



S-PLUS

Southern Photometric  
Local Universe Survey

# Search for compact dwarf galaxies in the Fornax cluster

Pierre Augusto Ré Martho, Laerte Sodré Jr.  
Instituto de Astronomia, Geofísica e Ciências Atmosféricas  
30 de junho de 2025



# Visão geral

---

# Visão geral

- Introdução
  - ▶ Contexto cosmológico
  - ▶ Estrutura em larga escala
  - ▶ Redshifts espectroscópicos e fotométricos
- Objetivo
- Dados
  - ▶ Fotométricos e espectroscópicos
  - ▶ Pré-processamento
  - ▶ Pesos por objeto
- Metodologia
  - ▶ Redes neurais
  - ▶ Redes neurais Bayesianas
  - ▶ Redes de mistura de densidades
- Resultados
  - ▶ Redshifts fotométricos
  - ▶ Funções de densidade de probabilidade
  - ▶ Estrutura em larga escala
- Conclusões e perspectivas

# Introdução

---

# O contexto cosmológico

Estudar a formação e evolução da estrutura em larga escala (LSS) é fundamental para entender o Universo em que vivemos. A partir dela podemos ter insights sobre:

- Matéria escura
- Energia escura
- Evolução cósmica
- Formação e evolução de galáxias
- Física a nível fundamental

# Redshifts espectroscópicos e suas limitações

Uma forma de determinar a distância de objetos celestes vêm da lei de Hubble Lemaître (válida para objetos próximos):

$$v_{res} = c \cdot z = H_0 \cdot D$$

Velocidade de recessão [km/s]      Redshift      Distância [Mpc]

Velocidade da luz [km/s]      Cte. Hubble [km/s/Mpc]

A observação de um espectro com alto sinal ruído demanda bastante tempo de observação, ainda mais para objetos fracos

É necessário encontrar uma alternativa

# A alternativa: redshifts fotométricos

Existem uma série de projetos que se baseiam em fotometria para produzir ciência.

- Pode ser considerada uma "aproximação" do espectro
- É obtida de forma muito mais rápida
- Pode alcançar magnitudes mais fracas
- É capaz de gerar dados para muitos objetos e grandes áreas

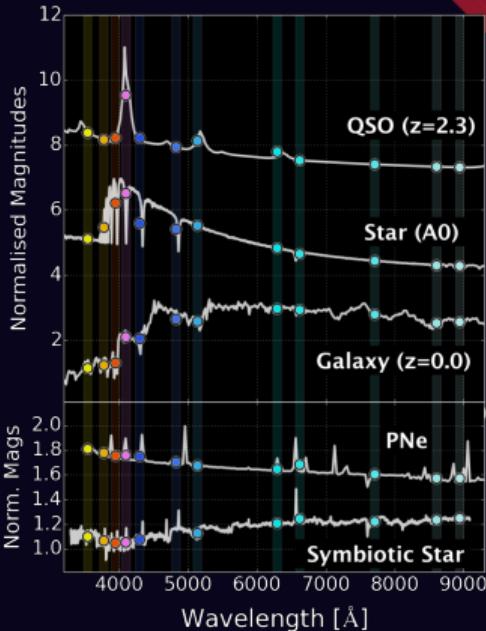


Figura 1: Mendes de Oliveira et al. (2019).

# A alternativa: redshifts fotométricos

Redshifts fotométricos são estimados por três tipos de algoritmos: aprendizado de máquina, ajuste de templates e códigos híbridos.

## Aprendizado de máquina (ML)

- ✓ Se baseiam no uso de uma amostra de treinamento
- ✓ Flexível em relação a modelos (RFs, KNNs, SVMs)
- ✓ Podem ser mais precisos e rápidos que modelos de ajuste de template
- ✗ Não fornecem uma classificação do objeto (exceto caso seja configurado para isso)
- ✗ Sujeito à vieses devido aos dados

## Ajuste de templates (TF)

- ✓ Fazem uma comparação entre a fotometria de um objeto e uma biblioteca de templates
- ✓ Fornecem uma classificação do objeto junto ao  $z_{\text{phot}}$
- ✓ Maior capacidade de extração
- ✗ Menos precisos, mais lentos
- ✗ Sujeito à vieses devido à escolha dos templates

# O objetivo

---

# O objetivo

## Redshifts fotométricos

- Determinação de redshifts fotométricos de alta precisão
- Funções de densidade de probabilidade bem calibradas
- Galáxias até  $z = 0.8$  e magnitude 21 na banda r

## Estrutura em larga escala

- Utilizar dados fotométricos para a reconstrução da LSS
- Obter um mapeamento similar ao visto usando  $z_{\text{spec}}$

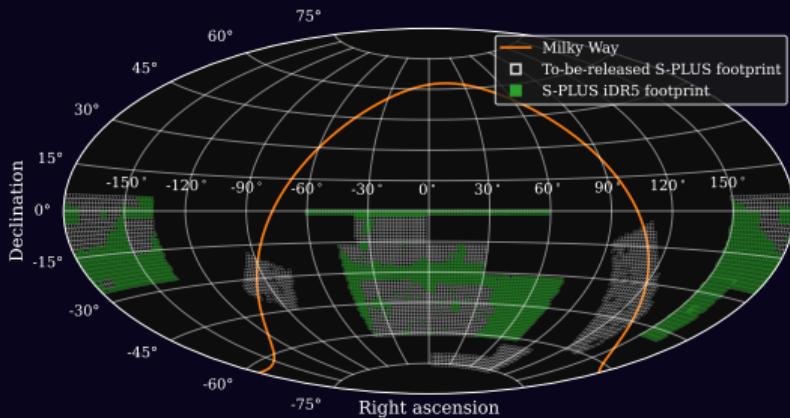
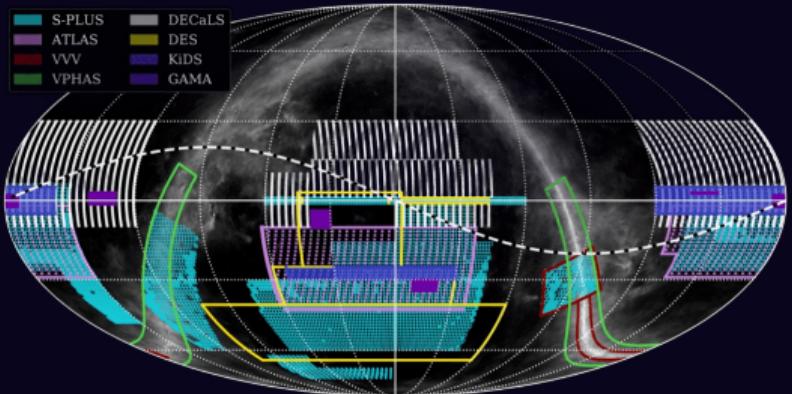
Expandir o conjunto de dados que podemos usar para estudos em diferentes áreas da astronomia

# Dados

---



# Dados fotométricos

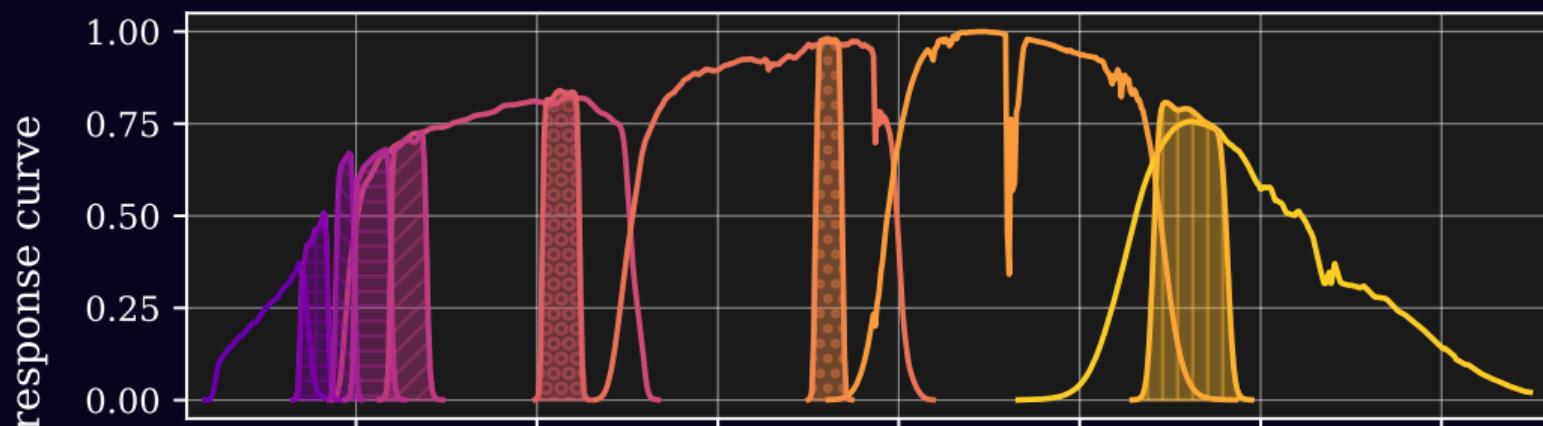


# Dados fotométricos

DR4: 3022,7 square degrees,

1629 Fields (each ~2 square degrees)

## Filters

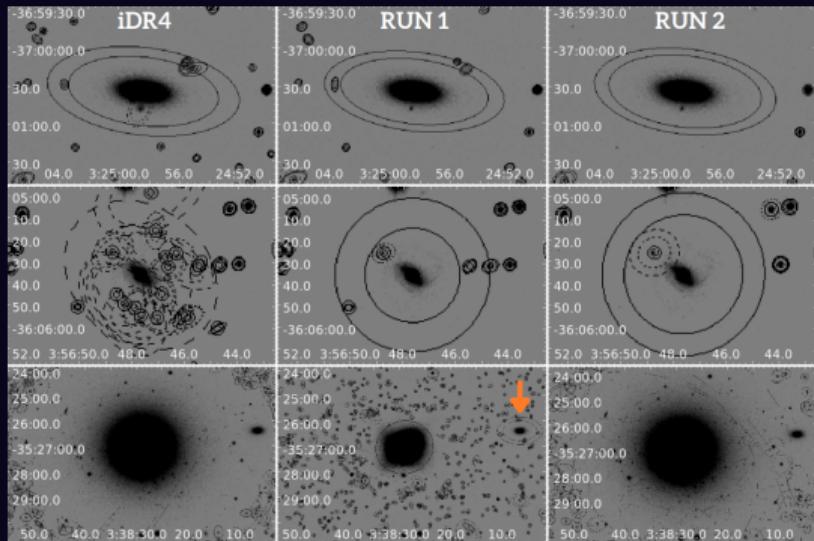


# Dados fotométricos

## Fornax

**Run 1:** Faint objects detected near bright galaxies at Fornax distance

**Run 2:** Better characterizes larger and brighter galaxies



## Extinction correction

- Galaxy dust affects photometric measurements. Python dustmaps (Green, 2018)
- Calculation of Extinction Coefficients. Python extinction (Barbary 2016)

Figura 2: Haack,R.F.(2024)-<https://arxiv.org/pdf/2404.10847>

# Cortes de qualidade na fotometria

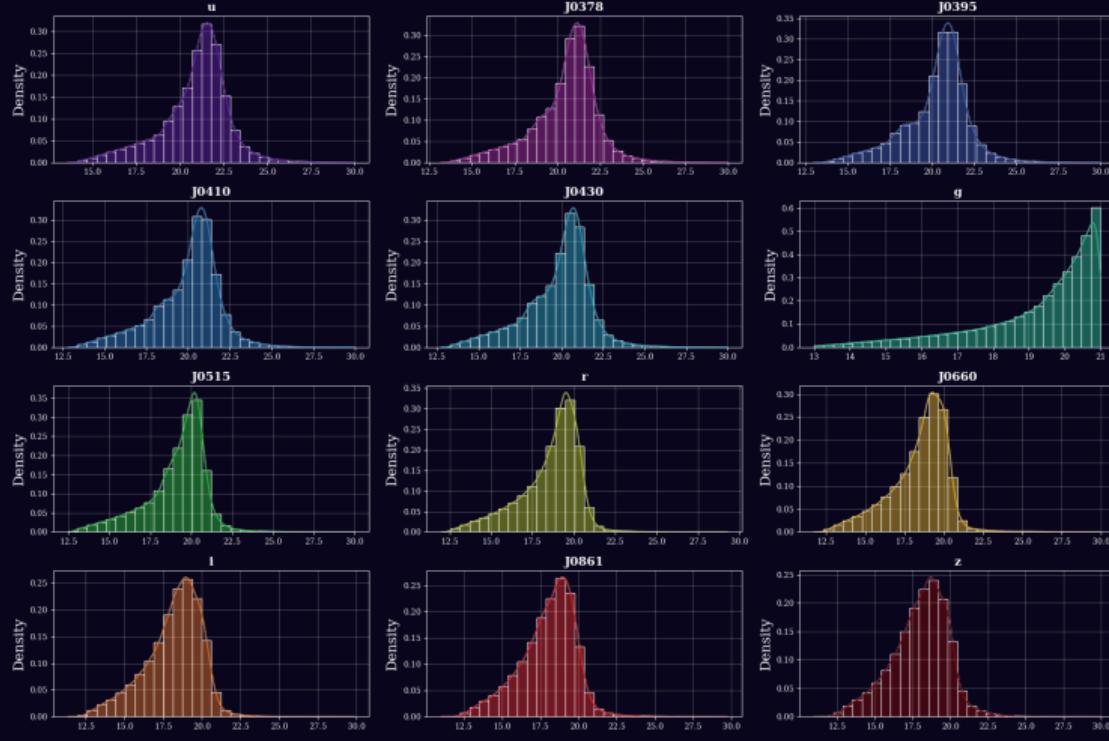
## Critérios de seleção

- Magnitudes maiores que 30 foram descartadas (baixa qualidade e erros altos).
- Objetos com  $\text{flag0} < 3$  também foram descartados.
- Banda g\_APER\_6 como referência principal:
  - ▶ Corte inferior:  $\text{g\_APER\_6} > 13$  (evitar saturação).
  - ▶ Corte superior:  $\text{g\_APER\_6} < 21$  (minimizar contaminação por aglomerados globulares).

## Resultados

- Total inicial: 2.9 milhões de objetos.
- Após cortes: 619.630 objetos restantes.

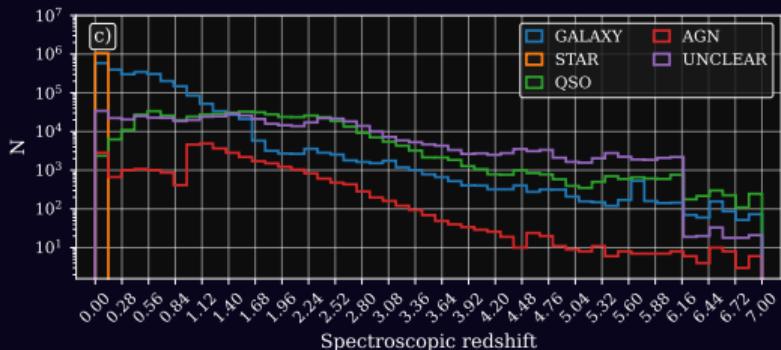
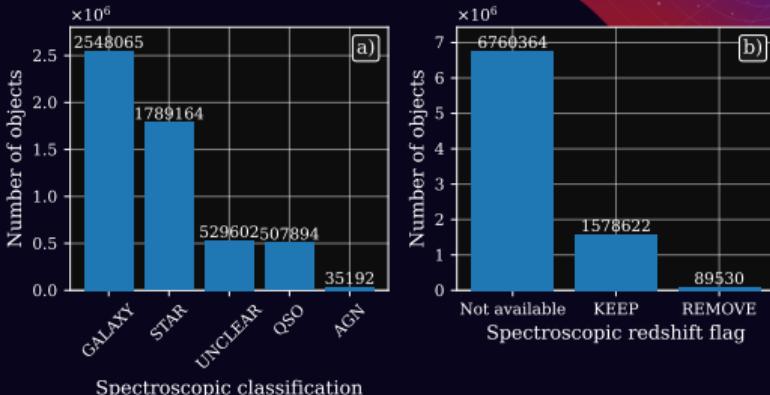
# Cortes de qualidade na fotometria



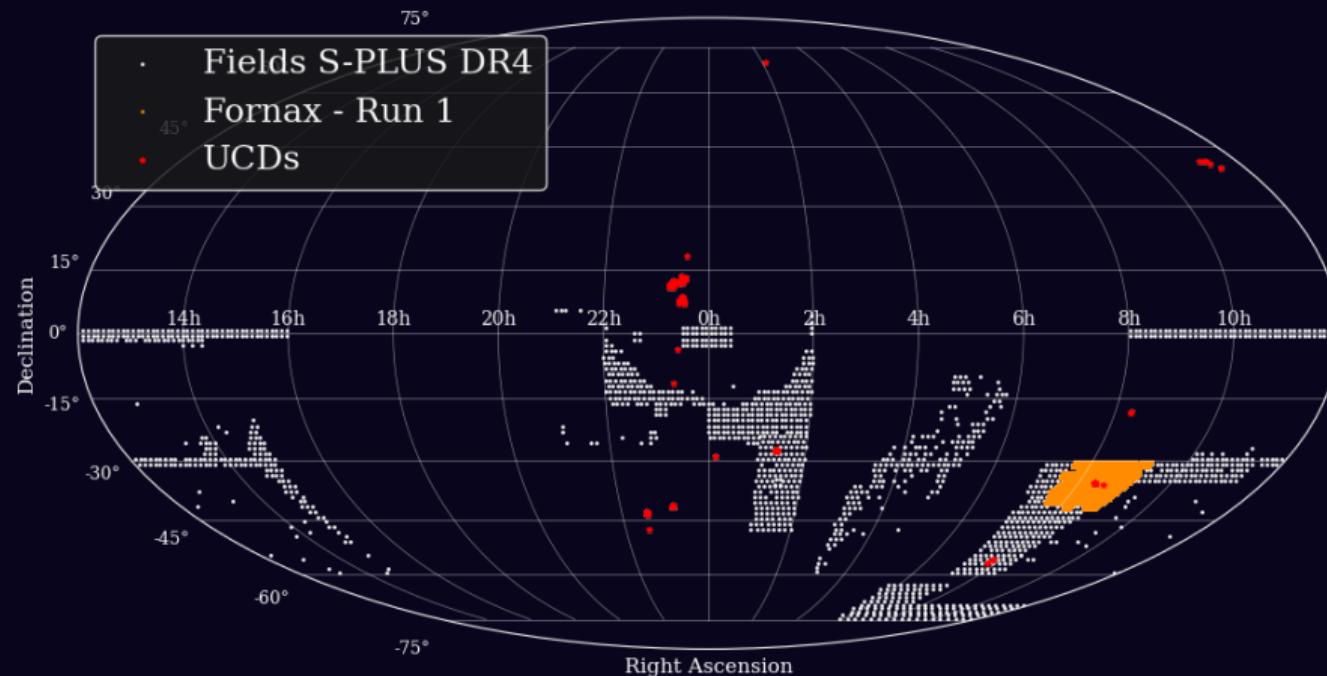
# Dados espectroscópicos ([https://github.com/ErikVini/specz\\_compilation](https://github.com/ErikVini/specz_compilation))

## Informações do compilado

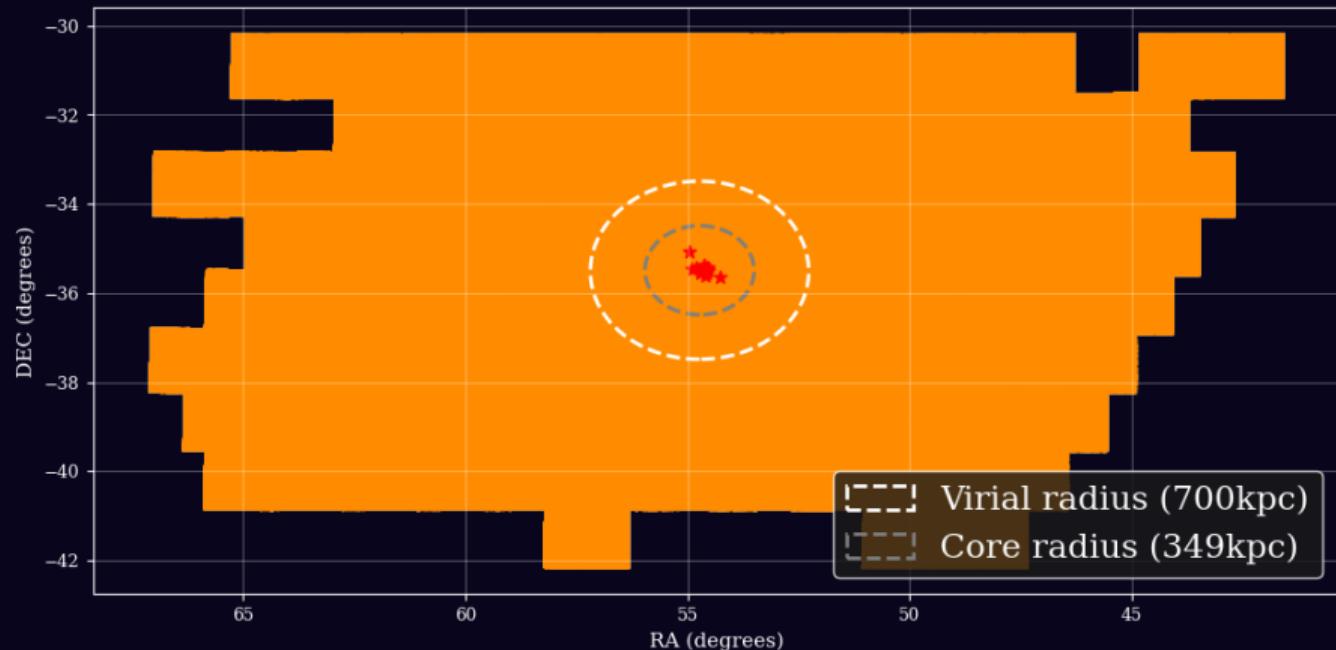
- Total de catálogos: 5097
- Catálogos usados: 1872
- Total de objetos: 8 437 460
- Pré era GPT!



# Distribuição das UCDs em Fornax

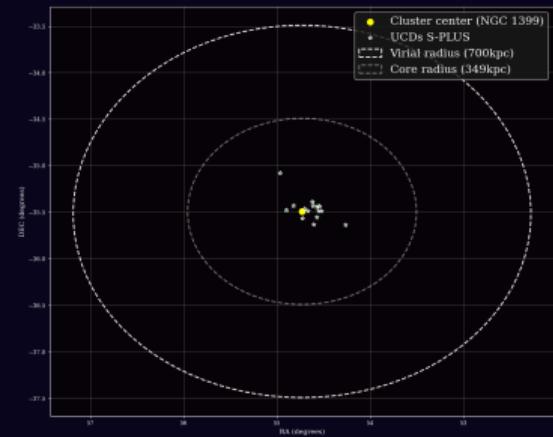
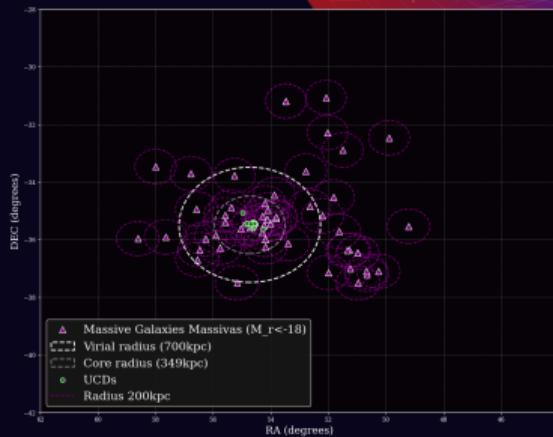


# Distribuição das UCDs em Fornax



# Distribuição das UCDs em Fornax

Parâmetro	Valor
Massa ( $M_{\odot}$ ) <sup>1</sup>	$7 \pm 2 \times 10^{13}$
Raio Virial ( $Mpc$ ) <sup>1</sup>	0.7
Raio Virial (graus) <sup>1</sup>	2
Raio interno ( $Mpc$ ) <sup>4</sup>	0.349
Raio interno (graus) <sup>4</sup>	1
$\sigma_v$ ( $km s^{-1}$ ) <sup>3</sup>	318
Módulo de distância (mag) <sup>2</sup>	31.51



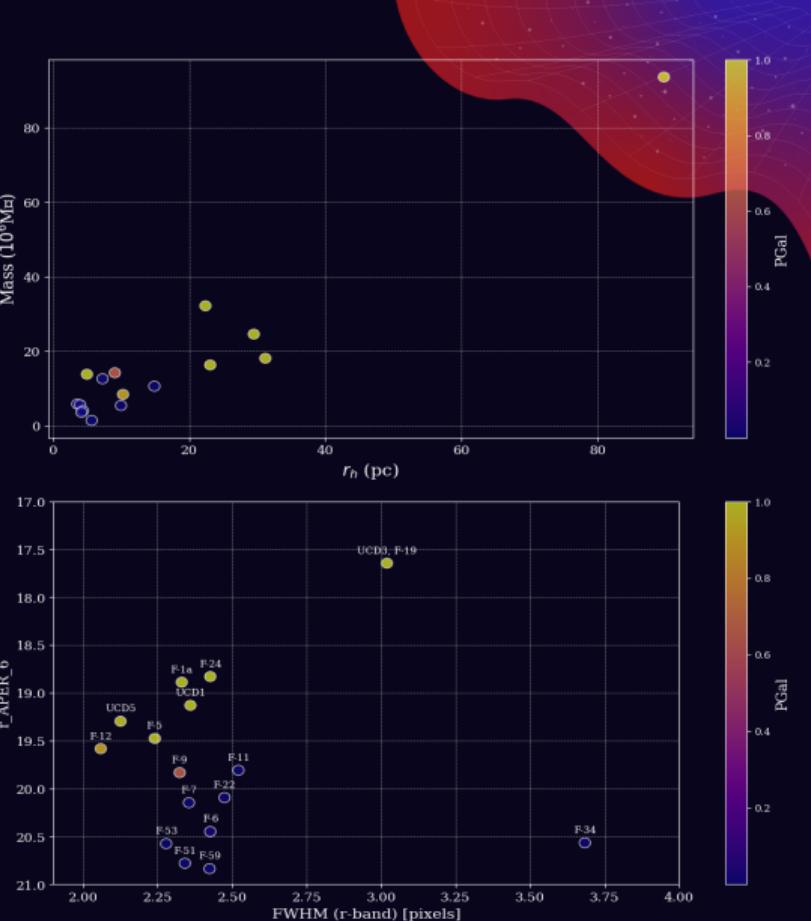
# Propriedades das UCDs

Tabela 1: Propriedades das UCDs de Mieske et al.(2008).

Nome	Massa Total ( $10^6 M_{\odot}$ )	$M/L_V$	[Fe/H] (dex)	$r_h$ (pc)	$\sigma$ (km/s)
UCD3, F-19	$93.6 \pm 14.0$	$4.69 \pm 0.70$	-0.4	89.7	22.8
UCD1	$32.1 \pm 3.9$	$4.99 \pm 0.60$	-0.7	22.4	27.1
F-24	$24.5 \pm 7.8$	$3.44 \pm 1.10$	-0.4	29.5	21.4
UCD5	$18.0 \pm 4.5$	$3.37 \pm 0.85$	-1.2	31.2	18.7
F-1a	$16.2 \pm 3.8$	$2.45 \pm 0.58$	0.0	23.1	18.7
F-9	$14.1 \pm 3.6$	$4.72 \pm 1.20$	-0.8	9.1	25.7
F-5	$13.7 \pm 2.4$	$3.16 \pm 0.55$	-0.3	5.0	34.5
F-6	$12.5 \pm 2.4$	$5.32 \pm 1.00$	0.2	7.3	27.3
F-7	$10.5 \pm 1.4$	$4.21 \pm 0.57$	-1.3	14.9	20.1
F-12	$8.3 \pm 2.9$	$2.36 \pm 0.83$	-0.4	10.3	22.9
F-11	$5.7 \pm 3.7$	$1.64 \pm 1.10$	-0.9	3.6	26.2
F-34	$5.5 \pm 1.3$	$3.17 \pm 0.74$	-0.9	4.0	24.6
F-22	$5.3 \pm 1.0$	$2.13 \pm 0.39$	-0.4	10.0	22.8
F-53	$3.9 \pm 1.0$	$2.66 \pm 0.69$	-0.9	4.4	19.6
F-51	$3.5 \pm 0.9$	$2.38 \pm 0.62$	-0.8	4.2	20.1
F-59	$1.3 \pm 0.6$	$0.94 \pm 0.43$	-2.1	5.7	9.8

# Propriedades das UCDs

Nome	spec - z	<i>g</i>	zml
UCD3	0.0053	18.47	0.03
UCD1	0.0052	19.75	0.08
F-24	0.0063	19.66	0.04
UCD5	0.0045	19.71	0.04
F-1a	0.0042	19.66	-
F-9	0.0058	20.85	0.07
F-5	0.0057	20.54	-
F-6	0.0037	20.48	-
F-7	0.0050	20.89	0.16
F-12	0.0055	20.37	-
F-11	0.0059	20.40	-
F-34	0.0054	20.79	-
F-22	0.0034	20.69	0.06
F-53	0.0020	21.57	-
F-51	0.0041	22.23	-
F-59	0.0061	21.47	-



# Aprendizado de máquina

---



# Procura de UCDs com Aprendizado de Máquina

## Objetivo

- Identificar candidatas a galáxias ultra-compactas no aglomerado de Fornax.
- Classificar objetos em duas categorias principais:
  - ▶ Compactos.
  - ▶ Extensos.
- Separar objetos com base em suas características morfológicas e fotométricas.

**Identificar dentro do grupo de objetos compactos, aqueles que possuem características fotométricas semelhantes às de galáxias extensas.**

# Amostra de Treino e Valores Faltantes

## Amostra de Treino

- Divisão em classes:
  - ▶ Classe 0: Compactos.
  - ▶ Classe 1: Extensos.
- Total de objetos: 545.267.
- Treinamento: 80% dos dados.
- Teste: 20% dos dados.

## Valores Faltantes

- Imputação com MICE (Método de Imputação Múltipla).
- 27% dos objetos possuem pelo menos 1 valor faltante.
- Dados faltantes tratados para garantir consistência.

# Distribuição de Valores Faltantes

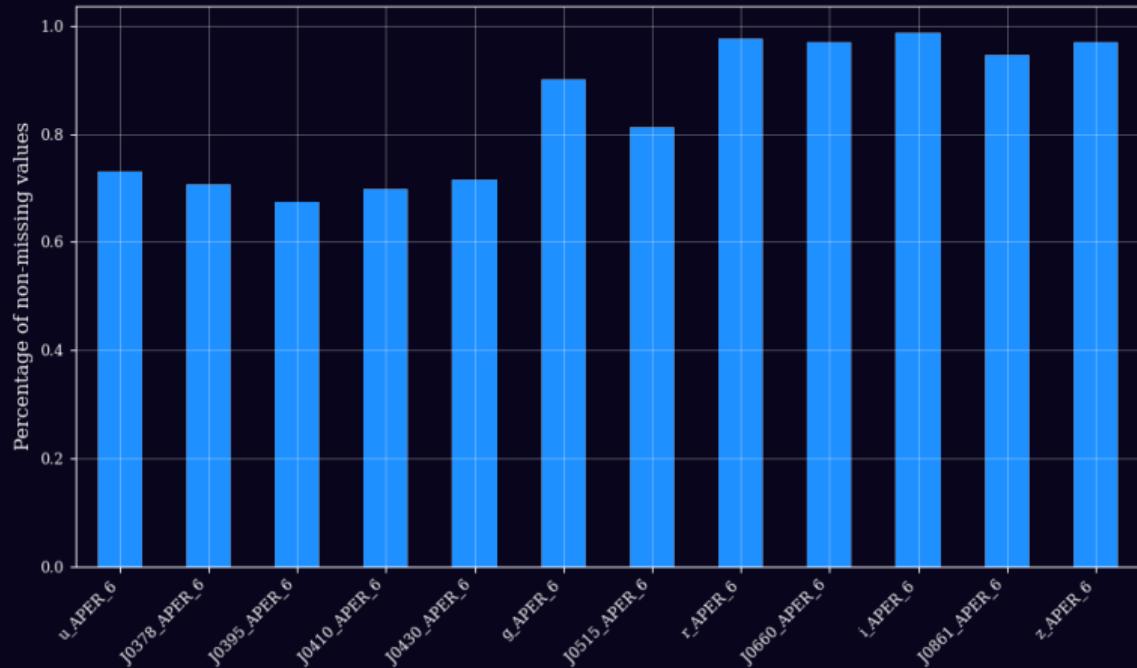
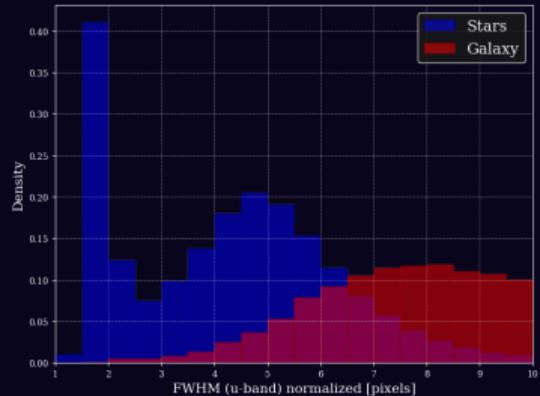
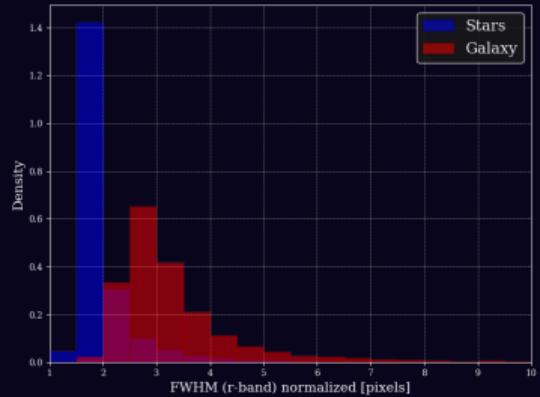


Figura 3: Distribuição de valores faltantes.

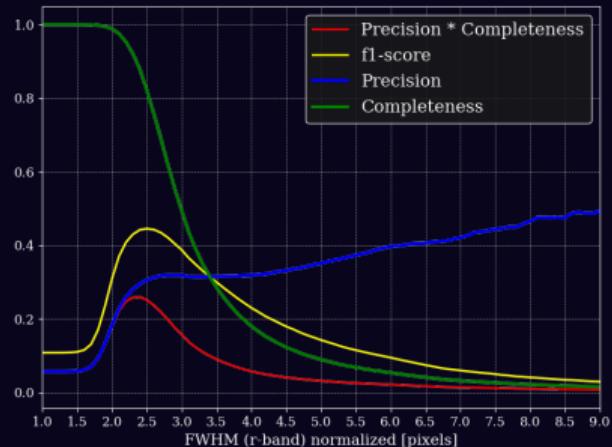
# Amostra de treino



- **Precisão:**  $\frac{TP}{TP+FP}$

- **Completeness:**  $\frac{TP}{TP+FN}$

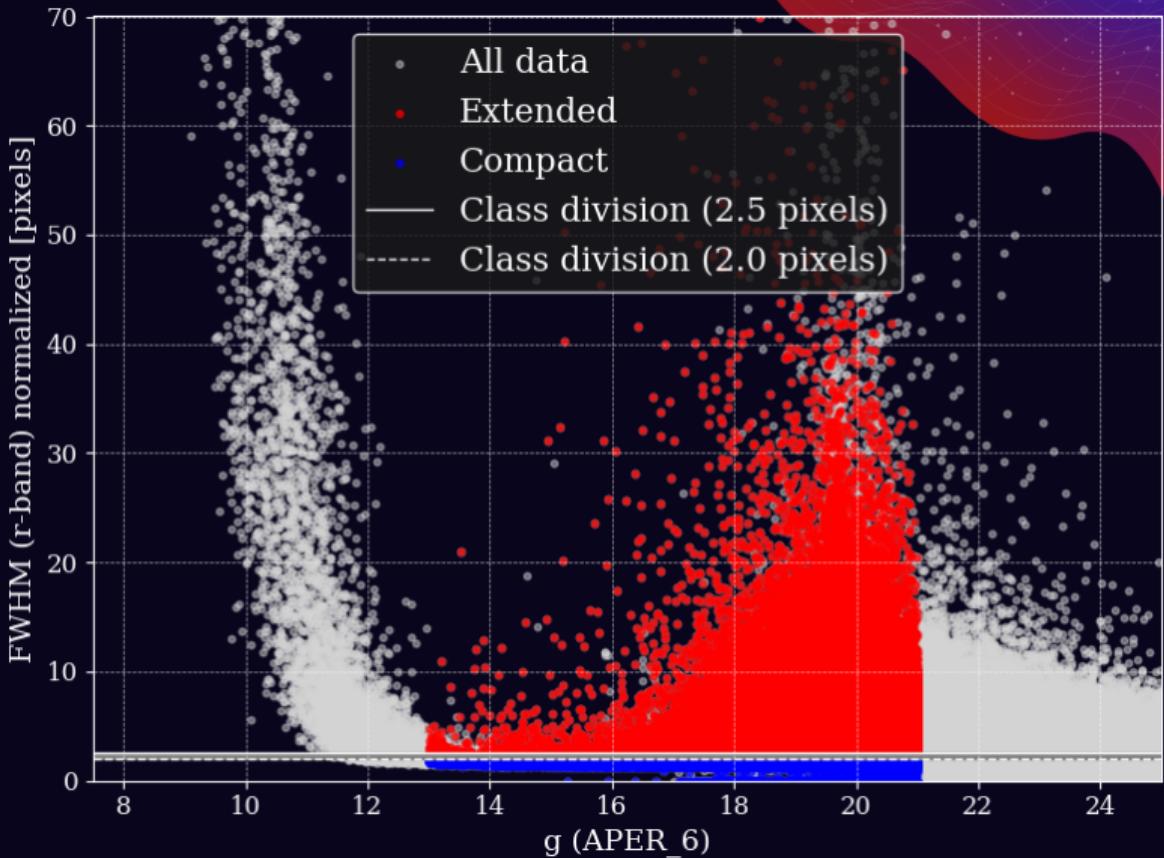
- **F1-Score:**  $2 \cdot \frac{\text{Precisão} \cdot \text{Completeness}}{\text{Precisão} + \text{Completeness}}$



# Amostra de treino

## Critérios de Classificação

- Compactos:  $\text{FWHM} < 2 \text{ pixels}$ .
- Extensos:  $\text{FWHM} > 2.5 \text{ pixels}$ .
- Dados entre 2 e 2.5 pixels não usados no treinamento.



# Amostra de Treino

## Divisão da Amostra

- Total de objetos: 545.267.
- Treinamento: 80% (436.213 objetos).
- Teste: 20% (109.054 objetos).
- Classe 0 (compactos): 242.085 no treino, 60.522 no teste.
- Classe 1 (extensos): 194.128 no treino, 48.532 no teste.

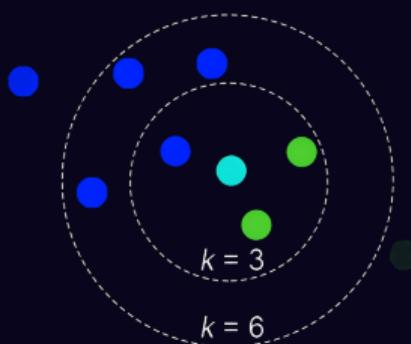
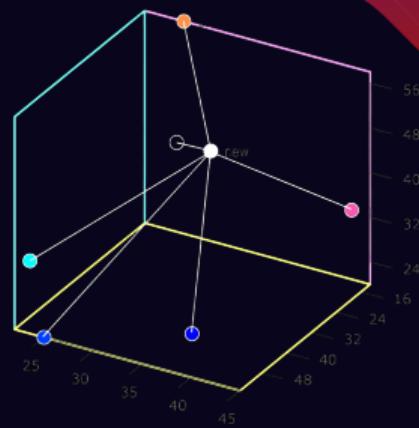
## Parâmetros Utilizados

- 12 magnitudes corrigidas pela extinção (APER\_6).
- 66 combinações possíveis de cores.
- Dados entre 2 e 2.5 pixels (FWHM) não usados no treinamento.

# Classificador KNN

## K-Nearest Neighbors (KNN)

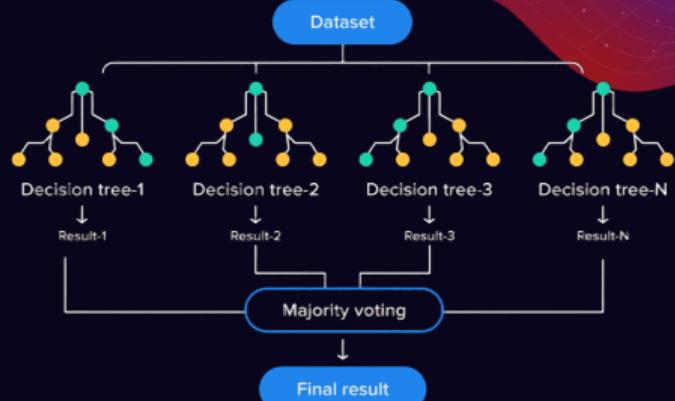
- Algoritmo supervisionado que classifica objetos com base nos vizinhos mais próximos.
- Simples e eficiente para conjuntos de dados menores.
- Utiliza a distância entre objetos no espaço de parâmetros para classificação.



# Classificador RF

## Random Forest (RF)

- Algoritmo supervisionado baseado em múltiplas árvores de decisão.
- Combina os resultados de várias árvores para melhorar a precisão.
- Robusto contra overfitting em muitos casos.
- Eficiente para conjuntos de dados grandes e com muitas features.



# Análise dos Classificadores

- Acurácia:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Precisão:

$$\frac{TP}{TP + FP}$$

- Completeness:

$$\frac{TP}{TP + FN}$$

- F1-Score:

$$2 \cdot \frac{\text{Precisão} \cdot \text{Completeness}}{\text{Precisão} + \text{Completeness}}$$

- **AUC-ROC**: Mede sensibilidade vs. taxa de falsos positivos.

- **MCC**:

$$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Real	Sim (Detectada)	Não (Detectada)
Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Tabela 2: Matriz de Confusão.

# Análise dos Classificadores

Tabela 3: Classificação binária - Métricas modelo KNN

Classe	Precisão	Compleza	F1-Score
0	0.92	0.88	0.90
1	0.86	0.91	0.88
AUC-ROC		0.95	
Coeficiente de Correlação de Matthews (MCC)		0.78	

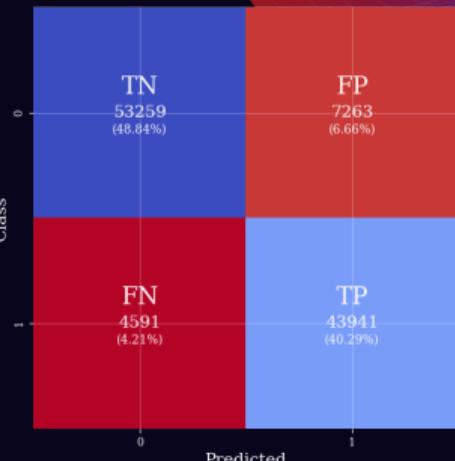
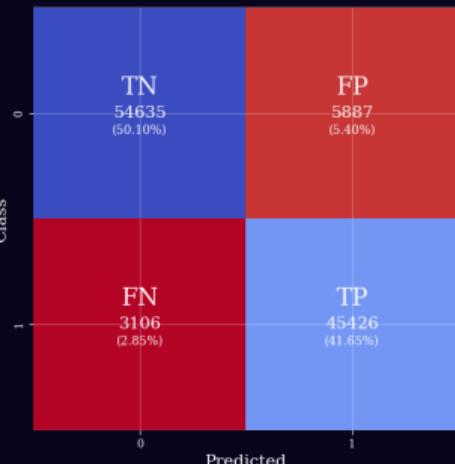
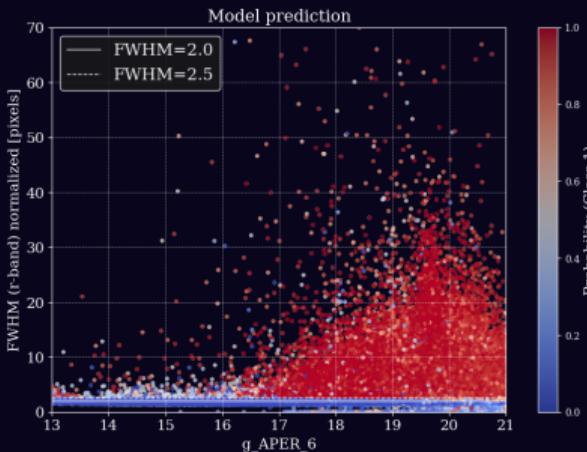
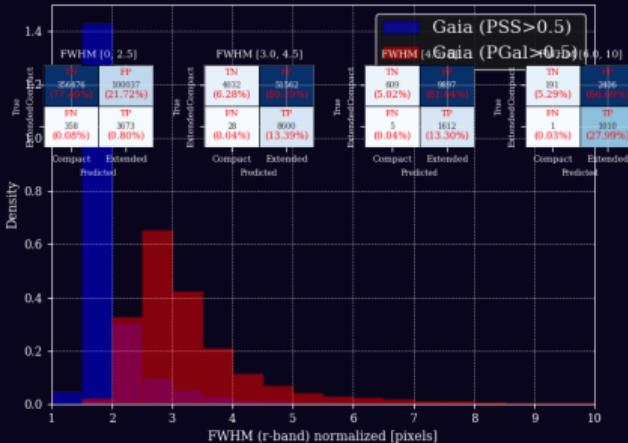


Tabela 4: Classificação binária - Métricas modelo RF

Classe	Precisão	Compleza	F1-Score
0	0.95	0.90	0.92
1	0.89	0.94	0.91
AUC-ROC		0.97	
Coeficiente de Correlação de Matthews (MCC)		0.84	



# Análise dos Classificadores



Nome	RF Predição
UCD3	0.36
UCD1	0.98
F-24	0.99
UCD5	0.38
F-1a	0.11
F-9	0.87
F-5	0.53
F-6	0.89
F-7	0.70
F-12	0.73
F-11	0.97
F-34	0.98
F-22	0.93
F-53	0.96
F-51	0.96
F-59	0.92

- **Total de objetos:** 1.803.561.
- **Extensos:** 1.411.803 (78,28%). Deles 311.846 (17,29%) com  $FWHM < 2.5$  pixels.

# Redshifts Fotométricos

Redshifts fotométricos são estimativas baseadas em fotometria multibanda.

$$v_{\text{res}} = c \cdot z = H_0 \cdot D,$$

(1)

## Modelo Treinado

- Photo-z de *Lima et al. 2022*.
- De 29000926, **290637** sem estimativa.
- $0.002 \leq z \leq 0.5$ ;  $15 \leq r_{\text{APER\_6}} \leq 21$ .
- Amostra final: 12.296 objetos.
- Treinamento com 66 cores (APER\_6).
- Divisão da amostra: 80% treino, 20% teste.
- Regressão com Random Forest (RF).

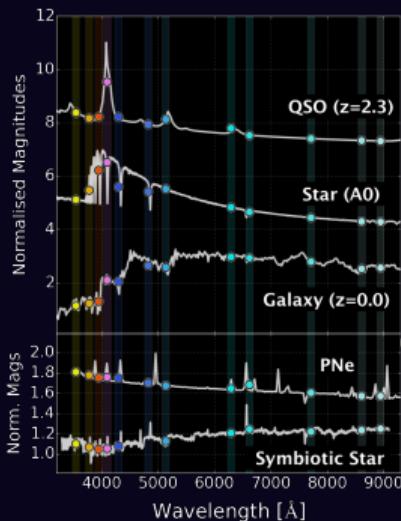


Figura 4: Adaptado de Mendes de Oliveira et al. (2019).

# Redshifts Fotométricos

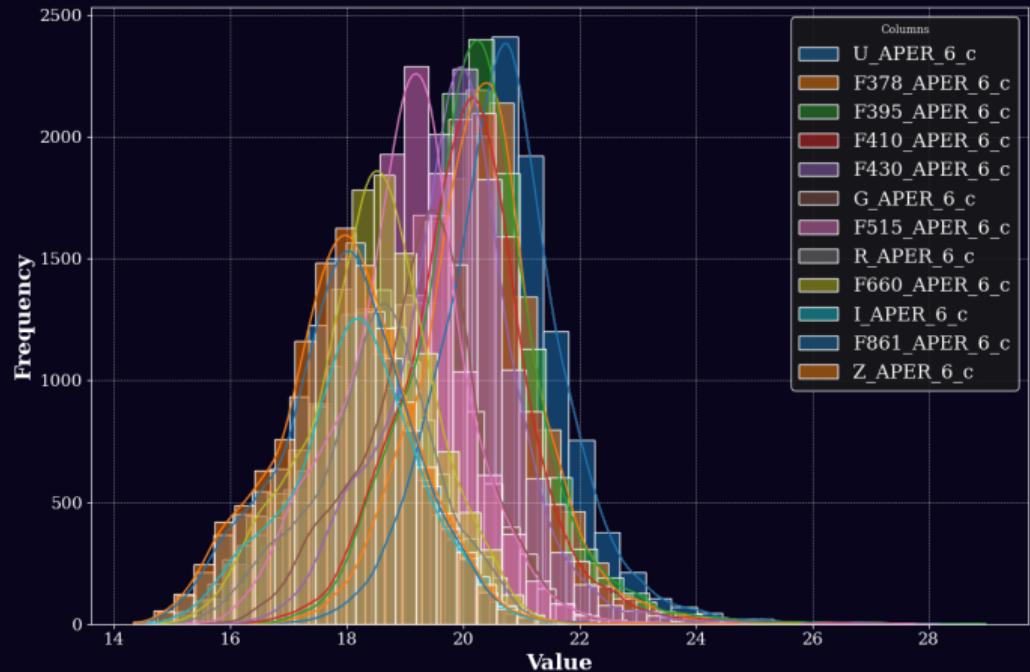
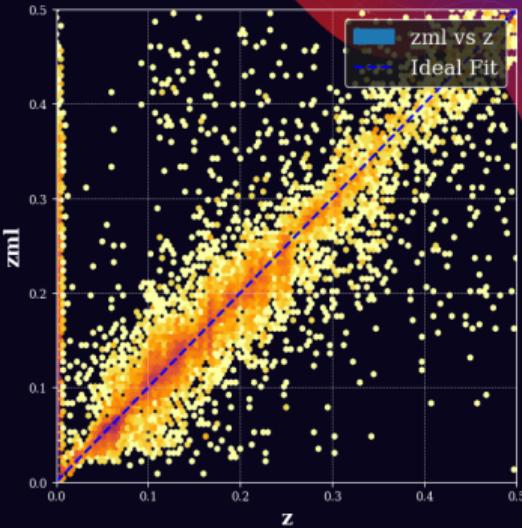
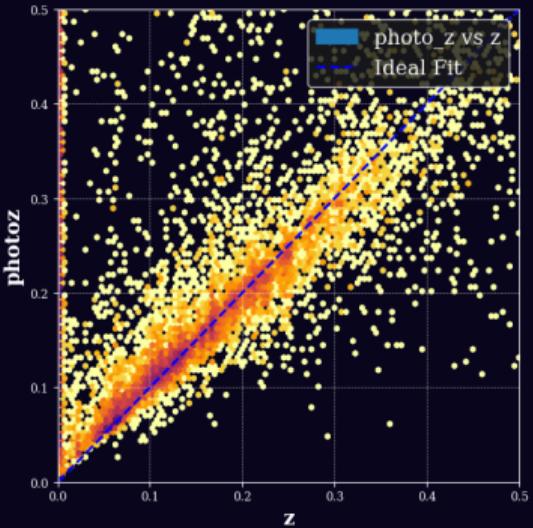
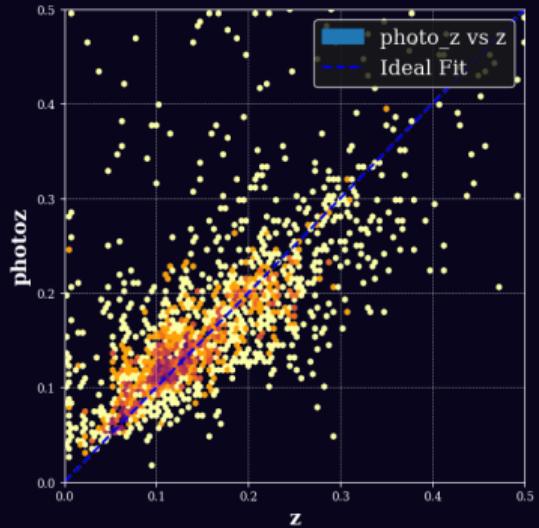


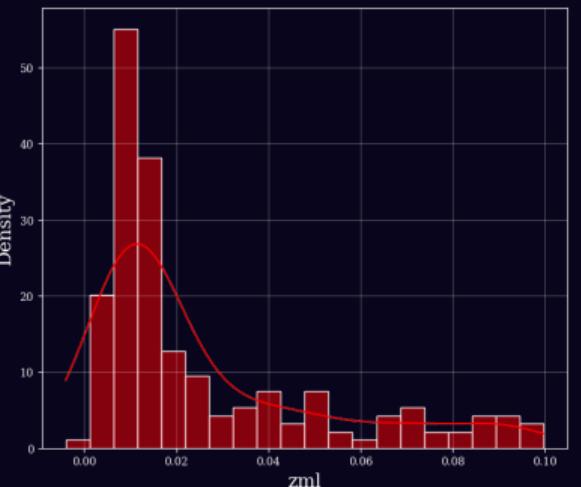
Figura 5: Distribuição magnitudes na amostra de treino *photo\_z*.

# Redshifts Fotométricos



- $EQR = 0.05$
- $R^2 = 0.6$
- $\sigma_{NMAD} = 0.03$

# Redshifts Fotométricos



Nome	$z_{phot}$	$z_{ml}$	$z_{spec}$
UCD3	0.07	0.03	0.0053
UCD1	0.09	0.08	0.0052
F-24	0.08	0.04	0.0062
UCD5	0.03	0.04	0.0045
F-1a	0.21	--	0.0042
F-9	0.09	0.07	0.0058
F-5	0.06	--	0.0057
F-6	0.10	--	0.0037
F-7	0.19	0.16	0.0050
F-12	0.07	--	0.0055
F-11	0.10	--	0.0059
F-34	0.07	--	0.0054
F-22	0.09	0.06	0.0034
F-53	0.31	--	0.0020
F-51	0.10	--	0.0041
F-59	0.06	--	0.0060