

Universidade de São Paulo
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Departamento de Astronomia

Lia Doubrawa

**Detecção e caracterização de aglomerados de
galáxias nos levantamentos fotométricos
S-PLUS e J-PAS**

São Paulo

2024

Lia Doubrawa

**Detecção e caracterização de aglomerados de
galáxias nos levantamentos fotométricos
S-PLUS e J-PAS**

Tese apresentada ao Departamento de Astronomia do Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo como requisito parcial para a obtenção do título de Doutor em Ciências.*

Área de Concentração: Astronomia

Orientador: Prof. Dr. Eduardo Serra Cypriano

São Paulo

2024

* Versão Corrigida. O original encontra-se disponível na Unidade.

À Bell, que chegou em nossas vidas quando mais precisávamos.

Agradecimentos

Gostaria de agradecer aos meus pais, Mirian e Francisco, e meu irmão, Jan, pela paciência, suporte e incentivo durante e pós-pandemia.

In memoriam, aos meus avós Lili e Norberto, Margarida e Francisco, e tia Wali pelo cuidado, carinho, canastras e cafés.

Ao meu orientador, Prof. Dr. Eduardo Serra Cypriano, pela paciência, apoio e dedicação ao trabalho desenvolvido.

Ao IAG, programa de Pós-Graduação, e a todos os colaboradores pela oportunidade e suporte para a realização do curso. Em especial, a equipe de suporte técnico, Ulisses, Luís e Marco.

À bolsa de estudos concedida pela coordenação Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (PROEX/CAPES), Código de Financiamento 001, sob número 88887.387950/2019-00.

Aos amigos que fiz ao longo dessa caminhada, antes e pós-pandemia. Certamente a atenção, risadas e carinho tornaram os momentos mais leves e tranquilos.

Aos meus amigos de sala, Isabel, Leonardo e Raquel, por me aguentarem conversar com o computador. E claro, às descontrações, chocolates e cafés.

Em especial, às queridas Joyce, Julise, Rafaela e Sandrine pelo suporte e carinho.

À Andresa, pelos ensinamentos e reflexões.

Aos colaboradores, Prof. Dr. Alexis Finoguenov, Prof. Dr. Paulo A. A. Lopes, Prof. Dr. Renato Dupke, Dr. Rogério Monteiro de Oliveira, pelas contribuições e ensinamentos.

Às colaborações S-PLUS e J-PAS.

Ao IAG/USP, pelo acesso ao Cluster do Laboratório de Astroinformática.

Esta tese/dissertação foi escrita em L^AT_EX com a classe IAGTESE, para teses e dissertações do IAG.

“ Toda a nossa ciência, comparada com a realidade, é primitiva e infantil - e, no entanto, é a coisa mais preciosa que temos.”

Albert Einstein

Resumo

Os aglomerados de galáxias desempenham um papel fundamental na compreensão das propriedades do nosso Universo, bem como na investigação de sua história dinâmica e evolução. Nesta tese de doutorado, investigamos os aglomerados de galáxias utilizando dados dos levantamentos fotométricos S-PLUS e miniJPAS, com o objetivo de aplicar técnicas desenvolvidas no levantamento J-PAS. Refinamos a detecção de aglomerados utilizando o algoritmo PZWav e desenvolvemos um estimador adaptativo, *Adaptive Membership Estimator* (AME), para atribuir probabilidades de pertinência a galáxias próximas aos centros dos aglomerados. Os resultados incluem análises com o AME aplicado a um catálogo simulado, proporcionando estimativas robustas de riqueza e destacando a aplicação competitiva do método em levantamentos fotométricos. A aplicação do PZWav ao miniJPAS, que revelou possíveis estruturas de massas tão baixas quanto $4 \times 10^{13} M_{\odot}$ a um *redshift* de 0.75, demonstrando a profundidade do levantamento e a sensibilidade de nossos algoritmos. No contexto do S-PLUS/iDR4, cobrindo 3 000 graus quadrados, a detecção revelou cerca de 190 000 candidatos a aglomerados de galáxias com $\text{SNR} > 4$. Comparações entre diferentes estudos revelou a completude do método e a qualidade dos dados.

Abstract

Galaxy clusters play a fundamental role in understanding the properties of our Universe, as well as investigating its dynamic history and evolution. In this thesis, we explore galaxy clusters using data from the photometric surveys S-PLUS and miniJPAS, with the aim of applying techniques developed in the J-PAS survey. We refine the detection of clusters using the PZWav algorithm and develop an Adaptive Membership Estimator (AME) to assign membership probabilities to galaxies near the cluster centers. The results include analyses with AME applied to a simulated catalog, providing robust estimates of richness and highlighting the competitive application of the method in photometric surveys. The application of PZWav to miniJPAS, revealed possible structures with masses as low as $4 \times 10^{13} M_{\odot}$ at a redshift of 0.75, emphasizing the depth of the survey and the sensitivity of our algorithms. In the context of S-PLUS/iDR4, covering 3 000 square degrees, the detection revealed about 190 000 candidates for galaxy clusters with $\text{SNR} > 4$. The comparison with the literature revealed the completeness of the method and the quality of the data.

Lista de Figuras

1.1	Aglomerado mJPC 2470-1771, observado no <i>redshift</i> 0.289. Essa é a estrutura mais massiva encontrada no levantamento miniJPAS (Seção 2.2). Em branco, o <i>redshift</i> fotométrico das galáxias próximas ao <i>redshift</i> da estrutura. Em amarelo os <i>redshifts</i> espectroscópicos das duas galáxias mais brilhantes. Em vermelho foram desenhadas as curvas críticas obtidas por lentes fortes. As curvas ciano mostram as linhas de isocontorno de emissão em raios-X no intervalo 0.5 – 2 keV obtidos com o telescópio espacial XMM-Newton. Figura retirada de Bonoli et al. (2021).	32
2.1	Sistema de filtros dos levantamentos fotométricos. Painel superior: S-PLUS (retirado de Mendes de Oliveira et al., 2019); Painel inferior: J-PAS (adaptado de Bonoli et al., 2021).	47
3.1	Exemplos de PDF de foto-z normalizadas pelo seu valor máximo obtidas pelos levantamentos S-PLUS (painel superior) e miniJPAS (inferior). Legenda mostra os números de identificação (ID) das galáxias em cada levantamento.	51
3.2	Mapa de densidades em 2D produzidos pelo PZWav. Cada pico que se destaca em relação ao ruído de campo é considerado um candidato a aglomerado.	54
3.3	Relação riqueza – massa obtida com o estimador de abertura fixa. Os losangos azuis mostram os valores médios dos aglomerados fictícios e os círculos avermelhados mostram os resultados FAE. As linhas de mesmos tons mostram a regressão linear para ambos os conjuntos de dados.	58

3.4	Exemplos de distribuição radial de galáxias e R_c . A área é calculada em unidades de kpc. A linha vertical vermelha destaca a descontinuidade detectada na densidade central. Adaptado de Doubrava et al. (2023).	60
3.5	Relação entre o número de membros verdadeiros do aglomerado (dados pela simulação), λ_{True} , e os valores médios obtidos com o método FAE aplicado sobre R_c , λ_{FAE} . A linha pontilhada representa a regressão linear. Essa relação pode ser usada para inferir o número mínimo de vizinhos para executar o HDBSCAN.	62
4.1	Propriedades do catálogo de aglomerados simulado. Painéis superiores: Da esquerda para a direita: distribuições de <i>redshift</i> e de massa do catálogo simulado de aglomerados de galáxias. Painéis inferiores: riqueza, λ_{AME} ; luminosidade óptica, L_λ ; e massa estelar total, M_λ^* . Tais quantidades são estimadas utilizando as probabilidades (P_{mem}) de cada galáxia pertencer ao dado aglomerado.	64
4.2	Distribuição das probabilidades calculadas por AME. Em azul, a distribuição quando consideramos apenas os membros verdadeiros.	66
4.3	Distribuição de riquezas. Painel superior: Valores de riqueza para coordenadas do céu distribuídas aleatoriamente em função do <i>redshift</i> . Os valores foram obtidos para os estimadores de abertura fixa (FAE) e adaptativa (AME), representados por diamantes em tons mais claros e círculos mais escuros, respectivamente. Painel inferior: Histograma de riqueza obtido com AME em escala logarítmica. Apenas 240 em 1000 coordenadas aleatórias têm um valor significativo (maior que 1).	68
4.4	Significância da riqueza em comparação entre a riqueza calculada com o estimador adaptativo (λ_{AME}) e a riqueza “verdadeira” fornecida pela simulação (λ_{True}). Cada intervalo de riqueza verdadeira contém pelo menos 10 aglomerados. Em pontos avermelhados as riquezas obtidas com <i>redshift</i> fotométrico (z_{phot}), e em diamantes azuis os resultados com o <i>redshift</i> fornecido pela simulação (z_{True}). Os pontos em cinza mostram o espalhamento da distribuição. A linha pontilhada preta mostra a relação um-para-um. . .	69

4.5	Relações de escala entre massa e traçadores ópticos. Destacamos os valores medianos em intervalos do traçador em questão. Os resultados do AME são apresentados como pontos avermelhados, os valores dados pela simulação como losangos azuis e suas respectivas regressões lineares, em linhas salmão e azul. Os valores dos coeficientes podem ser encontrados na Tabela 4.1. Do painel superior ao inferior: Relação Massa - Riqueza; Massa -Luminosidade óptica; Massa - Massa estelar.	71
4.6	Significância da pertinência da galáxia como uma comparação entre a fração de membros verdadeiros f_{true} e as probabilidades estimadas P_{mem} , para duas abordagens diferentes: o padrão 3σ e um corte mais rigoroso σ_{cut} com base em informações conhecidas da simulação.	77
4.7	Valores de completude e pureza aplicando diferentes limites de P_{mem} . Destacamos $P_{mem} > 0.50, 0.75, 0.90$. Para comparação, mostramos em tons avermelhados os resultados obtidos com o modelo padrão (3σ) e com a corte rigoroso (σ_{cut}) em azul.	79
4.8	Semelhante à Figura 4.7 para amostras distintas de aglomerados de galáxias. Painel superior: diferentes faixas de <i>redshift</i> com passos de $dz = 0.1$. Painel inferior: intervalos de massa com amplitude de $d\log_{10}(M) = 0.75$. Para ambos, destacamos os cortes $P_{mem} > 0.2, 0.5, 0.7$ e 0.8 como círculo, quadrado, losango e triângulo, respectivamente. Figura retirada de Doubrawa et al. (2023).	80
5.1	Distribuições de propriedades do catálogo de aglomerados gerado pelo PZWav. Painel superior: <i>Redshift</i> dos candidatos detectados. Painel inferior: Distribuição de SNR. Ambos os histogramas mostram a amostra antes (tons claros) e depois (vermelhas) da seleção em riqueza. Aplicamos λ_{thr} de 1.9 e 3.0 para os regimes de aglomerado e grupo (conforme discutido na Seção 5.2).	85

5.2	Distribuição das riquezas em diferentes cortes de magnitude absoluta. Painel superior: Comparação entre riquezas no regime de grupos ($M_r < -19.5$) para estruturas com $z < 0.3$, e no regime de aglomerados ($M_r < -21.25$). Painel inferior: Relação entre o SNR do PZWav e a riqueza. Alguns grupos que passaram despercebidos no regime de aglomerados podem apresentar uma riqueza significativa em cortes mais rasos de magnitude.	88
5.3	Distribuição de riqueza do catálogo detectado pelo algoritmo PZWav para os regimes de grupo (rosa) e aglomerado (vermelho escuro) - antes (tons mais claros) e depois (tons mais escuros) do limiar de riqueza. A linha tracejada indica a transição de <i>redshift</i> em $z = 0.3$	90
5.4	Distribuições de massa e <i>redshift</i> das detecções ópticas do PZWav que possuem uma contraparte em raios-X, representadas por diamantes roxos. Círculos pretos abertos exibem o catálogo de detecção de Erfanianfar para a área de sobreposição. No total, somam 17 grupos e aglomerados de galáxias com distância central correspondente menor que 500 kpc e diferença de <i>redshift</i> de $\Delta z < 0.05$. As massas são inferidas a partir da calibração de lentes fracas dadas por Leauthaud et al. (2010).	93
5.5	O segundo aglomerado mais massivo do catálogo estendido de raios-X, em $z = 0.41$. O círculo em cinza, de raio de 1.465 minutos de arco (R_c), está centralizado no pico de densidade encontrado pelo algoritmo PZWav.	94
5.6	Relações de escala obtidas para o levantamento miniJPAS. Painel superior: M_{200c} e riqueza, λ . Painel do meio: M_{200c} vs luminosidade óptica, L_λ . Painel inferior: M_{200c} vs massa estelar, M_λ^* . As linhas representam os valores de melhor ajuste. Os parâmetros de melhor ajuste podem ser encontrados na Tabela 5.4.	97
5.7	Diferença entre as coordenadas centrais dos aglomerados correspondentes identificados pelos três catálogos óticos, em uma comparação 2 por 2. As cores claras representam aglomerados com <i>redshifts</i> menores que $z \leq 0.55$, os escuros para $z > 0.55$. De cima para baixo: Vermelho: PZWav - VT; Azul: AMICO - PZWav; e Verde: VT - AMICO.	100
5.8	Igual à Figura 5.7 para a diferença entre <i>redshifts</i> das estruturas detectadas.	101

5.9	Distribuição da riqueza em função do SNR para os catálogos ópticos: PZWav em pontos cinza, correspondência entre PZWav - VT em verde, e PZWav - AMICO em azul. As correspondências entre os três catálogos são apresentadas por círculos abertos.	102
5.10	Os três algoritmos de detecção de aglomerados PZWav (vermelho), AMICO (azul) e VT (verde) – em comparação com o catálogo de galáxias massivas. As curvas resultantes fornecem a fração de galáxias de uma dada massa estelar sem qualquer associação à aglomerados/grupos.	105
5.11	Evolução da fração de galáxias sem associação à aglomerados/grupos em função do <i>redshift</i> para duas faixas de massa: $2 < M^* [10^{11} M_{\odot}] < 3$ e $3 < M^* [10^{11} M_{\odot}] < 4.2$. Em preto, as correspondências comuns entre catálogos ópticos, cinza exibe a amostra completa do PZWav, vermelho representa a sub-amostra após o corte em riqueza de PZWav, verde destaca VT, azul AMICO e roxo é a contribuição de todos os catálogos ópticos juntos (considerando a sub-amostra de PZWav).	106
5.12	A massa mínima do halo em função do <i>redshift</i> para garantir uma fração de galáxias isoladas inferior a 0.2. Em triângulos vermelhos são apresentados PZWav $\lambda > \lambda_{thr}$, VT como losangos verdes e quadrados azuis como valores obtidos para o catálogo produzido pelo AMICO. As massas dos halos são estimadas seguindo a relação M^*-M_h descrita em Gozaliasl et al. (2018). Os valores médios do <i>redshift</i> são deslocados artificialmente para uma melhor visualização.	108
6.1	Projeção do céu meridional demonstrando os resultados obtidos para cada uma das grandes áreas observadas pelo levantamento S-PLUS em diferentes cores. STRIPE 82 em azul escuro; SPLUS-S em índigo; SPLUS-N em carmesim; MC em magenta; e HYDRA em vermelho.	110
6.2	Histogramas das riquezas medidas pelo AME. As cores representam as área descritas na Figura 6.1.	112
6.3	Comparação entre catálogos de aglomerados gerados com o primeiro ($0.1 < z < 0.4$) e o quarto lançamento de dados do S-PLUS ($0.05 < z < 0.5$), correspondente a parte da área STRIPE 82.	114

6.4	Variação da fração de correspondências (<i>Match</i>) em função do <i>redshift</i> para os catálogos com as maiores diferenças dos valores medidos na Tabela 6.2. Em magenta escuro, Mehrrens et al. (2012); carmesim Takey et al. (2016, 2019); e azul escuro Rykoff et al. (2016).	116
6.5	Número de galáxias por área, em função da magnitude aparente, para a faixa de $0.1 < z < 0.4$. As linhas vermelhas representam S-PLUS/DR4, S-PLUS/DR1 em azul claro, e a simulação em azul escuro.	116
6.6	Valores médios e desvios padrão para as variações de centro (em kpc) e redshifts entre as estruturas encontradas pelo PZWav e os catálogos apresentados na Tabela 6.2 em duas faixas de redshift, $0.05 < z < 0.5$ (círculos roxos, e diamantes laranjas) e $0.1 < z < 0.4$ (estrelas azuis e pontos vermelhos). As linhas pontilhadas mostram a média das médias para $0.05 < z < 0.5$ para guiar os olhos.	117
8.1	Tensão nas restrições de σ_8 entre diversos estudos. Figura retirada de Abdalla et al. (2022).	126
A.1	Exemplo da aplicação do filtro de diferença de gaussianas (DoG). Painel superior esquerdo: Composição de 3 sobredensidades circulares adicionados a um ruído gaussiano. Painel superior direito: Resultado da aplicação do filtro DoG e a detecção dos picos utilizando a função <code>peak_local_max</code> do “scikit-image”. Painéis inferiores: da esquerda para direita, filtros gaussianos com $\sigma = 7$ e $\sigma = 15$	151
B.1	Exemplos de diferentes pontos centrais, e a distância do 6 ponto mais distante como raio do círculo ($k = 5$). Figura retirada do site <i>readthedocs</i> . . .	155
B.2	Exemplos de árvore de distâncias para $k = 5$. Figura retirada da documentação do código em <i>readthedocs</i>	156
B.3	Dendrograma do inverso da distâncias (λ) após a concentração da hierarquia em função do tamanho mínimo. Os braços circulados representam diferentes aglomerações selecionadas. Figura retirada da documentação do código em <i>readthedocs</i>	158

B.4 Identificação visual das aglomerações demarcadas na Figura B.3, com o mesmo código de cores. Figura retirada da documentação do código em <i>readthedocs</i>	158
--	-----

Lista de Tabelas

1.1	Traduzido e adaptado de Euclid Collaboration et al. (2019). Resumo de propriedades e nomes de algoritmos de detecção utilizados nos testes do levantamento Euclid. FL , função de luminosidade do aglomerado. P_{mem} representa se o código retorna ou não probabilidades de pertinência das galáxias aos candidatos detectados.	34
4.1	Valores de melhor ajuste da regressão linear para riqueza (λ), luminosidade óptica (L_λ) e massa estelar total (M_λ^*). O modelo massa-observável é descrito pela Equação 4.4. O_{piv} está sempre relacionado ao traçador verdadeiro (simulação). L_λ e M_λ^* são dados em unidades de L_\odot e M_\odot . Os resultados obtidos com a simulação são identificados por † após o traçador observável.	72
4.2	O erro relativo mediano entre os resultados de riqueza antes e depois dos deslocamentos ΔR e Δz , considerando distribuições gaussianas de 1σ , 2σ e 3σ	73
4.3	Comparação entre os resultados de riqueza de AME com PDFs de foto-z mais realistas como distribuição t de Student e diferentes frações de bimodalidade.	75
4.4	Valores de deslocamento de centro e <i>redshift</i> entre aglomerados da simulação/PZWav e os obtidos por galáxias aceitas como membro ($P_{mem} > 0$) em torno de seu centro.	81
5.1	Colunas e descrição do catálogo de detecção	84
5.2	Colunas e descrição do catálogo de membros.	86

5.3	Catálogo de raios-X estendido baseado nos centros ópticos do algoritmo PZWav. Estamos apresentando as primeiras e as últimas 4 linhas ordenadas por M_{200c} . L_x e M_{200c} são dados em unidades de $\text{erg cm}^{-2} \text{s}^{-1}$ e M_\odot respectivamente.	92
5.4	Valores de ajuste da regressão linear descrita pela Equação 4.4. L_λ e M_λ^* são dados em unidades de L_\odot e M_\odot para os dados do levantamento miniJPAS.	96
5.5	Valores medianos dos traçadores ópticos para cada catálogo de aglomerados e grupos de galáxias, divididos entre aglomerados correspondentes (<i>Match</i>) e não correspondentes (Nã-M), para riqueza, luminosidade óptica e massa estelar. M_λ é dado em unidades de $10^{11} M_\odot$ e L_λ em $10^{11} L_\odot$	103
6.1	Número de objetos utilizados para a análise em cada uma das grandes áreas observadas pelo levantamento S-PLUS, a quantidade de sobredensidades identificadas pelo algoritmo PZWav, tamanho da área, e as densidades de galáxias e detecções. Excluimos intencionalmente SPLUS-D, devido ao seu foco no disco da Via Láctea.	110
6.2	Comparação entre diferentes catálogos obtidos em diferentes comprimentos de onda, e a fração de correspondências de detecção com o catálogo produzido pelo algoritmo PZWav em duas faixas de <i>redshift</i> diferentes. O estudo de Hilton et al. (2021) utiliza diferentes fontes, e cortes em magnitude variam entre $r < 21$ e $r < 26$	113

Sumário

1. <i>Introdução</i>	27
1.1 Formação de estruturas	27
1.2 Aglomerados de galáxias	30
1.2.1 Algoritmos de detecção no óptico	33
1.2.2 Observações em raios-X	36
1.2.3 Lentes gravitacionais fracas	38
1.3 Relações de escala	40
1.4 Estrutura da tese	41
2. <i>Base de dados</i>	43
2.1 Catálogo simulado de galáxias	43
2.2 J-PAS	45
2.3 S-PLUS	46
3. <i>Metodologia</i>	49
3.1 <i>Redshifts</i> fotométricos	49
3.2 PZWav	53
3.3 Análise de pertinência das galáxias	55
3.3.1 O estado da arte em associação probabilística	55
3.4 Estimador de abertura fixa	56
3.5 Estimador adaptativo de pertinência de galáxias	58
4. <i>Resultados utilizando simulações</i>	63
4.1 Significado físico de R_c	65

4.2	Teste de sanidade	66
4.3	Significância da riqueza	67
4.4	Relação massa-observável	69
4.5	Perturbações no centro e <i>redshift</i> do aglomerado	72
4.6	PDFs não gaussianas	74
4.7	Medidas de pertinência individuais	75
4.8	Completude e pureza	76
4.8.1	Posições e <i>redshifts</i> de aglomerados a partir de P_{mem}	79
5.	<i>Resultados com dados do miniJPAS</i>	83
5.1	Catálogo de detecções	84
5.1.1	Catálogo de pertinência das galáxias membro	86
5.2	Regimes de grupo e aglomerado	86
5.3	Refinando o catálogo de detecções com cortes em riqueza	87
5.4	Comparações com o catálogo de raios-X do levantamento AEGIS	90
5.5	Catálogo de raios-X estendido	91
5.6	Relações de escala	95
5.7	Comparações entre catálogos ópticos	96
5.7.1	Correspondência entre os catálogos	98
5.7.2	Estatísticas dos traçadores de massa	99
5.8	Completude dos catálogos de aglomerados	103
6.	<i>Resultados com dados do S-PLUS</i>	109
6.1	Detecção de estruturas	109
6.2	Comparação com a literatura	111
6.2.1	Centros e <i>redshifts</i> a partir dos membros	117
7.	<i>Conclusões</i>	119
8.	<i>Perspectivas Futuras</i>	125
8.1	Cosmologia a partir de contagens de aglomerados	125
8.2	Aglomerados multimodais	127
	<i>Referências</i>	129

<i>Apêndice</i>	147
<i>A. Diferença de Gaussianas</i>	149
<i>B. HDBSCAN</i>	153
B.1 Transformação do espaço com base na densidade de pontos	153
B.2 Construção da árvore de distâncias mínimas	154
B.3 Ligação hierárquica dos componentes	155
B.4 Concentração da hierarquia com base no tamanho mínimo <code>min_samples</code>	156
B.5 Extração das aglomerações	157
<i>C. Linmix</i>	159
C.1 Funcionamento do código	160
<i>D. Acesso aos códigos desenvolvidos e catálogos</i>	163
<i>E. Artigos publicados</i>	165

Introdução

Os aglomerados de galáxias são poderosos traçadores da formação da estrutura cósmica, pois rastreiam picos de sobredensidade na distribuição de matéria. A abundância dessas estruturas, quando analisada em função da massa e do *redshift*, revela-se altamente sensível à densidade de matéria no Universo e sua evolução. Assim, os aglomerados são reconhecidos como notáveis ferramentas para a determinação dos parâmetros cosmológicos (e.g. Carlberg et al., 1996; Reiprich e Böhringer, 2002; Voit, 2005; Allen et al., 2011; Weinberg et al., 2013; Pacaud et al., 2016; Costanzi et al., 2019; Ider Chitham et al., 2020; Giocoli et al., 2021; Lesci et al., 2022b,a). Desta forma, ao combinar identificação e caracterização de aglomerados de galáxias em diferentes comprimentos de onda, juntamente com o uso de simulações numéricas, aprimoramos nossa compreensão da estrutura, evolução e física fundamental do Universo.

Nesta introdução, destacamos o processo de formação dessas grandes estruturas, e examinamos os métodos empregados para sua detecção a partir de levantamentos ópticos. Discutimos métodos de determinar a massa de grupos e aglomerados de galáxias, e apresentamos a aplicação das relações de escala com traçadores ópticos como meio de obter estimativas robustas e precisas de massa para as estruturas detectadas.

1.1 Formação de estruturas

Até o início do século 20, acreditava-se que o Universo era estático. Entretanto, entre 1927 e 1929 George Lemaître e Edwin Hubble mostraram que galáxias se afastam de nós com velocidades (v) diretamente proporcionais a sua distância (D), parametrizada por $v = H_0 D$, onde H_0 é a constante de Hubble. Essa discussão, levou a conclusão de que o

Universo está em expansão, assim como a existência de um instante inicial, conhecido como “Big Bang”. A teoria do Big Bang é amparada por outras duas observações: a abundância de elementos químicos prevista pela nucleossíntese primordial (Wagoner et al., 1967); e a emissão da radiação cósmica de fundo (sigla em inglês, CMB), que retrata um Universo em estágios iniciais, mostrando diferenças de temperatura tão pequenas quanto 10^{-5} K (Planck Collaboration et al., 2016c).

Dada a expansão contínua do Universo, seria esperado que a densidade média de matéria diminuísse ao longo do tempo. Entretanto, pequenas flutuações maiores que a densidade média inicial teriam variações de amplitude suficientes para atrair mais matéria de seus arredores, intensificando suas concentrações. Nesse contexto, podemos descrever tais flutuações, $\rho(\mathbf{r}, t)$ (em uma posição \mathbf{r} e tempo t), como contrastes de densidade, definido por

$$\delta(\mathbf{r}, t) = \frac{\rho(\mathbf{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (1.1)$$

sendo $\bar{\rho}(t)$ a densidade média do Universo no instante t .

Esse fenômeno, conhecido como instabilidade gravitacional, é considerado responsável pela formação de regiões onde nuvens de matéria escura e bárions colapsam, destacando-se totalmente da expansão do Universo (Schneider, 2006; Lima-Neto, 2022). Assim, em pequenas escalas, as flutuações de densidade presentes no Universo, em grandes escalas, se fundiriam sob a influência da gravidade para dar origem a estruturas como galáxias, e aglomerados de galáxias tal qual observamos (Blumenthal et al., 1984; White e Frenk, 1991).

Quando as inhomogeneidades são pequenas, isto é, $\delta(\mathbf{r}, t) \ll 1$, o crescimento das estruturas pode ser estudado através da aproximação linear, o que permite a simplificação das equações de dinâmica dos fluidos (Equações de Euler e Poisson). Entretanto, quando $\delta(\mathbf{r}, t) > 1$, como por exemplo para descrever as estruturas presentes no Universo atual, precisamos recorrer ao modelo do colapso esférico (Gunn e Gott, 1972). Devido ao seu difícil tratamento analítico, uma descrição precisa da evolução de $\delta(\mathbf{r}, t)$ é obtido por meio de simulações numéricas.

Supondo que a distribuição das flutuações de densidade em um determinado momento é um fenômeno aleatório, sem uma escala privilegiada, e com amplitude $|\delta_k|$, a distribuição assume a forma de,

$$|\delta_k|^2 \propto k^n \quad (1.2)$$

sendo k um parâmetro inversamente proporcional ao tamanho de uma estrutura.

Assim, definimos $P(k) \propto \langle |\delta_{\vec{k}}|^2 \rangle$ como o espectro de potência das flutuações, obtido pela média sobre todas as direções possíveis do vetor de onda \vec{k} . A evolução desse espectro depende da componente dominante no Universo em um dado momento (matéria escura, neutrinos, fótons e bárions), e pode ser descrita por uma função de transferência linear $T(k, z)$, que multiplica $P(k) \rightarrow P(k) = P(k_0) T(k, z)^2$.

A evolução temporal da função de transferência se descreve em função do *redshift* (Equação 178, Lima-Neto, 2022),

$$T(k, z) = \frac{\delta_k(0) D_+(z)}{\delta_k(z) D_+(0)} \quad (1.3)$$

de forma que $\delta_k(0)$ representa o espectro de potência logo após o processo de inflação¹, e D_+ é o fator de crescimento dependente da cosmologia utilizada para a descrição do Universo. Entende-se D_+ como uma das soluções da equação que descreve a evolução das perturbações sobre o efeito da pressão e gravitação utilizando a aproximação linear (Equação 150 em Lima-Neto, 2022).

Podemos também definir a flutuação de massa em um volume de raio R como a variância σ_M , que descreve a amplitude ou normalização do espectro de potência,

$$\sigma_M = \frac{\delta M}{M} \propto \left(\int |\delta_k|^2 k^2 dk \right)^{1/2} \quad (1.4)$$

Este valor pode ser derivado de observações e teoria. Medindo as flutuações de densidade em $z = 0$ em uma esfera de $8 h^{-1}$ Mpc, definimos um parâmetro bem conhecido na literatura como σ_8 . Estimativas usando a densidade de galáxias em aglomerados resultam em uma normalização $\sigma_8 = 0.83 \pm 0.15$ (Fan et al., 1997). Estudos realizados por Hinshaw et al. (2013) através de oscilações acústicas de bárions², reduziram significativamente a dispersão e encontraram um valor de $\sigma_8 = 0.821 \pm 0.023$. Nota-se que esse estudos são desatualizados. Com o aprimoramento da tecnologia e a melhoria na precisão de diferentes métodos, observou-se uma tensão significativa entre as medições. Ao considerar medidas

¹ A teoria da inflação conta com uma expansão exponencial do Universo inicial proposta por Guth (1981) e Linde (1982) para resolver fisicamente problemas encontrados no modelo do Big Bang clássico. Para uma revisão ver Capítulo 2.11 Lima-Neto (2022).

² Perturbações no campo de densidades que se formaram no período do desacoplamento ($z = 1100$) onde a temperatura do Universo inicial cai o suficiente para que radiação e matéria tenham comportamentos diferentes. Para uma revisão ver Capítulo 3.5 (Lima-Neto, 2022).

em relação ao universo primordial, CMB (do inglês, **cosmic microwave background**), essa tensão é ainda mais pronunciada. Pratt et al. (2019) discute algumas interpretações para esses resultados, e exemplos dessas discrepâncias podem ser encontrados na Seção 8.

Dado que o colapso gravitacional é influenciado pela distribuição de massa, as características da matéria escura desempenham um papel essencial nesse processo.

Apesar da ausência de informações conclusivas sobre a natureza da matéria escura, podemos tentar descrevê-la de duas formas distintas: como um sistema de partículas que se movem lentamente (matéria escura fria, sigla em inglês CDM), ou um sistema com velocidades relativísticas (matéria escura quente, sigla HDM). Nesse contexto, “frio/quente” referem-se as velocidades aleatórias das partículas inferiores/comparáveis à velocidade da luz (Liddle e Lyth, 1993).

O modelo predominante atualmente é conhecido como modelo hierárquico, dominado por matéria escura fria (White e Rees, 1978; Blumenthal et al., 1984; White e Frenk, 1991). Esse modelo prevê com grande precisão a forma do espectro de potência das estruturas em grande escala do Universo. Conforme sugerido por simulações, as primeiras estruturas a se formarem teriam massas bariônicas da ordem de $10^6 M_{\odot}$. Com sua evolução temporal, o colapso e eventual fusão com outras estruturas, estas se tornariam as galáxias e aglomerados que observamos atualmente, com massas totais que podem chegar até $10^{15} M_{\odot}$. Ao empregar um valor aproximado para o parâmetro de densidade $\Omega_M \approx 0.3$, observou-se que esse modelo reproduz de maneira precisa a formação e o número de estruturas.

Em modelos concorrentes, por exemplo, dominados pela presença de matéria escura quente (Ellis et al., 1984; Michael et al., 2006), a formação inicial de estruturas seria em escalas da ordem de aglomerados de galáxias, ou até superaglomerados. Posteriormente, galáxias e estruturas menores se formariam devido à fragmentação. Tal cenário resultaria em uma população reduzida de galáxias de campo, em contradição às observações.

1.2 Aglomerados de galáxias

Nesse contexto de formação hierárquica, os aglomerados de galáxias surgem como as maiores estruturas que atingiram um estágio avançado de maturidade dinâmica, caracterizado pelo equilíbrio virial³. Os aglomerados mais imponentes estão localizados nos pontos

³ Teorema do virial, que estabelece o equilíbrio entre energia cinética e gravitacional.

de interseção entre os filamentos que definem a estrutura em larga escala (por exemplo, Colless et al., 2001), agindo como “nós” dessa teia. Nessas regiões, o material tende a fluir em direção aos nós, na forma de grupos ou galáxias individuais, ou de estruturas de massas semelhantes. Considerando tal posição privilegiada no processo de formação de estruturas cósmicas, os aglomerados de galáxias representam um ponto de convergência entre a cosmologia e a astrofísica.

O estudo das distribuições dessas importantes estruturas se iniciou por meio da análise de padrões de grandes áreas do céu em placas fotográficas, onde encontraram-se evidências de regiões brilhantes que se destacavam em meio ao campo. No entanto, para um estudo mais detalhado se fazia necessário a utilização de métodos mais eficientes e de maior resolução, assim como telescópios ópticos com grandes aberturas e longos tempos de exposição. A Figura 1.1 apresenta o aglomerado mJPC 2470–1771 observado no espectro visível com um telescópio de 2.5m (levantamento fotométrico J-PAS (Benitez et al., 2014) Seção 2), com linhas de contorno de análises de outros métodos, como raios-X (ciano) e lentes gravitacionais fortes (vermelho).

Historicamente, para melhor classificação entre as diferentes estruturas observadas, distinguiu-se aglomerados e grupos de galáxias a partir do número de galáxias encontradas dentro de um determinado raio, denominadas como “membros do sistema”, ou “riqueza”. Podemos definir, por exemplo, um “grupo de galáxias” como estruturas com dezenas ($\lesssim 50$) de galáxias membros localizados em uma esfera de raio de $\lesssim 1$ Mpc, e de massas características que somam $M \sim 10^{13} M_{\odot}$. Enquanto “aglomerados de galáxias” seriam estruturas mais ricas, com dezenas a centenas de membros, e raios maiores que ~ 1 Mpc. Tais estruturas apresentam massas que variam entre $M \sim 10^{14} - 10^{15} M_{\odot}$ (Schneider, 2006; Jones e Lambourne, 2004).

Apesar dos grandes números, a contribuição das galáxias soma apenas $\sim 5\%$ da massa total do aglomerado. Observações na faixa do raio-X revelaram que o gás ionizado no meio intra-aglomerado (do inglês, *intra cluster medium*, ICM) representa aproximadamente 15% da massa total⁴. Os outros $\sim 80\%$ são associados ao halo de matéria escura, que permeia toda a estrutura. Apesar de não ser diretamente observado, seus efeitos podem ser determinados por diferentes técnicas, que incluem dispersões de velocidade do sistema

⁴ Esse gás emite radiação por meio do processo chamado *bremstrahlung* (mais detalhes na Seção 1.2.2), um fenômeno observado em plasmas de alta temperatura ($10^6 - 10^7$ K).

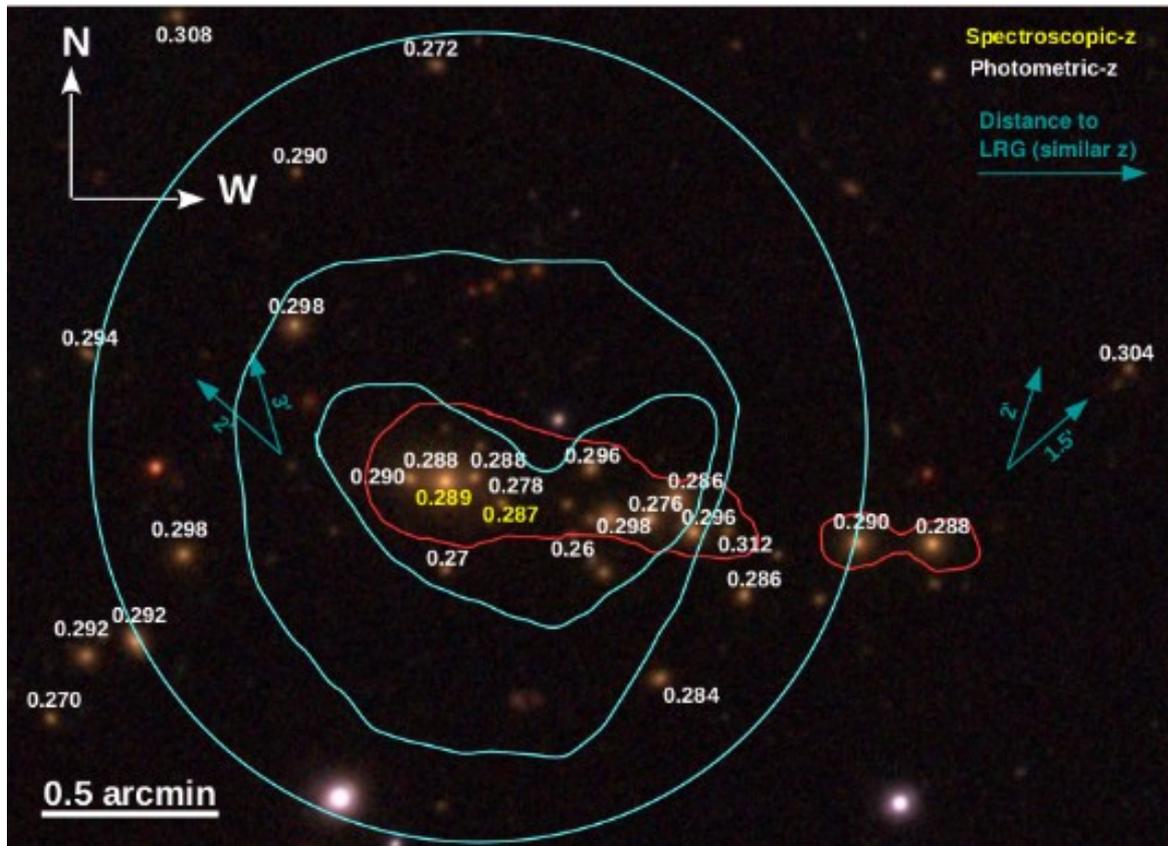


Figura 1.1: Aglomerado mJPC 2470-1771, observado no *redshift* 0.289. Essa é a estrutura mais massiva encontrada no levantamento miniJPAS (Seção 2.2). Em branco, o *redshift* fotométrico das galáxias próximas ao *redshift* da estrutura. Em amarelo os *redshifts* espectroscópicos das duas galáxias mais brilhantes. Em vermelho foram desenhadas as curvas críticas obtidas por lentes fortes. As curvas ciano mostram as linhas de isocontorno de emissão em raios-X no intervalo 0.5 – 2 keV obtidos com o telescópio espacial XMM-Newton. Figura retirada de Bonoli et al. (2021).

(Tempel e Tenjes, 2006; Elahi et al., 2018), emissão de raios-X (Suto et al., 1998; Chan, 2021) e lentes gravitacionais (Clowe et al., 2006; Shan et al., 2014).

Um dos catálogos de maior importância no contexto do início dos esquemas de classificação de aglomerados de galáxias é o Catálogo de Abell (Abell, 1958). Em seu método, Abell baseava-se na detecção de uma concentração de galáxias dentro de um ângulo sólido dado por $\theta_A = \frac{1'.7}{z}$. Nessa equação, o parâmetro z é uma estimativa do *redshift* da estrutura e situava-se na faixa de $0.02 < z < 0.2$. Esses limites são determinados pelo tamanho angular da placa fotográfica utilizada (limite superior) e pela capacidade de detecção dos instrumentos (limite inferior). Conforme esse critério, um aglomerado deveria conter no mínimo 50 galáxias, com magnitudes compreendidas na faixa de m_3 a $m_3 + 2$, sendo m_3 a representação da magnitude aparente da terceira galáxia mais brilhante dentro dessa área.

1.2.1 Algoritmos de detecção no óptico

Durante muitos anos, a principal tecnologia na astronomia óptica foi o uso de placas fotográficas, como o POSS (*Palomar Observatory Sky Survey*) e seus sucessores, para mapear todo o céu. No entanto, as placas fotográficas tinham suas limitações em termos de sensibilidade e linearidade. Isso, por sua vez, impulsionou o desenvolvimento de novas tecnologias, como os CCDs (dispositivos de carga acoplada) e telescópios de maior porte. O aumento na quantidade de dados resultante dessas melhorias exigiu o desenvolvimento de maior capacidade computacional e o aprimoramento de técnicas de análise para catalogar aglomerados de galáxias com maior precisão e eficiência.

Pesquisas cosmológicas de campo amplo, como KiDS (de Jong et al., 2013), DES (The Dark Energy Survey Collaboration, 2005), HSC (Miyazaki et al., 2012), WISE (Wright et al., 2010), LSST (Ivezić et al., 2019), EUCLID (Sartoris et al., 2016; Euclid Collaboration et al., 2022), entre outras, fornecem e/ou fornecerão conjuntos de dados contendo bilhões de galáxias. De particular interesse para a detecção de grupos e aglomerados de galáxias, foram desenvolvidos algoritmos sofisticados e técnicas de análise de dados que permitem identificar e catalogar estruturas em larga escala, bem como caracterizar propriedades físicas, como massa, morfologia, conteúdo de massa estelar e propriedades do gás. Tais algoritmos ficaram conhecidos como localizadores de aglomerados (em inglês “*cluster finders*”), e empregam diversas metodologias e definições computacionais para caracterizar aglomerados de galáxias.

Nome	Metodologia	Referência	Propriedades de aglomerados	Calibração de campo	P_{mem}
AMASCFI	Kernel adaptativo	Adami et al. (1998)	Tamanho típico e calibração m^*	Sim	Não
AMICO	Ajuste de filtros	Bellagamba et al. (2018)	FL e perfil de densidade	Não	Sim
HCFA	Localizador hierárquico	Díaz-Sánchez (in prep.)	Tamanho típico	Não	Sim
PZWav	Diferença de Gaussianas	Gonzalez (2014)	Tamanho típico e calibração m^*	Não	Não
sFoF	<i>Friends-of-friends</i>	Farrens et al. (2011)	–	Sim	Sim
WaZP	<i>Wavelet</i>	Benoist (2014)	Tamanho típico e calibração m^*	Sim	Sim
RedGOLD	Sequência vermelha	Licitra et al. (2016)	–	–	–
Voronoi	Diagrama de Voronoy	Iovino (in prep.)	–	–	–

Tabela 1.1 - Traduzido e adaptado de Euclid Collaboration et al. (2019). Resumo de propriedades e nomes de algoritmos de detecção utilizados nos testes do levantamento Euclid. *FL*, função de luminosidade do aglomerado. P_{mem} representa se o código retorna ou não probabilidades de pertinência das galáxias aos candidatos detectados.

Levantamentos de imagens que usam filtros de banda estreita são particularmente interessantes para estudos de aglomerados de galáxias. Esses levantamentos podem ser vistos como um meio-termo entre imagens de banda larga e levantamentos espectroscópicos em termos da precisão e acurácia dos *redshifts* fotométricos de galáxias (foto-z), como, por exemplo, os levantamentos J-PAS (Benitez et al., 2014; Bonoli et al., 2021) e S-PLUS (Mendes de Oliveira et al., 2019).

Essa transição está impactando diretamente os algoritmos de detecção, levando a adaptação de abordagens baseadas na cores de galáxias para outras dependentes, por exemplo, da detecção de sobredensidade de galáxias em diferentes *redshifts*.

Dessa ampla variedade de abordagens disponíveis na literatura, destacamos algumas na Tabela 1.1, que são consideradas no contexto do levantamento Euclid (Euclid Collaboration et al., 2019). Neste estudo, os pesquisadores exploram a detecção de aglomerados em levantamentos fotométricos, testando e selecionando os algoritmos que melhor se enquadram às características específicas do levantamento. Neste caso, destacaram-se os algoritmos AMICO e PZWav.

Neste trabalho, enfatizamos a utilização do algoritmo PZWav e realizamos uma comparação dos resultados obtidos com dois outros métodos de detecção, AMICO e VT, que foram gentilmente fornecidos por Matteo Maturi e Paulo Lopez, ambos da colaboração J-PAS, para analisar os dados do miniJPAS (Seção 5). Ressalta-se que os algoritmos passaram por modificações e adaptações em relação as versões utilizadas no levantamento Euclid. A seguir, fornecemos mais detalhes sobre cada um desses métodos. Na seção 3.2, apresentamos uma descrição mais aprofundada do algoritmo PZWav.

- AMICO (Maturi et al., 2019), busca por candidatos a aglomerado com um sistema de filtro dependente do *redshift*, que tem como objetivo amplificar o contraste entre os modelos de aglomerado e de ruído de campo. Ambos modelos utilizam a função de luminosidade assim como um perfil de densidade de galáxias para a definição do aglomerado. Utiliza-se a banda *H* para detecção como magnitude padrão, mas também aceita combinações de outras magnitudes. A distribuição de galáxias cria um mapa de amplitudes *3D*, no qual cada pico representa uma detecção. As coordenadas do céu e o *redshift* são estimados através do mapa *3D*.
- Voronoi (VT, Ramella et al., 2001; Lopes et al., 2004), adaptado para detecção de

aglomerados no espaço de *redshift* fotométrico. O algoritmo funciona em dois regimes diferentes: No primeiro, a amostra de galáxias é distribuída em fatias de *redshift*. Em cada fatia, o algoritmo seleciona as coordenadas do céu e calcula, para cada galáxia, a área coberta pelos vizinhos mais próximos. Define-se a densidade local como o inverso da área calculada. Dessa forma, apenas galáxias localizadas em regiões de densidade superior à média são mantidas como “sementes de aglomerados”. Em torno destas sementes, o código procura galáxias, crescendo até atingir um ponto de corte predefinido. Então, como segunda etapa, o algoritmo mescla resultados de diferentes *redshifts*, com base em coordenadas e *redshift*. Isso cria o catálogo de aglomerado final.

- PZWav (Gonzalez, 2014; Werner et al., 2022), é um localizador de aglomerados que identifica sobredensidades em escalas físicas fixas. Requer informações como coordenadas celestes para cada galáxias, *redshifts* e magnitudes. PZWav cria uma série de fatias de *redshift* inserindo cada galáxia ponderada pela probabilidade de estar no *redshift* determinado, criando mapas de densidade $2D$. Ao utilizar um kernel de diferença de Gaussianas, destacam-se as estruturas de tamanho similar aos aglomerados. Um segundo conjunto de mapas de densidade é gerado com coordenadas aleatórias, e é utilizado para estimar o ruído uniforme em função do *redshift*. Um candidato a aglomerado de galáxias é identificado em cada fatia do *redshift* como um pico no mapa $2D$ e é contabilizado se a estrutura estiver acima do limite de ruído.

1.2.2 Observações em raios-X

Observações de aglomerados de galáxias são conduzidas em várias partes do espectro eletromagnético, incluindo a faixa de raios-X. A emissão característica nessa faixa é produzida por um fenômeno conhecido como bremsstrahlung térmico, que surge devido à presença de plasmas de temperatura elevada e baixa densidade. Nesse processo, a interação de elétrons livres com íons resulta em emissão de um fóton e a redução da energia cinética do elétron (Jones e Lambourne, 2004). A energia do fóton costuma ser expressa em keV (quiloelétron-volt), tipicamente na faixa de 1 a 10 keV, o que corresponde a temperaturas na faixa de 10^7 a 10^8 K.

Ao contrário das linhas espectrais, que resultam das transições de elétrons entre níveis

de energia atômicos em gases de menor temperatura, os raios-X gerados pela desaceleração de elétrons livres exibem um espectro de emissão contínua (Sarazin, 1986). Essa emissividade, representada como ϵ , pode ser descrita por,

$$\epsilon = 2.4 \times 10^{-27} n_e^2 \sqrt{T} \text{ erg s}^{-1} \text{ cm}^{-3} \quad (1.5)$$

onde a temperatura do gás é representada por T , e a densidade eletrônica por n_e .

O gás presente no ICM é frequentemente caracterizado como um gás ideal totalmente ionizado, composto principalmente por elementos como hidrogênio ($\sim 75\%$) e hélio ($\sim 25\%$), somados a uma pequena contribuição de elementos resultantes da evolução estelar, chamados de “metais”. Esse gás que permeia a estrutura pode ser encontrado a alguns megaparsecs de distância do centro, fazendo com que sua emissão seja observada como fontes extensas.

As primeiras observações utilizando o satélite de raios-X *Uhuru* em 1971 (Gursky et al., 1971; Giacconi et al., 1971) confirmaram a notável luminosidade dos aglomerados de galáxias, expondo-os como objetos de destaque no céu. Telescópios posteriores como *Einstein* e *Rosat* permitiram a realização de observações com maior resolução, detectando também a presença de emissões de radiação provenientes de grupos de galáxias e aglomerados de menor massa (Jones e Forman, 1984; David et al., 1995; Voges et al., 1999). As informações mais recentes em raios-X são obtidas por meio de observações realizadas com os telescópios espaciais *XMM-Newton* (Gondoin et al., 1994), *Chandra* (Weisskopf, 1999), e atualmente, o *eRosita* (Predehl et al., 2021).

Essas medidas de emissão, somadas a suposições de equilíbrio hidrostático e simetria esférica, permitem estimar a massa de um aglomerado de galáxias. Entende-se que o gás presente no ICM tende a colapsar sob o poço de potencial gravitacional produzido pela massa presente, seja composta de matéria luminosa ou matéria escura. Esse colapso é evitado graças à pressão exercida pelo gás. Em um estado de equilíbrio hidrostático, a pressão varia conforme a distância em relação ao centro do aglomerado, portanto,

$$\frac{1}{\rho} \frac{dP}{dr} = -\frac{d\Phi}{dr} = -\frac{GM(r)}{r^2} \quad (1.6)$$

onde P é a pressão associada ao sistema, ρ é a densidade do gás e Φ o potencial gravitacional total. Dada a suposição que o gás presente é um gás ideal, podemos substituir $P = n_e k_B T = \rho k_B T / \mu m_p$, sendo n_e a densidade eletrônica. Dessa forma, a massa total pode

ser calculada por,

$$M(r) = -\frac{k_B T r}{G \mu m_p} \left(\frac{\ln \rho}{d \ln r} + \frac{d \ln T}{d \ln r} \right) \quad (1.7)$$

sendo $k_B T$ a constante de Boltzmann multiplicada pela temperatura do gás, μ o peso molecular médio⁵ e m_p a massa do próton.

Ao considerar apenas a matéria luminosa (galáxias e gás) as massas calculadas são significativamente menores do que aquelas obtidas a partir dos perfis de temperatura e dos gradientes de pressão observados (Smith, 1936; Zwicky, 1937). Esses resultados indicam que o campo gravitacional no aglomerado é mais forte do que o que seria previsto pelas componentes luminosas, sugerindo a presença de grandes quantidades de matéria escura envolvendo o aglomerado.

1.2.3 Lentes gravitacionais fracas

Uma abordagem independente para obter informações sobre aglomerados é observar os efeitos da gravidade sobre a luz. Quando um feixe passa perto o suficiente de um objeto massivo, sua trajetória sofre desvios (Bartelmann e Schneider, 2001). À medida que a massa do objeto defletor aumenta, a curvatura dos raios de luz se torna mais evidente. Dado que os aglomerados de galáxias possuem massas da ordem de $10^{14} M_\odot$, é esperado que atuem como lentes gravitacionais eficientes, desviando a trajetória da luz dos objetos situados atrás deles.

Como este efeito é uma consequência direta do campo gravitacional, ele é sensível a toda a massa presente no sistema, independentemente de se tratar de galáxias, gás ou matéria escura. Assim, a técnica de lentes gravitacionais apresenta uma vantagem sobre outros métodos, sem a necessidade de fazer suposições relacionadas à virialização ou equilíbrio hidrostático.

Além de deslocar a posição aparente do objeto em estudo, esse efeito pode resultar na distorção do objeto em forma de arcos, na formação de múltiplas imagens distorcidas, e no aumento de sua luminosidade. Por volta de 1980, diversos grupos de pesquisa identificaram galáxias em forma de arco em aglomerados em altos *redshifts*, cuja origem era desconhecida (Lynds e Petrosian, 1986; Soucaill et al., 1987; Lavery e Henry, 1988). Análises espectroscópicas do aglomerado A370 revelaram que os arcos apresentavam *redshifts* mai-

⁵ Em aglomerados de galáxias seu valor típico é de $\mu \sim 0.6$, representando um gás totalmente ionizado.

ores do que o próprio aglomerado, indicando que esses arcos correspondiam a fontes mais distantes situadas no campo de fundo.

Uma vez que as distorções dependem exclusivamente da distribuição de massa do aglomerado, o efeito também pode ser observado em menor escala. Nesses casos, dado que as distorções nas imagens de galáxias próximas umas das outras devem ser semelhantes, uma vez que o campo gravitacional que atua sobre elas é aproximadamente o mesmo, é possível calcular a média de várias imagens de galáxias e medi-las com precisão (Bartelmann e Schneider, 2001).

Esse alinhamento em primeiro plano de galáxias em torno de um aglomerado, fornece uma técnica poderosa para estimar a massa dos aglomerados de galáxias. No entanto, é crucial tratar cuidadosamente os erros sistemáticos, como a triaxialidade do halo (a orientação do aglomerado em relação ao seu eixo principal) e os efeitos de projeção de galáxias para minimizar a dispersão (Hoekstra et al., 2015; Melchior et al., 2017; Medezinski et al., 2018).

Assumindo um objeto isolado, a imagem observada das galáxias de fundo é distorcida tangencialmente ao plano do céu. Podemos definir tais distorções como,

$$\gamma_t = -(\gamma_1 \cos 2\phi + \gamma_2 \sin 2\phi) \quad (1.8)$$

onde ϕ é o ângulo azimutal da galáxia em relação ao centro da lente (no caso, o aglomerado) e (γ_1, γ_2) componentes complexas do cisalhamento γ .

Uma vez que o cisalhamento tangencial é fortemente dependente da distribuição de massa, seu valor médio a uma distância específica do centro da lente, representado por $\langle \gamma_t \rangle(r)$, pode ser interpretado como um contraste de massa (Miralda-Escude, 1991),

$$\langle \gamma_t \rangle(r) = \frac{\bar{\Sigma}(< r) - \bar{\Sigma}(r)}{\Sigma_{crit}} = \bar{\kappa}(< r) - \bar{\kappa}(r) \quad (1.9)$$

Nessa equação, a densidade superficial azimutal média em um raio de abertura r é representada por $\bar{\Sigma}(r)$ e $\bar{\Sigma}(< r)$ é a densidade superficial projetada em r . Assim, Σ pode ser expressa como $\Sigma = \Sigma_{crit} \langle \gamma_t \rangle$ se $\bar{\Sigma}(r)$ tender a zero, ou seja, se o raio no qual é realizado a medida é grande o suficiente para que o valor da densidade seja desprezível.

Outra abordagem para a medição de massas, denominada paramétrica, envolve o ajuste do perfil radial observado de $\langle \gamma_t \rangle$ às previsões desse valor feitas por modelos que descrevem o perfil de massa dos aglomerados, assim como em Cypriano et al. (2004); Vitorelli et al. (2018); McClintock et al. (2019).

1.3 Relações de escala

O estudo cosmológico utilizando aglomerados de galáxias enfrenta desafios pela necessidade de massas precisas para todas as estruturas identificadas, uma vez que estas não são diretamente observáveis. Uma estratégia consiste em buscar propriedades observacionais, como riqueza óptica ou sinal de raios-X, correlacionadas com a massa juntamente com a dispersão associada. Essa relação é inicialmente calibrada para um conjunto limitado de objetos e, posteriormente, aplicada à amostra completa.

Um traçador de massa ideal deve apresentar baixa dispersão intrínseca, pouca dependência com o estado dinâmico do aglomerado e ser facilmente acessível por meio de observações. Vários traçadores de massa têm sido explorados nas últimas décadas para minimizar incertezas e possíveis efeitos sistemáticos (Rozo et al., 2010; Vikhlinin et al., 2009; Becker e Kravtsov, 2011; Andreon, 2015; Palmese et al., 2020; Sampaio-Santos et al., 2020; Golden-Marx et al., 2022). No entanto, as estatísticas relacionadas a estruturas de baixa massa, como grupos de galáxias, ainda são limitadas e requerem investigações mais profundas.

Estudos focados principalmente na relação massa-riqueza enfrentam dois grandes desafios. O primeiro está relacionado ao baixo contraste da densidade de galáxias presentes nos grupos e aglomerados em relação às galáxias de campo, podendo gerar problemas de projeção e resultar em uma super (ao considerar galáxias de campo como membros) ou subestimação (perda de membros para o campo) da riqueza. O segundo desafio envolve a identificação precisa do centro dos aglomerados, sendo que um deslocamento incorreto pode impactar significativamente a estimativa de riqueza. Discutiremos tal ponto utilizando simulações na Seção 4.5.

Na literatura, diversas grandezas observáveis têm sido propostas como altamente correlacionadas com a massa de aglomerados. As incertezas estatísticas variam entre ~ 0.2 - 0.3 dex para massa de gás, temperatura, luminosidade em raios-X (Ettori, 2013; Sereno et al., 2019), sinal de efeito Sunyaev-Zel'dovich (Planck Collaboration et al., 2016b; Pratt e Bregman, 2020). Especialmente, quando se trata de dados ópticos, traçadores comuns incluem riqueza (Lopes et al., 2009; Bellagamba et al., 2019), luminosidade, massa estelar total e o ajuste do perfil de densidade das galáxias (Pearson et al., 2015; Huang et al., 2022).

Recentemente, Costanzi et al. (2019) e Ider Chitham et al. (2020) utilizaram com

sucesso a riqueza óptica para derivar restrições cosmológicas em pesquisas SDSS (Aihara et al., 2011) e CODEX (Finoguenov et al., 2020), respectivamente. Tais estudos se baseiam em amostras puras de estruturas de alta massa, como aglomerados de galáxias ($M > 10^{14} M_{\odot}$). Investigações para estruturas de menor massa, como grupos de galáxias, ainda são limitadas e requerem maior aprofundamento.

Levantamentos de imagens que usam filtros de banda estreita são particularmente interessantes para estudos de aglomerados de galáxias e sua caracterização. Portanto, também é necessário revisitar as técnicas de estimativa de pertinência e as estratégias que exploram os dados espectro-fotométricos como uma aplicação para traçadores de massa com baixa dispersão.

No entanto, um problema comum enfrentado nesse contexto, assim como em *redshifts* tradicionais produzidos por bandas largas, é a incerteza sobre a distribuição de galáxias no espaço de *redshift* (mais detalhes na Seção 3.1). Esse efeito pode levar à contaminação por galáxias próximas ao longo da linha de visada, assim como criar falsas detecções de aglomerados ao associar estruturas não relacionadas (Weinberg et al., 2013). O uso de observáveis menos discriminativos, como cores individuais de galáxias, pode agravar esse problema.

Pesquisas recentes, como as conduzidas por Castignani e Benoist (2016); Bellagamba et al. (2019); Lopes e Ribeiro (2020), abordam essa questão levando em consideração a associação de galáxias a aglomerados ou grupos. A utilização de probabilidades de pertencimento contínuas, em vez das tradicionais abordagens binárias (membro/não-membro), emerge como uma descrição mais precisa dos dados disponíveis.

1.4 Estrutura da tese

Levando em consideração a importância dos aglomerados de galáxias como traçadores da formação da estrutura cósmica, e a dificuldade em obter estimativas de massa para todas as estruturas detectadas, desenvolvemos um estimador adaptativo de pertinência de galáxias (do inglês, *adaptive membership estimator*, AME) cujos resultados são competitivos com a literatura, e que podem ser utilizados com traçadores que se relacionam com a massa.

Dessa forma, utilizamos os dados precisos dos levantamentos fotométricos miniJPAS e

S-PLUS, e detectamos a presença de sobredensidades com o algoritmo PZWav. A caracterização das estruturas é feita a partir da riqueza, massa estelar e luminosidade ópticas calculadas com AME.

Esse trabalho está dividido em três partes: A utilização de dados provenientes de simulações para teste e desenvolvimento do estimador adaptativo de pertinência das galáxias; A aplicação do código de detecção de estruturas e o AME em dados do levantamento miniJPAS; A aplicação do código de detecção de estruturas e o AME em dados do levantamento S-PLUS.

Apresentamos os dados utilizados na Seção 2. Os métodos são descritos em mais detalhes na Seção 3. Os resultados obtidos em cada parte são descritos em: Seção 4 para a simulação, que correspondem aos resultados e discussões publicados em Doubrawa et al. (2023); Seção 5 para o miniJPAS, referentes a Doubrawa et al. (2024); e Seção 6 para S-PLUS, como uma reanálise dos resultados encontrados em Werner et al. (2022) aplicados para o quarto lançamento de dados (DR4). Apresentamos as conclusões na Seção 7 e, na Seção 8 as possíveis aplicações e pesquisas em desenvolvimento.

Ao longo deste trabalho, adotamos uma cosmologia Λ CDM plana, com $h = 0.673$, $\Omega_m = 0.315$ e $\Omega_\Lambda = 0.685$, seguindo os parâmetros Planck Collaboration et al. (2014, 2016a), que representa a mesma cosmologia adotada nas simulações descritas em Seção 2.1. Magnitudes são fornecidas no sistema AB. Massas são dadas em relação a M_{200c} , caso contrário será especificado.

Base de dados

Neste capítulo descrevemos os dados utilizados ao longo do estudo, como por exemplo o catálogo simulado utilizado para preparação dos códigos de análise, assim como os dados reais provenientes dos levantamentos fotométricos S-PLUS e J-PAS.

2.1 *Catálogo simulado de galáxias*

Para avaliar o desempenho de nossos traçadores, utilizamos um catálogo simulado de galáxias e aglomerados produzido por Araya-Araya et al. (2021), feito para emular o primeiro lançamento dos dados do levantamento S-PLUS (Mendes de Oliveira et al., 2019, descrito em detalhes na Seção 2.3), correspondente à área Stripe 82.

Este catálogo fictício simula uma área projetada de 324 graus quadrados, usando galáxias sintéticas seguindo o modelo analítico Henriques et al. (2015) (SAM). O algoritmo usa a simulação do Millenium (Springel et al., 2005) como base para gerar um campo de densidade de matéria equivalente à estrutura cosmológica do Planck 1 (Planck Collaboration et al., 2014). A resolução de massa de cada partícula do halo é de $m_p = 9.6 \times 10^8 M_\odot/h$, mas apenas aquelas com massa estelar correspondente superior a $10^9 M_\odot/h$ são aceitas no catálogo.

Todos os halos dominantes de matéria escura com $M_{200c} \geq 10^{12.8} M_\odot$ são selecionados para o catálogo de aglomerados. Para garantir estimativas e estatísticas robustas de pertinência das galáxias, escolhemos para nosso estudo apenas as estruturas com pelo menos 3 galáxias associadas. As galáxias membros associadas aos halos de matéria escura (daqui em diante “membros verdadeiros”) são recuperados do histórico de fusão do sistema. Todas as galáxias que residiam formalmente em um halo e evoluíram para a formação de um

aglomerado recebem o ID de identificação de aglomerado “haloId”. Este ID nos permite identificar facilmente todas as galáxias que pertencem a um determinado aglomerado. As coordenadas dos aglomerados no céu são estimadas pelo valor mediano da distribuição das galáxias membros.

Redshifts verdadeiros observados nas simulações são criados adicionando *redshifts* cosmológicos (desvio para o vermelho nas linhas dos espectros das galáxias devido a expansão simulada) e *redshifts* Doppler criados por movimentos peculiares de cada galáxia fictícia.

O histórico de formação estelar (do inglês, SFH) e as massas estelares são extraídos dos resultados dos modelos de SAM. Uma vez que temos a informação da evolução estelar ao longo do tempo cósmico da simulação, é possível atribuir distribuições de energia espectral (do inglês, SED) a cada intervalo de tempo. Dependendo dos valores armazenados, pode representar uma determinada população estelar. Os SEDs, juntamente com os modelos de extinção de poeira, são usados para calcular as magnitudes aparentes u , g , r , i e z , de forma semelhante ao sistema de filtros Sloan AB (Fukugita et al., 1996).

Os *redshifts* fotométricos são simulados ao perturbar o seu valor verdadeiro adicionando um número aleatório de uma distribuição Normal. O desvio padrão utilizado, $\sigma_{MAD,z}(m_r)$, é o desvio absoluto mediano normalizado em função de diferentes magnitudes aparentes na banda r obtidos a partir da comparação entre foto- z s e *redshifts* espectroscópicos para uma amostra de galáxias encontradas pelo levantamento DR1/S-PLUS ($\sigma = MAD(r)$, Molino et al., 2020, com mediana de $\sigma = 0.026$).

Um procedimento semelhante gera a função de densidade de probabilidade do foto- z (PDF). Para cada galáxia, assumimos uma distribuição Normal com $\sigma_{MAD,z}(r)$ como desvio padrão, centrado no *redshift* fotométrico criado. Esta técnica considera a correlação entre as magnitudes da galáxia e seus erros fotométricos.

Para este estudo, limitamos o catálogo em $z < 0.45$, e assim, obtemos 238 grupos com massas entre $10^{12.8}$ – $10^{13.5}$ M_{\odot} , 358 aglomerados com $10^{13.5}$ – 10^{14} M_{\odot} , 249 com 10^{14} – $10^{14.5}$ M_{\odot} e 76 aglomerados massivos com $M > 10^{14.5}$ M_{\odot} .

Até onde sabemos, apenas Castignani e Benoist (2016), Bellagamba et al. (2019) e Werner et al. (2022) consideram amostras nesta mesma faixa de massa analisada ($\sim 10^{13}$ – 10^{15} M_{\odot}).

Com relação às limitações do catálogo, as estatísticas numéricas em baixas massas são restritas devido ao corte de riqueza aplicado (veja as distribuições resultantes na Fi-

gura 4.1).

2.2 J-PAS

Utilizamos os dados fotométricos do levantamento miniJPAS (Bonoli et al., 2021), que compreende um conjunto de 54 filtros de banda estreita, juntamente com três de banda larga (painel inferior da Figura 2.1). Esta pesquisa cobre uma área aproximada de 1 grau quadrado no campo *Extended Groth Strip*. O conjunto de dados miniJPAS serve como um desbravador para o levantamento cosmológico de campo amplo, o *Javalambre-Physics of the Accelerated Universe Astrophysical Survey* (J-PAS).

Conduzida a partir do Observatório Javalambre, na Espanha, a pesquisa emprega um telescópio dedicado de 2.5m e uma câmera de 1.2 Gpix capaz de capturar, em uma única tomada, uma área de 4.7 graus quadrados do céu (Benitez et al., 2014). A pesquisa completa visa observar 8500 graus quadrados usando o sistema de filtro descrito acima. Com sua configuração, espera-se que o J-PAS forneça uma imagem profunda do céu do hemisfério norte, com uma precisão de *redshift* fotométrico (z_{foto}) de $\Delta z/(1+z) \leq 0.003$ para galáxias com $r < 22.75$ AB mag.

Um estudo detalhado sobre as características de z_{foto} apresentadas por Hernán-Caballero et al. (2021), revelou um erro típico de $\sigma_{MAD,z} = 0.013$. Para atingir a precisão esperada de 0.003, seria necessário utilizar uma amostra mais refinada. No entanto, para o propósito deste estudo, priorizamos a completude do conjunto de dados. Uma descrição detalhada dos métodos utilizados para obter z_{foto} através do ajuste de modelos das SEDs podem ser encontrados na Seção 3.1.

Devido ao seu sistema de filtros estreitos, o conjunto de dados miniJPAS fornece estimativas confiáveis de SED. Em González Delgado et al. (2021, 2022), os autores detalham o procedimento para estimar a massa estelar das galáxias utilizando os espectros obtidos. A aplicação do código de ajuste BaySeAGal (González Delgado et al., 2021) permite determinar funções de densidade de probabilidade das propriedades da população estelar, como massa estelar, com base na luminosidade e perda de massa devida à evolução estelar.

Nesse contexto, aplicamos os seguintes critérios para a seleção dos dados para nossa análise. Para obter uma amostra predominantemente de galáxias, utilizamos o critério `total_prob_star < 0.5`. Esse parâmetro, desenvolvido por López-Sanjuan et al. (2019),

indica a probabilidade de um objeto ser uma estrela (1) ou uma galáxia (0), e é calculado com base em informações como perfil espacial e magnitude na banda r , aliadas aos dados de paralaxe do Gaia (Gaia Collaboration et al., 2018). Devido à sua distribuição altamente bimodal, o corte em 50% garante uma seleção precisa de galáxias.

Além disso, incluímos todas as galáxias com magnitude aparente na banda r `MAG_AUTO` ≤ 22.75 , garantindo a completude dos dados até $z_{foto} < 0.8$. Para evitar áreas com estrelas saturadas e outras condições que possam afetar a qualidade dos dados, aplicamos a condição `MASK_FLAGS=0`.

Este primeiro lançamento de dados do miniJPAS fornece um conjunto de dados ideal para refinar várias estratégias relacionadas à detecção e análise de aglomerados de galáxias, oferecendo previsões consistentes para o levantamento J-PAS.

2.3 S-PLUS

O levantamento fotométrico S-PLUS, do inglês *Southern Photometric Local Universe Survey*, é um projeto que tem como objetivo capturar imagens de mais de 9 000 graus quadrados da esfera celeste em 12 bandas ópticas por meio do telescópio T-80-Sul de 0.8m, localizado no Cerro Tololo, Chile. O telescópio está equipado com um detector de $9.2k \times 9.2k$ pixels, cobrindo 2 graus quadrados do céu. O levantamento abrange dois campos em altas latitudes galácticas $|b| > 30^\circ$, com 8 000 graus quadrados, e outras duas regiões no plano galáctico, totalizando 9 300 graus quadrados (Mendes de Oliveira et al., 2019).

Em seu 4^o lançamento de dados (DR4), o levantamento soma 1 629 campos observados, contemplando aproximadamente 3 000 graus quadrados do céu. Ele engloba as mesmas regiões dos lançamentos anteriores (DR1, 2 e 3), bem como novos dados provenientes das Nuvens de Magalhães, do disco da Galáxia e de alguns campos adicionais em altas latitudes galácticas.

O S-PLUS possui 12 filtros, incluindo 5 bandas largas e 7 bandas mais estreitas centradas em regiões de interesse especial (painel superior da Figura, 2.1). Esse sistema de filtros pode produzir estimativas de *redshifts* fotométricos de alta precisão, $\Delta z / (1 + z) \leq 0.02$, para galáxias com $r < 20$ mag AB.

Estudos das incertezas obtidas realizadas por Lima et al. (2022) mostram que para

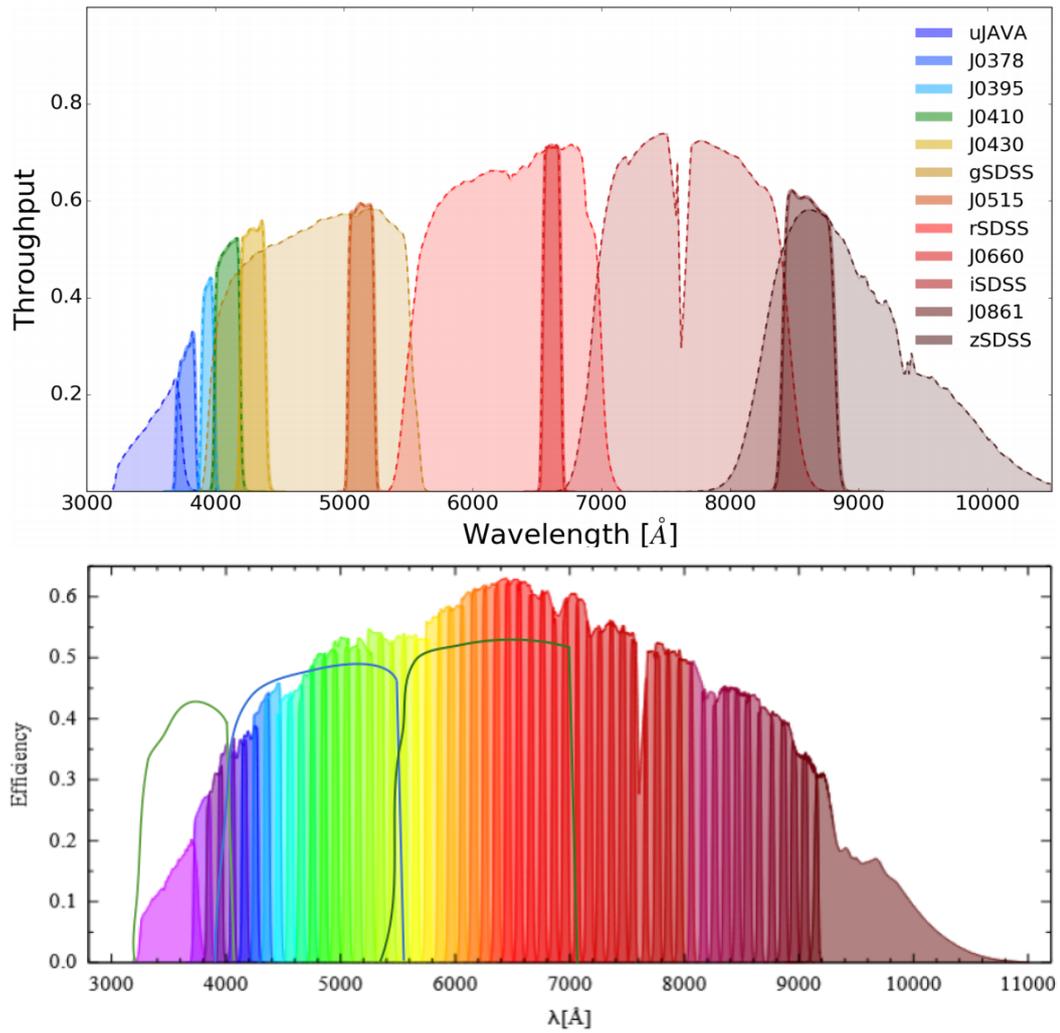


Figura 2.1: Sistema de filtros dos levantamentos fotométricos. Painel superior: S-PLUS (retirado de Mendes de Oliveira et al., 2019); Painel inferior: J-PAS (adaptado de Bonoli et al., 2021).

objetos com magnitude aparente na banda r menores que 21.3 apresentam dispersões $\sigma_{MAD} = 0.0189$, caindo para $\sigma_{MAD} = 0.006$ para objetos com $r \leq 17.5$. Tais valores apresentados estão de acordo com os valores esperados. Portanto, selecionamos nossa amostra garantindo a completude dos dados ao aplicar um corte em magnitude aparente na banda r , $r_auto < 21$, considerando os *redshifts* entre $0.05 < z_{foto} < 0.5$. Assim como apontado na seção anterior, detalhes sobre a obtenção dos valores de z_{foto} podem ser encontrados na Seção 3.1.

Neste trabalho enfatizamos cinco das seis principais regiões: HYDRA, MC, SPLUS-S, SPLUS-N e STRIPE 82. Como nosso interesse reside na distribuição das galáxias, excluimos intencionalmente SPLUS-D devido sua concentração no disco da Via Láctea. Além das informações obtidas diretamente das imagens como fotometria e astrometria (obtidas seguindo uma *pipeline* baseada no software SExtractor, Bertin e Arnouts, 1996), o levantamento também fornece em catálogos produzidos por análises subsequentes, redshifts fotométricos, classificação dos objetos e máscaras. Estes são concentrados nos “Catálogos de Valor Adicionado” (em inglês, *value added catalogues*, VAC).

Considerando as informações fornecidas pelo VAC, utilizamos a separação quasar/estrela/ galáxia para selecionar os objetos com uma probabilidade maior que 50% de ser uma galáxia ($PROB_GAL \geq 0.5$). Essa classificação é feita a partir de técnicas de aprendizado de máquina (e.g., *random forest*), treinada com dados fotométricos provenientes do levantamento SDSS, assim como detalhado em Nakazono et al. (2021).

Para obter uma amostra pura, aplicamos também o critério de seleção $SEX_FLAGS_r \neq 2$ referente ao parâmetro do algoritmo *SExtractor* que sinaliza um objeto superposto a um outro de forma que não é bem caracterizado pelo algoritmo.

Devido à pequenas sobreposições dos campos, também aplicamos um processo de limpeza que remove objetos espacialmente próximos (dentro de 10 segundos de arco) e de magnitudes aparente próximas $\Delta r_auto < 1$. A Tabela 6.1 mostra o número de objetos em cada área após os cortes assim como os números das detecções de aglomerados e grupos encontrados.

Metodologia

Neste capítulo descrevemos os métodos utilizados ao longo do estudo, assim como o código aplicado para a localização dos grupos e aglomerados de galáxias (PZWav) e os códigos desenvolvidos para as análises de pertinência e associação de galáxias (AME e FAE).

3.1 *Redshifts fotométricos*

Os *redshifts* fotométricos têm sido explorados em vários estudos, proporcionando estimativas que permitem a determinação de propriedades intrínsecas das galáxias. Foto-zs são essenciais para analisar a distribuição e a evolução das galáxias ao longo do tempo, oferecendo a capacidade de impor limites em modelos de evolução galáctica. Além disso, esses *redshifts* podem ser ferramentas valiosas para a seleção de objetos de interesse para estudos mais profundos utilizando espectroscopia, como, por exemplo, galáxias ou quasares. Em pesquisas cosmológicas, também desempenham um papel crucial na restrição de parâmetros cosmológicos, pois os observáveis usados como traçadores de massa muitas vezes dependem do *redshift* determinado.

Os limites de desempenho na determinação de *redshifts* são influenciados pela qualidade dos dados disponíveis. Em espectroscopia, o fluxo de um objeto é medido como uma função do comprimento de onda, com resolução ($R = \lambda/\Delta\lambda$) que varia entre 100 até 30 000 (Newman e Gruen, 2022). Quando diferentes características são obtidas em um espectro, assim como linhas de emissão e absorção, o *redshift* do objeto pode ser determinado com bastante precisão. Mesmo quando essas características não são bem resolvidas, a utilização de modelos espectrais permite determinar o *redshift* a partir da combinação de linhas mais

fracas.

Por outro lado, foto-zs são obtidos através de medidas de fluxos integrados em diferentes filtros. Quando se utilizam bandas largas os erros esperados nas medidas são da ordem de $\sigma_z \sim 0.05(1+z)$, com resolução de cerca de ~ 50 . Em contraste, ao combinar diversas bandas estreitas, como é feito no caso de observações como S-PLUS (Mendes de Oliveira et al., 2019) e J-PAS (Benitez et al., 2014), esses erros são reduzidos para $\sigma_z \sim 0.01(1+z)$.

Apesar de possuírem uma menor resolução em comparação com as medidas espectroscópicas, os foto-zs são essenciais, pois permitem a estimativa de *redshifts* para grandes amostras, incluindo galáxias de baixo brilho. Tal técnica é particularmente relevante, pois a obtenção dessas informações por meio de espectroscopia seria muito custosa do ponto de vista observacional. Um exemplo citado por Newman e Gruen (2022) destaca que, para observar a “amostra de ouro” do LSST contendo 4 bilhões de galáxias com $i < 25.3$, utilizando os instrumentos do DESI (DESI Collaboration et al., 2016, equivalente a um espectrógrafo de 5 000-fibras) seriam necessários 16 000 anos de observações contínuas em condições ideais (céu escuro e limpo), para atingir o sinal ruído obtidos pelo levantamento DEEP2 (espectroscopia) para objetos com $i < 22.5$.

A representação mais robusta da informação obtida pelos algoritmos de foto-zs é a função de densidade de probabilidade (PDF) do *redshift*. Isso se deve ao fato de que as distribuições podem ser multimodais ou altamente não-gaussianas. A Figura 3.1 mostra exemplos de PDFs obtidas para 4 galáxias, 2 sendo do levantamento S-PLUS e 2 do miniJ-PAS. A normalização pelo valor máximo da PDF destaca propriedades como 2 ou mais picos.

Existem diferentes métodos para obter tais distribuições, as baseadas em ajustes de modelos (*template-based*), tais como, LePhare (Arnouts et al., 1999), BPZ (Benítez, 2000), e EAZY (Brammer et al., 2008), e as baseadas em aprendizado de máquina, ANNz (Firth et al., 2003), TPZ (Carrasco Kind e Brunner, 2013), GPz (Almosallam et al., 2016), entre outros. Ambos os métodos apresentam suas qualidades e sofrem de problemas semelhantes que podem levar a resultados enviesados: A falta de modelos que descrevam todos os dados; Amostra de treinamento que não são representativas dos dados. No entanto, esses desafios não limitam a eficácia dos foto-zs.

Por exemplo, no levantamento miniJPAS, os *redshifts* fotométricos foram estimados usando uma versão personalizada do código LePhare (Arnouts e Ilbert, 2011), que foi es-

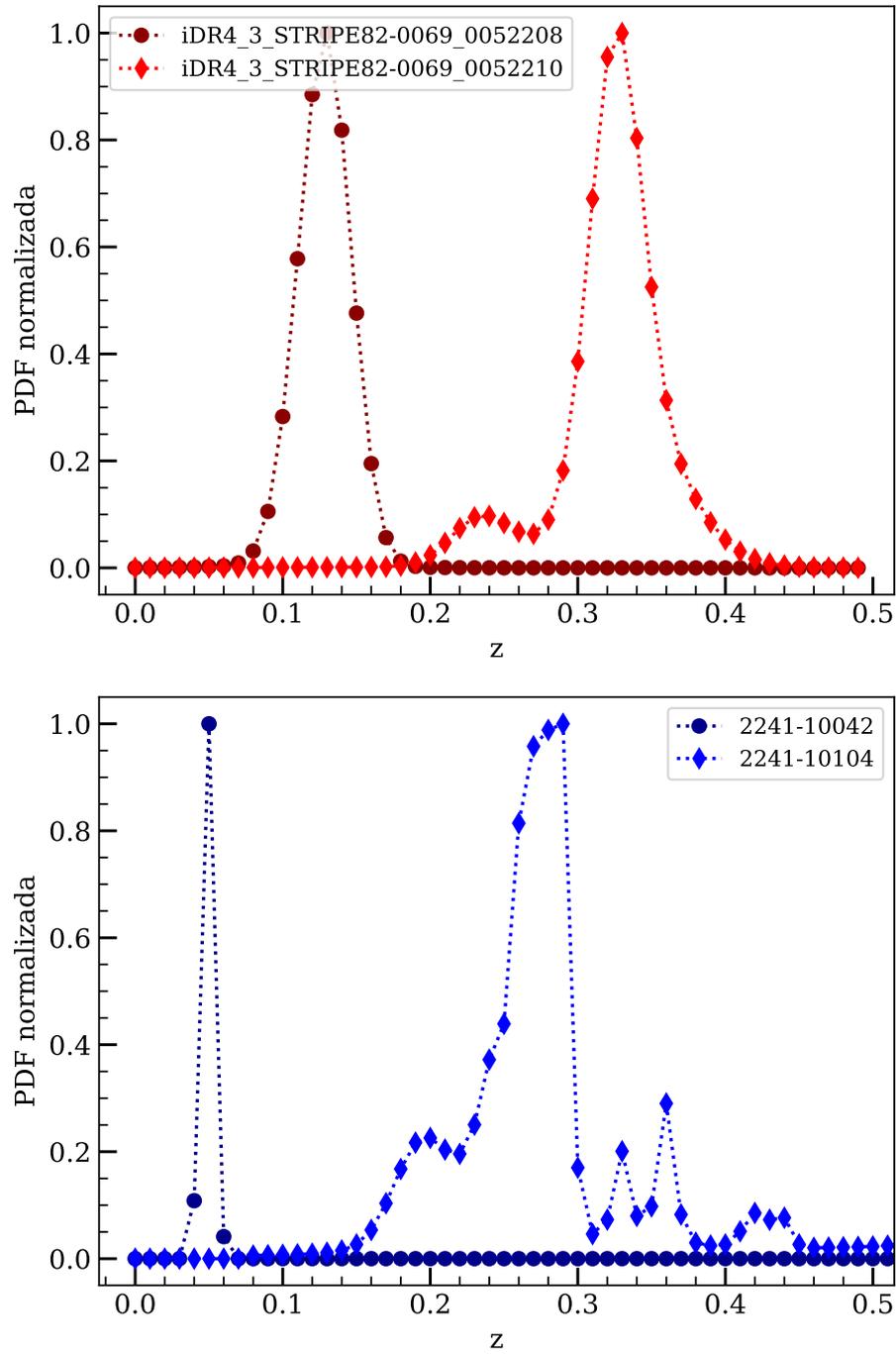


Figura 3.1: Exemplos de PDF de foto- z normalizadas pelo seu valor máximo obtidas pelos levantamentos S-PLUS (painel superior) e miniJPAS (inferior). Legenda mostra os números de identificação (ID) das galáxias em cada levantamento.

pecificamente otimizado para replicar o sistema de filtro do J-PAS. Este código calcula a PDF do z_{foto} ponderando a distribuição de log-verossimilhança com um prior que representa a distribuição do *redshift*, levando em consideração a magnitude aparente e a cor da galáxia. A função anterior é derivada da distribuição espectroscópica de *redshift* da galáxia provenientes do levantamento *VIMOS VLT Deep Survey* (Le Fèvre et al., 2005), que também fornece as funções de *redshift* e densidade de probabilidade do tipo espectral na magnitude da banda i .

Em contrapartida, no contexto do levantamento S-PLUS, os redshifts são produzidos seguindo o procedimento desenvolvido por Lima et al. (2022), baseado no método de aprendizagem de máquina *Bayesian Mixture Density Network*. O modelo é treinado utilizando as 12 bandas fotométricas disponíveis, juntamente com magnitudes nas bandas J, H, K, W1, e W provenientes de dois levantamentos distintos, 2MASS (Skrutskie et al., 2006) e unWISE (Lang, 2014). Tal método produz como resultados a PDF do *redshift* para todos os objetos, assim como uma estimativa precisa do *redshift* fotométrico, definido como o pico da PDF.

Melhorar o desempenho dos algoritmos pode ter diversas consequências, dependendo do contexto do estudo. Por exemplo, ao determinar a probabilidade de pertinência das galáxias a um determinado aglomerado, uma medida aprimorada do *redshift* reduziria as taxas de contaminação por galáxias ao longo da linha de visada, ao mesmo tempo que aprimoraria a completude da seleção.

Em situações envolvendo a determinação da função de distribuição de luminosidade ou de massa estelar, o modelamento das incertezas é crucial devido às dificuldades em relacionar a história de formação estelar com a luz emitida pela fonte (Bundy et al., 2017). Dessa forma, erros sistemáticos podem impactar as medidas de massa estelar em uma faixa de 0.1 a 0.5 dex (Courteau et al., 2014).

Nas medições de aglomerações de galáxias, aprimorar as estimativas de *redshift* amplia a detecção do sinal da sobredensidade em relação ao ruído de fundo (SNR), prevenindo a perda de galáxias dentro da faixa de *redshift* considerada devido a erros. Esse problema é reduzido ainda mais ao considerar medidas precisas das PDFs, em vez de estimativas pontuais, permitindo explorar todo o intervalo de possíveis *redshifts* do objeto.

Em aglomerados e grupos de galáxias, onde tanto a abundância esperada quanto as relações observadas com as propriedades físicas dependem do *redshift*, o desempenho dos

foto-zs impacta diretamente na medida de riqueza e no limite de massa das estruturas que podem ser distinguíveis dos objetos de campo.

Em estudos utilizando a técnica de lentes gravitacionais fracas, a amplitude do sinal depende das distâncias de diâmetro angular das galáxias lentesadas, que são determinadas pelo modelo cosmológico utilizado seu *redshift*. Portanto, métodos de calibração de massa que dependem dessa técnica também requerem uma caracterização precisa das distribuições de *redshift* dos objetos de campo (Mandelbaum, 2018).

3.2 PZWav

Nesta seção, oferecemos uma descrição detalhada do algoritmo utilizado para a localização dos aglomerados nesse projeto, indo além do resumo apresentado na Introdução (Seção 1.2.1). Apresentamos os parâmetros livres, os valores utilizados, e as razões por trás de suas escolhas.

Existe uma grande variedade de algoritmos para a detecção de aglomerados de galáxias usando dados ópticos (para uma revisão Gal, 2006; Euclid Collaboration et al., 2019). Entre eles, PZWav (Gonzalez, 2014; Werner et al., 2022) se destaca como particularmente adequado para análise de dados fotométricos devido à sua capacidade de lidar com PDFs de z_{foto} . Além disso, PZWav é um algoritmo baseado em sobredensidades que requer suposições mínimas sobre as propriedades do aglomerado.

O PZWav realiza a detecção de estruturas identificando sobredensidades em escalas físicas fixas associadas aos tamanhos típicos de aglomerados de galáxias. Requer informações como coordenadas celestes de galáxias, *redshifts* fotométricos e magnitudes aparentes. Com esses dados, o algoritmo cria uma série de fatias de *redshift*, com cada galáxia ponderada por um peso com base na probabilidade de seu *redshift* estar no respectivo intervalo analisado. Ao integrar as probabilidades sobre os limites de *redshift* de cada intervalo, o código gera mapas de densidade de galáxias 2D (Figura 3.2). Esses mapas são convoluídos com um *kernel* de diferença de Gaussianas (mais detalhes em Apêndice A), que destaca as estruturas do tamanho de aglomerado enquanto diminui o impacto de estruturas de menor e maior escala. Esse método é considerado uma forma de transformada *wavlet*, dando origem ao nome do algoritmo que combina as iniciais de “PZ” para PDFs de z_{foto} e “Wav” para *wavelet*. Além disso, para estimar o ruído relacionado, um conjunto separado

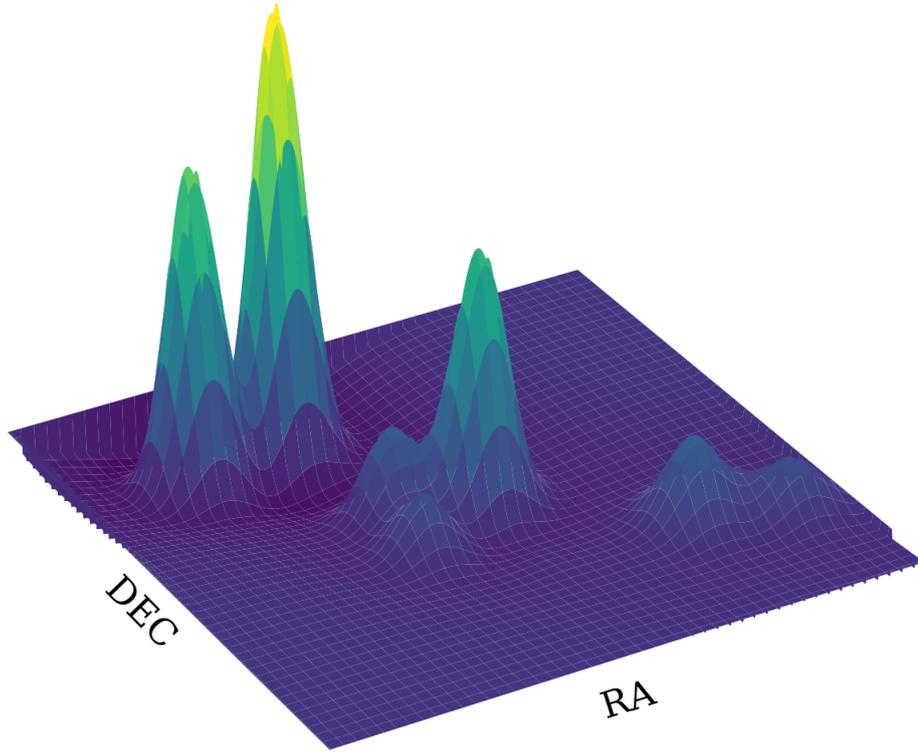


Figura 3.2: Mapa de densidades em 2D produzidos pelo PZWav. Cada pico que se desacata em relação ao ruído de campo é considerado um candidato a aglomerado.

de mapas de densidade é criado ao randomizar as posições projetadas das galáxias dentro do intervalo de *redshift*.

Em cada intervalo, um candidato a aglomerado de galáxias é identificado como um pico no mapa de densidade que excede o limite de ruído. O centro do aglomerado é definido como o centro do pico de densidade, obtido pela função `peak_local_max` que faz parte da biblioteca “scikit-image”¹. O *redshift* final do aglomerado é determinado pelo cálculo da mediana de z_{foto} das galáxias dentro de um intervalo $\Delta z = 0.02$ (equivalente a duas vezes o tamanho da fatia de *redshift*) e um raio físico de $R = 500$ kpc. Para identificar efetivamente as estruturas, definimos intervalos de *redshift* com uma largura de $dz_w = 0.01$ e restringimos as escalas de aglomerado para 400 e 1400 kpc.

A relação sinal ruído é definida como a amplitude do pico mais alto nos mapas de densidade em relação ao nível de ruído (valor médio sobre a dispersão dos picos dentro do intervalo). Para esta análise, o limite de detecção padrão é definido como $SNR_{thr} = 4$. Para evitar a dupla contagem de aglomerado passando por um possível evento de colisão,

¹ Essa função é comumente utilizada para encontrar picos em uma imagem.

definimos como parâmetros de fusão $dr_{lim} = 1500$ kpc e $dz_m = 0.03$ como a distância mínima entre duas estruturas tanto no plano do céu quanto no espaço de *redshift*. Esses valores foram encontrados a partir de testes com simulações.

3.3 Análise de pertinência das galáxias

3.3.1 O estado da arte em associação probabilística

Estudos que investigam a atribuição de galáxias como membros de grupos ou aglomerados usando *redshift* espectroscópico frequentemente partem de suposições iniciais sobre a sequência vermelha do aglomerado, sua distribuição radial ou função de luminosidade, aproveitando as medições precisas disponíveis. No entanto, ao lidar com *redshift* fotométrico, surge um desafio comum relacionado às incertezas em relação à extensão dos aglomerados de galáxias ao longo do espaço de *redshift*.

Diante desse cenário, é necessário considerar a contaminação por galáxias que não pertencem a um aglomerado ou grupo específico. Tal efeito pode ser abordado subtraindo a média das galáxias de campo próximas ao grupo, ou mesmo utilizando uma abordagem global. Outra estratégia é atribuir uma probabilidade a cada galáxia em relação à sua pertinência a um determinado grupo. Em seguida, as propriedades podem ser derivadas ao ponderar as galáxias de acordo com suas probabilidades.

Em 2000, Brunner e Lubin introduziu uma técnica que produz uma interpretação probabilística de pertinência sem depender diretamente de informações espectroscópicas. Primeiro, os autores atribuíram uma função de distribuição de probabilidade gaussiana para cada galáxia centrada no *redshift* fotométrico estimado. Então, a pertinência é obtida integrando a função de distribuição dentro do intervalo *redshift* do aglomerado de galáxias. Como veremos mais adiante (Seção 3.4) este é um método poderoso, mas que ainda apresenta uma contaminação significativa proveniente das galáxias de campo.

Este método foi aprimorado por George et al. (2011) e, posteriormente, difundido por Castignani e Benoist (2016). Os autores introduziram um formalismo Bayesiano que, juntamente com o foto-z, considera-se a distância projetada do centro do aglomerado da galáxia, as magnitudes aparentes e o tamanho relativo da população entre o campo e a densidade do aglomerado de galáxias. Esses estudos, adaptados para a amostra específica em questão, forneceram uma melhor compreensão da dependência da pertinência em função

da precisão do foto- z , magnitude das galáxias e propriedades de aglomerado. Destes, massa, riqueza, *redshift* e sua acurácia se mostram parâmetros estáveis, com variações menores que 10%. Quando consideradas galáxias de baixo brilho ($\sim 0.25 L_*$), pureza cai em aproximadamente 20%. Contaminação por galáxias de campo em R_{200} podem chegar em até 50%.

No contexto de estudos anteriores de detecção de aglomerados de galáxias e identificação da população de galáxias, os localizadores de aglomerados redMaPPer (Rykoff et al., 2014) e AMICO (Bellagamba et al., 2019; Maturi et al., 2023) avançaram significativamente ao utilizar um formalismo de filtragem radial e em *redshift*. Considerando as informações comentadas anteriormente, os métodos também incluíram modelos autosimilares para descrever a magnitude aparente esperada e as distribuições de galáxias em aglomerados. Essa modelagem permitiu uma melhor distinção entre galáxias membros e de campo, produzindo estimativas de membros com menor contaminação.

Neste trabalho, estudamos pertinências probabilísticas e sua aplicação como traçadores de massa baseados em fotometria, expandindo abordagens anteriores para explorar as PDFs dos foto- z s. Aqui, damos um passo para trás e revisitamos resultados que podem ser alcançados sem fortes hipóteses de modelagem que, *a priori*, podem não ser totalmente adequados aos nossos dados.

3.4 Estimador de abertura fixa

Para fazer uma primeira estimativa aproximada de riqueza, seguimos a ideia proposta por (Brunner e Lubin, 2000). Para cada aglomerado identificado em uma determinada posição nas coordenadas projetadas (celestiais) e *redshift*, ou o espaço $(2+1)D$, calculamos a sobredensidade de galáxias. Nesta subseção descrevemos o estimador de abertura fixa, FAE (do inglês, *fixed aperture estimator*).

As galáxias são selecionadas com base na distância do centro do aglomerado, dentro de um raio de corte de 500 kpc (como o valor mediano do raio característico R_c da amostra dos aglomerados simulados, mais detalhes na Seção 3.5 e 4.1), e intervalo *redshift*, $z_{cluster} \pm \Delta z$ (com $\Delta z = 3\sigma_{MAD,z}[1 + z_{cl}]$), onde $\sigma_{MAD,z}[1 + z_{cl}]$ é a estimativa da dispersão fotométrica no *redshift* em questão (semelhante ao $\sigma_{MAD,z}$ apresentado na Seção 2).

Uma probabilidade de pertinência pode então ser atribuída a cada uma das galáxias

selecionadas por,

$$P_{clz} = \int_{z_{cl}-\Delta z}^{z_{cl}+\Delta z} \text{PDF}_z(z') dz' \quad (3.1)$$

onde PDF_z é a função densidade de probabilidade do *redshift* fotométrico.

A riqueza em torno da posição do aglomerado é estimada como,

$$\lambda = \sum_i P_{clz,i} - \Sigma_{fld}(z_{cl}) \times A_{cl} \quad (3.2)$$

onde i é o índice que percorre todas as galáxias dentro da área do aglomerado (A_{cl}), e $\Sigma_{fld}(z)$ é a densidade superficial mediana das galáxias de campo em um determinado *redshift*. Calculamos a densidade superficial como a soma das probabilidades das galáxias por área para coordenadas do céu distribuídas aleatoriamente. Com um procedimento de corte de 3σ na densidade superficial, evitamos o viés devido à possível presença de aglomerados ou regiões de baixa densidade (vazios). Esta estratégia é adequada para pequenas e médias áreas de céu.

Outra abordagem possível para lidar com a contaminação de galáxias de campo é calcular a contribuição em anéis próximos ao centro do aglomerado. Este procedimento se destaca para grandes áreas do céu, onde uma a densidade superficial pode variar consideravelmente.

Na Figura 3.3, apresentamos relações de escala preliminares obtidas com o procedimento acima, comparadas com os valores reais do catálogo simulado (mais detalhes na Seção 2), como curvas avermelhadas e azuis, respectivamente. Os pontos vermelhos representam a riqueza de abertura fixa e os pontos azuis o catálogo simulado. Ambos os conjuntos de dados compartilham os mesmos valores de massas de halo. Os valores obtidos para a regressão são os seguintes: mock $a = 12.88 \pm 0.08$, $b = 1.32 \pm 0.11$, $\epsilon = 0.17 \pm 0.01$ e AME $a = 13.02 \pm 0.06$, $b = 0.84 \pm 0.12$, $\epsilon = 0.18 \pm 0.01$, sendo a a inclinação, b a intersecção e ϵ o espalhamento intrínseco da relação.

Enquanto a riqueza estimada pelo procedimento acima se correlaciona com a massa, a inclinação é consideravelmente plana dada a dispersão (0.84 ± 0.12) em comparação com o catálogo simulado (1.32 ± 0.11), tornando-o um traçador pouco atraente. A principal razão que encontramos para essa planicidade é o raio fixo que não leva em consideração diferenças nos tamanhos de aglomerados de diferentes massas.

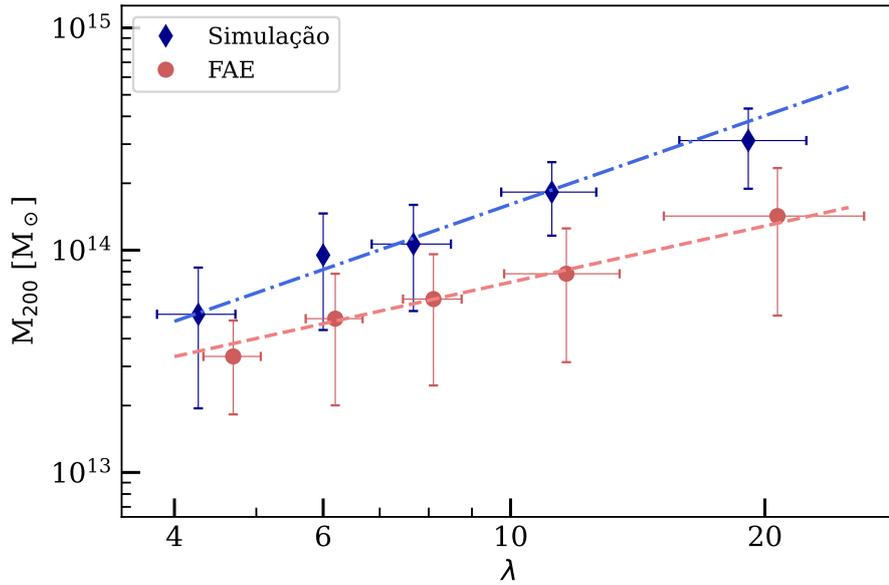


Figura 3.3: Relação riqueza – massa obtida com o estimador de abertura fixa. Os losangos azuis mostram os valores médios dos aglomerados fictícios e os círculos avermelhados mostram os resultados FAE. As linhas de mesmos tons mostram a regressão linear para ambos os conjuntos de dados.

3.5 Estimador adaptativo de pertinência de galáxias

Considerando a necessidade de ter diferentes raios para cada estrutura para reduzir o efeito da abertura na riqueza, propomos um procedimento de abertura adaptativa, chamado de AME (do inglês, *adaptive membership estimator*). Esse algoritmo leva em consideração a mesma ideologia do algoritmo PZWav: utilizar as informações precisas que obtemos dos levantamentos fotométricos na forma de PDF; e fazer uso de poucas restrições para definir propriedades do aglomerado, permitindo uma análise sem vieses relacionados a cores, luminosidades ou densidades. As entradas necessárias são as posições projetadas do catálogo de galáxias, PDFs do foto- z e a posição do aglomerado no espaço (2+1)D: z_{cl} , RA_{cl} e Dec_{cl} .

A seguir, apresentamos as principais etapas e detalhes do nosso algoritmo.

1. Remove os não-membros óbvios cortando todas as galáxias fora de um raio de 2.5 Mpc no plano do céu e com $|z_{foto,i} - z_{cl}| > 3\sigma_{MAD,z}(1 + z_{cl})$.
2. Calcula o perfil de densidade de galáxias. Um raio característico (R_c), será definido como uma quebra ou “joelho” neste perfil. Estima-se λ (Equação 3.2) dentro deste raio.

3. Assume um valor de *redshift* aleatório baseado na PDF do *redshift* fotométrico para cada galáxia dentro de R_c .
4. Calcula a dispersão de velocidades da amostra com os *redshifts* aleatórios, e aplica um corte baseado no processo de 3σ *clipping*.
5. Executa o HDBSCAN (Campello et al., 2013) no espaço 2D usando as galáxias restantes. Os parâmetros de entrada são as posições das galáxias e $\lambda(R < R_c)$, como o parâmetro de tamanho mínimo do aglomerado. A estrutura com mais galáxias é assumida como uma contraparte do aglomerado em questão.
6. Repete N vezes os passos 3–5. A probabilidade de cada galáxia ser um membro é o número de vezes que a galáxia é incluída como membro sobre N: $P_{mem} = N_{mem}/N$.

Na etapa (1.), tentamos garantir que nenhuma galáxia com possibilidades razoáveis de ser um membro do aglomerado ligado gravitacionalmente seja excluída da lista de possíveis membros do aglomerado estudado. Nesse sentido, espera-se que galáxias com $|z_{cl} - z_{gal}| > 0.04$, no caso do levantamento miniJPAS, não façam parte do aglomerado em questão.

Consistente com o processo orientado pelos dados, deixamos a distribuição radial das galáxias definir uma abertura espacial projetada (R_c) para as demais análises. Fazemos isso detectando uma descontinuidade no gradiente do perfil de densidade radial, vindo do centro, na etapa (2.). Espera-se que isso aconteça quando nossa amostra inicial deixar de ser dominada pelas galáxias do aglomerado para ser dominada pelo campo.

O perfil de densidade projetado é calculado no anel, a partir do centro, em etapas de sobreposição de 10 kpc e largura de escala logarítmica, que varia progressivamente entre 50–200 kpc. Os limites do levantamento são cuidadosamente levados em consideração ao estimar as áreas usando o método de Monte Carlo. Isso também fornece uma abordagem poderosa para lidar com possíveis áreas mascaradas em levantamentos. O raio do núcleo, R_c , é definido como uma queda/quebra repentina na densidade (por exemplo, um fator de 2 ordens de grandeza).

Para identificar esse ponto de “quebra” usamos o algoritmo Kneedle (Satopaa et al., 2011). De forma resumida, o código identifica um mínimo local contabilizando a diferença entre o perfil de densidade e uma linha reta conectando seus pontos inicial e final. R_c é então

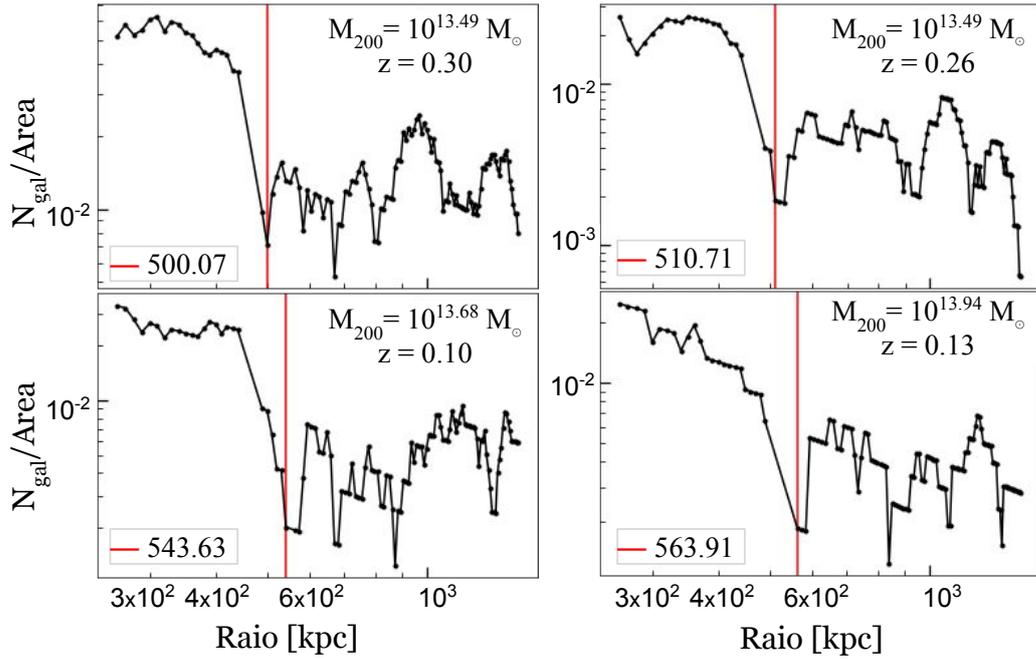


Figura 3.4: Exemplos de distribuição radial de galáxias e R_c . A área é calculada em unidades de kpc. A linha vertical vermelha destaca a descontinuidade detectada na densidade central. Adaptado de Doubrawa et al. (2023).

o primeiro (mais central) mínimo local detectado. Galáxias fora de R_c são descartadas neste ponto.

A Figura 3.4 mostra alguns exemplos da distribuição radial, e o ponto “quebra” (linhas verticais vermelhas). Apesar do comportamento ruidoso para raios maiores, a descontinuidade central é pronunciada. Testamos diferentes intervalos (largura do raio do anel) e passos (incremento do menor raio do anel). Degraus maiores podem introduzir uma diferença positiva de 30 a 60 kpc em toda a faixa de raios. Para *binning*, um esquema mais refinado não produz um viés significativo. A expectativa é que dentro desse raio o contraste sinal ruído entre as populações de galáxias de campo e aglomerado seja máximo. Na Seção 4.1 discutimos como R_c escala com R_{200} .

Nas etapas (3.) e (4.), usamos a PDF do *redshift* fotométrico em vez de sua estimativa pontual. Assumimos um valor para cada galáxia na amostra restante, imitando a realização de uma medição de *redshift* ideal.

Após esses recortes no espaço (2+1)D, ainda esperamos que a amostra esteja um pouco contaminada. Devido à qualidade do *redshift* fotométrico, os testes com galáxias membros verdadeiras produziram um valor semelhante na dispersão de velocidade. Esta análise indica que mesmo contando com uma abordagem interativa relacionando uma quantidade

simulada, por exemplo, riqueza, com a dispersão de velocidade esperada não produziria melhores resultados. A aplicação de um corte mais rigoroso (1σ) pode resolver esse problema para dados com uma PDF bem comportada, pois o limite se torna menor que a própria incerteza do *redshift*. No entanto, no caso de uma distribuição mais realista, esse procedimento pode eliminar uma parte significativa da informação. No entanto, apesar das limitações do código, podemos obter resultados precisos, conforme discutido abaixo.

Como etapa final (5.), usamos o algoritmo baseado em densidades HDBSCAN² (mais detalhes no Apêndice B). Ele conecta pontos em um espaço projetado baseando-se na proximidade mútua, considerando os mais isolados como contaminação.

Algoritmos semelhantes, como o mais conhecido DBSCAN (Ester et al., 1996), têm sido usados na astronomia para propósitos semelhantes (Bhattacharya et al., 2017; Olave-Rojas et al., 2018). DBSCAN, no entanto, requer dois parâmetros para executar: um parâmetro relacionado à distância (ϵ) e o número mínimo de vizinhos: (MIN_SAMPLES). O HDBSCAN estima ϵ por si só, variando e integrando-o na busca pelo valor mais estável. Ele ainda precisa do número mínimo de vizinhos como parâmetro de entrada.

Para fornecer essa entrada, usamos a relação entre o número de membros verdadeiros (riqueza verdadeira) definido em R_c , em vez de uma abertura fixa de 500 kpc, e λ (Equação 3.2). Mostramos na Figura 3.5 a relação entre o número de membros verdadeiros (λ_{True}) dado pela simulação e a média dos valores de $\lambda_{FAE}(R < R_c)$. Pode-se observar que os valores de riqueza que estamos obtendo são quase o dobro dos valores reais, indicando uma contaminação importante. Por outro lado, como existe uma relação linear entre essas quantidades, podemos usar $\lambda_{FAE}(R < R_c)$ para estimar MIN_SAMPLES. Como a riqueza depende do limite de magnitude absoluta escolhido, a inclinação dessa relação pode ser um pouco maior (inferior) para um limite mais raso (mais profundo).

A probabilidade final de associação é então definida executando o procedimento acima $N = 100$ vezes e estimando a probabilidade de pertinência de uma determinada galáxia pelo número de vezes que ela foi selecionada no número total de tentativas. Esse processo de iteração permite várias realizações de *redshift* das PDFs, fazendo seu uso total.

² Do inglês *Hierarchical Density-Based Spatial Clustering of Applications with Noise*

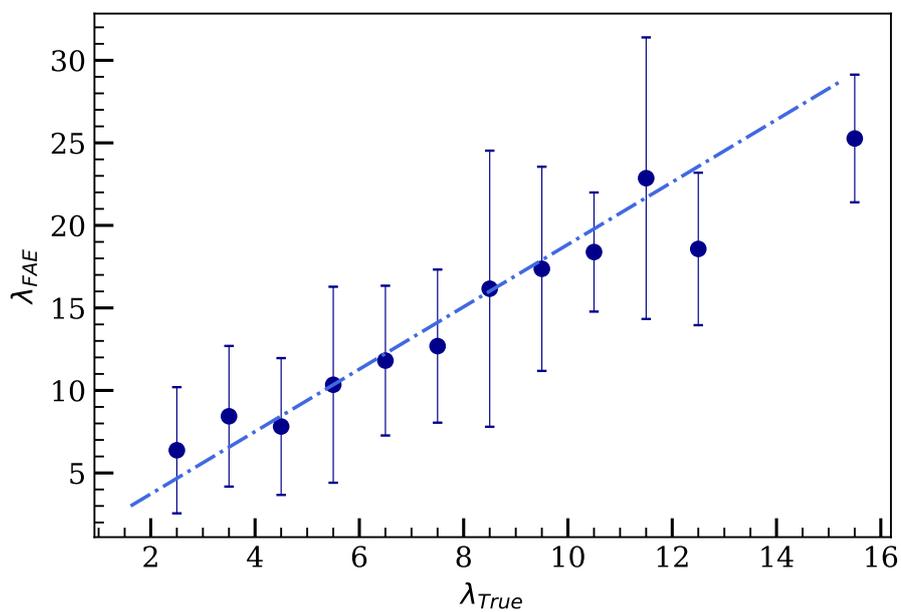


Figura 3.5: Relação entre o número de membros verdadeiros do aglomerado (dados pela simulação), λ_{True} , e os valores médios obtidos com o método FAE aplicado sobre R_c , λ_{FAE} . A linha pontilhada representa a regressão linear. Essa relação pode ser usada para inferir o número mínimo de vizinhos para executar o HDBSCAN.

Resultados utilizando simulações

Neste capítulo apresentamos os principais testes realizados e resultados utilizando o estimador probabilístico AME (Seção 3.5) quando aplicado ao catálogo simulado de galáxias, produzido para emular o primeiro lançamento dos dados do levantamento S-PLUS. Mostramos os resultados de desempenho das estimativas de riquezas, relações de escala com diferentes traçadores ópticos e os efeitos causados por pequenas perturbações no centro do aglomerado e *redshift*.

A partir da Seção 4.7 apresentamos também a significância da pertinência individual. Para isso, comparamos as probabilidades de pertencimento das galáxias com a fração de membros verdadeiros (definidos pela simulação), discutimos a completude e a pureza em termos de probabilidades de galáxias em diferentes *redshifts* e intervalos de massa de aglomerados. Por fim, comparamos as posições e *redshifts* dos aglomerados dados pela simulação com os valores estimados a partir das galáxias membro, além dos resultados obtidos utilizando o algoritmo de detecção PZWav.

Os resultados e discussões apresentados neste capítulo foram publicados em Doubrava et al. (2023).

Aqui, estamos considerando galáxias com $M_r < -20.25$. Dessa forma, temos uma amostra limitada por volume até $z = 0.45$. Calculamos as magnitudes absolutas simplesmente pela relação com a magnitude aparente, $M_r = m - 5\log_{10}(d_{\text{Mpc}}) - 25$, onde d_{Mpc} é a distância de luminosidade em Mpc.

Na Figura 4.1, mostramos os principais produtos deste exercício. O *redshift* e as distribuições de massa dos aglomerados simulados, juntamente com os principais traçadores de massa obtidos por meio do método de pertinência probabilística, AME:

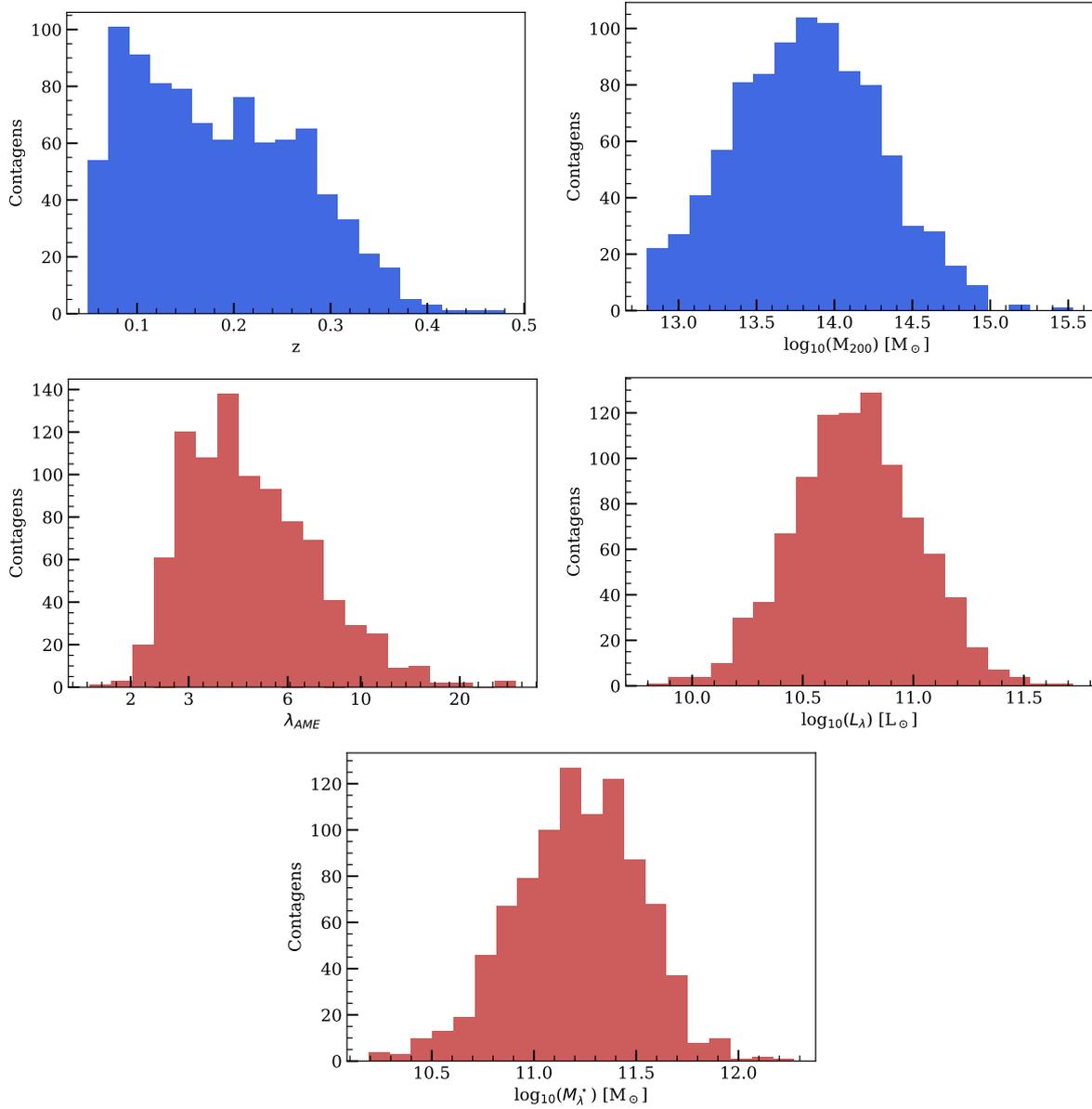


Figura 4.1: Propriedades do catálogo de aglomerados simulado. Painéis superiores: Da esquerda para a direita: distribuições de *redshift* e de massa do catálogo simulado de aglomerados de galáxias. Painéis inferiores: riqueza, λ_{AME} ; luminosidade óptica, L_λ ; e massa estelar total, M_λ^* . Tais quantidades são estimadas utilizando as probabilidades (P_{mem}) de cada galáxia pertencer ao dado aglomerado.

$$\text{Riqueza} \quad \lambda = \sum_i P_{\text{mem},i} \quad (4.1)$$

$$\text{Luminosidade óptica} \quad L_\lambda = \sum_i L_i P_{\text{mem},i} \quad (4.2)$$

$$\text{Massa estelar} \quad M_\lambda^* = \sum_i M_i^* P_{\text{mem},i} \quad (4.3)$$

Esses valores dependem de alguma forma do P_{mem} da galáxia. O total de galáxias que apresentam probabilidade maior do que zero de pertencer a algum grupo e/ou aglomerado corresponde apenas a 1.1% de todo o conteúdo de galáxias na simulação ao considerar o corte de magnitude absoluta de $M_r < -20.25$. A soma de todas as probabilidades resulta em 4664.58. A fração total de galáxias verdadeiras associadas a um aglomerado, conhecida pelas simulações e sem cortes, é de 1.75%. Esse valor cai para apenas 0.48% quando aplicados os cortes em magnitude e um número mínimo de galáxias verdadeiras associados (3, ver Seção 2.1). Sua soma total é de 4645. A discrepância entre os dois percentuais, em relação a P_{mem} (1.1%) e os membros verdadeiros (0.48%), evidencia a contaminação de nossa amostra. Entretanto, em números totais, a diferença é de apenas 21, enfatizando que apesar da contaminação, sua contribuição deve ser pequena. Vemos pela distribuição de P_{mem} apresentados na Figura 4.2, que a fração de falsos positivos aumenta para P_{mem} menores. Destacam-se em azul as probabilidades obtidas apenas para os membros verdadeiros. Discutimos na Seção 4.7 os resultados de completude e pureza de P_{mem} , bem como diferentes massas de aglomerados e intervalos de *redshift* para entender a origem da contaminação.

4.1 Significado físico de R_c

Normalmente, probabilidades de galáxias e estimativas de riqueza são contabilizadas dentro de algum raio de abertura específico relacionado ao tamanho físico do aglomerado. Os tamanhos dos aglomerados são comumente definidos como o raio dentro do qual a densidade é de 200 (500 ou outro) vezes a densidade crítica do Universo.

Para os resultados obtidos utilizando o estimador adaptativo de pertinência de galáxias (AME), também é necessário assumir um raio característico no qual o HDBSCAN atuará. Este processo é feito procurando uma quebra (R_c) no perfil de densidade numérica do

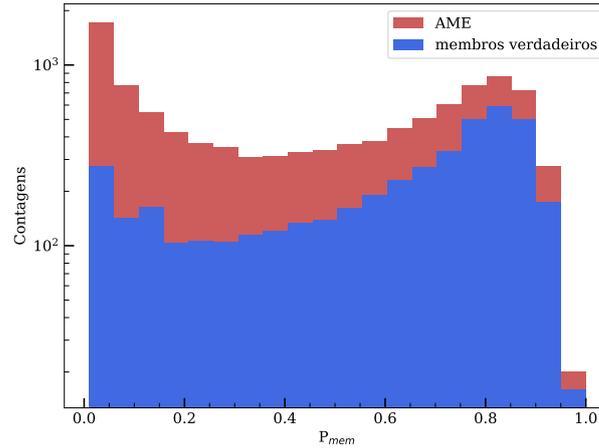


Figura 4.2: Distribuição das probabilidades calculadas por AME. Em azul, a distribuição quando consideramos apenas os membros verdadeiros.

aglomerado, conforme descrito acima (Seção 3.5). Uma comparação estatística entre os valores simulados de R_{200} e R_c revelou que R_c é, em média, 60% de R_{200} , com uma dispersão de $\sigma_{MAD} = 0.15$ Mpc. Portanto, R_c pode ser interpretado como um tamanho físico do aglomerado, próximo ao valor de R_{500} .

4.2 Teste de sanidade

Para garantir que os códigos estão funcionando corretamente, realizamos um “teste de sanidade” com um catálogo de coordenadas e *redshifts* aleatórios. Neste teste, é esperado que os valores de riqueza encontrados retornem zero, uma vez que essas posições, em princípio, não correspondem a sobredensidades significativas. Entretanto, por se tratar de uma distribuição aleatória, pode haver casos em que uma posição acaba coincidindo com uma estrutura específica.

O catálogo aleatório contém 1 000 pontos, com coordenadas do céu e *redshifts*, produzidos utilizando a função python *random.uniform*, com limites definidos pelos limites do catálogo simulado. Para R_c , não podemos aplicar simplesmente uma distribuição aleatória. Como os grupos geralmente são descritos com um raio menor e são mais numerosos em contagens, a função que descreve os raios em função da massa da estrutura segue um comportamento exponencial. Portanto, usamos os resultados R_c obtidos do catálogo de aglomerados simulado para derivar uma função de distribuição de probabilidade. Então, com *random.choice* podemos distribuir aleatoriamente valores que reproduzem os dados.

Para cada aglomerado aleatório, calculamos valores de riqueza para os estimadores de

abertura fixa e adaptativo. Os resultados são apresentados na Figura 4.3. Como esperado, o teste de sanidade mostra valores medianos próximos de zero, com σ_{MAD} de ~ 1.05 para o método FAE. Este resultado indica que valores de riqueza menores que 1.1 podem incluir sobreposições aleatórias. Na figura, observa-se também λ_{FAE} menor que zero. Esse comportamento ocorre quando a soma da probabilidade das galáxias de campo é maior que a soma das galáxias ao redor do raio fixo, indicando uma região de densidade de galáxias menor que o valor médio.

Para o AME, os valores são concentrados na linha zero sem barras de erro. No entanto, isso não significa que todos os pontos 1 000 aleatórios retornam zero. Neste caso, atingimos por acaso 240 vezes uma estrutura significativa ($R_{AME} > 1$), mas são insuficientes para contabilizar nas estatísticas (painel inferior da Figura 4.3).

Outra discussão importante para o AME é o tamanho mínimo do aglomerado obtido pelo ajuste linear. Como o parâmetro depende do estimador de abertura fixa, valores com $\lambda_{FAE} < \sigma_{MAD}$ retornam `MIN_SAMPLES` de ~ 0.6 . Mas HDBSCAN só funciona com `MIN_SAMPLES` > 2 . Se simplesmente arredondarmos os valores, forçaremos o código a encontrar pelo menos 2 galáxias espacialmente conectadas. Essa circunstância introduz um viés de pelo menos 3 para as detecções aleatórias em $z = 0.1$, que diminui para *redshifts* maiores. Para evitar essa situação, introduzimos um limite λ_{FAE} , onde $\lambda_{FAE} < \sigma_{MAD}$ HDBSCAN não pode ser executado, resultando em $\lambda_{AME} = 0$.

4.3 Significância da riqueza

O principal objetivo deste estudo é desenvolver uma metodologia que, utilizando informações fotométricas, produza um traçador de massa óptica com baixo espalhamento. Um método que pode ser aplicado a qualquer catálogo de aglomerados, variando de grupos a aglomerados de galáxias, sem depender de fortes hipóteses de modelagem.

No entanto, para usar a riqueza óptica como um traçador para relações de escala, precisamos ser capazes de quantificar a riqueza com boa precisão. Para verificar sua significância, podemos realizar uma análise em intervalos de riqueza.

Como conhecemos a riqueza verdadeira da amostra (λ_{True}), podemos comparar os valores médios dados pela simulação com a riqueza calculada pelo AME (λ_{AME}). Lembramos que, para uma comparação significativa, a riqueza verdadeira é dada como a soma das

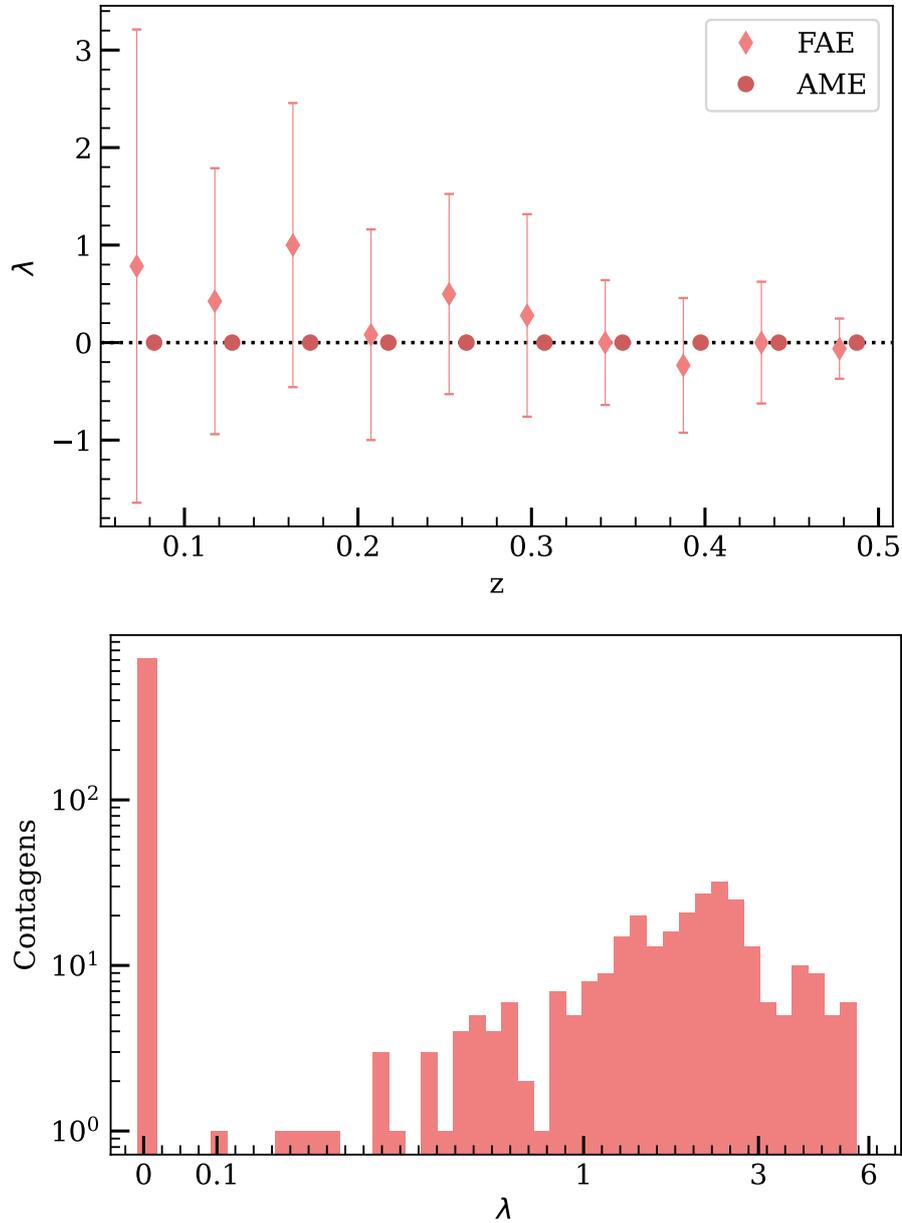


Figura 4.3: Distribuição de riquezas. Painel superior: Valores de riqueza para coordenadas do céu distribuídas aleatoriamente em função do *redshift*. Os valores foram obtidos para os estimadores de abertura fixa (FAE) e adaptativa (AME), representados por diamantes em tons mais claros e círculos mais escuros, respectivamente. Painel inferior: Histograma de riqueza obtido com AME em escala logarítmica. Apenas 240 em 1000 coordenadas aleatórias têm um valor significativo (maior que 1).

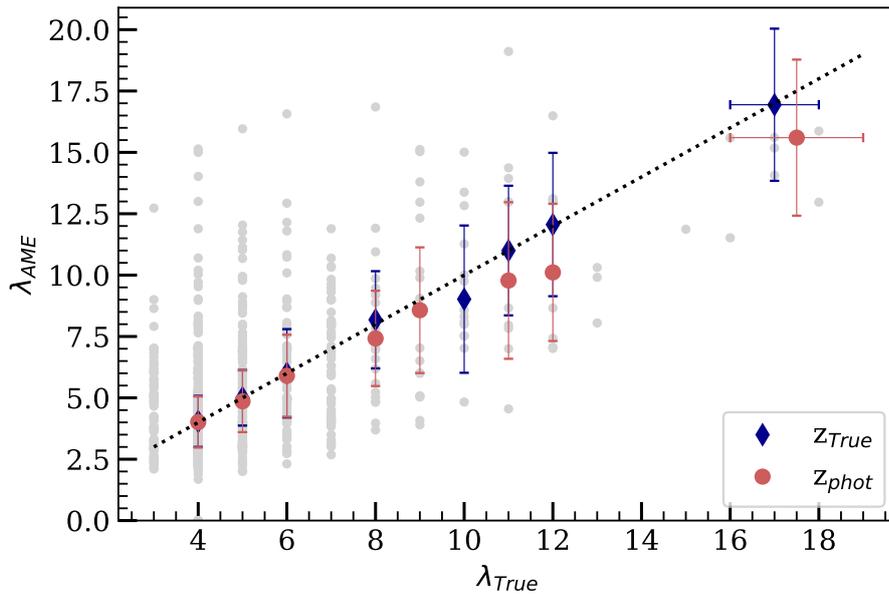


Figura 4.4: Significância da riqueza em comparação entre a riqueza calculada com o estimador adaptativo (λ_{AME}) e a riqueza “verdadeira” fornecida pela simulação (λ_{True}). Cada intervalo de riqueza verdadeira contém pelo menos 10 aglomerados. Em pontos avermelhados as riquezas obtidas com *redshift* fotométrico (z_{phot}), e em diamantes azuis os resultados com o *redshift* fornecido pela simulação (z_{True}). Os pontos em cinza mostram o espalhamento da distribuição. A linha pontilhada preta mostra a relação um-para-um.

galáxias identificadas pela simulação localizadas dentro de R_c do centro do aglomerado. A Figura 4.4 apresenta a concordância entre ambas as grandezas. Em pontos avermelhados estão representados os resultados obtidos com o *redshift* fotométrico (z_{phot}), e em diamantes azuis com o *redshift* verdadeiro fornecido pela simulação (z_{True}). Uma linha preta pontilhada mostra a linha um-para-um. O valor residual entre as quantidades, $\lambda_{True} - \lambda_{AME}$ é -0.004 ± 0.115 para z_{phot} . Valor que concorda com o melhor cenário (z_{True}) 0 ± 0.108 . Esse resultado baixo e imparcial reafirma a qualidade de nossa estimativa de riqueza. Por simplicidade e concordância, discutimos os próximos resultados utilizando z_{phot} . Nessa análise, garantimos que cada intervalo contenha pelo menos 10 aglomerados. Destacamos também a presença mais abundante de objetos de menores riquezas.

4.4 Relação massa-observável

A principal dificuldade em usar aglomerados de galáxias para estudos cosmológicos é a medição de suas massas para cada objeto. Uma alternativa é correlacionar a massa com outras propriedades observacionais, como riqueza óptica, luminosidade de raios-X ou massa estelar total. Essa relação entre massa e observável é geralmente calibrada utilizando uma

amostra limitada de objetos e depois estendida para o catálogo completo. No caso de pesquisas utilizando lentes gravitacionais, por exemplo, considera-se a amostra completa a partir da técnica do empilhamento. Portanto, um bom traçador de massa deve apresentar uma baixa dispersão intrínseca para produzir estimativas de massa robustas.

AME retorna, para cada galáxia, a probabilidade de estar conectada a um determinado aglomerado. Portanto, utilizando essa capacidade, podemos calcular outras propriedades, ponderadas pela pertinência, para caracterizar a amostra por aglomerado. Por exemplo, com a informação de magnitude na banda r podemos estimar a luminosidade óptica total, $L_\lambda = \sum L_i P_i = \sum 10^{0.4[4.65 - M_i]} P_i$ (Equação 4.2). A magnitude absoluta solar é dada por 4.65 na banda r (*SDSS_r*, Willmer, 2018), e a magnitude absoluta i -ésima galáxia na mesma banda por M_i . A massa estelar da galáxia simulada pode ser usada para derivar a massa estelar total, seguindo o mesmo procedimento descrito acima, $M_\lambda^* = \sum M_i^* P_i$, com M_i^* como a i -ésima massa estelar da galáxia (Equação 4.3).

Uma abordagem semelhante pode ser aplicada para derivar as quantidades simuladas. Neste caso, P_i é sempre 1 para os membros verdadeiros e zero caso contrário.

Com nossos traçadores de massa, podemos derivar as relações de escala. A regressão linear é feita por *Linmix* (Kelly, 2007, mais detalhes no Apêndice C). Este algoritmo usa uma abordagem Bayesiana e considera erros em ambos os parâmetros do processo de minimização, massas e traçadores, o que é ideal para dados reais. A relação é modelada como,

$$\log_{10} \left(\frac{M_{200}}{M_\odot} \right) = \alpha + \beta \log_{10} \left(\frac{O}{O_{piv}} \right) \pm \epsilon \quad (4.4)$$

onde os coeficientes são representados por α e β , O é o traçador de massa, O_{piv} é um valor pivô estimado como o valor médio do traçador simulado e ϵ é a dispersão aleatória intrínseca da regressão. A Tabela 4.1 mostra os parâmetros de melhor ajuste resultantes para a simulação, isto é, valores verdadeiros (marcadas com um símbolo †) e valores estimados com AME, como as variáveis sem marcações. A relação entre os valores medianos de massa e λ (painel superior), L_λ (meio) e M_λ^* (inferior), separados em intervalos do dado traçador, pode ser encontrada na Figura 4.5. Os losangos azuis representam a simulação e os círculos avermelhados os resultados obtidos com AME.

Estatisticamente, a distribuição mais dispersa é a relação riqueza-massa. Particularmente, os resultados obtidos a partir da simulação destacam a já presente e alta dispersão intrínseca para os pequenos grupos. Uma mesma estrutura com $\log_{10}(\lambda) = 0.6$ ($\lambda = 4$)

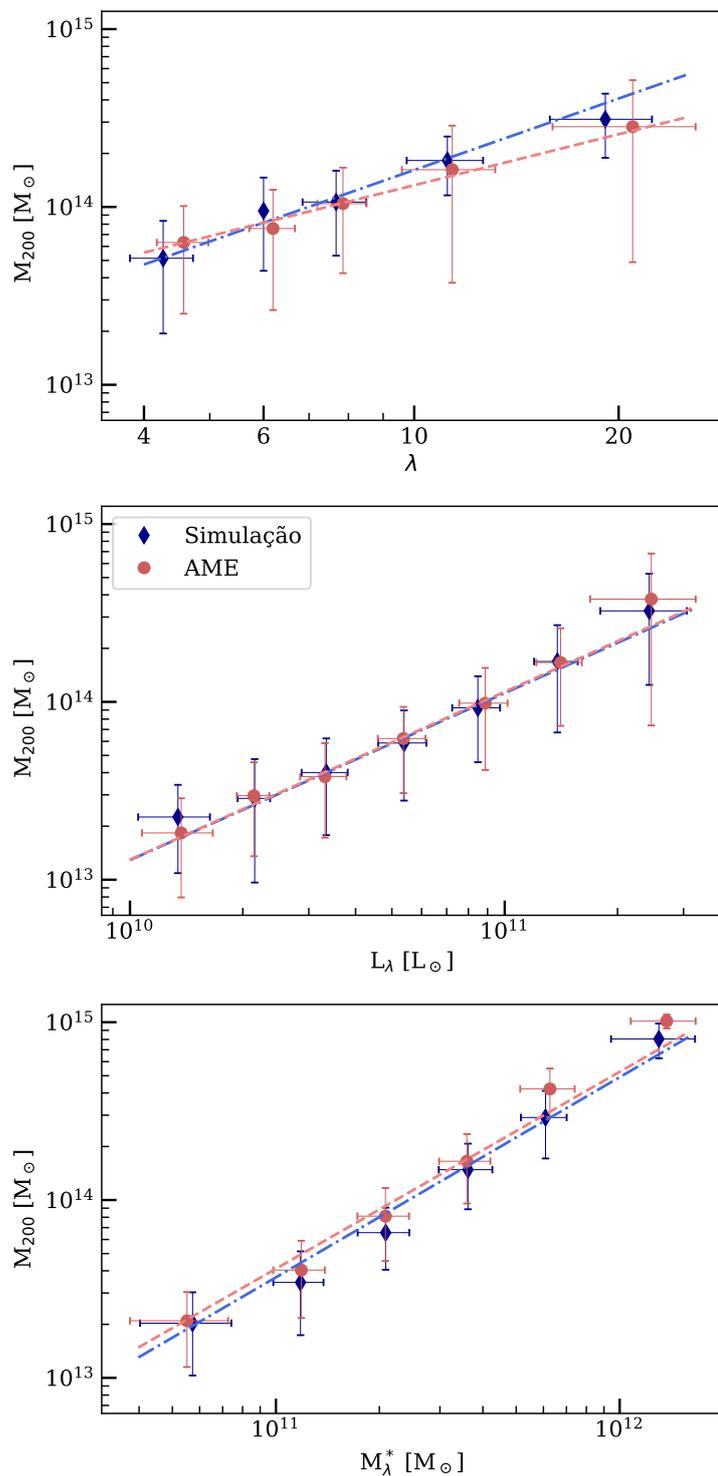


Figura 4.5: Relações de escala entre massa e traçadores ópticos. Destacamos os valores medianos em intervalos do traçador em questão. Os resultados do AME são apresentados como pontos avermelhados, os valores dados pela simulação como losangos azuis e suas respectivas regressões lineares, em linhas salmão e azul. Os valores dos coeficientes podem ser encontrados na Tabela 4.1. Do painel superior ao inferior: Relação Massa - Riqueza; Massa -Luminosidade óptica; Massa - Massa estelar.

Tabela 4.1 - Valores de melhor ajuste da regressão linear para riqueza (λ), luminosidade óptica (L_λ) e massa estelar total (M_λ^*). O modelo massa-observável é descrito pela Equação 4.4. O_{piv} está sempre relacionado ao traçador verdadeiro (simulação). L_λ e M_λ^* são dados em unidades de L_\odot e M_\odot . Os resultados obtidos com a simulação são identificados por † após o traçador observável.

traçador	α	β	ϵ	O_{piv}
λ^\dagger	13.85 ± 0.02	1.32 ± 0.07	0.180 ± 0.009	5.4
λ	13.86 ± 0.02	0.95 ± 0.10	0.181 ± 0.009	"
L_λ^\dagger	13.89 ± 0.01	0.94 ± 0.05	0.141 ± 0.007	6.7×10^{10}
L_λ	13.89 ± 0.01	0.95 ± 0.05	0.151 ± 0.007	"
$M_\lambda^*^\dagger$	13.92 ± 0.01	1.12 ± 0.04	0.092 ± 0.005	2.4×10^{11}
M_λ^*	13.99 ± 0.01	1.12 ± 0.03	0.097 ± 0.005	"

pode ter uma massa entre 10^{13} e $10^{13.9} M_\odot$. Apesar dos diferentes extremos de alta riqueza, ambas as regressões lineares (simulação e AME) apresentam o mesmo comportamento. A dispersão intrínseca observada é $\sigma_{\log_{10}(M|\lambda)} = 0.181 \pm 0.01$ dex, um valor semelhante ao melhor cenário, ou seja, simulações. A luminosidade óptica total é um parâmetro válido com espalhamento $\sigma_{\log_{10}(M|L_\lambda)} = 0.151$ dex, em comparação com $\sigma_{\log_{10}(M|L_\lambda)} = 0.141$ dex da simulação. Como L_λ depende principalmente das magnitudes e das probabilidades das galáxias, a mesma amplitude de massa observada na relação entre massa-riqueza é observada aqui. Vemos um pequeno desvio para estruturas de menor luminosidade, provavelmente devido a contaminação externa, que aumenta ligeiramente a dispersão intrínseca em relação ao melhor cenário. M_λ^* é uma opção interessante e apresenta a menor dispersão intrínseca. Ela fornece uma caracterização dos candidatos a aglomerados de galáxias em relação às propriedades físicas, ou seja, a massa estelar da galáxia. Esses resultados podem ser excessivamente otimistas, pois usamos os valores exatos obtidos pelos modelos. Em levantamentos ópticos, dispersões podem ser introduzidas pelo método de inferência e pelos erros sistemáticos, ao considerar, por exemplo, suposições assumidas para o cálculo das quantidades (equilíbrio do sistema, perfil esférico/radial, entre outros). O mesmo comportamento encontrado em L_λ para a extremidade inferior é visto aqui. Observamos também uma pequena diferença em β que introduz um deslocamento vertical de ~ 0.01 .

4.5 Perturbações no centro e redshift do aglomerado

Podemos testar a robustez do AME aplicando pequenas variações no *redshift* e nas coordenadas no catálogo simulado de aglomerados que imita as perturbações esperadas

Tabela 4.2 - O erro relativo mediano entre os resultados de riqueza antes e depois dos deslocamentos ΔR e Δz , considerando distribuições gaussianas de 1σ , 2σ e 3σ .

	ΔR [kpc]	Δz
$\bar{O} + 1\sigma$	0 ± 0.010	-0.02 ± 0.05
$\bar{O} + 2\sigma$	0 ± 0.011	-0.08 ± 0.10
$\bar{O} + 3\sigma$	0 ± 0.013	-0.18 ± 0.18

encontradas nas detecções de aglomerados de galáxias utilizando dados observacionais.

Werner et al. (2022) discute a aplicação do algoritmo de detecção de aglomerados PZWav para detectar aglomerados no S-PLUS/DR1 e analisa o desempenho do algoritmo comparando os resultados com o mesmo catálogo simulado usado em nosso trabalho. Devido à abordagem computacional do PZWav, que detecta subestruturas pela distribuição espacial das galáxias e estima o *redshift* também com base nas galáxias próximas ao centro do aglomerado, as variações encontradas nas distâncias centrais e *redshift* são notavelmente pequenas, uma diferença radial média de 10 kpc com um desvio padrão de 12 kpc, e média de 0.6×10^{-3} com desvios de 8.8×10^{-3} para *redshifts*.

Aqui, testamos diferentes cenários que incluem distribuições centradas no valor médio, e dispersões de 1σ , 2σ e 3σ , e comparamos os resultados de riqueza obtidos com os centralizados. Para as variações de coordenadas ΔR , aplicamos um deslocamento aleatório descrito por uma distribuição gaussiana 2D (tal como Johnston et al., 2007, para *miscentering*). Para orientar os deslocamentos atribuímos um valor aleatório entre 0 e 360.

Enquanto a variação radial é sempre positiva, Δz pode ser aplicado como deslocamento positivo ou negativo. Para isso, utilizamos distribuições normais que aleatoriamente somam ou subtraem do *redshift* do aglomerado.

Na Tabela 4.2, mostramos os erros relativos medianos, e σ_{MAD} , entre os resultados de riqueza para ΔR e Δz separadamente. Como não houve tendências significativas ao longo do *redshift*, apresentamos apenas um resultado contendo todos os aglomerados de galáxias para cada cenário.

Para a distância ao centro, não encontramos um desvio efetivo nos resultados, ou seja, os erros medianos são consistentes com zero. Esta análise mostra que o código é robusto para variações de centralização menores que um valor médio de $\Delta \bar{R} = 10$ kpc. Investigando mais a fundo, descobrimos um erro mediano menor que -2% até $\Delta \bar{R} = 250$ kpc, aumentando para -35% para $\Delta \bar{R} = 700$ kpc e -50% por $\Delta \bar{R} = 900$ kpc. O sinal negativo representa

uma subestimação da riqueza. Esses valores indicam que o código é sensível a diferenças na densidade de galáxias nos arredores do aglomerado e tende a escolher a maior concentração de galáxias como contraparte preferencial.

Para o *redshift*, observamos uma diminuição significativa (porém, dentro das barras de erro) nas medições de riqueza, passando de -8% para 2σ , para um desvio de -18% para 3σ . Então, afastando-se do centro do aglomerado, na linha de visada, com uma diferença tão grande quanto a incerteza do foto-z, faz com que se exclua uma contribuição significativa das PDFs das galáxias próximas ao *redshift* do aglomerado.

Ainda assim, em relação aos pequenos deslocamentos esperados, esta análise mostra que o código é robusto tanto para as variações de centro quanto para o *redshift*.

4.6 PDFs não gaussianas

Ao longo deste trabalho, aproximamos as PDFs de foto-z como distribuições gaussianas bem comportadas. No entanto, PDFs de observações podem ser mais complexas, muitas vezes apresentando longas caudas ou picos secundários. Como as PDFs de foto-z desempenham um papel central na definição da riqueza, testamos também os efeitos da não gaussianidade em nossa amostra.

Para abordar os efeitos das caudas longas, descrevemos a amostra de galáxias com distribuições t de Student. O valor central é dado pelo *redshift* fotométrico da galáxia, com largura $\sigma_{MAD,z}$ (Seção 2). Adotamos o parâmetro de “graus de liberdade” $\nu = 1$ para tornar as asas muito mais largas que as Normais. Este teste produziu resultados semelhantes em termos de riqueza, com um erro relativo de -2% , conforme apresentado na Tabela 4.3.

Para testar o efeito de uma PDF bimodal, utilizamos alguns exemplos do banco de dados S-PLUS. Definimos que uma PDF bimodal apresenta um pico secundário significativo quando a distância entre os picos é maior que 0.05, e a amplitude do pico secundário é de pelo menos 20% do pico primário. Para cada galáxia na simulação, atribuímos uma PDF de pico duplo semelhante ao S-PLUS, tendo como parâmetros de semelhança o *redshift* e magnitude das galáxias. Testamos diferentes frações de picos duplos, como 10%, 25%, 50% e 100%, e descobrimos que as medidas de riqueza são afetadas apenas para valores maiores que 50%, levando a subestimativas de 15%.

Tabela 4.3 - Comparação entre os resultados de riqueza de AME com PDFs de foto-z mais realistas como distribuição t de Student e diferentes frações de bimodalidade.

foto-z PDF	fração	Erros relativos medianos
Distribuição de t de Student	100%	-0.02 ± 0.03
Bimodalidade	10%	0.01 ± 0.02
	25%	-0.02 ± 0.05
	50%	-0.15 ± 0.06
	100%	-0.31 ± 0.08

Como este trabalho é inspirado em levantamentos fotométricos que utilizam vários filtros de banda estreita, não esperamos que o efeito de pico duplo seja um problema mais significativo do que nosso teste. Pesquisas como S-PLUS e J-PAS utilizam uma combinação de sistemas de filtro que permitem uma melhor restrição das PDFs de foto-z. O S-PLUS, por exemplo, apresenta fração de pico secundário inferior a 6%, enquanto esse valor para o levantamento J-PAS é menor que 8%.

4.7 Medidas de pertinência individuais

Nas seções anteriores, focamos nossa análise nos subprodutos do estimador adaptativo. Aqui vamos investigar o significado das probabilidades individuais obtidas utilizando uma abordagem semelhante a Castignani e Benoist (2016) e Lopes e Ribeiro (2020). A ideia é comparar a fração de membros verdadeiros, f_{true} , dada pela simulação, e o estimado P_{mem} . Os dados das galáxias são agrupados em P_{mem} , e para cada intervalo é calculado f_{true} . Para a comparação, estamos assumindo como galáxias membro verdadeiras as que foram identificadas pela simulação dentro de R_c .

Aqui também apresentamos um caso de teste baseado nas informações já conhecidas da simulação para observar como um limite mais estrito na dispersão de velocidades, apresentado no passo 4 da Seção 3.5, pode melhorar nossos resultados. Primeiro, executamos o AME utilizando apenas as galáxias membros verdadeiras em toda a amostra do aglomerado e estimamos a dispersão de velocidades “real” dados os fotos-zs. Como não se espera nenhuma evolução significativa com riqueza, tomamos o valor mediano da amostra. Este passo produz um σ menor do que considerando todas as galáxias (membros e não membros) dentro de R_c . Também modificamos o processo de recorte: ao invés do padrão 3σ , somente as galáxias dentro de 1.5σ são aceitas. Por fim, executar o AME com os

novos limites deve produzir resultados mais puros, porém menos completos.

As frações resultantes são apresentadas na Figura 4.6; uma linha pontilhada preta mostra a linha um-para-um, os pontos avermelhados representam nosso modelo padrão e os losangos azuis são os valores P_{mem} obtidos para a simulação com o rigoroso limite $\sigma_{cut} = 1.5\sigma$.

Em ambos os casos, detectamos uma fração ligeiramente maior de membros verdadeiros para P_{mem} baixos. Por outro lado, na extremidade alta de P_{mem} , a contaminação é maior do que o esperado. A combinação de ambos os efeitos produz uma curva achatada na Figura 4.6. Uma fração da contaminação da linha de visada é esperada, pois as incertezas fotométricas do *redshift* são uma ordem de grandeza maior do que a dispersão da velocidades do aglomerado. A fração média encontrada é de $\sim 11\%$, com maiores valores de contaminação para $P_{mem} > 75\%$. Esse comportamento indica que algumas galáxias de campo com o pico máximo das PDFs de foto-z próximos aos *redshifts* do aglomerado não podem ser removidos com o corte em 3σ . Ainda assim, um corte mais rigoroso poderia reduzir sua contribuição. Este efeito reflete na fração média de contaminação com o σ_{cut} que é diminuída para $\sim 6\%$. No entanto, o limite σ_{cut} é menor que a própria incerteza do *redshift*, restringindo P_{mem} a 85% e introduzindo uma pequena subestimação em nossos traçadores ópticos.

Ressaltamos que a contaminação por “intrusos” ainda ocorre mesmo em uma situação controlada onde conhecemos *a priori* a dispersão de velocidades da amostra. Apesar da maior fração de intrusos encontrada utilizando o limite mais raso orientado por dados, ainda obtemos resultados precisos que nos permitem produzir relações de escala com dispersão intrínseca competitiva. Uma melhor caracterização de como a contaminação afeta a pureza de nossa amostra é discutida abaixo (Seção 4.8).

4.8 Completude e pureza

Outra maneira de quantificar a robustez do nosso método é analisar a completude (C) e a pureza (P). Normalmente, a completude é descrita como a fração de membros verdadeiros classificados corretamente como um membro, e a pureza é a fração de membros verdadeiros em relação a todos os objetos selecionados como membros.

Utilizando a mesma notação de outros estudos de classificação (Castignani e Benoist,

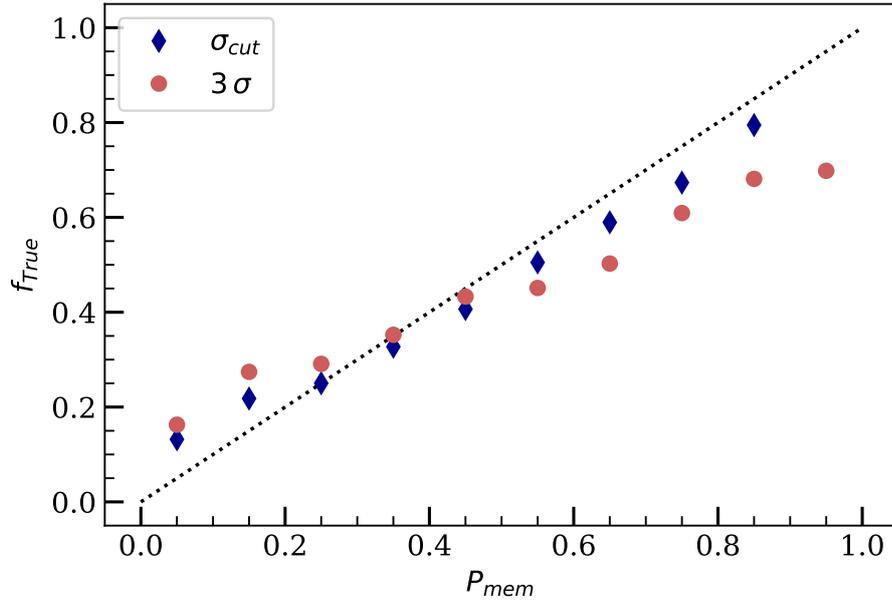


Figura 4.6: Significância da pertinência da galáxia como uma comparação entre a fração de membros verdadeiros f_{true} e as probabilidades estimadas P_{mem} , para duas abordagens diferentes: o padrão 3σ e um corte mais rigoroso σ_{cut} com base em informações conhecidas da simulação.

2016; Lopes e Ribeiro, 2020), estimamos C e P como,

$$C = \frac{N_{\text{verdadeiro}} - N_{\text{perdido}}}{N_{\text{verdadeiro}}} \quad (4.5)$$

$$P = \frac{N_{\text{selecionado}} - N_{\text{intrusos}}}{N_{\text{selecionado}}} \quad (4.6)$$

onde $N_{\text{selecionado}}$ é o número de galáxias que foram classificadas como membros, $N_{\text{verdadeiro}}$ representa o número de membros verdadeiros dado pela simulação, N_{intrusos} dá o número de objetos falsos positivos (erroneamente classificados como um membro verdadeiro) e N_{perdido} são os objetos falsos negativos (um membro verdadeiro com saída negativa).

Podemos explorar as variações tanto na pureza quanto na completude selecionando galáxias com probabilidades de associação maiores que um determinado valor limite. Na Figura 4.7, mostramos completude e pureza como uma função de diferentes limites P_{mem} , destacando $P_{mem} > 0.50, 0.75, 0.90$ em diamantes negros. Os resultados correspondentes obtidos para o modelo padrão (3σ) são $P = 0.58, 0.63, 0.66$ e $C = 0.91, 0.85, 0.38$. Nota-se que aumentar o valor limite acima de 0.75 não melhora significativamente a pureza (menos de 3%), sendo consistente com o achatamento da curva na Figura 4.7. Com o corte rigoroso (σ_{cut}), observa-se um ganho em completude e pureza de $\sim 5\%$.

Para identificar possíveis dependências com o *redshift* do aglomerado e a massa do halo,

dividimos a amostra de aglomerados em intervalos de *redshift* de $dz = 0.1$ e $d\log_{10}(M) = 0.75$. Figura 4.8 mostra os resultados destacando o limite P_{mem} de 0.2, 0.5, 0.7 e 0.8 como um círculo, quadrado, losango e triângulo, respectivamente. Em uma abordagem semelhante, Castignani e Benoist (2016), utilizando um $P_{mem} > 0.2$, observou valores médios com apenas algumas variações percentuais ($\sim 3\%$). Tomando para comparação $P_{mem} > 0.2$, vemos no painel superior da Figura 4.8 uma grande diferença entre $0.05 < z < 0.15$ (linha azul) e os outros intervalos de *redshift*. Essa diferença é de 9% em completude e sem ganho significativo de pureza (2%). Comportamento similar é observado na completude para os demais cortes P_{mem} ($\sim 12\%$), mas, nesses casos, observamos também uma melhora na pureza ($\sim 5\%$). Variações menores são observadas dentro de $0.15 < z < 0.35$, com melhorias crescentes tanto em P quanto em C ($\sim 5\%$, $\sim 2\%$). A linha vermelha $0.35 < z < 0.45$ mostra uma diminuição na completude para $P_{mem} > 0.8$ de 8.7%, com os maiores valores de pureza (81%).

Para a evolução de massa, no entanto, observamos uma diminuição em C e P para objetos mais massivos. Para $P_{mem} > 0.2$, C mostra uma variação de -4.5% e P a -6.6% (entre as linhas azul e laranja). Variações maiores para cortes P_{mem} altos ($C = -5.7, -8.3, -11.9$ e $P = -7.7, -8.1, -6.5$ para $P_{mem} > 0.5, 0.7, 0.8$, respectivamente).

Concluimos que este comportamento não se origina de aglomerados pobres, mas é devido à contaminação na periferia do aglomerado pelas galáxias de campo. Além da capacidade do código de remover pontos espacialmente dispersos, ainda são considerados aqueles espacialmente muito próximos aos verdadeiros membros da galáxia. Aglomerados massivos tendem a apresentar menor densidade central de galáxias do que a escala de grupos (Comerford e Natarajan, 2007; Merten et al., 2015). Esse efeito demonstra que o código estará mais suscetível a aceitar contribuições de galáxias de campo, aumentando a contaminação. Essa menor densidade central do aglomerado pode levar a uma negligência dos arredores. Objetos de menor massa tendem a ser mais concentrados, levando a resultados mais precisos.

Estudos de atribuições probabilísticas de pertinência normalmente selecionam um limite de probabilidade para contabilizar apenas membros confiáveis para o aglomerado. Neste trabalho, consideramos a contribuição de todas as galáxias para a estimativa de riqueza. Uma abordagem mais cuidadosa deve ser considerada no caso de análise de aglomerados únicos devido à fração não desprezível de intrusos. No entanto, para grandes

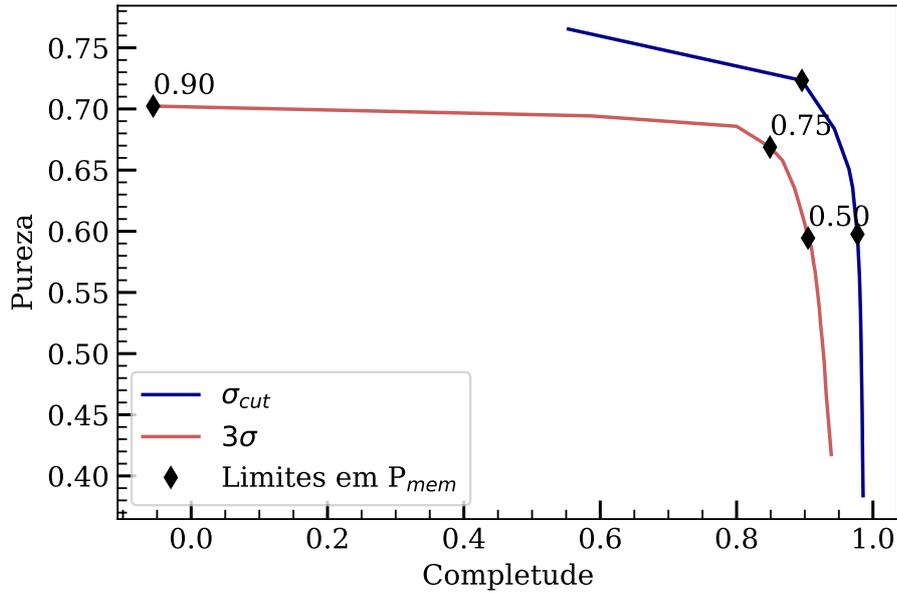


Figura 4.7: Valores de completude e pureza aplicando diferentes limites de P_{mem} . Destacamos $P_{mem} > 0.50, 0.75, 0.90$. Para comparação, mostramos em tons avermelhados os resultados obtidos com o modelo padrão (3σ) e com a corte rigoroso (σ_{cut}) em azul.

amostras, apresentamos resultados estatisticamente robustos com baixas diferenças medianas e pureza e completude que são comparáveis a outros traçadores ($P_{mem} > 0.5$ $C = 92\%$ $P = 69\%$, $P_{mem} > 0.5$ $C = 68\%$ $P = 69\%$, George et al., 2011; Castignani e Benoist, 2016, respectivamente).

4.8.1 Posições e redshifts de aglomerados a partir de P_{mem}

Outro teste que podemos realizar em relação ao nosso conhecimento sobre as galáxias associadas a um determinado aglomerado é a diferença entre o *redshift* derivado das galáxias membros e o valor obtido pelas galáxias verdadeiras, bem como as discrepâncias em relação aos centros.

Nesta análise, podemos também aproveitar o catálogo de aglomerados de galáxias produzido pelo algoritmo PZWav para a simulação e comparar como esses valores mudam em um caso real de detecção. Para esta análise, consideramos como “contraparte” as detecções do PZWav que possuem uma diferença de *redshift* com a simulação de 0.05 e distância central de 500 kpc.

Para *redshifts*, podemos calcular o valor médio ponderando cada contribuição de foto-z da galáxia por seu P_{mem} , como $z_{gals} = \sum z_i P_{mem,i} / \sum P_{mem,i}$. Esse procedimento garante que as galáxias fracamente relacionadas ao aglomerado tenham uma menor contribuição

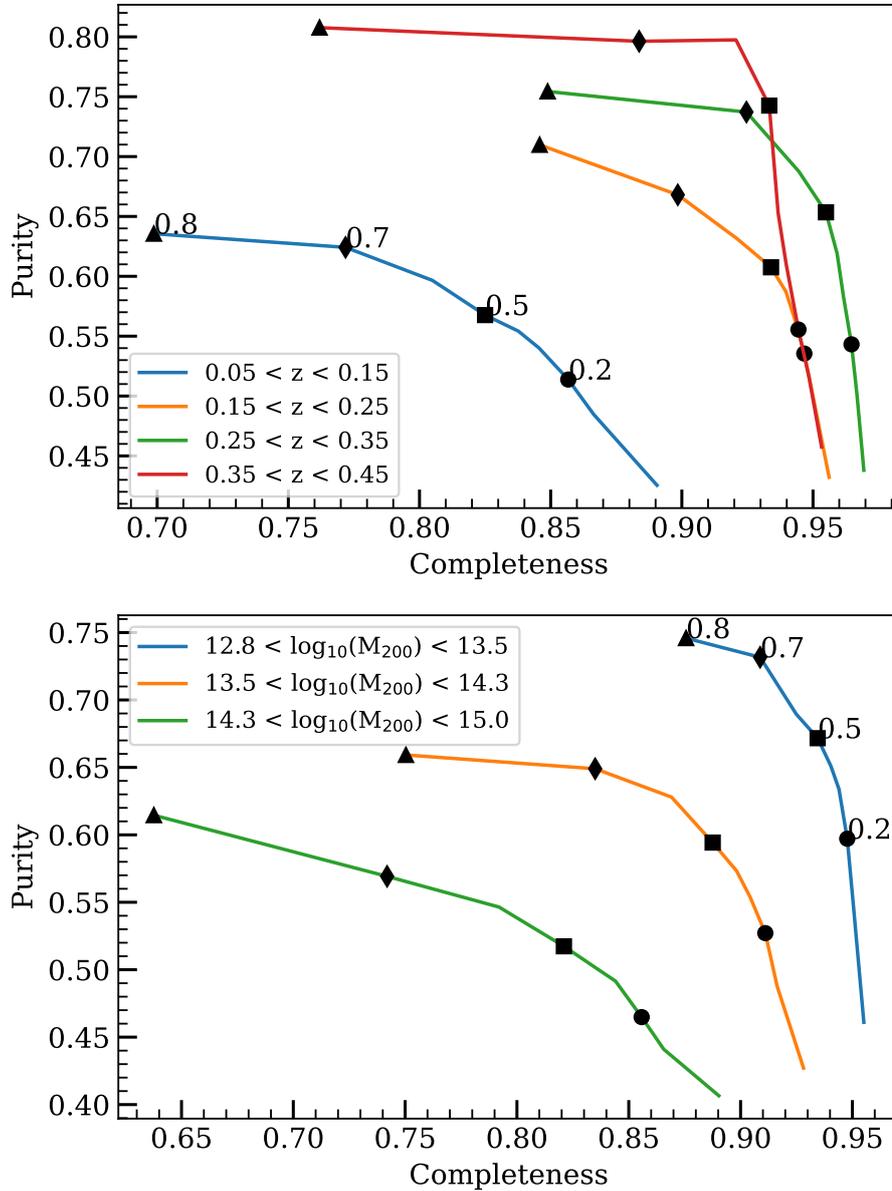


Figura 4.8: Semelhante à Figura 4.7 para amostras distintas de aglomerados de galáxias. Painel superior: diferentes faixas de *redshift* com passos de $dz = 0.1$. Painel inferior: intervalos de massa com amplitude de $d\log_{10}(M) = 0.75$. Para ambos, destacamos os cortes $P_{mem} > 0.2, 0.5, 0.7$ e 0.8 como círculo, quadrado, losango e triângulo, respectivamente. Figura retirada de Doubrava et al. (2023).

Tabela 4.4 - Valores de deslocamento de centro e *redshift* entre aglomerados da simulação/PZWav e os obtidos por galáxias aceitas como membro ($P_{mem} > 0$) em torno de seu centro.

	ΔR [kpc]	Δz
$O_{mock}-O_{gals}$	63.98 ± 34.76	-0.0006 ± 0.0049
$O_{PZWav}-O_{gals}$	128.38 ± 53.61	0.0013 ± 0.0061

nos cálculos. Para as estimativas de centro, seguimos o procedimento proposto por Castignani e Benoist (2016), onde os autores estimam o baricentro dos aglomerados provenientes de simulações pela média das coordenadas cartesianas das galáxias membros. Para isso, também introduzimos o peso das probabilidades P_{mem} . Então, para a ascensão reta, $\alpha_{gals} = \sum \alpha_i P_{mem,i} / \sum P_{mem,i}$, e similarmente para declinação.

Na Tabela 4.4, mostramos como $O_{mock}-O_{gals}$ a diferença entre a quantidade O (ΔR ou Δz) para o catálogo simulado e o valor médio obtido pelas galáxias aceitas como membro ($P_{mem} > 0$) em torno de seu centro e, da mesma forma, $O_{PZWav}-O_{gals}$ é a diferença entre a quantidade O para o catálogo PZWav e as galáxias selecionadas em torno dos centros obtidos pelo PZWav. Aqui também calculamos a média de todos os aglomerados uma vez que não encontramos tendências significativas na faixa de *redshift*.

Para ambos os cenários, as estimativas de *redshift* são altamente coerentes entre galáxias membros verdadeiras e selecionadas, dando uma diferença total consistente com zero, com $\sigma_{MAD} \leq 10^{-3}$. Este resultado destaca a abordagem computacional aplicada para derivar as estimativas de *redshift*. Tanto o PZWav quanto as simulações calculam z a partir da média das galáxias próximas.

Os centros, no entanto, apresentam discrepâncias substanciais, que podem ser explicadas de forma semelhante. As estimativas do centro nos aglomerados simulados são baseadas nos membros verdadeiros atribuídos independentemente da distância do centro da galáxia. Em nosso método, no entanto, as galáxias são selecionadas dentro de um raio de corte R_c , o que leva à remoção de alguns membros verdadeiros distantes da análise. Também destacamos a presença de galáxias contaminantes que podem deslocar o centro dependendo de seu P_{mem} . Ainda assim, o erro médio é de 64 ± 35 kpc. Para PZWav, os centros são estimados por sobredensidades, calculadas com base na distribuição de galáxias ponderada pela PDF integrada. Essas sobredensidades podem ter tamanhos físicos entre 400 e 1500 kpc. Então, a estimativa do centro considera a contribuição de todas as galáxias dentro do intervalo. Essa configuração retorna um erro médio de 128 ± 54 kpc.

Ressaltamos que esses valores podem ser dependentes do *redshift* ao considerar tamanhos angulares, nos quais, podemos subestimar os deslocamentos de aglomerados de baixo *redshift* e superestimar aglomerados de alto z . Nesse caso, encontramos uma variação de 1.3 minutos de arco em $z < 0.1$ que cai rapidamente para 0.6 minutos de arco em $z < 0.2$ e se estabiliza em 0.45 minutos de arco entre $0.3 < z < 0.45$.

Resultados com dados do miniJPAS

Neste capítulo, apresentamos os resultados derivados dos dados do levantamento miniJPAS. Os resultados aqui apresentados foram publicados em Doubrawa et al. (2024). Iniciamos nossa análise com as propriedades do catálogo de aglomerados de galáxias, obtido por meio do algoritmo PZWav, investigando o impacto da aplicação de diferentes limites de magnitude na caracterização da amostra. Além disso, descrevemos um método de seleção com base na riqueza das estruturas para aprimorar a pureza do catálogo.

Ao longo deste capítulo, exploramos e comparamos o catálogo de aglomerados e grupos com outros catálogos previamente apresentados na literatura. Direcionamos nossa atenção para a região sobreposta ao levantamento de raios-X AEGIS, focando na identificação de contrapartes ópticas. Este processo nos permitiu construir um novo catálogo de raios-X, utilizando nossas estimativas de centro. A aplicação da relação de escala proposta por Leauthaud et al. (2010) nos permitiu derivar as massas das estruturas e criar relações de escala com os traçadores ópticos fornecidos pelo AME.

Adicionalmente, comparamos os catálogos ópticos gerados pelos algoritmos VT e AMICO, investigando variações nas estimativas de centralização e *redshift*. Destacamos aglomerados de galáxias que apresentam contrapartes nos três localizadores de aglomerados. Finalmente, avaliamos a completude de cada catálogo óptico, utilizando os dados da galáxia como um traçador confiável. Essa abordagem proporcionou uma compreensão mais abrangente e precisa das propriedades dos aglomerados de galáxias analisados.

Tabela 5.1 - Colunas e descrição do catálogo de detecção

ID	Número de identificação do aglomerado
RA	Posição no céu em R.A.
DEC	Posição no céu em Dec.
z	<i>redshift</i>
SNR	Sinal-ruído da detecção
λ -19	Riqueza ($M < -19.5$)
λ _err-19	Incerteza da riqueza ($M < -19.5$)
Lum-19	Luminosidade óptica ($M < -19.5$)
Lum_err-19	Incerteza na Luminosidade óptica ($M < -19.5$)
Mstar-19	Massa estelar ($M < -19.5$)
Mstar_err-19	Incerteza na massa estelar ($M < -19.5$)
λ -21	Riqueza ($M < -21.25$)
λ _err-21	Incerteza da riqueza ($M < -21.25$)
Lum-21	Luminosidade óptica ($M < -21.25$)
Lum_err-21	Incerteza na Luminosidade óptica ($M < -21.25$)
Mstar-21	Massa estelar ($M < -21.25$)
Mstar_err-21	Incerteza na massa estelar ($M < -21.25$)

5.1 Catálogo de detecções

Ao aplicarmos o algoritmo PZWav na área do levantamento miniJPAS, produzimos um catálogo com 574 candidatos a grupos e aglomerados de galáxias dentro da faixa de *redshift* de $0.05 < z < 0.8$, com um SNR acima de 4. O catálogo oferece informações cruciais, incluindo número de identificação, coordenadas celestes, *redshift* fotométrico e SNR detectado para cada estrutura. Propriedades adicionais, como riqueza, massa estelar total e luminosidade óptica, são fornecidas para duas faixas diferentes de magnitude absoluta (como discutido na Seção 5.2). A Tabela 5.1 resume as colunas disponíveis no catálogo, juntamente com suas descrições correspondentes.

A Figura 5.1 apresenta as distribuições de *redshift* (painel superior) e SNR (painel inferior) do catálogo em forma de histograma. Destacamos duas amostras distintas: o catálogo completo, representado pelas barras em tons claros, que inclui todos os aglomerados detectados com $\text{SNR} > 4$, e a amostra mais pura (discutida na Seção 5.3), indicada pelas barras vermelhas.

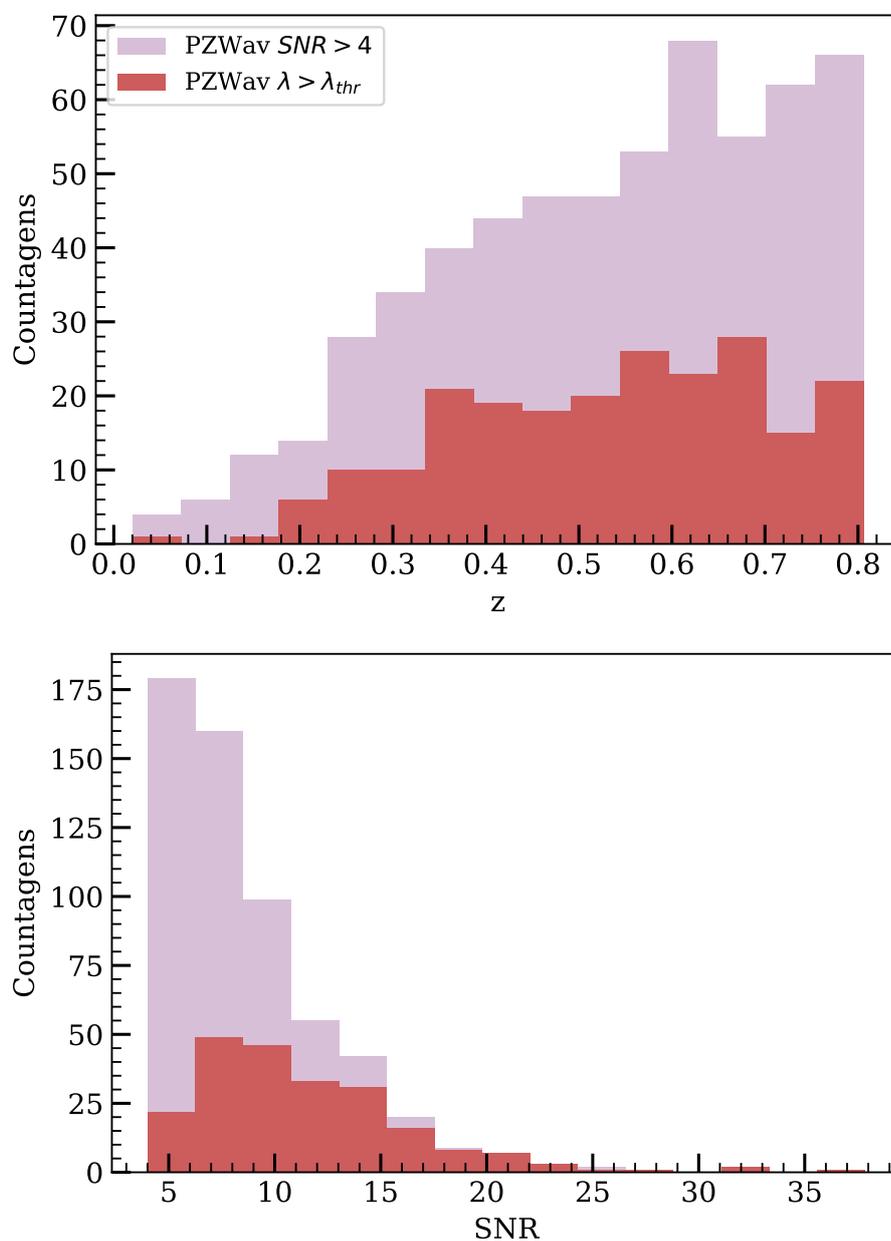


Figura 5.1: Distribuições de propriedades do catálogo de aglomerados gerado pelo PZWav. Painel superior: *Redshift* dos candidatos detectados. Painel inferior: Distribuição de SNR. Ambos os histogramas mostram a amostra antes (tons claros) e depois (vermelhas) da seleção em riqueza. Aplicamos λ_{thr} de 1.9 e 3.0 para os regimes de aglomerado e grupo (conforme discutido na Seção 5.2).

Tabela 5.2 - Colunas e descrição do catálogo de membros.

ID	Número de identificação
RA	Posição no céu em RA.
DEC	Posição no céu em DEC
photo_z	<i>redshift</i> fotométrico
PDF	PDF do <i>redshift</i> fotométrico
MAG_AUTO	Magnitude aparente na banda <i>r</i>
clusterID	ID do aglomerado correspondente
Pmem-19	Probabilidade de pertencimento para $M < -19$
Pmem-21	Probabilidade de pertencimento para $M < -21$

5.1.1 Catálogo de pertinência das galáxias membro

Compreender a composição de membros dos aglomerados de galáxias é crucial para uma análise abrangente de suas propriedades e dinâmicas. Nesta seção, apresentamos o catálogo de membros produzido pelo AME (Seção 3.5), um componente fundamental de nosso estudo, que nos permite identificar as galáxias que pertencem às estruturas detectadas.

O catálogo compreende 9 070 membros de aglomerados, representando 82% do catálogo total de galáxias disponível. A Tabela 5.2 fornece uma visão geral das informações-chave incluídas no catálogo de membros. Este conjunto de dados contém informações essenciais, como posições no céu, *redshift* fotométrico, PDF de foto-z e probabilidade de pertencimento. P_{mem} é listado para dois limites de magnitude absoluta, que serão discutidos em mais detalhes na próxima seção (Seção 5.2).

Além do catálogo de galáxias obtido pelo algoritmo PZWav, estendemos nossa análise para incluir catálogos de membros para os diversos catálogos de aglomerados e grupos investigados ao longo deste estudo.

5.2 Regimes de grupo e aglomerado

Para evitar erros sistemáticos nas estimativas de riqueza, luminosidade total e massa estelar, decorrentes da possível incompletude do catálogo de galáxias devido à faixa de *redshift*, optamos por trabalhar em dois regimes distintos:

- **Regime de aglomerados**, no qual consideramos galáxias com magnitude absoluta menores que $M_r < -21.25$, garantindo uma amostra limitada pelo volume até *redshift*

$z < 0.8$.

- **Regime de grupos**, com um limite de magnitude absoluta de $M_r < -19.5$, válido para $z < 0.3$.

Considerando suas diferentes propriedades de luminosidade, esses regimes nos permitem estudar estruturas com focos distintos, beneficiando estruturas tipo grupo de galáxias com um corte mais brilhante.

No painel superior da Figura 5.2, apresentamos as estimativas de riqueza para as estruturas concentradas na faixa de *redshift* de $0.05 < z < 0.3$. Comparamos as estimativas de riqueza obtidas com os dois cortes de magnitude absoluta, $M_r < -21.25$ (regime de aglomerados) e $M_r < -19.5$ (regime de grupos). Os diamantes pretos no painel superior mostram os valores medianos de riqueza, com barras de erro representando a dispersão da amostra. A figura destaca os valores mais altos na estimativa de riqueza ao considerar magnitudes mais fracas. Por exemplo, sistemas que são ignorados com uma riqueza de $\lambda = 0$ usando o corte de magnitude brilhante aumentam para $\lambda = 4.2$ quando galáxias mais fracas são incluídas.

O painel inferior da figura concentra-se na relação sinal-ruído (SNR) fornecida pelo PZWav para os candidatos selecionados. Ao analisar as estruturas detectadas no regime de aglomerados, algumas apresentam um valor de riqueza de $\lambda = 0$ e ainda mostram um alto SNR. Isso indica que esses candidatos têm uma alta amplitude de densidade em comparação com o ruído de fundo e não foram adequadamente descritos pelo corte de magnitude absoluta selecionado. No entanto, quando aplicamos o regime de grupos, os valores de riqueza são estimados de forma coerente, permitindo uma caracterização mais precisa.

Este teste destaca a importância de considerar diferentes faixas de magnitude, dependendo do nível de significância desejado.

5.3 Refinando o catálogo de detecções com cortes em riqueza

Um dos desafios ao detectar aglomerados de galáxias é avaliar a contaminação e a completude do catálogo resultante. Diferentemente da Seção 2.1, agora não sabemos a real distribuição das estruturas. Assim, a presença de falsas estruturas pode introduzir vieses

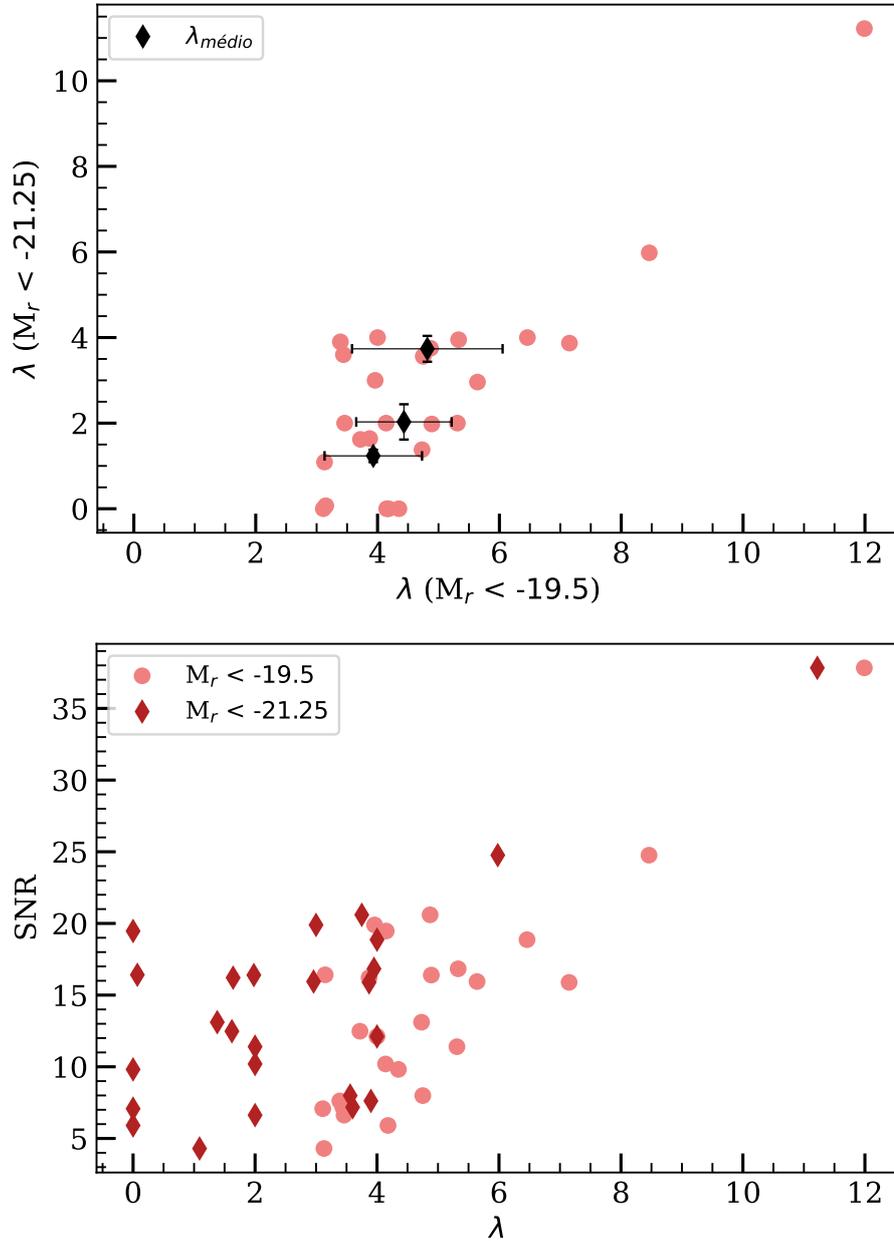


Figura 5.2: Distribuição das riquezas em diferentes cortes de magnitude absoluta. Painel superior: Comparação entre riquezas no regime de grupos ($M_r < -19.5$) para estruturas com $z < 0.3$, e no regime de aglomerados ($M_r < -21.25$). Painel inferior: Relação entre o SNR do PZWav e a riqueza. Alguns grupos que passaram despercebidos no regime de aglomerados podem apresentar uma riqueza significativa em cortes mais rasos de magnitude.

nas medições da abundância de aglomerados, afetando nossa compreensão dos parâmetros cosmológicos.

Estudos usando as simulações apresentadas na Seção 2.1 realizado por Werner et al. (2022), indicam que utilizar um $\text{SNR} > 3.3$ no catálogo produzido pelo PZWav alcança os melhores valores de completude e pureza (assim como discutido em Seção 2.3). Em contraste, outras abordagens, como a proposta por Rykoff et al. (2014), adotam um método de seleção diferente, usando um limite de riqueza de 20 (calculado pela soma das probabilidades de pertinência) para restringir a amostra para aglomerados de massa maior que $10^{14} M_{\odot}$.

Neste contexto, apresentamos uma abordagem alternativa para lidar com a presença de falsos positivos no catálogo. Seguindo uma metodologia semelhante à descrita por Klein et al. (2017a), desenvolvemos critérios de seleção com base na riqueza que nos permite remover aglomerados e grupos abaixo de um determinado limite.

O método envolve a comparação das estruturas do catálogo PZWav com pontos distribuídos aleatoriamente em posições no céu e *redshifts*. Assim como na Seção 4.2 esperamos encontrar pequenos valores de riqueza, uma vez que essas posições no céu não são privilegiadas. Entretanto, em alguns casos, essa distribuição pode atingir uma sobredensidade de galáxias e fornecer uma medida significativa. Para cada candidato detectado pelo algoritmo, avaliamos o número de pontos aleatórios dentro de um intervalo específico de *redshift* definido por $|z_{opt,i} - z_{rand}| < 0.05$. Em seguida, calculamos a fração desses pontos que apresentam um valor de riqueza menor do que a riqueza do candidato em questão. Esse processo fornece uma medida quantitativa da significância da riqueza calculada em comparação com os pontos distribuídos aleatoriamente. Ao definir um limite de significância da detecção de 90%, identificamos e removemos candidatos a aglomerados e grupos com potencial detecção por acaso, considerando o valor de riqueza associada.

Para reduzir a possibilidade de identificar erroneamente um grupo de galáxias como contaminação devido ao corte de magnitude (como discutido na Seção 5.2), repetimos a análise nos dois regimes de magnitude, obtendo dois valores distintos de riqueza. Dessa forma, para garantir um nível de significância de 90%, estabelecemos limites de riqueza de 1.9 e 3.0 para os regimes de aglomerados e grupos, respectivamente. Como resultado, o conteúdo final do catálogo foi reduzido para apenas 38% da amostra inicial de detecções. O impacto desse corte é evidenciado nas distribuições alteradas de *redshift* e SNR, conforme

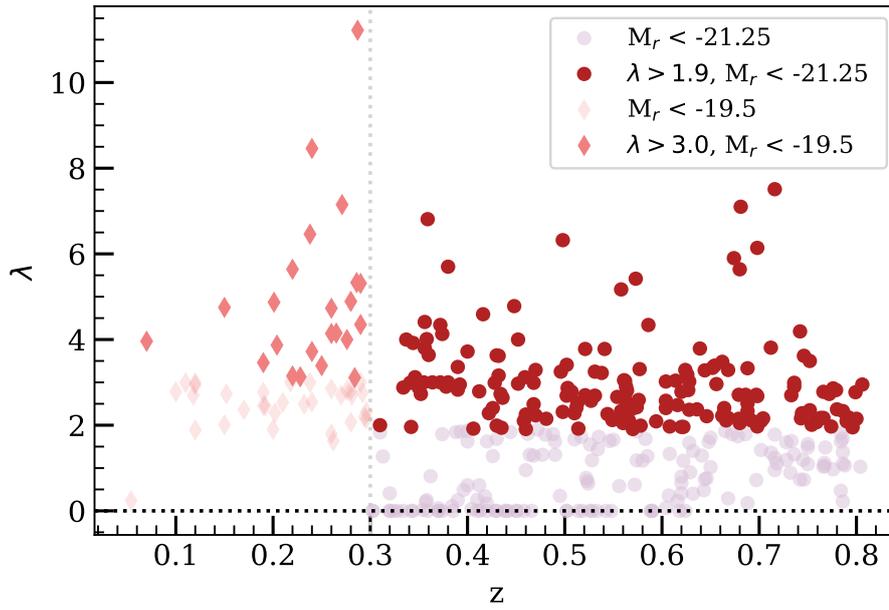


Figura 5.3: Distribuição de riqueza do catálogo detectado pelo algoritmo PZWav para os regimes de grupo (rosa) e aglomerado (vermelho escuro) - antes (tons mais claros) e depois (tons mais escuros) do limiar de riqueza. A linha tracejada indica a transição de *redshift* em $z = 0.3$.

ilustrado pelas barras vermelhas na Figura 5.1.

Na Figura 5.3, apresentamos a distribuição de riqueza em função do *redshift* para os regimes de grupo (rosa) e aglomerado (vermelho escuro). As cores mais claras representam a distribuição antes da aplicação do limite de riqueza, enquanto as mais escuras simbolizam a distribuição após sua aplicação. Dentro das estruturas identificadas, destaca-se um aglomerado específico, mJPC2470-1771 (Figura 1.1), que exibe um alto valor de riqueza de 11.5 ($M_r < -19.5$) em *redshift* de $z = 0.29$, e que foi objeto de estudo por Rodríguez-Martín et al. (2022).

Por fim, após a seleção de riqueza, as propriedades ópticas do nosso catálogo variam de $1.9 < \lambda < 12$ para a riqueza, $10.6 < \log_{10}(L_\lambda/L_\odot) < 11.8$ para a luminosidade óptica e $10.6 < \log_{10}(M_\lambda^*/M_\odot) < 12.2$ para as massas estelares.

5.4 Comparações com o catálogo de raios-X do levantamento AEGIS

Um aspecto interessante sobre o levantamento miniJPAS é sua intersecção com outros catálogos, como por exemplo, o levantamento AEGIS (Davis et al., 2007), o que permite uma comparação entre catálogos ópticos e de raios-X. O catálogo de raios-X, criado por Erfanianfar et al. (2013) usando dados dos telescópios Chandra e XMM-Newton, é com-

posto por 52 aglomerados no intervalo de *redshift* de $0.06 < z < 1.54$. Esses aglomerados e grupos apresentam uma faixa de massa entre $M_{200} \sim 1.34 \times 10^{13}$ e $1.33 \times 10^{14} M_{\odot}$. Quando impomos limites para centralizar nossa análise na área específica do miniJPAS e na faixa desejada de *redshift* ($0.05 < z < 0.8$), essa seleção reduz a amostra para 36 aglomerados.

Para estabelecer uma correspondência entre os catálogos óptico (PZWav) e de raios-X, realizamos um processo com base nas distâncias mais próximas entre os centros e no *redshift*. Aplicamos uma diferença máxima no centro de 0.5 Mpc e no *redshift* de $\Delta z = 0.05$.

Esse procedimento resultou em 17 correspondências. Dessas, 7 detecções estão localizadas em baixo *redshift* ($z < 0.55$), enquanto 10 estão dentro da faixa de *redshift* mais elevada de $0.55 > z > 0.8$. Os limites de *redshift* foram escolhidos de forma a reproduzir à análise feita por Maturi et al. (2023) na mesma área. Na Figura 5.4, apresentamos a distribuição de massa em função do *redshift* para os grupos de raios-X representados como círculos abertos.

Depois de identificarmos as correspondências, podemos calcular as diferenças na centralização e *redshift* para a subamostra correspondente. Essas diferenças têm um valor significativo, especialmente em análises de lentes gravitacionais, pois podem introduzir vieses que impactam nos resultados finais (Parroni et al., 2017; Sommer et al., 2022). No entanto, modelar com precisão os deslocamentos de centralização sem uma grande amostra pode ser desafiador. Portanto, este estudo se concentra em observar possíveis tendências em vez de fornecer um modelo dos efeitos de deslocamento.

Em geral, as variações medianas encontradas nas posições dos centros entre os catálogos são de 125 ± 60 kpc. Em relação aos *redshifts*, 90% das detecções correspondentes têm desvios dentro de 0.02 para ambas as faixas analisadas. Os valores medianos são semelhantes nas comparações, 0.001 ± 0.005 para toda a faixa de *redshift*. Esses resultados sugerem que os critérios de correspondência escolhidos são apropriados, e aumentar os limites não melhoraria significativamente o número de correspondências bem-sucedidas.

5.5 Catálogo de raios-X estendido

Para aprimorar a comparação, reanalisamos o mesmo mosaico observado pelo telescópio Chandra do campo AEGIS, implementando um limite de brilho superficial mais baixo, de 3σ , em comparação com 4σ em Erfanianfar et al. (2013). Essa nova análise, conduzida

Tabela 5.3 - Catálogo de raios-X estendido baseado nos centros ópticos do algoritmo PZWav. Estamos apresentando as primeiras e as últimas 4 linhas ordenadas por M_{200c} . L_x e M_{200c} são dados em unidades de $\text{erg cm}^{-2} \text{s}^{-1}$ e M_\odot respectivamente.

ID	RA	DEC	z	L_x	e L_x	M_{200c}	e M_{200c}
16	213.80150	52.09198	0.070	2.15×10^{41}	1.64×10^{41}	7.22×10^{12}	3.16×10^{12}
11	214.98826	53.11720	0.201	3.85×10^{41}	1.27×10^{41}	9.39×10^{12}	1.88×10^{12}
20	214.33039	52.59097	0.238	5.20×10^{41}	1.27×10^{41}	1.10×10^{13}	1.66×10^{12}
153	213.89208	52.19922	0.432	1.01×10^{42}	6.74×10^{41}	1.41×10^{13}	5.44×10^{12}
23	215.14407	53.13991	0.736	1.54×10^{43}	1.81×10^{42}	5.98×10^{13}	4.41×10^{12}
550	215.37229	53.46197	0.680	1.67×10^{43}	4.74×10^{42}	6.66×10^{13}	1.15×10^{13}
140	214.97817	53.22724	0.406	1.95×10^{43}	6.09×10^{42}	9.61×10^{13}	1.82×10^{13}
1	213.61502	51.93392	0.287	$2.1 \times 10^{+43}$	2.16×10^{42}	1.14×10^{14}	7.21×10^{12}

pelo prof. Dr. Alexis Finoguenov, aproveita a PSF (do inglês *point spread function*) do Chandra, que remove fontes pontuais contaminantes em uma ordem de grandeza de fluxo mais baixo. Além disso, adicionamos ao mosaico as observações do XMM-Newton que se sobrepõem ao campo do miniJPAS. Isso contribui não apenas para uma melhoria na subtração das estruturas em larga escala presentes no campo, mas também prova ser valioso para identificar fontes recém-descobertas. Dado nosso foco na amostra encontrada tanto na faixa do óptico quanto em raios-X, estabelecemos um procedimento sistemático para identificação de fontes de raios-X, garantindo a consistência e a replicabilidade de nossas análises.

Inicialmente, realizamos uma correspondência posicional entre as fontes ópticas do PZWav e as de raios-X, permitindo um deslocamento máximo entre centros de 0.5 Mpc. Algumas fontes de raios-X têm mais de uma correspondência óptica. Em tais casos de múltiplas correspondências, empregamos uma abordagem para avaliar a significância das correspondências. Grupos ópticos determinados como associações casuais com fontes de raios-X são removidos da análise. Aqui utilizamos o mesmo método aplicado para refinar o catálogo (Seção 5.3) descrito em Klein et al. (2017b).

Se uma fonte de raios-X ainda tiver várias identificações após esta etapa, ela é excluída da análise das relações de escala, a menos que uma correspondência apresente uma riqueza significativamente maior (por um fator superior a 1.5). Esse critério garante que a contaminação de outro grupo no fluxo de raios-X seja limitada a menos de 30%, uma incerteza

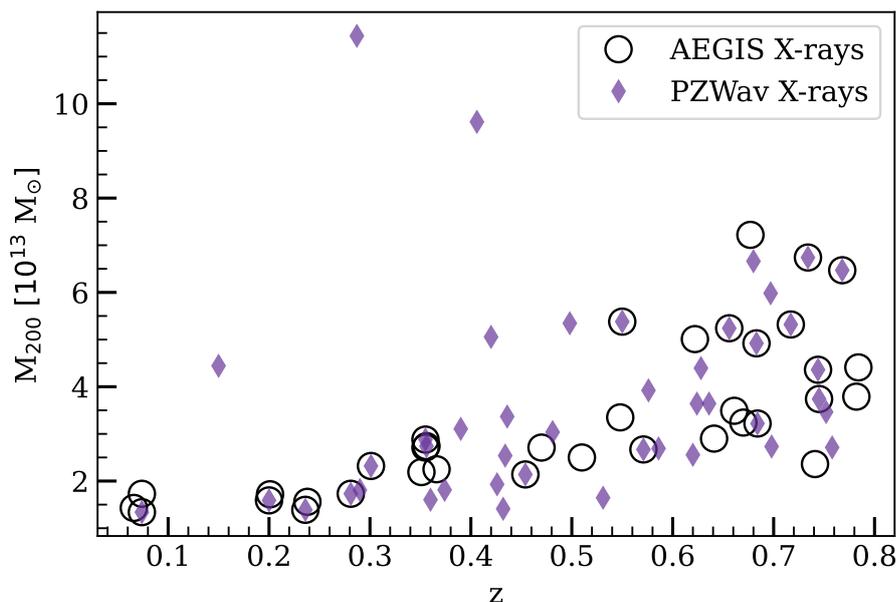


Figura 5.4: Distribuições de massa e *redshift* das detecções ópticas do PZWav que possuem uma contraparte em raios-X, representadas por diamantes roxos. Círculos pretos abertos exibem o catálogo de detecção de Erfanianfar para a área de sobreposição. No total, somam 17 grupos e aglomerados de galáxias com distância central correspondente menor que 500 kpc e diferença de *redshift* de $\Delta z < 0.05$. As massas são inferidas a partir da calibração de lentes fracas dadas por Leauthaud et al. (2010).

inferior ao erro estatístico típico associado às fontes recém-adicionadas.

Para o catálogo de fontes de raios-X identificadas de forma única, calculamos a luminosidade de raios-X seguindo os mesmos procedimentos apresentados em Erfanianfar et al. (2013) para correção do fluxo de abertura total e correção k . Por fim, a massa total foi inferida pela calibração de lentes fracas de Leauthaud et al. (2010).

Esses passos resultaram em um catálogo de 37 fontes de raios-X, incluindo 20 detecções dentro da faixa de *redshift* de $0.05 < z < 0.55$ e 17 detecções dentro da faixa de $0.55 < z < 0.8$. Dessas, 20 são novas detecções. Relatamos as propriedades das correspondências de raios-X utilizadas nesta análise na Tabela 5.3.

Na Figura 5.4, apresentamos os grupos de raios-X identificados no catálogo de Erfanianfar et al. (2013) (círculos abertos) juntamente com suas contrapartes correspondentes da amostra reanalisada pelo catálogo óptico PZWav (diamantes roxos). Os círculos abertos com um símbolo de diamante no centro representam as 17 correspondências no catálogo de Erfanianfar et al. (2013).

O sistema mais massivo, mJPC 2470-1771 (Figura 1.1), situado em $z = 0.29$ e exibindo a maior riqueza, possui uma massa de $M_{200} = (1.14 \pm 0.07) \times 10^{14} M_{\odot}$. Estimativas

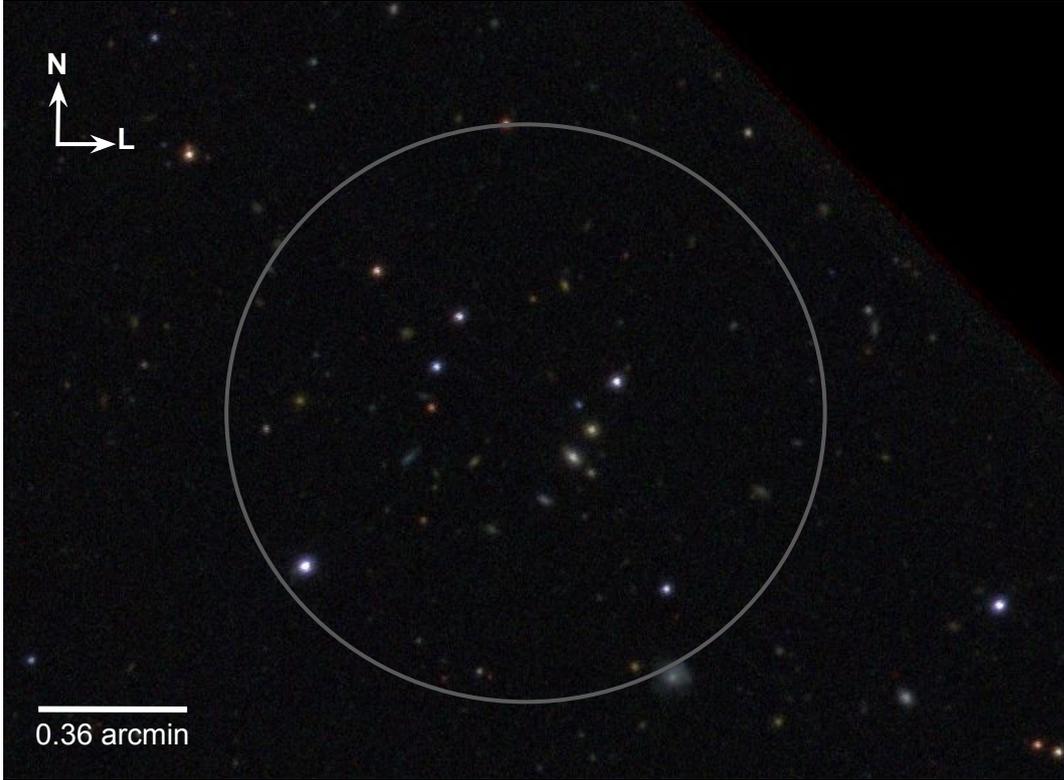


Figura 5.5: O segundo aglomerado mais massivo do catálogo estendido de raios-X, em $z = 0.41$. O círculo em cinza, de raio de 1.465 minutos de arco (R_c), está centralizado no pico de densidade encontrado pelo algoritmo PZWav.

anteriores da massa baseadas exclusivamente em dados de raios-X, apresentadas por Bonoli et al. (2021), resultaram em $M_{200} = (3.26 \pm 1.4) \times 10^{14} M_{\odot}$, um valor um pouco superior, mas com uma incerteza considerável ($\sim 50\%$). Utilizando observações de acompanhamento com o espectrógrafo Gemini GMOS, os mesmos autores obtiveram $M_{200} = (2.2 \pm 0.3) \times 10^{14} M_{\odot}$.

Encontramos um segundo candidato interessante em $z = 0.41$ (Figura 5.5), próximo às bordas do levantamento, com uma massa estimada de $9.6 \times 10^{13} M_{\odot}$. Esse objeto já havia sido observado por Hsieh et al. (2005) no âmbito do *Red-Sequence Cluster Survey* (Gladders e Yee, 2005), apresentando um *redshift* fotométrico semelhante ($z = 0.4$) e estando dentro de uma distância de 10 segundos de arco. Até o momento, não encontramos uma discussão dedicada a esse objeto na literatura, tornando-o um forte candidato para futuros estudos.

Os sistemas de menor massa variam de (0.72 ± 0.31) a $(1.60 \pm 0.24) \times 10^{13} M_{\odot}$ na faixa de *redshift* de $0.07 < z < 0.43$. Adicionalmente, em $z = 0.745$, detectamos um sistema com uma massa de $M_{200} = (3.75 \pm 0.5) \times 10^{13} M_{\odot}$ (também reportado por Hsieh et al., 2005, em um *redshift* similar e dentro de 10 segundos de arco.).

É importante destacar que a detecção bem-sucedida de grupos de baixa massa em nossa análise evidencia a profundidade da pesquisa miniJPAS e valida a eficácia de nossa metodologia, conforme também confirmado por Maturi et al. (2023).

5.6 Relações de escala

Usando LINMIX, um procedimento de regressão linear com abordagem Bayesiana (Kelly, 2007), somos capazes de realizar o processo de minimização considerando os erros tanto nas estimativas de massa em raios-X quanto nos traçadores de massa, assim como destacado em Seção 4.4. As riquezas utilizadas para essa análise considera o corte em magnitude absoluta de -21.25 .

Os parâmetros de melhor ajuste estão apresentados na Tabela 5.4, e os resultados estão resumidos na Figura 5.6. Com nossa amostra, não conseguimos identificar uma correlação forte entre as variáveis. Ao comparar os traçadores de massa, observa-se que a riqueza apresenta a maior dispersão para a inclinação. Esse comportamento é ainda mais acentuado pela faixa limitada de valores de riqueza na amostra, bem como pela presença de dois sistemas massivos com baixa riqueza que se destacam da distribuição geral. Em contraste, a luminosidade óptica apresenta uma inclinação relativamente íngreme, embora ainda dentro das barras de erro.

Destacamos que o tamanho de nossa amostra é relativamente pequena, o que compromete a robustez de nossos resultados estatísticos. Esse fato se torna evidente nos pequenos erros de espalhamento intrínseco. Aqui os erros nas variáveis são tão grandes que o valor de ϵ diminui. No entanto, com a continuação da levantamento J-PAS, esperamos obter um aumento no tamanho da amostra, proporcionando relações de escala mais sólidas e confiáveis.

Ainda assim, a melhora das relações de escala não depende apenas de uma maior quantidade de objetos, mas também das estimativas de massa. Os dados de raios-X profundos disponíveis para o miniJPAS nos permitem derivar massas na extremidade inferior do espectro, alcançando um limite de fluxo de 10^{-15} $\text{erg s}^{-1} \text{cm}^{-2}$. Em comparação, outras pesquisas, como o levantamento completo do céu ROSAT e o eRosita (Merloni et al., 2012), possuem limites de fluxo de $10^{-13}/10^{-12}$ e 10^{-14} $\text{erg s}^{-1} \text{cm}^{-2}$, respectivamente.

Para superar essa limitação, Maturi et al. (2023) propuseram abordagens alternativas,

Tabela 5.4 - Valores de ajuste da regressão linear descrita pela Equação 4.4. L_λ e M_λ^* são dados em unidades de L_\odot e M_\odot para os dados do levantamento miniJPAS.

Proxy	α	β	ϵ	\bar{O}
λ	13.47 ± 0.04	0.24 ± 0.51	0.039 ± 0.016	2.9
L_λ	13.48 ± 0.05	0.34 ± 0.18	0.076 ± 0.019	1.47×10^{11}
M_λ^*	13.48 ± 0.05	0.23 ± 0.18	0.080 ± 0.021	3×10^{11}

incluindo a reanálise dos dados arquivados do XMM-Newton sobre a área do J-PAS ou a incorporação de medidas de lentes gravitacionais fracas e dispersão de velocidades. Esses métodos oferecem soluções potenciais para aprimorar a estimativa de massa na presença de limitações de fluxo.

Como as estimativas de massa também podem ser afetadas pelo estado dinâmico da estrutura, filtrar a amostra para incluir apenas aglomerados mais relaxados (aproximadamente virializados) poderia aprimorar as relações de escala.

5.7 Comparações entre catálogos ópticos

Diversos algoritmos computacionais podem ser empregados para identificar aglomerados de galáxias em dados ópticos, cada um com suas próprias características e capacidades únicas.

Nas seções seguintes, apresentamos resultados comparativos entre o catálogo de grupos e aglomerados de galáxias encontrados por PZWav, com outros catálogos produzidos por VT e AMICO (Seção 1.2.1). Para isso, utilizamos o mesmo conjunto de dados do miniJPAS, com o objetivo de compreender as diferenças entre as amostras detectadas por todos os métodos ou apenas por um método específico. Para avaliar tais variações, utilizamos o traçador AME, que nos permite analisar quantitativamente as discrepâncias entre os diferentes catálogos. É importante salientar que nosso objetivo não é determinar qual catálogo é o mais correto (puro e/ou completo), mas sim fornecer uma compreensão mais profunda sobre diferentes propriedades da amostra de aglomerados. Aqui utilizamos as riquezas medidas no regime de aglomerados.

Além disso, ressaltamos que Maturi et al. (2023) apresenta uma descrição detalhada sobre o catálogo AMICO aplicado na região do miniJPAS e seus resultados. Este catálogo inclui 94 grupos e aglomerados com uma relação sinal ruído superior a 2.5. Neste trabalho,

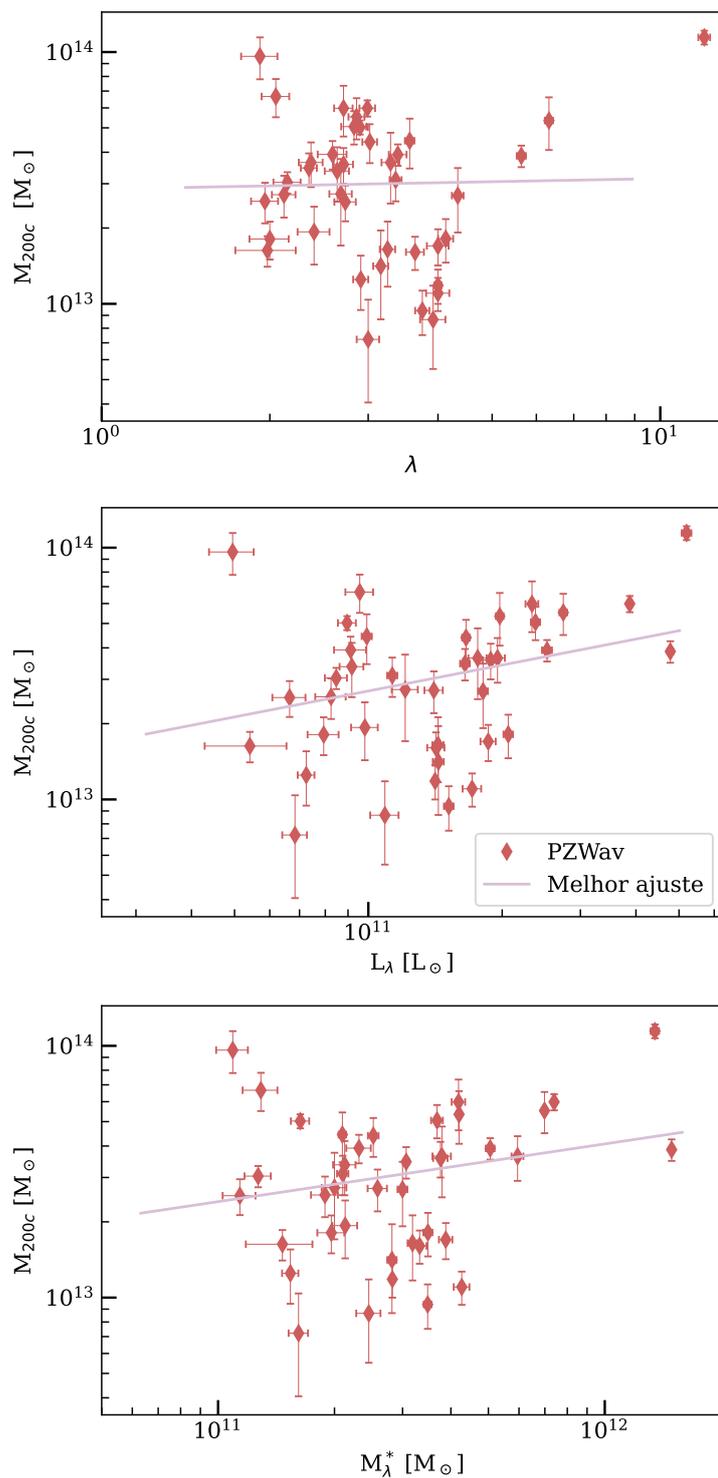


Figura 5.6: Relações de escala obtidas para o levantamento miniJPAS. Painel superior: M_{200c} e riqueza, λ . Painel do meio: M_{200c} vs luminosidade óptica, L_{λ} . Painel inferior: M_{200c} vs massa estelar, M_{λ}^* . As linhas representam os valores de melhor ajuste. Os parâmetros de melhor ajuste podem ser encontrados na Tabela 5.4.

os autores estabeleceram relações de escala para traçadores de massa, incluindo a amplitude A e estimativas de massa estelar. Ao construir um catálogo de membros de aglomerados por meio de associações probabilísticas, os autores demonstraram uma boa concordância com membros espectroscópicos para galáxias com $P > 0.2$ e identificaram as galáxias mais brilhantes dos grupos (BGGs) dentro da amostra. Utilizando AMICO, o estudo caracterizou grupos e aglomerados de galáxias, incluindo estruturas de pequeno porte ($\sim 10^{13}M_{\odot}/h$).

O catálogo utilizando o algoritmo VT foi produzido pelo Prof. Dr. Paulo A. A. Lopes (como parte da colaboração J-PAS). O catálogo é composto por 159 candidatos a aglomerados/grupos.

5.7.1 Correspondência entre os catálogos

Para a comparação entre os catálogos, mantemos os mesmos parâmetros de correspondência usados na comparação de raios-X: uma diferença máxima de 0.5 Mpc entre os centros e uma diferença máxima de *redshift* de $\Delta z = 0.05$. Através desse procedimento, identificamos um total de 43 aglomerados comuns aos três catálogos de identificação de aglomerados. As Figuras 5.7 e 5.8 fornecem uma comparação visual desses deslocamentos, mostrando a concordância entre as coordenadas dos centros e os *redshifts* em combinações de dois a dois. Nessas figuras, as barras vermelhas representam a comparação PZWav - VT, as barras azuis indicam AMICO - PZWav, e as barras verdes, VT - AMICO. A diferença da coloração em cada imagem representa diferentes faixas de *redshift*: baixo ($z \leq 0.55$) em tons mais claros; e alto ($z > 0.55$) mais escuros. Esse valor foi escolhido de forma a concordar com as análises realizadas em Maturi et al. (2023).

A Figura 5.7 demonstra que os centros dos aglomerados são bem caracterizados, com mais de 80% (35/43) dos aglomerados apresentando diferenças de centro inferiores a 300 kpc. Esses deslocamentos são menores para aglomerados de baixo *redshift*, com um valor médio de 92 ± 53 kpc para o caso PZWav - VT (31), em comparação com o valor médio de 112 ± 67 kpc para os aglomerados de alto *redshift* (12).

A Figura 5.8 mostra os resultados em termos de diferenças de *redshift*. Aqui, observamos que mais de 90% (40/43) das detecções apresentam uma variação inferior a ± 0.02 , aumentando para 95% (11/12) nos aglomerados de alto *redshift*. Os valores médios dessas diferenças de *redshift* são semelhantes entre os localizadores de aglomerados, com um valor

mediano de 0.001 ± 0.005 para toda a faixa de *redshift*.

Na Figura 5.9, analisamos a distribuição da relação sinal ruído do PZWav (SNR) em função da riqueza no regime de aglomerados. Lembramos que riqueza é definida como a soma das probabilidades de pertinência das galáxias. A distribuição revela uma correlação entre as quantidades. Ao aplicarmos o corte de riqueza, removendo principalmente estruturas com baixo SNR, observamos uma diminuição na densidade de pontos para valores menores de SNR. Esse comportamento também é evidente na forma modificada do histograma mostrado na Figura 5.1.

Observamos ainda que a maioria das correspondências possui um SNR superior a 6. As correspondências comuns entre os catálogos são indicadas por círculos abertos, enquanto os quadrados verdes e os diamantes azuis representam as correspondências 2 por 2 entre PZWav e AMICO, e VT, respectivamente. Essa análise mostra a sensibilidade do PZWav aos pequenos picos de densidade.

Ao aplicarmos o AME nos catálogos, podemos fornecer uma caracterização adicional das detecções usando os traçadores de massa e comparar os resultados com a amostra identificada por apenas um dos localizadores de aglomerados (Seção 5.7.2). Abaixo, focamos nos resultados de riqueza para calcular a fração de aglomerados correspondentes em relação ao número total de detecções.

Para o PZWav, 85% do catálogo apresenta similaridade com os outros para aglomerados com riqueza (λ) superior a 6. Esse percentual diminui para 30% para riqueza no intervalo de $3 < \lambda < 6$, e apenas 13% para λ inferior a 3. Tendências semelhantes são observadas para VT e AMICO, com 75%, 57%, 15% para VT e 100%, 62%, 34% para AMICO, respectivamente, nos mesmos intervalos de riqueza. Esses valores destacam a concordância entre os catálogos.

A importância de tais detecções em comparação com um catálogo de galáxias massivas é discutida mais detalhadamente na Seção 5.8.

5.7.2 Estatísticas dos traçadores de massa

Nesta subseção, exploramos as diferenças entre as propriedades de aglomerados e grupos de galáxias correspondentes e não correspondentes detectados por diferentes localizadores de aglomerados. A Tabela 5.5 exibe os valores médios e desvios de riqueza, luminosidade óptica e massa estelar para ambos os conjuntos no regime de aglomerados.

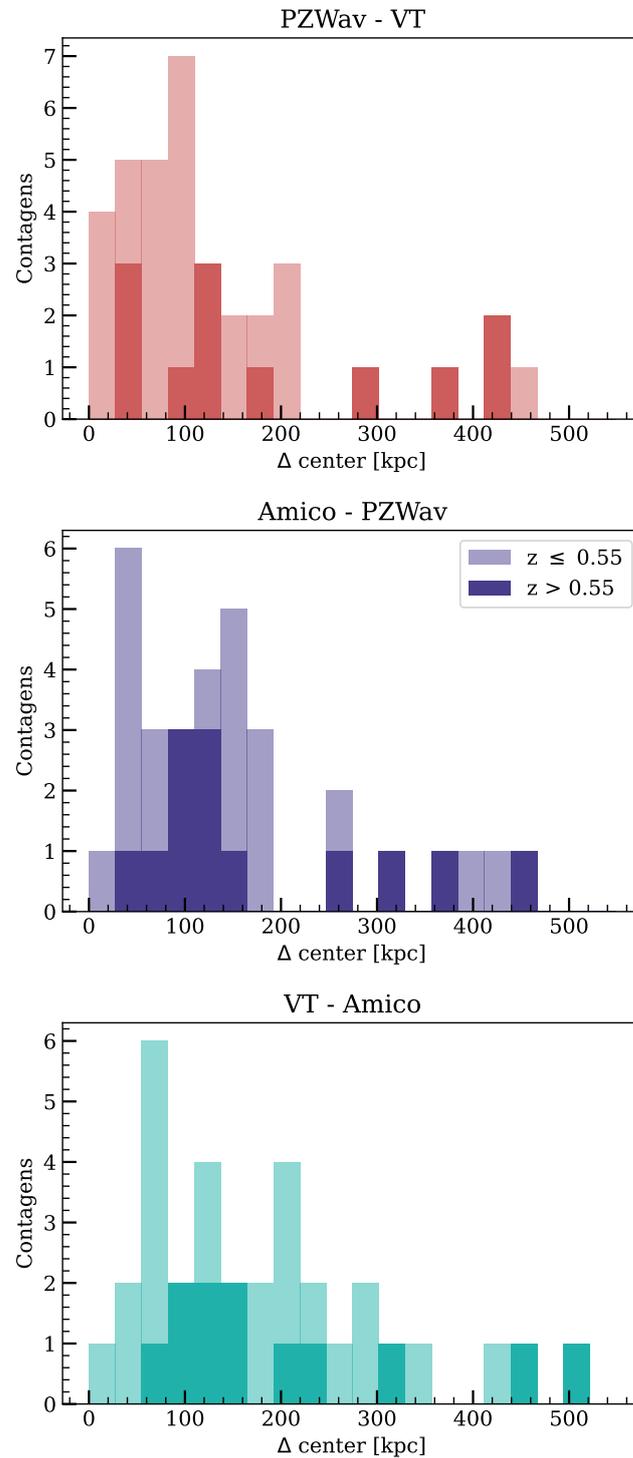


Figura 5.7: Diferença entre as coordenadas centrais dos aglomerados correspondentes identificados pelos três catálogos óticos, em uma comparação 2 por 2. As cores claras representam aglomerados com *redshifts* menores que $z \leq 0.55$, os escuros para $z > 0.55$. De cima para baixo: Vermelho: PZWav - VT; Azul: AMICO - PZWav; e Verde: VT - AMICO.

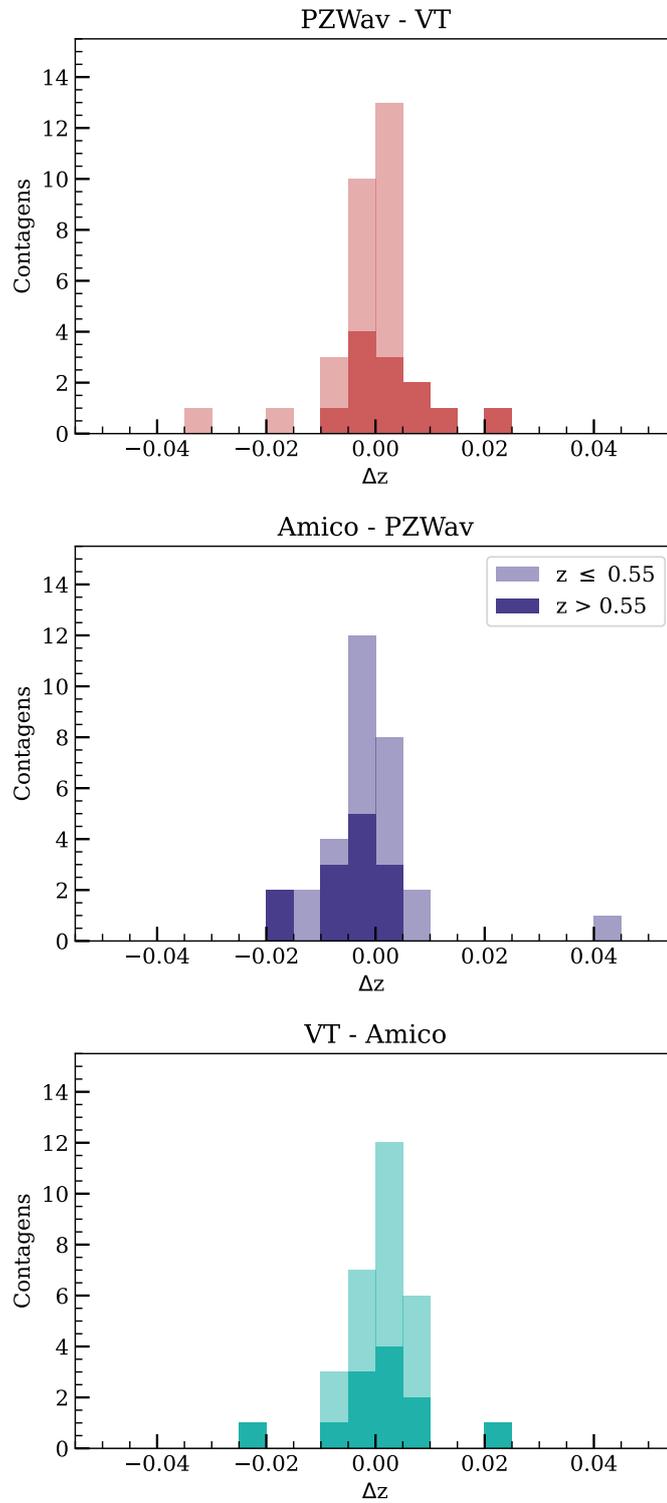


Figura 5.8: Igual à Figura 5.7 para a diferença entre *redshifts* das estruturas detectadas.

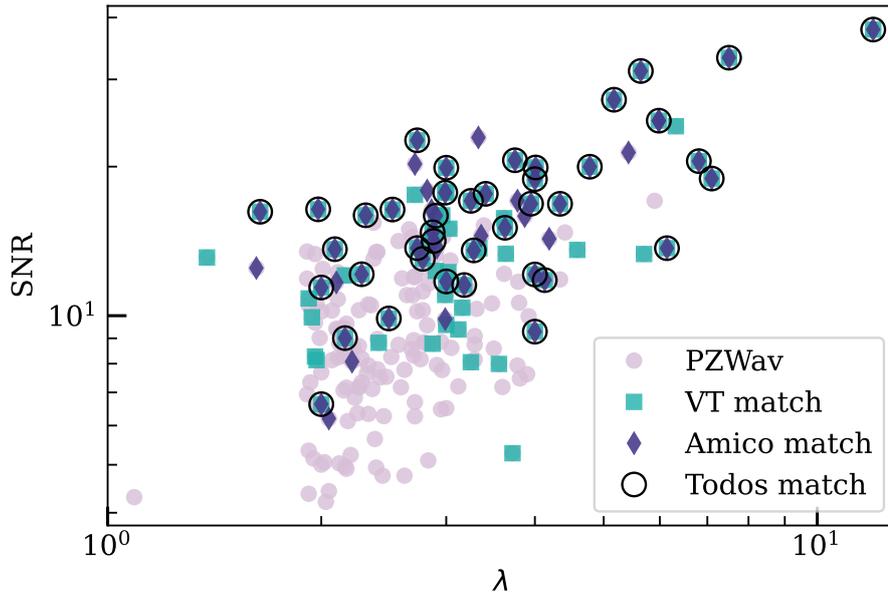


Figura 5.9: Distribuição da riqueza em função do SNR para os catálogos ópticos: PZWav em pontos cinza, correspondência entre PZWav - VT em verde, e PZWav - AMICO em azul. As correspondências entre os três catálogos são apresentadas por círculos abertos.

Para os aglomerados na amostra correspondente, observamos que os valores médios mostram semelhanças entre os diferentes localizadores, com variações mínimas dentro das barras de erro. Isso pode ser atribuído ao fato de que os aglomerados correspondentes têm posições de centro e *redshift* semelhantes, assim os traçadores são calculados e comparados dentro desses pequenos deslocamentos. Como vimos em Seção 4.5, apenas grandes variações de centro e *redshift* impactam significativamente a detecção de membros do aglomerado. Aqui, destacamos os valores encontrados para o PZWav: $\lambda = 3.18 \pm 0.06$, $L_\lambda = (1.65 \pm 0.07) \times 10^{11} L_\odot$ e $M_\lambda^* = (3.49 \pm 0.14) \times 10^{11} M_\odot$.

No entanto, diferenças substanciais surgem ao analisar a amostra não correspondente. Por exemplo, no caso do PZWav, a riqueza dos aglomerados não correspondentes (178) é consideravelmente menor, com um valor mediano de 2.58 ± 0.03 . Essa diferença é ainda mais pronunciada nos catálogos VT e AMICO, onde a diferença relativa excede 45%. Como o catálogo do PZWav é limitado a $\lambda > 1.9$, os deslocamentos gerais são menores. Tendências semelhantes podem ser observadas para os outros traçadores de massa.

Na Tabela 5.5 apresentamos os valores médios para toda a faixa de *redshift* para garantir clareza. No entanto, ao examinar a amostra separadamente para *redshifts* altos ($z > 0.55$) e baixos ($z < 0.55$), os deslocamentos se tornam mais significativos, com uma diferença de 75% para L_λ e uma diferença de 85% para M_λ^* para as estruturas sem correspondências.

Tabela 5.5 - Valores medianos dos traçadores ópticos para cada catálogo de aglomerados e grupos de galáxias, divididos entre aglomerados correspondentes (*Match*) e não correspondentes (Não-M), para riqueza, luminosidade óptica e massa estelar. M_λ é dado em unidades de $10^{11} M_\odot$ e L_λ em $10^{11} L_\odot$.

	nº of aglomerados Match / Não-M	λ	L_λ	M_λ^*
PZWav	43	3.18 ± 0.06	1.65 ± 0.07	3.49 ± 0.14
	178	2.58 ± 0.03	1.18 ± 0.02	2.13 ± 0.07
VT	43	3.22 ± 0.07	1.62 ± 0.07	3.24 ± 0.18
	116	1.73 ± 0.12	0.64 ± 0.06	1.12 ± 0.10
AMICO	43	3.00 ± 0.14	1.72 ± 0.10	3.64 ± 0.22
	51	1.87 ± 0.16	0.65 ± 0.09	1.22 ± 0.17

Apesar do desafio de distinguir detecções verdadeiras em observações, essa análise enfatiza as propriedades medianas de candidatos a aglomerados de galáxias identificados pelos três localizadores de aglomerados, sugerindo uma grande probabilidade de serem estruturas verdadeiras. Entretanto, para garantir essa conclusão, seria necessário recorrer a simulações. Observe que os valores medianos de riqueza para as detecções não correspondentes em VT e AMICO são semelhantes ao corte de riqueza aplicado durante a seleção de riqueza do catálogo do PZWav. Ao refazer a análise de riqueza, pode-se sugerir que os limites discutidos em Seção 5.3 podem ser utilizados para remover parte da contaminação.

É importante notar que as estruturas não correspondentes, com valores de λ significativamente menores, não indicam uma limitação nos códigos de detecção. Em vez disso, destaca a possível detecção de grupos de menor massa.

5.8 Completude dos catálogos de aglomerados

Os aglomerados de galáxias representam também extensos laboratórios para se estudar a evolução de suas galáxias, considerando os diversos processos que atuam nos componentes do aglomerado durante a evolução do halo (Evrard, 1997; Ettori et al., 2009; Allen et al., 2011; Dvorkin e Rephaeli, 2015). Assim, compreender a relação entre as propriedades luminosas das galáxias e os halos de matéria escura (M_h), ou a relação massa estelar para a massa do halo (do inglês *stellar-to-halo mass*, *SHM*), pode fornecer pistas sobre o papel de diferentes mecanismos físicos que afetam o ambiente. É possível quantificar a contribuição das galáxias mais brilhantes dos aglomerados/grupos (do inglês *brightest*

cluster/group galaxies, BCGs/BGGs) ao restringir uma relação SHM específica.

Estudos, como os conduzidos por Leauthaud et al. (012a,b) e Gozaliasl et al. (2018), fornecem métodos para estimar a massa do halo dada a BCG com base em sua massa estelar.

Seguindo a curva derivada por Gozaliasl et al. (2018), observamos que galáxias com massas estelares $M^* > 2 \times 10^{11} M_\odot$ são esperadas em halos com massas entre 1.4×10^{13} e $1.4 \times 10^{14} M_\odot$. Dado que a relação SHM apresenta uma dispersão intrínseca significativa ($\sigma_{\log M^*} \sim 0.25$, veja Gozaliasl et al., 2018, para detalhes), o limite aplicado não exclui a presença de galáxias menos massivas em halos, mas serve como uma ferramenta para investigar as massas esperadas dos halos em nossa pesquisa.

Embora a distribuição real de aglomerados na área de pesquisa seja desconhecida, podemos comparar a significância de nossos catálogos com a distribuição das próprias galáxias ao supor que tais galáxias massivas sejam as BCGs dos halos detectados.

Para avaliar a associação entre galáxias e nossos catálogos de aglomerados/grupos, realizamos um procedimento de correspondência selecionando todas as galáxias dentro de 1 Mpc do centro do aglomerado selecionado opticamente e dentro de uma faixa de *redshift* de $\Delta z < 0.05$. Uma vez identificadas e removidas, podemos estimar a fração de galáxias sem qualquer associação com os catálogos.

A Figura 5.10 apresenta algumas conclusões para baixos ($0.2 < z < 0.3$) e altos ($0.3 < z < 0.6$) *redshifts*. Observa-se que não temos resultados para $z < 0.2$, pois o catálogo não apresenta galáxias de alta massa nesses *redshifts*. Esse comportamento pode ser atribuído ao volume limitado do levantamento e à baixa probabilidade correspondente fornecida pela função de massa estelar da galáxia (Li e White, 2009). Como o PZWav é o maior catálogo da comparação, este também possui as melhores taxas de correspondência para massas estelares inferiores a $4 \times 10^{11} M_\odot$ em *redshifts* altos. Em geral, conseguimos recuperar aproximadamente 75% das galáxias. Esse valor aumenta para 83% na faixa de baixo *redshift*, indicando que apenas 17% das galáxias estão isoladas. Dentro da faixa de 3 a $5.5 \times 10^{11} M_\odot$, todas as galáxias têm um sistema óptico associado.

Analisamos a fração de galáxias associadas aos catálogos ópticos em função do *redshift* na Figura 5.11 em duas faixas de massa estelar: $2 < M^* [10^{11} M_\odot] < 3$ (correspondendo a uma massa de halo de 1.4 a $4.1 \times 10^{13} M_\odot$) e $3 < M^* [10^{11} M_\odot] < 4.2$ (0.41 a $1.4 \times 10^{14} M_\odot$). As linhas pretas representam a fração de galáxias com correspondência nos três catálogos

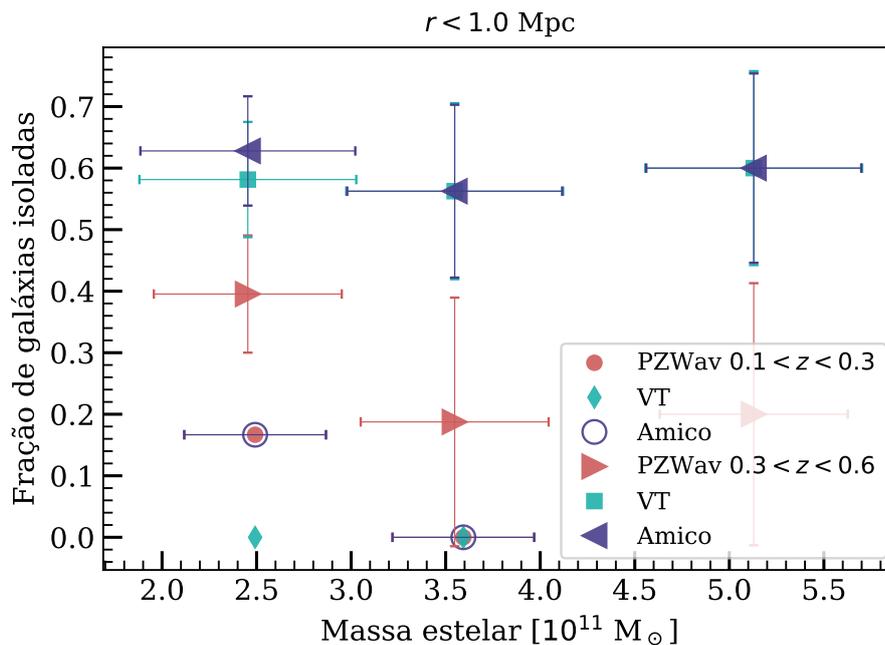


Figura 5.10: Os três algoritmos de detecção de aglomerados PZWav (vermelho), AMICO (azul) e VT (verde) – em comparação com o catálogo de galáxias massivas. As curvas resultantes fornecem a fração de galáxias de uma dada massa estelar sem qualquer associação à aglomerados/grupos.

ópticos, indicando uma identificação mais pura, com aproximadamente 20% das galáxias associadas a aglomerados para $z > 0.3$.

Na primeira faixa de massa, VT (verde) e AMICO (azul) apresentam tendências semelhantes, fornecendo boa cobertura até $z = 0.3$. PZWav (vermelho) tem melhor desempenho em maiores *redshifts* ($z > 0.5$). A curva roxa representa a contribuição combinada de todos os catálogos ópticos, demonstrando a melhoria geral na fração de galáxias isoladas. Além disso, a saída do PZWav com $\text{SNR} > 4$ também é mostrada como uma curva cinza, como um experimento do cenário mais completo. Nesse caso, apenas 20% das galáxias não têm correspondência óptica.

Para a faixa de massa estelar mais alta, $3 < M^* [10^{11} M_{\odot}] < 4.2$, VT e AMICO apresentam comportamento semelhante até $z = 0.6$, após o qual suas contribuições se espalham e mudam de posição para $z > 0.7$. Devido ao seu maior tamanho de amostra, PZWav apresenta melhores frações em toda a faixa de *redshift*, com um valor máximo de 40% de galáxias isoladas para $z > 0.65$.

Como uma análise adicional, podemos avaliar a significância em termos da massa do halo (M_h) seguindo a relação derivada entre a massa estelar e do halo apresentada por Gozaliasl et al. (2018). Para isso, seguimos uma metodologia semelhante à descrita acima

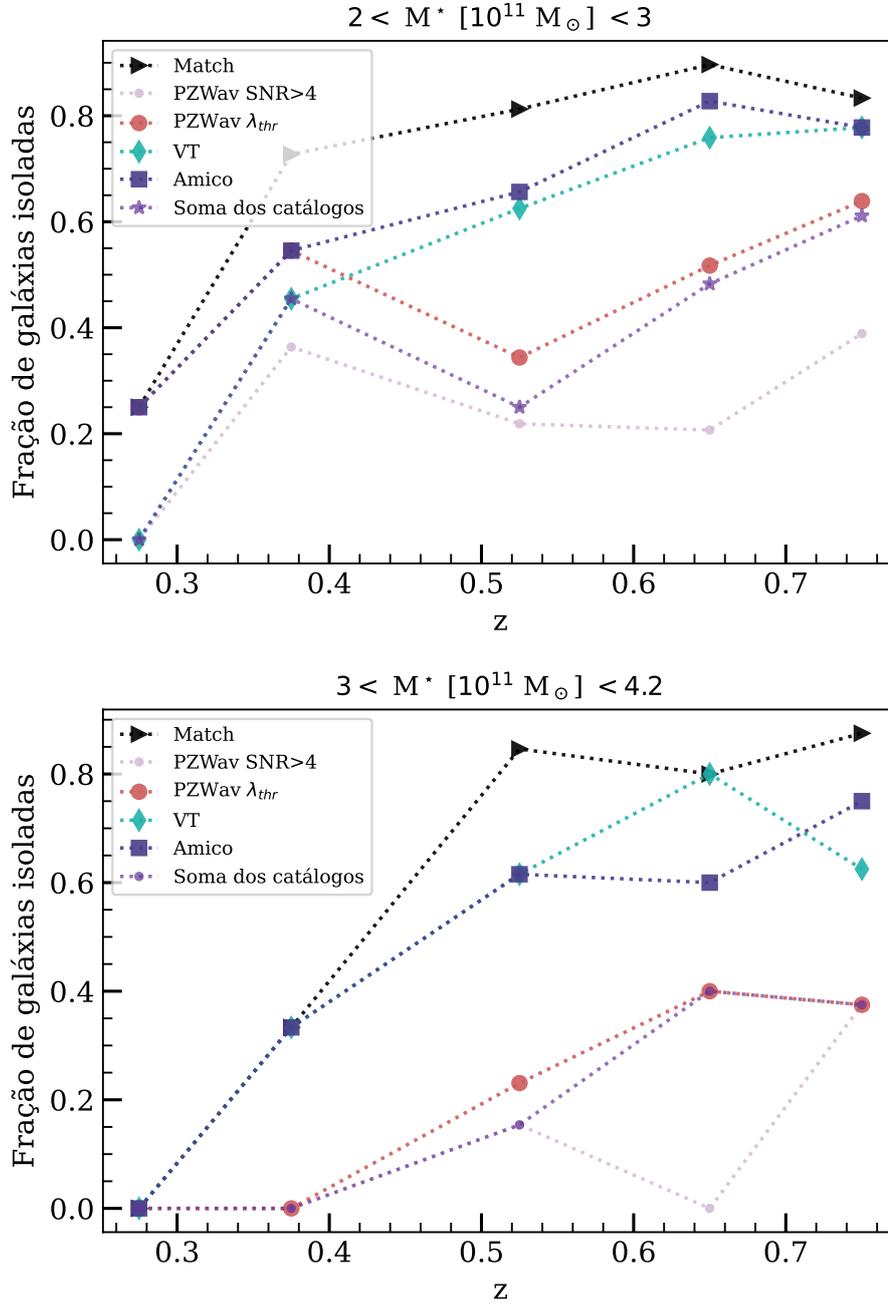


Figura 5.11: Evolução da fração de galáxias sem associação à aglomerados/grupos em função do *redshift* para duas faixas de massa: $2 < M^* [10^{11} M_{\odot}] < 3$ e $3 < M^* [10^{11} M_{\odot}] < 4.2$. Em preto, as correspondências comuns entre catálogos ópticos, cinza exhibe a amostra completa do PZWav, vermelho representa a sub-amostra após o corte em riqueza de PZWav, verde destaca VT, azul AMICO e roxo é a contribuição de todos os catálogos ópticos juntos (considerando a sub-amostra de PZWav).

para estimar as frações de galáxias isoladas dentro de faixas de *redshift* específicos, enquanto exploramos vários limites de massa estelar. Nossa abordagem envolve a adoção de faixas de *redshift* com largura de 0.2 com sobreposição de 0.1 e a identificação do limite ótimo de massa estelar que resulta em uma fração máxima de galáxias isoladas de 0.2. Esses valores foram escolhidos para garantir boa estatística devido aos baixos números de galáxias de alta massa na região. Com os resultados da massa estelar mínima, derivamos M_h .

Na Figura 5.12, ilustramos M_h em função do valor médio do *redshift* dentro de cada faixa para os três localizadores de aglomerados. PZWav $\lambda > \lambda_{\text{thr}}$ é representado por triângulos vermelhos, VT por diamantes verdes e AMICO por quadrados azuis. Para aprimorar a visualização, os valores médios de *redshift* são artificialmente deslocados em 0.01 e 0.02 para VT e AMICO, respectivamente. O gráfico mostra uma tendência semelhante entre os localizadores de aglomerados até $z = 0.4$, mas PZWav apresenta uma cobertura mais ampla para $z > 0.55$ devido ao maior número de detecções.

Esses resultados destacam as capacidades variadas de diferentes catálogos ópticos na detecção e associação de galáxias com aglomerados.

É importante notar que a fração de recuperação e as taxas de associação podem variar dependendo dos critérios específicos e dos procedimentos de correspondência utilizados na análise. No entanto, esta análise fornece uma compreensão mais profunda sobre as propriedades de aglomerados e grupos de galáxias e a distribuição espacial dentro da faixa de *redshift* estudada.

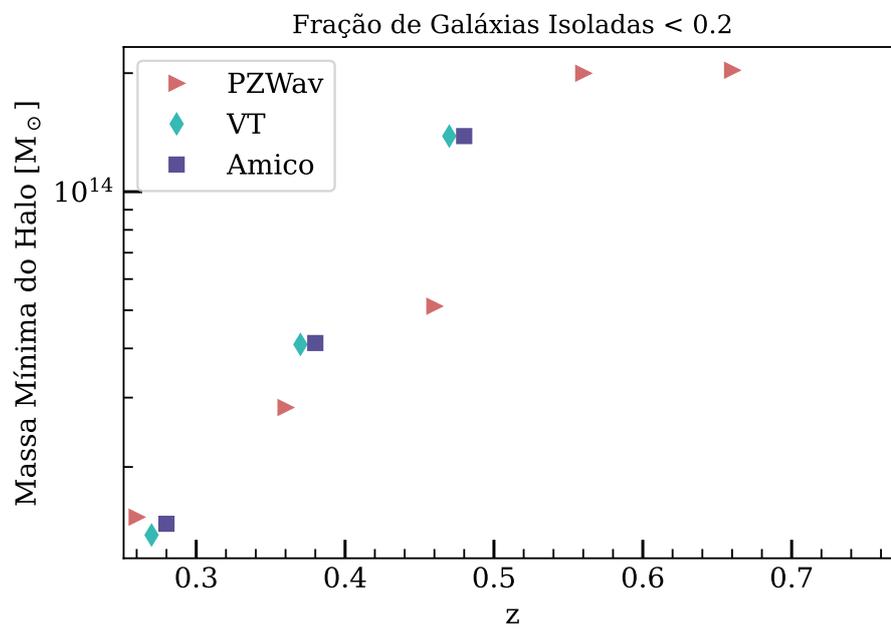


Figura 5.12: A massa mínima do halo em função do *redshift* para garantir uma fração de galáxias isoladas inferior a 0.2. Em triângulos vermelhos são apresentados PZWav $\lambda > \lambda_{thr}$, VT como losangos verdes e quadrados azuis como valores obtidos para o catálogo produzido pelo AMICO. As massas dos halos são estimadas seguindo a relação M^*-M_h descrita em Gozaliasl et al. (2018). Os valores médios do *redshift* são deslocados artificialmente para uma melhor visualização.

Resultados com dados do S-PLUS

Neste capítulo, apresentamos os resultados provenientes do levantamento fotométrico S-PLUS, destacando os números específicos de cada área principal observada. Além disso, conduzimos comparações com informações de levantamentos em diferentes comprimentos de onda, enfatizando a relevância e impacto dos resultados obtidos.

6.1 Detecção de estruturas

Utilizando o algoritmo PZWav (Seção 3.2), identificamos, em números totais, 138 220 grupos e aglomerados de galáxias dentro da faixa de *redshift* de $0.05 < z < 0.5$, com $\text{SNR} > 4$. O formato do catálogo resultante é similar ao apresentado em Seção 5.1, incluindo número de identificação, coordenadas celestes, *redshift* fotométrico e SNR detectado para cada estrutura, assim como propriedades adicionais obtidas com AME (Seção 3.5): riqueza e luminosidade óptica. A descrição das considerações utilizadas para a seleção das galáxias utilizadas pode ser encontrada na Seção 2.3.

Os resultados são apresentados separadamente, para cada área principal de estudo, na Tabela 6.1. Figura 6.1 demonstra, em uma projeção, os resultados em cada uma das grandes áreas: STRIPE 82 em azul escuro; SPLUS-S em índigo; SPLUS-N em carmesim; MC em magenta; e HYDRA em vermelho.

Para entender a pureza e completude de nossa amostras realizamos a comparação entre os catálogos ópticos obtidos com o PZWav com a simulação descrita na Seção 2.1 Definimos completude (C) como a fração de estruturas verdadeiras corretamente identificadas pelo algoritmo, com uma distância entre centros de $\Delta R < 500$ kpc, e diferença de *redshift* de $\Delta z < 0.05$. Pureza é a fração de estruturas verdadeiras em relação a todos os objetos

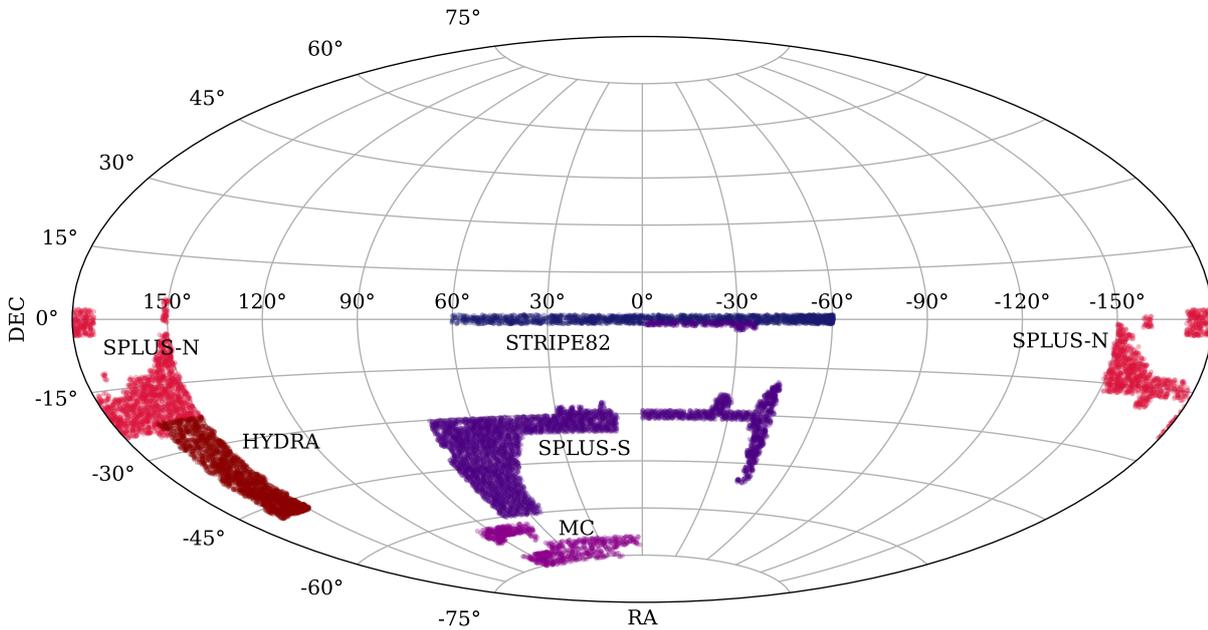


Figura 6.1: Projção do céu meridional demonstrando os resultados obtidos para cada uma das grandes áreas observadas pelo levantamento S-PLUS em diferentes cores. STRIPE82 em azul escuro; SPLUS-S em índigo; SPLUS-N em carmesim; MC em magenta; e HYDRA em vermelho.

Tabela 6.1 - Número de objetos utilizados para a análise em cada uma das grandes áreas observadas pelo levantamento S-PLUS, a quantidade de sobredensidades identificadas pelo algoritmo PZWav, tamanho da área, e as densidades de galáxias e detecções. Excluimos intencionalmente SPLUS-D, devido ao seu foco no disco da Via Láctea.

	Galáxias	Grupos e Aglomerados	area graus ²	densidade gal n_{gal}/graus^2	densidade cl n_{cl}/graus^2
STRIPE 82	855 727	12 153	345.13	2479.43	35.2
SPLUS-S	2 357 080	49 930	1412.04	1669.27	35.4
MC	413 421	8 738	582.84	709.32	15.0
SPLUS-N	1 537 483	56 356	744.72	2064.51	75.7
HYDRA	1 015 831	10 596	388.41	2615.3	27.3

identificados no catálogo (P). Além disso, conhecendo os valores de riqueza para todas as estruturas, simulação e as detecções obtidas pelo PZWav, podemos reconhecer e aplicar limites associados a riqueza. Dessa forma, para a região STRIPE 82, observamos que os catálogos produzidos pelo PZWav são 80% puro e 95% completos para aglomerados e grupos com riqueza superior a 2.9.

Como os aglomerados simulados também apresentam valores de massa, podemos relacionar massa e riqueza. Nesse contexto, o limite de 2.9 corresponde a uma massa de 10^{14} massas solares. Figura 6.2 mostra histogramas da distribuição das riquezas medidas pelo algoritmo desenvolvido AME em cada área. Lembramos que riqueza é definida como a soma das probabilidades de pertinência das galáxias.

Destacamos que tais valores de pureza e completeza se tornam projeções para as outras 4 grandes regiões (SPLUS-S, SPLUS-N, MC, HYDRA), uma vez que a simulação foi criada para reproduzir a primeira versão dos dados do S-PLUS (S-PLUS/DR1) que enquadra a região STRIPE 82. Dessa forma, as características projetadas podem não ser igualmente válidas. Assim, para valores mais precisos, precisaríamos de uma análise mais profunda com simulações atualizadas, que descrevam as bordas, variações de densidades e efeitos sistemáticos das observações.

6.2 Comparação com a literatura

Utilizando a área STRIPE 82, comparamos nossos resultados com outros estudos focados em diferentes abordagens. Para isso, buscamos correspondências entre as estruturas ao comparar as coordenadas do centros, que devem ser menores que 1 Mpc, e diferenças de *redshift* estimados menor que 0.05.

Nesse contexto, apresentamos o catálogo anterior, utilizando dados do DR1 (Werner et al., 2022), juntamente com 9 estudos, sendo 6 baseados em detecções no óptico utilizando dados provenientes do SDSS, assim como em Geach et al. (2011); Rykoff et al. (2014); Wen e Han (2015) e 3 voltados para a observação de características do meio intra-aglomerado: Merten et al. (2011) e Takey et al. (2016); Takey et al. (2019) em raio-X, obtidos pelo telescópio espacial XMM-Newton; e Hilton et al. (2021) que se baseia no efeito Sunyaev–Zel’dovich (SZ). As frações de correspondências são apresentadas na Tabela 6.2.

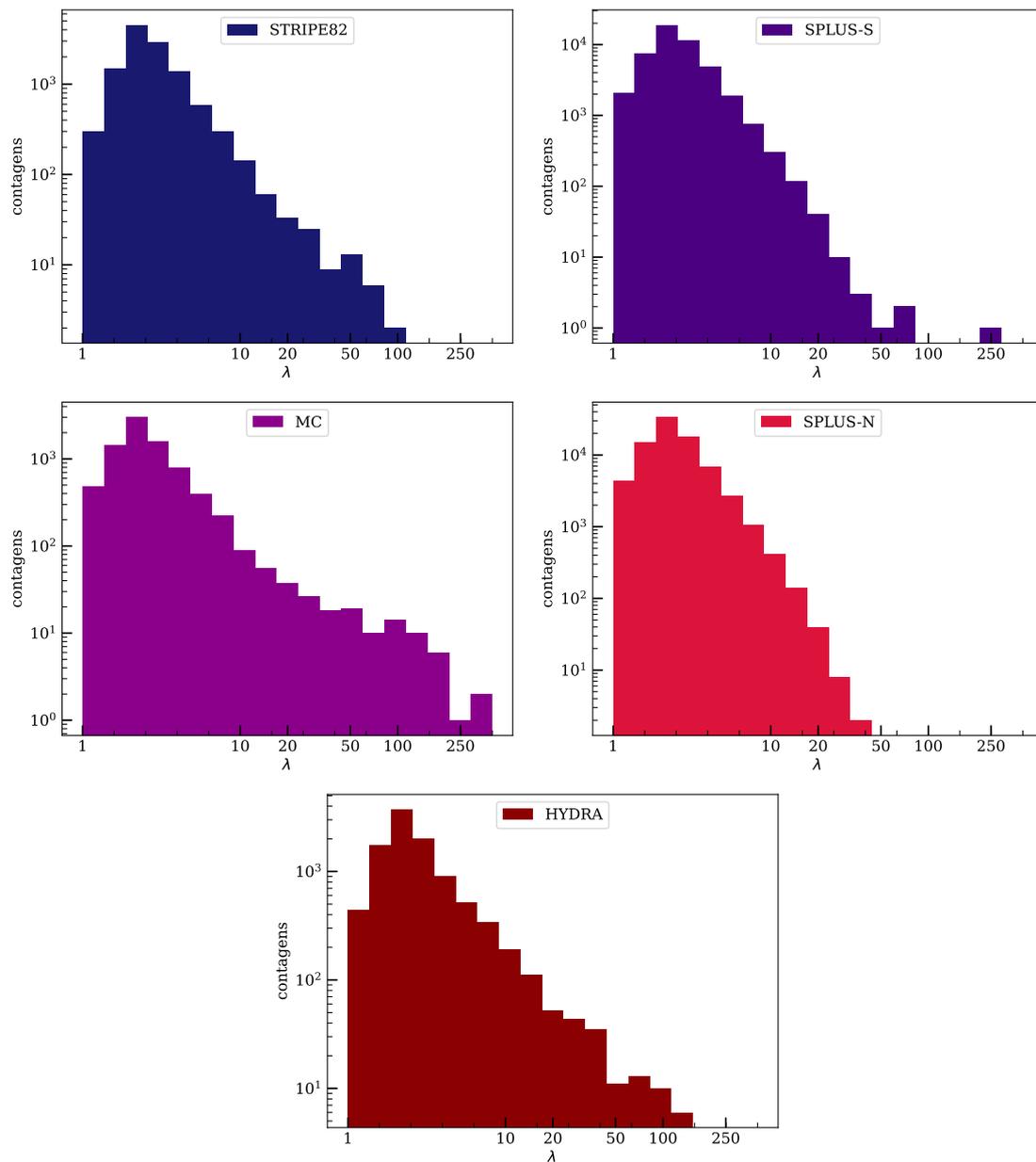


Figura 6.2: Histogramas das riquezas medidas pelo AME. As cores representam as área descritas na Figura 6.1.

Tabela 6.2 - Comparação entre diferentes catálogos obtidos em diferentes comprimentos de onda, e a fração de correspondências de detecção com o catálogo produzido pelo algoritmo PZWav em duas faixas de *redshift* diferentes. O estudo de Hilton et al. (2021) utiliza diferentes fontes, e cortes em magnitude variam entre $r < 21$ e $r < 26$.

Fonte	Catálogo	mag limite	f_{match} $0.05 < z < 0.5$	f_{match} $0.1 < z < 0.4$
Raios-X	Merten et al. (2011)	$r < 25$	74.2%	84%
Raios-X	Takey et al. (2016); Takey et al. (2019)	$r < 24$	78.9%	88.9%
SZ	Hilton et al. (2021)	–	86.5%	89.2%
Óptico	Geach et al. (2011)	$r < 24$	51.5%	54.2%
Óptico	Rykoff et al. (2014)	$i < 21$	83.3%	81.1%
Óptico	Wen e Han (2015)	$r < 22.2$	71.1%	72.5%
Óptico	Rykoff et al. (2016)	$i < 22$	72.8%	83.4%
Óptico	Aguena et al. (2021)	$i < 23$	43.1%	50%
Óptico	Werner et al. (2022)	$r < 21$	51.5%	48.3%

Em estudo recente produzido por Werner et al. (2022), os autores analisaram o primeiro lançamento de dados do levantamento S-PLUS. Para garantir a completeza da amostra de galáxias, os autores restringiram o intervalo de *redshift* entre $0.1 < z < 0.4$. Para a construção do catálogo de detecções, utilizam uma versão anterior do PZWav. Comparações realizadas com a simulação apresentada na Seção 2.1, mostraram que tal versão alcançava os melhores valores de completude e pureza (80%, 85%, respectivamente) para $M_{200} > 10^{14} M_{\odot}$ e PZWav SNR > 3.3 . Aqui, encontramos valores similares, utilizando um corte de SNR > 4 , para o intervalo $0.05 < z < 0.5$.

Para entender completamente os efeitos sistemáticos, seria necessário reanalisar as simulações em detalhes, o que vai além do escopo deste trabalho. Entretanto, ao aplicar os mesmos limites de *redshift* (coluna 4 da Tabela 6.2), vemos que os valores aumentam em média de 7.6%, representando uma pequena melhora. Esse valor pode chegar até 15% para a comparação com o catálogo apresentado por Takey et al. (2016); Takey et al. (2019).

A correspondência entre os catálogos produzidos para DR1 (com 5 533 detecções) e DR4 (com 6 660 detecções, utilizando os mesmos limites) é de apenas 48% da amostra. Na Figura 6.3, é apresentada uma parte da área correspondente à região STRIPE 82, com os catálogos de aglomerados de galáxias gerados a partir de DR1 em azul claro e de DR4 em vermelho. As correspondências entre os catálogos são representadas por círculos pretos. Em DR1, observa-se a presença de efeitos de borda próximos aos limites do levantamento, bem como um alinhamento em $RA \sim 318$, característico da sobreposição de bordas dos *tiles*

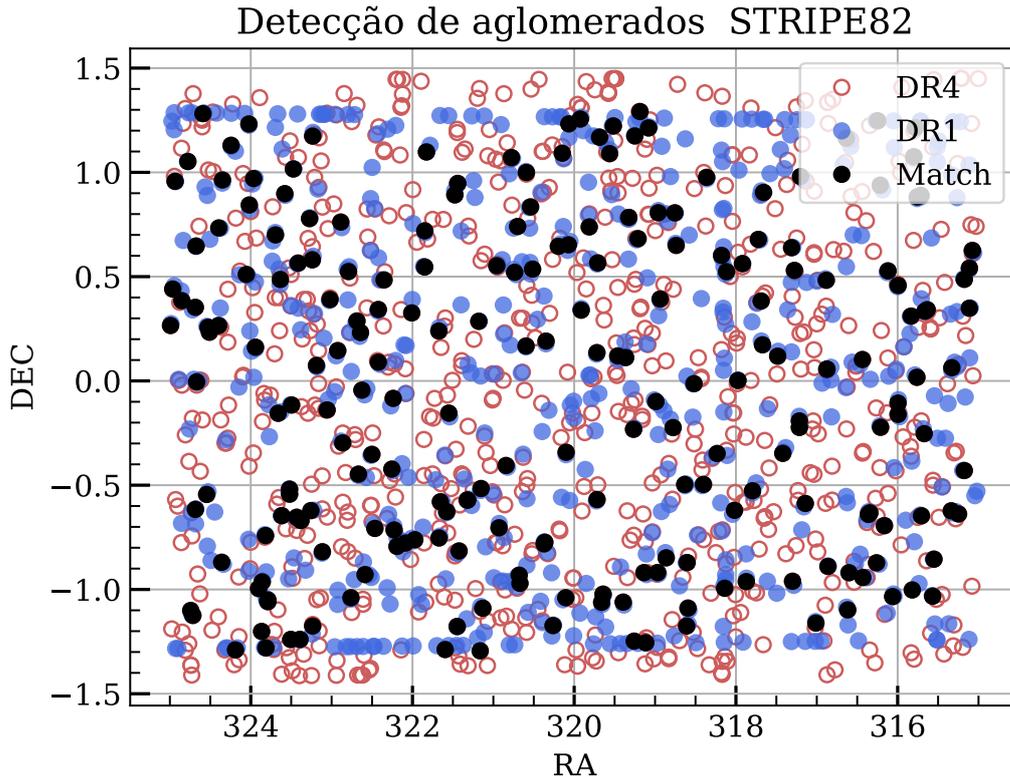


Figura 6.3: Comparação entre catálogos de aglomerados gerados com o primeiro ($0.1 < z < 0.4$) e o quarto lançamento de dados do S-PLUS ($0.05 < z < 0.5$), correspondente a parte da área STRIPE 82.

de observação. Esses problemas persistentes podem levar à falsa detecção de estruturas e prejudicar nossa análise. Identificada-se também sobreposições espaciais que não são consideradas correspondências. Nestes casos, as diferenças nos *redshifts* das estruturas podem ser consideráveis ($\Delta z > 0.05$). É importante destacar que diversos procedimentos de redução de dados, fotometria, separação estrela/galáxias e cálculo de *redshift* foram atualizados entre os dois lançamentos de dados, o que pode explicar essas diferenças entre os catálogos.

Por outro lado, as comparações com Hilton et al. (2021) (52 estruturas identificadas) demonstram que PZWav consegue recuperar 45 (86.5% do catálogo, coluna 3 da Tabela 6.2). Neste trabalho, os autores consideram aglomerados massivos, com massas de $M_{200} > 5.5 \times 10^{14} M_{\odot}$.

Dada a emissão típica de aglomerados e grupos de galáxias em tais comprimentos de onda, os catálogos apresentados baseados em raios-X e SZ são considerados puros. Assim, o número de correspondências obtidos representam um teste direto da completude do catálogo produzido pelo PZWav na faixa de massa coberta por eles.

Em contraste, Rykoff et al. (2014), um estudo realizado no óptico, que sofre de erros sistemáticos semelhantes, adota um método de seleção baseado no limite de riqueza de 20 galáxias. Tal corte, restringe a amostra para aglomerados de massa maior que $10^{14} M_{\odot}$, resultando em 657 estruturas. Apesar das dificuldades, os autores mostram que o catálogo é 95% puro e completo. Nesta comparação, PZWav recupera 83.3% (coluna 3 da Tabela 6.2) do catálogo de detecções.

A comparação em diferentes faixas de redshift, mostra que há uma correlação entre as frações medidas. Para analisar tais variações, consideramos os catálogos com as maiores diferenças de frações: Mehrrens et al. (2012), Takey et al. (2016, 2019), Rykoff et al. (2016). A Figura 6.4 mostra as diferenças em intervalos de $dz = 0.1$. Assim como discutido em Werner et al. (2022), observamos um comportamento similar para a completude do catálogo, que cai para altos redshifts ($z > 0.4$). Este pode estar relacionado à completude das galáxias para altos redshifts. A Figura 6.5 demonstra o número de galáxias por área, em diferentes faixas de magnitude e redshift, e comparam-se os valores obtidos para a simulação, DR4 e DR1 (que apresentam comportamento similar). Nesta imagem, os dados referentes às galáxias do STRIPE 82 do DR4 são representados por círculos vermelhos, enquanto as simulações são mostradas em diamantes azuis escuros e os dados do DR1 são representados por quadrados azuis claros. Para baixos *redshifts*, $0.05 < z < 0.16$, observa-se uma perda de galáxias mais fracas (inclinação descendente para $r > 19.8$). Enquanto em altos *redshifts*, $0.39 < z < 0.5$, o número obtido é aproximadamente a metade do valor esperado pelas simulações. Nota-se em alguns dos quadros que a densidade de galáxias para o DR4 é menor que DR1. Isso pode ocorrer devido aos cortes aplicados para remoção das sobreposições das observações, e seleções dos dados..

Surpreendentemente, ao considerar as variações entre os centros das estruturas e redshifts das estruturas detectadas pelo PZWav com os outros catálogos (dois a dois), vemos variações médias consistentes entre todos as comparações: para $0.05 < z < 0.5$ ($0.1 < z < 0.4$) $\Delta R = 187 \pm 96$ kpc (179 ± 91 kpc), e -0.004 ± 0.013 ($\Delta z = -0.002 \pm 0.012$). Esse valor, também é menor que o encontrado em Werner et al. (2022), de 240 ± 170 kpc. Figura 6.6 mostra as variações obtidas em relação ao *redshift* e distâncias ao centro para cada catálogo de comparação. Nesta figura, as linhas pontilhadas indicam a média das médias entre toda a faixa de redshift, $0.05 < z < 0.5$, para guiar os olhos.

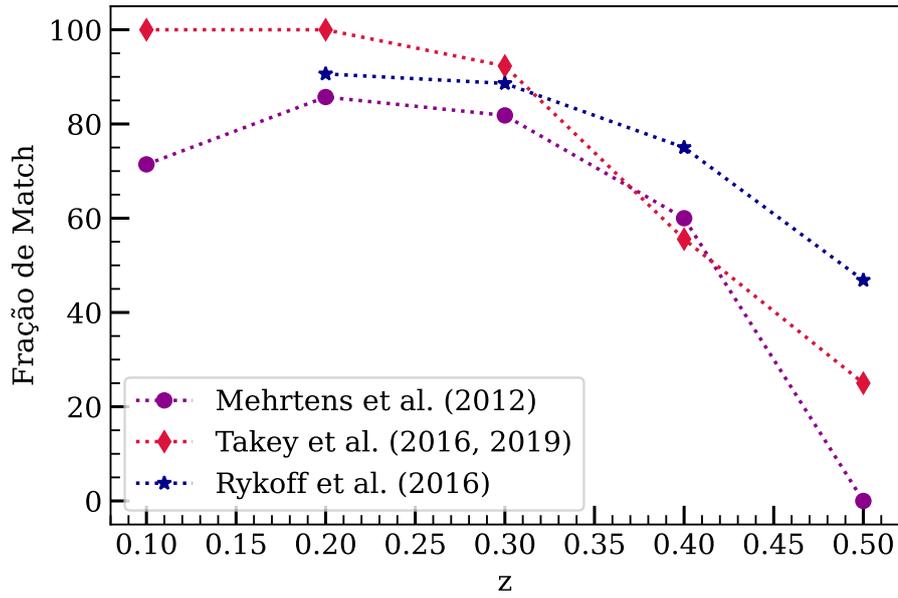


Figura 6.4: Variação da fração de correspondências (*Match*) em função do *redshift* para os catálogos com as maiores diferenças dos valores medidos na Tabela 6.2. Em magenta escuro, Mehrtens et al. (2012); carmesim Takey et al. (2016, 2019); e azul escuro Rykoff et al. (2016).

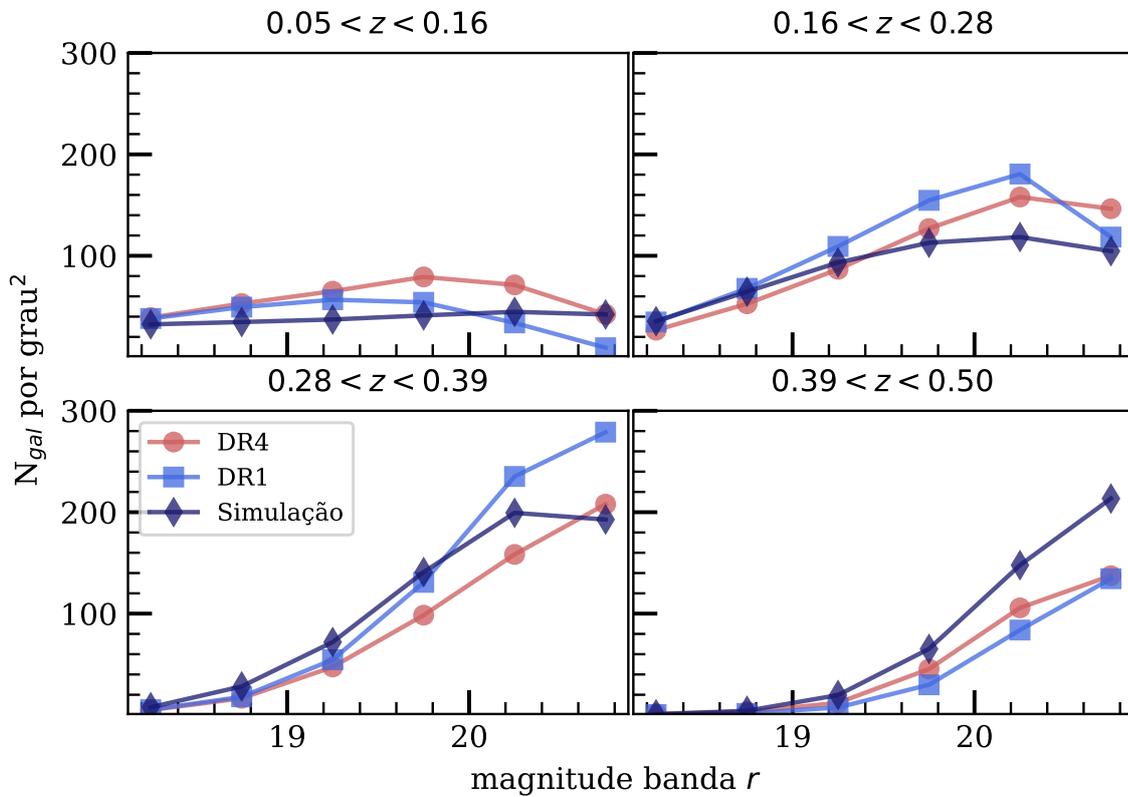


Figura 6.5: Número de galáxias por área, em função da magnitude aparente, para a faixa de $0.1 < z < 0.4$. As linhas vermelhas representam S-PLUS/DR4, S-PLUS/DR1 em azul claro, e a simulação em azul escuro.

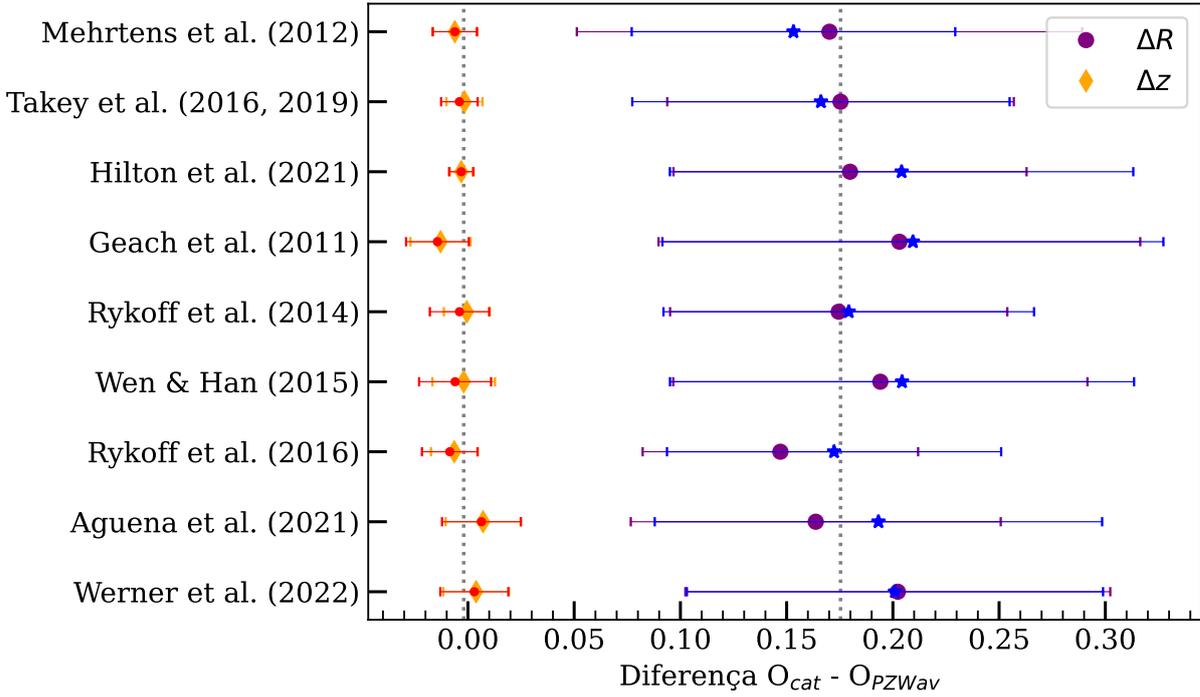


Figura 6.6: Valores médios e desvios padrão para as variações de centro (em kpc) e redshifts entre as estruturas encontradas pelo PZWav e os catálogos apresentados na Tabela 6.2 em duas faixas de redshift, $0.05 < z < 0.5$ (círculos roxos, e diamantes laranjas) e $0.1 < z < 0.4$ (estrelas azuis e pontos vermelhos). As linhas pontilhadas mostram a média das médias para $0.05 < z < 0.5$ para guiar os olhos.

6.2.1 Centros e redshifts a partir dos membros

Assim como em Seção 4.8.1, analisamos a diferença entre o *redshift* derivado das galáxias membros obtidos por AME e o valor obtida pelo algoritmo de detecção PZWav, bem como as discrepâncias em relação à determinação centros.

Para as diferenças de *redshifts*, calculamos o valor médio ponderando cada contribuição de foto-z da galáxia por seu P_{mem} , como $z_{gals} = \sum z_i P_{mem,i} / \sum P_{mem,i}$.

Para as estimativas de centro, estimamos o baricentro dos aglomerados introduzindo os pesos das probabilidades P_{mem} . Então, para a ascensão reta, $\alpha_{gals} = \sum \alpha_i P_{mem,i} / \sum P_{mem,i}$, e similarmente para declinação.

Aqui também calculamos a média de todos os aglomerados uma vez que não encontramos tendências significativas na faixa de *redshift*. O comportamento é também semelhante entre áreas. As diferenças são altamente coerentes entre o *redshift* definido pelo código de detecção e as galáxias selecionadas, resultando em uma diferença total consistente com zero, 1×10^{-4} , com $\sigma_{MAD} \leq 10^{-3}$ (similar aos resultados apresentados em Seção 4.8.1). Uma vez que PZWav também utiliza as galáxias próximas ao centro para estimar z , espera-se

que variações sejam pequenas.

Os centros apresentam menores diferenças do que as encontradas em Seção 4.8.1. Uma vez que o PZWav estima os centros pelo pico das sobredensidades, nosso método considera a distribuição das galáxias dentro do raio R_c . Como destacado anteriormente, a presença de galáxias falso-positivos podem deslocar o centro dependendo de seu P_{mem} . Ainda assim, o erro médio encontrado foi de 93 ± 37 kpc. Valores que aumentam em média apenas 12 kpc (105 ± 41 kpc, valor máximo) para a região MC e encontra valores mínimos como 91 ± 36 para HYDRA.

Conclusões

Neste trabalho, apresentamos um estimador de pertinência de galáxias (do inglês *adaptive membership estimator* AME), que usa parâmetros fotométricos para derivar estimativas confiáveis de riqueza, luminosidade óptica e massa estelar total para 919 estruturas simuladas com massas variando de grupos a aglomerados de galáxias. A abordagem assume pouca informação sobre a definição do aglomerado e usa a posição do céu dentro de um raio característico (R_c) para selecionar potenciais membros da galáxia.

Abaixo, destacamos as principais conclusões considerando o catálogo simulado:

- O raio característico R_c escala linearmente como $R_c = 0.6 R_{200}$, com um desvio absoluto mediano de 0.15 Mpc.
- Testes com pontos aleatórios distribuídos ao longo do *redshift* das simulações mostram que o método de abertura fixa FAE retorna σ_{MAD} de ~ 1.05 nas riquezas λ_{FAE} . Aplicando este valor como um limite de riqueza para o estimador adaptativo de pertinência de galáxias, λ_{AME} , obtemos corretamente $\lambda_{AME} = 0$.
- Considerando a amostra do grupo/aglomerado, as comparações entre λ_{mock} e λ_{AME} produzem uma relação linear em valores medianos, com desvio de -0.01 ± 0.12 .
- Com os resultados probabilísticos, derivamos com sucesso traçadores ópticos de massa que são simples, com baixo custo observacional e apresentam pequena dispersão intrínseca.

Das relações de escala:

$$\sigma_{\log_{10}(M|\lambda)} = 0.181 \pm 0.009 \text{ dex}$$

$$\sigma_{\log_{10}(M|L_\lambda)} = 0.151 \pm 0.007 \text{ dex}$$

$$\sigma_{\log_{10}(M|M_{\lambda}^*)} = 0.097 \pm 0.005 \text{ dex.}$$

- Mostramos que nosso estimador adaptativo é robusto para os deslocamentos de centro assim como para pequenos desvios de *redshift*. Os deslocamentos propostos por Werner et al. (2022) produzem variações de riqueza medidas inferiores a 1% nos nossos testes.
- Testes de complexidade da função de densidade de probabilidade foto-z revelam que a presença de caudas longas não afeta significativamente os resultados. No entanto, nos casos em que a PDF exibe uma distribuição bimodal, há uma potencial subestimação da riqueza em aproximadamente 15% quando essa incerteza afeta metade da amostra. Considerando que a fração de bimodalidade nos dados S-PLUS e miniJPAS é de apenas 8% e 10% respectivamente, prevemos que o fenômeno de picos duplos não é uma preocupação mais significativa do que nosso teste.

Em relação às probabilidades de pertencimento individuais:

- A fração de galáxias membro verdadeiras em intervalos de P_{mem} enfatiza a contaminação de intrusos em valores de probabilidade mais altas.

A melhor concordância entre pureza e integridade é $P = 63\%$ e $C = 85\%$, que é obtida com $P_{mem} \geq 0.75$. Os resultados estão de acordo com George et al. (2011); Castignani e Benoist (2016).

- Analisamos as distribuições em diferentes intervalos de *redshift* e massas e descobrimos que selecionamos melhor as galáxias membros verdadeiras em escalas de grupo e para intervalos de *redshift* mais altos. O resultado provavelmente está relacionado à distribuição espacial dos aglomerados, onde os grupos tendem a ser mais concentrados.
- A comparação entre os valores médios de centro e *redshift* obtidos através das galáxias membros selecionadas e as nominais, fornecidas pela simulação, mostra valores de *redshift* semelhantes, e variações de posições centrais em torno de 64 kpc entre simulação e galáxias selecionadas, e 128 kpc para PZWav e galáxias selecionadas.

Concluimos que nosso estimador é robusto a pequenos deslocamentos e produz traçadores ópticos de massa que são competitivos com outros observáveis tradicionais. Nosso método

pode ser aplicado em levantamentos fotométricos presentes e futuros, como S-PLUS e J-PAS, juntamente com qualquer lista de aglomerados ou grupos de galáxias produzidos por localizadores de aglomerados.

Na segunda parte do estudo conduzimos uma análise detalhada usando o algoritmo PZWav na pesquisa miniJPAS, servindo como um precursor para a versão completa do levantamento J-PAS. Para aprimorar nossa análise, utilizamos o AME para a caracterização dos candidatos a aglomerados.

Em resumo, as principais conclusões desta parte do estudo são as seguintes:

- O catálogo PZWav é composto por 574 detecções em um intervalo de *redshift* de $0.05 < z < 0.8$.
- Aplicando um método de descontaminação baseado na riqueza, resulta em 221 candidatos.
- Os regimes de grupos e aglomerados nos permitem caracterizar corretamente sistemas com baixo *redshift* ($z < 0.3$).
- Fazendo uso do AME, podemos calcular traçadores ópticos para relações de escala, como riqueza, luminosidade óptica e massa estelar, ponderados pela probabilidade de pertinência da galáxia.

Resultados preliminares das relações de escala mostram pequena correlação entre os traçadores.

- A sobreposição de áreas com o levantamento de raios-X AEGIS revelou 17 fontes com uma contraparte óptica.

Entre eles, recuperamos uma estrutura de $M_{200} = (3.75 \pm 0.5) 10^{13} M_{\odot}$ em $z = 0.745$.

A reanálise dos dados tomando como base os centros ópticos detectados pelo PZWav resultou em um novo catálogo de raios-X com 37 aglomerados. Dentro destes, 20 são novas detecções.

- O pareamento com diferentes algoritmos de localização no ópticos (PZWav, AMICO e VT) revelou 43 identificações comuns.

Redshifts e deslocamentos centrais estão dentro de 0.005 e 0.13 Mpc, respectivamente.

As estatísticas revelaram diferenças notáveis entre as propriedades de aglomerados de galáxias com e sem uma contraparte. Os valores médios dos traçadores de massa são consistentemente mais altos para os aglomerados que apresentam uma correspondência, indicando sua maior confiabilidade.

- A análise da fração de galáxias isoladas forneceu uma expectativa de completude do catálogo. Para halos de baixa massa, $C \sim 40\%$, enquanto os maiores apresentam $C \sim 50\%$.

Baseado nos números acima, concluímos que PZWav demonstra bom desempenho na detecção de aglomerados de galáxias através da utilização da função de densidade de probabilidade (PDF) de *redshifts* fotométricos. A capacidade de detectar estruturas de baixa massa em altos *redshifts* destaca a profundidade do levantamento. Esta análise preliminar, juntamente com outros estudos com foco na pesquisa miniJPAS, mostra o potencial da pesquisa J-PAS até as escalas de grupo.

Na terceira e última parte do estudo analisamos a distribuição de galáxias do 4º lançamento de dados do levantamento S-PLUS. Utilizamos o algoritmo PZWav para a detecção de estruturas nas 5 das 6 áreas principais, somando em torno de 3000 graus quadrados. Comparamos nossos resultados com a literatura, dando ênfase nas diferenças entre centros e *redshifts* encontrados.

- O catálogo produzido contém 194 129 grupos e aglomerados de galáxias dentro da faixa de *redshift* de $0.05 < z < 0.5$, com $\text{SNR} > 4$.
- Comparações com as simulações mostram que os catálogos são 80% puro e 95% completos para aglomerados com riqueza superior a 2.9, que corresponde a uma massa de $10^{14} M_{\odot}$.
- Devido a pureza do catálogo com base no efeito de Sunyaev–Zel’dovich, produzido por Hilton et al. (2021), destacamos a comparação com nossos resultados. Das 52 estruturas, recupera-se 45 representando 86.5% do catálogo entre os *redshifts* $0.05 < z < 0.5$.

- Encontramos uma redução da fração de correspondências para maiores *redshifts*, provavelmente relacionado a completude do catálogo de galáxias.
- A comparações 2-a-2 com os catálogos da literatura mostram variações $\Delta R = 187 \pm 96$ kpc (179 ± 91 kpc), e $\Delta z = -0.004 \pm 0.013$ (-0.002 ± 0.012) para $0.05 < z < 0.5$ ($0.1 < z < 0.4$).
- A comparação dos resultados utilizando as galáxias membro dos aglomerados mostram variações médias consistentes com zero em *redshift*, e 93 ± 37 kpc para o deslocamento dos centros.

A pesquisa S-PLUS proporciona uma grande oportunidade para explorar o Universo. Com sua ampla cobertura do céu e estimativas fotométricas precisas feitas por algoritmos de aprendizado de máquina permite a detecção de estruturas em larga escala, e oferece uma visão detalhada dos membros dos aglomerados. Ampliar as comparações com catálogos de aglomerados de raios-X e Sunyaev–Zel’dovich nas regiões HYDRA, MC, SPLUS-S e SPLUS-N tem o potencial de aprimorar os catálogos ópticos de aglomerados, e permite construir relações de escala para posterior uso em cosmologia.

Perspectivas Futuras

Neste capítulo mostramos exemplos da literatura que abordam a análise da imposição de limites dos parâmetros cosmológicos e seus resultados. Introduzimos também os resultados esperados dada nossa análise considerando riqueza óptica e as tensões encontradas nas medidas de σ_8 . Apresentamos também um estudo em andamento baseado nos resultados do AME aplicados ao levantamento fotométrico S-PLUS.

8.1 Cosmologia a partir de contagens de aglomerados

Dado que informações cosmológicas podem ser derivada a partir de previsões de relações de escala, é crucial minimizar a possível dispersão na relação massa-observável. Neste trabalho, apresentamos um algoritmo probabilístico para calcular os traçadores de massa, realizamos testes em simulação e suas aplicações nos levantamentos miniJPAS e S-PLUS.

Nos últimos anos, a cosmologia utilizando aglomerados de galáxias avançou significativamente com a introdução de novas amostras selecionadas em diferentes comprimentos de onda, assim como na melhorias nos métodos de tratamento das incertezas associadas às massas dos aglomerados. Esses estudos recentes têm destacado algumas pequenas divergências nas medições entre os métodos, como sumarizado na Figura,8.1. Pratt et al. (2019) discute diversas interpretações para tais tensões nas restrições, concentrando-se em questões sistemáticas, como calibração e modelagem instrumental, e levando em consideração a incerteza sistemática relacionada aos efeitos de seleção de aglomerados e determinação de massa.

Como próximos passos, destacamos para o levantamento S-PLUS as estimativas de massa para as detecções de aglomerados de galáxias, utilizando dados de elipticidade de

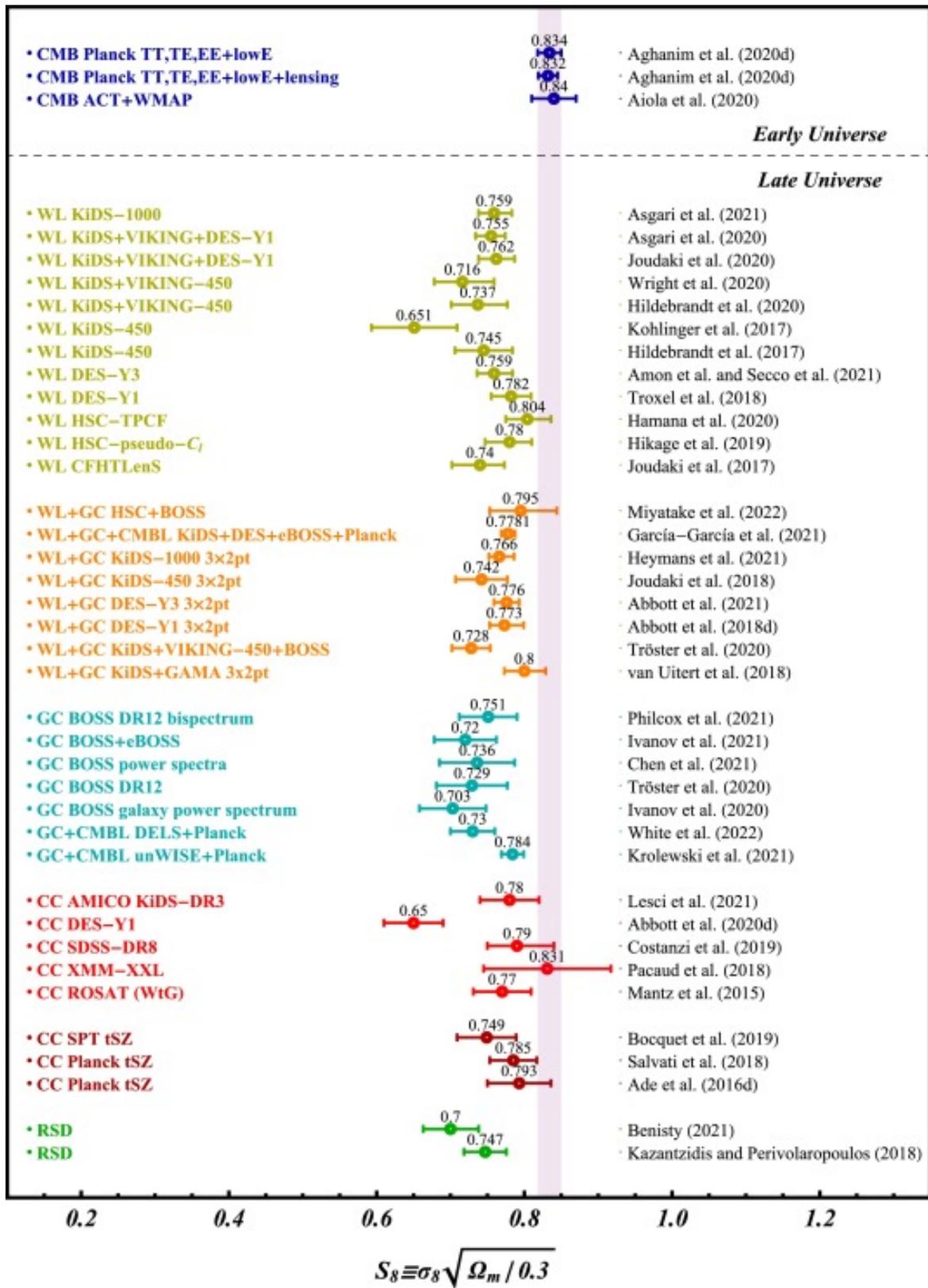


Figura 8.1: Tensão nas restrições de σ_8 entre diversos estudos. Figura retirada de Abdalla et al. (2022).

levantamentos como o CS82 (Leauthaud et al., 2017), DES (The Dark Energy Survey Collaboration, 2005) ou Kids (de Jong et al., 2013). Seguido da análise das funções de seleção, incertezas estatísticas e sistemáticas para finalmente restringir os parâmetros cosmológicos. No contexto do levantamento J-PAS, será necessário repetir toda a análise quando os dados estiverem completos e maduros. Seu primeiro lançamento de dados ocorreu em novembro de 2023, cobrindo aproximadamente 6.8 graus quadrados do céu.

Como exemplo desse desafio, destacamos o estudo realizado por Costanzi et al. (2019), no qual foram utilizadas amostras ópticas selecionadas a partir do algoritmo redMaPPer (Rykoff et al., 2014), aplicado ao Sloan Digital Sky Survey (DR8). Os perfis de massa obtidos a partir de lentes fracas, em intervalos de riqueza, revelaram um valor de $\sigma_8(\Omega_M/0.3)^{0.5} = 0.79_{-0.04}^{+0.05}$. Nesse estudo, os autores enfrentaram diversas incertezas estatísticas e erros sistemáticos, tais como o viés nas medidas de *redshift* e nas medidas de lentes gravitacionais (triaxialidade do halo, efeitos de projeção e má centralização). O modelo utilizado baseou-se em estatística Bayesiana e na amostragem por MCMC.

8.2 Aglomerados multimodais

Outra aplicação da informação obtida por meio das probabilidades de pertencimento das galáxias é a capacidade de detectar aglomerados multimodais, isto é, aglomerados que estejam passando por um processo de fusão. Alguns desses sistemas seriam selecionados para investigações mais aprofundadas, como imagens mais detalhadas, espectroscopia ou observações em raios-X.

O interesse principal reside na descoberta de novos sistemas, semelhantes ao *bullet cluster* (Clowe et al., 2006), onde o gás intra-aglomerado se destaca das galáxias e da matéria escura. Em princípio, essa pesquisa seria conduzida utilizando dados ópticos e técnicas como as descritas em Einasto et al. (2012); Wen et al. (2012); Tempel et al. (2017), que empregam informações sobre as posições no céu e em z das galáxias, ponderadas pela sua probabilidade de pertencimento, para identificar multimodalidades.

Como o AME foi construído levando em consideração as probabilidades das galáxias em relação à estrutura principal, sofreu pequenas alterações para fornecer resultados não apenas para a estrutura principal, mas também para as subestruturas identificadas pelo HDBSCAN. Essa análise está sendo conduzida por Giulya Souza em seu mestrado, uti-

lizando a verãõ específica do código, e os catálogos obtidos nesta tese para o S-PLUS (Capítulo 6), e tem apresentado resultados interessantes, como a detecção de multimodalidades com base na combinação das razões de riqueza entre a estrutura principal e as subestruturas.

Referências Bibliográficas

- Elcio Abdalla et al. Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies, *Journal of High Energy Astrophysics*, 2022, vol. 34, p. 49
- Abell G. O., The Distribution of Rich Clusters of Galaxies., *ApJS*, 1958, vol. 3, p. 211
- Adami C., Mazure A., Biviano A., Katgert P., Rhee G., The ESO nearby Abell cluster survey. IV. The fundamental plane of clusters of galaxies, *A&A*, 1998, vol. 331, p. 493
- M. Aguena et al. The WaZP galaxy cluster sample of the dark energy survey year 1, *MNRAS*, 2021, vol. 502, p. 4435
- Hiroaki Aihara et al. THE EIGHTH DATA RELEASE OF THE SLOAN DIGITAL SKY SURVEY: FIRST DATA FROM SDSS-III, *The Astrophysical Journal Supplement Series*, 2011, vol. 193, p. 29
- Allen S. W., Evrard A. E., Mantz A. B., Cosmological Parameters from Observations of Galaxy Clusters, *Annual Review of Astronomy and Astrophysics*, 2011, vol. 49, p. 409–470
- Almosallam I. A., Jarvis M. J., Roberts S. J., GPZ: non-stationary sparse Gaussian processes for heteroscedastic uncertainty estimation in photometric redshifts, *MNRAS*, 2016, vol. 462, p. 726
- Andreon S., Making the observational parsimonious richness a working mass proxy, *A&A*, 2015, vol. 582, p. A100

- Araya-Araya P., Vicentin M. C., Sodr e Laerte J., Overzier R. A., Cuevas H., Protocluster detection in simulations of HSC-SSP and the 10-yr LSST forecast, using PCcones, MNRAS, 2021, vol. 504, p. 5054
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North, MNRAS, 1999, vol. 310, p. 540
- Arnouts S., Ilbert O., , 2011 LePHARE: Photometric Analysis for Redshift Estimate Astrophysics Source Code Library, record ascl:1108.009
- Bartelmann M., Schneider P., Weak gravitational lensing, Phys. Rep., 2001, vol. 340, p. 291
- Becker M. R., Kravtsov A. V., On the Accuracy of Weak-lensing Cluster Mass Reconstructions, ApJ, 2011, vol. 740, p. 25
- Bellagamba F., Roncarelli M., Maturi M., Moscardini L., AMICO: optimized detection of galaxy clusters in photometric surveys, MNRAS, 2018, vol. 473, p. 5221
- Bellagamba F., Sereno M., Roncarelli M., Maturi M., Radovich M., Bardelli S., Puddu E., Moscardini L., Getman F., Hildebrandt H., Napolitano N., AMICO galaxy clusters in KiDS-DR3: weak lensing mass calibration, MNRAS, 2019, vol. 484, p. 1598
- Ben itez N., Bayesian Photometric Redshift Estimation, ApJ, 2000, vol. 536, p. 571
- N. Benitez et al. J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey, arXiv e-prints, 2014, p. arXiv:1403.5237
- Benoist C., How to best define cluster richness?. In Building the Euclid Cluster Survey - Scientific Program , Proceedings of a Conference held July 6–11 2014 at the Sexten Center for Astrophysics, 2014, p. 8
- Bertin E., Arnouts S., SExtractor: Software for source extraction., A&AS, 1996, vol. 117, p. 393
- Betancourt M., A Conceptual Introduction to Hamiltonian Monte Carlo, arXiv e-prints, 2017, p. arXiv:1701.02434

- Bhattacharya S., Mahulkar V., Pandaokar S., Singh P. K., Morphology of open clusters NGC 1857 and Czernik 20 using clustering algorithms, *Astronomy and Computing*, 2017, vol. 18, p. 1
- Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., Formation of galaxies and large-scale structure with cold dark matter., *Nature*, 1984, vol. 311, p. 517
- S. Bonoli et al. The miniJPAS survey: A preview of the Universe in 56 colors, *A&A*, 2021, vol. 653, p. A31
- Brammer G. B., van Dokkum P. G., Coppi P., EAZY: A Fast, Public Photometric Redshift Code, *ApJ*, 2008, vol. 686, p. 1503
- Brunner R. J., Lubin L. M., A Probabilistic Quantification of Galaxy Cluster Membership, *AJ*, 2000, vol. 120, p. 2851
- Bundy K., Leauthaud A., Saito S., Maraston C., Wake D. A., Thomas D., The Stripe 82 Massive Galaxy Project. III. A Lack of Growth among Massive Galaxies, *The Astrophysical Journal*, 2017, vol. 851, p. 34
- Campello R. J. G. B., Moulavi D., Sander J., Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining* , Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, p. 160
- Carlberg R. G., Yee H. K. C., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C. J., Galaxy Cluster Virial Masses and Omega, *ApJ*, 1996, vol. 462, p. 32
- Carrasco Kind M., Brunner R. J., TPZ: photometric redshift PDFs and ancillary information by using prediction trees and random forests, *MNRAS*, 2013, vol. 432, p. 1483
- Castignani G., Benoist C., A new method to assign galaxy cluster membership using photometric redshifts, *A&A*, 2016, vol. 595, p. A111
- Chan M. H., Two Analytic Relations Connecting the Hot Gas Astrophysics with the Cold Dark Matter Model for Galaxy Clusters, *The Astrophysical Journal*, 2021, vol. 923, p. 95

- Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., A Direct Empirical Proof of the Existence of Dark Matter, *ApJ*, 2006, vol. 648, p. L109
- M. Colless et al. The 2dF Galaxy Redshift Survey: spectra and redshifts, *MNRAS*, 2001, vol. 328, p. 1039
- Comerford J. M., Natarajan P., The observed concentration–mass relation for galaxy clusters, *Monthly Notices of the Royal Astronomical Society*, 2007, vol. 379, p. 190
- M. Costanzi et al. Methods for cluster cosmology and application to the SDSS in preparation for DES Year 1 release, *MNRAS*, 2019, vol. 488, p. 4779
- Courteau S., Cappellari M., de Jong R. S., Dutton A. A., Emsellem E., Hoekstra H., Koopmans L. V. E., Mamon G. A., Maraston C., Treu T., Widrow L. M., Galaxy masses, *Reviews of Modern Physics*, 2014, vol. 86, p. 47
- Cypriano E. S., Sodr e Laerte J., Kneib J.-P., Campusano L. E., Weak-Lensing Mass Distributions for 24 X-Ray Abell Clusters, *ApJ*, 2004, vol. 613, p. 95
- David L. P., Jones C., Forman W., Cosmological implications of ROSAT observations of groups and clusters of galaxies, *ApJ*, 1995, vol. 445, p. 578
- M. Davis et al. The All-Wavelength Extended Groth Strip International Survey (AEGIS) Data Sets, *ApJ*, 2007, vol. 660, p. L1
- de Jong J. T. A., Verdoes Kleijn G. A., Kuijken K. H., Valentijn E. A., The Kilo-Degree Survey, *Experimental Astronomy*, 2013, vol. 35, p. 25
- DESI Collaboration et al. The DESI Experiment Part II: Instrument Design, arXiv e-prints, 2016, p. arXiv:1611.00037
- L. Doubrawa et al. The miniJPAS survey: Optical detection of galaxy clusters with PZWav, arXiv e-prints, 2024, p. arXiv:2312.12197
- Doubrawa L., Cypriano E. S., Finoguenov A., Lopes P. A. A., Maturi M., Gonzalez A. H., Dupke R., Galaxy cluster optical mass proxies from probabilistic memberships, *Monthly Notices of the Royal Astronomical Society*, 2023, p. stad3024

- Duane S., Kennedy A., Pendleton B. J., Roweth D., Hybrid Monte Carlo, *Physics Letters B*, 1987, vol. 195, p. 216
- Dvorkin I., Rephaeli Y., Evolution of the gas mass fraction in galaxy clusters, *Monthly Notices of the Royal Astronomical Society*, 2015, vol. 450, p. 896
- Einasto M., Liivamägi L. J., Tempel E., Saar E., Vennik J., Nurmi P., Gramann M., Einasto J., Tago E., Heinämäki P., Ahvensalmi A., Martínez V. J., Multimodality of rich clusters from the SDSS DR8 within the supercluster-void network, *A&A*, 2012, vol. 542, p. A36
- Elahi P. J., Power C., Lagos C. d. P., Poulton R., Robotham A. S. G., Using velocity dispersion to estimate halo mass: Is the Local Group in tension with LCDM?, *Monthly Notices of the Royal Astronomical Society*, 2018, vol. 477, p. 616
- Eldridge J., Belkin M., Wang Y., Beyond Hartigan Consistency: Merge Distortion Metric for Hierarchical Clustering, *arXiv e-prints*, 2015, p. arXiv:1506.06422
- Ellis J., Hagelin J., Nanopoulos D., Olive K., Srednicki M., Supersymmetric relics from the big bang, *Nuclear Physics B*, 1984, vol. 238, p. 453
- G. Erfanianfar et al. X-Ray Groups of Galaxies in the AEGIS Deep and Wide Fields, *ApJ*, 2013, vol. 765, p. 117
- Ester M., Kriegel H.-P., Sander J., Xu X., A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *KDD* , 1996
- Ettori S., The generalized scaling relations for X-ray galaxy clusters: the most powerful mass proxy, *MNRAS*, 2013, vol. 435, p. 1265
- Ettori S., Morandi A., Tozzi P., Balestra I., Borgani S., Rosati P., Lovisari L., Terenziani F., The cluster gas mass fraction as a cosmological probe: a revised study, *A&A*, 2009, vol. 501, p. 61
- Euclid Collaboration et al. Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection, *A&A*, 2019, vol. 627, p. A23
- Euclid Collaboration et al. Euclid preparation. I. The Euclid Wide Survey, *A&A*, 2022, vol. 662, p. A112

- Evrard A. E., The intracluster gas fraction in X-ray clusters: constraints on the clustered mass density, *MNRAS*, 1997, vol. 292, p. 289
- Fan X., Bahcall N. A., Cen R., Determining the Amplitude of Mass Fluctuations in the Universe, *ApJ*, 1997, vol. 490, p. L123
- Farrens S., Abdalla F. B., Cypriano E. S., Sabiu C., Blake C., Friends-of-friends groups and clusters in the 2SLAQ catalogue, *MNRAS*, 2011, vol. 417, p. 1402
- A. Finoguenov et al. CODEX clusters. Survey, catalog, and cosmology of the X-ray luminosity function, *A&A*, 2020, vol. 638, p. A114
- Firth A. E., Lahav O., Somerville R. S., Estimating photometric redshifts with artificial neural networks, *MNRAS*, 2003, vol. 339, p. 1195
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., The Sloan Digital Sky Survey Photometric System, *AJ*, 1996, vol. 111, p. 1748
- Gaia Collaboration et al. Gaia Data Release 2. Summary of the contents and survey properties, *A&A*, 2018, vol. 616, p. A1
- Gal R. R., Optical Detection of Galaxy Clusters, arXiv e-prints, 2006, pp astro-ph/0601195
- Geach J. E., Murphy D. N. A., Bower R. G., 4098 galaxy clusters to $z \sim 0.6$ in the Sloan Digital Sky Survey equatorial Stripe 82, *MNRAS*, 2011, vol. 413, p. 3059
- Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin D., Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, 2013
- Gelman A., Carlin J. B., Stern H. S., Rubin D. B., Bayesian Data Analysis 2nd ed. edn. Chapman and Hall/CRC, 2004
- Geman S., Geman D., Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, vol. PAMI-6, p. 721
- Matthew R. George et al. Galaxies in X-Ray Groups. I. Robust Membership Assignment and the Impact of Group Environments on Quenching, *ApJ*, 2011, vol. 742, p. 125

- Giacconi R., Kellogg E., Gorenstein P., Gursky H., Tananbaum H., An X-Ray Scan of the Galactic Plane from UHURU, *ApJ*, 1971, vol. 165, p. L27
- Carlo Giocoli et al. AMICO galaxy clusters in KiDS-DR3. Cosmological constraints from large-scale stacked weak lensing profiles, *A&A*, 2021, vol. 653, p. A19
- Gladders M. D., Yee H. K. C., The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29, *ApJS*, 2005, vol. 157, p. 1
- Jesse B. Golden-Marx et al. The Observed Evolution of the Stellar Mass–Halo Mass Relation for Brightest Central Galaxies, *The Astrophysical Journal*, 2022, vol. 928, p. 28
- Gondoin P., van Katwijk K., Aschenbach B. R., Schulz N., Boerret R., Glatzel H., Citterio O., X-ray spectroscopy mission (XMM) telescope development. In *Space Optics 1994: Earth Observation and Astronomy*, vol. 2209 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1994, p. 438
- Gonzalez A., Cluster detection via wavelets. In *Building the Euclid Cluster Survey - Scientific Program*, 2014, p. 7
- Gonzalez R., Woods R., *Digital Image Processing*. Prentice Hall, 2008
- R. M. González Delgado et al. The miniJPAS survey. Identification and characterization of galaxy populations with the J-PAS photometric system, *A&A*, 2021, vol. 649, p. A79
- R. M. González Delgado et al. The miniJPAS survey. The role of group environment in quenching star formation, *A&A*, 2022, vol. 666, p. A84
- Gozaliasl G., Finoguenov A., Khosroshahi H. G., Henriques B. M. B., Tanaka M., Ilbert O., Wuyts S., McCracken H. J., Montanari F., Brightest group galaxies - II: the relative contribution of BGGs to the total baryon content of groups at $z \lesssim 1.3$, *MNRAS*, 2018, vol. 475, p. 2787
- Gunn J. E., Gott J. Richard I., On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution, *ApJ*, 1972, vol. 176, p. 1
- Gursky H., Kellogg E., Murray S., Leong C., Tananbaum H., Giacconi R., A Strong X-Ray Source in the Coma Cluster Observed by UHURU, *ApJ*, 1971, vol. 167, p. L81

- Guth A. H., Inflationary universe: A possible solution to the horizon and flatness problems, *Phys. Rev. D*, 1981, vol. 23, p. 347
- Hastings W. K., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 1970, vol. 57, p. 97
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses, *MNRAS*, 2015, vol. 451, p. 2663
- A. Hernán-Caballero et al. The miniJPAS survey: Photometric redshift catalogue, *A&A*, 2021, vol. 654, p. A101
- M. Hilton et al. The Atacama Cosmology Telescope: A Catalog of 4000 Sunyaev–Zel’dovich Galaxy Clusters, *The Astrophysical Journal Supplement Series*, 2021, vol. 253, p. 3
- G. Hinshaw et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results, *ApJS*, 2013, vol. 208, p. 19
- Hoekstra H., Herbonnet R., Muzzin A., Babul A., Mahdavi A., Viola M., Cacciato M., The Canadian Cluster Comparison Project: detailed study of systematics and updated weak lensing masses, *MNRAS*, 2015, vol. 449, p. 685
- Hsieh B. C., Yee H. K. C., Lin H., Gladders M. D., A Photometric Redshift Galaxy Catalog from the Red-Sequence Cluster Survey, *ApJS*, 2005, vol. 158, p. 161
- Huang S., Leauthaud A., Bradshaw C., Hearin A., Behroozi P., Lange J., Greene J., DeRose J., Speagle J. S., Xhakaj E., The outer stellar mass of massive galaxies: a simple tracer of halo mass with scatter comparable to richness and reduced projection effects, *MNRAS*, 2022, vol. 515, p. 4722
- Ider Chitham J., Comparat J., Finoguenov A., Clerc N., Kirkpatrick C., Damsted S., Kukkola A., Capasso R., Nandra K., Merloni A., Bulbul E., Rykoff E. S., Schneider D. P., Brownstein J. R., Cosmological constraints from CODEX galaxy clusters spectroscopically confirmed by SDSS-IV/SPIDERS DR16, *MNRAS*, 2020, vol. 499, p. 4768

-
- Željko Ivezić et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products, *ApJ*, 2019, vol. 873, p. 111
- Johnston D. E., Sheldon E. S., Wechsler R. H., Rozo E., Koester B. P., Frieman J. A., McKay T. A., Evrard A. E., Becker M. R., Annis J., Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation, *arXiv e-prints*, 2007, p. arXiv:0709.1159
- Jones C., Forman W., The structure of clusters of galaxies observed with Einstein, *ApJ*, 1984, vol. 276, p. 38
- Jones M. H., Lambourne R. J. A., *An Introduction to Galaxies and Cosmology*, 2004
- Kelly B. C., Some Aspects of Measurement Error in Linear Regression of Astronomical Data, *ApJ*, 2007, vol. 665, p. 1489
- M Klein et al. A multicomponent matched filter cluster confirmation tool for eROSITA: initial application to the RASS and DES-SV data sets, *Monthly Notices of the Royal Astronomical Society*, 2017a, vol. 474, p. 3324
- M Klein et al. A multicomponent matched filter cluster confirmation tool for eROSITA: initial application to the RASS and DES-SV data sets, *Monthly Notices of the Royal Astronomical Society*, 2017b, vol. 474, p. 3324
- Lang D., unWISE: UNBLURRED COADDS OF THE WISE IMAGING, *The Astronomical Journal*, 2014, vol. 147, p. 108
- Lavery R. J., Henry J. P., Two blue arcs associated with the cD galaxy of Abell 963, *ApJ*, 1988, vol. 329, p. L21
- O. Le Fèvre et al. The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with $17.5 \leq IAB \leq 24$, and the redshift distribution over $0 \leq z \leq 5$, *A&A*, 2005, vol. 439, p. 845
- Alexie Leauthaud et al. A Weak Lensing Study of X-ray Groups in the Cosmos Survey: Form and Evolution of the Mass-Luminosity Relation, *ApJ*, 2010, vol. 709, p. 97

- Leauthaud A., George M. R., Behroozi P. S., Bundy K., Tinker J., Wechsler R. H., Conroy C., Finoguenov A., Tanaka M., The Integrated Stellar Content of Dark Matter Halos, *ApJ*, 2012b, vol. 746, p. 95
- Alexie Leauthaud et al. Lensing is low: cosmology, galaxy formation or new physics?, *MNRAS*, 2017, vol. 467, p. 3024
- Alexie Leauthaud et al. New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from $z = 0.2$ to $z = 1$, *ApJ*, 2012a, vol. 744, p. 159
- G. F. Lesci et al. AMICO galaxy clusters in KiDS-DR3: Cosmological constraints from counts and stacked weak lensing, *A&A*, 2022a, vol. 659, p. A88
- G. F. Lesci et al. AMICO galaxy clusters in KiDS-DR3: Constraints on cosmological parameters and on the normalisation of the mass-richness relation from clustering, *A&A*, 2022b, vol. 665, p. A100
- Li C., White S. D. M., The distribution of stellar mass in the low-redshift Universe, *Monthly Notices of the Royal Astronomical Society*, 2009, vol. 398, p. 2177
- Licitra R., Mei S., Raichoor A., Erben T., Hildebrandt H., The RedGOLD cluster detection algorithm and its cluster candidate catalogue for the CFHT-LS W1, *MNRAS*, 2016, vol. 455, p. 3020
- Liddle A. R., Lyth D. H., The cold dark matter density perturbation, *Phys. Rep.*, 1993, vol. 231, p. 1
- E. V. R. Lima et al. Photometric redshifts for the S-PLUS Survey: Is machine learning up to the task?, *Astronomy and Computing*, 2022, vol. 38, p. 100510
- Lima-Neto G. B., *Astronomia Extragaláctica e Cosmologia, Notas de Aula*, 2022
- Linde A. D., A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems, *Physics Letters B*, 1982, vol. 108, p. 389

-
- Lopes P. A. A., de Carvalho R. R., Gal R. R., Djorgovski S. G., Odewahn S. C., Mahabal A. A., Brunner R. J., The Northern Sky Optical Cluster Survey. IV. An Intermediate-Redshift Galaxy Cluster Catalog and the Comparison of Two Detection Algorithms, *AJ*, 2004, vol. 128, p. 1017
- Lopes P. A. A., de Carvalho R. R., Kohl-Moreira J. L., Jones C., NoSOCS in SDSS - II. Mass calibration of low redshift galaxy clusters with optical and X-ray properties, *MNRAS*, 2009, vol. 399, p. 2201
- Lopes P. A. A., Ribeiro A. L. B., Reliable photometric membership (RPM) of galaxies in clusters - I. A machine learning method and its performance in the local universe, *MNRAS*, 2020, vol. 493, p. 3429
- C. López-Sanjuan et al. J-PLUS: Morphological star/galaxy classification by PDF analysis, *A&A*, 2019, vol. 622, p. A177
- Lynds R., Petrosian V., Giant Luminous Arcs in Galaxy Clusters. In *Bulletin of the American Astronomical Society*, vol. 18 of BAAS, 1986, p. 1014
- Mandelbaum R., Weak Lensing for Precision Cosmology, *ARA&A*, 2018, vol. 56, p. 393
- Maturi M., Bellagamba F., Radovich M., Roncarelli M., Sereno M., Moscardini L., Bardelli S., Puddu E., AMICO galaxy clusters in KiDS-DR3: sample properties and selection function, *MNRAS*, 2019, vol. 485, p. 498
- M. Maturi et al. The miniJPAS survey: clusters and galaxy groups detection with AMICO, *arXiv e-prints*, 2023, p. arXiv:2307.06412
- T. McClintock et al. Dark Energy Survey Year 1 results: weak lensing mass calibration of redMaPPer galaxy clusters, *MNRAS*, 2019, vol. 482, p. 1352
- Elinor Medezinski et al. Source selection for cluster weak lensing measurements in the Hyper Suprime-Cam survey, *PASJ*, 2018, vol. 70, p. 30
- P. Melchior et al. Weak-lensing mass calibration of redMaPPer galaxy clusters in Dark Energy Survey Science Verification data, *MNRAS*, 2017, vol. 469, p. 4899

- C. Mendes de Oliveira et al. The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters, *MNRAS*, 2019, vol. 489, p. 241
- A. Merloni et al. eROSITA Science Book: Mapping the Structure of the Energetic Universe, arXiv e-prints, 2012, p. arXiv:1209.3114
- J. Merten et al. Creation of cosmic structure in the complex galaxy cluster merger Abell 2744, *MNRAS*, 2011, vol. 417, p. 333
- J. Merten et al. CLASH: The Concentration-Mass Relation of Galaxy Clusters, *ApJ*, 2015, vol. 806, p. 4
- D. G. Michael et al. Observation of Muon Neutrino Disappearance with the MINOS Detectors in the NuMI Neutrino Beam, *Phys. Rev. Lett.*, 2006, vol. 97, p. 191801
- Miralda-Escude J., Gravitational Lensing by Clusters of Galaxies: Constraining the Mass Distribution, *ApJ*, 1991, vol. 370, p. 1
- Satoshi Miyazaki et al. Hyper Suprime-Cam. In Ground-based and Airborne Instrumentation for Astronomy IV , vol. 8446 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2012, p. 84460Z
- A. Molino et al. Assessing the photometric redshift precision of the S-PLUS survey: the Stripe-82 as a test-case, *MNRAS*, 2020, vol. 499, p. 3884
- L. Nakazono et al. On the discovery of stars, quasars, and galaxies in the Southern Hemisphere with S-PLUS DR2, *MNRAS*, 2021, vol. 507, p. 5847
- Newman J. A., Gruen D., Photometric Redshifts for Next-Generation Surveys, *ARA&A*, 2022, vol. 60, p. 363
- Olave-Rojas D., Cerulo P., Demarco R., Jaffé Y. L., Mercurio A., Rosati P., Balestra I., Nonino M., Galaxy pre-processing in substructures around $z \sim 0.4$ galaxy clusters, *MNRAS*, 2018, vol. 479, p. 2328
- F. Pacaud et al. The XXL Survey. II. The bright cluster sample: catalogue and luminosity function, *A&A*, 2016, vol. 592, p. A2

- A. Palmese et al. A Statistical Standard Siren Measurement of the Hubble Constant from the LIGO/Virgo Gravitational Wave Compact Object Merger GW190814 and Dark Energy Survey Galaxies, *ApJ*, 2020, vol. 900, p. L33
- Carolina Parroni et al. Next Generation Virgo Cluster Survey. XXI. The Weak Lensing Masses of the CFHTLS and NGVS RedGOLD Galaxy Clusters and Calibration of the Optical Richness, *ApJ*, 2017, vol. 848, p. 114
- Pearson R. J., Ponman T. J., Norberg P., Robotham A. S. G., Farr W. M., On optical mass estimation methods for galaxy groups, *MNRAS*, 2015, vol. 449, p. 3082
- Planck Collaboration et al. Planck 2013 results. XVI. Cosmological parameters, *A&A*, 2014, vol. 571, p. A16
- Planck Collaboration et al. Planck 2015 results. XIII. Cosmological parameters, *A&A*, 2016a, vol. 594, p. A13
- Planck Collaboration et al. Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts, *A&A*, 2016b, vol. 594, p. A24
- Planck Collaboration et al. Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters, *A&A*, 2016c, vol. 594, p. A11
- Pratt C. T., Bregman J. N., SZ Scaling Relations of Galaxy Groups and Clusters Near the North Ecliptic Pole, *ApJ*, 2020, vol. 890, p. 156
- Pratt G. W., Arnaud M., Biviano A., Eckert D., Ettori S., Nagai D., Okabe N., Reiprich T. H., The Galaxy Cluster Mass Scale and Its Impact on Cosmological Constraints from the Cluster Population, *Space Sci. Rev.*, 2019, vol. 215, p. 25
- P. Predehl et al. The eROSITA X-ray telescope on SRG, *A&A*, 2021, vol. 647, p. A1
- Ramella M., Boschin W., Fadda D., Nonino M., Finding galaxy clusters using Voronoi tessellations, *A&A*, 2001, vol. 368, p. 776
- Reiprich T. H., Böhringer H., The Mass Function of an X-Ray Flux-limited Sample of Galaxy Clusters, *ApJ*, 2002, vol. 567, p. 716

- J. E. Rodríguez-Martín et al. The miniJPAS survey. Galaxy populations in the most massive cluster in miniJPAS: mJPC2470-1771, *A&A*, 2022, vol. 666, p. A160
- Rozo E., Wechsler R. H., Rykoff E. S., Annis J. T., Becker M. R., Evrard A. E., Frieman J. A., Hansen S. M., Hao J., Johnston D. E., Koester B. P., McKay T. A., Sheldon E. S., Weinberg D. H., Cosmological Constraints from the Sloan Digital Sky Survey maxBCG Cluster Catalog, *ApJ*, 2010, vol. 708, p. 645
- Rykoff E. S., Rozo E., Busha M. T., Cunha C. E., Finoguenov A., Evrard A., Hao J., Koester B. P., Leauthaud A., Nord B., Pierre M., Reddick R., Sadibekova T., Sheldon E. S., Wechsler R. H., redMaPPer. I. Algorithm and SDSS DR8 Catalog, *ApJ*, 2014, vol. 785, p. 104
- E. S. Rykoff et al. The RedMaPPer Galaxy Cluster Catalog From DES Science Verification Data, *ApJS*, 2016, vol. 224, p. 1
- H Sampaio-Santos et al. Is diffuse intracluster light a good tracer of the galaxy cluster matter distribution?, *Monthly Notices of the Royal Astronomical Society*, 2020, vol. 501, p. 1300
- Sarazin C. L., X-ray emission from clusters of galaxies, *Reviews of Modern Physics*, 1986, vol. 58, p. 1
- Sartoris B., Biviano A., Fedeli C., Bartlett J. G., Borgani S., Costanzi M., Giocoli C., Moscardini L., Weller J., Ascaso B., Bardelli S., Maurogordato S., Viana P. T. P., Next generation cosmology: constraints from the Euclid galaxy cluster survey, *MNRAS*, 2016, vol. 459, p. 1764
- Satopaa V., Albrecht J., Irwin D., Raghavan B., Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior. In 2011 31st International Conference on Distributed Computing Systems Workshops , 2011, p. 166
- Schneider P., *Extragalactic Astronomy and Cosmology*, 2006
- Sereno M., Ettori S., Eckert D., Giles P., Maughan B. J., Pacaud F., Pierre M., Valageas P., The XXL Survey. XXXVIII. Scatter and correlations of X-ray proxies in the bright XXL cluster sample, *A&A*, 2019, vol. 632, p. A54

- Shan H. Y., Kneib J.-P., Comparat J., Jullo E., Charbonnier A., Erben T., Makler M., Moraes B., Van Waerbeke L., Courbin F., Meylan G., Tao C., Taylor J. E., Weak lensing mass map and peak statistics in Canada-France-Hawaii Telescope Stripe 82 survey, *MNRAS*, 2014, vol. 442, p. 2534
- M. F. Skrutskie et al. The Two Micron All Sky Survey (2MASS), *AJ*, 2006, vol. 131, p. 1163
- Smith S., The Mass of the Virgo Cluster, *ApJ*, 1936, vol. 83, p. 23
- Sommer M. W., Schrabback T., Applegate D. E., Hilbert S., Ansarinejad B., Floyd B., Grandis S., Weak lensing mass modeling bias and the impact of miscentring, *MNRAS*, 2022, vol. 509, p. 1127
- Soucail G., Fort B., Mellier Y., Picat J. P., A blue ring-like structure, in the center of the A 370 cluster of galaxies, *A&A*, 1987, vol. 172, p. L14
- V. Springel et al. Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature*, 2005, vol. 435, p. 629
- Storopoli J., , 2021 *Estatística Bayesiana com R e Stan: Markov Chain Monte Carlo – MCMC*
- Suto Y., Sasaki S., Makino N., Gas Density and X-Ray Surface Brightness Profiles of Clusters of Galaxies from Dark Matter Halo Potentials: Beyond the Isothermal B-Model, *The Astrophysical Journal*, 1998, vol. 509, p. 544
- Takey A., Durret F., Mahmoud E., Ali G. B., The 3XMM/SDSS Stripe 82 Galaxy Cluster Survey. I. Cluster catalogue and discovery of two merging cluster candidates, *A&A*, 2016, vol. 594, p. A32
- Takey A., Durret F., Márquez I., Ellien A., Molham M., Plat A., The 3XMM/SDSS Stripe 82 Galaxy Cluster Survey – II. X-ray and optical properties of the cluster sample, *Monthly Notices of the Royal Astronomical Society*, 2019, vol. 486, p. 4863
- Tempel E., Tenjes P., Line-of-sight velocity dispersions and a mass-distribution model of the Sa galaxy NGC 4594, *Monthly Notices of the Royal Astronomical Society*, 2006, vol. 371, p. 1269

- Tempel E., Tuvikene T., Kipper R., Libeskind N. I., Merging groups and clusters of galaxies from the SDSS data. The catalogue of groups and potentially merging systems, *A&A*, 2017, vol. 602, p. A100
- The Dark Energy Survey Collaboration The Dark Energy Survey, arXiv e-prints, 2005, pp astro-ph/0510346
- Vikhlinin A., Kravtsov A. V., Burenin R. A., Ebeling H., Forman W. R., Hornstrup A., Jones C., Murray S. S., Nagai D., Quintana H., Voevodkin A., Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints, *ApJ*, 2009, vol. 692, p. 1060
- Vitorelli A. Z., Cypriano E. S., Makler M., Pereira M. E. S., Erben T., Moraes B., On mass concentrations and magnitude gaps of galaxy systems in the CS82 survey, *MNRAS*, 2018, vol. 474, p. 866
- W. Voges et al. The ROSAT all-sky survey bright source catalogue, *A&A*, 1999, vol. 349, p. 389
- Voit G. M., Tracing cosmic evolution with clusters of galaxies, *Rev. Mod. Phys.*, 2005, vol. 77, p. 207
- Wagoner R. V., Fowler W. A., Hoyle F., On the Synthesis of Elements at Very High Temperatures, *ApJ*, 1967, vol. 148, p. 3
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., Observational probes of cosmic acceleration, *Phys. Rep.*, 2013, vol. 530, p. 87
- Weisskopf M. C., The Chandra X-Ray Observatory (CXO): An Overview, arXiv e-prints, 1999, pp astro-ph/9912097
- Wen Z. L., Han J. L., Calibration of the Optical Mass Proxy for Clusters of Galaxies and an Update of the WHL12 Cluster Catalog, *ApJ*, 2015, vol. 807, p. 178
- Wen Z. L., Han J. L., Liu F. S., A Catalog of 132,684 Clusters of Galaxies Identified from Sloan Digital Sky Survey III, *ApJS*, 2012, vol. 199, p. 34
- S. V. Werner et al. S-PLUS DR1 galaxy clusters and groups catalogue using PzWav, arXiv e-prints, 2022, p. arXiv:2211.04243

-
- White S. D. M., Frenk C. S., Galaxy formation through hierarchical clustering, *ApJ*, 1991, vol. 379, p. 52
- White S. D. M., Rees M. J., Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering., *MNRAS*, 1978, vol. 183, p. 341
- Willmer C. N. A., The Absolute Magnitude of the Sun in Several Filters, *The Astrophysical Journal Supplement Series*, 2018, vol. 236, p. 47
- Edward L. Wright et al. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance, *AJ*, 2010, vol. 140, p. 1868
- Zwicky F., On the Masses of Nebulae and of Clusters of Nebulae, *ApJ*, 1937, vol. 86, p. 217

Apêndice

Diferença de Gaussianas

O método de Diferença de Gaussianas (DoG) é uma técnica utilizada em processamento de imagem e visão computacional para detecção de características ou objetos em uma imagem. Abaixo, descrevemos o processo de maneira superficial. Para uma visão geral de filtragem e diferentes métodos ver Gonzalez e Woods (2008).

A ideia principal do método é identificar áreas de uma imagem onde ocorrem mudanças significativas de intensidade, indicativas de bordas, cantos ou outros padrões. Isso é feito por meio da aplicação de operações de suavização e subtração. O passo a passo é destacado a seguir:

- Suavização 1: A imagem original é suavizada utilizando um filtro Gaussiano de desvio padrão σ_1 . O filtro Gaussiano é uma função matemática que realiza uma média ponderada dos valores dos pixels na vizinhança de cada pixel, atribuindo maior peso aos pixels mais próximos ao pixel central.
- Suavização 2: A imagem original é novamente borrada com um filtro Gaussiano de um tamanho σ_2 diferente.
- Subtração: A imagem resultante é a subtração da “Suavização 1” e “Suavização 2”, pixel a pixel. Esse processo destaca as regiões onde ocorreram mudanças significativas de intensidade entre as duas etapas de suavização.

Considerando a utilização de dois σ diferentes, o filtro DoG é visto como um filtro “passa-banda”, isto é, apenas pixels dentro de um intervalo são considerados, resultando na remoção de ruídos.

A Figura A.1 mostra uma composição de 3 sobredensidades circulares (2 em sobreposição) adicionados a um ruído gaussiano normalizado (painel superior esquerdo), e o

resultado após a aplicação do filtro DoG (painel superior direito), os pontos vermelhos mostram o pico das distribuições obtidas pela função “`peak_local_max`” biblioteca “`scikit-image`”, com distância mínima entre picos de 10, em unidades da imagem. As duas imagens inferiores representam a imagem suavizada por dois filtros gaussianos diferentes, com $\sigma = 7$ e $\sigma = 15$, respectivamente. Nestas, são destacadas as diferenças produzidas por cada filtro: em $\sigma = 7$, a contribuição da distribuição maior é reduzida; enquanto para $\sigma = 15$ as estruturas são completamente borradas, criando apenas 1 distribuição.

Considerando a aplicação desse método no PZwav, é necessário converter a entrada de tamanho físico dos aglomerados (400 e 1400 kpc), máximos e mínimos em unidades de pixel. Isso é feito ao transformar a unidade para segundos de arco e então dividir pela escala de pixel, definido por `pscale = 1/pix_per_deg`, sendo `pix_per_deg=300`.

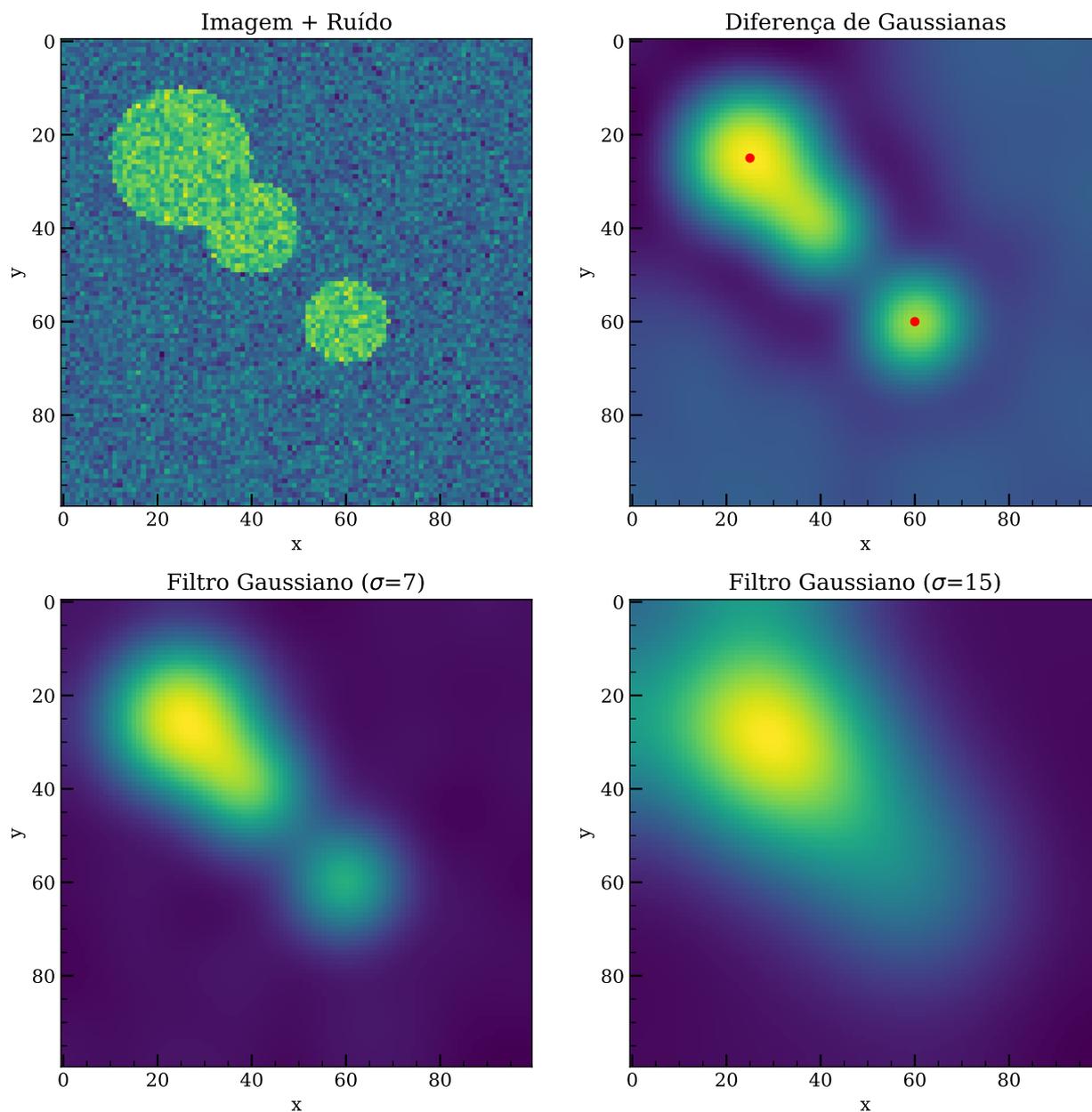


Figura A.1: Exemplo da aplicação do filtro de diferença de gaussianas (DoG). Painel superior esquerdo: Composição de 3 sobredensidades circulares adicionados a um ruído gaussiano. Painel superior direito: Resultado da aplicação do filtro DoG e a detecção dos picos utilizando a função `peak_local_max` do “scikit-image”. Painéis inferiores: da esquerda para direita, filtros gaussianos com $\sigma = 7$ e $\sigma = 15$.

HDBSCAN

Utilizamos HDBSCAN devido sua habilidade de conectar uma distribuição de pontos de forma hierárquica para extrair possíveis aglomerações, uma característica interessante na detecção de galáxias membro.

HDBSCAN é um algoritmo de sobredensidades desenvolvido por Campello, Moulavi e Sander em 2014 (Campello et al., 2013). Para maiores informações sobre instalação e aplicações diretas do código, visitar a página de documentação: https://hdbscan.readthedocs.io/en/latest/basic_hdbscan.html (último acesso em 08/01/2024).

Abaixo apresentaremos em mais detalhes o funcionamento do código, baseando-nos na página “*How HDBSCAN Works*” do link. A estrutura se baseia em 5 passos principais:

- Transformação do espaço com base na densidade de pontos
- Construção da árvore de distâncias mínimas
- Ligação hierárquica dos componentes
- Concentração da hierarquia com base no tamanho mínimo `min_samples`
- Extração das aglomerações

B.1 Transformação do espaço com base na densidade de pontos

Para a identificação dos agrupamentos se faz necessário reconhecer os pontos de maior e menor densidade na distribuição, e a presença de ruído. Dados reais podem ser bagunçados, com a presença de valores discrepantes, vazios e ruído. Na identificação de sobredensidades ruído é importante de ser considerado uma vez que um único ponto erroneamente presente entre duas distribuições pode ligá-las.

Para realizar essa separação de forma otimizada, utilizamos a distância entre um ponto e seu k -ésimo vizinhos mais próximos como um parâmetro de densidade. Para formar uma árvore kd, do inglês “kd-tree”, definimos um ponto central x , e calculamos a distância entre este e os vizinhos, denominando $core_k(x)$. Dessa forma a distância de “acessibilidade” entre dois pontos a e b pode ser descrita como,

$$d_{mreach-k}(a, b) = \max[core_k(a), core_k(b), d(a, b)] \quad (\text{B.1})$$

sendo $d(a, b)$ a distância métrica inicial. Ao considerar essa descrição, os pontos de menor distância ao centro (maior densidade) permanecem com as mesmas distâncias. Os mais esparsos são penalizados e afastados para estarem a pelo menos 1 distância ao centro ($core_k(x)$) de todos os outros pontos.

Um dos pontos negativos dessa abordagem é sua dependência com k : valores maiores de k tendem a identificar mais pontos como de menor densidade e, portanto, são mais penalizados.

Figura B.1 exemplifica 3 distâncias ao centro diferentes (círculos azul, vermelho e verde) entre 6 vizinhos ($k = 5$). Os pontos centrais azul-verde estão ligados pelo círculo verde, mas não para o azul, assim essa ligação seria marcada pelo raio do círculo verde. Em comparação, a distância dos pontos vermelho-verde é maior que os raios, esse é marcado como a distância simples entre os 2.

Segundo os autores do texto, e mencionando o estudo de Eldridge et al. (2015), a distância de acessibilidade mútua como uma transformação do espaço funciona bem para descrever a hierarquia da distribuição de densidades de qualquer amostra.

B.2 Construção da árvore de distâncias mínimas

Com essa nova métrica, se torna possível buscar por diferentes concentrações. Conceitualmente isso é feito ao considerar os dados como pontos em um gráfico (como por exemplo as galáxias de um aglomerado) ligados por arestas que representam a distância de acessibilidade mútua (distância espacial entre as galáxias). Queremos que cada ponto seja ligado por apenas uma aresta, representada pelo valor de mínima distância, assim como apresentado na Figura B.2. Por fim, teremos uma hierarquia de componentes conectados ou desconectados em diferentes limites.

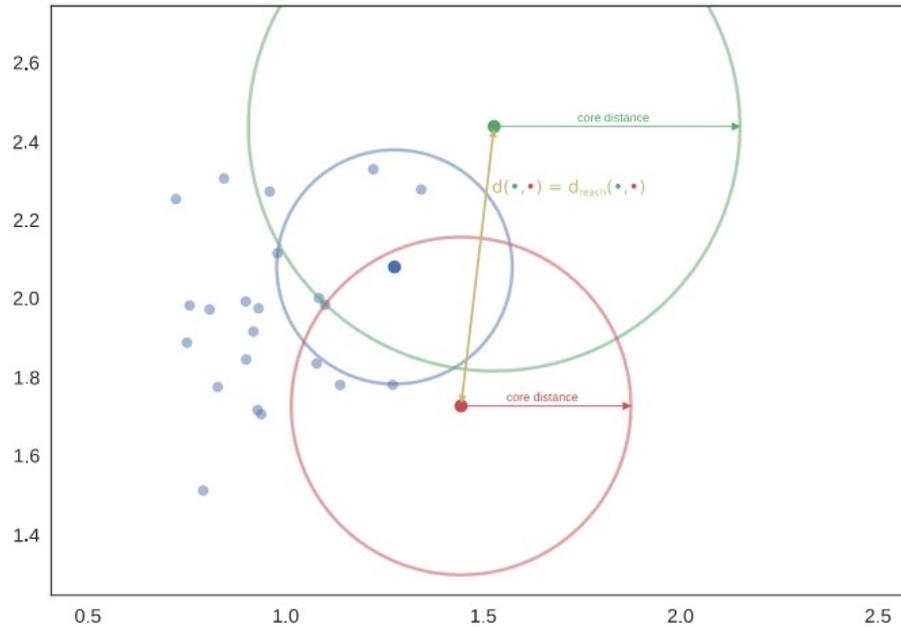


Figura B.1: Exemplos de diferentes pontos centrais, e a distância do 6 ponto mais distante como raio do círculo ($k = 5$). Figura retirada do site *readthedocs*.

Uma forma otimizada de realizar esse procedimento é utilizando o algoritmo de Prim (ou também conhecido como algoritmo de Jarník's na ciência da computação), este permite construir a árvore uma aresta por vez, adicionando a aresta de menor peso que conecta a árvore atual a um vértice que ainda não está na árvore.

B.3 Ligação hierárquica dos componentes

Uma vez que temos a árvore de distâncias, convertemos a informação na forma de hierarquia de componentes conectados. Para isso, cada concentração de pontos tem suas distâncias ordenadas de forma crescente. Cada maior distância é então conectada a outra, formando um dendrograma. A escolha de cada união é feita pela estrutura de dados *union-find* (também conhecido como *disjoint-set* em ciências da computação).

Uma forma de selecionar quais dados estão conectados é traçar uma linha horizontal que corta esse dendrograma. As aglomerações são selecionados pela linha que os corta, definindo então uma distância máxima de acessibilidade mútua.

Uma das dificuldades nesse processo é selecionar a linha de corte. Em DBSCAN, esse valor é definido pelo usuário como um parâmetro fixo.

O próximo passo aborda como HDBSCAN permite cortar o dendrograma em diferentes

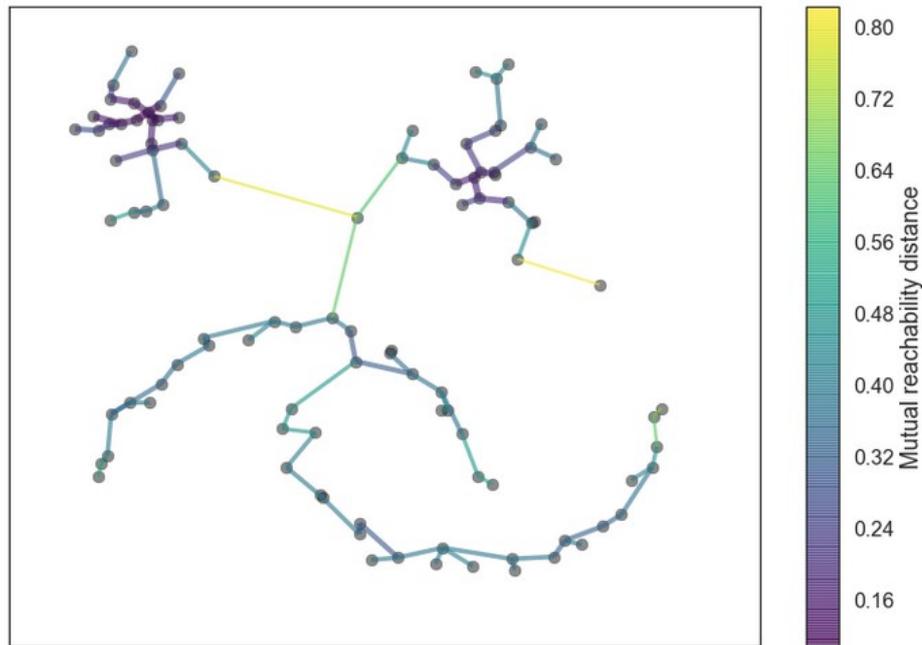


Figura B.2: Exemplos de árvore de distâncias para $k = 5$. Figura retirada da documentação do código em *readthedocs*.

valores, aumentando a robustez do processo, e permitindo a identificação de aglomerações de diferentes distâncias máximas, isto é, de densidades variáveis.

B.4 Concentração da hierarquia com base no tamanho mínimo *min_samples*

A ideia principal é condensar a hierarquia em um dendrograma menor, porém mantendo as informações sobre como o tamanho da aglomeração muda em função da distância.

Para isso, defini-se um parâmetro de tamanho mínimo "min_samples" da aglomeração e, a cada divisão da hierarquia, considera-se o número de pontos restantes:

- **Número de pontos menor do que o tamanho mínimo:** a aglomeração maior retem a identidade da aglomeração inicial, anotando quais pontos "saíram da distribuição" e em que valor de distância isso aconteceu.
- **Número de pontos maior do que o tamanho mínimo:** considera-se como uma verdadeira divisão, mantendo a divisão na árvore.

B.5 Extração das aglomerações

Finalmente, o processo de escolha e identificação das aglomerações se dá pela sua estabilidade na árvore hierárquica. Figura B.3 demonstra o dendrograma de distâncias após a concentração da hierarquia em função do tamanho mínimo. Cada distribuição circulada representa uma distribuição estável, que foi demarcada como aglomerada. O círculo vermelho demonstra uma das condições das seleções: ao definir uma aglomeração, não se pode escolher nenhum de seus descendentes.

Para definir estabilidade, primeiro precisamos apresentar um outro parâmetro λ , que é inversamente proporcional à distância. Cada aglomeração terá seus $\lambda_{\text{nascimento}}$ e λ_{morte} , representando o momento em que a aglomeração se separa do braço principal e quando se divide em aglomerações menores, respectivamente. A estabilidade é então calculada como a soma dos valores $\lambda_p - \lambda_{\text{nascimento}}$ para cada ponto p na árvore de hierarquia.

Assim, o algoritmo seleciona as aglomerações de forma ascendente na árvore (ver Figura B.3), comparando a estabilidade da aglomeração com a soma das estabilidades de seus braços de divisão. Se a soma das estabilidades dos braços for maior do que a estabilidade da aglomeração, então consideramos as divisões como distintas aglomerações (por exemplo, os círculos azul e verde na figura). Se, por outro lado, a soma da estabilidade da aglomeração for maior do que a soma dos braços, então consideramos a aglomeração e desmarcamos as divisões (por exemplo, o círculo vermelho na figura).

Qualquer ponto que não esteja em uma aglomeração selecionada é marcado como ruído e recebe a identificação de -1 . A Figura B.4 mostra as aglomerações selecionadas, com as mesmas cores da figura anterior, e em cinza os pontos de ruído. Com esse passo a passo, conseguimos identificar aglomerações de diferentes densidades, apenas atribuindo um tamanho mínimo "min_samples".

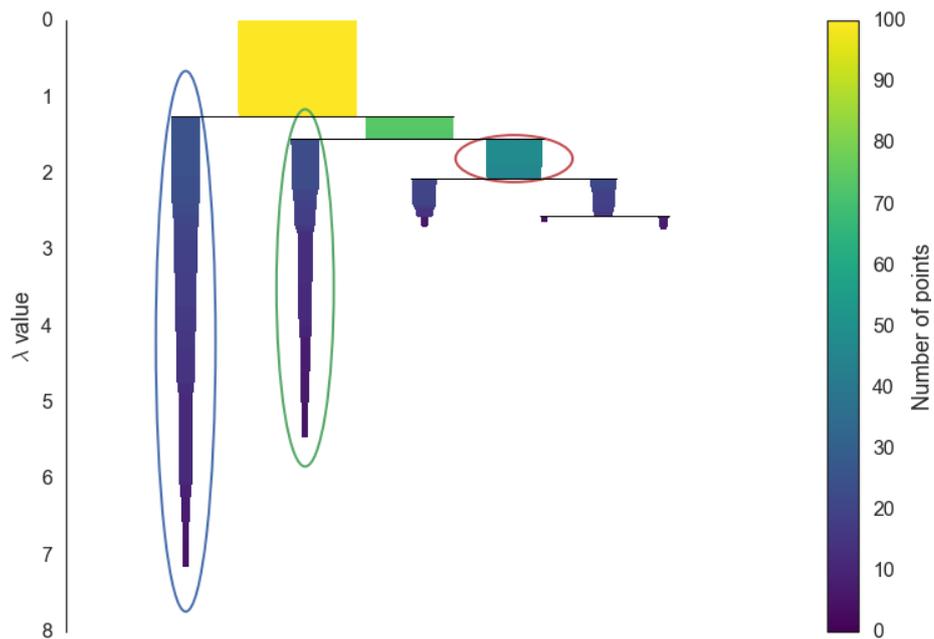


Figura B.3: Dendrograma do inverso da distâncias (λ) após a concentração da hierarquia em função do tamanho mínimo. Os braços circulados representam diferentes aglomerações selecionadas. Figura retirada da documentação do código em *readthedocs*.

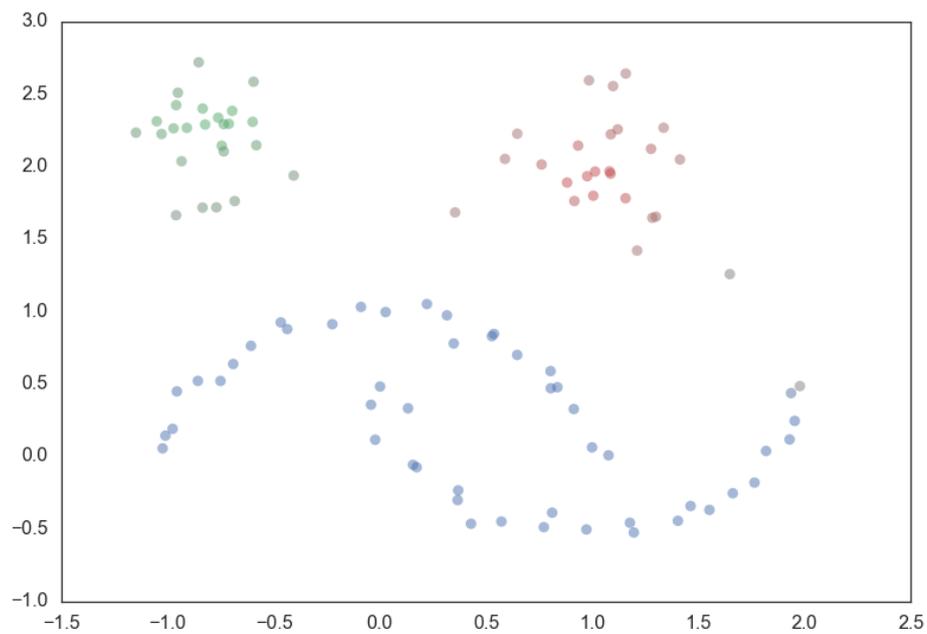


Figura B.4: Identificação visual das aglomerações demarcadas na Figura B.3, com o mesmo código de cores. Figura retirada da documentação do código em *readthedocs*.

Apêndice C

Linmix

Linmix é um algoritmo de regressão linear baseado no trabalho Kelly (2007). Apesar do artigo descrever modelos uni e multivariados, a implementação se dá apenas através do modelo univariado. Em <https://linmix.readthedocs.io/en/latest/index.html> (último acesso em 23/01/2024) os autores providenciam uma documentação que apresenta o processo de instalação e exemplos básicos de utilização.

Aqui, utilizamos a seção *linmix Math* do *site* citado acima para descrever a matemática utilizada pelo algoritmo.

Inicialmente, descrevem-se as variáveis independente (x) e dependente (y) como uma distribuição gaussiana bidimensional (N_2), com média $\mu = (\xi, \eta)$ e matriz de covariância $\Sigma = (\sigma_x^2, \sigma_{xy}, \sigma_{xy}, \sigma_y^2)$. Nessa notação, σ_x e σ_y representam os erros gaussianos de $1-\sigma$ nas variáveis x e y , e o termo cruzado σ_{xy} sua covariância.

A relação entre as variáveis verdadeiras não observadas dependente e independente se descreve por,

$$\eta \sim N(\alpha + \beta\xi, \sigma^2) \tag{C.1}$$

sendo α a interceptação da linha de regressão linear e β a inclinação. A dispersão intrínseca gaussiana de η é representada por σ^2 , de média zero e variância `sigsqr`. Nesse contexto, as variáveis verdadeiras não observadas (ξ, η) representam os valores teóricos que estão por trás das observações sujeitas a erros.

Considerado o modelo Bayesiano, os priores dos parâmetros de regressão são dados como uniformes, isto é, com combinações igualmente prováveis dentro de determinados limites.

O modelo utilizado para descrever a variável independente é uma mistura de gaussianas, que representa uma forma flexível e é computacionalmente viável. Podemos descrever o

modelo por,

$$Pr(\xi|\pi, \mu, \tau^2) = \sum_{k=1}^K \frac{\pi_k}{\sqrt{2\pi\tau_k^2}} \exp\left(-\frac{1}{2} \left(\frac{\xi - \mu_k}{\tau_k}\right)^2\right) \quad (\text{C.2})$$

de forma que $Pr(\xi|\pi, \mu, \tau^2)$ representa a probabilidade condicional da soma ponderada de termos Gaussianos π , de médias μ e variância τ^2 .

Os parâmetros π , μ e τ apresentam um descrição adicional: $\pi \sim \text{Dirichlet}(1, \dots, 1)$; $\mu \sim \mathcal{N}(\mu_0, u^2)$, de forma que u^2 é limitado entre $[0, 1.5\text{Var}(x)]$; $\mu_0 \sim \text{U}(\min(x), \max(x))$; $u^2, \tau^2 \sim \mathcal{N}(0, w^2)$, sendo $w^2 \sim \text{U}(0, \infty)$.

A notação “N” representa uma distribuição normal (gaussiana), “U” é distribuição uniforme descrevendo uma distribuição de probabilidade contínua.

C.1 Funcionamento do código

Considerando o modelo proposto, o código escrito em `Python` calcula a regressão linear seguindo o passo a passo descrito a seguir:

1. **Inicialização:** As variáveis fornecidas pelo usuário (x , y , σ_x , σ_y) são configuradas, inicializando-se também os parâmetros da regressão como α , β (coeficientes linear e angular), `sigsqr` (espalhamento intrínseco), `mu0` (média da gaussiana), `usqr` (variância), `wsqr` (escala típica, `usqr-mu0`), entre outros.
2. **Execução do MCMC:** O espaço de parâmetros é explorado utilizando múltiplas cadeias MCMC de forma a buscar amostras que representem a distribuição posterior. Cada etapa consiste em atualizações dos parâmetros do modelo e o cálculo da condição de convergência RHAT (fator de escala potencial).
3. **Verificação de Convergência:** O código verifica se as condições de convergência são atendidas ($\text{RHAT} < 1.1$, assim como definido por Gelman et al., 2004). O processo é concluído ao atingir a convergência ou o número máximo de iterações.
4. **Armazenamento de Cadeias de MCMC:** Para evitar “burn-in”, apenas a última metade de cada cadeia MCMC é concatenada para formar a variável “chain”. Esta variável contém informações sobre todos os parâmetros do modelo.
5. **Resultados:** Os resultados finais são obtidos a partir de “chain”, e são acessados na forma de:

```
alpha, alpha_err = lm.chain['alpha'].mean(), lm.chain['alpha'].std()
```

da mesma forma para “beta” e “sigsqr”

Sendo `mean()` e `std()` funções que calculam a média e o desvio padrão das cadeias.

O código utiliza técnicas de *Markov Chain Monte Carlo* (MCMC), como a amostragem de Gibbs, para realizar amostragens iterativas das distribuições a posteriori dos parâmetros do modelo. O MCMC é uma ferramenta poderosa para explorar distribuições de probabilidade complexas, especialmente quando métodos analíticos não são viáveis. Alguns dos algoritmos comumente encontrados na literatura incluem o Metropolis-Hastings (Hastings, 1970), a Amostragem de Gibbs (Geman e Geman, 1984, ou *Gibbs sampler* em inglês), e o Monte Carlo Hamiltoniano (Duane et al., 1987).

Em resumo, o algoritmo de Metropolis-Hastings utiliza um passeio aleatório (do inglês, *random walk*) para gerar uma amostra aleatória de valores e, em seguida, aplica uma regra de aceitação ou rejeição para encaminhar a convergência em direção à distribuição-alvo. Devido à natureza do método, é necessário realizar um grande número de iterações para encontrar resultados mais prováveis, especialmente em espaços de alta dimensão, onde a exploração eficiente é desafiadora.

Para contornar a alta taxa de rejeição e possibilitar a exploração de espaços multidimensionais, Josiah Willard Gibbs desenvolveu o algoritmo de amostragem de Gibbs. Este método pode ser considerado como uma forma especializada do algoritmo de Metropolis-Hastings, onde a regra de rejeição é substituída por uma abordagem condicional. O algoritmo de Gibbs aproveita a probabilidade condicional dos outros parâmetros do modelo para gerar amostras do parâmetro desejado. Diferentemente do Metropolis-Hastings, o algoritmo de Gibbs não rejeita propostas, tornando-o particularmente eficiente em espaços de alta dimensionalidade.

Para uma descrição mais detalhada dos métodos e aplicações ver Storopoli (2021); Gelman et al. (2013); Betancourt (2017), e o site “Course Handouts for Bayesian Data Analysis Clas” https://bookdown.org/marklhc/notes_bookdown/ (último acesso 23/01/24).

Acesso aos códigos desenvolvidos e catálogos

Este apêndice fornece acesso aos códigos-fonte utilizados na implementação dos métodos discutidos neste trabalho, bem como aos catálogos resultantes das análises realizadas.

Os links abaixo direcionam para os repositórios online onde os códigos e os catálogos podem ser encontrados:

- *Adaptive Membership Estimator*, AME:

`https://github.com/liadoubrawa/Probabilistic_membership_estimators`

- Catálogos Resultantes:

S-PLUS: `https://github.com/liadoubrawa/splus_cluster_catalogues`

J-PAS: `https://github.com/liadoubrawa/Optical_detections_in_minijpas`

Para acessar o catálogo simulado apresentado nesta tese, entre em contato com Pablo Araya-Araya (e-mail: paraya-araya@usp.br), solicitando o conjunto de dados específico.

Para o catálogo de magnitudes absoluta e/ou massas estelar presentes nas análises relacionadas ao miniJPAS, entre em contato com Rosa Gonzalez Delgado (e-mail: rosa@iaa.es).

Artigos publicados

Abaixo, encontram-se as produções bibliográficas desenvolvidas ao longo do curso de doutorado, das quais: quatro artigos publicados resultantes da produção do AME, participação do projeto J-PAS e DR1/S-PLUS, e um artigo aceito para publicação sobre os resultados obtidos com os dados miniJPAS; participação em evento que resultou em um *proceeding*.

Publicados:

1. DOUBRAWA, L; CYPRIANO, E S; FINOGUENOV, A; LOPES, P A A; MATURI, M; GONZALEZ, A H; DUPKE, R. Galaxy cluster optical mass proxies from probabilistic memberships. **Monthly Notices of the Royal Astronomical Society**, v. 526, p. 4285-4295, 2023
2. MATURI, M.; FINOGUENOV, A.; LOPES, P. A. A.; GONZÁLEZ DELGADO, R. M.; DUPKE, R. A.; CYPRIANO, E. S.; CARRASCO, E. R.; DIEGO, J. M.; PENNA-LIMA, M.; DOUBRAWA, L; VÍLCHEZ, J. M.; MOSCARDINI, L.; MARRA, V.; BONOLI, S.; RODRÍGUEZ-MARTÍN, J. E.; ZITRIN, A.; MÁRQUEZ, I.; HERNÁN-CABALLERO, A.; et al. The miniJPAS survey. **Astronomy & Astrophysics**. , v.678, p.A145 - , 2023.
3. WERNER, S V ; CYPRIANO, E S ; GONZALEZ, A H ; DE OLIVEIRA, C MENDES ; ARAYA-ARAYA, P ; DOUBRAWA, L; DE OLIVEIRA, R LOPES ; LOPES, P A A ; VITORELLI, A Z ; BRAMBILA, D ; COSTA-DUARTE, M ; TELLES, E ; KANAAN, A ; RIBEIRO, T ; SCHOENELL, W ; GONÇALVES, T S ; MENÉNDEZ-DELMESTRE, K ; BOM, C R ; NAKAZONO, L. S-PLUS DR1 galaxy clusters and

groups catalogue using PzWav. **Monthly Notices of the Royal Astronomical Society**, v. 519, p. 2630-2645, 2022.

4. BONOLI, S.; MARÍN-FRANCH, A.; VARELA, J.; VÁZQUEZ RAMIÓ, H.; ABRAMO, L. R.; CENARRO, A. J.; DUPKE, R. A.; VÍLCHEZ, J. M.; CRISTÓBAL-HORNILLOS, D.; GONZÁLEZ DELGADO, R. M.; HERNÁNDEZ-MONTEAGUDO, C.; LÓPEZ-SANJUAN, C.; MUNIESA, D. J.; CIVERA, T.; EDEROCLITE, A.; HERNÁN-CABALLERO, A.; MARRA, V.; BAQUI, P. O.; CORTESI, A.; CYPRIANO, E. S.; DAFLON, S.; DE AMORIM, A. L.; DÍAZ-GARCÍA, L. A.; DIEGO, J. M.; MARTÍNEZ-SOLAEICHE, G.; et al. The miniJPAS survey: A preview of the Universe in 56 colors. **Astronomy & Astrophysics.** , v.653, p.A31 - , 2021.

Aceito para publicação:

1. DOUBRAWA, L; CYPRIANO, E S; FINOGENOV, A; LOPES, P A A; MATURI, M; GONZALEZ, A H; DUPKE, R; GONZÁLEZ DELGADO, R M; BONOLI, S; HERNÁN-CABALLERO, A; ABRAMO, R; BENITEZ, N; CARNEIRO, S; CENARRO, J; CRISTÓBAL-HORNILLOS, D; EDEROCLITE, A; LÓPEZ-SANJUAN, C; MARÍN-FRANCH, A. MENDES DE OLIVEIRA, C; MOLES, M; SODRÉ JR., L; TAYLOR, K; VARELA, J; VÁZQUEZ RAMIÓ, H . The miniJPAS survey: Optical detection of galaxy clusters with PZWav. **Astronomy & Astrophysics.**, aceito em 14 de Fevereiro de 2024.

Proceedings:

1. DOUBRAWA, L; CYPRIANO, E S ; FINOGENOV, A. . Galaxy cluster finding and membership assignments in the mini J-PAS survey, 2023. **Boletim da Sociedade Astronomica Brasileira**, 2023. v. 34. p. 214-218.