Boletim da Sociedade Astronômica Brasileira, **34**, no. 1, 214-218

Galaxy cluster finding and membership assignments on the mini J-PAS survey

L. Doubrawa[1], E. S. Cypriano[1], & A. Finoguenov[2]

[1] Departamento de Astronomia, Instituto de Astronomia, Geofísica e Ciências Atmosféricas da USP, Cidade Universitária, 05508-900, São Paulo, SP, Brazil e-mail: lia.doubrawa@usp.br
[2] Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

Abstract. The wide-field photometric survey Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS) is expected to map the Universe with excellent accuracy. In this work, we present preliminary results of galaxy cluster detection obtained with PZwav using the photo-z from the mini J-PAS survey as preparation for J-PAS. Our optical catalog has 144 galaxy clusters distributed over a redshift range of $0.05 < z < 0.8$. With a new probabilistic membership estimator, we characterized the cluster candidates by measuring richness, total optical luminosity, and total stellar mass. As the mini J-PAS has an overlapping area with the AEGIS fields, we produced an X-ray catalog from Chandra data at the optical detections. The resulting catalog has 41 galaxy clusters within a mass range of $8.6\,10^{12}$ to $1.1\,10^{14}$ $M_\odot$. Scaling relations showed that all three observable parameters have a good performance. Total luminosity has a lower intrinsic scatter of 0.053 ± 0.018, but can be compared to richness and stellar mass within error-bars, both with 0.062 ± 0.021. These preliminary results demonstrate the quality of the data obtained with the mini J-PAS, highlighting the potential of the J-PAS survey.

Resumo. Espera-se que o levantamento fotométrico avalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS) mapeie o Universo com excelente precisão. Neste trabalho, apresentamos resultados preliminares de detecção de aglomerados de galáxias obtidos com PZwav usando o redshift fotométrico do levantamento mini J-PAS como uma preparação para J-PAS. O catálogo óptico apresenta 144 aglomerados de galáxias distribuídos em uma faixa de redshift de $0.05 < z < 0.8$. Com um novo estimador de riqueza probabilístico, caracterizamos os candidatos a aglomerados de galáxias medindo riqueza, luminosidade óptica total e massa estelar total. Como o mini J-PAS tem uma área de sobreposição com o campo AEGIS, produzimos um catálogo de raios-X a partir dos dados do Chandra com base nas detecções ópticas. O catálogo resultante é composto de 41 aglomerados de galáxias com massas que variam de $8.6\,10^{12}$ a $1.1\,10^{14}$ $M_\odot$. Relações de escala mostraram que todos os observáveis apresentam um bom desempenho. Luminosidade óptica total apresenta a menor dispersão intrínseca de 0.053 ± 0.018, mas é comparável à riqueza e massa estelar, ambos com 0.062 ± 0.021. Esses resultados preliminares demonstram a qualidade dos dados obtidos com o mini J-PAS, destacando o potencial do J-PAS.

Keywords. Galaxies: clusters: general – Galaxies: groups: general – Methods: statistical

1. Introduction

The Universe is composed of large structures such as galaxies, groups and galaxy clusters, and low-density regions (Bahcall 1988). Currently, the most accepted model for the formation scenario is the hierarchical formation theory, where low mass structures merge earlier, giving rise to larger systems (Press & Schechter 1974; Blumenthal et al. 1984). In this process, galaxy clusters are the most significant virialized gravitationally bound structures in the Universe. Due to its sensitivity to the underlying cosmology, studies of the cluster abundance and its evolution provide a sensitive probe of the cosmological parameters (e.g. Voit 2005; Allen et al. 2011).

In the last decades, several surveys have been aiming the galaxy cluster detection on different wavelengths, such as X-rays, optical and millimetric bands. In the near future, wide-field surveys such as The Javalambre Physics of the Accelerating Universe Astrophysical Survey (J-PAS, Bonoli et al. 2021) will provide a unique dataset.

There are numerous methods to build a galaxy cluster catalog from data. Those methods differ due to the computational algorithms and the in-built properties that define a galaxy cluster. Usually, optical catalogs utilize spectroscopic information to determine the galaxy clusters' sky positions and make assumptions about their galaxy members, as in Rykoff et al. (2014). However, the refined filter system used for estimating the new photometric redshifts allow us to transition between spectroscopic and photometric redshifts. On J-PAS, for example, the expected errors in redshift are comparable to spectroscopic estimates Benitez et al. (2014).

In this work, we briefly describe PZwav (Euclid Collaboration et al. 2019). This algorithm takes advantage of photometric surveys by identifying overdensities using all available photometric information in the form of probability density functions (PDFs). We introduce a new probabilistic membership estimator that also uses PDFs and allows us to characterize the optical catalog with richness estimates, total optical luminosity, and total stellar mass. Present mass estimates from X-ray analysis based on the PZwav detections. And discuss the best mass proxy. The relationship that is essential for cosmological studies once masses can't be obtained directly from observations.

2. PZwav algorithm

PZwav (Euclid Collaboration et al. 2019; Werner et al. 2022) is a density-based algorithm that only requires information such as sky coordinates of galaxies, photometric redshifts, and magnitudes.

The general idea is to find galaxies that are close enough in plane-of-the-sky and redshift space. For this, the data is divided into redshift slices ($dz = 0.01$) and projected. These pro-

Table 1. Final galaxy cluster catalog and columns description. It is a combination of the PZwav algorithm output and the calculated parameters obtained with the probabilistic membership estimator (identified by *).

ID	Cluster identification number
RA	Position in the sky in R.A.
DEC	Position in the sky in Dec.
z	Estimated redshift
SNR	Detection signal-to-noise
rich_mag-21	Estimated richness*
err_mag-21	Uncertainty in richness*
L_λ	Optical luminosity weighted by richness*
L_λ_err	Uncertainty in optical luminosity*
M^*_λ	Stellar mass weighted by richness*
M^*_λ_err	Uncertainty in stellar mass*

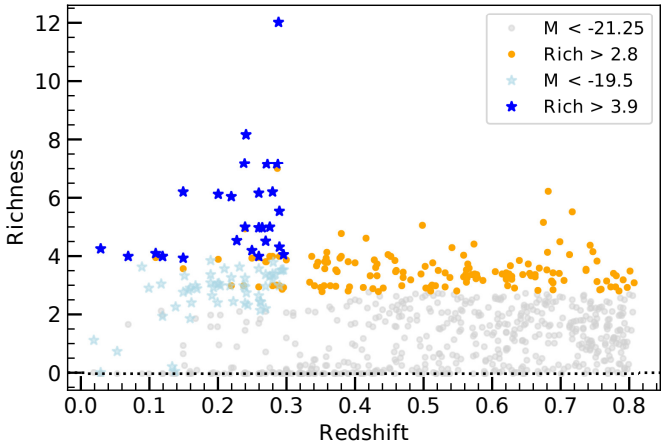

Figure 1. Richness results through all redshift range. In light blue (complete sample) and blue (Rich > 3.9) for the group regime ($M < -19.5, z < 0.3$); and in gray (complete sample) and orange (Rich > 2.9), for the cluster regime ($M < -21.25$).

jections create 2D galaxy density maps, whose contribution of each galaxy is weighted by integrating the $P(z)$ over the limits of the bin. A galaxy cluster candidate is then identified if the density peak rises above the noise threshold and has galaxy typical cluster scales, larger than 400 and smaller than 1400 kpc. If the conditions are satisfied, the redshift of the candidate is calculated as the 2σ-clipped median z_{phot} computed from the galaxies lying within $\Delta z = 0.12$ and $R = 500$ kpc. To avoid possible double counting for clusters passing through a collision event, the code allows setting values as the minimum distance between two substructures. We use as merging parameters $dr_{lim} = 1500$ kpc and $dz_{lim} = 0.03$.

In table 1 we describe the columns given by PZwav alongside the ones that we use to characterize the detection (marked with a *, more details below). The "ID" that identifies the cluster detections, sky positions "RA, DEC" corresponding to the peak location of each detected overdensity, the calculated photometric redshift "z" and "SNR" signal-to-noise ratio that is defined as the amplitude of the highest peak in the density maps in units of the noise level.

3. Probabilistic membership

One limitation of the PZwav algorithm is that it does not return a list of likely member galaxies for each cluster candidate. This is an interesting property that can be used to characterize the object, for example, by measuring the richness. In this work we present a probabilistic membership code with this capability, that is based on the Density-Based Clustering Based on Hierarchical Density Estimates (HDBSCAN, Campello et al. 2014).

HDBSCAN is a clustering algorithm that allows the identification and removal of spatially sparse structures, depending on the minimum number of neighbors and an optimal distance. The distance-related parameter (η) is calibrated by the code trough varying and integrating η in a search for the best stability value. Therefore we only need to define the minimum size of galaxies ($n_{\rm gal}$) expected for the cluster. Once a fixed value is not adequate to describe the variation between groups and galaxy clusters, $n_{\rm gal}$ can be calibrated using mock simulations of the survey.

We chose to run HDBSCAN over a characteristic radius (R_c), based on the galaxy's radial density profile for each cluster candidate. It is defined as a sudden drop in density, when moving away from the center of the cluster, identifying and delimiting their borders. This value is usually 0.6 R_{200} (A more detailed discussion can be found in Doubrawa et al. in prep.).

To estimate the richness, we follow a series of steps: given a cluster candidate with redshift, sky coordinates, and R_c, we select galaxies around the cluster center until R_c. Then, for each galaxy, we draw a random redshift value based on its redshift PDFs. We calculate the cluster velocity dispersion with a 3-σ clipping process, to avoid excessively high/low redshift values. And finally, Run HDBSCAN. The structure with most galaxies should be the cluster in question.

As HDBSCAN does not assign a probabilistic membership, we repeat this procedure N times. Thus, the probability of each galaxy being a member is the number of times that the galaxy is included as a member over N, $P_{mem} = N_{mem}/N$. The richness is the sum of all probabilities.

For the error estimation on the richness, for example, we run the same steps over 50 randomly distributed points, at the candidate redshift. The error is calculated as the standard error of the mean.

4. Results and discussion

We use the photometric data from the mini J-PAS survey (Bonoli et al. 2021), of ~ 1 square degrees, that was observed within the AEGIS field Davis et al. (2007). This is a small sample of J-PAS that allows for testing the scientific potential of the survey. Mini J-PAS typical errors in photometric redshifts are within $\sigma_{NMAD} = 0.013$ (López-Sanjuan et al. 2021).

Below we present the galaxy cluster catalog obtained by running the PZwav algorithm, the mass estimates from X-rays analysis, and the resulting scaling relations.

4.1. Galaxy cluster catalog

Running the PZwav algorithm over the mini J-PAS data, we obtained a catalog with 650 galaxy clusters candidates in the redshift range of $0.05 < z < 0.8$. Contamination and completeness of the catalog are common concerns when dealing with optical detection algorithms. PZwav studies within mock sky areas, as in Werner et al. (2022) show that a cut in signal-to-noise of SNR > 3.3 achieves the best agreement between completeness and purity.

Here, we deal with contamination using a different approach. Following Klein et al. (2018), we create a richness estimator that allows us to remove cluster candidates below a given threshold. The method compares the PZwav catalog with random distributions of sky positions and redshifts. Each random point has

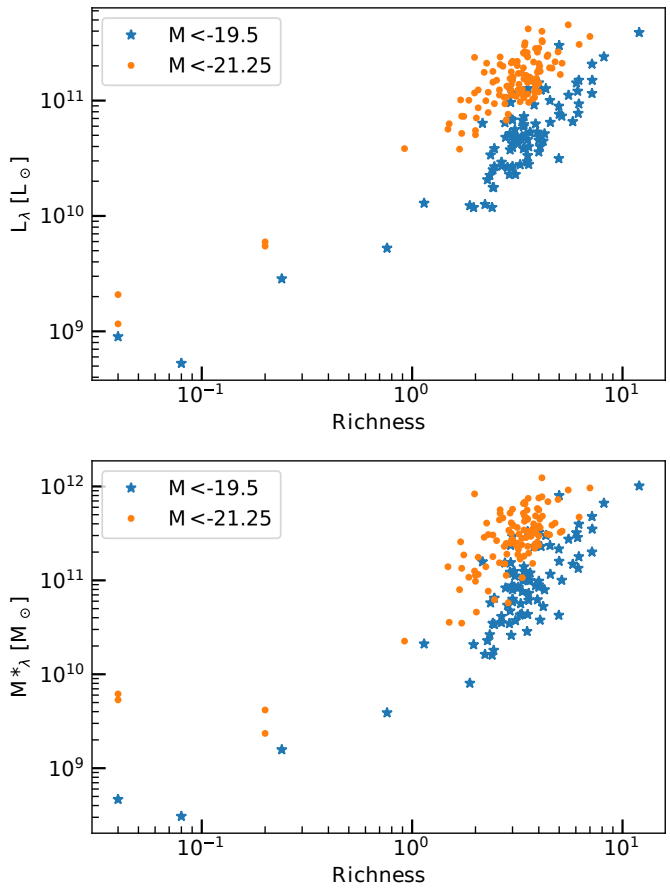

Figure 2. Top panel: Richness and Optical luminosity, L_λ, distributions. Bottom panel: richness and total stellar mass, M^*_λ. The colors differentiate the cluster (orange) and group regimes (blue).

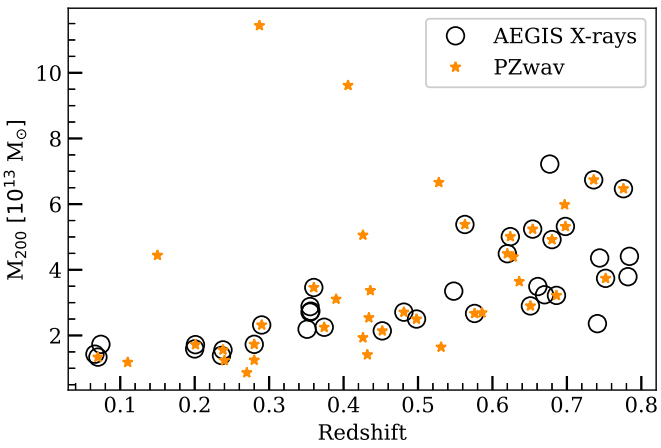

Figure 3. Erfanianfar et al. (2013) galaxy cluster catalog from the AEGIS field (open circles) and the sample of optical galaxy cluster candidates with an X-ray counterpart (orange markers), within $0.05 < z < 0.8$. The total mass was inferred from Leauthaud et al. (2010).

estimates of richness. For each PZwav candidate, we count the number of random points within $\Delta z = 0.05$ the cluster redshift, that has a richness lower than the candidate. This provides a numerical probability of random points with richness lower than the PZwav detected galaxy clusters. We remove potential contamination by choosing a probability threshold of 95%.

Another concern related to optical surveys and richness estimates is the incompleteness of the galaxy sample. Here, we chose to work with two different absolute magnitudes cuts: the "cluster regime" that uses a magnitude cut of $M < -21.25$ and includes objects visible until redshift $z < 0.8$; And the "group regime" with $M < -19.5$, valid only for $z < 0.3$ (Zheng & Shen 2021).

Applying the probability threshold for both regimes, we have a richness cut of 2.9 and 3.9 for the cluster and group regimes, respectively. Different constraints for each regime are interesting once a galaxy group can be affected by the magnitude cut and erroneously pointed out as contamination. Therefore, the final catalog presents 144 galaxy cluster candidates. Figure 1 shows the richness, calculated by the probabilistic membership estimator, and the redshift distribution of the PZwav catalog, before (lighter colors) and after the richness cuts. Blue markers indicate the group regime and orange ones the cluster regime.

The probabilistic estimator returns the probability of each galaxy belonging to a certain cluster. Hence, we are able to calculate other proprieties that characterize the cluster sample.

Mini J-PAS has magnitude information for several bands, here we use the r-band to calculate the total optical luminosity weighted by the galaxy pertinence, $L_\lambda = \sum L_i P_i = \sum 10^{0.4[4.42-R_i]} P_i$. Where 4.42 is the solar absolute magnitude in r-band (Binney & Merrifield 1998), and R_i is the i-th galaxy absolute magnitude in the same band. Studies of the same area also permitted estimates of the galaxy's stellar masses with great accuracy González Delgado et al. (2021). Those values allow us to estimate the total stellar mass similarly as described above, $M^*_\lambda = \sum M^*_i P_i$, with M^*_i indicating the i-th galaxy stellar mass.

The resulting values can be found in Figure 2. The top panel highlights the relationship between richness and L_λ, while the bottom panel presents the richness and stellar mass. For both panels, blue points indicate the group regime, and orange ones are the cluster regime. We can see a small gap between the lines due to the different richness values, but a similar behavior in scattering.

4.2. X-rays analysis

As mentioned before, mini J-PAS has an overlapping area with the AEGIS field. An interesting study done in the same area by Erfanianfar et al. (2013) revealed 52 diffuse X-ray emissions cataloged as galaxy clusters. The detailed analysis provided several useful information about the sources, for example, the mass of the structures.

Matching our optical to the X-ray catalog could provide clues to the mass of the objects found by PZwav and would also allow us to make scaling relations. But only a fraction of the X-ray sources were located within the redshift range, $0.05 < z < 0.8$, and the same survey limits. This resulted in 36 clusters for the comparison and only 22 real matches. A match is defined as the maximum center distance between the optical and the X-ray counterparts of 0.5 Mpc, and a redshift difference of $\Delta z = 0.05$.

To improve the comparison, we repeated the analysis done by Erfanianfar et al. over the Chandra mosaic of the AEGIS field, lowering the detection threshold of 4σ to 3σ. In order to contribute to the background subtraction of the new sources, we included the XMM-Newton mosaic observations in the overlap-

Table 2. Linear regression best fitting values. The mass-observable model is described by Equation 1. L_λ and M^*_λ are given in units of $L_\odot$ and $M_\odot$.

Proxy	α	β	ϵ
λ	12.95 ± 0.30	0.82 ± 0.55	0.062 ± 0.021
L_λ	6.78 ± 2.53	0.59 ± 0.23	0.053 ± 0.018
M^*_λ	9.69 ± 2.21	0.32 ± 0.19	0.062 ± 0.021

ping area. The steps resulted in a catalog with 41 X-rays sources with a matched optical counterpart.

The X-ray source identification procedure takes into account the PZwav optical catalog. Initially, we search for matching X-ray sources within 0.5 Mpc. To avoid multiple combinations by chance, we followed the same Klein et al. prescription. Once applied the probabilistic threshold cut, we remove the optical groups that were matched to X-ray sources by chance. After this step, X-ray sources with more than one counterpart were analyzed. If one of the multiple counterparts has a substantially larger richness, exceeding by a factor of 1.5, the match is accepted, if not it is removed from the final catalog. This procedure guarantees that the contamination to X-ray flux from the multiple counterparts will be less than 30%. A value lower than the expected statistical error of the newly identified sources. For each source, we compute the rest-frame X-ray luminosity as described in Erfanianfar et al. (2013), following the same aperture radius for total flux correction and K-correction. The masses were deduced from Leauthaud et al. (2010) weak lensing calibration. Figure 3 shows the mass distribution along the redshift of the resulting X-ray catalog, as orange markers, and the Erfanianfar et al. catalog that has an overlapping area with mini J-PAS, as open circles. Besides the 22 matches between the catalogs, we introduce 19 new sources. We are able to recover structures within a mass range of $8.6\,10^{12}$ to $1.1\,10^{14}$ $M_\odot$. We highlight the detection of a galaxy group at $z = 0.75$ with a mass of $3.7\,10^{13}$ $M_\odot$. The identification of such low masses groups at high redshifts, using photometric data, is only possible due to the quality of the sample combined with PZwav's capabilities.

4.3. Scaling relations

As previously mentioned, masses can't be obtained directly from observations. One approach is a search around the observational properties that can be correlated with mass, i.e. the relationship between the observable (optical richness or X-ray signal) and the mass. This relationship is calibrated for a limited number of objects and then applied to the full sample.

With our X-ray mass estimates and mass proxies, i.e. richness, L_λ and M^*_λ, calculated with the probabilistic membership estimator, we are able to derive these relations. This is done using LINMIX, a linear regression procedure with Bayesian approach (Kelly 2007), that performs the minimization process taking into account errors in both parameters: X-rays mass estimates and proxies. The relation is modeled as,

$$log(M_{200}) = \alpha + \beta\, log\,(O) + \epsilon \quad (1)$$

where α and β are the coefficients, O is the mass proxy and ϵ the intrinsic random scatter about the regression. The best-fit parameters are given in Table 2, and results are summarized in Figure 4.

In the richness-mass relation, the observed richness range is relatively small, once data seems concentrated within 2.5 to 4.1. This distribution is reflected in the slope error which is the largest of all mass proxies. Yet, statistically, all of them exhibit a comparable intrinsic scatter, so richness cannot be discarded.

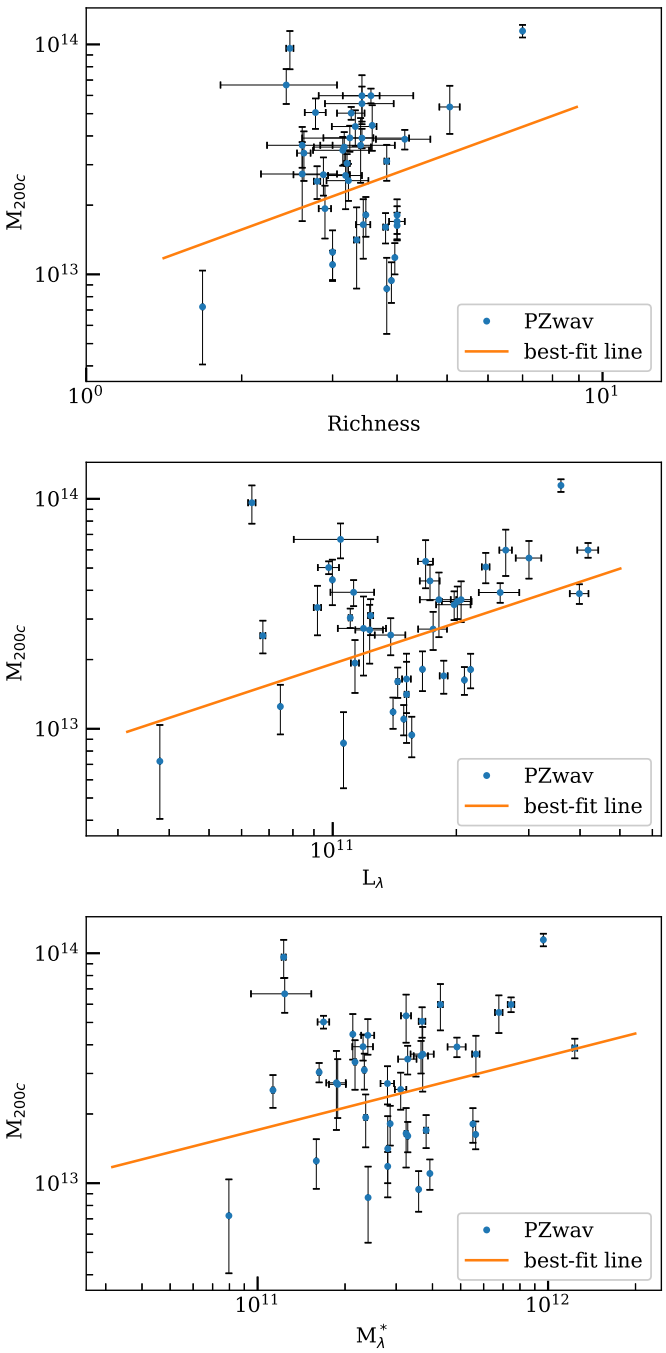

Figure 4. Scaling relations between mass and observable obtained with PZwav optical galaxy cluster candidates which has an X-ray counterpart. Top panel: M_{200} and richness distributions and estimated errors. Middle panel: M_{200c} and L_λ. Bottom panel: M_{200c} and M^*_λ. The orange lines show the best-fitting results. Coefficients can be found in Table 2.

One interesting test would be to make combinations of parameters to see if a better mass estimate could be achieved. The total optical luminosity is an interesting parameter as it presents the lowest ϵ. The excellent coverage of the filter system from mini J-PAS, as also all J-PAS survey, provides good magnitude estimations. As L_λ depends mainly on magnitudes and galaxy probabilities, eventual systematic biases due to the membership analysis may be lower than the expected statistical error. M^*_λ is also a valuable option, even presenting a slightly larger scatter than L_λ, once it provides a characterization of the galaxy clus-

ters candidates in terms of physical properties, i.e., the stellar mass. Besides our effort in increasing our limited number of objects with mass estimates, it is important to highlight that we are working with small galaxy groups populated with a low number of bright galaxies. Some results may be biased due to the sample selection. We will investigate the relations further as more data is available.

5. Conclusion

We created a sample of 144 galaxy cluster candidates within the redshift range of $0.05 < z < 0.8$. The candidates were detected by the cluster finder algorithm PZwav using the photometric information from the mini J-PAS survey. As contamination is a valuable concern on optical surveys we applied a decontamination process that allows removing possible detections by chance. We analyzed our catalog in two different magnitude regimes: the cluster regime with an absolute magnitude cut of $M < -21.25$, and the group regime with $M < -19.5$. This avoids an incorrect classification as "contamination" for small galaxy groups affected by the magnitude cut. We described a new probabilistic membership estimator, based on HDBSCAN, that identifies and attributes probabilities for each galaxy of being a member of a galaxy cluster. Using sky positions and photo-zs we estimated properties such as richness, total optical luminosity, and total stellar mass. Using the Chandra mosaic of the AEGIS field alongside XMM-Newton mosaic observations, we identified 41 optically detected clusters with an X-ray counterpart, 19 of them being newly detected sources. Weak lensing calibrations from Leauthaud et al. (2010) for the same area permitted to infer masses. This analysis showed that PZwav was recovering structures within $8.6\,10^{12}$ to $1.1\,10^{14}$ $M_\odot$. With the resulting catalog, we fitted mass-observable relations. We showed that all mass proxies present a similar performance. Optical luminosity is an interesting observable parameter, with the lowest intrinsic scatter of 0.053 ± 0.018, in comparison to richness and stellar mass, both with a slightly larger value of 0.062 ± 0.021. These preliminary results, down to the galaxy group regime, demonstrate the quality of the data obtained with the mini J-PAS survey, highlighting the potential of the J-PAS survey.

Acknowledgements. This work made use of the computing facilities of the Laboratory of Astroinformatics (IAG/USP, NAT/Unicsul), whose purchase was made possible by the Brazilian agency FAPESP (grant 2009/54006-4) and the INCT-A. ESC acknowledges the support of the funding agencies CNPq (309850/2021-5) and FAPESP (2019/19687-2). This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001.

References

Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409. doi: 10.1146/annurev-astro-081710-102514

Bahcall, N. A. 1988, ARA&A, 26, 631. doi:10.1146/annurev.aa.26.090188.003215

Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv:1403.5237

Binney, J. & Merrifield, M. 1998, Galactic astronomy / James Binney and Michael Merrifield. Princeton, NJ : Princeton University Press, 1998. (Princeton series in astrophysics) QB857 .B522 1998

Blumenthal, G. R., Faber, S. M., Primack, J. R., et al. 1984, Nature, 311, 517. doi:10.1038/311517a0

Bonoli, S., Marín-Franch, A., Varela, J., et al. 2021, A&A, 653, A31. doi: 10.1051/0004-6361/202038841

Campello, R., Moulavi, D. & Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. *Advances In Knowledge Discovery And Data Mining*. pp. 160-172 (2013)

Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJ, 660, L1. doi:10.1086/517931

Erfanianfar, G., Finoguenov, A., Tanaka, M., et al. 2013, ApJ, 765, 117. doi:10.1088/0004-637X/765/2/117

Euclid Collaboration, Adam, R., Vannier, M., et al. 2019, A&A, 627, A23. doi:10.1051/0004-6361/201935088

González Delgado, R. M., Díaz-García, L. A., de Amorim, A., et al. 2021, A&A, 649, A79. doi:10.1051/0004-6361/202039849

Kelly, B. C. 2007, ApJ, 665, 1489. doi:10.1086/519947

Klein, M., Mohr, J. J., Desai, S., et al. 2018, MNRAS, 474, 3324. doi: 10.1093/mnras/stx2929

Leauthaud, A., Finoguenov, A., Kneib, J.-P., et al. 2010, ApJ, 709, 97. doi: 10.1088/0004-637X/709/1/97

López-Sanjuan, C., Yuan, H., Vázquez Ramió, H., et al. 2021, A&A, 654, A61. doi:10.1051/0004-6361/202140444

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425. doi:10.1086/152650

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104. doi: 10.1088/0004-637X/785/2/104

Voit, G. M. 2005, Reviews of Modern Physics, 77, 207. doi:10.1103/RevModPhys.77.207

Werner, S. V., Cypriano, E. S., Gonzalez, A. H., et al. 2022, MNRAS. doi:10.1093/mnras/stac3273

Zheng, Y.-L. & Shen, S.-Y. 2021, ApJ, 911, 105. doi:10.3847/1538-4357/abeaa2