB	0.96/1.00	0.95/1.00	0.95/1.00
XGB with -99.9	0.96/1.00	0.93/1.00	0.94/1.00
<i>NUV, r, J, K</i>			
LDA	0.90/0.99	0.90/0.99	0.90/0.99
LDA + SMOTE	0.66/1.00	1.00/0.96	0.80/0.98
XGB	0.96/1.00	0.95/1.00	0.95/1.00
XGB with -99.9	0.95/0.99	0.91/1.00	0.93/1.00

Notes. The upper block refers to the training with all available rest-frame magnitude bands, the bottom block to the one using only the four with the highest decision impact. Testing set sample size and red/blue ratio: 94236 galaxies with 7.1% red/blue ratio; 98659 with 7.2% red/blue ratio, when missing values are imputed.

for the minority class via feature interpolation between existing data. This generally helps improve decision boundaries and reduces bias toward the majority class. We tested the classification with LDA both with and without the generation of synthetic data. The results of this test are shown in Table 1 and discussed in Sect. 3.4. The LDA method also requires the data to be scaled and normalized before training, which was done with the StandardScaler method from `scikit-learn`.

3.2. Training and testing dataset

We chose to train and test the ML models on a galaxy sample based on the COSMOS2015 galaxy catalog (Laigle et al. 2016). Choosing this sample ensures a realistic representation of the features and properties of the galaxies that we want to classify. We chose not to train the algorithm on a subsample of COSMOS2025 since we aim to explore the full sample without compromising statistics and without introducing the risk of repeated data points in the training and target set, which may lead to overfitting. By choosing COSMOS2015 as the training set, we have no data-point overlap with the target dataset. Additional details about the training/testing set choice can be found in Appendix A.

The first step to create a training and testing set is to clean the parent catalog to reduce the possibility of training the models on misclassified or spurious objects. For this reason, we performed a conservative cleaning applying the following cuts:

- We removed stars (`type=1`) and unclassified objects due to SED fitting failure;
- We selected only data with high-quality photometry according to the SExtractor flagging system in optical broad bands and UltraVISTA bands: `[BAND]_FLAGS < 4`;
- We kept only safe and unmasked objects, namely those with `FLAG_PETER` in `(0, 4, 6)` (for details on this selection, see Sect. 2 of Toni et al. 2024) and additionally removed any objects falling inside the AMICO-COSMOS visibility mask, described in Sect. 4.3 of Toni et al. (2024);

- We removed galaxies located within 0.05 degrees of pairs of a bright star halo (or a halo-border pair), resulting in a more conservative mask where the field is more fragmented;
- Finally, we cut the dataset at $H < 25$ (corresponding to the mode of the magnitude distribution) to discard faint sources with large magnitude errors and homogenize completeness. This cleaning yielded a sample of approximately 300 000 galaxies in our training/testing set, across the range $0 < z < 6$. In order to establish the “ground truth” in the classification of galaxies for the training/testing set, we used the classification label yielded by several classical methods. We proceeded as follows. First, we selected four classical methods widely used in the literature to discriminate between the two galaxy classes:
 - NUVrJ (Ilbert et al. 2013): a rest-frame color-color cut at $NUV - r = 3(r - J) + 1$ and $NUV - r = 3.1$;
 - sSFR threshold (Salim et al. 2018): a fixed cut in sSFR at $10^{-10.5} \text{ yr}^{-1}$, as adopted for instance by Euclid Collaboration: Humphrey et al. (2023) and Bisigello et al. (2020);
 - Sa-color evolution (Andreon 2006; Radovich et al. 2020): a cut given by the evolution in z of the observed color of a modeled Sa-type galaxy. Colors bracket the 4000 Å break;
 - NUVrK (Davidzon et al. 2016): a rest-frame color-color cut at $NUV - r = 1.37(r - K) + 2.6$ and $NUV - r = 3.15$ and $r - K = 1.3$.

We classified galaxies into three categories according to these four methods: the galaxy is flagged as quiescent if it satisfies most of the four criteria (7% of the sample). The same procedure is followed to classify star-forming galaxies (92% of the sample). If the classification is ambiguous (e.g., 2 vs. 2), the galaxy is stored in the third class. This third class allows us to identify galaxies with the following characteristics: those with uncertain classifications, because of missing UV photometry and/or ambiguous SED fitting, as well as those in a physical transition phase between the RS and the blue cloud, a regime commonly referred to as the green valley. This category of objects represents ~1% of the full sample. Given the small statistics, it could not be introduced as a third class in the learning process and the classification. Therefore, we decided to remove these objects from the training and testing sets. This reduces ambiguity in the classification and results in an improvement of the performances we describe in Sect. 3.4 of about 2–5% in all evaluation metrics for the minority class. The final selected sample was then randomly divided into training (2/3) and testing set (1/3) using the `scikit-learn train_test_split` function. Additional information about the classical method thresholds, their properties, and the distribution of labeled galaxies in redshift and mass can be found in Appendix A.

3.3. Feature engineering

In ML, features are the measurable properties available in the data that we select to make predictions on the objects we want to target. Feature engineering is the ML pre-processing step of selecting, evaluating, and adjusting features, including the handling of missing data, so that the ML model can better capture patterns in the data to make predictions. First of all, we selected all rest-frame magnitudes available in both COSMOS2015 and COSMOS2020. These are the magnitudes in the following bands: *NUV, u, r, i, z, Y, J, H* and *K*. To get an estimate of the impact of features on the decision-making process, we performed an initial training with the XGB algorithm, which easily allows the evaluation of feature importance. The results are shown in Fig. 1. The importance is measured in terms of the

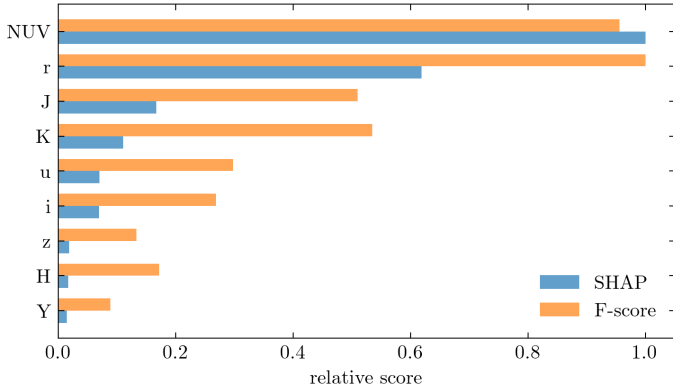


Fig. 1. Feature importance for an initial XGB training, measured by the F score (relative split frequency; orange bars), highlights bands directly used to define ground-truth labels are the most frequent in decision splits. To better assess feature impact, we evaluate SHAP scores (Lundberg & Lee 2017), which quantify the relative feature contribution to predictions (blue bars). Both approaches confirm that rest-frame magnitudes in r , NUV , K , and J provide the strongest constraining power.

F score (also known as weight, displayed by the orange bars), which is the number of times a feature appears in the decision tree splitting. The r , NUV , J , and K bands are the most frequent features used in the decision-making process. This result is expected, as these bands are directly used in assigning the ground-truth labels, as described in Sect. 3.2. To gain deeper insight into feature impact, we also evaluated the SHAP score (Lundberg & Lee 2017), which indicates how much each feature contributes to the model’s predictions (blue bars). The rest-frame magnitudes in r , NUV , J , and K are confirmed as the features with the greatest constraining power in the model predictions. These four magnitude bands are also the only rest-frame magnitudes available in the current release of the COSMOS2025 galaxy catalog at the moment of writing. Since the selection of features is not trivial, and often a few features with high importance simplify the model and lead to better performance than a large number of features, we chose to test and compare results using all magnitudes and with only the best-ranked four. In Sect. 3.4, we describe the differences in performance depending on the set of chosen features in more detail.

The XGB algorithm is sparsity-aware, meaning it can naturally handle missing data. This can be particularly useful in regions with reduced photometric coverage, as it allows us to include galaxies that are missing some rest-frame magnitudes rather than discarding them completely. Missing values are imputed with an arbitrary constant (-99.9), chosen to be clearly distinguishable from real values. Because XGB does not create artifacts from such artificial flags, this strategy is effective; indeed, Euclid Collaboration: Humphrey et al. (2023) showed that using a flagging number, like -99.9 , outperforms other imputation methods such as the mean, median, or minimum. This approach enables the algorithm to learn from both fully and partially characterized galaxies. We tested the XGB algorithm with and without imputation of a constant value to recover missing data.

3.4. Classification performance and results

To evaluate the algorithms and configurations, we relied on the standard metric used to evaluate classification performance

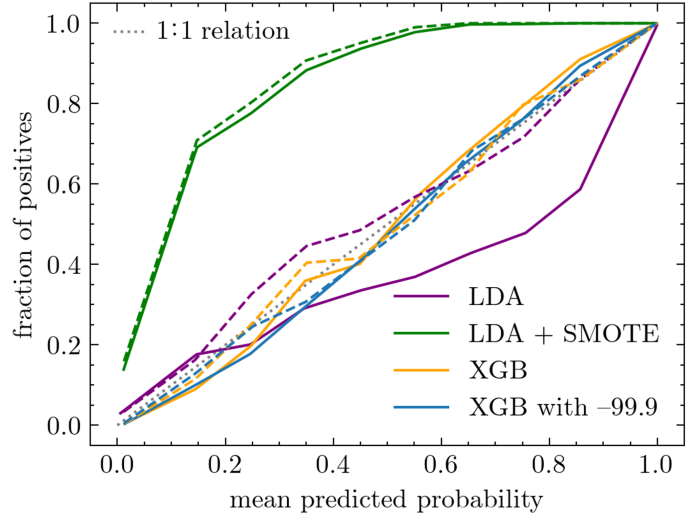


Fig. 2. Reliability curves showing the bias between predicted probability and fraction of actual positive labels (star-forming galaxies) using all rest-frame magnitude bands (solid lines) and the four most impacting in the decision (dashed line). Different colors represent different methods and configurations, as in the legend, and the dotted gray line represents the ideal one-to-one relation.

against the testing set: precision, recall, and F1 score. Precision (P) is the fraction of correctly identified objects among all objects assigned to a class, analogous to purity in astronomical samples. Recall (R) is the fraction of true objects correctly identified, analogous to completeness. The F1 score combines both precision and recall via their harmonic mean (i.e., $2PR/P + R$). These three quantities range between 0 and 1.

The performances evaluated against the testing set are summarized in Table 1, in terms of the metrics just introduced, both for the minority (quiescent galaxies, red) and for the majority class (star-forming galaxies, blue). The different methods yield comparable results for the classification of star-forming galaxies, whereas for the identification of quiescent galaxies, the XGB algorithm tends to perform better overall. Interestingly, applying SMOTE to help LDA with class imbalance increases recall for the minority class but significantly reduces precision. This may be due to an overlap between the two class distributions in the feature space, happening when classes share some similar properties and boundaries are not clearly separable, which is expected to cause overconfidence toward the minority class when synthetic data are injected (e.g., Chawla et al. 2011; Lu et al. 2024).

However, in this application, we are concerned not only with the quality of the binary classification of our target sample, but particularly with the reliability of the probability assigned to each object of belonging to a given class, since these probabilities are used to estimate quiescent fractions and identify the RS. Therefore, in addition to recall, precision, and F1 score, we analyzed how well the probabilities are estimated. Figure 2 shows the so-called reliability curve (also known as calibration curve), which plots the fraction of actual positives (in this case, “true” star-forming galaxies) against the mean predicted probability for that class, for all four methods, in bins of predicted probability. This diagnostic illustrates how closely the predicted probabilities match the true outcomes. A reliable, unbiased model produces a curve that follows the one-to-one diagonal in the plot: for example, if the mean predicted probability in a given bin is 0.7, then roughly 70% of the objects in that bin should belong to that class.

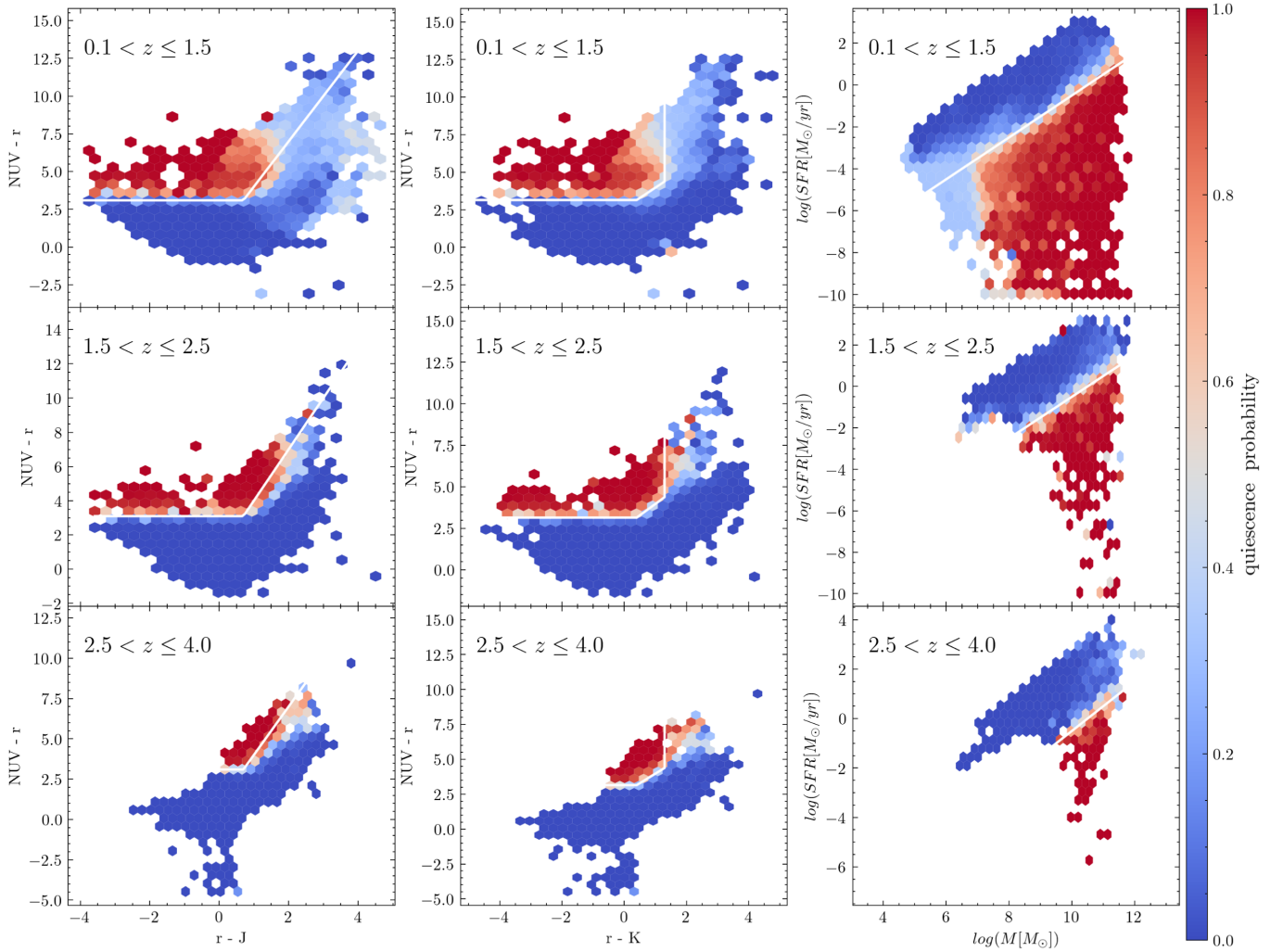


Fig. 3. Results of our probabilistic classification of COSMOS-Web galaxies with XGB with imputation, as in the NUVrJ (left column), NUVrK (middle column), and SFR– M (right column) planes. Each hexagon is color-coded by the mean quiescent probability (scale bar on the right), with rows corresponding to different redshift bins. As expected from our training, the method consistently reproduces classical cuts (white lines), while also providing additional insight near the classification boundaries. Classification is ambiguous for some low-mass galaxies at low redshift (light blue region in the top panels).

One case in which good classification performance does not imply good calibration is the model trained using data augmentation to balance the classes. Addressing the imbalance between star-forming and quiescent galaxies can lead the model to underestimate the probabilities of the majority class (positives), producing a reliability curve well above the one-to-one line (see the green curves in Fig. 2). The reliability curves also reveal the sensitivity of LDA to the number of selected features. For simple linear models, using a smaller set of strong features can yield better results than relying on a more complex set of galaxy properties.

By implementing the imputation of the missing values instead of complete discard, we observed a slight decrease in the performance for minority class recall, but a more reliable probability estimate. For this application, given the importance of an unbiased probability and the stability of the algorithm with respect to the number of features used, we decided to opt for an XGB-based classification with the imputation of a constant value for missing data. In Fig. 2, this configuration is represented by the dashed blue line, which is the most stable and closest to the one-to-one relation.

The results of the application of our chosen model for the probabilistic classification, to the COSMOS2025 dataset used for the COSMOS-Web group search, are shown in Fig. 3: each hexagon is color-coded according to the mean probability to be quiescent (bar on the right) inside that hexagonal bin; each column shows a different diagram used in classical methods, while rows refer to different redshift intervals. As expected from the labeling of our training set, our method is consistently mapping classical empirical cuts (displayed as white lines), while providing additional insight, particularly in regions near the classification boundaries. The algorithm struggles in the classification of galaxies with masses below $10^7 M_\odot$ in the first redshift bin, populating the light blue region in the top panels. These galaxies mostly have $z < 0.2$, and are expected to have a small impact on our analysis, as we are particularly interested in evolution at earlier cosmic times.

In a recent parallel study, Asadi et al. (2025) classified COSMOS2020 galaxies (Weaver et al. 2022) from the Farmer catalog release (Weaver et al. 2023b) using a gradient boosted tree algorithm trained on a mock galaxy catalog based on semi-analytical models and employing 28 colors. Their approach

differs from ours in several key aspects: it relies on simulated rather than observed data for training, uses apparent colors instead of rest-frame magnitudes, and labels the training set solely through an sSFR threshold. Since their trained model is publicly available, we applied it to the new COSMOS-Web data. The outcomes align with their findings for COSMOS2020: specifically, their model returns a higher fraction of quiescent galaxies at all redshifts compared to classical thresholds such as NUVrJ. Because our method is based on the consensus among multiple classical thresholds (see Sect. 3.2 and Appendix A), the model of Asadi et al. 2025 also predicts a consistently higher quiescent fraction relative to the approach presented in this work.

4. Quiescent galaxy fractions

We studied the fraction of quiescent (red) galaxies in groups and its dependencies on group richness and redshift. The quiescent fractions were computed with two different methods: pure membership (with testing of secondary association subtraction) and cylinder background subtraction.

Following common practice to reduce biases due to color uncertainties at the faint end and fully sample typical ranges occupied by RS galaxies, we introduced a redshift-dependent luminosity limit to define the quiescent fraction. This limit is often defined at 2 or 3 magnitudes fainter than an evolving characteristic magnitude or the magnitude of a reference galaxy, such as the brightest one (e.g., Stott et al. 2009; Cerulo et al. 2016; Radovich et al. 2020). The depth of COSMOS-Web allows us to push this limit deeper than in previous studies and to adopt a threshold at $m_*(z_j) + 4$, where z_j is the redshift of the group. In other words, for each group, we analyzed all galaxies up to 4 magnitudes fainter than the magnitude of a typical elliptical galaxy at the knee of the group luminosity function, $m_*(z_j)$, in our reference band (NIRCam F150W), which was estimated by evolving a massive passively evolving elliptical galaxy (see Toni et al. 2025, for details) with evolutionary synthesis models (Kotulla et al. 2009). This limit allows for a deep sampling of sub- m_* galaxies (galaxies fainter than the luminosity function knee), which include the strongest signatures of RS buildup (e.g., Stott et al. 2007, 2009). It should be noted that the $m_* + 4$ limit is not a sharp cut for the galaxy sample, because it depends on the redshift of the group the galaxies are associated with, so galaxies are limited in general by the survey's completeness limit (see Shuntov et al. 2025; Toni et al. 2025). However, at $z \sim 2.5$, the evolving $m_*(z_j)$ reaches the sample magnitude limit. Beyond this redshift, the $m_* + 4$ cut is no longer significant, because the faint-end is determined by the survey's limit. This implies that some of the faintest galaxies in our last two redshift bins – in particular, low-luminosity star-forming galaxies – might be undetected because beyond the survey limit, which could lead to a slight overestimation of the quiescent fraction in those bins with respect to lower redshifts. However, as discussed later, this effect appears not to be significant in our sample, indicating that this selection yields a consistent population across cosmic time while fully leveraging the depth and statistical power of COSMOS-Web. In this section, we briefly introduce the three approaches to compute the quiescent fraction and present the results for this group sample.

4.1. Pure membership and secondary association subtraction

Our sample includes the probability of each galaxy belonging to a given group, as provided by the AMICO algorithm. More-

over, we have the probability of each galaxy being quiescent as we estimated in Sect. 3. The simplest way to compute quiescent fractions is by summing the probabilities to be quiescent, conditional on the probability of belonging to a group, and dividing it by the total, regardless of the class attributed in Sect. 3. Given that $P_{i,j}$ is the AMICO probability of the i -th galaxy of belonging to the j -th group (see Maturi et al. 2019, for details) and that $P_{\text{red},i}$ and $P_{\text{blue},i}$ are the (complementary) probabilities of the galaxy to be quiescent or star-forming, respectively, the quiescent fraction of the j -th group, $f_{q,j}$, can be estimated as

$$f_{q,j} = \frac{\sum P_{\text{red},i} P_{i,j}}{\sum P_{\text{red},i} P_{i,j} + \sum P_{\text{blue},i} P_{i,j}} = \frac{\sum P_{\text{red},i} P_{i,j}}{\sum P_{i,j}}. \quad (1)$$

This computation relies on AMICO membership probabilities, which are model-dependent, computed during the detection process and based on the chosen model of that specific group search, which is built on an NFW profile (Navarro et al. 1997) and a Schechter luminosity function (Schechter 1976), for typical low-mass clusters or large groups (Toni et al. 2024, 2025). The AMICO association probability assignment also takes into account the possibility for a galaxy to be associated with multiple group detections. Thus, we introduced a small correction in the computation, which consists of considering only the main (highest probability) association for each galaxy and subtracting from it the probabilities to be associated with other detections (secondary associations). The probability, $P_{i,j}$, in Eq. (1) is therefore substituted by

$$P_{i,j-\text{new}} = P_{i,j-\text{main}} - \sum_{N_{\text{assoc}}-1} P_{i,j-\text{secondary}}. \quad (2)$$

However, this correction turned out to be minimal, and the results with and without correction are statistically compatible.

The results are reported in Fig. 4, where we show the weighted quiescent fraction, that is, the quiescent fraction expressed by Eq. (1), weighted by the purity of the sample in the corresponding redshift and richness bin. Purity was estimated against realistic data-driven mocks we have generated with SinFoniA (Maturi et al. 2019), which exploits a Monte Carlo approach. All details about the mock generation and the purity and completeness estimation can be found in Toni et al. (2025). For comparison, we included the plot in Fig. 4 without purity weighting in Appendix B.

Figure 4 shows that the population of quiescent galaxies seems to start assembling consistently shortly before $z = 2$, with earlier and accelerated buildup for the richest systems. The slight increase in the quiescent fraction observed at $z \geq 3$ may be influenced by the survey's magnitude limit, as previously discussed. However, this rise is minimal and also consistent with the expected contribution from a few rare, early RS galaxies that may already be emerging at these high redshifts (see Sect. 5.1). For this and the subsequent analyses, we chose richness bin sizes to homogenize the number counts whenever possible, and we estimated uncertainties using bootstrap resampling with 1000 realizations.

4.2. Cylinder background subtraction

The AMICO algorithm produces model-dependent membership probabilities for the group member galaxies. In some cases, when quantities used to model the distribution of galaxies in the template are studied, this might lead to biased considerations due to the underlying model chosen for the detection. This is the case for radial profiles and luminosity functions (see e.g., Puddu et al.

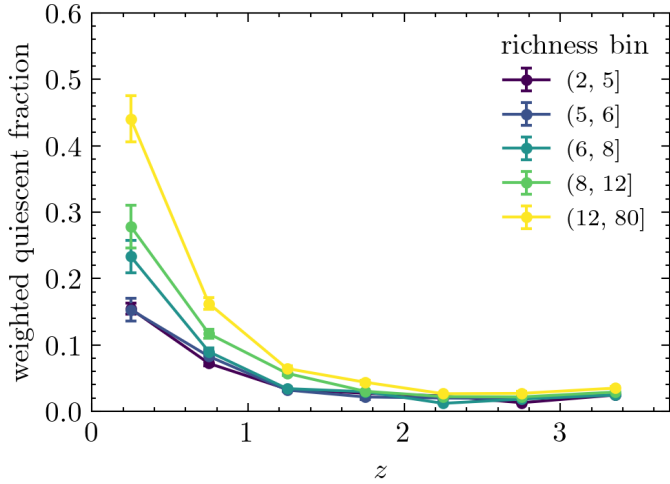


Fig. 4. Purity-weighted quiescent fraction vs. redshift in different richness (λ_*) bins, as in the legend. The richer the group, the faster its quiescent population buildup.

2021). In Toni et al. (2025), we proposed a new method based on a cylinder-based background subtraction technique to retrieve model-independent results for individual and stacked analyses. For the study of the quiescent fraction, we tested the cylinder method, performing background subtraction and estimating the quiescent fraction in a model-independent way. The implementation proceeds as follows:

1. The first step is the definition of cylindrical volumes around each group. We chose to make the cylinder extend $\pm 0.01(1+z_j)$ from the group center in the redshift direction, where z_j is the group redshift. On the sky-plane, we chose a variable radius given by the root mean square (RMS) of the group-centric galaxy distribution. The choice of the cylinder parameters is described in detail in Appendix B. We call the volumes inside the cylinders “group regions”.
2. Then, we must consider that not all volume is filled with galaxies, as areas can be masked due to the presence of bright stars and/or image artifacts and patterns. We estimated effective areas on the entire field (i.e., subtracting masked areas), using the same mask adopted in the detection process in each redshift slice (using the AMICO detection redshift resolution, $\Delta z = 0.01$). We accounted for masked areas both inside and outside the cylinders, namely for “group regions” and “field regions”, respectively.
3. We estimated quiescent and star-forming galaxy counts as the sum of the probability to be quiescent (P_{red}) or star-forming (P_{blue}), weighted by the value of the galaxy redshift probability distribution at the group redshift. Thus, the red and blue counts are expressed by

$$N_{\text{red}} = \sum_i P_{\text{red},i} p_i(z_j), \quad \text{and} \quad N_{\text{blue}} = \sum_i P_{\text{blue},i} p_i(z_j), \quad (3)$$

where the $p_i(z_j)$ is the redshift probability of the i -th galaxy at the redshift of the host group detection (z_j).

4. We did this computation for galaxies in both group and field regions and normalized counts by the volumes ($V_{(1)}$ and $V_{(0)}$, respectively), in all redshift slices crossed by the cylinder;
5. The final galaxy density within a group is the density in that group region minus the density in the field region at that redshift. Given that the subscripts (1) and (0) refer to galaxies and volumes in the group and field regions, respectively, the

final density of quiescent (red) galaxies (equivalent for the star-forming, or blue) in the j -th group is given by

$$\rho_{\text{red},j} = \sum_{(1)} P_{\text{red},i} p_i(z_j) V_{(1)}^{-1} - \sum_{(0)} P_{\text{red},i} p_i(z_j) V_{(0)}^{-1}. \quad (4)$$

6. Finally, the quiescent galaxy fraction is defined as the background-subtracted density of quiescent galaxies (Eq. (4)) over the total density. Galaxies and respective volumes located in the cylinder redshift slices, outside the cylinder volume, and along the line of sight of it, are rejected and not included in the calculation to avoid contamination. Thus, the final expression for the cylinder quiescent galaxy fraction, $f'_{\text{q},j}$, is given by

$$f'_{\text{q},j} = \frac{\rho_{\text{red},j}}{\rho_{\text{red},j} + \rho_{\text{blue},j}} \quad (5)$$

$$= \frac{\rho_{\text{red},j}}{\sum_{(1)} p_i(z_j) V_{(1)}^{-1} - \sum_{(0)} p_i(z_j) V_{(0)}^{-1}}. \quad (6)$$

Cylinder-based quiescent fractions were computed only for groups with a fraction of masked area not larger than 40% to reject extremely masked detections. We also removed groups in the extreme redshift bins to avoid border effects when considering the cylinders. Additional discussion about this computation and the method we used to deal with possible negative densities can be found in Appendix B. To retrieve these results, as well as those using the membership probability, we weighted the quiescent fraction obtained as just described by the value of the purity in the corresponding bins of redshift and richness. Results without purity-weighting are reported in Appendix B. Results for the cylinder-extracted quiescent fractions are shown in Fig. 5 for four different bins of richness. For comparison, we report in grayscale the results for the pure-membership quiescent fractions in the corresponding richness bins (Fig. 4). Apart from having larger uncertainties due also to lower statistics and to the scatter given by the background, the trend with redshift of the cylinder-based fractions is in line with a rapid buildup of the quiescent population starting at $z \sim 2$, and $z \sim 2.5$ for the richest groups, as seen when using pure memberships. However, when present, the quiescent population seems to be more dominant in rich groups when estimated through the cylinder-based method, which may be interpreted as a consequence of an effective field-galaxy contamination removal or an effect of background fluctuation, since the cylinder method retrieves fractions in comparison to the field population at that redshift. The discrepancy is less prominent in the low-richness bins. In conclusion, the model-dependent pure-membership method and the nearly model-independent² cylinder method, despite their different assumptions and computation procedures, produce strikingly consistent trends over nearly 12 Gyr of cosmic time.

4.3. X-ray luminosity of groups and environmental effects

In recent years, growing attention has been dedicated to the observed differences and expected biases between X-ray- and

² It should be noted that, although AMICO does not explicitly use color selections like RS-based detection algorithms, biases may still arise from the intrinsic properties of the data. In particular, quiescent galaxies tend to be massive and systematically brighter than star-forming ones. Even after removing model (and therefore magnitude) dependence via the cylinder method, the analysis may remain affected by the generally higher photo- z accuracy of quiescent galaxies compared to star-forming ones.

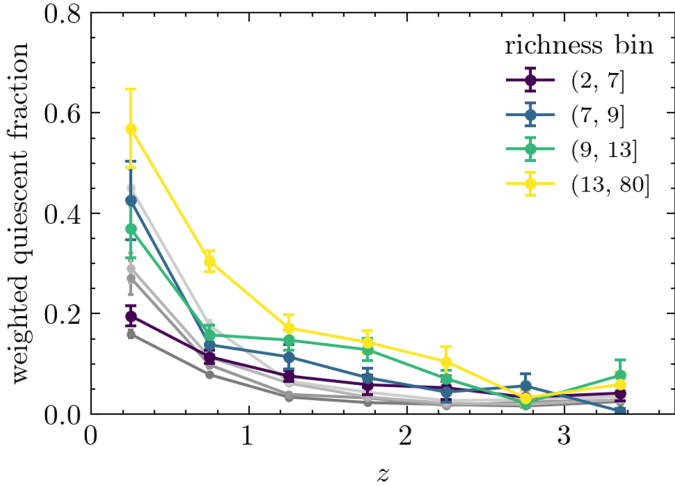


Fig. 5. Purity-weighted quiescent fraction as estimated with the cylinder background subtraction, as a function of redshift. Different colors show different richness bins, as in the legend. The grayscale trend lines refer to the fractions obtained with the pure membership method in corresponding richness bins (Fig. 4).

optically selected samples of galaxy clusters and groups. It is now increasingly recognized that X-ray–bright and X-ray–faint (or X-ray-undetected) systems often display distinct physical properties, indicating that different selection methods can preferentially identify different underlying populations. These discrepancies trace physical signatures of processes such as AGN feedback and halo assembly history, commonly referred to as halo assembly bias, and their impact becomes especially significant when transitioning to the galaxy group regime (e.g., Lovisari et al. 2021; Eckert et al. 2021; Popesso et al. 2024; Marini et al. 2025).

In this Section, we consider the AMICO-COSMOS group sample (Toni et al. 2024) comprising 622 candidate groups with X-ray flux above the significance level. Among these, 222 detections have a robust estimate of X-ray properties that was used to estimate group mass and to calibrate mass-proxy scaling relation in Toni et al. (2024). In this sample, around half of the detections have no significant X-ray emission, and this is fairly independent of S/N or redshift cuts. We leverage the availability of X-ray flux and X-ray flux significance (i.e., σ_X) in this sample to study the quiescent fraction and its evolution for the population of X-ray bright groups and those for which a significant X-ray emission was not detected. We calculated the quiescent fraction with the pure membership method, for simplicity, dividing it into the population of robust X-ray emitters ($\sigma_X > 1$ and good quality flag from Toni et al. 2024; 164 groups) and the population with X-ray flux under the significance limit ($\sigma_X \leq 1$; 359 groups). The difference between quiescent fraction for the two populations is shown in Fig. 6 as a function of the redshift (top panel) and AMICO signal-to-noise ratio, S/N_{noct} (bottom panel), for two different richness bins (blue and orange lines). The result is that groups with clear X-ray emission tend to have a quiescent fraction consistently higher than or at most equal to that of groups without significant X-ray emission. In particular, rich groups detected with high S/N and groups at $z \lesssim 0.5$ seem to have significantly more dominant population of quiescent galaxies when they are bright in the X-rays.

In a recent study, Popesso et al. (2024) compare the GAMA and eROSITA samples by analyzing the properties of X-ray-undetected clusters and groups. Among the considered proper-

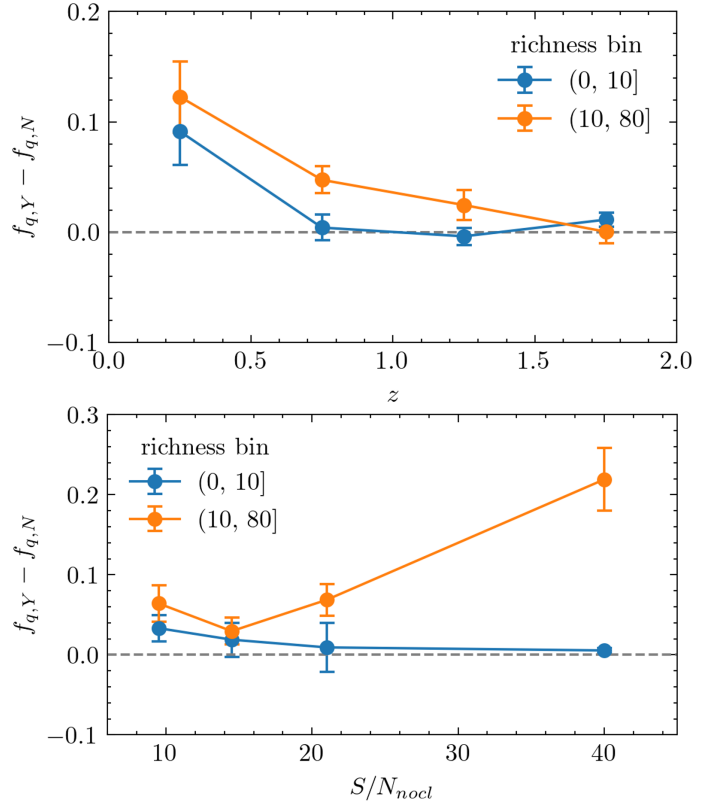


Fig. 6. Quiescent fraction difference between X-ray bright (Y) and X-ray faint (N) groups as a function of redshift (top panel) and AMICO S/N_{noct} (bottom panel), in two bins of richness as in the legend.

ties, they study the position of the systems with respect to the large-scale structure (LSS), finding that $\sim 90\%$ of low-emission clusters and groups, often undetected in the X-rays, are located in filaments (and sheets) rather than in the nodes of the cosmic web. Interestingly, this might be correlating with the properties of the hosted galaxy population, like the pre-accretion processing and therefore the star formation activity. Groups residing in nodes are more likely to accrete galaxies pre-processed in filaments, and therefore are more quenched on average. On the other hand, groups hosted by filaments accrete galaxies from the field that are therefore less processed. To check whether this is the case for our sample as well, we compared the position of our groups with that of galaxies that have been found to be located in filaments or in clusters based on local density in LSS maps. For this, we used data from the COSMOS LSS study by Darvish et al. (2014)³, which covers the redshift interval $0.1 \leq z \leq 1.4$ and the more recent data based on the COSMOS2020 galaxy catalog, which extends from $z = 0.5$ to beyond the higher end of our group sample (Taamoli et al., in prep.).

The results of this test are shown in Fig. 7 for the COSMOS2020-based LSS; results for the LSS map by Darvish et al. (2014) show very similar trends. In Fig. 7, the horizontal axis represents the filament signal, while the vertical axis indicates the probability density distribution smoothed and normalized using Kernel Density Estimation (KDE). To compute the filament signal, we counted galaxies in the central $0.25 \text{ Mpc}/h$ of each group and defined the filament signal as the

³ The density field was quantified through a scale-independent Multi-scale Morphology Filter (MMF) algorithm, on which the extraction of LSS components (field, filaments and clusters/nodes) was based (see Darvish et al. 2014, Sect. 3 for details).

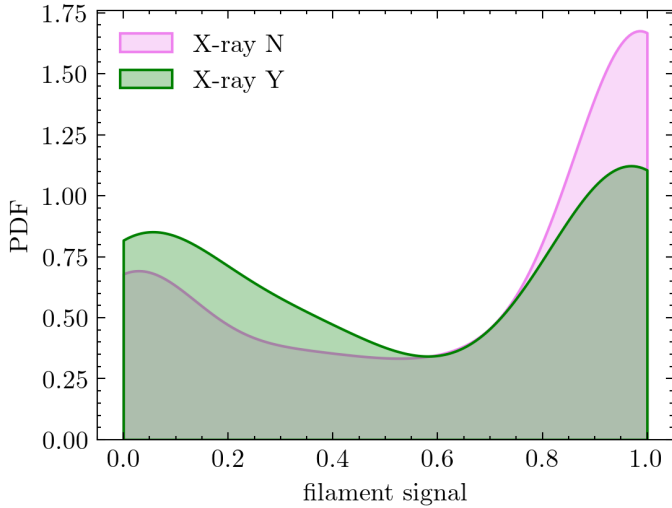


Fig. 7. Filament signal (ratio of number of filament galaxies to number of filament and cluster galaxies) probability density function for X-ray bright (*Y*) and X-ray faint (*N*) groups, using the LSS data and galaxy classification by Taamoli et al., in prep. Consistent results were obtained using the LSS information provided by Darvish et al. (2014).

ratio of the number of galaxies identified as part of a filament to the total (number of filament galaxies + number of cluster/node galaxies). In this way, groups with a filament signal close to 0 are likely hosted by a node (they are rich in cluster/node galaxies), while groups with filament signals close to 1 sit further away from nodes, likely found in filamentary environments connecting nodes. Groups without significant X-ray emission (pink curve) have around 2 times higher density in filaments (filament signal > 0.8) with respect to X-ray bright ones (green curve). The latter are instead more likely to be found in denser environments in correspondence or close to cosmic-web nodes (filament signal < 0.2).

Interesting insights can be gained from the investigation of group properties through a comparison between optically selected and X-ray selected clusters and groups. A comprehensive analysis of this aspect, for what concerns the AMICO-COSMOS groups and that incorporates the X-ray characterization of the new COSMOS-Web group catalog, will be presented in a forthcoming dedicated paper.

5. Red sequence

We now turn our attention to how the RS in our group sample evolves over a wide range of redshifts, from $z = 0$ to $z = 3.7$. To study this, we need to identify the RS and characterize it through the main parameters of the RS ridgeline.

5.1. Observed red sequence with rest-frame matching

The first step to robustly detect the observed RS is to identify the appropriate combination of colors and magnitude that, at a given redshift, allows a clean identification of the sequence in the CMD. This can be done following the matched rest-frame photometry method (Blakeslee et al. 2006; Stott et al. 2009), namely by choosing two magnitude bands that bracket a characterizing spectral feature, like the 4000 Å break, and the reddest of the two as the reference magnitude for the CMD. This reduces the impact of the k correction and the redshift dependence of the

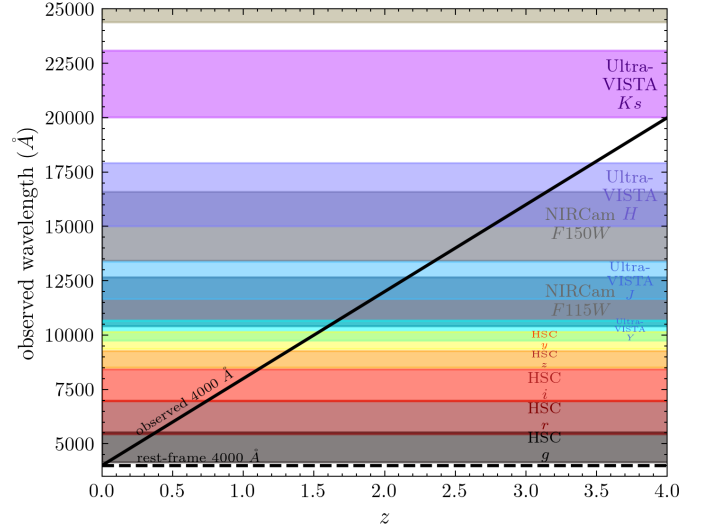


Fig. 8. Sampling in the COSMOS-Web filters of the observed 4000 Å, the spectral feature used for RS detection (solid black line). Different colors indicate different bands, as in the labels in the plot. Pairs of bands bracketing the break are reported in Table 2. The complementarity of COSMOS-Web dataset filters allows for a contiguous coverage up to $z \sim 3.5$.

Table 2. Red sequence selection method: observed color and magnitude choices by redshift interval.

Redshift Range	Color	Magnitude
$0.01 < z < 0.38$	$g - r$	r
$0.38 < z < 0.70$	$r - i$	i
$0.70 < z < 1.10$	$i - z$	z
$1.10 < z < 1.40$	$z - Y$	Y
$1.40 < z < 1.70$	$Y - J$	J
$1.70 < z < 2.35$	$F115W - F150W$	$F150W$
$2.35 < z < 3.25$	$F150W - H$	H
$3.25 < z < 3.70$	$H - Ks$	Ks

colors, straddling a feature that strongly traces different stellar populations in RS and non-RS galaxies.

The COSMOS-Web galaxy catalog features a comprehensive set of magnitude bands spanning from optical to mid-infrared wavelengths. This coverage is uniform and continuous thanks to the complementarity of the JWST NIRCcam filters and the UltraVISTA filters, making this the perfect opportunity to study observed RS evolution over a wide interval of redshifts, using the matched rest-frame photometry technique. This can be seen in Fig. 8, which shows the observed wavelength of the rest-frame 4000 Å break (solid black line) at different redshifts over the studied interval, and which filters are sampling and bracketing it. As clearly visible, the filter coverage of this spectral feature is continuous until at least $z \sim 3.5$, namely virtually covering the full COSMOS-Web group sample. Our selection of the 4000 Å-bracketing band pairs and the relative intervals are shown in Table 2. Once the RS is sampled in consistent observed colors, the ridgeline can be identified as the best-fitting line representing the locus of the quiescent member galaxies in the CMD.

As in Sect. 4, we performed this RS analysis over 4 magnitudes from the $m_*(z)$ in the F150W band, which, given the depth of

the galaxy catalog, guarantees redshift-dependent completeness until at least $z \sim 2.5$, without compromising the statistics. We computed the RS parameters by performing a weighted least squares (WLS) fitting analysis of the group CMDs on all galaxies attributed to a group with a probability larger than the minimum threshold imposed by AMICO; that is, 0.005. In a similar way to what we have done for the red fraction in Sect. 4.1, we weighted each galaxy with the probability of being quiescent, conditional on the probability of being a group member. This represents the first guess for the position of the RS, which is nevertheless still potentially contaminated by central and/or bright star-forming galaxies that might draw the ridgeline guess toward bluer colors. To refine the position guess, we performed an iterative cleaning of outliers by fitting with a 3σ clipping. Once convergence is reached (over a maximum of five iterations), the following RS best-fit parameters are stored for each group: the scatter computed on final retained galaxies, the average color, which is the color of the RS ridgeline computed at $m_*(z)$ and the number of RS galaxies, which we defined as the number of quiescent member galaxies ($P_{\text{red}} > 50\%$ and $P > 50\%$) lying within a typical ± 0.3 scatter (e.g., De Lucia et al. 2007; Martinet et al. 2015) from the best-fit line.

Besides the direct count of RS galaxies, we also computed λ_{RS} , which is the sum of the membership probability of these RS galaxies, in analogy to the AMICO richness. Figure 9 shows the fraction of λ_{RS} to the λ (sum of all membership probabilities of galaxies used for the initial guess) as a function of redshift, for the sample of groups with an RS identified with $\lambda_{\text{RS}} > 2$. This is an arbitrary limit that roughly corresponds to having at least 3 quiescent galaxies on the RS ridgeline, and therefore to a robust RS detection. The yellow line and points indicate the average trend of the RS fraction with redshift for the full sample. The size of the red points is proportional to the S/N of the group detection, which is used as a weight for the average. The gray trend line is reported from the quiescent fraction analysis for the central richness bin (see Fig. 4), while the blue trend line shows the same as done with the apparent magnitude but using rest-frame J and K_s , which we describe in Sect. 5.2. Quiescent fractions and RS fractions differ in terms of how they are defined and computed. The former is based on the probability of a galaxy being a member and being quiescent, while the latter also leverages the color homogeneity of the quiescent population (extracted by fitting a tight relation in the CMD). We find consistent trends between the two independent computations.

In the bottom plot of Fig. 9, we show the fraction of groups with an identified RS with $\lambda_{\text{RS}} > 2$ over the total number of detections per redshift bin, for S/N -cut at 6, 7, and 10 (i.e., purity level of 77%, 80%, and 90%). The overall trend follows that of the quiescent fraction and RS fraction, with the buildup of the RS settling between $z = 2$ and $z = 2.5$. The peak at $z \sim 0.7\text{--}0.8$ may be due to the presence of the COSMOS Wall (Iovino et al. 2016), which hosts more mature systems with established RSs. The total number of groups with an identified RS in this way is 214.

The highest-redshift RS detected in this study is located at $z = 3.4$ and it contains three quiescent member galaxies with very similar colors within around 250 kpc/h of each other. This high- z object is among the new protocluster cores not known in the literature before, which we recently discovered (Toni et al. 2025)⁴. The protocluster core is rich ($\lambda_* \sim 25$, $\lambda \sim 119$) and contains five galaxies with spectroscopic redshift, according to

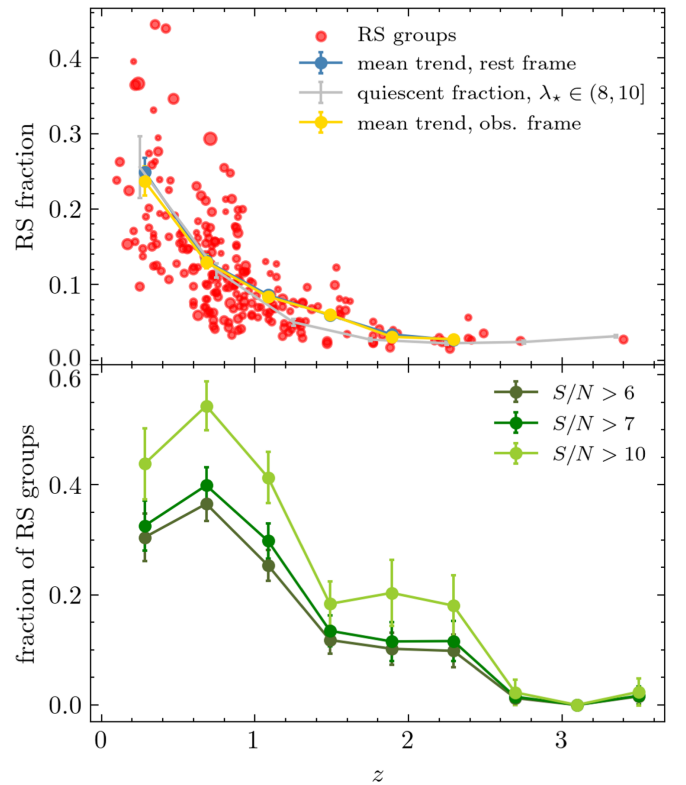


Fig. 9. Top panel: Fraction of the selected RS galaxies over the total ($\lambda_{\text{RS}}/\lambda$) in each group, as a function of redshift for the sample with identified RS. The yellow and blue lines indicate the mean fraction in each z -bin weighted by the group S/N (size of red points is proportional to the group S/N), using matched observed and rest-frame colors for the fit, respectively. Bottom panel: Fraction of groups with identified RS at different S/N (or purity) levels, as a function of redshift, showing the first consistent appearance of RS in groups at $z \sim 2\text{--}2.5$.

our spectroscopic counterpart assignment in the COSMOS spec- z compilation (Khostovan et al. 2025). The spectroscopic redshifts are compatible with that assigned by AMICO, solely based on photometry, which confirms the presence of an overdensity of galaxies at this location. However, none of the quiescent galaxies of the RS we found have a public spec- z available. If confirmed, this would represent one of the highest-redshift examples of early RS detected to date (see the $z \sim 4$ protocluster reported by Tanaka et al. 2024, and the overdensity discovered at $z = 3.44$ by Jin et al. 2024). The CMD of this $z = 3.4$ detection (ID CW117) is shown in Fig. 10, where we plot AMICO member galaxies with size proportional to their membership probability and their color referring to the probability of being quiescent and field galaxies as gray points. In the plot, we also report the main parameters of the RS ridgeline and the RS richness (λ_{RS}), while the RS galaxies are marked by black circles around the points. The dashed orange line indicates the expected color of a population passively evolving since $z = 5$, which is our reference model. This represents the color of a typical luminosity-function-knee galaxy (with magnitude m_*), according to the same configuration chosen for the AMICO model used in the group search, and mentioned earlier. The model was computed with the evolutionary synthesis model generator GALEV (Kotulla et al. 2009), relying on a Kroupa initial mass function (Kroupa 2002) and adopting a chemically consistent approach (see Kotulla et al. 2009, for further details) for a massive

⁴ And detected independently in a parallel work presented by Hung et al. (2025).

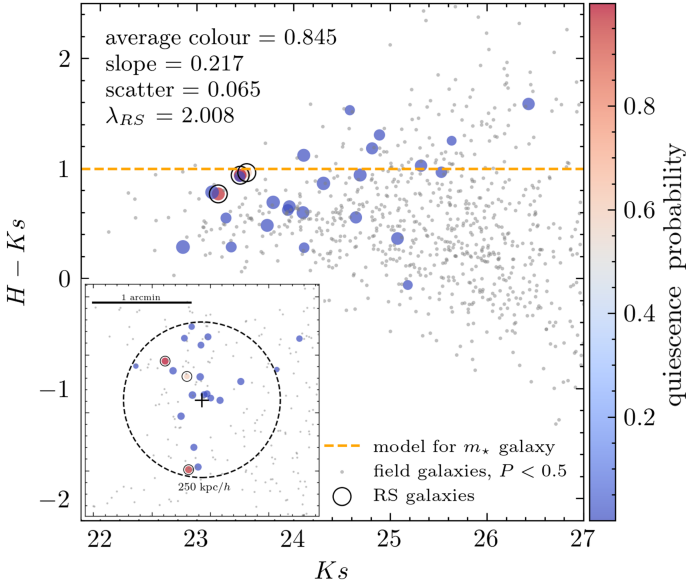


Fig. 10. Observed-frame CMD of a rich protocluster core at $z = 3.4$ (CW117) with three quiescent galaxies (red points with black contours) with colors consistent with RS evolutionary synthesis models for a typical m_* elliptical (orange line). This is the highest- z RS we detected, which, once confirmed, would represent one of the earliest hints of RS ever discovered.

elliptical formed at $z_f = 8$ with a more recent medium-intensity star formation burst at $z_b = 5$. This approach is similar to the one already used by Castignani et al. (2022, 2023). In the bottom left of the panel, we report the sky distribution of the galaxies in the central region of this high- z detection, with the same notation as in the CMD plot.

5.2. Results and comparison with rest-frame red sequence

We performed RS fitting with the same procedure described above (3σ -clipped fit on WLS initial guess) based on the LePhare rest-frame magnitudes in the UltraVISTA- J and $-Ks$ bands available in the COSMOS2025 release. We confirmed all RS detected with the observed magnitudes, mostly with compatible parameters. We additionally detected RS (mainly with low richness, $\lambda_{RS} \leq 3$) for another 11 groups.

The RS parameters as a function of redshift for the RS ridgelines detected in this study are shown in Fig. 11. When computing the fitted relations and mean trends, each group candidate is weighted by its S/N to suppress the impact of the smallest and least reliable detections on the overall fit. The size of the points is proportional to the S/N of the groups. The top panel shows the average observed color for all the detections in the COSMOS-Web group catalog, compared to the reference model we described above in Sect 5.1 (orange lines). Colors are computed in the associated redshift-dependent bands, as summarized in Table 2, which explains the jumps in average color when the 4000 \AA switches between adjacent bands. Besides this effect, the trend of the model is consistent with the evolution of the observed colors in our sample until at least $z \sim 2.4$. For the last two redshift intervals, the observed color of the ridgeline is not consistent with a prototypical quiescent galaxy, which is expected since at this redshift we do not fit any RS for most of the groups, and the ridgeline parameters are tentative and mainly driven by bright star-forming galaxies populating groups and cores at such high redshifts. The location in the plot of the

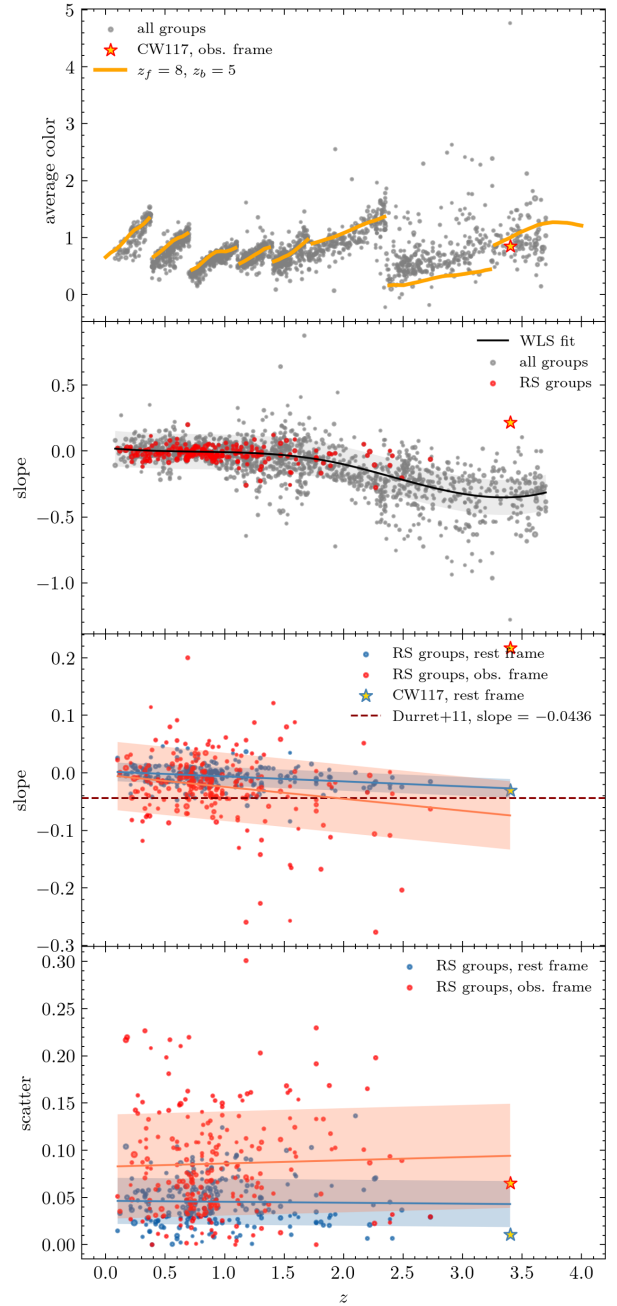


Fig. 11. Evolution with redshift of the main RS parameters for our group sample. From top to bottom: the first panel shows the average observed color as in Table 2, compared to our reference $m_*(z)$ model (orange lines). The star shows the CW117 detection, which has RS color compatible with that expected at that redshift for typical RS galaxies; the second panel shows the slope, with all groups in gray and those with RS in red; the third panel is a zoom on the slope of RS-groups only, with the parameters as retrieved using observer-frame (red points and orange fit line) and rest-frame magnitudes (blue point and fit line); the fourth panel shows the same but for the scatter of RS-groups. The yellow star shows the RS parameters of CW117, in the observer frame (red contours) and the rest frame (blue contours).

highest- z RS presented above (CW117) is indicated by a yellow star. As already mentioned, the average observed color of this RS is consistent with the model.

The second and third panels show the ridgeline slope. In the second plot, we include all groups in the sample (gray points)

with the relative total fitted trend (black line) and its 1σ error (shaded area). Red points mark groups for which it was possible to identify a RS with $\lambda_{RS} > 2$. This latter sub-sample is shown in the third plot from the top, where we compare the slope computed with observed rest-frame-matched colors (red points and orange line and shaded area) with that based on rest-frame J and K_s magnitudes (blue points and blue line and shaded area). The two trends are consistent within confidence intervals, with the rest-frame colors offering a stronger constraining power. Under the assumption of negligible k -correction effect, the RS slope trend reflects metallicity evolution (e.g., Stott et al. 2009). We found a slight increase in the absolute value of the slope with redshift. However, the steepening of the RS over the studied redshift range is very mild, and there is no clear evidence of a significant evolution. The average RS slope evolves of only $\Delta\text{slope} \sim 0.05$ across 12 Gyrs. This weak trend is consistent with little or no evolution in the mass–metallicity relation of quiescent galaxies, in agreement with, for instance, what was found for clusters up to $z \sim 1.5$ by Mei et al. (2009) and Cerulo et al. (2016). Therefore, our results support this picture, indicating that the mass–metallicity relation was already in place and has remained largely preserved since even earlier cosmic times. A slope for the RS typically adopted in the literature (Durret et al. 2011; Martinet et al. 2015; Takey et al. 2019) is shown as a dashed dark red line for comparison. Once again, the location of CW117 is indicated by a yellow star, with red contours for observed colors and blue contours for the rest frame. The slope of the RS in CW117 is a clear outlier in the observed colors (red star in the second and third panels of Fig. 11), because the quiescent galaxies lie very close together in the CMD (see Fig. 10), which drives the fitted slope to become large and positive.

As for the slope, our results for the RS scatter (fourth panel of Fig. 11) show little to no evolution over the studied redshift range. The RS scatter is found to be below 0.1 mag for most of the detections ($\sim 64\%$), especially when measured in the rest frame ($\sim 98\%$). The persistence of small scatters might suggest that the bulk of stars in RS galaxies formed early and over relatively short timescales, with limited diversity in their subsequent star formation histories. However, a more quantitative interpretation, such as estimating age dispersion, depends sensitively on photometric uncertainties, color choice, RS definition, group assembly history, luminosity thresholds, and survey depth; therefore, it is not straightforward to draw firm conclusions about the absolute age spread of the population from the observed RS scatter. However, the average scatter values and trend behavior are consistent with previous studies based on both simulations and semi-analytical models (e.g., Menci et al. 2008; Romeo et al. 2015) and on observations (e.g., Hao et al. 2009; Mei et al. 2009; Cerulo et al. 2016). We found that the scatter little-to-no evolution reported by Cerulo et al. (2016) for clusters up to $z = 1.5$ persists at higher redshifts, with an average rest-frame value of $\approx 0.045 \pm 0.025$ mag, consistent with that found by Mei et al. (2009) for elliptical (0.042 ± 0.021 mag) and for elliptical + lenticular galaxies (0.062 ± 0.015 mag). When considering observed colors, average RS scatter and its spread increase with respect to rest-frame measurements ($\approx 0.084 \pm 0.054$), consistent with the low-redshift observations by Hao et al. (2009) and semi-analytical model predictions (Menci et al. 2008).

6. Conclusions

We conducted a study of the evolution in redshift and the dependence on group richness of the quiescent galaxy fraction and of the RS parameters for COSMOS groups detected

with the AMICO algorithm, across around 12 Gyrs of cosmic history, from $z = 0$ to $z = 3.7$. We used multiple complementary methods to estimate the quiescent fraction and detect the RS, finding consistent results across different approaches. In total, we identified 225 RS ridgelines using rest-frame colors, with 214 also consistently characterized using rest-frame-matched observed colors. The main results of our study are as follows:

- Machine learning classifiers (both linear and nonlinear) provide a powerful framework for leveraging photometric multi-band survey data to robustly distinguish between quiescent and star-forming galaxies. In particular, gradient boosting decision trees, such as XGB, offer efficient handling of missing data and class imbalance, while delivering reliable probabilistic classifications. In our application, when tested against a testing dataset, this approach achieved an F1 score exceeding 93%, even in the case of limited availability of rest-frame magnitudes.
- The AMICO algorithm identifies galaxy groups without relying explicitly on color information. As a result, the AMICO catalogs enable a robust analysis of member galaxy properties, such as colors and luminosities, in both model-dependent and model-independent ways. Our results on the buildup of the quiescent galaxy population are found to be consistent across these methods.
- We found that the quiescent galaxy population in groups builds up steadily starting from $z = 1.5$ – 2 in all richness bins, with evidence of an earlier and accelerated growth for the richest systems, consistent with the scenario proposed, for instance, by Cerulo et al. (2016).
- By analyzing X-ray characterized AMICO detection, we found an indication that X-ray faint groups show, on average, lower quiescent fractions than X-ray bright ones. This might be connected to their location in the cosmic web, as X-ray faint groups are more likely to be found in filaments, where the environmental pre-processing of galaxies is lower compared to cosmic nodes, as well as the local density.
- We detected and characterized 225 RS ridgelines in COSMOS-Web groups, providing an extensive redshift coverage. We report the discovery of a rare compact (~ 250 kpc/ h scale) protocluster core with quiescent galaxies at $z = 3.4$, which, if confirmed, may represent one of the most distant early RSs detected to date, not known in the literature before our recent group search in COSMOS-Web.
- We found that the observed average colors of the RS ridgelines in our sample are consistent with those predicted by a passively evolving elliptical galaxy model. Regarding the slope and scatter of the RS, we observed no significant evolution with redshift, suggesting that the absence of significant trends found in previous studies of more massive galaxy clusters and at lower redshift may persist in the group regime and at earlier epochs.

This study extends the detection and characterization of the RS and quiescent galaxy fractions beyond $z > 1.5$ and into the regime of galaxy groups. Such an analysis is a natural application of a deep and well-characterized dataset such as the COSMOS-Web group catalog. Especially for high- z candidates and RS evolution trends, we stress the necessity of spectroscopic confirmation, which we plan to carry out in a dedicated campaign. In forthcoming work, we shall further investigate the correlation between galaxy properties and X-ray emission (i.e., intracluster and intragroup medium properties) and its connection with the LSS. Additionally, we plan to explore the properties and role of the central brightest group galaxies, examining

how their evolution relates to the environment and the assembly history of their hosts.

Acknowledgements. The authors thank the anonymous referee for the careful reading and the very useful suggestions, which helped improve the robustness and clarity of this manuscript. GT thanks David P. Petri for the valuable discussion and useful suggestions regarding ML. LM acknowledges the financial contribution from the PRIN-MUR 2022 20227RNLY3 grant “The concordance cosmological model: stress-tests with galaxy clusters” supported by Next Generation EU and from the grant ASI n. 2024-10-HH.0 “Attività scientifiche per la missione Euclid-fase E”.

References

- Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, *PASJ*, 74, 247
- Alberts, S., Pope, A., Brodwin, M., et al. 2016, *ApJ*, 825, 72
- Andreon, S. 2006, *MNRAS*, 369, 969
- Andreon, S., & Huertas-Company, M. 2011, *A&A*, 526, A11
- Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, *MNRAS*, 329, 355
- Arnouts, S., Le Floch, E., Chevallard, J., et al. 2013, *A&A*, 558, A67
- Asadi, V., Chartab, N., Zonoozi, A. H., et al. 2025, *ApJ*, 993, 123
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
- Balogh, M. L., Schade, D., Morris, S. L., et al. 1998, *ApJ*, 504, L75
- Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, *MNRAS*, 393, 1324
- Bellagamba, F., Maturi, M., Hamana, T., et al. 2011, *MNRAS*, 413, 1145
- Bellagamba, F., Roncarelli, M., Maturi, M., & Moscardini, L. 2018, *MNRAS*, 473, 5221
- Bertin, E., Schefer, M., Apostolakis, N., et al. 2022, *Astrophysics Source Code Library* [record ascl:2212.018]
- Bisigello, L., Kuchner, U., Conselice, C. J., et al. 2020, *MNRAS*, 494, 2337
- Blakeslee, J. P., Holden, B. P., Franx, M., et al. 2006, *ApJ*, 644, 30
- Boselli, A., & Gavazzi, G. 2006, *PASP*, 118, 517
- Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, *MNRAS*, 254, 601
- Brodwin, M., Stanford, S. A., Gonzalez, A. H., et al. 2013, *ApJ*, 779, 138
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Capak, P., Abraham, R. G., Ellis, R. S., et al. 2007, *ApJS*, 172, 284
- Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2023, *ApJ*, 954, 31
- Castignani, G., Radovich, M., Combes, F., et al. 2022, *A&A*, 667, A52
- Castignani, G., Radovich, M., Combes, F., et al. 2023, *A&A*, 672, A139
- Catinella, B., Schiminovich, D., Cortese, L., et al. 2013, *MNRAS*, 436, 34
- Cerulo, P., Couch, W. J., Lidman, C., et al. 2016, *MNRAS*, 457, 2209
- Chartab, N., Mobasher, B., Darvish, B., et al. 2020, *ApJ*, 890, 7
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2011, *SMOTE: Synthetic Minority Over-sampling Technique*, [arXiv:1106.1813]
- Chen, T., & Guestrin, C. 2016, *XGBoost: A Scalable Tree Boosting System*, [arXiv:1603.02754]
- Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
- Darvish, B., Sobral, D., Mobasher, B., et al. 2014, *ApJ*, 796, 51
- Darvish, B., Mobasher, B., Sobral, D., et al. 2016, *ApJ*, 825, 113
- Davidzon, I., Cucciati, O., Bolzonella, M., et al. 2016, *A&A*, 586, A23
- de Graaff, A., Setton, D. J., Brammer, G., et al. 2025, *Nat. Astron.*, 9, 280
- De Lucia, G., Poggianti, B. M., Aragón-Salamanca, A., et al. 2007, *MNRAS*, 374, 809
- De Lucia, G., Weinmann, S., Poggianti, B. M., Aragón-Salamanca, A., & Zaritsky, D. 2012, *MNRAS*, 423, 1277
- Dressler, A. 1980, *ApJ*, 236, 351
- Durret, F., Laganá, T. F., & Haider, M. 2011, *A&A*, 529, A38
- Eckert, D., Gaspari, M., Gastaldello, F., Le Brun, A. M. C., & O’Sullivan, E. 2021, *Universe*, 7, 142
- Euclid Collaboration (Adam, R., et al.) 2019, *A&A*, 627, A23
- Euclid Collaboration (Humphrey, A., et al.) 2023, *A&A*, 671, A99
- Euclid Collaboration (Bhargava, S., et al.) 2025, *A&A*, in press (Euclid Q1 SI), <https://doi.org/10.1051/0004-6361/202554937>, [arXiv:2503.19196]
- Fisher, R. A. 1936, *Ann. Eugenics*, 7, 179
- Franco, M., Casey, C. M., Koekemoer, A. M., et al. 2025, *COSMOS-Web: Comprehensive Data Reduction for Wide-Area JWST NIRCIm Imaging*, [arXiv:2506.03256]
- Franx, M., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2008, *ApJ*, 688, 770
- Friedman, J. H. 2000, *Ann. Stat.*, 29, 5
- Fritz, A., Scodreggio, M., Ilbert, O., et al. 2014, *A&A*, 563, A92
- Gallazzi, A., Charlot, S., Brinchmann, J., & White, S. D. M. 2006, *MNRAS*, 370, 1106
- Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, *ApJ*, 788, 72
- Gladders, M. D., & Yee, H. K. C. 2000, *Astron. J.*, 120, 2148
- Gould, K. M. L., Brammer, G., Valentino, F., et al. 2023, *AJ*, 165, 248
- Gozaiala, G., Finoguenov, A., Tanaka, M., et al. 2019, *MNRAS*, 483, 3545
- Gunn, J. E., & Gott, J. R., III 1972, *ApJ*, 176, 1
- Hao, J., Koester, B. P., Mckay, T. A., et al. 2009, *ApJ*, 702, 745
- Harish, S., Kartaltepe, J. S., Liu, D., et al. 2025, *ApJ*, 992, 45
- Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, *ApJS*, 172, 29
- Hatamnia, H., Mobasher, B., Taamoli, S., et al. 2025, *Large-Scale Structure in COSMOS-Web: Tracing Galaxy Evolution in the Cosmic Web up to $z \sim 7$ with the Largest JWST Survey*, [arXiv:2511.10727]
- Hausman, M. A., & Ostriker, J. P. 1978, *ApJ*, 224, 320
- Hennig, C., Mohr, J. J., Zenteno, A., et al. 2017, *MNRAS*, 467, 4015
- Hung, D., Lemaux, B. C., Cucciati, O., et al. 2025, *ApJ*, 980, 155
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, 690, 1236
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- Ilbert, O., Arnouts, S., Le Floch, E., et al. 2015, *A&A*, 579, A2
- Iovino, A., Petropoulou, V., Scodreggio, M., et al. 2016, *A&A*, 592, A78
- Ito, K., Valentino, F., Farcy, M., et al. 2025, *A&A*, 697, A111
- Jin, S., Sillassen, N. B., Magdis, G. E., et al. 2024, *A&A*, 683, L4
- Kawinwanichakij, L., Papovich, C., Quadri, R. F., et al. 2017, *ApJ*, 847, 134
- Khostovan, A. A., Kartaltepe, J. S., Salvato, M., et al. 2025, *arXiv e-prints* [arXiv:2503.00120]
- Knobel, C., Lilly, S. J., Iovino, A., et al. 2012, *ApJ*, 753, 121
- Kodama, T., & Arimoto, N. 1997, *A&A*, 320, 41
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- Kotulla, R., Fritze, U., Weilbacher, P., & Anders, P. 2009, *MNRAS*, 396, 462
- Kravtsov, A. V., & Borgani, S. 2012, *ARA&A*, 50, 353
- Kroupa, P. 2002, *Science*, 295, 82
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Lovisari, L., Etori, S., Gaspari, M., & Giles, P. A. 2021, *Universe*, 7, 139
- Lu, X., Ye, X., & Cheng, Y. 2024, *Engineering Applications of Artificial Intelligence*, 133, 108107
- Lundberg, S., & Lee, S.-I. 2017, *A Unified Approach to Interpreting Model Predictions*, [arXiv:1705.07874]
- Marini, I., Popesso, P., Dolag, K., et al. 2025, *A&A*, 698, A191
- Martinet, N., Durret, F., Guennou, L., et al. 2015, *A&A*, 575, A116
- Maturi, M., Bellagamba, F., Radovich, M., et al. 2019, *MNRAS*, 485, 498
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- Mei, S., Holden, B. P., Blakeslee, J. P., et al. 2009, *ApJ*, 690, 42
- Menci, N., Rosati, P., Gobat, R., et al. 2008, *ApJ*, 685, 863
- Moneti, A., McCracken, H. J., Hudelot, W., et al. 2023, *VizieR Online Data Catalog*, 2373, II/373. <https://ui.adsabs.harvard.edu/abs/2023yCat.2373..0M>
- Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, *Nature*, 379, 613
- Moutard, T., Malavasi, N., Sawicki, M., Arnouts, S., & Tripathi, S. 2020, *MNRAS*, 495, 4237
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, 490, 493
- Pacifici, C., Kassim, S. A., Weiner, B. J., et al. 2016, *ApJ*, 832, 79
- Paul, S., John, R. S., Gupta, P., & Kumar, H. 2017, *MNRAS*, 471, 2
- Pearson, W. J., Pistis, F., Figueira, M., et al. 2023, *A&A*, 679, A35
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
- Peng, Y.-J., Lilly, S. J., Kovač, K., et al. 2010, *ApJ*, 721, 193
- Popesso, P., Biviano, A., Bulbul, E., et al. 2024, *MNRAS*, 527, 895
- Puddu, E., Radovich, M., Sereno, M., et al. 2021, *A&A*, 645, A9
- Radovich, M., Tortora, C., Bellagamba, F., et al. 2020, *MNRAS*, 498, 4303
- Rieke, M. J., Kelly, D. M., Misselt, K., et al. 2023, *PASP*, 135, 028001
- Romeo, A. D., Kang, X., Contini, E., et al. 2015, *A&A*, 581, A50
- Rudnick, G., von der Linden, A., Pelló, R., et al. 2009, *ApJ*, 700, 1559
- Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, *ApJ*, 785, 104
- Rykoff, E. S., Rozo, E., Hollowood, D., et al. 2016, *ApJS*, 224, 1
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, 859, 11
- Sawicki, M., Arnouts, S., Huang, J., et al. 2019, *MNRAS*, 489, 5202
- Schechter, P. 1976, *AJ*, 203, 297
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Scoville, N., Arnouts, S., Aussel, H., et al. 2013, *ApJS*, 206, 3
- Shimakawa, R., Koyama, Y., Röttgering, H. J. A., et al. 2018, *MNRAS*, 481, 5630
- Shuntov, M., Akins, H. B., Paquereau, L., et al. 2025, *A&A*, 704, A339
- Siudek, M., Mezua, M., & Krywult, J. 2023, *MNRAS*, 518, 724
- Smolčić, V., Novak, M., Bondi, M., et al. 2017, *A&A*, 602, A1
- Spitler, L. R., Labbé, I., Glazebrook, K., et al. 2012, *ApJ*, 748, L21
- Stanford, S. A., Eisenhardt, P. R., & Dickinson, M. 1998, *ApJ*, 492, 461
- Stott, J. P., Smail, I., Edge, A. C., et al. 2007, *ApJ*, 661, 95
- Stott, J. P., Pimblet, K. A., Edge, A. C., Smith, G. P., & Wardlow, J. L. 2009, *MNRAS*, 394, 2098
- Strazzullo, V., Gobat, R., Daddi, E., et al. 2013, *ApJ*, 772, 118

- Strazzullo, V., Daddi, E., Gobat, R., et al. 2016, *ApJ*, 833, L20
- Taamoli, S., Nezhad, N., Mobasher, B., et al. 2024, *ApJ*, 977, 263
- Takey, A., Durret, F., Márquez, I., et al. 2019, *MNRAS*, 486, 4863
- Tanaka, M., Onodera, M., Shimakawa, R., et al. 2024, *ApJ*, 970, 59
- Taniguchi, Y., Kajisawa, M., Kobayashi, M. A. R., et al. 2015, *PASJ*, 67, 104
- Toni, G., Maturi, M., Finoguenov, A., Moscardini, L., & Castignani, G. 2024, *A&A*, 687, A56
- Toni, G., Gozaliasl, G., Maturi, M., et al. 2025, *A&A*, 697, A197
- Wang, T., Elbaz, D., Daddi, E., et al. 2016, *ApJ*, 828, 56
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11
- Weaver, J. R., Davidzon, I., Toft, S., et al. 2023a, *A&A*, 677, A184
- Weaver, J. R., Zalesky, L., Kokorev, V., et al. 2023b, *ApJS*, 269, 20
- Wetzel, A. R., Tinker, J. L., & Conroy, C. 2012, *MNRAS*, 424, 232
- Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, *MNRAS*, 432, 336
- Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, *ApJ*, 735, 86
- Wright, R. H., Sabatke, D., & Telfer, R. 2022, *SPIE*, 12180, 121803P
- Wu, P.-F., van der Wel, A., Gallazzi, A., et al. 2018, *ApJ*, 855, 85
- Zavala, J. A., Casey, C. M., Scoville, N., et al. 2019, *ApJ*, 887, 183
- Zheng, Y., Xu, K., Zhao, D., et al. 2025, *ApJ*, 984, 193
- Ziparo, F., Popesso, P., Finoguenov, A., et al. 2014, *MNRAS*, 437, 458
- ¹⁰ Department of Astronomy, The University of Texas at Austin, 2515 Speedway Blvd Stop C1400, Austin, TX 78712, USA
- ¹¹ Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
- ¹² Department of Physics, University of California, Santa Barbara, Santa Barbara, CA 93106, USA
- ¹³ Cosmic Dawn Center (DAWN), Denmark
- ¹⁴ Department of Physics and Astronomy, University of Hawaii, Hilo, 200 W Kawili St, Hilo, HI 96720, USA
- ¹⁵ Caltech/IPAC, MS 314-6, 1200 E. California Blvd., Pasadena, CA 91125, USA
- ¹⁶ Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA
- ¹⁷ DTU Space, Technical University of Denmark, Elektrovej 327, 2800 Kgs Lyngby, Denmark
- ¹⁸ Department of Physics and Astronomy, University of Kentucky, 505 Rose Street, Lexington, KY 40506, USA
- ¹⁹ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
- ²⁰ Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom
- ²¹ Niels Bohr Institute, University of Copenhagen, Jagtvej 128, DK-2200, Copenhagen, Denmark
- ²² Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France
- ²³ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91001, USA
- ²⁴ Department of Physics and Astronomy, UCLA, PAB 430 Portola Plaza, Box 951547, Los Angeles, CA 90095-1547, USA
- ²⁵ Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
- ²⁶ National Research Institute of Astronomy and Geophysics (NRIAG), Cairo, Egypt
- ²⁷ University of Geneva, 24 rue du Général-Dufour, 1211 Genève 4, Switzerland
- ²⁸ University of Massachusetts Amherst, 710 North Pleasant Street, Amherst, MA 01003-9305, USA
-
- ¹ University of Bologna, Department of Physics and Astronomy “Augusto Righi” (DIFA), Via Gobetti 93/2, I-40129 Bologna, Italy
- ² INAF – Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, I-40129 Bologna, Italy
- ³ Zentrum für Astronomie, Universität Heidelberg, Philosophenweg 12, D-69120 Heidelberg, Germany
- ⁴ Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany
- ⁵ INFN – Sezione di Bologna, Viale Berti Pichat 6/2, I-40127 Bologna, Italy
- ⁶ Department of Computer Science, Aalto University, P.O. Box 15400, FI-00076 Espoo, Finland
- ⁷ Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland
- ⁸ Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
- ⁹ Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191 Gif-sur-Yvette, France

Appendix A: Training-set choice and labeling

A.1. Additional details about the training-set choice

Here we provide additional details on the selection and use of the training set, highlighting its strengths and limitations in the context of our analysis. Firstly, we chose not to train our model directly on the target dataset because doing so would require extracting data from it and splitting it into training and testing subsets, thereby drastically reducing the statistical completeness of the final catalog that we aim to fully analyze. Instead, we train the model on an independent labeled dataset (COSMOS2015) while applying it to the full target sample (COSMOS2025). Using real data from the same field ensures that the model is exposed to the true photometric uncertainties, systematics, blending issues, and observational conditions of our survey, which simulations might fail to fully reproduce.

COSMOS2025 lies entirely within the COSMOS2015 footprint, so some galaxies appear in both catalogs. However, the two surveys differ in instrument sets, depths, PSFs, and photometric methods, resulting in different galaxy characterization, including different rest-frame magnitudes. From the ML perspective, what matters is overlap in feature space, not sky coordinates. We verified that no galaxy has identical or effectively duplicate feature vectors in both datasets, meaning there is no risk of data leakage or overfitting. COSMOS2015 thus provides an independent, realistic training set for our application to COSMOS2025.

The redshift range of the training set extends to $z = 6$, while the target set reaches $z \sim 13$. However, galaxies with $z > 6$ constitute only 0.6% of the target sample and are negligible for our analysis. The main gap in the training set we chose in our approach lies instead in the depth mismatch: the training set is limited to objects brighter than $H = 25$ mag, while the target set extends to 27.3 mag. This forces the model to extrapolate beyond its training domain, where faint objects have larger photometric uncertainties and somewhat broader color distributions. Consequently, classification performance may degrade at the faint end, potentially increasing systematic misclassifications.

To assess the practical impact, we examined galaxies fainter than $H = 25$ in the classified target set, using the widely adopted NUVrJ diagram as a reference. We analyzed (1) the distribution of predicted probabilities and (2) classification metrics in classifying quiescent galaxies⁵ (precision, recall, F1 score) for faint versus bright objects. The probability distribution for faint galaxies was found to be broadly consistent with that of the brighter sample, with a slight increase in intermediate probabilities ($P_{\text{blue}} \sim 70\%$), reflecting model caution in ambiguous cases. Compared to the bright end, the faint sample shows increased precision ($\sim +10\%$) but substantially reduced recall ($\sim -30\%$), leading to a net modest decrease in the F1 score ($\sim -10\%$). This indicates that the classifier becomes more conservative at faint magnitudes: quiescence predictions are generally correct, but a larger fraction of quiescent galaxies are missed. In other words, the model favors purity over completeness in the unrepresented faint regime. We stress that this comparison measures consistency with NUVrJ rather than ground-truth validation, so it does not provide a direct correction for faint-end behavior.

Large uncertainties are in general expected in this regime, given the larger photometric errors at the faint end, and our model's conservative response is appropriate. Importantly, only $\sim 25\%$ of group member galaxies fall in this faint regime (reduc-

ing to $\sim 14\%$ when considering confident associations with $P_{\text{mem}} > 75\%$), limiting its impact on our main scientific goals. In conclusion, when deriving quiescent fractions, the faint end may be characterized by a selection that is purer but less complete, consistent with the intrinsic limitations of the data. Overall, the results indicate that the model effectively copes with fainter, noisier data, becoming more conservative, without compromising the analysis of galaxy population trends.

A.2. Additional details about the training-set labeling thresholds

Here we provide further information about the choice of labeling criteria and thresholds and their z -dependence.

The NUVrJ criterion classifies galaxies using rest-frame $NUV-r$ and $r-J$ colors, identifying as star-forming the galaxies with $NUV-r < 3(r-J) + 1$ or $NUV-r < 3.1$. The method was introduced by [Ilbert et al. \(2013\)](#) as an alternative to the very commonly adopted UVJ diagnostic. It was developed as an alternative with extended validity up to high redshift, given its larger dynamical range, lower sensitivity to uncertainties, and better sampling at $z > 2$. The NUVrJ criterion has been widely tested and employed for galaxy classification up to $z \sim 4$ and beyond (e.g., [Ilbert et al. 2013](#); [Weaver et al. 2023a](#); [Gould et al. 2023](#); [Shuntov et al. 2025](#); [Pearson et al. 2023](#)).

Another variation that complements the NUVrJ criterion is the NUVrK threshold ([Arnouts et al. 2013](#); [Davidzon et al. 2016](#)), which replaces the $r-J$ with the $r-K$ color. Galaxies are classified as star-forming when $NUV-r < 1.37(r-K) + 2.6$ or $NUV-r < 3.15$ or $r-K > 1.3$. The NUVrK criterion, while with similar behavior and z -validity to the NUVrJ, combines the sensitivity of the $NUV-r$ color to the recent star formation history of the galaxy with the sensitivity of the $r-K$ color to stellar aging and dust attenuation, and helps also to better separate the influence of AGNs (e.g., [Siudek et al. 2023](#); [Moutard et al. 2020](#); [Pearson et al. 2023](#)).

As for what concerns the sSFR cut, this has been widely used in the literature by adopting different thresholds, generally between 10^{-10} yr^{-1} (e.g., [Wu et al. 2018](#)) and 10^{-11} yr^{-1} (e.g., [Ilbert et al. 2013](#)). We decided to adopt an intermediate fixed cut at $10^{-10.5} \text{ yr}^{-1}$ which is the same adopted in Euclid ([Euclid Collaboration: Humphrey et al. 2023](#)), following the work of [Bisigello et al. \(2020\)](#). However, it is true that the sSFR threshold to separate quiescent and star-forming galaxies is not uniform and can vary with cosmic time and with galaxy properties (e.g., [Bisigello et al. 2020](#); [Pacifci et al. 2016](#); [Franx et al. 2008](#); [Pearson et al. 2023](#); [Wetzel et al. 2013](#)).

In particular, [Franx et al. \(2008\)](#) have proposed a threshold proportional to the Hubble parameter, $H(z)$, where quiescent galaxies have $\text{sSFR} < 0.3H(z)$. In an analogous way, [Pacifci et al. \(2016\)](#) have more recently proposed a threshold dependent on the inverse of the age of the Universe, $t_U(z)$, with quiescent galaxies having $\text{sSFR} < 0.2/t_U(z)$. Figure A.1 shows the density of the selected COSMOS2015 star-forming (blue) and quiescent (red) galaxies according to the standard NUVrJ diagnostic in the sSFR vs. z diagram. The solid black line indicates the fixed threshold in sSFR we applied in this work, while the dashed lines mark the time-dependent threshold suggested by [Pacifci et al. \(2016\)](#), in green, and the one by [Franx et al. \(2008\)](#), in purple. The fixed and time-dependent thresholds diverge significantly at $z > 2$, when low-sSFR galaxies are, however, expected to be much rarer. The thresholds become more consistent with each other around the crossing-point at $z \sim 0.7 - 0.9$ where most of the galaxies in COSMOS

⁵ the performances for star-forming galaxies, which are the majority class, are not relevant here because they do not vary with magnitude, remaining stable above $\sim 96\%$.

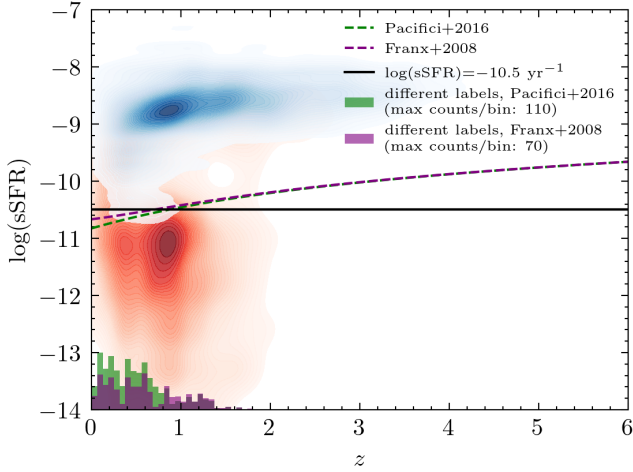


Fig. A.1. Density of COSMOS2015 star-forming (blue) and quiescent (red) galaxies (according to NUVrJ classification) in the sSFR– z plane. The solid black line shows the fixed sSFR threshold used in this work; dashed lines indicate the time-dependent thresholds of Pacifici et al. 2016 (green) and Franx et al. 2008 (purple). Thresholds diverge at $z > 2$, but converge near $z \sim 0.7$ – 0.9 , where most COSMOS galaxies lie. The bottom histograms show the redshift distribution of galaxies that would change classification under the two time-dependent thresholds, as reported in the legend.

are (e.g., Iovino et al. 2016), as visible in the plot. At the bottom of the panel in Figure A.1, we also show the distribution in redshift of all galaxies that would change their label if a time-dependent threshold was adopted. Counts for specific changes between the three classes of galaxies are also reported in Table A.1. Two main conclusions can be drawn from these results: (1) the impact of a time-dependent sSFR is significant only below the crossing-points ($z \lesssim 0.7$) and affecting a small fraction of galaxies (overall ~ 0.2 – 0.4% of the galaxy sample); (2) this effect is mainly manifested in slightly reducing the number of quiescent galaxies (~ 1 – 2%), and the number of green valley galaxies (~ 1 – 8%) in favor of a few hundreds (~ 0.1 – 0.3%) of new examples of star-forming galaxies (especially when applying Pacifici et al. (2016)’s threshold). Given that the effect is limited in redshift, affecting only a small fraction of the galaxies, and that it reduces examples for the minority class (quiescent galaxies) and increases those for the majority class (star-forming galaxies) close to the decision boundary, it is not expected to drastically impact the learning process and performance. In the redshift range we analyzed, the simple choice of $10^{-10.5} \text{ yr}^{-1}$ to separate the two populations is therefore well justified and closer to time-dependent relations with respect to, for instance a threshold at 10^{-11} yr^{-1} , which would largely overestimate the number of star-forming galaxies.

Table A.1. Number of galaxies which would have been labeled differently if the threshold by Franx et al. (2008) / Pacifici et al. (2016) was considered instead of a fixed threshold at $10^{-10.5} \text{ yr}^{-1}$. Rows indicate labels assigned in this study, and columns indicate new labels that would have been assigned with time-dependent thresholds.

	Quiescent	Star-forming	Green
Quiescent	–	255 / 458	27 / 84
Star-forming	19 / 17	–	204 / 123
Green	42 / 26	229 / 427	–

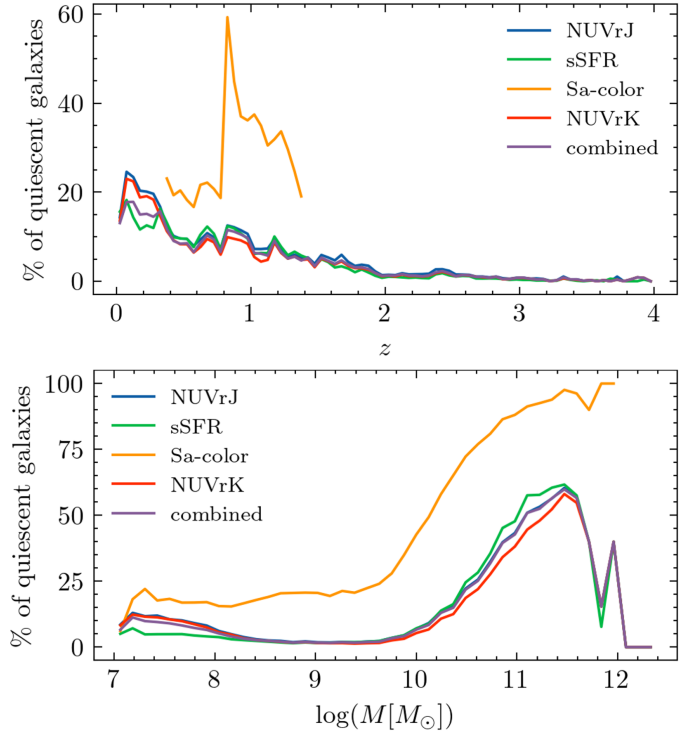


Fig. A.2. Variation in bins of redshift (upper panel) and galaxy stellar mass (lower panel) of the percentage of classified quiescent galaxies over the total, for the NUVrJ color-color cut (blue), the sSFR threshold (green), the Sa-color method (orange), and the NUVrK color-color cut (red). The purple line shows the final combined labeling we adopted for the training and testing sets.

Finally, for the Sa-color evolution diagram (Andreon 2006; Radovich et al. 2020), a few more aspects have to be taken into account. The classification method is interesting because it uses apparent colors and not rest-frame colors. The use of an evolving Sa-model is what makes it redshift-consistent: each galaxy is classified as quiescent or star-forming, given the typical color of a Sa-type galaxy (a spiral galaxy with properties close to those of lenticular and elliptical galaxies according to Hubble’s morphological classification scheme) at that specific redshift. However, the main limitation of this method is in the model-dependence itself and the fact that the chosen color must bracket the 4000\AA break and therefore change accordingly as redshift increases.

In this study, we modeled the evolution of Sa-colors with the GALEV evolutionary synthesis interface (Kotulla et al. 2009), choosing typical Sa parameters and formation redshift. We decided to use this criterion only over a limited redshift range; that is, $0.38 \leq z \leq 1.4$. The lower bound is due to the absence of g -band photometry in the COSMOS2015 catalog caused by poor seeing (Laigle et al. 2016), which would be the blue-bracket of the 4000\AA break at $z < 0.38$. The upper limit is instead given by the 4000\AA break shifting into the near-infrared. Given that the uncertainties of our model increase in this regime, we chose to be conservative and have the ML algorithms rely only on the other three methods at higher redshifts.

Figure A.2 shows the variation with redshift (upper panel) and with galaxy stellar mass (lower panel) of the percentage of classified quiescent galaxies over the total, for the four methods and for the final combined labeling we adopted (purple). As expected, even if quiescent galaxies are only 7% of the overall sample, the percentage over the total is very redshift- and

mass-dependent. The panels also clearly illustrate the strong consistency between the color–color cuts and the sSFR threshold across all redshifts and stellar masses, while highlighting a noticeable overestimate in the number of quiescent galaxies derived from the Sa-color method compared to the other approaches. The peak around $z \sim 0.8$ is likely influenced by a mismatch between our model and the observed colors when transitioning from the i to the z band. Apart from this effect, the number of quiescent galaxies remains systematically 15 – 20% higher across nearly all redshifts and masses – an excess over color-color cuts similar to that recently reported by Asadi et al. (2025). Interesting insights on this discrepancy could be gained by investigating a complete and spectroscopically characterized sample of green valley galaxies, which comprise most of the ambiguous and transitional cases in the classification.

In our specific application, as shown in Figure A.2, the resulting overall galaxy classification (purple line) is, by construction, driven by the consensus among methods and is therefore effectively constrained by the other three criteria. Nevertheless, the Sa-color method provides valuable information in ambiguous or edge cases and is particularly helpful when UV photometry is missing. Training the ML models on a labeled dataset that integrates multiple independent and complementary methods, while basing the labeling on their consensus, is one of the key strengths of our approach.

Appendix B: Cylinder method and purity weighting of the quiescent fraction

In Sect. 4.2, we have introduced a new method to characterize AMICO galaxy group members that (nearly) eliminates the model-dependence (and therefore the dependence on magnitude and position), while retaining the information of the redshift probability distribution. In short, this method exploits the concept of considering a volume around each detection and subtracting the density of field galaxies at the corresponding redshift to remove field contamination. Two key requirements come with this formalism: the appropriate choice of the cylinder size and the handling of negative densities, since these may arise from the group–field subtraction.

The choice of the depth of the cylinder is motivated by the typical photometric redshift uncertainty in the redshift range of interest. The depth is variable and redshift dependent, set to $\pm 0.01(1+z)$, and allows for a consistent sampling of members along the line of sight. Since galaxies are weighted by their redshift probability distribution, the exact choice of depth is generally less constraining and less impactful than the choice of the sky radius.

For choosing the cylinder sky radius, we started from an initial guess of $r = 0.5$ Mpc/h, and tested both r and $2r$ as fixed cylinder radii. The differences in the resulting measurements and their richness dependence suggested that an adaptive radius, varying with the distribution of galaxies in each group – and therefore tracing its richness and spatial extent – would be a more robust and physically motivated choice. To determine this adaptive radius, we computed the RMS of all galaxies associated with a group, weighting each galaxy by its membership probability. The distribution of the resulting RMS radii and their relation to richness (λ_*) are shown in Fig. B.1. Although the top panel indicates that 0.5 Mpc/h was a reasonable initial guess, the adaptive radius consistently captures the group region across different group sizes, as demonstrated by its trend with richness (bottom panel).

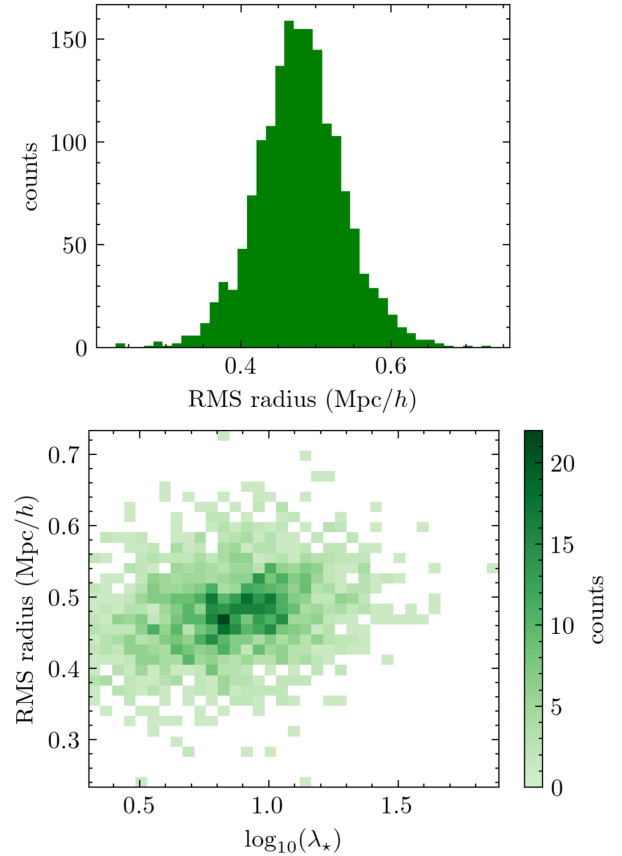


Fig. B.1. Top panel: distribution of RMS radii in Mpc/h for the studied group sample. Bottom panel: trend of RMS radii with richness, $\log_{10}(\lambda_*)$, shown as counts per bin.

As mentioned above, the subtraction of average field-galaxy densities from the group densities within the cylinder can theoretically produce negative densities (see Eq. 4). In other words, in the ratio of Eq. 5: (1) the numerator, ρ_{red} , can be negative (the density of red galaxies is higher outside the cylinder than inside, i.e., a particularly blue group); (2) the term ρ_{blue} can be negative (the density of blue galaxies is higher outside the cylinder than inside, i.e., a particularly red group); (3) the denominator can be negative overall (the group appears as an underdensity relative to the global field, but overdense with respect to the local field).

Cases (1) and (2), affecting a total of about 800 groups, are physically meaningful. We therefore flagged them and included them in our analysis by replacing negative densities with close-to-zero positive constants to avoid divergences. Case (3) affects only 10 low-richness groups (when using the adaptive RMS radius). If real, such systems might correspond to overdensities embedded within locally underdense regions, such as filament edges or voids. However, these detections are excluded from the present analysis, as the cylinder-based estimate cannot be determined reliably.

Figure B.2 shows the weighted quiescent fraction as a function of redshift for different richness bins using 0.5 Mpc/h (left column), 1.0 Mpc/h (middle column), and the adaptive RMS radius (right column), comparing the results obtained by discarding all groups with any case of negative densities (top row) to those obtained after correcting negative densities in cases (1) and (2) as described above (bottom row). The RMS-based quiescent fraction with these corrections is the final method adopted

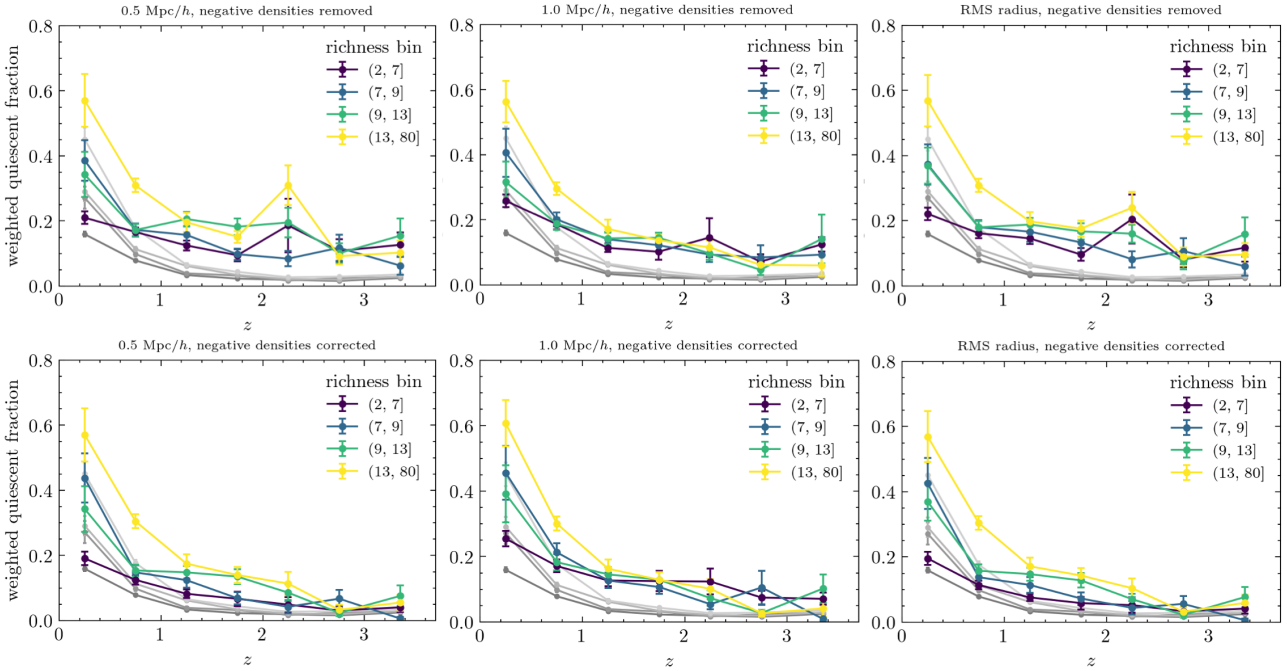


Fig. B.2. Purity-weighted quiescent fraction as a function of redshift for different richness bins, using fixed radii of 0.5 Mpc/h (left), 1.0 Mpc/h (middle), and the RMS radius (right). In the top panels, we discard all systems with negative densities, while in the bottom panels, we apply the corrections described in the text. Grayscale trend lines are reported using the pure-membership method in the same bins, for comparison. The RMS-based measurement with corrected negative densities is the method adopted in Sect. 4.2 and shown in Fig. 5.

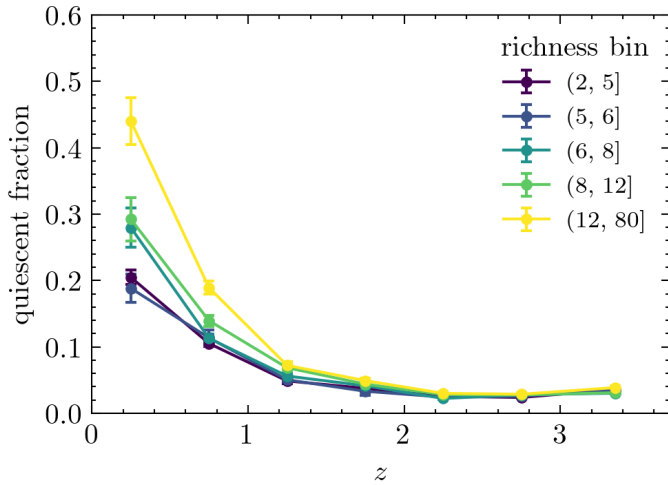


Fig. B.3. Same as in Fig. 4 but without applying purity weighting.

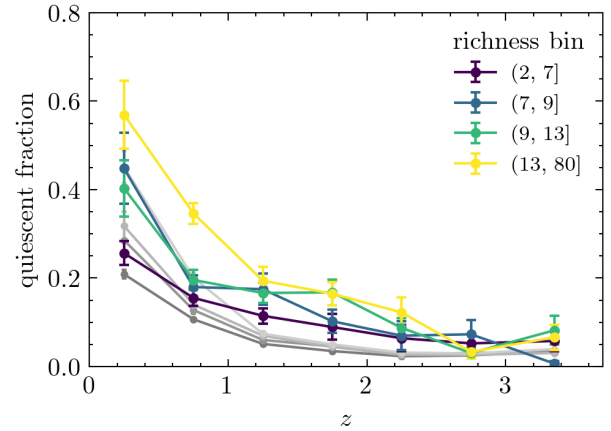


Fig. B.4. Same as in Fig. 5 but without applying purity weighting.

for the results presented and discussed in Sect. 4.2 and shown in Fig. 5.

B.1. Quiescent fraction purity-weighting

For completeness, Figs. B.3 and B.4 present the same measurements shown in Figs. 4 and 5, but without applying purity weighting in the corresponding bins for the pure-membership and cylinder methods, respectively. As expected, given the overall high purity of the COSMOS-Web catalog, all global trends are preserved. The purity weighting mainly influences the lowest-richness bins, which are naturally more prone to lower reliability and purity.